To Reviewer 3

The authors developed and applied two adaptive strategies within a paleoclimate data assimilation (PDA) framework. In particular, these two strategies relate to the observation error and to localization. The latter determines the extent to which local records influence remote regions. I found the manuscript interesting, as it proposes adaptive strategies to improve climate reconstructions using PDA. However, in its current form, the manuscript mainly presents these two strategies without drawing clear conclusions regarding their application to PDA frameworks. The evaluation is too limited to clearly quantify the benefits of the two approaches. In addition, by introducing new parameters to be estimated, and therefore increasing the number of degrees of freedom, these strategies also raise the potential issue of overfitting. This issue should be clearly stated, along with possible ways to assess and mitigate it.

In general, the authors should moderate some of their statements regarding the improvement over the standard framework. First, the improvement appears to be modest. Second, part of the improvement may result from overfitting. A more detailed evaluation (for instance focusing on other variables than sea surface temperature) as well as the evaluation of the potential overfitting occurrence are therefore required before robustly claiming that these strategies represent a clear improvement and provide substantial added value.

Answer: We thank the reviewer for these constructive and challenging comments. We have taken them very seriously and have made substantial revisions to address them. First, we have moderated our language throughout the manuscript, replacing subjective terms such as "extreme," "substantial," and "superior" with quantitative statements. For example, the abstract now reads: "...AOEI yields a mean RMSE reduction of $-0.0195°C$ but introduces local degradations (e.g., over the North Pacific), while HAOEI with optimal tuning ($\delta = 0.38$) gives a larger mean reduction ($-0.0253°C$, approximately 5.8%) with more spatially uniform improvements." Second, regarding the modest magnitude of improvements, we now explicitly discuss this in the revised manuscript. We note that the 5–6% RMSE reduction is consistent with or even exceeds the typical improvements reported in similar methodological studies in paleoclimate data assimilation. While the improvements are not dramatic, they are statistically robust and spatially consistent, which is valuable for applications where every increment in reconstruction accuracy matters (e.g., detecting forced climate signals against natural variability). Third, we acknowledge that the introduction of adaptive parameters increases the flexibility of the system and may raise the risk of overfitting. To mitigate this risk, we have: (1) used statistically motivated parameter choices (e.g., δ based on confidence intervals, h via Scott's rule) rather than arbitrary tuning; (2) systematically evaluated three different δ values (0.67, 0.38, and 0.13) to demonstrate that the performance trend is consistent and the optimal value is identifiable; (3) cross-validated our results against independent verification data (HadCRUT4.6) that were not used in the assimilation. We also note that the adaptive localization method using model-prior correlations (E_AL2) shows similar improvement patterns to the reanalysis-driven version (E_AL1), suggesting that the method's effectiveness is not dependent on overfitting to a specific reference dataset. Fourth, regarding the limited evaluation scope, we acknowledge that our study focuses on surface temperature over the tropics. We have added a sentence in the Conclusion and

Outlook section noting that future work should extend the evaluation to other variables (e.g., precipitation, atmospheric circulation) and to extratropical regions to assess the generalizability of the methods. We believe these revisions substantially strengthen the manuscript by providing a more balanced, quantitative, and critically reflective assessment of the proposed methods.

Specific comments:

1.The observation error needs to be described more clearly. Some would argue that the observation error cannot be considered as a tune parameter because it relates the uncertainties related to the observations. In a DA framework, I would argue that this error is not only related to the observational uncertainty itself, but also arises from the model-data comparison, as the model does not resolve all the processes that contribute to the proxy signal (in addition to the fact that most of PDA studies might underestimate the observational error in their system).

Answer: It is noticed in the introduction "However, this method typically underestimates true proxy uncertainty, as it fails to account for calibration error (such as non-stationarity between past and instrumental periodes), structural error (arising from missing physics, biology, or chemistry in the PSM), and representation error (due to the mismatch between point-scale proxies and gridded state variables).".

2.The chosen values for the threshold parameter δ should be justified.

Answer: We agree that the choice of δ requires justification. Recognizing that our original experimental setup was insufficient, we have supplemented the HAOEI experiments by expanding them to three groups with $\delta = 0.67$, 0.38, and 0.13 (see Table 1). Compared with the original manuscript, the new experimental results more clearly demonstrate a trend in which RMSE first decreases and then increases with varying δ, providing a more explicit comparison. Specifically, E_HAOEI(0.67) yields a mean RMSE reduction of -0.0210°C, E_HAOEI(0.38) achieves the largest reduction of -0.0253°C, while E_HAOEI(0.13) gives a slightly smaller reduction of -0.0208°C. Based on this comprehensive assessment, we selected $\delta = 0.38$ as the value that yields the optimal reconstruction performance. This is now discussed in detail in Section 3.2.1.

3.The introduction of the correlation coefficient at line 133 adds another layer of complexity and uncertainty. In particular, the relationship between two remote regions may vary substantially over time.

Answer: We thank the reviewer for raising this valid concern. The introduction of empirical parameters (bandwidth h and minimum radius $L_{\min}$) does add some complexity. However, we would like to clarify that these parameters are either data-adaptive or easy to set with clear physical justification, and they do not fundamentally hinder the generalizability of the method. 1. The bandwidth h is not free – it is determined by Scott's rule. In our kernel density estimation, the bandwidth h is not an arbitrary tuning parameter. We use Scott's rule:

$$h = n^{-1/6} \cdot \hat{\sigma}$$

where $\hat{\sigma}$ is the standard deviation of observation locations. This is a standard, automatic, and data-driven choice that balances bias and variance in density estimation. It does not require manual tuning by the user. Therefore, while h is a parameter in the mathematical sense, it is determined uniquely from the data and does not introduce additional user-specific uncertainty. 2. $L_{\min}$ is physically meaningful and easy to set. The minimum localization radius $L_{\min}$ is introduced to prevent the radius from becoming zero in very data-dense regions. Even in the densest proxy network, a zero radius would discard all observations, which is clearly undesirable. We set $L_{\min}$ to a small but reasonable value (e.g., 500 km in our experiments). This value can be guided by the typical decorrelation length scale of the climate variable (here, tropical temperature). For other variables (e.g., precipitation) or regions, a different $L_{\min}$ could be chosen based on known physical scales. The method does not rely on an exquisitely precise value; the results are stable over a range of reasonable choices, as we have verified through sensitivity tests (not shown for brevity). We will add a short paragraph in the revised manuscript (Section 2.3) to discuss the parameter choices and provide practical recommendations, acknowledging that any adaptive method involves some parameterization but that our choices are either data-driven or physically motivated. In summary, while our method does require setting a few parameters, they are either automatically determined from the data (h) or have a clear physical interpretation and can be set within a safe range ($L_{\min}$). Besides, we acknowledge that teleconnection strength can vary over time. However, our correlation-based weighting is implemented within an offline EnKF framework, where the background ensemble is static—drawn from climate model simulations covering the entire reconstruction period—rather than evolving through a continuous forecast cycle. Given the sparse nature of proxy data, estimating time-varying correlations from individual years would introduce substantial sampling noise and increase the risk of overfitting. We therefore compute correlations over the full 1880–2000 period as a robust first-order approximation, providing statistically stable estimates of the dominant covariability patterns (e.g., ENSO-related teleconnections). Importantly, the correlation-based weight is applied only to observations outside the adaptive localization radius, combined with distance-based weighting, and subjected to a threshold of 0.2 to exclude weak relationships. This conservative implementation limits the risk of introducing unstable adjustments. We have acknowledged this limitation in Section 2.4 and identified time-varying correlation estimation as a promising direction for future work.

4.It is unclear to me whether the PDA strategies used by the authors are time-dependent. Since the number of proxies varies strongly over time, using a time-dependent strategy would make sense. However, this could also increase the risk of overfitting, as mentioned above.

Answer: We thank the reviewer for this comment. In response, we have added a description of the offline data assimilation method in Section 2.1 of the revised manuscript. In the offline EnKF approach, the background ensemble is generated by random sampling from climate model simulations, and the same background ensemble is used across all assimilation years, i.e., the ensemble does not vary with the assimilation year. Therefore, although the adaptive observation error inflation and adaptive localization strategies are adjusted at each assimilation step based on the current innovation and observational density, the assimilation system does not involve temporal recursive dependence because the background field does not evolve over time. This implies that, even though the number of proxies varies with time, the adjustment of adaptive parameters only reflects spatial changes in observation distribution and does not accumulate additional degrees of freedom through temporal recursion. Consequently, under the offline framework of this study, the practical impact of time-dependent strategies on the degrees of freedom is limited, and the risk of overfitting remains relatively controllable. We have clearly stated this in the revised manuscript.