

OPTIMIZING OPTICAL FILTERS FOR JOVIAN PLANETS
TRANSMISSION SPECTROSCOPY



A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Physics
Suranaree University of Technology
Academic Year 2021

แผ่นกรองแสงในช่วงคลื่นแสงที่เหมาะสมสำหรับการทำทรานสมิสชัน
สเปกโทรสโกปีของดาวเคราะห์โจเวียน



นางสาวพัชรวีร์ หมั่นสระเกษ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาฟิสิกส์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2564

OPTIMIZING OPTICAL FILTERS FOR JOVIAN PLANETS TRANSMISSION
SPECTROSCOPY

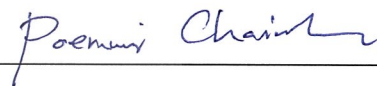
Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for a Master degree.

Thesis Examining Committee



(Dr. Utane Sawangwit)

Chairperson



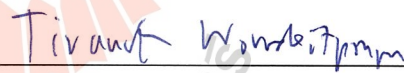
(Asst. Prof. Dr. Poemwai Chainakun)

Member (Thesis Adviser)



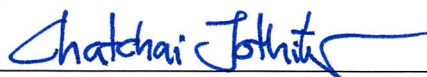
(Dr. Supachai Awiphan)

Member (Thesis Co-Advisor)



(Asst. Prof. Dr. Tirawut Worakitpoonpon)

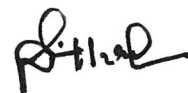
Member



(Assoc. Prof. Dr. Chatchai Jothityangkoon)

Vice Rector for Academic Affairs

and Quality Assurance



(Prof. Dr. Santi Maensiri)

Dean of Institute of Science

พัชรวีร์ หมั่นสระเกษ : แผ่นกรองแสงในช่วงคลื่นแสงที่เหมาะสมสำหรับการทำทรานสมิสมิชันสเปกโทรสโกปีของดาวเคราะห์โจเวียน (OPTIMIZING OPTICAL FILTERS FOR JOVIAN PLANETS TRANSMISSION SPECTROSCOPY) อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร. เพิ่มวัย ชัยนะกุล, 46 หน้า.

คำสำคัญ: ชั้นบรรยากาศของดาวเคราะห์นอกระบบสุริยะ, ทรานสมิสมิชันสเปกโทรสโกปี, เทคนิคการดึงข้อมูล, Random Forest Regression

คุณสมบัติของดาวเคราะห์นอกระบบสุริยะสามารถหาได้โดยใช้เทคนิคการดึงข้อมูลจากทรานสมิสมิชันสเปกตรัม อย่างไรก็ตามวิธีการคำนวณแบบดั้งเดิม (เช่น MCMC หรือ nested sampling) ยังคงใช้เวลานาน ในปัจจุบัน machine learning ได้ถูกนำมาใช้เพื่อลดระยะเวลาในการคำนวณ นอกจากนี้การเลือกใช้ชุดแผ่นกรองแสงจากระบบแผ่นกรองต่าง ๆ เช่น ระบบแผ่นกรองแสงในช่วงคลื่นแสงที่ตามองเห็น Johnson-Cousins และ Sloan Digital Sky Survey (SDSS) ยังส่งผลต่อเวลาที่ใช้เช่นกัน โดยการทำทรานสมิสมิชันสเปกตรัมจะใช้แผ่นกรองแสงที่มีอยู่ทั้งหมดเพื่อให้ได้ข้อมูลที่มากที่สุดที่เป็นไปได้ จากปัญหาดังกล่าวงานนี้จึงนำอัลกอริทึม Random Forest Regression มาใช้เพื่อศึกษาคุณสมบัติของดาวเคราะห์นอกระบบสุริยะประเภทดาวเคราะห์โจเวียนในช่วงความยาวคลื่นแสงที่ตามองเห็นเปรียบเทียบกับวิธีการหาแบบดั้งเดิมของ PLATON ซึ่งพบว่า Random Forest Regression สามารถทำนายรัศมีของดาวเคราะห์นอกระบบได้อย่างแม่นยำ ($R^2_{Fit} = 0.999$) ในขณะที่การทำนายมวล อุณหภูมิ และความเป็นโลหะของชั้นบรรยากาศของดาวเคราะห์นั้นอยู่ในเกณฑ์ที่ยอมรับได้ แต่ไม่สามารถทำนายอัตราส่วนของคาร์บอนต่อออกซิเจนในชั้นบรรยากาศได้ และพบว่า Random Forest Regression ใช้เวลาในการหาค่าพารามิเตอร์น้อยกว่าวิธีของ PLATON นอกจากนี้ยังมีการจัดลำดับของแผ่นกรองที่เหมาะสมสำหรับการศึกษาระดับชั้นบรรยากาศของดาวเคราะห์นอกระบบประเภทโจเวียนในช่วงคลื่นแสงที่ตามองเห็น โดยพิจารณาจากค่า feature importance ซึ่งแผ่นกรองแสงที่มีความสำคัญต่ำที่สุดจะถูกตัดทิ้งและแผ่นกรองแสงที่เหลือจะถูกนำไปใช้คำนวณต่อในรอบถัดไป โดยลำดับของแผ่นกรองที่ได้เรียงจากมากไปน้อยดังนี้ I, U, g', i', B, V, z', u', r', และ R และเมื่อพิจารณาความแม่นยำในการทำนายค่าพารามิเตอร์ของชุดแผ่นกรองแสงในแต่ละรอบพบว่าชุดแผ่นกรองแสง 7 ตัว (I, U, g', i', B, V และ z') ให้ผลการทำนายใกล้เคียงกับชุดแผ่นกรองจำนวนสูงสุด 10 ตัว

สาขาวิชาฟิสิกส์
ปีการศึกษา 2564

ลายมือชื่อนักศึกษา พัชรวีร์ หมั่นสระเกษ
ลายมือชื่ออาจารย์ที่ปรึกษา เพิ่มวัย ชัยนะกุล

PATCHARAWEE MUNSAKET : OPTIMIZING OPTICAL FILTERS FOR JOVIAN PLANETS TRANSMISSION SPECTROSCOPY. THESIS ADVISOR : ASST. PROF. POEMWAI CHAINAKUN, Ph.D. 46 PP.

Keyword: exoplanet atmosphere, transmission spectroscopy, retrieval technique, random forest regression

Exoplanetary atmospheric research using transmission spectroscopy is one of the most active fields in exoplanetary research. A retrieval technique is used to extract the transmission spectra for an understanding of the exoplanet atmospheres. However, traditional retrieval methods (e.g., MCMC and nested sampling) take a long time to retrieve parameters. To save computation time, machine learning methods were used to retrieve exoplanet atmosphere properties. Apart from that, there are also various filter selection choices, such as the Johnson-Cousins and Sloan Digital Sky Survey (SDSS) systems. In the transmission spectroscopy observation, all available filters are used to receive all possible information without filter ordering or filter selection. The goal of this research is to use random forest regression, a supervised machine learning algorithm, to obtain the atmosphere characteristics of Jovian planets from broad-band transmission spectra at an optical wavelength instead of the traditional method. The random forest regressor has been shown to have the best accuracy in predicting planetary radius ($R^2_{\text{Fit}} = 0.999$) while also having acceptable accuracy in predicting planetary mass, temperature, and metallicity of the planetary atmosphere. Random forest regression takes significantly less time to process while providing results that are comparable to the PLATON package's nested sampling retrieval. Furthermore, the feature importance is used to rank the filters. Every iteration, the filter with the lowest significance is dropped. The filters are arranged in descending sequence as I, U, g', i', B, V, z', u', r', and R. Our results suggest that, the optimal number of filters that yield the closest prediction to the highest filters is 7.

School of Physics
Academic Year 2020

Student's Signature พัชรวิทย์ นพินสรนภะ
Advisor's Signature โพยมวไล ชัยนาคุน

ACKNOWLEDGEMENTS

The completion of my thesis would not have been possible without the assistance, support, and encouragement of the following people:

First and foremost, I would like to express my deepest appreciation to my thesis advisors, Asst. Prof. Dr. Poemwai Chainakun and Dr. Supachai Awiphan, from the National Astronomical Research Institute of Thailand (NARIT), for their support, assistance, thoughtfulness, and kindly guidance throughout the study. I would like to extend my sincere thanks to Dr. Eamonn Kerins from the University of Manchester, who inspired this thesis. Thanks should also go to Dr. Nuanwan Sanguansak, who gave me the beginnings of my path in astronomy.

I also want to express my gratitude to the lecturers in the School of Physics who have taught me and helped me expand my knowledge and development over the years. A special thanks to our astronomy and astrophysics group members for their interesting discussions and comments. Many thanks to Kittipong Wangnok and Jaruchit Siripak, members of our research group, who always give me their support.

I would like to express my gratitude to Dr. Utane Sawangwit and Asst. Prof. Dr. Tirawut Worrakitpoonpon for sitting on the thesis committee and providing me with valuable advice.

I also would like to acknowledge the funding resources from the Development and Promotion of Science and Technology Talents Project (DPST), Suranaree University of Technology (SUT), National Astronomical Research Institute of Thailand (NARIT), and Spectroscopy and Photometry of Exoplanetary Atmospheres Research Network (SPEARNET).

Last but not least, I want to express my heartfelt appreciation to my parents and my supporter, Tawanchat Simantathammakul, for their sympathy, encouragement, and support during my studies, as well as for inspiring me to finish the thesis.

Patcharawee Munsaket

CONTENTS

	Page
ABSTRACT IN THAI	I
ABSTRACT IN ENGLISH	II
ACKNOWLEDGEMENTS	III
CONTENTS	IV
LIST OF TABLES	VI
LIST OF FIGURES	VII
LIST OF ABBREVIATIONS	X
CHAPTER	
I INTRODUCTION	1
II LITERATURE REVIEW	3
2.1 Exoplanets and their atmosphere	3
2.2 Photometric system	4
2.3 Transmission spectroscopy	4
2.4 Retrieval	7
2.5 Previous random forest regressor in exoplanet atmospheric retrieval	7
III METHODOLOGY	12
3.1 Synthesize transmission spectra	12
3.2 Machine learning method	14
3.2.1 Random forest regressor	14
3.2.2 Tuning hyperparameters	19
3.2.3 Model evaluation	20
3.3 Optimal filter investigation	20
3.3.1 Feature importance	20
3.3.2 Filter selection	21
IV RESULT	22
4.1 Machine learning model	22
4.2 Model efficiency	22
4.3 Filter optimization	23

CONTENTS (Continued)

	Page
V DISCUSSION AND CONCLUSION	31
REFERENCES	33
APPENDIX	37
CURRICULUM VITAE	47



LIST OF TABLES

Table		Page
3.1	The input parameters for generating the spectra using the transit depth calculator module in PLATON. Note that M_p is in Jupiter mass (M_J) and R_p is in earth radius (R_{\oplus}).	13
4.1	The performance of our random forest regressor model is compared to that of the PLATON retrievals with 50 and 100 live points. Note that the models are run on a machine with a Core i9-10900 CPU and 32 GB of RAM that runs Ubuntu 18.04.2 LTS (Munsaket et al., 2021).	24



LIST OF FIGURES

Figure		Page
2.1	Johnson-Cousins U, B, V, R _c and I _c (Granzer, 2014a).	5
2.2	Sloan u', g', r', i' and z' filter curves (Granzer, 2014b).	5
2.3	The transit depth is increased due to the absorption of the exoplanet atmosphere in different wavelengths.	6
2.4	Transmission spectra of four metallicity variation cases (1x, 10x, 100x, and 1000x solar) (Kawashima and Ikoma, 2019).	8
2.5	Transmission spectra of four C/O ratio variation cases (0.5, 1, 10, and 1,000) (Kawashima and Ikoma, 2019).	9
2.6	Transmission spectra of two temperature variation cases (790 K and 1,290 K) (Kawashima and Ikoma, 2019).	10
3.1	The flowchart for retrieving exoplanet atmospheric parameters from the synthetic transmission spectra in the optical wavelength using a random forest regressor and for the filter optimization.	13
3.2	The variation of transmission spectrum due to the variation of temperature.	15
3.3	The variation of transmission spectrum due to the variation of 10-based logarithm of metallicity.	16
3.4	The variation of transmission spectrum due to the variation of carbon-oxygen ratio.	17
3.5	A synthesized transmission spectrum (red) obtained by simulated transmission spectra from PLATON (turquoise) weighted by the transmission profiles of the filters (b) (Munsaket et al., 2021).	18
4.1	The cross-validation score (R ²) obtained by tuning hyperparameters using the GridSearchCV with the 5-fold cross veridations (Munsaket et al., 2021).	23

LIST OF FIGURES (Continued)

Figure		Page
4.2	The real and predicted values of five planetary parameters: (a) planetary mass, (b) planetary radius, (c) planetary atmospheric temperature, (d) metallicity of the planetary atmosphere, and (e) Carbon to Oxygen ratio in the atmosphere, of 20,000 test sets using a random forest regressor (light blue), the PLATON nested sampling retrieval with a number of live points of 50 (blue) and 100 (green). The red dashed lines show the perfect prediction lines (Munsaket et al., 2021).	25
4.3	The feature importance obtained by the random forest regressor model varies with the number of filters ranging from 10 to 2. The white area represents the eliminated filters for each iteration.	26
4.4	The parameters' prediction accuracy changes with the number of filters. The x-axis is the number of filters in each prediction, and the y-axis is the accuracy of the prediction. The blue, orange, green, red, and purple dash-dotted lines are the accuracy of planetary mass, planetary radius, temperature, logarithmic metallicity, and carbon-oxygen ratio, respectively.	27
4.5	The cross-validation score of each random forest regressor model with the different number of filters.	28
4.6	The filter optimization predictions' trend.	29
4.7	The filter transmission curve for each filter combination where the filter that is least importance is eliminated, one by one.	30
1	Real and predicted values of five planetary parameters for 10 filters (U, B, V, R, I, u', g', r', i' and z').	38
2	Real and predicted values of five planetary parameters for 9 filters (U, B, V, I, u', g', r', i' and z').	39
3	Real and predicted values of five planetary parameters for 8 filters (U, B, V, I, u', g', i' and z').	40

LIST OF FIGURES (Continued)

Figure		Page
4	Real and predicted values of five planetary parameters for 7 filters (U, B, V, I, g', i' and z').	41
5	Real and predicted values of five planetary parameters for 6 filters (U, B, V, I, g' and i').	42
6	Real and predicted values of five planetary parameters for 5 filters (U, B, I, g' and i').	43
7	Real and predicted values of five planetary parameters for 4 filters (U, I, g' and i').	44
8	Real and predicted values of five planetary parameters for 3 filters (U, I and g').	45
9	Real and predicted values of five planetary parameters for 2 filters (U and I).	46

LIST OF ABBREVIATIONS

CPU	Central Processing Unit
GB	Gigabyte
LTS	Long-Term Support
MCMC	Markov Chain Monte Carlo
MSE	Mean Squared Error
NARIT	National Astronomical Research Institute of Thailand
PC	Personal Computer
PLATON	PLanetary Atmospheric Transmission for Observer Noobs
RAM	Random Access Memory
SDSS	Sloan Digital Sky Survey
WFC3	Wide Field Camera 3



มหาวิทยาลัยเทคโนโลยีสุรนารี

CHAPTER I

INTRODUCTION

Humanity's suspicion of the existence of extraterrestrial life in the universe has given birth to the study of extrasolar planets. Some people may have heard of these questions: how many stars out of the billions are habitable for life, or what is a suitable environment that supports life? More than 4,500 exoplanets (Exoplanet Exploration: Planets Beyond Our Solar System, 2021) have been confirmed to date due to the desire to find answers to these issues. An exoplanet is a planet that exists outside our solar system and orbits the host star. Up to now, exoplanets have been classified with many criteria. For example, they can be classified into four types which are Gas Giants, Neptune-like, Super-Earth, and Terrestrial.

Most confirmed exoplanets are now discovered by indirect detections since planets are often near their host stars, which are significantly brighter than the planets. For the mentioned reason, direct detection of an exoplanet is very challenging. Astrometry, radial velocity, gravitational microlensing, and transit methods are all indirect detection techniques. The transit method is powerful because we can observe the brightness of a host star decrease when a planet passes in front of it periodically.

The study of planetary atmospheres is a rapidly expanding area in the field of exoplanet research which investigates the characteristics of the weather on these extrasolar planets. The exoplanet atmospheric study is the expanding of knowledge in atmospheric science. Transmission spectroscopy can be used to examine the atmospheres of these exoplanets from both ground-based and space-based observations. Transmission spectroscopy is a typical approach for studying exoplanetary atmospheres by measuring the variation of transit depths with wavelengths that relies on the characteristics of components in the planetary atmosphere (Seager and Sasselov, 2000). The transmission filters are an important tool in the transmission spectroscopy technique.

Atmospheric retrieval techniques can provide information on the exoplanet's atmosphere. Nevertheless, typical retrieval techniques such as MCMC (Foreman-Mackey et al., 2013) and nested sampling (Shaw et al., 2007) are time-

consuming. Hence, machine learning algorithms have been used to retrieve exoplanet atmospheric characteristics to minimize computing time. For example, random forest regression, a supervised machine learning algorithm, has previously been used to estimate the variables of an exoplanet's atmosphere using Hubble Space Telescope Wild Field Camera 3 data (WFC3) (Márquez-Neila et al., 2018). Other than devoting a significant amount of time to retrieval, there are also many options for filter selection, such as the Johnson-Cousins and the Sloan Digital Sky Survey (SDSS) system. All available filters in the observation for transmission spectroscopy are used to collect all possible important information without filter ordering or filter selection. Observations in this research field commonly use one filter per night, which is another time-consuming factor.

The objective of this work is to examine the optimal method for selecting filters suited for investigating the broad-band transmission spectra of hot Jupiters. The transmission spectra of hot Jupiters are simulated by the PLanetary Atmospheric Transmission for Observer Noobs (PLATON) (Zhang et al., 2020). The filter transmissions are then used to weigh the spectra. The atmospheric parameters are predicted using the python Scikit-learn package's random forest regression model (Pedregosa et al., 2011). The results are compared to those produced using standard PLATON's nested sampling retrieval. The filter selection is carried out by considering feature importance, where each feature is each optical filter in the specific waveband. The filter selection presented here will assist observatory planning and minimize the time required for data processing.

CHAPTER II

LITERATURE REVIEW

2.1 Exoplanets and their atmosphere

In our solar system, all planets orbit around the Sun. Planets that orbit around other stars are called extrasolar planets. “A host star” or “a parent star” is the name given to the star in the system. Many planets can orbit a host star in a single system. Finding extraterrestrial life is one goal of exoplanet research. The first exoplanet was confirmed in 1992 in the orbit of the pulsar PSR1257+12 using a radio pulsar timing technique (Wolszczan and Frail, 1992).

There are several techniques for detecting exoplanets. Firstly, radial velocity technique that observes the star wobble around the mutual center of mass of the system due to an effect of the gravitational interaction of planets with their host star. The changing of star radial velocity is detected from the changing of the star’s spectral lines as an effect of the Doppler phenomena. Secondly, the direct imaging method can be use to observe the reflecting light of an exoplanet from the host star during its orbit. Furthermore, the astrometry method is used to observe a star’s position changing over time due to the gravitational interaction between the host star and the exoplanet that makes both of them orbit around the center of mass. This method satisfies the face-on system plane, the massive planet or the planet that orbits a low-mass star such as a brown dwarf. Another method is a gravitational microlensing method which is based on the gravitational lens effect. It occurs when the lensing star passes in front of the source star and causes the light trajectory that passes through the gravitational field of the lensing star to be curved by gravity. The last and most famous method is the transit method that observes the periodic changing of the starlight caused by a planet passing in front of the host star. This method is the main idea of the transmission spectroscopy technique which is used in this work (see 2.3).

In 2001, the first exoplanet atmosphere was discovered (Charbonneau et al., 2002). Exoplanet atmosphere research is a branch of astronomy that aims to learn

more about the nature of exoplanets. The investigation of exoplanet atmospheres is part of the search for suitable conditions for the existence of life beyond the solar system. Most atmosphere detections at the moment are of hot Jupiters or hot Neptunes that orbit very close to their host star and therefore have warm and extended atmospheres (Madhusudhan et al., 2014). Transmission spectroscopy is one of the exoplanet atmosphere study methods. This method detects the light that travels through an exoplanet's atmosphere when it transits in front of its host star.

2.2 Photometric system

The light from the star emits in a broad range of wavelengths depending on their characteristics and temperature. To study the atmosphere of an exoplanet via transmission spectroscopy, the broad wavelength has to be truncated into the narrow wavelength. Specific filters are used to restrict the light that transmits through the filters. The Johnson-Cousins system and the Sloan system are the optical filters that we used in this research. The Johnson-Cousins filters consist of U, B, V, R_c , and I_c filter. The Sloan filters consist of u' , g' , r' , i' and z' filters. The different filters have different transmission properties. The transmission properties of the filters are represented by the filter profile. The filter profile of the Johnson-Cousins and Sloan transmission profiles represent in figure 2.1 and 2.2.

2.3 Transmission spectroscopy

Transits of exoplanets provide a chance to investigate the composition of their atmospheres from the light that passes through them. The hot Jupiter exoplanets are the most accomplished atmospheric characterizations (Sing et al., 2016). However, the atmosphere of the Neptune-size and the super-Earth size planets were also studied (Fraine et al., 2014; Kreidberg et al., 2014). Transmission spectroscopy is one method that is widely used in exoplanet atmosphere studies. This method measures the change in the transit depth of a light curve as a function of wavelength. The atmosphere of an exoplanet filters the light from the host star during the transit. Because of the absorption characteristics of the atmosphere's composition, a specific portion of starlight is absorbed. At a wavelength with more significant absorption characteristics, the observed planet disk will appear larger due to the planet's at-

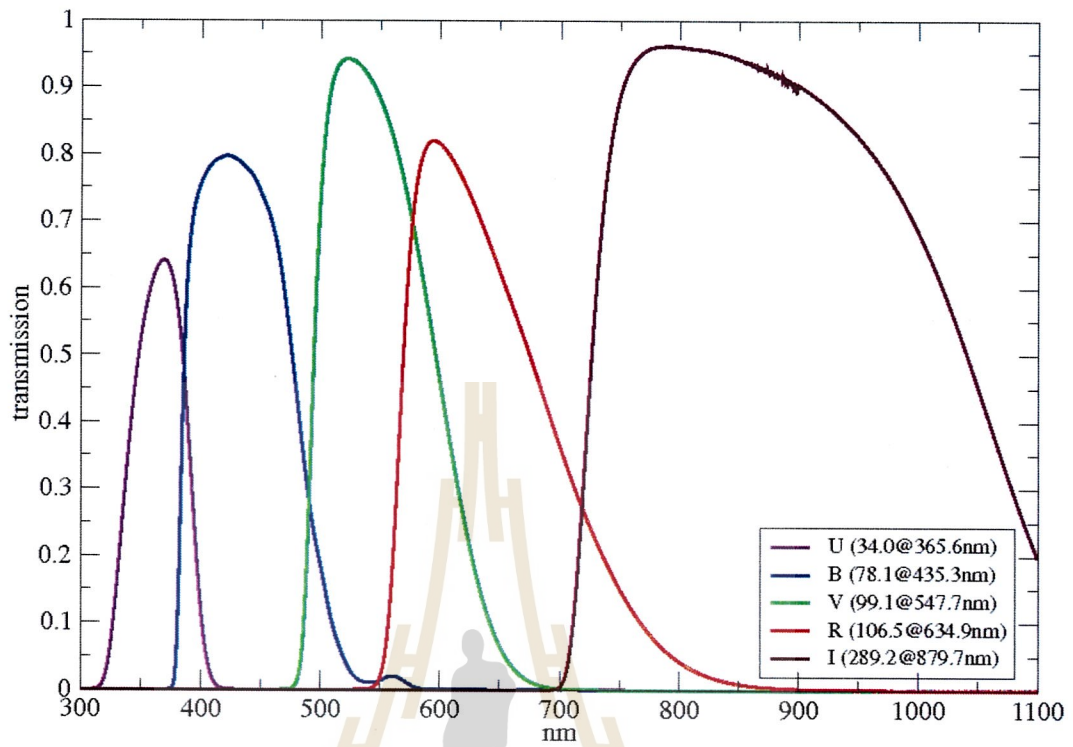


Figure 2.1 Johnson-Cousins U, B, V, R_c and I_c (Granzer, 2014a).

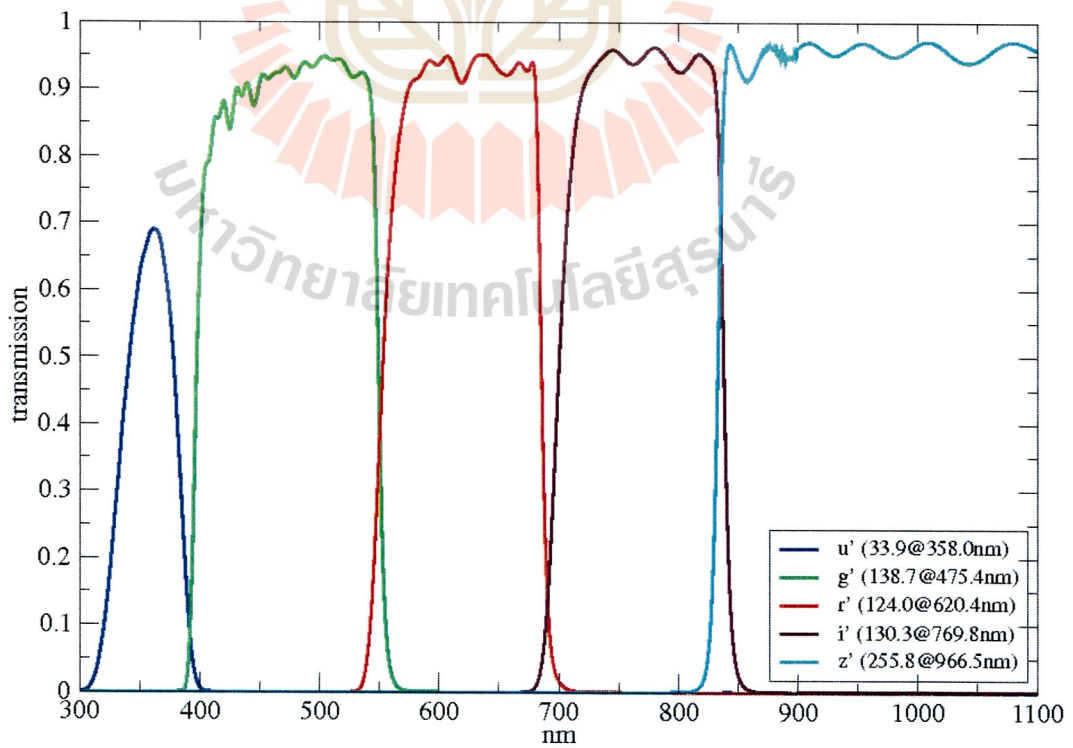


Figure 2.2 Sloan u' , g' , r' , i' and z' filter curves (Granzer, 2014b).

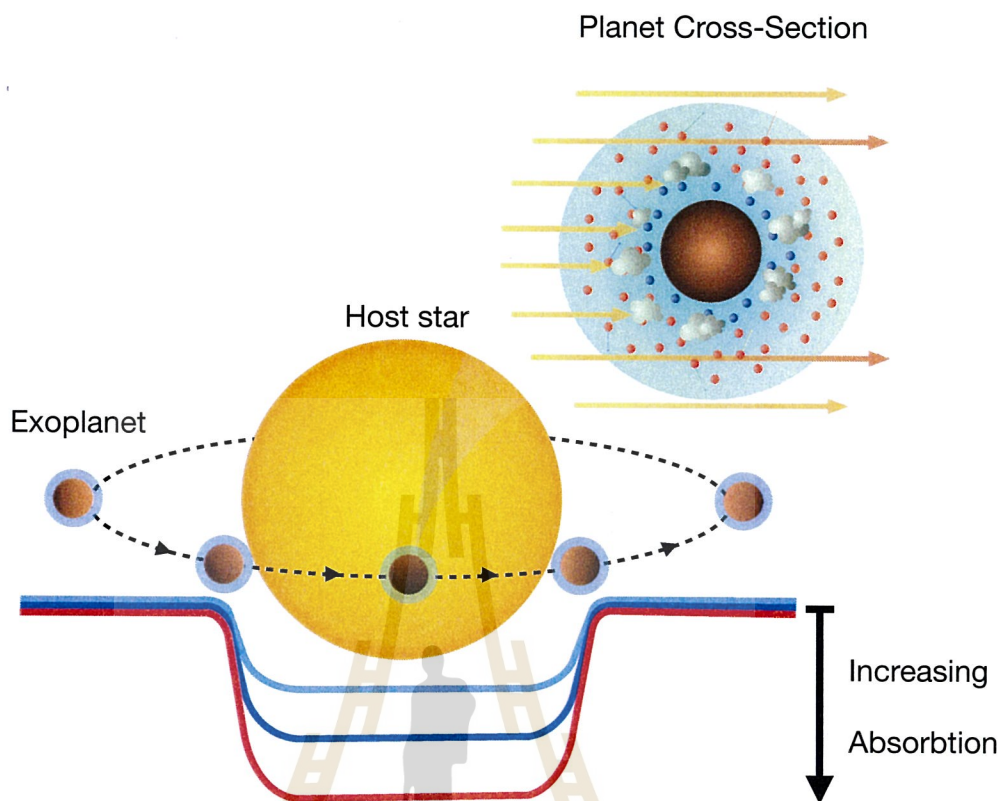


Figure 2.3 The transit depth is increased due to the absorption of the exoplanet atmosphere in different wavelengths.

atmosphere being more opaque from the absorption of the molecules or elements. As a result of the higher absorption, the transit light curve will be deeper (figure 2.3). Using many broad band filters, spectral features in the planetary spectrum can be detected by transmission spectroscopy. Therefore, the atmosphere's composition can be inferred.

The transmission spectra and the atmospheric properties of exoplanets depend on model parameters such as planetary mass, planetary radius, atmospheric temperature, metallicity, and carbon-to-oxygen ratio. The metallicity and the carbon-to-oxygen ratio are model parameters that we use in the transmission spectra simulation by PLATON. The sensitivity of metallicity, C/O ratio, and atmospheric temperature in the transmission spectra are reported in the literature (Kawashima and Ikoma, 2019). In the case of the sensitivity to metallicity, an increase in metallicity leads to a decrease in the atmospheric scale height, which is represented by the

transit depths (figure 2.4). This is because the hydrocarbons' photodissociation rate and haze monomer production rate decrease at high metallicity. Note that the photodissociation rate is a chemical process in which photons break down a chemical molecule and the haze monomer production rate. At high metallicity, the haze effect is pronounceable only at a short wavelength because the mass density of the haze particles decrease. For the clear atmosphere at increasing metallicity, the absorption strength is decreasing due to a decrease in the atmospheric scale height.

For C/O ratio sensitivity as shown in figure 2.5, the transit depth increases with increasing C/O. The higher C/O means the monomer production rate, the mass density, and the average size of haze particles are larger. Although the C/O is extremely high (i.e. $C/O = 10^{10}$), the photodissociation rate of hydrocarbons is not limited by the amount of carbon but depends on the amount of incoming photon flux. The temperature-dependent transmission spectra are shown in figure 2.6. In any case of haze, the higher temperature provides a higher transit depth and the spectral features are more pronounced because of the larger atmospheric scale height.

2.4 Retrieval

A spectrum stores information about the atmosphere's various interconnected physicochemical processes and properties, which are discovered by their effect on the radiation that passes through the atmosphere before reaching the observer. Chemical composition, temperature structure, atmospheric circulation, and clouds/hazes are all properties that leave their fingerprints on the spectrum. Exoplanetary atmospheric retrieval is the process of inferring an exoplanet's atmospheric parameters from its observed spectrum. The returned properties can be used to look deeply into the physical and chemical processes in the atmosphere, as well as their formation history.

2.5 Previous random forest regressor in exoplanet atmospheric retrieval

Nowadays, machine learning has become one of the research tools in exoplanet study and also in the study of the exoplanetary atmospheric spectra. For

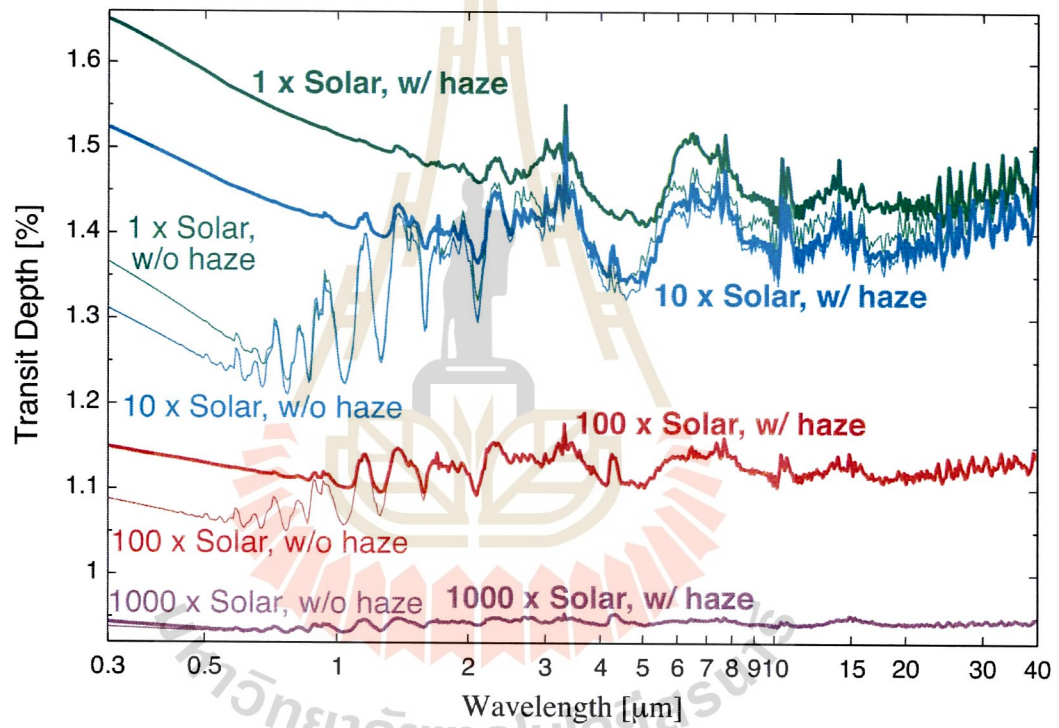


Figure 2.4 Transmission spectra of four metallicity variation cases (1x, 10x, 100x, and 1000x solar) (Kawashima and Ikoma, 2019).

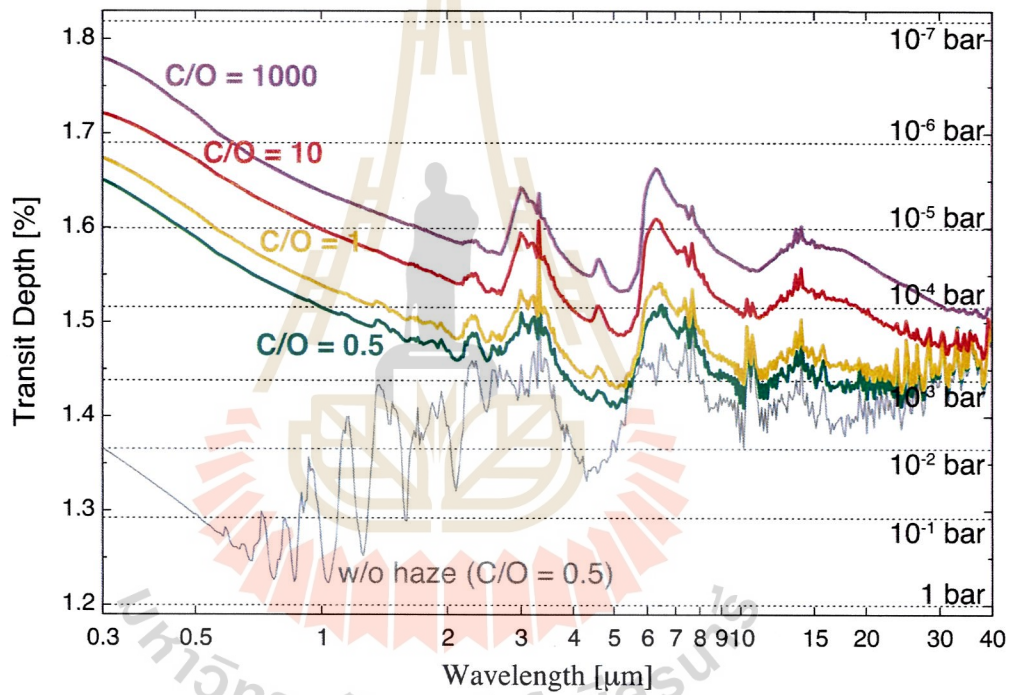


Figure 2.5 Transmission spectra of four C/O ratio variation cases (0.5, 1, 10, and 1,000) (Kawashima and Ikoma, 2019).

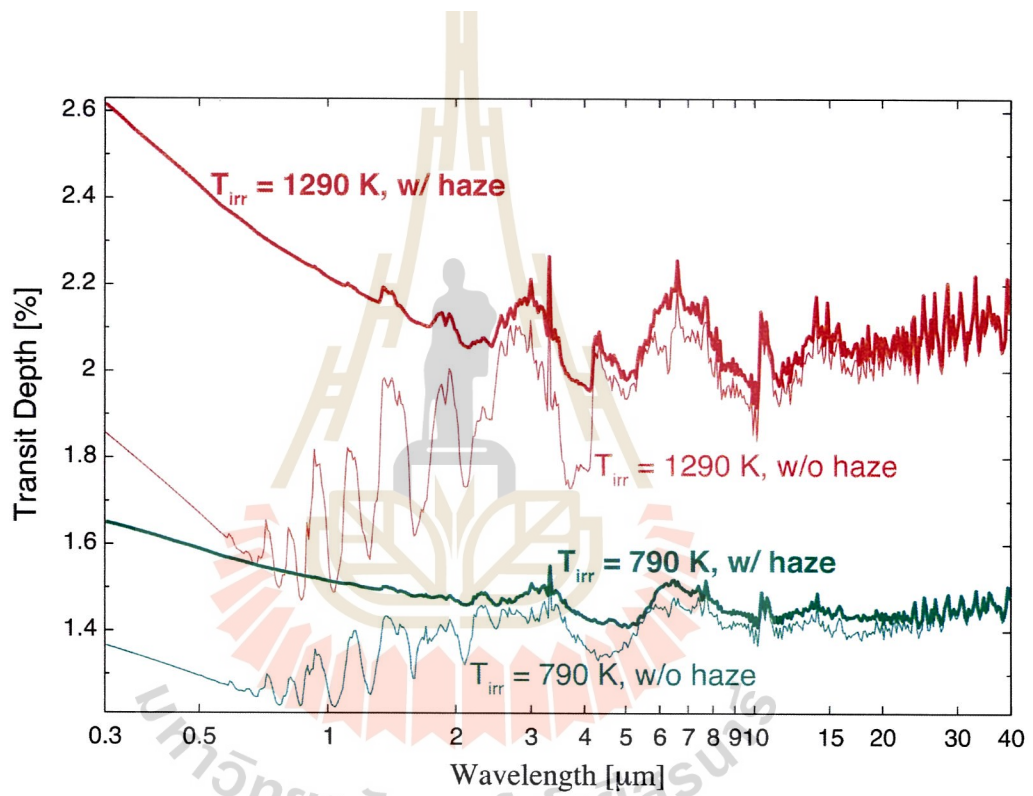
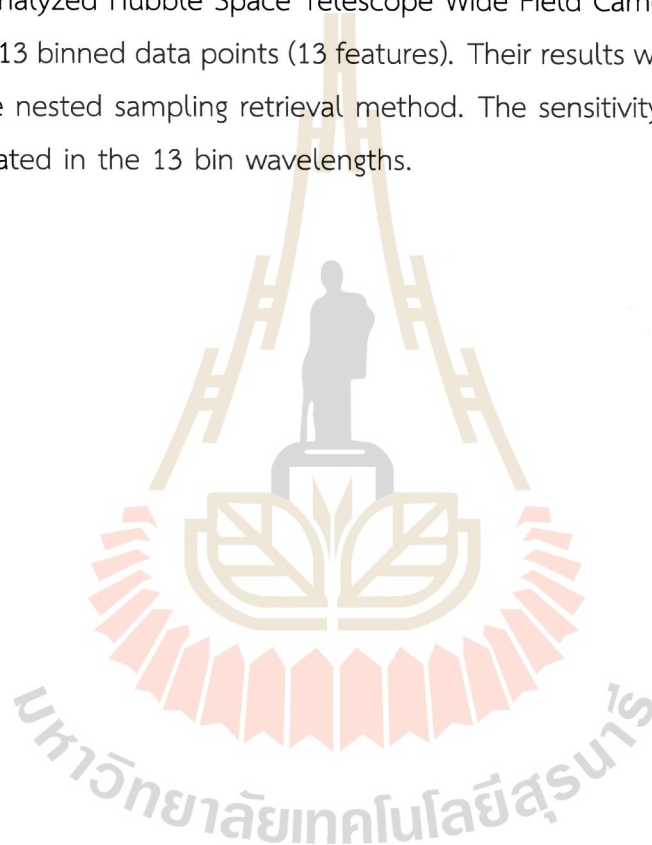


Figure 2.6 Transmission spectra of two temperature variation cases (790 K and 1,290 K) (Kawashima and Ikoma, 2019).

example Márquez-Neila et al. (2018) used the random forest regressor by training the model with a pre-compute grid of atmospheric models and predicting the abundance of the molecules and the cloud opacity. The model is applied to the spectrum of WASP-12b. The transmission spectra of WASP-12b are generated by randomly selecting the values of 5 parameters from the parameter ranges which consist of temperature, cloud opacity, and the volume mixing ratios or relative abundance of water, ammonia, and hydrogen cyanide. These transmission spectra were obtained from the previously analyzed Hubble Space Telescope Wide Field Camera 3 (WFC3) and were binned into 13 binned data points (13 features). Their results were consistent with the result of the nested sampling retrieval method. The sensitivity of model parameters was investigated in the 13 bin wavelengths.



CHAPTER III

METHODOLOGY

In this research, we optimize the photometric observation's filter transmission for the hot Jupiter atmosphere using the transmission spectroscopy technique. Firstly, the 100,000 transmission spectra are simulated using the PLATON (Zhang et al., 2020) where the parameters cover the range of the hot Jovian planet. The synthetic photometric observations are then created by binning and weighting spectra that include noise, explained in section 3.1. In section 3.2, we explain the atmospheric parameters are estimated using a machine learning model from Scikit-learn (Pedregosa et al., 2011). Section 3.3 describes the optimal filter investigation where the feature importance of the machine learning model is used to examine the optimal filter set. The calculations are carried out on a machine with a Core i9-10900 CPU and 32 GB of RAM that runs Ubuntu 18.04.2 LTS at National Astronomical Research Institute of Thailand (Public Organization) (NARIT). The flowchart explaining overall methods is represented in figure 3.1.

3.1 Synthesize transmission spectra

This thesis mainly focuses on hot Jupiters, gas giant exoplanets with extremely short orbital periods of $P < 10$ days (Gaudi et al., 2005). Planet mass (M_p), planet radius (R_p), planet atmosphere temperature (T), the metallicity of the planetary atmosphere ($\log Z$), and atmospheric carbon to oxygen ratio (C/O) are the key parameters studied here. Using the PLATON forward model, 100,000 hot Jupiter transmission spectra are generated by uniformly, randomly selecting the values of these parameters within the range specified in table 3.1. The chosen values for planetary mass and radius are within 68 and 95 percent confidence ranges of the mass-radius relationship of Jovian exoplanets, respectively (Chen and Kipping, 2016). The planets are thought to revolve around a solar-analog star. The planetary atmospheric temperature investigated here covers the temperature range of the hot Jovian planet (Madhusudhan et al., 2014). The values of metallicity and the carbon-oxygen

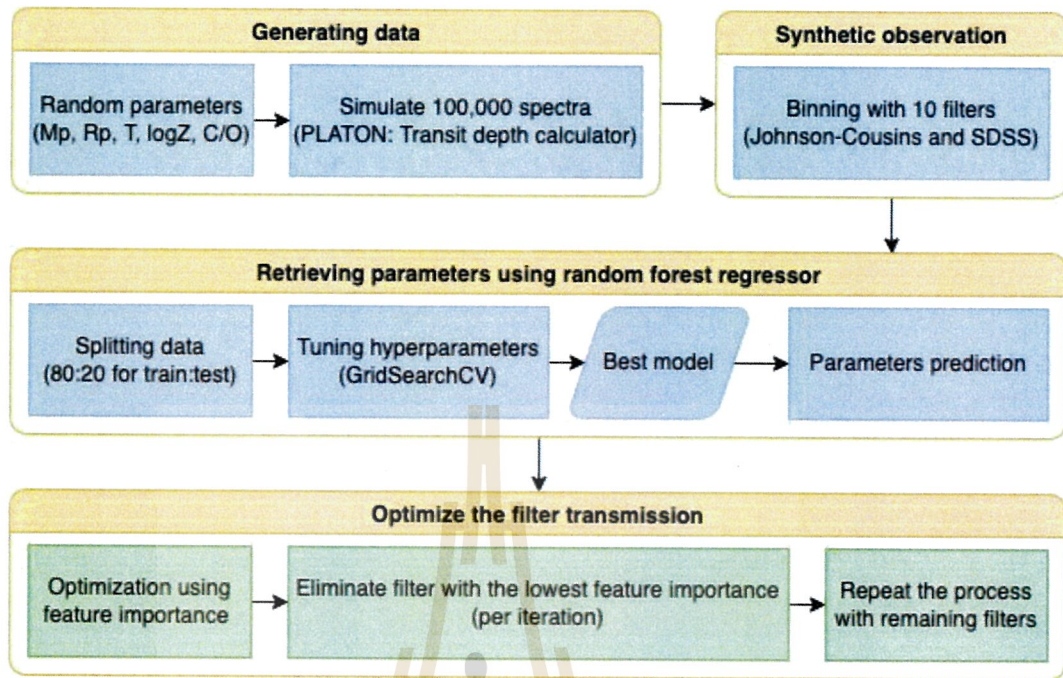


Figure 3.1 The flowchart for retrieving exoplanet atmospheric parameters from the synthetic transmission spectra in the optical wavelength using a random forest regressor and for the filter optimization.

ratio of the atmosphere used here cover all possible values of those parameters in PLATON (Zhang et al., 2020). The test of the transmission spectra variation due to the parameters changing is also produced by PLATON. The changes of transmission spectra by the variation of temperature, the 10-based logarithm of metallicity, and carbon-oxygen ratio are shown as an example, in figures 3.2, 3.3, and 3.4 respectively.

Table 3.1 The input parameters for generating the spectra using the transit depth calculator module in PLATON. Note that M_p is in Jupiter mass (M_J) and R_p is in earth radius (R_\oplus).

Parameters	Value	References
Planetary mass (M_p)	$0.3M_J - 3M_J$	Stevens and Gaudi (2013); Chen and Kipping (2016)
Planetary radius (R_p)	$8R_\oplus - 20R_\oplus$	Borucki et al. (2011); Chen and Kipping (2016)
Planetary atmospheric temperature (T)	1,300 – 3,000K	Madhusudhan et al. (2014)
Metallicity of planetary atmosphere ($\log Z$)	-1 – 3	Zhang et al. (2020)
Atmospheric Carbon-Oxygen ratio (C/O)	0.35 – 1	Zhang et al. (2020)

Several transit observations have been made from the ground using broad-band optical filters. The optical filters chosen in this study are Johnson-Cousins (U, B, V, R, I) and Sloan (u' , g' , r' , i' , z'). To synthesize transmission spectra of optical observations, the transmission profiles of the filters are utilized to weight the transmission spectra simulated using PLATON, as presented in figure 3.5. The Johnson-Cousins transmission curves are acquired using the PyAstronomy package's photometric transmission curves class (Czesla et al., 2019). Sloan Digital Sky Survey Data Release 7 (SDSS Data Release 7: <http://classic.sdss.org/dr7/instruments/imager/#filters>) provides the Sloan filter profile. The weighted transmission curves represent the transit depths of the relevant filters.

3.2 Machine learning method

The science of training machines to learn is known as machine learning. In the learning problem, the set of n samples of data is considered for the prediction of the properties of unknown data. If each sample contains the multiple input data, it is said to have several attributes or features. Learning problems are divided into two categories: supervised and unsupervised learning. Supervised learning refers to data that contains additional properties that we want to predict. The problem of unsupervised learning occurs when training data consists of a set of input data with no corresponding target values.

3.2.1 Random forest regressor

A Random Forest is a supervised learning method that can be used for classification and regression. It works by training many decision trees. The output of the individual trees for the classification and the regression are the class (discrete value) and predicted value (continuous value), respectively. The random forest regression method is applied in this research to retrieve the parameters from transmission spectra. The spectra from 100,000 synthetic transmission spectra of 10 Johnson-Cousins and Sloan filters were divided into two groups: 80,000 for training and 20,000 for testing. The model features are ten binned transit depths, whereas the labels are five planetary parameters (M_p , R_p , T , $\log Z$, and C/O). To decide which attribute value would result in the "best split", the decision tree regression method examines all

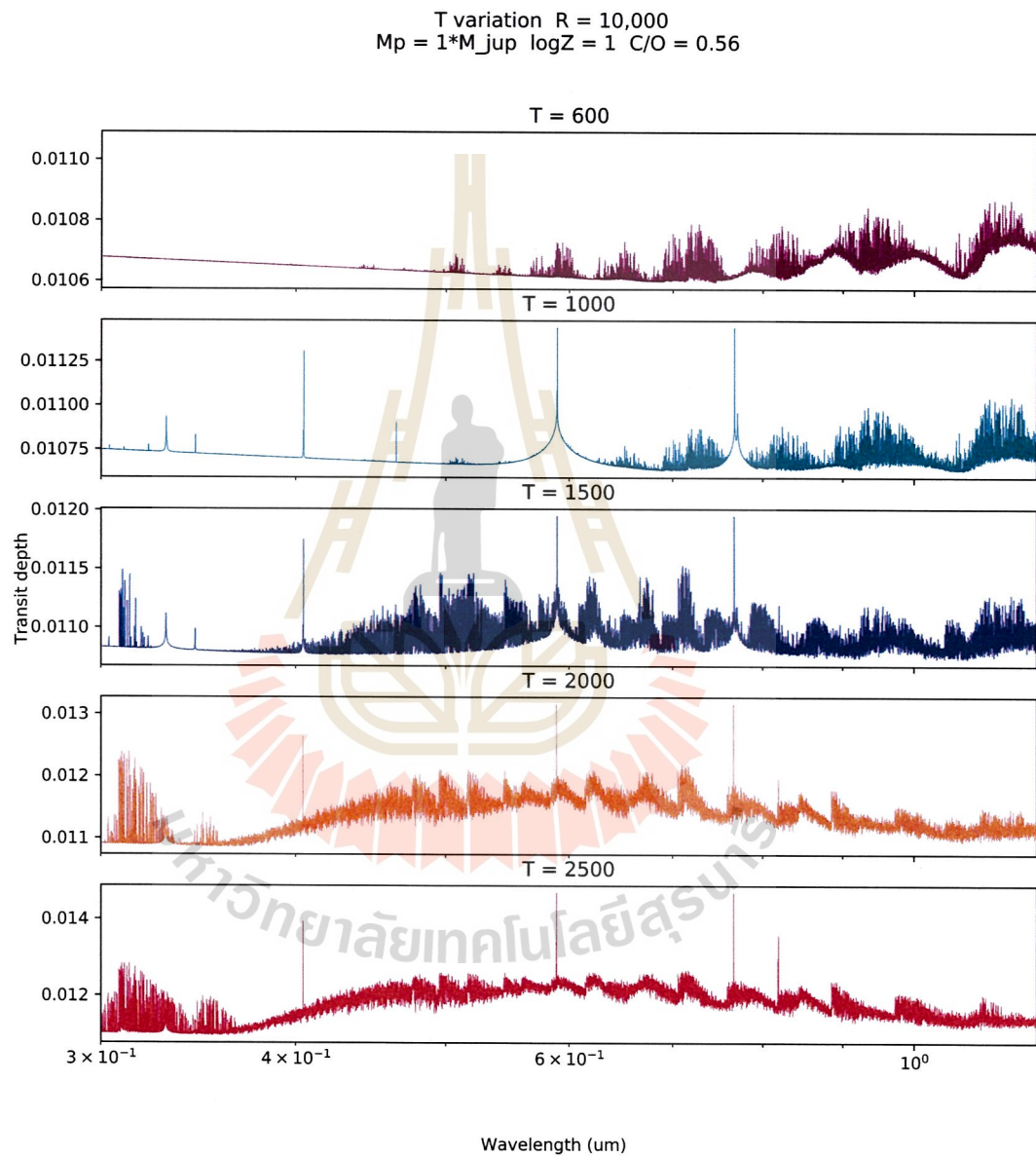


Figure 3.2 The variation of transmission spectrum due to the variation of temperature.

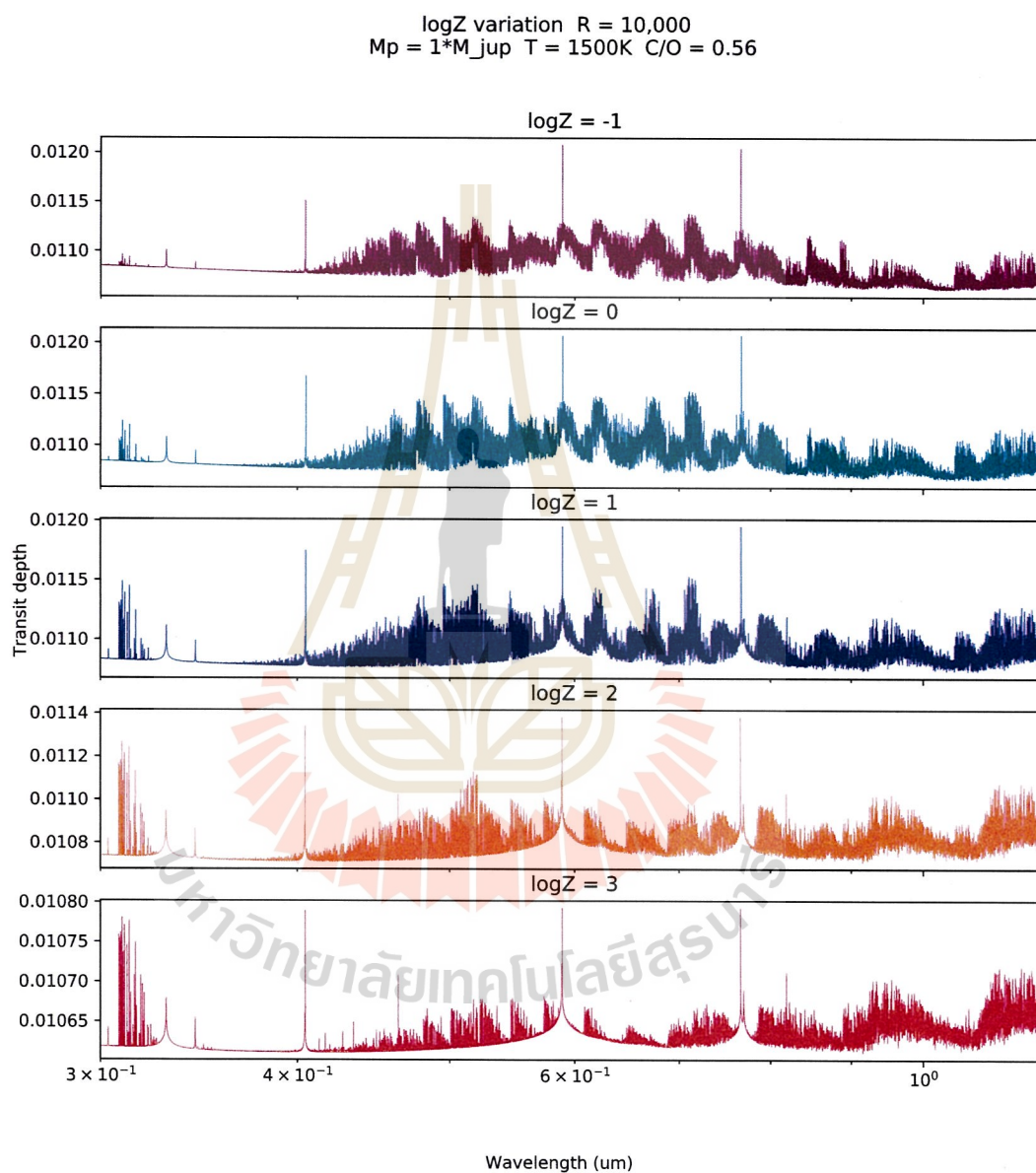


Figure 3.3 The variation of transmission spectrum due to the variation of 10-based logarithm of metallicity.

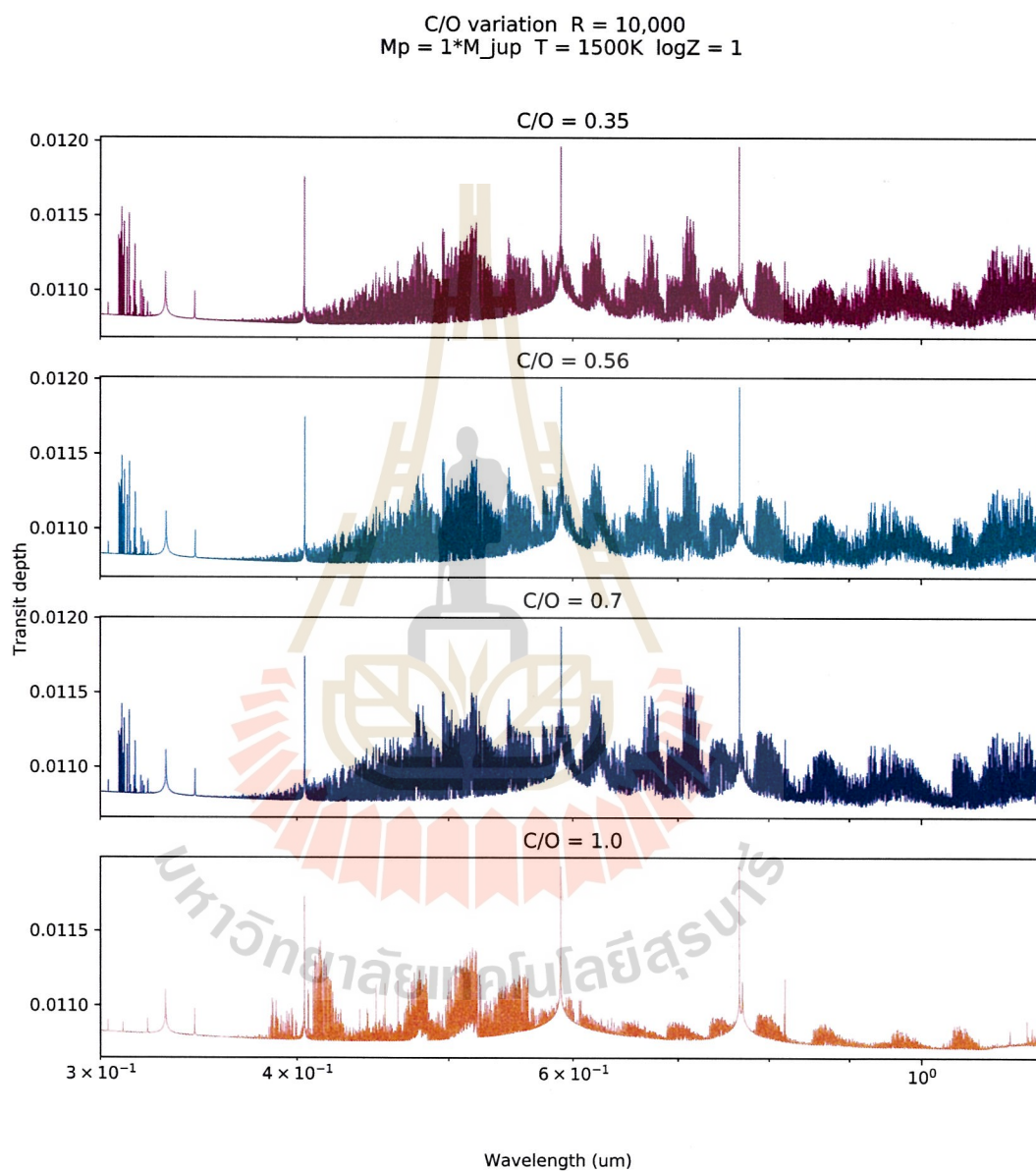


Figure 3.4 The variation of transmission spectrum due to the variation of carbon-oxygen ratio.

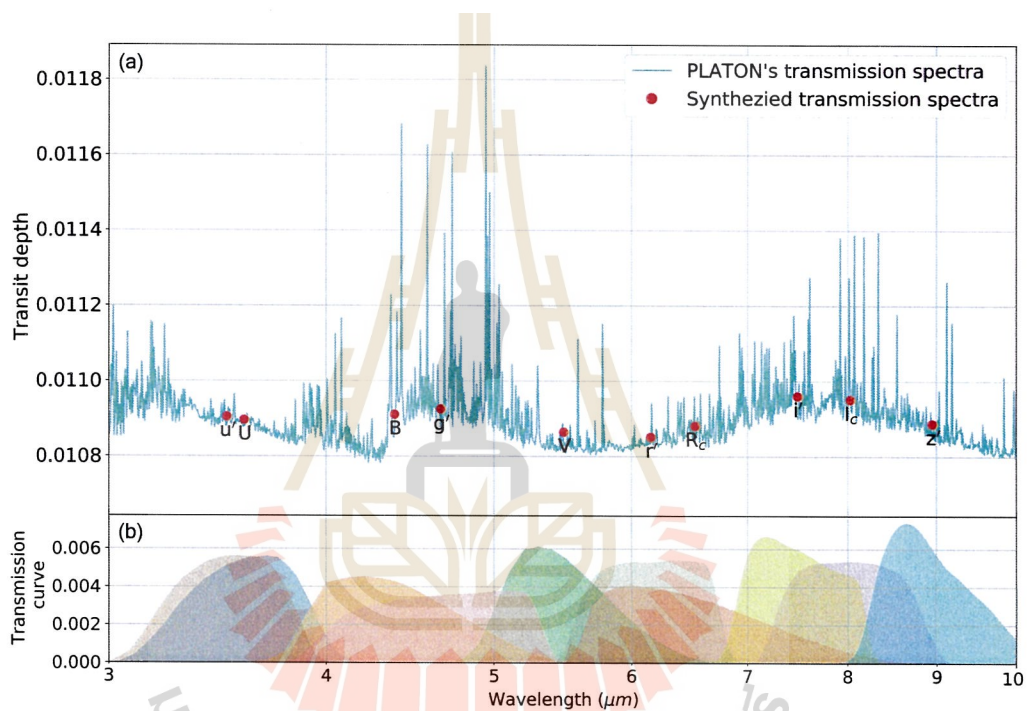


Figure 3.5 A synthesized transmission spectrum (red) obtained by simulated transmission spectra from PLATON (turquoise) weighted by the transmission profiles of the filters (b) (Munsaket et al., 2021).

attributes and their values. In the case of regression problems, the algorithm considers the mean squared error (MSE) as the objective or cost function that should be minimized. The MSE over the n samples is calculated as follows

$$\text{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2, \quad (3.1)$$

where \hat{y}_i is the predicted value of the i^{th} sample and y_i is the corresponding true value.

3.2.2 Tuning hyperparameters

Hyperparameters are the parameters that cannot be directly learnt by the machine. The key parameters for random forest regressor are the number of trees (number of estimators) and the maximum depth of each tree. Both hyperparameters are tuned using the GridSearchCV module in Scikit-learn (Pedregosa et al., 2011). When training the model, cross-validation is performed through the use of GridSearchCV. Note that we divide the data into two parts before training the model, which are training data (80%) and test data (20%). The procedure of cross-validation is to split the training data further into two parts: the training data and the validation data. In our research, we use K-fold Cross-validation. It is an iterative process that partitions the training data into K parts. During each iteration, one part is kept for testing (validation), while the remaining $K-1$ partitions are used to train the model. In the next iteration, the next part will be used as test data, and the remaining $K-1$ will be used as train data, and so on. It records the model's performance in each iteration and provides the average of the performance at the end. This is a time-consuming process, so we use the default 5-fold ($K=5$) cross-validation of the GridSearchCV module in Scikit-learn. The best random forest regressor model obtained from the hyperparameter tuning is applied to predict the parameters for the test data set. Then, the predicted results and the computational time required are compared to the results of the standard PLATON retrieval. Note that both random forest regressor and PLATON are carried out on a machine with a Core i9-10900 CPU and 32 GB of RAM that runs Ubuntu 18.04.2 LTS at NARIT.

3.2.3 Model evaluation

To evaluate the random forest regressor model, we use the coefficient of determination (R^2) which indicates the goodness of the machine learning model's prediction. The model with the highest R^2 is our best model. The estimated R^2 defines as

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.2)$$

where \hat{y}_i is the predicted value of the i^{th} sample, y_i is the corresponding true value for total n samples and

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (3.3)$$

Note that $R^2 = 1$ means all predicted values equal to the real values. $R^2 = 0$ means the predicted values are all the same as the average values (\bar{y}). Negative R^2 occurs when the average value (\bar{y}) explain the data better than the model.

3.3 Optimal filter investigation

3.3.1 Feature importance

From ten optical filters including Johnson-Cousins (U, B, V, R, I) and Sloan (u' , g' , r' , i' , z'), we do not know how each filter affects the prediction of random forest regressor. Therefore, the feature importance is applied to investigate the filter set suitable for the hot Jupiter atmospheric study. The importance of the feature i (f_i) in Scikit-learn is calculated as

$$f_i = \frac{\sum_j n_{ij}}{\sum_k n_{ik}}, \quad (3.4)$$

where j is the node that splits on feature i and n_{ij} is the importance of a node j in a single decision tree. The index k is all nodes. The n_{ij} is calculated by

$$n_{ij} = w_j C_j - w_{\text{left}(j)} C_{\text{left}(j)} - w_{\text{right}(j)} C_{\text{right}(j)}. \quad (3.5)$$

The w_j is the weighted number of samples reaching node j , C_j is the impurity value of node j . The subscript left and right indicate the parameter of child node from left

and right split on node j , respectively. Note that the child node is a possible split node from the root node. If it cannot be further split, this node becomes a leaf node. To normalize the importance of the feature i to a value between 0 and 1, the equation 3.4 is divided by the sum of all feature (l) importance values,

$$\|f_i\| = \frac{f_i}{\sum_l f_l}. \quad (3.6)$$

Finally, the feature importance of a random forest model (RF_i) is calculated by the sum of the feature's importance value on each tree divided by the total number of trees (T),

$$RF_i = \frac{\sum_m \|f_{i,j}\|}{T}, \quad (3.7)$$

where $\|f_{i,j}\|$ is the normalized feature importance for feature i in tree j .

3.3.2 Filter selection

The transmission data binned into the optical filters are considered as the features of the machine learning model. As mentioned earlier, there are ten filters in total from Johnson-Cousins and Sloan filter sets. In this research, we use the feature selection process to reduce the number of filters based on their importance. The process is done by calculating the significant impact of all filters on the machine learning model's prediction. For the random forest regression model, the importance of the features is used to select the filters. Higher score indicates that the considering feature is more significant compared to the one providing lower score. The process of the first iteration starts with tuning the hyper-parameters for ten filters and training the machine to obtain the best model, then predicts the labels with the best model and finds the importance of ten features. The feature importance package from Skit-learn (Pedregosa et al., 2011) used here performs the calculations as shown in equation 3.4. The filter with the lowest importance is eliminated for the next iteration, then the process will repeat for the remained filters, and so on until the last two, most important, filters left. Then, the rank of important filters for observations can be obtained by considering the order of their importance.

CHAPTER IV

RESULT

4.1 Machine learning model

The hyperparameter tuning is performed using the GridSearchCV package in Scikit-learn. Figure 4.1 shows the accuracy (R^2) of predicted results for various numbers of estimators and maximum depths of the tree, which are represented in the x and y axes, respectively. The lighter color represents the higher accuracy of the corresponding machine learning model. For the maximum depth of 320 and the number of estimators of 2,560, the best cross-validation score is obtained ($R^2 = 0.5949$). Therefore, the planetary parameters from 20,000 spectra in our test data set are predicted by the random forest regressor model with these best hyperparameter values.

In figure 4.2, the comparisons between real and predicted values of five parameters from a random forest model are represented in light blue points. The red dashed lines represent the case that real values equal predicted values. We also plotted the retrieved results of nested sampling in PLATON with the number of live points 50 and 100, represented by blue and green points, respectively, to compare with the random forest regressor prediction. The random forest regressor model predicts R_p with the highest accuracy ($R^2 = 0.999$), and M_p , T , and $\log Z$ with intermediate accuracy ($R^2 = 0.742$, 0.571 , and 0.638 , respectively). However, our model is unable to predict the C/O ratio because the transmission features caused by the C/O variation are not significant enough in the optical bands.

4.2 Model efficiency

To consider the performance of our random forest regressor model, the results from our model are compared to those obtained from the nested sampling fitting model in the standard PLATON. A sample of 100 spectra is simulated and retrieved with the PLATON. The number of live points in the nested sampling model was selected to be 50 and 100. All models are tested on a PC with a Core i9-10900

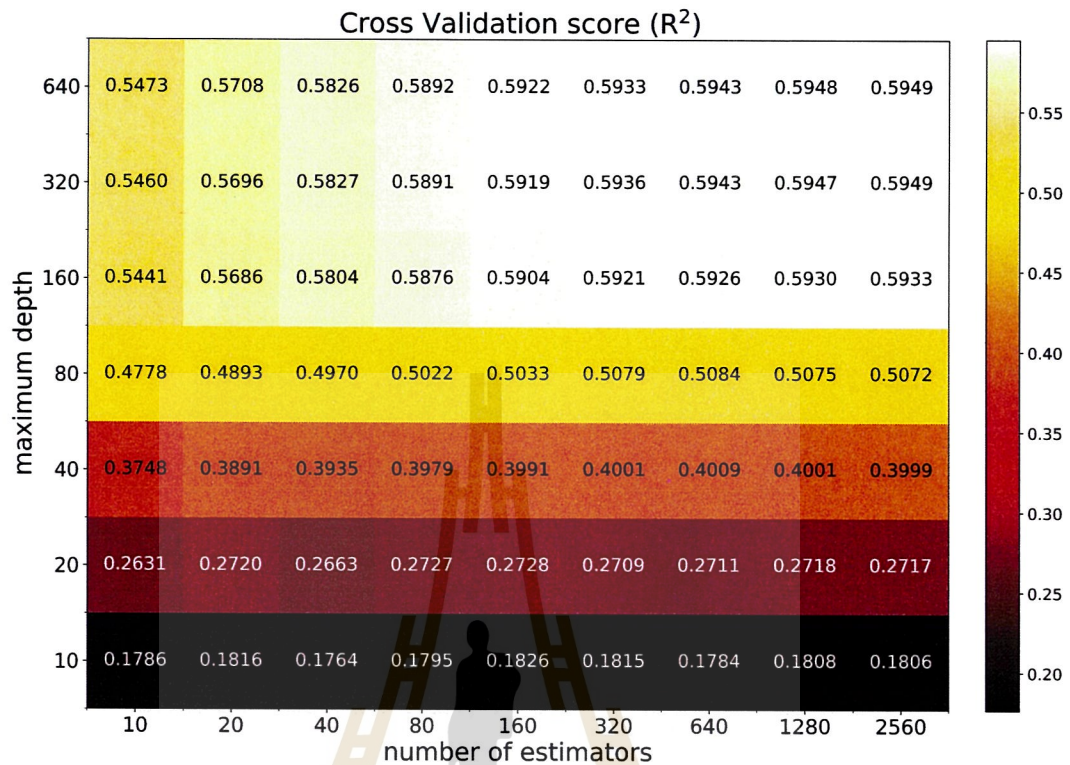


Figure 4.1 The cross-validation score (R^2) obtained by tuning hyperparameters using the GridSearchCV with the 5-fold cross validations (Munsaket et al., 2021).

CPU and 32 GB of RAM running Ubuntu 18.04.2 LTS to compare computational time. The results are shown in the table 4.1. The random forest regressor outperforms the PLATON nested sampling model with 160,000 and 270,000 times faster for 50 and 100 live points, respectively. Accordingly, the accuracy obtained by the regressor model is comparable to that of the PLATON retrievals.

4.3 Filter optimization

We use the feature importance to arrange the filters by their importance values. From the total number of filters, the process starts by tuning the hyperparameters using GridSearchCV, making prediction by using the tuned model and finding the importance of features (filters), eliminating the filter with the lowest importance for the next iteration, and the remaining filters will be used in the next iteration. The

Table 4.1 The performance of our random forest regressor model is compared to that of the PLATON retrievals with 50 and 100 live points. Note that the models are run on a machine with a Core i9-10900 CPU and 32 GB of RAM that runs Ubuntu 18.04.2 LTS (Munsaket et al., 2021).

Approach	Data sets	Time spent (Hr:Min:Sec)	Coefficient of determination (R_{Fit}^2)				
			M_p	R_p	T	log Z	C/O
Random forest regressor	20,000	0:0:14	0.742	0.999	0.571	0.638	0.129
PLATON (nlive=50)	100	3:12:8	0.780	0.999	0.546	0.659	-0.842
PLATON (nlive=100)	100	5:16:45	0.791	0.999	0.616	0.591	-0.604

process is repeated until only two filters remain. The result of this step is shown in figure 4.3. The x-axis represents the number of filters from the first (10 filters) to the last iteration (2 filters). The y-axis represents the filters, starting from the first eliminated filter and so on. The empty area represents the eliminated filter in each iteration. The colors correspond to the value of the importance, where the darker color represents the lower feature importance. The Johnson-Cousins U and I are the two remaining in the final, which are the filters of the highest importance.

To investigate the optimal number of filters, we considered the feature importance changing with the number of filters. In figure 4.4, the data points at each number of filters are the according to form predictions using the best model of the same filters' number in the figure 4.3. The prediction accuracy of the planetary radius gives the highest value, which is consistent with the figure 4.2. For other parameters, the accuracy is almost flat between 10 and 7 filters. After that, the accuracy significantly decreases. The accuracy in the prediction of temperature and carbon-oxygen ratio was negative at two filters. We also check the GridSearchCV best score of each random forest regressor model, as shown in figure 4.5. The plot of the cross-validation score appears to be flat between 10 and 7 filters, consistent with the figure 4.4. Figure 4.6 shows the prediction trends of the filter optimization, which represents the performance trend of each filter combination. Figure 4.7 shows the filter transmission profiles of the filter combinations obtained from the filter optimization using the lowest feature importance elimination.

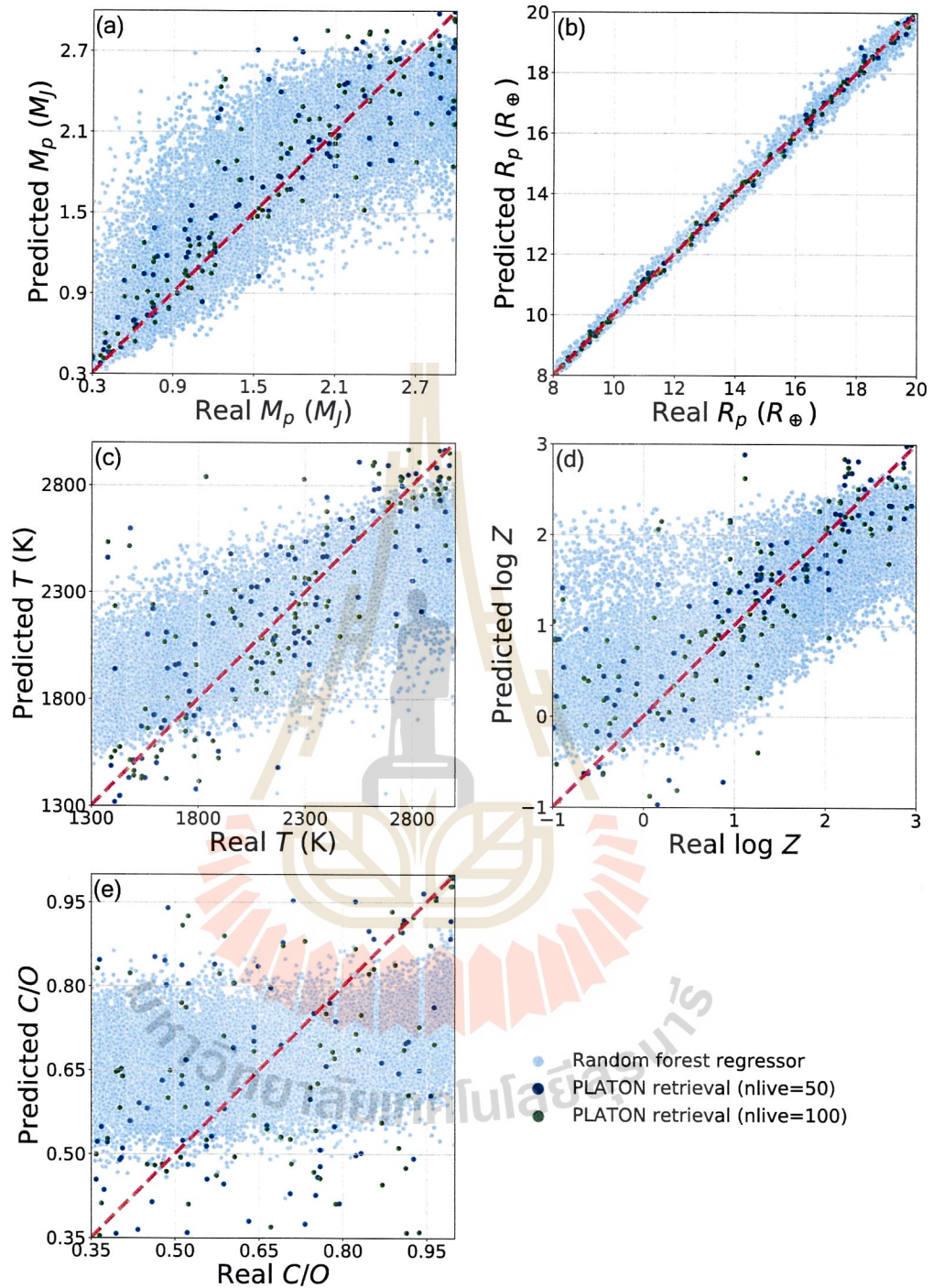


Figure 4.2 The real and predicted values of five planetary parameters: (a) planetary mass, (b) planetary radius, (c) planetary atmospheric temperature, (d) metallicity of the planetary atmosphere, and (e) Carbon to Oxygen ratio in the atmosphere, of 20,000 test sets using a random forest regressor (light blue), the PLATON nested sampling retrieval with a number of live points of 50 (blue) and 100 (green). The red dashed lines show the perfect prediction lines (Munsaket et al., 2021).

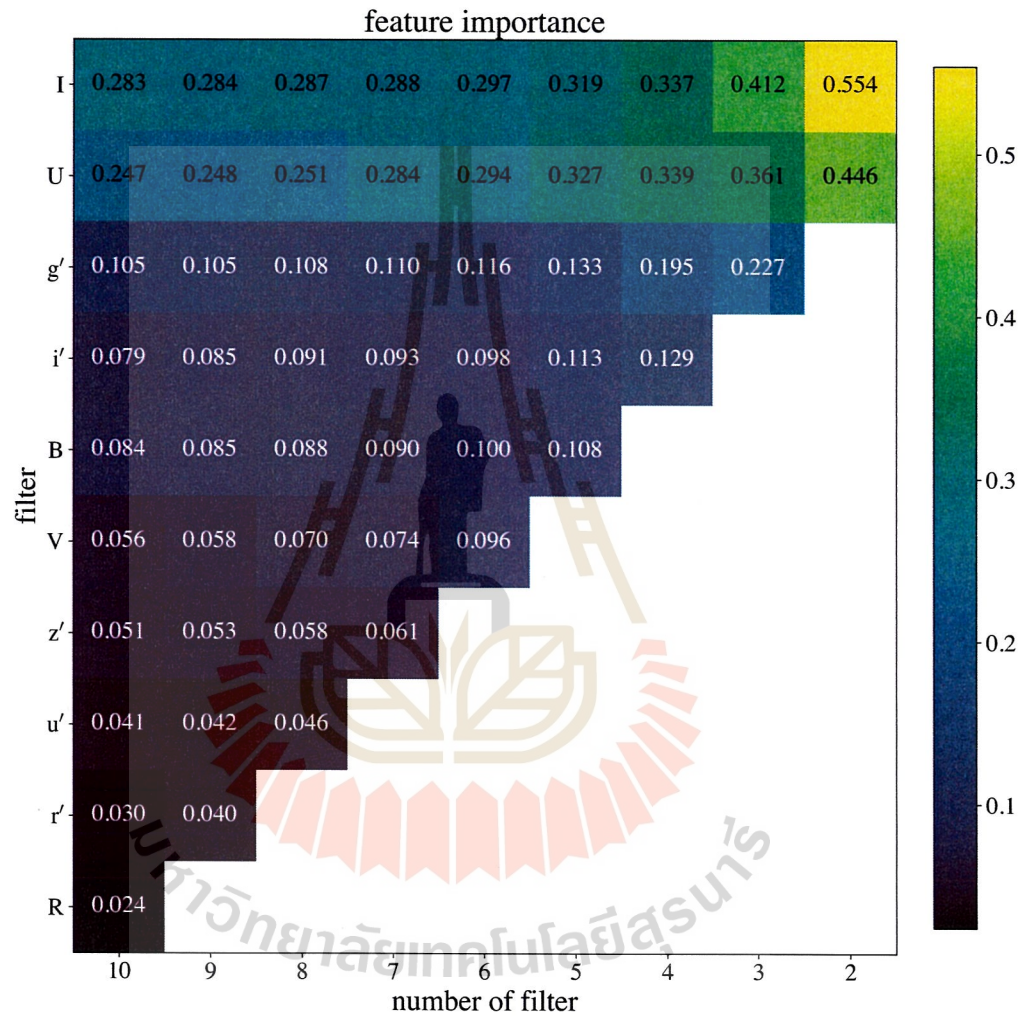


Figure 4.3 The feature importance obtained by the random forest regressor model varies with the number of filters ranging from 10 to 2. The white area represents the eliminated filters for each iteration.

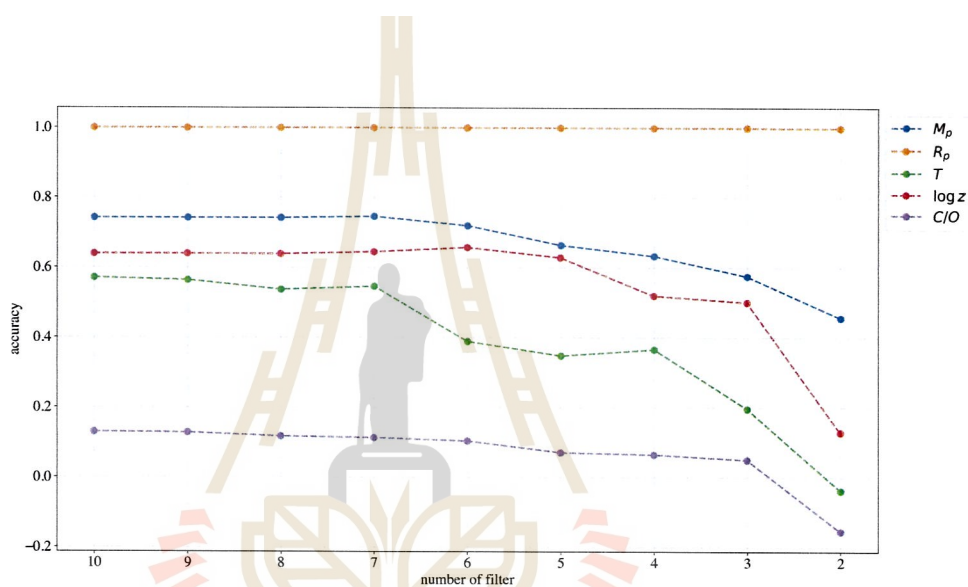


Figure 4.4 The parameters' prediction accuracy changes with the number of filters. The x-axis is the number of filters in each prediction, and the y-axis is the accuracy of the prediction. The blue, orange, green, red, and purple dash-dotted lines are the accuracy of planetary mass, planetary radius, temperature, logarithmic metallicity, and carbon-oxygen ratio, respectively.

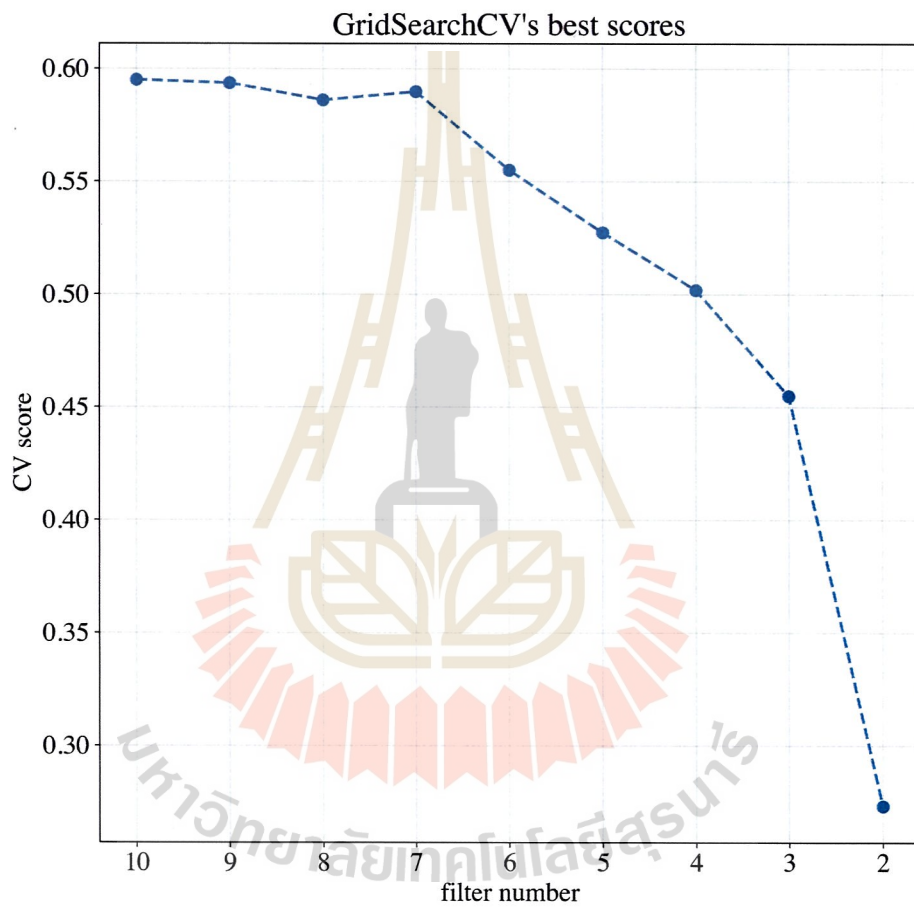


Figure 4.5 The cross-validation score of each random forest regressor model with the different number of filters.

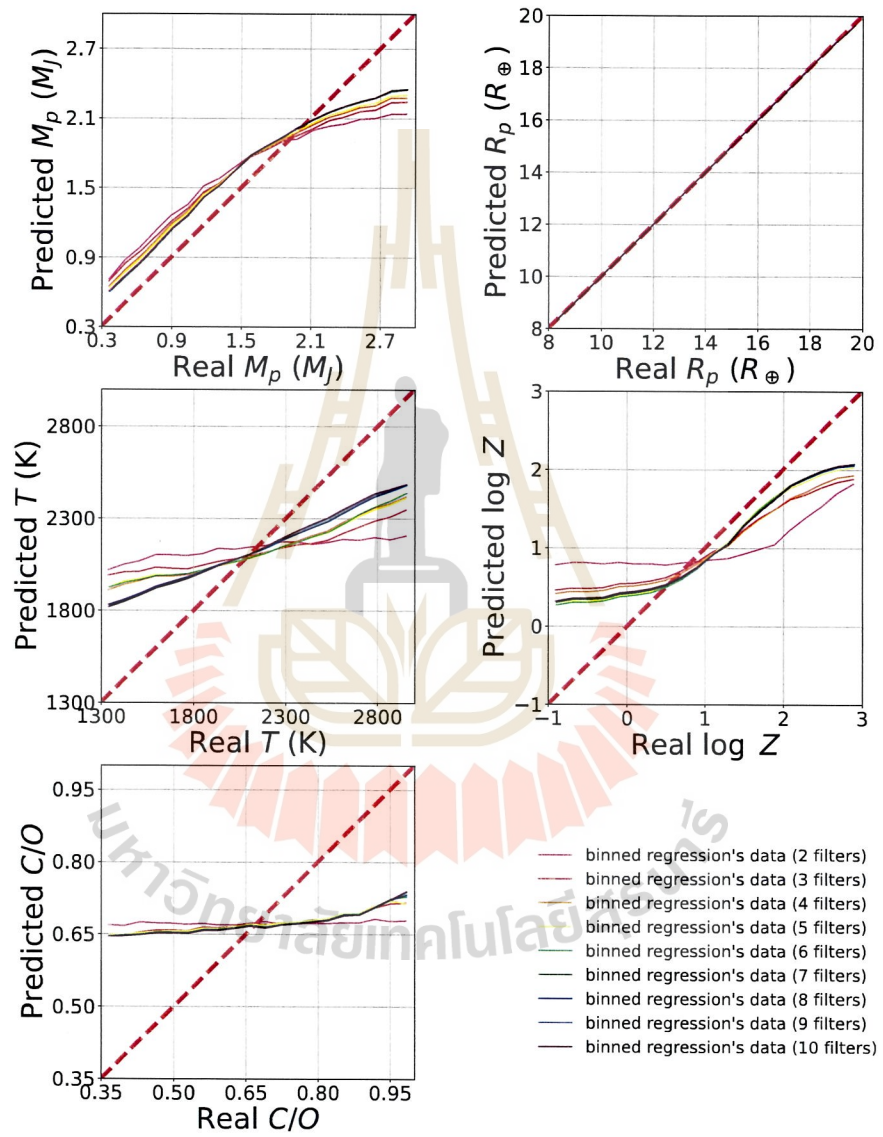


Figure 4.6 The filter optimization predictions' trend.

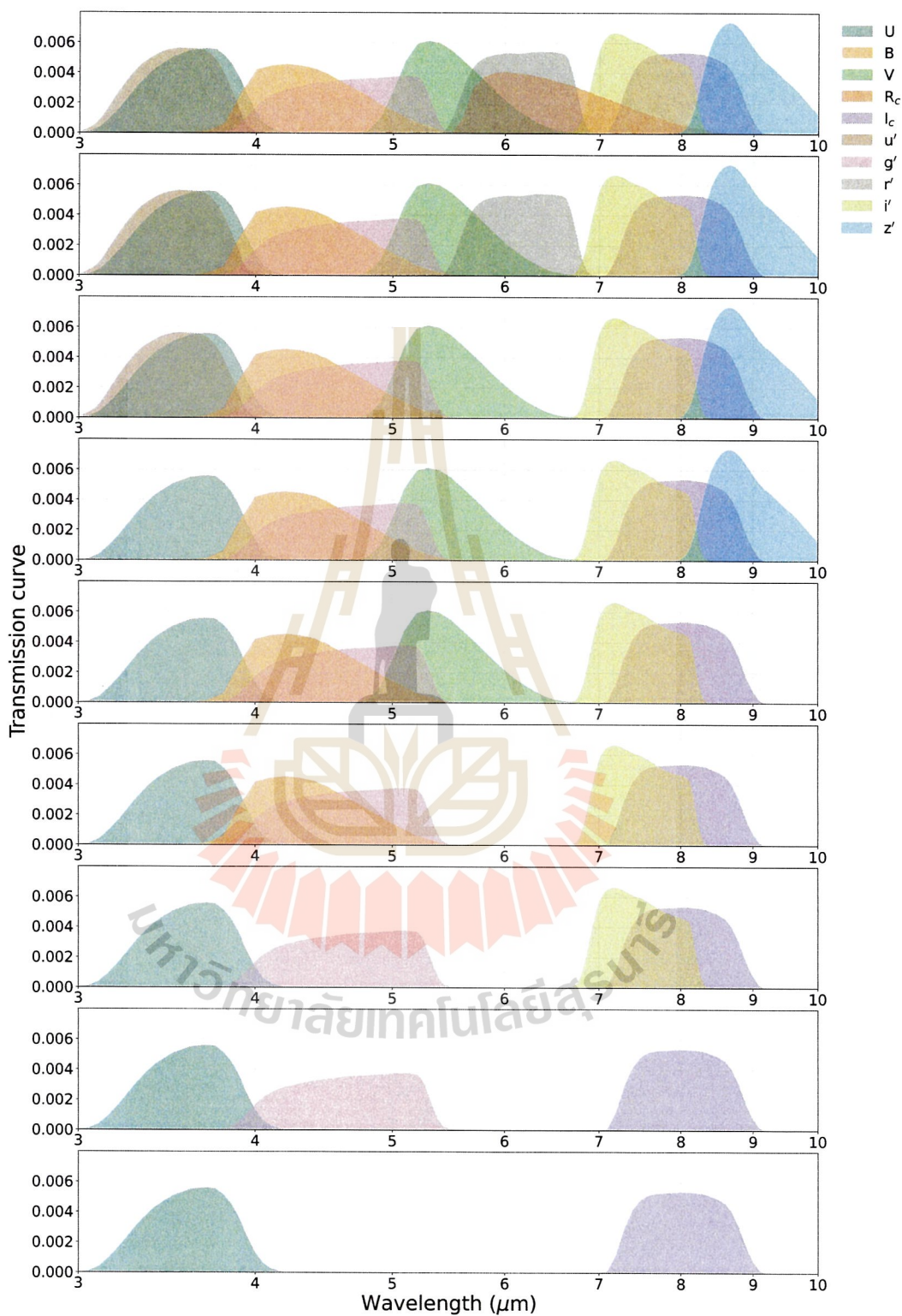


Figure 4.7 The filter transmission curve for each filter combination where the filter that is least importance is eliminated, one by one.

CHAPTER V

DISCUSSION AND CONCLUSION

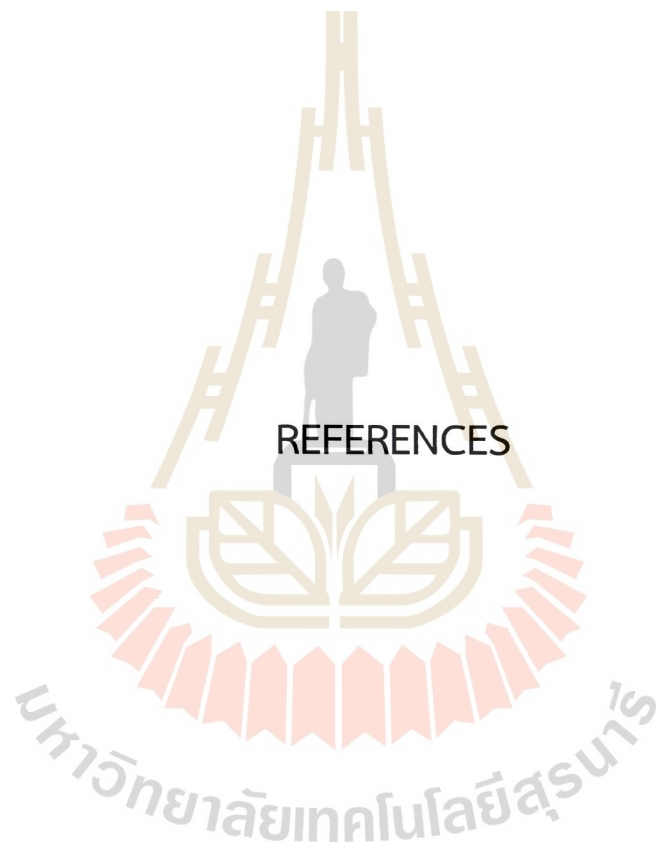
The best random forest regressor model with 80,000 training data and 20,000 test data consists of several estimators of 2,560 and a maximum depth of 320. The results compare with the estimated results from the nested sampling fitting model of PLATON with the number of live points set at 50 and 100. From the comparison, the accuracy obtained by the regressor model is comparable to that of the PLATON retrievals. The random forest regressor outperforms the PLATON nested sampling model using 50 and 100 live points by 160,000 and 270,000 times faster, respectively. The random forest regressor model accurately predicts the planetary radius and the acceptable accuracy in predicting planetary mass, temperature, and metallicity of the planetary atmosphere. Temperature predictions obtain a similar trend as the previous Márquez-Neila et al. (2018)'s work, with an overlap temperature range. The model does not correctly predict C/O due to the variation of C/O, which results in low sensitivity in the transmission spectra.

In filter optimization, we use the feature importance to rearrange the optical filters. The filter with the lowest importance is eliminated in every iteration. We obtain the descending order of the filters as I, U, g', i', B, V, z', u', r', and R. The filters with low importance have a negligible effect on the prediction of the exoplanet's parameters. When the parameters' prediction accuracy is considered, the optimal number of filters that yields the closest prediction to the highest number of filters is 7. The negative values in the predictions of log Z and C/O are seen at the lowest number of filters, which means the average value explains the data better than the model. From the prediction trend of the filter's optimization, the prediction trend of the planet mass decrease when the mass is bigger than 2.1 M_J . For the prediction of the logarithm of the metallicity, the model gives a high accuracy when the logarithm of metallicity is between 0.5 and 2. The temperature prediction performance decreases when the number of filters decreases.

In the prediction with the lowest accuracy of the C/O ratio, we expect that our coefficient of determination will be higher if the C/O ratio is not a model's

feature. We hope that this interesting point will be investigated in the future. Other machine learning models and hyper-parameters are not tuned in this work, which might also help improve the prediction efficiency. All synthesized transmission spectra are simulated from only one random seed. Hence, the error bar of accuracy is still not included in this work. All of these can be investigated more in future work.





REFERENCES

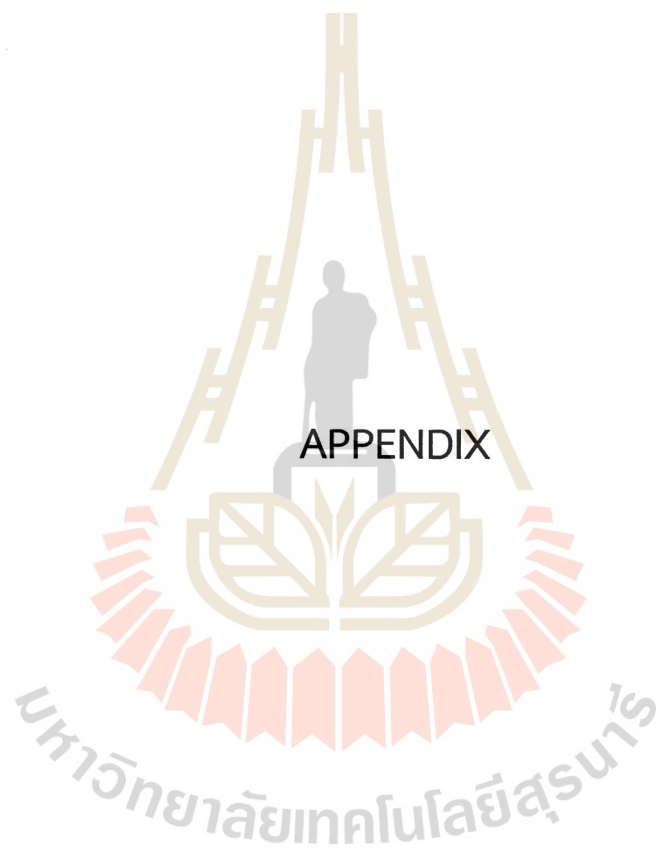
REFERENCES

- Borucki, W. J., Koch, D. G., Basri, G., Batalha, N., Brown, T. M., Bryson, S. T., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., and DeVore, E. (2011). Characteristics of planetary candidates observed by kepler. ii. analysis of the first four months of data. *The Astrophysical Journal*, 736(1), 19.
- Charbonneau, D., Brown, T. M., Noyes, R. W., and Gilliland, R. L. (2002). Detection of an extrasolar planet atmosphere. *The Astrophysical Journal*, 568(1), 377–384.
- Chen, J. and Kipping, D. (2016). Probabilistic forecasting of the masses and radii of other worlds. *The Astrophysical Journal*, 834(1), 17.
- Czesla, S., Schröter, S., Schneider, C. P., Huber, K. F., Pfeifer, F., Andreasen, D. T., and Zechmeister, M. (2019). Pya: Python astronomy-related packages.
- Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: the mcmc hammer. *Publications of the Astronomical Society of the Pacific*, 125(925), 306.
- Fraine, J., Deming, D., Benneke, B., Knutson, H., Jordán, A., Espinoza, N., Madhusudhan, N., Wilkins, A., and Todorov, K. (2014). Water vapour absorption in the clear atmosphere of a neptune-sized exoplanet. *Nature*, 513(7519), 526–529.
- Gaudi, B. S., Seager, S., and Mallen-Ornelas, G. (2005). On the period distribution of close-in extrasolar giant planets. *The Astrophysical Journal*, 623(1), 472.
- Granzer, T. (2014a). Johnson-cousins ubvri filter curves.
- Granzer, T. (2014b). Sloan u'g'r'i'z' filter curves.
- Kawashima, Y. and Ikoma, M. (2019). Theoretical transmission spectra of exoplanet atmospheres with hydrocarbon haze: Effect of creation, growth, and settling of haze particles. ii. dependence on uv irradiation intensity, metallicity, c/o ratio, eddy diffusion coefficient, and temperature. *The Astrophysical Journal*, 877(2), 109.

- Kreidberg, L., Bean, J. L., Désert, J.-M., Benneke, B., Deming, D., Stevenson, K. B., Seager, S., Berta-Thompson, Z., Seifahrt, A., and Homeier, D. (2014). Clouds in the atmosphere of the super-earth exoplanet gj 1214b. *Nature*, 505(7481), 69–72.
- Madhusudhan, N., Knutson, H., Fortney, J., and Barman, T. (2014). *Exoplanetary atmospheres*, page 739. University of Arizona Press, Arizona.
- Márquez-Neila, P., Fisher, C., Sznitman, R., and Heng, K. (2018). Supervised machine learning for analysing spectra of exoplanetary atmospheres. *Nature astronomy*, 2(9), 719–724.
- Munsaket, P., Awiphan, S., Chainakun, P., and Kerins, E. (2021). Retrieving exoplanet atmospheric parameters using random forest regression. *Journal of Physics: Conference Series*, 2145(1), 012010.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12, 2825–2830.
- Seager, S. and Sasselov, D. (2000). Theoretical transmission spectra during extrasolar giant planet transits. *The Astrophysical Journal*, 537(2), 916.
- Shaw, J., Bridges, M., and Hobson, M. (2007). Efficient bayesian inference for multi-modal problems in cosmology. *Monthly Notices of the Royal Astronomical Society*, 378(4), 1365–1370.
- Sing, D. K., Fortney, J. J., Nikolov, N., Wakeford, H. R., Kataria, T., Evans, T. M., Aigrain, S., Ballester, G. E., Burrows, A. S., and Deming, D. (2016). A continuum from clear to cloudy hot-jupiter exoplanets without primordial water depletion. *Nature*, 529(7584), 59–62.
- Stevens, D. J. and Gaudi, B. S. (2013). A posteriori transit probabilities. *Publications of the Astronomical Society of the Pacific*, 125(930), 933.
- Wolszczan, A. and Frail, D. A. (1992). A planetary system around the millisecond pulsar psr1257+ 12. *Nature*, 355(6356), 145–147.

Zhang, M., Chachan, Y., Kempton, E. M.-R., and Knutson, H. A. (2020). Platon ii: New capabilities and a comprehensive retrieval on hd 189733b transit and eclipse data. *The Astrophysical Journal*, 899(1), 27.





APPENDIX

REAL-PREDICTION PLOTS OF FILTER OPTIMISATION

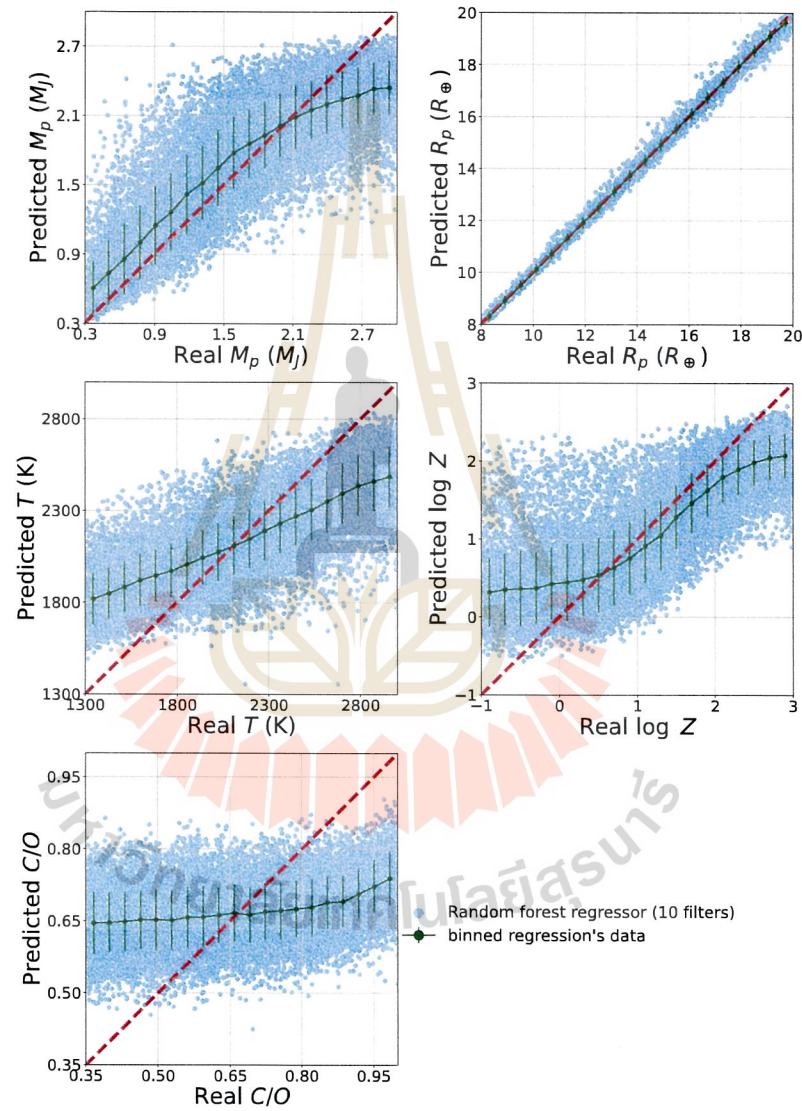


Figure 1 Real and predicted values of five planetary parameters for 10 filters (U, B, V, R, I, u', g', r', i' and z').

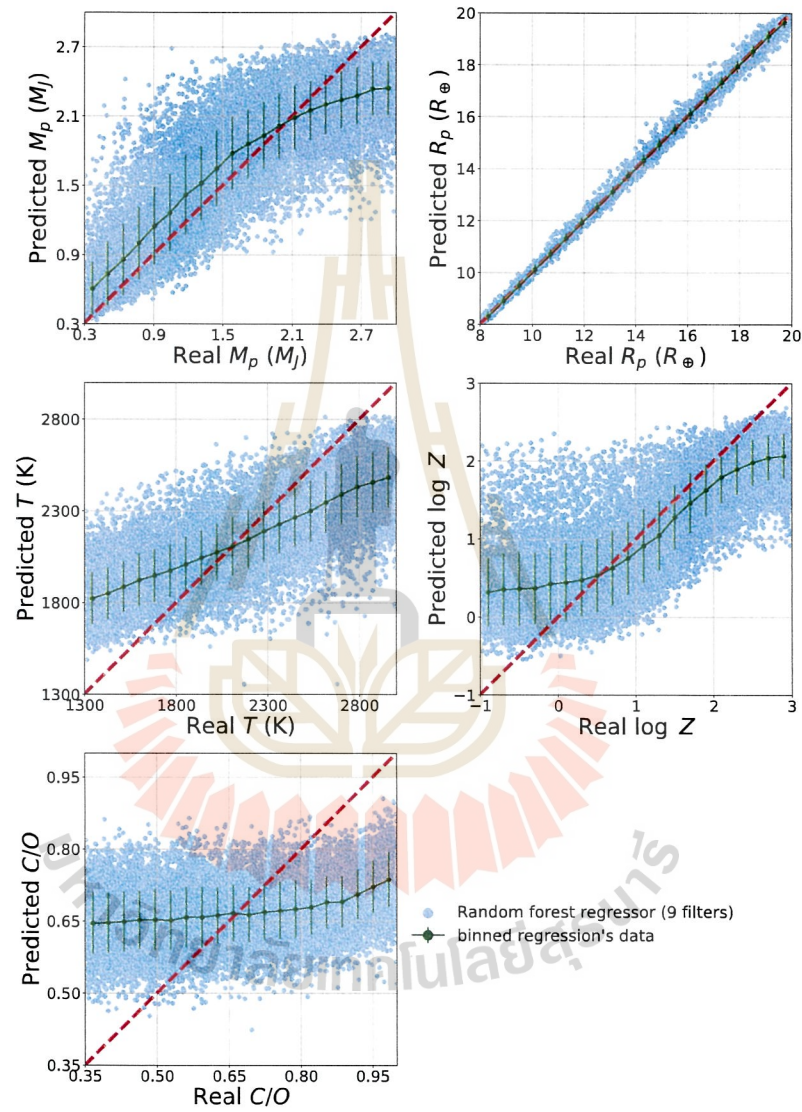


Figure 2 Real and predicted values of five planetary parameters for 9 filters (U, B, V, I, u', g', r', i' and z').

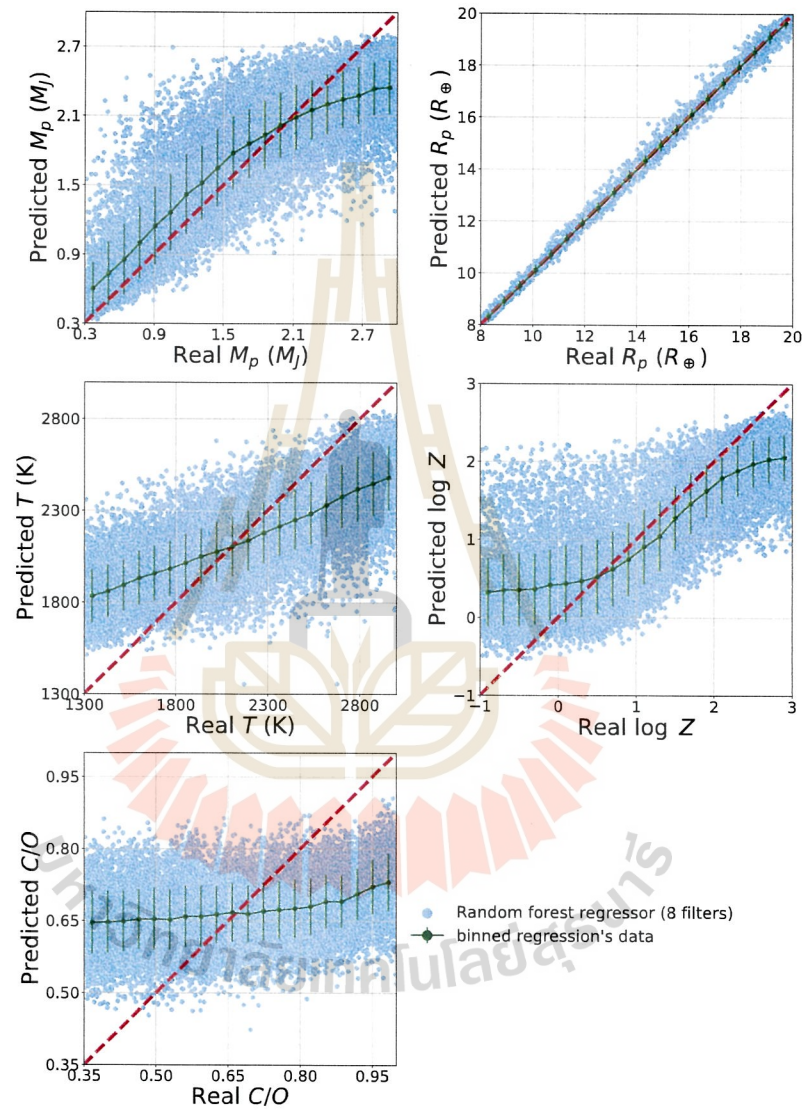


Figure 3 Real and predicted values of five planetary parameters for 8 filters (U, B, V, I, u' , g' , i' and z').

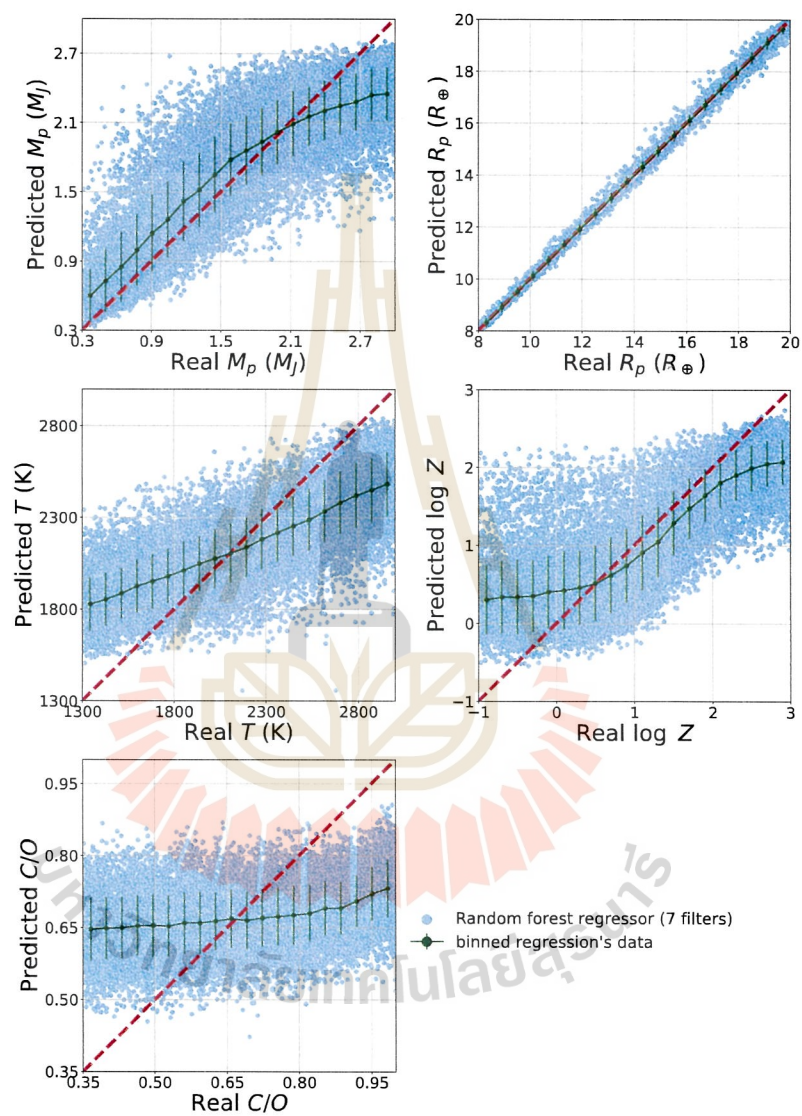


Figure 4 Real and predicted values of five planetary parameters for 7 filters (U, B, V, I, g', i' and z').

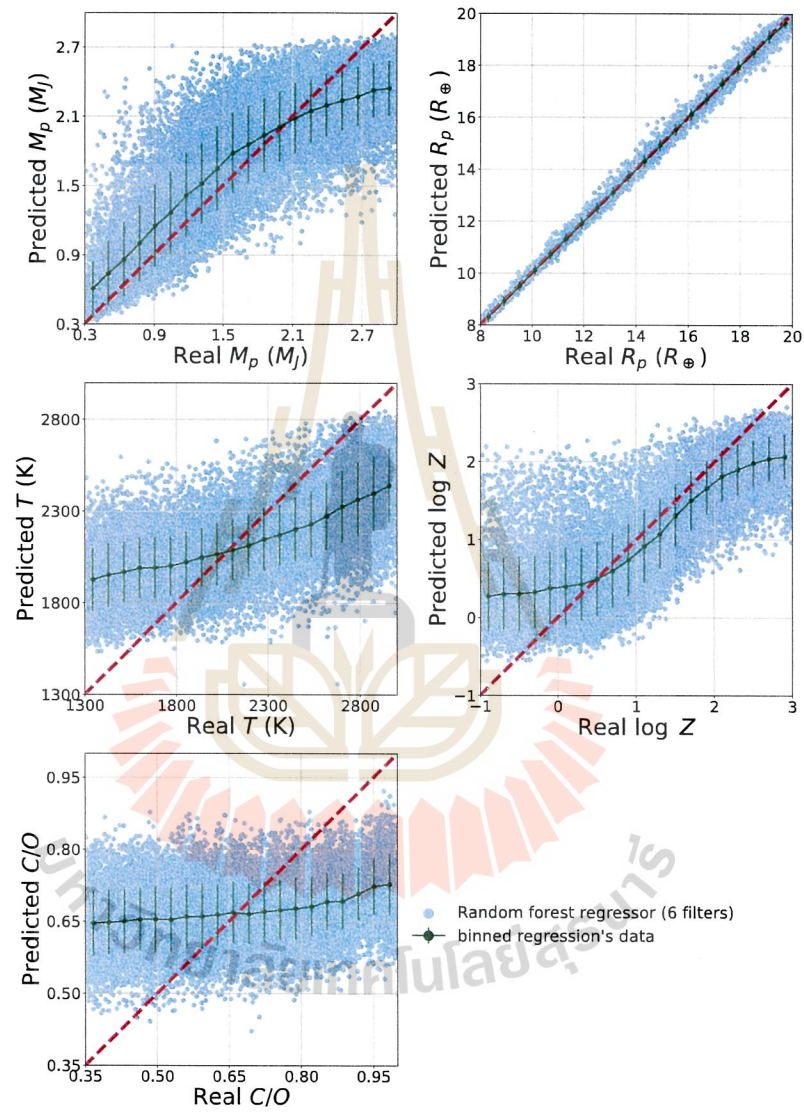


Figure 5 Real and predicted values of five planetary parameters for 6 filters (U, B, V, I, g' and i').

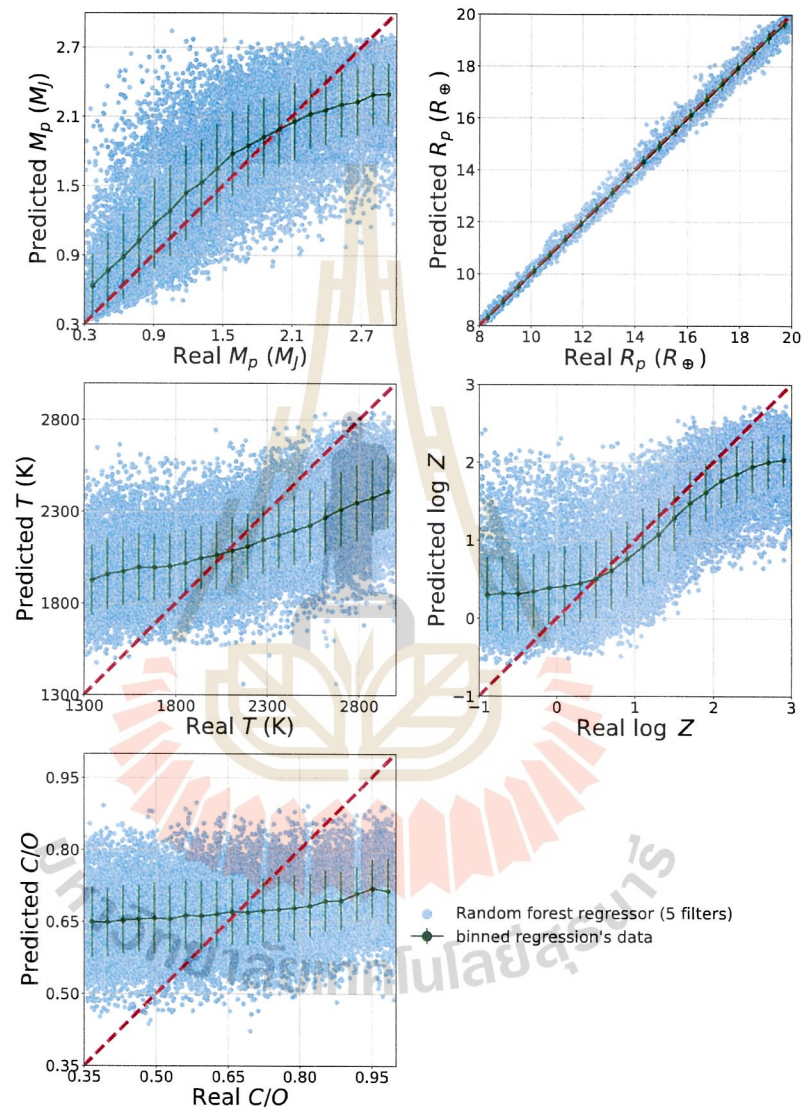


Figure 6 Real and predicted values of five planetary parameters for 5 filters (U, B, I, g' and i').

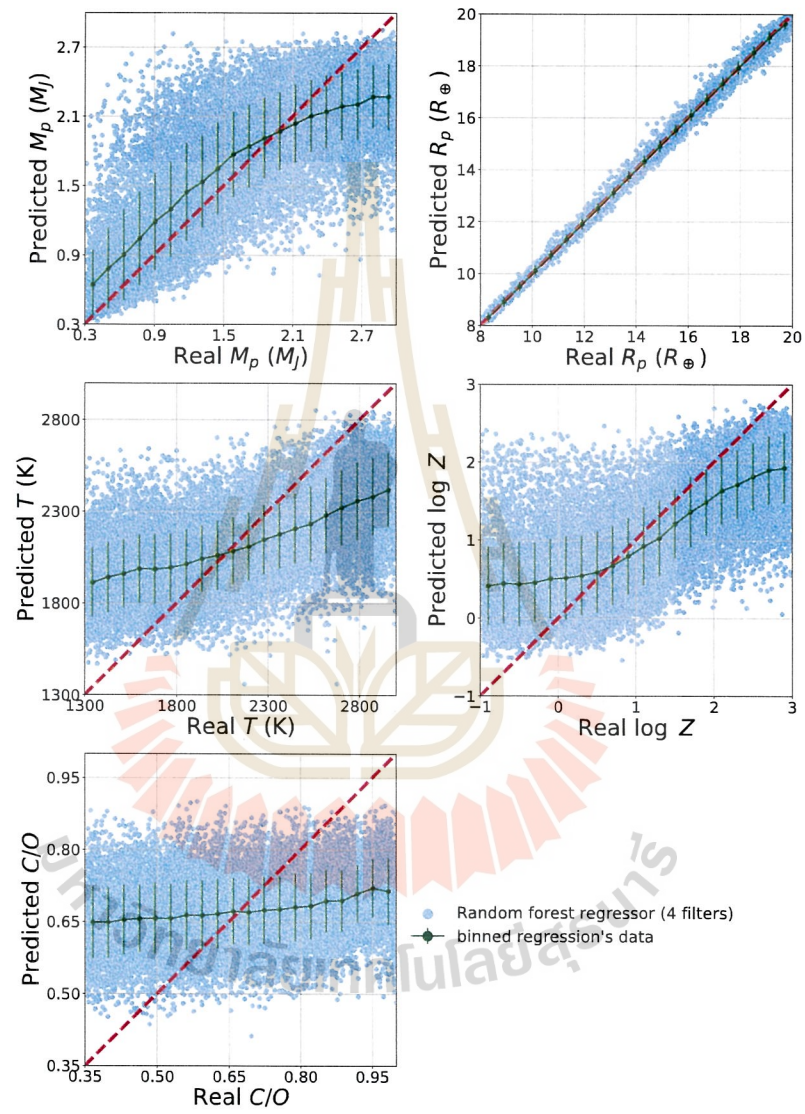


Figure 7 Real and predicted values of five planetary parameters for 4 filters (U, I, g' and i').

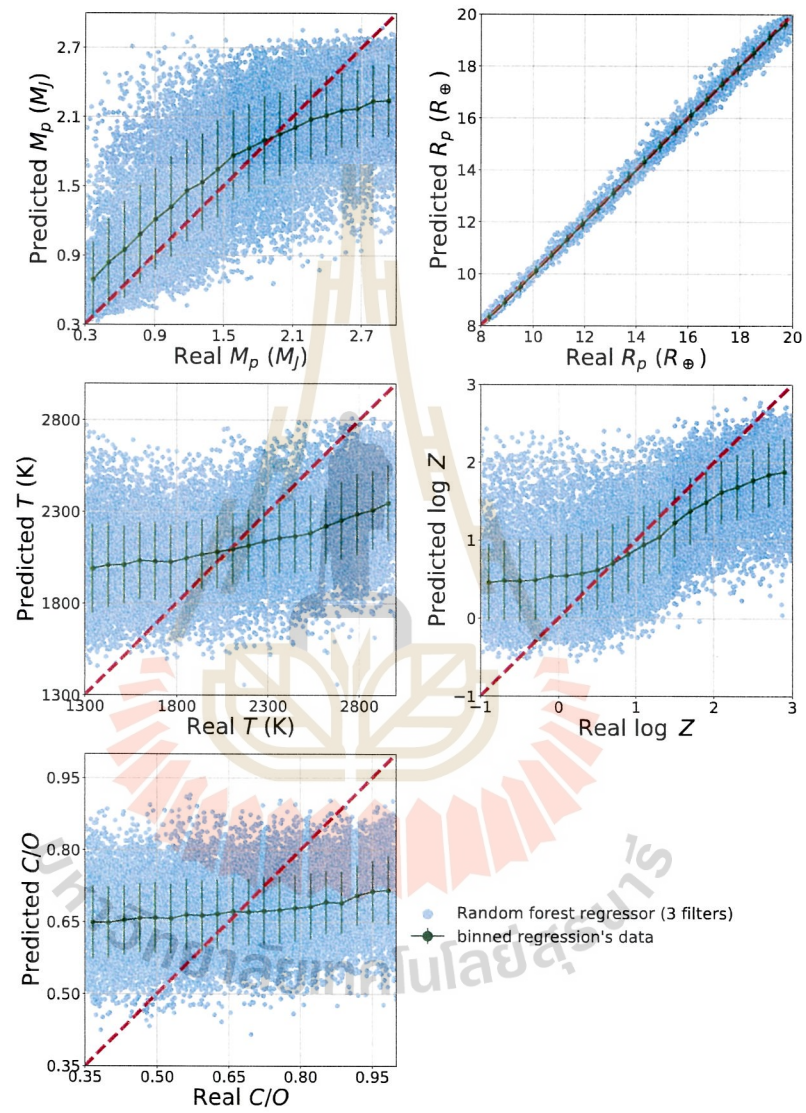


Figure 8 Real and predicted values of five planetary parameters for 3 filters (U, I and g').

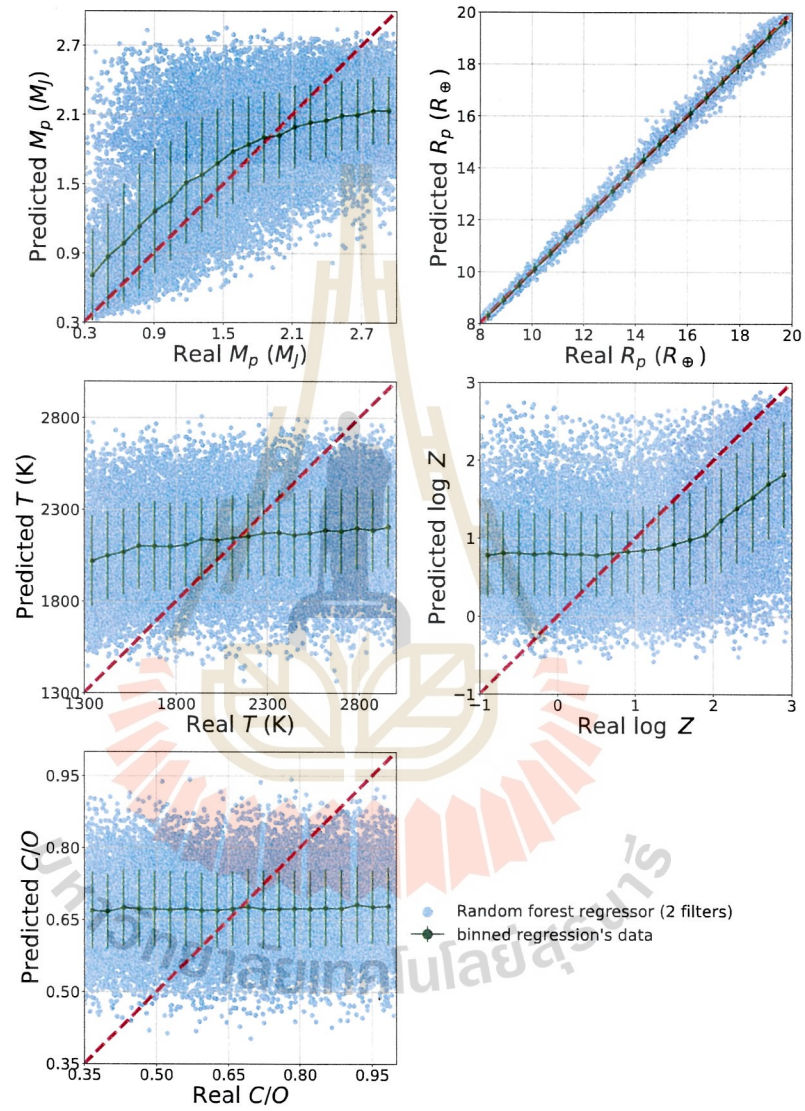


Figure 9 Real and predicted values of five planetary parameters for 2 filters (U and I).

CURRICULUM VITAE

Name : Ms. Patcharawee Munsaket

Date of Birth : August 27, 1996

Place of Birth : Nakhon Ratchasima, Thailand

Education :

2009 - 2015 High school at Suranari Wittaya school, Thailand

2015 - 2019 Bachelor of Science (Honors Programs), School of Physics,
Institute of science, Suranaree University of Technology,
Thailand

2019 - 2021 Master of Science, School of Physics, Institute of Science,
Suranaree University of Technology, Thailand

Grants and Fellowships :

2012 - present Scholarships from Development and Promotion of Science
and Technology Talents Project (DPST)

Publications :

1. Munsaket, P., Awiphan, S., Chainakun, P., and Kerins, E. (2021). Retrieving exoplanet atmospheric parameters using random forest regression. *Journal of Physics: Conference Series*, 2145(1), 012010.