LATENT STOCHASTIC DIFFERENTIAL EQUATIONS FOR MODELING QUASAR VARIABILITY AND INFERRING BLACK HOLE PROPERTIES

Joshua Fagin[1,2,3,*], **Ji Won Park**[4,*], **Henry Best**[1,2,3], **K.E. Saavik Ford**[1,3,5,6], **Matthew J. Graham**[6,7], **V. Ashley Villar**[8,9,10], **Shirley Ho**[6,11], **James Hung-Hsu Chan**[2,3], and **Matthew O'Dowd**[1,2,3]

[1]The Graduate Center of the City University of New York
[2]Department of Physics and Astronomy, Lehman College of the CUNY
[3]Department of Astrophysics, American Museum of Natural History
[4]SLAC National Accelerator Laboratory
[5]Department of Science, CUNY Borough of Manhattan Community College
[6]Center for Computational Astrophysics, Flatiron Institute
[7]California Institute of Technology
[8]Department of Astronomy and Astrophysics, Pennsylvania State University
[9]Institute for Computational & Data Sciences, The Pennsylvania State University
[10]Institute for Gravitation and the Cosmos, The Pennsylvania State University
[11]Department of Astrophysical Sciences, Princeton University
[*]Corresponding authors. These authors contributed equally to this work.
jfagin@gradcenter.cuny.edu
jp281@slac.stanford.edu

ABSTRACT

Active galactic nuclei (AGN) are believed to be powered by the accretion of matter around supermassive black holes at the centers of galaxies. The variability of an AGN's brightness over time can reveal important information about the physical properties of the underlying black hole. The temporal variability is believed to follow a stochastic process, often represented as a damped random walk described by a stochastic differential equation (SDE). With upcoming wide-field surveys set to observe 100 million AGN in multiple bandpass filters, there is a need for efficient and automated modeling techniques that can handle the large volume of data. Latent SDEs are well-suited for modeling AGN time series data, as they can explicitly capture the underlying stochastic dynamics. In this work, we modify latent SDEs to jointly reconstruct the unobserved portions of multivariate AGN light curves and infer their physical properties such as the black hole mass. Our model is trained on a realistic physics-based simulation of ten-year AGN light curves, and we demonstrate its ability to fit AGN light curves even in the presence of long seasonal gaps and irregular sampling across different bands, outperforming a multi-output Gaussian process regression baseline.

1 INTRODUCTION

Active galactic nuclei (AGN) are among the brightest objects in the universe and play a crucial role in galaxy evolution. They are thought to be powered by the conversion of gravitational potential energy into thermal radiation by feeding on matter in the hot accretion disks of super massive black holes (SMBH) at the center of galaxies (Salpeter, 1964; Zel'dovich, 1964). Luminous AGN with unobscured accretion disks are known as quasars. These objects are so luminous they remain observable at extreme cosmological distances (Mortlock et al., 2011; Bañados et al., 2018) making them exceptional probes of the early universe.

The stochastic variability of quasar brightness has been studied extensively since their discovery (Greenstein, 1963; Hazard et al., 1963; Matthews & Sandage, 1963; Oke, 1963; Schmidt, 1963).

While the accretion disk of an AGN is too small to resolve at extragalactic distances, its variability can provide insight into the intrinsic physical properties of the AGN. For instance, the amplitude increases with decreasing luminosity, rest-frame wavelength, and Eddington ratio (Wills et al., 1993; Giveon et al., 1999; Berk et al., 2004). Efforts have been made to constrain the correlation of variability parameters with black hole mass, but their robustness has been inconclusive, with studies claiming positive or negative relations depending on the degrees of observational bias present in the fit data (Wold et al., 2008; MacLeod et al., 2010; Simm et al., 2015). Inferring the physics of black holes from AGN light curves, the brightness of quasars over time, can offer valuable insights into the evolution of the universe and the nature of dark matter and dark energy (Khadka & Ratra, 2020).

The Legacy Survey of Space and Time (LSST) at the Vera Rubin Observatory marks an unprecedented improvement in data quality and volume. The LSST main survey alone (10,000 deg^2) is projected to yield 100 million AGN light curves with six optical bandpass filters (*ugrizy*) at $55-185$ samplings per band over a ten year period (Abell et al., 2009). Machine learning algorithms are well suited to analyze the large amount of data expected from LSST; however, these light curves pose challenges for existing techniques such as multiple bands, long gaps of missing data, non-uniform sampling, and photometric and systematic noise.

The temporal variability of AGN is thought to approximately follow a damped random walk (DRW), a type of Gaussian process also known as the Ornstein-Uhlenbeck process (Rasmussen & Williams, 2006; Zu et al., 2013). A DRW $X(t)$ is governed by the stochastic differential equation (SDE):

$$\mathrm{d}X(t) = -\frac{1}{\tau}X(t)\,\mathrm{d}t + \sigma\sqrt{\mathrm{d}t}\,\epsilon(t) + b\,\mathrm{d}t \tag{1}$$

where $\epsilon(t)$ is a white noise process with mean zero and variance one, τ is the characteristic timescale, b is related to the mean of the process $\overline{X} = b\tau$, and σ is related to the standard deviation of $X(t)$, defined by the asymptotic structure function $\mathrm{SF}_\infty = \sigma\sqrt{\tau/2}$ (Kelly et al., 2009). Gaussian process regression (GPR) can be used to measure the variability parameters SF_∞ and τ, which may be correlated to the black hole mass (MacLeod et al., 2010; Suberlak et al., 2021). A smaller τ yields more high frequency variations since the variability happens on a shorter timescale. Fig. 3 shows example DRWs for two values of τ.

Here, we model quasar variability using the latent SDE (Li et al., 2020), a type of generative model that can model continuous-time stochastic dynamics. Latent SDEs can be viewed as infinite-dimensional variational auto-encoders (VAEs; Kingma & Welling, 2013; Rezende et al., 2014) with an SDE-induced process as the latent. We train the latent SDE to model the light curve dynamics and simultaneously predict the black hole masses and variability parameters (SF_∞ and τ). Unlike GPR on individual light curves, latent SDEs can model the variability across an entire training set, potentially generalizing from a well-sampled light curve to a poorly-sampled one, and are not restricted to a specific choice of kernel. Previous works have modelled quasar variability using recurrent neural networks (RNNs) (Tachibana et al., 2020; Sá nchez-Sáez et al., 2021), stochastic RNNs (Sheng et al., 2022), and attentive neural processes (Park et al., 2021).

2 METHOD

2.1 TRAINING SET

Our training set is comprised of a realistic, physics-motivated simulation of LSST 10-year light curves. We also include an additional year to assess our ability to extrapolate forward or backward in time. We first realize the DRW for a given SF_∞, τ, and mean magnitude. To simulate the reaction of the accretion disk to a driving DRW flux, a transfer function is calculated for each bandpass filter (Blandford & McKee, 1982). All of the physical properties of the black hole and quasar geometry are included in the transfer functions derived from a general relativistic ray-traced accretion disk simulation (Best et al., 2022; Bursa, 2017; 2018). In particular, the black hole mass is related to a time delay between bands of order hours to weeks. These transfer functions are calculated from a vertically thin, optically thick accretion disk (Shakura & Sunyaev, 1973), irradiating the photons from an X-ray source (the irradiated disk model; Sergeev et al., 2005; Cackett et al., 2007). The DRW is converted from magnitude to flux and then convolved with the transfer function kernel to

obtain simulated light curves for a set of black hole parameters. We then converted back from flux to magnitude [1]. The time series is then degraded to LSST-like observations with seasonal gaps and photometric noise (see A.2) and discretized to three day intervals. We train on 100,000 mock light curves per epoch and test on 10,000. The training set is randomly regenerated on the fly each epoch during training to improve generalization.

2.2 Model architecture

The input into the network is the apparent magnitude (which can be interpreted as the brightness) and uncertainty for a total of 12 features at each time step (6 bands and 6 uncertainties) with unobserved time steps masked out. The network architecture consists of three components: the RNN encoder, the latent SDE solver and decoder, and the multi-layer perceptron (MLP) parameter estimator. More details can be found in A.4.

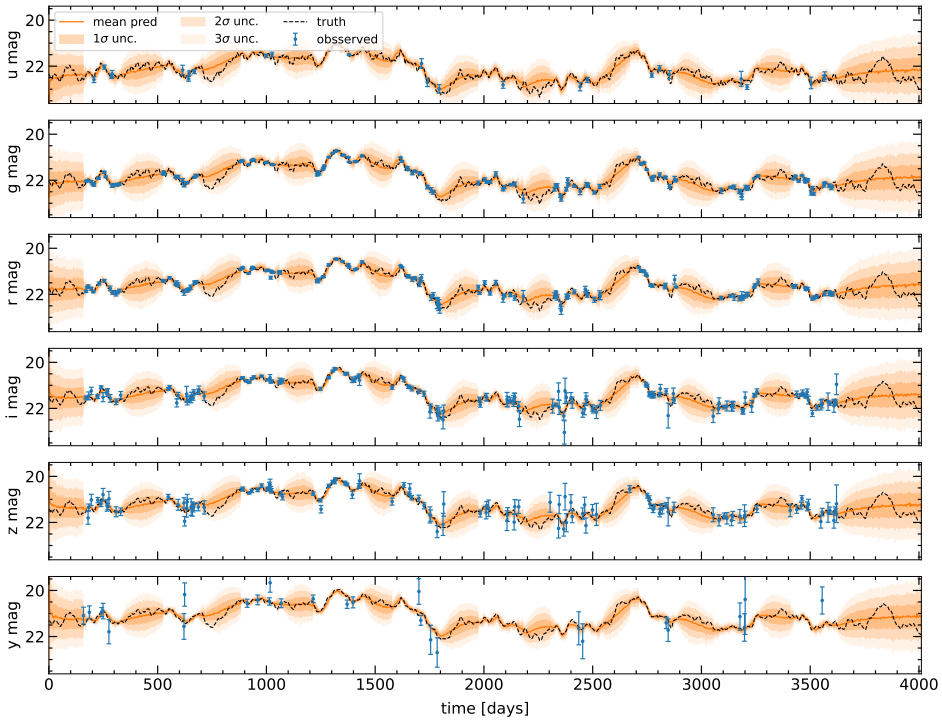

Figure 1: The latent SDE reconstructs the light curve at unobserved time points (orange) given partial observations, or context points (blue). The σ levels are calculated from 250 posterior samples.

2.3 Uncertainty quantification and loss

We parameterize the posterior on our parameters as a multivariate Gaussian distribution and minimize its log-likelihood. For light curve reconstruction, we minimize the evidence lower bound (ELBO) described in Li et al. (2020) with a Gaussian likelihood averaged across the time steps. To ensure that our mean prediction closely matches the context points, we additionally evaluate the negative Gaussian log-likelihood (NGLL) parameterized with the input uncertainty at the mean prediction, average over the observed times, and append this to the loss. Details about the parameter labels can be found in A.3 and about training the model in A.5.

3 Results and discussion

Fig. 1 shows the light curve reconstruction for a test example. The model is more uncertain at time points farther away from the context points, as expected, and the network is able to infer across

[1] The light curves are converted from flux to AB magnitude: $m_{\rm AB} = -2.5 \log_{10}(f_\nu) - 48.60$.

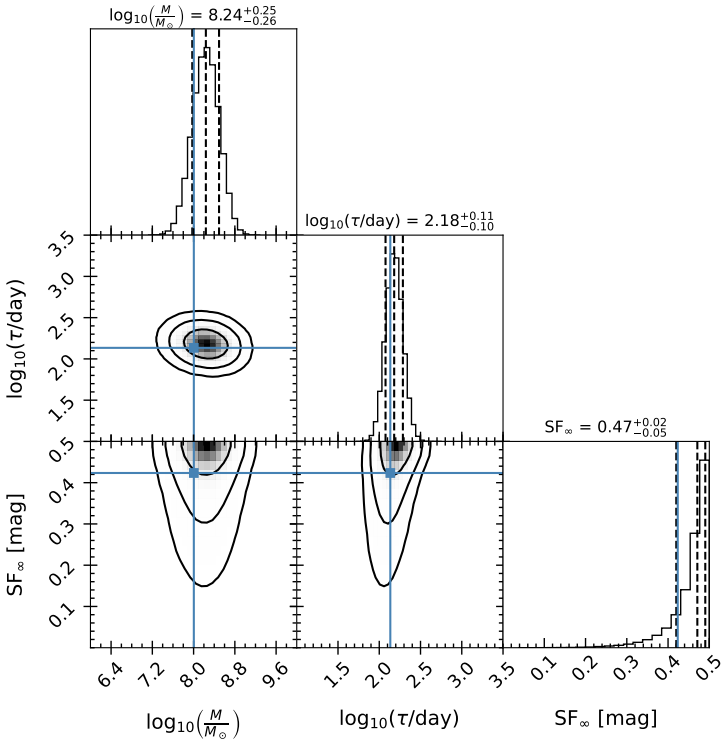

Figure 2: Example posterior on the parameters for the test light curve shown in Fig. 1. The diagonal plots display marginal distributions with the median and 1σ level indicated by dotted lines. The central plots depict 1, 2, and 3σ levels. The true parameter values are overlaid in blue.

bands. Fig 2 shows a predicted parameter posterior for the same test example. The parameter recovery performance across the test set is summarized in the confusion matrices of Fig. 4 and the uncertainty calibration in Fig. 5 (see A.3 for details). Table 1 compares latent SDEs light curve reconstruction performance with a fixed-noise, multi-task GPR baseline (see A.6). The latent SDE performs better than the GPR baseline on light curve reconstruction in terms of root mean square error (RMSE), mean absolute error (MAE), and NGLL.

Table 1: Light curve reconstruction performance of latent SDE and a multi-output Gaussian process regression baseline. Values reported are the median $\pm$ median absolute deviation for each metric across our test set of 10,000 light curves. Lower is better for all the metrics.

Model	RMSE [mag]	MAE [mag]	NGLL
latent SDE	$\mathbf{0.090 \pm 0.052}$	$\mathbf{0.061 \pm 0.034}$	$\mathbf{-1.35 \pm 0.57}$
GPR baseline	0.096 ± 0.054	0.066 ± 0.036	-1.16 ± 0.56

Latent SDEs offer a promising framework for joint light curve reconstruction and black hole parameter inference. Running inference on a trained latent SDE yields better performance and is faster than fitting a GP on each light curve individually—an important scaling advantage as we look forward to the 100 million light curves expected from LSST. In addition, standard GPR does not have the capacity for joint parameter inference. It only has access to the light curve after convolution with the transfer function, which loses information, whereas our model can be trained on the variability parameters of the quasar source (pre-convolution) magnitude as the labels. In future work, we hope to extend our parameter predictions to more black hole properties, such as the spin and inclination angle. Latent SDEs can also serve as an engine for anomaly detection to identify physically interesting quasars with variability that differs from the training set. The method presented

here is completely general and can be applied to any multivariate time series with missing data and irregular sampling.

ACKNOWLEDGMENTS

This research was made possible by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program. Matthew Graham acknowledges support from National Science Foundation (NSF) AST-2108402. V. Ashley Villar acknowledges support from NSF under grant AST-2108676. The data used in this publication were collected through the MENDEL high performance computing (HPC) cluster at the American Museum of Natural History. This HPC cluster was developed with NSF Campus Cyberinfrastructure support through Award #1925590. We thank Xuechen Li and David Duvenaud for their thoughtful comments on our work.

REFERENCES

Paul A Abell, David L Burke, Mario Hamuy, Martin Nordby, Tim S Axelrod, David Monet, Bojan Vrsnak, Paul Thorman, DR Ballantyne, Joshua D Simon, et al. Lsst science book, version 2.0. Technical report, LSST, 2009.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization, 2016. URL https://arxiv.org/abs/1607.06450.

Eduardo Bañados, Bram P. Venemans, Chiara Mazzucchelli, Emanuele P. Farina, Fabian Walter, Feige Wang, Roberto Decarli, Daniel Stern, Xiaohui Fan, Frederick B. Davies, Joseph F. Hennawi, Robert A. Simcoe, Monica L. Turner, Hans-Walter Rix, Jinyi Yang, Daniel D. Kelson, Gwen C. Rudie, and Jan Martin Winters. An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5. , 553(7689):473–476, January 2018. doi: 10.1038/nature25180.

Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. Botorch: A framework for efficient monte-carlo bayesian optimization. 2019. doi: 10.48550/ARXIV.1910.06403. URL https://arxiv.org/abs/1910.06403.

Daniel E Vanden Berk, Brian C Wilhite, Richard G Kron, Scott F Anderson, Robert J Brunner, Patrick B Hall, Željko Ivezić, Gordon T Richards, Donald P Schneider, Donald G York, et al. The ensemble photometric variability of˜ 25,000 quasars in the sloan digital sky survey. *The Astrophysical Journal*, 601(2):692, 2004.

Henry Best, Joshua Fagin, Georgios Vernardos, and Matthew O'Dowd. Resolving the vicinity of supermassive black holes with gravitational microlensing, 2022. URL https://arxiv.org/abs/2210.10500.

R. D. Blandford and C. F. McKee. Reverberation mapping of the emission line regions of Seyfert galaxies and quasars. , 255:419–439, April 1982. doi: 10.1086/159843.

Michal Bursa. Numerical implementation of equations for photon motion in Kerr spacetime. In *RAGtime 17-19: Workshops on Black Holes and Neutron Stars*, pp. 7–21, December 2017.

Michal Bursa. SIM5: Library for ray-tracing and radiation transport in general relativity. Astrophysics Source Code Library, record ascl:1811.011, November 2018.

Edward M. Cackett, Keith Horne, and Hartmut Winkler. Testing thermal reprocessing in active galactic nuclei accretion discs. , 380(2):669–682, September 2007. doi: 10.1111/j.13652966.2007.12098.x.

Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values, 2016. URL https://arxiv.org/abs/1606.01865.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014. URL https://arxiv.org/abs/1412.3555.

Hao Fu, Chunyuan Li, Xiaodong Liu, Jianfeng Gao, Asli Celikyilmaz, and Lawrence Carin. Cyclical annealing schedule: A simple approach to mitigating kl vanishing, 2019.

Jacob R Gardner, Geoff Pleiss, David Bindel, Kilian Q Weinberger, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Advances in Neural Information Processing Systems*, 2018.

Uriel Giveon, Dan Maoz, Shai Kaspi, Hagai Netzer, and Paul S Smith. Long-term optical variability properties of the palomar—green quasars. *Monthly Notices of the Royal Astronomical Society*, 306(3):637–654, 1999.

Jesse L. Greenstein. Red-Shift of the Unusual Radio Source: 3C 48. , 197(4872):1041–1042, March 1963. doi: 10.1038/1971041a0.

Ryan-Rhys Griffiths, Jiachen Jiang, Douglas J. K. Buisson, Dan Wilkins, Luigi C. Gallo, Adam Ingram, Alpha A. Lee, Dirk Grupe, Erin Kara, Michael L. Parker, William Alston, Anthony Bourached, George Cann, Andrew Young, and S. Komossa. Modeling the multiwavelength variability of mrk 335 using gaussian processes. *The Astrophysical Journal*, 914(2):144, jun 2021. doi: 10.3847/1538-4357/abfa9f. URL https://doi.org/10.3847%2F1538-4357%2Fabfa9f.

C. Hazard, M. B. Mackey, and A. J. Shimmins. Investigation of the Radio Source 3C 273 By The Method of Lunar Occultations. , 197(4872):1037–1039, March 1963. doi: 10.1038/1971037a0.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. URL https://arxiv.org/abs/1512.03385.

Željko Ivezić and Kahn et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. , 873(2):111, March 2019. doi: 10.3847/1538-4357/ab042c.

Brandon C. Kelly, J. Bechtold, and A. Siemiginowska. Are the Variations in Quasar Optical Flux Driven by Thermal Fluctuations? In *American Astronomical Society Meeting Abstracts #213*, volume 213 of *American Astronomical Society Meeting Abstracts*, pp. 422.06, January 2009.

Narayan Khadka and Bharat Ratra. Using quasar X-ray and UV flux measurements to constrain cosmological model parameters. *Monthly Notices of the Royal Astronomical Society*, 497(1): 263–278, 06 2020. ISSN 0035-8711. doi: 10.1093/mnras/staa1855. URL https://doi.org/10.1093/mnras/staa1855.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013. URL https://arxiv.org/abs/1312.6114.

Xuechen Li, Ting-Kam Leonard Wong, Ricky T. Q. Chen, and David Duvenaud. Scalable gradients for stochastic differential equations. *International Conference on Artificial Intelligence and Statistics*, 2020.

Ch L MacLeod, Ž Ivezić, CS Kochanek, S Kozłowski, B Kelly, E Bullock, A Kimball, B Sesar, D Westman, K Brooks, et al. Modeling the time variability of sdss stripe 82 quasars as a damped random walk. *The Astrophysical Journal*, 721(2):1014, 2010.

Wesley J. Maddox, Maximilian Balandat, Andrew Gordon Wilson, and Eytan Bakshy. Bayesian optimization with high-dimensional outputs, 2021. URL https://arxiv.org/abs/2106.12997.

Thomas A. Matthews and Allan R. Sandage. Optical Identification of 3C 48, 3C 196, and 3C 286 with Stellar Objects. , 138:30, July 1963. doi: 10.1086/147615.

Daniel J. Mortlock, Stephen J. Warren, Bram P. Venemans, Mitesh Patel, Paul C. Hewett, Richard G. McMahon, Chris Simpson, Tom Theuns, Eduardo A. Gonzáles-Solares, Andy Adamson, Simon Dye, Nigel C. Hambly, Paul Hirst, Mike J. Irwin, Ernst Kuiper, Andy Lawrence, and Huub J. A. Röttgering. A luminous quasar at a redshift of z = 7.085. , 474(7353):616–619, June 2011. doi: 10.1038/nature10159.

J. B. Oke. Absolute Energy Distribution in the Optical Spectrum of 3C 273. , 197(4872):1040–1041, March 1963. doi: 10.1038/1971040b0.

Ji Won Park, Ashley Villar, Yin Li, Yan-Fei Jiang, Shirley Ho, Joshua Yao-Yu Lin, Philip J. Marshall, and Aaron Roodman. Inferring black hole properties from astronomical multivariate time series with bayesian attentive neural processes, 2021. URL https://arxiv.org/abs/2106.01450.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006. ISBN 026218253X.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pp. 1278–1286. PMLR, 2014.

E. E. Salpeter. Accretion of Interstellar Matter by Massive Objects. , 140:796–800, August 1964. doi: 10.1086/147973.

P. Sá nchez-Sáez, H. Lira, L. Martí, N. Sánchez-Pi, J. Arredondo, F. E. Bauer, A. Bayo, G. Cabrera-Vives, C. Donoso-Oliva, P. A. Estévez, S. Eyheramendy, F. Förster, L. Hernández-García, A. M. Muñoz Arancibia, M. Pérez-Carrasco, M. Sepúlveda, and J. R. Vergara. Searching for changing-state AGNs in massive data sets. i. applying deep learning and anomaly-detection techniques to find AGNs with anomalous variability behaviors. *The Astronomical Journal*, 162 (5):206, oct 2021. doi: 10.3847/1538-3881/ac1426. URL https://doi.org/10.3847%2F1538-3881%2Fac1426.

M. Schmidt. 3C 273 : A Star-Like Object with Large Red-Shift. , 197(4872):1040, March 1963. doi: 10.1038/1971040a0.

S. G. Sergeev, V. T. Doroshenko, Yu. V. Golubinskiy, N. I. Merkulova, and E. A. Sergeeva. Lag-Luminosity Relationship for Interband Lags between Variations in B, V, R, and I Bands in Active Galactic Nuclei. , 622(1):129–135, March 2005. doi: 10.1086/427820.

N. I. Shakura and R. A. Sunyaev. Reprint of 1973A&A....24..337S. Black holes in binary systems. Observational appearance. *Astronomy and Astrophysics*, 500:33–51, June 1973.

Xinyue Sheng, Nicholas Ross, and Matt Nicholl. Legacy Survey of Space and Time cadence strategy evaluations for active galactic nucleus time-series data in Wide-Fast-Deep field. *Monthly Notices of the Royal Astronomical Society*, 512(4):5580–5600, 03 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac803. URL https://doi.org/10.1093/mnras/stac803.

T Simm, R Saglia, M Salvato, R Bender, WS Burgett, KC Chambers, PW Draper, H Flewelling, N Kaiser, R-P Kudritzki, et al. Pan-starrs1 variability of xmm-cosmos agn-i. impact on photometric redshifts. *Astronomy & Astrophysics*, 584:A106, 2015.

Krzysztof L. Suberlak, Ž eljko Ivezić, and Chelsea MacLeod. Improving damped random walk parameters for SDSS stripe 82 quasars with pan-STARRS1. *The Astrophysical Journal*, 907 (2):96, feb 2021. doi: 10.3847/1538-4357/abc698. URL https://doi.org/10.3847%2F1538-4357%2Fabc698.

Yutaro Tachibana, Matthew J. Graham, Nobuyuki Kawai, S. G. Djorgovski, Andrew J. Drake, Ashish A. Mahabal, and Daniel Stern. Deep modeling of quasar variability. *The Astrophysical Journal*, 903(1):54, nov 2020. doi: 10.3847/1538-4357/abb9a9. URL https://doi.org/10.3847%2F1538-4357%2Fabb9a9.

Beverley J Wills, MS Brotherton, D Fang, Charles C Steidel, and Wallace LW Sargent. Statistics of qso broad emission-line profiles. i. the c iv lambda 1549 line and the lambda 1400 feature. *The Astrophysical Journal*, 415:563, 1993.

M Wold, MS Brotherton, and Z Shang. Black hole mass and variability in quasars. In *AIP Conference Proceedings*, volume 1053, pp. 55–58. American Institute of Physics, 2008.

Ya. B. Zel'dovich. The Fate of a Star and the Evolution of Gravitational Energy Upon Accretion. *Soviet Physics Doklady*, 9:195, September 1964.

Ying Zu, C. S. Kochanek, Szymon Kozłowski, and Andrzej Udalski. IS QUASAR OPTICAL VARIABILITY a DAMPED RANDOM WALK? *The Astrophysical Journal*, 765(2):106, feb 2013. doi: 10.1088/0004-637x/765/2/106. URL https://doi.org/10.1088%2F0004-637x%2F765%2F2%2F106.

A Appendix

A.1 Example damped random walk

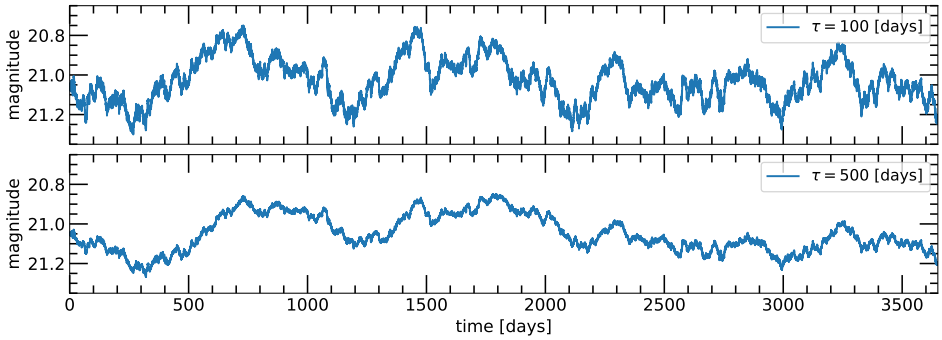

Figure 3: Damped random walk for two different values of τ with $\overline{X} = 21$ mag and $\mathrm{SF}_\infty = 0.1$ mag. Both DRWs are generated using the same random state so that their stochastic dynamics are identical.

A.2 LSST cadence and input uncertainties

The simulated time series is degraded to LSST-like observations. Observation times and photometric noise are taken from `rubin_sim`[2] by sampling 10,000 mock observation schemes from across the sky with between 750 and 1000 total observations across the ten years. The systematic uncertainty is set to 0.005 mag, the maximum value expected for LSST (Ivezić & et al., 2019) and added in quadrature with the photometric noise.

A.3 Parameter prediction labels and performance

When simulating the light curves, each parameter is drawn uniformly from a given range. We scale the labels from their physical values to between zero and one and then take the logit (scaling the labels from $-\infty$ to ∞) before taking the log-likelihood of our posterior. Once samples are drawn from the posterior we take the sigmoid (forcing the predictions to be between zero and one) and then scale back to the original physical range. This prevents posterior probability from being wasted in physically impossible parameter space (such as negative SF_∞) or outside the range of the training set. The scaling process is why the posterior in Fig. 2 is skewed away from the edges of the parameter space instead of being Gaussian.

Fig. 4 shows confusion matrices for our parameter predictions. The network was able to predict the variability parameters $\log_{10}(\tau)$ and SF_∞ well across a broad parameter range. The black hole

[2]https://github.com/lsst/rubin_sim

mass is harder to predict, especially at low mass, since it only affects the transfer function kernel that is convolved with the DRW and must be extrapolated from time delays between bandpasses. These time delays can be as little as a few hours in the low mass and redshift case but up to several weeks. Fig. 5 shows that our reported uncertainty of the parameter posteriors is, on average, correctly calibrated across the test set.

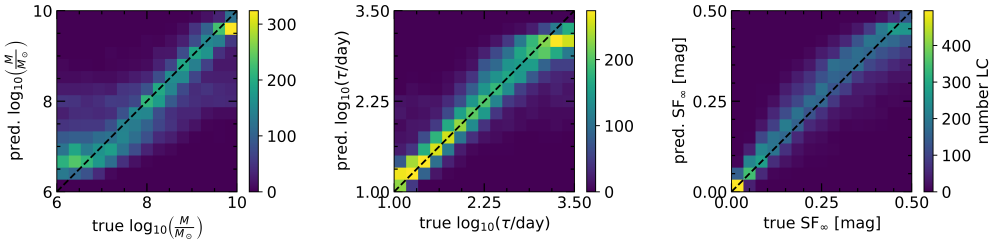

Figure 4: Confusion matrices for our mean parameter predictions vs. the true parameters. The black dashed lines across the diagonal represent perfect predictions.

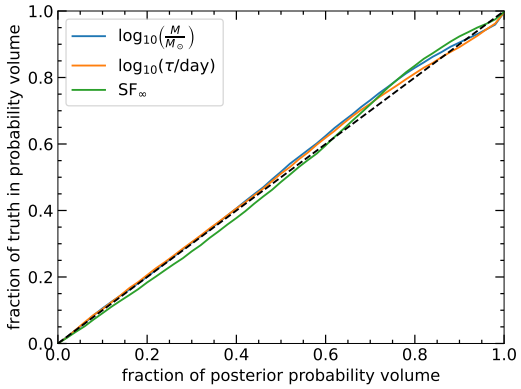

Figure 5: The fraction of the truth encompassed by some probability volume (credible interval) across our test set. Perfect calibration is shown by the black dashed line across the diagonal.

A.4 Model architecture

The neural network architecture uses a hidden size of 128 and a context size of 64. A latent size of 8 is used for the SDE with an additional 8 given to the RNN projector to aid in uncertainty estimation. Each hidden layer is followed by a LeakyReLU activation and layer normalization (Ba et al., 2016). The network is built using `PyTorch` (Paszke et al., 2019) and the SDE solver uses the `TorchSDE` [3] (Li et al., 2020) implementation. It has a total of 903,597 trainable parameters. The light curves are normalized to have mean zero and variance one during training (only using observed points to calculate the mean and variance).

The encoder starts with a GRU-D (Che et al., 2016) layer, a modified version of the Gated Recurrent Unit (GRU) (Chung et al., 2014) that includes both masking and irregular time intervals. This is followed by two GRU layer and two dense layers that output the context vector. A residual skip connection (He et al., 2015) is included between the GRU-D layer and the output of the second GRU layer. Two more dense layers are used to produce the mean μ_{q_z} and variance $\sigma^2_{q_z}$ of the variational distribution q_z from the context vector. We then draw the latent vector $z_0 \sim \mathcal{N}(\mu_{q_z}, \sigma^2_{q_z})$, the initial conditions of our latent SDE.

The decoder includes an Itô SDE with diagonal noise containing networks governing the drift, diffusion, and prior drift (Li et al., 2020). We use the Euler-Maruyama numerical approximation scheme.

[3] https://github.com/google-research/torchsde

The drift and prior drift networks are both MLPs with a residual skip connection. There are three dense layers pre-skip and two post-skip. The diffusion network consists of three dense layers and is applied element wise to satisfy the diagonal noise. In order to model both the dynamics of the time series as well as produce uncertainties on predictions, we use an RNN to project the output of the SDE, the latent vector, and the input uncertainties of the light-curve onto the mean and log-variance of the observation space. The projector consists of a GRU-D layer followed by two GRU layers. There is also a skip-connection of two dense layers between the output of the SDE and output of the RNN.

The parameter estimation is done using the same MLP architecture as the drift networks and outputting the mean and Cholesky factor of the covariance matrix for our multivariate Gaussian posterior. The input into the parameter estimator is the final time step of the context vector, μ_{q_z}, $\sigma^2_{q_z}$, the output of the drift, diffusion, and prior drift networks at μ_{q_z}, and the mean and standard deviation of the unnormalized light curve.

A.5 Training

Our network is trained for 50 epochs using the Adam optimizer (Kingma & Ba, 2017) with an initial learning rate of 0.0025, an exponential decay of 0.97, and a batch size of 125. Linear KL-annealing (Fu et al., 2019) is used for the first 25 epochs as well as annealing for the parameter predictions. Training took around seven days with an NVIDIA V100 GPU.

A.6 Gaussian process regression baseline

We use an exact, multi-task Gaussian process regression (GPR) as the baseline for our light curve reconstruction (Maddox et al., 2021). The multi-task GPR can infer correlation between output variables, in our case the different LSST bandpass filters. The bands must be modeled jointly, as they are related to the same X-ray driving DRW process and LSST observations are sparse and distributed asynchronously across the different bands. Since the observation noise is known at each point and is set from the LSST observation strategy, we use the fixed-noise multi-task GPR implementation in `BoTorch`[4] (Balandat et al., 2019), based on the GP library `GPyTorch` (Gardner et al., 2018). We use the absolute exponential kernel (equivalent to the Matérn-1/2 kernel) which corresponds to the DRW process. This kernel has also been empirically shown to best fit optical and UV AGN variability compared to the Matérn-3/2, Matérn-5/2, rational quadratic, and squared exponential kernels (Griffiths et al., 2021). Each band of the light curve is independently normalized to have zero mean and unit variance, using only observed points, when fitting the GP. We found this to work better than normalizing all bands together.

[4]https://botorch.org/