EPSC Abstracts
Vol. 13, EPSC-DPS2019-1512-1, 2019
EPSC-DPS Joint Meeting 2019

EPSC
EPSC-DPS Joint Meeting 2019
15–20 September 2019 | Geneva, Switzerland

Supervised Machine Learning for Analysing Spectra of Exoplanetary Atmospheres

Chloe Fisher (1), Kevin Heng (1), Pablo Márquez-Neila (2) and Raphael Sznitman (2)
(1) Center for Space and Habitability, University of Bern, Switzerland, (2) ARTORG Center for Biomedical Engineering, University of Bern, Switzerland (chloe.fisher@csh.unibe.ch)

Abstract

Based on our previous publication of this work [1], we report an adaptation of the "random forest" method of supervised machine learning [2, 3], trained on a pre-computed grid of atmospheric models, which retrieves full posterior distributions of the abundances of molecules and the cloud opacity. We demonstrate our technique on a transmission spectrum of the hot gas-giant exoplanet WASP-12b using a five-parameter model (temperature, a constant cloud opacity and the volume mixing ratios or relative abundance by number of water, ammonia and hydrogen cyanide) [4]. We obtain results consistent with the standard nested-sampling retrieval method. Additionally, we can estimate the sensitivity of the measured spectrum to constraining the model parameters and we can quantify the information content of the spectrum. Our method can be straightforwardly applied using more sophisticated atmospheric models and also to interpreting an ensemble of spectra without having to retrain the random forest.

1. Introduction

The use of machine learning is becoming ubiquitous in astronomy [5, 6, 7], but remains rare in the study of the atmospheres of exoplanets. Given the spectrum of an exoplanetary atmosphere, a multi-parameter space is swept through in real time to find the best-fit model [8, 9, 10]. Known as "atmospheric retrieval", it is a technique that originates from the Earth and planetary sciences [11]. Such methods are very time-consuming and by necessity there is a compromise between physical and chemical realism versus computational feasibility. Machine learning has previously been used to determine which molecules to include in the model, but the retrieval itself was still performed using standard methods [12]. Here, we present `HELA`, our random forest retrieval algorithm for exoplanetary atmospheres, and demonstrate its use on the HST WFC3 transmission spectrum for the hot Jupiter WASP-12b.

2. Methods

We generated a set of 100,000 examples of transmission spectra with varying parameters for temperature, molecular abundances and cloud opacities, using a model validated for HST resolution spectra [4]. We included the molecules H_2O, HCN, and NH_3. These spectra were binned down to the resolution matching the data from Kreidberg et al. 2015 [13], and divided into a training and testing set. We then trained 1000 regression trees on this data, using bootstrapping and bagging techniques to ensure no biases. We then used the testing set to check the performance of our forest.

3. Results

Once trained and tested we were able to use our forest to predict on the WASP-12b data. We could then compare our results with the standard retrieval method of "nested-sampling", and found a good agreement. We obtained a water volume mixing ratio of $\log X_{H_2O} = -2.8^{+1.4}_{-3.6}$ and a temperature of 952^{+412}_{-151} K, which are broadly consistent with the previous analysis [13]. Figure 1 shows our posteriors for WASP-12b.

A great advantage of using the random forest is the information content analysis known as "feature importance", which is easily computed in the training process. This feature importance gives the relative weight of each data point when constraining each parameter. We find a good agreement with some of our intuition, for the molecules for example, where the feature importance appears to match the opacities. However, temperature and cloud opacity are perhaps less intuitive parameters, and thus the feature importance for these is enlightening.

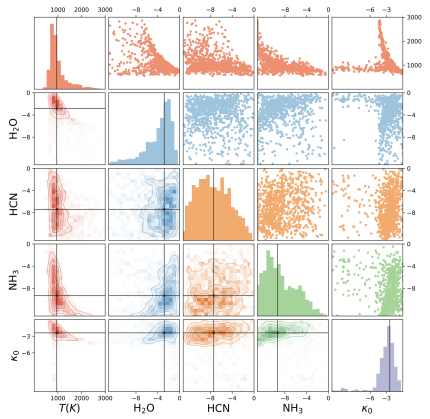

Figure 1: Posterior distributions of the relative molecular abundances (volume mixing ratios), temperature and cloud opacity obtained from the machine-learning retrieval analysis of the WFC3 transmission spectrum of WASP-12b. Shown are the logarithm (base 10) of the volume mixing ratios and cloud opacity. Within each scatter plot, each dot is an individual prediction of a single tree in the random forest. The straight lines indicate the median values of the parameters

4. Conclusion

We have demonstrated our random forest retrieval technique for exoplanetary atmospheres, and validated its performance against the tradition nested-sampling technique. We have obtained predictions for the hot Jupiter WASP-12b, and produced feature importance analysis for the parameters in our model.

Acknowledgements

We acknowledge partial financial support from the Center for Space and Habitability, the University of Bern International 2021 PhD Fellowship, the PlanetS National Center of Competence in Research, the Swiss National Science Foundation, the European Research Council via a Consolidator Grant, and the Swiss-based MERAC Foundation.

References

[1] Márquez-Neila, P., Fisher, C., Sznitman, R., & Heng, K. 2018, Nature Astronomy, 2, 719-724

[2] Ho, T.K. 1998, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 832-844

[3] Breiman, L. 2001, Machine Learning, 45, 5-32

[4] Heng, K., & Kitzmann, D. 2017, MNRAS, 470, 2972-2981

[5] Banerji, M. et al. 2010, MNRAS, 406, 342-353

[6] Graff, P., Feroz, F., Hobson, M.P., & Lasenby, A. 2014, MNRAS., 441, 1741-1759

[7] Pearson, K.A., Palafox, L., Griffith, C.A. 2018, MNRAS, 474, 478-491

[8] Madhusudhan, N., & Seager, S. 2009, ApJ., 707, 24-39

[9] Benneke, B., & Seager, S. 2012, ApJ., 753, 100

[10] Line, M.R. et al. 2012, ApJ., 749, 93

[11] Rodgers, C.D. 2000, World Scientific

[12] Waldmann, I.P. 2016, ApJ., 820, 107

[13] Kreidberg, L. et al. 2015, ApJ, 814, 66