OPTICAL FILTER RANKING FOR THE HOT JUPITERS TRANSMISSION SPECTROSCOPY

Patcharawee Munsaket[a,*], Poemwai Chainakun[a,b], Supachai Awiphan[c], Eamonn Kerins[d], Napaporn A-thano[c]

[a] *School of Physics, Institute of Science, Suranaree University of Technology, 111 University Avenue, Muang, Nakhon Ratchasima, 30000, Thailand.*

[b] *Centre of Excellence in High Energy Physics and Astrophysics, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand*

[c] *National Astronomical Research Institute of Thailand (Public Organization), 260 Moo 4, Donkaew, Mae Rim, Chiang Mai, 50180, Thailand*

[d] *Jodrell Bank Centre for Astrophysics, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom*

*Corresponding Author's E-mail: patcharawee.m@grad.sut.ac.th

Backgrounds

- Transmission spectroscopy probes exoplanet atmospheres by measuring wavelength-dependent variations in transit depth, which reveal atmospheric composition.
- Hot Jupiters are the most accessible ground-based targets because of their large radii.

Motivations

- Several filter sets are available for broadband transmission spectroscopy, including Johnson—Cousins and the Sloan Digital Sky Survey (SDSS), but there is no established method for selecting the optimal filters.
- Traditional retrievals such as Markov chain Monte Carlo (MCMC) and nested sampling require significant computational time.

Objectives

- Apply supervised machine learning, using Random Forest Regressor (RFR)[1], to predict atmospheric properties of hot Jupiters and compare its computational efficiency with the traditional Tau Retrieval for Exoplanets (TauREx3)[2] retrieval.
- Use the RFR to rank optical filters that are suitable for measuring the transmission spectra of hot Jupiters.

Methods

(a) Simulated transmission spectra

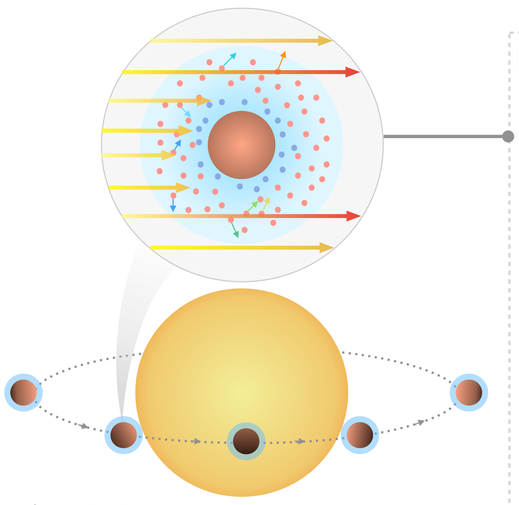


Figure 1. An exoplanet orbits its host star and passes in front during transit. Starlight crossing the atmosphere undergoes wavelength-dependent absorption and scattering, revealing atmospheric composition.

Random sampling parameters

- Planetary radius: R_p
- Planetary Temperature: T_p
- TiO and VO gas mixing ratio: X_{TiO}, X_{VO}

Simulate spectrum with **TauREx3 forward model**

Weight and bin with filter profiles

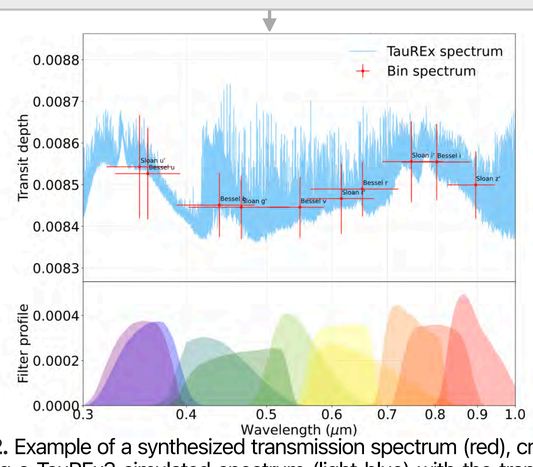


Figure 2. Example of a synthesized transmission spectrum (red), created by weighting a TauREx3 simulated spectrum (light blue) with the transmission profiles of the filters shown below.

(b) Machine learning

Stage 1 result: R_p

Stage 2 results: T_p, X_{TiO}, X_{VO}

Hyperparameter tuning and training

`GridSearchCV` + `RandomForestRegressor`

Data preprocessing

- Normalization
- Standard Scaler
- Logarithmic transformation

(c) TauREx3 retrieval

`Optimizer.Nestle`

Solution: R_p, X_{TiO}, T_p, X_{VO}

(d) Filter ranking

Measure feature importance of 10 features

Eliminate the lowest importance

Repeat the process until the two highest-importance filters remain

Results

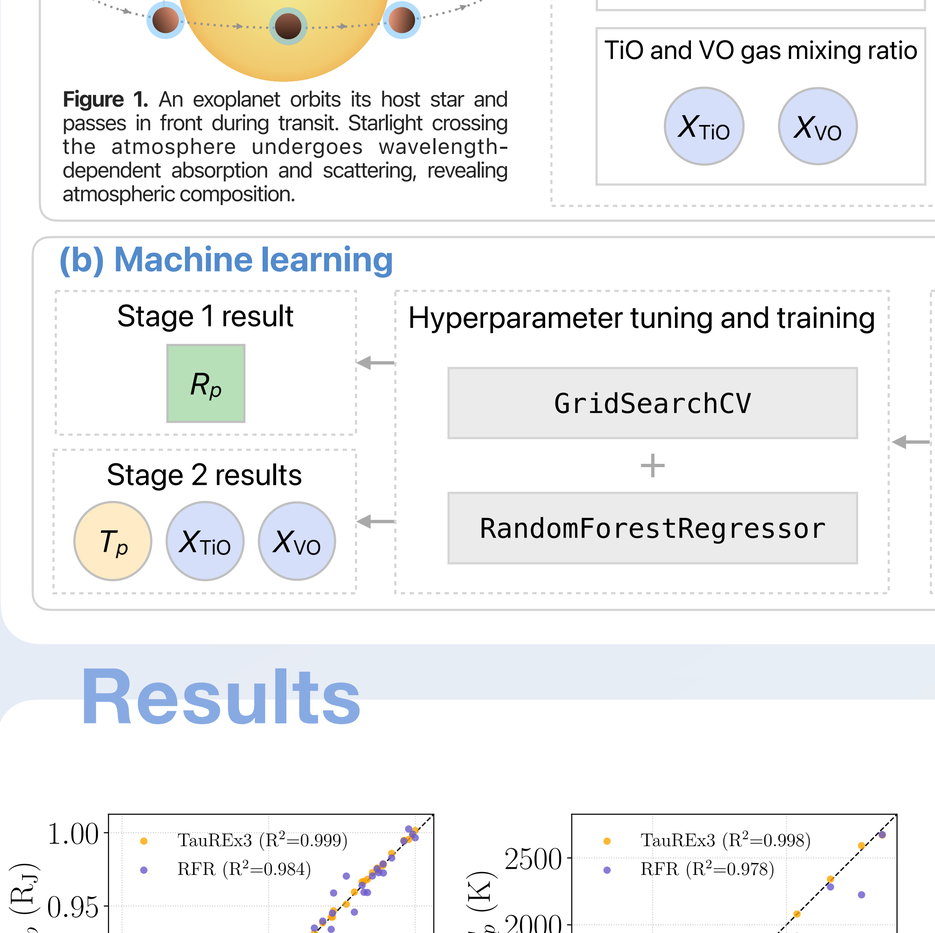


Figure 3. Comparing the performance of TauREx3 and RFR in retrieving four planetary parameters (R_p, T_p, X_{TiO} and X_{VO}) from transmission spectra of 50 exoplanets in the [0.8, 1.0) R_J range. Orange points represent TauREx3 results, and purple points represent RFR predictions. The dashed line is the one-to-one line, where predicted and true values are equal.

Table 1. Processing time of RFR compared with the TauRex3 retrieval for the dataset in the [0.8, 1.0)R_J range.

Approach	Number of data	Process time (hr:min:sec:ms)
RFR	100	00 : 00 : 00 : 852
TauREx3 retrieval (nlive=50)	100	11 : 37 : 26 : 00

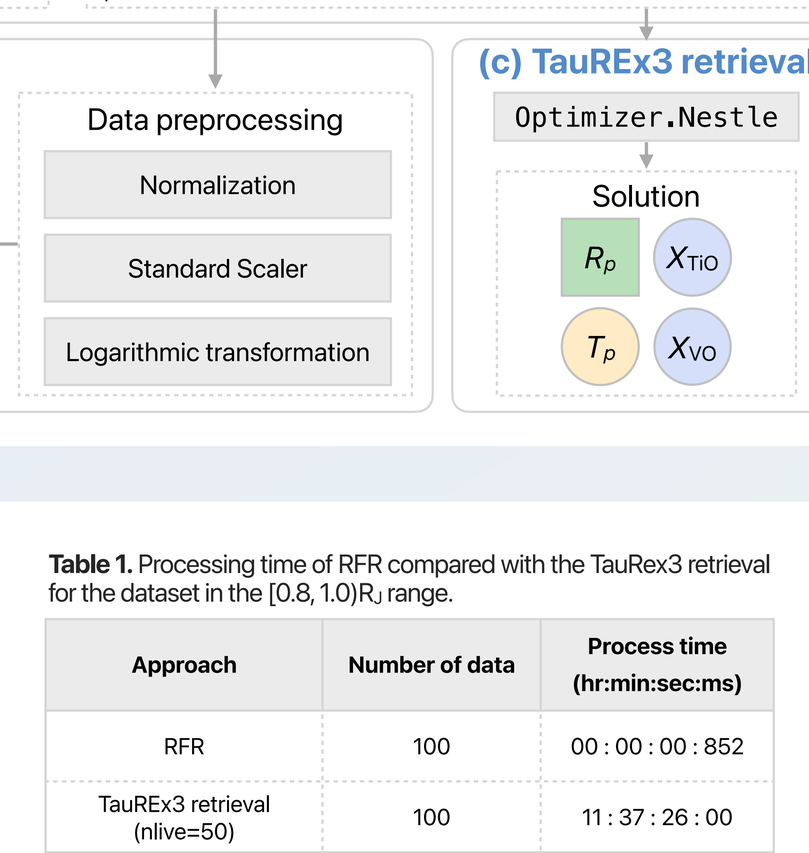


Figure 4. Example filter ranking for planets in the range [0.8, 1.0) R_J. The process begins with 10 filters and eliminates the lowest-importance feature at each step (importance shown in the heatmap). The final rank is the y-axis label.

Table 2. The filter ranking of the optical filter for each bin of the planetary radius. The leftmost filter is the most important for each bandpass.

R_p range (R_J)	1	2	3	4	5	6	7	8	9	10
[0.8, 1.0)	*V*	*u′*	*z′*	*B*	*i′*	*r′*	*I*	*g′*	*U*	*R*
[1.0, 1.2)	*V*	*u′*	*z′*	*i′*	*B*	*r′*	*I*	*g′*	*U*	*R*
[1.2, 1.4)	*V*	*u′*	*i′*	*z′*	*r′*	*B*	*I*	*g′*	*U*	*R*
[1.4, 1.6)	*V*	*u′*	*i′*	*r′*	*z′*	*B*	*I*	*U*	*g′*	*R*
[1.6, 1.8]	*V*	*u′*	*r′*	*i′*	*z′*	*B*	*U*	*I*	*g′*	*R*

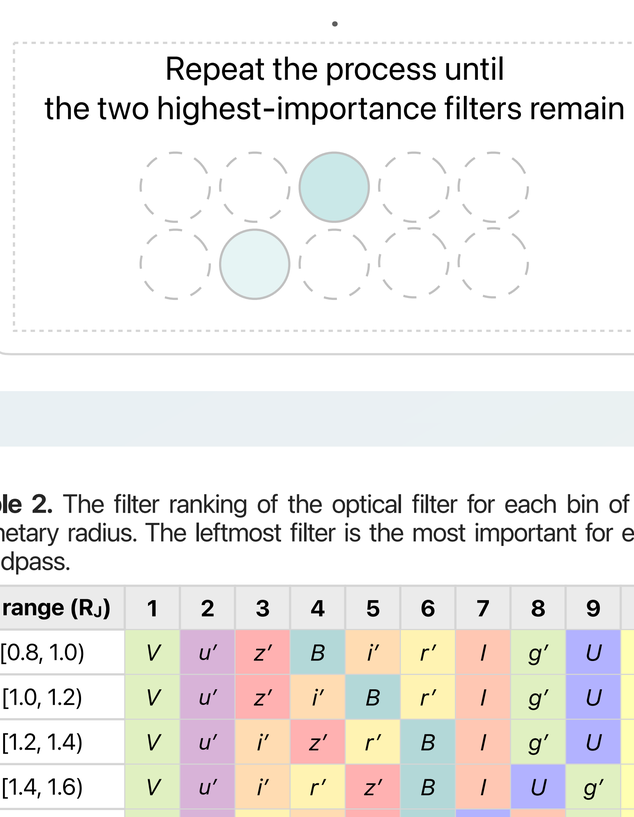


Figure 5. The cross-validation score, along with the reduction in the number of filters for each planet radius range. The dashed lines represent the filter-elimination trends for each R_p, and the black solid line is their mean value.

Conclusions

- A RFR trained on TauREx3 synthetic spectra predicts R_p, T_p, X_{TiO} and X_{VO} with high accuracy and is approximately 400,000 times faster than traditional retrieval methods.
- Feature-ranking shows that 5 optical filters achieve nearly the same performance as all 10, enabling more efficient observations.
- Limitations include the need for a two-stage model and applicability only to hot Jupiters; future work should explore a single model, multi-parameter machine learning retrievals, and broader exoplanet classes.

References

[1] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

[2] Al-Refaie, A. F., Changeat, Q., Waldmann, I. P., & Tinetti, G. (2021). TauREx 3: A fast, dynamic, and extendable framework for retrievals. The Astrophysical Journal, 917(1), 37.

Acknowledgements

We gratefully acknowledge the funding and facilities support from the Development and Promotion of Science and Technology Talents Project (DPST), the National Astronomical Research Institute of Thailand (NARIT) and Spectroscopy and Photometry Exoplanetary Atmosphere Research Network (SPEARNET). This research would not have been possible without their generous contributions.