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.