To Editor

1.When checking the paper for the further process, I have noticed that the doi dataset reference (LINK, xxx et al.) is missing (p. 17, l280). Please make sure that this reference is included in the data availability section and also in the reference list during the next revision.

2.Include the spelling of the acronyms in the abstract (AOEI and HAOEI). In the Code and data availability section please clarify that the datasets are hosted within the cfr and maybe add the link to it: cfr (https://zenodo.org/record/10575537).

3.Notification to the authors: The "Correspondence" section with the corresponding author's details (name and email address) is missing from the title page of the *.pdf manuscript. Please add this section.

4.Another editing issue I just realize: Editorship CI! The authors have the following competing interests: At least one of the (co-)authorsis a member of the editorial board of Geoscientific Model Development Can you correct to: The authors have the following competing interests: At least one of the (co-)authors is a member of the editorial board of Geoscientific Model Development

5. it is an interesting statement, could you clarify if it it is: one coauthor is a member of the editorial board of Geoscientific Model Development or Several coauthors are members of the editorial board of Geoscientific Model Development

6. in Financial support: Edit to: National Key Research and Development Program of China (grant no. 2023YFF0804703).

7.In Acknowledgements: We acknowledge financial support from the National Key Research and Development Program of China

Answer: Thank you for your feedback. I have made revisions to the above content.

To Reviewer 1

This manuscript proposes an adaptive observation error inflation and an adaptive localization for paleoclimate data assimilation. Since paleoclimate has large uncertainties for prior estimates and proxies, it is important to constrain the state by using the proxy information as much as possbile. Thus, covariance inflation and localization are essential components for paleoclimate data assimilation. The manuscript is well written. Please see my specific comments as below.

1. I think a flavor of EnKF is used for proxy data assimilation. Please give a brief introduction of the assimilation method, and how inflation and localization are applied within the assimilation framework.

Answer: We will discuss and then add an introduction to the EnKF method.

2. Sections 2.1 and 2.2, compared to AOEI, HAOEI introduces a new parameter delta to scale the inflation, to avoid too aggressive inflation. However, HAOEI needs an empirically determined parameter delta, which adds additional uncertainties.

Answer: We thank the reviewer for raising this important point. We acknowledge that the introduction of an empirically determined parameter δ in the HAOEI scheme may raise concerns about added uncertainty. Our response is as follows: 1. Statistical basis of δ: The choice of δ is not arbitrary. As noted in the manuscript, δ can be chosen based on statistical significance levels. For example, $\delta \approx 1.645$ corresponds to the 90% confidence interval of a standard normal distribution. This provides a theoretically grounded starting point, rather than purely empirical tuning. 2. Trade-off between flexibility and robustness: While δ introduces an additional degree of freedom, it also offers flexibility. Unlike AOEI, which applies a quadratic inflation to any large innovation (often too aggressive), HAOEI only inflates errors linearly when the normalized innovation exceeds a threshold. This makes HAOEI more robust and less prone to mis-inflating accurate observations. The parameter δ allows users to control how "conservative" or "aggressive" the inflation should be, depending on the confidence in the proxy data and model. 3. Managing the added uncertainty: We agree that any empirical parameter introduces some uncertainty. However, in practice, δ can be selected using cross-validation or calibrated against reference periods (e.g., the instrumental era) when available. In summary, while HAOEI introduces a parameter δ, this choice is statistically motivated, provides valuable flexibility, and can be managed through established calibration methods. The benefit of a more stable performance (compared to AOEI) justifies the added complexity. We will expand this discussion in the revised manuscript to make these points clear.

3. Although I personally can see the purpose of HAOEI, the foundation for HAOEI, or advantages of HAOEI over AOEI, needs more interpretation.

Answer: We thank the reviewer for this constructive suggestion. We will expand the manuscript to provide a clearer interpretation of the foundation of HAOEI and its advantages over AOEI. Below is a summary of the additions we plan to make. Foundational basis of HAOEI: HAOEI is rooted in robust statistics, specifically Huber's robust estimation (Huber, 1992). In paleoclimate data assimilation, proxy observations are sparse and subject to multiple, often poorly quantified errors. Large innovations may arise not only from observation errors but also from large model biases. AOEI inflates the observation error variance quadratically with the innovation, which can over-penalize observations in regions where the model prior is poor, leading to local degradation. HAOEI addresses this by first normalizing the innovation by the combined forecast and observation uncertainty. Only when this normalized residual exceeds a predefined threshold δ (e.g., corresponding to a 90% confidence interval) does it inflate the error variance, and the inflation is linear rather than quadratic. This follows the logic of Huber's estimationr: observations within the expected range are trusted, while those outside are down-weighted gradually, not abruptly or excessively. Advantages of HAOEI over AOEI: The key advantages of HAOEI become clear when comparing their behavior and experimental results. First, AOEI is highly sensitive because its inflation grows with the square of the innovation. This aggressive strategy can produce large local corrections where model and proxy strongly disagree, as seen in the equatorial ENSO region. However, it also risks mis-classifying accurate observations as outliers when the prior is poor, leading to new biases (e.g., over the North Pacific and Atlantic in Figure 2b). HAOEI, in contrast, applies a more moderate, linear inflation only after a threshold is crossed. This design makes it a robust stabilizer: it consistently improves performance across wider areas by avoiding extreme adjustments, even if it does not achieve the same peak correction in the most mismatched locations. Our quantitative results support this interpretation. The mean RMSE reduction of HAOEI with $\delta = 0.38$ (-0.0253°C) is smaller than that of AOEI (-0.0195°C). More importantly, the spatial pattern of improvement is more uniform with HAOEI, with fewer areas showing degradation. The CE metric also confirms that HAOEI delivers higher skill over a broader region. Thus, while AOEI may excel in isolated spots, HAOEI provides more smooth, reliable, and spatially consistent enhancement – a highly desirable property for paleoclimate reconstructions where data coverage is heterogeneous and model errors vary regionally. We will incorporate these points into the revised manuscript, specifically in Sections 2.2 and 3.2.1, to ensure readers fully understand the methodological rationale and practical benefits of HAOEI over AOEI.

4. Section 2.3, for both localization radius L based on KDE and correlation estimated from time series of variables and proxies, it is unclear why these two techniques are proposed, and why they are superior to GC and other existing adaptive localization methods.

Answer: We thank the reviewer for this important clarification request. We will revise Section 2.3 to more clearly explain the rationale behind the proposed adaptive localization scheme and its advantages over the standard Gaspari-Cohn (GC) function as well as other existing adaptive methods. In paleoclimate data assimilation, proxy data are extremely sparse and unevenly distributed. The conventional approach uses a fixed localization radius with the GC function, which tapers covariances based solely on physical distance. However, this method has two major limitations when applied to PDA: 1. Fixed radius ignores varying data density: In data-dense regions (e.g., Australia, Central America), a large fixed radius unnecessarily smooths fine-scale variability and introduces spurious correlations. In data-sparse regions (e.g., South America), a small radius would discard too much information, while a large radius risks including irrelevant remote proxies. An adaptive radius that responds to local observational density can better balance these competing needs. 2. Distance alone misses teleconnections: Important climate phenomena such as ENSO involve strong correlations over large distances. A purely distance-based cutoff (like GC) may either exclude these physically meaningful teleconnections (if the radius is too small) or include many spurious ones (if the radius is too large). By incorporating correlation-based weights, we can selectively retain remote observations that are statistically relevant while suppressing those that are not. Thus, we propose a two-component adaptive strategy: (i) a density-dependent localization radius to handle heterogeneous data coverage, and (ii) correlation-based weighting to identify and preserve genuine teleconnections beyond the distance cutoff. Compared to the standard fixed-radius GC function, our method is superior because it is spatially adaptive and information-aware. GC applies the same tapering everywhere regardless of how many proxies are available, which is suboptimal for heterogeneous networks. Our KDE-based radius shrinks in data-rich areas to preserve local details and expands in data-poor areas to gather more information. The correlation-based component further adds value by recovering remote signals that GC would completely zero out. Compared to other existing adaptive localization methods (e.g., Anderson 2007, 2012; Bishop & Hodyss 2009a,b; Ménétrier et al. 2015a,b), our approach is specifically tailored to the challenges of PDA. Many of those methods were developed for dense observing networks (e.g., satellite radiances or radiosondes) and rely on ensemble-derived statistics that can be poorly sampled when the ensemble size is small (here 100 members) and observations are few. Our method instead uses kernel density estimation (a purely geometric approach) and time-series correlations (which can be robustly estimated from the long reconstruction period, 1880–2000). This makes it less sensitive to the small ensemble sizes typical of PDA. Moreover, the hybrid formulation (Eq. 11) preserves the stability of distance-based localization within the adaptive radius while selectively adding distant, highly correlated observations – a feature not present in most existing adaptive schemes. In summary, our adaptive localization method is pragmatic, easy to implement, and specifically designed for sparse, unevenly distributed proxy networks. It directly addresses the two key weaknesses of fixed-radius GC (ignoring data density and teleconnections) and offers a practical alternative to more complex ensemble-based adaptive methods that may perform poorly with small ensembles. We will expand the discussion in Section 2.3 to include these points and cite relevant literature for comparison.

5. Moreover, as in my previous comment, there are empirical parameters introduced here too, as the bandwidth h, Lmin. Then more uncertainties are introduced, and it is hard to generalize the usage of such an adaptive method. **Answer:** We thank the reviewer for 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 ($L_{\min}$). We believe this does not undermine the generalizability of the method, especially given the clear performance improvements demonstrated.

6. Section 3.2.1, the choice of the parameter delta and associated impacts on the assimilation results needs be discussed here. Why the adaptive observation error inflation has impact on the region with large RMSEs around (20N 30E)?

Answer: We thank the reviewer for this specific and helpful comment. We have expanded the discussion in Section 3.2.1 to address both the role of δ and the regional impact around 20°N, 30°E. Below is the additional interpretation we have included. In HAOEI, δ controls the threshold for triggering inflation. To systematically evaluate its impact, we conducted experiments with three different δ values: 0.67, 0.38, and 0.13 (see Table 1 and Figures 3–5). A smaller δ makes the method more sensitive—more observations are flagged as potential outliers and receive linear inflation. However, the relationship is not monotonic: the RMSE first decreases and then increases as δ decreases. Specifically:

- E_HAOEI(0.67) yields a mean RMSE reduction of −0.0210°C (approximately 4.8%);
- E_HAOEI(0.38) achieves the largest reduction of −0.0253°C (approximately 5.8%);
- E_HAOEI(0.13) gives a slightly smaller reduction of −0.0208°C (approximately 4.7%).

The CE results show a similar pattern, with E_HAOEI(0.38) achieving the highest mean CE (0.1514), compared to 0.1275 for E_HAOEI(0.67), 0.1297 for E_HAOEI(0.13), 0.1201 for E_AOEI, and −0.2106 for the baseline E_20000. This indicates that $\delta = 0.38$ is the optimal in our setup: it is aggressive enough to correct large model-data mismatches, but not so aggressive that it mis-inflates marginally consistent observations. The existence of an optimal δ (rather than a monotonic improvement with decreasing δ) suggests that there is a trade-off between correcting genuine large errors and over-penalizing observations that are only moderately inconsistent with the prior. We have added this discussion to Section 3.2.1, along with a practical guideline that δ can be selected via cross-validation against reference data when available. The region roughly over North Africa and the adjacent Atlantic already exhibits relatively large RMSEs in the baseline experiment E_20000 (see Figure 3a), giving more room for improvement. In contrast, in regions where the model already performs well (e.g., the equatorial Pacific with dense proxy coverage), the adaptive inflation has little effect because the innovations are already small. We have included this explanation in the revised manuscript (Section 3.2.1).

7. Lines 213-216, do you use online or offline ensemble-based data assimilation? If it is offline, how does the imbalance occur?
Answer: We thank the reviewer for this clarifying question. Yes, we use an "offline" ensemble-based data assimilation approach. This is a standard practice in paleoclimate data assimilation (e.g., Hakim et al. 2016; Tardif et al. 2019), because proxy records are time-averaged (e.g., annual means) and the model climatology is used as the prior rather than a continuously cycled forecast. The term "unbalanced structures" in our discussion (lines 213–216) refers to "spatial imbalance" – rather than dynamical imbalance (such as gravity waves or geostrophic adjustment). If the localization radius is set too small, each grid point is influenced by only a very limited number of nearby proxies. In regions with sparse or uneven data coverage, this can produce abrupt spatial gradients, discontinuities, or checkerboard-like patterns that are physically unrealistic. To avoid confusion, we rephrased the sentences.

8. Furthermore, paleoclimate data assimilation is quite different from the convective-scale DA, due to sparse proxies and large uncertainties. Please explain why a small localization lengthscale expected?
Answer: We thank the reviewer for this important clarification question. We did not claim that a small localization lengthscale is expected. In our discussion, we noted that a very small fixed radius (1,000 km) leads to degraded spatial patterns (Figure 6d). This is because paleoclimate proxies are sparse, and a radius that is too small for the global network causes each grid point to rely on too few (or even zero) observations, producing noisy and fragmented analyses. Therefore, "for sparse proxy networks, a small localization radius is generally NOT expected or desirable" – this is exactly why PDA traditionally uses large radii. However, we did observe that among the fixed-radius experiments, the 1,000 km case produced the lowest RMSE (Figure 5). This seemingly contradictory result arises because RMSE is a local, point-wise metric: a small radius reduces spurious long-distance influences and yields better statistical accuracy at individual grid points, but it does so at the cost of distorting the large-scale coherent modes (e.g., ENSO pattern), as shown in Figure 6. This trade-off motivated our adaptive localization design: we aim to retain the "RMSE benefit" of smaller effective radii in data-dense areas while maintaining the "large-scale structural fidelity" offered by larger radii in data-sparse areas. A note that if the target variable were precipitation (which has shorter spatial correlation scales), a smaller radius might be more appropriate – but for temperature reconstruc-

tions, as in this study, larger scales are physically justified.

To Reviewer 2

This short study presents two potential improvements in paleoclimate data assimilation: adaptive observation error inflation and adaptive localization. The reasoning to propose those improvements is sound and the description of the methods and results is clear, except in some place as detailed below.

I still have one major concern. While the analyses and figures present relevant quantitative information, the discussion on the interest of the method is mainly qualitative and appears to overstate the impact of the proposed modifications. This should be changed in a revised version. For instance, the abstract and conclusion give no number. My understanding is that the proposed modifications only bring relatively modest improvements, reducing the errors of the order of a few percent only compared to the standard methodology for the majority of the diagnostics proposed. This is a key information for potential users to determine if implementing new methods for localization and error inflation worth the effort compared to other developments. Consequently, I consider that this quantitative information on the magnitude of improvements compared to the standard method should be discussed in more details, comparing the diagnostics to identify where this improvement is small or negligible and quantifying where it is considered more substantial. Some examples are given also below but this is general comments that is applicable for all the sections.

Specific points.

1. Lines 6-7. As mentioned in the general comment, a qualitative sentence like 'which yields significant extreme improvements in specific regions but introduces notable local biases, and Huber Robust Estimation (HAOEI) method, which provides more robust and spatially consistent enhancement overall ', should be supported by numbers to quantify the improvement. 'Extremes' is also too strong seeing the results of the experiments.
Answer: We will revise the sentence to include numbers and avoid "extreme". For example: "...AOEI yields a mean RMSE reduction of –0.0195°C but introduces local degradations (e.g., over the North Pacific), while HAOEI (δ =0.67) gives a slightly larger mean reduction (–0.0210°C) with more spatially uniform improvements, increasing CE by 2–3% over the baseline in most regions." We will add similar quantitative statements throughout the abstract and conclusions.

2. Line 30. EnKF has also limitations that could be mentioned briefly
Answer: We agree. We will add a brief sentence, e.g.: "However, EnKF also suffers from sampling errors due to finite ensemble size, assumes near-Gaussian error distributions, and requires covariance localization to suppress spurious long-range correlations – issues that are particularly acute in paleoclimate applications with sparse proxies."

3. In the method section, I was a bit confused with the usage of the variable x. I guess xf in equation 1 is the state vector (forecast) but it is not mentioned. Then xi is the position in equation 6 and xi,t in equation 10 the value of a variable at position i at time t but there I was lost and do not know the value of which variable.
Answer: In Eq. (1-2), x_f refers to the state vector. Meanwhile, in Section 2.3,

the variable x refers to the model grid points. Specifically, x_i denotes the i-th model grid point, and $x_{i,t}$ represents the value of the i-th grid point at time t. We will clearly define: In Eqs. (1) and (2), x_f is the state vector (forecast) containing all grid-point values of the variable (here, surface temperature). In Section 2.3, x_i denotes the i-th model grid point (spatial location), and $x_{i,t}$ is the temperature value at that grid point and time t. We will add these definitions right before Eq. (6) and Eq. (10).

4.For section 2.3, maybe add a sketch to illustrate more clearly the method?
Answer: We agree that a schematic would aid understanding. We will add a flowchart of the adaptive localization method in Figure 1.

5. Line 122. It is not clear the standard deviation of what variable.
Answer: We will clarify that σ is the standard deviation of the geographic locations (longitude/latitude) of all proxy observations used in the density estimation. The revised sentence will read: "where σ is the standard deviation of the observation locations (computed from their longitudes and latitudes)."

6. At the end of page 6, explain what is the cfr framework.
Answer: We will add a brief introduction at the end of Section 3.1: "which is a Python-based package that provides a suite of ensemble-based data assimilation and reconstruction tools, including EnKF implementation, proxy system models, and verification diagnostics. It is specifically designed for paleoclimate applications and is openly available."

7. Line 203 'superior performance' is mentioned but, if I see well, the improvement is very small (that would be useful to give percentages for instance). This should be mentioned.
Answer: We will replace "superior performance" with a quantitative statement: "The largest RMSE values in E_20000 are located in the equatorial ENSO region. E_AOEI produces a noticeable reduction in RMSE relative to E_20000 (mean difference = -0.0195 °C, approximately 4.5%). This improvement stems primarily from its aggressive error-inflation strategy, which strongly down-weights observations that deviate sharply from the model prior-most effectively in regions such as the ENSO zone where model biases are large. However, the same mechanism can also suppress accurate observations where the prior is poor, leading to local error increases, as seen over the North Pacific and Atlantic.The HAOEI method applies a more moderate, threshold-dependent inflation. E_HAOEI(0.67) yields widespread and modest improvements, with a mean RMSE difference of -0.0210 °C (a reduction of 4.8%). E_HAOEI(0.38) delivers a stronger overall error reduction (mean difference = -0.0253 °C, a reduction of approximately 5.8%) and outperforming both E_AOEI and E_HAOEI(0.67). However, E_HAOEI(0.13) still achieves some improvement (RMSE difference = -0.0208 °C, a reduction of 4.7%), its performance is inferior to that of E_HAOEI(0.38). Similar performance patterns are reflected in the CE results (Figure 4). While E_AOEI yields considerably higher CE values than E_20000 (0.1201 compared to -0.2106), tuning the δ parameter can adjust the skill, the mean CE values in E_HAOEI(0.67), E_HAOEI(0.38) and E_HAOEI(0.13) are 0.1275, 0.1514 and 0.1297, respectively, demonstrating a comprehensively better performance of E_HAOEI(0.38) overall."

8. Line 213-214. Again, the differences are small and should be quantified and discussed in the text.

Answer: We will rewrite that paragraph "As the localization radius grows, the RMSE decreases from 0.4377 in E_20000 to 0.4234 in E_5000, and increases again to 0.4307 in E_1000. This behavior is consistent with other studies (e.g., Zeng and Janjic (2016)). More importantly, regarding the leading EOF mode as shown in Figure 7, smaller radii significantly distort the spatial pattern of variability, although they explained variance values closer to those of the observations (66.7% in E_20000 and 27.3% in E_1000 compared to 27.4% in observations). This seemingly paradoxical behavior—where RMSE first decreases and then increases as the localization radius shrinks, while the EOF spatial pattern becomes progressively distorted—reflects the fundamental trade-off between sampling error and the retention of physically meaningful large-scale teleconnections in ensemble-based data assimilation. A large radius (e.g., 20,000 km) allows each grid point to be influenced by distant observations, preserving coherent large-scale structures (such as the ENSO pattern) and yielding a spatially reasonable EOF mode, but it also introduces spurious long-distance correlations from sampling noise, which inflates both the RMSE and the explained variance of the leading EOF. Reducing the radius to an intermediate value (e.g., 5,000 km) effectively suppresses many of these spurious correlations, leading to a lower RMSE and a more physical analysis. However, when the radius becomes too small (e.g., 1,000 km), the assimilation overly relies on local observations, cutting off genuine large-scale connections and fragmenting the spatial structure; although this yields a notably lower RMSE and an explained variance that matches the observations, the leading EOF mode becomes spatially distorted and physically unrealistic—demonstrating that numerical metrics like RMSE and variance explained are insufficient indicators of physical fidelity.'

9. Line 235. Please quantify 'considerably lower'. Considerably seems too strong from my understanding of the figures.

Answer: We will rewrite that paragraph "As shown in Figure 10 and Figure 11, E_AL1 outperforms E_20000, yielding smaller RMSE (E_20000: 0.4377°C; E_AL1: 0.4211°C; mean difference = -0.0166°C, approximately 3.8%) and generally higher CE (E_20000: -0.2106; E_AL1: -0.1179; mean difference = 0.0927). The improvements scale with data density: moderate in data-rich regions (e.g., Australia, Central America), more substantial in medium-density areas like the equatorial ENSO region, and largely neutral in sparse-data regions such as South America. Similarly, E_HAOEI0.38_AL1 achieves lower RMSE (0.4166°C) and higher CE (0.1638) than its adaptive-localization counterpart E_HAOEI0.38 (RMSE: 0.4233°C; CE: 0.1514), indicating that the combination of HAOEI and adaptive localization yields further improvements. Furthermore, Figure 12 reveals that the leading EOF modes of E_AL1 and E_HAOEI0.38_AL1 exhibit a spatial structure similar to observations but with amplified variability in the ENSO region. Although their explained variance (E_AL1: 42.7%; E_HAOEI0.38_AL1: ∼40%) remains substantially higher than observed (27.4%), it is considerably lower than that of E_20000 (66.7%, Figure 7, indicating a more faithful representation of the spatial variance distribution. Overall, the application of adaptive localization is beneficial for enhancing reconstruction quality, both in terms of statistical accuracy and the fidelity of dominant climate modes."

10. Line 251. Please quantify 'robust performance'.

Answer: We will revise the conclusion with specific numbers for performance.

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 δ = 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.