New classes of climate model emulators to improve paleoclimate reconstructions

Auguste Gaudin[1], Myriam Khodri[1]

[1]LOCEAN-IPSL, Sorbonne-Université, CNRS/IRD/UPMC/MNHN, Paris, France

Correspondence to: Auguste Gaudin (auguste.gaudin@locean.ipsl.fr)

RC1: 'Comment on egusphere-2026-1337', Anonymous Referee #1, 25 Apr 2026

General

This study analyzes the limitations of the LIM-EOF model in traditional paleoclimate reconstruction methods. To address these issues, an improved climate model simulator is proposed, targeting the structural deficiencies of the traditional method with three specific improvements. The study evaluates whether these innovative measures enhance the simulation accuracy of large-scale climate variability, improve predictive capability, and reduce simulation errors in extreme events. The research is detailed, thoroughly elaborating on the dimensionality reduction algorithm and prediction model, describing the architecture and implementation of the simulator, and using extensive datasets and climate models to evaluate its multifaceted performance. Simulations based on the CMIP6 model suite demonstrate that the proposed improvements significantly outperform the traditional LIM-EOF model in terms of prediction accuracy, dynamic performance, avoidance of error accumulation, etc. In addition to a detailed analysis of the advantages of these improvements, the study also discusses certain limitations of the simulator under specific conditions, while proposing directions for further research. The work contributes notably to enhancing the accuracy of paleoclimate reconstruction and holds strong exploratory significance for advancing various research methods in the field.

RC1: Reply

We would like to thank the reviewer for their helpful comments and feedback on our manuscript. Below are the reviewer comments in bold text followed by our response. Additional figures supporting our responses are provided at the end of this document.

R1C1: When analyzing the correlation matrix and variance structure of latent vectors from EOF and AE, why are both 30-dimensional and 16-dimensional analyses conducted?

ANS: The dimensionality of the latent space used in a particle filter must be carefully calibrated: it should retain a sufficiently large fraction of the climate variability to limit particle degeneracy while remaining compact enough to avoid computational costs associated with high dimension problems. In line with this requirement, Jebri and Khodri (2023) used a truncated 30-dimensional space, which explains approximately 70 % of the variance. We therefore first conducted the analysis using 30 latent vectors, in order to remain consistent with this previous reference configuration.

However, as shown in Fig. 2, performance across most metrics tends to plateau relatively quickly. A 16-dimensional representation already provides good overall skill, suggesting that it offers a suitable compromise: it preserves substantial predictive ability while keeping the latent space sufficiently low-dimensional for future integration into particle-filter framework.

R1C2: To assess the impact of differences in physical mechanisms among models on simulation results, the study uses surface temperature data from 52 CMIP6 models. What aspects were primarily considered in selecting these data, and what characteristics make these data suitable for the evaluation work in this study?

ANS: To test the robustness of our emulators and to perform the sensitivity experiments, we applied them to a broad CMIP6 multi-model ensemble. We selected models from all CMIP6 modelling institutions represented in the available ESGF archive, using practical availability criteria: each retained model had to provide monthly near-surface air temperature (tas) for both a piControl simulation and at least one 165-year historical run. This selection provides substantial physical diversity, as illustrated in Fig. A11 by the large inter-model differences in the tropical Pacific nonlinearity coefficient α. A few models had to be excluded because of missing data or inconsistencies in the available files. Nevertheless, the final dataset preserves institutional coverage, with at least one model retained from each CMIP6 modelling institution initially considered. We also thank the reviewer for this question, as it led us to identify duplicated entries at the end of Table A1. This issue has been corrected in the revised manuscript.

We also clarify that a more restricted subset of models was used for the specific analysis addressing the question: "Can RC-based emulators adapt to models with varying degrees of nonlinearity?" This subset was not selected on the basis of predefined physical characteristics, but according to a practical requirement: the models had to provide a sufficiently large training dataset for the RC-based emulators. As shown in Fig. 6, RC skill is strongly sensitive to training-set size, and RC-based emulators require substantially more training data than LIM to reach stable performance. We therefore retained models with at least 10 historical ensemble members for this analysis, in order to avoid interpreting undertrained RC behavior as a physical model-dependent effect.

The subsequent comparison between IPSL-CM6A-LR and MIROC6 was used as an illustrative contrast between a model with relatively low ENSO nonlinearity, as measured by the α coefficient, and a model with higher ENSO nonlinearity.

R1C3: Could you summarize which specific CMIP6 models show significant improvements in prediction accuracy and strong adaptability due to the proposed improvements in this study, and provide the relevant constraints?

ANS: The first point is that, regardless of the CMIP6 model considered, AERCn provides significant skill improvements when the training dataset is sufficiently large. For example, among the eight models with 20 historical runs available for training, all show improved skill with AERCn compared with LIM (Fig R1.1). The spatial patterns of improvement, however, differ across models and appear to be related to model-specific predictability. For instance, IPSL-CM6A-LR, ACCESS-ESM1-5, and CanESM5 show broad AERCn skill gains over the tropical Pacific, especially in its eastern part, whereas in MIROC6, MIROC-ES2L, and the GISS models, the improvement is more localized over the western tropical Pacific.

Nevertheless, we do not conclude that AERCn is only applicable to a specific subset of CMIP6 models. Rather, we consider that AERCn can in principle be applied to any climate model, but that the construction of the training dataset is critical. Beyond having enough training data, it is preferable to train on past2k simulations, as done for IPSL-CM6A-LR, so that the emulator learns under forcing conditions closer to the target reconstruction period.

Finally, our results suggest that, across models, the largest source of predictability comes from the AE-based dimensionality reduction. This confirms that the way the climate state is represented is a key component of the proposed improvement over the classical LIM-EOF framework.

Supplementary figures for Reviewer 1

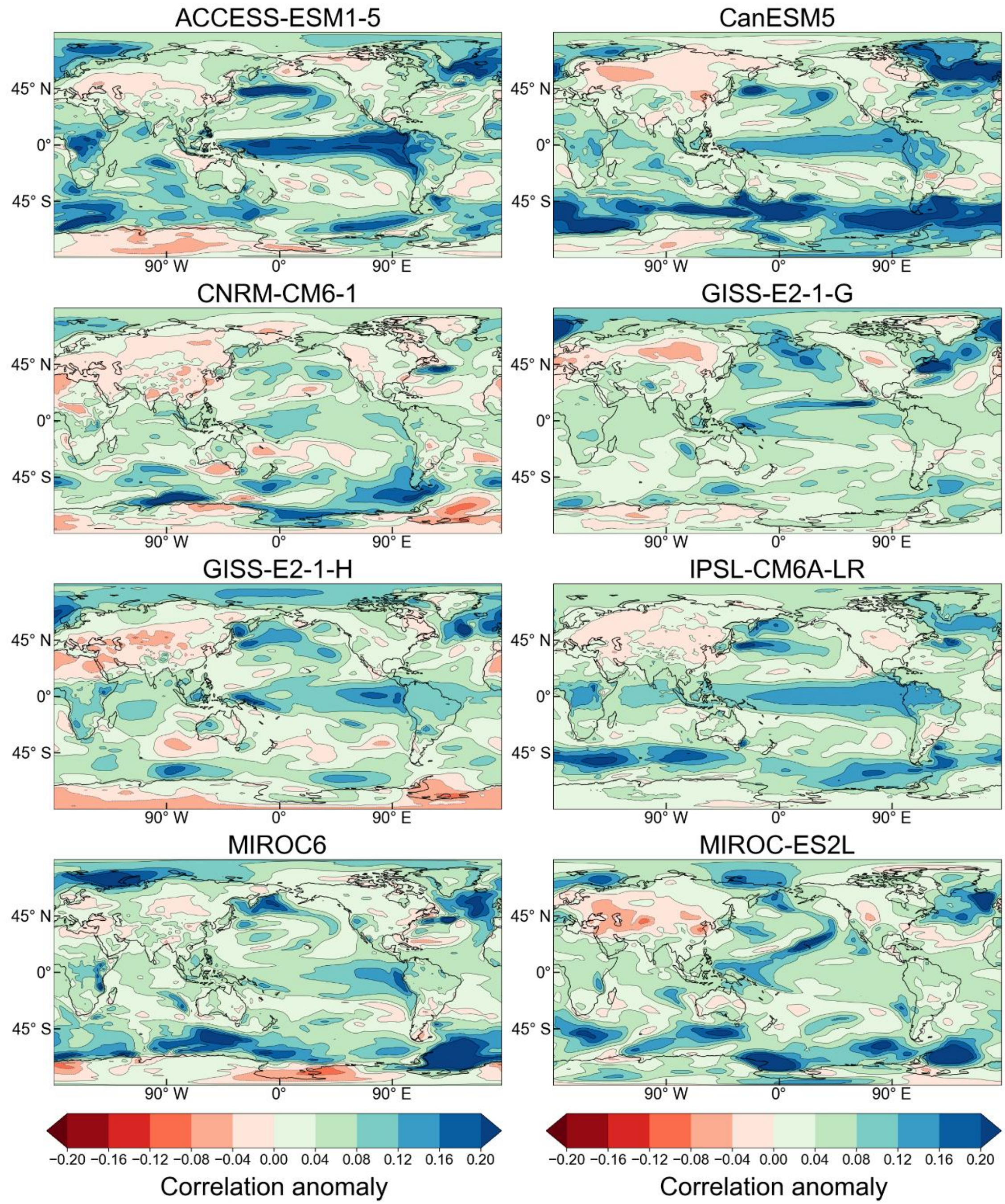

Figure R1.1: Spatial differences in one-year lead correlation skill between AERCn and LIM for the eight CMIP6 models with 20 training members. Positive values indicate regions where AERCn outperforms LIM.