

## Learning Machine Learning with Lorenz-96

**Dhruv Balwada<sup>1</sup>, Ryan Abernathey<sup>1</sup>, Shantanu Acharya<sup>2</sup>, Alistair Adcroft<sup>3</sup>, Judith Brener<sup>12</sup>, V Balaji<sup>15</sup>, Mohamed Aziz Bhour<sup>4</sup>, Joan Bruna<sup>2, 13</sup>, Mitch Bushuk<sup>9</sup>, Will Chapman<sup>12</sup>, Alex Connolly<sup>4</sup>, Julie Deshayes<sup>10</sup>, Carlos Fernandez-Granda<sup>2, 13</sup>, Pierre Gentine<sup>4, 14</sup>, Anastasiia Gorbunova<sup>6</sup>, Will Gregory<sup>3</sup>, Arthur Guillaumin<sup>5</sup>, Shubham Gupta<sup>8</sup>, Marika Holland<sup>12</sup>, J Emmanuel Johnsson<sup>6</sup>, Julien Le Sommer<sup>6</sup>, Ziwei Li<sup>2</sup>, Nora Loose<sup>3</sup>, Feiyu Lu<sup>9</sup>, Paul O’Gorman<sup>11</sup>, Pavel Perezhogin<sup>3</sup>, Brandon Reichl<sup>9</sup>, Andrew Ross<sup>2</sup>, Aakash Sane<sup>3</sup>, Sara Shamekh<sup>4</sup>, Tarun Verma<sup>3</sup>, Janni Yuval<sup>11</sup>, Lorenzo Zampieri<sup>7</sup>, Cheng Zhang<sup>3</sup>, and Laure Zanna<sup>2</sup>**

**1** Lamont Doherty Earth Observatory, Columbia University **2** Courant Institute of Mathematical Sciences, New York University **3** Program in Atmospheric and Oceanic Sciences, Princeton University **4** Earth and Environmental Engineering, Columbia University **5** Queen Mary University of London **6** Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, INRAE, IGE, 38000 Grenoble, France **7** Ocean Modeling and Data Assimilation Division, Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici - CMCC **8** Tandon School of Engineering, New York University **9** NOAA Geophysical Fluid Dynamics Laboratory **10** Sorbonne Universités, LOCEAN Laboratory, Paris, France **11** Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology **12** National Center for Atmospheric Research **13** Center for Data Science, New York University **14** Columbia Climate School, Columbia University **15** Schmidt Futures

DOI: [10.21105/jose.0XXXX](https://doi.org/10.21105/jose.0XXXX)

### Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 01 January XXXX

Published: 01 January XXXX

### License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

## Summary

Machine learning (ML) is a rapidly growing field that is starting to touch all aspects of our lives, and science is not immune to this. In fact, recent work in the field of scientific ML, ie combining ML and with conventional scientific problems, is leading to new breakthroughs in notoriously hard problems, which might have seemed too distant till a few years ago. One such age-old problem is that of turbulence closures in fluid flows (described later). This closure or parameterization problem is particularly relevant for environmental fluids, which span a large range of scales from the size of the planet down to millimeters, and remains a big challenge in the way of improving forecasts of weather and projections of climate.

The climate system is composed of many interacting components (e.g., ocean, atmosphere, ice) and is described by complex nonlinear equations. To simulate, understand and predict climate, these equations are solved numerically under a number of simplifications, therefore leading to errors. The errors result from numerics used to solve the equations and the lack of appropriate representations of processes occurring below the resolution of the climate model grid (i.e., subgrid processes).

This book aims to conceptualize the problems associated with climate models within a simple and computationally accessible framework, and show how some basic ML methods can be used to approach these problems. We will introduce the readers to climate modeling using a simple tool, the Lorenz (1995) (L96) two-timescale model. We discuss the numerical aspects of the L96 model, the approximate representation of subgrid processes

(known as parameterizations or closures), and simple data assimilation problems (a data-model fusion method). We then use the L96 results to demonstrate how to learn subgrid parameterizations from data with machine learning, and then test the parameterizations offline (apriori) and online (aposteriori), with a focus on the interpretability of the results. This book is written primarily for climate scientists and physicists, who are looking for a gentle introduction to how they can incorporate machine learning into their work. However, it may also help machine learning scientists learn about the parameterization problem in a framework that is relatively simple to understand and use.

The material in this Jupyter book is presented over five sections. The first section, Lorenz 96 and General Circulations Models, describes the Lorenz-96 model and how it can work as a simple analog to much more complex general circulation models used for simulating ocean and atmosphere dynamics. This section also introduces the essence of the parameterization or closure problem. In the second section, Neural Networks with Lorenz-96, we introduce the basics of machine learning, how fully connected neural networks can be used to approach the parameterization task, and how these neural networks can be optimized and interpreted. No model, even the well parameterized ones, is perfect, and the way we keep computer models close to reality is by guiding them with the help of observational data. This task is referred to as data assimilation, and is introduced in the third section, Data Assimilation with Lorenz-96. Here, we use the L96 model to quickly introduce the concepts from data assimilation, and show how machine learning can be used to learn data assimilation increments to help reduce model biases. While neural networks can be great functional approximators, they are usually quite opaque, and it is hard to figure out exactly what they have learnt. Equation discovery is a class of machine learning techniques that tries to estimate the function in terms of an equation rather than as a set of weights for a neural network. This approach produces a result that is far more interpretable, and can potentially even help discover novel physics. These techniques are presented in the fourth section, Equation Discovery with Lorenz-96. Finally, we describe a few more machine learning in section five, Other ML approaches for Lorenz-96, with the acknowledgment that there are many more techniques in the fast-growing ML and scientific ML literature and we have no intention of providing a comprehensive summary of the field.

The book was created by and as part of M2LInES, an international collaboration supported by Schmidt Futures, to improve climate models with scientific machine learning. The original goal for these notebooks in this Jupyter book was for our team to work together and learn from each other; in particular, to get up to speed on the key scientific aspects of our collaboration (parameterizations, machine learning, data assimilation, uncertainty quantification) and to develop new ideas. This was done as a series of tutorials, each of which was led by a few team members and occurred with a frequency of roughly once every 2 weeks for about 6-7 months. This Jupyter book is a collection of the notebooks used during these tutorials, which have only slightly been edited for continuity and clarity. Ultimately, we are happy to share these resources with the scientific community to introduce our research ideas and foster the use of machine learning techniques for tackling climate science problems.

## Acknowledgements

This work is supported by the generosity of Eric and Wendy Schmidt by recommendation of Schmidt Futures, as part of its Virtual Earth System Research Institute (VESRI). MAB acknowledges support from National Science Foundation's AGS-PRF Fellowship Award (AGS2218197).

## References

Lorenz, E. N. (1995). Predictability: A problem partly solved. *Seminar on Predictability*, 1, 1–18. <https://www.ecmwf.int/node/10829>