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.