UNIVERSITY OF COPENHAGEN
NIELS BOHR INSTITUTE

COSMOS2020

Insights into Galaxy Assembly and Evolution over the First 10 Billion Years

John R. **Weaver**

DAWN

UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE
NIELS BOHR INSTITUTE

COSMOS2020

Insights into Galaxy Assembly and Evolution over the First 10 Billion Years

JOHN RAYMOND WEAVER IV

A dissertation submitted for the degree of
Doctor of Philosphy (PhD)
Cosmic Dawn Center, Niels Bohr Institute
Faculty of Science, Københavns Universitet

Academic Advisor:

Prof. Sune Toft

Assessment Committee:

Prof. Danilo Marchesini

Prof. Karina I. Caputi

Prof. Pascal Oesch

Autumn 2022

In nova fert animus mutatas dicere formas
corpora; di, coeptis (nam vos mutastis et illas)
adspirate meis primaque ab origine mundi
ad mea perpetuum deducite tempora carmen

You heavenly powers, since you were responsible for those changes, as for all else, look favourably on my attempts, and spin an unbroken thread of verse, from the earliest beginnings of the world, down to my own times.

- Ovid, *Metamorphoses* (*1955, 29*)

Cover image credits:
ESO/H. J. McCracken
Tycho Brahe (1598), *Astronomiae instauratae mechanica*

ABSTRACT

Long before the invention of the telescope, early humans looked up and marvelled at the beauty of the night sky. Our inescapable curiosity has led us to chart those distant points of light, to measure their properties, and understand them through physical law. This cosmic cartography was revolutionized in the 16th century with the first scientific, and in many senses modern, star catalogs built by Tycho Brahe from his observatory in the middle of the windswept Danish straits.

Then almost four centuries later in 1995, the Hubble Space Telescope stared deep into the abyss and captured the light of thousands of galaxies seen as they were in their infancy, when our universe was only a fraction of its current age. This was a watershed moment, and since then, astronomers have mapped successively larger and deeper regions of our universe.

Such is the COSMOS survey, observed over a region of the sky a bit larger than the full moon by some of the foremost telescopes and studied by a worldwide collaboration of astronomers. Through its discoveries, COSMOS has earned its place as a cornerstone of modern galaxy evolution studies.

This dissertation is centered around the COSMOS2020 catalogs, our latest cartographic effort to measure these increasingly deeper and complex images. New innovations in measurement techniques developed for this task have proven their worth: instead of measuring the light of galaxies in circular apertures, whole models are fitted that produce more accurate measures of their brightness, masses, and distances than ever before. We have used these new catalogs to measure how galaxies assemble and transform over 75% of cosmic history.

To our surprise, we have discovered a handful of ultra-luminous galaxies seen only 600 million years after the big bang — cosmic beasts — whose incredible mass and maturity defies our understanding of how galaxies form and evolve. Soon, their mysterious origins will be revealed by some of the first observations by the James Webb Space Telescope.

DANSKE RESUME

Længe før opfindelsen af teleskopet kiggede mennesket op og beundrede nattehimlens skønhed. Vores uopsættelige nysgerrighed har fået os til at kortlægge de fjerne lyspunkter, måle deres egenskaber og forstå dem ved hjælp af fysik. Denne kosmiske kartografi blev revolutioneret i det 16. århundrede med de første videnskabelige, og på flere måder moderne, stjernekataloger konstrueret af Tycho Brahe fra hans observatorium på Hven, midt i det vindomsuste danske Øresund. Næsten fire århundreder senere, i 1995, stirrede rumteleskopet Hubble ud i dybet og fangede lyset fra tusinder af galakser, set som de var i deres spæde ungdom da vores Univers kun var en brøkdel af dets nuværende alder. Dette var et skelsættende øjeblik, og siden da har astronomer kortlagt stadig større og dybere områder af vores Univers.

En sådan kortlægning er det observationelle program COSMOS, som består af snesevis af dybe billeder over et område på himlen lidt større end fuldmånen, observeret af et verdensomspændende samarbejde mellem teleskoper og astronomer. I kraft af dets opdagelser har COSMOS fortjent sin plads som en hjørnesten i moderne studier af galaksernes udvikling.

Denne afhandling er centreret omkring COSMOS2020-katalogerne, vores seneste kartografiske indsats for at måle disse stadig dybere og mere komplekse billeder. Nye innovationer inden for måleteknikker udviklet til denne opgave har bevist deres værd: I stedet for at måle lyset fra galakser i simple, cirkulære områder centreret på galakserne, har vi tilpasset modeller, der giver mere nøjagtige mål for deres lysstyrke, masser og afstande end nogensinde før. Vi har brugt disse nye kataloger til at måle, hvordan galakser dannes og udvikler sig henover 75% af Universets historie.

Til vores overraskelse har vi opdaget en håndfuld ultra-lysstærke galakser set kun 600 millioner år efter Big Bang — kosmiske monstre — hvis utrolige masse og udviklingstrin trodser vores forståelse af, hvordan galakser dannes og udvikler sig. Deres mystiske oprindelse vil snart blive afsløret af nogle af de første observationer fra rumteleskopet James Webb.

ACKNOWLEDGEMENTS

First and foremost, I would like to convey my gratitude to Sune Toft. You have been a source of endless enthusiasm and encouragement on good days and bad. Your unwavering support of my interests, curiosity, and independence is deeply appreciated, and your watchful guidance has kept me afloat. It has been said that luck is what happens when preparation meets opportunity; you supplied both and for that have my eternal thanks.

Of course, this work was not done in isolation. I wish to thank the the team of scientists who have braved these waters with us. To particular members of the Cosmic Dawn Center Family with whom I've had the fortune of working: Iary Davidzon, Peter Capak, Gabe Brammer, Georgios Magdis, Charles Steinhardt, Francesco Valentino, Seiji Fujimoto, Bo Milvang-Jensen, Peter Lausen, and Charlotte Mason. To the COSMOS team for their invaluable insight and support: Henry J. McCracken, Andrea Moneti, Olivier Ilbert, Caitlin Casey, Andreas Faisst, Olivier Kauffmann, and Marko Shuntov. And to the H20 team for their enthusiasm and vision: Dave Sanders, Bahram Mobasher, Istvan Szapudi, Nao Suzuki, Conor McPartland, Andrew Repp, Lukas Zalesky, Nima Chartab, and Sina Taamoli. Without your generous advice, mutual respect, and curiosity, this work would surely not have been possible. A special thanks goes out to Stephane Rouberol for keeping our CPUs on CANDIDE running through all the long tasks and buggy scripts we've put it through. I also want to thank Henry, Dave, and Peter in particular for hosting me during my visits abroad. Lastly, thank you Lukas, Andreas, Vasily, Natalie, and Nathaniel Stickley for being the best farmhands this farmer could have asked for; you've helped opened the (barn) door to the distant universe.

I would not be here if it were not for the mentors who taught me the tools of our trade and provided an opportune environment for scientific growth: Bernd Husemann, Michael Maseda, Stella Kafka, Michael West, and Michael Gregg. I owe a special thanks to Vivienne Wild for teaching me that curiosity is the heart of science, and that we should seek to answer questions rather than focus on the means to do so. Another special thanks is owed to Keith Horne for not only your endless patience over the last six years, but also for imparting an all too uncommon sense of rigorous statistical analysis. Thank you also to Johan Fynbo and Troels Petersen for inspiring a life-long enthusiasm for teaching. Looking back, I would have never begun on this path if it were not for the encouragement of the astronomical and scientific community of Southeastern Connecticut, namely Leslie Brown, Scott MacNeill, Bruce Levine, and Bruce MacMahon. Thank you all for instilling in me a sense for the mysterious.

And to my dearest friends and colleagues for the many protracted discussions and incoherent ramblings on science and life: Alastair Marshall, Callum Fairbairn, Hayden Goodfellow, Didrik Willadssen, Lukas Zalesky, Vasily Kokorev, Kate Gould, Albert Sneppen, and Vadim Rusakov. Your encouragement and willingness to hear me out has not gone unnoticed or unappreciated.

Much of this research was performed during the COVID-19 pandemic. I wish to thank all of those whose sacrifices have allowed us to carry on during these difficult times. The people of Denmark, in particular, deserve thanks for not only making this a wonderful and safe place, but also for their investment in scientific pursuits and institutions like the Danish National Research Foundation, from which much of this work, as well as the Cosmic Dawn Center, was funded. I am especially grateful for the expert administrative support and friendship of Guarn Nissen and Helena Baungaard-Sørensen.

This work stands as the latest in a series of ever advancing frontiers into the origins of the cosmos. From the first star catalogs of Tycho Brahe in the late 1600s to the discovery of the 'island' galaxies by Edwin Hubble, Milton Humeson, Vesto Slipher, and Henrietta Leavitt some five centuries later, catalog makers have charted heavens above to reveal the unfolding of our universe and its mysteries. You stared into the abyss, and brought forth a field of study which has captured my attention and curiosity. This work is dedicated to you.

I am forever grateful to have Alisa Danilenko as my best friend and partner in life. You have been with me for the whole journey, through thick and thin, in the best of times and the worst of times. You've been there for me in triumph as well as defeat, and helped me to forge ahead regardless.

And lastly, to my parents, for their ceaseless encouragement, care, and love. Not to mention long drives to extracurricular classes, early morning pick-up at the observatory, and the dozens of books which line nearly every shelf in the house. Who would have thought that a brief glance through a telescope could have led to all of this? I am and will be forever grateful for the lifetime of wonder that you have given to me. For all, my love and eternal gratitude.

LIST OF PUBLICATIONS

Results from these manuscripts are used in this dissertation

Weaver, J. R., G. Brammer, C. M. Casey, N. Chartab, I. Davidzon, A. L. Faisst, S. L. Finkelstein, S. Fujimoto, K. M. L. Gould, O. Ilbert, C. K. Jespersen, O. B. Kauffmann, G. Magdis, H. J. McCracken, C. McPartland, B. Milvang-Jensen, B. Mobasher, P. Oesch, D. B. Sanders, C. Scarlata, M. Shuntov, C. L. Steinhardt, D. K. Stern, I. Szapudi, S. Toft, F. M. Valentino, and L. Zalesky. *Beasts in the Bubbles: Characterizing ultra-luminous galaxies at Cosmic Dawn*. JWST Proposal. Cycle 1. Mar. 2021.

Kauffmann, O. B., O. Ilbert, **J. R. Weaver**, H. J. McCracken, B. Milvang-Jensen, G. Brammer, I. Davidzon, O. Le Fèvre, D. Liu, B. Mobasher, A. Moneti, M. Shuntov, S. Toft, C. M. Casey, J. S. Kartaltepe, A. M. Koekemoer, D. B. Sanders, and L. Tresse. "COSMOS2020: UV selected galaxies at $z \geq 7.5$." Submitted to Astronomy & Astrophysics. 2022.

Weaver, J. R., I. Davidzon, S. Toft, O. Ilbert, H. J. McCracken, K. M. L. Gould, C. K. Jespersen, C. Steinhardt, C. Lagos, P. L. Capak, N. Chartab, C. Laigle, O. B. Kauffmann, A. M. Koekemoer, B. Milvang-Jensen, C. J. McPartland, B. Mobasher, A. Moneti, D. B. Sanders, M. Shuntov, and L. Zalesky. "COSMOS2020: The Galaxy Stellar Mass Function." Prepared for Astronomy & Astrophysics. 2022.

Weaver, J. R., O. B. Kauffmann, O. Ilbert, H. J. McCracken, A. Moneti, S. Toft, G. Brammer, M. Shuntov, I. Davidzon, B. C. Hsieh, C. Laigle, A. Anastasiou, C. K. Jespersen, J. Vinther, P. Capak, C. M. Casey, C. J. R. McPartland, B. Milvang-Jensen, B. Mobasher, D. B. Sanders, L. Zalesky, S. Arnouts, H. Aussel, J. S. Dunlop, A. Faisst, M. Franx, L. J. Furtak, J. P. U. Fynbo, K. M. L. Gould, T. R. Greve, et al. "COSMOS2020: A Panchromatic View of the Universe to $z \sim 10$ from Two Complementary Catalogs." In: *The Astrophysical Journal Supplement Series* 258.1, 11 (Jan. 2022), p. 11.

Weaver, J. R., L. Zalesky, V. Kokorev, C. J. R. McPartland, N. P. Allen, N. Chartab, K. M. L. Gould, M. Shuntov, I. Davidzon, A. L. Faisst, N. Stickley, P. Capak, S. Toft, D. Masters, B. Mobasher, D. B. Sanders, O. B. Kauffmann, O. Ilbert, A. Moneti, G. Brammer, and H. J. McCracken. "The Farmer: A reproducible profile-fitting photometry package for deep galaxy surveys." Prepared for The Astrophysical Journal. 2022.

Publications not directly part of this dissertation

Martin, C. L., L. F. Barrientos, A. L. Faisst, J. I. Gonzalez Lopez, Y. Harikane, W. Hu, L. Infante, C. Jiang, J. Kartaltepe, O. B. Kauffmann, A. A. Khostovan, A. M. Koekemoer, S. Malhotra, H. J. McCracken, M. Ouchi, J. Rhoads, S. Toft, J. Wang, **J. R. Weaver**, I. Wold, and Z. Zheng. *Galaxy Protoclusters as Drivers of Cosmic Reionization*. JWST Proposal. Cycle 1, ID. #1635. Mar. 2021.

Davidzon, I., K. Jegatheesan, O. Ilbert, S. de la Torre, S. K. Leslie, C. Laigle, S. Hemmati, D. C. Masters, D. Blanquez-Sese, O. B. Kauffmann, G. E. Magdis, K. Małek, H. J. McCracken, B. Mobasher, A. Moneti, D. B. Sanders, M. Shuntov, S. Toft, and **J. R. Weaver**. "COSMOS2020: Manifold Learning to Estimate Physical Parameters in Large Galaxy Surveys." In: *arXiv e-prints* (June 2022), arXiv:2206.06373.

Euclid Collaboration, A. Moneti, H. J. McCracken, M. Shuntov, O. B. Kauffmann, P. Capak, I. Davidzon, O. Ilbert, C. Scarlata, S. Toft, **J. R. Weaver**, R. Chary, J. Cuby, A. L. Faisst, D. C. Masters, C. McPartland, B. Mobasher, D. B. Sanders, R. Scaramella, D. Stern, I. Szapudi, H. Teplitz, L. Zalesky, A. Amara, N. Auricchio, C. Bodendorf, D. Bonino, E. Branchini, S. Brau-Nogue, M. Brescia, et al. "Euclid preparation. XVII. Cosmic Dawn Survey: Spitzer Space Telescope observations of the Euclid deep fields and calibration fields." In: *A&A* 658, A126 (Feb. 2022), A126.

Euclid Collaboration, S. E. van Mierlo, K. I. Caputi, M. Ashby, H. Atek, M. Bolzonella, R. A. A. Bowler, G. Brammer, C. J. Conselice, J. Cuby, P. Dayal, A. Díaz-Sánchez, S. L. Finkelstein, H. Hoekstra, A. Humphrey, O. Ilbert, H. J. McCracken, B. Milvang-Jensen, P. A. Oesch, R. Pello, G. Rodighiero, M. Schirmer, S. Toft, **J. R. Weaver**, S. M. Wilkins, C. J. Willott, G. Zamorani, A. Amara, N. Auricchio, M. Baldi, et al. "Euclid preparation: XXI. Intermediate-redshift contaminants in the search for $z > 6$ galaxies within the Euclid Deep Survey." In: *arXiv e-prints*, arXiv:2205.02871 (May 2022).

Faisst, A. L., R. R. Chary, S. Fajardo-Acosta, R. Paladini, B. Rusholme, N. Stickley, G. Helou, **J. R. Weaver**, G. Brammer, A. M. Koekemoer, and H. Miyatake. "Joint Survey Processing. I. Compact Oddballs in the COSMOS Field-Low-luminosity Quasars at $z > 6$?" In: *ApJ* 929.1, 66 (Apr. 2022), p. 66.

Jin, S., E. Daddi, G. E. Magdis, D. Liu, **J. R. Weaver**, Q. Tan, F. Valentino, Y. Gao, E. Schinnerer, A. Calabro, Q. Gu, and D. Blanquez Sese. "Diagnosing deceivingly cold dusty galaxies at 3.5<z<6: a substantial population of compact starbursts with high infrared optical depths." In: *arXiv e-prints*, arXiv:2206.10401 (June 2022), arXiv:2206.10401.

Lagattuta, D. J., J. Richard, F. E. Bauer, C. Cerny, A. Claeyssens, L. Guaita, M. Jauzac, A. Jeanneau, A. M. Koekemoer, G. Mahler, G. Prieto Lyon, A. Acebron, M. Meneghetti, A. Niemiec, A. Zitrin, M. Bianconi, T. Connor, R. Cen, A. Edge, A. L. Faisst, M. Limousin, R. Massey, M. Sereno, K. Sharon, and **J. R. Weaver**. "Pilot-WINGS: An extended MUSE view of the structure of Abell 370." In: *MNRAS* 514.1 (July 2022), pp. 497–517.

Shuntov, M., H. J. McCracken, R. Gavazzi, C. Laigle, **J. R. Weaver**, I. Davidzon, O. Ilbert, O. B. Kauffmann, A. Faisst, Y. Dubois, A. M. Koekemoer, A. Moneti, B. Milvang-Jensen, B. Mobasher, D. B. Sanders, and S. Toft. "COSMOS2020: The cosmic evolution of the stellar-to-halo mass relation for central and satellite galaxies up to $z \sim 5$." In: *arXiv e-prints*, arXiv:2203.10895 (Mar. 2022).

Sneppen, A., C. L. Steinhardt, H. Hensley, A. S. Jermyn, B. Mostafa, and **J. R. Weaver**. "Implications of a Temperature-dependent Initial Mass Function. I. Photometric Template Fitting." In: *ApJ* 931.1, 57 (May 2022), p. 57.

Steinhardt, C. L., A. Sneppen, H. Hensley, A. S. Jermyn, B. Mostafa, **J. R. Weaver**, G. Brammer, T. H. Clark, I. Davidzon, A. C. Diaconu, B. Mobasher, V. Rusakov, and S. Toft. "Implications of a Temperature Dependent IMF III: Mass Growth and Quiescence." In: *arXiv e-prints*, arXiv:2206.01750 (June 2022).

Valentino, F., G. Brammer, S. Fujimoto, K. E. Heintz, **J. R. Weaver**, V. Strait, K. M. L. Gould, C. Mason, D. Watson, P. Laursen, and S. Toft. "The Archival Discovery of a Strong Lyα and [C II] Emitter at $z = 7.677$." In: *ApJ* 929.1, L9 (Apr. 2022), p. L9.

Weaver, J. R. and K. Horne. "Dust and the intrinsic spectral index of quasar variations: hints of finite stress at the innermost stable circular orbit." In: *MNRAS* 512.1 (May 2022), pp. 899–916.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ACS	Advanced Camera for Surveys
ALMA	The Atacama Large Millimeter/submillimeter Array
AGN	Active Galactic Nucleus
CCD	Charge-coupled device
CDF	Cumulative Distribution Function
CLAUDS	CFHT Large Area U-band Deep Survey
CFHT	Canada-France-Hawaii Telescope
EAzY	Easy and Accurate z(photometric redshift) from Yale
EoR	Epoch of Reionization
FIR	Far Infrared
FWHM	The full width of distribution at half of maximum height
GALEX	The Galaxy Evolution Explorer
GOODS	Great Observatories Origins Deep Survey
(G)SMF	The galaxy stellar mass function
H20	Hawaii 20 deg^2 Survey
HSC	The Hyper Suprime-cam Instrument
HST	Hubble Space Telescope
ICM	Intracluster medium
ICM	Intergalactic medium
ISM	Interstellar medium
IFS/U	Integral field spectroscopy / unit
IMF	The initial mass function
IRAC	The Infrared Array Camera
JWST	The James Webb Space Telescope
MCMC	Markov Chain Monte Carlo
NIR	Near Infrared
MIR	Mid Infrared
PSF	The point spread function
RMS	Width of a distribution measured by root mean square
SC	The Suprime-cam Instrument

SED	Spectral Energy Distribution
SFH	Star-formation history
SFE	Star-formation efficiency
SFR	Star-formation rate
SNR	signal-to-noise ratio
SPS	stellar population synthesis
VISTA	The Visible and Infrared Survey Telescope for Astronomy
WFC3	Wide Field Camera 3

1 INTRODUCTION

The Cosmos is all that is or was or ever will be. Our feeblest contemplations of the Cosmos stir us – there is a tingling in the spine, a catch in the voice, a faint sensation, as if a distant memory, of falling from a height. We know we are approaching the greatest of mysteries.

- Carl Sagan, *1980*

1.1 EARLY COSMOLOGIES

Cosmos - the term first appears in the writings of the 5th century BC Greek Philosopher Pythagoras - κόσμος - to mean the orderly arrangement of the world[1]. The Greeks meant this in a more powerful sense too: the universe is in fact orderly, and behaves according to certain rules. These rules are worth careful study and understanding, and may be revealed given sufficient analytical investigation. An account of the achievements of our species, from the first pages, has been not simply that of humanity as tool-makers, but as a deeply curious species driven to explore and expand the frontiers of our world, and uncover these hidden rules which govern this cosmos. This study is *Cosmology* or κόσμος-λόγος.

The earliest cosmological systems known to us emerged in the first agrarian civilizations of the Indus River Valley, namely Babylonia and Phoenica, with perhaps simultaneous developments associated with the Hindu and Jainastic traditions of India. While colorful, these cosmologies were not predictive, but instead were mythological and sought to explain natural phenomena in connection to the everyday experiences of life itself.

Some of the earliest surviving records of astronomical catalogs come from dynastic China c. 300-400 BC, and depict the locations of the brightest stars,

Bibliographic documentation for material presented in Sections 1.1-1.4 is provided at the end of the chapter.

1 Curiously, the word only joined the English lexicon less than two centuries ago with the work of the German scientist Alexander von Humboldt, from the translation of his five-volume treatise *Kosmos*.

and the constellations of the Chinese culture. The two oldest surviving European texts come from c. 8$^{\text{th}}$ BC century Archaic Greece: Homer's *Odyssey* and *Iliad*. They include some of the first written references to named stars and constellations, which are still in use today such as Ursa Major and Orion. It is now well understood that these and other cosmological and astronomical precepts recorded in these early Greek writings were adopted from much earlier developments in the Near East that were imported into Greecian culture by interactions with Phoencian sailors. Our modern western astronomical systems descends therefore from these earliest civilizations, whose invention of sedentary agriculture not only permitted the formation of specialized workers, but required precise knowledge of the tides and seasons. The success of these early civilization depended upon knowing the time of year in order to plant crops, derived from the appearance and disappearance of the constellations. This inextricable link elevated astronomical observations to matters of great importance within these ancient city-states.

Astronomy in Archaic Greece followed from the Pythagorean school, established by the teachings of Pythagoras of Samos c. 500 BC, and was considered an integral component to mathematical studies. The Pythagoreans devised idealized geometric models to *predict* the motions of the heavenly bodies, including the Sun, Moon and stars. Like the Babylonians before them, the Greeks wished to utilize these models to understand the influence of the heavens on humanity. Importantly, their descriptions and theories were based on natural laws, as opposed to supernatural causes as had been previously assumed.

From these and other observations there emerged a more complicated picture, whereby certain bright stars were seen to wander, and even move in retrograde against an otherwise fixed array of stars. These are the planets, from the greek πλανήτης or *wanderers*. Only the brightest amongst them can be observed with the naked eye: Hermes, Aphrodite, Ares, Zeus, and Chronus in the Greek tradition; their Roman counterparts being Mercury, Venus, Mars, Jupiter, and Saturn, respectively.

Astronomy was developed further in Classical Greece (c. 400 BC), dominated by Plato, and the Academy. Plato contributed the first scientific cosmological model, largely displacing the previous focus on stars with predictions of planetary motions. He conceived that the apparently chaotic paths of the known planets could be understood as uniform circular motions along simple, two-dimensional planar orbits centred about an unmoving Earth.

This idealized geometric picture introduced by Plato was further developed by Eudoxus of Cnidus, who proposed proposed that heavenly bodies were instead carried through the sky by fixed, concentric three-dimensional translucent rotating spheres centred about the Earth. This was the first instance where mathematics entered cosmology. By adjusting the axial tilt and period of each sphere, Eudoxus established a model which could roughly approximate the movements of the planets. The furthest sphere contained the fixed stars, and was considered the edge of the universe. No works of Eudoxus have survived the wear of time; we can only gleam these hints from other sources, including Plato's more famous student, Aristotle.

While the system of Eudoxus improved over the simplistic Platonic model, it suffered from several deficiencies. For example, the changing speed and retrograde orbits could not be predicted, nor their apparent change in brightness. To address these flaws, Apollonius of Perga (c. 200 BC) exchanged the concentric spheres for deferent circles offset from Earth, each of which carry a smaller epicyclic circle that carries the respective planet.

1.2 CHARTING THE HEAVENS

The earliest star catalog correspondent to western culture was produced by Hipparchus c. 150 BC. Not only did he use these observations to improve on the model of Apollonius, but in the process developed some of the first astronomical instruments including the armillary sphere, and possibly the astrolabe. The catalog contains the positions and brightnesses of stars, and is speculated to have first introduced a system of apparent magnitudes to classify their brightnesses[2].

Although Hipparchus is known to have considered a heliocentric system, scant records suggest that Aristarchus of Samos first formulated the heliocentric model already c. 300 BC. He also is known to have first attempted to estimate the size of the Sun and its distance from Earth. However, these measurements were relative to the radius of Earth and so Eratosthenes, working at the Library of Alexandria, attempted to measure the size of the Earth itself. He did so by comparing shadows cast in Alexandria and Thebes to deduce the angle of arc subtended by the distance between the two cities, estimating the circumference and hence radius of Earth - within a few per cent of their modern values.

2 This systems stands today, although defined on a logarithmic system by Pogson in 1856 to preserve Hipparchus' standard, as the human eye responds to light logarithmically.

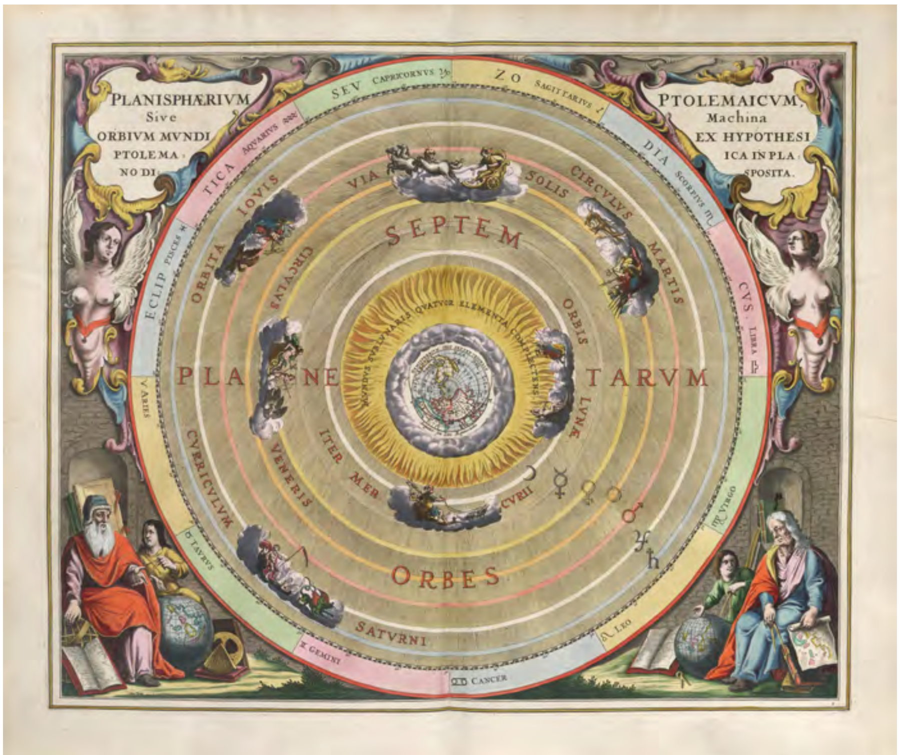

Figure 1.1: Planisphere of Claudius Ptolemy's geocentric model. *From the Haromia Macrocosmica of Andreas Cellarius, 1661.*

By the first century AD, the rise of the Roman Empire spelled the collapse of Hellenistic Greece, and the center of knowledge was re-established in Alexandria, under Roman Rule. It was during the height of Alexandrian learning that Claudius Ptolemy produced the first comprehensive mathematical theory of planetary motion. Inspired by the earlier geocentic traditions of Classical Greece, he fine-tuned the deferents and epicycles of Apollonius to produce exacting mathematical predictions of the movement of the Sun, Moon, and planets. In his view, heavenly bodies were not only perfect and unchanging, but must follow uniform circular motions in accordance to their divine nature (Figure 1.1).

The sacking of Rome by the Germanic tribes sank Europe into the dark ages for more than a millennia. Meanwhile, scientific contributions flourished in the Islamic caliphates of the Near East. Much of Greek learning, colored by the geocentric theories of Ptolemy, were rescued by Islamic scholars. Not only did they preserve the writings of Ptolemy in the *Almaghest*, literally *the greatest*, but developed even more sophisticated astronomical measuring tools, in part to precisely measure the times of worship and the direction of Mecca. They combined the mathematical developments of India (such as the concept of zero) with a revised numerical system still in use today.

Enlightenment would not be restored to Europe until the Middle Ages, c. 1500 AD. Nicolas Copernicus, a church deacon and educated astronomer in Poland, proposed a heliocentric model of the universe[3]. Futhermore, he proposed that the planets, including Earth, orbited the sun in circular motions. Unlike the prevailing Ptolemaic model, a heliocentric system naturally explained many of the phenomena which no longer required complex epicycles. In addition, the motion of the stars was plainly evident, and they were removed to a fixed background that separated them from the realm of the planets. Yet, Copernicus and his work were little known, and his model was in fact *less* accurate than Ptolemys at predicting the locations of planets. Furthermore, the geocentric teachings Ptolemy were not just the accepted viewpoint of the age, but were fact according to the dogma of the Catholic Church. Although essentially finished by 1532, Copernicus waited to publish this work until he was on his deathbed in 1543.

Figure 1.2: Sketch of the Mural Quadrant at the Urianiborg Observatory. It depicts Tycho Brahe (center) directing an observation by an assistant (right) of the altitude of a star. An assistant (lower center) reads off the time and the observation is recorded by a scribe (lower left). *Credit: Tycho Brahe (1598), Astronomiae instauratae mechanica, Wandsbeck.*

3 He did so apparently without knowledge of the much earlier proposal by Aristachus.

But in the 15^{th}, the Earth-centred dogma of the Church dominated public opinion. That did not stop Tycho Brahe, working from his observatory on the Danish island of Hven, from shattering the concentric (glass) spheres which conveyed the planets along their paths about the Earth. He discovered a number of comets, whose movement seemed to pass through these spheres. His observatory of Uraniborg established some of the most precise star catalogs of the era, as well as precise tracking of the planets enabled by new instruments and methods (as illustrated in Figure 1.2). He constructed a new but still geocentic model whereby the planets and the Sun orbited a fixed Earth using a complex set of epicycles and a fixed stellar sphere. Although he was aware of the Copernican model, his precise observations of the stars did not reveal any seasonal parallax, as had been predicted[4]. Yet, Brahe's vision was mathematically equivalent to the Copernican model but with a different frame of reference (Figure 1.3).

Then in 1597, Tycho fell out of favor with the new Danish King Christian IV and fled to Prague to continue his research. There he met Johannes Kepler, who recognized the importance of this vast collection of observational data. Passing soon after in 1604, Brahe left his library and collections to Kepler. In a sharp break with over two millenia of astronomical thought, Kepler used this data to question the largely accepted views of his time. He proved that the orbits of the planets did not follow the uniform circular motions introduced by Plato and Eudoxus, but rather followed elliptical orbits with predictable but changing speeds. This model was simpler and more accurate than that of Ptolemy. Beyond his three laws, Kepler's great philosophical contribution was the notion that the natural law that governed the Earth applied equally to the heavenly realm.

Not long after, the Italian scientist Galileo Galilei turned the telescope, which then was a new instrument of war, towards the night sky. What he saw stood in stark contrast from the geocentric viewpoint held by the academics of his day. The moon was not a perfect sphere, but rather cratered and scarred. Furthermore, he found four satellite moons of Jupiter which he named the Medicean stars, after his patron, Grand Duke Cosmio de Medici of Tuscany. He showed that they orbit Jupiter, not Earth, and that Venus underwent phases - facts which could not be explained by even the most complex modification to the geocentric model. The Church, however, condemned Galileo for his independent and heretical interpretation of scripture, and under the threat of torture, forced him to denounce his views[5]. He lived out the rest of his life under house arrest.

4 The lack of parallax seen by Brahe is due to the enormous distance between Earth and even the nearest star, and such measurements were out of reach by even the most sophisticated instruments accessible at the time.

5 Only in 1992, Pope John Paul II publicly apologized on behalf of the Church and cleared Galileo of any wrongdoings.

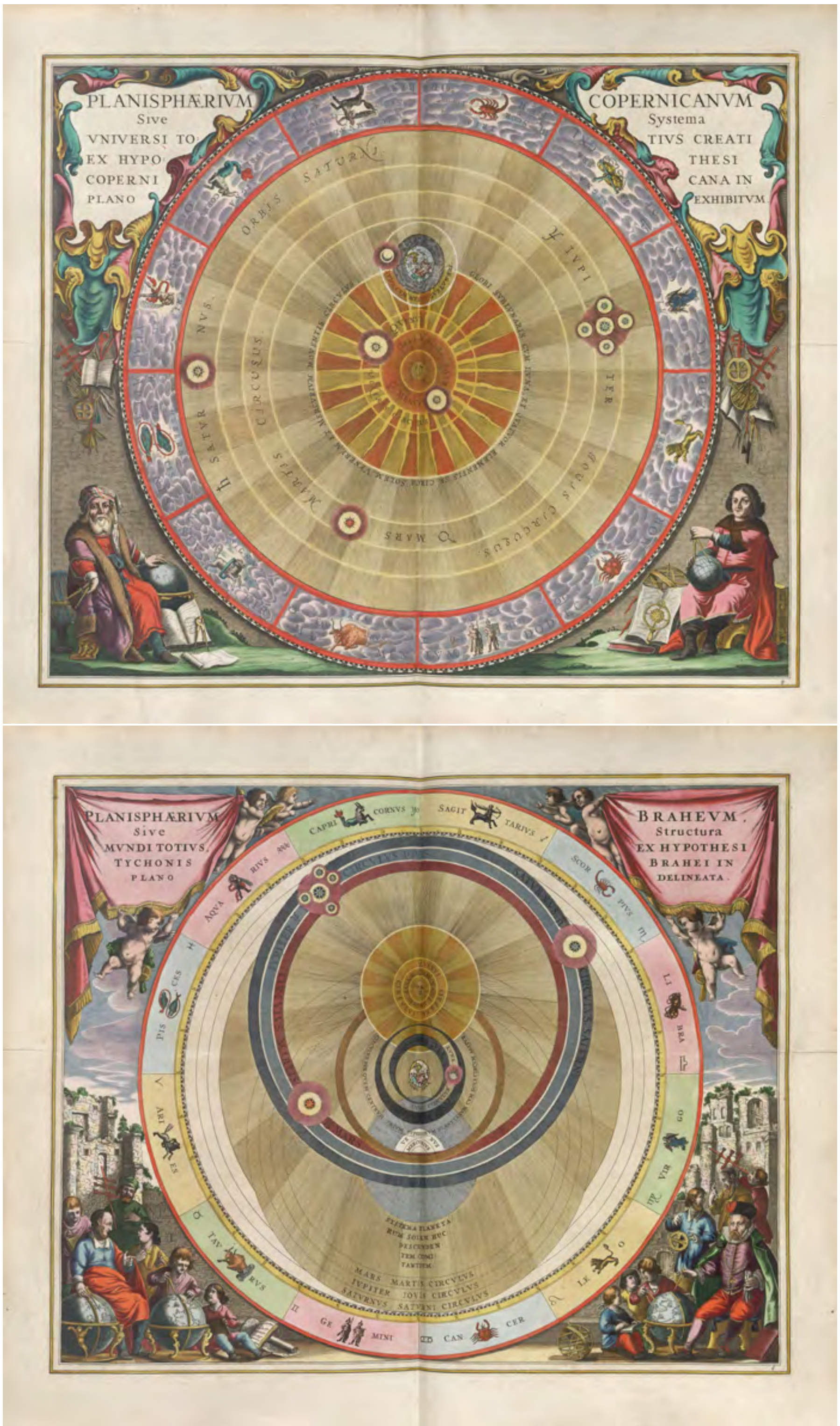

Figure 1.3: *Top*: Planisphere of the heliocentric Copernican model. *Bottom*: Planisphere of Tycho Brahe's geocentric model. *From the Haromia Macrocosmica of Andreas Cellarius, 1661.*

This mechanical universe was cemented by the work of Newton and his contemporaries in the 16th century. Newton recognized that the action of gravity which caused bodies to fall towards the Earth (as noted by Galileo) was in fact the same force which caused the orbits of the planets around the sun. By expanding Keplers laws, he showed with mathematical rigor that the cosmos was just that, an *orderly* arrangement of the world. Yet, there were still observations which did not match this grand theory. Frenchman Pierre-Simon Laplace, writing in the late 18th century, noted that although these deviations where small, they could potentially be significant over time and destabilize the solar system. However, he also recognized that measurements can be flawed, and so described them by an error function in an attempt to reconcile observations with theory. Laplace was operating under the prevailing wisdom that we live in a deterministic universe that could be measured to infinite precision, given sufficient instruments. Although little fanfare followed, this was the first instant that the tools of statistics, newly invented, entered into the lexicon of cosmology and astronomy. By the close of the century, however, increasingly powerful instrumentation led to more precise measurements, and further errors. Only gradually would this deterministic world-view be overturned, replaced by the notion of a universe that, although orderly, can only be meaningfully understood through *statistics*.

1.3 THE REALM OF THE NEBULAE

The most prestigious field of discovery in the 18th century was comet hunting, with prizes being offered by the nobilities of Europe. Charles Messier, working from Paris around 1780, attempted to find and catalog these faint flutters of light. Although he discovered 13 comets, he and Pierre Mećhain found 110 objects with similar visual appearance as comets, but which did not move from night to night. The pair noted them down to avoid wasting time re-observing them later on. These are known today as the Messier Objects, and consist of a mixture of galaxies, nebulae, and star clusters. However, with the telescopes of his day (and hindered by increasing cataracts), Messier could not easily distinguish between them let alone understand their nature.

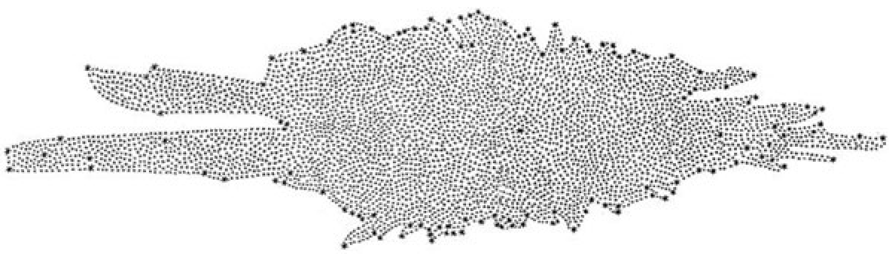

Figure 1.4: Sketch of the Milky Way as observed by William Herschel. *Credit: On the Construction of the Heavens. By William Herschel, Esq. F. R. S. Philosophical Transactions of the Royal Society of London, Vol. 75. (1785), pp. 213-266.*

The mystery deepened with even grander observations by William Herschel. Using his own suite of enormous telescopes, he was able to discern clusters of individual stars from these Messier objects that were previously thought to only be diffuse regions of light[6] (Herschel, 1786). He went further, and supposed that the band of the Milky Way was similarly composed of many thousands of stars (Figure 1.4). Although he failed to grasp the true size of the Milky Way, Herschel conclusively displaced Earth from the center of the known universe to merely one stellar system among myriad others like it. Our conception of the Universe, like it or not, had expanded.

Still, some of these diffuse regions could not be resolved. Instead, he postulated, in accordance with earlier theories introduced by the philosopher Immanuel Kant as well as by Laplace, that they too may be distant galaxies similar to our own (Kant, 1755; Laplace, 1796). Over the latter half of the 18th century, Herschel conducted one of the first deep sky surveys which resulted in a catalog of more than 2 000 nebulae[7] (Herschel, 1789). Yet, despite having constructed some of the most powerful telescopes of that century, Herschel was unable to confirm his hypothesis of neighboring galaxies.

The observations introduced by Herschel were soon expanded upon by Lord Rosse and his 'Leviathan of Parsonstown' in the middle of the 19th century (Rosse, 1844), and later by George Ellery Hale from Mount Wilson equipped with a fleet of large telescopes. The ever more powerful instruments showed stars within these nebulae, but that was not conclusive evidence for their nature. By the second decade of the 20th century, two camps had emerged: those who believed that the nebulae were within our Galaxy, and those who believed that, at least some of them, lie far distant as separate 'island universes'. The matter was put to debate organized by the US National Academy of Sciences in 1920. Harlow Shapley would argue for the nebulae as being local phenonema, while the more established Heber Curtis would challenge this picture. This event became known as the 'Great Debate'. Shapley had been measuring the distances to globular clusters in order to estimate the size of our Galaxy, which he estimated around 50,000 light years, $\sim 10\times$ larger than any previous estimate. Much of his argument therefore was an objection the extreme distances, sizes, and luminosities required to explain observations of Andromeda, and similar nebulae, as objects beyond the now much larger Milky Way. If they were to be believed, then our Galaxy is just one of potentially infinite array of galaxies. Curtis on the other hand accepted this reality, even if it meant that humanity would be pushed further from center stage. However, both sides lacked the concrete data required to justify their claims. Although the debate did little to settle this matter, it highlighted the need for further observations.

It wasn't until Edwin Hubble, working from Mount Wilson in the first half of the 20th century, that this mystery was finally solved. Again, new

6 In addition to star clusters, Herschel discovered the planet Uranus as well as infrared spectroscopy.
7 Herschel's deep sky objects were later expanded by John Dreyer and incorporated into the New General Catalog (NGC), still in use today.

technology enabled scientific advancement. Using the newly completed 100-inch (2.5m) Hooker Telescope[8], Hubble identified pulsating variable stars in the direction of the Andromeda Nebula (Figure 1.5) as well as Triangulum. He was then able to apply the recent method discovered earlier by Henrietta Leavitt to determine its luminosity and from it, its distance (Leavitt et al., 1912). What he found was that the distances to these objects were each more than a million light years, significantly larger than the estimated size of the Milky Way. Furthermore, they must contain hundreds of millions of stars in order to be visible from so far away. The Andromeda Nebula was really the Andromeda *Galaxy*. Leavitt, despite enabling this breakthrough, died in 1924 before she could accept her nomination for a Nobel Prize. Only in 2021 did the American Astronomical Society recognize her contribution by publicly supporting naming the relationship the *Leavitt Law*.

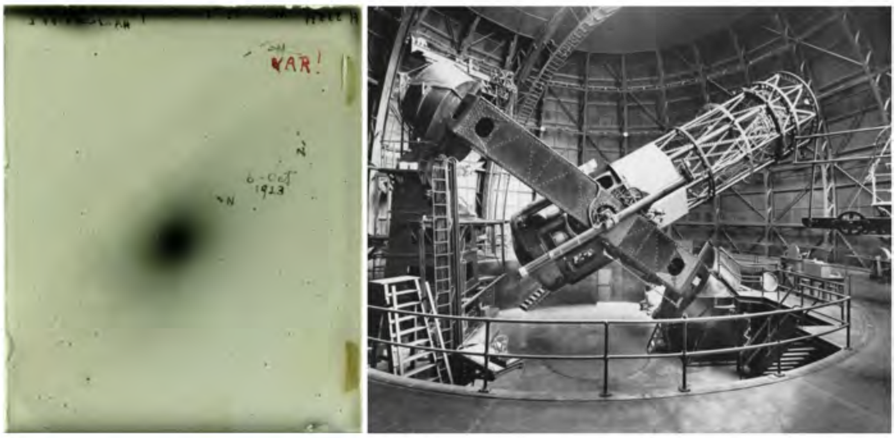

Figure 1.5: *Right*: photographic plate taken by Edwin Hubble showing the location of a variable star in Andromeda. *Left:* 100-inch Hooker Telescope at Mount Wilson. *Credit: Carnegie Observatories.*

Hubble's findings were first published in the *New York Times* in November 1924, and announced shortly after during a meeting of the American Association for the Advancement of Science at the close of the year. Hubble, wanting more conclusive data, did not formally publish the results until 1929, entitled *The Realm of the Nebulae* (Hubble, 1929). The debate was over, and the island universes had won.

1.4 OUR EXPANDING UNIVERSE

Cosmology entered the 20th century confident in an static and eternal universe. Einstien had published his General Theory of Relativity in 1911. In essence, Einstein theorized that presence of matter warps space-time and in doing do generates a gravitational field. While similar ideas were already introduced

8 The Hooker Telescope is a reflector, but was underused at Wilson by more senior staff who believed that traditional refractors were more reliable, despite considerable difficulty in building larger refractors.

earlier by Laplace, Einstein's mathematical theory provided concrete predictions about the gravitational conditions around massive objects like stars and even black holes. Immediately, General Relativity showed that it could predict the precession of Mercury, the innermost planet, which Newton's theory could not.

In an attempt to further test Einstein's theory, teams from Argentina, Germany, and the US set out to Crimea to observe the solar eclipse of 1914 in order to measure the deflection of background stars by the gravitational field of the Sun. However, a culmination of the first World War and bad weather prevented the observations. Although attempts were made in the interim, it was not until the eclipse of 1919 that a team of observers from Cambridge University succeeded in making the measurements from Príncipe Island, off of West Africa, led by Sir Arthur Eddington (Figure 1.6). Einstein had expanded his theory in 1915, and the observations matched. Newtonian gravity had been displaced by General Relativity.

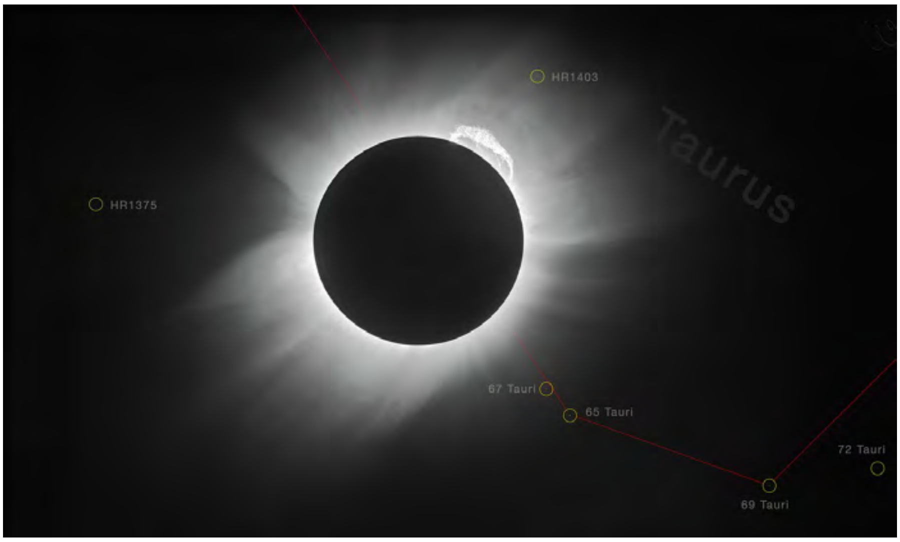

Figure 1.6: Image of the 1919 solar eclipse observed by Sir Arthur Eddington and others. Notable stars are indicated. *Credit: ESO/Landessternwarte Heidelberg-Königstuhl/F. W. Dyson, A. S. Eddington, & C. Davidson*

The confirmation of Einstein's General Relativity opened the floodgates to new theoretical explorations of cosmology. For Einstein, however, the results were disappointing. Over a series of discussion with Willem de Sitter, Einstein realized that his equations suggested that the Universe should be in a state of gravitational collapse. Our very existence suggests this is not the case, and so Einstein added a pressure term which perfectly balanced the collapse[9] - preserving a static and eternal universe. Although Einstein admitted that the solution was ugly, the cosmological community was satisfied with this solution.

9 Newton too found that his theory predicted a similar collapse, and in the fashion of his time, invoked divine intervention to sate the dynamics of the cosmos.

However, like the epicycles of Ptolemy, Einstein's *cosmological constant* was justified only by the prevailing wisdom of the time. Having read Einstein's early papers, Aleksandr Friedmann[10] discarded this constant and demonstrated that there are a range of possible dynamics depending on the gravitational content of the Universe. Einstein, as well as the Western scientific establishment, was displeased and initially attempted to censure his results citing mathematical flaws. Friedmann did not bow to the pressure; to him these 'curious facts' were simply an interesting theory, even if they were not yet supported by data. He died suddenly in 1925 in relative obscurity.

Working independently, the Belgian Georges Lemaître rediscovered the models of Friedmann. Lemaître realized that the apparently expanding universe predicted by the equations would imply a beginning. Being a Catholic Priest, he saw this result as evidence for biblical creation. His 'primeval atom' would have, in this view, expanded to form the present day universe. Furthermore, Lemaître suggested that the initial explosion would have emitted detectable radiation, connecting this theory with *observable* phenomena. Yet, despite his connections to prominent cosmologists and astronomers such as Eddington, he received a similarly cold response from the community. In an encounter with Einstein at the 1927 Solvay Conference, the cosmological giant quipped: "Your calculations are correct, but your physics is abominable".

The cosmological community, led by Einstein, had succeeded in preserving the notion of an eternal, static universe. However, cracks were already beginning to show. Vesto Silpher, working from the 24-inch refractor at Lowell Observatory in 1912, obtained spectra of several nebulae. In one of the first long-exposure measurements taken at that time, Silpher obtained a 40-hour exposure of the Andromeda Nebula and found that it was blueshifted by $\sim 300\,\mathrm{km\,s^{-1}}$. Swayed by the common view at the time, he doubted his measurements. As an independent check, he performed a similar observation of another nebula (now known as the Sombero Galaxy) and found that it was redshifted by $\sim 1000\,\mathrm{km\,s^{-1}}$. He continued these measurements for another decade, finding only four blue-shifted galaxies out of dozens of red-shifted ones. Silpher had accidentally discovered that the universe was in fact, *dynamic*. This 'drift' hypothesis was opposed by de Sitter, whose static universe predicted that distant objects would emit a different frequency of light. The same was not true of Einstein's solutions. Their distance, however, was still unknown and so the models could not be differentiated.

Hubble, not to be left out, repeated similar observations from the 100-inch at Mount Wilson, 17× more powerful than the Lowell refractor. To resolve this mystery, Hubble set about measuring both the recession velocities of galaxies *and* their distances (using the Cepheid method). He received considerable assistance from Milton Humason[11], who performed the careful spectroscopic

10 Friedmann's independence was due to circumstance rather than choice as a result of isolation that ensued following the 1917 Revolution.

11 Milton Humason started work at Mount Wilson as a mule driver and later a janitor. He soon proved himself a careful and motivated observer, and was promoted to staff. Without so much as a high school diploma, Humason was instrumental in Hubble's discoveries.

observations to determine the velocities. Hubble worked on the distances. They were aided by newly available equipment, including a more sensitive film which allowed the previously multi-night observations to be taken in only a few hours. By 1926, the pair had established a tentative relationship between the distance and recession velocity of 47 galaxies (Figure 1.7). The further the galaxy, the faster it was moving away. Modestly, Hubble published these early results in a six-page paper *A Relation between the Distance and Radial Velocity among Extra-Galactic Nebulae* (Hubble, 1929). He was reluctant to attach any particular interpretation to the data, and rightly so; while the relationship was clear, it was not statistically robust and he did not want to embroil himself in a theoretical argument. The pair continued to refine their measurements and obtain new ones at even greater distances. The relationship held, with both Einstein's and de Sitter's models seemingly now demolished.

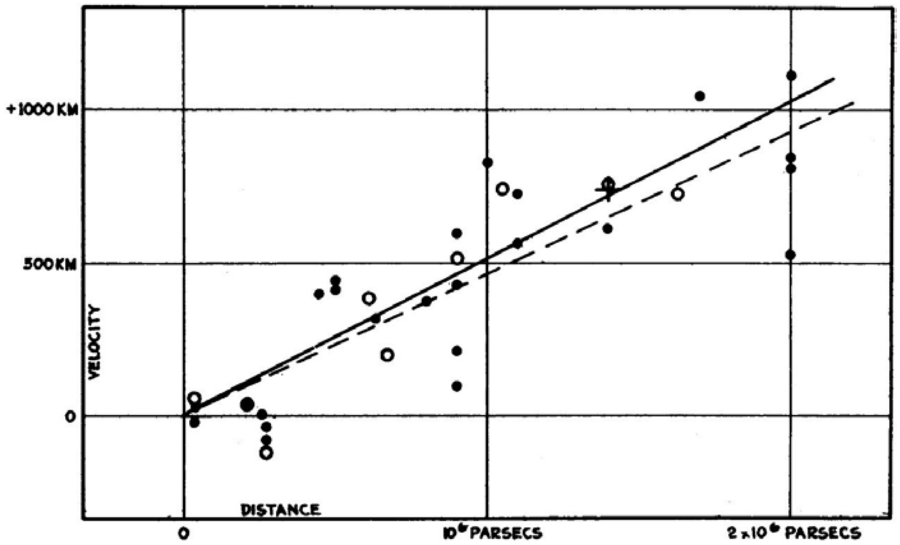

Figure 1.7: Hubble's 1929 velocity-distance relation for 46 nebulae. The black dots and the solid line represent the solution obtained from the 24 nebulae for which individual distances were determined, using them separately. Empty circles and the dashed line represent the solution obtained by combining the nebulae into groups. The cross represents the mean velocity for a set of 22 nebulae whose distances could not be estimated individually. *Credit: Hubble 1929.*

While the cosmological community had not been swayed by previous findings of dynamical events, such as novae, Hubble's findings suggested dynamics on the largest scales. The relationship not only implied a dynamic, expanding universe, but also that at some previous time the galaxies were much closer. Furthermore, the relating constant H_0 could be used to infer the age of the Universe itself, which, by the time Hubble had published the full dataset in 1929, was estimated at 1.8 Billion years. However, the theoretical basis of Hubble's findings had already been predicted by Lemaître, and subsequently ignored. Lemaître saw the observational data as a resounding confirmation of a dynamic universe, and his primeval atom.

Tides began to shift when Einstein visited Hubble while on sabbatical at the California Institute of Technology in 1931. After seeing Hubble's data with his own eyes, Einstein swiftly and very publicly renounced his belief in a static universe in favor of an expanding universe and a moment of creation. Yet others were not so easily swayed. In fact, the estimated age of the universe at the time was significantly shorter than the ages more confidently measured of rocks on Earth. Several notable figures attempted to reverse the momentum of Hubble's discovery, including Eddington himself.

Subsequent investigations by the cosmological community clarified the ramifications of this picture. While the galaxies themselves are held together by their immense self-gravity, it is the space between them that is expanding. Furthermore, the expansion is universal and experienced by any observer in any galaxy. Earth does not sit in the center of the universe, rather the universe has no center at all. In addition, scrutiny and subsequent recalibration of Hubble's observations by Walter Baade and Allan Sandage in the middle of the 20$^{\text{th}}$ century pushed the age of the universe between 10 and 20 Billion years (Baade, 1944; Sandage, 1958), which was now older that of any known object.

The emerging field of nuclear research joined the fray, hoping to explain why light atoms (e.g., Hydrogen and Helium) are more abundant than heavier 'metals'. Intense theoretical work by George Gamow, Ralph Alpher, and Robert Herman established a plausible timeline following creation (Alpher et al., 1948)[12]. Their theory of primordial nucleosynthesis was sufficient to explain the relative abundances of Hydrogen and Helium. However, they could not explain the emergence of heavier metals, which were later shown to be formed in stellar nucleosynthesis.

At the same time, Fred Hoyle, Thomas Gold, and Hermann Bondi developed a seemingly impossible compromise between the two camps. Inspired by the plot of the 1946 movie *Dead of Night*, the trio proposed the *Steady State Model*. The universe was expanding, but newly formed galaxies developed in the increasingly wide voids between existing galaxies (Bondi et al., 1948; Hoyle, 1948). The model permitted that the universe could be dynamic on small scales, but that its overall cosmological appearance would be a steady state. Compatible with Hubble's observations, the traditionalists replaced their view with a steady, but nonetheless eternal Universe. In the ensuring discussions, Hoyle would famously coin Lemaître's theory of creation a *big bang*[13].

Like the matter of the Nebulae, this new debate would be settled by new instrumentation. The results from Gamow, Alpher, and Herman's theory also predicted that the incredibly hot plasma that pervaded the early universe would eventually cool until electrons could be held onto by atomic nuclei

12 The resulting paper was titled "The Origin of Chemical Elements", but is more commonly known as the Alpher-Bethe Gamow paper for the resemblence of the authors' names to the greek letters alpha, beta, and gamma. Bethe, in fact, was added primarily to complete the title and had virtually no role in the research.

13 Hoyle's term is a misnomer. While Lemaître's model predicted a beginning, it was neither 'big' nor a 'bang', as the universe was small and there was no medium in existence through which a bang could yet travel - yet, the name stuck.

(recombination). The photons, no longer being scattered by resonant interactions with the electrons, were free to transit space (photon decoupling). They suggested that although redshifted to lower energies, this primordial radiation should be still detectable today. Finding this *cosmic microwave background* (CMB) would be a seemingly indisputable confirmation of the big bang.

Their predictions largely went ignored until 1964 when David Wilkinson, Peter Roll, and Robert Dickie began work building an instrument to detect this likely faint microwave radiation. Unknown to Dickie's team, Arno Penzias and Robert Wilson, were working at the Holmdel Horn Antenna under Bell Labs investigating the feasibility of satellite communication. They could not make progress in their tests, however, as the measurements were dominated by a strong noise which they could not remove nor explain[14]. Stumped, they called Dickie for suggestions to solve their problem. Hearing this, Dickie turned to his team and said famously, "Boys, we've been scooped!". The elusive cosmic microwave background had been found (Penzias et al., 1965). Penzias and Wilson were awarded the 1978 Nobel Prize for their discovery.

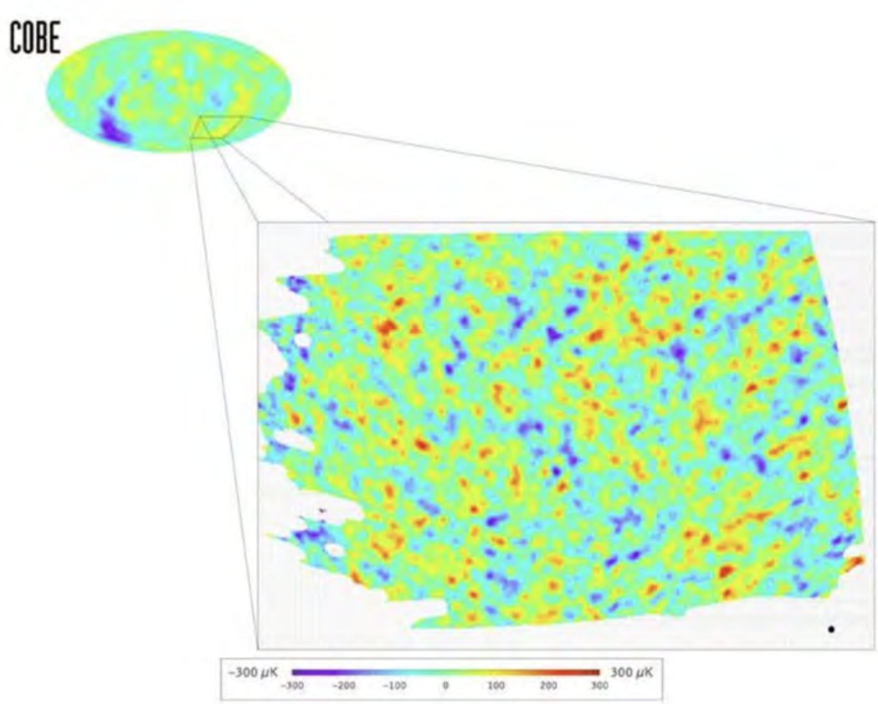

Figure 1.8: The balloon-launched BOOMERanG experiment provided better resolution than COBE over a small patch of the Cosmic Microwave Background, allowing a more detailed probe of the anisotropy structure. *Credit: US National Aeronautics and Space Administration.*

A series of subsequent observations by increasingly capable telescopes have sharpened this first view. Most famously, the 1992 COBE satellite provided the first all sky image of the background, confirmed the ambient temperature and blackbody form of the radiation (Mather et al., 1994), and demonstrated the

14 They went so far as to destroy a roost of pigeons living in the antenna!

scale of the anisotropies in the photon field at decoupling (Smoot et al., 1992). The existence of this anisotropy ruled out Hoyle's steady state model and its prediction of a entirely homogeneous universe. An era of precision cosmology had begun. Narrower observations by the balloon-launched BOOMERanG (Netterfield et al., 2002) and MAXIMA (Hanany et al., 2000) experiments provided detailed observations of smaller scale anisotropies not resolved by COBE (Figure 1.8). Furthermore, they suggested that Friedmann's model of a flat universe was most likely, and provided evidence to support existing theories that universe went through an initial period of inflation necessary to explain the structure of the anisotropies (Guth, 1981), which in this view developed as a result of quantum fluctuations within Lemaître's primeval atom. Follow-up with higher-resolution space-borne observatories WMAP (Spergel et al., 2003) and Planck (Collaboration et al., 2014) have probed anisotropies on even smaller scales, with the latter providing significant tension with cosmological measurements made from sources in the local universe (e.g. Type 1a SNe, see Scolnic et al. 2018).

However, radiation is not the only observed component of the universe. Already in 1933, Fritz Zwicky found that the mass of the Coma Cluster must be about 400 times larger than what can be accounted for by visible starlight in order to be gravitationally bound[15]. Soon after, early attempts to measure the dynamical masses of nearby galaxies from their rotation curves separately by Horace Babcock (Andromeda, Babcock 1939) and Jan Oort (NGC 3115, Oort 1940) found that stars in the outermost regions of these systems appear to move with similar speed to stars closer to the center in the deeper gravitational potential. This 'flat' rotation curve is at odds with Keplerian orbits, which until that point explained nearly all gravitational orbital phenomena. To Babcock, this meant either that significant starlight was being absorbed in the outer disc, or there was truly missing matter. Stronger evidence was found in the 1970's using more precise measurements by Vera Rubin and Kent Ford of nearly edge-on spiral galaxies where the stellar velocity could be directly measured without projection effects, extending the rotation curve measurements out to larger radii with significantly less uncertainties (Rubin et al., 1970). Further explorations using radio dishes to map the extended neutral hydrogen (HI) gas discs revealed that the flatness extended to some 20-30 kpc in radii (Gottesman et al., 1966; Roberts, 1966; Roberts et al., 1975). By the 1980's, the existence of dark matter was recognized as a significant challenge worth investigating (Rubin et al., 1980). Arguably the best evidence to date for the existence of dark matter is the Bullet Cluster, which remains a primary challenge to modified theories of Newtonian gravity (MOND, Milgrom, 1983; Angus et al., 2006; Lee et al., 2010; Thompson et al., 2015). While in optical light there are two galaxy clusters with opposing group velocities, X-ray observations indicate clumps of shocked hot gas which must have collided as the collision-less galaxies passed by each other (Clowe et al., 2004; Markevitch et al., 2004). Dark matter

15 Today that value is about 10 times lower due to revisions in the Hubble constant that affects subsequent cosmological measurements.

can be indirectly measured by the extent to which background sources are gravitationally lensed and revealed that the dark matter is co-incident with the galaxies, suggesting that dark matter interacts gravitationally with zero electromagnetic cross-section.

Further unexpected disruption to the classical picture of our universe came in 1998 from recession velocities of $z \approx 0.2 - 1$ galaxies whose distances have been independently secured by observations of Type Ia supernovae. As reported in Riess et al. (1998), Perlmutter et al. (1998), and Schmidt et al. (1998), the expansion of space has been *accelerating* over the last 4 Gyr. The discovery re-instated Einstein's cosmological constant, but with a positive contribution. While the nature of dark energy and the surprisingly recent start of the acceleration era have been intensely studied, a consensus view has yet to emerge. Yet, these pieces of evidence lead to our best phenomenological model to date –Λ cold dark matter (ΛCDM) – which has been nearly ubiquitously adopted as the concordance model of cosmology. According to ΛCDM, the matter density has shaped the geometry of the universe to be nearly at the critical density ($\Omega \approx 1$ or 'flat') and is composed of 68.5% dark energy and 31.7% matter; 85% of which is composed of dark matter and only 15% is baryonic matter (Planck Collaboration et al., 2020).

1.5 OUR EVOLVING COSMOS

The discovery of the island universes by Hubble in 1929 not only enlarged our Universe, but gave way to a fundamentally new area of study. Our galaxy was one of an untold multitude of galaxies; and Hubble's discovery meant that we had a birds' eye view of galaxies *from without*. Already by 1926 it was clear to Hubble, and others, that the observed properties of galaxies are broadly bimodal (Figure 1.9). In what is now referred to as the Hubble sequence (or Hubble's *tuning fork*), galaxies are first separated into classes, ellipticals and spirals, out of a continuity of galaxy morphologies. The nomenclature referring to ellipticals as 'early-type' and spirals as 'late-type' has led to a common, but false belief that Hubble was proposing an evolutionary sequence. On the contrary, Hubble (1927) clearly states that no temporal progression is intended (see Baldry 2008).

Figure 1.9: Image of the variety of galaxies in the Hickson Compact Group 87 from *Hubble* and Gemini-South. *Credit: Judy Schmidt.*

Hubble made a distinction between spiral galaxies that feature a central stellar bar (SB) and those that did not (S). He further distinguished them by the number and winding density of their spiral arms from many tightly wound arms (Sa or SBa) to a few loosely wound (Sc or SBc). They constitute $\approx$ 60% of galaxies in the present day universe and are not found in dense cluster environments. Morphologically, spiral galaxies are disc-like. Stars follow coherent orbital motion with low velocity dispersion (i.e. $(v/\sigma)_* > 1$) and are formed from cold molecular gas clouds that undergo compression by coherent density waves within the plane of the disc, requiring clouds of molecular hydrogen within the interstellar medium (ISM) which are both thermally cold (Jeans, 1902) and kinematically cold (Toomre, 1964) to enable collapse (Sanders et al., 1984, 1985). Massive O- and B-type stars are short lived ($\lesssim 100\,$Myr), and so with few exceptions[16] effectively trace these density waves, manifesting as spiral arms. Similarly structured dust lanes are observed in the disc. A relatively blue continuum and weak Balmer break indicate that the luminosity is dominated by these O- and B-type stars. Strong Hydrogen emission lines indicate photoionization of gas clouds by ionizing emission from these same stars, although contribution from harder ionization sources is possible. Dust obscuration varies significantly from relatively dust-free star-formation ($A_V \lesssim 1$ or $\mathrm{N}_{\mathrm{H}_2} \lesssim 10^{20}\,\mathrm{cm}^{-2}$) to deeply obscured regions ($A_V >$ 10 000 or $\mathrm{N}_{\mathrm{H}_2} \gtrsim 10^{26}\,\mathrm{cm}^{-2}$; e.g. in the nucleus of Arp 220 from Scoville et al. 2017b). Some star-forming galaxies host central bulges dominated by older

16 Fast gamma ray bursts are theorized to originate from high-mass stars, but are observed far outside the disc, implying a kicking action.

stellar populations in typically more axisymmetric orbits with a shallower surface brightness profile relative to the disc ($n \sim 4$ and ~ 1, respectively; Sérsic 1963a). As such, the bulge and discs can be separated not only by their profiles and colors, but also by their kinematics. Their mass ranges from $\mathcal{M} \approx 10^9\,\mathcal{M}_\odot$ spirals to massive $\mathcal{M} \approx 10^{12}\,\mathcal{M}_\odot$, with even lower masses being typically dominated by irregular star-forming galaxies lacking coherent spiral structure and associated kinematics.

Hubble distinguished ellipticals by their shape, more specifically their ellipticities[17] whereby spheroid are classified as E0 and highly elliptical objects as E7. They constitute $\approx 40\%$ of galaxies in the present universe and tend to be found in dense cluster environments. Ellipticals feature red optical colors, implying a predominantly red, older stellar population of low-to-intermediate mass stars which outlive more massive, blue stars. The observed ubiquity of strong Hydrogen absorption features (notably Hδ) from stellar atmospheres, lack of > 13.6 eV ionizing flux from O- and B-types, and strong 4000Å break together confirm the typical older stellar demographics within elliptical systems. A lack of strong Hydrogen emission lines driven by photoionization have revealed a typically weak or nonexistent cold molecular gas resevoir. Their optical light tends to be only weakly attenuated by dust grains entrained in the gas, with correspondingly minimal far-infrared emission from reprocessing by this dust. The lack of particularly cold molecular gas available to form stars is sufficient to explain the lack of star-formation, implying a strong correlation between quiescence and elliptical morphology. The expected Gyr lifetimes for the remaining old, red stars make weak gravitational deflections more probable, which impart a non-zero normal velocity component. Such dynamical heating helps explain the non-axisymmetric orbits of old stars, and why the kinematical structure of ellipticals is dominated by random motions (i.e. $(v/\sigma)_* < 1$). However, a more precise examination of their overall projected angular momentum of their stars found that most follow coherent rotation (so-called 'fast rotators') with a minority of systems showing disturbed kinematical signatures such as de-coupled cores ('slow rotators') (Emsellem et al., 2007, 2011; Khochfar et al., 2011; Weijmans et al., 2014). Ellipticals generally have a smooth light profile that falls off more quickly than in spirals ($n \sim 4$; Sérsic 1963a). While their sizes, measured by effective radius (typically taken at half of the total light), vary considerably, elliptical galaxies are an exceptionally homogeneous population. We observe an apparent correlation for elliptical galaxies between stellar velocity dispersion (σ_*), effective radius ($R_{\rm eff}$), and surface brightness known as the Fundamental Plane: $R_{\rm eff} \propto \sigma_*\, I_{\rm eff}$ (Faber et al., 1976; Djorgovski et al., 1987; Dressler et al., 1987; Bernardi et al., 2003). Their mass ranges from $\mathcal{M} \approx 10^7\,\mathcal{M}_\odot$ dwarf ellipticals to massive $\mathcal{M} \approx 10^{13}\,\mathcal{M}_\odot$ ellipticals, with mass-to-light ratios that increase with dominant stellar age.

17 $e = 1 - b/a$ where a and b are the semi-major and semi-minor axes, respectively.

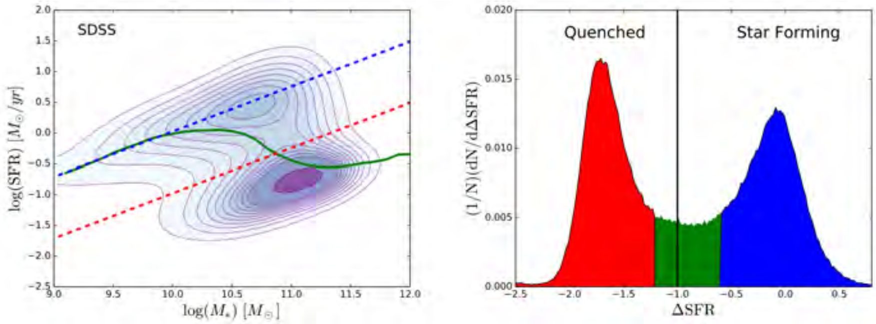

Figure 1.10: *Left:* Joint distribution of stellar mass (M_*) and star-formation rate (SFR) of a sample of $z \sim 0$ galaxies from the Sloan Digital Sky Survey. While the blue locus indicates the average SFR of star-forming galaxies at a given M_*, the red locus shows the same for quenched galaxies, with the overall median SFR at a given M_* is shown by the solid green curve. *Right:* Distribution of the distance of galaxies from the blue locus. *From Bluck et al. 2016*.

The observation that galaxy bimodality (Figure 1.10) is extensively correlated across morphology, age, kinematics, environment, and gas content is as surprising as it is fundamental. The fact that only rarely do we witness objects with a mixture of properties implies that the transformation from star-forming blue spiral to quiescent red elliptical must be rapid and appears to be in some way dependent on local environment (Hogg et al., 2003, 2004) and stellar mass (Baldry et al., 2006). In addition, star-formation requires available cold molecular gas with sufficient density to form stars (Kennicutt, 1998; Bigiel et al., 2008; Leroy et al., 2008; de los Reyes et al., 2019; Kennicutt et al., 2021). The precise role and interplay between these variables is a matter of intense investigation requiring a host of multi-wavelength data, spectro-kinematics, and numerical modelling.

Progress as to the causal connections behind these correlations came by recognizing that intrinsically brighter star-forming galaxies tend to be bluer (and be more FIR luminous), and so they are correlated along a sequence in a magnitude-color diagram. This observational space is broadly physically analogous to mass and star-formation where the sequence in $M_* - SFR$ is preserved, known as the Main Sequence (MS, Brinchmann et al. 2004; Daddi et al. 2007; Elbaz et al. 2007; Noeske et al. 2007; Salim et al. 2007; see also Figure 1.11 from Popesso et al. 2022 and references therein). Importantly, the existence of the MS indicates that the essentially instantaneous characteristic SFR is connected to its previous mass accumulated up to the epoch of observation. This relation persists across a wide range of masses and extends back to earlier epochs where gas supply was generally greater with correspondingly faster assembly timescales (Magdis et al., 2010; Reddy et al., 2012; Whitaker et al., 2012; Rodighiero et al., 2014; Speagle et al., 2014; Whitaker et al., 2014; Ilbert et al., 2015a; Renzini et al., 2015; Santini et al., 2017; Davidzon et al., 2018; Leslie et al., 2020b). The slope of the MS is easily understood since more massive star-forming galaxies generally have more molecular gas ($M_{\rm mol} - M_*$ relation, Bothwell et al., 2014; Saintonge et al., 2017), and so can simply form more stellar mass in a given epoch. The slope is surprisingly stable over look-back time. A variation of the same diagram removes this dependency by normalizing the *SFR* by mass ($sSFR \equiv SFR/\mathcal{M}_*$), such that the MS by *sSFR* has little if any dependence on mass, noting that some studies have found a turnover at extremely high masses (e.g. Lee et al., 2015; McPartland et al., 2019). It's normalization, however, has been steadily decreasing over time such that local starbursting galaxies which lie above the MS have *SFR*s more similar to main-sequence galaxies at earlier times. Although originally interpreted as an evolutionary sequence by which galaxies grow in *SFR* with mass (a natural feedback mechanism), the fact that the normalization decreases with time up-ends this picture entirely. Instead, as pointed out by Gladders et al. (2013) and Abramson et al. (2016), star-forming galaxies at have tended to *lose SFR* over time (either at fixed mass or with increasing mass), with the notable exception of starbursts[18]. Although dependent on the tracer used to calculate *SFR*, the width of the MS appears to surprisingly narrow and is likely to have some causal connection with gas supply (Tacchella et al., 2016, 2020) with progress coming from detailed modelling of star-formation histories (Abramson et al., 2016; Iyer et al., 2018; Caplar et al., 2019). Other regulatory mechanisms have been proposed to explain this remarkable stability in slope and evolution in normalization including accretion by dark matter halos (Dutton et al., 2010; Forbes et al., 2014), and increasing metalicities (Mannucci et al., 2010; Lilly et al., 2013).

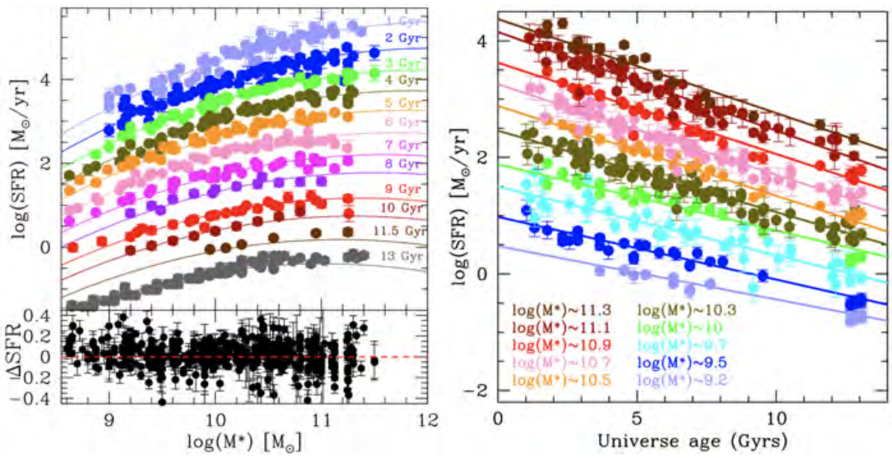

Figure 1.11: *Left*: Star-forming Main Sequence estimates at various cosmic ages and best-fit model, with residuals below. *Right*: Typical star-formation rates along the main sequence in bins of mass as a function of cosmic age. For clarity, relations in both panels are displaced by 0.3 dex. *From Popesso et al. 2022.*

Relative to star-forming galaxies at the same epoch, quiescent ellipticals tend to be more massive and less star-forming, and so lie below the main

18 This is debated; e.g., Robaina et al. (2009) finds that while some mergers at $z \approx 0.6$ produce large bursts, they contribute only a modest *SFR* enhancement over the entire merger timescale.

sequence[19]. It is thought that these systems went through a period of star-formation earlier in the history of the universe and ceased their star-formation more than 100 Myr earlier (i.e. the main sequence lifetime of B-type stars). Reasons for this apparent cessation of star-formation (or 'quenching') are not yet fully understood. One promising target population to investigate this transition are so-called post-starburst galaxies[20] (Zabludoff et al., 1996; Goto, 2005; Wild et al., 2009; French, 2021; Almaini et al., 2017/Aug/04/). Post-starburst galaxies are rare especially in the local universe (Wild et al., 2016) but occur at all environments and redshifits. They have historically been identified spectroscopically[21] by their intermediate 4000 Åbreak strength and strong Hδ absorption indicative of a middling stellar population dominated by A+F type stars . Morphologies of post-starburst galaxies are typically early-type, with an overall higher rate of occurance in overdense environments (Zabludoff et al., 1996; Wilkinson et al., 2017; Wilkinson et al., 2021; Werle et al., 2022). The distribution of shapes measured by a variety of structural parameters (e.g., concentration, asymmetry, clumpiness; Conselice 2003) indicates elevated rates of post-merger signatures compared to control samples (Pawlik et al., 2016). While historically only the brightest post-starburst galaxies could be spectroscopically confirmed and studied en masse, new selection techniques are lessening this observational bias by including less luminous sources (Maltby et al., 2016; Wilkinson et al., 2021).

The strongest post-starburst galaxies in the local universe (i.e. those most easily identified) appear to be ubiquitously related to gas-rich merger remnants, and predominantely feature post-starburst characteristics in their central regions (Bekki et al., 2005; Wild et al., 2009; Snyder et al., 2011). While large photometric galaxy surveys have been fruitful in establishing large statistical relations, they tend to marginalize over the profile of the galaxy with limited ability to understand e.g., radial stellar population gradients. Although belabored by difficulties in sample identification and selection biases (Greene et al., 2021), investigations into the radial stellar population gradients of post-starburst galaxies have provided a treasure trove of valuable insights into the starburst configuration and radial evolution (Belfiore et al., 2018; Rowlands et al., 2018; Chen et al., 2019). Recent discovery of surprisingly large, sustained molecular gas resevoirs in nearby post-starburst galaxies by French et al. (2015) suggest that simple depletion of gas is not alone responsible for ceasing star-formation in these systems, and puts into question the simple assumption that post-starbursts are a coherent quenching population (in agreement with simulations, see Pawlik et al. 2019). Generally, the gas has found to be in a state unfit for star-formation (Ellison et al., 2018).This has led to wide speculation about their future pathway(s) by coupling spatially-resolved star-formation

19 Although they seem to form a so-called 'Red Sequence', this is due to the observationally imposed upper limit in SFR measurements for these low-SFR systems, and they likely form a continuum towards even lower SFRs.

20 Post-starburst galaxies are also called 'E+A' because their spectra are like a juxtaposition of those of ellipticals and A-type stars (Dressler et al., 1983).

21 Exceptions exist especially at $z \gtrsim 1$ where candidates are typically selected by color-color (Maltby et al., 2016; Wu et al., 2020), see also Wild et al. (2016).

with gas surface densities resolved on similar scales (Masters et al., 2019a; French et al., 2022; Smercina et al., 2022) with mechanisms including turbulent heating and shearing of gas clouds possibly playing a role in suppressing residual star-formation. Valuable context as to the physical processes responsible continues to emerge from detailed simulations (e.g. Rodríguez Montero et al., 2019; Zheng et al., 2020).

New insights into these and other spatially complex systems have been provided by relatively new integral field spectroscopic instruments (e.g. SAURON, VIMOS, MUSE) and related surveys (e.g. Fornax3D, Sarzi et al. 2018; TIMER, Gadotti et al. 2019) as well as multiplexed spectroscopic surveys of more distant objects (e.g. Atlas3D, Cappellari et al. 2011; CALIFA,Sánchez et al. 2012; MaNGA, Bundy et al. 2015) have demonstrated that galaxies are far more complex than previous thought, even compared to previous studied using single fiber instruments (e.g. SDSS DR7, Brinchmann et al. 2004). Multiplexed low-resolution grism surveys have also been made possible from *Hubble* (e.g. 3D-HST, Brammer et al. 2012; Weaver et al. 2017; FIGS, Pirzkal et al. 2018) containing many thousands of galaxies, the most distant of which are only seen by their emission lines. Although dedicated spatially-resolved spectroscopy is now being taken on galaxy samples at $z \approx 2$ (e.g. KMOS, Wisnioski et al. 2015; MOSDEF, Kriek et al. 2015) and a handful of lensed systems at $z \sim 1-4$ (e.g. GLASS, Treu et al. 2015; REQIUEM, Akhshik et al. 2020), the vast majority of high redshift galaxies will be studied photometrically, and so these insights from spatially-resolved spectroscopic studies of the low-z universe (as well as simulations) are crucial in interpreting the physical processes shaping their properties.

Figure 1.12: The archetypal major merger remnant NGC 7252 as seen by the MPG/ESO telescope (*left*), with a zoom-in by *Hubble* (*right*). Tidal tails, in-situ star-formation, and shell structures are visible. The nuclear region is dominated by a star-forming mini-spiral with strong dust obscuration. *For details, see Weaver et al. 2018.*

One of the most dramatic mechanisms responsible for transforming galaxies is galaxy-galaxy mergers (Figure 1.12). Known since some of the first deep sky images (Arp, 1966), the study of galaxy mergers both observationally (Schweizer, 1982; Le Fèvre et al., 2000; Bell et al., 2006b; Darg et al., 2010; Whitmore et al., 2010; Lambas et al., 2012; Schweizer et al., 2013; Weaver et al., 2018) and theoretically (Barnes et al., 1991; Borne et al., 1991; Barnes et al., 1992a, 1996; Mihos et al., 1996; Chilingarian et al., 2010; Renaud et al., 2019) has resulted in a broad understanding of the physical processes at play. Minor mergers involving a lower-$\mathcal{M}$ galaxy being accreted onto a more massive galaxy, while essentially destroying the former also generally leaves the latter unchanged. Major mergers involving galaxies of similar masses is much more powerful and dramatic. Early work by Toomre et al. (1972) established a sequence of events from the first pericenter passage of the merging systems, subsequent passes, coalescence, and relaxation. Major mergers fall into two broad categories. One one hand, gas-poor galaxies undergo so-called 'dry mergers' which are effective at increasing mass but do not induce star-formation (Bell et al., 2006a; Ciotti et al., 2007; Khochfar et al., 2009; Ruszkowski et al., 2009). Gas-rich galaxies, on the other hand, are capable of increasing the mass of the remnant and form new stars from the colliding gas (Lambas et al., 2003; Nikolic et al., 2004; Barton et al., 2007; Lin et al., 2007, 2008). Observations of major mergers along the Toomre Sequence reveals ubiquitous debris streams, potentially with knots of new star-formation (Larson et al., 1978; Barnes et al., 1992b; Whitmore et al., 1995; Barton et al., 2000). While less efficient than the merger interaction itself (Barnes, 1988; Wright et al., 1990; Hernquist et al., 1991; Hibbard et al., 1996), latent gravitational heating by repeated weak interactions cause orbits of remaining stellar populations to loose coherence leading to dispersion-dominated systems (Nordström et al., 2004; Leaman et al., 2017; Shetty et al., 2020). Although the coalescence phase and associated kinematical destiny of the remnant is highly dependent on the merger orientation (Cox et al., 2008), tidal forces rapidly cause gas clouds loose angular momentum such that they fall deeper into the potential of the remnant (Emsellem et al., 2015; Combes, 2019) and can become ro-vibrationally excited by shocks (Farage et al., 2010; Alatalo et al., 2016; U et al., 2019). Remaining gas that collects within the core region may ignite a nuclear star-burst or even a star-forming disc (Schweizer, 1982; U et al., 2019), with new dust being produced in-situ (Leśniewska et al., 2019; Martínez-González et al., 2022). A fraction of this material may be driven further into the core where it can be accreted by the central supermassive black hole to form a luminous accretion disc seen as an AGN (Sanders et al., 1988; Di Matteo et al., 2005; Croton et al., 2006; Hewlett et al., 2017/May/10/). Powerful winds and/or jets may be launched by the black hole which expels heated gas from the disc into the halo whereupon it is unavailable to produce stars over potentially Gyr timescales (Murray et al., 1995; Proga et al., 2000), although cold gas ejections have been observed (Tremblay et al., 2018). This still emerging paradigm is far from complete, due in part to the exceptionally wide range in properties seen in merger remnants which combined with their rarity makes their study difficult. While a consensus has not yet emerged, it seems likely that mergers can transform star-forming spirals into more massive quiescent ellipticals (e.g. Duc et al., 2015), although other pathways certainly exist (e.g. Martig et al., 2009). The future of the gas reservoir and associated star-formation, and therefore the transition from blue to red, is generally unclear (French et al., 2015; Weaver et al., 2018; Smercina et al., 2022). Furthermore, the formation, structure, behaviour, and impact of the active galactic nucleus and nuclear star-formation remain a matter of debate (e.g. Neumayer et al., 2007; Bundy et al., 2008; Dubois et al., 2013). Yet it seems unlikely that major mergers are common enough to be the dominant mechanism for transforming galaxies (Weigel et al., 2017), and so additional mechanisms are needed.

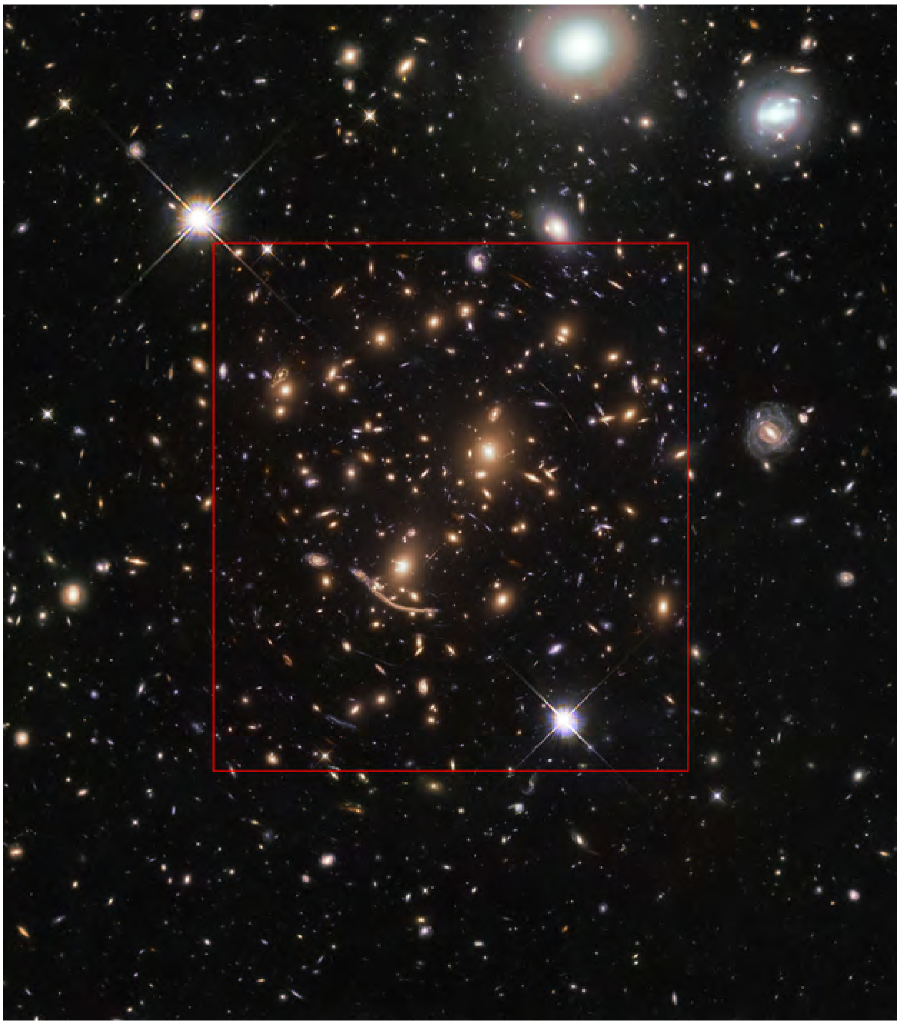

Figure 1.13: Composite image of the Abell 370 galaxy cluster. The innermost region of the cluster is dominated by massive quiescent galaxies, whose combined baryonic and dark mass creates a gravitational lens. Several bright, lensed background objects can be seen, making such systems invaluable for accessing galaxies in the distant universe which would be considerably fainter and unresolved in typical wide-field imaging. Recently, the BUFFALO program (Steinhardt et al., 2020a) has expanded *Hubble* existing imaging on Abell 370 (red box), as well as on five other clusters, to better understand the cluster properties, map out dark matter substructures, and to identify additional lensed objects. *Credit: NASA, ESA, A. Koekemoer, M. Jauzac, C. Steinhardt, and the BUFFALO team.*

The fact that galaxy clusters are dominated by quiescent elliptical galaxies was already recognized by Hubble et al. (1931). Subsequent observations of the local universe revealed that higher density environments, like clusters, preferentially host quiescent elliptical galaxies (Abell, 1958; Schneider et al., 1983; Dressler, 1984; Abell et al., 1989), see Figure 1.13. This apparent correlation between environmental density and quiescence suggests that perhaps mergers or other physical processes which manifest in cluster environments may be responsible for the growth and quiescence of these cluster systems (Butcher et al., 1984; Balogh et al., 2000; Gnedin, 2003). Ram pressure stripping, for example, has been observed in gas-rich satellite galaxies falling into the cluster potential causing shock-induced star-formation and stripping of the gas reservoir by the hot intracluster medium (ICM; see Figure 1.14 and Weaver et al. 2016), followed by thermal evaporation of the remaining gas (Chanamé et al., 2000; Lewis et al., 2002; Lee-Waddell et al., 2018; Boselli et al., 2022). Tidal effects such as harassment and truncation may also play a role in cluster environments (Gott, 1975; Moore et al., 1996, 1998; Bekki et al., 2001; Marcillac et al., 2007; Bialas et al., 2015). While the material from infalling systems may eventually become accreted by the more massive, quiescent systems near the core and thereby contribute to incremental mass growth (Hausman et al., 1978; Lin et al., 2004; Iodice et al., 2017), the majority of this material likely remains in ICM (Burke et al., 2015; Morishita et al., 2017). Together, these mechanisms help explain the observation that the closer a galaxy is to the cluster core, the less likely it is to retain star-formation (Balogh et al., 1999).

Figure 1.14: The Fornax Cluster is one of the foremost laboratories for studying accelerated galaxy evolution in dense environments. Here the immense gravitational potential created by the massive central-most elliptical galaxies has attracted the dwarf irregular galaxy NGC 1427A. Its once dormant supply of gas has been ignited likely as a result of ram-pressure stripping by the hot intracluster medium, making it one of the most UV-bright sources in the night sky. Understanding the repercussions of such interactions is key to uncovering the late time evolution of galaxy clusters. *Credit: CTIO/NOIRLab/DOE/NSF/AURA.*

Despite significant effort, the remarkable mass assembly of the innermost, quiescent cluster galaxies in the present-day universe remains an unsolved problem preventing us from realizing a unified theory of galaxy evolution (De Lucia et al., 2007; Toft et al., 2014). Hints from the limited number of rare galaxy clusters in the local universe may be expanded by understanding their formation and assembly histories (Thomas et al., 2005). To do so requires looking back to a younger universe where such virialized structures have only just begun to assemble.

1.6 TOWARDS A CENSUS OF GALAXY EVOLUTION

Until recently, detailed observations of distant galaxies and their derived properties were limited to spectroscopic measurements which could provide a redshift from which observed properties could be translated into physical ones through empirical relations. However, spectroscopy is incredibly time consuming and so limited the speed by which large statistical samples could be acquired. While nearly all of the brightest nearby galaxies were thoroughly dissected with spectroscopy, there was a dearth of data on nearby low-surface brightness galaxies as well as distant galaxies with faint apparent brightnesses. The result was a strong degree of bias and incompleteness in available data. Plagued by non-negligible selection effects, studies struggled to generalize their findings based on only a handful of galaxies to understand the demographics and behaviour of the larger population.

Figure 1.15: *Left*: Patch of sky in Ursa Major with the Hubble Deep Field footprint. *Right*: The Hubble Deep Field was the first blank field deep observation by *Hubble*. Taken in 1995, the field covers ~ 5 sq. armin and yet demonstrates the surprising diversity of galaxy morphologies in the distant universe, which until *Hubble* could not be discerned from ground-based observations. It remained the deepest astronomical image available until 2012, when it was eclipsed by the Hubble eXtreme Deep Field. *Credit: R. Williams, STScI, and NASA.*

Large representative samples of millions of galaxies are now routinely measured over cosmologically significant volumes capable of probing a range of environments. The largest of these include the Sloan Digital Sky Survey (SDSS; York et al., 2000; Abazajian et al., 2009; Ahn et al., 2014), the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Kaiser et al., 2000; Kaiser et al., 2010), the Dark Energy Survey (DES; Flaugher, 2005; Diehl et al., 2012; Flaugher et al., 2012), the Dark Energy Spectroscopic Instrument Legacy Surveys (Dey et al., 2019a), and soon the Legacy Survey of Space and Time (LSST; Tyson et al., 2003; Ivezić et al., 2019). While these surveys have been enormously effective at returning representative samples of the low-redshift universe ($z \lesssim 2$) as well as rare systems including major mergers, active galactic nuclei (AGN), quasars, and post-starburst galaxies, they lack the photometric depth and near-infrared sensitivities required to explore the distant, high-redshift universe at $z \gg 2$. Space-borne observatories such as *Hubble* have revealed much in this direction. By using a space-based platform above Earth's atmosphere, *Hubble* is able to efficiently capture the light from faint sources. Not only that, but the diffraction-limited imaging from *Hubble* has permitted precise studies tracing the morphological evolution of galaxies into the distant universe (e.g. Oesch et al., 2010b) and has resolved structures within distant ($z \approx 2-5$) galaxies, out of reach by even the largest ground-based facilities. The invent of blind, deep fields in 1995 with the *Hubble Deep Field* (Williams et al., 1996) was a watershed moment and provided not only fundamentally new avenues of research into both distant and low surface-brightness galaxies alike (Figure 1.15). Successively deeper imaging further explored this little known regime, such as the *Hubble Ultra-Deep Field* containing a sample of $\sim 10\,000$ galaxies (Beckwith et al., 2006). This survey strategy was industrialized by the Cosmic Assembly Near-infrared Deep Extragalactic Survey (CANDELS, Grogin et al., 2011; Koekemoer et al., 2011), which provided deep near-infrared imaging over five fields totalling $\sim 800\,\mathrm{arcmin}^2$, enabled by the then new Advanced Camera for Surveys (ACS) instrument. Selection in the near-infrared directly measured the rest-frame stellar bulk out to $z \lesssim 3$ and enabled mass-selected samples from space (Tomczak et al., 2014; Santini et al., 2015). The fields were augmented over time with some of the deepest multi-wavelength data from X-rays to radio. Although CANDELS revolutionized extragalactic astronomy in a number of ways, not least the study of galaxy morphologies (Lee et al., 2013; van der Wel et al., 2014) and the study of distant ($z > 4$) galaxies (Bouwens et al., 2015), it probed a relatively small volume and consequently suffered from a limited ability to assess possible biases in sample properties (so-called "cosmic variance", Trenti et al. 2008a; Driver et al. 2010; Moster et al. 2011a).

At the same time, the Cosmic Evolution Survey (COSMOS, Scoville et al. 2007b) was begun but with a markedly different strategy: instead of focusing on several small fields, COSMOS would drill into a single 2 deg² field in order to probe a range of contiguous environments and effectively counter cosmic variance. Although a 600 arcmin² portion was covered by multiple near-infrared bands with CANDELS, the larger 1.6 deg² received only F814W imaging at comparably lower depth (Koekemoer et al., 2007a). Consequently, COSMOS was proven as a test-bed for pioneering measurements of weak lensing and cosmic shear, improving our understanding of dark matter and large scale structure (Leauthaud et al., 2007). Ground-based facilities, in lower demand than *Hubble*, filled in the gaps by providing deep UV and optical coverage (CFHT, Subaru/SC and HSC) as well as near-infrared imaging (VISTA), which, provided by the UltraVISTA survey (McCracken et al., 2012), makes

COSMOS the deepest near-infrared survey over a contiguous 2 sq. degrees[22] (Figure 1.16). Several dedicated surveys from *Spitzer's* InfraRed Array Camera (IRAC) provided important infrared imaging coverage to measure dust ($z < 2$), the Balmer/4000Å break ($z \approx 2$), rest-frame near-infrared light from the stellar bulk ($2 \lesssim z \lesssim 5$), and rest-frame UV light from star-forming regions and AGN ($z \gtrsim 8$), on top of comparably shallower mid-infrared imaging (see compilation by Moneti et al., 2021, for details). A similar approach was taken to explore the Spitzer XMM-Newton Deep Field (SXDF, e.g. Mehta et al. 2018), which to date is the only comparable degree-scale field with sufficiently deep near-infrared coverage to establish mass-selected samples out to high-redshift ($z \gg 2$).

22 UltraVISTA is complemented by the shallow 1 500 sq. degree VIKING survey (Edge et al., 2013) and medium depth 12 sq. degree VIDEO survey (Jarvis et al., 2013).

Figure 1.16: *Upper:* The ESO UltraVISTA Public Survey is a near-infrared $YJHK_s$ survey of the COSMOS field McCracken et al. (2012). It covers 2 deg^2, including NB118 coverage in four stripes Milvang-Jensen et al. (2013). As of Data Release 4 (Moneti et al., 2019), UltraVISTA is the longest-running ESO survey spanning 2009-2016. *Lower:* Six zoom-in panels illustrate the diversity of galaxies and environments found in COSMOS by UltraVISTA. *Credit: H. J. McCracken, A. Moneti, and ESO.*

The compilation of such broad multi-wavelength data freed galaxy studies from their reliance on spectroscopic redshifts. Until this point, identifying high-redshift galaxies from optical data relied on so-called 'dropout' technique whereby the $912 - 1000$ Å Lyman break (with absorption by the neutral IGM at $z \gtrsim 5 - 6$) causes source light to drop out of successively blue bands (e.g., r-band corresponds to $z \approx 4$, Steidel et al. 1996a). Although first introduced by Baum (1962) to estimate cluster distances, only within the last two decades have photometric redshifts become the standard approach to measuring distances of galaxies[23] (Mobasher et al., 2004, 2007). The fact that redshfits could be massively multiplexed using photometry with low spectral resolution ($R = \lambda/\Delta\lambda \sim 5 - 10$) provided a cheap alternative which enabled the computation of redshifts and associated physical parameters for many millions of galaxies simultaneously. Early work by Connolly et al. (1995) established iso-z layers in the multidimensional color-color space of observed photometry, from which reliable redshifts could be estimated. While the underlying approach has remained, the practical techniques have evolved significantly. The modern approach is to match the photometrically measured spectral energy distributions (SEDs) to one or a combination of many spectral templates selected from a sufficiently complete library of galaxy spectral templates (e.g. Leitherer et al., 1999; Bruzual et al., 2003; Conroy et al., 2009a, 2010a), with redshift as one of potentially dozens of fitted parameters. The reliability of the redshift estimates, as well as intrinsic physical parameters such as $\mathcal{M}_*$ and *SFR* (for techniques, see Salvato et al. 2019 and Conroy 2013, respectively), depend strongly not only on the information contained in the SEDs (e.g., emission lines, number of filters, depths, and wavelength domain), but also on the templates assumed. Generally there are two approaches. Historically, libraries of observed spectra provided high-resolution templates. Yet while they excel at describing the SEDs of known objects, they are poorly suited to describe objects whose SEDs are not described in the library, and so discovery spaces are bent towards systems which are already known. A better understanding of stellar evolution, gas, and dust from new facilities enabled multi-wavelength SEDs built from theoretical frameworks. While these synthetic galaxy templates are able to describe any SED within the bounds of known astrophysics, they suffer from highly uncertain physical recipes, not the least of which is the injection of gas-excitation emission lines which are physically de-coupled from the bulk of the stellar continuum. Likewise, the number of metal absorption lines treated in these recipes is surely fewer than those that exist in real systems. The application of dust screens remains largely phenomenological with prescriptions derived from only a handful of nearby galactic systems which already contain significant variance in attenuation at different wavelengths (Nandy et al., 1975; Seaton, 1979; Fitzpatrick et al., 1986a; Fitzpatrick, 1986; Calzetti et al., 2000a; Gordon et al., 2003). Recently, a multitude of machine learning applications have been applied to address these shortcomings, and although there has been intense development in this area with promising results, no consensus exists on which are best (Collister et al., 2004; Carliles et al., 2010; Carrasco Kind et al., 2013; D'Isanto et al., 2018; Pasquet et al., 2019). They will be needed soon for large-scale deep surveys of the coming decade, such as *Rubin*-LSST (Brescia et al., 2021).

While the worth of photometric redshfits have been proven, their performance is strongly dependent on the quality of the photometric calibration and sources of uncertainty from the initial reductions through to the pho-

23 The term *photmetric redshift* was first used in Puschell et al. (1982) who estimated distances to faint radio galaxies with broad-band photometry, appearing later in the title of Loh et al. (1986).

tometry and associated corrections. Uncertainties too are important, as are non-detections and upper limits. Calibration of photometry (and/or training in the case of machine learning) is typically performed to maximize the performance of the photometric redshifts against a library of spectroscopic redshifts, although some exceptions exist (e.g. Brammer et al., 2008). This reliance on a known spectroscopic sample is a significant limitation, as like with the empirical templates, can potentially bias the results. In particular, photometric redshifts and associated physical parameters of galaxies in the so-called "redshift desert" ($z \approx 1.4 - 2.5$, where emission lines were historically not available to facilities at the time, see Steidel et al. 2005) can suffer from high calibration uncertainties and poorly understood redshift performance. Several highly multiplexed surveys have been initiated to address this lack of spectroscopy (C3R2, Masters et al. 2017; MOONS, Cirasuolo et al. 2014; PFS, Takada et al. 2014) which will improve calibration of photometric redshifts.

The gain from wide application of photometric redshifts cannot be overstated. Only recently have large, representative samples from a range of environmental densities been acquired from surveys such as SXDS (Furusawa et al., 2008; Mehta et al., 2018) and COSMOS (Scoville et al., 2007a; Laigle et al., 2016; Weaver et al., 2022a). Simultaneously, narrow field surveys of the very distant and especially faint universe have complemented these large-area deep fields by extending samples to intrinsically fainter luminosities, especially from lensed galaxies at the most low-luminosity end[24]. The importance is not to characterize any one galaxy in great detail, but instead leverage the enormous statistical weight of carefully selected samples to precisely estimate statistical relations and distributions from which the physical processes governing galaxy evolution can be inferred and/or compared with theoretical predictions. Similarly, such representative parent samples are ideal for identifying galaxies for targeted spectroscopic follow-up (e.g., ALPINE, Le Fèvre et al. 2019), with surveys such as the VIMOS Ultra-Deep Survey (VUDS, Le Fèvre et al. 2015) and zCOSMOS (Lilly et al., 2007a) providing important spectroscopically precise confirmations of previously photometrically derived relations and distributions at $0.2 < z \lesssim 6$, providing an anchor for photometric surveys of increasingly fainter and more distant samples.

One of the first statistical tools accessible to early, monochromatic galaxy surveys was the number count. Measuring the number density of galaxies within a given survey footprint as a function of observed-frame magnitude, they inform about the distribution of galaxy luminosities (Hubble, 1936a,b; Kiang, 1961; Christensen, 1975; Schechter, 1976a) and have been useful in testing cosmological theories (Brown et al., 1974; Kron, 1980a; Shanks et al., 1984). While magnitude-limited samples extracted from single-band surveys are straight-forward to obtain, galaxies in a given magnitude bin span a range of redshifts and so that bin corresponds to a range of different rest-frame luminosities. This effectively smears out the astrophysical information in

24 Although these analyses develop significant uncertainties when combining data from different fields and selections, see Chapter 4.

with the redshifts. Multi-wavelength surveys and the advent of photometric redshifts helped to break this degeneracy such that distributions of consistent rest-frame luminosities could be compared across redshifts (Benson et al., 2003; Cooray et al., 2005; Marchesini et al., 2007; Marchesini et al., 2012). Such luminosity functions measure the distribution of galaxy luminosities at various fixed redshift ranges. They are, however, more sensitive to selection effects and reshift systematics. Largely relegating number counts to a calibration exercise, luminosity functions continue to be used to understand samples across the electromagnetic spectrum (e.g. Branchesi et al., 2006; Aird et al., 2010) as well as the contribution of various populations (e.g., AGN, Caputi et al. 2007; Georgakakis et al. 2015). Yet, the luminosity function of the rest-frame near-infrared ($1 - 2\,\mu$m) is special as the luminosity directly corresponds to the light emitted by the bulk of the stellar mass (Jenkins et al., 1991; Mobasher et al., 1993; Glazebrook et al., 1994). The first near-infrared, mass-selected samples with redshifts were built by Songaila et al. (1994), and Glazebrook et al. (1995), with Cowie et al. (1996) securing the first mass-complete samples at $z \approx 2 - 3$. In essence, the resulting luminosity function (Kochanek et al., 2001; Caputi et al., 2006b) can be transformed into a stellar mass function by adopting a mass-to-light ratio corresponding to a given IMF (Cole et al., 2001; Bell et al., 2003), although more sophisticated SED fitting codes are able to effectively estimate the stellar mass parametrically[25].

Thanks to these earlier studies, modern galaxy catalogues are now capable of measuring the galaxy stellar mass function (SMF; see Chapter 4), which traces the volume number density distribution of galaxy stellar masses over time. Informing about the mass built and retained in galaxies at a given epoch, the SMF encodes the mass assembly history of the universe, and is sensitive to feedback processes including supernovae, black holes, and dynamical heating in dark matter halos. At $z < 1$, the SMF is described by a shallow (logarithmic) slope which turns over at a characteristic mass (the 'knee', $\mathcal{M}^* \approx 10^{10.5}\,\mathcal{M}_\odot$) beyond which increasingly massive galaxies are exponentially rarer per unit volume (see Figure 1.17). However, this picture is far from the universe whose galaxies assembled from cold baryonic accretion onto dark matter haloes (White et al., 1991; Navarro et al., 1996, 1997; Mo et al., 1998). The resulting halo mass function (HMF) derived from theoretical considerations describes a power law-like shape whereby massive galaxies are rare and low-mass (satellite) galaxies are extraordinary common, effectively loosing the 'knee' (Moore et al., 1999; Jenkins et al., 2001; Tinker et al., 2008; Bocquet et al., 2016; Wechsler et al., 2018). Numerous confrontations between observers and theorists struggled to produce meaningful results (Klypin et al., 1999; Simon et al., 2007; McGaugh et al., 2021). Were the mass functions incomplete, as argued by the cosmologists, or were there other physical processes at play in addition to halo accretion? Building on intense effort undertaken by a number of large observational programs, Peng et al. (2010b, 2012) established the

25 Even in flux-limited samples without rest-frame NIR light, SED fitting codes can use flanking rest-frame optical and IR light to robustly estimate mass.

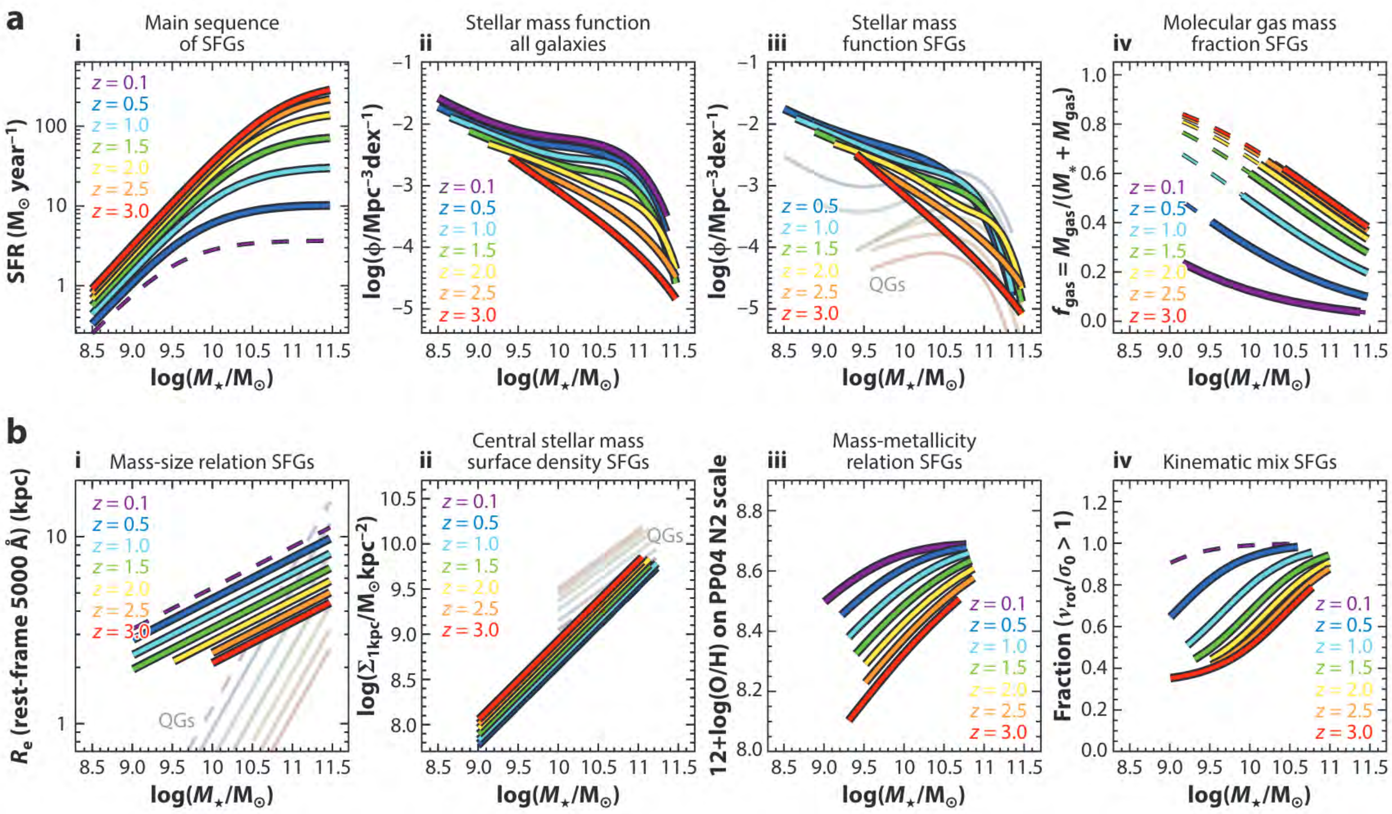

Figure 1.17: Evolution of eight key galaxy scaling relations from $z = 3 \rightarrow 1$, colored by redshift. For data which is not available, dashed lines are used. (*a*) The panels show (*i*) the star-forming main sequence and the galaxy stellar mass functions from Tomczak et al. (2014, 2016) for all galaxies (*ii*) , (*iii*) that divided into star-forming and quiescent samples, (*iv*) molecular gas mass fractions for star-forming galaxies (Tacconi et al., 2020). (*b*) The panels show (*i*) the mass-size relation (van der Wel et al., 2014), (*ii*) the relation between mass and central stellar mass surface density (Barro et al., 2017), (*iii*) stellar mass versus gas-phase metallicity (Wuyts et al., 2014; Zahid et al., 2014), and (*iv*) the fraction of rotation dominated galaxies from ionized gas kinematics (Kassin et al., 2012; Simons et al., 2017; Turner et al., 2017; Wisnioski et al., 2019). *From Figure 7 of Zavala et al. 2021.*

current consensus view. By studying the large volume probed by combining the SDSS and COSMOS surveys, Peng et al. showed that while massive galaxies cease star-formation (i.e. 'quench') regardless of environment, galaxies in dense environments cease star-formation regardless of mass. This simple action is sufficient to describe the growth of quiescent galaxies as seen by the SMF: they seem to dominate at high-masses at $z < 1$ (the limit of the study), there is a rapid growth of quiescent systems with a vanishing fraction seen at earlier times (see panels *ii* − *iii* in Figure 1.17, as well as Ilbert et al. 2013; Davidzon et al. 2017; Muzzin et al. 2013/Aug/16/). This follows with the expected time-scale for the first dense, virialized clusters to form and provide an environment to affect star-formation in low-mass satellite systems. Other mechanisms have been proposed to complement this otherwise phenomenological picture, including most notably the effect of AGN feedback in massive galaxies (Di Matteo et al., 2005; Schawinski et al., 2006; Cheung et al., 2016) and supernovae in low-mass systems (Larson, 1974; Yepes et al., 1997; Springel et al., 2005), with the latter being likely to explain star-formation cessation outside dense environments. If the SMF should coincide with the HMF at any point, it seems likely only in the very early universe ($z \gg 6$) when there has not been sufficient time for these physical processes to take hold (Steinhardt et al., 2016a). Although *JWST* will soon produce the first mass-selected samples of such early galaxies, it is currently unclear whether or not the SMF will coincide with the expected HMF. While the rest-frame UV luminosity functions at $z > 6$ (see Section 1.8) tend to suggest a more power law-like form in agreement with the HMF, translations from the rest-frame UV to stellar mass suffer from increasingly severe observational uncertainties and selection function effects (Harikane et al., 2016; Song et al., 2016; Stefanon et al., 2021b).

Yet, the SMF does not describe all of galaxy evolution. While it indeed tells us valuable information about the mass built up in galaxies by their respective epoch, that mass growth comes predominantly from star-formation. Therefore, tracing the cosmic star-formation rate density (CSFRD) over time therefore provides valuable details to complement the SMF (Madau et al., 1998). This is not straight-forward as instantaneous star-formation traced by UV continuum (which unlike individual emission lines can be directly measured by photometric surveys) can be diminished by dust attenuation within the galaxy (Zavala et al., 2021). While estimating the dust content is possible, far less uncertain is to estimate the grain-processed light which is re-emitted in the FIR. As established in the review by (Madau et al., 2014), the cosmic star-formation rate rose steadily from the early universe to a peak at $z \approx 2$, and has been in decline since then. Notably, the occurrence of mergers is thought to peak around $z \approx 2$, strengthening the links between merger-induced star-formation and rapid cessation thereafter. While the universe was assembling at $z > 2$, galaxy growth and evolution in present-day universe are slow in comparison.

The evolutionary role of massive starbursting galaxies during the peak of CSFRD remains a mystery. Understanding their population demographics, as well as their future state are key questions as they may be progenitors of massive quiescent galaxies in the local universe (Caputi et al., 2006a; Marchesini et al., 2014). Surveying them is made difficult by their large dust attenuation (Caputi et al., 2012, 2015; Whitaker et al., 2017), with observations from ALMA revealing surprising numbers missed in earlier optical-NIR surveys (e.g. Casey et al., 2021; Manning et al., 2022). Our ability to predict their future state depends on understanding the root cause of their star-formation (Kokorev et al., 2021). Do these systems have particularly massive molecular gas reservoirs, or are they just more efficient in forming stars? While investigations into rare local analogues – (Ultra-)Luminous Infrared Galaxies (LIRGs) – have been fruitful, observational challenges (e.g., CO conversion rates, Magdis et al. 2011; Cortzen et al. 2019) have thus far prevented a clear answer at higher redshifts. As demonstrated by Figure 1.18 from Zavala et al. (2021), the contribution of (U)LIRGs to the CSFRD is substantial. Therefore, continued investments in infrared and submillimeter facilities like ALMA and NOEMA will help clarify their role and constrain the star-formation history of the universe.

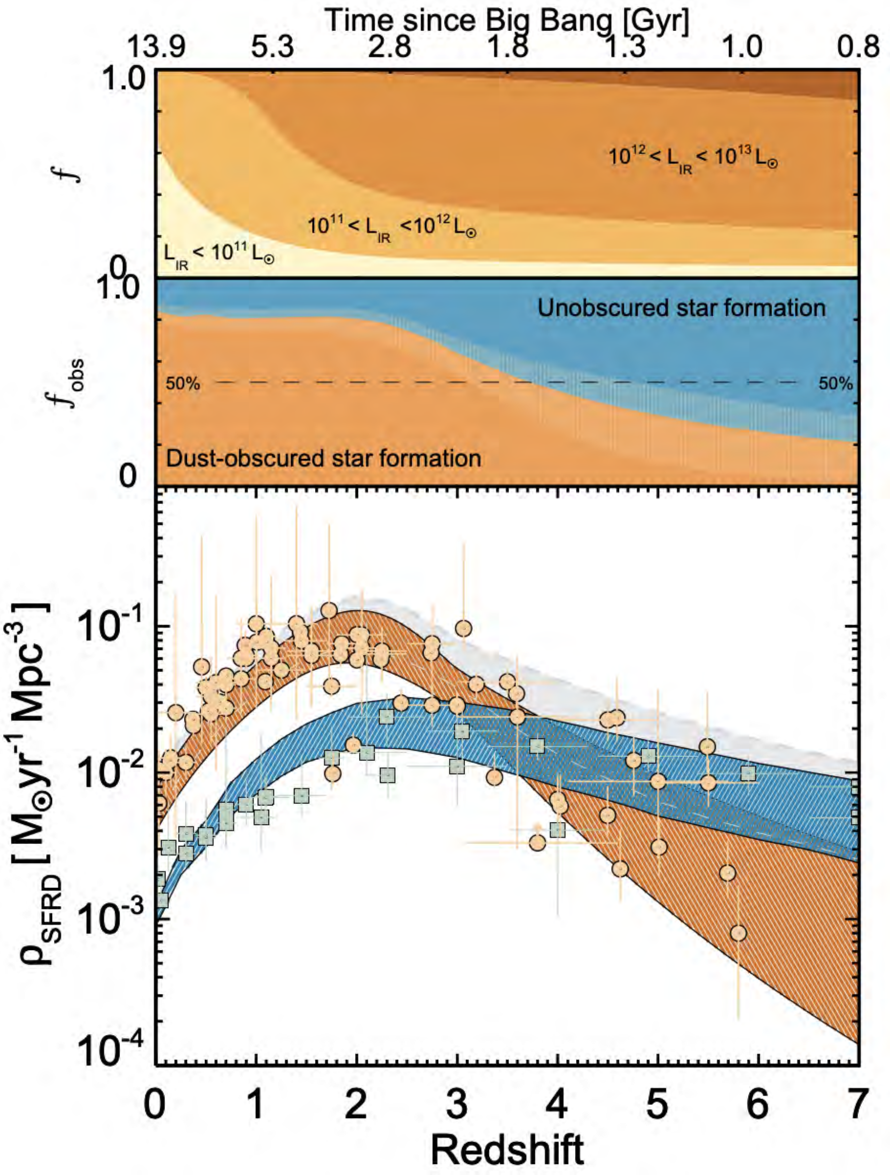

Figure 1.18: *Lower:* Evolution of the cosmic star-formation rate density (CSFRD) inferred from UV-selected surveys, uncorrected for attenuation (blue), and from IR/submillimeter suveys (orange). The total CSFRD is shown in grey. *Middle:* Fraction of obscured star-formation as a function of redshift. *Upper:* Contribution to the CSFRD from galaxies of differing luminosities, being dominated by (U)LIRGs. *From Figure 7 of Zavala et al. 2021*.

There appears to be an inextricable link between these star-bursting systems and the formation of AGN. As put forward by Sanders et al. (1988), LIRGs are observed to host nuclear starbusts as well as AGN activity, and are thought to be the result of major mergers of gas-rich disc-like systems (see also Sanders 1998; Sanders et al. 2009). Both the nuclear starburst and AGN activity is fuelled by disc gas being funnelled into the central region (~ 1 kpc) following the merger activity (similar to Weaver et al. 2018, above). The infrared luminosity has been observed to originate from a compact region co-spatial with the nuclear starburst, likely being emitted from heated dust grains entrained in the gas clouds (Genzel et al., 1998). This physically intuitive picture may help explain the observed correlation between the mass of central supermassive black holes and the masses of the galaxy bulges that host them (Ferrarese et al., 2000; Gebhardt et al., 2000; Gültekin et al., 2009), suggesting a co-evolution between galaxies and their supermassive black holes sensitive to nuclear star-formation (Wild et al., 2010; Martin-Navarro et al., 2018; Pope et al., 2019b); although this is debated (e.g. Cisternas et al., 2011). Furthermore, it is reasonable that LIRGs (and other starburst systems) may transition into quiescent ellipticals as their star-formation may be cut off after such a catastrophic starburst episode and/or from AGN-driven outflows. LIRGs are preferentially found in overdense regions (Caputi et al., 2009; Casey, 2016), and with a similar number density as massive quiescent galaxies such that they are candidate progenitors of massive elliptical galaxies in the present-day universe (Casey et al., 2019). A variety of optically and x-ray selected AGN systems (Seyfert I & II, quasars) as well as radio-loud AGN[26] are found in similarly overdense environments, again suggesting a more general connection between environmentally-driven mergers, starburst episodes, luminous AGN, and (possibly) an eventual fate as massive quiescent galaxies (Kauffmann et al., 2000; Hopkins et al., 2008; Su et al., 2019). While studies have shown conclusively that only a small fraction of galaxies are expected to undergo major mergers, the AGN fraction among merger galaxies at $z = 0$ is elevated by a factor of 10 compared to the general population (Weigel et al., 2018), although such an analysis is complicated by the variability in AGN luminosity (e.g., Ulrich et al. 1997; Delvecchio et al. 2020, see also Weaver et al. 2022b). Although easily studied (being UV and/or IR luminous) starburst galaxies are likely following an extreme, but rare evolutionary pathway unlike that experience by the general population.

1.7 MASSIVE GALAXIES: CHALLENGES & OPEN QUESTIONS

Despite decades of observations and intense theoretical efforts, massive galaxies at all epochs remain an enigma. In the local universe, massive galaxies are overwhemingly quiescent and are found in the centers of galaxy clusters. Lying in the pits of deep gravitational potentials, they indelibly affect the evolution of satellite systems. Archaeological studies of the most massive present-day ellipticals have revealed that the bulk of their mass formed already in the early universe (e.g Aragon-Salamanca et al., 1998; Webb et al., 2015), suggesting that identifying their progenitors in the initial assembly phase would provide currently missing insight as to their remarkable assembly history and ultimately illuminating the origins of galaxy bimodality.

The first samples of massive quiescent galaxies in the early universe ($z \sim 2$ were identified from the Faint InfraRed Extragalactic Survey (FIRES, Franx et al., 2000) from the VLT which provided missing ground-based near-infrared

26 For a schematic of AGN varieties, see Table 1 of Urry et al. 1995.

imaging over Hubble Deep Field South (HDF-S, Franx et al., 2003; Daddi et al., 2004; van Dokkum et al., 2004; Toft et al., 2005), which at that time had only been surveyed in optical bands. Franx et al. (2003) used observed-frame near-infrared colors ($J_s - K_s > 2.3$), which selects not only truly quiescent populations but naturally also includes dusty systems (Förster Schreiber et al., 2004; Reddy et al., 2005; Papovich et al., 2006; Webb et al., 2006). While their remarkably red colors suggested a strong 4000Å break consistent with a Gyr-old stellar populations indicative of rapid star-formation cessation, their number density was found to be similar to that of star-forming Lyman break systems at $z \sim 3$ (Steidel et al., 1996a,b). Suddenly, it became plausible that quiescent galaxies were not only in place 3 Gyr after the big bang ($z \sim 2$, co-incident with the peak of star-formation), but that they may be descendants of star-forming systems identified at even earlier times. This was yet another watershed moment in galaxy evolution which spawned a still expanding literature on these remarkably young, mature galaxies in the early universe (from HDF-S alone: Cimatti et al., 2004; Daddi et al., 2005a; Labbé et al., 2005; Kriek et al., 2006a; Toft et al., 2007; Williams et al., 2009).

The discovery of such early and seemingly quiescent galaxies has since lead to a number of surprises. Although unresolved from ground-based imaging, follow-up with the early NIR NICMOS instrument aboard *Hubble* revealed their remarkable compact sizes and high stellar densities (Daddi et al., 2005a; Papovich et al., 2005; Toft et al., 2005; Trujillo et al., 2006; Toft et al., 2007; Trujillo et al., 2007; van Dokkum et al., 2008). The installation of WFC3 in 2009 enabled even more precise size measurements (Newman et al., 2012; Szomoru et al., 2012; van Dokkum et al., 2015) suggesting that compact quiescent systems comprise a significant fraction of all massive galaxies at $z \sim 2$. If they are to be a progenitor population of local ellipticals, they must undergo a four-fold increase in mass likely involving future mergers (Kriek et al., 2008; Toft et al., 2009). The increased sensitivity of near-infrared spectroscographs on the largest available facilities (e.g. Gemini-S/GNIRS, VLT/ISAAC, and Keck/NIRSPEC) allowed for < 1 night integration times reaching $V \approx 24 - 25$ AB depths sufficient to reach the stellar continuum in such UV-faint systems (e.g. Kriek et al., 2006a, see Figure 1.19). Although with limited sample sizes, such high-resolution spectroscopy confirmed resolutely the surprisingly mature stellar populations in these massive quiescent systems, as indicated by strong 4000Å break strength and Balmer absorption (e.g., $H\delta$). Futhermore, detailed kinematical studies revealed high velocity dispersons ($\sigma \approx 25 - 300$ km/s) similar to those found in present day elliptical systems (Stockmann et al., 2020). These results have prompted fruitful investigations into the location of massive quiescent galaxies on the Fundamental Plane (e.g. Bezanson et al., 2015) and extend it to higher redshifts, finding that passive evolution alone cannot provide sufficient size growth for such systems to evolve into present day (cluster) ellipticals (Stockmann et al., 2021). Attempts to identify similar galaxies at higher redshifts ($z \approx 2 - 4$) where star-formation cessation must have quickly followed their rapid formation epochs have been successful (Carnall et al. 2020; Stevans et al. 2021, see also Lustig et al. 2022). However, detailed kinematical measurements are feasible from only the largest facilities and *Hubble* (Glazebrook et al., 2017; Schreiber et al., 2018a; Tanaka et al., 2019; Forrest et al., 2020; Valentino et al., 2020a). Although limited to only a handful of gravitationally lensed systems, evidence appears to indicate that massive quiescent galaxies at these earlier times may be rotationally supported discs (Toft et al., 2017; Newman et al., 2018), perhaps suggesting that the growth in stellar velocity dispersion for the earliest massive quiescent galaxies occurs after their star-formation ceases.

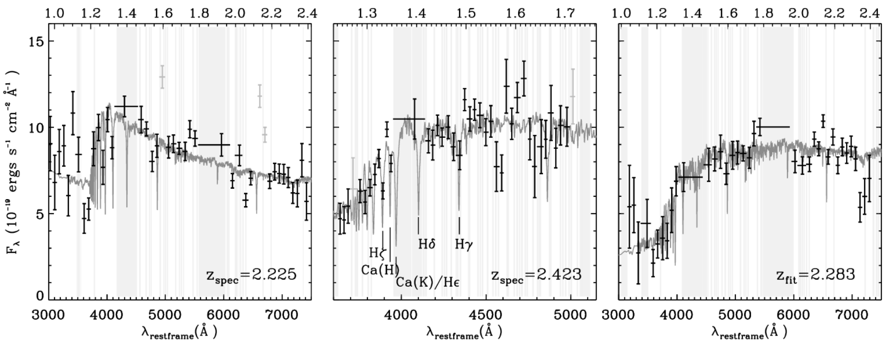

Figure 1.19: Some of the first spectroscopic measurements of the stellar continuum of $z \approx 2$ massive quiescent galaxies. Measurements of the 4000Å break region (black) are compared to the best-fit template spectrum (grey), with grey bars indicating data loss to atmospheric absorption. *From Figure 1 of Kriek et al. 2006b.*

Despite notable progress in the past decade since the first mass-complete samples (Marchesini et al., 2010), attempting to identify, let alone study massive and/or quiescent galaxies at $z \gtrsim 3$ remains a challenge. Firstly, robust mass estimates depend on well-sample rest-frame NIR measurements, which in the case of low-z galaxies requires NIR facilities. Already by $z \sim 2-3$ the observed-frame K-band light becomes a tracer of star-formation meaning that further infrared facilities are required. Ground-based IR observations are ineffective due to strong water absorption by the atmosphere. Plane-borne solutions such as SOFIA have helped in this direction, but remain secondary to space-borne platforms such as *Spitzer* and *Herschel*. However, cryogenic requirements of the detectors limits mission lifetimes (e.g., 6 years cryogenic for *Spitzer*) thereby restricting survey investment in terms of depth and/or area – critical for obtaining representative samples of massive galaxies at $z \gg 2$ (Caputi et al., 2011). Although the diffraction-limited, the meter-sized optical systems of *Spitzer* and Herschel coupled with IR wavelengths produces a characteristically low image resolution that makes sources difficult to characterize and measure. Consequently, mass estimates derived from these IR facilities suffer. By $z \sim 7-8$, sources become redshifted to the point where even the 4.5μm channel of IRAC is no longer a suitable mass probe.

Secondly, quiescent galaxies are photometrically degenerate with dusty star-forming galaxies. While low-z studies have relied upon a combination of morphology, spectroscopy, and ancillary FIR imaging to break this dust-age degeneracy, these tools are typically not available for objects at higher-z, and certainly not for large samples. Attempts to resolve this dust-age degeneracy for $z \gg 1$ have invoked color-color diagrams that attempt to orthogonalize star-formation and attenuation by dust. Using observed-frame colors $B - z$ and $z - K$ (*BzK*, Daddi et al., 2004) established the first reliable method to select quiescent systems at $z \approx 2$ (see also Daddi et al. 2005a,b). However, such an observed-frame selection has limited use to surveys at other redshifts, and imposes a complex definition of quiescence which evolves with redshift. Thanks to the advent of multi-wavelength surveys with detailed SED fitting, it has been possible to directly estimate rest-frame colors; UVJ (Williams et al., 2009), $NUVrJ$ (Ilbert et al., 2013), and $NUVrK$ (Arnouts et al., 2013a) are three of the most common. However, each implies a different definition of quiescence with respect to a cut in SFR, which in itself has been used to select quiescent galaxies (see Leja et al. 2019b for a review). Investment is beginning to be made to apply non-parametric clustering algorithms to better identify quiescent systems within the high-z domain (e.g. Steinhardt et al., 2020b), although further work is required.

A slew of additional obstacles bar a clean selection of massive (quiescent) galaxies. Morphological information of such distant and hence low angular size objects helps reveal the tell-tale signs of merger activity, but requires space-based imaging which is currently restricted to relatively small footprints containing only a fraction of the known massive high-z galaxies, and so dedicated follow-up from *Hubble* is currently industry standard. Furthermore, the lack of bright emission lines from quiescent galaxies means that they can only be spectroscopically confirmed from deep continuum spectroscopy which is incredibly expensive and has thus only been possible to a handful of sources taken from some of the most oversubscribed facilities (e.g., Keck, VLT, and *Hubble*). Cool stars from within the Milky Way have similar colors as both dusty star-forming galaxies and quiescent systems and must be identified from photometry as most $z > 2$ galaxies appear unresolved in ground-based imaging. Furthermore, stellar mass estimates are degenerate with the assumed IMF, which even in the local universe sees significant variation between star-forming and quiescent systems. Although investigated in earlier studies (e.g. van Dokkum, 2008), only recently have large, well-measured samples permitted implementing a variable IMF in SED modelling (Sneppen et al., 2022; Steinhardt et al., 2022a,b).

These pressing challenges define the horizons of our current observational landscape of $z > 2$ massive galaxy studies. Thankfully, advances in simulations, from both computational and physical recipe standpoints, continue to complement these existing observational results. By enabling linkages of large-scale properties across different epochs, as well as describing the assembly history of baryons in individual haloes, simulations of cosmological volumes excel at describing the dynamic and evolving universe. The discovery and subsequent study of rare massive galaxies has led to tension between observations and large-scale simulations (e.g. Nagamine et al., 2005; Feldmann et al., 2015). In the case of detailed hydrodynamical simulations (e.g., EAGLE, Furlong et al. 2015; ILLUSTRIS, Vogelsberger et al. 2014; ILLUSTRISTNG, Pillepich et al. 2018; HORIZON-AGN, Laigle et al. 2019; SIMBA, Davé et al. 2019; MAGNETICUM, Remus et al. 2014), their relatively small volumes are often too small to produce statistically meaningful numbers of massive $10^{10.5}\,\mathcal{M}_{\odot}$ galaxies (and if they did they would likely be *over*produced; see Figure 4.15). While semi-analytical simulations (e.g., Santa Cruz, Somerville et al. 2008; SHARK, Lagos et al. 2018) are able to overcome these volume limitations, their treatment of the small-scale physics is prescriptive and not necessarily physically complete. This is especially important for *quiescent* massive galaxies, wherein early simulations under-predicted observed quiescent fractions of massive systems. The recognition that action by AGN may play a role in driving down star-formation through gas heating and/or ejection prompted revision, the results of which shown better agreement with observed quiescent fractions. Since then, improvements to physical recipe as well as an increased number of resolution elements has produced increasingly better agreement with observations (which themselves are becoming more precise), especially in the predicted quiescent fraction at fixed mass. Yet uncertainties still exist; for example, is AGN energy injection pressure or density driven?. Not only is our understanding of AGN feedback incomplete, but the related signatures seen in simulations are under-constrained by current observational measurements. Further refinement of these simulations has been enormously fruitful; due in part to increased communication and collaboration between observers and theorists. Most notably ILLUSTRISTNG, and its three volumes TNG50, 100, and 300, has been shown to reproduce much of the assembly history and quiescent fraction evolution within the last 12 Myr ($z \lesssim 3$; Lustig et al. 2022, see also Chapter 4), although with an increasingly uncertain performance at higher redshifts due partly to larger observational uncertainties. Recently, Lovell et al. (2021) has been shown that a greater dynamical range of masses and environments can be achieved for less computational time by re-simulating many over- and under-dense subregions identified in cGpc-sized parent volumes, thereby quantifying predictions for extremely massive systems ($\mathcal{M} > 10^{10.5}\,\mathcal{M}_{\odot}$) which is currently not possible for single continuous volumes (see also Vijayan et al. 2021, 2022).

Despite heroic progress since the first plates taken by Hubble almost a century ago, galaxy evolution remains in its infancy. The means by which galaxies transform kinematically and morphologically, their relation to star-formation, mass, and environment has only recently come into focus. While new technologies such as integral field spectroscopy have enabled us to dissect nearby galaxies in unprecedented detail, the advent of infrared and submillimeter astronomy has enabled us to glimpse the distant universe and trace back its history to within $\sim 500\,\mathrm{Myr}$ of the big bang. Soon we will finally leverage the awesome vision of *JWST* to not only see the infrared universe in stunning morphological detail, but also accurately measure rest-frame optical properties of

$z > 6$ galaxies which before could only be studied in the UV. Furthermore, we will probe the stellar continuum of massive quiescent galaxies with sufficient detail to reconstruct their assembly histories and reveal the fantastic speed at which they formed, evolved, and slipped into quiescence. Doing so, however, requires that we understand the evolution of the early universe pervaded by neutral hydrogen gas – within which these first massive galaxies are thought to have formed.

1.8 THE EPOCH OF REIONIZATION

Modern cosmology[27] predicts that after the big bang, the rapid expansion of the universe allowed it to cool. After a span of ∼380 000 years, the universe cooled to the point where bayronic matter could de-couple from the electromagnetic field and form electronically bound states whose emission is visible as the CMB. Although with trace abundances of primordial Helium and Lithium, ∼ 95% of the baryonic material was in the form of neutral Hydrogen (Boesgaard et al., 1985). However, non-baryonic dark matter, it is thought, did not interact with the electromagnetic field and so was able to condense immediately after the big bang. The quantum-scale inhomogeneities frozen in by inflation caused dark matter to be anisotropically distributed such that it pooled under gravity to form the primordial cosmic web of immense filamentary structures which intersect to form the most overdense environments in the early universe (Blumenthal et al., 1984). Once baryonic matter was no longer coupled with light, it collected under the gravity of dark matter thereby tracing the underlying dark matter density field (White et al., 1978; Davis et al., 1985; Springel et al., 2005; Massey et al., 2007). The initial angular momentum of the in-falling gas led to the first accretion processes by which the first galaxies formed (White et al., 1991; Kauffmann et al., 1993; Bromm et al., 2011). Dominated by dark matter, these first self-gravitating baryonic structures are expected to have been similar in mass to today's globular clusters ($\mathcal{M} \sim 10^6\ \mathcal{M}_\odot$, Press et al. 1974; Binney 1977; Rees et al. 1977). Within them, clouds of neutral hydrogen cooled and formed the first molecular clouds. They were unlike those seen in the Milky Way, enriched from eons of heavy metals ejected from dead stars. Yet despite assistance from metal line cooling, these pristine gas clouds eventually fragmented to birth the first generation of stars (Population III), devoid of enriched heavy elements (Bromm, 2013; Karlsson et al., 2013).

27 For a summary, see Peacock (1999).

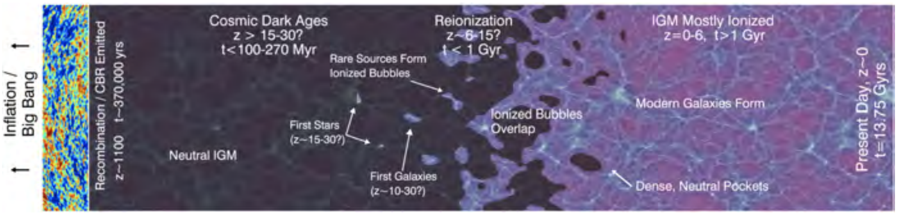

Figure 1.20: Schematic timeline of reionization from the formation of structure following recombination, the formation of the first galaxies, and their impact on the reionization of the neutral intergalactic medium (IGM). *From Figure 1 of Robertson et al. 2010.*

The connection between massive galaxies and overdense environments is well known in the local universe, and so it is reasonable that this connection exists back at their formation epoch (e.g. Hogg et al., 2003). Theory predicts that the first galaxies formed in the most overdense environments (De Lucia et al., 2007), the nodes where dark matter first collected and amassed the most extensive collections of hydrogen gas. The galaxies that formed within these massive dark matter haloes had enormous gas resevoirs and so could grow rapidly to become some of the most massive galaxies of their epoch (Collins et al., 2009). Their early formation was likely marked by violent *in situ* star-formation on the order of $\sim 100\, M_{\odot}\, \mathrm{yr}^{-1}$, making them *ultra*-luminous (Finkelstein et al., 2013; Mutch et al., 2016). The sheer ionizing radiation emitted by these young stars had an indelible impact on their environment, ripping the electrons from their neutral hydrogen cocoons (Furlanetto et al., 2004; Borthakur et al., 2014). The dark ages was over, and the epoch of reionization had begun. These first reionized bubbles, originating from the most overdense regions, continued expanding over the next 600 Myr, ending by $z \sim 6$ (Becker et al., 2001) leaving the IGM the transparent plasma we see today (Figure 1.20).

Observational insight into this little known but formative era in galaxy formation remains limited. Surveys of the distant universe must contend with interloper populations of low-z dust obscured systems and late-type stars which mimic Lyman breaks of bone fide $z > 6$ systems (Euclid Collaboration et al. 2022, Allen et al., in prep.). Furthermore, the rarity of these overdensities, varying over degree scales, implies that these first galaxies are also rare, but also highly clustered, making reionization correspondingly patchy (Trac et al., 2008a; Pentericci et al., 2014; Sobacchi et al., 2014). Blind photometric and grism spectroscopy surveys have been productive in identifying significant samples which are beginning to span cosmologically meaningful volumes (Zheng et al., 2017; Maseda et al., 2018). Spectroscopic follow-up at high-resolution have efficiently confirmed Lyα line emission and the ionizing continuum flux, the incidence of which measure the neutral fraction and hence trace the development of reionization (Mason et al., 2018b; Bolan et al., 2021); see Figure 1.21. From these and other measurements (e.g., absorption lines, Becker et al. 2015), studies are beginning to bridge the gap between theoretical predictions and observed properties and piece together a timeline of reionization from the first galaxies $z \sim 15(?)$ to $z \sim 5-6$ (Fan et al., 2006b; Kashikawa et al., 2006; Ouchi et al., 2010; Robertson et al., 2015; Bañados et al., 2018).

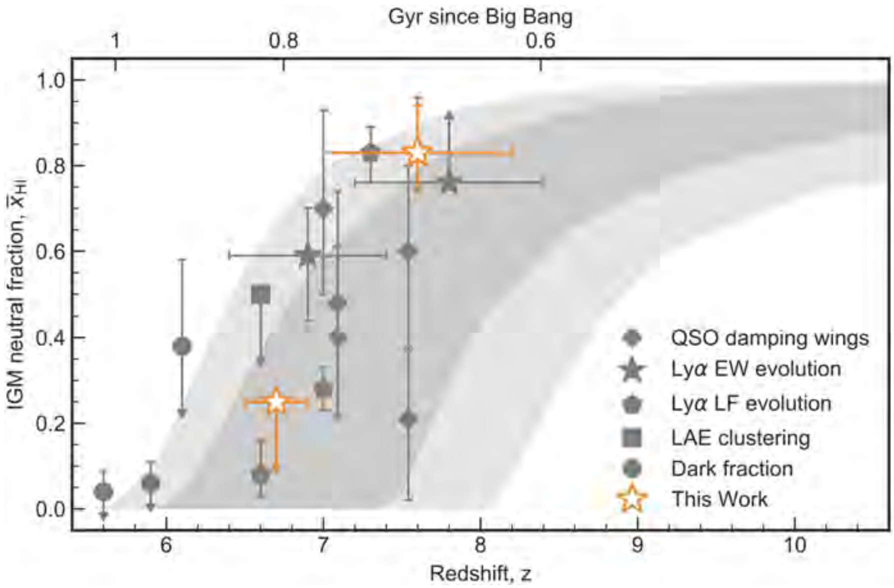

Figure 1.21: Observational constraints on how quickly the intergalactic medium (IGM) became neutral during reionization, and the 69 and 95% ranges computed from the cosmic microwave background constraints (Mason et al., 2019). *From Figure 7 of Bolan et al. 2021.*

One of the principle questions is the role of these first massive galaxies in reionizing their environments. If reionization was driven by a few ultra-luminous galaxies, outshining less luminous neighbors born within the same overdense region (Stark et al., 2017; Mason et al., 2018a; Naidu et al., 2020; Leonova et al., 2021; Larson et al., 2022), then reionization should be proceeded quickly and late. On the other hand, collections of many less luminous neighbors could be more powerful in the long run (Atek et al., 2015; Castellano et al., 2016; Anderson et al., 2017; Livermore et al., 2017) such that reionization started rapidly from many smaller regions. Moreover, these massive galaxies may undergo rapid evolution under the action of a fast-forming supermassive black hole, which imparts particularly hard ionizing radiation, potentially outdoing the relatively less potent photoionization from O- and B-type stars (Fan et al., 2006a; Alvarez et al., 2007; Bolton et al., 2007; Davies et al., 2016). The ionization budget, constrained by the necessary end by $z \sim 6$, can be meaningfully compared to the integrated UV emission from the rest-frame UV luminosity function (UVLF; Bouwens et al. 2007, 2011, 2015; Finkelstein et al. 2015a; Mason et al. 2015). However, construction of the UVLF is marred by systematics and uncertainties in sample selection, as well as the wide dynamic range in volumes and luminosities required to be measured (McLeod et al., 2016; Adams et al., 2020).

While constraints on reionization from $z \approx 6-7$ are now routinely assessed, there remains significant debate as to the onset of reionizaton emanating from the rarest overdensities at even earlier times. While the mid- to low-luminosity end of the $z \approx 8-10$ UVLF has been well measured in small, deep fields such as CANDELS by *Hubble* (Oesch et al., 2010a; Roberts-Borsani et al., 2016; Bouwens et al., 2021a), constraining the number density of the considerably rarer sources that constitute the luminous end has been far more challenging (Bouwens et al., 2006; Oesch et al., 2012, 2016, 2018). The necessarily wide infrared imaging that is also deep enough ($\lesssim 26$ AB) to identify these ultra-luminous *cosmic beasts* exists in only a handful of fields (e.g., COSMOS, SXDF). Recent improvements in imaging depth (e.g. UltraVISTA DR4, Moneti et al. 2019) has led to a number of studies reporting surprisingly large numbers of such ultra-luminous ($M_{UV} \lesssim -22$) galaxies in the EoR (Bowler et al. 2015; Stefanon et al. 2019a; Bowler et al. 2020a; Harikane et al. 2022; Leethochawalit et al. 2022, Kauffmann et al. in prep., see Appendix 6). While tension with theory is difficult to assess at the current state of uncertainties on both sides (Ren et al., 2019), observation supports a scenario where the form of the UVLF as traced by these ultra-luminous sources are too common relative to a Schechter function, as has been traditionally extrapolated from the low-luminosity end. This apparent excess thereby makes these ultra-luminous systems (and the few deep surveys wide enough to find them) of particular interest for testing reionization models. Interpretation is made difficult in part due to the comparably low spatial resolution and crowding of ground-based near-infrared images as well as those from *Spitzer*/IRAC that can blend two galaxies together which if endemic to these samples will naturally shift low luminosity sources to higher luminosity bins (Bowler et al., 2017a; Stefanon et al., 2021b). Although dust attenuation at such early times was thought unlikely, observations have proven otherwise (Bouwens et al., 2009; Watson et al., 2015; Laporte et al., 2017) suggesting that definitive interpretations of the UVLF may be skewed by dust obscuration (Bouwens et al., 2012a; Finkelstein et al., 2015b; Schouws et al., 2022). Additional contamination of broad-band colors by emission lines may contribute to red rest-frame optical colors (Labbé et al., 2013). Furthermore, while their redshift and Lyα properties can be spectroscopically measured by VLT and Keck (Bouwens et al., 2006), whether or not the complexity of sources seen in these ground based images is due to noise, blending, or genuine morphological disturbances can only be confirmed by *Hubble* (and soon *JWST*); for example, merging or multi-component morphologies of several $z \approx 7$ galaxies have been reported by Bowler et al. (2017b). Furthermore, the fact that the high- and low-luminosity samples are gathered from different volumes and selection functions makes the task of navigating the systematics (which in the case of volumes are zeroth order) more complex. Nonetheless, the current observational landscape appears to point to a genuine excess of ultra-luminous, and presumably massive galaxies already within the first billion years.

While their role in reionization is all but certain, the possible existence of the first class of massive galaxies is tantalizing. Reaching back to these epochs with detailed rest-frame UV and optical observations from *JWST* (Naidu et al., 2021; Stefanon et al., 2021a; Weaver et al., 2021, e.g.) will finally clarify the formation and evolutionary histories of the most massive galaxies found at $z \sim 3-5$, lying mysterious in the dark reaches of the cosmic past.

1.9 SUMMARY & OUTLINE

This thesis aims to further just a fraction of this grand story concerning the evolution of galaxies.

To explore this distant and little understood era of cosmic history, we have constructed the COSMOS2020 Galaxy Catalog containing over a million galaxies over 2 deg^2 of the COSMOS field. Source detection and multi-wavelength photometry is performed for 1.7 million sources across the 2 deg^2 of the COSMOS field, ~966 000 of which are measured from all available broad-band data. We adopt a near-infrared (NIR) selection function by detecting on an co-added $izYJHK_s$ image, which includes the deepest NIR images yet obtained over 2 deg^2. Although we have become sensitive to incredibly faint sources in these deep images, their emergence from the noise noticeably increases the source density which makes accurate source identification and subsequent photometry more challenging. Future deep ground-based surveys (e.g., Rubin-LSST, HSC-SSP) will only heighten this challenge. To overcome these obstacles we built The Farmer, which measures photometry by fitting model profiles, as opposed to summing fluxes in apertures that can become corrupted in crowded fields. Chapter 2 presents an overview of The Farmer, including key demonstrations that highlight its photometric capabilities.

In Chapter 3, we present photometry of COSMOS2020 sources using The Farmer. We also extract photometry using the industry standard aperture-approach as a complimentary benchmark. Photometric redshifts are computed for all sources in each catalog utilizing two independent photometric redshift codes, finding similar performance in predicting known spectroscopic redshifts. In all four cases the $i < 21$ sources have sub-percent photometric redshift accuracy and even the faintest sources at $25 < i < 27$ reach a precision of 5 %. Compared to COSMOS2015, reaches the same photometric redshift precision at almost one magnitude deeper. Spectacularly, we find that The Farmer provides generally more accurate photometry of faint, distant galaxies and so we can confidently capitalize on recent improvements in imaging depths to obtain more complete samples out to extremely high redshifts ($z \sim 10$ or 450 Myr after the big bang), and more accurate measurements of distances, stellar masses, and other physical properties than ever before.

In Chapter 4, we provide the latest constraints on the form and evolution of the SMF to $z \leq 7.5$; valuable not only to help piece together the assembly history of galaxies, but also to calibrate and test predictions from theoretical models and simulations. We uncover a fantastically consistent operation of the assembly of galaxy stellar mass, and continue to find surprising numbers of massive galaxies already in the early universe ($z > 3$); comparisons with state-of-the-art hydrodynamical and semi-analytical galaxy evolution models show agreement, and point towards areas of improvement. While these are the tightest constraints yet on the evolution of the SMF across ~ 10 billion years, obtaining direct measurements of galaxy masses at $z > 7$ requires *JWST*.

However, directly measuring star-formation (i.e. mass growth) at $z > 7$ is feasible now. UV emission from O- and B-type stars is critical in understanding how nascent galaxies assembled their mass through star-formation, and to what extent the associated radiation ionized the neutral hydrogen of their environments. We have leveraged the wide area of COSMOS2020 to identify dozens of these rare, ultra-luminous $z > 7$ *cosmic beasts*. They must have been some of the first galaxies to have lit up the early universe, assembling their mass in a luminous burst of star-formation < 600 Myr after the big bang. Yet, their tremendously rapid growth challenges the standard paradigm of galaxy evolution. Are we are witnessing a formation stage of seemingly unrestricted growth, before the subversive effects of feedback have taken hold? Chapter 5 introduces the 14.4 hour *JWST* program BEASTS, which will use the awesome vision of *JWST* and the Near-infrared Spectrograph's integral field unit to finally reveal their origins.

Bibliography for material presented in Sections 1.1-1.3

Silk, J. *The Big Bang*. en. W.H. Freeman, 2001.

Sagan, C. *Cosmos*. Random House, Inc., 2002 ed. New York: Random House, 2002.

Bartusiak, M. *Archives of the Universe: 100 Discoveries That Transformed Our Understanding of the Cosmos*. en. Knopf Doubleday Publishing Group, 2006.

Freedman, R. and W. J. Kaufmann. *Universe*. English. 8th edition. New York, NY: W. H. Freeman, May 2007.

Bartusiak, M. *The Day We Found the Universe*. en. Vintage Books, 2010.

Singh, S. *Big Bang*. en. HarperCollins UK, Nov. 2010.

Kelley, D. H. and E. F. Milone. *Exploring Ancient Skies: A Survey of Ancient and Cultural Astronomy*. en. Springer Science & Business Media, Feb. 2011.

2

LEARNING HOW TO FARM

In former times when one invented a new function it was for a practical purpose; today one invents them purposely to show up defects in the reasoning of our fathers and one will deduce from them only that.

- Henri Poincaré, *1899*

Adapted from:

The Farmer: A reproducible profile-fitting photometry package for the next-generation of ultra-deep galaxy surveys

Prepared for The Astrophysical Journal, 2022

J. R. Weaver, L. Zalesky, V. Kokorev, C. J. R. McPartland, N. P. Allen, N. Chartab, K. M. L. Gould, M. Shuntov, I. Davidzon, A. L. Faisst, N. Stickley, P. Capak, S. Toft, B. Mobasher, D. B. Sanders, O. B. Kauffmann, O. Ilbert, A. Moneti, G. Brammer, and H. J. McCracken

2.1 INTRODUCTION

For most of its history, astronomy has been defined by the use of electromagnetic waves to measure sources detected in the night sky. What began as a purely visual study was transformed in the late 19$^{\text{th}}$ century with the advent of photographic plates that enabled precise observations from which the brightness of sources could be measured (Bigourdan, 1888). It was with such comparatively primitive technology that the first variable stars in Andromeda were identified, leading to the discovery of the 'island universes' and later the expansion of the universe (Hubble, 1926; Hubble, 1929). Now almost a century later, all scientific astronomical observations are captured on Charge-Coupled Devices, or CCDs (Lesser, 2015), further enhancing the accuracy and precision of photometry.

Photometry itself has for decades been performed using apertures. That is, the integrated flux or total brightness of a source is computed within apertures of a fixed size. This is especially useful for isolated, unresolved, point-like sources like stars, quasars, and distant galaxies whose spatial appearance is well-described by the point-spread function (PSF) determined by the optical train of the telescope. While larger apertures ensure all of the light is captured and are less susceptible to noise, they may unintentionally capture light from other nearby sources which is usually mitigated by smaller apertures, although with typically greater uncertainties. Images with high source density, arising either from physically compact structures (e.g., star clusters) or from background and foreground sources appearing in close proximity on the sky, may require apertures smaller than the PSF (or alternative mitigation strategies, see Stetson 1987; Bertin et al. 1996). Recovering the total flux in such cases requires scaling the aperture-integrated flux proportional to the total extent of the PSF, which often involves complicated strategies to characterize the PSF stability across the detector or co-added mosaic. Transitioning from monochrome photometry of a single band to photometering multi-wavelength images presents its own challenge as PSFs typically vary with the filter as well as telescope, instrument, and observing conditions. The solution has been a procedure known as PSF homogenization whereby each image is convolved with a kernel that maps the PSF of that particular image to that of a target PSF. The choice of the target PSF is not always well-defined, especially in cases where the PSF characteristics vary significantly between bands.

For applications in extragalactic studies, the deepest wide-field ground-based near-infrared survey at the time of writing is UltraVISTA (McCracken et al., 2012) which at a uniform $K_s \approx 26$ AB depth captures ~ 150 sources per arcmin^2 over 2 deg^2 with resolution set by its 0.51″ PSF at FWHM. Consequently, modest apertures of even 3″diameter are often contaminated by neighboring sources. In the corresponding source catalog of Weaver et al. (2022a), 2″diameter apertures are adopted when measuring photometry to be used in spectral fitting, which in the case of some high-redshift ($z > 7$) galaxies remain contaminated such that interloping blue light does not permit a high-redshift solution (O. Kauffmann et al., 2022, submitted). While manually removing such interlopers in small samples is possible, doing so for several thousand becomes impractical, and risks imposing human biases. Until the operation of space observatories such as *Euclid* and *Roman*, surveys with the large area and near-infared bands necessary to detect rare, high-redshift galaxies will continue to be conducted by ground-based facilities at significantly lower spatial resolution, and so these challenges to aperture-based methods will only become more difficult. As we will demonstrate, apertures have self-imposed limitations with respect to survey depth.

These challenges must be met with appropriate solutions now if we are to continue exploring the high-redshift universe, particularly from ground-based facilities. Doing so successfully will require continued investment in large area near-infrared surveys complemented by UV, optical, and infrared imaging. The approaches which will ultimately solve these problems must not be as susceptible to contaminating fluxes from neighboring sources, and they must be consistently applicable over a wide range of spatial resolutions.

Precise, modern photometry has led to the development of photometric redshifts whose success lie in the ability of broadband photometry to constrain the stellar continuum, and possibly line emission. As introduced in the seminal paper of Fisher (1922), the Fisher Information metric provides that the information contained in any signal is proportional to its derivative. This can be readily appreciated in astronomy: a featureless continuum is much less informative than a high-resolution spectrum with well-measured emission and absorption lines. Hence, determining the redshifts of sources which lack strong emission lines (which may be detected precisely in narrow-band imaging) relies on accurate colors. If the aperture corrections applied to photometry in one band are miscalibrated then the related color terms will be biased, potentially diminishing or constructing the appearance of a Lyman or Balmer break and leading to an incorrect redshift estimate and corresponding physical parameters. The situation only worsens with redshift in inverse proportion to apparent brightness and signal-to-noise.

An attractive class of alternative photometric techniques called "profile-fitting" photometry has enjoyed great success overcoming these very challenges. They work by fitting a model (parametric or non-parametric) which describes the surface brightness profile of a source. Usually the total brightness is a parameter of that model, or can somehow be derived from it. Commonly-used parametric implementations of profile-fitting involves a source model parameterized by flux, position, and for resolved sources also size, axis ratio, position angle, and profile gradient (Sérsic index, Sérsic 1963b) which is then convolved with a known PSF and optimized to describe the surface brightness profile a given source. This approach has significant advantages over traditional apertures. Firstly, the flux reported is the total brightness of the source in that particular band, avoiding aperture corrections and related systematics. Secondly, the PSF is a property of the model which is a more tractable solution compared to PSF homogenisation which manipulates the measurement image.

This also means that the fitted properties of resolved sources only assume the PSF, but are not convolved; e.g., the estimated effective radius is that of the intrinsic source, not the apparent one. Thirdly, positions are fitted for and not simply determined as the peak or centroid of an image, subsequently achieving greater precision over commonly-used peak-finding routines in photometry software (e.g., Source Extractor, Bertin et al. 1996). Lastly, sources that have some fraction of their flux overlapping can be accurately photometered by fitting an appropriate number of simultaneous models. This forward-modeling ability to de-blend sources is unique to profile-fitting photometry and means that sources easily differentiated in high-resolution images can be accurately photometered in low-resolution bands such as *Spitzer*/IRAC.

The Tractor (Lang et al., 2016a) is one such profile-fitting tool. Given a set of initial positions, model profiles (e.g. point-like versus resolved sources), and image information with per-pixel uncertainties, The Tractor optimizes those models for a given set of images whose sources have been already identified from some existing detection image. The models are parametric, and provide estimates for source flux and position, in addition to effective radius, axis ratio, and position angle appropriate for resolved sources. The key distinction when utilizing such parametric models is that we can derive a likelihood for the particular model parametrization given the data, as well as estimate uncertainties on those parameters. Key implementations of The Tractor include Lang et al. (2016b), Faisst et al. (2021), and Stevans et al. (2021). In addition, Nyland et al. (2017) explored for the first time the capabilities of The Tractor to photometer highly blended IRAC sources using models derived from higher resolution VISTA imaging.

We develop a pipeline to perform reproducible profile-fitting photometry based on The Tractor called The Farmer, which adopts the similar principles outlined in previous work using model-based photometry including `HSCPipe` (Aihara et al., 2019), the `DECaLS` pipeline (Dey et al., 2019b), and `GaLight` (Ding et al., 2021). The Farmer provides a larger framework within which The Tractor can be scaled to large galaxy surveys where source detection must be handled in an automated, statistical manner. Crucially, The Farmer includes built-in parallelization methods which enable efficient computational runtimes. The Farmer utilizes the optimization routines already provided by The Tractor to obtain estimates of source flux and positions, as well as galaxy shapes for resolved sources. At no point are fluxes derived through integration over an aperture. Instead, the fluxes are derived directly from the normalization factor required to scale a unit-normalized model to best describe a given source. Parameter uncertainties, including flux, are derived as minimum-variance estimates according to the Cramér-Rao bound. For point-like sources, this equates to the classical variance derived when fitting a pattern using inverse-variance weights.

The work presented here is independent of any assumed cosmology. All magnitudes are expressed in the AB system (Oke, 1974), for which a flux f_ν in μJy (10^{-23} erg cm^{-1}s^{-1}Hz^{-1}) corresponds to AB$_\nu$ = 23.9 − 2.5 log$_{10}$(f_ν/μJy).

2.2 REVIEW OF THE TRACTOR

The Tractor is a recent development led by Dustin Lang, David Hogg, and others to provide a generalized framework for fitting the surface brightness profiles of sources in an image. The approach is generative, that is, The Tractor attempts to construct a predictive model based on the image calibration parameters such as the science image, a corresponding PSF, and a per-pixel noise estimate (typically a weight map), and optionally a background sky model; as well as initial guesses as to the model parameters such as source positions, shapes, and fluxes. In practice, The Tractor optimizes these initial parameters to produce a model image which describes input image within the bounds of the properties provided, separating the source signal from the background noise.

The flux of a given source α is not measured with apertures, but is rather obtained directly as the normalization of a unit-normalized model profile $G(\phi)_i$, where ϕ is the subset of parameters describing the position and shape of the overall model $m(\theta)_i$ defined over every pixel i and convolved with the PSF:

$$m(\theta)_i = \alpha\, G(\phi)_i \circledast \mathrm{PSF_i} \tag{2.1}$$

The flux for a single isolated point source is essentially computed as a mean of the input image x_i and the model image $m(\theta)_i$ normalized to unity and inversely weighted by pixel variance σ_i^2. In other words, flux is the value required to scale a unit-normalized model image of a point source to describe the real point source. The Tractor attempts to maximize the likelihood $\mathcal{L}(x|\theta)$ of the data x given the free parameters θ, and uses the quadrature addition of the weighted residual image (i.e. χ), which is analogous to a χ^2 minimization as $\mathcal{L} \sim \exp(-\frac{1}{2}\chi^2)$ but in two spatial dimensions, ignoring pixel-pixel covariances[1]:

$$\mathcal{L}(x|\theta) = \exp\left(-\sum_i \frac{(m(\theta)_i - x_i)^2}{2\sigma_i^2}\right) \tag{2.2}$$

One immediate advantage of this approach is that it avoids the need for PSF homogenization as the PSF is included in convolution with the source profile. Another advantage is that as long as the model is normalized to unity including the wings, it may be truncated in numerical processing without biasing the estimated flux. Therefore while an aperture over the model realized in some restricted image dimensions will return a flux less than the true flux, the flux determined by scaling the unit normalized (but truncated) model will remain accurate. This is especially useful when considering numerical and computational limitations.

1 We assume Gaussian statistics which in many cases may not be appropriate.

Of perhaps equal importance are parameter uncertainties[2]. According to the documentation of THE TRACTOR, the uncertainty estimates produced are related to the Cramér–Rao bound, which is a lower bound on the variance of any unbiased estimator $\hat{\theta}$:

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{I(\theta)} \tag{2.3}$$

where $I(\theta)$ is the Fisher information,

$$I(\theta) = E_x\left[\left(\frac{\partial \log \mathcal{L}(x|\theta)}{\partial \theta}\right)^2\right] \tag{2.4}$$

$$= -E_x\left[\frac{\partial^2 \log \mathcal{L}(x|\theta)}{\partial \theta^2}\right] \tag{2.5}$$

The log likelihood is therefore

$$\log \mathcal{L}(x|\theta) = -\sum_i \frac{(m(\theta)_i - x_i)^2}{2\sigma_i^2} \tag{2.6}$$

with first derivative

$$\frac{\partial}{\partial \theta} \log \mathcal{L}(x|\theta) = -\sum_i \frac{m(\theta)_i - x_i}{\sigma_i} \frac{1}{\sigma_i} \frac{\partial}{\partial \theta} m(\theta)_i \quad , \tag{2.7}$$

which should equal zero when the likelihood has been maximized.

The second derivative is

$$\frac{\partial^2}{\partial \theta^2} \log \mathcal{L}(x|\theta) =$$
$$-\sum_i \left[\frac{m(\theta)_i - x_i}{\sigma_i} \frac{1}{\sigma_i} \frac{\partial^2}{\partial \theta^2} m(\theta)_i + \left(\frac{1}{\sigma_i} \frac{\partial}{\partial \theta} m(\theta)_i\right)^2\right] \quad , \tag{2.8}$$

where the first term is zero at the optimum. Returning to the Cramér–Rao bound, we have

$$\frac{1}{\mathrm{var}(\hat{\theta})} \leq -E_x\left[\frac{\partial^2 \log \mathcal{L}(x|\theta)}{\partial \theta^2}\right] \tag{2.9}$$

and since our second derivative (equation 2.9) is independent of x, the expectation collapses and we get

$$\frac{1}{\mathrm{var}(\hat{\theta})} \leq \sum_i \left(\frac{1}{\sigma_i} \frac{\partial}{\partial \theta} m(\theta)_i\right)^2 \quad , \tag{2.10}$$

which is the inverse-variance estimate reported by THE TRACTOR. In the important case of estimating flux where $\theta \equiv \alpha$, the derivative of the model with respect to flux is just the profile of the model. Hence, the uncertainty estimate on flux for point-like sources is based entirely upon the PSF and the per-pixel error estimates σ_i from the weight map.

2 Note: The derivation here is adapted from documentation provided for THE TRACTOR.

We can gain a better understanding of The Tractor, both its functionality and limitations, through progressively complex examples.

The simplest example is an isolated, point-like galaxy. The Tractor is supplied with the image, a weight map, a PSF, and a known position for the source; The Tractor does not provide means to detect sources, and so a list of initial source positions is required beforehand. While the image calibration parameters (image, weight map, PSF) must be kept fixed, we may also fix the position parameter so that only the flux is allowed to vary. This one parameter optimization is linear in the case of a source. However, profile-fitting photometry is sensitive to offsets in source positions requiring greater precision than is typically needed for accurate aperture photometry. One can address this by simply allowing the model position to also vary, and The Tractor has built-in functionality to deal with this. This three parameter optimization is a non-linear procedure, although the degeneracy between the position and flux parameters should be virtually zero. The result is not only an estimate of the flux, but also the source position. The source may also be photometered in many bands in a single joint optimization where the shape and position are shared but flux is now a vector with an element for each band.

A more complicated example is an isolated, resolved source. The Tractor includes a library of discrete parametric models which include, in order of simplicity, point source profiles taken from the PSF stamp (as assumed in the previous example), resolved profiles exponential and de Vaucouleurs profiles, and composite profiles by superimposing exponential and de Vacouleurs profiles. As before, the image calibration parameters are kept fixed. We also may fix the position, for simplicity, leaving the source shape and flux to vary. The question then is how to decide which shape parameterization to use? The Tractor does not provide an answer; rather it is up to the user to choose a model type ahead of the optimization. A resolved model type is appropriate in this case, and so now our optimization returns source fluxes and shapes (e.g., effective radius, axis ratio, and position angle). Photometry of other images taken with different filters is usually of interest and so by fixing the model shape we can perform 'forced photometry'. Although it is possible to allow the shape to vary with each band, this comes at the cost of potentially overfitting our model.

An even more complicated example is an image containing many sources of various morphological presentations and crowding. This is typically what is encountered in deep galaxy surveys, and presents a serious challenge. We have already understood that The Tractor does not provide source detection, and so the degree to which the photometry succeeds is dependent on the performance of some external detection procedure. Once we have somehow supplied source centroids to The Tractor, we are still left to determine the appropriate model type for each source. Although it may be feasible to assign model types manually for small regions of interest occupied by a small number of sources, this is typically not practical for large surveys containing thousands of sources. Assuming this can be done in some way, The Tractor will optimize all source models simultaneously on that given image to produce optimized shapes which can then be fixed to performed forced photometry on other bands of interest. Alternatively, one can use all the bands of interest to optimize the model and simultaneously obtain measurements of fluxes, although this adds significant complexity that in some conditions the optimization may fail to converge.

The situation does not improve much even if there is only one source of interest amongst a crowded field. Although one may try to instantiate a single model at that source position, THE TRACTOR uses information from every pixel in the image that has non-zero weight. That means that the presence of every other source in the image counts against the likelihood. One option is to restrict the weight map to only the pixels belonging to that source. However, deciding the extent of such a region is non-trivial. Regions that are too large may include flux from a neighbor which are unaccounted for by our one source model, and may bias the photometry typically towards higher fluxes. Having too small an region is suboptimal, and ill-defined as you would need to know the extent any neighbors beforehand. Another option is to continue instantiating models (defined by centroids and model types) for all nearby sources until it is possible to cleanly define a contiguous region whose boundaries do not contain light from other sources (i.e., an isolated group of sources). Such a manual approach may work, but only in limited cases where the user is heavily involved, severely limiting reproducibility. Even if this can be done, it remains unclear how best to optimize this potentially large group of nearby sources. Should they be optimized simultaneously? This approach is straight-forward but computationally expensive. Perhaps they should be optimized one by one, subtracting the best-fit model each time? This is usually computationally faster, but induces hysteresis which can bias photometry.

A generalized version of this dilemma is useful in proving this point. In Figure 2.1, eight point sources are injected into a Gaussian noise field at signal strengths ranging from $\sim 3 - 10\sigma$ and arranged in a circle. A total of six cases are constructed (A, B, C, D, E, F) by varying the radial distance to each source such that at one extreme they are separated and overlapping at the other.

As a baseline, fluxes are summed in $2''$apertures that do not overlap in case A and so recover accurate fluxes. However, a bias grows towards case F where the apertures become confused and eventually include the flux from all eight sources in each aperture. This highlights the limitations of apertures in crowded fields, after which one must appeal to statistical mitigation strategies afterwards af to re-scale fluxes (as is done in SOURCE EXTRACTOR). We move on to profile-fitting photometry in the subsequent rows. The most direct approach is to model each galaxy individually in series, but by case C succumbs to the same confusion as the apertures and multiply counts each source per model. An attractive solution is to also subtract the model each time. While this is certainly more successful in that the measured flux is typically accurate, it comes at the cost of an increased failure rate that worsens with crowding. This arises because the first model tends to overestimate the flux that continues until eventually a negative residual is left for measurement by the last few models.

The optimal way is to model each source simultaneously, which allows the *joint* model to recognize that there are neighbors that it can describe. This approach does not suffer from the drawbacks of fixed apertures, or of fitting models individually or with subtraction. It recovers accurate photometry in cases A, B, C, and D, leaving only a minor bias in cases E and F. Yet, these most extreme cases are pathological ones, and it is unlikely that a source detection procedure would be able to separate the signal into even two centroids, let alone all eight. Therefore the most extreme cases remain a problem, but one which will have to be addressed by innovations in source detection and associated de-blending techniques. Although fitting multiple nearby sources simultaneously is clearly the optimal approach, it is also the most computationally expensive one, and for that reason it cannot be so readily scaled up to large area surveys without first developing efficient algorithms which can be utilized successfully by high performance computing facilities.

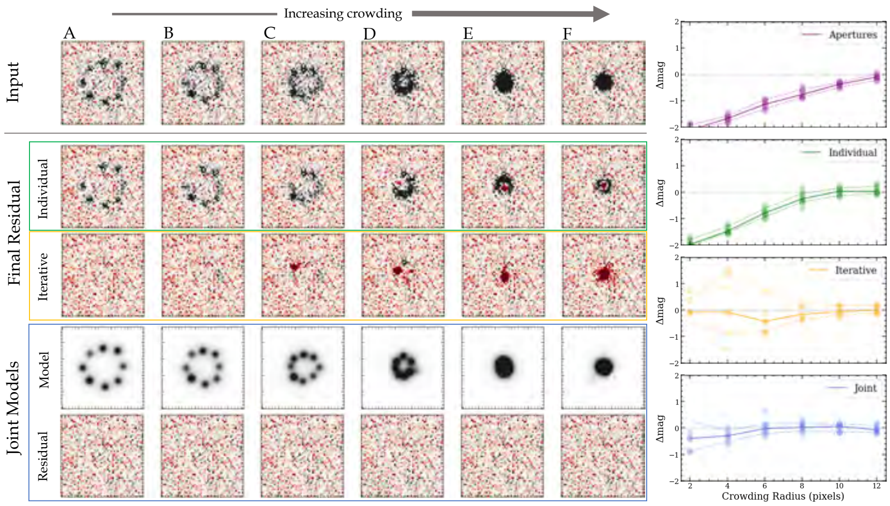

Figure 2.1: Eight similarly bright pointsources are injected into a simulated noise field over six scenarios of increasing degrees of circularly symmetric crowding. Assuming source positions are known beforehand, fluxes are measured in four ways: 2′′apertures, profile-fitting each source independently, iterative profile fitting each source with successive model subtraction, and jointly fitting all models simultaneously. The degree of success of each method is shown on the right measured in the difference in magnitude Δmag between the input and measured magnitudes as a function of source crowding, with a median Δmag and 68% ranges indicated for each scenario. Only joint fitting provides both precise and accurate recovery of crowded sources.

As we can see, The Tractor is a powerful tool for determining best-fit values corresponding to parametric models of sources, but it requires significant manual attention in all but the simplest cases. Therefore there is a considerable gap between the function of The Tractor and what is required for front-to-back catalog pipelines. Developing such pipelines is not only time consuming, but independently developed pipelines perform differently (e.g., that of Nyland et al. 2017 is different the pipeline of Dey et al. 2019b). While each implementation may be optimized for a certain task, the overwhelming success of software like Source Extractor is that they are immediately accessible, reproducible, and easy to use. However, the matters of source detection, model type decisions, which groups of sources to model and how best to model them, as well as computational efficiency are challenges that must be addressed if we are to construct such a generalized pipeline that applies The Tractor to the incredibly deep, crowded fields to be explored by the next generation of galaxy surveys.

2.3 THE FARMER: A GENERAL DESCRIPTION

The Farmer is a generalized, flexible, and reproducible framework that uses the model library from The Tractor and its optimization engine to photometer detected sources, measure their shapes, produce output catalogs and ancillary images, as well as provide supporting diagnostics. The Farmer overcomes the issue of how to assign model types by identifying natural groups of nearby sources and determines the best model type using a decision tree in a time efficient, optimal way whilst mitigating related pathological situations[3]. It includes a significant organizational capacity such that images can be divided up into sections for massively parallelized computation. Here we walk though the process of The Farmer from image preparation to the output catalogs.

2.3.1 *Image preparation*

At bare minimum, The Farmer requires a single science image containing sources of interest. A corresponding inverse variance weight map is ideal, but not required. Lacking weight information, The Farmer can measure noise directly from the images or assume equal weights.

In this basic case, The Farmer will detect sources, model them, and perform forced photometry all on the same monochromatic image. In more typical, complex cases it is desirable to produce a separate detection image. For surveys of faint sources, the `CHI-MEAN` approach (Szalay et al., 1999a; Bertin, 2010a) has been widely adopted (e.g. Laigle et al., 2016; Weaver et al., 2022a), or a similar signal-to-noise image co-add.

Masking is especially important in profile-fitting photometry for the reason that it is inadvisable to attempt to model large, saturated stars, nebulae, or nearby galaxies which are essentially nuisance foreground contamination. While apertures have the advantage of being able to efficiently sum fluxes in whatever regions of an image are of interest, models must attempt to describe the image as it is. Attempting to model such nuisance sources, which lie outside the reach of our parametric models, will never reach a satisfactory

3 With thanks to the general outline provided by the DESI legacy survey pipeline: `https://github.com/legacysurvey/legacypipe`

fit even after several hundred CPU hours, if at all. That being said, without extensive background modelling, sources within bright star halos will not be photometered accurately with apertures either.

A useful recipe is to stack all bands which will be used to detect sources, and mask out the full extent of such nuisance foreground objects, and possibly also the edges of the mosaic or detector. The Farmer can be configured to apply a mask before or after source detection. The latter is preferred in virtually all cases, as mask edges can produce spurious sources. Applying a mask after source detection simply removes the sources from the catalog and their corresponding segments are zeroed out.

The Farmer includes several ways to measure image backgrounds and per-pixel noise based on SEP, and can be configured by the user. Backgrounds can be measured as global medians or spatially varying (following the methods of Source Extractor; see Bertin et al. 1996), with per-pixel noise being estimated directly from the RMS of the image. The background and per-pixel noise estimates can be produced with and without the mask in order to mitigate the adverse effects of bright stars and foreground galaxies.

2.3.2 *PSF creation*

With the images and weights in hand, The Farmer needs a PSF for each band of interest. There are many way of generating PSF stamps, including as realizations of spatially varying models, and The Farmer can be supplied with several PSF types.

The most common is a constant PSF stamp sampled at the same pixel scale as the its corresponding image; these can be readily produced by packages such as PSFEx (Bertin, 2013). One may also use PSFEx to generate spatially varying PSF models, all flavours of which (e.g., Gauss-Laugere or pixel bases) are understood by The Farmer (and importantly also by The Tractor). While this can be achieved through using PSFEx by itself, The Farmer is able to run PSFEx in a semi-automatic way using built-in functions. First The Farmer runs Source Extractor to identify bright sources. Candidates for point sources are then selected either automatically by PSFEx, or more efficiently by a pre-selection by the user based on source FWHM and brightness. The user can also declare which bands should use a constant PSF and which should be spatially varying, and The Farmer will automatically reconfigure PSFEx in each case.

In some cases the PSF varies too quickly across an image to be accurately characterized by a smoothly varying surface as used by PSFEx. It is possible therefore for the user to supply a set of PSFs and a file which maps each one to a coordinate so that The Farmer can use the nearest sampled PSF for a given source. The assumption of a smoothly varying PSF can thereby be avoided, and the user is free to choose the grid geometry according to their requirements. This 'PSF Grid' approached was developed in Weaver et al.

(2022a) to characterize the photometry of the Subaru Suprime-Cam mosaics in COSMOS.

The images of Spitzer's Infrared Array Camera (IRAC) feature a highly variable PSF which is generally triangular in shape. The PRFMap package (A. Faisst, private communication[4]) attempts to characterize this highly irregular PSF by mapping the pixel of each stacked image back to the locations on the CCD of the constituent images. It then uses the spatially-dependent calibration PSFs to construct a combined PSF for the stacked image. Similar to the PSF Grid approach, PRFMap produces a library of individual PSFs corresponding to a fixed grid of sampling coordinates. This output can be used with The Farmer to measure IRAC photometry.

One important caveat to note is that in all cases the PSF must be measured into its wings and not be truncated. This is for two reasons. Firstly, profile-fitting models generally benefit from the wings of the PSF being in tact. This can be immediately appreciated in the case of unresolved sources fit with point-source models for which The Tractor uses the PSF stamp for the model profile: if the wings of the point-source model do not describe the full spatial profile of the source of interest then the fit will suffer and the measured flux may be biased. Secondly, the pixel values of a PSF which has been truncated and then normalized to unity will be larger than those of the full PSF normalized to unity, and so its optimal scaling coefficient (i.e. flux) will be smaller for the same source, introducing a bias. Therefore it is strongly advised to sample the entire PSF profile out to radii where the wings are indistinguishable from noise (in most cases corresponding to a radius of several arcseconds).

2.3.3 *Source detection*

The first step in catalog creation is source detection. The Farmer utilizes SEP to provide source detection, segmentation maps, background, and noise estimation with near identical performance as classical Source Extractor. Detection parameters for SEP can be configured directly with The Farmer, and related diagnostic images are supplied indicating source centroids on the detection image. It is also possible to hand The Farmer a catalog of source coordinates and a corresponding segmentation map from e.g., Source Extractor, or any other similar detection software.

The Farmer performs all functions on discrete sections of the total mosaic called "bricks". An example is shown in Figure 2.2. Each brick is cut out of the total mosaic image, weight, and mask with equal dimensions, and includes a buffer region on each side. Sources detected with centroids in the buffer region are removed from the source catalog of the brick, and their segments are set to zero. They are not lost, however, as they are found again in the main region of a neighboring brick. This 'fuzzy boundary' approach means The Farmer can construct unique source catalogs for each brick which have no

4 Available on github: `https://github.com/cosmic-dawn/prfmap`.

overlap with neighboring bricks, thus accounting for every source without duplication or loss. Although the buffer regions of the segment map are also set to zero, The Farmer keeps segment pixels in the buffer region of sources whose centroids are in the main region of the brick. This behavior allows sources which are near the buffer zone to be modelled with all of their pixels, as opposed to a strict cut-off at the buffer boundary where their profiles would be truncated.

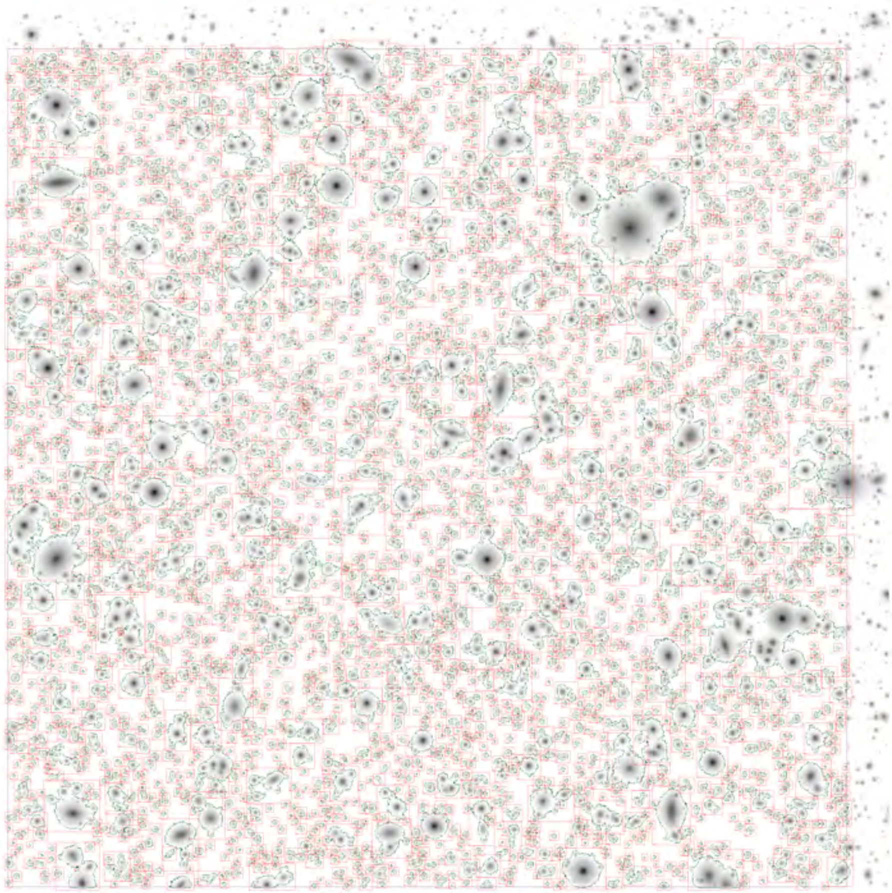

Figure 2.2: Example of groups detected over a brick. The brick lies at the edge of a mosaic and so has boundaries with two adjacent bricks. Groups are outlined by red boxes and their footprint of owned pixels are shown by the red borders.

Following the creation of the brick's preliminary source catalog and cleaning of the buffer regions, The Farmer attempts to identify natural groups of detected sources which would benefit from being simultaneously modelled. Groups are identified by dilating the original segments to form contiguous non-zero regions. Sources which are not in crowded areas form singularly occupied groups, whereas sources in crowded regions end up members of larger groups to be modelled simultaneously. See Section 2.4.1 for further discussions.

2.3.4 *Model type and shape determination*

A model must now be determined for each source in a given group. The goal is to not only determine the most suitable model for each source, but also its best-fit parameters. While the number of possible decision tree architectures is virtually infinite, The Farmer relies on a balanced architecture consisting of five discrete models to describe resolved and unresolved, stellar and extragalactic sources:

1. **PointSource** models are taken directly from the PSF used. They are parameterized by flux and centroid position and are appropriate for unresolved sources.
2. **SimpleGalaxy**[5] models use an exponential light profile with a fixed user-defined effective radius such that they describe marginally resolved sources and mediate the choice between PointSource and a resolved galaxy model. They are parameterized also by flux and centroid position.
3. **ExpGalaxy** models use an exponential light profile. They are parameterized by flux, centroid position, effective radius, axis ratio, and position angle.
4. **DevGalaxy** models use a de Vaucouleurs light profile. They are parameterized by flux, centroid position, effective radius, axis ratio, and position angle.
5. **FixedCompositeGalaxy** models use a combination of ExpGalaxy and DevGalaxy models. They are concentric, and hence share one centroid. There is a total flux parameter as well as a parameter for the fraction of total flux assigned to the DevGalaxy component. Each component has their own effective radius, axis ratio, and position angle.

In practice, the parameters of these spatially-resolved models are softened such that there are no forbidden regions in the fitting space. Ellipticities are processed through a sigmoid-like function $1 - \exp(-|e|)$, and effective radii are fit in $\log_e$ space, which widens their domains to $(-\inf, \inf)$.

5 SimpleGalaxy models are not included in Tractor by default, see https://github.com/legacysurvey/legacypipe.

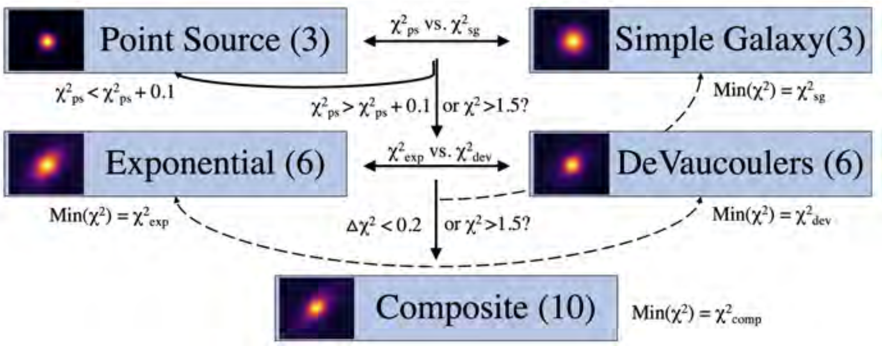

Figure 2.3: Schematic of the decision tree used by The Farmer to determine the most suitable model type for a given source. The five models are tested in order of increasing number of parameters as indicated by the number rightward of the model name. Solid arrows indicate forward progression to an equally or more complex model while dashed lines indicate that a less complex model is preferred. Values shown are examples, and can be configured by the user.

These five models form The Farmer's decision tree, whose goal is to both determine the most suitable model for a given source, and provide an optimized set of parameters to describe the shape and position of that source. To ensure that crowded regions do not suffer from poor modelling as a result of the model of a particular source being constrained by light from neighboring source, the models are determined simultaneously at each stage of the decision tree. The values for the decision tree parameters quoted here are taken from Weaver et al. (2022a) but can and should be tuned by the user for other data sets. An example of a group containing two sources progressing through the decision tree is shown in Figure 2.4.

The Tractor uses a likelihood cost function to score the performance of the joint model containing all of the invididual models of the sources in a group. All weight pixels outside the group footprint are set to zero such that nearby sources which are not part of the group cannot influence the likelihood. However, we also need to be able to assess the performance of an individual model for a given source in our group. The Farmer adopts χ^2_N as its goodness of fit statistic, which is calculated by quadrature addition of the residual image pixels belonging to a particular source by its original segment and then reduced by dividing by the number of free parameters N, taken as the difference between the number of pixels and the number of free parameters.

The Farmer begins by considering PointSource models for every source in a group, using centroids and fluxes estimated by SEP as initial conditions. The Tractor then performs an optimization to maximize the combined likelihood of the entire joint model, after which The Farmer computes the χ^2_N for each source in the group. Next, SimpleGalaxy models are considered for all sources in a group with the same initial conditions as before. The models are optimized and the χ^2_N per source is computed. The Farmer then tries to place each source into one of three categories: either the source is well fit by the PointSource and is fixed as a PointSource, it is fit well by a SimpleGalaxy, or neither model is appropriate. Satisfying either of the last two categories advances the source down the decision tree towards more complex, resolved model types. The role of the SimpleGalaxy here is not to be a commonly chosen model, but rather an indicator of a resolved source. Unlike comparing a PointSource model to a more complex ExpGalaxy model, the comparison with the SimpleGalaxy is not only computationally faster but is statistically fair since the number of parameters for both PointSource and SimpleGalaxy are the same, as are the number of data points. Sources which are best-fit with PointSource models will be assigned a PointSource model hereafter, which in the case of a one source group will terminate the decision tree. A source that is better fit by a SimpleGalaxy model by only a slim margin is typically sufficiently modelled by a PointSource also. It is desirable therefore to prefer a PointSource in these cases as a better fit. SimpleGalaxy model triggers the more complex tiers of the decision tree, meaning that the overall group model becomes more complex which requires even longer computational times. THE FARMER therefore penalizes the SimpleGalaxy models in χ^2_N by 0.1 such that a SimpleGalaxy model must have a lower χ^2_N by a margin of 0.1 or better in order to not choose a PointSource. A PointSource will also not be chosen if produces a bad fit, assessed by $\chi^2_N > 1.5$. If a PointSource is not chosen, the source continues to the next level of the decision tree. A SimpleGalaxy may still be chosen in the end, but only if it is still favored after the assessment of more complex models.

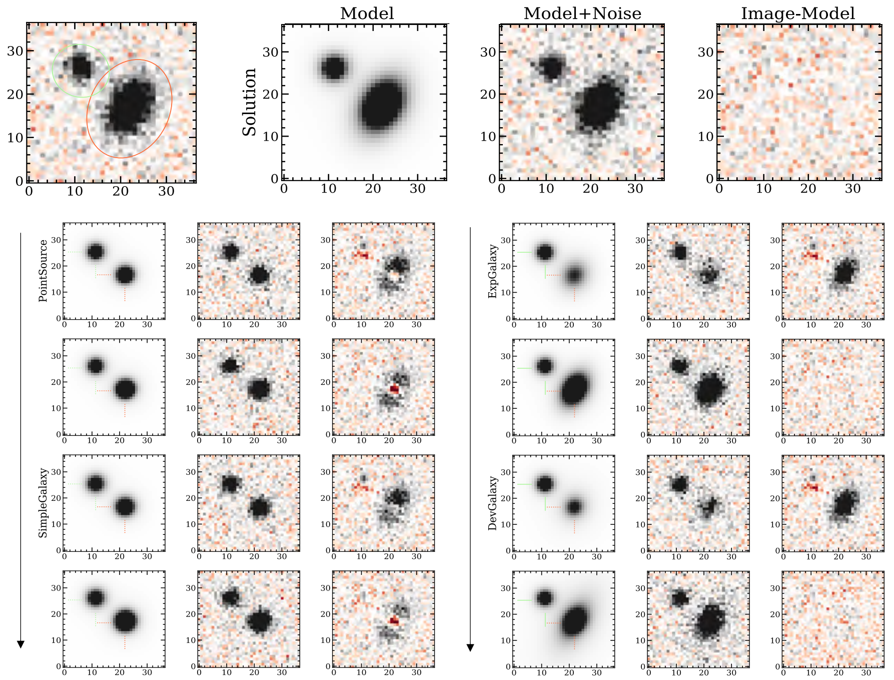

Figure 2.4: Example of a decision tree process for a group containing two identified sources. The input z-band image (top left) and model images (leftmost columns in each set) are scaled by $\log_{10}$ to highlight morphologies. The models are shown realized in a noise field sampled from the weight map RMS for illustration purposes (center columns) and residuals are computed by image - model (right columns). Both are scaled by $\pm 3\sigma$ to highlight faint signal and oversubtraction. Each row is a set whereby the marked model type is introduced to undetermined models and the optimization result is shown in the panels below it. This particular pair of sources satisfied the decision tree before reaching the FixedCompositeGalaxy stage. The final optimization is shown at the top.

The next stage of the decision tree determines the general Sérsic light profile of resolved sources whose model types remain unfixed, choosing between ExpGalaxy or DevGalaxy. At this stage, fixed sources can only have been assigned PointSources. The Farmer starts by considering ExpGalaxy models for all other unfixed sources, performs the optimization, and determines χ^2_N for each. Initial guesses for shape parameters are initialized borrowing from the SEP measurements estimated at detection. Then The Farmer performs the same computation but with DevGalaxy models on all unfixed sources. Again the χ^2_N is a fair comparison as the number of degrees of freedom are identical between the two model types. The Farmer allows the model parameters to remain variable for all sources, regardless of whether they have been assigned a final model type, at each stage of the decision tree (e.g., fixed PointSource models still re-optimize their flux). Sources whose ExpGalaxy and DevGalaxy models both fail to achieve a lower χ^2 than the SimpleGalaxy are fixed as SimpleGalaxy models, unless the SimpleGalaxy also fails to achieve a χ^2_N of 1.5 in which case that source advances down the decision tree to the third tier. The choice between ExpGalaxy and DevGalaxy models is determined by the lowest χ^2_N, without any penalties. However, if the absolute difference in χ^2_N

between the two models is less than 0.2, or neither ExpGalaxy or DevGalaxy achieves a χ^2_N of 1.5, the source also advances to the third tier.

All sources have typically been assigned a fixed model by this stage, especially those which have smooth light profiles or are unresolved, and the decision tree ends without trying more complex, time intensive models which The Farmer has already determined are not required for a sufficient fit. However, highly spatially resolved sources that have reached the third tier without an assigned model are fit assuming the most complex FixedCompositeGalaxy models. If the FixedCompositeGalaxy model fails to achieve a better χ^2_N than either ExpGalaxy or DevGalaxy, the source is fixed to the model type with the lowest χ^2_N overall.

Now that models for all sources belonging to a given group are assigned, The Farmer optimizes the models a final time. This is an important step as it is possible for an otherwise pathological case to arise whereby two assigned models were never optimized at the same time and their fits may influence each other. By computing this final optimization, the overall likelihood of the model set for the group of sources tends to improve slightly.

Forced photometry

The objective of forced photomety is to measure fluxes and estimate their uncertainties in other bands of interest, for already known, detected sources. Now that models types have been assigned and their parameters optimized for each source in a given group, it is straight forward to apply these parametric models to photometer these sources in other bands of interest. In most applications, it is advisable to fix the model shape parameters and only allow flux (α in Equation 2.1) to vary. However, The Tractor provides the flexibility to allow shapes and positions to vary as well; they can be unbounded or limited by a Gaussian prior. For example, it may be desirable to allow the shape to change in the presence of morphological differences between the model bands and the forced photomety band, or allow the position to vary if there are significant astrometric offsets (see discussions in Section 2.4.3). The Farmer enables the user to choose which parameters (if any) are fixed during the forced photometry stage.

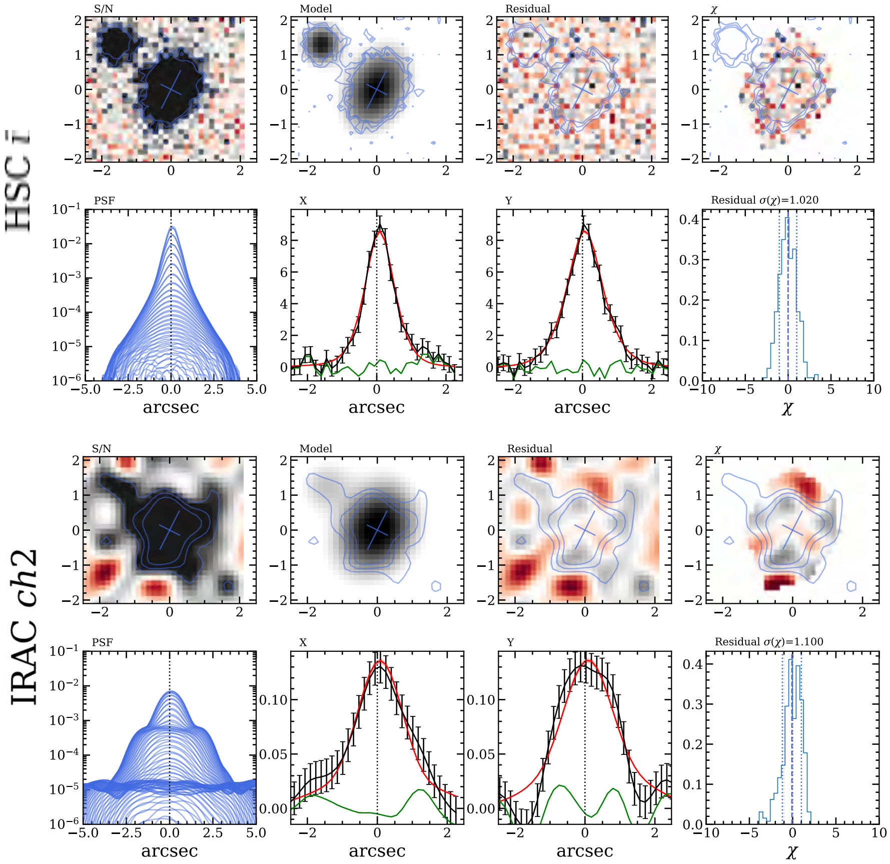

Figure 2.5: Example of results from forced photometry for the brightest source in Figure 2.4 measured in i and channel 2. The image of each band is shown next to the best-fit model from The Farmer. Brightness contours and principal axes are overlaid on the model in blue. The residuals are shown $\pm 3\sigma$ (same as the image) to highlight faint signal and any over subtraction. The rightmost panel shows the per-pixel χ image scaled ± 3 computed within the bounds of source segment Bottom rows for each band show a wiremesh representation of the PSF profile; slices though the source (black), best-fit model (red), and residual (green); and the distribution of χ values over the group pixels which on expectation should be normally distributed. The median and 68% range of the distribution is shown for illustration.

Optimization proceeds on a group-by-group basis so that the forced photometry can benefit from the same advantage as in the model stage by simultaneously optimizing all models belonging to a given group. Each model is convolved with the PSF of the band of interest and realized into the frame of the image, including images of different pixel scales to that of the detection image[6]. The group models are then simultaneously optimized until their joint likelihood converges, or until some maximum iteration set by the user. Figure 2.5 shows the results of forced photometry using the same sources from Figure 2.4. While this procedure is generally faster than the model stage, forcing photometry on dozens of images may approach a similar computational expense. Computational strategies are discussed in Section 2.4.6.

6 Currently, The Farmer requires pixel scale homogenization, but this restriction will be removed in a future release.

Catalogs and other output

After the modeling stage, The Farmer produces an intermediate catalog containing the source IDs, including their brick and group numbers, followed by the detection parameters from SEP. For each source, the best-fit model type (e.g., PointSource or ExpGalaxy) are recorded, as well as their best-fit parameters and associated uncertainties. Shapes and sizes are not measured for sources assigned unresolved models. Fluxes are also measured for each source in every band used in the modelling stage.

A number of residual statistics are also included that provide valuable insight into robustness of a given model for a given source and band. In order to minimize contamination with neighbours, we consider only the pixels belonging to the source segment in the computing these estimates (same as in the decision tree). The primary statistic is χ^2, already discussed in Section 2.3.4. Three other related statistics are produced by measuring the moments of the inverse variance weighted χ_i images where each i pixel value indicates the significance of the residual in units of per-pixel uncertainty σ_i: the median $\mu(\chi)$, standard deviation $\sigma(\chi)$, and D'Agostino's K^2 test which measures the normality of the residual by combining estimates of skew and kurtosis[7] (D'Agostino, 1970; D'Agostino et al., 1990). These statistics can also be combined to separate reliable models from poor fits and blends, as shown in Figure 2.6.

Once forced photometry is completed, The Farmer appends the measurements to (a copy of) the existing model catalog. This can be done on a band by band basis, or for all bands simultaneously. Output includes fluxes, as well as other parameters including band-specific positions and shapes if the user has allowed them to vary. Residual statistics are also included for every source in each measurement band.

The Farmer has an additional diagnostic ability to measure photometry of these known, detected sources with concentric circular apertures of various diameters specified by the user. This is especially useful for constructing profiles of the radial flux growth. The photometry can be measured on the science images (to get basic comparisons with the profile-fitting results, although without PSF homogenization), and it can go further by forcing the same apertures on the residual image and weight images. Most interestingly, these apertures can be forced on images constructed by realizing the entire group of models into pixel-space which can be readily compared with fluxes measured on the same apertures on the image itself. Similarly, apertures can be forced on single sources realized into the pixel-space of the image in complete isolation; these measurements can be compared with the total flux reported by The Tractor[8]. Similarly, The Farmer can also use SEP to produce forced photometry measurements similar to `MAG_AUTO` from Source Extrac-

7 The K^2 test is stable only for sources which have more than 8 pixels in their segment.

8 However, if the model is severely truncated by being realized into an image whose dimensions are much smaller than the full extent of the model then the integrated flux in large apertures will underestimate the total, correct flux measured by the scaling coefficient.

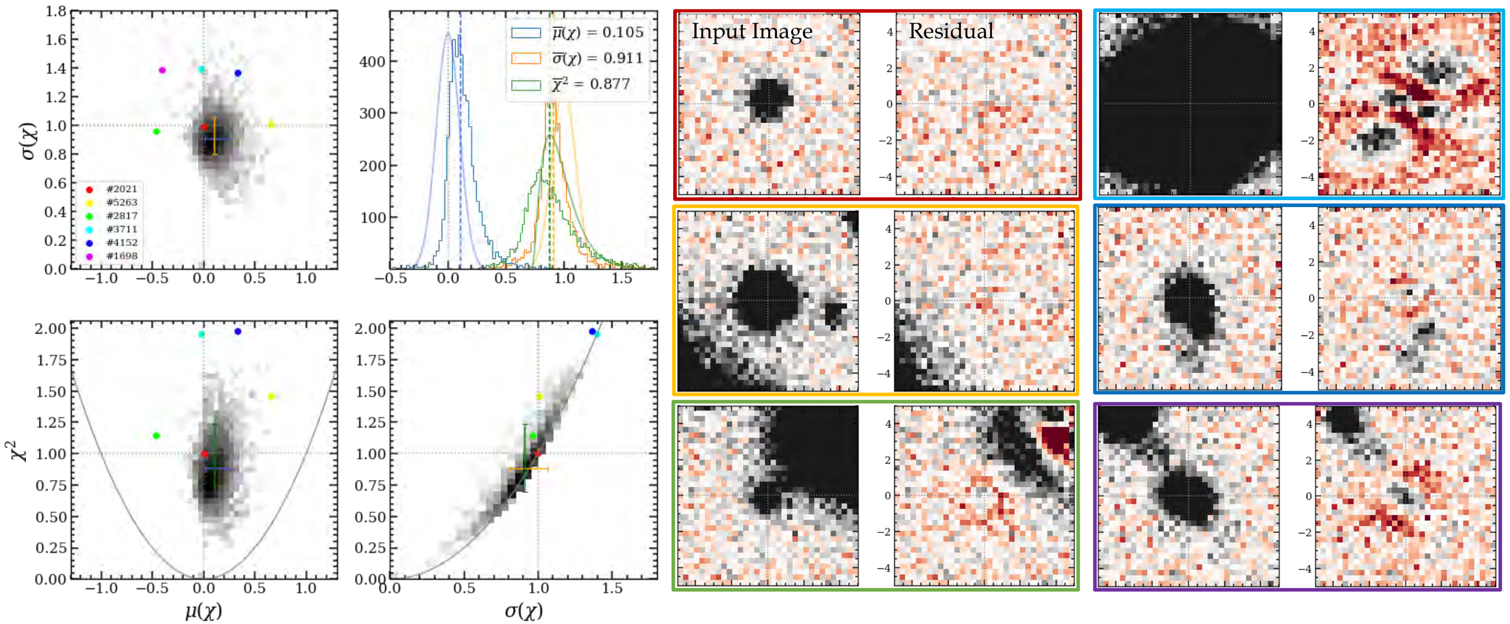

Figure 2.6: Residual statistics can be used to identify bad models and blends. *Left:* (joint) distributions of the χ^2 (green), the median χ distribution ($\mu(\chi)$; blue), and the standard deviation of the χ distribution ($\sigma(\chi)$; orange). Solid curves indicate distributions for expectation: $\langle\chi^2\rangle = 1$, $\mu(\chi) = 0$, and $\sigma(\chi) = 1$. Each statistic is measured from the residual of each source within its unique segment pixels. Grey curves in the show the allowed regions of the joint distributions in the lower panels. *Right:* Six sources are chosen in representative regions of the space, whose input image and residual (scaled to $\pm 3\sigma$) are shown, with red being at expectation for all three statistics.

TOR, including estimates of Kron radii (Kron, 1980b). Together, these aperture measurements can help diagnose model inaccuracies and bias, providing an effective means to internally validate the results of The Farmer.

Diagnostic images can be incredibly useful. The Farmer can be configured to produce pixel-level background and RMS maps in addition to source and group segmentation maps. Importantly, The Farmer can realize the entire model library of a brick as a reconstructed pixel-level model image from which corresponding residual and weighted significance χ images can be produced. Since catastrophic failures can result in models spanning large regions of the reconstructed model images, The Farmer allows the user to automatically filter models based on χ^2_N or axis ratio such that they are not included in the reconstructed model, residual, or χ images (especially useful for cleaning residuals when searching for undetected signal). Also, models with negative fluxes will create positive flux in residual images; these can also be automatically filtered. Although removing sources at this level introduces incompleteness, it is likely that the measurements of these problematic sources are not scientifically useful. To account for the missing area, The Farmer also provides an effective mask image which flags pixels belonging to removed sources according to their segment ownership and computes the effective area of that mask. Although laborious, this is an optimal system for precisely determining the effective area from which a cleaned sample has been selected. Caveats regarding these reconstructed images are discussed in Section 2.4, below.

2.4 CONSIDERATIONS, ASSUMPTIONS, AND LIMITATIONS

Although The Farmer effectively extends the functionality of The Tractor to include source detection and grouping, model assignments, catalog creation, and computational efficiency, these advantages come with considerable limitations which are discussed below.

2.4.1 *Image preparation & source grouping*

Several aspects of the image preparation and group identification stages are unique to The Farmer.

How should one determine how many bricks should a mosaic be broken into? This is primarily a computational concern. We can understand why by considering the combined perimeter of all the bricks; it is large when bricks are small and vice versa. The larger the perimeter, the greater the chance that the brick will split across a group of sources which should be ideally modelled simultaneously in the same brick. In general this should be avoided, and so bricks should be made as large as possible. One constraint is that each brick can be operated on by The Farmer independently, which means they can be parallelized across computational nodes[9]. The brick also needs to be read into memory at runtime, and so should be sized appropriately for the memory capacity of a given computational facility. Bricks from forced photometry are typically the largest files as they contain all bands of interest, their weight maps, and masks and so can become tens of gigabytes for even modest dimensions.

What about sources near the edges of bricks that extend into the next brick? It is up to the user to determine how large the brick buffer should be. In general, the buffer should be large enough that the largest sources of interest, placed at the brick edge, would not extend beyond the buffer. While one can set a large brick buffer, doing so comes at the cost of memory and computational overhead. Excessive brick buffers should be avoided where possible.

How should one assess if groups are correctly identified? As discussed in Section 2.3.3, groups of sources are identified by joining source segments which have been dilated by some morphological structure whose size dictates the extent of the dilation. The segments are constructed from the detection stage, and so one should only consider the detection image when assessing the identification of groups. Dilation is necessary because the segmentation extent in Source Extractor (and SEP) is tied to the significance level set for the detection. In some cases the segments may not capture the full extent of two neighboring sources such that they should be simultaneously modelled, but their segments do not touch. Hence the size of the dilation structure should be set so that these kinds of nearby sources are correctly assigned into one group. This is most easily assessed by inspection, and tuned in successive trials.

It is important to note that morphological dilation can destroy segments nearby to larger ones. For this reason the dilation is carried out on a copy of the segmentation image which has been made binary such that pixels assigned to the background are set to zero, and those active pixels assigned to sources are set to 1. Segments which are already touching are now indistinguishable, and the dilation simply enlarges the footprint of contiguous regions of active pixels. The small segments remain identifiable from the segmentation image. This is important, because the group pixels belong to the group itself; no one group pixel belongs to a single source. That ownership is retained in the original segmentation image. This is essential because while the joint likelihood maximized by The Tractor is computed over the group pixels, we maintain the ability to judge the fit of individual sources from the χ^2 computed over their uniquely owned pixels.

In some cases the segments produced at detection may be too large and so over-group sources. While this is not a problem scientifically, it increases the computational complexity of the fit which can lead to poor model performance, or worse, the joint model may even fail to converge altogether. However, unlike morphological dilation which cannot destroy groups of pixels, morphological erosion can destroy the smallest segments typically containing one source.

9 Or even processed by different computing facilities altogether.

This is a limit that must be avoided in order for The Farmer's decision tree to work. More work is required to address this case.

A limitation of this approach is that groups are defined based on the detection image, its effective resolution, as well as the depths and properties of constituent bands. A group determined on well-resolved optical images will likely miss pixels with significant flux when applied to typically wider sources in lower resolution images. This can be overcome by further dilating each group on a band-by-band basis such that all of the relevant pixels are now constraints on the model. This comes with a dilemma, however, as crowding is worse in low resolution images of the same depth and so light from sources not originally included in the group may now contribute. Yet they are not described by the group model, and so leaving their flux unaccounted for may instigate a bias in the photometry. The only tractable option seems to be to join these groups and perform the forced photometry in a simultaneous optimization. However, the shapes of these models were never determined together, and so it is uncertain how well the new group of models would perform. Worse, most sources in the deepest IRAC images are blended to some degree and so keeping to this philosophy of joint optimization of all overlapping sources would require possibly every source to be simultaneously fit, which is maximally complex and computationally prohibitive. This dilemma will be addressed in future work.

What sets the buffer sizes for groups? Although groups of sources are limited to their footprint whose pixels are identified by dilating source segments, the groups themselves are saved in memory as rectangular arrays whose dimensions are set by the maximum extent of the group footprint. Although pixels outside the footprint (which can often be fractal-like in shape) do not provide any constraining power as their weight is set to zero, the models are still realized onto these exterior pixels during the optimization. It is generally best if these models are not truncated whatsoever, and so The Farmer adds a buffer by enlarging the dimensions of the group array. This is not only for numerical reasons internal to The Tractor, but also is a requirement if post-processing apertures are to recover the full extent of the joint model image of the group sources. Truncation of that joint model will mean that the wings will not be realized and so the largest apertures will underestimate the true flux. Thankfully, if the models are correctly normalized then the truncation will *not* affect the best-fit scaling coefficient from which the source flux from The Tractor is derived. It is therefore advisable for the group buffers to be large enough so that the PSF stamp would not be truncated for a source placed near the edge of main group footprint.

2.4.2 *Selection functions and image depth*

One must be cognisant regarding which band should be used to determine the models and their best-fit parameters. In fact, this is not a free choice.

Using a band outside of those used in the detection image is inappropriate because there may be sources identified in the detection image which do not have flux in the chosen modelling image. If one is to maintain the selection function constructed by the detection strategy, then it must be guaranteed that a detected source has sufficient signal to constrain its model. Otherwise sources without models cannot be photometered, and so the selection function changes in a non-trivial way. For the same reason, it is also inadvisable to use only one band of a multi-band detection image, or even the bands which define the spectral domain of the detection image. Nor is it advisable to attempt to model sources in a co-added image as the effective PSF is not easily characterized, and the FWHM of the constituent PSFs can produce additional variation in the surface brightness profiles. Therefore, it is strongly recommended that the models be produced from precisely the same bands and images that were used or combined to make the detection image.

Measuring photometry of a source in an image which contains additional sources outside the selection function (by virtue of not being detected) presents another often encountered dilemma, although common also to aperture-based methods. This is because signal from an additional, undetected source is not described by the set of models assigned to a group. For instance, a red source which is undetected in a predominantly blue selection function may in a red band appear next to the known blue source. Although fixing model shapes helps avoid contamination, it is possible that the likelihood will be maximized by increasing the flux parameter of the blue model such that some of the flux from the new, red source is inadvertently accounted for, thus biasing the photometry for the blue source in that red band. Often times these cases can be identified afterwards from diagnostics provided by The Farmer, although not guaranteed.

A similar situation is encountered when forcing photometry onto deeper bands of the same wavelength as the detection, and although such images typically can provide better photometric constraints, they may at the same time introduce bias. This means that ideally all sources in a forced photometry image should be modelled. However, identifying these new sources automatically ahead of photometry is not practical as lists of detected sources will differ due the blends; the two catalogs must then somehow be reconciled and segmentation maps merged. Doing so in limited numbers is possible with careful supervision, typically with the assumption that new sources are unresolved to avoid re-processing the decision tree (as used to photometer optically dark sources in Jin et al., 2022, submitted). This potentially pathological issue will be addressed in future work.

2.4.3 *Models, morphological corrections, and drifting*

One significant complication with The Farmer is that the decision tree needs to be tuned. Because the central operation of the decision tree is to separate resolved and unresolved sources, its parameters are most sensitive to the resolution of the image. Size correlates strongly with apparent brightness, and so sources in deep images typically become unresolved around a certain magnitude threshold. In order to succeed, the decision tree needs to be tuned such that it correctly assigns unresolved models to essentially all sources fainter than this limit, in addition to bright point-sources. A photometric bias can develop if instead the decision tree assigns resolved models to unresolved sources, or vice versa. This can be readily diagnosed from number counts which should be smooth and increase monotonically with decreasing brightness. If the decision tree is not providing adequate model type assignments, the number counts of the detection bands will either contract towards a sharp rise or flatten into a plateau around the resolution threshold. An unresolved model assigned to a resolved source tends to produce an underestimated flux, thereby moving these typically bright but comparatively rare sources towards fainter magnitudes thus creating a plateau. In this case it is likely that the decision tree is tuned so that PointSource models are too easily assigned, and so the χ^2 criterion of the PointSource should be lowered. A resolved model assigned to an unresolved source tends to produce overestimated flux, thereby moving these typically faint but abundant sources towards brighter magnitudes thus creating a sharp rise in counts. In this case the χ^2 criterion of the PointSource should be increased so that it is easier for sources to be assigned an unresolved model. Number counts are not as sensitive to which resolved model is assigned to a resolved source (e.g., ExpGalaxy or DevGalaxy) and so the corresponding parameters are most easily tuned by examining residuals of bright, resolved sources.

It may not be possible to assign a simple parametric model to a particular source. It might be that the source is actually two blended together. Meanwhile, the brightest sources tend to be resolved and have features such as spiral arms, bars, and starbursts that are not described by the smooth models from The Tractor. As such, model performance tends to decrease for bright, resolved models. This is especially true for space-based imaging (e.g., *HST*) where the space spanned by models from The Tractor are divorced from the real space spanned by highly resolved galaxies. In such cases one may find success with traditional aperture-based approaches for all but the faintest, least resolved but can and should be tuned by the user for other data sets.

Chromatic changes in morphology presents a challenge for The Farmer. The model for a given source during the modelling stage may be simultaneously constrained by multiple bands, but The Tractor allows only one shape shared between the bands. Therefore the shapes reported by The Farmer from the modelling stage are averaged. Forced photometry in regular operation proceeds by only allowing only the flux to vary with the shape fixed, meaning that changes in morphology are not accounted for by the model. However, The Farmer makes it possible to perform forced photometry on each band separately so that the shape parameters can be allowed to vary in each case with or without priors, albeit at greater computational expense and danger of over-fitting. It is important to note that forcing models derived from well-resolved bands onto images of lower resolution is typically successful as the larger PSF of the forced photometry band makes the photometric measurement less sensitive to morphology. However, forcing models derived from low resolution images onto bands at high spatial resolution typically results in a poor performance as the band of interest contains more information than the model can describe.

As discussed in Portillo et al. (2020), flux and shape estimates can suffer from biases introduced from inadequate centroiding. Given the great number of multi-wavelength images and facilities involved in modern surveys, astrometric offsets due to even minor bulk flows can impact the measurements derived from model fitting. Hence, The Farmer allows the user to unfix the centroid position of each model and introduces a Gaussian prior on its position, on a band-by-band basis. This prior acts to penalize the likelihood of the model fit if the model obtains a centroid (i.e. 'drifts') that is beyond the distance set by the prior. This drifting can be especially prevalent in the case of a known faint source next to an undetected bright neighbor which because it is not accounted for by the model will cause the model of the faint source to move towards the bright source, whose unaddressed presence counts against the likelihood more than the original, fainter source. Priors can be set on the position, although their widths are usually determined by successive trials.

It is important to appreciate that the grouping of sources imparts a significant advantage. Because groups of sources are photometered in isolation, a failure of the model in one group does not affect any other group. Let us consider the unfortunate example in which a galaxy is assigned an inaccurate model whose wings extend beyond the group. While those wings will be a problem for the source in question, and perhaps its group members as well, they will not affect any other group in the image. Hence, while this is an issue for the residual map, there is no reason to be concerned about the photometry of the other group as they were fitted in an entirely separate optimization in isolation of the problematic source. However, this advantages effectively decouples the reconstructed brick-level residual image from the photometry and so complicates searches for sources in residuals. As mentioned in Section 2.3.1, The Farmer has built-in functions to filter out these problematic models.

2.4.4 *Source de-blending*

While profile-fitting photometry can be used to de-blend two sources, they first must be identified as separate sources which is dependent on the original source detection. As such, de-blending sources at the detection stage is not a problem which profile-fitting photometry can (or should) solve, which instead is well-suited to address the related, but distinct issue of accurately measuring the flux of two *identified* but blended sources. It is essential, therefore, to understand that if two nearby sources are not successfully detection de-blended, then profile-fitting techniques should not be expected to reliably de-blend them either.

This concept is demonstrated in Figure 2.7 whereby a point source is placed in the vicinity of a bright resolved galaxy and appropriate models are assumed to be known. Several cases are set up by varying their distance and relative brightness. Attempting to photometer both of them with only one model produces expectedly poor fits in several cases. The system is then evolved by allowing it to subtract the first source, find the brightest residual source, and re-fit using two centroids which in turn improves the performance in cases where the residual source can be identified. However, in practice this is dangerous one does not know beforehand if there is another source or if the model for the one source was simply a poor fit. Lastly, the two sources are fit by two appropriate models which results in accurate photometry at all distances and relative brightness. Hence, profile-fitting photometry succeeds if and only if it is first provided with the correct number of models (and centroids) to optimize for a given group of sources.

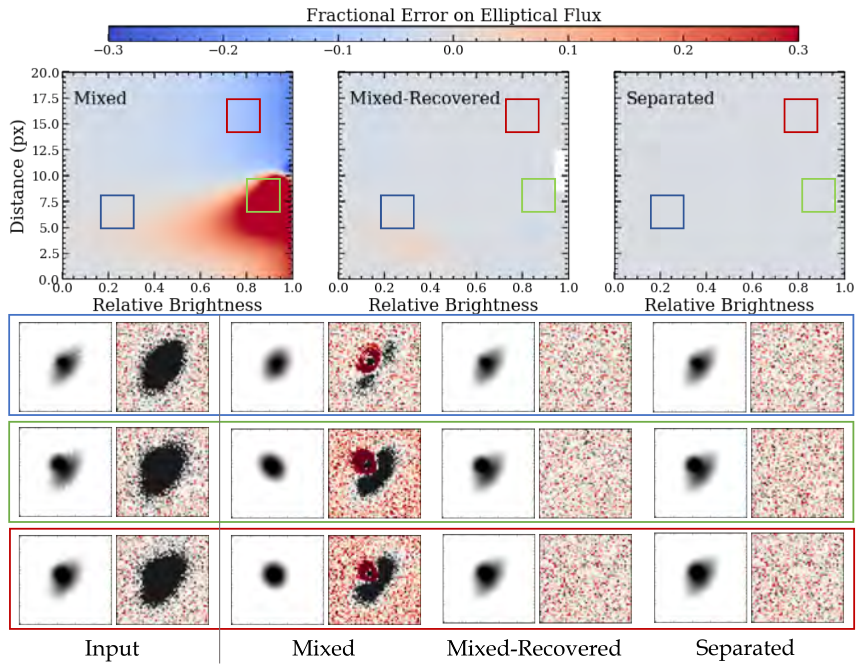

Figure 2.7: A point source is simulated in the vicinity of a large, central elliptical galaxy. Models are fitted for three cases: the sources are blended and have only one centroid (Mixed), the sources are blended, modelled, and then the missing source is recovered from the residual image and modelled (Mixed-Recovered), and lastly both sources are a priori known and simultaneously fit (Separated). Each measurement is repeated over a grid of relative brightness (0 meaning that the point source is negligable) and the distance from the central elliptical to the point source (20 means the point source is in the top left corner). While grey areas indicate successful recovery of the input elliptical flux, red areas indicate that the flux of the elliptical is underestimated. White areas in the middle panel indicate where the point source is not detected in the mixed residual. The model and residual of three situations are shown for each measurement strategy.

What is little appreciated, however, is that this behavior is undoubtedly an advantage. While aperture techniques do not make any assumption about source morphology and are hence extremely powerful in the face of resolved structural features in galaxies, their ability to identify cases of sources blended at detection or quantify contamination in photometry of photometrically blended sources is severely limited. Parametric profile-fitting techniques suffer from neither of these drawbacks. So long as intrinsically blended sources are not well-described also by a single profile[10], then the optimized model will not achieve a satisfactory fit. These cases may be confidently identified a posteriori using statistics such as those discussed in Section 2.3.4.

2.4.5 *Comparison to similar methods*

These advantages and limitations hold mainly for purely parametric models. There exists another, related class which uses a cutout of a given source as its models that can be used to photometer other bands by first convolving it with an appropriate kernel to translate its native PSF to that of the band of interest, and then scaling the unit normalized model to match the source in that band. These 'stamps' has a distinct advantage over parametric models in that they can exploit the resolution of the cutout image to capture structural features not describable by typically smooth parametric models. Without shape parameters to constrain, these stamps can be extremely efficient in measuring fluxes as essentially a scaling factor between the PSF-transformed stamp and the source in question. While simpler than purely parametric models, this approach requires a deep high-resolution image which contains every detected source (if not the same image). More so, the PSF must be well-understood to provide a kernel to map the original PSF to that of the lower resolution images, a drawback not shared by parametric models. The stamp must also be resampled to match the pixel scale of the image to be photometered. For example, an HST-derived stamp of a marginally resolved source applied to Spitzer provides no significant advantage over a parameteric model. Worse, the morphology described by the stamp is assumed to be constant across bands, and so there can arise significant effects between the wavelength of the stamp image and that of the image to be photometered. Such stamp-based profile-fitting software include `TFIT` (Laidler et al., 2007), `T-PHOT` (Merlin et al., 2015, 2016), `PyGFIT` (Mancone et al., 2013), `Morfometryka` (Ferrari et al., 2015), `LAMBDAR` (Wright et al., 2016), and GOLFIR (Kokorev et al., in prep.). Each one takes its own approach to the problem of flux estimation in terms of available models, parametrization, algorithm speed, flexibility, and accessibility. As discussed in Section 2.4.3, purely parametric models can overcome the limitations of these stamp-based codes by freely fitting the shape of the model, possibly with some prior constraints (e.g., `ProFit`, Robotham

10 in which case they cannot be identified as blended without higher resolution ancillary data

et al. 2017; SExtractor++, Bertin et al. 2020; Kümmel et al. 2020; GALAPAGOS-2, Häußler et al. 2022).

One of the most similar photometry frameworks to THE FARMER is HSCPipe, in part because they both inherit the profile-fitting approach of SDSS (Stoughton et al., 2002). As discussed in Aihara et al. (2019), HSCPipe provides model-based photometry by fitting both point-source (PSF) and composite galaxy (cModel) profiles to each galaxy individually. Even though both resolved and point-like models are tried, unlike THE FARMER they are tried for each source independent of their neighbors, which for blended sources can lead to inconsistencies (as demonstrated in Figure 2.1). Furthermore, HSCPipe does not choose a best-fit model type for each source and instead provides fluxes measured from each profile assuming independence from neighbors. While this is computationally faster than a decision tree, it is also inefficient to fit unresolved sources with highly parameterized composite models (which risk overfitting). As of version 8 of HSCPipe[11], only likelihood of the CModel fits are reported and so a consistent statistical comparison with the PSF models is not possible, which leaves only a binary extendedness flag to indicate a resolved source. THE FARMER provides not only a best-fit model type for each source, but also suite of statistics from which the reliability of that model can be assessed.

Although limited to low resolution IRAC images, the IRACLEAN software (Hsieh et al., 2012) measures photometry by iteratively subtracting PSFs at detected source centroids until the residual is clean of signal to some user defined level. Although broadly similar to THE FARMER, IRACLEAN does not perform model-fitting in a classical sense as an iterative subtraction of the PSF is equivalent to an model with effectively unlimited parameters. More so, the order in which sources are processed can introduce hysteresis in crowded regions. There is also the danger of overfitting, as IRACLEAN will continue subtracting a scaled PSF stamp until a given segment has no more signal, which in the case of a blend will combine the flux of the two sources into one photometric measurement. THE FARMER's parametric models act as a prior which can, in some cases, ignore the flux of a neighbor which is left in the residual, and report statistics flagging the problem to the user. Further discussions and comparisons with IRACLEAN are presented in Weaver et al. (2022a).

2.4.6 *Computational considerations*

Computation of sources scales exponentially with the number of sources fit simultaneously as well as the number of free parameters, meaning that these techniques require significantly longer runtimes compared to aperture photometry. Fitting N sources in a given optimization with models of an average number of parameters $\langle\theta\rangle$ requires a computational complexity on the

11 https://hsc.mtk.nao.ac.jp/pipedoc/pipedoc_8_e/index.html

order of $O^{N\langle\theta\rangle}$. In the context of modern deep surveys containing millions of sources (many of which are resolved), this would be enormously complex and by current computational power is essentially impossible. However, a high degree of paralleization can be achieved so long as the source density and resolution allow for distinct groups of sources to be identified and fit separately, requiring only $NO^{\langle\theta\rangle}$, which is easily achievable with modern computers. A practical approach is to process each brick independently. Source groups are constructed at runtime and kept in memory only, so they are ideal for being run in parallel, e.g. across central processing units (CPUs) of a given cluster node.

However, computational time still increases exponentially with the number of free parameters. As such, the modelling stage is not only more complicated than forced photometry because of the several trials of the decision tree, but also because shapes are left to vary in some stages. It is for this reason that the decision tree starts with simple models and moves towards complexity, or in other words, computational expense. If the conditions of the decision tree are satisfied for every source, then the models are assigned without moving to the next stage. For example, an isolated point source should only be tried out with a PointSource and SimpleGalaxy model whereupon it should satisfy the PointSource criterion and stop. Each of these model types have three parameters (two for position and one for flux) and so are incredibly quick compared to a FixedCompositeGalaxy with ten parameters.

Unfortunately, computational time increases exponentially with source crowding. The reason is essentially covariance. Many separate sources can be modelled independently and in parallel without a loss of accuracy. However, because deep images of crowded fields are best photometered when groups of nearby sources are simultaneously modelled, the complexity and computational expense is greater than if the same number of sources were fit separately. As such, it is strongly advised that typical source groups contain as few members as possible without breaking across two blended sources. Unfortunately, the situation of source crowding will only become more difficult as surveys grow deeper. While apertures will eventually hit a limit, profile-fitting photometry can forge ahead, albeit with a greater computational cost.

2.5 BENCHMARK AND VALIDATION

In this section we test and validate the performance of The Farmer using a set of simulated deep images with COSMOS-like properties.

2.5.1 *Construction of mock images*

The construction of the mock images used here follows the approach presented in L. Zalesky et al. (2022, in prep.). Images are created to include a number of realistic features. The noise in each filter is matched to the RMS measured on real images used in Weaver et al. (2022a). Galaxy-type sources are included with random positions and orientations using the open-source code `GalSim` (Rowe et al., 2015) via the `RealGalaxy` class, which allows the user to inject images of real sources observed by *HST* in the COSMOS field. Unfortunately, the morphology of these sources is only available at the resolution of *HST* in one filter (F814W). In order to simulate wavelength-dependent profiles, we use parametric model representations of these galaxies (bulge+disk composites), and give red spectral energy distributions to bulge components and blue spectral energy distributions to disk components; this is handled internally within `GalSim` by the `RealGalaxy` class. To ensure a realistic colors for each galaxy, we have cross-matched the *HST* catalog internal to `GalSim` to the COSMOS2020 catalog, and re-scaled the flux in each band that we simulate to that of the matched source.

The shape of the galaxy number counts is fixed by the internal `GalSim` catalog, and all we modify is the normalization, such that resolved galaxies comprise $\sim 2/3$ of all sources at intermediate magnitudes ($20 < m_{\mathrm{i}} < 24.5$). The counts are incomplete beyond $m_{\mathrm{i}} > 25.5$, and so we inject PSF-models with The Tractor, assuming a constant PSF in each band. The fluxes of these point sources are tuned such that together with the galaxy sources, the total sample yields a complete sample in the HSC-*i* band to 28.5 magnitude. Colors of point sources are assigned by randomly selecting sources of similar flux (within $\pm$ 0.1 mag) from the COSMOS2020 catalog and scaling the flux a given filter to match the color. Finally, the number counts are calibrated and scaled according to the number counts of the COSMOS2020 catalog and those in the empirical mock catalog of Girelli et al. (2020), and shown in grey in Figure 2.13.

Each image is simulated at the same scale as the images used in Weaver et al. (2022a) (0.15″/pixel), aside from the mid-IR Spitzer images. For these mid-IR images, we simulate them at the typical native resolution of 0.6″/pixel, and use `SWarp` (Bertin, 2010a) to resample them to the scale of the other images. This extra step is taken in order to introduce the correlated noise present in all Spitzer imaging that has been resampled, which is the common method for handling those data when measuring multi-wavelength photometry.

Although the galaxies in our simulated images are parametric representations, it should be noted that resolved galaxies feature structures such as spiral arms and star-bursting regions that are not captured by these models. As such, the performance of The Farmer for the brightest sources assessed on this simulation is likely overestimated compared to real galaxy images.

2.5.2 *Procedure*

We follow the general procedure outlined in Section 2.3. For simplicity and to ease the interpretation of our tests, we adopt the input PSFs used to produce the simulated images. No backgrounds are subtracted. Sources are detected on a izK_s `CHI-MEAN` image created using SWARP (Bertin, 2010a). Models are assigned according to the same decision tree structure as described in Figure 2.3. We tune the decision tree to produce smoothly increasing number counts, and then tune further by spot checking residuals of individual sources. The modelling stage is run, which assigns models and optimizes their parameters on a group basis.

Models are then forced on the bands of interest: $rizKs$ + $channel$ 1 + 2. Figure 2.8 shows the reconstructed model images and residuals produced by THE FARMER over a region of the simulated i and channel 1 mosaics. The vast majority of sources are well modelled with only a handful of failed fits which are left in the residual map. While the value of visual inspection of residuals cannot be understated, a rigorous statistical analysis can provide direct insight.

2.5.3 *Model and decision tree performance*

Now we use the suite of statistics provided by THE FARMER to asses the performance of the models and decision tree.

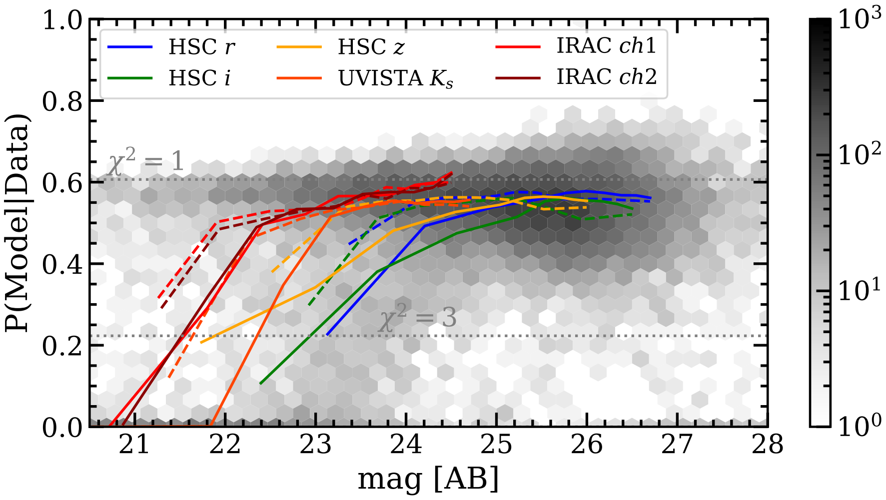

Figure 2.9: Probability of models as a function of apparent magnitude. Results from i are shown in grey histograms, summarized by binned medians of unresolved (solid) and resolved models (dashed). Other band are similarly summarized by binned medians.

As demonstrated by Figure 2.9, the probability of the model given the data (inversely proportional to χ^2_N) is greatest for faint sources across all bands. For images of high spatial resolution (riz), the model performance degrades for both resolved and unresolved models at magnitudes brighter than 24 AB, although with considerable variance. These bright sources are smooth in our simulations, however they are still more complex than the models supplied by THE TRACTOR. More likely, brighter sources tend to be larger and so reside in more complex groups where blending makes accurate photometry more challenging.

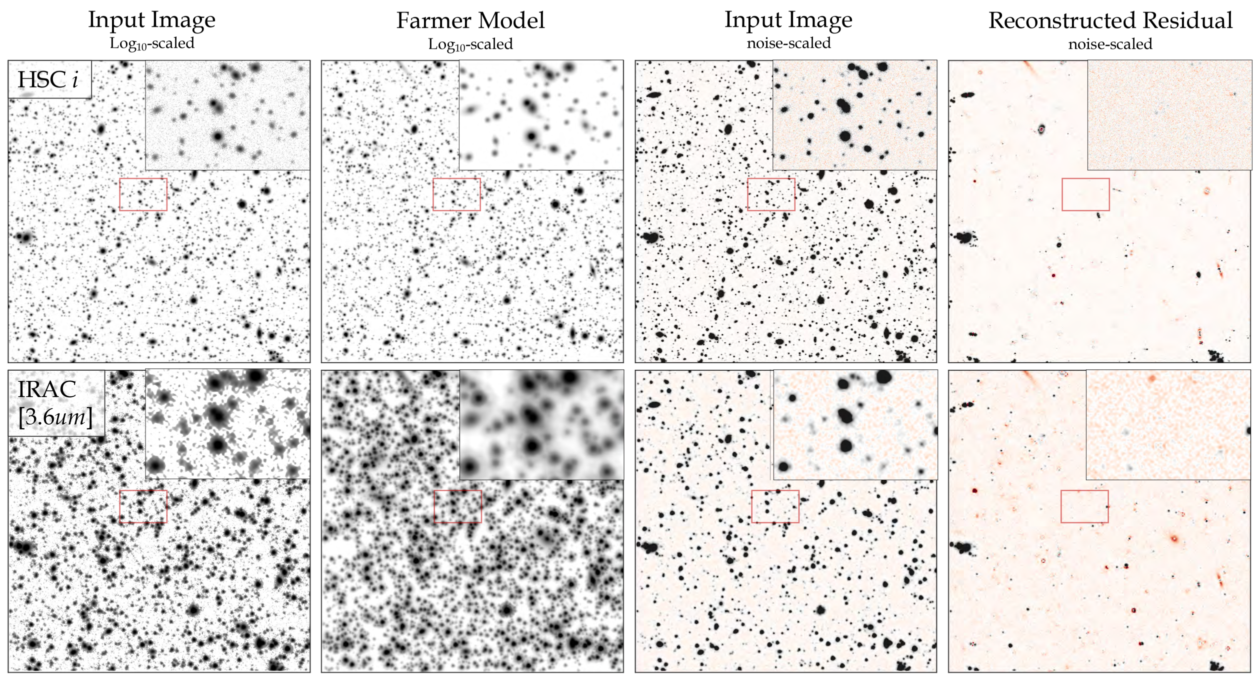

Figure 2.8: Results of forced photometery by The Farmer on simulated fields of depths similar to COSMOS. The same is $i + z$-detected and modelled on i and z jointly, then forced on other bands including channel 1. Models can be compared to input images in the two leftmost panels in $\log_{10}$ scaling where morphology is visible. Insets show a zoom-in of a smaller region. Residuals can be compared to input images in the two rightmost panels in $\pm 3\sigma$ scaling to highlight faint signal above and below the background.

The NIR and IR bands (K_s and IRAC) have slightly better performance at bright magnitudes. This is because their resolution threshold is at a brighter magnitude and so these particular bands contain a higher fraction of bright sources which appear unresolved. Whether or not THE FARMER assigned resolved or unresolved models to these sources, the PSF is large enough that they are effectively unresolved. Photometry is then made easier because there is little dependence on accurate model shapes.

The key insight therefore is that the effectiveness of profile-fitting photometery is not dependent on source magnitude directly, but rather on the size of the source and whether or not is is resolved, with some lesser dependence on the resolution of the bands used to derive the models.

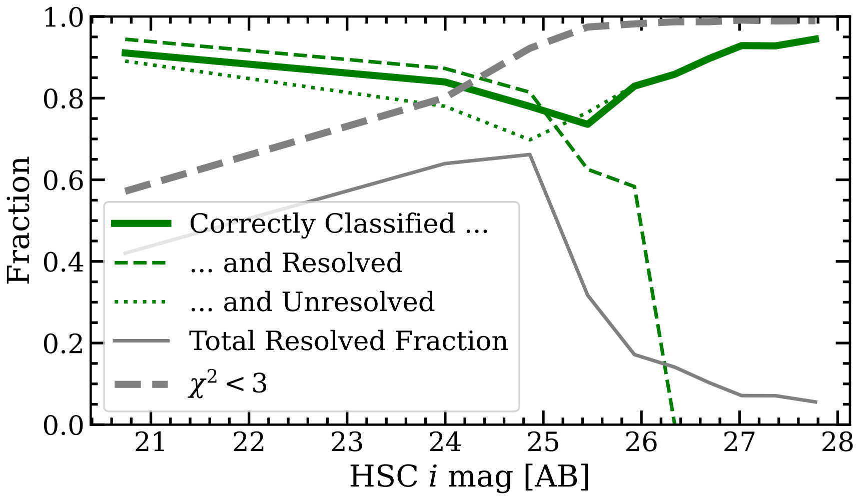

Figure 2.10: Fraction of unresolved (resolved) sources correctly assigned an unresolved (resolved) model by THE FARMER as a function of apparent i-band magnitude is shown by the solid green curve, and broken down into resolved and unresolved subsets by the dashed and dotted curves, respectively. We consider only sources with $\chi^2 < 3$ in the i-band as they are considered reliable, the fraction of which is shown by the grey dashed curve. Our simulated field at $i > 25$ uses mostly point-like sources to reflect real conditions in a COSMOS-like survey; the total resolved fraction is shown by the solid grey curve.

As shown in Figure 2.10, THE FARMER is generally able to correctly assign resolved models to sources which are injected as resolved galaxies, and unresolved models to those which are injected as unresolved point sources. As alluded to earlier, the resolution threshold averaged over the modelling bands (~ 25 for izK_s) is where it is most difficult for THE FARMER to distinguish between resolved and unresolved sources and so ultimately the fine tuning of the decision tree is aimed at improving performance in this regime. Based on our tuning, THE FARMER correctly assigns $> 75\%$ of marginally resolved sources.

While it appears that THE FARMER is not able to correctly assign resolved models to injected resolved galaxies at $i > 25$, this is almost certainly because these sources actually appear unresolved in our simulated images. It should be noted, therefore, that while a given source in the simulated images corresponds to either a resolved galaxy or unresolved point source model, the former may be be effectively unresolved in the image if its size is similar or smaller to that of the PSF. While identifying such cases in the i, z, and K_s bands is therefore of interest as The Farmer should not be expected to assign them a resolved model. These cases cannot be cleanly identified a priori, nor is it possible to identify them a posteriori with full confidence. As a result, the performance of The Farmer may be expected to be better than it appears around the $i \sim 25$ resolution threshold.

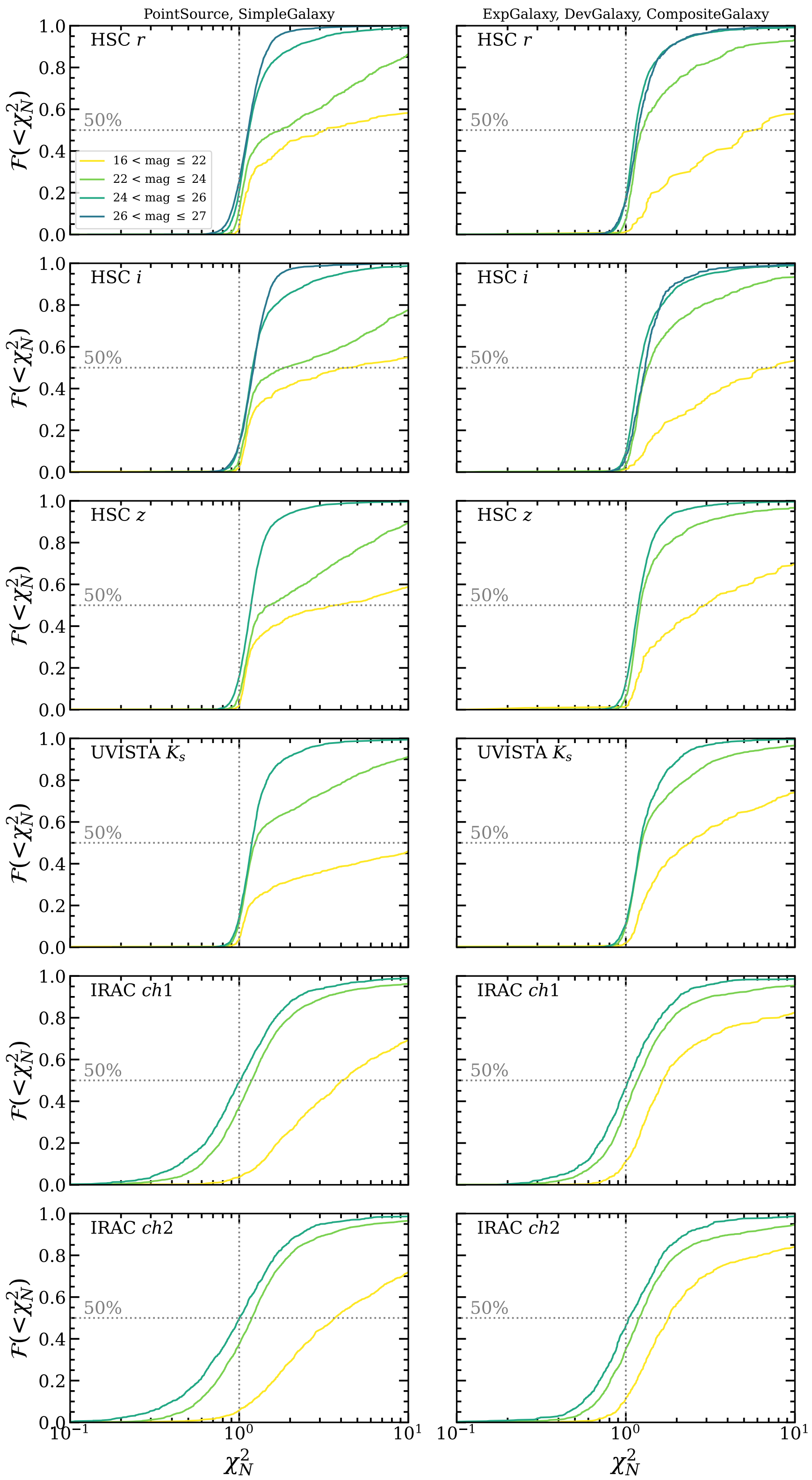

Figure 2.11: Fraction of sources below a certain χ^2 as a function of band and magnitude for unresolved (left) and resolved models (right).

The performance of models optimized in forced photometry is also generally better at faint magnitudes where sources are typically unresolved. Figure 2.11 shows the fraction of sources below a given reduced χ^2_N in four ranges of magnitude for each band separated into unresolved and resolved model types. A sample which is χ^2 distributed reduced by N degrees of freedom should have an expectation value of unity. It cumulative distribution should therefore be approximately evenly divided around $\chi^2_N \approx 1$. It should be noted that χ^2 is a measurement of significance and is therefore dependent on accurate per-pixel errors.

The performance of models for the well-resolved bands (r, i, z, K_s) is better for faint sources irrespective of resolved or unresolved models. Overall these distributions seem slightly shifted towards larger values of χ^2_N. Inspection of the residuals suggest these models are well fit, and so this shift may be due to inaccurate per-pixel errors, or pixel covariance which is not accounted for by χ^2 which assumes independent, Gaussian distributed data. For bright sources, a tail develops at $\chi^2_N > 10$ which also suggests an increased fraction of bad models. By inspection, we confirm that the complexity of the injected galaxies is not always well-captured by the smooth models from The Tractor (as would happen in real images). Source crowding may also play a role for these typically large, bright sources who may have fainter sources near their wings which if not detected may cause a photometric bias.

The two infrared bands (channel 1 and 2) appear to have slightly better performance at faint magnitudes. There does not seem to be a shift, which relative to the bluer bands may be due to greater degree of signal covariance relative to the bluer bands (from the larger PSF) whereby a good fit in one pixel means one can expect to achieve a good fit in the adjacent pixels. A tail does not develop for bright sources, which instead are shifted towards higher χ^2. This systematic behavior suggests that The Farmer has the greatest difficulty modelling the bright IRAC sources in general. This is not a surprise given the high degree of crowding in these IRAC images, which for bright, large sources means not only more complex groups, but also a higher likelihood of flux from neighboring groups falling in. Because this extra light is not expected by the group model, it may lead to a photometric bias.

The Farmer also provides accurate shape measurements for all resolved sources. Figure 2.12 demonstrates the recovery of axis ratio and position angle of the simulated galaxies, with agreement below 1 per cent. There are no obvious biases in either statistic, whether compared to itself, source magnitude, Sérsic index, or local source density. The only notable deviations are expected: circular sources with $b/a \sim 0$ where the axis ratio signal is very weak and small sources where $R_{\rm eff}$ approaches the pixel scale of the image (0.15″/px). The insensitivity to local source density gives The Farmer a considerable advantage over shapes estimated from Source Extractor, and are reliable enough to constrain galaxy alignments in filaments (Laigle et al., 2022, in prep.).

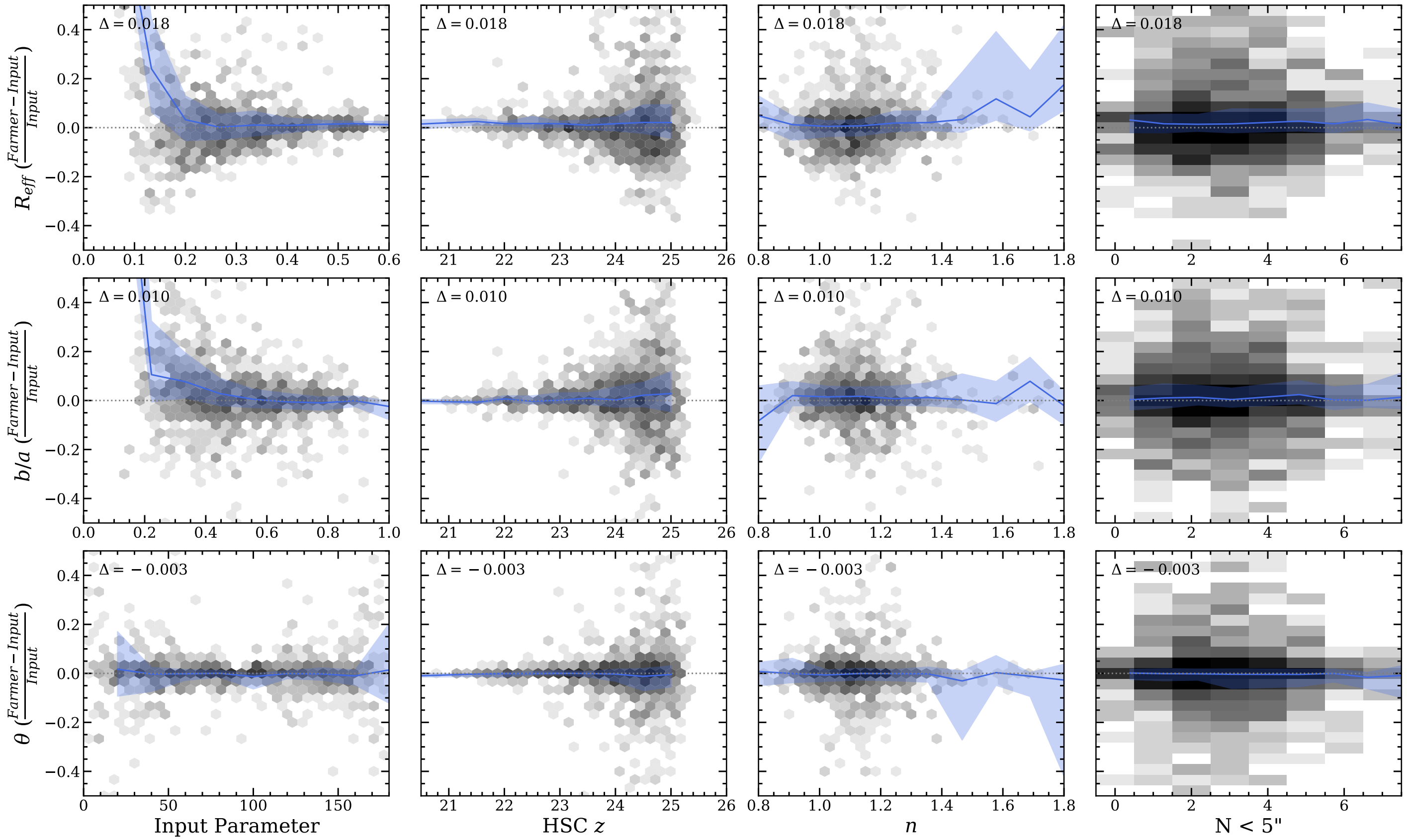

Figure 2.12: Recovery of effective radius (R_{eff}; top row), axis ratio (b/a; middle row) and position angle (θ; bottom row) as a function of input parameter, z-band magnitude, Sérsic index (n), and the local source density (number of sources within $5''$). Greyscale distributions are summarized by binned medians (colored curves) with 68% range indicated by the envelope.

2.5.4 *Counts and photometric accuracy*

Credible survey science ultimately rests on a foundation of complete samples and accurate photometry. We characterize the relevant performance of The Farmer in the following assessments.

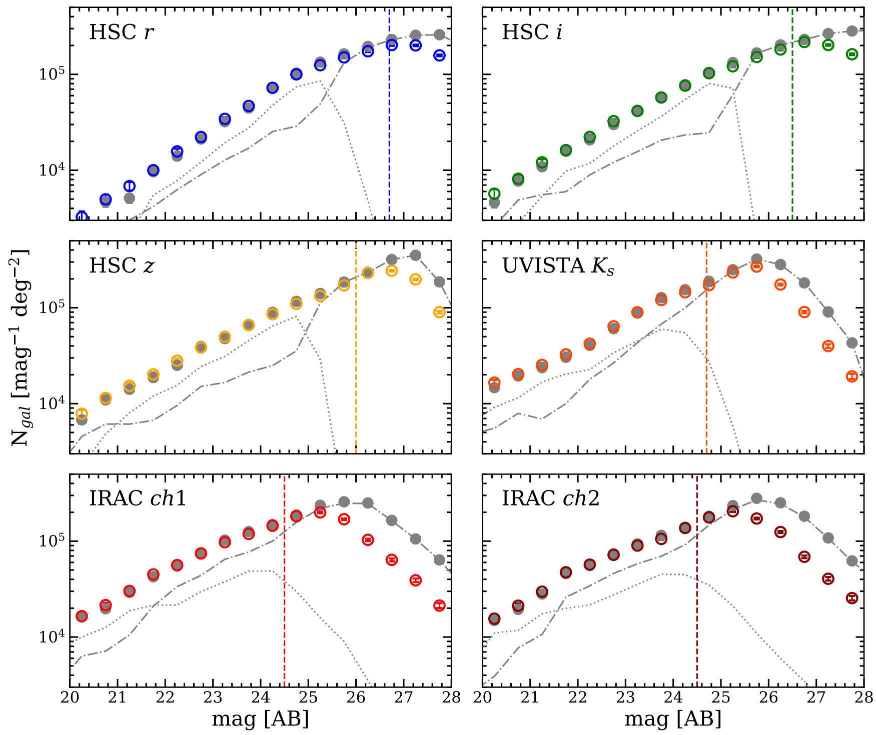

Figure 2.13: Number counts are shown for each band corresponding to the simulation input for all sources (filled grey points), resolved sources (grey dotted curve), and unresolved sources (grey dash-dot curve). This is compared to output from The Farmer (unfilled colored points with Poisson uncertainties) for an *iz*-selected sample. Nominal depths are shown by the vertical colored lines.

Source number counts not only diagnose issues in sample selection and incompleteness, but are also sensitive to photometric accuracy. The number counts of injected sources in our simulated images are shown alongside those recovered by The Farmer in Figure 2.13. The recovery of number counts is generally excellent. They are complete up the limiting magnitude of each band, which is most important for the i, z, K_s bands used in sample selection as incompleteness in other bands may be driven by selection effects. For instance, a small fraction of faint r-band sources is missing from our sample as expected given the simulation includes real galaxy colors and these predominantly blue sources are likely faint in our redder detection image. We can trust that The Farmer's decision tree is performing well given that there are no extended plateaus or sharp rises present anywhere in the number counts, in combination with other available diagnostics (e.g., residuals, χ^2, etc.).

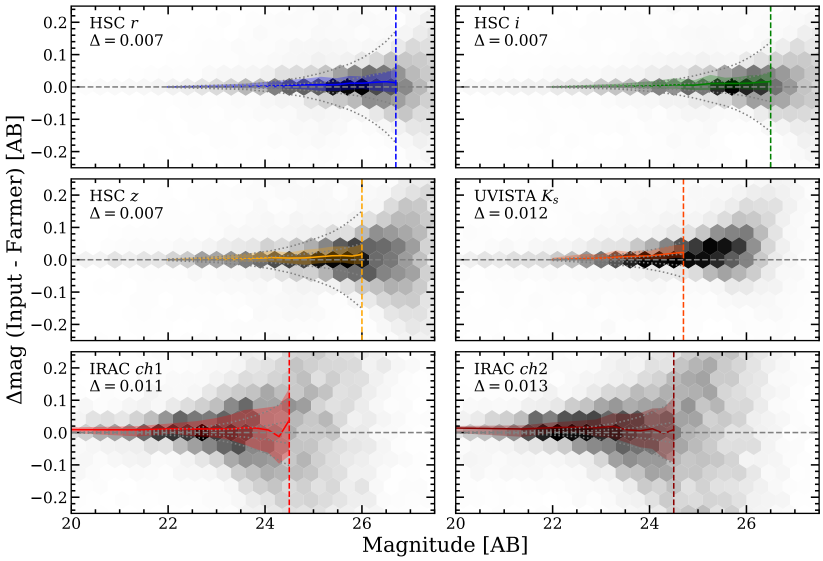

Figure 2.14: Photometry produced by THE FARMER is compared with true fluxes of simulated sources for all bands. Differences in magnitude as a function of input magnitude (grey histograms) are summarized by binned medians (colored curves) with 68% ranges indicated by the colored envelope out to the nominal depth limit of each band (vertical colored lines). ± 1 and 3σ uncertainties on Δmag are computed as medians from the THE FARMER uncertainties.

The most important measurement is ultimately photometry. As shown in Figure 2.14, the photometry measured by THE FARMER is seen on median expectation to be accurate below 0.05 AB in all bands, including IRAC. There are no significant systematic biases, with only a small trending towards overestimated fluxes for faint sources in K_s. The 68% scatter is similar to the typical magnitude uncertainty at a given magnitude for r, i, z, and K_s. For IRAC bands, the scatter is about three times larger than the typical magnitude uncertainty, suggesting that the photometric uncertainties may be underestimated. This may be expected given the high spatial covariance of noise in IRAC images.

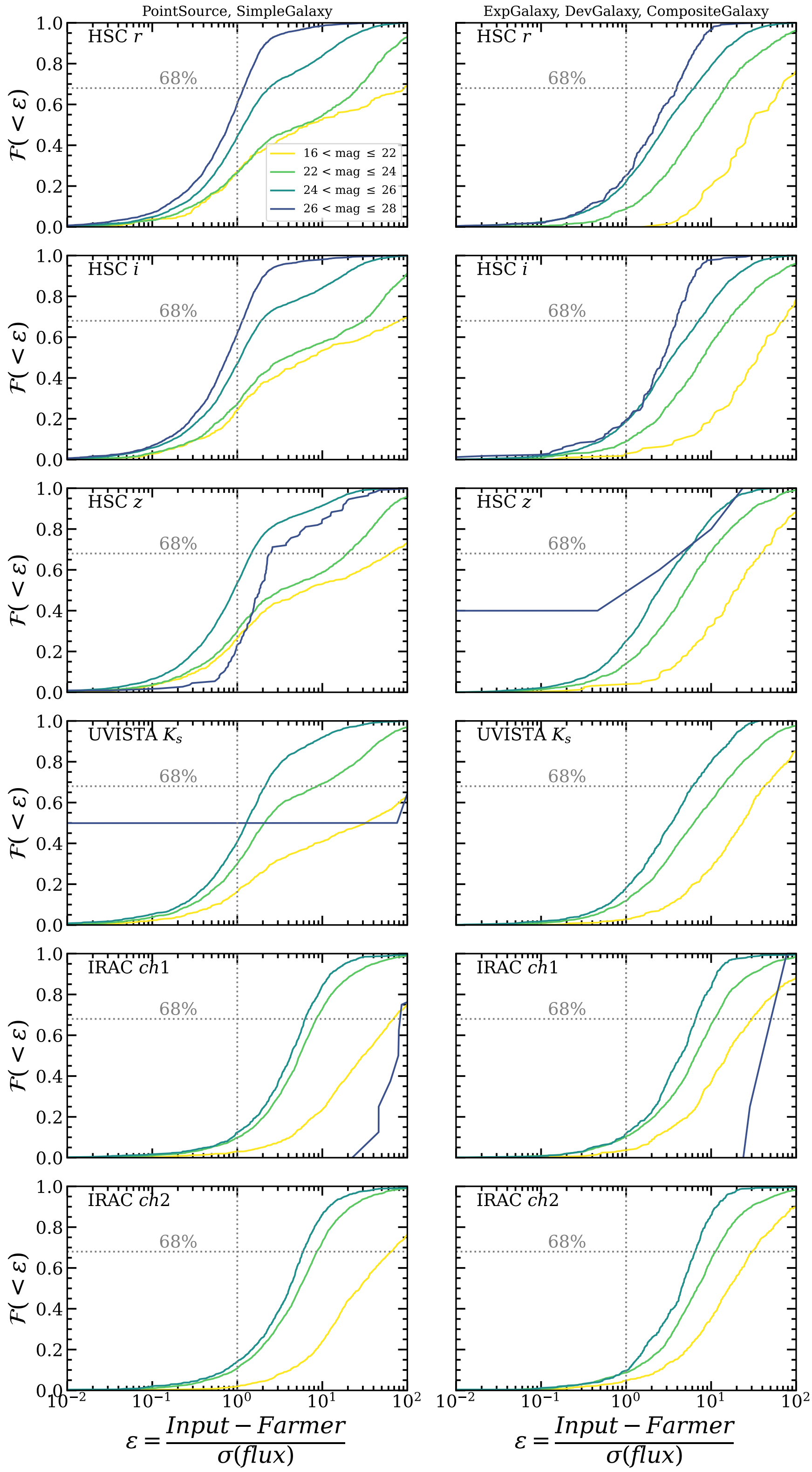

Figure 2.15: Fraction of sources whose relative photometric error ϵ is less than a certain value, broken down by resolved (left panels) and unresolved models (right panels) for each band. On expectation, $\epsilon < 1$ for 68% of sources where a significant departure may indicate under- or over-estimation of photometric uncertainties.

Photometric measurements are more appropriately assessed by directly examining the cumulative distributions of relative error ϵ. These are shown in Figure 2.15 broken down by band and separated into resolved and unresolved model types. On expectation, 68% of sources should be contained by $\epsilon \leq 1$. Given the lack of bias in our photometry, deviations of the ϵ CDFs from this expectation can be directly attributed to innapropriate flux uncertainties resulting from miscalibrated weights and/or spatially covariant noise.

We see a similar picture to the χ^2 CDFs in Figure 2.11 whereby photometry of faint sources measured in the high spatially-resolved bands (r, i, z, and K_s) better follow expectation compared to photometry of bright sources. The distribution of ϵ for bright point sources has a tail as even the smallest biases are expected to yield large ϵ values as the typical flux uncertainties are small. However, the same is not true for the resolved models which are systematically shifted towards larger ϵ with increasing brightness. This may suggest poor modelling performance of the brightest sources, in accord with previous results.

The ϵ CDFs for the IRAC bands are significantly shifted towards higher values in agreement with the results from Figure 2.14. This is further evidence that the weights from IRAC may produce underestimated photometric uncertainties. This is not an immediate confirmation, however, because both χ^2 and ϵ assume independent, gaussian distributed data which may not be the case in instances of significant pixel covariance as in the case of IRAC as it has been up-sampled such that the PSF is correlated across more pixels. Future work aims to evaluate the effect of resampling on the derived photometric uncertainties from The Farmer, which may confirm the expectation that photometry should be performed on images without significant manipulation such as pixel re-sampling.

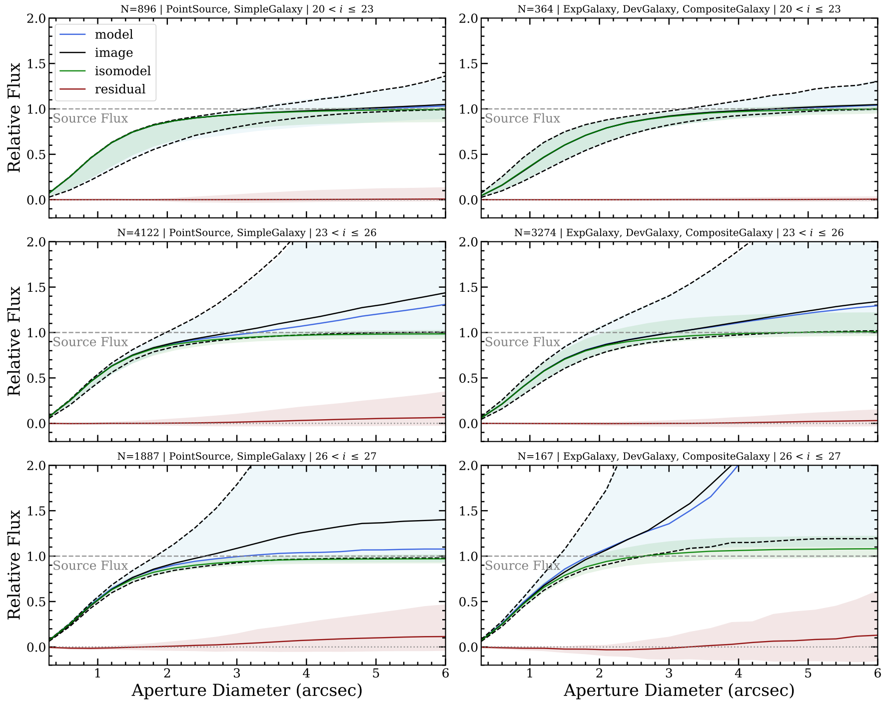

Figure 2.16: Source profiles are estimated by measuring fluxes are in several circular apertures of increasing diameter for unresolved (left panels) and resolved sources (right panels) from the *i*-band mosaic where each flux measurement is normalized to the input flux. Measurements taken from the simulated input image (black), THE FARMER joint model (blue), the farmer model of a given source realized in isolation (green), and residual (red) broken down into three *i*-band magnitude bins and summarized by a median and 68% range.

Another way to investigate typical model accuracy is demonstrated in Figure 2.16. As described in Section 2.3.4, THE FARMER can be configured to extract flux in circular concentric apertures at every source position. We have measured fluxes in several aperture sizes with sub-arcsecond steps for both resolved and unresolved models computed on the group images, models, and residuals. Fluxes are also measured consistently for each source individually, such that they are realized in isolation of other sources (the 'isomodel'). The largest aperture is 6″in diameter which is sufficient to capture flux from all neighboring sources in an *i*-band image used here.

Bright sources are typically large on the sky such that the largest apertures are dominated by the bright source with insignificant contributions from faint neighbors. The apertures measured on the image, model, and isomodel agree well for both resolved and unresolved bright sources, and tend towards agreement with the true input flux at large radii (a value of 1 on the y-axis).

The behavior is different for fainter sources. While their image and model fluxes continue growing even at large apertures, the flux of the isomodel stops growing around 3″as no new flux is captured by the apertures but in agreement with the true input flux. The situation changes again for the faintest sources where on average there is blending at radii smaller than 3″as shown by the divergence of the black image and blue model flux growth curves from that of the isomodel in green that on average agrees with the true input flux. Hence, while there is blending of sources within even 2-3″apertures in *i*-band, the approach used by The Farmer produces fluxes which are not typically affected by blending[12].

2.5.5 *Deblending in IRAC*

The aperture diagnostics are a valuable tool to assess model performance and photometric accuracy, but they also provide a powerful demonstration of the de-blending power of The Farmer. Here we demonstrate this more thoroughly in the context of our simulated IRAC images in Figure 2.17.

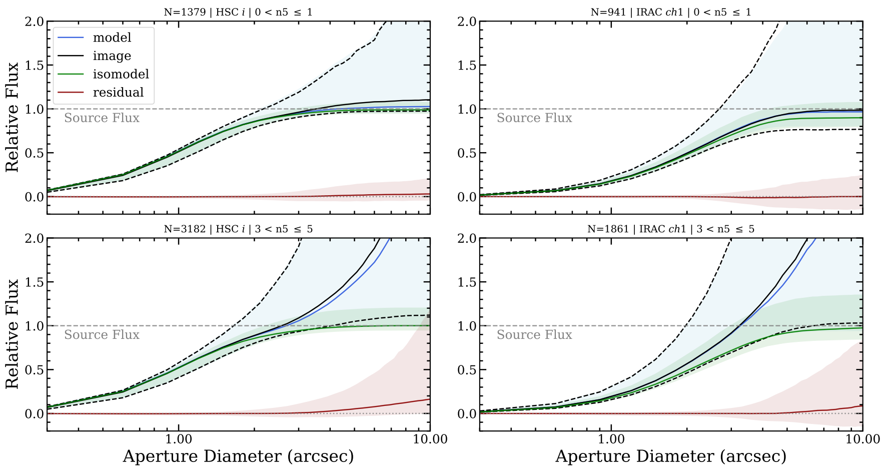

Figure 2.17: Summary of de-blending power of The Farmer. Similar layout to Figure 2.16, but for sources photometered in *i*-band (left) and channel 1 (right) broken down by local density *n*5 defined by the number of sources within 5″.

Similar to Figure 2.16, photometry is measured in apertures forced on source positions computed on the images, models, isomodels, and residuals. As a baseline, growth of flux for sources measured in *i* are in agreement between the images, model, and isomodel, as well as the true total flux for isolated sources. However, for sources in crowded regions the flux measured on the image and model grows exponentially whereas that of the isomodel flattens out around 4″ in agreement with the true flux.

Although IRAC images have very different properties compared to HSC's *i* band, the behavior for isolated sources is similar. The only difference being that larger apertures are required to encompass the total flux of IRAC sources. Aperture photometry measured in crowded regions of IRAC images, however, quickly become contaminated by the flux of neighbors so that no aperture diameter can cleanly measure the total flux of the central source. While the encompassed flux from both the image and model apertures grows

12 This will not be true in cases where blended sources are not separated by detection, see Section 2.4

exponentially, that of the isomodel finds good agreement with the true total flux of the simulated source. What is incredible is that the flux growth curve of the isomodels in green deviates from that of the total group images in black and their joint models in blue already below $2''$, meaning that de-blending is typically significant in our IRAC images even on these small scales. As such, the only tenable way to obtain accurate, high signal-to-noise photometry of IRAC sources is with a profile-fitting approach which, crucially, provides for joint modelling with neighboring sources as employed by The Farmer.

2.6 SUMMARY AND OUTLOOK

While deep galaxy surveys from space-based facilities offer exquisitely resolved images, ground-based surveys are capable of efficiently obtaining similar depths over significantly larger areas where searches for rare populations can be conducted, although at the cost of resolution. Already such survey images contain source densities which demand increasingly smaller aperture photometry to avoid crowding, which results in more uncertain measurements (Laigle et al., 2016; Weaver et al., 2022a). As we have demonstrated, apertures face a self-imposed depth limit, especially in ground-based imaging. Investments in deep, ground-based surveys will continue in the coming decade and so it should be expected that the magnitude of these challenges will only increase. Profile-fitting methods have been a longstanding technique for measuring low-resolution infrared images as they are less susceptible to source crowding. However, their advantages are now needed in the optical and near-infrared regimes. The Farmer attempts to answer this call.

We have explored the methodology of The Tractor whose photometry does not require that images be PSF homogenized, and total fluxes are reported solely based on the scaling of the model profile; avoiding the need for often ill-posed aperture corrections. However, we highlighted several obstacles preventing us from directly applying The Tractor to deep, crowded galaxy fields. These problems were solved by developing The Farmer which leverages an efficient albeit complex decision tree to assign models to sources in an optimal and less pathological way compared to simpler approaches. The decision tree is shown to be more than a useful algorithm, but indeed a required development in overcoming challenges related to blending in deep fields. The Farmer is also a means by which to organize survey data so that one can utilize massively paralellized computing facilities to streamline computational time from potentially years down to only a few weeks. Profile-fitting photometry is, however, more complicated than apertures and comes with its own drawbacks and considerations ranging from selection functions to image resolution, and from deblending capabilities to computational limits.

In a series of validation tests, we examined the ability of The Farmer to photometer sources in realistically simulated images. We found no significant biases in photometry in any band. Further, we illustrated the unique advantage of The Farmer in de-blending sources in low-resolution images like IRAC. Still, bright and potentially resolved sources will continue present a limitation when employing smooth model profiles. On the other extreme, The Farmer has been shown to provide incredibly accurate photometery of the faintest unresolved sources, and in this sense it helps open the door to the distant universe.

Still, challenges in profile-fitting photometry remain and many difficult problems are yet unsolved. While we have demonstrated that The Farmer will provide accurate photometry for the next generation of deep, crowded fields, we must continue to innovate as we move towards deeper and more complex surveys promising even greater discoveries.

ACKNOWLEDGEMENTS

The authors thank Dustin Lang, Charles Steinhardt, Keith Horne, Dan Masters, and Ranga-Ram Chary for helpful discussions.

The HST COSMOS program was supported through NASA grant HST-GO-09822. More information on the COSMOS survey is available at `http://www.astro.caltech.edu/cosmos`.

The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. ST, GB and JW acknowledge support from the European Research Council (ERC) Consolidator Grant funding scheme (project ConTExt, grant No. 648179). OI acknowledges the funding of the French Agence Nationale de la Recherche for the project "SAGACE". HJMcC acknowledges support from the PNCG. ID has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 896225. This work used the CANDIDE computer system at the IAP supported by grants from the PNCG and the DIM-ACAV and maintained by S. Rouberol.

3 COSMOS2020

We must not cease from exploration and the end of all our exploring will be to arrive where we began and to know the place for the first time.

- T. S. Elliot, *Little Gidding*

Adapted from:

COSMOS2020: A panchromatic view of the Universe to $z \sim 10$ from two complementary catalogs

Published in The Astrophysical Journal Supplement Series,
Volume 258, Issue 1, id.11, 30 pp. (2022)

J. R. Weaver, O. B. Kauffmann, O. Ilbert, H. J. McCracken, A. Moneti, S. Toft, G. Brammer, M. Shuntov, I. Davidzon, B.-C. Hsieh, C. Laigle, A. Anastasiou, C. K. Jespersen, J. Vinther, P. Capak, C. M. Casey, C. J. R. McPartland, B. Milvang-Jensen, B. Mobasher, D. B. Sanders, L. Zalesky, S. Arnouts, H. Aussel, J. S. Dunlop, A. Faisst, M. Franx, L. J. Furtak, J. P. U. Fynbo, K. M. L. Gould, T. R. Greve, S. Gwyn, J. S. Kartaltepe, D. Kashino, A. M. Koekemoer, V. Kokorev, O. Le Fèvre, S. Lilly, D. Masters, G. Magdis, V. Mehta, Y. Peng, D. A. Riechers, M. Salvato, M. Sawicki, C. Scarlata, N. Scoville, R. Shirley, J. D. Silverman, A. Sneppen, V. Smolčić, C. Steinhardt, D. Stern, M. Tanaka, Y. Taniguchi, H. I. Teplitz, M. Vaccari, W.-H. Wang, G. Zamorani

3.1 INTRODUCTION

Photometric surveys are an essential component of modern astrophysics. The first surveys of the sky with photographic plates (Bigourdan, 1888) permitted a quantitative understanding of our Universe; longer exposures on increasingly larger telescopes led to the first accurate understanding of the true size and scale of our Universe (Hubble, 1934). Recent breakthroughs have been enabled by the advent of wide-field cameras capable of covering several square degrees at a time such as MegaCam, Boulade et al., 2003, coupled with wide-field spectroscopic instruments capable of collecting large numbers of spectroscopic redshifts like the Visible Multi-Object Spectrograph (VIMOS; Le Fèvre et al., 2003) and the Multi-Object Spectrograph For Infrared Exploration (MOSFIRE; McLean et al., 2012).

The launch of the Hubble Space Telescope (HST) led to the first Hubble Deep Field catalog (HDF; Williams et al., 1996) which, although limited to an area of 7.5 arcmin^2 in four optical bands to ~ 28 AB depth, revealed the morphological complexity of the distant Universe. This first step gave way to an explosion of data from similar surveys (see Madau et al., 2014, and references therein). The installation of the Advanced Camera for Surveys (ACS) on HST led to a dramatic increase in the field-of-view and sensitivity of optical observations from space. This advancement laid the groundwork for the Great Observatories Origins Deep Survey (GOODS; Giavalisco et al., 2004) which captured multi-band ACS observations over two 16×10 arcmin fields, totaling over 40 times more area than the original HDF. These observations provided groundbreaking insights into the nature of high-redshift galaxies, their rest-frame properties, and helped guide the development of methods to select different classes of objects. Although deep ground-based near-infrared imaging achieved notable successes (e.g., FIRESurvey; Labbé et al., 2003), the installation of the near-infrared camera WFC3 on HST in 2009 expanded our ability to probe the distant Universe. This allowed, for the first time, spatially-resolved measurements of rest-frame optical light at early cosmic times to depths unreachable from ground-based facilities, because of the high infrared sky background. The combined power of ACS and WFC3 yielded the deepest 'blank-field' image of the Universe, the Hubble Ultra Deep Field (HUDF; Beckwith et al., 2006; Ellis et al., 2013; Illingworth et al., 2013; Teplitz et al., 2013), observed over the course of a decade in 13 filters, some reaching depths $\sim 29.5 - 30$ AB. Together with ground-based spectroscopy, it was then possible to confirm some of the most distant galaxies which likely contributed to the reionization of the Universe (e.g., Robertson et al., 2013; Ishigaki et al., 2018). However, the transformative power of these forerunner observations was limited by their small area, complicating efforts to detect and characterize populations of rare high-redshift galaxies. To combat the effects of cosmic variance, the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al., 2011; Koekemoer et al., 2011) placed observations over five separate fields, covering ~ 100 times more area than the HUDF with ACS and WFC3/IR in multiple filters to depths $\sim 28-29$ AB, which enabled precise measurements of the physical parameters of galaxies over cosmic time. Despite these significant advantages and the groundbreaking science they allowed, their individual areas proved still too small to fully combat cosmic variance to the extent required to probe large numbers of galaxies at high-redshift.

The Cosmic Evolution Survey (COSMOS; Scoville et al., 2007b) began in 2003 with a 1.7 deg^2 mosaic with ACS over 583 HST orbits, reaching a 5σ depth of 27.2 AB in the F814W band (Koekemoer et al., 2007b; Scoville et al., 2007a). This was the largest single allocation of HST orbits at the time and remains the largest contiguous area mapped with HST to date. Since then, the field has been covered with deep observations by virtually all major astronomical facilities which have consistently invested in extragalactic studies.

While various HST observations have been carried out with other bands in COSMOS, the programs completed to date generally cover no more than a few percent of the field. Ground-based broad- and narrow-band observations with Subaru Suprime-Cam were some of the first to be performed over the entire area in 2006, providing one of the largest imaging data sets available at that time (Capak et al., 2007). Mid-infrared observations of the entire COSMOS field were also taken using the Spitzer Space Telescope (Sanders et al., 2007).

The key to exploiting these multi-wavelength data sets has been 'photometric redshift' estimation (hereafter photo-z), in which template spectral energy distributions (SEDs) are fit to photometry to estimate distances and physical parameters of galaxies (see Salvato et al., 2019, for a review). This has enabled the construction of large statistical samples of galaxies with well-characterized photometric redshifts calibrated to subsets of galaxies with accurate spectroscopic redshifts. COSMOS has been a benchmark testing ground for photo-z measurement techniques, due to its unrivaled multi-wavelength imaging data and thousands of measured spectroscopic redshifts.

Over the years, several COSMOS photometric catalogs have been publicly released (Capak et al., 2007; Ilbert et al., 2009, 2013; Muzzin et al., 2013a; Laigle et al., 2016). Each of these releases followed new availability of progressively deeper data, such as the intermediate band Subaru/Suprime-Cam data (Taniguchi et al., 2015) and the VISTA near-infrared coverage (McCracken et al., 2012; Milvang-Jensen et al., 2013). The most recent release, COSMOS2015 (Laigle et al., 2016), contains half a million galaxies detected in the combined $zYJHK_s$ images from the Subaru and VISTA telescopes. Four ultra-deep stripes in VISTA and Spitzer, although non-uniform, cover a total area of 0.62 deg^2 (e.g., Ashby et al., 2018). The reported photometric redshifts reach a sub-percent precision at $i < 22.5$. This methodology was applied also to the Subaru-XMM Deep Field (Mehta et al., 2018), the only other deep degree-scale field to feature similarly deep near- and mid-infrared coverage.

For more than a decade, the COSMOS field has occupied an outstanding position in the modern landscape of deep surveys, and has been relied upon to address fundamental scientific questions about our Universe. The 2 deg^2 of

COSMOS have been used to trace large-scale structure (Scoville et al., 2013; Laigle et al., 2018), discover groups and clusters (e.g., Capak et al., 2011; Casey et al., 2015; Hung et al., 2016; Cucciati et al., 2018), and link galaxies to their dark matter halos (e.g., Leauthaud et al., 2007; McCracken et al., 2015; Legrand et al., 2019). The COSMOS photo-z distribution is used as reference to establish the true redshift distribution in redshift slices in the Dark Energy Survey (DES; Troxel et al., 2018), a crucial component when estimating cosmological parameters with weak lensing (e.g., Mandelbaum, 2018). COSMOS demonstrated feasibility of combining space-based shape measurements with ground-based photometric redshifts to map the spatial distribution of dark matter (Massey et al., 2007), a method which will be used by the *Euclid* mission (Laureijs et al., 2011). COSMOS is already being used to prepare essential spectroscopic observations for the mission (Masters et al., 2019b) and to study biases in shape analyses. COSMOS photometric data are being used to predict the quality of *Euclid* photo-z (Duprez et al., in prep.), as well as the number of [Oii] and Hα emitters expected for future dark energy surveys (Saito et al., 2020). Hence, the photometric catalogs created in COSMOS continue to play a crucial role in cosmic shear surveys (Albrecht et al., 2006).

The combination of its depth in the visible and near-infrared, and the wide area covered, makes COSMOS ideal for identifying the largest statistical samples of the rarest, brightest, and most massive galaxies, such as ultra massive quiescent galaxies up to $z \sim 4$ (e.g., Schreiber et al., 2018a; Stockmann et al., 2020; Valentino et al., 2020a), as well as extremely luminous $z \sim 5-6$ starbursts (e.g., Riechers et al., 2010, 2014; Pavesi et al., 2018; Casey et al., 2019; Riechers et al., 2020), quasars (e.g., Prescott et al., 2006; Heintz et al., 2016), and UV-bright star-forming galaxies at $6 < z < 10$ (e.g., Caputi et al., 2015; Stefanon et al., 2019b; Bowler et al., 2020a). With rich multi-wavelength coverage at all accessible wavelengths from the X-ray (Civano et al., 2016a) to the radio (Smolčić et al., 2017), an accurate picture of the galaxy stellar mass assembly was established with this data set, including numerous estimates of the galaxy stellar mass function e.g., Ilbert et al., 2013; Muzzin et al., 2013b; Davidzon et al., 2017, star formation rate density (e.g., Gruppioni et al., 2013; Novak et al., 2017), mass and star formation rate relation (Karim et al., 2011; Rodighiero et al., 2011; Ilbert et al., 2015b; Lee et al., 2015; Leslie et al., 2020a), and star formation quenching (e.g., Peng et al., 2010b). A large number of follow-up programs have been conducted, including extensive spectroscopic coverage (e.g., Lilly et al., 2007b; Le Fèvre et al., 2015; van der Wel et al., 2016; Hasinger et al., 2018), integral field spectroscopy (e.g., Förster Schreiber et al., 2009), and ALMA observations (Scoville et al., 2017a; Le Fèvre et al., 2019).

This paper presents 'COSMOS2020', the latest release of the COSMOS catalog. The principal additions comprise new ultra-deep optical data from the Hyper Suprime-Cam (HSC) Subaru Strategic Program (SSP) PDR2 (SSP; Aihara et al., 2019), new Visible Infrared Survey Telescope for Astronomy (VISTA) data from DR4 reaching at least one magnitude deeper in the K_s

band over the full area, and the inclusion of all Spitzer IRAC data ever taken in COSMOS. Additionally, even deeper u^* and new u band imaging from the Canada-France-Hawaii Telescope (CHFT) program CLAUDS (Sawicki et al., 2019) provides uniform, deep coverage over greater area than available in 2015. Legacy data sets (such as the Suprime-Cam imaging) have also been reprocessed. All imaging data is now aligned with Gaia DR1 (Gaia Collaboration et al., 2016) for the optical and near-infrared data and DR2 (Gaia Collaboration et al., 2018) for the U bands and IRAC data (see Moneti et al., submitted). This is reflected in band-to-band astrometric precision, which is comparably better than in Laigle et al. (2016). Taken together, these additions result in a doubling of the number of detected sources and an overall increase in photometric and astrometric homogeneity of the full data set.

Previous COSMOS photometric catalogs were created with `SExtractor` (Bertin et al., 1996), wherein each image is first homogenized to a common 'target' point-spread function (PSF). Fluxes are then extracted within circular apertures (Capak et al., 2007; Ilbert et al., 2009; Laigle et al., 2016). While this approach is widely applied in the literature (e.g., Sawicki et al., 1998; Hildebrandt et al., 2012), other approaches avoid this homogenization process in order to preserve the original PSFs. The most common alternative involves using a model profile to estimate fluxes, with a wide variety of implementations and variations thereof (e.g Mobasher et al., 1996; Fernández-Soto et al., 1999; Labbé et al., 2006; Hsieh et al., 2012; Labbé et al., 2015). Of recent popularity are prior-based techniques (e.g., De Santis et al., 2007; Laidler et al., 2007; Merlin et al., 2016) which use the highest resolution image as a prior, convolve it with the corresponding PSF kernel of the lower resolution images and utilize the normalization of the PSF convolved prior image to estimate the flux in the lower resolution images. Such an approach was instrumental to extract Spitzer/IRAC photometry in the CANDELS catalogs. Recently, The Tractor (Lang et al., 2016a) has been developed to perform profile-fitting photometry. Instead of a prior cut from a high resolution image (e.g., HST), The Tractor derives entirely parametric models from one or more images containing some degree of morphological information. This has two immediate advantages in that The Tractor does not require a high resolution image from HST and can hence be readily and consistently applied to ground-based data sets, nor does it require that all the images are aligned on the same or integer-multiple pixel grid. Because the models are purely parametric, The Tractor can provide shape measurements for resolved sources in addition to fluxes. The Tractor has already been applied to several deep imaging surveys (Nyland et al., 2017; Dey et al., 2019b), the methods of which have greatly influenced this work.

For COSMOS2020, two independent catalogs are created using different techniques. One is created using the same standard method as Laigle et al. (2016) where aperture photometry is performed on PSF-homogenized images, with the exception of IRAC where PSF-fitting with the `IRACLEAN` software (Hsieh et al., 2012) is used. This is the Classic catalog. The other catalog is created with The Farmer (Weaver et al., in prep.), a software package which generates a full multi-wavelength catalog utilizing The Tractor to perform the modeling. In this sense, The Farmer provides broadly reproducible source detection and photometry which The Tractor, requiring a custom driving script, cannot do by itself. Detailed comparisons of both photometric catalogs and the quality of the photo-z derived from each of them are presented. By utilizing these two methods in tandem it is possible to evaluate the reliability of COSMOS2020. This work presents a detailed analysis of the advantages of each method and provide quantitative arguments which could guide photometric extraction choices for future photometric surveys. The most compelling advantage, however, lies not in discriminating between the catalogs but rather in using them constructively to evaluate the significance, accuracy, and precision of scientific results, a feature which has not yet been possible from a single COSMOS catalog release.

The paper is organized as follows. In Section 3.2, the imaging data set and the data reduction are presented. Section 3.3 describes the source extraction and photometry. The photometry from the two photometric catalogs are compared in Section 3.4. Section 3.5 presents the photometric redshift measurements. In Section 3.6, the physical parameters of the sources in the catalog are presented. Section 3.7 presents our summary and conclusions.

The two catalog files contain the position, extracted multi band photometry, matched ancillary photometry, area flags, derived photometric redshifts and physical parameters. Details of the catalog files including column names and descriptions will purposely not be presented in this paper, as at the time of writing the two catalog files have a combined 1, 116 columns. Instead, reliable and up-to-date information corresponding to the particular catalog release version can be found in their accompanying `README` file and separate release documentation currently in preparation. More information can be found in Section 3.7.

The results presented in this paper adopt a standard ΛCDM cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\mathrm{m},0} = 0.3$ and $\Omega_{\Lambda,0} = 0.7$. All magnitudes are expressed in the AB system (Oke, 1974), for which a flux f_ν in μJy ($10^{-29}\,\mathrm{erg\,cm^{-2}s^{-1}Hz^{-1}}$) corresponds to $\mathrm{AB}_\nu = 23.9 - 2.5\,\log_{10}(f_\nu/\mu\mathrm{Jy})$.

3.2 OBSERVATIONS AND DATA REDUCTION

3.2.1 *Overview of included data*

The principal improvements in COSMOS2020 compared to previous catalogs are the significantly deeper optical and near-infrared images from ongoing Subaru-HSC and VISTA-VIRCAM surveys. In addition, this release contains the definitive reprocessing of all Spitzer data ever taken on COSMOS. 'Legacy' or pre-existing data sets present in COSMOS2015 have been reprocessed to take advantage of improved astrometry from Gaia (the only exceptions being external ancillary data such as GALEX). All images are resampled to make final stacks with a $0\farcs15$ pixel scale. These stacks are aligned to the COSMOS tangent point, which has a right ascension and declination (J2000) of (10h00m27.92s $+02°12'03\farcs50$).

Figure 3.1 illustrates the footprint of the observations in the COSMOS field. Complete details of included data are listed in Table 2. Image quality of the optical and near-infrared data, typically reported as the full-width at half-maximum (FWHM) of a Gaussian fit to the light profile, is excellent; with the exception of the Suprime-Cam g^+ band stack, FWHM values are all between $0\farcs6$ and $1\farcs0$. Figure 3.2 shows the filter transmission curves. Figure 3.3 indicates the depths of the photometric data and provides a comparison with the COSMOS2015 depths. The depth computations are explained in Section 3.3.1 and follow largely the methods in Laigle et al. (2016). As in previous releases, in each band the image and the corresponding weight-map is resampled on the same tangent point using `SWarp` (Bertin et al., 2002). These images will be made publicly available through the COSMOS website at the NASA/IPAC Infrared Science Archive[1] (IRSA).

1 https://irsa.ipac.caltech.edu/Missions/cosmos.html

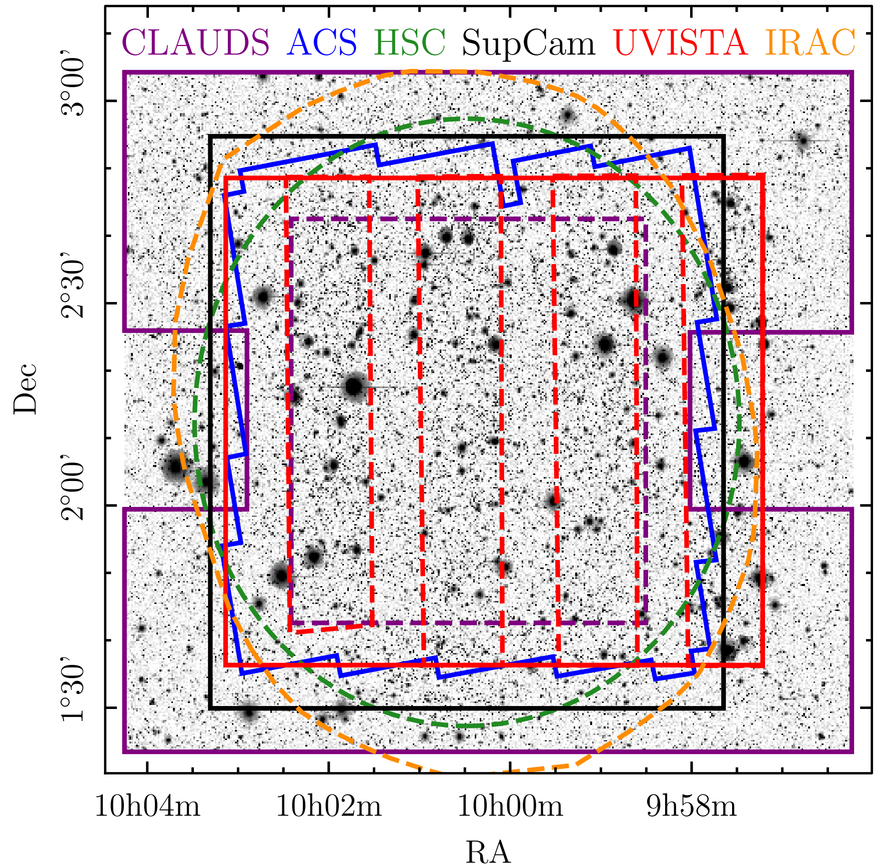

Figure 3.1: Schematic of the COSMOS field. The background image corresponds to the $izYJHK_s$ detection image. The solid lines represent survey limits, and the dashed lines indicate the deepest regions of the images. In the case of UltraVISTA, the dashed lines illustrate the 'ultra-deep' stripes. In the case of CLAUDS, the solid line shows the limit of the u band image and the dashed line shows the deepest region of the u^* band image.

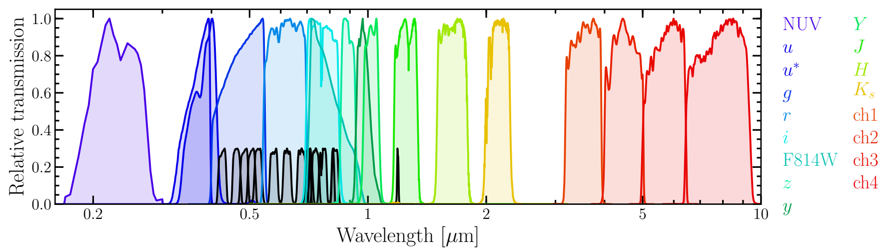

Figure 3.2: Relative transmission curves for the photometric bands used. The effect of atmosphere, telescope, camera optics, filter, and detector are included. The black curves represent medium and narrow-bands. The profiles are normalized to a peak transmission of 1.0 for the broad-bands, and to 0.3 for the medium and narrow-bands.

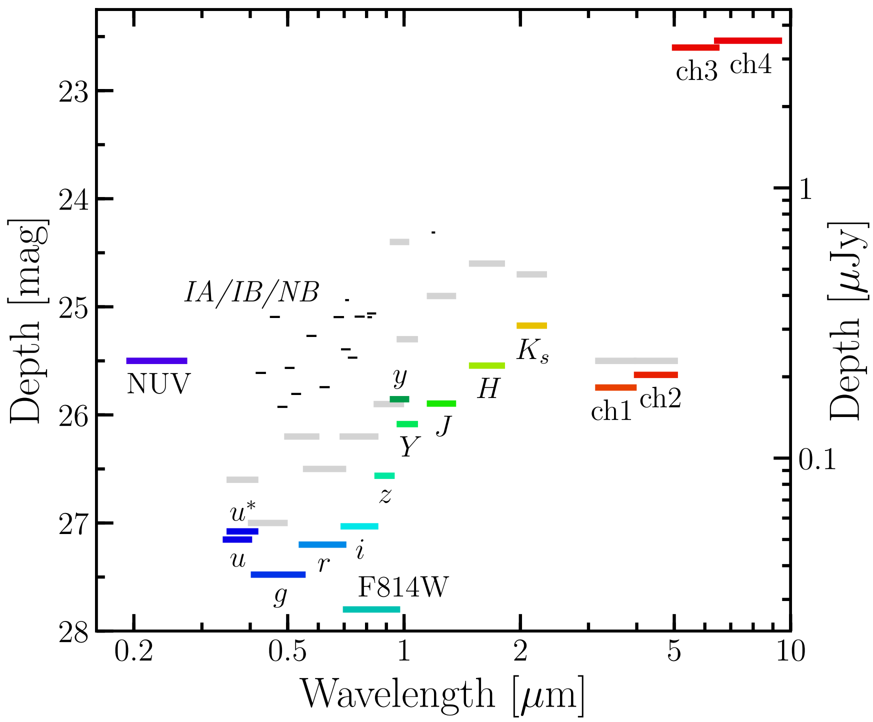

Figure 3.3: Depths at 3σ measured in empty $3''$ diameter apertures in PSF-homogenized images, except for NUV and IRAC images. The NUV depth is from Zamojski et al. (2007) and the F814W 3σ depth is derived from the 5σ value in Koekemoer et al. (2007b). For the Y, J, H, K_s bands, the depths in the ultra-deep regions are indicated. The length of each segment is the FWHM of the filter transmission curve. The thin black segments show the depths of the medium and narrow-bands. The grey segments indicate the depths of the images used in Laigle et al. (2016) for comparison.

Table 2: UV-optical-IR data used in the catalogs

Instrument /Telescope (Survey)	Band	Central[a] λ [Å]	Width[b] [Å]	Depth[c] (2″/3″) ±0.1	Error Fact.[d] (2″/3″) ±0.1
GALEX	FUV	1526	224	26.0[e]	-
	NUV	2307	791	26.0[e]	-
MegaCam /CFHT	u	3709	518	27.8/27.2	1.7/2.0
	u^*	3858	598	27.7/27.1	1.4/1.6
ACS/*HST*	F814W	8333	2511	27.8[f]	-
HSC /Subaru HSC-SSP PDR2	g	4847	1383	28.1/27.5	1.4/1.8
	r	6219	1547	27.8/27.2	1.4/1.7
	i	7699	1471	27.6/27.0	1.5/1.9
	z	8894	766	27.2/26.6	1.4/1.7
	y	9761	786	26.5/25.9	1.4/1.7
Suprime-Cam /Subaru	B	4488	892	27.8/27.1	1.5/1.8
	g^+	4804	1265	26.1/25.6	5.5/5.8
	V	5487	954	26.8/26.2	2.1/2.3
	r^+	6305	1376	27.1/26.5	1.6/1.9
	i^+	7693	1497	26.7/26.1	1.5/1.8
	z^+	8978	847	25.7/25.1	1.5/1.7
	z^{++}	9063	1335	26.3/25.7	2.3/2.6
	$IB427$	4266	207	26.1/25.6	2.0/2.2
	$IB464$	4635	218	25.6/25.1	3.1/3.3
	$IA484$	4851	229	26.5/25.9	1.5/1.7
	$IB505$	5064	231	26.1/25.6	1.6/1.8
	$IA527$	5261	243	26.4/25.8	1.7/2.0
	$IB574$	5766	273	25.8/25.3	2.4/2.5
	$IA624$	6232	300	26.4/25.7	1.4/1.7
	$IA679$	6780	336	25.6/25.1	2.5/2.7
	$IB709$	7073	316	25.9/25.4	2.2/2.3
	$IA738$	7361	324	26.1/25.5	1.5/1.7
	$IA767$	7694	365	25.6/25.1	2.1/2.2
	$IB827$	8243	343	25.6/25.1	2.4/2.6
	$NB711$	7121	72	25.5/24.9	1.2/1.4
	$NB816$	8150	120	25.6/25.1	2.3/2.5
VIRCAM /VISTA UltraVISTA DR4	$Y^{\rm UD}$	10216	923	26.6/26.1	2.8/3.1
	$Y^{\rm Deep}$			25.3/24.8	2.7/2.8
	$J^{\rm UD}$	12525	1718	26.4/25.9	2.7/2.9
	$J^{\rm Deep}$			25.2/24.7	2.5/2.7
	$H^{\rm UD}$	16466	2905	26.1/25.5	2.6/2.9
	$H^{\rm Deep}$			24.9/24.4	2.4/2.6
	$K_s^{\rm UD}$	21557	3074	25.7/25.2	2.4/2.6
	$K_s^{\rm Deep}$			25.3/24.8	2.4/2.6
	$NB118$	11909	112	24.8/24.3	2.8/2.9
IRAC /*Spitzer*	ch1	35686	7443	26.4/25.7	-
	ch2	45067	10119	26.3/25.6	-
	ch3	57788	14082	23.2/22.6	-
	ch4	79958	28796	23.1/22.5	-

[a] Median of the transmission curve.
[b] Full width of the transmission curve at half maximum.
[c] 3σ depth computed on PSF-homogenized images (except for IRAC images) in empty apertures with the given diameter, averaged over the UltraVISTA area.
[d] Multiplicative correction factor for photometric flux uncertainties in the Classic catalog, averaged over the UltraVISTA area (see Section 3.3.1).
[e] 3σ depth derived from the 5σ depth from http://cesam.lam.fr/galex-emphot/.
[f] 3σ depth derived from the 5σ depth in Koekemoer et al. (2007b).

3.2.2 *U band data*

Several programs have observed the COSMOS field in the *U* band using the Canada-France-Hawaii telescope (CFHT) and the MegaCam instrument, the most efficient wide-field *U* band instrument. For COSMOS2020, all archival MegaCam COSMOS *U* data are recombined in addition to new data taken as part of the CFHT Large Area *U* band Deep Survey[2] (CLAUDS), which use a new bluer *u* filter (Sawicki et al., 2019) which lacks the red ~ 5000 Å leakage present in the older and now retired u^* filter. The methodology employed in the reprocessing is similar to that used by CLAUDS. For completeness, u^* corresponds to the *u* band used in Laigle et al. (2016). The depths[3] of the *u* and the u^* images are reported in Table 2. The main motivations in reprocessing these data are to make deeper *U* band images for the field, to make use of the new improved Gaia astrometric reference, and to resample each individual image onto the same COSMOS tangent point.

Starting with the complete data set in both filters, these data were pre-processed by the *Elixir* pipeline (Magnier et al., 2004) at the CFHT before being ingested into the Canadian Astronomy Data Center, where the astrometric and photometric calibrations are recomputed using the image stacking pipeline *MegaPipe* (Gwyn, 2008). Images with sky fluxes above $\log_{10}(\mathrm{ADU/sec}) > -0.1$ were rejected. The images were visually inspected and those with obvious flaws (bad tracking, bad seeing) were rejected. Several images were rejected during the calibration stage, having seeing worse than $1''.4$. In total, there were 649 u^* band images and 500 *u* band images. The median seeing of this final sample is $0''.9$. The two final stacked images were separately resampled onto the COSMOS tangent point and pixel scale and each were combined using a weighted 2.8 sigma clipping. The astrometric calibration used the Gaia DR2 reference catalog (Gaia Collaboration et al., 2018). The final images have an absolute astrometric uncertainty of 20 mas. The *u* band calibration has been improved over earlier versions by carefully mapping the zeropoint variation across the mosaic for each observing run. Without this correction, the zeropoint could vary as much as 0.05 mag across the field. After the correction, the variation is reduced to an estimated 0.005 mag, a 10-fold improvement. This correction does not alter the average zeropoint. While the Sloan Digital Sky Survey (SDSS) is used as the photometric reference, it is not used as in-field standards to avoid propagating any local errors in the SDSS *u* band calibration. Instead, zero-points are computed per night using all available images. Images taken on photometric nights were used to calibrate data taken in non-photometric conditions (see Section 3 of Sawicki et al. 2019 for more details). In summary, both *u* and u^* images have equivalent average depths; however the newer *u* images do not cover the entire COSMOS field but have two gaps at the left and right middle edges of the field (Figure 3.1). However, compared to the older u^* data which is around 0.3 mag deeper in the field center and substantially shallower outside of it, the newer *u* data have uniform depth over the whole survey area.

2 https://www.ap.smu.ca/~sawicki/sawicki/CLAUDS.html

3 The reported u^* band depth is deeper than COSMOS2015 because this work averages over the UltraVISTA layout, compared to the entire field in Laigle et al. (2016).

3.2.3 *Optical data*

Wide-field optical data have played a key role in measuring COSMOS photometric redshifts. The commissioning of Subaru's 1.8 deg^2 Hyper Suprime-Cam (HSC; Miyazaki et al. 2018) instrument has enabled more efficient and much deeper broad-band photometric measurements over the entire COSMOS area. HSC/y data were already included in Laigle et al. (2016). COSMOS2020 uses the second public data release (PDR2) of the HSC Subaru Strategic Program (HSC-SSP) comprising the g, r, i, z, y bands (Aihara et al., 2019).

The public stacks in COSMOS suffer from scattered light from the presence of bright stars in the field and the small dithers used. These are not removed at the image combination stage. Therefore, all the individual calibrated pre-warp CCD images (*calexp* data) from the SSP public server are processed. These images were recombined with `SWarp` using `COMBINE_TYPE` set to `CLIPPED` with a 2.8σ threshold (see Gruen et al. 2014 for details). This removes a large fraction of the scattered light and satellite trails. As for the other data, images are centered on the COSMOS tangent point with a $0\farcs15$ pixel scale. The Gaia DR1 astrometric solution computed by the HSC-SSP team agrees well with the solutions used here in other bands.

Finally, the Subaru Suprime-Cam data used in COSMOS2015 are retained for this work (Taniguchi et al., 2007; Taniguchi et al., 2015), including 7 broad-bands ($B, g^+, V, r^+, i^+, z^+, z^{++}$), 12 medium-bands ($IB427$, $IB464$, $IA484$, $IB505$, $IA527$, $IB574$, $IA624$, $IA679$, $IB709$, $IA738$, $IA767$, $IB827$), and two narrow-bands ($NB711$, $NB816$). However, because the COSMOS2015 stacks had been computed with the old COSMOS astrometric reference, it was necessary to return to the individual images and recompute a new astrometric solution using Gaia DR1 with `Scamp` (Bertin, 2006). The opportunity was taken to perform a tile-level PSF homogenization on the individual images. (see Section 3.3.1).

3.2.4 *Near-infrared data*

The $YJHK_s$ broad-band and $NB118$ narrow-band data from the fourth data release[4] (DR4) of the UltraVISTA survey (McCracken et al., 2012; Moneti et al., 2019) are used. This release includes the images taken from December 2009 to June 2016 with the VIRCAM instrument on the VISTA telescope. Compared to DR2, the images are up to 0.8 mag deeper in the ultra-deep stripes for the J and H bands, and 1 mag in the deep stripes for the K_s band, effectively homogenizing the K_s depth across the full field. The additional $NB118$ narrow-band image only covers the ultra-deep region. Characterization of the $NB118$ filter is in Milvang-Jensen et al. (2013). Only the publicly available stacks are used. These public stacks are aligned to the COSMOS tangent point described previously and have a $0\farcs15$ pixel scale. Gaia DR1 has been used to compute the astrometric solution.

4 http://ultravista.org/release4/dr4_release.pdf

3.2.5 *Mid-infrared data*

The infrared data comprise Spitzer/IRAC channel 1,2,3,4 images from the Cosmic Dawn Survey (Moneti et al., submitted). This consists of all IRAC data taken in the COSMOS field up to the end of the mission in January 2020. This includes the Spitzer Extended Deep Survey (SEDS; Ashby et al. 2013), the Spitzer Large Area Survey with Hyper Suprime-Cam (SPLASH; Steinhardt et al. 2014), the Spitzer-Cosmic Assembly Deep Near-infrared Extragalactic Legacy Survey (S-CANDELS; Ashby et al. 2015), and the Spitzer Matching Survey of the UltraVISTA ultra-deep Stripes survey (SMUVS; Ashby et al. 2018). The resulting images have a $0\farcs6$ pixel scale, and are resampled to the $0\farcs15$ pixel scale of the optical and near-infrared images. The astrometric calibration used the Gaia DR2 reference. This work adopts the processed mosaics with stellar sources removed. A full listing of included programs and details of this processing are given in Moneti et al. (submitted).

3.2.6 *X-ray, ultraviolet, and HST data*

The COSMOS2020 catalog provides basic measurements from ancillary datasets in COSMOS, including data unchanged from various source catalogs. Sources in COSMOS2020 are matched with ancillary photometric catalogs using positional cross-matching within a conservative radius of $0\farcs6$ consistently for all ancillary catalogs, adopting only the most reliable sources, as described below. Measurements of the near-UV (0.23 μm) and far-UV (0.15 μm) are taken from the COSMOS GALEX catalog (Zamojski et al., 2007), and X-ray photometry are taken from the *Chandra* COSMOS Legacy survey (Civano et al., 2016b; Marchesi et al., 2016). With the exception of the GALEX near-UV photometry of Zamojski et al. (2007), these ancillary data are not used in deriving photo-z, or physical parameters. Sources with significant X-ray detections are not used to assess photo-z performance, presented in Section 3.5. HST/ACS morphological measurements are used in identifying stellar contaminants. Summaries of the ancillary photometric datasets can be found in the `README` files accompanying the COSMOS2020 catalogs. Also included are column descriptions and corresponding reference literature where details of this ancillary data including their construction and caveats can be found.

The *HST*/ACS F814W high-resolution photometry from Leauthaud et al. (2007) covering 1.64 deg^2 of the COSMOS field are included for only unblended sources, as well as their morphological parameters. The ACS observations in the F475W and F606W bands cover about 5 % of the field, so these are not included in the catalog.

Unlike Laigle et al. (2016), far-infrared to millimeter photometry from the COSMOS Super-deblended catalog (Jin et al., 2018) are not included as ancillary data in COSMOS2020. This is because the photometry was computed partly using a higher resolution prior catalog from COSMOS2015 and as such the identification of correct matches with COSMOS2020 is uncertain. Future work including *Spitzer*/MIPS (24 μm), *Herschel*/PACS (100, 160 μm) and SPIRE (250, 350, 500 μm), JCMT/SCUBA2 (850 μm), ASTE/AzTEC (1.1 mm), IRAM/MAMBO (1.2 mm) and VLA (1.4, 3 GHz) photometry will be provided in an updated Super-deblended catalog using the COSMOS2020 positions as priors (Jin et al., in prep.).

3.2.7 *Masking*

Photometric extraction of sources can be significantly affected by the spurious flux of nearby bright stars, galaxies, and various other artifacts in the images. Thus, it is of interest to mark these sources. For this purpose, the COSMOS2020 catalogs provide flags for objects in the vicinity of bright stars, and for objects affected by various artifacts.

The bright-star masks from the HSC-SSP PDR2 (Coupon et al., 2018) are used to flag these sources. In particular, masks are taken from the Incremental Data Release 1 revised bright-star masks that uses Gaia DR2 as a reference star catalog, where stars brighter than $G = 18$ mag are masked. About 18 % of sources in the catalog are found within the masked regions in the vicinity of bright stars. Furthermore, artifacts in the Suprime-Cam images are masked using the same masks as in COSMOS2015.

Masks indicating the area covered by the observations for the UltraVISTA deep and ultra-deep regions are provided as shown in Figure 3.1. Also included is a mask corresponding to coverage by Suprime-Cam. A conservative combined mask is prepared for sources within 1.27 deg^2 which have coverage from HSC, UltraVISTA, and IRAC but which are not close to bright stars or large artifacts.

The most up-to-date descriptions of these masks and their respective flags can be found in the `README` files which accompany the catalogs.

3.2.8 *Astrometry*

The astrometry in the previous COSMOS catalogs was based on radio interferometric data. However, with the advent of Gaia, a new, highly precise astrometric reference is available. For COSMOS2020, Astrometric solutions were computed using Gaia data for every data set described here. In the case where data presented in previous papers is included, the astrometric solutions were recomputed and data resampled. The UltraVISTA, HSC, and the reprocessed Suprime-Cam images were calibrated using the Gaia DR1 astrometric reference (Gaia Collaboration et al., 2016). Figure 3.4 shows the difference in position between sources in the catalog with HSC i band total magnitudes between 14 and 19 magnitudes and sources in Gaia DR2. The agreement with the reference catalog is excellent, with a standard deviation in both axes of ~ 10 mas and an offset of ~ 1 mas. This is much better than any previous COSMOS catalog; for example, the size of the residuals shown in Figure 9 of Laigle et al. (2016) are ~ 100 mas. Furthermore, there are no systematic trends of these offsets in either Right Ascension or Declination over the entire field, unlike previous catalogs. Consequently, this improved astrometric precision enables photometric measurements in smaller apertures for faint, unresolved sources.

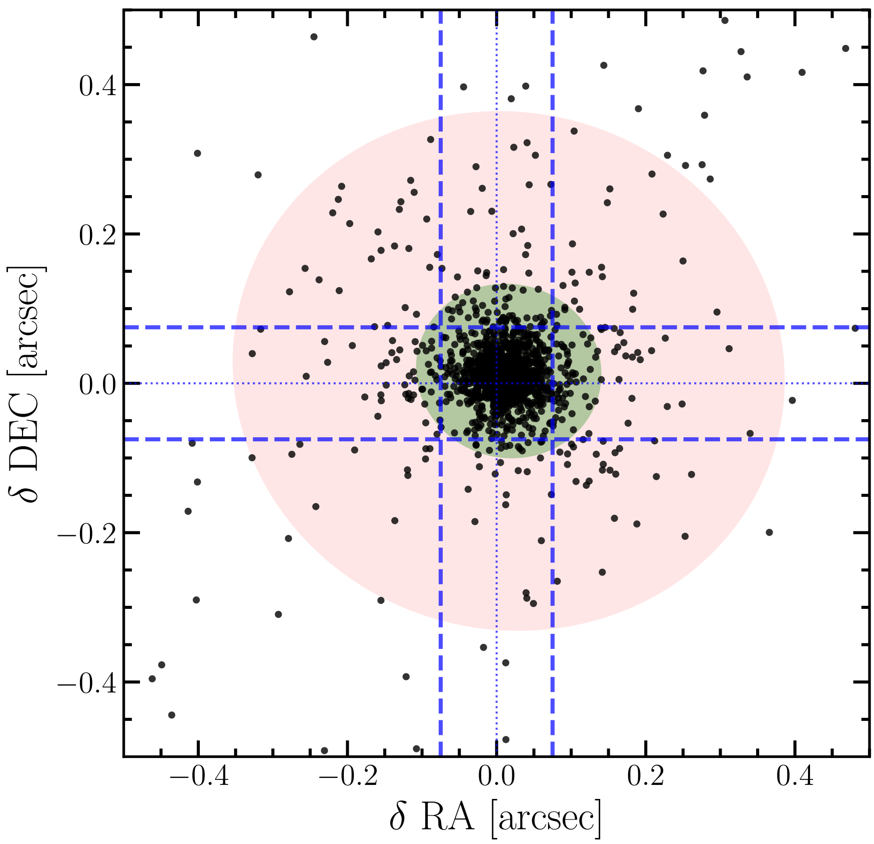

Figure 3.4: Coordinate offset between sources in the Gaia DR1 catalog and sources extracted in the combined detection image as measured in the aperture-based Classic catalog (see Section 3.3.1) The spacing between the dashed lines corresponds to the linear dimension of a pixel in the resampled images. Light and dark shaded regions are ellipses containing 68 % and 99 % of all sources respectively. For clarity, only one in ten sources are plotted.

3.2.9 *Spectroscopic data*

The spectroscopic data are collected from several spectroscopic surveys, conducted with different target selection criteria and instruments. In this paper, the spectroscopically-confirmed redshifts (spec-z hereafter) are used to evaluate the accuracy of the photo-z. Therefore, this work only includes spec-z with the highest confidence level. If the observation of one object is duplicated, only the spec-z associated to the highest confidence level is used.

The spectroscopic surveys presented below share a common system to define the confidence level in the redshift measurement (Lilly et al., 2007b; Le Fèvre et al., 2015; Hasinger et al., 2018; Kashino et al., 2019; Masters et al., 2019b; Rosani et al., 2019). They follow a flagging system described in section 6 of Le Fèvre et al. (2005). Each spectrum is inspected visually by two team members, who attribute a flag to the spec-z, depending on the robustness of the measurement. A flag 3 or 4 is associated to the spec-z if several prominent spectral features (e.g. emission and absorption lines, continuum break) support the same spec-z. While such flagging system is subjective, a posteriori analysis based on duplicated spectroscopic observations indicate that the confidence level of flag 3 and 4 spec-z is above 95%.

Two large programs were conducted at ESO-VLT with the VIMOS instrument (Le Fèvre et al., 2003) to cover the COSMOS field. The zCOSMOS survey (Lilly et al., 2007b) gathered 600 h of observation and is split into a bright and a faint component. The zCOSMOS-bright surveys targeted 20 000 galaxies selected at $i^* \leq 22.5$, which by construction is highly representative of bright sources. The zCOSMOS-faint survey (Kashino et al., in prep) targeted star-forming galaxies selected with $B_J < 25$ and falling within the redshift range $1.5 \lesssim z \lesssim 3$. The VIMOS Ultra Deep Survey (VUDS; Le Fèvre et al., 2015) includes a randomly selected sample of galaxies at $i < 25$, as well as a pre-selected component at $2 < z < 6$. Included are 8 280, 739 and 944 galaxies from the zCOSMOS-bright, zCOSMOS-faint and VUDS surveys, respectively.

Data from the Complete Calibration of the Color-Redshift Relation Survey (C3R2; Masters et al., 2019b) are also used. The galaxies were selected to fill the color space using the self-organising map algorithm (Kohonen, 1982). Depending on the expected redshift range, various instruments from the Keck telescopes were used, specifically LRIS, DEIMOS, and MOSFIRE. While this sample of 2 056 galaxies is representative in colors, it is not designed to be representative in brightness.

A large sample of 4 353 galaxies taken at Keck with DEIMOS, with various selections over a large range of wavelengths from the X-ray to the far-infrared and radio (Hasinger et al., 2018) are used. Such diversity of selection is crucial to estimate the quality of the photo-z for specific populations known to provide less robust results (e.g., Casey et al., 2012).

The FMOS near-infrared spectrograph at Subaru enables tests of the photo-z in the redshift range $1.5 < z < 3$ sometimes referred to as the "redshift desert" (e.g Le Fèvre et al., 2013). The sample from Kashino et al. (2019) contains 832 bright star-forming galaxies at $z \sim 1.6$ with stellar masses $\log_{10}(M_{\rm sim}/M_\odot) > 9.5$ following the star-forming main sequence.

Also adopted are 447 sources observed with MUSE at ESO/VLT (Rosani et al., 2020). The sample includes faint star-forming galaxies at $z < 1.5$ and Lyman alpha emitters at $z > 3$, and can be used to test the photo-z in a magnitude regime as faint as $i > 26$.

Finally, other smaller size samples are added including Darvish et al. (in prep.) and Chu et al. (in prep.) with MOSFIRE, passive galaxies at $z > 1.5$ (Onodera et al., 2012), and star-forming galaxies at $0.8 < z < 1.6$ from Comparat et al. (2015). The full compilation of spec-z in the COSMOS field, including the contributing survey programs, is described in Salvato et al. (in prep.).

3.3 SOURCE DETECTION AND PHOTOMETRY

3.3.1 *The* Classic *catalog*

Source detection

The "chi-squared" $izYJHK_s$ detection image (Szalay et al., 1999b) is created with `SWarp` from the combined original images without PSF-homogenization using the `CHI_MEAN` option. The inclusion of the HSC/i, z band data increases the catalog completeness for bluer objects. In particular, the HSC/i band image is very deep and has excellent seeing of around 0.″6. The previous 2015 catalog (Laigle et al., 2016) did not include i band data in their detection image. The inclusion of the deep i band in this detection strategy is the main reason for the higher number of sources detected in the COSMOS2020 catalog compared to COSMOS2015, likely driven by small, blue galaxies at low and intermediate redshift. The increased depth of the near-infrared bands also contributes to the greater number of detected sources.

For the Classic catalog, the detection is performed using `SExtractor` (Bertin et al., 1996) with parameters listed in Table 5. The main difference with respect to COSMOS2015 is `DETECT_MINAREA` set to 5 pix instead of 10 pix, which is made possible thanks to the lower number of spurious sources in the detection image compared to COSMOS2015, owing to the addition of the i band and deeper imaging in general. The number of detected sources reaches 1 720 700 over the whole field, with 790 579 sources in the UltraVISTA region outside the HSC bright star masks.

Point spread function homogenization

The procedure to homogenize the PSF in the optical/near-infrared images is similar to the one presented in Laigle et al. (2016). In the first step, `SExtractor` is used to build a catalog of bright sources. Stars are identified by cross-matching coordinates with point-like sources from the *HST*/ACS catalog in COSMOS (Koekemoer et al., 2007b; Leauthaud et al., 2007). Saturated stars are removed in the masks (see Section 3.2.7). Bright, but not saturated stars, are identified by their position in the half-light radius versus apparent magnitude diagram. The PSF of each image is modeled using `PSFEx` (Bertin, 2013) adopting the polar shapelet basis functions (Massey et al., 2005). The same code also provides a convolution kernel that can modify the image's response into a "target PSF", which is modeled as a Moffat profile (Moffat, 1969) with parameters $\theta = 0\farcs8$ and $\beta = 2.5$ (the former being the FWHM while β is the atmospheric scattering coefficient). These two parameters are identical to Laigle et al. (2016), whereas the `PSF_SAMPLING` parameter is now set to `1` in order to fix the kernel pixel scale. The core of the homogenization process consists in convolving the entire images with these kernels, so that all of them are affected by the same Moffat-shaped PSF.

Figure 3.5 illustrates the precision of the PSF homogenization as a function of distance from the center of the source. The integral of the best-fitting PSF within different apertures is plotted for every band, before and after the homogenization; all these functions are normalized by the integral of the target Moffat profile within the same apertures. The ratios of the integrals differ from 1 by less than 5 % for all apertures with the exception of Suprime-Cam/g^+, which has a particularly broad initial PSF. In this case, PSF homogenization kernels can still be consistently computed even when the input PSF is wider than the target PSF, and will give a fraction of the weight to the wings (as opposed to the central region) of the PSF. Although the difference between the Suprime-Cam/g^+ PSF and the target PSF is below 10 % in all apertures, it is poor enough that the band is excluded from SED fitting.

In principle, spatial variability of the PSF should be taken into account. For CLAUDS, HSC and UltraVISTA bands this effect is negligible. However, for Suprime-Cam bands the resulting impact of the PSF variability on aperture photometry can be as high as 0.1 mag (as discussed in Laigle et al. 2016). As an example, Figure 3.6 presents the variation of the PSF across the sky for the Suprime-Cam/*IB*464 band, which has the greatest spatial variability before homogenization among the considered bands.

In this work, the spatially dependent PSF homogenization of Suprime-Cam bands is performed starting from individual exposures, as they cover different patches of the field. First, the single exposure files (SEFs) at the original pixel scale of $0\farcs2$ are resampled to the target tangent point with the pixel scale of $0\farcs15$, to remove astrometric distortions. Then, the bright object extraction, PSF modeling, and kernel computation are done in the same way as for the other images. Stars are identified in the half-light radius versus apparent magnitude diagram, automatically adjusting the radius threshold using sigma clipping. The PSF-homogenized SEFs are finally coadded to build the final stacks. Frames with high sky noise ($> 3.5\times$ the median noise) are rejected, representing 1, 5, 28, 16, and 4 images in the B, g^+, z^+, z^{++}, and $NB816$ bands, respectively, out of a total of 2219 images. In these high noise images, only a few objects are detected making it difficult to compute an astrometric solution.

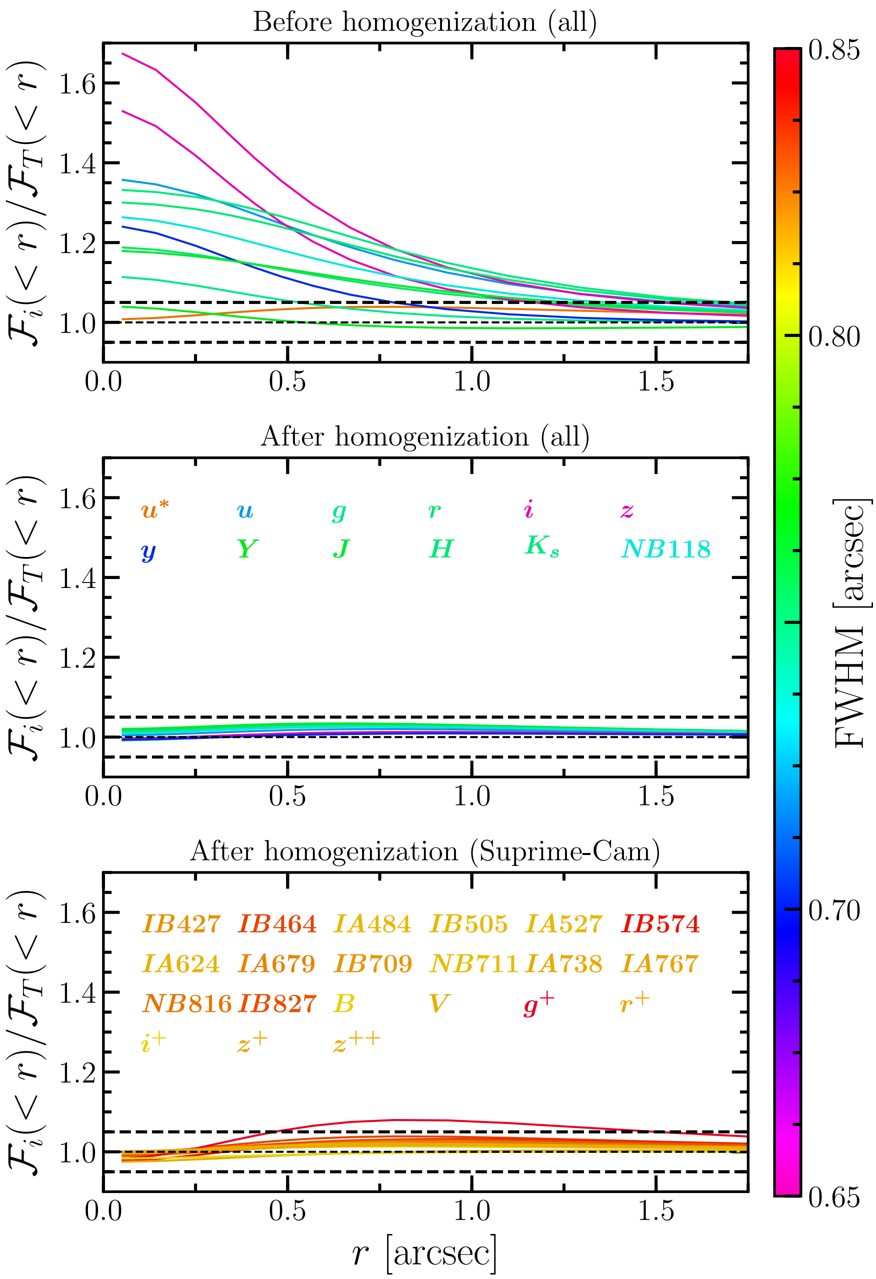

Figure 3.5: Best-fitting Moffat profile PSF integrated in circular apertures, $\mathcal{F}_i$, normalized to the target PSF $\mathcal{F}_T$, as a function of the aperture radius for all bands. *Top:* Before PSF-homogenization, for all bands except Suprime-Cam. *Middle:* After PSF-homogenization, for all bands except Suprime-Cam. *Bottom:* After PSF-homogenization, for Suprime-Cam bands. The horizontal dashed lines indicate $\pm 5\,\%$ relative offset. The color map reflects the PSF FWHM before homogenization for all bands and after homogenization for the Suprime-Cam bands.

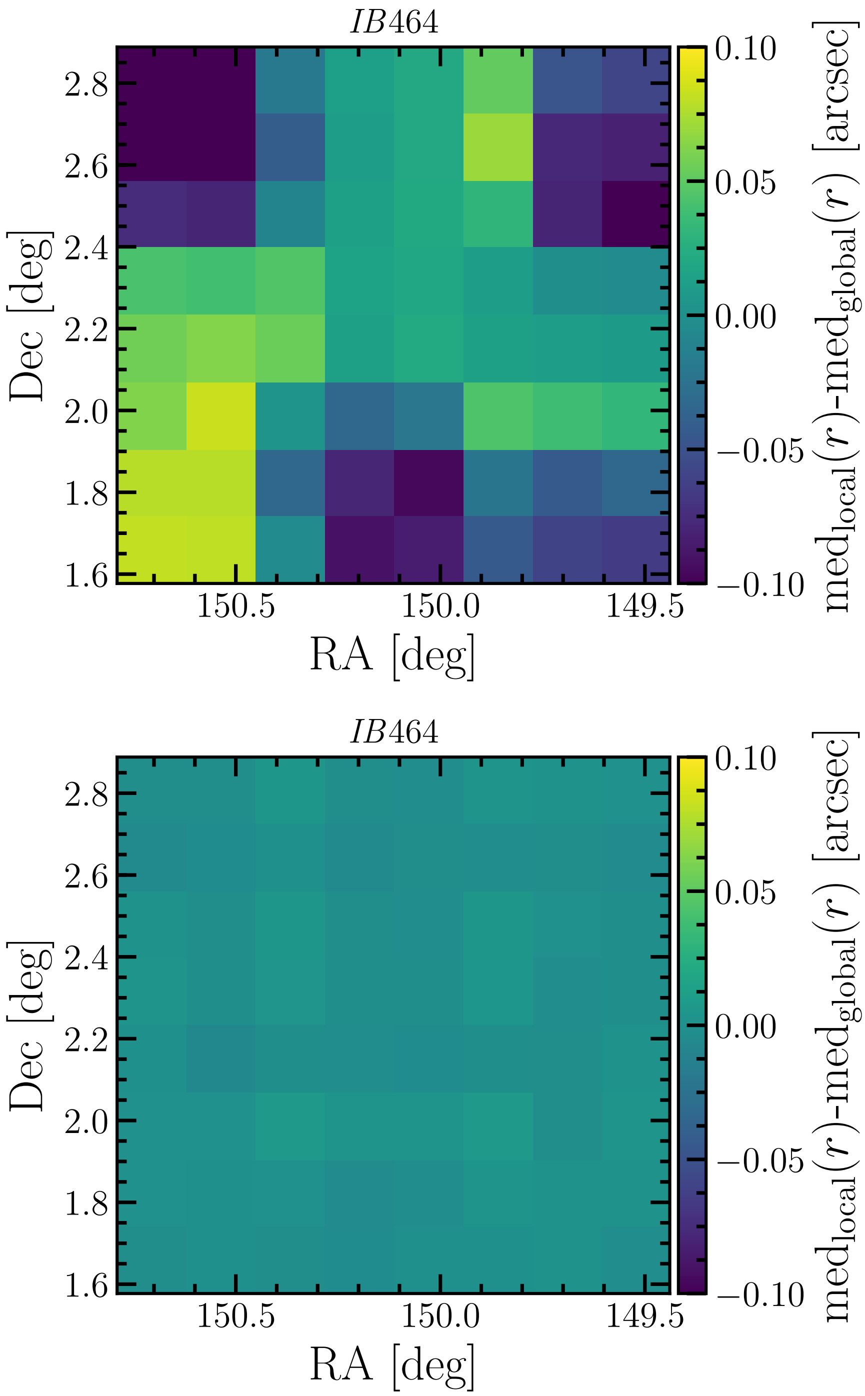

Figure 3.6: Distribution of the difference between the local and the global median half-light radius for the selected stars in the *IB*464 band, as a function of position, before (*top*) and after (*bottom*) PSF-homogenization.

Aperture photometry

Optical and near-infrared fluxes measured in 2″ and 3″ diameter apertures are extracted using `SExtractor` in "dual-image" mode from PSF-homogenized images, using the `CHI_MEAN` as the detection image. Fixed apertures ensure that the same structures are sampled in different bands for each source, which is necessary for reliable measurement of colors and photometric redshifts.

The photometric errors computed with `SExtractor` are underestimated in the case of correlated noise in the image (e.g., Leauthaud et al., 2007). The aperture flux errors and magnitude errors are therefore re-scaled with band-dependent correction factors applied to all sources (Bielby et al., 2012); see Mehta et al. (2018) for a detailed description. In the PSF-homogenized images, the flux is measured in empty apertures (using the segmentation map estimated in each image) randomly placed over the field. The depths are computed from the standard deviation (3σ clipped) of the fluxes in empty apertures inside the UltraVISTA area. The correction factors are then the ratio between the standard deviations of the fluxes measured in empty apertures and the median flux errors in the source catalog, as in Laigle et al. (2016). This is performed separately for $2''$ and $3''$ diameter apertures, and in the case of UltraVISTA photometry, the deep and ultra-deep regions are treated separately. 3σ depth estimates for each band computed over the central UltraVISTA area are listed in Table 2 and illustrated in Figure 3.3. Also included in Table 2 are the photometric uncertainty correction factors used in the CLASSIC catalog. The flux and the magnitude errors are already corrected in the CLASSIC catalog, as it was done for the COSMOS2015 catalog. The 3σ depth of the IRAC bands are computed using the same approach, after tuning the `SExtractor` configuration to the IRAC images.

Aperture photometry may underestimate the total flux of the sources. Optical and near-infrared aperture fluxes (and flux uncertainties) are converted to total fluxes using a source-dependent correction equivalent to the one adopted by Laigle et al. (2016). The correction for each object is computed from the pseudo-total flux $f_{\rm AUTO}$, provided by `SExtractor` and defined as the flux contained within the band-independent Kron radius (Kron, 1980b) as set by `PHOT_AUTOPARAMS` (see Table 5), and the aperture flux $f_{\rm APER}$, also provided by SExtractor. The ratio of these two measurements are then averaged over the HSC/g,r,i,z,y and UltraVISTA/Y,J,H,K_s broad-bands and weighted by the inverted quadratic sum of the pseudo-total and the aperture signal-to-noise:

$$o = \frac{1}{\sum_i w_i} \sum_i \left(\frac{f_{\rm AUTO}}{f_{\rm APER}} \right)_i w_i, \tag{3.1}$$

where the weights are defined as

$$w_i = \frac{1}{\left(\frac{\sigma_{\rm AUTO}}{f_{\rm AUTO}} \right)_i^2 + \left(\frac{\sigma_{\rm APER}}{f_{\rm APER}} \right)_i^2}, \tag{3.2}$$

with $\sigma_{\rm AUTO}$ the $f_{\rm AUTO}$ uncertainties, and $\sigma_{\rm APER}$ the $f_{\rm APER}$ uncertainties (corrected for correlated noise). The sum only includes the filters in which both $f_{\rm AUTO}$ and $f_{\rm APER}$ are positive and unsaturated. As a result, the optical and near-infrared colors remain unaffected. Since photometry from GALEX and

IRAC are measured in total fluxes, this step is required in order to obtain meaningful colors using these bands. Offsets are available (in magnitude units) in the Classic catalog for both 2″ and 3″ diameter apertures.

IRAC photometry

Photometry is performed on the *Spitzer*/IRAC channels 1 and 2 images using the IRACLEAN software (Hsieh et al., 2012). The infrared images of IRAC have a larger PSF (with FWHM between 1.″6 and 2.″0) compared to the optical data and are significantly affected by source confusion which prevents reliable photometric extraction. To tackle this issue, IRACLEAN uses a high-resolution image (and its segmentation map) as a prior to identify the centroid and the boundaries of the source, and iteratively subtract a fraction of its flux ('cleaning') until it reaches some convergence criteria specified by the user. IRACLEAN works in the approximation that an IRAC source can be modeled as a scaled Dirac delta function convolved with the PSF.

For each source identified in the segmentation map, the software uses a box of fixed size as a filter in the low-resolution image to find the centroid and estimate the flux within a given (square) aperture. The PSF is convolved with a Dirac delta function with an amplitude equal to a fraction of that aperture flux, and then subtracted from the image. Filtering and centroid positioning are executed within the object's boundaries as defined by the prior high-resolution segmentation map. This procedure is repeated on the residual image produced by the previous iteration until the flux of the treated source becomes smaller than a specified threshold. In this case, a minimum signal-to-noise ratio of 2.5 is set so that an object will be considered completed once its aperture flux, compared to the background, becomes smaller than that value. This also implies that not all the sources detected in the prior image will be extracted by IRACLEAN. Moreover, since the global sky background is recomputed at each iteration, the signal of a faint source – initially disregarded – may emerge from the background after several passes on the nearby objects. The iterative procedure of centroid positioning within the object's boundaries allows extended sources to be treated, and the fact that the flux is subtracted by convolving the PSF with a Dirac delta function centered on the centroid controls the contamination by neighbors. For more detail on the workings of IRACLEAN, the reader is referred to Sect. 7 of Hsieh et al. (2012), and their Figure 16 for an example of residual images.

User-controlled parameters are the threshold below which to stop cleaning, the filtering box size, the square aperture to measure IRAC flux and the fraction of flux to subtract at each iteration. In this configuration, a box of size 7×7 pixel is adopted to filter and to find the centroid, and a square aperture of size 9×9 pixel to estimate the aperture flux; the fraction of flux subtracted for each cleaning step is 20 %. The final flux of each object is the sum of the fluxes subtracted at each step. Since the centroid position is allowed to change at every iteration, the source is eventually modeled by a combination of Dirac delta functions that are not necessarily centered at the same point. The flux error is computed using the residual map by measuring the fluctuations in a local area around the object.

This implementation adopts the high-resolution $izYJHK_s$ detection image and its segmentation map produced by `SExtractor`. In order to parallelize the processing of the images, a mosaic of 14 × 14 tiles is made with a 0.′3 overlap in each direction. The PSF is modeled on a grid with spacing of 29″ across the full IRAC image in order to take into account its spatial variation using the software `PRFmap` (A. Faisst, private communication). When modeling the PSF at each grid point, the code takes into account that the final IRAC mosaic is made of multiple overlapping frames that can have different orientations with a PSF that is not rotationally symmetric. `PRFmap` models the PSF in each of the frames that overlap at a grid point and stacks them to produce the PSF model of the mosaic at that location. `IRACLEAN` thus provides photometry in channel 1 and 2 for more than a million sources over the whole field.

3.3.2 *The Farmer catalog*

Source detection

The source detection step is entirely equivalent to the procedure adopted for the Classic catalog. The Farmer utilizes the `SEP` code (Barbary, 2016) to provide source detection, extraction, and segmentation, as well as background estimation with near identical performance as classical `SExtractor`. Given their near identical performance, The Farmer uses `SEP` as both are written in Python and hence `SEP` is readily integrated into the existing workflow.

The detection parameters are configured identically between `SExtractor` and `SEP` where possible. Crucially, given that model-based photometry from The Farmer cannot be readily applied to saturated bright stars and sources contaminated by stellar halos, the HSC PDR2 bright star masks are adopted a priori to ensure the reliability of the derived photometry (see Section 3.2.7). Photometric extraction with The Farmer for COSMOS2020 is limited to the UltraVISTA footprint as this area contains all the bands used in the detection image, which are used by The Farmer to construct galaxy models. Including areas which lack complete $izYJHK_s$ coverage introduces undesirable inhomogeneities to the model constraints, and hence may adversely change the selection function. Photometry of sources within the HSC bright star masks is also not attempted with The Farmer as the halo light and the saturated stars are difficult to account for in a model, resulting in poor measurements and exponentially longer computational times. While there are 964 506 sources in the entire The Farmer catalog, only 816 944 sources lie within UltraVISTA footprint but outside the conservative HSC bright star halo masks. This is marginally larger than the number of sources detected in the Classic catalog (difference ∼3 %). Of these, ∼95 % have counterparts in the Classic catalog within 0.″6. Conversely, virtually all (>99 %) Classic catalog sources have counterparts in The Farmer catalog within the same radius over the same area. Generally, sources only included in The Farmer catalog are concentrated around unmasked bright star halos and their diffraction spikes (further underscoring the need for accurate a priori masking), and which are unlikely to possess well-fit models and so are easily flagged. Some, however, appear to be result of comparably more accurate de-blending of nearby sources by SEP, which given the ability to easily identify nonphysical detections, is advantageous for the important reason that two blended sources will not be well-fit by models unless they are identified as separate objects at detection. This will be further discussed in the context of The Farmer in Weaver et al., (in prep).

Once sources are detected, The Farmer identifies crowded regions with multiple nearby sources which although deblended at detection (i.e. have their own centroids), may have some overlapping flux which must be separated by the models. Hence, to avoid double-counting flux and to achieve the most robust modeling possible, these sources are modeled simultaneously. Such crowded regions are identified by dilating the source segmentation map, which assigns pixels to sources, in order to form groups of sources defined by contiguous dilated pixels. Sources which are not in crowded areas are expected to be a group of one source, whereas sources in crowded regions end up as members of larger groups to be modeled together.

PSF creation

In contrast with the PSF-homogenization strategy employed in the Classic catalog for all optical and NIR bands, The Tractor does not operate on images which are PSF-homogenized. Since the models it uses are purely parametric, The Tractor can simply convolve a given model with the PSF of a given band, which is generally a more tractable operation than PSF homogenization. The approach to generate PSFs for The Farmer catalog follows similarly with that of Classic, using spatially constant PSFs for the broad-bands and spatially varying PSFs for the Subaru medium-bands and IRAC bands.

A spatially constant PSF is computed for u, u^*, as well as all HSC and UltraVISTA bands with PSFEx. Point-source candidates are selected as described in Section 3.3.1. Since models are sensitive to the wings of sources, The Farmer benefits from particularly large PSF renderings. Typical unsaturated point-sources in optical and NIR images in this work are well-described by PSF stamps generated with 201 pixel diameters (30.″15).

Another consideration, introduced for the Classic catalog in section 3.3.1, is the highly variable PSF of the Suprime-Cam medium-bands. Although The Farmer does not use any kind of PSF-homogenization procedure and hence cannot overcome this variability in the same way as for the Classic catalog, it is still possible to overcome highly variable PSFs in model-based photometry by providing a particular PSF to a group of sources, similar to PRFMap which produces a theoretical PSF sampled over a fixed grid. However, this exact approach cannot be readily replicated for other bands, since there is a lack of sufficient theoretical PSFs for the Subaru medium-bands. Instead, a spatial grid is constructed using the PSF FWHM measured from a sample of point-like sources nearest to each grid point. The FWHM distribution is then discretized to form a set of PSFs at a gauge small enough to provide accurate PSFs for each grid point while maintaining the spatial sampling required to describe the variations across the field. Hence, for each medium-band a 20×20 grid consisting of 10 PSFs is built with a typical resolution of less than a tenth of a pixel. Then for a particular group of sources THE FARMER provides the nearest PSF sample to be used in the forced photometry modeling.

Lastly for IRAC, THE FARMER employs `PRFMap` to provide a spatially-varying PSF to each group of sources based on their nearest PRF sampling point, consistent with the `IRACLEAN` procedure described in section 3.3.1. The PSFs are then re-sampled to match the 0.″15 pixel scale of the mosaics.

Model determination

Details of the model determination procedure will be found in Weaver et al. (in prep.). This is a brief summary. THE FARMER employs five discrete models to describe resolved and unresolved, stellar and extragalactic sources:

1. **PointSource** models are taken directly from the PSF used. They are parameterized by flux and centroid position and are appropriate for unresolved sources.

2. **SimpleGalaxy** models use a circularly symmetric, exponential light profile with a fixed 0.″45 effective radius such that they describe marginally resolved sources and mediate the choice between PointSource and a resolved galaxy model. They are parameterized also by flux and centroid position.

3. **ExpGalaxy** models use an exponential light profile. They are parameterized by flux, centroid position, effective radius, axis ratio, and position angle.

4. **DevGalaxy** models use a de Vaucouleurs light profile. They are parameterized by flux, centroid position, effective radius, axis ratio, and position angle.

5. **CompositeGalaxy** models use a combination of ExpGalaxy and DevGalaxy models. They are concentric, and hence share one centroid. There is a total flux parameter as well as a fraction of total flux parameter to distribute the flux between the two components. Components have their own effective radii, axis ratios, and position angles.

These five models form THE FARMER's decision tree, whose goal is to both determine the most suitable model for a given source, and provide an optimized set of parameters to describe the shape and position of the source. Unlike some other model-based photometric techniques, the models in THE TRACTOR are purely parametric and hence do not require a high-resolution image stamp which must undergo PSF kernel convolution when photometering a different band. Although the exact implementation of the modeling can vary (e.g., choice of bands, library of models, etc.), for the present catalog The Farmer attempts to jointly model a group of nearby sources, using simultaneous constraints from each of the six individual $izYJHK_s$ bands used in the detection image. This ensures that the selection function is preserved by providing a model even for sources detected from one band.

The Farmer then uses its decision tree to select the most appropriate model type for each source in the group. The decision tree starts with unresolved or marginally resolved models (1,2) and moves towards more complex, resolved ones (3,4,5). Each level of the decision tree assumes the same initial conditions, excepting that some sources may already be assigned a model type in the latter stages. The tree must be tuned according to the data being used. In this work, marginally resolved SimpleGalaxy models must achieve a lower χ^2_N by a margin of 0.1 compared a unresolved PointSource model, thereby preferring the PointSource model whenever possible. If either model achieves a $\Delta\chi^2_N < 1.5$, then the next level is tried. If the ExpGalaxy and DevGalaxy models are not indistinguishable by $|\Delta\chi^2_N| = 0.2$ or neither achieves a $\chi^2_N < 1.5$, the most complex CompositeGalaxy is tried (see Weaver et al., in prep. for more details). Once a model type has been assigned to each source, the final ensemble of models is re-optimized to ensure that the derived model parameters reflect the actual model ensemble. If instead the parameters were adopted during the initial stages of the decision tree, then it would be possible for one source which has not yet been fit with the appropriate model type to influence the parameters of another nearby source. By re-computing the model parameters at the very end, when all the model types have been assigned, this case is avoided.

An example of the modeling procedure is shown in Figure 3.7, whereby two models are jointly determined for two nearby sources using each of the individual $izYJHK_s$ bands, simultaneously. It is stressed that the models are not constructed on the detection image itself, which suffers from PSF inhomogeniety which makes it not suitable for deriving morphologically-sensitive model constraints. The *i* band is shown as it is the deepest high-resolution band in the detection image and hence provides the greatest constraints on the morphology. Forced photometry on IRAC channel 1 (See Section 3.3.2) is shown to demonstrate the extent to which the prior information derived jointly from $izYJHK_s$ can adequately model IRAC flux, even for the most severely blended sources which apertures cannot accurately photometer.

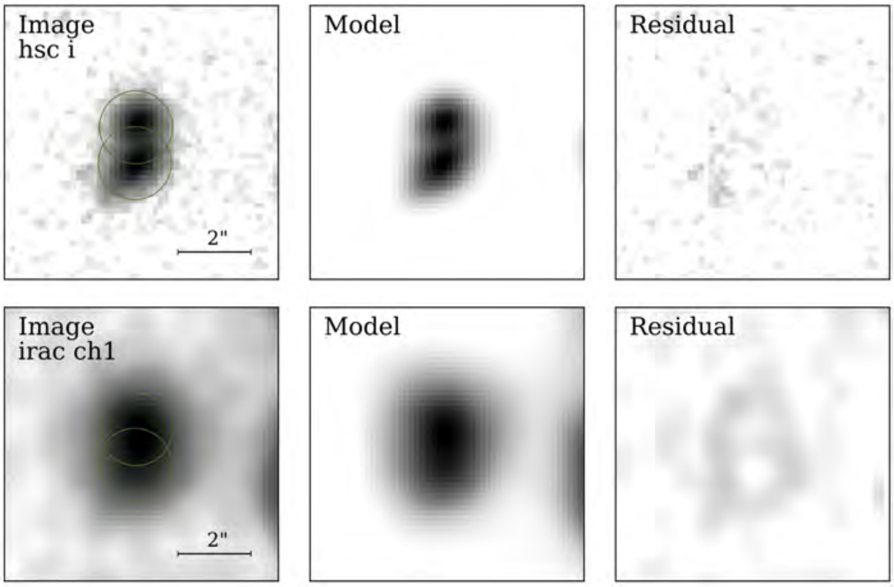

Figure 3.7: Demonstration of the model-fitting method from The Tractor. A pair of detected but overlapping sources is shown in the HSC i band (*top*). They are jointly modeled using The Farmer with constraints from each of the $izYJHK_s$ images in order to provide a parameterized solution which is suitably optimized, and from which the total flux is measured. The same pair of sources are shown in the less resolved IRAC channel 1 (*bottom*), where the two models are convolved with the channel 1 PSF and re-optimized using the channel 1 image to measure the flux contributed by each source. The extremely blended nature of this pair is underscored by the overlapping 2″apertures, consistent with the methodology of the Classic catalog. Pixel values are logarithmically scaled between the rms level and 95 % of the peak flux per pixel.

Forced photometry

With the model catalog complete for all detected sources, The Farmer can measure total model fluxes for every band of interest. The Farmer does this in a "forced photometry" mode, similar to the "dual-image" mode in `SExtractor`. In brief, the model catalog of a given group is initialized with the optimized parameters from the preceding stage. For each band, model centroids are allowed to vary with a strict Gaussian prior of 0.3 pix to prevent catastrophic failures. By doing so, The Farmer can overcome subtle offsets in astrometric frames between different images, and this can be done on an object-by-object basis to even overcome spatially varying offsets which may arise due to bulk flows in the astrometry. The optimization of these models produces total fluxes and flux uncertainties for each band of interest, keeping the shape parameters fixed. The flux measurement is obtained directly from the scaling factor required to match the models, which are normalized to unity, to the source in question. However, the flux uncertainties are derived by computing a quadrature sum over the weight map, weighted by the unit profile of the model, producing a similar result as traditional aperture methods but where the model profile is used in place of a fixed aperture. The weight maps are the same as those used by Classic. Importantly, the flux uncertainties reported in The Farmer catalog are not corrected with empty apertures, in contrast with the Classic catalog (see Section 3.3.1). The aperture-derived procedure used in Classic is inappropriate for model-based photometry, and although it may be expected that model-based methods would produce more precise measurements, they may still underestimate the true extent of correlated noise in the images and hence underestimate the uncertainty. This will be further discussed in Weaver et al. (in prep), and briefly evaluated later in Section 3.5.3 in terms of photometric redshift precision.

Photometry is performed with The Farmer for all CFHT, HSC, VISTA, and IRAC bands, as well as the Suprime-Cam intermediate bands. As such, there are two main differences with respect to Classic. Firstly, the older Suprime-Cam broad bands suffer from high spatial PSF variability, which is resolved in the Classic catalog by PSF-homogenizing each tile (see Section 3.3.1). However, this cannot be done for profile-fitting methods like The Farmer that do not operate on psf-homogenized images. Combined with the fact that these broad bands are eclipsed by deeper imaging from Hyper Suprime-Cam in almost all cases, they contribute very little to improving photo-z precision and can indeed even decrease accuracy if the PSF variability is not properly controlled. For these reasons the Suprime-Cam broad bands are only used when deriving photo-zs from the Classic photometry using `LePhare`, as described in Section 3.5. Secondly, photometry for IRAC channels 3 and 4 are performed with The Farmer to extend the wavelength baseline. This is largely due to the significantly cheaper computational power required for The Farmer relative to `IRACLEAN`. Although relatively shallow, in limited cases they can help place constraints on the rest-frame optical emission of potentially high-z sources. Details as to precisely which bands are available with each catalog can be found in associated `README` files.

Advantages and caveats

An important distinction between the two catalogs is that The Farmer provides total fluxes natively, without the need to correct for aperture sizes or perform PSF-homogenization. Since this advantage can be leveraged over different resolution regimes, The Farmer computes photometric measurements which are self-consistent. Additional metrics are also readily available from The Farmer. This includes the goodness-of-fit reduced χ^2_N estimate computed for the best-fit model of each source on a per-band basis, obtained by dividing the χ^2 value by the number of degrees of freedom, i.e. the pixels belonging to the segment for each source minus the number of fitted parameters. Measurements of source shape are provided for resolved sources, and as such they yield estimates of effective radii, axis-ratios, and position angles. These measurements are directly fitted in The Farmer, unlike in `SExtractor` where they are estimated from moments of the flux distribution. Uncertainties on shape parameters are deliverable as well, in the sense that they are a fitted parameter which is the result of a likelihood maximization and not a directly calculated quantity.

Likewise, centroids for both the modeling and forced photometry stages are also fitted parameters, and are delivered with associated uncertainties.

Another important consideration is that given the diversity of galaxy shapes and source crowding across ultra-deep imaging, it is inevitable that a model, or group of models, will fail to converge. Often it is due to either a bright, resolved source not being well described by smooth light profiles, an extremely dense group of sources, or a failure at detection to separate nearby sources (and hence assign the correct number of models to use), or a combination of all three. This problem is endemic to these methods and one which cannot be practically solved by manually tuning each fit; nor at this time by selecting tuning parameter based on statistics, which are unlikely to be effective in the most ill-conditioned cases. Thankfully, like in `SExtractor` which indicates failures by a combination of boolean flags, model based photometry can also be accompanied by a flag to indicate a failure to converge. Importantly, for those which do converge, however, model-based methods can provide more information about untrustworthy measurements than any aperture-based method by leveraging the statistical properties of the residual pixel distribution (e.g. χ^2 and other χ-pixel statistics) to precisely indicate the extent of these failures, and hence convey in comparably greater detail the extent to which the user can rely on any given measurement.

3.4 PHOTOMETRY COMPARISON

With the photometry from the two independent methods in hand, this section presents a comparison of the photometric catalogs as measured by differences in magnitudes, colors, and photometric uncertainties. In addition, a comparison is made with literature results of galaxy number counts. The primary motivation for these tests is to validate the two catalogs, in particular the performance of the relatively newer photometry from The Tractor generated with The Farmer. The performance of The Tractor code has been demonstrated previously (see Lang et al. 2016a), hence this work focuses on additional validation of the performance particular to The Farmer configuration used here. Additional validation of The Farmer where its performance is benchmarked against simulated galaxy images is provided in Weaver et al (in prep.).

A matched sample of sources common to both The Farmer and Classic is constructed consisting of 854 734 sources matched within 0.″6, for which The Farmer obtained a valid model and hence has extracted photometry. The sample contains 95.8 % of valid The Farmer sources, most of which are matched well below 0.″6. As explained in Section 3.3.2, those which are unmatched are typically marginally detected sources, or blends which are de-blended by only one of the detection procedures.

3.4.1 *Magnitudes*

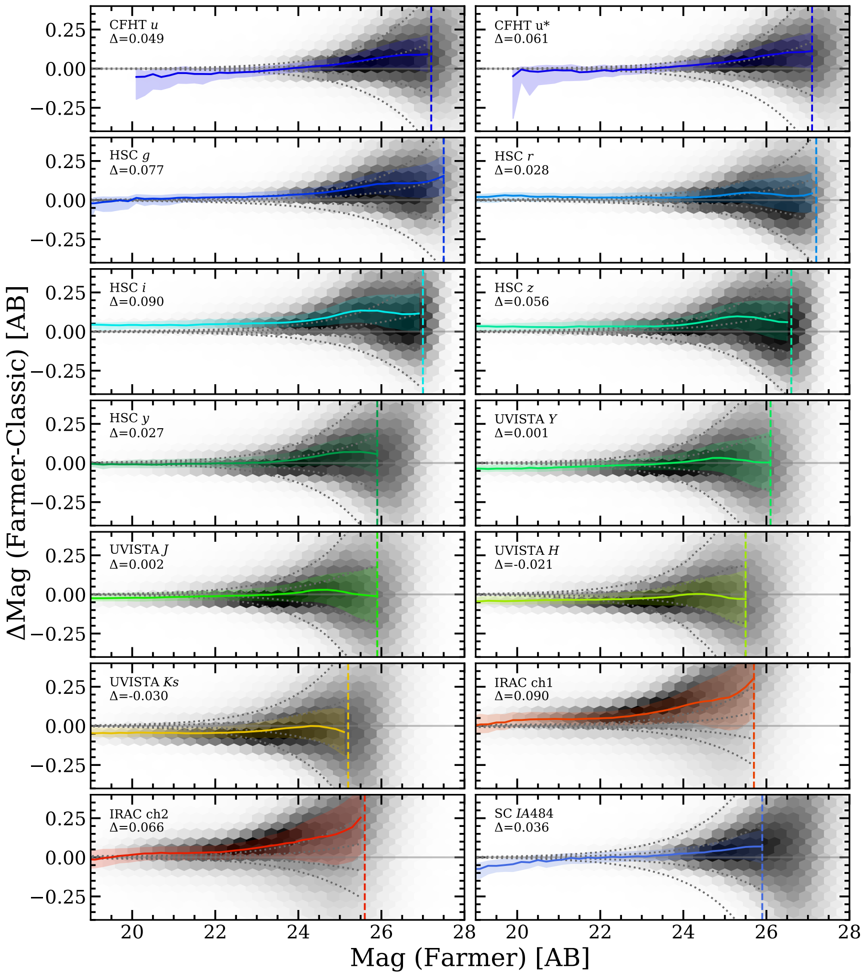

Figure 3.8: Summary of the difference between broad-band magnitudes measured by THE FARMER and CLASSIC catalogs, ΔMag. Magnitudes for CLASSIC are the re-scaled 2″total magnitudes. For UltraVISTA, sources in both the ultra-deep and deep regions are shown. Agreement for individual sources is shown by the underlying density histogram which is described by the overlaid median binned by 0.2 AB with an envelope containing 68 % of points per bin (solid line and shaded area). 1σ and 3σ photometric uncertainty estimates on ΔMag are indicated by the grey dotted curves. The 3σ depths measured with 3″ diameter apertures as reported in Table 2 are shown by vertical dashed lines. The median Δ magnitudes for sources brighter than the depth limit are reported in each panel.

A comparison of broad-band magnitudes derived independently with the two methods is shown in Figure 3.8. One medium-band is included for reference. Here the re-scaled 2″ total aperture magnitudes are used to compare with the model magnitudes from THE FARMER. The comparison is limited only to sources brighter than the 3σ depth as reported in Table 2 and indicated by the vertical dashed lines. For bands not included in the detection `CHI_MEAN`, these depths are upper bounds. The quadrature combined $\pm 3\sigma$ and $\pm 1\sigma$ uncertainty envelopes on ΔMag, computed by quadrature addition of the photometric uncertainties from both catalogs, are shown for reference by the grey dotted curves.

In general, there is excellent agreement between the photometric measurements from the two methods. As shown in Figure 3.8, the median systematic difference taken over all magnitudes is typically below 0.1 mag in all bands, and in some cases is noticeably smaller. If one were to remove this systematic median difference, then the remaining median differences in each magnitude bin would, for all bands, lie within the 3σ uncertainty threshold expected given the stated photometric uncertainties. In other words, the two sets of photometry are consistent within the expected uncertainties. The largest median differences occur for the faintest sources, but in most cases this is found to be $\lesssim 0.25$ mag, which is on the order of the expected uncertainty at these magnitudes. There is also noticeably low scatter between the measurements, as illustrated by the tight 68 % range envelopes about the medians. In most cases, the 68 % range envelope on the median spans the same range as the expected $\pm 1\sigma$ uncertainty envelope, the coincidence of which provides the first evidence validating the photometric uncertainties, discussed in full later in this section. Hence, it is established by multiple quantitative means that the two photometric measurements are broadly consistent.

A closer inspection, however, reveals a minor second-order curvature observed in all comparisons (including *IA*484) at the threshold where sources become unresolved in our ground-based NIR detection images, around $\sim$ 24.5 mag. At these magnitudes, photometry from The Farmer tends to be slightly fainter than that reported by `SExtractor` (or `IRACLEAN` for channel 1 and channel 2). However, these differences are generally very small and by median estimate are within the 3σ uncertainties for all bands. The fact that these features occur around the magnitude of each band where increasingly fainter sources are more likely to be point-sources may suggest that these sources are inadequately modeled because The Farmer chose a resolved model for a point-source, or conversely an unresolved model for a resolved source. If a resolved source is fitted with an unresolved model then the flux may be underestimated. Differences (in bands other than IRAC) may also arise from imperfections in rescaling the $2''$ apertures to total fluxes, compared to the native total fluxes obtained with The Tractor. This is particularly relevant given the high density of sources which can led to inaccurate estimates of object size, consequently producing inaccurate total flux measurements.

Regarding the IRAC photometry, which was obtained in both instances by profile-fitting techniques, discrepancies for faint sources cannot arise from aperture corrections. However, whereas `IRACLEAN` performs iterative subtraction of the PSF until convergence and sums all of the flux which has been subtracted, The Farmer solves for the flux as a model parameter without iterative subtraction. Yet, there is no evidence that any residual flux remaining from The Farmer fitting is significant enough to explain the observed discrepancy. Another potential difference which might explain the trend with brightness is that `IRACLEAN` performs iterative local background subtraction whereas The

Farmer performs a static background subtraction before performing photometry. However, it remains unclear as to exactly which methodology is most accurate. Definitively elucidating the cause of this observed discrepancy can only be obtained through simulation and is hence included in detail in Weaver et al. (in prep).

3.4.2 *Colors*

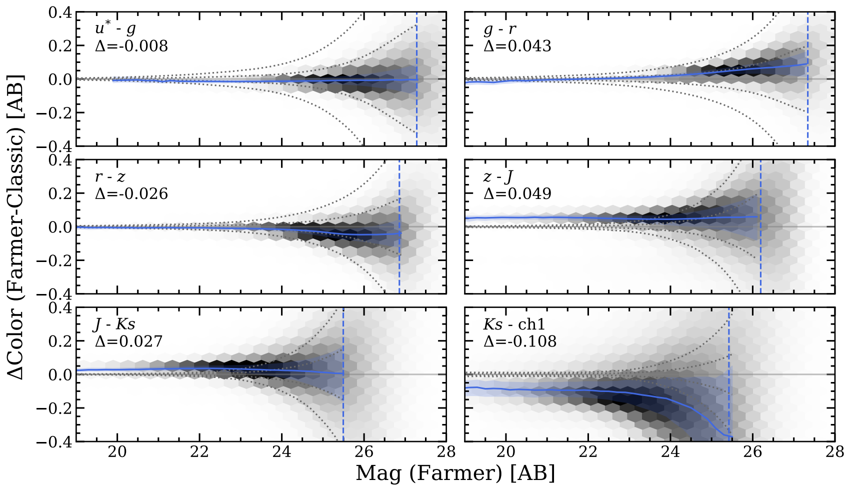

Figure 3.9: Comparison of broad-band colors between the The Farmer and Classic catalogs, Δcolor. The Farmer magnitudes of the first color term in each panel are shown on the x-axis. Colors for individual sources are shown by the underlying density histogram which is described by the overlaid median binned by 0.2 AB with a 68 % confidence interval. 1σ and 3σ photometric uncertainty estimates on the colors are indicated by the grey dotted curves and the mean 3σ depth computed from both bands of interest and measured with 3″ diameter apertures as reported in Table 2 are shown by vertical dashed lines, brighter than the median Δ are reported.

A comparison of six colors which contribute significantly to constraining a SED is shown in Figure 3.9. In similar fashion to the previous comparison, the distributions are described with a running median and 68 % range up to the nominal 3σ depth which is averaged for the two bands of interest. The expected $\pm 3\sigma$ and $\pm 1\sigma$ uncertainty thresholds on ΔColor, computed by the quadrature addition of the color uncertainties for each catalog, are shown by the grey dotted curves.

There is excellent agreement in colors, in some cases well-beyond the level of agreement achieved between individual bands. The median difference in color Δ is below 0.1 mag for all colors, with the best agreement seen by $u^* - g$, $g - r$, and $r - z$. Indeed, there is a lack of systematic difference in color and the observed scatter is well below the 1σ uncertainty expected for the color difference. The remaining panels show some level of systematic disagreement which is significant for bright sources. However, colors for faint sources are statistically consistent as they lie within the $\pm 1\sigma$ thresholds on the color uncertainty. This may be helped by the fact that the CLASSIC catalog does not require aperture-to-total rescaling to compute colors, thereby eliminating any relevant uncertainties present when comparing magnitudes only. In general, there is no evidence for a significant systematic difference in colors obtained by the two methods. Second-order curvatures are only visible at the faintest magnitudes, and are not significant even at the 1σ level after correcting for median shifts. The most significant deviation in color shown here is K_s − ch1, which features a relatively large systematic offset for bright sources and a strong second-order curvature for faint sources whereby THE FARMER obtains systematically bluer colors. Given that K_s magnitudes are well-matched between the two catalogs, this discrepancy in color must originate from the disagreement in faint IRAC channel 1 fluxes demonstrated in Figure 3.8. However, after correcting for the systematic median offset, the median curvature of the K_s − ch1 lies between the 3σ color uncertainty thresholds.

3.4.3 *Photometric uncertainties*

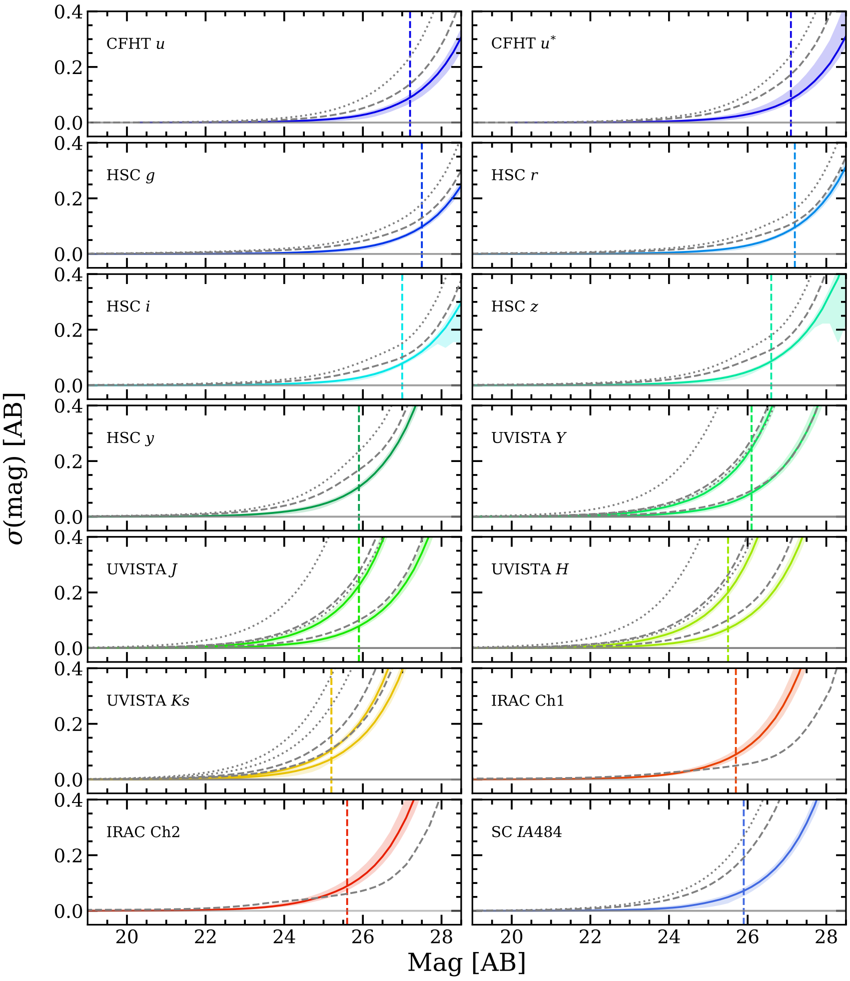

Figure 3.10: Growth of photometric uncertainties as a function of magnitude. The colored curves indicate the distributions for individual sources in The Farmer catalog, described by the running median and a tight envelope containing 68 % of sources. The grey curves represent the median growth of uncertainty for the total magnitudes in the Classic catalog derived from 2″ aperture photometry, shown by the dashed and dotted curves for the uncorrected and corrected uncertainties, respectively. The 3σ depths measured with 3″ diameter apertures as reported in Table 2 are shown by vertical dashed lines. The two curves shown for each band in $YJHK_s$ are due to different depths of the deep and ultra-deep regions.

One critically important aspect to compare is photometric uncertainties. The uncertainties from `SExtractor` are measured by quadrature summation of the $1/\sigma^2$ inverse-variance per-pixel (i.e. weight) map corresponding to the aperture on the source in the image. In contrast, The Tractor reports minimum variance estimates on the photometric uncertainty, although still using the same weight map. The Tractor computes flux uncertainties by a quadrature summation of weight map pixels, weighted by the unit-normalized model profile, which for point-sources is simply the PSF. This thereby prioritizes the per-pixel uncertainty directly under the peak of the model profile and places less weight on the per-pixel uncertainty near the edges of the model.

Figure 3.10 shows a comparison of magnitude uncertainties between The Farmer and the Classic catalogs. Unlike the magnitude and color comparisons, the sources which constitute this particular comparison are not matched between catalogs. They are however restricted to sources within the UltraVISTA area and clear of stellar halos indicated by the HSC bright star masks.

The distributions of magnitude uncertainties as a function of magnitude as measured by The Farmer for the primary broad-bands, as well as a medium-band for reference, are shown by colored binned medians with an envelope enclosing 68 % of sources per bin. The uncertainties in the UltraVISTA bands grow more quickly for the deep region compared to the ultra-deep region, and hence they are visualized here separately. The greatest differences between the rate of growth of uncertainties can be seen most noticeably for the Y, J, and H bands which feature the greatest difference in depth (see Table 2). K_s does not feature a significantly different growth rate between the deep and ultra-deep regions due to the near homogeneous coverage in DR4, a fact which will be useful when determining the mass completeness of the catalog.

For comparison, binned medians on the uncorrected magnitude uncertainties from the Classic catalog are indicated by the grey dashed curves. As described in Section 3.3.1, the uncertainties for most bands were then corrected using empty apertures, and are indicated by the grey dotted curves. The exception is IRAC, where the uncertainties for Classic are computed with `IRACLEAN` (see Section 3.3.1). Like with The Farmer, the magnitude uncertainties for UltraVISTA bands are split by depth. The faster growing curve is from the deep region, and the slower is from the ultra-deep region.

Photometric uncertainties smoothly and monotonically increase for fainter sources. For The Farmer, there is no evidence for discontinuities related to the transition between the resolved and unresolved regimes. There is, however, a difference between the magnitude uncertainties in that those measured with `SExtractor` and corrected are always larger than those from The Farmer for all bands except IRAC, where `IRACLEAN` was used. Yet in the case of the initial, uncorrected `SExtractor` uncertainties, this difference is much smaller. Moreover the two sets of uncertainties are in better agreement in the bluest bands (e.g., u, u^*, and HSC) where the spatial resolution is generally better than in the UltraVISTA bands. The opposite is true when comparing IRAC photometry, whereby The Farmer reports larger uncertainties than `IRACLEAN`. However, a noticeable level of consistency is achieved by The Farmer in that uncertainties from IRAC are similar to those from UltraVISTA, which should be expected given the similarity in the depths reported in Table 2. This consistency is not present in the Classic catalog, due to the difference between the methods of extraction from UltraVISTA and IRAC images.

Given that the photometric uncertainties measured with The Farmer are intrinsically linked to the underlying weight map, it is possible to quantify the internal consistency of these uncertainties using the reduced χ^2_N statistic, described in Section 3.3.2. In general, χ^2_N values are roughly unity for all bands. While this provides one measure of internal consistency, both the uncertainties reported by The Farmer and the χ^2 statistics fail to take into account pixel co-variance, which may be quite large, particularly in the lower resolution UltraVISTA mosaics which have been upsampled from their native 0.″34 per pixel to 0.″15 per pixel. It is then reasonable to conclude on this basis that although the uncertainties provided by The Farmer may be underestimated, they are indeed internally consistent with measurements which likewise ignore correlated noise, such as χ^2, and are in general suitable for use in SED-fitting. Additional correction of the photometric uncertainties from both The Farmer and Classic catalogs appropriate for SED-fitting is discussed further in Section 3.5.

3.4.4 *Galaxy number counts*

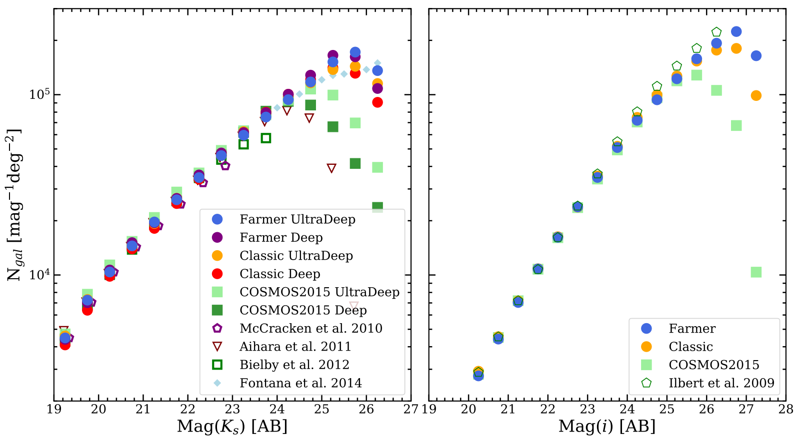

Figure 3.11: i- and K_s band galaxy number counts of the $izYJHK_s$-detected galaxies in the UltraVISTA ultra-deep and deep regions, compared to a selection of literature measurements, including previous COSMOS catalogs. The bins follow increments of 0.5 mag, with the exception of Fontana et al. (2014a) which uses 0.25 mag.

The galaxy number counts measured in COSMOS2020 are now compared to measurements in the literature. Figure 3.11 shows the galaxy number counts measured for bands on the bluest and reddest ends of the `CHI_MEAN` detection image, namely K_s (left panel) and i (right panel). The star-galaxy classification is adopted from the photometric redshift code `LePhare`, as described in Section 3.5.1, and is carried out similarly for both catalogs.

The effective area of The Farmer catalog is smaller than that of the Classic as photometry is not returned in the case of model failure with The Tractor, most often due to the presence of unexpected bright stars or large resolved galaxies which cannot be adequately modeled with one of the assumed smooth galaxy profiles (see Section 3.3.2). In this case the effective survey area is

Table 3: Bin centers and values of the $izYJHK_s$-selected logarithmic galaxy number counts shown in Figure 3.11 for both the The Farmer and Classic catalogs, in units of mag^{-1} deg^{-2} with bin widths of 0.5 mag.

	K_s-Deep		K_s-UltraDeep		i	
Mag	Farmer	Classic	Farmer	Classic	Farmer	Classic
19.25	3.64	3.61	3.65	3.66	3.01	3.04
19.75	3.85	3.80	3.86	3.86	3.23	3.27
20.25	4.03	3.99	4.02	4.02	3.44	3.47
20.75	4.18	4.14	4.16	4.16	3.64	3.66
21.25	4.29	4.26	4.29	4.29	3.85	3.86
21.75	4.42	4.40	4.42	4.43	4.03	4.03
22.25	4.56	4.53	4.54	4.54	4.21	4.21
22.75	4.68	4.66	4.66	4.68	4.38	4.38
23.25	4.79	4.78	4.78	4.80	4.54	4.55
23.75	4.90	4.90	4.88	4.90	4.71	4.71
24.25	5.00	5.00	4.97	4.99	4.86	4.87
24.75	5.11	5.08	5.07	5.07	4.97	5.00
25.25	5.22	5.15	5.18	5.14	5.08	5.10
25.75	5.21	4.12	5.24	5.16	5.20	5.19
26.25	5.03	4.96	5.13	5.06	5.29	5.25
26.75	-	-	-	-	5.35	5.26
27.25	-	-	-	-	5.22	5.00

corrected by subtracting the area occupied by sources for which a model is not available. Galaxy counts from COSMOS2015 are included for the deep and ultra-deep regions as the detection and photometry are equivalent to the Classic approach. The left panel of Figure 3.11 shows the $izYJHK_s$-detected K_s band galaxy number counts computed over the $0.812/0.757\,\mathrm{deg}^2$ of the HSC-masked ultra-deep region of UltraVISTA and over the $0.592/0.536\,\mathrm{deg}^2$ of the deep region as measured by photometry from both the Classic and The Farmer catalogs, respectively (see the corresponding `README` file for most up-to-date areas). There is good agreement with previous studies both within COSMOS (McCracken et al., 2012; Laigle et al., 2016) and from other surveys (Aihara et al., 2011; Bielby et al., 2012; Fontana et al., 2014a) over the regime where comparison is possible. The counts from both COSMOS2020 catalogs are in excellent agreement. The Farmer counts have slightly better completeness which may be due to the larger number of deblended sources at faint magnitudes. Notably, the COSMOS2020 completeness limit is ~ 1 mag deeper compared to COSMOS2015, which is due to a combination of both deeper infrared data and a much deeper detection image.

Similarly, the right panel of Figure 3.11 shows the $izYJHK_s$-detected i band galaxy number counts computed over the entire $1.403/1.234\,\mathrm{deg}^2$ of the HSC-masked UltraVISTA region for the Classic and The Farmer catalogs, respectively. Literature results from the i-selected counts of Ilbert et al. (2009) are included for reference. At the bright end, these counts are in excellent agreement with our measurements. At the faint end, however, Ilbert et al. are above our COSMOS2020 measurements. To identify the cause of this disagreement, a representative sample of $24 < i < 25$ objects detected only in Ilbert et al. were visually inspected in the detection `CHI_MEAN`, i, and K_s images, finding virtually all to be within the halos of bright foreground objects and stars. This is especially true for the K_s image, whose halos are significantly more extended relative to i, which in a `CHI_MEAN` construction can lead to noise structures resembling real sources even at $i \sim 25$. A reasonable explanation, therefore, is that the higher counts of Ilbert et al. are due to spurious sources created by an overly-aggressive deblending threshold.

3.5 PHOTOMETRIC REDSHIFTS

Photometric redshifts are computed using both Classic and The Farmer catalogs. First, photometric measurements are corrected for Galactic extinction at each object position using the Schlafly et al. (2011) dust map[5]. In the next sections, photometric redshifts are computed using both `LePhare` (Arnouts et al., 2002; Ilbert et al., 2006) and `EAZY` (Brammer et al., 2008), followed by a comparison between the two methods.

3.5.1 *LePhare*

The first set of photo-z is computed following the same method as in Laigle et al. (2016). Both galaxy and stellar templates are fitted to the observed photometry using the code `LePhare`[6] (Arnouts et al., 2002; Ilbert et al., 2006) with the same configuration as Ilbert et al. (2013).

Before fitting, 0.02 mag is added in quadrature to the photometric errors of the data in the optical, 0.05 mag for J, H, K_s, ch1, and the three narrow-bands, and 0.1 mag for ch2. Such an approach is common in numerous surveys (e.g. Arnouts et al., 2007), i.e., to include uncertainties in the color-modeling (more important near-infrared and in the narrow-bands due to the emission lines). Fluxes are used to perform the fit (as opposed to magnitudes), with the clear advantage of not introducing upper-limits. Given the uncertainties in the calibration of the Suprime-Cam/g^+, and the availability of deeper HSC images covering the same wavelength, this band is not included. Similarly, the shallow z^+ photometry is not used, since the Suprime-Cam/z^{++} and HSC/z images are deeper and already cover this wavelength range. IRAC channel 3 and 4 are not included given the difficulty to model the emission from polycyclic aromatic hydrocarbon (PAH) in the mid-infrared[7] and their shallower depth (Sanders et al., 2007).

Stellar templates include the library from Pickles (1998), the white dwarf templates of Bohlin et al. (1995), and the brown dwarf templates from Chabrier et al. (2000), Baraffe et al. (2015, BT-Settl/CIFIST2011_2015) and Morley et al. (2012, 2014). All the brown dwarf templates extend to at least 10 μm in the infrared. The blue limit of these templates is between 0.3 and 0.6 μm, and the flux density at bluer wavelengths is set to zero. Indeed, cool brown dwarfs belong to the very faint population of sources, and are expected to

5 Schlafly et al. (2011) re-scaled the entire Schlegel et al. (1998) dust map by a factor of 0.86.
6 https://www.cfht.hawaii.edu/~arnouts/LEPHARE/lephare.html
7 The 6.2 μm and 7.7 μm PAH lines contribute to the IRAC channel 4 photometry at $z < 0.3$, and the 3.3 μm line to both channels 3 and 4 with a lower contribution.

not be detected in the optical. Stellar templates with an effective temperature $T_{\mathrm{eff}} < 4000\,\mathrm{K}$ are rejected in the case that the physical parameters do not satisfy the constraints from Saumon et al. (2008).

Regarding galaxy templates, the original library (Ilbert et al., 2009) includes elliptical and spiral galaxy models from Polletta et al. (2007) interpolated into 19 templates to increase the resolution, and 12 blue star-forming galaxy models from Bruzual et al. (2003, hereafter BC03). Two additional BC03 templates with exponentially declining star-formation rate (SFR) were added to improve the photo-z of quiescent galaxies (Onodera et al., 2012). Extinction is a free parameter with reddening $\mathrm{E}(B-V) \leq 0.5$, and the considered attenuation curves are those of Calzetti et al. (2000b), Prevot et al. (1984), and two modifications of the Calzetti law including the bump at 2175 Å (Fitzpatrick et al., 1986b) with two different amplitudes. Emission lines are added using the relation between the UV luminosity and [Oii] emission line flux, as well as fixed ratios between dust-corrected emission lines following Ilbert et al. (2009). It is imposed that the absolute magnitude in the rest-frame Suprime-Cam/B band is $M_B \geq -24\,\mathrm{mag}$ which acts as a unique prior. The predicted fluxes for the templates are computed using a redshift grid with a step of 0.01 and a maximum redshift of 10.

Also included are a set of templates to account for active galactic nuclei (AGN) as well as quasars (see Table 3 of Salvato et al. 2009, and Salvato et al. 2011 for details). A measure of the goodness of fit and photo-z are provided for the best-fit AGN template which can be readily compared with that of the galaxy template to identify cases where the SED can be explained by emission from an AGN. This is especially important when considering stellar mass estimates, which can be inflated in the case of an undiagnosed AGN where the stellar continuum emission is unknowingly contaminated.

An initial run of `LePhare` fitting galaxies with spectroscopic redshifts provides a method to optimize the absolute calibration in each band. The method is the same as Ilbert et al. (2006): after having fixed the redshift to the spec-z value, the photometric offset of each band is derived by minimizing the difference between the predicted and observed fluxes. This procedure is applied iteratively until the offsets converge. The offset values are given in Table 4.

A key output of the photo-z code is the likelihood of the observed photometry given the redshift, $\mathcal{L}(\mathrm{data}|z)$, after having marginalized over the template set. The official photo-z estimate included in the catalog, noted z_{phot} hereafter, is defined as the median of the likelihood distribution. The z_{phot} error bar is comprised between $z_{\mathrm{phot}}^{\mathrm{min}}$ and $z_{\mathrm{phot}}^{\mathrm{max}}$, which are defined as 34 % of the likelihood surface below and above the median, respectively. The galaxy spectroscopic sample can be used to verify that these error bars actually represent 68 % confidence level intervals (see Section 3.5.3 for more details).

Galaxies are separated from stars and AGN in `LePhare` by combining morphological and SED criteria. The stellar sequence is isolated by comparing half-light radii and magnitude for bright sources in the *HST*/ACS and Subaru/HSC images. All the point-like sources falling on this sequence are classified as stars at $i < 23$ and $i < 21.5$ for ACS and HSC images, respectively. Point-like AGN sources are also removed by this criterion. Sources with $\chi^2_{star} < \chi^2_{gal}$ are also classified as a star, with χ^2_{star} and χ^2_{gal} being the best χ^2 obtained using the stellar and galaxy templates, respectively. This criterion is applied only for sources detected at 3σ in the K_s band or the IRAC channel 1, since the lack of near-infrared data could increase the risk of stellar contamination in the galaxy sample (Daddi et al., 2004; Coupon et al., 2009). We do not apply the criteria based on the χ^2 if the source is resolved, to avoid creating incompleteness in the galaxy sample.

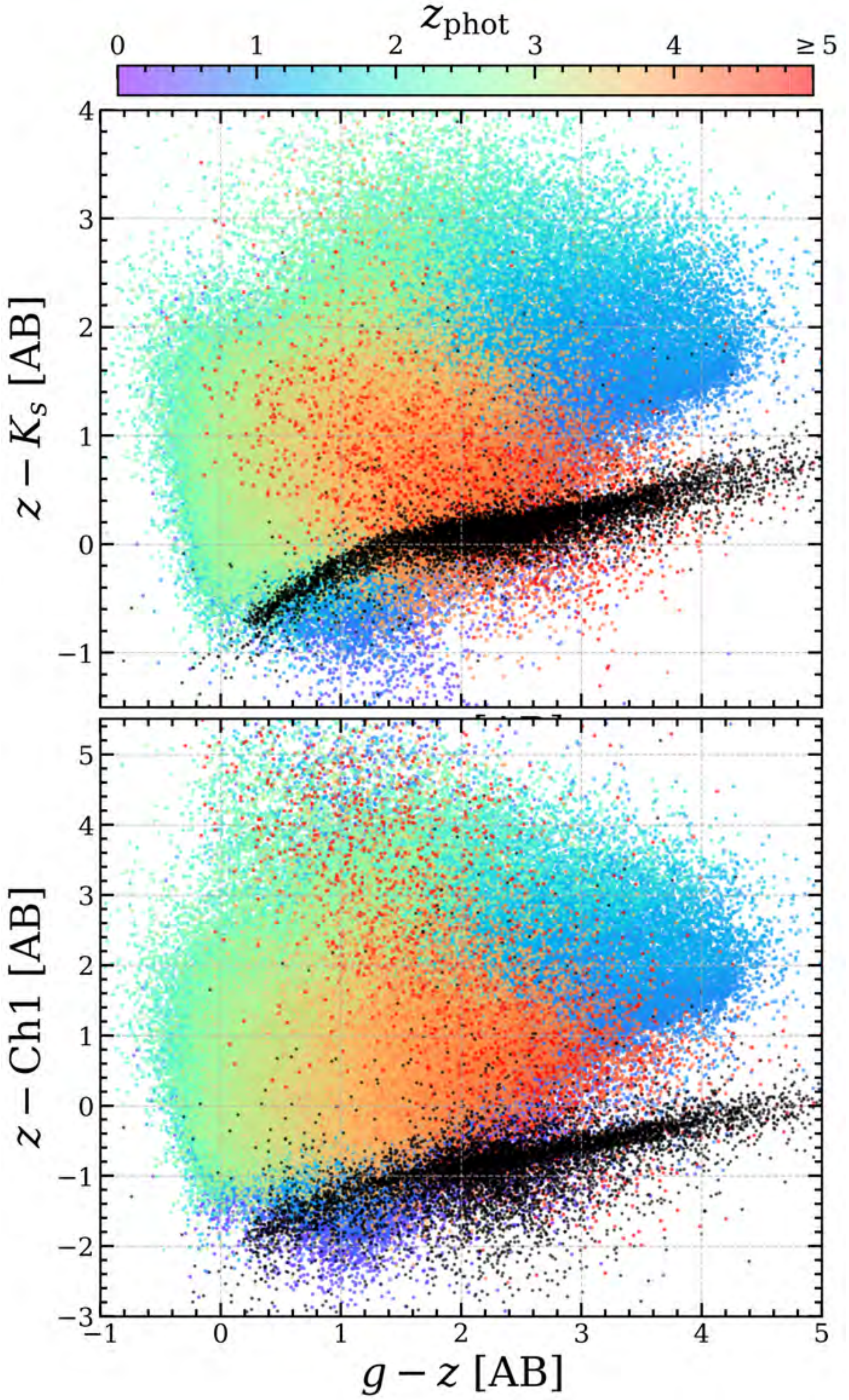

Figure 3.12: Color-color diagrams showing stars (black) and galaxies (colored by $z_{\rm phot}$) classified by `LePhare` for The Farmer photometry, shown in gzK_s (*top*) and gzch1 (*bottom*) color-color diagrams. For simplicity, galaxies with $z_{\rm phot} > 5$ are shown also by red points. Only sources with S/N > 3 in g, z, K_s and ch1 in the UltraVISTA area outside the HSC bright star halos are shown.

The result of this star-galaxy separation is shown in Figure 3.12. Here again The Farmer photometry is used, and the result is unchanged with Classic. Most of the sources classified as stars fall on the expected stellar locus of the two color-color diagrams (e.g., Figure 2 of Arcila-Osejo et al., 2013).

Although these classifications are made available in the catalogs (and explained in detail in the accompanying release documentation), it should be cautioned that this precise classification scheme may be sub-optimal for certain science investigations (e.g. where galaxies with stellar-like SEDs are science targets). Hence, this star-galaxy separation method is aimed at providing a baseline, conservative galaxy population from which to demonstrate the overall effectiveness of these catalogs, for instance with the galaxy number counts in Figure 3.11.

3.5.2 *EAZY*

Photometric redshifts are computed along with physical parameters using an updated version of the EAZY code[8] (Brammer et al., 2008) rewritten in Python. EAZY shares much of the strategy outlined for LePhare in the previous section, with the primary difference being the source of the population synthesis templates and how they are fit to the observed photometry. This computation uses a set of 17 templates derived from the Flexible Stellar Population Synthesis models (Conroy et al., 2009a, 2010a) with a variety of dust attenuation and ages from log-normal star formation histories that are chosen to broadly span the rest-frame UVJ color-space populated by galaxies over $0 < z < 3$. For each galaxy in the catalog, EAZY fits a non-negative linear combination of these templates integrated through the redshifted filter bandpasses to the observed flux densities and associated uncertainties. In this way, EAZY fits combinations of dust attenuation and star-formation histories to efficiently span the continuous color space populated by the majority of galaxies across the survey. For the EAZY photo-z estimates, the Subaru Suprime-Cam broad-band photometric measurements are not used, as these are generally significantly shallower than other nearby filters. Furthermore, the GALEX FUV and NUV are ignored, as these bands are relatively shallow and have broad PSFs that are difficult to combine with the other deeper filters.

As with LePhare, EAZY iteratively derives multiplicative corrections to the individual photometric bands (Table 4). For such a task, galaxies without a spec-z are also used, to mitigate the possible bias due to selection effects in the spectroscopic sample. At each step of the iteration, the median fractional residual is computed both for all bands individually and for all measurements in all bands sorted as a function of rest-frame wavelength. With many filters that overlap in the observed frame and galaxies across a broad range of redshifts, the catalog can largely break the degeneracy between systematic offsets in individual filters (e.g., from poor photometric calibration) and system-

8 https://www.github.com/gbrammer/eazy-py

atic effects resulting from the properties of the template set (e.g., continuum shape and emission line strengths). The correction routine is stopped after five iterations, where the updates are generally less than 1 %. For the final photometric redshift estimates, EAZY uses the "template error function" and apparent magnitude prior as described by Brammer et al., 2008.

Regarding star-galaxy separation, the current Python implementation of EAZY provides functionality for fitting stellar templates to the observed photometry, similar to LePhare. By default, EAZY uses a set of theoretical PHOENIX BT-Settl stellar templates (Allard et al., 2012) spanning a range of effective temperatures and calculates the χ^2 goodness of fit for each template individually (i.e., not as linear combinations). Included in the catalog is the minimum χ^2 of the fits to the stellar templates, as well as the effective temperature of the best-fit stellar model, which together may be used to separate stars from galaxies, possibly with the addition of morphological information to determine point-like sources.

Table 4: Values of the magnitude offsets used to optimize the absolute calibration in each band, derived with LePhare and EAZY for both photometric catalogs. When no value is indicated, the band was not used in the fit. The relative calibrations are normalized in K_s. Although included in The Farmer catalog, IRAC channels 3 and 4 are not used during the zeropoint calibration by LePhare. Observed photometry may be corrected by adding the appropriate values.

Band	LePhare The Farmer	LePhare Classic	EAZY The Farmer	EAZY Classic
NUV	−0.145	0.005	...	...
u	−0.092	0.001	−0.128	−0.097
u^*	−0.002	0.058	−0.182	−0.151
g	0.058	0.133	−0.010	0.020
r	0.081	0.133	0.046	0.057
i	0.018	0.102	0.006	0.054
z	0.019	0.090	0.038	0.078
y	0.070	0.105	0.091	0.103
B	...	−0.069	...	...
V	...	0.128	...	...
r^+	...	0.044	...	...
i^+	...	0.058	...	...
z^{++}	...	0.101	...	...
$IB427$	−0.111	−0.007	−0.187	−0.135
$IB464$	−0.057	0.014	−0.119	−0.094
$IA484$	−0.036	0.027	−0.086	−0.066
$IB505$	−0.035	0.031	−0.074	−0.051
$IA527$	−0.062	0.009	−0.092	−0.066
$IB574$	−0.104	−0.027	−0.120	−0.089
$IA624$	−0.015	0.037	−0.027	−0.012
$IA679$	0.145	0.213	0.146	0.174
$IB709$	−0.043	0.015	−0.036	−0.017
$IA738$	−0.054	0.009	−0.047	−0.021
$IA767$	−0.052	−0.009	−0.038	−0.032
$IB827$	−0.087	0.007	−0.060	−0.008
$NB711$	−0.030	0.028	...	...
$NB816$	−0.082	−0.016	...	...
Y	0.039	0.055	0.065	0.058
J	0.005	0.028	0.037	0.050
H	−0.049	−0.043	−0.029	−0.023
K_s	0.000	0.000	0.000	0.000
$NB118$	−0.034	−0.013	...	...
ch1	−0.184	−0.067	−0.127	−0.119
ch2	−0.186	−0.091	−0.200	−0.174
ch3	...	...	−0.168	...
ch4	...	...	−0.265	...

3.5.3 *Photometric redshift validation*

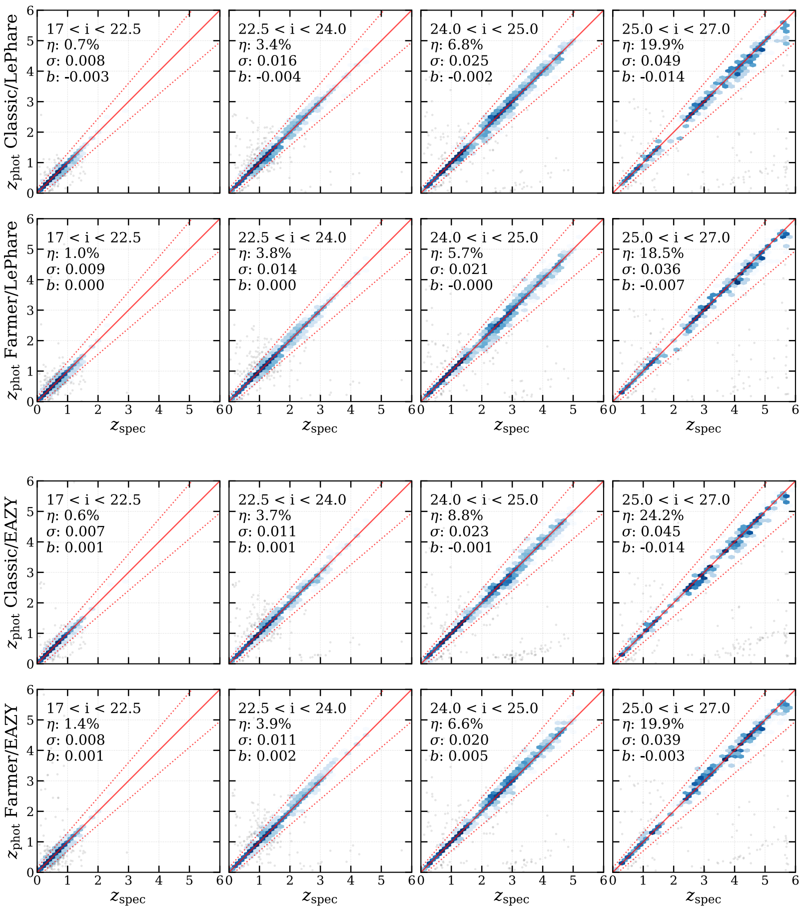

Figure 3.13: Photometric redshifts computed with `LePhare` and `EAZY`, split by apparent magnitude bin (from $i < 22.5$ on the left to $25 < i < 27$ on the right). *Top:* photo-z versus spec-z for the Classic and The Farmer photometric catalogs computed with `LePhare`. *Bottom:* photo-z versus spec-z for the Classic and The Farmer photometric catalogs computed with `EAZY`. The red solid line corresponds to the one-to-one relation, and the dashed lines correspond to the photo-z at $\pm 0.15(1 + z_{\rm spec})$. The fraction of sources outside the dashed lines (noted η), the precision measured with the normalized absolute deviation (noted σ), and the overall bias (noted b) are indicated in each panel. The nature of the off-diagonal points, shown individually, are discussed in the text. Bin color increases on a $\log_{10}$ scale. spec-z of $i > 26$ comprise 18% of sources shown in the rightmost $25 < i < 27$ panels.

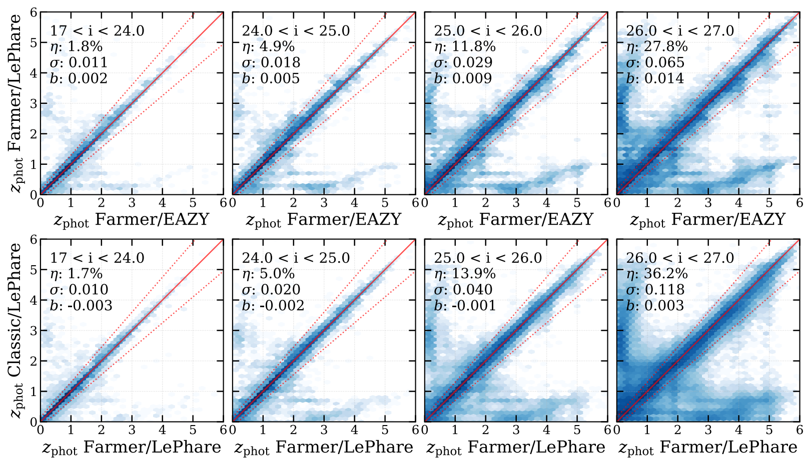

Figure 3.14: Photometric redshifts computed with `LePhare` and `EAZY` for the Classic and The Farmer photometric catalogs, split by apparent magnitude bin (from $i < 22.5$ on the left to $25 < i < 27$ on the right). *Top:* Comparison between the photometric redshifts computed with `LePhare` and `EAZY` for the full The Farmer photometric catalog. *Bottom:* Comparison between the photo-z derived from the Classic and The Farmer full catalogs computed with `LePhare` (excluding masked regions). The nature of the two groups of off-diagonal points is discussed in the text. Bin color increases on a $\log_{10}$ scale. Note that the magnitude bins are different than in Figure 3.13, to illustrate the behavior at faint magnitudes.

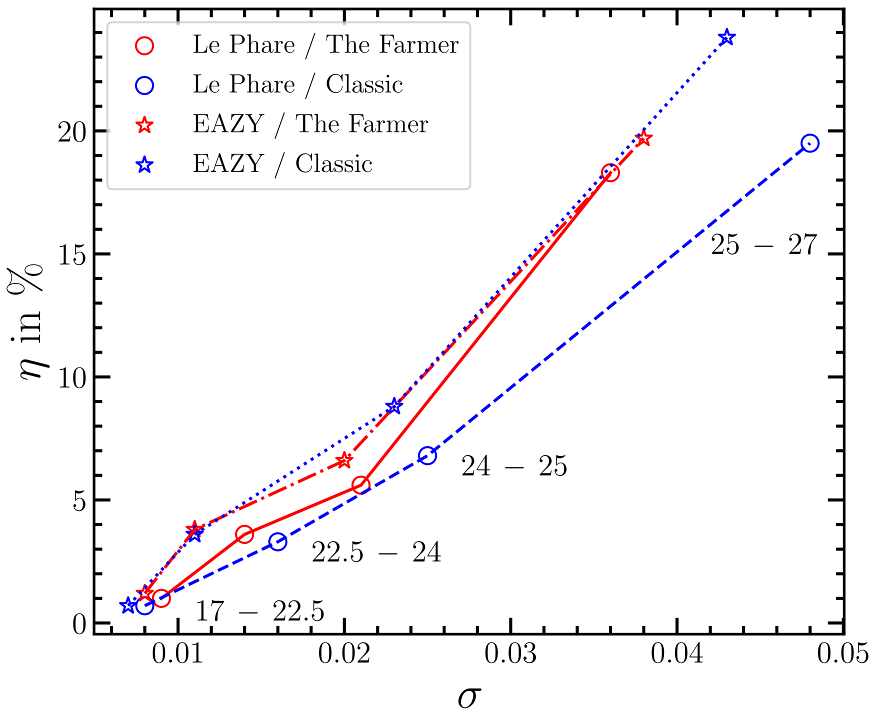

Figure 3.15: Comparison between the precision (σ_{NMAD}) and the outlier fraction for the two catalogs (the CLASSIC in blue and THE FARMER in red), and for the two photo-z codes (`LePhare` with circles and `EAZY` with stars). The statistics are computed per i band apparent magnitude bin, as indicated on the side of the points.

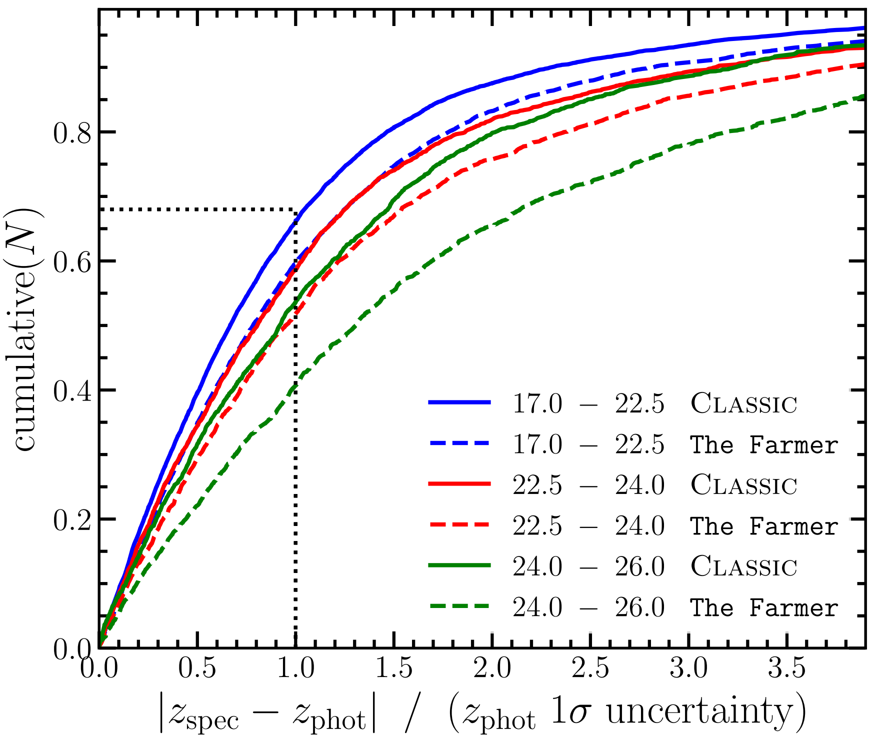

Figure 3.16: Cumulative distribution of the ratio between $|z_{\rm phot} - z_{\rm spec}|$ and the photo-z 1σ uncertainty, for both photometric catalogs and using `LePhare`. The photo-z 1σ uncertainty is taken as the maximum between $(z_{\rm phot} - z_{\rm phot}^{\rm min})$ and $(z_{\rm phot}^{\rm max} - z_{\rm phot})$. The solid and dashed lines correspond to the uncertainties from the Classic and The Farmer catalogs, respectively. For an unbiased estimate of the photo-z 1σ uncertainties, the cumulative number should reach 0.68 when the ratio equals 1 (black dotted line). The distributions are shown per bin of i band magnitude.

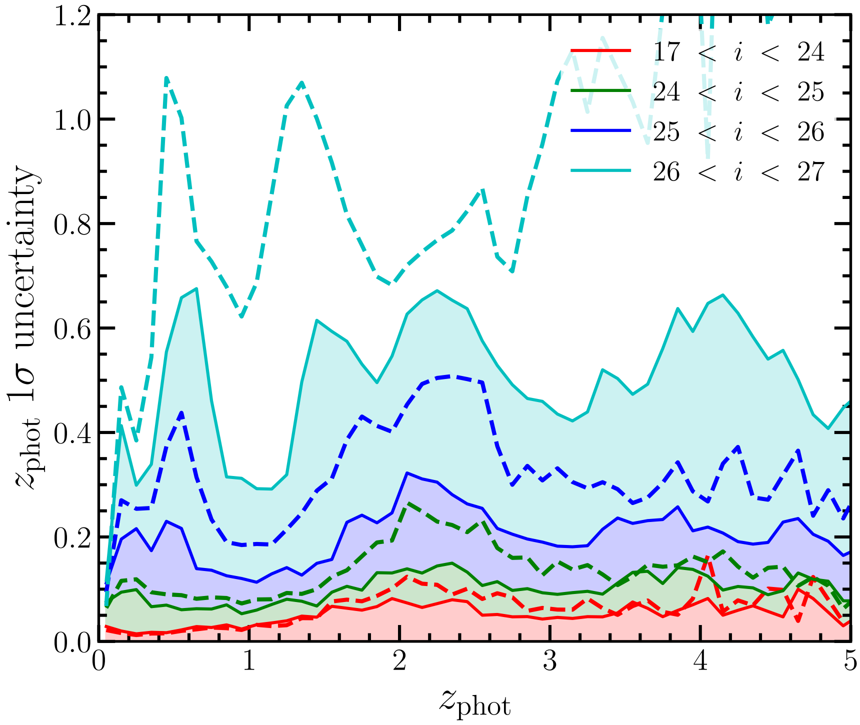

Figure 3.17: Median of the photo-z 1σ uncertainties (defined as in Section 3.5.3) shown as a function of redshift. The shaded areas correspond to the COSMOS2020 Classic catalog computed with `LePhare` and the dashed lines correspond to the COSMOS2015 catalog. The distributions are shown per bin of i band magnitude.

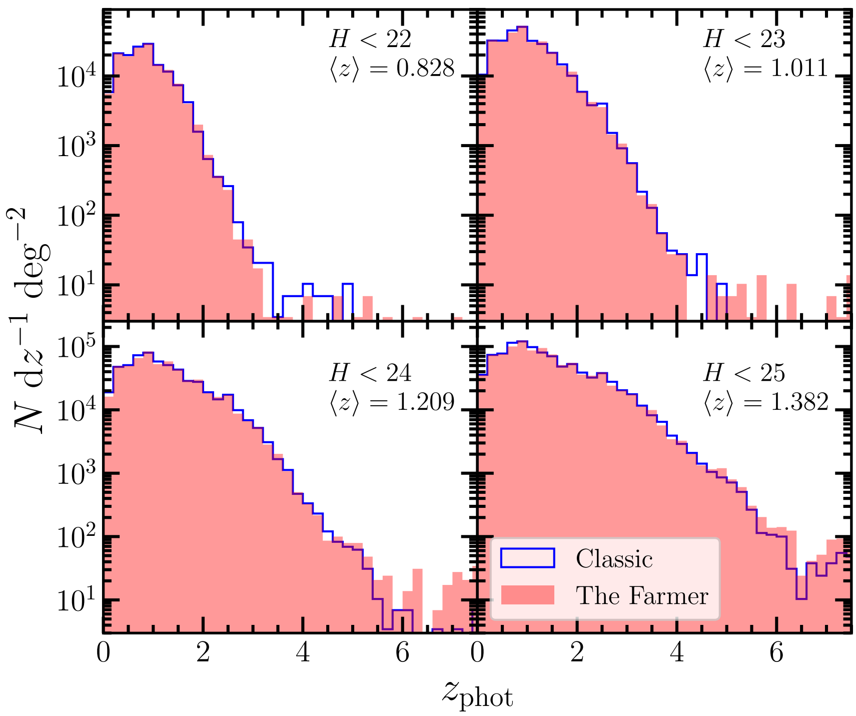

Figure 3.18: Redshift distribution for the Classic (blue) and The Farmer (red) full catalogs computed with LePhare. Each panel corresponds to a different magnitude limit in H band from The Farmer.

One unique aspect of this work different from Laigle et al. (2016) is the availability of two photometric catalogs created with different photometric extraction methods (see Section 3.3). By applying the same photo-z code to the Classic and The Farmer catalogs, it is possible to assess if one method to extract the photometry produces better results than the other. This is done by quantifying the precision of the photo-z using the normalized median absolute deviation (NMAD, Hoaglin et al., 1983), defined as

$$\sigma_{\mathrm{NMAD}} = 1.48 \times \mathrm{median}\left(\frac{|\Delta z - \mathrm{median}(\Delta z)|}{1 + z_{\mathrm{spec}}}\right), \tag{3.3}$$

following Brammer et al. (2008) as it is less sensitive to outliers compared to the normal definition Ilbert et al. (e.g. 2006). The fraction of outliers is noted η and defined, following Hildebrandt et al. (2012), as galaxies whose photo-z deviate from their spec-z by $|\Delta z| > 0.15\,(1 + z_{\mathrm{spec}})$. Lastly, the bias b and is computed as the median difference between photo-z and spec-z.

Comparisons between photo-z and spec-z are shown for both Classic and The Farmer catalogs in combination with LePhare and EAZY in Figure 3.13 and summarized in Figure 3.15. In general, the photo-z precision (given by σ_{NMAD}) is on the order of 0.01 $(1+z)$ at $i < 22.5$, and the precision is degraded at fainter magnitudes, but is still better than 0.025 $(1+z)$ at $i < 25$. For both catalogs, there is a population of galaxies with $z_{\mathrm{spec}} > 2$ and $z_{\mathrm{phot}} < 1$. This population is explained by the mis-identification between the Lyman and

Balmer breaks in the observed SED. This degeneracy appears clearly when comparing the photo-z derived for the full catalogs in Figure 3.14, especially for fainter objects where the lower signal-to-noise is not sufficient to constrain the identity of the break. The figure provides a straightforward demonstration of the remarkable similarity between the catalogs computed using the same photo-z code (LePhare) and the photo-z codes with the same catalog (The Farmer). The photo-z quality is similar between both catalogs, with a slight trend of having better results at $i < 22.5$ for the Classic catalog, while The Farmer catalog provides better results at fainter magnitudes.

The photo-z uncertainties are also an important aspect of the photo-z quality. If correctly estimated (i.e., representing the 1σ uncertainty) the fraction of spec-z which belong to the interval $[z_{\rm phot}^{\rm min}, z_{\rm phot}^{\rm max}]$ should be 0.68. Initially, this fraction was significantly smaller due to the photometric uncertainties being underestimated; therefore the error bars associated with the observed fluxes have been multiplied by a factor of 2× for the SED fitting. Figure 3.16 shows the cumulative distribution of the ratio between $|z_{\rm phot} - z_{\rm spec}|$ and the 1σ uncertainty derived for the LePhare photo-z solutions after boosting the flux error bars. The 1σ uncertainty is defined as the maximum between $(z_{\rm phot} - z_{\rm phot}^{\rm min})$ and $(z_{\rm phot}^{\rm max} - z_{\rm phot})$. The cumulative distribution of the bright sample ($i < 22.5$) now reaches 0.68 as expected, while the photo-z uncertainties of objects at $i > 22.5$ are still underestimated. This effect was already discussed in Laigle et al. (2016) and is seen also in EAZY. Since it is limited to faint galaxies, it may be due to a selection bias in the spectroscopic sample rather than a problem in the photo-z uncertainties (see Laigle et al., 2019). For this reason, no further correction is applied to the uncertainties of $i > 22.5$ objects. The effect is more pronounced in the The Farmer catalog since its photometric uncertainties are typically smaller, as they are not re-scaled to the same extent as in the Classic catalog (see Section 3.4.3). These larger uncertainties explain the more realistic photo-z errors in Classic, and may also help to explain the lower precision for faint sources as the photo-z are more uncertain.

Figure 3.17 illustrates the evolution with redshift of the 1σ photo-z uncertainties in several i band magnitude bins, as derived from the LePhare photo-z. There is an increase of the 1σ uncertainty between $z < 1$ and $1.5 < z < 2.5$. This increase is explained by the Balmer break being shifted out of the medium-band coverage, as well as blue galaxies at high redshift with low signal-to-noise in the near-infrared bands. Since the photo-z based on the Classic catalog are estimated using similar techniques as Laigle et al. (2016), the photo-z uncertainties computed with both catalogs can be compared. For this comparison, the photo-z uncertainties in both catalogs are re-scaled in order to make them consistent with 68 % of the spec-z falling into the 1σ error[9]. The result is that the photo-z are improved at $1.4 < z < 3$ at all magnitudes owing to the gain in UltraVISTA depth, and at faint magnitudes ($i > 25$) over the full

9 The COSMOS2020 photo-z uncertainties are re-scaled by a factor $1 + 0.1\,(i - 21)$ for the galaxies fainter than $i > 21$. Applying the same method and using the new spec-z sample, the COSMOS2015 photo-z uncertainties are re-scaled by a factor 1.3.

redshift range thanks to the new HSC and CFHT data. While COSMOS2015 photo-z were unreliable at $i > 26$, the new catalog can be used also at fainter magnitudes, depending on the scientific application. In summary, photo-z uncertainties reported in COSMOS2020 match those found 0.7 magnitudes brighter in COSMOS2015, a considerable gain.

Figure 3.18 shows the photo-z distribution of sources common to both the Classic and The Farmer catalogs in four selections of H band magnitude. As expected, the mean redshift increases toward faint magnitude from $z \sim 0.82$ at $H < 22$ to $z \sim 1.37$ at $H < 25$. There is an excellent agreement between the mean redshifts of both catalogs, within $\sim 0.01 - 0.02$. The mainly near-infrared selection in $izYJHK_s$ allows for the detection of a significant sample of galaxies above $z > 6$ ($100 - 300$ at $H < 25$ depending on the catalog). The Farmer catalog includes a higher density of $z > 6$ sources (by a factor almost two in the faintest bin). This is discussed in detail in Kauffmann et al. (in prep.).

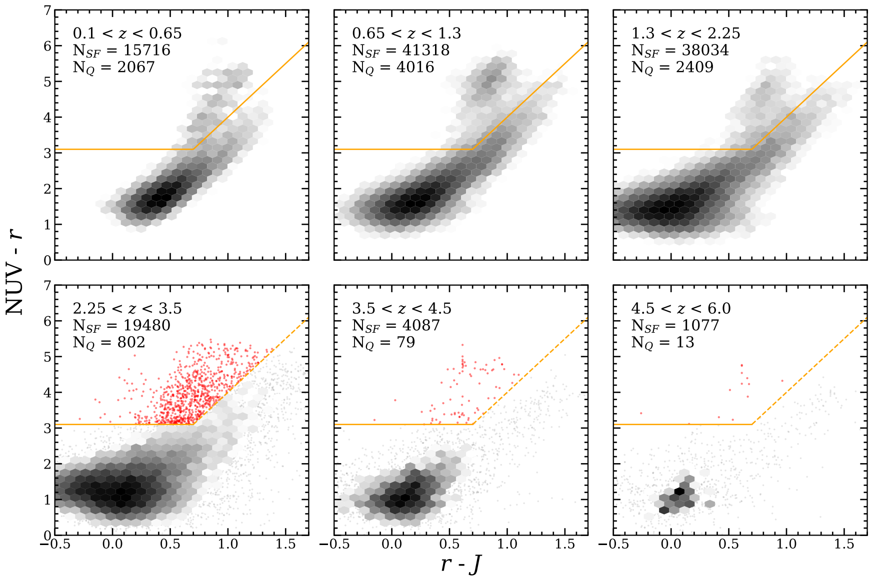

Figure 3.19: Identification of quiescent galaxies in bins of redshift by selection in rest frame NUV $- r$ and $r - J$ colors using the LePhare results, computed with The Farmer for sources which lie above their respective mass completeness limit. The selection is made using the prescription of Ilbert et al. (2013) shown in orange. For clarity, quiescent galaxies at $z > 2.25$ are shown by individual red points. $r - J$ colors are highly uncertain at $z > 2.6$ where the rest-frame J band is extrapolated redward of the available photometry, and hence have an uncertain classification marked by an orange dashed line.

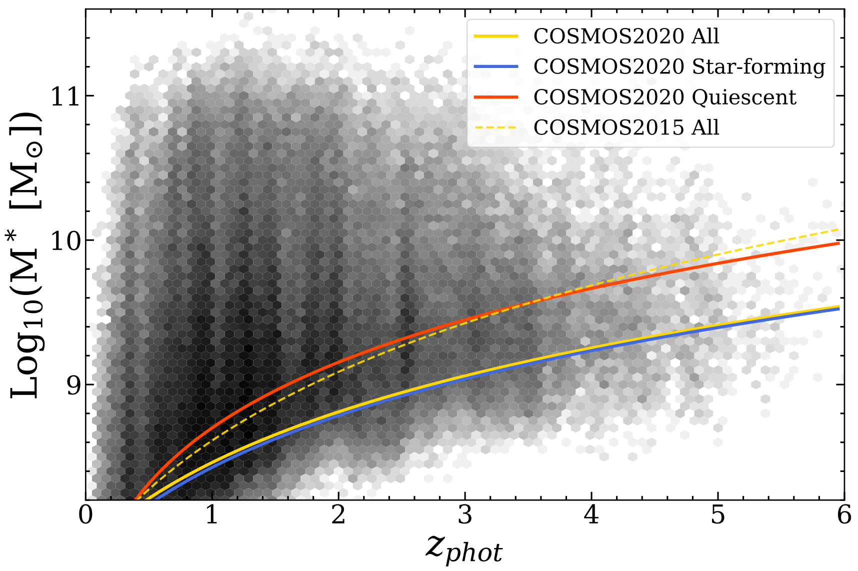

Figure 3.20: Mass completeness for the total sample (yellow), as well as the star-forming (blue) and quiescent (red) populations using quantities derived from THE FARMER and `LePhare` considering magnitude limits of IRAC channel 1. Limits are calculated based on the method introduced in Pozzetti et al. (2010) in a manner consistent with COSMOS2015 (Davidzon et al., 2017, yellow dashed). For clarity, the total sample limit has been raised by 0.02 dex so that both it and the star-forming limit are visible.

3.6 PHYSICAL PROPERTIES OF COSMOS GALAXIES

Now a first characterization of the sources classified as galaxies in Section 3.5.1 can be presented. Physical properties such as absolute magnitudes and stellar mass are computed using `LePhare` with the same configuration as COSMOS2015: a template library generated by BC03 models is fit to the observed photometry after fixing the redshift of each target to the photo-z estimated in the previous `LePhare` run (for more details, see Laigle et al., 2016). It should be noted that this standard configuration has been selected to be consistent with previous SED fitting results, even though recent work shows that the resulting stellar masses could be underestimated. For example, Leja et al. (2019a) finds that M^* estimates are $0.1 - 0.3$ dex larger when using complex SFHs to build their library, instead of the standard templates of FAST (Kriek et al., 2018). However, integrated fluxes (as provided by these catalogs) merges together the light of young stellar populations outshining the older ones. Sorba et al. (2018) show that when these different stellar components can be resolved (e.g. in the *Hubble* eXtreme Deep Field), a pixel-by-pixel SED fitting results in a galaxy stellar mass a factor $2 - 5\times$ larger (see also Abdurro'uf et al., 2018; Mosleh et al., 2020). On the other hand, tests with mock galaxy catalogs in Laigle et al. (2019) did not find such a significant bias, with an underestimation $< 20\%$.

The present analysis is limited to a classification of COSMOS2020 galaxies between star forming and quiescent, and a subsequent determination of their stellar mass completeness as a function of redshift; further investigation is deferred to future studies. Moreover, the following illustrates only the results generated with The Farmer and LePhare to provide the most direct comparison to Laigle et al. (2016) template fitting while demonstrating the effectiveness of the new The Farmer photometry. There are no significant differences when repeating the analysis with either Classic photometry or with EAZY.

3.6.1 *Galaxy classification*

Previous studies have devised a variety of techniques to identify quiescent galaxies using broad-band photometry. Williams et al. (2009) provides a prescription utilizing $U-V$ and $V-J$ rest-frame colors which has been broadly adopted in the literature (e.g., Muzzin et al., 2013b; Tomczak et al., 2014). Ilbert et al. (2013) and Arnouts et al. (2013b) proposed improving the selection by replacing $U-V$ with $\mathrm{NUV}-r$, since the latter can better separate galaxies with different star formation histories (see also Leja et al., 2019b).

This analysis adopts the rest-frame $\mathrm{NUV}-r$ vs. $r-J$ diagram described in Ilbert et al. (2013), where quiescent galaxies are defined to be those with $M_{\mathrm{NUV}} - M_r > 3\,(M_r - M_J) + 1$ and $M_{\mathrm{NUV}} - M_r > 3.1$. Measurements are provided by LePhare by convolving the best-fit template with the appropriate pass-band in the observed frame. Figure 3.19 shows the rest-frame $NUVrJ$ color-color diagram in six redshift bins from $z = 0.1$ to 6. The assembly of the quiescent population at late cosmic times is evident. Quiescent galaxies are rare at $z > 2$ (e.g., Ilbert et al., 2013; Muzzin et al., 2013b; Tomczak et al., 2014; Davidzon et al., 2017) but the large cosmic volume probed by COSMOS allows us to identify a significant number of candidates. However, a portion of them are expected to be star-forming galaxies that contaminate the high-z quiescent locus due to large uncertainties in their rest-frame colors (especially at $z > 2.6$ where M_J corresponds to observed wavelengths redder than channel 2).

3.6.2 *Stellar mass completeness*

The stellar mass completeness of our galaxy sample is empirically computed following the method described in Pozzetti et al. (2010), discriminating between star-forming and quiescent populations. This method is commonly used in the literature (e.g., Ilbert et al., 2013; Moustakas et al., 2013). It converts the detection limit of a given survey, given by the apparent magnitude m_{lim}, into a redshift-dependent threshold in stellar mass M_{lim} computed using the mass-to-light ratio of galaxies brighter than m_{lim}. Their stellar masses, estimated via template fitting, are re-scaled by a factor $10^{-0.4(m_i - m_{\mathrm{lim}})}$, where m_i is the magnitude of the i-th galaxy. One can determine M_{lim} in a given redshift bin from the distribution of such re-scaled masses: e.g., their 95^{th}

percentile can define the smallest mass at which most of the objects would still be observable.

The case of COSMOS2020 is more complicated because it is now possible to quantify $m_{\rm lim}$ not in a single band but for the `CHI_MEAN` $izYJHK_s$ detection image itself. Adopting the sensitivity limit in the K_s band (Table 2) is a conservative choice that disregards the numerous NIR-faint objects detected thanks to the deep HSC photometry. This bias has already been discussed for COSMOS2015 (see Davidzon et al., 2017) and it is now more relevant after the addition of the i band in the `CHI_MEAN` image which was not considered in 2015. Therefore the analysis proceeds as in Davidzon et al. (2017) by computing $m_{\rm lim}$ in IRAC channel 1, using the CANDELS-COSMOS catalog (Nayyeri et al., 2017a) as a reference parent catalog[10]. Source completeness in channel 1 is related not only to the properties of the IRAC mosaic itself, but also to the depth of the $izYJHK_s$ image, which is used as a prior for source extraction (Section 3.3.1 and 3.3.2). The choice to use channel 1 over K_s is motivated by the fact that channel 1 probes the bulk of stellar mass at $z > 2.5$, where the Balmer break is shifted beyond the optical-NIR bands. While `LePhare` and `EAZY` estimate broadly similar masses, this particular mass completeness is computed with masses reported by `LePhare` with The Farmer. Other combinations may produce a marginally different mass completeness limit, and should be re-derived for specific science applications.

A common sample is constructed by cross-matching IRAC channel 1 sources of COSMOS2020 to the deeper CANDELS catalog in the $\sim 200\,\mathrm{arcmin}^2$ where the two overlap. At $m_{\rm lim} = 26\,\mathrm{mag}$, about 75 % of the CANDELS sources are also recovered by The Farmer[11]; the completeness at that magnitude was $< 50\,\%$ in COSMOS2015. With $m_{\rm lim}$ in hand, galaxy masses are re-scaled to compute $M_{\rm lim}$ in bins of redshift (see Figure 3.20), to which a polynomial function in $1 + z$ is fitted The result is:

$$M_{\rm lim}(z) = -1.51 \times 10^6 (1 + z) + 6.81 \times 10^7 (1 + z)^2$$

for $z < 6$, which is more complete by ~0.5 dex compared to Davidzon et al. (2017). Since the boundary used here is the 95th percentile of the re-scaled mass distribution and the choice of $m_{\rm lim}$ already implied that about 25 % of the objects are missing, it is expected that $M_{\rm lim}$ corresponds to a 70 % completeness threshold.

The procedure is repeated separately for the star-forming and the quiescent sample, both shown in Figure 3.20. Quiescent galaxies start to be incomplete at stellar masses ~ 0.4 dex higher than the total sample since they have larger mass-to-light ratios. $M_{\rm lim}$ at $z < 2.5$ is additionally computed starting from the K_s limit (Table 2) and following precisely the procedure of Laigle et al. (2016). However, due to the nearly uniform coverage of the new data set, there is not

10 In the COSMOS field, CANDELS detection image *HST*/F160W has a 5σ limit at 27.56 mag within $0.''34$ diameter apertures, corresponding to twice the PSF FWHM.

11 The fraction of recovered CANDELS sources is the same with the Classic catalog.

a significant difference between the completeness limits of the ultra-deep and deep regions. The K_s-based completeness is well-described by the function:

$$M_{\mathrm{lim}}(z) = -3.55 \times 10^8(1+z) + 2.70 \times 10^8(1+z)^2$$

for $z < 2.5$ and is more complete by ~ 0.5 dex compared to the same threshold found in COSMOS2015 (Laigle et al., 2016).

3.7 CONCLUSIONS

This paper describes the creation and validation of COSMOS2020, a new set of two multi-wavelength catalogs of the distant Universe, each of which includes photometric redshifts and other physical parameters computed from two independent codes. COSMOS2020 builds on more than a decade of panchromatic observations on the COSMOS field. Compared to previous releases, COSMOS2020 features significantly deeper optical, infrared, and near-infrared data all tied to a highly precise astrometric reference frame, Gaia.

Starting from a very deep multi band detection image and using two different photometric extraction codes, one based on aperture photometry and one based on a profile-fitting technique, two photometric catalogs have been extracted. These photometric catalogs were then used to estimate photometric redshifts and stellar masses using two different codes, `LePhare` and `EAZY`. This enables us, for the first time, to make a robust estimate of the systematic errors introduced by photometric extraction and photometric redshift estimation over a large redshift baseline with an unprecedented number of objects over 2 deg^2. Our results show that all methods are in remarkable agreement. COSMOS2020 gains almost one order of magnitude in photometric redshift precision compared to COSMOS2015 (Laigle et al., 2016). In the brightest bin, $i < 22.5$, the catalogs reach redshift precision and outlier fraction are both below 1%. Even in the faintest $25 < i < 27$ bins, photometric redshift precision is still $\sim 4\,\%$ with an outlier fraction of $\sim 20\,\%$. A detailed comparison in Section 3.5 shows that at bright magnitudes the classic aperture catalog is marginally superior whereas at faint magnitudes the trend is reversed with the profile fitting technique providing a better result. This close agreement provides a unique validation of our measurement and photometric redshift techniques. Superseding our previous catalogs, COSMOS2020 represents an unparalleled deep and wide picture of the distant Universe. It will be of invaluable assistance in preparing for the next generation of large telescopes and surveys.

One can already start to imagine what COSMOS2025 might contain. After fifteen years of observations, the UltraVISTA survey will have been completed, providing an unparalleled near-infrared view of COSMOS. These data, combined with the *Spitzer* data presented here, will lay the foundation for a next-generation catalog combining deep high-resolution optical and infrared imaging data from *Euclid* and the *James Webb* Space Telescope with ultra-deep optical data from *Rubin*. Such a catalog will be an important step towards producing a mass-complete survey comprising every single galaxy in a representative volume from the present day to the epoch of reionization.

Acknowledgements

This paper is dedicated to Olivier Le Fèvre. Spectroscopic redshifts from his VIMOS instrument (often collected in surveys that he designed and led) played an invaluable role in preparing this catalog.

The authors thank Nathaniel Strickley, Dustin Lang, Clara Giménez Arteaga, Istvan Szápudi, Andrew Repp, and Emmanuel Bertin for helpful discussions. We are also grateful for the many helpful and constructive comments from the anonymous referee. We gratefully acknowledge the contributions of the entire COSMOS collaboration consisting of more than 100 scientists. The HST COSMOS program was supported through NASA grant HST-GO-09822. More information on the COSMOS survey is available at http://www.astro.caltech.edu/cosmos.

The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. ST, GB and JW acknowledge support from the European Research Council (ERC) Consolidator Grant funding scheme (project ConTExt, grant No. 648179). OI acknowledges the funding of the French Agence Nationale de la Recherche for the project "SAGACE". HJMcC acknowledges support from the PNCG. ID has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 896225. This work used the CANDIDE computer system at the IAP supported by grants from the PNCG and the DIM-ACAV and maintained by S. Rouberol. BMJ is supported in part by Independent Research Fund Denmark grant DFF - 7014-00017. CMC thanks the National Science Foundation for support through grants AST-1714528, AST-1814034 and AST-2009577, and additionally the University of Texas at Austin College of Natural Sciences, and the Research Corporation for Science Advancement from a 2019 Cottrell Scholar Award sponsored by IF/THEN, an initiative of Lydia Hill Philanthropies. The work of DS was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. GEM acknowledges the Villum Fonden research grant 13160 "Gas to stars, stars to dust: tracing star formation across cosmic time". DR acknowledges support from the National Science Foundation under grant numbers AST-1614213 and AST-1910107. D.R. also acknowledges support from the Alexander von Humboldt Foundation through a Humboldt Research Fellowship for Experienced Researchers. MS acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC)

The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. This work is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO program ID 179.A-2005 and on data products produced by CALET and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium. This work is based in part on observations made with the NASA/ESA *Hubble* Space Telescope, obtained from the Data Archive at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. Some of the data presented herein were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. This research is also partly supported by the Centre National d'Etudes Spatiales (CNES). These data were obtained and processed as part of the CFHT Large Area U-band Deep Survey (CLAUDS), which is a collaboration between astronomers from Canada, France, and China described in Sawicki et al. (2019, [MNRAS 489, 5202]). CLAUDS is based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the CFHT which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. CLAUDS uses data obtained in part through the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories, Chinese Academy of Sciences, and the Special Fund for Astronomy from the Ministry of Finance of China. CLAUDS uses data products from TERAPIX and the Canadian Astronomy Data Centre (CADC) and was carried out using resources from Compute Canada and Canadian Advanced Network For Astrophysical Research (CANFAR).

Authors contributed to the paper as follows: AM, HJMcC, PC, SG processed the imaging data; JW, OK, ID, MSh, BCH produced the photometric catalogs; JW, OI, GB produced the photometric redshifts and physical parameters catalogs; HJMcC, ST supervised this study. All these authors contributed to the validation and testing of the catalogs. The second group of authors (CL to ZL) covers those who have either made a significant contribution to assemble the data products or to the scientific analysis. The remaining authors (SA to GZ) contributed in a some way to the data products, conceptualization, validation, and/or analysis of this work.

data release

Both the Classic and The Farmer catalogs detailed in this work is publicly available in FITS format through the ESO Phase 3 System (https://www.

eso.org/qi/) and through servers at the Institut d'Astrophysique de Paris (https://cosmos2020.calet.org). Each catalog includes object positions, region mask flags, photometry, limited ancillary data (e.g., *HST*/ACS, GALEX), as well as photometric redshifts and physical parameters measured by both LePhare and EAZY, for each set of photometry. Four additional files will contain the redshift probability distributions for the two photometric catalogs in combination with both photometric redshift codes. Corresponding documentation will include information about the use of mask flags, and their respective regions. This dataset will also be made available through the IPAC-IRSA and CDS VizieR systems. Each catalog is a distinct item in the Digital Object Identifier (DOI) system: in work relying on COSMOS2020 data, the DOI name(s) should be cited, in addition to a reference to the present article, to keep track of which file(s) are actually used.

Science investigators who publish software analyzing these catalogs are encouraged to link their (e.g., github) repository to the website https://paperswithcode.com/; in this way the code will be also visible in the arXiv entry of the present publication under the section "Code & Data – Community Code".

3.A SOURCE DETECTION PARAMETERS

Table 5: `SExtractor` parameters used for the aperture detection and photometry.

Name	Value
ANALYSIS_THRESH	1.5
BACKPHOTO_THICK	30
BACKPHOTO_TYPE	LOCAL
BACK_FILTERSIZE	3
BACK_SIZE	128
BACK_TYPE	AUTO
CLEAN	Y
CLEAN_PARAM	1.0
DEBLEND_MINCONT	0.00001
DEBLEND_NTHRESH	32
DETECT_MAXAREA	100000
DETECT_MINAREA	5
DETECT_THRESH	1.5
DETECT_TYPE	CCD
FILTER	Y
FILTER_NAME	gauss_4.0_7x7.conv
GAIN	band-dependent
MAG_ZEROPOINT	band-dependent
MASK_TYPE	CORRECT
PHOT_APERTURES	13.33,20.00,47.33
PHOT_AUTOAPERS	13.3,13.3
PHOT_AUTOPARAMS	2.5,3.5
PHOT_FLUXFRAC	0.2,0.5,0.8
RESCALE_WEIGHTS	N
SATUR_LEVEL	30000
THRESH_TYPE	ABSOLUTE
WEIGHT_GAIN	N
WEIGHT_TYPE	MAP_WEIGHT,MAP_WEIGHT

3.B COMPARISON WITH REFERENCE PHOTOMETRY

The comparisons shown in Section 3.4 are here supplemented by comparing selected bands in this work to two well-known COSMOS-field literature catalogs for which this work is readily comparable: CANDELS (Nayyeri et al., 2017a, using UltraVISTA DR1 and IRAC/SPLASH) and COSMOS2015 (Laigle et al., 2016, using UltraVISTA DR2 and IRAC/SPLASH). As shown in Figure 3.21, broad-band K_s and IRAC channel 1 magnitudes and their colors are compared up to the depth limit of the shallower literature data set indicated by the vertical dashed line. For fairness, the sample includes only the ∼18 000 sources which are common to all three catalogs with 0.″6.

A brief analysis reveals three main points. Firstly, the COSMOS2020 depths in the bands considered exceed both those in CANDELS and COSMOS2015, as indicated by the vertical dashed and dotted lines, which manifests in the high scatter beyond the brightest magnitude limit. This restricts a meaningful comparison to sources below this limit. Secondly, the comparison with COSMOS2015 looks identical to the comparison of those bands between The Farmer and Classic, both in terms of offset and any trends with magnitude. This suggests that the Classic photometry is highly consistent with COSMOS2015, as verified directly during the catalog preparation process. Finally, the comparison of the The Farmer photometry with CANDELS is broadly similar. Although the K_s offset is larger than in comparisons with COSMOS2015 and Classic, the trend with magnitude in channel 1 is more constant than with either COSMSO2015 or Classic. The differences in K_s and channel 1 are similarly reflected in the colors, being more constant when comparing with CANDELS but not COSMOS2015. The similarity in the comparison with COSMOS2015 and Classic is expected, since both employed the same methodologies, by design. Similarly, the model-fitting employed in the IRAC photometry in CANDELS is more similar to that used by The Farmer and hence their agreement is unsurprising.

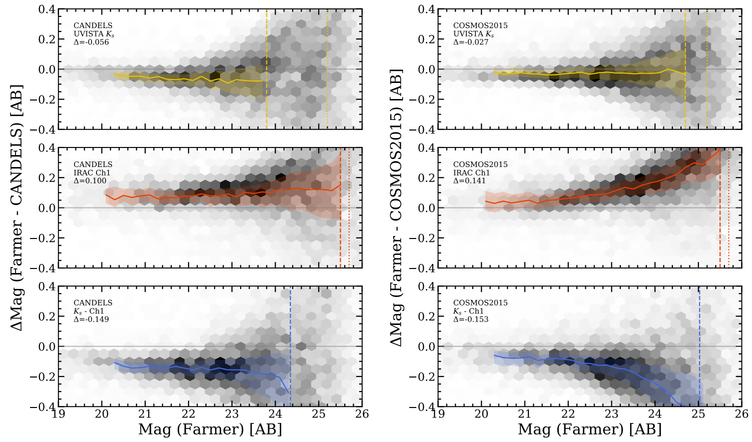

Figure 3.21: Comparison of broad-band K_s and IRAC channel 1 magnitudes and color between the The Farmer catalog of this work with those of CANDELS (Nayyeri et al., 2017a) and COSMOS2015 (Laigle et al., 2016). Individual sources are shown by the underlying density histogram which is described by the overlaid median binned by 0.2 AB with an envelope containing 68 % of sources per bin. For the magnitudes, depths are shown for the comparison sample (dashed) and for COSMOS2020 (dotted), corresponding to 3σ depths measured with 3″ diameter apertures. For colors, averaged 3σ depth computed from both bands of interest measured with 3″ diameter apertures. The median Δ magnitude offsets are reported for sources below the dashed magnitude limit.

4

THE GALAXY STELLAR MASS FUNCTION

The art of drawing conclusions from experiments and observations consists in evaluating probabilities and in estimating whether they are sufficiently great or numerous enough to constitute proofs. This kind of calculation is more complicated and more difficult than it is commonly thought to be.

- Antoine Lavoisier, *French Chemist, 1743-1794*

Adapted from:

COSMOS2020: the galaxy stellar mass function on the assembly and star-formation cessation of massive galaxies at $0.2 < z \leq 7.5$

Prepared for publication in Astronomy & Astrophysics, 2022

J. R. Weaver, I. Davidzon, S. Toft, O. Ilbert, K. M. L. Gould, C. K. Jespersen, C. Steinhardt, C. Lagos, P. L. Capak, N. Chartab, C. Laigle, O. B. Kauffmann, A. M. Koekemoer, B. Milvang-Jensen, H. J. McCracken, C. J. McPartland, B. Mobasher, A. Moneti, D. B. Sanders, M. Shuntov, and L. Zalesky

4.1 INTRODUCTION

The galaxy stellar mass function (SMF) is defined as the number density of galaxies $\Phi(\mathcal{M}, z)$ in bins of stellar mass $\Delta\mathcal{M}$, and is a fundamental cosmological observable in the study of the statistical properties of galaxies. Understanding its shape and evolution with cosmic time is a requirement to formulating a complete picture of the evolution of galaxies. The SMF ultimately informs us about the growth of the baryonic content of the universe, and can help infer star formation activity across cosmic time. Its integral over $\mathcal{M}$, the galaxy stellar mass density (SMD, or $\rho_*(z)$) describes the cumulative mass assembled by a given epoch. Their measurement has underpinned a number of pivotal discoveries related to the formation and evolution of galaxies.

Remarkable progress has been made since the first mass selected measurements of the local $z \approx 0$ SMF by Cole et al. (2001). The intervening years have been marked by order-of-magnitude increases in sample size, photometric precision, and redshift accuracy which have enabled more precise determinations of the local SMF as well as groundbreaking extensions to increasingly higher redshifts (Drory et al., 2009; Marchesini et al., 2009; Peng et al., 2010b; Pozzetti et al., 2010; Ilbert et al., 2013; Muzzin et al., 2013b; Vulcani et al., 2013; Grazian et al., 2015; Song et al., 2016; Davidzon et al., 2017; Wright et al., 2018; Adams et al., 2021; McLeod et al., 2021; Santini et al., 2021; Stefanon et al., 2021b).

A coherent picture has emerged from these studies. For instance, the shape of the SMF for star-forming galaxies is well-described by the empirically constructed Schechter Function (Schechter, 1976b), which features an exponential downturn at high masses at a cut-off characteristic stellar mass $\mathcal{M}^*$, with a low-mass end that declines with a slope α; both seeming to remain constant out to $z \approx 2$ (Marchesini et al., 2009; Ilbert et al., 2013; Tomczak et al., 2014; Davidzon et al., 2017; Adams et al., 2021). This constancy in the shape of the SMF implies that star formation activity has acted consistently over the past 10 Gyr to transform baryons from cold gas into stars, and appears to have formed $> 75\%$ of the total mass of the Universe in ≈ 10 Gyr. Only the normalization of the SMF $\Phi^*(z)$ has been confidently shown to evolve with cosmic time, and its rapid evolution at earlier times points directly to enhanced rates of mass growth at earlier epochs (Popesso et al., 2022, and references therein).

Which physical processes are responsible for the behaviour of the SMF, and the extent of their influence at different cosmic epochs, are not fully understood even in the low redshift universe ($z < 1$). Examples include feedback from supermassive black holes, supernovae, galaxy-galaxy mergers, stellar and gas dynamics, and gaseous inflows and outflows. As a result, significant effort has been undertaken to build detailed cosmological simulations to identify and understand the role of physical processes that underpin the observed shape and evolution of the SMF (e.g., Furlong et al., 2015; Lagos et al., 2018; Pillepich et al., 2018; Davé et al., 2019; Laigle et al., 2019; Lovell et al., 2021). Hence, the importance of obtaining precise measurements of the SMF and SMD is that they are key observables utilized in most (but not all, see e.g. Dubois et al. 2014) large-scale cosmological simulations to tune input parameters such that the SMF of the local universe ($z \approx 0$) is recovered. Comparisons of predicted and observed SMFs at earlier cosmic ages can point to the physical process at play and/or suggest recipes in need of refinement (Torrey et al., 2014; Furlong et al., 2015).

Owing to the challenges associated with large-scale simulations, purely analytical, data-driven models have enjoyed great popularity with the introduction of the first large galaxy samples with photometric redshifts (photo-z). For example, Peng et al. (2010b) constructed a phenomenological model to explain the bimodality of galaxy types (star-forming and quiescent) as seen from their mass functions, Mpc-scale environment, and star formation activity. That is, as a consequence of two mechanisms driving star formation cessation (often referred to as "quenching"): massive galaxies cease forming stars irrespective of environment ("mass quenching"), and galaxies in dense environments cease forming stars irrespective of mass ("environmental quenching"). Several discrete mechanisms have been proposed to explain the latter, such as ram pressure stripping (e.g., Gunn et al., 1972), gas strangulation (e.g., Larson et al., 1980; Balogh et al., 2000), and dynamical heating of gas within haloes (e.g., Gabor et al., 2015). Proposed mechanisms for the former must, by definition, involve secular processes. This includes star formation cessation due to structural changes, or heating and/or ejection of gas by central super-massive black holes for the most massive galaxies (radiative AGN feedback), or by supernovae for less massive systems as they are more weakly self-gravitating. Peng et al. (2010b) hypothesize that while environmental effects reproduce the single Schechter shape observed for star-forming, blue galaxies in the local universe, it is through a combination of both environmental and mass effects which reproduce the two-component Schechter shape observed for quiescent, red galaxies. A wide variety of extensions have been applied to this model to directly incorporate other measurements such as gas fraction (Bouché et al., 2010), and wholly new models continue to be developed (e.g., Peng et al., 2015; Belli et al., 2019; Suess et al., 2021; Varma et al., 2021).

Directly observed measurements of stellar mass and redshift are not in themselves the intrinsic, underlying SMF. Instead, they are bridged by statistical inferences between carefully selected samples and a vast, practically unobservable parent population from which a sample is taken. Insufficient control of biases and systematics obscure such inferences by creating unrepresentative samples, and hence weaken comparisons with simulation and analytical models (Fontanot et al., 2009; Marchesini et al., 2009). Great effort has been expended over the past 20 years to strip away a number of these biases, and their obscuring effects (Cole et al., 2001; Marchesini et al., 2009; Pozzetti et al., 2010; Davidzon et al., 2017; Leja et al., 2019a; Adams et al., 2021). Examples of these biases include sample incompleteness (including Malquist biases and mass completeness), the so-called "cosmic" variance relating to sampling over-and under-dense regions of large scale structure, and Eddington bias which acts to overestimate the number density of the most massive galaxies (Eddington, 1913; Malmquist, 1920, 1922). While many studies incorporate only poisson noise (Fontanot et al., 2009, although this is steadily improving with time), other notable uncertainties exist including sample variance arising from poisson noise and effective volume size, as well as uncertainties on photo-z and stellar mass; the latter two items may also introduce bias. Since the ultimate goal of any survey is to generalize the observed properties of a specific sample to that of the entire population, ignoring any of these important biases or uncertainties severely complicate this effort.

Studies of the SMF to increasingly higher redshift have been made possible due to advances in facilities, and continued investment in deep, primarily photometric surveys of the distant universe. Building on work by Songaila et al. (1994) and Glazebrook et al. (1995), Cowie et al. (1996) secured the first mass-complete samples at $z \approx 2-3$, selected in the near-infrared (K) that directly constrains the stellar bulk at $1-2\mu$m. More recently, precise mass estimates from similarly selected samples have been obtained by fitting observed spectral energy distributions (Tomczak et al., 2014; Straatman et al., 2015; Martis et al., 2016), from the ground (with VISTA, UKIRT) and space (*HST*/WFC3, *Spitzer*/IRAC). This strategy enables a higher degree of mass completeness, and in turn provides a more representative understanding of the high-z universe unrestrained by the selection biases which plagued forerunner surveys. Although limited in area, samples recovered by *HST* have continued to pose new and significant challenges to existing paradigms by highlighting a series of stark changes in the shape of the SMF that indicates earlier galaxy populations were fundamentally different to those in the present day universe. Additional studies at these early times ($z \gtrsim 2$) by degree-scale surveys capable of finding the rarest sources have revealed the existence of surprisingly mature, massive quiescent galaxies whose mass has already been assembled by $z \sim 4$ (e.g., Ilbert et al., 2013), challenging the assumed timeline typical for galaxy assembly (Steinhardt et al., 2016b; Behroozi et al., 2018a). Limited numbers of them have been confirmed by detailed spectroscopic follow-up (e.g., Schreiber et al., 2018a; Tanaka et al., 2019; Valentino et al., 2020a). Likewise, several studies have placed constraints on the massive end of the SMF at the highest-redshifts ($z > 6$) through the measurement of rest-frame UV continuum luminosity (e.g., Harikane et al., 2016; Song et al., 2016), who utilize empirically calibrated $L_{\rm UV} - \mathcal{M}$ relations to estimate stellar mass from UV luminosity.

Although tantalizing, deriving these UV-selected mass estimates involves significant uncertainties; and so until the operational deployment of the James Webb Space Telescope (*JWST*) expected in 2022, there will exist a dearth of facilities capable of deep infrared spectroscopy to permit the observation of the rest-frame stellar bulk ($\lambda > 4000$Å, although ideally $\sim 1\mu$m) required to more directly assess stellar mass at these redshifts. Only recently have deep ($K_s \approx 25$) near-infrared photometric studies enabled the first mass-selected samples at $z > 2$, although with mass uncertainties increasing with redshift (Retzlaff et al., 2010; Fontana et al., 2014b; Laigle et al., 2016). Importantly, photometric surveys of galaxies has – and is expected to remain – the primary means by which these measurements will be made; obtaining even elementary parameters for large ($N > 100\,000$) samples with deep spectroscopy, while precise, is simply too expensive even when utilizing highly multiplexed instrumentation. For the time being, spectroscopy of the distant universe will remain a follow-up exercise to strengthen larger photometrically-selected samples.

Of these deep photometric surveys, those with large areas have a key advantage: they probe a wider range of environments (i.e., density) compared to more narrow, so-called 'pencil-beam' surveys. As such, they provide more representative samples that are significantly less likely to be affected by field-to-field "cosmic" variances (and, for the same reason, are more suited to find rare, massive galaxies). Although one may resort to combining disparate data sets into a single analysis to combat this (e.g., Moster et al., 2013; Henriques et al., 2015; Volonteri et al., 2016; McLeod et al., 2021), the unknown systematics between survey selection functions, filter profiles, and depths make the interpretation of these joined samples more uncertain.

Stitching together surveys of low- and high-z samples carries similar concerns. A lack of uniform selection owing to different survey areas, detection bands, and photo-z determination strategies (i.e., "dropout" selection) can likewise complicate interpretations arising from such multi-component samples (e.g., Leja et al., 2019a; Adams et al., 2021; McLeod et al., 2021). Worse, these photometric samples may have been processed with different SED fitting codes which assume different templates, emission line recipes, and dust attenuation laws. Differing choices of cosmology and initial mass function, while reversible, nonetheless add complexity.

Accurate estimates of the galaxy stellar mass function at increasingly higher redshifts requires complementary deep observations from near-infrared selected samples to ensure both mass completeness and data reliability. Currently the widest near-infrared selected survey with the requisite depth to probe large samples of $z \gg 3$ galaxies is the Cosmic Evolution Survey (COSMOS; Scoville et al., 2007b). This work takes advantage of the latest NIR-selected catalog of the COSMOS field, COSMOS2020 (Weaver et al., 2022a). We use profile-fitting photometry from The Farmer paired with photometric redshifts, stellar mass estimates, and rest-frame magnitudes from `LePhare`. From this data we construct a consistently measured SMF from $z = 0.2 - 7.5$, identify quiescent and star-forming systems, and study the build-up and assembly of stellar mass over 10 Gyr of cosmic history. This includes a detailed study of key moments in the development of galaxy populations from the Epoch of Reionization ($z > 6$) to the peak of star formation activity at Cosmic Noon ($z \sim 2$), up to the rich bimodality of star-forming and quiescent galaxies at the present day.

This paper is organized as follows. Section 4.2 introduces the dataset chosen for this analysis from which samples are drawn and possible uncertainties discussed in Section 4.3. Section 4.4 provides a brief overview of the Schechter Function. Results are presented in Section 4.5, including the presentation and fitting of the total, star-forming, and quiescent mass functions. These are compared to literature measurements in Section 4.6, whereupon further discussion is had with regards to galaxy assembly, star formation cessation, and connections to dark matter halos. This work concludes in Section 4.7.

These results are computed adopting a standard ΛCDM cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\mathrm{m},0} = 0.3$ and $\Omega_{\Lambda,0} = 0.7$ throughout, such that the dimensionless Hubble Parameter $h_{70} \equiv H_0/(70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}) = 1$. Galaxy stellar masses, when derived from SED fitting, scale as the square of the luminosity distance (i.e., D_L^2); hence a factor of h_{70}^{-1} is retained implicitly for all relevant measurements, unless explicitly noted otherwise (see Croton 2013 for an overview of h and best practices). Masses computed from `LePhare` assume a Chabrier (2003) Initial Mass Function. All magnitudes are expressed in the AB system (Oke, 1974), for which a flux f_ν in μJy (10^{-23} erg cm^{-1}s^{-1}Hz^{-1}) corresponds to $\mathrm{AB}_\nu = 23.9 - 2.5\,\log_{10}(f_\nu/\mu\mathrm{Jy})$.

4.2 DATA: COSMOS2020

The most recent release of the COSMOS catalog is COSMOS2020 (Weaver et al., 2022a), comprised of $\sim 1\,000\,000$ galaxies selected from a near-infrared $izYJHK_s$ `CHI-MEAN` image (Szalay et al., 1999a; Bertin, 2010b), with near-infrared depths approaching 26 AB required to ensure a mass-selected sample complete down to $10^9\,\mathcal{M}_\odot$ at $z \approx 3$. This sample is complemented by extensive supporting photometry ranging from the UV down to 8μm over an area of 2 deg^2, making it the widest near-infrared selected multi-wavelength catalog with this depth. This is made possible by the latest release of UltraVISTA survey (DR4; McCracken et al., 2012; Moneti et al., 2019), which is the longest running near-infrared survey to date, and complemented by even deeper optical *grizy* data provided by Subaru's Hyper Suprime-Cam instrument (HSC PDR2; Aihara et al., 2019). Flanking this core imaging are u band measurements made by the CLAUDS survey from the Canada-France-Hawaii Telescope (Sawicki et al., 2019) and deep Spitzer/IRAC imaging in all four channels, reprocessed to include nearly every exposure taken in COSMOS for use in this catalog (Moneti et al., 2021). As in the previous COSMOS catalog by Laigle et al. (2016), several intermediate and narrow bands from both Subaru/Suprime-Cam and VISTA (Milvang-Jensen et al., 2013) are used to provide precise determinations of photometric redshifts.

In addition to the comprehensive set of data, photometry is performed by two methods: aperture photometry with SExtractor (Bertin et al., 1996) for the optical/near-infrared and IRACLEAN (Hsieh et al., 2012) for the infrared (as in Laigle et al.) and with The Farmer (Weaver et al. 2022, in prep.), which uses The Tractor (Lang et al., 2016a) to construct and apply parametric models to extract source fluxes through profile-fitting consistently in more than 30 bands. Each photometric catalog is paired with photometric redshift and physical parameter estimates from both `LePhare` (Arnouts et al., 2002; Ilbert et al., 2006) and `EAZY` (Brammer et al., 2008) resulting in 4× measurements of photo-z and stellar mass per source. The combination of comprehensive deep images, independent photometry techniques, and independent SED fitting enables an unparalleled control of systematics and uncertainties, and hence further insight when inferring the underlying true nature of galaxy populations. See Weaver et al. (2022a) for details.

4.3 Sample selection, completeness, and bias

4.3.1 *Selection function*

Galaxies in COSMOS2020 are selected from a near-infrared $izYJHK_s$ `CHI-MEAN` image (Szalay et al., 1999a; Bertin, 2010b). The deepest band is i, with a 3σ sensitivity limit at 27.0 mag, and the shallowest is K_s at 25.9 mag[1]. This presents a significant improvement over COSMOS2015 as the comparably deeper K_s imaging translates directly into a higher degree of completeness in r.f. Balmer continuum flux across lookback time. Because of this, low-z samples are more mass complete down to lower masses as the rest-frame stellar bulk is directly measured. At higher z however, the deeper NIR and IRAC imaging equates to reliable detections of the reddest, oldest, and heavily dust obscured sources compared to COSMOS2015.

Although the nominal area of the COSMOS survey is 2 deg^2, the most secure region comprises 1.279 deg^2 after removing contamination due to bright star halos and requiring a union of the deep near-infrared UltraVISTA imaging and the Subaru Suprime-Cam intermediate bands, where the photo-z performance is generally best. Compared to Davidzon et al. (2017) who use UltraVISTA DR2 (McCracken et al., 2012) imaging to probe the near-IR, here we rely on UltraVISTA DR4 (Moneti et al., 2019), reaching greater and simultaneously uniform NIR depth across the entire field (especially K_s). These non-uniform depths presented a severe limitation in Davidzon et al., who restricted their stellar mass functions to only the deepest stripes in UltraVISTA spanning only 0.62 deg^2. The gains leveraged in this work are hence five-fold. Firstly, the total number of recovered sources at all apparent magnitudes increases due to the additional area, improving statistical margins especially of rare sources. Secondly, the increased optical and near-infrared depths permit the recovery of fainter sources, improving flux completeness. Thirdly, the consistent depth and more controlled imaging from Hyper Suprime-Cam provides significantly more precise, deeper, and less biased photometry relative to Suprime-Cam. Forth, the increased depth in both optical and near-infrared imaging provides more accurate, less biased photo-z and $\mathcal{M}$. Lastly, the wider field makes it less likely to become biased due to specific structures along the line-of-sight (e.g.,

1 Computed at 3σ from $3''$apertures randomly placed in regions without detected sources; see Table 1 of Weaver et al. (2022a)

clusters, voids) and instead probes a greater variety of environments affecting the evolution of galaxies within them.

As presented in Fig. 13 of Weaver et al. (2022a), the best photo-z performance at $i > 22.5$ is achieved by the combination of The Farmer and `LePhare` with a precision of $< 1\%$ at $i \sim 20$ and $< 4\%$ at $i \sim 26$. For this reason, and to expedite comparisons with similar work in the literature (Ilbert et al., 2013; Davidzon et al., 2017), this work adopts the photometry from The Farmer paired with the photo-z, stellar masses ($\mathcal{M}$), and rest-frame magnitudes estimated by `LePhare`. This combination will be henceforth referred to as COSMOS2020, unless explicitly stated otherwise. We choose to adopt $\mathcal{M}$ estimates from `MASS_MED` for our masses as the median of the mass likelihood distributions are generally less susceptible to template fitting systematics than `MASS_BEST` taken at the minimum $\chi^2_{\rm SED}$. We find that the two agree at all redshifts within 0.01 dex with a narrow scatter whose 68% range is below 0.05 dex.

Identification of stellar contaminants proceeds as described in Section 5.1 of Weaver et al. (2022a). Briefly, `LePhare` computes the best-fit stellar template to each SED from a range of templates, including white and brown dwarfs. Sources that achieve a χ^2 from a stellar template fit lower than any galaxy template are removed. An additional criterion on morphology is used, identifying likely stars as point-like sources from the COSMOS *HST*/ACS mosaics (Koekemoer et al., 2007b) and Subaru/HSC (PDR2, Aihara et al., 2019) for the $i < 23$ and $i < 21.5$ AB, respectively. Bright, resolved sources likely to be galaxies are not removed.

An initial sample of 638 423 galaxies[2] in the range $0.2 < z \leq 6.5$ is selected from the contiguous $1.27\,\mathrm{deg}^2$ `COMBINED` region combining the UltraVISTA and Subaru/SC footprints but removing regions around bright stars[3]. However, the source density at $6.5 < z \leq 7.5$ becomes noticeably greater in the ultra-deep stripes and so we restrict sources in this particular z-bin to only the $0.716\,\mathrm{deg}^2$ of the ultra-deep stripes also covered by Subaru/SC but away from bright stars, finding 1 327 galaxies. The construction of a robust stellar mass function requires precise redshifts and accurate stellar masses and so 57 515 sources with ambiguous redshifts are removed, quantified by having more than $> 32\%$ of their redshift probability lying outside $z_{\rm phot} \pm 0.5$. Since an IRAC channel 1 detection is required to accurately measure galaxy stellar masses at $z \gtrsim 2-3$, where this study seeks to cover new ground, an additional 184 292 sources with $m_{\rm ch1} > 26$ AB (i.e., $S/N \lesssim 5$) are removed. Finally, we remove a further 2 037 sources with unreliable SED fits $\chi^2 > 10$. These three necessary cuts[4] create a final sample of 395 906 galaxies with precise redshifts and accurate stellar masses.

As such, this work presents the deepest near-infrared selected galaxy stellar mass function studied over a single contiguous field spanning $1.27\,\mathrm{deg}^2$ and is supported by some of the most precise photo-z achieved to date.

2 `lp_type=0`
3 `FLAG_COMBINED=0`
4 There is considerable overlap sources removed these three criteria, see Figure 4.18.

4.3.2 *Quiescent galaxy selection*

Accurate identification of quiescent galaxies has been a longstanding problem (see e.g. Leja et al., 2019b; Shahidi et al., 2020; Steinhardt et al., 2020b). While quiescent galaxies can be identified by their low star formation rates, SFR estimates from fitting broad-band photometry are subject to large uncertainties and are dependent on the particular SED template library (Pacifici et al., 2015; Carnall et al., 2019a; Leja et al., 2019a, Pacifici et al. 2022, submitted.). The other most successful approach is that of rest-frame colors, as the lack of blue O- and B-type stars in quiescent populations implies a dearth of UV emission and a red optical contintuum. Similar photometric SEDs are measured for heavily dust-enshrouded but otherwise star-forming systems, making the challenge of color-color selection that of distinguishing truly quiescent galaxies from dusty star-forming ones. Popular rest-frame color-color selections include $(U-V)$ vs. $(V-J)$, $(NUV-r)$ vs. $(r-J)$, $(NUV-r)$ vs. $(r-K)$, and $(FUV-r, r-J)$ (Williams et al., 2009; Ilbert et al., 2010; Arnouts et al., 2013a; Leja et al., 2019b, respectively). However, each implies a different definition of quiescence (see Leja et al., 2019b; Shahidi et al., 2020). These pairs of rest-frame colors are well measured only when there exists densely sampled observed frame photometry over the range from UV to IR with sufficient depth, and in those cases it may be measured directly from the observed photometry. However, they can also be derived from the best-fit templates which should largely agree with observed-band photometry, and can also be meaningfully extrapolated in cases where appropriate observed-frame data are not available. While the former may be adversely affected by uncertainties and systematics propagated from the observed photometry, the latter assumes that the SED template library or basis contains a suitable model. It is worth noting that efforts are being made to employ machine learning techniques to lessen the dependence and bias impact from model assumptions (e.g., Steinhardt et al., 2020b), although no such method has gained comparable use as yet.

Recent work by Leja et al. (2019b) has clarified the efficacy of these color-color selections, proposing to extend the typically adopted $U-V$ vs. $V-J$ baseline further into the UV on the basis of stronger correlation with intrinsic sSFR as measured in simulated photometry (see also Arnouts et al., 2013a). As such, catalogs containing reliable and deep FUV and NUV photometry have the advantage of identifying low-z quiescent populations that can be consistently identified up to the highest redshifts even when the rest-frame U band is no longer measurable in the observed-frame. Although the rest-frame J band is crucial for measuring the rest-frame stellar bulk, it becomes easily redshifted out of most deep surveys with IRAC photometry by $z \sim 2-3$, thus rendering quiescent selections highly dependent on template libraries even at these intermediate redshifts.

This work selects quiescent galaxies following the criteria introduced by Ilbert et al. (2013):

$$(NUV - r) > 3\,(r - J) + 1 \text{ and } (NUV - r) > 3.1 \tag{4.1}$$

whereby the slant line runs perpendicular to the direction of increasing sSFR and parallel to an increase in dust attenuation to separate truly quiescent systems from otherwise dusty star-forming contaminants. Generally, this cut has been found to approximate a cut in sSFR $\lesssim 10^{-11}\,\mathrm{yr}^{-1}$ (Davidzon et al., 2018).

It is important to note that `LePhare` computes rest-frame magnitudes by attempting to find an observed frame band which directly probes the rest-frame band, which for the reasons listed above is less model-dependent than simply adopting the rest-frame colors of the best-fit galaxy template and hence conveys a view of the true variety of observed galaxies that is less biased by our assumptions. Although the rest-frame photometry are estimated using a K-correction and color-term, they are most strongly dependent on the best-fit model at redshifts where observed photometry are not well matched the rest-frame band (see Appendix 1 of Ilbert et al. 2005 for details). As such, the most readily comparable studies of quiescent galaxies are those from previous COSMOS catalogs: Ilbert et al. (2013) and Davidzon et al. (2017).

Unbiased sample selection becomes increasingly difficult with redshift. By $z \approx 3$, rest-frame J-band fluxes are no longer directly measured even by IRAC channel 2[5] and so become increasingly model-dependent and uncertain. While rest-fame NUV remains constrained even at $z > 6$, rest-frame r must be extrapolated by $z \approx 5$ whereby differentiation between quiescent and dusty star-forming galaxies becomes statistically impossible. Because of this, selection of quiescent galaxies in this work is limited to $z \leq 5.5$, noting that selection between $3 < z \leq 5.5$ is subject to a significantly higher degree of uncertainty. Advancement in this selection will only be made possible with deeper infrared imaging provided by future facilities (e.g., *JWST*).

Selection of quiescent galaxies is presented in Fig. 4.1, showing star-forming and quiescent galaxies whose masses are above their respective mass limits (see Section 4.3.3). Error bars on the rest-frame colors are estimated from the quadrature addition of the observed-frame filter nearest to that of the rest-frame. Given that even rest-frame photometric uncertainties are generally inversely proportional to mass at given redshift, these error bars are most representative for median mass systems but are overestimated for bright, massive galaxies. In addition, uncertainties become increasingly underestimated at $z > 3$ as the extrapolation based on the best-fit template is not propagated. However, it is already obvious that the dividing power of the slant line is significantly diminished at $z > 3$ where the rest-frame J band is no longer directly observed. Both the quiescent and dusty star-forming regions appear to be devoid of sources by $z \approx 6$, such that no intrinsically quiescent (misclassified or not) are found at these early times. Whether this is an intrinsic feature of galaxy populations or merely a selection effect is discussed below in Sec-

5 Channel 3 and 4 are too shallow to provide useful constraints at $z \gtrsim 3$

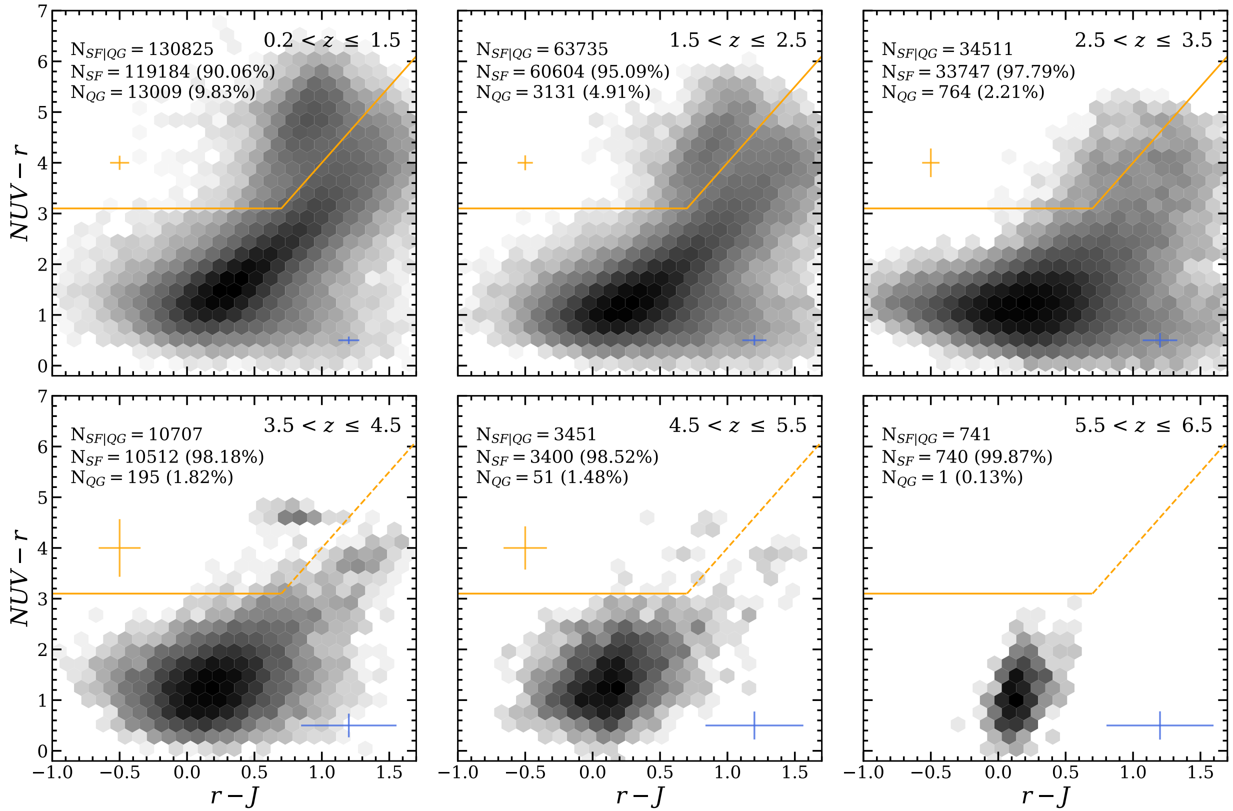

Figure 4.1: Galaxies are classified as either star-forming or quiescent in bins of redshift by selection in rest frame NUV $- r$ and $r - J$ colors, following Ilbert et al. (2013). The rest-frame J band is not directly measured at $z \gtrsim 3.5$ and so the sloped part of the selection box is dashed to indicate that the $r - J$ colors are strongly model dependent. Typical uncertainties for the rest-frame colors of the star-forming and quiescent subsamples are estimated as medians of the nearest observed-frame bands, consistent with the derivation of the rest-frame colors. See the text for details.

tion 4.3.3. For these reasons this work restricts distinction between quiescent and star-forming galaxies to $z \lesssim 5.5$. Galaxies at $z \gtrsim 5.5$ should be considered as star-forming.

4.3.3 *Completeness*

Accurate estimates of completeness are crucial when inferring general properties about a population from an otherwise incomplete sample. While advancements in near-infrared facilities have enabled breakthroughs in selecting representative sample of galaxies by measuring their stellar bulk, samples are still mass limited and these mass limits evolve with redshfit owing to the faintness of increasingly distant galaxies. As such, mass limits are highly dependent on accurate estimates of survey depths and their impact on the selection function as it relates to the detection of the lowest mass galaxies.

There are three known populations which are expected to be missed by a near-infrared selection whose deepest images are on the blue-end of the selection function.

Firstly, the faintest star-forming blue galaxies in the local universe (e.g., dwarf irregular starbursts) may have detectable fluxes blueward of i, they will not be included in existing $izYJHK_s$ selections as their near-infrared emission is too faint. However, their contribution is expected to be limited to only the low-mass end of the galaxy stellar mass function, and can henceforth be well-characterised in large numbers by deeper 'pencil-beam' surveys (e.g., CANDELS, Grogin et al., 2011; Koekemoer et al., 2011).

Secondly, quiescent galaxies are generally more difficult to detect than star-forming galaxies owing to a lack of rest-frame UV and blue optical emission, meaning their detection relies upon deep near-infrared imaging at both low and high-redshifts. For this reason, an $izYJHK_s$ detection will be insufficient to detect the quiescent systems compared to star-forming ones of the same mass, meaning that the mass completeness limit of the quiescent sample will naturally be lower than that of the star-forming sample. We compute their respective mass completeness limits separately and apply them consistently throughout.

Lastly, and similar to quiescent galaxies, the most heavily dust obscured star-forming galaxies ($A_V \gg 5$) at high redshift ($z \gtrsim 2$) will not present any appreciable optical or near-infrared fluxes to be detected in COSMOS2020, but unlike quiescent galaxies are ubiquitously and efficiently detected in far-infrared, sub-millimeter, and radio surveys (e.g., Schreiber et al., 2018c; Jin et al., 2019; Wang et al., 2019; Fudamoto et al., 2020; Casey et al., 2021; Fudamoto et al., 2021; Sun et al., 2021; Shu et al., 2022). The nature and extent of this recently discovered population remains difficult to quantify, owing to a combination of complex selection functions and serendipitous detections. They are likely to be missed by an $izYJHK_s$ selection function even at the depths of COSMOS2020 (see discussions in Section 4.6.3).

Although mass completeness estimates are presented in Weaver et al. (2022a), they are derived for a comparably less secure sample than used in this work. For this reason the mass completeness limits are re-derived in identical fashion following the method of Pozzetti et al. (2010). A critical advantage of this method is that it does not rely on the theoretical mass distribution of galaxies fainter than the magnitude limit, but assumes that those just above the threshold share similar properties with the undetected ones, modulo a rescaling factor as detailed below. This contrasts with studies that estimate mass completeness either through injection-recovery of simulated sources, or extrapolation of the observed distribution itself below the magnitude limit. The latter approach, in particular, may underestimate the stellar mass limit if the galaxy sample is sparse due purely to astrophysical reasons and not genuine incompleteness. In our case, the sample in each z-bin is first cleaned by discounting the top 1% worst-fit sources via χ^2. The stellar masses $\mathcal{M}$ of the 30% faintest galaxies in channel 1 are then re-scaled such that their observed channel 1 apparent magnitude $m_{\rm ch1}$ matches the IRAC sensitivity limit $m_{\rm lim} = 26$:

$$\mathrm{Log}_{10}(\mathcal{M}_{\rm resc}) = \mathrm{Log}_{10}(\mathcal{M}) + 0.4\,(m_{\rm ch1} - 26.0)\,. \tag{4.2}$$

The limiting mass $\mathcal{M}_{\rm lim}$ is then taken to be the 95^{th} percentile of the $\mathcal{M}_{\rm resc}$ distribution. Finally 2$^{\rm nd}$ order expansions in $(1+z)$ are fitted to each $M_{\rm lim}$ per z-bin such that a mass complete sample may still be precisely identified even in the absence of z-bins. Limits above which samples are $\sim 70\%$ mass complete (see justification below) are derived consistently for the total sample from $0.2 < z \leq 7.5$ and for the star-forming, and quiescent samples from $0.2 < z \leq 5.5$:

$$\text{Total}: -\,3.61\times10^{7}\,(1+z) + 7.95\times10^{7}\,(1+z)^2 \tag{4.3}$$

$$\text{Starforming}: -\,5.77\times10^{7}\,(1+z) + 8.66\times10^{7}\,(1+z)^2 \tag{4.4}$$

$$\text{Quiescent}: -\,3.79\times10^{8}\,(1+z) + 2.98\times10^{8}\,(1+z)^2 \tag{4.5}$$

and shown in Fig. 4.2. Despite the more conservative selection adopted in this work, the derived mass completeness limits are essentially identical to those derived in Weaver et al. (2022a), which indicates the robustness of these limits against sample selections.

There remains an additional incompleteness arising from the fact that mass completeness is derived from IRAC channel 1 photometry, and yet our $izYJHK_s$ selection function does not include sources which are only identified in IRAC. As discussed in Davidzon et al. (2017), despite this drawback it is also disadvantageous to any one of the six $izYJHK_s$ bands used in the catalog. The reddest, K_s, samples the rest-frame stellar bulk only out to $z \lesssim 2$, and will tend to underestimate stellar masses at $z \gtrsim 2$. Thankfully, it is possible to estimate the impact of this additional incompleteness by examining a sample of galaxies common to this work and those of the comparably deeper $\sim$200 arcmin^2

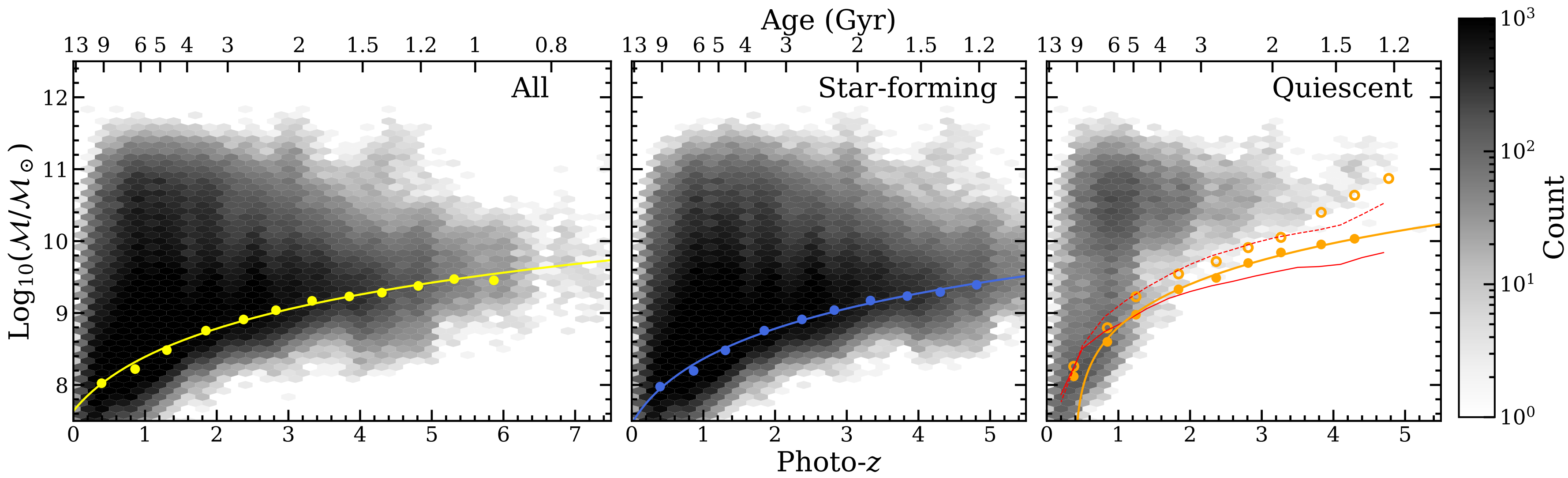

Figure 4.2: The three panels show, from left to right, the density of the total, star-forming, and quiescent samples with their respective mass completeness limits. The limits are derived in discrete z bins following Pozzetti et al. (2010) with magnitude limits adopted from IRAC channel 1 (filled circles), then interpolated with Eqs. 4.3, 4.4, and 4.5, respectively, resulting in the solid lines. Consistent estimates using K_s are also shown (empty circles). Conservative estimates of stellar mass from simulated spectra assuming a delta-like burst at $z = 15$ and normalized to the K_s and channel 1 magnitude limits are shown by the dotted and solid red lines, respectively. See the text for details.

CANDELS-COSMOS catalog of Nayyeri et al. (2017b). This analysis is performed and discussed at length in Weaver et al. (2022a). In summary, 75% of CANDELS sources at $M_{\rm lim}$ are recovered by both The Farmer and Classic, which combined with the choice of $M_{\rm lim}$ being the 95^{th} percentile of $M_{\rm resc}$ implies that $M_{\rm lim}$ of the total sample corresponds to a mass completeness of $\sim 70\%$.

As shown in Fig. 4.2, both the mass limits of the total sample are almost identical with those of the star-forming sample at all redshifts. This is unsurprising as star-forming galaxies generally dominate galaxy demographics. As in our $NUVrJ$ selection from Fig. 4.1, we see the lack of quiescent systems at $z > 5$ and so we consider the star-forming subsample to be statistically equivalent to the total sample at $z > 5$.

Deriving a consistent mass completeness limit for quiescent galaxies is more challenging. It is well known that the predominantly older, redder stellar populations of quiescent systems imply higher mass-to-light ratios than found in star-forming galaxies which in turn imply a lower degree of mass completeness at the same flux limit. Fig. 4.2 shows that our channel 1 derived mass completeness falls $\gtrsim 0.2$ dex lower than the bulk of the quiescent sources. Taken at face value, this would appear to indicate a lack of quiescent systems at low-intermediate masses at $z > 2$. Again, we have cause for concern as channel 1 is not in our selection function, and because of the predominantly red optical spectral slopes of quiescent galaxies, continuum emission in K_s should be lower than in channel 1, implying that one would need a deeper K_s magnitude limit to detect the same sources from shallower channel 1 imaging. In other words, it is expected that there are red sources visible only in channel 1 which are not included in our $izYJHK_s$ selection function. However, even K_s becomes more sensitive to star formation and by $z \approx 3.5$ no longer measures Balmer continuum flux in the rest-frame but rather the comparably less luminous UV Lyman continuum (channel 1 does so by $z \approx 8$). Worse, even the Lyman continuum in these systems is expected to be faint (even given 'frostings' of star formation in $NUVrJ$-selected post-starburst systems), making any assessment of stellar mass from K_s at $z > 3.5$ unreliable. Therefore, while deriving a mass completeness from channel 1 magnitude limits is not appropriate due to possible selection effects with regards to a $izYJHK_s$-selected sample, we cannot turn to K_s as it is no longer a mass indicator at $z > 3.5$.

One solution is modify our selection function by incorporating $izYJHK_s$ undetected IRAC-only sources into our sample. While a systematic search for IRAC only sources is ongoing (Blasquez et al., in prep.), their identification is made difficult not only because of the significantly lower resolution and consquently higher source crowding in necessarily deep IRAC images, but these sources also lack optical/NIR data which is, by definition, insufficiently deep to identify low-z interlopers. Deep MIR data redward of IRAC do not currently exist, making a determination of redshift, mass, and quiescent nature of these IRAC-detected sources problematically uncertain.

For the time being, whether or not this absence of intermediate-low mass quiescent systems is astrophysical cannot be determined. Literature measurements do not yield more mass-complete quiescent samples as the UltraVISTA DR4 NIR depths are now similar to those from even the deepest small-field NIR imaging, $H_{160} \approx 25.9$ (e.g. CANDELS, Tomczak et al., 2014), and even so, comparisons are hampered by field-to-field and photometric systematics. Comparisons with other stellar mass functions measured more consistently in COSMOS (e.g. Ilbert et al., 2013; Davidzon et al., 2017) are still affected by systematics from comparably less certain measurements from previous, shallower data, despite the lessened impact of cosmic variance. Additionally, comparisons with Davidzon et al. (2017) are complicated by the fact that they adopt different methodologies for measuring masses at $z < 3$, which are taken directly from Laigle et al. (2016), and at $z \geq 3$, which are computed following a modified version of `LePhare` which produces lower number densities of massive galaxies compared to Laigle et al. directly, as demonstrated by a recent comparison by Lustig et al. (2022).

We turn to theoretical frameworks to investigate this further. Namely, we use Flexible Stellar Population Synthesis (`FSPS`; Conroy et al., 2009b, 2010b) to estimate the stellar masses of model spectra normalized to match our K_s and channel 1 magnitude limits (25.5 and 26.0 mag, respectively). We assume a Charbier IMF with $\log_{10}(Z/Z_\odot) = -0.3$. For a given redshift, we derive observed-frame K_s and channel 1 magnitudes corresponding to the most conservative (least mass complete) scenario resulting from an extraordinarily old stellar population produced by a delta-burst evaluated at $z = 15$. As shown in by the red solid (dashed) curves in Fig. 4.2, this conservative scenario defines an upper limit for mass completeness as derived by channel 1 (K_s). While they agree with the estimates using the method of Pozzetti et al. at $z \lesssim 2$, they favor a higher degree of mass completeness at $z \gtrsim 2$. If we reduce our estimate for the channel 1 magnitude limit by 0.5 to 25.5 mag, we find that the channel 1 mass completeness as predicted by our delta-burst SFH aligns with our channel 1 mass limit. This suggests that our estimates based on Pozzetti et al. are at least consistent with simple theoretical predictions. We caution however, that the assumption that these models accurately describe real high-z quiescent galaxies is becoming increasingly dubious. Such a system formed in a monolithic delta-burst just following the big bang cannot reach quiescence (as defined by $NUVrJ$) above $z \approx 5.3$, and yet remarkably mature systems at $z \approx 4 - 5$ have already been reported in the literature (e.g., Schreiber et al., 2018a; Tanaka et al., 2019; Valentino et al., 2020a). Worse, the spread of mass-to-light ratios found in quiescent systems means that that a single mass completeness limit for all quiescent galaxies at a given redshift is ill-defined even for consistently selected (e.g., UVJ, $NUVrJ$, $sSSFR$) samples, resulting in a non-negligible selection effects.

Nonetheless, we stress that the difference in completeness between the effectively flux-complete K_s-derived mass limit and the mass limit derived from channel 1 magnitudes is only ~ 0.3 dex, which is typically less than a single bin in our analysis. In light of this considerable uncertainty, we adopt the optimistic quiescent galaxy mass completeness limits derived via Pozzetti et al. (2010) from channel 1 magnitudes with the caveat that the lowest mass bins in each measurement at $z > 3$ are potentially incomplete. We indicate the K_s mass limit for quiescent samples throughout. Also, we note that our quiescent mass limit, although consistently determined in Davidzon et al. (2017), is more comparably more conservative despite the deeper NIR data. As will be discussed in Section 4.5.2, we attribute this to an overestimate of the quiescent galaxy mass completeness by Davidzon et al.

The difference in mass completeness between the star-forming and quiescent samples presents an additional complication. Because the star-forming galaxies can be reliably detected to lower masses than quiescent galaxies in our $izYJHK_s$ selection function, the low-mass regime of the total SMF does not include contributions from quiescent systems. Therefore the total SMF at masses below the quiescent mass limit are effectively the low-mass end of the star-forming SMF. Although extrapolations from the evolution in the number density of $z \approx 0.2 - 2.5$ low-mass quiescent galaxies predict a SMF at least 1 dex lower than that of the star-forming galaxies at $z > 3$, there can be, in principle, a contribution from quiescent systems which could steepen the low-mass slope. Uncertainties are still above the 10% level at low-masses even at $z \approx 2$, and so unless the low-mass quiescent population is comparable in number to star-forming galaxies at the same redshift and mass, we can safely neglect their contribution.

4.3.4 *Derivation of the* $1/V_{\rm max}$ *correction*

Intrinsically faint galaxies at any given redshift are more likely to be missed by survey selection functions compared to brighter sources. For a NIR-selected sample, this equates to a mass bias by which low-luminosity, low-mass galaxies can only be detected in smaller volumes relative to brighter, more massive ones which could be otherwise detected if they were at higher redshifts. This is the well-known Malmquist Bias (Malmquist, 1920, 1922).

The most straight-forward approach to correct for such a bias is the $1/V_{\rm max}$ method of Schmidt (1968), which has enjoyed significant popularity owing to its simplicity. Briefly, the $1/V_{\rm max}$ method statistically corrects for selection incompleteness by weighting each detected object by the maximum comoving volume in which it can be observed, given the characteristics of the telescope survey. The $V_{\rm max}$ estimate per individual object is computed after finding the maximum redshift $z_{\rm max}$ by which the best-fit SED would no longer be observable[6] because of the survey's flux limit. On the other hand, the minimum redshift ($z_{\rm min}$) should be the one at which the source would become too bright and saturate the camera, although in practice is the lower boundary of the z

6 using ALF (Ilbert et al., 2005)

bin in which the considered galaxy lies ($z_{\rm low}$). Therefore, for the i-th galaxy inside the bin $z_{\rm low} < z < z_{\rm high}$, the maximum observable volume is

$$V_{\max,i} = \frac{4\pi}{3}\frac{\Omega}{\Omega^{\rm sky}}\left(D_{\rm cov}(\min(z_{\max,i}, z_{\rm high}))^3 - D_{\rm cov}(\max(z_{\min,i}, z_{\rm low}))^3\right), \tag{4.6}$$

where Ω is the solid angle subtended by the sample, $\Omega^{\rm sky} \equiv 41\,253\,\deg^2$, the solid angle of a sphere, and $D_{\rm cov}(z)$ is the co-moving distance at z (Hogg, 1999). If $z_{\max}$ exceeds the upper boundary of the redshift bin, the latter is used instead, meaning that the brightest sources are often assigned a weight that corresponds to the full volume of that redshift slice. As such, this correction is expected to be significant for only the faintest, lowest-mass sources in a given z-bin. While it is non-parametric and does not assume a functional form of the SMF, the $1/V_{\max}$ technique does assume that samples are drawn from a uniform spatial distribution, which is not accurate in the case of over- or under-dense environments Efstathiou et al. (1999). However, the assumption of a uniform spatial distribution is expected to be problematic only at $z < 1$, where large-scale structures have fully assembled, or in narrower surveys that can be biased by structures at smaller scales. Other methods exist which do not make this assumption such as STY (Sandage et al., 1979) and SWML (Efstathiou et al., 1988), a parametric and non-parametric maximum likelihood method, respectively. Already Davidzon et al. (2017) found that the constraints provided by COSMOS2015 were sufficiently strong for the $1/V_{\max}$ method as well as more complex methods (e.g. STY, SWML) to essentially converge. With even stronger constraints and larger effective area provided by COSMOS2020, we can expect even better agreement, with minimal advantages to the more complex methods. More extensive discussions on strengths and weaknesses of the various approaches can be found in the literature (e.g., Binggeli et al., 1998; Takeuchi, 2000; Ilbert et al., 2004; Johnston, 2011; Weigel et al., 2016).

4.3.5 *Further considerations of uncertainty and bias*

We adopt a statistical error budget on the SMF number density Φ consisting of the quadrature addition of Poisson noise (σ_N), cosmic variance fluctuations ($\sigma_{\rm cv}$), and uncertainties on masses induced by SED fitting ($\sigma_{\rm SED}$) such that $\sigma_\Phi = (\sigma_N^2 + \sigma_{\rm cv}^2 + \sigma_{\rm SED}^2)^{1/2}$. Fig. 4.3 shows the composition of the total error budget from $z = 1.1$ to 6.5 as a function of stellar mass for mass-complete bins.

Poisson noise

Poisson noise arises from processes wherein the abundance of a discrete quantity (or counts) is measured. Although in the limit of many events a Poisson process becomes indistinguishable with that of a Gaussian, in small number counts it can be the dominant source of uncertainty. Here we compute the Poisson error σ_N for each mass bin as $\sqrt{N}$ where N is the number of objects

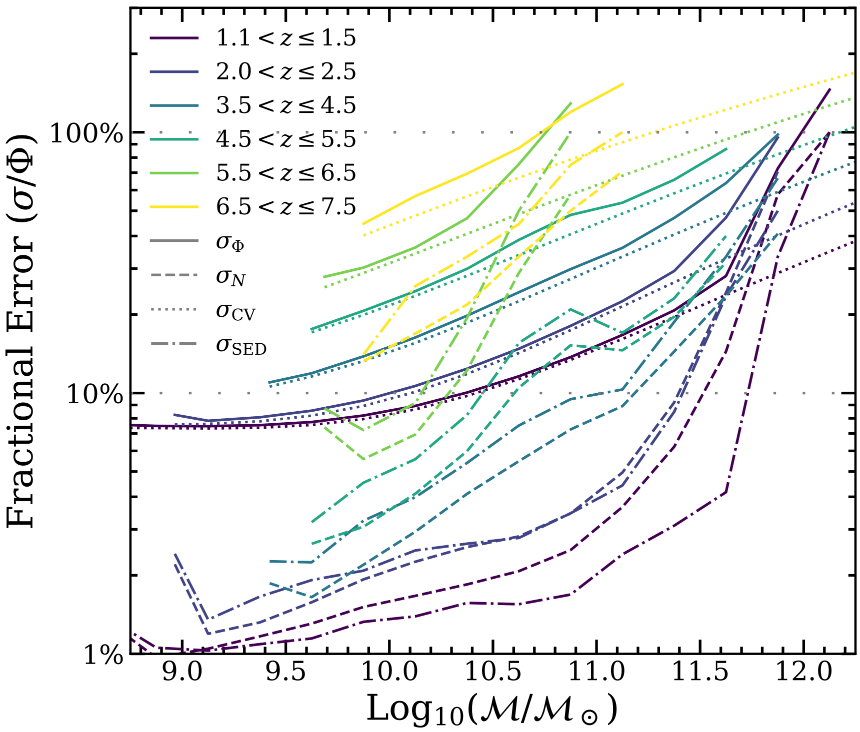

Figure 4.3: Adopted estimates for the total uncertainty σ_Φ (solid) as a function of stellar mass at several redshifts (by color), for mass-complete samples only. Contributions include uncertainties from Poisson noise σ_N (dashed), Cosmic Variance $\sigma_{\rm CV}$ (dotted), and SED fitting $\sigma_{\rm SED}$ (dash-dot).

in that bin. These values are re-computed for the star-forming and quiescent sub-samples separately. As shown in Fig. 4.3, the fractional contribution from σ_N increases with mass and redshift with the largest contribution at $\mathcal{M} > 10^{11.5}\mathcal{M}_\odot$.

The discrete nature of Poisson processes allows us to also provide upper limits on bins containing zero detected galaxies. Following Table 1 of Gehrels (1986), the statistical upper limit on Φ $(N = 0)$ for a given observed volume V is $\sigma_{N,\text{limit}} = 0.841/V$. See Ebeling (2003) and Weigel et al. (2016) for details and further discussions.

Cosmic Variance

It is well established that galaxy properties are correlated with environmental density (i.e. clustering). Galaxy clusters, while being an important laboratory for galaxy evolution, are not typical of galaxy environments. Because of their density, they impart a higher overall normalization to the stellar mass function. More noticeably, they tend to inflate the massive end of the mass function as they preferentially contain the most massive systems. This environmental field-to-field bias (so-called 'Cosmic Variance') is a topic of intense study, and is a key component to accurately assessing sample uncertainties when trying to infer universal or intrinsic properties of galaxies. There are many published methods to estimate Cosmic Variance, based on numerical simulations (Bhowmick et al., 2020; Ucci et al., 2021), analytical models calibrated to observations solved either using linear theory (Moster et al., 2011a; Trapp et al., 2020) or on forward simulation corrections to linear theory (Steinhardt et al., 2021), and observationally (e.g., Driver et al., 2010). Trenti et al. (2008b) combines results from cosmological simulations with analytical predictions.

Cosmic variance σ_{cv} is estimated following Steinhardt et al. (2021), who adapt the methods of Moster et al. (2011a) which, importantly, scale with stellar mass (up to $10^{11.25}\,\mathcal{M}_\odot$) and are commonly adopted for use in $0.1 \lesssim z \lesssim 3.5$ measurements as that is the redshift range in which Moster et al. calculator was devised. However, above $z \approx 3.5$ these estimates become increasingly underestimated, so Steinhardt et al. use linear perturbation theory to extend this work more reasonably to the early universe, while maintaining sub permille agreement at $z < 3$. Although environmental density has known covariance with star formation (e.g. Bolzonella et al., 2010; Davidzon et al., 2016), we assume cosmic variance is equivalent between star-forming and quiescent sub-samples. As shown in Fig. 4.3, the fractional contribution from σ_{CV} increases with mass and redshift with dominance at $\mathcal{M} \gtrsim 10^{11.5}\mathcal{M}_\odot$.

SED fitting uncertainties and bias

Another consideration is the uncertainty on the stellar mass estimate provided by the SED fit. Fig. 4.4 shows the likelihood distributions on stellar mass at fixed redshifts and masses produced by `LePhare` at $\Delta \log_{10}(\mathcal{M}/\mathcal{M}_\odot) = 0.025$ dex sampling. Trends with width of the likelihood distributions indicate

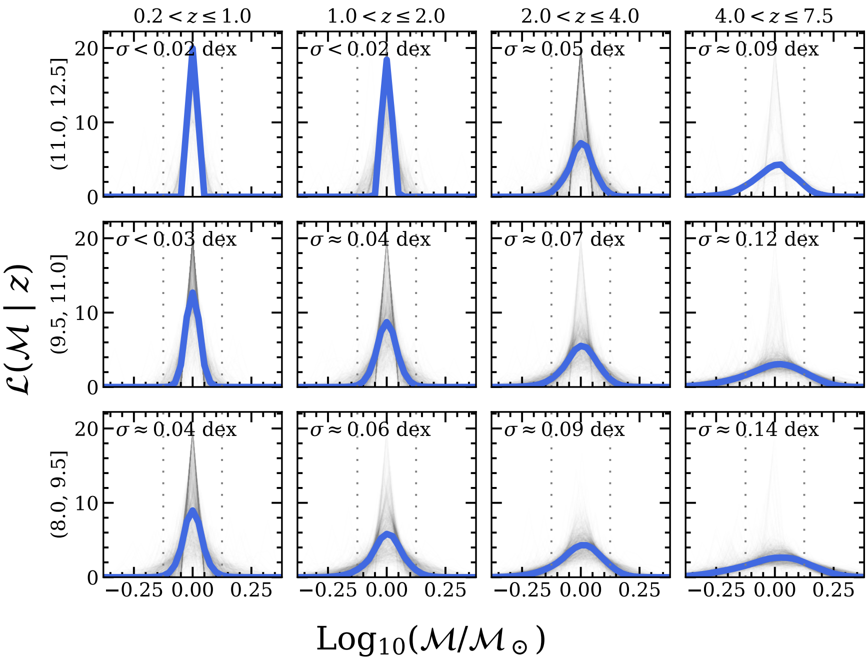

Figure 4.4: Likelihood distributions of galaxy stellar mass are derived from SED fitting with `LePhare` and assume a fixed redshift. Individual distributions (grey) are summarized by a median stack (blue) grouped by redshift and stellar mass (indicated in ranges of $\log_{10}(\mathcal{M}/\mathcal{M}_{\odot})$). Estimates of standard deviation σ are shown. The size of a typical mass bin used in this work is 0.25 dex, indicated by the pair of dotted grey lines in each panel.

that mass is best constrained for low-redshift, massive (i.e., bright) sources. Although there is non-zero skew and kurtosis in individual cases, the overall median distribution is symmetric. This is expected, as the uncertainty on stellar mass is essentially a measurement of the range of allowable templates and their normalization in the fitting procedure, and thus σ_{SED} scales with photometric uncertainties. However, these likelihood distributions on stellar mass should be treated as lower limits as they do not take into account any covariance with redshift.

While these typical per-bin distributions can be valuable, especially for injecting noise into measurements from simulations, attempting to compute its contribution to the SMF, σ_{SED}, using the typical width in a given bin is suboptimal as the wings of neighboring mass-bins contribute asymmetrically. To address this, we use the individual mass likelihood distributions to draw 1 000 independent realizations of the galaxy stellar mass function and thereby directly estimate the variance produced by the mass uncertainties, which we take as the 68% range about the median number density per bin of mass. As shown in Fig. 4.3, the contribution from σ_{SED} become dominant only at $\mathcal{M} > 10^{11.5}\mathcal{M}_{\odot}$, in some cases becoming larger than unity. They are comparable to contributions from σ_N across the entire mass range.

It is important to note that this does not account for systematic biases arising from SED fitting, such as assumptions of the stellar initial mass function[7] or potential photo-z offsets that propagate into the stellar mass determination. However, concerning the latter case, we show in Weaver et al. (2022a) that the combination of The Farmer and LePhare achieves a sub-percent photo-z bias even for faint, high-z sources (-0.004 at $25 < i < 27$) improving over other works including COSMOS2015. Systematic errors cannot be combined with random errors, and so additionally complicate measurements of the SMF. Given this indication of relatively low bias arising from the photo-z and significantly better constrained SEDs relative to previous measurements, we omit these considerations in the present work. See Marchesini et al. (2009) and Davidzon et al. (2017) for detailed discussions of various sources of bias and their effect on the SMF.

Eddington bias

The number of low-mass galaxies is orders of magnitude larger that of the highest-mass systems, and so a randomly chosen galaxy is overwhelmingly likely to be lower-mass. If even a small fraction of such truly low-mass systems scatter towards high-mass (owing to a $\mathcal{M}$ overestimate) it can significantly change the poisson-dominated high-mass number density estimate. The converse situation, while depleting the high-mass end, is far less likely and would have virtually no effect on the low-mass estimates. This is the well-known Eddington bias (Eddington, 1913). While generally understood to mean that there is a net bias leading to the overestimation of the density of massive galaxies, a small, but highly asymmetric uncertainty on the mass of low-mass systems can similarly generate such a bias which effects the shape of the low-mass regime (Grazian et al., 2015).

Effectively correcting for Eddington biases has been a leading point of discussion in recent literature, generally favouring the convolution of the fitting function with a kernel that describes the uncertainty in stellar mass (e.g., Ilbert et al., 2013; Davidzon et al., 2017). Recently, Adams et al. (2021) compared the effect of using three different forms for the convolution kernel finding a significant difference in the inferred intrinsic SMF. Alternative approaches have been proposed, e.g. Leja et al. (2019a) developed a non-parametric formalism for incorporating σ_Φ into an unbinned Likelihood fitting, whereby multiple realizations of the parent catalog are made, each time sampling stellar mass from the mass likelihood distributions of each galaxy. In this work we primarily adopt the traditional convolution kernel method to estimate Eddington bias. At the same time, we also fit the mass function using the method of Leja et al., and so follow their approach where relevant.

7 For example, $\mathcal{M}$ computed assuming Salpeter (1955) is on average 0.24 dex larger than those computed assuming Chabrier (2003).

4.4 THE SCHECHTER FUNCTION

Galaxy luminosity and stellar mass functions can be described empirically by the parametric formulation first introduced by Schechter (1976b) in the context of the local universe. Since then, the Schechter function has been adopted ubiquitously in statistical studies of galaxy mass assembly. It is more convenient to express the number density of galaxies per logarithmic mass bin $d\log\mathcal{M}$ as given by Weigel et al. (2016):

$$\begin{aligned} n_{gal} &= \Phi\, d\log\mathcal{M} \\ &= \ln(10)\,\Phi^*\, e^{-10^{\log\mathcal{M}-\log\mathcal{M}^*}} \times (10^{\log\mathcal{M}-\log\mathcal{M}^*})^{\alpha+1}\, d\log\mathcal{M}\,, \end{aligned} \tag{4.7}$$

which describes a power law of slope α at masses smaller than a characteristic stellar mass $\mathcal{M}^*$, whereupon the function is cut-off by a high-mass exponential tail. The overall normalization is set by Φ^*, which corresponds to the number density at $\mathcal{M}^*$. Although $\Phi(\mathcal{M})$ is expected to evolve smoothly with redshift such that $\Phi(\mathcal{M}, z)$ can be mapped from secondary functions $\mathcal{M}^*(z), \Phi^*(z), \alpha(z)$, most literature applications fit for each parameter independently at each redshift. A notable exception is Leja et al. (2019a).

Many studies have since found evidence that the galaxy population at low-z is better described as the co-addition of two Schechter Functions (e.g. Peng et al., 2010b) whereby the low-mass and high-mass end acquire individual normalizations (Φ_1^* and Φ_2^*) and low-mass slopes (α_1 and α_2) but retaining a single characteristic stellar mass $\mathcal{M}^*$. This so-called 'Double' Schechter form is as follows:

$$\begin{aligned} \Phi\, d\log\mathcal{M} = \ln(10)\, & e^{-10^{\log\mathcal{M}-\log\mathcal{M}^*}} \\ & \times \Big[\, \Phi_1^*\,(10^{\log\mathcal{M}-\log\mathcal{M}^*})^{\alpha_1+1} \\ & + \Phi_2^*\,(10^{\log\mathcal{M}-\log\mathcal{M}^*})^{\alpha_2+1} \Big]\, d\log\mathcal{M}\,. \end{aligned} \tag{4.8}$$

The Double Schechter function has been adopted in most of the studies in the local universe and up to $z \sim 2$. At higher redshift it is less clear whether this is a better description of Φ than a single Schechter function. Moreover, deviations have been recently been reported at $z > 7$ (e.g. Bowler et al., 2020a; Stefanon et al., 2021b) in which the observed stellar mass function is better described by a power law than by any Schechter profile.

4.5 RESULTS

Now having established the selections and methods adopted in this work, we present the resulting measurement of the SMF for the total, star-forming, and quiescent samples. We investigate also the evolution of number densities and

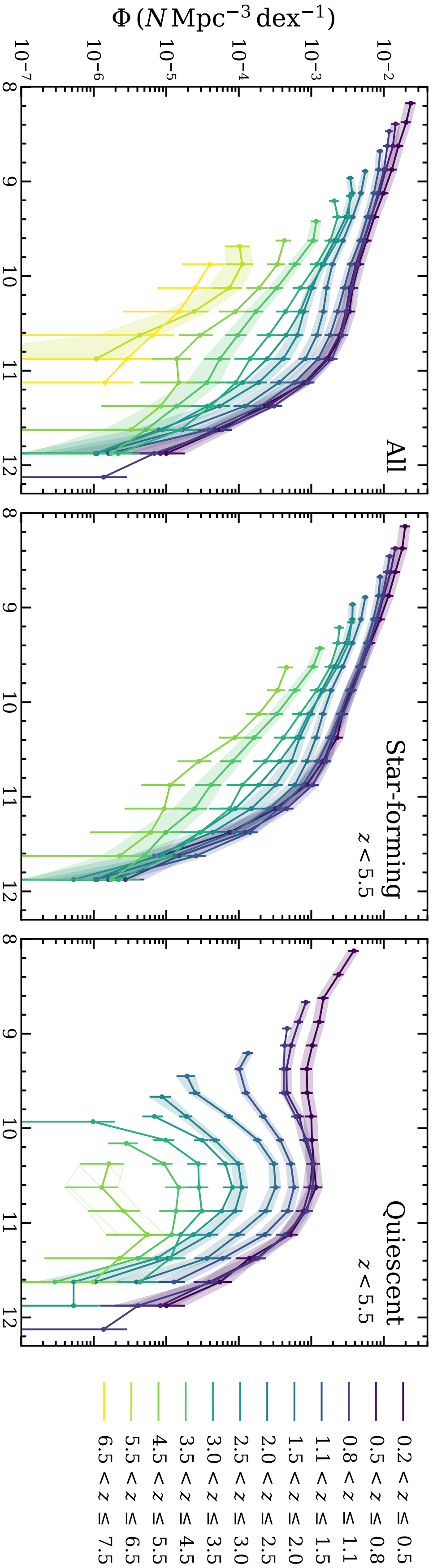

Figure 4.5: Evolution of the galaxy stellar mass function over 12 redshift bins ($0.2 < z \leq 7.5$) for the total, star-forming, and quiescent samples. Mass incomplete measurements are not shown.

quiescent fractions at fixed mass, and then fit the SMF for each sample with several methods to derive the evolution of the best-fit Schechter parameters.

4.5.1 *The Total Galaxy stellar mass function*

We measure the SMF for our total (star-forming and quiescent) sample divided in 12 redshift bins from $z = 0.2$ to 7.5. Shown in the left panel of Fig. 4.5, the shape and normalization of the SMF changes considerably over the ~ 10 billion years corresponding to this redshift range. At $z \lesssim 3$, the SMF features a smooth, monotonically decreasing low-mass number density which flattens before falling off steeply at $\mathcal{M} \approx 10^{11}\,\mathcal{M}_{\odot}$, around the so-called 'knee' of the function. Its overall shape and normalization remains roughly constant until $z \sim 1.5$ indicating a lack of mass growth at recent times, consistent with the decline of the cosmic star formation rate (Madau et al., 2014). However, by $z \sim 1.5$ the normalization decreases dramatically; on the order of $\sim 1\%$ of their $z \approx 0.5$ level at the knee, with the fastest growth occuring on the low mass end, consistent with mass 'downsizing in time' (Cowie et al., 1996; Neistein et al., 2006; Fontanot et al., 2009). At the same time, the knee itself becomes difficult to distinguish as the SMF takes the form of a smooth power law, and we become increasingly unable to constrain the low-mass end of the SMF due to worsening mass completeness (mass incomplete points are omitted in the figure). As expected from Fig. 4.3, the overall uncertainty grows significantly with increasing redshift and mass. We note that at a few epochs (e.g., $z \sim 4$–6), the SMF is not strictly monotonic, likely driven by systematics rather than a real physical phenomenon (see discussions in Section 4.6).

This evolution of the total SMF is compared with literature results in Fig. 4.6. We begin at $z \approx 0.2$ by comparing with two SMFs previously measured in the same field: Ilbert et al. (2013) and Davidzon et al. (2017); both use photo-z and $\mathcal{M}$ computed with `LePhare` over COSMOS, out to $z \approx 4$ and $z \approx 5.5$, respectively, with the nearly same binning scheme as we use. We note one exception that Ilbert et al. bins sources at $3.0 < z < 4.0$ whereas Davidzon et al. uses $3.0 < z \leq 3.5$ and $3.5 < z \leq 4.5$, and so we opt to follow the scheme of Davidzon et al. as it allows a comparison up to higher redshift and omit the comparison with the highest-z measurement of Ilbert et al. Additionally, Davidzon et al. only considered sources in the ultra-deep regions of COSMOS2015 (corresponding to UltraVISTA DR2) as the spatial inhomogeniety of the NIR bands implies a significant variation in selection function and mass completeness between the deep and ultra-deep regions. Thankfully, COSMOS2020 (corresponding to UltraVISTA DR4) contains nearly uniform NIR coverage ($\Delta \approx 0.4$ mag, Moneti et al. 2019) and so can leverage an area almost $2\times$ larger out to at least $z \approx 6.5$, beyond which the source density becomes clearly different between the deep and ultra-deep stripes. Thus, for the $6.5 < z \leq 7.5$ bin we use the 0.72 deg^2 subset of our primary area covered by the UltraVISTA deep stripes.

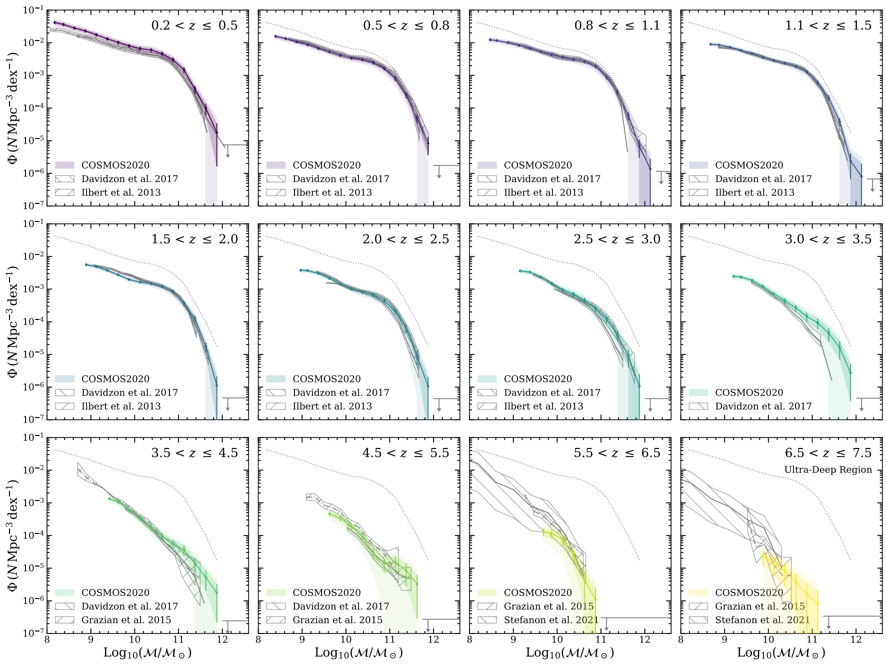

Figure 4.6: Evolution of the galaxy stellar mass function across 12 redshift bins ($0.2 < z < 7.5$). Upper limits for empty bins are shown by the horizontal grey line with an arrow. The $0.2 < z < 0.5$ SMF from the first redshift bin is repeated in each panel for reference shown by the purple dotted curve. Two other COSMOS stellar mass functions from Ilbert et al. (2013) and Davidzon et al. (2017) are shown for comparison, along with Grazian et al. (2015) from UDS/GOODS-S/HUDF and the recent work of Stefanon et al. (2021b) from GREATS at $z > 6$. Mass incomplete measurements are not shown. Upper limits for empty bins are shown by the horizontal grey line with an arrow.

At $z < 2.5$, in the range that all three can be directly compared, we find excellent agreement with Ilbert et al. and Davidzon et al. This is unsurprising, as measurements from Davidzon et al. at $z < 2.5$ are adopted directly from COSMOS2015 computed nearly identically as Ilbert et al. but with deeper imaging. Our measurements are similar, but with visibly less structure around the knee especially between $0.8 < z \leq 1.1$ where the volume density of sources is slightly higher, with the greatest increase at $\mathcal{M} \sim 10^{10}\,\mathcal{M}_{\odot}$. However, at $z \approx 2.5$ Davidzon et al. predicts a lower volume density than either Ilbert et al. or our measurements, which lie in agreement. Nor is this unsurprising, as Davidzon et al. made considerable changes to the SED library of `LePhare` and so our work is naturally more similar to that of Ilbert et al. (a trend made clearer when looking at quiescent systems in the Section 4.5.2). At $z > 3.0$, we predict significantly higher volume densities of massive galaxies compared to Davidzon et al., although the general shape of the low-mass end of the SMF remains similar. However, we find better agreement if we similarly limit the SMF to only the Ultra-Deep region. This may be due to the presence of clusters preferentially in the Deep region at $3 < z \leq 4$, one of which has been recently spectroscopically confirmed by McConachie et al. (2021). Constraints from Grazian et al. (2015) can be introduced at $3.5 < z \leq 4.5$, although at a significantly higher degree of caution as Grazian et al. uses a combination of smaller CANDELS fields (GOODS-South, UDS, HUDF) with an overall area of $\sim 12\times$ smaller than of the present study which implies a higher degree of uncertainty from cosmic variances and Poisson noise, as well as a reduced constraint on rare, high-mass systems found only in larger volumes. They are, however, completely independent as Grazian et al. do not include the CANDELS-COSMOS field. Where a comparison is possible ($3.5 < z \leq 7.5$), our results are consistent with those of Grazian et al. within the stated uncertainties. Similarly at $z > 5.5$, we are able to compare with the recent measurements of the low-to-intermediate mass regime from Stefanon et al. (2021b), with which the present work also is consistent within the stated uncertainties.

We do not include comparisons to stellar mass functions inferred from L_{UV}-selected samples (e.g. Harikane et al., 2016; Song et al., 2016) as they more directly measure the abundance of UV-bright sources as opposed to the general galaxy population. Moreover, these studies often rely on color-color selections which are less certain than photo-z, in addition to being susceptible to a number of systematic effects when translating a UV luminosity function to a SMF by means of a $L_{\mathrm{UV}} - \mathcal{M}$ relation. We refer the reader to Davidzon et al. (2017) for basic comparisons and discussion.

While the smaller area surveys used by Grazian et al. (2015) and Stefanon et al. (2021b) are effectively mass complete at lower masses, COSMOS2020 contains the widest contiguous NIR imaging at these depths and consequently provides a larger volume than any previous study (including Davidzon et al.) and so introduces new constraints on the most massive systems at all redshifts $z \leq 7.5$. Yet, these constraints are only marginally higher than our upper limits, and so it appears that we are quickly approaching the statistical limit beyond which a larger survey is needed to find more massive systems, if they exist (see discussions in McPartland et al., in prep). The nature of these sources, their authenticity, and their potential Eddington bias is further assessed in Section 4.6.3.

4.5.2 *The Star-forming and Quiescent Galaxy stellar mass functions*

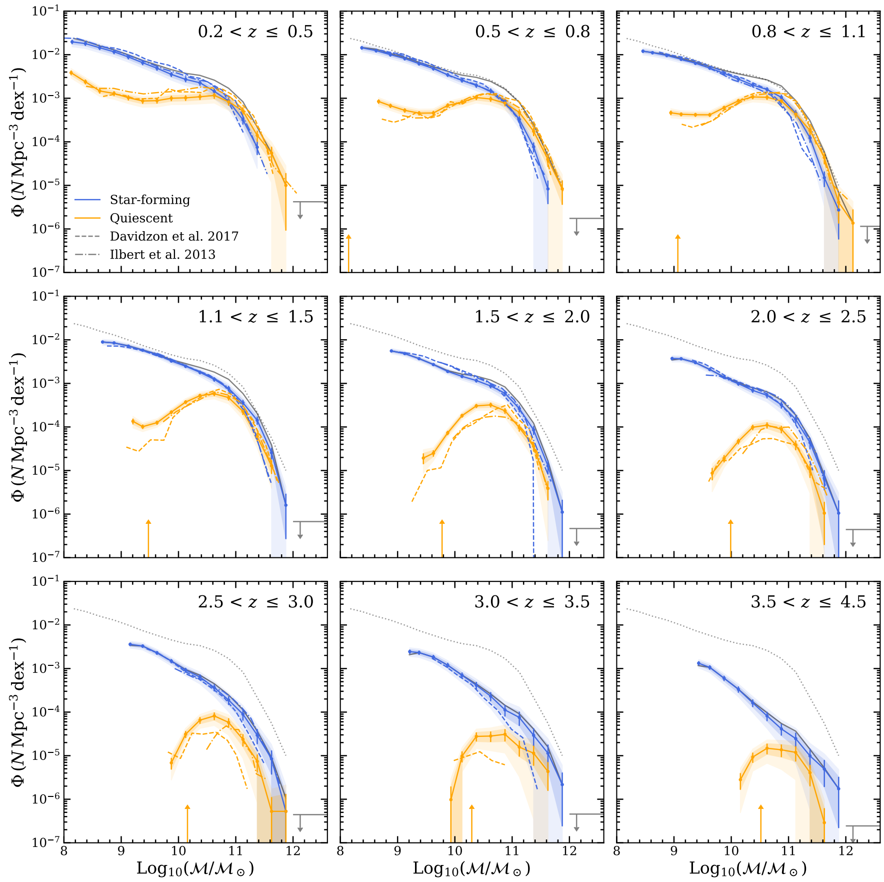

Figure 4.7: Evolution of the star-forming (blue) and quiescent (orange) galaxy components of the Galaxy Stellar Mass Function in nine bins of redshift ($0.2 < z < 4.5$) with 1 and 2σ uncertainty envelopes, as well as similarly colored comparisons with other literature studies in COSMOS from Ilbert et al. (2013) and Davidzon et al. (2017). For reference, the total SMF is shown in each bin (solid grey) and the $0.2 < z < 0.5$ total SMF is repeated in each panel (purple dotted). Mass incomplete measurements as defined by the ch1 limiting magnitude are not shown, with the mass limits corresponding to the K_s limiting magnitude are shown by the orange arrows. Upper limits for empty bins are shown by the horizontal grey line with an arrow.

In Section 4.3.2, we used an $NUVrJ$ color-color selection to distinguish quiescent systems from star-forming ones and then estimated their respective mass completeness limits, which will differ since their $\mathcal{M}/L$ ratios are not the same. The corresponding SMFs are shown in Fig. 4.5 and compared to literature in Fig. 4.7. Their corresponding Schechter parameters are reported in Tables 7 and 8 for the star-forming and quiescent samples, respectively. Detailed fitting results can be found in Appendix 4.C, and discussed below in Section 4.4. Fractional uncertainties on mass and cosmic variance are adopted from the total sample, leaving only the impact of poisson noise differentiated between the star-forming and quiescent samples (see Section 4.3.5).

We can follow the development of quiescent systems out to $z \approx 5$, although with significant uncertainty at $z \gtrsim 4$. As evidenced by Fig. 4.1, only ~ 200 of quiescent systems are found at $z \gtrsim 4$, dropping precipitously to only one candidate by $z \sim 6$. This is partly driven by mass completeness. At $z < 3$ the difference between the IRAC channel 1 mass limit and the effective mass completeness dictated by our $izYJHK_s$ selection function is well below 1 dex in mass. This difference becomes significant at $z \sim 3.5$ when K_s falls blueward of the Balmer break causing the mass completeness to shift upwards to higher masses, indicated by the hatched region of the SMF of $4.5 < z \leq 5.5$ quiescent systems in Fig. 4.5. At this point the identification of quiescent systems is reliant on only a few bands, and is dependent on the particular SED templates (as suggested by the two distinct clusters in the upper right corner of the $4.5 < z \leq 5.5$ $NUVrJ$ diagram), making such a distinction hazardous and susceptible to interlopers.

Our measurements of the star-forming and quiescent SMFs are generally in good agreement with other literature measurements in COSMOS, namely Davidzon et al. (2017) and Ilbert et al. (2013) as they are the only NIR-selected, mass-complete samples from which star-forming and quiescent sub-samples are identified by $NUVrJ$. Other selections may induce additional systematics, and other separation methods (e.g., UVJ, BzK, $sSFR$) implicitly adopt a different criteria for quiescence (see Davidzon et al., 2018; Leja et al., 2019b). As shown in Fig. 4.7, while our sample provides similar quiescent number densities compared to Ilbert et al., we find significantly more low-mass quiescent galaxies compared to Davidzon et al. at $0.5 < z \leq 2.0$. Since our sample is derived from comparably deeper NIR data, it is expected to be complete down to lower masses relative to Davidzon et al. Given the increased quiescent galaxy number densities near the mass completeness limit in our work, we conclude that the 70-80% completeness threshold of Davidzon et al. is underestimated by a factor of $\sim 2-10$ across this z-range, and is more likely only $15-35\%$ complete. We note, however, that this in agreement with worst-case scenario discussed by the authors (see Section 4.2 of Davidzon et al. 2017).

The SMFs of star-forming and quiescent galaxies have remarkably different shapes and evolutionary histories. As shown by Fig. 4.5, star-forming galaxies at $0.2 < z \leq 0.5$ follow a double Schechter form (Equ. 4.7) with a characteristic stellar mass $10^{10} < \mathcal{M}_* < 10^{11}\,\mathcal{M}_\odot$. By $z \sim 3$ it appears to flatten into a smooth powerlaw-like form with a lower overall normalization. The SMF of quiescent galaxies at the same epoch is known to be different; it follows the form of a double Schechter (Equ. 4.8) with a low-mass upturn and a similarly positioned $\mathcal{M}^*$ (Moutard et al., 2016). The form appears to loose its lower-mass Schechter component around $z \sim 2$. Although this may be physical, the fact that the quiescent sample is less mass complete at a given z means that the low-mass end of the total sample reflects only the contribution from star-forming systems. However, the contribution from low-mass quiescent systems, if they exist, can be expected to be $< 1\%$.

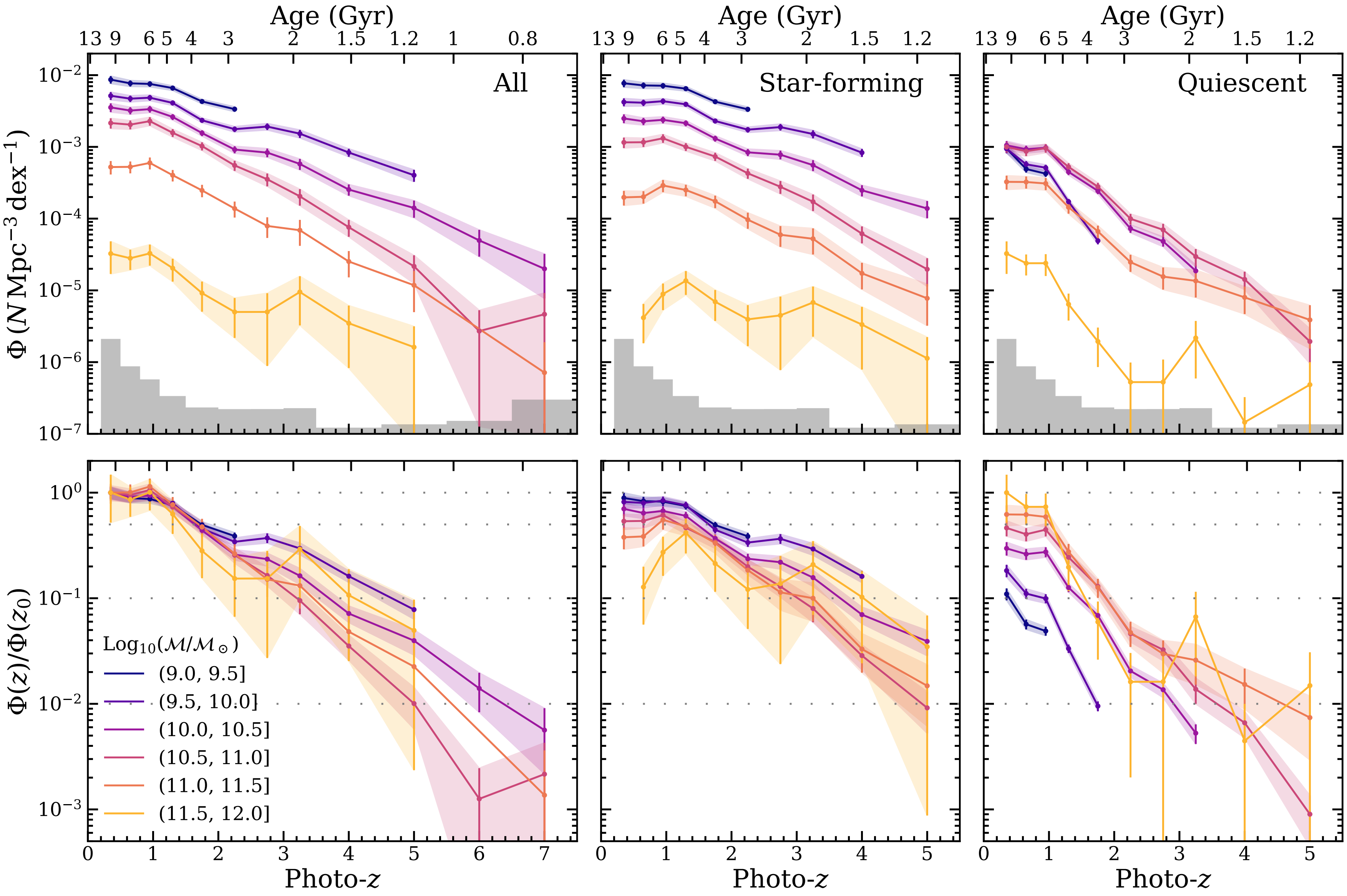

Figure 4.8: *Upper:* Evolution of the volume number density Φ of the total, star-forming, and quiescent samples at fixed mass, showing only mass complete bins. 1σ uncertainties are indicated by the colored envelopes. Upper limits at each redshift are indicated by the grey shaded region, with those from a CANDELS-like survey shown by grey dotted curve above it. *Lower:* Evolution of the fractional change in the Galaxy Stellar Mass Function at fixed mass, relative to $z_0 \equiv 0.2 < z \leq 0.5$ of the total sample.

This evolutionary picture is perhaps more easily understood by Fig. 4.8. Here galaxies have been binned by mass instead of redshift, allowing for a more direct comprehension of the growth, or lack thereof, in the number density of galaxies at fixed mass. We first notice a strikingly constant rate of growth in the number densities at all masses from $z = 7.5 \rightarrow 1$ (i.e. similar slopes in each mass bin), constituting ~ 5 Gyr or $\sim 36\%$ of the history of the universe, consistent with recent findings by Wright et al. (2018). This consistent growth is made especially clear in the lower left panel where the growth is relative to $z_0 \equiv 0.2 < z \leq 0.5$. Although the uncertainties are considerable, there may be a hint that the growth is fastest (i.e., the slope is higher) for systems at $\mathcal{M} \sim 10^{10.5-11.0} \mathcal{M}_\odot$ or at the 'knee' of the mass function where star formation is hypothesized to be the most efficient (Behroozi et al., 2013; Moster et al., 2013; Wechsler et al., 2018). However, we note that Eddington Bias remains uncorrected, and so the apparently slow evolution of the most massive bins may be a bias. While the least massive systems are always more common than more massive ones, this is simply a consequence of the monotonic shape of the SMF.

Another interesting feature is the loss of number density in $\mathcal{M} = 10^{9.5-10.0}$ and $\mathcal{M} = 10^{10.0-10.5} \mathcal{M}_\odot$ systems at $z \sim 2$, as shown in Fig. 4.6. There are fewer such systems in this work relative to both Davidzon et al. and Ilbert et al., and so this indicates a sample incompleteness, or alternatively a bias in z and/or $\mathcal{M}$. Determining the origin of this difference is non-trivial, and so we simply caution against over-interpretation.

The highest-z constraints are difficult to interpret due to the incompleteness of low-mass systems (which are omitted) and the rarity of the most massive sources. The fact that the constraints at $z \gtrsim 5$ overlap with the $\Phi(N = 0)$ upper limit region in grey indicates that even the comparably large, deep volume of COSMOS is insufficient to provide robust measurements of the number density of massive galaxies at these early times. Future constraints will be derived from even larger volumes, which are expected to be delivered soon by near-infrared wide-area deep surveys such as *Roman* and *Euclid*.

This nearly uniform growth in the number density of sources slows down at $z \lesssim 1$ in all but the least massive bin, with almost no growth for $\mathcal{M} > 10^{10.5} \mathcal{M}_\odot$ systems. To understand this further, we examine the growth of the number densities of the star-forming and quiescent sub-samples. As seen in the middle column of Fig. 4.8, star-forming systems maintain their monotonic mass-ranked order. In addition, their abundance grows at a nearly constant rate until $z \sim 1$, when it starts to decrease. As shown in the rightmost column, quiescent galaxies follow a different pattern. Instead of a monotonic, mass ranked growth in number density, quiescent systems around the knee at $10^{10.5} \mathcal{M}_\odot$ are always the most numerous, with a slower rate of growth compared to the least massive and most massive bins. While the evidence for massive quiescent systems at $z > 2$ is hampered by the limited volume of COSMOS, we can confidently see that they appear to grow quickly in number at $z < 2$. By $z \sim 1$ their number densities stabilize $\sim 3 \times 10^{-5}\, N\, \mathrm{Mpc}^{-3}\, \mathrm{dex}^{-1}$. Reasons for this are discussed in Section 4.6.3.

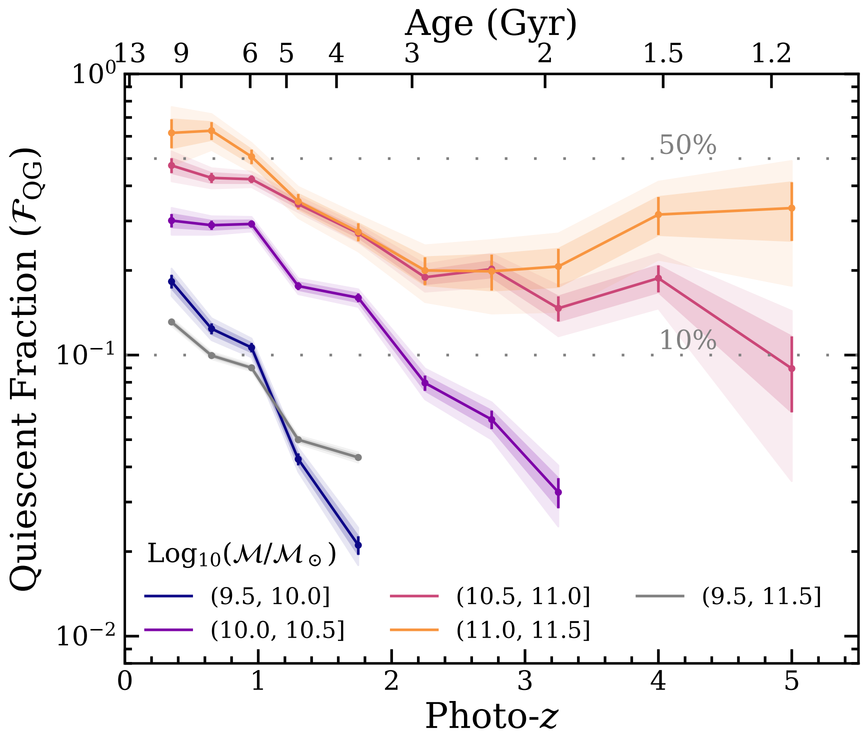

Figure 4.9: Evolution of the fraction of quiescent galaxies as a function of redshift in four bins of mass. Uncertainty envelopes correspond to 1 and 2σ limits. Points which are mass incomplete are not shown.

Fig. 4.9 shows the evolution of the fractional contribution of quiescent systems in four 0.5 dex bins at fixed mass and one wider 2 dex bin. The fraction of low mass systems increases with time monotonically, and at a higher rate of growth than those of the highest mass, at least for $z < 2$ where bins are mass complete. The growth in the fraction of the most massive systems is more severely complicated by Eddington Bias than low-mass systems and as such should be taken as upper limits. Nevertheless, it seems that between 10–30% of $\mathcal{M} > 10^{11}\, \mathcal{M}_\odot$ galaxies are quiescent from $z \approx 5 \rightarrow 1$, growing above 50% and plateauing at $z < 1$.

4.5.3 *Inferring the intrinsic mass function*

So far we have accounted for three key sources of statistical uncertainty in Section 4.3.5, however Eddington Bias remains a source of systematic uncertainty which has not yet been removed.

To do so, we fit the observed total, star-forming, and quiescent SMFs with Schechter functions convolved with a kernel which attempts to describe the mass uncertainty for a given mass and redshift interval. This approach is common in the literature (Ilbert et al., 2013; Davidzon et al., 2017; Adams et al.,

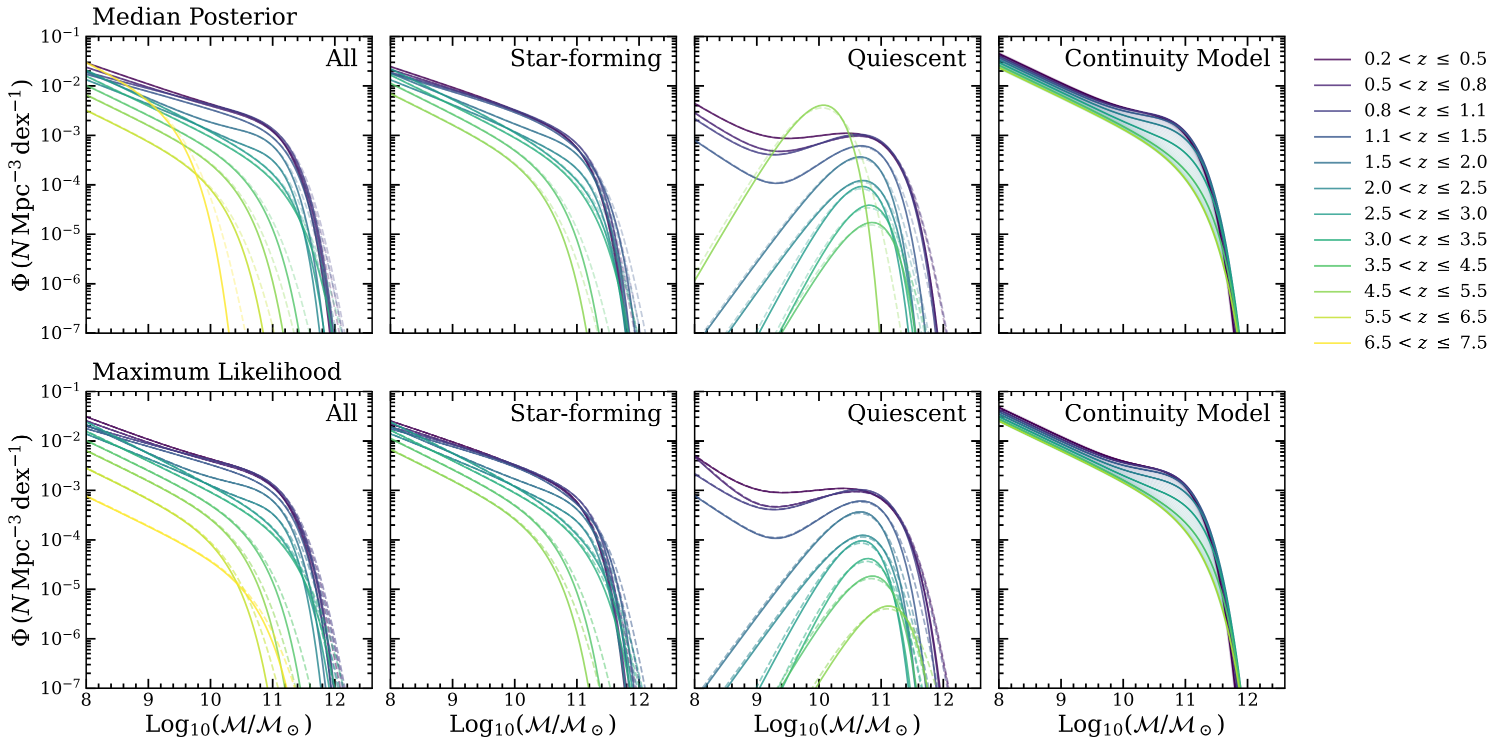

Figure 4.10: Summary of inferred galaxy stellar mass function evolution. Best-fit parameters are estimated by a Markov Chain Monte Carlo fitting with a Schechter function at fixed redshift convolved with a redshift dependent kernel (colored dashed curves) from which the inferred intrinsic mass function is recovered with the unconvolved Schechter fit (colored solid curves) for the total (leftmost), star-forming (center left), and quiescent (center right) samples. Estimates are shown corresponding to both the median posterior parameter values (upper row) and the single set of maximum likelihood parameters (bottom row). The rightmost panel shows the results of fitting our measurements with the Continuity Model at $z \leq 3$.

2021). We adopt a two component parametric kernel of the form introduced by Ilbert et al. (2013):

$$\mathcal{D}(\mathcal{M}_0, \sigma_{\text{Edd}}, \tau_{\text{Edd}}) = \frac{1}{2\pi} exp\left(\frac{-\mathcal{M}_0}{2\sigma_{\text{Edd}}}^2\right) \times \frac{\tau_{\text{Edd}}}{2\pi}\frac{1}{(\tau_{\text{Edd}}/2)^2 + x^2)} \tag{4.9}$$

which describes a core Gaussian component with a constant width σ_{Edd} in product with a Lorenztian component which provides wide wings that are crucial to capturing catastrophic errors in SED measurements. The other free parameter τ_{Edd} is redshift dependent such that $\tau_{\text{Edd}} = \tau_c\,(1+z)$, widening the wings with increasing z. Both σ_{Edd} and τ_{c} are nuisance parameters and leaving them free to vary is likely to produce over-fitting. Instead of fitting the kernel directly to the $\mathcal{L}(\mathcal{M}|z)$ distributions (as in Davidzon et al. 2017 and Adams et al. 2021), we opt to directly fit the total SMF leaving the kernel free to vary independently at each redshift. Overall, we find approximate median values for $\sigma_{\text{Edd}} \approx 0.2$ and $\tau_{\text{c}} \approx 0.1$. The resulting kernel is significantly wider than the $\mathcal{L}(\mathcal{M}|z)$ distributions shown earlier in Fig. 4.4. This is because the latter assume known redshifts and therefore underestimate the full $\mathcal{M}$ uncertainty as there is likely to be a covariance with z. Additionally, Eddington Bias at one redshift interval is expected to have contributions from other intervals. In addition, there is a mild evolution of $\mathcal{L}(\mathcal{M}|z)$ with $\mathcal{M}$, which we have omitted from our kernel form for the sake of simplicity and to avoid overfitting. Interestingly, while $\sigma_{\text{Edd}} \approx 0.2$ is smaller than that found by Davidzon et al. (0.35), $\tau_{\text{c}} \approx 0.1$ is conspicuously larger (0.04, same as in Ilbert et al. 2013). If we instead fit to the $\mathcal{L}(\mathcal{M}|z)$ distributions directly, we find $\sigma_{\text{Edd}} \approx 0.05$ and $\tau_{\text{c}} \approx 0.04$. These are much smaller than Davidzon et al., which may be explained in part by the slightly different approach they used to estimate $\mathcal{L}(\mathcal{M})$ that more directly incorporates uncertainties on photo-z (see Section 4.1 of Davidzon et al. 2017). Nevertheless, we opt for the pessimistic case and fix the kernel shape and evolution with z to $\sigma_{\text{Edd}} \approx 0.2$ and $\tau_{\text{c}} \approx 0.1$ for the subsequent analysis of the total, as well as the star-forming and quiescent SMFs; their shapes are shown in the inset panels in Fig. 4.21.

At the present, precisely determining the shape of this kernel, and its evolution with z as well as $\mathcal{M}$ is problematically difficult. In addition, the results stemming from these kernel-convolved fits suffer a degree of degeneracy with the kernel parameters (fixed or unfixed). We are not alone in issuing this caution; although the SMF measurements of Davidzon et al. (2017) is similar to those of Grazian et al. (2015), their different choice of convolution kernel caused their inferred Schechter parameters to differ. Recently, Adams et al. (2021) explored the impact of the assumed shape of the kernel, as well as various other systematics at $z < 2$. Pushing measurements of the SMF to higher redshifts, where fewer constraints are available, naturally increases the influence of the kernel shape, and so the results derived here for α, $\mathcal{M}_*$, and

Φ^* (Tables 6, 7, 8) should be taken in conjunction with our assumed kernel shape and its evolution with z.

When fitting each sample (total, star-forming, and quiescent) we assume a double Schechter parametric form (Equ. 4.8) and move to a single Schechter form (Equ. 4.7) when it achieves a lower χ^2 per degree of freedom. The point at which this change occurs, and when various parameters are fixed, are different for each of the three samples as it depends on not only their shape but also $\mathcal{M}_{\text{lim}}$. The fitted points follow from before: we account for minor incompleteness on the low-mass end using the $1/V_{\text{max}}$ correction and include only mass complete $\mathcal{M}$-bins adopting the uncertainty budget from Section 4.3.5. We proceed in two stages, fitting first using a simple and efficient χ^2 minimization routine (MINUIT, James et al., 1975) whose resulting best-fit parameters are used to set the initial positions of walkers in a secondary and longer Markov Chain Monte Carlo (MCMC) optimization (EMCEE, Foreman-Mackey et al., 2013). It is generally appropriate to initialize walkers at well-defined initial positions, assuming each chain is linearly independent (in this case scattering them by 1% of the initial parameter values) and allowed to converge well past its respective autocorrelation length such that it is not sensitive to those initial conditions (Hogg et al., 2010). Flat priors are used for each parameter, with generous limits which are typically never encountered during the fit. The MCMC walkers are programmed to seek the state of maximum likelihood derived similarly from the minimum χ^2, and do so following the "stretch move" (Goodman et al., 2010) until converged, defined as having every chain iterate for at least 10× their autocorrelation length, and every autocorrelation length having changed by $< 1\%$ of their previous value. We do not see any significant differences by using a different move style (e.g., Red-Blue, Metropolis-Hastings), suggesting that the data provide sufficient constraining power. Although we include the original χ^2 results throughout, we focus on the MCMC results which provide full posterior sampling, which are taken from the last 90% of each chain to avoid only the initial burn-in. The MCMC and χ^2 methods generally find consistent results.

The evolution of the Schechter parameters inferred from the total SMF are shown in the leftmost panels Fig. 4.11, and compared with Davidzon et al. in the center panels. Table 6 records the best-fit parameters, with detailed fits shown in Fig. 4.6 of Appendix 4.C.

As evidenced by the low χ^2 values, the double Schechter well describes the observed SMF at $z \leq 3$. No parameters are fixed other than those of the kernel. A single Schechter finds a better fit at $z > 3$. However, the increasing mass completeness limit weakens the constraints on the low-mass slope α_1, which we fix to the value from the previous redshift interval for the χ^2 and MCMC methods independently. This minor difference in α_1 drives a small difference at higher redshifts.

It is important to note that from χ^2 minimization we obtain a single set of parameters which have minimized the χ^2 as well as their symmetric, Gaussian uncertainties. On the other hand, MCMC provides only posterior distributions that can be used to estimate parameter uncertainties but do not imply a unique definition of "best-fit parameters". Commonly, the median is the best-fit statistic of choice, bounded by a 68% one-parameter confidence interval (which ignores covariance). However, for highly skewed posteriors the median may lie out in the wings and is therefore not a typical value for that parameter. In this case, the most obvious choice may be the parameters corresponding to the model which has found the maximum likelihood. However, MCMC cannot provide an uncertainty on these maximum likelihood parameters, limiting its use. Worse, the model uncertainty cannot be derived from the posterior widths, as they also are covariant and so the resulting error envelopes will be meaningless. The most obvious choice then is to compute the medians and widths of the vertical distribution models produced by all of the chains after burn-in. However, the curve traced by the median is not guaranteed to reflect the actual model function, and the 68% confidence envelopes may not contain the maximum likelihood nor the median posterior. We therefore strongly advise that models are reconstructed using the maximum likelihood parameters, and that caution should be exercised when interpreting best-fit parameters from median posteriors. Thankfully, the situation is less severe for symmetric posteriors, which for the total SMF are generally symmetric at $z \lesssim 6$. At $6.5 < z \leq 7.5$, the posteriors are highly skewed as the relatively weak constraints produce widespread parameter covariance. The maximum likelihood results in all cases appear reliable.

The evolution of the fitted Schechter models for both the median posterior and maximum likelihood is shown in Fig. 4.10. A divergence at $z \sim 7$ can be clearly seen. Furthermore, the two methods show a different evolution of the low-mass slope, but this can be attributed to fitting each redshift interval independently at $z < 3$ where low-mass constraints are available and then necessarily fixing the slightly different α_1 values thereafter. We note that determining the cosmic stellar mass density is highly dependent on α_1, and so this minor discrepancy will propagate accordingly.

In addition to the fits at fixed redshift, we also use the method of Leja et al. (2019a) to fit both the shape *and* evolution of the SMF simultaneously. This so-called 'Continuity Model' is fitted on the unbinned distribution of mass-complete sources in z and $\mathcal{M}$[8]. Double Schechter parameters $\mathcal{M}^*$, Φ_1, and Φ_2 are treated as continuous functions over time described by second order polynomial expansions in z. The slopes α_1 and α_2 are assumed to be constant. These 11 parameters are joined by a minor parameter σ_{eff} which attempts to capture the cosmic variance. The effects of Eddington Bias are incorporated by resampling the entire catalog by 10 random draws from their $\mathcal{L}(\mathcal{M}|z)$, which is wider than the kernels used in fits at fixed redshift and so allow for a relatively greater degree of scatter. Only the mass-complete points are fitted, which allows sources near the mass limit to scatter in and out of the SMF realizations. Importantly, this method accounts for the significant covariances in the Schechter function between adjacent redshift bins which is

8 This is not formally an STY method, as the normalization is constrained as a fitted parameter.

neglected when binning in redshift, and exploits it to provide generally tighter parameter constraints; see Leja et al. for details.

We employ EMCEE to determine the posterior distributions and maximum likelihood fit of the Continuity Model, noting the aforementioned caveats. Initially we tried to fit the entire SMF out of $z = 7.5$, but the fits were inaccurate due to the limited range of evolution which can be described by a second order expansion. Already known from Davidzon et al. (2017), Φ rises quickly before slowing down towards $z \sim 0$. The second order expansion in z can either fit the quick rise or the slow down, but not both. We therefore only consider galaxies at $z \leq 3$ in our Continuity Model fit, as Leja et al. do, and leave modifications of the fitting functions to future work. The expansion shapes are determined not by their coefficients, but rather by the amplitudes of three anchor points at fixed redshifts. The location of these anchor points is largely arbitrary, and so we follow Leja et al. in choosing $z = 0.2$, 1.6, and 3.0. Given the larger parameter space, we randomly initialize 100 walkers which explore the space until converged as before. We cannot apply the $1/V_{\rm max}$ correction directly to the Continuity Model, and so we exclude sources in the lowest 0.25 dex from the $\mathcal{M}_{\rm lim}$ from Equ. 4.3.

The evolution of the Schechter parameters inferred from the star-forming and quiescent SMFs are shown in the rightmost panels in Fig. 4.11. Tables 7 and 8 record the best-fit parameters, with fits shown in Fig. 4.7 in Appendix 4.C. Adapting the Continuity Model to the star-forming and quiescent SMFs is left to future work.

The treatment for the star-forming SMF fit is similar to that of the total form. We begin with a double Schechter form at $0.2 < z \leq 0.5$, and transition to a single Schechter form at $3.0 < z \leq 5.5$ with a fixed low-mass slope α_1. The best-fit model describes the observed SMF reasonably well until $z \approx 3.5 - 5.5$ where an excess of high-mass star-forming systems is observed that cannot be described by a single Schechter form. Possible reasons for this are discussed in Section 4.6.3.

The quiescent SMF behaves noticeably differently from that of the star-forming sample and therefore requires a different strategy. We begin with a double Schechter form at $0.2 < z \leq 1.5$ but fix the low-mass slope α_2 at $1.1 < z \leq 1.5$ as the constraints on the low-mass regime deteriorate. With no significant low-mass constraints, we transition to a single Schechter at $1.5 < z \leq 5.5$. We note that this disregards a possible low-mass contribution and so the extrapolation to low-masses is an underestimate at $z \approx 2 - 3$. At $z \approx 5$, however, the effective K_s selection makes the sample less mass complete as K_s falls blueward of the Balmer break, and so the SMF is constrained by only three $\mathcal{M}$-bins. With effectively no low-mass regime observed, we fix the low-mass slope. Although the χ^2 and Maximum Likelihood best-fit models are reasonable, the posterior distributions are highly skewed and covariant. Not only does this cause the median likelihood model to be unreasonable, but also results in a wide range of possible fits. Consequently, the $4.5 < z \leq 5.5$ median posterior parameter estimates are unreliable.

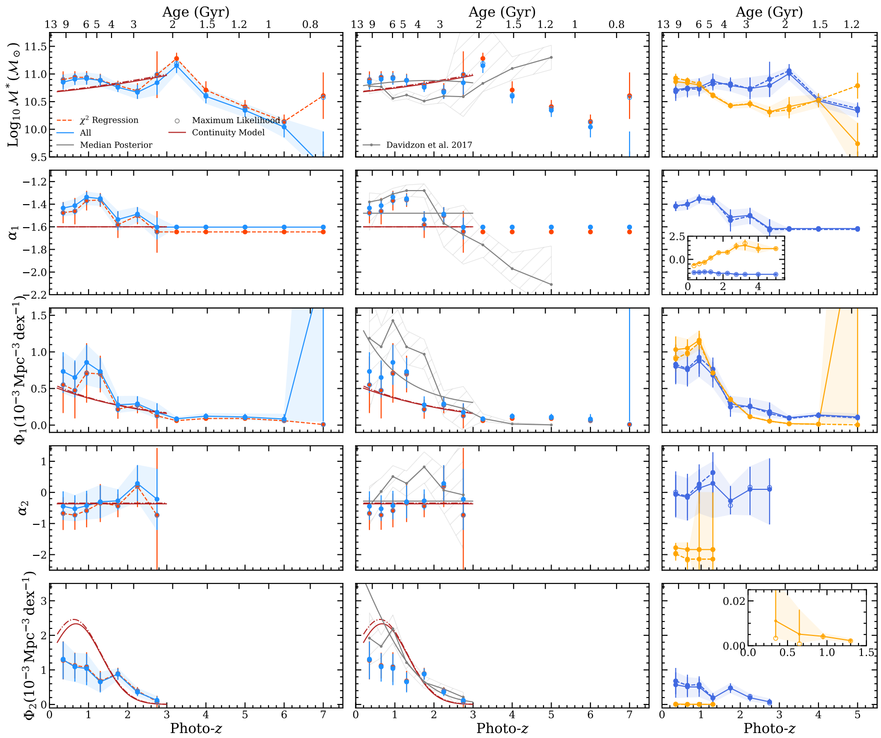

Figure 4.11: *Left:* Evolution of best-fit Schechter parameters computed from fits to the total SMF from χ^2 regression (red), as well as likelihood methods using fixed redshifts bins (blue), and the continuity model (maroon). While for the fixed redshift likelhood technique, median posterior values and $\pm34\%$ uncertainties are shown by the solid points with error bars, the maximum likelihood values are shown by the unfilled points. Error bars are not shown when parameters are fixed. For the continuity technique, median posterior value curves are solid, and maximum likelihood are dash-dot. *Center:* Same points as on the left panels compared to literature values from Davidzon et al. (2017) (grey points with hatching) and Leja et al. (2019a) (grey curves)). *Right:* Evolution of the best-fit Schechter parameters computed from fits to the star-forming (blue) and quiescent (orange) SMFs from the χ^2 regression (dashed) and fixed redshift likelihood (median posterior: filled points with $\pm34\%$ envelope, maximum likelihood: unfilled points).

The evolution of the inferred total, star-forming, and quiescent SMFs resulting from the fixed redshift and Continuity Model fits are shown in Fig. 4.10 of Appendix 4.C. Furthermore, the evolution of the Schechter Function parameters inferred from each sample are shown in Fig. 4.11. Although the evolutionary physics inferred from the fitting will be discussed in Section 4.6.1, we briefly describe the immediate result here.

Parameters derived at fixed redshift from the χ^2, maximum likelihood, and median posterior agree well for the total, star-forming, and quiescent SMFs, with few exceptions.

The characteristic mass is found to be $\mathcal{M}^* \approx 10^{10.7-11.0}\,\mathcal{M}_\odot$ with very little significant evolution until $z \approx 3$ when it begins to decrease with increasing z, with the Continuity Model suggesting a potentially increasing value with z. This contrasts with results from Davidzon et al. (2017), who find an initial decrease in $\mathcal{M}^*$ which increases at $z > 3$. The results of the Continuity Model of Leja et al. (2019a) agree well with the lack of evolution suggested by our measurements, although in some tension with our Continuity Model results. $\mathcal{M}^*$ of the star-forming sample similar in shape but generally larger than that of the quiescent sample, except at $0.2 < z \leq 0.7$ where the quiescent sample finds a slightly higher $\mathcal{M}^*$. The likeness of the total $\mathcal{M}^*$ at fixed redshift is therefore proportional to the sub-population which dominates around the knee in that interval.

The low-mass slope α_1 is roughly constant with time, although may experience a maximum at $z \approx 1$, in agreement with Davidzon et al. However, whereas Davidzon et al. finds a steeply decreasing α_1 at $z > 2$, we find a constant evolution which we then fix at $z > 3$. This is in good agreement with the α_1 found from our Continuity Model, as well as that of Leja et al. While α_1 of the star-forming and quiescent are similar to that of the total sample at $z < 1$, the α_1 of the quiescent sample rises sharply with z and turns down at $z \approx 1$. It appears to continue to turn down until $z \sim 3$ when it may stabilize, although a definitive assessment is not possible with our current constraints.

The low-mass normalization Φ_1 has little evolution at $0.2 < z \leq 1$, afterwards rapidly decreasing until it appears to approach zero asymptotically. Although Davidzon et al. finds slightly larger values of Φ_2 at low-z, both measurements generally agree on the rapid decline of the low-mass normalization. Φ_2 as derived from the Continuity Model finds still lower values and hence a more gradual evolution[9]. On the other hand, these parameter values agree with our fixed redshift measurements better. While Φ_2 of the quiescent sample is higher at $z < 1$, it declines more rapidly than that of the star-forming sample and remains subdominant at $z > 2$.

The high-mass slope α_2 is statistically consistent with being constant with z, but may rise slightly towards $z \sim 3$, thus flattening the low-mass regime. This is broadly comparable with Davidzon et al. within the stated uncertainties. Whereas α_2 derived from our Continuity Model closely follows that of the fixed redshift estimates at low-z, that of Leja et al. agrees better with our average α_2. Again, the α_2 of the star-forming and quiescent samples diverge around that of the total sample with α_2 of the quiescent sample being subdominant and that of the star-forming sample fluctuating about $\alpha_2 = 0$.

Lastly, the high-mass normalization Φ_2 generally declines over $z = 0.2 \rightarrow 3.0$. Interestingly, the Continuity Model finds an initial increase in Φ_2 towards $z \approx 1$, declining afterwards in agreement with the fixed redshift estimates. We find lower values compared to those of Davidzon et al. and Leja et al., which are in general agreement in part because they both use COSMOS2015. This suggests that the different evolution in Φ_2 between this work and the literature may be driven by subtle differences between the catalogs, sample selection, or both.

9 We note that a second order expansion cannot describe an exponential tail as seen here, which justifies $z \sim 3$ as the rightmost domain limit reachable by our 12 parameter Continuity Model.

Overall we find good agreement with the most directly comparable literature measurements from Davidzon et al. (2017) and Leja et al. (2019a). The implications for these evolutionary trends, and their relation to the growth of massive quiescent galaxies, are explored in detail in Sections 4.6.3 and 4.6.4.

4.6 DISCUSSION

In this section we focus on some of the open issues in galaxy evolution that can be addressed from a statistical perspective, using the results of Sect. 4.5 as a starting point. Besides the assembly of $z \approx 3-5$ massive galaxies and their quiescent fractions, we also discuss derivative measurements such as the cosmic stellar mass density ρ_*, as well as the relation between stellar and dark matter halo mass functions. We include several comparisons with simulations to provide not only physical insight but also to highlight areas for improvement from both sides.

4.6.1 *Cosmic stellar mass density evolution*

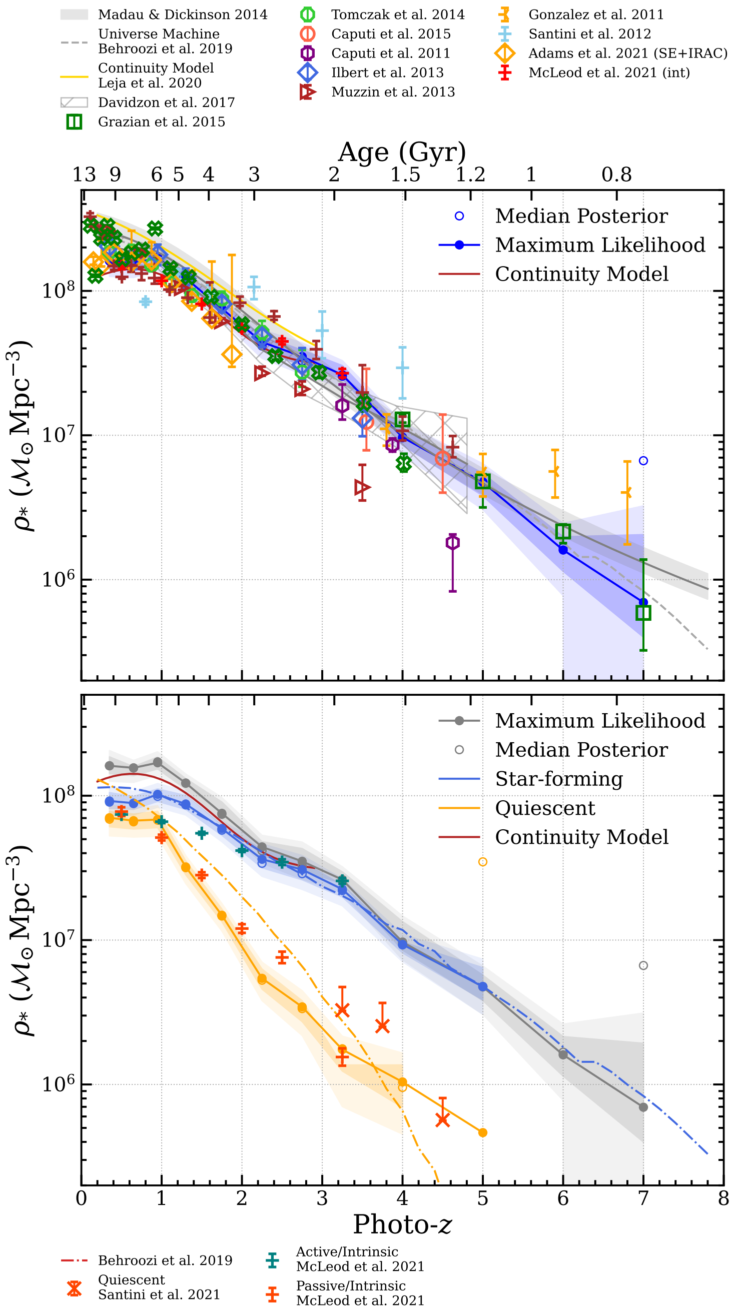

Figure 4.12: *Upper:* Evolution of the Cosmic Stellar Mass Density of the total sample computed from the best-fit likelihood models (blue) and continuity model (maroon) integrated above 10^8 $\mathcal{M}_\odot$ in both cases. Literature results from observational studies of mass-selected samples (Caputi et al., 2011; Santini et al., 2012; Ilbert et al., 2013; Muzzin et al., 2013b; Tomczak et al., 2014; Caputi et al., 2015; Grazian et al., 2015; Wright et al., 2018; Adams et al., 2021; McLeod et al., 2021) and mass inferred from rest-frame UV measurements (González et al., 2011). By integrating their SFRD functions, we can plot ρ_* from (Behroozi et al., 2013) and (Madau et al., 2014). In both cases we assume a return fraction of 41% (based on Chabrier's IMF). For Madau et al. (2014), we include a shaded area based on return fractions between 25-50% (the latter value is similar to the one given by Salpeter's IMF). *Lower:* Evolution of the Cosmic Stellar Mass Density of the total (grey, repeated from above), star-forming (blue), and quiescent (orange) samples compared to recent literature measurements (Behroozi et al., 2019; McLeod et al., 2021; Santini et al., 2021).

Galaxy mass assembly and growth is inextricably related to star formation. As such, reconciling direct measurements of galaxy mass growth with the behaviour of the star formation main sequence (e.g., Brinchmann et al., 2004; Daddi et al., 2007; Noeske et al., 2007; Salim et al., 2007; Whitaker et al., 2012, 2014) is of great interest. Accurate assessments of the empirical models of galaxy growth which link the two (e.g. Peng et al., 2010b; Behroozi et al., 2019) depends on unbiased, precise measurements of the integrated mass density ρ_* and its evolution with z.

We integrate the SMF measurements presented in Section 4.5 to derive an estimate of the stellar mass density (ρ_*) for each bin of z. Although definitions vary, ρ_* is commonly integrated down to $10^8\ \mathcal{M}_\odot$. Given that our $\mathcal{M}_{\rm lim}$ at all z is larger than $10^8\ \mathcal{M}_\odot$, and larger than $10^9\ \mathcal{M}_\odot$ above $z \approx 2$, we integrate the Schechter models for the total sample (Fig. 4.10) down to a mass range where the fit has been extrapolated. The resulting ρ_* is compared with other literature measurements in Fig. 4.12, converting to a Chabrier IMF where relevant. All SMF-based literature measurements of ρ_* have been re-integrated consistently to the same mass limit. We also show ρ_* derived from integrating the star formation rate density (SFRD) function of Madau et al. (2014), assuming a 41% return fraction.

We find remarkably good agreement with previous observational studies. At $z < 3$, the agreement seems to be limited by systematics as these are generally well-measured, secure samples. However, ρ_* at $z > 3$ is dominated by significantly less certain measurements, due both to the size of the samples and their typically noisy photometry. Our measurements place ρ_* near the midpoint of the scatter, in closest agreement with Davidzon et al. (2017) ($z < 5$) and Grazian et al. (2015) ($z < 7$). We report relatively large, but confidently robust uncertainties on ρ_* beyond $z \approx 4$, which stem from the increasing degeneracy of $\mathcal{M}_*$ and ϕ (α being fixed at $z > 3$). Because of their skew, median posterior results for the $z \sim 7$ Likelihood fit are unreliable, and so consequently so too is the corresponding ρ_*.

Our estimated median ρ_* as well as that of Grazian et al. are lower than the predictions of Madau et al. (2014) that based on measurements of the SFRD. This discrepancy could suggest that while star formation is high, the mass growth is lagging behind. As intriguing as this is, we stress that the measurements are consistent within 1σ, and that the SFRD constructed in Madau et al. are constrained by only two surveys at $z > 5$ (Bouwens et al., 2012a,b; Bowler et al., 2012). Further work utilizing larger, more mass complete samples will be required to confirm this divergence at $z > 7$.

In addition to ρ_* of the total sample, the lower panel of Fig. 4.12 includes ρ_* for the star-forming and quiescent samples. While stellar mass is overwhelmingly concentrated in star-forming systems from $z \approx 7 \rightarrow 3$, the mass density of quiescent systems grows rapidly until flattening at $z \approx 1$, consistent with Fig. 4.8. From $z = 1 \rightarrow 0$, there is little growth in ρ for both star-forming and quiescent systems where the former remains dominant over the latter.

While ρ_* of the star-forming sample is well constrained out to lower masses, the same cannot be said about ρ_* of the quiescent sample as its SMF is less mass complete in comparison. Since we cannot directly determine its low-$\mathcal{M}$ shape, we assume that the low-$\mathcal{M}$ Schechter component effectively vanishes at $z > 1.5$. If this is not the case, then we underestimate ρ_* for the quiescent sample. This may help explain the differences observed with respect to Behroozi et al. (2019) who report generally larger values of ρ_* for their quiescent sample, which features a significant quiescent population at low $\mathcal{M}$ (see Fig. 4.19). Although in agreement at $0.2 < z \leq 0.5$, Behroozi et al. report higher quiescent fractions at $< 10^{10.7}\mathcal{M}_\odot$ with redshift, finding more than 10× as many $10^9\ \mathcal{M}_\odot$ by $z \approx 1$. UNIVERSE MACHINE primarily defines quiescence as $sSFR < 10^{-11}\ \mathcal{M}yr^{-1}$, which according to Davidzon et al. (2018) is comparable our $NUVrJ$-selection, and so this overproduction of quiescent galaxies may indeed be at odds with our observations. Although not readily available, a consistently $NUVrJ$-selected sample from UNIVERSE MACHINE would clarify this sensitive comparison. Comparisons with the results of Santini et al. (2021) and McLeod et al. (2021), although generally in agreement, are also complicated by differences in selection. Whereas Santini et al. (2021) adopts a novel, multi-component SFR-driven selection, McLeod et al. (2021) adopts a UVJ selection following Carnall et al. (2018a). Therefore, we stress that estimates of ρ_* are particular to $NUVrJ$-selected quiescent systems, and our assumption of a single Schechter description at $z > 1.5$ may drive our relatively low estimates. Finally, despite differences in selection and constraining power, these observational studies collectively indicate a general agreement as to the evolution of ρ_* for quiescent systems.

4.6.2 *Comparison to simulations*

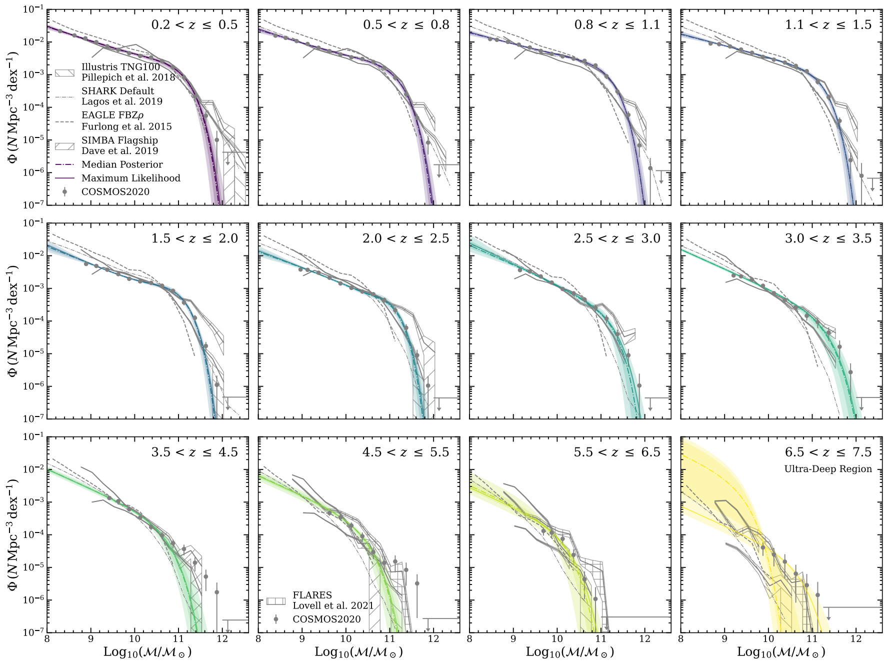

Figure 4.13: Comparison of observed and inferred galaxy stellar mass function (grey points, and colored curve with 1 and 2σ envelopes, respectively) to the reference flavours of four simulations: Illustris (TNG100; Pillepich et al., 2018), Eagle (FBZρ; Furlong et al., 2015), Shark (Default; Lagos et al., 2018), and Simba (Flagship; Davé et al., 2019). Upper limits for empty bins are shown by the horizontal grey line with an arrow. Mass incomplete measurements are not shown.

Observational constraints on the shape and evolution of the SMF may be useful in their own right, but we can also gain meaningful insight by comparing them with SMFs produced by simulations. Fig. 4.13 shows the SMF constraints from this work and the inferred SMF derived by the binned MCMC fits evaluated at maximum likelihood, with the flagship or reference versions of simulated SMFs overlaid: Illustris TNG100[10], Shark, Eagle, and Simba (Furlong et al., 2015; Lagos et al., 2018; Pillepich et al., 2018; Davé et al., 2019, respectively) as well as Flares ($z > 4.5$; Lovell et al., 2021; Vijayan et al., 2021, 2022). Since the estimates from our observed SMF are likely inflated by Eddington bias, the simulations should be considered primarily in comparison to our inferred SMF. We acknowledge that each simulation has multiple flavors with variations to their physical recipes. Although these variations provide additional insight, care must be taken when making these more complicated comparisons. As the goal of this work is to provide constraints on the observed and inferred SMF, we reserve comparisons to the variations of simulated SMFs for future work.

10 We choose this version as a compromise between resolution and volume. See Pillepich et al. 2018; Donnari et al. 2021a,b

Overall, we find the best agreement with Simba and Illustris TNG100. At $z < 1.5$ there is a slight preference towards Simba as Illustris TNG100 underestimates the number densities at $\mathcal{M}_*$ for $z \leq 1.5$. The situation is different at $1.5 < z \leq 3.0$ where Simba significantly overestimates the high-$\mathcal{M}$ end. While this could be explained by a systematic bias in observed masses or a missing high-$\mathcal{M}$ population (see Section 4.6.3), it could also be due to Simba potentially over-grouping dark matter haloes or insufficient AGN feedback (D. Narayanan, priv. comm.). Over this same range, it is apparent that Eagle suffers from volume limitations and thereby does not contain the most massive galaxies. Shark fares better as its semi-analytical prescriptions are able to produce high-$\mathcal{M}$ systems. In some z-bins, both Shark and Eagle assemble low-$\mathcal{M}$ systems too early relative to the observed SMF where there are direct constraints. We find a lesser degree of agreement at $z > 3.5$: while both Shark and Eagle underproduce at all $\mathcal{M}$, there is significant scatter between Illustris TNG, Simba, and Flares. Overall, Flares, Shark, and Eagle perform best below $\mathcal{M}_*$, with Simba and Flares performing best above $\mathcal{M}_*$. Meanwhile the volume limitations of Illustris TNG100 become significant at $z > 6$ (see Pillepich et al. 2018 for comparisons with TNG300).

It is remarkable that all of the simulations reproduce the $0.2 < z < 0.5$ SMF[11] , and yet they have severe disagreements at higher redshifts with number counts in disagreement by more than a factor of ten. While this should not come as a surprise given that simulations are typically tuned to reach an end state in agreement with the local universe, it suggests that their initial conditions and early evolutionary behavior are strikingly different. Hence, variously tuned physical recipes and initial conditions can produce the same end state. This apparent attractor behavior of modern simulations highlights the needed for continued deep observations to improve constraints on the SMF at high-z where simulations can be critically tested, and their physical thereby recipes refined. Our constraints indicate that while it appears that simulations are able to reproduce the observed abundance of high-$\mathcal{M}$ galaxies, the production of early low-$\mathcal{M}$ systems needs improvement.

4.6.3 *Abundant massive galaxies at $z \sim 3-5$*

One of the most striking results highlighted in Fig. 4.6 is the number density of massive $\mathcal{M} > 10^{11}\,\mathcal{M}_\odot$ galaxies not only at $z > 3.5$, but also at all redshifts. Although few in number, their identification in COSMOS2020 is a direct consequence of utilizing the larger 1.27 deg² now accessible with deep, homogeneous NIR coverage. While no $\mathcal{M} > 10^{11}$ galaxies are observed at $z > 5$, their growth since then has been similar to galaxies at other masses, as shown by Fig. 4.8. The majority of $\mathcal{M} > 10^{11}\,\mathcal{M}_\odot$ systems are star-forming at $z > 1$, as shown by Fig. 4.7 and quantified in Fig. 4.9, with only $z < 1$ systems at $\mathcal{M} \approx 10^{10.5-11}\,\mathcal{M}_\odot$ being equally divided between star-forming and quies-

11 With the exception Flares which is limited to $z > 4.5$.

cent states. We find evidence for a sustained population of massive quiescent systems at $z > 2$, but their number densities are dwarfed by that star-forming systems by a factor of ~ 10. The existence of these massive quiescent systems seems to defy the timescales expected for mass assembly (Steinhardt et al., 2016b; Schreiber et al., 2018b; Cecchi et al., 2019), and so their tremendously early formation and growth are a topic of great interest (Caputi et al., 2011; Hill et al., 2017; Toft et al., 2017; Carnall et al., 2018b; Tanaka et al., 2019; Valentino et al., 2020a; Whitaker et al., 2021; Akhshik et al., 2022; Marsan et al., 2022), including a more focused investigation into the origins of similarly selected massive galaxies from COSMOS2020 by Gould et al. (2022, in prep.).

Despite nearly identical sample selections, we still find greater number densities of massive galaxies compared to Davidzon et al. (2017) over $3 < z \leq 5.5$. By comparing to COSMOS2015 (matched within 0.6″), we find that 78% of $\mathcal{M} > 10^{11}\ \mathcal{M}_{\odot}$ galaxies in our sample are also found in COSMOS2015 where they are exclusively high-z (91% agree within $\Delta z \pm 0.5$) and high-mass (78% agree within $\Delta\mathcal{M} \pm 0.5\ \mathcal{M}_{\odot}$). The remaining 23% are new sources found only in COSMOS2020. As shown by Fig. 4.14, they are predominantly faint, having a median K_s magnitude ~ 24.2 AB compared to the median sample brightness $K_s \sim 23.4$ AB with a photometric uncertainty $0.05 - 0.1$ AB (see Fig 10 of Weaver et al. 2022a). They are also remarkably red, having a median $H - K_s$ color of ~ 1.2 AB compared to that of the sample overall (~ 0.8 AB). The most probable explanation for the new faint, red sources is that the deeper UltraVISTA $YJHK_s$ (as well as HSC iz) have enabled a the detection of these faint red sources which in COSMOS2015 could not be detected. Interestingly, we find that their distribution in mass is consistent with that of the total sample such that that number densities at all masses $\mathcal{M} > 10^{11}\ \mathcal{M}_{\odot}$ are proportionately represented.

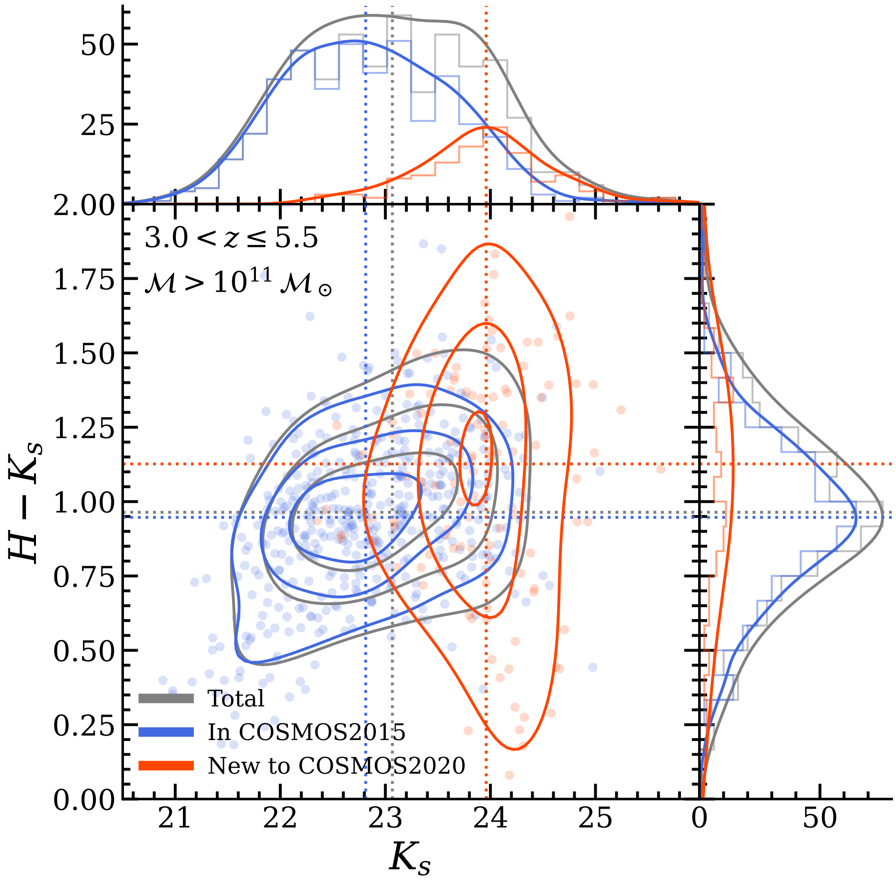

Figure 4.14: K_s magnitude and $H - K_s$ colors of 125 new massive $\mathcal{M} > 10^{11}\ \mathcal{M}_\odot$ galaxy candidates $3.0 < z \leq 5.5$ found in COSMOS2020 (orange), relative to the total sample used in this work (grey) and the 444 galaxies found in COSMOS2015 (blue). Measurements are taken from COSMOS2020, although they are similar to those from COSMOS2015, where matched. Number densities are summarised using gaussian kernel density estimators to produce smoothed distributions and contours for each sample.

Visual inspection of photometry and SED fits indicate that nearly all of the $3 < z \leq 5.5$ $\mathcal{M} > 10^{11}\ \mathcal{M}_\odot$ galaxies have red colors. About 75% of them are selected as star-forming by $NUVrJ$, $\gtrsim 80\%$ of which are dust-obscured ($A_V > 1$; compared to only $\sim 10\%$ across entire sample). The remaining $\sim 25\%$ are classified as quiescent. The red colors appear to be genuine and not driven by blends, as confirmed by visually inspecting the imaging for each galaxy. Their broad-band photometry lacks the strong spectral features that contribute to a secure photo-z: there is no detectable flux contamination by nebular emission lines and both the Lyman and Balmer breaks are weak. However, they are surprisingly well fit by `LePhare` ($\langle \chi^2_N \rangle \approx 1.5$). Their likelihood redshift distributions $\mathcal{L}(z)$ are narrow as $> 68\%$ of the probability is typically contained within $\Delta z \approx 1$, with similarly well constrained $\mathcal{L}(\mathcal{M}|z)$. Recently, Lower et al. (2020) has shown how the relatively simple parametric star formation histories assumed by most template-based SED fitting codes are susceptible to biases on the order of 0.5 dex in mass, which suggests that these uncertainties are likely underestimated.

One possible explanation for smooth SEDs with a power-law slope is contamination by AGN. The COSMOS2020 results computed by `LePhare` include classifications for sources with strong X-ray detections[12] based on Civano et al. (2016b). In our sample we use only those sources identified as galaxies, which excludes these X-ray detections. Their inclusion would only serve to increase these already surprisingly large number densities. We do not attempt to quantify this, as the expected accretion disc light from Type 1 Seyferts and quasars make estimates of photo-z and $\mathcal{M}$ unreliable and susceptible to catastrophic failures. See Weaver et al. (2022a) and Salvato et al. (2011) for details.

Identification of AGN (including X-ray faint AGN) from broadband colors has been explored in the literature (Stern et al., 2005; Donley et al., 2012; Hviding et al., 2022), and their discussion is a standard component of SMF studies at these redshifts (see Grazian et al., 2015; Davidzon et al., 2017), although a consensus has yet to be reached. In general, AGN selection criteria rely on colors derived from (near-)infrared wavelengths from most notably *Spitzer*/IRAC. While the Donley et al. criteria have been used successfully at $z < 3$, they require constraints in all four IRAC bands, which is not the case for COSMOS2020 where channel 3 and channel 4 lack sufficient depth to detect these sources. Even if deeper IRAC images could be taken, the selection criteria rely on the four bands sampling the continuum shape at restrame $2 - 10\,\mu$m but at $z > 3$ instead sample the rest-frame stellar bulk at $\sim 2\,\mu$m. While the MIPS $24\,\mu$m data from S-COSMOS (Sanders et al., 2007) is an attractive solution, it also is too shallow (20.2 AB at 3σ, Jin et al., 2018) to fully constrain the restrame $2-10\,\mu$m continuum at $3 < z \leq 5$: only galaxies $H \lesssim 20$ with large AGN fractions can be positively identified, and both low-fraction AGN at $H \approx 20$ and normal galaxies at $H < 20$ cannot be classified with MIPS. Full spectral fitting including far-infrared measurements is challenging without the constraints from channel 3, channel 4, and $24\,\mu$m, and attempts to gain further insight using Stardust (Kokorev et al., 2021) were broadly unsuccessful with possible contamination on the order of $10 - 30\%$. Removing these sources provides no significant change in the SMF. We are also able to leverage the elementary AGN template fitting from `LePhare` to statistically assess the spectral similarity between the best-fit AGN and galaxy templates for each source. While only 10% of galaxies in the total sample are best-fit with AGN templates, this fraction grows to 30% for all mass complete galaxies $3.5 < z < 5.5$ and then to 50% for those with $\mathcal{M} > 10^{10.8}\,\mathcal{M}_\odot$ with broad wings, having $> 80\%$ of these sources statistically indistinguishable as either galaxy or AGN ($|\Delta\chi^2| < 0.5$). Turning to further infrared data of individual sources, we find 15% of the sample are detected by VLA-COSMOS (Smolčić et al., 2017). While not conclusive, we find 5% of these massive $\mathcal{M} > 10^{11}\,\mathcal{M}_\odot$ galaxies at $3 < z \leq 5$ are detected in the ALMA maps of A^{3}COSMOS (Liu et al., 2019), which currently covers ~5% of the COSMOS field, highlighting the need for further observations. While similar sample statistics for individual sources have been extrapolated in the literature (e.g., Santini et al., 2021), anticipated

12 `lp_type=2`

surveys such as the TolTEC Ultra-Deep Galaxy Survey (Pope et al., 2019a) will expand and deepen the far-infrared coverage in COSMOS so that these samples can be more conclusively investigated.

A recent X-ray and radio stacking analysis of similarly selected $z > 1.5$ $\mathcal{M} > 10^{10}\, \mathcal{M}_\odot$ COSMOS2020 galaxies by Ito et al. (2022) revealed that low-luminosity AGN are likely ubiquitous in massive quiescent galaxies from $1.5 < z \leq 5$, even if individually they are not X-ray or radio detected. They also estimate the AGN contribution to optical/NIR continuum and find that at these redshifts, e.g., the rest-frame B-band luminosities of their quiescent galaxies are a factor of 30× larger than the expected rest-frame AGN luminosity. Ito et al. also find that the AGN luminosities for quiescent galaxies are significantly larger than those of star-forming galaxies. Together, their findings provide further confidence that our redshifts and stellar mass estimates for our X-ray undetected sample are not strongly contaminated by AGN light.

The lack of obvious systematics or likely only weak AGN contamination increases our confidence that these sources, or at least a part of them, are truly massive at $z \gtrsim 3$. In agreement with Ito et al. (2022), we also find that $> 60\%$ of $\mathcal{M} > 10^{11}\, \mathcal{M}_\odot$ galaxies appear to be star-forming, and it is likely that at least some of them are also dust obscured (DSFGs; Casey et al., 2014; Zavala et al., 2021, and references therein). Since there are a number of sources found in the deeper NIR images of COSMOS2020, it makes sense that we are now more sensitive to fainter red sources than ever before (see Fig. 4.14). It is precisely this class of galaxy which are efficiently captured by infrared facilities such as IRAC and ALMA, until now being optically "dark" (Schreiber et al., 2018c; Wang et al., 2019; Fudamoto et al., 2021; Sun et al., 2021; Shu et al., 2022). The existence of this population points to an incompleteness on the *massive*-end of the SMF in previous studies, and also highlights the sensitive interplay between the shape of their SEDs and the selection function consisting of the bands, their depths, and the detection methodology (see Fig. 3 of Weaver et al. 2022a); our deeper, redder NIR selection providing a greater degree of sample completeness by capturing these extremely red sources. Although detailed simulations will be explored in future work, qualitatively this could explain why the excess of sources relative to a Schechter at $z \approx 3-5$ diminishes at $z \approx 6$ as similarly red sources become too faint to still be detected. Optically-dark galaxies selected from 2mm ALMA observations in COSMOS from Casey et al. (2021) and Manning et al. (2022) are constrained to similar redshifts, stellar masses, and number densities. Importantly, Manning et al. studied two systems with star formation rates above $200\, \mathcal{M}_\odot\, \mathrm{yr}^{-1}$ but a gas depletion timescale $< 1\,\mathrm{Gyr}$, suggesting rapid star formation cessation and a potential transition to massive quiescent galaxies by $z \sim 3-4$. Casey et al. (2021) find that DSFGs are responsible for $\sim 30\%$ of the integrated star formation rate density at $3 < z < 6$ and that 2mm selection is an efficient way to identify larger samples in future surveys (see also Cooper et al. 2022). Ongoing theoretical work by Long et al. (in prep.) explores the effect of these obscured DSFGs by incorporating a dust-obscured component to the stellar mass function that extends beyond the knee to produce a total mass function remarkably similar to our findings. Lastly, recent work by Viero et al. (2022) found evidence for hot dust in similarly obscured systems at $z > 4$ in a stacking analysis of warm dust emission.

As introduced in Section 4.6.2 and shown in Fig. 4.13, there is generally good agreement between number densities of massive galaxies found in several simulations and in this work. This may be surprising given the range of physical recipes utilized in these simulations, and may suggest that several different tuning of physical recipes (mainly those of radiative AGN feedback) can reproduce observations. However from $2.5 < z \leq 5.5$, we observe galaxies with mass 1.8× larger than the most massive galaxies in any of these simulations. While bias and underestimated mass uncertainties may contribute, the simulations are usually considered to be limited by their volume. Fig. 4.15 compares the observed number densities of massive galaxies (including quiescent) to the upper limits set by volumes of different surveys as well as simulations. The fully hydrodynamical codes (Eagle, TNG100, and Simba) have the smallest volumes comparable to CANDELS and are similar to the observed number densities of $\mathcal{M} > 10^{11}\,\mathcal{M}_\odot$ galaxies. However, the fact that they contain a large enough volume to catch the most massive galaxy seen in our observations may suggest that their volumes are adequate, but that their DM halo growth physics are not providing the large halos from which they can grow. Alternatively, recent observations of a highly star-forming galaxy at $z = 6.9$ ($SFR \approx 2900\,\mathcal{M}_\odot\,\mathrm{yr}^{-1}$, Marrone et al., 2018) have presented a direct challenge to the capacity of simulations to produce similar systems expected evolve into the most massive systems at $z \approx 2-4$ (Lower et al. 2022, in prep.). This is especially true of Shark, whose volume is similar to that of COSMOS itself. More massive galaxies, if they exist, are not likely to be found in COSMOS, and so larger volumes (such as that probed by the two Euclid Deep Fields North and Fornax, as well as *Roman*) will be required (McPartland et al. 2022, in prep.).Visual inspection of photometry and SED fits indicate that nearly all of the $3 < z \leq 5.5$ $\mathcal{M} > 10^{11}\,\mathcal{M}_\odot$ galaxies have red colors. About 75% of them are selected as star-forming by $NUVrJ$, $\gtrsim 80\%$ of which are dust-obscured ($A_V > 1$; compared to only $\sim 10\%$ across entire sample). The remaining $\sim 25\%$ are classified as quiescent. The red colors appear to be genuine and not driven by blends, as confirmed by visually inspecting the imaging for each galaxy. Their broad-band photometry lacks the strong spectral features that contribute to a secure photo-z: there is no detectable flux contamination by nebular emission lines and both the Lyman and Balmer breaks are weak. However, they are surprisingly well fit by `LePhare` ($\langle \chi^2_N \rangle \approx 1.5$). Their likelihood redshift distributions $\mathcal{L}(z)$ are narrow as $> 68\%$ of the probability is typically contained within $\Delta z \approx 1$, with similarly well constrained $\mathcal{L}(\mathcal{M}|z)$. Recently, Lower et al. (2020) has shown how the relatively simple parametric star formation histories assumed by most template-based SED fitting codes are susceptible to biases on the order of 0.5 dex in mass, which suggests that these uncertainties are likely underestimated.

One possible explanation for smooth SEDs with a power-law slope is contamination by AGN. The COSMOS2020 results computed by `LePhare` include classifications for sources with strong X-ray detections[13] based on Civano et al. (2016b). In our sample we use only those sources identified as galaxies, which excludes these X-ray detections. Their inclusion would only serve to increase these already surprisingly large number densities. We do not attempt to quantify this, as the expected accretion disc light from Type 1 Seyferts and quasars make estimates of photo-z and $\mathcal{M}$ unreliable and susceptible to catastrophic failures. See Weaver et al. (2022a) and Salvato et al. (2011) for details.

Identification of AGN (including X-ray faint AGN) from broadband colors has been explored in the literature (Stern et al., 2005; Donley et al., 2012; Hviding et al., 2022), and their discussion is a standard component of SMF studies at these redshifts (see Grazian et al., 2015; Davidzon et al., 2017), although a consensus has yet to be reached. In general, AGN selection criteria rely on colors derived from (near-)infrared wavelengths from most notably *Spitzer*/IRAC. While the Donley et al. criteria have been used successfully at $z < 3$, they require constraints in all four IRAC bands, which is not the case for COSMOS2020 where channel 3 and channel 4 lack sufficient depth to detect these sources. Even if deeper IRAC images could be taken, the selection criteria rely on the four bands sampling the continuum shape at restrame $2 - 10\,\mu$m but at $z > 3$ instead sample the rest-frame stellar bulk at $\sim 2\,\mu$m. While the MIPS $24\,\mu$m data from S-COSMOS (Sanders et al., 2007) is an attractive solution, it also is too shallow (20.2 AB at 3σ, Jin et al., 2018) to fully constrain the restrame $2-10\,\mu$m continuum at $3 < z \leq 5$: only galaxies $H \lesssim 20$ with large AGN fractions can be positively identified, and both low-fraction AGN at $H \approx 20$ and normal galaxies at $H < 20$ cannot be classified with MIPS. Full spectral fitting including far-infrared measurements is challenging without the constraints from channel 3, channel 4, and $24\,\mu$m, and attempts to gain further insight using Stardust (Kokorev et al., 2021) were broadly unsuccessful with possible contamination on the order of $10-30\%$. Removing these sources provides no significant change in the SMF. We are also able to leverage the elementary AGN template fitting from `LePhare` to statistically assess the spectral similarity between the best-fit AGN and galaxy templates for each source. While only 10% of galaxies in the total sample are best-fit with AGN templates, this fraction grows to 30% for all mass complete galaxies $3.5 < z < 5.5$ and then to 50% for those with $\mathcal{M} > 10^{10.8}\,\mathcal{M}_\odot$ with broad wings, having $> 80\%$ of these sources statistically indistinguishable as either galaxy or AGN ($|\Delta\chi^2| < 0.5$). Turning to further infrared data of individual sources, we find 15% of the sample are detected by VLA-COSMOS (Smolčić et al., 2017). While not conclusive, we find 5% of these massive $\mathcal{M} > 10^{11}\,\mathcal{M}_\odot$ galaxies at $3 < z \leq 5$ are detected in the ALMA maps of A^{3}COSMOS (Liu et al., 2019), which currently covers ~5% of the COSMOS field, highlighting the need for further observations. While similar sample statistics for individual sources have been extrapolated in the literature (e.g., Santini et al., 2021), anticipated

13 `lp_type=2`

surveys such as the TolTEC Ultra-Deep Galaxy Survey (Pope et al., 2019a) will expand and deepen the far-infrared coverage in COSMOS so that these samples can be more conclusively investigated.

A recent X-ray and radio stacking analysis of similarly selected $z > 1.5$ $\mathcal{M} > 10^{10}\,\mathcal{M}_\odot$ COSMOS2020 galaxies by Ito et al. (2022) revealed that low-luminosity AGN are likely ubiquitous in massive quiescent galaxies from $1.5 < z \leq 5$, even if individually they are not X-ray or radio detected. They also estimate the AGN contribution to optical/NIR continuum and find that at these redshifts, e.g., the rest-frame B-band luminosities of their quiescent galaxies are a factor of 30× larger than the expected rest-frame AGN luminosity. Ito et al. also find that the AGN luminosities for quiescent galaxies are significantly larger than those of star-forming galaxies. Together, their findings provide further confidence that our redshifts and stellar mass estimates for our X-ray undetected sample are not strongly contaminated by AGN light.

The lack of obvious systematics or likely only weak AGN contamination increases our confidence that these sources, or at least a part of them, are truly massive at $z \gtrsim 3$. In agreement with Ito et al. (2022), we also find that $> 60\%$ of $\mathcal{M} > 10^{11}\,\mathcal{M}_\odot$ galaxies appear to be star-forming, and it is likely that at least some of them are also dust obscured (DSFGs; Casey et al., 2014; Zavala et al., 2021, and references therein). Since there are a number of sources found in the deeper NIR images of COSMOS2020, it makes sense that we are now more sensitive to fainter red sources than ever before (see Fig. 4.14). It is precisely this class of galaxy which are efficiently captured by infrared facilities such as IRAC and ALMA, until now being optically "dark" (Schreiber et al., 2018c; Wang et al., 2019; Fudamoto et al., 2021; Sun et al., 2021; Shu et al., 2022). The existence of this population points to an incompleteness on the *massive*-end of the SMF in previous studies, and hence why our deeper data provides a greater degree of sample completeness by capturing redder sources. If genuine, their existence not only points at an incompleteness in previous studies, but also highlights the sensitive interplay between the shape of their SEDs and the selection function consisting of the bands, their depths, and the detection methodology (see Fig. 3 of Weaver et al. 2022a). Although detailed simulations will be explored in future work, qualitatively this could explain why the excess of sources relative to a Schechter at $z \approx 3-5$ diminishes at $z \approx 6$ as similarly red sources become too faint to still be detected. Optically-dark galaxies selected from 2mm ALMA observations in COSMOS from Casey et al. (2021) and Manning et al. (2022) are constrained to similar redshifts, stellar masses, and number densities. Importantly, Manning et al. studied two systems with star formation rates above $200\,\mathcal{M}_\odot\,\mathrm{yr}^{-1}$ but a gas depletion timescale $< 1\,\mathrm{Gyr}$, suggesting rapid star formation cessation and a potential transition to massive quiescent galaxies by $z \sim 3-4$. Casey et al. (2021) find that DSFGs are responsible for $\sim 30\%$ of the integrated star formation rate density at $3 < z < 6$ and that 2mm selection is an efficient way to identify larger samples in future surveys (see also Cooper et al. 2022). Ongoing theoretical work by Long et al. (in prep.) explores the effect of these obscured DSFGs by incorporating a dust-obscured component to the stellar mass function that extends beyond the knee to produce a total mass function remarkably similar to our findings. Lastly, recent work by Viero et al. (2022) found evidence for hot dust in similarly obscured systems at $z > 4$ in a stacking analysis of warm dust emission.

As introduced in Section 4.6.2 and shown in Fig. 4.13, there is generally good agreement between number densities of massive galaxies found in several simulations and in this work. This may be surprising given the range of physical recipes utilized in these simulations, and may suggest that several different tuning of physical recipes (mainly those of radiative AGN feedback) can reproduce observations. However from $2.5 < z \leq 5.5$, we observe galaxies with mass $1.8\times$ larger than the most massive galaxies in any of these simulations. While bias and underestimated mass uncertainties may contribute, the simulations are usually considered to be limited by their volume. Fig. 4.15 compares the observed number densities of massive galaxies (including quiescent) to the upper limits set by volumes of different surveys as well as simulations. The fully hydrodynamical codes (EAGLE, TNG100, and SIMBA) have the smallest volumes comparable to CANDELS and are similar to the observed number densities of $\mathcal{M} > 10^{11}\,\mathcal{M}_{\odot}$ galaxies. However, the fact that they contain a large enough volume to catch the most massive galaxy seen in our observations may suggest that their volumes are adequate, but that their DM halo growth physics are not providing the large halos from which they can grow. Alternatively, recent observations of a highly star-forming galaxy at $z = 6.9$ ($SFR \approx 2900\,\mathcal{M}_{\odot}\,\mathrm{yr}^{-1}$, Marrone et al., 2018) have presented a direct challenge to the capacity of simulations to produce similar systems expected evolve into the most massive systems at $z \approx 2-4$ (Lower et al. 2022, in prep.). This is especially true of SHARK, whose volume is similar to that of COSMOS itself. More massive galaxies, if they exist, are not likely to be found in COSMOS, and so larger volumes (such as that probed by the two Euclid Deep Fields North and Fornax, as well as *Roman*) will be required (McPartland et al. 2022, in prep.).

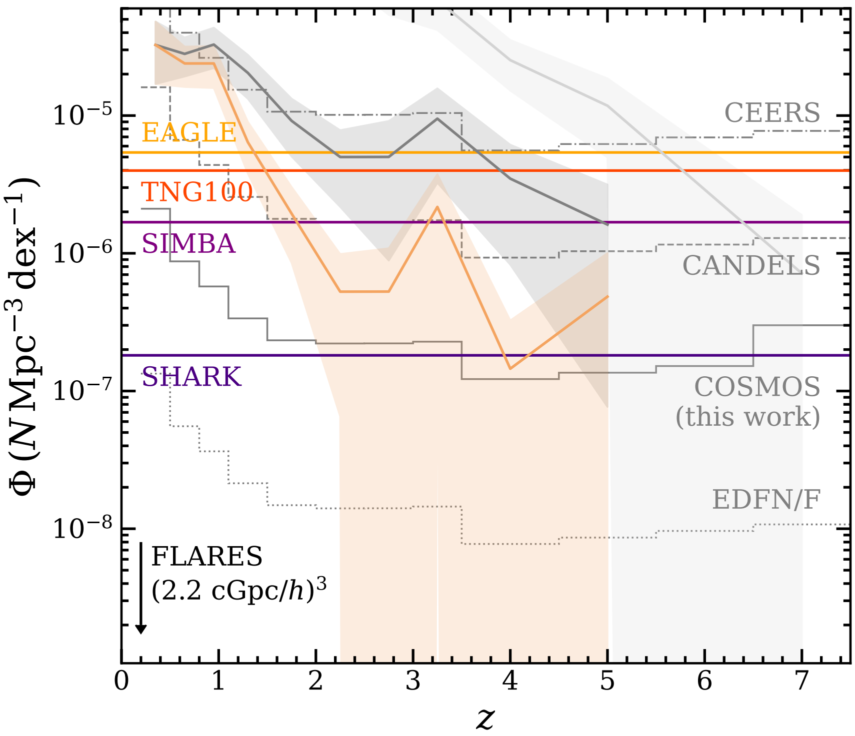

Figure 4.15: Comparison of observed number densities and upper limits on the rarity probed by various observed (grey steps) and simulated (colored lines) volumes. Number densities correspond to the two most massive bins from the total sample (light/dark grey), and the most massive bin from the quiescent sample (orange) from Fig. 4.8. Upper limits are computed following Gehrels (1986), which for the observed volumes are dependent on widths of redshift bins.

Naturally, massive galaxies in the $z > 3$ universe are of great interest as targets for *JWST*. The widest deep-field ERS program is the 100 arcmin² Cosmic Evolution Early Release Science Survey (CEERS; Finkelstein et al., 2017), and based on our estimates it stands to find approximately 22, 16, 5 for $\mathcal{M} > 10^{10.5}\,\mathcal{M}_{\odot}$, and 6, 4, 2 for $> 10^{11.0}\,\mathcal{M}_{\odot}$) galaxies at $3 < z \leq 3.5$, $3.5 < z \leq 4.5$, and $4.5 < z \leq 5.5$, respectively, although the small area will equate to a stronger density bias from cosmic variance. The widest galaxy survey field of any GO program is the 0.6 deg² COSMOS-Web[14] (Kartaltepe et al., 2021), which is expected to find 493 (137), 340, (94), 103 (40) galaxies for the same redshift ranges (and masses). Although *JWST* will be instrumental in studying these sources in exquisite spatial resolution and with efficient spectroscopy (Barrufet et al., 2021; Carnall et al., 2021; Glazebrook et al., 2021), ground-based NIR observations that can efficiently identify them in wide-area surveys will retain their importance.

14 Formerly COSMOS-Webb

4.6.4 *Rise of Quiescent Galaxies*

As shown by Fig. 4.7, low-z quiescent galaxies are well described by a two component Schechter function whose low-$\mathcal{M}$ component diminishes rapidly from $z \approx 0.2 \rightarrow 1$ (see Fig. 4.11). Simultaneously our sample becomes less mass complete with redshift, doing so more rapidly for quiescent systems due to their red color characteristic of their low mass-to-light ratios. Despite the considerable uncertainties on the completeness of our low-$\mathcal{M}$ quiescent sample outlined in Section 4.3.3, it seems likely that the shape of the quiescent stellar mass function in this work and in previous literature in fact does turn down at low-masses, with selection effects playing a comparably minor role (see also Ilbert et al. 2010). Given their low apparent brightness and rarity, this work is the first to quantify the number densities of such low-mass quiescent systems as it is based on the deepest NIR data taken over a degree-scale area required to find them in sufficient numbers. As seen in Fig. 4.8, the rate of growth in the number density of low-mass ($\mathcal{M} \lesssim 10^{10}\,\mathcal{M}_\odot$) quiescent galaxies has been seemingly rapid over the past ~ 10 billion years. Still, they constitute only a minor fraction of low-mass sources by $z \sim 0.2$, and by extrapolating the number densities to $z > 2$ one may expect to find none within COSMOS by $z \sim 3$. This is typically interpreted by the phenomenological model of Peng et al. (2010b) to mean that the processes which act to halt star formation cessation in low-$\mathcal{M}$ systems are inefficient at early times. For example the lack of virialized at $z > 2-3$ structure makes influence from environmental effects unlikely.

As shown by Fig. 4.8, the apparent lack of growth in the abundance of massive systems at $z < 1$ is the result of as a decline star-forming galaxies simultaneous with an increase in quiescent number densities. As noted by Ilbert et al. (2013), this decrease in the star-forming population is consistent with star formation cessation becoming extremely efficient, to the extent that massive star-forming galaxies are becoming quiescent faster than they can be replaced. Therefore, the mass assembly of massive star-forming systems at $z < 1$ is slower than the cessation of their star formation. However, there is also a slowing down in the rate of growth in the number density of massive quiescent systems themselves. While this may suggest high incidences of dry mergers or even rejuvination, it must also relate to evolving demographics: from $z \approx 1.5 \rightarrow 0.3$ there become fewer and fewer high-$\mathcal{M}$ star-forming galaxies *available* to become quiescent. While these the number density growth of massive systems seems to have stalled out, that of lower-mass systems continue to grow; this is the so-called phenomenon of "downsizing with time" (Cowie et al., 1996; Neistein et al., 2006; Fontanot et al., 2009).

The quiescent mass function at $\mathcal{M} > 10^{11}\mathcal{M}_\odot$ does not change much from $z \approx 4.5 \rightarrow 2.0$, with surprisingly elevated quiescent fractions (Fig. 4.9) being above 20% at $z < 5$. Fig. 4.8 shows why: while the number densities of star-forming galaxies at $z \approx 5$ is lower than at $z \approx 2$, quiescent galaxies are roughly constant in density over this same range. However, we caution that their number densities are only marginally above what should be possible to find the volume of COSMOS at these redshifts, and so it is plausible that they are dominated by noise and/or photo-z bias. The question of whether or not this behavior is genuine can only be explored in future large volume surveys, as demonstrated by Fig. 4.15. Even though the most massive galaxies at $z > 3-4$ are typically too faint to obtain continuum features, spectroscopic follow-up and supporting SED fitting will continue to provide valuable insights (Glazebrook et al., 2017; Schreiber et al., 2018a; Valentino et al., 2020a).

Fig. 4.16 shows the fraction of quiescent galaxies in bins of mass for three epochs: $z \approx 0.3$, 1.3, and 2.7. A key advantage of examining fractions is that they are less sensitive to the overall normalization of the simulation and biases from observations (e.g., in simulations: overproduction of all galaxy masses; in observations: systematics in effective survey volume), and provide additional insight which is obscured by simply comparing mass functions. Although comparisons to Universe Machine have been already discussed in Section 4.6.2, we introduce two samples of quiescent galaxies which we selected from Eagle and Shark with an $NUVrJ$ selection consistent with our methodology. While Eagle underproduces quiescent systems by $10-20\%$ at all masses, Shark overproduces in all but the most massive bins at $z \approx 0.3-1.3$ but underproduces at $z \approx 2.7$. Roughly, the fraction of quiescent galaxies in Shark at $z \approx 1.3$ matches the observed fractions at $z \approx 0.3$. This may suggest that Shark's physical recipes that halt star formation in lower mass galaxies are too aggressive. These same systems are also overproduced (see Fig. 4.13), and so they assemble *and* mature too early. For additional figures and details, see Appendix 4.B.

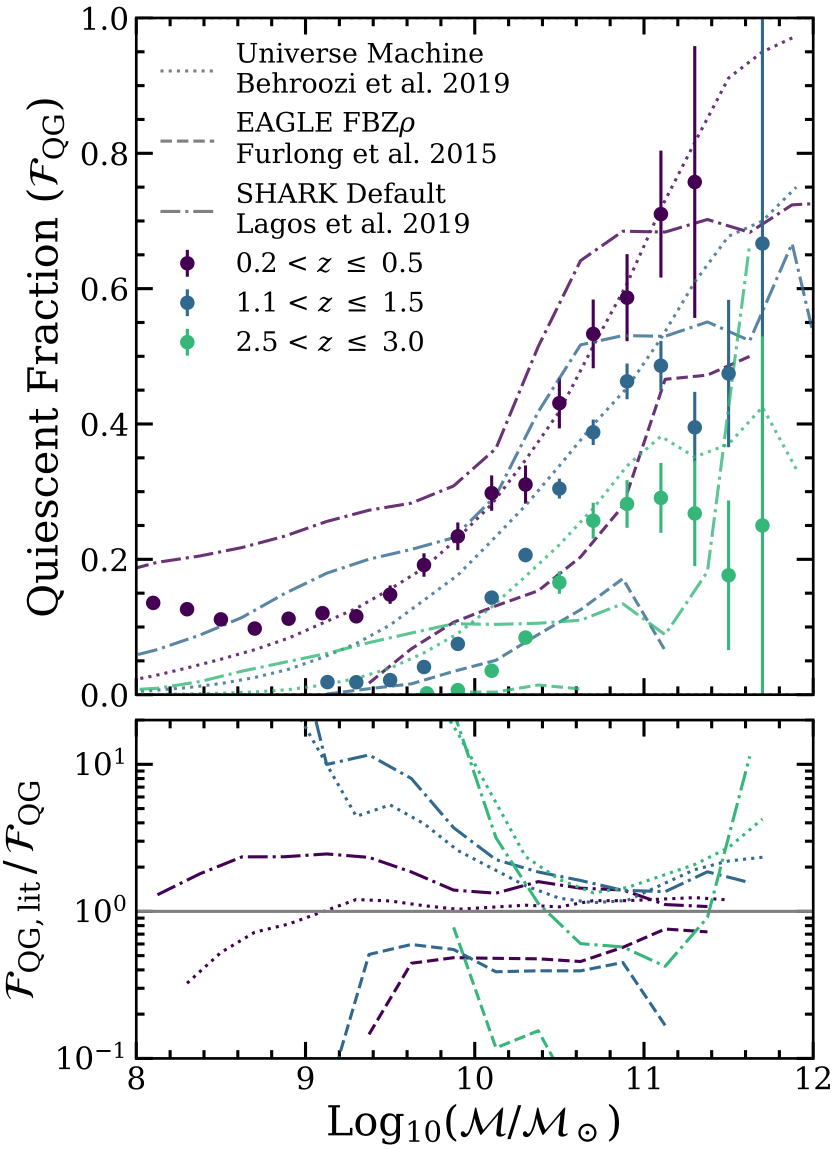

Figure 4.16: *Upper:* Fraction of quiescent galaxies as a function of mass for three redshift ranges compared with predictions from EAGLE (Furlong et al., 2015) and SHARK (Lagos et al., 2018) as well as measurements from UNIVERSE MACHINE (Behroozi et al., 2019) are shown for comparison. *Lower*: Fractional difference between this work and literature predictions/measurements.

4.6.5 *Dark matter halo connection*

The mass assembly of a galaxy is inherently connected to the dark matter halo in which it formed and grew (see Wechsler et al. 2018 for a review). Yet, stellar masses $\mathcal{M}$ have been observed to be < 20% of their halo mass $\mathcal{M}_h$, which point to galaxy formation as a strikingly inefficient process (Mandelbaum et al., 2006; Conroy et al., 2009c; Behroozi et al., 2010). This has led to investigations into the role of dark matter halo mass in driving star formation cessation, including promoting thermal heating and/or gas expulsion by AGN

(Main et al., 2017), as well as virial shock heating of in-falling molecular gas whose Jeans mass inhibits the formation of star-forming molecular clouds (so-called hot-halo mode, Birnboim et al., 2003). Hence, the gravitational influence of the halo mass on the cold gas reservoir regulates the ability of a galaxy to form stars, and hence the stellar-to-halo mass relationship (SHMR). Therefore it is no surprise that there is a similar evolution between the specific mass increase rate of the halo by accretion ($sMIR \equiv \mathcal{M}_{\rm h}^{-1}\partial\mathcal{M}_{\rm h}/\partial t$, Neistein et al. 2008; Saintonge et al. 2013) and the instantaneous mass growth by star formation ($sSFR$) (see discussions in Lilly et al., 2013). While in this work we restrict ourselves to comparisons with theoretical HMFs, another paper in this series computes a self consistent SHMR based on measuring the halo occupation distribution directly from angular correlation functions and SMFs of COSMOS2020 galaxies (Shuntov et al., 2022). For an investigation into the SHMR split by star-forming and quiescent samples, see Cowley et al. (2019), which is also based on galaxies from the COSMOS field.

As shown in Fig. 4.17, we compare our observed and inferred SMFs to the halo mass function (HMF) of Tinker et al. (2008)[15] from $z = 1.5 \rightarrow 7.5$. We choose not to show $z < 1.5$ as these comparisons have been thoroughly explored by previous investigations (e.g., Davidzon et al., 2017; Legrand et al., 2019) and we do not observe any significant differences. The effects of feedback can be seen in the first panel of Fig. 4.17 at $1.5 < z \leq 2.0$ that explain the relatively lower number densities at both low- and hig-$\mathcal{M}$ systems, with those around $\mathcal{M}_*$ being most similar to HMF. This apparent tension is a well known feature and lies at the foundation of the contemporary galaxy evolution paradigm, involving halting star formation by secular (internal) and/or environmental (external) action on the gas reservoir such a thermal heating, dynamical turbulence, and/or removal. Whereas the growth of galaxies $\mathcal{M} < \mathcal{M}_*$ can be impeded by secular (e.g., supernovae and stellar winds) and environmental processes (e.g., ram-pressure stripping, thermal evaporation), that of massive $\mathcal{M} > \mathcal{M}_*$ galaxies can be impeded by secular (e.g., AGN, bar formation) and environmental (e.g., major mergers) processes (see Peng et al., 2010b, 2012; Peng et al., 2015; Wechsler et al., 2018; Förster Schreiber et al., 2020). In this context, the characteristic knee is the result of a build-up of massive galaxies which can no longer sustain mass growth. At $z \approx 3$, the low-mass end is still considerably lower than the scaled HMF but the number density of massive systems comes into agreement. Although the stellar mass function slightly lies above the SHMR-scaled HMF at some points, we caution that this should not be taken as a challenge to theory as it assumes that the SHMR at $z = 0$ is appropriate at higher-z (which is unlikely; see Legrand et al., 2019) and small modifications can reconcile this difference. We note

15 The Tinker et al. (2008) HMF is computed according to our cosmology ($\sigma_8 = 0.82$) at from the mid-point in each z-bin using Colussus (Diemer, 2018), which explicitly takes into account that these mass functions derived were originally derived from spherical overdensities, which is not universal with redshift (See Equations 3-8 of Tinker et al.). We adopt the definition for a halo as the DM mass contain in a region 200× the mean matter density, and using other options (e.g., friends-of-friends, spherical overdensity, virial radius)

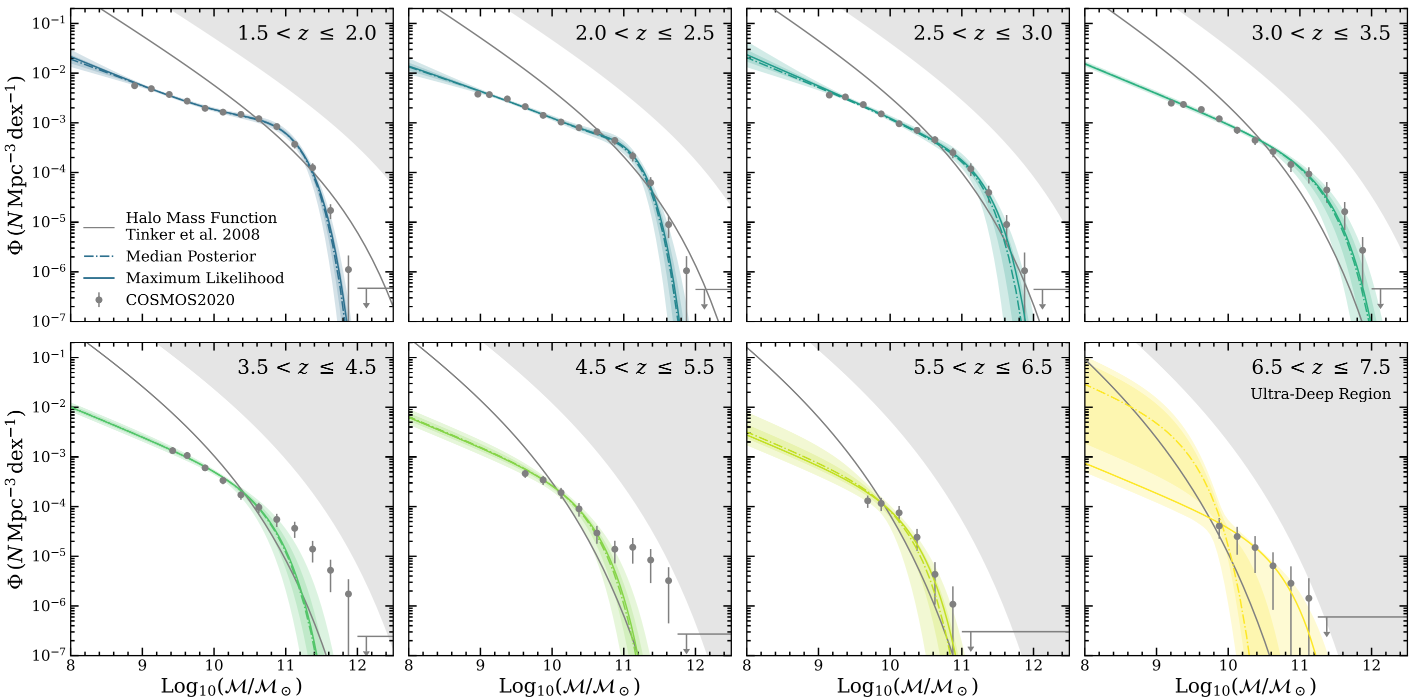

Figure 4.17: Comparison of the inferred $z > 1.5$ Galaxy Stellar Mass Function (colored curve with 1 and 2 σ envelopes) fitted to observed measurements (grey points). We superimpose an SMF in grey by scaled by the Halo Mass Function (HMF) of Tinker et al. (2008) by 0.018 according to the stellar-to-halo mass relation at $z = 0$ and $\mathcal{M}_{\rm h} = \mathcal{M}_{\rm h}^{*}$ (Behroozi et al., 2013). Upper limits on the SMF at each redshift are derived from the HMF assuming a fixed baryon fraction (0.166). Upper limits for empty bins are shown by the horizontal grey line with an arrow. Mass incomplete measurements are not shown.

that Davidzon et al. finds no such offset (see Fig. 18) using the same scaling, but we are unable to reproduce their precise result.

We derive an upper limit for the baryonic matter distribution by re-scaling the HMF by $\Omega_b/\Omega_m = 0.166$, which for our adopted cosmology is redshift independent, and assume a 100% efficient baryon-to-stellar mass conversion. This is the maximum SMF physically allowed under our simple assumptions. This upper limit becomes relevant especially at $z > 3.5$ where our observed number densities exceed those inferred by the Schechter model. While a large Eddington bias or selection systematic could explain this excess (see Section 4.6.3), we stress that our inferred SMF assumes the applicability of Schechter formalism, which cannot accurately describe the observed number densities at $3.5 < z \leq 5.5$. Nevertheless, the inferred stellar mass function agrees well with the SHMR-scaled HMF up to $z \approx 7$. This suggests that the most efficient haloes during these early epochs are not around $\mathcal{M}_*$, but rather the most massive ones, and with little observed evolution consistent with the findings from Stefanon et al. (2021b). One interpretation is that this high star formation efficiency in massive haloes is the result of diminished feedback from AGN, with stellar mass growing similarly to the host halo at these early times. Indeed, this is consistent with findings of inefficient radiative AGN feedback from simulations (Roos et al., 2015; Bieri et al., 2017; Kaviraj et al., 2017; Laigle et al., 2019; Habouzit et al., 2022), as well as FIR/radio observations of AGN activity at $z > 3$ (Maiolino et al., 2012; Cicone et al., 2014, 2015; Padovani et al., 2015; Vito et al., 2018).

At no point do our mass functions, observed or inferred, exceed this upper limit. Therefore we do not report evidence of "impossibly early galaxies" introduced by Steinhardt et al. (2016b) who point out that there appears to be too many massive galaxies at $z > 4$ compared to the dark matter haloes that should host them. However at $z > 6.5$, where Steinhardt et al. predicts that the effect will be most obvious, we report observed number densities approaching this upper limit and in clear excess of the SHMR-scaled HMF. We caution that these sources are the most vulnerable to misclassification and bias, being constrained by only a handful of NIR bands, and their Schechter fits are proportionately uncertain. Extrapolation to $\mathcal{M} > 10^{11.5}$ would place their number density below that which can be probed in a volume contained by the 0.716 deg^2 area of the Ultra-Deep region, and so COSMOS2020 is unlikely to find them if they exist. While they are not "impossibly early galaxies", their surprising abundance hint that explorations of $z > 7$ with future deep, large-volume surveys may provide the evidence necessary to firmly challenge theoretical frameworks.

4.7 SUMMARY & CONCLUSIONS

Following on from COSMOS2020 photo-z catalog (Weaver et al., 2022a), we study the shape and evolution of the galaxy stellar mass function. Our measurements span three decades in mass ($z = 0$) across 10 billion years of cosmic history ($z = 7.5 \rightarrow 0.2$), including the most mass complete sample of quiescent galaxies at $z > 2$ enabled by our unprecedented, homogeneous NIR depths across an effective 1.27 deg². We probe a volume nearly 2× that of Davidzon et al. (2017) which not only improves sample statistics but also finds new galaxies of still greater mass at all redshifts. Complementary deep IRAC coverage has allowed us to directly measure the stellar bulk, and hence galaxy mass, in a less biased way and to higher redshifts compared to K_s-based measurements. We developed a robust, mass-dependent error budget with contributions from poisson, stellar mass, and cosmic variance, and account for the effects of Eddington bias by fitting a kernel-convolved Schechter function to our observed SMF. We use three fitting techniques, most notably the continuity model of Leja et al. (2019a), finding good agreement with literature measurements with smaller bin-to-bin variance with z. We stress that our derived parameters are highly dependent on the assumed Eddington correction, and while the inferred SMF evaluated at maximum likelihood and associated parameters are robust, parameters (and their uncertainties) derived from the median of posterior distributions can be unreliable if constraints are weak.

Although literature comparisons on the shape of the SMF at fixed redshift show good agreement, the novel advantage of COSMOS2020 is the extended historical baseline over which the mass functions (as well as many other properties) can be consistently measured. Not only do we examine the evolution of the integrated mass density ρ_* over this time to find hints of tension with SFR measurements, we also examine the growth of galaxies at fixed mass over time to conclusively demonstrate the remarkably consistent rate of growth at all masses from $z \approx 7 \rightarrow 1$, whereupon the most massive star-forming galaxies become quiescent faster than they can be replaced. Similarly, we find evidence for the sharp rise in low-mass quiescent systems consistent with the phenomenological model of Peng et al. (2010b) probed to $z \approx 2.5$ where our sample become incomplete. Furthermore, we highlight three main results:

– Comparisons with five leading simulations indicates an exceptional degree of agreement for the most massive galaxies out to high-z. This comes despite the surprisingly high number densities of massive galaxies at $z \approx 3 - 5$ in excess of a Schechter function, suggesting that existing physical recipes are assembling massive $\mathcal{M} \approx 10^{10-11.5}\,\mathcal{M}_\odot$ systems in sufficient quantity at early times. However, a comparison of volume sizes suggests that simulations are sufficiently large to capture the most massive $\mathcal{M} > 10^{11.5}\,\mathcal{M}_\odot$ galaxies seen in observations, pointing to a need for even larger dark matter halos and/or more efficient star formation. In order to explore star formation cessation and feedback modes, we identify quiescent galaxies out to $z \approx 5.5$ by means of a $NUVrJ$ selection and compare them to consistently selected quiescent samples produced by two simulation codes, finding evidence for delayed assembly of low-mass quiescent systems in Eagle, and too rapid assembly in the Shark.

We also find tentative evidence for a sustained $> 20\%$ fraction of high-mass quiescent systems from $z \approx 5 \rightarrow 2$.

– A closer examination of these massive systems reveals that a quarter are not found in COSMOS2015. Not only are they K_s-faint, but their extremely red colors challenge SED fitting templates. We find no strong evidence for AGN contamination, although we stress the need for future infrared facilities with deep surveys capable of measuring the rest-frame MIR light at $z \approx 3 - 5$. Recent findings of optically-dark galaxies from IRAC and ALMA suggest that previous studies have missed contributions from dust-obscured star-forming galaxies. Their brightness, redshifts, mass, and number densities are consistent with our findings, suggesting that the $K_s \sim 26$ depth of UltraVISTA DR4 may indeed be sufficient to reach out into the tail end of this population missed by previous optical-NIR selections. Further work is required to conclusively establish the nature of these massive galaxy candidates.

– Lastly, we investigate the connection to dark matter halos by comparing both our observed and inferred SMFs to constraints provided by the HMF. While we confirm the divegence of the SMF from the HMF at both low- and high-mass regimes which has been historically intepreted as evidence for feedback processes, the massive end of the inferred, Schechter-fit SMF comes into agreement with the HMF at $z \gtrsim 2$. While we find no evidence of tension which would challenge theoretical models, our observed number densities at $z \approx 3 - 5$ approach the upper limit for fully efficient star formation in the most massive halos. Larger volume surveys capable of detecting rarer population of even more massive systems, if they exist, may be able to challenge these models, especially at $z > 6 - 7$.

The launch of *JWST* has opened the door on a new era, and it will soon be flanked by efficient survey facilities from space (*Euclid*, *Roman*) and the ground (*Rubin*). While massive quiescent systems may exist at $z \sim 5$ and perhaps even at earlier times (Mawatari et al., 2016, 2020), their identification is beyond the reach of COSMOS2020. They may be identified soon by deep degree-scale *JWST* surveys i.e., COSMOS-Web (Kartaltepe et al., 2021) and possibly by narrower ones including the *JWST* Advanced Deep Extragalactic Survey (JADES, Eisenstein et al., 2017; Ferruit, 2017), the Cosmic Evolution Early Release Science Survey (CEERS, Finkelstein et al., 2017), the Next Generation Deep Extragalactic Exploratory Public Survey (NGDEEP, Finkelstein et al., 2021)[16], and Ultra-deep NIRCam and NIRSpec Observations Before the Epoch of Reionization (UNCOVER, Labbe et al., 2021) to name a few.

It remains to be seen if such low mass quiescent systems exist at $z > 2.5$. By linearly extrapolating these observations, $< 1\%$ of $\mathcal{M} \approx 10^{9.5-10.0}\,\mathcal{M}_\odot$ are expected to be quiescent by $z \sim 2.5$, become even rarer at earlier times. While

16 Formerly WDEEP

identifying even one quiescent, low-mass system at $z > 2$ in the absence of virialized structure would present a significant challenge to paradigm of Peng et al., performing a statistically meaningful survey of them will require incredibly deep, degree-scale NIR surveys, placing them out of reach by current facilities. While the deepest degree-scale surveys from *JWST* (COSMOS-Web), *Roman*, and *Euclid* stand to establish the rarity of these systems at $z \approx 2-3$, it seems that no currently planned survey will be able to probe their contribution at $z > 3$.

While explorations of mass-selected samples at $z > 7$ will be made possible by *JWST*, we can already explore UV-selected sources at these extraordinarily early times (Kauffmann et al., submitted). Doing so will not only allow us to measure the star formation rate and dust content of the first ultra-luminous galaxies from deep within the epoch of reionization, but also to directly identify the progenitors of $z \sim 3-4$ massive galaxies while still in their formation stage. Cycle 1 observations with *JWST* from the Beasts in the Bubbles program (BEASTS, Weaver et al., 2021) will explore in detail the properties and assembly history of five such ultra-luminous galaxies at $z \sim 9$. Despite the incredible promises of highly resolved NIR imaging and spectroscopy from *JWST*, the rarest and potentially most informative populations that can challenge and thereby improve galaxy formation models will remain the domain of wide-field surveys.

ACKNOWLEDGEMENTS

The authors thank Joel Leja, Vadim Ruskov, Francesco Valentino, Sidney Lower, Deskia Narayanan, and Caitlin Casey for helpful discussions.

The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. ST and JW acknowledge support from the European Research Council (ERC) Consolidator Grant funding scheme (project ConTExt, grant No. 648179). OI acknowledges the funding of the French Agence Nationale de la Recherche for the project “SAGACE”. HJMcC acknowledges support from the PNCG. ID has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 896225. This work used the CANDIDE computer system at the IAP supported by grants from the PNCG and the DIM-ACAV and maintained by S. Rouberol. BMJ is supported in part by Independent Research Fund Denmark grant DFF - 7014-00017. GEM acknowledges the Villum Fonden research grant 13160 “Gas to stars, stars to dust: tracing star formation across cosmic time”. DR acknowledges support from the National Science Foundation under grant numbers AST-1614213 and AST-1910107.

The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. This work is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO program ID 179.A-2005 and on data products produced by CALET and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium. This work is based in part on observations made with the NASA/ESA *Hubble* Space Telescope, obtained from the Data Archive at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. Some of the data presented herein were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation.

This research is also partly supported by the Centre National d'Etudes Spatiales (CNES). These data were obtained and processed as part of the CFHT Large Area U-band Deep Survey (CLAUDS), which is a collaboration between astronomers from Canada, France, and China described in Sawicki et al. (2019, [MNRAS 489, 5202]). CLAUDS is based on observations obtained with MegaPrime/ MegaCam, a joint project of CFHT and CEA/DAPNIA, at the CFHT which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. CLAUDS uses data obtained in part through the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories, Chinese Academy of Sciences, and the Special Fund for Astronomy from the Ministry of Finance of China. CLAUDS uses data products from TERAPIX and the Canadian Astronomy Data Centre (CADC) and was carried out using resources from Compute Canada and Canadian Advanced Network For Astrophysical Research (CANFAR).

4.A SAMPLE SELECTION

As discussed in Section 4.3, we select sources from the 1.27 deg² `COMBINED` region for sources $0.2 < z \leq 6.5$, and impose an additional restriction of the ultra-deep stripe for sources $6.5 < z \leq 7.5$. The three remaining criteria are aimed to ensure secure photo-z and $\mathcal{M}$ based on the IRAC ch1 magnitude, SED χ^2 fit quality, and $\mathcal{L}(z)$. As shown in Figure 4.18, there is considerable overlap between these three criteria, with $m_{ch1} > 26$ accounting for 94% of the sample removed by combining these three.

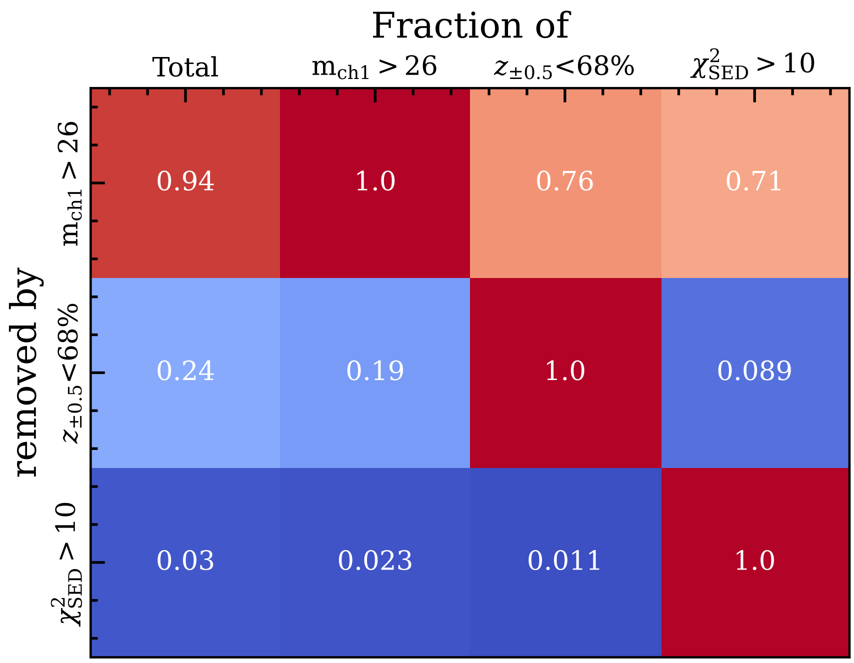

Figure 4.18: Fraction of sources removed by one criteria which are also removed by another, in addition to the primary criteria on area and photo-z.

4.B FURTHER COMPARISONS WITH SIMULATIONS

Here we include additional figures comparing the total, star-forming, and quiescent mass functions (Fig. 4.19) and quiescent fractions (Fig. 4.20) from this work to those of UNIVERSE MACHINE Behroozi et al. (2019, $sSFR < 10^{-11}\, M_{\odot}\,\mathrm{yr}^{-1}$) as well as those of consistently $NUVrJ$-selected quiescent galaxies from EAGLE (Furlong et al., 2015) and SHARK (Lagos et al., 2018).

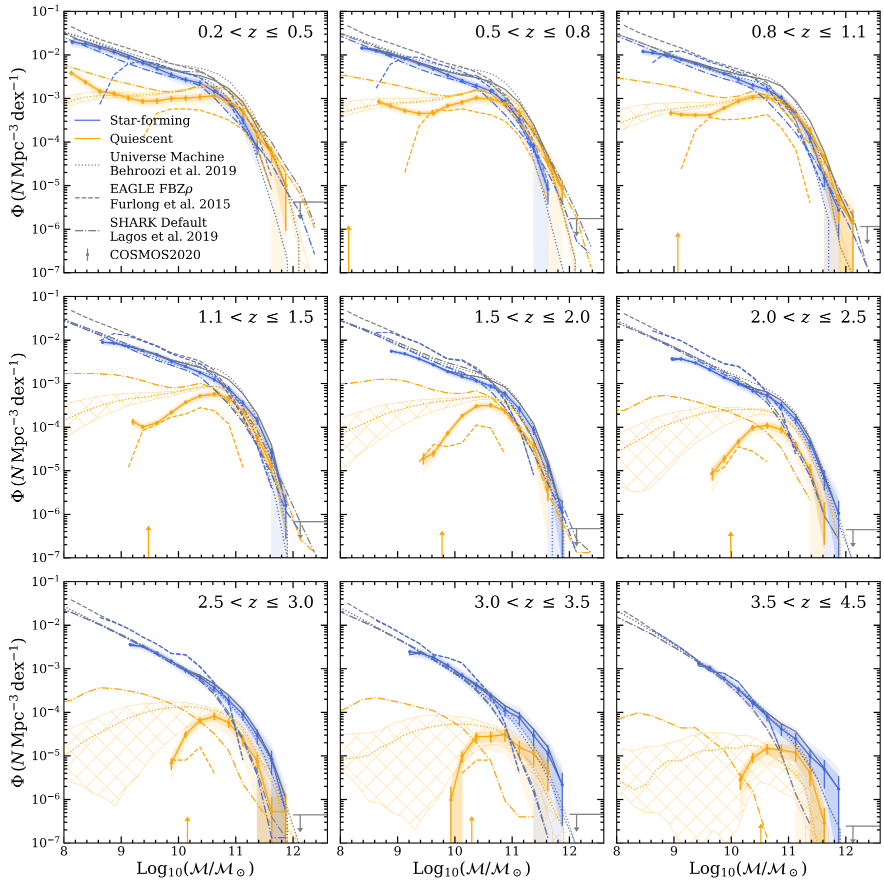

Figure 4.19: Galaxy stellar mass functions for star-forming (blue) and quiescent (orange) samples from this work compared to those from Universe Machine Behroozi et al. (2019, $sSFR < 10^{-11}\ M_\odot\,\mathrm{yr}^{-1}$), as well as consistently $NUVrJ$-selected samples from Eagle (Furlong et al., 2015) and Shark (Lagos et al., 2018). Upper limits for empty bins are shown by the horizontal grey line with an arrow. Mass incomplete measurements are not shown.

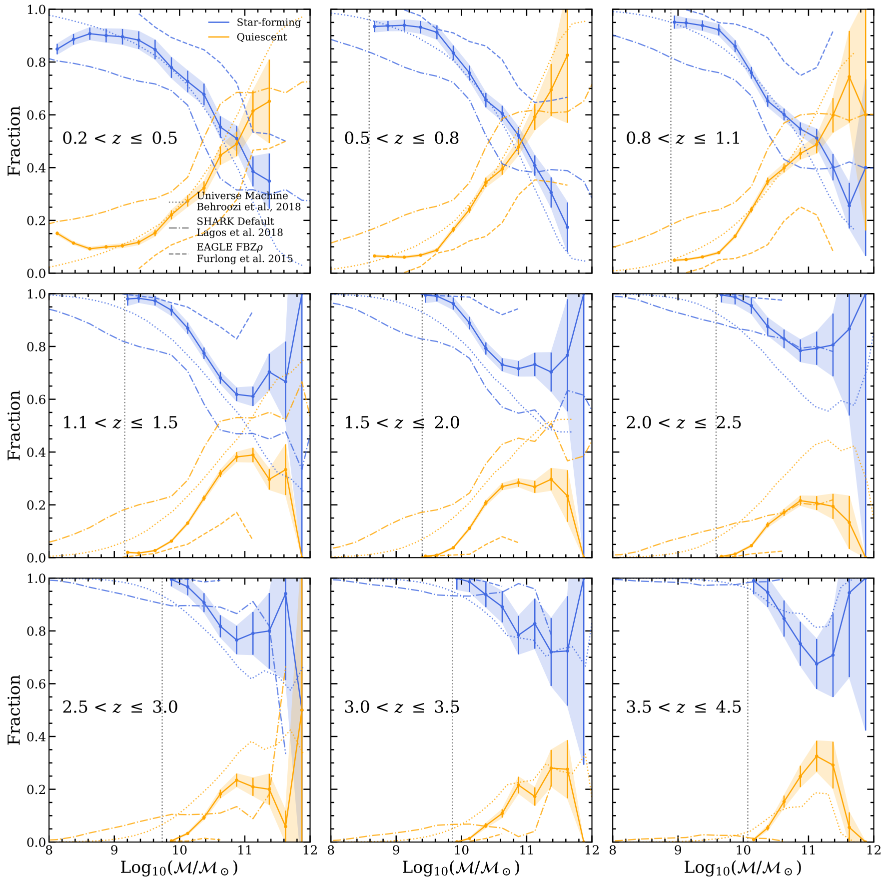

Figure 4.20: Fraction of star-forming (blue) and quiescent (orange) galaxy samples as a function of mass at for three redshift ranges. Measurements of consistently $NUVrJ$-selected samples from Eagle (Furlong et al., 2015) and Shark (Lagos et al., 2018) are included for comparison.

4.C FITTING RESULTS

Here we include the inferred total, star-forming, and quiescent mass functions (Figures 4.21, 4.22, and 4.23) optimized by both χ^2 and Likelihood maximization; their derived parameters are contained in Tables 6, 7, and 8, respectively. These results are summarized together in Fig. 4.10. Shown in the same figure are the SMFs realized by evaluation of the Continuity Model (based on Leja et al., 2019a), with related parameters contained in Table 9. Although not shown, full Markov chains and parameter corner diagrams are available upon request.

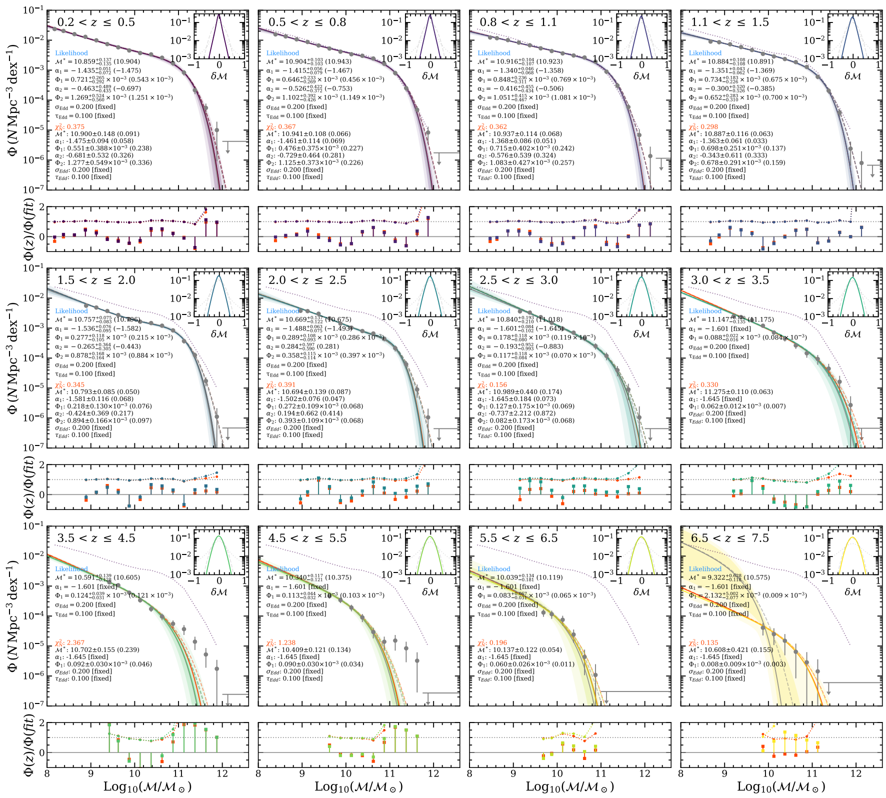

Figure 4.21: Results of fitting a double ($z < 3$) and single Schechter ($z \geq 3$) functional forms to the observed Galaxy Stellar Mass Function, binned in redshift, inferred from both χ^2 minimization (red) and maximum likelihood estimator (evolving colors) methods. Both Schechter functions are convolved with parameterized, redshift-dependent kernels in δM (upper right subpanels) to account for Eddington bias (dashed lines) and is then removed in the corrected fit (solid lines). Differences between the kernel convolved fit and the data are shown by both the relative fractional difference (dashed curve) and the difference weighted by the uncertainty (square points) are shown in the lower panels.

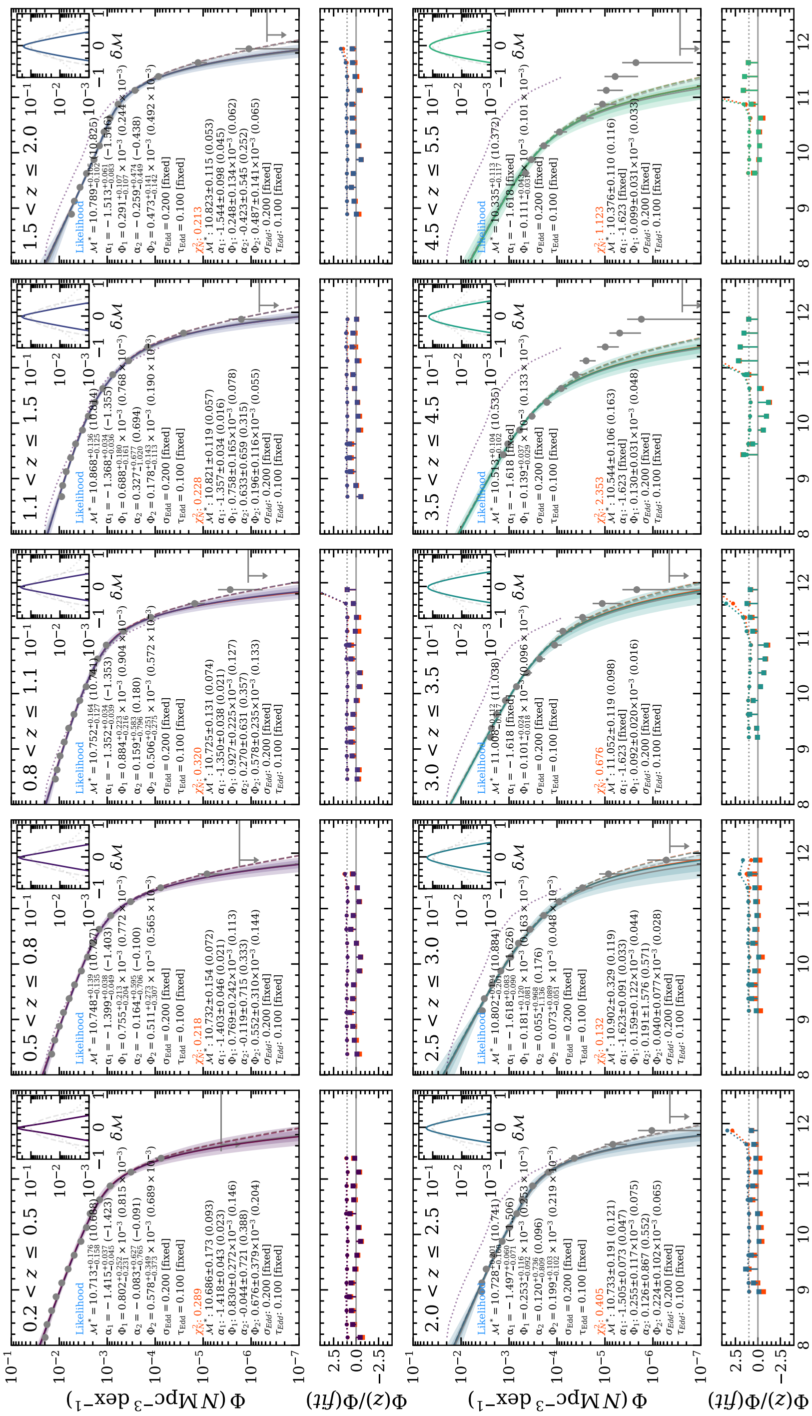

Figure 4.22: Results of fitting a double ($z \leq 3.0$) and single Schechter ($z > 3.0$) functional forms to the star-forming sample. Same layout as Fig. 4.21.

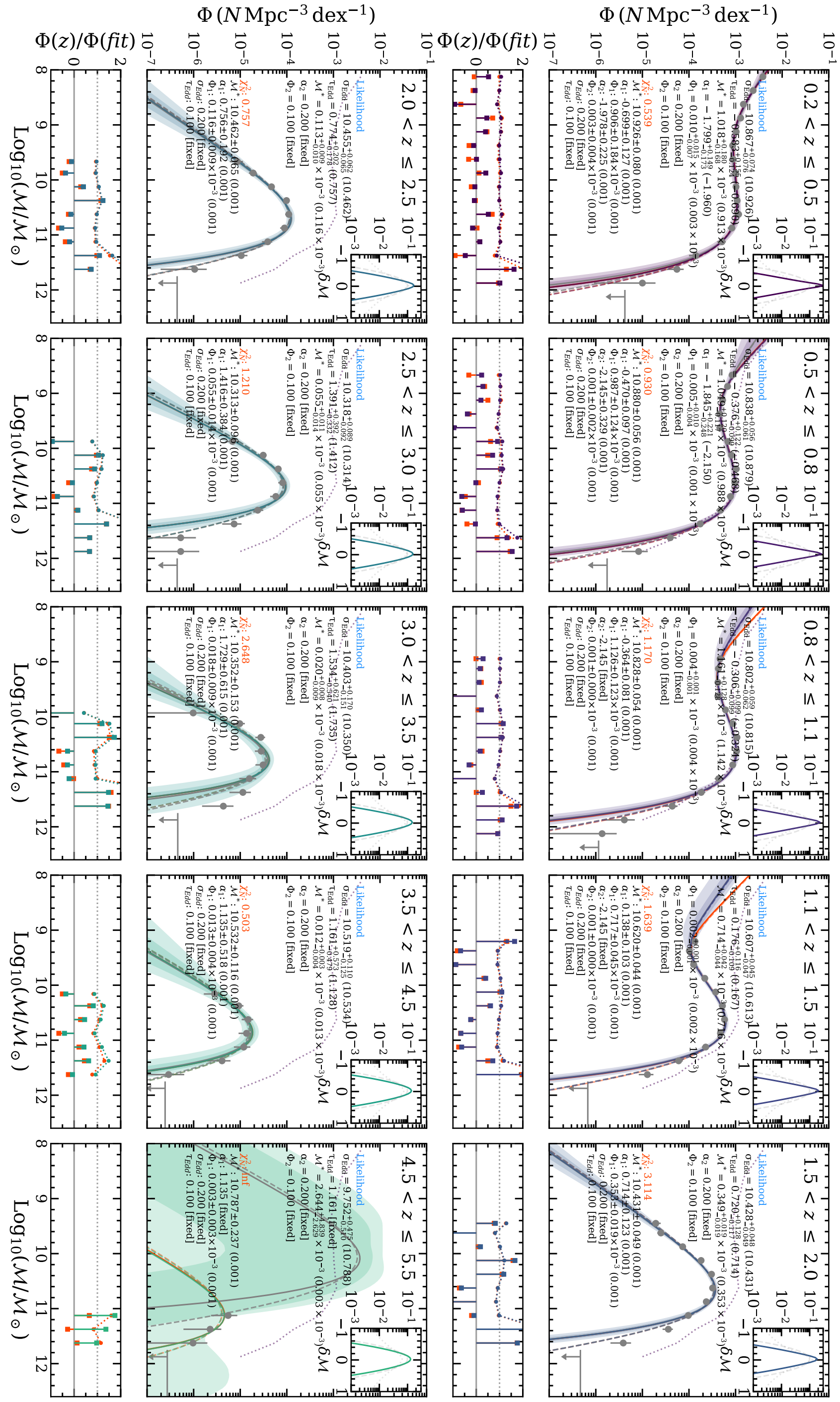

Figure 4.23: Results of fitting a double ($z \leq 1.5$) and single Schechter ($z > 1.5$) functional forms to the quiescent sample. Same layout as Fig. 4.21.

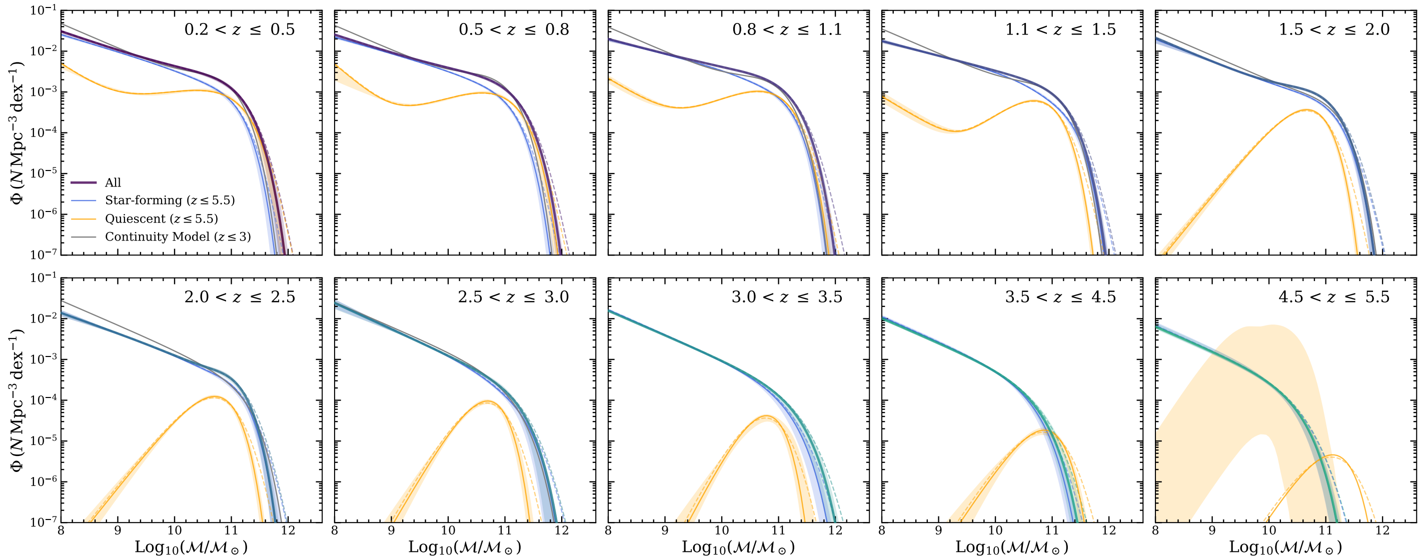

Figure 4.24: Best fit Schechter models corresponding to maximum likelihood for the total, star-forming, and quiescent samples. Fits for the median posterior will be similar for symmetric parameter posterior distributions, which are found in all but the last bin shown here. Kernel convolved fits are shown by the dashed curves.

z-bin	$\text{Log}_{10}\,\mathcal{M}^*$ ($\mathcal{M}_\odot$)	α_1	$\Phi_1 \times 10^3$ ($\text{Mpc}^{-3}\,\text{dex}^{-1}$)	α_2	$\Phi_2 \times 10^3$ ($\text{Mpc}^{-3}\,\text{dex}^{-1}$)	$\rho_*(\mathcal{M} \geq 10^8) \times 10^{-7}$ ($\mathcal{M}_\odot\,\text{Mpc}^{-3}$)
Likelihood Fit						
$0.2 < z \leq 0.5$	$10.86^{+0.14}_{-0.14}[10.90]$	$-1.44^{+0.05}_{-0.07}[-1.47]$	$0.72^{+0.27}_{-0.29}[0.54]$	$-0.46^{+0.49}_{-0.44}[-0.70]$	$1.27^{+0.52}_{-0.56}[1.25]$	$15.91^{+1.95}_{-1.66}$
$0.5 < z \leq 0.8$	$10.90^{+0.10}_{-0.10}[10.94]$	$-1.41^{+0.06}_{-0.08}[-1.47]$	$0.65^{+0.23}_{-0.27}[0.46]$	$-0.53^{+0.42}_{-0.37}[-0.75]$	$1.10^{+0.39}_{-0.43}[1.15]$	$15.17^{+1.64}_{-1.40}$
$0.8 < z \leq 1.1$	$10.92^{+0.10}_{-0.11}[10.92]$	$-1.34^{+0.05}_{-0.07}[-1.36]$	$0.85^{+0.24}_{-0.31}[0.77]$	$-0.42^{+0.45}_{-0.43}[-0.51]$	$1.05^{+0.41}_{-0.47}[1.08]$	$16.79^{+1.62}_{-1.47}$
$1.1 < z \leq 1.5$	$10.88^{+0.11}_{-0.11}[10.89]$	$-1.35^{+0.04}_{-0.06}[-1.37]$	$0.73^{+0.18}_{-0.23}[0.68]$	$-0.30^{+0.52}_{-0.53}[-0.39]$	$0.65^{+0.28}_{-0.31}[0.70]$	$11.95^{+1.01}_{-0.85}$
$1.5 < z \leq 2.0$	$10.76^{+0.07}_{-0.08}[10.80]$	$-1.54^{+0.08}_{-0.10}[-1.58]$	$0.28^{+0.12}_{-0.11}[0.21]$	$-0.26^{+0.36}_{-0.31}[-0.44]$	$0.88^{+0.17}_{-0.16}[0.88]$	$7.52^{+0.66}_{-0.64}$
$2.0 < z \leq 2.5$	$10.67^{+0.14}_{-0.12}[10.67]$	$-1.49^{+0.06}_{-0.07}[-1.49]$	$0.29^{+0.11}_{-0.09}[0.29]$	$0.28^{+0.60}_{-0.59}[0.28]$	$0.36^{+0.12}_{-0.11}[0.40]$	$4.36^{+0.51}_{-0.48}$
$2.5 < z \leq 3.0$	$10.84^{+0.19}_{-0.21}[11.02]$	$-1.60^{+0.08}_{-0.10}[-1.65]$	$0.18^{+0.12}_{-0.08}[0.12]$	$-0.19^{+0.95}_{-0.99}[-0.88]$	$0.12^{+0.12}_{-0.08}[0.07]$	$3.42^{+0.47}_{-0.39}$
$3.0 < z \leq 3.5$	$11.15^{+0.12}_{-0.14}[11.17]$	-1.60	$0.09^{+0.02}_{-0.02}[0.08]$	–	–	$2.79^{+0.25}_{-0.33}$
$3.5 < z \leq 4.5$	$10.59^{+0.14}_{-0.13}[10.61]$	-1.60	$0.12^{+0.04}_{-0.03}[0.12]$	–	–	$0.98^{+0.13}_{-0.12}$
$4.5 < z \leq 5.5$	$10.34^{+0.12}_{-0.12}[10.38]$	-1.60	$0.11^{+0.04}_{-0.03}[0.10]$	–	–	$0.46^{+0.09}_{-0.07}$
$5.5 < z \leq 6.5$	$10.04^{+0.13}_{-0.18}[10.12]$	-1.60	$0.08^{+0.07}_{-0.03}[0.07]$	–	–	$0.15^{+0.05}_{-0.04}$
$6.5 < z \leq 7.5$	$9.32^{+0.63}_{-0.12}[10.58]$	-1.60	$2.13^{+5.00}_{-2.08}[0.01]$	–	–	$0.09^{+0.12}_{-0.05}$
χ^2 Fit						
$0.2 < z \leq 0.5$	$10.90 \pm 0.09(0.15)$	$-1.47 \pm 0.06(0.09)$	$0.55 \pm 0.24(0.39)$	$-0.68 \pm 0.33(0.53)$	$1.28 \pm 0.34(0.55)$	$16.17^{+8.08}_{-4.30}$
$0.5 < z \leq 0.8$	$10.94 \pm 0.07(0.11)$	$-1.46 \pm 0.07(0.11)$	$0.48 \pm 0.23(0.38)$	$-0.73 \pm 0.28(0.46)$	$1.13 \pm 0.23(0.37)$	$16.36^{+5.64}_{-5.76}$
$0.8 < z \leq 1.1$	$10.94 \pm 0.07(0.11)$	$-1.37 \pm 0.05(0.09)$	$0.71 \pm 0.24(0.40)$	$-0.58 \pm 0.32(0.54)$	$1.08 \pm 0.26(0.43)$	$18.45^{+6.89}_{-5.14}$
$1.1 < z \leq 1.5$	$10.89 \pm 0.06(0.12)$	$-1.36 \pm 0.03(0.06)$	$0.70 \pm 0.14(0.25)$	$-0.34 \pm 0.33(0.61)$	$0.68 \pm 0.16(0.29)$	$12.01^{+2.80}_{-2.34}$
$1.5 < z \leq 2.0$	$10.79 \pm 0.05(0.08)$	$-1.58 \pm 0.07(0.12)$	$0.22 \pm 0.08(0.13)$	$-0.42 \pm 0.22(0.37)$	$0.89 \pm 0.10(0.17)$	$8.22^{+2.55}_{-2.20}$
$2.0 < z \leq 2.5$	$10.69 \pm 0.09(0.14)$	$-1.50 \pm 0.05(0.08)$	$0.27 \pm 0.07(0.11)$	$0.19 \pm 0.41(0.66)$	$0.39 \pm 0.07(0.11)$	$4.57^{+1.67}_{-1.37}$
$2.5 < z \leq 3.0$	$10.99 \pm 0.17(0.44)$	$-1.64 \pm 0.07(0.18)$	$0.13 \pm 0.07(0.17)$	$-0.74 \pm 0.87(2.21)$	$0.08 \pm 0.07(0.17)$	$3.29^{+2.64}_{-2.35}$
$3.0 < z \leq 3.5$	$11.28 \pm 0.06(0.11)$	-1.64	$0.06 \pm 0.01(0.01)$	–	–	$2.70^{+0.41}_{-0.43}$
$3.5 < z \leq 4.5$	$10.70 \pm 0.24(0.16)$	-1.64	$0.09 \pm 0.05(0.03)$	–	–	$0.81^{+1.25}_{-0.48}$
$4.5 < z \leq 5.5$	$10.41 \pm 0.13(0.12)$	-1.64	$0.09 \pm 0.03(0.03)$	–	–	$0.44^{+0.40}_{-0.19}$
$5.5 < z \leq 6.5$	$10.14 \pm 0.05(0.12)$	-1.64	$0.06 \pm 0.01(0.03)$	–	–	$0.17^{+0.04}_{-0.03}$
$6.5 < z \leq 7.5$	$10.61 \pm 0.15(0.42)$	-1.64	$0.01 \pm 0.00(0.01)$	–	–	$0.07^{+0.05}_{-0.03}$

Table 6: Double ($z \leq 3$) and single ($z > 3$) Schechter parameters derived for the total mass complete sample from both Likelihood and χ^2 regression fitting. For the Likelihood fit, values are shown for the median posterior distributions with even-tailed 68% range and the values corresponding to the maximum likelihood solution in brackets. For the χ^2 regression fit, uncertainties on parameter values are shown multiplied by $\sqrt{\chi^2_N}$ with formal uncertainties following in brackets.

z-bin	$\text{Log}_{10}\,\mathcal{M}^*$ ($\mathcal{M}_\odot$)	α_1	$\Phi_1 \times 10^3$ ($\text{Mpc}^{-3}\,\text{dex}^{-1}$)	α_2	$\Phi_2 \times 10^3$ ($\text{Mpc}^{-3}\,\text{dex}^{-1}$)	$\rho_*(\mathcal{M} \geq 10^8) \times 10^{-7}$ ($\mathcal{M}_\odot\,\text{Mpc}^{-3}$)
Likelihood Fit						
$0.2 < z \leq 0.5$	$10.71^{+0.18}_{-0.16}[10.90]$	$-1.42^{+0.04}_{-0.05}[-1.47]$	$0.80^{+0.25}_{-0.23}[0.54]$	$-0.08^{+0.63}_{-0.76}[-0.70]$	$0.58^{+0.35}_{-0.37}[1.25]$	$16.17^{+8.08}_{-4.30}$
$0.5 < z \leq 0.8$	$10.75^{+0.14}_{-0.14}[10.94]$	$-1.40^{+0.04}_{-0.05}[-1.47]$	$0.76^{+0.21}_{-0.20}[0.46]$	$-0.16^{+0.59}_{-0.71}[-0.75]$	$0.51^{+0.27}_{-0.31}[1.15]$	$16.36^{+5.64}_{-5.76}$
$0.8 < z \leq 1.1$	$10.75^{+0.16}_{-0.13}[10.92]$	$-1.35^{+0.03}_{-0.04}[-1.36]$	$0.88^{+0.22}_{-0.22}[0.77]$	$0.16^{+0.58}_{-0.80}[-0.51]$	$0.51^{+0.25}_{-0.27}[1.08]$	$18.45^{+6.89}_{-5.14}$
$1.1 < z \leq 1.5$	$10.87^{+0.14}_{-0.12}[10.89]$	$-1.37^{+0.03}_{-0.04}[-1.37]$	$0.69^{+0.18}_{-0.16}[0.68]$	$0.33^{+0.68}_{-1.02}[-0.39]$	$0.18^{+0.14}_{-0.11}[0.70]$	$12.01^{+2.80}_{-2.34}$
$1.5 < z \leq 2.0$	$10.79^{+0.11}_{-0.10}[10.80]$	$-1.51^{+0.06}_{-0.08}[-1.58]$	$0.29^{+0.11}_{-0.11}[0.21]$	$-0.26^{+0.47}_{-0.45}[-0.44]$	$0.47^{+0.14}_{-0.14}[0.88]$	$8.22^{+2.55}_{-2.20}$
$2.0 < z \leq 2.5$	$10.73^{+0.20}_{-0.16}[10.67]$	$-1.50^{+0.06}_{-0.07}[-1.49]$	$0.25^{+0.12}_{-0.09}[0.29]$	$0.12^{+0.74}_{-0.81}[0.28]$	$0.20^{+0.10}_{-0.10}[0.40]$	$4.57^{+1.67}_{-1.37}$
$2.5 < z \leq 3.0$	$10.80^{+0.19}_{-0.20}[11.02]$	$-1.62^{+0.08}_{-0.09}[-1.65]$	$0.18^{+0.12}_{-0.08}[0.12]$	$0.05^{+0.97}_{-1.14}[-0.88]$	$0.07^{+0.09}_{-0.05}[0.07]$	$3.29^{+2.64}_{-2.35}$
$3.0 < z \leq 3.5$	$11.01^{+0.11}_{-0.12}[11.17]$	-1.62	$0.10^{+0.02}_{-0.02}[0.08]$	–	–	$2.70^{+0.41}_{-0.43}$
$3.5 < z \leq 4.5$	$10.51^{+0.10}_{-0.10}[10.61]$	-1.62	$0.14^{+0.04}_{-0.03}[0.12]$	–	–	$0.81^{+1.25}_{-0.48}$
$4.5 < z \leq 5.5$	$10.33^{+0.11}_{-0.12}[10.38]$	-1.62	$0.11^{+0.04}_{-0.03}[0.10]$	–	–	$0.44^{+0.40}_{-0.19}$
$5.5 < z \leq 6.5$	10.33	-1.62	0.11	–	–	$0.17^{+0.04}_{-0.03}$
$6.5 < z \leq 7.5$	10.33	-1.62	0.11	–	–	$0.07^{+0.05}_{-0.03}$
χ^2 Fit						
$0.2 < z \leq 0.5$	$10.69 \pm 0.09(0.17)$	$-1.42 \pm 0.02(0.04)$	$0.83 \pm 0.15(0.27)$	$-0.04 \pm 0.39(0.72)$	$0.68 \pm 0.20(0.38)$	$16.17^{+8.08}_{-4.30}$
$0.5 < z \leq 0.8$	$10.73 \pm 0.07(0.15)$	$-1.40 \pm 0.02(0.05)$	$0.77 \pm 0.11(0.24)$	$-0.12 \pm 0.33(0.71)$	$0.55 \pm 0.14(0.31)$	$16.36^{+5.64}_{-5.76}$
$0.8 < z \leq 1.1$	$10.73 \pm 0.07(0.13)$	$-1.35 \pm 0.02(0.04)$	$0.93 \pm 0.13(0.23)$	$0.27 \pm 0.36(0.63)$	$0.58 \pm 0.13(0.24)$	$18.45^{+6.89}_{-5.14}$
$1.1 < z \leq 1.5$	$10.82 \pm 0.06(0.12)$	$-1.36 \pm 0.02(0.03)$	$0.76 \pm 0.08(0.16)$	$0.63 \pm 0.31(0.66)$	$0.20 \pm 0.06(0.12)$	$12.01^{+2.80}_{-2.34}$
$1.5 < z \leq 2.0$	$10.82 \pm 0.05(0.11)$	$-1.54 \pm 0.05(0.10)$	$0.25 \pm 0.06(0.13)$	$-0.42 \pm 0.25(0.54)$	$0.49 \pm 0.07(0.14)$	$8.22^{+2.55}_{-2.20}$
$2.0 < z \leq 2.5$	$10.73 \pm 0.12(0.19)$	$-1.51 \pm 0.05(0.07)$	$0.25 \pm 0.07(0.12)$	$0.13 \pm 0.55(0.87)$	$0.22 \pm 0.06(0.10)$	$4.57^{+1.67}_{-1.37}$
$2.5 < z \leq 3.0$	$10.90 \pm 0.12(0.33)$	$-1.62 \pm 0.03(0.09)$	$0.16 \pm 0.04(0.12)$	$0.19 \pm 0.57(1.58)$	$0.04 \pm 0.03(0.08)$	$3.29^{+2.64}_{-2.35}$
$3.0 < z \leq 3.5$	$11.05 \pm 0.10(0.12)$	-1.62	$0.09 \pm 0.02(0.02)$	–	–	$2.70^{+0.41}_{-0.43}$
$3.5 < z \leq 4.5$	$10.54 \pm 0.16(0.11)$	-1.62	$0.13 \pm 0.05(0.03)$	–	–	$0.81^{+1.25}_{-0.48}$
$4.5 < z \leq 5.5$	$10.38 \pm 0.12(0.11)$	-1.62	$0.10 \pm 0.03(0.03)$	–	–	$0.44^{+0.40}_{-0.19}$
$5.5 < z \leq 6.5$	10.38	-1.62	0.10	–	–	$0.17^{+0.04}_{-0.03}$
$6.5 < z \leq 7.5$	10.38	-1.62	0.10	–	–	$0.07^{+0.05}_{-0.03}$

Table 7: Double ($z \leq 3$) and single ($z > 3$) Schechter parameters derived for the star-forming mass complete sub-sample from both Likelihood and χ^2 fitting. Values are provided as described in Table 6.

z-bin	$\mathrm{Log}_{10}\,\mathcal{M}^*$ ($\mathcal{M}_\odot$)	α_1	$\Phi_1 \times 10^3$ ($\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$)	α_2	$\Phi_2 \times 10^3$ ($\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$)	$\rho_*(\mathcal{M} \geq 10^8) \times 10^{-7}$ ($\mathcal{M}_\odot\,\mathrm{Mpc}^{-3}$)
Likelihood Fit						
$0.2 < z \leq 0.5$	$10.87^{+0.07}_{-0.08}[10.90]$	$-0.58^{+0.15}_{-0.12}[-1.47]$	$1.02^{+0.18}_{-0.17}[0.54]$	$-1.80^{+0.15}_{-0.17}[-0.70]$	$0.01^{+0.02}_{-0.01}[1.25]$	$6.92^{+0.88}_{-0.83}$
$0.5 < z \leq 0.8$	$10.84^{+0.06}_{-0.06}[10.94]$	$-0.38^{+0.12}_{-0.10}[-1.47]$	$1.05^{+0.12}_{-0.12}[0.46]$	$-1.85^{+0.22}_{-0.25}[-0.75]$	$0.00^{+0.01}_{-0.00}[1.15]$	$6.59^{+0.65}_{-0.71}$
$0.8 < z \leq 1.1$	$10.80^{+0.06}_{-0.06}[10.92]$	$-0.31^{+0.10}_{-0.10}[-1.36]$	$1.16^{+0.13}_{-0.13}[0.77]$	-1.85	$0.00^{+0.00}_{-0.00}[1.08]$	$6.78^{+0.73}_{-0.77}$
$1.1 < z \leq 1.5$	$10.61^{+0.04}_{-0.05}[10.89]$	$0.18^{+0.12}_{-0.11}[-1.37]$	$0.71^{+0.04}_{-0.04}[0.68]$	-1.85	$0.00^{+0.00}_{-0.00}[0.70]$	$3.20^{+0.31}_{-0.37}$
$1.5 < z \leq 2.0$	$10.43^{+0.05}_{-0.05}[10.80]$	$0.72^{+0.13}_{-0.12}[-1.58]$	$0.35^{+0.02}_{-0.02}[0.21]$	–	–	$1.53^{+0.15}_{-0.19}$
$2.0 < z \leq 2.5$	$10.45^{+0.06}_{-0.07}[10.67]$	$0.77^{+0.20}_{-0.18}[-1.49]$	$0.11^{+0.01}_{-0.01}[0.29]$	–	–	$0.53^{+0.07}_{-0.08}$
$2.5 < z \leq 3.0$	$10.32^{+0.09}_{-0.09}[11.02]$	$1.39^{+0.39}_{-0.33}[-1.65]$	$0.05^{+0.01}_{-0.01}[0.12]$	–	–	$0.33^{+0.07}_{-0.05}$
$3.0 < z \leq 3.5$	$10.40^{+0.17}_{-0.15}[11.17]$	$1.53^{+0.62}_{-0.54}[-1.60]$	$0.02^{+0.01}_{-0.01}[0.08]$	–	–	$0.18^{+0.08}_{-0.05}$
$3.5 < z \leq 4.5$	$10.52^{+0.11}_{-0.13}[10.61]$	$1.16^{+0.57}_{-0.48}[-1.60]$	$0.01^{+0.00}_{-0.00}[0.12]$	–	–	$0.09^{+0.03}_{-0.02}$
$4.5 < z \leq 5.5$	$9.75^{+0.48}_{-0.51}[10.38]$	1.16	$2.64^{+4.84}_{-2.63}[0.10]$	–	–	$0.01^{+0.08}_{-0.01}$
χ^2 Fit						
$0.2 < z \leq 0.5$	$10.93 \pm 0.06(0.08)$	$-0.70 \pm 0.09(0.13)$	$0.91 \pm 0.13(0.18)$	$-1.98 \pm 0.16(0.22)$	$0.00 \pm 0.00(0.00)$	$6.88^{+1.61}_{-1.39}$
$0.5 < z \leq 0.8$	$10.88 \pm 0.05(0.06)$	$-0.47 \pm 0.09(0.10)$	$0.99 \pm 0.12(0.12)$	$-2.15 \pm 0.32(0.33)$	$0.00 \pm 0.00(0.00)$	$6.48^{+1.43}_{-1.29}$
$0.8 < z \leq 1.1$	$10.83 \pm 0.06(0.05)$	$-0.36 \pm 0.09(0.08)$	$1.13 \pm 0.13(0.12)$	-2.15	$0.00 \pm 0.00(0.00)$	$6.88^{+1.50}_{-1.51}$
$1.1 < z \leq 1.5$	$10.62 \pm 0.06(0.04)$	$0.14 \pm 0.13(0.10)$	$0.72 \pm 0.06(0.04)$	-2.15	$0.00 \pm 0.00(0.00)$	$3.11^{+0.56}_{-0.50}$
$1.5 < z \leq 2.0$	$10.43 \pm 0.09(0.05)$	$0.71 \pm 0.22(0.12)$	$0.35 \pm 0.03(0.02)$	–	–	$1.49^{+0.51}_{-0.49}$
$2.0 < z \leq 2.5$	$10.46 \pm 0.06(0.06)$	$0.76 \pm 0.17(0.19)$	$0.12 \pm 0.01(0.01)$	–	–	$0.52^{+0.14}_{-0.10}$
$2.5 < z \leq 3.0$	$10.31 \pm 0.11(0.10)$	$1.42 \pm 0.42(0.38)$	$0.06 \pm 0.02(0.01)$	–	–	$0.37^{+0.54}_{-0.21}$
$3.0 < z \leq 3.5$	$10.35 \pm 0.25(0.15)$	$1.73 \pm 1.00(0.61)$	$0.02 \pm 0.02(0.01)$	–	–	$0.11^{+0.58}_{-0.10}$
$3.5 < z \leq 4.5$	$10.53 \pm 0.08(0.12)$	$1.14 \pm 0.37(0.52)$	$0.01 \pm 0.00(0.00)$	–	–	$0.09^{+0.07}_{-0.03}$
$4.5 < z \leq 5.5$	$11.69 \pm 1.53(1.69)$	$-0.81 \pm 0.84(0.93)$	$0.00 \pm 0.00(0.00)$	–	–	$0.01^{+0.76}_{-0.02}$

Table 8: Double ($z \leq 1.5$) and single ($z > 1.5$) Schechter parameters derived for the quiescent mass complete sub-sample from both Likelihood and χ^2 fitting. Values are provided as described in Table 6.

z-fix	$\mathrm{Log}_{10}\,\mathcal{M}^*$ ($\mathcal{M}_\odot$)	α_1	$\Phi_1 \times 10^3$ ($\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$)	α_2	$\Phi_2 \times 10^3$ ($\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$)	$\rho_*(\mathcal{M} \geq 10^8) \times 10^{-7}$ ($\mathcal{M}_\odot\,\mathrm{Mpc}^{-3}$)
Continuity Model						
0.2	$10.68^{+0.04}_{--0.02}(10.69)$	$-0.37^{+0.06}_{--0.03}(-0.34)$	$1.80^{+0.25}_{--0.13}(2.03)$	$-1.60^{+0.00}_{--0.00}(-1.60)$	$0.50^{+0.03}_{--0.02}(0.53)$	12.54(14.15)
1.5	$10.80^{+0.03}_{--0.01}(10.82)$	$-0.37^{+0.06}_{--0.03}(-0.34)$	$0.86^{+0.07}_{--0.04}(0.89)$	$-1.60^{+0.00}_{--0.00}(-1.60)$	$0.29^{+0.01}_{--0.01}(0.29)$	8.61(9.23)
3.0	$10.98^{+0.06}_{--0.03}(11.00)$	$-0.37^{+0.06}_{--0.03}(-0.34)$	$0.00^{+0.00}_{--0.00}(0.01)$	$-1.60^{+0.00}_{--0.00}(-1.60)$	$0.16^{+0.02}_{--0.01}(0.17)$	3.11(3.49)

Table 9: Double Schechter Function parameters derived for the total sample from the Continuity Model fitting. Values are provided as described in Table 6 for the three fixed z points.

5 BEASTS IN THE BUBBLES: A JWST CYCLE 1 GO PROGRAM

As for me, I am tormented with an everlasting itch for things remote. I love to sail forbidden seas, and land on barbarous coasts.

- Herman Melville, *Moby Dick*

Adapted from:

Beasts in the Bubbles: Characterizing ultra-luminous galaxies at Cosmic Dawn

Selected for Cycle 1 General Observing with the James Webb Space Telescope

J. R. Weaver (PI), C. M. Casey, S. Toft, S. L, Finkelstein, S. Fujimoto, K. M. L. Gould, O. Ilbert, C. K. Jespersen, O. B. Kauffmann, H. J. McCracken, B. Milvang-Jensen, B. Mobasher, A. L. Faisst, G. Magdis, F. M. Valentino, N. Chartab, I. Davidzon, C. McPartland, P. Oesch, D. B. Sanders, C. Scarlata, M. Shuntov, C. L. Steinhardt, D. K. Stern, I. Szapudi, L. Zalesky, G. Brammer

Based on work from:

COSMOS2020: UV selected galaxies at $z \geq 7.5$

Submitted to Astronomy & Astrophysics; included in Appendix 6.

O. B. Kauffmann, O. Ilbert, J. R. Weaver, H. J. McCracken, B. Milvang-Jensen, G. Brammer, I. Davidzon, O. Le Fèvre, D. Liu, B. Mobasher, A. Moneti, M. Shuntov, S. Toft, C. M. Casey, J. S. Dunlop,J. S. Kartaltepe, A. M. Koekemoer, D. B. Sanders, and L. Tresse

5.1 OVERVIEW

In the last decade, surveys have discovered galaxies at $z \sim 11$ (Oesch et al., 2016), massive quasars at $z \sim 7.5$ (Bañados et al., 2018) and massive quiescent galaxies at $z > 4$ (Valentino et al., 2020b). The existence of these surprisingly evolved galaxies at the highest redshifts available challenges the standard paradigm of structure formation (Steinhardt et al., 2016a; Behroozi et al., 2018b). However, it has not been possible to spectroscopically confirm whether these galaxies are truly so mature and at such high redshift.

Here, we propose NIRSpec IFU/Prism observations of a complete sample of the five rare, ultra-luminous ($-22 < M_{UV} < -21$), massive ($10 < logM_*/M_\odot < 10.5$) galaxies at $z > 9$ selected from the COSMOS2020 catalog (Weaver et al., 2022a). The wide area of COSMOS is essential to find such extremely rare galaxies. Only five are found over the 2 deg^2 COSMOS field, so *none* are expected in the entire planned JWST GTO/ERS programs; JADES being 24 times smaller in area. COSMOS is the *only* survey which can deliver large samples of these cosmic beasts as it includes both wide and deep imaging in 39 bands from UV to 4.5μm. As a result, the redshift probability distributions peak only at high redshift, in contrast to other less secure samples. Therefore,this is the most secure sample of $z > 9$ galaxy candidates available. We will confirm the redshifts *and* characterize the properties of this five-galaxy sample in only 14.4 hours.

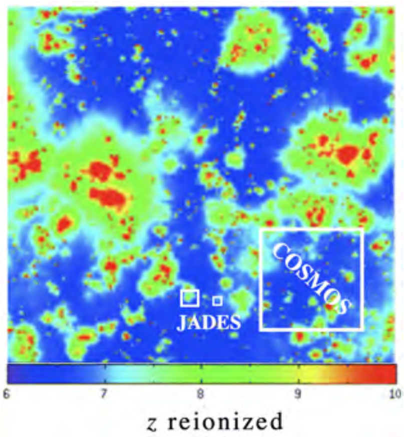

Figure 5.1: 25 deg^2 simulation of reionization, adapted from Trac et al. (2008b). Reionization is patchy on degree scales, starting around the brightest galaxies at $z \sim 10$ and growing until $z \sim 6$ when the universe is reionized. Only wide surveys like COSMOS (2 deg^2) can detect multiple bright $z > 9$ galaxies. Smaller surveys like JADES are likely to miss them.

5.1.1 *The first massive galaxies challenge galaxy formation models*

The most massive galaxies at $9 < z < 10$ are still in their assembly phase; between initial collapse and subsequent evolution (e.g. Bouwens et al., 2014b). Contemporary cosmological models predict that such massive galaxies form via hierarchical merging, in the most overdense regions of the primordial web of dark matter and encased in a thick fog of neutral hydrogen from which they grow. At lower redshifts, the UV Luminosity Function (UVLF) is described by a Schechter function (Schechter, 1976b) with an exponential cutoff above a characteristic luminosity (i.e. mass) attributed to "feedback" mechanisms that inhibit galaxy growth (e.g Peng et al., 2010b). However at $9 < z < 10$, galaxies had little time to mature and so closely betray the conditions under which they formed. If they grew out of baryonic cooling onto dark matter halos, as predicted (e.g. Gnedin et al., 2004), then the UVLF at $z \gg 6$ should resemble a powerlaw without an exponential cutoff, pointing to a pre-feedback formation stage (Finkelstein et al., 2015b).

It is thought that only the most massive among these early galaxies emit enough UV emission to reionize a significant region of their prenatal hydrogen cocoon so quickly (Naidu et al., 2020). Alternatively, collections of smaller galaxies co-inhabiting these overdense regions may have provided the necessary UV emission (Bouwens et al., 2012b). Whichever is the case, the first "bubbles" of the Epoch of Reionization formed in the vicinity of only the most massive galaxies and hence trace the underlying fluctuations in the primordial web of dark matter. Given that these fluctuations are expected to vary on degree scales, the topology of reionization is expected to be correspondingly patchy (Trac et al., 2008b), making the most luminous sources at these epochs incredibly rare (Moster et al., 2011b; Steinhardt et al., 2021).

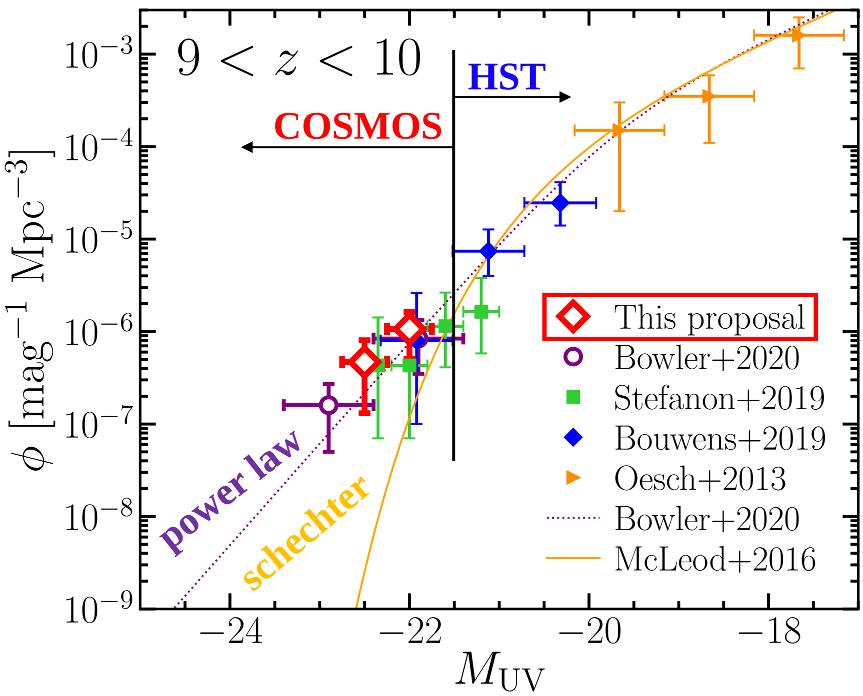

Figure 5.2: Predicted constraints with this *spectroscopic* proposal (red diamonds) on the $9 < z \leq 10$ UV luminosity function. Points show *photometric* data. Only JWST will allow us to differentiate models of galaxy evolution: a Schechter function (yellow) or a double power law (black).

5.1.2 *Area is more important than depth*

Recent studies of the UVLF have introduced evidence for this pre-feedback formation stage (Figure 5.2, Bowler et al. 2020a, Kauffmann in prep.). Although studies have established the low-luminosity end of the UVLF with *HST* (Bouwens et al., 2014a, 2021a), the most luminous galaxies are common only on degree scales. Hence, only ground-based NIR surveys with $> 1\,\mathrm{deg}^2$ in area have been able to secure significant candidate samples. Only these extreme galaxies can indicate the physics of galaxy formation and yet no spectroscopic sample of confirmed $9 < z < 10$ galaxies currently exists.

Like *HST*, *JWST* will not survey several degrees, and no planned GTO or ERS program is expected to find even a *single* such galaxy, an effect exacerbated by cosmic variance which makes the median case significantly worse than the average one. Estimates of $9 < z < 10$ galaxy number densities *including cosmic variance* indicate one would need to perform a JADES-like survey 50 times to expect to find a galaxy at $M_{UV} \sim -22$, or $> 1,000\times$ at $M_{UV} \sim -23$ (e.g. Moster et al., 2011b). In fact, the discovery of the most distant known galaxy, GN-z11 (Oesch et al., 2016), rules out the possibility of undetected galaxies with $M_{UV} < -22.1$ at $z < 11$ in GOODS-N (part of JADES), wherein lie valuable constraints. COSMOS is one of the only existing surveys with sufficient multi-wavelength *wide and deep* coverage to find these most massive and extremely rare cosmic beasts.

5.1.3 *A census of luminous $z \sim 9$ galaxies*

We select a complete sample of the five brightest, most extreme and most secure $z \sim 9$ candidates between $-23 < M_{UV} < -21$ out of the nearly 1 000 000 galaxies in the new COSMOS2020 catalog. Each source was measured by *both* aperture and profile-fitting photometry in 39 bands spanning the UV to 4.5μm, and each set of photometry has been fit with two photometric redshift codes (`EAZY` and `Le Phare`). Therefore, each candidate has four photometric redshift estimates, advantageous to the goals of this proposal as their remarkable agreement demonstrates their robustness and mitigates catastrophic biases arising in more simplistic analyses.

Although the sample established by Bowler et al. (2020b) uses the latest NIR UltraVISTA DR4 release (Moneti et al., 2019), they are not selected against the latest ultra-deep HSC PDR2 (Aihara et al., 2019) *grizy* images which we use in our sample to better identify low-z interlopers. Because of this, photometric redshift solutions of our sample are uniquely constrained to high redshift, in contrast to other similar samples. Crucially, the criteria used to establish our sample *rejects* the bright $z \sim 9$ source of (Bowler et al., 2020b) with non-zero optical flux, shown in Figure 5.2.

No object shows any significant optical flux in any of the 15 bands blueward of 1.2 μm down to 27.5 AB in g and 26.0 AB in Y (Fig. 5.3). They are consistent with photometric redshift solutions between $9 < z < 10$. None have coincident 24 μm emission and are thus unlikely to be contaminated by low-redshift dusty galaxies. Comparisons with brown dwarfs rule out stellar contamination. These five galaxies constitute the most secure sample of extremely luminous $z \sim 9$ galaxy candidates complete over a contiguous 2 deg^2.

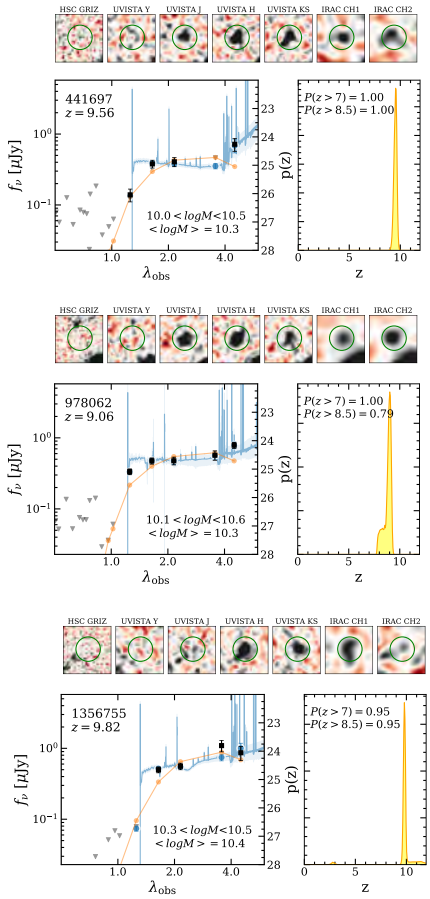

Figure 5.3: A complete sample of five bright $9 < z < 10$ candidates selected by both aperture- and profile-fitting photometric extraction techniques, shown here with EAZY for the latter. $4'' \times 4''$ cutouts are shown scaled by ± 3 S/N with a $2''$ diameter aperture demonstrate the Y band dropping out, indicative of a $9 < z < 10$ solution. The best-fitting template spectrum (blue) is fit to the observed photometry (black) including upper limits (grey arrows), corresponding to the peak of the redshift probability distribution (yellow). Also indicated is the range and median stellar masses derived using two different codes on the two sets of photometry, as well as the best-fit stellar template (orange) which in all cases is *significantly* less likely than the galaxy template.

5.2 TECHNICAL JUSTIFICATION

We propose to achieve all of the aforementioned science goals with a small program utilizing the IFU-mode of NIRSpec dispersed through the $R \sim 100$ prism to observe five of the most distant, luminious galaxies yet known. At $z \sim 9$ this will obtain simultaneous *and fast* restframe UV and optical spectroscopy. It is not known if galaxies at $z \sim 9$ emit strongly in Lyα as spectroscopy of a $z \sim 11$ source (i.e. GN-z11) has shown only continuum UV emission. Regardless, only extreme equivalent width Lyα emission could be detected for our proposed observations due to the effective R~40 resolution at $\sim 1.3\,\mu$m with the PRISM. Nonetheless, we will use UV+optical spectroscopy to obtain a redshift determination from strong optical emission lines *and* a secure Lyman break signature guaranteed by our observation strategy. If Lyα is detected then we will have up to three ways to confirm redshift.

Previous attempts to confirm even small samples of galaxies at $z > 9$ with existing facilities have been unsuccessful. Their rarity, coupled with their faintness and uncertain line emission complicate ground-based spectroscopic confirmation. Despite achieving greater degrees of success at marginally lower redshifts ($z \sim 7-8$), ALMA cannot detect [OIII 88μm] at $9 < z < 9.6$ as it falls within the 150 GHz atmospheric absorption gap. Meanwhile, Keck would require several nights for each target to reach even optimistic Lyα emission, if it is there at all, and it would leave all science objectives out of reach. These are the brightest sources found at $z \sim 9$, and so there is little hope for shorter redshift-confirmation observations with even brighter sources.

Space-based multiplexed grism spectroscopic surveys (e.g. 3D-HST, FIGS) have also been unable to identify significant samples of $z > 9$ sources for similar reasons, in addition to the optical lines being redshifted out of reach by WFC3. Like Keck and ALMA, the science objectives of this proposal are impossible with *HST*.

As mentioned, JADES-MEDIUM is not expected to find any significant samples to probe the bright end of the UVLF where both the physics of galaxy formation and the process of reionization can be simultaneously studied. Only spectroscopy of this sample, collected over the 2 deg^2 of COSMOS, can provide a secure means of addressing these hitherto unsolved mysteries.

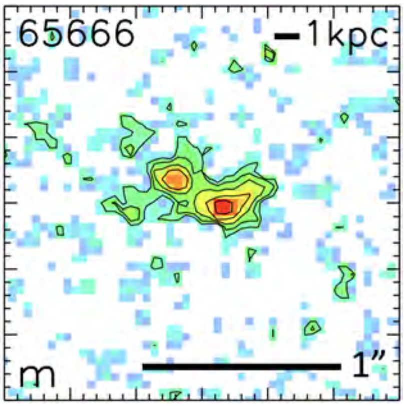

Figure 5.4: An example of a $z \sim 7$ galaxy as seen by *HST*/*WFC3* with multiple components, but unresolved in ground-based surveys. Adapted from Bowler et al. (2017a).

IFU studies pioneered in the local universe have been transformative in studying the spatially resolved properties of galaxies at $z < 1$. Now studies at the highest redshifts are possible with NIRSpec/IFU. Recent *HST* observations have revealed that similarly bright $z \sim 7$ are resolved and surprisingly spatially extended with disturbed morphologies indicative of ongoing major mergers (see Fig. 5.4). If $z > 9$ candidates have similar multi-component morphologies or are indeed multiple sources, the IFU will resolve them spatially and we will be able to confirm them as ongoing mergers. Moreover, IFU-mode provides free background estimates simultaneously, crucial to establishing the significance of our measurements. Observing with only slits, which demands the costly overheads from pre-imaging, is therefore sub-optimal in efficiently harvesting all the available information. Our rationale in choosing IFU spectroscopy is then both practical *and* scientific, as NIRSpec IFU can quickly and efficiently reveal the structural components of these galaxies while simultaneously exploring their spectral properties in an incredible leap forward. For these reasons, IFU spectroscopy is an effective strategy to extract comprehensive measurements of the high-redshift universe.

Our exposure times are calculated such that we will probe the light blueward of the Lyman break to the nominal depth of the UltraVISTA *Y* band, at 26.0 AB in order to secure a spectroscopic redshift in the unexpected scenario where all emission lines are absent. In addition, we have explored modelling simulated NIRSpec observations of the best-fit `EAZY` templates for each galaxy in the sample with `BAGPIPES` (Carnall et al., 2019b), with an example shown in Figure 5.5. Given the similar apparent magnitudes of our sources, a common exposure time of 1h38m will be sufficient to perform measurements with the necessary contrast required to secure a unambiguous detection of the Lyman break *and* address all of the science objectives in an efficient, small program. The total time for the observations of the five-galaxy sample with NIRSpecc PRISM IFU spectroscopy is 14.4h.

We choose a `SPARSE-CYCLE LARGE` dither patter using positions 1, 2, 3, 4 in order to obtain optimal pixel sub-sampling, background subtraction in case

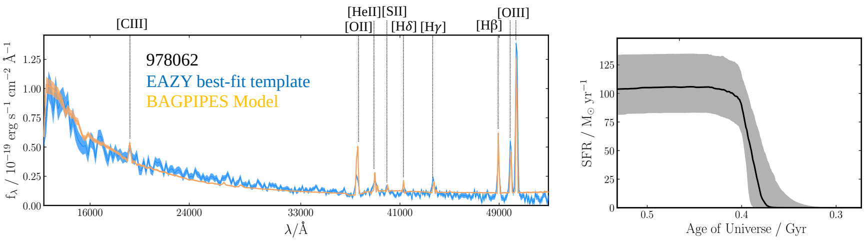

Figure 5.5: A sample simulated 5836 s observation of a $z \sim 9$ galaxy with BAGPIPES, using the best-fit EAZY template as a proxy. We demonstrate the feasibility of our analysis by performing several trials on randomly generated noisy spectra from which we recover consistent redshift, mass, and star-formation history solutions.

of extended emission (<0.3-0.4″), and to place the targets safely within the IFU for each exposure, taking into account expected pointing uncertainties of 0.1-0.2″. The MSA slits, even when closed, are not fully opaque and so allow parasitic light from particularly bright stars in the field to fall onto the detector. These parasitic spectra can be identified and removed by performing an observation with the MSA slits open. The expected level of contamination for PRISM/CLEAR is < 1%, and so we will acquire an open MSA observation once per object of an exposure time equal to that one one integration group.

This incredibly rich set of science objectives only scratches the surface of what is possible with this sample. Deeper spectroscopic followup of these sources could not only detect nominal Lyα emission, but resolve its profile to determine the location of the galaxy within its reionized bubble (cf. Mason et al., 2020). This would provide the first tantalizing constraints the topology of reionization. High-resolution spectroscopy of the 4000Å region would enable the first ultra-precise constraints on the stellar populations, star-formation histories, and metallicities of galaxies at these early times. A follow-up program to obtain all of these measurements will be made feasible by this proposal. They will provide fundamental constrains to precisely anchor theories of galaxy formation at high redshift, the results of which would feed back to the entire extragalactic community.

5.3 ANALYSIS STRATEGY & SCIENCE RETURNS

5.3.1 *Confirming the redshifts*

The primary goal of this program is to "derive record breaking redshifts" for this sample of five rare and ultra-luminous galaxies seen only 500 Myr following the Big Bang. Doing so will place them amongst the most distant objects ever confirmed, and enable further exploration into the formation of these mysterious galaxies. The constraints from our robust, four-fold SED-fitting analysis places these galaxies at z 9 and predicts UV continuum between 24-25 AB. The observing time for each target has been calculated to measure continuum emission to a depth exceeding 26 AB, in order to confirm their Lyman Break by the absence of continuum blueward of 1.2um. This redshift confirmation can be supported by the detection and identification of multiple emission lines ([OII], Hδ, Hγ, Hβ, [OIII], and possibly Lyα) which will be immediately revealed by the spectral analysis discussed below.

5.3.2 *Testing theories of galaxy formation and reionization*

Spectroscopic constraints on the luminous end of the $z \approx 9$ UVLF will enable us to differentiate theories of galaxy formation and understand the sources of reionization. In Kauffmann et al., we used the largest and most secure photometric sample of $z = 8 - 10$ galaxies from COSMOS to place the first robust constraints on the massive end of the UVLF. We will now spectroscopically measure the UVLF with $z \sim 9$ observations from *JWST*.

Are we witnessing a pre-feedback stage?
At lower redshifts, the UV Luminosity Function (UVLF) is described by a Schechter function with an exponential cutoff above a characteristic luminosity (i.e. mass) attributed to "feedback" mechanisms which inhibit galaxy growth. However at $9 < z \lesssim 10$, galaxies have only just collapsed, and so closely betray the conditions under which they formed. If they grew out of baryonic cooling onto dark matter halos, as predicted, then the UVLF at $z \gg 6$ should resemble a powerlaw without an exponential cutoff, pointing to a *pre*-feedback formation stage (Finkelstein et al., 2015b). Our work in Kauffmann et al. measures the UV emission of the brightest $8 < z \lesssim 10$ galaxies selected from COSMOS2020 and the most compelling evidence to date: luminous, massive galaxies may have formed and assembled within 500 Myr of the Big Bang. Spectroscopically confirming the redshifts and UV luminosities of these five galaxies with JWST will show the power-law like shape of the UVLF, thus demonstrating this scenario for the first time.

Did such ultra-luminous beasts drive reionization?
Total illumination will only come by consistently measuring *both* the high- and low-luminosity ends of the $z \approx 9$ UVLF. This means we must not only confirm the existence and nature of these cosmic beasts, but also identify hitherto undiscovered less-luminous galaxies which surround them. We expect to identify $\sim$ 500 less luminous ($M_{UV} > -21$) galaxies at $z \approx 9$ over the combined 0.6 deg^2 NIRCam and 0.2 deg^2 MIRI mosaics from the 200 hr *JWST* Cycle 1 program COSMOS-Web, and be able to assess their UV luminosities. By doing so we will complete the first precise picture of the $z \approx 9$ UVLF from the massive beasts down to the smaller ones in their shadows. The consistency of measuring across the same field (and hence same cosmic structure) will help to *remove systematic biases* and solidify the finding of a pre-feedback stage. Insufficient ionizing flux from collections of less luminous sources will establish the dominant role of the most luminous beasts as the drivers of reionization.

5.3.3 *The demographics of the first massive galaxies*

Statistical diagnostics like the UVLF are extremely powerful to test models and understand *average* properties. However, the formation of the first galaxies, like reionization itself, is likely to have a high *variance*. We will leverage our team expertise to characterize the diverse properties of these five beasts to perform the first precision demographics survey of galaxies within the EoR.

What are their formation histories?
We see galaxies as snapshots; thankfully the consistent operation of physical law allows us to employ spectral models to reconstruct their recent past. A key advantage of IFU spectroscopy is the ability to collapse the cube spatially to produce a high SNR spectrum of the entire object. For these high-SNR global spectra we will leverage bleeding-edge stellar population fitting codes such as Bagpipes (Carnall et al., 2018a) and Prospector (Leja et al., 2017) to produce precise measurements of the global stellar properties including the stellar mass and star-formation history, bulk stellar ages from the Balmer break, dust attenuation from the UV continuum slope, and other diagnostics from emission line fitting. Although the detection of Lyα at $R < 100$ would imply an enormous equivalent width, the finding would fundamentally change our view of star-formation in early galaxies. These valuable constraints will enable us to gain fundamentally new insights into the formation of these the first massive galaxies.

Furthermore, for the first time we will be able to establish a *timeline* for the potentially diverse assembly histories of the these galaxies. Constant star-formation would indicate that physical mechanisms have not yet suppressed their growth, and *directly* corroborate the UVLF. Alternatively, if these mechanisms are acting to halt their star-formation then we can find out why by e.g., using the [OIII]/Hβ ratio to indicate hard ionization from a central massive black hole. The discovery of an active black hole at $z \approx 9$ would be a record-breaker, and would suggest that they too are contributing to reionization *alongside* star-formation.

What are their proprieties? Are they merger-driven?
Recent *HST* observations caught several $z \sim 7$ luminous galaxies in the process of merging (Fig. 5.4, Bowler et al. 2017b). It is well known that mergers indelibly alter the properties and destiny of a galaxy, and even spur on star-formation (e.g. Weaver et al., 2018). Can the extreme UV luminosity of $z \sim 9$ galaxies be explained by star-formation brought upon by mergers?

Broadband morphologies for each object can be reconstructed by collapsing spectral slices of the datacubes. We will do this for both the UV and Optical regimes separately, and employ surface profile fitting tools Galfit (Peng et al., 2002, 2010a) and The Tractor (Lang et al., 2016a) to model the continuum morphologies of each galaxy. Key measurements such as effective radius, sersic index, axis ratio, and position angle will be obtained, enabling additional means of understanding and contextualizing the findings from the emission line and continuum fitting. Additional codes already prepared by this team will constrain five additional key morphological indices: Concentration, Asymmetry, Clumpiness, Gini, M20, and Shape Asymmetry (Pawlik et al., 2016, and references therein) which together will enable a complete understanding of the dynamical state of each object in both the UV and optical. This will allow us to assess extent to which mergers are contributing to their rapid formation and place their assembly in the context of other massive galaxies seen at later epochs of the universe.

The most valuable measurements provided *only* by spatially resolved spectroscopy are the surface maps of various physical quantities, which reveal the variance of these properties across the galaxy. To do this, we will use PyParadise (Walcher et al., 2015; Weaver et al., 2018) to separate the stellar continuum from the emission lines, per spaxel, for each galaxy. Then we will use the Voronoi algorithm (Cappellari et al., 2003) to bin the stellar continuum spaxels into a smaller number of voxels, each with a similar SNR. PyParadise will again be used to measure the stellar continuum properties: dust attenuation and > 10 Myr average SFR (UV continuum), stellar mass (optical continuum), and bulk stellar ages (Balmer break).

The remaining emission line spectra will have sufficient SNR to avoid being Voronoi binned. Hence we will use PyParadise again to measure the emission line fluxes of several key lines, where they may appear: [CIII], [OII], HeII, SII, Hδ, Hγ, Hβ, and [OIII], and hence trace the instantaneous star-formation rate ([OIII]), dust attenuation (Hβ/Hγ) metalicity, and ionization conditions arising from AGN ([OIII]/Hb) of the gas at the full spatial resolution afforded by NIRSpec. While PyParadise is able to provide elementary kinematical measurements on the emission lines (the spectral resolution will not permit absorption line measurements with sufficient SNR to measure stellar kinematics), we will also employ the pPXF code (Cappellari, 2017) as a cross-check on gas kinematics.

Future follow-up spectra with NIRSpec at higher spectral resolution will provide detailed constraints of the stellar population and kinematics, enabled by these forerunner measurements. As shown in the Fig. 5.6, Weaver et al. (2018) used these techniques to conduct a similar study of a major merger remnant in the nearby universe, which these galaxies are likely to resemble albeit at *much* lower resolution. These valuable results, taken together, will finally provide a clear view of the aftermath of a galaxies' initial collapse needed to refine simulations of galaxy formation.

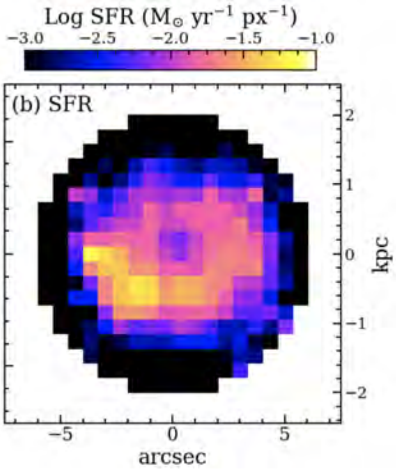

Figure 5.6: An integral field spectroscopic map of the star-formation rates across the central 5 kpc star-forming disc of the nearby major merger remnant NGC 7252. Adapted from Weaver et al. (2018).

A unique aspect of this program is the completeness of the sample. With precise global measurements as well as spatially resolved maps, the properties of this complete sample can be explored. The potential diversity, or lack thereof, will provide additional strength in understanding these first galaxies as a population, with greater constraining power than can be provided by comparably smaller, and less complete samples. This will involve both traditional correlation analyses, as well as applications of bleeding-edge machine learning techniques such as UMAP (McInnes et al., 2018). These analyses will provide the first comprehensive overview of early galaxy formation and evolution by comparing their key measurements both globally and spatially resolved, and will uncover how these beasts grew rapidly from the most overdense regions and began the process of reionizing the universe. Their diverse properties, now unveiled, will shed light on the precise physical mechanisms responsible for the shape of the UVLF and the formation of the first reionized bubbles.

5.3.4 *Galaxies as tracers of large scale structure*

These five galaxies are scattered across the 2 deg^2 of COSMOS. Their expected nature as the most UV-luminous galaxies at $z \approx 9$ means they must have formed within massive dark matter haloes existing in the most overdense regions in the early universe, imparting significant constraints on the structure of dark matter, and large scale structure by extension. We intend to use extreme value statistics to place the first constraints on the matter density distribution at $z \approx 9$, comparing with merger histories obtained from semi-analytic models (e.g., those of Santa Cruz, Somerville et al. 2008). We will use the spatial location of each galaxy in tandem with their physical properties to infer valuable constraints on large scale cosmology. In addition, we can use their UV emission measurements (including Lyα, possibly) to place the first constraints on the $z \approx 9$ ionization budget and hence test theories about the development of the Epoch of Reionization. Hence we will validate similar measurements which will be provided by other Cycle 1 programs.

5.4 ANCILLARY OBSERVATIONS

Galaxy evolution stands to thrive in the rapidly approaching era of JWST. Significant progress is expected to follow even the first observations in all areas of astrophysics. It is particularly exciting to have the opportunity to leverage the first space-based IFU to push the frontier of integral field spectroscopic studies out to the first billion years. No doubt that there will be obstacles and challenges to overcome, but, at the same time there are discoveries waiting on the other side.

Table 10: Summary of ancillary observations for four of five BEASTS.
1. PIs: Casey, Mobasher, and Sanders. 2. PI: Fujimoto. 3. PI: Hashimoto

ID (Classic)	H-band	Keck MOSFIRE[1]	VLT/Xshooter[2]	Subaru/SWIMS[2]	ALMA[3]
COS-1356755	24.54	1 N (1/22/22)	4hr	3/4 N	queue
COS-441697	24.89	1 N (1/22/22)	4hr	3/4 N	queue
COS-852845	25.25	1 N (2/2/22)	5hr		
COS-564423	25.4	3hr archival data	queue		

Since the identification of these five beasts, there has been a flurry of observational follow-ups. Table 10 provides an overview of each object and which programs have been active in obtaining measurements. While COS-978062 has been confirmed by REBELS (ALMA Large Program, Bouwens et al., 2021b) to be at $z = 7.7$, which is not only still in the high redshift discovery space, but is statistically consistent with our constraints. The other four remain undetected. We expect additional investment particularly from the WERLS program (Keck NASA strategic program, PIs: Casey & Kartaltepe) in the near future, in support of the COSMOS-Web program (which has an overlap with COS-564423). Given that four of five remain unconfirmed highlights the difficulties faced by ground-based observatories, and the awesome, transformative vision of JWST on the horizon.

6 OUTLOOK

> *We find them smaller and fainter, in constantly increasing numbers, and we know that we are reaching into space, farther and farther, until, with the faintest nebulae that can be detected with the greatest telescopes, we arrive at the frontier of the known universe.*
>
> - Edwin Hubble, *The Realm of the Nebulae, 1936*

Catalogs have been the backbone of astronomy since time immemorial. They represent our journey into the furthest reaches of our universe, and our attempts to understand it with the tools available to us. From the first collections of stars by Tycho Brahe to the unfathomable images of the deep fields captured by *Hubble*, advances in our understanding of technology, physical law, and statistics have preceded and enabled some our greatest discoveries.

Extragalactic astronomy entered the 21$^{\mathrm{st}}$ century reliant on a slew of color-color diagrams, number counts, and elementary methods of photometry. Two decades later and the field is unrecognizable. We have replaced three-band optical surveys over a few arcminutes with deep near-infrared selected samples spanning degree scales complemented by a wealth of multi-wavelength data spanning the entire electromagnetic spectrum. The recognition that redshifts could be estimated from coarsely sampled photometry transformed statistical studies of galaxies, enabling the study of rest-frame properties (L_{UV}, $\mathcal{M}$) before which were only accessible spectroscopically from biased samples. Photometry in particular has grown beyond apertures to include more detailed model-based fitting – in many ways thanks to improvements in computational technologies – which this work (Chapter 2) advances an immensely insightful yet *complementary* technique.

As we continue to gain an ever greater foothold on the distant universe, our increasingly deeper images become crowded. Already the deepest wide-area near-infrared imaging from UltraVISTA demands a careful assessment of source confusion which until now was the domain of infrared surveys from *Spitzer* and *Herschel*. Coming deep surveys from *Roman*, Rubin, *Euclid*, and even *JWST* will present rich galaxy fields dense enough to challenge both aperture and model-based photometry alike. Source detection and segmentation too will only become more difficult. The Farmer is one of a scant few attempts to confront them (Chapter 2). Although more work is required to perfect these profile-fitting techniques, they offer an attractive solution worth pursuing. After all, failing to recognize and address these challenges will mean forfeiting a complete understanding of galaxy evolution, from the earliest times down to the present day.

COSMOS is one of the foremost fields for studying the distant universe. Its recognition as such is due in no small part to the continued investment by some of the most in-demand astronomical facilities ever constructed. Our current paradigm of star-formation cessation and its relation to mass and environment were postulated based on COSMOS observations (Peng et al., 2010b); the survey pioneered weak lensing techniques now widely adopted as standard in the next generation of weak lensing surveys (Leauthaud et al., 2007; Massey et al., 2007); it has formed the foundation of our understanding of the growth of large scale structure; and has provided many if not most of the most distant galaxy candidates known thus far (Stefanon et al., 2019b; Bowler et al., 2020a; Weaver et al., 2021). The fact COSMOS was selected as the only large area deep field to be observed shortly by *JWST* in Cycle 1 further cements its legacy in galaxy evolution studies. Likewise, the COSMOS2020 catalog (Chapter 3) stands as a milestone in the twenty year long history of the COSMOS survey: two sets of photometry computed from two independent techniques, each processed by two independent SED fitting codes, yielding four-fold estimates of photometric redshifts for each of the ~1 million near-infrared selected galaxies. Although involving considerably more effort compared to single measurements, this increasingly necessary standard – exemplified by COSMOS2020 – reveals the misunderstandings of the past and dramatically increases our confidence in future efforts.

Despite continued innovation and progress, our understanding of galaxy evolution remains incomplete. In particular, massive quiescent galaxies – glimpsed from first deep sky observations by Messier and later identified by Hubble more than a century ago – are still a mystery. While resolved in astounding detail in the local universe, and dissected by integral field spectroscopy (e.g. Weaver et al., 2018), our insights into their past assembly and growth are limited to archaeological studies of star-formation histories with a strong reliance on simulations. While recently identified submilimeter-bright dusty star-forming galaxies assembling rapidly only a few Gyr after the big bang ($z \approx 2-5$; Casey et al. 2021) are an attractive progenitor population, their evolutionary lineage and variance therein are only just beginning to be studied. The unexpected finding of massive quiescent galaxies already matured also at $z \approx 2-5$ point to a yet unknown progenitor population at even earlier times which must have undergone a luminous assembly followed by a rapid star-formation cessation. Identifying large samples of massive quiescent galaxies remains challenging, due not only to their rarity requiring degree-scale surveys, but also in part to well-known photometric degeneracies with dust-reddened galaxies at lower-z. By utilizing COSMOS2020 to measure the galaxy stellar mass function out to $z \approx 7$ (Chapter 4), we contribute to the growing body of evidence that such massive galaxies indeed exist improbably early in the history of our universe. While the abundance of these systems appears comparable to those predicted by theory, definitive comparisons will require larger samples from both observations *and* simulations. Furthermore, a portion of these are newly discovered in COSMOS2020 due to their extremely red colors; possibly being a bridge population to other optically-dark systems being reported in FIR surveys. Confirming such a 'dark' region of the $z \approx 3-5$ galaxy stellar mass function occupied by massive star-forming galaxies obscured by dust may help explain the observed populations of massive quiescent galaxies in the local universe, and confidently establish them as a key transition stage in the assembly of massive quiescent galaxies.

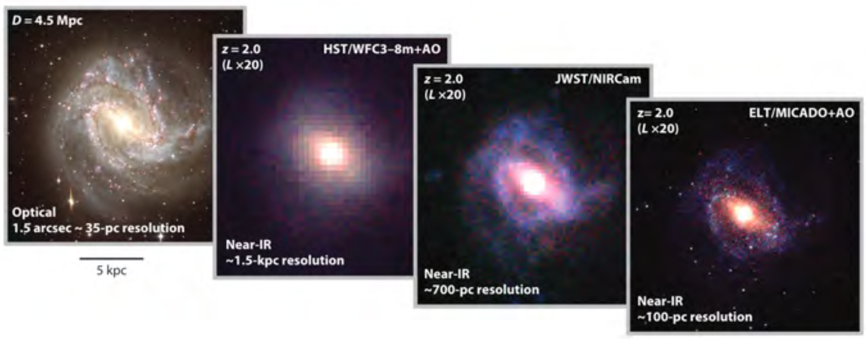

Figure 6.1: Our view into the universe has been ever-improving, thanks to advances in technology in concert with significant investments in facilities capable of pushing measurements to new heights. This is particularly true in the near-infrared, from which the distant universe is made accessible. *From Figure 11 of Förster Schreiber et al. 2020.*

JWST will spearhead the effort to understand the origins of massive quiescent galaxies, as well as the overdense environments in which they formed. While photometric surveys such as COSMOS-Web, PRIMER, JADES, CEERS, NGDEEP, and UNCOVER (Eisenstein et al., 2017; Finkelstein et al., 2017; Dunlop et al., 2021; Finkelstein et al., 2021; Kartaltepe et al., 2021; Labbe et al., 2021, respectively) will doubtless provide newly precise measurements, successfully interpreting them depends on follow-up spectroscopy of select sources. The immense mirror of *JWST* will enable studies beyond the $z \gg 5$ emission line astrophysics championed by *Hubble* by accessing the rest-frame optical stellar continuum for the first time. Chapter 5 introduces the Beasts in the Bubbles (PI:Weaver), a *JWST* Cycle 1 program which represents our deepest look into the formation *and diversity* of ultra-luminous and presumably massive cosmic beasts seen only moments after the big bang – the expected progenitors of $z \approx 4-5$ massive quiescent galaxies. By pushing *JWST* to its limits, these observations will mark a new era of precision astrophysics of galaxies within the Epoch of Reionization. Yet despite the awesome vision of *JWST* (Figure 6.1), the total area expected to be observed within its nominal 5-year mission is no more than a few square degrees – insufficient to identify *new* and sufficiently *large* samples of massive galaxies.

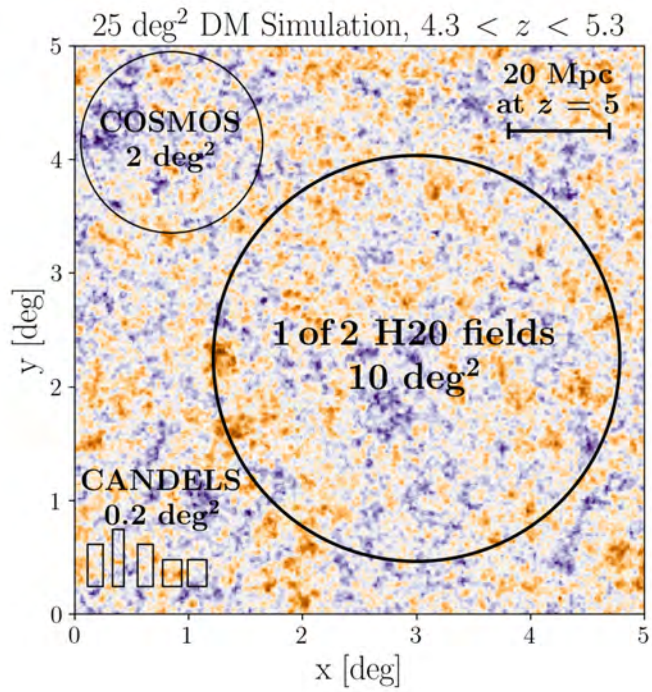

Figure 6.2: The dark matter density distribution at $z \approx 4-5$ from the simulations of Rácz et al. (2021). Smaller fields like CANDELS and even COSMOS contain limited power to assess the full range of environments expected to vary on degree-scales at these early times. They will be dwarfed by the two 10 deg^2 Euclid Deep Fields with optical imaging made possible by the Hawaii 2-0 Survey. *From Figure 2 of McPartland et al., in prep.*

Future investigations into the assembly of the most massive galaxies within this formative epoch will be conducted over still larger volumes than probed even by COSMOS. This is not only to improve the poisson noise statistics within the known high-$\mathcal{M}$ domain measured by existing wide-area NIR surveys, but also to access even rarer populations of ultra-massive galaxies, if they exist. In this respect, the launch of *Euclid* will mark a new era in galaxy evolution studies by providing tens of millions of galaxies measured with comparable precision to existing deep fields. Such large samples will mitigate both poisson noise and cosmic variance leaving systematic biases as the dominant source of error. While *Euclid* will provide high-resolution optical images used for definitive weak lensing measurements across much of the sky, it will also provide particularly deep NIR coverage over its deep fields. Already our team has acquired unique deep optical and infrared imaging from the Hawaii 2-0 (PI: Sanders; see McPartland et al. in prep.) and Spitzer Legacy surveys (PI: Capak, see Moneti et al. 2021), respectively. This dataset, unprecedented over a combined 20 deg^2 area split equally between the Euclid Deep Field North and South (Figure 6.2), represents the largest sample of galaxies with directly measured stellar masses within a cosmologically significant volume available for at least the next decade. Early results based on photometry measured with The Farmer have already led to the identification of dozens of $z > 3$ protoclusters which, spectroscopically confirmed, are a sizable fraction of all protoclusters reported in the literature.

The next generation of wide-area imaging facilities will certainly outpace multiplexed spectroscopic surveys in discovering new galaxies such that the fraction of spectroscopically confirmed galaxies will remain less than 1% of known sources. This is a situation unlikely to change even in the far future. Thankfully, photometric samples are typically less biased and more complete, and lend themselves to estimates of key functions and scaling relations with constantly improving statistical precision. However, such functions and relations can only tell us so much; they provide limited insight into the physics underlying them, which in themselves are often degenerate and sensitive to observational biases. Differences in sample selections make comparisons with other observational results hazardous and further slow the pace towards scientific consensus. Worse, the otherwise invaluable insights of cosmological simulations are often difficult to interpret due to the widespread incongruity in defining and obtaining physical parameters, in addition to complications from numerical limitations such as volume, spatial resolution, and object differentiation. These challenges remain endemic to our study; and yet overcoming them promises a consensus view of key observables, meaningful insight into the underlying physics and, ultimately, the emergence of a unified theory of galaxy evolution.

In the heaven that most of his light receives have I been, and I have seen things, to recount which, descending, I neither know how nor have the power.
For nearing its desired end, our intellect sinks into an abyss so deep that memory fails to follow it.

- Dante, *Paradiso I*

BIBLIOGRAPHY

Abazajian, K. N., J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, et al. In: *ApJS* 182.2 (June 2009), pp. 543–558.

Abdurro'uf and M. Akiyama. In: *Monthly Notices of the RAS* 479.4 (Oct. 2018), pp. 5083–5100.

Abell, G. O. In: *The Astrophysical Journal Supplement Series* 3 (May 1958), p. 211.

Abell, G. O., H. G. Corwin Jr., and R. P. Olowin. In: *The Astrophysical Journal Supplement Series* 70 (May 1989), p. 1.

Abramson, L. E., M. D. Gladders, A. Dressler, A. Oemler Jr., et al. In: *The Astrophysical Journal* 832 (Nov. 2016), p. 7.

Adams, N. J., R. A. A. Bowler, M. J. Jarvis, B. Häußler, and C. D. P. Lagos. In: *Monthly Notices of the RAS* 506.4 (Oct. 2021), pp. 4933–4951.

Adams, N. J., R. A. A. Bowler, M. J. Jarvis, B. Häußler, et al. In: *Monthly Notices of the Royal Astronomical Society* 494 (May 2020), pp. 1771–1783.

Ahn, C. P., R. Alexandroff, C. Allende Prieto, F. Anders, et al. In: *ApJS* 211.2 (Mar. 2014), p. 17.

Aihara, H., C. Allende Prieto, D. An, S. F. Anderson, et al. In: *Astrophysical Journal, Supplement* 193.2, 29 (Apr. 2011), p. 29.

Aihara, H., Y. AlSayyad, M. Ando, R. Armstrong, et al. In: *Publications of the Astronomical Society of Japan* 71 (Dec. 2019), p. 114.

Aird, J., K. Nandra, E. S. Laird, A. Georgakakis, et al. In: *Monthly Notices of the Royal Astronomical Society* 401 (Feb. 2010), pp. 2531–2551.

Akhshik, M., K. E. Whitaker, G. Brammer, G. Mahler, et al. In: *The Astrophysical Journal* 900 (Sept. 2020), p. 184.

Akhshik, M., K. E. Whitaker, J. Leja, J. Richard, et al. In: *arXiv e-prints*, arXiv:2203.04979 (Mar. 2022), arXiv:2203.04979.

Alatalo, K., S. L. Cales, J. A. Rich, P. N. Appleton, et al. In: *The Astrophysical Journal Supplement Series* 224 (June 2016), p. 38.

Albrecht, A., G. Bernstein, R. Cahn, W. L. Freedman, et al. In: *arXiv e-prints*, astro-ph/0609591 (Sept. 2006), astro–ph/0609591.

Allard, F., D. Homeier, and B. Freytag. In: *Philosophical Transactions of the Royal Society of London Series A* 370.1968 (June 2012), pp. 2765–2777.

Almaini, O., V. Wild, D. Maltby, W. Hartley, et al. In: *Monthly Notices of the Royal Astronomical Society* 472.2 (2017/Aug/04/), pp. 1401–1412.

Alpher, R. A., H. Bethe, and G. Gamow. In: *Physical Review* 73 (Apr. 1948), pp. 803–804.

Alvarez, M. A. and T. Abel. In: *Monthly Notices of the Royal Astronomical Society* 380 (Sept. 2007), pp. L30–L34.

Anderson, L., F. Governato, M. Karcher, T. Quinn, and J. Wadsley. In: *Monthly Notices of the Royal Astronomical Society* 468 (July 2017), pp. 4077–4092.

Angus, G. W., B. Famaey, and H. S. Zhao. In: *Monthly Notices of the Royal Astronomical Society* 371.1 (Aug. 2006), pp. 138–146.

Aragon-Salamanca, A., C. M. Baugh, and G. Kauffmann. In: *Monthly Notices of the Royal Astronomical Society* 297 (June 1998), pp. 427–434.

Arcila-Osejo, L. and M. Sawicki. In: *Monthly Notices of the RAS* 435.1 (Oct. 2013), pp. 845–860.

Arnouts, S., E. Le Floc'h, J. Chevallard, B. D. Johnson, et al. In: *Astronomy and Astrophysics* 558, A67 (Oct. 2013), A67.

Arnouts, S., E. Le Floc'h, J. Chevallard, B. D. Johnson, et al. In: *Astronomy and Astrophysics* 558, A67 (Oct. 2013), A67.

Arnouts, S., L. Moscardini, E. Vanzella, S. Colombi, et al. In: *Monthly Notices of the Royal Astronomical Society* 329 (Jan. 2002), pp. 355–366.

Arnouts, S., C. J. Walcher, O. Le Fèvre, G. Zamorani, et al. In: *Astronomy and Astrophysics* 476.1 (Dec. 2007), pp. 137–150.

Arp, H. In: *The Astrophysical Journal Supplement Series* 14 (Nov. 1966), p. 1.

Ashby, M. L. N., K. I. Caputi, W. Cowley, S. Deshmukh, et al. In: *The Astrophysical Journal Supplement Series* 237 (Aug. 2018), p. 39.

Ashby, M. L. N., S. P. Willner, G. G. Fazio, J. S. Dunlop, et al. In: *The Astrophysical Journal Supplement Series* 218 (June 2015), p. 33.

Ashby, M. L. N., S. P. Willner, G. G. Fazio, J.-S. Huang, et al. In: *The Astrophysical Journal* 769 (May 2013), p. 80.

Atek, H., J. Richard, M. Jauzac, J.-P. Kneib, et al. In: *The Astrophysical Journal* 814 (Nov. 2015), p. 69.

Baade, W. In: *The Astrophysical Journal* 100 (Sept. 1944), p. 137.

Babcock, H. W. In: *Lick Observatory Bulletin* 498 (Jan. 1939), pp. 41–51.

Baldry, I. K., M. L. Balogh, R. G. Bower, K. Glazebrook, et al. In: *Monthly Notices of the Royal Astronomical Society* 373 (Dec. 2006), pp. 469–483.

Baldry, I. K. In: *Astronomy and Geophysics* 49.5 (Oct. 2008), pp. 5.25–5.26.

Balogh, M. L., S. L. Morris, H. K. C. Yee, R. G. Carlberg, and E. Ellingson. In: *The Astrophysical Journal* 527 (Dec. 1999), pp. 54–79.

Balogh, M. L., J. F. Navarro, and S. L. Morris. In: *Astrophysical Journal* 540.1 (Sept. 2000), pp. 113–121.

Bañados, E., B. P. Venemans, C. Mazzucchelli, E. P. Farina, et al. In: *Nature* 553.7689 (Jan. 2018), pp. 473–476.

Baraffe, I., D. Homeier, F. Allard, and G. Chabrier. In: *Astronomy and Astrophysics* 577, A42 (May 2015), A42.

Barbary, K. In: *The Journal of Open Source Software* 1.6 (Oct. 2016), p. 58.

Barnes, J. E. In: *The Astrophysical Journal* 331 (Aug. 1988), p. 699.

Barnes, J. E. and L. Hernquist. In: *Annual Review of Astronomy and Astrophysics* 30 (1992), pp. 705–742.

Barnes, J. E. and L. Hernquist. In: *Nature* 360 (Dec. 1992), pp. 715–717.

Barnes, J. E. and L. Hernquist. In: *The Astrophysical Journal* 471 (Nov. 1996), p. 115.

Barnes, J. E. and L. E. Hernquist. In: *The Astrophysical Journal* 370 (Apr. 1991), p. L65.

Barro, G., S. M. Faber, D. C. Koo, A. Dekel, et al. In: *The Astrophysical Journal* 840 (May 2017), p. 47.

Barrufet, L., P. Oesch, Y. Fudamoto, G. D. Illingworth, et al. JWST Proposal. Cycle 1. Mar. 2021.

Barton, E. J., J. A. Arnold, A. R. Zentner, J. S. Bullock, and R. H. Wechsler. In: *The Astrophysical Journal* 671 (Dec. 2007), pp. 1538–1549.

Barton, E. J., M. J. Geller, and S. J. Kenyon. In: *The Astrophysical Journal* 530 (Feb. 2000), pp. 660–679.

Baum, W. A. In: *Problems of Extra-Galactic Research* 15 (Jan. 1962), p. 390.

Becker, G. D., J. S. Bolton, and A. Lidz. In: *Publ. Astron. Soc. Aust.* 32 (2015), e045.

Becker, R. H., X. Fan, R. L. White, M. A. Strauss, et al. In: *The Astronomical Journal* 122 (Dec. 2001), pp. 2850–2857.

Beckwith, S. V. W., M. Stiavelli, A. M. Koekemoer, J. A. R. Caldwell, et al. In: *Astronomical Journal* 132.5 (Nov. 2006), pp. 1729–1755.

Behroozi, P. and J. Silk. In: *Monthly Notices of the RAS* 477.4 (July 2018), pp. 5382–5387.

Behroozi, P. and J. Silk. In: *Monthly Notices of the RAS* 477.4 (July 2018), pp. 5382–5387.

Behroozi, P., R. H. Wechsler, A. P. Hearin, and C. Conroy. In: *Monthly Notices of the RAS* 488.3 (Sept. 2019), pp. 3143–3194.

Behroozi, P. S., C. Conroy, and R. H. Wechsler. In: *Astrophysical Journal* 717.1 (July 2010), pp. 379–403.

Behroozi, P. S., R. H. Wechsler, and C. Conroy. In: *Astrophysical Journal* 770.1, 57 (June 2013), p. 57.

Bekki, K., W. J. Couch, Y. Shioya, and A. Vazdekis. In: *Monthly Notices of the Royal Astronomical Society* 359 (May 2005), pp. 949–965.

Bekki, K., W. J. Couch, and Y. Shioya. In: *Publications of the Astronomical Society of Japan* 53.3 (June 2001), pp. 395–400.

Belfiore, F., R. Maiolino, K. Bundy, K. Masters, et al. In: *Monthly Notices of the Royal Astronomical Society* 477 (July 2018), pp. 3014–3029.

Bell, E. F., D. H. McIntosh, N. Katz, and M. D. Weinberg. In: *The Astrophysical Journal Supplement Series* 149 (Dec. 2003), pp. 289–312.

Bell, E. F., T. Naab, D. H. McIntosh, R. S. Somerville, et al. In: *The Astrophysical Journal* 640 (Mar. 2006), pp. 241–251.

Bell, E. F., S. Phleps, R. S. Somerville, C. Wolf, et al. In: *The Astrophysical Journal* 652 (Nov. 2006), pp. 270–276.

Belli, S., A. B. Newman, and R. S. Ellis. In: *Astrophysical Journal* 874.1, 17 (Mar. 2019), p. 17.

Benson, A. J., R. G. Bower, C. S. Frenk, C. G. Lacey, et al. In: *The Astrophysical Journal* 599 (Dec. 2003), pp. 38–49.

Bernardi, M., R. K. Sheth, J. Annis, S. Burles, et al. In: *The Astronomical Journal* 125 (Apr. 2003), pp. 1866–1881.

Bertin, E. In: *Astronomical Data Analysis Software and Systems XV*. Ed. by C. Gabriel, C. Arviset, D. Ponz, and S. Enrique. Vol. 351. Astronomical Society of the Pacific Conference Series. July 2006, p. 112.

Bertin, E. and S. Arnouts. In: *A&AS* 117 (June 1996), pp. 393–404.

Bertin, E., M. Schefer, N. Apostolakos, A. Álvarez-Ayllón, et al. In: *Astronomical Data Analysis Software and Systems XXIX*. Ed. by R. Pizzo, E. R. Deul, J. D. Mol, J. de Plaa, and H. Verkouter. Vol. 527. Astronomical Society of the Pacific Conference Series. Jan. 2020, p. 461.

Bertin, E. Oct. 2010.

Bertin, E. Oct. 2010.

Bertin, E. In: *Astrophysics Source Code Library* (Jan. 2013), ascl:1301.001.

Bertin, E., Y. Mellier, M. Radovich, G. Missonnier, et al. In: *Astronomical Data Analysis Software and Systems XI*. Ed. by D. A. Bohlender, D. Durand, and T. H. Handley. Vol. 281. Astronomical Society of the Pacific Conference Series. Jan. 2002, p. 228.

Bezanson, R., M. Franx, and P. G. van Dokkum. In: *The Astrophysical Journal* 799 (Feb. 2015), p. 148.

Bhowmick, A. K., R. S. Somerville, T. Di Matteo, S. Wilkins, et al. In: *Monthly Notices of the RAS* 496.1 (July 2020), pp. 754–766.

Bialas, D., T. Lisker, C. Olczak, R. Spurzem, and R. Kotulla. In: *Astronomy and Astrophysics* 576 (Apr. 2015), A103.

Bielby, R., P. Hudelot, H. J. McCracken, O. Ilbert, et al. In: *antike und abendland* 545 (Sept. 2012), A23.

Bieri, R., Y. Dubois, J. Rosdahl, A. Wagner, et al. In: *Monthly Notices of the RAS* 464.2 (Jan. 2017), pp. 1854–1873.

Bigiel, F., A. Leroy, F. Walter, E. Brinks, et al. In: *The Astronomical Journal* 136 (Dec. 2008), pp. 2846–2871.

Bigourdan, G. In: *Bulletin Astronomique, Serie I* 5 (Jan. 1888), pp. 303–309.

Binggeli, B. and H. Jerjen. In: *Astronomy and Astrophysics* 333 (May 1998), pp. 17–26.

Binney, J. In: *The Astrophysical Journal* 215 (July 1977), pp. 483–491.

Birnboim, Y. and A. Dekel. In: *Monthly Notices of the RAS* 345.1 (Oct. 2003), pp. 349–364.

Bluck, A. F. L., J. T. Mendel, S. L. Ellison, D. R. Patton, et al. In: *Monthly Notices of the Royal Astronomical Society* 462 (Nov. 2016), pp. 2559–2586.

Blumenthal, G. R., S. M. Faber, J. R. Primack, and M. J. Rees. In: *Nature* 311.5986 (Oct. 1984), pp. 517–525.

Bocquet, S., A. Saro, K. Dolag, and J. J. Mohr. In: *Monthly Notices of the Royal Astronomical Society* 456 (Mar. 2016), pp. 2361–2373.

Boesgaard, A. M. and G. Steigman. In: *Annual Review of Astronomy and Astrophysics* 23 (1985), pp. 319–378.

Bohlin, R. C., L. Colina, and D. S. Finley. In: *Astronomical Journal* 110 (Sept. 1995), p. 1316.

Bolan, P., B. C. Lemaux, C. Mason, M. Bradač, et al. Nov. 2021.

Bolton, J. S. and M. G. Haehnelt. In: *Monthly Notices of the Royal Astronomical Society* 374 (Jan. 2007), pp. 493–514.

Bolzonella, M., K. Kovač, L. Pozzetti, E. Zucca, et al. In: *Astronomy and Astrophysics* 524, A76 (Dec. 2010), A76.

Bondi, H. and T. Gold. In: *Monthly Notices of the Royal Astronomical Society* 108 (Jan. 1948), p. 252.

Borne, K. D. and D. O. Richstone. In: *The Astrophysical Journal* 369 (Mar. 1991), p. 111.

Borthakur, S., T. M. Heckman, C. Leitherer, and R. A. Overzier. In: *Science* 346 (Oct. 2014), pp. 216–219.

Boselli, A., M. Fossati, and M. Sun. In: *Astronomy and Astrophysics Review* 30.1 (Dec. 2022), p. 3.

Bothwell, M. S., J. Wagg, C. Cicone, R. Maiolino, et al. In: *Monthly Notices of the RAS* 445.3 (Dec. 2014), pp. 2599–2620.

Bouché, N., A. Dekel, R. Genzel, S. Genel, et al. In: *Astrophysical Journal* 718.2 (Aug. 2010), pp. 1001–1018.

Boulade, O., X. Charlot, P. Abbon, S. Aune, et al. In: ed. by M. Iye and A. F. M. Moorwood. Vol. 4841. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Mar. 2003, pp. 72–81.

Bouwens, R. J., L. Bradley, A. Zitrin, D. Coe, et al. In: *Astrophysical Journal* 795.2, 126 (Nov. 2014), p. 126.

Bouwens, R. J., G. D. Illingworth, M. Franx, R. .-. Chary, et al. In: *The Astrophysical Journal* 705 (Nov. 2009), pp. 936–961.

Bouwens, R. J., G. D. Illingworth, M. Franx, and H. Ford. In: *The Astrophysical Journal* 670 (Dec. 2007), pp. 928–958.

Bouwens, R. J., G. D. Illingworth, P. A. Oesch, M. Franx, et al. In: *Astrophysical Journal* 754.2, 83 (Aug. 2012), p. 83.

Bouwens, R. J., G. D. Illingworth, P. A. Oesch, I. Labbé, et al. In: *The Astrophysical Journal* 737 (Aug. 2011), p. 90.

Bouwens, R. J., G. D. Illingworth, P. A. Oesch, I. Labbé, et al. In: *Astrophysical Journal* 793.2, 115 (Oct. 2014), p. 115.

Bouwens, R. J., G. D. Illingworth, P. A. Oesch, M. Trenti, et al. In: *Astrophysical Journal, Letters* 752.1, L5 (June 2012), p. L5.

Bouwens, R. J., P. A. Oesch, M. Stefanon, G. Illingworth, et al. In: *Astronomical Journal* 162.2, 47 (Aug. 2021), p. 47.

Bouwens, R. J., R. Smit, S. Schouws, M. Stefanon, et al. In: *arXiv e-prints*, arXiv:2106.13719 (June 2021), arXiv:2106.13719.

Bouwens, R., G. Illingworth, P. Oesch, M. Trenti, et al. In: *The Astrophysical Journal* 803.1 (2015), p. 34.

Bouwens, R. J. and G. D. Illingworth. In: *Nature* 443 (Sept. 2006), pp. 189–192.

Bowler, R. A. A., J. S. Dunlop, R. J. McLure, H. J. McCracken, et al. In: *Monthly Notices of the Royal Astronomical Society* 426.4 (Nov. 2012), pp. 2772–2788.

Bowler, R. A. A., J. S. Dunlop, R. J. McLure, H. J. McCracken, et al. In: *Monthly Notices of the Royal Astronomical Society* 452 (Sept. 2015), pp. 1817–1840.

Bowler, R. A. A., J. S. Dunlop, R. J. McLure, and D. J. McLeod. In: *Monthly Notices of the RAS* 466.3 (Apr. 2017), pp. 3612–3635.

Bowler, R. A. A., J. S. Dunlop, R. J. McLure, and D. J. McLeod. In: *\mnras* 466.3 (Apr. 2017), pp. 3612–3635.

Bowler, R. A. A., M. J. Jarvis, J. S. Dunlop, R. J. McLure, et al. In: *Monthly Notices of the RAS* 493.2 (Apr. 2020), pp. 2059–2084.

Bowler, R. A. A., M. J. Jarvis, J. S. Dunlop, R. J. McLure, et al. In: *Monthly Notices of the RAS* 493.2 (Apr. 2020), pp. 2059–2084.

Brammer, G. B., P. G. van Dokkum, and P. Coppi. In: *Astrophysical Journal* 686.2 (Oct. 2008), pp. 1503–1513.

Brammer, G. B., P. G. van Dokkum, M. Franx, M. Fumagalli, et al. In: *The Astrophysical Journal Supplement Series* 200 (June 2012), p. 13.

Branchesi, M., I. M. Gioia, C. Fanti, R. Fanti, and R. Perley. In: *Astronomy and Astrophysics* 446.1 (Jan. 2006), pp. 97–111.

Brescia, M., S. Cavuoti, O. Razim, V. Amaro, et al. In: *Front. Astron. Space Sci.* 8 (June 2021), p. 658229.

Brinchmann, J., S. Charlot, S. D. M. White, C. Tremonti, et al. In: *Monthly Notices of the RAS* 351.4 (July 2004), pp. 1151–1179.

Bromm, V. In: *Reports on Progress in Physics* 76 (Nov. 2013), p. 112901.

Bromm, V. and N. Yoshida. In: *Annual Review of Astronomy and Astrophysics, vol. 49, issue 1, pp. 373-407* 49.1 (Sept. 2011), p. 373.

Brown, G. S. and B. M. Tinsley. In: *The Astrophysical Journal* 194 (Dec. 1974), pp. 555–558.

Bruzual, G. and S. Charlot. In: *Monthly Notices of the Royal Astronomical Society* 344 (Oct. 2003), pp. 1000–1028.

Bundy, K., M. A. Bershady, D. R. Law, R. Yan, et al. In: *The Astrophysical Journal* 798 (Jan. 2015), p. 7.

Bundy, K., A. Georgakakis, K. Nandra, R. S. Ellis, et al. In: *The Astrophysical Journal* 681 (July 2008), pp. 931–943.

Burke, C., M. Hilton, and C. Collins. In: *Monthly Notices of the Royal Astronomical Society* 449 (May 2015), pp. 2353–2367.

Butcher, H. and A. Oemler Jr. In: *The Astrophysical Journal* 285 (Oct. 1984), pp. 426–438.

Calzetti, D., L. Armus, R. C. Bohlin, A. L. Kinney, et al. In: *The Astrophysical Journal* 533 (Apr. 2000), pp. 682–695.

Calzetti, D., L. Armus, R. C. Bohlin, A. L. Kinney, et al. In: *The Astrophysical Journal* 533 (Apr. 2000), pp. 682–695.

Capak, P., H. Aussel, M. Ajiki, H. J. McCracken, et al. In: *The Astrophysical Journal Supplement Series* 172 (Sept. 2007), pp. 99–116.

Capak, P. L., D. Riechers, N. Z. Scoville, C. Carilli, et al. In: *Nature* 470.7333 (Feb. 2011), pp. 233–235.

Caplar, N. and S. Tacchella. In: *Monthly Notices of the RAS* 487.3 (Aug. 2019), pp. 3845–3869.

Cappellari, M. and Y. Copin. In: *Monthly Notices of the Royal Astronomical Society* 342.2 (June 2003), pp. 345–354.

Cappellari, M. In: *Mon. Not. R. Astron. Soc.* 466.1 (Apr. 2017), pp. 798–811.

Cappellari, M., E. Emsellem, D. Krajnović, R. M. McDermid, et al. In: *Monthly Notices of the Royal Astronomical Society* 413 (May 2011), pp. 813–836.

Caputi, K. I., M. Cirasuolo, J. S. Dunlop, R. J. McLure, et al. In: *Monthly Notices of the Royal Astronomical Society* 413.1 (May 2011), pp. 162–176.

Caputi, K. I., H. Dole, G. Lagache, R. J. McLure, et al. In: *Astronomy and Astrophysics* 454.1 (July 2006), pp. 143–150.

Caputi, K. I., J. S. Dunlop, R. J. McLure, J. .-. Huang, et al. In: *The Astrophysical Journal* 750 (May 2012), p. L20.

Caputi, K. I., O. Ilbert, C. Laigle, H. J. McCracken, et al. In: *Astrophysical Journal* 810.1, 73 (Sept. 2015), p. 73.

Caputi, K. I., K. Kovač, M. Bolzonella, S. J. Lilly, et al. In: *The Astrophysical Journal* 691 (Jan. 2009), pp. 91–97.

Caputi, K. I., G. Lagache, L. Yan, H. Dole, et al. In: *The Astrophysical Journal* 660 (May 2007), pp. 97–116.

Caputi, K. I., R. J. McLure, J. S. Dunlop, M. Cirasuolo, and A. M. Schael. In: *Monthly Notices of the Royal Astronomical Society* 366 (Feb. 2006), pp. 609–623.

Carliles, S., T. Budavári, S. Heinis, C. Priebe, and A. S. Szalay. In: *The Astrophysical Journal* 712 (Mar. 2010), pp. 511–515.

Carnall, A. C., R. J. McLure, J. S. Dunlop, and R. Davé. In: *Monthly Notices of the RAS* 480.4 (Nov. 2018), pp. 4379–4401.

Carnall, A. C., R. J. McLure, J. S. Dunlop, and R. Davé. In: *Monthly Notices of the RAS* 480.4 (Nov. 2018), pp. 4379–4401.

Carnall, A. C., S. Walker, R. J. McLure, J. S. Dunlop, et al. In: *Monthly Notices of the Royal Astronomical Society* 496 (July 2020), pp. 695–707.

Carnall, A., R. A. Begley, A. Cimatti, F. Cullen, et al. JWST Proposal. Cycle 1. Mar. 2021.

Carnall, A. C., J. Leja, B. D. Johnson, R. J. McLure, et al. In: *Astrophysical Journal* 873.1, 44 (Mar. 2019), p. 44.

Carnall, A. C., J. Leja, B. D. Johnson, R. J. McLure, et al. In: *Astrophysical Journal* 873.1, 44 (Mar. 2019), p. 44.

Carrasco Kind, M. and R. J. Brunner. In: *Monthly Notices of the Royal Astronomical Society* 432 (June 2013), pp. 1483–1501.

Casey, C. M., S. Berta, M. Béthermin, J. Bock, et al. In: *Astrophysical Journal* 761.2, 140 (Dec. 2012), p. 140.

Casey, C. M., A. Cooray, P. Capak, H. Fu, et al. In: *Astrophysical Journal, Letters* 808.2, L33 (Aug. 2015), p. L33.

Casey, C. M. In: *The Astrophysical Journal* 824 (June 2016), p. 36.

Casey, C. M., D. Narayanan, and A. Cooray. In: *Physics Reports* 541.2 (Aug. 2014), pp. 45–161.

Casey, C. M., J. A. Zavala, M. Aravena, M. Béthermin, et al. In: *Astrophysical Journal* 887.1, 55 (Dec. 2019), p. 55.

Casey, C. M., J. A. Zavala, S. M. Manning, M. Aravena, et al. In: *Astrophysical Journal* 923.2, 215 (Dec. 2021), p. 215.

Castellano, M., P. Dayal, L. Pentericci, A. Fontana, et al. In: *The Astrophysical Journal* 818 (Feb. 2016), p. L3.

Cecchi, R., M. Bolzonella, A. Cimatti, and G. Girelli. In: *The Astrophysical Journal* 880 (July 2019), p. L14.

Chabrier, G., I. Baraffe, F. Allard, and P. Hauschildt. In: *Astrophysical Journal* 542.1 (Oct. 2000), pp. 464–472.

Chabrier, G. In: *Publications of the ASP* 115.809 (July 2003), pp. 763–795.

Chanamé, J., L. Infante, and A. Reisenegger. In: *The Astrophysical Journal* 530 (Feb. 2000), pp. 96–106.

Chen, Y.-M., Y. Shi, V. Wild, C. Tremonti, et al. In: *Monthly Notices of the Royal Astronomical Society* 489 (Nov. 2019), pp. 5709–5722.

Cheung, E., K. Bundy, M. Cappellari, S. Peirani, et al. In: *Nature* 533 (May 2016), pp. 504–508.

Chilingarian, I. V., P. Di Matteo, F. Combes, A.-L. Melchior, and B. Semelin. In: *Astronomy and Astrophysics* 518 (July 2010), A61.

Christensen, C. G. In: *The Astronomical Journal* 80 (Apr. 1975), p. 282.

Cicone, C., R. Maiolino, S. Gallerani, R. Neri, et al. In: *Astronomy and Astrophysics* 574, A14 (Feb. 2015), A14.

Cicone, C., R. Maiolino, E. Sturm, J. Graciá-Carpio, et al. In: *Astronomy and Astrophysics* 562, A21 (Feb. 2014), A21.

Cimatti, A., E. Daddi, A. Renzini, P. Cassata, et al. In: *Nature* 430 (July 2004), pp. 184–187.

Ciotti, L., B. Lanzoni, and M. Volonteri. In: *The Astrophysical Journal* 658 (Mar. 2007), pp. 65–77.

Cirasuolo, M., J. Afonso, M. Carollo, H. Flores, et al. In: *Ground-based and Airborne Instrumentation for Astronomy V* 9147 (July 2014), 91470N.

Cisternas, M., K. Jahnke, K. J. Inskip, J. Kartaltepe, et al. In: *The Astrophysical Journal* 726 (Jan. 2011), p. 57.

Civano, F., S. Marchesi, A. Comastri, M. C. Urry, et al. In: *Astrophysical Journal* 819.1, 62 (Mar. 2016), p. 62.

Civano, F., S. Marchesi, A. Comastri, M. C. Urry, et al. In: *The Astrophysical Journal* 819 (Mar. 2016), p. 62.

Clowe, D., A. Gonzalez, and M. Markevitch. In: *Astrophysical Journal* 604.2 (Apr. 2004), pp. 596–603.

Cole, S., P. Norberg, C. M. Baugh, C. S. Frenk, et al. In: *Monthly Notices of the RAS* 326.1 (Sept. 2001), pp. 255–273.

Collaboration, P., P. a. R. Ade, N. Aghanim, C. Armitage-Caplan, et al. In: *Astronomy and Astrophysics* 571 (Nov. 2014), A16.

Collins, C. A., J. P. Stott, M. Hilton, S. T. Kay, et al. In: *Nature* 458.7238 (Apr. 2009), pp. 603–606.

Collister, A. A. and O. Lahav. In: *Publications of the Astronomical Society of the Pacific* 116 (Apr. 2004), pp. 345–351.

Combes, F. Nov. 2019.

Comparat, J., J. Richard, J.-P. Kneib, O. Ilbert, et al. In: *Astronomy and Astrophysics* 575, A40 (Mar. 2015), A40.

Connolly, A. J., I. Csabai, A. S. Szalay, D. C. Koo, et al. In: *The Astronomical Journal* 110 (Dec. 1995), p. 2655.

Conroy, C. In: *Annual Review of Astronomy and Astrophysics* 51.1 (2013), pp. 393–455.

Conroy, C. and J. E. Gunn. In: *Astrophysical Journal* 712.2 (Apr. 2010), pp. 833–857.

Conroy, C. and J. E. Gunn. In: *Astrophysical Journal* 712.2 (Apr. 2010), pp. 833–857.

Conroy, C., J. E. Gunn, and M. White. In: *Astrophysical Journal* 699.1 (July 2009), pp. 486–506.

Conroy, C., J. E. Gunn, and M. White. In: *Astrophysical Journal* 699.1 (July 2009), pp. 486–506.

Conroy, C. and R. H. Wechsler. In: *Astrophysical Journal* 696.1 (May 2009), pp. 620–635.

Conselice, C. J. In: *The Astrophysical Journal Supplement Series* 147 (July 2003), pp. 1–28.

Cooper, O. R., C. M. Casey, J. A. Zavala, J. B. Champagne, et al. In: *The Astrophysical Journal* 930 (May 2022), p. 32.

Cooray, A. and M. Milosavljević. In: *The Astrophysical Journal* 627 (July 2005), pp. L89–L92.

Cortzen, I., J. Garrett, G. Magdis, D. Rigopoulou, et al. In: *Monthly Notices of the Royal Astronomical Society* 482 (Jan. 2019), pp. 1618–1633.

Coupon, J., O. Ilbert, M. Kilbinger, H. J. McCracken, et al. In: *Astronomy and Astrophysics* 500.3 (June 2009), pp. 981–998.

Coupon, J., N. Czakon, J. Bosch, Y. Komiyama, et al. In: *Publications of the ASJ* 70, S7 (Jan. 2018), S7.

Cowie, L. L., A. Songaila, E. M. Hu, and J. G. Cohen. In: *Astronomical Journal* 112 (Sept. 1996), p. 839.

Cowley, W. I., K. I. Caputi, S. Deshmukh, M. L. N. Ashby, et al. In: *Astrophysical Journal* 874.2, 114 (Apr. 2019), p. 114.

Cox, T. J., P. Jonsson, R. S. Somerville, J. R. Primack, and A. Dekel. In: *Monthly Notices of the Royal Astronomical Society* 384 (Feb. 2008), pp. 386–409.

Croton, D. J. In: *Publications of the Astron. Soc. of Australia* 30, e052 (Oct. 2013), e052.

Croton, D. J., V. Springel, S. D. M. White, G. De Lucia, et al. In: *Monthly Notices of the Royal Astronomical Society* 365.1 (Jan. 2006), pp. 11–28.

Cucciati, O., B. C. Lemaux, G. Zamorani, O. Le Fèvre, et al. In: *Astronomy and Astrophysics* 619, A49 (Nov. 2018), A49.

D'Agostino, R. B. In: *Biometrika* 57.3 (Dec. 1970), pp. 679–681.

D'Agostino, R. B. and A. Belanger. In: *The American Statistician* 44.4 (1990), pp. 316–321.

D'Isanto, A. and K. L. Polsterer. In: *Astronomy & Astrophysics, Volume 609, id.A111* 609 (Jan. 2018), A111.

Daddi, E., A. Cimatti, A. Renzini, A. Fontana, et al. In: *Astrophysical Journal* 617.2 (Dec. 2004), pp. 746–764.

Daddi, E., M. Dickinson, R. Chary, A. Pope, et al. In: *The Astrophysical Journal* 631 (Sept. 2005), pp. L13–L16.

Daddi, E., M. Dickinson, G. Morrison, R. Chary, et al. In: *Astrophysical Journal* 670.1 (Nov. 2007), pp. 156–172.

Daddi, E., A. Renzini, N. Pirzkal, A. Cimatti, et al. In: *The Astrophysical Journal* 626 (June 2005), pp. 680–697.

Darg, D. W., S. Kaviraj, C. J. Lintott, K. Schawinski, et al. In: *Monthly Notices of the Royal Astronomical Society* 401 (Jan. 2010), pp. 1552–1563.

Davé, R., D. Anglés-Alcázar, D. Narayanan, Q. Li, et al. In: *Monthly Notices of the RAS* 486.2 (June 2019), pp. 2827–2849.

Davidzon, I., O. Cucciati, M. Bolzonella, G. De Lucia, et al. In: *Astronomy and Astrophysics* 586, A23 (Feb. 2016), A23.

Davidzon, I., O. Ilbert, C. Laigle, J. Coupon, et al. In: *Astronomy and Astrophysics* 605, A70 (Sept. 2017), A70.

Davidzon, I., O. Ilbert, A. L. Faisst, M. Sparre, and P. L. Capak. In: *Astrophysical Journal* 852.2, 107 (Jan. 2018), p. 107.

Davies, F. B., S. R. Furlanetto, and M. McQuinn. In: *Monthly Notices of the Royal Astronomical Society* 457 (Apr. 2016), pp. 3006–3023.

Davis, M., G. Efstathiou, C. S. Frenk, and S. D. M. White. In: *The Astrophysical Journal* 292 (May 1985), pp. 371–394.

de los Reyes, M. A. C. and J. Kennicutt Robert C. In: *Astrophysical Journal* 872.1, 16 (Feb. 2019), p. 16.

De Lucia, G. and J. Blaizot. In: *Monthly Notices of the Royal Astronomical Society* 375 (Feb. 2007), pp. 2–14.

De Santis, C., A. Grazian, A. Fontana, and P. Santini. In: *New Astronomy* 12.4 (Feb. 2007), pp. 271–288.

Delvecchio, I., E. Daddi, J. Aird, J. R. Mullaney, et al. In: *The Astrophysical Journal* 892 (Mar. 2020), p. 17.

Dey, A., D. J. Schlegel, D. Lang, R. Blum, et al. In: *Astronomical Journal* 157.5, 168 (May 2019), p. 168.

Dey, A., D. J. Schlegel, D. Lang, R. Blum, et al. In: *Astronomical Journal* 157.5, 168 (May 2019), p. 168.

Di Matteo, T., V. Springel, and L. Hernquist. In: *Nature* 433 (Feb. 2005), pp. 604–607.

Diehl, H. T. and Dark Energy Survey Collaboration. In: *American Astronomical Society Meeting Abstracts #219*. Vol. 219. American Astronomical Society Meeting Abstracts. Jan. 2012, 413.05, p. 413.05.

Diemer, B. In: *Astrophysical Journal, Supplement* 239.2, 35 (Dec. 2018), p. 35.

Ding, X., S. Birrer, T. Treu, and J. D. Silverman. In: *arXiv e-prints*, arXiv:2111.08721 (Nov. 2021), arXiv:2111.08721.

Djorgovski, S. and M. Davis. In: *The Astrophysical Journal* 313 (Feb. 1987), p. 59.

Donley, J. L., A. M. Koekemoer, M. Brusa, P. Capak, et al. In: *Astrophysical Journal* 748.2, 142 (Apr. 2012), p. 142.

Donnari, M., A. Pillepich, G. D. Joshi, D. Nelson, et al. In: *Monthly Notices of the Royal Astronomical Society* 500 (Jan. 2021), pp. 4004–4024.

Donnari, M., A. Pillepich, D. Nelson, F. Marinacci, et al. In: *Monthly Notices of the Royal Astronomical Society* 506 (Oct. 2021), pp. 4760–4780.

Dressler, A. In: *Annual Review of Astron and Astrophys* 22 (Jan. 1984), pp. 185–222.

Dressler, A. and J. E. Gunn. In: *The Astrophysical Journal* 270 (July 1983), pp. 7–19.

Dressler, A., D. Lynden-Bell, D. Burstein, R. L. Davies, et al. In: *The Astrophysical Journal* 313 (Feb. 1987), p. 42.

Driver, S. P. and A. S. G. Robotham. In: *Monthly Notices of the Royal Astronomi-*

cal Society 407.4 (Sept. 2010), pp. 2131–2140.

Drory, N., K. Bundy, A. Leauthaud, N. Scoville, et al. In: *Astrophysical Journal* 707.2 (Dec. 2009), pp. 1595–1609.

Dubois, Y., C. Pichon, C. Welker, D. Le Borgne, et al. In: *Monthly Notices of the RAS* 444.2 (Oct. 2014), pp. 1453–1468.

Dubois, Y., R. Gavazzi, S. Peirani, and J. Silk. In: *Monthly Notices of the Royal Astronomical Society* 433.4 (Aug. 2013), pp. 3297–3313.

Duc, P.-A., J.-C. Cuillandre, E. Karabal, M. Cappellari, et al. In: *Monthly Notices of the Royal Astronomical Society* 446 (Jan. 2015), pp. 120–143.

Dunlop, J. S., R. G. Abraham, M. L. N. Ashby, M. Bagley, et al. In: *JWST Proposal. Cycle 1* (Mar. 2021), p. 1837.

Dutton, A. A., F. C. van den Bosch, and A. Dekel. In: *Monthly Notices of the RAS* 405.3 (July 2010), pp. 1690–1710.

Ebeling, H. In: *Monthly Notices of the RAS* 340.4 (Apr. 2003), pp. 1269–1278.

Eddington, A. S. In: *Monthly Notices of the RAS* 73 (Mar. 1913), pp. 359–360.

Edge, A., W. Sutherland, K. Kuijken, S. Driver, et al. In: *The Messenger* 154 (Dec. 2013), pp. 32–34.

Efstathiou, G. and J. R. Bond. In: *Monthly Notices of the RAS* 304.1 (Mar. 1999), pp. 75–97.

Efstathiou, G., R. S. Ellis, and B. A. Peterson. In: *Monthly Notices of the RAS* 232 (May 1988), pp. 431–461.

Eisenstein, D. J., P. Ferruit, and M. J. Rieke. JWST Proposal. Cycle 1. June 2017.

Elbaz, D., E. Daddi, D. Le Borgne, M. Dickinson, et al. In: *Astronomy and Astrophysics* 468.1 (June 2007), pp. 33–48.

Ellis, R. S., R. J. McLure, J. S. Dunlop, B. E. Robertson, et al. In: *Astrophysical Journal, Letters* 763.1, L7 (Jan. 2013), p. L7.

Ellison, S. L., B. Catinella, and L. Cortese. In: *Monthly Notices of the Royal Astronomical Society* 478 (Aug. 2018), pp. 3447–3466.

Emsellem, E., M. Cappellari, D. Krajnović, K. Alatalo, et al. In: *Monthly Notices of the RAS* 414.2 (June 2011), pp. 888–912.

Emsellem, E., M. Cappellari, D. Krajnović, G. van de Ven, et al. In: *Monthly Notices of the RAS* 379.2 (Aug. 2007), pp. 401–417.

Emsellem, E., F. Renaud, F. Bournaud, B. Elmegreen, et al. In: *Monthly Notices of the Royal Astronomical Society* 446 (Jan. 2015), pp. 2468–2482.

Euclid Collaboration, S. E. van Mierlo, K. I. Caputi, M. Ashby, et al. May 2022.

Faber, S. M. and R. E. Jackson. In: *The Astrophysical Journal* 204 (Mar. 1976), pp. 668–683.

Faisst, A. L., R. R. Chary, S. Fajardo-Acosta, R. Paladini, et al. In: *arXiv e-prints*, arXiv:2103.09836 (Mar. 2021), arXiv:2103.09836.

Fan, X., C. L. Carilli, and B. Keating. In: *Annual Review of Astronomy and Astrophysics* 44.1 (Sept. 2006), pp. 415–462.

Fan, X., M. A. Strauss, R. H. Becker, R. L. White, et al. In: *The Astronomical Journal* 132 (July 2006), pp. 117–136.

Farage, C. L., P. J. McGregor, M. A. Dopita, and G. V. Bicknell. In: *The Astrophysical Journal* 724 (Nov. 2010), pp. 267–284.

Feldmann, R. and L. Mayer. In: *Monthly Notices of the Royal Astronomical Society* 446.2 (Jan. 2015), pp. 1939–1956.

Fernández-Soto, A., K. M. Lanzetta, and A. Yahil. In: *Astrophysical Journal* 513.1 (Mar. 1999), pp. 34–50.

Ferrarese, L. and D. Merritt. In: *The Astrophysical Journal* 539 (Aug. 2000), pp. L9–L12.

Ferrari, F., R. R. de Carvalho, and M. Trevisan. In: *Astrophysical Journal* 814.1, 55 (Nov. 2015), p. 55.

Ferruit, P. JWST Proposal. Cycle 1. July 2017.

Finkelstein, S. L., C. Papovich, M. Dickinson, M. Song, et al. In: *Nature* 502 (Oct. 2013), pp. 524–527.

Finkelstein, S. L., J. R. E. Ryan, C. Papovich, M. Dickinson, et al. In: *ApJ* 810 (Sept. 2015), p. 71.

Finkelstein, S. L., M. Dickinson, H. C. Ferguson, A. Grazian, et al. JWST Proposal ID 1345. Cycle 0 Early Release Science. Nov. 2017.

Finkelstein, S. L., C. Papovich, N. Pirzkal, M. Bagley, et al. JWST Proposal. Cycle 1. Mar. 2021.

Finkelstein, S. L., J. Ryan Russell E., C. Papovich, M. Dickinson, et al. In: *Astrophysical Journal* 810.1, 71 (Sept. 2015), p. 71.

Fisher, R. A. In: *Philosophical Transactions of the Royal Society of London Series A* 222 (Jan. 1922), pp. 309–368.

Fitzpatrick, E. L. and D. Massa. In: *The Astrophysical Journal* 307 (Aug. 1986), pp. 286–294.

Fitzpatrick, E. L. and D. Massa. In: *The Astrophysical Journal* 307 (Aug. 1986), pp. 286–294.

Fitzpatrick, E. L. In: *The Astronomical Journal* 92 (Nov. 1986), pp. 1068–1073.

Flaugher, B. In: *International Journal of Modern Physics A* 20.14 (Jan. 2005), pp. 3121–3123.

Flaugher, B. L., T. M. C. Abbott, R. Angstadt, J. Annis, et al. In: *Ground-Based and Airborne Instrumentation for Astronomy IV*. Ed. by I. S. McLean, S. K. Ramsay, and H. Takami. Vol. 8446. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Sept. 2012, 844611, p. 844611.

Fontana, A., J. S. Dunlop, D. Paris, T. A. Targett, et al. In: *Astronomy and Astrophysics* 570, A11 (Oct. 2014), A11.

Fontana, A., J. S. Dunlop, D. Paris, T. A. Targett, et al. In: *Astronomy and Astrophysics* 570, A11 (Oct. 2014), A11.

Fontanot, F., G. De Lucia, P. Monaco, R. S. Somerville, and P. Santini. In: *Monthly Notices of the RAS* 397.4 (Aug. 2009), pp. 1776–1790.

Forbes, J. C., M. R. Krumholz, A. Burkert, and A. Dekel. In: *Monthly Notices of the RAS* 443.1 (Sept. 2014), pp. 168–185.

Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman. In: *PASP* 125 (2013), pp. 306–312.

Forrest, B., M. Annunziatella, G. Wilson, D. Marchesini, et al. In: *The Astrophysical Journal* 890 (Feb. 2020), p. L1.

Förster Schreiber, N. M., R. Genzel, N. Bouché, G. Cresci, et al. In: *Astrophysical Journal* 706.2 (Dec. 2009), pp. 1364–1428.

Förster Schreiber, N. M., P. G. van Dokkum, M. Franx, I. Labbé, et al. In: *The Astrophysical Journal* 616 (Nov. 2004), pp. 40–62.

Förster Schreiber, N. M. and S. Wuyts. In: *Annual Review of Astron and Astrophys* 58 (Aug. 2020), pp. 661–725.

Franx, M., A. Moorwood, H. .-. Rix, K. Kuijken, et al. In: *The Messenger* 99 (Mar. 2000), pp. 20–22.

Franx, M., I. Labbé, G. Rudnick, P. G. van Dokkum, et al. In: *The Astrophysical Journal* 587 (Apr. 2003), pp. L79–L82.

French, K. D. In: *Publications of the Astronomical Society of the Pacific* 133 (July 2021), p. 072001.

French, K. D., A. Smercina, K. Rowlands, A. Tripathi, et al. Apr. 2022.

French, K. D., Y. Yang, A. Zabludoff, D. Narayanan, et al. In: *The Astrophysical Journal* 801 (Mar. 2015), p. 1.

Fudamoto, Y., P. A. Oesch, A. Faisst, M. Béthermin, et al. In: *Astronomy and Astrophysics* 643, A4 (Nov. 2020), A4.

Fudamoto, Y., P. A. Oesch, S. Schouws, M. Stefanon, et al. In: *Nature* 597.7877 (Sept. 2021), pp. 489–492.

Furlanetto, S. R., M. Zaldarriaga, and L. Hernquist. In: *The Astrophysical Journal* 613 (Sept. 2004), pp. 1–15.

Furlong, M., R. G. Bower, T. Theuns, J. Schaye, et al. In: *Monthly Notices of the RAS* 450.4 (July 2015), pp. 4486–4504.

Furusawa, H., G. Kosugi, M. Akiyama, T. Takata, et al. In: *ASTROPHYS J SUPPL S* 176.1 (May 2008), pp. 1–18.

Gabor, J. M. and R. Davé. In: *Monthly Notices of the RAS* 447.1 (Feb. 2015), pp. 374–391.

Gadotti, D. A., P. Sánchez-Blázquez, J. Falcón-Barroso, B. Husemann, et al. In:

Monthly Notices of the Royal Astronomical Society 482 (Jan. 2019), pp. 506–529.

Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, et al. In: *Astronomy & Astrophysics* 595 (Nov. 2016), A2.

Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, et al. In: *Astronomy & Astrophysics* 616 (Aug. 2018), A1.

Gebhardt, K., R. Bender, G. Bower, A. Dressler, et al. In: *The Astrophysical Journal* 539 (Aug. 2000), pp. L13–L16.

Gehrels, N. In: *Astrophysical Journal* 303 (Apr. 1986), p. 336.

Genzel, R., D. Lutz, E. Sturm, E. Egami, et al. In: *The Astrophysical Journal* 498 (May 1998), pp. 579–605.

Georgakakis, A., J. Aird, J. Buchner, M. Salvato, et al. In: *Monthly Notices of the Royal Astronomical Society* 453 (Oct. 2015), pp. 1946–1964.

Giavalisco, M., H. C. Ferguson, A. M. Koekemoer, M. Dickinson, et al. In: *Astrophysical Journal, Letters* 600.2 (Jan. 2004), pp. L93–L98.

Girelli, G., L. Pozzetti, M. Bolzonella, C. Giocoli, et al. In: *Astronomy and Astrophysics* 634, A135 (Feb. 2020), A135.

Gladders, M. D., A. Oemler, A. Dressler, B. Poggianti, et al. In: *Astrophysical Journal* 770.1, 64 (June 2013), p. 64.

Glazebrook, K., J. A. Peacock, C. A. Collins, and L. Miller. In: *Monthly Notices of the Royal Astronomical Society* 266 (Jan. 1994), pp. 65–91.

Glazebrook, K., T. Nanayakkara, J. Esdaile, J. M. Espejo, et al. JWST Proposal. Cycle 1. Mar. 2021.

Glazebrook, K., J. A. Peacock, L. Miller, and C. A. Collins. In: *Monthly Notices of the Royal Astronomical Society* 275 (July 1995), pp. 169–184.

Glazebrook, K., C. Schreiber, I. Labbé, T. Nanayakkara, et al. In: *Nature* 544.7648 (Apr. 2017), pp. 71–74.

Gnedin, O. Y. In: *The Astrophysical Journal* 582 (Jan. 2003), pp. 141–161.

Gnedin, O. Y., A. V. Kravtsov, A. A. Klypin, and D. Nagai. In: *Astrophysical Journal* 616.1 (Nov. 2004), pp. 16–26.

González, V., I. Labbé, R. J. Bouwens, G. Illingworth, et al. In: *Astrophysical Journal, Letters* 735.2, L34 (July 2011), p. L34.

Goodman, J. and J. Weare. In: *Communications in Applied Mathematics and Computational Science* 5.1 (Jan. 2010), pp. 65–80.

Gordon, K. D., G. C. Clayton, K. A. Misselt, A. U. Landolt, and M. J. Wolff. In: *The Astrophysical Journal* 594 (Sept. 2003), pp. 279–293.

Goto, T. In: *Monthly Notices of the Royal Astronomical Society* 357.3 (2005), pp. 937–944.

Gott III, J. R. In: *The Astrophysical Journal* 201 (Oct. 1975), pp. 296–310.

Gottesman, S. T., R. D. Davies, and V. C. Reddish. In: *Monthly Notices of the RAS* 133 (Jan. 1966), p. 359.

Grazian, A., A. Fontana, P. Santini, J. S. Dunlop, et al. In: *Astronomy and Astrophysics* 575 (Mar. 2015), A96.

Greene, O. A., M. R. Anderson, M. Marinelli, K. Holley-Bockelmann, et al. In: *The Astrophysical Journal* 910 (Apr. 2021), p. 162.

Grogin, N. A., D. D. Kocevski, S. M. Faber, H. C. Ferguson, et al. In: *Astrophysical Journal, Supplement* 197.2, 35 (Dec. 2011), p. 35.

Gruen, D., S. Seitz, and G. M. Bernstein. In: *Publications of the ASP* 126.936 (Feb. 2014), p. 158.

Gruppioni, C., F. Pozzi, G. Rodighiero, I. Delvecchio, et al. In: *Monthly Notices of the RAS* 432.1 (June 2013), pp. 23–52.

Gültekin, K., D. O. Richstone, K. Gebhardt, T. R. Lauer, et al. In: *The Astrophysical Journal* 698 (June 2009), pp. 198–221.

Gunn, J. E. and I. Gott J. Richard. In: *Astrophysical Journal* 176 (Aug. 1972), p. 1.

Guth, A. H. In: *Physical Review D* 23 (Jan. 1981), pp. 347–356.

Gwyn, S. D. J. In: *Publications of the ASP* 120.864 (Feb. 2008), p. 212.

Habouzit, M., R. S. Somerville, Y. Li, S. Genel, et al. In: *Monthly Notices of the RAS* 509.2 (Jan. 2022), pp. 3015–3042.

Hanany, S., P. Ade, A. Balbi, J. Bock, et al. In: *The Astrophysical Journal* 545 (Dec. 2000), pp. L5–L9.

Harikane, Y., A. K. Inoue, K. Mawatari, T. Hashimoto, et al. In: *The Astrophysical Journal* 929 (Apr. 2022), p. 1.

Harikane, Y., M. Ouchi, Y. Ono, S. More, et al. In: *Astrophysical Journal* 821.2, 123 (Apr. 2016), p. 123.

Hasinger, G., P. Capak, M. Salvato, A. J. Barger, et al. In: *Astrophysical Journal* 858.2, 77 (May 2018), p. 77.

Hausman, M. A. and J. P. Ostriker. In: *The Astrophysical Journal* 224 (Sept. 1978), pp. 320–336.

Häußler, B., M. Vika, S. P. Bamford, E. J. Johnston, et al. In: *arXiv e-prints*, arXiv:2204.05907 (Apr. 2022), arXiv:2204.05907.

Heintz, K. E., J. P. U. Fynbo, P. Møller, B. Milvang-Jensen, et al. In: *Astronomy and Astrophysics* 595, A13 (Oct. 2016), A13.

Henriques, B. M. B., S. D. M. White, P. A. Thomas, R. Angulo, et al. In: *Monthly Notices of the RAS* 451.3 (Aug. 2015), pp. 2663–2680.

Hernquist, L. and J. E. Barnes. In: *Nature* 354 (Nov. 1991), pp. 210–212.

Herschel, W. In: *Philosophical Transactions of the Royal Society of London Series I* 76 (Jan. 1786), pp. 457–499.

Herschel, W. In: *Philosophical Transactions of the Royal Society of London Series I* 79 (Jan. 1789), pp. 212–255.

Hewlett, T., C. Villforth, V. Wild, J. Mendez-Abreu, et al. In: *Monthly Notices of the Royal Astronomical Society* 470.1 (2017/May/10/), pp. 755–770.

Hibbard, J. E. and J. H. van Gorkom. In: *The Astronomical Journal* 111 (Feb. 1996), p. 655.

Hildebrandt, H., T. Erben, K. Kuijken, L. van Waerbeke, et al. In: *Monthly Notices of the RAS* 421.3 (Apr. 2012), pp. 2355–2367.

Hill, A. R., A. Muzzin, M. Franx, B. Clauwens, et al. In: *Astrophysical Journal* 837.2, 147 (Mar. 2017), p. 147.

Hoaglin, D. C., F. Mosteller, and J. W. Tukey. In: *Wiley Series in Probability and Mathematical Statistics, New York: Wiley, 1983, Edited by Hoaglin, David C.; Mosteller, Frederick; Tukey, John W.* Wiley Series in Probability and Mathematical Statistics, John Wiley, 1983.

Hogg, D. W. In: *arXiv e-prints*, astro-ph/9905116 (May 1999), astro–ph/9905116.

Hogg, D. W., M. R. Blanton, J. Brinchmann, D. J. Eisenstein, et al. In: *The Astrophysical Journal* 601 (Jan. 2004), pp. L29–L32.

Hogg, D. W., M. R. Blanton, D. J. Eisenstein, J. E. Gunn, et al. In: *The Astrophysical Journal* 585 (Mar. 2003), pp. L5–L9.

Hogg, D. W., J. Bovy, and D. Lang. In: *arXiv e-prints*, arXiv:1008.4686 (Aug. 2010), arXiv:1008.4686.

Hopkins, P. F., L. Hernquist, T. J. Cox, and D. Kereš. In: *The Astrophysical Journal Supplement Series* 175 (Apr. 2008), pp. 356–389.

Hoyle, F. In: *Monthly Notices of the Royal Astronomical Society* 108 (Jan. 1948), p. 372.

Hsieh, B.-C., W.-H. Wang, C.-C. Hsieh, L. Lin, et al. In: *The Astrophysical Journal Supplement Series* 203.2 (Nov. 2012), p. 23.

Hubble, E. P. In: *Astrophysical Journal* 64 (Dec. 1926), pp. 321–369.

Hubble, E. P. In: *The Observatory* 50 (Sept. 1927), pp. 276–281.

Hubble, E. In: *Proceedings of the National Academy of Science* 15.3 (Mar. 1929), pp. 168–173.

Hubble, E. In: *Astrophysical Journal* 79 (Jan. 1934), p. 8.

Hubble, E. In: *The Astrophysical Journal* 84 (Sept. 1936), p. 158.

Hubble, E. In: *The Astrophysical Journal* 84 (Oct. 1936), p. 270.

Hubble, E. and M. L. Humason. In: *Astrophysical Journal* 74 (July 1931), p. 43.

Hung, C.-L., C. M. Casey, Y.-K. Chiang, P. L. Capak, et al. In: *Astrophysical Journal* 826.2, 130 (Aug. 2016), p. 130.

Hviding, R. E., K. N. Hainline, M. Rieke, S. Juneau, et al. In: *arXiv e-*

prints, arXiv:2203.11217 (Mar. 2022), arXiv:2203.11217.

Ilbert, O., S. Arnouts, E. Le Floc'h, H. Aussel, et al. In: *åp* 579 (July 2015), A2.

Ilbert, O., S. Arnouts, E. Le Floc'h, H. Aussel, et al. In: *Astronomy and Astrophysics* 579, A2 (July 2015), A2.

Ilbert, O., S. Arnouts, H. J. McCracken, M. Bolzonella, et al. In: *antike und abendland* 457 (Oct. 2006), pp. 841–856.

Ilbert, O., P. Capak, M. Salvato, H. Aussel, et al. In: *Astrophysical Journal* 690.2 (Jan. 2009), pp. 1236–1249.

Ilbert, O., H. J. McCracken, O. Le Fèvre, P. Capak, et al. In: *Astronomy and Astrophysics* 556 (Aug. 2013), A55.

Ilbert, O., M. Salvato, E. Le Floc'h, H. Aussel, et al. In: *The Astrophysical Journal* 709 (Feb. 2010), pp. 644–663.

Ilbert, O., L. Tresse, S. Arnouts, E. Zucca, et al. In: *Monthly Notices of the RAS* 351.2 (June 2004), pp. 541–551.

Ilbert, O., L. Tresse, E. Zucca, S. Bardelli, et al. In: *Astronomy and Astrophysics* 439.3 (Sept. 2005), pp. 863–876.

Illingworth, G. D., D. Magee, P. A. Oesch, R. J. Bouwens, et al. In: *Astrophysical Journal, Supplement* 209.1, 6 (Nov. 2013), p. 6.

Iodice, E., M. Spavone, M. Capaccioli, R. F. Peletier, et al. In: *The Astrophysical Journal* 839 (Apr. 2017), p. 21.

Ishigaki, M., R. Kawamata, M. Ouchi, M. Oguri, et al. In: *Astrophysical Journal* 854.1, 73 (Feb. 2018), p. 73.

Ito, K., M. Tanaka, T. Miyaji, O. Ilbert, et al. In: *arXiv e-prints*, arXiv:2203.04322 (Mar. 2022), arXiv:2203.04322.

Ivezić, Ž., S. M. Kahn, J. A. Tyson, B. Abel, et al. In: *Astrophysical Journal* 873.2, 111 (Mar. 2019), p. 111.

Iyer, K., E. Gawiser, R. Davé, P. Davis, et al. In: *Astrophysical Journal* 866.2, 120 (Oct. 2018), p. 120.

James, F. and M. Roos. In: *Computer Physics Communications* 10.6 (Dec. 1975), pp. 343–367.

Jarvis, M. J., D. G. Bonfield, V. A. Bruce, J. E. Geach, et al. In: *Monthly Notices of the Royal Astronomical Society* 428 (Jan. 2013), pp. 1281–1295.

Jeans, J. H. In: *Philosophical Transactions of the Royal Society of London Series A* 199 (Jan. 1902), pp. 1–53.

Jenkins, A., C. S. Frenk, S. D. M. White, J. M. Colberg, et al. In: *Monthly Notices of the Royal Astronomical Society* 321 (Feb. 2001), pp. 372–384.

Jenkins, C. R. and I. N. Reid. In: *The Astronomical Journal* 101 (May 1991), p. 1595.

Jin, S., E. Daddi, G. E. Magdis, D. Liu, et al. In: *Astrophysical Journal* 887.2, 144 (Dec. 2019), p. 144.

Jin, S., E. Daddi, D. Liu, V. Smolčić, et al. In: *The Astrophysical Journal* 864 (Sept. 2018), p. 56.

Johnston, R. In: *Astronomy and Astrophysics Reviews* 19, 41 (Aug. 2011), p. 41.

Kaiser, N., J. L. Tonry, and G. A. Luppino. In: *Publications of the ASP* 112.772 (June 2000), pp. 768–800.

Kaiser, N., W. Burgett, K. Chambers, L. Denneau, et al. In: *Ground-Based and Airborne Telescopes III*. Ed. by L. M. Stepp, R. Gilmozzi, and H. J. Hall. Vol. 7733. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July 2010, 77330E, 77330E.

Kant, I. Jan. 1755.

Karim, A., E. Schinnerer, A. Martínez-Sansigre, M. T. Sargent, et al. In: *Astrophysical Journal* 730.2, 61 (Apr. 2011), p. 61.

Karlsson, T., V. Bromm, and J. Bland-Hawthorn. In: *Reviews of Modern Physics* 85 (Apr. 2013), pp. 809–848.

Kartaltepe, J., C. M. Casey, M. Bagley, A. Bongiorno, et al. JWST Proposal. Cycle 1. Mar. 2021.

Kashikawa, N., K. Shimasaku, M. A. Malkan, M. Doi, et al. In: *The Astrophysical Journal* 648 (Sept. 2006), pp. 7–22.

Kashino, D., J. D. Silverman, D. Sanders, J. Kartaltepe, et al. In: *Astrophysical Journal, Supplement* 241.1, 10 (Mar. 2019), p. 10.

Kassin, S. A., B. J. Weiner, S. M. Faber, J. P. Gardner, et al. In: *The Astrophysical Journal* 758 (Oct. 2012), p. 106.

Kauffmann, G., S. D. M. White, and B. Guiderdoni. In: *Monthly Notices of the Royal Astronomical Society* 264 (Sept. 1993), pp. 201–218.

Kauffmann, G. and M. Haehnelt. In: *Monthly Notices of the Royal Astronomical Society* 311 (Jan. 2000), pp. 576–588.

Kaviraj, S., C. Laigle, T. Kimm, J. E. G. Devriendt, et al. In: *Monthly Notices of the RAS* 467.4 (June 2017), pp. 4739–4752.

Kennicutt Robert C., J. and M. A. C. De Los Reyes. In: *Astrophysical Journal* 908.1, 61 (Feb. 2021), p. 61.

Kennicutt Jr., R. C. In: *The Astrophysical Journal* 498 (May 1998), pp. 541–552.

Khochfar, S. and J. Silk. In: *Monthly Notices of the Royal Astronomical Society* 397.1 (July 2009), pp. 506–510.

Khochfar, S., E. Emsellem, P. Serra, M. Bois, et al. In: *Monthly Notices of the RAS* 417.2 (Oct. 2011), pp. 845–862.

Kiang, T. In: *Monthly Notices of the Royal Astronomical Society* 122 (Jan. 1961), p. 263.

Klypin, A., A. V. Kravtsov, O. Valenzuela, and F. Prada. In: *The Astrophysical Journal* 522 (Sept. 1999), pp. 82–92.

Kochanek, C. S., M. A. Pahre, E. E. Falco, J. P. Huchra, et al. In: *The Astrophysical Journal* 560 (Oct. 2001), pp. 566–579.

Koekemoer, A. M., H. Aussel, D. Calzetti, P. Capak, et al. In: *ApJS* 172 (Sept. 2007), p. 196.

Koekemoer, A. M., H. Aussel, D. Calzetti, P. Capak, et al. In: *ApJS* 172 (Sept. 2007), p. 196.

Koekemoer, A. M., S. M. Faber, H. C. Ferguson, N. A. Grogin, et al. In: *Astrophysical Journal, Supplement* 197.2, 36 (Dec. 2011), p. 36.

Kohonen, T. In: *Biol. Cybern.* 43 (1982), pp. 59–69.

Kokorev, V. I., G. E. Magdis, I. Davidzon, G. Brammer, et al. In: *Astrophysical Journal* 921.1, 40 (Nov. 2021), p. 40.

Kriek, M., A. E. Shapley, N. A. Reddy, B. Siana, et al. In: *The Astrophysical Journal Supplement Series* 218 (June 2015), p. 15.

Kriek, M., A. van der Wel, P. G. van Dokkum, M. Franx, and G. D. Illingworth. In: *The Astrophysical Journal* 682 (Aug. 2008), pp. 896–906.

Kriek, M., P. G. van Dokkum, M. Franx, N. M. Förster Schreiber, et al. In: *The Astrophysical Journal* 645 (July 2006), pp. 44–54.

Kriek, M., P. G. van Dokkum, M. Franx, R. Quadri, et al. In: *The Astrophysical Journal* 649 (Oct. 2006), pp. L71–L74.

Kriek, M., P. G. van Dokkum, I. Labbé, M. Franx, et al. Mar. 2018.

Kron, R. C. In: *Physica Scripta* 21 (June 1980), pp. 652–661.

Kron, R. G. In: *ApJS* 43 (June 1980), pp. 305–325.

Kümmel, M., E. Bertin, M. Schefer, N. Apostolakos, et al. In: *Astronomical Data Analysis Software and Systems XXIX*. Ed. by R. Pizzo, E. R. Deul, J. D. Mol, J. de Plaa, and H. Verkouter. Vol. 527. Astronomical Society of the Pacific Conference Series. Jan. 2020, p. 29.

Labbe, I., R. Bezanson, H. Atek, G. Brammer, et al. JWST Proposal. Cycle 1. Mar. 2021.

Labbé, I., P. A. Oesch, R. J. Bouwens, G. D. Illingworth, et al. In: *The Astrophysical Journal* 777 (Nov. 2013), p. L19.

Labbé, I., P. A. Oesch, G. D. Illingworth, P. G. van Dokkum, et al. In: *Astrophysical Journal, Supplement* 221.2, 23 (Dec. 2015), p. 23.

Labbé, I., R. Bouwens, G. D. Illingworth, and M. Franx. In: *Astrophysical Journal, Letters* 649.2 (Oct. 2006), pp. L67–L70.

Labbé, I., M. Franx, G. Rudnick, N. M. Förster Schreiber, et al. In: *Astronomical Journal* 125.3 (Mar. 2003), pp. 1107–1123.

Labbé, I., J. Huang, M. Franx, G. Rudnick, et al. In: *The Astrophysical Journal* 624 (May 2005), pp. L81–L84.

Lagos, C. d. P., R. J. Tobar, A. S. G. Robotham, D. Obreschkow, et al. In:

Monthly Notices of the RAS 481.3 (Dec. 2018), pp. 3573–3603.

Laidler, V. G., C. Papovich, N. A. Grogin, R. Idzi, et al. In: *Publications of the ASP* 119.861 (Nov. 2007), pp. 1325–1344.

Laigle, C., I. Davidzon, O. Ilbert, J. Devriendt, et al. In: *Monthly Notices of the RAS* 486.4 (July 2019), pp. 5104–5123.

Laigle, C., H. J. McCracken, O. Ilbert, B. C. Hsieh, et al. In: *ApJS* 224 (June 2016), p. 24.

Laigle, C., C. Pichon, S. Arnouts, H. J. McCracken, et al. In: *Monthly Notices of the RAS* 474.4 (Mar. 2018), pp. 5437–5458.

Lambas, D. G., S. Alonso, V. Mesa, and A. L. O'Mill. In: *Astronomy and Astrophysics* 539 (Mar. 2012), A45.

Lambas, D. G., P. B. Tissera, M. S. Alonso, and G. Coldwell. In: *Monthly Notices of the Royal Astronomical Society* 346 (Dec. 2003), pp. 1189–1196.

Lang, D., D. W. Hogg, and D. Mykytyn. In: *Astrophysics Source Code Library* (Apr. 2016), ascl:1604.008.

Lang, D., D. W. Hogg, and D. J. Schlegel. In: *Astronomical Journal* 151.2, 36 (Feb. 2016), p. 36.

Laplace, P. S. Jan. 1796.

Laporte, N., R. S. Ellis, F. Boone, F. E. Bauer, et al. In: *ApJL* 837.2 (Mar. 2017), p. L21.

Larson, R. B. and B. M. Tinsley. In: *The Astrophysical Journal* 219 (Jan. 1978), pp. 46–59.

Larson, R. B., B. M. Tinsley, and C. N. Caldwell. In: *Astrophysical Journal* 237 (May 1980), pp. 692–707.

Larson, R. L., S. L. Finkelstein, T. A. Hutchison, C. Papovich, et al. In: *The Astrophysical Journal* 930 (May 2022), p. 104.

Larson, R. B. In: *Monthly Notices of the Royal Astronomical Society* 169 (Nov. 1974), pp. 229–246.

Laureijs, R., J. Amiaux, S. Arduini, J. .-. Augueres, et al. In: *arXiv e-prints*, arXiv:1110.3193 (Oct. 2011), arXiv:1110.3193.

Le Fèvre, O., R. Abraham, S. J. Lilly, R. S. Ellis, et al. In: *Monthly Notices of the Royal Astronomical Society* 311 (Jan. 2000), pp. 565–575.

Le Fèvre, O., M. Béthermin, A. Faisst, G. Jones, et al. In: *arXiv e-prints*, arXiv:1910.09517 (Oct. 2019), arXiv:1910.09517.

Le Fèvre, O., P. Cassata, O. Cucciati, B. Garilli, et al. In: *Astronomy and Astrophysics* 559, A14 (Nov. 2013), A14.

Le Fèvre, O., L. A. M. Tasca, P. Cassata, B. Garilli, et al. In: *Astronomy and Astrophysics* 576, A79 (Apr. 2015), A79.

Le Fèvre, O., G. Vettolani, B. Garilli, L. Tresse, et al. In: *Astronomy and Astrophysics* 439.3 (Sept. 2005), pp. 845–862.

Le Fèvre, O., M. Saisse, D. Mancini, S. Brau-Nogue, et al. In: ed. by M. Iye and A. F. M. Moorwood. Vol. 4841. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Mar. 2003, pp. 1670–1681.

Leaman, R., J. T. Mendel, E. Wisnioski, A. M. Brooks, et al. In: *Monthly Notices of the Royal Astronomical Society* 472 (Dec. 2017), pp. 1879–1896.

Leauthaud, A., R. Massey, J.-P. Kneib, J. Rhodes, et al. In: *ApJS* 172.1 (Sept. 2007), p. 219.

Leavitt, H. S. and E. C. Pickering. In: *Harvard College Observatory Circular* 173 (Mar. 1912), pp. 1–3.

Lee, B., M. Giavalisco, C. C. Williams, Y. Guo, et al. In: *The Astrophysical Journal* 774 (Sept. 2013), p. 47.

Lee, J. and E. Komatsu. In: *Astrophysical Journal* 718.1 (July 2010), pp. 60–65.

Lee, N., D. B. Sanders, C. M. Casey, S. Toft, et al. In: *Astrophysical Journal* 801.2, 80 (Mar. 2015), p. 80.

Lee-Waddell, K., P. Serra, B. Koribalski, A. Venhola, et al. In: *Monthly Notices of the Royal Astronomical Society* 474 (Feb. 2018), pp. 1108–1115.

Leethochawalit, N., G. Roberts-Borsani, T. Morishita, M. Trenti, and T. Treu. May 2022.

Legrand, L., H. J. McCracken, I. Davidzon, O. Ilbert, et al. In: *Monthly Notices of the RAS* 486.4 (July 2019), pp. 5468–5481.

Leitherer, C., D. Schaerer, J. D. Goldader, R. M. G. Delgado, et al. In: *The Astrophysical Journal Supplement Series* 123 (July 1999), pp. 3–40.

Leja, J., B. D. Johnson, C. Conroy, P. van Dokkum, et al. In: *Astrophysical Journal* 877.2, 140 (June 2019), p. 140.

Leja, J., B. D. Johnson, C. Conroy, P. G. van Dokkum, and N. Byler. In: *Astrophysical Journal* 837.2, 170 (Mar. 2017), p. 170.

Leja, J., S. Tacchella, and C. Conroy. In: *Astrophysical Journal, Letters* 880.1, L9 (July 2019), p. L9.

Leonova, E., P. A. Oesch, Y. Qin, R. P. Naidu, et al. Dec. 2021.

Leroy, A. K., F. Walter, E. Brinks, F. Bigiel, et al. In: *Astronomical Journal* 136.6 (Dec. 2008), pp. 2782–2845.

Leslie, S., E. Schinnerer, D. Liu, B. Magnelli, et al. In: *arXiv e-prints*, arXiv:2006.13937 (June 2020), arXiv:2006.13937.

Leslie, S. K., E. Schinnerer, D. Liu, B. Magnelli, et al. In: *Astrophysical Journal* 899.1, 58 (Aug. 2020), p. 58.

Leśniewska, A. and M. J. Michałowski. In: *Astronomy and Astrophysics* 624 (Apr. 2019), p. L13.

Lesser, M. In: *Publications of the Astronomical Society of the Pacific* 127.957 (2015), pp. 1097–1104.

Lewis, I., M. Balogh, R. De Propris, W. Couch, et al. In: *Monthly Notices of the Royal Astronomical Society* 334 (Aug. 2002), pp. 673–683.

Lilly, S. J., O. Le Fèvre, A. Renzini, G. Zamorani, et al. In: *\apjs* 172 (Sept. 2007), pp. 70–85.

Lilly, S. J., O. Le Fèvre, A. Renzini, G. Zamorani, et al. In: *Astrophysical Journal, Supplement* 172 (Sept. 2007), pp. 70–85.

Lilly, S. J., C. M. Carollo, A. Pipino, A. Renzini, and Y. Peng. In: *Astrophysical Journal* 772.2, 119 (Aug. 2013), p. 119.

Lin, L., D. C. Koo, B. J. Weiner, T. Chiueh, et al. In: *The Astrophysical Journal* 660 (May 2007), pp. L51–L54.

Lin, L., D. R. Patton, D. C. Koo, K. Casteels, et al. In: *The Astrophysical Journal* 681 (July 2008), pp. 232–243.

Lin, Y.-T. and J. J. Mohr. In: *The Astrophysical Journal* 617 (Dec. 2004), pp. 879–895.

Liu, D., P. Lang, B. Magnelli, E. Schinnerer, et al. In: *Astrophysical Journal, Supplement* 244.2, 40 (Oct. 2019), p. 40.

Livermore, R. C., S. L. Finkelstein, and J. M. Lotz. In: *The Astrophysical Journal* 835 (Feb. 2017), p. 113.

Loh, E. D. and E. J. Spillar. In: *The Astrophysical Journal* 303 (Apr. 1986), p. 154.

Lovell, C. C., A. P. Vijayan, P. A. Thomas, S. M. Wilkins, et al. In: *Monthly Notices of the RAS* 500.2 (Jan. 2021), pp. 2127–2145.

Lower, S., D. Narayanan, J. Leja, B. D. Johnson, et al. In: *Astrophysical Journal* 904.1, 33 (Nov. 2020), p. 33.

Lustig, P., V. Strazzullo, R.-S. Remus, C. D'Eugenio, et al. In: *arXiv e-prints*, arXiv:2201.09068 (Jan. 2022), arXiv:2201.09068.

Madau, P. and M. Dickinson. In: *Annual Review of Astron and Astrophys* 52 (Aug. 2014), pp. 415–486.

Madau, P., L. Pozzetti, and M. Dickinson. In: *The Astrophysical Journal* 498 (May 1998), pp. 106–116.

Magdis, G. E., D. Elbaz, H. S. Hwang, E. Daddi, et al. In: *Astrophysical Journal, Letters* 720.2 (Sept. 2010), pp. L185–L189.

Magdis, G. E., E. Daddi, D. Elbaz, M. Sargent, et al. In: *The Astrophysical Journal* 740 (Oct. 2011), p. L15.

Magnier, E. A. and J. .-. Cuillandre. In: *Publications of the ASP* 116.819 (May 2004), pp. 449–464.

Main, R. A., B. R. McNamara, P. E. J. Nulsen, H. R. Russell, and A. N. Vantyghem. In: *Monthly Notices of the RAS* 464.4 (Feb. 2017), pp. 4360–4382.

Maiolino, R., S. Gallerani, R. Neri, C. Cicone, et al. In: *Monthly Notices of the RAS* 425.1 (Sept. 2012), pp. L66–L70.

Malmquist, K. G. In: *Meddelanden fran Lunds Astronomiska Observatorium Serie I* 96 (Aug. 1920), pp. 1–11.

Malmquist, K. G. In: *Meddelanden fran Lunds Astronomiska Observatorium Serie I* 100 (Mar. 1922), pp. 1–52.

Maltby, D., O. Almaini, V. Wild, N. Hatch, et al. In: *Monthly Notices of the Royal Astronomical Society: Letters* 459.1 (2016), pp. L114–L118.

Mancone, C. L., A. H. Gonzalez, L. A. Moustakas, and A. Price. In: *Publications of the ASP* 125.934 (Dec. 2013), p. 1514.

Mandelbaum, R. In: *Annual Review of Astron and Astrophys* 56 (Sept. 2018), pp. 393–433.

Mandelbaum, R., U. Seljak, G. Kauffmann, C. M. Hirata, and J. Brinkmann. In: *Monthly Notices of the RAS* 368 (May 2006), pp. 715–731.

Manning, S. M., C. M. Casey, J. A. Zavala, G. E. Magdis, et al. In: *Astrophysical Journal* 925.1, 23 (Jan. 2022), p. 23.

Mannucci, F., G. Cresci, R. Maiolino, A. Marconi, and A. Gnerucci. In: *Monthly Notices of the RAS* 408.4 (Nov. 2010), pp. 2115–2127.

Marchesi, S., F. Civano, M. Elvis, M. Salvato, et al. In: *The Astrophysical Journal* 817 (Jan. 2016), p. 34.

Marchesini, D., P. van Dokkum, R. Quadri, G. Rudnick, et al. In: *The Astrophysical Journal* 656 (Feb. 2007), pp. 42–65.

Marchesini, D., A. Muzzin, M. Stefanon, M. Franx, et al. In: *The Astrophysical Journal* 794 (Oct. 2014), p. 65.

Marchesini, D., M. Stefanon, G. B. Brammer, and K. E. Whitaker. In: *The Astrophysical Journal* 748 (Apr. 2012), p. 126.

Marchesini, D., P. G. van Dokkum, N. M. Förster Schreiber, M. Franx, et al. In: *Astrophysical Journal* 701.2 (Aug. 2009), pp. 1765–1796.

Marchesini, D., K. E. Whitaker, G. Brammer, P. G. van Dokkum, et al. In: *The Astrophysical Journal* 725 (Dec. 2010), pp. 1277–1295.

Marcillac, D., J. R. Rigby, G. H. Rieke, and D. M. Kelly. In: *The Astrophysical Journal* 654 (Jan. 2007), pp. 825–834.

Markevitch, M., A. H. Gonzalez, D. Clowe, A. Vikhlinin, et al. In: *Astrophysical Journal* 606.2 (May 2004), pp. 819–824.

Marrone, D. P., J. S. Spilker, C. C. Hayward, J. D. Vieira, et al. In: *Nature* 553.7686 (Jan. 2018), pp. 51–54.

Marsan, Z. C., A. Muzzin, D. Marchesini, M. Stefanon, et al. In: *The Astrophysical Journal* 924 (Jan. 2022), p. 25.

Martig, M., F. Bournaud, R. Teyssier, and A. Dekel. In: *The Astrophysical Journal* 707 (Dec. 2009), pp. 250–267.

Martin-Navarro, I., J. Brodie, A. Romanowsky, T. Ruiz-Lara, and G. van de Ven. In: *Nature* 553.7688 (2018), pp. 307–309.

Martínez-González, S., R. Wünsch, G. Tenorio-Tagle, S. Silich, et al. June 2022.

Martis, N. S., D. Marchesini, G. B. Brammer, A. Muzzin, et al. In: *Astrophysical Journal, Letters* 827.2, L25 (Aug. 2016), p. L25.

Maseda, M. V., R. Bacon, M. Franx, J. Brinchmann, et al. In: *The Astrophysical Journal* 865 (Sept. 2018), p. L1.

Mason, C. A. and M. Gronke. In: *Monthly Notices of the RAS* 499.1 (Sept. 2020), pp. 1395–1405.

Mason, C. A., R. P. Naidu, S. Tacchella, and J. Leja. In: *Monthly Notices of the Royal Astronomical Society* 489 (Oct. 2019), pp. 2669–2676.

Mason, C. A., M. Trenti, and T. Treu. In: *The Astrophysical Journal* 813 (Nov. 2015), p. 21.

Mason, C. A., T. Treu, S. de Barros, M. Dijkstra, et al. In: *The Astrophysical Journal* 857 (Apr. 2018), p. L11.

Mason, C. A., T. Treu, M. Dijkstra, A. Mesinger, et al. In: *The Astrophysical Journal* 856 (Mar. 2018), p. 2.

Massey, R. and A. Refregier. In: *Monthly Notices of the Royal Astronomical Society* 363 (Oct. 2005), pp. 197–210.

Massey, R., J. Rhodes, R. Ellis, N. Scoville, et al. In: *Nature* 445.7125 (Jan. 2007), pp. 286–290.

Masters, D. C., D. K. Stern, J. G. Cohen, P. L. Capak, et al. In: *The Astrophysical Journal* 841 (June 2017), p. 111.

Masters, D. C., D. K. Stern, J. G. Cohen, P. L. Capak, et al. In: *\apj* 877.2 (June 2019), p. 81.

Masters, D. C., D. K. Stern, J. G. Cohen, P. L. Capak, et al. In: *Astrophysical Journal* 877.2, 81 (June 2019), p. 81.

Mather, J. C., E. S. Cheng, D. A. Cottingham, R. E. Eplee Jr., et al. In: *The Astrophysical Journal* 420 (Jan. 1994), p. 439.

Mawatari, K., A. K. Inoue, T. Hashimoto, J. Silverman, et al. In: *Astrophysical Journal* 889.2, 137 (Feb. 2020), p. 137.

Mawatari, K., T. Yamada, G. G. Fazio, J.-S. Huang, and M. L. N. Ashby. In: *Publications of the ASJ* 68.3, 46 (June 2016), p. 46.

McConachie, I., G. Wilson, B. Forrest, Z. C. Marsan, et al. In: *arXiv e-prints*, arXiv:2109.07696 (Sept. 2021), arXiv:2109.07696.

McCracken, H. J., B. Milvang-Jensen, J. Dunlop, M. Franx, et al. In: *Astronomy and Astrophysics* 544 (Aug. 2012), A156.

McCracken, H. J., M. Wolk, S. Colombi, M. Kilbinger, et al. In: *Monthly Notices of the RAS* 449.1 (May 2015), pp. 901–916.

McGaugh, S. S. and P. van Dokkum. In: *Research Notes of the American Astronomical Society* 5 (Feb. 2021), p. 23.

McInnes, L., J. Healy, and J. Melville. In: *arXiv e-prints*, arXiv:1802.03426 (Feb. 2018), arXiv:1802.03426.

McLean, I. S., C. C. Steidel, H. W. Epps, N. Konidaris, et al. In: *Ground-Based and Airborne Instrumentation for Astronomy IV*. Ed. by I. S. McLean, S. K. Ramsay, and H. Takami. Vol. 8446. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Sept. 2012, 84460J, 84460J.

McLeod, D. J., R. J. McLure, and J. S. Dunlop. In: *MNRAS* 459.4 (July 2016), pp. 3812–3824.

McLeod, D. J., R. J. McLure, J. S. Dunlop, F. Cullen, et al. In: *Monthly Notices of the RAS* 503.3 (May 2021), pp. 4413–4435.

McPartland, C., D. B. Sanders, L. J. Kewley, and S. K. Leslie. In: *Monthly Notices of the Royal Astronomical Society: Letters* 482.1 (Jan. 2019), pp. L129–L133.

Mehta, V., C. Scarlata, P. Capak, I. Davidzon, et al. In: *Astrophysical Journal, Supplement* 235.2, 36 (Apr. 2018), p. 36.

Merlin, E., N. Bourne, M. Castellano, H. C. Ferguson, et al. In: *Astronomy and Astrophysics* 595, A97 (Nov. 2016), A97.

Merlin, E., A. Fontana, H. C. Ferguson, J. S. Dunlop, et al. In: *Astronomy and Astrophysics* 582, A15 (Oct. 2015), A15.

Mihos, J. C. and L. Hernquist. In: *The Astrophysical Journal* 464 (June 1996), p. 641.

Milgrom, M. In: *ApJ* 270 (July 1983), p. 371.

Milvang-Jensen, B., W. Freudling, J. Zabl, J. P. U. Fynbo, et al. In: *Astronomy and Astrophysics* 560, A94 (Dec. 2013), A94.

Miyazaki, S., Y. Komiyama, S. Kawanomoto, Y. Doi, et al. In: *Publications of the ASJ* 70, S1 (Jan. 2018), S1.

Mo, H. J., S. Mao, and S. D. M. White. In: *Monthly Notices of the Royal Astronomical Society* 295 (Apr. 1998), pp. 319–336.

Mobasher, B., P. Capak, N. Z. Scoville, T. Dahlen, et al. In: *The Astrophysical Journal Supplement Series* 172 (Sept. 2007), pp. 117–131.

Mobasher, B., R. Idzi, N. Benítez, A. Cimatti, et al. In: *The Astrophysical Journal* 600 (Jan. 2004), pp. L167–L170.

Mobasher, B., M. Rowan-Robinson, A. Georgakakis, and N. Eaton. In: *Monthly Notices of the RAS* 282.1 (Sept. 1996), pp. L7–L13.

Mobasher, B., R. M. Sharples, and R. S. Ellis. In: *Monthly Notices of the Royal Astronomical Society* 263 (Aug. 1993), p. 560.

Moffat, A. F. J. In: *Astronomy and Astrophysics* 3 (Dec. 1969), p. 455.

Moneti, A., H. J. McCracken, S. Rouberol, O. Herent, et al. ESO. Mar. 2019.

Moneti, A., H. J. McCracken, M. Shuntov, O. B. Kauffmann, et al. In: *arXiv e-prints*, arXiv:2110.13928 (Oct. 2021), arXiv:2110.13928.

Moore, B., S. Ghigna, F. Governato, G. Lake, et al. In: *The Astrophysical Journal* 524 (Oct. 1999), pp. L19–L22.

Moore, B., N. Katz, G. Lake, A. Dressler, and A. Oemler. In: *Nature* 379 (Feb. 1996), pp. 613–616.

Moore, B., G. Lake, and N. Katz. In: *The Astrophysical Journal* 495 (Mar. 1998), pp. 139–151.

Morishita, T., L. E. Abramson, T. Treu, K. B. Schmidt, et al. In: *The Astrophysical Journal* 846 (Sept. 2017), p. 139.

Morley, C. V., J. J. Fortney, M. S. Marley, C. Visscher, et al. In: *Astrophysical Journal* 756.2, 172 (Sept. 2012), p. 172.

Morley, C. V., M. S. Marley, J. J. Fortney, R. Lupu, et al. In: *Astrophysical Journal* 787.1, 78 (May 2014), p. 78.

Mosleh, M., S. Hosseinnejad, S. Z. Hosseini-ShahiSavandi, and S. Tacchella. In: *Astrophysical Journal* 905.2, 170 (Dec. 2020), p. 170.

Moster, B. P., R. S. Somerville, J. A. Newman, and H.-W. Rix. In: *The Astrophysical Journal* 731.2 (2011), p. 113.

Moster, B. P., R. S. Somerville, J. A. Newman, and H.-W. Rix. In: *The Astrophysical Journal* 731.2 (2011), p. 113.

Moster, B. P., T. Naab, and S. D. M. White. In: *Monthly Notices of the RAS* 428.4 (Feb. 2013), pp. 3121–3138.

Moustakas, J., A. L. Coil, J. Aird, M. R. Blanton, et al. In: *Astrophysical Journal* 767.1, 50 (Apr. 2013), p. 50.

Moutard, T., S. Arnouts, O. Ilbert, J. Coupon, et al. In: *antike und abendland* 590 (May 2016), A102.

Murray, N., J. Chiang, S. A. Grossman, and G. M. Voit. In: *The Astrophysical Journal* 451 (Oct. 1995), p. 498.

Mutch, S. J., C. Liu, G. B. Poole, P. M. Geil, et al. In: *Monthly Notices of the Royal Astronomical Society* 463 (Dec. 2016), pp. 3556–3562.

Muzzin, A., D. Marchesini, M. Stefanon, M. Franx, et al. In: *The Astrophysical Journal Supplement Series* 206 (May 2013), p. 8.

Muzzin, A., D. Marchesini, M. Stefanon, M. Franx, et al. In: *Astrophysical Journal* 777.1, 18 (Nov. 2013), p. 18.

Muzzin, A., D. Marchesini, M. Stefanon, M. Franx, et al. In: *The Astrophysical Journal* 777.1 (2013/Aug/16/), p. 18.

Nagamine, K., R. Cen, L. Hernquist, J. P. Ostriker, and V. Springel. In: *ApJ* 618.1 (Jan. 2005), p. 23.

Naidu, R., R. Bouwens, C. Conroy, R. S. Ellis, et al. In: *JWST Proposal. Cycle 1* (Mar. 2021), p. 2279.

Naidu, R. P., S. Tacchella, C. A. Mason, S. Bose, et al. In: *Astrophysical Journal* 892.2, 109 (Apr. 2020), p. 109.

Nandy, K., G. I. Thompson, C. Jamar, A. Monfils, and R. Wilson. In: *Astronomy and Astrophysics* 44 (Nov. 1975), pp. 195–203.

Navarro, J. F., C. S. Frenk, and S. D. M. White. In: *The Astrophysical Journal* 462 (May 1996), p. 563.

Navarro, J. F., C. S. Frenk, and S. D. M. White. In: *The Astrophysical Journal* 490 (Dec. 1997), pp. 493–508.

Nayyeri, H., S. Hemmati, B. Mobasher, H. C. Ferguson, et al. In: *Astrophysical Journal, Supplement* 228.1, 7 (Jan. 2017), p. 7.

Nayyeri, H., S. Hemmati, B. Mobasher, H. C. Ferguson, et al. In: *Astrophysical Journal, Supplement* 228.1, 7 (Jan. 2017), p. 7.

Neistein, E. and A. Dekel. In: *Monthly Notices of the RAS* 388.4 (Aug. 2008), pp. 1792–1802.

Neistein, E., F. C. van den Bosch, and A. Dekel. In: *Monthly Notices of the RAS* 372.2 (Oct. 2006), pp. 933–948.

Netterfield, C. B., P. A. R. Ade, J. J. Bock, J. R. Bond, et al. In: *The Astrophysical Journal* 571 (June 2002), pp. 604–614.

Neumayer, N., M. Cappellari, J. Reunanen, H. .-. Rix, et al. In: *The Astrophysical Journal* 671 (Dec. 2007), pp. 1329–1344.

Newman, A. B., S. Belli, R. S. Ellis, and S. G. Patel. In: *The Astrophysical Journal* 862 (Aug. 2018), p. 126.

Newman, A. B., R. S. Ellis, K. Bundy, and T. Treu. In: *The Astrophysical Journal* 746 (Feb. 2012), p. 162.

Nikolic, B., H. Cullen, and P. Alexander. In: *Monthly Notices of the Royal Astronomical Society* 355 (Dec. 2004), pp. 874–886.

Noeske, K. G., B. J. Weiner, S. M. Faber, C. Papovich, et al. In: *Astrophysical Journal, Letters* 660.1 (May 2007), pp. L43–L46.

Nordström, B., M. Mayor, J. Andersen, J. Holmberg, et al. In: *Astronomy and Astrophysics* 418 (May 2004), pp. 989–1019.

Novak, M., V. Smolčić, J. Delhaize, I. Delvecchio, et al. In: *Astronomy and Astrophysics* 602, A5 (June 2017), A5.

Nyland, K., M. Lacy, A. Sajina, J. Pforr, et al. In: *Astrophysical Journal, Supplement* 230.1, 9 (May 2017), p. 9.

Oesch, P. A., R. J. Bouwens, G. D. Illingworth, C. M. Carollo, et al. In: *The Astrophysical Journal* 709 (Jan. 2010), pp. L16–L20.

Oesch, P. A., R. J. Bouwens, G. D. Illingworth, V. Gonzalez, et al. In: *The Astrophysical Journal* 759 (Nov. 2012), p. 135.

Oesch, P. A., G. Brammer, P. G. van Dokkum, G. D. Illingworth, et al. In: *Astrophysical Journal* 819.2, 129 (Mar. 2016), p. 129.

Oesch, P. A., C. M. Carollo, R. Feldmann, O. Hahn, et al. In: *The Astrophysical Journal* 714 (May 2010), pp. L47–L51.

Oesch, P., R. Bouwens, G. Illingworth, I. Labbé, and M. Stefanon. In: *The Astrophysical Journal* 855.2 (2018), p. 105.

Oke, J. B. In: *ApJS* 27 (Feb. 1974), p. 21.

Onodera, M., A. Renzini, M. Carollo, M. Cappellari, et al. In: *Astrophysical Journal* 755.1, 26 (Aug. 2012), p. 26.

Oort, J. H. In: 91 (Apr. 1940), p. 273.

Ouchi, M., K. Shimasaku, H. Furusawa, T. Saito, et al. In: *The Astrophysical Journal* 723 (Nov. 2010), pp. 869–894.

Pacifici, C., E. da Cunha, S. Charlot, H.-W. Rix, et al. In: *Monthly Notices of the RAS* 447.1 (Feb. 2015), pp. 786–805.

Padovani, P., M. Bonzini, K. I. Kellermann, N. Miller, et al. In: *Monthly Notices of the RAS* 452.2 (Sept. 2015), pp. 1263–1279.

Papovich, C., L. A. Moustakas, M. Dickinson, E. Le Floc'h, et al. In: *The Astrophysical Journal* 640 (Mar. 2006), pp. 92–113.

Papovich, C., M. Dickinson, M. Giavalisco, C. J. Conselice, and H. C. Ferguson. In: *The Astrophysical Journal* 631 (Sept. 2005), pp. 101–120.

Pasquet, J., E. Bertin, M. Treyer, S. Arnouts, and D. Fouchez. In: *Astronomy & Astrophysics, Volume 621, id.A26, <NUMPAGES>15</NUMPAGES> pp.* 621 (Jan. 2019), A26.

Pavesi, R., D. A. Riechers, C. E. Sharon, V. Smolčić, et al. In: *Astrophysical Journal* 861.1, 43 (July 2018), p. 43.

Pawlik, M. M., S. McAlpine, J. W. Trayford, V. Wild, et al. In: *Nat Astron* 3.5 (May 2019), pp. 440–446.

Pawlik, M. M., V. Wild, C. J. Walcher, P. H. Johansson, et al. In: *Monthly Notices of the RAS* 456.3 (Mar. 2016), pp. 3032–3052.

Peacock, J. A. Jan. 1999.

Peng, C. Y., L. C. Ho, C. D. Impey, and H.-W. Rix. In: *Astronomical Journal* 124.1 (July 2002), pp. 266–293.

Peng, C. Y., L. C. Ho, C. D. Impey, and H.-W. Rix. In: *Astronomical Journal* 139.6 (June 2010), pp. 2097–2129.

Peng, Y., R. Maiolino, and R. Cochrane. In: *Nature* 521.7551 (May 2015), pp. 192–195.

Peng, Y.-j., S. J. Lilly, K. Kovač, M. Bolzonella, et al. In: *Astrophysical Journal* 721.1 (Sept. 2010), pp. 193–221.

Peng, Y.-j., S. J. Lilly, A. Renzini, and M. Carollo. In: *Astrophysical Journal* 757.1, 4 (Sept. 2012), p. 4.

Pentericci, L., E. Vanzella, A. Fontana, M. Castellano, et al. In: *The Astrophysical Journal* 793 (Oct. 2014), p. 113.

Penzias, A. A. and R. W. Wilson. In: *The Astrophysical Journal* 142 (July 1965), pp. 419–421.

Perlmutter, S., G. Aldering, M. della Valle, S. Deustua, et al. In: *Nature* 391.6662 (Jan. 1998), pp. 51–54.

Pickles, A. J. In: *Publications of the ASP* 110.749 (July 1998), pp. 863–878.

Pillepich, A., D. Nelson, L. Hernquist, V. Springel, et al. In: *Monthly Notices of the RAS* 475.1 (Mar. 2018), pp. 648–675.

Pirzkal, N., B. Rothberg, R. E. Ryan, S. Malhotra, et al. In: *The Astrophysical Journal* 868 (Nov. 2018), p. 61.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, et al. In: *Astronomy and Astrophysics* 641, A6 (Sept. 2020), A6.

Polletta, M., M. Tajer, L. Maraschi, G. Trinchieri, et al. In: *The Astrophysical Journal* 663 (July 2007), pp. 81–102.

Pope, A., I. Aretxaga, D. Hughes, G. Wilson, and M. Yun. In: *American Astronomical Society Meeting Abstracts #233*. Vol. 233. American Astronomical Society Meeting Abstracts. Jan. 2019, 363.20, p. 363.20.

Pope, A., L. Armus, E. Murphy, S. Aalto, et al. In: *arXiv:1903.05110* [*astro-ph*] (Mar. 2019).

Popesso, P., A. Concas, G. Cresci, S. Belli, et al. In: *arXiv:2203.10487* [*astro-ph*] (Mar. 2022).

Portillo, S. K. N., J. S. Speagle, and D. P. Finkbeiner. In: *Astronomical Journal* 159.4, 165 (Apr. 2020), p. 165.

Pozzetti, L., M. Bolzonella, E. Zucca, G. Zamorani, et al. In: *Astronomy and Astrophysics* 523, A13 (Nov. 2010), A13.

Prescott, M. K. M., C. D. Impey, R. J. Cool, and N. Z. Scoville. In: *Astrophysical Journal* 644.1 (June 2006), pp. 100–115.

Press, W. H. and P. Schechter. In: *The Astrophysical Journal* 187 (Feb. 1974), pp. 425–438.

Prevot, M. L., J. Lequeux, E. Maurice, L. Prevot, and B. Rocca-Volmerange. In: *Astronomy and Astrophysics* 132 (Mar. 1984), pp. 389–392.

Proga, D., J. M. Stone, and T. R. Kallman. In: *The Astrophysical Journal* 543 (Nov. 2000), pp. 686–696.

Puschell, J. J., F. N. Owen, and R. A. Laing. In: *The Astrophysical Journal* 257 (June 1982), pp. L57–L61.

Rácz, G., I. Szapudi, I. Csabai, and L. Dobos. In: *Monthly Notices of the Royal Astronomical Society* 503 (June 2021), pp. 5638–5645.

Reddy, N., M. Dickinson, D. Elbaz, G. Morrison, et al. In: *Astrophysical Journal* 744.2, 154 (Jan. 2012), p. 154.

Reddy, N. A., D. K. Erb, C. C. Steidel, A. E. Shapley, et al. In: *The Astrophysical Journal* 633 (Nov. 2005), pp. 748–767.

Rees, M. J. and J. P. Ostriker. In: *Monthly Notices of the Royal Astronomical Society* 179 (June 1977), pp. 541–559.

Remus, R.-S., K. Dolag, L. K. Bachmann, A. M. Beck, et al. In: *Proc. IAU* 10.S309 (July 2014), pp. 145–148.

Ren, K., M. Trenti, and C. A. Mason. In: *The Astrophysical Journal* 878 (June 2019), p. 114.

Renaud, F., F. Bournaud, O. Agertz, K. Kraljic, et al. In: *Astronomy and Astrophysics* 625 (May 2019), A65.

Renzini, A. and Y.-j. Peng. In: *Astrophysical Journal, Letters* 801.2, L29 (Mar. 2015), p. L29.

Retzlaff, J., P. Rosati, M. Dickinson, B. Vandame, et al. In: *Astronomy and Astrophysics* 511, A50 (Feb. 2010), A50.

Riechers, D. A., P. L. Capak, C. L. Carilli, P. Cox, et al. In: *Astrophysical Journal, Letters* 720.2 (Sept. 2010), pp. L131–L136.

Riechers, D. A., C. L. Carilli, P. L. Capak, N. Z. Scoville, et al. In: *Astrophysical Journal* 796.2, 84 (Dec. 2014), p. 84.

Riechers, D. A., J. A. Hodge, R. Pavesi, E. Daddi, et al. In: *Astrophysical Journal* 895.2, 81 (June 2020), p. 81.

Riess, A. G., A. V. Filippenko, P. Challis, A. Clocchiatti, et al. In: *Astronomical Journal* 116.3 (Sept. 1998), pp. 1009–1038.

Robaina, A. R., E. F. Bell, R. E. Skelton, D. H. McIntosh, et al. In: *The Astrophysical Journal* 704.1 (Oct. 2009), pp. 324–340.

Roberts, M. S. and R. N. Whitehurst. In: *Astrophysical Journal* 201 (Oct. 1975), pp. 327–346.

Roberts, M. S. In: *Astrophysical Journal* 144 (May 1966), p. 639.

Roberts-Borsani, G. W., R. J. Bouwens, P. A. Oesch, I. Labbe, et al. In: *ApJ* 823 (June 2016), p. 143.

Robertson, B. E., R. S. Ellis, J. S. Dunlop, R. J. McLure, and D. P. Stark. In: *Nature* 468.7320 (Nov. 2010), pp. 49–55.

Robertson, B. E., R. S. Ellis, S. R. Furlanetto, and J. S. Dunlop. In: *\apjl* 802.2 (Apr. 2015), p. L19.

Robertson, B. E., S. R. Furlanetto, E. Schneider, S. Charlot, et al. In: *Astrophysical Journal* 768.1, 71 (May 2013), p. 71.

Robotham, A. S. G., D. S. Taranu, R. Tobar, A. Moffett, and S. P. Driver. In: *Monthly Notices of the RAS* 466.2 (Apr. 2017), pp. 1513–1541.

Rodighiero, G., E. Daddi, I. Baronchelli, A. Cimatti, et al. In: *Astrophysical Journal, Letters* 739.2, L40 (Oct. 2011), p. L40.

Rodighiero, G., A. Renzini, E. Daddi, I. Baronchelli, et al. In: *Monthly Notices of the RAS* 443.1 (Sept. 2014), pp. 19–30.

Rodríguez Montero, F., R. Davé, V. Wild, D. Anglés-Alcázar, and D. Narayanan. In: *Monthly Notices of the Royal Astronomical Society* 490.2 (Dec. 2019), pp. 2139–2154.

Roos, O., S. Juneau, F. Bournaud, and J. M. Gabor. In: *Astrophysical Journal* 800.1, 19 (Feb. 2015), p. 19.

Rosani, G., G. B. Caminha, K. I. Caputi, and S. Deshmukh. In: *VizieR Online Data Catalog*, J/A+A/633/A159 (Nov. 2019), J/A+A/633/A159.

Rosani, G., G. B. Caminha, K. I. Caputi, and S. Deshmukh. In: *Astronomy and Astrophysics* 633, A159 (Jan. 2020), A159.

Rosse, E. O. In: *Philosophical Transactions of the Royal Society of London Series I* 134 (Jan. 1844), pp. 321–324.

Rowe, B. T. P., M. Jarvis, R. Mandelbaum, G. M. Bernstein, et al. In: *Astronomy and Computing* 10 (Apr. 2015), pp. 121–150.

Rowlands, K., T. Heckman, V. Wild, N. L. Zakamska, et al. In: *Monthly Notices of the Royal Astronomical Society* 480 (Oct. 2018), pp. 2544–2561.

Rubin, V. C., N. Thonnard, and W. K. Ford Jr. In: *ApJ* 238 (June 1980), p. 471.

Rubin, V. C. and W. K. Ford Jr. In: *ApJ* 159 (Feb. 1970), p. 379.

Ruszkowski, M. and V. Springel. In: *ApJ* 696.2 (May 2009), pp. 1094–1102.

Saintonge, A., B. Catinella, L. J. Tacconi, G. Kauffmann, et al. In: *Astrophysical Journal, Supplement* 233.2, 22 (Dec. 2017), p. 22.

Saintonge, A., D. Lutz, R. Genzel, B. Magnelli, et al. In: *Astrophysical Journal* 778.1, 2 (Nov. 2013), p. 2.

Saito, S., S. de la Torre, O. Ilbert, C. Dubois, et al. In: *Monthly Notices of the RAS* 494.1 (Mar. 2020), pp. 199–217.

Salim, S., R. M. Rich, S. Charlot, J. Brinchmann, et al. In: *Astrophysical Journal, Supplement* 173.2 (Dec. 2007), pp. 267–292.

Salpeter, E. E. In: *Astrophysical Journal* 121 (Jan. 1955), p. 161.

Salvato, M., G. Hasinger, O. Ilbert, G. Zamorani, et al. In: *Astrophysical Journal* 690.2 (Jan. 2009), pp. 1250–1263.

Salvato, M., O. Ilbert, G. Hasinger, A. Rau, et al. In: *Astrophysical Journal* 742.2, 61 (Dec. 2011), p. 61.

Salvato, M., O. Ilbert, and B. Hoyle. In: *Nature Astronomy* 3 (June 2019), pp. 212–222.

Sánchez, S. F., R. C. Kennicutt, A. Gil de Paz, G. van de Ven, et al. In: *Astronomy & Astrophysics, Volume 538, id.A8, <NUMPAGES>31</NUMPAGES> pp.* 538 (Feb. 2012), A8.

Sandage, A., G. A. Tammann, and A. Yahil. In: *Astrophysical Journal* 232 (Sept. 1979), pp. 352–364.

Sandage, A. In: *The Astrophysical Journal* 127 (May 1958), p. 513.

Sanders, D. B. In: 192 (May 1998), p. 62.06.

Sanders, D. B., J. S. Kartaltepe, L. J. Kewley, V. U, et al. In: 408 (Oct. 2009), p. 3.

Sanders, D. B., M. Salvato, H. Aussel, O. Ilbert, et al. In: *Astrophysical Journal, Supplement* 172.1 (Sept. 2007), pp. 86–98.

Sanders, D. B., N. Z. Scoville, and P. M. Solomon. In: *Astrophysical Journal* 289 (Feb. 1985), pp. 373–387.

Sanders, D. B., B. T. Soifer, J. H. Elias, B. F. Madore, et al. In: *The Astrophysical Journal* 325 (Feb. 1988), p. 74.

Sanders, D. B., P. M. Solomon, and N. Z. Scoville. In: *Astrophysical Journal* 276 (Jan. 1984), pp. 182–203.

Santini, P., A. Fontana, A. Grazian, S. Salimbeni, et al. In: *Astronomy and Astrophysics* 538 (Feb. 2012), A33.

Santini, P., M. Castellano, E. Merlin, A. Fontana, et al. In: *Astronomy and Astrophysics* 652, A30 (Aug. 2021), A30.

Santini, P., H. C. Ferguson, A. Fontana, B. Mobasher, et al. In: *The Astrophysical Journal* 801 (Mar. 2015), p. 97.

Santini, P., A. Fontana, M. Castellano, M. Di Criscienzo, et al. In: *Astrophysical Journal* 847.1, 76 (Sept. 2017), p. 76.

Sarzi, M., E. Iodice, L. Coccato, E. M. Corsini, et al. In: *Astronomy & Astrophysics, Volume 616, id.A121, <NUMPAGES>23</NUMPAGES> pp.* 616 (Aug. 2018), A121.

Saumon, D. and M. S. Marley. In: *Astrophysical Journal* 689.2 (Dec. 2008), pp. 1327–1344.

Sawicki, M., S. Arnouts, J. Huang, J. Coupon, et al. In: *Monthly Notices of the Royal Astronomical Society* 489 (Nov. 2019), pp. 5202–5217.

Sawicki, M. and H. K. C. Yee. In: *Astronomical Journal* 115.4 (Apr. 1998), pp. 1329–1339.

Schawinski, K., S. Khochfar, S. Kaviraj, S. K. Yi, et al. In: *Nature* 442 (Aug. 2006), pp. 888–891.

Schechter, P. In: *ApJ* 203 (Jan. 1976), pp. 297–306.

Schechter, P. In: *Astrophysical Journal* 203 (Jan. 1976), pp. 297–306.

Schlafly, E. F. and D. P. Finkbeiner. In: *The Astrophysical Journal* 737 (Aug. 2011), p. 103.

Schlegel, D. J., D. P. Finkbeiner, and M. Davis. In: *The Astrophysical Journal* 500 (June 1998), pp. 525–553.

Schmidt, B. P., N. B. Suntzeff, M. M. Phillips, R. A. Schommer, et al. In: *Astrophysical Journal* 507.1 (Nov. 1998), pp. 46–63.

Schmidt, M. In: *Astrophysical Journal* 151 (Feb. 1968), p. 393.

Schneider, D. P., J. E. Gunn, and J. G. Hoessel. In: *The Astrophysical Journal* 264 (Jan. 1983), pp. 337–355.

Schouws, S., M. Stefanon, R. Bouwens, R. Smit, et al. In: *The Astrophysical Journal* 928 (Mar. 2022), p. 31.

Schreiber, C., K. Glazebrook, T. Nanayakkara, G. G. Kacprzak, et al. In: *Astronomy and Astrophysics* 618, A85 (Oct. 2018), A85.

Schreiber, C., K. Glazebrook, T. Nanayakkara, G. G. Kacprzak, et al. In: *Astronomy & Astrophysics, Volume 618, id.A85, <NUMPAGES>39</NUMPAGES> pp.* 618 (Oct. 2018), A85.

Schreiber, C., I. Labbé, K. Glazebrook, G. Bekiaris, et al. In: *Astronomy and Astrophysics* 611, A22 (Mar. 2018), A22.

Schweizer, F. In: *The Astrophysical Journal* 252 (Jan. 1982), pp. 455–460.

Schweizer, F., P. Seitzer, D. D. Kelson, E. V. Villanueva, and G. L. Walth. In: *The Astrophysical Journal* 773 (Aug. 2013), p. 148.

Scolnic, D. M., D. O. Jones, A. Rest, Y. C. Pan, et al. In: *The Astrophysical Journal* 859 (June 2018), p. 101.

Scoville, N., R. G. Abraham, H. Aussel, J. E. Barnes, et al. In: *Astrophysical Journal, Supplement* 172.1 (Sept. 2007), pp. 38–45.

Scoville, N., S. Arnouts, H. Aussel, A. Benson, et al. In: *Astrophysical Journal, Supplement* 206.1, 3 (May 2013), p. 3.

Scoville, N., H. Aussel, M. Brusa, P. Capak, et al. In: *ApJS* 172.1 (Sept. 2007), p. 1.

Scoville, N., N. Lee, P. Vanden Bout, T. Diaz-Santos, et al. In: *Astrophysical Journal* 837.2, 150 (Mar. 2017), p. 150.

Scoville, N., L. Murchikova, F. Walter, C. Vlahakis, et al. In: *Astrophysical Journal* 836.1, 66 (Feb. 2017), p. 66.

Seaton, M. J. In: *Monthly Notices of the Royal Astronomical Society* 187 (Jan. 1979), p. 73.

Sérsic, J. L. In: *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina* 6 (Feb. 1963), pp. 41–43.

Sérsic, J. L. In: *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina* 6 (Feb. 1963), pp. 41–43.

Shahidi, A., B. Mobasher, H. Nayyeri, S. Hemmati, et al. In: *Astrophysical Journal* 897.1, 44 (July 2020), p. 44.

Shanks, T., P. R. F. Stevenson, R. Fong, and H. T. MacGillivray. In: *Monthly Notices of the Royal Astronomical Society* 206 (Feb. 1984), pp. 767–800.

Shetty, S., M. A. Bershady, K. B. Westfall, M. Cappellari, et al. In: *ApJ* 901.2 (Sept. 2020), p. 101.

Shu, X., L. Yang, D. Liu, W.-H. Wang, et al. In: *Astrophysical Journal* 926.2, 155 (Feb. 2022), p. 155.

Shuntov, M., H. J. McCracken, R. Gavazzi, C. Laigle, et al. In: *arXiv e-prints*, arXiv:2203.10895 (Mar. 2022).

Simon, J. D. and M. Geha. In: *The Astrophysical Journal* 670 (Nov. 2007), pp. 313–331.

Simons, R. C., S. A. Kassin, B. J. Weiner, S. M. Faber, et al. In: *The Astrophysical Journal* 843 (July 2017), p. 46.

Smercina, A., J.-D. T. Smith, K. D. French, E. F. Bell, et al. In: *ApJ* 929.2 (Apr. 2022), p. 154.

Smolčić, V., M. Novak, M. Bondi, P. Ciliegi, et al. In: *Astronomy and Astrophysics* 602, A1 (June 2017), A1.

Smoot, G. F., C. L. Bennett, A. Kogut, E. L. Wright, et al. In: *The Astrophysical Journal* 396 (Sept. 1992), p. L1.

Sneppen, A., C. L. Steinhardt, H. Hensley, A. S. Jermyn, et al. In: *ApJ* 931.1, 57 (May 2022), p. 57.

Snyder, G. F., T. J. Cox, C. C. Hayward, L. Hernquist, and P. Jonsson. In: *The Astrophysical Journal* 741 (Nov. 2011), p. 77.

Sobacchi, E. and A. Mesinger. In: *Monthly Notices of the Royal Astronomical Society* 440 (May 2014), pp. 1662–1673.

Somerville, R. S., P. F. Hopkins, T. J. Cox, B. E. Robertson, and L. Hernquist. In: *\mnras* 391.2 (Dec. 2008), pp. 481–506.

Song, M., S. L. Finkelstein, M. L. N. Ashby, A. Grazian, et al. In: *Astrophysical Journal* 825.1, 5 (July 2016), p. 5.

Songaila, A., L. L. Cowie, E. M. Hu, and J. P. Gardner. In: *The Astrophysical Journal Supplement Series* 94 (Oct. 1994), p. 461.

Sorba, R. and M. Sawicki. In: *Monthly Notices of the RAS* 476.2 (May 2018), pp. 1532–1547.

Speagle, J. S., C. L. Steinhardt, P. L. Capak, and J. D. Silverman. In: *The Astrophysical Journal Supplement Series* 214 (Oct. 2014), p. 15.

Spergel, D. N., L. Verde, H. V. Peiris, E. Komatsu, et al. In: *The Astrophysical Journal Supplement Series* 148 (Sept. 2003), pp. 175–194.

Springel, V., T. Di Matteo, and L. Hernquist. In: *Monthly Notices of the Royal Astronomical Society* 361 (Aug. 2005), pp. 776–794.

Stark, D. P., R. S. Ellis, S. Charlot, J. Chevallard, et al. In: *Monthly Notices of the Royal Astronomical Society* 464 (Jan. 2017), pp. 469–479.

Stefanon, M., M. Aravena, R. Bouwens, R. A. A. Bowler, et al. In: *JWST Proposal. Cycle 1* (Mar. 2021), p. 1626.

Stefanon, M., R. J. Bouwens, I. Labbé, G. D. Illingworth, et al. In: *arXiv e-prints*, arXiv:2103.16571 (Mar. 2021), arXiv:2103.16571.

Stefanon, M., I. Labbé, R. J. Bouwens, P. Oesch, et al. In: *Astrophysical Journal* 883.1, 99 (Sept. 2019), p. 99.

Stefanon, M., I. Labbé, R. J. Bouwens, P. Oesch, et al. In: *Astrophysical Journal* 883.1, 99 (Sept. 2019), p. 99.

Steidel, C., A. Shapley, M. Pettini, K. Adelberger, et al. In: *Multiwavelength Mapping of Galaxy Formation and Evolution*. Ed. by A. Renzini and R. Bender. Jan. 2005, p. 169.

Steidel, C. C., M. Giavalisco, M. Dickinson, and K. L. Adelberger. In: *The Astronomical Journal* 112 (Aug. 1996), p. 352.

Steidel, C. C., M. Giavalisco, M. Pettini, M. Dickinson, and K. L. Adelberger. In: *The Astrophysical Journal* 462 (May 1996), p. L17.

Steinhardt, C. L., M. Jauzac, A. Acebron, H. Atek, et al. In: *The Astrophysical Journal Supplement Series* 247.2 (2020), p. 64.

Steinhardt, C. L., C. K. Jespersen, and N. B. Linzer. In: *Astrophysical Journal* 923.1, 8 (Dec. 2021), p. 8.

Steinhardt, C. L., A. Sneppen, H. Hensley, A. S. Jermyn, et al. June 2022.

Steinhardt, C. L., A. Sneppen, B. Mostafa, H. Hensley, et al. In: *The Astrophysical Journal* 931 (May 2022), p. 58.

Steinhardt, C. L., J. S. Speagle, P. Capak, J. D. Silverman, et al. In: *The Astrophysical Journal Letters* 791 (Aug. 2014), p. L25.

Steinhardt, C. L., J. R. Weaver, J. Maxfield, I. Davidzon, et al. In: *Astrophysical Journal* 891.2, 136 (Mar. 2020), p. 136.

Steinhardt, C. L., P. Capak, D. Masters, and J. S. Speagle. In: *Astrophysical Journal* 824.1, 21 (June 2016), p. 21.

Steinhardt, C. L., P. Capak, D. Masters, and J. S. Speagle. In: *Astrophysical Journal* 824.1, 21 (June 2016), p. 21.

Stern, D., P. Eisenhardt, V. Gorjian, C. S. Kochanek, et al. In: *Astrophysical Journal* 631.1 (Sept. 2005), pp. 163–168.

Stetson, P. B. In: *Publications of the Astronomical Society of the Pacific* 99 (Mar. 1987), p. 191.

Stevans, M. L., S. L. Finkelstein, L. Kawinwanichakij, I. Wold, et al. In: *Astrophysical Journal* 921.1, 58 (Nov. 2021), p. 58.

Stockmann, M., I. Jørgensen, S. Toft, C. J. Conselice, et al. In: *The Astrophysical Journal* 908 (Feb. 2021), p. 135.

Stockmann, M., S. Toft, A. Gallazzi, S. Zibetti, et al. In: *Astrophysical Journal* 888.1, 4 (Jan. 2020), p. 4.

Stoughton, C., R. H. Lupton, M. Bernardi, M. R. Blanton, et al. In: *Astronomical Journal* 123.1 (Jan. 2002), pp. 485–548.

Straatman, C. M. S., I. Labbé, L. R. Spitler, K. Glazebrook, et al. In: *Astrophysical Journal, Letters* 808.1, L29 (July 2015), p. L29.

Su, K.-Y., P. F. Hopkins, C. C. Hayward, X. Ma, et al. In: *Monthly Notices of the Royal Astronomical Society* 487 (Aug. 2019), pp. 4393–4408.

Suess, K. A., M. Kriek, S. H. Price, and G. Barro. In: *Astrophysical Journal* 915.2, 87 (July 2021), p. 87.

Sun, F., E. Egami, P. G. Pérez-González, I. Smail, et al. In: *Astrophysical Journal* 922.2, 114 (Dec. 2021), p. 114.

Szalay, A. S., A. J. Connolly, and G. P. Szokoly. In: *Astronomical Journal* 117 (1999), p. 68.

Szalay, A. S., A. J. Connolly, and G. P. Szokoly. In: *Astronomical Journal* 117.1 (Jan. 1999), pp. 68–74.

Szomoru, D., M. Franx, and P. G. van Dokkum. In: *The Astrophysical Journal* 749 (Apr. 2012), p. 121.

Tacchella, S., A. Dekel, C. M. Carollo, D. Ceverino, et al. In: *Monthly Notices of the RAS* 457.3 (Apr. 2016), pp. 2790–2813.

Tacchella, S., J. C. Forbes, and N. Caplar. In: *Monthly Notices of the Royal Astronomical Society* 497 (Sept. 2020), pp. 698–725.

Tacconi, L. J., R. Genzel, and A. Sternberg. In: *Annual Review of Astronomy and Astrophysics, vol. 58, p.157-203* 58 (Aug. 2020), p. 157.

Takada, M., R. S. Ellis, M. Chiba, J. E. Greene, et al. In: *Publications of the Astronomical Society of Japan* 66 (Feb. 2014), R1.

Takeuchi, T. T. In: *Astrophysics and Space Science* 271 (Jan. 2000), p. 213.

Tanaka, M., F. Valentino, S. Toft, M. Onodera, et al. In: *Astrophysical Journal, Letters* 885.2, L34 (Nov. 2019), p. L34.

Taniguchi, Y., N. Scoville, T. Murayama, D. B. Sanders, et al. In: *The Astrophysical Journal Supplement Series* 172 (Sept. 2007), pp. 9–28.

Taniguchi, Y., M. Kajisawa, M. A. R. Kobayashi, Y. Shioya, et al. In: *Publications of the Astronomical Society of Japan* 67 (Dec. 2015), p. 104.

Teplitz, H. I., M. Rafelski, P. Kurczynski, N. A. Bond, et al. In: *Astronomical Journal* 146.6, 159 (Dec. 2013), p. 159.

Thomas, D., C. Maraston, R. Bender, and C. Mendes de Oliveira. In: *The Astrophysical Journal* 621 (Mar. 2005), pp. 673–694.

Thompson, R., R. Davé, and K. Nagamine. In: *Monthly Notices of the RAS* 452.3 (Sept. 2015), pp. 3030–3037.

Tinker, J., A. V. Kravtsov, A. Klypin, K. Abazajian, et al. In: *Astrophysical Journal* 688.2 (Dec. 2008), pp. 709–728.

Toft, S., M. Franx, P. van Dokkum, N. M. Förster Schreiber, et al. In: *The Astrophysical Journal* 705 (Nov. 2009), pp. 255–260.

Toft, S., V. Smolčić, B. Magnelli, A. Karim, et al. In: *The Astrophysical Journal* 782 (Feb. 2014), p. 68.

Toft, S., P. van Dokkum, M. Franx, I. Labbe, et al. In: *The Astrophysical Journal* 671 (Dec. 2007), pp. 285–302.

Toft, S., P. van Dokkum, M. Franx, R. I. Thompson, et al. In: *The Astrophysical Journal* 624 (May 2005), pp. L9–L12.

Toft, S., J. Zabl, J. Richard, A. Gallazzi, et al. In: *Nature* 546.7659 (June 2017), pp. 510–513.

Tomczak, A. R., R. F. Quadri, K.-V. H. Tran, I. Labbé, et al. In: *Astrophysical Journal* 783.2, 85 (Mar. 2014), p. 85.

Tomczak, A. R., R. F. Quadri, K.-V. H. Tran, I. Labbé, et al. In: *The Astrophysical Journal* 817 (Feb. 2016), p. 118.

Toomre, A. In: *Astrophysical Journal* 139 (May 1964), pp. 1217–1238.

Toomre, A. and J. Toomre. In: *The Astrophysical Journal* 178 (Dec. 1972), pp. 623–666.

Torrey, P., M. Vogelsberger, S. Genel, D. Sijacki, et al. In: *Monthly Notices of the RAS* 438.3 (Mar. 2014), pp. 1985–2004.

Trac, H., R. Cen, and A. Loeb. In: *The Astrophysical Journal* 689 (Dec. 2008), p. L81.

Trac, H., R. Cen, and A. Loeb. In: *Astrophysical Journal, Letters* 689.2 (Dec. 2008), p. L81.

Trapp, A. C. and S. R. Furlanetto. In: *Monthly Notices of the RAS* 499.2 (Dec. 2020), pp. 2401–2415.

Tremblay, G. R., F. Combes, J. B. R. Oonk, H. R. Russell, et al. In: *The Astrophysical Journal* 865 (Sept. 2018), p. 13.

Trenti, M. and M. Stiavelli. In: *Astrophysical Journal* 676.2 (Apr. 2008), pp. 767–780.

Trenti, M. and M. Stiavelli. In: *Astrophysical Journal* 676.2 (Apr. 2008), pp. 767–780.

Treu, T., K. B. Schmidt, G. B. Brammer, B. Vulcani, et al. In: *The Astrophysical Journal* 812 (Oct. 2015), p. 114.

Troxel, M. A., N. MacCrann, J. Zuntz, T. F. Eifler, et al. In: *Physical Review D* 98.4, 043528 (Aug. 2018), p. 043528.

Trujillo, I., C. J. Conselice, K. Bundy, M. C. Cooper, et al. In: *Monthly Notices of the Royal Astronomical Society* 382 (Nov. 2007), pp. 109–120.

Trujillo, I., N. M. Förster Schreiber, G. Rudnick, M. Barden, et al. In: *The Astrophysical Journal* 650 (Oct. 2006), pp. 18–41.

Turner, O. J., M. Cirasuolo, C. M. Harrison, R. J. McLure, et al. In: *Monthly Notices of the Royal Astronomical Society* 471 (Oct. 2017), pp. 1280–1320.

Tyson, J. A., D. M. Wittman, J. F. Hennawi, and D. N. Spergel. In: *Nuclear Physics B Proceedings Supplements* 124 (July 2003), pp. 21–29.

U, V., A. M. Medling, H. Inami, L. Armus, et al. In: *The Astrophysical Journal* 871 (Feb. 2019), p. 166.

Ucci, G., P. Dayal, A. Hutter, G. Yepes, et al. In: *Monthly Notices of the RAS* (May 2021).

Ulrich, M.-H., L. Maraschi, and C. M. Urry. In: *Annual Review of Astronomy and Astrophysics* 35 (Jan. 1997), pp. 445–502.

Urry, C. M. and P. Padovani. In: *Publications of the Astronomical Society of the Pacific, v.107, p.803* 107 (Sept. 1995), p. 803.

Valentino, F., M. Tanaka, I. Davidzon, S. Toft, et al. In: *Astrophysical Journal* 889.2, 93 (Feb. 2020), p. 93.

Valentino, F., M. Tanaka, I. Davidzon, S. Toft, et al. In: *Astrophysical Journal* 889.2, 93 (Feb. 2020), p. 93.

van der Wel, A., M. Franx, P. G. van Dokkum, R. E. Skelton, et al. In: *The Astrophysical Journal* 788 (June 2014), p. 28.

van der Wel, A., K. Noeske, R. Bezanson, C. Pacifici, et al. In: *Astrophysical Journal, Supplement* 223.2, 29 (Apr. 2016), p. 29.

van Dokkum, P. G. In: *The Astrophysical Journal* 674 (Feb. 2008), pp. 29–50.

van Dokkum, P. G., M. Franx, N. M. Förster Schreiber, G. D. Illingworth, et al. In: *The Astrophysical Journal* 611 (Aug. 2004), pp. 703–724.

van Dokkum, P. G., M. Franx, M. Kriek, B. Holden, et al. In: *The Astrophysical Journal* 677 (Apr. 2008), p. L5.

van Dokkum, P. G., E. J. Nelson, M. Franx, P. Oesch, et al. In: *The Astrophysical Journal* 813 (Nov. 2015), p. 23.

Varma, S., M. Huertas-Company, A. Pillepich, D. Nelson, et al. In: *Monthly Notices of the RAS* (Nov. 2021).

Viero, M. P., G. Sun, D. T. Chung, L. Moncelsi, and S. S. Condon. Mar. 2022.

Vijayan, A. P., C. C. Lovell, S. M. Wilkins, P. A. Thomas, et al. In: *Monthly Notices of the RAS* 501.3 (Mar. 2021), pp. 3289–3308.

Vijayan, A. P., S. M. Wilkins, C. C. Lovell, P. A. Thomas, et al. In: *Monthly Notices of the RAS* (Feb. 2022).

Vito, F., W. N. Brandt, G. Yang, R. Gilli, et al. In: *Monthly Notices of the RAS* 473.2 (Jan. 2018), pp. 2378–2406.

Vogelsberger, M., S. Genel, V. Springel, P. Torrey, et al. In: *Monthly Notices of the Royal Astronomical Society* 444 (Oct. 2014), pp. 1518–1547.

Volonteri, M. and A. E. Reines. In: *Astrophysical Journal, Letters* 820.1, L6 (Mar. 2016), p. L6.

Vulcani, B., B. M. Poggianti, A. Oemler, A. Dressler, et al. In: *Astronomy and Astrophysics* 550, A58 (Feb. 2013), A58.

Walcher, C. J., P. R. T. Coelho, A. Gallazzi, G. Bruzual, et al. In: *Astronomy and Astrophysics* 582, A46 (Oct. 2015), A46.

Wang, T., C. Schreiber, D. Elbaz, Y. Yoshimura, et al. In: *Nature* 572.7768 (Aug. 2019), pp. 211–214.

Watson, D., L. Christensen, K. K. Knudsen, J. Richard, et al. In: *Nature* 519 (Mar. 2015), pp. 327–330.

Weaver, J., B. Husemann, H. Kuntschner, I. Martín-Navarro, et al. In: *Astronomy & Astrophysics* 614 (2018), A32.

Weaver, J. R. and M. Gregg. In: *American Astronomical Society Meeting Abstracts*. Vol. 227. 2016.

Weaver, J. R., O. B. Kauffmann, O. Ilbert, H. J. McCracken, et al. In: *Astrophysical Journal, Supplement* 258.1, 11 (Jan. 2022), p. 11.

Weaver, J. R. and M. V. Maseda. In: *American Astronomical Society Meeting Abstracts*. Vol. 229. 2017.

Weaver, J. R. and K. Horne. In: *Monthly Notices of the Royal Astronomical Society* (Jan. 2022).

Weaver, J. R., G. Brammer, C. M. Casey, N. Chartab Soltani, et al. JWST Proposal. Cycle 1. Mar. 2021.

Webb, T. M. A., A. Muzzin, A. Noble, N. Bonaventura, et al. In: *The Astrophysical Journal* 814 (Dec. 2015), p. 96.

Webb, T. M. A., P. van Dokkum, E. Egami, G. Fazio, et al. In: *The Astrophysical Journal* 636 (Jan. 2006), pp. L17–L20.

Wechsler, R. H. and J. L. Tinker. In: *Annual Review of Astron and Astrophys* 56 (Sept. 2018), pp. 435–487.

Weigel, A. K., K. Schawinski, E. Treister, B. Trakhtenbrot, and D. B. Sanders. In: *Monthly Notices of the Royal Astronomical Society* 476.2 (May 2018), pp. 2308–2317.

Weigel, A. K., K. Schawinski, and C. Bruderer. In: *Monthly Notices of the RAS* 459.2 (June 2016), pp. 2150–2187.

Weigel, A. K., K. Schawinski, N. Caplar, A. Carpineti, et al. In: *The Astrophysical Journal* 845 (Aug. 2017), p. 145.

Weijmans, A.-M., P. T. de Zeeuw, E. Emsellem, D. Krajnović, et al. In: *Monthly Notices of the Royal Astronomical Society* 444 (Nov. 2014), pp. 3340–3356.

Werle, A., B. Poggianti, A. Moretti, C. Bellhouse, et al. In: *The Astrophysical Journal* 930 (May 2022), p. 43.

Whitaker, K. E., M. Franx, J. Leja, P. G. van Dokkum, et al. In: *Astrophysical Journal* 795.2, 104 (Nov. 2014), p. 104.

Whitaker, K. E., A. Pope, R. Cybulski, C. M. Casey, et al. In: *The Astrophysical Journal* 850 (Dec. 2017), p. 208.

Whitaker, K. E., P. G. van Dokkum, G. Brammer, and M. Franx. In: *Astrophysical Journal, Letters* 754.2, L29 (Aug. 2012), p. L29.

Whitaker, K. E., C. C. Williams, L. Mowla, J. S. Spilker, et al. In: *Nature* 597.7877 (Sept. 2021), pp. 485–488.

White, S. D. M. and M. J. Rees. In: *Monthly Notices of the Royal Astronomical Society* 183 (May 1978), pp. 341–358.

White, S. D. M. and C. S. Frenk. In: *The Astrophysical Journal* 379 (Sept. 1991), p. 52.

Whitmore, B. C., R. Chandar, F. Schweizer, B. Rothberg, et al. In: *The Astronomical Journal* 140 (July 2010), pp. 75–109.

Whitmore, B. C. and F. Schweizer. In: *The Astronomical Journal* 109 (Mar. 1995), p. 960.

Wild, V., O. Almaini, J. Dunlop, C. Simpson, et al. In: *Monthly Notices of the Royal Astronomical Society* 463.1 (2016), pp. 832–844.

Wild, V., T. Heckman, and S. Charlot. In: *Monthly Notices of the Royal Astronomical Society* 405 (June 2010), pp. 933–947.

Wild, V., J. Walcher, P. Johansson, L. Tresse, et al. In: *Monthly Notices of the Royal Astronomical Society* 395.1 (2009), pp. 144–159.

Wilkinson, A., O. Almaini, V. Wild, D. Maltby, et al. In: *Monthly Notices of the Royal Astronomical Society* 504 (July 2021), pp. 4533–4550.

Wilkinson, C., K. Pimbblet, and J. Stott. In: *Monthly Notices of the Royal Astronomical Society* (2017).

Williams, R. J., R. F. Quadri, M. Franx, P. van Dokkum, and I. Labbé. In: *Astrophysical Journal* 691.2 (Feb. 2009), pp. 1879–1895.

Williams, R. E., B. Blacker, M. Dickinson, W. V. D. Dixon, et al. In: *Astronomical Journal* 112 (Oct. 1996), p. 1335.

Wisnioski, E., N. M. Förster Schreiber, M. Fossati, J. T. Mendel, et al. In: *The Astrophysical Journal* 886 (Dec. 2019), p. 124.

Wisnioski, E., N. M. Förster Schreiber, S. Wuyts, E. Wuyts, et al. In: *The Astrophysical Journal* 799 (Feb. 2015), p. 209.

Wright, A. H., S. P. Driver, and A. S. G. Robotham. In: *Monthly Notices of the RAS* 480.3 (Nov. 2018), pp. 3491–3502.

Wright, A. H., A. S. G. Robotham, N. Bourne, S. P. Driver, et al. In: *Monthly Notices of the RAS* 460.1 (July 2016), pp. 765–801.

Wright, G. S., P. A. James, R. D. Joseph, and I. S. McLean. In: *Nature* 344 (Mar. 1990), pp. 417–419.

Wu, P.-F., A. van der Wel, R. Bezanson, A. Gallazzi, et al. In: *ApJ* 888.2 (Jan. 2020), p. 77.

Wuyts, E., J. Kurk, N. M. Förster Schreiber, R. Genzel, et al. In: *The Astrophysical Journal* 789 (July 2014), p. L40.

Yepes, G., R. Kates, A. Khokhlov, and A. Klypin. In: *Monthly Notices of the Royal Astronomical Society* 284 (Jan. 1997), pp. 235–256.

York, D. G., J. Adelman, J. E. Anderson Jr., S. F. Anderson, et al. In: *The Astronomical Journal* 120.3 (Sept. 2000), pp. 1579–1587.

Zabludoff, A. I., D. Zaritsky, H. Lin, D. Tucker, et al. In: *The Astrophysical Journal* 466 (July 1996), p. 104.

Zahid, H. J., D. Kashino, J. D. Silverman, L. J. Kewley, et al. In: *The Astrophysical Journal* 792 (Sept. 2014), p. 75.

Zamojski, M. A., D. Schiminovich, R. M. Rich, B. Mobasher, et al. In: *The Astrophysical Journal Supplement Series* 172 (Sept. 2007), pp. 468–493.

Zavala, J. A., C. M. Casey, S. M. Manning, M. Aravena, et al. In: *Astrophysical Journal* 909.2, 165 (Mar. 2021), p. 165.

Zheng, Y., V. Wild, N. Lahén, P. H. Johansson, et al. In: *Monthly Notices of the Royal Astronomical Society* 498.1 (2020), pp. 1259–1277.

Zheng, Z.-Y., J. Wang, J. Rhoads, L. Infante, et al. In: *The Astrophysical Journal* 842 (June 2017), p. L22.

SUPPLEMENTARY 1 | HISTORY AND DESTINY OF AN EMERGING EARLY-TYPE GALAXY

Includes the following:

History and destiny of an emerging early-type galaxy new IFU insights on the major-merger remnant NGC 7252

Published in Astronomy & Astrophysics, 2018, 614, A32

J. R. Weaver, B. Husemann, H. Kuntschner, I. Martín-Navarro, F. Bournaud, P.-A. Duc, E. Emsellem, D. Krajnovć, M. Lyubenova, and R. M. McDermid

A&A 614, A32 (2018)
https://doi.org/10.1051/0004-6361/201732448
© ESO 2018

Astronomy & Astrophysics

History and destiny of an emerging early-type galaxy

New IFU insights on the major-merger remnant NGC 7252

J. Weaver[1,2], B. Husemann[2], H. Kuntschner[3], I. Martín-Navarro[2,4], F. Bournaud[5], P.-A. Duc[5,6,7], E. Emsellem[3,8], D. Krajnović[9], M. Lyubenova[3], and R. M. McDermid[10,11]

[1] SUPA, School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK
e-mail: jrw20@st-andrews.ac.uk
[2] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
[3] European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany
[4] University of California Observatories, 1156 High Street, Santa Cruz, CA 95064, USA
[5] Laboratoire AIM Paris-Saclay, CEA/IRFU/SAp, Université Paris Diderot, 91191 Gif-sur-Yvette Cedex, France
[6] Université de Strasbourg, CNRS, Observatoire Astronomique de Strasbourg, UMR 7550, 67000 Strasbourg, France
[7] Université Paris Diderot, AIM, Sorbonne Paris Cité, CEA, CNRS, 91191 Gif sur Yvette, France
[8] Observatoire de Lyon, Centre de Recherche Astrophysique de Lyon and Ecole Normale Supérieure de Lyon, Université Lyon 1, 9 Avenue Charles André, 69230 Saint-Genis Laval, France
[9] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany
[10] Department of Physics and Astronomy, Macquarie University, Sydney, NSW 2109, Australia
[11] Australian Astronomical Observatory, PO Box 915, Sydney, NSW 1670, Australia

Received 11 December 2017 / Accepted 23 January 2018

ABSTRACT

Context. The merging of galaxies is one key aspect in our favourite hierarchical ΛCDM Universe and is an important channel leading to massive quiescent elliptical galaxies. Understanding this complex transformational process is ongoing.
Aims. We aim to study NGC 7252, which is one of the nearest major-merger galaxy remnants, observed ∼1 Gyr after the collision of presumably two gas-rich disc galaxies. It is therefore an ideal laboratory to study the processes inherent to the transformation of disc galaxies to ellipticals.
Methods. We obtained wide-field IFU spectroscopy with the VLT-VIMOS integral-field spectrograph covering the central 50″ × 50″ of NGC 7252 to map the stellar and ionised gas kinematics, and the distribution and conditions of the ionised gas, revealing the extent of ongoing star formation and recent star formation history.
Results. Contrary to previous studies, we find the inner gas disc not to be counter-rotating with respect to the stars. In addition, the stellar kinematics appear complex with a clear indication of a prolate-like rotation component which suggests a polar merger configuration. The ongoing star formation rate is $2.2 \pm 0.6\ M_\odot\ \mathrm{yr}^{-1}$ and implies a typical depletion time of ∼2 Gyr given the molecular gas content. Furthermore, the spatially resolved star formation history suggests a slight radial dependence, moving outwards at later times. We confirm a large AGN-ionised gas cloud previously discovered ∼5 kpc south of the nucleus, and find a higher ionisation state of the ionised gas at the galaxy centre relative to the surrounding gas disc. Although the higher ionisation towards the centre is potentially degenerate within the central star forming ring, it may be associated with a low-luminosity AGN.
Conclusions. Although NGC 7252 has been classified as post-starburst galaxy at the centre, the elliptical-like major-merger remnant still appears very active. A central kpc-scale gas disc has presumably re-formed quickly within the last 100 Myr after final coalescence. The disc features ongoing star formation, implying Gyr long timescale to reach the red sequence through gas consumption alone. While NGC 7252 is useful to probe the transformation from discs to ellipticals, it is not well-suited to study the transformation from blue to red at this point.

Key words. galaxies: individual: NGC 7252 – galaxies: elliptical and lenticular, cD – galaxies: formation – galaxies: interactions

1. Introduction

Mergers of galaxies are a natural consequence of the hierarchical build-up of large-scale structure, and emergent from our concordance Λ cold dark matter (ΛCDM) cosmological model. This model has been shown to provide a good working paradigm to describe the formation and evolution of dark matter haloes and their galaxies within (e.g. Springel et al. 2005b). In particular, the merger of two late-type (disc) galaxies has become a major scenario to assemble massive early-type (elliptical) galaxies, as shown by early numerical simulations (e.g. Hernquist & Barnes 1991; Barnes 1992). Despite the success of this scenario, we have yet to fully understand the various processes at play during major mergers, made complex by intricate baryonic physics, radiative feedback, and the hydrodynamics and chemistry of the gas.

Since the pioneering work to understand the role of hierarchical merging in the formation of early-type galaxies, numerical simulations have continued to mature due to increased resolution, implementation of more refined and extended prescriptions of baryonic physics, and larger statistical samples (e.g. Naab & Burkert 2003; Bournaud et al. 2005; Di Matteo et al. 2008; Genel et al. 2014; Vogelsberger et al. 2014a,b; Schaye et al. 2015; Tsatsi et al. 2017; Li et al. 2018). These and other inclusions now allow for quantitative predictions of various merger remnant properties. One major issue identified by the simulations is that massive early-type galaxies appear too blue, due to high continuous star formation resulting from gas inflow during the merger (e.g. Hopkins et al. 2013). This has often been addressed by invoking energetic feedback from active galactic nuclei (AGN) which too are ignited following gas inflow, but subsequently expel gas from their host galaxy core and hence suppress star formation (e.g. Di Matteo et al. 2005; Springel et al. 2005a; Somerville et al. 2008). This has lead to a popular scenario in which AGN play a key role in shaping the properties of early-type galaxies during their assembly, following from a two disc major merger (e.g. Sanders et al. 1988; Hopkins et al. 2008; Schawinski et al. 2010).

Observationally, integral-field spectroscopy has provided a significant step forwards in understanding early-type galaxy formation by recovering the 2D projected internal dynamics and properties, and hence a fossil record of their formation process. Integral-field unit (IFU) surveys covering a large number of early-type galaxies, such as SAURON (de Zeeuw et al. 2002), ATLAS^{3D} (Cappellari et al. 2011), CALIFA (Sánchez et al. 2012), SAMI (Croom et al. 2012), MaNGA (Bundy et al. 2015), or the MASSIVE surveys (Ma et al. 2014), have revealed a rich variety of kinematic properties and sub-structures (e.g. Emsellem et al. 2004, 2011; Krajnović et al. 2008, 2013; Falcón-Barroso et al. 2017; Veale et al. 2017; van de Sande et al. 2017). Based on the 2D mapped kinematics, the specific angular momentum has been proposed as a metric for classifying fast and slow rotating early-type galaxies (Cappellari et al. 2007; Emsellem et al. 2011). While dedicated numerical simulations have shown that this classification does not reflect a bimodality in their merger histories (Naab et al. 2014), differences between these populations remain. Observationally, the fraction of fast to slow rotators seems to be a strong function of stellar mass (Emsellem et al. 2011; Veale et al. 2017) and fast rotators frequently show strong evidence of disc-like structures, often with increased metallicity and younger ages (Krajnović et al. 2008; Kuntschner et al. 2010). These populations also have been identified as having different star formation histories and star formation quenching timescales (Smethurst et al. 2018).

While the transformation from discs to an elliptical galaxy can be readily addressed through simulations guided mainly by gravitation, the star formation histories are related to complex baryonic physics which are difficult to fully implement. Indeed, early-type galaxies are known to show a great variety of ionised (e.g. Sarzi et al. 2006; Singh et al. 2013), atomic (e.g. Serra et al. 2012) and molecular gas content (e.g. Young et al. 2011), which can have an internal or external origin, as discussed in Davis et al. (2011), which depend on environment and merger history. This highlights the complexity in the star formation histories and associated processes responsible for transforming blue (star-forming) disc galaxies into red (quiescent) elliptical galaxies. While fast quenching of early-type galaxies has been often proposed in light of AGN feedback (e.g. Schawinski et al. 2010), inconsistent quenching timescales of several Gyr have also been observed (e.g. Weigel et al. 2017; Smethurst et al. 2018).

It is therefore still necessary to observe and characterise ongoing major mergers at various evolutionary phases in order to establish a complete picture of their transformation. Given the decreasing major merger rate towards lower redshift (e.g. Lotz et al. 2011; Rodriguez-Gomez et al. 2015), these events are particularly rare in the local Universe. This scarcity has driven detailed studies of spatially resolved nearby major mergers such as the Mice (NGC 4676; e.g. Barnes 2004; Wild et al. 2014), the Antennae (NGC 4038, NGC 4039; e.g. Whitmore et al. 2010; Ueda et al. 2012), or the Atoms-for-peace galaxy (NGC 7252; e.g. Schweizer 1982; Wang et al. 1992; Hibbard et al. 1994; Schweizer et al. 2013). While these individual galaxies are not necessarily representative of the entire population and the total parameter space concerning major mergers, they nonetheless provide important information about the transformational processes.

In this article, we focus on NGC 7252 ($z = 0.016$) aiming to uncover the history and destiny of this enigmatic merger remnant based on new wide-field optical IFU observations. The system is well studied with many ancillary observations available, providing a well-defined framework from which to interpret our new observations. Bright tidal tails are clearly visible around NGC 7252, which are rich in atomic H I gas (Hibbard et al. 1994). Numerical simulations suggest that these features are produced by a major merger of two gas-rich galaxies which rarely last longer than ~500 Myr (e.g. Duc & Renaud 2013) after the first passage of the galaxies. A dedicated numerical simulation of NGC 7252 has been presented by Chien & Barnes (2010) to study the star formation history during this merger. Their simulation predicts a rise in the star formation rate (SFR) when the two galaxies approached the first close passage at pericentre about ~620 Myr ago. Another burst in star formation is associated with the second encounter about ~260 Myr ago, followed by the final coalescence of the nuclei about 215 Myr ago. A prolonged star-formation episode lasting for ~60 Myr is predicted by the simulations without significant levels of star formation thereafter. While the stellar population at the very central galaxy core is indeed characterised by a post-starburst spectrum (Fritze-v. Alvensleben & Gerhard 1994), it is surrounded by a rotating disc of molecular (Wang et al. 1992; Ueda et al. 2014) and ionised gas (Schweizer 1982) within ~8" (2.4 kpc) which is still actively forming stars. The surface brightness profile of NGC 7252 has already become well-described by a $r^{1/4}$ law (e.g. Schweizer 1982; Rothberg & Joseph 2004), suggesting that NGC 7252 is close to finishing the transformational process from two disc galaxies to an elliptical galaxy.

The paper is organised as follows: in Sect. 2 we give a brief report of the observations and data reduction. This is followed by the Sect. 3 where we present the analysis and results which are discussed in more detail in Sect. 4. The paper finishes with a summary and conclusion in Sect. 5. Throughout the paper we adopt a concordance cosmological model with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{m} = 0.3$, and $\Omega_\Lambda = 0.7$.

2. Observations and data reduction

We observe the central 50″ × 50″ of NGC 7252 with the VIsible MultiObject Spectrograph (VIMOS) instrument (Le Févre et al. 2003) using the integral-field unit (IFU) mode in October 2011 as part of programme 088.B-0224 (PI: H. Kuntschner). With the high-resolution blue grating of VIMOS we cover a wavelength range from 4130 Å to 6200 Å with a spectral resolution of $R \sim 2550$ over a 27″ × 27″ field-of-view per individual pointing. Hence, we cover the main body of NGC 7252 with four pointings that slightly overlap to always cover the centre of the galaxy as shown in Fig. 1. Each pointing is observed twice for 1500 s together with 500 s long blank sky field exposure. Three lamp flat exposures as well as one arc lamp exposure are observed as night time calibrations. Standard star observations for our setup are taken as part of the standard calibration plan.

The data reduction is performed entirely with the Py3D data reduction package (Husemann et al. 2013) that has been developed for fibre-fed IFUs as part of the Calar Alto Large Integral

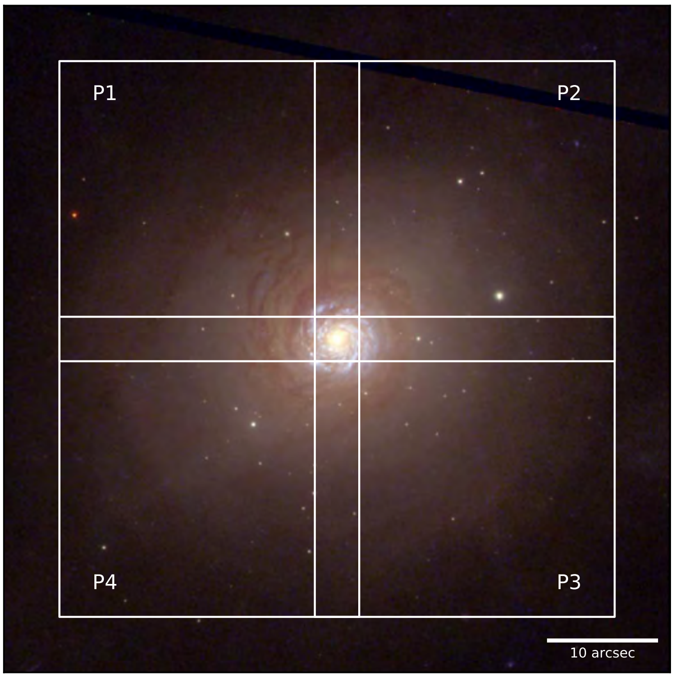

Fig. 1. Field-of-view of the four independent VIMOS pointings. A white box denotes each pointing, overlaid on a colour image of NGC 7252 taken with WFC3 aboard *Hubble* in the bands F336W, F475W and F775W (Bastian et al. 2013).

Field Area survey (CALIFA, Sánchez et al. 2012). It has already been successfully used to reduce several VIMOS IFU data sets (Husemann et al. 2014; Woo et al. 2014; Richtler et al. 2017) and we follow the same scheme here, as briefly described below.

After bias subtraction, cosmic rays are detected with PyCosmic (Husemann et al. 2013) and flagged throughout the processing. Fibres are identified in the combined lamp images and the cross-dispersion width of fibre profiles are estimated as a function of wavelength to allow for an optimal extraction (Horne 1986) of the fibre counts, even in the presence of severe cross-talk between fibres. From the arc lamp observations we determine the wavelength solution for each fibre as well as the spectral resolution which can vary significantly across the four independent spectrographs. We applied the wavelength solution to the data after extraction and adaptively smooth the data to a common spectral resolution of 3 Å, corresponding to the worst resolution found among the data sets. Fibre-to-fibre transmission differences are corrected using a fibre-flat created from the continuum lamp observations of each corresponding observing block. Flexure of the instrument is handled by measuring the shifts of sky line spots in the science data cross-dispersion and dispersion direction. Flux calibration is performed by reducing standard star observations in exactly the same way as the data from which a response function is determined by comparing the observed counts with the reference spectrum.

In post-processing we subtract the sky background as measured from the reduced blank sky field closest in time. Due to varying levels of stray light, the background is subtracted independently in each spectrograph quadrant. The effect of differential atmospheric refraction is accounted for by tracing the bright centre of the galaxy as a function of wavelength in each observation. The final datacube measuring $50'' \times 50''$ is then re-constructed from the individual flux-calibrated and sky-subtracted pointings using the DRIZZLE algorithm (Fruchter & Hook 2002). The galaxy centre serves as the reference point to align all pointings over the entire wavelength range.

When computing absolute quantities from the resulting spectra in the IFU data, a good absolute spectrophotometric calibration is key. We check the accuracy of our absolute spectrophotometric accuracy by comparing the IFU data with the photometry of the available *Hubble* images in the F475W filter, corresponding to Sloan g band. Our VIMOS spectral range covers more than 90% of the effective F475W filter band. We measure $m_{\rm IFU} = 13.0$ mag (AB) compared to $m_{\rm HST} = 12.9$ mag (AB) within an aperture of $30''$ in diameter centred on the nuclear region of NGC 7572. Hence, we adopt a systematic photometric error of $\leq$10% for all absolute quantities inferred from the VIMOS data.

3. Analysis and results

3.1. Stellar and emission line modelling

The signal-to-noise ratio (S/N) of the stellar continuum varies significantly over the VIMOS field-of-view (FoV). Thus, in order to obtain the best stellar kinematics for the entire field, spectra are co-added using the Voronoi binning algorithm (Cappellari & Copin 2003) with a target $S/N \sim 50$. This produces a total of 135 co-added spectra, with the largest bins near the edges of the field where the brightness is significantly less than at the field centre, as expected.

Continua of the 135 spectra are then fitted with PyParadise, an extended Python version of Paradise (Walcher et al. 2015). We adopt the input library of single stellar population spectra from the Medium-Resolution *Isaac Newton* Telescope Library of Empirical Spectra (MILES; Sánchez-Blázquez et al. 2006; Vazdekis et al. 2010; Falcón-Barroso et al. 2011). The upper panel of Fig. 2 is presented as an example, whereby the spectrum of the most central and brightest spaxel, which also corresponds to a single Voronoi bin, is fitted with the continuum model.

Emission line regions are masked during the continuum fitting and normalisation, highlighted in grey. The width of these exclusion regions allows for an emission line to remain masked despite the peculiar velocities of each bin. One may note the O I bright sky line residual around 5577 Å (observed frame), which is also excluded. Although the spectra of most Voronoi bins agree with the target continuum S/N, the spectra of the edge regions, even though heavily binned, still exhibit unacceptably low S/N which is related to limited area per bin at the edges of the VIMOS FoV. These bins are discarded for clarity. The resulting stellar radial velocity and velocity dispersion maps are shown in panels a and b of Fig. 3, respectively.

Emission lines are extracted with PyParadise after the best-fit stellar continuum model is subtracted from the initial data. Given the higher S/N in the emission lines, we repeat the stellar continuum fitting on the unbinned cube but use the previously inferred Voronoi-binned stellar kinematics field as a fixed prior. Hence, only the linear superposition of stellar-population synthesis spectral basis is performed per spaxel and the best-fitting continuum model, at fixed stellar kinematics, is subtracted after which the emission lines are fitted in the residual spectrum. An example of the emission-line modelling for this two-step fitting process is shown in the lower panel of Fig. 2 for the innermost spaxel, which also represents a single Voronoi bin given its high S/N. The emission lines are selected a priori in the rest frame and treated as Gaussian profiles. Kinematic error estimates are derived from a 30 trial bootstrapping on each spectrum. The

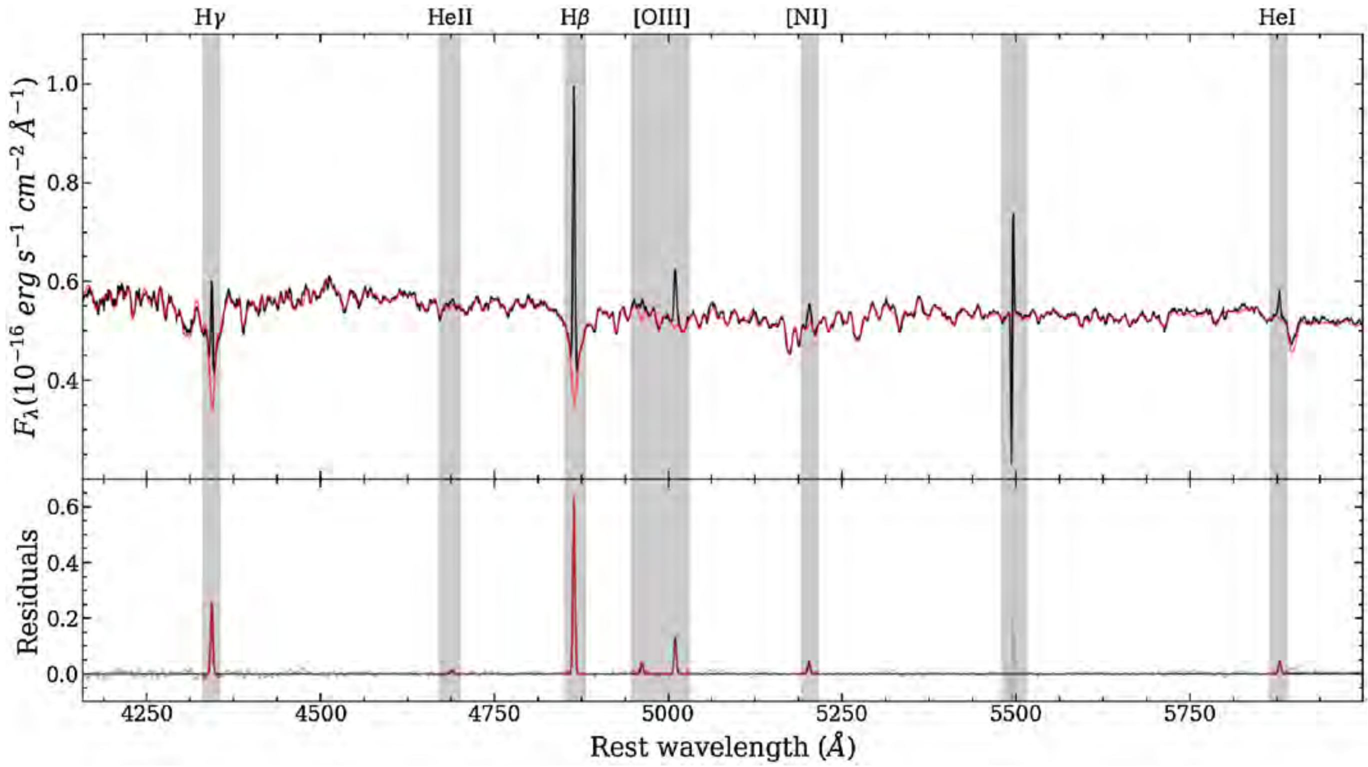

Fig. 2. Example spectral modelling of the brightest spaxel at the galaxy centre. The *top panel* shows the stellar continuum (black), fitted with a model spectrum (red). Regions around emission lines (grey) are omitted from the continuum fit. The bright O I sky line at 5577 Å (observed frame) is also omitted. The *bottom panel* shows the emission line model (red), obtained by fitting the residuals (grey) of the spaxel and continuum model. Rest-frame transformation is calculated from cosmological redshift alone (z = 0.016).

resulting gas radial velocity and velocity dispersion maps are shown in panels e and f of Fig. 3. We also obtain line fluxes with PyParadise. The Hβ and [O III] λ5007 (shortly [O III] hereafter) surface brightness maps are shown in panels c and g, respectively.

As with the continua fitting, the central regions produce high S/N line detections which rapidly deteriorate towards the edges of the field. Spaxels with $S/N < 3$, a radial velocity error >60 km s^{-1}, or FWHM error >150 km s^{-1} are deemed insignificant and discarded. The pixel retention for the gas kinematic maps in panels e and f is carried over from the [O III] flux map in panel g.

3.2. Ionised gas kinematic modelling

One of the most interesting features of NGC 7252 is a central rotating gas disc (see Fig. 3 panels e and f) which exhibits a low velocity dispersion. The p–v diagram (Fig. 4 upper panel) is typical for an inclined rotating disc and we examine the disc kinematics by fitting the radial velocity map using DiskFit (Spekkens & Sellwood 2007; Sellwood & Spekkens 2015). From DiskFit, we obtain a disc kinematic position angle $\mathrm{PA_{gas}}$ = $-65° \pm 1°$ (counter-clockwise with respect to north, solid arrow in Fig. 3 panels e and f), an ellipticity of 0.10 ± 0.04, and a disc inclination of $26° \pm 5°$. Our disc inclination value agrees well with the most recent findings by Ueda et al. (2014), whereby the molecular gas disc is shown to have an inclination angle of $23° \pm 3°$. However, earlier results by Schweizer (1982), obtained with slit spectra, find a conflicting inclination angle of $41° \pm 9°$.

We utilise the inclination-corrected aperture velocity measurements from DiskFit to investigate the velocity curve of the inner 6″ (2 kpc) of the disc. As shown in Fig. 4, the velocity curve flattens around 1.25 kpc to an inclination-corrected circular velocity of $v_{\rm circ} = 229 \pm 45$ km s^{-1}. The apparent decline of velocity in the outer annuli is significantly affected by higher noise and beam smearing effects at the edge of the visible gas disc. Whether the velocity curve would remain flat or really declines is therefore not possible to verify. The velocity of the inner gas disc leads to an enclosed dynamical mass of $M_{\rm dyn}(R < 1.75\,{\rm kpc}) = (2.1 \pm 0.9) \times 10^{10}\,M_\odot$.

In addition to the inner gas disc, we detect outer gas streams to the north-east (NE) and south-west (SW) (Fig. 3 panel e) which also appear to have a low velocity dispersion. As seen in the narrow-band imaging of Schweizer et al. (2013), these gas streams may be related to the larger gas tails which encircle the nucleus. The outer gas streams appear to have a velocity gradient nearly perpendicular to that of the inner gas disc, which initially led to the conclusion that the inner gas disc is counter-rotating with respect to the main body of the galaxy (e.g. Wang et al. 1992). However, if these gas streams are indeed related to the tidal tails this conclusion is not necessarily valid, as discussed further in Sect. 3.3.

The significantly increased velocity dispersion of the gas disc towards the galaxy centre may be either related to shock-like ionisation as indicated by emission line diagnostics (see Sect. 3.4), or simply caused by the beam smearing of the strong gas velocity gradient at the centre. The connection between gas excitation and velocity dispersion as commonly observed in the diffuse ionised gas phase in and around galaxies (e.g. Monreal-Ibero et al. 2010; Ho et al. 2014) is consistent with the first scenario. Indeed, it could be related to weak activity of an AGN as discussed in Sect. 4.3. On the other hand, high central gas dispersions are also

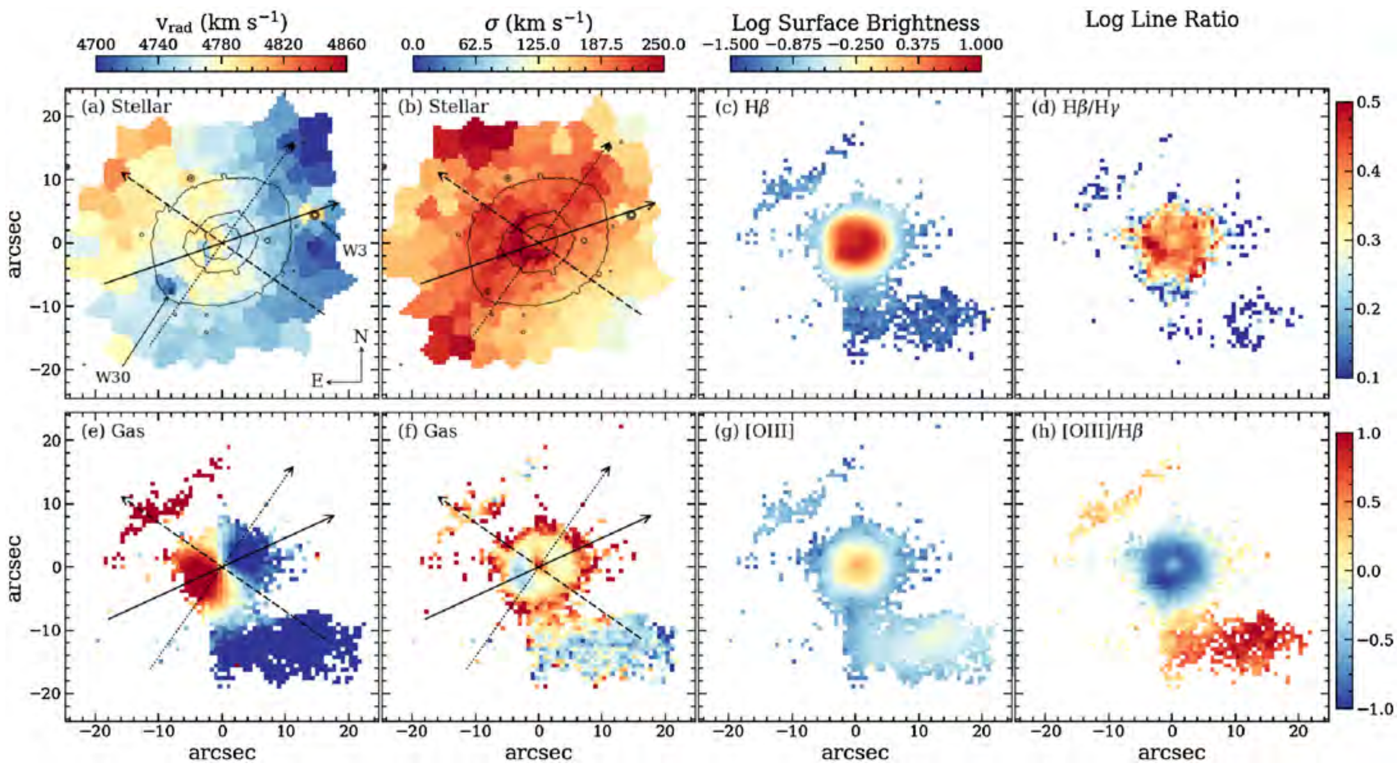

Fig. 3. Kinematic and emission line overview. The stellar maps in *panels a* and *b* are derived from kinematic measurements on 135 Voronoi co-added spectra using PyParadise. Overlaid black contours at 24, 25, and 26 mag arcsec^{-2} (AB) are constructed from smoothed F775W *Hubble* imaging (Bastian et al. 2013). The gas kinematic maps in *panels e* and *f* are derived from emission line measurements on each spaxel, excluding pixels with low S/N or high velocity and dispersion errors. Position angles are indicated by arrows. The major axis of gas and stellar kinematics agree within 6°, shown both as solid lines. The axis of the prolate-like rotation is shown as the dashed line and the photometric PA is shown by the dotted line as measured from elliptical isophotes at >15″. The gas flux maps are shown in *panels c* and *g*, in units of Log(10^{-16} erg s^{-1}cm^{-2}). The Balmer ratio map in *panel d* serves as an approximate extinction map and the variation in ionisation hardness is highlighted by the [OIII]/Hβ ratio in *panel g*.

a common feature in rotating gas discs due to the observational limitation of the beam smearing. Quantifying if beam smearing alone can explain the peak in velocity dispersion with sufficient precision is not possible in this case as we lack information on the intrinsic unsmoothed line flux distribution and a model of the PSF for these observations. In reality, we expect that both effect, beam smearing and change in physical condition, contribute to the high line dispersion at the nucleus.

3.3. Stellar kinematics

The stellar radial velocity map shows a clear east-west velocity gradient with a similar position angle and rotation as the circum-nuclear gas disc. This clearly suggests that the gas disc is not counter-rotating and indeed shares the same angular momentum vector as the stellar component. This rules out the gas counter-rotating with the stars as suggested by Wang et al. (1992).

Similar to the methods described in Sect. 3.2 for the gas disc, we use DiskFit to obtain the position angle of the stellar radial velocity component PA$_{stellar}$, measured over the same inner 6″ (2 kpc). This measurement yields PA$_{stellar}$ = −71° ± 10° (counter-clockwise with respect to north, solid arrow in Fig. 3 panels a and b). PA$_{stellar}$ is readily consistent with PA$_{gas}$. We find that the inclination as derived from the stellar radial velocity map (i = 18° ± 23°) are consistent with results from the gas disc (i = 26° ± 5°), but suffer from slightly larger error bars.

To check whether NGC 7252 can be classified as a fast or slow rotator (e.g. Emsellem et al. 2007, 2011), we computed the projected specific angular momentum λ_R as described in (Emsellem et al. 2007) from the stellar velocity and velocity dispersion maps. Ancillary information on the effective radius (R_{eff}), ellipticity (ϵ) and photometric position angle (PA$_{phot}$) was obtained through ellipse isophotal fitting of the archival *Hubble* F775W and F475W images using the Python Photutils package (Bradley et al. 2017). Those measurements lead to R_{eff} = 4.9 ± 0.3 kpc, ϵ_{eff} = 0.21 ± 0.03 and $\lambda_{R_{eff}}$ = 0.17 ± 0.3 as well as PA$_{phot}$ = −35° (counter-clockwise with respect to north, dotted arrow in Fig. 3). In Fig. 5, we compare our measurements with those of early-type galaxies from ATLAS^{3D} as published by Emsellem et al. (2011). We find that NGC 7252 lies very close to the dividing line, with no robust way to determine its future evolution considering dynamical relaxation and stellar population ageing. Only by comparing observations with matched simulations can we obtain a good prediction, with the caveat that it is difficult and often ambiguous to connect the kinematics to the merger histories (Naab et al. 2014).

However, there is also a noticeable velocity gradient in the north-south direction, that is prolate-like rotation, which exhibits the same sense of motion as the outer gas streams, but with lower amplitude in velocity. The PA of this component is roughly 55° (counter-clockwise with respect to north, Fig. 3) as measured by connecting the highest residual velocity bins. Hence, the stars appear to be a superposition of two different kinematic structures with nearly orthogonal angular momentum vectors. Such a kinematic structure could arise from a polar galaxy merger (Tsatsi et al. 2017) or the spin angular momentum in a radial galaxy merger (Li et al. 2018). Hence, a large range of initial merger conditions can lead to prolate-like rotation (Ebrova & Lokas 2017),

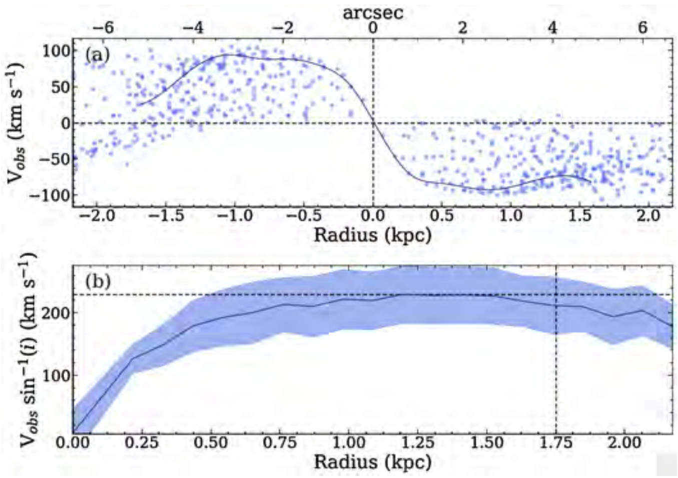

Fig. 4. Velocity pattern of the nuclear star-forming disc. *Panel a: p–v* diagram of the ionised gas kinematics of all pixels within the central 2 kpc. It shows a nice rotational pattern with zero velocity along the minor axis. The velocity profile along the major axis from the DiskFit model is shown as the solid line. *Panel b:* inclination-corrected velocity curve derived from aperture fittings using DiskFit. The curve flattens around 1.25 kpc from the centre of the disc, at a circular velocity of $\upsilon_{\rm circ} = 229 \pm 45\,{\rm km\,s^{-1}}$. We attribute the fall off at large radii to noise. The intersecting dotted black lines indicate the point from which the dynamical mass of the disc is calculated, as discussed in the text.

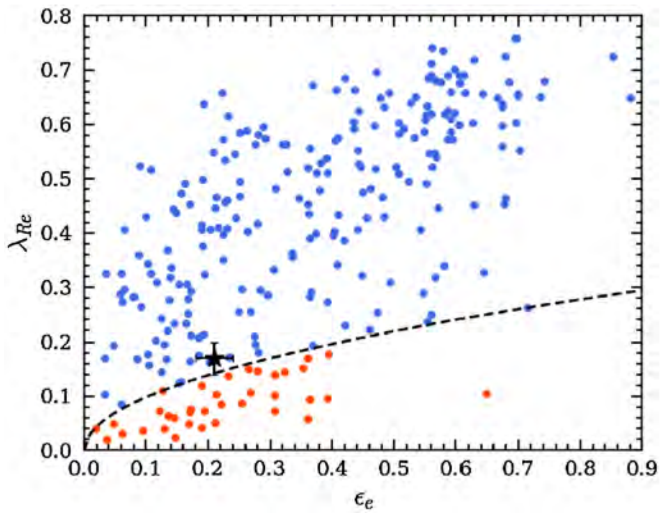

Fig. 5. Specific angular momentum $\lambda_{R_{\rm eff}}$ against ellipticity $\epsilon_{\rm eff}$. The measured value for NGC 7252 $\lambda_{R_{\rm eff}} = 0.17 \pm 0.03$ is shown as the black star and compared to the fast (blue) and slow (red) rotating early-type galaxies as obtained from the ATLAS^{3D} survey (Emsellem et al. 2011). The black dashed line represents the proposed dividing line $\lambda_{R_{\rm eff}} = 0.31\sqrt{\epsilon_{\rm eff}}$ proposed by Emsellem et al. (2011).

which needs to be explored to infer the particular conditions for NGC 7252 and is beyond of the scope of this paper.

This merger configuration is likely reflected in the stellar velocity dispersion map which reveals an elongation spanning some 40″ (13 kpc). The primary axis of this high velocity dispersion elongation seems to be orientated nearly along the isophotal semi-major axis within the nearly orthogonal axes of oblate and prolate-like stellar kinematics. Hence, the high velocity dispersion might be primarily caused by the superposition of two different stellar bodies co-existing within intersecting orbital planes. This is qualitatively supported by the structure in the stellar population history as shown later in Fig. 9.

Another notable feature in the stellar kinematics is an isolated bin just west of the nucleus, which seems to be orbiting counter to the adjacent stars. When compared to photometry, this bin is revealed to be the extensively studied super star cluster W3 (e.g. Maraston et al. 2004; Bastian et al. 2013; Cabrera-Ziri et al. 2016). The connection between the star formation history of W3 and that of NGC 7252 will be explored later in Sect. 3.6. A similar bin is also found southeast of the nucleus, corresponding to the super star cluster W30.

3.4. Ionisation conditions

The inner regions of NGC 7252 are rich in ionised gas, particularly in the central star-forming nuclear star forming disc, the [O III] nebulae to the south-west, and a north-easterly gas stream, as shown in Fig. 3. The innermost regions show intense Hβ emission, as shown in panel c of Fig. 3. Most of the Hβ surface brightness is relatively flat, and concentrated within the star-forming disc. It is also asymmetric, with increased brightness to the south and east. There is slight decrease in the Hβ surface brightness within the central 1″, which has been previously found in high-resolution narrow-band imaging (Schweizer et al. 2013). The Hβ distribution falls off quickly towards the edges of the nuclear star forming disc, and remains relatively flat in the annulus immediately outside. Hβ does not, however, feature strongly in the two outer gas streams.

In contrast, the [O III] flux in panel g of Fig. 3 is centrally concentrated. The distribution falls off much more steeply at all radii, never remaining flat. This may suggest a different ionisation mechanism in the central <1″. [O III] does not reach far beyond the nuclear star forming disc, and hence does not feature strongly in the annulus immediately outside. However, the [O III] flux is dominant in the SW gas stream.

To explore the ionisation conditions of NGC 7252, we construct a line ratio diagnostic diagram (Fig. 6). Since the wavelength coverage of VIMOS prohibits us from constructing the classical [O III]/Hβ versus [N II]/Hα diagram (Baldwin et al. 1981), we replace [N II]/Hα with the ([N I] $\lambda\lambda$5197, 5200)/Hβ line ratio. This surrogate line ratio has been exploited successfully by Sarzi et al. (2010).

While the S/N for all lines is high enough in individual spaxels within the central 12″, examining the properties of the extended gas streams requires co-adding spaxels in order to obtain a reasonable S/N. However, even co-added, the spectra lack detection of the weak [N I] doublet. Hence, we determine 3σ upper limits based on the noise of the co-added spectra. The resulting line ratios for the NE and SW gas streams are indicated in Fig. 6, respectively.

To infer the relevant ionisation mechanisms for each region in the galaxy, we compare the measured line ratios with dusty AGN photoionisation models (Groves et al. 2004), shock-ionisation models (Allen et al. 2008), and starburst photoionisation models (Levesque et al. 2010). The outer regions of the nuclear star forming disc feature shock-driven photoionisation. Moving inwards, we see a smooth transition to towards star formation as the dominant ionisation source. However, the trend reverses again at ∼2 arcsec (0.7 kpc) at which the gas excitation is increasing again towards shock and AGN ionisation with decreasing distance from the centre. We suspect this is due to an additional ionisation mechanisms at play in the centre of NGC 7252, possibly due to a low-luminosity AGN, which is discussed in Sect. 4.3. At very large radii the co-added emission

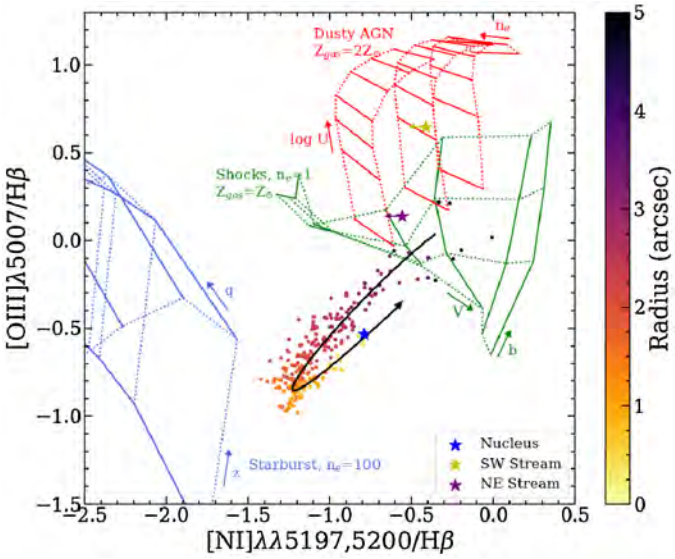

Fig. 6. Diagnostic diagram for the inner 5″ of NGC 7252. Each point corresponds to a pixel, coloured by radius from the centre of the star forming disc. We highlight the systematic trend of the changing-line ratios with decreasing radius by the curved black arrow. Due to low S/N, we are forced to co-add the outer SW and NE gas streams, and determine an upper limit on N I. For comparison we highlight model grids based on AGN photoionisation (red, Groves et al. 2004), shock ionisation (green, Allen et al. 2008), starburst photoionisation (blue, Levesque et al. 2010) for an instantaneous burst with an age of 1 Myr.

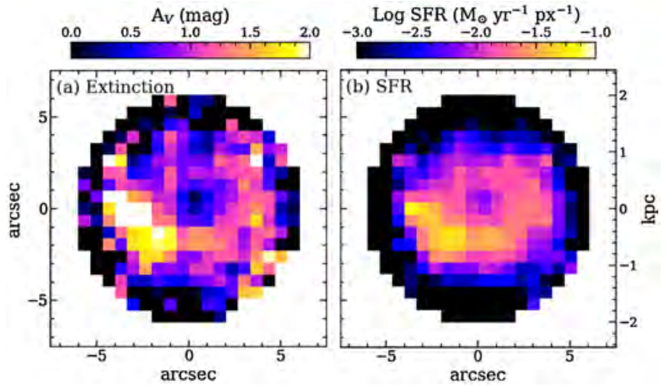

Fig. 7. Extinction and SFR maps of the central starburst derived from gas flux measurements with PyParadise. *Panel a:* A_V extinction calculated from the observed Hβ/Hγ line ratios based on the Milky Way extinction law (Cardelli et al. 1989). *Panel b:* estimated map of the SFR per pixel based on the extinction-corrected Hβ luminosity and the prescription of Calzetti (2013) after conversion to Hα luminosity. No signature of ongoing star formation have been identified outside of the central 12″ based on emission-line diagnostics.

lines ratios of the NE and SW gas streams both reveal signs of either shocked- or AGN-dominated ionisation, consistent with the bright [O III] flux seen in panels g and h of Fig. 3. In particular, the high ionization of the SW stream is consistent with the results of the narrow-band imaging of Schweizer et al. (2013), who argued that this region may be the light echo of a recent AGN episode.

3.5. Conditions for ongoing star formation

The nuclear star forming disc is the only site of active star formation in the 50″ × 50″ FoV, as evidenced by the aforementioned diagnostic plot shown in Fig. 6. Here, we determine the SFR based on the dust-corrected Balmer line fluxes. While usually Hα is employed to count the number of OB stars from their well-known ionising flux, we use the Hβ line converted to the corresponding Hα flux.

We can begin to understand the spatially resolved dust extinction for NGC 7252 by examining the Hβ/Hγ line ratio map shown in panel d of Fig. 3. Going further, the dust extinction map shown in panel a of Fig. 7 is derived from the measured Hγ/Hβ line ratio for which we assume an intrinsic line ratio of 0.468 by adopting Case B recombination with T_e = 10^4 K and n_e = $10^2\,\mathrm{cm}^{-3}$ (Osterbrock & Ferland 2006), and the Milky Way extinction law of Cardelli et al. (1989). The mean extinction within the central 12″ region is found to be 0.65 mag.

In examining both the Balmer ratio and the dust extinction map, one may note the apparent ring structure first seen in studies of the ionised gas by Schweizer (1982) and then of the molecular gas by Wang et al. (1992). We also note a relatively high extinction region in the western portion of the ring, which corresponds to dust features seen in optical imaging, as well as the asymmetric Hβ distribution mentioned in Sect. 3.1.

To use Hα as a SFR calibration, we first compute the dust-corrected Hα flux from Hβ adopting an intrinsic Hα/Hβ line ratio of 2.863 which is based consistently on Case B conditions with T_e = 10^4 K and n_e = $10^2\,\mathrm{cm}^{-3}$ (Osterbrock & Ferland 2006). This leads to an integrated dust-corrected Hα luminosity of $L_{\mathrm{H}\alpha} = (3.9 \pm 0.4) \times 10^{41}\,\mathrm{erg\,s}^{-1}$ from which we then calculate the SFR using the prescription of Calzetti (2013) given by

$$\mathrm{SFR(H}\alpha)/[M_\odot\,\mathrm{yr}^{-1}] = 5.5 \times 10^{-42} L_{\mathrm{H}\alpha}(\mathrm{erg\,s}^{-1}), \qquad (1)$$

which assumes a Kroupa initial mass function (Kroupa 2001) and Case B conditions consistent with our dust correction procedure. The resulting star formation rate map is shown in panel b of Fig. 7 which is clipped to show only pixels predominately ionised by star formation, as indicated by the diagnostic plot in Fig. 6. The corresponding SFR within the VIMOS FoV is computed to be SFR(Hα) = 2.2 ± 0.6 $M_\odot\,\mathrm{yr}^{-1}$. Since star formation is also known to happen far away from the galaxy centre in the tidal tails of NGC 7252 (e.g. Lelli et al. 2015), we also estimated the total SFR. FIR fluxes were measured by *AKARI* (Kawada et al. 2007) which are $f_{90\mu\mathrm{m}}$ = 5.1 Jy and $f_{140\mu\mathrm{m}}$ = 6.4 Jy as listed in the point source catalogue (Yamamura et al. 2010). Following the prescription of Takeuchi et al. (2010) we compute an FIR-based SFR of SFR(FIR) = 6.7 $M_\odot\,\mathrm{yr}^{-1}$ also for a Kroupa IMF. Hence, a substantial fraction of the total star formation is confined to the central few kpc.

CO(1–0) measurements of NGC 7252 obtained by ALMA have provided a molecular gas mass of 4.3×10^9 $M_\odot$ for the inner gas disc (Ueda et al. 2014). Hence, we calculate a molecular depletion time of t_{dep} = 1.9 ± 0.6 Gyr. This is surprisingly close to the average molecular depletion time of 2.35 Gyr for normal star forming disc galaxies (Bigiel et al. 2011), but shorter than expected from quiescent elliptical galaxies (e.g. Saintonge et al. 2012). While gas-rich early-type galaxies (Davis et al. 2014) exhibit on average a factor of ~2.5 lower depletion times, the depletion time can be as short as for normal galaxies if the molecular gas is located more the flat part of galaxies rotation curve.

Based on the results above we explore the location of NGC 7252 in the sSFR–M_* plane (Fig. 8). We use the archival *Hubble* WFC3 images of NGC 7252 in the F475W (Sloan *g*) and

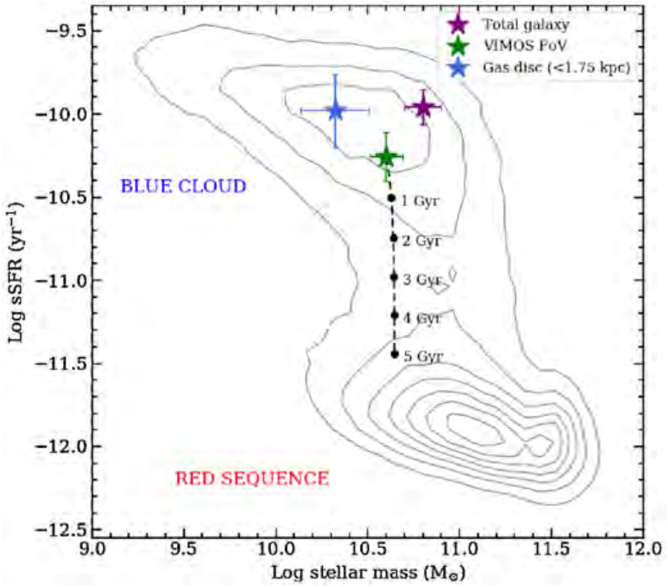

Fig. 8. Specific SFR (sSFR) against stellar mass for NGC 7252 corresponding to the total galaxy (purple), VIMOS field-of-view (green), and nuclear star forming disc (blue). The bimodal galaxy distribution from SDSS MPA-JHU catalog (Kauffmann et al. 2003; Brinchmann et al. 2004) is shown as contours for comparison. Assuming a constant depletion timescale of $t_{\rm dep}$ = 1.9 ± 0.6 Gyr for NGC 7252 based on the molecular gas mass derived from CO(1–0) (Ueda et al. 2014) we predict the position of NGC 7252 for the next 5 Gyr in case of simple gas consumption.

F775W (Sloan i) filters to infer a brightness of m_g = 12.56 mag (AB) and m_i = 11.80 mag (AB), respectively, within an aperture of 30″ radius. Following the empirical stellar mass calibration of Taylor et al. (2011),

$$\log(M_*/[M_\odot]) = 1.15 + 0.7(g - i) - 0.4M_i \tag{2}$$

we compute a stellar mass for NGC 7252 of $\log(M_*/[M_\odot])$ = 10.6 ± 0.1. Compared to the total stellar mass through a 80″ radius aperture, the VIMOS FoV actually covers 70% of the stellar mass. Together with the SFR computed for the VIMOS FoV and total galaxy as described above, we compute specific star formation rates (sSFRs) of $(5.4 \pm 1.9) \times 10^{-11}$ yr^{-1} and $(11 \pm 3) \times 10^{-11}$ yr^{-1} for the VIMOS FoV and the total galaxy, respectively. Comparing this to the distribution of local galaxies from the SDSS MPA/JHU galaxy catalogs (e.g. Kauffmann et al. 2003; Brinchmann et al. 2004) we find that, NGC 7252 is still located in the blue cloud of star forming galaxies. We note that the region of ongoing star formation is much smaller than the host galaxy. Hence, we calculate the sSFR using the dynamical mass of the nuclear star forming disc as described in Sect. 3.2, which is marked in Fig. 8. Since the mass of the star forming disc is smaller than the total mass by a factor of three within the VIMOS FoV, the sSFR of this region is right on the main sequence of star formation. This suggests that the galaxy still contains a settled disc component, which has strong implications for the future evolution of NGC 7252 as discussed later in Sect. 4.4.

3.6. Star formation history

We use the STEllar Content and Kinematics from high-resolution galactic spectra via Maximum A Posteriori code (STECKMAP; Ocvirk et al. 2006a,b) to measure star formation

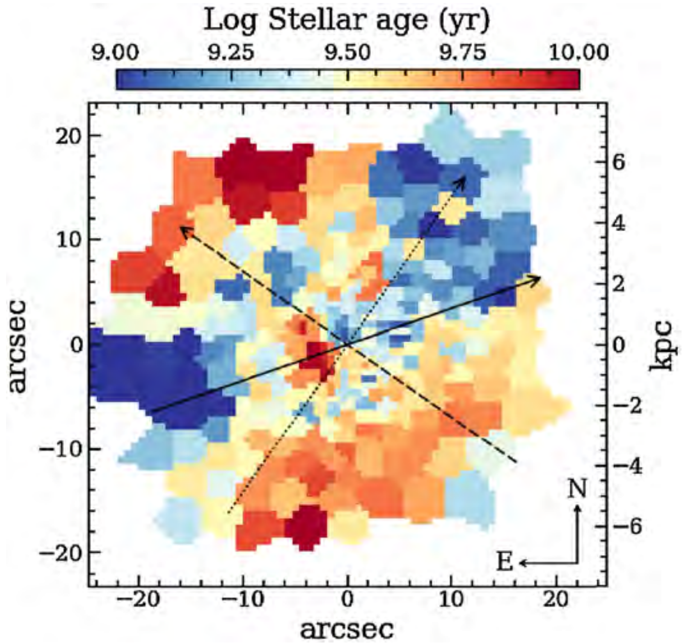

Fig. 9. Luminosity-weighted mean stellar ages from PyParadise. Ages are determined from the stellar population templates used to model the stellar continuum from each Voronoi co-added bin. The younger blue population is spatially distinct from the older population in red. Arrows indicating position angles of the primary stellar rotation axis (solid), primary isophotal axis (dashed), and prolate-like rotation (dotted) are borrowed from Fig. 3.

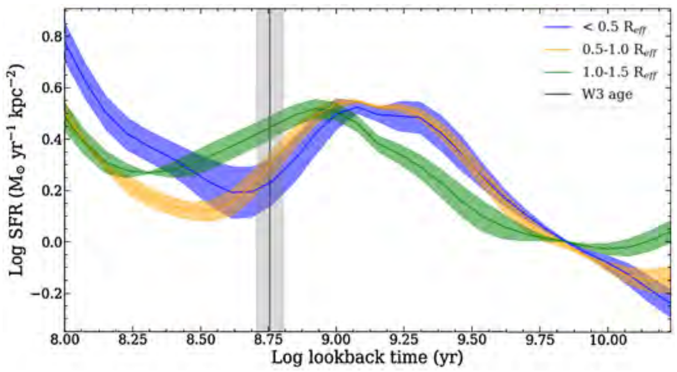

Fig. 10. Star formation histories binned by effective radius. Each curve is binned by 0.5 effective radius, with errors determined from the standard deviation within each spatial bin at a given age. The black vertical line indicates the age of the associated super star cluster W3 as reported by Cabrera-Ziri et al. (2016), with the uncertainty shown in grey.

histories (SFHs) across the VIMOS FoV. STECKMAP recovers age and metallicity distributions in a non-parametric way, by fitting the full observed spectrum. In practice, STECKMAP minimizes the objective function Q_μ defined as

$$Q_\mu = \chi^2(\mathbf{x}, \mathbf{Z}, \mathbf{g}) + P_\mu(\mathbf{x}, \mathbf{Z}, \mathbf{g}) \tag{3}$$

where $\mathbf{x}$ and $\mathbf{Z}$ are the age and metallicity distributions, and $\mathbf{g}$ is the discretised line of sight velocity distribution (LOSVD). The term P_μ is the penalisation function

$$P_\mu = \mu_{\mathbf{x}} P(\mathbf{x}) + \mu_{\mathbf{Z}} P(\mathbf{Z}) + \mu_{\mathbf{v}} P(\mathbf{g}). \tag{4}$$

This penalisation favours smooth solutions and, effectively, it acts as a Gaussian prior $f_{\rm prior} = e^{-\mu P}$ (Ocvirk et al. 2006b).

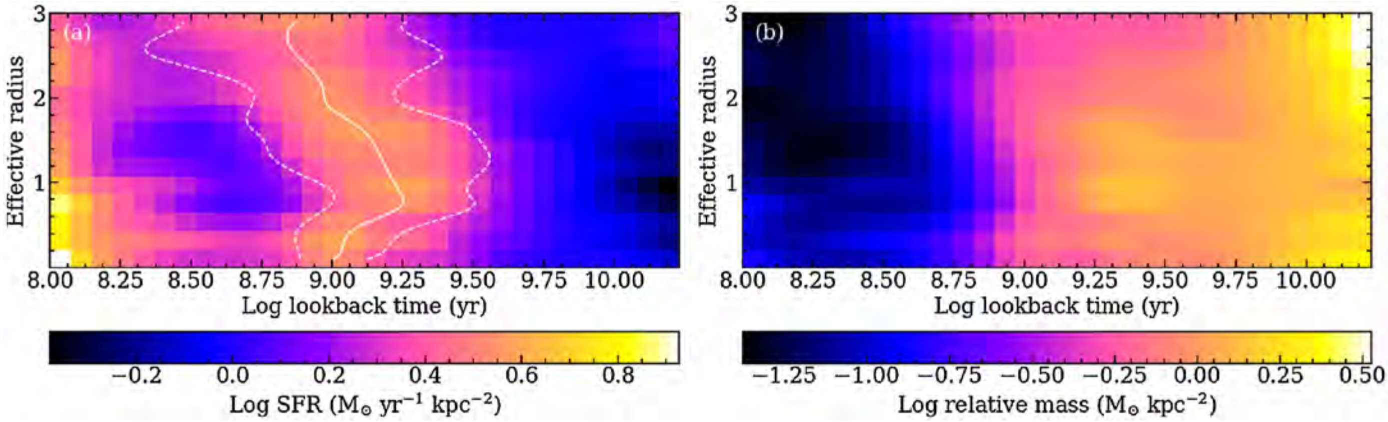

Fig. 11. Summary of the star formation history of NGC 7252 obtained with STECKMAP, normalised over each annulus. *Panel a:* SFR as a function of effective radius, measured across cosmic time. The white solid and dotted lines outline the mean and standard deviation of the initial star formation episode. The current star formation episode is clearly visible in the lower left-hand corner. Similarly, *panel b* shows the stellar mass gained at a given time. As expected, most of the mass was built up slowly, beginning at early times with a second peak around 1 Gyr.

Notice that the penalisation coefficients μ are set separately for the age, metallicity, and LOSVD distributions.

Although STECKMAP allows for a simultaneous recovery of both stellar populations and kinematical properties, it has been shown that metallicity and stellar velocity dispersion are degenerate parameters (Koleva et al. 2008). However, this degeneracy can be safely broken by first measuring the discretised LOSVD, and then fixing it while measuring the stellar population properties (Sánchez-Blázquez et al. 2011). The LOSVD measurements are already performed with PyParadise as described in Sect. 3.3 and used as inputs for STECKMAP.

We feed STECKMAP with the MILES stellar populations synthesis models (Vazdekis et al. 2010). This set of models ranges from −1.3 to −0.22 dex in metallicity, and from 67 Myr to 17 Gyr in age. We note that the MILES base models follow the solar neighbourhood abundance pattern, meaning that they are [Mg/Fe]-enhanced at low metallicities, but [Mg/Fe] ~ 0 for [Fe/H] $\gtrsim 0$. We keep the stellar initial mass function fixed to that measured in the Milky Way (Kroupa 2001). Since the S/N drops significantly on the blue end of the spectra, we limit our analysis to wavelengths between $\lambda = 4500$ Å and $\lambda = 5500$ Å. For this work, we assume $\mu_{\mathbf{x}} = \mu_{\mathbf{Z}} = 20$. This choice is motivated by a joint comparison between the minimum χ^2 value and the structure of the SFHs. Large μ values lead to smoother solutions, therefore losing time-resolution, but assuming low μ values would produce artefacts on the recovered age and metallicity distributions. The choice of a reliable μ vector is not trivial, but does not affect the overall shape of the SFHs. We have assessed this point by varying the adopted μ values, finding no significant differences in our results.

Finally, the age distributions recovered using STECKMAP are translated to mass fractions and SFRs as a function of lookback time, as demonstrated in Figs. 10 and 11. The recovered SFHs strengthen the story of NGC 7252 as a merger remnant. Fig. 10 demonstrates that the SFR peaked roughly 1 Gyr ago whereby the burst occurred slightly earlier in the innermost region (blue) than in the outer regions (orange, then green) at a similar SFR. The current episode of rapid star formation is significantly stronger in the inner region.

Not only does the merger age estimate of 0.6 Gyr as suggested by Schweizer (1982) fit well into this picture, but also the estimated age of 570 Myr for star cluster W3 as inferred by Cabrera-Ziri et al. (2016). In contrast, the star formation history of the entire galaxy appears significantly extended over several 100 Myr during the merger period, more than just a single burst. We will discuss the relation of this extended star formation with its spatial distribution as a result of the merger in the following section.

4. Discussion

4.1. Prolate-like rotation from a gas-rich major merger

Early-type galaxies are usually rotation around their minor axis (oblate rotation), if they regularly rotate at all. Evidence for systems with rotation around the major axis (prolate-like rotation) instead is currently scarce. After many unsuccessful searches (e.g. Bertola et al. 1988), there are currently about a dozen early-type galaxies known to have clear signatures of prolate-like rotation (e.g. Schechter & Gunn 1979; Wagner et al. 1988; Krajnović et al. 2011; Emsellem et al. 2014). IFU surveys like CALIFA (Sánchez et al. 2012), MaNGA (Bundy et al. 2015) or the Most Massive MUSE Galaxies (M3G) survey (PIs: Emsellem & Krajnović) are working to increase the number of galaxies with signatures of prolate-like rotation.

The nature and origin of the prolate-like rotation in early-type galaxies is still under debate. Satellite accretion, as well as major mergers of nearly equal mass have become the two leading formation scenarios (Ebrova & Lokas 2017; Li et al. 2018). Tsatsi et al. (2017) has discovered ten early-type galaxies in CALIFA with signatures of prolate-like rotation, which corresponds to 9% of all early-type galaxies in the same volume. By comparing the observations with N-body simulations they showed that the prolate-like rotation can be produced by a major polar merger. As NGC 7252 has formed from the merger of two disc galaxies, it is very likely that its formation mechanism and observed prolate-like rotation are indeed related. We therefore expect that some part of the angular momentum vectors, composed of the orbital and rotational vectors, should be perpendicular to each other. Whether the prolate-like rotation is predominately produced by one of those angular momentum components is unclear at this point. However, the vast majority of presently known galaxies with prolate-like rotation are found in dense environments and have little or no gas. NGC 7252 is in both ways different from those, suggesting its formation mechanism is different than other known cases. Alternatively, with NGC 7252 we are witnessing the early, still gas rich, stage of the formation of a future classical elliptical with complex kinematics.

Binary mergers are known to produce galaxies with prolate-like rotation (e.g. Naab & Burkert 2003), where the actual kinematics, such as the existence of a kinematically decoupled core (KDC), or the change from an oblate-like (centre) to a prolate-like rotation (outskirts) is dependent on the mergers orbit, spins of the progenitors and the gas content (Hoffman et al. 2010; Bois et al. 2011). NGC 7252 still has a considerable amount of gas, $4.5 \times 10^9\ M_\odot$ of H I (Hibbard et al. 1994) and $3.5 \times 10^9\ M_\odot$ of H_2 (Wang et al. 1992), and its kinematic structure could be viewed as an inner KDC exhibiting an oblate-like (regular) rotation, which changes beyond 10″ (3.3 kpc) into a prolate-like rotation. The velocity map of NGC 7252 resembles those of binary merger simulations of Hoffman et al. (2010) with gas fractions of 20% or more. In these simulations, the gas is responsible for building the short-axis tubes and the regular rotation in the centre, while the outer prolate-like rotation (built of long-axis tubes) can appear in various combinations of the orbital parameters and progenitor spins, and does not necessarily require a polar orbit. Similar conclusions are also found in Bois et al. (2011).

The first numerical simulations of NGC 7252 by Borne & Richstone (1991) and Mihos et al. (1993) employed a retrograde merger of disc galaxies to account for the presumably counter-rotating gas disc reported by Schweizer (1982). Later, Hibbard & Mihos (1995) presented a new self-consistent *N*-body simulation based solely on the large-scale H I kinematics obtained with the Very Large Array. In this model, the inclination of two progenitor galaxies are $i_1 = -40°$ and $i_2 = 70°$ with respect to the merger plane, which would imply perpendicular orbital angular momentum vectors, consistent with a polar merger geometry. However, various other configurations may also be able to produce the observed stellar kinematics that need to be systematically tested by varying the initial conditions of this merger in terms of mass ratio, impact parameter, orbital and spin angular momentum. As stars are collisionless particles, in contrast to the gas, they can probe the kinematics down to the centre of the remnant galaxy, void of H I gas. While performing matched *N*-body simulations is beyond the scope of this paper, our stellar kinematic field should provide valuable constraints on merger parameters for future hydrodynamical simulations.

4.2. The star formation history during the merger event

NGC 7252 provides a snapshot in the evolution of two spiral galaxies into an elliptical. Dedicated numerical simulations of this particular merger remnant have constrained the time since the pericentre of the two discs to be about 770 Myr in the simulations by Borne & Richstone (1991) and 620 Myr in the simulations by Chien & Barnes (2010). This timescale is supported by the ages of associated massive star clusters lying in the range of 400–600 Myr (Schweizer & Seitzer 1998). The age of the well-studied cluster W3, for example, has been found at 570 Myr (Cabrera-Ziri et al. 2016). The centre of NGC 7252 exhibits a post-starburst spectrum with strong Balmer lines (Fritze-v. Alvensleben & Gerhard 1994), implying a strong starburst event less than 1 Gyr ago and an extended period of star formation starting even before pericentre passage. With our IFU data we are able to explore the star formation history of NGC 7252 for the first time, elucidating the nature of the intense star formation episode coincident with the merger.

One main result of our IFU observation is that the starburst episode is significantly extended over several hundred Myr and suggests star formation activity before, during, and after the time of first pericentre passage. In particular, we show that the peak in star formation activity first occurred near the galaxy centre ($<1R_{\mathrm{eff}}$), moving radially outwards at later times. The extended star formation activity prior to the anticipated time of pericentre passage could be explained by a starburst which had been already triggered in the galaxy core, as the galaxies were approaching the closest point of their first pass. Such enhanced star formation activity at the centre of galaxies has already been measured over large samples of galaxies at separations of $\lesssim$20 kpc (Li et al. 2008; Ellison et al. 2008; Wong et al. 2011; Scudder et al. 2012; Patton et al. 2013; Davies et al. 2015). Hence, it is observationally expected that star formation is already enhanced before the final pericentre passage. This is consistent with numerical simulations of nearly equal mass mergers of disc galaxies, exhibiting the highest SFR excess a few hundred Myr before the final pericentre passage (e.g. Moreno et al. 2015) and a suppression after final coalescence.

Our spatially resolved SFH also recovers the period of low star formation activity after pericentre, which reaches a minimum around ~300 Myr ago. The decline in SFR is less effective for the regions which end up at larger radii possibly due to a higher angular momentum than the remnant centre. Those regions may therefore originate primarily from the outskirts of the progenitors rather than the cores and may subject to continuous or enhanced star formation at later time during the merger. The relatively young ages and galacto-centric radius of the star clusters such as W3 matches well with the extended episode of star formation in the outskirts of the merging system. Such a shift in SFR peak from the core to the large scale regions is also reproduced in the numerical simulations by Moreno et al. (2015).

Although the nucleus of NGC 7252 exhibits post-starburst features, ongoing star formation in the central 2 kpc is clearly present. This is readily seen in the current epoch by an increased SFR at small radii, with a similar SFR level appearing to have existed previously ~1 Gyr ago. This on-going star formation activity suggest that the final coalescence of the major merger occurred relatively recently, certainly within the last 200 Myr: numerical simulations of major mergers with masses and dynamical times show that merger-induced star formation stops within 100–200 Myr after the coalescence, even without AGN quenching (Springel et al. 2005a; Bournaud et al. 2011; Torrey et al. 2012; Powell et al. 2013; Moreno et al. 2015). High-resolution hydrodynamic simulations resolving dense gas clouds within merging galaxies have shown that triggered star formation can last longer than previously thought, but only before the coalescence, and rapidly decreases in 100–200 Myr afterwards (e.g. Teyssier et al. 2010; Renaud et al. 2014). Observations also suggest that merger-induced star formation events are short-lived (Ellison et al. 2013).

This also holds for systems at high redshift and/or with high gas fractions (Fensch et al. 2017). Dedicated simulations fitting NGC 7252 by Chien & Barnes (2010) also predict rapidly falling SFR after the starburst phases. In detail these simulations do not predict any recent star formation activity, but their resolution limited to 170 pc for the gravitational softening with a potentially larger SPH smoothing length may miss local events in the nucleus itself – alternatively the true orbit, structure, and kinematics of the progenitor models which merged to form NGC 7252 may somewhat differ from those used in this specific simulation.

Another line of evidence for recent coalescence comes from the bright tidal features visible around NGC 7252: such features rarely last more than ~500 Myr (e.g. Duc & Renaud 2013).

The study of tidal debris in post-merger systems by Ji et al. (2014) indicates that such debris are likely visible at magnitudes brighter than 25 mag arcsec^{-2} during 600 Myr or less for systems with the mass of NGC 7252: debris in NGC 7252 is both much brighter than this limit and fairly continuous over tens of kpc, suggesting that the previous value is a strong upper limit to the age of this post-merger system.

These lines of evidence show that the presumably re-formed disc at the centre of NGC 7252 is young, most likely 100–200 Myr old. Such rapid re-formation of a relaxed and symmetric cold gas disc is consistent with high-resolution simulations of major mergers (e.g. Bournaud et al. 2011; Renaud et al. 2014). The settled gas must have been located in-situ in or near the galaxy nuclei or at least in the inner progenitor galaxy discs, as accretion from the outskirts would require much longer timescales.

Although the SFH of NGC 7252 agree well with generic predictions of merger simulations, some aspects remain puzzling. We anticipate that the spatially resolved SFH of NGC 7252 will provide useful constraints on the initial conditions of gas content and distribution for tailored simulations of this particular major merger system.

4.3. Low-luminosity AGN or LINER?

One of the most recent discoveries from the study of NGC 7252 is the detection of a bright [O III] nebulosity ~5 kpc south-west of the galaxy centre by Schweizer et al. (2013). No indication of ongoing black hole (BH) activity at the nucleus has been yet confirmed in the radio and X-rays, and so Schweizer et al. (2013) concluded that NGC 7252 may have hosted an AGN at its centre that has drastically decreased in luminosity in the recent past. NGC 7252 would therefore fall in the same category as Hanny's Voorwerp (Lintott et al. 2009) and other galaxies that display potential AGN light echos (Keel et al. 2015). Our IFU observations also cover this [O III] nebulosity and we confirm the high excitation of the gas. Based on our emission-line diagnostic diagrams as discussed in Sect. 3.4, we also find evidence of an AGN as the dominant ionisation mechanism in this particular region.

In addition, our VIMOS IFU data provides additional constraints on the gas excitation around the nucleus. As shown in Fig. 7, the dust attenuation map and SFR distribution reveal a drop in dust content and Hα luminosity at the galaxy centre. This agrees with a previously reported deficit within the central 1–2″ of Hα from narrow-band images (Schweizer 1982) and molecular gas (Wang et al. 1992). However, the centrally concentrated [O III] emission leads to an enhanced [O III]/Hβ line ratio. The emission-line diagnostic plot (Fig. 6) confirms that the line ratios indeed increase towards higher excitation at the galaxy centre. Schweizer et al. (2013) also compared the line ratios from narrow-slit spectra of the galaxy centre, but only considered a nuclear and a star forming disc size aperture, missing the gradient apparent in the emission profiles. Given the low spatial resolution of our data we cannot avoid a strong contamination of the nuclear emission lines by the bright star-forming disc, in particular for Hβ. The intrinsic line ratios of the nucleus are then likely located in either the shock or the low-luminosity AGN region of the diagram. Only IFU observations at higher spatial resolution may be able to spatially separate the line ratios from the nucleus and the star forming disc.

If we nevertheless interpret the extinction-corrected ($A_V \sim 1.35$ mag) [O III] line flux of $(3.17 \pm 0.3) \times 10^{-15}$ erg s^{-1} cm^{-2} within the central 4″ × 4″ to be excited by a low-luminosity AGN, we would estimate an AGN bolometric luminosity of $L_{\rm bol} = (2.4 \pm 0.3) \times 10^{41}$ erg s^{-1} following the empirical correlation between [O III] and bolometric luminosity from Stern & Laor (2012). The corresponding 2–10 keV luminosity would be 5.9×10^{40} erg s^{-1}, adopting a bolometric correction of $L_{\rm [O\,III]}/L_{2-10\,\rm keV} \sim 0.03$ from Heckman et al. (2005). This is perfectly consistent with the X-ray luminosity of the nucleus $\log(L_X/[\rm erg\ s^{-1}]) = 40.75^{+0.03}_{-0.04}$ as reported by Nolan et al. (2004) from *Chandra* observations. Hence, the nuclear [O III] and X-ray emission indeed follows the relation implied for BH activity. Schweizer et al. (2013) concluded that most of the X-ray luminosity may originate from star formation from the circumnuclear disc. However, they assume a SFR of 6 $M_\odot$ yr^{-1} for the star-forming disc, which is a factor of three higher than we infer from the IFU data as discussed in Sect. 3.5.

In any case, the putative AGN luminosity is still much lower than the luminosity required to excite the [O III] nebulosity as discussed by Schweizer et al. (2013) and their scenario of a recently faded AGN remains valid. These kind of changing-look AGN are observed with increasing numbers (e.g. Denney et al. 2014; LaMassa et al. 2015; Merloni et al. 2015; McElroy et al. 2016). In particular, Mrk 1018 is not only a galaxy in an advanced major merger stage, but also shows AGN flickering on a timescale of a few years. This flickering may possibly be related to the interaction of a binary BH system, accretion disc wind regulation, or other currently unknown scenarios (Husemann et al. 2016). Whether the variability in BH accretion on 10^3–10^5 yr timescales is particularly enhanced during particular phases of a major merger or simply reflects the usual variability of AGN (e.g. Schawinski et al. 2015) remains to be tested.

Alternatively, the emission-line ratios may also point to a Low-Ionisation Nuclear Emission Region (LINER) for which AGN are not necessarily the primary ionisation mechanism. Also, shocks from supernova or the gravitational infall of gas would lead to line ratios consistent with slow shocks. Furthermore, post-AGB stars have become a more favourable interpretation of LINERs in early-type galaxies (e.g. Singh et al. 2013). The high central concentration of [O III] emission also agrees with such a scenario given the steep surface brightness profile of the stellar continuum corresponding to a de Vaucouleurs law (e.g. Schweizer 1982). Again, given our poor spatial resolution, we cannot readily distinguish between a pure point-like or slightly extended [O III] flux distribution. Also, the contamination of Hβ due to the star forming ring is too high to test whether the expected amount of ionising photons from post-AGB stars is sufficient to explain the excitation, as we are using the equivalent width of Hβ (as a surrogate for Hα; Cid Fernandes et al. 2010). In this case, the estimated AGN bolometric luminosity from the [O III] line would be a very strong upper limit, strengthening the fading AGN scenario.

4.4. When will NGC 7252 reach the red sequence?

While NGC 7252 is still located within the blue cloud of star-forming galaxies in colour-magnitude and sSFR-M_* plane, the high Sérsic index of the merger remnant is consistent with the standard evolutionary picture in which the major merger of two star-forming disc galaxies is responsible for the morphological transformation to an elliptical galaxy. Also, a high SFR is predicted to occur at several times during the merger, as discussed above. However, the process of transitioning from a blue cloud to a red sequence galaxy is less clear. On which timescales the stellar populations evolve is a key question, being either a fast or a slow process. The relatively low number density of green valley galaxies implies that the transition phase should be short, lasting much less than 1 Gyr (e.g. Schawinski et al. 2007, 2014). On the contrary, Martin et al. (2007) found a much longer timescale of 1–6 Gyr for the transition from the blue cloud to the red sequence. Only the latter scenario appears consistent with the gas consumption timescale, which may be the main regulator for the future evolution of NGC 7252. This implies that the transition phase for NGC 7252 is either significantly prolonged and unexpected for gas-rich major mergers, or that other processes such as AGN feedback have yet to kick-in. The later must occur in order to support a rapid quenching of star formation (e.g. Springel et al. 2005a; Di Matteo et al. 2005).

The long transformation time we infer is based on the assumption that the $t_{\rm dep}$ remains unchanged until all the molecular gas is consumed. This timescale may be significantly shortened in the event of another burst in star formation, an increasing fraction of the molecular gas being locked up in the diffuse atomic gas phase and not able to form stars, dynamical suppression of star formation or an expulsion of gas by an AGN. The first option is less likely given that the merger has already passed the coalescence phase and no additional triggering of an enhanced star formation phase is expected in numerical simulations (e.g. Chien & Barnes 2010). In line with the second option, H I-rich elliptical galaxies have been detected and represent roughly 40% of the local ellipticals (Serra et al. 2012) outside of clusters, so it is possible that the final remnant of NGC 7252 will be a H I-rich elliptical galaxy. The dynamical suppression of star formation in early-type galaxies was put forward by Davis et al. (2014), who found that the depletion time was significantly longer for molecular gas located in the rising part of the rotation curve. This mechanisms seem not to apply in the case of NGC 7252 because most of the molecular gas in the star-forming disc is located flat part of the rotation curve as traced by Hα emission (see Fig. 4). Hence, additional tidal forces would be required to bring the gas closer in towards the nucleus on short timescale, which is unlikely to be happen given the advanced evolutionary stage of the merger.

AGN triggering during a major merger is often invoked to rapidly quench star formation by expelling much of the gas with powerful outflows (e.g. Di Matteo et al. 2005). The strong [O III] nebulosity of NGC 7252 already shows that an AGN was likely present in the recent past, as convincingly argued in Schweizer et al. (2013). This is also consistent with the presumed time delay of a few 100 Myr between the peak of star formation and AGN accretion (e.g. Wild et al. 2010; Schawinski et al. 2010). However, no kinematic signature of an AGN-driven outflow on galaxy-scales has yet been detected for NGC 7252. Apparently, the previous AGN phase was either too brief or too weak to develop a galactic outflow. It is unknown when the next AGN phase will be triggered and if it will be powerful enough to develop a sufficiently strong outflow to rapidly quench star formation. The IFU data we present is consistent with the presence of a low-luminosity AGN and an increasing number of AGN are known to change their luminosity on rather short timescales (e.g. Lintott et al. 2009; Denney et al. 2014; LaMassa et al. 2015; McElroy et al. 2016). Even if an AGN will turn on again, it may not be able to affect the settled gas disc, as some simulations have shown that the AGN energy may be able to escape without removing much gas in this configuration (e.g. Gabor & Bournaud 2014). Hence, NGC 7252 already passed the presumed AGN blow-out phase for a purely merger-driven AGN quenching scenario. If the modest gas reservoir currently available to form stars is not first removed by an AGN phase or rapidly depleted by a burst of star formation triggered by dynamical effects (e.g. minor mergers), we conclude that NGC 7252 will eventually die out and become a red elliptical galaxy in several Gyr as predicted for slow transformation scenario (e.g. Martin et al. 2007).

5. Summary and conclusions

Using VLT-VIMOS IFU observations covering the central $50'' \times 50''$ of the recent merger remnant NGC 7252, we obtain maps detailing the stellar and gaseous properties of the galaxy, including kinematics and line strengths. Our major results based on a detailed analysis of these spatially resolved properties can be summarised by the following:

– We confirm the presence of a central rotating gas disc with a kinematic inclination and maximum velocity in agreement with Ueda et al. (2014) and no signature of counter-rotation as initially reported.
– We find clear signatures of both oblate- and prolate-like rotation which may be caused by a polar merger configuration or a radial orbit with different spin angular momentum vectors.
– A study of the SFH of NGC 7252 indicates an extended period of elevated SFR extending until 1 Gyr ago, reaching beyond the anticipated time of first passage, which highlights a complex radially dependent SFH produced throughout the galaxy merger.
– We compute a SFR of $2.2 \pm 0.6\,M_\odot\,{\rm yr}^{-1}$ as measured by the dust-corrected Hα luminosity within the central 7 kpc, corresponding to a molecular gas depletion time of ~2 Gyr given the measured molecular gas mass in this region.
– Higher ionisation conditions of the gas are found at the very centre of NGC 7252 from which we draw an upper limit on $L_{\rm bol} < 2.4 \times 10^{41}\,{\rm erg\,s^{-1}}$ based on the [O III] luminosity, consistent with the X-ray luminosity of the nucleus.
– NGC 7252 is still within the blue cloud of galaxies as seen in the sSFR–M_* plane and assuming a constant molecular gas depletion time we infer a transitional time of about 5 Gyr until it reaches the red sequence through gas consumption.

Given that NGC 7252 is one of the nearest major merger remnants of two disc galaxies, it has been employed as a Rosetta stone in understanding the transformational process from star-forming disc galaxies to quiescent elliptical galaxies. Our IFU observations provide a suite of new diagnostics such as prolate-like rotation in the stellar components and a complex spatially resolved star formation history. All of this information needs to be interpreted alongside dedicated numerical simulations of major mergers, which can now be better tuned to NGC 7252, beyond the scope of this work.

The past history of NGC 7252 seems more and more settled as simulations and observations are converging to a common picture. The final destiny however, still remains open. While NGC 7252 has already transformed into an early-type galaxy based on the light profile, it remains as a blue cloud galaxy based on colours and SFR. A detailed analysis of the star formation history and tidal debris suggests that the central gas disc has presumably re-formed quickly in within 100 Myr since central coalescence of the major merger, most likely from in-situ material rather than infall from the galaxy outskirts. Hence, the transformation from two blue disc galaxies to a quiescent elliptical is far from complete as a significant portion of the transformation has not yet begun. It seems the destiny of NGC 7252 cannot be constrained nor predicted at this point. The most plausible scenario is that the galaxy will likely evolve on the timescale of several Gyr as it slowly consumes the remaining gas. However, we cannot predict further events, such as an intense AGN period, that could rapidly accelerate this process.

Although NGC 7252 provides a strong proof of concept that two disc galaxies can be transformed into an elliptical on rather short timescales, it cannot be used to study the complete transformational process to quiescence. Detailed observations of similar major merger remnants within a range of several 100 Myr before and a few Gyr after first pericentre passage are still required to constrain physical prescriptions of merger simulations and fully understand the evolution in star formation in galaxy mergers.

Acknowledgements. We thank the anonymous referee for their comments which improved our original manuscript. We also extend a special thanks to Glenn van de Ven for helpful discussion about the project. JW acknowledges financial support and thanks Max Planck Institute for Astronomy for the great hospitality during his two month summer student internship in 2017, during which the majority of this work was performed. IMN acknowledges funding from the Marie Skłodowska-Curie Individual Fellowship 702607, and from grant AYA2013-48226-C3-1-P from the Spanish Ministry of Economy and Competitiveness (MINECO). RMcD is the recipient of an Australian Research Council Future Fellowship (project number FT150100333).

References

Allen, M. G., Groves, B. A., Dopita, M. A., Sutherland, R. S., & Kewley, L. J. 2008, ApJS, 178, 20
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Barnes, J. E. 1992, ApJ, 393, 484
Barnes, J. E. 2004, MNRAS, 350, 798
Bastian, N., Schweizer, F., Goudfrooij, P., Larsen, S. S., & Kissler-Patig, M. 2013, MNRAS, 431, 1252
Bertola, F., Galletta, G., Capaccioli, M., & Rampazzo, R. 1988, A&A, 192, 24
Bigiel, F., Leroy, A. K., Walter, F., et al. 2011, ApJ, 730, L13
Bois, M., Emsellem, E., Bournaud, F., et al. 2011, MNRAS, 416, 1654
Borne, K. D., & Richstone, D. O. 1991, ApJ, 369, 111
Bournaud, F., Jog, C. J., & Combes, F. 2005, A&A, 437, 69
Bournaud, F., Chapon, D., Teyssier, R., et al. 2011, ApJ, 730, 4
Bradley, L., Sipocz, B., Robitaille, T., et al. 2017, astropy/photutils: v0.4
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7
Cabrera-Ziri, I., Bastian, N., Hilker, M., et al. 2016, MNRAS, 457, 809
Calzetti, D. 2013, Star Formation Rate Indicators (Cambridge: Cambridge University Press), 419
Cappellari, M., & Copin, Y. 2003, MNRAS, 342, 345
Cappellari, M., Emsellem, E., Bacon, R., et al. 2007, MNRAS, 379, 418
Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, MNRAS, 413, 813
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Chien, L.-H., & Barnes, J. E. 2010, MNRAS, 407, 43
Cid Fernandes, R., Stasińska, G., Schlickmann, M. S., et al. 2010, MNRAS, 403, 1036
Croom, S. M., Lawrence, J. S., Bland-Hawthorn, J., et al. 2012, MNRAS, 421, 872
Davies, L. J. M., Robotham, A. S. G., Driver, S. P., et al. 2015, MNRAS, 452, 616
Davis, T. A., Alatalo, K., Sarzi, M., et al. 2011, MNRAS, 417, 882
Davis, T. A., Young, L. M., Crocker, A. F., et al. 2014, MNRAS, 444, 3427
de Zeeuw, P. T., Bureau, M., Emsellem, E., et al. 2002, MNRAS, 329, 513
Denney, K. D., De Rosa, G., Croxall, K., et al. 2014, ApJ, 796, 134
Di Matteo, P., Bournaud, F., Martig, M., et al. 2008, A&A, 492, 31
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Duc, P.-A., & Renaud, F. 2013, in Lecture Notes in Physics, eds. J. Souchay, S. Mathis, & T. Tokieda (Berlin: Springer Verlag), 861, 327
Ebrova, I., & Lokas, E. L. 2017, ApJ, 850, 144
Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008, AJ, 135, 1877
Ellison, S. L., Mendel, J. T., Patton, D. R., & Scudder, J. M. 2013, MNRAS, 435, 3627
Emsellem, E., Cappellari, M., Peletier, R. F., et al. 2004, MNRAS, 352, 721
Emsellem, E., Cappellari, M., Krajnović, D., et al. 2007, MNRAS, 379, 401
Emsellem, E., Cappellari, M., Krajnović, D., et al. 2011, MNRAS, 414, 888
Emsellem, E., Krajnović, D., & Sarzi, M. 2014, MNRAS, 445, L79
Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95
Falcón-Barroso, J., Lyubenova, M., van de Ven, G., et al. 2017, A&A, 597, A48
Fensch, J., Renaud, F., Bournaud, F., et al. 2017, MNRAS, 465, 1934
Fritze-v. Alvensleben, U., & Gerhard, O. E. 1994, A&A, 285, 775
Fruchter, A. S., & Hook, R. N. 2002, PASP, 114, 144
Gabor, J. M., & Bournaud, F. 2014, MNRAS, 441, 1615
Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175
Groves, B. A., Dopita, M. A., & Sutherland, R. S. 2004, ApJS, 153, 9
Heckman, T. M., Ptak, A., Hornschemeier, A., & Kauffmann, G. 2005, ApJ, 634, 161
Hernquist, L., & Barnes, J. E. 1991, Nature, 354, 210
Hibbard, J. E., & Mihos, J. C. 1995, AJ, 110, 140
Hibbard, J. E., Guhathakurta, P., van Gorkom, J. H., & Schweizer, F. 1994, AJ, 107, 67
Ho, I.-T., Kewley, L. J., Dopita, M. A., et al. 2014, MNRAS, 444, 3894
Hoffman, L., Cox, T. J., Dutta, S., & Hernquist, L. 2010, ApJ, 723, 818
Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš D. 2008, ApJS, 175, 356
Hopkins, P. F., Cox, T. J., Hernquist, L., et al. 2013, MNRAS, 430, 1901
Horne, K. 1986, PASP, 98, 609
Husemann, B., Jahnke, K., Sánchez, S. F., et al. 2013, A&A, 549, A87
Husemann, B., Jahnke, K., Sánchez, S. F., et al. 2014, MNRAS, 443, 755
Husemann, B., Urrutia, T., Tremblay, G. R., et al. 2016, A&A, 593, L9
Ji, I., Peirani, S., & Yi, S. K. 2014, A&A, 566, A97
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33
Kawada, M., Baba, H., Barthel, P. D., et al. 2007, PASJ, 59, S389
Keel, W. C., Maksym, W. P., Bennert, V. N., et al. 2015, AJ, 149, 155
Koleva, M., Prugniel, P., Ocvirk, P., Le Borgne, D., & Soubiran, C. 2008, MNRAS, 385, 1998
Krajnović, D., Bacon, R., Cappellari, M., et al. 2008, MNRAS, 390, 93
Krajnović, D., Emsellem, E., Cappellari, M., et al. 2011, MNRAS, 414, 2923
Krajnović, D., Alatalo, K., Blitz, L., et al. 2013, MNRAS, 432, 1768
Kroupa, P. 2001, MNRAS, 322, 231
Kuntschner, H., Emsellem, E., Bacon, R., et al. 2010, MNRAS, 408, 97
LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, ApJ, 800, 144
Le Févre, O., Saisse, M., Mancini, D., et al. 2003, in Proc. SPIE, eds. M. Iye, & A. F. M. Moorwood, 4841, 1671
Lelli, F., Duc, P.-A., Brinks, E., et al. 2015, A&A, 584, A113
Levesque, E. M., Kewley, L. J., & Larson, K. L. 2010, AJ, 139, 712
Li, C., Kauffmann, G., Heckman, T. M., White, S. D. M., & Jing, Y. P. 2008, MNRAS, 385, 1915
Li, H., Mao, S., Emsellem, E., et al. 2018, MNRAS, 473, 1489
Lintott, C. J., Schawinski, K., Keel, W., et al. 2009, MNRAS, 399, 129
Lotz, J. M., Jonsson, P., Cox, T. J., et al. 2011, ApJ, 742, 103
Ma, C.-P., Greene, J. E., McConnell, N., et al. 2014, ApJ, 795, 158
Maraston, C., Bastian, N., Saglia, R. P., et al. 2004, A&A, 416, 467
Martin, D. C., Wyder, T. K., Schiminovich, D., et al. 2007, ApJS, 173, 342
McElroy, R., Husemann, B., Croom, S. M., et al. 2016, A&A, 593, L8
Merloni, A., Dwelly, T., Salvato, M., et al. 2015, MNRAS, 452, 69
Mihos, J. C., Bothun, G. D., & Richstone, D. O. 1993, ApJ, 418, 82
Monreal-Ibero, A., Arribas, S., Colina, L., et al. 2010, A&A, 517, A28
Moreno, J., Torrey, P., Ellison, S. L., et al. 2015, MNRAS, 448, 1107
Naab, T., & Burkert, A. 2003, ApJ, 597, 893
Naab, T., Oser, L., Emsellem, E., et al. 2014, MNRAS, 444, 3357
Nolan, L. A., Ponman, T. J., Read, A. M., & Schweizer, F. 2004, MNRAS, 353, 221
Ocvirk, P., Pichon, C., Lançon, A., & Thiébaut, E. 2006a, MNRAS, 365, 74
Ocvirk, P., Pichon, C., Lançon, A., & Thiébaut, E. 2006b, MNRAS, 365, 46
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei, eds. D. E. Osterbrock, & G. J. Ferland, 2nd ed. (Sausalito, CA: University Science Books)
Patton, D. R., Torrey, P., Ellison, S. L., Mendel, J. T., & Scudder, J. M. 2013, MNRAS, 433, L59
Powell, L. C., Bournaud, F., Chapon, D., & Teyssier, R. 2013, MNRAS, 434, 1028
Renaud, F., Bournaud, F., Kraljic, K., & Duc, P.-A. 2014, MNRAS, 442, L33
Richtler, T., Husemann, B., Hilker, M., et al. 2017, A&A, 601, A28
Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, MNRAS, 449, 49
Rothberg, B., & Joseph, R. D. 2004, AJ, 128, 2098
Saintonge, A., Tacconi, L. J., Fabello, S., et al. 2012, ApJ, 758, 73
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Sánchez-Blázquez, P., Ocvirk, P., Gibson, B. K., Pérez, I., & Peletier, R. F. 2011, MNRAS, 415, 709
Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2012, A&A, 538, A8
Sanders, D. B., Soifer, B. T., Elias, J. H., Neugebauer, G., & Matthews, K. 1988, ApJ, 328, L35
Sarzi, M., Falcón-Barroso, J., Davies, R. L., et al. 2006, MNRAS, 366, 1151
Sarzi, M., Shields, J. C., Schawinski, K., et al. 2010, MNRAS, 402, 2187
Schawinski, K., Thomas, D., Sarzi, M., et al. 2007, MNRAS, 382, 1415
Schawinski, K., Dowlin, N., Thomas, D., Urry, C. M., & Edmondson, E. 2010, ApJ, 714, L108

Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, MNRAS, 440, 889
Schawinski, K., Koss, M., Berney, S., & Sartori, L. F. 2015, MNRAS, 451, 2517
Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
Schechter, P. L., & Gunn, J. E. 1979, ApJ, 229, 472
Schweizer, F. 1982, ApJ, 252, 455
Schweizer, F., & Seitzer, P. 1998, AJ, 116, 2206
Schweizer, F., Seitzer, P., Kelson, D. D., Villanueva, E. V., & Walth, G. L. 2013, ApJ, 773, 148
Scudder, J. M., Ellison, S. L., Torrey, P., Patton, D. R., & Mendel, J. T. 2012, MNRAS, 426, 549
Sellwood, J. A., & Spekkens, K. 2015, ArXiv e-prints [arXiv:1509.07120]
Serra, P., Oosterloo, T., Morganti, R., et al. 2012, MNRAS, 422, 1835
Singh, R., van de Ven, G., Jahnke, K., et al. 2013, A&A, 558, A43
Smethurst, R. J., Masters, K. L., Lintott, C. J., et al. 2018, MNRAS, 473, 2679
Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, MNRAS, 391, 481
Spekkens, K., & Sellwood, J. A. 2007, ApJ, 664, 204
Springel, V., Di Matteo, T., & Hernquist, L. 2005a, ApJ, 620, L79
Springel, V., White, S. D. M., Jenkins, A., et al. 2005b, Nature, 435, 629
Stern, J., & Laor, A. 2012, MNRAS, 426, 2703
Takeuchi, T. T., Buat, V., Heinis, S., et al. 2010, A&A, 514, A4
Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, MNRAS, 418, 1587
Teyssier, R., Chapon, D., & Bournaud, F. 2010, ApJ, 720, L149
Torrey, P., Cox, T. J., Kewley, L., & Hernquist, L. 2012, ApJ, 746, 108
Tsatsi, A., Lyubenova, M., van de Ven, G., et al. 2017, A&A, 606, A62
Ueda, J., Iono, D., Petitpas, G., et al. 2012, ApJ, 745, 65
Ueda, J., Iono, D., Yun, M. S., et al. 2014, ApJS, 214, 1
van de Sande, J., Bland-Hawthorn, J., Brough, S., et al. 2017, MNRAS, 472, 1272
Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J., et al. 2010, MNRAS, 404, 1639
Veale, M., Ma, C.-P., Greene, J. E., et al. 2017, MNRAS, 471, 1428
Vogelsberger, M., Genel, S., Springel, V., et al. 2014a, Nature, 509, 177
Vogelsberger, M., Genel, S., Springel, V., et al. 2014b, MNRAS, 444, 1518
Wagner, S. J., Bender, R., & Moellenhoff, C. 1988, A&A, 195, L5
Walcher, C. J., Coelho, P. R. T., Gallazzi, A., et al. 2015, A&A, 582, A46
Wang, Z., Schweizer, F., & Scoville, N. Z. 1992, ApJ, 396, 510
Weigel, A. K., Schawinski, K., Caplar, N., et al. 2017, ApJ, 845, 145
Whitmore, B. C., Chandar, R., Schweizer, F., et al. 2010, AJ, 140, 75
Wild, V., Heckman, T., & Charlot, S. 2010, MNRAS, 405, 933
Wild, V., Rosales-Ortega, F., Falcón-Barroso, J., et al. 2014, A&A, 567, A132
Wong, K. C., Blanton, M. R., Burles, S. M., et al. 2011, ApJ, 728, 119
Woo, J.-H., Cho, H., Husemann, B., et al. 2014, MNRAS, 437, 32
Yamamura, I., Makiuti, S., Ikeda, N., et al. 2010, VizieR Online Data Catalog, II/298
Young, L. M., Bureau, M., Davis, T. A., et al. 2011, MNRAS, 414, 940

SUPPLEMENTARY 2 | A METHOD TO DISTINGUISH QUIESCENT AND DUSTY STAR-FORMING GALAXIES WITH MACHINE LEARNING

Includes the following:

A method to distinguish quiescent and dusty star-forming galaxies with machine learning

Published in The Astrophysical Journal, 2020, 891, 136

C. L. Steinhardt, J. R. Weaver, J. Maxfield, I. Davidzon, A. L. Faisst, D. Masters, M. Schemel, and S. Toft

The Astrophysical Journal, 891:136 (12pp), 2020 March 10 https://doi.org/10.3847/1538-4357/ab76be
© 2020. The American Astronomical Society. All rights reserved.

A Method to Distinguish Quiescent and Dusty Star-forming Galaxies with Machine Learning

Charles L. Steinhardt[1,2], John R. Weaver[1,2], Jack Maxfield[1,2,3], Iary Davidzon[3,4], Andreas L. Faisst[3,4], Dan Masters[3,5], Madeline Schemel[2,3], and Sune Toft[1,2]

[1] Cosmic Dawn Center (DAWN), Denmark; steinhardt@nbi.ku.dk
[2] Niels Bohr Institute, Lyngbyvej 2, DK-2100 København Ø, Denmark
[3] California Institute of Technology, 1216 East California Boulevard, Pasadena, CA 91125, USA
[4] Infrared Processing and Analysis Center, 1216 East California Boulevard, Pasadena, CA 91125, USA
[5] Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

Received 2019 June 13; revised 2020 February 12; accepted 2020 February 13; published 2020 March 13

Abstract

Large photometric surveys provide a rich source of observations of quiescent galaxies, including a surprisingly large population at $z > 1$. However, identifying large, but clean, samples of quiescent galaxies has proven difficult because of their near-degeneracy with interlopers such as dusty, star-forming galaxies. We describe a new technique for selecting quiescent galaxies based upon t-distributed stochastic neighbor embedding (t-SNE), an unsupervised machine-learning algorithm for dimensionality reduction. This t-SNE selection provides an improvement both over *UVJ*, removing interlopers that otherwise would pass color selection, and over photometric template fitting, more strongly toward high redshift. Due to the similarity between the colors of high- and low-redshift quiescent galaxies, under our assumptions, t-SNE outperforms template fitting in 63% of trials at redshifts where a large training sample already exists. It also may be able to select quiescent galaxies more efficiently at higher redshifts than the training sample.

Unified Astronomy Thesaurus concepts: Astronomy data analysis (1858); Computational astronomy (293); Galaxy quenching (2040); Quenched galaxies (2016); Galaxy classification systems (582); Star formation (1569)

1. Introduction

Although not highly celebrated, perhaps the most influential discovery in the history of modern astronomy has been gradually finding that galaxies do not take on arbitrary properties, spanning the entire range of theoretically possible spectra. As a result, it has been possible to produce meaningful surveys of faint galaxies using photometry, with only a very limited amount of information about the full spectral energy distribution (SED). Most commonly, a series of templates (Bruzual & Charlot 2003; Maraston et al. 2009; Brown et al. 2014) are fit to photometric colors with one of several competing techniques (see Arnouts et al. 1999; Brammer et al. 2008; Kriek et al. 2009) in order to produce a best-fit set of parameters.

Fundamentally, the goal of photometry is to map observed colors to galaxy properties. The validity of this technique therefore requires two additional assumptions. First, the mapping between observed colors and galaxy properties must be surjective, i.e., any specific combination of colors must only be produced by one set of galaxy properties. Otherwise, the colors are insufficient to break degeneracies between different possible galaxy models. Second, due to the complexity of calculating synthetic templates, current codes use a precompiled library of discrete models. Therefore, it is also necessary to assume that similar colors map to similar properties, to the point that it is possible to interpolate between nearby points with a known mapping.

Interpolation presents a considerable challenge, because there are often ∼10 galaxy parameters that one would like to fit, and this produces too large of a search space. Fortunately, we have discovered a series of scaling and other relations between observed galaxy parameters, including the "fundamental plane" (Gudehus 1973; Pahre et al. 1998; Bernardi et al. 2003) between radius, velocity dispersion, and surface brightness, the "star-forming main sequence" (Brinchmann et al. 2004; Noeske et al. 2007; Peng et al. 2010; Speagle et al. 2014), and a similar sequence for quasar accretion (Steinhardt & Elvis 2010, 2011). Because galaxies do not span the entirety of this ∼10-dimensional space, it is natural to consider first mapping to a smaller space that can be entirely searched, then running similar algorithms.

Previous work has shown that dimensionality reduction via a self-organizing map (SOM; Kohonen 1982) can be used to map photometry to a two-dimensional space suitable for redshift determination (Masters et al. 2015; Hemmati et al. 2019). The SOM spreads objects out approximately equally, dedicating more cells to more common types of objects. In this work, we use a related technique, t-Distributed Stochastic Neighbor Embedding (t-SNE; van der Maaten & Hinton 2008; Van Der Maaten 2014), which similarly produces a map with reduced dimension, but will produce a sparser mapping in an attempt to preserve structure and relative distance.

It might be hoped that combining such a map with observed spectroscopic redshifts will provide the basis for unsupervised machine-learning-derived photometric redshifts. In practice, redshifts determined by the SOM may produce a lower bias, suitable for several applications to *Euclid* (Massey et al. 2013; Masters et al. 2015, 2017, 2019; Hemmati et al. 2019) and other upcoming surveys. For relatively common objects where high-quality training data is available, it is also possible to directly use machine learning to model other galaxy parameters (Krakowski et al. 2016; Siudek et al. 2018; Davidzon et al. 2019). However, at present, photometric redshifts derived from template fitting remain competitive with those from unsupervised machine learning and for exotic outliers with few counterparts in a training sample are typically superior (Masters et al. 2015; Hemmati et al. 2019). In effect, at this point, there is more information in the theoretical templates derived from current models than there is from observed spectroscopic redshifts for rare but well-understood objects. However, it has been possible to apply machine-learning techniques to a wide variety of astronomical problems for which observations indeed provide more information than theoretical models. Recent work has included the use of t-SNE to derive stellar chemical abundances (Anders et al. 2018) and spectral information (Traven et al. 2017), as well as the use of Convolutional Neural Networks to measure galaxy morphology (Dieleman et al. 2015; Domínguez Sánchez et al. 2018; Hausen & Robertson 2019; Cheng et al. 2020) and shape (Ribli et al. 2019), perform light profile fitting (Tuccillo et al. 2018), identify mergers (Bottrell et al. 2019), estimate cluster masses (Ho et al. 2019), and classify supernovae (Muthukrishna et al. 2019).

For the same reason, unsupervised machine learning should be expected to perform better than template fitting for objects that are poorly modeled by current theory. In this work, we use a combination of t-SNE and current observations to develop a new selection for high-redshift, quiescent galaxies. In Section 2, the underlying assumptions and a more formal definition of a quiescent estimator is given. The new estimator is described in Section 3. We then evaluate the success of this algorithm in Section 4.

2. Quiescence Estimators and Varying Assumptions

In this work, we develop a method based upon the unsupervised machine-learning algorithm t-SNE to select quiescent galaxies from photometric surveys. Let $\{\boldsymbol{x}_i \in \mathcal{R}^k\}$ be the set of photometric measurements provided by the survey, where each individual $\boldsymbol{x}_i$ has k components, representing one object in k bands. Each specific galaxy might be quiescent; denote this by $q_i \in \{0,1\}$. A method of quiescent galaxy detection consists of an estimator[6] $Q(\{\boldsymbol{x}_i\})$ for $\{q_i\}$. Although the true $q_i \in \{0,1\}$, some estimators instead return a probability of quiescence $q_i \in [0,1]$.

Current estimators use information learned from past analysis of galaxies to produce a static mapping $Q(\boldsymbol{x}_i) \rightarrow q_i$. The most commonly used example is a color–color selection such as a *UVJ* diagram (Strateva et al. 2001; Baldry et al. 2004; Wuyts et al. 2007; Williams et al. 2009; Muzzin et al. 2013; van der Wel et al. 2014; Leja et al. 2019), which is a mapping from only three of the k bands (chosen or adjusted to approximate U, V, and J bands in rest-frame colors) to a quiescence estimator, selecting a region of ratios between adjacent bands that is populated primarily by quiescent galaxies. A far more complex mapping is produced by photometric template fitting (Arnouts et al. 1999; Brammer et al. 2008; Kriek et al. 2009), which produces spectra for various combinations of model parameters and uses them to find a best-fit spectrum for each $\boldsymbol{x}_i$. The model parameters corresponding to that spectrum determine the specific star-formation rate (sSFR; or SFR per unit stellar mass), and applying a threshold to the sSFR produces a quiescent estimator. Both methods assume that our knowledge of *typical* galaxies is sufficient to produce a mapping Q that will be valid for *all* galaxies, and in the case of template fitting, that knowledge includes a mapping from model parameters to spectrum through stellar population synthesis (see Conroy et al. 2009; Conroy 2013).

The unsupervised machine-learning method developed here does not require stellar population synthesis or any other astrophysics but rather attempts to let the data itself determine the proper quiescence estimators. Conceptually, the success of such an algorithm relies upon three key assumptions:

1. There is a surjection $Q(\boldsymbol{x}_i) \rightarrow q_i$ from photometric measurements in the survey bands to quiescence. Unlike the methods described above, *it is not necessary to know any properties of that surjection* but merely that one exists.
2. If x_1 and x_2 are nearby in the k-dimensional photometric color space, it is very likely that $q_1 = q_2$.
3. There is a mapping $T : \mathcal{R}^k \rightarrow \mathcal{R}^2$ from the k-dimensional vector space to a lower-dimensional space, in this case two-dimensional, in which the previous two properties continue to hold.

The first assumption is straightforward, as it is minimum necessary foundation for photometry as a valid astronomical technique. In regions where the second assumption does not hold, it means that quiescent and non-quiescent galaxies will be very nearly degenerate. One example would be the near-degeneracy between age and extinction in photometric template fitting (Gallazzi et al. 2005). Thus, photometry will be insufficient to determine quiescence with high certainty for such galaxies regardless of the methodology employed.

The third assumption is necessary because unsupervised machine learning requires a training set, and galaxy photometry is sparse in $\mathcal{R}^k$. Thus, we first map galaxies T: $\{\boldsymbol{x}_i\} \rightarrow \{\boldsymbol{x}_i'\}$ into $\mathcal{R}^2$, then find a surjection $Q(\{\boldsymbol{x}_i\}) \rightarrow \{q_i\}$ that produces a quiescence estimator from a reduced space in which individual galaxies are likely to have many close neighbors.

If all of these assumptions hold, it is then possible to produce an estimator for whether any specific galaxy should be classified as quiescent by looking at neighboring galaxies for which q has been well measured and letting those neighbors vote. We evaluate the correctness of this estimator in Section 4, finding that, on average, it is more successful than template fitting, more strongly so toward high redshift. However, there is ultimately value in using both approaches, one which is based upon astrophysical knowledge about the physics of galaxy evolution and another which is entirely ignorant of that physics and only given examples of quiescent and non-quiescent galaxies, letting the data alone predict quiescence.

3. Using t-SNE to Select Quiescent Galaxies

The most successful existing methods for photometric quiescent galaxy selection are variations on color–color selection, in which galaxies are mapped into a two-dimensional space based upon the two slopes described by a set of three specific rest-frame photometric bands, then a region is identified that is populated primarily by quiescent galaxies (Williams et al. 2009). Every galaxy within that region is selected as quiescent, and the remainder as selected as non-quiescent.

The machine-learning method here, although is it constructed from a very different toolkit, is essentially an improved version of this familiar color–color selection. It similarly finds

[6] Note that this is a more general definition than required for previous estimators, which can be given an individual $\boldsymbol{x}_i$ and produce $Q(\boldsymbol{x}_i) = q_i$. The machine-learning classification developed here can only operate on the entire set $\{\boldsymbol{x}_i\}$ simultaneously, and is meaningless for individual objects.

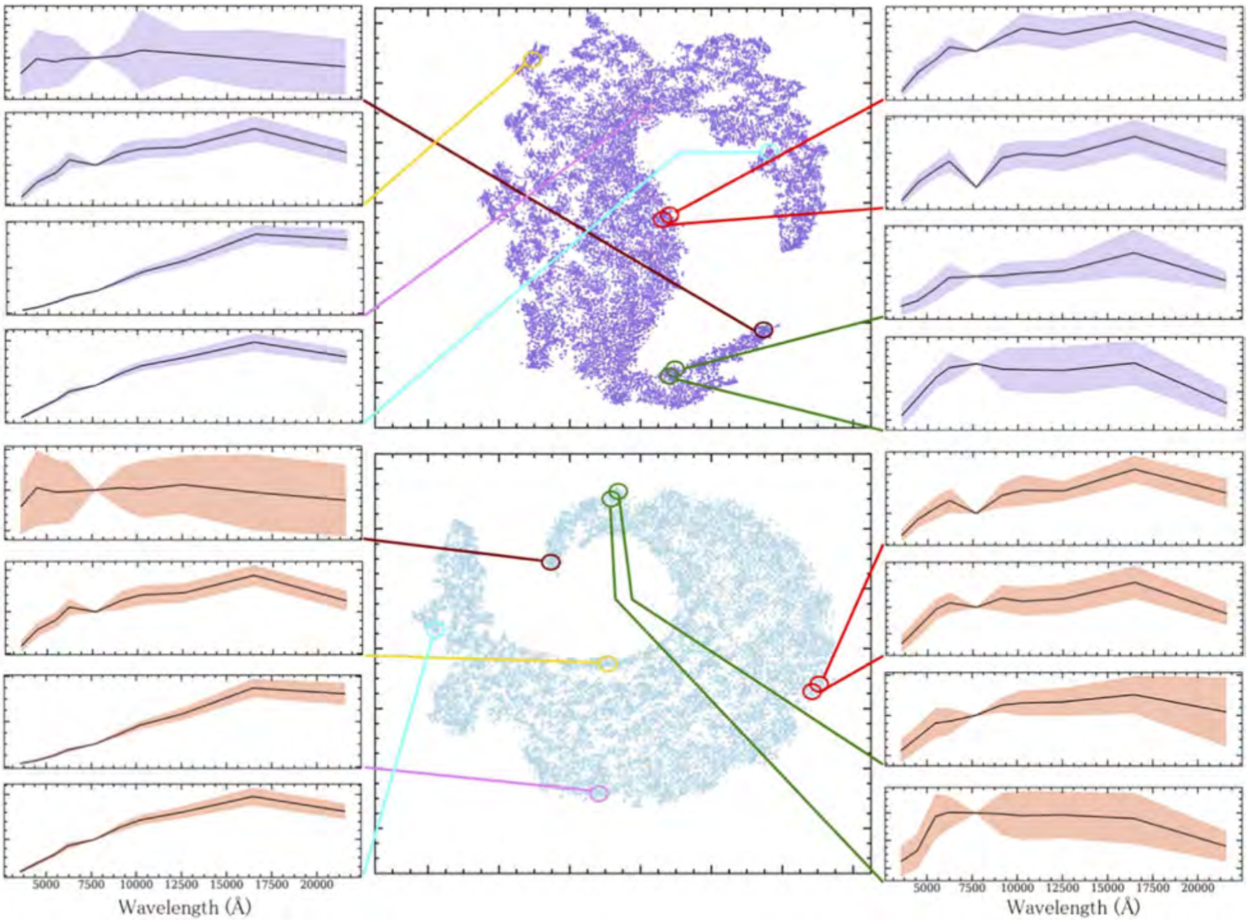

Figure 1. (a) t-SNE map of all galaxies at $1.0 < z < 1.1$ in the ULTRAVISTA catalog (McCracken et al. 2012) for which MIPS coverage is available. A narrow redshift range is necessary, since otherwise, the primary structure shown in the map would indicate redshift. Galaxies with similar SEDs end up in neighboring locations, and galaxies with dissimilar SEDs end up far apart. Mean SEDs with 1σ envelopes are calculated from the black circles indicated. (b) A second t-SNE map for the same catalog, produced with a different random seed and initial ordering of the sample. The eight sample SEDs shown are the same for both maps, with corresponding colors identifying the same SEDs. In each case, the range of SEDs within the indicated circle is shown, with all SEDs normalized to a common z-band flux.

regions populated primarily by quiescent galaxies, but using all available bands, so that all available information can be used in the selection. Further, quiescent galaxies are selected from any of a potentially large number of tiny regions. As a result, with a sufficiently high-quality training sample, it becomes possible to exclude dusty star-forming galaxies that are *nearly* identical, but not entirely identical, to quiescent galaxies and avoid selecting them. Naturally, if star-forming galaxies are truly degenerate with quiescent galaxies, no algorithm can distinguish them, in which case the first assumption (Section 2) would be violated because $Q(\boldsymbol{x}_i) \rightarrow q_i$ would no longer be surjective. Similarly, if, in practice, measurement uncertainties are large enough to partially or fully restore degeneracies, it will again become impossible to select all quiescent galaxies but no interlopers.

The algorithm we select, t-SNE (van der Maaten & Hinton 2008; Van Der Maaten 2014), is an unsupervised machine-learning algorithm for dimensionality reduction designed for the visualization of high-dimensional data sets. We first use t-SNE to produce a map T in which galaxies with similar photometric SEDs are placed in nearby locations, while galaxies with dissimilar photometric SEDs are further away (Figure 1(a)). For the figures shown in this work, maps were constructed in rest-frame magnitude space, and different features of the data set would be revealed using different units or distance metrics.

Note that the two coordinates produced by t-SNE are arbitrary and do not represent any sort of basis for the space. A galaxy further in the x-direction has not become more “x-like,” but is merely dissimilar from galaxies further to the left. Further, t-SNE is a randomized algorithm, and running it on the same data set with different initial conditions will produce the same topology but a different map (Figure 1(b)).

To this point, the t-SNE map has been produced without any direct information about galaxy quiescence or any other astronomical properties. However, because photometry is an indicator of astrophysical quantities, the map resulting from arranging galaxies based upon their photometry has also arranged them by these quantities. It is therefore possible to predict the properties such as stellar mass that would be found by photometric template fitting without the need to run template fitting codes, merely by looking at the results of running those codes on nearby galaxies (Figure 2, left panel).

Most importantly for our purposes, the same is very likely true for quiescence. It is certainly true that a t-SNE map can predict whether template fitting will determine that a galaxy is quiescent, but this has limited utility. After all, the primary advantage of unsupervised machine learning is avoiding the

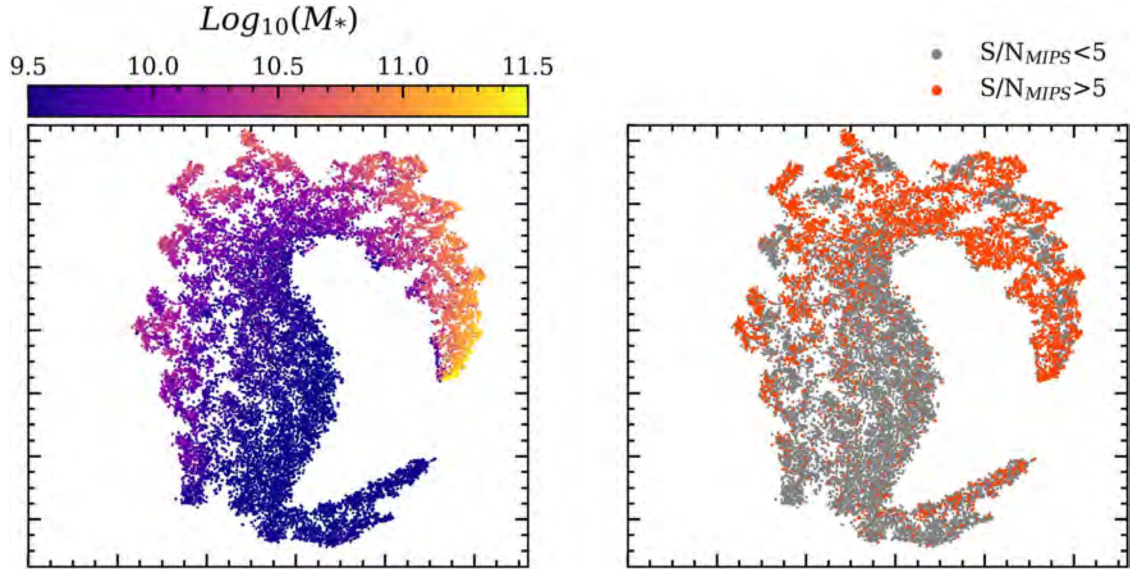

Figure 2. (Left panel) Best-fit stellar mass for the galaxies in the ULTRAVISTA sample from Figure 1. Galaxies with similar best-fit stellar masses cluster together. Thus, it is possible to predict the photometric template fitting-determined stellar mass of a galaxy without actually running those template fitting codes on the object in question by instead looking at the masses of its neighbors. The same is true of many other galaxy properties that currently require template fitting. (Right panel) Summary of MIPS detections for the low-z sample. Sources with credible detections ($S/N > 5$; orange) are clustered with respect to those without credible detections ($S/N < 5$; gray).

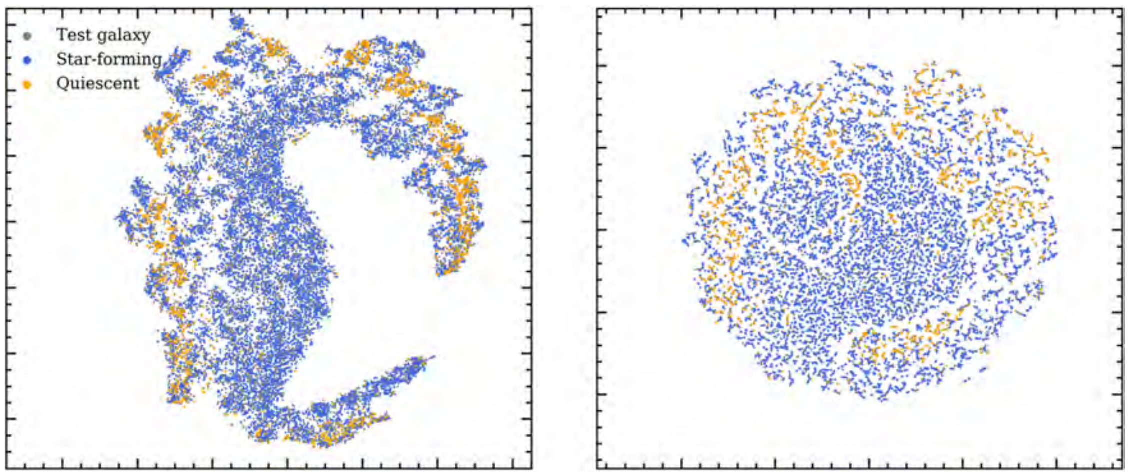

Figure 3. (Left panel) Mapping produced by t-SNE for a combination of training (blue-labeled star-forming and orange-labeled quiescent) and test (gray) sets, both taken from $z \sim 1$. The quiescent galaxies cluster within the map, implying that galaxies in the test sample should be labeled as star-forming or quiescent in a similar manner. (Right panel) A second mapping produced with a smaller value of perplexity, which effectively optimizes the map for finding more local rather than global structures. Different types of structure will be produced by different choice of t-SNE hyperparameters, so using t-SNE for predicting astronomical properties requires careful tuning. Maps are produced with perplexity hyperparameters of 30 (left) and 7 (right).

need to make the assumptions required to produce templates. Instead, we train our predictor on the most successful technique applied to the ULTRAVISTA catalog, combining a successful (rest-frame) *UVJ* color selection with a non-detection in the *Spitzer Space Telescope* MIPS 24 μm band (Rieke et al. 2004), an independent indicator of hot dust and likely also a high star-formation rate (Rieke et al. 2009).

Galaxies with MIPS detections lie primarily within a set of contiguous regions on the t-SNE map (Figure 2). Further, nearly every galaxy within that region has an MIPS detection, while nearly every galaxy lying elsewhere does not. Similarly, galaxies classified as quiescent (*UVJ*-selected but no MIPS detection) lie within distinct regions (Figure 3). Therefore, it is natural to produce an estimator that examines the objects within a small neighborhood on the map with known classification as quiescent or non-quiescent and lets them vote on whether a new object is likely to be quiescent and on the confidence in that prediction. If the number of such neighborhoods were known a priori, that information could be used to produce a further improved predictor (see Turner et al. 2019). However, avoiding the imposition of this condition allows t-SNE to attempt to detect of all types of quiescent galaxies and all types of dusty star-forming galaxies without imposing any prior expectations on how many distinct types exist.

3.1. Information Used for t-SNE Mapping

The Laigle et al. (2016) catalog includes over 30 bands from NUV out NIR , some which overlap with alternative bands at similar wavelength. However, due to differences in on-sky coverage, many of those bands are only available for a fraction of the catalog. Because t-SNE relies on Euclidean distances, which cannot be calculated for vectors of different dimension, we restrict our analysis to u, B, r, i^+, z^{++}, Y, J, H, K_s bands, which are available as statistically significant detections for most of the catalog. As described in Table 1 of Laigle et al. (2016), their 3σ depths vary across the field and range from 23.4 (K_s) to 27.0 (B). Due to completeness, a mass cut is made between $8.5 < \log_{10}(M_*) < 11.5$. Based on photometric redshift solutions, we isolate a low-redshift sample within $0.9 < z < 1.1$ containing 19,774 galaxies (17% quiescent) and a high-redshift sample within $1.9 < z < 2.1$ containing 7524 galaxies (6% quiescent). When training and test samples are made, in every case, objects are randomly selected from within the relevant redshift boundaries.

3.2. Definition of "True" Quiescent Galaxies

What any learning algorithm attempts to do is predict how new data would have been labeled if it were part of that training sample. A significant issue in training t-SNE selection is therefore that the labeling, in the ideal case, will ultimately be exactly as good as the training sample. Thus, using t-SNE to select quiescent galaxies relies on a good selection of the ground truth of quiescence for the training sample.

Ideally, this would be done spectroscopically, using specific lines as tracers of SFR and stellar mass. However, spectroscopy is available for only about 1% of the Laigle et al. (2016) catalog. Further, objects with spectroscopic follow up have often been targeted because of specific photometric properties, so that a complete spectroscopic sample would be even smaller. It is therefore necessary to rely on photometry to select "true" quiescent galaxies in the training and test samples. Several methods have been proposed for using photometry to select quiescent galaxies, including the new method developed here. Of these, each has known flaws:

1. Static color selection, such as *UVJ*, provides a fairly good proxy for quiescence while minimizing contamination. However, some objects with very high SFR (and MIPS flux) will pass the *UVJ* selection (Figure 7(a) and related discussion). Quiescent galaxies can also lie outside the *UVJ* region (Domínguez Sánchez et al. 2016). A possible solution to these limitations is replacing the rest-frame U band with near-UV (NUV) and increase the color leverage by using NUV $- r$ versus $r - K$ (or $r - J$, see, respectively, Arnouts et al. 2013; Ilbert et al. 2013).
2. MIPS is only successful at selecting the galaxies that are brightest at 24 μm (COSMOS does not have uniform MIPS coverage, so the detection threshold varies significantly). Therefore, many star-forming galaxies, including ones that, at low mass but high redshift, lie *above* the main sequence, will not be detected by MIPS.
   At $z \sim 2$, this potential mischaracterization of lower-SFR (likely lower mass) star-forming galaxies will be significantly worse than at $z \sim 1$. Thus, there is a strong redshift dependence in this definition of a true quiescent galaxy. This effect will therefore underestimate the quality of a selection trained at one redshift but tested at another (Section 4.2).
3. Photometric template fitting and the resulting sSFR attempts to calculate a quantity that can be most directly interpreted as "true" quiescence. However, photometric template fitting is also known to produce significant errors in SFR (see Laigle et al. 2019) and, therefore, even less reliable sSFR, dividing that by an estimated stellar mass. Often, there is insufficient multiwavelength coverage to use any other method (except for *UVJ* or some other two-color selection), and photometric template fitting is used by default. However, as shown in Section 4, these flaws in using template fitting to estimate SFR mean that best-fit sSFR is not a particularly good proxy for quiescence, even though true sSFR would be.

Given the available options, in this work, a combination of *UVJ* and MIPS selection is used to define ground truth for quiescence. When a sufficiently large spectroscopic sample can be produced, it would be ideal to then recalibrate the t-SNE predictor based upon this improved definition of true quiescence.

3.3. Training and Test Samples

Since t-SNE has no knowledge of astronomy, it must be provided with a training set consisting of identified quiescent and non-quiescent galaxies in order to produce a predictor. Unlike algorithms such as a self-organizing map (Kohonen 1982; Masters et al. 2015), t-SNE does not produce a static transformation from the higher-dimensional space to the lower-dimensional one but rather produces a mapping that extremizes a global penalty score for the entire sample. Thus, adding a new object requires recalculating the entire t-SNE map, and may alter the positions of every object. This means that t-SNE is a poor choice for real-time analysis, because it is not possible to precompute a static surjection $Q(\boldsymbol{x})$. However, if the training sample is already large compared with the test sample, the entire test sample can be processed in approximately the same time as one object.

Therefore, the estimator developed here first uses t-SNE to arrange the union of both training and test samples. Galaxies described by their rest-frame photometry in the higher-dimension space are mapped by t-SNE with a perplexity of 30 over 1000 iterations. Once converged, labels are applied to the training set to denote quiescence (Figure 3). It should be noted that the choice of perplexity and other settings (conventionally called hyperparameters in order to distinguish them from parameters, which instead belong to the model) makes a substantial difference and is part of the t-SNE tuning process. Perplexity formally is defined in terms of the Shannon entropy of the system (van der Maaten & Hinton 2008), so that higher values produce a wider search, which results in more strongly weighting global structure, and lower values similarly reveal more local structure. The choice of perplexity is effectively a prediction of the number of neighbors that should be used in determining the properties of a galaxy and, therefore, depends not only on the underlying distribution of galaxy properties but also on sample size and selection.

Because the quiescent galaxies cluster within the training sample, if galaxies in the test sample are drawn from a very similar population, they should have the same label as their neighbors (Figure 3). This is also a corollary of the three assumptions listed in Section 2. Objects in the test sample are therefore classified as quiescent when the quiescent fraction of m neighboring training galaxies is $f_Q > f_{\min}$.

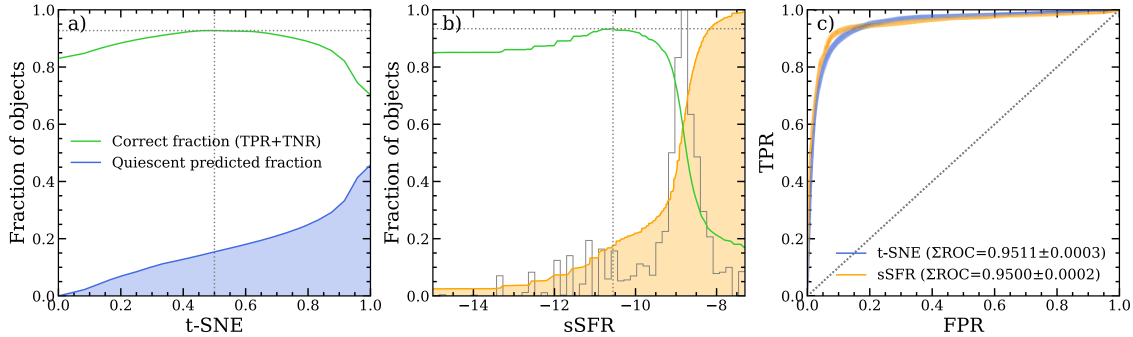

Figure 4. Panels (a)–(b) Correct (sum of of true positive and true negative) prediction fractions for (a) t-SNE and (b) photometric template fitting (green) as a function of the threshold used for both training and test sample drawn from the same $0.9 < z_{\rm phot} < 1.1$ ULTRAVISTA catalog. The sSFR distribution (black, panel (b)) and cumulative distribution (blue/orange in panels (a)/(b)) are also shown, as well as the maximum TPR+TNR achieved and their corresponding thresholds (gray dotted lines). (c) Receiver operating characteristic (ROC) curve for both methods over 10,000 random draws of training and test samples. t-SNE (blue) outperforms photometric template fitting (orange) in 63% of trials. With different tuning, t-SNE selection could be constructed to outperform sSFR selection either for high-quantity samples (illustrated here) or instead for high-quality samples but not for both use cases simultaneously. A typical threshold used to identify quiescent galaxies based on $\log_{10}$(sSFR) is −10.

A natural choice of $f_{\rm min}$ might seem to be 0.5, letting a majority of nearby objects determines the label. However, in practice, the optimal choice depends upon how many objects are scattered via measurement errors to incorrectly be close neighbors. Therefore, the optimal choice of $f_{\rm min}$ depends upon a combination of uncertainties and the underlying true fraction of quiescent galaxies in the training sample. The choice of $f_{\rm min}$ also depends upon the desired relationship between quantity and quality in the resulting catalog; a higher value of $f_{\rm min}$ will result in fewer false positives but more false negatives (e.g., Figure 4). It is common to see this tradeoff referred to in machine-learning literature as one between precision (true positives/total positives), a measure of sample quality, and either recall or sensitivity, two terms referring to the true positive rate, which is a measure of sample quantity.

4. Comparison with Template Fitting

We construct a series of tests to compare the effectiveness of t-SNE in selecting "quiescent" galaxies (defined here as *UVJ* selected but not MIPS detected) with that of photometric template fitting. Since our fiducial definition of a quiescent galaxy includes a two-color (*UVJ*) selection, it is difficult to test t-SNE against two-color selection. However, in Section 4.3, we evaluate whether t-SNE is more likely to discard dusty star-forming galaxies with MIPS detections than a standard two-color selection.

In all cases, the comparison is done on the ULTRAVISTA photometric catalog (McCracken et al. 2012), with data drawn from the Laigle et al. (2016) catalog providing improved reductions and additional ancillary data. Currently, the most reliable method for differentiating between dusty, star-forming interlopers and bona fide quiescent galaxies in ULTRAVISTA requires additional observations, using *Spitzer*/MIPS to search for 24 μm emission characteristic of hot dust, and therefore a dusty star-forming galaxy. A successful classification is therefore defined as predicting correctly from u, B, r, i^+, z^{++}, Y, J, H, K_s bands whether a galaxy will both be selected by (rest-frame) *UVJ* and have no discernible MIPS 24 μm detection with a S/N > 5, at which threshold it would instead be considered a dusty star-forming galaxy.

The sample used is described in detail in Laigle et al. (2016), which provides robust multiwavelength rest-frame magnitudes. Galaxies are then cross-matched with an FIR/mm catalog (Sanders et al. 2007; Jin et al. 2018) over the same footprint. Objects with MIPS SNR < 3 are considered too faint for an MIPS non-detection to be a reliable indicator of quiescence and are not included in the sample. Although color selection and template fitting are static methods, t-SNE requires a training sample. Therefore, for t-SNE, the catalog must be divided into a training sample and test sample.

Template fitting and t-SNE selection are compared in their ability to solve two different problems. First, to examine the case of a well-understood domain, the $0.9 < z < 1.1$ population is divided up into equal-sized, disjoint training and test samples. t-SNE is given the entire training sample (in principle, so are other methods, but they do not change based upon new information) and a list of which training objects have been identified as quiescent. The methods are also given the entire test sample and its NIR photometry, but no information about which test objects have MIPS detections. Each method produces a catalog of test objects classified as quiescent and is evaluated on both false-positive and false negative rates.

Second, template fitting and t-SNE selection are also evaluated on their ability to determine which galaxies are quiescent in an unexplored domain. The training sample consists of the entire $0.9 < z < 1.1$ catalog, but the test sample consists of the $1.9 < z < 2.1$ catalog. Because the $z \sim 2$ galaxy population is not identical to that at $z \sim 1$ (see Speagle et al. 2014 for star-forming galaxies and van der Wel et al. 2014 for quiescent ones), this presents a far more difficult problem for machine learning, which has no knowledge of astronomy or any expected redshift evolution. Methods are given rest-frame colors from the Laigle et al. (2016) catalog, so any errors in photometric redshift determination for the test sample will result in all three methods making predictions from incorrect inputs at the same rate. Errors in photometric redshift determination for the training sample will degrade the efficiency of t-SNE but not the other methods.

4.1. Comparison of Estimators in a Well-explored Domain

We first consider these estimators in a domain that is already well-explored. Both the training and test samples are drawn from the same catalog at the same photometric redshift of $0.9 < z_{\rm phot} < 1.1$, with "true" quiescent galaxies defined as in Section 3.2. A typical use case might be producing a catalog of quiescent galaxies for a large photometric survey with limited

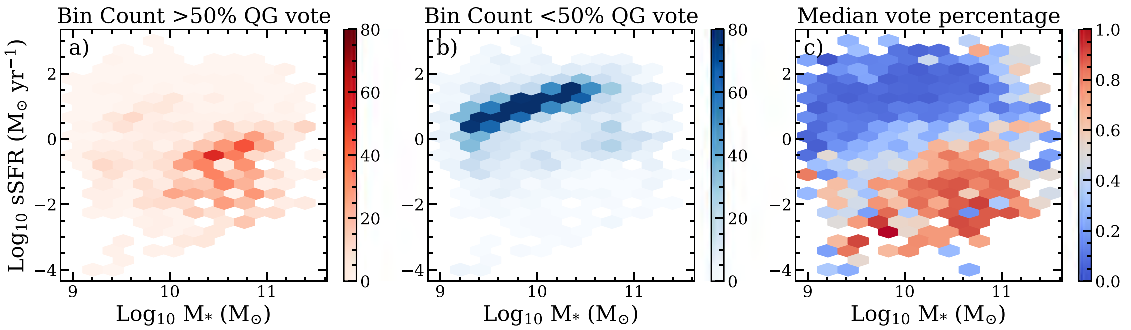

Figure 5. Panels (a)–(b) Distribution of stellar mass and star-formation rate for galaxies selected as quiescent or non-quiescent using the t-SNE threshold indicated in Figure 4(a), with shading indicating the number of objects selected. Panel (c) Median quiescence score using t-SNE selection for the test sample, with a score of 1.0 indicating the highest certainty that an object is quiescent. Objects with the most uncertain classification lie either near the boundary between star-forming and quiescent, as well as at low mass, where fainter galaxies have higher measurement uncertainties.

spectroscopy. In that case, the additional bands or spectroscopic follow up sufficient to produce confirmed quiescent galaxies would only exist for a small fraction of the full catalog, but could provide a high-quality training sample.

Color selection has no free parameters and, therefore, is completely defined and produces a fixed error rate, both for false positives and false negatives. The other two methods do have tunable parameters. For photometric template fitting, a large number of selections (choice of templates, grid spacing, other fit parameters and hints) were made in the ULTRA-VISTA catalog used (described in detail in Laigle et al. 2016) and cannot be altered for this test. However, the choice of sSFR threshold used to determine quiescence is an additional parameter, and a higher threshold will reduce both true and false positives for quiescence (Figure 4(b)).

For t-SNE, there are similarly several hyperparameters required to produce an estimator. The most significant for producing a map is perplexity, which governs the relative importance of local neighbors compared with more distant ones. Once the map is produced, the definition of close neighbors, how many training sample neighbors are chosen, and threshold fraction of quiescent training sample neighbors are additional choices. These choices must be made differently for every specific use case, because two identical galaxies will end up at different distances on the t-SNE map depending upon properties of other galaxies, sample selection, sample size, perplexity, and t-SNE grid size. For this specific test, we experimentally determined that perplexity 30 maximized ΣROC (defined below) for the figures shown. Then, setting the threshold for the required fraction of quiescent neighbors to label a galaxy as quiescent produces a similar tradeoff between true and false-positive rates as for template fitting (Figure 4(a)). The total accuracy shown for each estimator is a combination of the true quiescent galaxy fraction and a false-positive fraction, (TP+TN)/total. The same accuracy could be produced from, e.g., more true quiescent galaxies with more false positives or fewer of each. Thus, the maximum accuracy will lie at the cutoff where a marginal change in cutoff will result in an equal change in both true and false positives.

The appropriate tool for comparing these estimators is a receiver operating characteristic (ROC) curve, a tool that is common for assessing the quality of medical diagnostics producing a Boolean answer (Albeck & Børgesen 1990; Baker 2003; Fawcett 2006). A random estimator containing no information can be produced lying anywhere along the dashed diagonal, e.g., randomly selecting 40% of galaxies to be quiescent will result in a 40% of quiescent galaxies selected as quiescent (true positive) as well as 40% of star-forming galaxies selected as quiescent (false positive). The best estimators have ROC curves lying as close as possible to the top left, corresponding to 100% true positives with no false positives.

We find that the t-SNE ROC curve is comparable to the template fitting ROC curve (Figure 4(c)), with a slightly different shape. For some desired true positive rates, a t-SNE method will produce a corresponding sample with fewer false positives, but for some true positive rates, template fitting is more successful. With different hyperparameters, t-SNE selection could be constructed to outperform sSFR selection either for any specific true positive rate desired (illustrated for high-quantity samples in Figure 4(c)) but not for all true positive rates simultaneously and with a lower overall success rate.

A related statistic[7] is $\Sigma\mathrm{ROC} \equiv \int_0^1 \mathrm{ROC(TPR)}\,d\mathrm{FPR}$, commonly used in machine learning to consider the quality of estimators across all possible thresholds. Prior to selecting a threshold, both template fitting (using sSFR) and t-SNE (using fraction of quiescent neighbors) in a ranked ordering of the entire test sample by likelihood that the object is quiescent. The selection of a threshold then divides the sample into two groups, labeling the more likely group with $q = 1$ and the remainder as $q = 0$. The ΣROC corresponds to the probability that a randomly selected quiescent galaxy is ranked higher than a randomly selected star-forming galaxy (see Bradley 1997) and produces a similar result the Wilcoxon-Mann-Whitney rank sum test (Wilcoxon 1945; Mann & Whitney 1947). A random estimator will rank the quiescent galaxy higher half of the time, for a ΣROC of 0.5.

The t-SNE ΣROC is 0.951 and the photometric template fitting ΣROC selecting quiescent galaxies as those with a low best-fit sSFR is 0.950. Both are substantially better than random, and t-SNE outperforms sSFR in 63% of trials from random draws of training and test samples. The objects for which the t-SNE estimator produces an incorrect classification, as might be expected, are primarily those in two categories: (1) galaxies with high measurement uncertainties, typically fainter and, thus, lower-mass galaxies; and (2) galaxies near the boundary between star-forming and quiescent, for which even

[7] In previous literature, in medical diagnostics and, later, machine learning, this is described as "area under the curve," or AUC, rather than in terms of the integral. Since astronomers should be more comfortable with a description in terms of calculus, we choose to do so instead.

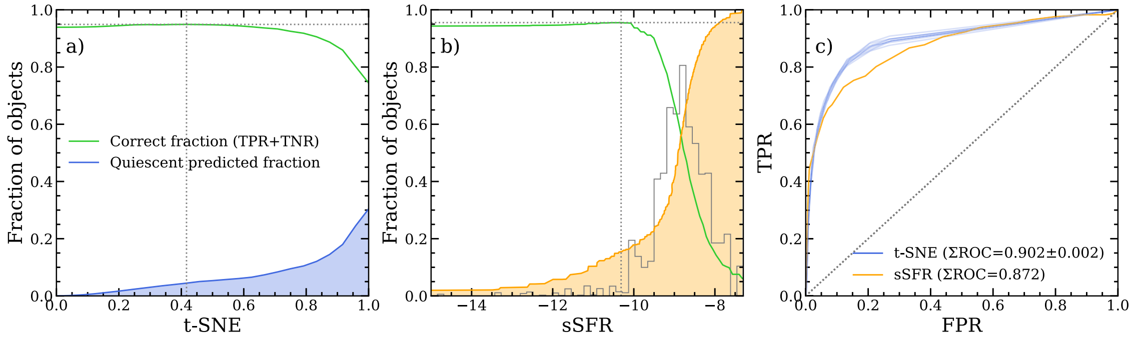

Figure 6. (a)–(b) Correct (sum of of true positive and true negative) prediction fractions for (a) t-SNE and (b) photometric template fitting (green) as a function of the threshold used for a training sample drawn from the $0.9 < z < 1.1$ ULTRAVISTA rest-frame catalog and test sample drawn at $1.9 < z < 2.1$ from the same catalog. The sSFR distribution (black, panel (b)) and cumulative distribution (blue/orange in panels (a)/(b)) are also shown, as well as the maximum TPR+TNR achieved and their corresponding thresholds (gray dotted lines). (c) Receiver operating characteristic (ROC) curve for both methods. t-SNE (blue) is dominant over photometric template fitting (orange) for any choice of optimal true positive rate, with different choices of initial conditions for t-SNE algorithm having negligible impact on the ROC curve.

small differences in SFR would change their classification (Figure 5).

In Figure 4, we have attempted to optimize the choice of hyperparameters for t-SNE to produce a clean separation. With a different choice of hyperparameters, the quality of the estimator would be decreased, much as it would be if inferior templates were used for photometric template fitting. The ideal choice of hyperparameters depends upon the sample, but can be estimated using the following heuristics:

1. The perplexity should represent the number of neighbors that should be considered informative as to the average galaxy. Thus, if a survey is doubled in an area with otherwise identical detection properties, the optimal perplexity will also approximately double.
2. The neighborhood radius for determining which objects are sufficiently close on t-SNE map to vote on quiescence should be chosen so that the typical number of objects within that radius matches the perplexity.
3. The threshold can then be set to different values depending on the relative importance of completeness and quality. A threshold set near the total fraction of quiescent galaxies, or equivalently, the voting score that would result if neighbors are chosen randomly from the entire sample rather than from the t-SNE map, will typically maximize the sum of TPR and TNR.

A reasonable interpretation is that the training sample used by t-SNE has comparable, but slightly more, information about quiescent galaxies than the models used to produce templates. Depending upon which regime is most useful, that information can be used to make either slightly better high-quality or high-completeness samples, but t-SNE must be tuned for that specific purpose.

4.2. Comparison of Estimators in a Novel Domain

We now consider these estimators in a domain that is primarily unexplored. Both the training and test samples are drawn from the same rest-frame ULTRAVISTA catalog, but the training sample is drawn at a photometric redshift of $0.9 < z_{phot} < 1.1$ and the test sample is drawn at $1.9 < z_{phot} < 2.1$. This test is designed to explore the utility of these estimators in finding quiescent galaxies in a new, higher-redshift regime, on the basis of what has been learned about them at lower redshifts. As a result, although hyperparameters of the t-SNE map were carefully chosen in order to optimize the lower-redshift quiescent galaxy selection, we have then frozen them rather than selecting new hyperparameters for the $z \sim 2$ test in order to provide a fair test of a truly unexplored domain in which there is no training sample to calibrate against.

All estimators considered are predicated on the idea that high-redshift quiescent galaxies look sufficiently similar to low-redshift counterparts that it will be possible to recognize them without high-redshift examples. For photometric template fitting, the assumption is that quiescent galaxies at different redshifts might possibly have dissimilar properties apart from their low star-formation rates, but that they will be driven by similar astrophysics. Therefore, the same stellar population synthesis codes, extinction laws, etc., can be used to produce valid templates. The other two estimators make no direct assumptions about the underlying astrophysics and instead assume that high-redshift quiescent galaxies will have similar SEDs, in a holistic way for t-SNE and in specific bands for *UVJ* selection.

As in Section 4.1, photometric template fitting and t-SNE both require a choice of threshold and can be assessed through analyzing the true and false-positive rates as a function of threshold (Figure 6). The optimal threshold is lower here, because the overall fraction of true quiescent galaxies is lower in the sample. In general, the optimal threshold will lie close to the point at which neighbors are consistent with being randomly drawn from the full sample, including both quiescent and non-quiescent galaxies. This also means that the optimal threshold will depend not just on redshift, but also the detection limit. Even in a novel domain, some prior expectation about the fraction of quiescent galaxies is required for optimal t-SNE selection.

The resulting ROC curve indicates that t-SNE is a dominant selection mechanism, and allows substantially larger high-quality samples. For this unexplored domain, the t-SNE ΣROC is 0.915 and the template fitting ΣROC is 0.871. In this case, t-SNE even with fixed parameters is dominant over sSFR selection; for any choice of ideal true positive rate, t-SNE selection will have a lower false-positive rate than photometric template fitting (Figure 4(c)). The information advantage in favor of color space rather than model space is now sufficiently large that the t-SNE hyperparameters no longer need to be tuned for a specific use case in order to substantially outperform template fitting.

Both values of ΣROC are substantially lower than in the well-sampled domain explored at Section 4.1. This is due to a combination of several factors. For both methods, $z \sim 2$ galaxies are generally fainter, and therefore more poorly measured than at $z \sim 1$. Additionally, both methods in different ways assume that high-redshift galaxies look like those at low redshift. For t-SNE, the assumption is indeed that they are identical in (rest-frame) color space. For template fitting, this same assumption is instead expressed in the choice of templates, with the assumption that physical models developed using a combination of theory and observed spectra of more local quiescent galaxies continue to describe those at higher redshift. Finding that both methods are broadly successful at $z \sim 2$ confirms that the colors of quiescent galaxies perhaps change slightly but do not change substantially between $z = 1$ and $z = 2$.

4.3. Improvement over Two-color Selection

For completeness, it is also important to show that t-SNE indeed provides an improvement over color–color selection. It should be expected that t-SNE will perform better than, e.g., *UVJ* selection because t-SNE is using more information and is allowed to construct its quiescent locus by combining many small regions of quiescent galaxies rather than one, continuous region. Thus, it should be hoped that t-SNE will be able to remove many of the *UVJ*-selected objects with MIPS detections.

However, it is perhaps not obvious a priori that these advantages must provide a significant improvement. The colors of quiescent galaxies are dominated by very old stellar populations, and the color of an aging stellar population changes quickly for young populations but is nearly constant for very old ones. As a result, the observed SEDs of quiescent galaxies look very similar to each other, and predicting the full SED of a quiescent galaxy from a small number of bands is easier than doing so for a star-forming galaxy. Thus, the additional bands might be mostly redundant information. If the quiescent locus is tight, there may be negligible benefit to instead describing it as the union of many small neighborhoods and omitting galaxies that lie in between. Thus, it is necessary to confirm that t-SNE selection truly outperforms *UVJ*.

Since the quiescent galaxy training sample already includes a two-color (*UVJ*) selection, t-SNE is trained in part with the goal of reproducing that selection. Indeed, if *UVJ* selection were the only criterion used, t-SNE would reproduce it almost exactly, since *UVJ* selection also corresponds to a region of the full color space described by all bands. Instead, we test whether t-SNE provides an improvement over *UVJ* by examining the interlopers with MIPS detections indicating hot dust and, therefore, that the object is not a quiescent galaxy.

For *UVJ* selection, these interlopers are well mixed with true quiescent galaxies, so that it would not be possible to produce a high-quality sample with a more restrictive cut in *UVJ* (Figure 7(a)). However, t-SNE is able to find some regions that do have lower interloper densities, so that it can produce a higher-quality sample (Figure 7(b)).

Selecting quiescent galaxies from a catalog of only *UVJ*-selected galaxies is a much more difficult problem than selecting them from the full catalog. The full catalog contains many galaxies that can be very easily rejected as quiescent candidates, whereas the interlopers that pass *UVJ* selection look far more similar to true quiescent galaxies. Thus, t-SNE is less effective at this separation, with a ΣROC of 0.762, compared with $\Sigma\text{ROC} = 0.952$ for the full catalog. However, this is still a significant improvement upon *UVJ* selection; it means that given the (rest-frame) optical photometry for a true quiescent galaxy and a dusty star-forming galaxy that also passed *UVJ* selection and asked to select which one is truly quiescent, t-SNE will make the correct selection 76.2% of the time. Template fitting also provides an improvement over *UVJ* selection, but a smaller one, with a ΣROC of 0.664. Both are an improvement upon *UVJ* selection alone, which makes no distinction between objects that pass its selection and, thus, would correctly identify the MIPS detection exactly 50% of the time.

Clearly the best estimator of whether MIPS will detect an object is based on MIPS observations, and in practice neither t-SNE nor sSFR estimators would be used instead. However, coverage is not always available, or is not available at sufficient depth, when determining quiescent candidates from photometry. Indeed, the rationale behind using template fitting is that it should be possible to predict 24 μm flux from optical and near-infrared photometry. These results show that t-SNE is a much better estimator of 24 μm flux than photometric template fitting.

5. Results and Discussion

In this work, we develop a new, machine-learning-based method for selecting quiescent galaxies from optical photometry. This method provides an improved catalog over two-color selection by identifying and rejecting many (but not all) of the dusty star-forming galaxies that contaminate two-color samples. The choice of t-SNE threshold also provides a tradeoff between sample size and sample quality that does not exist in two-color classification.

The efficiency of this t-SNE is compared with photometric template fitting, which similarly allows a tradeoff between quality and quantity, with more restrictive cuts on best-fit sSFR or SFR providing a smaller but higher-quality sample. We find that in a well-explored domain, in which there is already a large training sample available at the same redshift, t-SNE outperforms photometric template fitting in 63% of trials. In a novel domain, using $z \sim 1$ galaxies as a training sample to select $z \sim 2$ quiescent candidates, t-SNE is dominant over template fitting, in that, for any choice of sample size, t-SNE will produce a sample with fewer interlopers.

It should also be stressed that t-SNE was not evaluated here under optimal conditions. The mapping was based upon a limited number of bands, with IR bands available to template fitting excluded from the t-SNE mapping (since t-SNE was asked in part to make a prediction of 24 μm luminosity), whereas other surveys often provide more information. Further, because t-SNE needs to compare objects on a similar vector basis, it was necessary to provide it with rest-frame optical measurements. For objects with catastrophic errors in photometric redshifts, t-SNE simply provided incorrect data, so that some of the objects for which template fitting fails automatically failed t-SNE as well. Still, the result is that t-SNE provides an improvement over both color selection and photometric template fitting under essentially all conditions for which it was tested.

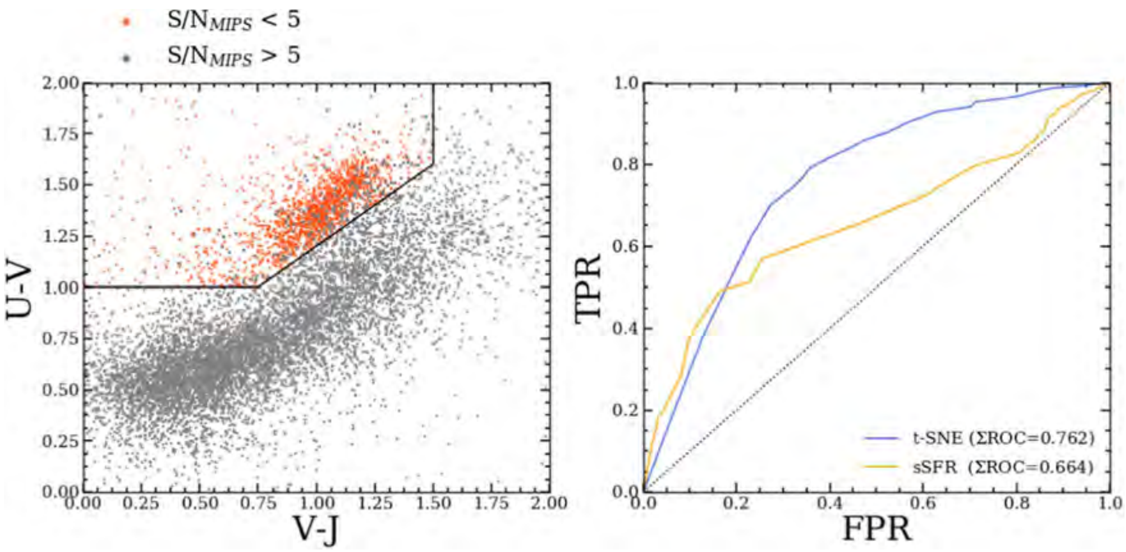

Figure 7. (a) *UVJ* diagram for ULTRAVISTA galaxies at $z \sim 1$. Galaxies with MIPS detections (gray) are scattered throughout the region and well mixed with the remainder of the population (red). (b) Receiver operating characteristic (ROC) curve for t-SNE when selecting non-MIPS detected galaxies from a sample that was previously *UVJ*-selected. The t-SNE estimator is still trained on a training sample including all types of galaxies, but the ROC curve is only calculated from *UVJ*-selected galaxies. Given a random pair of *UVJ*-selected galaxies, one with an MIPS detection and one with no MIPS detection, t-SNE will identify which one has the MIPS detection 76.2% of the time. Template fitting also provides an improvement, but a smaller one, correctly identifying the MIPS detection for 66.4% of possible pairs. For comparison, *UVJ* alone makes no distinction between objects that pass its selection, so it would correctly identify the MIPS detection exactly 50% of the time and would correspond to the dotted line.

An additional issue for both template fitting and t-SNE is computational complexity. Two-color selection is very simple (although in practice, two-color selection is done in the rest frame and may require template fitting to be performed prior to selection), and adding objects to an existing sample is also quick. Photometric template fitting is a slow process, with the quality of the fit often determined by the limited set of templates that can be considered given available computing time. However, adding one new object to an existing catalog only requires running template fitting on that one object, which can typically be done in minutes at reasonable quality. Because t-SNE does not produce a static map, the addition of even one new object requires reoptimizing the entire map, which is the time-consuming step. Formally, t-SNE runs in $\mathcal{O}(n^2)$, and the runtime is already prohibitive for the $\sim 10^4$ objects in the samples shown here. Approximations (e.g., Barnes–Hut; Barnes & Hut 1986) exist to produce reductions to $d \leqslant$ three-dimensions in $\mathcal{O}(n \log n)$ and were used here to produce a runtime of a few minutes. However, this would still be an issue for the entire $>10^6$ object COSMOS catalog or upcoming catalogs from LSST, *Euclid*, etc. For such surveys, an alternative algorithm such as the self-organizing map (Masters et al. 2015), which produces a static mapping is likely a better choice if real-time decision-making is required for an individual object.

5.1. Where Does the Improvement Come from?

Understanding this improvement requires evaluating how much information is being used and how useful that information is for quiescent selection. The improvement in t-SNE over color selection is straightforward, since t-SNE is using all of the available information instead of only some bands and doing so in a more flexible manner. The ability of t-SNE to do this indicates that the three assumptions in Section 2 are generally true, which should not be surprising because these same assumptions are required for photometry to be capable of building a catalog.

Photometric template fitting and t-SNE are much less similar, since they arise from two different ways of modeling galaxies. Photometry describes galaxies in physical parameter space, using astrophysical modeling to transform those parameters into colors on the basis of the knowledge that human astronomers have built up about galaxy evolution. t-SNE models galaxies purely empirically in color space, with no astrophysical knowledge used at any stage of the process, so that the information depends purely on the sample size.

The greater success of t-SNE means that at present, our astrophysical models provide less information about quiescent galaxies and dusty star-forming interlopers than photometric catalogs, with the gap increasing toward high redshift. This is the opposite of the current situation for redshift determination, for which astrophysical models provide more information and templates produce better fits. Whether this remains true going forward will depend upon the rate at which models improve compared with the rate at which catalogs become larger and of higher quality.

On the other hand, it should be noted that the performance of the two-color selection can be improved without resorting to machine learning. In case of a rich multiwavelength baseline as in COSMOS, the astrophysical model limitation mentioned above can be minimized by estimating rest-frame magnitudes from the nearest observer's frame band (see, e.g., Davidzon et al. 2017). Moreover, defining the quiescent *locus* in the $\mathrm{NUV}-r$ versus $r-J$ diagram (NUVrJ) instead of *UVJ* dramatically reduces the contamination fraction, since $\mathrm{NUV}-r$ probe shorter star-formation timescales and it is more sensitive to fast quenching processes (see discussion in Moutard et al. 2016 and Valentino et al. 2020 for applications at higher redshift). To show that, we compare the NUVrJ catalog of quiescent galaxies provided by Laigle et al. (2016) with a fiducial t-SNE selection resulting from a threshold equal

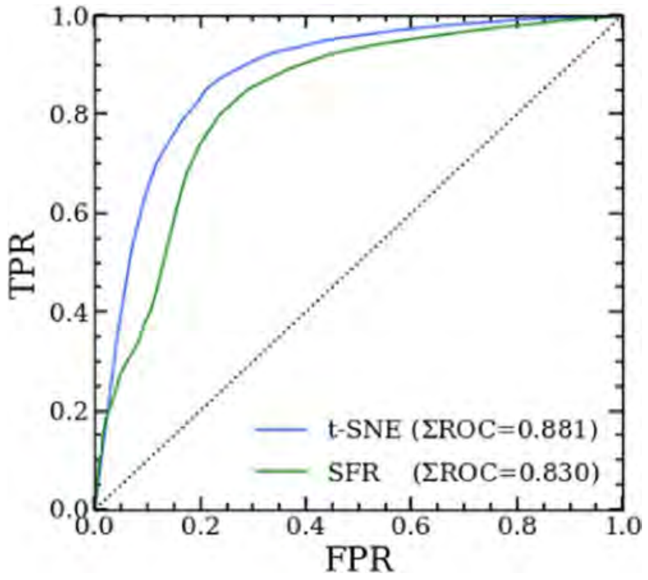

Figure 8. Receiver operating characteristic (ROC) curve for t-SNE selection and photometric template fitting's best-fit SFR for selecting MIPS detected galaxies from the full $0.9 < z < 1.1$ sample.

to 0.46, which maximizes TPR+TNR at $0.9 < z < 1.1$. For the same validation sample used above (1685 quiescent galaxies), the two methods have a similar fraction of interlopers (~22% in both cases) and a comparable level of completeness: NUVrJ recovers 84% of the "true" quiescent galaxies while t-SNE (in the 0.46 threshold configuration) about 79%.

It is perhaps more surprising that training t-SNE at $z \sim 1$ and predicting whether $z \sim 2$ galaxies are quiescent still outperforms template fitting. This means that at least in (rest-frame) color space, galaxies at $z \sim 2$ are nearly identical to those at $z \sim 1$. If this were to continue to hold at much higher redshifts, it means that t-SNE would be a good method for selecting high-redshift quiescent candidates from rest-frame optical data using, e.g., the *James Webb Space Telescope*. However, one should use considerable caution here: because it has no astrophysical knowledge, t-SNE is only capable of selecting high-redshift quiescent galaxies *which look like low-redshift examples*. If high-redshift quiescent galaxies have different astrophysical properties and, therefore, exhibit different colors, they will not be selected by t-SNE but could still be selected by template fitting if these new properties are well described by models. Similarly, if the low-redshift training sample definition of "true" quiescent galaxies is flawed or incomplete, then t-SNE will attempt to faithfully reproduce that flawed selection at high redshift.

5.2. Additional Considerations

It should be noted that, as suggested by Figure 2, t-SNE may be used to estimate many galaxy properties apart from quiescence. A full discussion of these possibilities is beyond the scope of this paper. However, as one example, t-SNE is used to estimate MIPS detection alone (for both quiescent and star-forming galaxies). In this way, we can evaluate the method against a quantity that has been measured directly from telescope images.

For MIPS detections as well, t-SNE is able to produce a meaningful predictor (Figure 8). Since an MIPS detection corresponds to a high SFR, not sSFR, it is compared against the set of galaxies with best-fit SFR above some threshold using photometric template fitting. The comparison has a simple qualitative purpose: the best-fit SFR is, by construction, a poorer predictor of 24 μm emission, as the latter also depends on the amount of dust (which span a large range in galaxies at a given SFR). The t-SNE ΣROC is 0.881 and the template fitting ΣROC is 0.830.

In any regime in which target galaxies look similar to an existing catalog of examples, dimensionality reduction provides an alternative selection mechanism and alternative method for determining physical properties. Because quiescent galaxies are all characterized by similar (very old) stellar populations with little active galactic nucleus contamination, their selection presents an ideal use case for these new methods, and at this point, dimensionality reduction provides superior classification to existing techniques.

It should be noted that it is not possible to provide a simple prescription of applying dimensionality reduction and subsequent selection to a new catalog. Rather, doing so effectively requires carefully tuning t-SNE hyperparameters to match the expected properties of that catalog. For example, perplexity needs to be tuned in order to ensure that the number of objects strongly influencing the locations on the t-SNE mapping matches the expected number of meaningful neighbors. Similarly, the choice of t-SNE threshold depends upon the expected fraction of true quiescent galaxies in the sample. Properly applying t-SNE to, e.g., CANDELS (Grogin et al. 2011; Koekemoer et al. 2011) will almost certainly yield an improved estimator, but additional optimization will be required to produce that estimator.

Finally, it should be stressed that other techniques within the family of machine-learning methods hold the possibility of substantial further improvement. It can be more difficult to understand where the improvement is coming from using t-SNE, but an initial exploration (beyond the scope of this current paper) suggests that it may be possible to produce a further improved estimator. The best methods for selecting quiescent galaxies in poorly explored domains such as at high redshift, contrary to conventional wisdom, might not rely on improved model making or on expensive observations of a few specimens. Instead, future photometric surveys will probe those domains with enough statistics so that the galaxy color space, albeit unclassified, might be analyzed by means of t-SNE or other manifold learning algorithms to identify galaxy classes with no need for templates.

The authors wish to thank Johann Bock Severin, Gabe Brammer, Beryl Hovis-Afflerbach, Adam Jermyn, Christian Kragh Jespersen, Vasily Kokorev, Allison Man, Georgios Magdis, and Jonas Vinther for useful discussions. C.L.S. and S. T. are supported by ERC grant 648179 "ConTExt." The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. J.M. is supported by the Jonathan Baker Excellence in Physics Fund.

ORCID iDs

Charles L. Steinhardt https://orcid.org/0000-0003-3780-6801
John R. Weaver https://orcid.org/0000-0003-1614-196X
Iary Davidzon https://orcid.org/0000-0002-2951-7519
Andreas L. Faisst https://orcid.org/0000-0002-9382-9832
Dan Masters https://orcid.org/0000-0001-5382-6138

Sune Toft https://orcid.org/0000-0003-3631-7176

References

Albeck, M., & Børgesen, S. 1990, Ugeskr. Laeg., 152, 1650
Anders, F., Chiappini, C., Santiago, B. X., et al. 2018, A&A, 619, A125
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Arnouts, S., Le Floc'h, E., Chevallard, J., et al. 2013, A&A, 558, A67
Baker, S. G. 2003, J. Natl. Cancer Inst., 95, 511
Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, ApJ, 600, 681
Barnes, J., & Hut, P. 1986, Natur, 324, 446
Bernardi, M., Sheth, R. K., Annis, J., et al. 2003, AJ, 125, 1866
Bottrell, C., Hani, M. H., Teimoorinia, H., et al. 2019, MNRAS, 490, 5390
Bradley, A. P. 1997, Pattern Recognit., 30, 1145
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Brown, M. J. I., Moustakas, J., Smith, J. D. T., et al. 2014, ApJS, 212, 18
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Cheng, T.-Y., Conselice, C. J., Aragón-Salamanca, A., et al. 2020, MNRAS, in press (doi:10.1093/mnras/staa501s)
Conroy, C. 2013, ARA&A, 51, 393
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, A&A, 605, A70
Davidzon, I., Laigle, C., Capak, P. L., et al. 2019, MNRAS, 489, 4817
Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441
Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, MNRAS, 476, 3661
Domínguez Sánchez, H., Pérez-González, P. G., Esquej, P., et al. 2016, MNRAS, 457, 3743
Fawcett, T. 2006, PaReL, 27, 861
Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Gudehus, D. H. 1973, AJ, 78, 583
Hausen, R., & Robertson, B. 2019, arXiv:1906.11248
Hemmati, S., Capak, P., Masters, D., et al. 2019, ApJ, 877, 117
Ho, M., Rau, M. M., Ntampaka, M., et al. 2019, ApJ, 887, 25
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
Jin, S., Daddi, E., Liu, D., et al. 2018, ApJ, 864, 56
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kohonen, T. 1982, Biol. Cybern., 43, 59
Krakowski, T., Małek, K., Bilicki, M., et al. 2016, A&A, 596, A39
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Laigle, C., Davidzon, I., Ilbert, O., et al. 2019, MNRAS, 486, 5104
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Leja, J., Tacchella, S., & Conroy, C. 2019, ApJL, 880, L9
Mann, H. B., & Whitney, D. R. 1947, Ann. Math. Stat., 18, 50
Maraston, C., Strömbäck, G., Thomas, D., Wake, D. A., & Nichol, R. C. 2009, MNRAS, 394, L107
Massey, R., Hoekstra, H., Kitching, T., et al. 2013, MNRAS, 429, 661
Masters, D., Capak, P., Stern, D., et al. 2015, ApJ, 813, 53
Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2017, ApJ, 841, 111
Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, ApJ, 877, 81
McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156
Moutard, T., Arnouts, S., Ilbert, O., et al. 2016, A&A, 590, A103
Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, PASP, 131, 118002
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJL, 660, L43
Pahre, M. A., Djorgovski, S. G., & de Carvalho, R. R. 1998, AJ, 116, 1591
Peng, Y.-j., Lilly, S. J., Kovač, K., et al. 2010, ApJ, 721, 193
Ribli, D., Dobos, L., & Csabai, I. 2019, MNRAS, 489, 4847
Rieke, G. H., Alonso-Herrero, A., Weiner, B. J., et al. 2009, ApJ, 692, 556
Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, ApJS, 154, 25
Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, ApJS, 172, 86
Siudek, M., Małek, K., Pollo, A., et al. 2018, A&A, 617, A70
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Steinhardt, C. L., & Elvis, M. 2010, MNRAS, 402, 2637
Steinhardt, C. L., & Elvis, M. 2011, MNRAS, 410, 201
Strateva, I., Ivezić, Ž, Knapp, G. R., et al. 2001, AJ, 122, 1861
Traven, G., Matijevič, G., Zwitter, T., et al. 2017, ApJS, 228, 24
Tuccillo, D., Huertas-Company, M., Decencière, E., et al. 2018, MNRAS, 475, 894
Turner, S., Kelvin, L. S., Baldry, I. K., et al. 2019, MNRAS, 482, 126
Valentino, F., Tanaka, M., Davidzon, I., et al. 2020, ApJ, 889, 93
Van Der Maaten, L. 2014, J. Mach. Learn. Res., 15, 3221
van der Maaten, L., & Hinton, G. 2008, J. Mach. Learn. Res., 9, 2579
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
Wilcoxon, F. 1945, Biom. Bull., 1, 80
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Wuyts, S., Labbé, I., Franx, M., et al. 2007, ApJ, 655, 51

SUPPLEMENTARY 3 | DUST AND THE INTRINSIC SPECTRAL INDEX OF QUASAR VARIATIONS

Includes the following:

Dust and the intrinsic spectral index of quasar variations: hints of finite stress at the innermost stable circular orbit

Published in Monthly Notices of the Royal Astronomical Society, 2022, stac248

J. R. Weaver & K. Horne

Monthly Notices of the Royal Astronomical Society
MNRAS 512, 899–916 (2022)
Advance Access publication 2022 January 29
https://doi.org/10.1093/mnras/stac248

Dust and the intrinsic spectral index of quasar variations: hints of finite stress at the innermost stable circular orbit

John R. Weaver[1,2,3]★ and Keith Horne[3]
[1]*Cosmic Dawn Center (DAWN)*
[2]*Niels Bohr Institute, University of Copenhagen, Jagtvej 128, DK-2200 Copenhagen, Denmark*
[3]*SUPA School of Physics & Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK*

Accepted 2022 January 24. Received 2022 January 24; in original form 2021 November 26

ABSTRACT

We present a study of 9 242 spectroscopically confirmed quasars with multiepoch *ugriz* photometry from the SDSS Southern Survey. By fitting a separable linear model to each quasar's spectral variations, we decompose their five-band spectral energy distributions into variable (disc) and non-variable (host galaxy) components. In modelling the disc spectra, we include attenuation by dust on the line of sight through the host galaxy to its nucleus. We consider five commonly used attenuation laws, and find that the best description is by dust similar to that of the Small Magellanic Cloud, inferring a lack of carbonaceous grains from the relatively weak 2175-Å absorption feature. We go on to construct a composite spectrum for the quasar variations spanning 700–8000 Å. By varying the assumed power-law $L_\nu \propto \nu^\alpha$ spectral slope, we find a best-fitting value $\alpha = 0.71 \pm 0.02$, excluding at high confidence the canonical $L_\nu \propto \nu^{1/3}$ prediction for a steady-state accretion disc with a $T \propto r^{-3/4}$ temperature profile. The bluer spectral index of the observed quasar variations instead supports the model of Agol & Krolik, and Mummery & Balbus, in which a steeper temperature profile, $T \propto r^{-7/8}$, develops as a result of finite magnetically induced stress at the innermost stable circular orbit extracting energy and angular momentum from the black hole spin.

Key words: accretion, accretion discs – methods: statistical.

1 INTRODUCTION

The optical identification of quasi-stellar objects (quasars hereafter) by Matthews & Sandage (1963) enabled, for the first time, studies of the distant universe at $z > 0.1$. Quasars are now recognized as high-luminosity examples of active galactic nuclei (AGN), powered by accretion on to a supermassive black hole (SMBH) (Lynden-Bell 1969; Shakura & Sunyaev 1973). The continuum variability of quasars, known soon after their discovery, allows us to peer directly into their central engines. Varying by 10–20 per cent over time-scales of months to years, the intrinsic variability of quasar continuum emission has long been theorized to be caused by changes in the environment close to the SMBH.

Quasar spectral energy distributions (SEDs) provide insight into their underlying physics. Spanning the full range from gamma-rays to radio, quasar SEDs exhibit both thermal (accretion disc, dust) and non-thermal (corona, jet) components. In the rest-frame UV–optical, thermal emission from the accretion disc is thought to manifest as the 'Big Blue Bump' (Shields 1978; Malkan & Sargent 1982), described by a sum of blackbody spectra over a range of temperatures from $\sim 10^3$ K for the cool outer edges of the disc to perhaps $\sim 10^5$ K near the innermost stable circular orbit (ISCO). A related feature is the 'Small Blue Bump', caused by closely packed FeII emission and the Balmer recombination continuum (Elvis, Wilkes & Tananbaum 1985; Wills, Netzer & Wills 1985)

For a geometrically thin steady-state accretion disc (Shakura & Sunyaev 1973), the effective temperature profile is $T_{\rm eff} \propto \left(M\,\dot{M}\right)^{1/4} r^{-3/4}$, where M is the black hole mass, $\dot{M}$ is the accretion rate, and r is the radial distance from the black hole. The corresponding spectrum, obtained by summing blackbody spectra weighted by solid angle, is $L_\nu \propto \left(M\,\dot{M}\right)^{2/3} \nu^{1/3}$. This power-law spectrum applies in the spectral range corresponding to the minimum and maximum disc temperatures, $k\,T_{\rm min} \ll h\,\nu \ll k\,T_{\rm max}$, where h and k are the Planck and Boltzmann constants. For a more general power-law temperature profile, $T \propto r^{-b}$, the disc spectrum is $L_\nu \propto \nu^\alpha$ with $\alpha = (3\,b - 2)/b$. Thus, measuring the disc's spectral slope α determines the power-law slope b of its temperature profile and tests the accretion disc theory. If the theoretical power-law slope, $\alpha = 1/3$ is confirmed, the results measure the product $M\,\dot{M}$. Moreover, since the disc spectrum scales with inclination angle i and luminosity distance D_L via $\cos i / D_L^2$, we may potentially be able to measure quasar luminosity distances.[1]

Several obstacles stand in the way of realizing these motivating goals. First, there may be significant extinction and reddening due to dust along the line of sight. Correcting for dust in our Milky Way galaxy is relatively straightforward (e.g. Schlegel, Finkbeiner & Davis 1998; Schlafly & Finkbeiner 2011). More difficult is to correct for the adverse effects of reddening caused by scattering and absorption by dust grains within the host galaxy, a complication shared by

[1]For Type 1 AGN ($i < 60^\circ$), the mean ± rms of $\cos i$ is $3/4 \pm \sqrt{1/48} = 0.75 \pm 0.14$.

★ E-mail: john.weaver.astro@gmail.com

© 2022 The Author(s)
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

the use of Type Ia supernovae as standard candles. Reddening along the line of sight presents a degeneracy since dust grains can redden quasar spectra with a wavelength dependence similar to the power law form expected for the disc spectrum. However, carbonaceous grains produce an absorption feature prominent in the dust extinction law observed in the Milky Way Galaxy (Nandy et al. 1975; Allen & Glass 1976; Seaton 1979) and the Large Magellanic Cloud (Fitzpatrick 1986). This absorption feature, described by a 'Drude' profile centred at 2175 Å (Fitzpatrick & Massa 1986; Draine & Malhotra 1993), can largely resolve the degeneracy, given sufficient spectral coverage. The 2175-Å absorption is weak or absent in other notable systems including the Small Magellanic Cloud (Gordon et al. 2003) and local starburst galaxies (Calzetti et al. 2000), which suffer from the full strength of this degeneracy. Moreover, it has been postulated that quasar dust may differ from the varieties studied closely in the local universe, for example, by lacking small grains that are evaporated by the quasar luminosity (e.g. Gaskell et al. 2004).

Secondly, the observed spectra of quasars are generally redder than the predicted disc spectrum, hinted at already by Sandage (1965). This is due in part to the comparably red starlight of the host galaxy. Quasars are often too distant to directly resolve the host galaxy, which means that their measurements are contaminated by host galaxy starlight captured within the aperture from which the photometry is performed. While this problem can be mitigated in small samples of nearby, resolved quasars where the host galaxy's light profile can be modelled, or extrapolated inward and subtracted from images, this approach fails for larger samples of more distant, unresolved quasars. The advent of large multiwavelength monitoring campaigns provides a viable workaround. Instead of attempting to subtract host galaxy contributions from imaging data, a time-series of images or spectra can be used to isolate the spectrum of the variable light arising from the central engine and the accretion disc. By this method, one can extract separate spectra for the variable accretion disc and the non-varying host galaxy components for a large number of unresolved quasars, provided multiwavelength photometric monitoring with sub-year cadence over a sufficiently long baseline to probe the variations.

Several successful and innovative campaigns have marked the previous twenty years of the study of AGN. Most recently, assemblages of multiepoch, multiwavelength photometric data sets have been observed by the Sloan Digital Sky Survey (SDSS; York et al. 2000) and more recently the Zwicky Transient Facility (ZTF; Bellm et al. 2018; Graham et al. 2019), which have enabled fundamentally new comparisons with theoretical models of accretion disc structure and behaviour with statistically significant samples. Additional observations from the Rubin Observatory's Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) will greatly increase both sample sizes and epoch baselines, expected to be underway in 2024. In addition, precise, spectroscopic monitoring campaigns such as the Sloan Reverberation Mapping Project (SDSS-RM; e.g. Shen et al. 2015) are providing valuable details on the variability of continuum and line emission, although with smaller samples.

Several techniques have been applied to interpret these photometric data sets. For example, MacLeod et al. (2012) employed a damped random walk model to describe the stochastic variations for an ensemble of $\sim 10^4$ quasars from SDSS, finding good agreement as a viable description of the optical continuum variability. For the same dataset, Kokubo et al. (2014) employed a 'flux-flux correlation' technique to derive the colour of the flux difference spectrum, which was used to infer an accretion disc spectral slope of $L_\nu \propto \nu^{1/3}$, consistent with standard steady-state accretion models. Parallel work has been undertaken with this and similar data sets to determine the extinction law most appropriate for quasars (e.g. Hopkins et al. 2004; Krawczyk et al. 2015), the results of which influence work on variability.

The objective of this work is to directly probe the accretion disc light and test theories of accretion physics. This will be accomplished as follows. In Section 2, we develop our method that leverages the source variability to isolate the accretion disc light. In Section 3, we apply this to a sample of 9 242 quasars observed with multiepoch multiwavelength photometry during the Sloan Digital Sky Survey Stripe 82 quasar campaign, including a de-reddening of the isolated accretion spectrum with five commonly used dust extinction laws. Section 4 presents the composite spectra. Our results are then discussed in Section 5 and our conclusions made in Section 6.

We adopt in this paper a concordance cosmological model with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{M} = 0.3$, and $\Omega_\Lambda = 0.7$. All magnitudes are in the AB$_\nu$ system (Oke 1974), for which a flux f_ν in mJy (10^{-26} erg cm^{-2}s^{-1}Hz^{-1}) corresponds to $\mathrm{AB}_\nu = 16.4 - 2.5\,\log_{10}(f_\nu/\mathrm{mJy})$.

2 ISOLATING THE ACCRETION DISC LIGHT

In this section, we describe our method using a separable linear model to fit the photometric variations of a quasar observed with multiwavelength photometry. The method is illustrated for a particular SDSS quasar in Fig. 1, which we discuss below as we outline the steps of the analysis.

A quasar is observed at N_t times t in N_λ photometric bands, each labelled by its pivot wavelength[2] λ. The observations at time t are considered simultaneous if measured within a time interval so short that changes in the state of the accretion disc can be neglected. For UV and optical observations of quasars, this typically means measurements on the same night, or even over a few nights.

We fit the observed spectral flux variations with the following separable linear model:

$$F(\lambda, t) = A(\lambda)\,X(t) + B(\lambda). \tag{1}$$

Here, $F(\lambda, t)$ can be F_ν or F_λ, or indeed any suitable flux unit. The dimensionless light-curve shape $X(t)$ is shifted to zero mean and scaled to unit root mean square (rms):

$$\langle X\rangle_t = 0\,, \qquad \left\langle X^2\right\rangle_t = 1, \tag{2}$$

where $\langle\cdot\rangle_t$ denotes a suitably weighted time average. With this normalization, the model's amplitude spectrum $A(\lambda)$ is the rms of the flux variations about the mean background spectrum $B(\lambda)$. This model has $2\,N_\lambda + N_t$ parameters, which are $A(\lambda)$, $B(\lambda)$, and $X(t)$. These are constrained by $N_t \times N_\lambda$ flux measurements plus two normalization constraints. For observations at a single wavelength, $N_\lambda = 1$, the model fits the N_t flux measurements exactly. For multiband observations, the model parameters are over constrained by the data, which permits optimizing the model parameters by fitting the data, and testing the validity of the model assumptions.

This model fitting is illustrated for a particular SDSS quasar in Fig. 1. The light curve of this quasar is sampled at $N_t = 62$ epochs with SEDs measured by $N_\lambda = 5$ bands, as shown in Fig. 1(b). Note here that the data follow a light-curve shape that is similar for all five bands. Fitting the model to these data, by minimizing χ^2, the corresponding set of linear equations is solved to determine the model parameters $A(\lambda)$, $B(\lambda)$, and $X(t)$, with corresponding uncertainties. This is done in practice by using iterated linear regression fits. Start by constructing an initial guess for $X(t)$, for example using one

[2] See A2.1 of Bessell & Murphy (2012) for details.

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

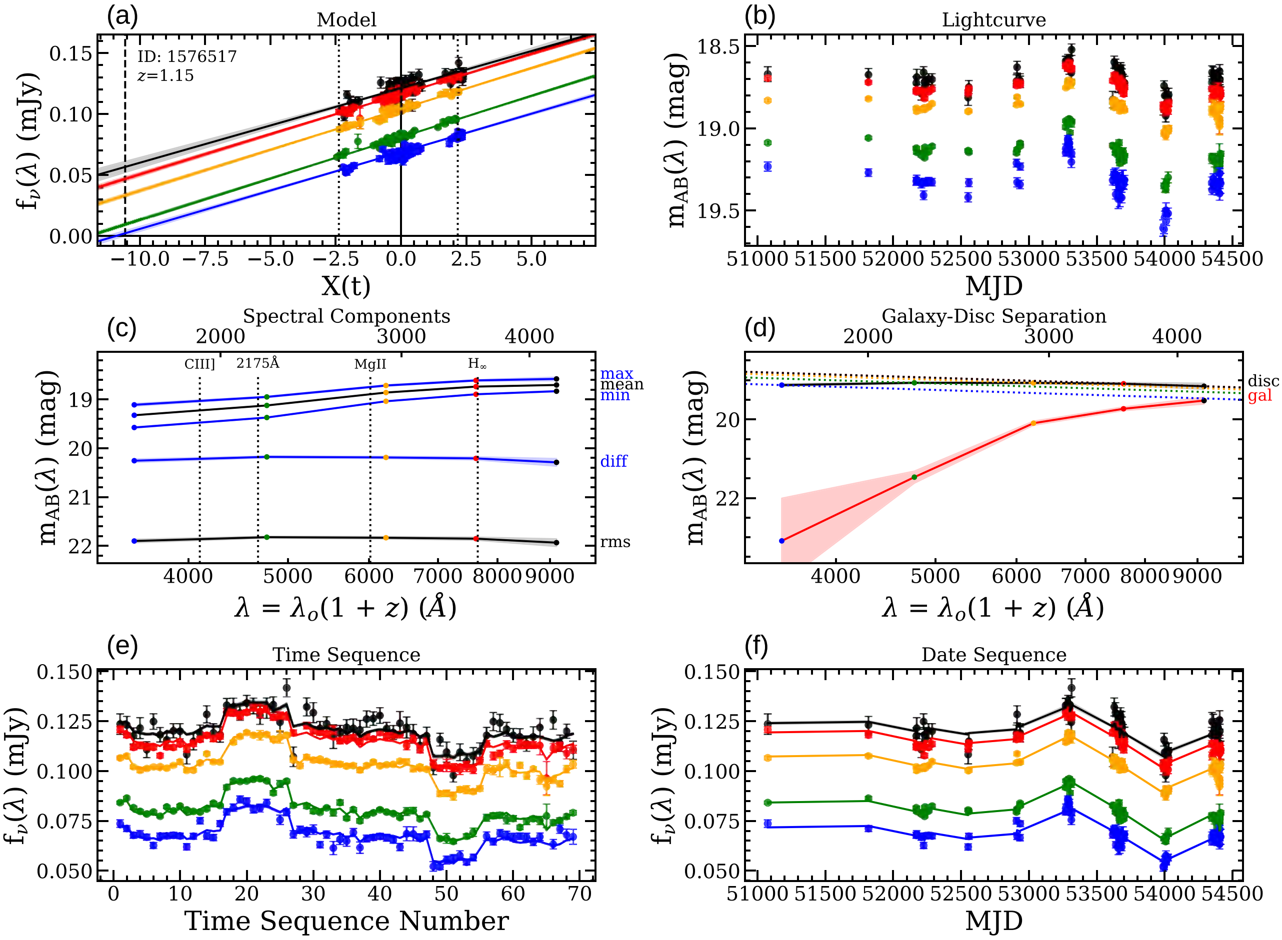

Figure 1. An example illustrating our light-curve decomposition method. Panel (b) shows the $N_\lambda = 5$-band *ugriz* light-curve data at $Nt = 62$ epochs for the SDSS quasar ID 1576517 at redshift $z = 1.15$. Panel (a) shows that the flux variations are well fitted by a linear model, $F(\lambda, t) = B(\lambda) + A(\lambda)\,X(t)$, where $B(\lambda)$ is the mean flux, $A(\lambda)$ is the rms amplitude of the flux variations, and $X(t)$ is the light-curve shape, assumed to be the same for all bands, normalized to $\langle X\rangle = 0$ and $\langle X^2\rangle = 1$. The maximum and minimum brightness states are indicated in Panel (a) by vertical dotted lines on either side of the mean state at $X = 0$. Extrapolating to fainter levels, the *u*-band flux becomes negative just below $X \approx -10.5$. Having thus turned off the disc, we attribute the extrapolated fluxes at the vertical dashed line, where *u* is 1σ above 0, to the non-variable host galaxy. Panel (c) shows the resulting five-band SEDs extracted at the maximum, minimum, rms, and mean states, as well as the bright – faint difference spectrum, with vertical dotted lines marking the wavelengths of relevant spectral features. Panel (d) shows the SEDs extracted for the disc (black) and host galaxy (red). Dotted lines show $L_\nu \propto \nu^{1/3}$ fixed to each band and coloured accordingly. Finally, the light-curve data and fitted model are shown versus time sequence number in Panel (e) and by date of observation in Panel (f). All error envelopes are shown at $\pm 1\sigma$.

of the observed light curves, suitably normalized. Then use two-parameter linear regression to find $A(\lambda)$ and $B(\lambda)$ assuming $X(t)$ is known. Next, revise $X(t)$ assuming $A(\lambda)$ and $B(\lambda)$ are known. Impose the normalization constraints on $X(t)$. Finally, iterate to convergence. Some care may be needed to identify and down-weight or reject significant outliers, using a robust procedure such as sigma clipping.

The two-parameter linear regression fits that determine $A(\lambda)$ and $B(\lambda)$, with $X(t)$ assumed to be known, are presented in Fig. 1(a). The fitted linear models are shown as solid coloured lines with $\pm 1\sigma$ envelopes. The flux data with error bars are plotted versus the dimensionless $X(t)$. This tracks the changing brightness of the quasar above and below the mean flux level. For each band, the slope in this diagram is the rms amplitude $A(\lambda)$ of the flux variations above and below the mean spectrum $B(\lambda)$ at $X(t) = 0$. Note here that the quasar variations are well described by linear flux variations. In particular, there is no evident curvature that could indicate a change in the disc spectrum between the faint and bright states. The shape of the light-curve $X(t)$ and comparison of the fitted model with the light-curve data, are examined in Fig. 1(e), where the flux data are plotted versus time sequence number, and in Fig. 1(f) versus observation date. Here, the light-curve shape is determined as a weighted average of the variations seen in all bands.

The fitted model can now be used to predict fluxes expected at different variability states $X(t)$. Given N_λ observed bands, the fitted model predicts the SED for any variability state $X(t)$. Meaningful extrapolation is possible, above and below the range sampled by the monitoring data, with the usual caveat that the extrapolated model becomes progressively uncertain. Fig. 1(c) presents the SED obtained for several indicative states. The mean spectrum, $B(\lambda)$, is the SED derived for the mean state of the system, at $X(t) = 0$. Above and below the mean SED are SEDs for the faintest and brightest observed states, at $X_{\min}$ and $X_{\max}$, respectively. These SEDs are relatively red, F_ν rising to longer wavelengths. However, the difference SED between the brightest and faintest states, evaluated for $\Delta X = X_{\max} - X_{\min}$, and the SED of the rms variations, for $\Delta X = 1$, are comparably blue. Such quasar variations are often described as 'bluer when brighter'. However, the linearity seen in Fig. 1(a) shows that this is not due to

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

the disc spectrum becoming bluer when brighter, but rather to the relatively red (host galaxy) spectrum becoming dominant as light from the relatively blue disc dims.

Our linear model fit to the spectral variations determines the disc and host galaxy SEDs shown in Fig. 1(d). The host galaxy SED is obtained by extrapolating the linear model to fainter states until the disc is effectively turned off. Here we define this static point as the variable state $X_{\rm gal}$ at which the lower 1σ uncertainty envelope of any band is predicted to lie at zero flux. The model would be non-physical at any fainter state. With short-term variability assumed to arise by modulating the disc's SED, the variable disc's SED is the flux emitted at each band in excess of the galaxy's SED. Note in Fig. 1(d) that the disc SED is close to, but slightly redder than, the power-law $L_\nu \propto \nu^{1/3}$ spectra, indicated by dotted lines.

3 APPLICATION TO SDSS DATA

3.1 Sample selection: SDSS stripe 82 quasars

As described in the previous section and illustrated by Fig. 1, our disc + galaxy decomposition procedure requires multiwavelength coverage with a suitably long time baseline to adequately probe the variability of a given source. More importantly, the procedure is well-posed mathematically if and only if the multiwavelength coverage is near simultaneous ($\lesssim$1 night) as to constrain all relevant regimes of the SED at any one time. Thankfully, multiwavelength photometric coverage for transient surveys are usually performed on nightly basis, thereby providing multiple samplings in wavelength per source, per night.

A suitable survey satisfying these requirements is the Southern Sample of the Sloan Digital Sky Survey. The Southern Sample catalogue (MacLeod et al. 2012) contains re-calibrated *ugriz* light curves for all of the spectroscopcially confirmed quasars in SDSS DR7 Stripe 82. Summarily, the catalogue includes 9 258 quasars over $\sim$290 deg^2 with an observational baseline of $\sim$10 yr, observing each source for 2–3 consecutive months a year. The total number of epochs per source is $\sim$60 with photometric accuracy between 0.02 and 0.04 mag.

The original photometry was adopted from the official SDSS quasar catalogue (Schneider et al. 2010) using PSF magnitudes which were re-calibrated (see MacLeod et al. (2012) for details). According to Schneider et al. (2010), 97 per cent of these objects are registered as having point-like morphology, with the remaining 3 per cent limited to $z \lesssim 0.7$; $\sim$80 per cent of $z < 0.7$ sources are registered as point like. Future surveys such as LSST will be deeper and have higher angular resolution, relative to SDSS. As such, the task of accurately disentangling the nuclear quasar light from that of resolved host galaxies will require more detailed image modeling and/or aperture photometry.

Fig. 2 shows the photometric properties, redshift distribution, and sky density of sources within the catalogue. This sample provides a broad range in redshift, $0.1 < z < 6.0$, which extends the rest-frame spectral range deep into the ultraviolet and enables us to probe quasar variability and thus accretion disc structure out to remarkably early times. The photometry here is corrected for Galactic extinction using the coefficients provided in the catalogue, to thus be consistent with MacLeod et al. (2012).

As highlighted in Fig. 2(b) and (c), the typical observed-frame $g - i$ colour index of the SDSS quasars, taken from the initial catalogue prior to our decomposition analysis, is $\sim$0.5 mag. redder than a power-law $L_\nu \propto \nu^{1/3}$ spectrum. This may be expected due to contamination of the disc spectrum by light from the host galaxy and/or reddening due to dust on the line of sight to the quasar. In addition, the undulating redshift dependence of $g - i$ shown in Fig. 2(c) likely arises from quasar emission-line features redshifting into and out of the g and i passbands.

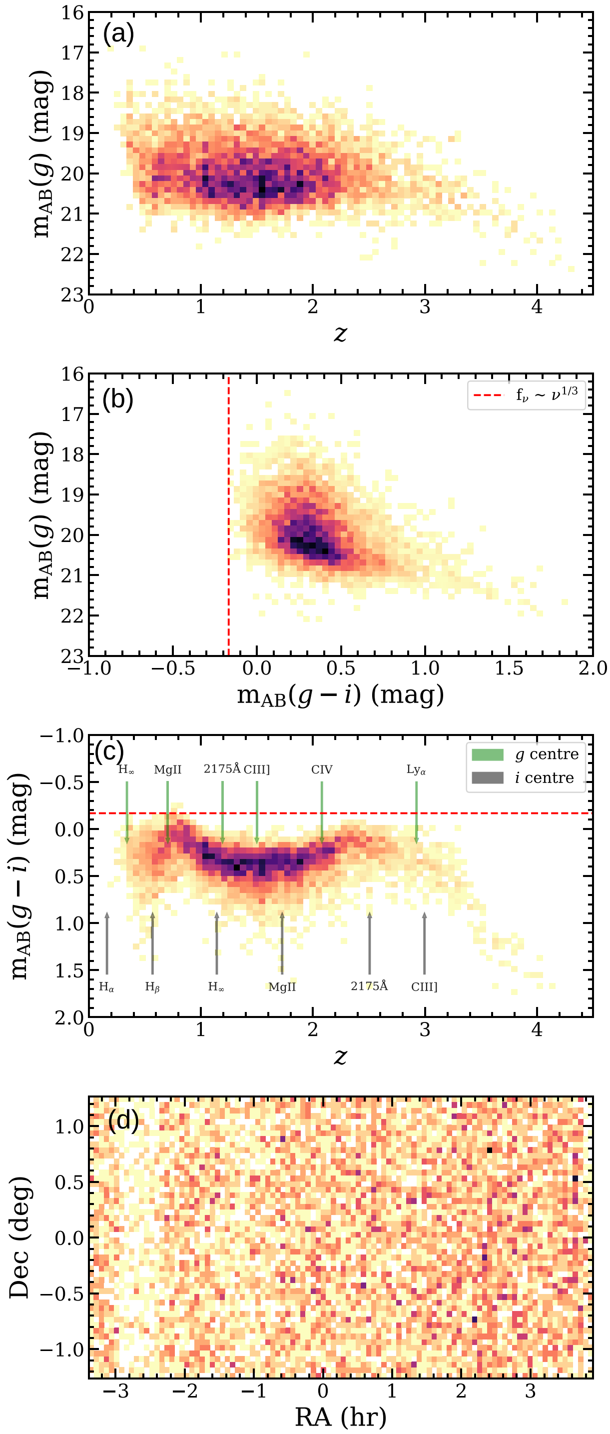

Figure 2. Summary of the raw input photometry from the SDSS Southern Sample. The g-band magnitude and $g - i$ colour distributions with redshift are shown in Panels (a) and (c). The g-band magnitude distribution is also shown against the $g - i$ colour distribution. The red dotted line indicates the expected $g - i$ colour from an $f_\nu \sim \nu^{1/3}$ spectrum. Panel (d) shows the density of sources on the sky.

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

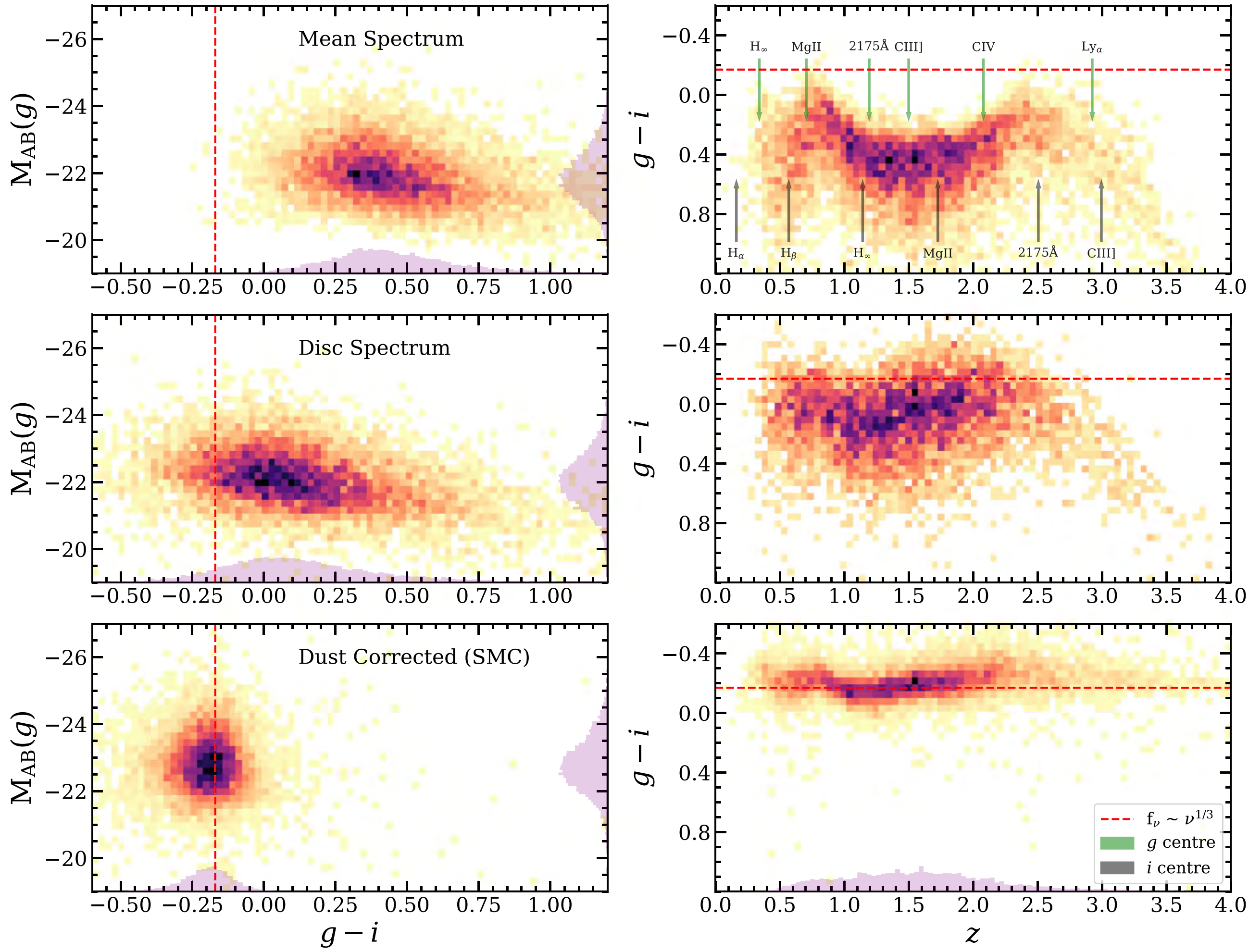

Figure 3. Luminosity (absolute g-band AB magnitude) versus $g - i$ colour index (left-hand panel) and $g - i$ versus redshift (right-hand panel) for the mean SED (top panel), disc SED (middle panel), and SMC de-reddened disc SED (bottom panel). Distributions are coloured darker with increasing density and shown as histograms projected on to each axis. The $g - i$ colour index for a power-law $L_\nu \sim \nu^{1/3}$ spectrum is indicated by the dashed red lines. Arrows in the top right-hand panel indicate the redshifts at which prominent spectral features are centred in the g- or i-band in green and grey, respectively.

3.2 Isolating disc SEDs using variations

For 99.8 per cent of the southern sample, the procedure explained in Section 2 succeeded in isolating the variable (accretion disc) and non-variable (host galaxy) SEDs. The analysis failed for just 0.2 per cent (16) of the sources, owing to either too sparsely sampled data (either in wavelength or epoch), insufficient variability signal, or a combination of the two.

The main effect of this disc + galaxy decomposition is evident for Object 1576517 in Fig. 1(c) and (d). Much of the red light is ascribed to the non-variable background galaxy component, shown in red in Fig. 1(d), thus isolating the relatively blue SED of the variable disc light, as shown in black in the same panel. In this case, the disc SED is slightly redder than the expected $L_\nu \propto \nu^{1/3}$ power-law SED, shown by dotted curves fixed to the observed magnitudes per band.

In Fig. 3, comparing the top and middle panels shows the effect of our disc + galaxy decomposition on $g - i$ colour indices over the full sample of SDSS quasars. The distribution of $g - i$ colours, for the mean and disc SEDs, are shown here as a function of magnitude and redshift. A red-dashed line marks the $g - i$ colour for the $f_\nu \propto \nu^{1/3}$ power law. The g-band absolute magnitude distribution is very similar for the disc and mean SEDs, indicating that these quasars are typically brighter at g than their host galaxies. The $g - i$ distributions differ significantly – the disc SEDs are generally bluer than the mean SEDs. While none of the SDSS quasars has a mean spectrum as blue as the $L_\nu \propto \nu^{1/3}$ power law, most of the disc SEDs have bluer $g - i$ colours, moving toward and in some cases beyond the $L_\nu \propto \nu^{1/3}$ power law colour. But the $g - i$ distribution is not simply translated, rather it appears to be stretched towards bluer colors, leaving behind a long red tail of somewhat fainter quasars with $g - i$ similar in their disc and mean SEDs. One possible interpretation of these redder and fainter SEDs is dust along the line of sight to the quasar disc.

Note also that the stark effect of the emission lines causing $g - i$ to undulate with redshift is stronger for the mean than for the disc SEDs. This is consistent with broad UV emission lines being less variable than the disc continuum.

3.3 Accounting for dust extinction and reddening

We now investigate the possibility of dust along the line of sight to the quasar discs. This dust could be absent or differ significantly from the dust along lines of sight to other parts of the host galaxy,

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

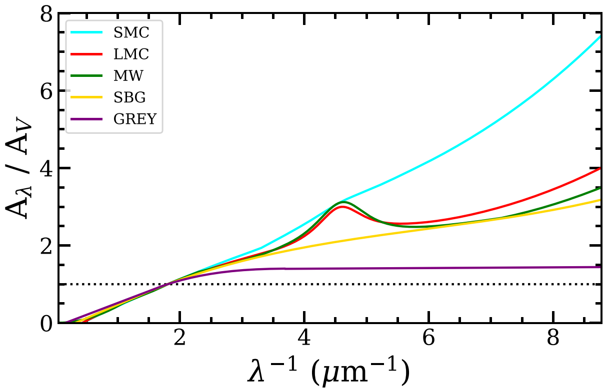

Figure 4. Attenuation curves of five commonly assumed dust attenuation laws used in this work. See Section 3.3 for details.

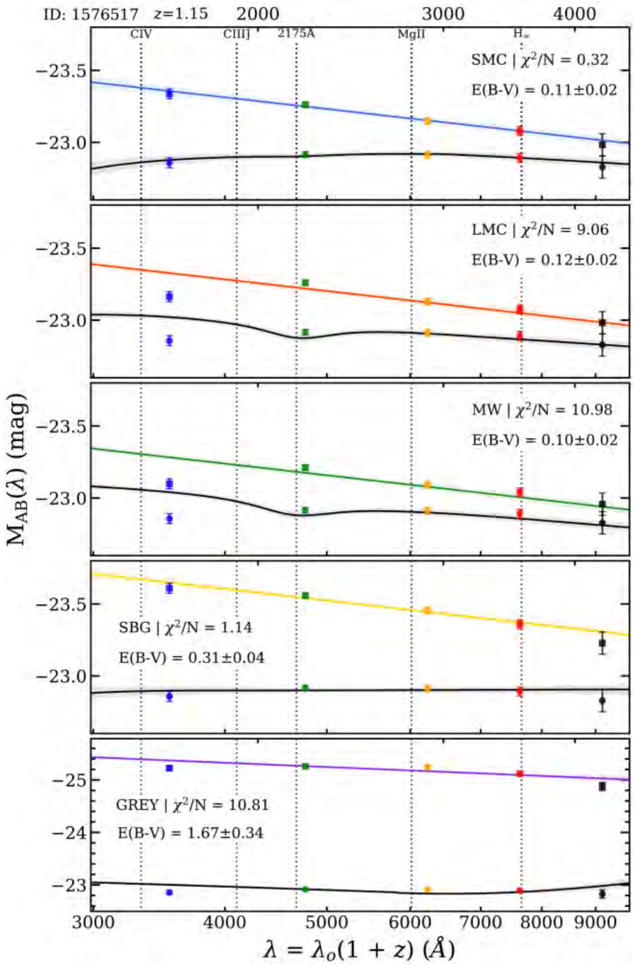

Figure 5. Dust-correcting the quasar disc spectrum (absolute AB magnitude versus observed- and rest-frame wavelength) to illustrate the procedure with each of the five dust laws in Fig. 4, for the same object as in Fig. 1. In each panel, the best-fitting dust-attenuated $L_\nu \propto \nu^{1/3}$ power-law spectrum (grey curve) is fitted to the observed disc fluxes (filled circles with error bars, coloured to correspond with the *ugriz* filters). The coloured lines show the same models after dust-correcting by setting $E(B-V)=0$. The corresponding coloured square points are similarly dust-corrected data. The best-fitting $E(B-V)$ and the reduced χ^2/N is shown in each panel. With two parameters fitting five data, there are $N = 5 - 2$ degrees of freedom.

since the quasar luminosity can heat and evaporate dust in its vicinity. Nevertheless, there is evidence from infrared interferometry (Hönig et al. 2013; Asmus 2019) for both polar dust and equatorial dust. While equatorial dust is thought to obscure the disc and associated broad emission-line regions in Type 2 AGN, polar dust may attenuate and redden the observed disc spectra even for more face-on discs.

Although there is extensive discussion in the literature (e.g. Gallerani et al. 2010; Krawczyk et al. 2015; Zafar et al. 2015), there is as yet no definitive evidence and certainly no consensus as to the correct or possibly universal dust extinction law for quasars. Given this uncertainty, we consider five possible dust laws. Their attenuation curves, A_λ/A_V, are shown in Fig. 4. The five dust laws are as follows:

(i) SMC – The Small Magellanic Cloud – a nearly smooth power-law-like curve with relatively high UV extinction due to small grains. Adopted from Gordon et al. (2003).

(ii) LMC – The Large Magellanic Cloud – a flatter UV extinction curve with a strong 2175(a)-Å graphite absorption feature. Adopted from Gordon et al. (2003).

(iii) MW – The Milky Way – similar to the LMC dust law with a strong 2175-Å feature. Adopted from Seaton (1979) fitted by Fitzpatrick (1986).

(iv) SBG – The Calzetti Starburst Law – a monotonic extinction curve similar to MW and LMC dust but lacking the graphite feature. Adopted from Calzetti et al. (2000).

(v) GREY – The Gaskell AGN Law – flattens in the UV due to absence of small grains. Adopted from Gaskell et al. (2004).

The dust-attenuated power-law spectrum model, expressed in absolute AB magnitude versus rest wavelength λ, is

$$M_{\mathrm{AB}}(\lambda) = M_0 + 2.5\,\alpha\,\log_{10}\left(\frac{\lambda}{\lambda_0}\right) + R(\lambda) \times E(B-V)\,. \qquad (3)$$

Here, $R(\lambda) \equiv -2.5\,\log_{10}(A_\lambda)/E(B-V)$ is the dust attenuation in magnitudes per colour excess $E(B-V)$. The intrinsic power-law spectrum is $L_\nu = L_0\,(\lambda_0/\lambda)^\alpha$, with disc theory predicting a power-law index $\alpha = 1/3$. With no dust, $E(B-V) = 0$, the model's absolute AB magnitude is M_0 at the fiducial rest wavelength $\lambda_0 = 2400$ Å, chosen because the vast majority of the SDSS quasars have rest-frame coverage at 2400 Å thus minimizing cases that pivot at wavelengths outside the observed *ugriz* range.

For all five dust laws, and for each SDSS quasar, we fit the observed five-band disc SED, holding $\alpha = 1/3$ fixed and minimizing χ^2 to estimate the two model parameters, M_0 and $E(B-V)$ in equation (3). Fig. 5 illustrates this fit and de-reddening procedure for the disc spectrum of SDSS ID 1576517 ($z = 1.15$) determined in Fig. 1. For each of the five dust laws, the de-reddened model spectrum, setting $E(B-V) = 0$, gives the intrinsic power-law $L_\nu \propto \nu^{1/3}$ fixed at the best-fitting value of M_0. This also allows the photometric data to be dust-corrected by compensating for the dust extinction at each wavelength. This analysis delivers best-fitting estimates for $E(B-V)$ and M_0, and a five-band dust-corrected SED, for each of the 9 242 quasar discs.

With 2 parameters fitted to five data, there are $N = 3$ residual degrees of freedom. If the data and model are reliable, the reduced χ^2/N should be $1 \pm \sqrt{2/3}$, helping to discriminate among the five dust laws. For the $z = 1.15$ quasar in Fig. 5, the *g*-band happens to sample the redshifted 2175 Å feature that arises from graphite grains and is prominent in the MW and LMC dust laws. The observed disc SED is relatively smooth about the *g*-band. This strongly disfavours the LMC and MW dust laws, $\chi^2/N = 9.06$ and 10.98, respectively, for which the *g*-band datum is above and the *u*-band datum is well

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

below the best-fitting model. For the GREY dust law, the best fit requires a larger dust correction compared with the other dust laws. Also, the GREY dust law leaves relatively large residuals, and so is also strongly disfavoured, $\chi^2/N = 10.81$. For this particular quasar, and for the assumed power-law index $\alpha = 1/3$, the SMC and SBG dust laws remain viable, with $\chi^2/N = 0.32$ and 1.14, respectively.

A secondary metric to consider is the best-fit colour excess $E(B - V)$, which quantifies the line of sight dust column density. A prior on the dust reservoir of the quasar host galaxy may be set by the relatively small values observed in most extragalactic systems, with the notable exception of dusty starbursts (Casey, Narayanan & Cooray 2014; Talia et al. 2021). In our analysis, with $E(B - V)$ a free parameter, a fit requiring a much higher $E(B - V)$ should be rightly disfavoured. In Fig. 5, the best fit with the GREY dust law gives $E(B - V) = 1.67 \pm 0.34$ mag, the SBG dust law gives 0.31 ± 0.04 mag, and the SMC, LMC, and MW dust laws are consistent with $E(B - V) \approx 0.11 \pm 0.02$ mag. Thus, importantly, the SMC is not only the best-fittinf dust-law as measured by χ^2/N, it also requires a significantly smaller $E(B - V)$ when compared to the next-best-fitting SBG dust law.

The similarities and differences among the dust law fits discussed above for SDSS ID 1576517 are found to hold statistically in the aggregate sample. For the SMC dust law, the lower panels of Fig. 3 demonstrate the dramatic tightening of the $g - i$ colour distribution effected by dust-correcting the quasar disc SEDs. For all five dust laws, Fig. 6 compares their dust-corrected colour–magnitude and colour–redshift distributions, reporting for each case the colour dispersion $\sigma(g - i)$ and the total χ^2 summed over all objects. The dust-corrected disc SEDs cluster around the assumed intrinsic $L_\nu \propto \nu^{1/3}$ power-law disc spectrum, with relatively mild dependencies on redshift. The tightest dispersions, $\sigma(g - i) \sim 0.23$ mag, are achieved similarly by the SMC, SBG, and GREY dust laws. This is closely followed by the LMC and MW dust laws, at 0.25 and 0.30 mag, respectively.

Despite their similar success in reducing the $g - i$ dispersion, we note several differences among the five dust laws. First, the GREY dust law, flat in the UV, is problematic as it spreads the dust-corrected disc SEDs over a wide range of implausibly large luminosities. In our view, this strongly disfavours the GREY dust law unless the intrinsic SEDs of quasar discs differ very substantially from a power-law spectrum.

For the LMC and MW, dust laws featuring graphite absorption at 2175 Å, the $g - i$ distribution has a tight core arising from the redshift range $0.9 < z < 1.6$, and broader wings from outside this range. This redshift structure stems from the 2175-Å feature redshifting across the g, r, and i bands, at $z \sim 1.2$, 1.9, and 2.5, respectively. At these redshifts, the evidence for absence of graphite absorption keeps $E(B - V)$ relatively small and better constrained than at intermediate redshifts where the feature falls between bands. This highly structured redshift dependence reduces the viability of our fits with these dust laws (see Appendix A).

In comparison, the SMC and SBG dust laws produce dust-corrected disc SEDs with tight distributions in both luminosity and colour, with small undulations in redshift that may plausibly be associated with emission-line features redshifting across the g and i bands, as indicated in the top right-hand panel of Fig. 6. Our fits with these dust laws also achieve the lowest total χ^2, 2.5×10^5 for SMC and 3.3×10^5 for SBG, compared with $(5.0, 7.3, \text{and } 8.8) \times 10^5$ for the (LMC, MW, and GREY) dust laws.

In conclusion, the SMC is our preferred dust law. It appears to be both reasonable and the best-fitting dust law overall, with a tight $g - i$ colour distribution centred about colour of an expected $L_\nu \propto \nu^{1/3}$ power law which is well constrained nearly equally at all redshifts. The SBG dust law is a close second choice, but with a somewhat higher χ^2. For individual sources, the SMC provides the best fit in 43 per cent of cases, followed by the LMC, MW, and SBG at around ~17 per cent each, and lastly by GREY at < 7 per cent (see Fig. 12). We continue with all five dust laws, but consider the SMC dust law to be the most appropriate for our subsequent analysis.

3.4 Host galaxy SEDs

As a check on our SED decomposition procedure, using variability to separate the variable disc and non-variable galaxy SEDs, we examine the resulting galaxy SEDs. If our linear extrapolation to (sometimes much) lower fluxes than observed is a poor approximation, the resulting galaxy SEDs could be distorted.

Fig. 7 shows the SEDs inferred for the quasar host galaxies, sorted by redshift and by dust extinction. The red curves show the galaxy SEDs. The blue curves show the corresponding dust-corrected disc SEDs. Higher redshift host galaxies appear to be more luminous than those at lower redshifts. This is a natural consequence of the SDSS quasar sample being approximately magnitude limited, with fainter objects being detectable only at lower redshifts. Note that at the highest redshifts, $3 < z < 6$, the galaxy SEDs are strongly affected by the Lyman break moving into and thus suppressing the luminosity in the u band.

The host galaxy SEDs may be expected to be fainter and redder in quasars for which a large $E(B - V)$ is inferred to produce a $L_\nu \propto \nu^{1/3}$ intrinsic disc spectrum. However, comparing the right two columns in Fig. 7, we see no strong trend in this direction. The host galaxies of more attenuated discs are perhaps a bit fainter, but not much redder. This implies that dust along the line of sight to the quasar disc is not strongly correlated with dust on lines of sight to stars in the host galaxy.

Turning to trends with redshift, at $z < 1$, the quasar host galaxy SEDs covering 2000–6000 Å all look similar. They are fainter than and intermediate in spectral shape between the red SED of NGC 7585 and the blue SED of Mrk 930, typical red sequence and blue cloud galaxies, shown for comparison in each panel of Fig. 7. At $z > 1$, the quasar host galaxy SEDs are brighter, and an increasing fraction of them exhibit a UV component producing a V-shaped *ugr* dip, with g fainter than u or r. This can be interpreted as a young stellar population as in star-forming (blue cloud) or intermediate (green valley) galaxies. They constitute a minority at $1 < z < 2$, and a majority at $2 < z < 3$, compatible with maximum star formation at cosmic noon, and decreasing thereafter. At $z > 3$, virtually all of the quasar host galaxies are blue-cloud starbursts, with strong UV emission and brighter than the SED of the compact blue starburst galaxy Mrk 930. The Lyman break appears to depress the galaxy SEDs on the blue end.

These trends with redshift accord with our current understanding of the star formation history of galaxies over cosmic time (Madau & Dickinson 2014; Förster Schreiber & Wuyts 2020). As summarized in Fig. 8, star-forming hosts become increasingly faint with time. We find no significant change if we remove the 3 per cent of sources registered with resolved morphologies, as they constitute $\lesssim$20 per cent of sources at $z \lesssim 0.7$, and $\ll$1 per cent at higher redshifts. While these trends could be affected by unknown selection biases, they are broadly consistent with the well-known fading of star formation between $z \sim 2$ and the present epoch. In contrast to the quasar host galaxies, the dust-corrected disc luminosities are remarkably stable across all epochs, $M_{\rm AB} \sim -22.6$ at $\lambda_0 = 2400$ Å, with a dispersion of ~0.4 mag, becoming less certain at z

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

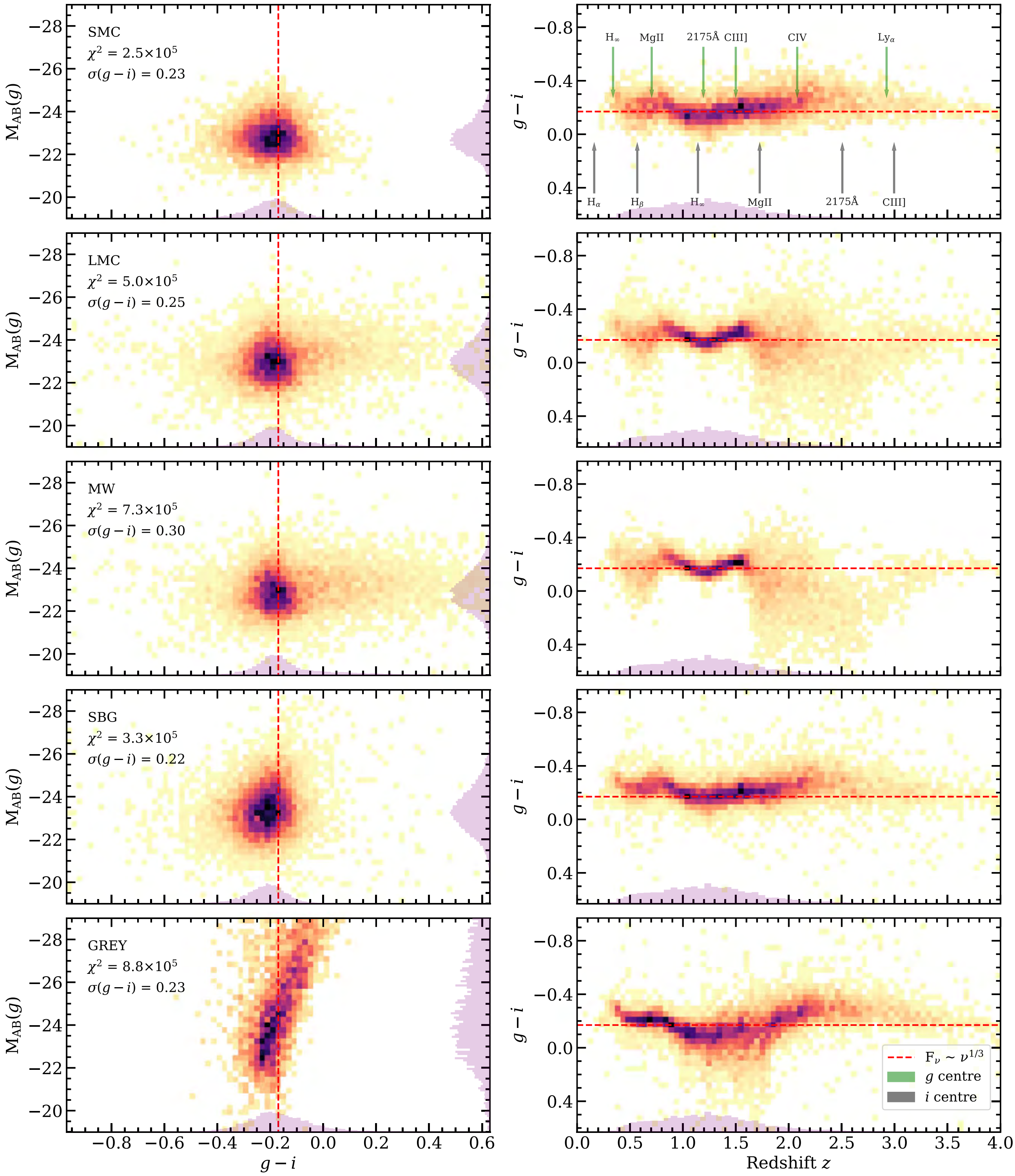

Figure 6. $g-i$ colour–magnitude (left-hand panel) and redshift-colour (right-hand panel) diagrams for each assumed dust law used to de-redden the 9156 *ugiz* disc SED. Distributions are coloured darker with increasing density and shown as a histogram projected on to each axis. The expected $F_\nu \sim \nu^{1/3}$ is indicated by the dotted red lines. Arrows indicate the presence of an emission feature in the center of the g- or i-band in green and grey, respectively. The reported values for χ^2 are total combined χ^2 statistics over the sample.

> 3 where extrapolation redward of the observed SED is required. Nevertheless, these encouraging results serve to validate our procedure using variability to separate the quasar disc and host galaxy light.

Some of our galaxy SEDs have u brighter than g, in fact rising more rapidly into the UV compared with a blue stellar population, perhaps even more rapidly than a Rayleigh–Jeans slope. This effect is likely a small flaw in our decomposition procedure. We currently

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

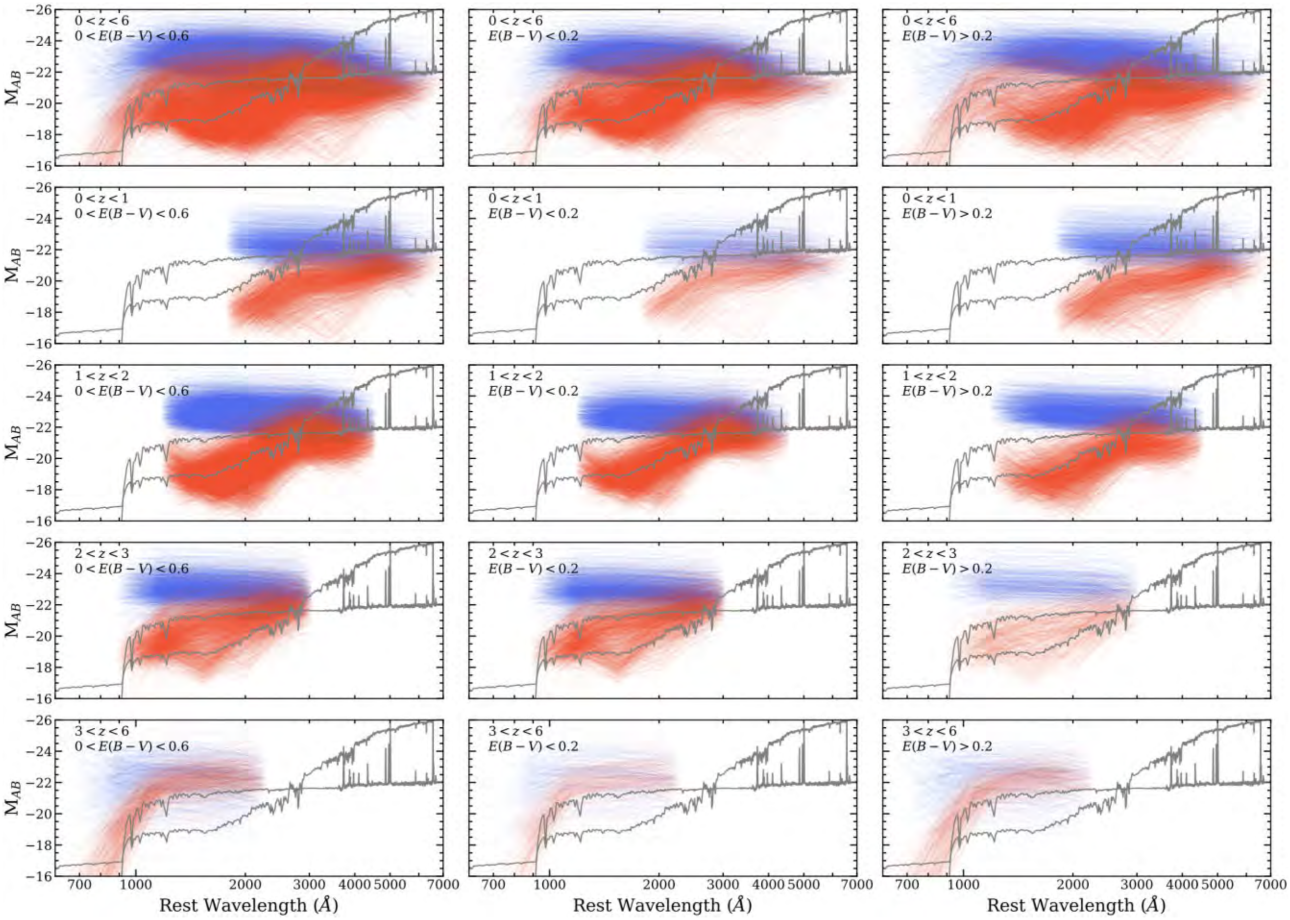

Figure 7. Quasar host galaxy SEDs (red) and dust-corrected disc SEDs (blue) sorted by redshift (columns) and $E(B-V)$ (rows). SEDs for typical galaxies in the red sequence (e.g. the lenticular shell galaxy NGC 7585) and blue cloud (e.g. the compact starburst Mrk 930), adopted from Brown et al. (2014), are shown in each panel for reference.

set $X_{\rm gal}$ at the lowest possible level, so that the extrapolated flux in one band, usually u or g, is 1σ above 0. A slightly higher level for $X_{\rm gal}$ could be used, thus moving a small fraction of the disc SED to the galaxy SED. The effect would be to make the V-shaped *ugr* dip in the galaxy SEDs less prominent in those cases where g is fainter than u, elevating the galaxy flux at g and moving the galaxy SEDs closer to the SED of a blue cloud galaxy. The disc SED would then have a correspondingly lower flux at g. We have not yet implemented this procedural tweak. We expect it to have a relatively small effect on the disc SEDs, which are much brighter than the galaxy.

These results follow the trends found in the analysis of Matsuoka et al. (2014), who performed a spatial decomposition to extract point-like quasar signals from their host galaxies, based on the same SDSS observations of Stripe 82. Limited to resolved sources at $z < 0.6$, they find that quasars are bluer than their host galaxies, with a quasar-to-host ratio of $\sim$8 in u and $\sim$1 in i. For our sources, at $z < 0.6$, host light is also typically fainter than our de-reddened discs, by a factor of $\sim$90 in u and $\sim$3 in i. However, a more equivalent comparison using our reddened (i.e. uncorrected) disc components produces a less extreme ratio of $\sim$30 in u and $\sim$2 in i, in better agreement with Matsuoka et al. The remaining discrepancy could be driven by a combination of selection effects, PSF-modelling biases in the analysis from Matsuoka et al. and that our definition of host galaxy may underestimate the host contribution at u. Regardless, it seems that the quasar discs are bluer than their hosts.

4 COMPOSITE SPECTRA OF VARIABLE QUASAR DISCS

The SDSS Stripe 82 quasar sample provides an unprecedented multiyear record of multiband quasar variations, but it is limited to just five optical bands (*ugriz*). Despite this drawback, we can leverage the cosmological redshift range to construct a composite quasar disc spectrum at somewhat finer spectral resolution and extending to much bluer rest-frame ultraviolet wavelengths. Our approach implicitly assumes that the spectral features of the accretion disc are universal, an assumption we made in the dust-correction procedure by assuming an $L_\nu \propto \nu^{1/3}$ power law for the intrinsic disc spectrum. As justified in Section 3.3, we assume that local extinction follows the SMC law for all sources when constructing our composite spectrum. It is also important to note that spectral features seen here will be smoothed out by the resolution of the filter profile of each band.

For each SDSS quasar, we have removed the host galaxy contribution by using the spectrum of the variable component, and corrected for possible dust extinction and SMC-like reddening in the host galaxy, assuming an $L_\nu \propto \nu^{1/3}$ power law for the intrinsic disc spectrum. We construct a composite disc spectrum by combining the resulting disc SEDs for quasars sampled across a continuous range of redshifts $0.1 < z < 6$. Our dust-correction procedure fits the power-law model, $L_\nu(\lambda) = L_0\,(\lambda_0/\lambda)^\alpha$, assuming $\alpha = 1/3$, to determine for each quasar the luminosity L_0 at reference rest wavelength $\lambda_0 =$

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

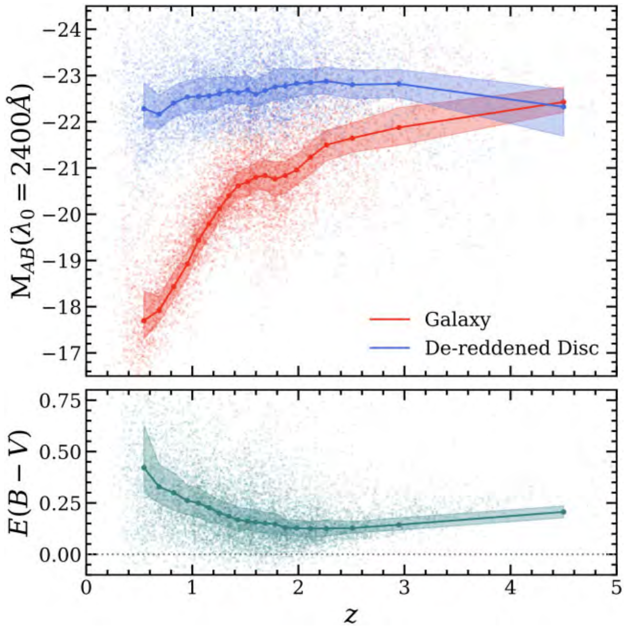

Figure 8. Upper panel: rest-frame 2400-Å absolute AB magnitudes as a function of z for the galaxy SED component (red) and disc component (blue), dust-corrected assuming an SMC-like attenuation, and $L_\nu \propto \nu^{1/3}$ power law. Lower panel*:* Dust attenuation $E(B-V)$ estimates assuming SMC. Bins have equal occupation at all redshifts and are used to determine the medians and 68 per cent ranges as shown.

2400 Å, and the required $E(B-V)$. The power-law spectrum provides a backbone for our composite disc model. We simply scale the L_0 for each quasar to a common value (-22.6 AB, as evidenced by Fig. 8), and scale the dust-corrected *ugriz* fluxes by the same scale factor. This provides five measurements on the composite spectrum, at the rest wavelengths of the *ugriz* bands at the redshift of that quasar. Doing that for the full sample gives ∼45 000 points, which we then average with a binned median to reduce the scatter.

In Section 4.1, we construct the composite disc spectrum assuming the canonical $L_\nu \propto \nu^{1/3}$ power law. In Section 4.2, we generalize this analysis by assuming $L_\nu \propto \nu^\alpha$ and solving for the best-fitting power-law slope α, thus testing the disc theory prediction $\alpha = 1/3$.

4.1 Composite disc spectrum for $L_\nu \propto \nu^{1/3}$

Fig. 9 presents the composite disc spectrum assuming a canonical $L_\nu \propto \nu^{1/3}$ power law. We remove 1 per cent of sources with the worst χ^2 (typically in excess of 100) to keep these poor fits from dominating the overall statistics. The top panel stacks the five-band dust-corrected disc SEDs for over the remaining 9 150 SDSS quasars, sorted by redshift, and coloured according to relative brightness, interpolating linearly between the pivot wavelengths of the *ugriz* bands. Despite somewhat larger noise in the u and z bands, there is clearly a general increase in brightness toward bluer rest-frame wavelengths. This reflects the assumed $L_\nu \propto \nu^{1/3}$ power-law adopted for the dust corrections.

The middle panel presents the composite disc spectrum, which undulates above and below the assumed power-law spectrum (red dashed line). Here, the 5 × 9 150 individual photometric measurements are summarized using a median with ∼300 points in each bin (blue curve with a 68 per cent envelope). We additionally show 1 in 5 (i.e. 20 per cent) of sources to illustrate the object-to-object variations. The panel below this uses a similar format (in green) to show residuals relative to the power law. The bottom panel shows the number of quasars contributing at each wavelength, nearly all 9150 in the middle at 2400 Å and dropping below 100 on the ends below 600 Å and above 7000 Å.

The power-law model provides a reasonable match to the data, with a reduced $\chi^2/N = 11.75$. Undulations around the power law are significant and plausibly attributed to variable spectral features such as the Balmer continuum emission around 3500 Å. Only a handful of low-redshift quasars contribute to the Hα peak at 6563 Å. The downturn blueward of 1200 Å is expected due to intervening Lyman α forest absorption in the u band at $z > 2$. The more dramatic drop blueward of 900 Å is the Lyman break arising from Lyman continuum absorption depressing the u-band flux in the highest redshift quasars at $z > 3$.

4.2 Consideration of alternative spectral slopes

The common assumption that the underlying power law should be $L_\nu \propto \nu^{1/3}$ originates from classical theory (Shakura & Sunyaev 1973) and has not been conclusively demonstrated to be the true underlying spectral power-law. This work so far has made the same assumption, and so now we question it. While the disc-decomposition procedure is entirely independent of the assumed underlying power law, the de-reddening procedure is not. Hence, we re-derive all dust reddening solutions for each of the five assumed dust laws assuming a range of underlying power-law slopes.

As before, we compute the aggregate χ^2 statistic for each assumed dust law *and* underlying power-law index α such that $L_\nu \propto \nu^\alpha$, for a coarse grid of α values. The results are shown in Fig. 10, measured both using the aggregate χ^2 of the sample (left-hand panel) and only the scatter in $g - i$ colours (right-hand panel). Each coloured curve corresponds to an assumed dust law and each point in α is coloured by the median $E(B-V)$ achieved under those two assumptions. Whereas red colours indicate high median dust extinction, greyscale colours denote the median $E(B-V) < 0$ is non-physical, although some objects in a given collection may still have $E(B-V) > 0$. The best-fitting α for each curve is calculated by fitting a seventh-order polynomial to the samples in α and computing the minimum χ^2. Estimates for the best-fitting values of α are shown also in Fig. 10 and are indicated by colour-corresponding vertical dotted lines with their uncertainty envelopes calculated where $\Delta\chi^2 = \chi^2_{\rm min}/N$, which are unexpectedly small for such a large data set. For reference, grey vertical dashed and dash–dotted lines are included to indicate α corresponding to the canonical $L_\nu \propto \nu^{1/3}$ as well as an alternative $L_\nu \propto \nu^{5/7}$, respectively. Table 1 summarizes the best-fitting parameters for each dust law, for fits with $\alpha = 1/3$, 5/7, and α optimized for each dust law.

From the left-hand panel of Fig. 10, showing the χ^2 landscape as a function of α, we find that the graphite-heavy dust laws for LMC and MW are strongly disfavoured. Their minimum χ^2 occurs for a red spectral slope ($\alpha < 0$) and under the assumption of $L_\nu \propto \nu^{1/3}$ they are disfavoured at high confidence in clear excess of 5σ. The best-fitting solution for the UV-flat GREY dust law occurs for an even redder spectrum, $\alpha = -0.39$, and with an non-physical median $E(B-V) = -0.67$ mag. LMC, MW, and GREY also have relatively high $\chi^2_{\rm min}$ values, corresponding to reduced χ^2/N values 11.19, 13.19, 13.80, respectively.

The SMC and SBG laws fare rather better, achieving lower best-fit $\chi^2_{\rm min}$ values of 5.99 and 7.37, respectively. The SMC dust law achieves its lowest overall $\chi^2_{\rm min}$ at $\alpha = 0.71 \pm 0.02$ and is tightly constrained relative to the SBG, which achieves its relatively higher $\chi^2_{\rm min}$ at $\alpha = 1.39 \pm 0.04$. While the SMC enjoys a smooth progression

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

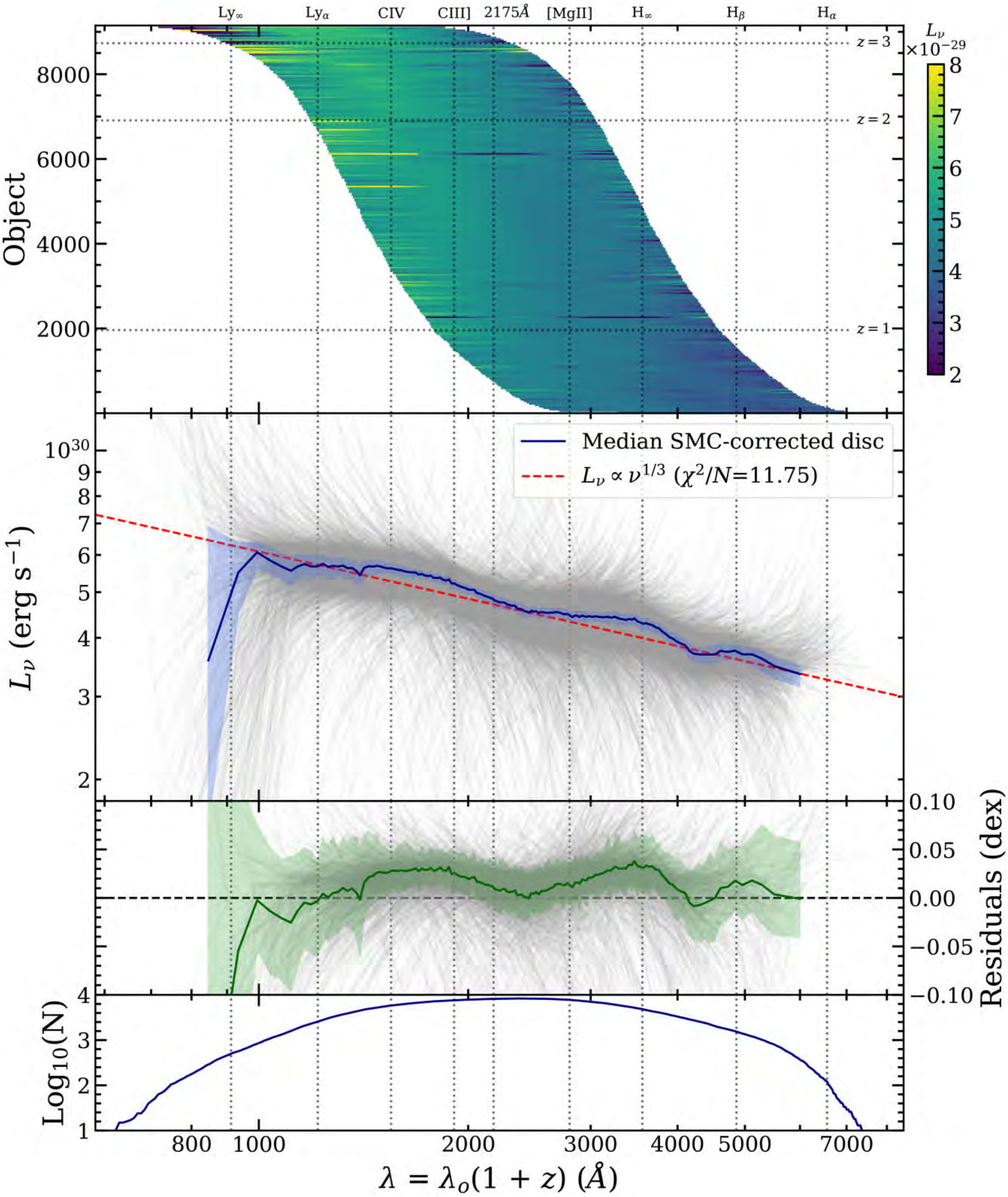

Figure 9. SEDs for the sample are shown at the top, shifted into the rest frame and ordered by z. Relevant emission lines are indicated by the vertical dotted lines. Shown in the middle is the composite spectrum of the de-reddened disc component computed with a binned median (blue, ∼300 points per bin) with a 68 percentile envelope indicating the width of the distribution at that point, with scatter indicated behind by 1-in-5 SEDs. We assume an SMC-like reddening and $L_\nu \propto \nu^{1/3}$ power law, overlaid in red. The residuals are shown in the lower panel in green, with an envelope likewise from above. The lower panel shows the distribution of sources contributing to any give rest-frame wavelength.

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

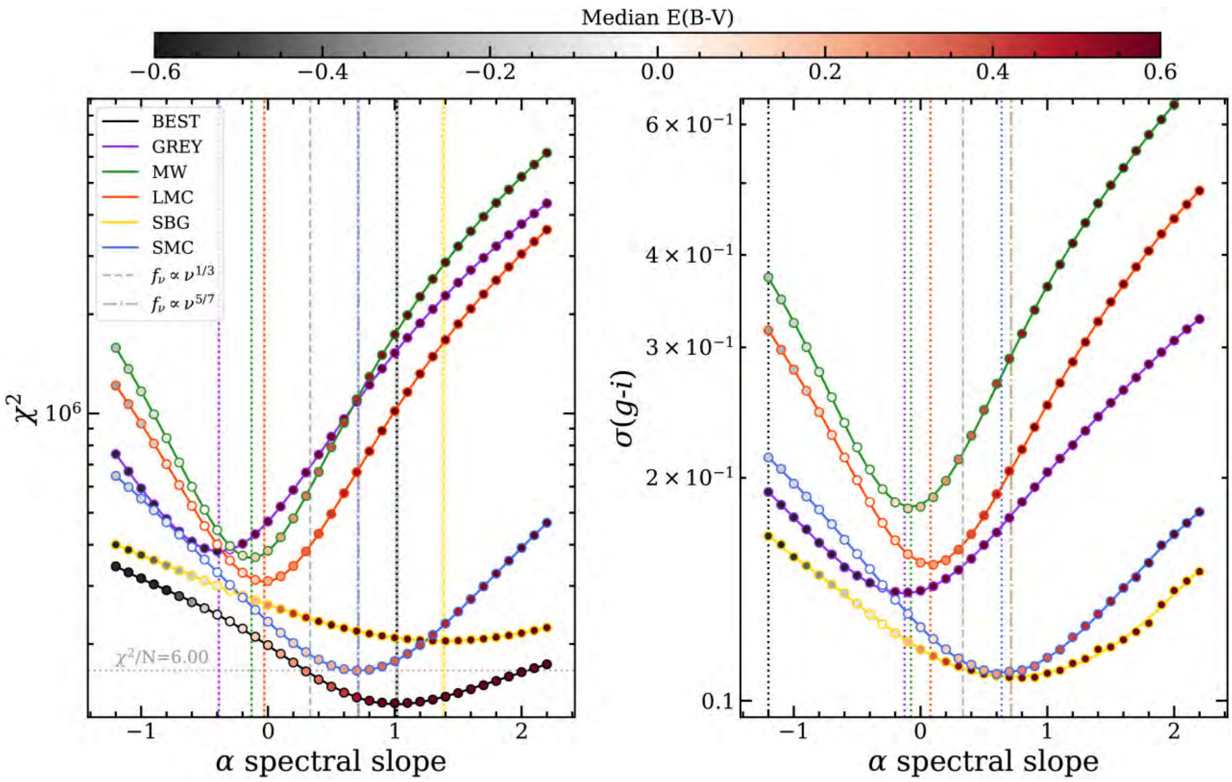

Figure 10. Two aggregate badness-of-fit (BoF) metrics, the total χ^2 (left-hand panel) and $\sigma(g-i)$ (right-hand panel), as functions of the accretion disc spectral index α. Points show the BoF values for each of the five candidate dust laws, fitting the α grid with seventh-order polynomials. For each dust law the median $E(B-V)$ is indicated by the filler of each point, red for positive values, and grey for (non-physical) negative attenuation. The best-fitting α values, at BoF minima, are marked on the left-hand panel by vertical dotted coloured lines within coloured bands denoting the uncertainty in α based on the $\Delta\chi^2 = \chi^2_{\rm min}/N$ criterion. Fiducial power-law models corresponding to $\nu^{1/3}$ and $\nu^{5/7}$ are marked by vertical grey lines. The best fit is achieved for SMC-like dust, median $E(B-V) = 0.28$ and $\alpha = 0.71 \pm 0.02$, close to 5/7, as detailed in Table 1. Lastly, we compute the total χ^2 at each α using value corresponding to the best-fitting dust law for each source, shown in black circles.

Table 1. Summary of best-fitting χ^2_N for the total sample of 9 242 sources, reduced by the number of degrees of freedom ($N = 3 \times 9\,242$), including corresponding median $E(B-V)$ and $\sigma(g-i)$ for each of the five assumed dust laws, in the case, where $L_\nu \propto \nu^{1/3}$, $\propto \nu^{5/7}$, and the best-fitting spectral slope α whose uncertainty is boosted by $\sqrt{\chi^2/N}$.

	α	χ^2/N	$E(B-V)$ Median	$E(B-V)$ $\sigma_{\rm MAD}$	$\sigma(g-i)$
		$L_\nu \propto \nu^{1/3}$			
SMC	0.33...	6.70	0.18	0.13	0.11
SBG	0.33...	8.61	0.55	0.38	0.11
LMC	0.33...	14.26	0.34	0.22	0.16
MW	0.33...	21.30	0.32	0.23	0.22
GREY	0.33...	24.80	6.18	7.10	0.15
		$L_\nu \propto \nu^{5/7}$			
SMC	0.71...	5.99	0.28	0.15	0.11
SBG	0.71...	7.88	0.86	0.37	0.11
LMC	0.71...	24.30	0.52	0.23	0.21
MW	0.71...	40.44	0.48	0.27	0.29
GREY	0.71...	39.62	11.63	11.61	0.18
		Best-fitting α			
SMC	0.71 ± 0.02	5.99	0.28	0.15	0.11
SBG	1.39 ± 0.04	7.37	1.43	0.39	0.12
LMC	-0.03 ± 0.01	11.19	0.18	0.21	0.15
MW	-0.13 ± 0.01	13.19	0.13	0.21	0.18
GREY	-0.39 ± 0.02	13.80	− 0.67	4.47	0.14

of median $E(B-V)$ values, the SBG requires even greater reddening for the same α despite turning over in $E(B-V)$ at the same α. Thus, the aggregate sample measured with χ^2 corroborates the aforementioned findings that the SMC provides the best-fitting dust solution to describe the de-reddened disc SED of the sample considered. However, we also measure the total χ^2 at each α by assigning each source a χ^2 corresponding to the best-fitting dust law, and recompute the best-fitting α (called 'BEST'). Unsurprisingly, this combined sample finds a minimum $\alpha = 1.01 \pm 0.02$, mid way between the SMC and SBG laws which are the two best-fitting dust laws for any given α. However, this best-fitting α is suffers from a high median $E(B-V)$ compared to the SMC law at $\alpha \sim 1$, and may suffer from noisy measurements (as they are not weighted here), preferential χ^2 arsing from unreasonably attenuated solutions, and other effects which make its interpretation non-trivial.

The constraining power of χ^2 in the left-hand panel of Fig. 10 is distinctly superior to that of $\sigma(g-i)$ in the right-hand panel. This makes sense, as χ^2 utilizes all five bands compared to only two bands in $\sigma(g-i)$. Nevertheless, comparing these may increase confidence in the results and deepen our understanding of the models. Note that the order of the best-fitting α values for different dust laws is the same for minima of χ^2 and minima of $\sigma(g-i)$. The best-fitting α values have a smaller range for minima of $\sigma(g-i)$. The MW, LMC, and GREY dust laws cluster around $\alpha = 0$, with relatively high $\sigma(g-i)$. The SMC and SBG are both close to $\alpha = 0.7$, with SBG achieving

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

a slightly lower $\sigma(g - i)$ than that for SMC. The $E(B - V)$ values at a given α are generally similar between the estimators.

Owing to the large sample size of this investigation, the constraint on the best-fitting slope α is remarkably tight, uncertain by of order 1 per cent for the χ^2 BoF. Both χ^2 and $\sigma(g - i)$ prefer a power-law spectral index α significantly bluer than the canonical $L_\nu \propto \nu^{1/3}$ accretion disc spectrum. The $L_\nu \propto \nu^{1/3}$ power law is statistically inconsistent with our results.

The merit of this result is further explored in Fig. 11 where we perform the same spectral composition procedure as shown by Fig. 9 but assume an $L_\nu \propto \nu^{5/7}$ while maintaining the assumption of an SMC-like power law. By doing so, we find an even more consistent picture with the $L_\nu \propto \nu^{5/7}$ power law, achieving a $\chi^2/N = 9.00$. This is lower than that achieved with the expected $L_\nu \propto \nu^{1/3}$, suggesting that $\alpha = 5/7$ is a more appropriate model. In addition, it is apparent that the bluest residuals have lessened, with the de-reddened disc spectrum now being consistent with a smooth $L_\nu \propto \nu^{5/7}$ power law within a 68 percentile range for wavelengths bluer than H_α and redder than the Lyman continuum break. We discuss the implications of this serendipitous finding in the following section.

5 DISCUSSION

5.1 Assumptions and caveats

Advantageous properties of the SDSS data set analysed here are its unprecedented number of quasars and the long time-span over which the five-band *ugriz* photometry has been obtained. For each quasar, we leverage its variable nature to separate the variable disc component from its static host galaxy. Our decomposition method treats each observation as an independent measurement of the galaxy+disc flux at some time-dependent dimensionless brightness level $X(t)$, where $X = 0$ is the mean level and $\Delta X = 1$ is the rms of the light-curve variations. This is illustrated in Fig. 1, where each flux measurement provides an independent constraint on the linear model, $F(\lambda, t) = B(\lambda) + A(\lambda)\,X(t)$. Here, the intercept $B(\lambda)$ is the mean galaxy + disc spectrum at $X = 0$ and the slope $A(\lambda)$ is the rms spectrum of the disc variations. Extrapolating the fit to fainter levels is assumed to effectively turn-off the variable disc light leaving just the galaxy spectrum at some minimum value of X. This point is somewhat arbitrary, particularly when the variations are small so that the extrapolation is a long one. In order to have a well-defined decomposition, we adopt the point at which the extrapolated flux is consistent with zero flux at 1σ, which we interpret as the limit below which the model is no longer physically meaningful.

Our estimates of $E(B - V)$ for five different extinction laws are computed for each source to quantify and compensate for extinction and reddening of the disc spectrum by dust along the line of sight to the accretion disc. This assumes that the observed disc spectrum is fainter and redder than the intrinsic disc spectrum due to line-of-sight reddening and extinction by dust, although the converse (requiring non-physical negative attenuation) is an allowed solution. For the intrinsic disc spectrum, we assume a power-law, $L_\nu \propto \nu^\alpha$. The estimate of $E(B - V)$ depends on the assumed power-law spectral index α, a bluer slope requires a larger $E(B - V)$. Our power-law disc model neglects possible contributions of emission lines and bound-free continua. The variability of this approximation is supported by Fig. 3, which shows that the variable disc spectrum has weaker emission features than the mean spectrum. However, as shown by Fig. 9, the final de-reddened disc spectrum, assuming an SMC-like extinction and $L_\nu \propto \nu^{1/3}$, shows that some emission features remain. These are of course smoothed by the broad bandwidths of the *ugriz* filters, leaving wide and weak rather than narrow and strong emission features in the residuals. Although visually the residual features corresponding to the $\alpha = 1/3$ composite (Fig. 9) appear similar to those of the $\alpha = 5/7$ (Fig. 11), the latter achieves a significantly better fit, $\Delta\chi^2 \sim 20\,000$ or $\chi^2/N = 11.75 \rightarrow 9.00$. However, we caution that overfitting and unseen systematics may contribute to this effect.

Despite the straight-forward interpretation that a combination of SMC-like dust and a bluer spectral slope describes the variable accretion disc spectra of quasars, we note several caveats. First, for the MW and LMC laws the $E(B - V)$ distribution of the SDSS quasars has an implausible redshift dependence caused by the strong rest-frame 2175-Å absorption feature moving across the centre of a band. A 'beating' pattern is observed where the estimates of $E(B - V)$ have a large scatter when the 2175-Å feature is not directly observed (see Appendix A). This highlights a shortcoming in the modeling of the dust when adopting the MW and LMC models. We considered addressing this by using a prior favouring models that make $E(B - V)$ a smoother function of redshift, but decided in the interest of simplicity to omit this complication in our modelling. Our model also places no limitation on the extent to which the variable component can be reddened, which may permit extreme reddening requiring extraordinary dust column densities. We note that, with the exception of the GREY law, there are no instances of problematically dusty attenuation estimates.

5.2 On attenuation laws

The dust laws investigated in this work broadly fall into two categories. Either they are well described by a smooth power-law-like curve (e.g. SMC, SBG), or a power-law-like curve with a strong graphite absorption feature at 2175 Å (LMC, MW). The outlying case is that of the Gaskell's dust law derived from a sample of AGN (GREY). Their differences are highlighted in Fig. 4.

As shown for a specific case in Fig. 5, the likelihood that a particular dust law is well suited for a particular source is assessed here using χ^2 to quantify the badness-of-fit and $E(B - V)$ to indicate cases where exceptional dust columns would be required. Similarly, to quantify the success in modelling the SDSS quasar sample as a whole, we employ two badness-of-fit metrics, χ^2 and $\sigma(g - i)$, along with the median $E(B - V)$. Given that the information presented by $\sigma(g - i)$ is contained in χ^2 and has generally less constraining power (see Fig. 10), we adopt χ^2 as the primary criterion, using $\sigma(g - i)$ and $E(B - V)$ for secondary considerations.

As presented in Section 3, assuming a $L_\nu \propto \nu^{1/3}$ power law, the least likely dust law is GREY which is statistically excluded at high confidence for the vast majority of the sources. However, 7 per cent of the sample finds a best-fitting solution with the GREY extinction law, but with an exceptionally large median $E(B - V) = 6.86$ mag for this sub-sample. Further, we find evidence to exclude the graphite absorption laws of the LMC and MW as a general best-fit, finding the best-fitting solution for only 17 and 16 per cent of the total sample, respectively.

We find the greatest success with the smooth power-law extinction laws, SMC and SBG. As measured by χ^2 in Fig. 10, the SMC provides the best fit to the sample as a whole and is consistent with a best-fitting $\alpha \sim 0.7$ from both the χ^2 and $\sigma(g - i)$ estimators. Assuming an $L_\nu \propto \nu^{1/3}$ power law, Fig. 12 shows that the SMC provides the best-fitting solution for 43 per cent of the sample while the SBG provides 17 per cent. Interestingly, although constrained with less information, the $\sigma(g - i)$ estimator finds that the SBG provides a solution similar to the SMC consistent with a $L_\nu \propto \nu^{5/7}$ power law. Thus, while we cannot exclude the SBG law outright, we

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

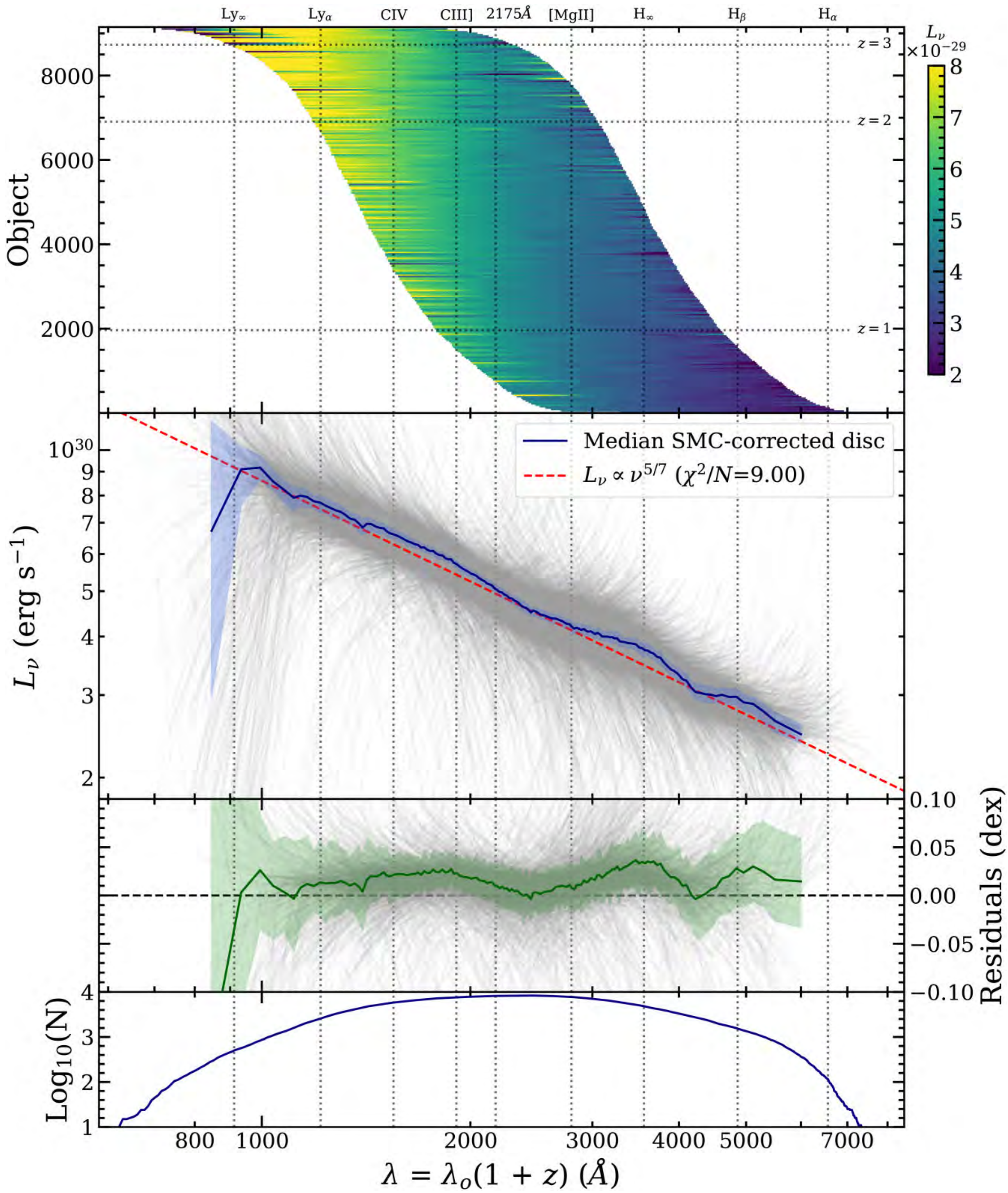

Figure 11. Same as Fig. 9, but assuming a $L_\nu \propto \nu^{5/7}$ power law.

none the less find the most likely dust law for this sample of quasars is the SMC. This result holds also in the case of an assumed $L_\nu \propto \nu^{5/7}$ power law.

Considering now the derived composite SEDs shown in Figs 9 and 11, under the assumption of an SMC-like dust law derived assuming either $L_\nu \propto \nu^{1/3}$ or $\propto \nu^{5/7}$, there are no discernible features consistent with strong Balmer absorption or emission at the 10 per cent level. However, given that the emission is smeared across the broad-band filter, only the highest equivalent width lines could be detected. Continuum consistent with thermal emission from an optically thick accretion disc is clear (i.e. the 'Big Blue Bump'; Malkan & Sargent 1982).

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

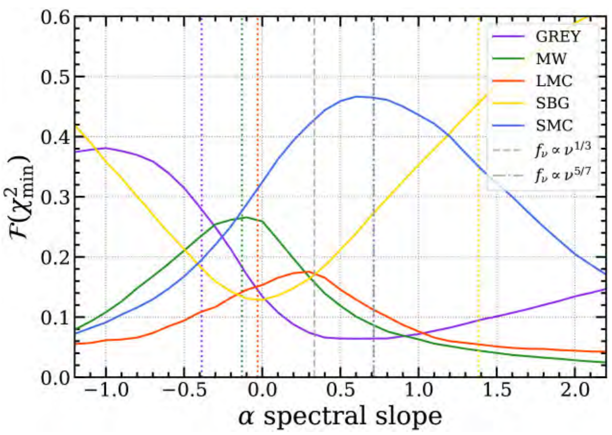

Figure 12. Fraction of sources per assumed spectral slope α best-fit by the given attenuation law, as measured by minimum χ^2. Vertical dotted lines indicate the best-fitting α for each dust law.

The southern sample data set contains several sources observed at $3 < z < 6$. At these redshifts, the rest-frame u band intersects the expected rest-frame UV turnover of the accretion disc spectrum at ~1000 Å. Although this subset constitutes only a small fraction of the total sample, the effect of the turnover is evident from Fig. 9. The handful of these sources approaching $z \sim 6$ may also be affected by the neutral intergalactic medium which may be contributing to this observed turnover with resonant absorption by hydrogen gas clouds along the line of sight (i.e. the well-known Lyman Forest). Regardless of the physical mechanism driving this highly significant turnover, it is not accounted for in the continuous power law form assumed for the disc spectrum. Consequently, the continuous power-law model is not appropriate for the bluest bands for $z > 3$ objects, whose $E(B - V)$ may be overestimated due to the turnover acting as an extreme reddening of the u band.

5.3 Best-fitting power-law exponent

Initially, we assumed the canonical $L_\nu \propto \nu^{1/3}$ power law before expanding to a range of power law exponents to determine the most suitable power-law index and extinction law combination as assessed by the badness-of-fit χ^2 statistic. The result is shown in Fig. 10. The SMC law shows remarkable agreement from both χ^2 and $\sigma(g - i)$ estimators finding a best-fitting blue slope $\alpha \sim 0.7$ in both cases. For the latter estimator, SBG finds its minimum also at $\alpha \sim 0.7$. Taking χ^2 to be the more robust and more precise estimator, the corresponding α for the best-fitting SMC is 0.71 ± 0.02. This is highly inconsistent with $L_\nu \propto \nu^{1/3}$ predicted for a geometrically thin steady-state disc, within theses strict uncertainties.

Given this unexpected result, we verified the robustness of the de-reddening procedure by measuring 9000 simulated disc SEDs with $\alpha = 1/3$ perturbed with random noise corresponding to that of the observed photometry. We successfully recovered a best-fitting power-law index of 1/3, confirming that the procedure is unbiased and that the result is indeed genuine.

This result contrasts with work from Kokubo et al. (2014), who used the same SDSS data set and employed a 'flux-flux' deomposition method most similar to ours in order to explore the variations in the photometric light curves. They found that the composite spectrum *before* decomposition features a red, $\alpha = -0.5$ slope relative to the much bluer $\alpha = 1/3$ 'difference' composite spectrum, in agreement with (Shakura & Sunyaev 1973). However, Kokubo et al. caution that their method does not attempt to estimate or account for non-variable components of the host galaxy. Despite this work being the closest analogue to the present study available in the literature, the still different methodologies and assumptions make a concrete comparison of derived spectral slopes hazardous. It is possible though that the bluer slope derived in this work is found because we removed the static host galaxy component, which would otherwise produce a redder best-fit spectral slope.

Entirely serendipitously, we find our best-fit spectral slope $\alpha = 0.71 \pm 0.02$ to be statistically consistent with the recent theoretical prediction $\alpha = 5/7 \simeq 0.71$ for a think disc heated predominantly by stresses conveyed by magnetic links between the inner disc and black hole (Agol & Krolik 2000; Mummery & Balbus 2020). We caution however that our result (1) assumes that all quasars have SMC-like dust and (2) is sensitive to the tail end of the χ^2 distribution used to compute the total χ^2 from Fig. 10. At face value, however, these stress-heated $\alpha = 5/7$ thick disc models cannot be ruled out by our results.

Our finding is intriguing, as it hints at additional accretion physics not considered in many previous studies. The possible importance of magneticc links connecting the inner disc with the black hole is noted by Thorne (1974). Consequences including a steeper disc temperature profile, $\tau \propto \lambda^{8/7}$, and bluer spectral shape, $L_\nu \propto \nu^{5/7}$, are worked out in detail for steady accretion by Agol & Krolik (2000), and for time-dependent accretion, as in tidal disruption events, by Mummery & Balbus (2020). These studies employ a fully relativistic framework of accretion discs, finding a temperature structure driven by energy liberated by both local viscous dissipation and by the torques arising from magnetic linksto the spinning black hole and /or to gas inside the ISCO plunging toward the event horizon. This non-vanishing ISCO stress term augments the gravitational energy released by the inspiraling disc material, thus steepening the temperature structure. The steeper $T \propto r^{-7/8}$ temperature profile results in a bluer spectral slope, $L_\nu \propto \nu^{5/7}$, than in the traditional steady-state accretion disc model with $T \propto r^{-3/4}$ and $L_\nu \propto \nu^{1/3}$.

5.4 Comparison with intensive disc reverberation mapping

An independent method being used to probe accretion disc temperature profiles is to obtain intensive (sub-day) monitoring and then measure interband time-delays (Cackett, Horne & Winkler 2007; Edelson et al. 2019). This intensive disc reverberation mapping method (IDRM) assumes light traveltime delays $\tau \approx r/c$, and blackbody emission peaking near $\lambda \approx (h\,c/k\,T)$. A reverberating disc with $T \propto r^{-b}$ gives time-delays $\tau \propto \lambda^{1/b}$, and disc spectra $L_\nu \propto \nu^{(2b-3)/b}$. The standard disc model with $b = 3/4$ predicts $\tau \propto \lambda^{4/3}$ and $L_\nu \propto \nu^{1/3}$, while a steeper temperature profile with $b = 7/8$ predicts $\tau \propto \lambda^{8/7}$ and $L_\nu \propto \nu^{5/7}$.

The IDRM results to date are typically described as consistent with the standard disc prediction, $b = 3/4$, but with uncertainties large enough to admit $b = 1$. The most accurate IDRM results to date, from the Space Telescope and Optical Reverberation Mapping campaign monitoring of NGC 5548 in 2014, give a best-fitting power-law slope $b = 1.03 \pm 0.12$ from cross-correlation lags (Fausnaugh et al. 2016) or $b = 0.99 \pm 0.03$ from detailed fitting of a reverberating disc model to the light curves (Starkey et al. 2017). This corresponds to a steeper temperature profile, closer to $b = 7/8$ ($\alpha = 5/7$) than to 3/4 (1/3).

A caveat, however, is that the disc sizes inferred from IDRM are typically larger than expected, by a factor of ~3, and an excess lag in the Balmer continuum region suggests that bound-free emission from the (larger) broad emission-line region (BLR) may be contributing

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

significantly to the cross-correlation lags (Korista & Goad 2001; Lawther et al. 2018). Perhaps the clearest example is the lag spectrum from *HST* monitoring of NGC 4593 (Cackett et al. 2018). Work is underway to understand how best to disentangle the disc and BLR contributions to the measured lags.

6 SUMMARY AND CONCLUSIONS

In this work, we have separated the variable accretion disc light from their static host galaxies using broad-band photometric light curves, de-reddened their SEDs to account for dust, and leveraged the results to test accretion disc physics.

(i) We developed a method for decomposing quasar light curves by separating the contribution of the variable light from the static, background emission. Each disc SED was then de-reddened to provide a dust-free estimation of the underlying accretion disc light.

(ii) Of the five dust laws examined in this work, we find that the featureless laws of the Small Magellanic Clouds (SMC; Gordon et al. 2003) and starburst galaxies (SBG; Calzetti et al. 2000) are the most reasonable attenuation models as measured by their χ^2/N, with the SMC being slightly preferable as it typically requires less attenuation than the SBG.

(iii) Assuming an SMC-like dust attenuation, the best-fitting spectral slope α is found to be inconsistent with a standard $L_\nu \propto \nu^{1/3}$ corresponding the steady-state accretion model (Shakura & Sunyaev 1973). Instead, we find significant evidence for a $L_\nu \propto \nu^{5/7}$ accretion slope based on our best-fitting $\alpha = 0.71 \pm 0.02$, in agreement with the proposed disc model of Mummery & Balbus (2020).

If it is indeed the case that these ISCO-stress models better reflect the reality of accretion physics compared to previous models, then these observational findings challenge commonly made assumptions about the thermal structure of quasar accretion dics. Moreover, they have implications for deriving key properties of black holes and their accretion discs, including the Eddington luminosity and black hole mass. The observed variability of quasar discs might be plausibly attributed to changes in the magnetic links modulating ISCO stress and black hole spin. Future work is needed to confirm this model, requiring independent observations and continued monitoring of key sources in a way which these results can be confidently replicated.

Finally, we note that the methodology developed here should be useful for analysis of quasar variability data from the LSST, and could also be applied to data from multiobject spectroscopic monitoring surveys such as SDSS-RM.

ACKNOWLEDGEMENTS

The authors would like to thank Chelsea Macleod, Marianne Vestergaard, Michael Goad, and Julian Krolik for helpful discussions. We would also like to thank the organisers of the 2019 Quasars in Crisis meeting for facilitating so many useful conversations. Lastly, we would like to thank the anonymous referee for their generous insight and helpful comments which improved this work.

The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. JRW acknowledges support from the European Research Council (ERC) Consolidator Grant funding scheme (project ConTExt, grant No. 648179), and from the University of St Andrews Undergraduate Research Assistant Scheme. KH acknowledges support from STFC grant ST/R000824/1.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

DATA AVAILABILITY

The data underlying this paper were accessed from Southern Survey of Stripe 82 Quasars as part of the Sloan Digital Sky Survey. The derived data generated in this research will be shared on a reasonable request to the corresponding author.

REFERENCES

Agol E., Krolik J., 2000, ApJ, 528, 161
Allen D. A., Glass I. S., 1976, ApJ, 210, 666
Asmus D., 2019, MNRAS, 489, 2177
Bellm E. C. et al., 2018, Publ. Astron. Soc. Pac., 131, 018002
Bessell M., Murphy S., 2012, Publ. Astron. Soc. Pac., 124, 140
Brown M. J. I. et al., 2014, ApJS, 212, 18
Cackett E. M., Horne K., Winkler H., 2007, MNRAS, 380, 669
Cackett E. M., Chiang C.-Y., McHardy I., Edelson R., Goad M. R., Horne K., Korista K. T., 2018, ApJ, 857, 53
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Casey C. M., Narayanan D., Cooray A., 2014, Phys. Rep., 541, 45
Draine B. T., Malhotra S., 1993, ApJ, 414, 632
Edelson R. et al., 2019, ApJ, 870, 123
Elvis M., Wilkes B. J., Tananbaum H., 1985, ApJ, 292, 357
Fausnaugh M. M. et al., 2016, ApJ, 821, 56
Fitzpatrick E. L., 1986, AJ, 92, 1068
Fitzpatrick E. L., Massa D., 1986, ApJ, 307, 286
Förster Schreiber N. M., Wuyts S., 2020, ARA&A, 58, 661
Gallerani S. et al., 2010, A&A, 523, A85
Gaskell C. M., Goosmann R. W., Antonucci R. R. J., Whysong D. H., 2004, ApJ, 616, 147
Gordon K. D., Clayton G. C., Misselt K. A., Landolt A. U., Wolff M. J., 2003, ApJ, 594, 279
Graham M. J. et al., 2019, Publ. Astron. Soc. Pac., 131, 078001
Hönig S. F. et al., 2013, ApJ, 771, 87
Hopkins P. F. et al., 2004, AJ, 128, 1112
Ivezić e. et al., 2019, ApJ, 873, 111
Kokubo M., Morokuma T., Minezaki T., Doi M., Kawaguchi T., Sameshima H., Koshida S., 2014, ApJ, 783, 46
Korista K. T., Goad M. R., 2001, in Peterson B. M., Pogge R. W., Polidan R. S., eds, ASP Conf. Ser. Vol. 224, Probing the Physics of Active Galactic Nuclei. Astron. Soc. Pac., San Francisco, p. 411
Krawczyk C. M., Richards G. T., Gallagher S. C., Leighly K. M., Hewett P. C., Ross N. P., Hall P. B., 2015, AJ, 149, 203

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

Lawther D., Goad M. R., Korista K. T., Ulrich O., Vestergaard M., 2018, MNRAS, 481, 533
Lynden-Bell D., 1969, Nature, 223, 690
MacLeod C. L. et al., 2012, ApJ, 753, 106
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Malkan M. A., Sargent W. L. W., 1982, ApJ, 254, 22
Matsuoka Y., Strauss M. A., Price Ted N. I., DiDonato M. S., 2014, ApJ, 780, 162
Matthews T. A., Sandage A. R., 1963, ApJ, 138, 30
Mummery A., Balbus S. A., 2020, MNRAS, 492, 5655
Nandy K., Thompson G. I., Jamar C., Monfils A., Wilson R., 1975, A&A, 44, 195
Oke J. B., 1974, ApJS, 27, 21
Sandage A., 1965, ApJ, 141, 1560
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Schneider D. P. et al., 2010, AJ, 139, 2360
Seaton M. J., 1979, MNRAS, 187, 73
Shakura N. I., Sunyaev R. A., 1973, A&A, 500, 33
Shen Y. et al., 2015, ApJS, 216, 4
Shields G. A., 1978, Nature, 272, 706
Starkey D. et al., 2017, ApJ, 835, 65
Talia M., Cimatti A., Giulietti M., Zamorani G., Bethermin M., Faisst A., Le Fèvre O., Smolçić V., 2021, ApJ, 909, 23
Thorne K. S., 1974, ApJ, 191, 507
Wills B. J., Netzer H., Wills D., 1985, ApJ, 288, 94
York D. G. et al., 2000, AJ, 120, 1579
Zafar T. et al., 2015, A&A, 584, A100

APPENDIX A: PHYSICAL INTERPRETATION OF ATTENUATION ESTIMATES

As discussed briefly in Section 5, we do not place any priors on the allowed ranges of $E(B-V)$ parameter in our dereddening procedure. As a consequence of the prominent 2175-Å feature in the LMC and MW dust laws, estimates of $E(B-V)$ are more similar for sources where the bump is directly constrained by one of the five bands, which is shown to undulate with redshift in Fig. A1 for the LMC. This undulation is not seen, however, in the smooth dust laws of the SMC and SBG. We also find $E(B-V)$ flares up for sources at $z>3$ where the u-band falls blueward of the Lyman continuum break. This discontinuity is not included in our continuous power-law model, and so mimics an extreme reddening. We do not interprete either feature as a genuine physical phenomena of accretion discs, but merely a limitation of our model. The consequences of this are described in Section 5.

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

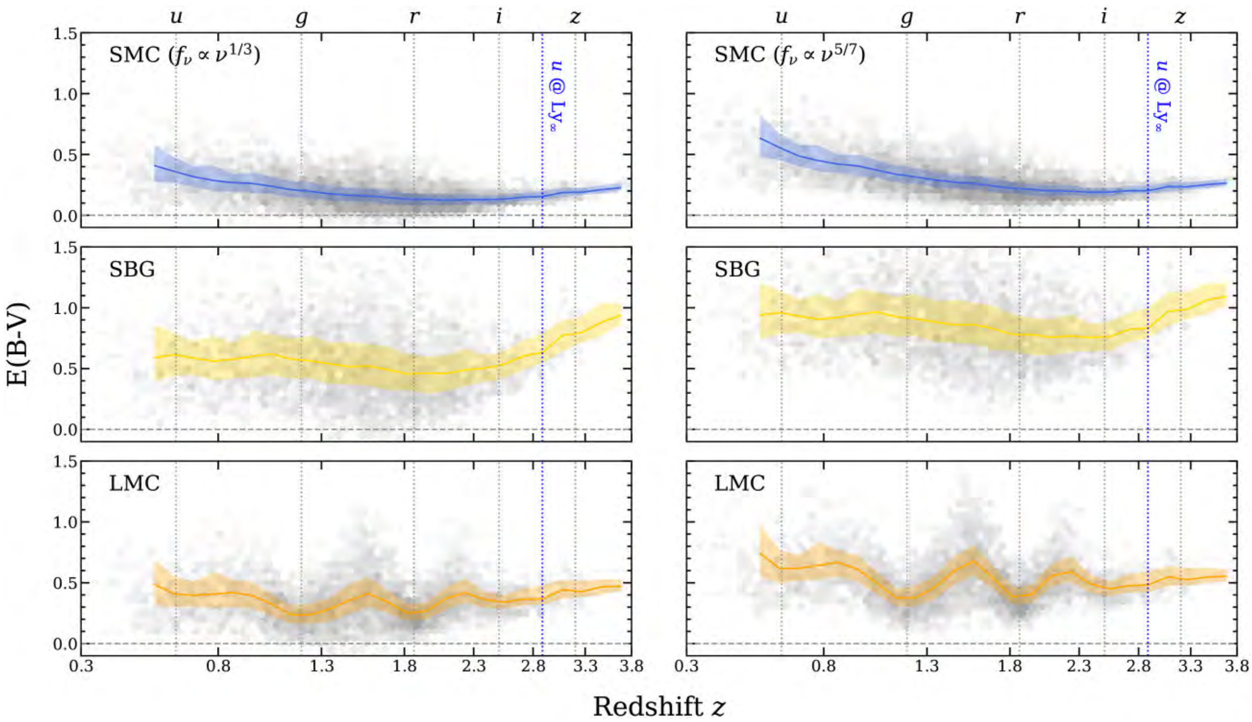

Figure A1. Binned median estimates of $E(B-V)$ over a range of redshift, assuming a $\boldsymbol{F}_\nu \sim \nu^{1/3}$ (left-hand panel) or a $\boldsymbol{F}_\nu \sim \nu^{5/7}$ (right-hand panel), for the two best-performing extinction laws (SMC, SBG) as well as the LMC which features significant absorption at 2175 Å. The redshift at which the 2175-Å feature is centred in each band is shown by the grey dotted lines. The coloured envelopes contain 68 per cent of sources in each bin.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/512/1/899/6517110 by Danmarks Veterinær- og Jordbrugsbibliotek user on 13 June 2022

SUPPLEMENTARY 4 | COSMOS2020: UV SELECTED GALAXIES AT $z \geq 7.5$

Includes the following:

COSMOS2020: UV selected galaxies at $z \geq 7.5$

Submitted to Astronomy & Astrophysics

O. B. Kauffmann, O. Ilbert, J. R. Weaver, H. J. McCracken, B. Milvang-Jensen, G. Brammer, I. Davidzon, O. Le Fèvre, D. Liu, B. Mobasher, A. Moneti, M. Shuntov, S. Toft, C. M. Casey, J. S. Dunlop, J. S. Kartaltepe, A. M. Koekemoer, D. B. Sanders, and L. Tresse

 ©ESO 2022

COSMOS2020: UV selected galaxies at $z \geq 7.5$

O. B. Kauffmann[1], O. Ilbert[1,2,3], J. R. Weaver[4,5], H. J. McCracken[6], B. Milvang-Jensen[4,5], G. Brammer[4,5], I. Davidzon[4,5], O. Le Fèvre[1], D. Liu[7], B. Mobasher[8], A. Moneti[6], M. Shuntov[6], S. Toft[4,5], C. M. Casey[9], J. S. Dunlop[10], J. S. Kartaltepe[11], A. M. Koekemoer[12], D. B. Sanders[13], and L. Tresse[1]

1 Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France
2 California Institute of Technology, Pasadena, CA 91125, USA
3 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
4 Cosmic Dawn Center (DAWN)
5 Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen, Denmark
6 Sorbonne Universités, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France
7 Max-Planck-Institut für Extraterrestrische Physik (MPE), Giessenbachstr. 1, D-85748 Garching, Germany
8 Department of Physics and Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA
9 The University of Texas at Austin 2515 Speedway Blvd Stop C1400 Austin, TX 78712 USA
10 Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, UK
11 School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester NY 14623, USA
12 Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
13 Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

Received ??? / Accepted ???

ABSTRACT

This paper presents a new search for $z \geq 7.5$ galaxies using the COSMOS2020 photometric catalogues. Finding galaxies at the reionization epoch through deep imaging surveys remains observationally challenging, and the larger area covered by ground-based surveys like COSMOS enables to discover the brightest galaxies at these high redshifts. Covering 1.4 deg^2, our COSMOS catalogues were constructed from the latest UltraVISTA data release (DR4) combined with the final *Spitzer*/IRAC COSMOS images and the Hyper-Suprime-Cam Subaru Strategic Program DR2 release. We identify 17 new $7.5 < z < 10$ candidate sources, and confirm 15 previously published candidates. Using deblended photometry extracted by fitting surface brightness models on multi-band images, we select four candidates which would be rejected using fixed aperture photometry. We test the robustness of all our candidates by comparing six different photometric redshift estimates. Finally, we compute the galaxy UV luminosity function in three redshift bins centred at $z = 8, 9, 10$. We find no clear evolution of the number densities of the brightest galaxies $M_{\mathrm{UV}} < -21.5$, in agreement with previous works. Rapid changes in the quenching efficiency or attenuation by dust could explain such lack of evolution between $z \sim 8$ and $z \sim 9$. A spectroscopic confirmation of the redshifts, already planned with *JWST* and the Keck telescopes, will be essential to confirm our results.

Key words. galaxies: high-redshift – galaxies: photometry – galaxies: distances and redshifts – galaxies: fundamental parameters – galaxies: evolution

1. Introduction

Star-forming galaxies are expected to be major contributors to the reionization process in the primordial Universe (e.g., Dayal & Ferrara 2018; Finkelstein et al. 2019). During this time, when the Universe was less than 1 Gyr old, Lyman continuum photons emitted by new-born massive stars ionised the intergalactic medium around them. The efficiency of this process is expected to depend on the density of galaxies, their star formation rate (SFR) and the fraction of ionising photons escaping from their interstellar medium (e.g., Bouwens et al. 2015a). The latest results from the Planck survey (Planck Collaboration et al. 2018) give the reionization redshift mid-point $z_{\mathrm{re}} = 7.67 \pm 0.73$, while the reionization probably ended around redshift $z = 6$ (e.g., Fan et al. 2006). However, an accurate timeline showing the interplay between the formation of the first galaxies and reionization is not established yet. At high redshifts, the galaxy UV luminosity function (UVLF) is currently the main observable to establish the cosmic star formation rate density (SFRD). The number of ionising photons generated by galaxies can be derived from the SFRD, assuming a number of Lyman continuum photons per second produced per unit SFR, and a fraction of them escaping from galaxies (e.g., Robertson et al. 2015; Vanzella et al. 2018). Furthermore, because of the fluctuation of the intergalactic medium (IGM) opacity on large scales (e.g., Kulkarni et al. 2019), the observed redshift of the end of the reionization may depend on sight-lines. Therefore, a census of star-forming galaxies at $z > 6$ in deep fields is essential to understand the origin and physics of reionization, as well as formation of the first galaxies.

Neutral hydrogen in the IGM, the circumgalactic medium and in the interstellar medium absorbs the flux blueward of the Lyman limit at 912 Å, with a flux expected to be consistent with zero below this limit (e.g., Bouwens et al. 2015b). Moreover, the IGM absorbs the light bluer than the Lyman alpha emission line at 1215.67 Å. These features have been used to select high-redshift galaxies in deep imaging surveys (Steidel et al. 1996). The search for these "Lyman Break Galaxies" (LBG) is carried out by using fluxes straddling the Lyman emission wavelength at any given redshift and detecting the amplitude of this feature (the height of the Lyman break). The colour redward of the break constrains the rest-frame UV slope, which is sensitive to star formation, dust and metallicity. Photometric redshifts are also commonly used to select high-redshift candidates, through spectral energy distribution (SED) fitting (e.g., McLure et al. 2013; Finkelstein et al. 2015; Bowler et al. 2015). This method presents the advantage of using more than three bands if available, as well as providing a redshift probability distribution function (hereafter zPDF) for each candidate.

Recent deep and wide-area surveys have provided the necessary multi-wavelength data to select high-redshift galaxies, in particular at near-infrared wavelengths ($1-2\,\mu$m). The Wide Field Camera 3 (WFC3) onboard the *Hubble* space telescope (*HST*), with its unmatched sensitivity at near-IR wavebands, has revolutionised our knowledge of the distant Universe. One of the major surveys in this context is the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011), covering a total of 750 arcmin^2 with deep *HST* imaging. The CANDELS fields include the Great Observatories Origins Deep Survey (GOODS; Giavalisco et al. 2004) with two fields (GOODS-South and GOODS-North), COSMOS (Scoville et al. 2007), the Extended Groth Strip (EGS; Davis et al. 2007), and the UKIDSS Ultra-deep Survey field (UDS; Cirasuolo et al. 2007). Furthermore, the *Hubble* Ultra-Deep Field (HUDF) covering 4.7 arcmin^2 (Illingworth et al. 2013) is the deepest astronomical images ever taken, reaching depths of 30 mag in the optical and near-infrared. Alternatively, the *Hubble* Frontier Fields (HFF; Lotz et al. 2017) include deep imaging of six massive galaxy clusters, which are used to study extremely faint background galaxies magnified through gravitational lensing. Using these multi-wavelength surveys, over a thousand galaxies with $z > 6$ were identified, reaching $z \sim 10$ (e.g., McLeod et al. 2016; Oesch et al. 2014, 2018; Bouwens et al. 2015b, 2019).

Based on galaxy candidates detected in these *HST* deep fields, there is evidence that the Schechter function (Schechter 1976) still provides a reasonable fit to the UV luminosity function (UVLF) at high redshift (e.g., Schmidt et al. 2014; Bouwens et al. 2015b). The shape of the UVLF appears to evolve with redshift, with a steepening of the faint-end slope α from -1.6 at $z = 4$ to -2.3 at $z = 10$ (e.g., Bouwens et al. 2015b; Finkelstein et al. 2015; Oesch et al. 2018). The observed steep slope for the UVLF at high redshifts predicts a high abundance of faint galaxies in the Universe during the reionization epoch. However, there is no consensus whether the observed evolution of the UVLF at $z > 6$ is driven by changes in the characteristic absolute magnitude or in the normalisation (e.g., McLure et al. 2013; Finkelstein et al. 2015). This may be due to an evolving shape of the LF towards a non-Schechter form. When converted in SFRD, most of the studies are consistent with a steady decrease of the SFRD from its peak at $z = 2-3$ to $z = 10$ (e.g., Madau & Dickinson 2014; Finkelstein 2016). Oesch et al. (2018) find that this decrease is even more rapid between $z = 8$ and $z = 10$ by more than a factor ten, possibly explained by the fast build-up of the underlying dark matter mass function in the primordial Universe.

While an extremely powerful technique to study the faint galaxy populations, the current deep pencil-beam surveys do not cover sufficient cosmological volumes to capture the rare and bright sources which would constrain the bright end of the UVLF. For instance, Oesch et al. (2018) analyse 800 arcmin^2 of archival data, but do not constrain the LF brighter than $M_{\rm UV} < -21.2$. Bridge et al. (2019) find eight galaxies brighter than $M_{\rm UV} < -21.5$ at $7 < z < 8$ in the Brightest of Reionising Galaxies (BoRG) survey, a parallel *HST* survey specially designed to observe the brightest galaxies. In this regime, deep ground-based imaging surveys are invaluable. The combination of deep optical (27 – 28 mag) and near-infrared (25 – 26 mag) imaging over degree-scale area has made it possible to isolate the brightest galaxies in the $z > 7-8$ Universe (e.g., Bowler et al. 2014). Sufficiently deep *Spitzer* coverage with the Infrared Array Camera (IRAC) at 3.6 and 4.5 μm complements these observations, allowing efficient removal of the intermediate-redshift contaminants (e.g., Oesch et al. 2012), as well as direct detection of high-redshift sources.

Using the available data from UltraVISTA survey (McCracken et al. 2012) in the COSMOS field, multiple samples of galaxy candidates were already identified at $z > 6$ (Bowler et al. 2014, 2015) and $z \geq 7.5$ (Stefanon et al. 2017, 2019, hereafter S19), imposing strong constraints on the bright end of the galaxy UVLF. Bowler et al. (2020, hereafter B20) find 27 LBGs over 6 deg^2 area combining data from the COSMOS and XMM-LSS fields. The bright end ($-23 < M_{\rm UV} < -21$) of the resulting galaxy number density is in excess compared to the exponential decline predicted from a Schechter parametrisation, suggesting a double power-law may be more appropriate. Early models have linked the UVLF evolution with the one of the dark matter halo mass functions (Cooray & Ouchi 2006; Bouwens et al. 2008; Tacchella et al. 2013). The change of shape at the bright end of the UVLF compared to the halo mass function is interpreted by the combination of several physical processes, including quenching and dust extinction (e.g., Harikane et al. 2022). Indeed, for a given massive galaxy, its host dark matter halo or its own intrinsic stellar content may not be massive enough to trigger internal star-formation quenching (Peng et al. 2010). Moreover, dust extinction is known to bend the bright end of the UVLF at intermediate redshift (Reddy et al. 2010). The formation of dust in the interstellar medium (ISM) of galaxies at $z > 6$ may still be immature, thus high SFR galaxies at this epoch may experience less attenuation (Finkelstein et al. 2015). This would result in a UVLF that is well populated at the bright end, describing a double power-law (Bowler et al. 2014, 2015).

With the COSMOS2020 photometric catalogues (Weaver et al. 2022), we have the opportunity to improve the search and identify the brightest star-forming galaxies at the epoch of reionization. Candidates are identified using deep near-infrared imaging from UltraVISTA DR4, complemented with IRAC images from the Cosmic Dawn Survey (Moneti et al. 2021) to detect the galaxy rest-frame optical emission. While B20 used the latest UltraVISTA release (DR4), we use in addition the latest deep optical images from the public DR2 of the Hyper Suprime-Cam (HSC) Subaru Strategic Program (HSC-SSP; Aihara et al. 2018) and the final *Spitzer* images in the mid-infrared. We also validate our results using the recently released DR3 of the HSC-SSP (Aihara et al. 2021). These data are essential to improve the purity of the high-redshift sample and to extend the area under investigation. Furthermore, COSMOS2020 photometric extractions have been made with two different techniques: a traditional approach using aperture photometry and a surface brightness profile-fitting technique using multi-band images. In this paper, we combine both photometric catalogues, and multiple photometric redshift codes, leading to a more robust final sample of candidates.

In this work, we search for galaxies at $z \geq 7.5$ in the COSMOS field using the COSMOS2020 photometric catalogue, and we estimate the galaxy rest-frame UVLF at $8 \leq z \leq 10$. The paper is structured as follows. Section 2 presents the imaging data used in this work, as well as the estimated source photometry. In Sect. 3 describes the high-redshift galaxy selection. The sample

Table 1: Depth of the deepest optical and infrared broad bands used in this work.

Band	$m_{3\sigma,\,\mathrm{AB}}$[a] ultra-deep[b]	$m_{3\sigma,\,\mathrm{AB}}$ deep	Source
u	27.8	...	CFHT/MegaCam
u^*	27.7	...	...
g	28.1	27.8	Subaru/HSC
r	27.8	27.4	...
i	27.7	27.2	...
z	27.2	26.7	...
y	26.5	25.8	...
Y	26.6	25.3	UltraVISTA
J	26.4	25.2	...
H	26.1	24.9	...
K_s	25.7	25.3	...
[3.6]	26.5	26.4	*Spitzer*/IRAC
[4.5]	26.5	26.2	...

[a] Depth at 3σ computed on PSF-homogenised images (except for IRAC images) in empty 2″ diameter apertures.
[b] The deep and ultra-deep regions are distinct for the HSC and UltraVISTA data.

of candidates is presented in Sect. 4 and compared with previous studies. The updated constraints on the high-redshift UVLF resulting from the selected galaxies are presented in Sect. 5. We discuss these results in Sect. 6 and summarise our conclusions in Sect. 7.

We adopt the standard ΛCDM cosmology with $\Omega_\mathrm{m} = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70$ km s^{-1} Mpc^{-1}. We use the initial mass function (IMF) from Chabrier (2003). Magnitudes are given the AB system (Oke 1974).

2. Data

This work is based on the COSMOS2020 photometric catalogues (Weaver et al. 2022), which includes the currently deepest optical to mid-infrared imaging in the COSMOS field. The imaging data, as well as the extracted photometry are briefly described here. Full details can be found in Weaver et al. (2022).

2.1. Imaging

The COSMOS field is covered by several multi-wavelength deep imaging surveys. While the photometric catalogue consists of observations in more than 35 bands, here we only discuss the deepest and reddest broad-band imaging, which are the most relevant datasets to search for high-redshift galaxies. The photometric depths of these broad-band images are given in Table 1. We stress that the depths are not homogeneous over the full field and differ by wavelength. We discuss the impact on the selection of the high-redshift candidates in Sect. 4.

The UltraVISTA survey (McCracken et al. 2012) provides deep near-infrared imaging over 1.5 deg^2 of the COSMOS field, as shown in Fig. 1. The exposure time is not homogeneous over the entire field. Four “ultra-deep” stripes across the field covering 0.62 deg^2 have deeper exposure times (vertical dark grey stripes in Fig. 1). We use the UltraVISTA DR4 data, with a near-infrared coverage in four broad bands, $YJHK_s$. The additional narrow-band $NB118$ covers the ultra-deep stripes.

The mid-infrared data comes from the Cosmic Dawn Survey (Moneti et al. 2021), in which all the *Spitzer* observations in the IRAC/[3.6], [4.5] bands are processed (the full list of included programs is given in Table C.1). This notably includes the COSMOS *Spitzer* survey (S-COSMOS; Sanders et al. 2007) over 2 deg^2, the *Spitzer* Large Area Survey with Hyper Suprime-Cam (SPLASH; Steinhardt et al. 2014) over 1.8 deg^2, the *Spitzer* Extended Deep Survey (SEDS; Ashby et al. 2013), the deep *Spitzer* Matching Survey of the UltraVISTA ultra-deep Stripes survey (SMUVS; Ashby et al. 2018) and Completing the Legacy of *Spitzer*/IRAC over COSMOS (COMPLETE, COMPLETE2; Labbe et al. 2016; Stefanon et al. 2018). The IRAC depths reported in Table 1 are separately computed inside the deep and ultra-deep UltraVISTA stripes. The 3σ depths of S-COSMOS data outside the SPLASH field (shown in Fig. 1) reach 25.3 mag in [3.6] and 25 mag in [4.5].

The optical data provided by the Hyper Suprime-Cam (HSC) Subaru Strategic Program (HSC-SSP; Aihara et al. 2018), include deep imaging in the g, r, i, z, y broad bands over 2.2 deg^2. We use the HSC-SSP DR2 images (Aihara et al. 2019), stacked for the COSMOS2020 catalogue. The central region of the COSMOS field includes ultra-deep HSC imaging within a 0.75 deg radius circle (blue circle in Fig. 1), and deep imaging in the extended COSMOS survey, about 0.5 mag shallower than in the centre. Since the COSMOS2020 catalogue was created, a third release of the HSC-SSP became available (Aihara et al. 2021). While these data are not included in our analysis, we used these images to insure the robustness of our candidates. In addition, we include the reprocessed Subaru Suprime-Cam images with 12 medium and two narrow bands in optical (Taniguchi et al. 2007, 2015), the u and u^* bands from the CFHT Large Area U-band Deep Survey (CLAUDS; Sawicki et al. 2019), and the UV photometry from GALEX (Zamojski et al. 2007).

2.2. Photometry

Source detection is performed in the combined `CHI_MEAN` image constructed with `SWarp` (Bertin et al. 2002) from the HSC/i, z and UltraVISTA/Y, J, H, K_s bands. This stacked detection image provides an advantage for faint, high-redshift sources compared to single images. Photometry is extracted following two independent approaches, leading to two separate photometric catalogues, called the Classic catalogue and the `The Farmer` catalogue. In the Classic approach, the photometry of the high-resolution images is performed with `SExtractor` (Bertin & Arnouts 1996) in dual-image mode. The photometry is extracted in fixed 2″ diameter apertures. To ensure that the apertures include the same features at all wavelengths, the point-spread functions (PSF) of the science images are homogenised. Multiple corrections are applied to the measured magnitudes, including the magnitude error scaling and the aperture-to-total magnitude corrections. The photometry of the low-resolution IRAC image is performed with the software `IRACLEAN` (Hsieh et al. 2012), using the high-resolution images as prior to extract the photometry of the confused sources. In this case, the IRAC PSF is iteratively subtracted from the IRAC images centred within the source boundary, as defined by the high-resolution prior. In the second approach, all bands are extracted with `The Farmer` (Weaver et al., in prep.), which uses the `The Tractor` software (Lang et al. 2016) to obtain more accurate photometry by fitting galaxy profiles with parametric models. The morphology of the sources is determined through a decision tree, separating point and extended sources. In contrast with the Classic approach, this method directly pro-

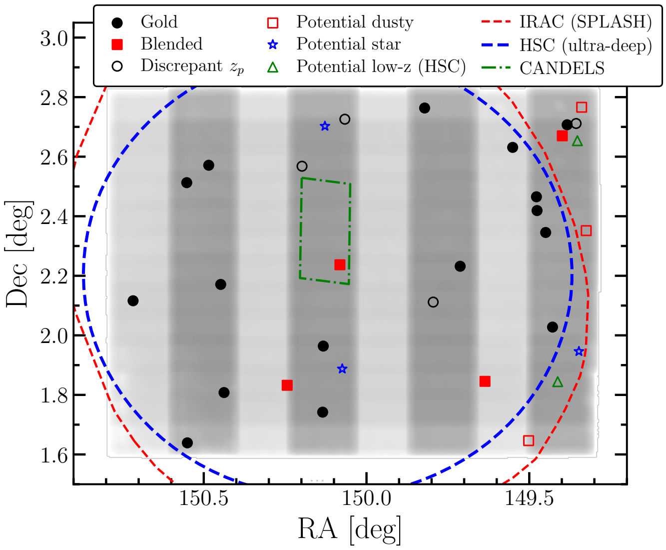

Fig. 1: Imaging data in COSMOS. The background image shows the UltraVISTA H-band weight map, where the vertical dark grey stripes represent the UltraVISTA ultra-deep stripes. The red dashed line indicates the *Spitzer*/IRAC [3.6] coverage of the SPLASH survey, the blue dashed line represents the ultra-deep region of the HSC images, and the green dot-dashed line corresponds to CANDELS. The filled circles indicate our gold sample (see Sect. 4.1). Open symbols indicate sources for which we can not exclude a low redshift or star solution based on template fitting, or because of a low S/N detection in one HSC band.

vides total magnitudes, performs an improved deblending in the high-resolution images, and extracts all the images consistently.

3. Galaxy selection criteria

Here we describe different steps taken to select a complete sample of galaxies at high redshifts. The resulting candidates are discussed in Sect. 4.

We exclusively search for candidates over the UltraVISTA area, by rejecting sources located in the masked regions near bright stars in the HSC images. This corresponds to an area of 1.404 deg^2. We require the candidates to be detected in three bands among the H, K_s, [3.6] and [4.5] bands. We require at minimum a 5σ detection in one of these bands, one at 3σ in a second band, and at 1σ in a third band. This ensures that at least two colours are reliable to estimate photo-z. Sources only detected in IRAC images are not included. We apply these selection criteria separately on both the Classic and `The Farmer` catalogues.

We reject candidates detected in any broad band blueward of the Lyman alpha break based on visual inspection, rather than explicit magnitude cuts. The measured flux may be inconsistent with zero in the case of a noise local maximum at the source coordinates. Moreover, nearby sources likely contaminate aperture fluxes for blended candidates, and generate a significant flux in the photometric apertures. In these cases, we estimate whether the observed flux originates from the detected source or from nearby sources. In the latter case, we flag these sources as blended. Artefacts are also rejected through the visual inspection of the science images, as discussed in Appendix A.

3.1. Photometric redshift selection

Our initial candidate lists are selected taking into account both the photometric redshifts and the posterior probability redshift distributions (zPDF) computed for all sources in COSMOS2020 using `LePhare` (Arnouts et al. 2002; Ilbert et al. 2006) with both the Classic and `The Farmer` catalogues. We first use the default COSMOS2020 configuration of `LePhare`, as done in Ilbert et al. (2013), Laigle et al. (2016) and Weaver et al. (2022), with the templates, dust attenuation curves and the recipe to add emission lines from Ilbert et al. (2009). The galaxy library includes 33 templates, covering various SED types, from quenched to starbursting galaxies. The most relevant ones for this work are 12 templates generated with the Bruzual & Charlot (2003) (hereafter BC03) stellar population synthesis models, with ages ranging from 30 Myr to 3 Gyr, and including sub-solar metallicities (Z = 0.004, 0.008, 0.02). We add the dust attenuation as a free parameter, allowing $E(B-V)$ to vary between 0 and 0.5 for two different attenuation curves: Calzetti et al. (2000) and Prevot et al. (1984). Several emission lines are added using an empirical relation between the UV luminosity corrected for dust attenuation and H_α emission line flux. Physically constrained ratios are considered between the intrinsic emission lines. We added [OII], H_β, [OIII], H_α emission lines, as well as Lyman$_\alpha$ despite the large uncertainties which potentially affect the modelling of this line. Dust attenuation is applied to the emission line fluxes using the same dust model as for the stellar continuum. The normalisation of the emission line fluxes are allowed to vary by a factor of two (using the same ratio for all lines) during the fitting procedure. IGM absorption is implemented following the analytical correction of Madau (1995).

As in Laigle et al. (2016), we perform the fit using fluxes (and not magnitudes), even when a source is extremely faint or non-detected in one band. Such approach is suitable as long as uncertainties are measured consistently. Therefore, we do not have to include upper-limits in the fitting procedure.

`LePhare` provides the redshift likelihood distribution for each source, after a marginalisation over the galaxy templates and the dust attenuation. We use it as the posterior redshift probability density function (zPDF), assuming a flat prior. The photo-z point estimate, $z_{\rm phot}$, is defined as the median of the zPDF. We also consider the photo-z which minimises the χ^2 over the full template library as an alternate photo-z point estimate, to strengthen our selection of high-redshift sources. We select any source with a solution at $z \geq 7.5$ with `LePhare`, in either the Classic or `The Farmer` catalogue. This is defined by imposing that both the median of the zPDF and the redshift which minimises the χ^2 are at $z \geq 7.5$.

3.2. Complementary template-fitting procedures

We produce photometric redshifts with other template-fitting procedures for the selected candidates. These estimates are not directly used to select the candidates, but are used in Sect. 4.1 to assess their robustness.

The photometric redshifts computed for all sources using `EAZY` (Brammer et al. 2008) are available in COSMOS2020. The adopted strategy is equivalent to `LePhare` and corresponds to the technique described in Weaver et al. (2022). The fitted galaxy library consists of 17 templates derived from the Flexible Stellar Population Synthesis models (Conroy et al. 2009; Conroy & Gunn 2010). Moreover, `EAZY` allows the combination of the templates in the fitting procedure.

Furthermore, we compute new photometric redshifts for the selected candidates using a different `LePhare` configuration, hereafter noted `LePhare` BC03. Such configuration is optimised to model sources with extreme colours. Since we select a really specific population with extreme properties, a larger coverage of the parameter space (in terms of dust, ages, star formation histories) should be explored to reject potential intermediate redshift contaminants. In this case, we include a set of BC03 templates assuming different star formation histories (exponentially declining and delayed), as described in Ilbert et al. (2015). Each of these 12 templates is generated at 43 ages from 50 Myr to 13.5 Gyr. During the fitting procedure, no template with an age older than the age of the Universe is considered. We assume two attenuation curves (Calzetti et al. 2000; Arnouts et al. 2013) with $E(B-V)$ varying from 0 to 2. The dust attenuation reaches $A_V \sim 8$, enabling potentially extremely dusty sources at lower redshift to be rejected (Dunlop et al. 2007). We add the emission line fluxes with a recipe described in Saito et al. (2020). For each template, we derive the number of ionising photons by integrating the SED blueward of the Lyman break, and convert it into H_β luminosity following Schaerer & de Barros (2009). Then, the ratios between H_β and other emission lines are given by Osterbrock & Ferland (2006); Anders & Fritze-v. Alvensleben (2003). Emission line fluxes are allowed to vary by a factor of two during the fitting procedure, to reproduce potential variations around the expected value. We do not use this configuration to compute photometric redshifts for the full COSMOS2020 catalogue, since the large parameter space covered by the templates increases the risk of degeneracy in the colour-redshift space and creates a larger fraction of catastrophic failures for the general population[1].

3.3. Brown dwarf contamination

The selected high-redshift candidates may be contaminated by cool Milky Way brown dwarfs, because of their similar near-infrared colours (e.g., Wilkins et al. 2014). Given the predictions by Ryan & Reid (2016), we could expect 277/deg^2 brown dwarfs at $J < 25$.

To isolate these contaminants, we fit brown dwarf templates to COSMOS2020 photometry. We include the simulated high-resolution brown dwarf templates from Baraffe et al. (2015, BT-Settl/CIFIST2011_2015) and Morley et al. (2012, 2014) in `LePhare`. The modelled emission extends to at least 10 μm in the mid-infrared. In addition, the nonphysical templates are rejected following constraints from Saumon & Marley (2008) in the effective temperature versus surface gravity space, based on the predicted evolution of cool brown dwarfs. Then, we compare the minimum χ^2 obtained with the galaxy templates (we use the standard `LePhare` configuration described in Sect. 3.1) and the one obtained with the stellar library including brown dwarf templates. We compute $\Delta\chi^2 = \chi^2_{\rm star} - \chi^2_{\rm gal}$ with $\chi^2_{\rm gal}$ being the minimum χ^2 obtained with the galaxy templates, and $\chi^2_{\rm star}$ the one obtained with the stellar templates. We select candidates with $\Delta\chi^2 > 0$.

3.4. The case of ID720309

The source ID720309 has the highest photo-z among the candidates extracted from the COSMOS2020 catalogue. However, we identified a bright star at the coordinates R.A.=09h59m10.81s, Dec.=+2d11m04.29s which could potentially generate a cross-talk signal (see Appendix A) at a position 09h59m10.81s +2d06m42.40s (assuming a native pixel scale of 0.341″ expected for the detector 7 of the VIRCAM camera). Each ultra-deep stripe is the combination of VIRCAM observations taken at three different declinations, each called "paw" (see McCracken et al. 2012). This source presents two components in one paw (paw 1), and just one component in the other (paw 2). It suggests that ID720309 in COSMOS2020 is the superposition of a cross-talk and a real source separated by 0.7″. It contributes to the signal in the 2″ aperture of the Classic catalogue. As the `The Farmer` photometry is centred on the real source, the cross-talk has less impact on the profile-fitting photometry. We also notice that the [3.6] and [4.5] positions are well aligned with the K_s position of the source in paw 2.

The star responsible for the artefact is not included in 80% of the images taken in the paw 2. Therefore, we created new stacks in Y, J, H, K_s using only these observations. This impacts the sensitivity of our dataset, but it ensures that the source is not affected by the cross-talk. Figure 2 shows the difference between the images on the full stack (top panels) and the stack not affected by the cross-talk (bottom). The main difference appears in H with a change in the shape of the source. We recomputed the flux using `The Farmer` on this new stack. The resulting magnitudes are given in Table 2, showing fainter magnitude in H and K_s, and a non-detection in the Y-band.

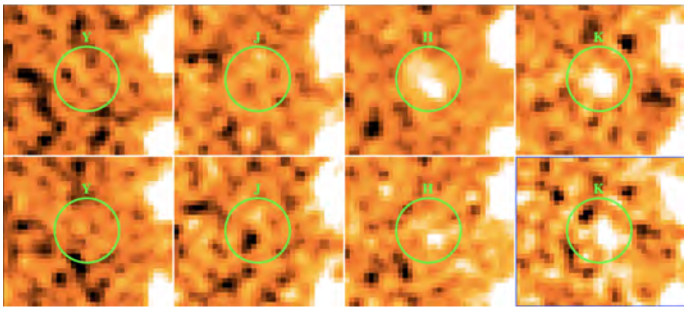

Fig. 2: Images of the source ID720309 in the four UltraVISTA broad bands (from left to right), in the full stack (top) and in the stack including only the images not contaminated by the potential cross-talk artefact (bottom). The circles represent the 2″ diameter aperture used in the Classic photometry.

4. Resulting galaxy sample

Our final galaxy sample comprises 32 candidates at $z \geq 7.5$. Based on the `LePhare`/`The Farmer` photometric redshifts, the sample includes 15 candidates in the range $7.5 < z < 8.5$, 11 candidates at $8.5 < z < 9.5$ and 1 candidates at $z > 9.5$ (some photo-z are below $z < 7.5$ in the `LePhare`/`The Farmer` configuration).

Tables 2 and 3 present the coordinates, photometry, photometric redshifts and absolute magnitudes of each object. The identifiers are from the Classic catalogue, and are always indicated starting with the letters "ID" in the following discussion. We report the magnitudes from `The Farmer`, corrected for the Milky Way extinction and systematic zero-point offsets. Appendix B describes the photometry, the best-fitting templates and the zPDF estimated with `LePhare` together with a detailed discussion for each candidate. Figure 1 illustrates the coordinates of the identified candidates over the COSMOS field. We emphasise an important aspect of our selection: four candidates are located in the UltraVISTA "deep stripes". These areas are one magni-

[1] Reducing the parameter space could be seen as a prior for the general population.

tude shallower than the ultra-deep stripes in YJH. This makes it more difficult to meet selection criteria requiring an unambiguous three-band detection.

We identify 17 new unpublished candidates at $z \geq 7.5$, many of which are located within the westernmost UltraVISTA ultra-deep stripe (at low right ascension), as shown in Fig. 1. This region is fully covered with the "deep" HSC-SSP DR2 survey, about 0.5 mag shallower on the outer part of the field compared to the centre. In contrast, the Suprime-Cam data did not cover this part of the field, which made impossible the search of $z > 7$ galaxies in this region in previous works, because of the lack of a sufficiently deep optical photometry.

We find some differences between the total apparent magnitudes between the Classic and `The Farmer` photometric COSMOS2020 catalogues, as shown in Appendix B. `The Farmer` catalogue has the advantage of photometric measurements performed consistently over the full wavelength range. Moreover, `The Farmer` measures fluxes more accurately in crowded fields. Therefore, we adopt this photometry by default for our UVLF analysis.

4.1. Robustness of the selected candidates

We produce several tests to establish the robustness of our candidates. Based on this step, we define a gold sample of 22 sources satisfying all our criteria as $z > 7.5$ galaxies (see Table 3). But we can not exclude a possible contaminant for the remaining sources, based on these additional criteria.

First, we compare several photo-z estimates to assess the sensitivity of our selection to the template-fitting approach. Table 3 presents the six photometric redshift estimates for each candidate, computed using the three different template-fitting procedures (`LePhare`, `EAZY` and `LePhare` BC03) applied to both the Classic and `The Farmer` catalogues, as described in Sect. 3.1 and 3.2. All template-fitting codes consolidate our selection for the first 16 candidates listed in Table 3. They represent the core of our gold sample. We find some discrepancies for the remaining candidates, split in three categories in Table 3.

Blended sources: for four of these candidates, the three photo-z estimates with the Classic catalogue are at $z < 3$. These sources are flagged as blended in the images (flag B in Table 3), as a bright foreground source is clearly identified and contaminates the 2″ aperture. Based on the Classic aperture photometry, all four of these candidates show a 3σ detection in at least one HSC optical band. The flux captured in the fixed aperture present a spatial offset. The Classic redshift probability distributions peak at $z \sim 2$ for all candidates, although one of them presents a secondary $z > 7$ solution. In contrast, the majority of the zPDF weights are located at $z > 7$ with `The Farmer` catalogue. It is precisely in these cases that `The Farmer` photometry is expected to be more reliable than the Classic aperture photometry, since these candidates are well deblended thanks to the profile-fitting photometry. Therefore, we consider these estimates with `The Farmer` as robust and we include these candidates in our gold sample. We note that ID859061 falls in the CANDELS region and Stefanon et al. (2019) confirmed the lack of emission in the visible from *HST* data (see their Figure 5).

Low-redshift galaxy based on `EAZY` photo-z: for two sources, we find one solution at $z < 3$ using `EAZY`. Since `EAZY` produces a high-redshift solution in either the Classic or the `The Farmer` catalogue, we decide to consider these candidates as robust and keep them in our gold sample.

Low-redshift galaxy based on `LePhare` photo-z: `EAZY` points toward a robust $z > 9$ solution with a single peak for ID1356755. However, all the solutions obtained with `LePhare` present a significant peak at $z < 3$, including more than 30% of the zPDF. Hence, we do not include this candidate in our gold sample. The configuration `The Farmer`/`LePhare` is our reference to derive the UVLF, but corresponds to a photo-z of $2.55^{+0.19}_{-0.18}$. Therefore, we do not indicate the corresponding absolute magnitude in Table 3 and this source is not used to derive the UVLF. For the source ID720309, all photo-z indicate a $z > 9$ solution with the original COSMOS2020 photometry. However, our photometry could be contaminated by a cross-talk (see Sect. 3.4). We recomputed the photo-z with `LePhare` using the new stack not contaminated by the cross-talk. The redshifts which minimise the χ^2 are $9.46^{+0.05}_{-6.59}$ and $9.47^{+1.53}_{-0.18}$ for the COSMOS and BC03 templates, respectively. We obtain $2.89^{+6.52}_{-0.43}$ and $9.46^{+0.75}_{-1.6}$ using the median of the zPDF. This galaxy remains a high-redshift candidate in our sample, but with a significant peak in the zPDF at $z \sim 2.8$.

Dusty star-forming galaxies at intermediate redshift: we checked that none of the selected candidates has $H - [3.6] > 2$ which would indicate a low-redshift galaxy with strong Balmer-break or high dust obscuration (Oesch et al. 2013). None of the candidates has coincident *Spitzer*/MIPS 24 μm emission nor at longer wavelengths in the COSMOS Super-deblended catalogue (Jin et al. 2018), and are thus unlikely to be contaminated by low-redshift dusty galaxies. We also checked that none of these sources were detected in SCUBA2 (Simpson et al. 2019) at less than 6″. Three of the candidates present a lower redshift $z < 6$ solution when fitted using the `LePhare` BC03 configuration described in 3.2. In this configuration, we allow for a possible attenuation reaching $A_V \sim 8$, which could reveal a possible solution associated to a dusty star-forming galaxies. These 3 candidates have also secondary peaks in their zPDF using other template-fitting configurations. They are located outside the ultra-deep HSC region, and outside the SPLASH coverage. Their mid-infrared photometry relies on S-COSMOS, which is about 1 mag shallower than in the centre of the field. The use of shallower data may explain the lower constraint on the zPDF resulting in multiple peaks. In addition, the width of the zPDF is systematically larger when using the Classic catalogue compared to `The Farmer`, because of larger flux errors for the same objects (as explained in Weaver et al. 2022). While our fits favour a high-redshift solution for these sources, we flag them as possible intermediate redshift contaminants and we do not include them in our gold sample.

Secondly, we also assess the robustness of the star-galaxy classification. We already rejected all candidates with a better χ^2 using the brown dwarf templates than with the galaxy templates. However, such sharp cut does not quantify the risk of degeneracy in the classification. Therefore, we generate 500 realisations of the candidates, and for each realisation we add noise to the observed flux corresponding to the photometric uncertainty in each band [2]. We compute $\Delta\chi^2$ for each realisation. The analysis of these 500 realisations allows us to quantify the robustness of the star-galaxy classification for a given source. For three sources ID485056, ID545752, ID1346929, we find that 31%, 28%, 23% (27%, 33%, 43%) of the $\Delta\chi^2$ distribution falls at $\Delta\chi^2 < 0$ with `The Farmer` (Classic). Therefore, these sources present a significant probability to be brown dwarf

[2] We note that even if the source is undetected in a given image, a flux is always provided with an associated uncertainty at the position indicated in the combined `CHI_MEAN` image.

Table 2: Coordinates and observed photometry of the selected $z \geq 7.5$ candidates. The first columns indicate the ID and coordinates from the Classic catalogue. The other columns give the photometry from `The Farmer` catalogue, corrected for Milky Way extinction and systematic zeropoint offsets. The upper-limits correspond to the 3σ depth of the image at the source coordinates (10σ for IRAC), for sources with S/N < 1.

ID Classic	R.A. [J2000]	Dec. [J2000]	ID `Farmer`	Y [mag]	J [mag]	H [mag]	K_s [mag]	[3.6] [mag]	[4.5] [mag]	Flag[a]
234500	10:02:12.08	01:38:20.22	153872	> 24.8	$25.76^{+0.13}_{-0.12}$	$25.60^{+0.10}_{-0.09}$	$25.27^{+0.10}_{-0.09}$	$24.67^{+0.03}_{-0.03}$	$24.22^{+0.02}_{-0.02}$	...
241443	09:58:00.45	01:38:46.82	228669	$27.12^{+0.41}_{-0.30}$	$25.31^{+0.08}_{-0.07}$	$24.87^{+0.07}_{-0.07}$	$24.41^{+0.06}_{-0.06}$	$23.97^{+0.03}_{-0.03}$	$23.83^{+0.02}_{-0.02}$	1
336101	10:00:32.32	01:44:31.22	499535	$26.83^{+0.15}_{-0.13}$	$25.68^{+0.06}_{-0.05}$	$25.28^{+0.05}_{-0.05}$	$25.59^{+0.10}_{-0.09}$	$24.74^{+0.02}_{-0.02}$	$24.44^{+0.02}_{-0.02}$	...
403992	10:01:45.04	01:48:28.42	214940	$27.87^{+0.52}_{-0.35}$	$25.96^{+0.08}_{-0.08}$	$26.11^{+0.13}_{-0.11}$	$25.99^{+0.15}_{-0.13}$	$25.85^{+0.11}_{-0.10}$	$25.00^{+0.05}_{-0.05}$	...
441697	09:57:39.01	01:50:40.05	598190	> 26.7	$26.02^{+0.09}_{-0.09}$	$24.96^{+0.05}_{-0.04}$	$24.89^{+0.06}_{-0.06}$	$26.16^{+0.24}_{-0.19}$	$24.30^{+0.05}_{-0.04}$	1
428351	10:00:58.48	01:49:55.97	615402	$26.79^{+0.26}_{-0.21}$	$25.27^{+0.07}_{-0.07}$	$25.51^{+0.12}_{-0.11}$	$25.26^{+0.10}_{-0.09}$	$25.01^{+0.04}_{-0.03}$	$25.07^{+0.04}_{-0.04}$	DB
442053	09:58:32.63	01:50:43.59	711973	> 25.5	$25.95^{+0.24}_{-0.20}$	$24.96^{+0.13}_{-0.11}$	$25.08^{+0.10}_{-0.09}$	$24.82^{+0.04}_{-0.04}$	$24.54^{+0.03}_{-0.03}$	DB
485056	10:00:17.89	01:53:14.39	184348	$26.99^{+0.16}_{-0.14}$	$25.99^{+0.07}_{-0.07}$	$26.29^{+0.13}_{-0.12}$	$27.24^{+0.52}_{-0.35}$	> 25.1	$27.03^{+0.21}_{-0.18}$	...
545752	09:57:23.39	01:56:45.93	758856	$26.20^{+0.10}_{-0.09}$	$24.92^{+0.03}_{-0.03}$	$24.89^{+0.04}_{-0.04}$	$25.09^{+0.07}_{-0.07}$	$25.60^{+0.17}_{-0.14}$	$25.52^{+0.09}_{-0.08}$	1
564423	10:00:31.87	01:57:50.12	78629	> 26.6	$25.95^{+0.07}_{-0.07}$	$25.40^{+0.06}_{-0.05}$	$25.43^{+0.08}_{-0.08}$	$26.18^{+0.08}_{-0.07}$	$24.83^{+0.02}_{-0.02}$	...
631862	09:57:42.84	02:01:39.64	747154	$27.27^{+0.28}_{-0.22}$	$25.89^{+0.08}_{-0.08}$	$25.59^{+0.08}_{-0.08}$	$25.42^{+0.10}_{-0.09}$	$25.27^{+0.04}_{-0.04}$	$24.80^{+0.03}_{-0.03}$	...
720309	09:59:10.82	02:06:41.96	859236	$27.76^{+0.42}_{-0.30}$	> 26.4	$25.12^{+0.05}_{-0.05}$	$24.88^{+0.05}_{-0.05}$	$24.36^{+0.02}_{-0.02}$	$24.06^{+0.02}_{-0.02}$	...
720309[b]	...	...	...	> 26.9	$27.89^{+0.59}_{-0.59}$	$25.53^{+0.1}_{-0.1}$	$25.1^{+0.1}_{-0.1}$	...	...	...
724872	10:02:52.10	02:06:57.91	729770	$26.56^{+0.36}_{-0.27}$	$25.07^{+0.09}_{-0.08}$	$24.74^{+0.09}_{-0.08}$	$25.45^{+0.12}_{-0.11}$	$25.04^{+0.04}_{-0.04}$	$24.71^{+0.04}_{-0.04}$	D
784810	10:01:47.48	02:10:15.43	749805	> 26.5	$25.72^{+0.06}_{-0.06}$	$25.69^{+0.08}_{-0.07}$	$25.82^{+0.12}_{-0.11}$	$26.02^{+0.09}_{-0.09}$	$25.16^{+0.04}_{-0.04}$	...
852845	09:58:50.94	02:13:55.09	371044	$27.13^{+0.23}_{-0.19}$	$26.01^{+0.09}_{-0.08}$	$25.25^{+0.05}_{-0.05}$	$25.32^{+0.08}_{-0.07}$	$25.03^{+0.03}_{-0.03}$	$24.30^{+0.02}_{-0.02}$	...
859061	10:00:19.59	02:14:13.28	443686	> 26.6	$26.29^{+0.11}_{-0.10}$	$25.95^{+0.11}_{-0.10}$	$26.14^{+0.19}_{-0.16}$	$24.84^{+0.03}_{-0.03}$	$24.88^{+0.03}_{-0.03}$	B
978062	09:57:47.90	02:20:43.55	764263	> 26.7	$25.07^{+0.04}_{-0.04}$	$24.71^{+0.03}_{-0.03}$	$24.72^{+0.05}_{-0.05}$	$24.60^{+0.03}_{-0.03}$	$24.20^{+0.02}_{-0.02}$	...
984164	09:57:18.00	02:21:05.90	774509	> 26.4	$26.01^{+0.09}_{-0.08}$	$25.53^{+0.08}_{-0.07}$	$25.46^{+0.10}_{-0.09}$	$24.53^{+0.24}_{-0.20}$	$23.48^{+0.10}_{-0.10}$	2
1055131	09:57:54.25	02:25:08.42	518572	> 26.3	$25.42^{+0.04}_{-0.04}$	$25.57^{+0.07}_{-0.07}$	$25.48^{+0.10}_{-0.09}$	$25.10^{+0.04}_{-0.03}$	$24.52^{+0.03}_{-0.03}$	...
1103149	09:57:54.69	02:27:54.95	71035	$28.39^{+0.75}_{-0.44}$	$25.81^{+0.06}_{-0.06}$	$25.91^{+0.09}_{-0.09}$	$25.86^{+0.14}_{-0.12}$	$25.05^{+0.03}_{-0.03}$	$24.92^{+0.05}_{-0.05}$	...
1151531	10:02:12.54	02:30:45.84	776124	$27.99^{+0.59}_{-0.38}$	$25.03^{+0.03}_{-0.03}$	$25.15^{+0.05}_{-0.05}$	$25.82^{+0.14}_{-0.13}$	$24.91^{+0.03}_{-0.03}$	$24.13^{+0.02}_{-0.02}$	...
1209618	10:00:47.53	02:34:04.50	270250	> 26.8	$25.85^{+0.07}_{-0.07}$	$25.66^{+0.08}_{-0.08}$	$26.48^{+0.26}_{-0.21}$	$25.65^{+0.07}_{-0.06}$	$25.19^{+0.04}_{-0.04}$	...
1212944	10:01:56.33	02:34:16.22	391218	$28.09^{+0.65}_{-0.40}$	$25.94^{+0.08}_{-0.08}$	$26.22^{+0.15}_{-0.13}$	$25.98^{+0.16}_{-0.14}$	$26.13^{+0.11}_{-0.10}$	$25.17^{+0.04}_{-0.04}$	...
1274544	09:58:12.23	02:37:52.34	50411	$25.69^{+0.14}_{-0.13}$	$25.09^{+0.09}_{-0.08}$	$25.01^{+0.12}_{-0.11}$	$24.87^{+0.07}_{-0.06}$	$23.86^{+0.01}_{-0.01}$	$23.50^{+0.01}_{-0.01}$	D
1297232	09:57:24.53	02:39:13.18	292281	$26.77^{+0.15}_{-0.13}$	$25.16^{+0.04}_{-0.04}$	$25.08^{+0.05}_{-0.05}$	$25.38^{+0.09}_{-0.08}$	$25.03^{+0.11}_{-0.10}$	$24.59^{+0.10}_{-0.09}$	2
1313521	09:57:35.64	02:40:12.09	454321	> 26.4	$25.86^{+0.08}_{-0.07}$	$25.21^{+0.06}_{-0.06}$	$25.21^{+0.08}_{-0.07}$	$24.03^{+0.03}_{-0.03}$	$23.66^{+0.05}_{-0.05}$	2B
1346929	10:00:30.65	02:42:09.10	764277	> 26.7	$25.71^{+0.07}_{-0.07}$	$25.36^{+0.07}_{-0.06}$	$26.11^{+0.18}_{-0.16}$	$25.26^{+0.05}_{-0.05}$	$26.76^{+0.22}_{-0.18}$	...
1352064	09:57:32.07	02:42:25.56	740295	$27.11^{+0.23}_{-0.19}$	$25.67^{+0.06}_{-0.06}$	$25.41^{+0.07}_{-0.06}$	$25.39^{+0.09}_{-0.08}$	$25.88^{+0.22}_{-0.18}$	$24.42^{+0.07}_{-0.07}$	2
1356755	09:57:25.45	02:42:41.22	9881	> 26.4	$26.67^{+0.15}_{-0.15}$	$24.65^{+0.32}_{-0.32}$	$24.53^{+0.40}_{-0.40}$	$23.68^{+0.03}_{-0.03}$	$23.87^{+0.05}_{-0.05}$	...
1371152	10:00:15.97	02:43:32.91	247885	> 26.7	$26.24^{+0.11}_{-0.10}$	$25.50^{+0.08}_{-0.07}$	$25.50^{+0.11}_{-0.10}$	$26.41^{+0.19}_{-0.16}$	$25.74^{+0.10}_{-0.09}$	...
1409328	09:59:17.15	02:45:48.22	666370	$26.60^{+0.18}_{-0.15}$	$25.90^{+0.11}_{-0.10}$	$25.50^{+0.10}_{-0.09}$	$25.31^{+0.11}_{-0.10}$	$25.30^{+0.06}_{-0.06}$	$24.38^{+0.03}_{-0.03}$	...
1412106	09:57:21.36	02:45:57.47	331814	> 26.2	$25.30^{+0.06}_{-0.06}$	$25.47^{+0.09}_{-0.08}$	$26.12^{+0.25}_{-0.20}$	$25.21^{+0.12}_{-0.11}$	$25.29^{+0.20}_{-0.17}$	2

[a] The flags are the following. 1: outside the HSC ultra-deep region, 2: outside the IRAC SPLASH region, B: blended, D: inside the UltraVISTA deep stripe

[b] UltraVISTA photometry measured on a new stack not affected by the contribution of a potential cross-talk (see Sect. 3.4).

contaminants, and we do not include them in our gold sample.

Thirdly, we inspect the HSC-SSP DR3 images in g, r, i and z for the 32 candidates. The increase in depth for DR3 compared DR2 is minor in the COSMOS field (Aihara et al. 2021). DR3 correspond to the final images taken on COSMOS and the sky subtraction has been improved in the SSP pipeline. In the z-band image at the position of each of the 32 candidates, we measure the photometric flux in a 2″ diameter aperture and the corresponding sky flux in a $2-4''$ diameter annulus using the python package `photutils`. To minimise the impact of objects in the sky annulus, we compute a sigma-clipped mean value. The noise per pixel is computed by aggressively detecting and removing all objects in the postage-stamp around the object and computing the standard deviation of the remaining pixels. We note that the precise signal-to-noise (hereafter S/N) values reported here are dependent on the exact details of object thresholding and background computation. However, we have verified that the conclusions presented below are largely robust to the exact parameter choice. In addition to the S/N values presented below, we examine both the pixels present in the measurement aperture and in the sky annulus around the object. Six candidates have S/N > 1. One of them is already identified as potential $z < 6$ dusty galaxy (ID241443). Two

Table 3: Photometric redshifts estimated with `LePhare` and `EAZY` as in Weaver et al. (2022), and with `LePhare` using BC03 templates, for the $z \geq 7.5$ candidates. The UV absolute magnitudes are derived from a Monte Carlo sampling of the galaxy zPDF, using `The Farmer` and `LePhare`. The three last columns list identifiers from S19 and B20, flags[a] and spec-z from Bouwens et al. (2021b).

ID	$z_{\rm phot}$						$M_{\rm UV}$	B20/S19	Flag	$z_{\rm spec}$
	`LePhare`		`EAZY`		`LePhare BC03`					
	Classic	`Farmer`	Classic	`Farmer`	Classic	`Farmer`				
All criteria satisfied (gold sample)										
234500	$9.15^{+0.83}_{-1.10}$	$8.20^{+0.43}_{-0.35}$	$9.40^{+0.36}_{-0.39}$	$8.14^{+0.23}_{-0.22}$	$7.53^{+2.20}_{-4.10}$	$8.48^{+0.78}_{-0.48}$	$-21.56^{+0.50}_{-0.17}$	...	...	...
336101	$7.45^{+0.42}_{-0.29}$	$7.51^{+0.09}_{-0.09}$	$7.30^{+0.16}_{-0.16}$	$7.46^{+0.07}_{-0.11}$	$7.69^{+0.45}_{-0.37}$	$7.77^{+0.07}_{-0.12}$	$-21.40^{+0.11}_{-0.10}$	213/Y3	...	7.306/R25
403992	$8.98^{+0.88}_{-0.87}$	$8.52^{+0.31}_{-0.51}$	$8.98^{+0.31}_{-0.12}$	$8.59^{+0.19}_{-0.58}$	$8.72^{+1.10}_{-3.70}$	$8.77^{+0.74}_{-0.60}$	$-21.20^{+0.25}_{-0.19}$	266	...	...
564423	$8.82^{+0.62}_{-0.61}$	$9.17^{+0.15}_{-0.17}$	$8.92^{+0.21}_{-0.17}$	$8.98^{+0.05}_{-0.03}$	$8.71^{+1.00}_{-0.73}$	$9.83^{+0.12}_{-0.19}$	$-22.00^{+0.12}_{-0.11}$	237/Y5	...	NIRSpec
631862	$8.62^{+0.67}_{-0.80}$	$8.51^{+0.46}_{-0.85}$	$8.81^{+0.26}_{-0.27}$	$7.55^{+0.10}_{-0.08}$	$8.68^{+0.96}_{-0.93}$	$8.38^{+1.00}_{-0.54}$	$-21.52^{+0.31}_{-0.30}$	...	...	...
724872	$8.03^{+1.40}_{-2.20}$	$8.15^{+0.51}_{-0.53}$	$8.57^{+0.53}_{-1.40}$	$7.64^{+0.07}_{-0.07}$	$7.53^{+1.80}_{-5.30}$	$8.04^{+0.66}_{-0.24}$	$-22.12^{+0.18}_{-0.38}$	...	D	...
784810	$8.37^{+0.52}_{-0.55}$	$8.68^{+0.23}_{-0.33}$	$8.36^{+0.29}_{-0.25}$	$8.79^{+0.10}_{-0.13}$	$8.58^{+0.77}_{-0.62}$	$9.05^{+0.72}_{-0.65}$	$-21.63^{+0.17}_{-0.16}$	598/Y10	...	...
852845	$8.50^{+0.82}_{-1.20}$	$9.16^{+0.17}_{-0.21}$	$8.74^{+0.31}_{-0.42}$	$8.96^{+0.10}_{-0.11}$	$7.83^{+1.50}_{-4.50}$	$9.29^{+0.54}_{-1.50}$	$-22.08^{+0.26}_{-0.13}$	...	...	NIRSpec
978062	$8.52^{+0.44}_{-0.55}$	$8.86^{+0.18}_{-0.36}$	$8.47^{+0.36}_{-0.42}$	$8.87^{+0.06}_{-0.07}$	$8.55^{+0.89}_{-0.51}$	$8.82^{+0.66}_{-0.63}$	$-22.05^{+0.03,c}_{-0.03}$	762/Y1	...	7.675/R18 NIRSpec
1055131	$8.69^{+0.58}_{-0.66}$	$8.36^{+0.28}_{-0.35}$	$8.86^{+0.14}_{-0.16}$	$8.40^{+0.14}_{-0.29}$	$8.72^{+0.95}_{-0.90}$	$8.67^{+0.44}_{-0.50}$	$-21.74^{+0.10}_{-0.09}$	839	...	...
1103149	$8.50^{+0.48}_{-0.85}$	$7.78^{+0.32}_{-0.16}$	$8.72^{+0.17}_{-0.19}$	$7.67^{+0.08}_{-0.07}$	$8.65^{+0.82}_{-0.88}$	$8.25^{+1.00}_{-0.38}$	$-21.31^{+0.13}_{-0.10}$	879	...	7.369/R19
1151531	$8.25^{+0.46}_{-0.53}$	$8.32^{+0.26}_{-0.37}$	$7.79^{+0.63}_{-0.16}$	$8.57^{+0.07}_{-0.09}$	$8.49^{+0.73}_{-0.55}$	$8.65^{+0.27}_{-0.61}$	$-22.02^{+0.05,c}_{-0.05}$	914/Y2	...	7.677/R36
1212944	$8.08^{+0.72}_{-0.88}$	$8.60^{+0.27}_{-0.40}$	$8.12^{+0.43}_{-0.43}$	$8.76^{+0.11}_{-0.17}$	$8.21^{+1.00}_{-4.70}$	$8.89^{+0.81}_{-0.43}$	$-21.10^{+0.28}_{-0.25}$	953/Y15	...	...
1274544	$7.95^{+0.75}_{-0.58}$	$7.36^{+0.12}_{-0.12}$	$8.00^{+0.58}_{-0.44}$	$7.34^{+0.06}_{-0.07}$	$7.99^{+1.10}_{-6.10}$	$7.67^{+0.10}_{-0.15}$	$-21.94^{+0.20}_{-0.18}$	...	D	...
1352064	$8.09^{+1.00}_{-4.20}$	$8.88^{+0.20}_{-0.41}$	$7.65^{+0.98}_{-0.42}$	$7.51^{+0.15}_{-0.06}$	$6.78^{+2.40}_{-5.30}$	$9.49^{+0.35}_{-1.50}$	$-21.86^{+0.38}_{-0.19}$	...	2	...
1409328	$9.40^{+0.84}_{-0.92}$	$8.57^{+0.57}_{-1.30}$	$9.52^{+0.23}_{-0.18}$	$8.93^{+0.11}_{-1.60}$	$7.52^{+2.30}_{-3.70}$	$7.77^{+1.80}_{-0.31}$	$-21.43^{+0.41}_{-0.55}$	...	...	...
Blended candidates (gold sample)										
428351	$1.41^{+0.38}_{-0.24}$	$7.59^{+0.14}_{-0.11}$	$1.52^{+0.12}_{-0.09}$	$7.55^{+0.05}_{-0.05}$	$1.15^{+0.82}_{-0.65}$	$7.81^{+0.74}_{-0.11}$	$-21.79^{+0.14}_{-0.13}$	301/Y4	DB	7.090/R27
442053	$2.47^{+1.80}_{-0.81}$	$8.97^{+0.50}_{-6.70}$	$2.50^{+0.59}_{-0.59}$	$9.17^{+0.16}_{-0.10}$	$2.22^{+1.10}_{-0.92}$	$8.38^{+1.20}_{-6.70}$	$-22.13^{+6.39}_{-0.47}$	...	DB	...
859061	$2.01^{+1.00}_{-0.30}$	$7.88^{+0.28}_{-0.17}$	$3.38^{+0.08}_{-0.21}$	$7.65^{+0.08}_{-0.08}$	$2.03^{+1.40}_{-0.41}$	$8.12^{+1.30}_{-0.27}$	$-20.93^{+0.28}_{-0.17}$	Y11	B	...
1313521	$2.53^{+6.60}_{-0.60}$	$8.07^{+0.42}_{-0.32}$	$2.05^{+0.30}_{-0.26}$	$1.57^{+0.03}_{-0.03}$	$2.27^{+2.80}_{-0.88}$	$7.98^{+0.32}_{-6.40}$	$-21.47^{+0.21}_{-0.16}$	...	2B	...
Potential low-redshift galaxy (based on `EAZY` photometric redshift) (gold sample)										
1209618	$9.48^{+0.72}_{-0.68}$	$8.38^{+0.34}_{-0.37}$	$0.07^{+0.02}_{-0.03}$	$8.37^{+0.28}_{-0.27}$	$8.16^{+1.90}_{-2.70}$	$8.39^{+0.61}_{-0.34}$	$-21.47^{+0.18}_{-0.24}$	Y8	...	...
1371152	$8.49^{+1.10}_{-1.50}$	$9.33^{+0.19}_{-0.20}$	$8.72^{+0.30}_{-0.27}$	$0.39^{+0.01}_{-0.01}$	$7.75^{+1.70}_{-5.30}$	$9.72^{+0.21}_{-0.25}$	$-21.89^{+0.16}_{-0.14}$	Y12	...	...
Potential low-redshift galaxy (based on `LePhare` photometric redshift)										
720309	$9.71^{+0.52}_{-0.40}$	$10.10^{+0.26}_{-0.16}$	$9.60^{+0.32}_{-0.27}$	$11.20^{+0.23}_{-0.40}$	$10.00^{+0.51}_{-2.10}$	$10.40^{+0.33}_{-0.27}$	$-22.42^{+0.12}_{-0.11}$	...	...	...
720309[b]	...	$2.89^{+6.52}_{-0.43}$	...	...	...	$9.46^{+0.75}_{-1.6}$	...	...	...	...
1356755	$9.11^{+0.41}_{-6.73}$	$2.55^{+0.19}_{-0.18}$	$9.62^{+0.35}_{-0.41}$	$9.46^{+0.06}_{-0.05}$	$8.89^{+0.97}_{-6.63}$	$2.51^{+0.25}_{-0.1}$	...	...	...	NIRSpec
Potential $z < 6$ dusty galaxies										
241443	$7.09^{+0.82}_{-5.00}$	$7.70^{+0.63}_{-0.21}$	$2.09^{+5.10}_{-0.37}$	$11.90^{+0.06}_{-4.30}$	$2.23^{+5.50}_{-1.10}$	$8.05^{+0.85}_{-0.25}$	$-21.86^{+8.77}_{-0.19}$	...	1	...
984164	$8.15^{+0.94}_{-3.60}$	$8.47^{+0.38}_{-0.44}$	$7.84^{+1.00}_{-5.60}$	$8.37^{+0.25}_{-0.34}$	$4.51^{+4.40}_{-3.40}$	$8.37^{+0.51}_{-0.33}$	$-21.54^{+0.19}_{-0.19}$	...	2	...
1412106	$7.88^{+1.10}_{-5.70}$	$8.10^{+0.37}_{-0.28}$	$8.34^{+0.36}_{-0.39}$	$7.95^{+0.25}_{-0.13}$	$4.37^{+4.60}_{-3.70}$	$8.47^{+0.65}_{-0.46}$	$-21.81^{+0.11}_{-0.11}$	...	2	...
Potential star										
485056	$7.81^{+0.88}_{-4.10}$	$7.45^{+0.11}_{-0.11}$	$8.17^{+0.40}_{-0.43}$	$7.34^{+0.07}_{-0.08}$	$7.46^{+1.60}_{-5.90}$	$7.65^{+1.10}_{-0.18}$	$-21.03^{+0.14}_{-0.15}$	356	...	...
545752	$7.79^{+0.94}_{-0.31}$	$7.59^{+0.05}_{-0.05}$	$7.51^{+0.09}_{-0.10}$	$7.54^{+0.03}_{-0.03}$	$8.41^{+0.93}_{-0.86}$	$7.74^{+0.04}_{-0.04}$	$-22.11^{+0.07}_{-0.06}$	...	1	...
1346929	$7.83^{+0.59}_{-0.35}$	$8.22^{+0.90}_{-0.41}$	$7.67^{+0.14}_{-0.11}$	$9.10^{+0.05}_{-0.03}$	$1.23^{+7.90}_{-0.74}$	$9.34^{+0.12}_{-1.30}$	$-21.37^{+0.14}_{-0.13}$	1032	...	...
Potential local noise in HSC DR3 images										
441697	$9.10^{+0.35}_{-0.67}$	$9.51^{+0.11}_{-0.16}$	$9.22^{+0.18}_{-0.21}$	$9.50^{+0.06}_{-0.09}$	$9.11^{+0.71}_{-7.60}$	$9.94^{+0.09}_{-0.22}$	$-22.47^{+0.10}_{-0.09}$	...	1	NIRSpec
1297232	$7.72^{+0.77}_{-1.10}$	$7.60^{+0.06}_{-0.07}$	$7.57^{+0.74}_{-0.55}$	$7.62^{+0.04}_{-0.04}$	$7.79^{+0.91}_{-6.50}$	$7.80^{+0.06}_{-6.20}$	$-21.91^{+0.08}_{-0.07}$	...	2	...

[a] The flags are the following. 1: outside the HSC ultra-deep region, 2: outside the IRAC SPLASH region, B: blended, D: inside the UltraVISTA deep stripes.
[b] `LePhare` photo-z computed using the UltraVISTA photometry uncontaminated by the cross-talk (see Sect. 3.4).
[c] Absolute magnitudes computed using the ALMA spec-z. We report the error associated with the H-band apparent magnitude.

galaxies (ID1297232 and ID441697) have a z-band S/N of 1.4 and 2.7, respectively, and are listed separately in Table 3. The three remaining S/N > 1 candidates are not considered since the bright pixels are shifted by 1″ with the expected source position, or the signal comes from only one pixel.

To summarise, the first 16 candidates satisfy all the criteria to be selected as $z > 7.5$ galaxies. Six additional candidates present discrepancies between the different photo-z procedures, explained by blending in the Classic catalogue and one discrepant photo-z in the EAZY run. We still consider them as robust and include them in our gold sample. For the last 8 candidates in Table 3, there is a significant probability that these sources are either intermediate redshift galaxies or brown dwarf contaminants.

4.2. Comparison with $z > 7.5$ candidates from the literature

We find 6 candidates with a match in the A^{3}COSMOS ALMA catalogue[3] (Liu et al. 2019) from the Reionisation Era Bright Emission Line Survey (REBELS, Bouwens et al. 2021b). Five of them present $[CII]_{158\mu m}$ line detection with a spectroscopic redshift from Schouws et al. (in prep.) provided in Table 3. Only one ALMA source (ID564423/REBELS-24) does not present a $[CII]_{158\mu m}$ line detection, either because of too weak SFR, or observations which are still missing for this source (Bouwens et al. 2021b). These **five** spec-z confirm the high-redshift nature $z > 7$ of our candidates. We find a systematic over-estimate of the photo-z. For the Classic catalogue, these spec-z are consistent with the zPDF (with a PIT[4] of 0.3, 0.05, 0.10, 0.13 for ID336101, ID978062, ID1103149, ID1151531, respectively).

We compare our candidates to those previously identified by S19 and B20 in the COSMOS field based on ground-based imaging. S19 (see also Stefanon et al. 2017) used the near-infrared broad and narrow bands from UltraVISTA DR3, all the available CFHT/MegaCam, Subaru/Suprime-Cam optical bands and *Spitzer*/IRAC channels 1 to 4 in the mid-infrared. This study also benefited from *HST*/WFC3 coverage from the Drift And SHift mosaic (DASH; Momcheva et al. 2016; Mowla et al. 2019), with an improved spatial resolution. The HSC imaging was not available at the time of their initial sample selection but the authors used the HSC-SSP DR1 data to validate their candidate list, checking that there was no significant detection in the optical. S19 identified 16 galaxy candidates in the COSMOS field, including 10 at $z \sim 8$ and 6 at $z \sim 9$.

B20 included the UltraVISTA DR4 data and the *Spitzer*/IRAC [3.6] and [4.5] images from SPLASH, SEDS, and SMUVS in the infrared. Optical data consisted of the CFHT/MegaCam u^*, g, r, i, z broad bands from CFHTLS, and the HSC-SSP DR1 g, r, i, z, y broad bands. The Suprime-Cam z' band, deeper than the HSC/z band in DR1 release, was also used. The search was performed in the HSC ultra-deep area. B20 recovered 7 candidates from S19 and selected 9 new candidates for a total of 16 LBG at $7.5 < z < 9.1$ in the COSMOS field, including 14 at $z \sim 8$ and 2 at $z \sim 9$.

We select 15 out of the 25 high-redshift candidates from S19 and B20. The majority of these candidates present single-peaked redshift probability distributions located at $z > 7$, and are classified as galaxies in both COSMOS2020 catalogues. The identifiers from S19 and B20 are shown in Table 3. The 10 rejected candidates from S19 and B20 are described in details in Appendix B.3. Four of those candidates include strong low-redshift solutions, and six are not detected in the combined $izYJHK_s$ image. Since these galaxies are not expected to be visible in the i and z bands, the signal may be diluted in the combined $izYJHK_s$ detection image. To test the impact of such approach, we matched these six sources with the official release catalogues from UltraVISTA DR4 (McCracken et al. 2012) with a detection performed in each VIRCAM band. These sources do not have any counterpart in the K_s catalogue. One source from S19 is detected only in the H selected catalogue (but not identified in COSMOS2020 or B20 because of the proximity of a bright source $J = 20.7$). Another source from S19 is only detected in the J-band selected catalogue with a magnitude of $J = 26.3 \pm 0.1$. At best, we could have retrieved one of these six candidates with a detection performed in each VIRCAM band, rather than our combined $izYJHK_s$ detection image.

Finally, we checked the $z \geq 7$ galaxy candidates identified in CANDELS (Bouwens et al. 2015b). None of these sources are in our sample considering our selection criteria at $z \geq 7.5$. Although deeper, CANDELS covers a smaller area of 151.9 arcmin^2, which is less efficient to select the brightest sources at $z > 7.5$.

4.3. Lensing magnification

Any high-redshift galaxy sample may be subject to gravitational lensing, and in particular lensing magnification, from massive, low-redshift galaxies (e.g., Mason et al. 2015; Barone-Nugent et al. 2015; Roberts-Borsani et al. 2016). While gravitational lensing preserves surface brightness, the apparent solid angle of the background source may increase, leading to an increased apparent flux.

We investigate the possible impact of lensing magnification on our selected galaxy candidates. We search the full COSMOS2020 catalogue for massive low-redshift galaxies within a 20″ radius for each candidate galaxy[5]. Lens galaxies are modelled as singular isothermal spheres (SIS). Stellar velocity dispersions are estimated from photometry through the Faber-Jackson relation (FJR; Faber & Jackson 1976) of Barone-Nugent et al. (2015), based on the rest-frame B-band absolute magnitude and calibrated using early-type galaxies with redshifts spanning $0 < z < 1.6$. Velocity dispersion uncertainties are dominated by the intrinsic scatter in the FJR, estimated at 46 km s^{-1}. We checked that these velocity dispersion estimates are consistent with the FJR of Bernardi et al. (2003) using i-band absolute magnitudes. Photometric redshifts and absolute magnitudes in the rest-frame Suprime-Cam/B are taken from The Farmer catalogue. We restrict the lens selection to galaxies with a velocity dispersion of at least $\sigma_v = 200$ km s^{-1}, because the spectroscopic samples used to calibrate the FJR become incomplete at lower values (Barone-Nugent et al. 2015). In addition, we only include lenses with a magnification of $\mu \geq 1.1$.

We find that 8 candidates out of 32 are probably magnified with $1.1 < \mu < 1.2$, and 5 galaxies with $\mu \geq 1.2$. This includes 5 already identified candidates from S19 and B20. However, we find no evidence of strongly lensed galaxies with multiple images. The most magnified candidate is ID441697, with a cumulative magnification of $\mu = 2.36 \pm 0.80$ from five lenses within 14″, representing a boost of 0.9 mag. The main contribution comes from a $z = 0.50^{+0.01}_{-0.01}$ galaxy located within 4.6″ and with a $\sigma_v = 231$ km s^{-1} velocity dispersion, leading to a

[3] https://sites.google.com/view/a3cosmos/data

[4] Probability Integral Transform (PIT, Dawid 1984) defined as PIT = $\int_0^{z_s} \mathcal{P}(z)\,dz$.

[5] We identify lenses leading to magnifications $\mu \geq 1.1$ up to a 15″ angular distance.

$\mu = 1.37 \pm 0.20$ magnification. The second most magnified candidate is ID442053 because of two sources, one $z = 1.50^{+0.03}_{-0.04}$ galaxy at 6.0″ with $\sigma_v = 218\,\mathrm{km\,s^{-1}}$ and $\mu = 1.14 \pm 0.07$, and one $z = 1.74^{+0.04}_{-0.04}$ galaxy at 4.0″ with $\sigma_v = 207\,\mathrm{km\,s^{-1}}$ and $\mu = 1.17 \pm 0.09$. The candidate Y8 (ID1209618) has a $z = 1.26^{+0.02}_{-0.02}$ galaxy at a 4.4″ distance with a velocity dispersion of $\sigma_v = 215\,\mathrm{km\,s^{-1}}$, which gives a $\mu = 1.21 \pm 0.11$ magnification. This candidate was already identified as magnified by S19, who found a $40\,\mathrm{km\,s^{-1}}$ higher velocity dispersion and a 0.3″ smaller angular separation[6], resulting in a higher magnification $\mu = 1.39$. Three other candidates (Y1, Y10, Y12) from S19 are moderately magnified, with $1.1 < \mu < 1.2$.

Consequently, every object in our $z \geq 7.5$ sample is potentially affected by lensing magnification. However, as we shall see in Sect. 6.2, the impact of this magnification remains limited on our luminosity function measurements.

The velocity dispersion values quoted above are extracted from the FJR and their uncertainties reach $46\,\mathrm{km\,s^{-1}}$ due to the scatter of the relation (Barone-Nugent et al. 2015). Moreover, the velocity dispersion may be overestimated, because of the calibration based on early-type galaxies, leading to overestimated magnifications. Inversely, we note that the velocity dispersion lower limit of $\sigma_v = 200\,\mathrm{km\,s^{-1}}$ significantly restricts the number of lens galaxies, so that the reported total magnifications may be underestimated. This will need to be further investigated with more precise velocity dispersion estimates at fainter luminosities.

5. The UV luminosity function

In this section, we use our $z \geq 7.5$ candidates to make a new measurement of the bright end of the UVLF. We select objects brighter than $H = 25.6$, equivalent to the 5σ depth in the ultra-deep stripes. We use the $0.812\,\mathrm{deg^2}$ of the HSC-masked UltraVISTA ultra-deep area, so we do not include the 4 candidates selected in the deep area. This ensures a homogeneous selection function across the field. We take the physical parameters estimated with the `LePhare`/`The Farmer` configuration. This is necessary because the blended candidates are not included in the Classic selection.

We split our galaxy sample into three redshift bins, centred at $z = 8, 9, 10$ with a $\Delta z = 1$ width. Since the photometric redshift estimates mainly rely on infrared broad-band imaging, the redshift probability distributions are relatively broad in the interval $7 < z < 10$, and so are the photometric redshift uncertainties. As a result, the candidates with photometric redshifts at the limit between two adjacent bins may be scattered in one bin or the other. Therefore, we develop a Monte Carlo simulation to propagate the photometric uncertainties into the absolute magnitude estimates. We generate 500 realisations of the high-redshift candidates. For each realisation, we add noise to the observed flux, according to the photometric uncertainty measured in the considered band. We base our estimate of the UVLF on the 500 catalogues for which the photo-z and absolute magnitudes have been derived.

We adopt the spec-z derived with REBELS, when possible. As a result, two sources among the five are included in our first redshift range, for which we recompute the FUV absolute magnitudes.

5.1. Completeness correction

To estimate the completeness of our sample of high-redshift galaxies, we simulate point-like sources with a Moffat profile (Moffat 1969) with parameter $\beta = 3$ and a FWHM of 0.9″. We generate these objects with a uniform distribution in magnitude over $22 < H < 27$ and in position over the ultra-deep stripes area and add them to the J, H, K_s bands, avoiding placing objects on masked areas or where object are present in the segmentation map derived from the `CHI_MEAN` image (see Sect. 2.2). We then generate a new `CHI_MEAN` detection image from these simulated J, H, K_s images combined with the original i, z, Y images. Object detection is repeated using the same parameters as the main catalogue and using this new `CHI_MEAN` image. The resulting catalogue is cross-correlated with the input simulated catalogue to the find the fraction of recovered sources as a function of magnitude.

We must assume colours for the simulated galaxies. We use `LePhare` and the BC03 templates to predict the expected colours. For $E(B-V) = 0.1$, we find average colours of $J - H = 0.05, 0.49, 4.94$ at $z = 8, 9, 10$, respectively, and $K_s - H = 0.00, -0.10, -0.09$ at $z = 8, 9, 10$, respectively. The predicted $K_s - H$ colours do not vary by more than 0.25 mag at $z > 7$, depending on the assumed attenuation. We simplify the simulation by assuming $J = H = K_s$ for galaxies at $z \sim 8$, and $H = K_s$ for galaxies at $z \sim 9$ and 10, with no flux contribution in the other bands.

We estimate the completeness as a function of the H-band magnitude, defined as the fraction of recovered sources in the `CHI_MEAN` image. At $z \sim 8$, we find a drop in completeness from 84% to 60% between $H = 25$ to 25.6, the latter being our selection limit. At $z \sim 9$ and 10, the completeness drops from 72% to 37% in the same magnitude range.

We do not attempt to correct for contamination/incompleteness in our selection method, considering that possible biases in the redshift estimate (as shown by the comparison with ALMA) dominate the uncertainty budget and can not be captured by a simulation. Moreover, the use of the zPDF to generate multiple realisations of the UVLF would make such correction difficult, since a single source is split in several redshift bins.

5.2. Binned luminosity function

Absolute UV magnitudes are computed with `LePhare` in the GALEX far-UV filter[7], as follows:

$$M_{\mathrm{UV}} = m_f - DM(z) - \mathrm{KC}(z, \mathrm{SED}) \quad (1)$$

with m_f the observed magnitude in the filter f, DM the distance modulus. KC is the sum of the k-correction k_f and a rest-frame colour term, all derived from the best-fit SED. The filter f is chosen among J, H, K_s, [3.6] to minimise the SED dependency of the KC term (Ilbert et al. 2005). For each source, we list in Table 3 the median of the M_{UV} distribution, as well as the associated 0.16 and 0.84 quantiles.

The binned luminosity function is calculated using the $V_{\max}$ estimator (Schmidt 1968). This estimator is non-parametric, al-

[6] We find an astrometric shift of 0.″34 in declination compared to S19 for the candidate Y8 (ID1209618). This is more than the shift of about 0.″15 induced by the change of astrometric calibration from UltraVISTA DR3 to DR4. We note that this candidate remains faint with $H > 25.6$, which may have led to uncertain coordinates in S19 who used DR3.

[7] With a central wavelength of 1526 Å and a full-width at half maximum of 224 Å.

though the number of bins and the bin widths are set, and implicitly assumes a uniform spatial distribution of galaxies. The number density in a given magnitude bin depends on the maximum volume $V_{\rm max}$ in which each galaxy could have been selected. This volume, for a given galaxy i, is computed as:

$$V_{\max,i} = \int_{\Omega}\int_{z_{\min,i}}^{z_{\max,i}} \frac{dV}{d\Omega dz} d\Omega dz, \tag{2}$$

where $z_{\min,i}$ and $z_{\max,i}$ are the lower and upper redshift limits in which a galaxy i can be included in the sample, and dV is the differential co-moving volume. The co-moving volume is defined as the shell between the limits of the redshift bins. The maximum redshift $z_{\max,i}$ is the redshift at which a galaxy with the same intrinsic properties would not be selected in our sample, with a flux limit set to $H = 25.6$. Thus, the luminosity function ϕ can be expressed as:

$$\phi(M)\Delta M = \frac{\mathbf{w_i}}{N_{\rm real}} \sum_{i=1}^{N(M)} \frac{1}{V_{\max,i}}, \tag{3}$$

where $N(M)$ is the number of galaxies in the magnitude bin centred at M and of width ΔM, considering all the $N_{\rm real} = 500$ realisations of the catalogues. w_i is the inverse of the completeness estimated at the H-band magnitude of the galaxy i. The associated Poisson uncertainties σ_ϕ are computed as Marshall (1985):

$$\sigma_\phi(M)\Delta M = \sqrt{\frac{1}{N_{\rm real}} \sum_{i=1}^{N(M)} \frac{\mathbf{w_i}}{V^2_{\max,i}}} \tag{4}$$

The total uncertainties are computed as the quadratic sum of the Poisson errors and cosmic variance errors, estimated following Trenti & Stiavelli (2008). Since the galaxy samples become incomplete at faint magnitudes, the LFs are computed brightward of $M_{\rm UV} = -21.5$.

The galaxy UV luminosity functions at $z = 8, 9, 10$ estimated from the selected galaxy candidates are represented in Fig. 3, 4, and 5, and tabulated in Table 4. In magnitude bins with no galaxies, we put an upper limit with the number density computed from one galaxy. Cosmic variance represents about 14 % of the Poisson uncertainties at $M_{\rm UV} < -23$, 20 % at $M_{\rm UV} = -22.75$, and 30 % at $M_{\rm UV} = -22.25$.

We show in Fig. 6 the three UVLFs. Given the magnitude limit at $H < 25.6$, we cover the brightest end of the UVLF at $-23 < M_{\rm UV} < -21.5$. There is no clear evolution of the number densities from $z = 8$ to $z = 9$, with an equivalent number of candidates at $-22.5 < M_{\rm UV} < -21.5$ in both redshift bins. At $z = 10$, the density in the single absolute magnitude bin is lower than at $z = 9$, but the median $M_{\rm UV}$ is shifted to brighter magnitudes and could follow a simple extrapolation of the LF at lower redshift. However, the constraint on the UVLF at $z \sim 10$ remains weak.

We provide a constraint on the brightest part of the UVLF and over 1.5 mag range, thus the shape of the UVLF can not be constrained from our data alone. Hence, we do not attempt a fit with a power law or a Schechter function. Nevertheless, we discuss the consistency with the various fitting forms when comparing with the literature in Sect. 5.3.

5.3. Comparison with UVLFs from the literature

The calculated UVLFs are in good agreement with results from the literature in particular at $M_{\rm UV} < -22$. This suggests that our sample is complete for the brighter magnitude bins. The most straightforward comparison is with B20 who used the same near-infrared images as this work (see Sect. 4.2). In the COSMOS field, we have 12 galaxy candidates in common. However, the authors analysed about 6 deg^2 of imaging data from both the COSMOS and the XMM-LSS fields, sampling larger co-moving volumes than in this work.

Table 4: Galaxy luminosity functions derived from all the $z \geq 7.5$ candidates. The columns indicate redshifts, central absolute magnitudes, magnitude bin widths, median absolute magnitude in the bin, number of galaxies averaged over the 500 realisations (see Sect. 5.2) and co-moving number densities.

z	$M_{\rm UV}$ [mag]	$\Delta M_{\rm UV}$ [mag]	$M_{\rm median}$ [mag]	N	ϕ [10^{-6} mag^{-1}Mpc^{-3}]
8	−22.75	0.5		0.0	< 0.35
	−22.25	0.5	−22.05	3.312	1.37 ± 0.76
	−21.75	0.5	−21.78	4.02	2.04 ± 1.12
9	−22.50	1.0	−22.12	2.714	0.82 ± 0.50
	−21.75	0.5	−21.86	3.32	3.58 ± 2.55
10	−22.50	1.0	−22.47	0.702	0.21 ± 0.26

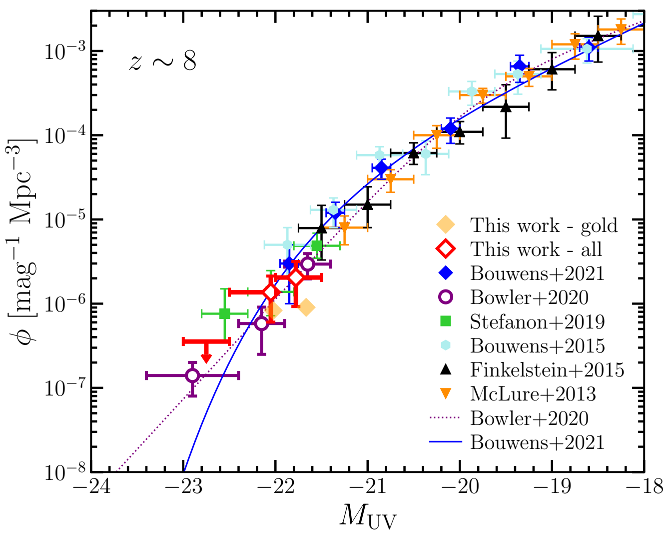

Fig. 3: Galaxy UV luminosity functions at $z = 8$. The red symbols show number densities from the galaxy sample presented in this work (uncorrected for incompleteness). The upper limits are at 1σ. The orange diamonds correspond to the measurement only based on the gold sample. The data points from B20; S19; Bouwens et al. (2021a); Bouwens et al. (2015b); Finkelstein et al. (2015); McLure et al. (2013) are represented. The best-fitting double power-law function from B20 and the Schechter function from Bouwens et al. (2021a) are displayed.

All candidates from B20 at $z < 8.5$ and $M_{\rm UV} < -22.5$ are outside the COSMOS field, and this is the reason for their better constraint on the bright end of the UVLF. Nonetheless, our number densities at $M_{\rm UV} < -22.5$ are in excellent agreement with B20 and S19.

At redshift $z \sim 10$, the number densities computed from the three candidates selected in this work are in agreement with the double power-law evolution from B20. The candidate XMM3-3085 from B20, identified in the XMM-LSS field with a photometric redshift of $z_{\rm phot} = 10.8 \pm 1.0$, is extremely bright with an absolute magnitude of $M_{\rm UV} = -23.7$ and $H = 23.9$. It is the brightest $z > 7$ galaxy candidate ever found in the litera-

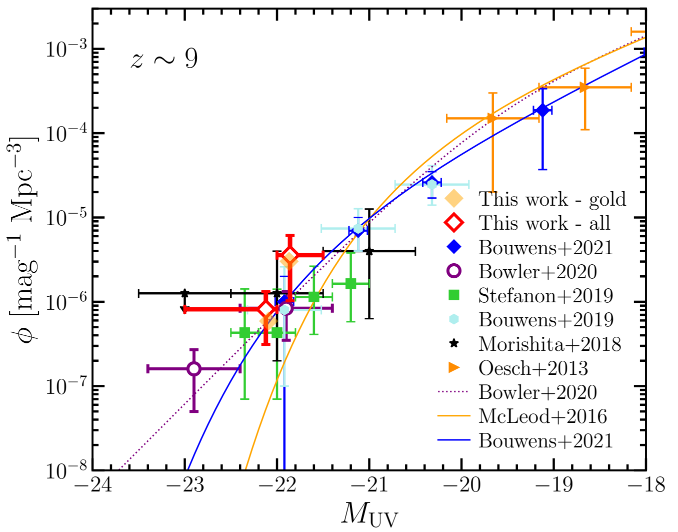

Fig. 4: Galaxy UV luminosity functions at $z = 9$ as in Fig. 3. Data points from B20; S19; Bouwens et al. (2021a); Bouwens et al. (2019); Morishita et al. (2018); Oesch et al. (2013) are shown. The best-fitting double power-law function from B20 and the Schechter function from McLeod et al. (2016) are displayed.

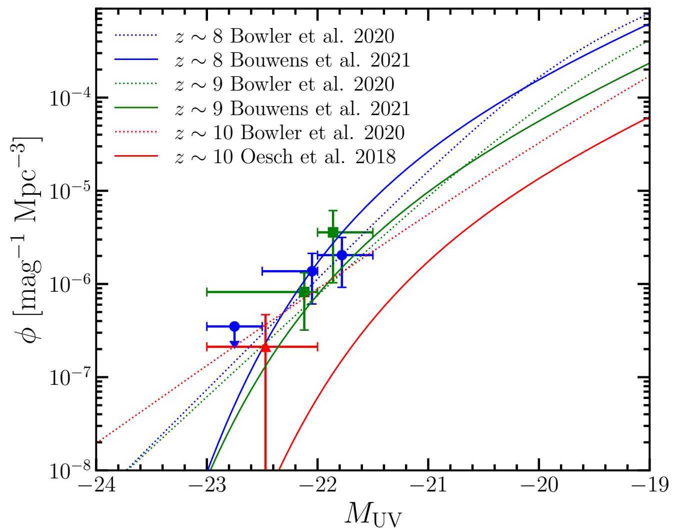

Fig. 6: Galaxy UV luminosity functions from $z = 8$ to $z = 10$. Our results are shown with blue, green, and red squares for $z = 8$, 9, 10, respectively. The dotted and solid lines are results from the literature using the same colours, for double power-law functions and Schechter functions, respectively.

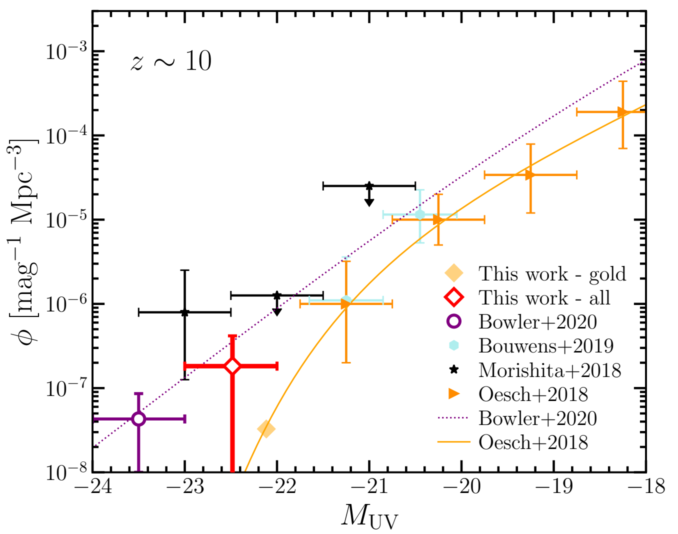

Fig. 5: Galaxy UV luminosity functions at $z = 10$, as in Fig. 3. Data points from B20; Bouwens et al. (2019); Morishita et al. (2018); Oesch et al. (2018) are shown. The best-fitting double power-law function from B20 and the Schechter function from Oesch et al. (2018) are displayed.

ture, although spectroscopic confirmation is still required. The $z \sim 10$ candidate 2140+0241–303 from Morishita et al. (2018) has an *HST*/F160W flux of 24.4 mag and an absolute magnitude of $M_{\rm UV} = -22.6$. The authors used the Brightest of Reionising Galaxies (BoRG[z9]) survey, including *HST* optical and near-infrared imaging in five broad bands over 370 arcmin^2, in addition to IRAC/[3.6] imaging. The resulting number density from that paper is an order of magnitude higher than our results at $M_{\rm UV} < -22.5$.

We compare our results with Bouwens et al. (2021a), who used the most comprehensive compilation of *HST* data taken on deep fields. Our results at $z \sim 8$ and 9 are consistent within the uncertainties. The faintest of our points at $z \sim 8$ falls below their fit by a factor two in density. At $z \sim 10$, they considered the results from Oesch et al. (2018) which indicate a sharp decline of the galaxy density at $M_{\rm UV} < -22$. Our data probe brighter absolute magnitudes (our only point is brighter than $M_{\rm UV} < -22$ while their brightest point is at $M_{\rm UV} = -21.25$). However, the lower density of our fit is unconstrained and consistent with zero. Therefore, we can not conclude on the difference with Oesch et al. (2018).

6. Discussion

6.1. Sources of contamination

The most obvious explanation for the high density of bright sources at $z \geq 7.5$ is the contamination by low-redshift galaxies or brown dwarfs wrongly classified as $z > 9$ candidates. Even if we combine several photometric redshift codes, and two different methods to extract the photometry, contamination of our sample remains a possibility. In Sect. 4.1, we found 10 candidates with a significant probability of being either dusty star-forming galaxies at intermediate redshift or brown dwarf contaminants. Seven over ten of these candidates fall on the westernmost ultra-deep stripe with a lower HSC and IRAC coverage (see Fig. 1). The larger surface density of sources in this stripe compared to the others points out toward a significant population of contaminants among these 10 galaxies.

In Fig. 3, Fig. 4 and Fig. 5, we present the UVLF after having removed these sources (gold squares). We find that the measurement at $7.5 < z < 8.5$ is severely affected by this removal, as expected since 7 sources have a photo-z estimated in this redshift range. Therefore, the density could be affected by a factor 2 if all these sources are in fact contaminants.

A systematic overestimate of the photo-z could also impact the UVLF. The 5 sources with a spec-z have a systematically higher photo-z (see Sect. 4.1). If this trend affects the whole sample, galaxies could move in a lower redshift bin. Given the shape of the redshift distribution, it would lead to an overestimation of the UVLF density. A better modelling of the galaxy properties (e.g., dust attenuation law, emission lines) could alleviate these biases. We are particularly sensitive to the modelling because the major feature used to derive the photo-z at this epoch, namely the continuum break around 1216 Å, can not be accurately located. Indeed, this break is redshifted within the gap between the ground-based Y and J-bands for galaxies between $z \sim 7.6$ and $z \sim 8.7$. Spectroscopic confirmation of the candidate redshifts is the solution to alleviate these uncertainties (Muñoz & Loeb 2008), as well as future space mission without such gap between near-infrared filters.

In addition, the uncertainties in the absolute UV magnitudes may affect the bright end through the Eddington bias (Eddington 1913). Because of the steep slope of the luminosity function, there are statistically more faint galaxies scattered into the brighter bins than the reverse, resulting in a flattened slope. The Eddington bias is also stronger for steep luminosity functions. In the selected galaxy sample, photometric redshift uncertainties can be relatively large, leading to large uncertainties once propagated to absolute magnitudes. This bias may be limited by using large magnitude bins.

The presence of Active Galaxy Nuclei (AGN) in the selected galaxy sample may also affect the estimated galaxy UVLF at the bright end. High-redshift AGN have Lyman alpha break features similar to star-forming galaxies without an AGN component. At intermediate redshifts $4 < z < 6$, the contribution from faint AGN dominates the number densities at $M_{\rm UV} < -23$, whereas it becomes negligible at $M_{\rm UV} > -22$ (Ono et al. 2018; Harikane et al. 2021). At $z > 6$, the number density of faint AGN is still uncertain. The evolution of the quasar spatial density is often parametrized as $\rho(z) \propto 10^{k(z-6)}$, with $k \simeq -0.47$ from $z = 3.5$ to $z = 5$ (Fan et al. 2001) and $k \simeq -0.72$ from $z = 5$ to $z = 6$ (Jiang et al. 2016). Recently, Wang et al. (2019) measured a consistent value $k \simeq -0.78$ from $z = 6$ to $z = 6.7$. With this accelerated redshift evolution, high-redshift AGN are sufficiently rare such that they have a negligible impact on the galaxy number density at $M_{\rm UV} = -23$ in a survey of this size. The faint-end slope of the quasar UV luminosity function is nonetheless poorly constrained at high redshift (Matsuoka et al. 2019).

6.2. Impact of magnification

As a consequence of gravitational lensing at very high redshifts, and in particular for steep luminosity functions (Mason et al. 2015), the bright end of the luminosity function is expected to be artificially higher. As discussed in Sect. 4.3, we find no evidence of strongly magnified galaxies in the selected sample, although candidates are still affected by multiple lenses leading to cumulative magnifications up to $\mu = 2.4$. We estimate the UVLF after having corrected the UV absolute magnitudes from the magnification (not attempting to correct the density). The binned UVLF at $z = 8$ and $z = 9$ remain unchanged. Thus, magnification bias does not explain the lack of evolution of the UVLF bright end at the probed magnitudes. In the $z = 10$ galaxy sample, the candidate ID441697 with a cumulative magnification of $\mu = 2.36$ becomes fainter than $M_{\rm UV} = -22$ after removing the effect of lensing. This would therefore lower by half the estimated number density at $-22.5 < M_{\rm UV} < -22$, leaving one candidate in this magnitude regime. Therefore, it would even lower down the density found at $z \sim 10$.

6.3. Shape of the UVLF

It is well established that the UVLF at $z < 6$ is well-fit with a Schechter function (e.g., Moutard et al. 2020; Bouwens et al. 2021a). At the same time, the very different shape of the bright and massive ends of the dark matter (DM) halo mass function from cosmological models and the observed galaxy luminosity function is currently explained with “quenching”, or any process which can halt star formation in these haloes and galaxies. This cessation of the star-forming activity can be due to numerous processes like AGN feedback (e.g., Croton et al. 2006; Hopkins et al. 2006) or halo quenching due to shock-heated gas in halos more massive than 10^{12}M$_\odot$ (e.g., Somerville et al. 2008; Gabor & Davé 2015). Peng et al. (2010) showed that if the quenching rate is proportional to the SFR, it would naturally lead to a galaxy luminosity function with a Schechter shape.

While the Schechter form is well established at $z < 6$, whether this holds at higher redshifts is unclear. Bowler et al. (2015) find a high density of sources at $M_{\rm UV} < -22$ and conclude that the UVLF at $z > 8$ is better fit with a power law than a Schechter function, without displaying an exponential cutoff at the bright end. We confirm the density measured by B20 and our points are superposed on their extrapolation of the UVLF with a power law. Our measurements are inconsistent with the Schechter function obtained by McLeod et al. (2016), and the density of galaxies we measure is ten times higher than expected by their extrapolation to bright magnitudes (or brighter by 0.4 mag). The brightest galaxies observed by McLeod et al. (2016) is at $M_{\rm UV} = -20.7$, not allowing them to constrain the UVLF in the bright regime probed by our data.

The latest estimates by Bouwens et al. (2021a) compile all *HST* measurements and find that the UVLF are consistent with a Schechter function at $z \sim 8-9$. The new Schechter parameters obtained by Bouwens et al. (2021a) shift the exponential cut-off at brighter magnitude than previous publications (e.g., McLeod et al. 2015). This is more consistent with the density of galaxies we find at $M_{\rm UV} < -22$. At $z \sim 8$, our faintest point falls below their fit, however such an effect could be explained by some residual incompleteness not corrected in our measurement. We find that our points are still consistent with their fit at $z \sim 9$. So, we cannot reject the Schechter fit by Bouwens et al. (2021a).

6.4. Lack of evolution at the bright end

The UVLF evolves rapidly at magnitudes fainter than $M_{\rm UV} > -20.5$. At a given density of $\phi \sim 10^{-4}$ mag^{-1}Mpc^{-3}, we expect a brightening of about 0.7 mag between $z \sim 9$ and $z \sim 8$ following the latest compilation from Bouwens et al. (2021a). Between $z \sim 9$ and $z \sim 10$, this trend is even more extreme with a brightening of 0.9 mag. This evolution is interpreted as the galaxy population following the growth of the dark matter halos with a constant star-formation efficiency (Oesch et al. 2018).

In contrast with the faint end, the density of galaxies at $M_{\rm UV} \sim -22$ is consistent with no evolution from $z \sim 8$ to $z \sim 9$ as shown in Fig. 6. This is consistent with the findings of B20.

One interpretation is that quenching efficiency increases between $z = 9$ and $z = 8$ at high mass and as a consequence we observe the building of the exponential cut-off of the UVLF. Peng et al. (2010) introduced a quenching rate proportional to the SFR to preserve a constant characteristic stellar mass around $10^{10.6}$ M$_\odot$. If we extrapolate the relation between $M_{\rm UV}$ and stellar mass found by Stefanon et al. (2021), we do not reach this regime at $z = 8$. Also, halo quenching is efficient above 10^{12} M$_\odot$ (Cattaneo et al. 2006). According to Stefanon et al. (2021), galaxies in our sample are hosted by lower mass halos. So, mass quenching is not expected to be efficient yet. However, the physical conditions at $z > 8$ may be different from those considered in lower redshift studies and the quenching mechanisms may be already effective for lower mass galaxies or DM halos.

Another possibility is that dust is generated in significant quantities between $z = 9$ and $z = 8$ (Finkelstein et al. 2015), sufficiently to decrease the light emerging from the brightest UV galaxies at $z = 8$, leading to the wrong interpretation that the density of bright galaxies remains the same between $z = 9$ and $z = 8$. Indeed, if dust is increasing with decreasing $M_{\rm UV}$, it would bend the bright end of the luminosity function, with a more pronounced dimming of the UV absolute magnitudes at bright magnitudes[8]. The mean dust content is generally expected to decrease with increasing redshift in particular for bright galaxies (e.g., Bouwens et al. 2016), because of the low metallicity of young stars in galaxies. Hence, the mean dust attenuation at $z > 9$ is often assumed to be zero (e.g., Bouwens et al. 2015b), so that the bright end of the UVLF is expected to reflect the recent star formation in the galaxy population.

7. Summary and conclusions

This paper presents one of the first science results from the COSMOS2020 catalogues: a search for candidate $z \geq 7.5$ galaxies at the epoch of the reionization of the Universe. The deep optical, near-infrared and mid-infrared imaging over the 1.4 square-degree field of COSMOS enables the detection of rare and bright galaxies, a complementary population to the much fainter galaxies found in deep pencil beam surveys like CANDELS or HUDF.

COSMOS2020 uses the latest and deepest optical, mid-infrared and near-infrared imaging in COSMOS. It includes the most recent UltraVISTA DR4 data together with new optical data from the HSC-SSP PDR2 release. We also use new images comprising all *Spitzer*/IRAC [3.6], [4.5] bands data ever taken on COSMOS from the Cosmic Dawn Survey. From these data, we extract two photometric catalogues using independent approaches: the Classic catalogue measures colours using circular aperture photometry, whilst `The Farmer` catalogue colours are measured based on a fit of surface brightness models. Sources are detected in a combined $izYJHK$ image.

The galaxy selection is primarily based on photometric redshifts estimated from SED-fitting with `LePhare` and their associated zPDF. We select sources with a robust solution at $z \geq 7.5$ with `LePhare` using either the Classic or `The Farmer` catalogues. The final sample consists of 32 candidates at $z \geq 7.5$, including 17 unpublished candidates. To assess the robustness of the COSMOS2020 photo-z estimates, we compute new sets of photometric redshifts with three different template-fitting procedures using `LePhare` and `EAZY`, with both the Classic and `The Farmer` catalogues. We isolate a gold sample of 22 out of 32 candidates which is more robust. Among them, four blended candidates are identified thanks to the profile-fitting photometry from `The Farmer`, where the contamination of light from nearby sources can be identified. These candidates have low photo-z solutions of $z < 3$ with the Classic catalogue, because of the optical flux of the nearby sources. This illustrates the effectiveness of profile-fitting techniques in deblending confused sources typically found in the search of distant galaxies. The final sample of high-redshift galaxies is therefore more complete.

From this unique list of $z \geq 7.5$ star-forming galaxies, we make a new determination of the bright end of the UV luminosity functions in three redshift bins, centred at $z = 8, 9, 10$. There is no clear evolution of the number densities from $z = 8$ to $z = 9$ between $-23 < M_{\rm UV} < -21.5$, in excellent agreement with B20. One interpretation is that quenching efficiency increases between $z = 9$ and $z = 8$. Another possibility is that dust is generated in significant quantity, sufficiently to decrease the light emerging from the brightest UV galaxies at $z = 8$. Another explanation for the high density of bright sources at $z \geq 7.5$ is the contamination by low-redshift galaxies or brown dwarfs, or a systematic overestimation of photometric redshifts. Spectroscopic confirmation is thus essential. Follow-up observations with JWST-NIRSpec (Weaver et al. 2021) are planned for five of our brightest candidates at $z > 8.5$ (see Table 3). We have also begun a systematic follow-up of these candidates with the WERLS Key Strategic Mission Support Program on Keck (Casey et al., in prep.).

This work has demonstrated the great potential of COSMOS2020 for finding rare, luminous objects in the distant Universe and the benefit from using multiple photometric extraction techniques and photometric redshifts codes to assess the robustness of the results. In the longer term, deep Euclid near-infrared observations of COSMOS and other Cosmic dawn survey fields, together with our approved 218 + 80 h, 0.6 deg^2 COSMOS-Webb proposal, will provide the ultimate survey of bright, luminous objects in the early Universe.

Acknowledgments. We are grateful to the referee for a careful reading of the manuscript and useful suggestions. This paper is dedicated to Olivier Le Fèvre who initiated this work and devoted so much energy and passion to building the COSMOS survey. We warmly acknowledge the contributions of the entire COSMOS collaboration consisting of more than 100 scientists. The HST-COSMOS program was supported through NASA grant HST-GO-09822. More information on the COSMOS survey is available at https://cosmos.astro.caltech.edu. This research is also partly supported by the Centre National d'Etudes Spatiales (CNES). OI acknowledges the funding of the French Agence Nationale de la Recherche for the project "SAGACE". HJMcC acknowledges support from the PNCG. ID has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 896225. This work used the CANDIDE computer system at the IAP supported by grants from the PNC, DIM-ACAV and the CNES and maintained by S. Rouberol. BMJ is supported in part by Independent Research Fund Denmark grant DFF - 7014-00017. The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. ST, GB and JW acknowledge support from the European Research Council (ERC) Consolidator Grant funding scheme (project ConTExt, grant No. 648179). CMC thanks the National Science Foundation for support through grants AST-1814034 and AST-2009577, and the Research Corporation for Science Advancement from a 2019 Cottrell Scholar Award sponsored by IF/THEN, an initiative of Lyda Hill Philanthropies. CMC thanks the National Science Foundation for support through grants AST-1814034 and AST-2009577, and the Research Corporation for Science Advancement from a 2019 Cottrell Scholar Award sponsored by IF/THEN, an initiative of Lyda Hill Philanthropies. The authors wish to recognise and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. This work is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO program ID 179.A-2005 and on data products produced by CALET and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium. This work is based in part on observations made with the NASA/ESA *Hubble* Space Telescope, obtained from the Data Archive at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555.

References

Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, Publications of the Astronomical Society of Japan, 71, 114

Aihara, H., AlSayyad, Y., Ando, M., et al. 2021, arXiv e-prints, arXiv:2108.13045

[8] We stress that dust is included in our modeling of the galaxy SED during the fitting process, but we do not correct the $M_{\rm UV}$ absolute magnitudes for dust attenuation.

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, Publications of the Astronomical Society of Japan, 70, S4
Anders, P. & Fritze-v. Alvensleben, U. 2003, A&A, 401, 1063
Arnouts, S., Le Floc'h, E., Chevallard, J., et al. 2013, A&A, 558, A67
Arnouts, S., Moscardini, L., Vanzella, E., et al. 2002, MNRAS, 329, 355
Ashby, M. L. N., Caputi, K. I., Cowley, W., et al. 2018, The Astrophysical Journal Supplement Series, 237, 39
Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, The Astrophysical Journal, 769, 80
Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, A&A, 577, A42
Barone-Nugent, R. L., Wyithe, J. S. B., Trenti, M., et al. 2015, MNRAS, 450, 1224
Bernardi, M., Sheth, R. K., Annis, J., et al. 2003, AJ, 125, 1849
Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 281, Astronomical Data Analysis Software and Systems XI, ed. D. A. Bohlender, D. Durand, & T. H. Handley, 228
Bouwens, R. J., Aravena, M., Decarli, R., et al. 2016, ApJ, 833, 72
Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2008, ApJ, 686, 230
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015a, ApJ, 811, 140
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015b, ApJ, 803, 34
Bouwens, R. J., Oesch, P. A., Stefanon, M., et al. 2021a, arXiv e-prints, arXiv:2102.07775
Bouwens, R. J., Smit, R., Schouws, S., et al. 2021b, arXiv e-prints, arXiv:2106.13719
Bouwens, R. J., Stefanon, M., Oesch, P. A., et al. 2019, ApJ, 880, 25
Bowler, R. a. A., Dunlop, J. S., McLure, R. J., et al. 2015, MNRAS, 452, 1817
Bowler, R. A. A., Dunlop, J. S., McLure, R. J., & McLeod, D. J. 2017, MNRAS, 466, 3612
Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2014, MNRAS, 440, 2810
Bowler, R. A. A., Jarvis, M. J., Dunlop, J. S., et al. 2020, MNRAS, 493, 2059
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Bridge, J. S., Holwerda, B. W., Stefanon, M., et al. 2019, ApJ, 882, 42
Bruzual, G. & Charlot, S. 2003, Monthly Notices of the Royal Astronomical Society, 344, 1000
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, The Astrophysical Journal, 533, 682
Cattaneo, A., Dekel, A., Devriendt, J., Guiderdoni, B., & Blaizot, J. 2006, MNRAS, 370, 1651
Chabrier, G. 2003, PASP, 115, 763
Cirasuolo, M., McLure, R. J., Dunlop, J. S., et al. 2007, MNRAS, 380, 585
Conroy, C. & Gunn, J. E. 2010, ApJ, 712, 833
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Cooray, A. & Ouchi, M. 2006, MNRAS, 369, 1869
Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11
Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, ApJL, 660, L1
Dawid, A. 1984, J R Stat Soc, 147, 278
Dayal, P. & Ferrara, A. 2018, Phys. Rep., 780, 1
Dunlop, J. S., Cirasuolo, M., & McLure, R. J. 2007, MNRAS, 376, 1054
Eddington, A. S. 1913, MNRAS, 73, 359
Faber, S. M. & Jackson, R. E. 1976, ApJ, 204, 668
Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, AJ, 132, 117
Fan, X., Strauss, M. A., Schneider, D. P., et al. 2001, AJ, 121, 54
Finkelstein, S. L. 2016, PASA, 33, e037
Finkelstein, S. L., D'Aloisio, A., Paardekooper, J., et al. 2019, ApJ, 879, 36
Finkelstein, S. L., Ryan, J. R. E., Papovich, C., et al. 2015, ApJ, 810, 71
Gabor, J. M. & Davé, R. 2015, MNRAS, 447, 374
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJL, 600, L93
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Harikane, Y., Ono, Y., Ouchi, M., et al. 2021, arXiv e-prints, arXiv:2108.01090
Harikane, Y., Ono, Y., Ouchi, M., et al. 2022, ApJS, 259, 20
Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, ApJS, 163, 1
Hsieh, B., Wang, W., Hsieh, C., et al. 2012, ApJS, 203, 23
Ilbert, O., Arnouts, S., Floc'h, E. L., et al. 2015, A&A, 579, A2
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, Astronomy and Astrophysics, 556, A55
Ilbert, O., Tresse, L., Zucca, E., et al. 2005, A&A, 439, 863
Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, ApJS, 209, 6
Jiang, L., McGreer, I. D., Fan, X., et al. 2016, ApJ, 833, 222
Jin, S., Daddi, E., Liu, D., et al. 2018, The Astrophysical Journal, 864, 56
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kulkarni, G., Keating, L. C., Haehnelt, M. G., et al. 2019, MNRAS, 485, L24
Labbe, I., Caputi, K., McLeod, D., et al. 2016, Completing the Legacy of Spitzer/IRAC over COSMOS, Spitzer Proposal ID 13094
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Lang, D., Hogg, D. W., & Mykytyn, D. 2016, Astrophysics Source Code Library, ascl:1604.008
Liu, D., Lang, P., Magnelli, B., et al. 2019, ApJS, 244, 40
Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97
Madau, P. 1995, ApJ, 441, 18
Madau, P. & Dickinson, M. 2014, ARA&A, 52, 415
Marshall, H. L. 1985, ApJ, 299, 109
Mason, C. A., Treu, T., Schmidt, K. B., et al. 2015, ApJ, 805, 79
Matsuoka, Y., Onoue, M., Kashikawa, N., et al. 2019, ApJL, 872, L2
McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156

McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156

McLeod, D. J., McLure, R. J., & Dunlop, J. S. 2016, MNRAS, 459, 3812
McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2015, MNRAS, 450, 3032
McLure, R. J., Dunlop, J. S., Bowler, R. a. A., et al. 2013, MNRAS, 432, 2696
Moffat, A. F. J. 1969, A&A, 3, 455
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Moneti, A., McCracken, H. J., Shuntov, M., et al. 2021, arXiv e-prints, arXiv:2110.13928
Morishita, T., Trenti, M., Stiavelli, M., et al. 2018, ApJ, 867, 150
Morley, C. V., Fortney, J. J., Marley, M. S., et al. 2012, ApJ, 756, 172
Morley, C. V., Marley, M. S., Fortney, J. J., et al. 2014, ApJ, 787, 78
Moutard, T., Sawicki, M., Arnouts, S., et al. 2020, MNRAS, 494, 1894
Mowla, L. A., van Dokkum, P., Brammer, G. B., et al. 2019, ApJ, 880, 57
Muñoz, J. A. & Loeb, A. 2008, MNRAS, 386, 2323
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2012, ApJ, 745, 110
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2013, ApJ, 773, 75
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2014, ApJ, 786, 108
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., & Stefanon, M. 2018, ApJ, 855, 105
Oke, J. B. 1974, ApJS, 27, 21
Ono, Y., Ouchi, M., Harikane, Y., et al. 2018, Publications of the Astronomical Society of Japan, 70, S10
Osterbrock, D. E. & Ferland, G. J. 2006, Astrophysics of gaseous nebulae and active galactic nuclei
Peng, Y.-j., Lilly, S. J., Kovač, K., et al. 2010, ApJ, 721, 193
Planck Collaboration et al. 2018, preprint, arXiv:1807.06209
Prevot, M. L., Lequeux, J., Maurice, E., Prevot, L., & Rocca-Volmerange, B. 1984, Astronomy and Astrophysics, 132, 389
Reddy, N. A., Erb, D. K., Pettini, M., Steidel, C. C., & Shapley, A. E. 2010, ApJ, 712, 1070
Roberts-Borsani, G. W., Bouwens, R. J., Oesch, P. A., et al. 2016, ApJ, 823, 143
Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJ, 802, L19
Ryan, R. E., J. & Reid, I. N. 2016, AJ, 151, 92
Saito, S., de la Torre, S., Ilbert, O., et al. 2020, MNRAS, 494, 199
Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, ApJS, 172, 86
Saumon, D. & Marley, M. S. 2008, ApJ, 689, 1327
Sawicki, M., Arnouts, S., Huang, J., et al. 2019, Monthly Notices of the Royal Astronomical Society, 489, 5202
Schaerer, D. & de Barros, S. 2009, A&A, 502, 423
Schechter, P. 1976, ApJ, 203, 297
Schmidt, K. B., Treu, T., Trenti, M., et al. 2014, ApJ, 786, 57
Schmidt, M. 1968, ApJ, 151, 393
Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2019, ApJ, 880, 43
Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, MNRAS, 391, 481
Stefanon, M., Bouwens, R. J., Labbé, I., et al. 2021, arXiv e-prints, arXiv:2103.16571
Stefanon, M., Labbe, I., Caputi, K., et al. 2018, COMPLETE2: Completing the Legacy of Spitzer/IRAC over COSMOS, Spitzer Proposal ID #14045
Stefanon, M., Labbé, I., Bouwens, R. J., et al. 2017, ApJ, 851, 43
Stefanon, M., Labbé, I., Bouwens, R. J., et al. 2019, ApJ, 883, 99
Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJ, 462, L17
Steinhardt, C. L., Speagle, J. S., Capak, P., et al. 2014, The Astrophysical Journal Letters, 791, L25
Tacchella, S., Trenti, M., & Carollo, C. M. 2013, ApJ, 768, L37
Taniguchi, Y., Kajisawa, M., Kobayashi, M. A. R., et al. 2015, Publications of the Astronomical Society of Japan, 67, 104
Taniguchi, Y., Scoville, N., Murayama, T., et al. 2007, The Astrophysical Journal Supplement Series, 172, 9
Trenti, M. & Stiavelli, M. 2008, ApJ, 676, 767
Vanzella, E., Nonino, M., Cupani, G., et al. 2018, MNRAS, 476, L15
Wang, F., Yang, J., Fan, X., et al. 2019, ApJ, 884, 30
Weaver, J. R., Brammer, G., Casey, C. M., et al. 2021, Beasts in the Bubbles: Characterizing ultra-luminous galaxies at Cosmic Dawn, JWST Proposal. Cycle 1
Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, ApJS, 258, 11
Wilkins, S. M., Stanway, E. R., & Bremer, M. N. 2014, MNRAS, 439, 1038
Zamojski, M. A., Schiminovich, D., Rich, R. M., et al. 2007, The Astrophysical Journal Supplement Series, 172, 468

Table A.1: Coordinates of the high-redshift galaxy candidates identified as cross-talk artefacts

ID CLASSIC	R.A. [J2000]	Dec. [J2000]
295952	10:01:56.01	01:42:08.37
327551	09:57:48.08	01:44:01.39
365776	10:02:16.98	01:46:16.88
454766	10:00:57.43	01:51:27.89

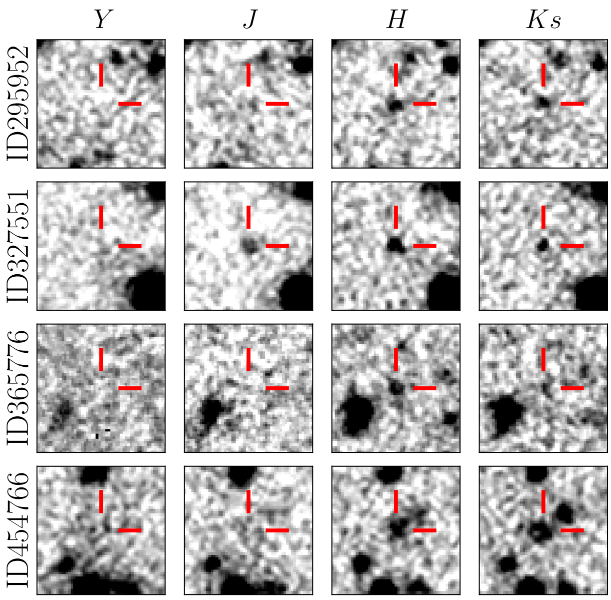

Fig. A.1: Same as Fig. B.2 for the cross-talk artefacts.

Appendix A: Artefacts

The cross-talk effect is an electronic artefact in which bright sources reappear at different locations in the detector focal plane (e.g., Bowler et al. 2017). Such artefacts are particularly problematic for high-redshift galaxy searches as they appear in the near-infrared channels only. We identified several high-redshift candidates which are most probably these inter-channel cross-talk artefacts in the UltraVISTA images. They are caused by bright sources located at the same RA with a Dec differing by $\pm k \times 128 \times 0\farcs339$ with integers $k = 1, ..., 15$ in the VIRCAM images. Here, 128 is the number of detector pixels in each of the 16 readout channels of the detector, and $0\farcs339$ is approximately the size of the original pixels. These sources are typically $11.9 - 13.8$ mag brighter than the artefacts, and were identified comparing single pawprint stacked images with similar total exposure times, in the UltraVISTA bands. Moreover, we find no evidence of an optical counterpart with HSC for these sources, or in the mid-infrared with IRAC. The coordinates of these candidates are reported in Table A.1. Figure A.1 shows the UltraVISTA stamps at the corresponding coordinates.

Appendix B: Detailed description of each candidate

From figure B.1 to B.6, we show the observed photometry of each candidate, together with the best-fitted galaxy and stellar templates, for both the CLASSIC and `The Farmer` catalogues. In the case where the flux is smaller than the flux uncertainty, for clarity the photometric measurement is replaced by a 3σ upper limit. The CLASSIC photometry and posterior redshift distributions are also shown. Figures B.2, B.3, B.5, B.7 and B.8 display stamp images centred on the candidate coordinates.

We note that there may be some offsets between the total apparent magnitudes between the CLASSIC and `The Farmer` photometric COSMOS2020 catalogues, even though the observed colours match each others. One of the main reasons for this is that `IRACLEAN` and `The Farmer` provide total fluxes, whereas fixed aperture fluxes are used in the CLASSIC catalogue. Hence, these aperture fluxes are rescaled using aperture-to-total corrections applied to all the aperture-extracted bands, and computed from the weighted mean difference between fixed aperture and pseudo-total fluxes (using Kron apertures) from `SExtractor`. While the corrections remain low (-0.29, -0.11, -0.01 mag for the 25%, 50%, 75% percentiles), this procedure introduces some additional noise for faint sources.

In IRAC, the flux are already total in both catalogues. Still, we find inconsistencies between the `IRACLEAN` and `The Farmer` photometry (e.g., for ID720309 and ID1103149), partially explained by the two different approaches in the algorithms. `The Farmer` assumes a parametric light-profile for a galaxy, convolved with the IRAC PSF before the fit. `IRACLEAN` repeatedly removes point-like source contributions from the residual map, until reaching a threshold with no pixel having a flux above a given signal-to-noise within the detection area, defined by the segmentation map. We identify that the presence of close-by objects induces large differences between the two IRAC photometry. The different deblending procedures in `The Farmer` and CLASSIC introduce small differences in the segmentation map, resulting in this case in large variations of the IRAC flux. We are not able to conclude the superiority of one catalogue over the other, highlighting the importance of using several methods to assess the robustness of the results.

Appendix B.1: New galaxy candidates

We identify 17 new candidates at $z \geq 7.5$ from the selection in both COSMOS2020 catalogues.

We find two candidates with $z_{\rm phot} > 9.5$ according to `LePhare/The Farmer` results. The candidate ID441697 is robustly detected in the J band, its photometric redshift with `The Farmer` is $z_{\rm phot} = 9.51^{+0.12}_{-0.15}$. In this case, the CLASSIC photometric redshift $z_{\rm phot} = 9.09^{+0.35}_{-0.68}$ suggests a lower redshift, nonetheless all the zPDF weight is at $z > 8$ in both catalogues. The other candidate ID720309 is detected in H and K_s bands, and has $z_{\rm phot} \sim 10.1$ for `The Farmer` photometry and $z_{\rm phot} \sim 9.7$ in the CLASSIC catalogue.

Four sources (ID241443, ID984164, ID1412106) have a primary solution at $z < 5$ when applying `EAZY` to the CLASSIC catalogue or allowing for more attenuation in the fit (BC03 configuration). We note that these sources are located outside the ultra-deep HSC region which could explain a looser constraint on the zPDF and multiple peaks in redshifts (in particular for the CLASSIC catalogue).

One new candidate, ID234500, is located in the deep stripe at the south-eastern edge of the field (at high RA). The south-eastern field edge has been covered by NIR data for UltraVISTA DR4 for the first time in COSMOS2020. This region had been previously masked because of the non-uniform quantum efficiency of the VISTA NIR detectors; this region has a higher noise, particularly in the Y-band. Moreover, new HSC DR2 data in this area allows us to select this candidate in a region which has already been studied by S19 and B20.

The observed photometry and images of best-fitted galaxy templates of two blended candidates ID442053 and ID1313521 are shown in Fig. B.4 and Fig. B.5, respectively.These candidates have magnitudes $25.0 < H < 25.5$. For these objects we find a photo-z solution at low redshift $z < 3$ with the Classic catalogue. From aperture photometry alone, all of these candidates have a 3σ detection in at least one HSC optical band. The Classic redshift probability distributions peak at $z \sim 2$ for all of them, although one shows a secondary $z > 7$ solution. In contrast, the majority of the zPDF weights are located at $z > 7$ with `The Farmer` catalogue.

Appendix B.2: Previously published galaxy candidates

The photometry of the 15 recovered, previously identified candidates are shown in Fig. B.6, and the corresponding stamp images are in Fig. B.7 and B.8.

Among the 25 previously identified candidates we find 15 of them to be at high redshift. This represents 12 of the 16 candidates selected by B20 in COSMOS, and 10 of the 16 candidates from S19 (including 3 selected in S19 but not in B20). The identifiers from S19 and B20 are given in Table 3.

For 9 of them, all photometric redshift estimates indicate a robust $z > 7$ sources, and 13 of them belong to our gold sample. The candidates 356 (ID485056) and 1032 (ID1346929) present a significant fraction of the $\Delta\chi^2$ distribution falling in the star classification (Sect. 3.3).

The candidates 301 (ID428351) and Y11 (ID859061) are blended with nearby sources and present significant optical fluxes. This was the reason for the candidate Y11 to be rejected from the B20 selection. With the Classic catalogue, the photometric redshifts are $z \sim 2$ for both sources. In contrast, these candidates are undetected in the optical with `The Farmer` catalogue, leading to $z \sim 8$. Similarly to the new blended candidates, only the photometry with `The Farmer` is considered for these sources. For the candidate 301, we note that the smaller 1.8″ diameter apertures used in B20 may have limited the impact of the nearby source on the optical photometry. For the candidate Y11, the H and K_s-band flux is clearly concentrated in one location, while the g-band emission comes from a distinct nearby source, which also contributes to the J-band aperture flux. In this case, a smaller aperture may also limit the contamination.

The estimated redshifts and absolute UV magnitudes are in excellent agreement with those from S19 and B20. We find that the absolute magnitudes from `The Farmer` are systematically fainter for blended candidates, which is expected since the profile-fitting photometry separates the fluxes from nearby sources. In contrast, we find brighter magnitudes with `The Farmer` for the two candidates 356 (ID485056) and 879 (ID1103149), which have no obvious neighbours. Nevertheless, resolved internal structures may be observed in the postage stamp images, and the Classic magnitudes are (in fact) in better agreement with B20. This situation may be due to the simple symmetric profile used to estimate the photometry in `The Farmer`, or to the background subtraction.

Candidate ID978062 is particularly bright, with a H-band magnitude of 24.5. Its relatively broad zPDF, with a full-width at half maximum (FWHM) larger than 1, results from the Lyman break located between the Y and the J band. The resulting photometric redshift is higher ($z = 8.86$ instead of 8.19) than in B20, and the absolute magnitude brighter ($M_{\mathrm{UV}} = -22.56$ instead of -22.36). This source has a spectroscopic redshift at 7.675 from the REBELS ALMA survey Bouwens et al. (2021b).

The candidate Y8 (ID1209618) has problematic `IRACLEAN` photometry with an extremely low [3.6] flux, which is not the case in `The Farmer` catalogue. In addition, the `IRACLEAN` photometric uncertainties in both [3.6] and [4.5] are much smaller than in the near-infrared bands for the Classic catalogue, so that the main constraint on the SED is an unexpectedly red IRAC colour. All but one best-fit galaxy templates produce a redshift at $z > 8$, except `LePhare` with the Classic catalogue.

Appendix B.3: Rejected candidates from the literature

We now describe the 10 candidates from S19 and B20 which were not selected in this work. For four of these candidates (Y6, Y13, 919, 1212), the estimated zPDF is double-peaked with strong low-redshift solutions. In addition, candidates Y6 and Y13 are detected at more than 2σ in the deep HSC y and z bands, respectively, and were already rejected in the B20 galaxy sample. We note that these sources were not listed as robust by S19 based on their zPDF. For these reasons, these candidates are not included. Candidate 1212 is the brightest high-redshift galaxy identified in the COSMOS field in B20, with a photometric redshift of $z = 9.1$ and an absolute UV magnitude of $M_{\mathrm{UV}} = -23$. For this candidate, both the Classic and `The Farmer` catalogues have a 3σ detection in the r band, although this is not clear from the associated postage stamp. As a consequence, it is not kept, even though it remains an interesting candidate.

Our combined detection image fails to recover six candidates (Y7, Y9, Y14, Y15, 634, 1043) from S19 and B20. The candidate Y7 is clearly visible in the $izYJHK$ detection image and listed as robust by S19 based on the zPDF; however, it is not identified as a distinct object because of two large, bright nearby sources (B20 arrived at the same conclusion), with a bright source $J = 20.7$ located at 0.4″. We matched these six sources with the official release catalogues from UltraVISTA DR4 (McCracken et al. 2012). The detection is performed in each VIRCAM band. These sources do not have any counterpart in the K_s catalogue. The source Y7 discussed previously is detected only in the H-band selected catalogue. The source Y14 from S19 is detected in the J-band selected catalogue with a magnitude of $J = 26.3 \pm 0.1$ mag, but not in the other bands.

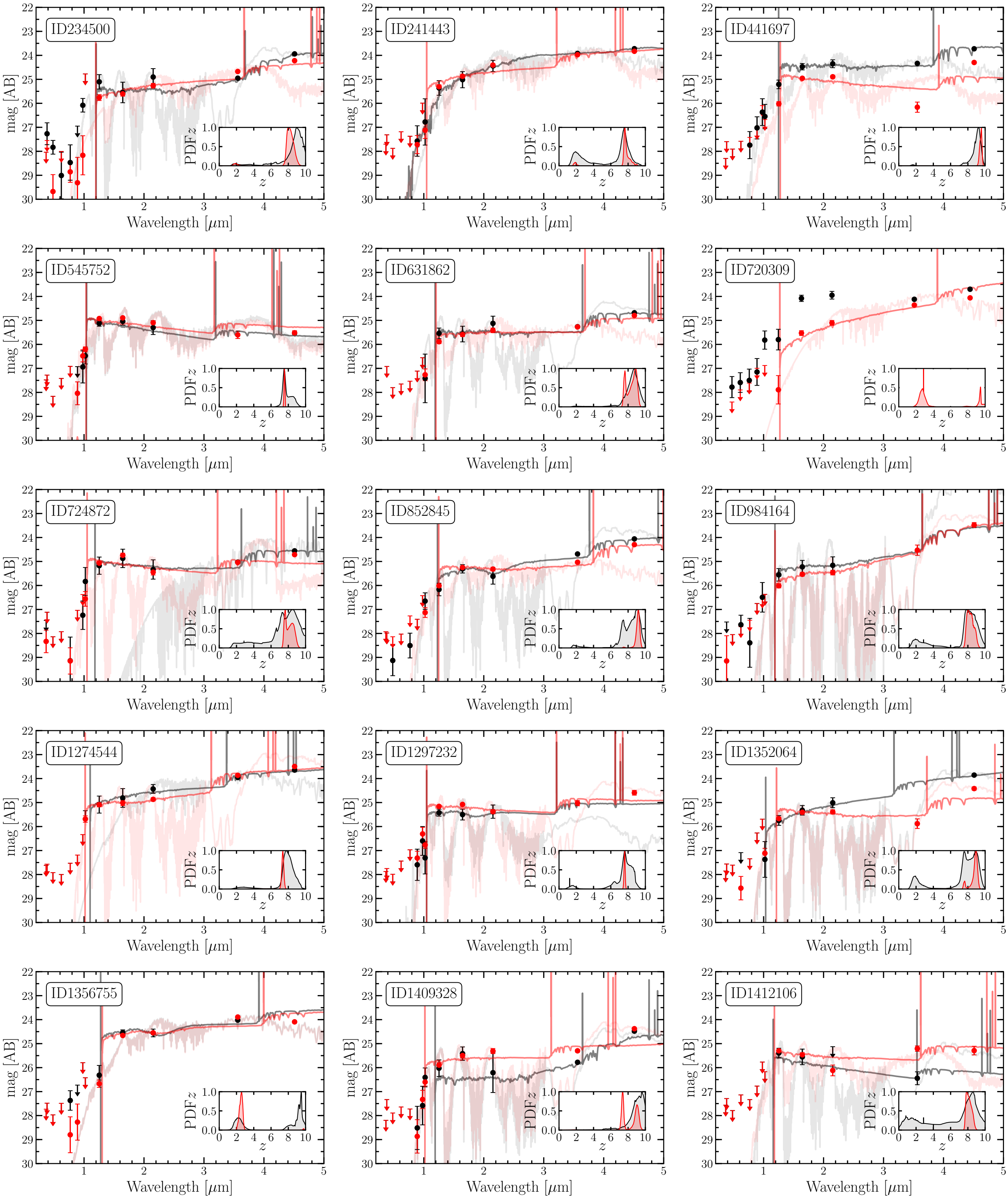

Fig. B.1: Observed photometry and best fitting templates for the new $z \geq 7.5$ candidates in the COSMOS field. The markers and lines are based on the Classic (black) `The Farmer` (red) photometry. The photometry is replaced by the 3σ upper limit for non-detection at 1σ. The bright lines show the best-fitting galaxy templates. The faint lines show the best-fitting stellar templates. The insets give the redshift probability distributions. For ID720309, `The Farmer` photometry is measured on the paw 2 stack, free from the potential cross-talk contamination. We show only the fit corresponding to this case (see Sect. 3.4).

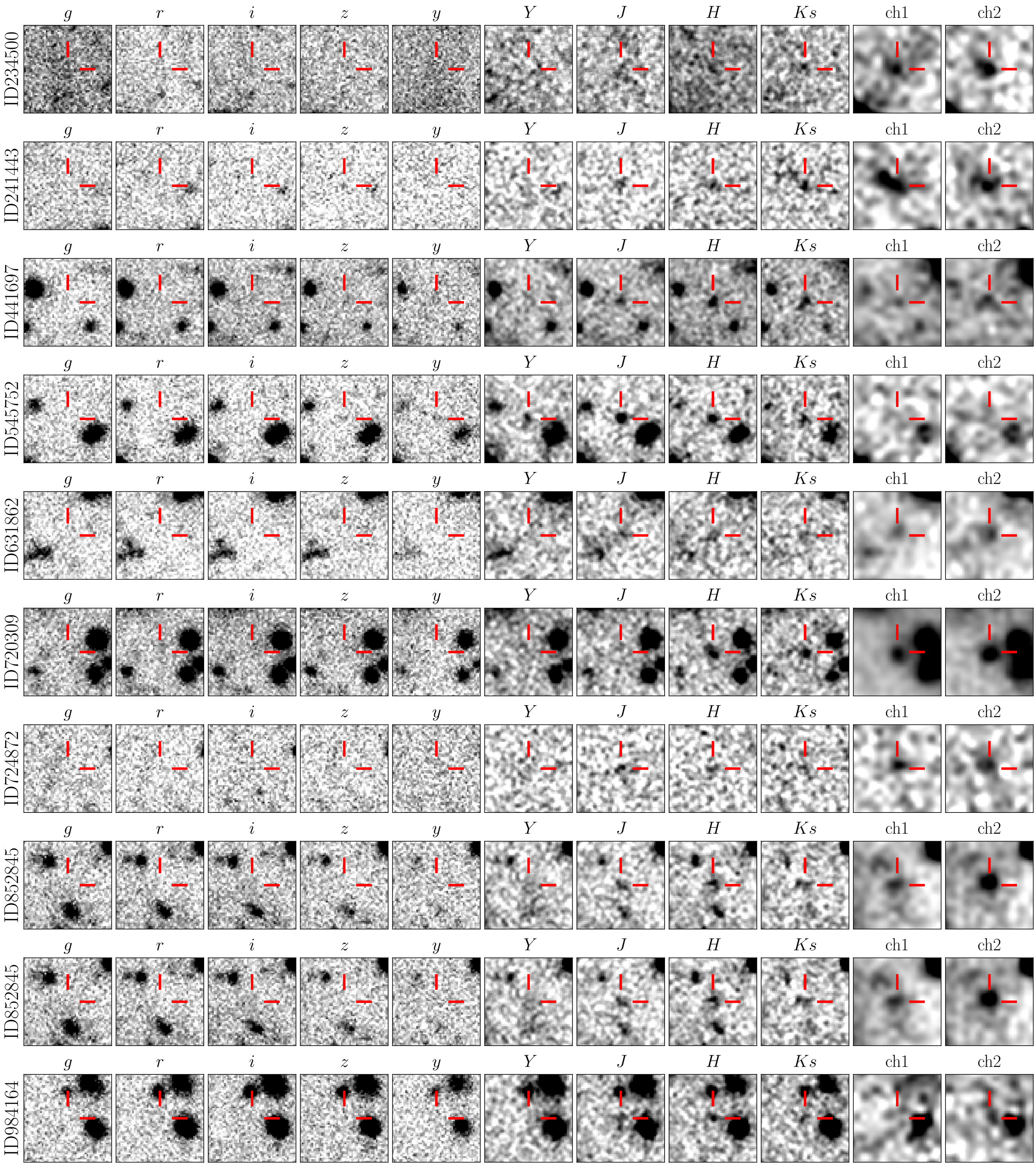

Fig. B.2: Stamp images of the $z \geq 7.5$ new candidates in the COSMOS field. Each candidate appears in one row of stamps. The stamps are 8″ wide, with North to the top and East to the left. The stamps are saturated beyond the range $[-1, 4]\sigma$, where σ is the 3σ clipped standard deviation of the pixel values in the stamp.

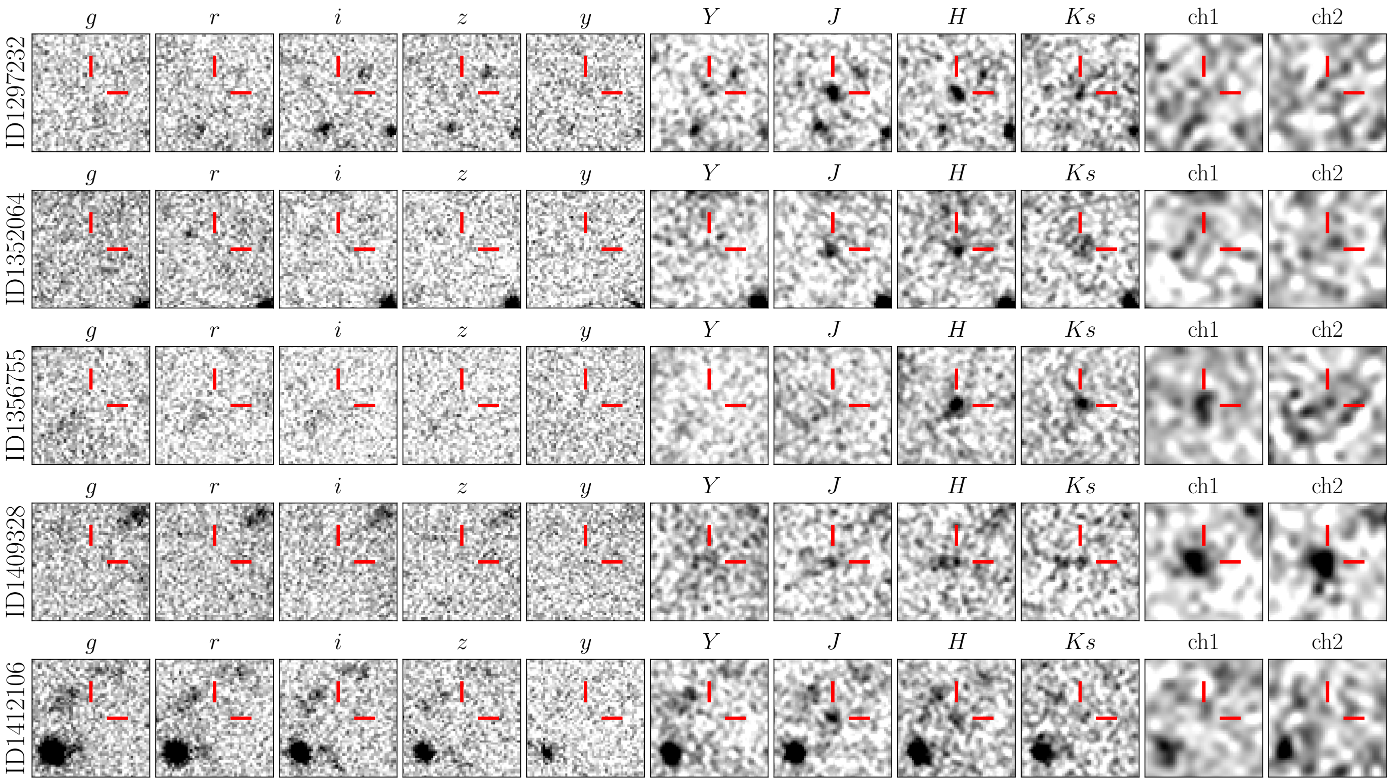

Fig. B.3: Continued

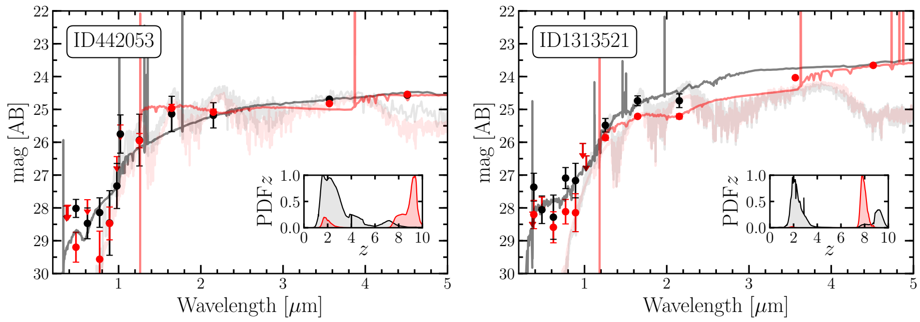

Fig. B.4: Same as Fig. B.1 for the new blended $z \geq 7.5$ galaxy candidates.

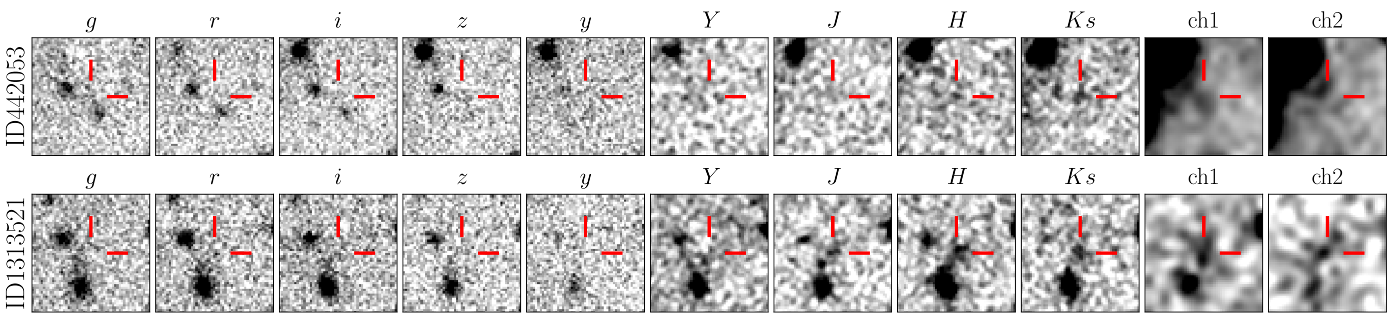

Fig. B.5: Same as Fig. B.2 for the new blended $z \geq 7.5$ galaxy candidates.

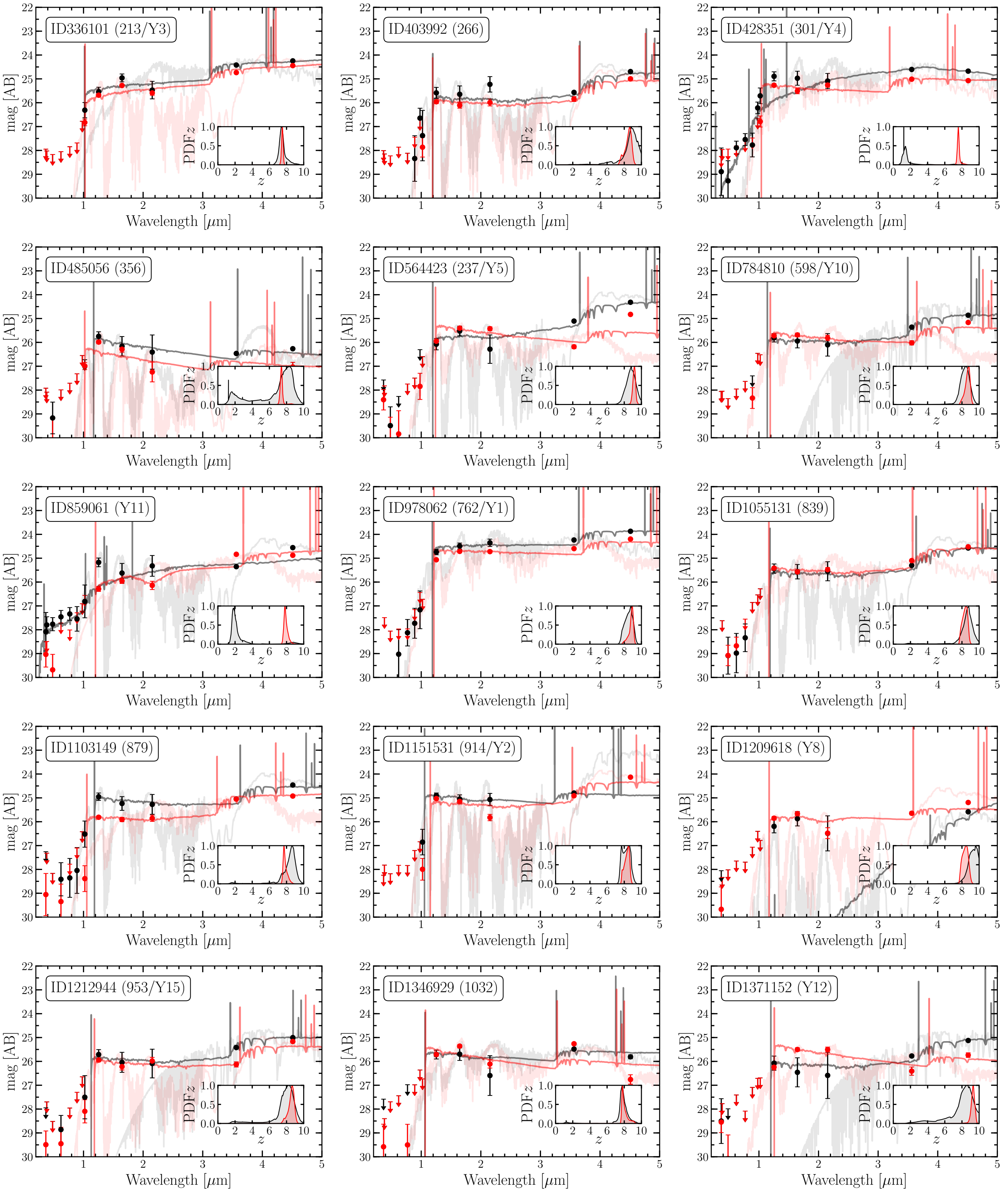

Fig. B.6: Same as Fig. B.1 for the $z \geq 7.5$ candidates from S19 and B20 which are recovered in the COSMOS2020 catalogue. The identifiers from those papers are indicated in parentheses.

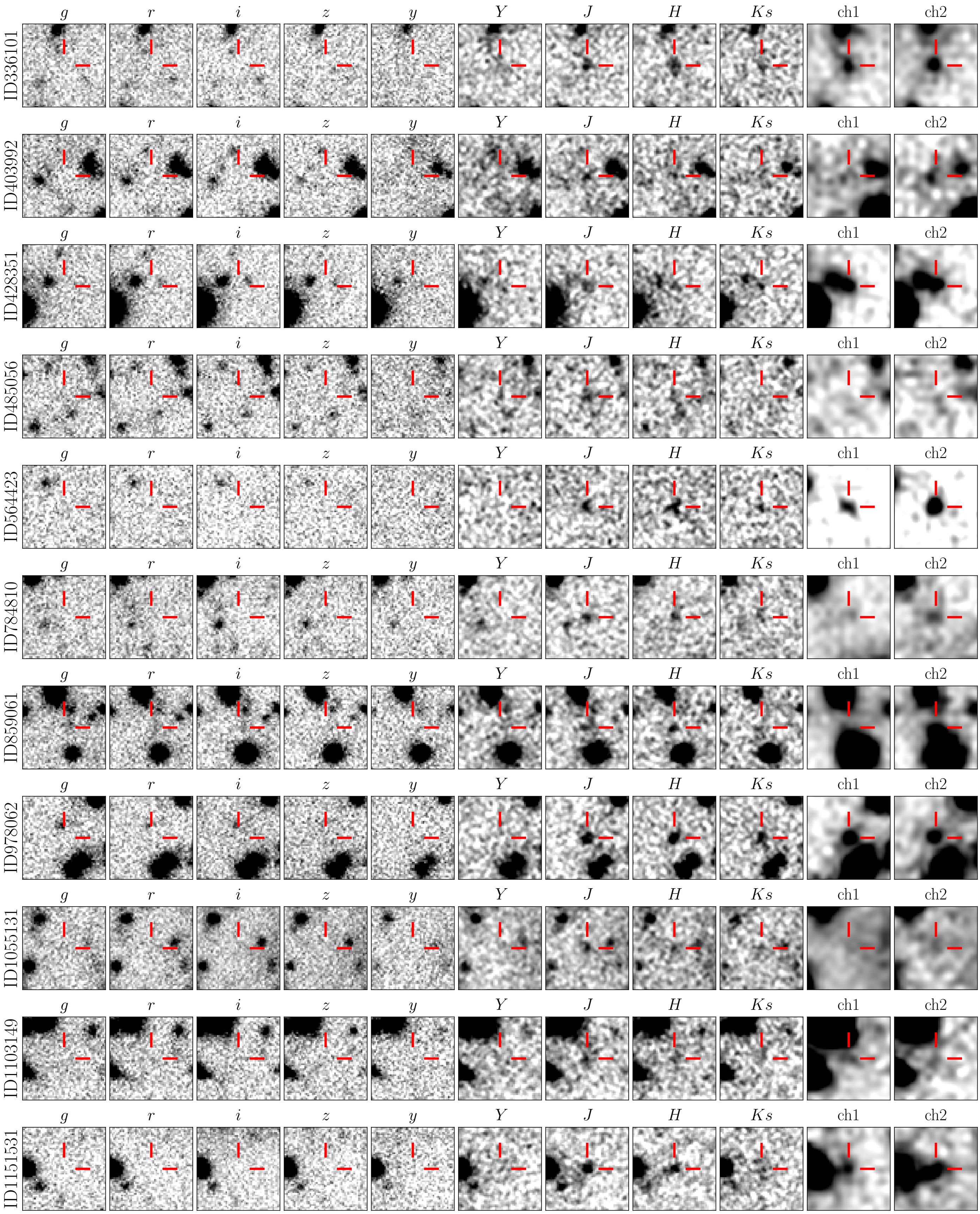

Fig. B.7: Same as Fig. B.2 for the $z \geq 7.5$ candidates from S19 and B20 which are recovered in the COSMOS2020 catalogue.