

---

# Bayesian Deep Learning for Exoplanet Atmospheric Retrieval

---

Frank Soboczenski,<sup>a</sup> Michael D. Himes,<sup>b</sup> Molly D. O’Beirne,<sup>c</sup> Simone Zorzan<sup>d</sup>  
Atılım Güneş Baydin,<sup>e</sup> Adam D. Cobb,<sup>e</sup> Yarin Gal,<sup>f</sup> Daniel Angerhausen<sup>g</sup>  
Massimo Mascaro,<sup>h</sup> Giada N. Arney,<sup>i,j</sup> Shawn D. Domagal-Goldman<sup>i,k</sup>

<sup>a</sup>School of Population Health and Environmental Sciences, King’s College London

<sup>b</sup>Planetary Sciences Group, Department of Physics, University of Central Florida

<sup>c</sup>Department of Geology and Environmental Science, University of Pittsburgh

<sup>d</sup>ERIN Department, Luxembourg Institute of Science and Technology

<sup>e</sup>Department of Engineering Science, University of Oxford

<sup>f</sup>Department of Computer Science, University of Oxford

<sup>g</sup>Center for Space and Habitability, University of Bern

<sup>h</sup>OCTO Applied AI, Google Cloud

<sup>i</sup>Virtual Planetary Laboratory Team, NASA Astrobiology Institute

<sup>j</sup>Planetary Systems Laboratory, NASA Goddard Space Flight Center

<sup>k</sup>Planetary Environments Laboratory, NASA Goddard Space Flight Center

frank.soboczenski@kcl.ac.uk; mhimes@knights.ucf.edu

mdobeirne@pitt.edu; simone.zorzan@list.lu

{gunes, acobb}@robots.ox.ac.uk; yarin@cs.ox.ac.uk

daniel.angerhausen@csh.unibe.ch; massy@google.com

{giada.n.arney, shawn.goldman}@nasa.gov

## 1 Introduction

Over the past decade, the study of extrasolar planets has evolved rapidly from plain detection and identification to comprehensive categorization and characterization of exoplanet systems and their atmospheres. Atmospheric retrieval, the inverse modeling technique used to determine an exoplanetary atmosphere’s temperature structure and composition from an observed spectrum, is both time-consuming and compute-intensive, requiring complex algorithms that compare thousands to millions of atmospheric models to the observational data to find the most probable values and associated uncertainties for each model parameter [12]. For rocky, terrestrial planets, the retrieved atmospheric composition can give insight into the surface fluxes of gaseous species necessary to maintain the stability of that atmosphere, which may in turn provide insight into the geological and/or biological processes active on the planet [16]. These atmospheres contain many molecules, some of them biosignatures, spectral fingerprints indicative of biological activity, which will become observable with the next generation of telescopes [5]. Runtimes of traditional retrieval models scale with the number of model parameters, so as more molecular species are considered, runtimes can become prohibitively long. Recent advances in machine learning (ML) and computer vision [11, 8] offer new ways to reduce the time to perform a retrieval by orders of magnitude [13, 23], given a sufficient data set to train with. Here we present an ML-based retrieval framework called *Intelligent exoplanet Atmospheric Retrieval* (INARA) that consists of a Bayesian deep learning model for retrieval and a data set of 3,000,000 synthetic rocky exoplanetary spectra generated using the NASA Planetary Spectrum Generator (PSG) [20]. Our work represents the first ML retrieval model for rocky, terrestrial exoplanets and the first synthetic data set of terrestrial spectra generated at this scale.

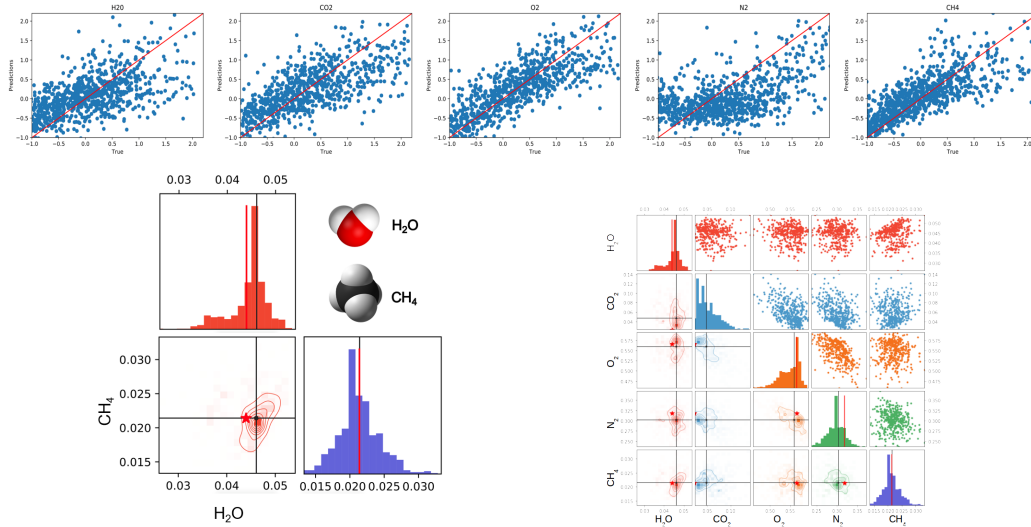


Figure 1: Predictions of H<sub>2</sub>O, CO<sub>2</sub>, O<sub>2</sub>, N<sub>2</sub>, CH<sub>4</sub> based on the best performing model; training limited to 64 epochs on 110,000 parameter–spectra pairs hence some uncertainties reflect calibration issues.

## 2 Background

Traditionally, the study of exoplanetary atmospheres has been done by fitting forward models to observational data, which is based on the relative decrease in flux when the exoplanet is in front of or behind its host star [3, 4]. This is usually performed using a Monte Carlo sampling method in a Bayesian framework to propose atmospheric models, simulate the spectrum, and compare it to the observed data [e.g., 17, 19]. Degeneracies among atmospheric parameters complicate this process, necessitating the evaluation of hundreds of thousands to millions of atmospheric models to fully explore the parameter space. This results in a posterior distribution which characterizes these degeneracies and informs the relative probability of the ranges of values considered for each model parameter. While these sampling methods are executed in parallel, this task still requires a significant amount of computational time [12].

Recently, the exoplanet community has begun to apply supervised ML methods to the problem of atmospheric retrieval. Waldmann [21] used a deep belief network to identify molecular species in an observed spectrum, paving the way for more advanced ML applications. Building upon this, two ML retrieval algorithms have been developed to date: ExoGAN [23] and HELA [13]. These produce results in seconds to minutes, compared to on the order of 100 CPU hours for the aforementioned traditional Monte Carlo sampling-based methods. ExoGAN utilizes a generative adversarial network (GAN) [7] to approximate the data distribution of realistic spectra and then uses the trained GAN to make predictions using inpainting to infer planetary conditions from observed spectra. HELA uses random forests [10] to similarly make predictions of planetary parameters from observed spectra. Both models produce results that are generally consistent with conventional retrieval methods. Note that these models specialize in hot Jupiters, a class of gas giant exoplanets on very short-period orbits, and consider less than a handful of molecules.

## 3 Methods

We train a deep neural network in a supervised setting to predict exoplanet atmospheric parameters  $\theta$  given an observed spectrum  $s$ , using a training set  $(s_i, \theta_i), i = 1, \dots, m$  that we generate by running the NASA PSG<sup>1</sup> [20] simulator to get  $s_i = \text{psg}(\theta_i)$ , where the parameters are sampled from a physically-motivated prior model  $\theta_i \sim p(\theta)$ . Spectra  $s_i$  are vectors (of length 4379) describing radiation intensity as a scalar function of wavelength and therefore we explore a series of 1D convolutional neural network (CNN) configurations. In order to train our CNN models, we generate a data set encompassing spectra based on a given planetary system model, where we consider F-, G-, K-,

<sup>1</sup><https://psg.gsfc.nasa.gov/>

Table 1: Comparison of atmospheric retrieval methods.

Method	CPU time for inference	Number of molecules retrieved
Traditional	Hundreds of hours	User-specified
ExoGAN [23]	Minutes	4
HELA [13]	Seconds	3
INARA	Seconds	12

and M-type main sequence stars. Observations are simulated using an instrument model of the Large UltraViolet/Optical/InfraRed Surveyor (LUVOIR), a design concept for a multi-wavelength space observatory, but with a much higher resolution. The prior model  $p(\theta)$  comprises planetary parameters (radius, mass, surface pressure, semi-major axis, pressure temperature profile) and atmospheric compositions. Planetary parameters are randomly selected from ranges and distributions consistent with our solar system and observations of other systems [1, 2, 14, 15, 18, 22]. The ranges for these parameters are chosen such that a planet in an Earth-like orbit can vary in temperature by a few hundred Kelvin. We consider 12 molecules based on the composition of atmospheres in our solar system as well as the observability of species [5]:  $\text{H}_2\text{O}$ ,  $\text{CO}_2$ ,  $\text{O}_2$ ,  $\text{N}_2$ ,  $\text{CH}_4$ ,  $\text{N}_2\text{O}$ ,  $\text{CO}$ ,  $\text{O}_3$ ,  $\text{SO}_2$ ,  $\text{NH}_3$ ,  $\text{C}_2\text{H}_6$ , and  $\text{NO}_2$ . Concentrations are randomly selected within a range based on the observed composition of atmospheres in our solar system. While cloud mixing ratios are calculated, clouds are ignored in our simulations due to the computational burden as even poor modeling efforts increase computational time by a factor of 50.

We use the Monte Carlo dropout approximation to produce predictive distributions over parameters  $\theta$ . Dropout is a common regularization technique in neural networks to prevent overfitting and allow for a more generalizable model [9]. It has recently been shown that applying dropout at both training and test time is equivalent to making a variational approximation to the posterior distribution over the network weights [6]. Each dropout mask removes a certain proportion of weights by setting them to zero during a forward pass. Therefore, multiple forward passes with different dropout masks for the same input gives a set of predictive samples that build a predictive distribution. Through implementing dropout both at training and test time, we are effectively sampling from the posterior over weights of the network. This distribution over the weights enables us to approximate a predictive distribution  $p(\theta|s)$  over the parameters of an exoplanet given an observed spectrum.

INARA is implemented in Python using PyTorch, and the source code and a Docker image are publicly available.<sup>2</sup> To interact with NASA Goddard PSG, the simulator at the core of our spectrum generation setup, we implemented a Python package called *pypsg*<sup>3</sup> that handles data generation in PSG format and http-based two-way communication with PSG servers. The INARA codebase covers the running of server instances for data generation, ML model training, and inference in a distributed fashion utilizing the Google Cloud infrastructure. For the generation of the data set of 3M parameter and spectrum pairs, we employed approximately 2,000 high-end VMs (groups of 16 INARA instances connected to one PSG node).

## 4 Preliminary Results, Discussion, and New Horizons

We performed a grid search over model architectures and training hyperparameters, exploring over 70 combinations of different architectures (linear regression, feed-forward neural networks, and CNNs), learning rates in [0.0001, 0.01], activation functions in {tanh, ReLU, ELU}, and optimization algorithms {ADAM, SGD, ADAdelta, RMSProp}. No dropout was used in this phase. Due to time constraints, model training was set to 64 epochs in all cases, using a training set of 110,000 parameter–spectra pairs. We used a mean square error (MSE) loss, and employed early stopping with a validation set of size 10,000 to avoid overfitting. 1D CNNs produced the best results, and we settled on a model with the configuration Conv1d(64)–tanh–MaxPool–Conv1d(64)–relu–MaxPool–Conv1d(128)–relu–MaxPool–Conv1d(256)–relu–FC(256)–relu–FC(12), which has approximately 18M trainable parameters.

Here we report results of the best 1D CNN model trained using 110,000 parameter–spectra pairs out of the 3M data set, leaving results with the full training set to future work. The prediction for 1,000

<sup>2</sup><https://gitlab.com/frontierdevelopmentlab/astrobiology/inara>

<sup>3</sup><https://gitlab.com/frontierdevelopmentlab/astrobiology/pypsg>

spectra is illustrated in Figure 1. H<sub>2</sub>O, CO<sub>2</sub>, O<sub>2</sub>, N<sub>2</sub> and CH<sub>4</sub> are shown in the five plots in Figure 1 top row, where each dot represents the average of 600 runs of our model with dropout for each planet. The details for predictive joint distributions for a random planet among those simulated is shown in the two bottom plots of Figure 1. The true value, indicated by the red star and the red line, falls within the predictive distribution for both parameters. Figure 2 in the appendix presents predictions for the full set of 12 molecules. INARA outperforms traditional Monte Carlo-based approaches by several orders of magnitude while computing a larger set of parameters and atmospheric molecules (Table 1).

Thanks to the computational resources we had access to, our present data set is the largest collection of rocky planet spectra to date. For the first time in ML atmospheric retrieval, we adopted Monte Carlo dropout [6], providing a predictive distribution comparable to the posterior distributions yielded by traditional, Bayesian approaches. Further investigation is necessary to determine how this predictive distribution compares to the posterior distributions of traditional methods. While we obtained good results, our search for the best model is incomplete, and a thorough exploration of different neural network architectures is desirable. In addition, a more detailed data set (i.e., in terms of wavelength, self-consistency, and the presence of clouds/hazes) could be used with INARA to generate more reliable and scientifically-informative models.

## Acknowledgements

This work originated at the NASA Frontier Development Lab (FDL), an accelerated research program focused on finding solutions for space-related scientific challenges using ML, with support by NASA, the SETI Institute, and industry partners Google Cloud, IBM, Intel, Nvidia, XPRIZE, KBRwyle, KX, Luxembourg Space Resources, and Lockheed Martin. We thank the NASA Astrobiology Institute for their support of the astrobiology challenges at FDL 2018; our amazing mentors for their fantastic support and guidance; Geronimo Villanueva at NASA Goddard Space Flight Center for offering extensive support in setting up PSG on Google Cloud; Sara Jennings, Shyla Spicer, and James Parr for organizing NASA FDL; the SETI Institute and NASA Ames Research Center for hosting us; and all industry partners involved in the program.

## References

- [1] T. S. Boyajian, K. von Braun, G. van Belle, C. Farrington, G. Schaefer, J. Jones, R. White, H. A. McAlister, A. Theo, S. Ridgway, et al. Stellar diameters and temperatures. III. Main-sequence A, F, G, and K stars: additional high-precision measurements and empirical relations. *The Astrophysical Journal*, 771(1):40, 2013.
- [2] T. S. Boyajian, K. Von Braun, G. Van Belle, H. A. McAlister, A. Theo, S. R. Kane, P. S. Muirhead, J. Jones, R. White, G. Schaefer, et al. Stellar diameters and temperatures. II. Main-sequence K-and M-stars. *The Astrophysical Journal*, 757(2):112, 2012.
- [3] I. J. M. Crossfield. Observations of Exoplanet Atmospheres. *Publications of the Astronomical Society of the Pacific*, 127:941, Oct. 2015.
- [4] L. D. Deming and S. Seager. Illusion and reality in the atmospheres of exoplanets. *Journal of Geophysical Research (Planets)*, 122:53–75, Jan. 2017.
- [5] Y. Fujii, D. Angerhausen, R. Deitrick, S. Domagal-Goldman, J. L. Grenfell, Y. Hori, S. R. Kane, E. Pallé, H. Rauer, N. Siegler, K. Stapelfeldt, and K. B. Stevenson. Exoplanet Biosignatures: Observational Prospects. *Astrobiology*, 18:739–778, June 2018.
- [6] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning*, pages 1050–1059, 2016.
- [7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.
- [8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [9] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- [10] T. K. Ho. Random decision forests. In *Third International Conference on Document Analysis and Recognition*, volume 1, pages 278–282. IEEE, 1995.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

- [12] N. Madhusudhan. Atmospheric retrieval of exoplanets. *Handbook of Exoplanets*, pages 1–30, 2018.
- [13] P. Márquez-Neila, C. Fisher, R. Sznitman, and K. Heng. Supervised machine learning for analysing spectra of exoplanetary atmospheres. *Nature Astronomy*, 2(9):719, 2018.
- [14] T. D. Robinson and D. C. Catling. Common 0.1 bar tropopause in thick atmospheres set by pressure-dependent infrared transparency. *Nature Geoscience*, 7(1):12, 2014.
- [15] L. A. Rogers. Most 1.6 earth-radius planets are not rocky. *The Astrophysical Journal*, 801(1):41, 2015.
- [16] E. W. Schwieterman, N. Y. Kiang, M. N. Parenteau, C. E. Harman, S. DasSarma, T. M. Fisher, G. N. Arney, H. E. Hartnett, C. T. Reinhard, S. L. Olson, V. S. Meadows, C. S. Cockell, S. I. Walker, J. L. Grenfell, S. Hegde, S. Rugheimer, R. Hu, and T. W. Lyons. Exoplanet Biosignatures: A Review of Remotely Detectable Signs of Life. *Astrobiology*, 18:663–708, June 2018.
- [17] J. Skilling. Nested sampling. In *AIP Conference Proceedings*, volume 735, pages 395–405. AIP, 2004.
- [18] C. Sotin, O. Grasset, and A. Mocquet. Mass–radius curve for extrasolar Earth-like planets and ocean planets. *Icarus*, 191(1):337–351, 2007.
- [19] C. J. ter Braak and J. A. Vrugt. Differential evolution Markov chain with snooker updater and fewer chains. *Statistics and Computing*, 18(4):435–446, 2008.
- [20] G. L. Villanueva, M. D. Smith, S. Protopapa, S. Faggi, and A. M. Mandell. Planetary spectrum generator: an accurate online radiative transfer suite for atmospheres, comets, small bodies and exoplanets. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2018.
- [21] I. P. Waldmann. Dreaming of atmospheres. *The Astrophysical Journal*, 820(2):107, 2016.
- [22] K. J. Zahnle and D. C. Catling. The cosmic shoreline: The evidence that escape determines which planets have atmospheres, and what this may mean for proxima centauri b. *The Astrophysical Journal*, 843(2):122, 2017.
- [23] T. Zingales and I. P. Waldmann. Exogan: Retrieving exoplanetary atmospheres using deep convolutional generative adversarial networks. *arXiv preprint arXiv:1806.02906*, 2018.

# Appendix

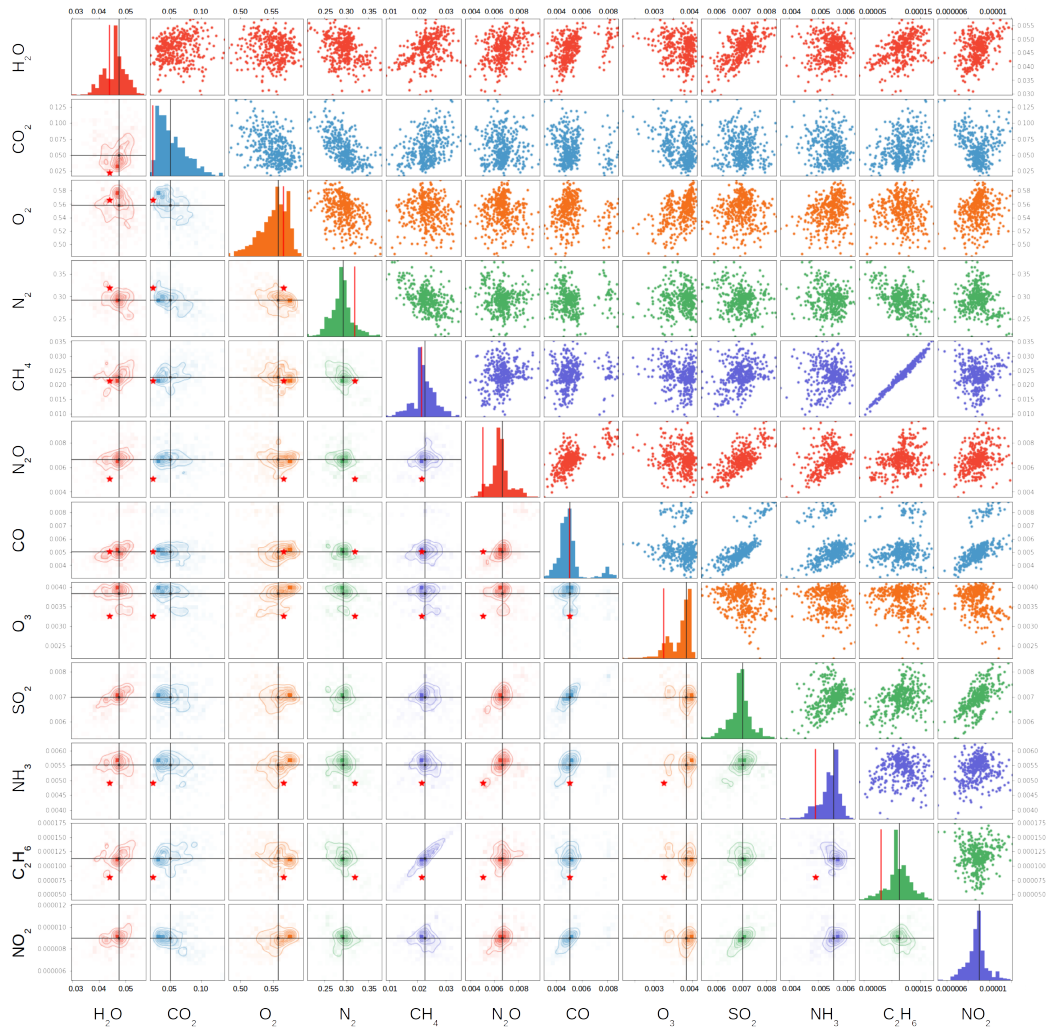


Figure 2: Posterior results with all 12 molecules in the model.