

Peer Review

Review of: "Spatiotemporal Forecasting in Climate Data Using EOFs and Machine Learning Models: A Case Study in Chile"

Huihui Zhang¹

1. Humboldt Universität Berlin, Germany

Spatiotemporal Forecasting in Climate Data Using EOFs and Machine Learning Models: A Case Study in Chile

The topic is interesting and useful. However, I have to say I was somewhat disappointed with the investigation presented. It is recommended to include a flowchart in the Materials and Methods section to enhance clarity. Moreover, the Introduction section should be revised, as the current structure makes it challenging to follow the paper's logical flow. From my point of view, this manuscript should be considered for a major **review** before its probable publication.

Abstract

1. Formulas are unnecessary in the abstract and may complicate the reader's understanding.

Introduction

1. The logical flow is unclear. It would be helpful if the abstract included a comparison with related studies, particularly clarifying the specific problems your research addresses that previous works have not.

The motivation for this study appears to rest solely on the fact that it does not require distinguishing between data-driven modes and physically meaningful structures.

The introduction could include a summary of progress in related research. However, I cannot find this information.

1. "This study utilizes a grid comprising 6355 points at a resolution of degrees, covering the entirety of Chile. Each point is associated with a time series spanning from 1980 to 2022, encompassing

climatic variables such as daily accumulated precipitation, maximum, mean, and minimum daily temperature, evapotranspiration, etc.” It would be more appropriate to move this content to the “Data” section instead of including it in the “Introduction”.

2. “(MODWT) algorithm developed by [13]”, “The data were obtained from ERA5 in [14] ”

For improved clarity, it is recommended to cite the reference using the author’s name and year format rather than a numbered citation alone.

Materials and methods

-

1. “Oceanic variables” The authors did not use oceanic variables in this paper, so it would be better to remove them.
2. “Each point is associated with historical data containing time series of climatic variables such as maximum, mean, and minimum temperature, precipitation, evapotranspiration, etc.” What does ‘etc.’ refer to? Please include the complete list of variables.
3. Table 1 : It would be better to convert it into a three-line table and to place the table title above the table. Additionally, please add descriptions for the rows and columns. All tables have this problem, please revise them.
4. “460 strategically selected geographic locations from

the data grid and across Chile in search of similarity patterns.” It would be better to add a figure to show the 460 selected locations.

1. Page 6: “see Fig. 4 for an example”. Figure 4 is placed too far from the relevant sentence, which may interrupt the flow of reading.
2. Figure 1, The authors use data from 1980–2022 in the study, yet this figure includes data from 1953–2022 without a clear explanation.
3. “short to medium horizons”. How do authors define short and medium horizons?
4. Page 8, “Only records from the months of May, June, July, and August (MJJA), which correspond to the rainy season in Chile, were considered.” This sentence would be better placed in the research area description section. Please also add a reference to support it.
5. Figure 2, It would be helpful if the authors could explain why the period 2018–2022 was chosen, as the rationale currently feels a bit abrupt or unclear. In addition, the authors should clarify what the x

and y represent. Each figure is recommended to be self-explanatory, allowing readers to understand it without referring to other figures.

6. Figure 3: The labels include “All”, “0–500 m”, “500–1500 m”, and “>500 m”. However, only ‘0–500 m’ and ‘>1500 m’ appear in the figure. What about ‘500–1500 m’ and ‘>500 m’? Is it supposed to be ‘>500 m’ or ‘>1500 m’? Please revise for consistency and clarity.
7. Table 3, authors use several models in the results section without any introduction in the “Material and Methods” section.
8. The model performance on both the training and testing datasets is missing.
9. Table 3. Accuracy table for some autoregressive models. Do XGBoost and Random Forest belong to autoregressive models?

Declarations

Potential competing interests: No potential competing interests to declare.