



REVERBERATION MAPPING OF NEARBY SEYFERT GALAXIES FOR COSMOLOGY.

Determining the dust torus time lag

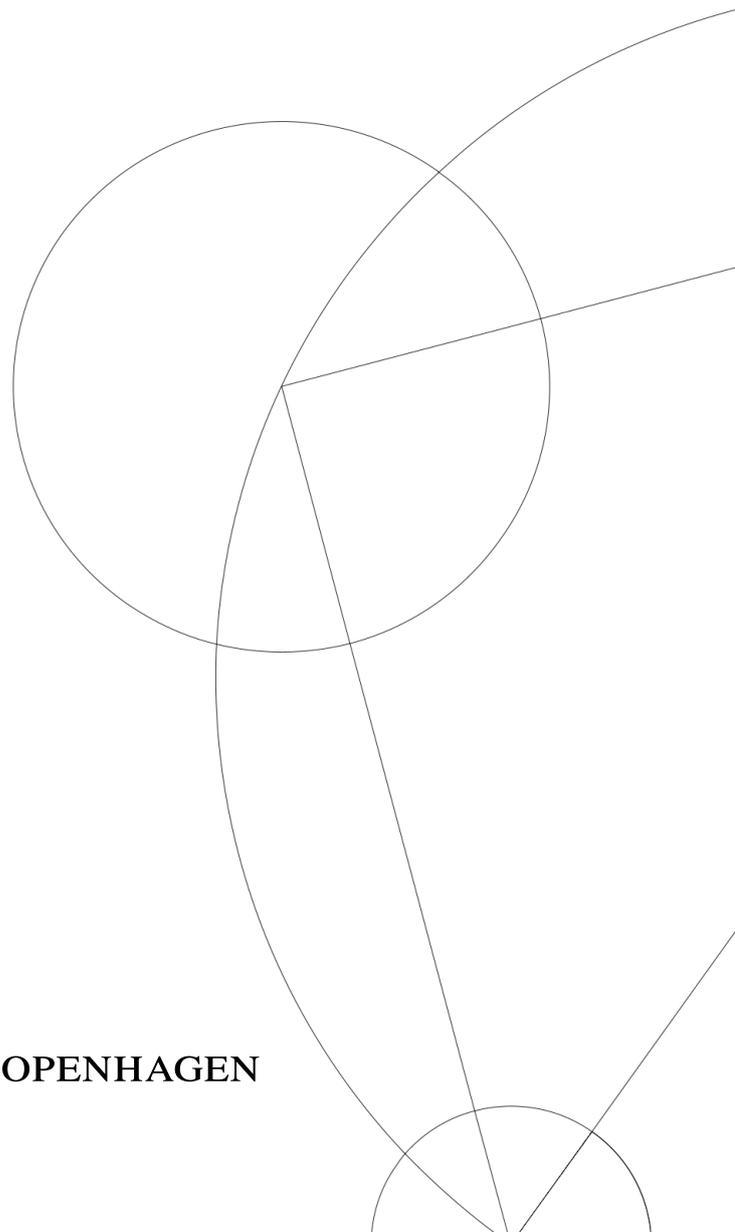
MASTER THESIS

Written by *Malte Brinch*

July 2, 2020

Under Supervision of
Darach Jafar Watson

UNIVERSITY OF COPENHAGEN





UNIVERSITY OF
COPENHAGEN

NAME OF INSTITUTE: University of Copenhagen

NAME OF DEPARTMENT: The Niels Bohr Institute

AUTHOR(S): Malte Brinch

EMAIL: mgb512@alumni.ku.dk

TITLE AND SUBTITLE: Reverberation mapping of nearby Seyfert galaxies
for cosmology.
- Determining the dust torus time lag

SUPERVISOR(S): Darach Jafar Watson

HANDED IN: 02.07.2020

DEFENDED: TBD Defended 17 August 2020

NAME _____

SIGNATURE _____

DATE _____

Abstract

There is currently a dispute as to the true value of the Hubble constant. The discrepancy is mainly between the value determined by the Planck collaboration using the Cosmic Microwave background and the value determined by using the traditional distance ladder by measuring type Ia supernovas in the local universe. The idea of this project is to determine an intermediary value for the Hubble constant that can corroborate either of the other values or perhaps hint at undiscovered physics. The idea behind this project is to determine the lag of the dusty torus in AGN, so that it can be used to determine the absolute distance to the object. The data to do this reverberation mapping will be from the Rapid Eye Mount telescope at the La Silla Observatory in Chile. The data is from the 3 J,H,K Near Infra Red filters and the 4 g,r,i,z optical filters. The main focus of this project is to create an algorithm that perform photometric reverberation mapping analysis using the 7 light curves without having any data for the driving light curve that drives the response seen in the observed data. The focus is on creating code that reliably and with reasonable compute time that determine the lag to the torus. To set up such an algorithm background knowledge of the structure of AGN and the mathematical models used to describe their light curves is presented. to efficiently determine the lag, a state of the art Markov Chain Monte Carlo Sampler is used and the setup for such a sampler is discussed. To test the reverberation mapping model a script is created that can create synthetic light curves to test the model. While the model was able to recover the lag with some uncertainty using the synthetic data, the model failed to work on data from actual AGN. The reason as to why this might be the case are discussed. It is concluded that more work needs to be done to understand the specific form of the transfer function used in the model and that the scatter in the light curves might hinder the determination of the lag.

Contents

1	Introduction	2
1.1	The Hubble constant and cosmological measurements	2
1.1.1	The distance ladder and measuring the Hubble constant	3
1.2	AGN	7
1.2.1	AGN Morphology	7
1.2.2	Categorizing AGN	14
1.3	Reverberation Mapping	15
1.3.1	The transfer function and convolution	17
2	Defining a Driving Light Curve	25
2.1	AR(1) and CAR(1) process	25
2.1.1	Structure function	28
2.2	Gaussian Processes	29
2.2.1	Regression and Gaussian processes	31
2.2.2	Function space view	32
2.2.3	The mean- and covariance function	32
2.3	A degeneracy in DRW modelling of AGN light curves	35
3	MCMC	37
3.1	The Monte Carlo method	38
3.2	What is sampling?	40
3.3	MCMC samplers	41
3.3.1	Advances samplers	42
3.3.2	Hamiltonian Monte Carlo and the No U-Turn Sampler	42
4	Data	43
4.1	Instruments: REMIR/ROS2	45
4.2	Objects	45
5	Creating synthetic data	46
6	Model setup and testing	52
6.1	Defining priors	53
6.2	Setting up the GP	54
6.3	Setting up the model parameters	56
6.4	Initializing the sampler	56
6.5	Testing the model on synthetic data	57
6.6	Testing the model on real data	60
7	Discussion	61

8 Conclusion	64
References	64
A Figures for Second synthetic model test	70
B Plots of other objects	70

1 Introduction

This paper investigates the use of Active Galactic Nuclei (AGN) light curves from the Rapid Eye Mount (REM) and REMIR 2 telescopes situated at the La Palma site located in Chile to determine the thermal and optical time lags of said AGN. The time lags can be used to determine a value of the Hubble constant H_0 independent from the standard distance ladder. The time lags are determined from the 7 observational bands (J,H,K,g,r,i,z) by using a Bayesian framework together with a Gaussian Process (GP) and a state of the art Markov Chain Monte Carlo (MCMC) sampler to sample not only the transfer function parameters used to determine the time lags for the AGN, but also the UV-continuum driving light curve.

The paper will go through the theoretical and practical considerations taken to create a Python script that can determine the driving light curve, the transfer functions and finally the time lags. To understand the use of the specific model I implement, a description of the physical setup for the AGN relevant to determining the time lags is provided. To test that the Python script can sample the posterior probability of the transfer function parameters and the driving light curve, I describe the creation of a data generation script that can generate both a specific synthetic driving light curve and synthetic observational band data for the 7 relevant bands. The synthetic data is created to simulate the real data from the REM telescope as closely as possible.

1.1 The Hubble constant and cosmological measurements

¹ Hubble's law describes how galaxies and astronomical objects not gravitationally bound will travel away from each other with a recessional velocity that is proportional to their distance. Hubble's law can be written as

$$z = \frac{H_0}{c}d, \quad z = \frac{\lambda_{ob}}{\lambda_{em}} - 1 \quad (1)$$

Where z is the redshift determined from the observed λ_{ob} and emitted λ_{em} wavelengths, H_0 is the Hubble constant and d is the distance to the galaxy. Interpreting the redshifts as Doppler shift, Hubble's law can be written as

$$v = H_0d \quad (2)$$

Hubble's law indicates that the universe is expanding and this movement of astronomical object due to expansion is called the Hubble flow. The velocity of galaxies induced by their gravitational interaction is called the peculiar velocity, whereas the recessional velocity is a result of the interplay between the curvature of spacetime and the movement of mass-energy as described by the theory of general relativity. Determining H_0 makes it possible to calculate the distance to galaxies that are not gravitationally bound

$$d \simeq \frac{c}{H_0} \frac{(z+1)^2 - 1}{(z+1)^2 + 1} \quad (3)$$

The distance is only approximate as galaxies will still have peculiar velocities, though as the distances to them increase the peculiar velocities become a smaller and smaller fraction of their overall velocity. The relative error due to the peculiar velocity is assumed to be 10% or less for a distance over 50 Mpc

¹This section relies on [Ryden \(2017\)](#) and [Sparke & Gallagher\(III\) \(2007\)](#).

(Jackson, 2015). To be able to use equation 3 an accurate estimate of H_0 needs to be determined. This process of determination has historically been difficult due to a number of factors, like having to rely on a sequence of secondary distance measurements (the "distance ladder") all with their own uncertainties and the aforementioned peculiar velocities to obtain distances to distant galaxies.

The Malmquist bias has also historically been a source of error for determining H_0 . Malmquist bias is introduced in magnitude limited samples due to the fact that only object under a sharp apparent magnitude limit are observed and this can lead to skewed statistics if not corrected for. The Tully-Fisher relation shows that the rotational speed of a edge-on spiral galaxy increases with its luminosity, roughly as $L \propto V_{max}^\alpha$ with V_{max} being the peak rotational speed and $\alpha \sim 4$. With the luminosity it is possible to obtain the distance to the object and therefore this method can be used to obtain an estimate of H_0 . When trying to determine the value of H_0 with the Tully-Fisher relation a nearby galaxy cluster can be used for calibration and then the relation can be applied to a more distant cluster. If direct regression is used and the Malmquist bias is not taken into account, the distance estimate will be systematically underestimated (HENDRY *et al.* , 1993) and subsequently H_0 will be overestimated.

In the present the value of the H_0 has come into contention, in what has come to be known as "The Hubble tension" (Poulin *et al.* , 2019). Measurements of H_0 in the local universe using the distance ladder approach (Cepheids/SNe Ia) from the SH₀ES team determine a value of H_0 around 74.03 ± 1.42 km/s/Mpc (Riess *et al.* , 2019). On the other hand a value of H_0 close to 67.27 ± 0.60 km/s/Mpc (Planck-Collaboration, 2018) is inferred from Cosmic Microwave Background (CMB) temperature and polarization power spectra plus CMB lensing from the Planck Collaboration assuming a spatially-flat 6-parameter Λ CDM cosmology. The discrepancy in the values of H_0 is at 4.4σ and therefore excludes the possibility of it being due to random chance (Knox & Millea, 2019). The values of H_0 determined from the distance ladder approach are generally higher than those derived from CMB approach. Neutron star mergers ($H_0 = 68.9^{+4.7}_{-4.6}$ km/s/Mpc) (Hotokezaka *et al.* , 2019) and Supernovas calibrated with the tip of the red-giant branch ($H_0 = 68.9 \pm 0.8(\text{stat}) \pm 1.7(\text{sys})$ km/s/Mpc) (Freedman *et al.* , 2019) have also yielded results that are consistent with the other results at the 2σ level. There has also been an attempt to explain the discrepancy between the values from these early and late universe approached by proposing a nearby 40 Mpc underdense region with half the density of the rest of the universe (Lombriser, 2020). The discrepancy could hint at unexplored physics not included in The Standard Model. The discrepancy motivates a search for a different approach that can close the gap between the late and early universe approaches and that does not suffer from the same problems the distance ladder approach has. This paper attempts to develop and apply such an approach using AGN.

1.1.1 The distance ladder and measuring the Hubble constant

² To obtain an accurate estimate of H_0 from equation 3, we need an accurate determination of the distance to the object and its recessional velocity. The redshift of the object can be determined from emission lines (equation 1) and used to determine the recessional velocity in the cases where the relative error from the peculiar velocity of the object is small. The distance to the object needs to be determined

²This section relies on Ryden (2017) and Jackson (2015).

through observable properties inherent to the object, in the local universe this is traditionally done with the distance ladder through the use of standard candles and rulers. Methods appropriate to closer objects like stars in our galaxy are used to calibrate the ladder so it can then be used on objects at greater distances.

Standard candles are objects with know luminosity's throughout spacetime. By knowing the luminosity of a specific object or specific type of objects you can measure the apparent flux f from the object to calculate the luminosity distance.

$$d_L = \left(\frac{L}{4\pi f} \right)^{1/2} \quad (4)$$

d_L has a very simple relation with the proper distance to the object $d_P(t_0)$ under the assumption that space is very nearly flat.

$$d_L = d_P(t_0)(1 + z) \quad (5)$$

A standard ruler on the other hand is an object whose proper length l is known. If you have a standard ruler in your line of sight with a known length, you can measure its redshift and angular extend on the sky. For small angular extends ($\delta\theta \ll 1$) you can use the small-angle formula to obtain the angular diameter distance.

$$d_A = \frac{l}{\delta\theta} \quad (6)$$

and again assuming a spatially flat universe gives a simple relation with the proper distance.

$$d_A = \frac{d_P(t_0)}{1 + z} \quad (7)$$

Standard rulers have generally been affected by observational difficulties, as an example the standard ruler must have an angular size that can be resolved by our telescopes. Assuming the cosmological Benchmark Model a standard ruler will have its angular size minimized at the critical redshift z_c and there the angular size will be,

$$\delta\theta_{min} = \frac{l}{d_{A,max}} = \frac{l}{1770Mpc} \approx 0.1arcsec \left(\frac{l}{1kpc} \right) \quad (8)$$

Galaxies and clusters of galaxies are in the size range where they would be useful as standard rulers, but they do not have clearly defined shapes and since these object are not isolated they change over time due mergers with their neighbors. This makes it difficult to assign a particular angular size to a corresponding physical size for these objects.

To understand the cosmic distance ladder, let us start by considering the smallest astronomical scales. At relatively short distances on a cosmic scale, distances to planets in our solar system can be measured by bouncing radio signals off of them. Correction due to the relative motion of the planets is small, since the radio waves are traveling at the speed of light. At distances greater than 10 AU this method starts to falter as the reflected radio waves become too faint to detect.

Next step is to measure the distance to stars within our galaxy, this is done using the method of trigonometric parallax. When observing a star from two different points separated by a baseline of

distance b , the star's apparent position on the sky changes by an angle θ . In the case where the baseline is perpendicular to the line of sight to the star the parallax distance is

$$d_{par} = 1pc \left(\frac{b}{1AU} \right) \left(\frac{\theta}{1arcsec} \right)^{-1} \quad (9)$$

Where 1 parsec is equivalent to 206265 AU. Using the Earth's as a baseline (2 AU) the limiting factor in using this approach is the accuracy involved in determining the parallax angle θ , since the size Earth's orbit is known with a high degree of accuracy due to radio measurements. To be able to measure θ for an object 100 Mpc away using Earth's orbit as a baseline would require an error of $< 0.01 \mu arcsec$ which is not currently feasible meaning this approach cannot be used to determine the cosmological distances of other galaxies.

It is possible to determine the distance to open clusters within our galaxy. The hundreds of stars in these cluster can be plotted on a Hertzsprung-Russell (HR) diagram of temperature (determined from their colour and Wien's law) against apparent luminosity to show the "main sequence" of stars where the stars burn hydrogen to helium. for close clusters all the stars are at the same distance and we are able to use the parallax approach to determine the distance with an error of $< 1\%$. This makes it possible to calibrate the main sequence so that it can be used to predict the absolute luminosity of a main-sequence star of a given colour. It is then possible to apply this method to other clusters not covered by the parallax approach to determine their distance with an error of a few $\%$. An example would be the closest objects outside our galaxy, the LMC and SMC, which are at a distance of 50-60 kpc.

Next step on the ladder are variable stars like RR-Lyrae and Cepheid's. Cepheid's are the most applicable to determine inter-galactic distances since they are intrinsically the most luminous, being that they are relatively hot (7000-8000 K) giant blue stars. The period of pulsation for Variable stars is proportional to their luminosity, which makes them standard candles. Cepheid's in open clusters close to us can be used to calibrate this approach so it can be applied to objects at greater distances by comparing their variability periods. This approach can be used up to a distance of 20-30 Mpc using the Hubble Space Telescope where the Cepheid's become too faint to detect. On this distance scale the Local Group is gravitationally attracted towards the Virgo cluster, giving it a peculiar velocity in that direction.

The last step in the traditional distance ladder is the use of Type Ia supernovae (SNe). Galaxies with a distance determined by Cepheid's are used to calibrate the supernovas in those galaxies so they can be compared with supernovas in galaxies at distances of > 50 Mpc, where the relative error from the peculiar velocities of galaxies is assumed to be under 10% and comparison between redshift and distance will give an accurate determination of H_0 . SNe Ia are produced by a binary star systems where a giant star is dumping mass on to a white dwarf which has already gone through its evolutionary process, collapsed and is hold together by electron-degeneracy pressure. When the white dwarfs obtains the critical Chandrasekhar mass of $1.4 M_{\odot}$ electron-degeneracy pressure can no longer stop the dwarf from collapsing leading to a supernova explosion. The light curve from the explosion has a characteristic rise and fall regardless of distance, since the physics of the explosion are the same for all SNe Ia. There is a strong correlation between the peak brightness and the degree to which the

supernova fades in brightness 15 afterwards (Δm_{15}).

Other approaches to obtaining distance measurements would be the aforementioned Tully-fisher relation for spiral galaxies or its elliptical galaxy equivalent the Faber-Jackson relation.

The problem with the distance ladder approach is that errors propagate through each step of the ladder making the final determination of H_0 uncertain. Ideally we would want an approach using a standard candle or ruler that could circumvent the problems associated with the traditional distance ladder. There are a number of one-step methods that do exactly this.

Megamaser systems involve clumps of gas which are typically located ~ 0.1 pc from the center of a galaxy, close to the central Super-Massive Black Hole (SMBH). These clumps radiate coherently in the water line at a frequency of approximately 22 GHz. With Very Long Baseline Interferometry (VLBI) techniques the line can be observed at the milliarcsec scale. The velocity and acceleration of the gas clumps can be determined and under the assumption of Keplerian motion the radius from the central SMBH can be calculated. With the physical radius of the gas clumps and the angular separation from the galaxy center, we have a standard ruler which can be used to determine an estimate of H_0 in one step. There can be systematic errors involving the modelling parameters (eccentricity, position angle, periapsis angle and inclination) not adequately describing the disk, but these can be simulated and are likely to be small (Humphreys *et al.*, 2013).

Other one step approaches are the use of gravitational lensing, the CMB approach (using the Sunyaev-Zel'dovich effect) mentioned before and gravitational waves from the coalescence of compact binaries (Chassande-Mottin *et al.*, 2019). All of these one-step approaches have their own set of errors, statistical and systematic, that have to be accounted for to determine an estimate of H_0 .

From all of these examples of approaches to determining H_0 , it becomes clear that there are a series of conditions inherent to the ideal object for measuring H_0 (Jackson, 2015):

- Has a property which allows it to be treated as either as a standard candle or as a standard ruler
- Can be used independently of other calibrations (i.e., in a one-step process)
- Lies at a large enough distance (a few tens of Mpc or greater) that peculiar velocities are small compared to the recession velocity at that distance
- Involves as little astrophysics as possible, so that the distance determination does not depend on internal properties of the object
- Provides the Hubble constant independently of other cosmological parameters.

AGN (especially Quasars) are among the most luminous objects in the universe, but they are unable to be used as standard candles due to the fact that their luminosity varies while their Spectral Energy Distribution (SED) remains mostly unchanged (Risaliti & Lusso, 2015). On the other hand since AGN are so luminous they are ideal objects to observe at high redshift, so if they could be used as standard rulers they would be ideal to obtain an intermediary value of H_0 between the one from the local distance ladder and the CMB. This project attempts to describe a method by which we can obtain a measure of

the physical size of the AGN through reverberation mapping. As will be described later the approach of obtaining the transfer function from AGN to use them as standard rulers fulfills multiple of the conditions for an ideal object, since it is a one-step process standard ruler that sits at a comparably large distance and does not use other cosmological parameters to determine H_0 .

1.2 AGN

³ Active Galactic Nuclei are the compact centers of galaxies that emit relatively strong radiation over the entire electromagnetic spectrum, even in regions of the spectrum where the stars of galaxies do not emit at all. AGN can have energy outputs of $E > 3 \times 10^{61} \text{erg}$ over their lifetimes. The emission is variable over time-frames of months to days and in some cases even hours, depending on the energy level of the emission such that higher energy emission varies more within a given timeframe. Variability on these timescales suggest the radiation comes from a region of space only a few light-weeks in diameter. It should be noted that the gas clouds in these region also move a lot faster than the galaxies stars, up to and around 50 times faster. The most plausible explanation is a black hole, where energy is released when gas, located in an accretion disk, falls into the central black hole.

Observationally AGN are identified through a number of criteria ([Peterson \(1997\)](#), [Peterson \(2008\)](#) and [Mo et al. \(2010\)](#)), though not all of these criteria need to be fulfilled for a given object to be classified as an AGN. The data provided for this project is of AGN that appear as point like objects with time varying emission in the center of spiral galaxies (Seyfert galaxies) and with SEDs that have properties unlike stars.

- The point like representation of an AGN upon an imaging detector
- Strong emission lines
- Strong X-RAY emission
- The Continuum Luminosity varies over time
- Evidence of strong non-stellar emission
- Radio emission
- Non-stellar UV through IR emission
- Broad emission lines in the UV through IR.

1.2.1 AGN Morphology

To be able to use AGN as standard rulers we need an accurate understanding of the morphology of the AGN to be able to differentiate objects that are suitable to obtain a physical distance measure and also to determine how such a method would be applied.

³Unless specified otherwise this section mainly relies on [Sparke & Gallagher\(III\) \(2007\)](#) and [Schneider \(2006\)](#) and occasionally on [Peterson \(1997\)](#), [Peterson \(2008\)](#) and [Mo et al. \(2010\)](#).

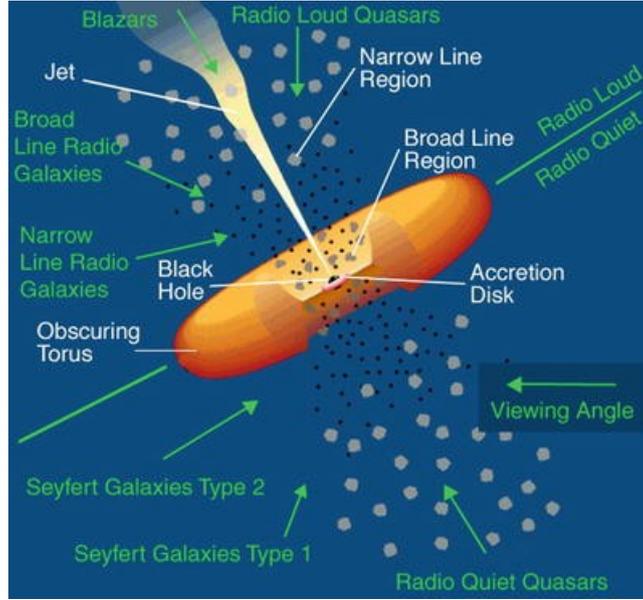


Figure 1: A simple model of the structure of an AGN. The black hole is located at the center, where gas from the surrounding accretion disk falls onto it. A Broad Line Region (BLR) is located around the black hole, with Gas clouds in bound orbits creating broad emission lines. farther out is a dusty torus and slower moving clouds creating relatively narrow emission lines. It should be noted that while radio plasma jets are shown in this picture, not every AGN has jets. Different types of AGN (1-2) are indicated by the angle they are observed from. This figure is originally from [Urry & Padovani \(1995\)](#), modified by NASA.

Figure 1 shows a simple model of the structure of a typical AGN. In the center A black hole is located. Material falling into the black hole will inevitably have some angular momentum and will form an accretion disk of rapidly spinning gas surrounding the black hole. The surrounding gas spirals closer towards the black hole and heats up due to viscose processes, thereby losing gravitational potential energy and releasing it as thermal radiation. Farther from the center gas clouds of varying densities are circulating, while the entire region is encompassed by a dusty torus.

1.2.1.1 The AGN Black Hole

The light from a AGN is strong enough to exert a considerable pressure on gas surrounding it. If the outward pressure is too strong the gas will be pushed away and the nucleus will be starved of fuel suffocating itself. This leads to a limit on the luminosity of the object. Assuming spherical symmetry it is possible to calculate the point where the radiation pressure is equal to the inward pull of gravity. Assuming the gas near the nucleus is fully ionized hydrogen, the outward force comes from the Thomson scattering of electrons (we can neglect proton scattering due to their mass). The Thomson cross-section of interaction for each electron is

$$\sigma_T = \frac{e^4}{6\pi\epsilon_0^2 c^4 m_e^2} \quad (10)$$

Where e is the charge of the electron and m_e is its mass. If the nucleus releases photons with luminosity L , they will have a momentum L/c and so an electron would for a given radius r gain a momentum

of $\sigma_T L / (4\pi r^2 c)$ each second. The electrostatic forces between the electron and proton makes it the electron drags the proton with it and so we must compare the outward force on them both with the inward pull of gravity on them both. if the nucleus has mass M the force balance point will when

$$\frac{GM(m_e + m_p)}{r^2} \approx \frac{GMm_p}{r^2} = \frac{\sigma_T L}{4\pi r^2 c} \quad (11)$$

with m_p being the mass of the proton. rearranging equation 11 we obtain the Eddington luminosity the largest luminosity where it is still possible for material to fall inward

$$L_E = \frac{4\pi GMm_p c}{\sigma_T} \quad (12)$$

It should be noted that the presence of jets in AGN makes it possible to exceed the Eddington luminosity, since this breaks the assumption of spherical symmetry. The total luminosity is not expected to be more than a few times the Eddington luminosity.

The bolometric luminosity L_{Bol} is the total luminosity of an object over all wavelengths and must also be equal to the energy output of the AGN, assuming that the majority of the light we observe from the object is a result of the radiative processes of the AGN. Generally this will not be equal to the Eddington luminosity. This is described by the equation

$$L_{Bol} = \eta \dot{M}_{BH} c^2, \quad (13)$$

where η is the efficiency of the black hole in converting the rest-mass of the infalling gas into energy (Mc^2) that is radiated away, \dot{M}_{BH} is the mass accretion rate of the black hole. Compare this to the luminosity of an AGN where the material falling into the black hole crosses a shell at radius r

$$L_{AGN} = \frac{GM_{BH}}{r} \dot{M}_{BH} \quad (14)$$

The Schwarzhild radius for a black hole is the radius where the escape velocity of the solution for the virial theorem is faster than c .

$$r_S = \frac{2GM_{BH}}{c^2} \quad (15)$$

Combining equations 13, 14 and 15 it is possible to obtain a measure of the conversion efficiency η

$$\eta = \frac{L_{AGN}}{\dot{M}_{BH} c^2} = \frac{1}{2} \frac{r_S}{r} \quad (16)$$

For free falling material without angular momentum the conversion efficiency would be 6%, though observation of continuum radiation from AGN suggest a primary origin at $r \sim 5r_S$ leading to a value around 10% (Peterson, 1997). This suggest that the black holes in AGN rotate. Theoretically for a rapidly spinning black hole the conversion efficiency could be as high as 42%. The conversion efficiency is considerably higher than nuclear fusion done in stars, which is at 0.7% leading further credence to the fact that the nucleus of a AGN cannot be a star.

1.2.1.2 The Accretion Disk

When gas falls onto a compact like a black hole it losses potential energy, as it is converted to kinetic energy. If there is nothing to prevent the infall the gas will not be able to radiate away its energy as

it falls into the black hole. We would expect the gas that falls towards the black hole to have some finite amount of angular momentum. This means that the gas will not be able to fall directly into the black hole and instead the gas particle will have friction amongst each other, leading to a transfer of momentum and the formation of a disk with its orientation perpendicular to the vector of the angular momentum. The viscous forces in the gas should be small compared to the force of gravity and therefore the disk should approximately be in Keplerian rotation. In a Keplerian disk the rotation is differential meaning the angular velocity depends on the radius from the center. The disk will be heated by friction between gas at different radii, like concentric rings grinding against each other. The friction also creates a deceleration leading to an inward movement of the gas. The potential energy is converted to kinetic energy from the inward movement and then into internal energy in the gas by friction. That internal energy is then released as thermal radiation. The virial theorem tells us that half of the potential energy released is converted into kinetic energy. One half of the potential energy goes to the rotational energy of the disk, while the other half of the potential energy can be converted into internal energy.

For a geometrically thin, optically thick accretion disk we can write the change in energy when a particle of mass m falls from radius $r + \Delta r$ to r

$$\Delta E = \frac{GM_{BH}m}{r} - \frac{GM_{BH}m}{r + \Delta r} \approx \frac{GM_{BH}m}{r} \frac{\Delta r}{r} \quad (17)$$

We assume that the black hole is gravitationally dominant in this case, neglecting the self-gravity of the disk. The virial theorem implies half of this energy is converted to heat. Assuming local emission the luminosity due to energy change is

$$\Delta L = \frac{GM_{BH}\dot{M}_{BH}}{2r^2} \Delta r \quad (18)$$

The mass accretion rate is independent of the radius otherwise there would be a buildup of mass at a certain radius. For a disk that is optically thick, the local emission corresponds to that of a black body. We can use Stefan-Boltzmann's law which states that the black-body radiant emittance is directly proportional to the fourth power of the black body's temperature to determine the luminosity a ring between r and $r + \Delta r$ emits

$$\Delta L = 2 \times 2\pi r \Delta r \sigma_{SB} T^4(r) \quad (19)$$

Where σ_{SB} is the Stefan-Boltzmann's constant and the factor of 2 comes from the fact that the disk has a bottom and top side. Combining equation 18 and 19 we can determine the temperature profile of the disk

$$T(r) = \left(\frac{GM_{BH}\dot{M}_{BH}}{8\pi\sigma_{SB}r^3} \right)^{1/4} \quad (20)$$

This approximate derivation using a geometrically thin, optically thick disk is close to the more accurate derivation that consider friction dissipation and heating also partially moved inwards in the disk. The correction factor is as follows

$$T(r) = \left(\frac{3GM_{BH}\dot{M}_{BH}}{8\pi\sigma_{SB}r^3} \right)^{1/4} \quad (21)$$

This equation holds for values $r \gg r_S$. It is possible to scale the temperature profile with r_S and introducing the definition from equation 15 we obtain the following equation

$$T(r) = \left(\frac{3GM_{BH}\dot{M}_{BH}}{8\pi\sigma_{SB}r_S^3} \right)^{1/4} \left(\frac{r}{r_S} \right)^{-3/4} = \left(\frac{3c^6}{64\pi\sigma_{SB}G^2} \right)^{1/4} \dot{M}_{BH}^{1/4} M_{BH}^{-1/2} \left(\frac{r}{r_S} \right)^{-3/4} \quad (22)$$

A couple of points become clear from inspecting equation 22. The temperature profile is not directly dependent on the viscosity and the temperature increases with a decreasing radius as $T(r) \propto r^{-3/4}$. We can approximate the emission of the accretion disk as a series of concentric black body rings at different temperatures, where temperatures are higher for higher accretion rate and lower for more massive BH. The spectrum we observe will not be a blackbody, but a broader and more flat distribution where the lower and upper bound of the frequency interval is determined by the lowest and highest temperature, which correspond to the outer and inner radius of the disk. The radiation from the accretion disk is continuum radiation and can be described by a power law continuum in the optical-UV/soft X-ray range of the spectrum. The mass dependence of equation 22 shows that more massive BH will have lower temperatures for a fixed ratio of (r/r_S) and this suggests that stellar mass object will have hotter accretion disk than a AGN BH and this is corroborated by the fact that the accretion disks of stars and stellar-mass BH emit in the hard X-ray range of the spectrum while AGN thermal disk emission goes to the UV/soft X-ray range.

1.2.1.3 The Broad Line Region

The Broad Line Region (BLR) is where dense gas clouds ($n_H \geq 10^{10}$ atoms cm^{-3}), moving at thousands of $\frac{km}{s}$, are located. The emission lines (primarily from hydrogen) we observe from that region is the sum of the Doppler-shifted emission from all the individual gas clouds. The BLR is in a halo surrounding the accretion disk and has strong line emission with velocity widths from 500 $\frac{km}{s}$ up to 10000 $\frac{km}{s}$, this dispersion is Doppler broadening due to the rotation of the BLR and can cause line blending leading to features in the AGN spectra to become unresolved. velocities of that magnitude suggest strong gravitational forces like those that would be present near a SMBH. For emission lines originating a distance r from the center the rotational velocity (assuming Keplerian rotation) would be

$$v_{rot} \sim \sqrt{\frac{GM_{BH}}{r}} = \frac{c}{\sqrt{2}} \left(\frac{r}{r_S} \right)^{-1/2} \quad (23)$$

For velocities that are at $v=c/30$ this suggests a origin point at a distance of $r \sim 500r_S$. For most Seyfert nuclei we observe continuum radiation shortward of the Lyman limit ($\lambda < 912\text{\AA}$). These continuum photons would be absorbed if they went through the BLR gas, so the BLR must only cover a small fraction of the central source. A comparison of continuum radiation and line emission gives the fraction of ionizing continuum photons which are absorbed by the BLR clouds. The BLR absorbs 10% of the emitted AGN energy and as the BLR clouds are optically thick they will cover 10% of the AGN solid angle.

1.2.1.4 The Narrow Line Region

Narrow emission lines can be observed from AGN such as [OII] and [OIII], these come from forbidden transitions that can only be observed in low density environments ($n_H \leq 10^8$ atoms cm^{-3}) as they would

be suppressed by collisional interaction otherwise. The velocity widths of these lines are below $1000 \frac{km}{s}$ and have not been observed to vary with the brightness of the nucleus of the AGN, this suggest that they originate from a region further away from the nucleus than the BLR. This region of origination is called the Narrow Line Region (NLR) and is generally a few kpc across, though in some cases ionized gas regions have been observed hundreds of kpc from the center nucleus. The gas is most likely shining in response to the active nucleus and material being ionized due to massive stars close to the AGN.

1.2.1.5 The Dusty Torus

⁴ Farther out from the accretion disk, a dusty torus is encompassing the inner region. The torus can block the BLR emissions from reaching us if we observe the object at an angle looking into the torus (reference figure 1). In fact the obscuring effect of the torus not only affects the optical broad lines but also the observed in X-ray and optical/UV continuum emission from the accretion disk and the environment close to it. Observation of polarized light in Seyfert 2 galaxies that showed broad emission lines in the optical while having no BLR emission (which is initially unpolarized) suggest that a dusty torus is responsible for scattering the BLR light (Nenkova *et al.*, 2002). An example of this is NGC1068 where a BLR is present but our direct view of it is obscured by absorbing material (Miller & Antonucci, 1983). The absorbing material will not fully encompass the BLR in every direction but only a solid angle of $<4\pi$ as seen from the central core. The direct light from the AGN completely outshines the scattered light, which is the reason why we cannot identify the BLR emission in the total flux. By observing the polarized light the direct radiation is suppressed and the BLR is then visible in the scattered light. It should be noted that not all AGN without direct BLR emission show a polarized flux, which could be due to the lack of a sufficient scattering medium for the BLR to be visible or the total lack of a BLR. By considering the abundances of galaxies with and without direct BLR emission (Seyfert 1 and Seyfert 2 galaxies) which is around a ratio of 1:2 we can determine an estimate of the solid angle fraction in which the view to the BLR is obscured, as seen from the AGN. The ratio shows that around 2/3 of the solid angle is covered by an absorbing material. This absorbing material is what we assume to be a thick dusty gas torus.

The shape of the dusty torus was originally though to be a big and smooth distribution of gas and dust, but by studying the IR SEDs of quasars the belief today is that the torus is less of a donut shape and more of a general circum-nuclear, geometrically- and optically-thick dust distribution with a characteristic scale height of $h/r \sim 1$ (Hoenig *et al.* (2010) & Hoenig & Kishimoto (2010)). The observed hydrogen column densities of optically obscured type 2 AGN range from around $10^{23} - 10^{25} cm^{-2}$ and is believed to come from the torus, accounting for most the mass budget there. The torus is thought to be an order-of-magnitude larger than the BLR and is filled with rather dense clouds where the dust reprocesses the optical/UV photons of the accretion disk and re-emits the received energy in the IR. This makes it possible to study the torus directly by studying the IR emission from AGN, except for the case where the AGN is radio-loud. It is possible to model the torus clouds by assuming a certain size distribution and observing if the models can reproduce the observed IR spectral shape. Taking into account the variation within individual AGN, the modelling suggest that around 25% of the total power emitted by an AGN emerges in the IR. This would naively imply that the material in the torus blocks the light in

⁴This section relies on Hoenig *et al.* (2010) & Hoenig & Kishimoto (2010)

a quarter of all directions, in reality this estimate is likely too simple. There is likely a broad probability distribution of absorbing optical depths for light rays traversing the torus region (Schneider, 2006).

The dust in the torus will thermally re-radiate the received energy in the IR at a temperature from 100K to around 1500K which is the dust sublimation temperature. Temperatures in that range are possible at a distance of 0.1 pc to 100 pc depending on the actual AGN luminosity, torus geometry, and dust composition (Hoenig & Kishimoto, 2010). The thermal radiation of the dusty torus gives rise to a "red bump" in the near- to mid-IR emission with a cut-off point around $1\mu m$ where the "Big Blue Bump" from the UV/optical accretion disk emission starts to take over. The Big Blue Bump spectral energy can be approximated as $\lambda F_\lambda \propto \lambda^{-4/3}$. The λF_λ energy-peak of the AGN is located in the UV. Most of the energy is emitted at those wavelengths and combined with the dust absorption being the strongest in the UV, it must mean that the driving light curve is in the UV. In type 1 AGN broad spectral line at around $10\mu m$ and $18\mu m$ caused by hot silicate dust is assumed to come from the inner part of the torus, though it should be noted that these silicate features also appear in absorption in most type 2 AGN where only cooler dust is observed.

As the torus cloud temperature and total intensity decrease with distance clouds at different distances from the AGN contribute different fractions to the total torus flux at a given wavelength. In general, clouds at the innermost torus region contribute most of the near-IR emission while clouds at larger radii are the dominant source of the mid-IR emission. It is possible to approximate the relationship between the radius and temperature of the torus as (Hoenig & Kishimoto, 2010)

$$\frac{r}{r_{sub}} = \left(\frac{T}{T_{sub}} \right)^{-2...-2.8} \quad (24)$$

for dust grains in radiative equilibrium. T_{sub} is the dust sublimation temperature (around 1500 K) and T is the dust temperature at a distance r from the AGN center. The range given for the exponent covers the range from black body grains (large, graphite) to typical ISM dust. The inner part of the torus will be at the dust sublimation temperature, so by knowing the thermal time lag it is possible to determine the dust sublimation radius r_{sub} .

Mid-IR imaging, interferometry and sub-mm observation constrain the torus size to a range of 0.1-10 pc. Radio observations also indicate a sub-pc scale maser disk which is a compact concentration of clouds orbiting the SMBH and emitting in the 22 GHz maser line. The general assumption is that the maser disk is occupying the same space as the dusty torus, but it is unclear whether it corresponds to its innermost part or to a geometrically thin disk which inflates in the outer part (Almeida & Ricci, 2017). The torus is interesting for us since it can be used as a standard ruler like the megamaser disk approach mentioned earlier, since the torus responds to the variations of the accretion disk with a time delay and can be probed using reverberation mapping. The luminous nature of AGN allows to determine the distance to object at redshift of $z=4-6$ (Watson *et al.*, 2011).

1.2.1.6 The Corona

One part of the AGN not shown in figure 1 is the corona. Through variability studies, along with gravitational microlensing, the inner region of AGN can be probed. The blackhole is surrounded by a small and hot (109K) spherical corona and appears pointlike relative to the accretion disk [Edelson *et al.* \(2015\)](#). X-ray emission in AGN is believed to be produced by Comptonization (Compton up-scattering) of optical/UV accretion disk photons by a corona of hot electrons located above and below the SMBH ([Haardt & Maraschi, 1991](#)). Some evidence suggest that the X-ray leads the variability in changing look AGN where the seyfert type changes from 2 to 1. It is possible but rare for changes to be so large that the broad-line components of $H\alpha$ and $H\beta$ appear in a Seyfert 2 or vanish from a Seyfert 1. Large changes in X-ray emission might be from the central BH accreting stars from the surrounding galaxy core. The natural explanation of the observations seems to be variability driven by X-ray irradiation of the BH accretion disk. It should be noted that models trying to describe this phenomenon does not include any dust reprocessing. If there is significant reprocessing of UV and optical photons on scales larger than the BLR, then the NIR would respond to the increase in X-ray flux on two distinct timescales, whereas only one is currently included in models of the region ([Shappee *et al.*, 2014](#)). Due to the fact that the corona has a relatively small part of the overall mass budget of the AGN it is unlikely that the corona X-ray emission could drive the response we observe in the AGN. From investigations of the Kepler AGN light curves the X-ray reprocessing is also unlikely to be a large contributing factor to the 0.1-10% optical variability of AGN ([Smith *et al.*, 2018](#)).

1.2.2 Categorizing AGN

⁵ AGN can be categorized into a couple of different objects for example the Seyferts which are in spiral galaxies. Seyferts are further split into Seyfert 1 and Seyfert 2 galaxies, where the first has BLR emissions and forbidden lines and the second only has NLR emission. Quasars are a sub-type of AGN that have an optical magnitude of $M_V < -22.5$. They are so bright that they can outshine their entire host galaxy and appear, as their name suggest, quasi-stellar when observing them. The luminosity of a Quasar can in some cases be $\sim 10^3$ higher than their host galaxy, so typically the host galaxy can only be observed at low redshifts. Quasars are in many ways similar to Seyferts, though they have weak stellar absorption features and have less NLR to BLR emission relative to Seyferts. Radio galaxies are defined as galaxies with power $P_{1.4GHz} \geq 2 \times 10^{23} WHz^{-1}$ and are almost all AGN ellipticals that like Seyferts can be classified into two groups the Broad-Line Radio Galaxies (BLRG) and Narrow-Line Radio Galaxies (NLRG). Besides being radio loud, which the Seyferts are not, these AGN can also have radio jets stretching out to kpc or Mpc from the host galaxy.

The differences in AGN can be explained by unification theory. The idea behind unification theory is that the observed differences in properties between type 1 and type 2 AGN are due to the inclination angle at which we observe the different systems instead of anything intrinsic to the AGN. Unification theory assumes that the reason behind the observational differences is that the AGN are axisymmetric like solar systems or the galaxies they reside in. Despite the differences between type 1

⁵This section relies in part on [Peterson \(1997\)](#) and [Bianchi *et al.* \(2012\)](#)

and 2 AGN there are also many similarities like the fact that both types of Seyfert galaxies have narrow line emission and corona line emission.

1.3 Reverberation Mapping

Simple models of the AGN detailing their individual parts can help to explain certain observed features in the AGN spectra, but they leave the specific geometry and dynamics largely unknown. Sub-parsec scale structures are unresolved in even the closest AGN, so any additional information about the structure must be obtained by indirect means (Fausnaugh *et al.*, 2014). Reverberation mapping (or Echo mapping) is the tool that allows us to probe the structure and kinematics of compact AGN structure on these scales without being limited by telescopic spatial resolution. It utilises spectroscopic and photometric observation to determine the light travel time between the different parts of the AGN more specifically the accretion disk and the outer parts of the AGN (Nuñez *et al.*, 2018).

There are two way to infer the structure/size of the torus either by Reverberation Mapping or by IR interferometry.

The principle behind reverberation mapping is to look for time-variable flux signals and their echoes in the light curves since they contain information related to the unresolved structures of the object. The reverberation mapping approach is mainly built upon three⁶ assumption about the object being observed. Those being (Peterson (1993) & Peterson (2001))

- The continuum originates in a single central source of isotropic radiation.
- The light-travel time $\tau = r/c$ is the most important time scale compared to dynamic timescale, which describe significant changes in the geometry and is assumed to be a great degree larger than τ or the timescale to reprocess the driving light, which is assumed virtually instantaneous.
- There is a simple, though not necessarily linear relationship between the observed continuum and the ionizing continuum.

An example of reverberation mapping are seen with the BLR where variations from the ionizing continuum flux from the accretion disk are reprocessed in the BLR gas to variable emission line flux with a time delay that scales with the BLR light crossing time. Measuring the time delay gives an estimate of the spatial extent of the BLR. An other example that is relevant for this project is how the dusty torus reprocesses the variations from light from the UV driving light curve of the AGN into variations in the different continuum bands, most prominently the NIR. This means that if the time delay of the torus can be measured between the UV/Optical and near-IR it can give us an estimate of the distance between the central engine and the inner radius of the dusty torus, which is the dust sublimation radius. This would be an estimate of the physical size of the AGN torus since the time delay can be interpreted as the light travelling time $r \sim c \times \tau$ where τ is the time delay. This is one part of the two parts needed for an object to be used as a standard ruler.

⁶For more criteria specific to the BLR, see Peterson (1993)

Lag times of 10-100 light days corresponding to 0.01-0.1 pc for the inner dust torus where the hot dust sublimates were obtained for several Seyfert galaxies from UV (with typical luminosity's of $10^{42} - 10^{44} \text{ erg s}^{-1}$). From optical and near-infrared monitoring observations, a strong correlation between lag time and both the UV and optical luminosity's was found to be consistent with $\Delta t_{dust} \propto L^{0.5}$. From obtaining lags from both the dusty torus and the BLR the lags for torii seem to be 4-5 larger than the ones from the BLR (Hönig (2014) & Mandal *et al.* (2019)). The equation for the dust sublimation temperature is (Koshida *et al.*, 2014)

$$R_{sub} = 1.3 \left(\frac{L_{UV}}{10^{46} \text{ erg/s}} \right)^{0.5} \left(\frac{T_{sub}}{1500\text{K}} \right)^{-2.8} \left(\frac{a_{dust}}{0.05\mu\text{m}} \right) \text{ pc}$$

where L_{UV} , T_{sub} and a_{dust} are the UV luminosity of the accretion disk, sublimation temperature of dust, and dust grain size, respectively. The sublimation temperature might be closer to 1700K and the grain size might be closer to $1\mu\text{m}$ instead of 1500K and $0.5\mu\text{m}$. The inner radii found with reverberation mapping seem to be systematically smaller by a factor of 3 from what we would expect from the dust sublimation radius for a temperature of 1600 K, though this discrepancy can be explained by a combination of many factors like the dust grain size, clumpy dusty torus, anisotropic illumination of the accretion disk, winds/outflows from the accretion disk etc. (Mandal *et al.*, 2019).

It is also possible to do reverberation mapping of the accretion disk itself. Depending on the temperature of the accretion disk, the time delays for different continuum bands can be interpreted as the light travel time across the accretion disk, so to a first approximation the time delays can give information about the size of the accretion disk and the temperature gradient across the accretion disk which are both important parameters to constrain for theoretical models of the accretion disk (Nuñez *et al.*, 2018). The standard geometrically thin, optically thick accretion disk (the α -disk model) has a temperature profile of $T \sim r^{-3/4}$ resulting in a spatial dependence of the black body emission of the accretion disk. The inner parts of the accretion disk will emit UV photons ($\sim 10 - 3000\text{\AA}$) and the cooler outer parts will emit in the optical and NIR ($\sim 3000 - 10000\text{\AA}$). Variations in the X-ray from the corona or the inner edge of the accretion disk can irradiate the outer parts of the accretion disk and drive variations at longer wavelength. This mean that there is an expected time delay between the UV and the optical continuum variation and they will be proportional to the size of the accretion disk (Fausnaugh *et al.*, 2014). While the -3/4 exponent is consistent with the 2σ level, Fausnaugh *et al.* (2014) and the physical modeling of Starkey *et al.* (2017) suggest that the data strongly prefers a steeper temperature profile of $T \sim r^{-1}$. This shows the ability of reverberation mapping to inform us about the disk structure without directly being able to resolve the accretion disk.

One of the main difficulties in applying the reverberation mapping method is to obtain continuous, well-sampled (of order 1 day or less) light curves on time scales of weeks to months or even years in the case of investigating the torus. Having to collect data over years is expensive and it is often not possible to have data that is sampled on the order of 1 day. Even more limiting is the fact that there might not be data available from the UV light curve that drives the response in the other optical/NIR continuum bands. The attempt is to try and provide a way to model the UV driving light curve/function and obtain an estimate of it without any UV data. If the lag (τ) is obtained it can be used to determine the angular diameter distance to the the AGN, accounting for time dilation and small angles (Hönig

et al. , 2014)

$$d_A(Mpc) = \frac{0.173\tau(days)}{(\delta\theta(mas)(1+z))}$$

The idea originally was to use the BLR, but its angular size is too small to be spatially resolved with todays equipment. The dust continuum emission from the dusty torus is 4X larger and it requires only photometric reverberation mapping.

1.3.1 The transfer function and convolution

⁷ To first order, the problem of torus reverberation mapping can be formulated as a deconvolution problem in which the flux in the continuum light curve (the reprocessor light curve), at a given wavelength λ , $F_c(t, \lambda)$ is given by a convolution of the AGN UV driving continuum light curve (driving function), over some wavelength range, $F_{driv}(t)$ with a transfer function $\Psi(\tau, \lambda)$ that encodes the physics and geometry of the torus,

$$F_c(t, \lambda) = \int_{-\text{inf}}^{\text{inf}} \Psi(\tau, \lambda) F_{driv}(t - \tau) d\tau \quad (25)$$

the transfer function is assumed to be linear. Compare this to the BLR where the flux in the emission-line light curve, at a given wavelength λ , $F_l(t, \lambda)$ is given by a convolution of the AGN continuum light curve, over some wavelength range, $F_c(t)$ with a transfer function $\Psi_{BLR}(\tau, \lambda)$ that encodes the physics and geometry of the BLR instead. The main difficulty we face with reverberation mapping of the torus compared to the BLR is the fact that we only have F_c where for the BLR it is possible to have both F_l and F_c .

Generally convolutions can be thought of as using one function to perform smoothing and averaging on another function. The convolution of two functions $g(t)$ and $f(t)$ is defined as (Osgood, 2007)

$$(g * f)(t) = \int_{-\text{inf}}^{\text{inf}} g(t - x) f(x) dx \quad (26)$$

The discrete convolution operation is defined as

$$(g * f)[n] = \sum_{m=-\infty}^{\infty} g[m] f[n - m] \quad (27)$$

with m and n being entries in the arrays g and f.

To get an idea of what convolution does it is instructive to think of the "flip and drag" example. Start by fixing a value of t so that the graph of the function $g(x - t)$ is the same shape as $g(x)$ but shifted to the right by t. Flip the function graph (left-right) about the line $x=t$ by forming $g(t - x)$. Next, multiply the two functions $f(x)$ and $g(t - x)$ and integrate with respect to x, this is the "dragging" part of the example. compare this to cross correlation where the function is not flipped before it is dragged. Averaging is being done due to the fact that computing the convolution $g * f$ at t replaces the value $f(t)$ with the weighted average of the values of f near t. The function f is being weighted by the values of the flipped and shifted function g . The smoothing comes into effect as changing t or "dragging" $g(t - x)$

⁷This section relies on Skielboe *et al.* (2015), Hönig *et al.* (2014) and Pancoast *et al.* (2011)

through different values of t repeats the operation. Moving t small amount to t' will mean $(g * f)(t')$ will be a new weighted average near t' with values of f that were part of the average near t . This means the two convolutions $(g * f)(t)$ and $(g * f)(t')$ will be closer to each other than the functions $f(t)$ and $f(t')$ and therefore the convolution is "smoothing" the function f as t changes. Keep in mind for these examples that any value of $(g * f)(t)$ includes all the values of f and g and if both functions reach a value of zero at some point then the convolution will as well, but if either function does not reach zero then neither will the convolution.

The Convolution Theorem is defined as follows

$$\mathcal{F}(g * f)(s) = \mathcal{F}g(s)\mathcal{F}f(s) \quad (28)$$

where \mathcal{F} is the Fourier transform. This states that convolution in the time domain corresponds to multiplication in the frequency domain.

A lot of work has been done on reverberation mapping the BLR and there exist multiple approaches for determining the BLR size and transfer function and it is worthwhile to consider them when trying to determine a method to obtain the transfer function:

- Directly using the convolution theorem of Fourier transforms, this requires very high data quality.
- Maximum entropy method that finds the solution for the transfer function that has the highest entropy, while still providing a good fit, it is computationally expensive, relies on assumptions about transfer function shape and it is hard to do extensive error analysis and model comparisons.
- The Subtractive Optimally-Localized Averages (SOLA) method, estimates the 2D response as a weighted average of the emission line light curve data points.
- Dynamical modelling in which a simplified physical model of the BLR is constructed, and its parameters are inferred within the framework of Bayesian statistics, requires long computation time and flexibility for good BLR description.
- Regularized Linear Inversion (RLI) rewrites equation 25 into a linear matrix equation, $L_{\Delta\lambda} = \Psi_{\Delta\lambda}C$ where $L_{\Delta\lambda}$ is the emission line light curve integrated over the wavelength range (spectral bin) $\Delta\lambda$, C is a matrix of continuum light curves, and $\Psi_{\Delta\lambda}$ is the transfer function corresponding to the given wavelength range. For noise free data solving the discrete transfer equation would just involve inverting C to obtain the transfer function Ψ . Due to noise the approach is instead to minimize χ^2 together with a smoothing condition so as to not fit the noise. This method has very few assumptions about transfer function shape and is a generally fast

From these examples we see that there are generally two classes of methods depending on the strategy adopted in solving the integral equation that links the variations of continuum and emission lines in RM. One class relies on directly inverting the integral equation, including early attempts based on the Fourier transform and RLI methods. The method using the Fourier transform was found to require high-fidelity data and has therefore not seen much application [Li et al. \(2016\)](#), whereas RLI seeks to

discretize the integral equation and employ a differencing operator to suppress noise. The downside of this class of methods is that measurement noises cannot be self-consistently incorporated. The result that this class may be very noise sensitive in some cases. The other class relies on indirectly inferring the “best” transfer function that fits the data by supposing prior limits or models on it. The maximum entropy method and BLR dynamical modeling method belong to this class. The maximum entropy method finds the “simplest” solutions for the transfer function by maximizing the entropy on the premise of a reasonable fit to the data. The method is limited by the fact that it does not allow for straightforward uncertainty estimates or model selection.

One example of these methods used on the torus is from [Landt *et al.* \(2019\)](#) which uses MEMEcho. This is a formalism based on the maximum-entropy method and infers the lag distribution given a driver and a reprocessor light-curve. MEMEcho fits a linearised echo model

$$\begin{aligned}
 F_{UV}(t) &= \bar{F}_{UV} + \Delta F_{UV}(t) \\
 F_v(t) &= \bar{F}_v + \Delta F_v(t) \\
 F_v(t) &= \bar{F}_v + \int_{-\text{inf}}^{\text{inf}} \Psi_v(\tau) \Delta F_{UV}(t - \tau) d\tau
 \end{aligned}$$

where both the observed driver light-curve, $F_{UV}(t)$, and the observed reprocessor light-curve, $F_v(t)$, are assumed to be composed of a constant and a variable component. The total flux of the reprocessor is modelled as the sum of the constant mean reprocessor flux, \bar{F}_v , and the convolution of the response function (the mean subtracted transfer function), $\Psi_v(\tau)$, with the variable part of the driver light-curve, $\Delta F_{UV}(t)$. MEMEcho can then recover a combination of the functions $\Psi_v(\tau)$, \bar{F}_v and $\Delta F_{UV}(t)$. Multiple combinations will satisfy the constraint of the data, so the selection is made based on the maximum total entropy of all three functions summed. MCMCRev (see CREAM from [Starkey *et al.* \(2016\)](#) for a similar approach) then applies a MCMC technique to sample posterior parameter distributions for a model response function given the driver and the reprocessor light curves. The response function is modelled with a log normal distribution and with a temperature-radius relationship for a standard accretion disc seen face on ($T \propto R^{-3/4}$). The relationship comes as a result of models of the dusty torus as a radiatively accelerated outflow launched from the outer regions of the accretion disc ([Landt *et al.*, 2019](#)). The result is that the temperature-radius relationship translates into a wavelength-lag relationship ($\tau \propto \lambda^{4/3}$).

The aim of this project is measure the distance to the inner part of the dust torus by assuming a parametric form for $\Psi(\tau, \lambda)$ and determine the best fit parameters by convolving the driving light curve with the transfer function to obtain a model response light curve that can then be compared to the data. This method provides a complete, although simple model for the reverberation mapping problem and it has the benefit of providing well defined error estimates for the model parameters. A similar approach is done by the program JAVELIN ([Zu *et al.*, 2013](#)), which uses a top hat model for the transfer function and includes modeling of the continuum light curve to determine a likelihood distribution for the time delay.

1.3.1.1 The shape of the transfer function

While JAVELIN uses a top hat function to model the transfer function, the aim is to provide a more detailed version of the transfer function for the model to more accurately describe the geometry of the AGN. Multiple independent approaches for reverberation mapping of the BLR has determined that the general shape of the transfer function is that of a log normal distribution see [Grier et al. \(2013\)](#), [Skjelboe et al. \(2015\)](#), [Li et al. \(2016\)](#), [Grier et al. \(2017\)](#) and [Starkey et al. \(2017\)](#). [Landt et al. \(2019\)](#) also assumes a lognormal shape for their transfer function when trying to map the torus of NGC 5548. It should be noted that some of these approaches are numerical in nature and therefore do not assume anything about the shape of the transfer function beforehand, leading more credence to the fact that the general shape of the transfer function does not come as a result of our prior held belief about its shape. Figures 2 and 3 show an example of a velocity resolved transfer function for the BLR and a comparison of transfer function from the different methods used respectively.

The general formula for the probability density function of the lognormal distribution is ([NIST/SEMATECH, 2019](#))

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{\ln(\frac{x-\theta}{m})^2}{2\sigma^2}\right) \quad x > \theta; m, \sigma > 0 \quad (29)$$

where σ is the shape parameter (and is the standard deviation of the log of the distribution), θ is the location parameter and m is the scale parameter (and is also the median of the distribution). It is common to parameterize the lognormal distribution using the conversion

$$\mu = \ln(m)$$

where μ is the mean of the log of the distribution. If the μ parameterization is used, the lognormal distribution becomes

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x-\theta) - \mu)^2}{2\sigma^2}\right) \quad (30)$$

Using a single lognormal as the parametric form of the transfer function will in fact not be entirely accurate, since both the accretion disk and the dusty torus contributes to the resulting response light curves we observe, meaning they both contribute to the transfer function. A more accurate form would be a sum of two lognormals, one for the accretion disk and one for the dusty torus,

$$\Psi(\tau, \lambda) = \frac{N_{DT}}{\tau\sigma_{DT}\sqrt{2\pi}} \exp\left(-\frac{(\ln(\tau - \theta_{DT}) - \mu_{DT})^2}{2\sigma_{DT}^2}\right) + \frac{N_{AD}}{\tau\sigma_{AD}\sqrt{2\pi}} \exp\left(-\frac{(\ln(\tau - \theta_{AD}) - \mu_{AD})^2}{2\sigma_{AD}^2}\right) \quad (31)$$

where DT stands for Dusty Torus and AD stands for Accretion Disk. N_{DT} and N_{AD} are the normalization functions for the torus and accretion disk respectively. The two lognormals have three free parameters each, meaning each observation band will have 6 parameters (disregarding the normalization functions for now). The goal of this project is primarily to determine μ_{DT} , since it describes the median lag time to the dusty torus. If the mode or the mean are more desirable then they can be obtained by knowing σ_{DT} as well, since the mode of the lognormal distribution is $\exp(\mu - \sigma^2)$ and the mean is $\exp(\mu + \frac{\sigma^2}{2})$.

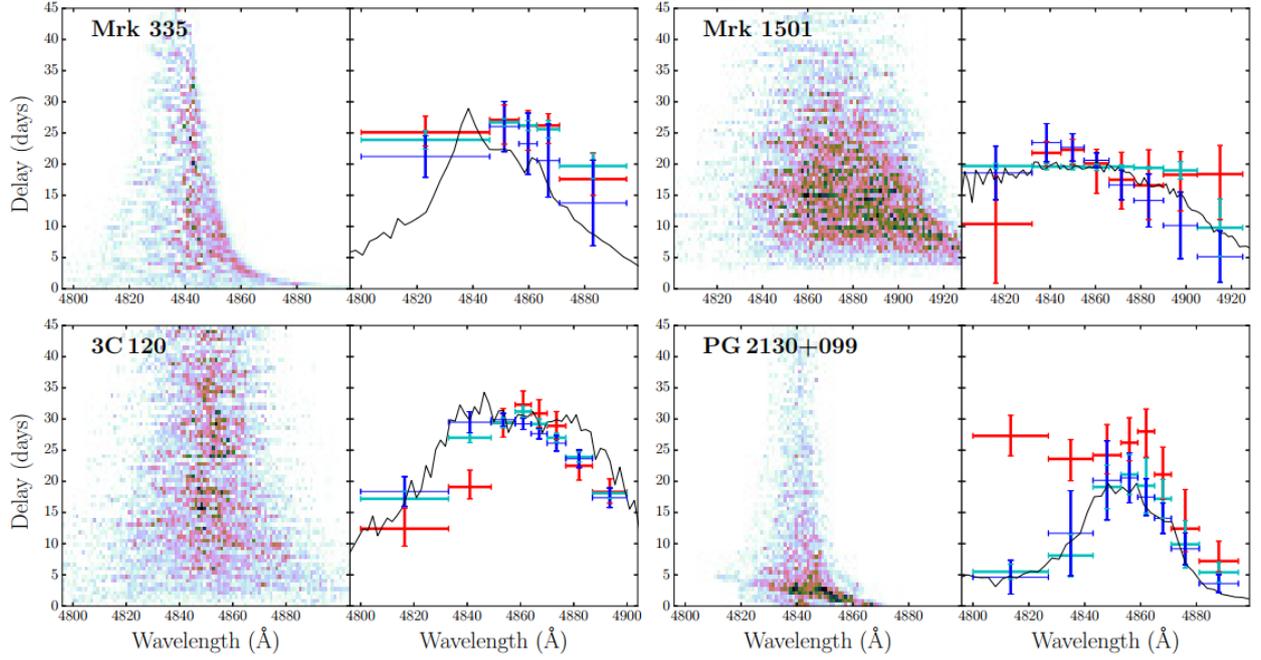


FIG. 4.— Transfer functions and velocity-resolved time delays for all four AGN. For each source, the left panel shows a representative transfer function drawn from the posterior PDF. The right panel for each source shows the velocity-resolved time delays for a number of wavelength bins. The solid black line shows the mean time delays computed by the model corresponding to the transfer function shown, and the blue crosses show the median lag values recovered from all of the model fits. To compare the model lags (blue and black) with the lags measured via cross correlation techniques, we also show mean time delays measured via cross correlation of the continuum light curves with light curves generated from the decomposed spectra (red crosses). We also cross correlated the model spectra from all of the samples with the continuum light curve and measured the median time lag in each bin (cyan crosses).

Figure 2: Example of log normal transfer function for the BLR. Figure and text is originally from [Grier et al. \(2017\)](#)

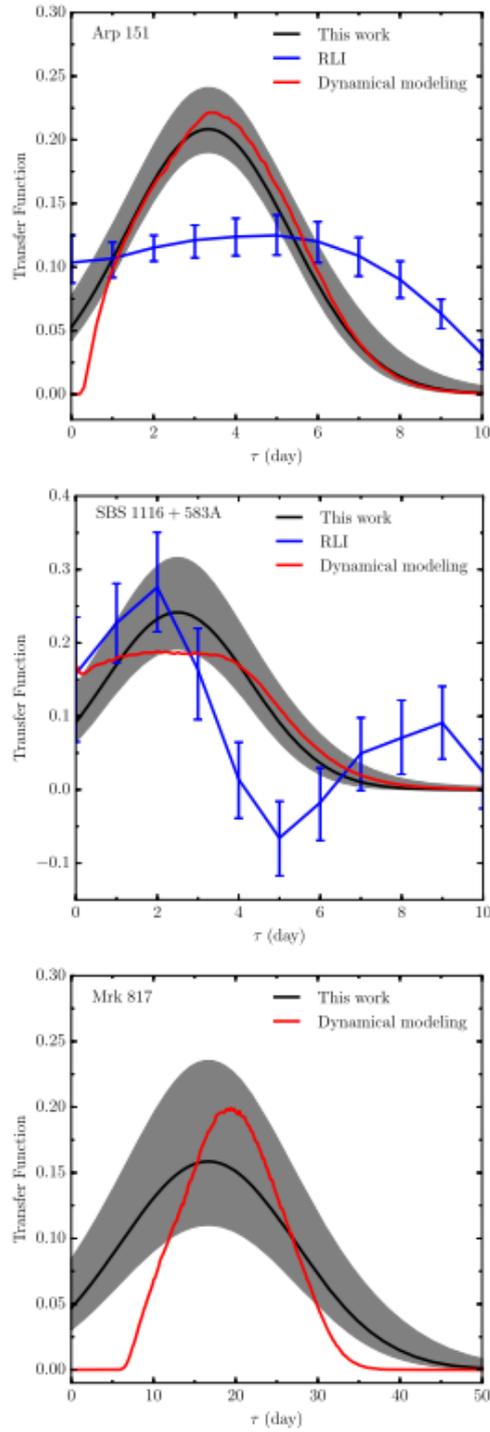


Figure 9. Comparisons of the transfer functions derived from different methods for Arp 151 (top), SBS 1116+583A (middle), and Mrk 817 (bottom). The black line with shaded areas is the transfer function from this work, the blue line with error bars is from the regularized linear inversion (RLI) method of Skielboe et al. (2015), and the red line is from the BLR dynamical modeling method of Li et al. (2013). Note that there is no RLI solution for Mrk 817. All the transfer functions are normalized on a common scale.

Figure 3: Comparisons of transfer functions for the BLR obtained using different methods. Figure and text originally from Li *et al.* (2016) showing their method for obtaining the transfer function which is to write it up as a sum of Gaussians.

1.3.1.2 The amplitude of the transfer function

The transfer function needs to be normalized if it is to be used to create response light curves that can be compared with data. It is also necessary to take the contribution from the accretion disk and dusty torus into account, since their normalization differs from each other and both are dependent on the wavelength range we are observing. It is assumed that the normalization of the accretion disk transfer function is a power law of the form

$$N_{AD} = F_{AD}(\lambda) = K_0 \left(\frac{\lambda}{\lambda_0} \right)^i \quad (32)$$

where K_0 is a normalization constant, λ_0 is a normalization wavelength chosen to be the same order of magnitude as the wavelengths of the different observational bands we are considering and i is the power law index that is typically in the range $-2 \lesssim i \lesssim -1$ (Peterson, 1997) when considering flux values measured in wavelengths. This normalization is used for the accretion disk because the observed characteristic of accretion disk and broad band SEDs of AGN continuum can generally be described by a power law. The wavelength λ chosen for each observation band will be the "pivot" wavelength at the midpoint of the band, since it would be very computationally costly to have a version of the transfer function for each wavelength in an observational band.

The normalization of the dusty torus is informed by our knowledge of the region. Equation 24 shows that the torus has a decreasing temperature with an increase in the distance from the center of the AGN. The thermal radiation of the torus is described by a series of concentric rings at varying temperatures each emitting black body radiation. For the simple parametric model used in this project, a sum of black body radiation at different temperature would be unnecessarily complicated, computationally expensive and dismisses the fact that the temperature and size of the torus is unknown beforehand. It is assumed that the torus region is uniform with a constant temperature and density, radiating as a single black body. This approximation is in fact fairly accurate while only requiring one free parameter, the black body temperature T . Figure 4 shows how well a combined black body spectrum made up of 10 black bodies of varying temperature can be approximated by a black body spectrum with a single temperature. The black body radiation, which is the power emitted per unit area of the body, per unit solid angle of emission, per unit wavelength, is described by Planck's law of radiation

$$BB(\lambda, T) = \frac{2hc^2}{\lambda^5} \left(\exp\left(\frac{hc}{\lambda k_B T}\right) - 1 \right)^{-1} \quad (33)$$

where h is the Planck constant and k_B is the Boltzmann constant. Since the transfer function is unitless the normalization should also be unitless. The transfer function also needs to be properly normalized to 1. Dividing the value of the black body at a specific wavelength with its peak value the normalization becomes unitless. The wavelength of the peak value of the black body is given by Wien's displacement law

$$\lambda_{peak} = \frac{b}{T} \quad (34)$$

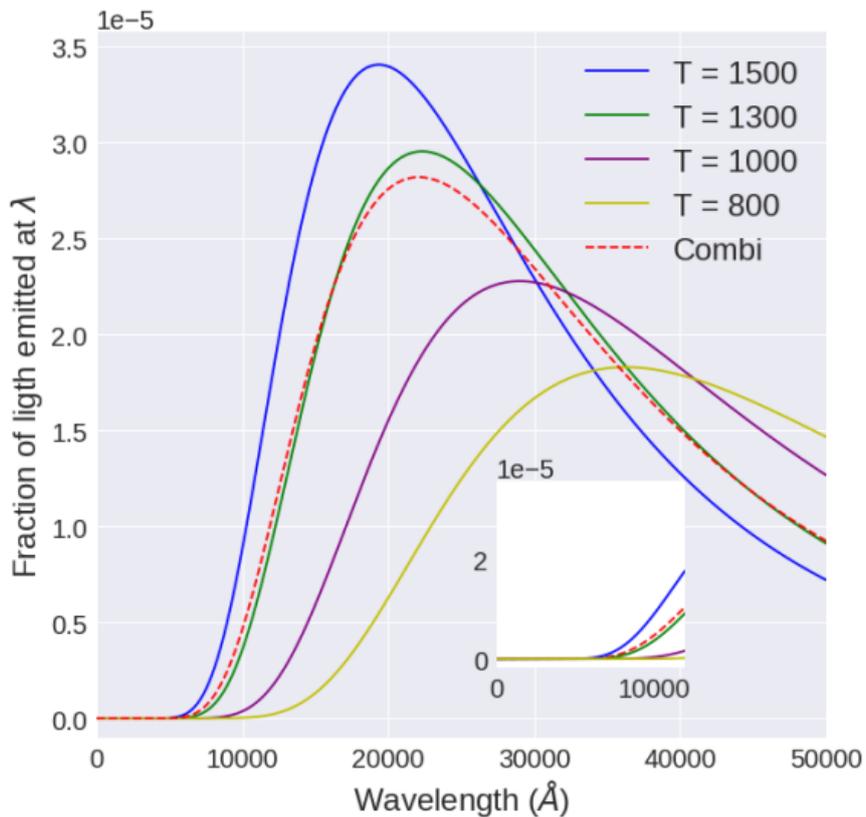


Figure 5.3: The fraction of Black Body emission being emitted at various wavelengths. The dotted red line is the simple temperature varying Dust Torus Black Body emission model, composed of 10, equally spaced temperatures of equal weight between $T = 800\text{K}$ and $T = 1500\text{K}$. The $T = 1300\text{K}$ Black Body emission aptly demonstrates the small gains at an increased computational cost by not simplifying the Torus model.

Figure 4: Figure and text is originally from [Lauritsen \(2018\)](#)

where b is Wien's displacement constant, equal to $2.898 \times 10^{-3} mK$. Inserting this peak wavelength into equation 33 reduces it to

$$BB_{peak}(T) = \frac{2hc^2}{\left(\frac{b}{T}\right)^5} (exp(4.967) - 1)^{-1} \quad (35)$$

dividing the black body by its peak value then results in the normalization of the dusty torus

$$N_{DT} = \frac{BB(\lambda, T)}{BB_{peak}(T)} = \frac{\left(\frac{b}{T}\right)^5}{\lambda^5} \frac{(exp(4.967) - 1)}{\left(exp\left(\frac{hc}{\lambda k_B T}\right) - 1\right)} \quad (36)$$

with both the normalization function of the accretion disk and the dusty torus in place, the full parametric form of the transfer function is able to be used in the deconvolution problem. The accretion disk normalization is characterized by two free parameters for each observational band, while all the torus normalization functions are characterized by the single black body temperature.

1.3.1.3 Cross correlation and the transfer function

Write about why we use the transfer function instead of the CFF.

2 Defining a Driving Light Curve

The transfer function is one piece of the puzzle when it comes to the deconvolution problem, The other is the UV driving light curve/function which drives the observed response light curves. The idea originally put forward by [Kelly et al. \(2009\)](#) working with quasars optical fluxes is to model the light curve as a stochastic process called a Damped Random Walk (DRW) using three free parameters: a characteristic timescale (decorrelation timescale), amplitude of short timescale variability, and the mean value of the light curve. Practically all quasars spectroscopically confirmed by the Sloan Digital Sky Survey are photometrically variable. Damped Random Walk (DRW) models seem to be a good statistical description of quasar variability for optical wavelengths ([Ivezi & MacLeod, 2013](#)). Since both quasars and Seyferts are AGN with the only difference being their brightness and distance from us, the application of a DRW should also work for the Seyferts. To understand how this DRW is defined and applied when simulating the driving light curve it is important to understand what an AR process is and its continuous equivalent the CAR process.

2.1 AR(1) and CAR(1) process

⁸ An Auto Regressive (AR) process describes a system whose future value can be predicted given a measure of its current state. that means that an AR process is a form of forecasting. In a typical linear regression ($y = mx + b$) it is possible to predict the value of a dependent variable (y) based on the value of the independent variable (x). In an AR process it is instead the case that the prediction of the future

⁸This section relies on [Kelly et al. \(2009\)](#), [Ivezi & MacLeod \(2013\)](#) and [Moreno et al. \(2019\)](#)

value of the system is based on the past value (or values) of the system itself. Therefore the dependent and independent variables are the same. An example of this would be

$$x_i = \mu + \phi_1 x_{i-1} + \epsilon_i \quad (37)$$

where ϕ_1 is the auto-regressive or "lag" coefficient that indicates how closely tied future values are to past values and where ϵ_i represents a source of noise. Equation 37 is a auto regressive model of the first-order or in short a "AR(1)" model where the current state of the system is linearly dependent on the immediately prior state plus a random perturbation which is described by ϵ_i . The random perturbations are an inherent part of the model and describes an intrinsic perturbation mechanism of the system in question, unlike measurement errors. An important case of the AR(1) process is the "random walk" which is when $\phi_1 = 1$ and where ϵ_i is defined by white noise. Note that the amplitude of an AR(1) system can grow arbitrarily large given enough time.

While less relevant for the model in this project a Moving Average (MA) process is when the future system value depends only on the shocks to the system. MA processes are bounded by the amplitude of the shock to the system, so they cannot grow arbitrarily large over time like the AR processes. These two types of processes can also be combined to make ARMA models, which combine short-memory AR responses and MA inputs that govern the "amplitude" of random perturbations at different timescales.

For AGN light curves the data is irregularly sampled due to night/seasonal gaps while the discrete AR(1) process is only defined for regularly sampled time series. It should also be noted that the underlying physical process that gives rise to the observed flux is continuous. To model the AGN driving light curve it is therefore necessary to use a first-order Continuous Auto Regressive (CAR(1)) process instead⁹. The CAR(1) process is described by the following stochastic differential equation

$$dX(t) = \frac{1}{\tau} X(t) dt + \sigma \sqrt{dt} \epsilon(t) + b dt \quad \tau, \sigma, t > 0 \quad (38)$$

where τ is called the "relaxation time" of the process $X(t)$, and $\epsilon(t)$ is a white noise process with zero mean and variance equal to 1. It is assumed in Kelly *et al.* (2009) that the white noise process is also Gaussian. $X(t)$ would be the flux values of the driving light curve. The mean value of $X(t)$ is $b\tau$ and the variance is $\tau\sigma^2/2$. This process is also referred to as an Ornstein–Uhlenbeck (O–U) process and is important for describing Brownian motion. The relaxation time, τ , can be interpreted as the time required for the time series to become roughly uncorrelated, and σ can be interpreted as describing the variability of the time series on timescales short compared to τ . τ could be thought of as a characteristic timescale for example the time required for diffusion to smooth out local accretion rate perturbations and σ could represent the variability resulting from local random deviations in the accretion disk structure, such as caused by turbulence and other random Magnetohydrodynamic (MHD) effects. The CAR(1) process is a DRW because it includes a damping term that pushes the time series back to its mean. This also means that the CAR(1) process will not grow arbitrarily large like the AR(1) process.

For astronomical time series analysis is it standard to interpret a light curve in terms of its autocorrelation function (ACF) and power spectrum. The autocovariance function at time t is defined to be

⁹It is possible to use an AR process if the light curve is regularly sampled, though this is rarely if ever the case.

the expected value of the product of $X(t)$ and $X(t + t')$, and the autocorrelation function is determined by dividing the autocovariance function by the variance of the time series. For a CAR(1) process the autocorrelation function is defined as

$$ACF(t') = \exp\left(-\frac{t'}{\tau}\right) \quad (39)$$

Equation 39 shows that the correlations in CAR(1) light curve fall off exponentially with lag t' , with an e-folding time equal to the relaxation time, τ . The power spectrum of a process is determined from the autocovariance function $rX(t)$ by the equation

$$P_X(f) = 4 \int_0^{\infty} rX(t') \cos(2\pi f t') dt', \quad f \geq 0 \quad (40)$$

The power spectrum or Power Spectral Density (PSD) function shows the relative power in the variability as a function of temporal frequency. in the case of a CAR(1) process the autocovariance function is $rX(t') = (\frac{\tau\sigma^2}{2})\exp(-\frac{t'}{\tau})$ so the power spectrum of a CAR(1) process is

$$P_X(f) = \frac{2\sigma^2\tau}{1 + (2\pi\tau f)^2} \quad (41)$$

Equation 41 shows that the power spectrum of a CAR(1) process falls off as $P_X(f) \propto 1/f^2$ on timescales short compared to the relaxation time, and flattens to white noise $P_X(f) \propto \text{constant}$ at timescales long compared to the relaxation time. The characteristic timescales of AGN light curves can in many cases be defined by a break in the power spectrum, lending more credence to the idea that the relaxation time τ is associated with a characteristic timescale.

The variability of AGN is a red noise process, meaning that successive samples are correlated in time (Smith *et al.*, 2018). The power spectra of these processes are well described by a power law, where the spectral density varies with the temporal frequency as $P_X(f) \propto f^\alpha$. There exist a break in the PSD of AGN that could be described by a steep power law at high frequencies and a shallower one at low frequencies. The break relates to a characteristic variability timescale. The power spectra of AGN optical light curves often fit well with the slope $\alpha = -2$ such that $P_X(f) \propto 1/f^2$, which is used for the high frequency regime. This is consistent with a DRW model for the AGN variability, it is also consistent with X-ray literature where the PSD slopes are typically well modelled by a bending power law with $P(f) \propto 1/f^\alpha$ with $\alpha \sim 1$ for low frequencies down to a high frequency break where $\alpha \sim 2$ afterwards (Kelly *et al.*, 2013). This suggests that a CAR(1) process should provide a good description of the UV driving light curve, with τ being on the order of the length of the light curves or longer. There also seems to be evidence that DRW works for data on timescales from months to years, whereas for a month or less there seems to be some deviation from the DRW model. For the highest frequencies the PSD slope is also steeper than $1/f^2$ (Ivezi & MacLeod, 2013) and therefore deviates from the CAR(1) model. Since this project works with data on the timescale of years and both the optical and X-ray seem to be well defined by it, the DRW model should be a good fit for modelling the UV driving light curve.

Multiple facts indicate that a CAR(1) process describes the driving light curve well. The slope

of the power spectrum, also called the PSD slope being -2, the lack of any peaks in the power spectra and the aperiodic and noisy appearance of AGN light curves indicates that AGN light curves are stochastic or chaotic in nature. The physical process in the accretion disk is continuous and because it allows for a way to handle irregular light curve sampling, The stochastic model is continuous, which is what a CAR(1) process is. It should be noted that the stochasticity of the solution for the driving light curve underlies the fact that the physics are complex, not that the process is not deterministic. The CAR(1) model is used because of its simplicity meaning it is easy to calculate and it allows for fast generation of a large number of light curve realizations, required in Monte Carlo sampling procedures when fitting a light curve model. It also allows statistical inference without having the results biased by the irregular sampling, measurement errors, and finite span of the time series. The CAR(1) model is the simplest of stationary continuous autoregressive processes, though it is possible to obtain additional flexibility through the addition of higher order derivatives.

Typically the parameters for a CAR(1) process are estimated by maximum likelihood directly from the observed time series. This gives an advantage over non parametric approaches, like the discrete power spectrum or the structure function. The observed power spectrum and structure function can both suffer from windowing effects due to the finite duration and sampling of the light curve. It is possible for power from high frequencies to leak to low frequencies (aliasing) and power at low frequencies can leak to high frequencies (red noise leak). [Kelly et al. \(2009\)](#) uses a MCMC approach to obtain the posterior probability distributions for the model parameters. In the context of MCMC It is possible to define priors for the model parameters and calculate the likelihood. The priors on the parameters are made to be uniform for b and σ . For τ it is noted that when data is regularly sampled the CAR(1) process becomes an AR(1) process with $\alpha_{AR} = e^{-\Delta t/\tau}$ Where Δt is the time sampling interval. This value gives the correlation between neighboring data points. [Kelly et al. \(2009\)](#) states that it is reasonable to assume that any value of α_{AR} is a priori likely, which mean there is no a priori assumption about the correlations between subsequent data points. The result is a uniform prior on α_{AR} from 0 to 1. This prior is non informative because all of the information on α_{AR} comes from the data. The aim for this project is to apply the same MCMC approach for the model parameters.

[Kelly et al. \(2009\)](#) also advises to work with the log flux or apparent magnitude since the Gaussian white noise gives positive and negative values for the light curve while the flux is strictly positive. It is also important to remember the cosmological time dilation between rest and observed frame for the model parameters. assuming antisymmetric errors is found to not help results for the model.

2.1.1 Structure function

One other way to look at AGN variability is to determine the Structure Function (SF), which is the standard deviation of the distribution of the magnitude difference $m(t_2)-m(t_1)$ evaluated at many different times t_1 and t_2 , so that the time lag is $\Delta t = t_2 - t_1$ (and divided by $\sqrt{2}$ due to differencing). The time dependence of the structure function was found to be consistent with the prediction based on

a DRW model

$$SF(\Delta t) = SF_\infty [1 - \exp(-\Delta t/\tau)]^{1/2}$$

for small time lags this is $SF(\Delta t) \propto \Delta t^{1/2}$, which is the same as a random walk and becomes SF_∞ for large time lags. The structure function is related to the autocorrelation function, which makes a Fourier pair with the PSD function.

$$PSD(f) = \frac{\tau^2 SF_\infty^2}{1 + (2\pi f\tau)^2}$$

The asymptotic value of the structure function, SF_∞ , is equal to 2σ , relating the two definitions of the power spectrum or PSD function together. Light curve modelling can help us constrain τ and SF_∞ .

2.2 Gaussian Processes

¹⁰ Following the procedure of [Pancoast et al. \(2011\)](#) and [Skjelboe et al. \(2015\)](#) The CAR(1) model of [Kelly et al. \(2009\)](#) can be rewritten from the stochastic differential equation 38 to a Gaussian Process (GP). Originally this idea of GP's came from neural network research The properties of a neural network with one hidden layer converge to a Gaussian process as the number of neurons tend to infinity, but GP's can be have a wide number of different applications. To be able to apply the GP to the model the driving light curve it is important to understand what a GP is and how it is defined. Gaussian processes are like the CAR(1) processes stochastic in nature. Gaussian processes are used for two main reasons. The first reason is that a Gaussian process is completely determined by its mean and covariance functions. This property makes GP's suited for model fitting as only the first- and second-order moments of the process require specification. The second reason is solving the prediction problem (obtaining the value of the GP at a point where there is no data) is relatively straightforward. The best predictor of a Gaussian process at an unobserved location is a linear function of the observed values and for many problems these functions can be computed rather quickly using recursive formulas.

GP's are generally used for regression, classification and clustering problems. For this project the problem is one of regression and therefore that is where the focus will lie. The idea with GP's is to predict new data points not in what is called the training set, which in the case of AGN is the data points of the light curves. The idea is to get a function f to make predictions for all input values, to do this an assumption about the function like the form must be made or all functions consistent with the data would be equally valid. One approach is to choose the class of possible functions, like using linear functions. An other approach is to give priors to all the possible functions. There are problems with the first approach as it is possible to choose the wrong class of functions that may fit the training data well, but makes bad predictions. The second type of approach could have problems with computation time, since there are infinitely many functions to work with. The solution to these problems is to use GP's. A GP is a generalization of the Gaussian probability distribution. A probability distribution describes random variables which are either scalars or vectors (for multivariate distributions) while a stochastic process governs the properties of functions. It is possible to determine the properties of a function by only sampling it at finite points and it would be like we would have taken all the infinitely many points into account.

¹⁰This section relies on [MacKay \(1998\)](#), [Davis \(2006\)](#) and [Rasmussen & Williams \(2006\)](#). There is also a wealth of web articles with graphs and interactive elements that serve as great introductions to Gaussian processes. The ones that were drawn from in this work are [Fonnesbeck \(2017\)](#), [Parikh \(2019\)](#), [Shi \(2019\)](#) and [Görtler et al. \(2019\)](#)

The procedure here is to introduce some of the background concepts necessary to understand GP's and then see how they apply to cases like linear- and non linear regression, then introduce how the GP's relates to the CAR(1) model.

2.2.0.1 Background concepts

First concept is the **Multivariate Gaussian Distribution** also known as the joint normal distribution. A Gaussian distribution can be specified using a mean μ , variance σ^2 and probability distribution function (PDF)

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

It is important as it is the basic element from which a Gaussian process is constructed. In the presence of multiple independent Gaussian distribution it is possible to combine them. The combined PDF will also be a Gaussian i.e a multivariate Gaussian. In the case of two Gaussian combined it is a bivariate Gaussian. Generalizing the PDF of multivariate Gaussian can be specific by the means vector $\bar{\mu}$ and a covariance matrix Σ

$$f(x, \mu, \Sigma) = \frac{1}{\sqrt{|\Sigma|(2\pi)^d}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right)$$

The mean vector μ describes the expected value of the distribution. Each entry in the mean vector describes the mean of the corresponding dimension. Σ describes the variance along each dimension and the correlation between different random variables (in the bivariate example it would be the variance of x,y and the correlation between them). The covariance matrix is always symmetric and positive semi-definite. The diagonal of Σ consists of the variance σ_i^2 of the i-th random variable and the off-diagonal elements σ_{ij} describe the correlation between the i-th and j-th random variable. Σ describes the shape of the distribution. Some of the notable properties of Multivariate Gaussian is that the sum of Gaussians is a Gaussian, the marginal (procedure extracting partial information from multivariate probability distributions) of a Gaussian is a Gaussian and conditioning (the probability of one variable depending on another variable) a Gaussian also results in a Gaussian.

Secondly is the concept of **Bayes rules**, which states that posterior probability is given by the prior probability multiplied with the likelihood and divided by the evidence

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (42)$$

here A is the hypothesis (model) being tested and B is the evidence (data) used to test the hypothesis. $P(A|B)$ is the posterior which is the probability of the hypothesis given the evidence. $P(B|A)$ is the likelihood which is the probability of observing the new evidence, given the initial hypothesis. $P(A)$ is the prior, which is the probability of the hypothesis without any additional prior information. Lastly is $P(B)$ referred to as the marginal likelihood¹¹ this is the total probability of observing the evidence and is the integral over all priors and all likelihoods. It functions mainly as a normalization constant, but is also used when comparing different types of models.

¹¹Sometimes $P(B)$ is also referred to as the evidence

2.2.1 Regression and Gaussian processes

There are several ways to interpret GP regression models. One can think of a Gaussian process as defining a distribution over functions, and inference taking place directly in the space of functions, the function-space view. It can be easier to come to grasp with the weight space view so that will be detailed first.

2.2.1.1 Linear regressions

to understand the formulation of GP's and how they relate to regression it a good idea to consider the what is referred as the "weight space view" with an example of linear regression

$$y = f(x) + \epsilon$$

$$y = xw_1 + w_2$$

where ϵ is a term for the noise, which is often incorporated into the weights. In this case the weights are the slope w_1 and the intercept w_2 . Changing the values of the weights creates a number of linear models. The point of a parametric model is to find the weight values that makes it so the model best fits the data. for a Bayesian linear regression priors are put on the weights and from observing the data (x,y) i.e. the training set the posterior probability which describes the eights that best fit the data, is determined using Bayes rule. The likelihood in this case is the function which is mostly likely for a given weight and training input x so $p(y|x, w)$. It is then possible to use the model on new data when the posterior for the weights are determined by using Gaussian marginalisation.

2.2.1.2 Non linear regression and feature space

The General idea for non linear parametric models is to empirically model relationships in high dimensions. For Bayesian analysis using a model for a set of data is the same as inference of a function given the data. The inference of the function $y(x)$ is given by the posterior probability distribution following Bayes rule.

For non-linear regression there is a function $y(x)$ that is assumed to describe the data x and we want to infer the function from the data and predict its value at new points. One approach to this problem is the parametric one, where the unknown function is expressed in terms of a non-linear function $y(x; w)$ with the parameters w . The function $y(x; w)$ is inferred by inferring the parameters w just like the linear case. The posterior probability is again given by Bayes rule. Both the likelihood and the prior are taken to be separable Gaussians. For the likelihood it is for each data point and for the prior it is for each parameter. Its should be noted that if y does not depend linearly on w then the posterior distribution will not generally be a Gaussian. The parameters can then be inferred using maximum likelihood or the posterior distribution can be sampled using MCMC.

There is also another to understand non linear regression, which will introduce the concept of the GP. For the case of a non linear and general function it is possible first project the inputs x into some high dimensional space using a set of basis feature space functions and then use the linear model there

instead of directly on x . An example of this approach is if a scalar input x could be projected into the space of powers of x using $\phi(x) = (1, x, x^2, x^3, \dots)^\top$ to implement polynomial regression. If the projections are fixed functions so that they are independent of the parameters w the model will still be linear in the parameters and this means it will have an analytic form. It is then possible to use Bayesian inferences to obtain the value function value (f^* or y^*) using $\phi(x^*)$. It is this transformation using basis functions where in general terms the move from a linear model to a GP happens.

A full derivation is given in [Rasmussen & Williams \(2006\)](#) but the idea is to use what is called the "kernel trick" which makes it possible to apply the Bayesian framework for linear models to any type of function. The kernel trick is defined as "for an algorithm defined solely in terms of inner products in input space then it can be lifted into feature space by replacing occurrences of those inner products by the covariance function or kernel $k(x, x')$ " ([Rasmussen & Williams, 2006](#)). What this means is that the trick makes it possible to directly use inputs in the function space by specifying the kernel and it therefore circumvents the need to specify weights w or basis functions $\phi(x)$.

2.2.2 Function space view

A Gaussian process is a generalization of the the multivariate normal distribution to infinite dimensions. It is defined as an infinite collection of random variables, with any marginal subset having a Gaussian distribution¹² ([Fonnesbeck, 2017](#)). The random variables represent the value of the function $f(x)$ at location x . Another way of thinking about an infinite vector is as a function. When writing a function that takes continuous values as inputs, it is implied that an infinite vector that only returns values (indexed by the inputs) when the function is called upon to do so. The idea of an infinite dimensional Gaussian represented as a function makes it possible to work with them computationally, as it is never required to store all the elements of the GP, it is only necessary to calculate them as they are needed. Therefore it is also possible to describe a GP as a distribution over functions. The simplest type of prior over functions one has is the GP.

2.2.3 The mean- and covariance function

In comparison a multivariate normal distribution is completely specified by a mean vector and covariance matrix, whereas a GP is fully specified by a mean function and a covariance function.

$$\begin{aligned}
 m(x) &= E[f(x)] \\
 K(x, x') &= E[(f(x) - m(x))(f(x') - m(x')))] \\
 p(x) &\sim GP(m(x), K(x, x'))
 \end{aligned}$$

Where $m(x)$ is the mean function, which can be thought of as the mean value of the function in the absence of any data. Usually the mean function is set to zero as it turns out the kernel mostly governs how the GP is updated. The kernel or covariance function $K(x, x')$ describes the shape of the distribution and therefore determines the characteristics of the function. In fact for a finite number of points, the GP becomes a multivariate normal, with the mean and covariance as the mean function and covariance

¹²A similar definition is given by [Rasmussen & Williams \(2006\)](#): A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

function, respectively, evaluated at those points. An example of a kernel is the Radial Basis Function (RBF), exponentiated quadratic or squared exponential kernel

$$K(x, x') = \sigma^2 \exp\left(-\frac{\|x - x'\|^2}{2l^2}\right) \quad (43)$$

where σ informs the average distance away from the functions mean (σ^2 is the variance) and l is the length scale. The length scale in the covariance function can be thought of as roughly the distance needed to move in input space before the function value can change significantly or in short the reach of influence on neighboring points. These parameters are referred to as hyperparameters.

The CAR(1) process for AGN light curves is described by a mean function detailing the long-term mean and the exponential covariance function (Pancoast *et al.*, 2011)

$$K_{AGN}(t, t') = \sigma^2 \exp\left(-\frac{\|t - t'\|}{\tau_{var}}\right) \quad (44)$$

where σ is the long-term standard deviation of the light curve and τ_{var} is the typical timescale of variations.

2.2.3.1 GP priors

When a GP prior is defined samples can be drawn from it and the result will be arbitrary function samples. Then given the training dataset it is possible to determine the posterior (y or $f(x)$, since $y=f(x)+\text{noise}$). A demonstration of how a realisation of a GP prior is obtained by evaluating a function over a number of points is a good way to understand the idea of how a GP is a distribution over functions. In this case the GP prior is sampled without any data being introduced. It is possible from this to create a functions that evaluates the covariance at specified points, giving the covariance matrix. These GP realizations can be generated sequentially one point at a time due to the aforementioned conditioning property of multivariate Gaussian distributions. The mean function will be zero and the covariance function will be the squared exponential with the hyperparameters $\sigma = 1$ and $l = \frac{1}{\sqrt{10}}$. Figure 5a shows the first realization without any points sampled. with the zero mean function the confidence band plotted represents one standard deviation from the mean. Figure 5b shows that given a starting point to sample the confidence band can be updated using the covariance function, conditional on the point being sampled. By sampling this point the probable location of further points have been constrained. The confidence can be continually updated one by one as seen in figure 5c or multiple points can be sampled at once like what is seen in figure 5d. Notice that as more and more point are sampled for the prior the confidence bounds shrink near where points are sampled and expands moving away from the points. The posterior can be updated in much the same way as the prior as more data is added, though real data (especially in astronomy) is corrupted by statistical noise and so the function $f(x)$ will not go directly through the data like what is seen with figure 5

2.2.3.2 Different GP approaches

There are two main approaches to implementing a GP into a model. The first is the marginal likelihood method and the second is the latent variable method. The more common type of GP is the marginal

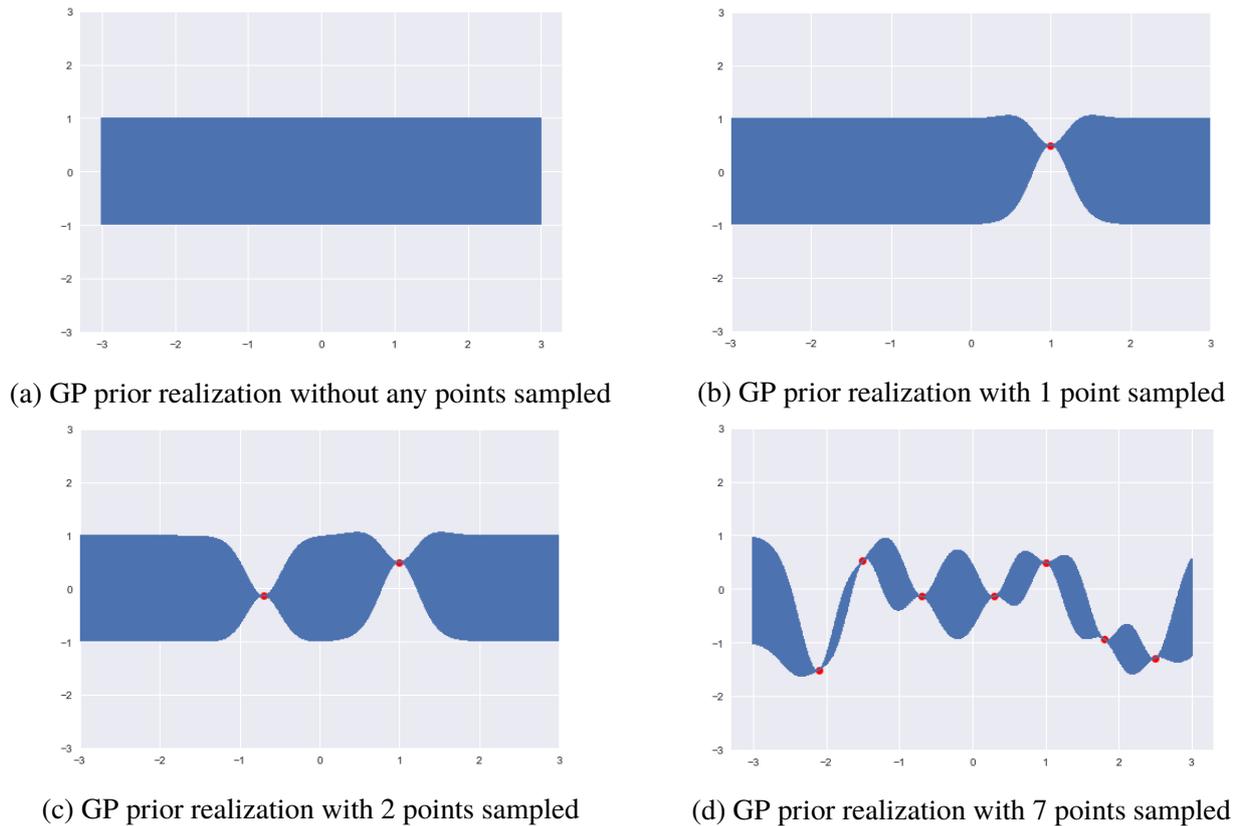


Figure 5: Code to generate these plots is originally from [Fonnesbeck \(2017\)](#)

likelihood one, where the observed data are the sum of a GP and Gaussian noise ($\epsilon \sim N(0, \Sigma)$). The unknown latent function $f(x)$ can be analytically integrated out of the product of the GP prior probability with a normal likelihood. This quantity is called the marginal likelihood. While the marginal likelihood approach is more common, is it not possible to do any type of mathematical operation on the GP since the latent function is integrated out. Instead the latent variable approach should be used. The latent variable approach is a direct implementation of a GP. The GP is latent because the underlying function values are treated as latent variables. These variables can be defined over a range of values making the approach ideal for working with time series data.

2.2.3.3 Limitations of GP's and the transfer function

[Duvenaud \(2014\)](#) The most standard kernels for GP regression or classification model are the Squared-Exponential or Rational Quadratic kernels. They generally work as a solution for interpolating smooth functions with no "kinks". In the case where the function happens to have a discontinuity or is discontinuous in its first few derivatives (an example of this would be the `abs()` function), then either the resulting length scale will be extremely short and the posterior mean will become zero almost everywhere or the posterior mean will have "ringing" effects. Even in the absence of hard discontinuities, the length scale will usually end up being determined by the smallest 'wiggle' in the function. It is therefore the case that the model can fail to extrapolate in smooth regions if there is even a small non smooth region in the data ([Duvenaud, 2014](#)).

This can have an effect on the combined model of the driving light curve and the transfer func-

tion, since the transfer function is specified as two lognormals with a location parameter θ . The problem with having a location parameter is it moves the transfer function and can create a discontinuity as the gradient changes from 0. To avoid this all the models will use $\theta = 0$ for both the torus and the accretion disk. A consequence of this is also a performance boost, since without the locations parameters the model will not have to check for cases where $\tau_i - \theta$ is less than 0.0 so there would be a logarithm of a negative value¹³.

2.3 A degeneracy in DRW modelling of AGN light curves

Kozłowski (2016) had the idea to test whether a non-DRW stochastic process would be successfully and well modelled as DRW, and return correct variability parameters Or rather return biased parameters, for example, longer time-scales for steeper SFs as in the SF analysis. This was done by testing the following covariance function

$$K_{AGN}(t, t') = \sigma^2 \exp\left(\left(-\frac{\|t - t'\|}{\tau}\right)^\beta\right) \quad (45)$$

a 1000 light curves were simulated with values $0.5 \leq \beta \leq 1.5$. The result was that β seems to weakly change from 1 to 1.2 for the brightest AGN leading to a bias for the parameters as β changes. Kozłowski (2016) concludes that as an upside the modest deviations from the DRW model seem to be nearly unimportant for the estimated variability parameters and they weakly affect the correlations with the physical AGN parameters. The downside is that typical AGN light curves are not good enough to notice the deviations from DRW and so it is possible to misinterpret parameters. Because SFs or PSDs are a model-independent means of estimating the shape of the covariance function of the signal, it is recommended to estimate β this way first and then use it as input parameter in direct light-curve modelling to obtain correct model parameters. The CAR(1) process is generally good as structure functions for 9000 AGN from the Sloan Digital Sky Survey have β near 1, but it should be kept in mind that some processes might not be DRW and give rise to other beta values.

Skjelboe *et al.* (2015) similarly tried to use different values of β between $1 < \beta < 2$ to test different amount of small scale structure on Arp 151 and found no significant difference between the different models when determining the response function using RLI.

Smith *et al.* (2018) investigated 21 Kepler AGN light curves and found their slopes to be steeper than this $\alpha = -2$ (see figure 6 for slope distribution) and they therefore disagree with the optical results. early studies using high signal-to-noise and high-cadence Kepler data also suggest the presence of steeper PSD slopes, up to -4.51 with timescales shorter than a few days. To obtain the PSD a discrete Fourier transform was performed on a interpolated light curve. The data was fitted with a line to remove a linear trend, which will remove the lowest-frequency component of the power spectrum. The mean was also subtracted before the Fourier transform. Monte Carlo methods is then used to obtain the best fit slope and its uncertainty. The slope that can be accepted with the highest confidence is the slope that

¹³Also this check can break the gradient, which is a problem for Hamiltonian MCMC smapler like NUTS. One way to get around this would be to approximate this "switch" with a sigmoid function.

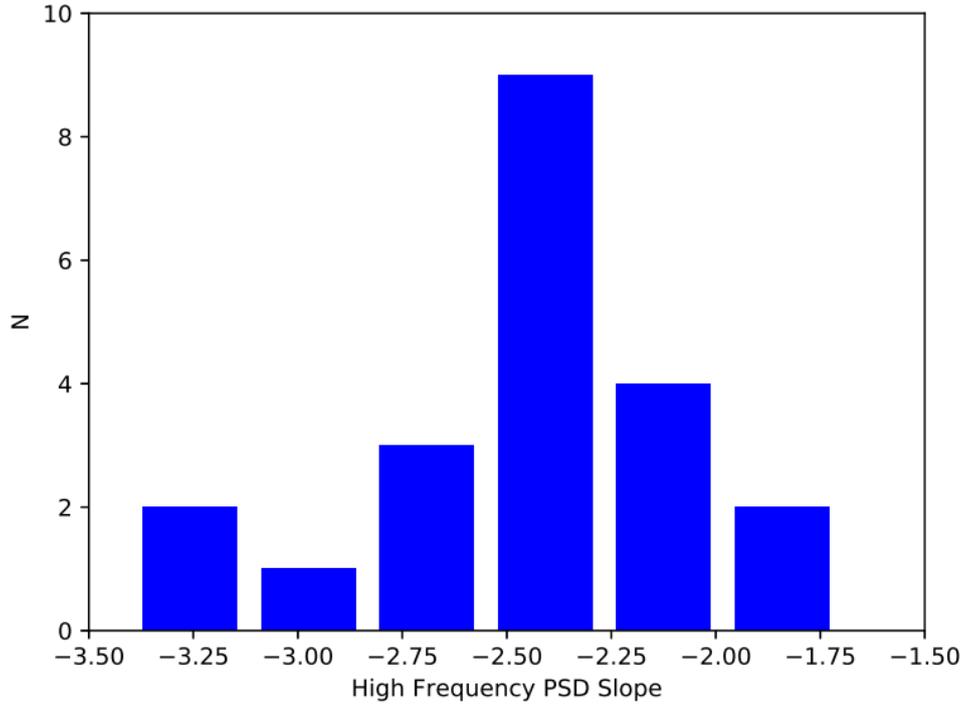


Figure 6: Histogram of the best-fitting high-frequency PSD slopes as measured by simulating light curves. Figure is originally from [Smith *et al.* \(2018\)](#)

is then adopted as the best fit for each object. While the PSD slopes tend to be steeper than $\alpha = -2$ for the space based Kepler data, they might still fit for ground based data.

[Smith *et al.* \(2018\)](#) also suggests it is possible that the break timescale τ could be the point at which the behavior switches from a steep, red variability to a damped random walk or similar. One possibility is that the PSD of an AGN becomes redder and redder due to a cyclic accretion-disk duty cycle similarly to what occurs in X-ray binaries. The AGN moves closer and closer to a critical moment when its behavior switches to a damped random walk. The time it takes for that switch to occur will depend on the black hole mass, with longer timescales for larger masses and larger accretion disks in general. If this is the case then objects with observed PSD breaks would have the reddest high-frequency slopes compared to the rest of the sample, which has been caught somewhere in the middle of its reddening phase. For the broken-PSD objects in the Kepler data of [Smith *et al.* \(2018\)](#) this is the case. The mean value for the broken-PSD objects in the Kepler data is $\langle\alpha\rangle = -3.0$, compared to $\langle\alpha\rangle = -2.3$ for the unbroken sample. $\alpha=2.0$ seems to be true for luminous objects like quasars.

The consequence of a steeper PSD slope is that the light curve has less structure on small scales, compared to what is predicted by the DRW model. While the Kepler data does not represent all AGN light curves (the SDSS light curves are indicative of that), the resulting PSD slopes do indicate that care should be taken then when interpreting small time scale AGN variability using DRW models. It is possible that any small scale structure fitted by a DRW is simply noise in the light curves, which is being unaccounted for by the estimated measurement errors. modeling noisy light curves

using a DRW has the possibility of overfitting the light curves resulting in the following (Skjelboe, 2016):

- If the additional noise is due to flux calibration issues affecting the continuum light curves in a single epoch, overfitting the light curves can introduce excess power in the time delay distribution at $\tau = 0$.
- For independent light curve noise, overfitting the continuum may still produce accurate time delay measurements, but the error on the time delay will be underestimated. This is due high frequency noise being propagated to the likelihood distribution, resulting in error estimation being done on small scale noise, rather than on the underlying likelihood signal.

The error bars can be multiplied by a factor to make them dominate over the variability's on the smallest scales for the light curves. This effectively removes the effect of high frequency noise and makes the likelihood distribution smoother. This preserves large scale behavior, which is what is needed to determine the time delay.

3 MCMC

¹⁴ Markov Chain Monte Carlo (MCMC) is important tool for integration and optimisation problems in large dimensional spaces. some examples are:

- Bayesian inference and learning. For some given unknown parameters and data the normalization, marginalisation and expectation (value) can be found.
- Statistical mechanics. compute the partition function for a system with a given number of states and a Hamiltonian
- Optimization. Chi squared minimization and the like.
- Penalised likelihood model selection. This involves two steps. First, one finds the maximum likelihood (ML) estimates for each model separately. Then one uses a penalisation term (for example MDL, BIC or AIC) to select one of the models. This is computationally heavy.

The first item is relevant for the this project, as the idea is to obtain the posterior probability distribution for the model parameters, especially the median lag of the dusty torus.

MCMC makes it possible to characterize a distribution without knowing all of the distribution's mathematical properties by randomly sampling values out of the distribution. A particular strength of MCMC is that it can be used to draw samples from distributions even when all that is known about the distribution is how to calculate the density for different samples. The 2 properties giving Markov Chain Monte Carlo its namesake is:

¹⁴This section mainly relies on Andrieu *et al.* (2003), D. van Ravenzwaaij (2018) and Hogg & Mackey (2018)

- The Markov Chain property. This is the idea that the random samples are generated by a special sequential process. Each random sample is used as a stepping stone to generate the next random sample, which is why it is called a chain. The current step is only dependant directly on the one before in the chain. In that sense the chain is "memory less".
- The Monte Carlo method. This is the practice of estimating the properties of a distribution by examining random samples from the distribution, instead of doing it directly from equations.

MCMC methods are methods for sampling probability distribution functions or probability density functions (pdfs). They only require the ability to be able to compute ratios of the pdf at pairs of locations. While MCMC methods can be less efficient than classic methods, but are more general and unbiased. MCMC works well with Bayesian inference due to its focus on posterior distributions (posterior pdf sampling), which are hard to derive analytically. A useful version of Bayes rule in this case is:

$$p(\theta|D) = \frac{p(D|\theta) \cdot p(\theta)}{p(D)}$$

Compare this form of Bayes rule to the general one from equation 42. Here θ indicates a (set of) parameter(s) of interest and D indicates the data, $p(\theta|D)$ indicates the posterior or the probability of θ given the data, $p(D|\theta)$ indicates the likelihood or the probability of the data given θ , and $p(\theta)$ indicates the prior or the apriori probability of θ . $p(D)$ is the evidence and is generally hard to calculate, but this is not needed if the pdf ratios for two points in parameter space are computed. $p(D|\theta) \cdot p(\theta)$ is what is referred to as the density. The way the data is used to update the prior belief is by examining the likelihood of the data given a certain (set of) value(s) of the parameter(s) of interest.

One thing to keep in mind when using MCMC is that it should not be used to search the parameter space for good models. MCMC should also not be used to optimize the posterior pdf or for certain choices of prior pdf, optimize the likelihood. MCMC is good at one thing, and one thing only, which is sampling ill normalized (or otherwise hard to sample) pdfs.

3.1 The Monte Carlo method

¹⁵ It is possible to express the solution to many problems in math as the integration of a function.

$$I = \int_{\Omega} f(x)dx$$

in which Ω is the domain of integration, the integral I can be related to an expectation of a random variable with respect to some probability measure. it is possible to rewrite the integral in terms of the expectation value for probability measures of a random variable X that have a density $\rho(x)$.

$$I = E(g(X))$$

Where $g(x) = \frac{f(x)}{\rho(x)}$. It is then possible to use the law of large number, which states that for a collection of independent identically distributed random variables $X_{i=1}^{\infty}$, to write

$$E(g(X)) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N g(X_i)$$

¹⁵This section relies heavily on [Atzberger \(2020\)](#) and is rather technical in nature. It can be skipped without a great loss to the overall context.

Which means I can be approximated with $I \approx \frac{1}{N} \sum_{i=1}^N g(X_i)$. This is the Monte Carlo method. The accuracy of the method is given by

$$error = \left| \frac{\sigma_g}{\sqrt{N}} \eta(0, 1) \right|$$

Where

$$\sigma_g^2 = \int_{\Omega} (g(x) - I)^2 \rho(x) dx$$

and $\eta(0,1)$ denotes a standard normal random variable (Gaussian random variable) with mean zero and variance 1. The convergence rate in The Monte Carlo method is strongly influenced by the prefactor σ_g which depends on the function $f(x)$ and the sampling distribution with density $\rho(x)$ that is used. The prefactor is the main why the accuracy can be improved.

For the Monte Carlo method to work there is a need for a way to generate random number or at least pseudo-random numbers there are a couple of ways to generate this:

- Linear Congruential Generator, which attempts to create a sequence of numbers in a range by recurrence.
- Lagged Fibonacci Generators, which also uses recurrence but a different form
- Transformation method, where samples for non-uniform random variables can be obtained from the samples generated for the uniform random variable, since transforming a random variables changes the underlying probability distribution.
- Exponentially Distributed Random Variables, with a probability density $\rho(x) = \lambda e^{-\lambda x}$
- Normally Distributed Random Variables: Box-Muller, with a probability density $\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- Sampling by Rejection, where the desired random variates (particular outcomes of random variables) can be obtained by generating candidate samples which are either accepted or rejected to obtain the desired distribution.

Variance reduction is an important parts of the Monte Carlo method and a way to reduce it is to use a probability density $\rho(x)$ for which generation of variates X_i is not too difficult while making σ_g^2 small. One method is Importance sampling, which is concerned with the choosing $\rho(x)$ for the random variates X_i so that regions which contribute significantly to the expectation of $g(X)$ are sampled with greater frequency. For a given value $f(x) > 0$ it is possible to find a $\rho(x)$ which makes $\sigma_g = 0$. This is possible if $\rho(x) = \frac{f(x)}{I}$ is chosen since then $g(x) = I$, but that would presuppose already knowing what I is. It is instead possible to choose a $\rho(x)$ which comes as close to this as possible. A way to do this is to use sampling with a Gaussian mixture where $\rho(x)$ is a multi-modal probability distributions. The method require generating the extra uniform random variate, but it also decreases σ_g more than a standard Gaussian distribution.

3.2 What is sampling?

When sampling a pdf $p(\theta)$ a number of K values is drawn and if K is large enough making a histogram would up to a normalization constant be the same as the original pdf. A pdf is always positive and its integral over the total parameter space is equal to 1, that is there is a 100% chance the parameters have some value. There can be many parameters and so the integral can be multidimensional. For a pdf $p(\theta)$ the expectation values for θ or for any value that can be expressed as a function of θ , $g(\theta)$ can be written as

$$E_{p(\theta)}[\theta] = \int \theta p(\theta) d\theta$$
$$E_{p(\theta)}[g(\theta)] = \int g(\theta) p(\theta) d\theta$$

These expectation values are the mean values of θ and $g(\theta)$ under the pdf. A good sampling makes the sampling approximation to these integrals accurate. The integrals can be replaced with sums over samples θ_k for a good sampling

$$E_{p(\theta)}[\theta] \approx \frac{1}{K} \sum_{k=1}^K \theta_k$$
$$E_{p(\theta)}[g(\theta)] \approx \frac{1}{K} \sum_{k=1}^K g(\theta_k)$$

Often when working with MCMC there is a badly normalized function $f(\theta)$ which is the pdf $p(\theta)$ multiplied by some unknown scalar that is hard to compute. Assuming that $f(\theta)$ is positive and that its integral is finite. When Sampling $f(\theta)$ the sampling-approximation expectation values are the expectation values under the properly normalized corresponding pdf, even in the case where the normalization is not determined.

$$E_{p(\theta)}[g(\theta)] = \frac{\int g(\theta) f(\theta) d\theta}{\int f(\theta) d\theta} \approx \frac{1}{K} \sum_{k=1}^K g(\theta_k)$$

3.2.0.1 MCMC chains and convergence

An MCMC chain converges to an invariant distribution (the posterior pdf) as long as the transition probability matrix (the probability of transitioning to other states) has two properties:

- Irreducibility. For any state of the Markov chain, there is a positive probability of visiting all other states. That is, the matrix T cannot be reduced to separate smaller matrices, which is also the same as stating that the transition graph is connected.
- Aperiodicity. The chain should not get trapped in cycles

MCMC samplers are irreducible and aperiodic Markov chains that have the target distribution as the invariant distribution.

3.3 MCMC samplers

¹⁶ The simplest algorithm for MCMC sampling is the Metropolis Hastings algorithm (MH MCMC). The MH MCMC algorithm requires two inputs. The first is the function $f(\theta)$, that is the function to be sampled, such that the algorithm can evaluate $f(\theta)$ for any value of the parameters θ . This function is given by the density $(p(D|\theta) \cdot p(\theta))$ evaluated at the observed data D . The second input is the proposal pdf function $q(\theta'|\theta)$ which can construct samples, so that the algorithm can draw a new position θ' in the parameter space given an "old" position θ . This is typical choice is a multi-variate Gaussian distribution for θ' centered on θ with some simple variance tensor. This second function must meet a symmetry requirement (detailed balance)¹⁷. This is also referred to as a reversibility condition, since $q(\theta'|\theta) = q(\theta|\theta')$ must be satisfied. It permits us to random-walk around the parameter space in a fair way. The easiest way to meet this requirement is to use pdfs that are symmetric in θ' and θ , like the Gaussian or centered uniform distribution.

The procedure to update a parameter with the function $f(\theta)$ in MH MCMC¹⁸ is as follows:

- Draw a proposal θ' from the proposal pdf $q(\theta|\theta')$.
- Draw a random number from $0 < r < 1$ from a uniform distribution.
- If $\frac{f(\theta')}{f(\theta)} > r$ then update the parameter to the proposal, otherwise keep it the same.

This algorithm (and indeed any MCMC algorithm) produces a biased random walk through parameter space. It is a random walk because each part in the chain only depends on the last, this is also why it is a Markov chain. It is biased because of the acceptance algorithm involving the ratios of function values; this acceptance rule biases the random walk such that the amount of time spent in the neighborhood of location θ is proportional to $f(\theta)$.

A thing to keep in mind is that the first steps in the Markov Chain should be ignored since the initial guess might be wrong and the answer needs to converge. These first steps are what is called the "burn in phase" The width of the proposal distribution is sometimes called a tuning parameter of this MCMC algorithm and can affect whether there is a lot of rejections or maybe the algorithm is stuck in a local maxima (similar problems occur with χ^2 optimization). For a too narrow proposal distribution almost all steps are accepted but it will take a long time to move anywhere. If the proposal distribution is too wide (it proposes steps too large) the moves will cover parameter space easily, but almost no steps will be accepted. The acceptance fraction should be 0.2-0.5 with a value around 0.234 for high dimensional problems (Hogg & Mackey, 2018). It is only possible to tune during burn in phase and not the final MCMC run since that would break the Markov property.

¹⁶For a straightforward and slightly humorous introduction to the MCMC samplers with animated examples, see [McElreath \(2017\)](#)

¹⁷It is possible to break detailed balance if the acceptance condition changed to take this into account, which is what is done in the MH algorithm

¹⁸Technically this is the Metropolis algorithm rather than the Metropolis Hastings algorithm, but the difference in this case is minute.

3.3.1 Advances samplers

For strongly correlated parameters a more advanced method than Metropolis Hastings is needed since correlations between parameters can lead to extremely slow convergence of sampling chains, and sometimes to nonconvergence (within any reasonable time). Gibbs sampling breaks down the problem by drawing samples for each parameter directly from that parameter's conditional distribution, that is the probability distribution of a parameter given a specific value of another parameter. For a combined Metropolis within Gibbs approach The key is that for a multivariate density, each parameter is treated separately: the propose/accept/reject steps are taken parameter by parameter.

Differential Evolution MCMC (DE MCMC) ([Braak, 2016](#)) can help improve the speed of convergence for correlated parameters relative to the Gibbs approach. The idea is that the chains are not independent, as they interact with each other during sampling, and this helps address the problems caused by parameter correlations. It does this by using the difference between other chains to generate new proposal values.

A similar type of sampler is an ensemble sampler. An example of such an algorithm is emcee from [Mackey *et al.* \(2013\)](#) where an ensemble of walkers (100 or more) are used. A single walker is updated using the current position of all the other walkers. It has the advantage that while anisotropic pdf are difficult for traditional MCMC, emcee does what is referred to as an affine transformation, so that the problem can be transformed to sampling from an isotropic pdf, this is because an algorithm that is affine invariant performs equally well under all linear transformations. It will therefore be insensitive to covariances among parameters. It should be noted though that ensemble samplers have problems sampling target distributions in moderate to high dimensions ([Huijser *et al.*, 2015](#)). Also for multi-modal target distributions walkers may become stuck and other samplers like Diffusive Nested Sampling (DNS) is needed. DNS [Brewer *et al.* \(2011\)](#) also allows for the computation of the evidence, but is generally more computationally expensive as a result.

3.3.2 Hamiltonian Monte Carlo and the No U-Turn Sampler

Hamiltonian MCMC (HMC) transforms the problem of sampling from a target distribution to simulating Hamiltonian dynamics. This is more efficient than a random-walk Metropolis algorithm, but it requires the gradient of the log-posterior and it needs a step size and number of leapfrog steps. Setting these two parameters can be difficult and expensive, requiring tuning runs first to determine the number of steps. The No-U-Turn Sampler (NUTS) proposed by [Hoffman & Gelman \(2011\)](#) can determine these two parameters, eliminating the need for tuning runs and hand tuning.

HMC works by also having momentum variables for every model parameter, which are independently drawn from a normal distribution. It works like a Hamiltonian physical system with the position, momentum, potential energy, kinetic energy and negative energy for a particle. The leapfrog integrator is used to evolve the system over time and update parameter proposals. This method is essentially setting a "particle" off in a random direction in parameter space, which acts as a frictionless surface. Samples are taken of the particles position along its path and used to update the particles position, before sending it off in a different random direction. Doing this repeatedly makes it possible to learn

about the shape of the surface in question.

The performance for a HMC is dependant upon the choice of step size and step number. For too large step sizes the the simulation will be inaccurate and the result will be a low acceptance probability and if it is too small computations are wasted by taking a lot of small steps. For too small step number the samples will be close to each other and the consequence will be a random walk behavior and if it is too large then the HMC will loop back on itself and either explore parameter space very slowly or get stuck.

NUTS can calculate when the particle simulation has gone for long enough so these problems do not occur. A binary tree of points is made forwards and backwards in fictions time until the leftmost and rightmost nodes start to turn back on themselves (make a u-turn), That is the distance between them start to decrease. Compared to the other samplers mentioned NUTS is able to be used on complicated models and efficiently sample from high dimensional target distributions without needing to be tuned to the shape of those distributions. The number of effective samples is also expected to be higher from NUTS than a standard random walk MH sampler, as the expected acceptance probability is around 0.8. Another advantage of using HMC is that it provides the user the ability to diagnose biased inferences by considering divergences. Divergences occur when the sampler encounters regions of high curvature in the target distribution which it cannot adequately explore. In those cases it is possible to change the acceptance probability, thereby forcing the sampler to have smaller step sizes so it can adequately explore parameter space or it may be necessary to reparameterize the model in question.

The two main MCMC samplers that have a NUTS algorithm at its cores is STAN [Stan \(2017\)](#) and PyMC3 ([Salvatier et al. , 2016](#)). STAN is implemented in multiple programming languages, while PyMC3 is programmed in Python. They have different ways of defining models, but both allows for the user to define parameter priors and likelihoods for models to perform Bayesian posterior inference. This projects uses PyMC3¹⁹, built on top of the Python library Theano, which allows for the defining, optimization, and evaluation of mathematical expressions involving multi-dimensional arrays efficiently.

4 Data

The data used in this project consist of NIR and optical data from the REMIR and ROS2 cameras respectively. The cameras are part of the Rapid Eye Mount (REM) telescope situated at the La Silla Observatory in Chile. The consist of 7 observational bands. The J,H,K bands from REMIR and the g,r,i,z bands from ROS2. Data reduction, photometry and identification of the best reference stars was performed by Bo Milvang-Jensen. I hereby provide a summary of this procedure.

The images used for the analysis are firstly the NIR images, which are the co-added images from the REM archive, which means that the 5 raw images were sky-subtracted and co-added, so for J,H,K there is only one magnitude per epoch. The Optical images are reduced images done by Bo, made by

¹⁹Great sources for an introduction to using PyMC3 are the official website <https://docs.pymc.io/> and the PyMC3 discourse forum <https://discourse.pymc.io/>.

applying the relevant monthly master bias and flat from the REM archive; the 3 individual images have not been co-added since it was unclear if the astrometry was good enough, so 3 magnitudes at each epoch are present in the photometry tables.

Firstly, photometry was done using the SExtractor software on each image in single-image mode. This produces a photometry catalogue for each image, with photometry for the objects that were detected in that image (varies from image to image), and with Right Ascension and Declination (RA,Dec) as given by the World Coordinate System (WCS) of that image. These coordinates could potentially be off by several arcsec or in some cases completely off. Then SExtractor measures the aperture magnitudes in apertures from 1 to 25 arcsec, in steps of 1 arcsec. All images were inspected and the AGN were manually identified as well as the sources in the image that seemed like potential photometric reference stars. usually the number of potential stars were in the order of 5. Next the RA,Dec coordinates were obtain from the Simbad Astronomical Database or in the cases where coordinates were not present in the database, the Vizier version of the Gaia catalogue was used as a substitute. The resulting catalogue from these two sources was cross correlated with all the individual SExtractor catalogues with a maximum matching distance of 10 arcsec. For this cross correlation to work a large value is needed as the WCS of the individual images are inconsistent at the several arcsec level, also since the objects in question are relatively bright, mismatching objects should be kept at a minimum. The SExtractor data was then reformatted so that N objects in the literature catalogue in each SExtractor catalogue becomes a single row in the output catalogue, which is done by giving all the columns for each object a suffix $_N$ (where $_1$ is the AGN, $_2$ the first potential reference star, etc.).

This procedure has been done for 6 fields analysing 6 different objects (AKN120, ESO323, F51, F9, NGC3783, NGC7213). For each field the two best best reference stars are determined. The the median subtracted differential magnitude is plotted for each AGN given the two reference stars (that is $m_{AGN} - m_{ref,1}$ and $m_{AGN} - m_{ref,2}$) as well as the difference in magnitude between the two reference stars ($m_{ref,1} - m_{ref,2}$) to compare the two. This is done for different aperture diameters raging from 6 to 25 arcsec, to find the one with the least scatter that incorporates the entire AGN into the image, without including too much of the background. The median is subtracted since this project is concerned with the variable part of the continuum. The error bars are computed based on the error bars from SExtractor. An upper limit of 0.1 mag is put on the acceptable error for the points of the light curve. Also to mitigate the effect of outliers in the light curves, every band is inspected individually and an upper and lower bound for the possible values of the light curve is chosen. It should be noted that it is possible that some stars either are too faint (resulting in high random noise) or too bright (resulting in systematic errors due to non linearity or even saturation), meaning that the same reference star cannot be used for all bands, though this does not seem to be the case with the objects investigated in this projects.

To work with the flux ratio between the AGN and the reference star $\left(\frac{F_{AGN}}{F_{ref}}\right)$ instead the (median subtracted) differential magnitude (Δm) can be converted using the following equation

$$\frac{F_{AGN}}{F_{ref}} = 10^{-\frac{\Delta m}{2.5}} \quad (46)$$

The accompanying uncertainty for the flux ratio is calculated using error propagation

$$\sigma_{\frac{F_{AGN}}{F_{ref}}} = \sigma_{\Delta m} \left(\frac{F_{AGN}}{F_{ref}} \right) \left(\frac{2.5}{\ln(10)} \right)^{-1} \quad (47)$$

where $\sigma_{\Delta m} = \sqrt{\sigma_{AGN}^2 + \sigma_{ref}^2}$ is the combined error of the AGN and the reference star. While the light curve of the median subtracted differential magnitude will vary around zero, when converted to the flux ratio the light curve will instead vary around one. This projects will use the flux ratio data as the model input. This is done because the interpretation of the driving light curve and transfer function is more straightforward and in line with the similar type of the work done in the field of reverberation mapping. As for the use of the flux ratio instead of the absolute flux (or magnitude), the differential magnitudes were obtained using a arbitrary zero points of 30 and as a result this does not ensure that the magnitude differences between the bands are calibrated. that would only be the case if the instruments were equally sensitive in the different bands and had identical detector gains. Though in reality this is not a problem for the determining the lag, since in that case the only concern are the differences in time between the peaks and troughs of the time variable component of the light curve, so the absolute flux (or magnitude) is irrelevant. Since only the variable part of the light curve is relevant for the analysis, a value of one can be subtracted from all the flux ratio light curves. whether to use data that is varying around one or zero is mostly relevant for the GP mean function, when defining the driving light curve, since it should be varying around the same value as the data. A constant mean function can be used to model the variation around one for the flux ratio. Both approaches were tried to test if it made a difference on the final result and that was not the case. This project will subtract one from the flux ratio light curve and use a zero mean function, though both approaches are equally valid.

4.1 Instruments: REMIR/ROS2

Information regarding the two cameras can be found on the official REM website [Molinari *et al.* \(2020\)](#). The infrared camera of REMIR is working from 1100 to 2400 nm, while the optical camera of ROS2 covers a range from 400 to 950 nm. The transfer functions are made to take a single wavelength for each band (modelling a range of wavelength would be computationally expensive and unnecessary), so to represent each bands wavelength a pivot wavelength is used, which is the wavelength midpoint of each filter. The response curves for all the filters are on the REM website and the optical filters of ROS2 are the same as the Sloan/SDSS g', r', i', z' filters. Table 1 shows the center wavelength from all the filters. The table also shows the average wavelength λ_0 , which is used in the power law normalization function for the accretion disk transfer function

4.2 Objects

Out of all the objects, two were found to have clear variation that were deemed suitable to test the model on. Those objects are F51 and AKN120. Figures 7 and 8 show the flux ratio for the two objects. see appendix B for plots of the other objects analysed. Keep in mind that for the optical data, in many cases multiple measurements were taken minutes apart on a given day, as the images were not co-added in the analysis. This shows the scatter of the data on very short timescales. An aperture diameter of 11

Band	Wavelength[nm]
J	1250.0
H	1625.0
K	2150.0
g	475.4
r	620.4
i	769.8
z	966.5
λ_0	1122.4

Table 1: Center wavelength for the J,H,K,g,r,i,z filters and their average wavelength used in the normalization function for the accretion disk transfer function.

arcsec was chosen for F51, while 12 arcsec was deemed appropriate for AKN120. Comparing the two objects, F51 seems more noisy than AKN120.

5 Creating synthetic data

²⁰ To be able to test if the model gives sensible and reproducible results, I have created a data generation program that can create a synthetic driving light curve, convolve it with transfer functions whose parameters I specify and then obtain synthetic continuum light curves for the 7 relevant observational bands.

To start with the synthetic driving light curve is created by specifying the number of point and the range of the light curve. The time span of the driving light curve is set to be around the same as the real data, so that the resulting synthetic data spans the time range of time as the real data. To distinguish plots of the synthetic data from the real data, I plot the synthetic data with a generic "Time" x-axis, while plotting the real data with a "MJD" x-axis. Next, the hyperparameters for the exponential covariance function are specified. Firstly the timescale of variation for the driving light curve ℓ and secondly the long term standard deviation of the light curve η (they are given different names here to differentiate from the transfer function parameter names). It should be noted that since this is created using PyMC3's covariance function the value ℓ is related to the true timescale of variation by $\ell_{true} = 2\ell^2$. The mean function is set to be zero since the real data is mean subtracted and should vary around 0. Both the mean- and covariance function are passed through a multivariate normal distribution to create the points for the synthetic light curve. Figure 9 shows an example of a generated driving light curve

Next is the creation of the transfer function. The general model for the transfer function is defined so that it can be called when creating the transfer function for the different observational bands. The script allows the user to specify all the parameters for the transfer function or to use pre-specified parameters. Table 2 shows the parameters used to create the transfer functions seen in figure 10.

²⁰The code to create the synthetic data can be found at <https://github.com/MalteBrinch/Master-Thesis-Code/blob/master/Generate%20fake%20light%20curves%20py.ipynb>

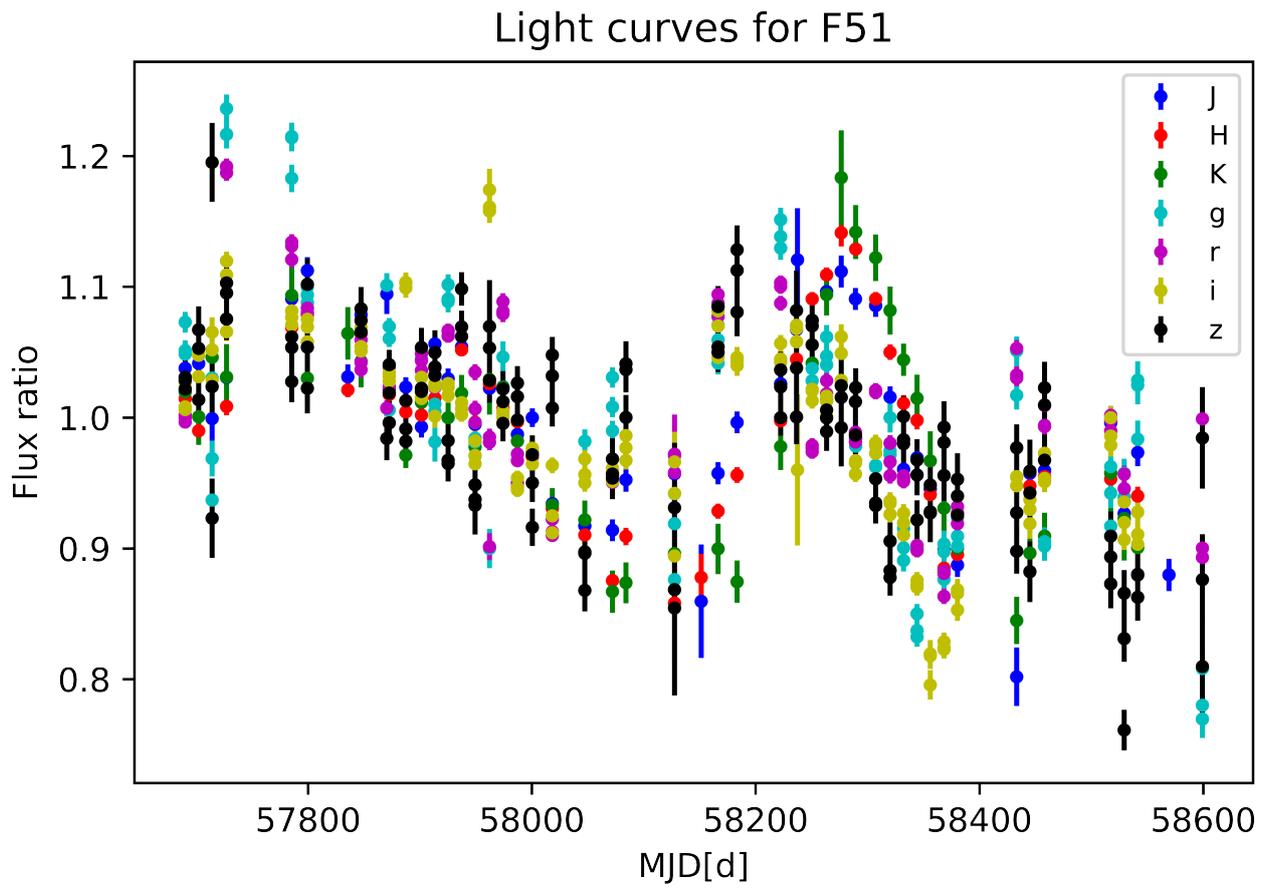


Figure 7: The flux ratio for the AGN F51. The 7 observational bands are plotted with different colors, indicated by the legend. MJD stands for Modified Julian Date.

Light curves for AKN120

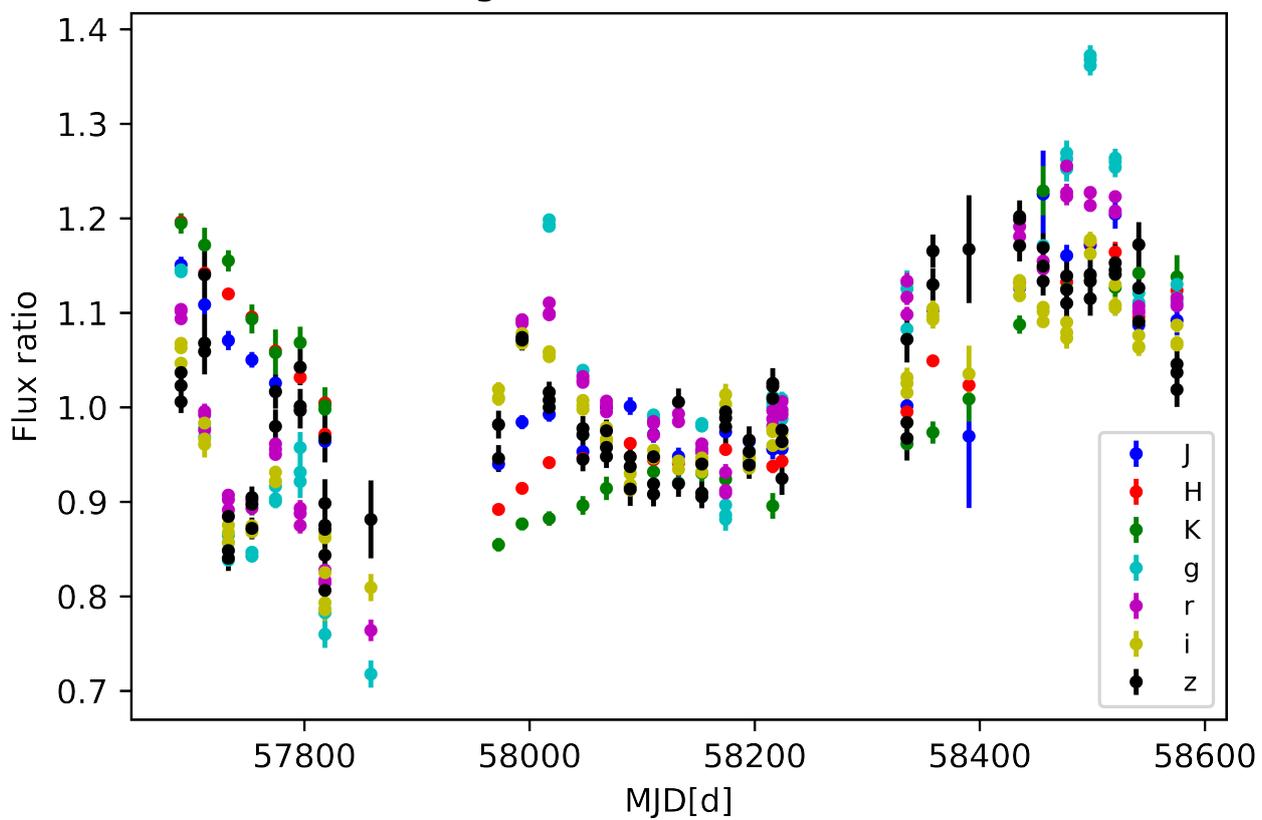


Figure 8: The flux ratio for the AGN AKN120. The 7 observational bands are plotted with different colors, indicated by the legend. MJD stands for Modified Julian Date.

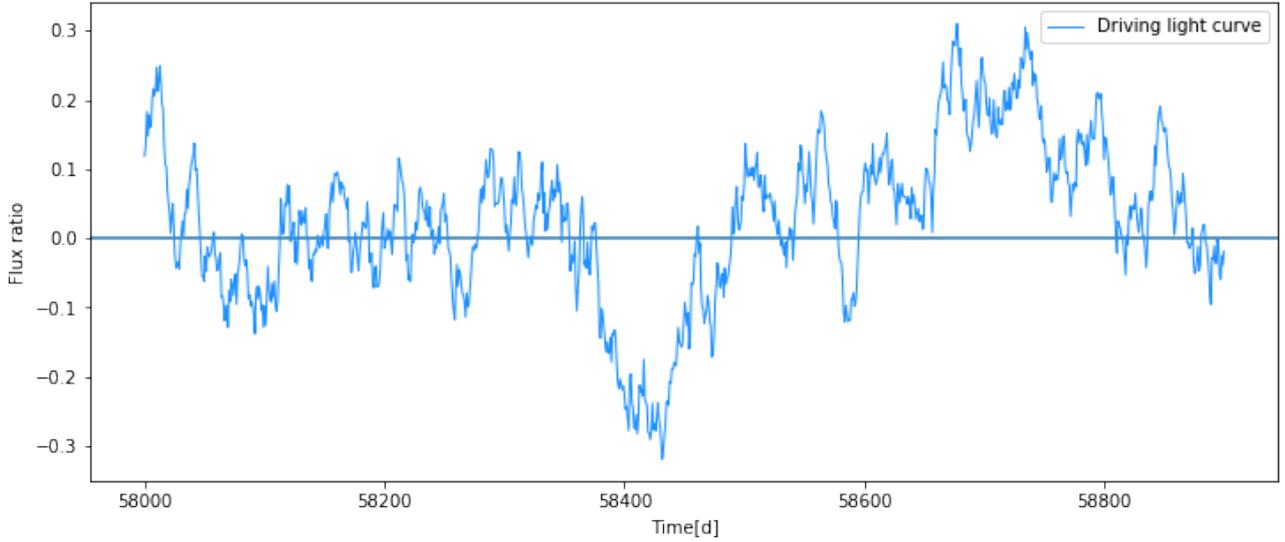


Figure 9: Plot of a 900 day synthetic driving light curve with a horizontal zero line created with the data generation script. Here $\ell = 13d^{1/2}$ (meaning $\ell_{true} = 338d$) and $\eta = 0.1mag$.

With both the driving light curve and transfer functions in hand it is then possible to create synthetic continuum data for the 7 bands. The two arrays (of equal length) are convolved for each band. Points for the synthetic continuum light curves are then picked at random to simulate the data gaps in the real data. The real data has around 30-50 points (excluding multiple points that are less than one day apart.), so 50 points are picked out from the synthetic light curves. Also the IR and optical points will be selected separately to model the different times they were measured at in the real data. To mimic the real data the measurement uncertainties need to be modelled. This is done by drawing random numbers from a truncated normal distribution. the mean and sigma values for the normal distribution are $\mu_{norm} = 0.2$ and $\sigma_{norm} = 0.1$, with the upper and lower cutoff are set to 0.005 and 0.1 (which is the upper selection limit for the real data) respectively. The truncated normal distribution was chosen since the distribution approximately resembles the distribution of measurement uncertainties in the real data. The real data is also corrupted by statistical noise, so to model this effect each point picked from the light curves will have added Gaussian noise to them. The Gaussian noise will be generated randomly from a distribution that has a mean of zero and a σ equal to the measurement uncertainty of the data point. since the noise is centered on zero, the noise will both have positive and negative values. The resulting data points with a comparison to the full uncorrupted light curves are seen in figure 11. With the appropriate synthetic data generated, it is now possible to test the convolution model before applying it to real data.

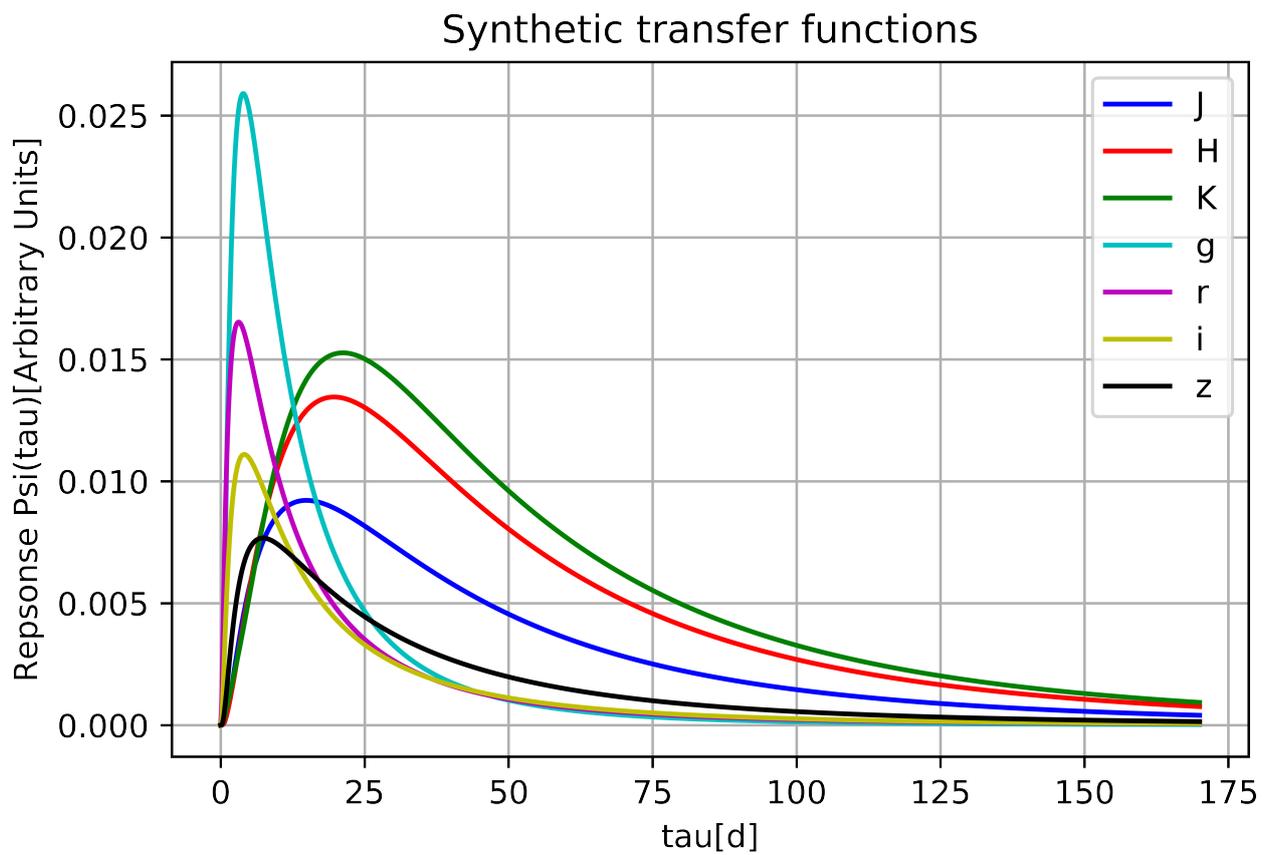


Figure 10: Plot of the transfer functions used to create synthetic data for the J,H,K,g,r,i,z bands. The parameters used to create the synthetic transfer functions are specified in table 2.

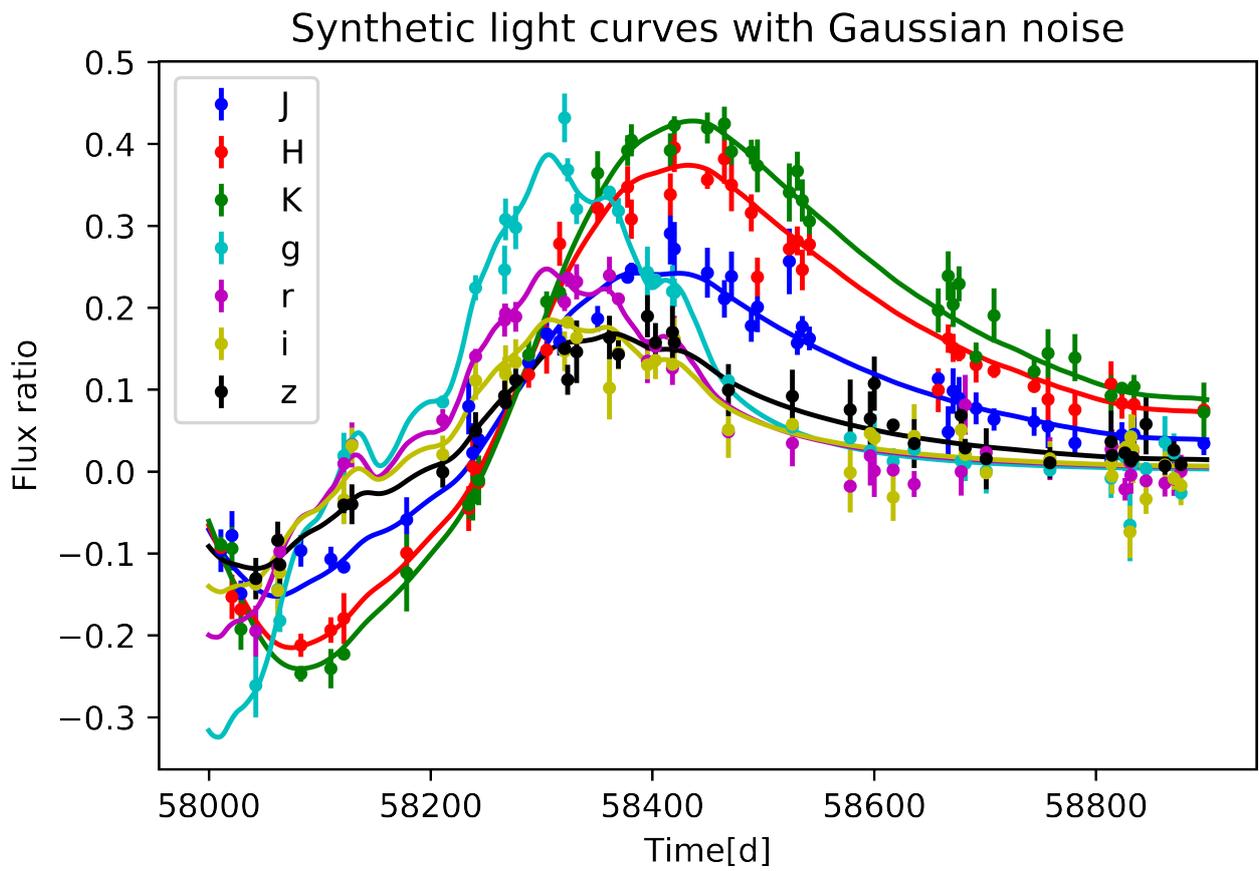


Figure 11: Synthetic continuum data for the J,H,K,g,r,i,z bands used to test the convolution model. The solid lines are the uncorrupted continuum light curves.

Name	Value
$\sigma(\eta)$	0.10 mag
$\sqrt{\frac{\tau}{2}}(\ell)$	$13.00 \sqrt{d}$
i	-1.20
K_0	0.15
T	1356 K
σ_{DT}	0.86
μ_{DT}	3.86
$\sigma_{J,AD}$	0.90
$\mu_{J,AD}$	2.87
$\sigma_{H,AD}$	0.95
$\mu_{H,AD}$	2.92
$\sigma_{K,AD}$	1.10
$\mu_{K,AD}$	2.98
$\sigma_{g,AD}$	1.10
$\mu_{g,AD}$	2.37
$\sigma_{r,AD}$	1.18
$\mu_{r,AD}$	2.53
$\sigma_{i,AD}$	1.12
$\mu_{i,AD}$	2.66
$\sigma_{z,AD}$	1.0
$\mu_{z,AD}$	2.82

Table 2: Parameters used to create the driving light curve of figure 9 and the transfer functions of figure 10. The parameters are rounded off to two decimals. Refer to chapter 1.3 for the definition of the different parameters.

6 Model setup and testing

²¹ With all the tools in place it is now possible to test the model of the driving light curve and the parametric form of the transfer function to determine the lag to the dusty torus. The model is defined inside PyMC3’s model context. To test if the model is self consistent it will be applied to the synthetic continuum data. If the result are within an acceptable range, especially when it comes to the parameters for the torus, then the next step is to apply them to the objects F51 and AKN120.

Firstly priors are defined for the hyperparameters and model parameters. Next the GP and transfer functions are defined. In the case of the GP over a time array spanning the data and for the transfer functions over a time delay array going from 0-170 to get most of the tail of the sum of the two lognormal distributions. The driving function and the transfer function are both passed to a convolution

²¹A full example of the code descipted in this section using the synthetic data can be found at <https://github.com/MalteBrinch/Master-Thesis-Code/blob/master/Example%20of%20model%20using%20synthetic%20data..ipynb>

function. This convolution has a slight quirk that has to be taken into account. PyMC3 is build on top of Theano and the function provided in Theano to apply convolution only has the border mode called "half" as opposed to the usual "same" border mode that gives an array the same size as the input arrays. the half border mode gives the same result as the same one only when the array size of the inputs are uneven. To take this into account the time array used to define the GP and the time delay used to define the transfer functions have an uneven number of entries. The resulting convolution is then linearly interpolated at the times specified by the data of each observation band to be able to compare them to the data. Lastly the model predicted points are compared with the data using a normal likelihood function. The likelihood is the same as the one used in [Pancoast et al. \(2011\)](#) and is chosen to be Gaussian, centered around the model-predicted continuum flux timeseries

$$p(D|\theta) = \prod_{i=1}^n \frac{\exp\left(-\frac{1}{2} \left(\frac{y_{i,cont} - m_i(\theta)}{\kappa\sigma_i}\right)^2\right)}{\kappa\sigma_i\sqrt{2\pi}} \quad (48)$$

where κ is a “noise boost” parameter to account for the presence of unknown systematic effects not included in the reported error bars, such as those due to flux calibration, wavelength calibration, and continuum subtraction. For the test example using the synthetic data κ will be set to one to test how well the parameters are recovered without changing the likelihood. the likelihood is defined separately for each band, since PyMC3 can handle defining multiple likelihoods in one model.

Once the sampling is done the trace plots for the hyperparameters and model parameters are inspected to check for any odd behavior during sampling. To compare the final model predicted points with the data, they are plotted together and their residual’s are determined as well to check if the model is updating properly. The reduced χ^2 is also calculated to determined how well the model predicted points fit well with the data. To check the transfer functions, The transfer function are plotted for each band with an accompanying error band. In the case of the test example using the synthetic data the transfer functions can be directly compared to their predetermined counterparts. Lastly, samples from the GP posterior are plotted to check how the driving light curve behaves. Again for the synthetic example the GP posterior is compared to the predetermined driving light curve.

6.1 Defining priors

Both the priors for the hyperparameters and the model parameters are defined by what is essentially a uniform prior, but with a twist. [Gelman \(2020\)](#) recommends not to use uniform priors or any hard constraints, unless the bounds represent true constraints. If a parameter for example is believed to be anywhere from 0 to 1, a normal distribution with a mean μ of 0.5 and a σ value of 0.5 should be used instead. For many of the parameters in the model there is a constraint that the parameter has to be positive. To achieve this without setting a bound, the parameter can initially be defined as its log value and then converted back to its linear values. This makes it so negative parameter values only ever gets closer and closer to 0. To help the sampler as much as possible to get the ideal set (where you would like the sampler to take samples from), all the parameters have also had their mean "taken out" when they are defined. While the NUTS sampler can adapt its path length, it can only do so for one scale, meaning it can potentially have problems exploring parameter space if the value of one or more

parameters are on a different scale than the rest of the parameters. This is also why it is advantageous to define the log of the parameter first and then convert it back when it is passed to the model to do sampling, as it brings all the parameters close to the same numerical scale.

Defining a prior this way is best explained with an example. Say a prior for a parameter it believed to be between a lower value 1 and an upper value 3 and that it is constrained to only positive values. Following the [Gelman \(2020\)](#) example the prior would be a normal with a mean of 2 and a σ value of 1. To convert this to the log value the following conversion is done

$$\mu_{\ln} = \frac{\ln(\text{upper}) + \ln(\text{lower})}{2} \quad (49)$$

$$\sigma_{\ln} = \frac{\ln(\text{upper}) - \ln(\text{lower})}{2} \quad (50)$$

then lastly the mean is subtracted so the final prior will be

$$\theta_{\text{prior}} = \exp(\mu_{\ln} + \text{Normal}(\mu = 0, \sigma = \sigma_{\ln})) \quad (51)$$

6.2 Setting up the GP

To start with the GP has to be defined. The GP is defined by a GP prior by specifying a number of points over a range of slightly before to slightly after the range of the data. The parameters for the GP defined at those points is treated the same way as the model parameters, where every new proposal for the continuum function related to the one before. Further interpolation is done linearly between the points in the GP so as to compare with the data. When considering how many points should be used for the GP prior two factors should be taken into consideration, resolution and computational speed. Both the synthetic data and the real data have a range around 900 days, meaning 900 points for GP prior would roughly be spaced one day apart. [Pancoast et al. \(2011\)](#) suggest using 500 points for their simulated light curve that spans roughly 120 days, stating that 500 continuum parameters is sufficient to render the distance between continuum flux points much smaller than the maximum monitoring cadence, allowing them to resort to linear interpolation only on scales not probed by the data. Following the example of [Pancoast et al. \(2011\)](#) I would need to define around 4500 parameters for to obtain the same resolution for the driving light curve, but there is a catch. Each point in the GP is treated as a parameter and the latent variable GP implementation has an $O(n^3)$ complexity, meaning adding new points to points to the model quickly bogs the sampler down as it tries to sample all the driving light curve parameters. To have a reasonable sampling time, the number of points for the GP has to be limited. MCMCrev used in [Landt et al. \(2019\)](#) takes around a day to sample, so I have chosen to specify the number of GP points that gives a sampling time of around a day. From testing with different data sets, both synthetic and real, this number seems to be around 450-500 points, meaning for a light curve spanning 900 days that is a point a little under every second day. Since I am trying to probe the large scale trend of the light curves the spacing of the GP should not provide a problem. It should also be noted that the simulated light curves in [Pancoast et al. \(2011\)](#) are sampled to a higher degree than the data in this project. The data in this project is spaced around 20-30 days apart.

Next the priors for the hyperparameters of the GP needs to be defined. Stan (2017) have specific recommendations for specifying the prior for the timescale of variation τ (length scale) and the long term standard deviation σ for the GP. For values of τ below the minimum spacing of the covariates (that is smaller than the spacing between the points in the GP) the GP likelihood reaches a plateau. If the likelihood is not regularized by a prior, the resulting flat likelihood induces considerable posterior mass at small values of τ where the observation variance drops to zero and the functions supported by the GP would be ones that exactly interpolate between the input data. The resulting posterior will not only significantly overfits to the input data, but also become difficult to accurately sample using Euclidean HMC. If we have high-frequency covariates in our fixed effects (A fixed effects model has fixed parameters or non-random quantities), then it may be necessary to regularize the GP away from high-frequency functions, which in turn means that smaller values of τ needs to be penalized. There is a way to do this by understanding how τ affects the frequency of the functions supported the GP. Imagine a case where a zero-mean GP is repeatedly drawn from with a variation timescale τ in a fixed-domain $[0, T]$ where T is the span of the light curve. In this case the result would be a distribution for the number of times each draw of the GP crossed the zero axis. The expectation of this random variable, the number of zero crossings, is given by $u = \frac{T}{\pi\tau}$. It is clear that as τ decreases, the expectation of the number of upcrossings increases as the GP is representing higher frequency functions. It is therefore a good idea to keep this statistic in mind when setting a lower bound for the prior on τ in the presence of high-frequency covariates. It should be noted that this statistic is only appropriate to use for one-dimensional inputs, though this the exact case for the AGN continuum light curves. The problem with applying this lower bound is that there is no data for the driving function that could inform the choice. The continuum data for the 7 bands could be used, but it is expected that the driving function varies more than the observed bands due to the smoothing from the transfer function and therefore u would be higher. The result would be an overestimation of the lower bound. Considering the data and resulting driving function from Landt *et al.* (2019) seems to suggest that the number of upcrossings for is twice the amount for the driving function that it is for the observed data. From generating a multitude of synthetic driving light curves, I have also come to the conclusion that having a values of $\tau < 50$ days seem to be give too high frequency functions. On the other end of the spectrum for a value of τ that is larger than the span of the data will give a GP posterior that is practically linear (with respect to the particular covariate) and increasing τ beyond that point has little impact on the likelihood. This will introduce nonidentifiability into the model as both the fixed effects and the GP will explain similar variation. In order to limit the amount of overlap between the GP and the linear regression the upper limit of τ is set to the length of the light curve. As for the prior for the long term standard deviation, the range of standard deviations of the 7 bands is used to inform prior. This is done under the assumption that the transfer functions are normalized to one. The highest and lowest value of all the standard deviations from the 7 bands is used for the prior, so if the range of standard deviation were from 0.1-0.2, then the prior would be between there. Both priors are defined as the their natural logarithm first and then converted back to linear units when setting up the covariance function. For the long term standard deviation this is done to only have positive values, since negative values of σ would also work for σ^2 (another way to around this is to model the long term variance instead). These are very rough estimates of the value for τ and more work needs to done to limit the possible values for the hyperparameters.

6.3 Setting up the model parameters

The priors for the model parameters should incorporate the general knowledge of AGN and the dusty torus and accretion disk. To start with are the σ values for the lognormal distributions. From testing multiple different values and considering the numerical values of lognormal distribution with different σ values, I have concluded that a prior from 0.1 to 2.0 for the σ values for both the torus and accretion disk should be adequate to model the possible shapes for the lognormal distributions. The prior for the logarithm of the median value of the lognormal distribution is split into two categories, one for the torus and one for all the accretion disks. For the torus value a prior of 10-100 days is set, informed by our general knowledge of the size of the dust sublimation radius for the torus. For the accretion disk a value between 1-20 days is used, again informed by our general knowledge of AGN accretion disks (([Hawkins, 2006](#)) and [Kokubo \(2018\)](#)). Remember the model is working with the lognormal distribution of equation 30, so the parameter is defined as the log value, but is not converted back to its linear value. Next are the priors for the normalization function parameters. The dust sublimation temperature for the torus black body is set to be between 1000-2000 Kelvin so that the mean is 1500 Kelvin. It is expected that the model is weakly dependent on the temperature of the Black body and the prior is kept fairly wide to take the uncertainty with regards to the dust sublimation temperature into account. For the power law the normalization constant prior is set to be between 0.1-10 and is expected to be of order unity. The power law index prior is set to be between (-1)-(-2) but like the other parameters has no boundary to allow its value to go outside that range.

6.4 Initializing the sampler

HMC samplers in general are not affine invariant (as opposed to emcee) and therefore the "metric" or "mass matrix" used has an effect on the performance of this method. PyMC3 allows for multiple different way to initialize the sampler and it is worthwhile to consider which one to use for this project to have a sampling that give good samples the posterior and is not too computationally expensive. The main choices for the sampler initialization are

- MAP: the Maximum A Posteriori tries to find the mode of the posterior distribution and start the sampler there. This should never be used for high dimensionality problems as it tends to have odd behavior in those cases.
- Jitter+Adapt_diag: This is the standard initialization in PyMC3. The Adapt_diag initialization starts with a identity mass matrix and then adapts a diagonal based on the variance of the tuning samples. All chains use the test value (usually the prior mean) as starting point. The Jitter+Adapt_diag initialization is the same, except for the fact that uniform jitter is added in [-1, 1] to the starting point in each chain. This initialization can have problems with bounded distribution, as the sampler can set it starting point outside the supported region.
- Adv+Adapt_diag: Runs ADVI and then adapts the resulting diagonal mass matrix based on the sample variance of the tuning samples. ADVI stands for Automatic Differentiation Variational Inference and can be seen as an approximate version of MCMC samplers. Instead of drawing samples from the posterior, these algorithms instead fit a distribution (e.g. normal) to the posterior,

turning a sampling problem into an optimization problem. This approach is not as flexible as MCMC but the result is a much faster algorithm and if the posterior is approximately Gaussian it provides a good way to initialize the sampler.

- Nuts: Runs NUTS and estimates posterior mean and mass matrix from the trace. also sets a full mass matrix to get rid of parameter correlations. While this method is the best out of all the choices here when it comes to determining the mass matrix, it is computationally expensive, as the sampler has to run the entire chain twice (though it is possible to set a lower number of sampler for the first run).
- Dense mass matrix: Dan Foreman Mackey has developed a way to determine the dense mass matrix (off diagonal elements in the mass matrix), which has led to performance increases in order of magnitudes for certain problems. This is not currently part of PyMC3, but Mackey's "exoplanet" package in Python and can interface with PyMC3.

From testing all of these initialization on the model in this project, it seems that Adv+Adapt_diag is the initialization that best fits the dual criterion of giving good samples and being the least computationally expensive. While the Dense mass matrix initialization seemed promising from initial testing, as the sampler was working a lot faster than in the other cases, the sampler either gets stuck or explores very little of the full parameters space. Furthermore, where the sampler gets stuck is off from the true value of the parameters when running the initialization with the test case.

To get the sampling done in a reasonable amount of time, 1000 tuning samples and 1000 actual drawn samplers will be used. This number might seem low compared to those usually used with a MH sampler, but NUTS also obtains a higher number of effective samples since the acceptance probability is higher. The tuning phase is what helps the sampler get the ideal set before it starts to take actual samples. Another important setup parameter is the number of chains that the sampler is going to sample. It is a good idea to sample multiple chains, as it can show problems with the sampler exploring parameter space, such as multimodal posterior distributions or initialization sensitivities. A value to keep in mind in this case is the Gelman–Rubin statistic, which roughly measures how similar these chains are. Ideally it should be close to 1 and anything over 1.4 indicates problems with the model specification. This project will use two chains when sampling.

6.5 Testing the model on synthetic data

To start with the model is run with the synthetic data seen in figure 11 created with the transfer function parameters of table 2 and the synthetic driving function of figure 9. The results of the sampling are seen in figure 12. The figure shows a table with the mean and standard deviation for the posterior distribution of all the parameters. hpd_3% and hpd_97% indicate the lower and upper value of the credible interval, which is the interval within which an unobserved parameter value falls with a particular probability. Think of the lower value as including 3% of the posterior and the upper value as including 97%. The next set of values "mcse" is the Markov Chain Standard Error, this indicates the error due to using the sampler. Longer chains will have a lower mcse. when compared to the standard deviation it is roughly around 1% of its value. "ess" stands for effective sample size. for MCMC the draws are correlated and

the ess will generally be lower than the draws. It is therefore used to determine if the MCMC model has reached convergence. Lastly r_{hat} is the Gelman-Rubin statistic. It is clear from the table that the two chains are close to each other for all the parameters (this is also the case for the GP parameters.). A value over 1.05 would indicate slight problems, but the highest value is 1.02.

Figure 13 shows the trace plots for all the model parameters and hyperparameters. Looking at the trace the sampler seems to explore the parameter space okay and for many of the parameters their true value is within the posterior distribution. The main outliers are the normalization function parameters (index, K_0 and T), the accretion disk function parameters for the H,K band and the long term standard deviation of the driving light curve (η). For the normalization function parameters, the lower index value will mean the difference between the optical and NIR bands will be greater for the amplitude of the accretion disk transfer functions. It could indicate problems with the model that multiple parameters are off, but importantly both of the true torus lognormal parameters are close to the mean of their posterior distribution (the log of the median is within the credibility interval).

To see whether the model is actually updating properly, figure 14 shows a comparison between the synthetic data and the model input, which is used to update the model. The model input points are the result of linearly interpolating the convolution of the driving function and transfer function. The residual indicates that the model is able to "fit" the data well to within 1σ . Table 3 shows the reduced chi square statistic between the model input and the data and for all the bands the model fits reasonably well, with a value near 1.

Figure 15 shows a comparison between the true transfer function seen in figure 10 and the model transfer function. The two transfer functions for each band seem to overlap fairly well as shown by the blue shaded area indicating the 1σ region for the model transfer functions. The two bands that are the most off appear to be the z-band and the J-band. The J band especially has a peak in the model transfer function near a time delay of 0 not seen in the true transfer function. This is because $\sigma_{J,AD}$ is larger for the model than its true value, pushing the accretion disk part of the transfer function towards smaller values of τ . It is also clear that due to the increased uncertainty of the lognormal parameters for the H- and K-bands accretion disk transfer functions, there is an increase in the uncertainty at the first 20 days of the total transfer functions. The sampler seems to struggle especially with the log of the median for these two bands, as they have the lowest amount of effective samples.

Lastly is the driving function seen in figure 16. With the true value of the long term standard deviation being off by approximately twice the value of the true driving function, it is clear that the comparison between the draws from the GP posterior and the true driving function is off. Figure 17 shows that if the scale of the GP is corrected for, it actually follows the overall trend of the true driving function fairly well.

In the appendix A is a similar run of the model, the main differences being that the time scale of variation ℓ prior is being defined as its linear value and the number of data points for the bands is slightly higher at 49 points. The comparison between the two sampler runs could suggest that the sampler is sensitive to whether the timescale of variation is defined as the log or linear value first, a point

	mean	sd	hpd_3%	hpd_97%	mcse_mean	mcse_sd	ess_mean	ess_sd	ess_bulk	ess_tail	r_hat
Index	-0.037	0.123	-0.260	0.193	0.004	0.003	1143.0	1143.0	1155.0	690.0	1.00
K_0	-2.176	0.111	-2.371	-1.964	0.004	0.003	638.0	638.0	635.0	672.0	1.00
T	-0.059	0.013	-0.085	-0.035	0.001	0.001	286.0	252.0	309.0	695.0	1.01
zm_AD	-0.732	0.229	-1.147	-0.305	0.013	0.009	307.0	307.0	301.0	624.0	1.00
zsigma_AD	0.639	0.202	0.265	1.025	0.010	0.007	407.0	407.0	402.0	663.0	1.00
im_AD	-0.945	0.258	-1.404	-0.479	0.016	0.011	262.0	262.0	260.0	432.0	1.00
isigma_AD	0.563	0.205	0.172	0.931	0.011	0.008	319.0	319.0	317.0	592.0	1.00
rm_AD	-1.099	0.299	-1.663	-0.564	0.019	0.013	255.0	255.0	248.0	403.0	1.00
rsigma_AD	0.720	0.213	0.343	1.142	0.014	0.010	227.0	226.0	228.0	467.0	1.00
gm_AD	-1.116	0.289	-1.614	-0.551	0.018	0.013	245.0	245.0	244.0	372.0	1.00
gsigma_AD	0.649	0.197	0.280	1.026	0.013	0.009	235.0	235.0	234.0	472.0	1.00
Km_AD	0.687	0.514	-0.128	1.756	0.071	0.051	52.0	52.0	125.0	33.0	1.02
Ksigma_AD	-0.163	0.524	-1.160	0.801	0.022	0.024	586.0	232.0	811.0	624.0	1.00
Hm_AD	1.828	0.891	0.103	3.524	0.134	0.096	44.0	44.0	107.0	31.0	1.02
Hsigma_AD	-0.514	1.298	-2.771	1.835	0.064	0.046	406.0	406.0	411.0	757.0	1.00
Jm_AD	-0.479	0.463	-1.312	0.336	0.031	0.025	222.0	172.0	170.0	231.0	1.02
Jsigma_AD	1.122	0.312	0.562	1.704	0.025	0.018	153.0	153.0	146.0	386.0	1.02
m_DT	0.334	0.053	0.226	0.427	0.002	0.002	458.0	452.0	457.0	875.0	1.00
sigma_DT	0.496	0.104	0.292	0.691	0.007	0.005	257.0	257.0	260.0	397.0	1.00
‡	0.571	0.448	-0.279	1.350	0.017	0.012	696.0	696.0	699.0	916.0	1.00
η	0.640	0.206	0.236	1.010	0.009	0.007	526.0	493.0	546.0	754.0	1.00

Figure 12: Table of the parameter values and statistics from the test using the synthetic data. The definitions of the different values are explained in the text. The statistics for the GP parameters are not shown, since there is simply too many for it to make sense.

to keep in mind when defining the prior using the real data. The synthetic sampler runs could indicate that there is a degeneracy in the model, where multiple amplitudes of the transfer functions and the driving function can provide equally good fits to the data. It is important to keep in mind that the main goal is to sample the posterior distribution for the log of the median μ_{DT} and the standard deviation of the log of the distribution σ_{DT} for the dusty torus transfer function. While it would be advantageous to sample the posterior of all the parameters, since they shed light on the physical characteristics of the object being investigated (like the dust sublimation temperature), it is not a necessary requirement to obtaining the lag. The results indicate that the sampler is able to recover many of the transfer function parameters and the overall transfer function with some uncertainty, which means it should be feasible to apply the model on real data, since the synthetic data has been created to mimic the real data, through the selection of a similar distribution for the uncertainties and the adding of statistical noise to the synthetic data.

Band	χ_{red}^2
J	1.334
H	1.378
K	1.065
g	1.200
r	1.192
i	1.550
z	0.809

Table 3: Reduced χ^2 statistic for the J,H,K,g,r,i,z filters using the synthetic test data. The values are computed from the model input and the data seen in figure 14.

6.6 Testing the model on real data

The procedure for using the real data is the same as the synthetic one. The first object considered is AKN120. Notably this run took longer than the one for the synthetic data (30 hours as supposed to 18 hours) and the loss function for the ADVI initialization was higher in this case (lower is better). It is clear going through figures 18 to 22 that the model has more problems in this case. The number of effective samples is lower for many parameters. The worst case is the accretion disk median lag for the K-band, where there is only 3 effective samples. Looking at the trances, some chains do not overlap and the exploration of parameter space is worse for many of the parameters, especially the K-band median accretion disk lag. looking at figure 20 the model has not properly updated at all and figure 21 confirms that the transfer function either have very small values or in the case of the i-, z- and K-bands no value at all, as the entire transfer function is zero across the specified time delay array. This is the due to the fact that the parameters are so off that the transfer function has left the time span supported by the time delay array. Interestingly, the driving function does follow some long term trend, but it is not as variable as the synthetic driving function. I am omitting plotting the run of F51, since they are essentially identical and suffer from the same problems.

Band	χ_{red}^2
J	91.535
H	317.873
K	120.609
g	494.698
r	600.017
i	168.448
z	36.835

Table 4: Same description as table 3, but uses the data seen in figure 8

7 Discussion

To try and determine what is causing the model not to update properly when working with the actual data, I have tried a multitude of different approaches, affecting different aspects of the model. This includes everything from the shape of the transfer function, the normalization functions, the driving function and the likelihood definition. A full list includes

- Using a single lognormal for the transfer function for each band.
- Using lognormals with a location parameter θ .
- Using a normal distribution instead of a lognormal.
- Using an Exponentially modified Gaussian. This has major problems with divergences when sampling (effectively every draw diverges), possibly due to the fact that the error function and complementary error function in Theano has problems calculating the gradient.
- Used a longer time delay array to define the transfer functions over. Multiple values were tried like 200, 250 and 500 days.
- Most of the calculations done by the model were rewritten to be done in the log first and then converted back to its linear value. This was done to make sure there were no problems with numerical precision when using the model.
- More points were added to the GP prior to better represent the driving function. The number was upped to 900 points, so that the separation between points was 1 day instead of 2 days. This upped the sampling time to around 70 hours.
- Tried to change the normalization of the dusty torus from a single normalized black body to a power law. The power law was defined like the one for the accretion disk with an order of unity constant in the front and an index to control the difference between different bands.
- Tried to change the normalization of the dusty torus to a set of generic parameters, one for each band.
- The Fourier series model of [Starkey et al. \(2016\)](#) for the driving function was implemented. The Fourier series needs a for-loop to work, which is expensive in Theano. The resulting sampling time was around 70 hours and the driving function gave erroneous results, so this approach was discarded.
- The noise boost factor κ in the likelihood from [Pancoast et al. \(2011\)](#) was set as a parameter with a prior somewhere between 1-5. The parameter value would have an absurdly high value after sampling, in the order of 20 and was therefore discarded as an option. Both a single value of κ for all bands and separate values for each band was tried, with similar results.
- The differential magnitude was used as input for the model instead. This was encouraged by [Kelly et al. \(2009\)](#), but the interpretation of the CAR(1) is unclear when working with the

magnitude. Also the changes to the normalization functions for the transfer functions are not well understood. The index of the power law changes from negative to positive when working with magnitudes, but what about the black body? If the differential magnitude is to be used as input for the model, more work needs to be done understanding how this changes the parameterization of the transfer function and the driving function.

- Different values for the hyperparameters were tried. The timescale of variation was set to both a very small value and a very high value, relative to the usual definition of the prior. The long term standard deviation was also set to be high and lower than what would be specified by the data.
- The optical data points that are less than a day apart were combined to test if the sampler had trouble working with data that close together in time.

For all of the different approaches tried above one or more of the following things happened when working with the real data.

- The Gelman-Rubin statistics for all or most of the parameters were over 1.4, around a value of 2.0 signaling clear model misspecification.
- The compression between the model input and the data showed that the model input had a very small value, with the K-band usually being having the value closest to the data, but still far off. The reduced χ^2 statistic was far from 1, around 70-600.
- In combination with the model input not updating properly, the transfer functions would either have very small values or they would be outside the times supported by the time delay array. Inserting a bound on the parameter priors would only push the posterior of the parameters right up against the bound.
- The shape of the driving function would often be very "spiky" with values much higher at certain points of the light curve. The scale could in some cases also be way off, with flux ratio values near 500-1000.

The question remains as why the model does not seem to update properly. The two main reasons for these problems I can think of is either it is related to the definition of the model (driving function or transfer functions) or it is related to the data. There is pretty good evidence for the CAR(1) model being good enough to represent the driving light curve, but it could be that the priors for the GP hyperparameters is off. I suspect the main culprit is the long term standard deviation, since trying to inform the prior by using the data could be a mistake. Having the driving function at the same scale as the data only works if the total transfer function is normalized to about 1. From creating multiple different synthetic transfer functions with my synthetic data generation script, I found that the transfer function are normalized to be anywhere from 0.1-3.0, which while not being too far off from 1.0, could still prove to be a problem when trying to sample the posterior for η .

The other part of the model that there could be a problem with is the transfer function. I do not think that the shape of the transfer function (both for the accretion disk and the dusty torus) is the problem, since a number of papers has found the shape of the transfer function to be roughly a

lognormal. It is more likely that it is the normalization functions that are the problem, especially the one for the dusty torus. For the power law normalization function the difference in the value of the normalization function between K- and g-band is at most one order of magnitude so from 0.1-1.0 for reasonable choice for K_0 and the power law index, while the torus normalization function it can be several orders of magnitude, so that the K-band value is close to 1.0, while the g-band value is 0.0001.

Why does this matter? Should we not expect the transfer function to be different between the torus and the accretion disk? Compare the synthetic of figure 11 and 25 with the data for F51 and AKN120 seen in figure 7 and 8. The light curves of different bands for the synthetic data seems to be more spread apart than for the real data. The different bands for the real data are fairly close together and they vary in a mostly similar fashion throughout the span of the light curves. This indicates that the value of the transfer functions are also fairly close to each other, at least the amplitude is. It is not possible for the transfer function to be so similar if the value of the torus transfer function for the optical bands is so small that the torus transfer function contributes almost nothing to the overall transfer function for the optical bands. Another way the similar variations could happen is if the transfer functions for the accretion disk was more similar to the torus ones, but that way betray our knowledge of AGN structure, since that would indicate that the accretion disk is as far out (or close to as far out) as the torus.

It is also clear that while the level of variability of the optical and the NIR light curves are closer for the real data, for the synthetic data there is a drop off at later times in the light curve, especially for the optical data. This difference can be explained by the fact that the optical transfer functions have less of a tail that goes towards larger time delays. In fact all the values for the transfer function decrease as later time delays and this coincides with all the light curves going towards zero at later times in the synthetic light curves. One way to account for this would be to have different time delay arrays for each band, but then it would require a prediction beforehand of how much of a tail the transfer function would have. A smaller time delay array could also be used for all the transfer functions, but then the tail from the NIR transfer functions would not be included. Both cases do still would not explain why the real data is so close together and varies mostly in the same way. Looking at the transfer functions of Landt *et al.* (2019), all the transfer function there are also defined over the same range of time delays. The simplest explanation seems to be that the normalization for the torus transfer function is wrong in some capacity, but in what way it should be changed is still unclear. I have tried to implement different versions of the normalization function for the torus, but they run into the same problems as the original one, as the model does not update properly. It is clear that more work needs to be done to figure out what the normalization should be when using a parametric form of the transfer function.

One way to help the sampler determine the hyperparameters better would be to have data for the driving function. The problem in this case (other than getting the data) is to determine what part of the spectrum is the driving function, other than the general knowledge that is it in the UV. Landt *et al.* (2019) showed the difference in the lag of the torus if the V band was used as the driver and difference was tens of days longer for the mean lag than if the driving function was inferred from the data. It is straightforward to introduce data from the driving function into the model, as the GP array

can be linearly interpolated to compare the model value with the driving function data.

Lastly it is important to consider any problems with the data itself. From an analysis of the objects by Bo Milvang-Jensen found that there are variation in the star magnitude difference with the Position Angle (PA), which points to the fact that there are variations in the Point Spread Functions (PSF) within each image used, which causes scatter as a result. It should be possible to construct aperture correction maps that potentially could remove this effect, though it is not trivial as the number of bright and isolated stars are low in any given image, but there is perhaps a pattern that is similar in different images. This scatter is likely to make it more difficult to determine the lag, but it should not be so severe that it would result in what is seen with AKN120.

8 Conclusion

This project has shown the process of creating a model using a state of the art MCMC sampler in the form of PyMC3 to determine the time lag to the dusty torus in AGN. The act of generating a good algorithm using computationally fast and reliable code that could recover the model parameters required extensive knowledge of physical structure of AGN, but also the mathematical models that make it possible to try and solve the deconvolution problem, like the CAR(1) model used to characterize the driving function or the specific form of the transfer function and its normalization. Determining the time lag would be a great step in using AGN as a standard rulers, so that an independent way of measuring the Hubble constant at mid to high redshift would become available and more light can be shed on the discrepancy between the early universe CMB approach and the late universe distance ladder approach to determining the Hubble constant. The creation of a synthetic data generation script gives an easy way to define a synthetic driving light curve and multiple synthetic transfer function and to convolve to obtain synthetic NIR and optical data. The synthetic test case showed that is it possible to construct a self consistent example where the lag can be recovered with some uncertainty given a set of synthetic light curves as input. Multiple runs of the sampler with synthetic data has also showed that this can be done reliably. as for the performance of the model, it is reasonably fast considering the complexity of the model and the number of parameters in play. The difference in sampling time when using the real data hits at the Folk Theorem of Statistical Computing which states "If you're having computational problems, then it is probably because your model is wrong.". it becomes even more apparent that something is wrong with the model when inspecting the resulting figures from running the model with the data from AKN120.

It is clear that more work needs to be done to either learn more about the hyperparameters for the GP and/or the normalization for the transfer function. Any future efforts to use a parametric form of the transfer function should put their focus there. Another potential improvement would be to introduce a more general covariance function, whose β value would be a parameter that could differ from 1. This would allow a smoother driving function, which may capture the variability of some objects better. A good way to improve on the data side would be to have better sampled light curves and to find a way to obtain UV data that could be used to learn about the driving function more directly.

References

- Almeida, C. R., & Ricci, C. 2017. Nuclear obscuration in active galactic nuclei. *Nature Astronomy*, **1**.
- Andrieu, C., de Freitas, N., Doucet, A., & Jordan, M. I. 2003. An Introduction to MCMC for Machine Learning. *Machine Learning*, **50**.
- Atzberger, Paul J. 2020. The Monte-Carlo Method.
- Bianchi, S., Maiolino, R., & Risaliti, G. 2012. AGN Obscuration and the Unified Model. *Advances in Astronomy*.
- Braak, C. J. F. T. 2016. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, **16**.
- Brewer, B. J., Pártay, L. B., & Csányi, G. 2011. Diffusive Nested Sampling. *Statistics and Computing*, **21**.
- Chassande-Mottin, E., Leyde, K., Mastrogiovanni, S., & Steer, D. A. 2019. Gravitational wave observations, distance measurement uncertainties, and cosmology. *Phys. Rev. D*, **100**.
- D. van Ravenzwaaij, P. Cassey, S. D. Brown. 2018. A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin & Review*, **25**.
- Davis, R. A. 2006. Gaussian Process, 3. Stochastic Modeling and Environmental Change. *Wiley Online Library*.
- Duvenaud, D. K. 2014. Automatic Model Construction with Gaussian Processes. *University of Cambridge, Pembroke College*.
- Edelson, R., Gelbord, J. M., Horne, K., McHardy, I. M., Peterson, B. M., Arévalo, P., Breeveld, A. A., Rosa, G. De, Evans, P. A., Goad, M. R., Kriss, G. A., Brandt, W. N., Gehrels, N., Grupe, D., Kennea, J. A., Kochanek, C. S., Nousek, J. A., Papadakis, I., Siegel, M., Starkey, D., Uttley, P., Vaughan, S., Young, S., Barth, A. J., Bentz, M. C., Brewer, B. J., Crenshaw, D. M., Bontà, E. Dalla, Lorenzo-Cáceres, A. De, Denney, K. D., Dietrich, M., Ely, J., Fausnaugh, M. M., Grier, C. J., Hall, P. B., Kaastra, J., Kelly, B. C., Korista, K. T., Lira, P., Mathur, S., Netzer, H., Pancoast, A., Pei, L., Pogge, R. W., Schimoia, J. S., Treu, T., Vestergaard, M., Villforth, C., Yan, H., & Zur, Y. 2015. SPACE TELESCOPE AND OPTICAL REVERBERATION MAPPING PROJECT. II. SWIFT AND HST REVERBERATION MAPPING OF THE ACCRETION DISK OF NGC 5548. *The Astrophysical Journal*, **806**.
- Fausnaugh, M. M., Peterson, B. M., Starkey, D. A., Horne, K., & the AGN STORM Collaboration. 2014. Continuum Reverberation Mapping of AGN Accretion Disks. *Front. Astron. Space Sci*.
- Fonnesbeck, C. 2017. Fitting Gaussian Process Models in Python. <https://blog.dominodatalab.com/fitting-gaussian-process-models-python/>.

- Freedman, W. L., Madore, B. F., D. Hatt, T. J. Hoyt, Jang, I., Beaton, R. L., Burns, C. R., Lee, M. G., Monson, A. J., Neeley, J. R., Phillips, M. M., Rich, J. A., & Seibert, M. 2019. The Carnegie-Chicago Hubble Program. VIII. An Independent Determination of the Hubble Constant Based on the Tip of the Red Giant Branch. *The Astrophysical Journal*, **882**.
- Gelman, A. 2020. Prior Choice Recommendations. *Stan Development Team*, <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>.
- Grier, C. J., Peterson, B. M., Horne, K., Bentz, M. C., Pogge, R. W., Denney, K. D., Rosa, G. De, Martini, P., Kochanek, C. S., Zu, Y., Shappee, B., Siverd, R., Beatty, T. G., Sergeev, S. G., Kaspi, S., Salvo, C. Araya, Bird, J. C., Bord, D. J., Borman, G. A., Che, X., Chen, C., Cohen, S. A., Dietrich, M., Doroshenko, V. T., S., Yu., Free, E. N., Ginsburg, I., Henderson, C. B., King, A. L., Mogren, K., Molina, M., Mosquera, A. M., Nazarov, S. V., Okhmat, D. N., Pejcha, O., Rafter, S., Shields, J. C., Skowron, J., Szczygiel, D. M., Valluri, M., & van Saders, J. L. 2013. THE STRUCTURE OF THE BROAD-LINE REGION IN ACTIVE GALACTIC NUCLEI. I. RECONSTRUCTED VELOCITY-DELAY MAPS. *The Astrophysical Journal*, **764**.
- Grier, C. J., Pancoast, A., Barth, A. J., Fausnaugh, M. M., Brewer, B. J., Treu, T., & Peterson, B. M. 2017. The Structure of the Broad-line Region in Active Galactic Nuclei. II. Dynamical Modeling of Data From the AGN10 Reverberation Mapping Campaign. *The Astrophysical Journal*, **849**.
- Görtler, J., Kehlbeck, R., & Deussen, O. 2019. A Visual Exploration of Gaussian Processes. <https://www.jgoertler.com/visual-exploration-gaussian-processes/>.
- Haardt, F., & Maraschi, L. 1991. A Two-Phase Model for the X-Ray Emission from Seyfert Galaxies. *Astrophysical Journal Letters*, **380**.
- Hawkins, M. R. S. 2006. Timescale of variation and the size of the accretion disc in active galactic nuclei. *Astronomy & Astrophysics*, **462**.
- HENDRY, M. A., SIMMONS, J. F. L., & NEWSAM, A. M. 1993. WHAT DO WE MEAN BY ‘MALMQUIST BIAS’ ? *IAP Conference*.
- Hoening, S. F., & Kishimoto, M. 2010. The dusty heart of nearby active galaxies – II. From clumpy torus models to physical properties of dust around AGN. *Astronomy & Astrophysics*, **523**.
- Hoening, S. F., Kishimoto, M., Gandhi, P., Smette, A., Asmus, D., Duschl, W., Polletta, M., & Weigelt, G. 2010. The dusty heart of nearby active galaxies. I. High-spatial resolution mid-IR spectro-photometry of Seyfert galaxies. *Astronomy & Astrophysics*, **515**.
- Hoffman, M. D., & Gelman, A. 2011. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, **15**.
- Hogg, D. W., & Mackey, D. F. 2018. Data Analysis Recipes: Using Markov Chain Monte Carlo. *The Astrophysical Journal Supplement Series*, **236**.

- Hotokezaka, K., Nakar, E., Gottlieb, O., Nissanke, S., Masuda, K., Hallinan, G., Mooley, K. P., & Deller, A. T. 2019. A Hubble constant measurement from superluminal motion of the jet in GW170817. *Nature Astronomy*, **3**.
- Huijser, D., Goodman, J., & Brewer, B. J. 2015. Properties of the Affine Invariant Ensemble Sampler in high dimensions. *arXiv:1509.02230*.
- Humphreys, E. M. L., Reid, M. J., Moran, J. M., Greenhill, L. J., & Argon, A. L. 2013. Toward a New Geometric Distance to the Active Galaxy NGC 4258. III. Final Results and the Hubble Constant. *The Astrophysical Journal*, **775**.
- Hönig, S. F. 2014. DUST REVERBERATION MAPPING IN THE ERA OF BIG OPTICAL SURVEYS AND ITS COSMOLOGICAL APPLICATION. *The Astrophysical Journal Letters*.
- Hönig, S. F., Watson, D., Kishimoto, M., & Hjorth, J. 2014. A dust-parallax distance of 19 megaparsecs to the supermassive black hole in NGC 4151. *Nature*, **515**.
- Ivezi, Z., & MacLeod, C. 2013. Optical variability of quasars: a damped random walk. *Proceedings of the International Astronomical Union, Multiwavelength AGN Surveys and Studies*, **9**.
- Jackson, N. 2015. The Hubble Constant. *Living Reviews in Relativity*, **18**.
- Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009. ARE THE VARIATIONS IN QUASAR OPTICAL FLUX DRIVEN BY THERMAL FLUCTUATIONS? *The Astrophysical Journal*, **689**.
- Kelly, B. C., Treu, T., Malkan, M., Pancoast, A., & Woo, J. H. 2013. ACTIVE GALACTIC NUCLEUS BLACK HOLE MASS ESTIMATES IN THE ERA OF TIME DOMAIN ASTRONOMY. *The Astrophysical Journal*, **779**.
- Knox, L., & Millea, M. 2019. The Hubble Hunter's Guide. *American Physical Society*.
- Kokubo, M. 2018. Constraints on accretion disk size in the massive type 1 quasar PG 2308+098 from optical continuum reverberation lags. *Publications of the Astronomical Society of Japan*, **70**.
- Koshida, S., Minezaki, T., Yoshii, Y., Kobayashi, Y., Sakata, Y., Sugawara, S., Enya, K., Suganuma, M., Tomita, H., Aoki, T., & Peterson, B. A. 2014. Reverberation Measurements of the Inner Radius of the Dust Torus in 17 Seyfert Galaxies. *The Astrophysical Journal*, **788**.
- Kozłowski, S. 2016. A degeneracy in DRW modelling of AGN light curves. *Monthly Notices of the Royal Astronomical Society*, **459**.
- Landt, H., Ward, M. J., Kynoch, D., Packham, C., Ferland, G. J., Lawrence, A., Pott, J.-U., Esser, J., Horne, K., Starkey, D. A., Malhotra, D., Fausnaugh, M. M., Peterson, B. M., Wilman, R. J., Riffel, R. A., Storchi-Bergmann, T., Barth, A. J., Villforth, C., & Winkler, H. 2019. The first spectroscopic dust reverberation programme on active galactic nuclei: the torus in NGC 5548. *Monthly Notices of the Royal Astronomical Society*, **489**.
- Lauritsen, Lyng R. B. 2018. Master Thesis, Dust reverberation mapping of AGN in the local universe. *University of Copenhagen, Niels Bohr Institute, Dark Cosmology Centre*.

- Li, Y. R., Wang, J. M., & Bai, J. M. 2016. A NON-PARAMETRIC APPROACH TO CONSTRAIN THE TRANSFER FUNCTION IN REVERBERATION MAPPING. *The Astrophysical Journal*, **831**.
- Lombriser, L. 2020. Consistency of the local Hubble constant with the cosmic microwave background. *Physics Letters B*, **803**.
- MacKay, D. J. C. 1998. Introduction to Gaussian processes. *Department of Physics, Cambridge University*.
- Mackey, D. F., Hogg, D. W., Lang, D., & Goodman, J. 2013. emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, **125**.
- Mandal, A. K., Rakshit, S., Pal, I., Stalin, C. S., Sagar, R., & Mathew, B. 2019. REMAP: Determination of the inner edge of the dust torus in AGN by measuring time delays. *arXiv: Astrophysics of Galaxies*.
- McElreath, R. 2017. Markov Chains: Why Walk When You Can Flow? <https://eleanth.org/blog/2017/11/28/build-a-better-markov-chain/>.
- Miller, J. S., & Antonucci, R. R. J. 1983. Evidence for a highly polarized continuum in the nucleus of NGC 1068. *Astrophysical Journal*, **271**.
- Mo, H., Bosch, F. Van Den, & White, S. 2010. GALAXY FORMAT ION AND EVOLUTION. *Cambridge University press*.
- Molinari, E., Covino, S., D'Alessio, F., Fugazza, D., Malaspina, G., Nicastro, L., Righi, C., Stefanon, M., Testa, V., & Vitali, F. 2020. Rapid Eye Mount. <http://www.rem.inaf.it/>, **21**.
- Moreno, J., Vogeley, M. S., Richards, G. T., & Yu, W. 2019. Stochastic Modeling Handbook for Optical AGN Variability. *Publications of the Astronomical Society of the Pacific*, **131**.
- Nenkova, M., Ivezić, Ž., & Elitzu, M. 2002. Dust Emission from Active Galactic Nuclei. *The American Astronomical Society*, **507**.
- NIST/SEMATECH. 2019. e-Handbook of Statistical Methods. <http://www.itl.nist.gov/div898/handbook/>.
- Núñez, F. P., Chelouche, D., & Kaspi, S. 2018. Accretion Disk Reverberation Mapping of Active Galactic Nuclei at Wise Observatory. *Proceedings of Science*.
- Osgood, B. 2007. Lecture Notes for EE261 The Fourier Transform and its Applications. *Electrical Engineering Department, Stanford University*.
- Pancoast, A., B. Brewer, Brendon, & Treu, T. 2011. Geometric and Dynamical Models of Reverberation Mapping Data. *The Astrophysical Journal*, **730**.
- Parikh, J. 2019. Prior over functions: Gaussian process. <https://towardsdatascience.com/prior-over-functions-gaussian-process-1c58e8c40272>.

- Peterson, B. M. 1993. Reverberation mapping of active galactic nuclei. *Publications of the Astronomical Society of the Pacific*, **105**.
- Peterson, B. M. 1997. An Introduction to Active Galactic Nuclei. *Cambridge University Press*.
- Peterson, B. M. 2001. Advanced Lectures on the Starburst-AGN Connection, 3. *World Scientific*.
- Peterson, B. M. 2008. The central black hole and relationships with the host galaxy. *New Astronomy Reviews*, **52**.
- Planck-Collaboration. 2018. Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics*.
- Poulin, V., Smith, T. L., Karwal, T., & Kamionkowski, M. 2019. Early Dark Energy Can Resolve The Hubble Tension. *PHYSICAL REVIEW LETTERS*.
- Rasmussen, C. E., & Williams, C. K. I. 2006. Introduction to Gaussian processes. *Massachusetts Institute of Technology*.
- Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019. Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics Beyond LambdaCDM. *The Astrophysical Journal*, **876**.
- Risaliti, G., & Lusso, E. 2015. A Hubble Diagram for Quasar. *The Astrophysical Journal*, **815**.
- Ryden, B. 2017. *Introduction to Cosmology*. Pearson Education, Inc., 2. Edition.
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*.
- Schneider, P. 2006. Extragalactic Astronomy and Cosmology - An Introduction. *Springer Berlin Heidelberg New York*.
- Shappee, B. J., Prieto, J. L., Grupe, D., Kochanek, C. S., Stanek, K. Z., Rosa, G. De, Mathur, S., Zu, Y., Peterson, B. M., Pogge, R. W., Komossa⁵, S., Im, M., Jencson, J., Holoién, T.W-S., Basu, U., Beacom, J. F., Szczygieł, D. M., Brimacombe, J., Adams, S., Campillay, A., Choi, C., Contreras¹¹, C., Dietrich, M., Dubberley, M., Elphick, M., Foale, S., Giustini, M., Gonzalez, C., Hawkins, E., Howell, D. A., Hsiao, E. Y., Koss, M., Leighly, K. M., Morrell, N., Mudd, D., Mullins, D., Nugent, J. M., Parrent, J., and G. Pojmanski⁹, M. M. Phillips, Rosing, W., Ross, R., Sand, D., Terndrup, D. M., Valenti, S., Walker, Z., & Yoon, Y. 2014. THE MAN BEHIND THE CURTAIN: X-RAYS DRIVE THE UV THROUGH NIR VARIABILITY IN THE 2013 ACTIVE GALACTIC NUCLEUS OUTBURST IN NGC 2617. *The Astrophysical Journal*, **788**.
- Shi, Y. 2019. Gaussian Processes, not quite for dummies. *The Gradient*, <https://thegradient.pub/gaussian-process-not-quite-for-dummies/>.
- Skielboe, A. 2016. Colossal creations of gravity; From clusters of galaxies to active galactic nuclei. *Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen*.

- Skielboe, A., Pancoast, A., Treu, T., Park, D., Barth, A. J., & Bentz, M. C. 2015. Constraints on the broad line region from regularized linear inversion: Velocity–delay maps for five nearby active galactic nuclei. *Monthly Notices of the Royal Astronomical Society*, **454**.
- Smith, K. L., Mushotzky, R. F., Boyd, P. T., Malkan, M., Howell, S. B., & Gelino, D. M. 2018. The Kepler Light Curves of AGN: A Detailed Analysis. *The Astrophysical Journal*, **857**.
- Sparke, L. S., & Gallagher(III), J. S. 2007. Galaxies in the Universe, An Introduction, SECOND EDITION. *Cambridge University Press*, **2**.
- Stan. 2017. Stan Modeling Language, User’s Guide and Reference Manual. *Stan Development Team*, <https://mc-stan.org/>.
- Starkey, D., Horne, K., Fausnaugh, M. M., Peterson, B. M., Bentz, M. C., Kochanek, C. S., Denney, K. D., Edelson, R., Goad, M. R., Rosa, G. De, Anderson, M. D., Arévalo, P., Barth, A. J., Bazhaw, C., Borman, G. A., Boroson, T. A., Bottorff, M. C., Brandt, W. N., Breeveld, A. A., Cackett, E. M., Carini, M. T., Croxall, K. V., Crenshaw, D. M., Bontà, E. Dalla, Lorenzo-Cáceres, A. De, Dietrich, M., Efimova, N. V., Ely, J., Evans, P. A., Filippenko, A. V., Flatland, K., Gehrels, N., Geier, S., Gelbord, J. M., Gonzalez, L., Gorjian, V., Grier, C. J., Grupe, D., Hall, P. B., Hicks, S., Horenstein, D., Hutchison, T., Im, M., Jensen, J. J., Joner, M. D., Jones, J., Kaastra, J., Kaspi, S., Kelly, B. C., Kennea, J. A., Kim, S. C., Kim, M., Klimanov, S. A., Korista, K. T., Kriss, G. A., Lee, J. C., Leonard, D. C., Lira, P., MacInnis, F., Manne-Nicholas, E. R., Mathur, S., McHardy, I. M., Montouri, C., Musso, R., Nazarov, S. V., Norris, R. P., Nousek, J. A., Okhmat, D. N., Pancoast, A., Parks, J. R., Pei, L., Pogge, R. W., Pott, J.-U., Rafter, S. E., Rix, H.-W., Saylor, D. A., Schimoia, J. S., Schnülle, K., Sergeev, S. G., Siegel, M. H., Spencer, M., Sung, H.-I., Teems, K. G., Turner, C. S., Uttley, P., Vestergaard, M., Villforth, C., Weiss, Y., Woo, J.-H., Yan, H., Young, S., Zheng, W., & Zu, Y. 2017. SPACE TELESCOPE AND OPTICAL REVERBERATION MAPPING PROJECT.VI. REVERBERATING DISK MODELS FOR NGC 5548. *The Astrophysical Journal*, **835**.
- Starkey, D. A., Horne, K., & Villforth, C. 2016. Accretion disc time lag distributions: applying CREAM to simulated AGN light curves. *Monthly Notices of the Royal Astronomical Society*, **456**.
- Urry, C. Megan, & Padovani, P. 1995. Unified Schemes for Radio-Loud Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, **107**.
- Watson, D., Denney, K. D., Vestergaard, M., & Davis, T. M. 2011. A New Cosmological Distance Measure Using Active Galactic Nuclei. *The Astrophysical Journal Letters*, **740**.
- Zu, Y., Kochanek, C. S., Kozłowski, S., & Udalski, A. 2013. IS QUASAR OPTICAL VARIABILITY A DAMPED RANDOM WALK? *The Astrophysical Journal*, **765**.

A Figures for Second synthetic model test

B Plots of other objects

Name	Value
$\sigma(\eta)$	0.10
$\sqrt{\frac{\tau}{2}}(\ell)$	13.00 \sqrt{d}
i	1.20
K_0	0.25
T	1356 K
σ_{DT}	0.76
μ_{DT}	3.86
$\sigma_{J,AD}$	0.90
$\mu_{J,AD}$	2.68
$\sigma_{H,AD}$	0.95
$\mu_{H,AD}$	2.75
$\sigma_{K,AD}$	1.10
$\mu_{K,AD}$	2.98
$\sigma_{g,AD}$	1.10
$\mu_{g,AD}$	2.37
$\sigma_{r,AD}$	1.18
$\mu_{r,AD}$	2.45
$\sigma_{i,AD}$	1.12
$\mu_{i,AD}$	2.50
$\sigma_{z,AD}$	1.20
$\mu_{z,AD}$	2.63

Table 5: Parameters used to create the driving light curve of figure 23 and the transfer functions of figure 24. The parameters are rounded off to two decimals. Refer to chapter 1.3 for the definition of the different parameters. Note that the power law index has a positive value, since that was the only way I could replicate the transfer functions from Landt *et al.* (2019).

Band	χ_{red}^2
J	0.983
H	1.256
K	0.802
g	1.204
r	1.430
i	1.566
z	1.124

Table 6: Same description as table 3, except the data used is seen in figure 28

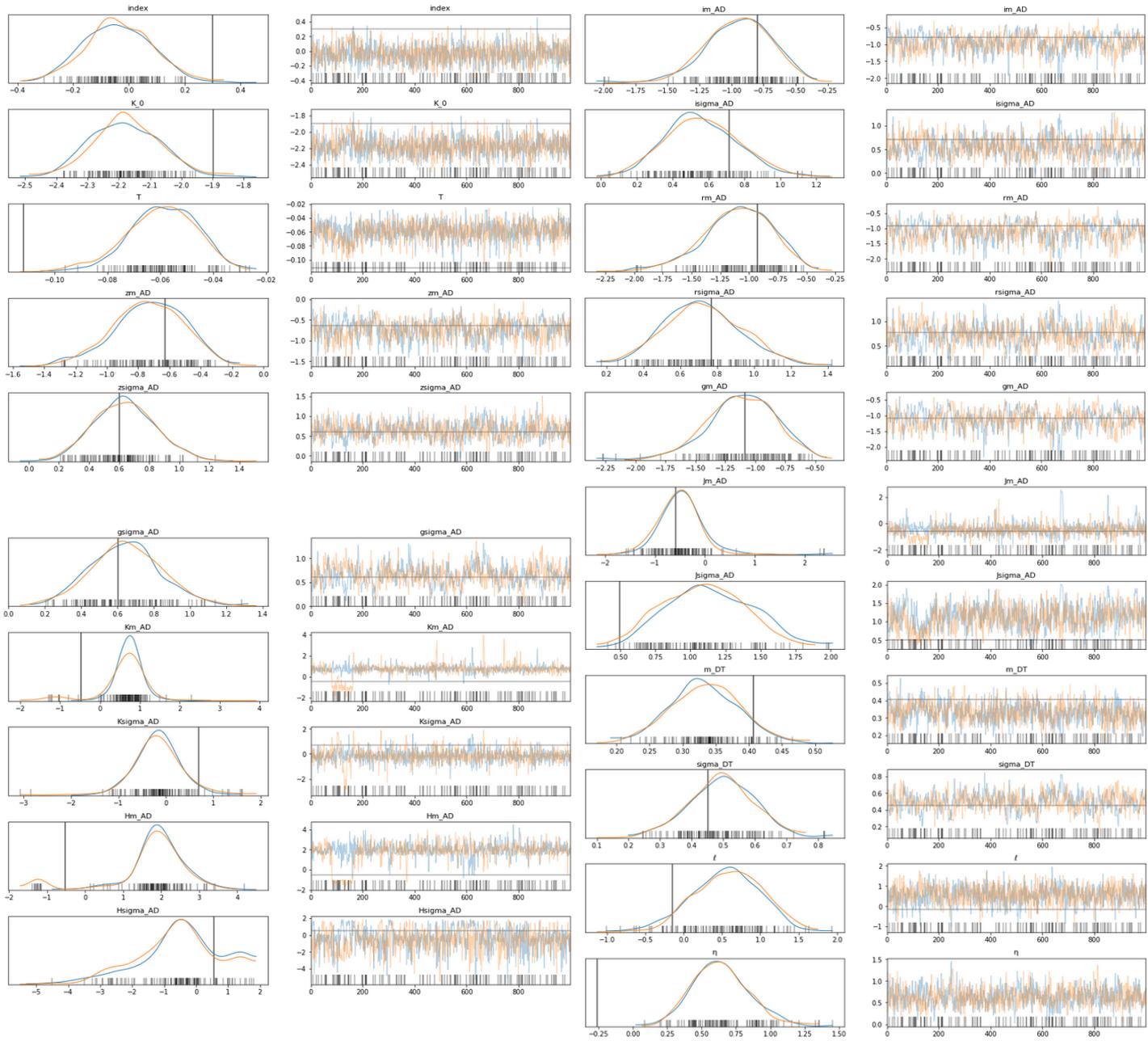


Figure 13: Trace plots for all model parameters and hyperparameters. Both chains are shown in different colors to distinguish them. The horizontal black line in all of the plots shows the actual value for the parameter. The black marks at the bottom of the plots indicates when a divergence occurs during the sampling. Remember that the true parameter values shown are relative to the prior mean.

Comparison between model input and data

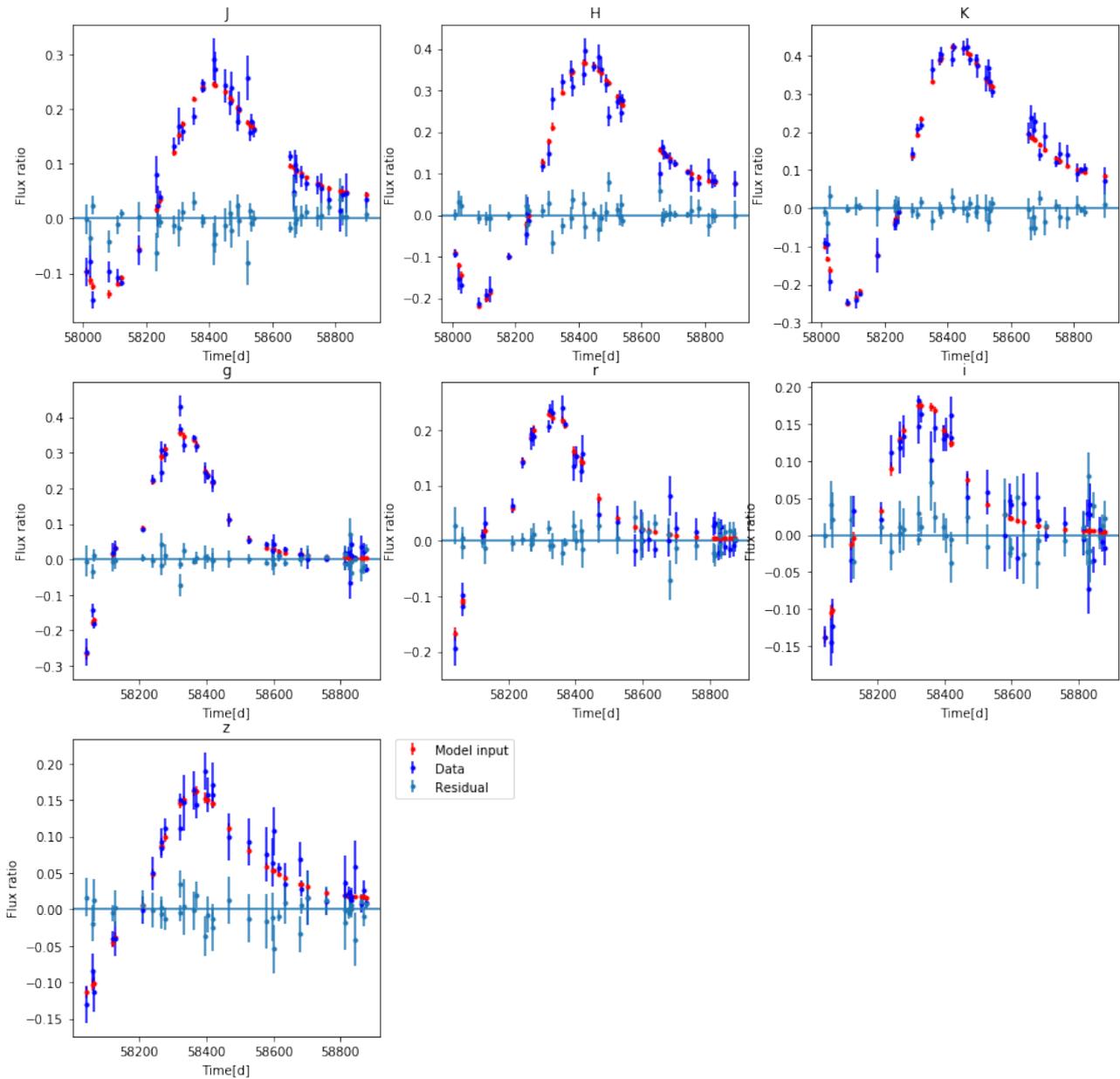


Figure 14: Comparison between the model input and the synthetic data. The model input "data" is compared with synthetic data in the likelihood to update the model. The residual between the two is used to gauge how well the sampling is done. Technically this is the Δ flux ratio plotted, since it is only the variable part of the light curves.

Transfer function comparisons

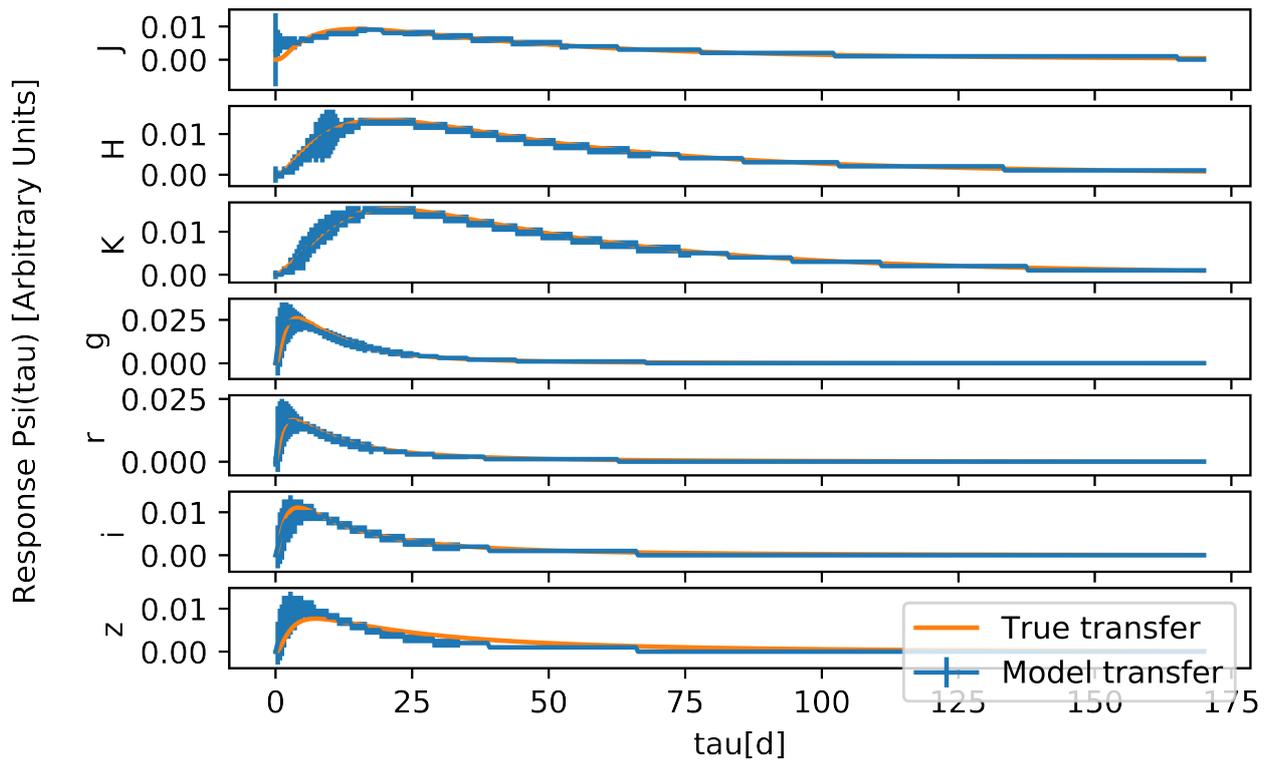


Figure 15: This shows a comparison between the transfer functions determined by the model (in blue) and the true transfer functions from figure 10 (in orange). The area marked by the model transfer functions is the 1σ error. Here τ is the time delay and $\Psi(\tau)$ is the value of the transfer function at a given time delay.

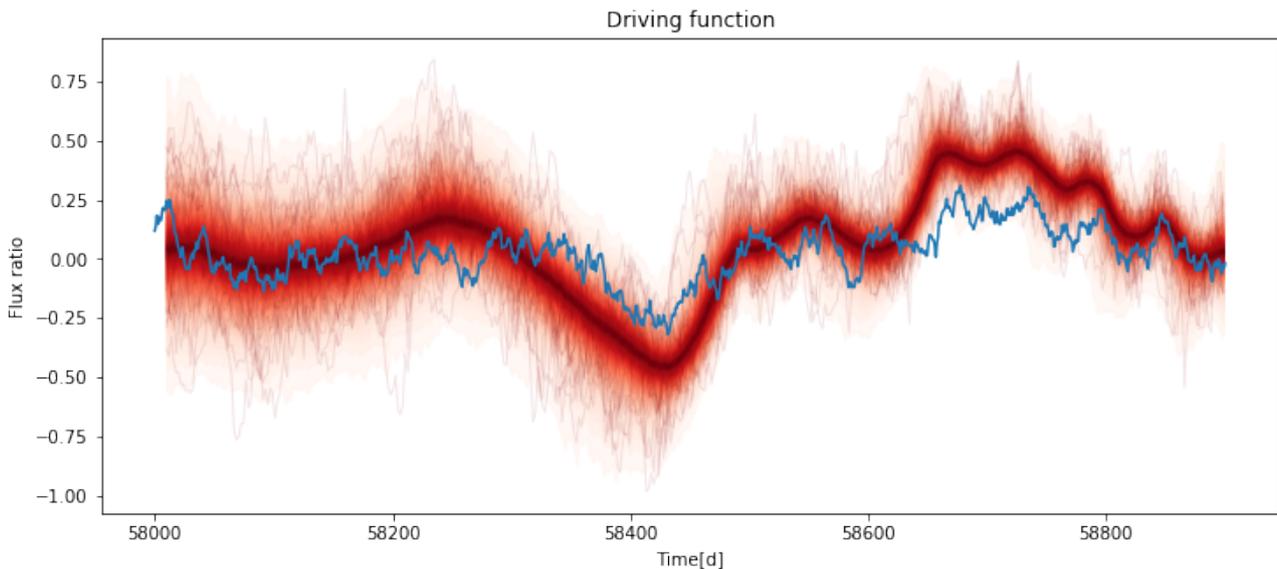


Figure 16: Comparison of the driving function determined by the sampling with the true driving function. The different lines red lines are draws from the posterior GP. The blue line is true driving function. Technically this is the Δ flux ratio plotted, since it is only the variable part of the driving light curve.

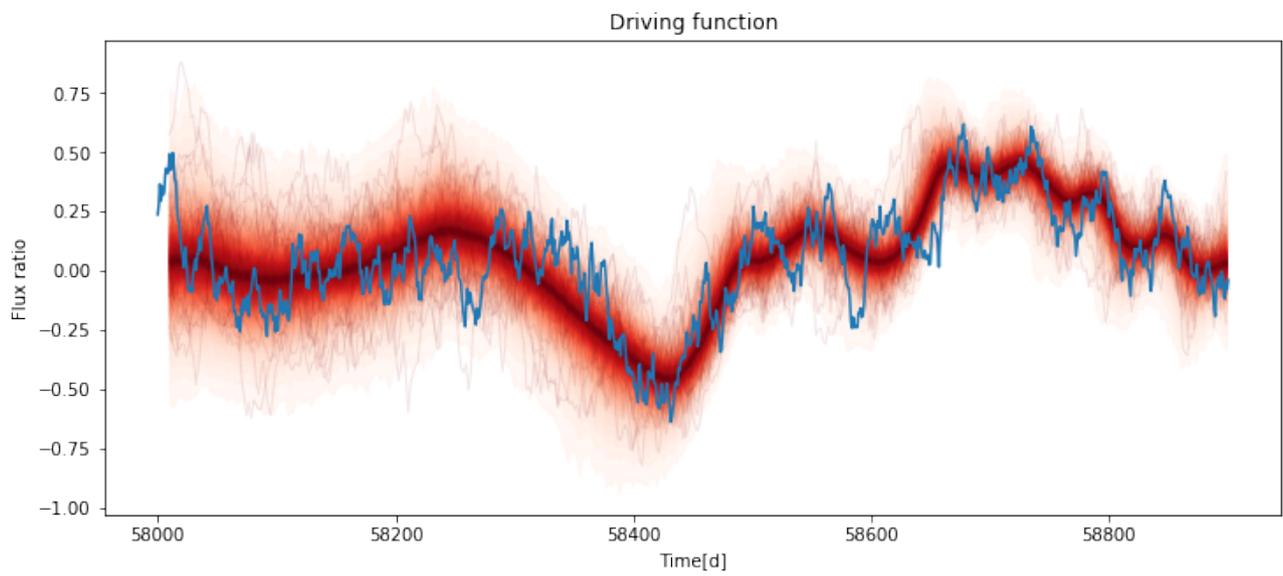


Figure 17: Same as figure 16, but the flux ratio values of the true driving function has been multiplied by 2.

	mean	sd	hpd_3%	hpd_97%	mcse_mean	mcse_sd	ess_mean	ess_sd	ess_bulk	ess_tail	r_hat
index	0.180	0.288	-0.336	0.693	0.020	0.014	204.0	204.0	212.0	279.0	1.01
K_0	-1.371	0.487	-2.223	-0.349	0.033	0.024	216.0	216.0	216.0	491.0	1.01
T	-0.017	0.350	-0.635	0.661	0.011	0.008	1001.0	1001.0	1015.0	1107.0	1.00
zm_AD	4.405	0.521	3.675	5.323	0.029	0.021	316.0	297.0	392.0	385.0	1.00
zsigma_AD	-1.947	1.085	-4.179	-0.095	0.047	0.033	540.0	540.0	552.0	645.0	1.00
lm_AD	4.443	0.508	3.689	5.348	0.024	0.017	468.0	460.0	509.0	527.0	1.00
lsigma_AD	-2.091	0.973	-3.886	-0.372	0.046	0.034	442.0	409.0	466.0	631.0	1.00
rm_AD	3.537	0.149	3.293	3.842	0.012	0.009	147.0	144.0	179.0	162.0	1.03
rsigma_AD	0.456	0.098	0.277	0.649	0.008	0.006	158.0	143.0	166.0	140.0	1.03
gm_AD	4.405	0.230	4.031	4.890	0.019	0.014	140.0	137.0	161.0	141.0	1.03
gsigma_AD	0.742	0.090	0.564	0.909	0.007	0.005	173.0	164.0	182.0	238.0	1.02
Km_AD	-2.589	0.616	-3.707	-1.809	0.379	0.305	3.0	3.0	3.0	65.0	1.83
Ksigma_AD	-2.621	0.962	-4.568	-0.877	0.170	0.121	32.0	32.0	25.0	328.0	1.09
Hm_AD	3.754	0.321	3.207	4.364	0.020	0.014	259.0	247.0	287.0	438.0	1.00
Hsigma_AD	0.553	0.226	0.130	0.960	0.008	0.006	751.0	707.0	749.0	721.0	1.00
Jm_AD	3.238	0.237	2.760	3.625	0.014	0.010	297.0	287.0	320.0	464.0	1.01
Jsigma_AD	0.331	0.195	-0.031	0.704	0.011	0.008	289.0	289.0	289.0	664.0	1.01
m_DT	2.452	0.462	1.723	3.250	0.023	0.016	402.0	399.0	430.0	716.0	1.00
sigma_DT	-2.157	1.034	-4.057	-0.278	0.043	0.031	567.0	562.0	566.0	784.0	1.00
η	-0.053	0.398	-0.791	0.751	0.021	0.015	368.0	368.0	371.0	470.0	1.00
t	6.871	6.856	-5.830	19.338	0.298	0.211	528.0	528.0	519.0	662.0	1.00

Figure 18: same description as figure 12, but for the data for the object AKN120.

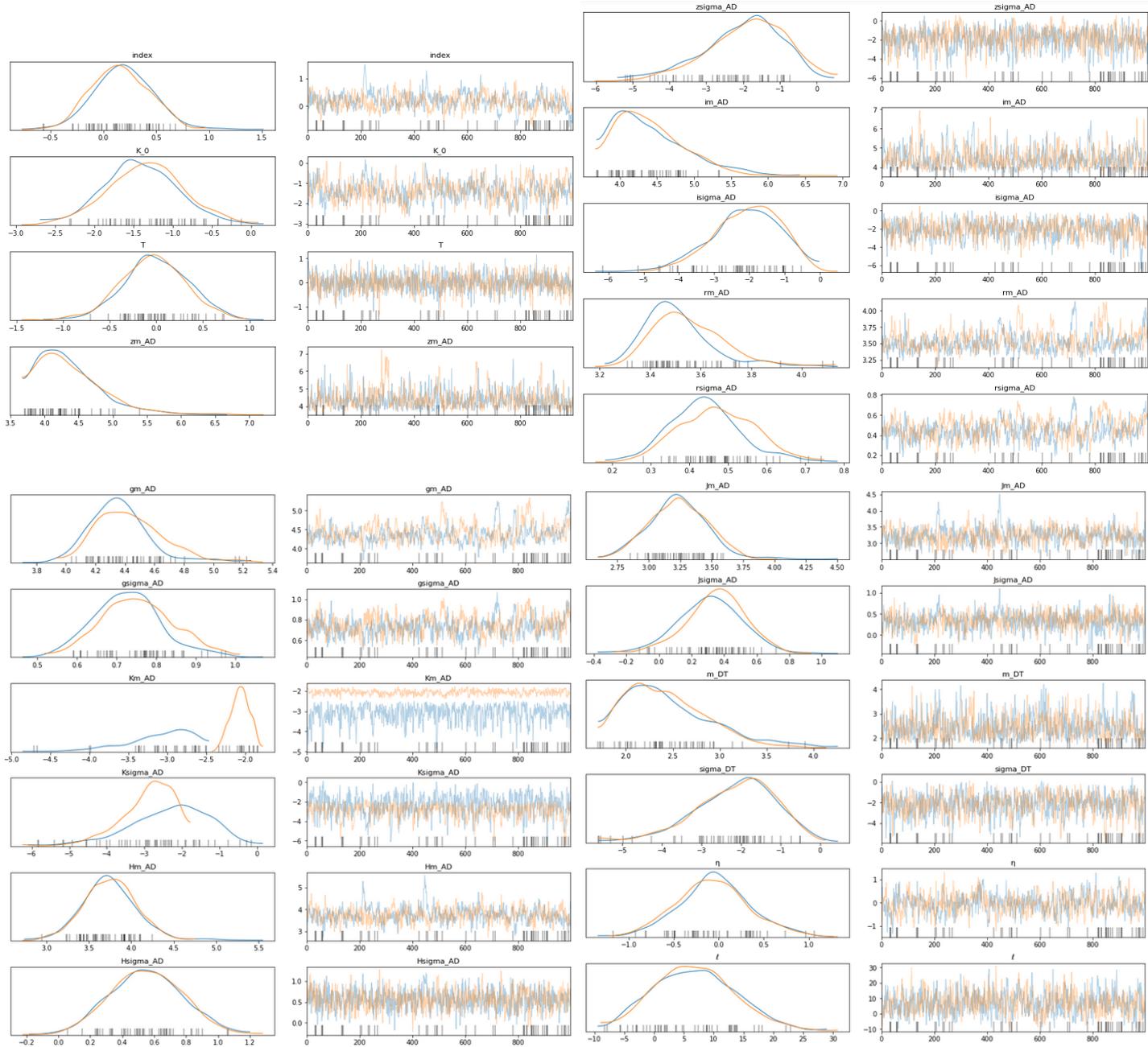


Figure 19: Same description as figure 13, but for the object AKN120.

Comparison between model input and data

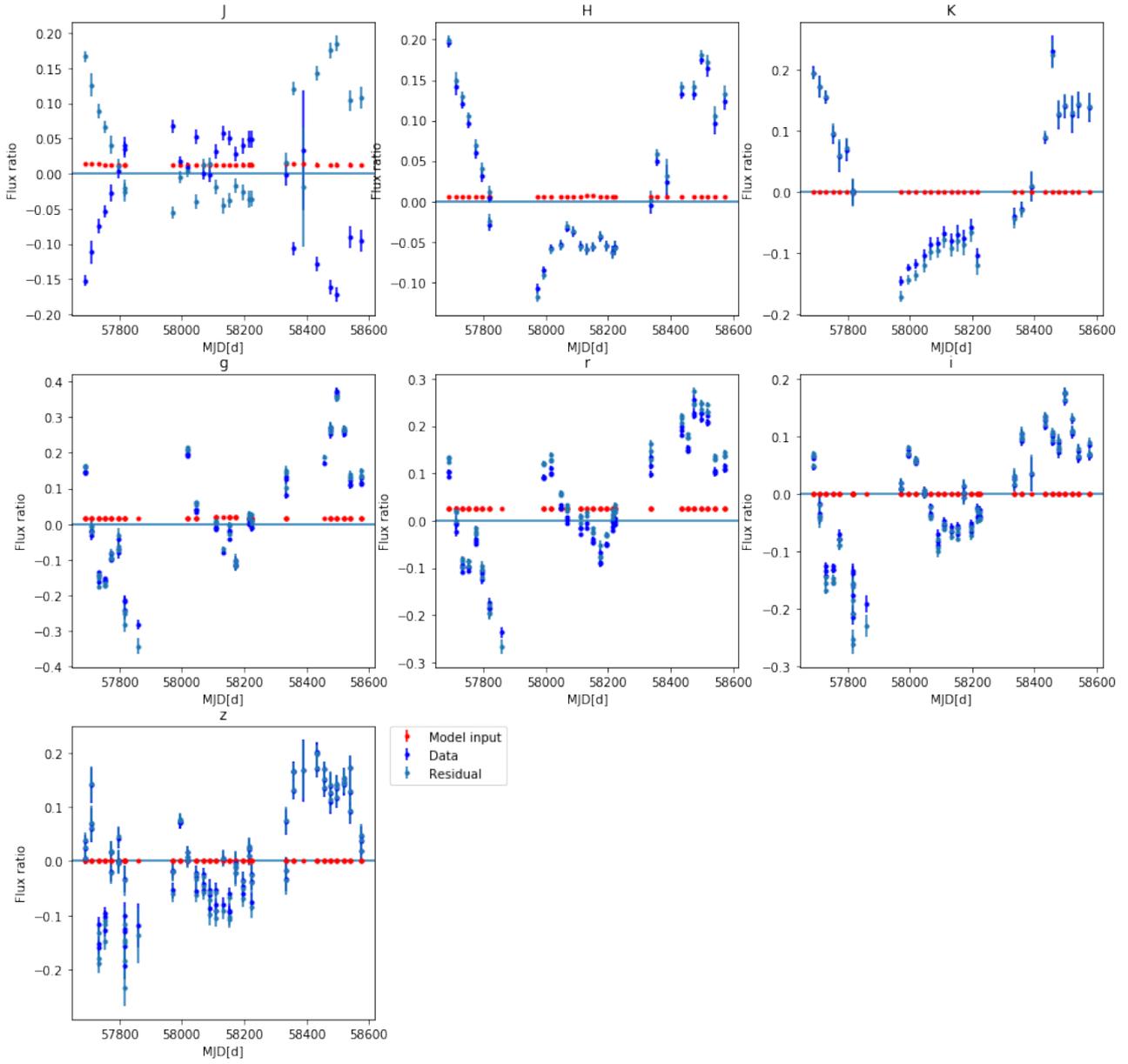


Figure 20: Same description as figure 14, but for the object AKN120

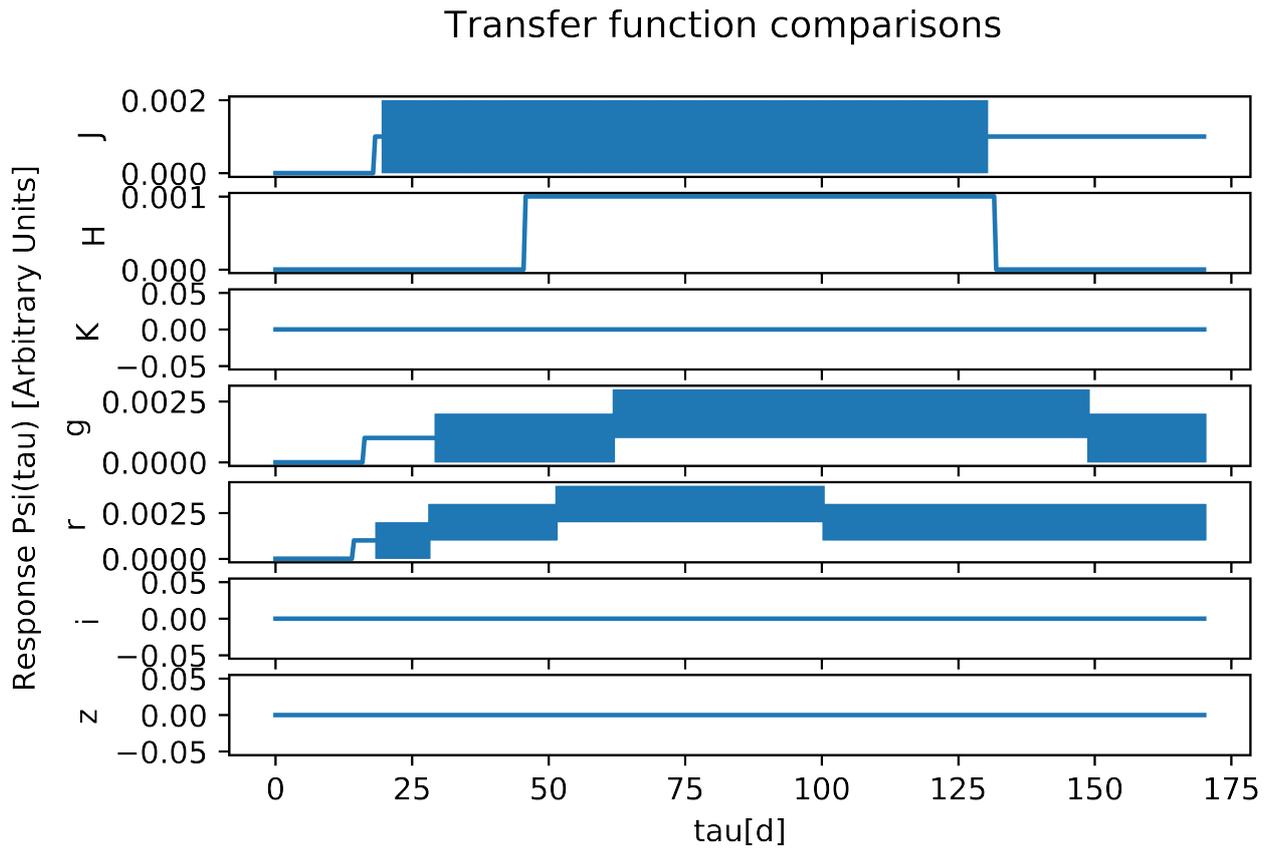


Figure 21: This shows a the transfer functions for AKN120 determined by the model. The area marked by the model transfer functions is the 1σ error. Here tau is the time delay and Psi(tau) is the value of the transfer function at a given time delay.

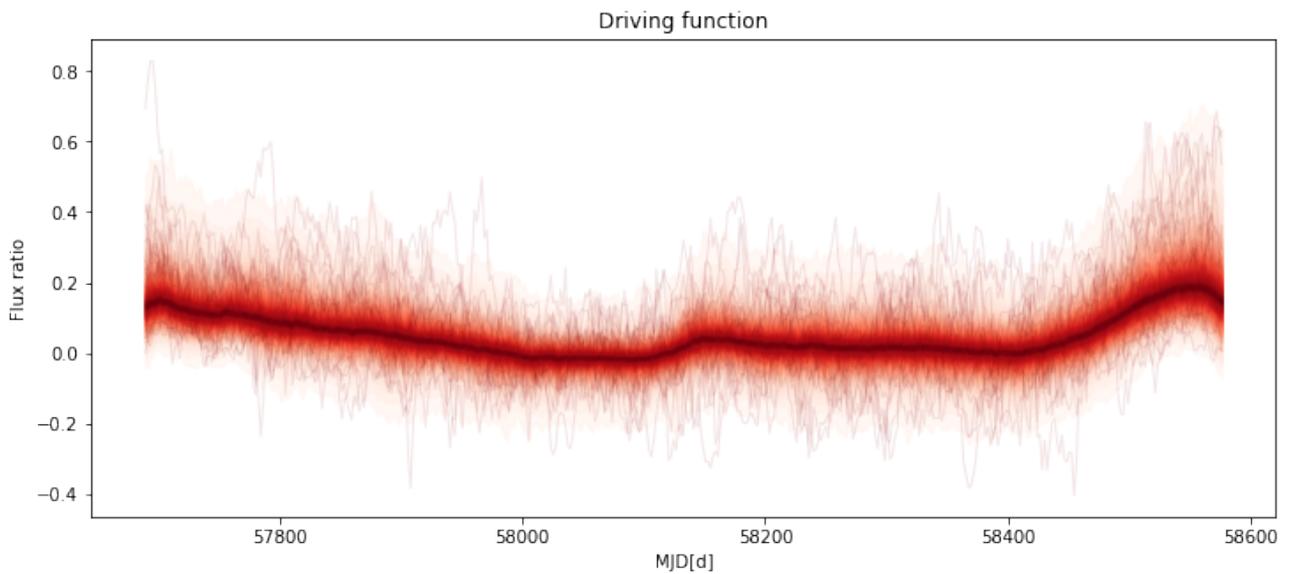


Figure 22: The driving function for AKN120 determined by the sampling. The different lines red lines are draws from the posterior GP. Technically this is the Δ flux ratio plotted, since it is only the variable part of the driving light curve.

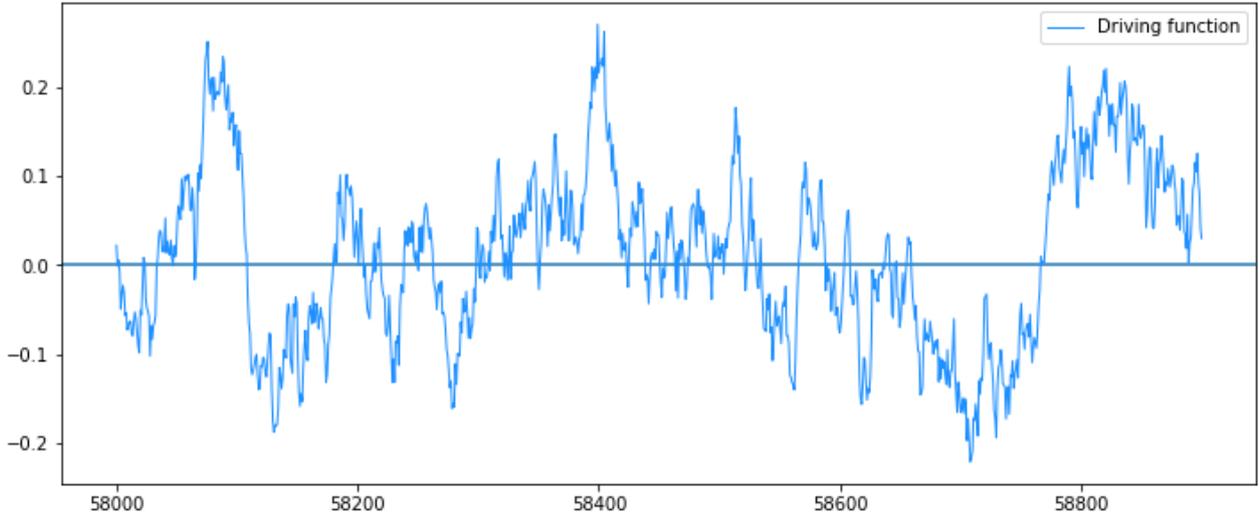


Figure 23: Plot of a 900 day synthetic driving light curve with a horizontal zero line created with the data generation script. Here $\ell = 13d^{1/2}$ (meaning $\ell_{true} = 338d$) and $\eta = 0.1$. This is from a different test of the model using synthetic data.

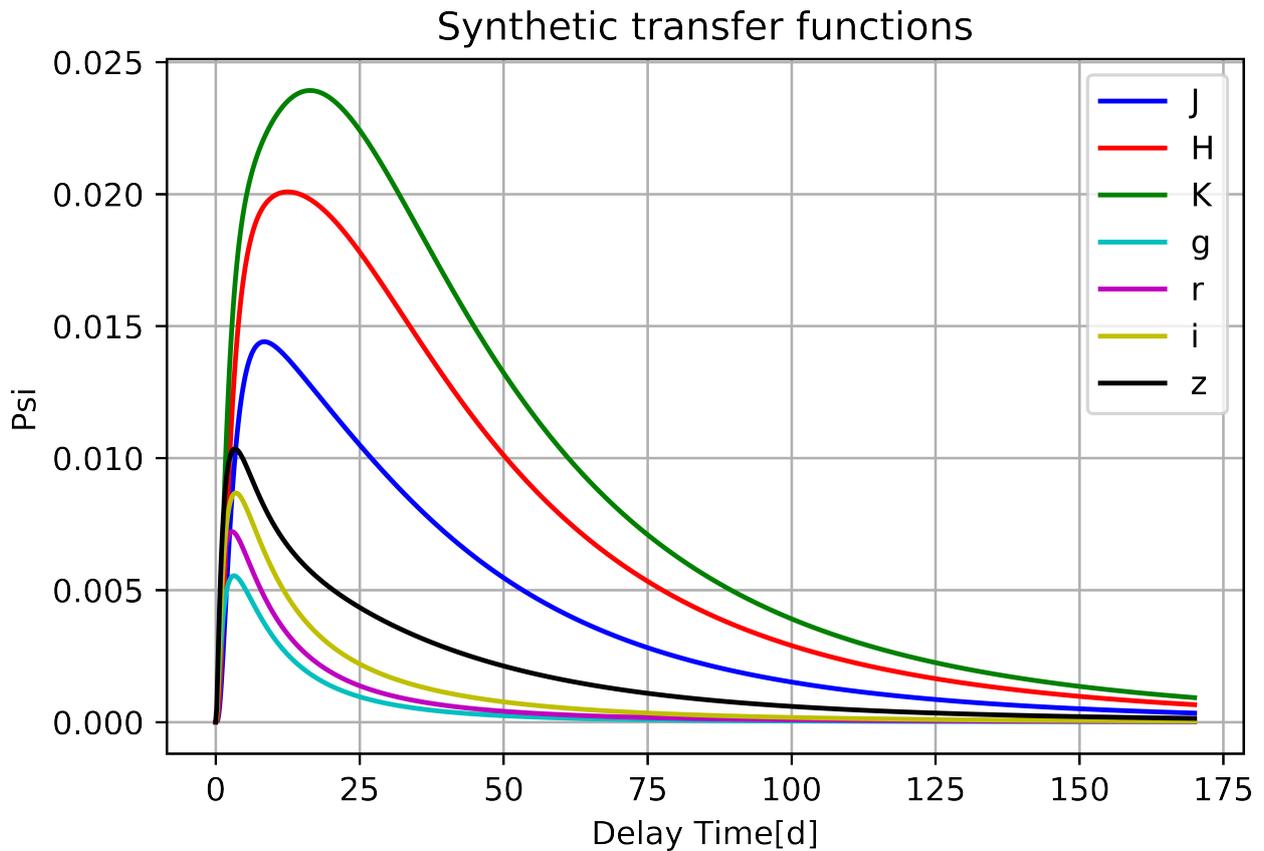


Figure 24: Plot of the transfer functions used to create synthetic data for the J,H,K,g,r,i,z bands. The parameters used to create the synthetic transfer functions are specified in table 5. The transfer functions are made to roughly compare to the ones found in Landt *et al.* (2019).

Synthetic light curves with Gaussian noise

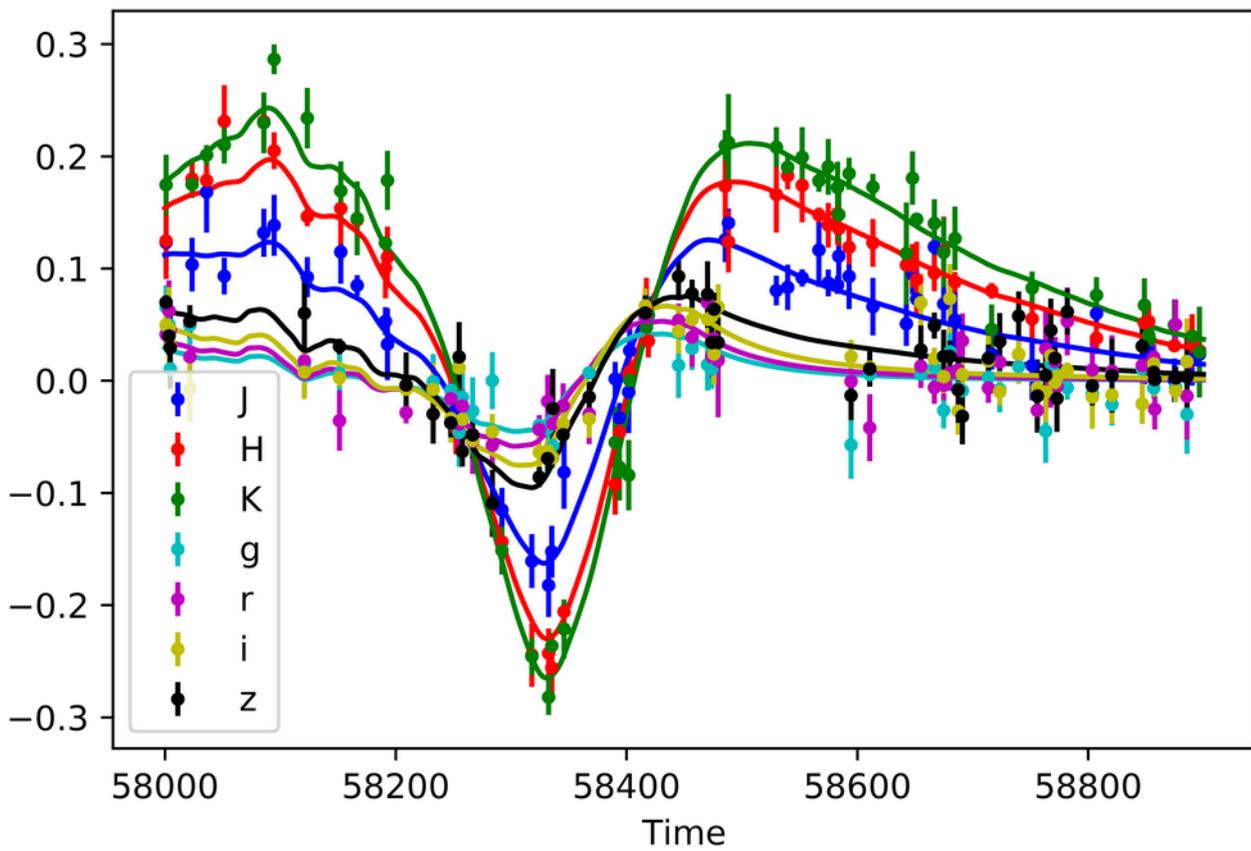


Figure 25: Same description as figure 11. This example uses 49 points instead of 40.

	mean	sd	hpd_3%	hpd_97%	mcse_mean	mcse_sd	ess_mean	ess_sd	ess_bulk	ess_tail	r_hat
index	-0.213	0.409	-0.976	0.485	0.014	0.010	913.0	913.0	763.0	651.0	1.01
K_0	-1.232	0.609	-2.261	-0.111	0.021	0.015	832.0	832.0	908.0	951.0	1.00
T	-0.146	0.060	-0.260	-0.040	0.002	0.001	929.0	929.0	869.0	1296.0	1.00
zm_AD	-0.087	0.344	-0.697	0.581	0.011	0.008	998.0	998.0	1032.0	1289.0	1.00
zsigma_AD	0.401	0.289	-0.089	0.974	0.009	0.006	1119.0	1119.0	1147.0	2023.0	1.00
im_AD	-0.533	0.254	-0.969	-0.008	0.007	0.005	1301.0	1301.0	1361.0	1572.0	1.01
isigma_AD	0.137	0.311	-0.393	0.758	0.009	0.007	1073.0	1073.0	1108.0	1750.0	1.00
rm_AD	-0.526	0.267	-0.975	0.021	0.006	0.004	2137.0	2137.0	2304.0	1606.0	1.00
rsigma_AD	0.148	0.343	-0.578	0.715	0.017	0.012	421.0	421.0	341.0	94.0	1.00
gm_AD	-0.500	0.327	-1.099	0.104	0.008	0.005	1816.0	1788.0	1637.0	2050.0	1.00
gsigma_AD	0.085	0.357	-0.507	0.750	0.012	0.009	859.0	859.0	645.0	267.0	1.00
Km_AD	1.023	0.430	0.219	1.873	0.009	0.006	2417.0	2417.0	2389.0	2314.0	1.00
Ksigma_AD	0.107	0.358	-0.566	0.771	0.011	0.008	1100.0	1100.0	1071.0	1361.0	1.00
Hm_AD	0.594	0.406	-0.177	1.381	0.012	0.008	1195.0	1195.0	1198.0	1795.0	1.00
Hsigma_AD	0.007	0.227	-0.398	0.472	0.006	0.004	1419.0	1419.0	1434.0	1320.0	1.00
Jm_AD	0.192	0.380	-0.496	0.878	0.013	0.009	894.0	894.0	946.0	1352.0	1.00
Jsigma_AD	0.051	0.218	-0.310	0.474	0.006	0.004	1298.0	1223.0	1433.0	1423.0	1.00
m_DT	0.569	0.239	0.137	1.037	0.007	0.005	1081.0	1081.0	1108.0	1297.0	1.00
sigma_DT	-0.162	0.186	-0.504	0.183	0.007	0.005	665.0	576.0	633.0	450.0	1.00
ℓ	3.857	4.076	-3.678	11.542	0.070	0.062	3386.0	2176.0	3392.0	1913.0	1.00
η	1.153	0.391	0.401	1.786	0.012	0.009	1006.0	1006.0	1204.0	1037.0	1.00

Figure 26: Same description as figure 12.

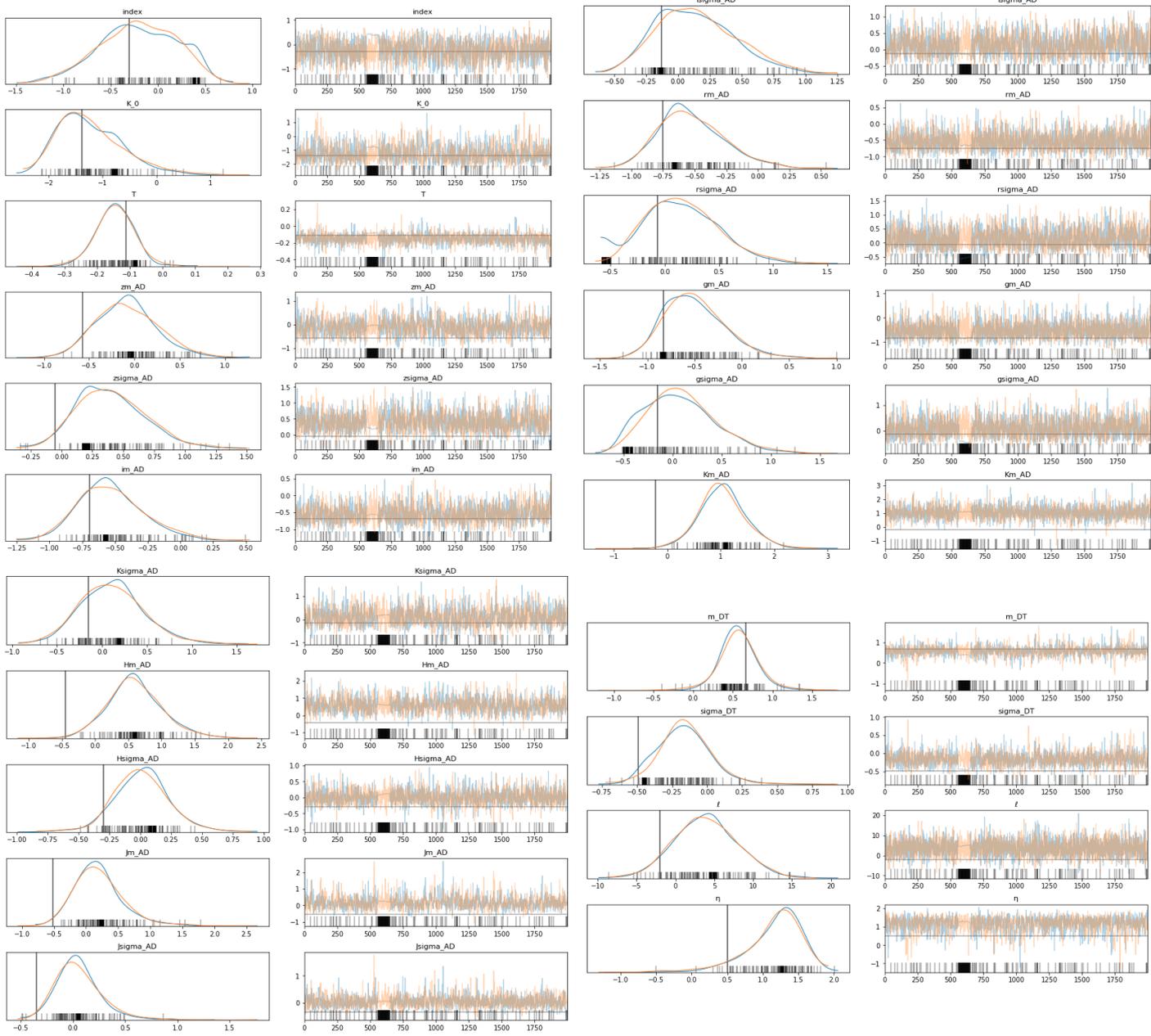


Figure 27: same description as figure 13

Comparison between model input and data

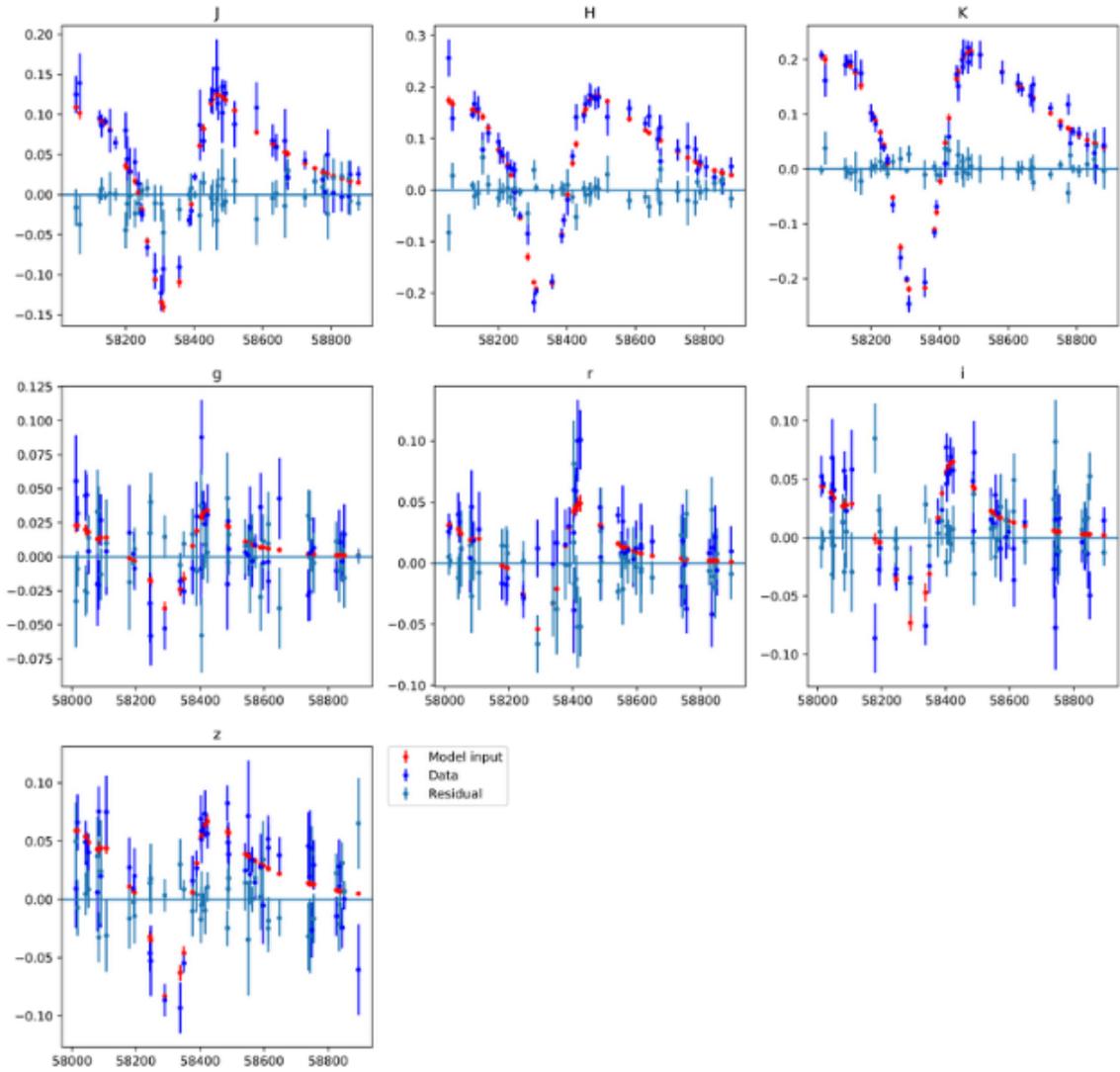


Figure 28: Same disription as figure 14.

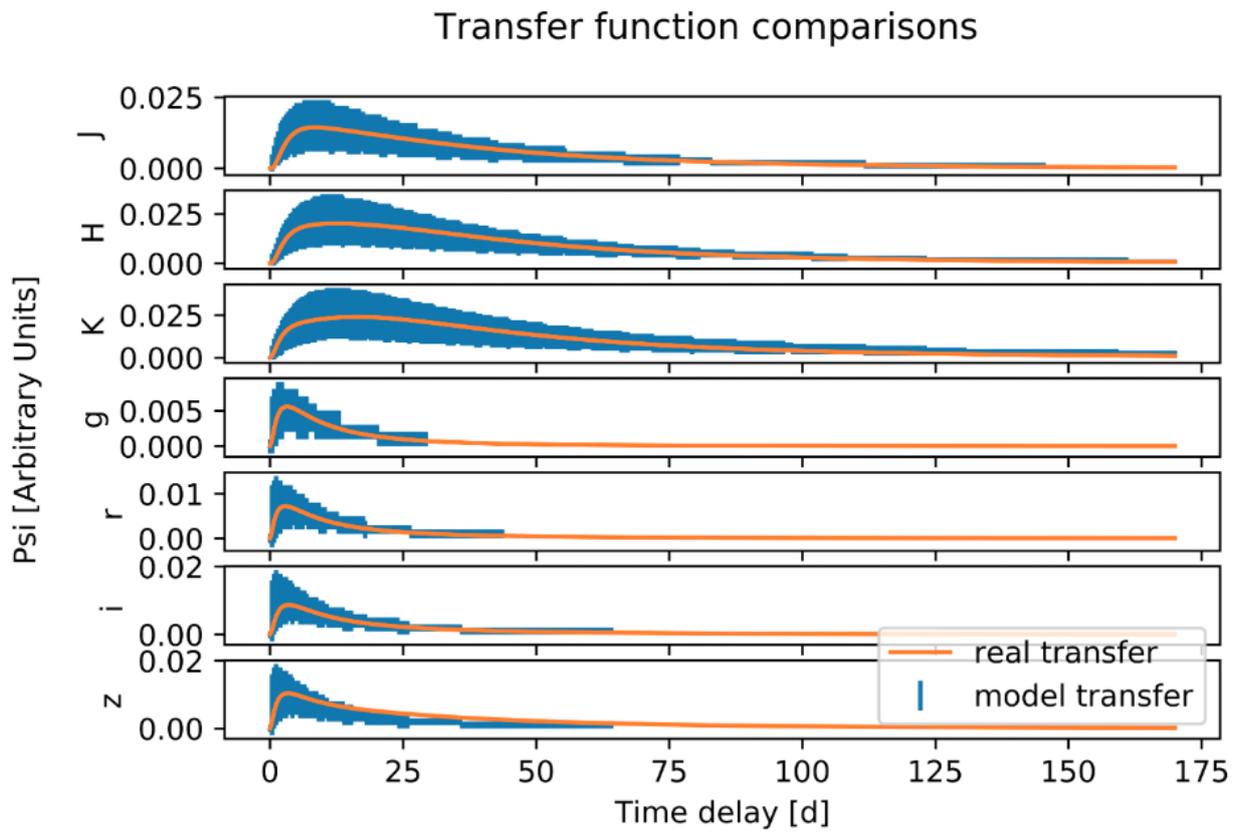


Figure 29: Same description as figure 15.

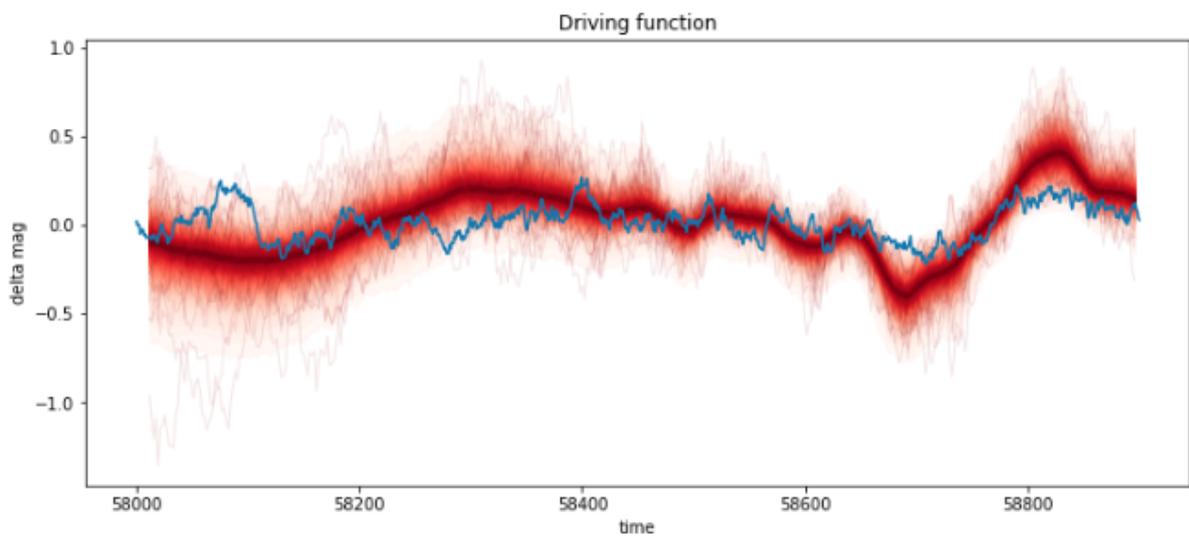


Figure 30: Same description as figure 16.

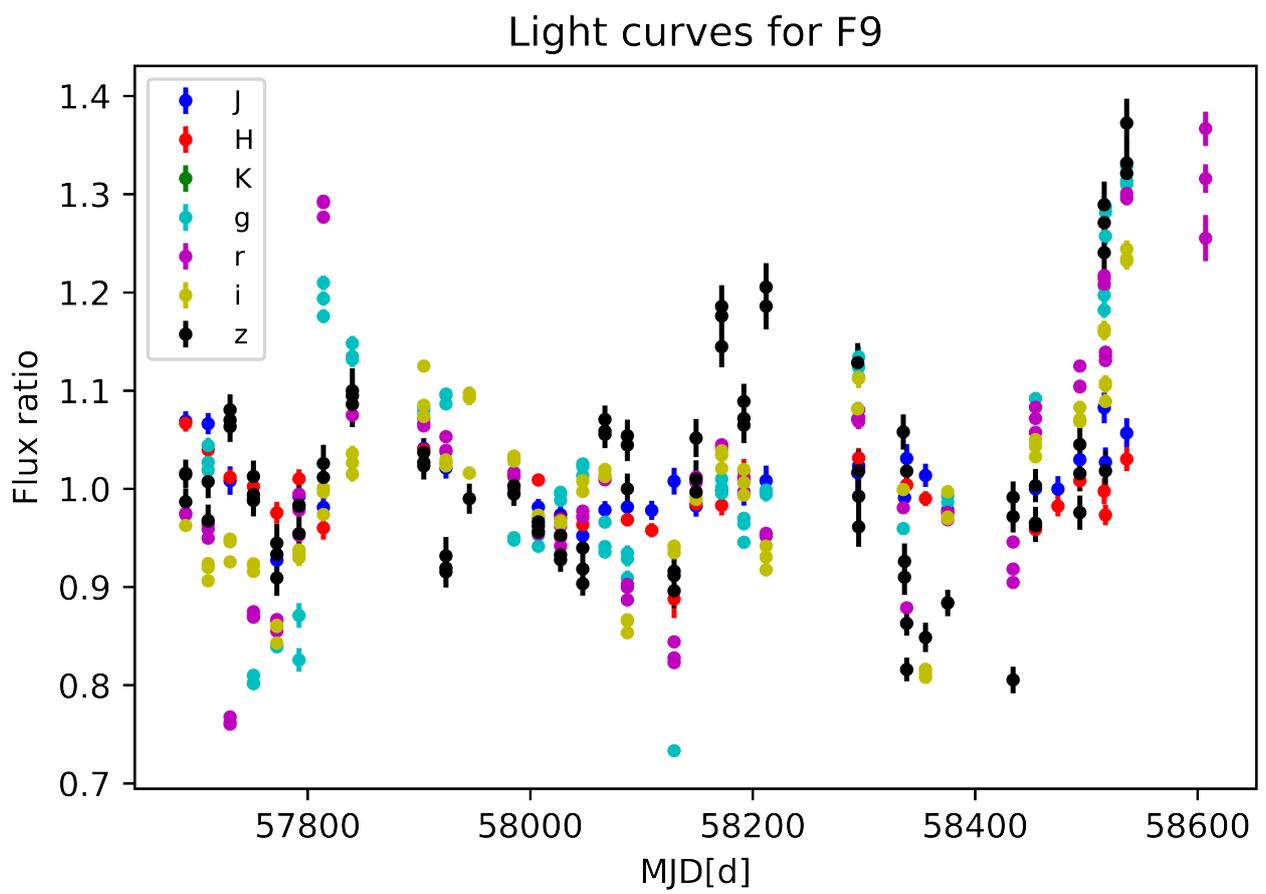


Figure 31: Same description as figure 7

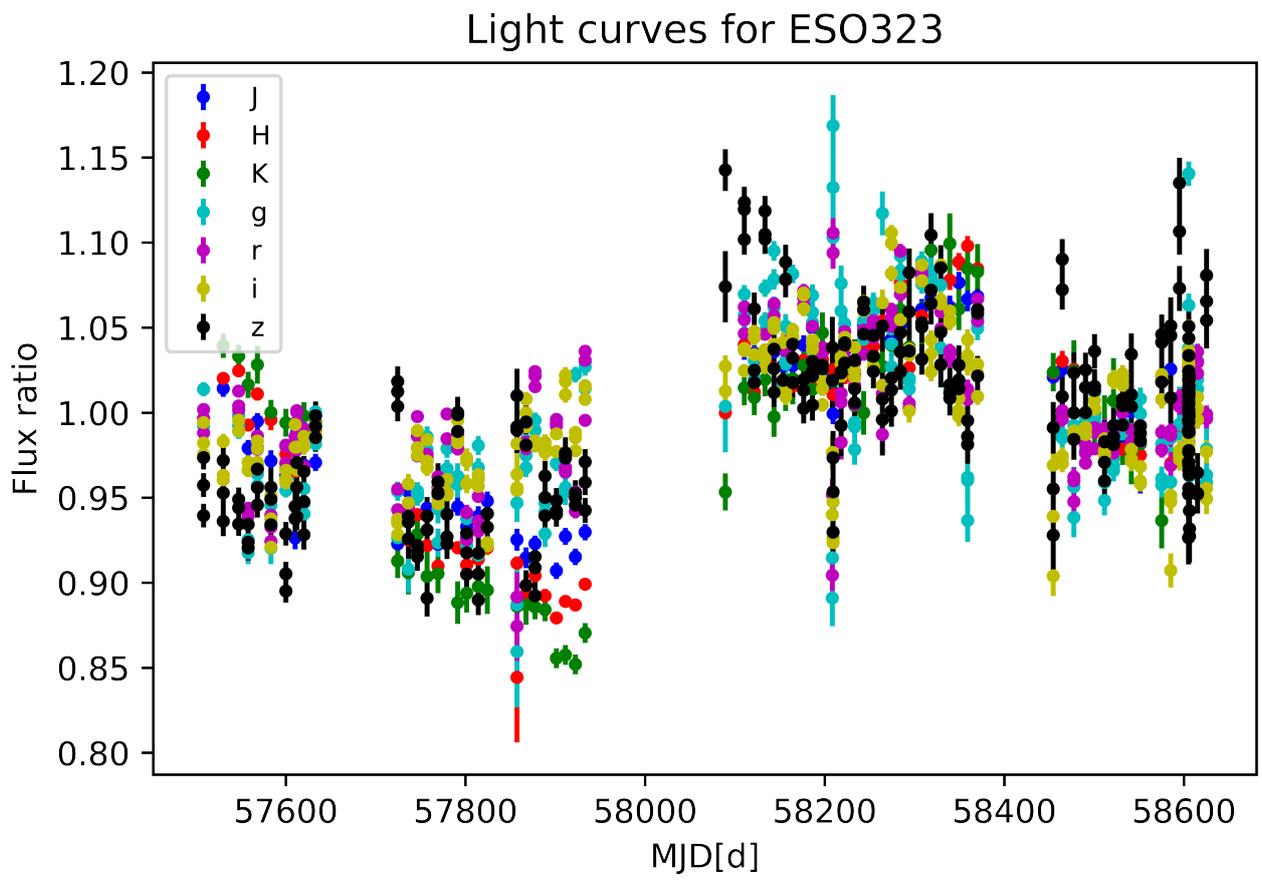


Figure 32: Same description as figure 7

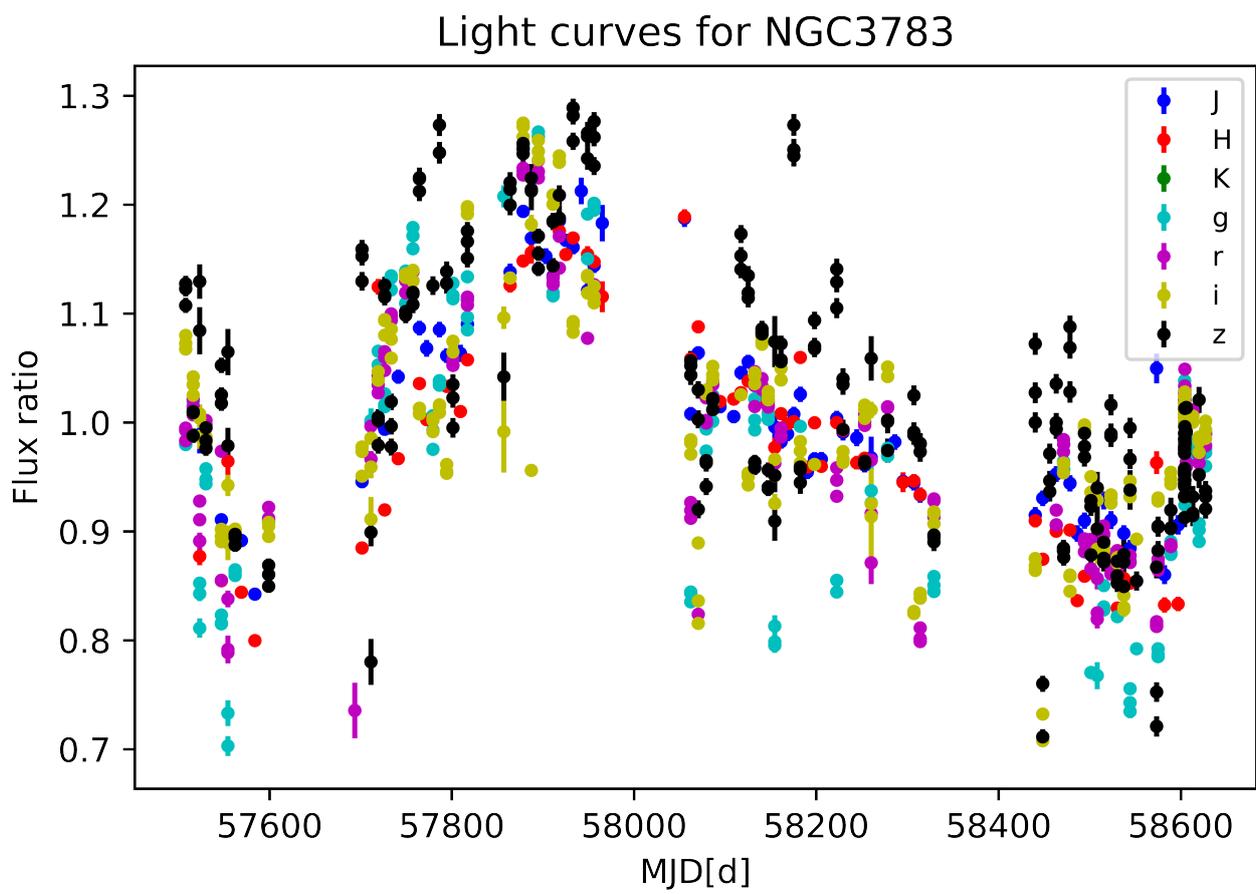


Figure 33: Same description as figure 7

Light curves for NGC7213

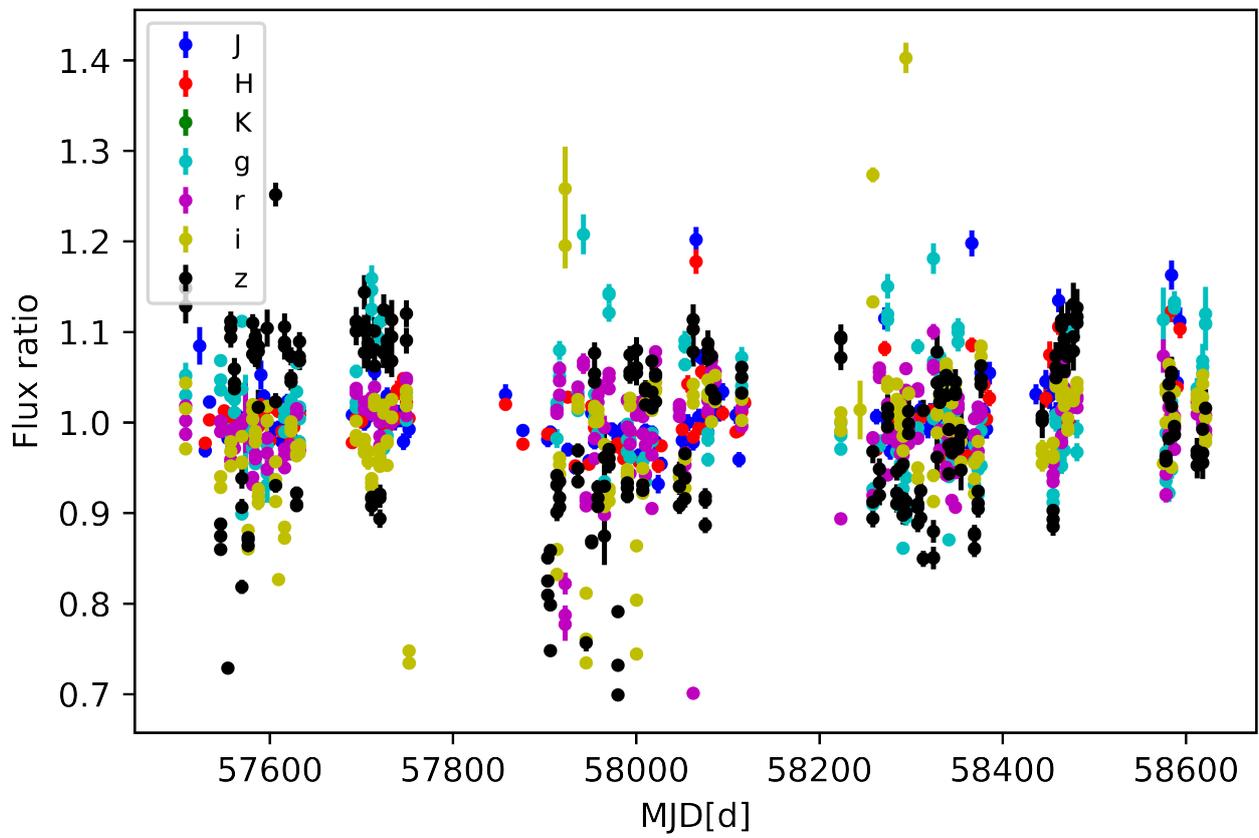


Figure 34: Same description as figure 7