Response to Reviewer#1

Manuscript number: https://doi.org/10.5194/essd-2025-717

Title: An AI-Driven Reconstruction of Global Surface Temperature with Emphasis on Refining the Antarctic Record

Comments from reviewer:

I thank the authors for the substantial revisions. The manuscript has improved considerably, in particular through the expanded validation and uncertainty analysis. I have only a few minor points remaining which would benefit from further clarification or some slight refinement before final acceptance.

Response: We sincerely thank the reviewer for the careful review of our manuscript and for providing detailed and constructive comments. The detailed responses and the specific revisions made to the manuscript are presented point-by-point below. In the revised manuscript, the text highlighted in blue indicates the revised content, the red strikethrough text indicates the deleted content, and the black text represents the original content.

Comments from reviewer:

1. The propagation of the HadSST ensemble members is a substantial improvement. However, the current text appears to treat this as a complete representation of the SST observational uncertainty. The authors should clarify more exactly which uncertainty components are represented by these ensemble members and whether spatially correlated uncertainty/covariance is explicitely accounted for. If this is not the case, this limitation should be stated clearly in the Methods and Discussion. Also, the reported uncertainty should be described as conditional on the adopted ensemble representation, rather than as a complete uncertainty budget.

Response: We have clarified the uncertainties represented by the HadSST ensemble members and explicitly described the unrepresented spatially correlated covariance errors in the Methods and Limitations sections. We have also revised the wording of the uncertainty to emphasize that it is conditional on the adopted ensemble representation.

Changes to the manuscript:

Line 199–201 in track change:

The HadSST4 ensemble members are generated using an observational error model that accounts for multiple sources of uncertainty, including Bucket measurement errors, ~~sampling~~ Engine-room intake measurement errors, and bias-adjustment uncertainties.

Line 231–236 in track change:

It should be noted that the total predictive uncertainty estimated in this study represents an approximation of the true predictive uncertainty and is conditional on the adopted ensemble representation. Although the observational uncertainty of the SST input data is propagated through the reconstruction model, the 200-member HadSST4 ensemble primarily represents the uncertainty associated with SST bias corrections and does not explicitly incorporate the spatially correlated error covariance information. Therefore, these spatially correlated covariance errors are not explicitly propagated in our uncertainty quantification framework.

Line 617–621 in track change:

A further limitation concerns the representation of predictive uncertainty. The uncertainty estimated in this study is conditional on the adopted ensemble representation and does not fully account for all sources of observational uncertainty. In particular, the spatially correlated error covariance of the HadSST4 ensemble and the observational uncertainties of C-LSAT2.1 are not explicitly propagated through the reconstruction framework. Therefore, the estimated total predictive uncertainty should be regarded as an approximate, potentially lower-bound estimate of the true uncertainty.

Comments from reviewer:

2. If the model is trained on near-surface air temperature while reconstructing a blended land-SAT/ocean-SST product

following Kadow et al's practice, the resulting mismatch of physical variables should be discussed at least, in particular close to coastlines and sea-ice margins. The treatment and interpretation of seasonal sea-ice-covered grid cells remains also somewhat unclear to me. If these issues cannot be adressed methodologically at this stage, they should at least be described more explicitely as limitations of the reconstruction.

Response: We have addressed this issue in Section 5 (Limitations and Future Perspectives).

Changes to the manuscript:

Line 610–616 in track change:

Another limitation concerns the potential mismatch between the physical variables used for model training and those represented in the final reconstruction. The AI model is trained primarily on near-surface air temperature, whereas, the final reconstruction is constructed as a blended land-SAT/ocean-SST product. This introduces a potential physical-variable mismatch between SAT and SST, which may be particularly relevant in coastal regions and near the seasonal sea-ice edge, where the distinction between land, ocean, and sea-ice surfaces can be important. Future work should investigate the effects of this SAT–SST mismatch and develop a more physically consistent treatment of land, ocean, and sea-ice-covered grid cells.

Comments from reviewer:

3. The ENSO evaluation in Section 3.1 should explicitly acknowledge closely related recent work. Qian et al. (2026) examined historical ENSO reconstruction using long-term Nino indices and associated spatial patterns, which closely overlaps with the evaluation framework adopted here. A direct citation and brief discussion of this study in Section 3.1, particularly in connection with Fig. 3, is therefore necessary for properly positioning the present analysis within the existing literature.

Response: We have cited the relevant literature and provided further clarification in Section 3.1.

Changes to the manuscript:

Line 304–306 in track change:

These results also corroborate the findings of Qian et al. (2026), who investigated historical ENSO variability and reconstruction using long-term Niño indices and their associated spatial patterns.