The Astrophysical Journal, 980:257 (14pp), 2025 February 20 https://doi.org/10.3847/1538-4357/adae85

Extracting the X-Ray Reverberation Response Functions from the Active Galactic Nucleus Light Curves Using an Autoencoder

Sanhanat Deesamutara[1], Poemwai Chainakun[1,2], Tirawut Worrakitpoonpon[1,2], Kamonwan Khanthasombat[1], Wasutep Luangtip[3,4], Jiachen Jiang[5,6], Francisco Pozo Nuñez[7], and Andrew J. Young[8]
[1] School of Physics, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand; pchainakun@g.sut.ac.th
[2] Center of Excellence in High Energy Physics and Astrophysics, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand
[3] Department of Physics, Faculty of Science, Srinakharinwirot University, Bangkok 10110, Thailand
[4] National Astronomical Research Institute of Thailand, Chiang Mai 50180, Thailand
[5] Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV 4 7AL, UK
[6] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[7] Astroinformatics, Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, D-69118 Heidelberg, Germany
[8] H.H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, UK
Received 2024 July 1; revised 2025 January 7; accepted 2025 January 24; published 2025 February 20

Abstract

We study the X-ray reverberation in active galactic nuclei (AGN) using the variational autoencoder (VAE), which is a machine learning algorithm widely used for signal processing and feature reconstruction. While the X-ray reverberation signatures that contain the information of the accretion disk and the X-ray-emitting corona are commonly analyzed in the Fourier domain, this work aims to extract the reverberation response functions directly from the AGN light curves. The VAE is trained using the simulated light curves that contain the primary X-rays from the lamppost corona, varying its height and the corresponding reflection X-rays from the disk. We use progressively more realistic light-curve models, such as those that include the effects of disk-propagating fluctuations and random noises, to assess the ability of the VAE to reconstruct the response profiles. Interestingly, the VAE can recognize the reverberation patterns on the light curves; hence, the coronal height can be predicted. We then deploy the VAE model on the XMM-Newton data of IRAS 13224–3809 and directly estimate, for the first time, the response functions of this source in various observations. The result reveals the corona changing its height between $3r_g$ and $20r_g$, which is correlated with the source luminosity and in line with previous literature. Finally, we discuss the advantages and limitations of this method.

Unified Astronomy Thesaurus concepts: Reverberation mapping (2019); X-ray astronomy (1810); Active galactic nuclei (16); Black hole physics (159)

1. Introduction

Active galactic nuclei (AGN) are the most luminous objects in the Universe, which are powered by an accretion of gas onto the central supermassive black hole. The accreting gas can form a disklike structure referred to as an accretion disk that converts the gravitational energy into thermal optical/UV radiation. The disk photons gain more energy by Compton upscattering with hot electrons inside the corona, raising the photon energy to X-rays (C. S. Reynolds & M. A Nowak 2003). The X-rays from the corona can travel back to the disk, be reprocessed, and reflect off the disk in the form of a reflection spectrum, carrying out the characteristic features of the emission lines to an observer (I. M. George & A. C. Fabian 1991; R. R. Ross et al. 1999; R. R. Ross & A. C. Fabian 2005; J. García & T. R. Kallman 2010; J. García et al. 2013). How much the Doppler and relativistic effects distort the lines determines how close the inner edge of the disk is to the black hole, hence allowing us to probe the innermost region of the AGN.

The X-ray variations in AGN occur on a wide range of timescales (e.g., S. Vaughan et al. 2003). The variabilities on long timescales relate to the fluctuations of the mass accretion rate that propagate inward along the disk, modulating the outer (spectrally softer) regions first before the inner (spectrally harder) regions, hence producing the hard lags (O. Kotov et al. 2001; P. Arévalo & P. Uttley 2006). On shorter timescales, the fluctuations in the direct continuum cause the reflection to reverberate, so the changes in the reflection-dominated energy bands are expected to lag behind the changes in the continuum-dominated band. The amplitude of the reverberation lags relates to the light-travel time between the corona and the reflecting region, providing us insights about the disk–corona geometry (see P. Uttley et al. 2014; E. M. Cackett et al. 2021, for a review).

Theoretical works in calculating the X-ray reverberation lags have been undertaken early on (C. S. Reynolds et al. 1999; A. J. Young & C. S. Reynolds 2000), before the first hints of the reverberation were seen in Ark 564 (I. M. McHardy et al. 2007), followed by the first robust detection in 1H 0707–495 (A. C. Fabian et al. 2009), where the fluctuations in the soft (reflection-dominated) energy band lagged behind those in the hard (continuum-dominated) band by ∼30 s. Since then, the number of AGN that exhibit the X-ray reverberation lags have increased (e.g., B. De Marco et al. 2013; E. Kara et al. 2016). While the nature of the corona is still uncertain, the dynamics of the corona were evident, e.g., in Mrk 335 (D. R. Wilkins & L. C. Gallo 2015), IRAS 13224–3809 (W. N. Alston et al. 2020; M. D. Caballero-García et al. 2020; P. Chainakun et al. 2022), 1H 0707–495 (S. Hancock et al. 2023; N. Mankatwit et al. 2023), and NGC 4051 (N. Kumari et al. 2023). By studying the time delays between UV and X-ray bands, a

change in the geometry of the system (e.g., the height of the corona) was also suggested (K. Kumari et al. 2024).

Meanwhile, the fully relativistic X-ray reverberation models were continuously developed based on the ray-tracing simulations that trace photon trajectories along the Kerr geodesics between the corona, the disk, and the observer (e.g., D. R. Wilkins & A. C. Fabian 2013; E. M. Cackett et al. 2014; D. Emmanoulopoulos et al. 2014; P. Chainakun et al. 2016; A. Epitropakis et al. 2016). In this way, we can calculate the flux of the reflection photons from different parts of the disk as seen by the observer and as a function of time after the direct continuum. This is known as the response function, which can be used to generate realistic light curves based on the convolution theorem. The lags between the light curves in different energy bands then are estimated in the Fourier frequency space (P. Uttley et al. 2014). Comparison of the lags between the observations and the models is traditionally performed in the Fourier domain, by fitting either the lag-frequency spectra (e.g., W. N. Alston et al. 2020; M. D. Caballero-García et al. 2018, 2020) or the power spectral density (PSD) obtained from the modulus square of the Fourier transform (e.g., D. Emmanoulopoulos et al. 2016; P. Chainakun et al. 2022).

This work aims to develop machine learning (ML) models to extract the response functions from the AGN light curves so that they can be directly fitted with the theoretical response profiles. We employ the variational autoencoder (VAE), which is based on the multilayer perceptron (MLP) for signal reconstruction and reverberation feature extraction. Several applications of autoencoding architectures have been applied to astronomical problems such as spectral energy distribution feature extraction (J. Frontera-Pons et al. 2017), strong lensing identification (T.-Y. Cheng et al. 2020), and X-ray binary pattern recognition (J. K. Orwat-Kapola et al. 2022). However, there was no previous work that has been applied to predict the X-ray reverberation response function. Thus, we are motivated by these previous studies to adopt the VAE to our problem and examine it in terms of performance and sensitivity.

Simulated AGN light curves in several scenarios are produced to test the effectiveness of the VAE model in reconstructing the response functions. This is very challenging since the AGN light curves are known to be stochastic (i.e., different realizations of the same underlying physical process may look entirely different). This work can provide an independent tool to discover the underlying process straightforwardly in the time domain and to constrain the key parameters of the system. In addition, we also apply the model to IRAS 13224–3809, which is the well-known AGN that exhibits strong and variable signatures of the X-ray reverberation due to its dynamic corona, where the coronal height seems to increase with the source luminosity (e.g., W. N. Alston et al. 2020; M. D. Caballero-García et al. 2020; P. Chainakun et al. 2022). In brief, IRAS 13224–3809 has been identified as a narrow-line Seyfert 1 galaxy (M. P. Véron-Cetty & P. Véron 2006), located at redshift z of 0.06580;[9] this corresponds to the luminosity distance D_L of 288 Mpc—in which the cosmological parameters of $H_0 = 73\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$, $\Omega_{\mathrm{matter}} = 0.27$, and $\Omega_{\mathrm{vacuum}} = 0.73$ have been assumed (J. Jiang et al. 2018). The source was well observed by the XMM-Newton observatory (F. Jansen et al. 2001) mainly in 2011 and 2016, in which its total exposure time is ∼2000 ks, making this a good observational template for this work. The ultimate goal of this work is to directly derive the response functions of IRAS 13224–3809 and discuss the potential use of this method.

The IRAS 13224–3809 observations and detailed data reduction are given in Section 2. The KYNXILREV model to produce the disk response functions is explained in Section 3. All light-curve models investigated here are presented in Section 4. The ML models to extract the response profiles from the light curves, as well as the signal similarity indicators, are presented in Section 5. The performance of the models and the results when applying to IRAS 13224–3809 are presented in Section 6. The discussion and conclusion are given in Section 7.

Table 1
XMM-Newton Data Set of IRAS 13224–3809 Used in This Work

Observation ID	Observed Date	Exposure Time[a] (ks)	L_{X}[b] ($\times 10^{42}$ erg s^{-1})	SNR[c]
0673580101	2011-07-19	50.26	4.50	26.81
0673580201	2011-07-21	64.63	3.72	50.06
0673580301	2011-07-25	67.67	2.59	50.75
0673580401	2011-07-29	98.81	6.10	62.23
0780561301	2016-07-10	121.21	3.42	55.46
0780561501	2016-07-20	113.82	2.79	70.17
0780561601	2016-07-22	114.98	5.38	112.06
0780561701	2016-07-24	112.39	3.14	66.60
0792180101	2016-07-26	119.64	2.58	121.11
0792180201	2016-07-30	118.99	3.55	72.62
0792180301	2016-08-01	103.28	2.56	79.20
0792180401	2016-08-03	104.90	6.86	130.71
0792180501	2016-08-07	111.18	4.23	62.44
0792180601	2016-08-09	114.51	5.92	93.38

Notes.
[a] The useful exposure time after the removal of high background flaring periods.
[b] The source X-ray luminosity in the 2–10 keV band; the data were obtained from the extended data Figure 2 in W. N. Alston et al. (2020).
[c] The SNR of data set determined using Equation (1) (see text for further details).

2. Observations and Data Reduction

To test our proposed method with real observational data, in this work we obtained the data of IRAS 13224–3809 previously observed by the XMM-Newton observatory from the XMM-Newton Science Archive;[10] they are listed in Table 1. The observational data were reprocessed using the science analysis software (SAS) version 21.0.0[11] with the up-to-date set of calibration files in 2024 April; the SAS task EPPROC with its default parameter values was used to generate the calibrated event lists for all observations; note that here only data sets from the pn detector were used in order to obtain good-quality timing data. In addition, we also removed the exposure periods that were affected by high background flaring events from the reprocessed event files; this was done manually by inspecting the instrument count rate in the high energy band (10–12 keV) and removing the observational periods having

[9] Fiducial redshift obtained from http://ned.ipac.caltech.edu.

[10] http://nxsa.esac.esa.int

[11] https://www.cosmos.esa.int/web/xmm-newton/sas

anomalously high count rates of >0.4 counts s^{-1}. This results in a factor of $\sim$20% reduction of observational time of each individual data set and a random building of the time gaps in the observational data, typically with the periods of $\sim$0.1–10 ks. The total remaining exposure time (i.e., a summation of good time intervals) of each observation is shown in column (3) of Table 1.

We then extracted background-subtracted light curves of the source with the temporal resolution of 1 s from each individual observation from the events that were flagged as #XMMEA_EP and PATTERN $\leqslant 4$; the source extracting area was defined as a circular region with a radius of 20″ centered on the AGN position, while that of the background was a 50″ radius circular region located in the source-free area close to and also on the same chip as the source's region; these criteria were applied to create light curves in the 0.3–1 keV (reflection-dominated) band for all observations.

Finally, we also estimate the quality of the obtained light curves by calculating their signal-to-noise ratio (SNR). We use a similar definition of the SNR to what would be employed when we add noises to the simulated light curves, which can be written as (M. D. Sacchi 2016)

$$\mathrm{SNR} = \frac{\|\boldsymbol{s}\|_2^2}{\alpha^2 \|\boldsymbol{n}\|_2^2}, \tag{1}$$

where $\boldsymbol{s}$ and $\boldsymbol{n}$ represent the signal and noise, respectively, and $\|\cdot\|_2^2$ denotes the squared l^2-norm of the parameters. This equation assumes that noise is composed of zero-mean, unit-variant Gaussian noise, while coefficient α amplifies the noise to the designated level of SNR. Here α was fixed at unity in the calculation, while the SNR was calculated from the signal and noise powers, i.e., the PSD, in the frequency range of 0.0625–1 mHz. This enables us to compare the quality of real observational data with our simulated data. The SNR result was presented in column (5) of Table 1. We note that while the noise defined in Equation (1) could represent stochastic noise produced by any processes relating to the source, indeed, the noise detected in observational data might be dominated by—or at least some fraction could be contributed from—nonstochastic noise, in particular the detector noise. However, this kind of noise should be ignorable in our case, as it would affect only on pn data below 250 eV.[12] Furthermore, the observational data used in this work also have good quality, in which their SNRs measured by the traditional method are $\gtrsim$100 (i.e., Table 1 of N. Nakhonthong et al. 2024), suggesting that the noise is unlikely to dominate the total signal. Examples of some observed and simulated light curves are presented in Appendix A.

3. Response Functions from KYNXILREV

The KYNXILREV model is a computational code for X-ray reverberation from the accretion disk around the central black hole (M. D. Caballero-García et al. 2018; M. Dovčiak et al. 2004a, 2004b). It assumes that the disk is geometrically thin, optically thick, and extended from the innermost stable circular orbit to $1000r_g$ ($1r_g = 1\ GM/c^2$, where G is the gravitational constant, M is the central mass, and c is the speed of light). In KYNXILREV, the lamppost corona is employed, and the XILLVER model (J. García & T. R. Kallman 2010; J. García et al. 2013) with the XSTAR code (T. Kallman & M. Bautista 2001) is used to calculate the reprocessing on the disk as it is irradiated with an isotropic flash of X-ray continuum (i.e., the primary X-rays) from the corona. The primary X-rays are characterized by a power law with the photon index (Γ) and with the high energy cutoff. The disk also has a constant density, and its ionization state varies accordingly with the incident flux.

By using KYNXILREV, the response function of the disk reflection is obtained from the ray-tracing simulations that trace photon paths along the Kerr geodesics between the source, the disk, and the observer. Since we aim to deploy the developed VAE model on IRAS 13224–3809, we fix the central mass, the inclination, and the black hole spin to the values found in previous studies, which are $M = 2 \times 10^6\ M_\odot$ (W. N. Alston et al. 2020), $i = 45°$ (M. D. Caballero-García et al. 2020), and $a = 0.998$ (J. Jiang et al. 2018). Note that Γ may vary among different observations, but to avoid the model degeneracy, we fix $\Gamma = 2$. In this work, we use `xillverD-5.fits` as a precalculated table model for XILLVER. Different choices of the table models can lead to different amplitudes of the obtained response functions. Any unspecified KYNXILREV parameters are set to their default values.

We choose to normalize the area under the response function and use the reflection fraction (R) as a normalization parameter to regulate the importance between the X-ray reflection and direct continuum components (e.g., P. Chainakun et al. 2022). While a lamppost assumption fixes the ratio of photons reaching the disk to those reaching infinity, we define R as the ratio of reflection flux to continuum flux. This allows R to vary even with fixed geometry, due to factors such as disk density and specific reflection models used (e.g., REFLIONX and XILLVER). Allowing R to vary produces dilution effects similar to when we vary, e.g., disk density or its ionization state (i.e., the first response time of the response function remains the same while the dilution or normalization changes). Therefore, instead of adjusting a large number of parameters in KYNXILREV that influence dilution effects, we have chosen to fix all of these and vary only R. This allows for some flexibility in our model while still maintaining the core structure of the lamppost geometry.

4. Modeled Light Curves

The AGN light curve in each energy band consists of the X-ray continuum and the disk reflection components. We investigate the progressively more realistic light-curve models when the effects of the disk-propagating fluctuations are excluded and included, referred to as the REV model and REV+PROP model, respectively. Then, we investigated the case where the random noise is included (REV+PROP+NG model) and the case where the Poisson noise is also taken into account (REV+PROP+NG+NP model). All models focus on the light curves in the soft (reflection-dominated, 0.3–1 keV) energy band given by $s(t)$.

4.1. The REV Model

Let us assume a driving signal $a(t)$, so that the soft-band light curves can be written as (D. Emmanoulopoulos et al. 2014;

[12] https://xmm-tools.cosmos.esa.int/external/xmm_user_support/documentation/uhb/epicdetbkgd.html; see also J. A. Carter & A. M. Read (2007).

P. Chainakun et al. 2016; A. Epitropakis et al. 2016)

$$s(t) = ba(t) + Ra(t) \otimes \psi(t), \quad (2)$$

where $a(t) \otimes \psi(t) = \int_0^t a(t')\psi(t - t')dt'$. The driving signal $a(t)$ in Equation (2) is generated using the `stingray.simulator` from the PSD in the form of a power-law red noise (D. Huppenkothen et al. 2019). The slope of the PSD is randomly varied between 1 and 3, covering the mean index of $\sim$2 as reported in O. González-Martín & S. Vaughan (2012). The parameter b is the normalization factor tied to the continuum flux contribution in this band:

$$b = \int_{0.3\ \mathrm{keV}}^{1\ \mathrm{keV}} E^{-\Gamma} dE. \quad (3)$$

Parameter $\psi(t)$ is the disk response function due to a lamppost source, and R is the normalization related to the reflection fraction. The first and the second terms on the right-hand side of the light-curve equation then represent the flux contributed by the direct continuum and the X-ray reflection, respectively.

Furthermore, the area under $\psi(t)$ is normalized to b, so that $R = 1$ means an equal contribution between the continuum and the reflected flux in the corresponding energy band. Given the complexity of the model and to avoid a prohibitively large number of parameters, the photon index of the X-ray continuum is fixed at $\Gamma = 2$ in this model.

We produce the model grid of $\psi(t)$, for an ensemble of 29 coronal heights: [h = 2.3, 3, 4, 5, 6, …, 30r_g]. For each grid of h, we change the seed in the light-curve simulator in `stingray.simulator` to produce 100 different $a(t)$, while uniformly randomly selecting the reflection fraction in the soft band within the range of $R \in [1, 3]$. Therefore, the data contain 2900 light curves in total. We then investigate the performance of the models to extract the response functions from these light curves. This represents the simplest model with which we begin our investigation. In the final model applied to the IRAS 13224–3809 data, the range of R is extended to [1, 10], and additional factors such as disk-propagating fluctuations and noise effects are incorporated.

4.2. The REV+PROP Model

This model is built by including additional top-hat response functions, TH(t), in the light-curve equation. The top-hat function represents the response due to the disk-propagating fluctuations that produce the hard lags on longer timescales than the soft reverberation lags (e.g., Y. E. Lyubarskii 1997; O. Kotov et al. 2001; P. Arévalo & P. Uttley 2006). The shape of the top-hat response is controlled by the centroid, τ, and the width, w. In this case, we generalize the light-curve equation so that it can represent the light curve in any energy bands of interest. The area under both $\psi(t)$ and TH(t) is normalized to 1. The light-curve formula is given by (e.g., W. N. Alston et al. 2014; P. Chainakun et al. 2023)

$$s(t) = a(t) + R\ a(t) \otimes \psi(t) + P\ a(t) \otimes \mathrm{TH}(t), \quad (4)$$

where the normalization R and P regulate the importance of the responses due to the reverberation and the disk-propagating fluctuation processes, respectively. A uniform distribution function is used to generate a random number for $R \in [1, 3]$ and $P \in [1, 3]$. When $R > P$, the light curve is dominated by the effects of the X-ray reverberation, and when $P > R$, the propagating fluctuations dominate. We note that this study does not account for the potential contribution of a warm corona to the X-ray soft excess (e.g., A. Kubota & C. Done 2018; R. Middei et al. 2018; D. Porquet et al. 2018; F. Ursini et al. 2020; X. Xu et al. 2021; D. R. Ballantyne et al. 2024). Incorporating this component may affect the dilution of time lags that play a similar role to adjusting the reflection fraction. See Appendix B for more discussion in the case that R of the real data is different from that of the training data.

We produce the model grids for $\psi(t)$ similar to what is done in the REV model. On the other hand, the TH(t) is produced by the uniform distribution function, where the centroid and the width are randomly selected from $\tau \in [10^4, 5 \times 10^4]$ s and $w \in [10^3, 5 \times 10^3]$ s. This ensures that the average response time and the width of the response of TH(t) are larger than those of $\psi(t)$, which is likely the case for the disk-propagating fluctuation lags that dominate on longer timescales. In this case, we simulate 2900 light curves using different driving signals (i.e., the slope of the underlying PSD of the light curve is randomly varied to be between 1 and 3, and the seed for generating $a(t)$ is also randomly varied).

4.3. The REV+PROP+NG Model

Now, we investigate the effect of random noise by introducing uncorrelated variability to the light curves. The Gaussian noise is added to the light-curve model under different conditions of SNR, following the code provided by M. D. Sacchi (2016). The light-curve equation can be written by

$$s(t) = a(t) + R\ a(t) \otimes \psi(t) + P\ a(t) \otimes \mathrm{TH}(t) + \alpha\ n(t). \quad (5)$$

According to Equations (1) and (5), α and SNR can be fixed or free to serve different purposes. We fix α to determine the SNR in Equation (1), while in Equation (5) α is instead determined as a function of SNR. We produce the light curves with various SNR scenarios, ranging from 0.1 to 1000, by varying α as a mixture weight to obtain the certain designated value of SNR.

4.4. The REV+PROP+NG+NP Model

A further additional feature to the light curve is Poisson noise. The motivation is to mimic the readout process of CCD to make the simulated light curves more realistic. Light curves are resampled using the built-in `NumPyrandom.poisson` function. These light curves are treated as the new signal, where the Gaussian noise is also included with a similar method to that mentioned in the previous section.

The light curves from the REV+PROP+NG+NP model are used to train the model that is applied to IRAS 13224–3809 data, as they closely resemble the observed ones. An example of light curves with the SNR corresponding to the observed data is shown in Appendix A. All other setups for the modeled light curve are consistent with those in the REV+PROP+NG model, except that we extend the reflection fraction range to $R \in [1, 10]$. Note that, in this study, we adopt simplified assumptions by fixing certain parameters (e.g., the photon index, black hole spin, inclination, and disk density) and employing a lamppost source with a flexible free parameter, R. This approach generates disk response functions with different dilution effects (different R), possibly induced by variations in other parameters that are held constant. While a true composite spectral and timing model may

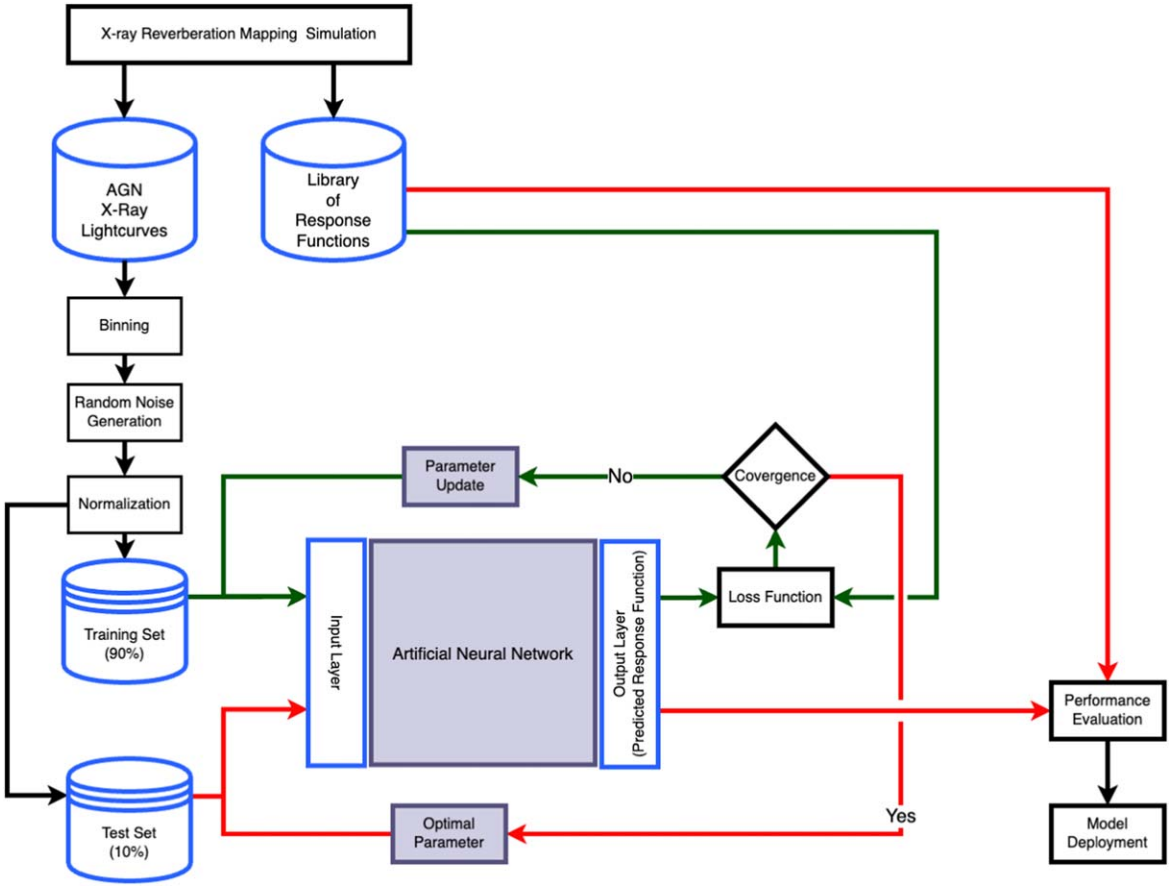

Figure 1. Generalized workflow for supervised ML. The light curves are binned with the bin sizes of 1 and 5 s for the REV and REV+PROP models, respectively. The green and red lines represent the pipeline for the optimization process and performance evaluation, respectively. See text for more details.

be difficult to resolve, these choices help avoid an excessive number of potentially degenerate parameters.

Nevertheless, we also show in Appendix B the case when the model is trained with varying R, while fixing, e.g., Γ, i, and M can still predict h of the lamppost source when testing with the new data generated from a "strict" lamppost case, where we do not employ the variable R but instead generate the data by varying Γ, i, and M within the uncertainty of previously observed values for IRAS 13224–3809. This suggests that the current R range is sufficient when applying the model to the specific AGN IRAS 13224–3809, as the test data set is tailored to variations in Γ, i, M within the observational uncertainties of this source.

5. ML Models for Extracting the Response Functions

The ML extraction pipeline involves supervised ML techniques, which is summarized in Figure 1. The data, which in our case are the simulated AGN light curves with X-ray reverberation features, are binned using 1 and 5 s time bins for REV and REV+PROP models, respectively, and are separated into training and test sets by the fraction of 90% of the training set and 10% of the test set. Note that there are no missing data in the simulated light curves used to train the machine. Therefore, when applied to IRAS 13224–3809 data, we employ the piecewise cubic Hermite interpolating polynomial (PCHIP; F. N. Fritsch & J. Butland 1984) to replace the missing data and readout error with estimated values. Training data are fed to the model, as it will learn to find the optimal transformation from input data to designated output feature (green line), by achieving the global minimum of the loss function via the numerical optimization process. Once the global optimum is reached, performance evaluation will take place by implementing the model in the test set to investigate whether the model is overfitted (red line). If it is not, the model is ready for deployment on the real observational data.

5.1. Autoencoder

The mathematical aspect of the autoencoder (AE) is based on the MLP that works by mapping the input feature ($\boldsymbol{x}$: light curve, in the case of our work) to the designated output feature space (I. Goodfellow et al. 2016), while each light curve consists of `INPUT_DIM` data points, which refers to the length of the simulated light curve (`INPUT_DIM` = 112400/`bin_size` in our study). The `bin_size` is assigned as 1 s for the REV model and 5 s for all other models. A combination of simple functions provides the learning pipeline for the representation of input features in terms of the hidden layer ($\boldsymbol{h}$). It can be simplified to the composition of linear mapping and activation functions, called fully connected layers, which can be mathematically represented by

$$\begin{aligned}
\boldsymbol{h}_1 &= f(\boldsymbol{A}_0\boldsymbol{x} + \boldsymbol{b}_0)\\
\boldsymbol{h}_2 &= f(\boldsymbol{A}_1\boldsymbol{h}_1 + \boldsymbol{b}_1)\\
&\vdots\\
\boldsymbol{h}_n &= f(\boldsymbol{A}_{n-1}\boldsymbol{h}_{n-1} + \boldsymbol{b}_{n-1})\\
\boldsymbol{y} &= f(\boldsymbol{A}_n\boldsymbol{h}_n + \boldsymbol{b}_n). \qquad (6)
\end{aligned}$$

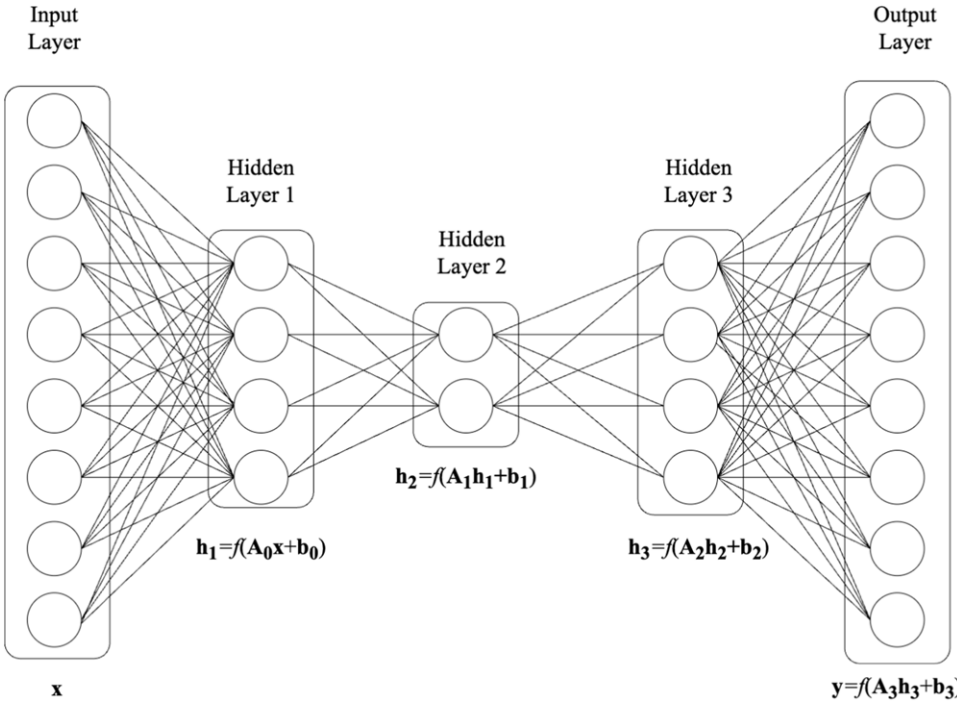

Figure 2. Example for autoencoding architecture with three hidden (fully connected) layers. The network consists of the input layer, hidden layers, and the output layer, resembling Equation (6).

The network in Equation (6) can be illustrated in Figure 2. Variables $\boldsymbol{A}$ and $\boldsymbol{b}$ are, respectively, the weight metric and the bias tensor, while their dimensions are determined by the number of input and output neurons for each layer. The function f is the so-called activation function, which adds nonlinearity to the model and allows the network to be suited to the complex system. It is often used in the form of a hyperbolic tangent (tanh), rectified linear unit (ReLU), or sigmoid function. Finally, the target layer ($\boldsymbol{y}$: response function), with length `OUTPUT_DIM` = 10,000 in our study, shall be returned from the output layer.

We implemented the shallow autoencoder via PyTorch (A. Paszke et al. 2019), consisting of two concatenated MLP networks, encoding layers, and decoding layers. Its architecture is shown in Table 2. The optimization process is performed to obtain the best-fit parameters for the network, subjected to the reconstruction loss. For example, the loss functions widely used in this architecture are L1 loss (Euclidean distance), mean square loss, and cross-entropy loss. The optimization rate is restricted by the user-defined learning rate, to avoid overfitting and to guarantee the convergence to the local minimum.

5.2. Variational Autoencoder

The mathematical expression of VAE can be written in a similar way to Equation (6), comprising two consecutive MLP networks as encoding and decoding layers. However, VAE is interposed by the reparameterization process. The encoder transforms the input feature into a lower dimension, achieving a mutual representation of the transformed features of the arithmetic mean (μ) and variance (σ^2) as the latent space. The decoder then transforms the resampled latent space by increasing the dimension of the feature to the designated output shape. Reparameterization is performed to resample the latent space, resulting in the latent vector $\boldsymbol{z}$, which is recovered from encoded μ, σ, and an additional random number (ϵ), derived from a Gaussian probability distribution, with respect to

$$z = \mu + \epsilon\ \sigma. \tag{7}$$

Table 2
Architecture for Standard AE

Layer	Input × Output
Input layer	`INPUT_DIM` × 400
Hidden layer 1	400 × 400
Hidden layer 2	400 × 200
Hidden layer 3	200 × 400
Hidden layer 4	400 × 400
Output layer	400 × `OUTPUT_DIM`

Note. The user must provide the number of data points for each input feature (`INPUT_DIM`), as well as that for the output feature (`OUTPUT_DIM`).

We also use the PyTorch library and adopt the code from M. J. Kang (2021) to construct the VAE as in Table 3. The optimization process is governed by a linear combination of binary cross-entropy loss (BCE), to determine how well the response functions are predicted ($\hat{y}(t)$) compared with its ground truths ($y(t)$), and Kullback–Leibler divergence loss (KL), to penalize the performance of the reparameterization process by considering the behavior of μ and σ in latent space, with a time-wise weight for each point of response:

$$\mathrm{BCE}(\hat{y},\, y) = -\sum_i w_i (y_i \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)),$$
$$\text{s.t.} \sum_i w_i = 1, \tag{8}$$

$$\mathrm{KL}(\mu,\, \sigma) = -\frac{1}{2}\sum_j (1 + \log(\sigma_j) - \mu_j^2 - \sigma_j). \tag{9}$$

An index i is the time-stamp index that runs from 1 to `OUTPUT_DIM`, while j is indexing every element in

Table 3
VAE Composition

Encoding Layers	
Layer	Input × Output
Input layer	`INPUT_DIM` × 400
Hidden layer	400 × 400
Mean	400 × 200
Variance	400 × 200
Decoding Layers	
Layer	Input × Output
Hidden layer 2	200 × 400
Hidden layer 3	400 × 400
Output layer	400 × `OUTPUT_DIM`

Note. The network performs a combination of linear mapping to translate the input feature to the target vector. Note that the VAE assumes the Gaussianity behavior in the data. Per mathematical definition, features computed by encoding layers shall be represented in terms of mean (μ) from the mean layer and variance (σ^2) from the variance layer, respectively. The mean and variance layers encoded the yield of the hidden layer from length 400 to 200. See text for more details.

representation tensors μ and σ (j runs between 1 and 200), as defined in Table 3. Combining Equations (8) and (9), the loss function for VAE can then be written as

$$\mathcal{L} = \mathrm{BCE}(\hat{y}, y) + \mathrm{KL}(\mu, \sigma). \tag{10}$$

5.3. Signal Similarity Indicator

In order to verify the closeness between a true response function and one predicted by the ML model, statistical indicators are used to determine the proximity between the predicted curves ($\hat{y}(t)$) and their ground truths ($y(t)$). Standard metrics such as R^2-score, mean absolute error (MAE), and mean square error (MSE) can also be used to measure the performance of prediction. However, we instead use cross-correlation (CC) and signal coherence as the performance metrics, as they are widely used to estimate similarity between two time-series data. CC is the measurement of the similarity between two time series under different time delays, with the governing equation

$$(\hat{y} \star y)(\tau) = \int_{-\infty}^{\infty} \overline{\hat{y}(t)} y(t - \tau) dt, \tag{11}$$

where $\overline{\hat{y}(t)}$ denotes the complex conjugate of $\hat{y}(t)$. We use the `scipy.signal.correlate` function to determine the correlation between the signals. Only zero-lag ($\tau = 0$) CC is considered to determine the prediction accuracy of the model.

On the other hand, the signal coherence is also used to estimate the relationship between two signals $\hat{\mathbf{y}}$ and $\boldsymbol{y}$ in a Fourier space, as the likelihood between two power spectra. The equation is defined as

$$C_{\hat{y}y}(f) = \frac{|G_{\hat{y}y}(f)|^2}{G_{\hat{y}\hat{y}}(f) G_{yy}(f)}. \tag{12}$$

$G_{\hat{y}y}$ is cross-spectral density between $\hat{\mathbf{y}}$ and $\boldsymbol{y}$, while $G_{\hat{y}\hat{y}}$ and G_{yy} are the autospectral density of $\hat{\mathbf{y}}$ and $\boldsymbol{y}$, respectively. In this work, the built-in function `scipy.signal.coherence` is used to compute the coherence at the 0 Hz component, as the response functions are composed of low-frequency elements.

Examples of the similarity measurements using the CC and signal coherence are presented in Figure 3. We produce mock response functions different in shape and response time. It can be seen that the signal coherence seems to drop faster, especially when their shapes are mismatched.

6. Results

6.1. Model Performance

To evaluate the performance of VAE, we set up three simultaneous experiments that handle different sources of data, comparing with standard AE as a baseline. To control the computation, we define the network to use a similar amount of trainable parameters, together with the Adam optimizer at a learning rate of 10^{-3} and a training process of 2000 epochs.

Table 4 shows the accuracy results in the cases of noise-free light curves generated by the REV and REV+PROP models, using driving signals with randomized geometrical parameters. It can be seen that, for the noise-free light curves, the accuracy of both the AE and VAE models to extract the reverberation response function is quite high (CC and signal coherence are more than $\sim$0.9) whether or not the effects due to the disk-propagating fluctuations are included. The examples of the extracted response functions compared to the real response profiles for The REV+PROP model are also presented in Figure 4. The models perform well for a broad range of coronal heights. The shape of the response function is best recovered in the case of $h = 20r_g$, as can be seen from the distinct double-peak feature that consists of a small peak at $\sim$400 s and a broader peak at $\sim$500 s. These double-peak features are not clearly visible in the predicted response functions for $h = 18r_g$, $17r_g$, and $12r_g$. For $h = 2.3r_g$ and $7r_g$, the recovered response functions are good with slightly larger delays and smaller delays, respectively.

Now, we investigate the REV+PROP+NG model, which is the case when random noises contaminated the light curve. The training and testing are performed under an SNR ranging from 0.1 to 1000. Note that the uncorrelated noise is added after the binning process. How the accuracy decreases with increasing SNR is presented in Figure 5. We can see that both CC and signal coherence gradually decline when noise becomes more dominant. The trend of decreasing accuracy with decreasing SNR is quite comparable between the training and test data sets, meaning that the models are not overfitting. However, the accuracy measured by CC seems to drop more rapidly than the coherence. These results suggested that the models perform well (accuracy from both indicators is $\gtrsim$0.9) if the SNR $\gtrsim$ 50.

We also used VAE with Poissonian resampled light curves and additional Gaussian noise (REV+PROP+NG+NP model) in the identical SNR range to train and test the model. CCs for training and test light curves agree at an accuracy of $\gtrsim$0.9 at SNR $\gtrsim$ 10. However, signal coherence gradually declines when SNR $\lesssim$ 50. Thus, it could be interpreted that the prediction sensitivity of autoencoder for the REV+PROP+NG +NP model also agrees at an accuracy of $\gtrsim$0.9 and SNR $\gtrsim$ 50, comparable to that of the REV+PROP+NG model.

We also evaluate the robustness of the model for the general applicability when the observed or test data have slightly different characteristics from the training data (i.e., have

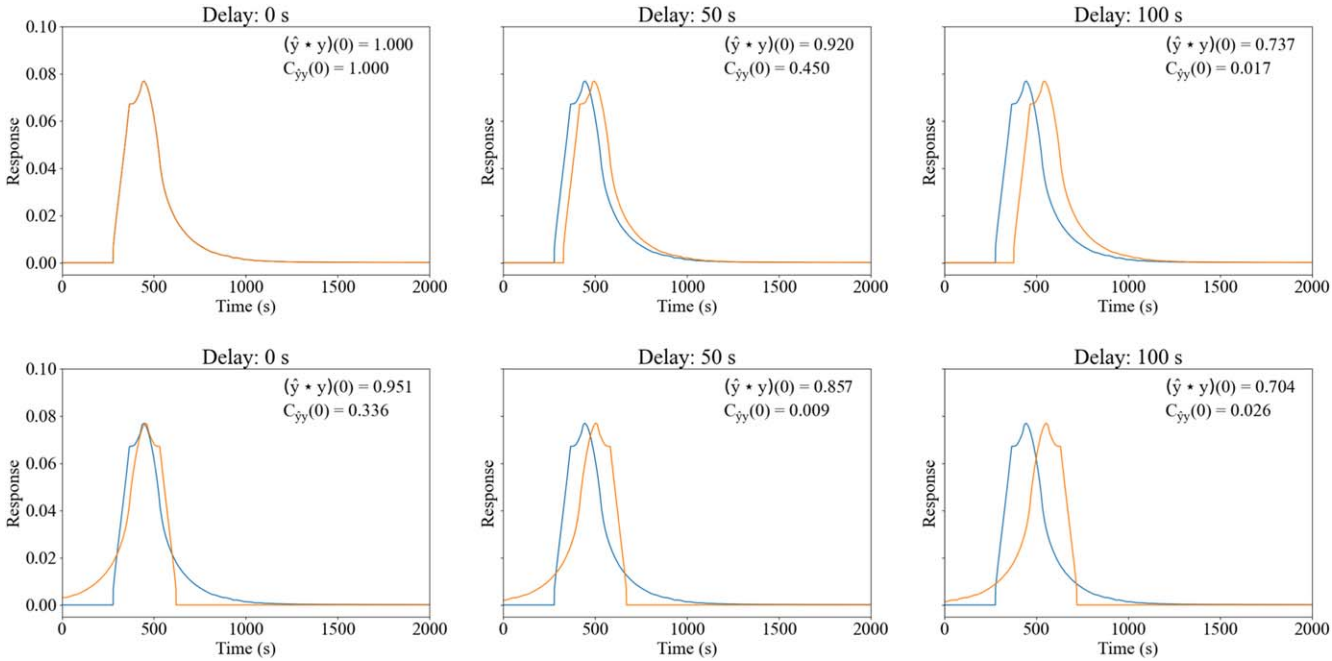

Figure 3. Similarity measurements (CC and signal coherence) of mock response functions under different conditions. Top panels: two response signals are different only by a specific time shift. Bottom panels: two responses are different by the time shift, and one is also the mirror image of the other. See text for more details.

Table 4
Prediction Likelihood for the Response Function Extracted by Training the Autoencoders with Simulated REV and REV+PROP Model Light Curves

Architecture	REV Model				REV+PROP Model			
	Cross-correlation		Signal Coherence		Cross-correlation		Signal Coherence	
	Training Set	Test Set	Training Set	Test Set	Training Set	Test Set	Training Set	Test Set
AE	$1.000^{+0.000}_{-0.000}$	$0.987^{+0.010}_{-0.055}$	$1.000^{+0.000}_{-0.000}$	$0.971^{+0.028}_{-0.287}$	$1.000^{+0.000}_{-0.000}$	$0.979^{+0.016}_{-0.077}$	$1.000^{+0.000}_{-0.003}$	$0.953^{+0.041}_{-0.267}$
VAE	$0.992^{+0.006}_{-0.014}$	$0.983^{+0.013}_{-0.046}$	$0.988^{+0.011}_{-0.056}$	$0.959^{+0.037}_{-0.195}$	$0.990^{+0.007}_{-0.019}$	$0.962^{+0.030}_{-0.081}$	$0.980^{+0.018}_{-0.090}$	$0.905^{+0.083}_{-0.339}$

different M, i, and Γ). Results and discussion are presented in Appendix B.

6.2. Model Deployment on Real Data

In this part, we used 13 light-curve observations from IRAS 13224–3809, with SNR greater than 50, to predict their respective response functions using the REV+PROP+NG+NP model. However, we emphasize here that the data SNR presented in this work was obtained from the definition in Equation (1), in which this is different from, and cannot be compared directly to, the value obtained from the typical calculation (see Table 1 of N. Nakhonthong et al. 2024). Indeed, while the value from the typical calculation is dependent on the numbers of source and background photons, the SNR determined in this work is, in principle, proportional to the total power of variability of the data, compared to the total power of frequency-independent, random (i.e., white) noise, regardless of the origin of variability whether it is from source or background. However, as the light curves used here are background subtracted, we argue that the detected variability is dominated by the source's photons. In fact, determining the SNR in this way is appropriate in this context, as this could ensure that the data sets used have a sufficient level of variability for our analysis.

Note that, prior to prediction, we use PCHIP to replace the missing values or gaps in the IRAS 13224–3809 data with an estimate. We compare the predicted result with the category of response functions of a lamppost corona situated at the height between $2.3r_g$ and $20.0r_g$. We investigated the likelihood between each predicted response and prescribed response using CC similarity. Thereafter, the expected scale height and its standard deviation shall be computed with the 80% rule adopted from F. Pozo Nuñez et al. (2023b), as shown in Figure 6. In other words, only coronal heights, yielding CC $\gtrsim$0.8 of maximum CC to each predicted response, are considered. The factor of 0.8 has been conventionally used in time delay measurement, as mentioned in B. M. Peterson et al. (2004). This criterion is adopted to this work to bestow the prediction uncertainty of the coronal height. Note that thresholding at 0.9 or 0.95 would leave only a few values in computation, leading to small uncertainty. On the other hand, a smaller threshold at, e.g., 0.7 would lead to larger uncertainty, with a slight change of expected coronal scale height. The monotonic relationship between X-ray luminosity and coronal height obtained here is comparable with, e.g., W. N. Alston et al. (2020), as shown in Figure 7, with the Spearman correlation coefficient of 0.68 (p-value: 0.011).

Figure 8 (left panel) shows, as an example, the predicted response function for IRAS 13224–3809 (Obs. ID 0792180501). We illustrated the predicted response with a blue solid line, while its best-matched ground-truth response corresponding to the case of $h = 9r_g$ is illustrated with the blue dotted lines. Other ground-truth responses that can lead to CCs

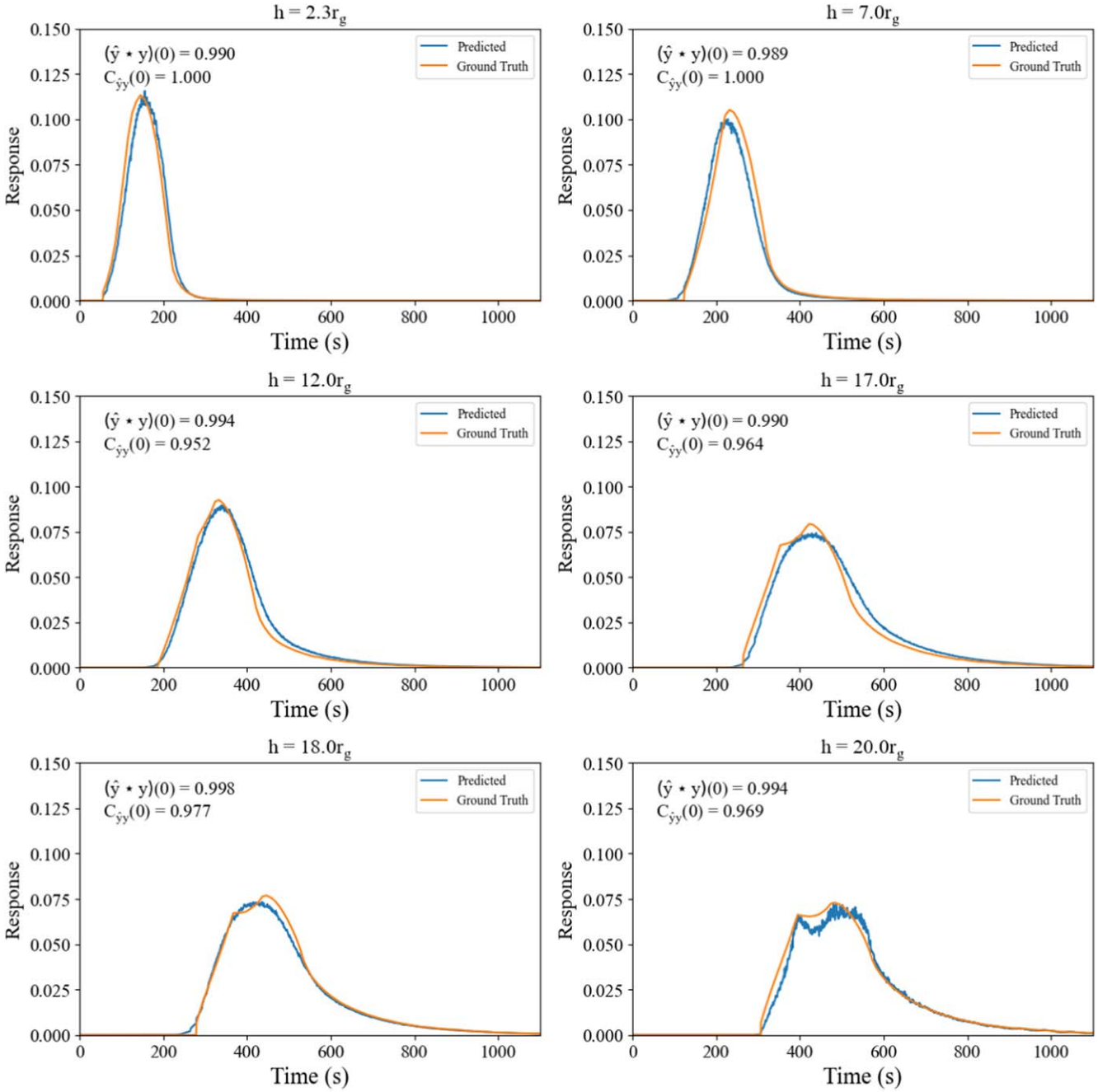

Figure 4. Example of the X-ray reverberation response functions predicted by AE (blue lines) compared with actual KYNXILREV response functions (orange lines) for the REV+PROP model.

more than 0.8 are also presented with the dotted lines in different colors, with the lowest and highest coronal height of $7r_g$ and and $11r_g$.

Figure 8 (right panel) shows the comparison between the height estimates for each observation in this work and those obtained in W. N. Alston et al. (2020). We also compare the cases when the standard deviation is computed using the threshold levels of 0.8 and 0.6. The constrained heights are loosely comparable within the estimated errors. While there is no clear trend compared to the one-to-one line, the individual mean values show some discrepancies, which might be because different assumptions are used in the models. For example, we fix Γ, while in W. N. Alston et al. (2020) it is derived from the 0.3–10 keV energy spectrum model fits and is varied among different observations. Here our approach only employs the soft-band (0.3–1 keV) light curve to derive the coronal height.

7. Discussion and Conclusion

Analyses of X-ray variability in AGN have typically been limited to Fourier space (e.g., by measuring the PSD and the lag spectrum). The X-ray reverberation time delays are usually measured from the phases of the cross-spectrum (e.g., A. C. Fabian et al. 2009; A. Zoghbi et al. 2010; D. R. Wilkins & A. C. Fabian 2013; E. M. Cackett et al. 2014; D. Emmanoulopoulos et al. 2014; P. Chainakun & A. J. Young 2015;

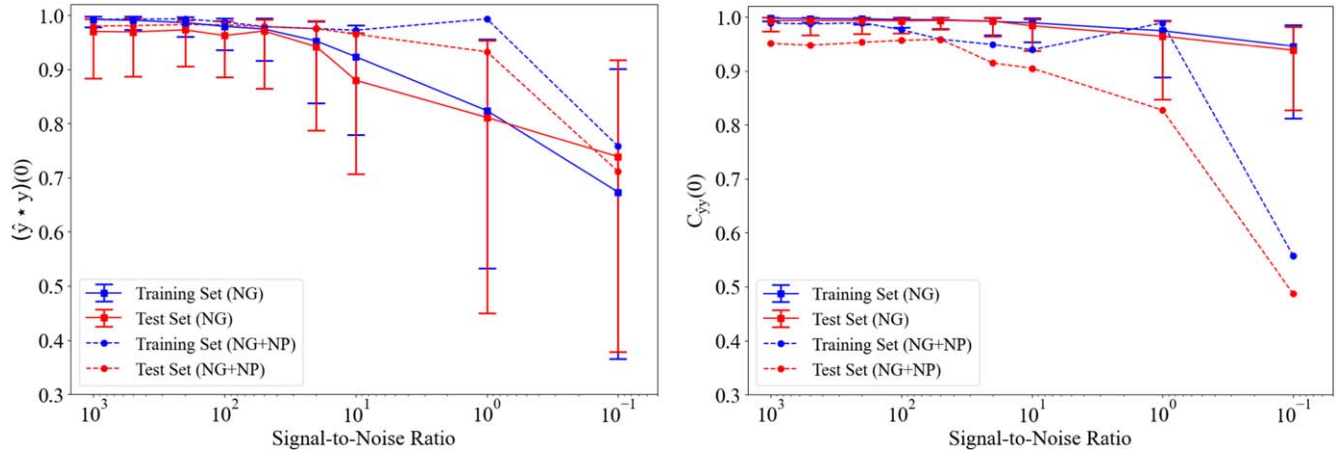

Figure 5. The evolution of response function likelihood with the SNR, using the CC (left panel) and signal coherence (right panel) as similarity indicators. Solid lines with squares are the likelihood for the REV+PROP+NG model, while dashed lines with circles are for the REV+PROP+NG+NP model; red and blue represent the training and test set, respectively. When SNR decreases, the CC seems to drop more rapidly than the coherence. The error bars correspond to the interquartile range between the first and third quartiles, in order to interpret how well the overall response functions are extracted.

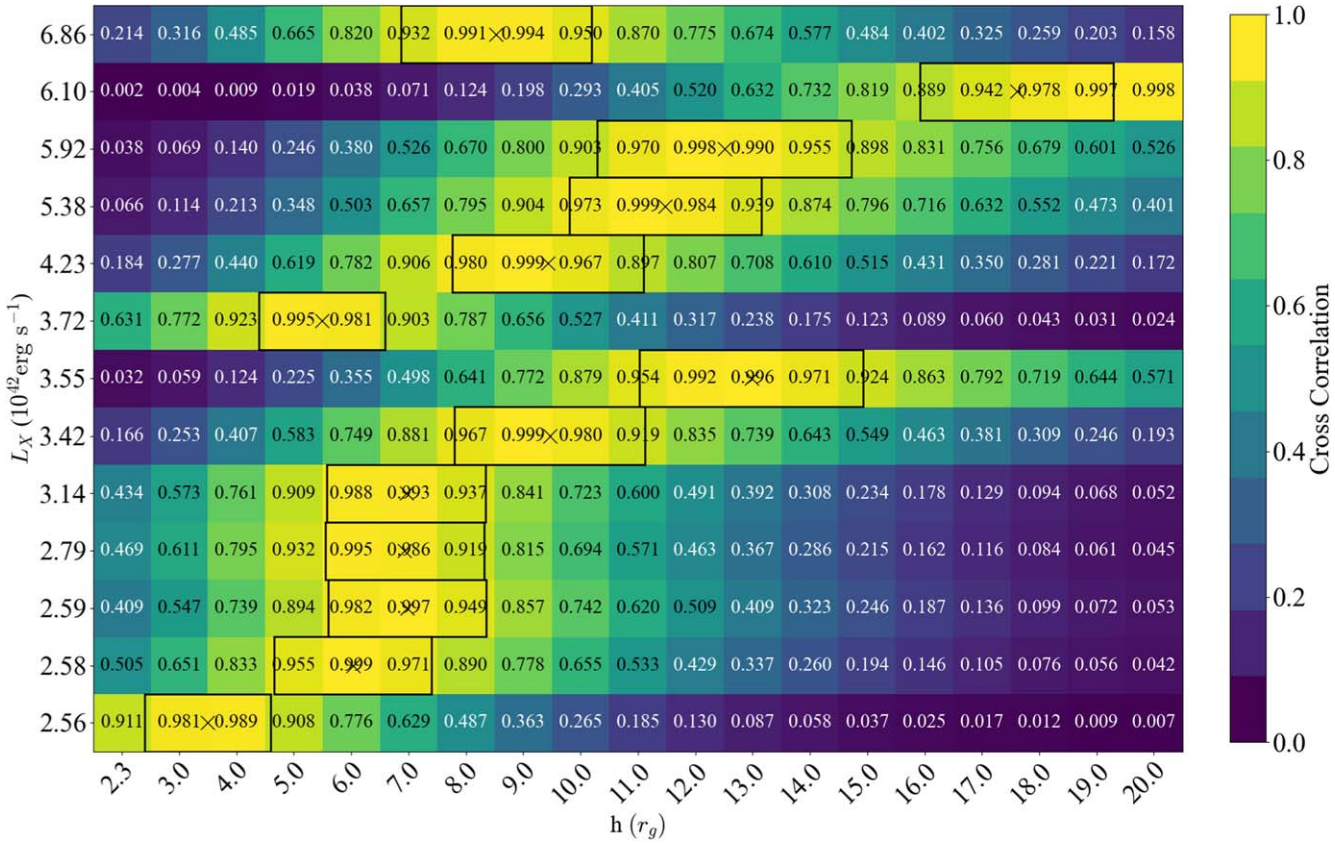

Figure 6. Matching score, by zero-lag CC, of the predicted response function using the REV+PROP+NG+NP model with the response function generated by KYNXILREV. The colors used in this plot represent the scoring level with the lowest value of 0 (dark blue) and the highest value of 1 (yellow). Cross marks and boxes in the illustration imply, respectively, the expected coronal height and standard deviation for each response function computed by the 80% rule.

A. Epitropakis et al. 2016), and no studies have been able to extract the associated response function directly from an AGN light curve. This work aims to tackle the problem in the time domain by developing the VAE model, which enables us to directly derive the response functions from the light curve. Subsequently, this technique offers an independent tool for mapping the disk–corona geometry straightforwardly from the estimated shape of the reverberation response function.

VAE is the neural network architecture sharing a similar mathematical basis to the MLP (I. Goodfellow et al. 2016; A. Paszke et al. 2019; M. J. Kang 2021), which is supervised learning. The network learns the transformation between light curves and response functions via numerical optimization. Using the standard autoencoder as the baseline, we find that VAE performed well in predicting the response functions using noise-free light curves generated by REV and REV+PROP

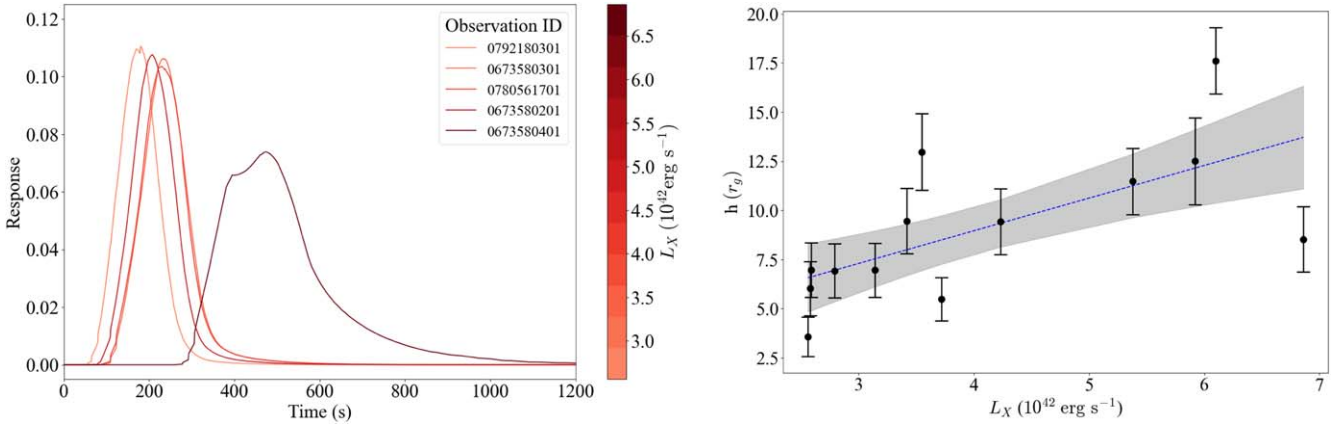

Figure 7. Left: examples of the reconstructed response functions of some IRAS 13224–3809 observations using the REV+PROP+NG+NP model. The colors of the profiles show the source luminosity measured in the 2–10 keV band (L_X). Right: predicted coronal heights (h) plotted against L_X in 13 observations of IRAS 13224–3809. The blue dotted line indicates the linear regression prediction of the coronal height with respect to X-ray luminosity, while the 3σ confidence interval is illustrated by the gray region.

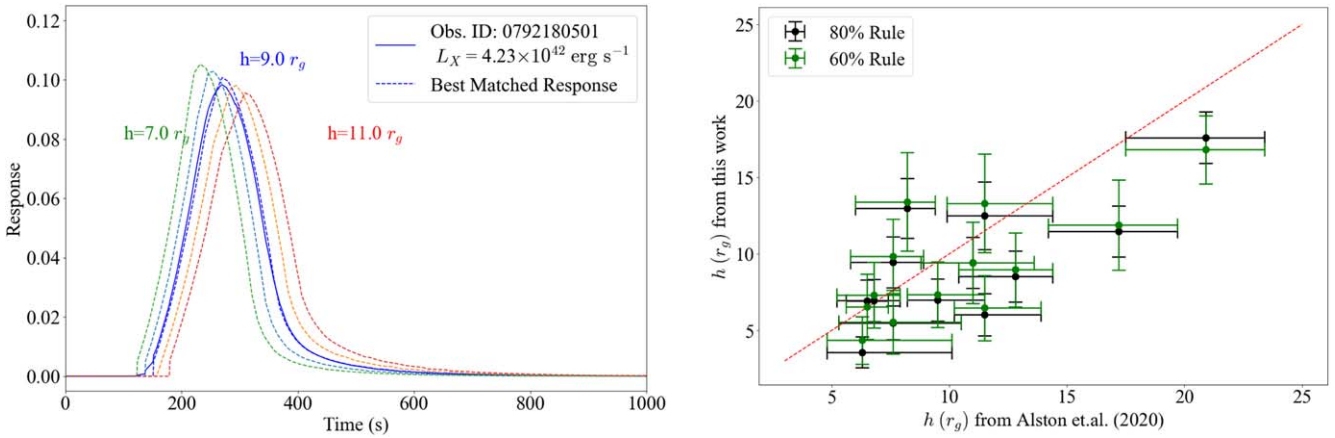

Figure 8. Left: predicted response functions of IRAS 13224–3809 (Obs. ID 0792180501; blue solid line) and some KYNXILREV response functions corresponding to different coronal heights that also show CCs >0.8 (dotted lines). The best-matched response in this case is $h = 9r_g$ (blue dotted line). Right: prediction of coronal height by our model compared with what was recovered by W. N. Alston et al. (2020). The one-to-one line is marked as a red dashed line. Prediction uncertainties by the 80% rule (black crosshairs) and 60% rule (green crosshairs) are illustrated. We note that our approach uses only a single energy band to estimate h, with Γ fixed at 2, which may limit the ability to fully capture the source's spectral evolution. Nevertheless, considering the significant uncertainties in the constrained h, our results remain broadly consistent with those of W. N. Alston et al. (2020).

models. We also examine the limitation of VAE by adding uncorrelated Gaussian noise to generated light curves. The SNR value is determined by the power spectral fraction between the signal and additional noise. The results show that the prediction accuracy declines gradually when SNR < 50. Therefore, we deploy VAE on IRAS 13224–3809 light curves with SNR $\gtrsim 50$. Data are preprocessed by binning and then interpolation to approximate the photon count in the gap area, as well as normalization, before the prediction of response functions. The best-matched response functions are determined by comparing the predicted response with KYNXILREV-simulated response functions in the archive. Only candidates with zero-lag CC $\gtrsim$0.8 are used to compute the expected coronal height and its standard deviation. The criteria are adopted from the 80% rule mentioned in F. Pozo Nuñez et al. (2023b).

With the developed VAE model, we can obtain an estimate of the shape of the reverberation response function of IRAS 13224–3809. Our findings demonstrate a dynamic corona changing its height with luminosity, from $\sim 3r_g$ to $20r_g$, which aligns with previous studies. W. N. Alston et al. (2020) discovered, through an analysis of the lag-frequency spectra, that the lamppost coronal height increases from $\sim 6r_g$ to $20r_g$ with source luminosity. By assuming $a = 0.99$ and analyzing combined spectral-timing data in various flux states, M. D. Caballero-García et al. (2020) found a trend for the source height rising from $\sim 3r_g$ to $10^{+10}_{-1}r_g$ with luminosity. Furthermore, due to the gravitational light-bending effects, the reflection component in IRAS 13224–3809 is less variable compared to the primary X-ray continuum (G. Miniutti & A. C. Fabian 2004; C.-Y. Chiang et al. 2015). The reverberation effects then act as a filter that suppresses the fractional excess variance of intrinsic variability, producing the reverberation dip on the PSD profiles, whose amplitude is stronger in the energy band that is more dominated by reflection (e.g., D. Emmanoulopoulos et al. 2016; I. Papadakis et al. 2016; P. Chainakun 2019). The coronal height correlated with luminosity was also suggested by the studies of these reverberation dips on the PSD data of IRAS 13224–3809 (P. Chainakun et al. 2022; N. Mankatwit et al. 2023). This work then provides further evidence supporting such a scenario.

While we use KYNXILREV, a part of the KYN package (M. Dovčiak et al. 2004a, 2004b; M. D. Caballero-García et al. 2018), which employs the XILLVER reflection model (J. García & T. R. Kallman 2010; J. García et al. 2013) to generate the disk response functions, it is also possible to use KYNREFREV, which relies on the REFLIONX reflection model (R. R. Ross et al. 1999; R. R. Ross & A. C. Fabian 2005). Different choices of these models may lead to different dilution effects on the amplitude of the lags, resulting in a bias in determining the coronal height (J. García et al. 2013; K. Khanthasombat et al. 2024). The alternative model for the reflection from a high-density disk is also possible for IRAS 13224–3809 (J. Jiang et al. 2022). We predict the response function using only soft X-ray light curves for both training and deployment, whereas W. N. Alston et al. (2020) incorporate both soft and hard X-rays for time-lag analysis. Additionally, there is evidence that Γ might be variable and flux dependent in IRAS 13224–3809 (e.g., W. N. Alston et al. 2020; M. D. Caballero-García et al. 2020). To maintain simplicity, we initially set $\Gamma = 2$ and employ free parameter R to compensate for the dilution effects possibly induced by variations in those parameters that are held constant. Although this simplified analysis process may overlook certain complex spectral behaviors, the h–L_X correlation trend can be seen, and our results are not drastically different from what was obtained by, for example, W. N. Alston et al. (2020). The uncertainties in coronal height measurements across different methods emphasize the need for more advanced measurement techniques, cross-validation between methods, and higher-quality observational data to better constrain the models. At the moment, the model tries to fit the overall shape of the response rather than locally concentrating on the feature surrounding the peak. Nevertheless, judging by the overall shape of the reconstructed response functions, h should still be accurately predicted by the model as long as SNR $\gtrsim$ 50. Predicting various parameters at once may require the model to fully capture the complex details around the response peak, which is subject to future study.

Nevertheless, we expect the model to work well even if it is designed to predict the TH-shaped response functions instead. In this case, it could be used to test the Comptonization models or disk propagation models. Each of these delays should be characterized by an energy-dependent response function. V. K. Jaiswal et al. (2023) investigated the time delays in AGN caused by both accretion disk and broad-line region (BLR) reprocessing. They concluded that the effects of BLR scattering behave similarly to those resulting from an increase in the height of the X-ray source above the disk. This introduces degeneracy that complicates the interpretation of time lags and may also lead to overestimated coronal heights. However, F. Pozo Nuñez et al. (2023a) have shown that the primary source of contamination contributing to the increased accretion disk time delay is predominantly related to diffuse continuum emission from the BLR (D. Chelouche et al. 2019), with little or negligible dependence on the height of the X-ray source. Therefore, the presence of this degeneracy remains a matter of debate. The lags in IRAS 13224–3809 may be explained by a vertically extended corona simplified using two lamppost sources (S. Hancock et al. 2023), and the corona itself may also vary within each individual observation (P. Chainakun et al. 2023; D. R. Wilkins 2023; N. Nakhonthong et al. 2024). In these aspects, the model can be improved by being trained to reproduce the combined realistic response functions from different reprocessors or different geometry assumptions. It is also worth training the model using arbitrary shapes of the response functions in order to test how the reconstructed responses deviate from what is expected from the lamppost reverberation.

While our focused information is limited to the coronal height, this work shows a potential way to extract the shape of the disk response function directly from the light curves. It is applicable to estimate the shape of the reverberation response in AGN within the scope of the lamppost assumption. Particularly, we note that the complex features around the response peak still cannot be well resolved with the current version of the model. This can be improved, in the future, in several ways, such as by providing more synthetic data for training, but this undoubtedly requires a large amount of computational time. Reconstructing the reverberation response functions simultaneously with the responses due to other processes is vital for testing the models and concluding a unique interpretation of the source geometry. Implementing another type of neural network, such as convolutional neural network and recurrent neural network, is another possibility to improve the prediction capability, flexibility, and sensitivity. A more comprehensive model that incorporates the warm corona is necessary for a full understanding of different emission regions in the accretion flow and their contributions to the observed spectral and timing properties of AGN.

Acknowledgments

This work was supported by (i) Suranaree University of Technology (SUT), (ii) Thailand Science Research and Innovation (TSRI), and (iii) National Science, Research and Innovation Fund (NSRF), grant No. 204265. J.J. acknowledges support from Leverhulme Trust, Isaac Newton Trust, and St Edmund's College, University of Cambridge. F.P.N. gratefully acknowledges the generous and invaluable support of the Klaus Tschira Foundation. F.P.N. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 951549).

Appendix A
Signal-to-noise Ratio

In this work, SNR is defined by the fraction between signal and noise power spectra, assuming noise involvement in all frequency domains, as mentioned in Equation (1). We also note that while the SNR obtained by the traditional method and that obtained by the method in this paper may not directly correlate, somehow they still have the correlation at some level. Indeed, the observations that have high SNR measured by the traditional method (e.g., Table 1 of N. Nakhonthong et al. 2024) are likely to have a relatively high SNR when it is calculated by our method (e.g., Obs. ID 0780561601, 0792180401, and 0792180601). Thus, we demonstrate how the behavior of light curves in the simulated, training REV +PROP+NG+NP data varies with SNR, as shown in Figure 9. For comparison, we present the light curve of IRAS 13224–3809 (Obs. ID 067358401) showing SNR = 62.23, which is almost similar in quality to our example light curve with a similar SNR.

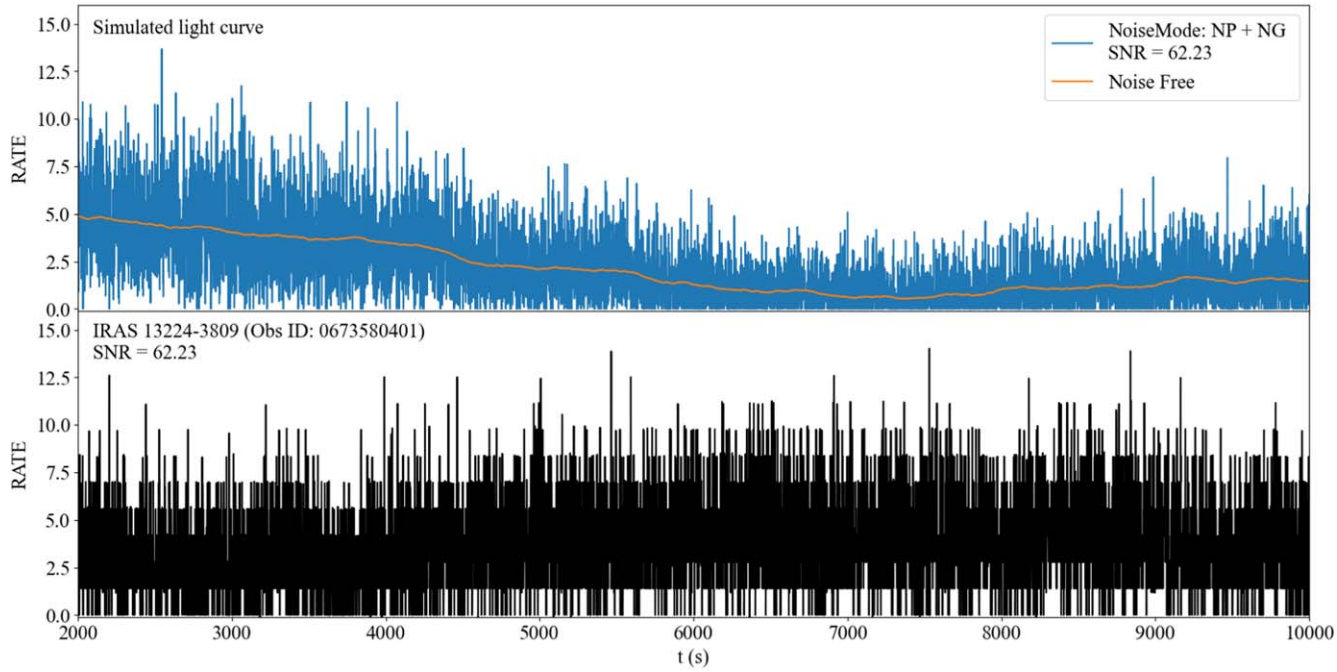

Figure 9. Top panel: simulated light curves. Noise-free and noisy light curves are labeled as orange and blue lines, respectively. Bottom panel: example of IRAS 13224–3809 light curve showing SNR = 62.23.

Appendix B
Investigation of Parameter Variation Effect on Prediction Efficiency

Furthering our performance investigation, we test for the general applicability of the model when the observed data have slightly different characteristics to the training data. First, we investigate the case when the reflection fraction R is outside the range used to train the model. Note that, in this investigation, the model is trained with $R \in [1, 3]$, while the R range is extended to [1, 10] in the final model applied to IRAS 13224–3809. We also produce some new response functions with different Γ, i, and M within the observed uncertainty reported in W. N. Alston et al. (2020).[13] We generate sets of convolved light curves corresponding to four scenarios: (1) only R is varied ∼25% beyond the trained values, (2) only Γ is varied, (3) Γ and i are varied, and (4) Γ, i, and M are varied. We test the original REV+PROP model (trained using the light curves with the fixed Γ, i, and M) on these light curves. The results are reported in Table 5. We find that the model can capture the essential aspects of the coronal height while accommodating potential variations in the reflection fraction. This is likely because, although the overall shape and amplitude of the response function vary, the first response time of the reverberation delays—which is closely related to the coronal height—remains unaffected. The model appears to use this initial response signature as a key indicator for determining the coronal height. While the response functions extracted by the model are still generally excellent in terms of signal coherence, we can see a decrease in the CC. This is expected since the training data have slightly different characteristics from the observed data, even though this should be the real scenario, where we do not exactly know the precise values of these parameters. Regardless, this R range is likely adequate for applying the model to the specific AGN IRAS 13224–3809, with CC $\gtrsim$0.8 and signal coherence $\gtrsim$0.9. Finally, it could be expected that the performance shall decline when Gaussian noise and Poisson noise are considered, in a similar trend to that shown in Figure 5.

[13] Varying parameters are $2 \leqslant \Gamma \leqslant 2.68$, $62° \leqslant i \leqslant 72°$, and $1.7 \leqslant M(10^6\ M_\odot) \leqslant 2.1$.

Table 5
Prediction Performance of the Original REV+PROP Model (Trained with the Light Curves Where Γ, i, and M Are Fixed) on the Test Data Produced in Four Scenarios

Scenario	Cross-correlation	Signal Coherence
Vary R outside the trained range	$0.939^{+0.047}_{-0.119}$	$0.964^{+0.016}_{-0.088}$
Vary Γ	$0.875^{+0.055}_{-0.095}$	$0.959^{+0.005}_{-0.078}$
Vary Γ and i	$0.861^{+0.080}_{-0.156}$	$0.959^{+0.009}_{-0.025}$
Vary Γ, i, and M	$0.848^{+0.116}_{-0.110}$	$0.959^{+0.007}_{-0.025}$

ORCID iDs

Sanhanat Deesamutara https://orcid.org/0000-0002-0964-0050
Poemwai Chainakun https://orcid.org/0000-0002-9099-4613
Tirawut Worrakitpoonpon https://orcid.org/0000-0002-0384-305X
Wasutep Luangtip https://orcid.org/0000-0002-4516-6042
Jiachen Jiang https://orcid.org/0000-0002-9639-4352
Francisco Pozo Nuñez https://orcid.org/0000-0002-6716-4179
Andrew J. Young https://orcid.org/0000-0003-3626-9151

References

Alston, W. N., Done, C., & Vaughan, S. 2014, MNRAS, 439, 1548
Alston, W. N., Fabian, A. C., Kara, E., et al. 2020, NatAs, 4, 597
Arévalo, P., & Uttley, P. 2006, MNRAS, 367, 801
Ballantyne, D. R., Sudhakar, V., Fairfax, D., et al. 2024, MNRAS, 530, 1603
Caballero-García, M. D., Papadakis, I. E., Dovčiak, M., et al. 2018, MNRAS, 480, 2650

Caballero-García, M. D., Papadakis, I. E., Dovčiak, M., et al. 2020, MNRAS, 498, 3184
Cackett, E. M., Bentz, M. C., & Kara, E. 2021, iSci, 24, 102557
Cackett, E. M., Zoghbi, A., Reynolds, C., et al. 2014, MNRAS, 438, 2980
Carter, J. A., & Read, A. M. 2007, A&A, 464, 1155
Chainakun, P. 2019, ApJ, 878, 20
Chainakun, P., Luangtip, W., Jiang, J., & Young, A. J. 2022, ApJ, 934, 166
Chainakun, P., Nakhonthong, N., Luangtip, W., & Young, A. J. 2023, MNRAS, 523, 111
Chainakun, P., & Young, A. J. 2015, MNRAS, 452, 333
Chainakun, P., Young, A. J., & Kara, E. 2016, MNRAS, 460, 3076
Chelouche, D., Pozo Nuñez, F., & Kaspi, S. 2019, NatAs, 3, 251
Cheng, T.-Y., Li, N., Conselice, C. J., et al. 2020, MNRAS, 494, 3750
Chiang, C.-Y., Walton, D. J., Fabian, A. C., Wilkins, D. R., & Gallo, L. C. 2015, MNRAS, 446, 759
De Marco, B., Ponti, G., Cappi, M., et al. 2013, MNRAS, 431, 2441
Dovčiak, M., Karas, V., Martocchia, A., Matt, G., & Yaqoob, T. 2004a, in Proc. of RAGtime 4/5: Workshops on Black Holes and Neutron Stars, ed. S. Hledík & Z. Stuchlík (Opava: Silesian Univ. Opava), 33
Dovčiak, M., Karas, V., & Yaqoob, T. 2004b, ApJS, 153, 205
Emmanoulopoulos, D., Papadakis, I. E., Dovčiak, M., & McHardy, I. M. 2014, MNRAS, 439, 3931
Emmanoulopoulos, D., Papadakis, I. E., Epitropakis, A., et al. 2016, MNRAS, 461, 1642
Epitropakis, A., Papadakis, I. E., Dovčiak, M., et al. 2016, A&A, 594, A71
Fabian, A. C., Zoghbi, A., Ross, R. R., et al. 2009, Natur, 459, 540
Fritsch, F. N., & Butland, J. 1984, SIAM J. Sci. Stat. Comput., 5, 300
Frontera-Pons, J., Sureau, F., Bobin, J., & Le Floch, E. 2017, A&A, 603, A60
García, J., Dauser, T., Reynolds, C. S., et al. 2013, ApJ, 768, 146
García, J., & Kallman, T. R. 2010, ApJ, 718, 695
George, I. M., & Fabian, A. C. 1991, MNRAS, 249, 352
González-Martín, O., & Vaughan, S. 2012, A&A, 544, A80
Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (Cambridge, MA: MIT Press)
Hancock, S., Young, A. J., & Chainakun, P. 2023, MNRAS, 520, 180
Huppenkothen, D., Bachetti, M., Stevens, A. L., et al. 2019, ApJ, 881, 39
Jaiswal, V. K., Prince, R., Panda, S., & Czerny, B. 2023, A&A, 670, A147
Jansen, F., Lumb, D., Altieri, B., et al. 2001, A&A, 365, L1
Jiang, J., Dauser, T., Fabian, A. C., et al. 2022, MNRAS, 514, 1107
Jiang, J., Parker, M. L., Fabian, A. C., et al. 2018, MNRAS, 477, 3711
Kallman, T., & Bautista, M. 2001, ApJS, 133, 221
Kang, M. J. 2021, Pytorch-VAE-tutorial: A simple tutorial of Variational AutoEncoders with Pytorch, GitHub, https://github.com/Jackson-Kang/Pytorch-VAE-tutorial/tree/master
Kara, E., Alston, W. N., Fabian, A. C., et al. 2016, MNRAS, 462, 511
Khanthasombat, K., Chainakun, P., & Young, A. J. 2024, MNRAS, 528, 3130
Kotov, O., Churazov, E., & Gilfanov, M. 2001, MNRAS, 327, 799
Kubota, A., & Done, C. 2018, MNRAS, 480, 1247
Kumari, K., Dewangan, G. C., Papadakis, I. E., & Singh, K. P. 2024, MNRAS, 527, 5668
Kumari, N., Jana, A., Naik, S., & Nandi, P. 2023, MNRAS, 521, 5440
Lyubarskii, Y. E. 1997, MNRAS, 292, 679
Mankatwit, N., Chainakun, P., Luangtip, W., & Young, A. J. 2023, MNRAS, 523, 4080
McHardy, I. M., Arévalo, P., Uttley, P., et al. 2007, MNRAS, 382, 985
Middei, R., Bianchi, S., Cappi, M., et al. 2018, A&A, 615, A163
Miniutti, G., & Fabian, A. C. 2004, MNRAS, 349, 1435
Nakhonthong, N., Chainakun, P., Luangtip, W., & Young, A. J. 2024, MNRAS, 530, 1894
Orwat-Kapola, J. K., Bird, A. J., Hill, A. B., Altamirano, D., & Huppenkothen, D. 2022, MNRAS, 509, 1269
Papadakis, I., Pecháček, T., Dovčiak, M., et al. 2016, A&A, 588, A13
Paszke, A., Gross, S., Massa, F., et al. 2019, PyTorch: An Imperative Style, High-Performance Deep Learning Library, arXiv:1912.01703
Peterson, B. M., Ferrarese, L., Gilbert, K. M., et al. 2004, ApJ, 613, 682
Porquet, D., Reeves, J. N., Matt, G., et al. 2018, A&A, 609, A42
Pozo Nuñez, F., Bruckmann, C., Deesamutara, S., et al. 2023a, MNRAS, 522, 2002
Pozo Nuñez, F., Gianniotis, N., & Polsterer, K. L. 2023b, A&A, 674, A83
Reynolds, C. S., & Nowak, M. A. 2003, PhR, 377, 389
Reynolds, C. S., Young, A. J., Begelman, M. C., & Fabian, A. C. 1999, ApJ, 514, 164
Ross, R. R., & Fabian, A. C. 2005, MNRAS, 358, 211
Ross, R. R., Fabian, A. C., & Young, A. J. 1999, MNRAS, 306, 461
Sacchi, M. D. 2016, Adding Noise with a Desired Signal-to-noise Ratio, Signal Analysis and Imaging Group, University of Alberta, https://sites.ualberta.ca/%7Emsacchi/SNR_Def.pdf
Ursini, F., Petrucci, P. O., Bianchi, S., et al. 2020, A&A, 634, A92
Uttley, P., Cackett, E. M., Fabian, A. C., Kara, E., & Wilkins, D. R. 2014, A&ARv, 22, 72
Vaughan, S., Edelson, R., Warwick, R. S., & Uttley, P. 2003, MNRAS, 345, 1271
Véron-Cetty, M. P., & Véron, P. 2006, A&A, 455, 773
Wilkins, D. R. 2023, MNRAS, 526, 3441
Wilkins, D. R., & Fabian, A. C. 2013, MNRAS, 430, 247
Wilkins, D. R., & Gallo, L. C. 2015, MNRAS, 449, 129
Xu, X., Ding, N., Gu, Q., Guo, X., & Contini, E. 2021, MNRAS, 507, 3572
Young, A. J., & Reynolds, C. S. 2000, ApJ, 529, 101
Zoghbi, A., Fabian, A. C., Uttley, P., et al. 2010, MNRAS, 401, 2419