UNIVERSITY OF COPENHAGEN

Estimating photometric redshifts and inferring quantities of dust in supernovae with machine learning

Dissertation submitted for the degree of
PHILOSOPHIÆ DOCTOR

Zoe Ansari

Supervisor
Assoc. Prof. Christa Gall
Co-supervisor
Dr. Adriano Agnello

Assessment Committee
Prof. Viviana Acquaviva
Dr. Mikako Matsuura
Assoc. Prof. Steen Hansen

Submitted on 20.02.2022


NIENSIS SIG
ATÆ 1537

PUBLICATIONS

This thesis contains the following publications:

- Mixture models for photometric redshifts
  Zoe Ansari, Adriano Agnello & Christa Gall
  Published on *Astronomy & Astrophyics*; AA 650, A90 (2021)

- Inferring properties of dust in supernovae with neural networks
  Zoe Ansari, Christa Gall, Roger Wesson & Oswin Krause
  Under review at *Astronomy & Astrophyics* ; submitted on 10/01/2022

ACKNOWLEDGEMENTS

I would like to acknowledge my supervisor, Christa Gall, for her massive patience, excellent guidance, and strong belief in me throughout my Ph.D. Her effort in pushing me to the very edge of my comfort zone helped me develop into a version of myself beyond my initial expectations. In the same manner, I am thankful to Adriano Agnello, for being helpful and supportive throughout my Ph.D. I am grateful for their heartwarming encouragements and sharp comments during my Ph.D, which felt like crawling through an uncomfortable dark tunnel in the pursuit of the unknown with only a torch in hand. They inspired me through this journey, which helped me not let the dirt on my shoulders keep me from reaching the finishing line and eventually come out of the tunnel with trained muscles. Moreover, I am thankful to my collaborators, Oswin Krause, Roger Wesson, Doogesh Kodi Ramanah, Jes Frellsen, and Kristoffer Stensbo-Smidt for many hours of discussions. I am thankful to Jens Hjorth for funding my research via Villum fonden. I am grateful for having the opportunity to be part of the DARK as a Ph.D. fellow. I am thankful to all my office mates, in particular our coordinators, Michelle and Evi, who made the every day physical or virtual experience at DARK pleasurable. I would like to thank all the lovely people in the YSE group, from whom I learned a lot about observational techniques. I would like to thank my fellow students, Sofie and Jo, and in particular Cecilie for helping me to form the Abstrakt of this thesis. I am thankful to Marianne Vestergaard for her constructive and helpful feedbacks, and in general, for being a supportive Ph.D. coordinator. I am also thankful to my family for sending me all the best wishes along with lots of digital media every now and then. In particular, I am thankful to my newly born niece for her carefree smiles on the majority of the recently sent media that boosted my energy level whenever was needed. I would also like to thank my friends outside of academia who sent me to guilt-trips whenever I canceled on them due to my workload, which made me be sure that I am on the right path as the effort I put into my research was beyond ordinary. I am thankful to those at KU, NBI, DARK, DAWN, DTU-compute, ITU, UCL, UCSC, STScI, KIPAC, ESO and several other institutions and organisations that, one way

or another, shaped my conscious and/or subconscious to continue my research wholeheartedly. I am also grateful for the smell of Denmark, from which I got motivated every day to get out of my apartment and keep the wheels of my research going. Except for the lockdown periods and when I got covid-19 and had to self-quarantine. I am thankful for many audio-books about mental health, in particular Matt Haig's books that helped me to not only tolerate but also grow through rough moments during my Ph.D. I am also thankful for the 'high impact Ph.D. meme' page that gave me smiles and sometimes loud and intense laughs during my "DARK" Ph.D. moments. I would like to thank Steen Hansen for being an inspiring director at DARK and for reviewing my thesis as the internal opponent. I extend my gratitude towards Mikako Matsura and Viviane Acquaviva, who accepted to be the external opponents and put hours into reviewing my thesis. Last but not least, I acknowledge the ones from under-represented groups whose empowering resistance against the ”norms” somewhere along the way made a path to an equal world for everyone. Your avant-garde action is the reason that I got the opportunity to pursue my passion. Therefore, I dedicate my thesis to you.

ABSTRACT

Extracting precise information from large observational data is playing a vital role in modern astronomy. In particular, machine learning as a promising approach for the efficiency in gaining precise estimations on large astronomical data has become a growing field. In this thesis I have applied different machine learning algorithms to estimate redshift and quantities of dust from large photometric data sets of different astrophysical objects. Hereby I also investigated, if such machine learning methods provide improved estimates of the quantities and properties, compared to other, commonly used methods. First, I estimate photometric redshifts by applying a method that contains two probabilistic algorithms on a data set from the Sloan Digital Sky Survey cross matched with the Wide-field Infrared Survey Explorer. I employ an infinite Gaussian mixture model (IGMM) to classify the observed objects in the data set as three different classes (i.e. stars, galaxies and quasars). Then, I estimate probability density functions (PDFs) of photometric redshifts for the objects in the data set, by feeding the outcome of the probablistic classification into a mixture density network (MDN). Secondly, to infer quantities of dust in supernovae, I use a simulated data set of supernova spectral energy distributions calculated by an advanced fully three-dimensional radiative transfer code, `MOCASSIN`, to design three scenarios for estimating quantities and properties of dust in and around supernovae with synthetic magnitude adjusted to James Webb Space Telescope (JWST) telescope bandpass filters. I developed a neural network to estimate the dust properties along with the corresponding predicted uncertainties that reflect the sufficiency of the given photometric information for representing the dust quantities. By exploring a rich parameters space that is covered by the simulated sample, this trained algorithm became a promising approach to estimate dust properties with JWST photometric observations. Moreover, by implementing a feature selection framework on the neural network predictions I have found a minimum set of JWST filters for inferring

dust in supernovae that could benefit future observational strategies. Moreover, as an ongoing project for improving estimating photo-zs, I apply two different methods as alternatives to two different parts of the photo-z estimator in the first project. First, instead of classifying objects in the feature space, I classify the objects in a latent space by implementing a variational auto-encoder instead of an IGMM. Secondly, for reflecting the measurement errors in the estimated PDF photo-zs, I employ an importance sampling method instead of including the measurement errors as additional input features to the MDN, which may lead to more precise photo-z estimations.

ABSTRAKT

At udvinde præcis information fra store observationelle datasæt spiller en vital rolle i moderne astronomi. Specielt Machine Learning vinder frem som en lovende tilgang til at få mere præcise estimater på kortere tid end ved brug af andre metoder. I denne afhandling, som indrammer to forskellige astrofysiske udfordringer, har jeg undersøgt muligheden for at anvende Machine Learning algoritmer til at estimere egenskaber og størrelser relateret til astrofysiske objekter via fotometriske observationer og simuleringer. Først til at estimere fotometriske rødforskydninger, anvender jeg en metode, der indeholder to probabilistiske algoritmer, på et datasæt fra Sloan Digital Sky Survey sammenkoblet med Wide-field Infrared Survey Explorer. Først anvender jeg en Infinite Gaussian Mixture Model (IGMM) til at inddele de observerede objekter i datasættet i tre forskellige klasser (stjerner, galakser og kvasarer). Dernæst estimerer jeg probability density functions (PDFs) for de fotometriske rødforskydninger af objekterne i datasættet ved at fodre resultaterne fra klassifikationen til et Mixture Density Network (MDN). For det andet, for at udlede mængder af støv i supernovaer, bruger jeg et simuleret datasæt af supernova spektrale energifordelinger, genereret af en avanceret fuldt tredimensionel Radiative Transfer kode, `MOCASSIN`, til at fabrikere tre scenarier til at estimere størrelser og egenskaber relateret til støv i og omkring supernovaer med syntetisk størrelse justeret til James Webb Space Telescope (JWST) filtre. Jeg har udviklet et neutralt netværk til at estimere støvets egenskaber samt de tilsvarende forventede usikkerheder, der reflekterer tilstrækkeligheden af den givne fotometriske information til at repræsentere støvets egenskaber. Ved at udforske et rigt parameterrum, der er dækket af det simulerede datasæt, blev denne trænede algoritme en lovende tilgang til at estimere egenskaber ved støv med JWST fotometriske observationer. Derudover, har jeg ved at implementere et framework til at udvælge kendetegn ved det neurale netværks forudsigelser fundet et mindste antal JWST filtre nødvendigt for at kunne udlede egenskaber ved støv i supernovaer som

kunne gavne fremtidige observationsstrategier. Endvidere, som et igangværende projekt til at forbedre fotometrisk rødforskydning, anvender jeg to forskellige metoder som alternativer til to forskellige dele af beregneren af fotometrisk rødforskydning i det første projekt. Først klassificerer jeg objekterne i et latent rum, i stedet for at klassificere objekter i komponentrummet, ved at implementere en variabel auto-encoder i stedet for en IGMM. Dernæst for at inkludere måleusikkerhederne i de estimerede fotometriske rødforskydninger fra sandsynlighedsdensitetsfunktionen anvendte jeg en Importance Sampling metode i stedet for at inkludere måleusikkerhederne som øvrig input til MDN dat kan leiden tot nauwkeuriger photo-z schattingen.

Contents

1 Introduction

> *"Your reward will be the widening of the horizon as you climb. And if you achieve that reward you will ask no other."*
>
> – Cecilia Payne-Gaposchkin[a]

---

[a] Who proposed in her 1925 doctoral thesis that stars were composed primarily of hydrogen and helium, which was initially rejected due to its contradiction to the scientific wisdom of the time.

1.1 Preface

WE can study many complex physical phenomena by observing outer space that holds a vast portion of the universe. Because the universe is so big, and because the speed of light is finite, we can look back in time by observing more distant objects. Therefore, astrophysical phenomena take our attention when it comes to expanding the edge of our knowledge in physics.

However, in the scale of the age of the universe, we had only a blink of time to develop, study and understand this complex system. Therefore, even with the most advanced instruments, we have made, deriving the fundamental astrophysical parameters from observations and cracking the code of the universe is still challenging. Understanding the laws of physics behind the universe with observations is still a challenge for several reasons: i) to study the entire sky on a large scale, tracking each astronomical object with the most precise observational recordings is costly and practically not feasible. ii) due to uncontrollable shortcomings such as the instruments' limits and weather conditions, observations might contain systematic errors. iii) to extract information from even the most accurate and complete observations, a clear understanding of the physics behind the observed phenomena is required.

As a remedy to the first issue, astronomers employ statistical methods to study a large amount of objects by sampling. These methods help us extract information from the entire sky efficiently. For the second issue, astronomers use probabilistic methods that reflect measurement uncertainties in the extracted information from observations. The reflection of measurement uncertainties is commonly shown by error bars on estimations. For the third issue, astronomers develop various astrophysical models with different hypotheses. The models with the best fits to the observations are considered as the possible description of reality.

Observing the universe is done by measuring the electromagnetic waves from objects that reach us. The measuring system is outlined in Section 1.2, while Section 1.3.1 describes our current understanding of the universe in large scale. Then Section 1.3 explains two current astrophysical and cosmological challenges regarding extracting information from observations. Section 1.4 describes machine learning methods and the usage of them for resolving the observational challenges. In Section 1.5, I briefly introduce the following chapters of the thesis in which I applied machine learning algorithms to tackle the challenges of extracting precise information from observations.

1.2 Observations

IN order to study astronomical objects or phenomena, we observe their radiations by ground-based and space telescopes. The observations are recorded in two main ways: spectroscopy and photometry. Spectroscopic measurements are recorded by a spectrograph that works similarly to a prism and splits an incoming beam into several components: a set of fluxes at different wavelengths, known as spectrum. The spectrum of an observed object contains a detailed fingerprint that helps to identify some properties of the object. For instance, the emission lines at specific wavelengths represent presence of corresponding chemical elements in the observed object. On the other hand, photometric measurement is a measure of the observed flux in a given wavelength interval defined by a bandpass filter (hereafter filter). Figure 1.1 shows a comparison between a set of filters and an observed spectrum by representing the spectrum of the star Vega and the wavelength intervals and transmission curves for u, g, r, i, z filters from the Sloan Digital Sky Survey (SDSS) as a ground-based telescope and for $W1, W2, W3, W4$ filters from

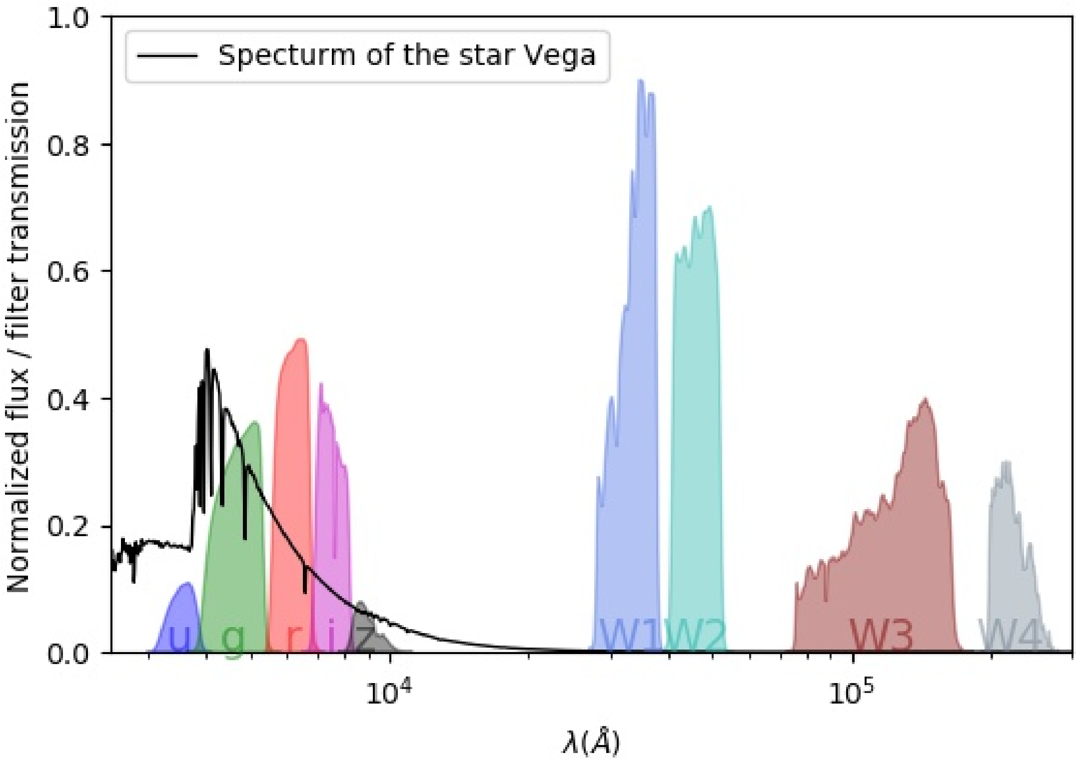

Figure 1.1: The spectrum of the star Vega (black line) together with SDSS and WISE filters. The transmission curves of SDSS u, g, r, i, z filters and WISE $W1, W2, W3, W4$ filters are shown in different colors. The figure is generated by `astroML` package with transmission curves fetched from Spanish virtual observatory.

Wide-field Infrared Survey Explorer telescope (WISE) as a space telescope. Photometric measurements from these filters have been used in Chapter 2.

In photometry, we measure the flux by determining the power collected within a defined aperture by a filter regarding to its transmission curve. An aperture is usually a circular shape that depending on the target and the background flux can be used in different sizes. Usually, the measured intensities by the filters are expressed in magnitude, a logarithmic unit that increases with decreasing intensities. The apparent magnitude of an observed object is defined as:

$$m_\lambda = -2.5 \log \left(\frac{F_\lambda}{F_{\lambda,0}} \right) \tag{1.1}$$

where F_λ and $F_{\lambda,0}$ are the observed flux and the reference flux at λ respectively. Astronomers created several magnitude systems to measure how relatively the brightness of sources is compared to different references. For instance, the Vega magnitude system uses star Vega as the reference;

$$m_{\text{Vega}} = 2.5\log\left(\frac{F}{F_{\text{Vega}}}\right) \tag{1.2}$$

where F is the observed flux of an object, and F_{Vega} is the observed star Vega flux. Therefore, by definition, the star Vega has approximately zero Vega magnitude at any wavelength. The AB magnitude system uses 3631 Jy[1] as the reference point for all the wavelengths;

$$m_{\text{AB}} = -2.5\log\left(\frac{F_\nu[Jy]}{3631}\right) = -2.5\log(F_\nu[Jy]) - 48.6 \tag{1.3}$$

The main focus of this thesis is using photometric observations (i.e. magnitudes) to efficiently estimate astrophysical properties by designing and employing machine learning algorithms, in order to tackle two current cosmological and astrophysical challenges.

1.3 Two current challenges in cosmology and astrophysics

1.3.1 Expansion of the universe

TODAY, the majority of scientists agree that our universe began from the Big Bang about 13.8 Gyr ago. However, up to 10^{-43} seconds after the Big Bang is a subject of speculation. It is widely accepted that the universe then started to expand exponentially and consequently its scale increased by many order of magnitudes. This epoch is known as inflation (Baumann, 2009), in which the universe was opaque since it was filled with an extremely hot and dense plasma. The universe was still expanding to the point that it started to cool down, which led to the formation of particles (Ryden, 2003; Baumann, 2009). However, it took about 380 000 years after the big bang for the universe to become cool enough (i.e. $T<3000$ K) to be able to form atoms of hydrogen and helium by combing ions and electrons in the plasma. This epoch is known as the epoch of

---

[1] 1 Jy = $10^{-23} erg s^{-1} cm^{-2} Hz^{-1}$= $10^{-26} W m^{-2} Hz^{-1}$

recombination. At this epoch the universe became visible, as the photons decoupled from matter and could travel through the universe. This era has been observed as the cosmic microwave background radiation at z=1100, by COsmic microwave Background Explorer (COBE; Smoot *et al.*, 1992), the Wilkinson Microwave Anisotropy Probe (WMAP; e.g. Dunkley *et al.*, 2009), and Planck satellites (Planck Collaboration; e.g. Planck Collaboration *et al.*, 2016) with different levels of precision.

The cosmological principle states the universe is homogeneous and isotropic on large scales (Cervantes-Cota and Smoot, 2011). The mathematical model that is used to describe how the homogeneous and isotropic universe expands is derived by four different scientists independently; Friedmann (Friedmann, 1922), Lemaître (Lemaıtre, 1931), Robertson (Robertson, 1935) and Walker (Walker, 1937). This metric is known as FLRW and expressed as follows:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + sin^2\theta d\phi^2)\right] \tag{1.4}$$

where r, ϕ and θ are spatial coordinates, t is the cosmic time and k represents the curvature of the universe; k equals to zero, $+1$ or -1 corresponding to the flat, open and closed universe. $a(t)$ is the scale factor that contains the temporal evolution of the universe. The scale factor $a(t)$, can be determined by general relativity proposed by Einstein in 1915 that states the curvature of space-time is described by the Einstein's field equations as follows:

$$\tag{1.5}$$

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{1.6}$$

where G and R are the Newton constant and Ricci scalar respectively. The $T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\partial L_{\mathrm{m}}\sqrt{-g}}{\partial g^{\mu\nu}}$ is the energy-momentum tensor where $g = |g_{\mu\nu}|$ is the determinant of the metric tensor. Friedmann then reduced $T_{\mu\nu}$ to $\rho + 3P$ by assuming that at large scales the universe can be approximated as a perfect isotropic and homogeneous fluid. The ρ and P represent the energy-density tensor and pressure of the fluid. Therefore, Friedmann introduced the following equations by approximating the Einstein's field equations:

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{1.7}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \tag{1.8}$$

where H is the Hubble parameter and the dots represent the time-derivatives. By assuming conservation of the energy-momentum tensor $\nabla^{\mu} T_{\mu\nu} = 0$, and describing most of the fluids in the universe by a simple equation of state $P = w\rho$ the cosmological evolution of the fluid can be determined as follows:

$$\rho \propto a^{-3(1+w)} \tag{1.9}$$

where w is the equation of state parameter and can be equal to zero, $1/3$, -1 and $w < -1/3$ corresponding to different substances of the universe, matter, radiation, dark energy if it corresponds to a cosmological constant, and dark energy. Each of the substances then can be expressed in a dimensionless density by dividing their density to a corresponding critical density that is the average density that a flat universe should have. Then the Friedmann equations can be written in terms of the dimensionless densities as follows:

$$\Omega = \Omega_{\mathrm{R}} + \Omega_{\mathrm{M}} + \Omega_{\Lambda} = 1 + \frac{k}{a^2 H^2} \tag{1.10}$$

where Ω_{R} represents the radiation component dominated by the CMB ($\approx 5 \times 10^5$), $\Omega_{\mathrm{M}} \equiv 8\pi G\rho_{\mathrm{m}}/3H^2$ and the cosmological constant $\Omega_{\Lambda} \equiv \Lambda/3H^2$ are characterising the matter content and the dark energy content of the universe. The cosmological constant represents the accelerated expansion of the universe.

To investigate the expansion of the universe, and the existence of dark energy that takes $\sim 70\%$ of the matter-radiation content of the universe, Dark energy survey mission[2] (DES; Abbott *et al.*, 2018) probes four different methods with Type Ia supernovae, Baryon acoustic oscillations, galaxy clusters and weak lensing. All these studies are highly dependent on the quality of the redshifts of the observed objects, as redshift can be seen as a measure of the expansion of the universe:

$$1 + z = \frac{a_0}{a(t)} \tag{1.11}$$

Therefore, estimating precise redshifts is playing a vital role for the success of the investigations on the expansion of the universe and dark energy.

---

[2] http://www.darkenergysurvey.org/science/

Photometric redshift

The most accurate way of measuring the redshift of astronomical objects is from the emission lines in their observed spectrum. The difference between the wavelength that a photon emit at ($\lambda_{\text{emit.}}$) and that of the photon observed at ($\lambda_{\text{obs.}}$) tells us how much the radiation from the object is stretched, which is the redshift and is calculated as follows:

$$z = \frac{\lambda_{\text{obs.}} - \lambda_{\text{emit.}}}{\lambda_{\text{emit.}}} \tag{1.12}$$

which shows how much the emission lines of elements in the object shifted compared to the emission lines from the same elements on the earth. Therefore, having an astronomical object's high-resolution spectrum lets us precisely measure its redshift. However, obtaining spectra of all observable objects is immensely time-consuming and it is more challenging to extract signal from spectroscopic observations compared to photometric ones. In particular, this is more challenging for distant galaxies and faint objects that require large integration time to record their spectra with an acceptable level of signal to noise ratio. Therefore, wide-field surveys such as DES, that aim to study the entire sky use photometric observations, which is less time-consuming and thus allows us to cover a wide region of the sky a lot faster. The measured redshift of an object with only photometric observations is called photometric redshift (hereafter photo-z).

Observing an object with photometric filters gives us only the general shape of the spectral energy distribution (SED) and does not contain all the details observed from the spectrum (e.g. emission and absorption lines). Therefore, another principle is used to estimate photo-zs; the Wien's displacement law which is derived from Planck's law, indicates the black-body radiation peaks at different wavelength proportional to the temperature of a black-body as follows:

$$\lambda_{\text{peak}} = T \times b \tag{1.13}$$

where T is the absolute temperate of the black-body in K and b is the Wien's displacement constant equals to 2.897 mm $\times K^{-1}$. Since different types of astronomical objects have different temperatures, we can estimate their photo-zs by measuring the shift in the λ_{peak} of their observed SEDs. However, the observed flux density from the objects are affected by both their nature and their distance from us. Thus, estimating one of them without knowing the other is challenging. A set of examples of objects is shown in Figure 1.2; in the top panel, spectra of a star, a galaxy, and a quasar contain detailed information that helps to identify their types. As shown in Figure 1.2, in this particular case, in the bottom panel, the shape of the photometric SED of a quasar is different from a star and a galaxy that helps to separate it. However, the shape of the SEDs observed from the galaxy and the star looks similar. Thus they are not easily distinguishable from one another by photometry.

The degeneracy between the type and redshift of objects is an example of degeneracy in the photometric parameter space that would be systematically suppressed by predicting point estimate photo-zs. These types of degeneracy, as well as the error measurements can be reflected in a probability density function (PDF) which provides complete information compared to point estimates. Moreover, PDF estimates provide a wide range of confidence on the reliability of estimated photo-zs, which is necessary to have, for example, in the cases where high precision is required. In particular, for deriving cosmological parameter measurements such as weak lensing studies, reliable photo-z PDF estimations are required to remove the calibration bias effect that is used to address systematic errors in observations. Thus, ensures a sufficient analysis of cosmological uncertainties (e.g. Viola *et al.*, 2015; Mandelbaum, 2018). Therefore, it is crucial to estimate PDF photo-zs that capture the complete photometric information for cosmological surveys such as DES, Kilo-Degree Survey Collabration (KiDs; de Jong *et al.*, 2013), and *Rubin*-Legacy Survey of Space and Time (LSST; Amiaux *et al.*, 2012).

Two main methods for estimating PDF photo-zs are template fitting methods (e.g. Arnouts *et al.*, 1999; Bolzonella *et al.*, 2000; Brammer *et al.*, 2010) and machine learning algorithms (e.g. Collister and Lahav, 2004; Vanzella *et al.*, 2004; Brescia *et al.*, 2013; Sadeh *et al.*, 2016; Almosallam *et al.*, 2016; Reis *et al.*, 2019; Wang *et al.*, 2008a; Han *et al.*, 2016). In template fitting methods, a set of referential templates are used from either direct observations or averaging the spectra of similar galaxies, or synthetic spectroscopy from physical modelings. The redshift of an observed object will be estimated from the similarity of its spectral energy distribution (SED) to the referential templates. These methods are computationally expensive due to a huge parameter space to search, and highly dependent on astrophysical assumptions behind the templates. In machine learning methods, an algorithm learns to map a photometric space onto the redshift using a large data set. This large data set should contain accurate spectroscopic redshifts as a

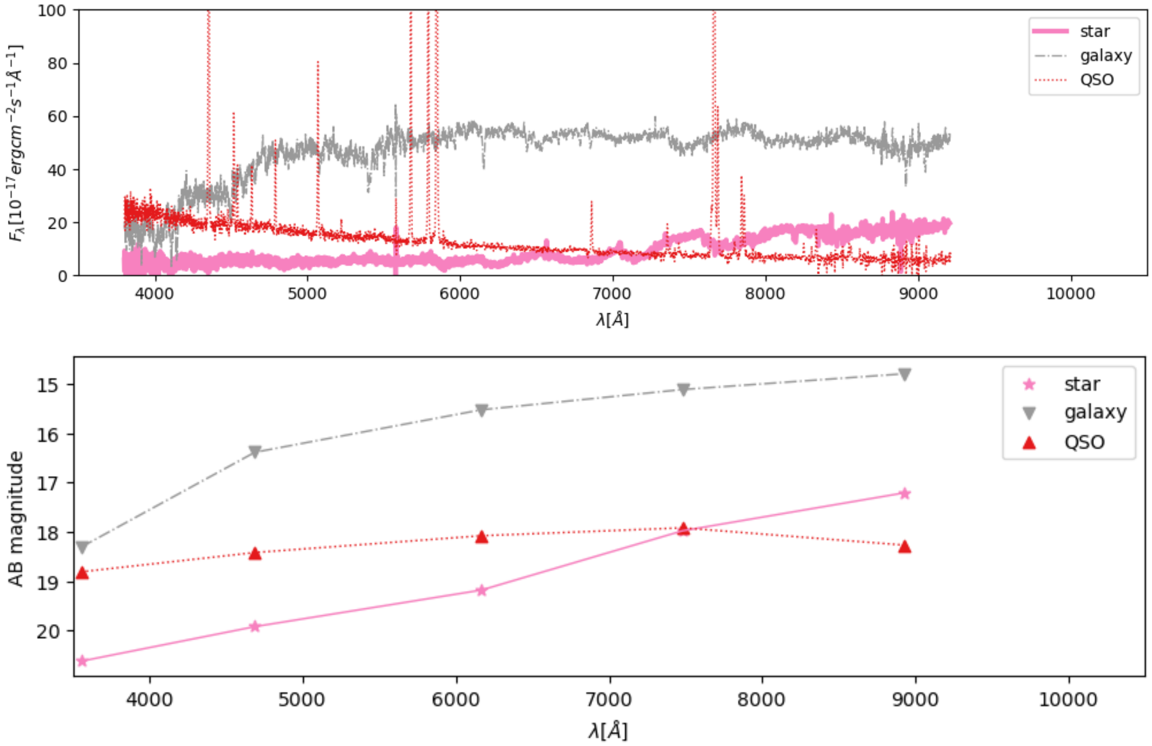

Figure 1.2: An example set of SED of star, galaxy and quasar. *Top panel:* The spectra of objects from SDSS database. The solid pink line represent a star at $z \approx 0.00002$. The dash-dotted gray line represent a galaxy at $z = 0.04407 \pm 0.00001$. The dotted red line represent a quasar at $z = 0.16742 \pm 0.00001$. *Bottom panel:* The corresponding magnitudes observed by u, g, r, i, z SDSS filters for the star ($*$ symbols), galaxy (downward triangles), and quasar (upward triangles).

prior knowledge so that the machine can learn the relationship between the photometric space and the redshift from it. Therefore, training machine learning algorithms as robust photo-z estimators requires a rich sample of observed SEDs containing all types of astronomical objects with accurate spectroscopic measurements of their redshifts. However, compared to template fitting methods, machine learning algorithms are fast to estimate the redshift of an observed object once they are well-trained.

In Chapter 2, I describe a method in which I employed two probabilistic machine learning algorithms to estimate PDF photo-zs, which is competitive with the state-of-the-art techniques. I trained the algorithm by a large sample of observed objects fetched from SDSS cross matched with WISE that made it a powerful tool to predict PDF photo-z for newly observed objects.

1.3.2 The origin of dust

Today we believe that our universe is made of visible (i.e. baryonic matter) and invisible matters (i.e. dark matter). The mass of the visible universe, which is in the interest of this work, contains 98% low-mass elements, hydrogen, and helium. It is believed that hydrogen and helium were produced in the hot, dense conditions of the birth of the universe itself. Heavier elements such as carbon, oxygen, and iron have since been continuously produced via fusion reactions in stars' birth, life, and death.

Cosmic dust (hereafter dust) refers to solid grains in space that form when atoms and molecules from the gas phase combine in a series of chemical reactions, and turn into clusters of increasing size. Dust in the size of a few molecules to a few microns have various chemical compositions including refractory compounds such as amorphous carbon, graphite, silicates (amorphous or crystalline), metal oxides (e.g. alumina), pure metals (e.g. iron and silicon) (Molster *et al.*, 2010)

The asymptotic giant branch (AGB) stars are often advocated to be the origin of cosmic dust (hereafter dust) in the local universe (e.g. Zhukovska and Henning, 2013), and provide the seeds for subsequent growth via accretion in the interstellar medium (ISM) (e.g. Draine, 2009). However, this scenario has been challenged with the recent observations of the Large Magellanic Cloud (LMC), the Small Magellanic Cloud (SMC) and the Milky Way (MW) in the local universe, and the observations of high redshift quasars in the early universe. Observations of the LMC, SMC and MW, identify a missing dust-mass problem that arises from the inability of existing AGBs to produce the dust mass observed in ISM (e.g. Matsuura *et al.*, 2009; Boyer *et al.*, 2012a; Matsuura *et al.*, 2013; Draine, 2009). Moreover, large amount of dust masses up to 10^8 $M_\odot$ discovered in high redshift galaxies (e.g. Carilli *et al.*, 2002; Omont *et al.*, 2003; Priddey *et al.*, 2003; Bertoldi *et al.*, 2003; Beelen *et al.*, 2006), can not have formed from the AGB stars since the evolutionary timescales of AGB stars are comparable to the age of the universe at $z < 4$.

It is known that core-collapse supernovae (hereafter CCSNe) are able to produce dust (e.g. Cernuschi *et al.*, 1967; Gehrz, 1989; Todini and Ferrara, 2001; Nozawa *et al.*, 2003; Barlow *et al.*, 2005). Supernovae are some the most energetic explosions in the universe that occur at the end of stars' lifetime. CCSNe are type of supernovae that are triggered

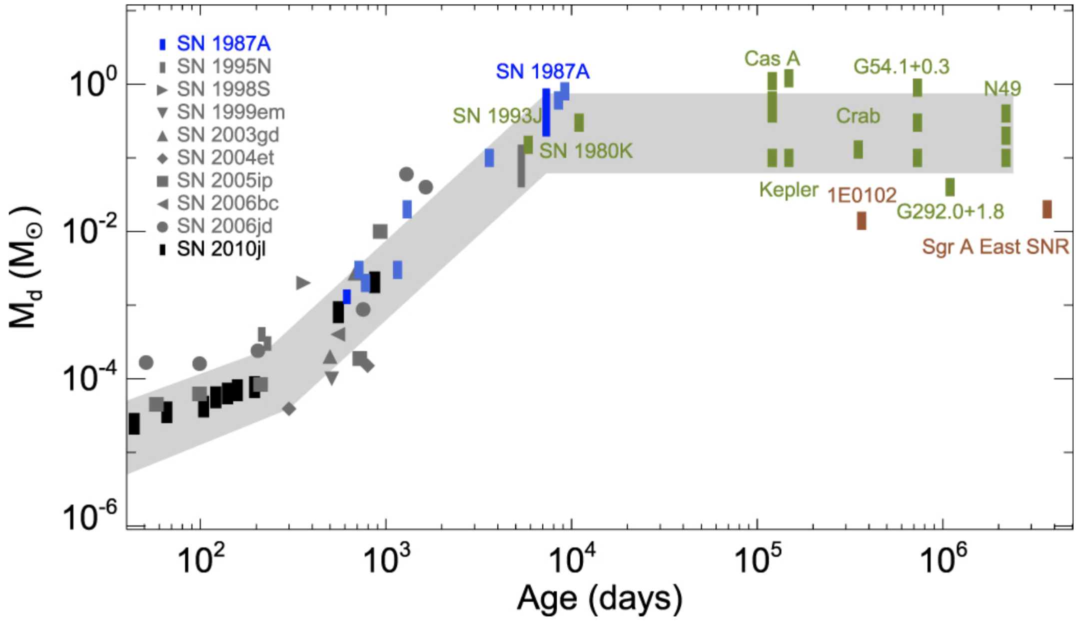

Figure 1.3: Evolution of dust mass in supernovae. The gray and black symbols are taken from Gall *et al.*, 2014. (reported measurements for SN 1987A are from Matsuura *et al.*, 2015; Lawrence, 2012; Indebetouw *et al.*, 2014a; Wesson *et al.*, 2014; Bevan and Barlow, 2015, for SN 1993J and SN 1980K are from Bevan *et al.*, 2016, for Cas A from Bevan *et al.*, 2016; Arendt *et al.*, 2014; De Looze *et al.*, 2016, for Kepler from Gomez *et al.*, 2009, for Crab from Gomez *et al.*, 2012b; Owen and Barlow, 2015; Temim and Dwek, 2013, for G54.1+0.3 from Temim *et al.*, 2017; Rho *et al.*, 2018, and for G292.0+1.8 from Ghavamian and Williams, 2016, for N49 Otsuka *et al.*, 2010, SNRs 1E0102 and Sgr A EAST SNR from Rho *et al.*, 2009b and Lau *et al.*, 2015). Detailing modeling suggests that dust mass either is formed instantly in optically tick clumps at early times (Dwek *et al.*, 2014), or increases over time (Gall *et al.*, 2014; Wesson *et al.*, 2015), as suggested here. The figure is taken from Gall and Hjorth, 2018.

by the gravitational collapse of the core. Supernovae type II (i.e. IIn, II-P, and II-L), Ib and Ic are known as CCSNe (Filippenko, 1997), and come from high-mass stars ($\sim$ 8-40 $M_\odot$). Therefore, it is argued that CCSNe due to the short life time of their progenitors are responsible for the presence of large amounts of dust in the early universe (e.g. Dwek and Scalo, 1980; Todini and Ferrara, 2001; Gall *et al.*, 2011b; Ferrara *et al.*, 2016).

There have been many attempts to model dust in CCSNe to investigate whether the CCSNe are the main contributors to the dust formation (e.g. Kozasa *et al.*, 1989; Todini and Ferrara, 2001; Nozawa *et al.*, 2003; Nozawa *et al.*, 2010; Bianchi and Schneider, 2007; Dwek and Cherchneff, 2011; Matsuura, 2017; Mauney and Lazzati, 2018; Sarangi *et al.*, 2018). The models state that in order to have CCSNe as the prime sources of dust formation in the early universe, a dust production of $0.1 - 1\mathrm{M}_{\odot}$ per supernova is required. In order to check the validity of the models, quantifying the dust in nearby CCSNe came to attention.

An excess thermal emission in the observed spectra is a sign of existence of the dust. The quantities of dust can be characterised by this thermal emission; from an observed flux density at frequency λ, by a method proposed by Hildebrand, 1983, the amount of dust can be determined as follows:

$$M_{\mathrm{d}} = \frac{F(\lambda) D_{\mathrm{L}}^{2}}{\kappa_{\mathrm{d}}(\lambda, a) B(\lambda, T_{\mathrm{d}})} \tag{1.14}$$

where $F(\lambda)$ is the total observed flux, D_{L} is the luminosity distance from an observer to the observed source, $B(\lambda, T_{\mathrm{d}})$ is the black-body function, T_{d} is the dust temperature of the source and $\kappa_{\mathrm{d}}(\lambda, a)$ is the dust absorption coefficient. For a spherical grain type the absorption coefficient is $(3/4)Q(\lambda, a)/(\rho a)$ where Q is the dust absorption efficiency, ρ is the density of the dust bulk, and a is the radius of dust particle.

Since the thermal emissions for hot, warm, and cold dust in supernovae occur at mid-IR to submillimetre wavelength ranges, telescopes that operate in these ranges such as the *Spitzer space telescope* (3–180 μ m), *Herschel space observatory* (55-672 μ m), WISE (3–25 μ m), and the Atacama Large Millimeter/submillimeter Array (ALMA) allowed us to investigate dust in the nearby supernovae. Figure 1.3 shows reported dust masses in SNe and SN remnants as a function of time after the explosion from different measurements via different observations. These measurements show dust production of ~ 0.1–$1\ \mathrm{M}_{\odot}$ per supernova that is consistent with most theoretical predictions (Nozawa *et al.*, 2003; Dwek and Cherchneff, 2011).

However, since supernovae are hostile environments, there is a concern that all the dust formed by supernovae might get destroyed by their reverse shocks. Some scenarios suggest that the reverse shocks may destroy dust, but might also influence either dust formation, re-formation or re-growth (e.g. Kochanek, 2011; Gall and Hjorth, 2018). The suggested mechanism for re-formation and re-growth in these scenarios is similar to the suggested mechanism of very early dust formation Type IIn SNe. As shown in Figure 1.4, early dust formation in Type IIn SNe are suggested to occur in a cold dense shell. Therefore, even in the case of having dust destruction in the reverse shock, the

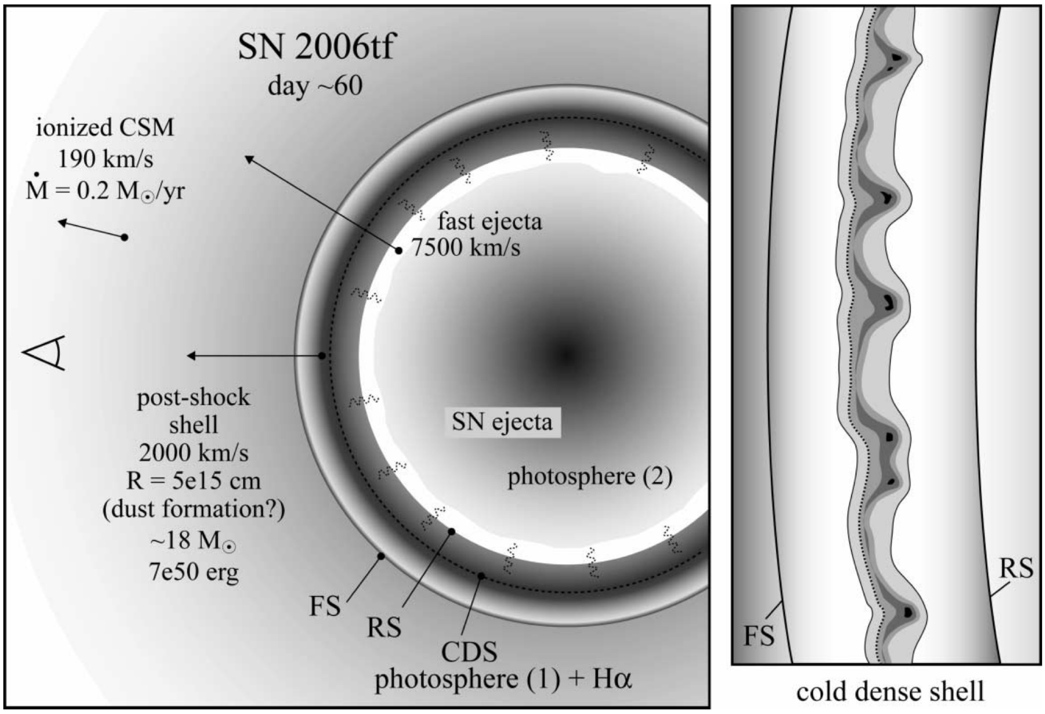

Figure 1.4: Schematic representation of the interaction of SN and circumstellar material (CSM). The cold dense shell (CDS) is divided from the shocked CSM and the shocked ejecta by the forward shock (FS) and the reveresed shock (RS) respectively. The figure is adapted from Smith *et al.*, 2008. Once the ejecta reaches the CSM around the supernova, as the consequence of their interaction, a reversed shock travels back into the expanding shell of the ejecta and a forward shock expands the CSM into ISM. Based on the Rayleeigh-Taylor instabilities, a cold dense shell between these two shocks will form that let the material in it cools down, mixes and form dust (Smith, 2016).

suggested mechanism of re-formation or re-growth may explain the dust budget produced by supernovae. Examples of these CCSNe are SN 1998S, 2005ip, 2006jd, 2006tf or 2010jl (e.g. Pozzo *et al.*, 2004; Stritzinger *et al.*, 2012; Smith *et al.*, 2008; Gall *et al.*, 2014). However, the theoretical models can not find evidence for dust re-formation in the post-reverse shock that may be due to assuming that the reverse shock destroyed all the dust in the first place. Although, Biscaro and Cherchneff, 2014 suggest that molecules re-form and whether or not dust gets destroyed may depend on the grain size distribution of the newly formed dust in supernova (e.g. Gall and Hjorth, 2018).

The methods that have been employed in these studies, either fitting a modified black body function or advanced physical modelings using Monte Carlo to find the best fit to the observed SEDs suffer from the influence of set of assumptions such as grain size distribution and dust composition. Therefore, the estimated quantities of dust from them usually contain some level of uncertainties. For example, as shown in Figure 1.5, fitting a modified black body function with carbon dust and silicate dust to SED of SN 2010jl that has been observed in a set of optical and near-IR filters are shown. Depending on the assumed composition, the estimated amount of dust formed by the supernova could vary up to one order of magnitude. As shown in Figure 1.5, sufficient observations in mid-IR wavelength range could allow to constrain the choice of the dust composition in the model which may help with inferring quantities of dust with more certainty.

Therefore, more observational data from supernovae are required in order to clear our understanding of dust formation and destruction mechanisms in supernovae. The James Webb Space Telescope (JWST) advanced instruments that will allow us to observe in 0.6–27 μm range with unprecedented resolution and sensitivity may let us obtain sufficient observations of supernovae and thus clarify the dust modelings in CCSNe.

In Chapter 3, I introduce a machine learning algorithm that I designed to infer properties of dust in supernovae with estimated uncertainties. I adjusted the algorithm for being used with JWST photometric observations. To train my algorithm, I used a simulated data calculated by a 3D radiative transfer code `MOCASSIN`. The data set covers the dust properties space properly and thus made the algorithm a powerful tool to estimate properties of dust in supernovae with.

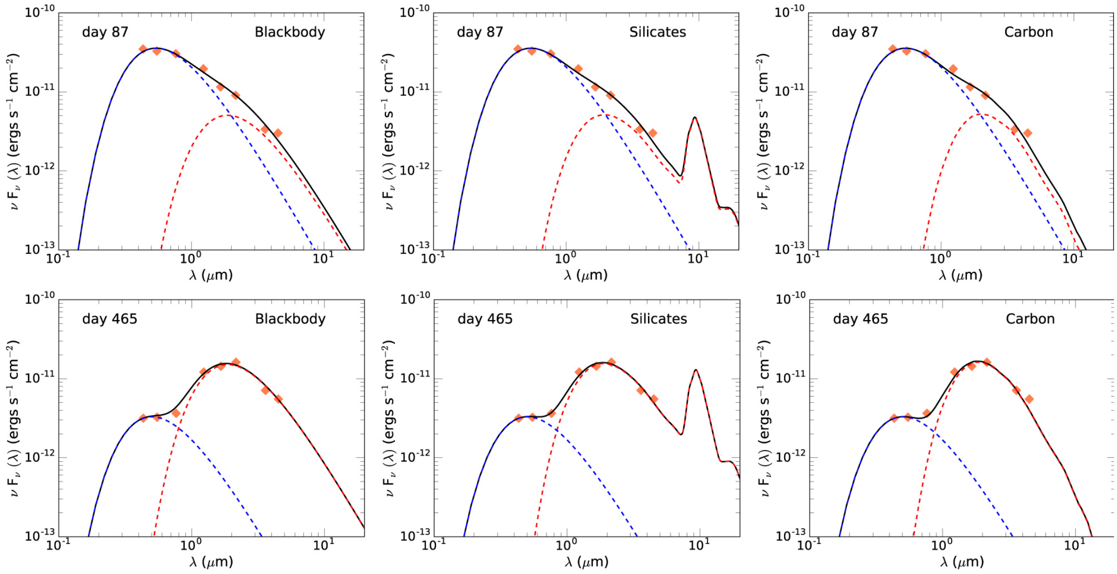

Figure 1.5: Fitted black bodies to SN 2010jl observations with optical and near-IR filters. The top and bottom panels represent the supernova at 87 and 456 days after the explosion, respectively. The left column illustrates a pure modified black body fitted to observations. The middle and right columns show a modified black body fitted to observations, with carbon dust and silicate dust, respectively. For a modified black body model with carbon dust at day 87 and 456, estimated dust masses 4.4×10^{-4} $M_\odot$ and $8.8\ \times10^{-4}$ $M_\odot$, and with silicate dust estimated amounts of dust are 1.8×10^{-3} $M_\odot$ and $4.7\ \times10^{-3}$ $M_\odot$. Due to lack of observations in mid-IR range, the choice of the dust composition can not be constrained and thus there is uncertainty in the estimated amount of dust. Figures are adopted from Sarangi *et al.*, 2018.

1.4 Machine learning

To perform computational tasks automatically, one can design and build a software program with a given set of instructions defined as an algorithm. When the tasks are governed by explicit rules, designing an algorithm for giving the instructions to a computer is explicit. Building a calculator is an example of this. However, when the rules that govern the tasks are difficult to specify clearly, we let the computer learn the rules using examples. For instance, reading handwritten digits is challenging for a computer since it follows vaguely specified rules. Although one can develop a set of rules such as the location of edges and curves corresponding to the different shapes of the digits, there is no guarantee that the program could read a new handwritten digit correctly. By using the MNIST[3] data set (Deng, 2012) that contains several thousands of examples of handwritten digits, one can design a program to learn to distinguish between different handwritten digits by discovering the rules automatically. The discovered rules by the program are known as learnable parameters. The learnable parameters are the parameters of a set of functions called a machine learning algorithm.

Although machine learning is an expanding field, as a rule of thumb, three main categories are proposed to define them; supervised learning, unsupervised learning, and reinforcement learning. In a nutshell, supervised learning is an algorithm that learns to process desired new information from past examples. Unsupervised learning looks at the structure of examples to identify the similar ones to each other in a feature space. In reinforcement learning, an algorithm learns by mistakes and develops its reactions based on cultivating rewards. The main focus of this thesis is on supervised learning.

1.4.1 Supervised learning

The aim of supervised learning is to train an algorithm $\mathscr{F}$, that can best approximates a function F, which maps a set of input features $X = (x_1, x_2, ..., x_i)$, onto a set of labels $Y(y_1, y_2, ..., y_j)$:

$$Y(y_1, y_2, ..., y_j) = F(X(x_1, x_2, ..., x_i)) \tag{1.15}$$

---

[3] Modified National Institute Standards and Technology

where i and j represent number of features and number of labels respectively. The most precise approximation of F, predicts Y most accurately for a set of X that has never been observed by it before. The mapping can be defined as:

$$Y = \mathscr{F}(X;\theta) \tag{1.16}$$

where θ is a set of parameters that the machine learns to be able to best approximate F, in a training process. The quality of approximation that is known as the performance of a machine learning algorithm is defined by a loss function $L(y_i, \hat{y}_i)$, where y_i is the ground truth set of labels and $\hat{y}_i$ is the predicted set of labels. The main objective of training an algorithm $\mathscr{F}$, is minimising the loss function, which will be equal to zero in the case of having perfect predictions. The choice of loss function depends on the problem at hand, the available features in training, and the desired label(s) to be predicted. Consequently, different forms of loss function are required for classification and regression problems.

As a branch of supervised learning, neural networks are powerful tools to approximate any arbitrary real valued continuous function f, with a well defined set of hyperparamters based on the universal approximation theorem. This is the reason why neural networks can generalise so well. More detailed descriptions on how neural networks work are provided in Section 3.4.1.

1.4.2 Application of machine learning in astrophysics and cosmology

Astronomical data sets are growing both in volume and complexity; upcoming surveys will provide imaging and spectroscopy of fainter sky than we presently observed, and thus increase the depth of observable distance. Consequently the amount of observational data will be increased by more than an order of magnitude. The massive production of data from wide-field surveys, along with an increase in computational powers and advancements in developing algorithms, makes a rich foundation for tackling astronomical challenges using machine learning.

In the domain of supervised learning, methods have been introduced and implemented for classifying galaxies (Connolly *et al.*, 1995), stellar spectra (Daniel *et al.*, 2011), transients (Bloom *et al.*, 2012, Wright *et al.*, 2015, Djorgovski *et al.*, 2012, Mahabal *et al.*, 2008), supernovae (Lochner *et al.*, 2016), and variable stars (Castro *et al.*, 2018, Naul *et al.*, 2018). Furthermore, there have been investigations on feature relevance in classifications (D'Isanto *et al.*, 2016), optimizing spectroscopic follow-up strategies (Ishida *et al.*, 2019), and attempts on estimating properties of astrophysical systems (e.g. Acquaviva, 2016; Acquaviva *et al.*, 2020). In the domain of unsupervised learning, variety of investigations have been done; anomaly detection or outlier detection in variable stars (Protopapas *et al.*, 2006), transients (Nun *et al.*, 2016), quasars (Agnello, 2017), dimensionality reduction (Vanderplas and Connolly, 2009), Automated Unsupervised Classification (Sánchez Almeida and Allende Prieto, 2013), and Knowledge Extraction from Complex Astronomical (D'Abrusco *et al.*, 2012).

However, when it comes to implementing machine learning in astrophysics and cosmology, a big concern is the lack of interpretability. On the contrary, the physical modeling commonly used in astrophysics can incorporate all the necessary astrophysical background knowledge. These models can be used for predictions by running Monte Carlo simulations. The physical models can explain the relationship between the observations and predictions, which is usually not provided by machine learning algorithms[4]. However, machine learning methods can be faster than physical modelling.

Moreover, the prediction quality of a physical model highly depends on the prior physical knowledge that may not be complete and suffer from simplifications. In contrast, machine learning approaches as data-driven modelings can outperform physical modelings regarding prediction accuracy. For instance, predicting photo-zs from machine learning approaches achieves better accuracy than template fitting methods (see Chapter 2).

In the case of quickly identifying observations worth a follow-up investigation based on a specific set of properties (i.e. a transient), machine learning approaches are indeed more practical. Furthermore, machine learning and physical modeling can be employed together to tackle astrophysical challenges. An example of such is training a machine learning algorithm with simulated data derived from a physical model that can predict a set of quantities of dust in and around supernovae (see Chapter 3).

While machine learning algorithms have their own problems, such as over-fitting and inability to extrapolate, they allow us to discover statistical relationships between quanti-

---

[4] However, recently, graph neural networks have been suggested as algorithms that can interpret the relationships between the observations and predictions (Cranmer *et al.*, 2019)

ties where the physical relationship is poorly understood either due to a set of chaotic and non-linear effects or too many interacting components involved in the system under the study.

1.5 This thesis in perspective

IN my thesis, I focus on applying machine learning algorithms to resolve two different astrophysical challenges motivated by the scientific goals explained in Section 1.3.

Both methods are designed for predicting astrophysical properties and quantities using photometric information. Therefore, both can be applied to wide-field surveys that provide their corresponding input features.

In Chapter 2, I present a probabilistic method to classify objects into stars, galaxies, and quasars and estimate PDF photo-zs. The method contains two machine learning algorithms. Both are fed with measurements uncertainties along with a set of observed SED with u, g, r, i, z filters from SDSS and $w1, w2$ from WISE. First, I review the importance of estimating PDF photo-z in Section 2.2. Then I describe the data in Section 2.3.1, and the Infinite Gaussian mixture model in Section 2.3.2 that I implemented for classifying the data into stars, galaxies, and stars. In Section 2.3.3, I describe the mixture density network that I fed with the outcome of classification to predict PDF photo-zs. I explain the results in Section 2.4, and conclude and discuss the outcome in Sections 2.5 and2.6 respectively. A review of state-of-the-art techniques can be found in Appendix 2.7.2. The corresponding paper is published in *Astronomy & Astrophyiscs*.

In Chapter 3, I present a novel machine learning approach to infer quantities and properties of dust in supernovae. I used a simulated set of SEDs achieved by a fully three-dimensional radiative transfer code `MOCASSIN` that I convolved with JWST filters to make a data set for training the designed algorithm. Section 3.2, reviews the literature on estimating dust formed by CCSNe as primary factories of dust. In Section 3.3.1, I describe how `MOCASSIN` provides a set of SEDs from simulations and what is the parameters' space that is covered by the data set. In Section 3.4.4, I explain how I deal with observational shortcomings. In Section 3.4.6, I describe a feature selection process that I employed to find the most suitable and the least set of JWST filters to predict the amount, temperature, and composition of dust in and around supernovae. In Sections 3.5 and 3.7, I explain how I interpret and evaluate the results on a simulated test set. Conclusion and the discussion of the work are summarised in Sections 3.9 and 3.8. The corresponding paper is submitted in *Astronomy & Astrophysics* and is under review.

In Chapter 4, I describe the works in progress for PDF photo-z estimation and future plans for improving usage of machine learning for inferring quantities of dust in supernovae.

2 Mixture models for photometric redshifts

"It is not only the question, but the way you try to solve it."
– Maryam Mirzakhani[a]

---

[a] The first woman to win the Fields Medal in mathematics

"Mixture models for photometric redshifts"

Published in *Astronomy & Astrophysics* (A&A) 650, A90 (2021)
Authors: Zoe Ansari, Adriano Agnello, Christa Gall

2.1 Abstract

Determining photometric redshifts (photo-zs) of extragalactic sources to a high accuracy is paramount to measure distances in wide-field cosmological experiments. With only photometric information at hand, photo-zs are prone to systematic uncertainties in the intervening extinction and the unknown underlying spectral-energy distribution of different astrophysical sources, leading to degeneracies in the modern machine learning algorithm that impacts the level of accuracy for photo-z estimates. Here, we aim to resolve these model degeneracies and obtain a clear separation between intrinsic physical properties of astrophysical sources and extrinsic systematics. Furthermore, we aim to have meaningful estimates of the full photo-z probability distribution, and their uncertainties. We performed a probabilistic photo-z determination using mixture density networks (MDN). The training data set is composed of optical ($griz$ photometric bands) point-spread-function and model magnitudes and extinction measurements from the SDSS-DR15 and *WISE* mid-infrared (3.4μm and 4.6μm) model magnitudes. We used infinite Gaussian mixture models to classify the objects in our data set as stars, galaxies, or quasars, and to determine the number of MDN components to achieve

optimal performance. The fraction of objects that are correctly split into the main classes of stars, galaxies, and quasars is 94%. Furthermore, our method improves the bias of photometric redshift estimation (i.e. the mean $\Delta z = (z_p - z_s)/(1 + z_s)$) by one order of magnitude compared to the SDSS photo-z, and it decreases the fraction of 3σ outliers (i.e. $3 \times rms(\Delta z) < \Delta z$) . The relative, root-mean-square systematic uncertainty in our resulting photo-zs is down to 1.7% for benchmark samples of low-redshift galaxies ($z_s < 0.5$). We have demonstrated the feasibility of machine-learning-based methods that produce full probability distributions for photo-z estimates with a performance that is competitive with state-of-the art techniques. Our method can be applied to wide-field surveys where extinction can vary significantly across the sky and with sparse spectroscopic calibration samples. The code is publicly available.

2.2 Introduction

The redshift of an astrophysical object is routinely determined from absorption or emission lines in its spectrum. In the absence of spectroscopic information, its photometric redshift (hereafter photo-z) can be estimated from the apparent luminosity measured in different photometric bands (see e.g. Salvato *et al.*, 2019, for a general review). Accurate photo-zs are needed by wide-field surveys that seek to probe cosmology through the spatial correlations of the matter density field, and are in fact a core limiting factor in the accuracy of these measurements (e.g. Knox *et al.*, 2006).

While large areas of the sky are covered by optical and near-IR imaging surveys, only a minority of objects have observed spectra, and hence secure redshifts from emission or absorption lines. The major problem is the rather narrow wavelength range covered by most photometric bands that introduces uncertainties and degeneracies in estimating the redshift. The Kilo-Degree Survey (KiDS; de Jong *et al.*, 2013) and Dark Energy Survey (DES; Abbott *et al.*, 2018) collaborations have used a combination of spectroscopic surveys to calibrate photo-zs (e.g. Hoyle *et al.*, 2018; Joudaki *et al.*, 2018), with the ultimate aim of measuring the matter content (Ω_m) and present-day root-mean-square (rms) matter density fluctuations (σ_8). The most used spectroscopic datasets are zCOSMOS (Lilly and Zcosmos Team, 2008), PRIMUS[1] (Coil *et al.*, 2011), VVDS (Garilli *et al.*,

---

[1] While it is not all present in the literature, PRIMUS has been used for instance by Hoyle *et al.* (2018), and Behroozi *et al.* (2019) suggest that its redshift uncertainties may be smaller than previously thought.

2008), DEEP2 (Newman *et al.*, 2013), VIPERS (Guzzo *et al.*, 2014), GAMA (Driver *et al.*, 2011), and SDSS (Ahn *et al.*, 2012). Albeit to a different extent, their completeness is generally affected by a non-trivial pre-selection in colour and morphology, and in some cases a limited footprint and depth. Hildebrandt *et al.* (2020) have identified the different calibrations of photo-zs, across DES and KiDS, to explain the difference in inferred cosmological parameters, claiming that the uncertainties in photo-zs are one outstanding challenge towards percent-level cosmology from weak lensing.

When only photometric information is available, a three-fold degeneracy between an object type, its redshift, and foreground extinction hinders the unambiguous determination of the redshift. Galametz *et al.* (2017) have quantified this effect explicitly in view of a possible synergy between the ESA-*Euclid* mission (Amiaux *et al.*, 2012) and *Rubin*-Legacy Survey of Space and Time (LSST; Amiaux *et al.*, 2012), which should cover more than half of the extragalactic sky to $\gtrsim 24$ mag depth in YJH-bands and $ugriz$-bands, respectively. Here, we explore a probabilistic approach to compute photo-zs that account for the existence of an indefinite number of astrophysical object types and their cross-contamination due to broad-band imaging information. Specifically, we trained a suite of mixture density networks (MDNs, Bishop, 1994) to predict the probability distribution of the photo-z of an object with measured magnitudes in multiple photometric bands as well as Galactic extinction. Following the standard nomenclature of machine-learning works, we alternatively refer to the photometric properties (magnitudes and extinction) as features in the rest of this paper. The MDN output is a sum of Gaussian functions in photo-z, whose parameters (i.e. the average, dispersion, amplitude) are non-linear combinations of the photometric inputs such as magnitude and extinction. Throughout the paper, we refer to these output Gaussians as 'branches'. In order to determine the number of branches that are needed to optimally parameterise the photo-z probabilities, we must determine the range of MDN branches that most accurately describe the data set. Hence, we explore infinite Gaussian mixture models (IGMM) on a photometric sample of which about 2% of the sources have spectroscopic redshifts (see sect. 2.3.1).

There are two main methods commonly used to estimate photometric redshifts: (I) template fitting and (II) machine learning algorithms. Template fitting methods specify the relation between synthetic magnitudes and redshift with a suite of spectral templates across a range of redshifts and object classes, through maximum likelihood (e.g. Fernández-Soto *et al.*, 1999) or Bayesian techniques (e.g. Benıtez, 2000; Brammer *et al.*, 2008; Ilbert *et al.*, 2006). Machine learning methods, using either images or a vector of magnitudes and colours, learn the relation between magnitude and redshift from a training data set of objects with known spectroscopic redshifts. In principle, template fitting techniques do not require a large sample of objects with spectroscopic redshifts for training, and can be applied to different surveys and redshift coverage. However, these methods are computationally intensive and require explicit assumptions on , for example, dust extinction, which can lead to a degeneracy in colour-redshift space. Moreover, template fitting techniques are only as predictive as the family of available templates. In the case of large samples of objects with spectroscopic redshifts, machine learning approaches such as artificial neural networks (ANNs; e.g. Amaro *et al.*, 2019; Shuntov *et al.*, 2020), k-nearest neighbours (kNN; e.g. Curran, 2020; Graham *et al.*, 2018; Nishizawa *et al.*, 2020), tree-based algorithms (e.g. Carrasco Kind and Brunner, 2013; Gerdes *et al.*, 2010) or Gaussian processes (e.g. Almosallam *et al.*, 2016) have shown similar or better performances than the template fitting methods. However, machine learning algorithms are only reliable in the range of input values of their training data set. Additionally, a lack of sufficient high redshift spectroscopic samples affects the performance of machine learning implementations on photo-z estimates. Another aspect is the production of photo-z probability distributions given the photometric measurements: While template-based methods can easily produce a probability distribution by combining likelihoods from different object templates, most of the machine-learning methods in the literature are only trained to produce point estimates, that is just one photo-z value for each object. For the sake of completeness, we summarise the state-of-the-art (and heterogeneous) efforts in the literature in Table 2.7 and their performance metrics evaluation in Table 2.8. We emphasise that most of the photo-z estimation methods above have been trained and tested purely on spectroscopic samples of different types of galaxies, often in a limited redshift range. Additionally, some of the spectroscopic galaxy samples were simulated entirely.

Here, we explore different kinds of mixture models to produce appropriate photo-z probability distributions that naturally account for the superposition of multiple, a priori unknown classes of astrophysical objects (e.g. stars, galaxies, quasars). There are multiple ways to describe a distribution of such objects in photometry space that consists of, for instance, magnitudes and extinction estimates (see Sect. 2.3.1) and that is also termed 'feature space' following the standard machine-learning terminology.

First, we use an IGMM (Teh, 2010) to separate the astrophysical objects in feature space. This approach allows the algorithm to cluster the objects based on all the available photometric information without forcing the algorithm to classify the objects in a predetermined way. Subsequently, the structure of the photometric (feature) space defines the number of Gaussian mixture components. Whenever a spectroscopic subsample of different types of astrophysical objects is available, IGMMs allow to separate this sample into classes, ideally representing each type of object. Secondly, we train MDNs to predict the photo-z probablity distributions of objects in our data set. To find the optimal results, we explore different MDN implementations, which all include the IGMM components and membership probabilities obtained in the first step next to the entire photomoetric (feature) space (Sect. 2.3.1).

In Section **??**, we describe our chosen training and test data sets as well as the IGMM and MDN implementations. The obtained accuracy of the classification along with the precision of the inferred photo-zs are provided in Section 2.4. In Section 2.6 we discuss our results, shortcomings and future improvements on our photo-z estimation alongside a comparison with other methods to estimate photo-zs from the literature.

2.3 Data and methods

To train our machine learning algorithms, we require a data set that contains: (I) morphological information from publicly available object catalogues (e.g. psf vs model magnitudes, or stellarity index), to aid the separation of stars from galaxies and quasars; (II) a wide footprint of the sky, to cover regions with sufficiently different extinction; (III) multiband photometry from optical to mid-IR wavelengths, possibly including u-band; and (IV) a spectroscopic sub-sample of different types of objects (here: stars, galaxies and quasars)

2.3.1 Data

Our photometric data set is composed of optical `PSF` and `model` $griz$-band magnitudes including $i-$band extinction measurements from the SDSS-DR15 (Aguado *et al.*, 2019). We combine these SDSS magnitudes with `w1mpro` and `w2mpro` magnitudes (hereafter $W1$, $W2$) from WISE (Wright *et al.*, 2010). We query the data in CasJobs[2] on the

---

[2] https://skyserver.sdss.org/casjobs/

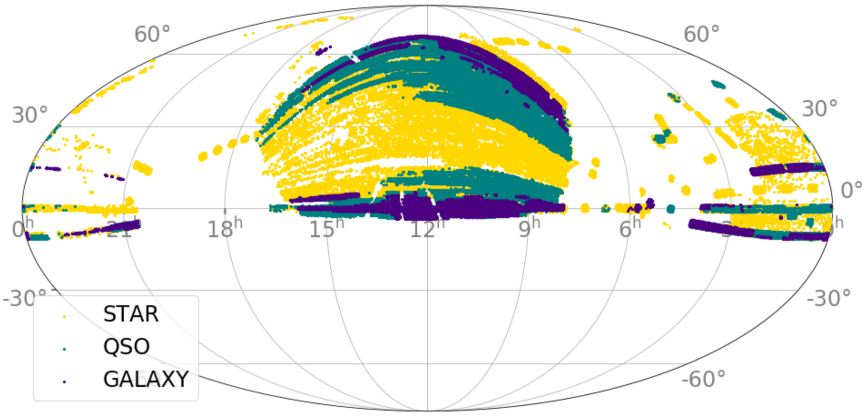

Figure 2.1: Spectroscopic data set in equatorial coordinates. Data are taken from SDSS-DR15 + WISE totalling about 245 000 objects of which there are 86412 stars (yellow), 83 119 galaxies (purple) and 75 955 quasars (green). The entire photometric data set is a sample of about 1 023 000 objects, of which 98% lack spectroscopic redshifts and classification.

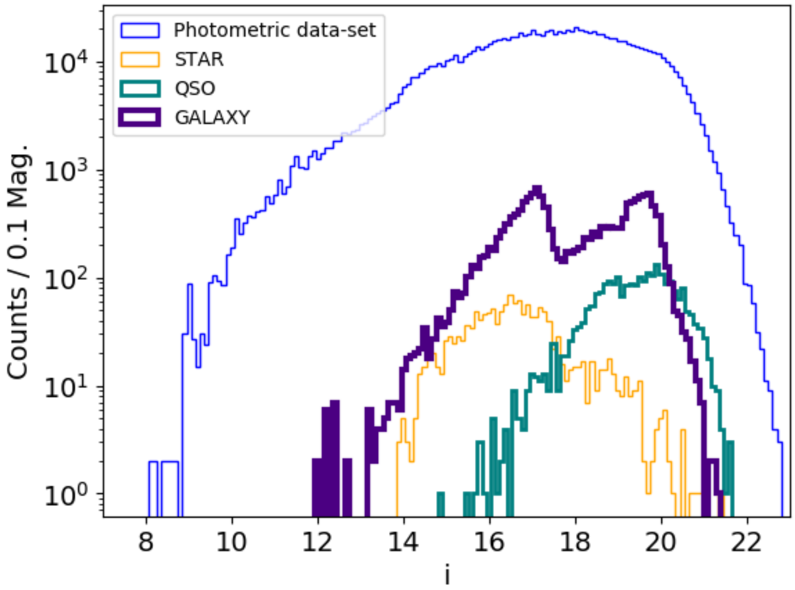

Figure 2.2: Histogram showing $i-$band magnitudes of the objects from the photometric (blue) and spectroscopic data sets for stars (yellow), galaxies (purple) and quasars (green), in 0.1 magnitude bins.

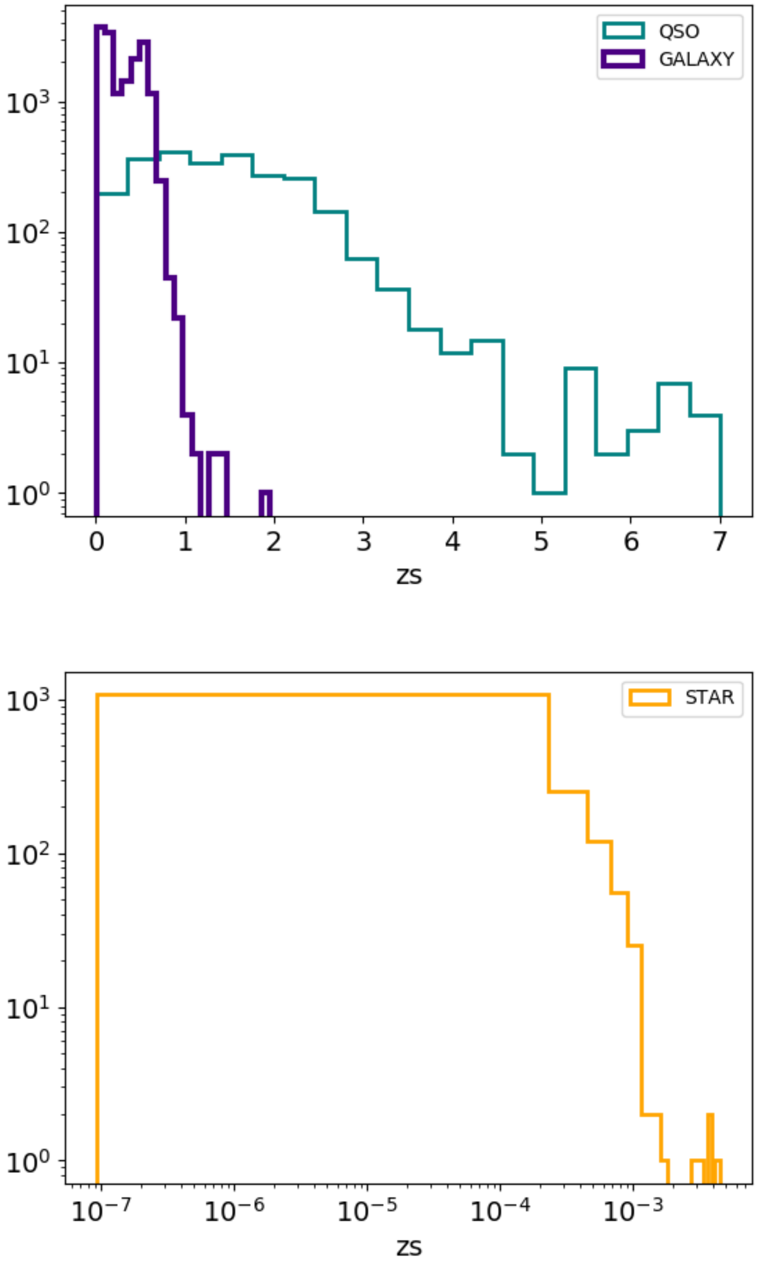

Figure 2.3: Redshift distribution of the spectroscopic data set. Top panel: galaxies (purple) and quasars (green) in 0.1 redshift bins. Bottom panel: stars (yellow) in 0.0001 redshift bins.

`PhotoObjAll` table with a SDSS-WISE cross-match, requiring magnitude errors lower than 0.3 mag and $i - W1 < 8$ mag. Adding $g - r$, $r - i$, $i - z$, $z - W1$ and $W1 - W2$ colours leaves us with 22 dimensions to be used by our MDNs. However, the colours are strictly speaking redundant as they are obtained from the same, individual photometric bands. While this introduces many null value Eigenvectors in the IGMM, additional combinations of measurements are enabled, which speeds up the MDN computations by detrending the magnitude-magnitude distribution. Our spectroscopic data set (from SDSS-DR15) includes only objects with uncertainties on their spectroscopic redshift (from the SDSS pipelines) smaller than 1%. For only one MDN training, we added $u-$band `PSF` as well as `model` magnitudes. Our complete data sets are composed of a photometric and a spectroscopic data set. For about 2% of the photometric data set we have spectroscopic information. This data set is called the 'spectroscopic data set'. For the IGMM, we have in total 1 022 731 unique sources from `PhotoObjAll` and WISE, with additional 11 358 unique galaxies from WiggleZ (Drinkwater *et al.*, 2010) cross-matched with `PhotoObjAll` and WISE. For the test samples, the spectroscopic data set contains 86 412 unique stars, 83 119 unique galaxies and 75 955 quasars from `SpecPhoto` and WISE according to the classification of their spectra by the SDSS pipelines (see Figure 2.1).

Figure 2.2 shows a general issue that is common to the literature on photo-zs, i.e. (by survey construction) the spectroscopic training sets do not reach the same depth as the photometric ones. This highlights the need for techniques that can extrapolate smoothly and with realistic uncertainties outside the ranges of a limited spectroscopic training set. Figure 2.3 shows the redshift distribution of different classes of objects from the spectroscopic data set. Evidently, galaxies are mainly placed at redshifts $\lesssim 1$, while quasars extend out to redshifts ~ 7.

2.3.2 Infinite Gaussian mixture models

In a Gaussian mixture model (GMM), the density distribution of objects in 'feature space' (equivalent to photometric space, see Sec 2.3.1) is described by a sum of Gaussian density components. The GMM is a probabilistic model which requires that a data set is drawn from a mixture of Gaussian density functions. Each Gaussian distribution is called a component. As the Gaussian distributions are defined in all the dimensions of the feature space, they are characterised by a mean vector and a covariance matrix.

The feature vector contains the photometric information of each astronomical source. To describe the GMM, whenever needed, we use the notation $\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$, where $k(\in \{1, ..., K\})$ is the component index, μ_k, Σ_k and π_k are the mean vector and the covariance matrix in feature space, and the weight of component k, respectively.

Since the GMM is a Bayesian method, it requires multiple sets of model parameters and hyperparameters. The model parameters (means, covariances) change across the Gaussian components, while the hyperparameters are common to all of the Gaussian components, because they describe the priors from which all Gaussian components are drawn. For the GMM, the number of Gaussian components is a fixed hyperparameter.

The IGMM is the GMM case with an undefined number of components, which is optimised by the model itself, depending on the photometric data set used. In particular, the IGMM describes a mixture of Gaussian distributions on the data population with an infinite (countable) number of components, using a Dirichlet process (Teh, 2010) to define a distribution on the component weights.

However, setting an initial number of Gaussian density components is required by the IGMM. Based on the weights that are given to each such component at the end of the model training, it is common practice to exclude the least weighted components and define the data population only by the highest-weighted components. To pursue a fully Bayesian approach, it is advisable to explore a set of model hyperparameters with different initial guesses for the number of components. Like its finite GMM counterpart, each realisation of IGMM estimates the membership probability of each data point to each component. Appendix 2.7.2 provides a summary of the IGMM formalism.

For this work, we used the built-in variational IGMM package from the `scikit-learn` library for our implementations. In practice, the variational optimiser uses a truncated distribution over component weights with a fixed maximum number of components, known as stick-breaking representation (Ferguson, 1973), with an expectation-maximisation algorithm (Dempster *et al.*, 1977). To optimise the model and find the best representation of the data set, we explored a range of parameters. We increased the maximum number of allowed Gaussian components from 10 to 100 in increments of 2 and we set the maximum number of iterations for expectation maximisation performance to 2 000. We used a Dirichlet concentration (γ) of each Gaussian component (k) on the weight distribution, of either 0.01, 0.05 or 0.0001 times the number of objects in the training

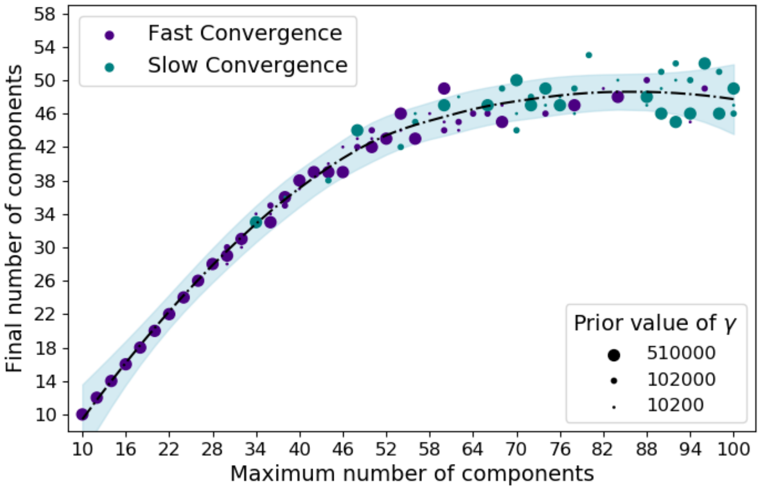

Figure 2.4: Maximum number of components vs. final number of components for different IGMM realisations, restricted to Gaussian components that contain at least 0.5% of the photometric data. Blue filled circles represent IGMM realisations that needed more than 2 000 iterations to converge, while purple filled circles mark IGMM realisations that needed less than 2 000 iterations. The size of the symbols scales with three different values of the prior of the Dirichlet concentration (γ). The light blue shaded region represents the confidence interval of 99% of regression estimation over the IGMM profiles by a multivariate smoothing procedure.

data set. The covariance matrix for each Gaussian component was defined as type 'full'; as per definition, each component has its own general covariance matrix. Furthermore, the prior on the mean distribution for each Gaussian component is defined as the median of the entries of the input vectors of the training data set (i.e. magnitudes, extinction).

Whenever needed, each object is assigned to the component to which its membership probability is maximal. In that case, we say that a component contains a data point.

The IGMM provides different possible representations of the same data set for each set of hyperparameters: here, we are interested in finding out the optimal number of components that can adequately describe the majority of the data. We then introduce a lower threshold on the number of sources that each component contains, and drop the components which contain less than the threshold. The threshold is defined by considering the size of the photometric sample and the highest value that we considered for the Dirichlet γ prior. The IGMM starts with components that contribute to 0.5% of the size of the photometric sample, since the highest γ value is 510 000 (see Appendix for further details), due to our chosen ranges of hyperparameters. Therefore, we use 0.5% of the size of the photometric data set as the threshold. Figure 2.4 shows that the final number of components converges to 48 ± 4. The convergence indicates that the models do not need more than 48 ± 4 components to describe the sample. Moreover, the initial 1:1 ramp-up in the figure shows that the final number of components is the same as the maximum tolerance, and so the model cannot adequately describe the data set; this trend breaks at about 44 components. To guide the eye, we determine a regression surface of all the IGMM profiles by a multivariate smoothing procedure[3]. In what follows, we choose 52 components.

The first IGMM implementation was fully unsupervised, as it was optimised to only describe the distribution of the objects in feature space. Subsequently, we trained different IGMMs considering additional spectroscopic information available for $\approx 2\%$ of the photometric sample. In particular, these partially supervised implementations are trained using the entire photometric feature space including either (I) spectroscopic classifications or (II) spectroscopic redshifts or (III) spectroscopic classifications and redshifts. Since the objects with additional spectroscopic information are a small part of the photometric training sample ($\approx 2\%$), the implementations ensure that the SDSS spectroscopic preselection does not bias the IGMM over the entire photometric sample. Finally, we calculate the membership probabilities to the 52 components for each object in the spectroscopic data set ($\approx 2.45 \times 10^5$ objects) from the optimised IGMM. This allows us to assign each object from the spectroscopic sample to one component. Thereafter, we label each of the IGMM components based on the percentage of spectroscopic classes that it contains.

Figure 2.5 shows the population of objects from the spectroscopic data set and their corresponding IGMM components in $g-r$ vs. $z-w1$ (upper panel) and $w2$ vs. $w1-w2$ (bottom panel) colour-colour and colour-magnitude diagrams. Each row from left to right shows the assigned components to stars, galaxies and quasars in the respective panels.

---

[3] https://has2k1.github.io/scikit-misc/loess.html

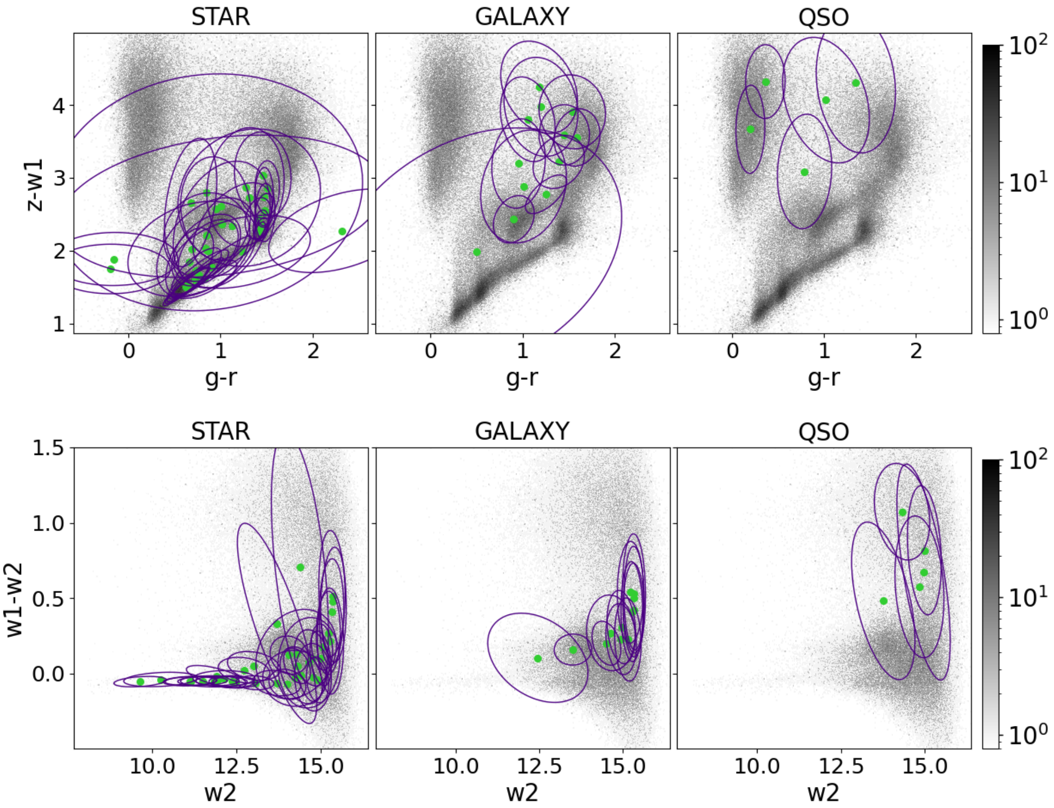

Figure 2.5: Colour-colour and colour-magnitude diagrams. Shown are $g-r$ vs $z-W1$ colour-colour diagrams (upper panel) and $W2$ vs $W1-W2$ colour-magnitude diagrams (bottom panel) for a populations of objects from the spectroscopic data set such as stars (left column), galaxies (middle column) and quasars (right column). The purple contours correspond to the 68-th percentile of each Gaussian IGMM component. The green filled circles correspond to the means μ_k of the Gaussian components. The grey scale indicates the abundance of the sources in each diagram.

2.3.3 Mixture density networks

MDNs are a form of ANNs, which are capable of arbitrarily accurate approximation to a function and its derivatives based on the Universal Approximation Theorem (Hornik, 1991). ANNs can be used for regression or classification purposes. ANNs are structured in layers of neurons, where each neuron receives an input vector from the previous layer, and outputs a nonlinear function of it that is passed on to the next layer. In MDNs, the aim is to approximate a distribution in the product space of input vectors of the individual sources ($\mathbf{f}_i$) and target values (e.g. $z_{s,i}$) as a superposition of different components. MDNs (Bishop, 1994) are trained to optimise the log-likelihood

$$\log \mathscr{L} = \sum_{i=1}^{N} \log \left(\sum_{k=1}^{N_c} \hat{p}_k(\mathbf{f}_i) \mathscr{N}(z_{s,i} | m_k(\mathbf{f}_i), s_k(\mathbf{f}_i)) \right) \tag{2.1}$$

by approximating the averages $m_k(\mathbf{f})$, amplitudes $\hat{p}_k(\mathbf{f})$ and widths $s_k(\mathbf{f})$. Here, N is the number of objects in the spectroscopic data set, while N_c denotes the number of output components (or branches) of the MDN.

Due to the limited information provided by the photometric space, a source of a specific spectroscopic class and low redshift can be confused with a different spectroscopic class and high redshift. Therefore, by providing distributions over a full range of redshifts, MDNs can cope with the fact that colours are not necessarily monotonic with redshift (as is the case e.g. in quasars). In order to avoid confusing MDN components with IGMM components, here we call MDN components branches.

For the sake of reproducibility, we use a publicly available MDN wrapper around the `keras` ANN module[4] and a simple MDN architecture. The MDN input layer contains the same photometric features (see 2.3.1) along with the membership probabilities of the IGMM, which carry additional information of the object classes (stars, galaxies and quasars). The dimension of the MDN input space is 74, of which 52 are the IGMM membership probabilities and 22 are the feature-space entries. The output layer of the MDN is defined by three neurons for each branch: the average redshift on the branch, the width of the branch and the membership probability of the source to the branch. The MDN is fully connected, that is the neurons in one layer are connected to all of

---

[4] https://github.com/cpmpercussion/keras-mdn-layer

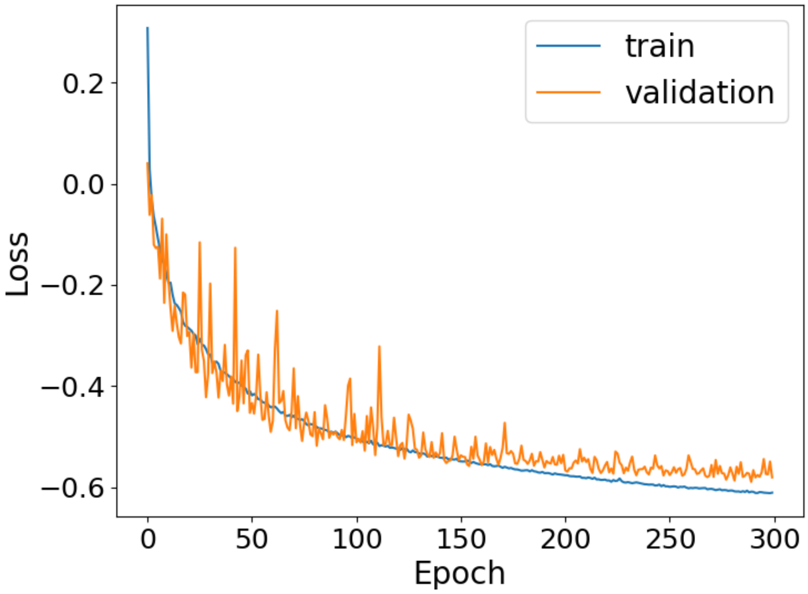

Figure 2.6: MDN Loss ($-\log(\mathscr{L})/N$) as a function of epoch. The loss obtained during the MDN training and validation are shown by blue and orange lines, respectively.

the neurons in the next layer. Due to the fact that the MDN input contains the IGMM membership probabilities, after MDN hyperparameter optimisation, we train one MDN for each of the four IGMM implementations as described in previous sections.

We randomly split the entire spectroscopic data set (2.3.1) and use 80% for training and 20% for validation of the MDN. In order to optimise the MDN, we explored neural networks with 0 to 3 hidden layers. Each layer with a discrete number of neurons in the interval [3, 7, 10, 74, 100, 156, 222, 300, 400, 500, 528, 600, 740]. Furthermore, we used a discrete number of branches for the MDN in the interval [10, 52, 56, 100, 300]. The standard rectified linear unit (ReLU, Nair and Hinton, 2010) and parametric rectified linear unit (PReLU, He *et al.*, 2015) are used as activation functions. We used `ADAM` as optimiser and batch learning with 64 objects per epoch with learning rates of 10^{-6}, 10^{-5}, 10^{-4} and 10^{-3} to mitigate local minima of the loss function.

By comparing the training and validation loss of MDNs with the previously defined set of hyperparameters, the resulting optimal set of hyperparameters contains a hidden layer with 528 neurons and 10 MDN branches. The activation function is PReLU and the learning rate for the ADAM optimiser is 10^{-4}. Figure 2.6 shows the loss function, $-\log(\mathscr{L})/N$, for the training and validation data set, for the MDN optimisation for

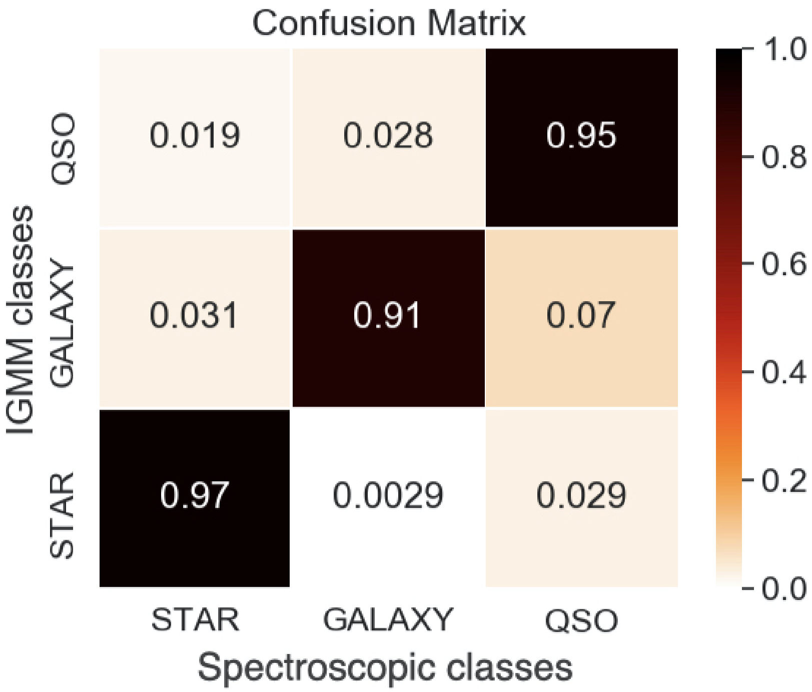

Figure 2.7: IGMM confusion matrix. The spectroscopic classifications are shown against the IGMM classes of the spectroscopic data set.

which membership probabilities are obtained from the partially supervised IGMM realisation that also considers the spectroscopic classes. As Figure 2.6 shows, the learning curve flattens roughly around 300 epochs. To mitigate overfitting, we concluded that 300 epochs are sufficient to train the model. Additionally to training MDNs with the redshifts as targets, we tested $\log(z_s)$ as a target and it led to an improvement in the z_p estimation.

2.4 Results

We trained an IGMM on the photometric data set (see sect. 2.3.1), using the optimal hyperparameters (sect. 2.3.2). Thereafter, we linked IGMM components to the three spectroscopic classes using a spectroscopic data set (2.3.1). Finally, we implemented MDNs on the spectroscopic data set using photometric features and membership probabilities from the IGMM to estimate the conditional probability distribution $p(z_p|\mathbf{f})$ of

Table 2.1: Percentage of objects from each spectroscopic class (stars, galaxies, quasars) within each IGMM component. The components highlighted in red lie between different spectroscopic class regions in photometric feature space, and can reduce the classification accuracy.

IGMM components	Stars	Galaxies	Quasars
1	85.42	0.27	14.31
2	99.98	0	0.02
3	97.46	0.06	2.48
4	100	0	0
5	1.57	0.54	97.88
6	99.86	0.05	0.1
7	3.7	86.06	10.24
8	100	0	0
9	97.45	0.05	2.5
10	8.94	71.67	19.39
11	1.97	90.16	7.87
12	6.95	52.25	40.80
13	99.6	0	0.4
14	100	0	0
15	42.38	43.22	14.41
16	55.39	0.43	44.18
17	99.93	0.01	0.06
18	96.75	2.48	0.77
19	6.58	36.44	56.98
20	99.89	0	0.11
21	1.14	94.51	4.35
22	98.02	0.07	1.90
23	99.94	0	0.06
24	3.69	89.54	6.77
25	100	0	0
26	99.94	0.01	0.05
27	97.48	0.47	2.05
28	100	0	0
29	12.31	20.04	67.65
30	100	0	0
31	1.02	96.60	2.38
32	11.13	35.58	53.28
33	99.96	0.02	0.02
34	99.71	0	0.29
35	100	0	0
36	99.8	0.1	0.1
37	34.23	42.05	23.72
38	100	0	0
39	8.43	51.74	39.83
40	99.91	0	0.09
41	99.51	0.04	0.45
42	100	0	0
43	4.43	88.61	6.97
44	0.56	98.18	1.25
45	90.3	0.83	8.87
46	79.57	1.22	19.21
47	2.87	65.41	31.72
48	0.73	0.05	99.21
49	100	0	0
50	60.24	0.74	39.02
51	44.52	26.33	29.15
52	95.64	0.04	4.32

photo-z values from the photometric inputs. In this section, we describe the evaluation methods and the resulting classification and photo-z estimations.

2.4.1 Classification

With our mixture models we address the common problem of cross contamination among different classes of objects due to the a priori unknown underlying spectral energy distribution. In the IGMM realisations, each object can belong to each of the components with a probability $p_{i,k} = w_k \mathcal{N}(\mathbf{f}_i|\mu_k, \Sigma_k)/\sum_l (w_l \mathcal{N}(\mathbf{f}_i|\mu_l, \Sigma_l))$, which we denote by membership probabilities in the following. As we introduced above (end of Sect. 2.3.2), the simplest way to assign an object (with feature vector $\mathbf{f}_i$) to a component is to consider the component index $\hat{k}$ for which $p_{i,\hat{k}}$ is maximised.

To parameterise the accuracy of the classification, we consider the usual quantification of true and false positives and true and false negatives (e.g. Fawcett, 2006), and build a confusion matrix to quantify the rate of correct classifications. Figure 2.7 shows the confusion matrix of the GMM-based classification for the spectroscopic data set. The true positive rates[5] for stars, galaxies and quasars are 0.97, 0.91 and 0.95, respectively. False positive rates for stars that are true galaxies and quasars are 0.0029 and 0.029. False negative rates for stars that are assigned to galaxies and quasars are 0.031 and 0.019 of all stars, respectively. The accuracy[6] is $\approx 94\%$. This means that the IGMM part of our mixture models can clean an extragalactic sample from most of stellar contaminants, and broadly separate galaxies from AGN-dominated objects. We compare the performance of our classification to that of a `HDBSCAN` classification (Campello *et al.*, 2013) of a subset of 39 447 objects from the SDSS-DR14 that have an accuracy of $\approx 95\%$ using $ugrizYHJKw1w2$-band magnitudes (Logan, C. H. A. and Fotopoulou, S., 2020). We find that our method achieves a comparable accuracy even without u-band and $YJHK$-band magnitudes.

Figure 2.5 demonstrates that the IGMM recognises the main behaviours of stars, galaxies and quasars in colour space and also identifies subclasses that are not highly represented in the spectroscopic sample, such as white dwarfs and brown dwarfs. On the other hand, some components happen to lie in regions of the colour-magnitude-extinction space that

---

[5] Defined as: TP/(TP+FN).

[6] Defined as: (TP+TN)/(TP+TN+FN+FP).

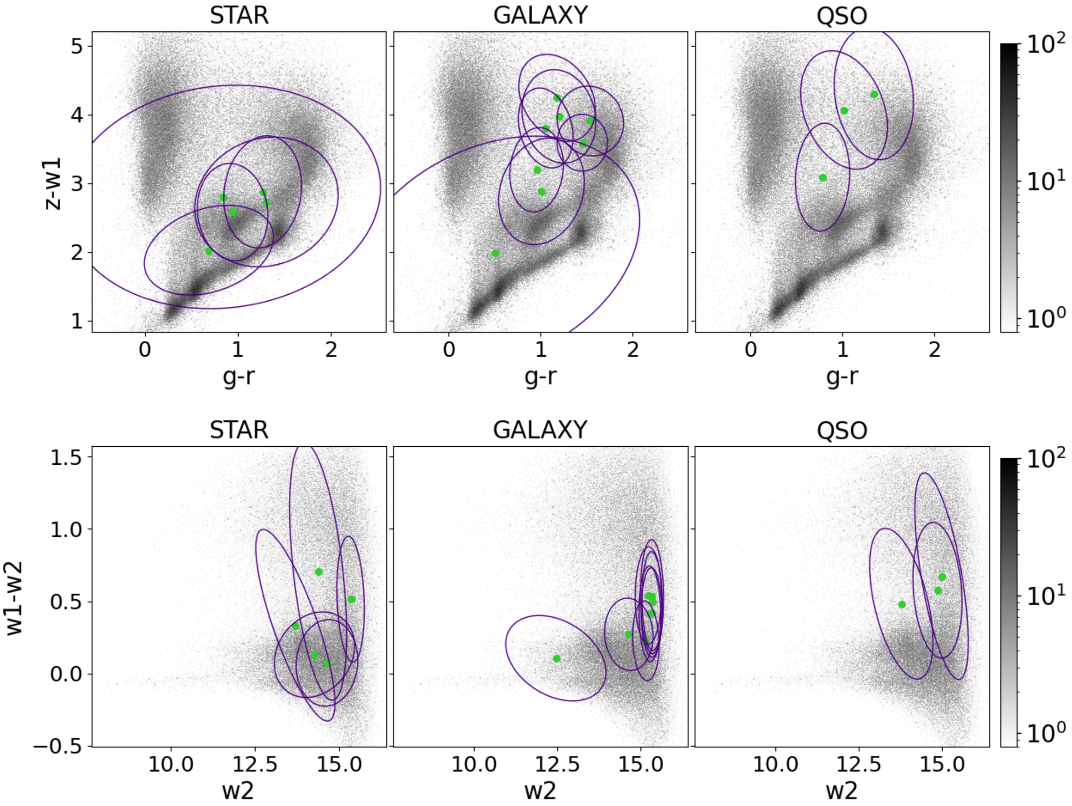

Figure 2.8: Colour-colour and colour-magnitude diagrams. Shown are $g-r$ vs $z-W1$ colour-colour diagrams (upper panel) and $W2$ vs $W1-W2$ colour-magnitude diagrams (bottom panel) for objects from the spectroscopic data set of the three spectroscopic classes such as stars (left column), galaxies (middle column) and quasars (right column). The purple contours correspond to the 68-th percentile of the problematic Gaussian components of the IGMM that are not dominated by objects of just one spectroscopic class. The green filled circles correspond to the means μ_k of these components. The grey scale indicates the number of sources in each diagram.

are not dominated by only one subclass. The overlap between different object classes in photometry can affect the classification performance and the output of the classification that is then used by the MDN regression. The components corresponding to regions of overlap between different classes are discussed below.

Approximately 30% of IGMM components that cover ≈ 15% of the spectroscopic data set, marked in red in Table **??**, contain a non-negligible fraction of objects from more than one of the three main classes. Figure 2.8 shows their position in the same colour-colour and colour-magnitude diagrams as Figure 2.5. We address these components as 'problematic components'.

As expected, the problematic components lie at the faint end (with higher magnitude uncertainties in WISE), or in intermediate regions of the colour space between AGN-dominated and galaxy-dominated systems. Additionally, the SDSS spectroscopic classification of some objects is ambiguous and for some cases the automatic classification (by the SDSS spectral pipelines) is either erroneous or has multiple incompatible entries[7]. These issues occur more frequent for fainter objects which have spectra with low signal-to-noise ratio[8]. However, since most of the objects are clustered in three main classes which are correctly identified by the IGMM components, uncertain spectroscopic labels are not a significant problem for our calculations.

2.4.2 Photometric redshifts

Here we discuss different metrics employed to evaluate the performance of our methods used to determine photometric redshifts. Most metrics are based on commonly used statistical methods as outlined. The prediction bias is defined as the mean of weighted residuals, $\Delta z = (z_p - z_s)/(1 + z_s)$, which is the same as the definition in Cohen *et al.*, 2000. The root-mean-square of the weighted residuals is defined as $\mathrm{rms}(\Delta z)$. The fraction of outliers is defined as the number of objects with $3 \times \mathrm{rms}(\Delta z) < \Delta z$.

For all methods, we excluded objects with spectroscopic redshift errors $\delta z_s > 0.01 \times (1 + z_s)$. For each source, the MDN determines a full photo-z distribution, which is a superposition of all branches, each with a membership probability, average, and dispersion. If one so-called point estimate is needed, there are at least two options to compute it. One option is the expectation value

$$\mathbb{E}(z_{p,i}|\mathbf{f}_i) = \frac{\sum_k \mu_k(\mathbf{f}_i)\hat{p}_k}{\sum_k \hat{p}_k} , \tag{2.2}$$

weighted across all branches that an object can belong to, according to its branch membership probabilities. Another choice, which we follow here, is the peak $\mu_r(\mathbf{f}_i)$ of the branch that gives the maximum membership probability of a given object. In what follows, we refer to this redshift value as the 'peak photo-z'.

---

[7] For example for `OBJID=1691188859137714176` from SDSS-DR15

[8] For example for `OBJID=743142903307593728` from SDSS-DR15

Table 2.2: MDN performance evaluation, without any clipping for the average and rms, without any threshold on branch membership probabilities.

IGMM implementation	photometry	$\langle\Delta z\rangle$ (all)	rms(Δz) (all)	3σ outliers (all)	$\langle\Delta z\rangle$, range1a	rms(Δz), range1a	3σ outliers, range1 a	rms(Δz), range2b	rms(Δz), range3c
Fully unsup.	$griz$, $W1$, $W2$	0.0152	0.2174	3.08%	0.0007	0.0177	0.28%	0.0988	0.0945
spec. class	$griz$, $W1$, $W2$	0.0111	0.2069	1.31%	0.0006	0.0167	0.41%	0.0822	0.0783
spec. classd	$griz$, $W1$, $W2$	0.0356	0.2300	1.35%	0.0110	0.0260	0.71%	0.0953	0.0903
redshift (z_s)	$griz$, $W1$, $W2$	0.0176	0.2131	3.21%	-0.0009	0.0174	0.38%	0.0896	0.0873
spec. class, z_s	$griz$, $W1$, $W2$	0.0047	0.1990	2.66%	0.0036	0.0181	0.57%	0.0675	0.0664
spec. class	$ugriz$, $W1$, $W2$	0.0135	0.1592	1.62%	0.0007	0.0160	0.23%	0.0601	0.0611

Spectroscopic sample for all IGMM implementations containing stars, galaxies and quasars.

aRestricted to galaxies with $z_s < 0.3$; bRestricted to galaxies with $z_s < 0.4$; cRestricted to galaxies with $z_s < 0.5$; dExpectation value.

Fully unsup.: Fully unsupervised IGMM implementation. Spec. class: uses spectroscopically classified objects (e.g. stars, galaxies, and quasars) for those which SDSS provides spectroscopic information (i.e. $\approx 2\%$ of the photometric data set) to classify the data. Redshift(z_s): uses the spectroscopic redshift of objects from $\approx 2\%$ of the photometric data set to classify the data. Spec. class,(z_s): uses both, spectroscopically classified objects and the spectrsocopic redshift of objects from $\approx 2\%$ of the photometric data set to classify the data.

Table 2.3: MDN performance evaluation exclusively for sources with MDN branch $weight_{max} > 0.8$, without any clipping for the average and rms.

IGMM implementation	photometry	$\langle \Delta z \rangle$ (all)	rms(Δz) (all)	3σ outliers (all)	$\langle \Delta z \rangle$, range1a	rms(Δz), range1a	3σ outliers, range1a	rms(Δz), range2b	rms(Δz), range3c
Fully unsup.	$griz$, $W1$, $W2$	0.0032	0.1165	1.00%	0.0007	0.0177	0.60%	0.0350	0.0360
spec. class	$griz$, $W1$, $W2$	0.0031	0.1244	0.93%	0.0006	0.0167	0.83%	0.0405	0.0391
redshift (z_s)	$griz$, $W1$, $W2$	0.0035	0.1076	0.79%	-0.0009	0.0174	0.52%	0.0299	0.0331
spec. class, z_s	$griz$, $W1$, $W2$	-0.0048	0.1170	1.02%	0.0036	0.0036	1.02%	0.0337	0.0314
spec. class	$ugriz$, $W1$, $W2$	0.0043	0.0934	0.66%	0.0007	0.0160	0.92%	0.0334	0.0341

aRestricted to galaxies with $z_s < 0.3$; bRestricted to galaxies with $z_s < 0.4$; cRestricted to galaxies with $z_s < 0.5$.

The IGMM implementation abbreviations are the same as for Table 2.2

In a fully Bayesian framework, one would also consider the maximum-a-posteriori, which in our approach would correspond with the photo-z that maximises the MDN output sum of Gaussians because all our priors are uniform. Strictly speaking, in principle the maximum-a-posteriori and the peak redshift are not the same, if the MDN Gaussian with the highest membership probability lies close to other MDN Gaussians. Here, for the sake of computational convenience we choose the peak redshift, leaving the comparison with other possible choices (including the maximum-a-posteriori, or non-uniform priors) to future investigation. We also note that this distinction becomes important for objects whose membership probabilities are not clearly dominated by one of the MDN output Gaussians. In the following, we examine the photo-z estimation performance on objects whose membership probabilities to one of the MDN Gaussians is $> 80\%$, in which case there is almost no difference between peak redshit and maximum-a-posteriori redshift.

Figure 2.9 shows the distribution of peak photo-zs (top) and expectation photo-zs (bottom) versus spectroscopic redshifts, z_s, for the MDN run with ten branches. One aspect to consider when determining photo-z in cosmological wide-field imaging surveys, is the availability of $u-$band magnitudes, which is currently available for KiDS but not for DES. The *Rubin* LSST is expected to deliver $u-$band photometry at the same depth of KiDS over $\approx 12\,300 \mathrm{deg}^2$ (Y1) and $\approx 14\,300 \mathrm{deg}^2$ (Y2) (The LSST Dark Energy Science Collaboration *et al.*, 2018). To test the effect, we re-trained one of our mixture models (IGMM spec. class) for a data set that includes $u-$band `PSF` and `model` magnitudes as additional input features (Fig. 2.10). The bias and root-mean-square residuals are provided in Table 2.2 for all objects and for galaxies with spectroscopic redshifts $z_s < 0.3$, $z_s < 0.4$, and $z_s < 0.5$. This test leads to a lower rms Δz and smaller fraction of 3σ outliers than for the same model without u-band magnitudes and can be considered an improvement in accuracy. One reason is that for low redshift objects, u-band contains information on the position and strength of the $4000\AA$ Balmer-break, which is not covered by other, longer wavelength bands. Furthermore, with respect to the cross-contamination problem, this model also improves the overall confidence level with which an object belongs to a branch. As demonstrated in Fig. 2.10, bottom panel, the MDN performs indeed better for objects with increased confidence level.

Table 2.4: Comparison between the photo-z evaluation on all objects from the spectroscopic sample with available SDSS photo-zs.

no k-fold in MDN; IGMM implementation	restriction on objects	number of objects	Bias	rms	3σ outliers
SDSS *griz* photo-z	None	38 632	0.0001 ± 0.0008	0.0458 ± 0.0483	(0.73 ± 0.22)%
MDN spec. class + *griz*, $W1$, $W2$	None	38 632	0.0018 ± 0.0014	0.0478 ± 0.0384	(0.49 ± 0.04)%
SDSS *griz* photo-z	$weight_{max} > 0.8$	18 355	-0.0038 ± 0.0040	0.0394 ± 0.0413	(0.46 ± 0.19)%
MDN spec. class + *griz*, $W1$, $W2$	$weight_{max} > 0.8$	18 355	-0.0002 ± 0.0009	0.0389 ± 0.0318	(0.55 ± 0.26)%

k-fold in MDN; IGMM implementation	restriction on objects	number of objects	Bias	rms	3σ outliers
SDSS *griz* photo-z	None	38 632	0.0001 ± 0.0020	0.0662 ± 0.0072	(0.79 ± 0.22)%
MDN spec. class + *griz*, $W1$, $W2$	None	38 632	0.0027 ± 0.0031	0.0724 ± 0.0139	(0.45 ± 0.10) %
SDSS *griz* photo-z	$weight_{max} > 0.8$	23 908	-0.0028 ± 0.0029	0.0506 ± 0.0096	(0.33 ± 0.14)%
MDN spec. class + *griz*, $W1$, $W2$	$weight_{max} > 0.8$	23 908	0.00011 ± 0.0042	0.0458 ± 0.0069	(0.28 ± 0.07)%

2.5 Conclusion

Performance evaluations of MDN photo-zs from different IGMM realisations are summarised in Table 2.2. The minimum MDN photo-z bias and rms of different IGMM realisations without using u-band magnitudes is reached by the IGMM realisation (Spec. class,(z_s)) that uses the spectroscopic classifications and spectroscopic redshifts of $\approx 2\%$ of the dataset. This implies that the IGMM can provide a reasonable description of the objects in their feature space even with very limited spectroscopic information. On the other hand, for this realisation the fraction of 3σ outliers is not as low as for other IGMM realisations (Tables 2.2 and 2.3) which have a slightly higher bias and rms. For the validation samples that are restricted to galaxies with a spectroscopic redshift $<$ 0.3, the minimum MDN photo-z bias, rms and fraction of 3σ outliers are achieved for the realisation that includes u-band magnitudes. The MDN performance of different IGMM realisations for objects that belong to one MDN branch with high confidence (i.e $weight_{max} > 0.8$) is summarised in Table 2.3. In order to compare our photo-z estimates to those from the SDSS, we select objects from the SDSS `SpecObjAll` table that have photo-zs (obtained with a kNN interpolation) from the SDSS `PhotoObjAll` (38 487 objects). We remark that our MDN yields the full photo-z PDF, so a choice must be made when comparing its results with point estimates from other methods in the literature. For this reason, we follow our previous choice and consider the peak photo-z's. We compare our photo-z's to the SDSS ones for the full sample (38 487 objects) and also for the subset of objects that belong to a MDN branch with high confidence ($weight_{max} > 0.8$, 18 355 objects). To account for uncertainties in bias and rms due to the finite sample size, we split the whole sample into five sub-sets, compute the bias and rms on each, and then report their average and standard deviation. As an alternative, we also evaluate the bias, rms and fraction of outliers when the MDN is trained with a k-fold cross-validation method (see e.g Altman and Bland, 2005) on the full SDSS training set ($\approx 245,000$ objects), where each of the k=5 folds exclude a fifth of the SDSS photo-z sample. As summarised in Table 2.4 and shown in Figure 2.11, our method improves the bias of photo-z estimates by about one order of magnitude compared to SDSS photo-zs for objects for which our model estimated the photo-zs at high confidence (i.e. $weight_{max} > 0.8$), and our method also decreases the rms and the fraction of outliers. All metrics are improved, with the added advantage that the MDN computes photo$-z$s for all objects (instead of just those with low stellarity) and can also cover the $z_s > 1$ range more accurately than the SDSS kNN. As a matter of fact, the

SDSS photo-zs hardly exceed $z_p \approx 1$, while our machinery is trained over a much wider redshift range.

As a general benchmark, the LSST system science requirements document [9] defines three photometric redshift requirements for a sample of four billion galaxies with $i < 25$ mag within $z_s < 0.3$ as follows: First, for the error in $(1 + z_s)$ the rms of residuals is lower than 0.02. Secondly, the fraction of 3σ ('catastrophic') outliers is lower than 10% of the sample and thirdly, the bias is lower than 0.003.

In our approach, these requirements are met if the MDN peak z_p is adopted. The rms Δz can be brought to 0.02 over $0 < z_s < 0.5$ both for all objects (Table 2.5) and for 'high-confidence' objects with > 0.8 membership probability to a branch (Table 2.6; called weight in Sect. 2.3.3). Although our training and evaluation samples from the SDSS do not reach as deep as the *Rubin* LSST is expected to, our method shows promising results as a starting point. A general remark, which holds for all photo-z estimation methods is that a re-training for the LSST regime may also need deeper photometric and more robust spectroscopic samples. Recently, Beck *et al.* (2020a) used neural networks to classify objects in the Pan-STARRS1 footprint, which is known to have a more accurate photometry than the SDSS (Magnier *et al.*, 2013), and evaluated photo-zs on objects with a probability $p > 0.8$ of being galaxies, obtaining rms(Δz)=0.03 over $0 < z_s < 1$. If we follow the same definitions and clipping[10] as by Beck *et al.* (2020a), then we obtain 1.7-2% relative rms over the $0 < z_s < 0.5$ redshift range. Our code for re-training and any further evaluation is publicly accessible on github.com/ZoeAnsari/MixtureModelsForPhotometricRedshifts

2.6 Discussion

Adding $u-$band information, as is the case with the SDSS and will be the case with the LSST, reduces the bias and fraction of outliers in all the redshift ranges considered. This is also because adding $u-$band magnitudes sharpens the MDN separation into branches and increases the fraction of objects with the highest weighted branch > 0.8, as can be seen in the bottom panels of Figure 2.9.

---

[9] https://docushare.lsstcorp.org/docushare/dsweb/Get/LPM-17

[10] Their clipping procedure removes objects with $|\Delta z| > 0.15$.

Table 2.5: MDN performance evaluation. The bias and rms are computed using the definition of clipped bias and rms in PS1-STR (Beck *et al.*, 2020a).

IGMM implementation	photometry	$\langle \Delta z \rangle$ $(z_s < 1)$	rms(Δz) $(z_s < 1)$	$\langle \Delta z \rangle$, range1[a]
Fully unsup.	*griz*, $W1$, $W2$	0.0007 ± 0.0006	0.0204 ± 0.0106	0.0025 ± 0.0005
spec. class	*griz*, $W1$, $W2$	0.0030 ± 0.0028	0.0216 ± 0.0104	0.0042 ± 0.039
redshift (z_s)	*griz*, $W1$, $W2$	0.0022 ± 0.0024	0.0207 ± 0.0113	0.0008 ± 0.0003
spec. class, z_s	*griz*, $W1$, $W2$	0.0001 ± 0.0005	0.0208 ± 0.0112	0.0058 ± 0.0006
spec. class	*ugriz*, $W1$, $W2$	0.0023 ± 0.0026	0.0181 ± 0.0091	0.0004 ± 0.0025

IGMM implementation	photometry	rms(Δz), range1[a]	rms(Δz), range2[b]	rms(Δz), range3[c]
Fully unsup.	*griz*, $W1$, $W2$	0.0215 ± 0.0013	0.0233 ± 0.0013	0.0238 ± 0.0015
spec. class	*griz*, $W1$, $W2$	0.0230 ± 0.0015	0.0245 ± 0.0019	0.0247 ± 0.0014
redshift (z_s)	*griz*, $W1$, $W2$	0.0219 ± 0.0014	0.0238 ± 0.0012	0.0242 ± 0.0013
spec. class, z_s	*griz*, $W1$, $W2$	0.0218 ± 0.0019	0.0235 ± 0.0019	0.0235 ± 0.0024
spec. class	*ugriz*, $W1$, $W2$	0.0179 ± 0.0015	0.0182 ± 0.0016	0.0187 ± 0.0019

[a] Restricted to galaxies with $z_s < 0.3$; [b] Restricted to galaxies with $z_s < 0.4$; [c] Restricted to galaxies with $z_s < 0.5$.

Table 2.6: MDN performance evaluation exclusively for sources with MDN branch $weight_{max} > 0.8$. The bias and rms are computed using the definition of clipped bias and rms in PS1-STR (Beck *et al.*, 2020a).

IGMM implementation	photometry	$\langle\Delta z\rangle$ ($z_s < 1$)	rms(Δz) ($z_s < 1$)	$\langle\Delta z\rangle$, range1a
Fully unsup.	$griz$, $W1$, $W2$	0.0004 ± 0.0010	0.0249 ± 0.0038	0.0014 ± 0.0005
spec. class	$griz$, $W1$, $W2$	0.0035 ± 0.0030	0.0273 ± 0.0045	0.0015 ± 0.0009
redshift (z_s)	$griz$, $W1$, $W2$	0.0018 ± 0.0013	0.0236 ± 0.0045	-0.0004 ± 0.0004
spec. class, z_s	$griz$, $W1$, $W2$	-0.0014 ± 0.0012	0.0264 ± 0.0054	0.0042 ± 0.0005
spec. class	$ugriz$, $W1$, $W2$	0.0036 ± 0.0026	0.0227 ± 0.0058	0.0009 ± 0.0005

IGMM implementation	photometry	rms(Δz), range1a	rms(Δz), range2b	rms(Δz), range3c
Fully unsup.	$griz$, $W1$, $W2$	0.0186 ± 0.0002	0.0201 ± 0.0003	0.0207 ± 0.0004
spec. class	$griz$, $W1$, $W2$	0.0181 ± 0.0002	0.0207 ± 0.0009	0.0214 ± 0.0007
redshift (z_s)	$griz$, $W1$, $W2$	0.0182 ± 0.0004	0.0195 ± 0.0005	0.0203 ± 0.0004
spec. class, z_s	$griz$, $W1$, $W2$	0.0189 ± 0.0013	0.0205 ± 0.0014	0.0207 ± 0.0014
spec. class	$ugriz$, $W1$, $W2$	0.0163 ± 0.0007	0.0167 ± 0.0008	0.0170 ± 0.0010

aRestricted to galaxies with $z_s < 0.3$; bRestricted to galaxies with $z_s < 0.4$; cRestricted to galaxies with $z_s < 0.5$.

We remark that throughout this work, we are simply adopting reddening values in the $i-$band (A_i), which the SDSS provides via a simple conversion of measured $E(B-V)$ values with a Milky-Way extinction law and $R_V = 3.1$. Our approach accounts for the systematic uncertainties due to the unknown extinction law by producing probability distributions and associate uncertainties for each photo-z value.

The combined information across the optical and infrared, through the SDSS and WISE magnitudes, helps reducing the overlap between different classes in colour-magnitude space. The WISE depth is not a major limiting factor in the sample completeness as long as samples from the SDSS are considered, but it can affect the completeness significantly for deeper surveys (Spiniello and Agnello, 2019). In view of performing the classification and photo-z estimation on the DES, and on the *Rubin* LSST later on, deeper mid-IR data are needed. The unWISE reprocessing of the WISE cutouts improved upon the original WISE depth (Lang, 2014). Further in the future, forced photometry of the unWISE cutouts from wide-field optical and NIR surveys may further increase the mid-IR survey depth (e.g. Lang *et al.*, 2014).

In general, separating objects into many sub-classes aids the photo-z regression, as each MDN branch only needs to consider a subset of objects with more homogeneous properties than the whole photometric sample. Furthermore, the approach that we used in this work is both in the realm of machine learning (hence less constrained by choices of templates) while it can also produce a full output distribution for the photo$-z$ given the available photometric information. Beyond their first implementation in this work, mixture models can be easily adapted so that they can account for missing entries and limited depth, as in the GMM implementation by Melchior and Goulding (2018).

Acknowledgements

This work is supported by a VILLUM FONDEN Investigator grant (project number 16599), a VILLUM FONDEN Young Investor Grant (project number 25501), and a Villum Experiment Grant (project number 36225). This project is partially funded by the Danish council for independent research under the project “Fundamentals of Dark Matter Structures”, DFF–6108-00570.

2.7 Appendix

2.7.1 Summary of Photometric Redshifts in the Literature

2.7.2 IGMM

Probability density distribution (PDF) formalisation by Gaussian mixture modelling for K components is defined as follows:

$$P(x|\mu_1,\ldots,\mu_K,\Sigma_1,\ldots,\Sigma_K) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k,\sigma_k) \tag{2.3}$$

where x is the data, π_k is the weight distribution of mixtures that is defined by a Dirichlet distribution and $\sum_{k=1}^{K} \pi_k = 1$.

IGMM is the GMM case with infinite number of components using Dirichlet process instead of Dirichlet distribution to define the prior over the mixture distribution. Dirichlet process is a distribution over distributions, parameterising by concentration parameter γ and a base distribution G_0. The base distribution is the Dirichlet distribution which is a prior over the locations of components in the parameter space (i.e. $\Theta = (\mu,\Sigma)$). The concentration parameter γ expresses the strength of belief in G_0 and affects the components weight (Görür and Edward Rasmussen, 2010).
Based on Bayes rule:

$$\gamma Z_i(k) = P(Z_i = k|x) = \frac{P(k)P(x|Z_i = k)}{P(x)} = \frac{\pi_k \mathcal{N}(x|\Theta_k)}{\sum_{k=1}^{k} \pi_k \mathcal{N}(x|\Theta_k)} \tag{2.4}$$

where $\underline{\pi}$ is considered as the Dirichlet process and Z_i is the latent variable. $\pi_k = N_k/N$ represents the effective number of data points assigned to the k-th mixture component. Despite the fact that we do not know the latent variable, there is information about it in the posterior.

Using an expectation-maximisation (EM) algorithm to find the maximum likelihood with respect to the model parameters includes two steps, estimation step (e-step) and

Table 2.7: Recent automated approaches to estimate photo-zs.

Ref.	Method^a	Photometric information	Objects	z_s range^b	Depth [mag]^c	Survey
1	kNN	$ugrizy^d$	Galaxies	$0 < z \leq 2$	$i < 25.3$	mock galaxy for LSST from DESC
2	ANN	$ugriz^e$, $E(B-V)$	Galaxies	$z < 0.4$	$r_{Petro} \leq 17.8$	SDSS-DR12
3	METAPHOr, ANN, template fitting	*ugri* GAaP	Galaxies	$z_s \leq 1$	$r \leq 21$	SDSS-DR9, KiDS ESO-DR3, GAMA-DR2, 2dFGRS
4	ANN	$ugriz^e$	Galaxies	$z_s \leq 0.4$	$r_{Petro} \leq 17.8$	SDSS/BOSS-DR12, GAMA-DR3
5	kNN	UV, *ugrizy*, $YJHK$	Galaxies	$0.3 < z_s < 3.0$	$i < 25$	mock galaxy catalogues for *Euclid*, RST, and/or CASTOR
6	kNN	UV, *ugriz*, $w1w2w3w4$	QSOs	—	$14.7 < r < 22.6$	SDSS-DR12, 2MASS, WISE
7	kNN	*grizy*	Galaxies	$z_s > 0.01$	$18.5 < i < 25$	SDSS/BOSS-DR14, DEEP2/3DR4, VANDELS-DR2, COSMOS, C3R2, COSMOS2015
8	tree based	*ugriz*, *BRI*	Galaxies	$0.02 \leq z_s \leq 0.3$	B_{AB} <24.1	SDSS/MGS-DR7, DEEP2-DR4
9	tree based	*ugriz*	Galaxies	$z_s \leq 0.55$	$r_{Petro} < 17.77$	SDSS-DR6, 2dF-SDSS LRG, 2SLAQ, DEEP2
10	Gaussian process	*griz*, *RIZ*, YJH	Galaxies	$0 \leq z_s \leq 2$	$RIZ < 25$	SDSS/BOSS
11	ensemble of ANNs, trees and kNN	*ugriz*	Galaxies	$z_s < 0.8$	$i_{AB} \lesssim 22.5$	SDSS/BOSS-DR10
12	ANN, Monte-Carlo, extrapolation	$grizy^f$, $E(B-V)^g$	Galaxies, QSOs, Stars	$z_s < 1.5$	$i \lesssim 23.1$	PS1 3π DR1, SDSS-DR14, DEEP2-DR4, VIPERS PDR-2, WiggleZ, zCOSMOS-DR3, VVDS

(1) Schmidt *et al.* (2020); (2) Pasquet *et al.* (2019); (3) Amaro *et al.*, 2019; (4) Shuntov *et al.*, 2020; (5) Graham *et al.*, 2018; (6) Curran, 2020; (7) Nishizawa *et al.*, 2020; (8) Carrasco Kind and Brunner, 2013; (9) Gerdes *et al.*, 2010; (10) Almosallam *et al.*, 2016; (11) Sadeh *et al.*, 2019; (12) Beck *et al.*, 2020b

(a)Acronyms are defined in the respective literature; (b)Spectroscopic redshift range. ; (c)Petrosian r-band magnitude, r_{Petro}; (d)Grey scale 48×48 pixel images; (e)Images in *ugriz*, 64×64 piexels in each band; (f)Magnitudes for PSF, Kron and seeing-matched apertures (FPSFMag, FKronMag and FApMag, respectively), as well as 3.00", 4.63" and 7.43" fixed-radius apertures (FmeanMagR5, FmeanMagR6 and FmeanMagR7); (g)PS1 and Planck extinction maps.

Table 2.8: Comparison of photo-z estimates.

Ref.	Method^a	Bias^b	rms^c	Fraction of outlier in %
1	(trainZ)	−0.2086	0.1808	0
	ANNz2	0.00063	0.0270	4.4
	BPZ	−0.00175	0.0215	3.5
	Delight	−0.00185	0.0212	3.8
	EAZY	−0.00218	0.0225	3.4
	FlexZBoost	−0.00027	0.0154	2.0
	GPz	0.00000	0.0197	5.2
	Lephare	−0.00161	0.0236	5.8
	METAPhoR	0.00000	0.0264	3.7
	CMNN	−0.00132	0.0184	3.5
	SkyNet	−0.00167	0.0219	3.6
	TPZ	0.00309	0.0161	3.3
2	Convolutional neural network (CNN)	0.0001	0.0456^d	0.31
3	METAPHOR	−0.004	0.065	0.98
	ANNz2	−0.008	0.078	1.60
	BPZ	−0.020	0.048	1.13
4	CNN + density field (mode)	0.0038d	—	0.83
	CNN + density field (median)	0.0045d		0.44
	CNN + density field (mean)	0.0066d		0.31
5	kNN	−0.0001 ± 0.0	0.0165 ± 0.0001	4.0
6	kNN	0.001^e	0.36	10.7^f
7	DEmP^g	-0.0291	0.1018	0.16
	DEmP^h	-0.0175	0.07	0.17
8	Trees and Random Forest(Regression mode)	-0.00008	0.0225	0
	Trees and Random Forest (Classification mode)	0.00218	0.0246	0
9	ArborZ	-0.006^e	0.985	1.9
10	GP-GL	0.0946	0.1420	5.3
	GP-VL	0.828	0.1251	5.5
	GP-VC	0.0294	0.0435	4.7
11	ensemble of ANNs, trees and KNN (nominal solution)	0.0002	0.034	0.105
	ensemble of ANNs, trees and KNN($< PDF >$)	0.00035	0.034	0.105
	ensemble of ANNs, trees and KNN(PDF)	0.00035	0.052	0.1
12	PS1-STRM (All validation) base estimate	0.0003	0.0342	2.88 ^{i}
	PS1-STRM (All validation) Monte-Carlo sampled	0.0010	0.0344	2.99
	PS1-STRM (Non-extrapolated) base estimate	0.0005	0.0322	1.89
	PS1-STRM (Non-extrapolated) Monte-Carlo sampled	0.0013	0.0323	2.00

(1) Schmidt *et al.* (2020); (2) Pasquet *et al.* (2019); (3) Amaro *et al.*, 2019; (4) Shuntov *et al.*, 2020; (5) Graham *et al.*, 2018; (6) Curran, 2020; (7) Nishizawa *et al.*, 2020; (8) Carrasco Kind and Brunner, 2013; (9) Gerdes *et al.*, 2010; (10) Almosallam *et al.*, 2016; (11) Sadeh *et al.*, 2019; (12) Beck *et al.*, 2020b Values are provided where information was available. aAcronyms are defined in the respective literature; bBias: defined as mean of $\Delta z = (z_p - z_s)/(1+z_s)$; crms($(z_p - z_s)/(1+z_s)$); d$\sigma_{MAD} = 1.4826 \times MAD$, where MAD (Median Absolute Deviation) is the median of $|\Delta z - Median(\Delta z)|$; eAverage of $\delta z = z_p - z_s$; fFraction of outliers defined as number of objects with $|\Delta z| > rms(\Delta z) \pm 0.5$; gExclusively using wide-band photometry from *Wide* fields of HSC (https://hsc.mtk.nao.ac.jp/ssp/) as additional photometric input; hExclusively using deep photometry from *Deep* and *UltraDeep* fields of HSC as additional photometric input; iFraction of outliers defined as number of objects with $|\Delta z| > 0.15$.

maximisation step (m-step). After initialising the model parameters and evaluating the log-likelihood, the e-step evaluates the posterior distribution of Z_i using the current model parameter values by equation (2.4). Then, the m-step updates the model parameters based on the calculated latent variable as follows:

$$\mu_k = \frac{\sum_{i=1}^{N} \gamma Z_i(k) x_i}{\sum_{i=1}^{N} \gamma Z_i(k)} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma Z_i(k) x_i \tag{2.5}$$

$$\Sigma_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma Z_i(k)(x_i - \mu_k)(x_i - \mu_k) \tag{2.6}$$

$$\pi_k = \frac{N_k}{N} \tag{2.7}$$

Where $N_k = \sum_{i=1}^{N} \gamma Z_i(k)$. Eventually, the algorithm detects the convergence by the lack of significant change in the log-likelihood value from one iteration to the next, using:

$$\log P(x|\mu, \Sigma, \pi) = \sum_{i=1}^{N} \log \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k) \right\} \tag{2.8}$$

where π_k, the mixture proportion, represents the probability of x_i belonging to the k-th mixture component.

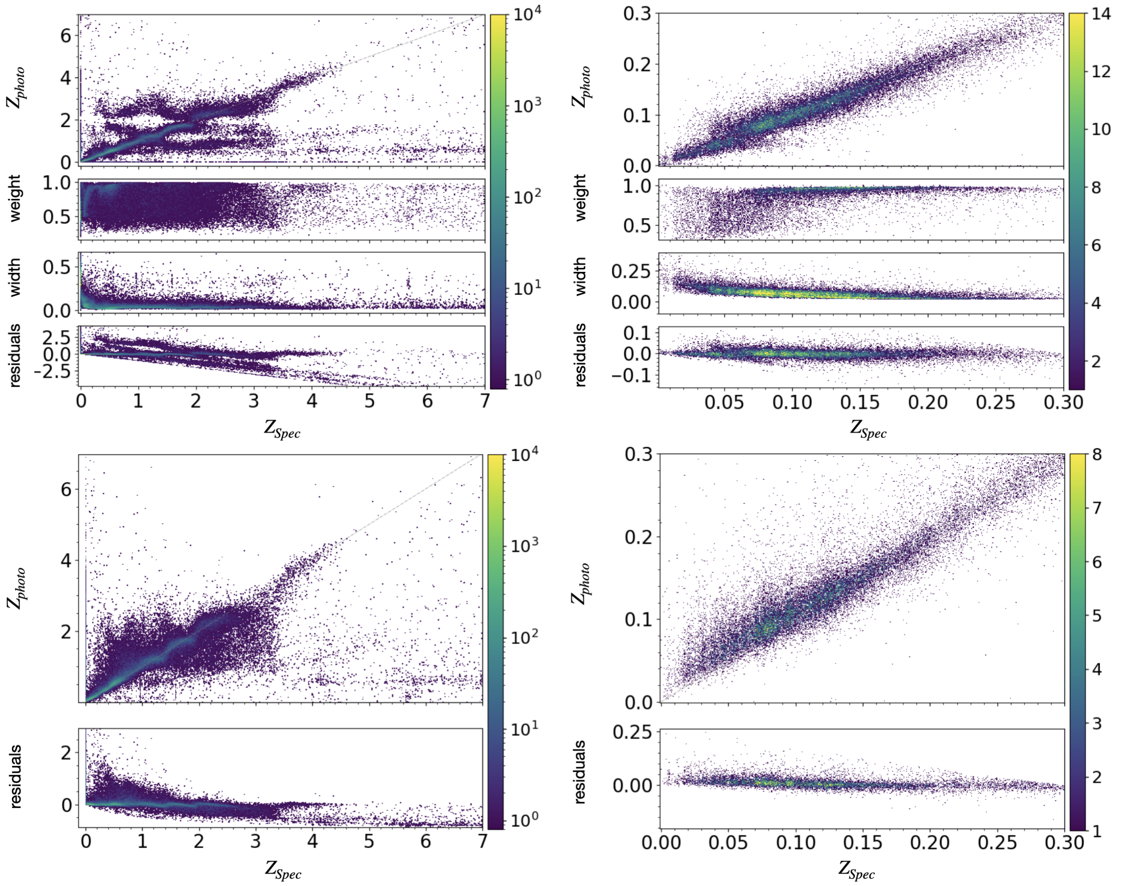

Figure 2.9: Comparison of spectroscopic vs. MDN photometric redshifts. The photometric redshifts are taken from the partially supervised ‘spec. class’ IGMM implementation (as described in Sec. 2.3.2). The colour-scales indicate the number of objects. Top panels: The predicted photometric redshifts that correspond to the branches with the highest weights. The single panels show the weights, dispersions (denoted by ’width’) and residuals from top to bottom. Bottom panels: The mean photometric redshifts of the predicted redshifts over all branches with respect to their weights. The lower panel shows the residuals. Left panels: Include all classes with $z_{spec} < 7$. Right panels: Include all galaxies with $z_{spec} < 0.3$.

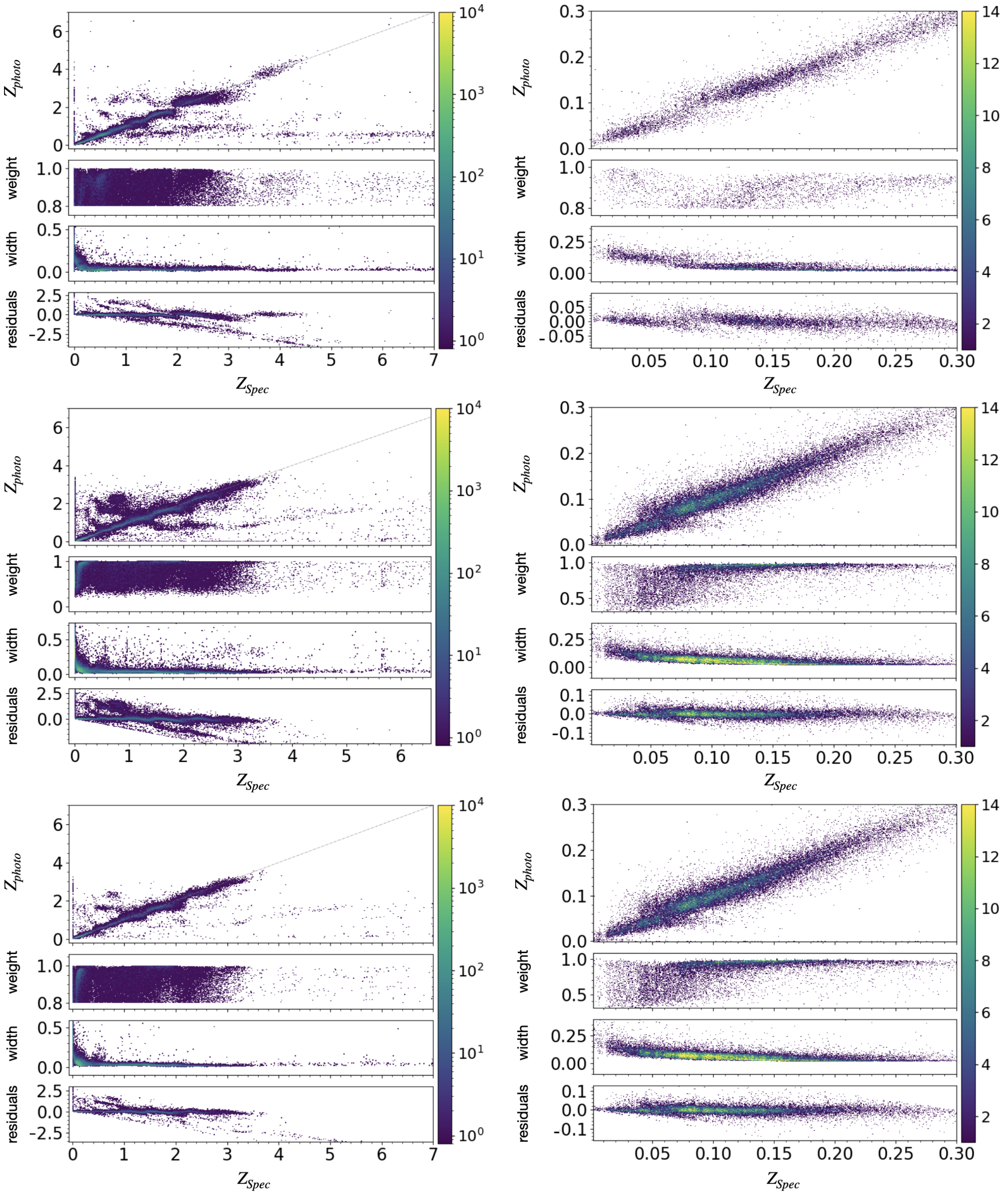

Figure 2.10: Photo-z performance of different MDN implementations. Top panel: Retaining only objects with $weight_{max} > 0.8$ membership probability to a MDN branch. Middle panel: Including $u-$band `PSF` and `model` magnitudes. Bottom panel: $u-$band magnitudes and MDN branch $weight_{max} > 0.8$. Left column: All types of objects in the spectroscopic data set. Right column: Only galaxies in $z_s < 0.3$. from the spectroscopic data set.

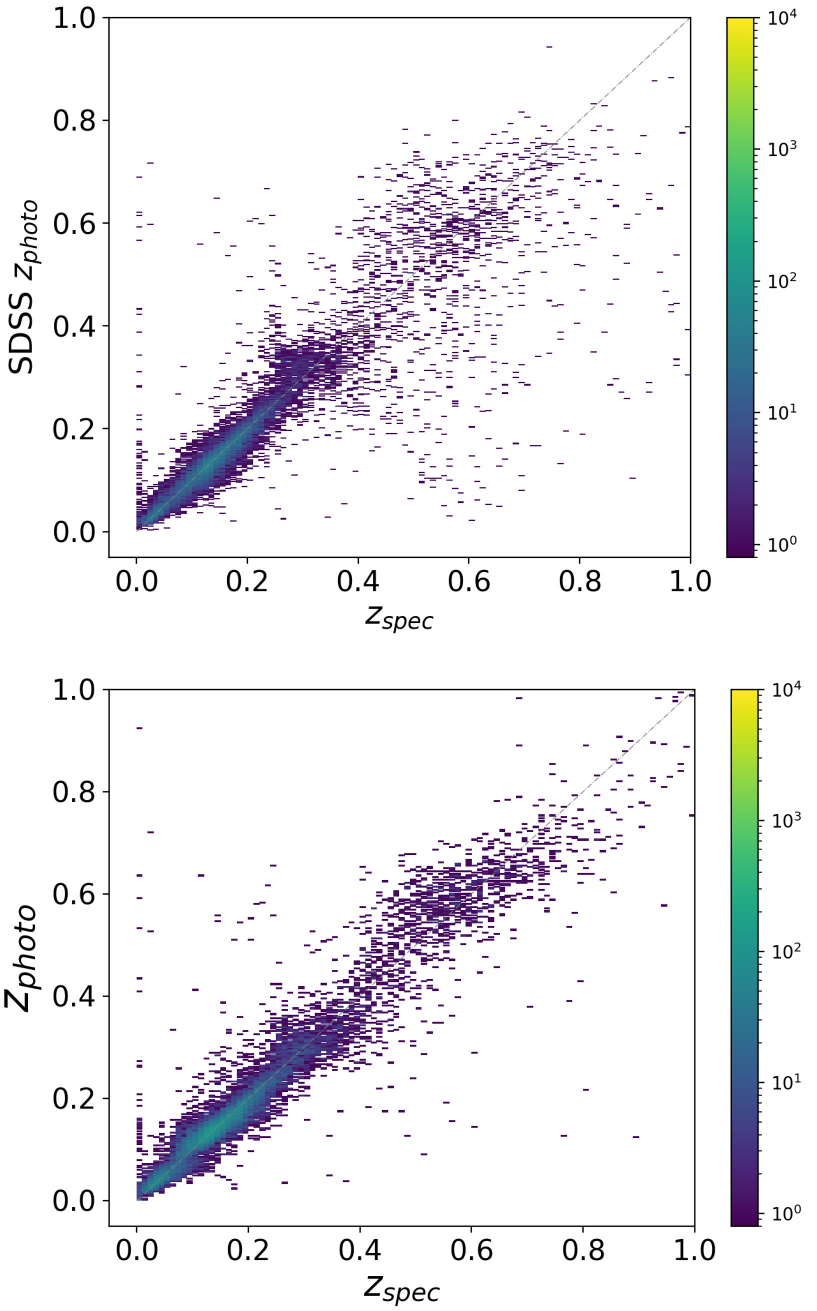

Figure 2.11: Top panel: SDSS spectroscopic redshift vs. SDSS photometric redshift. Bottom panel: spectroscopic redshift vs. photometric redshift (this work). Colour bars indicate the number of sources in the diagrams. The selection of sources is made by retaining objects with $weight_{max} > 0.8$ membership probability to a MDN branch.

3

Dust in supernovae with neural networks

> *"The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform."*
>
> – Ada Lovelace[a]

---

[a] The world's first computer programmer

"Inferring properties of dust in supernovae with neural networks"

Submitted in *Astronomy & Astrophysics* - under review

Authors: Zoe Ansari, Christa Gall, Roger Wesson, Oswin Krause

3.1 Abstract

Determining the properties of either pre-existing or newly formed dust in and around supernovae from observations remains challenging. This may be due to either incomplete coverage of data in wavelength or time but also due to often inconspicuous signatures of dust in the observed data. Here we address this challenge using modern machine learning methods to determine the amount and temperature of dust as well as its composition from a large set of simulated data. We aim to quantify, if such methods are suitable to infer quantities and properties of dust from future observations of supernovae. We develop a neural network consisting of eight fully connected layers and an output layer with specified activation functions that allow us to predict the dust properties and quantities as well as their respective uncertainties for each single supernova of a large set of simulated supernova spectral energy distributions (SEDs). We produce the large set of supernova SEDs for a wide range of different supernova and dust properties using the advanced, fully three-dimensional radiative transfer code `MOCASSIN`. We then convolve each SED with the entire suite of *James Web Space Telescope* (JWST) bandpass filters to synthesise a photometric data set. We split this data set into three subsets which are used to train, validate and test the neural network. To find out how accurately the neural network can predict the dust properties from the simulated data, we consider three different scenarios. First, we assuming a uniform distance of ~ 0.43 Mpc for all simulated SEDs. Next we uniformly distribute all simulated SEDs within a volume of 0.43–65 Mpc and finally, we artificially add random noise corresponding to a photometric uncertainty of about 0.1 mag. Lastly, we conduct a feature importance analysis via SHapley Additive exPlanations (SHAP) to find the minimum set of JWST bandpass filters required to predict the dust properties with an accuracy that is either comparable or better than achieved with standard methods in the literature. We find that our neural network performs best for the scenario in which all SEDs are at the same distance and for a minimum subset of 7 JWST bandpass filters within a wavelength range 3–25 μm. This results in rather small root-mean-square errors (RMSE) of ~ 0.08 dex and ~ 42 K for the most reliable predicted dust masses and temperatures, respectively. For the scenario in which SEDs are distributed out to 65 Mpc and contain synthetic noise, the most reliable predicted dust masses and temperatures achieve an RMSE of ~ 0.12 dex and ~ 38 K, respectively. Thus, in all scenarios, both predicted dust quantities have smaller predicted uncertainties compared to those in the literature achieved with common SED fitting methods of actual observations of supernovae. Moreover, our neural network can well distinguish between the different dust species included in our work, reaching a classification accuracy of up to 95% for carbon and 99% for silicate dust. Although we trained, validated and tested our neural network entirely on simulated SEDs, our analysis shows that a suite of JWST bandpass filters containing NIRCam F070W, F140M, F356W, F480M and MIRI F560W, F770W, F1000W, F1130W, F1500W, F1800W filters are likely the most important filters needed to derive the quantities and determine the properties of dust formed in and around supernovae from future observations.

3.2 Introduction

The origin of dust in galaxies in the universe remains debated. Large amounts of dust are observed in galaxies and quasars in the early and local universe (e.g. Bertoldi *et al.*, 2003; Priddey *et al.*, 2003; Michałowski *et al.*, 2010a; Watson *et al.*, 2015; Wang *et al.*, 2008b; Michałowski *et al.*, 2010b; Marrone *et al.*, 2018), some of which require a rapid and efficient dust formation process (e.g. Dwek *et al.*, 2007; Gall *et al.*, 2011a; Gall *et al.*, 2011b; Finkelstein *et al.*, 2012). Hence, slow dust production by for instance asymptotic giant branch stars (e.g. Boyer *et al.*, 2012b) and subsequent grain growth in the interstellar medium of galaxies (e.g. Hirashita, 2012) is unlikely. Instead, there is growing evidence that core collapse supernovae (CCSNe), which mark the death of short-lived massive stars, are efficient dust producers likely responsible for the observed large amounts of dust in galaxies (Gall *et al.*, 2011b; Gall *et al.*, 2014; Ferrara *et al.*, 2016; Gall and Hjorth, 2018; De Looze *et al.*, 2020).

Dust masses inferred from observations of CCSNe range from less than about 10^{-4} $M_{\odot}$ in young CCSNe of a few hundred days old to about 0.1–1.0 $M_{\odot}$ in old CCSN remnants of a few 100 – 1000 years of age. From a handful of CCSNe that have observationally been monitored over several years, it is evident that the amount of dust gradually increases over about 25–30 years (Gall *et al.*, 2011b; Gall *et al.*, 2014; Wesson *et al.*, 2015; Bevan and Barlow, 2016; Gall and Hjorth, 2018). Observations of older SN remnants such as Cas A (Niculescu-Duvaz *et al.*, 2021), N49 (Otsuka *et al.*, 2010), Sgr A East (~ 0.02 $M_{\odot}$, and $\sim 10\,000$ years old, Lau *et al.*, 2015), G11.2$-$0.3 (~ 0.34 $M_{\odot}$), G21.5$-$0.9 (~ 0.29 $M_{\odot}$), and G29.7$-$0.3 (~ 0.51 $M_{\odot}$) (Chawner *et al.*, 2019) confirm that on average of about ~ 0.3 $M_{\odot}$ of CCSN produced dust is sustained over a period of about 3000 years. This is sufficient to account for the total dust mass observed in local as well as high redshift galaxies (Gall and Hjorth, 2018).

Inferring dust properties from observations is challenging. Typically, the amount of dust and its temperature is determined by fitting the thermal dust emission in the near- to far-infrared (far-IR) wavelength range with dust models at different levels of complexity (e.g. Rho *et al.*, 2009a; Gall *et al.*, 2011b; Wesson *et al.*, 2015; Matsuura *et al.*, 2015; Matsuura *et al.*, 2019; Chen *et al.*, 2021). However, the most common dust species are rather featureless in this wavelength range with silicates having the most prominent emission feature at around 10–12 micron (Draine and Lee, 1984; Henning, 2010), which also could appear featureless for cold dust and/or dust with large grains. Due to limited computational power or insufficient data, the manifold of dust model parameters can often neither be fully explored nor constrained. This leads to dust mass estimates that may vary over an order of magnitude (Gall and Hjorth, 2018).

Warm and cold dust ($\lesssim 500\,\mathrm{K}$) in nearby SNe and SN-remnants has been detected in the mid- to far-IR wavelength range with telescopes such as WISE, SOFIA, ALMA or the *Herschel* mission (2009–2013) (e.g., Gomez *et al.*, 2012a; Indebetouw *et al.*, 2014b; De Looze *et al.*, 2019; Gall *et al.*, 2011b; Gall and Hjorth, 2018, and references therein), and notably the *Spitzer Space Telescope*, which observed during its cold (2003–2009) and warm phase (2009–2020) about 380 CCSNe out of about 1100 SNe in total (see for a summary Szalai *et al.*, 2019). The next telescope in line with the right sensitivity to observe warm and cold dust that either is newly formed or heated and to possibly constrain some dust species will be the *James Web Space Telescope* (JWST, Gardner *et al.*, 2006). With instruments onboard such as the Near-Infrared Camera and Spectrograph (NIRCam, NIRSpec), the Near-Infrared Imager and Slitless Spectrograph (NIRISS), and the Mid-Infrared Instrument (MIRI) imaging as well as spectroscopic observations of CCSNe in the wavelength range 0.6 – 30 μm will be possible.

In this paper, we investigate whether modern machine learning algorithms can be used to determine the properties of existing as well as newly formed dust grains in and around CCSNe. We train a neural network to predict various dust properties from a simulated set of spectral energy distributions (SEDs) of core collapse supernovae with different dust properties. The SEDs are produced using the fully three-dimensional photoionisation and dust radiative transfer code `MOCASSIN` [1] (Ercolano *et al.*, 2003a; Ercolano *et al.*, 2005) exploring a large parameter space of dust and SN properties. Assuming that the SNe are distributed within maximally 65 Mpc, we then convolve the SEDs with the suite of available JWST NIRCam (0.6–5.0 μm) and MIRI (5.0–30 μm) bandpass filters to synthesise a photometric data set. The use of simulated data is essential for this work since unfortunately the presently existing wealth of observational data of dust in and around SNe is insufficient.

The neural network is optimised to predict the total dust mass, dust temperature, and dust species. The data input to the neural network include the entire photometric data set, which consists of 293 236 SEDs and the redshift for each SED. To obtain a practical method, we perform a feature selection method to find the minimum number of JWST filters to estimate the dust properties. Furthermore, we train the neural network to obtain an estimate on the uncertainties of the predicted quantities (i.e. dust mass, temperature and species). We then identify the most reliable predictions using self-defined and common performance evaluation metrics which also provide information about the overall performance of the neural network.

---

[1] https://mocassin.nebulousresearch.org

In Section 3.3 we describe the simulated data set which sets the basis of our analysis and which we use to train our machine learning algorithm, which is described in Section 3.4. In Section 3.5 we describe the metrics that we employ to evaluate the performance of the neural network, and discuss possible caveats in Section 3.6. We present our results in Section 3.7 and discuss the implications of our results on future observations and the SN dust community in Section 3.8. We conclude in Section 3.9.

Throughout the paper we assume a ΛCDM model with $H_0 = 70$ (km/s) Mpc, and $\Omega_0 = 0.3$ (Abbott *et al.*, 2017) that we applied on our simulated data set whenever needed, via a built-in library from `astropy`[2].

3.3 Simulated data

Here, we describe the simulated data set, which consists of simulated SN SEDs from which we synthesise a photometric data set using the entire suite of JWST NIRCam and MIRI bandpass filters. We describe how we deal with either exceptionally faint or bright sources with respect to the JWST detection / sensitivity limits. Furthermore, we define three different scenarios, in each of which we derive a different data set from the simulated data set to train the neural network and test its performance for predicting the supernova dust quantities and properties.

3.3.1 `MOCASSIN`

`MOCASSIN` (Monte Carlo Simulations of Ionised Nebulae) is a fully three-dimensional radiative transfer code that propagates radiation packets using a Monte Carlo technique (Ercolano *et al.*, 2003a; Ercolano *et al.*, 2005). Arbitrary distributions of material can be represented within a Cartesian grid. The material can consist of gas, dust, or both. In each grid cell, the thermal equilibrium and ionisation balance equations are solved to determine the physical conditions. For dusty models, `MOCASSIN` uses standard Mie scattering theory to calculate the effective absorption and scattering efficiencies for a grain of radius a at wavelength λ, from the optical constants of the material. Any type of grain size distribution and mixture of materials may be specified.

---

[2] https://docs.astropy.org/en/stable/api/astropy.cosmology.FlatLambdaCDM.html#astropy.cosmology.FlatLambdaCDM

The material is illuminated by a radiation source or sources, which can be discrete point sources, or a diffuse source present within each grid cell. The spectral energy distribution of the illuminating source can be a simple blackbody or an arbitrary spectral shape such as a stellar atmosphere model. The radiation field is described by a composition of a discrete number of monochromatic packets of energy (Abbott and Lucy, 1985) for all sources. At each location, the Monte Carlo estimator (Lucy, 1999) derives the mean intensity of the radiation field. The contribution of each energy packet to the radiation field at each location is defined by its path through the grid.

To synthesise different SEDs of SNe with dusty shells (hereafter SN model SEDs) using `MOCASSIN` we define a set of parameters for the underlying radiation source (the SN), the dust itself and its location. Specifically, in our simulation, our chosen radiation source is a central blackbody, which is defined by a temperature (T_{BB}) and a luminosity (L_{BB}). The range of the two parameters follows typical measurements of SN photospheres up to a few hundred days past explosion. The range of radii used in our models covers both the expected radii of supernova ejecta up to $\sim$1000 days after explosion, as well as larger radii at which pre-existing dust flash-heated by a supernova explosion could give rise to infrared emission. For the dust, we consider two prominent grain species, which are amorphous carbon and astronomical silicates with optical constants taken from Zubko *et al.* (1996) and Draine and Lee (1984), respectively. For our simulations, we consider that all the dust consists of either 100% carbon, 100% silicates, or is a 50:50 mixture of the two dust species. The range of initial dust masses is limited to $10^{-5} - 10^{-1}$ $\mathrm{M}_{\odot}$. The upper dust mass limit is partly motivated by the long run-time of simulations of SN model SEDs have, if a lot of dust is present. Another reason is that the mean dust mass for SNe and SN remnants is 0.4 ± 0.07 $\mathrm{M}_{\odot}$ (Gall and Hjorth, 2018), but the dust temperatures for large dust masses ($\gtrsim 10^{-2}$ $\mathrm{M}_{\odot}$) in some SNe is $\lesssim 50$ K (Gall *et al.*, 2014). Even with JWST such cold dust will not be easily detected. Furthermore, we consider only single grain sizes ranging between 0.005 – 5 μm. Typically, such grain sizes are present in for example the Milky Way (Mathis *et al.*, 1977, e.g.) and observed in some SNe (e.g. Gall *et al.*, 2014; Wesson *et al.*, 2015; Bevan *et al.*, 2020). In total, our SN model SEDs are composed of seven parameters, for which we define either a set of distinct choices or a range of values (some are described above). A summary of the entire parameter space is presented in Table 3.1. To finally create our data set, each SN model SED is synthesised from a set of parameters that is stochastically generated from this parameter space. This method ensures that the entire parameter space is uniformly exploited.

Table 3.1: Input parameters for the `MOCASSIN` models.

Parameter	Value / Switch	Unit	Description
T_{BB}	$(4 – 14) \times 10^3$	K	BB-Temperature
L_{BB}	$(1 – 100) \times 10^4$	$L_\odot$	BB-Luminosity
R_{out}	$(1 – 20) \times 10^{16}$	cm	Outer radius of dusty shell
R_{in}	$0.8 \times R_{out}$	cm	Inner radius of dusty shell
M_d	$10^{-5} – 10^{-1}$	$M_\odot$	aDust mass
a	0.005 – 5	μm	aGrain size
κ, X	silicates, carbon 50:50 mixture		Grain species

a Values are chosen from a logarithmic distribution.

In this work, we use `MOCASSIN` version 2.02.73 to synthesise 293 236 model SEDs. We construct a cubical Cartesian grid with 11 cells on each side of the 3D grid to model the dusty shells, which are defined by an inner and outer radius of the shell, R_{in} and R_{out}, respectively. We model one-eighth of the grid cube (shell) with the illuminating source in one corner. Assuming spherical symmetry, this cube segment is then scaled to the full cube for an effective resolution of 21^3 cells (Ercolano *et al.*, 2003b). We use 10^6 energy packets in most of our simulations. This relatively low number (~750 energy packets per grid cell) ensures that the `MOCASSIN` models run very quickly. However, at wavelengths where only a few photons are emitted, the SEDs are affected by small number of statistics and hence dominated by noise. Therefore, for `MOCASSIN` models with dust masses lower than 10^{-4} $M_\odot$ in which few photons are reprocessed to longer wavelengths, we used 10 times as many energy packets to reduce the statistical noise in the SEDs at longer wavelengths (e.g. 5-30 μm).

For efficiency reasons, we set a maximum run-time of two minutes for each model. For most regions of the investigated parameter space, the `MOCASSIN` models have a run-time of a few seconds, but models with both a small shell radius ($\lesssim 4\times10^{16}$cm) and a high dust mass ($\gtrsim 10^{-2}M_\odot$) have very high optical depths and thus, time out. This results in a slightly non-uniform filling of the entire parameter space. Furthermore, any dust grains in a simulation which reach the sublimation temperature of its species (1400 K for silicate dust, 2200 K for carbon dust) are considered to have evaporated and are not included when calculating the SED. The final dust mass is then either lower than the input dust mass or dust may even be no longer existing. Consequently, for mixed-chemistry models, the composition is altered from 50:50 to a higher carbon fraction due to the higher sublimation temperature of carbon dust. `MOCASSIN` does not directly provide the

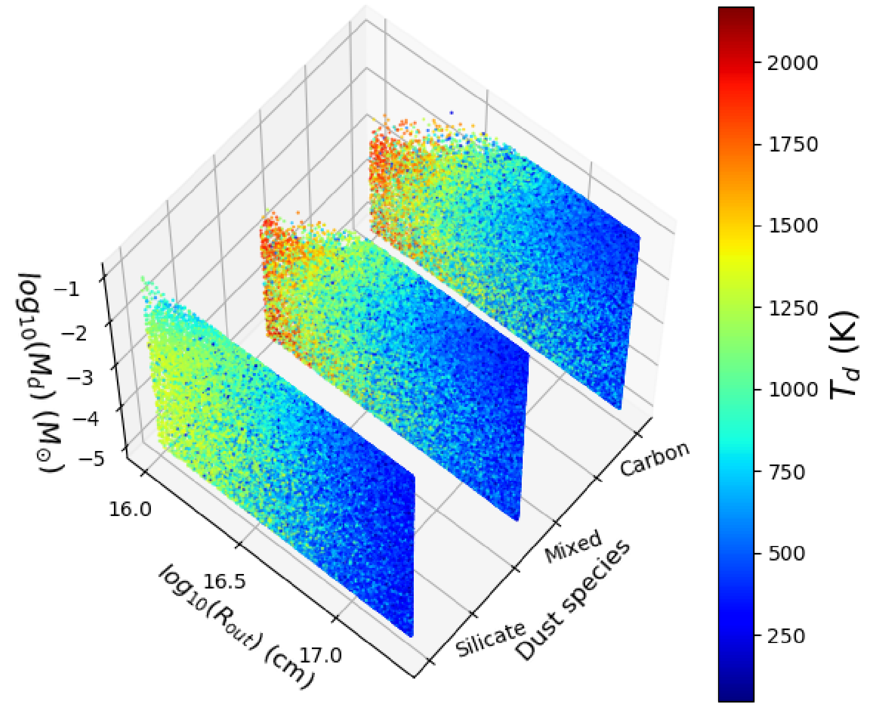

Figure 3.1: Coverage of SN model SEDs in M_{dust}, R_{out}, and dust species parameters space. The colour bar represents T_{dust} of the SN model SEDs .

final dust mass and composition if dust evaporation occurs; we calculate both of these from the output grid files by summing the dust masses in cells where the temperature is below the dust sublimation temperature. Some dust evaporation occurs in about 5 percent of our models. Figure 3.1 shows the final distribution of the SN model SEDs in the dust mass, temperature and species `MOCASSIN` output-parameter space.

Synthetic photometry of optical and mid-IR JWST bandpass filters

The JWST is equipped with two imaging cameras, NIRCam and MIRI. The two cameras have in total six narrow and 31 broad bandpass filters available that cover the wavelength ranges $0.6-5\ \mu m$ (NIRCam) and $5-30\ \mu m$ (MIRI). As a next step in preparing the data set for our neural network, we convolve the SN model SEDs with both NIRCam and MIRI bandpass filters (hereafter filters) in order to synthesize a photometric data set. To do so, we use the python program `Pyphot`[3]. This program has a built-in library of transmission

---

[3] https://mfouesneau.github.io/docs/pyphot/libcontent.html

curves of different filters. Since `Pyphot` also allows customised transmission curves, we import transmission curves for NIRCam and MIRI filters from the Spanish virtual observatory[4]. For each NIRCam and MIRI filter, we first calculate the integrated flux in units of Jansky via `Pyphot`, which then are converted to AB magnitudes as

$$M_{\mathrm{AB}} = 2.5 \times (23 - \log_{10}(F_\nu(\lambda_{\mathrm{obs}})) - 48.6, \tag{3.1}$$

following the definition of Hogg *et al.* (2002).

JWST detection limits

For the final step, we consider that our synthetic photometric data set contains magnitudes in some filters that would either be too bright or too faint to be detected with JWST. In order to filter out data with magnitudes that practically could not be observed (hereafter missing values), we adopt the pre-calculated detection limits (Greene *et al.*, 2017) that have been derived using the JWST exposure time calculator (ETC, Pontoppidan *et al.*, 2016) for a signal-to-noise ratio (S/N) of 10 and exposure times of 21.4 s and 10 000 s for the saturation and sensitivity limits, respectively. A visualisation of these limits for all NIRCam and MIRI filters is shown in the appendix in Figures 3.10 and 3.11, respectively.

3.3.2 Three scenarios

Typically, CCSNe occur in different types of galaxies at different distances. Consequently, distant CCSNe appear fainter than the same nearby CCSNe because their brightness decreases with distance as

$$F_\nu(\lambda_{\mathrm{obs}}) = \frac{L_\nu(\lambda_{\mathrm{emit}})(1+z)}{4\pi D_L^2} \tag{3.2}$$

with $F_\nu(\lambda_{\mathrm{obs}})$ the observed flux as a function of the observed wavelength, λ_{obs}, in units of Jy; $L_\nu(\lambda_{\mathrm{emit}})$ the emitted luminosity at the emitted wavelength, λ_{emit}; D_L the luminosity distance and z the redshift. The observed wavelength is given by $\lambda_{\mathrm{obs}} = \lambda_{\mathrm{emit}}(1+z)$.

---

[4] http://svo2.cab.inta-csic.es/svo/theory/fps3/

This implies that the SEDs of CCSNe are redshifted and a well-defined bandpass filter will sample the light from a bluer wavelength region of the intrinsic CCSN spectrum compared to the restframe wavelength range of the bandpass filter. In extreme cases (e.g. at high redshift) such an effect may cause a non-negligible degeneracy between dust properties and redshift. In what follows, we define three individual scenarios that are used to test if some quantities and properties of dust formed in and around CCSNe, such as the dust mass, $\mathrm{M}_{\mathrm{dust}}$, dust temperature, $\mathrm{T}_{\mathrm{dust}}$ and the dust species can be determined with neural networks.

For the first scenario, S1, we simply assume that all CCSNe are at the same, low redshift of $z = 0.0001$, which corresponds to a distance of ≈ 0.43 Mpc. For comparison, the distance of SN 1987A, the closest observed extragalactic CCSN is $\approx 0.49 \pm 0.0009$ (statistical) $\pm$ 0.0054 (systematic) Mpc (Pietrzyński *et al.*, 2019) and the next closest CCSN, SN 1885A (Fesen *et al.*, 1989), is ~ 0.765 Mpc away.

Placing all SN model SEDs at the same such short distance has the advantage that the observed model magnitudes are nearly identical to the intrinsic magnitudes of SN model SEDs and thus, free of any possible degeneracy between dust properties and distance. Hence, we expect this scenario to be an ideal test case for the neural network. Moreover, from this scenario we can identify the smallest amount of dust detectable with JWST (see Section 3.3.1). For simplicity reasons, here for S1, we only consider the upper sensitivity limit of the JWST filters but do not apply the lower saturation limits.

For the second scenario, S2, we assume that all our simulated CCSNe are uniformly distributed within the redshift range 0.0001–0.015, which corresponds to a distance range of $\sim$0.43–65 Mpc. The decrease in brightness and shift in wavelength with increasing distance, together with the sensitivity and saturation limits of the JWST filters (see Figures 3.10 and 3.11) place a limit on the distance out to which dust in CCSNe may be observed. Therefore, we choose $z = 0.015$ (i.e. $\sim$ 65 Mpc) as an upper limit. This limit is based on the SN model SEDs, for which the thermal dust emission of $10^{-5}\ \mathrm{M}_{\odot}$ carbon dust at a temperature of $\sim 2000\,\mathrm{K}$ remains detectable (see Section 3.3.1) in at minimum 10 out of 28 NIRCam filters.

The data sets of scenarios S1 and S2 solely consist of synthesised magnitudes of all available JWST filters without uncertainties. Therefore, as our third test scenario, S3, we use the data set of S2 and add synthetic photometric noise. We assume that each synthesised magnitude is ‘observed’ at S/N = 10, which translates into an uncertainty of 0.1 mag. This assumption is in line with what has been used to derive the detection limits (see Section 3.3.1). Hence, to create S3, we add randomly synthesised noise to the data of S2 as $m_{i,\mathrm{S3}} = m_{i,\mathrm{S2}} + \mathscr{N}(0, 0.1)$, with m_i the magnitude of each JWST filters, and $\mathscr{N}(0, 0.1)$ is a randomly generated number taken from a Gaussian distribution with zero mean and $\sigma = 0.1$.

3.4 Neural networks

Our analysis is based on training a deep neural network using simulated data (see Section 3.3). The goal is to predict three dust quantities and properties, $\mathrm{T_{dust}}$, $\mathrm{M_{dust}}$ and dust species, together with a prediction of their respective uncertainties. To conform with machine learning nomenclature, we refer to the set of photometric data that is synthesised from each SN model SED using JWST filters, along with the redshift of the SN model SED, as the *input features*. We also refer to each SN model SED that corresponds to each set of synthesised magnitudes, as a *data point*, since it is defined as a point in the input features’ space.

In the following sub-sections, we describe the artificial neural networks and the corresponding hyperparameters. We also describe the specific type of neural network that we use and its corresponding optimal set of hyperparameters as well as a pre-processing method to treat the missing values in our data set. Furthermore, we describe the training process of our neural network, in which we define target values for three dust quantities and properties. Thereafter, we explain an iterative feature selection procedure which we use to find the minimum set of the most important JWST filters with which the dust quantities and properties can still be predicted with an acceptable accuracy.

3.4.1 Artificial neural networks

An artificial neural network or in short, neural network, is a set of algorithms that is used to recognize relationships in a data set, and to find patterns. The structure of a neural network is inspired by biological neurons, and thus it mimics the methodology that biological neurons use to send signals to one another. Neural networks consist of one or more layers, known as hidden layers, between an input and an output layer. Each layer contains a set of neurons. The process of training a neural network consists of transferring information from the input layer to the output layer via a set of connections. Each connection is defined between each neuron in one layer to each neuron in the next layer.

There are different methods to connect neurons and to transfer information between them. In the classic framework, each neuron of a given layer is connected to all neurons in the next layer. Layers that follow this pattern are called fully connected layers. Another method for connecting neurons consists of convolutional layers, in which each neuron from a layer is only connected to a well defined set of neurons from the next layer. A neural network can be built using either one or a combination of different layers and different patterns. To transfer the information, each layer applies an activation function to a set of weights associated with a set of neurons in the layer.

The output vector of each layer is defined as follows:

$$\mathbf{a}_i^l = \mathscr{H}^l \left(\sum_{j=1}^{p} \mathrm{W}_{j,i}^l \cdot \mathbf{a}_j^{l-1} + \mathbf{b}^{l-1} \right) \tag{3.3}$$

where $\mathbf{a}^{l-1}$ is the input vector to layer l, $\mathrm{W}_{i,j}^l$ is a matrix that contains a set of weights from neuron j in layer $l-1$ to neuron i in layer l, p is the number of neurons in the layer $l-1$, $\mathbf{b}^{l-1}$ is a vector of constant values assigned to neurons of layer l, known as thresholds, and $\mathscr{H}^l$ is an activation function for layer l. For the input layer (i.e. l=0) $\mathbf{a}_i^l = \mathbf{x}$, where $\mathbf{x}$ is the input feature vector for the neural network.

The weights and the thresholds of neural networks are the model parameters that a neural network aims to optimise by improving its performance of estimating the target values. In a forward-propagation process of a neural network training, the prediction error is first calculated using random weights. The prediction errors are quantified by a ‘loss function’. In a subsequent back-propagation process (e.g. Rumelhart *et al.*, 1986), the weights are adjusted with the aim of minimising the loss. As the name suggests, the forward-propagation method iterates from the input via the hidden to the output layer, while the back-propagation is converse. This combination of forward- and back-propagation takes place within one epoch of training (hereafter epoch). Typically, several epochs are required to minimise the loss function and to improve the performance of the neural network.

Since the loss function can be non-convex, and finding a global minimum of a general non-convex function is NP-hard (Murty and Kabadi, 1987), a neural network can be considered optimised when the loss function is converged to a ‘good’ local minimum. To do so, minimisation algorithms, such as the classical gradient descent, are employed. The basic principle of such algorithms is to calculate the gradient of the loss function and step by step move in the direction as specified by the gradient, with the step size termed as the learning rate.

Choosing the right learning rate is important as for a high learning rate the calculated loss with updated model parameters can jump over the local minimum, therefore can not converge to it. On the other hand, using a low learning rate, the algorithm takes a long time to reach the local minimum of the loss function.

The batch gradient descent is a gradient descent optimisation method in which the neural network updates the weights only once per epoch for the entire training data set. Although this process is a fast approach for finding the local minimum of the loss function, the memory requirement for such computational task is large. A remedy to this is to employ a mini-batch gradient descent, which allows the neural network in each epoch to update the weights for a sub-sample of the data set separately. This sub-sample is called mini-batch, and the size of it is defined by the size of the mini-batch.

The classical gradient descent uses a fixed learning rate for the entire process. Since this is not optimal, other types of optimisation algorithms that can adjust the learning rate, such as Adaptive Moment Estimation (ADAM, Kingma and Ba, 2014) may be used instead.

Neural network parameters, such as the number of either hidden layers, neurons or epochs, the learning rate, the optimiser, the activation function for each layer and the size of the mini-batch, are referred to as hyperparameters. The hyperparameters affect the efficiency and performance of the neural network and, like the model parameters, need to be optimised to reach the best possible network performance. While the process of training a neural network adjusts the model parameters, usually the hyperparameters must be manually fine-tuned for each science case and data set in question (LeCun *et al.*, 1998; Bengio, 2012; You *et al.*, 2017; Rijn and Hutter, 2018; Weerts *et al.*, 2020).

3.4.2 Our neural network

We design a neural network to estimate a set of target values, along with their uncertainties. Our neural network aims to approximate a distribution for each target value with a given input feature, $\mathbf{x}$, of each data point and three target values correspond to three dust properties, $y^{\rm sim}_{M_{\rm dust}}$, $y^{\rm sim}_{T_{\rm dust}}$, and $y^{\rm sim}_{\kappa}$. The neural network implements this approximation by maximizing the log-likelihood of the target values under the assumption that the deviations follow a normal distribution, by approximating the mean (m_k) and standard deviation ($\sigma_k^{\rm pred}$), which is the expected squared difference between the $y^{\rm pred}$ and $y^{\rm sim}$, as follows:

$$\log \mathscr{L}(y^{\rm sim}, \mathbf{x}) = \sum_{n=1}^{N} \sum_{k=1}^{K} \log \left(\mathscr{N}(y^{\rm sim}_{k,n} | m_k(\mathbf{x}_n), \sigma_k^{\rm pred}(\mathbf{x}_n)) \right) \tag{3.4}$$

where N is the number of data points in the data set, while K represents the number of target values. Therefore, each target value $y_k^{\rm sim}$ is estimated by a mean m_k (hereafter $y_k^{\rm pred}$), and a standard deviation $\sigma_k^{\rm pred}$, that represents the estimated uncertainty of $y_k^{\rm pred}$.

3.4.3 Hyperparameter tuning

To find the optimal set of hyperparameters for our neural network, we first explore combinations of 3–12 convolutional and fully connected layers. Each layer can have of either 4, 16, 32, 64, 128, 256, 512 neurons. We use the standard Rectified Linear Units (ReLU, Maas *et al.*, 2013) and Parametric Rectified Linear Units (PReLU, He *et al.*, 2015) as non-linear activation functions between the input and the hidden layers. For the output layer, we use a linear activation function to predict the mean of the target values and an exponential linear unit (ELU, Clevert *et al.*, 2015) as activation function to predict the standard deviations of the target values. Using ELU as the activation function ensures that the estimated standard deviations are positive.

We use six different learning rates of 10^{-6}, 5×10^{-6}, 10^{-5}, 5×10^{-5}, 10^{-4}, and 10^{-3} for the ADAM optimiser (Kingma and Ba, 2014) to search for the local minimum of the loss function with mini-batch sizes of 32 and 64 data points. By comparing the validation and training loss of the neural network with different sets of hyperparameters, we find that the optimal set of hyperparameters consists of eight fully connected layers (one input and seven hidden layers) with 512, 256, 128, 64, 32, 16, 8, 4 neurons in the first to the 8th layer, respectively. Furthermore, ReLU activation functions are best used between the layers together with a learning rate of 10^{-5} for the ADAM optimiser with mini-batch size of 64 data points. The number of epochs is chosen to be 2000 in S1 and 1500 for S2 and S3, in which the training and validation loss are converged.

3.4.4 Missing data

Considering the sensitivity and saturation limits (see Section 3.3.2) for both MIRI and NIRCam filters, some SN model SEDs are not detectable in all filters over the entire wavelength range. For instance, particularly bright or faint SEDs (or parts of the SEDs) result in magnitudes that either exceed the sensitivity limit or remain below the saturation limits of some filters. In reality, such cases would not lead to detections (magnitude measurements) and hence, may be considered as ‘missing values’. Here, for each filter, we replace the synthesised magnitudes that fall outside the saturation and sensitivity limits with the magnitude of the sensitivity and saturation limits, respectively. This approach is inspired by the forced photometry measurement that is commonly used to study transients, for example for Pan-STAARS1[5]. In this method, when a source is detected in a filter at a specific location in the sky, photometric values are forced to be extracted in other filters. These forced photometric values are either the actual magnitudes of the source, or the limit magnitudes of the filters (i.e. the sensitivity or saturation limits).

3.4.5 Neural network training preparation

To train our neural network with the set of hyperparameters that are defined in Section 3.4.3, we create a ‘training-validation-test’ split from each of the data sets that are described in Section 3.3.2. Particularly, out of a total of 293 236 data points, we use 70% (193 536) as training-, 15% (49 850) as validation- and the other 15% as test-set.

---

[5] https://outerspace.stsci.edu/display/PANSTARRS/PS1+Forced+photometry+of+sources

We normalise $y^{\text{sim}}_{M_{\text{dust}}}$ and $y^{\text{sim}}_{T_{\text{dust}}}$ of all SN model SEDs as

$$g(y^{\text{sim}}) = \frac{y^{\text{sim}}}{y^{\text{sim, max}}}$$

with $y^{\text{sim, max}}_{M_{\text{dust}}} = 0.1\ M_{\odot}$, and $y^{\text{sim, max}}_{T_{\text{dust}}} = 2200\,\text{K}$. Moreover, we define a conditional function in which we arbitrarily assign each dust species (e.g. carbon, and silicate) a target value:

$$g_{\kappa}(y^{\text{sim}}) = \begin{cases} 1, & \text{if } \ y^{\text{sim}} = \text{silicate} \\ 0.75, & \text{if } \ y^{\text{sim}} = \text{a mix of carbon and silicate} \\ 0.5, & \text{if } \ y^{\text{sim}} = \text{carbon } \ . \end{cases}$$

We find that inferring the dust properties from SN model SEDs that contain no dust or only very small amounts of dust at cooler temperatures ($\text{M}_{\text{dust}} < 5 \times 10^{-5} M_{\odot}$ and $\text{T}_{\text{dust}} < 800\,\text{K}$) using neural networks is challenging (see Section 3.6 for further explanation). Therefore, to let the neural network differentiate between these SN model SEDs and SN model SEDs that contain recognisable dust, we define a dedicated target value for this group of ‘no-dust’ data points as $y^{\text{sim}}_{\text{no-dust}}$ = -0.5.

3.4.6 Feature selection

The SHapley Additive exPlanations (SHAP; Lundberg and Lee, 2017) is a framework that uses an additive feature attribution method to evaluate the importance of a certain input feature on the prediction of a neural network. In this framework, the Shapley values (Shapley, 2016) are calculated for each input feature based on cooperative game theory (Nash, 1953). In this theory, to calculate the contribution of each input feature to a model’s output, the average marginal effect of feature i is measured for all possible coalitions, which represents the effect of feature i on the model’s output. In an additive feature attribution method, for an input features vector $\mathbf{x}$, for a model f, a simplified local input features vector $\mathbf{x}'$, is defined for an explanatory model $\mathscr{F}$. The simplified local input features is a discrete binary vector, $\mathbf{x}' \in \{0,1\}^d$ (where d is the number of the input features), which represents that either features are included or excluded. The explanatory model $\mathscr{F}$ is defined as

$$\mathscr{F}(x') = \phi_0 + \sum_{i=1}^{d} \phi_i \mathbf{x}'_i \ ,$$

where ϕ_0 is the base value of the model in the absence of any information, that is defined by the average of the model's output, and ϕ_i is the explained effect of feature i, known as the attribution of feature i. The ϕ_i shows how much feature i changes the output of the model. The second term of the model $\mathscr{F}$ is the average over marginal contribution of each feature, over all possible coalitions. The absolute value of $\mathscr{F}(\mathbf{x}'_{n,i}) - \mathscr{F}(\mathbf{x}'_{n,-i})$ indicates the importance of the feature i, where $\mathbf{x}'_{n,-i}$ represents that feature i is not included in the input feature vector $\mathbf{x}'_{n,i}$. Therefore, the Shapley values are defined as:

$$\phi_{n,i} = \sum \frac{|S|!(d-|S|-1)!}{d!} (\mathscr{F}(\mathbf{x}'_{n,i}) - \mathscr{F}(\mathbf{x}'_{n,-i})) \tag{3.5}$$

when the summation is over all feature subsets $S \subseteq d$.

To calculate the Shapley values, all coalition values for all possible feature permutations must be sampled. Since the relation between the number of features and the number of possible feature permutations is exponential, for a large set of features the number of calculations in $\mathscr{F}$ is immense, and practically not feasible to implement. Therefore, the SHAP framework uses a fast approximation, Deep Learning Important FeaTures (DeepLIFT, Shrikumar *et al.*, 2016; Shrikumar *et al.*, 2017), in which a linear approximation of Taylor series is used to approximate

$$\mathscr{F}(\mathbf{x}'_{n,i}) - \mathscr{F}(\mathbf{x}'_{n,-i}) \ ,$$

in which the expectation values, $E[\mathbf{x}']$, are calculated for all features and are used as referenced values in the input features vector, when the feature is omitted during the calculations.

Since the variance of the expectation values for N data points is roughly $1/\sqrt{N}$, using approximately 1000 data points gives an acceptable estimation for expectation values[6]. Therefore, in this work, for each of our three test scenarios S1, S2 and S3, we use a sample

---

[6] https://shap-lrjball.readthedocs.io/en/latest/generated/shap.DeepExplainer.html

of 5 000 data points that we randomly choose from the training data set to approximate the expectation values for all features (i.e. $E[\mathbf{x}'_i]$ for $\forall i; i \in \mathbf{x}'$). Thereafter, we compute the Shapley values for 1 000 randomly selected data points from the validation data set (see Section 3.10.2 for the details of the computational cost). We select the sub-samples from the validation and training data sets with a random seed that we change for each step in the feature selection process. Therefore, we calculate the importance of each feature (i.e. filter) with index i via

$$\phi_i = \frac{1}{N} \sum_{n=1}^{N} |\phi_{n,i}| \quad , \tag{3.6}$$

where N=1 000. In each step, we remove the three filters that achieve the three lowest absolute Shapley values. Subsequently, in the next step, we train the neural network using the reduced filters set as the input features of the entire training data set and repeat the procedure. Considering that in each step, we remove the three least important filters, we perform the process for a total of 11 steps. Therefore, we are left with four filters out of 37 filters at the end of the process.

3.5 Evaluation

In this section we describe the chosen evaluation metrics to evaluate the performance of our trained neural network. We address how we interpret the resulting predictions for the dust species and how we treat the no-dust models in the performance evaluation. Furthermore, we define criteria to estimate the reliability of the predictions via the predicted standard deviations as the outputs of the neural network. Finally we describe the metrics for comparing the performance of the neural network in different steps of the feature selection process.

The performance evaluation of the predicted target values, $y^{\text{pred}}_{M_{\text{dust}}}$, $y^{\text{pred}}_{T_{\text{dust}}}$ and y^{pred}_{κ}, is applied on test data sets, and consists of three individual methods: root-mean-square error (RMSE), bias and 3σ outliers. For the dust temperature, the residual of data point n, is defined as $\Delta y_{T_{\text{dust}},n} = y^{\text{pred}}_{T_{\text{dust}},n} - y^{\text{sim}}_{T_{\text{dust}},n}$. For the dust mass, due to the logarithmic distribution of M_{dust} in the simulated data set we define the residual as $\Delta y_{M_{\text{dust}},n} =$

Table 3.2: Comparison of neural network performance for estimating M_{dust} and T_{dust} in different scenarios for 4 different cases. In case-1 and case-2, the training data set contains the preferred subset of JWST filters. In case-3 and case-4, the data set with the minimum subset of JWST filters is used to train our neural network. In case-1, and case-3 the evaluation metrics are applied on the entire test data set. In case-2, and case-4 the evaluation metrics are applied on the predictions of the test data set that have $\sigma_{\star}^{\mathrm{pred}}$.

		M_{dust} ($M_{\odot}$)			T_{dust} (K)		
Case	Scenario	bias (dex)	RMSE (dex)	3σ outliers (%)	bias	RMSE	3σ outliers (%)
1	S1a	-0.0084	0.1696	2.08	1.16	18.16	1.64
	S2b	-0.0786	0.3170	1.72	4.28	31.25	2.05
	S3c	-0.0135	0.3120	7.01	6.31	59.90	2.02
2	S1a	-0.0110	0.0541	12.80	1.61	14.14	1.45
	S2b	-0.0653	0.1056	15.00	4.79	18.85	2.13
	S3c	-0.0355	0.1136	16.22	4.60	30.12	1.56
3	S1d	0.0130	0.2075	8.19	1.39	56.52	1.92
	S2e	0.0537	0.3985	4.10	-0.62	37.92	2.01
	S3f	0.0325	0.5522	2.25	-2.69	78.55	2.34
4	S1d	0.0013	0.0847	8.78	2.77	42.65	1.41
	S2e	0.1018	0.1328	21.54	-1.95	20.80	1.72
	S3f	-0.0235	0.1257	15.72	-8.05	38.52	2.03

With the subset of JWST filters that are selected via the feature selection procedure as follows:

a NIRCam:$F070W$, $F115W$, $F140M$, $F150W$, $F210M$, $F300M$, $F335M$, $F360M$, $F430M$, $F444W$, $F460M$, $F466N$, and MIRI: $F560W$, $F770W$, $F1000W$, $F1130W$, $F1280W$, $F1500W$, $F1800W$, $F2100W$, $F2550W$

b NIRCam:$F070W$, $F115W$, $F140M$, $F150W$, $F182M$, $F187N$, $F200W$,$F250M$, $F277W$, $F300M$, $F322W2$, $F356W$, $F360M$,$F405N$ and MIRI: $F560W$, $F770W$, $F1000W$, $F1130W$, $F1280W$, $F1500W$, $F1800W$, $F2100W$

c NIRCam: $F070W$, $F115W$, $F140M$, $F356W$, $F460M$, $F480M$, and MIRI: $F560W$, $F770W$, $F1000W$, $F1130W$, $F1280W$, $F1500W$, $F1800W$

d NIRCam: $F460M$, and MIRI: $F560W$, $F770W$, $F1130W$, $F1280W$, $F1500W$, $F2100W$

e NIRCam: $F140M$, $F150W$, $F200W$, $F300M$, $F356W$, $F360M$, $F410M$, and MIRI: $F560W$, $F770W$, $F1000W$, $F1280W$, $F1500W$, $F1800W$

f NIRCam: $F070W$, $F140M$, $F356W$, $F480M$, and MIRI: $F560W$, $F770W$, $F1000W$, $F1130W$, $F1500W$, $F1800W$

Table 3.3: Comparison of neural network performance for classifying dust species, and the fraction of predictions of the test data set that have $\sigma_{\star}^{\text{pred}}$ to all the predictions from the test data set ($\sigma_{\star}^{\text{pred}}$), in different scenarios for four different cases. The definition of cases are the same as defined in Table 3.2

		classification accuracy rate (%)			
Case	Scenario	Carbon dust	Mixed dust	Silicate dust	$\sigma_{\star}^{\text{pred}}$ (%)
1	S1 a	86	95	100	-
	S2b	74	85	99	-
	S3c	72	75	98	-
2	S1 a	97	99	100	68
	S2 b	98	99	100	31
	S3c	97	94	100	12
3	S1 d	89	90	98	-
	S2e	73	79	98	-
	S3f	61	63	96	-
4	S1d	98	100	100	48
	S2e	94	95	99	34
	S3f	95	57	99	07

With the subset of JWST filters that are selected via the feature selection procedure as follows:

a-f: See the definitions in Table 3.2.

$log_{10}(y^{\mathrm{pred}}_{M_{\mathrm{dust}},n}/y^{\mathrm{sim}}_{M_{\mathrm{dust}},n})$. For both $\mathrm{M_{dust}}$ and $\mathrm{T_{dust}}$ the bias is defined as the mean of the residuals as

$$\frac{1}{N}\sum_{n=1}^{N}\Delta y_n \quad ,$$

and the RMSE is defined as

$$\sqrt{\frac{1}{N}\sum_{n=1}^{N}(\Delta y_n)^2},$$

where n represents each data point, and N is the number of data points in the test data set. Furthermore, for $\mathrm{M_{dust}}$ and $\mathrm{T_{dust}}$ we define the 3σ-outliers as the predictions with $|\Delta y_n| > 3 \times$ RMSE. Moreover, due to the numeric representation of all dust species (see Section 3.4.5) that are fed to the neural network, numeric target values are predicted. In order to interpret these numeric target values, we define each dust species as a 'class'. This way, we have the following classes: silicate, mixed, carbon and no-dust that we define by a conditional function as

$$\mathscr{G}_{\kappa}(y^{\mathrm{pred}}) = \begin{cases} \text{silicate,} & \text{if } y^{\mathrm{pred}} > 0.875 \\ \text{mixed,} & \text{if } 0.675 < y^{\mathrm{pred}} \leq 0.875 \\ \text{carbon,} & \text{if } 0.25 < y^{\mathrm{pred}} \leq 0.675 \\ \text{no-dust,} & \text{if } y^{\mathrm{pred}} \leq 0.25 \quad . \end{cases}$$

Furthermore, to evaluate how well the neural network predicts the dust species, we use the definition of true and false positives, and true and false negatives (e.g. Fawcett, 2006) to build a confusion matrix. The classification accuracy for each dust species class is defined as the fraction of correct predictions out of the total number of predictions of each class from the neural network.

Moreover, we evaluate the performance of the neural network on estimating the prediction uncertainties. To do so, we define criteria to estimate the reliability of predictions for each $\mathrm{M_{dust}}$ and $\mathrm{T_{dust}}$ as

$$a1 \times \ \sigma^{\mathrm{pred}}_{k,n} < a2 \times \ |y^{\mathrm{pred}}_{k,n}| \tag{3.7}$$

and for dust species as

$$a1 \times \ \sigma^{\mathrm{pred}}_{\kappa,n} < a2 \tag{3.8}$$

where $a1$ represents the limits in the 3σ rule for the estimated standard deviation ($\sigma_{k,n}^{\text{pred}}$), k corresponds to the index of a target value, and n is a data point from a test data set. The $1\sigma_{k,n}^{\text{pred}}$, $2\sigma_{k,n}^{\text{pred}}$, and $3\sigma_{k,n}^{\text{pred}}$ correspond to confidence intervals of 68%, 95%, and 99.7%, and $a2$ represents the coefficient of a relative estimated error of target value k for data point n, when the relative estimated error is defined as $Y_{n,k}^{\text{pred}} = \pm\, a2 \times y_{n,k}^{\text{pred}}$.

Here we introduce the reliable predicted standard deviation ($\sigma_{\star}^{\text{pred}}$). The $\sigma_{\star}^{\text{pred}}$ is a σ^{pred} that satisfies the Equations 3.7 and 3.8 with conditions $a2 = 0.2$ and $a1 = 1$. Therefore, the predicted target values that have $\sigma_{\star}^{\text{pred}}$, are considered as reliable predictions. On the contrary, for the target values that have σ^{pred} that are not satisfying the criterion for $\sigma_{\star}^{\text{pred}}$ that is defined by Equations 3.7 and 3.8, the predictions from the neural network are not considered as reliable predictions.

To compare the performance of the neural network in each step of the feature selection process, we use two values from the neural network output; i) the values that are reached by the loss function (i.e. Equation 3.4), for the training and validation data sets at the end of the training process, ii) the ratio of the number of predictions that have $\sigma_{\star}^{\text{pred}}$, to the total number of predictions of the test data set (hereafter $\sigma_{\star}^{\text{pred}}$).

Since we choose a fixed set of hyperparameters (see Section 3.4.3), for instance, a fixed number of epochs, the minimum loss achieved by the neural network in the training process in each step of the feature selection process can differ from the ‘absolute or true’ minimum that could be achieved, if the hyperparameters were to be re-adjusted for each step. This is independent of the chosen subset of JWST filters and happens in all scenarios. Ideally, in order to reach the absolute minimum loss possible one should re-adjust the hyperparameters for each step. However, this is a very time consuming process. Additionally, this would make the entire feature selection process dependent on the training data set as well as on the subset of the JWST filters, while not providing further relevant information for all the steps necessary to obtain the final preferred subset of JWST filters.

3.6 Caveats

Typically, very low amounts of dust (less than about 10^{-5} $\mathrm{M}_{\odot}$) are not easily observable in supernovae, since the thermal dust emission is rather weak at the expected wavelengths. This means that in some of our SN model SEDs that contain such low amounts of dust, the thermal dust emission in the simulated SEDs may either not be clearly discernible from the emission of the supernova or generally remains below the detection capabilities of JWST. Such SN model SEDs that exhibit barely noticeable or no dust signatures may therefore also remain largely unrecognised by our neural network.

In what follows, we train the neural network on the synthesised photometric data set for S1 to identify the SN model SEDs in this data set that have the lowest $\mathrm{M}_{\mathrm{dust}}$ and $\mathrm{T}_{\mathrm{dust}}$ that still can be recognised by the neural network. We find that for SN model SEDs with $\mathrm{M}_{\mathrm{dust}} < 5 \times 10^{-5}$ $\mathrm{M}_{\odot}$ and $\mathrm{T}_{\mathrm{dust}} < 800$ K the predicted dust properties have very large uncertainties (i.e. $|\Delta y_{T_{\mathrm{dust}}}| \gtrsim 0.5 \times y^{\mathrm{sim}}_{T_{\mathrm{dust}}}$), causing so called catastrophic outliers. Consequently, we train the neural network again, but this time to label such SN model SEDs as ‘no-dust’ data points (See Section 3.4.5), similar to the SN model SEDs that indeed contain no dust. Figure 3.2 presents an example set of such no-dust SN model SEDs with $\mathrm{M}_{\mathrm{dust}}$ and $\mathrm{T}_{\mathrm{dust}}$ below the aforementioned thresholds. Due to the fact that from the no-dust SN model SEDs the predicted dust properties including their uncertainties are highly unreliable, we do not include these models in subsequent performance evaluations of the dust properties.

3.7 Results

We investigate whether a neural network can be used as an effective tool to determine different properties of dust that formed in and around core-collapse SNe from its spectral energy distribution. Since the number of observed SNe is too sparse to be used for such an endeavour, we simulate a total of 293 236 SN SEDs (referred to as SN model SEDs), each with different dust properties. Then, we convolve each SN model SED with the entire suite of JWST NIRCam + MIRI banpass filters (see details in Section 3.3) to synthesise a photometric data set that is suitable for machine learning purposes.

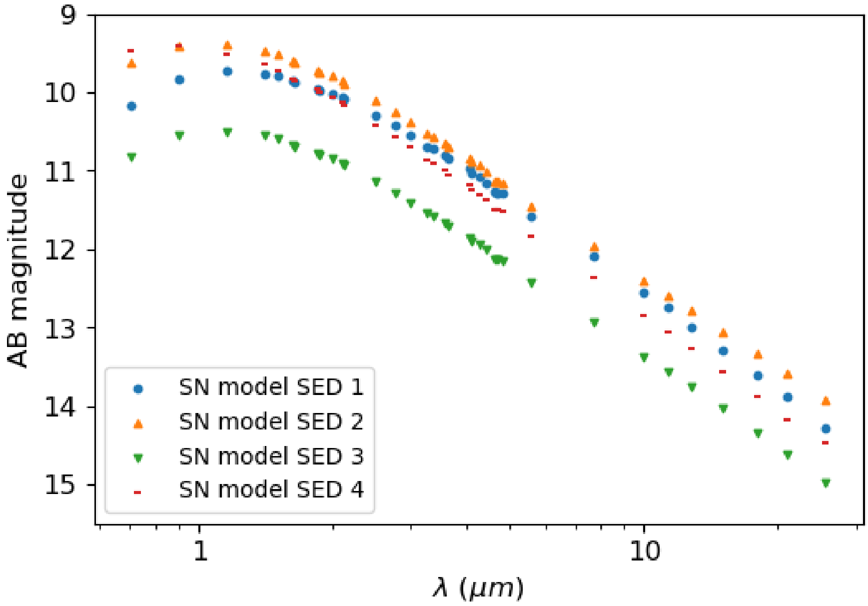

Figure 3.2: An example set of SN model SEDs with unrecognisable dust signatures. Each symbol represents a synthesised magnitude by a JWST filter and is shown at its central wavelength. The filled circles and triangles correspond to the SN model SED 1 and 2, that contain silicate dust, while filled dash and downward triangles correspond to SN model SED 3 and 4, respectively, both contain a mix of silicate and carbon dust. The amount and the temperature of dust in the SN model SEDs one to four, are about 2×10^{-5} $M_{\odot}$, 314 K, 5×10^{-5} $M_{\odot}$, 329 K, 10^{-5} $M_{\odot}$, 620 K, 10^{-5} $M_{\odot}$, 617 K respectively.

For a step by step analysis we consider three different scenarios, which are described in more detail in Section 3.3.2. In short, for the first scenario, S1, all SN model SEDs are placed at the same, low redshift, $z = 0.0001$. In the second scenario, S2, we uniformly distribute the SN model SEDs within the redshift range 0.0001–0.015. In the third scenario, S3, we use the data set of S2 and add random noise that corresponds to a photometric uncertainty of 0.1 mag (see the details in Section 3.3.1). Comparing the outcome of these scenarios allows us to examine how strongly the performances of the neural network and the feature importance change for our simulated data that are equipped with properties of real observations.

In our approach, we trained our neural network to predict the distribution of dust quantities given the SN model SEDs, $p(y^{\mathrm{sim}}|\mathbf{x}) \approx \mathcal{N}(y^{\mathrm{sim}}; y^{\mathrm{pred}}, \sigma^{\mathrm{pred}})$. To evaluate how well our estimated uncertainty align with the prediction error, we analyse the distribution of the normalised prediction errors $(y^{\mathrm{pred}} - y^{\mathrm{sim}})/\sigma^{\mathrm{pred}}$. Under perfect neural network modeling, the distribution of these normalised values must follow a standard normal distribution. Figure 3.3 shows histograms of the normalised prediction errors for $\mathrm{M}_{\mathrm{dust}}$ and $\mathrm{T}_{\mathrm{dust}}$ of a test data set predicted by the trained neural network with the entire set of JWST filters, in S3, excluding the predictions that the neural network classifies them as no-dust. By fitting a normal probability distribution function to the normalised prediction errors, we find that for $\mathrm{M}_{\mathrm{dust}}$ distribution, a mean of 0.04, and a standard deviation of 0.94 are inferred. The inferred values for $\mathrm{T}_{\mathrm{dust}}$ are mean and standard deviation of -0.001 and 0.76 respectively. Therefore, the inferred standard deviations corresponding to $\mathrm{M}_{\mathrm{dust}}$ and $\mathrm{T}_{\mathrm{dust}}$ are 6% and 24% lower than in a standard normal distribution. This might indicate that the predicted uncertainties, σ^{pred}, are overestimating the prediction errors.

For each scenario, S1, S2 and S3 we discuss four cases of a performance evaluation. For case-1 and case-2 we evaluate the performances of our neural network that is trained on data sets that consist of preferred subsets of JWST filters (see Section 3.8 for further discussions on the selection of preferred subsets). For case-3, and case-4 we evaluate the performances of our neural network that is trained with data sets that are constructed with a minimum subset of JWST filters (see definition section 3.5), with which the different dust quantities are predicted with an acceptable level of accuracy. The latter means that the fraction of reliable predictions, out of the entire test data set, is $\gtrsim 5\%$. Furthermore, for case-1 and case-3 we apply the evaluation metrics on the entire test data set. For case-2 and case-4, we apply the metrics only on the sub-sample of the test data set that satisfies the criteria for being reliable predictions as defined in Section 3.5.

Tables 3.2 and 3.3 summarise the outcome of the case by case performance evaluations of our neural network to predict $\mathrm{M_{dust}}$, $\mathrm{T_{dust}}$ and to classify the dust species for all three scenarios S1, S2 and S3. Out of all scenarios and all cases we find that in S1 and for case-2, the RMSE of both $\mathrm{M_{dust}}$ and $\mathrm{T_{dust}}$ is the smallest and $\sigma_{\star}^{\mathrm{pred}}$ is maximal. For case-2, the RMSE of $\mathrm{M_{dust}}$ increases from ~ 0.05 dex in S1, to ~ 0.1 dex in S2 and to ~ 0.11 dex in S3. However, for $\mathrm{T_{dust}}$ the RMSE increases from about 14 K in S1, only to ~ 18 K in S2. From S1 to S3, the RMSE of $\mathrm{T_{dust}}$ increases to ~ 30 K in S3. From both S1 to S3, in case-2, the fraction of 3σ outliers for $\mathrm{M_{dust}}$ target values increases.

The bias of the $\mathrm{T_{dust}}$ predictions for most of the scenarios for case-3 and case-4 is negative. This indicates that the neural network underestimates the $\mathrm{T_{dust}}$ target values (i.e. $y_{T_{\mathrm{dust}}}^{\mathrm{pred}}$). For case-1 and case-2 the bias is positive for $\mathrm{T_{dust}}$ in all scenarios. This indicate that the neural network overestimates the $\mathrm{T_{dust}}$ target values. For instance, in case-3 for S1, the bias of 0.013 (dex) for $\mathrm{M_{dust}}$ represents that the average of $\mathrm{M_{dust}}$ estimations over all the test data set is about $10^{0.013} \approx 1.03$ times more than the simulated $\mathrm{M_{dust}}$. For $\mathrm{T_{dust}}$ the average of $\mathrm{T_{dust}}$ estimations over all the test data set is about 1 K more than the simulated $\mathrm{T_{dust}}$ values.

As shown in Table 3.3, the highest classification accuracy for dust species is achieved for S2 for case-2. For this, we find a classification accuracy of 97%, 98%, and 100% for carbon, mixed and silicate dust, respectively. Comparing the classification accuracy for each dust species, we find that for all scenarios and all cases, silicate dust is predicted with the highest accuracy. Carbon dust is predicted least accurately in all scenarios and cases, except in S3 for case-4. There the SN model SEDs that are labeled as mixed dust are predicted with the lowest accuracy (57%). In case-4 and S3, 42% of the mixed dust species are predicted as carbon dust.

In Figures 3.4, 3.5, and 3.6, the performance of the neural network is shown for case-1 and case-2 for all scenarios. Overall, the performance of the neural network for case-2 is better than for case-1. As illustrated in the top panels of Figures 3.4 and 3.5, the dispersion of the predictions around the diagonal line that represents predicted values equal to simulated values, increases from S1 to S2 for both target values $\mathrm{M_{dust}}$ and $\mathrm{T_{dust}}$.

Moreover, as summarised in Table 3.3 the classification accuracy decreases for all dust species from S1 to S3 in case-1.

As shown in Figure 3.4, for S1 the reliable predictions for M_{dust} and T_{dust} range between about $6 \times 10^{-5} - 10^{-1}$ $M_\odot$ and 100–1400 K, respectively. However, Figure 3.5 shows that in S2, the reliable predictions only range between about $10^{-4} - 5 \times 10^{-2}$ $M_\odot$, and 250–1200 K for M_{dust} and T_{dust}, respectively. Figure 3.6 shows that in S3, the reliable predictions for M_{dust} are within $5 \times 10^{-4} - 10^{-1}$ $M_\odot$ and 250–1000 K for T_{dust}. This means that the dust mass and temperature range of the reliable predictions for all cases shrinks from S1 to S2 to S3, and thus with the increased complexity of the scenarios.

3.7.1 Feature selection

Figure 3.7 presents the performance of the neural network with the subsets of the JWST filters that are selected in each step of the feature selection process. The bottom panel compares the training losses obtained for the last epoch at all feature selection steps for all three scenarios. The validation losses for S3 are also included in Figure 3.7. It is evident that for S3, the validation loss closely follows that of the training loss. We find the same for the other two scenarios, although the loss values vary more drastically from step to step. The absolute local minimum of both the validation and training loss appears to be reached in step zero of the feature selection process for S3, while for both S1 and S2 the absolute local minimum is reached in step five. However, we find for S3 that both the training and validation loss slowly increase from step zero to eight by about 5%.

The first three panels of Figure 3.7 show the performance evaluation of the neural network for the test data sets. It is evident that the RMSE for T_{dust} varies only minimally around a mean value of about 58 ± 9 K, after which it increases to about 240 K in the last step. The RMSE of M_{dust} behaves similarly constant over the first eight steps except for step two, and increases from step eight to eleven by about 0.45 dex. The classification accuracy for carbon dust and the mixed composition also only changes minimally over the first eight steps, but appears to decrease from step eight to eleven from about 70% to about 50%. For silicate dust the classification accuracy remains nearly 100% over all steps.

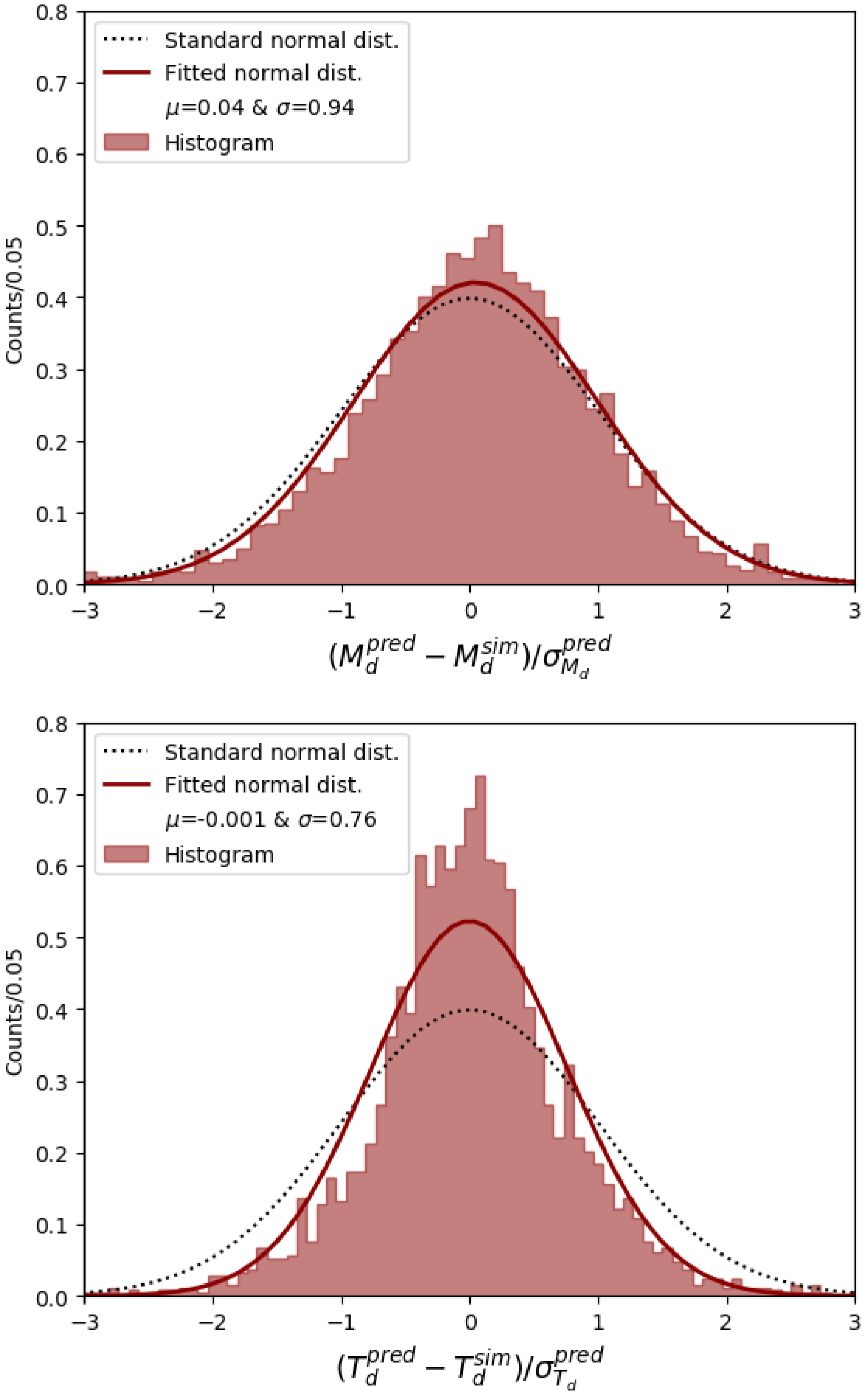

Figure 3.3: Comparison of the distribution of normalised prediction errors to a standard normal distribution. The histograms represent the distributions of $(y^{\text{pred}} - y^{\text{sim}})/\sigma^{\text{pred}}$ for M_{dust} (*top panel*), and T_{dust} (*bottom panel*), for a test data set predicted by the trained neural network with the entire suite of JWST filters, in S3. The dotted curves represent the standard normal distributions (i.e. $\mathcal{N}(0,1)$). The solid curves represent the normal distributions fitted to each of the histograms with μ=0.04 and σ=0.94 for M_{dust}, and μ=-0.001 and σ=0.76 for T_{dust}.

Table 3.4: The preferred and minimum subsets of JWST filters obtained from the feature selection process and used to estimate M_{dust}, T_{dust}, and dust species. This is shown for all three scenarios. Columns termed 'Pref.' and 'Min.' stand for preferred and minimum subset of JWST filters, respectively.

JWST filters	S1		S2		S3	
	Pref.	Min.	Pref.	Min.	Pref.	Min.
$F2550W$	●					
$F2100W$	●	●	●			
$F1800W$	●		●	●	●	●
$F1500W$	●	●	●	●	●	●
$F1280W$	●	●	●	●	●	
$F1130W$	●	●			●	●
$F1000W$	●		●	●	●	●
$F770W$	●	●	●	●	●	●
$F560W$	●	●	●	●	●	●
$F480M$					●	●
$F470N$						
$F466N$	●					
$F460M$	●	●			●	
$F444W$	●					
$F430M$	●					
$F410M$			●	●		
$F405N$			●			
$F360M$	●		●	●		
$F356W$			●	●	●	●
$F335M$	●					
$F323N$						
$F322W2$			●			
$F300M$	●		●	●		
$F277W$			●	●		
$F250M$			●			
$F212N$						
$F210M$	●					
$F200W$			●			
$F187N$			●			
$F182M$			●			
$F164N$	●					
$F162M$						
$F150W$	●		●	●		
$F140M$	●		●	●	●	●
$F115W$	●		●		●	
$F090W$						
$F070W$	●		●		●	●

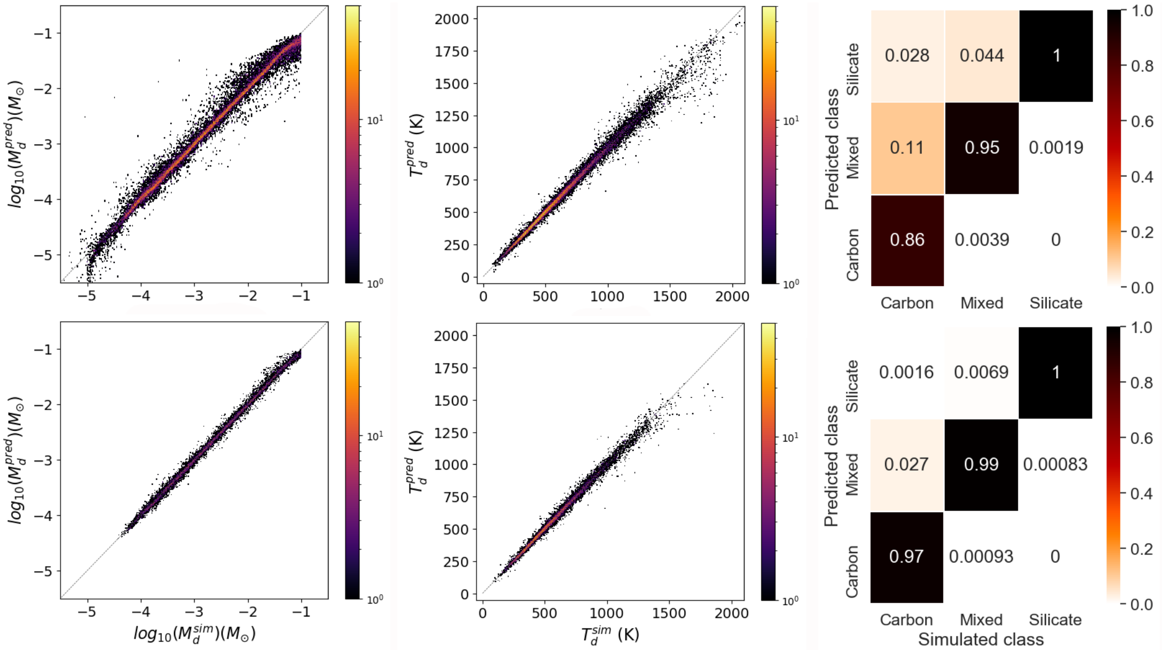

Figure 3.4: The performance of the neural network with the preferred subset of JWST filters, for S1. The M_{dust} (*left column*), and T_{dust} estimates (*middle column*), and dust species classification (*right column*), are shown for all the predictions of the test data set (*top panel*), and the reliable predictions of the test data set (*bottom panel*). The dust species classifications are shown in the format of confusion matrices that represent the simulated dust species against the predicted dust species. The colour-bars in the left and middle diagrams indicate the number of predictions. The dashed lines represent the predicted values equal to the simulated values of M_{dust} and T_{dust}.

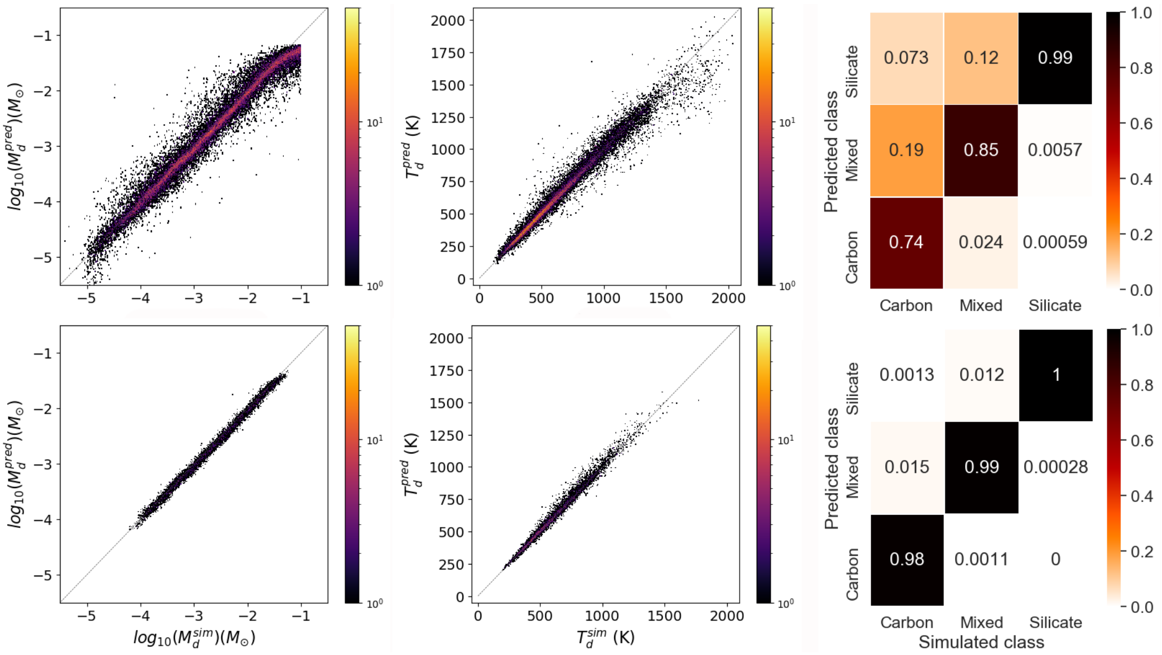

Figure 3.5: The performance of the neural network with the preferred subset of JWST filters, for S2. The definition of the panels, the variables, the dashed lines and the colour bars are the same as the definitions in Figure 3.4.

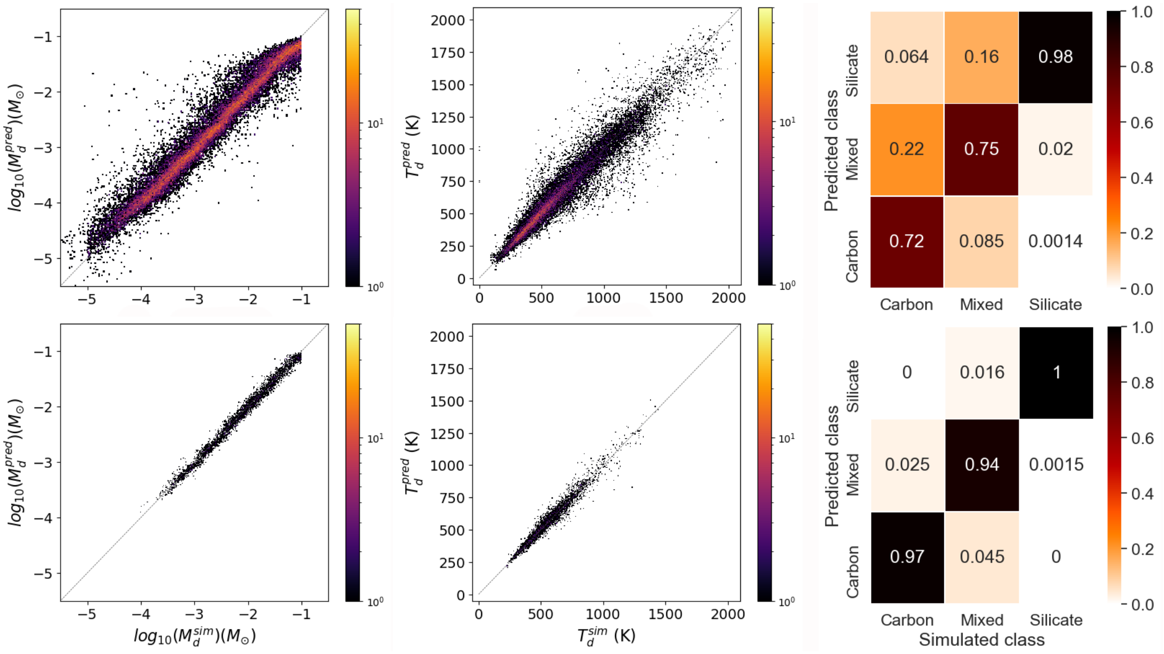

Figure 3.6: The performance of the neural network with the preferred subset of JWST filters, for S3. The definition of the panels, the variables, the dashed lines and the colour bars are the same as the definitions in Figure 3.4.

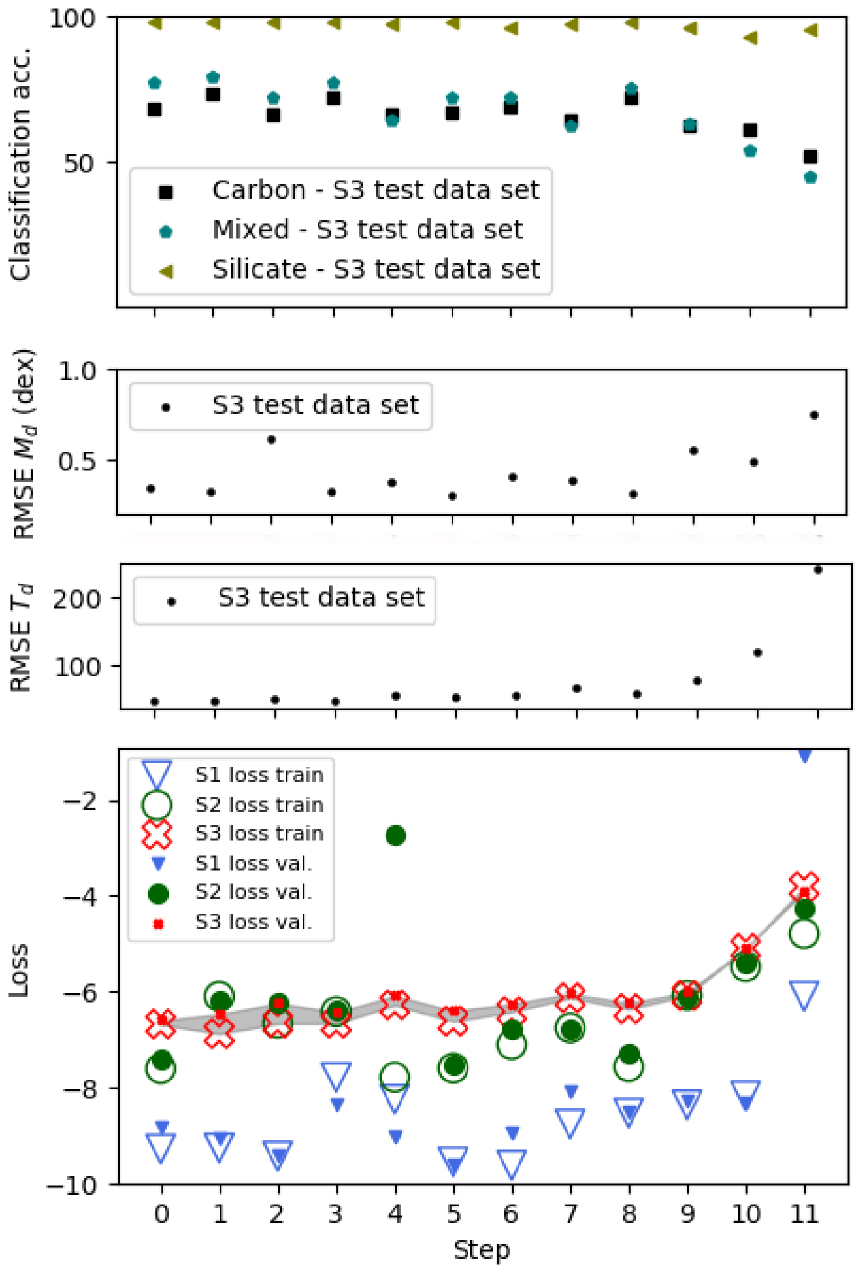

Figure 3.7: The performance of our neural network with each filter set that are obtained at each step of the feature selection process. *Bottom Panel*: The loss values that are achieved by the training and validation data sets at the end of each training process of the neural network in S1 (downward triangles), S2 (circles), and S3 (X symbols). The empty symbols represent the training loss. The filled symbols represent the validation loss. The grey shaded region shows the area between the training and validation loss in S3. The single panels show the RMSE of T_{dust} (K), and M_{dust} ($M_\odot$), and the classification accuracy (%) for predicting the dust species for the test data sets in S3, from bottom to top. The classification accuracy for predicting carbon and silicate dust species, and a mixture of them are shown with circles, triangles, and dashes respectively.

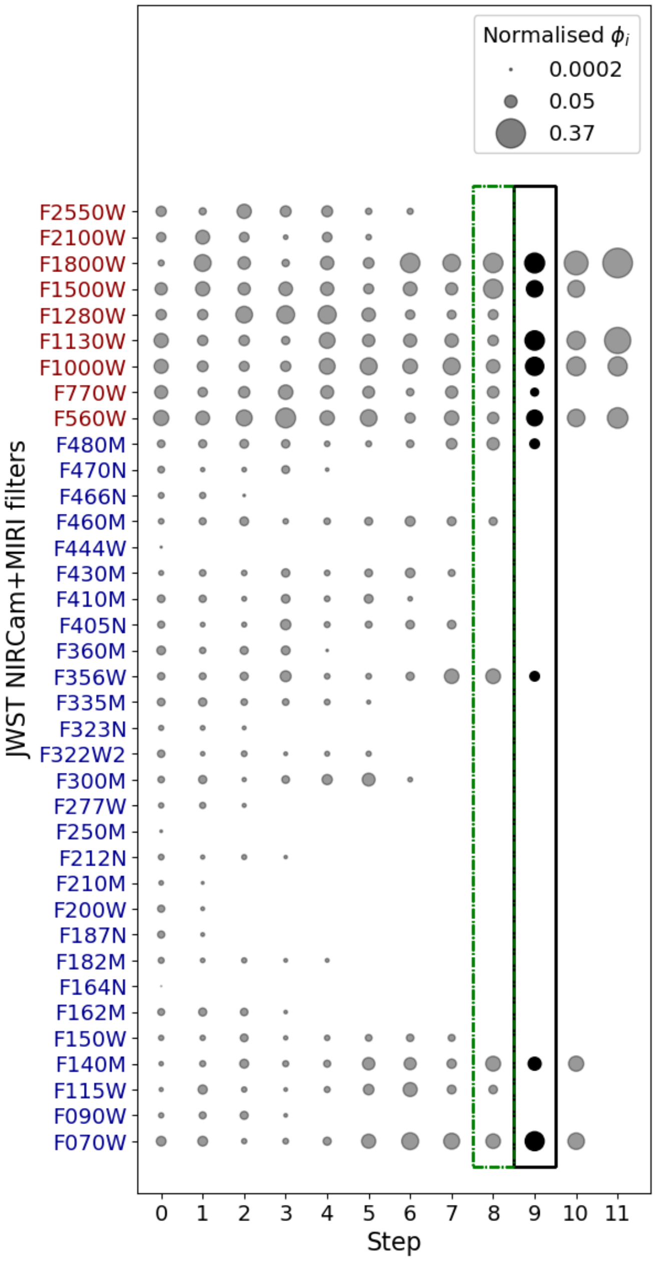

Figure 3.8: Importance of JWST filters for estimating the amount, temperature and the dust species, in S3. The normalised feature importance (ϕ_i) of each NIRCam (blue) and MIRI (red) filters in each step of the feature selection process is shown by the size of the filled circles that are scaled to three values in the legend. The preferred and minimum subsets of JWST filters are highlighted with dash-dotted-line and solid-line boxes, respectively.

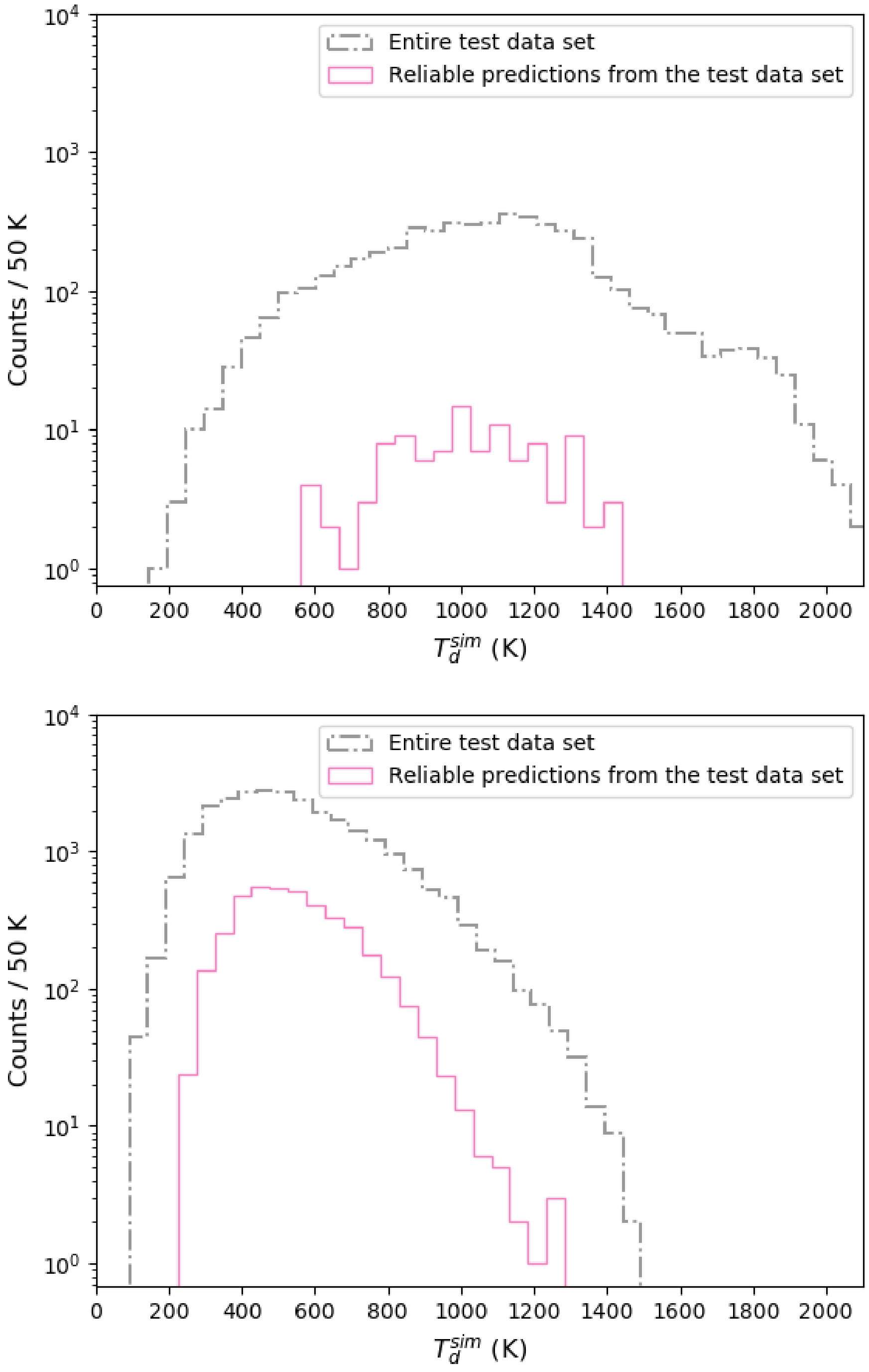

Figure 3.9: Distribution of SN model SEDs in T_{dust} in 50 K bins. The dash-dotted line represents the distribution of the entire test data set and the solid line represents the distribution of the reliable predictions from the test data set. *Top panel*: The SN model SEDs with $R_{in} \lesssim 5 \times 10^{16}$ cm. *Bottom panel*: The SN model SEDs with $R_{in} \gtrsim 5 \times 10^{16}$ cm.

3.8 Discussion

The performance evaluation of our trained neural network, which is designed to predict dust properties such as M_{dust}, T_{dust} and different dust species, demonstrates that neural networks can be a powerful tool, if a sufficiently large data set is at hand. One advantage of using such a method is that it is possible to obtain a good estimate on the prediction uncertainties for each dust property under consideration. For other common methods, such as fitting a simple modified black body function or combination of thereof, uncertainties of the fitted dust mass or dust temperature are often not obtained (e.g. Gall *et al.*, 2011b, and references therein). Furthermore, due to the fact that for such fitting methods assumptions about the dust composition need to be made a-priori to fitting, the parameter range can be large and often not explored in all detail. The reasons for this may include insufficient data quality, but also time and computational limitations. These issues also apply to more sophisticated dust models such as `MOCASSIN`, when used to fit observational data to obtain the amount and temperature of dust in and around supernovae (see e.g., Wesson *et al.*, 2015).

Our performance evaluation demonstrates that for all scenarios and cases (see Table 3.2), the obtained prediction error, RMSE, for M_{dust} is smaller than ~ 0.55 (dex), and is smaller than ~ 78 K for T_{dust}. These RMSE are obtained for case-3 and S3 and are the maximum RMSE values out of all scenarios and cases. This is because in case-3, the evaluation metrics are applied only onto the test data-sets and minimum subsets of JWST filters of each scenario S1, S2 and S3. Moreover, for the evaluation of case-3 the test data sets are used without prior $\sigma_{\star}^{\rm pred}$ cut and hence, contain predictions with larger uncertainties. Additionally, S3 is the most complex scenario of all scenarios. However, compared to other works in the literature with inferred amounts and temperatures of dust from observed supernovae, we find that even the worst performance here in this work constitutes a very good performance. For example, we can compare to other works that estimate the amount of dust, with *Spitzer Space Telescope* observations up to about 25 μm for SNe such as SN 2004et (Kotak *et al.*, 2009) and SN 1987A (Ercolano *et al.*, 2007). For SN 2004et, the estimated range for dust mass and dust temperature at 300, 464 and 795 days after the explosion, are about 0.37 dex and 500 K, 0.26 dex and 250 K, and 0.38 dex and 80 K. For SN 1987A, the amount of carbon dust at day 615 has been estimated with an uncertainty of 0.81 dex.

We now turn to the performance evaluations of the most reliable predictions, which are drawn from case-1 and case-3 data sets that have $\sigma_\star^{\mathrm{pred}}$, evaluated as case-2 and case-4. Comparing the RMSE for $\mathrm{M_{dust}}$ and $\mathrm{T_{dust}}$ between case-1 and case-2 (cases with the data sets that contain the preferred subsets of JWST filters) across the three scenarios, S1, S2 and S3, shows that the prediction errors are reduced by up to a factor of about 2–3 in case-2 where the predictions that do not have $\sigma_\star^{\mathrm{pred}}$ are excluded. Since this is an expected, but not guaranteed, consequence of including only the predictions that have $\sigma_\star^{\mathrm{pred}}$, which removes 'bad' predictions that do not fulfill the criterion to have $\sigma_\star^{\mathrm{pred}}$, the same is expected for the cases with the data sets that contain the minimum subsets of filters (case-3 and case-4). Our evaluations show that the effect of excluding the unreliable predictions for $\mathrm{M_{dust}}$ estimations is even stronger than that for $\mathrm{T_{dust}}$, meaning that the RMSE (in dex) of the dust mass is smaller by about a factor of 4–5 in case-2 and case-4, compare to case-1 and case-3, while for $\mathrm{T_{dust}}$ the decrease is only about factor of 2. The classification accuracy of classifying the different dust species shows the same behavior, which is higher for nearly all species and scenarios for case-2 and case-4 than for case-1 and case-3. Particularly for silicate dust, the classification accuracy is close to or at 100%. However, in case-4 and S3, there is a bias in predicting the mixed dust species towards the carbon dust species. The evaluation method using the $\sigma_\star^{\mathrm{pred}}$ definition demonstrates that the dust mass and temperature predictions that have been under the scrutiny of the $\sigma_\star^{\mathrm{pred}}$ criterion can truly be considered as reliable predictions.

On the other hand, as shown by $\sigma_\star^{\mathrm{pred}}$ in Table 3.3, the number of predictions that satisfy the $\sigma_\star^{\mathrm{pred}}$ criteria in case-2 and case-4 is smaller than the predictions using the entire data set as in case-1 and case-3. Since there are 3σ outliers (as defined in Section 3.5) also for case-2 and case-4, the fractions of the best reliable predictions for S1, S2 and S3 in case-2 and case-4 are smaller than $\sigma_\star^{\mathrm{pred}}$. For instance, in case-2 for S1 the fraction of the best reliable predictions is still about 59% while in case-4 and S3 it shrinks to only about 5.8%.

Comparing the number of 3σ outliers between $\mathrm{M_{dust}}$ and $\mathrm{T_{dust}}$, we find that for nearly all setups of cases and scenarios, the $\mathrm{M_{dust}}$ evaluations result in a larger number of 3σ outliers than the $\mathrm{T_{dust}}$ evaluations. This is because the dispersion of $\mathrm{M_{dust}}$ residuals is larger than $\mathrm{T_{dust}}$ residuals.

Since observing supernovae with all the JWST filters at the same time is practically not feasible, we are interested in finding the smallest set of filters with which an acceptable performance can be achieved. To do so, we utilize a feature selection process as described in Section 3.4.6. From this we obtain two sets of filters for each scenario, one preferred set of filters and one minimum set of filters. The preferred filter set is chosen based on the absolute minimum reached by both the training and the validation loss while the minimum filter set is chosen based on criterion that the fraction of the number of reliable predictions to the total number of predictions is larger than 5%. It turns out that for S1 and S2 the preferred filter set is reached early in the filter selection process, step 5, and thus still contains a large number of filters (22 filters). The minimum filter set is obtained in steps eight or ten, and thus contain fewer, between seven to thirteen, filters. Looking at the performance evaluation from the two filter sets, for example case-1 and case-3 or case-2 and case-4 in Table 3.2, then while as expected, the performance of the neural network is overall better with the preferred set of filters. The performance with the minimum filter set is only minimally decreased. Hence, as demonstrated in Table 3.2, accurate predictions of $\mathrm{T_{dust}}$ and $\mathrm{M_{dust}}$ can be achieved with the minimum set of filters.

For scenario S3, the preferred and the minimum subset of filters are chosen from step eight and nine and are thus very close to each other. It is important to note that in this case the preferred set is chosen to be at step 8 instead of step zero, where the loss reached the absolute minimum. However, we do not consider step zero as ‘preferred’. Since both the training and validation loss remains rather stable until step 8 as pointed out in Section 3.7.1, step eight can be considered as preferred.

As illustrated in Figure 3.7, in S3 compared to S2 there are insignificant changes of loss values in each step of the feature selection process up to step 9. This stability of the performance of the neural network in S3, regardless of the number of filters that are used as the input features can be due to the training of the neural network with additional noise. This is because the training of a neural network with additional noise can be equivalent to a regularisation (Bishop, 1995), which helps the neural network to react less to the variation of input features. Therefore, in S3 compared to S2, the training and validation losses that are achieved by the neural network with smaller sets of filters than the entire filter set, do not significantly change in each step of the feature selection process up to step 9.

Figure 3.8 visualises the resulting Shapley values obtained for each step in scenario S3. Figures 3.12, and 3.13 show the same for S1 and S2, respectively. It is interesting to note that for all three scenarios, none of the narrow-band JWST filters are amongst the minimum subsets of the JWST filters. However, for S1 and S2 two such narrow-band filters are included in the preferred filter set albeit with small Shapley values and hence, marginal importance.

This implies that real observations of supernovae with such JWST narrow-band filters would have the least impact on estimating dust properties with our neural network. As shown in Figure 3.8, and expected, the MIRI filters that cover the longer wavelength region are crucially important to estimate the dust properties while the shorter wavelength NIRCam filters seem not to play a significant role.

One of the most pressing questions of course is: ‘Is it technically feasible to construct an observing run with the minimum subset of filters?’ The NIRCam instrument uses a dichroic to split the incoming radiation into two wavelength ranges, $\lambda < 2.5\ \mu$m and $\lambda > 2.5\ \mu$m, known as short and long wavelength channels (Horner and Rieke, 2004). This setup allows to simultaneously obtain two images with two different filters, each from one of the channels. Since, in the minimum subset for S3, two selected NIRCam filters, $F070W$ and $F140M$, are in the short wavelength channel of NIRCam and two, $F356W$ and $F480M$, are in the long wavelength channel, two separate runs are required to observe a supernova with all four NIRCam filters. For MIRI, observations can only be conducted with one filter at a time. The entire observing time needed for all selected MIRI filters of the preferred subset may in the end depend on the brightness of the supernova, the desired / best possible signal-to-noise ratio or the phase of the supernova.

Figure 3.9 shows the histogram of test data set over T_{dust}, in S3, for the entire test data set (dashed line), and the sub-sample of test data with reliable estimated standard deviations (solid line). In the top panel, the distribution is shown for the SN model SEDs with $R_{\mathrm{in}} \lesssim 5 \times 10^{16}$ cm, while the bottom panel represents the SN model SEDs with $R_{\mathrm{in}} \gtrsim 5 \times 10^{16}$ cm. This cutoff represents an approximate division of models into those which are representative of newly-formed dust in supernova ejecta, and those which are representative of pre-existing circumstellar dust, flash-heated by a supernova explosion. Pre-existing grains at radii less than 5×10^{16} cm are likely to be evaporated by the supernova explosion (Gall *et al.*, 2014). Meanwhile, supernova ejecta expanding at ~ 6000 km/s would reach this radius after ~1150 days; a supernova following the bolometric evolution of SN1987A (Seitenzahl *et al.*, 2014) would have a luminosity of

$\sim$10,000 $L_\odot$, the lowest considered in our models, at a similar epoch. Therefore, any dust located at distances $\gtrsim 5 \times 10^{16}$ cm (i.e. the top panel in Figure 3.9) is assumed to be pre-existing dust, and dust located at any distance $\lesssim 5 \times 10^{16}$ cm (i.e. the bottom panel in Figure 3.9) is assumed to be newly-formed dust. The newly-formed dust could be located in the SN ejecta, or in the case of Type IIn SNe, the interaction between the SN ejecta and sufficiently dense circumstellar material (CSM) that formed prior to the supernova explosion, and is located at $\sim 10^{16}$ cm from the explosion, could result in dust formation in a cold dense shell (Smith *et al.*, 2008). By comparing the covered areas in both panels, we find that for real observations, our neural network is better at estimating the dust properties for pre-existing dust than newly formed dust.

To use modern machine learning algorithms effectively, large data sets are essential. Presently ongoing wide-field surveys such as ZTF (Bellm, 2014), YSE (Jones *et al.*, 2021) or SkyMapper (Scalzo *et al.*, 2017) are discovering hundreds to thousands of SNe and other transients per year and are building up a wealth of optical photometric as well as spectroscopic data of various different types of CCSNe that will be further advanced in future surveys such as LSST Ivezić *et al.*, 2019. While near- to mid-IR observations of CCSNe will likely boom with the launch of JWST and possibly other, future instruments on ground-based telescopes, they are rare at present, and will most likely not reach the level required to train machine learning algorithms on observational near- to mid-IR data.

The Open Supernova Catalog[7] reported the discovery of 450 Type II, 102 Type Ib, and 60 Type Ic SNe in the year 2021, from which only a few have mid-IR data. Albeit this is a large amount of observed CCSNe in just one year, collecting a data set of the size, wavelength range and degree of variation used in our study will also in future not be easily feasible. This 'data-size' limitation is especially important for estimating the dust properties of the types of SNe we used here in this work, where we had simulate 293 236 SN SEDs covering a wavelength range from 0.7 to 30 μm. Finally, as the dust properties and quantities cannot directly be measured from the observational data, thus are unknown, well advanced simulations with known dust properties are highly valuable. Therefore, applying a neural network that is well trained on a rich set of highly advanced simulated data exploring a large parameter space may be a promising way to determine dust quantities and properties of future observations. This work also allows testing in

---

[7] https://sne.space

what detail quantities and properties of dust can be inferred from observational data. Furthermore future observational data, if included in the training of the neural network, can be used to validate the neural network and thus, will improve its performance and outcome.

3.9 Conclusion

In this work, we present a first test for using neural networks to estimate different quantities and properties of dust located in and around supernovae including their predicted uncertainties. We aim to predict the temperature and amount of dust and to differentiate between three dust compositions. To do so, we simulate an extensive data set of 293 236 SN model SEDs using the 3D photoionisation and dust radiative transfer code `MOCASSIN` (Ercolano *et al.*, 2003a; Ercolano *et al.*, 2005). We convolve the simulated data set with JWST MIRI and NIRCam bandpass filters. We consider the instrument's detection limits as well as the estimated magnitude uncertainties to make the trained neural network suitable for predicting the properties of dust in supernovae from future observations of this instrument. We define three different scenarios to examine the feasibility and accuracy of inferring the dust properties by our neural network. In the first scenario, we assume that all SN model SEDs have the same low redshift. In the second and third scenarios, we distribute all SN model SEDs within the redshift range of $0.0001 - 0.015$, in which at least 7 JWST bandpass filters of all SN model SEDs are within the sensitivity and saturation limits that are calculated for a S/N of 10. Additionally, in the third scenario, we add random noise to the distributed SN model SEDs within a redshift range of $0.0001 - 0.015$. Thereafter, we select a preferred and a minimum subset of JWST filters from the feature selection process which is based on the SHAP framework. We use these filter subsets to estimate the amount, temperature and dust species with our neural network.

From the outcome of our trained neural network in S3, which is the closest scenario to real observations, we find the minimum subset of JWST filters needed to estimate dust quantities and properties consists of NIRCam: $F070W$, $F140M$, $F356W$, $F480M$, and MIRI: $F560W$, $F770W$, $F1000W$, $F1130W$, $F1500W$, $F1800W$ filters.

As presented in Table 3.2, our neural network can well predict the dust quantities and properties for approximately 7% of SN model SEDs from the entire test data set. This fraction has a RMSE of ~ 0.12 dex, and ~ 38 K for M_{dust} and T_{dust}. The classification accuracy is 95%, 99% and 57% for carbon, silicate and a mix of carbon and silicate dust, respectively.

We find that the dust quantities and properties are best predicted by our neural network for SN model SEDs that approximately range in T_{dust} between 250–1000 K, and M_{dust} between $5 \times 10^{-4} - 10^{-1}$ $M_{\odot}$, and are dominated by astronomical silicates. Given our simulated data set, these SN model SEDs can be associated with a pre-existing dust scenario.

Acknowledgements

We thank Dr. Doogesh Kodi Ramanah and Dr. Adriano Agnello for helpful discussions. This work is supported by a VILLUM FONDEN Investigator grant (project number 16599) and a VILLUM FONDEN Young Investor Grant (project number 25501). RW acknowledges support from European Research Council (ERC) Advanced Grant 694520 SNDUST. This work has made use of the Horizon Cluster hosted by Institut d'Astrophysique de Paris, and an HPC facility funded by a grant from VILLUM FONDEN (project number 16599.)

3.10 Appendix

3.10.1 Sensitivity and saturation limits for NIRCam and MIRI

Figures 3.10 and 3.11 represent the sensitivity and saturation limits for observing with NIRCam and MIRI filters with a minimum signal-to-noise ratio of 10.

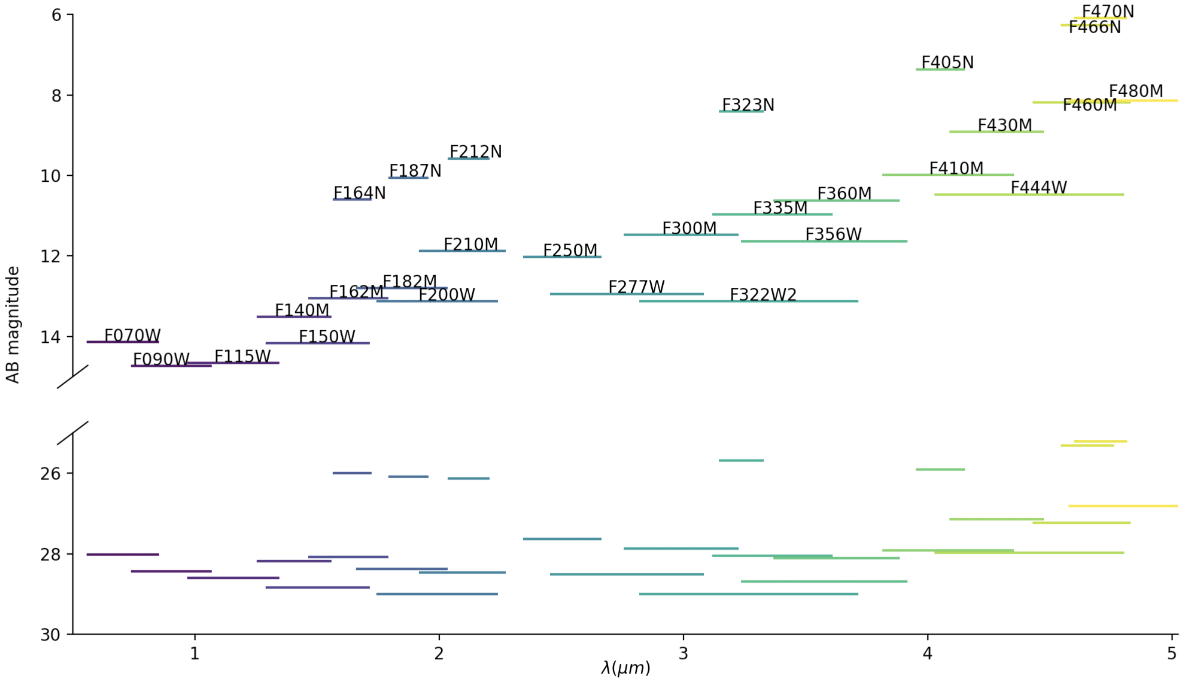

Figure 3.10: NIRCam saturation magnitudes in 10 000 seconds exposure time, and point source sensitivity for 21.4 seconds exposure time. The sizes roughly represent the wavelength range of each filter. There is a break in y-axis (15–25 AB magnitude) to save a large blank space between the sensitivity and saturation limits.

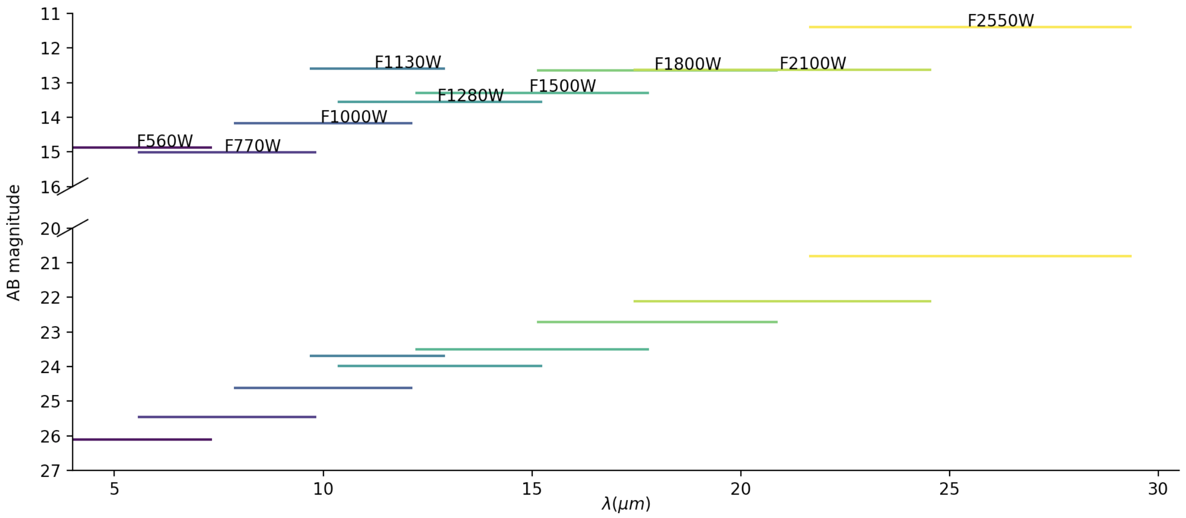

Figure 3.11: MIRI saturation magnitudes in 10 000 seconds exposure time, and point source sensitivity for 21.4 seconds exposure time. The sizes roughly represent the wavelength range of each filter. There is a break in y-axis (16–20 AB magnitude) to save a large blank space between the sensitivity and saturation limits.

Table 3.5: Computational costs of calculating Shapley values using DeepLIFT for different sub-samples of validation and training data sets. The instance highlighted in blue is the selected size in this work.

Type	Size (%)	Time consumption (s)
Computed	0.1	7
Computed	0.2	14.7
Computed	0.5	86.5
Computed	1	347
Computed	~ 2	3041
Estimated	4	1.9×10^{5}
Estimated	5	1.5×10^{6}
Estimated	10	4.5×10^{10}
Estimated	100	1.4×10^{91}

3.10.2 Computational caveats

Computational cost

Here we address the computational costs of implementing the DeepLIFT from SHAP framework on a photometric data set with a full set of JWST filters and redshift (i.e. 38 features). We compute the time consumption of calculating Shapley values for several sub-samples with different sizes from the training and the validation data sets. We fit an exponential function to the computed time consumption and the corresponding size of the sub-samples as the percentage of the training and validation data sets. We found the following function as the computational cost function for our data set:

$$\mathscr{Q}(x) = a + (b \times e^{-cx}) \ ,$$

where $a \approx -59.3$, $b \approx 51$, and $c \approx -2.06$. Table 3.5 summarises the computed and estimated computational costs as the time consumption for sub-samples with different sizes. The estimated times are calculated by $\mathscr{Q}(x)$ for a given x as the percentage of the training and validation data sets used to calculate the expectation and Shapley values, respectively. The computed computational costs are derived from an implementation of the algorithm on a MacBook Pro with 9 GHz 6-Core Intel Core i9 processor, and 32 GB 2400 MHz DDR4 memory.

Reproducibility

For the sake of reproducibility we made our code publicly accessible on GitHub[8], for any further evaluation and/or optimisation purposes. However, a specific distribution of computations over the processors has an effect on the training. Therefore, depending on each specific machine in use the reproduced outcome of the neural network can be slightly different (e.g. Bhojanapalli *et al.*, 2021).

3.10.3 S1, S2 feature importance

Figures 3.12 and 3.13 represent the resulting Shapley values obtained for each step in scenarios S1 and S2.

---

[8] https://github.com/ZoeAnsari/InferringSNdustwithNN

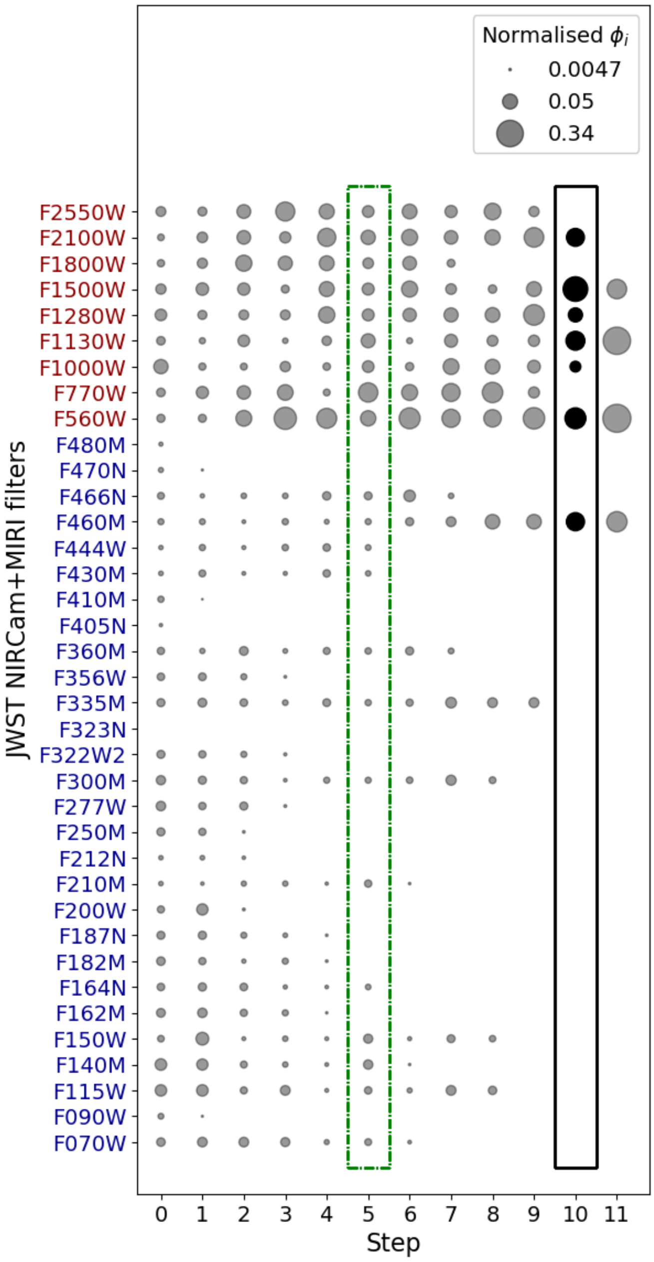

Figure 3.12: Importance of JWST filters for estimating the amount, temperature and the dust species, in S1. The symbols, relative sizes and colour codes are the same as defined in Figure 3.12.

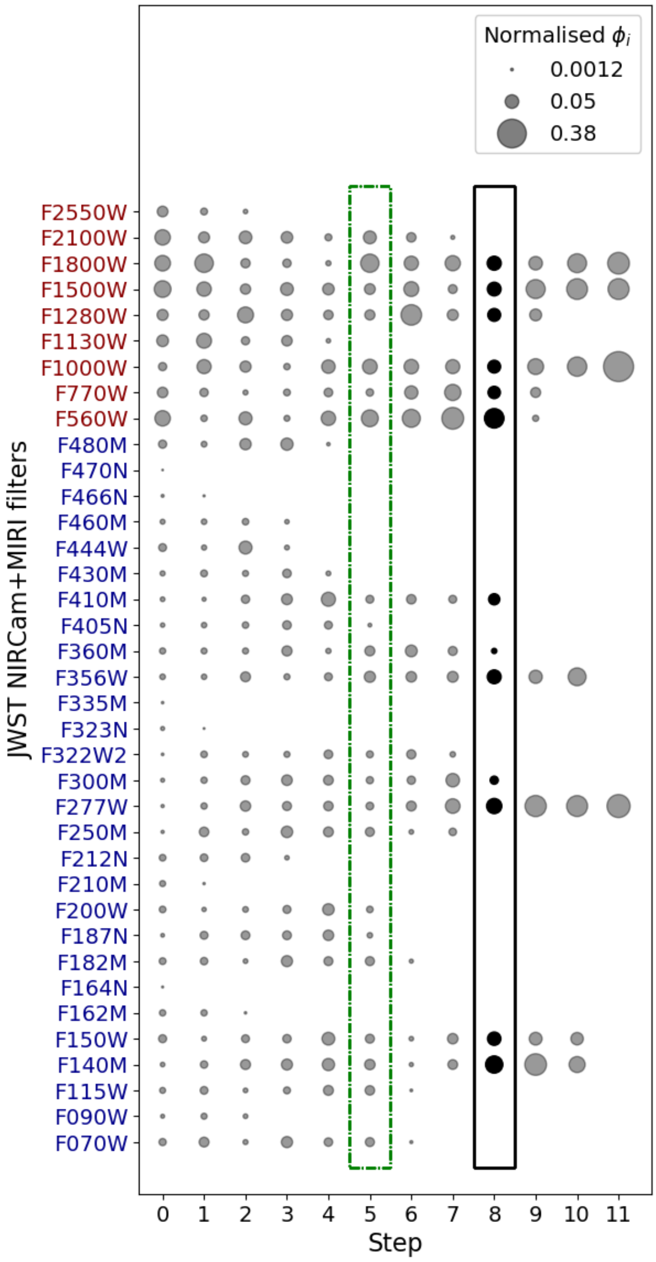

Figure 3.13: Importance of JWST filters for estimating the amount, temperature and the dust species, in S2. The symbols, relative sizes and colour codes are the same as defined in Figure 3.13.

Outlook

4

> *"In my view, all that is necessary for faith is the belief that by doing our best we shall succeed in our aims: the improvement of mankind."*
>
> -— Rosalind Franklin[a]

---

[a] Who made critical contributions to the understanding of the fine molecular structures of DNA

Understanding the physics that governs the universe via large volumes of astronomical observations requires efficient statistical methods to extract the information from this large amount of data. Machine learning algorithms as promising approaches in accurately estimating the astronomical quantities and properties with a minimum required computational time have become more demanded than traditional computational approaches. This thesis contributes to this emerging field in two science cases: a) estimating PDF photo-z; b) quantities and properties of dust in and around supernovae. The accuracy gained by estimations in this work competes with the estimations from the state-of-the-art techniques in both cases. Further improvements can be achieved by including richer data sets and improving the methods.

4.1 Photometric redshift

In Chapter 2, by implementing a probabilistic classifier followed by a probabilistic photo-z estimator, I included the uncertainties about identifying the type of objects in the process of estimating photo-z. Moreover, I let the method consider the effect of measurement errors on estimating PDF photo-z by including measurement errors as additional input features to the observed magnitudes in the process. Consequently, PDF photo-z estimates from the method reflect both the uncertainty of the object's type and the measurement uncertainties. Therefore, this trained machine learning algorithm that can estimate PDF photo-zs with a high level of accuracy can be employed to quickly estimate the redshift of newly observed objects in wide-field surveys for any astrophysical investigation sensitive to the type of objects and their error magnitudes.

4.1.1 Ongoing project: estimating PDF photo-zs by variational auto-encoders

In order to improve the reflection of measurement errors in estimated photo-z, as an alternative to feeding the measurements errors as additional input features to a machine learning algorithm, *importance sampling* can be used to include the measurement errors in the training process. In this approach, in each epoch of training of the algorithm, the observed magnitudes of each object in all filters will be updated by a random value within the range of their corresponding measurement errors as follows:

$$m_{i,\mathrm{up}} = m_i + \mathcal{N}(0, m_{i,\mathrm{error}}^2) \tag{4.1}$$

where $m_{i,\mathrm{up}}$ is the updated magnitude of each filter, and $\mathcal{N}(0, m_{i,\mathrm{error}})$ is a randomly generated number taken from a normal distribution with zero mean and σ equals to the corresponding measurement error. Letting the algorithm learn the relationships between the redshift and set of magnitudes that vary within the measurements errors range at each epoch of training is expected to lead to a robust and well-calibrated uncertainties on the photo-z estimations.

Moreover, to improve the separation accuracy between stars, galaxies, and quasars, an alternative approach to IGMM can be variational auto-encoders (VAEs). A VAE is a method that can learn to represent a set of input features as a set of typically low-dimensional latent variables via a non-linear embedding such as neural networks. Therefore, instead of implementing a classifier that separates the objects in the input feature space, by using VAEs, we can explore the possibility of a clearer separation of the objects. This is because the encoder can use highly non-linear transformations to position the data in the latent space. VAEs may improve the classification in particular due to their capability of capturing an extraordinary amount of information in their latent space.

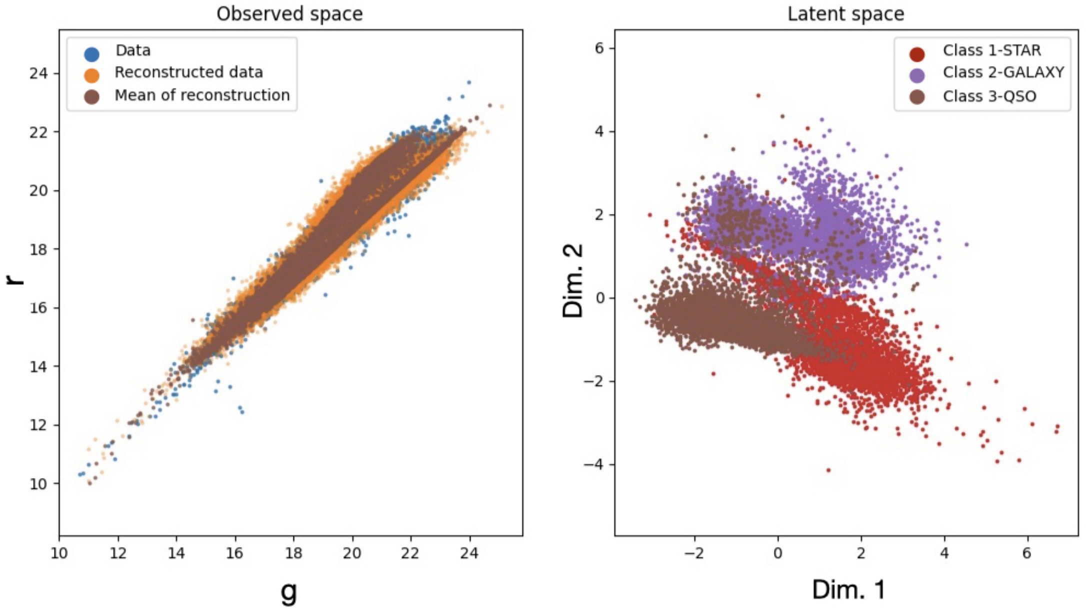

Figure 4.1: *Left panel*: Objects from the data set in g vs r feature space. The blue dots correspond to input features to the VAE (i.e. $\mathbf{x}$). The orange dots represent the reconstructed features from VAE (i.e. $\tilde{x}$). The brown dots represent the mean of reconstructed data.
Right panel: Objects from thee data set in the 2D latent space. The red, purple and brown dots represent stars, galaxies and quasars respective.

In an ongoing project, I implemented a VAE for classifying $\approx 20\%$ of the data set in Chapter 2 to stars, galaxies, and quasars in a latent space via neural networks. The current VAE consists of two neural networks as encoder and decoder. The encoder transforms the set of input features to a set of latent variables via learning the likelihood function $q_\theta(z_{\text{latent}}|x)$ that is defined by the encoder neural network, in which $\mathbf{x}$ is a set of input features, z_{latent}, is a set of latent variables. The decoder transforms the latent variables to a set of *reconstructed* features via defined likelihood function by the decoder neural network, $p_\phi(\tilde{x}|z_{\text{latent}})$. The $\tilde{x}$ is the predicted set of input features that should be equal to $\mathbf{x}$ in an ideal situation that the loss function of VAE reaches the global minimum. Both encoder and the decoder neural networks in our current VAE consist of three fully connected layers with ReLU as their activation functions, and we defined two dimensions for the latent space.

As shown in Figure 4.1, a promising separation of the data set in the 2D latent space is achieved. However, a further attempt to optimise the hyperparameters of the method may lead to a better separation.

As the second step in this approach, instead of adding the measurement errors as additional features in the input features set, I used importance sampling. In this approach, the photo-z estimates achieved RMSE of 0.14, for objects with $weight_{max} > 0.8$ membership probability to a MDN branch (as defined in Chapter 2). For galaxies with zs < 0.3, the RMSE is 0.01. The corresponding RMSE from the method explained in Chapter 2, are 0.11 and 0.01, respectively. Although the sample employed in this approach contains only 20% of the data set used in Chapter 2, the results look very promising. Therefore, including all the data set in the training and further optimisation for the encoder and decoder hyperparameters that can improve the outcome should be the next step in this project. Building a more robust photo-z estimator using VAEs and importance sampling helps to better capture the measurements errors as well as underlying degeneracy in the data. Eventually, estimated PDF photo-zs will be better that lead to more precise photo-z calibrations and more precise restrictions on cosmological parameters. This help us to investigate cosmological models more efficiently.

4.2 Dust in supernovae

In Chapter 3, by developing and training a neural network adjusted for JWST photometric observations, the amount, temperature, and composition of dust in and around supernovae can be estimated. Since I trained our neural network with a simulated data set that adequately covers the parameters space of dust properties, it became a reliable tool that operates quicker than other methods in the literature for predicting properties of dust from new observations.

Moreover, the minimum set of filters for inferring dust properties that I found in chapter 3 helps design observational strategies more efficiently for investigating the mechanism of dust formation in supernovae. Furthermore, I found that, in particular, our neural network is better at inferring newly formed dust than pre-existing dust, which in the future may lead to justifications in the approaches for inferring pre-existing dust in order to resolve this shortcoming.

4.2.1 Future plans

One way of improving the accuracy of predicting quantities of the dust is enriching the training data set. For example, by adding SEDs of supernovae with clumpy dust simulated by `MOCASSIN` to the training data set, the algorithm will learn the more complex relationships between the SEDs and the quantities of dust. Consequently, the trained algorithm will be well trained to predict dust quantities from the real observations that are more complex than symmetric dusty shells. However, in the case of optically thick clumps that would absorb any UV and optical radiation, the feasibility of estimating actual dust mass will still be under debate.

Moreover, hyperparameter optimization for each step in the feature selection process may increase the prediction accuracy and let the number of filters in the minimum set of JWST filters for predicting quantities of dust decrease. This lets us investigate mechanism of dust formation in supernovae more efficiently.

4.3 Postface

In the realm of large volumes of observations, employing machine learning algorithms is key to accurately estimate properties and quantities of astrophysical systems. Especially that recent improvements in computing powers let us implement machine learning algorithms more efficiently. As advanced upcoming ground-based and space-based telescopes such as the LSST and JWST are designed to bring larger volumes of observations than the current amount, getting ready to quickly and accurately extract information from them is vital. Deriving various properties from large samples let us understand the universe statistically. Machine learning methods that show promising outcomes in the sense of accuracy and time consumption are becoming a number one choice of selection for studying the evolution of the universe. This is also true for understanding astrophysical phenomena such as forming dust in supernovae that let us understand the evolution of cosmic dust.

At the same time, it is crucial to continue using one-object oriented observations to extract physics behind objects that we have not fully understood, either due to a lack of sufficient observations or confronting new phenomena that we have not observed yet.

Bibliography

5

Abbott, B. P., R. Abbott, T. D. Abbott, *et al.* (Nov. 2017). „A gravitational-wave standard siren measurement of the Hubble constant“. In: 551.7678, pp. 85–88. arXiv: `1710.05835 [astro-ph.CO]`.

Abbott, D. C. and L. B. Lucy (Jan. 1985). „Multiline transfer and the dynamics of stellar winds.“ In: 288, pp. 679–693.

Abbott, T. M. C., F. B. Abdalla, S. Allam, *et al.* (Dec. 2018). „The Dark Energy Survey: Data Release 1“. In: 239.2, 18, p. 18. arXiv: `1801.03181 [astro-ph.IM]`.

Acquaviva, Viviana (Feb. 2016). „How to measure metallicity from five-band photometry with supervised machine learning algorithms“. In: 456.2, pp. 1618–1626. arXiv: `1510.08076 [astro-ph.IM]`.

Acquaviva, Viviana, Chistopher Lovell, and Emille Ishida (Nov. 2020). „Debunking Generalization Error or: How I Learned to Stop Worrying and Love My Training Set“. In: *arXiv e-prints*, arXiv:2012.00066, arXiv:2012.00066. arXiv: `2012.00066 [astro-ph.IM]`.

Agnello, Adriano (Oct. 2017). „Quasar lenses and galactic streams: outlier selection and Gaia multiplet detection“. In: 471.2, pp. 2013–2021. arXiv: `1705.08900 [astro-ph.GA]`.

Aguado, D. S., Romina Ahumada, Andrés Almeida, *et al.* (Feb. 2019). „The Fifteenth Data Release of the Sloan Digital Sky Surveys: First Release of MaNGA-derived Quantities, Data Visualization Tools, and Stellar Library“. In: 240.2, 23, p. 23. arXiv: `1812.02759 [astro-ph.IM]`.

Ahn, Christopher P., Rachael Alexandroff, Carlos Allende Prieto, *et al.* (Dec. 2012). „The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey“. In: 203.2, 21, p. 21. arXiv: `1207.7137 [astro-ph.IM]`.

Almosallam, Ibrahim A., Sam N. Lindsay, Matt J. Jarvis, and Stephen J. Roberts (Jan. 2016). „A sparse Gaussian process framework for photometric redshift estimation“. In: 455.3, pp. 2387–2401. arXiv: `1505.05489 [astro-ph.IM]`.

Altman, Douglas G and J Martin Bland (2005). „Standard deviations and standard errors“. In: *BMJ* 331.7521, p. 903. eprint: `https://www.bmj.com/content/331/7521/903.full.pdf`.

Amaro, V., S. Cavuoti, M. Brescia, *et al.* (Jan. 2019). „Statistical analysis of probability density functions for photometric redshifts through the KiDS-ESO-DR3 galaxies“. In: 482.3, pp. 3116–3134. arXiv: `1810.09777 [astro-ph.IM]`.

Amiaux, J., R. Scaramella, Y. Mellier, *et al.* (Sept. 2012). „Euclid mission: building of a reference survey“. In: *Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*. Vol. 8442. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 84420Z. arXiv: `1209.2228 [astro-ph.IM]`.

Arendt, Richard G., Eli Dwek, Gladys Kober, Jeonghee Rho, and Una Hwang (Apr. 2014). „INTERSTELLAR AND EJECTA DUST IN THE CAS A SUPERNOVA REMNANT“. In: *The Astrophysical Journal* 786.1, p. 55.

Arnouts, Stephane, Stefano Cristiani, Lauro Moscardini, Sabino Matarrese, Francesco Lucchin, Adriano Fontana, and Emanuele Giallongo (1999). „Measuring and modeling the redshift evolution of clustering: The Hubble Deep Field North“. In: *Mon. Not. Roy. Astron. Soc.* 310, pp. 540–556. arXiv: `astro-ph/9902290`.

Barlow, M. J., B. E. K. Sugerman, J. Fabbri, *et al.* (June 2005). „Detection of a Massive Dust Shell around the Type II Supernova SN 2002hh“. In: *The Astrophysical Journal* 627.2, pp. L113–L116.

Baumann, Daniel (July 2009). „TASI Lectures on Inflation“. In: *arXiv e-prints*, arXiv:0907.5424, arXiv:0907.5424. arXiv: `0907.5424 [hep-th]`.

Beck, Róbert, István Szapudi, Heather Flewelling, Conrad Holmberg, Eugene Magnier, and Kenneth C. Chambers (Aug. 2020a). „PS1-STRM: Neural network source classification and photometric redshift catalogue for PS1 3π DR1“. In: arXiv: `1910.10167 [astro-ph.GA]`.

Beck, Róbert, István Szapudi, Heather Flewelling, Conrad Holmberg, Eugene Magnier, and Kenneth C. Chambers (Aug. 2020b). „PS1-STRM: Neural network source classification and photometric redshift catalogue for PS1 3π DR1“. In: arXiv: `1910.10167 [astro-ph.GA]`.

Beelen, Alexandre, Pierre Cox, Dominic J. Benford, C. Darren Dowell, Attila Kovács, Frank Bertoldi, Alain Omont, and Chris L. Carilli (May 2006). „350 μm Dust Emission from High-Redshift Quasars“. In: 642.2, pp. 694–701. arXiv: `astro-ph/0603121 [astro-ph]`.

Behroozi, Peter, Risa H. Wechsler, Andrew P. Hearin, and Charlie Conroy (Sept. 2019). „UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from z = 0-10“. In: 488.3, pp. 3143–3194. arXiv: `1806.07893 [astro-ph.GA]`.

Bellm, E. (Jan. 2014). „The Zwicky Transient Facility“. In: *The Third Hot-wiring the Transient Universe Workshop*. Ed. by P. R. Wozniak, M. J. Graham, A. A. Mahabal, and R. Seaman, pp. 27–33. arXiv: `1410.8185 [astro-ph.IM]`.

Bengio, Yoshua (2012). „Practical Recommendations for Gradient-Based Training of Deep Architectures“. In: *Neural Networks: Tricks of the Trade: Second Edition*. Ed. by Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 437–478.

Benıtez, Narciso (June 2000). „Bayesian Photometric Redshift Estimation“. In: 536.2, pp. 571–583. arXiv: `astro-ph/9811189 [astro-ph]`.

Bertoldi, F., C. L. Carilli, P. Cox, X. Fan, M. A. Strauss, A. Beelen, A. Omont, and R. Zylka (July 2003). „Dust emission from the most distant quasars“. In: 406, pp. L55–L58. eprint: `astro-ph/0305116`.

Bevan, A. M., K. Krafton, R. Wesson, J. E. Andrews, E. Montiel, M. Niculescu-Duvaz, M. J. Barlow, I. De Looze, and G. C. Clayton (May 2020). „Disentangling Dust Components in SN 2010jl: The First 1400 Days“. In: 894.2, 111, p. 111. arXiv: `2004.01503 [astro-ph.SR]`.

Bevan, A. and M. J. Barlow (Feb. 2016). „Modelling supernova line profile asymmetries to determine ejecta dust masses: SN 1987A from days 714 to 3604“. In: 456, pp. 1269–1293. arXiv: `1509.00858 [astro-ph.SR]`.

Bevan, Antonia and M. J. Barlow (Dec. 2015). „Modelling supernova line profile asymmetries to determine ejecta dust masses: SN 1987A from days 714 to 3604“. In: *Monthly Notices of the Royal Astronomical Society* 456.2, pp. 1269–1293. eprint: `https://academic.oup.com/mnras/article-pdf/456/2/1269/18510750/stv2651.pdf`.

Bevan, Antonia, M. J. Barlow, and D. Milisavljevic (Nov. 2016). „Dust masses for SN 1980K, SN1993J and Cassiopeia A from red–blue emission line asymmetries“. In: *Monthly Notices of the Royal Astronomical Society* 465.4, pp. 4044–4056. eprint: `https://academic.oup.com/mnras/article-pdf/465/4/4044/10254692/stw2985.pdf`.

Bhojanapalli, Srinadh, Kimberly Wilber, Andreas Veit, Ankit Singh Rawat, Seungyeon Kim, Aditya Menon, and Sanjiv Kumar (Feb. 2021). „On the Reproducibility of Neural Network Predictions“. In: *arXiv e-prints*, arXiv:2102.03349, arXiv:2102.03349. arXiv: `2102.03349 [cs.LG]`.

Bianchi, Simone and Raffaella Schneider (July 2007). „Dust formation and survival in supernova ejecta“. In: 378.3, pp. 973–982. arXiv: `0704.0586 [astro-ph]`.

Biscaro, Chiara and Isabelle Cherchneff (Apr. 2014). „Molecules and dust in Cassiopeia A. I. Synthesis in the supernova phase and processing by the reverse shock in the clumpy remnant“. In: 564, A25, A25. arXiv: `1401.5594 [astro-ph.SR]`.

Bishop, Chris M. (1995). „Training with Noise is Equivalent to Tikhonov Regularization“. In: *Neural Computation* 7.1, pp. 108–116.

Bishop, Christopher M. (1994). „Mixture density networks“. Technical Report. Birmingham.

Bloom, J. S., J. W. Richards, P. E. Nugent, *et al.* (Nov. 2012). „Automating Discovery and Classification of Transients and Variable Stars in the Synoptic Survey Era“. In: 124.921, p. 1175. arXiv: `1106.5491 [astro-ph.IM]`.

Bolzonella, M., J. -M. Miralles, and R. Pelló (Nov. 2000). „Photometric redshifts based on standard SED fitting procedures“. In: 363, pp. 476–492. arXiv: `astro-ph/0003380 [astro-ph]`.

Boyer, M. L., S. Srinivasan, D. Riebel, I. McDonald, J. Th. van Loon, G. C. Clayton, K. D. Gordon, M. Meixner, B. A. Sargent, and G. C. Sloan (Mar. 2012a). „THE DUST BUDGET OF THE SMALL MAGELLANIC CLOUD: ARE ASYMPTOTIC GIANT BRANCH STARS THE PRIMARY DUST SOURCE AT LOW METALLICITY?“ In: *The Astrophysical Journal* 748.1, p. 40.

Boyer, M. L., S. Srinivasan, D. Riebel, I. McDonald, J. T. van Loon, G. C. Clayton, K. D. Gordon, M. Meixner, B. A. Sargent, and G. C. Sloan (Mar. 2012b). „The Dust Budget of the Small Magellanic Cloud: Are Asymptotic Giant Branch Stars the Primary Dust Source at Low Metallicity?“ In: 748, 40, p. 40. arXiv: `1201.5384 [astro-ph.SR]`.

Brammer, Gabriel B., Pieter G. van Dokkum, and Paolo Coppi (Oct. 2008). „EAZY: A Fast, Public Photometric Redshift Code“. In: 686.2, pp. 1503–1513. arXiv: `0807.1533 [astro-ph]`.

Brammer, Gabriel B., Pieter G. van Dokkum, and Paolo Coppi (Oct. 2010). *EAZY: A Fast, Public Photometric Redshift Code*. ascl: `1010.052`.

Brescia, M., S. Cavuoti, R. D’Abrusco, G. Longo, and A. Mercurio (Aug. 2013). „Photometric Redshifts for Quasars in Multi-band Surveys“. In: 772.2, 140, p. 140. arXiv: `1305.5641 [astro-ph.IM]`.

Campello, Ricardo J. G. B., Davoud Moulavi, and Joerg Sander (2013). „Density-Based Clustering Based on Hierarchical Density Estimates“. In: *Advances in Knowledge Discovery and Data Mining*. Ed. by Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, and Guandong Xu. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 160–172.

Carilli, C. L., P. Cox, F. Bertoldi, K. M. Menten, A. Omont, S. G. Djorgovski, A. Petric, A. Beelen, K. G. Isaak, and R. G. McMahon (Aug. 2002). „Imaging Low-Order CO Emission from the z=4.12 Quasi-Stellar Object PSS J2322+1944“. In: 575.1, pp. 145–149. arXiv: `astro-ph/0204253 [astro-ph]`.

Carrasco Kind, Matias and Robert J. Brunner (June 2013). „TPZ: photometric redshift PDFs and ancillary information by using prediction trees and random forests“. In: 432.2, pp. 1483–1501. arXiv: `1303.7269 [astro-ph.CO]`.

Castro, Nicolás, Pavlos Protopapas, and Karim Pichara (Jan. 2018). „Uncertain Classification of Variable Stars: Handling Observational GAPS and Noise“. In: 155.1, 16, p. 16. arXiv: `1801.09732 [astro-ph.IM]`.

Cernuschi, F., F. Marsicano, and S. Codina (Feb. 1967). „Contribution to the theory on the formation of cosmic grains“. In: *Annales d'Astrophysique* 30, p. 1039.

Cervantes-Cota, Jorge L. and George Smoot (Oct. 2011). „Cosmology today–A brief review“. In: *VIII Workshop of the Gravitation and Mathematical Physics Division of the Mexican Physical Society*. Ed. by Luis Arturo Ureña-López, Hugo Aurelio Morales-Técotl, Román Linares-Romero, Elı Santos-Rodrıguez, and Sendic Estrada-Jiménez. Vol. 1396. American Institute of Physics Conference Series, pp. 28–52. arXiv: `1107.1789 [astro-ph.CO]`.

Chawner, H., K. Marsh, M. Matsuura, H. L. Gomez, P. Cigan, I. De Looze, M. J. Barlow, L. Dunne, A. Noriega-Crespo, and J. Rho (Feb. 2019). „A catalogue of Galactic supernova remnants in the far-infrared: revealing ejecta dust in pulsar wind nebulae“. In: 483.1, pp. 70–118. arXiv: `1811.00034 [astro-ph.GA]`.

Chen, T. -W., S. J. Brennan, R. Wesson, *et al.* (Sept. 2021). „SN 2018bsz: significant dust formation in a nearby superluminous supernova“. In: *arXiv e-prints*, arXiv:2109.07942, arXiv:2109.07942. arXiv: `2109.07942 [astro-ph.HE]`.

Clevert, Djork-Arné, Thomas Unterthiner, and Sepp Hochreiter (Nov. 2015). „Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)“. In: *arXiv e-prints*, arXiv:1511.07289, arXiv:1511.07289. arXiv: `1511.07289 [cs.LG]`.

Cohen, Judith G., David W. Hogg, Roger Blandford, Lennox L. Cowie, Esther Hu, Antoinette Songaila, Patrick Shopbell, and Kevin Richberg (July 2000). „Caltech Faint Galaxy Redshift Survey. X. A Redshift Survey in the Region of the Hubble Deep Field North“. In: 538.1, pp. 29–52. arXiv: `astro-ph/9912048 [astro-ph]`.

Coil, Alison L., Michael R. Blanton, Scott M. Burles, *et al.* (Nov. 2011). „The PRIsm MUlti-object Survey (PRIMUS). I. Survey Overview and Characteristics“. In: 741.1, 8, p. 8. arXiv: `1011.4307 [astro-ph.CO]`.

Collister, Adrian A. and Ofer Lahav (Apr. 2004). „ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks“. In: 116.818, pp. 345–351. arXiv: `astro-ph/0311058 [astro-ph]`.

Connolly, A. J., A. S. Szalay, M. A. Bershady, A. L. Kinney, and D. Calzetti (Sept. 1995). „Spectral Classification of Galaxies: an Orthogonal Approach“. In: 110, p. 1071. arXiv: `astro-ph/9411044 [astro-ph]`.

Cranmer, Miles D., Rui Xu, Peter Battaglia, and Shirley Ho (Sept. 2019). „Learning Symbolic Physics with Graph Networks“. In: *arXiv e-prints*, arXiv:1909.05862, arXiv:1909.05862. arXiv: `1909.05862 [cs.LG]`.

Curran, S. J. (Mar. 2020). „QSO photometric redshifts from SDSS, WISE, and GALEX colours“. In: 493.1, pp. L70–L75. arXiv: `2001.06514 [astro-ph.IM]`.

D'Abrusco, R., G. Fabbiano, G. Djorgovski, C. Donalek, O. Laurino, and G. Longo (Aug. 2012). „CLaSPS: A New Methodology for Knowledge Extraction from Complex Astronomical Data Sets“. In: 755.2, 92, p. 92. arXiv: `1206.2919 [astro-ph.IM]`.

D'Isanto, A., S. Cavuoti, M. Brescia, C. Donalek, G. Longo, G. Riccio, and S. G. Djorgovski (Feb. 2016). „An analysis of feature relevance in the classification of astronomical transients with machine learning methods“. In: *Monthly Notices of the Royal Astronomical Society* 457.3, pp. 3119–3132. eprint: `https://academic.oup.com/mnras/article-pdf/457/3/3119/8001547/stw157.pdf`.

Daniel, Scott F., Andrew J. Connolly, Jeff Schneider, Jake Vanderplas, and Liang Xiong (Oct. 2011). „Classification of Stellar Spectra with LLE“. In: *arXiv e-prints*, arXiv:1110.4646, arXiv:1110.4646. arXiv: `1110.4646 [astro-ph.SR]`.

de Jong, Jelte T. A., Gijs A. Verdoes Kleijn, Konrad H. Kuijken, and Edwin A. Valentijn (Jan. 2013). „The Kilo-Degree Survey“. In: *Experimental Astronomy* 35.1-2, pp. 25–44. arXiv: `1206.1254 [astro-ph.CO]`.

De Looze, I., M. J. Barlow, R. Bandiera, A. Bevan, M. F. Bietenholz, H. Chawner, H. L. Gomez, M. Matsuura, F. Priestley, and R. Wesson (Sept. 2019). „The dust content of the Crab Nebula“. In: 488.1, pp. 164–182. arXiv: `1906.02203 [astro-ph.HE]`.

De Looze, I., M. J. Barlow, B. M. Swinyard, J. Rho, H. L. Gomez, M. Matsuura, and R. Wesson (Nov. 2016). „The dust mass in Cassiopeia A from a spatially resolved Herschel analysis“. In: *Monthly Notices of the Royal Astronomical Society* 465.3, pp. 3309–3342. eprint: `https://academic.oup.com/mnras/article-pdf/465/3/3309/8541915/stw2837.pdf`.

De Looze, I., I. Lamperti, A. Saintonge, *et al.* (Aug. 2020). „JINGLE - IV. Dust, H I gas, and metal scaling laws in the local Universe“. In: 496.3, pp. 3668–3687. arXiv: `2006.01856 [astro-ph.GA]`.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). „Maximum Likelihood from Incomplete Data Via the EM Algorithm“. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1, pp. 1–22. eprint: `https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1977.tb01600.x`.

Deng, Li (2012). „The mnist database of handwritten digit images for machine learning research“. In: *IEEE Signal Processing Magazine* 29.6, pp. 141–142.

Djorgovski, S. G., A. A. Mahabal, C. Donalek, M. J. Graham, A. J. Drake, B. Moghaddam, and M. Turmon (Sept. 2012). „Flashes in a Star Stream: Automated Classification of Astronomical Transient Events“. In: *arXiv e-prints*, arXiv:1209.1681, arXiv:1209.1681. arXiv: `1209.1681 [astro-ph.IM]`.

Draine, B. T. (Dec. 2009). „Interstellar Dust Models and Evolutionary Implications“. In: *Cosmic Dust - Near and Far*. Ed. by T. Henning, E. Grün, and J. Steinacker. Vol. 414. Astronomical Society of the Pacific Conference Series, p. 453. arXiv: `0903.1658 [astro-ph.GA]`.

Draine, B. T. and H. M. Lee (Oct. 1984). „Optical Properties of Interstellar Graphite and Silicate Grains“. In: 285, p. 89.

Drinkwater, Michael J., Russell J. Jurek, Chris Blake, *et al.* (Jan. 2010). „The WiggleZ Dark Energy Survey: survey design and first data release“. In: 401.3, pp. 1429–1452. arXiv: `0911.4246 [astro-ph.CO]`.

Driver, S. P., D. T. Hill, L. S. Kelvin, *et al.* (May 2011). „Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release“. In: 413.2, pp. 971–995. arXiv: `1009.0614 [astro-ph.CO]`.

Dunkley, J., E. Komatsu, M. R. Nolta, *et al.* (Feb. 2009). „Five-Year Wilkinson Microwave Anisotropy Probe Observations: Likelihoods and Parameters from the WMAP Data“. In: 180.2, pp. 306–329. arXiv: `0803.0586 [astro-ph]`.

Dwek, E., F. Galliano, and A. P. Jones (June 2007). „The Evolution of Dust in the Early Universe with Applications to the Galaxy SDSS J1148+5251“. In: 662, pp. 927–939. arXiv: `0705.3799`.

Dwek, E. and J. M. Scalo (July 1980). „The evolution of refractory interstellar grains in the solar neighborhood“. In: 239, pp. 193–211.

Dwek, Eli and Isabelle Cherchneff (Feb. 2011). „The Origin of Dust in the Early Universe: Probing the Star Formation History of Galaxies by Their Dust Content“. In: 727.2, 63, p. 63. arXiv: `1011.1303 [astro-ph.CO]`.

Dwek, Eli, Johannes Staguhn, Richard G. Arendt, Attila Kovacks, Ting Su, and Dominic J. Benford (June 2014). „DUST FORMATION, EVOLUTION, AND OBSCURATION EFFECTS IN THE VERY HIGH-REDSHIFT UNIVERSE“. In: *The Astrophysical Journal* 788.2, p. L30.

Ercolano, B., M. J. Barlow, and P. J. Storey (Sept. 2005). „The dusty MOCASSIN: fully self-consistent 3D photoionization and dust radiative transfer models“. In: 362.3, pp. 1038–1046. arXiv: `astro-ph/0507050 [astro-ph]`.

Ercolano, B., M. J. Barlow, P. J. Storey, and X. -W. Liu (Apr. 2003a). „MOCASSIN: a fully three-dimensional Monte Carlo photoionization code“. In: 340.4, pp. 1136–1152. arXiv: `astro-ph/0209378 [astro-ph]`.

Ercolano, B., M. J. Barlow, and B. E. K. Sugerman (Mar. 2007). „Dust yields in clumpy supernova shells: SN 1987A revisited“. In: 375.3, pp. 753–763. arXiv: `astro-ph/0611719 [astro-ph]`.

Ercolano, B., C. Morisset, M. J. Barlow, P. J. Storey, and X. -W. Liu (Apr. 2003b). „Three-dimensional photoionization modelling of the planetary nebula NGC 3918“. In: 340.4, pp. 1153–1172. arXiv: `astro-ph/0209417 [astro-ph]`.

Fawcett, Tom (2006). „An introduction to ROC analysis.“ In: *Pattern Recognit. Lett.* 27.8, pp. 861–874.

Ferguson, Thomas S. (Mar. 1973). „A Bayesian Analysis of Some Nonparametric Problems“. In: *Ann. Statist.* 1.2, pp. 209–230.

Fernández-Soto, Alberto, Kenneth M. Lanzetta, and Amos Yahil (Mar. 1999). „A New Catalog of Photometric Redshifts in the Hubble Deep Field“. In: 513.1, pp. 34–50. arXiv: `astro-ph/9809126 [astro-ph]`.

Ferrara, A., S. Viti, and C. Ceccarelli (Nov. 2016). „The problematic growth of dust in high-redshift galaxies“. In: 463, pp. L112–L116. arXiv: `1606.07214`.

Fesen, Robert A., Andrew J. S. Hamilton, and Jon M. Saken (June 1989). „Discovery of the Remnant of S Andromedae (SN 1885) in M31“. In: 341, p. L55.

Filippenko, Alexei V. (Jan. 1997). „Optical Spectra of Supernovae“. In: 35, pp. 309–355.

Finkelstein, S. L., C. Papovich, B. Salmon, *et al.* (Sept. 2012). „Candels: The Evolution of Galaxy Rest-frame Ultraviolet Colors from z = 8 to 4“. In: 756, 164, p. 164. arXiv: `1110.3785`.

Friedmann, A. (Jan. 1922). „Über die Krümmung des Raumes“. In: *Zeitschrift fur Physik* 10, pp. 377–386.

Galametz, Audrey, Roberto Saglia, Stéphane Paltani, Nikolaos Apostolakos, and Pierre Dubath (Feb. 2017). „SED-dependent galactic extinction prescription for Euclid and future cosmological surveys“. In: 598, A20, A20. arXiv: `1609.08624 [astro-ph.CO]`.

Gall, C., A. C. Andersen, and J. Hjorth (Apr. 2011a). „Genesis and evolution of dust in galaxies in the early Universe. II. Rapid dust evolution in quasars at z 6“. In: 528, A14, A14. arXiv: `1101.1553`.

Gall, C. and J. Hjorth (Nov. 2018). „Maximally Dusty Star-forming Galaxies: Supernova Dust Production and Recycling in Local Group and High-redshift Galaxies“. In: 868.1, 62, p. 62. arXiv: `1809.11032 [astro-ph.GA]`.

Gall, C., J. Hjorth, and A. C. Andersen (Sept. 2011b). „Production of dust by massive stars at high redshift“. In: 19, 43, p. 43. arXiv: `1108.0403 [astro-ph.CO]`.

Gall, C., J. Hjorth, D. Watson, E. Dwek, J. R. Maund, O. Fox, G. Leloudas, D. Malesani, and A. C. Day-Jones (July 2014). „Rapid formation of large dust grains in the luminous supernova 2010jl“. In: 511, pp. 326–329. arXiv: `1407.4447 [astro-ph.SR]`.

Gardner, Jonathan P., John C. Mather, Mark Clampin, *et al.* (Apr. 2006). „The James Webb Space Telescope“. In: 123.4, pp. 485–606. arXiv: `astro-ph/0606175 [astro-ph]`.

Garilli, B., O. Le Fèvre, L. Guzzo, *et al.* (Aug. 2008). „The Vimos VLT deep survey. Global properties of 20,000 galaxies in the I_{AB} ¡ 22.5 WIDE survey“. In: 486.3, pp. 683–695. arXiv: `0804.4568 [astro-ph]`.

Gehrz, R. (Jan. 1989). „Sources of Stardust in the Galaxy“. In: *Interstellar Dust*. Ed. by Louis J. Allamandola and A. G. G. M. Tielens. Vol. 135, p. 445.

Gerdes, David W., Adam J. Sypniewski, Timothy A. McKay, Jiangang Hao, Matthew R. Weis, Risa H. Wechsler, and Michael T. Busha (June 2010). „ArborZ: Photometric Redshifts Using Boosted Decision Trees“. In: 715.2, pp. 823–832. arXiv: `0908.4085 [astro-ph.CO]`.

Ghavamian, Parviz and Brian J. Williams (Nov. 2016). „THE COLD DUST CONTENT OF THE OXYGEN-RICH SUPERNOVA REMNANT G292.01.8“. In: *The Astrophysical Journal* 831.2, p. 188.

Gomez, H. L., L. Dunne, R. J. Ivison, E. M. Reynoso, M. A. Thompson, B. Sibthorpe, S. A. Eales, T. M. DeLaney, S. Maddox, and K. Isaak (July 2009). „Accounting for the foreground contribution to the dust emission towards Kepler's supernova remnant“. In: *Monthly Notices of the Royal Astronomical Society* 397.3, pp. 1621–1632. eprint: `https://academic.oup.com/mnras/article-pdf/397/3/1621/3755157/mnras0397-1621.pdf`.

Gomez, H. L., O. Krause, M. J. Barlow, *et al.* (Nov. 2012a). „A Cool Dust Factory in the Crab Nebula: A Herschel Study of the Filaments“. In: 760.1, 96, p. 96. arXiv: `1209.5677 [astro-ph.GA]`.

Gomez, H. L., O. Krause, M. J. Barlow, *et al.* (Nov. 2012b). „A COOL DUST FACTORY IN THE CRAB NEBULA: AHERSCHELSTUDY OF THE FILAMENTS“. In: *The Astrophysical Journal* 760.1, p. 96.

Görür, Dilan and Carl Edward Rasmussen (2010). „Dirichlet Process Gaussian Mixture Models: Choice of the Base Distribution“. In: *Journal of Computer Science and Technology* 25.4, pp. 653–664.

Graham, Melissa L., Andrew J. Connolly, Željko Ivezić, Samuel J. Schmidt, R. Lynne Jones, Mario Jurić, Scott F. Daniel, and Peter Yoachim (Jan. 2018). „Photometric Redshifts with the LSST: Evaluating Survey Observing Strategies“. In: 155.1, 1, p. 1. arXiv: `1706.09507 [astro-ph.CO]`.

Greene, Thomas P., Douglas M. Kelly, John Stansberry, Jarron M. Leisenring, Everett A. Schlawin, Eiichi Egami, Laurie Chu, Klaus W. Hodapp, and Marcia J. Rieke (2017). „=2.4 to 5m spectroscopy with the James Webb Space Telescope NIRCam instrument“. In: *Journal of Astronomical Telescopes, Instruments, and Systems* 3.3, pp. 1–10.

Guzzo, L., M. Scodeggio, B. Garilli, *et al.* (June 2014). „The VIMOS Public Extragalactic Redshift Survey (VIPERS). An unprecedented view of galaxies and large-scale structure at 0.5 ¡ z ¡ 1.2“. In: 566, A108, A108. arXiv: `1303.2623 [astro-ph.CO]`.

Han, Bo, Hongpeng Ding, Yanxia Zhang, and Yongheng Zhao (Oct. 2016). „Improving Accuracy of Quasars’ Photometric Redshift Estimation by Integration of KNN and SVM“. In: *IAU Focus Meeting* 29A, pp. 209–209.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (Feb. 2015). „Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification“. In: *arXiv e-prints*, arXiv:1502.01852, arXiv:1502.01852. arXiv: `1502.01852 [cs.CV]`.

Henning, Thomas (Sept. 2010). „Cosmic Silicates“. In: 48, pp. 21–46.

Hildebrand, R. H. (Sept. 1983). „The determination of cloud masses and dust characteristics from submillimetre thermal emission.“ In: 24, pp. 267–282.

Hildebrandt, H., F. Köhlinger, J. L. van den Busch, *et al.* (Jan. 2020). „KiDS+VIKING-450: Cosmic shear tomography with optical and infrared data“. In: 633, A69, A69. arXiv: `1812.06076 [astro-ph.CO]`.

Hirashita, H. (May 2012). „Dust growth in the interstellar medium: how do accretion and coagulation interplay?“ In: 422, pp. 1263–1271. arXiv: `1202.1345`.

Hogg, David W., Ivan K. Baldry, Michael R. Blanton, and Daniel J. Eisenstein (Oct. 2002). „The K correction“. In: *arXiv e-prints*, astro-ph/0210394, astro–ph/0210394. arXiv: `astro-ph/0210394 [astro-ph]`.

Horner, Scott D. and Marcia J. Rieke (2004). „The near-infrared camera (NIRCam) for the James Webb Space Telescope (JWST)“. In: *Optical, Infrared, and Millimeter Space Telescopes*. Ed. by John C. Mather. Vol. 5487. International Society for Optics and Photonics. SPIE, pp. 628–634.

Hornik, K. (1991). „Approximation capabilities of multilayer feedforward networks“. In: *Neural Networks* 4, pp. 251–257.

Hoyle, B., D. Gruen, G. M. Bernstein, *et al.* (July 2018). „Dark Energy Survey Year 1 Results: redshift distributions of the weak-lensing source galaxies“. In: 478.1, pp. 592–610. arXiv: `1708.01532 [astro-ph.CO]`.

Ilbert, O., S. Arnouts, H. J. McCracken, *et al.* (Oct. 2006). „Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey“. In: 457.3, pp. 841–856. arXiv: `astro-ph/0603217 [astro-ph]`.

Indebetouw, R., M. Matsuura, E. Dwek, *et al.* (Jan. 2014a). „DUST PRODUCTION AND PARTICLE ACCELERATION IN SUPERNOVA 1987A REVEALED WITH ALMA“. In: *The Astrophysical Journal* 782.1, p. L2.

Indebetouw, R., M. Matsuura, E. Dwek, *et al.* (Feb. 2014b). „Dust Production and Particle Acceleration in Supernova 1987A Revealed with ALMA“. In: 782.1, L2, p. L2. arXiv: `1312.4086 [astro-ph.SR]`.

Ishida, E. E. O., R. Beck, S. González-Gaitán, *et al.* (Feb. 2019). „Optimizing spectroscopic follow-up strategies for supernova photometric classification with active learning“. In: 483.1, pp. 2–18. arXiv: `1804.03765 [astro-ph.IM]`.

Ivezić, Ž., S. M. Kahn, J. A. Tyson, *et al.* (Mar. 2019). „LSST: From Science Drivers to Reference Design and Anticipated Data Products“. In: 873, 111, p. 111. arXiv: `0805.2366`.

Jones, D. O., R. J. Foley, G. Narayan, *et al.* (Feb. 2021). „The Young Supernova Experiment: Survey Goals, Overview, and Operations“. In: 908.2, 143, p. 143. arXiv: `2010.09724 [astro-ph.HE]`.

Joudaki, Shahab, Chris Blake, Andrew Johnson, *et al.* (Mar. 2018). „KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering“. In: 474.4, pp. 4894–4924. arXiv: `1707.06627 [astro-ph.CO]`.

Kingma, Diederik P. and Jimmy Ba (Dec. 2014). „Adam: A Method for Stochastic Optimization“. In: *arXiv e-prints*, arXiv:1412.6980, arXiv:1412.6980. arXiv: `1412.6980 [cs.LG]`.

Knox, Lloyd, Yong-Seon Song, and Hu Zhan (Dec. 2006). „Weighing the Universe with Photometric Redshift Surveys and the Impact on Dark Energy Forecasts“. In: 652.2, pp. 857–863. arXiv: `astro-ph/0605536 [astro-ph]`.

Kochanek, C. S. (Dec. 2011). „The Astrophysical Implications of Dust Formation during the Eruptions of Hot, Massive Stars“. In: 743.1, 73, p. 73. arXiv: `1109.2596 [astro-ph.SR]`.

Kotak, R., W. P. S. Meikle, D. Farrah, *et al.* (Oct. 2009). „Dust and The Type II-Plateau Supernova 2004et“. In: 704, pp. 306–323. arXiv: `0904.3737 [astro-ph.SR]`.

Kozasa, T, H Hasegawa, and K Nomoto (Nov. 1989). „Thermal radiation from dust grains formed in the ejecta of SN 1987A“. In: *Astrophysical Journal; (USA)* 346.

Lang, Dustin (May 2014). „unWISE: Unblurred Coadds of the WISE Imaging“. In: 147.5, 108, p. 108. arXiv: `1405.0308 [astro-ph.IM]`.

Lang, Dustin, David W. Hogg, and David J. Schlegel (Oct. 2014). „WISE photometry for 400 million SDSS sources“. In: *arXiv e-prints*, arXiv:1410.7397, arXiv:1410.7397. arXiv: `1410.7397 [astro-ph.IM]`.

Lau, R. M., T. L. Herter, M. R. Morris, Z. Li, and J. D. Adams (Apr. 2015). „Old supernova dust factory revealed at the Galactic center“. In: *Science* 348.6233, pp. 413–418. arXiv: `1503.07173 [astro-ph.SR]`.

Lawrence, A. (May 2012). „The UV peak in active galactic nuclei: a false continuum from blurred reflection?“ In: *Monthly Notices of the Royal Astronomical Society* 423.1, pp. 451–463. eprint: `https://academic.oup.com/mnras/article-pdf/423/1/451/18604923/mnras0423-0451.pdf`.

LeCun, Yann, Leon Bottou, Genevieve B. Orr, and Klaus -Robert Müller (1998). „Efficient BackProp“. In: *Neural Networks: Tricks of the Trade*. Ed. by Genevieve B. Orr and Klaus-Robert Müller. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 9–50.

Lemaıtre, G. (Mar. 1931). „Expansion of the universe, The expanding universe“. In: 91, pp. 490–501.

Lilly, S. and Zcosmos Team (Dec. 2008). „The zCOSMOS Data Release 2: the “zCOSMOS-bright 10k-sample” and structure in the Universe out to redshifts of order unity“. In: *The Messenger* 134, pp. 35–40.

Lochner, Michelle, Jason D. McEwen, Hiranya V. Peiris, Ofer Lahav, and Max K. Winter (Aug. 2016). „Photometric Supernova Classification with Machine Learning“. In: 225.2, 31, p. 31. arXiv: `1603.00882 [astro-ph.IM]`.

Logan, C. H. A. and Fotopoulou, S. (2020). „Unsupervised star, galaxy, QSO classification - Application of HDBSCAN“. In: *A&A* 633, A154.

Lucy, L. B. (May 1999). „Improved Monte Carlo techniques for the spectral synthesis of supernovae“. In: 345, pp. 211–220.

Lundberg, Scott and Su-In Lee (May 2017). „A Unified Approach to Interpreting Model Predictions“. In: *arXiv e-prints*, arXiv:1705.07874, arXiv:1705.07874. arXiv: `1705.07874 [cs.AI]`.

Maas, Andrew L., Awni Y. Hannun, and Andrew Y. Ng (2013). „Rectifier nonlinearities improve neural network acoustic models“. In: *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*.

Magnier, E. A., E. Schlafly, D. Finkbeiner, *et al.* (Apr. 2013). „The Pan-STARRS 1 Photometric Reference Ladder, Release 12.01“. In: 205.2, 20, p. 20. arXiv: `1303.3634 [astro-ph.IM]`.

Mahabal, A., S. G. Djorgovski, M. Turmon, J. Jewell, R. R. Williams, A. J. Drake, M. G. Graham, C. Donalek, E. Glikman, and Palomar-QUEST Team (Mar. 2008). „Automated probabilistic classification of transients and variables“. In: *Astronomische Nachrichten* 329.3, pp. 288–291. arXiv: `0802.3199 [astro-ph]`.

Mandelbaum, Rachel (Sept. 2018). „Weak Lensing for Precision Cosmology“. In: 56, pp. 393–433. arXiv: `1710.03235 [astro-ph.CO]`.

Marrone, D. P., J. S. Spilker, C. C. Hayward, *et al.* (Jan. 2018). „Galaxy growth in a massive halo in the first billion years of cosmic history“. In: 553, pp. 51–54. arXiv: `1712.03020`.

Mathis, J. S., W. Rumpl, and K. H. Nordsieck (Oct. 1977). „The size distribution of interstellar grains.“ In: 217, pp. 425–433.

Matsuura, M., M. J. Barlow, A. A. Zijlstra, *et al.* (June 2009). „The global gas and dust budget of the Large Magellanic Cloud: AGB stars and supernovae, and the impact on the ISM evolution“. In: 396, pp. 918–934. arXiv: `0903.1123`.

Matsuura, M., E. Dwek, M. J. Barlow, *et al.* (Feb. 2015). „A Stubbornly Large Mass of Cold Dust in the Ejecta of Supernova 1987A“. In: 800, 50, p. 50. arXiv: `1411.7381 [astro-ph.SR]`.

Matsuura, Mikako (2017). „Dust and Molecular Formation in Supernovae“. In: *Handbook of Supernovae*. Ed. by Athem W. Alsabti and Paul Murdin, p. 2125.

Matsuura, Mikako, James M. De Buizer, Richard G. Arendt, *et al.* (Jan. 2019). „SOFIA mid-infrared observations of Supernova 1987A in 2016 - forward shocks and possible dust re-formation in the post-shocked region“. In: 482.2, pp. 1715–1723. arXiv: `1810.03615 [astro-ph.SR]`.

Matsuura, Mikako, Paul M. Woods, and Patrick J. Owen (Mar. 2013). „The global gas and dust budget of the Small Magellanic Cloud“. In: 429.3, pp. 2527–2536. arXiv: `1212.1468 [astro-ph.GA]`.

Mauney, Christopher M. and Davide Lazzati (Sept. 2018). „The formation of astrophysical Mg-rich silicate dust“. In: *Molecular Astrophysics* 12, pp. 1–9. arXiv: `1803.04323 [astro-ph.SR]`.

Melchior, P. and A. D. Goulding (Oct. 2018). „Filling the gaps: Gaussian mixture models from noisy, truncated or incomplete samples“. In: *Astronomy and Computing* 25, pp. 183–194. arXiv: `1611.05806 [astro-ph.IM]`.

Michałowski, M. J., E. J. Murphy, J. Hjorth, D. Watson, C. Gall, and J. S. Dunlop (Nov. 2010a). „Dust grain growth in the interstellar medium of 5 z 6.5 quasars“. In: 522, A15, A15. arXiv: `1006.5466`.

Michałowski, M. J., D. Watson, and J. Hjorth (Apr. 2010b). „Rapid Dust Production in Submillimeter Galaxies at z 4?“ In: 712, pp. 942–950. arXiv: `1002.2636 [astro-ph.CO]`.

Molster, FJ, LBFM Waters, and F Kemper (2010). „The mineralogy of interstellar and circumstellar dust in galaxies“. In: *Astromineralogy*, pp. 143–201.

Murty, Katta G. and Santosh N. Kabadi (1987). „Some NP-complete problems in quadratic and nonlinear programming“. In: *Mathematical Programming* 39, pp. 117–129.

Nair, Vinod and Geoffrey E. Hinton (2010). „Rectified Linear Units Improve Restricted Boltzmann Machines“. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML'10. Haifa, Israel: Omnipress, pp. 807–814.

Nash, John (1953). „Two-Person Cooperative Games“. In: *Econometrica* 21.1, pp. 128–140.

Naul, Brett, Joshua S. Bloom, Fernando Pérez, and Stéfan van der Walt (Nov. 2018). „A recurrent neural network for classification of unevenly sampled variable stars“. In: *Nature Astronomy* 2, pp. 151–155. arXiv: `1711.10609 [astro-ph.IM]`.

Newman, Jeffrey A., Michael C. Cooper, Marc Davis, *et al.* (Sept. 2013). „The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts“. In: 208.1, 5, p. 5. arXiv: `1203.3192 [astro-ph.CO]`.

Niculescu-Duvaz, Maria, M. J. Barlow, A. Bevan, D. Milisavljevic, and I. De Looze (June 2021). „The dust mass in Cassiopeia A from infrared and optical line flux differences“. In: 504.2, pp. 2133–2145. arXiv: `2103.12705 [astro-ph.GA]`.

Nishizawa, Atsushi J., Bau-Ching Hsieh, Masayuki Tanaka, and Tadafumi Takata (Feb. 2020). „Photometric Redshifts for the Hyper Suprime-Cam Subaru Strategic Program Data Release 2“. In: *arXiv e-prints*, arXiv:2003.01511, arXiv:2003.01511. arXiv: `2003.01511 [astro-ph.GA]`.

Nozawa, Takaya, Takashi Kozasa, Nozomu Tominaga, Keiichi Maeda, Hideyuki Umeda, Ken'ichi Nomoto, and Oliver Krause (Apr. 2010). „Formation and Evolution of Dust in Type IIb Supernovae with Application to the Cassiopeia A Supernova Remnant“. In: 713.1, pp. 356–373. arXiv: `0909.4145 [astro-ph.SR]`.

Nozawa, Takaya, Takashi Kozasa, Hideyuki Umeda, Keiichi Maeda, and Ken'ichi Nomoto (Dec. 2003). „Dust in the Early Universe: Dust Formation in the Ejecta of Population III Supernovae“. In: 598.2, pp. 785–803. arXiv: `astro-ph/0307108 [astro-ph]`.

Nun, Isadora, Pavlos Protopapas, Brandon Sim, and Wesley Chen (Sept. 2016). „Ensemble Learning Method for Outlier Detection and its Application to Astronomical Light Curves“. In: 152.3, 71, p. 71.

Omont, A., A. Beelen, F. Bertoldi, P. Cox, C. L. Carilli, R. S. Priddey, R. G. McMahon, and K. G. Isaak (Feb. 2003). „A 1.2 mm MAMBO/IRAM-30 m study of dust emission from optically luminous z ~2 quasars“. In: 398, pp. 857–865. arXiv: `astro-ph/0211655 [astro-ph]`.

Otsuka, M., J. Th. van Loon, K. S. Long, *et al.* (July 2010). „Dust in the bright supernova remnant N49 in the LMC“. In: 518, L139, p. L139. arXiv: `1005.2787 [astro-ph.CO]`.

Owen, P. J. and M. J. Barlow (Mar. 2015). „THE DUST AND GAS CONTENT OF THE CRAB NEBULA“. In: *The Astrophysical Journal* 801.2, p. 141.

Pasquet, Johanna, E. Bertin, M. Treyer, S. Arnouts, and D. Fouchez (Jan. 2019). „Photometric redshifts from SDSS images using a convolutional neural network“. In: 621, A26, A26. arXiv: `1806.06607 [astro-ph.IM]`.

Pietrzyński, G., D. Graczyk, A. Gallenne, *et al.* (Mar. 2019). „A distance to the Large Magellanic Cloud that is precise to one per cent“. In: 567.7747, pp. 200–203. arXiv: `1903.08096 [astro-ph.GA]`.

Planck Collaboration, R. Adam, N. Aghanim, *et al.* (Dec. 2016). „Planck intermediate results. XLVII. Planck constraints on reionization history“. In: 596, A108, A108. arXiv: `1605.03507 [astro-ph.CO]`.

Pontoppidan, Klaus M., Timothy E. Pickering, Victoria G. Laidler, *et al.* (July 2016). „Pandeia: a multi-mission exposure time calculator for JWST and WFIRST“. In: *Observatory Operations: Strategies, Processes, and Systems VI*. Ed. by Alison B.

Peck, Robert L. Seaman, and Chris R. Benn. Vol. 9910. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 991016. arXiv: `1707.02202 [astro-ph.IM]`.

Pozzo, M., W. P. S. Meikle, A. Fassia, T. Geballe, P. Lundqvist, N. N. Chugai, and J. Sollerman (Aug. 2004). „On the source of the late-time infrared luminosity of SN 1998S and other Type II supernovae“. In: 352.2, pp. 457–477. arXiv: `astro-ph/0404533 [astro-ph]`.

Priddey, R. S., K. G. Isaak, R. G. McMahon, E. I. Robson, and C. P. Pearson (Oct. 2003). „Quasars as probes of the submillimetre cosmos at z 5 - I. Preliminary SCUBA photometry“. In: 344, pp. L74–L78. eprint: `astro-ph/0308132`.

Protopapas, P., J. M. Giammarco, L. Faccioli, M. F. Struble, R. Dave, and C. Alcock (June 2006). „Finding outlier light curves in catalogues of periodic variable stars“. In: 369.2, pp. 677–696. arXiv: `astro-ph/0505495 [astro-ph]`.

Reis, Itamar, Dalya Baron, and Sahar Shahaf (Jan. 2019). „Probabilistic Random Forest: A Machine Learning Algorithm for Noisy Data Sets“. In: 157.1, 16, p. 16. arXiv: `1811.05994 [astro-ph.IM]`.

Rho, J., H. L. Gomez, A. Boogert, M. W. L. Smith, P. -O. Lagage, D. Dowell, C. J. R. Clark, E. Peeters, and J. Cami (Oct. 2018). „A dust twin of Cas A: cool dust and 21 μm silicate dust feature in the supernova remnant G54.1+0.3“. In: 479.4, pp. 5101–5123. arXiv: `1707.08230 [astro-ph.GA]`.

Rho, J., W. T. Reach, A. Tappe, U. Hwang, J. D. Slavin, T. Kozasa, and L. Dunne (July 2009a). „Spitzer Observations of the Young Core-Collapse Supernova Remnant 1E0102-72.3: Infrared Ejecta Emission and Dust Formation“. In: 700, pp. 579–596.

Rho, J., W. T. Reach, A. Tappe, U. Hwang, J. D. Slavin, T. Kozasa, and L. Dunne (July 2009b). „SPITZEROBSERVATIONS OF THE YOUNG CORE-COLLAPSE SUPERNOVA REMNANT 1E0102-72.3: INFRARED EJECTA EMISSION AND DUST FORMATION“. In: *The Astrophysical Journal* 700.1, pp. 579–596.

Rijn, Jan N. van and Frank Hutter (2018). „Hyperparameter Importance Across Datasets“. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '18. London, United Kingdom: Association for Computing Machinery, pp. 2367–2376.

Robertson, H. P. (Nov. 1935). „Kinematics and World-Structure“. In: 82, p. 284.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams (Oct. 1986). „Learning representations by back-propagating errors“. In: 323.6088, pp. 533–536.

Ryden, Barbara (2003). *Introduction to cosmology*.

Sadeh, I., F. B. Abdalla, and O. Lahav (Oct. 2016). „ANNz2: Photometric Redshift and Probability Distribution Function Estimation using Machine Learning“. In: 128.968, p. 104502. arXiv: `1507.00490 [astro-ph.CO]`.

Sadeh, Iftach, Filipe B. Abdalla, and Ofer Lahav (Oct. 2019). *ANNz2: Estimating photometric redshift and probability density functions using machine learning methods*. ascl: `1910.014`.

Salvato, Mara, Olivier Ilbert, and Ben Hoyle (June 2019). „The many flavours of photometric redshifts“. In: *Nature Astronomy* 3, pp. 212–222. arXiv: `1805.12574 [astro-ph.GA]`.

Sánchez Almeida, J. and C. Allende Prieto (Jan. 2013). „Automated Unsupervised Classification of the Sloan Digital Sky Survey Stellar Spectra using k-means Clustering“. In: 763.1, 50, p. 50. arXiv: `1211.5321 [astro-ph.SR]`.

Sarangi, Arkaprabha, Eli Dwek, and Richard G. Arendt (May 2018). „Delayed Shock-induced Dust Formation in the Dense Circumstellar Shell Surrounding the Type IIn Supernova SN 2010jl“. In: 859.1, 66, p. 66. arXiv: `1804.06878 [astro-ph.SR]`.

Scalzo, R. A., F. Yuan, M. J. Childress, *et al.* (July 2017). „The SkyMapper Transient Survey“. In: 34, e030, e030. arXiv: `1702.05585 [astro-ph.IM]`.

Schmidt, S. J., A. I. Malz, J. Y. H. Soo, *et al.* (Jan. 2020). „Evaluation of probabilistic photometric redshift estimation approaches for LSST“. In: *arXiv e-prints*, arXiv:2001.03621, arXiv:2001.03621. arXiv: `2001.03621 [astro-ph.CO]`.

Seitenzahl, Ivo R., F. X. Timmes, and Georgios Magkotsios (Sept. 2014). „The Light Curve of SN 1987A Revisited: Constraining Production Masses of Radioactive Nuclides“. In: 792.1, 10, p. 10. arXiv: `1408.5986 [astro-ph.SR]`.

Shapley, L. S. (2016). „17. A Value for n-Person Games“. In: *Contributions to the Theory of Games (AM-28), Volume II*. Ed. by Harold William Kuhn and Albert William Tucker. Princeton University Press, pp. 307–318.

Shrikumar, Avanti, Peyton Greenside, and Anshul Kundaje (Apr. 2017). „Learning Important Features Through Propagating Activation Differences“. In: *arXiv e-prints*, arXiv:1704.02685, arXiv:1704.02685. arXiv: `1704.02685 [cs.CV]`.

Shrikumar, Avanti, Peyton Greenside, Anna Shcherbina, and Anshul Kundaje (May 2016). „Not Just a Black Box: Learning Important Features Through Propagating Activation Differences“. In: *arXiv e-prints*, arXiv:1605.01713, arXiv:1605.01713. arXiv: `1605.01713 [cs.LG]`.

Shuntov, M., J. Pasquet, S. Arnouts, *et al.* (Apr. 2020). „PhotoWeb redshift: boosting photometric redshift accuracy with large spectroscopic surveys“. In: 636, A90, A90. arXiv: `2003.10766 [astro-ph.GA]`.

Smith, Nathan (2016). „Interacting Supernovae: Types IIn and Ibn“. In: *arXiv: High Energy Astrophysical Phenomena*.

Smith, Nathan, Ryan Chornock, Weidong Li, Mohan Ganeshalingam, Jeffrey M. Silverman, Ryan J. Foley, Alexei V. Filippenko, and Aaron J. Barth (Oct. 2008). „SN 2006tf: Precursor Eruptions and the Optically Thick Regime of Extremely Luminous Type IIn Supernovae“. In: 686.1, pp. 467–484. arXiv: `0804.0042 [astro-ph]`.

Smoot, G. F., C. L. Bennett, A. Kogut, *et al.* (Sept. 1992). „Structure in the COBE Differential Microwave Radiometer First-Year Maps“. In: 396, p. L1.

Spiniello, C. and A. Agnello (Oct. 2019). „VEXAS: VISTA EXtension to Auxiliary Surveys. Data Release 1. The southern Galactic hemisphere“. In: 630, A146, A146. arXiv: `1908.11392 [astro-ph.GA]`.

Stritzinger, Maximilian, Francesco Taddia, Claes Fransson, *et al.* (Sept. 2012). „Multiwavelength Observations of the Enduring Type IIn Supernovae 2005ip and 2006jd“. In: 756.2, 173, p. 173. arXiv: `1206.5575 [astro-ph.CO]`.

Szalai, Tamás, Szanna Zsıros, Ori D. Fox, Ondřej Pejcha, and Tomás Müller (Apr. 2019). „A Comprehensive Analysis of Spitzer Supernovae“. In: 241.2, 38, p. 38. arXiv: `1803.02571 [astro-ph.HE]`.

Teh, Yee Whye (2010). „Dirichlet Process“. In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, pp. 280–287.

Temim, Tea and Eli Dwek (Aug. 2013). „THE IMPORTANCE OF PHYSICAL MODELS FOR DERIVING DUST MASSES AND GRAIN SIZE DISTRIBUTIONS IN SUPERNOVA EJECTA. I. RADIATIVELY HEATED DUST IN THE CRAB NEBULA“. In: *The Astrophysical Journal* 774.1, p. 8.

Temim, Tea, Eli Dwek, Richard G. Arendt, Kazimierz J. Borkowski, Stephen P. Reynolds, Patrick Slane, Joseph D. Gelfand, and John C. Raymond (Feb. 2017). „A Massive Shell of Supernova-formed Dust in SNR G54.10.3“. In: *The Astrophysical Journal* 836.1, p. 129.

The LSST Dark Energy Science Collaboration, Rachel Mandelbaum, Tim Eifler, *et al.* (Sept. 2018). „The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document“. In: *arXiv e-prints*, arXiv:1809.01669, arXiv:1809.01669. arXiv: `1809.01669 [astro-ph.CO]`.

Todini, Paolo and Andrea Ferrara (Aug. 2001). „Dust formation in primordial Type II supernovae“. In: 325.2, pp. 726–736. arXiv: `astro-ph/0009176 [astro-ph]`.

Vanderplas, Jake and Andrew Connolly (Nov. 2009). „Reducing the Dimensionality of Data: Locally Linear Embedding of Sloan Galaxy Spectra“. In: 138.5, pp. 1365–1379. arXiv: `0907.2238 [astro-ph.IM]`.

Vanzella, E., S. Cristiani, A. Fontana, *et al.* (Aug. 2004). „Photometric redshifts with the Multilayer Perceptron Neural Network: Application to the HDF-S and SDSS“. In: 423, pp. 761–776. arXiv: `astro-ph/0312064 [astro-ph]`.

Viola, M., M. Cacciato, M. Brouwer, *et al.* (Oct. 2015). „Dark matter halo properties of GAMA galaxy groups from 100 square degrees of KiDS weak lensing data“. In: 452.4, pp. 3529–3550. arXiv: `1507.00735 [astro-ph.GA]`.

Walker, A. G. (Jan. 1937). „On Milne's Theory of World-Structure“. In: *Proceedings of the London Mathematical Society* 42, pp. 90–127.

Wang, D., Y. Zhang, and Y. Zhao (Aug. 2008a). „Estimating Photometric Redshifts of Quasars Using Support Vector Machines“. In: *Astronomical Data Analysis Software and Systems XVII*. Ed. by R. W. Argyle, P. S. Bunclark, and J. R. Lewis. Vol. 394. Astronomical Society of the Pacific Conference Series, p. 509.

Wang, R., C. L. Carilli, J. Wagg, *et al.* (Nov. 2008b). „Thermal Emission from Warm Dust in the Most Distant Quasars“. In: 687, 848-858, pp. 848–858. arXiv: `0806.3022`.

Watson, D., L. Christensen, K. K. Knudsen, J. Richard, A. Gallazzi, and M. J. Michałowski (Mar. 2015). „A dusty, normal galaxy in the epoch of reionization“. In: 519, pp. 327–330. arXiv: `1503.00002`.

Weerts, Hilde J. P., Andreas C. Mueller, and Joaquin Vanschoren (July 2020). „Importance of Tuning Hyperparameters of Machine Learning Algorithms“. In: *arXiv e-prints*, arXiv:2007.07588, arXiv:2007.07588. arXiv: `2007.07588 [cs.LG]`.

Wesson, R., M. J. Barlow, M. Matsuura, and B. Ercolano (Nov. 2014). „The timing and location of dust formation in the remnant of SN 1987A“. In: *Monthly Notices of the Royal Astronomical Society* 446.2, pp. 2089–2101. eprint: `https://academic.oup.com/mnras/article-pdf/446/2/2089/9389396/stu2250.pdf`.

Wesson, R., M. J. Barlow, M. Matsuura, and B. Ercolano (Jan. 2015). „The timing and location of dust formation in the remnant of SN 1987A“. In: 446, pp. 2089–2101. arXiv: `1410.7386 [astro-ph.SR]`.

Wright, D. E., S. J. Smartt, K. W. Smith, *et al.* (May 2015). „Machine learning for transient discovery in Pan-STARRS1 difference imaging“. In: 449.1, pp. 451–466. arXiv: `1501.05470 [astro-ph.IM]`.

Wright, Edward L., Peter R. M. Eisenhardt, Amy K. Mainzer, *et al.* (Dec. 2010). „The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance“. In: 140.6, pp. 1868–1881. arXiv: `1008.0031 [astro-ph.IM]`.

You, Yang, Igor Gitman, and Boris Ginsburg (Aug. 2017). „Large Batch Training of Convolutional Networks“. In: *arXiv e-prints*, arXiv:1708.03888, arXiv:1708.03888. arXiv: `1708.03888 [cs.CV]`.

Zhukovska, S. and T. Henning (July 2013). „Dust input from AGB stars in the Large Magellanic Cloud“. In: 555, A99, A99. arXiv: `1305.3521 [astro-ph.GA]`.

Zubko, V. G., V. Mennella, L. Colangeli, and E. Bussoletti (Oct. 1996). „Optical constants of cosmic carbon analogue grains - I. Simulation of clustering by a modified continuous distribution of ellipsoids“. In: 282.4, pp. 1321–1329.