SYMBOLIC REGRESSION MODEL FOR THE PARAMETER RELATIONS IN THE X-RAY REVERBERATING AGN

PRATHAK THONGKONSING

A Thesis Submitted in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Physics

Suranaree University of Technology

Academic Year 2023

แบบจำลองการวิเคราะห์ถดถอยเชิงสัญลักษณ์เพื่อหาความสัมพันธ์ระหว่างตัวแปรของนิวเคลียสดาราจักรกัมมันต์ที่สะท้อนรังสีเอกซ์

นายประทักษ์ ทองก้อนสิงห์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาฟิสิกส์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2566

SYMBOLIC REGRESSION MODEL FOR THE PARAMETER RELATIONS IN THE X-RAY REVERBERATING AGN

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for a Master's Degree.

Thesis Examining Committee

Wasutep Luangtip

(Asst. Prof. Dr. Wasutep Luangtip)

Chairperson

Poemwai Chainakun.

(Asst. Prof. Dr. Poemwai Chainakun)

Member (Thesis advisor)

Tirawut Worrakitpoonpon

(Asst. Prof. Dr. Tirawut Worrakitpoonpon)

Member (Thesis Co-advisor)

Panu Yimmuang

(Asst. Prof. Dr. Panu Yimmuang)

Member

Ittipon.

(Dr. Ittipol Fongkaew)

Member

Yupaporn Clst

(Assoc. Prof. Dr. Yupaporn Ruksakulpiwat)

Vice Rector for Academic Affairs

and Quality Assurance

Situl

(Prof. Dr. Santi Maensiri)

Dean of Institute of Science

ประทักษ์   ทองก้อนสิงห์ :   แบบจำลองการวิเคราะห์ถดถอยเชิงสัญลักษณ์เพื่อหาความสัมพันธ์ระหว่างตัวแปรของนิวเคลียสดาราจักรกัมมันต์ที่สะท้อนรังสีเอกซ์   (SYMBOLIC REGRESSION MODEL FOR THE PARAMETER RELATIONS IN THE X-RAY REVERBERATING AGN) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.เพิ่มวัย   ชัยนะกุล, 40 หน้า

คำสำคัญ: การวิเคราะห์ถดถอยเชิงสัญลักษณ์, นิวเคลียสดาราจักรกัมมันต์, การวิเคราะห์ข้อมูล

นิวเคลียสดาราจักรกัมมันต์ (Active Galactic Nucleus, AGN) คือบริเวณใจกลางของดาราจักรซึ่งถูกเชื่อว่ามีหลุมดำมวลยิ่งยวดอยู่ ณ ใจกลาง การศึกษาวิวัฒนาการของระบบ AGN และปรากฏการณ์ทางฟิสิกส์ที่เกี่ยวข้อง   สามารถทำได้ผ่านการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรของ AGN งานวิทยานิพนธ์นี้จะเน้นศึกษา AGN ที่มีการสะท้อนรังสีเอกซ์จำนวน 20 ระบบจากฐานข้อมูลของกล้องโทรทัศน์อวกาศ XMM-Newton

การวิเคราะห์ถดถอยเชิงสัญลักษณ์ (Symbolic Regression, SR) เป็นการเรียนรู้ของเครื่องแบบไร้การกำกับดูแลที่ใช้ในการค้นหาความสัมพันธ์ระหว่างตัวแปรของ AGN ในกลุ่มตัวอย่างโดยไม่มีความลำเอียงของสมการที่จะค้นหา งานวิจัยในอดีตค้นพบว่ามีความสัมพันธ์ระหว่างค่าหน่วงเวลา (Time Lag) และมวลของหลุมดำ ซึ่งจะเป็นสิ่งที่ศึกษาในงานนี้เช่นกัน ผลการศึกษาพบว่า SR เสนอความสัมพันธ์ของค่าหน่วงเวลาและมวลในรูปสมการพหุนาม (Polynomial Equation) ซึ่งอธิบายว่าตำแหน่งของโคโรน่า (Corona) ในกลุ่มตัวอย่าง AGN ของงานนี้มีการกระจายที่หลากหลาย นอกจากนี้ได้ทำการศึกษาความสัมพันธ์ระหว่างมวลของดวงดาวในดาราจักรนั้น ๆ และหลุมดำ ณ ใจกลางดาราจักรโดยค้นพบว่า SR แนะนำความสัมพันธ์ในรูปแบบไม่ใช่เชิงเส้น (Non-linear) ระหว่างมวลของดวงดาวและหลุมดำ   ซึ่งสามารถเขียนได้ในรูปสมการเส้นตรงในมาตราส่วนลอการิทึมที่มีค่าความชันเป็น 0.40 ผลการศึกษาการใช้ตัวแปรอื่นประกอบด้วยค่าอัตราส่วนการสะท้อน (Reflection Fraction) ค่าดัชนีโฟตอน (Photon Index) และค่าอัตราส่วนของเอ็ดดิงตัน (Eddington Ratio) ในการวิเคราะห์ระบบ AGN ทว่าตัวแปรเหล่านี้ไม่ได้ช่วยเพิ่มให้ประสิทธิภาพของสมการที่ SR ค้นพบนั้นดีขึ้นในแง่ความแม่นยำของการทำนายผลลัพธ์

สาขาวิชาฟิสิกส์   ลายมือชื่อนักศึกษา_____ประทักษ์_____
ปีการศึกษา 2566   ลายมือชื่ออาจารย์ที่ปรึกษา_____
                                                       ลายมือชื่ออาจารย์ที่ปรึกษาร่วม_Tiraawat Woramethrat_

PRATHAK THONGKONSING : SYMBOLIC REGRESSION MODEL FOR THE PARAMETER RELATIONS IN THE X-RAY REVERBERATING AGN. THESIS ADVISOR : ASSOC. PROF. POEMWAI CHAINAKUN, PH.D. 40 PP

Keyword: symbolic regression, active galactic nuclei, data analysis

An active galactic nucleus (AGN) is the region at the center of a galaxy, supposedly hosting a supermassive black hole (SMBH). The evolution of the AGN system and the underlying physical phenomena around them can be understood by analyzing the relationship between the AGN parameters. This work focuses on 20 samples of AGN with X-ray reverberation features taken from the observational data in the XMM-Newton archives. Previously, a relationship between the time lag and the black hole mass has been suggested, which is also investigated in this work. Using symbolic regression (SR), an unsupervised machine learning algorithm for relating AGN parameters in the unbiased form. Our result shows that the SR prefers a relation between the lags and the masses in a polynomial form, indicating that the corona distance of the AGN in our samples varies significantly. The relationship between stellar mass and black hole mass is also investigated. Here, the SR prefers a non-linear relation in the form of a linear equation in logarithmic scale with slope equal to 0.40. Other parameters including reflection fraction, photon index, and Eddington ratio are also analyzed, yet they do not provide significant improvement of the SR.

School of Physics	Student's Signature __ประทักษ์__
Academic Year 2023	Advisor's Signature __เพิ่มไว ชัยนกูล__
	Co-advisor's Signature _Tiranat Worrakit_

ACKNOWLEDGEMENTS

CONTENTS

CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF ABBREVIATIONS

AGN   Active galactic nuclei

BH   Black hole

SMBH   Supermassive black hole

SR   Symbolic regression

RF   Reflection fraction

LPM   Lamp-post model

GP   Genetic programming

MSE   Mean square error
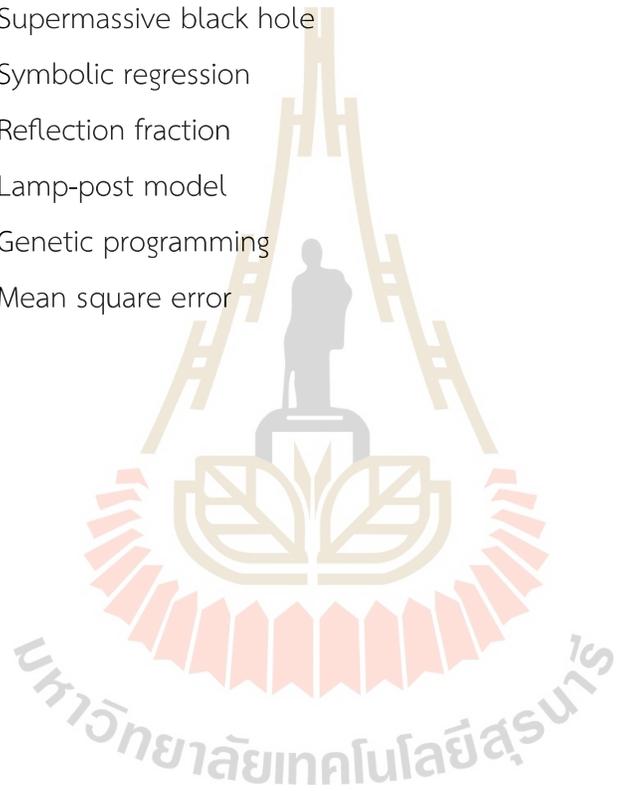
CHAPTER I

INTRODUCTION

Active Galactic Nucleus (AGN) is a region around the center of the galaxy that hosts supermassive black hole (SMBH) which produces powerful X-ray radiation due to an accretion of gas onto the central black hole. The SMBH has the mass of $10^5 - 10^{10} \ M_\odot$ (Bentz and Katz, 2015). The AGN is composed of different components, but the main part that contributes to the observable X-rays is the corona region above the accretion disc of the accreting SMBH.

The AGN also possesses a dusty torus on the outer region which can obscure the central SMBH. In principle, different types of the AGN can be classified by different viewing angles of the observer. For example, the Seyfert type 1 is the AGN that has a small inclination angle as presented in Figure 1. Its center has not been obscured by the dusty torus, making it a good candidate for observing the activity around the innermost region closest to the event horizon of the black hole.

The accreting SMBH also released high-energetic plasma known as the relativistic jet dominated in the radio waveband. Based on the orientation of the jet, the AGN can be divided into a radio-loud and radio-quiet AGN. The radio-loud AGN produces a jet of plasma that moves along the line of sight of an observer, causing a very high observed radio luminosity from relativistic beaming effect. If emission from the jet dominates the emission from other wavebands, it is classified as the blazar. On the other hand, the radio-quiet AGN does not have strong jets pointing towards the observer.
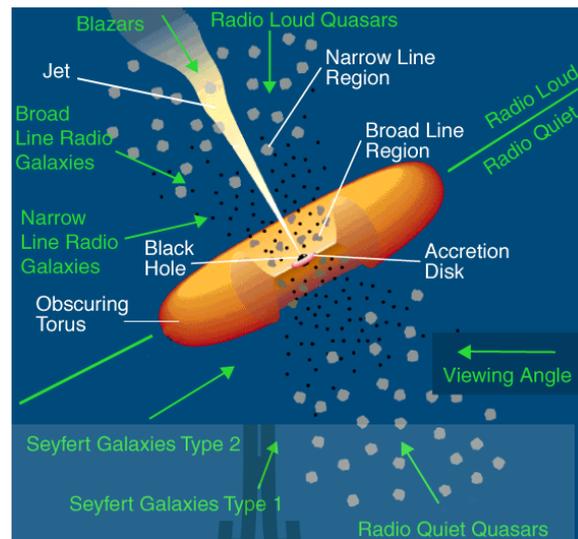
Figure 1 The classification of different AGN types based on the viewing angle.
Figure adapted from Urry and Padovani, 1995.

In high-energy astrophysics, the active area of research is commonly focus on understanding the interaction between the accretion disc, the corona, and the central black hole in order to understand the evolution and physics of different processes that occurred around the AGN. In this work, the study of the global relations between the observed parameters of the AGN has been conducted. The symbolic regression will be used to derive the mathematical expressions of these parameters. The test for robustness for the obtained relations also has been performed.

The scope of this work is as follows: The AGN samples is mainly focus on the Seyfert I galaxy since it has not been obscured by the dust, hence it allows probing of the physics of inner accretion flows near the SMBH. Since the innermost region produced X-rays, the samples are based on the observed X-ray data from the XMM-Newton space-based telescope. The samples are focused on the X-ray reverberating AGN (i.e., the AGN that shows the time lags between the direct and reflection X-rays). This work will provide a global look on these AGN and provide more insights towards the relationship between AGN parameter and their underlying physics.

CHAPTER II

THEORY AND LITERATURE REVIEW

2.1 Spectral analysis

The components around AGN can be studied by analyzing the time-average spectrum, a profile of the intensity of photons in each energy band, obtained from the spectroscopy technique. The total AGN spectrum is commonly modeled using different combinations of spectral features being contributed from different parts of the AGN as seen in Figure 2.
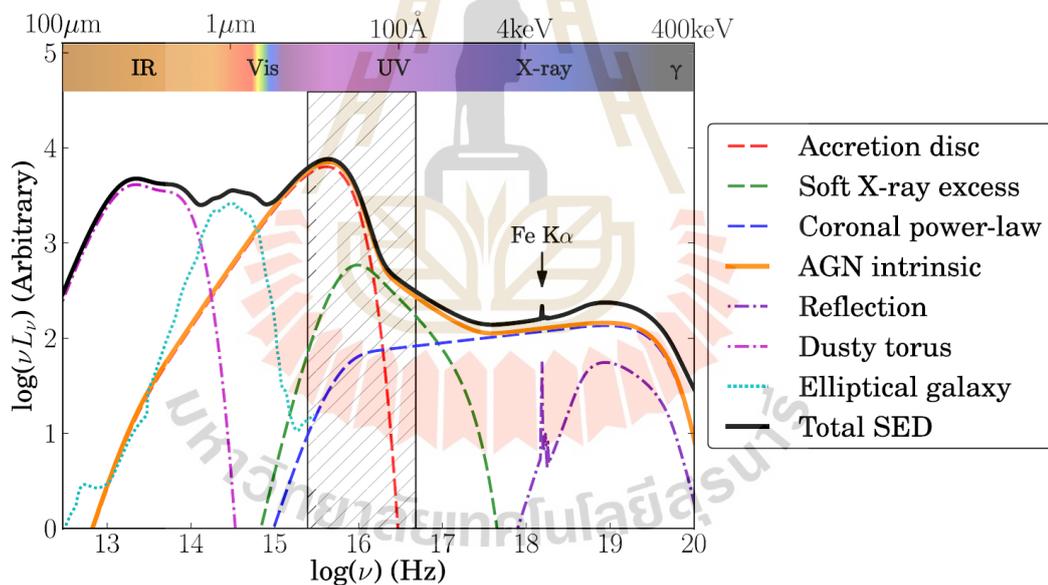
Figure 2 The spectral energy distribution (SED) of the AGN.

Figure from Collinson et al., 2016.

The accretion disc spectrum is modeled with multiple blackbody radiation with different temperatures which differ for each radial position of the accretion disc. The disc spectrum is dominant in the range of optical to UV band. However, some parts of the spectral profile have energy exceeding UV band where it cannot be physically produced by the photons from the accretion disc alone.

The X-ray component of the spectrum is commonly modeled by a power-law continuum and a reflection spectrum. The power-law part is the contribution from the photons that have been Compton up-scattered by the electron in a corona or from a base of jet (Haardt, 1993). The reflection spectrum is believed to come from the continuum power-law irradiating the surface of the accretion disc. The interaction between the photons and element in the accretion disc can be understood from the absorption and emission process. Studying the reflection component can lead to an understanding of the properties of the accretion disc such as the structure, the temperature, and the composition of accreting gas. The reflection profile can be simulated using the spectral model, such as XILLVER, which is a public model for the reflection component of the photons that are reprocessed by the accretion disc (García et al., 2013).

2.1.1 Photon index

The X-rays are thought to be produced by the inverse-Compton scattering of the disc photons (lower energy) with the relativistic electrons inside the corona (higher energy). The disc photons then gain energy from the relativistic electrons, boosting energy up to X-rays. The coronal X-ray emission has its spectrum in the form of a cut-off power law with various parameters which can be fitted to the observed spectrum. The cut off energy in the power law is linked to the energy of electrons in the corona region. Figure 3 shows examples of the reflection spectrum varying with the photon index, Γ. The value of photon index affects the slope of the power law component, as we can see that the reflection spectrum become less inclined with reducing the photon index. The prominence of the emission feature depends on the ionization parameters which involve the heating and cooling process due to an interaction between the photon and the accretion disc. The observed parameter Γ is one of the parameters that we are going to investigate.
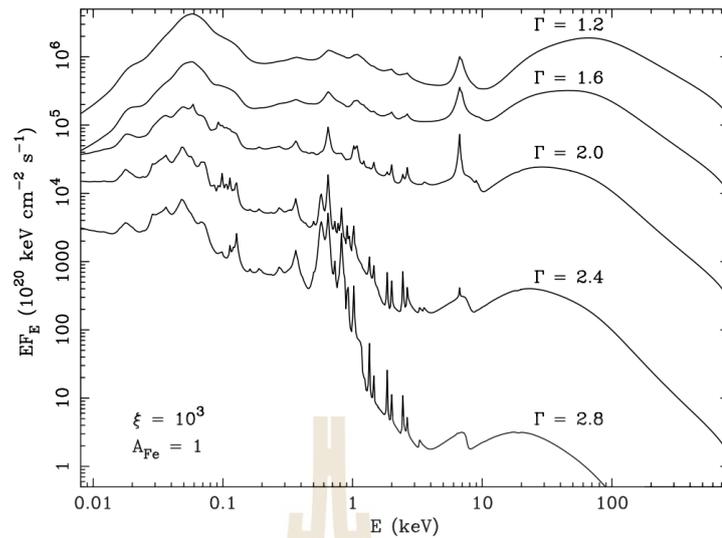
Figure 3 The reflection spectrum when varying Γ.

   Figure from Ross and Fabian, 2005.

2.1.2 Reflection fraction

   The reflection fraction ($\mathbf{RF}$) is defined as the ratio of coronal photons that hit the disc to those that reach infinity. The $\mathbf{RF}$ then can also be derived through the spectral modeling using the ray tracing simulation (Dauser et al., 2016). The RF is also one of the observed AGN parameters used in this work.

   For simplicity, the corona is modeled as a point source located at a specific height on the rotational axis of the SMBH. This corona configuration is referred to as the lamp post model (LPM) which represents the corona that resides in a compact region (George and Fabian, 1991). The simple illustration of the LPM model is shown in Figure 4. Here, we employ the LPM to derive furthermore parameters relating to the AGN geometry.
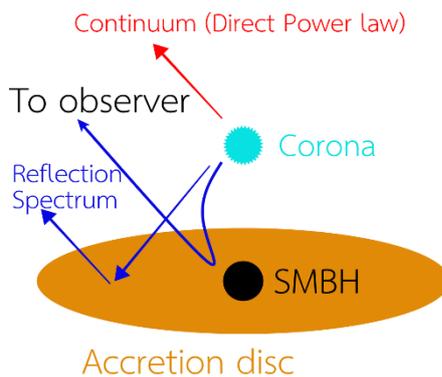
Figure 4 The diagram of the LPM which models the corona as a point source.
Figure adapt from Emmanoulopoulos, Papadakis, Dovčiak, and McHardy,
2014.

2.1.3 Eddington limit

The Eddington limit is defined by L_{bol}/L_{edd} which describe the ratio of intrinsic luminosity of AGN to the Eddington limit. There is also indirect evident that the massive BH gain mass via accreting close to the Eddington limit (Kollmeier et al., 2006). Figure 5 shows relationships between the BH mass and AGN bolometric luminosity at various redshift (Z). We can see that the AGN with higher redshift is also more likely to exhibiting higher mass and L_{bol}.
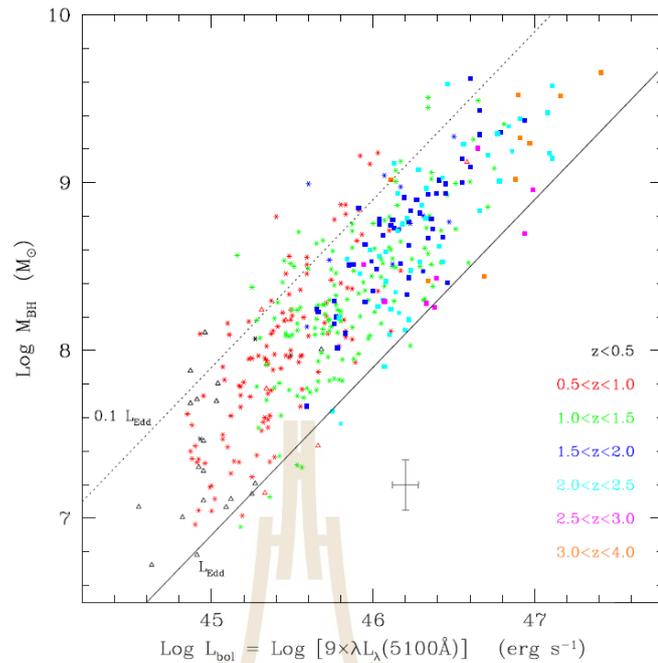
Figure 5 The relationship between the BH mass and AGN bolometric luminosity at
various redshift. Each dot represents the AGN at a different redshifts
represented by different colors.

Figure from Kollmeier et al., 2006.

2.1.4 Stellar mass

The estimation of the stellar mass utilizes CIGALE module and
procedure as described by Boquien et al., 2019. The CIGALE can fit a galaxy parameter,
i.e., stellar mass, star formation rate, etc., from the SED profile which can be obtained
from observations covering multiple wavelengths from far-ultraviolet to the radio
domain. The stellar mass is a parameter used in this work to determine its correlation
with the BH mass.

2.2 Timing analysis

The dynamics of phenomena around AGN can be studied through the timing analysis of the light curve profiles (Figure 6). The light curves from the observational data are the photon counts in the desired energy band as a function of time which can be further analyzed in the Fourier frequency domain via the Fourier transform technique.
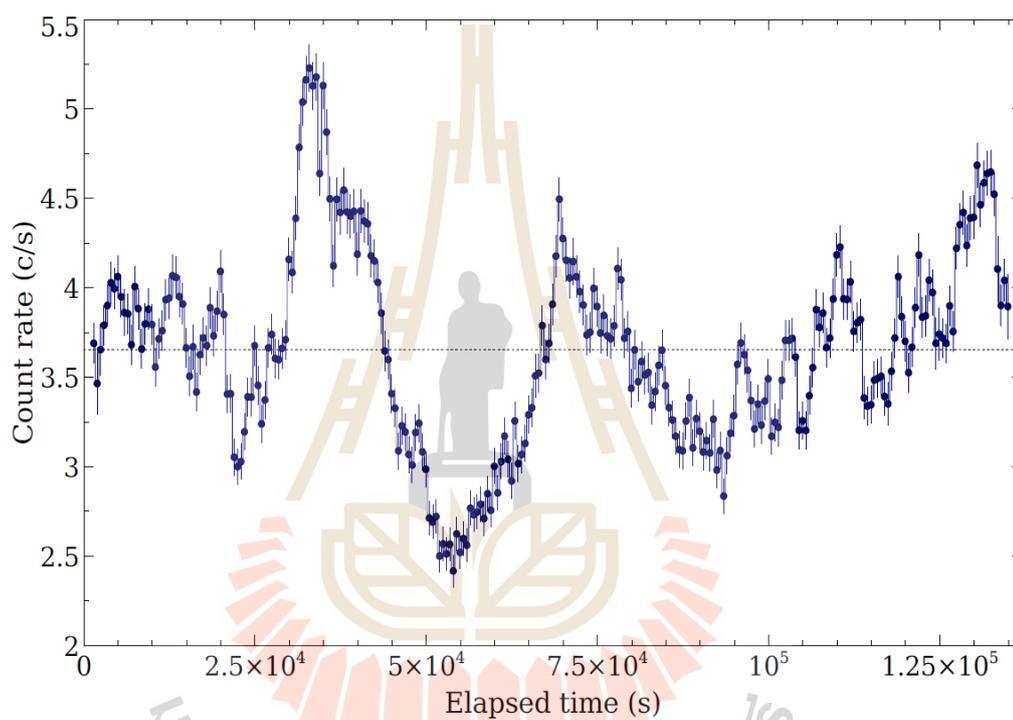
Figure 6 The light curve of the AGN Mkn 478 in the energy band 0.2 – 10 keV.

Figure from Barua, Jithesh, Misra, Medhi, and Adegoke, 2022.

2.2.1 X-ray reverberation

The X-ray reverberation is a phenomenon that caused by a difference in the path that light traveled from the same source, resulting in the difference in time that they traversed to the observer. The source of X-rays is believed to come from the mechanism where the photons from the disc has been up scattered by the electrons in the corona, making their energy reach to the X-ray band. The region that can cause the reverberation can be divided into several part based on their distance from the BH as seen in Figure 7. In this work, we are interested in the closest region which contains information about the corona geometry that can be probed using X-ray reverberation.
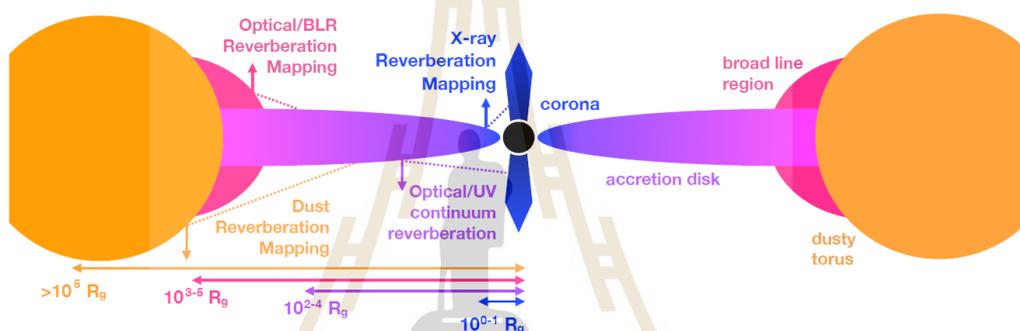
Figure 7 The schematic describing different types of reverberation around the AGN. Figure from Cackett, Bentz, and Kara, 2021.

2.2.2 Time lag

The time lag relates to the difference in time that photons travel in the X-ray reverberation phenomenon. The two paths that we consider are the one where photons from the source travel directly to the observer and another where photons from the source reprocessed with the accretion disc before traveling to the observer.

Under the LPM, the time delays (or X-ray reverberation time lags) between the direct coronal photons and the reflection photons can be observed. The variations of the light curves in the reflection dominated band (e.g. ~ 0.3 – 0.8 keV) lag behind those of the corona-emission dominated band (e.g. ~ 1 - 4 keV). The amplitude of the lags depends on the light-travel distance between the corona and the disc, then allowing us to probe the source geometry. Here, we select to use the light curves extracted in

0.3 – 0.8 keV and 1 - 4 keV bands as representatives of the reflection dominated and the corona-emission dominated band, and referred to as the soft and the hard band, respectively.

Note that it is difficult to measure the reverberation lags directly from the light curves. Instead, the time lag is normally measured in the Fourier frequency domain as described in Nowak, Vaughan, Wilms, Dove, and Begelman, 1999 with some of the application in De Marco et al., 2013, by computing the Fourier phase lag, $\phi(f)$, from the argument of the cross power spectrum between the Fourier forms of the soft and the hard band light curves

$$\phi(f) = \arg[C(f)], \text{ where } C(f) = < S^*(f)H(f) >. \qquad [1]$$

The time lag, $\tau(f)$, can be computed by

$$\tau(f) = \phi(f)/2\pi f . \qquad [2]$$

We use the sign convention where the positive time lag indicates the hard light curve lags the soft light curve, and vice versa.

2.2.3 Lag-frequency spectrum

An example of the Lag-frequency spectrum which can be affected by the height of the corona is shown in Figure 8. The X-ray reverberation lag is an observed parameter from the timing data that will be considered here. The lag-frequency describes the frequency where we detect time lag between two energy band. As we can see in Figure 9, the maximum lag-frequency where the soft reverberation lags are seen is anti-correlate with the corona height, while the lag amplitude is correlate with height.
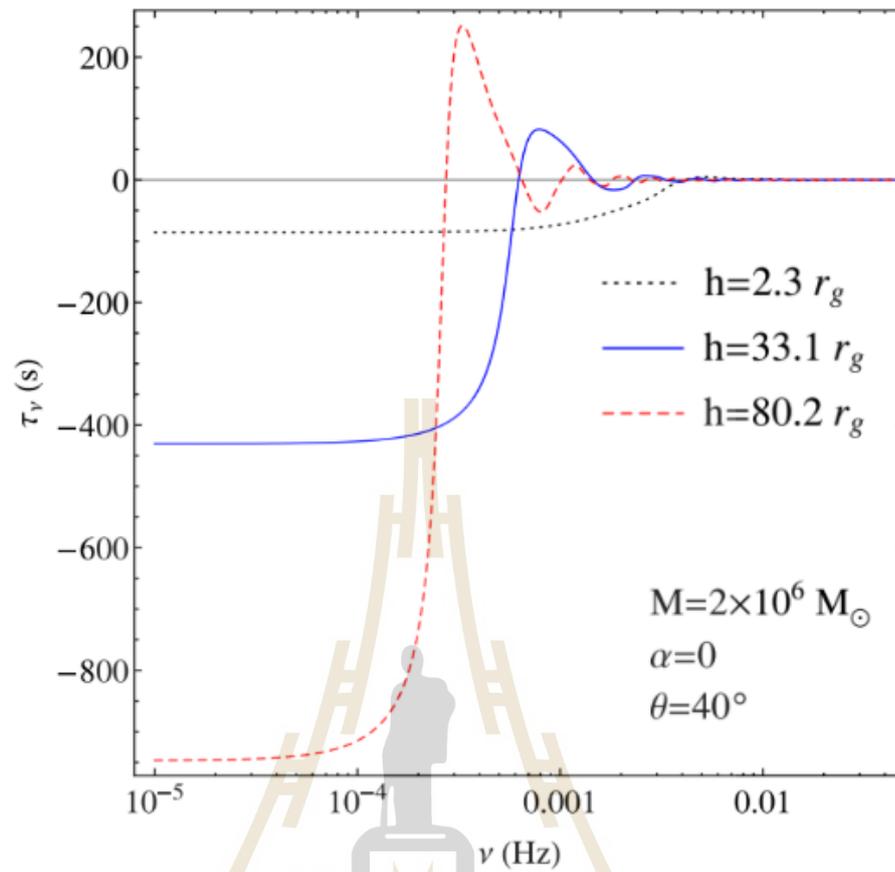
Figure 8 The example of the lag-frequency spectrum while varying the coronal
height (h) under the LPM.

Figure adapted from Emmanoulopoulos et al., 2014.

2.3 Discovered relation between AGN parameter

2.3.1 Time lag and BH mass

The relation between the time lag and the BH mass (M_{BH}) of the AGN has been analyzed by De Marco et al., 2013. They analyzed time lag from lag-frequency spectrum and discovered its strong correlation with the BH mass. They reported Spearman rank correlation coefficient between BH mass and time lag to be 0.9. They also employ Monte Carlo simulation to minimize the uncertainty when fitting linear equation. The discovered equation can be re-arranged to be

$$\log(M_{\mathrm{BH}}/M_{\odot}) = 8.81 + 1.69\log(\mathrm{Lag}). \qquad [3]$$

Note that we also derived the equation using the BH mass in the unit of solar mass for comparison purpose.

2.3.2 Stellar mass and BH mass

The previously discovered relation between the stellar mass and the BH mass has been analyzed by Reines and Volonteri, 2015 and Shankar et al., 2020 where discovered relations from these work are

$$\log(M_{\mathrm{BH}}/M_{\odot}) = 8.35 + 1.31\log\left(\frac{M_*}{10^{11}M_{\odot}}\right) \text{ and} \qquad [4]$$

$$\log(M_{\mathrm{BH}}/M_{\odot}) = 7.45 + 1.05\log\left(\frac{M_*}{10^{11}M_{\odot}}\right) \qquad [5]$$

respectively. The type of samples in Reines and Volonteri, 2015 is the AGN in local universe ($z < 0.055$) while Shankar et al., 2020 studied wider variety of galaxy types since they also want to investigate how a bias from selection effects can affect the relationship between the BH mass and the stellar mass.

2.4 Symbolic regression

The symbolic regression (SR) algorithm is an unsupervised machine learning model that uses genetic algorithms which mimic the natural selection process in nature to find the best population of mathematical expression of the input data. The SR operates on genetic programming (GP) as described in Koza, 1989 to find the best candidate of mathematical expression for the input by representing the equation as a tree of expression (Figure 9), and evolving the population composed of such tree with natural selection inspired process as the following procedure:

- Stochastic replacement of the sub tree by i.e., changing the operators, changing the variable, etc.
- Evaluate the prediction accuracy of the expression
- Find the survival of the fittest by eliminating the expression with poor prediction accuracy
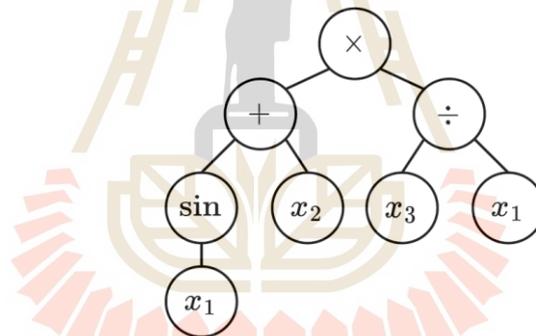
Figure 9 Example of a tree that represents the expression $f(\mathbf{x}) = (\sin(x_1) + x_2) \times x_3/x_1$.
Figure from Virgolin and Pissis, 2022.

The discovered equation can be analyzed to understand the principle of physics underlying the system. It has been applied in many fields of research, for example, astrophysics (Lemos, Jeffrey, Cranmer, Ho, and Battaglia, 2022) and mechanical engineering (Kronberger, Kommenda, Promberger, and Nickel, 2021; Virgolin, Wang, Alderliesten, and Bosman, 2020).

In this work, we will use the public SR model from Cranmer et al., 2020 for evaluating discovered equation to study the relations between observable spectral

and timing parameters of AGN. This can lead to a better understanding of the physics behind the accretion flows and X-ray reverberation phenomenon in the AGN.

2.5 Bootstrap resampling

Bootstrap resampling is one of a simple Monte Carlo simulation for resampling the data on assumption that the data is independent and identically distributed. The bootstrap involves generating a synthetic data set from the original data set by drawing arbitrary number of sample N from the sample with the size N, causing some data to appear more than once while some can disappeared as seen in Figure 10. The bootstrapping is suitable in the case where we do not know true distribution of the data set (William, 1992), such as our AGN sample.

Initial sample		New sample			
X	Y	X	Y	X	Y
1	2.5	1	2.5	2	1.3
2	1.3	1	2.5	3	4.6
3	4.6	2	1.3	4	8.2
4	8.2	3	4.6	4	8.2

Figure 10 The example of a result from 2 iteration of bootstrap resampling.

CHAPTER III

METHODOLOGY

3.1 The data

The data of X-ray reverberating AGN that we use are the observational data from the XMM-Newton archives and were previously analyzed by Hancock, Young, and Chainakun, 2022. The data investigated here then consists of 20 AGN samples in total. The samples of our interest are those that exhibit reverberation features. All data investigated here are summarized in Table 1. Note that the observed parameters obtained from the spectral analysis, which can be analyzed from the X-ray spectral fitting include the reflection fraction ($\mathbf{RF}$), photon index (Γ), Eddington ratio (λ_{Edd}), and stellar mass (M_*), while the parameters from the timing technique, which can be analyzed using the reverberation technique, include the BH mass (M_{BH}), time lag (τ), and Lag-frequency (υ). The lags are measured between 0.3 – 0.8 keV and 1 – 4 keV energy bands.

Table 1 The data of X-ray reverberating AGN that is used to train the SR in this work. Table from Thongkonsing et al. (in prep.)

AGN name	$\log(M_{\mathrm{BH}}/M_\odot)$	RF	Γ	τ (s)	ν (Hz)	λ_{Edd}	$\log(M_*/M_\odot)$
1H0707–495	6.31 (1)	$2.14^{+0.15}_{-0.15}$	$3.38^{+0.025}_{-0.02}$	29.1 ± 3.6	1.55×10^{-3}	1.05	10.88
Ark564	6.27 (2)	$0.64^{+0.44}_{-0.26}$	$2.36^{+0.06}_{-0.03}$	36.2 ± 10.5	6.07×10^{-4}	0.976	9.91
IRAS13224–3809	6.8 (3)	$3.20^{+0.36}_{-0.37}$	$3.25^{+0.04}_{-0.02}$	39.3 ± 9.6	5.06×10^{-4}	6.91	10.90
MCG–6–30–15	6.3 (4)	$10.00^{+0.00}_{-4.99}$	$2.00^{+0.02}_{-0.12}$	15.9 ± 5.9	9.66×10^{-4}	0.478	10.61
Mrk335	7.23 (R)	$10.00^{+0.00}_{-4.38}$	$2.82^{+0.28}_{-0.20}$	132.7 ± 36.4	2.65×10^{-4}	0.588	10.33
Mrk766	6.822 (4)	$4.33^{+0.41}_{-0.08}$	$1.89^{+0.01}_{-0.01}$	23.9 ± 6.7	9.66×10^{-4}	0.233	11.28
Mrk841	8.52 (2)	$10.00^{+0.00}_{-5.00}$	$2.00^{+0.49}_{-0.25}$	265.9 ± 217.5	1.02×10^{-4}	0.166	10.18
NGC1365	7.6 (3)	$10.00^{+0.00}_{-6.50}$	$1.59^{+0.04}_{-0.14}$	144.2 ± 113.4	7.27×10^{-5}	0.0195	11.51
NGC3516	7.395 (R)	$8.88^{+1.12}_{-0.54}$	$1.96^{+0.05}_{-0.07}$	256.6 ± 144.4	7.27×10^{-5}	0.0623	10.56
NGC4051	6.13 (R)	$10.00^{+0.00}_{-4.78}$	$1.84^{+0.05}_{-0.04}$	17.2 ± 7.1	9.66×10^{-4}	0.0107	8.44
NGC4151	7.65 (R)	$10.00^{+0.00}_{-4.53}$	$1.61^{+0.10}_{-0.12}$	488.0 ± 278.6	1.39×10^{-4}	0.0182	12.33
NGC4395	5.449 (R)	$0.40^{+0.00}_{-0.18}$	$1.06^{+0.00}_{-0.04}$	23.9 ± 16.2	5.06×10^{-4}	0.00423	10.29
NGC5548	7.718 (R)	$6.09^{+3.91}_{-1.29}$	$1.77^{+0.36}_{-0.36}$	156.7 ± 55.9	2.65×10^{-4}	0.0937	11.31
NGC6860	7.6 (3)	$2.19^{+1.68}_{-0.89}$	$3.20^{+0.20}_{-0.31}$	186.7 ± 192.5	1.94×10^{-4}	0.0102	11.35
NGC7314	6.7 (5)	$0.68^{+0.14}_{-0.13}$	$2.09^{+0.04}_{-0.06}$	1.6 ± 5.8	1.84×10^{-3}	0.0151	9.51
NGC7469	6.956 (R)	$0.30^{+0.14}_{-0.08}$	$2.39^{+0.14}_{-0.29}$	82.2 ± 51.1	3.71×10^{-4}	1.11	9.37
PG1211+143	7.61 (2)	$2.61^{+2.15}_{-1.18}$	$2.06^{+0.07}_{-0.07}$	215.6 ± 112.7	8.33×10^{-5}	2.88	10.15
PG1244+026	7.26 (6)	$7.61^{+2.39}_{-3.29}$	$1.94^{+0.33}_{-0.27}$	54.5 ± 20.3	5.06×10^{-4}	0.182	9.66
PG1247+267	8.92 (7)	$10.00^{+0.00}_{-6.41}$	$2.53^{+0.57}_{-0.28}$	498.6 ± 513.2	1.16×10^{-4}	2.09	14.37
REJ1034+396	6.18 (8)	$10.00^{+0.00}_{-4.19}$	$1.54^{+0.24}_{-0.27}$	55.6 ± 68.5	2.65×10^{-4}	0.660	10.42

3.2 The model

The SR model used here is adopted from Cranmer et al., 2020. The hyperparameters that we need to consider for fine-tuning the SR model are listed below:

- Binary operators: the basic algebraic operator in the equation which are $+, -, \times,$ and $\div$
- Unary operators: more complex algebraic operators such as log, exp, etc.
- Population: specify the initial population, which affects the diversity of the discovered equation
- Maxsize: control the maximum complexity of the discovered equation
- Niterations: control the total iteration for the algorithm to run
- Constraints: limit the maximum complexity for the argument of the unary operator

- Weights: describe how to weigh the mean squared error (MSE) based on the error for each input.

Note that the weight parameter regulates the importance between the AGN samples since it affects the loss of the discovered equation. For example, the sample with larger errors should be less important (less confidence) than the samples with small errors. We then define the weight factor in the SR model as

$$\text{weight} = \frac{1}{\sigma^2}.$$

[6]

where σ is a standard deviation.

The best equation is evaluated based on the calculated score by comparing how loss of the discovered equation reduced with increment in complexity:

$$\text{score} = \frac{\Delta\text{MSE}_c}{\Delta c},$$

[7]

where the MSE_c is the mean absolute error of respective equation with complexity c which depends on the number of variables and operators in the obtained equation.

3.3 The test for robustness

We investigate the correlation between parameter using the Spearman's rank correlation since the relationships among our parameters are likely to be non-linear. The p-value is also calculated alongside the correlation to check the statistical significance of the correlation value. The correlation and its corresponding p-value is computed by using the SciPy module in Python.

We evaluate the robustness of parameters by using the bootstrap resampling. Due to a small sample size and large scatter of the data, the bootstrap resampling test is used to analyze whether the obtained correlation and relation between the observed parameters is reliable. We use bootstrapping to resample the data set then analyze the robustness of the parameter relations from the distribution of the result.

The resampling is conducted by using the resample function from the scikit-learn module in Python.

CHAPTER IV

RESULT

We investigate the Spearman correlation between the BH mass and all AGN variables to find the suitable parameters for deriving the BH-mass equations via the SR technique. The result is shown in Table 2. The variables that show a strong correlation with the BH mass and have acceptable p-value (<0.05) are log(Lag), log(Lag-frequency) and $\log(M_*)$ while the RF, Γ, and λ_{Edd} show weak and insignificant correlation with the BH mass ($p > 0.05$). The log(Lag) has a strong correlation with mass which agrees with analysis from previous research (De Marco et al., 2013; Kara et al., 2016) since the size of the BH system scales with the mass, and the gravitational distance and time is proportional to the mass which results in increasing reverberation lags with mass.

Table 2 The Spearman correlation coefficients of the $\log(\mathrm{M_{BH}}/M_\odot)$ and AGN parameters.

	Variables	Spearman correlation coefficient	p-value
$\log(M_{\mathrm{BH}}/M_\odot)$	log(Lag)	0.859	1.22E-06
$\log(M_{\mathrm{BH}}/M_\odot)$	log(Lag-frequency)	-0.724	3.03E-04
$\log(M_{\mathrm{BH}}/M_\odot)$	$\log(M_*)$	0.479	3.26E-02
$\log(M_{\mathrm{BH}}/M_\odot)$	RF	0.309	1.85E-01
$\log(M_{\mathrm{BH}}/M_\odot)$	Γ	0.109	6.47E-01
$\log(M_{\mathrm{BH}}/M_\odot)$	λ_{Edd}	0.077	7.48E-01

4.1 Predicting the BH mass with single variable

We find the equation describing the BH mass using variables with strongest correlation which included $\log(\text{Lag})$, $\log(\text{Lag-frequency})$ and $\log(M_*)$. The SR model with single input variable to predict the BH mass is fine-tuned by setting the binary operator to be $+, -, \times,$ and $\div$ and unary operators include exponential, logarithm, square root, and square. The hyperparameters that control behavior of genetic algorithm are set as population=15, niteration=15, maxsize=20, and annealing=True. The weight is set as $1/\sigma^2$ where σ is error of the input parameter. We also set a constraint for complexity of unary operators. Note that when fitting the SR model using multiple variables, we changed the populations to 40 and removed the constraint parameter to increase diversity and complexity of the discovered equations.

The prediction of the BH mass using $\log(\text{Lag})$ from the equation with highest score is shown in Figure 11. The discovered equations in Table 3 show that the best equation is in the linear and polynomial form which might be possible to interpret for the physical meaning. Note that these equations are hand picked from the results of the SR since we want the equation that is suitable for discussion with other work and contain physical meaning.
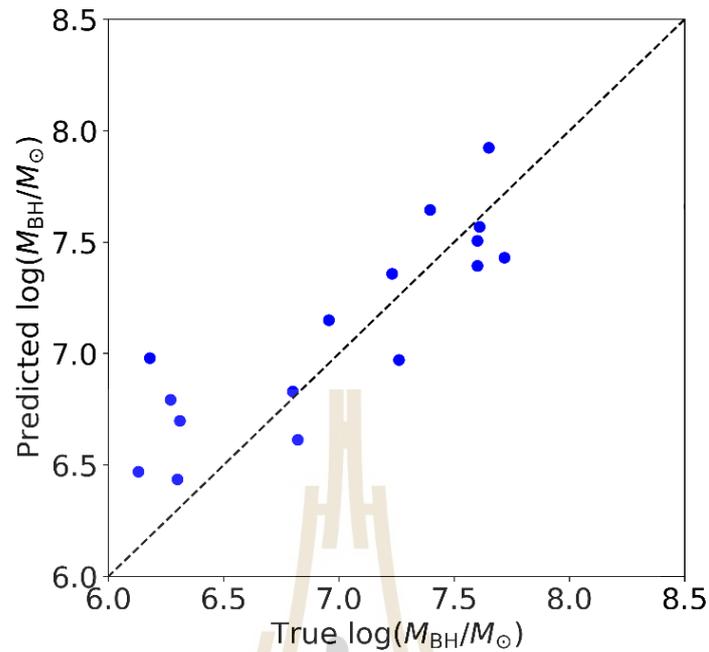
Figure 11 Predicting the BH mass with $\log(\text{Lag})$. The loss is 0.316.

Table 3 Best 3 equations sorted by their score from the SR model for predicting the $\log(\text{M}_{\text{BH}}/M_{\odot})$ using the $\log(\text{Lag})$.

C	equation	loss	score
4	$\log(M_{\text{BH}}/M_{\odot}) = \log(\text{Lag}) + 5.24$	0.32	0.39
6	$\log(M_{\text{BH}}/M_{\odot}) = 0.35(\log(\text{Lag}))^2 + 5.79$	0.32	0.35
7	$\log(M_{\text{BH}}/M_{\odot}) = \log(\text{Lag}) + \sqrt{\log(\text{Lag})} + 4.56$	0.25	0.11

The prediction of the BH mass with $\log(\text{Lag-frequency})$ from the equation with highest score is shown in Figure 12. Using $\log(\text{Lag})$ is better in terms of both accuracy and efficiency since the loss of the discovered equations as seen in Table 4 is worse and has higher complexity.
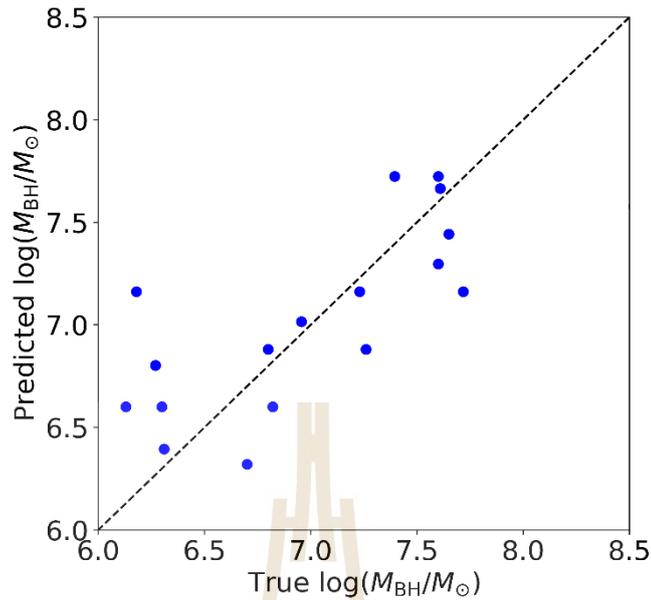
Figure 12 Predicting the BH mass with $\log(\text{Lag-frequency})$. The loss is 0.369.

Table 4 Best 3 equations sorted by the obtained SR score for predicting the BH mass using the $\log(\text{Lag-frequency})$.

C	equation	loss	score
3	$\log(M_{\text{BH}}/M_{\odot}) = 3.59 - \log(\text{Lag-frequency})$	0.369	0.309
9	$\log(M_{\text{BH}}/M_{\odot}) = 3.27 - 1.81\left(\sqrt{\log(\text{Lag-frequency}) + 3.95}\right)^2 + 5.67$	0.207	0.252
5	$\log(M_{\text{BH}}/M_{\odot}) = 2.39 - 1.34\log(\text{Lag-frequency})$	0.347	0.031

The prediction of the BH mass using the $\log(M_*/10^{11}M_{\odot})$ from the equation with highest score is shown in Figure 13. The best discovered equation with their respective loss and score is shown in Table 5. Note that we use stellar mass in unit of $10^{11}M_{\odot}$ to compare our result with previous literature.
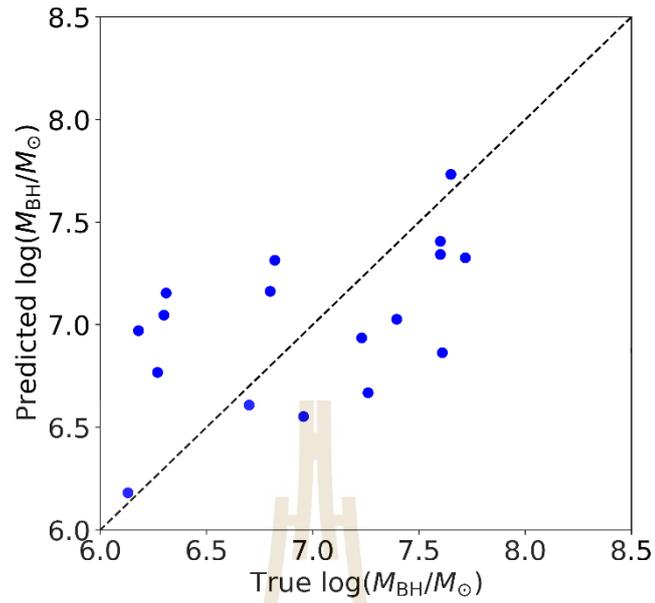
Figure 13 Predicting the BH mass with $\log(M_*/10^{11}M_\odot)$. The loss is 0.369.

Table 5 Best 3 equations sorted by the obtained SR score for predicting the BH mass using the $\log(M_*/10^{11}M_\odot)$.

C	equation	loss	score
5	$\log(M_{\mathrm{BH}}/M_\odot) = 0.399\log(M_*/10^{11}M_\odot) - 7.20$	0.455	0.296
14	$\log(M_{\mathrm{BH}}/M_\odot) = \sqrt[4]{\log\left(\dfrac{M_*}{10^{11}M_\odot}\right) + 0.12} + \sqrt{2\log\left(\dfrac{M_*}{10^{11}M_\odot}\right) + 6.52} + 3.77$	0.389	0.045
4	$\log(M_{\mathrm{BH}}/M_\odot) = \sqrt{\log(M_*/10^{11}M_\odot) + 50.3}$	0.611	0.038

4.2 Predicting the BH mass with multiple variables

We use multiple variables for predicting the BH mass to investigate improvement in accuracy of the prediction. Firstly, we use $\log(\text{Lag})$ and $\log(\text{Lag-frequency})$ to predict the BH mass. The results show that by using $\log(\text{Lag})$ and $\log(\text{Lag-frequency})$, the loss is reduced by 32% compared to using only $\log(\text{Lag})$ as seen in Figure 14. Note that the best equation in this case is hand-picked since the result does not always include both parameters in the discovered equation.

Since the discovered equation does not always include two variables in all solutions, the score cannot be easily compared between these equations. We then select the obtained equation that contains two variables (the $\log(\text{Lag})$ and $\log(\text{Lag-frequency})$) that has lowest loss and complexity.
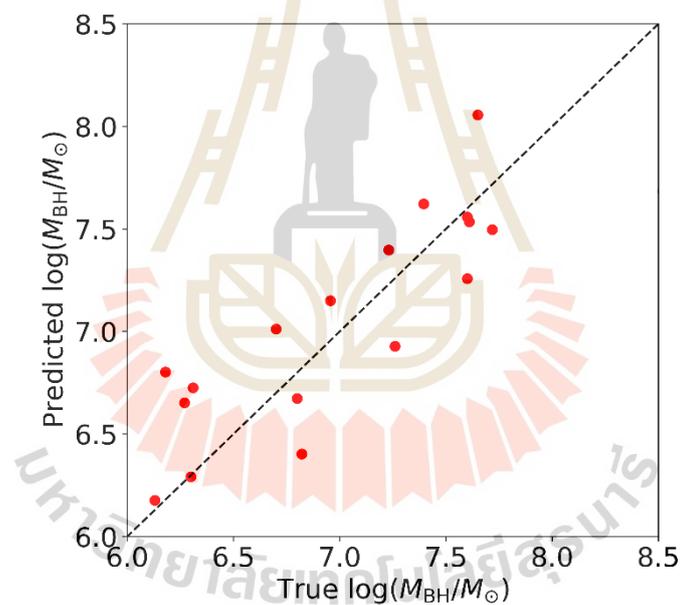
Figure 14 Predicting the BH mass with $\log(\text{Lag})$ and $\log(\text{Lag-frequency})$.
The loss is 0.160.

The result of predicting the BH mass with the $\log(\text{Lag})$ and RF is shown in Figure 15. For this case, we additionally define two input variable ($\log(\text{Lag}) \times \text{RF}$) and ($\log(\text{Lag}) \div \text{RF}$) as a single variable since the SR algorithm only includes the $\log(\text{Lag})$ in the best solution, which is probably because the RF has a weak

correlation with the BH mass. The score is significantly less than when we use only $\log(\text{Lag})$ to predict the BH mass.
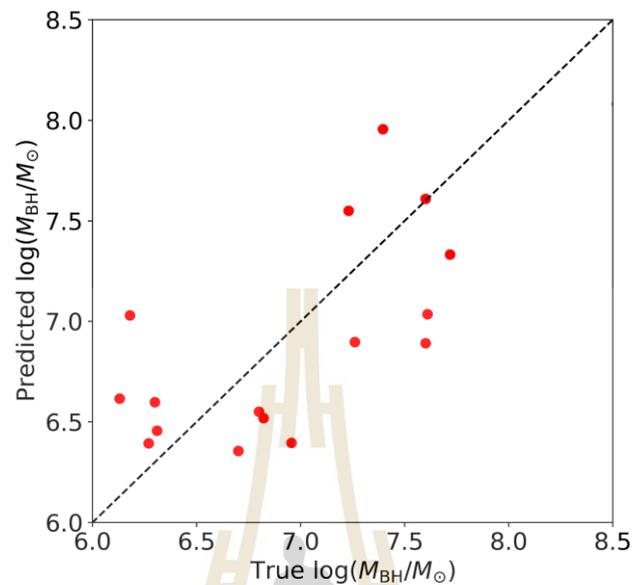
Figure 15 Predicting the BH mass with the Lag and RF. The loss is 0.262.

We also use three variables including the Lag, RF, and **Lag-frequency** to predict the mass (see Figure 16 and obtained equations in Table 6). It is clear that the mass prediction is more accurate than using a single variable, but the accuracy is not significantly different than predicting the mass using only the Lag and **Lag-frequency**. This confirms that the RF is not helpful in predicting the BH mass. In other words, none of the spectral data investigated here is significantly useful for predicting the BH mass.
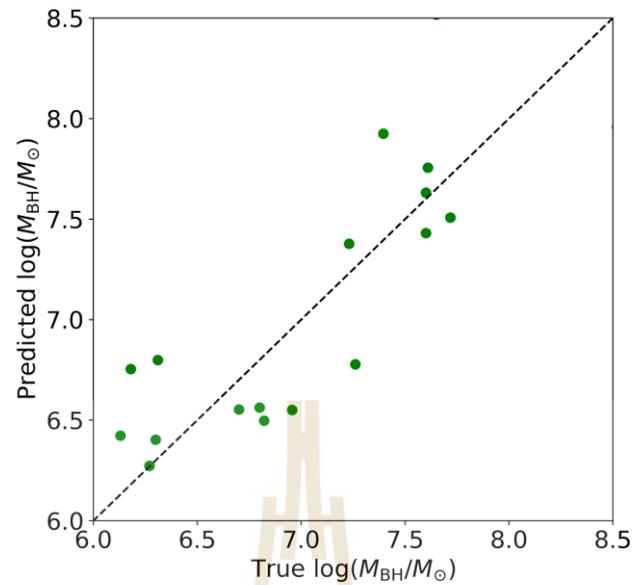
Figure 16 Predicting the BH mass with **Lag**, **RF**, and **Lag-frequency**.
The loss is 0.144.

Table 6 Best 3 equations sorted by the obtained loss for predicting mass with the
Lag, **RF**, and **Lag-frequency**.

C	equation	loss	score
16	$\log(M_{\mathrm{BH}}/M_\odot) = (-\,9.62\,\mathrm{Lag} - \dfrac{7.97}{0.0261 - \mathrm{Lag\text{-}frequency}} + \dfrac{132}{\mathrm{RF}})$	0.144	0.050
10	$\log(M_{\mathrm{BH}}/M_\odot) = (-\,9.97\,\mathrm{Lag} - \dfrac{0.451}{0.00245 - \mathrm{Lag\text{-}frequency}})$	0.165	0.154
6	$\log(M_{\mathrm{BH}}/M_\odot) = \log(\,\mathrm{Lag} + 5.52) + 2.08$	0.194	0.249

4.3 Predicting timing data using spectral data

In this section, we use the parameters in the spectral data (e.g., the RF and Γ) to predict the timing data (the Lag), so it could be possible to link the parameters from the spectral data to the timing data. We begin by using only the RF to predict the Lag and the result shows that the accuracy is poor as seen in Figure 17. The overall accuracy and score of the best equation are likewise subpar as shown in Table 7.
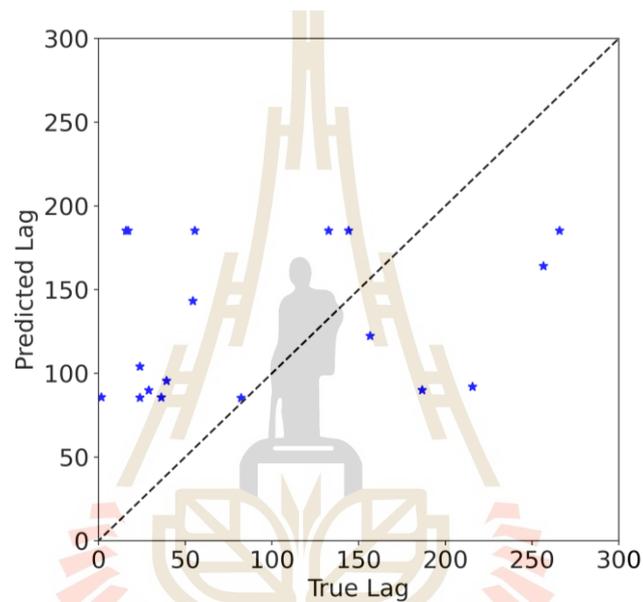
Figure 17 Predicting the Lag with the RF. The loss is 1.72E4.

Table 7 Best 3 equations sorted by score for predicting the Lag with the RF.

C	equation	loss	score
4	$Lag = RF^2 + 85.2$	1.72E4	0.166
12	$Lag = RF(RF - 4.68)\log(RF + 2.61) + 76.8$	1.62E4	0.0311
5	$Lag = 15.8\,RF + 41.9$	1.68E4	0.0198

Now, we use two spectral parameters which are RF and Γ to predict the Lag. The result shows that the accuracy is not significantly better than when using only RF

as seen in Figure 18. The summary of the best equations based on the prediction accuracy is shown in Table 8.
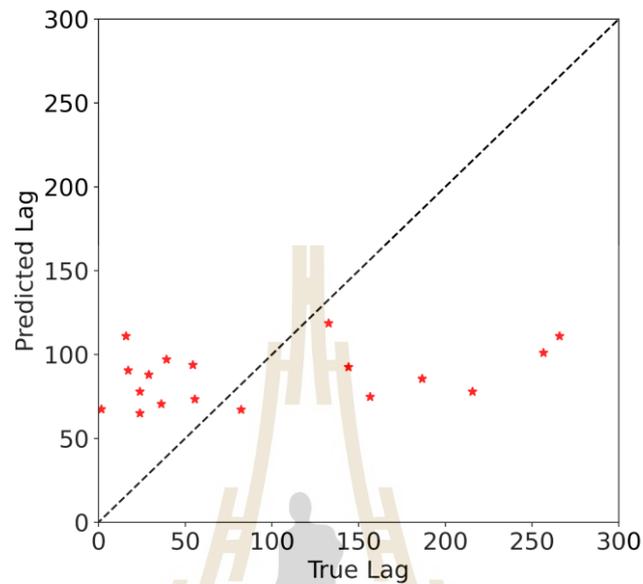
Figure 18 Predicting Lag with RF and Γ. The loss is 6.10E3.

Table 8 Best 3 equations sorted by accuracy for predicting the Lag with the RF and Γ.

C	equation	loss	score
18	$\mathrm{Lag} = \mathrm{RF}\log(3.99\,\Gamma^2 - 10.3)^2 + \dfrac{\mathrm{RF}}{2.56 - \Gamma} + 63.4$	6.10E3	0.386
17	$\mathrm{Lag} = 8.23\left(0.349(\mathrm{RF} + \Gamma) + 1 - \dfrac{0.0133}{(0.12(\mathrm{RF} - \Gamma) - 1)^2}\right)^2 + 18.8$	8.98E3	0.241
13	$\mathrm{Lag} = \dfrac{RF}{2.56 - \Gamma} + 36.4\sqrt{(0.345\,RF + 1)^2}$	1.19E4	0.260

4.4 Verifying the robustness with bootstrapping

4.4.1 Correlation between the BH mass and AGN parameter

The robustness of the AGN parameters has been verified by bootstrapping to find the correlation and its corresponding p-value of each parameters with the BH mass. Figure 19 shows the result after bootstrapping the data by 4000 iterations. In each iteration, the Spearman's rank correlation coefficient and their respective p-value of each parameter pair as described in Table 2 has been evaluated. The result show that only first three parameters pairs from the top panels of Figure 19 are the most robust parameters for our data set. These parameters show high correlation value are likely to have the p-value < 0.05.
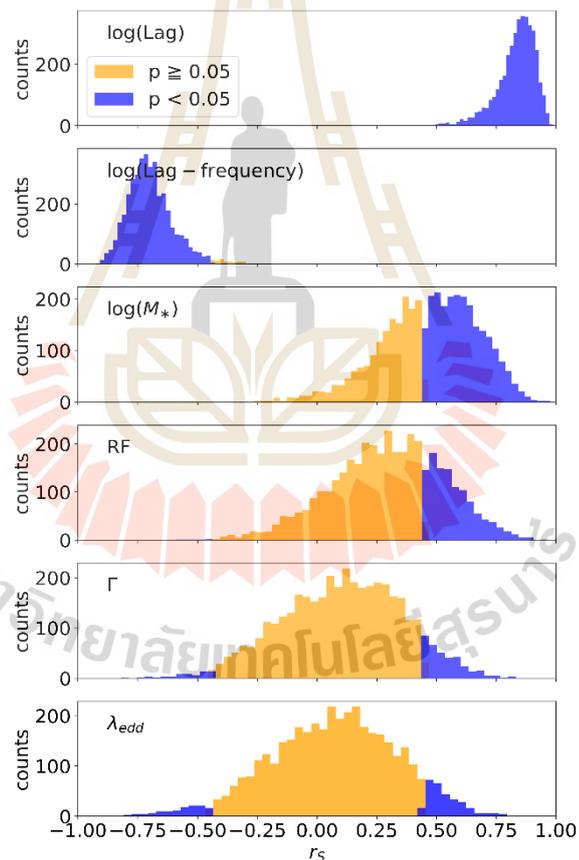
Figure 19 The Spearman's rank correlation and the counts where result derived from 4000 iterations of bootstrap resampling. The p-values of < 0.05 and ≥ 0.05 are described as blue and yellow, respectively.

4.4.2 Relationship between the BH mass and time lag

We verify the linear and polynomial relationship between the BH mass and time lag as discovered by the SR. The bootstrap resampling is performed by 2000 iteration. The equation is fitted using the linear and polynomial equation in **log-log** scales. The result from bootstrapping with the linear equation is shown in Figure 20 and we define the equation as

$$\log\left(M_{\mathrm{BH}}/M_{\odot}\right) = \alpha + \beta \log\left(\mathrm{Lag}\right) \qquad [8]$$

Where the α is intercept and the β is slope. We find that the parameter in this form with high prediction accuracy (high R^2 score) is cluster in the range of $\alpha \sim 4.9 - 5.5$ and $\beta \sim 0.7 - 1.2$ which agree with the result from SR.
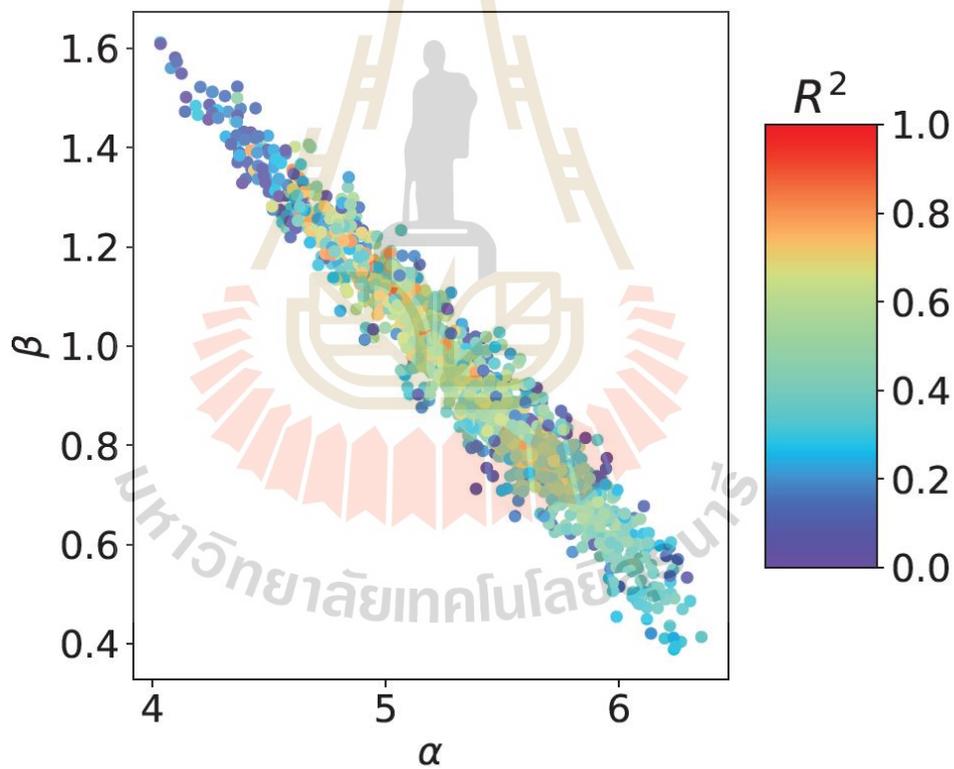
Figure 20 The best fitting parameter of linear equation from 2000 iteration of
  bootstrapping.

We also fit the data with the polynomial equation:

$$\log(M_{\mathrm{BH}}/M_{\odot}) = \alpha + \beta \, (\log(\mathrm{Lag}))^{\gamma}, \qquad [9]$$

where the α is intercept, the β is slope, and γ is the power. The result from 2000 iteration of bootstrap resampling is shown in Figure 21. The parameters with high prediction accuracy (high R^2 score) are cluster in the range of $\alpha \sim 5.0 - 6.0$, $\beta \sim 0.1 - 0.6$, and $\gamma \sim 1.7 - 2.2$ which also agree with the SR. From bootstrapping, we also see that the M_{BH}-Lag relation prefers the polynomial form with $\gamma \neq 1$.
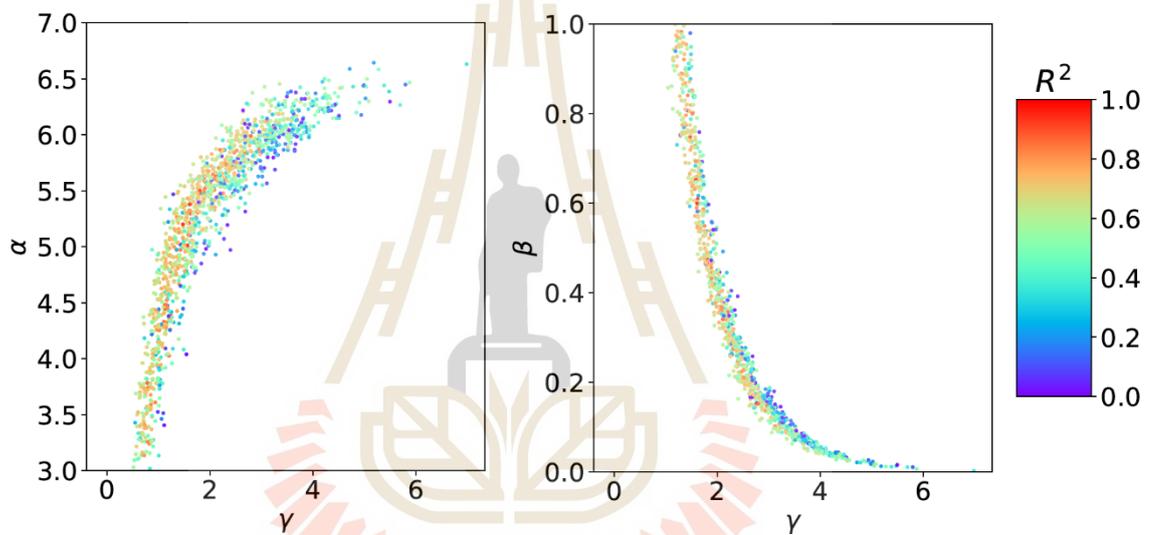
Figure 21 The best fitting parameter of the polynomial equation from 2000 iteration of bootstrapping. Left panel shows the best fitting values of α and right panel shows best fitting values of β, with their respective γ.

CHAPTER V
DISCUSSION AND CONCLUSION

The strong correlation between time lag and BH mass also agrees with what discovery by De Marco et al., 2013. We confirm that the strong correlation between time lag and BH mass is likely certain from our bootstrapping result. We also discovered the parameter value in the linear equation from the SR model to be slightly differed from their work. However, our bootstrap and SR also suggest that polynomial form of relation between the time lag and the BH mass might be better for describing the relationship between these parameters.

The height of the corona and the BH mass can affect the light travel time, causing the measured time lag to differ. The non-linear relationship between these parameters from our SR model and bootstrapping suggests that the geometry of corona in our sample is quite varied (it is not at the same gravitational distance for these AGN).

The comparison of stellar mass and the BH mass relation with Reines and Volonteri, 2015 and Shankar et al., 2020 also shows that slope of the linear equation from our SR model is greatly differed from their work, as seen in Figure 22. This might come from difference in types of AGN samples.
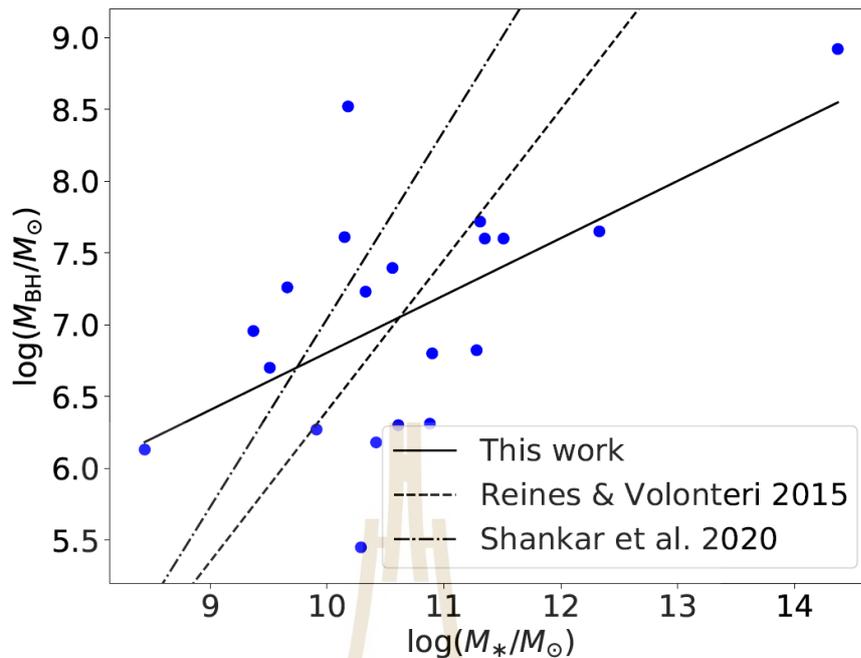
Figure 22 The comparison of the linear equation describing the BH mass and the stellar mass between our work and other literature.

There are other literatures that employ different methods for determining the BH mass using the parameters of AGN. Akylas, Papadakis, and Georgakakis, 2022 use the excess variance to estimate the BH mass. With appropriate quality of light curve determined from exposure time of 80-100 ks and signal to noise ratio of 3, one can measure the BH mass from excess variance with a good accuracy (average order of uncertainty is 0.3 dex). Our SR model is worse than this literature in terms of the prediction accuracy, but we do not have a strong requirement for the quality of the light curve.

Chainakun, Fongkaew, Hancock, and Young, 2022 focused on using the time lag in the Fe-K band and the excess variance to predict the BH mass using a neural network model. They achieved high prediction accuracy with predicted value differed by ~3.5% from true value. The performance of our SR model in predicting the BH mass is also worse than their neural network model in terms of accuracy. However, the SR model can uncover the explicit equation describing the time lag and the BH mass, while it is not straightforward for the neural network to do the same. These works show us that

it might be worth conducting analysis on using the excess variance to assist in prediction of the BH mass with the SR model in the future.

In conclusion, while the prediction accuracy of the SR model is not that great compared to other methods, it can describe the relationship between the input parameters without specific criterion of the samples. The time lag is the best candidate among our AGN parameter for predicting the BH mass since including other parameters show insignificant improvement in performance of the SR model. The relationship between the stellar mass and the BH mass from our SR model also differed from previous literature which need more detailed analysis in the future. We also need a larger amount of AGN samples with reverberation feature to draw more robust conclusion from the result of our work.

On the other hand, we can use the SR to derive the equation in terms of $\mathrm{Lag}(M_{BH}, RF, \Gamma)$ rather than $M_{BH}(\mathrm{Lag}, RF, \Gamma)$. In this way, the nature of time lags dependence on other parameters can be studied. We perform this test and find, e.g., that RF can act as a correction term in the lag-mass scaling relation. These results will be reported in Thongkonsing et al. (submitted to MNRAS, under revision).
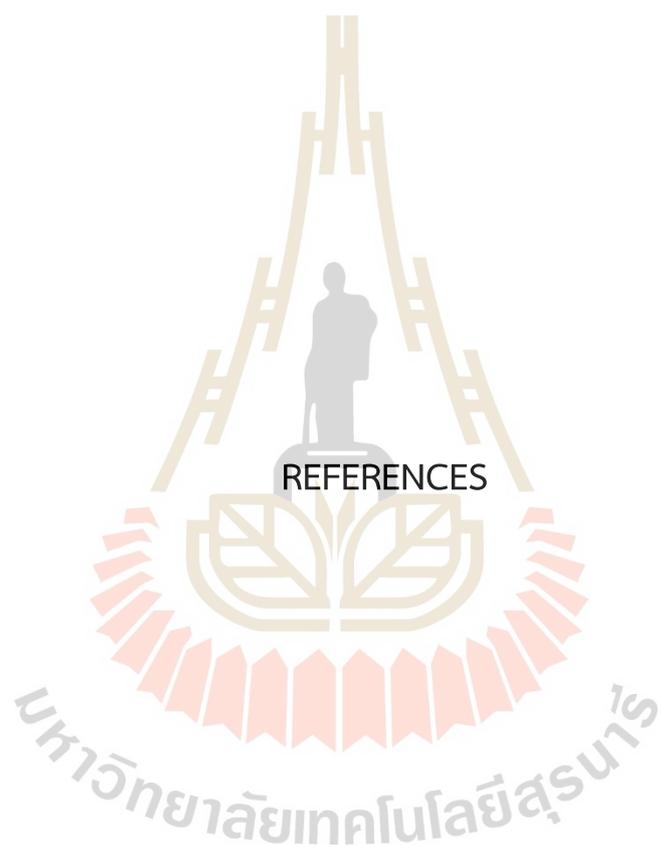
REFERENCES

REFERENCES

Akylas, A., Papadakis, I., and Georgakakis, A. (2022). Black hole mass estimation using X-ray variability measurements in Seyfert galaxies. *Astronomy and Astrophysics, 666*, A127. doi:10.1051/0004-6361/202244162

Barua, S., Jithesh, V., Misra, R., Medhi, B., and Adegoke, O. (2022). *Correlated variability of the reflection fraction with the X-ray flux and spectral index for Mkn 478*.

Bentz, M. C., and Katz, S. (2015). The AGN Black Hole Mass Database. *Publications of the Astronomical Society of the Pacific, 127*(947), 67. doi:10.1086/679601

Boquien, M., Burgarella, D., Roehlly, Y., Buat, V., Ciesla, L., Corre, D., . . . Salas, H. (2019). CIGALE: a python Code Investigating GALaxy Emission. *Astronomy and Astrophysics, 622*, A103. doi:10.1051/0004-6361/201834156

Cackett, E. M., Bentz, M. C., and Kara, E. (2021). Reverberation mapping of active galactic nuclei: from X-ray corona to dusty torus. *iScience, 24*, 102557. doi:10.1016/j.isci.2021.102557

Chainakun, P., Fongkaew, I., Hancock, S., and Young, A. J. (2022). Predicting the black hole mass and correlations in X-ray reverberating AGNs using neural networks. *Monthly Notices of the Royal Astronomical Society, 513*(1), 648-660. doi:10.1093/mnras/stac924

Collinson, J., Ward, M., Landt, H., Done, C., Elvis, M., and McDowell, J. (2016). Reaching the Peak of the quasar spectral energy distribution - II. Exploring the accretion disc, dusty torus and host galaxy. *Monthly Notices of the Royal Astronomical Society, 465*. doi:10.1093/mnras/stw2666

Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., and Ho, S. (2020). *Discovering Symbolic Models from Deep Learning with Inductive Biases*.

Dauser, T., García, J., Walton, D. J., Eikmann, W., Kallman, T., McClintock, J., and Wilms, J. (2016). Normalizing a relativistic model of X-ray reflection. Definition of the reflection fraction and its implementation in relxill. *Astronomy and Astrophysics, 590*, A76. doi:10.1051/0004-6361/201628135

De Marco, B., Ponti, G., Cappi, M., Dadina, M., Uttley, P., Cackett, E. M., . . . Miniutti, G. (2013). Discovery of a relation between black hole mass and soft X-ray time lags in active galactic nuclei. *Monthly Notices of the Royal Astronomical Society, 431*, 2441-2452. doi:10.1093/mnras/stt339

Emmanoulopoulos, D., Papadakis, I. E., Dovčiak, M., and McHardy, I. M. (2014). General relativistic modelling of the negative reverberation X-ray time delays in AGN. *Monthly Notices of the Royal Astronomical Society, 439*, 3931-3950. doi:10.1093/mnras/stu249

García, J., Dauser, T., Reynolds, C. S., Kallman, T. R., McClintock, J. E., Wilms, J., and Eikmann, W. (2013). X-RAY REFLECTED SPECTRA FROM ACCRETION DISK MODELS. III. A COMPLETE GRID OF IONIZED REFLECTION CALCULATIONS. *The Astrophysical Journal, 768*(2), 146. doi:10.1088/0004-637X/768/2/146

George, I. M., and Fabian, A. C. (1991). X-ray reflection from cold matter in Active Galactic Nuclei and X-ray binaries. *Monthly Notices of the Royal Astronomical Society, 249*, 352. doi:10.1093/mnras/249.2.352

Haardt, F. (1993). Anisotropic Comptonization in Thermal Plasmas: Spectral Distribution in Plane-Parallel Geometry. *The Astrophysical Journal, 413*, 680. doi:10.1086/173036

Hancock, S., Young, A. J., and Chainakun, P. (2022). X-ray timing and spectral analysis of reverberating active galactic nuclei. *Monthly Notices of the Royal Astronomical Society, 514*(4), 5403-5421. doi:10.1093/mnras/stac1653

Kara, E., Alston, W., Fabian, A. C., Cackett, E., Uttley, P., Reynolds, C., and Zoghbi, A. (2016). A global look at X-ray time lags in Seyfert galaxies. doi:10.17863/CAM.8076

Kollmeier, J. A., Onken, C. A., Kochanek, C. S., Gould, A., Weinberg, D. H., Dietrich, M., . . . Stern, D. (2006). Black Hole Masses and Eddington Ratios at 0.3 andlt; z andlt; 4*. *The Astrophysical Journal, 648*(1), 128. doi:10.1086/505646

Koza, R. (1989). Hierarchical genetic algorithms operating on populations of computer programs.

Kronberger, G., Kommenda, M., Promberger, A., and Nickel, F. (2021). Predicting Friction System Performance with Symbolic Regression and Genetic Programming with Factor Variables. *arXiv e-prints*, arXiv:2107.09484. Retrieved from https://ui.adsabs.harvard.edu/abs/2021arXiv210709484K

Lemos, P., Jeffrey, N., Cranmer, M., Ho, S., and Battaglia, P. (2022). Rediscovering orbital mechanics with machine learning. *arXiv e-prints*, arXiv:2202.02306. Retrieved from https://ui.adsabs.harvard.edu/abs/2022arXiv220202306L

Nowak, M. A., Vaughan, B. A., Wilms, J., Dove, J. B., and Begelman, M. C. (1999). Rossi X-Ray Timing Explorer Observation of Cygnus X-1. II. Timing Analysis. *The Astrophysical Journal, 510*, 874-891. doi:10.1086/306610

Reines, A. E., and Volonteri, M. (2015). RELATIONS BETWEEN CENTRAL BLACK HOLE MASS AND TOTAL GALAXY STELLAR MASS IN THE LOCAL UNIVERSE. *The Astrophysical Journal, 813*(2), 82. doi:10.1088/0004-637X/813/2/82

Ross, R. R., and Fabian, A. C. (2005). A comprehensive range of X-ray ionized-reflection models. *Monthly Notices of the Royal Astronomical Society, 358*, 211-216. doi:10.1111/j.1365-2966.2005.08797.x

Shankar, F., Allevato, V., Bernardi, M., Marsden, C., Lapi, A., Menci, N., . . . Sheth, R. K. (2020). Constraining black hole-galaxy scaling relations and radiative efficiency from galaxy clustering. *Nature Astronomy, 4*, 282-291. doi:10.1038/s41550-019-0949-y

Urry, C. M., and Padovani, P. (1995). Unified Schemes for Radio-Loud Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific, 107*, 803. doi:10.1086/133630

Virgolin, M., and Pissis, S. P. (2022). Symbolic Regression is NP-hard. *arXiv e-prints*, arXiv:2207.01018. doi:10.48550/arXiv.2207.01018

Virgolin, M., Wang, Z., Alderliesten, T., and Bosman, P. (2020). Machine learning for the prediction of pseudorealistic pediatric abdominal phantoms for radiation dose reconstruction. *Journal of Medical Imaging, 7*, 1. doi:10.1117/1.JMI.7.4.046501

William, H. P. (1992). *Numerical recipes in C : the art of scientific computing*: Second
edition. Cambridge [Cambridgeshire] ; New York : Cambridge University Press,
1992.
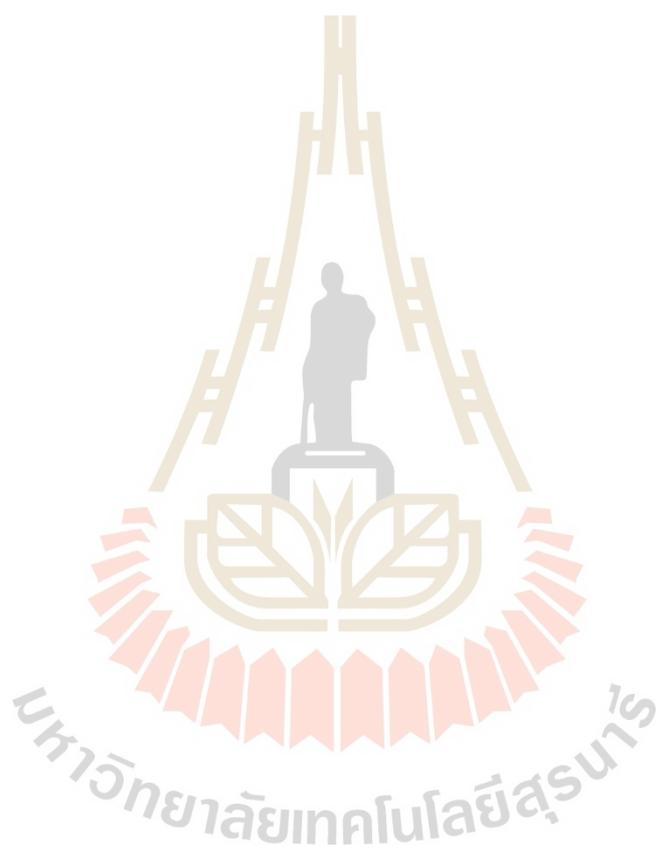
CIRRICULUM VITAE

Name: Prathak Thongkonsing   **Date of Birth**: 02/12/1998

Place of Birth: Nakhonratchasima, Thailand

Residence: 299/177, Sikhio, Sikhio, Nakhonratchasima, Thailand 30140.

Tel: +66-986407087

Education

2017-2020: Bachelor of Science, School of Physics, Institute of Science, Suranaree University of Technology

2020-present: Master of Science, School of Physics, Institute of Science, Suranaree University of Technology

Publication

1. Thongkonsing, P., Chainakun, P., Worrakitpoonpon, T., and Young, A. J. (2024). Investigating scaling relations in X-ray reverberating AGN using symbolic regression. Monthly Notices of the Royal Astronomical Society, 527(2), 1950-1961.

2. Chainakun, P., Mankatwit, N., Thongkonsing, P., and Young, A. J. (2021). Machine learning application to detect light echoes around black holes. Monthly Notices of the Royal Astronomical Society, 506(4), 5318-5327.

3. Chainakun, P., Luangtip, W., Young, A. J., Thongkonsing, P., and Srichok, M. (2021). Evolution of the truncated disc and inner hot-flow of GX 339-4. Astronomy & Astrophysics, 645, A99.