
Joint Modeling of Quasar Variability and Accretion Disk Reprocessing using Latent Stochastic Differential Equations

Joshua Fagin¹²³ James Hung-Hsu Chan²³ Henry Best¹²³⁴ Matthew O’Dowd¹²³ K. E. Saavik Ford¹²⁵⁶
Matthew J. Graham⁷ Ji Won Park⁸⁹ V. Ashley Villar¹⁰

Abstract

Quasars are bright active galactic nuclei powered by the accretion of matter around supermassive black holes at the center of galaxies. With LSST set to monitor tens of millions of quasars over the next ten years, efficient techniques like machine learning must be developed. Quasar variability is believed to be driven by an X-ray corona reprocessed by the accretion disk and emitted as UV/optical variability. We are the first to introduce an auto-differentiable simulation of the accretion disk and reprocessing that we use as a direct component of our neural network to jointly model the driving variability and reprocessing in simulated LSST light curves. We encode the light curves using a Transformer, and the driving variability is reconstructed using latent stochastic differential equations, a physically motivated generative deep learning method to model continuous-time stochastic dynamics. By embedding the physical processes into our network, we achieve a model that is more robust and interpretable.

¹The Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016, USA ²Department of Astrophysics, American Museum of Natural History, Central Park West and 79th Street, NY 10024-5192, USA ³Department of Physics and Astronomy, Lehman College of the CUNY, Bronx, NY 10468, USA ⁴Department of Theoretical Physics and Astrophysics, Faculty of Science, Masaryk University, Kotlářská 2, CZ-602 00 Brno, Czech Republic ⁵Department of Science, CUNY Borough of Manhattan Community College, 199 Chambers St, New York, NY 10007, USA ⁶Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA ⁷Department of Physics, Maths and Astronomy, California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA ⁸SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA ⁹Prescient Design, Genentech, South San Francisco, CA 94080, USA ¹⁰Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138-1516, USA. Correspondence to: Joshua Fagin <jfagin@gradcenter.cuny.edu>.

Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

1. Introduction

Quasars are bright and unobscured active galactic nuclei (AGN) thought to be powered by the accretion of matter around supermassive black holes, ranging from millions to billions of solar masses, and found at the center of galaxies (Salpeter, 1964). They are some of the brightest known objects, making them powerful probes of the early Universe (Bañados et al., 2018) and are thought to play an important role in galaxy evolution (Hoshi et al., 2024). Their stochastic brightness variability has been studied since their discovery (Greenstein, 1963; Hazard et al., 1963) and reflects the physical properties of the black hole and accretion disk that powers them. Studies of quasar variability have found accretion disk sizes to be larger than the standard thin-disk model by a factor of $\sim 2-4$ (Mudd et al., 2018; Guo et al., 2022; Jha et al., 2022). New measurements and methods are needed to test accretion disk models and enhance our understanding of quasar emissions.

The UV/optical variability of quasars is typically modeled as an X-ray driving variability source corona above the black hole that illuminates the accretion disk (Cackett et al., 2007). The reprocessing of the driving variability on the UV/optical emitting regions of the accretion disk is represented by the *transfer function* and introduces wavelength-dependent time lags ranging from less than a day for lower mass black holes to several tens of days (Blandford & McKee, 1982). This reprocessing is modeled through the convolution of the driving signal with the transfer functions:

$$F_{\lambda}(t, \lambda) = \bar{F}_{\lambda}(\lambda) + \Delta F_{\lambda}(\lambda) \int_0^{\infty} X(t-\tau)\psi(\tau|\lambda)d\tau, \quad (1)$$

where λ is the wavelength, $F_{\lambda}(t, \lambda)$ is the flux, $\bar{F}_{\lambda}(\lambda)$ is the mean flux, $\Delta F_{\lambda}(\lambda)$ is the amplitude of the variable flux, $X(t)$ is the normalized driving variability (mean zero and variance one), and $\psi(\tau|\lambda)$ is the transfer function kernels (Cackett et al., 2007; Starkey et al., 2015). These time lags can be measured through continuum reverberation mapping of UV/optical light curves to probe the size of the emitted regions and are interconnected to properties of the accretion disk (Cackett et al., 2021).

Upcoming wide-field surveys such as the Rubin Observa-

tory Legacy Survey of Space and Time (LSST) will observe an unprecedented quantity of data. The LSST main survey is projected to monitor tens of millions of quasars over a ten-year period with six UV/optical bandpass filters (*ugrizy*) at 55 – 185 samplings per band or around 800 total visits across the ten years (Collaboration, 2009). Machine learning (ML) algorithms are well suited to analyze the vast amounts of data. Quasar light curves from LSST pose challenges for traditional ML techniques, however, due to being multivariate, stochastic, irregularly sampled, and noisy.

UV/optical light curve data is most commonly modeled using Gaussian process regression (GPR), where they are fit with a specific kernel of the Gaussian process. For example, using the kernel associated with the damped random walk (DRW; Zu et al., 2013). The codebase JAVELIN (Zu et al., 2011; 2016) uses a DRW kernel with top hat transfer functions to simultaneously model the variability and time delays, that can in turn be used to measure the accretion disk size. CREAM (Starkey et al., 2015) is similar to JAVELIN but uses the thin-disk transfer functions directly. Instead of treating the bluest band as the effective driving variability, CREAM explicitly reconstructs the driving variability by modeling it as a Fourier series. These methods, however, require computationally expensive Markov chain Monte Carlo optimization and would be infeasible to apply to the entire LSST sample. They also cannot learn from features across an entire sample of light curves, unlike ML methods.

Fagin et al. (2024) introduced latent stochastic differential equations (SDEs) as a method to reconstruct quasar light curves and simultaneously predict accretion disk and variability parameters. Latent SDEs are a type of generative neural network (NN) that can model continuous-time stochastic dynamics (Li et al., 2020). They are physically motivated by the fact that UV/optical quasar variability is often modeled by SDEs such as the DRW or higher-order processes (Yu et al., 2022). In Fagin et al. (2024), simulated LSST observations are fitted, and parameter inference is performed based on the latent space of the SDE and context.

In this work and the full paper Fagin et al. (2025), we are the first ML method to combine the light curve reconstruction and parameter inference into a self-consistent, unified framework. This is achieved by developing an auto-differentiable simulation of the accretion disk and reprocessing, and including it in the architecture of our ML model. Within our NN, we use a latent SDE to generate the mean X-ray driving variability. We then predict the accretion disk parameters, which are converted to the corresponding transfer functions using our auto-differentiable simulation and convolved with the latent SDE generated driving signal to produce the mean best fit reconstruction for each observed band. The mean time of the predicted transfer functions gives a measurement of the continuum time lags. We also quantify the uncertainty

in the driving signal, UV/optical, time delays, and parameter reconstructions. We improve the driving variability parameter predictions through analyzing the power spectrum of the reconstructed driving signal by leveraging the slope and intercept of linear fits applied to the power spectrum broken up into five sections. In addition, we test using recurrent inference machines (RIM; Putzky & Welling, 2017) to iteratively improve our reconstructions and parameter inference.

2. Method

2.1. Training set and auto-differentiable simulation

Our training set is comprised of a realistic simulation of LSST 10-year light curves. The stochastic driving variability is generated using the method of Timmer & Koenig (1995) using a bending broken power-law (BPL) power spectral density (Czerny et al., 2023). The disk model is used for both the training set and the auto-differentiable simulation that is incorporated into the NN. To generate the transfer functions, we use the thin-disk plus lamppost model (Novikov & Thorne, 1973; Cackett et al., 2007) including a wind model accretion rate to have a power-law temperature profile (Sun et al., 2018; Fagin et al., 2024), along with gravitational redshifting and Doppler shifting caused by general relativity (Cunningham & Bardeen, 1973). We model the brightness of each band by simulating the spectrum with our continuum and including emission lines, host galaxy contamination, extinction, Lyman forest, and Lyman limit (Temple et al., 2021). The spectrum is integrated across the LSST bandpass response functions. The time series is generated in discrete daily intervals and degraded to LSST-like observations and photometric noise (see section 2.4 of Fagin et al., 2024). We have 12 total parameters: 7 for the accretion disk and 5 for the driving variability.

2.2. Neural network architecture

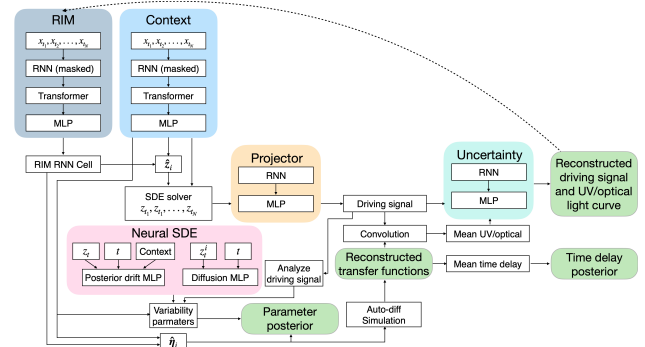


Figure 1. Diagram of our ML model. The data products are highlighted in green. The dashed line indicates the reconstructed UV/optical light curve that can optionally be used in the RIM procedure to update the predicted latent vector of the SDE \hat{z}_i and accretion disk parameters $\hat{\eta}_i$.

Our overall goal is to solve the inverse problem in Equa-

tion (1) related to reconstructing the driving variability and transfer function kernels (j), given only a set of noisy and sparsely sampled observations. Since the reprocessing is modeled as a convolution of the driving variability and transfer functions, we are essentially training a NN to solve a blind deconvolution. To make the problem tractable, we parameterize the transfer function via our differentiable simulation, ($j ; \hat{\alpha}$), where $\hat{\alpha}$ is the vector of accretion disk parameters. The driving signal is reconstructed using a latent SDE (Li et al., 2020; Fagin et al., 2024), parameterized by the context at each time and the latent vector \hat{z} . We then solve for the flux by embedding the physics of the reprocessing of the driving variability into our NN architecture by:

$$F_{\text{predicted}}(t; \{z, j, \hat{\alpha}\}) = \int_0^{Z_1} X_{\text{latent SDE}}(t; \{z, j, \hat{\alpha}\}) \cdot |_{\text{auto di transfer function}}(\{z, j, \hat{\alpha}\}) dt; \quad (2)$$

where this convolution is evaluated numerically within our ML model. Our ML model also quantifies the uncertainty in the mean reconstructions $\hat{F}(t; \hat{\alpha})$, $\hat{X}(t; \hat{\alpha})$, the variability parameters, and time delays coming from each reconstructed transfer function ($j ; \hat{\alpha}$). We use RIM to compare the reconstruction of our light curve with the observations, and iteratively adjust the accretion disk parameters by $\hat{\alpha}_{i+1} = \hat{\alpha}_i + \Delta \hat{\alpha}_i$ and the latent space of the latent SDE by $\hat{z}_{i+1} = \hat{z}_i + \Delta \hat{z}_i$ in Equation (2) with iteration i and initial values of zero.

A diagram of our ML model architecture is shown in Figure 1. The input into the network is the brightness at each observation with uncertainty. Unobserved time steps are set to a dummy value to be masked by the network. The network has two main encoder components: the context network of the latent SDE and a network used to predict the parameter posteriors and latent space. Both are bidirectional recurrent NNs (RNNs) and begin with a GRU-D layer (Che et al., 2016) that can handle the masked input, followed by two GRU layers (Chung et al., 2014). We also use a Transformer encoder after the RNN layers that is concatenated with the output (Vaswani et al., 2023) followed by two fully connected layers. There is an additional RNN without the Transformer to produce uncertainty in the reconstructed driving and UV/optical variability. Furthermore, there are multi-layer perceptrons (MLPs) to produce the mean and bias of the light curve, the variability parameters, and the uncertainty in the time delay estimates. There are also two MLPs in the latent SDE: the posterior drift function that decodes the context and latent vector, and the diffusion network that is applied element-wise to satisfy the diagonal noise (Li et al., 2020). Each MLP consists of four fully connected layers. The RNNs use tanh activation, Transformers use GELU (Hendrycks & Gimpel, 2023), and fully connected layers use LeakyReLU (Maas et al., 2013).

Whenever appropriate, we include residual skip connections (He et al., 2015) and layer norm (Ba et al., 2016). We use a hidden size of 256, Transformer encoder size of 512, context size of 128, and latent size of 16. The Transformer encoders have 5 layers, 8 heads, and use a sinusoidal time embedding. Our model is built in PyTorch (Paszke et al., 2019) and has 38,551,054 parameters.

2.3. Loss function and training

We parameterize the posterior of the parameters as a Gaussian mixture model with ν multivariate Gaussians and minimize its negative log-likelihood (NLL). We take the sigmoid of the posterior to restrict the parameter estimates into our uniform priors. We also quantify uncertainty in the mean time delays between bands, obtained from the reconstructed transfer functions, as a multivariate Gaussian. For the light curve reconstruction, we again minimize a Gaussian NLL averaged across time. We also evaluate the Gaussian NLL at each observation with respect to the photometric errors to force the mean reconstruction to be close to the context points. The loss function is the weighted sum of the four components. We train with 100,000 light curves per epoch for 30 epochs, randomly regenerated on the fly to avoid overfitting. We reserve 10,000 light curves per test set. Our ML model is trained for 30 epochs using the Adam optimizer (Kingma & Ba, 2017). We train with 4 A100 GPUs in multiple stages. We first train without RIM to train faster. We use an initial learning rate of 10^{-4} . The learning rate is exponentially decayed by 0.95 each epoch, and a batch size of 26 per GPU. In the final 4 epochs, we use 3 RIM iterations with a batch size of 14 to test the RIM technique. Training took around 3 weeks.

3. Results and conclusions

We show an example light curve reconstruction in Figure 2 from our nominal test set. The model is more certain near observations and infers across the bands and driving signal. We show in Appendix A for the nominal test set that our inferred posteriors are well calibrated and superior to the GPR baseline, and our ML model's ability to infer accretion disk and driving signal parameters. We find that the RIM procedure yields a minor improvement, with the average loss on the nominal test set decreasing from 0.619 to 0.580 and then to 0.579, respectively, with each RIM iteration.

We apply our trained ML model to several test sets to evaluate its performance on out-of-distribution variability, including using BPL, DRW, BPL+Sine, Sine, sawtooth, and square wave driving signals. The BPL is the same as the training set. We compare the NLL light curve reconstruction to an exact multitask GPR baseline with DRW kernel (see section 4 of Fagin et al., 2024). The latent SDE model outperforms the GPR baseline across all cases, summarized in

Figure 2. The left panels show the reconstruction driving variability and z bands (orange) given the set of LSST-like observations (blue). The UV/optical bands are reprocessed from the driving signal using the reconstructed transfer functions given in the right panel, with mean time delays given by the dashed lines.

Table 1. Light curve reconstruction performance of our latent SDE model compared to the GPR baseline in terms of the Gaussian NLL for six test sets with different types of driving signals. The values reported are the median absolute deviation on the median.

Driving Signal	Latent SDE		GPR	
BPL	1:528	0:007	1:409	0:007
DRW	1:214	0:007	1:129	0:007
BPL+Sine	1:276	0:007	1:111	0:007
Sine	1:847	0:009	1:765	0:012
Sawtooth	1:221	0:007	0:971	0:008
Square Wave	1:098	0:007	0:933	0:007

Figure 3. Median predicted black hole masses compared to the true values by applying our pretrained ML model to test sets with out-of-distribution driving signals.

Table 1. Furthermore, we demonstrate that we can robustly infer the black hole mass for each test set in Figure 3. This learning approach retains. Although our model is demanding to train, the average inference time using a batch size of 64 is about 27 minutes per million light curves using 3 iterations of the RIM, or 7 minutes using 1 iteration, six to seven orders of magnitude faster than AVELIN or CREAM (Li et al., 2024).

Unlike GPR, which is applied to light curves individually, our latent SDE model generalizes across the entire training set. For example, it learns from the brightness of each band as modeled in the training set. As a result, GPR and simple We aim to incorporate as much physics into our NN as

possible and test its robustness. The latent SDE captures the physics of the stochastic driving variability, while the auto-differentiable simulation models the reprocessing of the driving signal on the accretion disk. In previous work, there is nothing that forces the reconstruction and time delays to actually correspond to the inferred accretion disk parameters. Within our ML model, we fit the power spectrum of the reconstructed driving signal with several linear fits to better estimate its parameters. By embedding these physical processes into our network, we achieve a model that is more robust and interpretable compared to traditional black-box parameter estimators. For example, our predicted time delays can be compared to measurements obtained via curve-shifting techniques, providing a way to confirm the accuracy of our model on data. Our ML method is general and may be adapted to solve other blind deconvolution or inverse problems with irregularly sampled time series. See Fagin et al. (2025) for the full paper.

Software and Data

Our ML model and auto-differentiable and GPU-accelerated accretion disk simulation are open-sourced and available at: <https://github.com/JFagin/QuasML>.

This work relied on the following open source software: PyTorch (Paszke et al., 2019), torchsde (Li et al., 2020), BoTorch (Balandat et al., 2020), Matplotlib (Hunter, 2007), Numpy (Harris et al., 2020), SciPy (Virtanen et al., 2020), Astropy (Astropy Collaboration et al., 2018), corner.py (Foreman-Mackey, 2016), speclite (Kirkby et al., 2023).

Acknowledgements

Support was provided by Schmidt Sciences, LLC. for J.F., J.C., H.B., and M.O. H.B. acknowledges the GA Junior Star grant No. GM24-10599M for support. The authors would like to thank Sajesh Singh for computing support and Matthew Temple for useful discussions. We thank the anonymous referee for useful comments. This work used resources available through the National Research Platform (NRP) at the University of California, San Diego. NRP has been developed, and is supported in part, by funding from National Science Foundation, from awards 1730158, 1540112, 1541349, 1826967, 2112167, 2100237, and 2120019, as well as additional funding from community partners. M.J.G. acknowledges support from National Science Foundation award AST-2108402. K.E.S.F. acknowledges support from NSF AST-2206096, NSF AST-1831415, and Simons Foundation grant No. 533845.

References

- Astropy Collaboration, Price-Whelan, A. M., Sipz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., Ginsburg, A., VanderPlas, J. T., Bradley, L. D., Pérez-Sánchez, D., de Val-Borro, M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., Ardelean, C., Babej, T., Bach, Y. P., Bachetti, M., Bakanov, A. V., Bamford, S. P., Barentsen, G., Barmby, P., Baumbach, A., Berry, K. L., Biscani, F., Boquien, M., Bostroem, K. A., Bouma, L. G., Brammer, G. B., Bray, E. M., Breytenbach, H., Buddelmeijer, H., Burke, D. J., Calderone, G., Cano Rodríguez, J. L., Cara, M., Cardoso, J. V. M., Cheedella, S., Copin, Y., Corrales, L., Crichton, D., D'Avella, D., Deil, C., Depagne, É., Dietrich, J. P., Donath, A., Droettboom, M., Earl, N., Erben, T., Fabbro, S., Ferreira, L. A., Finethy, T., Fox, R. T., Garrison, L. H., Gibbons, S. L. J., Goldstein, D. A., Gommers, R., Greco, J. P., Greenfield, P., Groener, A. M., Grollier, F., Hagen, A., Hirst, P., Homeier, D., Horton, A. J., Hosseinzadeh, G., Hu, L., Hunkeler, J. S., Ivezić, Z., Jain, A., Jenness, T., Kanarek, G., Kendrew, S., Kern, N. S., Kerzendorf, W. E., Khvalko, A., King, J., Kirkby, D., Kulkarni, A. M., Kumar, A., Lee, A., Lenz, D., Littlefair, S. P., Ma, Z., Macleod, D. M., Mastroiello, M., McCully, C., Montagnac, S., Morris, B. M., Mueller, M., Mumford, S. J., Muna, D., Murphy, N. A., Nelson, S., Nguyen, G. H., Ninan, J. P., O'Neil, M., Ogaz, S., Oh, S., Parejko, J. K., Parley, N., Pascual, S., Patil, R., Patil, A. A., Plunkett, A. L., Prochaska, J. X., Rastogi, T., Reddy Janga, V., Sabater, J., Sakurikar, P., Seifert, M., Sherbert, L. E., Sherwood-Taylor, H., Shih, A. Y., Sick, J., Silbiger, M. T., Singanamalla, S., Singer, L. P., Sladen, P. H., Sooley, K. A., Sornarajah, S., Streicher, O., Teuben, P., Thomas, S. W., Tremblay, G. R., Turner, J. E. H., Terrón, V., van Kerkwijk, M. H., de la Vega, A., Watkins, L. L., Weaver, B. A., Whitmore, J. B., Woillez, J., Zabalza, V., and Astropy Contributors. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *The Astronomical Journal* 156(3):123, September 2018. doi: 10.3847/1538-3881/aabc4f.
- Ba, J. L., Kiros, J. R., and Hinton, G. E. Layer normalization, 2016. URL <https://arxiv.org/abs/1607.06450>.
- Bañados, E., Venemans, B. P., Mazzucchelli, C., Farina, E. P., Walter, F., Wang, F., Decarli, R., Stern, D., Fan, X., Davies, F. B., Hennawi, J. F., Simcoe, R. A., Turner, M. L., Rix, H.-W., Yang, J., Kelson, D. D., Rudie, G. C., and Winters, J. M. An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5. *Nature* 553(7689):473–476, January 2018. doi: 10.1038/nature25180.

Balandat, M., Karrer, B., Jiang, D. R., Daulton, S., Letham,

- B., Wilson, A. G., and Bakshy, E. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In *Advances in Neural Information Processing Systems* 33, 2020. URL <http://arxiv.org/abs/1910.06403>.
- Blandford, R. D. and McKee, C. F. Reverberation mapping of the emission line regions of Seyfert galaxies and quasars. *The Astrophysical Journal*, 255:419–439, April 1982. doi: 10.1086/159843.
- Cackett, E. M., Horne, K., and Winkler, H. Testing thermal reprocessing in active galactic nuclei accretion discs. *Monthly Notices of the Royal Astronomical Society*, 380(2):669–682, September 2007. doi: 10.1111/j.1365-2966.2007.12098.x.
- Cackett, E. M., Bentz, M. C., and Kara, E. Reverberation mapping of active galactic nuclei: From x-ray corona to dusty torus. *iScience*, 24(6):102557, June 2021. ISSN 2589-0042. doi: 10.1016/j.isci.2021.102557. URL <http://dx.doi.org/10.1016/j.isci.2021.102557>.
- Che, Z., Purushotham, S., Cho, K., Sontag, D., and Liu, Y. Recurrent neural networks for multivariate time series with missing values, 2016. URL <https://arxiv.org/abs/1606.01865>.
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014. URL <https://arxiv.org/abs/1412.3555>.
- Collaboration, L. S. Lsst science book, version 2.0, 2009. URL <https://arxiv.org/abs/0912.0201>.
- Cunningham, C. T. and Bardeen, J. M. The Optical Appearance of a Star Orbiting an Extreme Kerr Black Hole. *The Astrophysical Journal*, 183:237–264, July 1973. doi: 10.1086/152223.
- Czerny, B., Panda, S., Prince, R., Kumar Jaiswal, V., Zajaček, M., Martínez Aldama, M. L., Kozłowski, S., Kovacevic, A. B., Ilic, D., Popović, L. Č., Pozo Nuñez, F., Hönig, S. F., and Brandt, W. N. Expectations for time-delay measurements in active galactic nuclei with the Vera Rubin Observatory. *Astronomy & Astrophysics*, 675:A163, July 2023. doi: 10.1051/0004-6361/202345844.
- Fagin, J., Park, J. W., Best, H., Chan, J. H. H., Ford, K. E. S., Graham, M. J., Villar, V. A., Ho, S., and O’Dowd, M. Latent stochastic differential equations for modeling quasar variability and inferring black hole properties. *The Astrophysical Journal*, 965(2):104, apr 2024. doi: 10.3847/1538-4357/ad2988. URL <https://dx.doi.org/10.3847/1538-4357/ad2988>.
- Fagin, J., Chan, J. H.-H., Best, H., O’Dowd, M., Saavik Ford, K. E., Graham, M. J., Park, J. W., and Villar, V. A. Joint modeling of quasar variability and accretion disk reprocessing using latent stochastic differential equations. *The Astrophysical Journal*, 988(1):59, jul 2025. doi: 10.3847/1538-4357/addabc. URL <https://dx.doi.org/10.3847/1538-4357/addabc>.
- Foreman-Mackey, D. corner.py: Scatterplot matrices in python. *The Journal of Open Source Software*, 1(2):24, jun 2016. doi: 10.21105/joss.00024. URL <https://doi.org/10.21105/joss.00024>.
- Greenstein, J. L. Red-Shift of the Unusual Radio Source: 3C 48. *Nature*, 197(4872):1041–1042, March 1963. doi: 10.1038/1971041a0.
- Guo, W.-J., Li, Y.-R., Zhang, Z.-X., Ho, L. C., and Wang, J.-M. Accretion disk size measurements of active galactic nuclei monitored by the zwicky transient facility. *The Astrophysical Journal*, 929(1):19, April 2022. ISSN 1538-4357. doi: 10.3847/1538-4357/ac4e84. URL <http://dx.doi.org/10.3847/1538-4357/ac4e84>.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. Array programming with NumPy. *Nature*, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.
- Hazard, C., Mackey, M. B., and Shimmins, A. J. Investigation of the Radio Source 3C 273 By The Method of Lunar Occultations. *Nature*, 197(4872):1037–1039, March 1963. doi: 10.1038/1971037a0.
- He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition, 2015. URL <https://arxiv.org/abs/1512.03385>.
- Hendrycks, D. and Gimpel, K. Gaussian error linear units (gelus), 2023. URL <https://arxiv.org/abs/1606.08415>.
- Hoshi, A., Yamada, T., Kokubo, M., Matsuoka, Y., and Nagao, T. The relationship of supermassive black holes and host galaxies at $z < 4$ in the deep optical variability-selected active galactic nuclei sample in the cosmos field. *The Astrophysical Journal*, 969(1):11, jun 2024. doi: 10.3847/1538-4357/ad414c. URL <https://dx.doi.org/10.3847/1538-4357/ad414c>.
- Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

- Jha, V. K., Joshi, R., Chand, H., Wu, X.-B., Ho, L. C., Rastogi, S., and Ma, Q. Accretion disc sizes from continuum reverberation mapping of AGN selected from the ZTF survey. *Monthly Notices of the Royal Astronomical Society*, 511(2):3005–3016, 01 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac109. URL <https://doi.org/10.1093/mnras/stac109>.
- Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization, 2017.
- Kirkby, D., Robitaille, T., Weaver, B. A., Tollerud, E., Droettboom, M., Sipőcz, B., Moustakas, Bray, E. M., Bradley, L., Park, A., Burgess, J. M., Bailey, S., Craig, M., Deil, C., Lim, P. L., Kopusov, S., Barbary, K., kgage, Ginsburg, A., Alvarez, M., Kerzendorf, W., dcunning11235, du Mas des Bourboux, H., Crawford, S., Zabalza, V., Sanchez, J., and Günther, H. M. desihub/speclite: 2mass filters, September 2023. URL <https://doi.org/10.5281/zenodo.8347108>.
- Li, J. I.-H., Johnson, S. D., Avestruz, C., Jarugula, S., Shen, Y., Kesler, E., Liu, Z. W., and Mishra, N. Fast and flexible inference framework for continuum reverberation mapping using simulation-based inference with deep learning, 2024. URL <https://arxiv.org/abs/2407.14621>.
- Li, X., Wong, T.-K. L., Chen, R. T. Q., and Duvenaud, D. Scalable gradients for stochastic differential equations. *International Conference on Artificial Intelligence and Statistics*, 2020.
- Maas, A. L., Hannun, A. Y., Ng, A. Y., et al. Rectifier nonlinearities improve neural network acoustic models, 2013.
- Mudd, D., Martini, P., Zu, Y., Kochanek, C., Peterson, B. M., Kessler, R., Davis, T. M., Hoormann, J. K., King, A., Lidman, C., Sommer, N. E., Tucker, B. E., Asorey, J., Hinton, S., Glazebrook, K., Kuehn, K., Lewis, G., Macaulay, E., Moeller, A., O’Neill, C., Zhang, B., Abbott, T. M. C., Abdalla, F. B., Allam, S., Banerji, M., Benoit-Lévy, A., Bertin, E., Brooks, D., Carnero Rosell, A., Carollo, D., Carrasco Kind, M., Carretero, J., Cunha, C. E., D’Andrea, C. B., da Costa, L. N., Davis, C., Desai, S., Doel, P., Fos-alba, P., García-Bellido, J., Gaztanaga, E., Gerdes, D. W., Gruen, D., Gruendl, R. A., Gschwend, J., Gutierrez, G., Hartley, W. G., Honscheid, K., James, D. J., Kuhlmann, S., Kuropatkin, N., Lima, M., Maia, M. A. G., Marshall, J. L., McMahan, R. G., Menanteau, F., Miquel, R., Plazas, A. A., Romer, A. K., Sanchez, E., Schindler, R., Schubnell, M., Smith, M., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Swanson, M. E. C., Tarle, G., Thomas, D., Tucker, D. L., Walker, A. R., and DES Collaboration. Quasar Accretion Disk Sizes from Continuum Reverberation Mapping from the Dark Energy Survey. *The Astrophysical Journal*, 862(2):123, August 2018. doi: 10.3847/1538-4357/aac9bb.
- Novikov, I. D. and Thorne, K. S. Astrophysics of black holes. In *Black Holes (Les Astres Occlus)*, pp. 343–450, January 1973.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.
- Putzky, P. and Welling, M. Recurrent inference machines for solving inverse problems, 2017.
- Salpeter, E. E. Accretion of Interstellar Matter by Massive Objects. *The Astrophysical Journal*, 140:796–800, August 1964. doi: 10.1086/147973.
- Starkey, D. A., Horne, K., and Villforth, C. Accretion disc time lag distributions: applying CREAM to simulated AGN light curves. *Monthly Notices of the Royal Astronomical Society*, 456(2):1960–1973, 12 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv2744. URL <https://doi.org/10.1093/mnras/stv2744>.
- Sun, M., Xue, Y., Trump, J. R., and Gu, W.-M. Winds can ‘blow up’ AGN accretion disc sizes. *Monthly Notices of the Royal Astronomical Society*, 482(2):2788–2794, oct 2018. doi: 10.1093/mnras/sty2885. URL <https://doi.org/10.1093/mnras/sty2885>.
- Temple, M. J., Hewett, P. C., and Banerji, M. Modelling type 1 quasar colours in the era of Rubin and Euclid. *Monthly Notices of the Royal Astronomical Society*, 508(1):737–754, 09 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab2586. URL <https://doi.org/10.1093/mnras/stab2586>.
- Timmer, J. and Koenig, M. On generating power law noise. *Astronomy & Astrophysics*, 300:707, August 1995.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need, 2023. URL <https://arxiv.org/abs/1706.03762>.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A.,

Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Yu, W., Richards, G. T., Vogeley, M. S., Moreno, J., and Graham, M. J. Examining AGN UV/optical variability beyond the simple damped random walk. *The Astrophysical Journal*, 936(2):132, sep 2022. doi: 10.3847/1538-4357/ac8351. URL <https://doi.org/10.3847/1538-4357/ac8351>.

Zu, Y., Kochanek, C. S., and Peterson, B. M. An alternative approach to measuring reverberation lags in active galactic nuclei. *The Astrophysical Journal*, 735(2):80, jun 2011. doi: 10.1088/0004-637X/735/2/80. URL <https://dx.doi.org/10.1088/0004-637X/735/2/80>.

Zu, Y., Kochanek, C. S., Kozłowski, S., and Udalski, A. 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/10.1088/0004-637x/765/2/106>.

Zu, Y., Kochanek, C. S., Kozłowski, S., and Peterson, B. M. Application of stochastic modeling to analysis of photometric reverberation mapping data. *The Astrophysical Journal*, 819(2):122, March 2016. ISSN 1538-4357. doi: 10.3847/0004-637x/819/2/122. URL <http://dx.doi.org/10.3847/0004-637x/819/2/122>.

A. Uncertainty calibration and parameter inference

In Figure 4 top panel, we show that the uncertainties we predict in our UV/optical light curve reconstruction are well calibrated for the nominal test set, while the GPR baseline is misaligned. In the bottom panel, we show that the uncertainties we predict in our parameter posteriors are well calibrated, although overall slightly underconfident. Figure 5 shows the median predictions compared to the true value for each parameter in the nominal test set. The parameters σ/mag (variability amplitude), $\log_{10}(\nu_b/\text{day}^{-1})$ (driving signal break frequency), α_L (driving signal lower power-law slope), $\alpha_H - \alpha_L$ (driving signal additional high-frequency slope), $\log_{10}(M/M_\odot)$ (black hole mass), a (dimensionless spin), θ_{inc} (inclination angle), $(H - R_{\text{in}})/R_g$ (corona height), f_{lamp} (lampost strength), β (temperature slope), z (redshift), and $\log_{10}(\lambda_{\text{Edd}})$ (Eddington ratio) were sampled uniformly between the ranges in Figure 5. Some parameters (i.e., a , f_{lamp} , $(H - R_{\text{in}})/R_g$, and θ_{inc}) are difficult to infer due to having little impact on the transfer functions, causing the median predictions to be close to the mean of the prior.

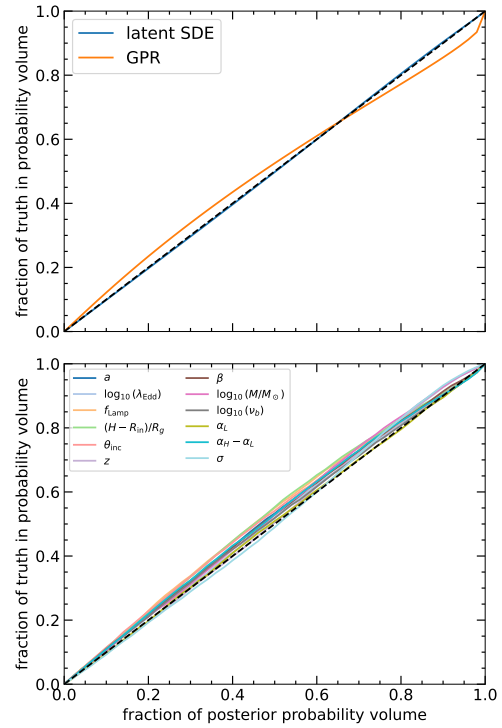


Figure 4. Evaluation of how well the uncertainties are calibrated for the light curve reconstruction (top panel) and parameter inference (bottom panel) by showing the fraction of the truth encompassed within the posterior probability volume across the nominal test set. Perfect uncertainty calibration is shown by the black dashed line along the diagonal.

