Studying the interior evolution of rocky exoplanets using machine learning

Chandan Kumar Sahu[1] **Maitrey Sharma**[1]
[1]School of Physical Sciences
National Institute of Science Education and Research, HBNI
Jatni-752050, India
chandankumar.sahu@niser.ac.in
maitrey.sharma@niser.ac.in

Abstract

With the advancements of observation techniques, many Earth-like exoplanets have been discovered. But, these techniques can only probe exoplanetary atmospheres and thus, the mechanisms of interior are difficult to quantify. Studying time-evolution of interiors of these exoplanets might give us a better understanding of how planets like Earth formed. We present a machine-learning based approach to investigate and predict time-evolution of interiors of rocky exoplanets using neural networks (NN).

1 Introduction

The past two decades have proved to be the heyday for exoplanetary science. Thousands of planets have been discovered and characterized, and now the focus is shifting towards habitability and life. Most exoplanets are too far from us and any direct mission is unfeasible in human lifetime. Thus, one can only analyse their atmospheric spectrum. However, the atmosphere is always in interaction with the interior. As the interior cannot be probed with any means as of now, we are left with analytical models to predict how interiors may affect the atmospheres. There are two approaches for developing models to study exoplanetary interiors. The first concerns with the prediction of the interior structure, that is, composition and radial variation of the interior and then studying how it interacts with the atmosphere. This approach has been investigated the most as of now (Bower et al. [2018], Lorenzo [2018], Suissa et al. [2018], Wang et al. [2019], Huang et al. [2022] and many more). The second approach involves looking at the big picture, i.e., the time-evolution of the interior structure. This approach is more comprehensive, and can provide insight how planets like ours, came into being. Very few analytical models exist which simulate time-evolution of exoplanetary interiors. Furthermore, they are limited in their working; with multiple factors coming into play, the complexity also increases.

In recent times, machine-learning based approaches are being used as substitute for analytical models as they are often less computationally intensive and might go beyond limitations. Recently, Zhao and Ni [2021, 2022] used mixture density networks (MDNs) to predict interior structures of rocky and gaseous exoplanets. This involved using analytical models of Lorenzo [2018] for rocky exoplanets and adapted equations of hydrostatic equilibrium and rotation for gaseous planets to train the MDN.

In this paper, we have tried to go with a similar approach for modeling the interior evolution of exoplanets. Here we describe a machine-learning surrogate model for interior evolution using the code package `VPLanet` first described in Barnes et al. [2020].

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

2 Exoplanet Interiors and VPLanet

VPLanet is a software package developed to simulate the fundamental aspects of planetary evolution over Gyr timescales, focusing on Earth-like habitable worlds (Barnes et al. [2020]). It can model the atmospheric, internal, orbital, rotational, stellar, and galactic processes and mimic existing results. In this work, we use the MagmOc, an interior evolution module of VPLanet which uses simple physics to model the evolution of a solar system like rocky exoplanet right from its formation to atmospheric desiccation. This module is based on the previous work by Schaefer et al. [2016] and Elkins-Tanton [2008], and includes various parameters such as stellar heat, radioactive decay, and tidal interactions.

The simulation starts with a completely molten mantle that solidifies with time, causing change in various physical properties such as mantle potential temperature, water content in melt and atmosphere, escape of gases, etc. MagmOc is coupled with the STELLAR code of VPLanet which couples the stellar evolution too with the planet evolution. The code computes the evolution of the mantle potential temperature, solidification radius, water mass in the solid, magma ocean and atmosphere, oxygen mass in the solid, magma ocean and atmosphere, atmospheric pressure due to H_2O and O_2, Mass fraction of FeO and Fe_2O_3 in the magma ocean, atmospheric net flux and the tidal and radiogenic heat sources. A typical run for the planet TOI776 b is shown in figure 1.

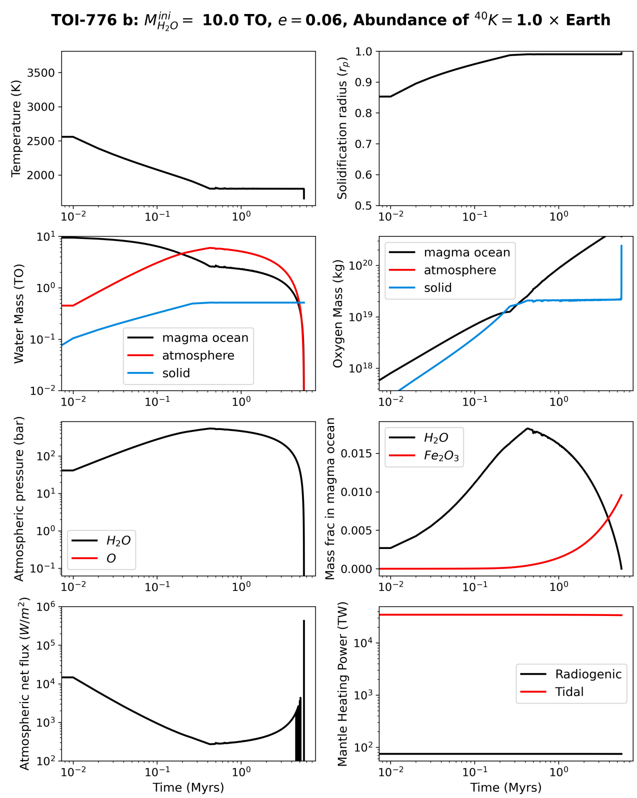

Figure 1: VPLanet *simulation for TOI776 b, a super earth planet with mass about 4 times of Earth and radius of 1.85 times of earth at a initial water mass of 10 TO. The figure shows how various properties of the planet must have evolved till the desiccation of the panet's atmosphere.*

The code can run the simulation till complete desiccation of the atmosphere or till the planet reaches habitable zone. However, it takes several minutes to run till the desiccation, and sometimes hours to simulate till the habitable zone, especially when the proportion of water is high or the planet is far from the host star. In these cases, the output file becomes large in size. We aim to produce a surrogate model to the code using machine learning to reduce the runtime of the models.

3 Neural Networks

Neural networks are a class of machine learning algorithms inspired by the human brain, designed to solve a variety of complex problems. They consist of interconnected layers of artificial neurons where each neuron receives input from one or more neurons in the previous layer, and produces an output that is transmitted to one or more neurons in the next layer. By adjusting the strength of these connections, the neural network can learn to recognise patterns in the input data and make predictions or classifications based on those patterns.

Multi-layer perceptron (MLP) is a type of neural network used for supervised learning. It is composed of input, hidden, and output layers of nodes, and each node is connected to every node in the next layer. The network uses weights and biases to adjust the strengths of connections between nodes, which are updated during training using an optimization algorithms. The main feature of using MLP in our model is that it can handle both linear and nonlinear relationships between inputs and outputs, making it a powerful tool for our work.

4 Implementation of the model

4.1 Dataset

Instead of using all output parameters of `MagmOc`, we selected five which are the most fundamental to govern the evolution of the system, viz., the mass of the star m_s (in M_{sun} units), the star-planet distance a (semi major axis of the orbit in AU), mass of the planet m_p (in M_{earth} units), radius of the planet r_p (in R_{earth} units), and the initial water content on the magma ocean m_w (in TO= terrestrial ocean units, 1TO=1.39×10^{21} kg). Note that r_p was calculated using values of m_p at constant density of 4000 kgm^{-3}. The density of the magma ocean melt is kept constant at a density of 4000 kgm^{-3}.

The complete dataset consists of three parts, first part based on the TRAPPIST planetary system ($m_s = \{0.1, 0.2\}; a = \{0.01, 0.03, 0.06\}; m_p = (0.5, 1, \ldots 3.5, 4); r_p$, and $m_w = \{1.0, 2.0\}$), second part based on our solar system ($m_s = \{0.8, 1.0, 1.2\}; a = \{0.1, 0.5, 1.0\}; m_p = (0.5, 1, \ldots 3.5, 4); r_p$, and $m_w = \{0.5, 1.0, 2.0\}$), and third part covers the intermediate values so that our model can be trained on most of the permutations to model Earth-sized exoplanets that can potentially host life.

The output of the simulation is stored in arrays containing the evolution of the parameters in each row in two separate text files for both the evolution of the star and the planet. We split the data randomly in the ratio of 80:20 for training:testing data.

4.2 Methodology

We started by training our datasets with a three hidden-layer MLP model from `sklearn` (Pedregosa et al. [2011]). The model contains 16 nodes in each layer. The input parameters include the five parameters mentioned in section 4.1 and an additional property whose evolution in the next step has to be evaluated. The activation function used in the model is `relu` with the `adam` optimizer. With a maximum iterations of 10000, the code optimizes for the given set of input parameters in few seconds. The code can be found at our GitHub repository.

We ran our model on the first dataset for trial basis as this was our smallest dataset in terms of size. Since each successive row of the output files represents a time-series evolution of the fate of planet, MLP was used to predict the second row of the evolution series given the five input parameters and the first row of the time-series (which is first 1000 years of evolution).

We train the model on to predict six planetary properties: (1) mantle potential temperature, (2) Net flux coming out of the atmosphere, (3) Pressure of water in the atmosphere, (4) mass of water in the magma ocean and the atmosphere combined, (5) mass of O_2 in the magma ocean and the atmosphere combined, and (6) the mass of H_2 escaping into space. The results are shown in figure 2.

When we trained the model to predict the 11^{th} row given the 10^{th} row, the accuracy improved significantly as shown in figure 3. We see that in figure 2, the accuracy of some of the properties is negative, indicating imperfect correlation between training and test data. However, in figure 3, we get a better correlation and higher accuracy solution. We claim that this behaviour occurs because initially, the variation in the evolving parameters is not effective due to which the model is not able to

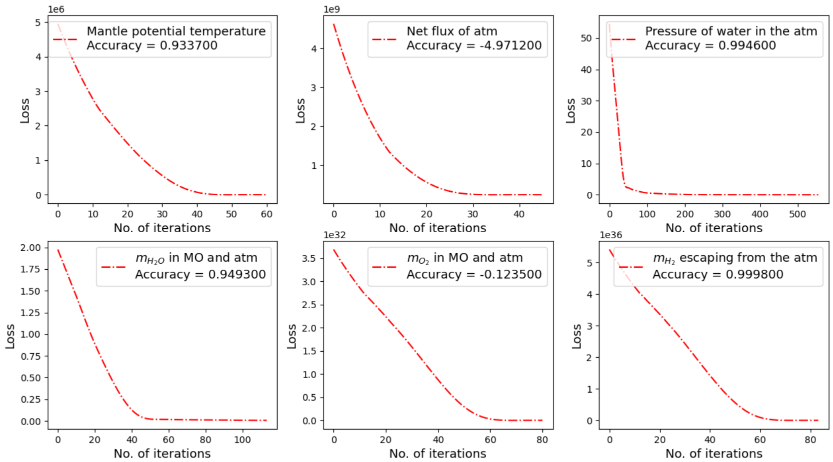

Figure 2: *Loss function vs iterations of various output parameters when trained on* 2^{nd} *row*

correctly predict the output. However, when it is trained on the later data, the variation in most of the parameters is at the peak, and hence the model gives a good output.

We also tried to apply RNN and LSTM as part of our baselines, but the structure of our data limited us as RNN are more commonly used for test and image data. However, we plan to explore RNN more efficiently in the next half of the semester.

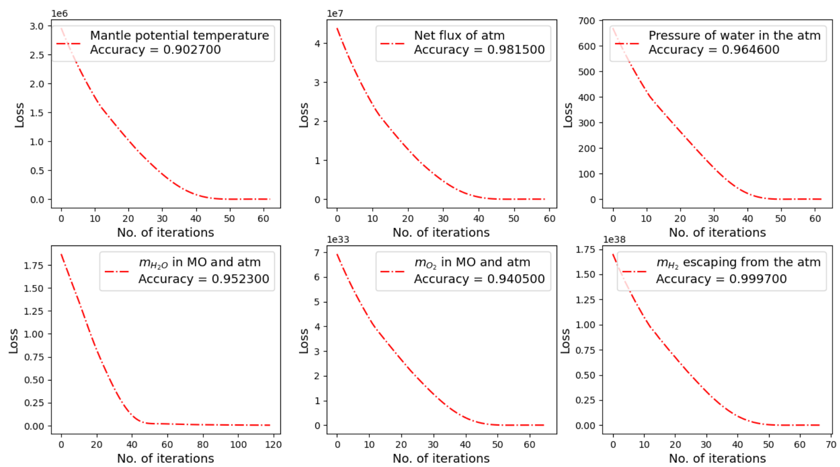

Figure 3: *Loss function vs iterations of various output parameters when trained on* 11^{th} *row*

5 Future works

The complete plan is to develop a model that gets the five input parameters and predict the evolution over the entire timescale of the planet. For this, we plan to train the model using MLP to predict the first few rows and proceed with RNN thereafter. This is because the RNN consumes more memory and lacks accuracy for uncertain data like ours. Moreover, as this is a time-series data, we have an extra component to take care while training the model.

In the second half of the semester, we plan to optimize all the parameters for the first row with increased accuracy using MLP and RNN for the entire dataset. We will also look for the possibility of exploring other NN algorithms if required. The next step would be to apply the complete model on dataset (3) and achieve higher accuracy. Thereafter we may explore for a more detailed dataset varying more input parameters (that have been fixed to Earth-like values presently).

References

Rory Barnes, Rodrigo Luger, Russell Deitrick, Peter Driscoll, Thomas R. Quinn, David P. Fleming, Hayden Smotherman, Diego V. McDonald, Caitlyn Wilhelm, Rodolfo Garcia, Patrick Barth, Benjamin Guyer, Victoria S. Meadows, Cecilia M. Bitz, Pramod Gupta, Shawn D. Domagal-Goldman, and John Armstrong. VPLanet: The Virtual Planet Simulator. *Publications of the Astronomical Society of the Pacific*, 132(1008):024502, February 2020. ISSN 0004-6280, 1538-3873. doi: 10.1088/1538-3873/ab3ce8. URL https://iopscience.iop.org/article/10.1088/1538-3873/ab3ce8.

Dan J. Bower, Patrick Sanan, and Aaron S. Wolf. Numerical solution of a non-linear conservation law applicable to the interior dynamics of partially molten planets. *Physics of the Earth and Planetary Interiors*, 274:49–62, January 2018. ISSN 00319201. doi: 10.1016/j.pepi.2017.11.004. URL http://arxiv.org/abs/1711.07303. arXiv:1711.07303 [astro-ph].

L. T. Elkins-Tanton. Linked magma ocean solidification and atmospheric growth for Earth and Mars. *Earth and Planetary Science Letters*, 271(1):181–191, July 2008. ISSN 0012-821X. doi: 10.1016/j.epsl.2008.03.062. URL https://www.sciencedirect.com/science/article/pii/S0012821X08002306.

Chenliang Huang, David R. Rice, and Jason H. Steffen. MAGRATHEA: an open-source spherical symmetric planet interior structure code. *Monthly Notices of the Royal Astronomical Society*, page stac1133, April 2022. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stac1133. URL http://arxiv.org/abs/2201.03094. arXiv:2201.03094 [astro-ph].

Alejandro M. Lorenzo, Jr. *ExoPlex: A New Python Library For Detailed Modeling Of Rocky Exoplanet Internal Structure And Mineralogy*. PhD thesis, September 2018. URL https://ui.adsabs.harvard.edu/abs/2018MsT.........25L. Pages: 25 Publication Title: Masters Thesis ADS Bibcode: 2018MsT.........25L.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Laura Schaefer, Robin D. Wordsworth, Zachory Berta-Thompson, and Dimitar Sasselov. PREDICTIONS OF THE ATMOSPHERIC COMPOSITION OF GJ 1132b. *The Astrophysical Journal*, 829(2):63, September 2016. ISSN 1538-4357. doi: 10.3847/0004-637X/829/2/63. URL https://iopscience.iop.org/article/10.3847/0004-637X/829/2/63.

Gabrielle Suissa, Jingjing Chen, and David Kipping. A hardcore model for constraining an exoplanet's core size. *Monthly Notices of the Royal Astronomical Society*, 476(2):2613–2620, May 2018. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/sty381. URL https://academic.oup.com/mnras/article/476/2/2613/4858395.

H S Wang, F Liu, T R Ireland, R Brasser, D Yong, and C H Lineweaver. Enhanced constraints on the interior composition and structure of terrestrial exoplanets. *Monthly Notices of the Royal Astronomical Society*, 482(2):2222–2233, January 2019. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/sty2749. URL https://academic.oup.com/mnras/article/482/2/2222/5127054.

Yong Zhao and Dongdong Ni. Machine learning techniques in studies of the interior structure of rocky exoplanets. *Astronomy & Astrophysics*, 650:A177, June 2021. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/202140375. URL https://www.aanda.org/10.1051/0004-6361/202140375.

Yong Zhao and Dongdong Ni. Understanding the interior structure of gaseous giant exoplanets with machine learning techniques. *Astronomy & Astrophysics*, 658:A201, February 2022. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/202142874. URL https://www.aanda.org/10.1051/0004-6361/202142874.