MASTER IN ASTROPHYSICS AND SPACE SCIENCE

Erasmus Mundus Master
in Astrophysics and Space Science

Master's thesis

Application of Conditional Neural Flows to Multiband Reverberation Mapping of Active Galactic Nuclei

Supervisors

Prof. Anđelka Kovačević
Astrophysics Department, University of Belgrade

Prof. Médéric Boquien
Laboratoire Lagrange, Université Côte d'Azur

Prof. Dragana Ilić
Astrophysics Department, University of Belgrade

Author

Ana Cecilia Luis Ramirez

Academic year 2025/2026

TOR VERGATA UNIVERSITÀ DEGLI STUDI DI ROMA
UNIVERSITY OF BELGRADE
University of Bremen
UNIVERSITÉ CÔTE D'AZUR

Cofunded by the European Union

This Master thesis is submitted in partial fulfillment of the requirements for the degree of Master in Astrophysics and Space Science as part of a multiple degree awarded in the framework of the Erasmus Mundus Joint Master in Astrophysics and Space Science — MASS jointly delivered by a Consortium of four Universities: Tor Vergata University of Rome, University of Belgrade, University of Bremen, and Université Cote d'Azur, regulated by the MASS Consortium Agreement and funded by the EU under the call ERASMUS-EDU-2021-PEX-EMJM-MOB.

УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

Ana Cecilia Luis Ramirez

Primena uslovnih normalizujućih neuronskih tokova na višefrekventno reverberaciono mapiranje aktivnih galaktičkih jezgara

Master Thesis

Београд, 2026.

Abstract

Active galactic nuclei (AGN) exhibit stochastic variability over a wide range of timescales across the entire electromagnetic spectrum. These brightness variations carry information about the physical structure of the accretion flow and the coupling between the central ionizing source and the surrounding disk, making time-domain studies of AGN light curves an important tool for studying black hole growth and accretion physics on otherwise unresolved scales. Wide-field surveys such as the Zwicky Transient Facility (ZTF), and the forthcoming Legacy Survey of Space and Time (LSST) at the Vera C. Rubin Observatory, now deliver optical light curves for large numbers of AGN.

However, these light curves are irregularly sampled and can contain substantial gaps, which makes reconstructing the underlying signal an underdetermined problem, better addressed by a probabilistic method that returns a distribution of plausible light curves rather than a single estimate. This thesis investigates the use of neural stochastic flows (Kiyohara et al. 2025), a recently proposed class of generative models that learn the transition density of a stochastic process directly, for the reconstruction of these optical light curves modelled as a damped random walk. The method is trained and evaluated on simulated light curves with physically motivated parameters, and validated against the analytic transition of the damped random walk, which the learned model should reproduce, and against a Gaussian process with the same kernel, the Bayes-optimal reconstruction baseline.

On low-noise simulations, at the 0.01 mag photometric precision expected of LSST, the method recovers the analytic transition and reconstructs light curves to within about three per cent of the Bayes-optimal Gaussian process in accuracy, with a calibration that is close to it though not exact, reaching a coverage of 0.61 where the Gaussian process reaches 0.70. When the sampling and the heteroscedastic noise are instead matched to those of ZTF, the predictive intervals become systematically overconfident, with a coverage near 0.3 against a nominal 0.68, and the accuracy falls to roughly twice the error of the Gaussian process. This is traced to the model treating each observed magnitude as if it were noise-free, with no channel through which the per-epoch uncertainties can enter it, and anchoring too tightly to noisy observations. Applied to real ZTF light curves through a masking experiment, in which withheld observations are predicted and the measurement error of each is folded into the comparison, the reconstruction is close to calibrated on the less noisy bands and retains a deficit on the noisiest. A Gaussian process fitted to each curve, which uses the per-epoch uncertainties directly in both its fit and its predictions, is the more consistent baseline, and both improve on linear interpolation. These results identify the handling of measurement noise as the principal limitation of the approach, and motivate a latent formulation, in which the flow describes an unobserved signal related to the observations through an explicit noise model, as the natural next step.

Contents

List of Figures

List of Tables

1 Introduction

The first quasar discovered, 3C 273 (Edge et al. 1959), had been documented in photographic plates at least as early as 1887 (Slavcheva-Mihova et al. 2013). This emblematic radio-loud object showed variable emission in the optical and radio bands on timescales from days to years (Smith and Hoffleit 1963; Dent 1965), indicating that the size of the emitting region spanned only a few light-days in size. Its compactness, coupled with the finding of its distant redshift (Schmidt 1963), was very difficult to explain by dynamics that operate in normal galaxies, and variability thus became one of the key arguments that the enormous luminosities of quasars had to originate from an exceptionally compact and energetic central engine (Salpeter 1964; Zeldovich 1964).

Now we know that variability in all wavelengths is an ubiquitous property of active galactic nuclei (AGN), the broader class of which quasars are the most luminous members. Its study has been essential in understanding the physics of their nuclear regions, which, in general, cannot be resolved by current telescopes. This had been strictly true only until the observations of Sgr A* and M87 (The Event Horizon Telescope Collaboration et al. 2019) and, more recently, with the GRAVITY interferometer on the VLT resolving the outer BLR in 3C 273 (Gravity Collaboration et al. 2018); still, these surveys are limited to massive sources in the very nearby universe. For the vast majority of AGN, variability studies therefore rely on the *light curve*, a time series recording how the brightness (or flux) of the AGN, measured within a given wavelength band, varies over a monitoring period. The time scales, spectral changes, and the correlations and delays between variations extracted from these light curves provide crucial information on the nature and location of these components, described in Chapter 2, and on their interdependencies (Ulrich et al. 1997).

While variability-based methods are, generally speaking, more effective than other methods (e.g., SED-based) at providing a clearer signature of the nuclear activity against host galaxy emission (Yu et al. 2025), deriving the features and their uncertainties from an irregularly sampled light curve is very challenging; thereon, high quality light curves are indispensable to enable its robust characterization (Kelly et al. 2014).

The challenge is now pressing because of the scale of modern time-domain surveys. Wide-field surveys such as the Zwicky Transient Facility (Bellm et al. 2019; § 3.3.1), and the forthcoming Legacy Survey of Space and Time (Ivezić et al. 2019; § 3.3.2), monitor large areas of the sky repeatedly and deliver optical light curves for enormous numbers of AGN at once. These light curves are irregularly sampled and interrupted by gaps from the day-night cycle, weather, and the seasonal visibility of each field. Reconstructing the signal across these gaps is an underdetermined problem, since many different light curves can fit the same sparse observations; the problem therefore calls for a probabilistic method, one that returns a distribution of plausible light curves rather than a single estimate.

This work investigates the use of *neural stochastic flows* (Kiyohara et al. 2025), a recently proposed class of generative models that learn the transition density of a stochastic process directly, as a flexible method for reconstructing AGN optical light curves. The aim is not to show that such a method surpasses existing approaches, but to establish whether it can be applied to this problem, to validate it against the analytic properties of the damped random walk commonly used to model the variability, and to characterize its behaviour as the simulated data are made progressively more realistic and it is finally applied to real light curves from the Zwicky Transient Facility.

This thesis is organized as follows. Chapter 2 reviews the structure of active galactic nuclei and the emission from each of their components. Chapter 3 describes the physical origin of AGN variability and the tools used to characterize it, introducing the damped random walk and the time-domain surveys relevant to this work. Chapter 4 sets out the simulation framework, deriving the transition probability of the damped random walk and defining the survey regimes reproduced in the simulations. Chapter 5 presents the reconstruction method, developing neural stochastic flows and the bridge construction used to fill the gaps. Chapter 6 reports the results on simulated and real ZTF light curves, and Chapter 7 summarizes the findings and outlines future work.

2 Active Galactic Nuclei

Compared to ordinary galaxies, some have extraordinarily bright central regions, which can sometimes even outshine the host-galaxy emission by many orders of magnitude; these galaxies are called *active galaxies* and their nuclei are referred to as *active galactic nuclei*. The physical processes necessary to create such high luminosities ($L = 10^{43-48}$erg s^{-1}; e.g., Lusso et al. 2012) within such a small volume were a subject of debate for years, but now are widely accepted to be connected to accretion onto the galaxies' central black holes (BHs; Salpeter 1964; Zeldovich 1964). These accreting objects have masses $M_{BH} \sim 10^{6-10} M_{\odot}$ (e.g., Peterson and Horne 2004) and an Eddington ratio exceeding λ_{Edd}[1] $\approx 10^{-5}$. This threshold, chosen somewhat arbitrarily, excludes galactic nuclei with no significant accretion activity, such as that of the Milky Way, while still covering their full range of observed luminosities (Netzer 2015).

The "unified model" (Antonucci 1993; Urry and Padovani 1995) discussed in Section 2.1, posits the black hole at the center of the AGN, surrounded by several distinct components, including an accretion disk, gas and dust clouds, and, in some cases, a relativistic jet (Figure 2.4). Each of these regions contributes emission into different wavelength regimes, resulting in these objects having detectable emission across the whole electromagnetic spectrum, as illustrated by the spectral energy distribution (SED) in Figure 2.1.

The gravitational energy released as the circumnuclear material falls inwards and loses angular momentum forms the accretion flow around the black hole, which acts as the primary source of AGN continuum emission. The disk gets heated up via viscous heating (Peterson 1993), producing a thermal spectrum that has a blackbody temperature peaking in the ultra violet (UV; $\approx 10 - 400$ nm) for a "typical" AGN (i.e., with $M_{BH} \approx 10^8 M_{\odot}$; $\lambda_{Edd} \approx 0.1$); this is often referred to as the "big blue bump" (Harrison 2014; Figure 2.1)

The thermal disk photons, also called *seed photons*, are Comptonized by highly energetic electrons that are located at the *corona* (§2.1.2) to form the X-ray power-law spectrum observed in Figure 2.1. These high energy photons can also reflect off the accretion disk and/or the *torus* (§2.1.4), producing the additional hard X-ray "reflection" component. In a majority of AGN X-ray spectra, there is an additional "soft excess" (Singh et al. 1985; Arnaud et al. 1985) at energies less than ~ 2 keV. This emission is above what would be expected from simple accretion disk models and its origin is still a topic of discussion. Two proposed models (which could be simultaneously true) to explain this soft excess are: relativistically blurred ionized reflection from the

[1]The ratio L/L_{Edd}, where L is the bolometric luminosity and $L_{Edd} = 1.5 \times 10^{38} M_{BH}/M_{\odot}$ erg s^{-1} is the Eddington luminosity for a solar composition gas. This is the maximum isotropic luminosity a body can achieve while maintaining balance between radiation pressure and gravitational force.

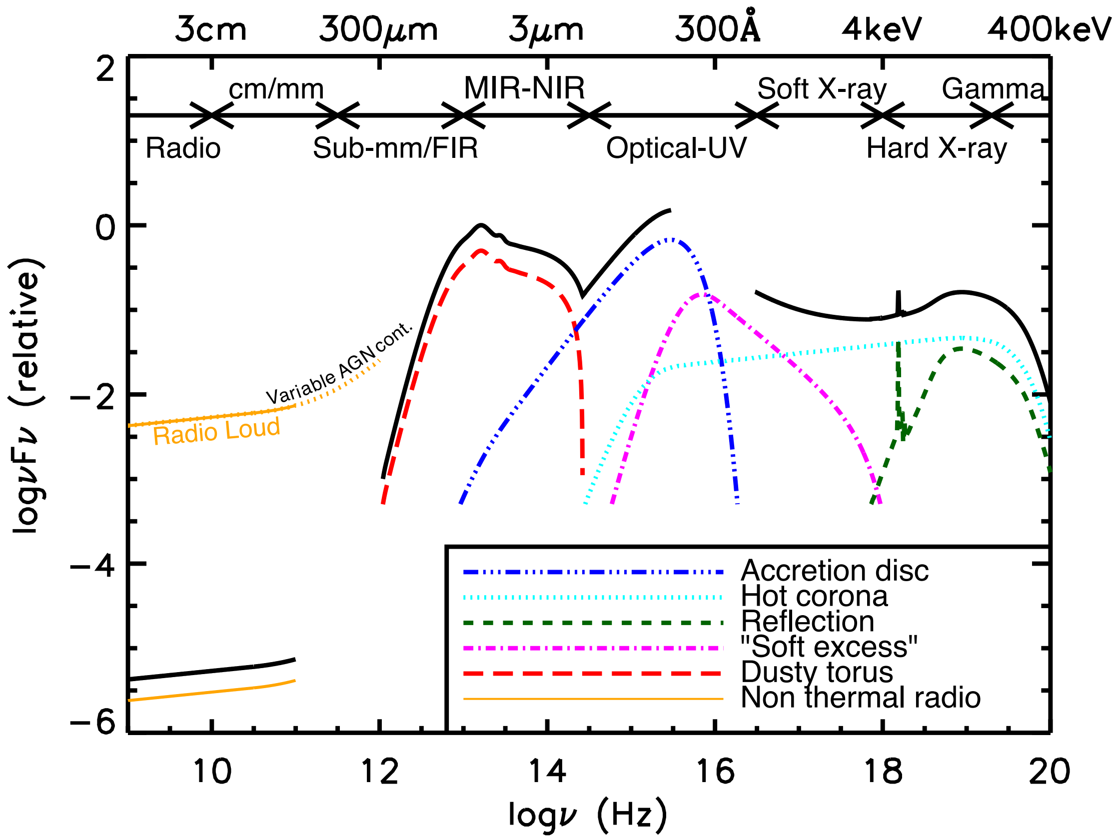


Figure 2.1: AGN SED from radio frequencies to hard X-ray, loosely based on the observed SEDs of radio-quiet quasars (e.g., Elvis et al. 1994; Richards et al. 2006). The black solid curve represents the total emission and the colored curves (shifted down for clarity) represent the individual constituents. The intrinsic shape of the SED in the mm-far infrared (FIR) regime is uncertain; however, it is credited to have a minimal contribution (to an overall galaxy SED) compared to star formation (SF), except in the most intrinsically luminous quasars and powerful jetted AGN. It should be noted that the relative contribution of each of the components can vary dramatically for different AGN types. Adapted from Harrison (2014).

inner accretion disc (e.g., Crummy et al. 2006), or Comptonization in a "warm corona" located near the surface of the accretion flow (e.g., Ballantyne and Xiang 2022).

The infrared (IR; $\approx 1 - 1000$ μm) band, is mostly sensitive to the reprocessed emission of the accretion disk by the torus. Although the shape of this IR emission varies from source to source, the peak of this dusty component is usually around $\lambda \approx 20 - 50$ μm and falls steeply at longer wavelengths (Harrison 2014). In the sub-mm/far-infrared (FIR), or sometimes even in the mid-infrared (MIR), stellar heated emission from cooler dust dominates over AGN heated emission (Lutz 2014).

AGN cover a wide range of radio luminosities, which can differ by several orders of magnitude between radio-quiet (RQ) and radio-loud (RL) sources, now more aptly referred to as non-jetted and jetted, respectively [2] (orange lines in Figure 2.1). The latter type makes up only a minority of the AGN population (e.g., Padovani 2011) with objects that present highly collimated, relativistic radio *jets* (§2.1.5) and extended radio structures such as lobes and hotspots, that together emit strong non-thermal radiation and, occasionally, a γ-ray jet. On the other hand, the spectrum of non-jetted objects is dominated by thermal emission in most bands and the origin

[2]While RL/RQ has been the historical terminology, radio-quiet sources are not truly radio-silent but rather "radio-faint" and they are, however, γ-ray quiet (Padovani 2017).

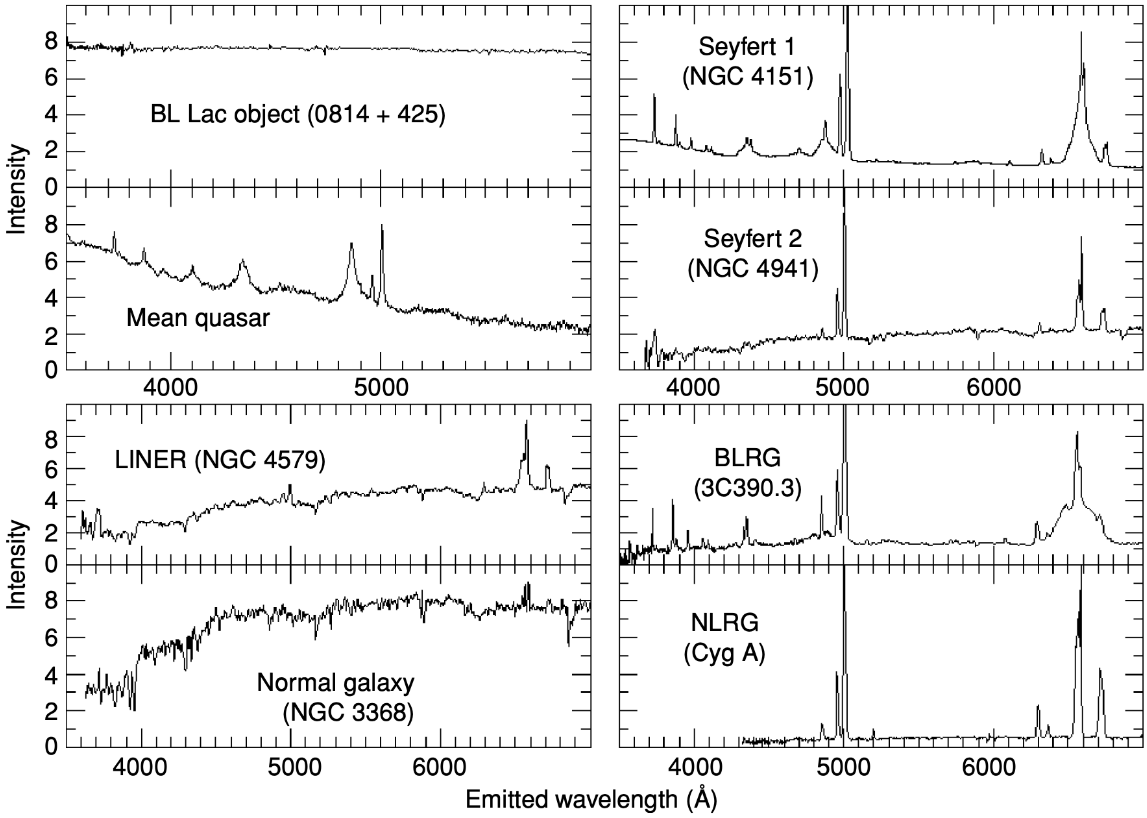


Figure 2.2: Optical spectra of representative AGN classes, from the (essentially featureless) BL Lac objects (Top left) through the radio galaxies and QSOs (Seward and Charles 2010). Note the presence of both broad and narrow emission lines in quasars, Seyfert 1 galaxies and broad-line radio galaxies (BLRG), while Seyfert 2 galaxies, low-ionisation nuclear emission region (LINER) galaxies and narrow-line radio galaxies (NLRG) only show narrow permitted and forbidden lines. A normal galaxy spectrum (Bottom left) is shown for comparison.

of their radio emission has not been clearly established. Some possibilities include coronal activity or synchrotron radiation from particles accelerated in shocks (see e.g., Laor and Behar 2008; Ishibashi and Courvoisier 2011).

2.1 AGN structure

Because different wavebands probe different physical elements, AGN selected at particular wavelengths often exhibit distinct properties, leading to a rich phenomenological classification (Figure 2.2). Reality, however, is much simpler because now we know that most of these classes differ from each other only by a small number of parameters, such as: orientation, accretion rate, the presence of jets, and possibly the host galaxy and environment (see Padovani et al. (2017) and references there-in). This suggests that the numerous brands do not necessarily correspond to fundamentally different physical objects, but instead motivate a unifying framework capable of explaining their apparent diversity.

This unification idea arose shortly after the discovery of quasars, when the observed variety of AGN prompted questions about the role of the observer's orientation relative to the AGN axis. Blandford and Rees (1978) proposed that BL Lac objects were radio galaxies viewed down the axis of a relativistic jet, where *relativistic beaming* (§2.1.5) would cause the non-thermal continuum to

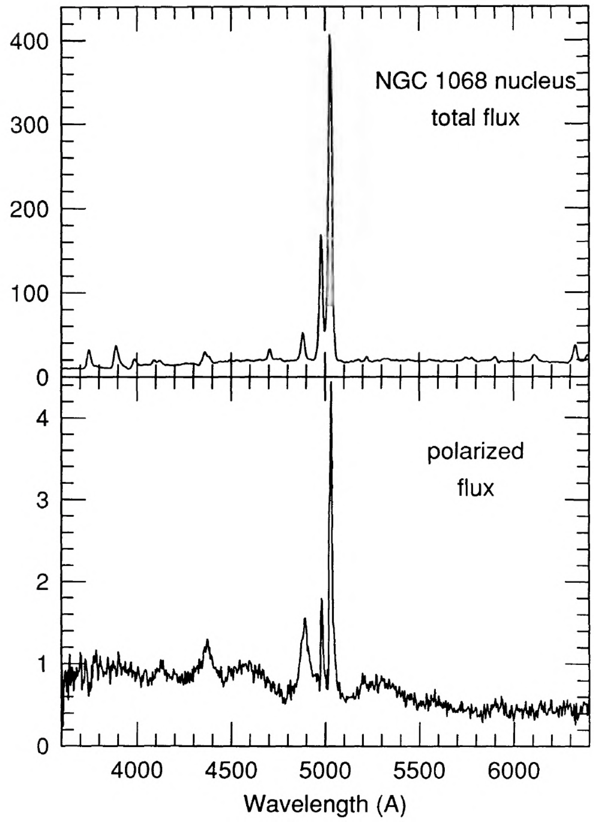


Figure 2.3: Optical spectra of the nucleus of NGC 1068 (Miller, Goodrich, and Mathews 1991). The total spectrum (Top) shows the narrow emission lines expected for Type 2 AGN, while the polarized flux spectrum (Bottom) reveals the broad permitted emission lines characteristic of unobscured Type 1 AGN.

appear very bright and the emission lines weak in comparison. The same object, viewed from the side, would have normal emission-line equivalent widths, and the radio structure would be dominated by the extended lobes rather than the core (Shields 2004).

Further evidence came from spectropolarimetry, the measurement of the polarization state of the incoming light from a source as a function of wavelength. Seyfert galaxies (Seyfert 1943) are primarily divided into two types based on their optical spectra: Type 1 objects exhibit both broad permitted emission lines such as Hα and Hβ (FWHM $>$ 1000 km s^{-1}) and narrow forbidden lines such as [O III] and [Ne III] (FWHM $<$ 1000 km s^{-1}), while Type 2 objects show only the narrow component. Crucially, both types share the same high- and low-ionization narrow forbidden lines with similar line ratios, suggesting they are powered by the same intrinsic engine.

Antonucci and Miller (1985) confirmed this by analyzing the polarized flux of NGC 1068, a Seyfert 2 galaxy[3], resembled that of a Seyfert 1 spectrum, as can be seen in Figure 2.3. This was interpreted as the broad-line region (BLR; §2.1.3) and the central continuum source being blocked from direct view by a dusty torus, while free electrons above the nucleus scatter a fraction of the nuclear light towards the observer, polarizing it in the process. The broad lines had gone unnoticed in direct light because this scattered component was too faint relative to the strong narrow-line emission from the narrow-line region (NLR; §2.1.6), which extends beyond the obscuring torus and is therefore visible regardless of orientation.

Multiple efforts such as those recounted above (e.g., Antonucci 1984; Orr and Browne 1982)

[3]It should be noted that a subset of Type 2 AGN ($\sim$ 30%), sometimes referred to as "true Type 2" objects, show no detectable broad lines even in polarized light and little X-ray absorption (see Netzer 2015 and references there-in).

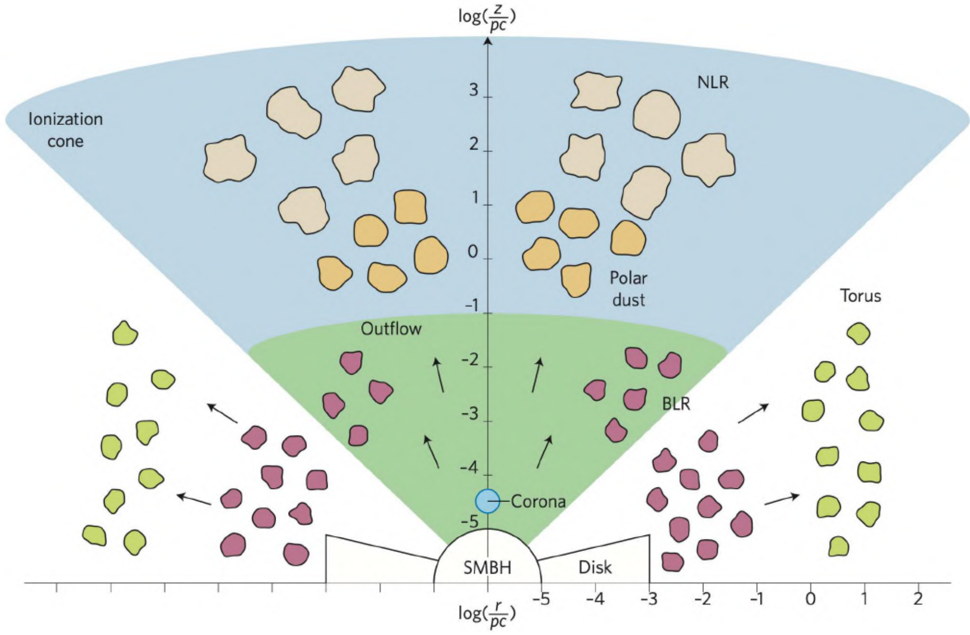


Figure 2.4: Overview of the innermost regions around a non-jetted accreating supermassive black hole (SMBH). Typical scales along the polar and equatorial directions (for nearby AGN) are shown. Different colors depict regions with different densities or compositions. Taken from Ramos Almeida and Ricci 2017.

were fundamental to the development of the "favoured" unified model of AGN introduced in this section. Whilst it has been succesful in homogenizing many kinds of active nuclei, recent works point to the conclusion that a picture based primarily on orientation and obscuring material is incomplete (e.g., Bianchi, Maiolino, and Risaliti 2012). Nonetheless, the major aspects of the model are sufficient for the purposes of this thesis; in the sections below, the primary components are explained briefly, while additional features such as polar dust and large-scale outflows (also visible in Figure 2.4) are discussed shortly when relevant but are not treated as separate subsections.

2.1.1 Accretion Disk

The standard framework of the optically thick, geometrically thin accretion disk (Shakura and Sunyaev 1973) described above applies to SMBHs with moderate to high Eddington ratios ($10^{-2} \lessapprox \lambda_{Edd} \lessapprox 1$; Ricci 2026), where the accretion flow is considered radiatively efficient. This scheme describes the disk as a composition of Keplerian rings, nested and coupled to each other by viscous forces, likely driven by turbulence and magnetic instabilities, the magneto-rotational instability (MRI; Balbus and Hawley 1998) being the most widely quoted. The outer, slower ring brakes the inner faster one, transporting the angular momentum outwards and allowing material to spiral inward. Friction between the rings produces heat that is dissipated as blackbody emission with a characteristic maximum temperature of $T \sim 10^5$ K for the "typical" AGN characterized above (Netzer 2013), yielding the multi-temperature spectrum depicted in Figure 2.1. The continuum emission of the accretion disk varies with time depending on the dynamics of the infalling matter (see Chapter 3).

Most AGN in the local universe are thought to accrete through thin disks, yet contrasting accretion scenarios are possible depending on the Eddington ratio (Figure 2.5). At $\lambda_{Edd} \lesssim 10^{-2}$, the disk becomes optically thin and with an Advection Dominated Accretion Flow (ADAF, Narayan

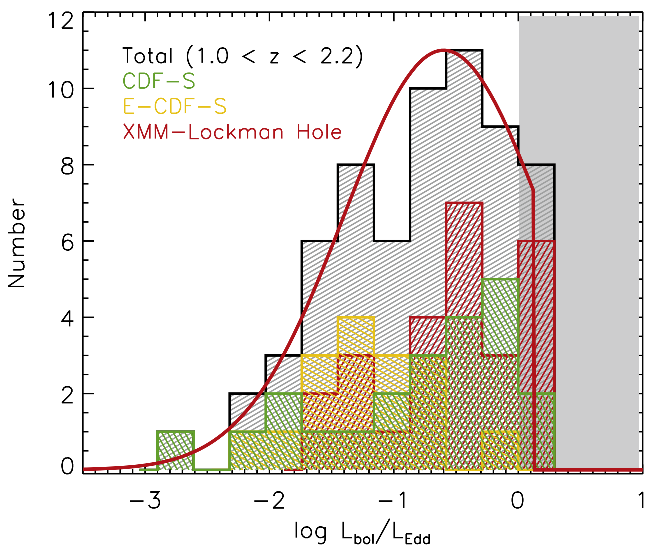


Figure 2.5: Eddington ratio distribution of broad-line AGNs at $1.0 < z < 2.2$. Different X-ray surveys are represented as different color histograms, and the black histogram pictures all the combined distributions. The gray shade illustrates the Eddington limit ($\lambda_{Edd} = 1$) and the red solid line corresponds to a log-normal fit with a peak of $\log L/L_{Edd} = -0.6$ (Suh et al. 2015).

and Yi 1994), radiating weakly and with minimal optical/UV emission. Conversely, at $\lambda_{Edd} \gtrsim 1$, theoretical works have predicted that the SMBHs accrete rapidly through "slim" (or "thick") disks, showing very different SEDs than their slower counterparts and powerful outflows (e.g., Kubota and Done 2018; Giustini and Proga 2019).

2.1.2 Corona

Numerical simulations of magnetohydrodynamical accretion disks have shown that the dissipation of the magnetic energy built up by the MRI naturally produces a hot, active corona located within a few gravitational radii ($R_g = GM/c^2$) of the accretion disk (Balbus and Hawley 1998; Miller and Stone 2000; Fabian et al. 2009). This compact region of ionized gas reaches temperatures of $T \sim 10^9$ K (Kara and García 2025), and its rapid variability (see Chapter 3) is consistent with such a small emitting region. The corona contributes approximately 5% of the AGN bolometric output (Ricci 2026), a fraction that decreases with increasing λ_{Edd}, with lower λ_{Edd} systems showing stronger X-ray emission relative to the disk (Padovani et al. 2017).

2.1.3 Broad-line Region

The broad-line region consists of high density gas clouds ($\gtrsim 10^9$ cm^{-3}) predominantly in Keplerian motion at luminosity-dependent distances of $0.01 - 1$ pc from the central black hole (Netzer 2015), well within the *dust sublimation radius* (see §2.1.4). The BLR is therefore dust free, emitting broad permitted and semi-forbidden emission lines with velocity widths ranging from $1,000$ to $20,000$ km s^{-1} (Ricci 2026), wideness caused by *Doppler broadening*[4]. Ionizing photons from the accretion disk and corona illuminate these clouds, which subsequently exhibit ionization stratification with respect to radius, with higher ionization species located closer to the central source (Netzer 2013). The physical origin of the BLR remains an open question, some of the models

[4]Doppler broadening refers to the superposition of emission from clouds moving at different directions/velocities along the line of sight produces a single broadened line whose width reflects the velocity dispersion of the gas.

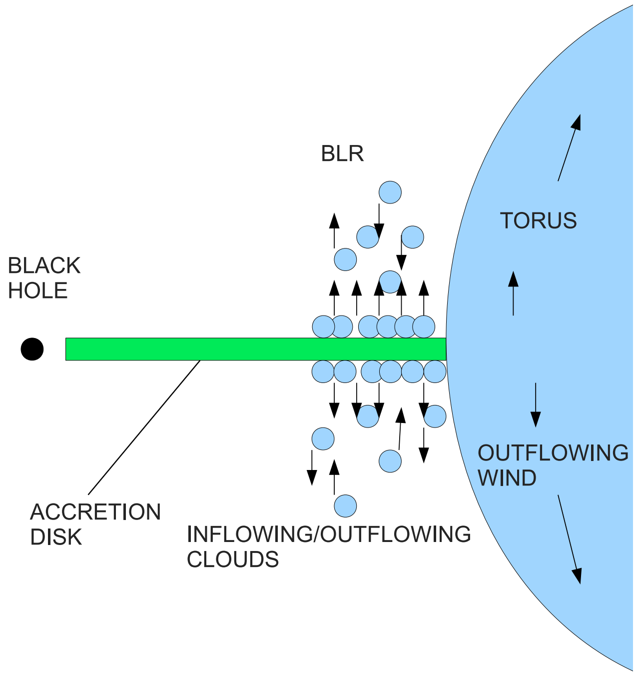


Figure 2.6: Schematic of the BLR geometry and its location relative to the accretion disk and dusty torus. The figure also illustrates the failed wind scenario of Czerny and Hryniewicz (2011): dust forms in the atmosphere of the inner accretion disk at radii where $T_{eff} < T_{sub}$, is launched as a wind, but then is suppressed by the radiation from the central source and partially falls back, giving rise to the inflowing and outflowing clouds that make up the BLR.

presented in the literature include radiation pressure driven disk winds (Murray et al. 1995) and failed dust-driven wind from the disk surface (Figure 2.6; Czerny and Hryniewicz 2011). Notably, the distance of the BLR from the central source scales with the AGN luminosity as $R_{\mathrm{BLR}} \propto L^{1/2}$ (Kaspi et al. 2000), a relationship that is fundamental in reverberation mapping studies (§3.1.1).

2.1.4 Dusty Torus

As posed previously, the presence of an assymetric "dusty obscurer" has successfully unified apparently different AGN classes. When observed edge-on (i.e., at high inclination angles), this torus[5] obscures the innermost regions of the AGN and hides the broad emission lines; conversely, when viewed face-on (at lower inclination angles), there is a clear line of sight to the BLR and accretion disk (Seyfert 1 vs Seyfert 2 in Figure 2.2). This structure is conceptualized as an optically thick structure surrounding the central engine at scales larger than that of the BLR and luminosity dependent dimensions ranging from $0.1 - 10$ pc (Netzer 2015). Its inner boundary coincides is the dust sublimation radius (R_{sub}), beyond which the dust can survive the radiation from the central engine. This radius corresponds to the distance at which the dust temperature reaches the sublimation threshold ($T_{sub} \sim 1400 - 1800$ K for silicate and graphite grains respectively; e.g., Barvainis 1987; Netzer 2015).

In addition to being responsible for the IR reprocessed emission on the AGN SED (Figure 2.1), the torus acts as a gas and dust reservoir that regulates the material available for accretion onto the black hole. Its geometry, believed to contain a central opening, also collimates the AGN radiation

[5]For convenience, we keep referring to this obscuring matter as the torus despite the fact that it could have a different (and more complex) geometrical distribution (e.g., Nenkova et al. 2008; Stalevski et al. 2012).

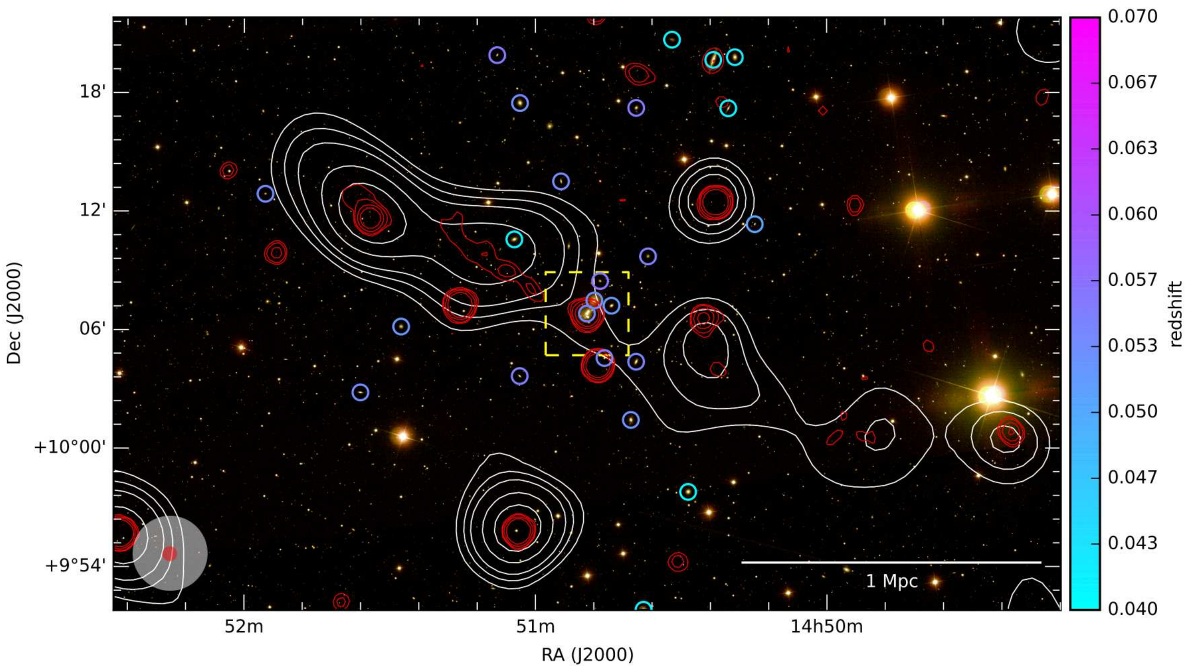


Figure 2.7: SDSS optical image (composite from bands g, r and i) of the galaxy group UGC 9555 overlaid with NVSS contours (red) and LOFAR MSSS radio contours (white) revealing the size of the giant radio galaxy (GRG). The white bar indicates a physical scale of 1 Mpc, illustrating how AGN jets can extend well beyond the boundaries of their host galaxies. Adapted from Clarke et al. 2017.

into bi-conical ionization cones along the polar direction (Pogge 1988), which illuminate the NLR (§2.1.6). Furthermore, AGN drive large scale winds and outflows along these directions, powered by radiation and magnetic pressure from the central engine, which can interact with and regulate the host galaxy interstellar medium (Ricci 2026).

2.1.5 Relativistic Jets

Jets are thought to be powered by the extraction of rotational energy from the SMBH via strong electromagnetic fields (Blandford and Znajek 1977), and can extend to Mpc scales from their host galaxy (Figure 2.7; Clarke et al. 2017).

Their launch point coincides with the location of the corona, in the immediate vicinity of the accreting object, and their apparent velocities range from mildly to highly relativistic motion. The broadband emission of jets is characterized by a double-peaked SED: the low-energy peak, spanning radio to optical wavelengths, arises from synchrotron radiation of relativistic electrons spiralling in the jet magnetic field, while the high-energy peak, from X-ray to γ-ray energies, is thought to result from inverse-Compton scattering of soft photons from the BLR and accretion disk by the same relativistic electrons (Netzer 2013).

Their observed properties depend strongly on orientation because relativistic beaming could make an approaching jet appear significantly brighter than a receding one. This effect, known as *Doppler boosting*, arises when the jet plasma has bulk relativistic motion that concentrates the emitted radiation into a narrow cone in the direction of motion, amplifying the observed flux and variability for jets viewed close to face-on and explainig the extreme characteristics observed on blazars (Urry and Padovani 1995). Jets are also thought to play a role in galaxy evolution by injecting energy into the surrounding environment, a process referred to as radio-mode feedback (Fabian 2012).

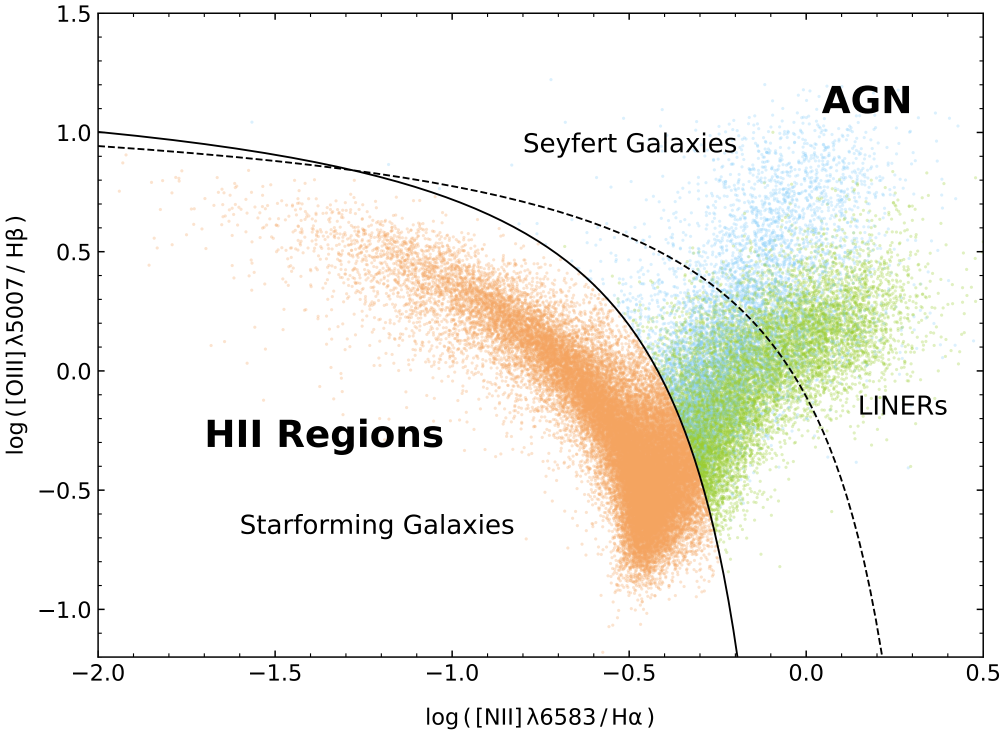


Figure 2.8: Select galaxy objects from the Sloan Digital Sky Survey (SDSS) catalog shown in the BPT diagram. It plots the intensity of the [OIII] $\lambda5007$ forbidden transition relative to that of the Hβ line versus the ratio of [NII] $\lambda6583$ to Hα on logarithmic axes. Using this to infer ionization states and abundances of the emitting medium enables us to reconstruct the origin of the dominant radiation in the galactic host, which has led to various approaches for the classification of these objects. Depicted is the theoretical delimiter by Kewley et al. (2001) as well as the empirical line by Kauffmann et al. (2003) to distinguish predominantly starforming galaxies from AGN. Additionally, the Schawinski et al. (2007) boundary sorts AGN into efficiently accreting Seyfert and inefficient LINER subtypes. Courtesy of Fritz A. Agildere.

2.1.6 Narrow-line Region

This assymetric region is extended enough to be resolved optically (e.g., Fischer et al. 2013; Polack et al. 2024), spanning hundreds or even thousands parsecs along the ionization cones (Figure 2.4). The NLR, just as the BLR, is photoionized by the central engine and exhibits optical emission lines, kinematically broadened with typical widths of 10^{2-3} km s^{-1} (Bennert 2005), with most lines falling within the $350 - 500$ km s^{-1} range (Ricci 2026).

Compared to the BLR, the NLR has a lower electron density ($\sim 10^{2-3}$ cm^{-3}), allowing forbidden transitions to occur without being collisionally suppressed. The presence of these forbidden lines, together with the high ionization state of the gas, can trace AGN activity and makes the NLR emission diagnostically distinct from that of star-forming regions, a separation commonly visualised using emission line ratio diagrams such as the Baldwin-Phillips-Terlevich (BPT) diagram (Baldwin, Phillips, and Terlevich 1981) (Figure 2.8).

3 Variability of AGN

3.1 Physical Origin and Timescales

Early campaigns quickly revealed that AGN vary much more rapidly in X-rays than in the optical (Winkler and White 1975), sometimes on timescales as short as a few minutes (McHardy et al. 1981), and that variability at different wavelengths is correlated (Ulrich et al. 1997). This stochastic behaviour can generally be explained by the *propagating fluctuations* theory (Lyubarskii 1997). Local fluctuations in the disk's surface density, presumably driven by the MRI (§ 2.1.1), occur at all radii and propagate inwards at the local viscous timescale[1]. As these fluctuations travel, they modulate the local mass accretion rate at progressively smaller radii, coupling the variability of the outer disk to that of the hotter inner regions, where the shorter local viscous timescale drives faster fluctuations in the emitted luminosity (Uttley et al. 2023). The X-ray variability produced in the innermost regions (i.e., in the corona; Krolik et al. 1991) then irradiates the outer accretion disk and is reprocessed, driving the optical and UV variability observed at longer wavelengths with a characteristic time delay, a *reverberation*, as discussed below.

Naturally, both propagating fluctuation lags and reverberation lags are simultaneously present in the data, but dominate at different temporal frequencies and have opposite energy dependences (Figure 3.1): at low frequencies ($\nu \lesssim 10^{-5}$ Hz for a $10^6\, M_\odot$ black hole, or equivalently, $\gtrsim 1$ day), hard X-ray variations lag soft ones as accretion fluctuations propagate inward; at high frequencies ($\nu \gtrsim 10^{-3}$ Hz, timescales of minutes), soft X-rays lag hard ones as the disk responds to the driving corona (Kara and García 2025). Fourier analysis is essential for disentangling the two contributions.

Beyond this persistent stochastic variability, more extreme variability phenomena also exist, including *changing-look* AGN, which undergo dramatic changes in their broad emission lines and continuum on timescales of months to years, possibly driven by rapid changes in accretion rate or obscuration (LaMassa et al. 2015), as well as quasi-periodic eruptions (QPEs) and tidal disruption events (TDEs); these represent transient phenomena superimposed on the underlying stochastic variability and are not discussed further here.

[1]The viscous timescale, $t_{\rm visc} \sim r^2/\nu$, where ν is the kinematic viscosity, indicates the time for material to drift radially through the disk and increases with radius.

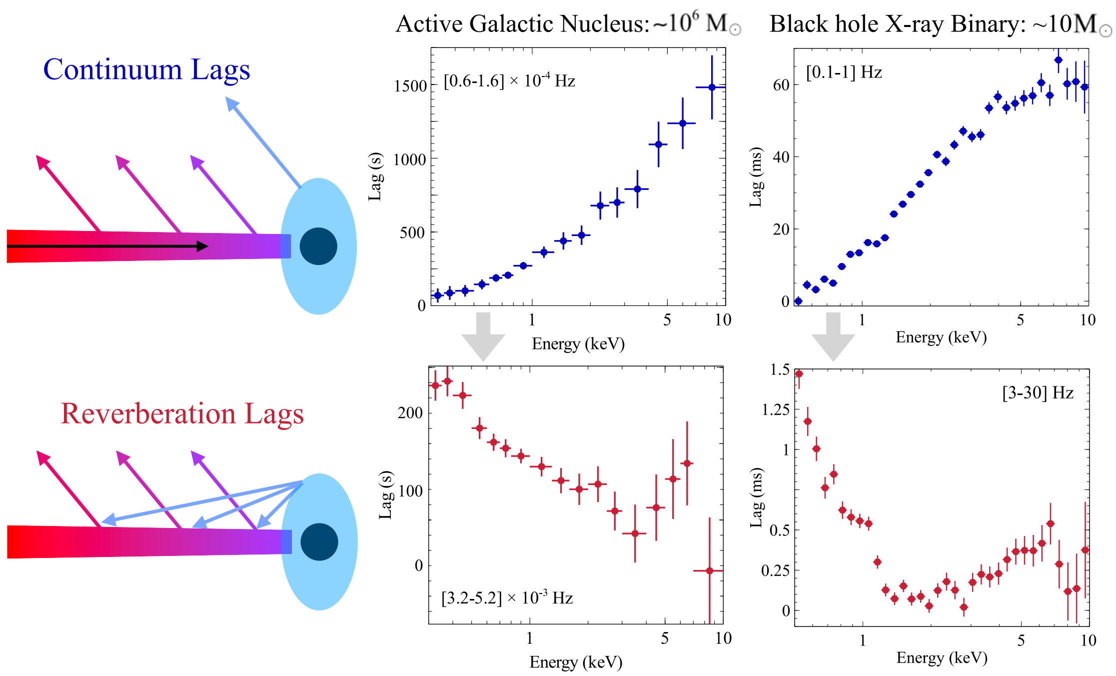


Figure 3.1: Frequency-resolved X-ray time lags as a function of energy for an AGN ($\sim 10^6\, M_\odot$, left) and a black hole X-ray binary ($\sim 10\, M_\odot$, right), illustrating the universality of the two lag regimes. At low Fourier frequencies (top, blue), hard X-rays lag soft ones, consistent with inward-propagating accretion fluctuations. At high Fourier frequencies (bottom, red), soft X-rays lag hard ones, consistent with disk reverberation. Taken from Kara and García 2025.

3.1.1 Reverberation Mapping

The reverberation delays introduced above are not only a complication in the interpretation of light curves, they are also a powerful tool. Because the different regions of an AGN cannot be spatially resolved, their sizes cannot be measured directly, but the light-travel time between a driving source and a region that reprocesses its emission provides an indirect gauge. *Reverberation mapping* (RM; Blandford and McKee 1982) exploits this by monitoring the delay with which one part of the AGN responds to variations in another: a region at a distance R from the driving source responds on average after the light-crossing time $\tau_{\rm lag} \sim R/c$, so that measuring the lag yields the size of the region. The characteristic scales and their corresponding light-crossing times for the main AGN regions are summarized in Table 3.1, ranging from hours for the X-ray corona to years for the dusty torus.

Table 3.1: Order of magnitude size scales and their equivalent light-crossing times for the different regions of an AGN. Adapted from Cackett et al. 2021.

Region	Size scale	Light-crossing time	
	(R_g)	$10^7\, M_\odot$	$10^9\, M_\odot$
X-ray corona	10	500 s	14 hr
UV/optical disk	10^2–10^4	0.06–6 days	6–600 days
BLR	10^3–10^5	0.6–60 days	60–6000 days
Dusty torus	$> 10^5$	>60 days	>6000 days

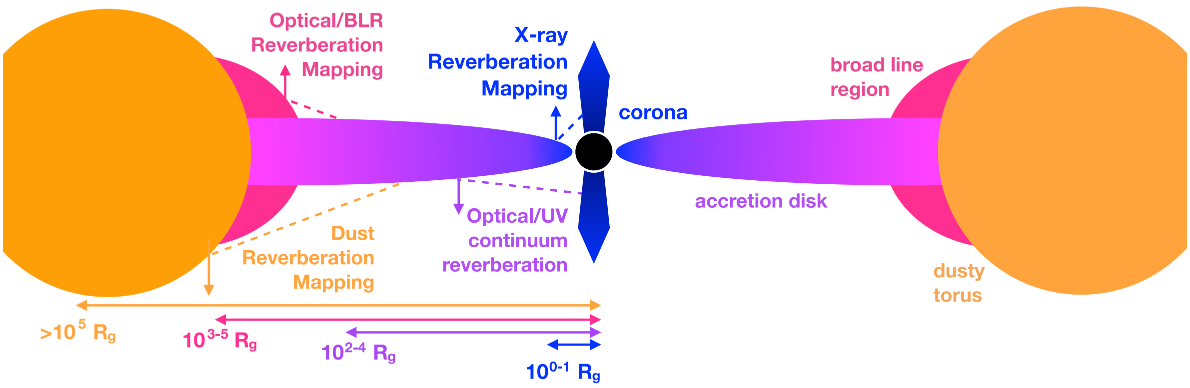


Figure 3.2: Cross-sectional schematic of an AGN depicting the four main types of reverberation mapping and their corresponding radial scales from the central black hole. Taken from Cackett et al. 2021.

The two most established applications probe different regions. In *emission-line RM*, the broad emission lines respond to variations in the ionising continuum with a delay that measures the size of the BLR; combined with the velocity width of the lines, this yields an estimate of the black hole mass and underlies the radius-luminosity relation introduced in § 2.1.3 (Peterson 1993). In *continuum RM*, the delays are measured between different wavebands of the disk continuum itself, mapping the temperature structure of the accretion disk (Collier et al. 1999). For a standard thin disk the lag increases with wavelength as:

$$\tau_\lambda \propto \lambda^{4/3} \left[\frac{M_{\rm BH}\dot{M}}{(3 + f_{\rm irr})} \right]^{1/3}, \tag{3.1}$$

since longer wavelengths originate at larger, cooler radii, with the normalization set by the black hole mass, the accretion rate $\dot{M}$, and the fraction of irradiating flux $f_{\rm irr}$ (Cackett et al. 2021).

Formally, the observed response is described as the convolution of the driving light curve with a *transfer function* that encodes the geometry and kinematics of the responding region (Blandford and McKee 1982). Recovering this transfer function, or even the lag alone, is an ill-posed problem: the light curves are irregularly sampled and noisy, and different underlying signals can reproduce the same observations to within their uncertainties. This degeneracy is not unique to RM but is a general feature of inference from AGN light curves, and it motivates treating the reconstruction of a light curve probabilistically, returning a distribution of signals consistent with the data rather than a single best-fit curve. It is this probabilistic reconstruction of the individual light curve that is the subject of this thesis.

3.2 Variability Characterization

Variability has been studied under a variety of methods, starting from simple statistical quantities tracing the amplitude and timescales of the observed variability to more sophisticated approaches in both the time and frequency domains. In general, these statistics attempt to measure the absolute or fractional variance of the light curve over some timescale Δt, or express the degree of correlation between different timescales or photometric bands (see Paolillo and Papadakis 2025 and references there-in).

The distribution of variability power over temporal frequencies (or, equivalently, 1/timescale)

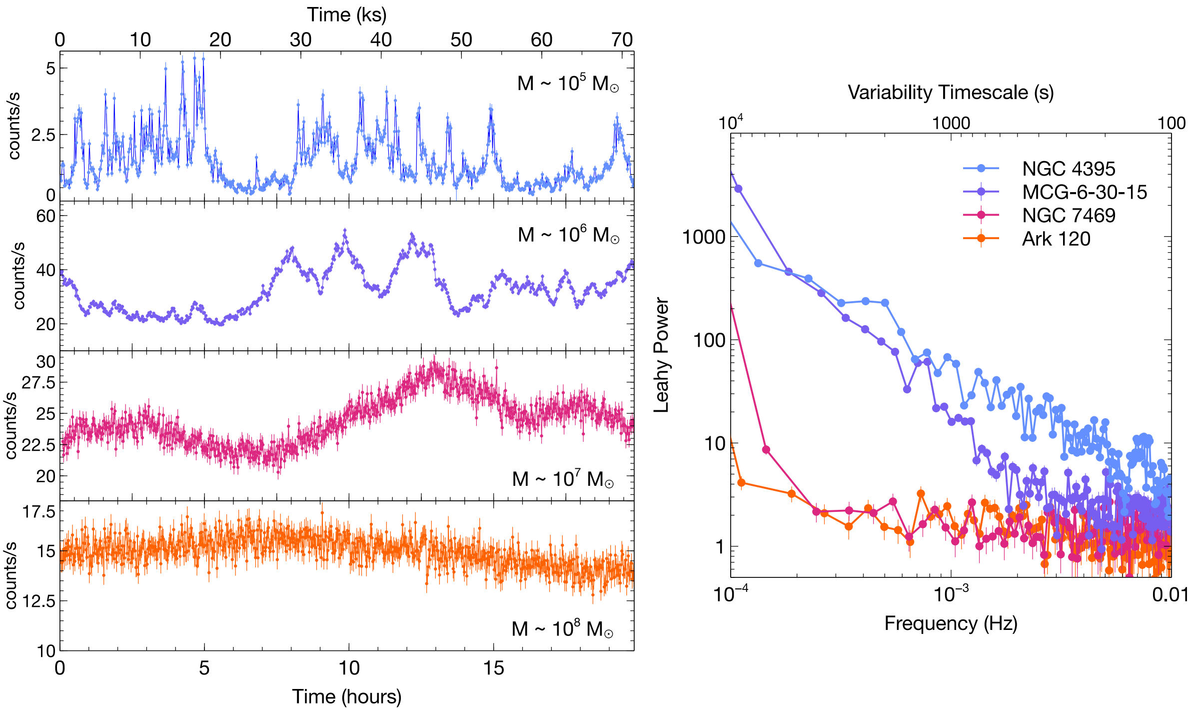


Figure 3.3: Left: X-ray light curves of four Seyfert 1 AGN with black hole masses spanning 10^5–10^8 $M_\odot$, observed with XMM-Newton over 70 ks. Smaller black holes vary more rapidly, with the 10^5 $M_\odot$ AGN (NGC 4395, top) showing variability on timescales of tens of seconds, while the 10^8 $M_\odot$ AGN (Ark 120, bottom) varies on timescales of hours. Right: The corresponding PSDs, showing the red-noise power-law shape characteristic of AGN X-ray variability. Taken from Kara and García 2025.

is quantified by the *power spectral density* (PSD), defined as the squared amplitude of the flux (e.g., Uttley et al. 2002). AGN PSDs typically follow a "red-noise" power law, $P(\nu) \propto \nu^{-\alpha}$ with $\alpha \sim 1$ at low frequencies, steepening to $\alpha \gtrsim 2$ above a characteristic break frequency ν_b, consequently displaying larger amplitude variations on longer timescales. This break timescale scales with black hole mass and accretion rate (McHardy et al. 2006), with smaller black holes varying more rapidly than bigger ones for the same observing duration (Figure 3.3). The anticorrelation can be interpreted as a consequence of the fact that the most luminous AGN are likely to host larger mass black holes (if all quasars accrete at similar Eddington rates), in which case, their variability should be diluted by the larger emitting region (Paolillo and Papadakis 2025).

The PSD also depends strongly on the observing waveband, with optical PSDs showing steeper slopes and lower normalization than X-ray PSDs, consistent with the propagating fluctuations picture where outer disk regions contribute power at lower frequencies and inner regions at higher ones (Kara and García 2025).

Both the amplitude and the timescale of the variability correlate with the physical properties of the accreting system. The *damping timescale* τ, which determines when the flux decorrelates from its past, increases with black hole mass and, more weakly, with rest-frame wavelength (e.g., Collier and Peterson 2001; Burke et al. 2021) linking the variability timescale and the thermal timescale of the accretion disk. The *variability amplitude* σ, in turn, decreases towards longer wavelengths, so that AGN vary with larger fractional amplitude in the bluer bands (MacLeod et al. 2010), consistent with the bluer emission arising from the hotter, more compact inner disk. These scaling relations, and in particular the dependence of τ on black hole mass, are used in Chapter 4 to assign physically motivated parameters to the simulated light curves.

A complementary characterization is provided in the time domain by the *structure function*

(SF), which measures the average magnitude variation as a function of timescale (or time lag Δt separating pairs of observations), adequate for the sparse, irregularly sampled light curves typical of optical surveys, where a direct estimate of the PSD is difficult (Kozłowski 2016b). On short lags the SF rises as a power law, reflecting the growing amplitude of the variability, and flattens above the damping timescale τ to a constant level set by the long-term variability amplitude, mirroring in the time domain the same turnover seen as the PSD break at $\nu_b \sim 1/\tau$.

Nevertheless, the SF has well-known limitations because its points are heavily correlated since the underlying light curve points are themselves correlated, which prevents proper model fitting via chi-square minimization and can produce spurious breaks that depend on the sampling pattern and light curve length (Emmanoulopoulos, McHardy, and Uttley 2010)

A further related tool is the *autocorrelation function* (ACF), which quantifies the degree of "self-correlation" between lagged values of a time series as a function of the time lag Δt, normalized to unity at zero lag. The ACF and the PSD are related by the Wiener-Khinchin theorem[2], making them equivalent representations of the same underlying variability process. The shape of the SF, the PSD break, and the ACF decay therefore all encode the same two quantities that characterize AGN variability as a whole: an amplitude and a damping timescale. These are precisely the parameters of the *damped random walk* (DRW) adopted to model the light curves in the following chapter.

3.3 Time-Domain Surveys

The advent of new facilities and major advances in instrumentation have led to outstanding progress in time-domain astronomy. As discussed at the start of this chapter, the nuclear regions of AGN cannot be spatially resolved for all but a handful of sources, and for the vast majority the only tracer of their inner structure is the variation of their brightness over time. This is what Wide-Field surveys can provide, rather than resolving individual objects, they monitor large areas of sky repeatedly over long baselines, building light curves for enormous numbers of sources at once.

This capability has grown over successive generations of surveys. The Sloan Digital Sky Survey (York et al. 2000) assembled the first large photometric and spectroscopic samples of quasars over a wide area and provided some of the earliest large-scale variability measurements. Dedicated time-domain surveys followed, trading depth for cadence and area. The Palomar Transient Factory (Law et al. 2009) and its successor at the same telescope, the Zwicky Transient Facility, scanned the northern sky at a rate of hundreds of square degrees per hour, while surveys such as Pan-STARRS (Chambers et al. 2016) extended monitoring to fainter magnitudes. Each step increased the number of sources monitored and the regularity with which they were revisited, and the forthcoming Legacy Survey of Space and Time represents the next.

What can be extracted from these light curves is set by the survey itself, by its cadence, the number of bands it observes, and the photometric precision it reaches in each. These are also its limitations. A ground-based survey observes only at night and in clear weather, each field is visible for only part of the year, and the depth reached in a single exposure bounds the precision of every measurement, so the light curves it delivers are irregularly sampled, interrupted by seasonal gaps, and carry a per-epoch uncertainty that varies with the observing conditions. Two surveys are relevant to this work: the Zwicky Transient Facility, from which the real light curves reconstructed in Chapter 6 are drawn, and the Legacy Survey of Space and Time.

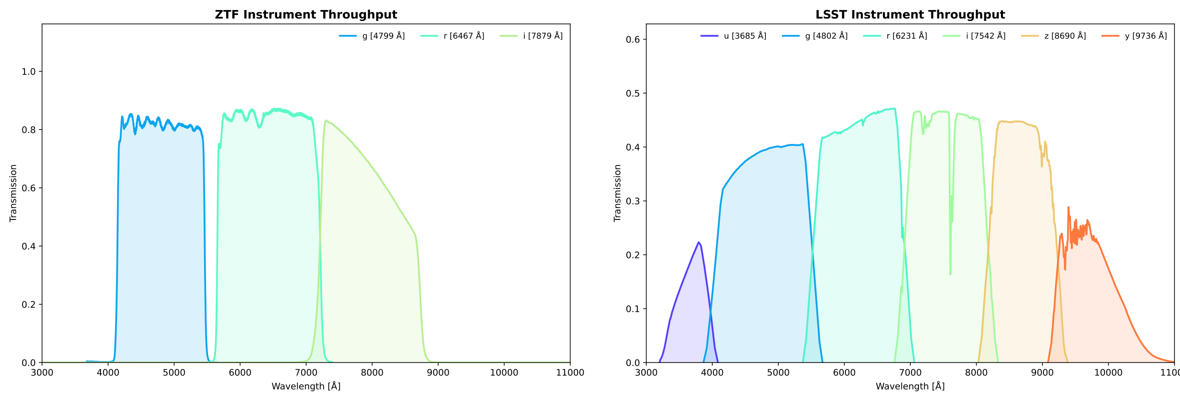
[2] The Wiener-Khinchin theorem states that the PSD of a wide-sense stationary process is the Fourier transform of its ACF: $P(\nu) = \int_{-\infty}^{\infty} \mathrm{ACF}(\Delta t)\, e^{-2\pi i \nu \Delta t}\, d\Delta t$.

3.3.1 The Zwicky Transient Facility

The Zwicky Transient Facility (ZTF) (Bellm et al. 2019; Graham et al. 2019) is an optical time-domain survey carried out with the 48-inch Samuel Oschin Schmidt Telescope at Palomar Observatory, operating since 2018. Its wide-field camera images tens of square degrees at once, covering the entire visible northern sky every few nights. Observations are taken in three custom bands, g, r, and i, with central wavelengths of 4799, 6467, and 7879 Å (Figure 3.4a). The public Northern Sky Survey observes in g and r on a cadence of about three nights, reaching median five-sigma depths of 20.8 and 20.6 mag in 30 second exposures. The i band is observed within a separate programme, on a slower cadence and to a shallower depth of about 20.0 mag, so it is more sparsely sampled than the other two. This uneven coverage between bands carries directly into the light curves used in this work, in which the i band contains the fewest epochs and the widest gaps.

3.3.2 Legacy Survey of Space and Time

The Legacy Survey of Space and Time (LSST) (Ivezić et al. 2019), to be conducted at the Vera C. Rubin Observatory, will survey the southern sky over ten years. It uses a telescope with an 8.4 meter primary mirror (6.5 m effective aperture) and a 9.6 square degree field of view, imaged by a 3.2-gigapixel camera, in six bands, u, g, r, i, z, and y, spanning 320 to 1050 nm (Figure 3.4b). In the main Wide-Fast-Deep survey each of roughly 18,000 square degrees is visited on the order of eight hundred times over the ten years, distributed among the six bands, with a single-exposure five-sigma depth of about 24.0 mag in r. The photometric accuracy of the survey is specified at 10 mmag[3], and over its lifetime it is expected to deliver light curves for more than ten million AGN, each sampled from tens to a few hundred times per band.



(a) ZTF

(b) LSST

Figure 3.4: Instrument throughput of the two surveys. **(a)** The three ZTF bands g, r, and i, with central wavelengths 4799, 6467, and 7879 Å. **(b)** The six LSST bands u, g, r, i, z, and y, spanning 320 to 1050 nm, including atmospheric transmission, optics, and detector sensitivity.

[3] `https://rubinobservatory.org/for-scientists/rubin-101/key-numbers`

4 Simulation Framework

The reconstruction method developed in this thesis is trained and tested, first, on simulated light curves, for which the underlying signal is known exactly. This chapter sets out the framework used to generate them. It begins with the stochastic process and damped random walk, derives the transition probability that the reconstruction method will later learn, and then describes the simulation itself, including the way physically motivated timescales are assigned and the sampling regime.

4.1 Stochastic Processes

The variability described in Chapter 3 is aperiodic and aleatory, so it is not well described by a deterministic model with a fixed set of parameters. It is instead treated as a realization of a *stochastic process*: a collection of random variables indexed by time, whose statistical properties, rather than whose exact values, are the object of study (Ross 1995). What is predicted at a given epoch is then a distribution, conditioned on the known values, rather than a single brightness value.

The simplest stochastic model consistent with the observed variability of AGN is the damped random walk, also known as an Ornstein-Uhlenbeck process. It is described by the stochastic differential equation

$$df(t) = -\frac{1}{\tau}(f(t) - \mu)\,dt + \sigma\sqrt{\frac{2}{\tau}}\,dW_t, \tag{4.1}$$

where $f(t)$ is the light curve at time t, τ is the damping timescale, μ the long-term mean, σ the stationary standard deviation, i.e. the variability amplitude of the light curve[1], and $dW_t \sim \mathcal{N}(0, dt)$ is an infinitesimal increment of a standard Wiener process, a continuous-time stochastic process which is random at every timescale. The first one is a restoring term that pulls $f(t)$ back towards μ on the timescale τ, so that excursions do not grow without bound, while the second perturbs it with stochastic fluctuations. The competition between damping and driving produces the characteristic behavior of AGN light curves, which vary coherently on timescales shorter than τ and decorrelate into noise on timescales much longer than τ.

[1]The SDE is more commonly written with a driving amplitude σ_{d} in place of $\sigma\sqrt{2/\tau}$. That form is less convenient here because σ_{d} is not directly observable, whereas this parametrization expresses the SDE in terms of the variability amplitude σ introduced in Chapter 3.

The DRW provides a good statistical description of optical AGN light curves (e.g., Kelly et al. 2009; Wilkins 2019; McLaughlin et al. 2024) and, importantly, it is a *Gaussian process*: a distribution over functions such that the values at any finite set of times follow a multivariate normal distribution, specified by a mean function and a covariance function (Rasmussen and Williams 2006). This is also why $f(t)$ is taken to be a magnitude rather than a flux: the flux light curves of accreting sources are approximately lognormal (e.g., Giebels and Degrange 2009; De Cicco et al. 2022) and so close to Gaussian in magnitude, matching the Gaussian process assumption (Yu et al. 2025). For the DRW the mean is constant at μ and the covariance function, the *kernel*, is exponential,

$$k(\Delta t) = \sigma^2 \exp\left(-\frac{|\Delta t|}{\tau}\right), \tag{4.2}$$

with $\Delta t = |t_i - t_j|$. Because any finite collection of values is jointly Gaussian, conditioning on observed values yields also a Gaussian predictive distribution in closed form, which is what makes such processes analytically tractable for interpolation. Equation 4.2 is the $\beta = 1$ member of a broader family of autocorrelation functions,

$$\mathrm{ACF}(\Delta t) = \exp\left(-\left(\frac{|\Delta t|}{\tau}\right)^{\beta}\right), \tag{4.3}$$

related to the covariance by $k(\Delta t) = \sigma^2\,\mathrm{ACF}(\Delta t)$, with $\beta = 1$ recovering the DRW exponential decay and $\beta = 2$ a Gaussian correlation. The exponent controls the smoothness of the realizations: $\beta = 1$ gives paths that are continuous but nowhere differentiable, consistent with the Wiener driving term of Equation 4.1, whereas $\beta = 2$ gives realizations that are infinitely differentiable and therefore too smooth to represent AGN variability.

The DRW case corresponds to a power spectrum $P(\nu) \propto \nu^{-2}$ at high frequencies, flattening to white noise below the break frequency $\nu_b = 1/(2\pi\tau)$. In the structure function representation, the DRW produces a power-law rise of slope $\gamma = 0.5$ on short timescales, flattening above τ (Figure 4.1); real AGN light curves are consistent with exponents scattered around this value (Kozłowski 2016b), which is one of the empirical justifications for adopting the DRW as a baseline.

The damped random walk is also the simplest member of *continuous-time autoregressive moving average* (CARMA) processes (Kelly et al. 2014), corresponding to the case CARMA$(1, 0)$. A general CARMA(p, q) process is the solution of a linear stochastic differential equation of order p, driven by white noise through an order q moving-average term, and can reproduce a wider range of power-spectrum shapes than the damped random walk alone. In particular the CARMA$(2, 1)$ process, also known as the damped harmonic oscillator, has been found to describe some observed AGN light curves more faithfully (Yu et al. 2022).

Both of these, however, are only phenomenological models and the precise PSD of intrinsic AGN UV/optical variability remains under debate (e.g., Petrecca et al. 2024; Beard et al. 2025). The damped random walk is nonetheless retained throughout this work as the generative model, both because it remains an adequate approximation on the timescales probed by the surveys considered here and because its analytic tractability provides a controlled setting in which to develop and test the reconstruction method.

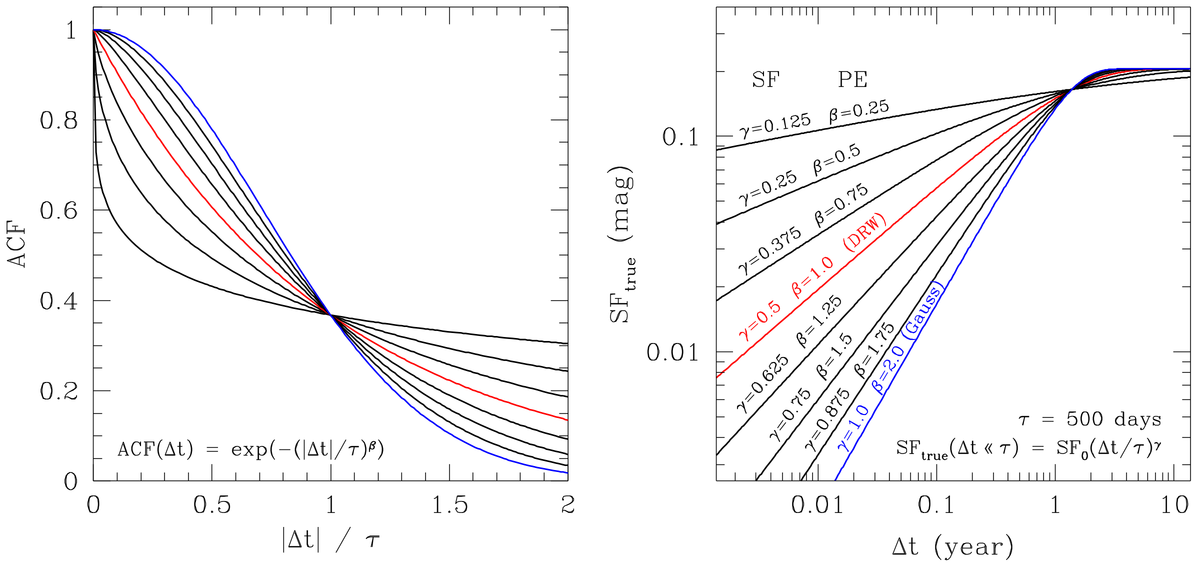


Figure 4.1: Left: autocorrelation functions of the family of Equation 4.3 for a range of the decay exponent β. Right: the corresponding structure functions, with short-timescale slope γ. The DRW ($\beta = 1$, red) produces a structure function of slope $\gamma = 0.5$ below the damping timescale, flattening above it. All curves cross at $\Delta t = \tau$, illustrating the role of the damping timescale as the characteristic decorrelation lag. Adapted from Kozłowski 2016b.

4.1.1 The Transition Probability

Conditioning a Gaussian process on N observations gives a Gaussian predictive distribution whose mean is a weighted combination of all of them. For the DRW this collapses to a much simpler statement, because Equation 4.1 is first-order and therefore Markovian[2]: the most recent value contains all the information relevant to predicting later times, so conditioning on a single point discards nothing. Setting $N = 1$ in the conditioning formulae, with $k(0) = \sigma^2$ and $k(\Delta t) = \sigma^2 e^{-\Delta t/\tau}$, gives the *transition probability*

$$f(t + \Delta t) \mid f(t) \ \sim\ \mathcal{N}\big(\mu + e^{-\Delta t/\tau}\big(f(t) - \mu\big),\ \sigma^2\big(1 - e^{-2\Delta t/\tau}\big)\big)\,, \tag{4.4}$$

derived in full in Appendix A. The mean relaxes exponentially towards μ on the timescale τ, while the variance grows from zero at $\Delta t = 0$ to the stationary value σ^2 as $\Delta t \gg \tau$, at which point the memory is lost and the distribution approaches the stationary marginal $\mathcal{N}(\mu, \sigma^2)$. Note that the variance depends only on Δt and not on $f(t)$, so two light curves observed with the same cadence receive the same uncertainty regardless of their measured values.

Equation 4.4 is the object that the reconstruction method of Chapter 5 learns from data, and it also provides the exact Gaussian process baseline against which the method is compared. Throughout this work the light curves are centered so that $\mu = 0$, reducing the mean to $f(t)\, e^{-\Delta t/\tau}$.

A note on conventions: throughout this work σ denotes the stationary standard deviation of the light curve, following the EzTao convention (Yu and Richards 2022). The driving amplitude $\sigma_{\rm d}$ of Equation 4.1 is the quantity denoted $\sigma_{\rm KBS}$ by Kelly et al. (2009), related to the stationary value by $\sigma^2 = \tau\, \sigma_{\rm d}^2/2$, while the structure function amplitude defined in MacLeod et al. 2010 is $\mathrm{SF}_\infty = \sqrt{2}\,\sigma$. The damping timescale τ refers to the rest-frame value throughout.

[2] A process is Markovian if its future depends only on its present state and not on the path by which it was reached.

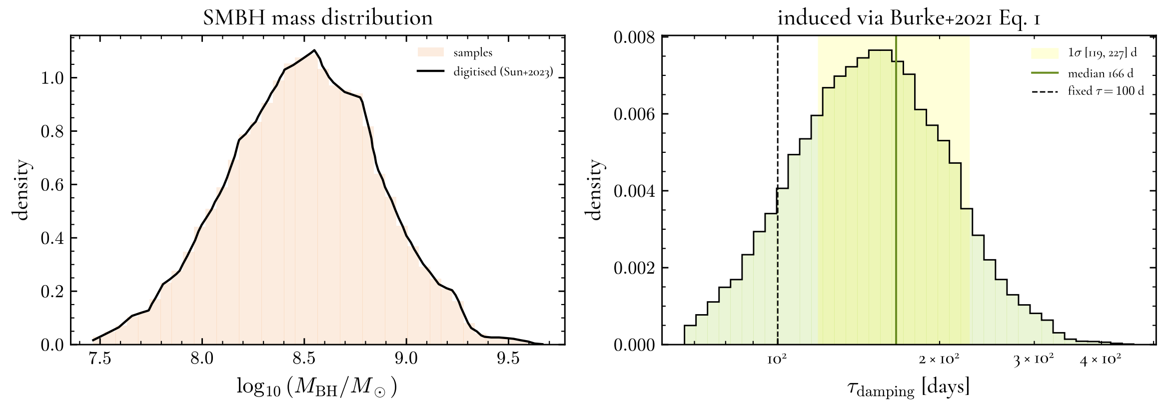


Figure 4.2: Left: the black hole mass distribution used to assign timescales, digitized from Sun 2023, with samples drawn from it by inverse-transform sampling. Right: the damping-timescale distribution induced by passing these masses through the mass-timescale relation of Equation 4.5, with median 166 d and the marked 1σ range $[119, 227]$ d. The fixed value $\tau = 100$ d used in the controlled experiment is shown for reference.

4.2 Physically Motivated Timescales

For the simulated light curves to resemble a real population, the damping timescale is not fixed at a single value but drawn from a distribution representative of observed AGN. As introduced in Chapter 3, τ increases with black hole mass, therefore timescale is assigned in two steps: first, a black hole mass is first drawn from the bias-corrected mass distribution of quasars of Sun (2023), and the corresponding damping timescale is then obtained from the mass-timescale relation of Burke et al. (2021),

$$\tau = 107^{+11}_{-12}\,\text{days}\left(\frac{M_{\rm BH}}{10^8\,M_\odot}\right)^{0.38^{+0.05}_{-0.04}}. \tag{4.5}$$

The mass distribution is digitized from the observed sample and sampled by *inverse-transform sampling*. This method draws from an arbitrary distribution using only its cumulative distribution function, first, the digitized histogram is numerically integrated (a cumulative sum over the bins) and normalized to give the cumulative distribution $F(\log M)$, which rises monotonically from zero to one, and a value is drawn by generating a uniform random number $u \in [0, 1)$ and inverting it, $\log M = F^{-1}(u)$. Because F is steep where the distribution is dense and shallow where it is sparse, uniform samples in u map to masses distributed according to the original histogram. The sampled masses are then passed through the mass-timescale relation of Equation 4.5, so the resulting timescales inherit the shape of the mass distribution, including its high-mass tail. The induced distribution has a median of $\tau \approx 166$ d, with the central 68% of curves falling between 119 and 227 d (Figure 4.2).

This procedure defines the varying-timescale configuration used for most of the experiments, in which each light curve is assigned its own τ and the reconstruction method must generalize across the range shown in Figure 4.2. A simpler configuration was also used, in which the timescale was held fixed at $\tau = 100$ d for all curves and served as a controlled setting in which the sampling and noise can be studied in isolation from any variation in τ. The varying-timescale case is the more realistic and more demanding test, and is the one against which the main results are reported.

4.3 Light Curve Generation

Light curves are generated with the EzTao package (Yu and Richards 2022), which draws damped random walk realizations as a Gaussian process and evaluates it through the *celerite* method (Foreman-Mackey et al. 2017). A direct evaluation of the Gaussian process likelihood requires inverting the covariance matrix, at a cost that scales as $\mathcal{O}(N^3)$ in the number of epochs and is prohibitive for long light curves; celerite reduces this to $\mathcal{O}(N)$ for the damped random walk kernel, which makes the simulation of large samples tractable. For each simulated object a damping timescale is assigned as described in Section 4.2, the stationary amplitude is set to $\sigma = 0.112$ mag, a value representative of quasar optical variability (Yu et al. 2025), and a densely sampled realization is produced to represent the true underlying signal.

The observed light curve is formed from this dense realization rather than interpolated from it. An irregular cadence is drawn beforehand, and the realization is evaluated on the union of the dense grid and the observed epochs in a single simulation, so that the magnitudes at the sampled times can be read off directly from the same realization that defines the true signal. These sampled magnitudes are then perturbed with photometric noise, giving for each light curve a set of observed epochs, magnitudes, and per-epoch uncertainties, $(t_{\text{obs}}, y_{\text{obs}}, \sigma_{\text{obs}})$. The observations and the underlying signal are therefore two views of one realization, which gives access to the true signal at every epoch and allows the reconstruction to be evaluated exactly. The noise is added as a Gaussian perturbation at each epoch, with a single-visit uncertainty of 0.01 mag, matching the photometric accuracy expected of the Legacy Survey of Space and Time (§ 3.3.2). An example of each of the two configurations is shown in Figure 4.3.

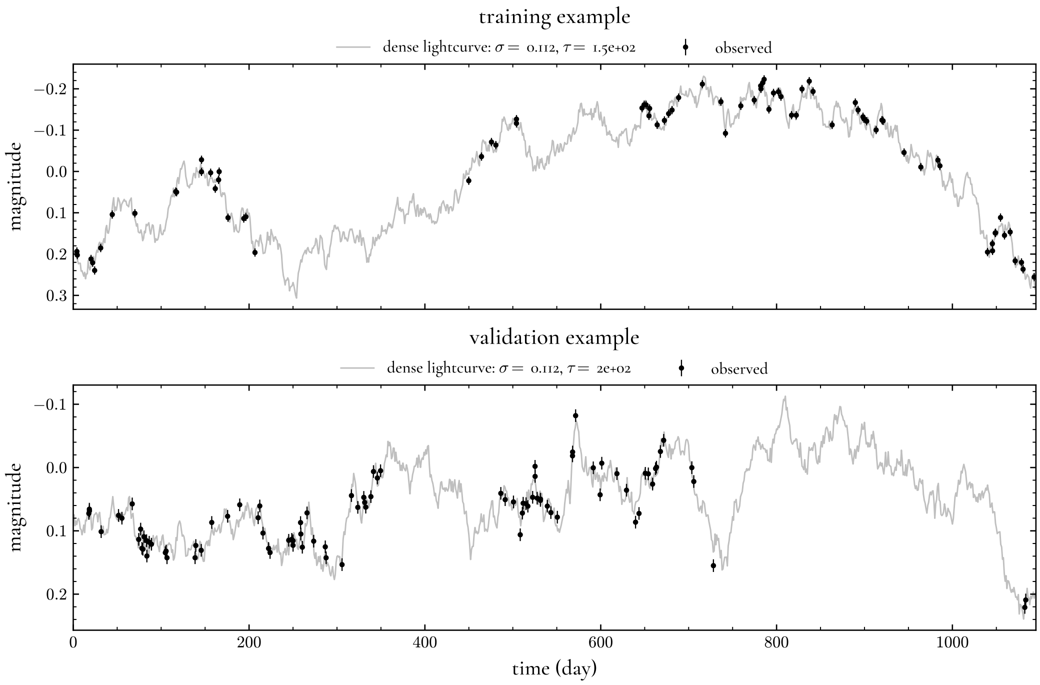


Figure 4.3: Two simulated light curves, one from the training set and one from the validation set, each showing the dense underlying realization (grey) and the sparse, noisy observations drawn from it (black).

The generation of the light curves is implemented in `LightCurves.py`, and the simulations used in the controlled experiments are produced from the notebook `Stage0.ipynb`.

4.4 Sampling Regime

The difficulty of gap filling is controlled by a single dimensionless quantity, the gap length in units of the damping timescale, $\Delta t/\tau$, since a damped random walk loses memory as $e^{-\Delta t/\tau}$. The sampling parameters are therefore quoted in units of τ rather than in days. With a three-year baseline ($t_{\max} = 1095$ d), 130 attempted epochs, and a gap fraction of 0.35 distributed over three windows, at the fiducial timescale $\tau = 100$ d the baseline spans $10.9\,\tau$, the mean spacing between consecutive epochs is 8.4 d $= 0.084\,\tau$, and the mean gap is 128 d $= 1.28\,\tau$. Consecutive epochs are therefore strongly correlated, while the gaps straddle the decorrelation scale, the regime in which reconstruction is neither trivial nor impossible. Figure 4.4 shows the resulting distributions of the consecutive spacing and of the largest gap in each curve, in units of τ.

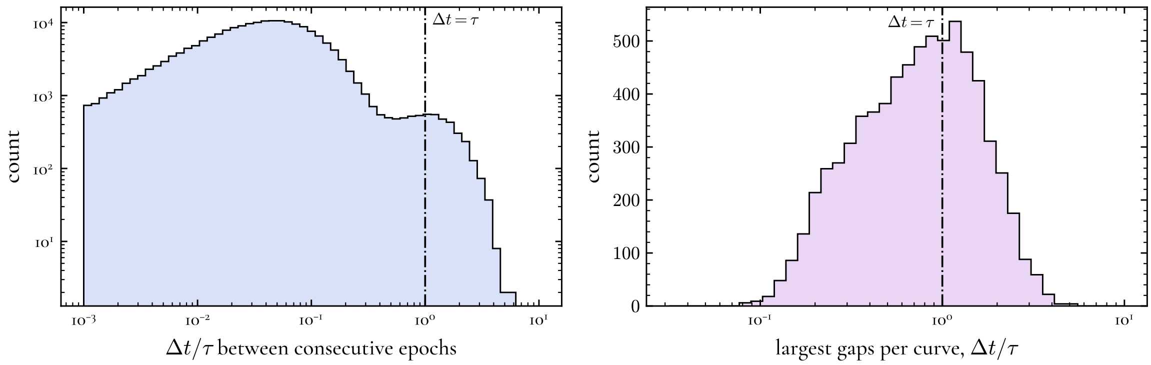


Figure 4.4: Left: the distribution of the spacing between consecutive epochs, in units of the damping timescale. Right: the distribution of the single largest gap in each light curve, in the same units. Most consecutive spacings lie well below τ, while the largest gaps straddle τ, the regime in which reconstruction is most demanding. The line marks $\Delta t = \tau$.

This regime is bounded by two thresholds that appear in the literature as biases on the recovery of τ. The timescale is underestimated when the baseline is shorter than roughly $10\,\tau$ (Kozłowski 2016a), and overestimated when it is shorter than about five times the mean cadence (Hu 2024). Here they mark the range over which a gap carries information at all: below the cadence threshold consecutive epochs are effectively uncorrelated and no interpolation is possible, so a model can only return the prior, while above the baseline threshold the signal barely decorrelates over the observing window and every gap is trivially interpolable. A fixed timescale of $\tau = 100$ d places the fiducial experiment near the centre of this range. When the timescale is instead drawn from the physically motivated distribution of Section 4.2, its bulk between roughly 100 and 230 d falls within the same informative range, while its tails extend toward the two thresholds, so that the varying-timescale experiment samples the full span over which reconstruction is meaningful.

A single photometric band is used throughout. Surveys such as ZTF and LSST observe in several bands, and the coordinated multi-band variability discussed in Chapter 3 is in principle a rich source of additional information, but the treatment here is confined to single-band light curves, with the extension to multiple bands left for future work.

5 Neural Stochastic Flows

As motivated in § 3.1.1, the reconstruction of a gappy light curve is underdetermined, and is best approached by returning a distribution of signals consistent with the data rather than a single curve. The damped random walk introduced in Chapter 4 already provides such a probabilistic description because it is a Gaussian process and the transition between any two epochs has a known analytic form. The method adopted in this thesis replaces that transition with a flexible one learned directly from data, using neural stochastic flows (NSF; Kiyohara et al. 2025). The motivation is twofold, flexibility and efficiency. A learned transition can in principle capture departures from the damped random walk that a fixed kernel cannot, and, more generally, neural stochastic flows are *solver-free*: they model the transition density of a stochastic differential equation directly, without the numerical integration that conventional neural SDEs require. In the damped random walk case considered here the transition is analytically known, which provides an exact reference against which the learned model can be validated (see Chapter 6).

5.1 Normalizing Flows

A *normalizing flow* (NF) transforms a simple probability distribution, typically a standard normal $z \sim p_Z(z)$, into a more complex one through a sequence of invertible and differentiable mappings (Rezende et al. 2016; Kobyzev et al. 2021). A sample is generated by passing the base variable through the map, $y = g_\theta(z)$, the *generative direction*. Because the map is invertible, one can also go back from y to z through $z = f_\theta(y) = g_\theta^{-1}(y)$, the *normalizing direction*, and the density of y follows from the change-of-variables formula,

$$p_Y(y) = p_Z\big(f_\theta(y)\big)\,\big|\det J_{f_\theta}(y)\big|, \tag{5.1}$$

where J_{f_θ} is the Jacobian of the inverse map. The base density p_Z is known and the Jacobian term, the *volume correction*, accounts for the change of volume induced by the transformation, so Eq. 5.1 yields an exact density for y (Figure 5.1). This exactness, together with the ability both to sample and to evaluate densities, is what makes normalizing flows well suited to modelling complex distributions, and it means the model can be trained simply by maximising the likelihood of the data under Eq. 5.1 (Kobyzev et al. 2021).

For Eq. 5.1 to be practical, the transformation must satisfy two requirements: it must be easily invertible, and its Jacobian determinant must be cheap to compute (Kobyzev et al. 2021). A construction that meets both, and the one on which the model used here is built, is the *coupling flow*

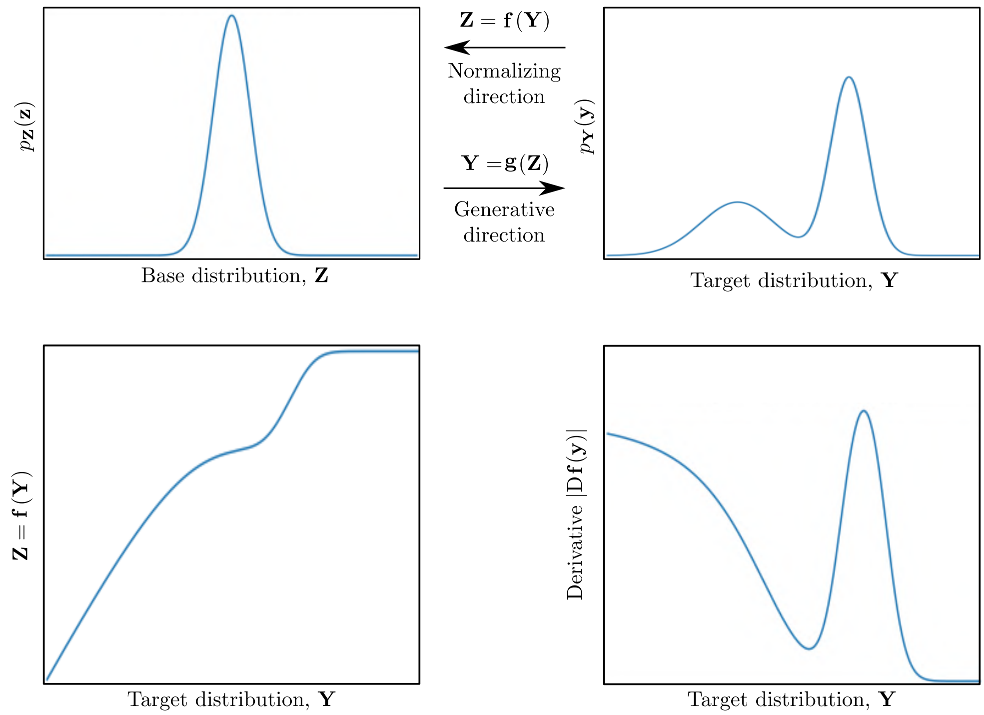


Figure 5.1: The change of variables of Eq. 5.1. Top left: the base density p_Z. Top right: the target density p_Y. A bijective map g transforms one into the other, with inverse f. Bottom left: the inverse map f. Bottom right: its absolute Jacobian, which rescales the density so that probability is conserved. Adapted from Kobyzev et al. 2021.

(Dinh, Krueger, and Bengio 2015). The input is split into two partitions; one is left unchanged, and the other is transformed by a bijection whose parameters are produced by a *conditioner* network taking the unchanged partition as input,

$$y_{\mathrm{A}} = x_{\mathrm{A}}, \qquad y_{\mathrm{B}} = x_{\mathrm{B}} \odot \exp\big(s(x_{\mathrm{A}})\big) + t(x_{\mathrm{A}}), \tag{5.2}$$

where s and t are the scale and shift functions and $\odot$ is the elementwise product. This is the affine coupling layer of RealNVP (Dinh et al. 2017). It is trivially inverted, since the unchanged partition $x_{\mathrm{A}} = y_{\mathrm{A}}$ recovers the arguments of s and t, and its Jacobian is triangular, so its determinant is simply the product of the diagonal, $\prod_j \exp(s(x_{\mathrm{A}}))_j$. Because neither the inverse nor the determinant requires inverting or differentiating s and t, these functions can be arbitrarily complex neural networks. A single such layer leaves one partition untouched, so the partitions are alternated between successive layers, and stacking many layers yields an expressive invertible map.

5.2 From Static Densities to Transition Laws

Neural stochastic flows extend this construction from a single *static* density to the *transition law* of a stochastic process (Kiyohara et al. 2025). Rather than modelling one distribution $p_Y(y)$, they model the conditional distribution of a future state given a current state and an elapsed time,

$$p_\theta\big(x_t \mid x_s, \Delta t\big), \qquad \Delta t = t - s, \tag{5.3}$$

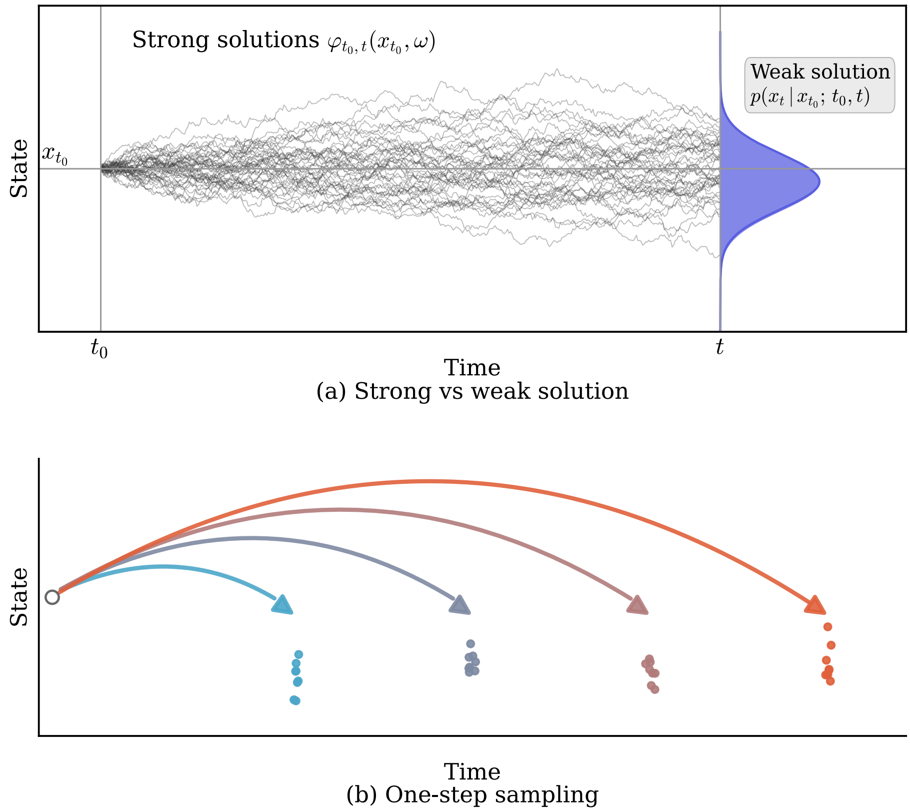


Figure 5.2: (a) The *strong solution* of a stochastic differential equation is an ensemble of paths $\varphi_{t_0,t}(x_{t_0},\omega)$ diverging from a common initial state; the *weak solution* is the transition density $p(x_t \mid x_{t_0}; t_0, t)$ they define at a later time, shown at the right edge. Neural stochastic flows model this density directly. (b) Because the density is modelled directly, samples at any elapsed time can be drawn in a single step, without the iterative integration that path-based methods require. Adapted from Kiyohara et al. 2025.

that is, the distribution of the state x_t at time t given that the process was at x_s at the earlier time s. This conditional is defined in the same way as an ordinary flow, but with the transformation now conditioned on the starting state and the elapsed time: a Gaussian noise variable $\epsilon \sim \mathcal{N}(0, I)$ is passed through a conditional normalizing flow,

$$x_t = F_\theta\big(x_s, \Delta t, \epsilon\big). \tag{5.4}$$

For fixed x_s and Δt, different draws of ϵ produce different possible future states x_t, and together they define the conditional density $p_\theta(x_t \mid x_s, \Delta t)$ of Eq. 5.3. Modelling this density directly, rather than the individual paths of the process, is what allows a state at any elapsed time to be sampled in a single step, without iterative integration (Figure 5.2).

Concretely, the sample is drawn by first initializing a state-dependent Gaussian that mirrors an Euler-Maruyama step, with a mean displaced from x_s by a drift scaled by Δt and a variance scaled by $\sqrt{\Delta t}$,

$$z = \underbrace{x_s + \Delta t\, \mathrm{MLP}_\mu(c)}_{\mu(c)} + \underbrace{\sqrt{\Delta t}\, \mathrm{MLP}_\sigma(c)}_{\sigma(c)} \odot \epsilon, \qquad c := (x_s, \Delta t), \tag{5.5}$$

and then passing it through a sequence of coupling layers of the form of Eq. 5.2, whose scale and shift functions are now additionally conditioned on c and multiplied by Δt. The map F_θ

is constrained in this way so that the transition density inherits the defining properties of the solution of a stochastic differential equation (Kiyohara et al. 2025). Two are relevant here. The *identity* property requires that a zero time gap returns the input state, $p_\theta(x_t \mid x_s, 0) = \delta(x_t - x_s)$; the explicit factors of Δt in Equations 5.5 and 5.2 enforce this, since both the Gaussian and every coupling layer collapse to the identity as $\Delta t \to 0$. The *stationarity* property requires that, for a time-homogeneous process, the transition depend only on the gap Δt and not on absolute time; since the damped random walk is autonomous, the conditioner omits absolute time and depends on Δt alone. A third property, discussed in the following section, ties transitions of different lengths together.

This learned transition plays the role of the damped random walk step in place of the analytic transition of the Ornstein-Uhlenbeck process (Chapter 4), the flexible map F_θ models the variability between any two epochs, while reducing to the correct transition when the data are generated by a damped random walk.

5.3 Triplets and Stochastic Bridges

Reconstructing a gap requires the distribution of a state that lies *between* two observations, not merely one that follows from a single earlier state. This intermediate conditional is modelled by a *stochastic bridge*. Following the training strategy of Kiyohara et al. (2025), from each light curve one randomly samples triplets of times (t_i, t_j, t_k) with $t_i < t_j < t_k$, drawn as described in §5.7, and two models are trained jointly on them. The *flow* model learns the transitions between states, both the direct endpoint transition $p_\theta(x_k \mid x_i, \Delta t_{ik})$ and the two one-step transitions $p_\theta(x_j \mid x_i, \Delta t_{ij})$ and $p_\theta(x_k \mid x_j, \Delta t_{jk})$. The *bridge* model learns the intermediate conditional[1]

$$q_\xi\big(x_j \mid x_i, x_k, t_i, t_j, t_k\big), \tag{5.6}$$

the distribution of the middle state given both surrounding states and their times. It is this bridge, conditioned on the two bracketing observations, that performs the gap filling at inference.

The two models are tied together by the third inherited property, the *flow property*, which requires that the transition density satisfy the Chapman-Kolmogorov relation: the direct transition from x_i to x_k must equal the composition of the two one-step transitions through any intermediate state,

$$p_\theta\big(x_k \mid x_i, \Delta t_{ik}\big) = \int p_\theta\big(x_k \mid x_j, \Delta t_{jk}\big)\, p_\theta\big(x_j \mid x_i, \Delta t_{ij}\big)\, \mathrm{d}x_j. \tag{5.7}$$

Equivalently, the joint distribution of the pair (x_j, x_k) given x_i can be factorized either as a one-step flow from x_i to x_k followed by the bridge that places x_j between them, or as a two-step flow $x_i \to x_j \to x_k$; both describe the same distribution and must agree. Enforcing this relation is what forces the flow and the bridge to represent a single, self-consistent stochastic process rather than two unrelated conditionals.

The two sides of Eq. 5.7 cannot be compared directly, since the two-step side requires marginalizing over the intermediate state x_j. The bridge q_ξ is therefore introduced as an auxiliary variational distribution, which makes the comparison tractable through variational upper bounds on the Kullback-Leibler divergence between the one-step and two-step transitions, taken in both

[1] The bridge is denoted b_ξ in the original formulation of Kiyohara et al. (2025); it is written q_ξ here to emphasize its role as an auxiliary variational distribution and for consistency with the notation used throughout this work.

directions since the divergence is asymmetric. The forward bound, from the one-step to the two-step transition, is

$$\mathcal{L}_{\text{1-to-2}} = \mathbb{E}\left[\log \frac{p_\theta(x_k \mid x_i)\, q_\xi(x_j \mid x_i, x_k)}{p_\theta(x_k \mid x_j)\, p_\theta(x_j \mid x_i)}\right], \tag{5.8}$$

with the expectation over $p_\theta(x_k \mid x_i)\, q_\xi(x_j \mid x_i, x_k)$, and the reverse bound is

$$\mathcal{L}_{\text{2-to-1}} = \mathbb{E}\left[\log \frac{p_\theta(x_j \mid x_i)\, p_\theta(x_k \mid x_j)}{q_\xi(x_j \mid x_i, x_k)\, p_\theta(x_k \mid x_i)}\right], \tag{5.9}$$

with the expectation over $p_\theta(x_j \mid x_i)\, p_\theta(x_k \mid x_j)$. Their sum is the flow-consistency loss, $\mathcal{L}_{\text{flow}} = \mathcal{L}_{\text{1-to-2}} + \mathcal{L}_{\text{2-to-1}}$. The explicit form of these bounds, as a difference of log-densities evaluated on Monte Carlo samples for a single triplet, is given in Appendix B, together with the architecture of the bridge model.

5.4 Training Objective

The model is trained by minimising a loss with two terms (Kiyohara et al. 2025),

$$\mathcal{L}(\theta, \xi) = \mathcal{L}_{\text{data}}(\theta) + \lambda\, \mathcal{L}_{\text{flow}}(\theta, \xi), \tag{5.10}$$

where $\mathcal{L}_{\text{data}}$ is the negative log-likelihood of the observed transitions under the flow, evaluated through the change-of-variables formula of Eq. 5.1, and $\mathcal{L}_{\text{flow}}$ is the flow-consistency term of Equations 5.8 and 5.9, weighted by a hyperparameter λ. For each triplet the data term is evaluated over all three pairs of states, the two one-step transitions $x_i \to x_j$ and $x_j \to x_k$ and the direct endpoint transition $x_i \to x_k$, so that the long transition across the full span of the triplet is trained directly on the data rather than only reproduced by composing shorter ones. This is implemented in `nllTriplet` (in `NSF.py`). The data term anchors the model to the observed light curves, while the consistency term enforces that the learned transitions compose correctly across time. Both terms are necessary: the consistency term alone admits self-consistent but arbitrary transitions, and it is the data likelihood that ties the learned process to the damped random walk actually present in the light curves. The balance between the two terms is set by λ, but the results (available in `Stage1Tests.ipynb`) are insensitive to its precise value; varying λ over more than an order of magnitude, from 0.03 to 1.0, changes the coverage on the low-noise validation set by at most a few per cent, from 0.62 to 0.60, and leaves the RMSE unchanged at 0.18. A value of $\lambda = 0.1$ is used throughout.

5.5 Reconstruction

Once trained, the bridge performs the gap filling. Reconstructing a gap means conditioning on the observations at *both* of its ends, which is precisely what the bridge does. The flux is reconstructed on a dense grid between each pair of consecutive observations. For a grid point lying between two observations, the bracketing observations provide the conditioning states x_i and x_k and their times, and the bridge q_ξ of Eq. 5.6 gives the distribution of the flux at the intermediate time. Because this distribution can be sampled in a single pass, the reconstruction is built by drawing a fixed number of samples from the bridge at each of a set of evenly spaced grid points between every consecutive pair of observations, as illustrated in Figure 5.3. This is carried out by `fillGaps` (in `GapFill.py`), which returns the full set of Monte Carlo realizations. The reconstruction is summarized by the pointwise median of these samples, taken as the central estimate, and by the interval between the 16th and 84th percentiles, which gives the central 68% predictive band. At the observations themselves the reconstruction is pinned to the measured values, so the band closes to zero width there and widens across the interior of an interval, where the elapsed time to the nearest observation is longer and the transition is correspondingly more uncertain. It is this predictive distribution, rather than any single reconstructed curve, that the evaluation of Chapter 6 assesses.

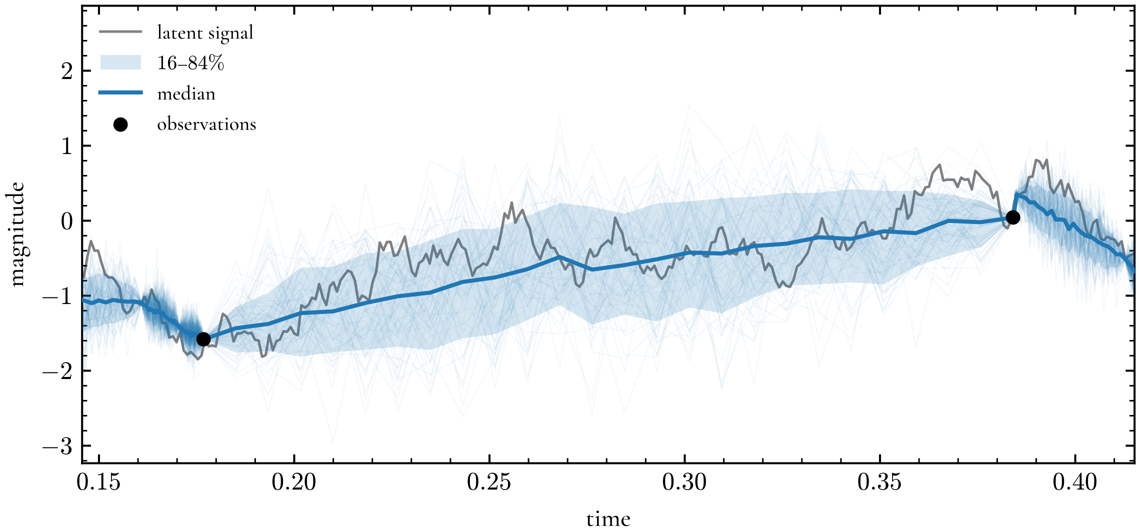


Figure 5.3: Monte Carlo reconstructions of a single gap in a varying-timescale validation curve. Each thin line is one sample drawn from the bridge conditioned on the two bracketing observations (black); their pointwise median (blue) and central 68% interval (shaded) summarize the ensemble. The samples fan out across the interior of the gap and converge at the endpoints, and together they bracket the latent signal (grey).

The flow and bridge are conditioned on the flux values and times alone: the per-epoch measurement uncertainties enter neither their conditioning, nor the training objective, nor the sampling at inference. The predictive intervals of the reconstruction therefore express the variance of the learned process and carry no information about the measurement error of the observations on which they are conditioned. The Gaussian process baseline, by contrast, includes the per-epoch uncertainties directly in its covariance, and it is this difference that underlies the comparison of their calibration in Chapter 6.

5.6 Evaluation

The reconstructions are assessed by three quantities, computed on points held out from each light curve. The *coverage* is the fraction of held-out points that fall within a predictive interval of a given nominal probability. A well-calibrated model reproduces the nominal level, so that a central 68% interval contains the truth 68% of the time, and a coverage below the nominal level indicates intervals that are too narrow, that is, overconfident. The *root-mean-square error* (RMSE) measures the accuracy of the point reconstruction, taken as the posterior median, against the truth. The *predictive log-likelihood* is the mean Gaussian log-density assigned to the held-out points by the predictive distribution, and it combines the other two. Unlike coverage, it cannot be improved by widening the intervals, since intervals that are too wide are penalized as heavily as those that are too narrow. Coverage and the RMSE therefore isolate the two ways a reconstruction can fail, in its uncertainty and in its accuracy, while the predictive log-likelihood captures both together.

All three quantities are computed on the interiors of the gaps, excluding the observation epochs themselves. At an observation the reconstruction is pinned to the measured value and its interval collapses to zero width, so both the reconstruction and the baseline match the data there trivially; including these points would inflate the coverage and deflate the error for every method alike and obscure the performance across the gaps, where reconstruction is actually tested. The scoring is carried out per curve by `evalCurve` (in `GapFill.py`).

These quantities are used throughout Chapter 6. What the reconstruction is scored *against* depends on the experiment. On simulated data the held-out points are compared with the known underlying signal, while on real data, where no underlying signal is available, they are compared with withheld observations.

5.7 Implementation

The model is implemented in JAX (Bradbury et al. 2018), building on the publicly available `jax_nsf` codebase accompanying Kiyohara et al. (2025), from which the flow and bridge architectures and the core training machinery are taken. The change-of-variables likelihood of Eq. 5.1 and the variational bounds of Equations 5.8 and 5.9 are differentiated automatically, so the gradients of the loss are obtained straightforwardly, and the full training step is compiled once and reused across iterations, which makes the repeated evaluation of the loss over many triplets efficient. The flow and bridge are represented as `equinox` modules (Kidger and Garcia 2021), and only their floating-point array leaves are treated as trainable parameters, with the remaining static structure held fixed during optimisation.

Both models use four affine-coupling layers, and the multilayer perceptrons that produce the base-distribution parameters and the coupling scale and shift have a width of 64 and a depth of two, with $\tanh$ activations. The flow and bridge are conditioned only on elapsed times and never on absolute time, matching the stationarity of the damped random walk. They are constructed by `makeModels` (in `NSF.py`).

Before training, the light curves are normalized so that times are expressed in units of the observing baseline and magnitudes in units of the variability amplitude, carried out by `loadDataset` (in `Tensorize.py`). This step is necessary rather than cosmetic. Because the flow conditions jointly on a state and an elapsed time, presenting it with raw values, times of order a thousand days against magnitudes of order a tenth, prevents it from learning the transition at all.

Training minimizes the objective of Eq. 5.10 using the Adam optimizer (Kingma and Ba 2017) as implemented in `optax` (DeepMind et al. 2020), with a learning rate of 2×10^{-3}, and is carried out by `train` (in `NSF.py`). At each step a batch of sixteen light curves is drawn, and from each a set of thirty-two triplets (t_i, t_j, t_k) is sampled by `sampleTriplets` (in `Tensorize.py`), with the span $t_k - t_i$ drawn log-uniformly and the intermediate time t_j placed uniformly between the endpoints. Drawing the span this way ensures that both short and long endpoint separations are represented, so that the direct transition $p_\theta(x_k \mid x_i, \Delta t_{ik})$ is trained on wide gaps as well as narrow ones. The loss is averaged over the batch and its gradient taken through the compiled step. The zero-padding introduced when the ragged light curves are stored as fixed-size arrays (Chapter

4) is excluded from the loss through a mask, so that only genuine observations contribute.

The simulation of the light curves and the training of the model are carried out in two separate environments. The simulator relies on `EzTao` (Yu and Richards 2022), whose dependencies conflict with those of `jax_nsf`, so the two cannot share an environment. The simulated light curves are written to disk and read back for training, with the saved dataset acting as the only interface between the two stages. This separation keeps the generation of the training set reproducible and independent of the model, so a dataset can be regenerated or reused without retraining.

The contributions of this thesis are in the surrounding framework rather than in the flow architecture itself: the simulation of damped random walk light curves with physically motivated parameters (Chapter 4), the construction of a training set matched to the sampling and noise of a real survey (Chapter 6), and the evaluation of the learned model against the analytic transition and a Gaussian process baseline.

6 Results

This chapter reports the performance of the reconstruction method. It proceeds from the most controlled setting to the most realistic. A fixed-timescale, low noise experiment first establishes that the method learns the analytic damped random walk transition and reconstructs light curves close to the theoretical, optimal case. A varying-timescale experiment then introduces a distribution of damping timescales. The sampling regime of the Zwicky Transient Facility (§3.3.1) is then measured and reproduced, and it is in this regime, with higher and epoch-dependent noise, that a limitation appears in the way the model treats measurement error. The method is finally applied to real ZTF light curves through a masking experiment.

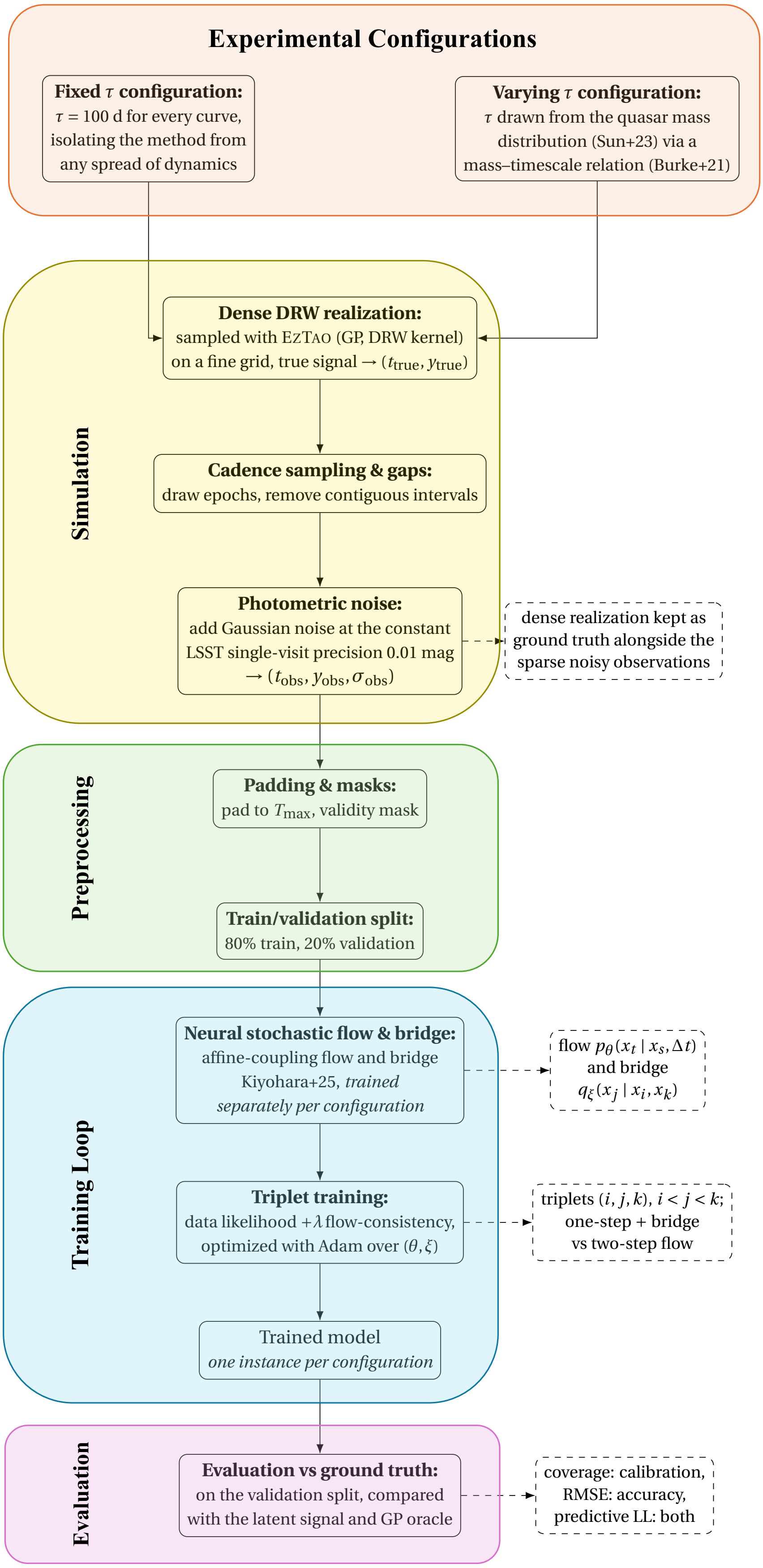
6.1 Experimental Setup

The experiments are summarized in Table 6.1. The fixed-timescale experiment holds $\tau = 100$ d for every curve, with a constant photometric uncertainty of 0.01 mag against a variability amplitude of $\sigma = 0.112$ mag, a uniform ratio $y_{\mathrm{err}}/\sigma = 0.089$. The varying-timescale experiment instead draws τ from the distribution of Section 4.2 at the same low noise, the ZTF-matched experiment replaces the sampling and noise with those of the survey, and the final experiment applies the method to the real ZTF light curves. Each simulated configuration produces 2400 light curves for training and 600 for validation, an 80/20 split.

Table 6.1: The four experiments, in increasing order of realism. Each adds one element of a real survey to the previous: a distribution of timescales, then survey-level sampling and noise, then real data. The reference against which each is scored changes accordingly, from the Bayes-optimal Gaussian process on simulated data to a fitted Gaussian process on real data.

Experiment	Timescale	Noise	Ground truth	Reference	Section
Fixed τ	$\tau = 100$ d	low, constant	latent signal	GP oracle	6.2.1
Varying τ	mass-prior	low, constant	latent signal	GP oracle	6.2.2
Simulated ZTF	mass-prior	ZTF, per-epoch	latent signal	GP oracle	6.4
Real ZTF	unknown	ZTF, real	none	fitted GP, interp.	6.5

The two experimental pipelines, for the simulated and the ZTF-matched experiments, are summarized in Figures 6.1 and 6.7.



3

Figure 6.1: Pipeline of the idealized simulation experiments.

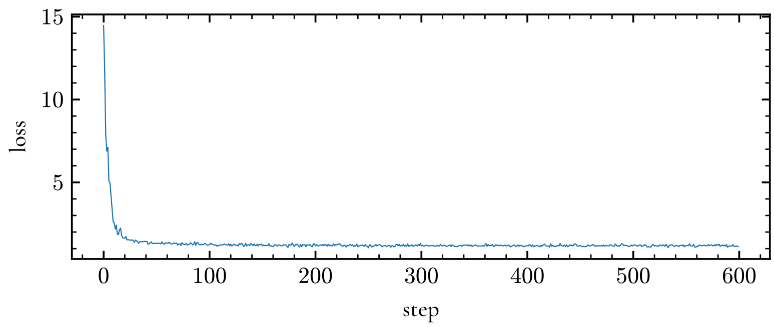
6.2 Low Noise Simulations

On the simulated data the reconstruction is measured against a Gaussian process with the damped random walk kernel, given the true parameters of each curve. As shown in Section 4.1.1, conditioning the damped random walk on the observations yields a Gaussian predictive distribution in closed form. Supplied with the true timescale and amplitude, this is the Bayes-optimal predictor, since no method conditioned on the same observations can achieve a lower expected error, and it therefore sets the ceiling against which the reconstruction is measured.

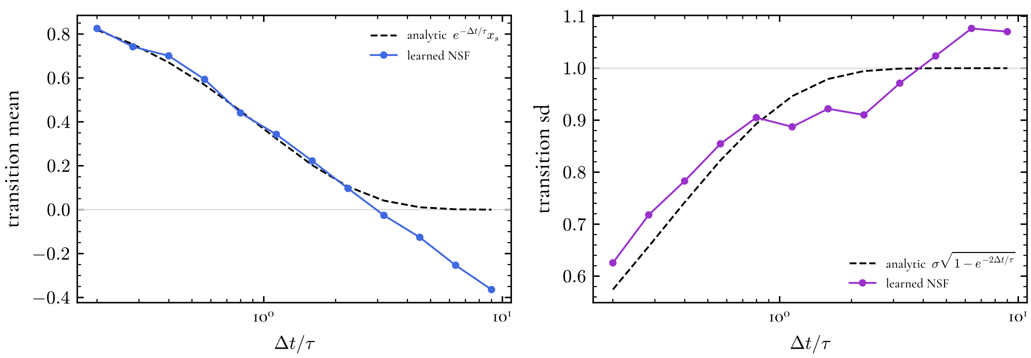
6.2.1 Fixed Timescale

The most controlled experiment fixes the damping timescale at $\tau = 100$ d for every light curve, removing any variation in dynamics so that the behaviour of the method can be studied in isolation. The training objective converges within a few hundred steps (Figure 6.2a).

The first quantity examined is the learned transition itself, independently of any reconstruction. The trained flow defines a transition density $p_\theta(x_{t+\Delta t} \mid x_t, \Delta t)$, and because the true transition of the damped random walk is known in closed form, the two can be compared directly. The comparison is constructed as follows. A single starting state x_s is fixed, and a set of elapsed times Δt is chosen spanning from a small fraction of the damping timescale to several times it. For each elapsed time, a large sample, here four thousand draws, is taken from the flow conditioned on x_s and that elapsed time, and the empirical mean and standard deviation of the samples are formed.



(a) Training loss as a function of optimization step.



(b) Mean (left) and standard deviation (right) of the learned transition, in units of the damping timescale, against the analytic damped random walk (dashed).

Figure 6.2: Training and validation of the fixed-timescale model.

These are compared with the analytic mean $e^{-\Delta t/\tau}x_s$ and standard deviation $\sqrt{\sigma^2(1-e^{-2\Delta t/\tau})}$ of Equation 4.4, evaluated at the same elapsed times. Repeating this over the range of Δt traces out the learned mean and standard deviation as functions of the elapsed time, which is expressed in units of the damping timescale since it is this ratio that governs the transition. The procedure is carried out by `checkTransition` (in `NSF.py`), and all quantities are computed in the normalized units in which the model is trained. Figure 6.2b shows the result. The learned mean relaxes towards zero and the learned standard deviation grows towards its stationary value, both following the analytic curves across the whole range, so the flow has recovered the transition law of the process it was trained on.

Applied to full light curves, the method reconstructs the signal on the dense grid between the observations, drawing two hundred samples from the bridge at twelve grid points between each consecutive pair, and returning a posterior median and a predictive interval at every point. An example is shown in Figure 6.3. The reconstruction follows the underlying signal through the gaps, and the interval widens across the wider gaps as the bracketing observations become less informative. The accuracy of the reconstruction is quantified by the error against the true dense realization as a function of gap width, in units of the damping timescale (Figure 6.4). The error grows smoothly with the gap, and tracks that of the Gaussian process across the whole range, converging with it at the widest gaps, where the observations no longer constrain the signal and the correct prediction for both methods is the stationary prior. The difficulty of the reconstruction is therefore set by the size of the gap relative to the correlation length, as expected when the underlying process is recovered correctly.

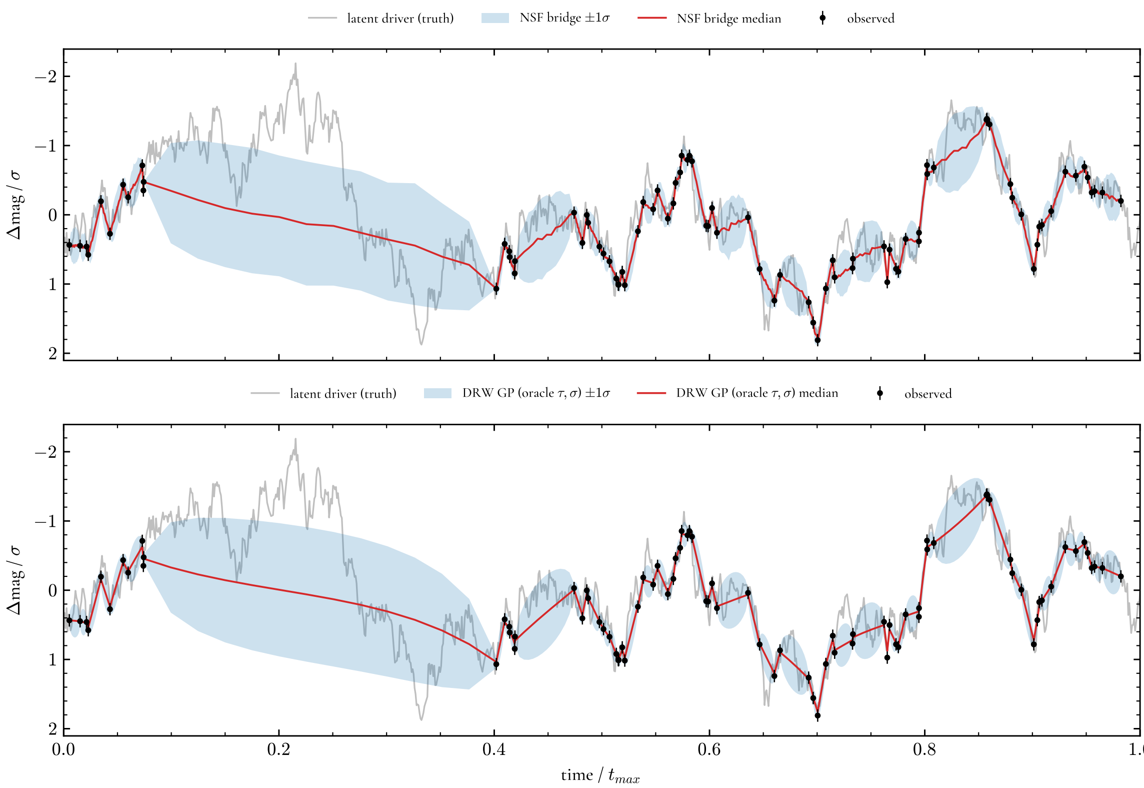


Figure 6.3: Reconstruction of a light curve from the fixed-timescale experiment, showing the posterior median (blue) and the central 68% interval (shaded) against the true dense realization (grey). The observations are shown in black, and the interval widens across the gaps between them.

This is the setting in which the reconstruction and the Gaussian process agree most closely across the experiments, as expected in the absence of any spread of timescales or survey-level noise, and it serves as the reference against which the effect of each added element of realism is measured in the sections that follow.

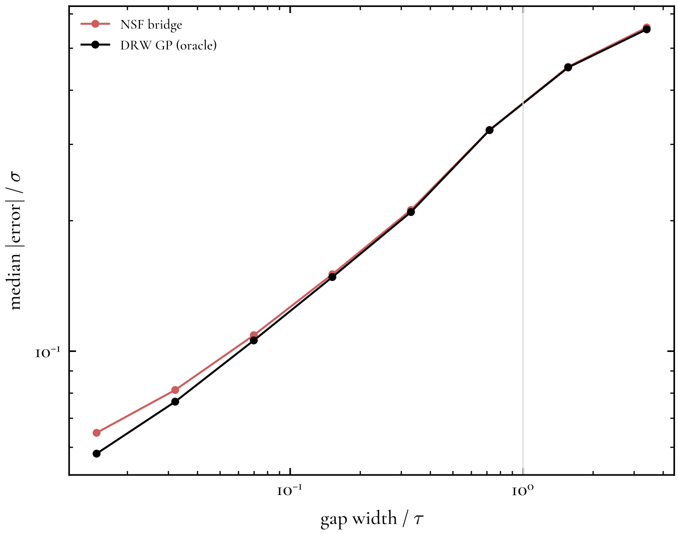


Figure 6.4: Reconstruction error as a function of gap width in units of the damping timescale, compared with the Gaussian process. The two track each other and converge at the widest gaps, where the correct prediction is the prior.

6.2.2 Varying Timescale

The timescale is next drawn from the physically motivated distribution of Section 4.2, so that each light curve carries its own damping timescale and the flow is trained on a mixture of dynamics rather than a single one. This is the more demanding and more realistic test, since the model must now represent a family of transitions rather than a single one, with the timescale of each curve inferred implicitly from its observations.

Because the transition now depends on the timescale of the individual curve, there is no longer a single learned transition to compare against the analytic form. It is instead evaluated at the median timescale of the distribution, using the same procedure as before with the damping timescale set to the median value. The learned transition remains close to the analytic one over most of the range, departing only at the longest separations, where the mean reaches -0.30 at nine damping timescales rather than relaxing to zero, and the standard deviation settles near 1.11 rather than unity (Figure 6.5). This residual is small and appears at the separations that are least represented in the training data, since gaps spanning many damping timescales are rare even when the span of the sampled triplets is drawn log-uniformly. The same slight overshoot of the mean past zero and inflation of the standard deviation reappear, greatly amplified, once survey-level noise is introduced in the ZTF experiments, where they become the signature of the model's difficulty with noisy observations.

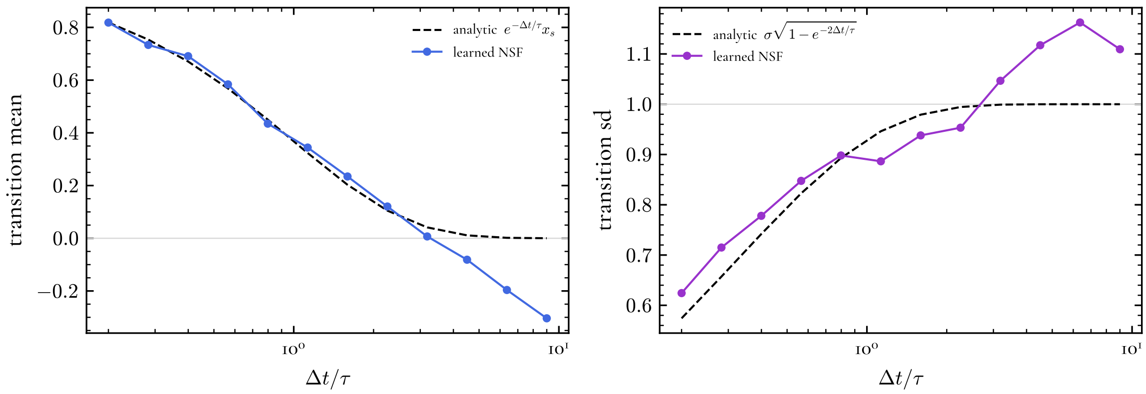


Figure 6.5: The mean (left) and standard deviation (right) of the learned transition, evaluated at the median damping timescale, compared with the analytic damped random walk (dashed). A small residual bias remains at the longest separations.

In reconstruction the method stays close to the Bayes-optimal reference across the whole distribution of timescales. Each validation curve is reconstructed in the same way and scored on its gap interiors against its own true signal, with the Gaussian process baseline given that curve's true parameters, and the results are aggregated over the validation set. The method reaches a median coverage of 0.61 at the 68% level, with a central range of $[0.55, 0.67]$ across curves, against 0.70 $[0.66, 0.73]$ for the Gaussian process, and a median RMSE of 0.172 $[0.151, 0.204]$ against 0.167 $[0.144, 0.200]$, placing the reconstruction about 3% above the optimal reference in error. The coverage varies systematically with the timescale of the curve. Splitting the validation set at the median timescale, the curves with shorter timescales reach a median coverage of 0.57 and those with longer timescales 0.64. This follows from the sampling. A short timescale decorrelates faster, so for a fixed cadence the gaps span more correlation lengths and the reconstruction is harder, whereas a long timescale keeps neighbouring epochs correlated and the reconstruction is better constrained.

Figure 6.6 compares the reconstruction with the Gaussian process on the same light curve. The two are closely matched where the observations are dense, and across the wider gaps both widen into broad intervals that follow the same shape, with the reconstruction and the oracle assigning almost the same uncertainty to the unobserved stretches. The individual Monte Carlo samples that make up such a reconstruction were shown earlier in Figure 5.3.

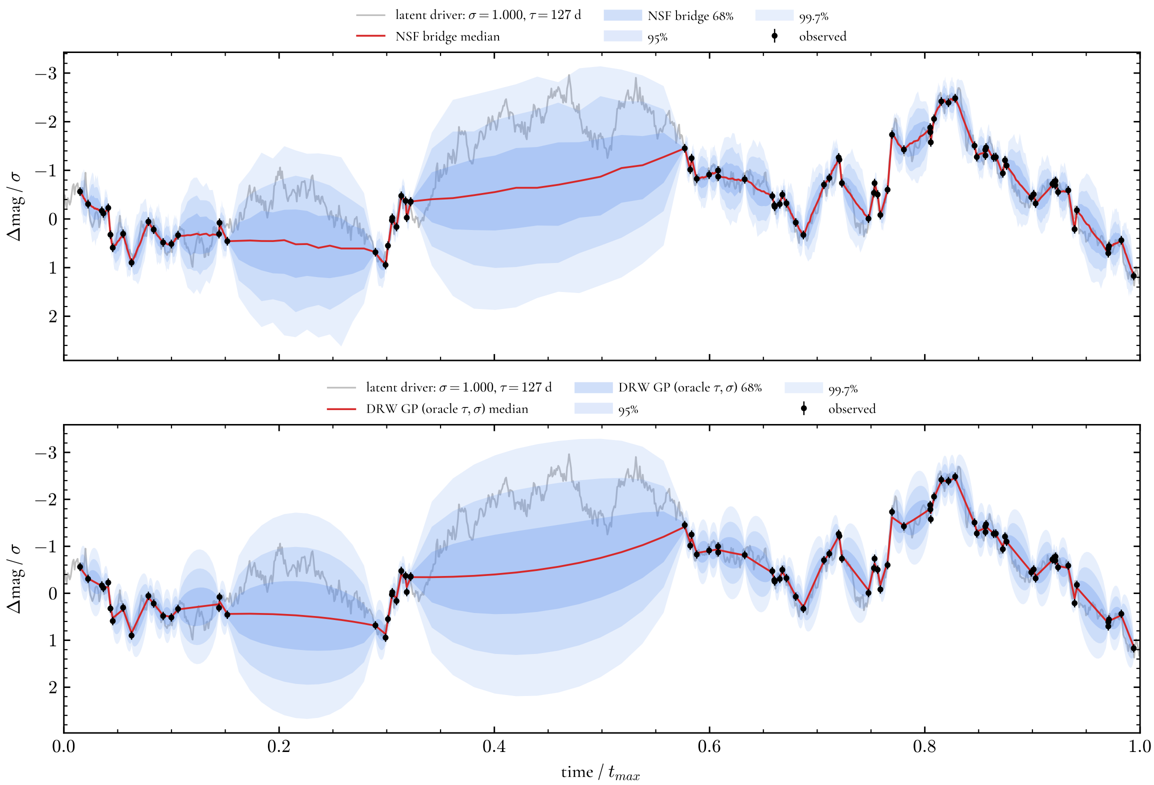


Figure 6.6: Reconstruction of a varying-timescale validation curve by the neural stochastic flow (top) and the Gaussian process given the true parameters (bottom), shown against the true dense realization (grey), with the observations in black. Each shows the posterior median (red) and the nested central 68, 95, and 99.7% intervals (shaded). The two assign closely matching uncertainty across the gaps.

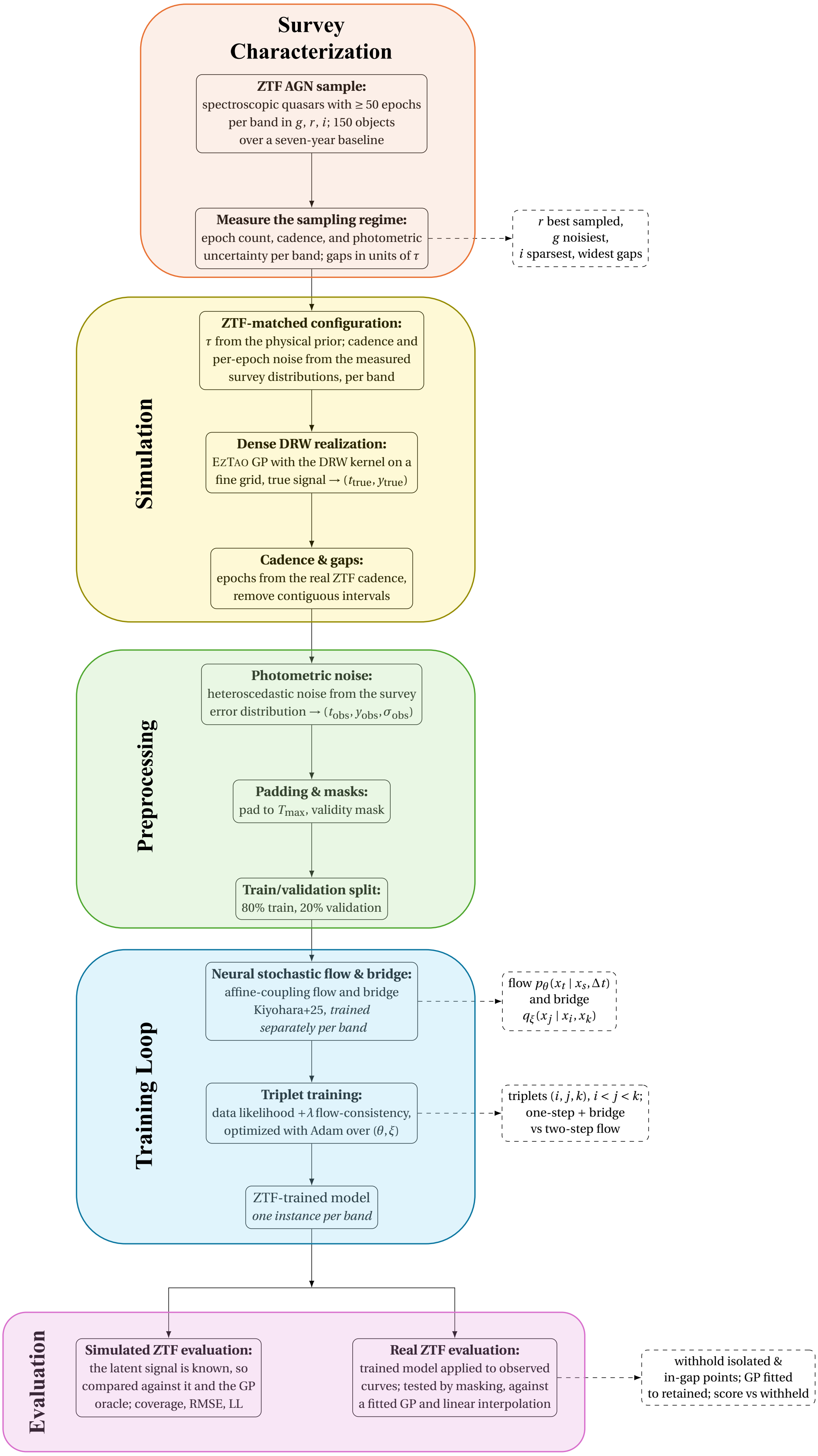


Figure 6.7: Pipeline of the ZTF experiments.

6.3 ZTF Experiments

The sampling regime that the simulations must reproduce is measured from a sample of real AGN, assembled and characterized in the notebook `ZTF_Stage0.ipynb`. The sample is built from the spectroscopically confirmed quasars of the Sloan Digital Sky Survey (York et al. 2000). Objects are selected from the SDSS spectroscopic catalogue as those classified as quasars with a redshift below $z = 0.5$, which keeps the sources bright enough to be well measured by ZTF.

For each selected quasar, its light curve is retrieved from ZTF by a cone search about its SDSS position, a query that returns all ZTF detections within a small angular radius (3 arcseconds) of the given coordinates, restricted to detections that pass the survey's photometric quality flags. The returned detections are separated by band into g, r, and i. Within each band, observations separated by less than 0.3 days are then treated as duplicates and only the first is kept, since on the timescales of interest such closely spaced points are almost perfectly correlated under the damped random walk and add little information while distorting the cadence statistics. A quasar is retained in the sample only if at least 50 epochs remain in each of the three bands, and the sample is grown in this way until it reaches 150 AGN, spanning a baseline of roughly seven years. An example is shown in Figure 6.8, with the irregular cadence and seasonal gaps characteristic of a ground-based survey.

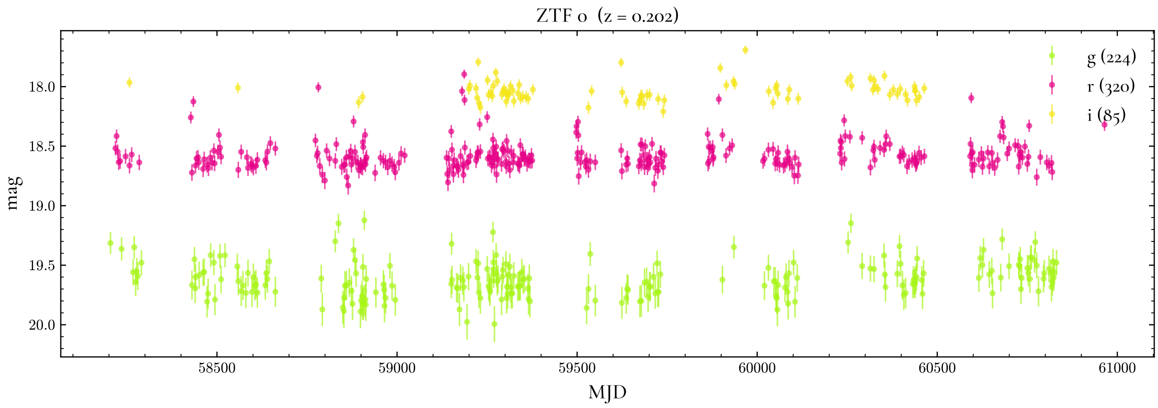


Figure 6.8: A real ZTF light curve of one AGN in the sample, showing the irregular cadence and seasonal gaps of a ground-based survey.

For each band the number of epochs, the cadence, and the photometric uncertainty are measured across the sample, and their distributions differ between the three bands (Figure 6.9). The r band is the most densely sampled, the g band carries the largest photometric uncertainties, since ZTF is shallower in g (Section 3.3.1), and the i band is the most sparsely sampled. From these measurements, three quantities per band are used to construct the ZTF-matched simulations: the median cadence, the range of epoch counts, and the 16th and 84th percentiles of the photometric error distribution.

These ZTF-matched simulations, for which the underlying signal is known, are used to test the method under survey noise before it is applied to the real light curves.

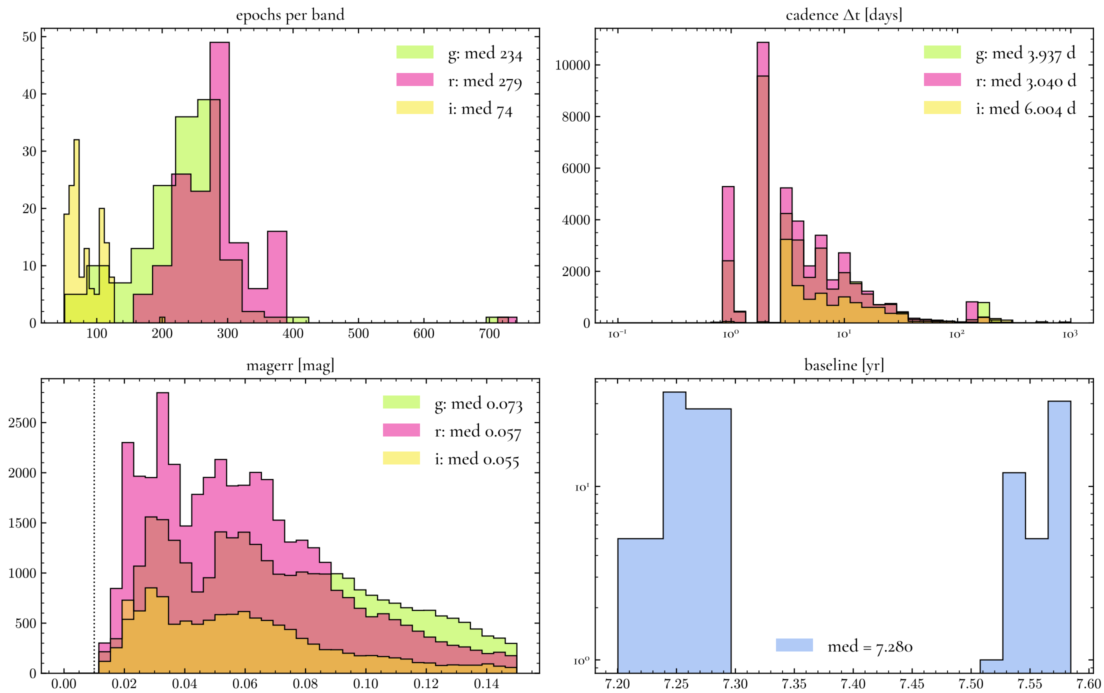


Figure 6.9: Distributions of the number of epochs, cadence, and photometric uncertainty across the 150 ZTF AGN in the g, r, and i bands.

6.4 Calibration on Simulated ZTF Data

The ZTF-matched simulations reproduce the sampling and the photometric noise of the survey, band by band, as can be appreciated on the example from Figure 6.10. The observing times of each simulated curve are taken directly from a real ZTF object in the corresponding band, chosen at random from the sample, with a random 85 to 95% of its epochs retained and each epoch jittered by half a day. In this way the simulated light curves inherit the true structure of the survey rather than an approximation of it, and no synthetic gaps are imposed.

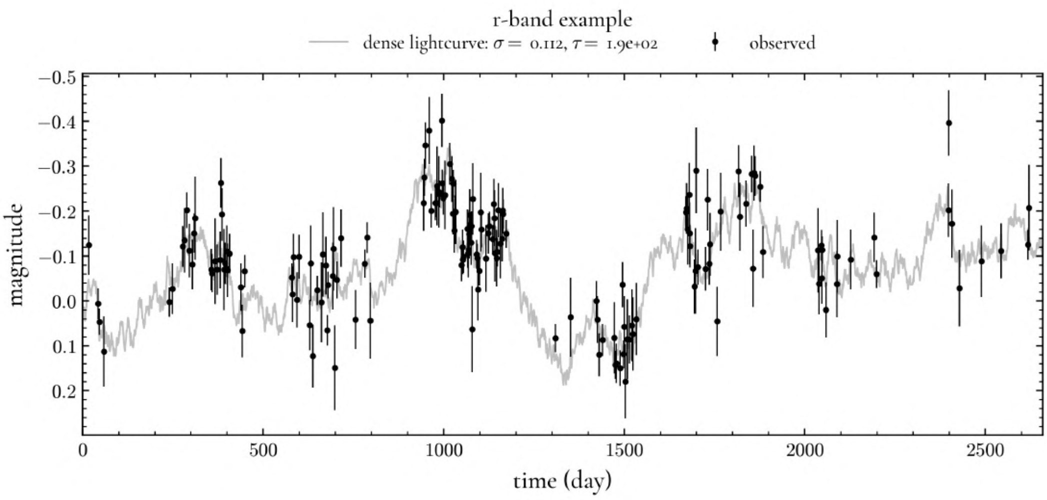
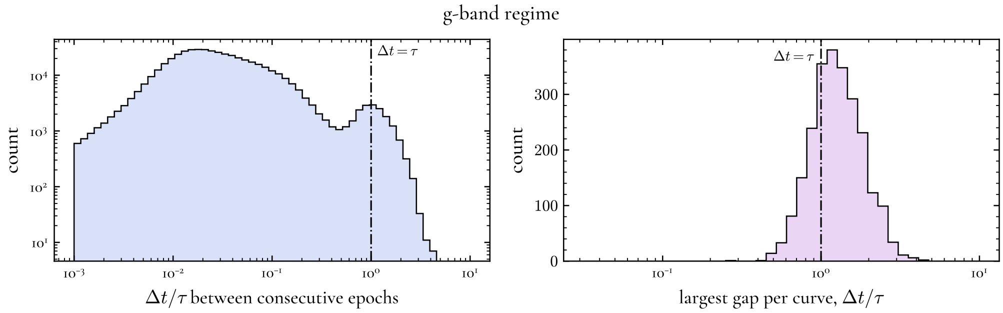


Figure 6.10: A simulated light curve from the ZTF-matched configuration, with the dense underlying realization and the sparse noisy observations.

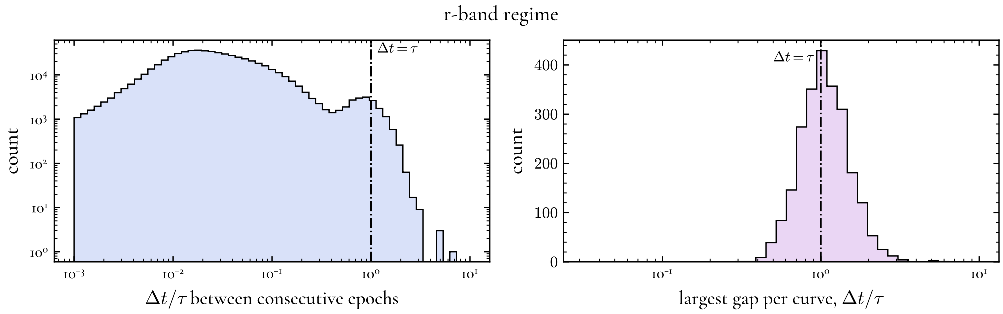
Each epoch is then assigned its own photometric uncertainty, drawn log-uniformly between the 16th and 84th percentiles of the measured error distribution of that band. The noise is therefore heteroscedastic, varying from epoch to epoch across the central range of the real errors, in contrast to the controlled experiments of Section 6.2. Expressed relative to the variability amplitude, the resulting per-epoch uncertainty has a median of 0.60 in g, 0.48 in r, and 0.46 in i, and reaches values above unity in g, so that in the noisiest band the measurement error is at times comparable to the intrinsic variability itself.

As in the low-noise experiments, the reconstruction is measured against a Gaussian process given the true parameters of each curve, which remains the Bayes-optimal reference on simulated data.

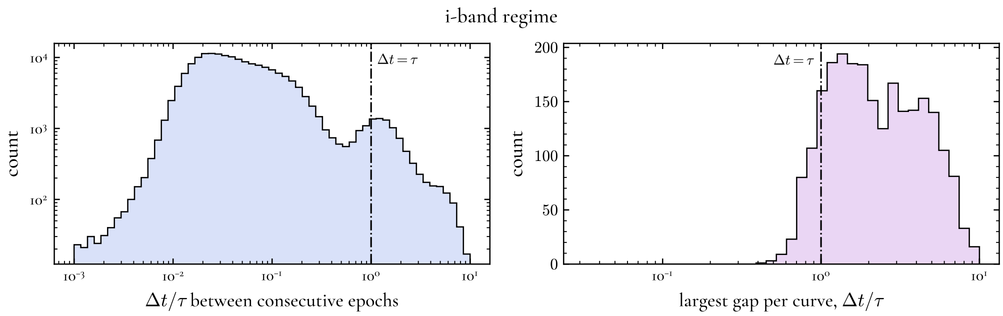
The size of the gaps relative to the damping timescale, which sets the difficulty of reconstruction as described in Section 4.4, differs noticeably between the three simulated bands (Figure 6.11).



(a) g band



(b) r band



(c) i band

Figure 6.11: The spacing between consecutive epochs (left) and the single largest gap per curve (right), in units of the damping timescale, for the ZTF-matched simulations in each band. The line marks $\Delta t = \tau$.

In the g band the median, largest gap per curve is $1.24\,\tau$, with 73% of curves containing a gap wider than the damping timescale, in the r band the median largest gap is $1.05\,\tau$, with 56% exceeding τ, and in the i band the median largest gap is $2.10\,\tau$, with 89% exceeding τ and a tail reaching almost seven damping timescales. The i band therefore contains the widest gaps, some large enough that the endpoints retain essentially no information about the intervening signal, and it is the band in which single-band reconstruction is most strongly limited.

Trained and evaluated on these light curves, the method continues to reconstruct the underlying signal, but its predictive intervals become overconfident. To establish that this is a property of the model rather than of a particular training run, the training is repeated across five random seeds and the results are given in Table 6.2. The coverage at the 68% level is below the nominal value in every band, at 0.30, 0.33, and 0.42 in g, r, and i, and the scatter across seeds is a few per cent, so the overconfidence is systematic. The predictive log-likelihood worsens with the noise level of the band, from -3.29 in the least noisy band to -10.63 in the noisiest, since the noisiest band produces the narrowest intervals relative to the scatter of the truth and is penalized most heavily for the points that fall outside them.

Table 6.2: Median coverage at the 68% level, RMSE, and predictive log-likelihood on the ZTF-matched validation set, as mean and standard deviation across five training seeds.

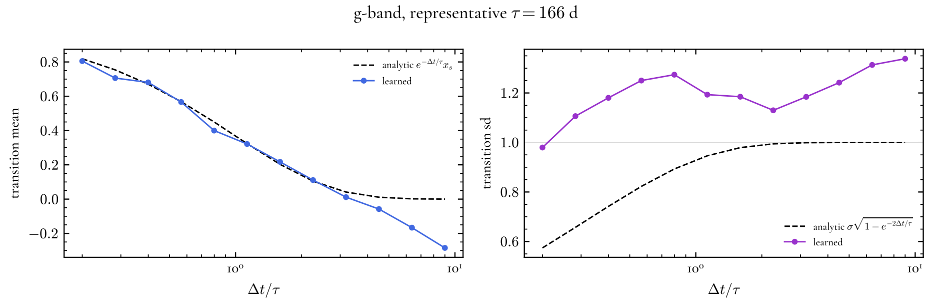
Band	Coverage @ 68%	RMSE	Predictive LL
g	0.301 ± 0.026	0.569 ± 0.004	-10.63 ± 2.34
r	0.331 ± 0.020	0.456 ± 0.003	-8.84 ± 1.43
i	0.419 ± 0.020	0.466 ± 0.002	-3.29 ± 0.44

The overconfidence is present at every level of the predictive interval. Table 6.3 compares the empirical coverage of the reconstruction with that of the Gaussian process at four nominal levels. The Gaussian process is close to calibrated throughout, its empirical coverage matching each nominal level, while the reconstruction falls short at all of them, reaching only 0.46 where 0.90 is expected in the g band. The RMSE is roughly twice that of the Gaussian process in every band, so the method loses accuracy as well as miscalibrating its uncertainty.

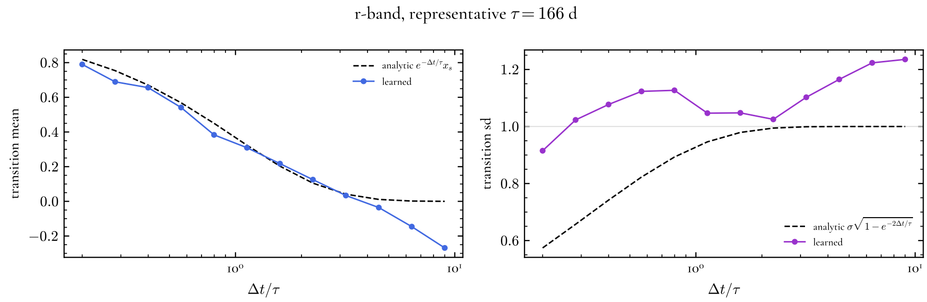
Table 6.3: Empirical coverage at four nominal levels, with RMSE, for the reconstruction and the Gaussian process given the true parameters, on the ZTF-matched validation set.

Band	Method	Coverage at nominal level				RMSE
		0.50	0.68	0.90	0.95	
g	NSF	0.21	0.31	0.46	0.51	0.572
	GP	0.51	0.68	0.89	0.94	0.296
r	NSF	0.23	0.33	0.50	0.56	0.458
	GP	0.51	0.69	0.90	0.94	0.264
i	NSF	0.31	0.43	0.61	0.67	0.470
	GP	0.50	0.67	0.89	0.94	0.318

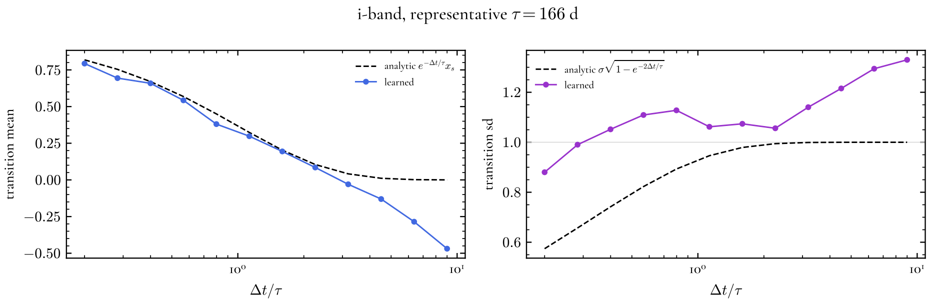
The origin of the overconfidence is visible in the learned transition, which is no longer recovered as it was at low noise. Figure 6.12 repeats the transition comparison for models trained in each band. The agreement with the analytic form is lost: the learned mean overshoots, passing below zero at large separations instead of relaxing to it, and the learned standard deviation inflates to about 1.33 at nine damping timescales and does not saturate at unity.



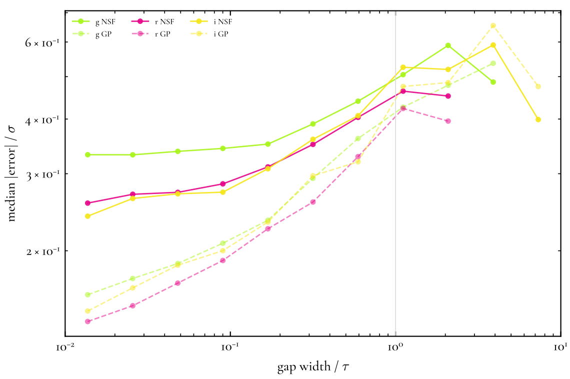
(a) g band



(b) r band



(c) i band



(d) Reconstruction error against gap width

Figure 6.12: Behaviour of the reconstruction in the ZTF regime. Panels (a)–(c): the learned transition for models trained in each band, as in Figure 6.2b. Panel (d): the reconstruction error as a function of gap width in units of the damping timescale, for the three bands, compared with the Gaussian process.

This is the same departure seen at varying timescale in Section 6.2.2, now much larger, and it is consistent with the bridge anchoring too tightly to the noisy observations at the gap endpoints, treating these noisy magnitudes as if it were the true signal.

That the deficit is one of noise handling rather than of gap filling is indicated by the way it depends on the sampling. The coverage worsens as the number of epochs per curve increases (Figure 6.13), the opposite of what more data should do, because each additional noisy epoch is treated as another exact constraint rather than another uncertain measurement, and the sparsely sampled i band is the least overconfident of the three. This, together with the loss of the learned transition under noise, points to the absence of per-epoch uncertainty handling. The reconstruction error itself grows smoothly with gap width and converges to the Gaussian process at the widest gaps (Figure 6.12d), so the point reconstruction, though less accurate than the optimal reference throughout, still behaves orderly as a function of gap size.

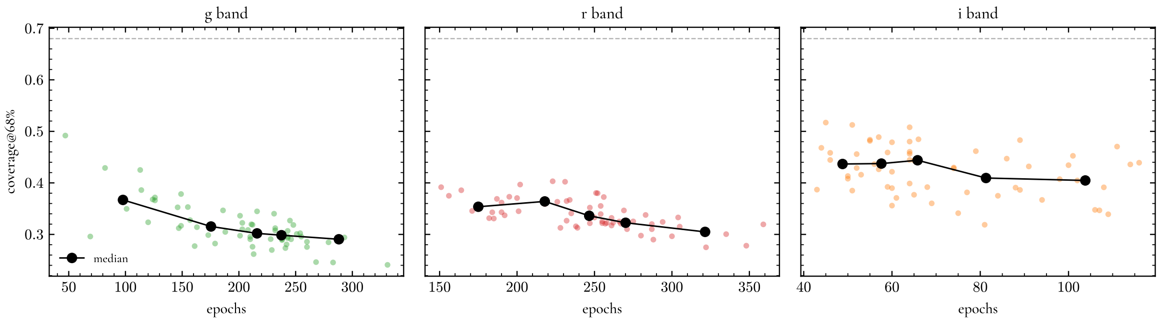


Figure 6.13: Coverage at the 68% level as a function of the number of epochs per curve, for the three ZTF bands, with the median taken in quantile bins. The coverage worsens as the number of epochs grows, consistent with the reconstruction anchoring more tightly to a larger number of noisy observations.

Example reconstructions are shown for each band in Figures 6.14 to 6.16, each compared with the Gaussian process oracle on the same curve.

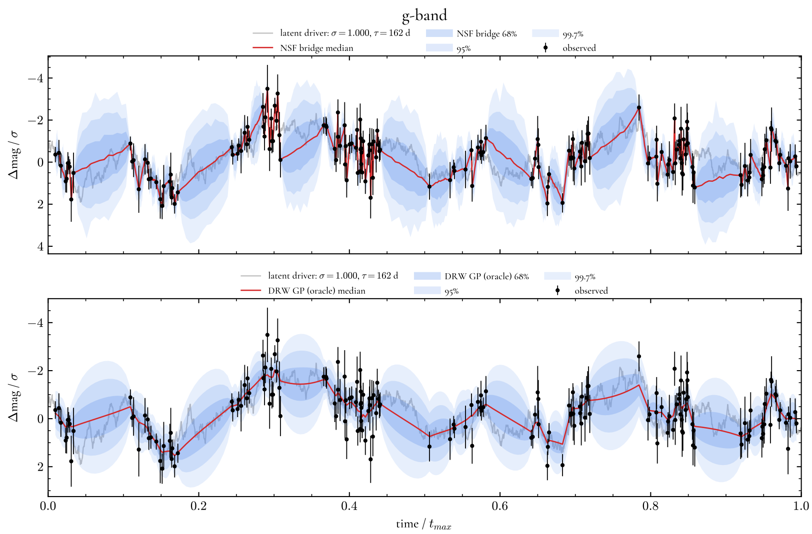


Figure 6.14: Reconstruction of a simulated ZTF-matched light curve in the g band by the neural stochastic flow (top) and the Gaussian process given the true parameters (bottom), shown against the true dense realization (grey), with the observations and their uncertainties in black. Each panel shows the posterior median (red) and the nested central 68, 95, and 99.7% intervals (shaded).

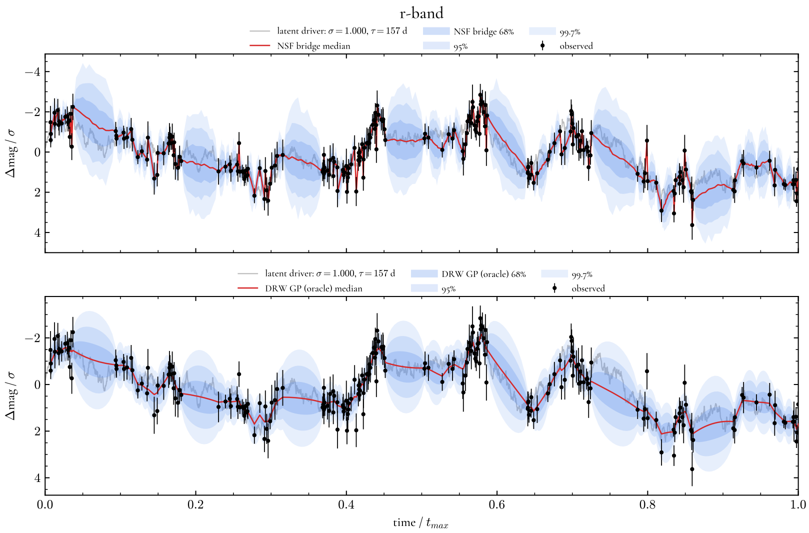


Figure 6.15: As Figure 6.14, for a simulated ZTF-matched light curve in the r band.

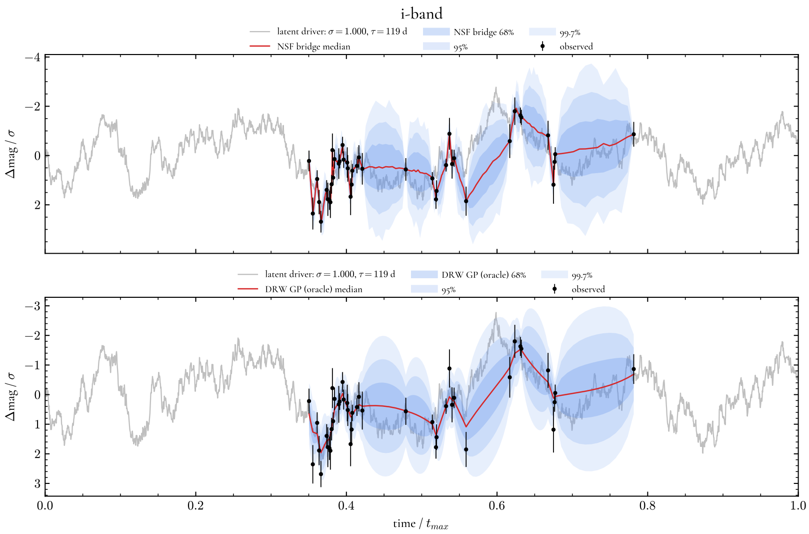


Figure 6.16: As Figure 6.14, for a simulated ZTF-matched light curve in the i band.

In a single example the reconstruction seems plausible, with the median following the signal and the intervals widening across the gaps; the overconfidence reported in Table 6.3 is a property of the ensemble, more apparent in the coverage over the full validation set.

6.5 Application to Real ZTF Light Curves

The method is finally applied to the real ZTF light curves. Since the true signal is unavailable, the reconstruction is tested by masking, in the notebook `ZTF_Stage1.ipynb`. A fraction of the observations in each curve is withheld in two forms. Isolated single points, removed at random and bracketed by nearby retained observations, test the reconstruction where it is well constrained, while points removed inside contiguous intervals test it across gaps. Before masking, outliers are removed by comparing each point with a median of its neighbours and discarding those that deviate by more than 5×s the median absolute deviation of the residualst, so that a single discrepant epoch is not mistaken for a real excursion.

Because the true parameters are unknown for the real light curves, the Gaussian process can no longer act as an oracle. Its damping timescale and amplitude are instead fitted to the retained observations of each curve by maximum likelihood, maximizing the Gaussian process likelihood of the retained points under the damped random walk kernel, with the per-epoch measurement uncertainties included in the covariance. Unlike in the simulated experiments, the Gaussian process must therefore estimate the parameters it conditions on, from the same data it reconstructs. The normalization of each curve is likewise fitted to the retained points alone, so that the withheld observations play no part in the reconstruction.

Furthermore, a linear interpolation between the retained observations is added as a trivial reference, providing a point prediction with no associated uncertainty. Each masked curve is reconstructed by the flow, by this fitted Gaussian process, and by the interpolation, and the three are compared with the withheld observations.

The withheld observations are themselves noisy, so a reconstruction of the latent signal should scatter around them rather than pass through them exactly, and the predictive intervals must be broadened by the measurement error of each withheld point before they can be compared. For the Gaussian process, whose predictive distribution is Gaussian, this is done by adding the variance of the withheld point to the predictive variance. For the reconstruction, whose predictive distribution is represented by samples and need not be Gaussian, the same broadening is applied by adding independent Gaussian noise of the appropriate variance to each sample and recomputing the percentiles of the noisy samples. Both treatments form the predictive distribution of a noisy observation rather than of the latent signal, in the manner appropriate to each representation. Since the withheld points enter neither reconstruction, this broadening is applied only at the scoring stage, and it is applied equally to both methods.

The results are given in Table 6.4, separately for the isolated points and for those inside gaps. The behaviour depends on the band, and it does so in a way that follows the noise level. On the two less noisy bands, r and i, the reconstruction is close to calibrated on the isolated points, with coverage of 0.68 and 0.73, and over covers on the harder in-gap points, reaching 0.85 against the nominal 0.68, so that its intervals are wider than necessary across real gaps. On the noisiest band, g, the reconstruction instead under-covers on the isolated points, at 0.59, and is only close to calibrated in the gaps, at 0.72. This is the band in which the intervals were most overconfident against the latent signal, and the deficit persists here once the measurement noise is accounted for. The fitted Gaussian process is more consistent across the three bands, with coverage between 0.69 and 0.75 throughout. In accuracy, both probabilistic methods predict the withheld observations to within about one variability amplitude, the Gaussian process being consistently the more accurate, and both improve on the linear interpolation while additionally providing an uncertainty, which the interpolation does not. Example reconstructions in the three bands are shown in Figure 6.17.

Table 6.4: Coverage at the 68% level and RMSE for held-out prediction on real ZTF light curves, scored against the withheld noisy observations, shown separately for randomly removed isolated points and for points removed inside contiguous gaps. The Gaussian process uses parameters fitted to the retained points, and the predictive intervals of both probabilistic methods are broadened by the measurement uncertainty of each withheld observation. RMSE is in units of the fitted variability amplitude.

		Coverage @ 68%		RMSE	
Band	Points	NSF	GP	NSF	GP
g	isolated	0.59	0.71	0.855	0.829
	in-gap	0.72	0.74	1.003	0.941
r	isolated	0.68	0.71	0.995	0.887
	in-gap	0.85	0.75	1.080	1.012
i	isolated	0.73	0.70	0.987	0.940
	in-gap	0.85	0.69	1.101	0.999

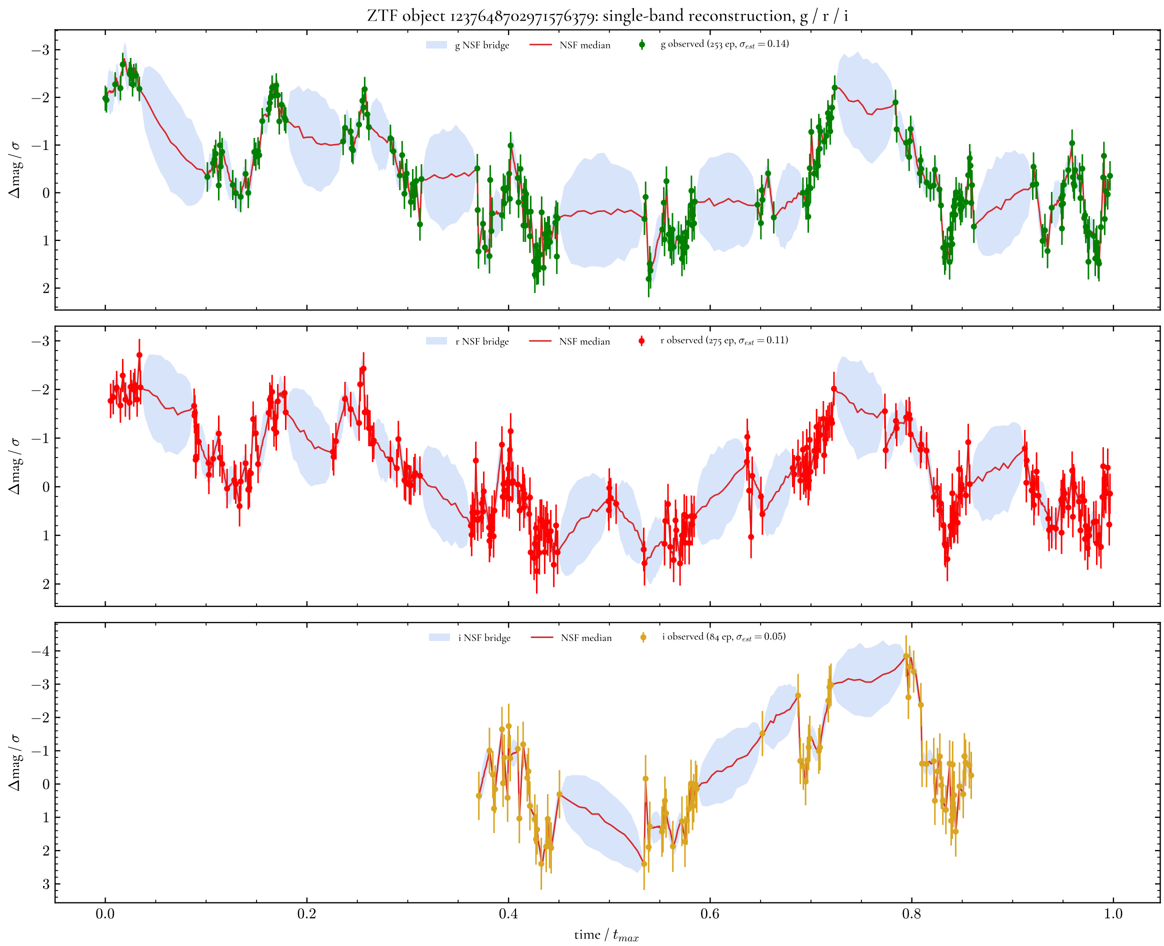


Figure 6.17: Reconstruction of a real ZTF light curve in the g, r, and i bands. The retained observations are shown in colour, the posterior median in red, and the central 68% interval shaded, with the interval widening across the wider gaps in the real cadence.

The result on real data appears at first to sit uneasily with the simulated evaluation, where the intervals were overconfident in every band. The two are not in conflict. In the simulated experiment the intervals are compared with the noise-free underlying signal, while here they are compared with noisy observations and are broadened by the measurement error, so the two evaluations measure the coverage of different quantities. On the less noisy bands, this broadening is enough to bring the reconstruction to nominal coverage on the isolated points and beyond it in the gaps, while on the noisiest band the underlying overconfidence is large enough to remain visible even after the broadening. What holds across both evaluations is that the calibration of the reconstruction depends on how the measurement noise is treated, and that the fitted Gaussian process, which uses the per-epoch uncertainties directly, is the more consistent baseline on real data.

6.6 Discussion

The four experiments, taken in increasing order of realism, help locate the strengths and the limitations of the method. On simulated data at low noise, at both a fixed timescale and a distribution of timescales, the reconstruction recovers the analytic transition of the damped random walk and performs within a few per cent of the Bayes-optimal Gaussian process, in both coverage and accuracy. The model and its training objective are therefore solid, even across a range of dynamics. In this regime the reconstruction error grows smoothly with the width of the gap and converges to the prior at the widest gaps, where the observations no longer constrain the signal, the orderly behaviour expected of a well-posed reconstruction.

The limitation appears only under the higher, heteroscedastic noise of the ZTF regime. There the reconstruction error still grows smoothly with gap width, but it sits above that of the Gaussian process throughout, by roughly a factor of two, and the predictive intervals become overconfident, systematically across training seeds and at every level of the interval. The loss of accuracy is approximately uniform across gap widths rather than a breakdown at the widest gaps, which already suggests that the difficulty lies not in filling the gaps but in the treatment of the noise. Three results confirm this. The learned transition, recovered cleanly at low noise, is lost, with the mean overshooting and the standard deviation inflating rather than saturating. The predictive log-likelihood worsens with the noise level of the band. And the coverage worsens as the number of epochs per curve grows, the opposite of what more data should do, since each additional noisy epoch is treated as an exact constraint rather than an uncertain measurement. The flow and bridge are conditioned on the fluxes and times alone, with no channel for the per-epoch uncertainties, so the model treats each noisy magnitude as the true signal, and the overconfidence follows directly.

The real-data result appears at first to disagree with the simulated one, but the two must be read together and carefully. Against the noise-free latent signal, in the simulated experiment, the intervals are overconfident in every band. Against noisy held-out observations, in the masking experiment, the intervals are broadened by the measurement error before scoring, and this is enough to bring the reconstruction to nominal coverage on the two less noisy bands, while the noisiest band keeps a deficit. The two are not in conflict, since they measure the coverage of different quantities, and the band most overconfident against the signal is the one whose deficit survives the broadening. What holds throughout is that the calibration of the reconstruction depends on how the measurement noise is handled, and that the fitted Gaussian process, which uses the per-epoch uncertainties directly, is the more consistent baseline on real data. Both probabilistic methods improve on linear interpolation and, unlike it, provide an uncertainty with each prediction.

7 Conclusions

This thesis investigated the use of neural stochastic flows for the reconstruction of gaps in the optical light curves of active galactic nuclei, a problem made underdetermined by the irregular sampling of time-domain surveys and therefore suited to a probabilistic treatment. The damped random walk provided a favourable setting for this investigation, because its transition probability is known in closed form and it admits an exact Gaussian process treatment. The method could therefore be validated against the analytic transition of the very process it was trained to learn, and measured against a Gaussian process that is optimal by construction. Few of the processes on which such generative models are usually demonstrated allow either check.

On simulated light curves at low noise, at the 0.01 mag photometric precision expected of the LSST, the method recovered the analytic transition and reconstructed the light curves to within about three per cent of the Bayes-optimal Gaussian process in accuracy, at a median coverage of 0.61 against 0.70. This held at both a fixed damping timescale and a distribution of timescales drawn from a physical prior, and it was insensitive to the weighting between the two terms of the training objective, so the result does not depend on a tuned hyperparameter. The method therefore learns the underlying dynamics and interpolates the gaps well when the observations are precise, with a reconstruction error that grows smoothly with the width of the gap and converges to the prior where the observations no longer constrain the signal.

When the sampling and the heteroscedastic noise were matched to those of the Zwicky Transient Facility, the reconstruction was no longer competitive. Its error grew to about twice that of the Gaussian process, and its predictive intervals became overconfident, with a coverage near 0.3 against a nominal 0.68, systematically across training seeds and at every level of the interval. Results traced this to the treatment of measurement noise: the learned transition was lost, the predictive log-likelihood worsened with the noise level of the band, and the coverage worsened as more noisy observations were added, the opposite of what more data should do. Applied to real ZTF light curves through a masking experiment, in which the intervals were broadened by the measurement error before scoring, the reconstruction was close to calibrated on the two less noisy bands and kept a deficit on the noisiest, while a Gaussian process fitted to each curve, which uses the per-epoch uncertainties directly, remained the more consistent baseline; both improved on linear interpolation. The calibration of the reconstruction depends throughout on how the measurement noise is handled, and it is the noisiest data that expose the limitation most clearly.

In the current setup, the flow and bridge are conditioned on the observed magnitudes and times alone, the per-epoch uncertainties are not among the model's inputs, so it treats each noisy magnitude as the true signal and anchors too tightly to it. A model that instead consumes the per-epoch uncertainties, treating each observation as a noisy measurement of an unobserved latent state and propagating that uncertainty into its predictions, would address the limitation at its source. Such a latent formulation, in which the flow describes the evolution of the underlying signal and a separate term relates it to the noisy observations, is the principal direction left for future work. The results of this thesis both motivate it and provide the controlled setting in which it can be tested.

A second direction, motivated by the widest gaps found in the i band, is the extension to multiple bands. The variability of an AGN is correlated across wavelength, as discussed in Chapter 3, so that a gap in one band typically coincides with observations in another. A model that reconstructs the bands jointly could use this coordinated variability to constrain a gap in one band with the data available in the others, recovering signal where single-band reconstruction is most strongly limited. This multiband extension is also the step that connects the present work to its original aim. The reconstruction developed here is a prerequisite for reverberation mapping, in which the delays between the variations of different bands are measured to map the structure of the accretion flow, and reliable inter-band delays can only be recovered once the gaps in each band are filled in a way that is both accurate and calibrated. This thesis has established and characterized that reconstruction step in the single-band case; extending it to several bands, and applying it to the measurement of reverberation lags, is the natural continuation. With the Legacy Survey of Space and Time due to deliver optical light curves for more than ten million AGN across six bands, a reconstruction method that is calibrated under survey noise and able to combine bands would provide the foundation for reverberation mapping at a scale far beyond what is possible today.

Code Availability

The code that produces the results of this thesis is publicly available at `https://github.com/anacecilialr1/MASS-Thesis`. The research code is organized into two parallel sets of scripts, `framework/` for the low-noise simulation experiments and `experiments/` for the experiments matched to the Zwicky Transient Facility. Each contains the simulation, model, and evaluation code, `LightCurves.py`, `NSF.py`, `GapFill.py`, and `Tensorize.py`, together with the notebooks in which the experiments are carried out. The `data/` directory holds the simulated training and validation sets and the digitized quasar mass distribution of Sun (2023), from which the damping timescales are drawn.

The experiments are organized as follows. `Stage0.ipynb` generates and validates the controlled simulations. `Stage1.ipynb` trains the neural stochastic flow and evaluates the reconstruction on the low-noise validation set, and `Stage1Tests.ipynb` contains the sensitivity tests on the loss weighting. `ZTF_Stage0.ipynb` characterizes the ZTF sample and constructs the ZTF-matched simulations, and `ZTF_Stage1.ipynb` evaluates the calibration on those simulations and applies the method to the real ZTF light curves through the masking experiment. The reconstruction method builds on the publicly available `jax_nsf` codebase of Kiyohara et al. (2025). Further detail on the repository structure is given in its `README`.

Bibliography

Antonucci, Robert (Mar. 1984). "Optical spectropolarimetry of radio galaxies." In: *The Astrophysical Journal* 278, pp. 499–520. DOI: 10.1086/161816.

— (Jan. 1993). "Unified models for active galactic nuclei and quasars." In: *Annual Review of Astronomy and Astrophysics* 31, pp. 473–521. DOI: 10.1146/annurev.aa.31.090193.002353.

Antonucci, Robert and J. S. Miller (Oct. 1985). "Spectropolarimetry and the nature of NGC 1068." In: *The Astrophysical Journal* 297, pp. 621–632. DOI: 10.1086/163559.

Arnaud, K. A. et al. (Nov. 1985). "EXOSAT observations of a strong soft X-ray excess in MKN 841." In: *Monthly Notices of the Royal Astronomical Society* 217, pp. 105–113. DOI: 10.1093/mnras/217.1.105.

Balbus, Steven A. and John F. Hawley (Jan. 1998). "Instability, turbulence, and enhanced transport in accretion disks". In: *Reviews of Modern Physics* 70.1, pp. 1–53. DOI: 10.1103/RevModPhys.70.1.

Baldwin, J. A., M. M. Phillips, and R. Terlevich (Feb. 1981). "Classification parameters for the emission-line spectra of extragalactic objects." In: *Publications of the Astronomical Society of the Pacific* 93, pp. 5–19. DOI: 10.1086/130766.

Ballantyne, David and Xin Xiang (Apr. 2022). "A Model for the X-ray Soft Excess in Active Galactic Nuclei: Combining the Effects of Reflection and a Warm Corona". In: *High Energy Astrophysics Division*. Vol. 19, 106.39, p. 106.39.

Barvainis, Richard (Sept. 1987). "Hot Dust and the Near-Infrared Bump in the Continuum Spectra of Quasars and Active Galactic Nuclei". In: *The Astrophysical Journal* 320, p. 537. DOI: 10.1086/165571.

Beard, Max et al. (Jan. 2025). *Testing disc reprocessing models for AGN optical variability by comparison of X-ray and optical power spectra of NGC 4395.* arXiv: 2501.02664 [astro-ph.HE]. URL: https://arxiv.org/abs/2501.02664.

Bellm, Eric C. et al. (Jan. 2019). "The Zwicky Transient Facility: System Overview, Performance, and First Results". In: *Publications of the Astronomical Society of the Pacific* 131.995, p. 018002. DOI: 10.1088/1538-3873/aaecbe.

Bennert, Nicola (2005). "The narrow-line region of active galaxies". Doctoral Thesis. Ruhr-Universität Bochum, Universitätsbibliothek.

Bianchi, Stefano, Roberto Maiolino, and Guido Risaliti (Jan. 2012). "AGN Obscuration and the Unified Model". In: *Advances in Astronomy* 2012, pp. 1–17. ISSN: 1687-7977. DOI: 10.1155/2012/782030.

Blandford, R. D. and C. F. McKee (Apr. 1982). "Reverberation mapping of the emission line regions of Seyfert galaxies and quasars." In: *The Astrophysical Journal* 255, pp. 419–439. DOI: 10.1086/159843.

Blandford, R. D. and M. J. Rees (Jan. 1978). "Some comments on radiation mechanisms in Lacertids." In: *BL Lac Objects*. Ed. by A. M. Wolfe, pp. 328–341.

Blandford, R. D. and R. L. Znajek (May 1977). "Electromagnetic extraction of energy from Kerr black holes." In: *Monthly Notices of the Royal Astronomical Society* 179, pp. 433–456. DOI: 10.1093/mnras/179.3.433.

Bradbury, James et al. (2018). *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. URL: http://github.com/jax-ml/jax.

Burke, Colin J. et al. (Aug. 2021). "A characteristic optical variability time scale in astrophysical accretion disks". In: *Science* 373.6556, pp. 789–792. DOI: 10.1126/science.abg9933.

Cackett, Edward M. et al. (June 2021). "Reverberation mapping of active galactic nuclei: From X-ray corona to dusty torus". In: *iScience* 24.6, p. 102557. ISSN: 2589-0042. DOI: 10.1016/j.isci.2021.102557.

Chambers, K. C. et al. (Dec. 2016). "The Pan-STARRS1 Surveys". In: *arXiv e-prints*, arXiv:1612.05560, arXiv:1612.05560. DOI: 10.48550/arXiv.1612.05560. arXiv: 1612.05560 [astro-ph.IM].

Clarke, A. O. et al. (May 2017). "LOFAR MSSS: Discovery of a 2.56 Mpc giant radio galaxy associated with a disturbed galaxy group". In: *Astronomy and Astrophysics* 601, A25, A25. DOI: 10.1051/0004-6361/201630152.

Collier, Stefan and Bradley M. Peterson (July 2001). "Characteristic Ultraviolet/Optical Timescales in Active Galactic Nuclei". In: *The Astrophysical Journal* 555.2, pp. 775–785. DOI: 10.1086/321517.

Collier, Stefan et al. (Jan. 1999). "A new direct method for measuring the Hubble constant from reverberating accretion discs in active galaxies". In: *Monthly Notices of the Royal Astronomical Society* 302.1, pp. L24–L28. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.1999.02250.x.

Crummy, J. et al. (Feb. 2006). "An explanation for the soft X-ray excess in active galactic nuclei". In: *Monthly Notices of the Royal Astronomical Society* 365.4, pp. 1067–1081. DOI: 10.1111/j.1365-2966.2005.09844.x.

Czerny, B. and K. Hryniewicz (Jan. 2011). "The origin of the broad line region in active galactic nuclei". In: *Astronomy and Astrophysics* 525, L8, p. L8. DOI: 10.1051/0004-6361/201016025. arXiv: 1010.6201 [astro-ph.CO].

De Cicco, D. et al. (Aug. 2022). "A structure function analysis of VST-COSMOS AGN". In: *Astronomy and Astrophysics* 664, A117, A117. DOI: 10.1051/0004-6361/202142750.

DeepMind et al. (2020). *The DeepMind JAX Ecosystem*. URL: http://github.com/google-deepmind.

Dent, W. A. (Nov. 1965). "Variation in the Radio Emission of 3C273 and Other Quasi-Stellar Sources." In: *Astronomical Journal* 70, p. 672. DOI: 10.1086/109792.

Dinh, Laurent, David Krueger, and Yoshua Bengio (2015). *NICE: Non-linear Independent Components Estimation*. arXiv: 1410.8516 [cs.LG].

Dinh, Laurent et al. (2017). *Density estimation using Real NVP*. arXiv: 1605.08803 [cs.LG].

Edge, D. O. et al. (Jan. 1959). "A survey of radio sources at a frequency of 159 Mc/s." In: *Memoirs of the Royal Astronomical Society* 68, pp. 37–60.

Elvis, Martin et al. (Nov. 1994). "Atlas of Quasar Energy Distributions". In: *The Astrophysical Journal Supplement Series* 95, p. 1. DOI: 10.1086/192093.

Emmanoulopoulos, D., I. M. McHardy, and P. Uttley (May 2010). "On the use of structure functions to study blazar variability: caveats and problems". In: *Monthly Notices of the Royal Astronomical Society* 404.2, pp. 931–946. DOI: 10.1111/j.1365-2966.2010.16328.x.

Fabian, A. C. (Sept. 2012). "Observational Evidence of Active Galactic Nuclei Feedback". In: *Annual Review of Astronomy and Astrophysics* 50, pp. 455–489. DOI: 10.1146/annurev-astro-081811-125521.

Fabian, A. C. et al. (2009). "Broad line emission from iron K- and L-shell transitions in the active galaxy 1H 0707-495". In: *Nature* 459.7246, pp. 540–542. DOI: 10.1038/nature08007.

Fischer, T. C. et al. (Nov. 2013). "Determining Inclinations of Active Galactic Nuclei via their Narrow-line Region Kinematics. I. Observational Results". In: *The Astrophysical Journal Supplement Series* 209.1, 1, p. 1. DOI: 10.1088/0067-0049/209/1/1.

Foreman-Mackey, Daniel et al. (Nov. 2017). "Fast and Scalable Gaussian Process Modeling with Applications to Astronomical Time Series". In: *The Astronomical Journal* 154.6, p. 220. ISSN: 1538-3881. DOI: 10.3847/1538-3881/aa9332.

Giebels, B. and B. Degrange (Sept. 2009). "Lognormal variability in BL Lacertae". In: *Astronomy and Astrophysics* 503.3, pp. 797–799. DOI: 10.1051/0004-6361/200912303.

Giustini, Margherita and Daniel Proga (2019). "A global view of the inner accretion and ejection flow around super massive black holes: Radiation-driven accretion disk winds in a physical context". In: *Astronomy and Astrophysics* 630, A94. ISSN: 1432-0746.

Graham, Matthew et al. (May 2019). "The Zwicky Transient Facility: Science Objectives". In: *Publications of the Astronomical Society of the Pacific* 131, p. 078001. DOI: 10.1088/1538-3873/ab006c.

Gravity Collaboration et al. (Nov. 2018). "Spatially resolved rotation of the broad-line region of a quasar at sub-parsec scale". In: *Nature* 563.7733, pp. 657–660. DOI: 10.1038/s41586-018-0731-9. arXiv: 1811.11195 [astro-ph.GA].

Harrison, Christopher (2014). "Observational constraints on the influence of active galactic nuclei on the evolution of galaxies". PhD Thesis. Durham University, UK.

Hu, Xu-Fan (2024). "How Can the Optical Variation Properties of Active Galactic Nuclei Be Unbiasedly Measured?" In: *The Astrophysical Journal* 961. DOI: 10.3847/1538-4357/ad072f.

Ishibashi, W. and T. J.-L. Courvoisier (Jan. 2011). "Synchrotron radio emission in radio-quiet AGNs". In: *Astronomy and Astrophysics* 525, A118, A118. DOI: 10.1051/0004-6361/201014987.

Ivezić, Željko et al. (Mar. 2019). "LSST: From Science Drivers to Reference Design and Anticipated Data Products". In: *The Astrophysical Journal* 873.2, 111, p. 111. DOI: 10.3847/1538-4357/ab042c.

Kara, Erin and Javier García (2025). *Supermassive Black Holes in X-rays: From Standard Accretion to Extreme Transients*. arXiv: 2503.22791 [astro-ph.HE].

Kaspi, Shai et al. (Apr. 2000). "Reverberation Measurements for 17 Quasars and the Size-Mass-Luminosity Relations in Active Galactic Nuclei". In: *The Astrophysical Journal* 533.2, pp. 631–649. DOI: 10.1086/308704.

Kauffmann, Guinevere et al. (Dec. 2003). "The host galaxies of active galactic nuclei". In: *Monthly Notices of the Royal Astronomical Society* 346.4, pp. 1055–1077. DOI: 10.1111/j.1365-2966.2003.07154.x.

Kelly, Brandon C. et al. (May 2009). "Are the Variations in Quasar Optical Flux Driven by Thermal Fluctuations?" In: *The Astrophysical Journal* 698.1, p. 895. DOI: 10.1088/0004-637X/698/1/895.

Kelly, Brandon C. et al. (2014). "Flexible and Scalable Methods for Quantifying Stochastic Variability in the Era of Massive Time-Domain Astronomical Data Sets". In: *The Astrophysical Journal* 788.1. ISSN: 1538-4357. DOI: 10.1088/0004-637x/788/1/33.

Kewley, L. J. et al. (July 2001). "Theoretical Modeling of Starburst Galaxies". In: *The Astrophysical Journal* 556.1, pp. 121–140. DOI: 10.1086/321545.

Kidger, Patrick and Cristian Garcia (2021). *Equinox: Neural Networks in JAX via callable PyTrees and filtered transformations*. eprint: 2111.00254.

Kingma, Diederik P. and Jimmy Ba (2017). *Adam: A Method for Stochastic Optimization*. URL: https://arxiv.org/abs/1412.6980.

Kiyohara, Naoki et al. (2025). "Neural Stochastic Flows: Solver-Free Modelling and Inference for SDE Solutions". In: *Advances in Neural Information Processing Systems*.

Kobyzev, Ivan et al. (Nov. 2021). "Normalizing Flows: An Introduction and Review of Current Methods". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.11, pp. 3964–3979. ISSN: 1939-3539. DOI: 10.1109/tpami.2020.2992934.

Kozłowski, Szymon (2016a). "Limitations on the recovery of the true AGN variability parameters using Damped Random Walk modeling". In: *arXiv: Astrophysics of Galaxies*. URL: https://api.semanticscholar.org/CorpusID:119237192.

— (Aug. 2016b). "Revisiting Stochastic Variability of AGNs with Structure Functions". In: *The Astrophysical Journal* 826.2, 118, p. 118. DOI: 10.3847/0004-637X/826/2/118.

Krolik, J. H. et al. (Apr. 1991). "Ultraviolet Variability of NGC 5548: Dynamics of the Continuum Production Region and Geometry of the Broad-Line Region". In: *The Astrophysical Journal* 371, p. 541. DOI: 10.1086/169918.

Kubota, Aya and Chris Done (Oct. 2018). "A physical model of the broad-band continuum of AGN and its implications for the UV/X relation and optical variability". In: *Monthly Notices of the Royal Astronomical Society* 480.1, pp. 1247–1262. DOI: 10.1093/mnras/sty1890.

LaMassa, Stephanie M. et al. (Feb. 2015). "The Discovery of the First "Changing Look" Quasar: New Insights into the Physics and Phenomenology of Active Galactic Nuclei". In: *The Astrophysical Journal* 800.2, p. 144. DOI: 10.1088/0004-637X/800/2/144.

Laor, Ari and Ehud Behar (Oct. 2008). "On the origin of radio emission in radio-quiet quasars". In: *Monthly Notices of the Royal Astronomical Society* 390.2, pp. 847–862. ISSN: 1365-2966. DOI: 10.1111/j.1365-2966.2008.13806.x.

Law, Nicholas M. et al. (Dec. 2009). "The Palomar Transient Factory: System Overview, Performance, and First Results". In: *Publications of the Astronomical Society of the Pacific* 121.886, p. 1395. DOI: 10.1086/648598.

Lusso, E. et al. (Sept. 2012). "Bolometric luminosities and Eddington ratios of X-ray selected active galactic nuclei in the XMM-COSMOS survey". In: *Monthly Notices of the Royal Astronomical Society* 425.1, pp. 623–640. DOI: 10.1111/j.1365-2966.2012.21513.x.

Lutz, Dieter (Aug. 2014). "Far-Infrared Surveys of Galaxy Evolution". In: *Annual Review of Astronomy and Astrophysics* 52, pp. 373–414. DOI: 10.1146/annurev-astro-081913-035953.

Lyubarskii, Yu. E. (Dec. 1997). "Flicker noise in accretion discs". In: *Monthly Notices of the Royal Astronomical Society* 292.3, pp. 679–685. DOI: 10.1093/mnras/292.3.679.

MacLeod, C. L. et al. (Oct. 2010). "Modeling the Time Variability of SDSS Stripe 82 Quasars as a Damped Random Walk". In: *The Astrophysical Journal* 721.2, pp. 1014–1033. DOI: 10.1088/0004-637X/721/2/1014.

McHardy, I. M. et al. (Dec. 1981). "The Ariel V (3A) catalogue of X-ray sources. II." In: *Monthly Notices of the Royal Astronomical Society* 197, pp. 893–919. DOI: 10.1093/mnras/197.4.893.

McHardy, I. M. et al. (Dec. 2006). “Active galactic nuclei as scaled-up Galactic black holes”. In: *Nature* 444.7120, pp. 730–732. DOI: 10.1038/nature05389.

McLaughlin, Summer A. J. et al. (Apr. 2024). “Using Gaussian Processes to detect AGN flares”. In: *Monthly Notices of the Royal Astronomical Society* 529.3, pp. 2877–2892. DOI: 10.1093/mnras/stae721.

Miller, J. S., R. W. Goodrich, and William G. Mathews (Sept. 1991). “Multidirectional Views of the Active Nucleus of NGC 1068”. In: *The Astrophysical Journal* 378, p. 47. DOI: 10.1086/170406.

Miller, Kristen A. and James M. Stone (May 2000). “The Formation and Structure of a Strongly Magnetized Corona above a Weakly Magnetized Accretion Disk”. In: *The Astrophysical Journal* 534.1, p. 398. DOI: 10.1086/308736.

Murray, N. et al. (Oct. 1995). “Accretion Disk Winds from Active Galactic Nuclei”. In: *The Astrophysical Journal* 451, p. 498. DOI: 10.1086/176238.

Narayan, Ramesh and Insu Yi (June 1994). “Advection-dominated Accretion: A Self-similar Solution”. In: *The Astrophysical Journal Letters* 428, p. L13. DOI: 10.1086/187381.

Nenkova, Maia et al. (Sept. 2008). “AGN Dusty Tori. I. Handling of Clumpy Media”. In: *The Astrophysical Journal* 685.1, pp. 147–159. DOI: 10.1086/590482.

Netzer, Hagai (2013). *The Physics and Evolution of Active Galactic Nuclei.*

— (Aug. 2015). “Revisiting the Unified Model of Active Galactic Nuclei”. In: *Annual Review of Astronomy and Astrophysics* 53, pp. 365–408. DOI: 10.1146/annurev-astro-082214-122302.

Orr, M. J. L. and I. W. A. Browne (Sept. 1982). “Relativistic beaming and quasar statistics.” In: *Monthly Notices of the Royal Astronomical Society* 200, pp. 1067–1080. DOI: 10.1093/mnras/200.4.1067.

Padovani, P. (Mar. 2011). “The microjansky and nanojansky radio sky: source population and multiwavelength properties”. In: *Monthly Notices of the Royal Astronomical Society* 411.3, pp. 1547–1561. DOI: 10.1111/j.1365-2966.2010.17789.x.

— (July 2017). “On the two main classes of active galactic nuclei”. In: *Nature Astronomy* 1.8. ISSN: 2397-3366. DOI: 10.1038/s41550-017-0194.

Padovani, P. et al. (Aug. 2017). “Active galactic nuclei: what’s in a name?” In: *Astronomy & Astrophysics Review* 25.1, 2, p. 2. DOI: 10.1007/s00159-017-0102-9.

Paolillo, Maurizio and Iossif Papadakis (Aug. 2025). “Continuum optical-UV and X-ray variability of AGN: current results and future challenges”. In: *La Rivista del Nuovo Cimento* 48.8, pp. 537–621. ISSN: 1826-9850. DOI: 10.1007/s40766-025-00072-5. URL: http://dx.doi.org/10.1007/s40766-025-00072-5.

Peterson, B. M. (Mar. 1993). “Reverberation Mapping of Active Galactic Nuclei”. In: *Publications of the Astronomical Society of the Pacific* 105, p. 247. DOI: 10.1086/133140.

Peterson, B. M. and K. Horne (Mar. 2004). “Echo mapping of active galactic nuclei”. In: *Astronomische Nachrichten* 325.3, pp. 248–251. DOI: 10.1002/asna.200310207.

Petrecca, V. et al. (2024). “Ensemble power spectral density of SDSS quasars in UV/optical bands”. In: *Astronomy and Astrophysics* 686, A286. DOI: 10.1051/0004-6361/202449161.

Pogge, Richard W. (May 1988). “An Extended Ionizing Radiation Cone from the Nucleus of the Seyfert 2 Galaxy NGC 1068”. In: *The Astrophysical Journal* 328, p. 519. DOI: 10.1086/166309.

Polack, Garrett E. et al. (Nov. 2024). “Determining the Extents, Geometries, and Kinematics of Narrow-line Region Outflows in Nearby Seyfert Galaxies”. In: *The Astrophysical Journal* 975.1, 129, p. 129. DOI: 10.3847/1538-4357/ad71c3.

Ramos Almeida, Cristina and Claudio Ricci (Oct. 2017). "Nuclear obscuration in active galactic nuclei". In: *Nature Astronomy* 1, pp. 679–689. DOI: 10.1038/s41550-017-0232-z. arXiv: 1709.00019 [astro-ph.GA].

Rasmussen, Carl Edward and Christopher Williams (2006). "Gaussian Processes for Machine Learning". In: 2.

Rezende, Danilo Jimenez et al. (2016). *Variational Inference with Normalizing Flows*. arXiv: 1505.05770 [stat.ML]. URL: https://arxiv.org/abs/1505.05770.

Ricci, C. (2026). *Unification models of Active Galactic Nuclei*. arXiv: 2606.12511 [astro-ph.GA]. URL: https://arxiv.org/abs/2606.12511.

Richards, Gordon T. et al. (Oct. 2006). "Spectral Energy Distributions and Multiwavelength Selection of Type 1 Quasars". In: *The Astrophysical Journal Supplement Series* 166.2, pp. 470–497. DOI: 10.1086/506525.

Ross, Sheldon M. (1995). *Stochastic Processes*. 2nd. New York: John Wiley & Sons. ISBN: 978-0-471-12062-9.

Salpeter, E. E. (Aug. 1964). "Accretion of Interstellar Matter by Massive Objects." In: *The Astrophysical Journal* 140, pp. 796–800. DOI: 10.1086/147973.

Schawinski, Kevin et al. (Dec. 2007). "Observational evidence for AGN feedback in early-type galaxies". In: *Monthly Notices of the Royal Astronomical Society* 382.4, pp. 1415–1431. DOI: 10.1111/j.1365-2966.2007.12487.x.

Schmidt, M. (Mar. 1963). "3C 273 : A Star-Like Object with Large Red-Shift". In: *Nature* 197.4872, p. 1040. DOI: 10.1038/1971040a0.

Seward, Frederick D. and Philip A. Charles (2010). *Exploring the X-Ray Universe*. 2nd. Chap. Active Galactic Nuclei, pp. 277–305.

Seyfert, Carl K. (Jan. 1943). "Nuclear Emission in Spiral Nebulae." In: *The Astrophysical Journal* 97, p. 28. DOI: 10.1086/144488.

Shakura, N. I. and R. A. Sunyaev (Jan. 1973). "Black holes in binary systems. Observational appearance." In: *Astronomy and Astrophysics* 24, pp. 337–355.

Shields, Gregory A. (2004). *A Brief History of Active Galactic Nuclei*. https://ned.ipac.caltech.edu/level5/Sept04/Shields/frames.html. Accessed: 2026-05-28.

Singh, K. P. et al. (Oct. 1985). "Observations of Soft X-Ray Spectra from a Seyfert 1 and a Narrow Emission-Line Galaxy". In: *The Astrophysical Journal* 297, p. 633. DOI: 10.1086/163560.

Slavcheva-Mihova, L. et al. (2013). *3C 273 - half a century later*. arXiv: 1401.5987 [astro-ph.HE].

Smith, Harlan J. and Dorrit Hoffleit (May 1963). "Light Variations in the Superluminous Radio Galaxy 3C273". In: *Nature* 198.4881, pp. 650–651. DOI: 10.1038/198650a0.

Stalevski, Marko et al. (Mar. 2012). "3D radiative transfer modelling of the dusty tori around active galactic nuclei as a clumpy two-phase medium". In: *Monthly Notices of the Royal Astronomical Society* 420.4, pp. 2756–2772. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2011.19775.x.

Suh, Hyewon et al. (Dec. 2015). "Eddington Ratio Distribution of X-ray-selected Broad-line AGNs at 1.0 < z < 2.2". In: *The Astrophysical Journal* 815.2, p. 129. DOI: 10.1088/0004-637X/815/2/129.

Sun, Mouyuan (Mar. 2023). "The mass distribution of quasars in optical time-domain surveys". In: *Monthly Notices of the Royal Astronomical Society* 521.2, pp. 2954–2961. ISSN: 1365-2966. DOI: 10.1093/mnras/stad740.

The Event Horizon Telescope Collaboration et al. (Apr. 2019). "First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole". In: *The Astrophysical Journal Letters* 875.1, p. L4. DOI: 10.3847/2041-8213/ab0e85.

Ulrich, Marie-Helene et al. (1997). “Variability of active galactic nuclei”. In: *Ann. Rev. Astron. Astrophys.* 35, pp. 445–502. DOI: 10.1146/annurev.astro.35.1.445.

Urry, C. Megan and P. Padovani (Sept. 1995). “Unified Schemes for Radio-Loud Active Galactic Nuclei”. In: *Publications of the Astronomical Society of the Pacific* 107, p. 803. DOI: 10.1086/133630.

Uttley, P. et al. (May 2002). “Measuring the broad-band power spectra of active galactic nuclei with RXTE”. In: *Monthly Notices of the Royal Astronomical Society* 332.1, pp. 231–250. DOI: 10.1046/j.1365-8711.2002.05298.x.

Uttley, Phil et al. (2023). *Large and complex X-ray time lags from black hole accretion disks with compact inner coronae*. arXiv: 2312.08302 [astro-ph.HE].

Wilkins, D R (Aug. 2019). “Low-frequency X-ray timing with Gaussian processes and reverberation in the radio-loud AGN 3C 120”. In: *Monthly Notices of the Royal Astronomical Society* 489.2, pp. 1957–1972. ISSN: 1365-2966. DOI: 10.1093/mnras/stz2269.

Winkler Jr., P. F. and A. E. White (Aug. 1975). “A sudden increase in the X-ray flux from Centaurus A.” In: *The Astrophysical Journal Letters* 199, pp. L139–L142. DOI: 10.1086/181868.

York, Donald G. et al. (Sept. 2000). “The Sloan Digital Sky Survey: Technical Summary”. In: *Astronomical Journal* 120.3, pp. 1579–1587. DOI: 10.1086/301513.

Yu, Weixiang and Gordon T. Richards (Jan. 2022). *EzTao: Easier CARMA Modeling*. Astrophysics Source Code Library, record ascl:2201.001. ascl: 2201.001.

Yu, Weixiang et al. (June 2022). “Examining AGN UV/optical Variability Beyond the Simple Damped Random Walk”. In: *American Astronomical Society Meeting #240*. Vol. 240. American Astronomical Society Meeting Abstracts, 129.01, p. 129.01.

Yu, Weixiang et al. (2025). *Scalable and Robust Multiband Modeling of AGN Light Curves in Rubin-LSST*. DOI: https://doi.org/10.3847/1538-4357/ae28d3. arXiv: 2511.21479 [astro-ph.GA].

Zeldovich, Ya. B. (Apr. 1964). *Fate of a star and liberation of gravitational energy at accretion <sud’ba zbezdy i vydeleniye gravitatsionnoy energii pri akkretsii<*. NASA Technical Translation, NASA/STI Accession number: 19640015209.

A Derivation of the Damped Random Walk Transition Probability

The transition probability of Equation 4.4 follows from solving the Ornstein-Uhlenbeck stochastic differential equation. Here X_t denotes the process $f(t)$ of Equation 4.1, rewritten using the standard Itô notation $\mathrm{d}X_t$ and the damping rate $\theta = 1/\tau$:

$$\mathrm{d}X_t = \theta\,(\mu - X_t)\,\mathrm{d}t + \sigma\sqrt{2\theta}\,\mathrm{d}W_t, \tag{A.1}$$

where μ is the long-term mean, θ the mean-reversion rate, σ the stationary standard deviation, and W_t a Wiener process. The equation is solved by variation of parameters, using the integrating factor $e^{\theta t}$. Defining $Y_t = X_t\,e^{\theta t}$ and applying Itô's lemma,

$$\mathrm{d}Y_t = \theta X_t\,e^{\theta t}\,\mathrm{d}t + e^{\theta t}\,\mathrm{d}X_t = e^{\theta t}\,\theta\mu\,\mathrm{d}t + \sigma\sqrt{2\theta}\,e^{\theta t}\,\mathrm{d}W_t, \tag{A.2}$$

where the drift terms in X_t cancel on substituting Equation A.1. Integrating from 0 to t,

$$X_t\,e^{\theta t} - X_0 = \theta\mu\int_0^t e^{\theta s}\,\mathrm{d}s + \sigma\sqrt{2\theta}\int_0^t e^{\theta s}\,\mathrm{d}W_s = \mu\left(e^{\theta t} - 1\right) + \sigma\sqrt{2\theta}\int_0^t e^{\theta s}\,\mathrm{d}W_s, \tag{A.3}$$

and multiplying through by $e^{-\theta t}$ gives the solution

$$X_t = X_0\,e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \sigma\sqrt{2\theta}\int_0^t e^{-\theta(t-s)}\,\mathrm{d}W_s. \tag{A.4}$$

The first two terms are deterministic, since X_0, μ, and θ are fixed. The final term is an Itô integral of a deterministic function against the Wiener process, and is therefore Gaussian: it can be approximated as a weighted sum of independent Gaussian increments $\mathrm{d}W_s \sim \sqrt{\mathrm{d}s}\,\varepsilon_s$ with $\varepsilon_s \sim \mathcal{N}(0,1)$, and any linear combination of Gaussians is itself Gaussian. The state X_t conditioned on X_0 is thus normally distributed, and it remains to compute its mean and variance.

The mean is obtained by taking the expectation of Equation A.4. The Itô integral has zero mean, so

$$\mathbb{E}[X_t \mid X_0] = X_0\,e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right). \tag{A.5}$$

The variance comes from the Itô isometry, which turns the variance of the stochastic integral into an ordinary integral of the squared integrand,

$$\mathrm{Var}[X_t \mid X_0] = 2\theta\sigma^2 \int_0^t e^{-2\theta(t-s)}\,\mathrm{d}s = 2\theta\sigma^2 \cdot \frac{1-e^{-2\theta t}}{2\theta} = \sigma^2\left(1-e^{-2\theta t}\right). \tag{A.6}$$

Collecting the mean and variance, the transition probability is

$$X_t \mid X_0 \;\sim\; \mathcal{N}\big(\mu + (X_0-\mu)\,e^{-\theta t},\;\; \sigma^2\left(1-e^{-2\theta t}\right)\big). \tag{A.7}$$

Writing the elapsed time as Δt and the damping timescale as $\tau = 1/\theta$, this is Equation 4.4. The factor $\sqrt{2\theta}$ in the driving term of Equation A.1 is what makes σ the stationary standard deviation directly: as $t \to \infty$ the variance tends to σ^2, independently of τ.

B Neural Stochastic Flow in Depth

This appendix supplements the description of the reconstruction method in Chapter 5, following Kiyohara et al. 2025. It first gives the architecture of the bridge model, the counterpart of the flow whose construction is set out in the main text, and then makes explicit the triplet-consistency loss through which the two models are trained jointly.

B.1 The Bridge Architecture

The flow F_θ of Chapter 5 models the transition between two states separated by an elapsed time, and its construction, a state-dependent Gaussian base distribution followed by conditioned affine-coupling layers, is given in Equations 5.5 and 5.2. The bridge q_ξ of Equation 5.6 is built in the same way, but it models a different conditional where the intermediate state x_j given *both* bracketing states x_i and x_k and the three times, rather than a future state given a single past one. It is this conditioning on both endpoints that makes the bridge suitable for gap filling, where the observations on either side of a gap are known.

The bridge conditioner is

$$c_b := \big(x_i,\, x_k,\, \Delta t_{ij},\, \Delta t_{jk}\big), \tag{B.1}$$

carrying both surrounding states and the two elapsed times to the intermediate point, in place of the single pair $(x_s, \Delta t)$ used by the flow. As with the flow, a sample of the intermediate state is drawn by first initializing a Gaussian base distribution and then transforming it through a sequence of L conditioned affine-coupling layers (Dinh, Krueger, and Bengio 2015),

$$x_j = G_\xi(z_b, c_b) = g_L(\,\cdot\,; c_b, \xi_L) \circ \cdots \circ g_1(z_b; c_b, \xi_1), \qquad z_b \sim \mathcal{N}\big(\mu_b(c_b),\, \sigma_b^2(c_b)\big), \tag{B.2}$$

where the mean and variance of the base distribution and the scale and shift of each coupling layer are produced by multilayer perceptrons taking the conditioner c_b as input, in the same manner as Equations 5.5 and 5.2 for the flow. The bridge and the flow therefore share the same affine-coupling structure and differ only in what they are conditioned on, so that both admit a tractable likelihood through the change-of-variables formula of Equation 5.1 and can be sampled in a single pass.

B.2 The Triplet-Consistency Loss

The training objective of Equation 5.10 combines a data term and a flow-consistency term, both evaluated on the triplets (t_i, t_j, t_k) sampled from each light curve. This section writes out both terms explicitly for a single triplet, the data term as a sum of transition log-densities and the flow-consistency term as a difference of log-densities computed on Monte Carlo samples drawn from the models.

B.2.1 Data Term

The data term is the negative log-likelihood of the transitions of the triplet under the flow. All three ordered pairs are included, the two one-step transitions and the direct endpoint transition, so that

$$\ell_{\text{data}} = -\frac{1}{3}\Big[\log p_\theta(x_j \mid x_i, \Delta t_{ij}) + \log p_\theta(x_k \mid x_j, \Delta t_{jk}) + \log p_\theta(x_k \mid x_i, \Delta t_{ik})\Big], \tag{B.3}$$

each log-density being evaluated through the change-of-variables formula of Equation 5.1. Including the endpoint transition $x_i \to x_k$ alongside the two one-step transitions is what trains the flow on the full span of the triplet directly, rather than leaving the long transition to be reproduced by composing shorter ones. The data term of the objective is the expectation of ℓ_{data} over the triplets.

B.2.2 Flow-Consistency Term

The flow-consistency term of Chapter 5 is written in Equations 5.8 and 5.9 as two variational bounds on the Kullback-Leibler divergence between the one-step and two-step transitions. What follows makes explicit how those bounds are evaluated in practice, as a difference of log-densities computed on Monte Carlo samples, for a single triplet.

Consider one light curve and three times $t_i < t_j < t_k$, with corresponding states x_i, x_j, and x_k. The flow defines the one-step transition densities

$$p_\theta(x_k \mid x_i, \Delta t_{ik}), \qquad p_\theta(x_j \mid x_i, \Delta t_{ij}), \qquad p_\theta(x_k \mid x_j, \Delta t_{jk}), \tag{B.4}$$

and the bridge defines the intermediate conditional $q_\xi(x_j \mid x_i, x_k)$. The joint distribution over the intermediate and final states (x_j, x_k) can be written in two ways, and the loss enforces that they assign consistent log-densities to samples generated from the models.

Forward Direction

In the forward direction the final state is drawn from the one-step flow and the intermediate state from the bridge,

$$x_k \sim p_\theta(\,\cdot \mid x_i, \Delta t_{ik}), \qquad x_j \sim q_\xi(\,\cdot \mid x_i, x_k), \tag{B.5}$$

which corresponds to the factorization

$$p_\theta(x_k \mid x_i, \Delta t_{ik})\; q_\xi(x_j \mid x_i, x_k). \tag{B.6}$$

The same joint distribution can be written as a two-step flow through the intermediate state,

$$p_\theta(x_j \mid x_i, \Delta t_{ij})\, p_\theta(x_k \mid x_j, \Delta t_{jk}). \tag{B.7}$$

Given the sample (x_j, x_k), the contribution to the loss is the difference of the log-densities of the two factorizations,

$$\ell_{\text{1-to-2}} = \log p_\theta(x_k \mid x_i, \Delta t_{ik}) + \log q_\xi(x_j \mid x_i, x_k) - \log p_\theta(x_j \mid x_i, \Delta t_{ij}) - \log p_\theta(x_k \mid x_j, \Delta t_{jk}). \tag{B.8}$$

Reverse Direction

In the reverse direction both states are generated by two consecutive flow transitions,

$$x_j \sim p_\theta(\,\cdot \mid x_i, \Delta t_{ij}), \qquad x_k \sim p_\theta(\,\cdot \mid x_j, \Delta t_{jk}), \tag{B.9}$$

and this is compared with the factorization that uses the bridge and the direct one-step flow, $q_\xi(x_j \mid x_i, x_k)\, p_\theta(x_k \mid x_i, \Delta t_{ik})$. The corresponding log-density difference is

$$\ell_{\text{2-to-1}} = \log p_\theta(x_j \mid x_i, \Delta t_{ij}) + \log p_\theta(x_k \mid x_j, \Delta t_{jk}) - \log q_\xi(x_j \mid x_i, x_k) - \log p_\theta(x_k \mid x_i, \Delta t_{ik}). \tag{B.10}$$

Combined Loss

The total consistency loss for the triplet is the sum of the two directions,

$$\ell_{\text{flow}} = \ell_{\text{1-to-2}} + \ell_{\text{2-to-1}}, \tag{B.11}$$

and the flow-consistency term of the objective is its expectation over triplets and over the Monte Carlo samples drawn from the models. Combined with the data term of Equation B.3, the full per-triplet loss is

$$\ell = \ell_{\text{data}} + \lambda\, \ell_{\text{flow}}, \tag{B.12}$$

matching the objective of Equation 5.10. When the flow and bridge are mutually consistent, the two factorizations of the joint distribution over (x_j, x_k) coincide, and the expected log-density differences of Equations B.8 and B.10 vanish.