Modeling the Multiwavelength Variability of Mrk 335 Using Gaussian Processes

Ryan-Rhys Griffiths[1] , Jiachen Jiang[2,3], Douglas J. K. Buisson[4], Dan Wilkins[5], Luigi C. Gallo[6], Adam Ingram[7] , Alpha A. Lee[1], Dirk Grupe[8] , Erin Kara[9], Michael L. Parker[10] , William Alston[11], Anthony Bourached[12], George Cann[13], Andrew Young[14], and S. Komossa[15]

[1] Department of Physics, University of Cambridge, JJ Thomson Avenue, CB3 0HE Cambridge, UK
[2] Department of Astronomy, Tsinghua University, Beijing 100084, People's Republic of China
[3] Tsinghua Center for Astrophysics, Tsinghua University, Beijing 100084, People's Republic of China
[4] Department of Astronomy, University of Southampton, Southampton SO17 1BJ, UK
[5] Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA
[6] Department of Astronomy and Physics, Saint Mary's University, 923 Robie Street, Halifax, NS, B3H 3C3, Canada
[7] Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK
[8] Department of Physics, Earth Sciences, and Space System Engineering, Morehead State University, Morehead, KY 40514, USA
[9] Kavli Institute for Astrophysics and Space Research, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[10] Institute of Astronomy, University of Cambridge, Madingley Road, CB3 0HA Cambridge, UK
[11] European Space Astronomy Centre (ESAC), Science Operations Department, E-28692 Villanueva de la Canada, Madrid, Spain
[12] UCL Queen Square Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK
[13] Department of Space and Climate Physics, University College London, Mullard Space Science Laboratory, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK
[14] School of Physics, Tyndall Avenue, University of Bristol, Bristol BS8 1TH, UK
[15] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53111 Bonn, Germany

Abstract

The optical and UV variability of the majority of active galactic nuclei may be related to the reprocessing of rapidly changing X-ray emission from a more compact region near the central black hole. Such a reprocessing model would be characterized by lags between X-ray and optical/UV emission due to differences in light travel time. Observationally, however, such lag features have been difficult to detect due to gaps in the lightcurves introduced through factors such as source visibility or limited telescope time. In this work, Gaussian process regression is employed to interpolate the gaps in the Swift X-ray and UV lightcurves of the narrow-line Seyfert 1 galaxy Mrk 335. In a simulation study of five commonly employed analytic Gaussian process kernels, we conclude that the Matern $\frac{1}{2}$ and rational quadratic kernels yield the most well-specified models for the X-ray and UVW2 bands of Mrk 335. In analyzing the structure functions of the Gaussian process lightcurves, we obtain a broken power law with a break point at 125 days in the UVW2 band. In the X-ray band, the structure function of the Gaussian process lightcurve is consistent with a power law in the case of the rational quadratic kernel while a broken power law with a break point at 66 days is obtained from the Matern $\frac{1}{2}$ kernel. The subsequent cross-correlation analysis is consistent with previous studies and furthermore shows tentative evidence for a broad X-ray-UV lag feature of up to 30 days in the lag-frequency spectrum where the significance of the lag depends on the choice of Gaussian process kernel.

1. Introduction

Active galactic nuclei (AGNs) show strong and variable emission across multiple wavelengths. The UV emission from an AGN is believed to be dominated by thermal emission from an accretion disk close to the central supermassive black hole (SMBH; e.g., Pringle 1981). The variability of optical and UV AGN[16] emission is stochastic and described by random Gaussian fluctuations (e.g., Welsh et al. 2011; Gezari et al. 2013; Zhu et al. 2016; Sánchez-Sáez et al. 2018; Smith et al. 2018; Xin et al. 2020) with the autocorrelation functions of such fluctuations adhering to the "damped random walk" model. The X-ray emission from an AGN is often found to show faster variability relative to emission at longer wavelengths (e.g., Mushotzky et al. 1993; Gaskell & Klimek 2003) and originates from a more compact region (e.g., Morgan et al. 2008; Chartas et al. 2017).

The relationship between the UV and X-ray emission has been well studied. For instance, correlations between the variability in two energy bands have been seen in some individual sources (e.g., Shemmer et al. 2001; Buisson et al. 2017) while others do not show significant evidence for similar correlation (e.g., Smith & Vaughan 2007; Buisson et al. 2018). In sources where correlation is found, lags that are related to the light travel time between two emission regions are frequently observed. These lags are often found to be on timescales of days and are longer than those predicted by classical disk theories (Shakura & Sunyaev 1973). Such lag amplitudes indicate a disk of size a few times larger than expected (e.g., Edelson et al. 2000; Shappee et al. 2014; Troyer et al. 2016; Buisson et al. 2017). Alternatively, some modified models have been proposed for the underestimation of lags by the classical thin disk model, e.g., disk turbulence (e.g., Cai et al. 2020), additional varying FUV illumination (e.g., Gardner & Done 2017), a tilted or inhomogeneous inner disk (e.g., Dexter & Fragile 2011; Starkey et al. 2017), or an extended coronal region (e.g., Kammoun et al. 2021). Much shorter lags,

---

[16] AGNs with UV and optical luminosity change of more than 1 mag, such as changing-look AGNs, are not discussed in this work; see Jiang et al. (2021) for details.

e.g., hundreds of seconds, in agreement with the Shakura & Sunyaev (1973) model, have been rarely observed by comparison (e.g., in NGC 4395; McHardy et al. 2016).

The Neil Gehrels Swift Observatory has been monitoring the X-ray sky in the past decade in tandem with simultaneous pointings in the optical and UV band. In this work, we focus on the X-ray and UVW2 ($\lambda = 212$ nm) lightcurves of the narrow-line Seyfert 1 galaxy (NLS1; e.g., Gallo 2018) Mrk 335 obtained by XRT and UVOT, the soft X-ray and UV/optical telescopes on Swift. Mrk 335 was one of the brightest X-ray sources prior to 2007, before its flux diminished by $10\times$–$50\times$ its original brightness (Grupe et al. 2007). The X-ray brightness has not recovered since. During this low X-ray flux period, the UV brightness remains relatively unchanged rendering Mrk 335 X-ray weak (Tripathi et al. 2020). The behavior has been explained as a possible collapse of the X-ray corona, e.g., Gallo et al. (2013, 2015; Parker et al. 2014), and/or increased absorption in the X-ray-emitting region, e.g., Grupe et al. (2012; Longinotti et al. 2013, 2019; Parker et al. 2019).

Mrk 335 has been continuously monitored since 2007, making it one of the best-studied AGNs with Swift. Previous studies from the Swift monitoring program can be found in Grupe et al. (2007, 2012), Gallo et al. (2018), Tripathi et al. (2020), and Komossa et al. (2020). The X-rays are constantly fluctuating and regularly display large-amplitude flaring, e.g., Wilkins et al. (2015). The UV are significantly variable, but at a much smaller amplitude than the X-rays. Gallo et al. (2018) found tentative evidence for lags of ≈ 20 days based on cross-correlation analyses, suggesting a potential reprocessing mechanism of the more variable X-ray emission in the UV emitter of this source. One challenge faced by the Swift monitoring program is that the lightcurves are not continuously sampled and hence standard Fourier techniques cannot be applied. This uneven sampling of the lightcurves is imposed by limited telescope time.

In the context of cross-correlation analysis, methods have been developed to address the problem of unevenly sampled lightcurves. In Reynolds (2000), the method of Press et al. (1992) is extended to interpolate the lightcurve gaps using a model of the covariance function, or equivalently the power spectrum, of the lightcurve. In Bond et al. (1998), Miller et al. (2010), and Zoghbi et al. (2013) a maximum likelihood approach is taken to fit models of the lightcurve power spectra, which accounts for the correlation between the lightcurves. In this paper we focus on a relatively new approach to tackle unevenly sampled lightcurves.

Gaussian processes confer a Bayesian nonparametric framework to model general time-series data (Roberts et al. 2013; Tobar et al. 2015) and have proven effective in tasks such as periodicity detection (Durrande et al. 2016) and spectral density estimation (Tobar 2018). More broadly, Gaussian processes (GPs) have recently demonstrated modeling success across a wide range of spatial and temporal application domains including robotics (Deisenroth & Rasmussen 2011; Greeff & Schoellig 2020), Bayesian optimization (Shahriari et al. 2015; Grosnit et al. 2020; Cowen-Rivers et al. 2021; Grosnit et al. 2021), as well as areas of the natural sciences such as molecular machine learning (Nigam et al. 2021; Griffiths & Hernández-Lobato 2020; Moss & Griffiths 2020; Thawani et al. 2020; Griffiths et al. 2021; Hase et al. 2020, Bartók et al. 2010), genetics (Moss et al. 2020), and materials science (Cheng et al. 2020; Zhang et al. 2020). In the context of astrophysics there is a recent trend favoring nonparametric models

such as GPs due to the flexibility afforded when specifying the underlying data modeling assumptions. Applications have arisen in lightcurve modeling (Luger et al. 2021a, 2021b), continuous-time autoregressive moving average (CARMA) processes Yu & Richards (2021), modeling stellar activity signals in radial velocity data (Rajpaul et al. 2015), lightcurve detrending (Aigrain et al. 2016), learning imbalances for variable star classification (Lyon et al. 2020), inferring stellar rotation periods (Angus et al. 2018), estimating the dayside temperatures of hot Jupiters (Pass et al. 2019), exoplanet detection (Czekala et al. 2017; Jones et al. 2017; Gordon et al. 2020; Langellier et al. 2021), spectral modeling (Gibson et al. 2012; Nikolov et al. 2018; Diamond-Lowe et al. 2020), as well as blazar variability studies (Karamanavis 2015, 2017; Covino et al. 2020; Yang et al. 2021).

It has recently been demonstrated in lightcurve simulations by Wilkins (2019) that a GP framework can compute time lags associated with X-ray reverberation from the accretion disk that are longer and observed at lower frequencies than can be measured by applying standard Fourier transform techniques to the longest available continuous segments. It is for this principal reason that we choose to employ GPs for our timing analysis. Further desirable facets of GPs include the fact that, unlike parametric models, they do not make strong assumptions about the shape of the underlying lightcurve (Wang et al. 2012). Additionally, we may perform Bayesian model selection at the level of the covariance function or kernel allowing us to quantitatively compare different models of the lightcurve power spectrum. Finally in the cross-correlation analysis, we may make a weaker modeling assumption than (Zoghbi et al. 2013) in treating the X-ray and UV lightcurves as being independent (Wilkins 2019).

The paper is outlined as follows: In Section 2, we provide the background on GPs including discussion of different kernels as well as Bayesian model selection, the criterion used to choose between kernels. In Section 3 we describe the procedures used to fit GPs to the X-ray and UVW2 bands including aspects such as identification of the flux distribution, consideration of measurement noise, as well as a simulation study to determine the appropriate kernels. In Section 4 we compare the structure functions of the GP-interpolated light-curves with the observational structure functions from Gallo et al. (2018). In Section 5 we present a cross-correlation analysis of the X-ray and UVW2 bands using the GP-interpolated lightcurves. Finally, in Section 6 we provide concluding remarks about the discrepancy between the observational and GP-derived structure functions as well as the implications of the cross-correlation analysis, namely that the broad lag features suggest an extended emission region of the disk in Mrk 335 during the reverberation process. All code for reproducing the analysis is available at https://github.com/Ryan-Rhys/Mrk_335.

2. Gaussian Processes

We may define a GP as a collection of random variables, any finite number of which have a joint Gaussian distribution. When the GP is used as a prior over functions, the aforementioned random variables consist of function values $f(t)$ at different points in time t. In our setting f represents flux or count rate. The GP is characterized by a mean function

$$m(t) = \mathbb{E}[f(t)] \tag{1}$$

and a covariance function

$$k(t, t') = \mathbb{E}[(f(t) - m(t))(f(t') - m(t'))]. \tag{2}$$

The process is written as follows:

$$f(t) \sim \mathcal{GP}(m(t), k(t, t')). \tag{3}$$

The mean function is set to the empirical mean of the standardized observational data in the cases we consider. Standardization, in this case, refers to the common practice of subtracting the mean and dividing by the standard deviation of the data when fitting the GP in order to facilitate the identification of appropriate hyperparameters (Murray 2008). The standardization is reversed once the fitting procedure is complete in order to obtain predictions on the original scale of the data. $m(t) = 0$ will be assumed henceforth for the sake of the current presentation. The covariance function computes the pairwise covariance between two random variables (function values). In the GP literature, the covariance function is commonly referred to as the kernel and is denoted as

$$k(t, t') = \text{cov}(f(t), f(t')). \tag{4}$$

Informally, the kernel is responsible for determining the smoothness of the functions which the GP is capable of fitting. The inductive bias created by the choice of kernel is an important consideration in Gaussian process modeling.

2.1. Kernels

The most widely known kernel is the squared exponential (SE) or radial basis function kernel:

$$k_{\text{SQE}}(t, t') = \sigma_f^2 \exp\left(\frac{-|t - t'|^2}{2\ell^2}\right), \tag{5}$$

where σ_f^2 is the signal amplitude hyperparameter (vertical lengthscale) and ℓ is the (horizontal) lengthscale hyperparameter. For such hyperparameters, we will adopt the notation of θ to represent the set of kernel hyperparameters. It has been argued by Stein (2012) that the smoothness assumptions of the SE kernel are unrealistic for many physical processes. As such, kernels such as the Matern,

$$k_{\text{Matern}}(t, t') = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu} - |t - t'|}{\ell}\right)^\nu K_\nu$$

$$\times \left(\frac{\sqrt{2\nu} - |t - t'|}{\ell}\right), \tag{6}$$

are more commonly seen in machine learning literature. Here K_ν is a modified Bessel function of the second kind, Γ is the gamma function, and ν is a nonnegative parameter of the kernel, which is typically taken to be either $\frac{3}{2}$ or $\frac{5}{2}$ (Rasmussen & Williams 2006). The lengthscale hyperparameter ℓ can be thought of loosely as a decay coefficient for the covariance between inputs as they become increasingly far apart in the input space; the farther apart the inputs are, the less correlated they will be. The final kernel used in this work is the rational quadratic (RQ) kernel,

$$k_{RQ}(t, t') = \left(1 + \frac{|t - t'|^2}{2\alpha\ell^2}\right)^{-\alpha}. \tag{7}$$

where $\alpha, \ell > 0$. The RQ kernel can be viewed as a scale mixture of SE kernels with different characteristic lengthscales. All kernels used in this work are stationary kernels, and as such, it should be stated that this reflects a modeling assumption that the underlying time series is stationary. The extension of the current work to include nonstationary kernels will be discussed in Section 6.

2.2. Prediction with GPs

To illustrate the homoscedastic (constant noise) GP predictive model we use X-ray timing as an example. We wish to model the count rate $f(t)$. We place a GP prior over f,

$$p(\mathbf{f}(\mathbf{t})|\theta) = \mathcal{GP}(0, K(\mathbf{t}, \mathbf{t})), \tag{8}$$

where $\mathbf{f}$ denotes the vector of function values evaluated at the set of times $\{t_i\}_{i=1}^N$, $N \in \mathbb{N}$. $K(\mathbf{t}, \mathbf{t})$ is a kernel matrix where entries are computed by the kernel function as $[K]_{ij} = k(t_i, t_j)$. θ represents the set of kernel hyperparameters. The GP prior is written as

$$p(\mathbf{y}(\mathbf{t})|\theta) = \mathcal{GP}(0, K(\mathbf{t}, \mathbf{t}) + I\sigma_y^2), \tag{9}$$

where $I\sigma_y^2$ represents the variance of the Gaussian noise on the observations $\mathbf{y}$. The applicability of such a noise model, also known as a Gaussian likelihood, will be discussed further in Section 3.2. Once we have observed some data $\mathbf{y}$, the joint distribution over the observed data $\mathbf{y}$ and the predicted function values $\mathbf{f}^*$ at test locations $\mathbf{t}^*$ may be written as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f_*} \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(\mathbf{t}, \mathbf{t}) + \sigma_y^2 I & K(\mathbf{t}, \mathbf{t_*}) \\ K(\mathbf{t_*}, \mathbf{t}) + \sigma y^2 & K(\mathbf{t_*}, \mathbf{t_*}) \end{bmatrix}\right), \tag{10}$$

where $\mathcal{N}$ is the multivariate Gaussian probability density function. The joint prior may be conditioned on the observations through

$$p(\mathbf{f_*}|\mathbf{y}) = \frac{p(\mathbf{f_*}, \mathbf{y})}{p(\mathbf{y})}, \tag{11}$$

which enforces that the joint prior agrees with the observed target values $\mathbf{y}$. The predictive distribution is thus given as

$$p(\mathbf{f_*}|\mathbf{t}, \mathbf{y}, \mathbf{t_*}) = \mathcal{N}(\bar{\mathbf{f}}_*, \text{cov}(\mathbf{f_*})), \tag{12}$$

with the predictive mean at test locations $\mathbf{t_*}$ being

$$\bar{\mathbf{f}}_* = K(\mathbf{t_*}, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \sigma_y^2 I]^{-1}\mathbf{y} \tag{13}$$

and the predictive uncertainty being

$$\text{cov}(\mathbf{f_*}) = K(\mathbf{t_*}, \mathbf{t_*}) - K(\mathbf{t_*}, \mathbf{t})$$

$$\times [K(\mathbf{t}, \mathbf{t}) + \sigma_y^2 I]^{-1} K(\mathbf{t}, \mathbf{t_*}). \tag{14}$$

Analyzing the form of this expression one may notice that the first term $K(\mathbf{t_*}, \mathbf{t_*})$ in the expression for the predictive uncertainty $\text{cov}(\mathbf{f}^*)$ may be viewed as the prior uncertainty and the second term $K(\mathbf{t_*}, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \sigma_y^2 I]^{-1} K(\mathbf{t}, \mathbf{t_*})$ can be thought of as a subtractive factor that accounts for the reduction in uncertainty when observing the data points $\mathbf{y}$.

2.3. Bayesian Model Selection

One desirable property of GPs and Bayesian models in general is the ability to carry out hierarchical modeling

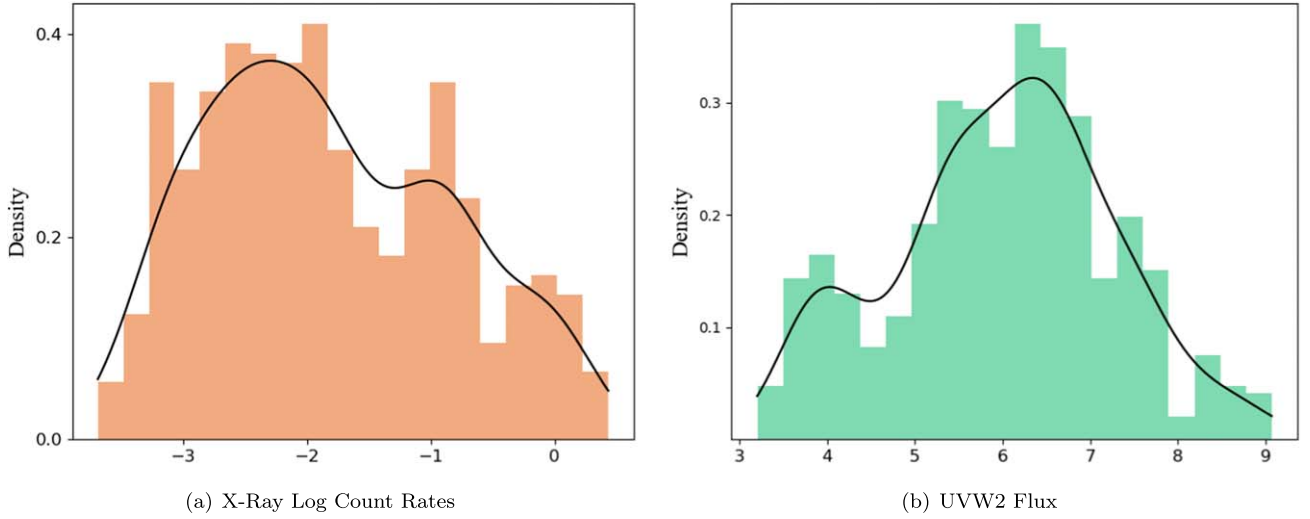(a) X-Ray Log Count Rates

(b) UVW2 Flux

Figure 1. Histograms of the observed Swift X-ray log count rate and UVW2 flux overlaid with Gaussian kernel density estimates. The raw UVW2 flux values have been scaled by $1e^{14}$.

(MacKay 1992; van der Wilk 2019). The three tiers of the modeling hierarchy are:

1. Model Parameters,
2. Model Hyperparameters,
3. Model Structures.

In the case of the nonparametric GP framework, parameters do not have the same meaning as in parametric Bayesian models and are instead obtained from the posterior distribution over functions. Hyperparameters are typically parameters of the kernel function such as signal amplitudes and lengthscales. An important entity for hyperparameter optimization in GPs is the log marginal likelihood or evidence (MacKay 1991):

$$\log p(\boldsymbol{y}|\boldsymbol{t}, \theta) = \underbrace{-\frac{1}{2}\boldsymbol{y}^{\top}(K_{\theta}(\boldsymbol{t}, \boldsymbol{t}) + \sigma_y^2 I)^{-1}\boldsymbol{y}}_{\text{encourages fit with data}}$$

$$\underbrace{-\frac{1}{2}\log|K_{\theta}(\boldsymbol{t}, \boldsymbol{t}) + \sigma_y^2 I|}_{\text{control smodel capacity}} - \frac{N}{2}\log(2\pi), \qquad (15)$$

where N is the number of observations and the subscript notation on the kernel matrix $K_{\theta}(\boldsymbol{t}, \boldsymbol{t})$ is chosen to indicate the dependence on the set of hyperparameters θ. The two terms in the expression for the marginal likelihood embody Occam's Razor (Rasmussen & Ghahramani 2001) in their preference for selecting models of intermediate capacity. The first term in Equation (15) acts as a term that penalizes functions that do not fit the data well whereas the second term acts as a regularizer, disfavoring overly complex models. In this work, kernel hyperparameters are chosen to optimize the marginal likelihood. At the level of model structures, the fit achieved by different kernels can be quantitatively assessed by comparing the values of the optimized log marginal likelihood objective.

3. GP Modeling of Mrk 335

In this paper we consider the Swift X-ray and UVW2 lightcurves in time bins of one day. We refer the reader to Gallo et al. (2018) for details of the data reduction processes. The observational measurements used in this work run from 54327

to 58626 modified Julian days and comprise 509 data points for the X-ray band and 498 data points for the UVW2 band. We consider the latest UVOT sensitivity calibration file ("swusen-scorr20041120v006.fits") to account for the sensitivity loss with time in the UVW2 band.[17]

3.1. Identifying the Flux Distribution

In order to assess the applicability of GPs in modeling the flux distribution of the X-ray and UVW2 bands of Mrk 335, we perform a series of graphical distribution tests to determine the sample distribution. The histograms of the log count rates for the X-ray, and flux for the UV bands, of Mrk 335 are shown in Figure 1. The histograms show that the distribution of the UVW2 flux is approximately Gaussian-distributed whereas the X-ray count rate distribution appears to be log-Gaussian-distributed, in line with the general observation of Uttley & McHardy (2005) that fluxes from accreting black holes tend to follow log-Gaussian distributions. We provide further graphical distribution tests based on probability–probability (PP) plots and empirical cumulative distribution functions (ECDFs) in Appendix A.

Furthermore, following Wilkins (2019) we perform a Kolmogorov–Smirnov test for goodness of fit where the null hypothesis is that the sample was drawn from a Gaussian distribution. For the UVW2 flux values we obtain a p-value of 0.164. We obtain a p-value of $1.017e^{-20}$ for the raw X-ray count rates and a p-value of 0.028 for the log-transformed X-ray count rates. As such, we cannot reject the null hypothesis that either UVW2 flux or log-transformed X-ray count rates are drawn from a Gaussian distribution at the 1% level of significance. We may however reject the null hypothesis in the case of the raw X-ray count rates, providing evidence that the raw X-ray count rates should be log-transformed in order to be well modeled by a Gaussian distribution. As such, we log transform the raw X-ray count rates and leave the UVW2 flux values unchanged.

---

[17] The most up-to-date calibration files may be found at https://heasarc.gsfc.nasa.gov/docs/heasarc/caldb/swift. We consider only UVW2 data collected by UVOT because the UVW2 filter was most frequently used in the archival observations.
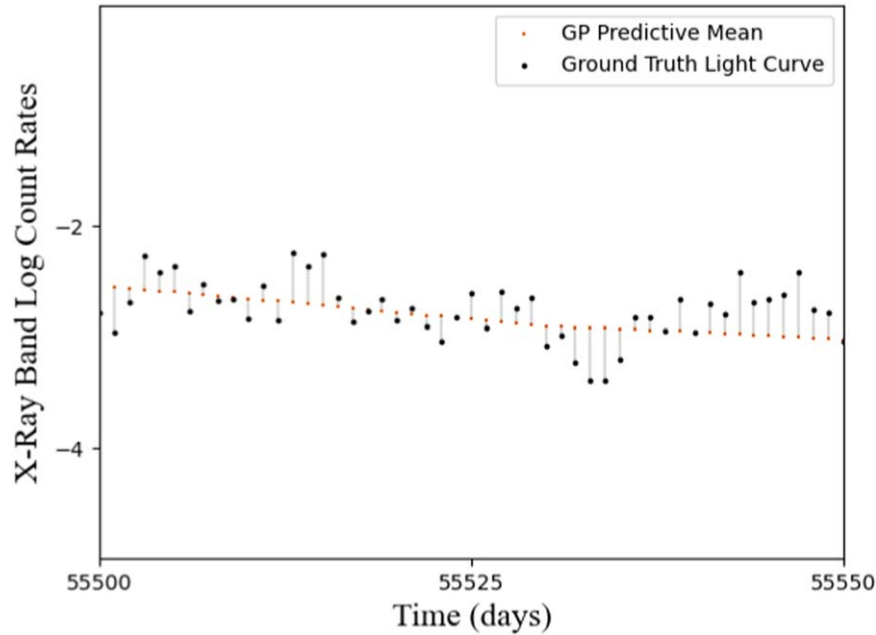
Figure 2. Residual plot. The normalized RSS metric is the sum of squared residuals divided by the total number of discretized points (4390) comprising the simulated lightcurve. A residual in this case represents the difference between the Gaussian process predictive mean and the ground-truth value of the simulated lightcurve.

3.2. Noise

As noted by Wilkins (2019) fitting a GP to the logarithm of the count rate is appropriate only in the limit of a large signal-to-noise ratio. In the case of Mrk 335, the Poisson (shot) noise intrinsic to the photon detectors used to obtain the flux measurements is over an order of magnitude smaller than the flux measurement itself. As such the choice of the log-GP would appear to be justified.

3.3. Simulations

We undertake a simulation study in order to quantitatively assess the abilities of different kernels to interpolate gapped simulated lightcurves. Observational power spectral densities (PSDs) of AGNs are well described by (broken) power laws (Mchardy et al. 2004). As such, our simulations employ a power-law PSD with an index fit to the observational data. Our goals with the study are twofold: first, although we cannot be sure of the true PSD for the observational data, we hope that the simulations may afford a good proxy for identifying performant kernels based on the fact that AGNs typically exhibit power-law-like PSDs, and second, we wish to test whether a kernel's ability to reconstruct the full simulated lightcurve correlates with its marginal likelihood value for the gapped data on which it is trained. If there is a correlation, we may use the marginal likelihood as a metric for identifying the appropriate kernel on the observational data.

One thousand simulated lightcurves with gaps are generated for the Mrk 335 X-ray and UV bands using the method of Davies & Harte (1987), first applied in astrophysics by Timmer & König (1995). For each lightcurve we have access to the ground-truth functional form of the lightcurve before the introduction of gaps. Computationally, the ground-truth light-curve is evaluated on a fine, discrete grid of 4390 time points whereas the gapped lightcurves are evaluated on a coarser, unevenly spaced grid of 498 time points for the UV simulations and 509 time points for the X-ray simulations in line with the

number of observational data points. We then quantify how well each GP kernel performs in recovering the ground-truth lightcurve by measuring the normalized residual sum of squared errors,

$$\text{RSS} = \frac{1}{N}\sum_{i=1}^{N}(f(t_i) - y_i)^2, \tag{16}$$

where $f(t_i)$ is the GP prediction at grid point t_i and y_i is the true simulated count rate value. The RSS values are averaged over the 1000 simulated lightcurves. We provide an illustration of the RSS metric in Figure 2. In addition, we compute the averaged negative log marginal likelihood (NLML) for each kernel:

$$\text{NLML} = \frac{1}{2}\boldsymbol{y}^{\top}(K_{\theta}(\boldsymbol{t},\boldsymbol{t})$$
$$+\sigma_y^2 I)^{-1} + \boldsymbol{y}\frac{1}{2}\log|K_{\theta}(\boldsymbol{t},\boldsymbol{t})$$
$$+\sigma_y^2 I| + \frac{N}{2}\log(2\pi). \tag{17}$$

The NLML in this case is the negative of the quantity given in Equation (15). Kernel hyperparameters were selected via optimization of the NLML using the scipy optimizer of GPflow (Matthews et al. 2017). The jitter level was fixed at 0.001, a small positive number to ensure numerical stability. The output values (flux or the logarithm of the count rate) were standardized according to their empirical mean and standard deviation. We use a constant mean function set to the empirical mean of the data following standardization as discussed in Section 2.

We report the results of this simulation study in Table 1. The NLML values show correlation with RSS, thus providing evidence that NLML is an appropriate metric for determining the GP kernel for the real observational data (for which the ground-truth lightcurve is of course not available). A paired t-test was conducted to determine whether the RSS results were significant
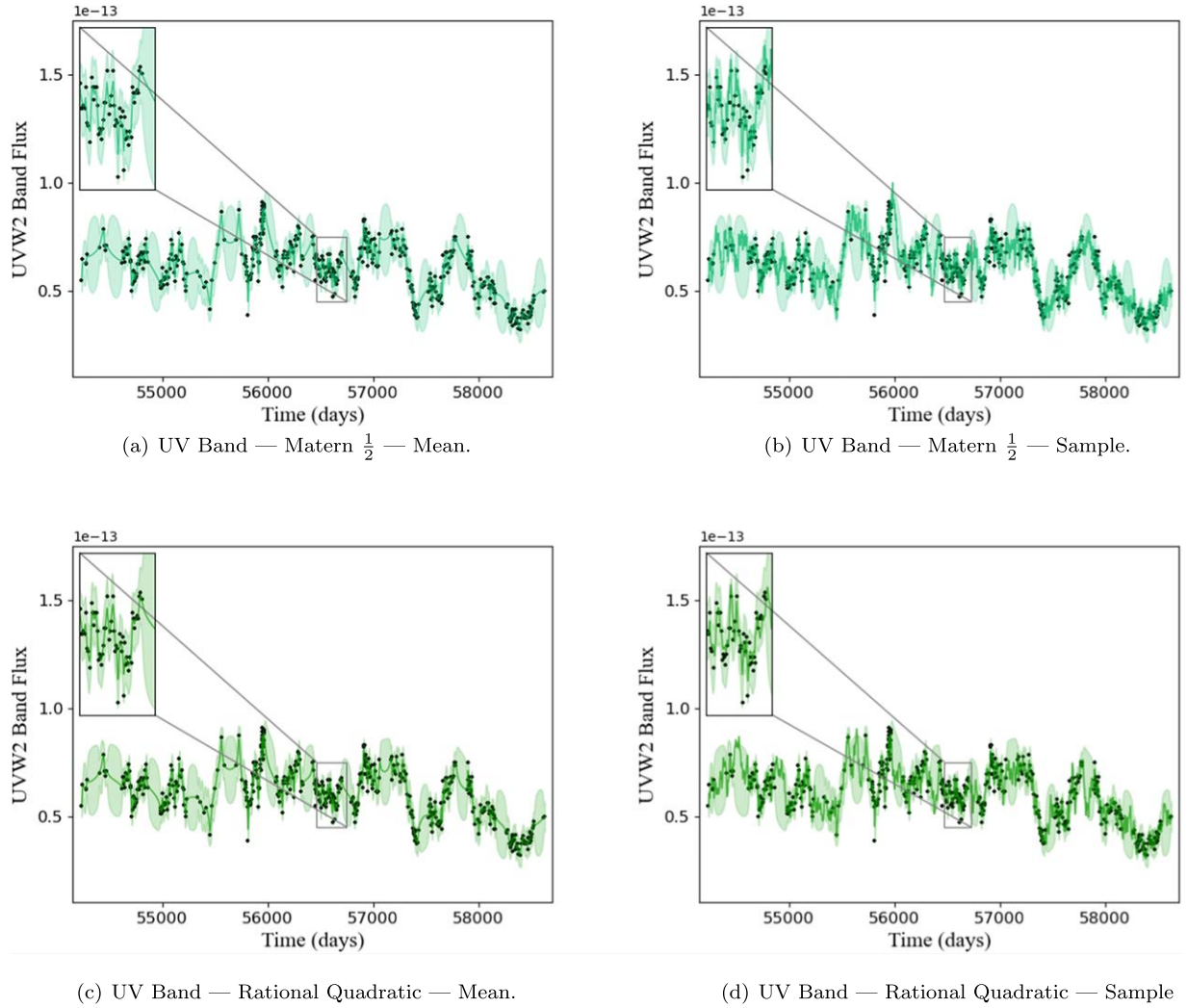
(a) UV Band — Matern $\frac{1}{2}$ — Mean.

(b) UV Band — Matern $\frac{1}{2}$ — Sample.

(c) UV Band — Rational Quadratic — Mean.

(d) UV Band — Rational Quadratic — Sample

Figure 3. GP lightcurves for the UVW2 band. The shaded regions denote the GP 95% confidence interval. We show both the GP mean and a sample from the GP posterior in separate plots. The insets are included to highlight the variability of the fit.

Table 1
Performance Comparison of Kernels Based on the NLML on the Simulated Gapped X-Ray and UV Lightcurves and Normalized Residual Sum of Squared Errors (RSS) on the Ground-truth Simulated Lightcurves

Kernel	NLML	RSS
X-Ray		
Matern $\frac{1}{2}$	**180.2 ± 3.8**	0.121 ± 0.002
Matern $\frac{3}{2}$	420.7 ± 3.3	0.309 ± 0.003
Matern $\frac{5}{2}$	523.5 ± 2.9	0.374 ± 0.003
Rational Quadratic	**184.2 ± 3.6**	**0.117 ± 0.002**
Squared Exponential	632.1 ± 1.5	0.554 ± 0.004
UVW2		
Matern $\frac{1}{2}$	−399.0 ± **5.2**	**2.9 ± 0.08**
Matern $\frac{3}{2}$	−298.3 ± 6.0	7.9 ± 0.25
Matern $\frac{5}{2}$	−219.6 ± 6.5	17.0 ± 0.41
Rational Quadratic	−349.2 ± 5.4	3.4 ± 0.09
Squared Exponential	−65.0 ± 7.4	32.8 ± 0.55

Note. The mean NLML and RSS across 1000 simulations are reported with the standard error. UVW2 RSS values have an exponent of −30. Bold values indicate best performance.

in terms of identifying the best kernel. For the X-ray simulations, a t-statistic of 9 was obtained corresponding to a two-sided p-value of 5^{-20}. For the UVW2 simulations, a t-statistic of −22 was obtained corresponding to a two-sided p-value of 9^{-85}. As such, the null hypothesis that the performance discrepancy between kernels on the RSS metric is due to chance variation across 1000 simulations, may be rejected at the 1% level of significance. We offer further rationalization in Appendix B for why the top two performing kernels in the simulation study are the Matern $\frac{1}{2}$ and RQ kernels.

3.4. GP Fits

The fits to the observational data for the UVW2 and X-ray bands are shown in Figures 3 and 4, respectively. In an analogous fashion to the simulation experiments we evaluate five stationary kernels: Matern $\frac{1}{2}$, Matern $\frac{3}{2}$, Matern $\frac{5}{2}$, RQ, and SE. We choose to display the two kernels, RQ and Matern $\frac{1}{2}$, which performed best in the simulation study in their abilities to model power-law-like PSDs. These kernels also have the most favorable values under the NLML metric for the observational data. We again use a constant mean function set to the empirical mean of the data following standardization. We optimize all
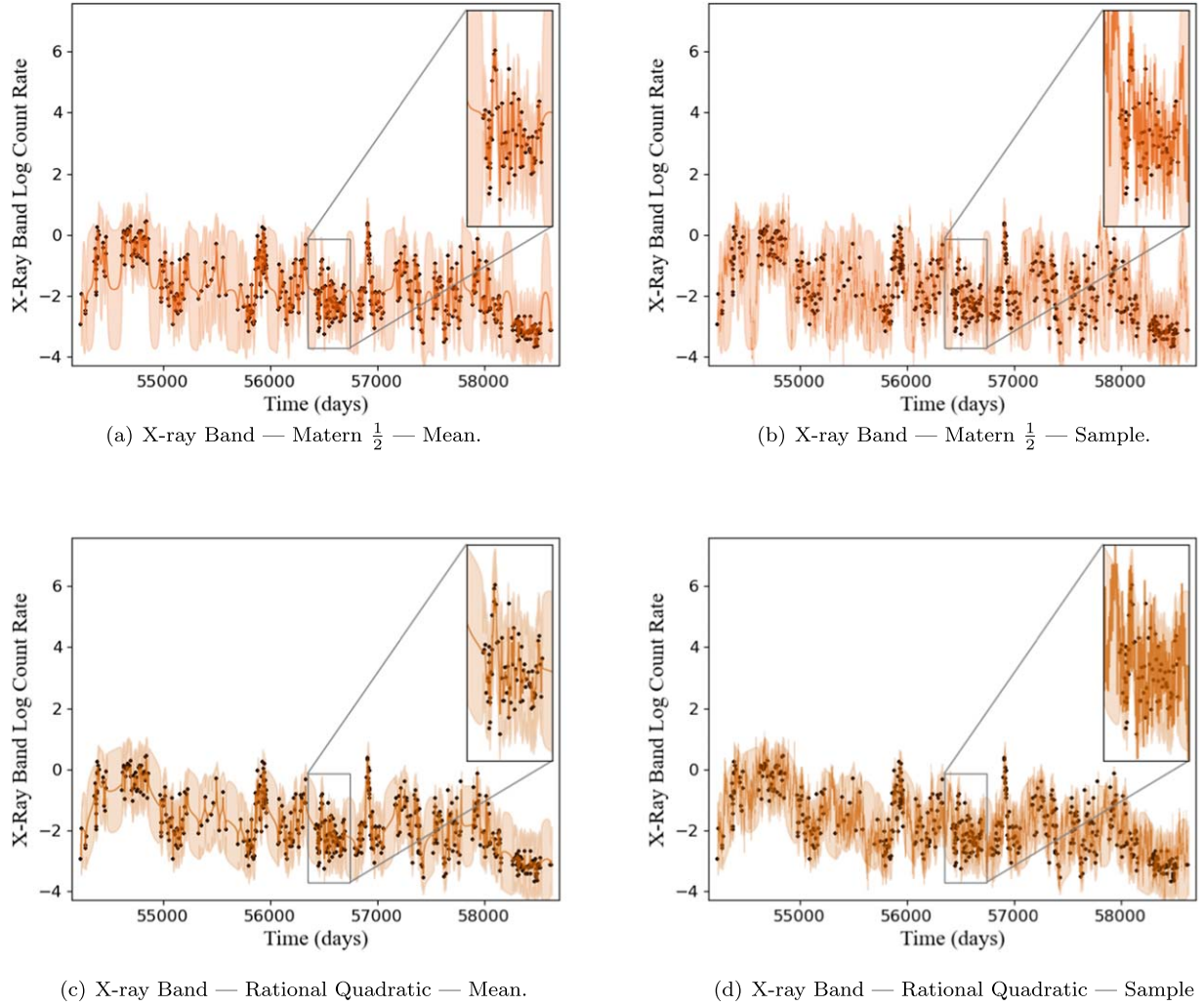
(a) X-ray Band — Matern $\frac{1}{2}$ — Mean.

(b) X-ray Band — Matern $\frac{1}{2}$ — Sample.

(c) X-ray Band — Rational Quadratic — Mean.

(d) X-ray Band — Rational Quadratic — Sample

Figure 4. GP lightcurves for the X-ray band. The shaded regions denote the GP 95% confidence interval. We show both the GP mean and a sample from the GP posterior in separate plots. The insets are included to highlight the variability of the fit.

kernel hyperparameters under the marginal likelihood save for the noise level which we fix to a constant value in the standardized space. This constant noise value is computed by dividing the mean output value in the standardized space by the mean signal-to-noise ratio in the original space.

4. Structure Function Analysis

Ideally we would like to examine the PSD of the GP fits to the observational data. The PSD characterizes the distribution of power over frequencies of a given emission band and properties of the PSD can be linked to underlying physical processes in the accretion disk. Computation of the PSD, while possible, can be complicated by the uneven sampling of the observational data, leading previous studies to instead perform a structure function analysis on the Mrk 335 data (Gallo et al. 2018). While it is possible to extract the PSD from the learned kernel (Wilkins 2019), in this work we choose to perform a structure function analysis of the GP lightcurves in order to compare directly against the results of Gallo et al. (2018). We follow the method described in Simonetti et al. (1985), Hughes et al. (1992), di Clemente et al. (1996), Collier & Peterson (2001), and Gallo et al. (2018). The binned structure function is

defined as

$$\mathrm{SF}(\tau) = \frac{1}{N(\tau)} \sum_i \left[f(t_i) - f(t_i + \tau) \right]^2, \qquad (18)$$

where $\tau = t_j - t_i$ is the distance between pairs of points i and j such that $t_j > t_i$. The structure function is binned according to τ where the centers of each bin are given by $\tau_i = (i - \frac{1}{2})\delta$. δ in this instance is the structure function resolution. We use the same δ as in Gallo et al. (2018), namely 5.3 days for the structure function computation over both the X-ray and UVW2 bands. $f(t_i)$ gives the count rate value at time point t_i and $N(\tau)$ is the number of structure function pairs in each bin i with center τ_i. Accounting for measurement noise by subtracting twice the mean noise variance from each structure function bin, as performed in Gallo et al. (2018), was found to have a negligible effect on the GP structure functions and so we ignore it. As in Gallo et al. (2018), we normalize the structure function values by the global lightcurve variance.

The GP structure functions for the interpolated lightcurves are shown in Figure 5. The 1σ GP error bars are obtained by computing the structure function over 50 samples from the GP posterior. Each sample gives rise to highly similar structure
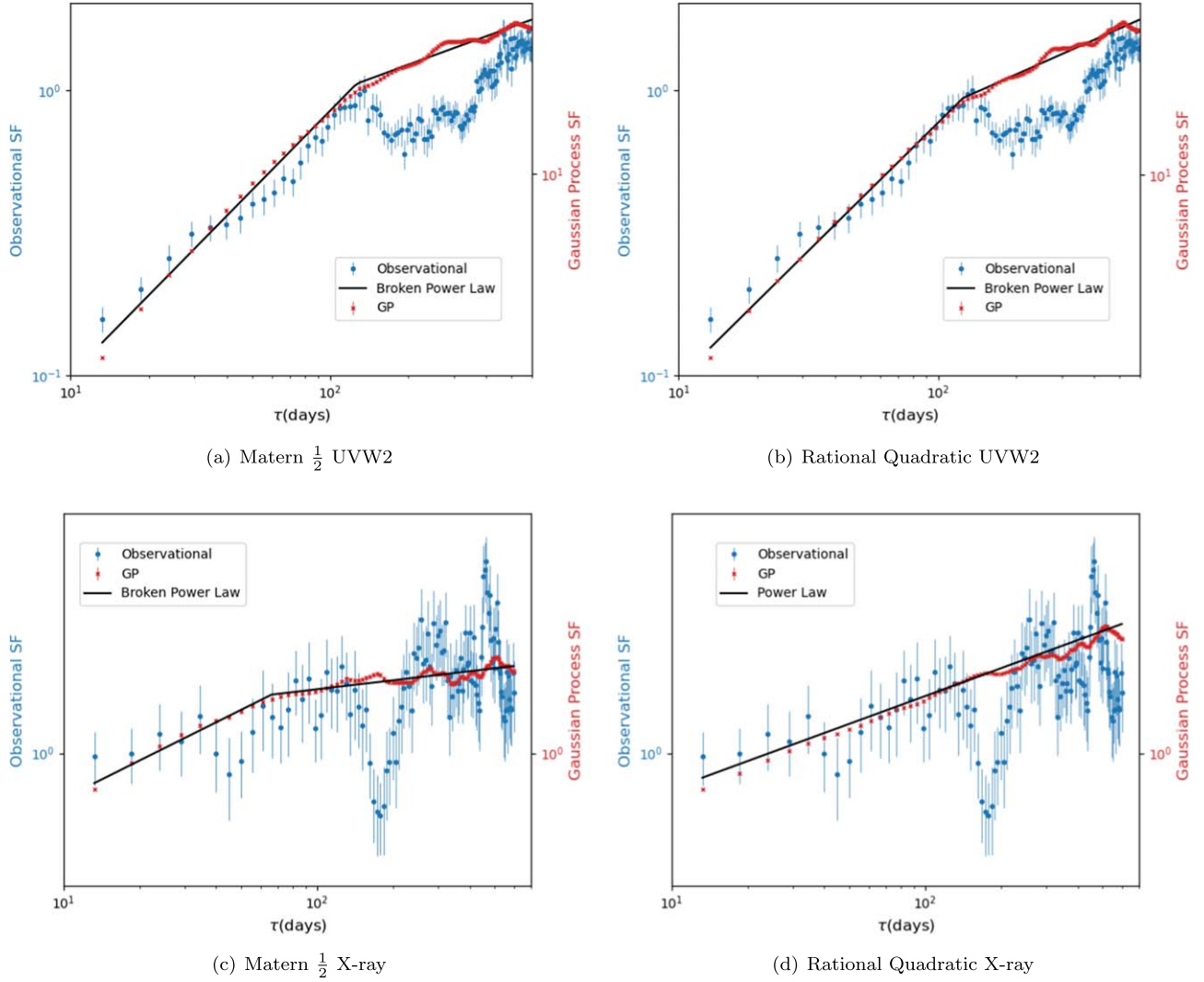
(a) Matern $\frac{1}{2}$ UVW2

(b) Rational Quadratic UVW2

(c) Matern $\frac{1}{2}$ X-ray

(d) Rational Quadratic X-ray

Figure 5. Comparison of observational and GP structure functions. The GP structure functions are consistent with those calculated from the observational data in the non-noise-dominated regions. The dip at ca. 200 days in the observational X-ray structure function is potentially a sampling artifact as demonstrated by simulation in Figure 6.

functions and so the errors are not visible on the plot. The structure functions computed from the observational data, 509 and 498 data points for the X-ray and UV bands of Mrk 335, respectively, are included for reference. In contrast to the GP structure function error bars, in the case of the observational data the error bars are computed as $\sigma_i / \sqrt{(\frac{N_i}{2})}$, where σ_i is the noise standard deviation in bin i and N_i is the number of pairs in bin i.

The GP structure functions are compared against the observational structure functions in Figure 5. In addition, we plot broken power-law fits to the GP structure functions, the parameters of which are given in Table 2. In the UVW2 band, both GP kernels yield structure functions possessing a consistent break point at ca. 125 days. In the X-ray band the Matern $\frac{1}{2}$ kernel yields a break point at 66 days whereas the RQ kernel fit yields an unbroken power law. Given the discrepancy between GP kernels, we do not find definite evidence for a break in the X-ray power law.

Of particular interest is whether the dip in the X-ray structure function is an expected feature of the latent lightcurve or a

Table 2
Parameters for the Broken Power-law Fits to the GP Structure Function

Wave Band	Kernel	α_1	α_2	τ_{char}
UVW2	Matern $\frac{1}{2}$	-0.72 ± 0.03	-0.26 ± 0.01	127 ± 8
UVW2	Rational Quadratic	-0.62 ± 0.01	-0.28 ± 0.01	125 ± 5
X-ray	Matern $\frac{1}{2}$	-0.29 ± 0.03	-0.07 ± 0.004	66 ± 8
X-ray	Rational Quadratic	-0.21 ± 0.002	N/A	N/A

Note. α_1 and α_2 are the indices for the power law before and after the break point τ_{char}. The break point τ_{char} is reported in days. Errors are computed using 200 bootstrap samples of the data points corresponding to the GP structure functions. The X-ray rational quadratic structure function is fit using a power law and as such only has a single index as a parameter. The astropy library (Astropy Collaboration et al. 2013, 2018) is used to compute the (broken) power-law fits using the simplex algorithm and least-squares statistic for optimization with the GP structure function uncertainties used as weights in the fitting.
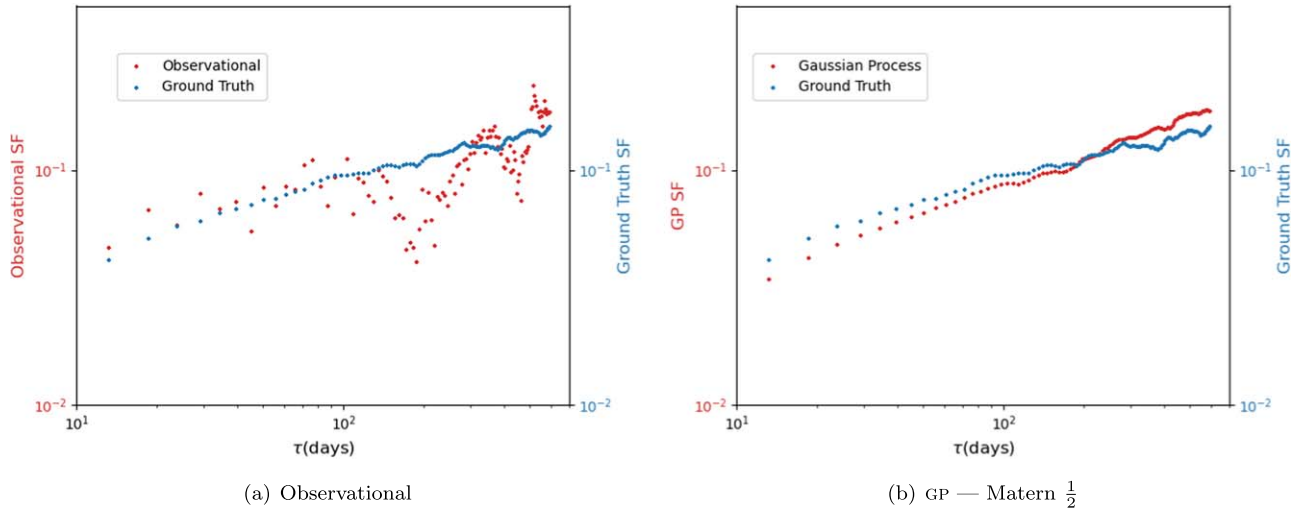
(a) Observational　　　　　　　　　　　　　　　　　　　　　(b) GP — Matern $\frac{1}{2}$

Figure 6. Structure function simulations. Pseudo-observational lightcurves are produced by introducing gaps into the simulated lightcurves. The structure function for the gapped lightcurve is shown in red in (a) whereas the structure function of the GP interpolation is shown in red in (b). Both structure functions are compared against the ground-truth structure function obtained from the full simulated ground-truth lightcurve (no gaps). The dips at $\tau = 200$ days and $\tau = 400$ days in the structure function derived from the gapped observational simulation in 6(a) are artifacts of the uneven sampling.

measurement artifact arising from uneven sampling. In order to assess the potential for the dip to be a sampling artifact, we perform simulations using the Timmer and Konig algorithm from Section 3.3. In this case we compute structure functions of gapped lightcurves and compare against structure functions derived from the ground-truth lightcurves with no gaps. One representative simulation is depicted in Figure 6. In this instance a similar dip to that found in the observational data is observed in the X-ray band simulation. This highlights the possibility that the dip seen in the observational X-ray structure function is a sampling artifact arising from gaps in the lightcurve.

5. Lag and Coherence

In this section, we calculate the coherence between the UVW2 and X-ray emission from Mrk 335 in search of evidence of lag features in the Fourier frequency domain. The coherence and lag spectra are estimated from 1000 pairs of UVW2 and X-ray GP lightcurve samples drawn from the GP posterior for each kernel. The lags are defined as the phase lags divided by the corresponding Fourier frequency. A similar approach has been used in other disciplines (e.g., Fabian et al. 2009; Kara et al. 2013). We consider both Matern $\frac{1}{2}$ and RQ kernels. The results are shown in Figure 7. Positive lags imply that the X-ray variability leads the UVW2 variability. The error bars in the figure are the standard errors of the corresponding measurements for the 1000 samples.

The coherence between the UVW2 and X-ray emission decreases with frequency, suggesting more coherent variability at lower frequency. Positive lag features are shown at the low frequencies in the range $f = 0.005$–0.025 day^{-1}. The absolute value of the lag at $f = 0.0039 \pm 0.0014$ days^{-1} is estimated to be 19 ± 22 days for the Matern $\frac{1}{2}$ kernel applied to both lightcurves and 29 ± 19 days for the RQ kernel; however, both measurements are consistent with zero lag in the 2σ uncertainty range.

Tentative evidence of a shorter time lag at a higher frequency of $f = 0.018 \pm 0.006$ day^{-1} is also found. The longer lag feature at a lower frequency would correspond to a more

extended emission region while the shorter lag feature at a higher frequency would correspond to a more compact region. This could be explained by the presence of an extended UV emission region on the disk where reverberation happens.

Given that the lags are consistent with zero lag within 2σ uncertainty ranges, we conclude that only tentative evidence for a broad lag feature is found by applying GPs to the UVW2 and X-ray lightcurves of Mrk 335. Previous attempts to identify lags between two wavelengths of Mrk 335 based on cross-correlation analysis in the time domain suggest similar results (e.g., Gallo et al. 2018).

6. Conclusions

Following the interpolation of the unevenly sampled light-curves of Mrk 335 using GPs, tentative evidence for broad lag features is found in the Fourier frequency domain. The magnitude of the lags is consistent with previous cross-correlation analyses. In addition, the broad lag features might suggest an extended emission region, e.g., of the disk in Mrk 335 during the reverberation processes. If the corona is compact within 5 R_{g} in Mrk 335 (Wilkins et al. 2015), our data suggest a possibly wide range of UVW2 emission radii.

The structure functions computed from the GP-interpolated lightcurves are consistent with those derived from the observational data and furthermore, illicit potential insights into the properties of the latent lightcurves. In particular, we show through a simulation study that it is possible that dips in the X-ray structure function may be produced by sampling artifacts arising from gaps in the lightcurve. In contrast, the GP structure functions show no dips. While this is not proof that the dip in the observational X-ray structure function is due to a sampling artifact, it does allude to the possibility. The UVW2 GP structure functions do not exhibit strong dependence on the choice of kernel with both Matern $\frac{1}{2}$ and RQ showing up as a broken power law with breaks at 139 and 155 days, respectively. The X-ray structure functions however do show up differences between kernels with the RQ kernel predicting a power law and the Matern 1/2 kernel predicting a broken power law.

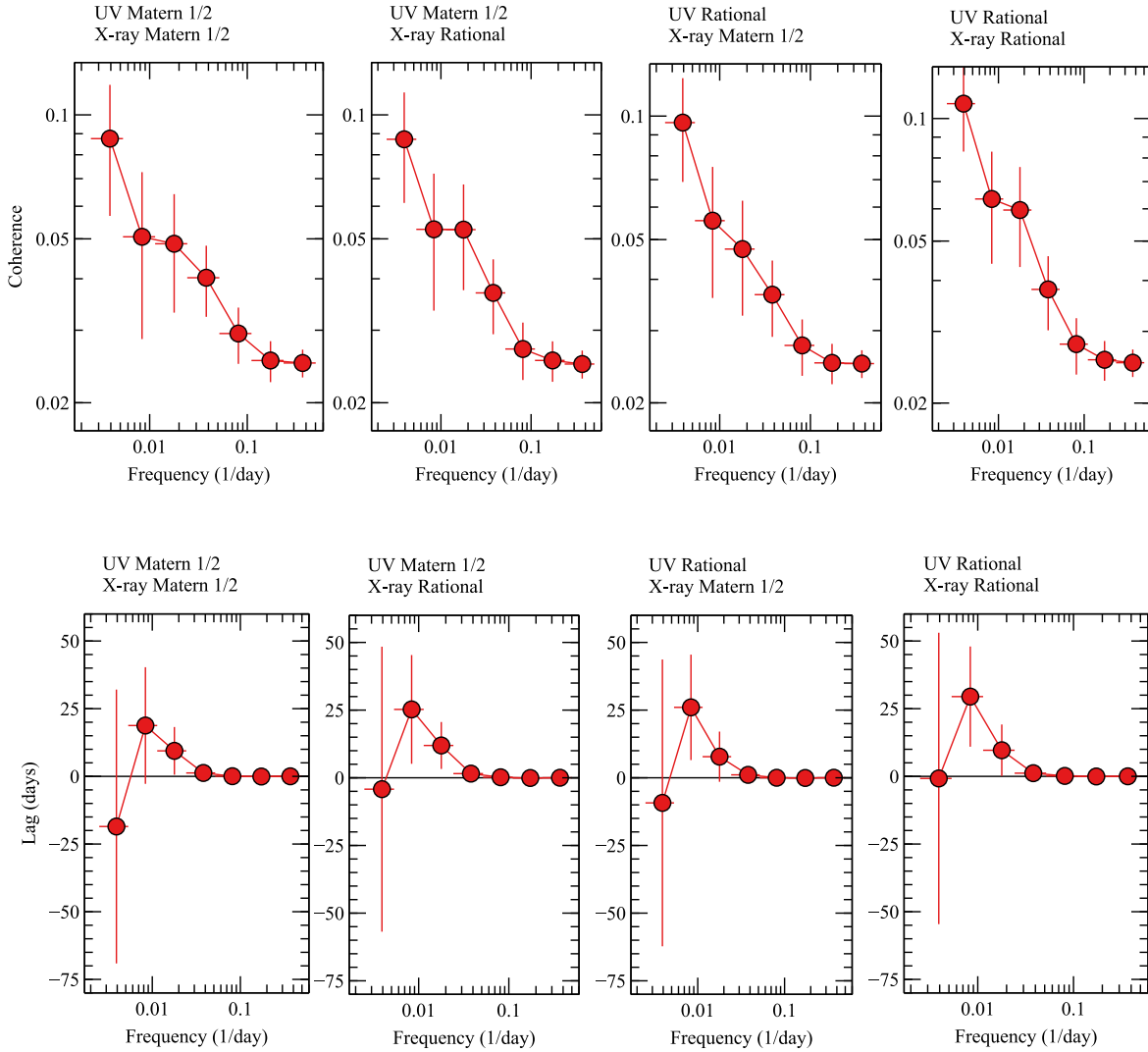**Figure 7.** The coherence and lag spectra for Mrk 335. They are calculated by using 1000 pairs of GP lightcurve samples fit to the observed lightcurves. The error bars are the standard errors of the corresponding measurements for the 1000 samples. Different panels are for different kernels. Positive lags imply that the X-ray band leads the UVW2 band.

From the GP modeling perspective, the ability to carry out Bayesian model selection affords a quantitative means of comparing analytic kernels under the marginal likelihood. It may be possible to incorporate further flexibility into the fitting procedure by making use of more sophisticated methods of kernel design (Duvenaud 2014) in order to allow the assessment of fits of sums and products of analytic kernels or by leveraging advances in transforming GP priors via deep GPs (Damianou & Lawrence 2013) or normalizing flows (Maroñas et al. 2020). Such approaches could be validated using simulation studies. Additionally, modeling the cross-correlation using multi-output GPs (de Wolff et al. 2021) may be an interesting avenue for

comparison against the approach taken here. Lastly, Bayesian spectral density estimation (Tobar 2018) may afford further flexibility through nonparametric modeling of the power spectral density in addition to nonparametric modeling of the lightcurve in the time domain. These improvements in Bayesian modeling machinery may help to minimize model misspecification and as such, enable more robust inferences to be made about the functional forms of the latent lightcurves.
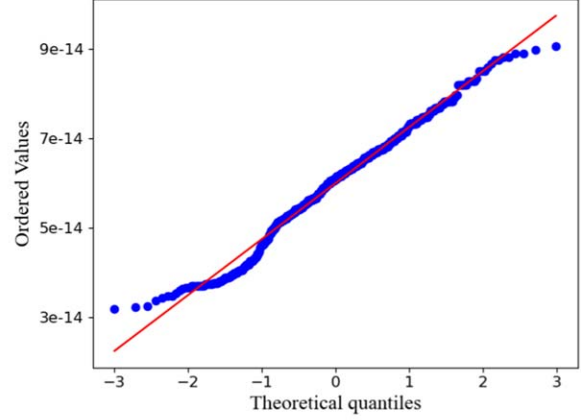
Appendix A
Additional Graphical Tests for Identifying the Flux Distribution

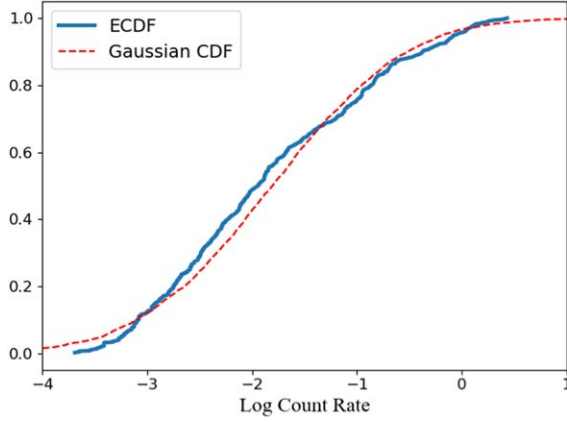In Figure 8 we show P-P plots and ECDFs as graphical distribution tests for Gaussianity. It may be observed qualitatively that both X-ray band log count rates and UVW2 flux are well modeled by a Gaussian distribution.
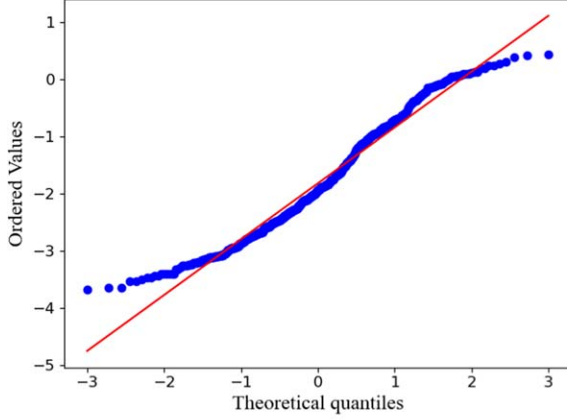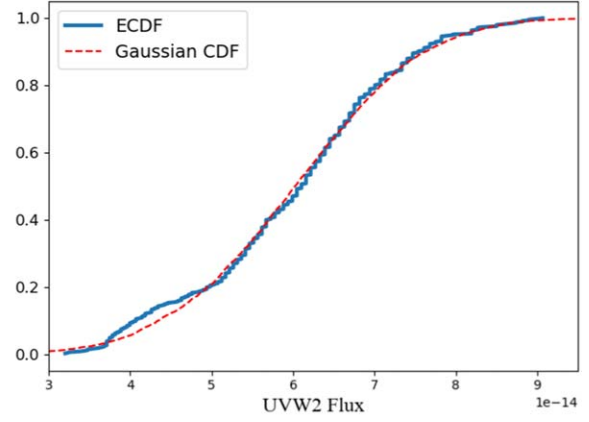

(a) PP plot for X-ray log count rates

(b) PP plot for UVW2 flux

(c) ECDF for X-ray log count rates

(d) ECDF for UVW2 flux

Figure 8. P-P plots and ECDFs for X-ray log count rates and UVW2 flux. Graphical tests of Gaussianity. In the case of the P-P plots, proximity to the line is an indicator of Gaussianity. In the case of the ECDF plots, resemblance to the cumulative distribution function of a Gaussian is indicative of Gaussianity.
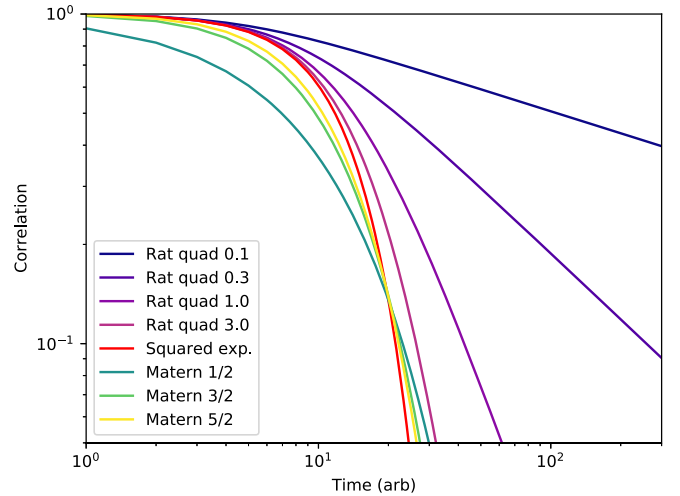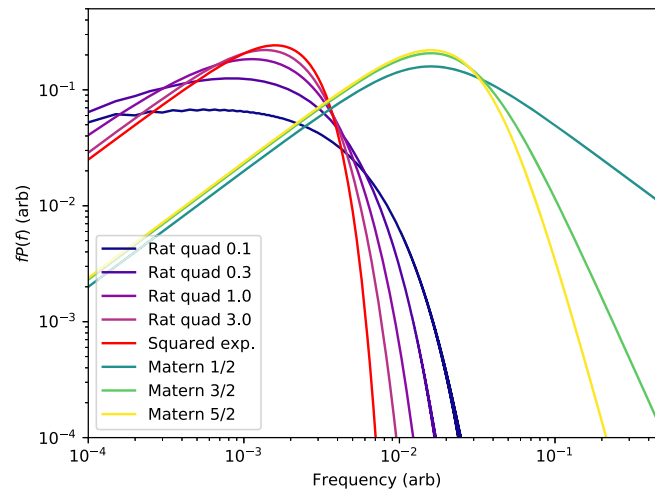
Appendix B
Kernels

In Figure 9, we show plots of analytic kernel autocorrelation functions as well as their associated PSDs.
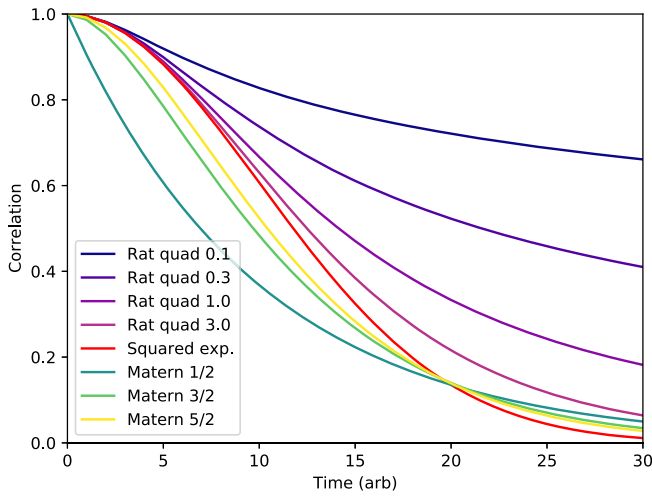
(a) Kernel autocorrelation functions

(b) Kernel log autocorrelation functions

(c) Kernel PSDs

Figure 9. Kernel autocorrelation functions and PSDs. The rational quadratic kernel is plotted for different values of the α parameter. The Matern kernel plots in the PSD figure are offset by a factor of 10 for clarity. A PSD of f^{-2} will match the high-frequency part of the Matern $\frac{1}{2}$ kernel and the rational quadratic is endowed with additional flexibility to model PSDs by virtue of its α parameter. Such characteristics may explain why these kernels are preferred in the simulation study.

ORCID iDs

Ryan-Rhys Griffiths (ID) https://orcid.org/0000-0003-3117-4559
Adam Ingram (ID) https://orcid.org/0000-0002-5311-9078
Dirk Grupe (ID) https://orcid.org/0000-0002-9961-3661
Michael L. Parker (ID) https://orcid.org/0000-0002-8466-7317
S. Komossa (ID) https://orcid.org/0000-0002-9214-4428

References

Aigrain, S., Parviainen, H., & Pope, B. 2016, MNRAS, 459, 2408
Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., & Rajpaul, V. 2018, MNRAS, 474, 2094
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Bartók, A. P., Payne, M. C., Kondor, R., & Csányi, G. 2010, PhRvL, 104, 136403
Bond, J., Jaffe, A. H., & Knox, L. 1998, PhRvD, 57, 2117
Buisson, D., Lohfink, A., Alston, W., et al. 2018, MNRAS, 475, 2306
Buisson, D. J. K., Lohfink, A. M., Alston, W. N., & Fabian, A. C. 2017, MNRAS, 464, 3194
Cai, Z.-Y., Wang, J.-X., & Sun, M. 2020, ApJ, 892, 63
Chartas, G., Krawczynski, H., Zalesky, L., et al. 2017, ApJ, 837, 26
Cheng, B., Griffiths, R.-R., Wengert, S., et al. 2020, Acc. Chem. Res., 53, 1981
Collier, S., & Peterson, B. M. 2001, ApJ, 555, 775

Covino, S., Landoni, M., Sandrinelli, A., & Treves, A. 2020, ApJ, 895, 122
Cowen-Rivers, A. I., Lyu, W., Tutunov, R., et al. 2021, arXiv:2012.03826
Czekala, I., Mandel, K. S., Andrews, S. M., et al. 2017, ApJ, 840, 49
Damianou, A., & Lawrence, N. D. 2013, in Proc. of the 16th Int. Conf. on Artificial Intelligence and Statistics (Scottsdale, AZ: PMLR), 207
Davies, R. B., & Harte, D. 1987, Biometrika, 74, 95
de Wolff, T., Cuevas, A., & Tobar, F. 2021, Neurocomput., 424, 49
Deisenroth, M., & Rasmussen, C. E. 2011, in ICML'11: Proc. of the 28th Int. Conf. on Machine Learning, ed. L. Getoor & T. Scheffer (Madison, WI: Omnipress), 465
Dexter, J., & Fragile, P. C. 2011, ApJ, 730, 36
di Clemente, A., Giallongo, E., Natali, G., Trevese, D., & Vagnetti, F. 1996, ApJ, 463, 466
Diamond-Lowe, H., Berta-Thompson, Z., Charbonneau, D., Dittmann, J., & Kempton, E. M.-R. 2020, AJ, 160, 27
Durrande, N., Hensman, J., Rattray, M., & Lawrence, N. D. 2016, PeerJ Comput. Sci., 2, e50
Duvenaud, D. 2014, PhD thesis, Computational and Biological Learning Laboratory, Univ. Cambridge
Edelson, R., Koratkar, A., Nandra, K., et al. 2000, ApJ, 534, 180
Fabian, A. C., Zoghbi, A., Ross, R. R., et al. 2009, Natur, 459, 540
Gallo, L. 2018, arXiv:1807.09838
Gallo, L. C., Blue, D. M., Grupe, D., Komossa, S., & Wilkins, D. R. 2018, MNRAS, 478, 2557
Gallo, L. C., Fabian, A. C., Grupe, D., et al. 2013, MNRAS, 428, 1191
Gallo, L. C., Wilkins, D. R., Bonson, K., et al. 2015, MNRAS, 446, 633
Gardner, E., & Done, C. 2017, MNRAS, 470, 3591
Gaskell, C. M., & Klimek, E. S. 2003, A&AT, 22, 661
Gezari, S., Martin, D. C., Forster, K., et al. 2013, ApJ, 766, 60
Gibson, N., Aigrain, S., Roberts, S., et al. 2012, MNRAS, 419, 2683
Gordon, T. A., Agol, E., & Foreman-Mackey, D. 2020, AJ, 160, 240
Greeff, M., & Schoellig, A. P. 2020, IEEE Control Syst. Lett., 5, 1121
Griffiths, R.-R., & Hernández-Lobato, J. M. 2020, Chem. Sci., 11, 577
Griffiths, R.-R., Aldrick, A. A., Garcia-Ortegon, M., Lalchand, V. R., & Lee, A A. 2021, arXiv:1910.07779
Grosnit, A., Cowen-Rivers, A. I., Tutunov, R., et al. 2020, arXiv:2012.08240
Grosnit, A., Tutunov, R., & Maraval, A. M. 2021, arXiv:2106.03609
Grupe, D., Komossa, S., & Gallo, L. C. 2007, ApJL, 668, L111
Grupe, D., Komossa, S., Gallo, L. C., et al. 2012, ApJS, 199, 28
Hase, F., Roch, L. M., & Aspuru-Guzik, A. 2020, arXiv:2003.12127
Hughes, P., Aller, H., & Aller, M. 1992, ApJ, 396, 469
Jiang, J., Cheng, H., Gallo, L. C., et al. 2021, MNRAS, 501, 916
Jones, D. E., Stenning, D. C., Ford, E. B., et al. 2017, arXiv:1711.01318
Kammoun, E. S., Dovčiak, M., Papadakis, I. E., Caballero-García, M. D., & Karas, V. 2021, ApJ, 907, 20
Kara, E., Fabian, A. C., Cackett, E. M., et al. 2013, MNRAS, 434, 1129
Karamanavis, V. 2017, Galax, 5, 19
Karamanavis, V. V. 2015, PhD thesis, Universität zu Köln
Komossa, S., Grupe, D., Gallo, L., et al. 2020, A&A, 643, L7
Langellier, N., Milbourne, T., Phillips, D., et al. 2021, AJ, 161, 287
Longinotti, A. L., Kriss, G., Krongold, Y., et al. 2019, ApJ, 875, 150
Longinotti, A. L., Krongold, Y., Kriss, G. A., et al. 2013, ApJ, 766, 104
Luger, R., Foreman-Mackey, D., & Hedges, C. 2021a, arXiv:2102.01697
Luger, R., Foreman-Mackey, D., & Hedges, C. 2021b, arXiv:2102.01774
Lyon, R., Hosenie, Z., Mootovaloo, A., Stappers, B., & McBride, V. 2020, MNRAS, 493, 6050
MacKay, D. J. 1992, Neural Computation, 4, 415
MacKay, D. J. C. 1991, PhD thesis, California Institute of Technology
Maroñas, J., Hamelijnck, O., Knoblauch, J., & Damoulas, T. 2020, arXiv:2011.01596
Matthews, A. G. D. G., van der Wilk, M., Nickson, T., et al. 2017, JMLR, 18, 1
Mchardy, I., Papadakis, I., Uttley, P., Page, M. J., & Mason, K. 2004, MNRAS, 348, 783

McHardy, I. M., Connolly, S. D., Peterson, B. M., et al. 2016, AN, 337, 500
Miller, L., Turner, T., Reeves, J., et al. 2010, MNRAS, 403, 196
Morgan, C. W., Kochanek, C. S., Dai, X., Morgan, N. D., & Falco, E. E. 2008, ApJ, 689, 755
Moss, H., Leslie, D., Beck, D., Gonzalez, J., & Rayson, P. 2020, in Advances in Neural Information Processing Systems 33 (NeurIPS 2020), ed. H. Larochelle et al. (Red Hook, NY: Curran Associates, Inc.), 15476
Moss, H. B., & Griffiths, R.-R. 2020, arXiv:2010.01118
Murray, I. 2008, Introduction to Gaussian Processes (Univ. Toronto)
Mushotzky, R. F., Done, C., & Pounds, K. A. 1993, ARA&A, 31, 717
Nigam, A., Pollice, R., Hurley, M. F. D., et al. 2021, arXiv:2102.11439
Nikolov, N., Sing, D. K., Fortney, J. J., et al. 2018, Natur, 557, 526
Parker, M. L., Longinotti, A. L., Schartel, N., et al. 2019, MNRAS, 490, 683–97
Parker, M. L., Wilkins, D. R., Fabian, A. C., et al. 2014, MNRAS, 443, 1723
Pass, E. K., Cowan, N. B., Cubillos, P. E., & Sklar, J. G. 2019, MNRAS, 489, 941
Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992, ApJ, 385, 404
Pringle, J. E. 1981, ARA&A, 19, 137
Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, MNRAS, 452, 2269
Rasmussen, C. E., & Ghahramani, Z. 2001, in Advances in Neural Information Processing Systems 13 (NIPS 2000) (Red Hook, NY: Curran Associates, Inc.), 294
Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning (Cambridge, MA: MIT Press)
Reynolds, C. S. 2000, ApJ, 533, 811
Roberts, S., Osborne, M., Ebden, M., et al. 2013, RSPTA, 371, 20110550
Sánchez-Sáez, P., Lira, P., Mejía-Restrepo, J., et al. 2018, ApJ, 864, 87
Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. 2015, IEEEP, 104, 148
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 500, 33
Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, ApJ, 788, 48
Shemmer, O., Romano, P., Bertram, R., et al. 2001, ApJ, 561, 162
Simonetti, J., Cordes, J., & Heeschen, D. 1985, ApJ, 296, 46
Smith, K. L., Mushotzky, R. F., Boyd, P. T., et al. 2018, ApJ, 857, 141
Smith, R., & Vaughan, S. 2007, MNRAS, 375, 1479
Starkey, D., Horne, K., Fausnaugh, M. M., et al. 2017, ApJ, 835, 65
Stein, M. L. 2012, Interpolation of Spatial Data: Some Theory for Kriging (Berlin: Springer Science Business Media)
Thawani, A. R., Griffiths, R.-R., Jamasb, A., et al. 2020, arXiv:2008.03226
Timmer, J., & König, M. 1995, A&A, 300, 707
Tobar, F. 2018, in Advances in Neural Information Processing Systems 31 (NeurIPS 2018), ed. S. Bengio et al. (Red Hook, NY: Curran Associates, Inc.), 10127
Tobar, F., Bui, T. D., & Turner, R. E. 2015, in Advances in Neural Information Processing Systems 28 (NIPS 2015) (Red Hook, NY: Curran Associates Inc.), 3501
Tripathi, S., McGrath, K. M., Gallo, L. C., et al. 2020, MNRAS, 499, 1266
Troyer, J., Starkey, D., Cackett, E. M., et al. 2016, MNRAS, 456, 4040
Uttley, P., & McHardy, I. M. 2005, MNRAS, 363, 586
van der Wilk, M. 2019, PhD thesis, Univ. Cambridge
Wang, Y., Khardon, R., & Protopapas, P. 2012, ApJ, 756, 67
Welsh, B. Y., Wheatley, J. M., & Neil, J. D. 2011, A&A, 527, A15
Wilkins, D. R. 2019, MNRAS, 489, 1957
Wilkins, D. R., Gallo, L. C., Grupe, D., et al. 2015, MNRAS, 454, 4440
Xin, C., Charisi, M., Haiman, Z., & Schiminovich, D. 2020, MNRAS, 495, 1403
Yang, S., Yan, D., Zhang, P., Dai, B., & Zhang, L. 2021, ApJ, 907, 105
Yu, W., & Richards, G. T. 2021, AAS Meeting Abstracts, 237, 541.08
Zhang, Y., Tang, Q., Zhang, Y., et al. 2020, NatCo, 11, 1706
Zhu, F.-F., Wang, J.-X., Cai, Z.-Y., & Sun, Y.-H. 2016, ApJ, 832, 75
Zoghbi, A., Reynolds, C., & Cackett, E. 2013, ApJ, 777, 24