Geosci. Model Dev., 19, 7893–7909, 2026
https://doi.org/10.5194/gmd-19-7893-2026

Geoscientific Model Development
Open Access

Paleoclimate data assimilation with adaptive observation error inflation and adaptive localization

Ge Luo[1,2], **Yuefei Zeng**[1,2], **Feng Zhu**[3], **and Jiuwei Zhao**[1,2]

[1]State Key Laboratory of Climate System Prediction and Risk Management/Key Laboratory of Meteorological Disaster, Ministry of Education/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044, China
[2]School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China
[3]Climate and Global Dynamics Laboratory, NSF National Center for Atmospheric Research, Boulder, CO, USA

Correspondence: Yuefei Zeng (yuefei.zeng@nuist.edu.cn)

Received: 14 March 2026 – Discussion started: 1 June 2026
Revised: 29 July 2026 – Accepted: 16 August 2026 – Published: 25 August 2026

Abstract. Paleoclimate data assimilation methods significantly enhance the accuracy, spatiotemporal continuity, and global relevance of climate reconstructions by integrating Earth system models with proxy records. In this study, we further improve the algorithm by implementing two adaptive strategies – adaptive observation error inflation and adaptive localization – and systematically evaluate their performance in reconstructing temperature data over equatorial regions. For the adaptive observation error inflation experiments, two distinct methods were employed: the adaptive observation error inflation (AOEI) method yields significant improvements in specific regions but also introduces local biases, whereas the Huber Robust Estimation (HAOEI) method provides more robust and spatially consistent enhancements overall. In the adaptive localization experiments, the localization radius and weight matrix at each grid point are dynamically adjusted based on observational density and correlation information. This strategy effectively utilizes sparse observational data, suppresses spurious teleconnections, accurately reproduces the spatial structure of dominant climate variability modes, and thereby enhances the overall stability of the analyzed field.

1 Introduction

Reconstructing paleoclimate states is a crucial link in understanding the mechanisms and impacts of Earth system evolution. Currently, this field primarily relies on two methods: paleoclimate proxy records and Earth system model simulations. However, both approaches have significant limitations when reconstructing past climate states. On one hand, proxy records (such as tree rings, ice cores, speleothems, corals and etc.) play an indispensable role in reconstructing climate over the past millennia due to their long temporal span, but they often suffer from issues such as uneven spatiotemporal distribution, discontinuity, and sparsity. Furthermore, climate signals are easily contaminated by noise during preservation, making it extremely challenging to establish quantitative relationships between proxy indicators and the true climate state (Mann et al., 1998, 2008; Tingley et al., 2012; Liu et al., 2014; Guillot et al., 2015; Wang et al., 2015), and reconstructions of the same climate state from different proxy sources often show inconsistencies. On the other hand, climate models based on physical processes can provide physically consistent, globally complete climate fields with full spatiotemporal coverage, effectively capturing large-scale features of the climate system, however, due to inaccuracies in parameterizing physical processes and uncertainties in response to external forcings, model simulations may exhibit systematic biases, struggling to fully reproduce observed climate variability and leading to limited predictive skills (Widmann et al., 2010; Phipps et al., 2013; Kageyama et al., 2021). Therefore, to overcome the limitations of individual methods, paleoclimate data assimilation (PDA) techniques have emerged as a powerful tool to optimally merge information from proxy records – sparse, noisy, and indirect indicators of past climate – with the dynamical constraints of climate

Published by Copernicus Publications on behalf of the European Geosciences Union.

models (Von Storch et al., 2000; Hakim et al., 2016; Goosse, 2017). By doing so, the PDA method produces spatially complete and physically consistent climate field estimates, akin to reanalysis products for the instrumental era, while also quantifying reconstruction uncertainties. Among other methods, the Ensemble Kalman Filter (EnKF, Evensen, 1994), has gained prominence in PDA due to its easy implementation and capability to handle high-dimensional nonlinear systems (Hakim et al., 2016; Tardif et al., 2019; Osman et al., 2021; Sun et al., 2025). 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.

Despite the great potential of PDA, it faces unique challenges distinct from modern meteorological data assimilation, primarily due to the fact that proxy records are fairly sparse in time and space, and often represent time-averaged climatic signals (e.g., annual means) rather than instantaneous observations. Furthermore, proxy data are subject to complex errors arising from measurement inaccuracies, chronological uncertainties, and the imperfect relationship between proxy signals and the target climate variables. In the PDA, observation error variance is often specified empirically from the residual variance of a Proxy System Model (PSM, Evans et al., 2013; Dee et al., 2015; Zhu et al., 2023) calibrated against instrumental data. 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). To compensate, observation error inflation is frequently used (Sun et al., 2025). Given the strong spatial and temporal heterogeneity of proxy errors and information content, in PDA adaptive inflation methods are generally favored over a fixed inflation approach. While modern data assimilation has adopted techniques like Adaptive Observation Error Inflation (AOEI, Minamide and Zhang, 2017), its applicability to the specific challenges of the PDA remains insufficiently discussed and requires further exploration.

A critical factor for the success of the EnKF is its use of covariance localization. Early implementations typically employed a fixed localization radius (Houtekamer and Mitchell, 1998, 2001). Within this framework, a common approach is to taper the sampled covariances – between observations and model states, or between different model states – using a smooth, distance-dependent function. A standard choice for this function is the Gaussian-like Gaspari-Cohn (GC) function (Gaspari and Cohn, 1999). However, the Gaussian-like tapering function is not necessarily optimal as shown in several studies (Anderson, 2007; Lei and Anderson, 2014a, b). To address limitations of the fixed localization, a range of adaptive localization methods have been developed. Those methods adjust the localization often in response to the state-dependent correlation patterns or ensemble-estimated errors. Examples include techniques developed by Anderson (2007, 2012), Bishop and Hodyss (2009a, b), and Ménétrier et al. (2015a, b). However, these methods have yet to be applied to PDA context. In PDA, a large covariance localization radius is typically employed due to the sparse distribution of proxy records (Tardif et al., 2019). While necessary, an excessively large localization radius can introduce spurious long-distance correlations, ultimately degrading analysis quality. This limitation suggests that localization with a fixed radius is suboptimal for PDA applications. Instead, an adaptive localization, which can optimize the influence radius of observations based on the underlying spatial correlation structures, applying broader scales in regions with strong large-scale covariability and tighter constraints in data-rich or locally forced regions, is preferred. Currently, research on adaptive localization techniques in the PDA is still very limited.

In this paper, adaptive observation error inflation and adaptive covariance localization will be developed and explored within an ensemble-based PDA framework, aiming to create a more flexible and accurate assimilation system. As a methodological paper, a pseudoproxy experiment might be more informative than a real-word one. But we directly conduct real-world proxy experiments using a climate model to systematically evaluate whether these adaptive strategies can enhance reconstruction skill – particularly in sparse data regimes – improve uncertainty quantification, and provide a more generalizable approach for assimilating the heterogeneous and uncertain proxy records that characterize paleoclimatology.

2 Methods

2.1 Offline Data Assimilation

The offline EnKF is a widely used approach for paleoclimate data assimilation, popularized in the paleoclimate context by Steiger et al. (2014), Hakim et al. (2016), and Tardif et al. (2019). "Offline" means that the background ensemble is static – typically drawn from climate model simulations covering the entire reconstruction period – rather than evolving through a continuous forecast cycle. The state update equation is given by

$$\boldsymbol{x}_a = \boldsymbol{x}_b + \mathbf{K}\left[\boldsymbol{y} - \boldsymbol{y}_e\right], \tag{1}$$

where $\boldsymbol{x}_b$ is the ensemble prior. The vector $\boldsymbol{y}$ represents the assimilated proxy data, and $\boldsymbol{y}_e = H(\boldsymbol{x}_b)$ is the vector of proxy estimates derived from the prior through the forward operator H. $\mathbf{K}$ is the Kalman gain matrix:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right]^{-1}, \tag{2}$$

where $\mathbf{B}$ is the prior covariance matrix, $\mathbf{R}$ is the error covariance matrix of the proxy data, and $\mathbf{H}$ is the linearization of the forward operator H about the prior mean. The above Eq. (1) is solved using the ensemble square-root filter (EnSRF, Whitaker and Hamill, 2002). $\mathbf{R}$ is taken as a diagonal matrix (uncorrelated observation errors), with the diagonal elements representing the error variance for each assimilated proxy record (Tardif et al., 2019). This allows for serial processing of observations, in which observations are assimilated one at a time, greatly simplifying the implementation of covariance localization. For a single jth proxy data $\boldsymbol{y}_j$, the ensemble mean is updated by

$$\overline{\boldsymbol{x}}_a = \overline{\boldsymbol{x}}_b + \frac{w_{\text{loc}} \circ \operatorname{cov}\left(\boldsymbol{x}_b, \boldsymbol{y}_{\text{e},j}\right)}{\operatorname{var}\left(\boldsymbol{y}_{\text{e},j}\right) + R_j}\left(\boldsymbol{y}_j - \overline{\boldsymbol{y}}_{\text{e},j}\right), \tag{3}$$

where $\overline{\boldsymbol{y}}_{\text{e},j} = \overline{H(\boldsymbol{x}_b)}_j$ is the prior estimate of the jth proxy from the ensemble mean, R_j is the observation error variance for the jth proxy record, and cov() and var() denote the covariance and variance functions. Covariance localization is represented by the Schur product (denoted by $\circ$, i.e., element-wise multiplication) in the equation, which acts as a distance-weighted filter w_{loc} on the prior covariance matrix to suppress spurious long-distance correlations. The jth ensemble perturbation is updated by

$$\boldsymbol{x}_a' = \boldsymbol{x}_b' - \left[1 + \sqrt{\frac{R_j}{\operatorname{var}\left(\boldsymbol{y}_{\text{e},j}\right) + R_j}}\right]^{-1} \frac{w_{\text{loc}} \circ \operatorname{cov}\left(\boldsymbol{x}_b, \boldsymbol{y}_{\text{e},j}\right)}{\operatorname{var}\left(\boldsymbol{y}_{\text{e},j}\right) + R_j}\left(\boldsymbol{y}_{\text{e},j}'\right). \tag{4}$$

2.2 Adaptive observation error inflation

The AOEI was first introduced and systematically applied in the context of satellite radiance data assimilation for numerical weather prediction (Minamide and Zhang, 2017). For the jth proxy, it operates as in Eq. (5) by inflating the observation error variance when the squared innovation – the difference between the observation and the model simulation – exceeds the ensemble spread of the model simulation:

$$R_j = \max\left\{\sigma_{\text{o},j}^2, \left[\boldsymbol{y}_j - \overline{\boldsymbol{y}}_{\text{e},j}\right]^2 - \sigma_{\boldsymbol{y}_e,j}^2\right\}, \tag{5}$$

where $\sigma_{\text{o},j}^2$ is original observation error variance, $\boldsymbol{y}_j$ is the jth proxy value, $\overline{\boldsymbol{y}}_{\text{e},j} = \overline{H(\boldsymbol{x}_b)}_j$ is the prior estimate of the jth proxy from the ensemble mean, H is the observation forward operator, $[\boldsymbol{y}_j - \boldsymbol{y}_{\text{e},j}]$ is the innovation, and $\sigma_{\boldsymbol{y}_{\text{e},j}}^2$ is the ensemble spread in observation space.

The AOEI is designed to limit erroneous analysis increments where there are large representativeness errors in either the forecast model and/or from the observation itself. However, its aggressive adjustment strategy (inflation grows with the square of innovation) requires careful application in data-sparse regions or areas with uncertain observations to avoid amplifying noise.

2.3 Huber Robust Estimation

In this study, a new adaptive observation error inflation scheme based on the Huber robust estimation method (Huber, 1992) is introduced (hereafter called "HAOEI"), which aims to mitigate the undue influence of large outliers. This method computes a normalized innovation – the difference between the observation and the model forecast, normalized by the square root of the sum of the model forecast error variance and the baseline observational error variance – and compares it against a threshold δ. The baseline observational error variance is then inflated for values exceeding this threshold. For the jth proxy, the simplified formulation is as follows:

$$r_j = \frac{\boldsymbol{y}_j - \overline{\boldsymbol{y}}_{\text{e},j}}{\sqrt{\sigma_{\text{o},j}^2 + \sigma_{\boldsymbol{y}_{\text{e},j}}^2}}, \tag{6}$$

$$R_j = \begin{cases} \sigma_{\text{o},j}^2, & \text{for } |r_j| \le \delta, \\ \sigma_{\text{o},j}^2 \cdot \frac{|r_j|}{\delta}, & \text{for } |r_j| > \delta. \end{cases} \tag{7}$$

where $\sigma_{\text{o},j}^2$ is the uninflated observation error variance, $\boldsymbol{y}_j$ is the jth proxy value, $\overline{\boldsymbol{y}}_{\text{e},j} = \overline{H(\boldsymbol{x}_b)}_j$ is the model estimated proxy using the ensemble mean of climatological samples, H is the observation forward operator and $\sigma_{\boldsymbol{y}_e,j}^2$ is the ensemble spread in observation space. The threshold parameter δ defines the boundary between normal observations and potential outliers, typically chosen based on statistical significance levels (e.g., $\delta \approx 1.645$ corresponds to the 90 % confidence interval of a standard normal distribution). Observations with normalized innovations within $\pm\delta$ are considered statistically consistent with the forecast and retain their original error variance, while those beyond this range are treated as potential outliers and undergo error variance inflation.

Although δ introduces an additional degree of freedom, 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. In practice, δ can be selected using cross-validation or calibrated against reference periods (e.g., the instrumental era) when available. Compared to AOEI that inflates the observation error variance quadratically with the innovation, which can over-penalize observations in regions with poor model priors and thus lead 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. Therefore, the HAOEI 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.

2.4 Adaptive localization

Due to the uneven spatial distribution of proxy data, localization is required to constrain their spurious influence on remote regions. Within a fixed localization radius PDA framework, a weight matrix was introduced by Tardif et al. (2019) to adjust the covariance. This weight matrix is defined by the Gaspari-Cohn function (Gaspari and Cohn, 1999):

$$w_{ij} = \mathrm{GC}(d_{ij}, L) \tag{8}$$

where w_{ij} is the weight assigned to observation $\boldsymbol{y}_j$ when updating grid point $\boldsymbol{x}_i$, d_{ij} is the distance between grid point and observation, L is a specified cut-off (or localization) radius, and GC denotes the Gaspari-Cohn fifth-order polynomial:

$$\mathrm{GC}(z_{ij}) = \begin{cases} 1 - \frac{5}{3}z_{ij}^2 + \frac{5}{8}z_{ij}^3 + \frac{1}{2}z_{ij}^4 - \frac{1}{4}z_{ij}^5, & 0 \le z_{ij} \le 1 \\ -\dfrac{2}{3}z_{ij}^{-1} + 4 - 5z_{ij} + \dfrac{5}{3}z_{ij}^2 + \dfrac{5}{8}z_{ij}^3 - \dfrac{1}{2}z_{ij}^4 + \dfrac{1}{12}z_{ij}^5, & 1 < z_{ij} \le 2 \\ 0, & z_{ij} > 2 \end{cases} \tag{9}$$

where $z_{ij} = \frac{d_{ij}}{L/2}$.

This method 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, a large fixed radius unnecessarily smooths fine-scale variability and introduces spurious correlations. In data-sparse regions, 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). Thus, a two-component adaptive strategy is proposed: (1) a density-dependent localization radius to handle heterogeneous data coverage. (2) Correlation-based weighting to identify and preserve genuine teleconnections beyond the distance cutoff.

Observational density is estimated at each grid point using kernel density estimation (KDE) with a Gaussian kernel (Parzen, 1962). For a given grid point $\boldsymbol{x}_i$, the density is computed as:

$$\rho^*(\boldsymbol{x}_i) = \frac{1}{nh^2}\sum_{j=1}^{n} K\left(\frac{\|\mathbf{o}_i - \mathbf{o}_j\|}{h}\right) \tag{10}$$

where n is the total number of proxy observations, $\mathbf{o}_i$ denotes the location of the grid point $\boldsymbol{x}_i$, $\mathbf{o}_j$ denotes the location of the jth observation, $\|\cdot\|$ denotes the euclidean distance between two points, h is the bandwidth parameter, and $K(\cdot)$ is the Gaussian kernel function:

$$K(u) = \frac{1}{2\pi}e^{-\frac{u^2}{2}}. \tag{11}$$

The bandwidth h is determined using Scott's rule:

$$h = n^{-1/6} \cdot \hat{\sigma}, \tag{12}$$

where $\hat{\sigma}$ is the standard deviation of the observation locations (computed from their longitudes and latitudes). This choice provides a data-adaptive smoothing that balances bias and variance in density estimation, yielding a continuous density field. This density field is then normalized by its global maximum value to obtain $\rho(\boldsymbol{x}_i) \in [0, 1]$, representing the relative data density at grid point $\boldsymbol{x}_i$. Based on this, the localization radius $L(\boldsymbol{x}_i)$ at grid point $\boldsymbol{x}_i$ is defined as a function of this normalized density.

$$L(\boldsymbol{x}_i) = \begin{cases} L \cdot \left(1 - \rho(\boldsymbol{x}_i)^{1/2}\right), & \text{if } L \cdot \left(1 - \rho(\boldsymbol{x}_i)^{1/2}\right) \ge L_{\min} \\ L_{\min}, & \text{otherwise} \end{cases} \tag{13}$$

where L is the original fixed localization radius (attained in observation-sparse regions where $\rho(\boldsymbol{x}_i) \to 0$), the exponent $1/2$ controls the curvature of the density-to-radius mapping, and $L_{\min}$ is a lower bound enforced to maintain a minimum influence range even in observation-dense regions.

However, adjusting the localization radius solely based on density may further reduce the number of effective observations available for some grid points. While this helps suppress spurious teleconnections, it can also limit the use of already sparse paleoclimate proxy data, potentially undermining assimilation performance.

To better balance the use of observational information and the suppression of spurious correlations, we further incorporate correlation information into the weighting strategy, i.e., for observations outside the localization radius, their influence weight depends on the correlation ($w_{ij} \sim$ corr). The correlation coefficient r_{ij} between grid point $\boldsymbol{x}_i$ and observation $\boldsymbol{y}_j$ is computed as the Pearson correlation coefficient between their respective time series over the reconstruction period:

$$r_{ij} = \frac{\sum_{t=1}^{T}\left(\boldsymbol{x}_{i,t} - \overline{\boldsymbol{x}}_i\right)\left(\boldsymbol{y}_{j,t} - \overline{\boldsymbol{y}}_j\right)}{\sqrt{\sum_{t=1}^{T}\left(\boldsymbol{x}_{i,t} - \overline{\boldsymbol{x}}_i\right)^2 \sum_{t=1}^{T}\left(\boldsymbol{y}_{j,t} - \overline{\boldsymbol{y}}_j\right)^2}}, \tag{14}$$

where $\boldsymbol{x}_{i,t}$ is the value at grid point $\boldsymbol{x}_i$ at time t, $\boldsymbol{y}_{j,t}$ is the value at observation $\boldsymbol{x}_j$ at time t, $\overline{\boldsymbol{x}}_i$ and $\overline{\boldsymbol{y}}_j$ are the temporal means of the respective series, and T is the length of the time series. This coefficient quantifies the statistical relationship between distant locations and helps identify teleconnections that are physically meaningful rather than spurious.

The final weight w_{ij} assigned to observation $\boldsymbol{x}_j$ when updating grid point $\boldsymbol{x}_i$ is then defined as a hybrid of distance-based and correlation-based weighting:

$$w_{ij} = \begin{cases} \frac{1}{2} \cdot |r_{ij}|, & \text{if } d_{ij} > L_i \text{ and } |r_{ij}| > 0.2 \\ \mathrm{GC}(d_{ij}, L_i), & \text{else} \end{cases} \tag{15}$$

where d_{ij} is the distance between grid point $\boldsymbol{x}_i$ and observation $\boldsymbol{y}_j$, $L_i = L(\boldsymbol{x}_i)$ is the adaptive localization radius at grid point $\boldsymbol{x}_i$ computed from Eq. (13), r_{ij} is the correlation coefficient, 0.2 is the set expression threshold.

Figure 1 illustrates the flowchart of this method. Compared to the standard fixed-radius GC function, the new method is superior because it is spatially adaptive and information-aware. The GC applies the same tapering everywhere regardless of how many proxies are available, which is suboptimal for heterogeneous networks. The 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, 2012; Bishop and Hodyss, 2009a, b; Ménétrier et al., 2015a, b), this approach is specifically tailored to the challenges of the PDA. Many of those methods were developed for dense observing networks (e.g., satellite radiances or radiosondes). The new method 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. It is specifically designed for sparse, unevenly distributed proxy networks and directly addresses the two key weaknesses of fixed-radius GC (ignoring data density and teleconnections).

Finally, it is worth noting that h is a parameter in the mathematical sense, which is determined uniquely from the data and does not introduce additional user-specific uncertainty. $L_{\min}$ 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. $L_{\min}$ does not rely on an exquisitely precise value but it should be set to a reasonably small value that can be guided by the typical decorrelation length scale of the climate variables.

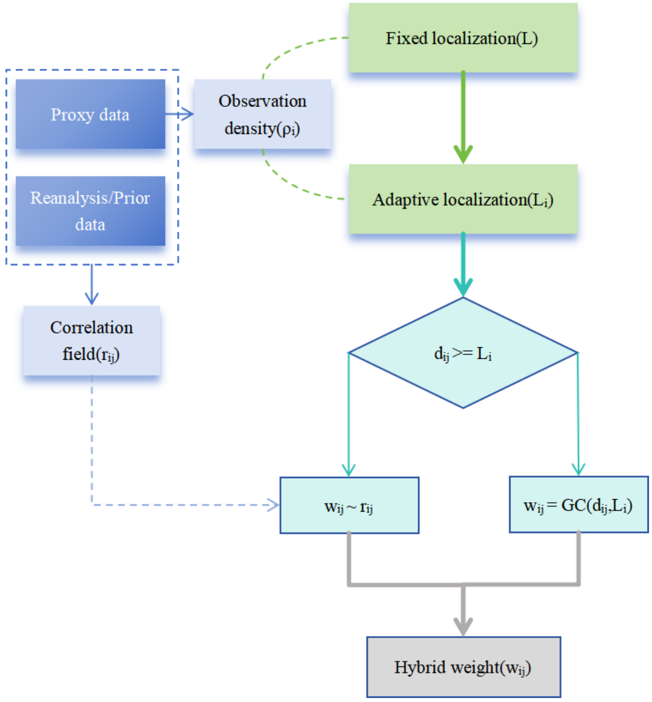

Figure 1. Flowchart of the adaptive localization method.

3 Experimental design and results

3.1 Experimental design

To evaluate the effectiveness of adaptive inflation and adaptive localization methods, a set of sensitivity experiments (Table 1) is conducted using the cfr framework (v2024.1.26; Zhu et al., 2024b), 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. The goal is to reconstruct tropical (30° S–30° N) surface temperature anomalies from 1880 to 2000, with respect to the 1951–1980 climatological baseline. The prior ensemble is drawn from the "iCESM1" last millennium and historical simulations (Brady et al., 2019), which provide physically consistent climate backgrounds spanning 850–2000 CE. Coral-based proxy records from the PAGES 2k Phase 2 database (PAGES 2k Consortium, 2017) – calibrated against NASA GISTEMP v4 (Lenssen et al., 2019) – are assimilated. These records are selected due to their strong temperature sensitivity and predominant distribution within the tropics (Fig. 2). Focusing on the tropics is motivated by two main reasons: first, the high temperature sensitivity of coral proxies enhances reconstruction reliability, whereas incorporating other proxy types (e.g., tree rings, ice cores) at a global scale could increase uncertainty due to their generally lower temperature sensitivity. Second, the tropics contain key climate systems such as ENSO, frequently studied in paleoclimatology, allowing for a clear assessment of methodological performance while reducing computational costs. It is important to note that the methodology presented here can, in principle, be extended to global applications. In the following, the impact of adaptive observation error is first assessed by comparing two inflation schemes – AOEI and HAOEI. Next, sensitivity experiments investigate the role of localization. Initial tests use fixed localization radii, which are then compared to two experiments employing the adaptive localization method described in Sect. 2.4 – one estimating proxy-state correlations based on reanalysis data GISTEMP v4, and the other one based on the model priors from "ICESM1" last millennium simulations. All data assimilation experiments cover 1880–2000, use a 1-year assimilation window, and an ensemble size of 100.

Three metrics are computed to evaluate the reconstruction analyses: the Root Mean Square Error (RMSE), the Coefficient of Efficiency (CE), and the Empirical Orthogonal Function (EOF). The RMSE for a given model grid point q is defined as:

$$\mathrm{RMSE}(q) = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{v}_{q,t} - \overline{\mathbf{x}}_{q,t}\right)^2}, \tag{16}$$

where n is the number of time samples, $\mathbf{v}_{q,t}$ is the verification data at grid point q and time t, and $\overline{\mathbf{x}}_{q,t}$ is the corresponding prior or posterior ensemble mean.

The CE evaluates the skill of the reconstruction relative to the climatological mean of the verification data. For a given model grid point q, it is calculated as:

$$\mathrm{CE}(q) = 1 - \frac{\sum_{t=1}^{n}\left(\mathbf{v}_{q,t} - \overline{\mathbf{x}}_{q,t}\right)^2}{\sum_{t=1}^{n}\left(\mathbf{v}_{q,t} - \overline{\mathbf{v}}_{q}\right)^2}, \tag{17}$$

where $\overline{\mathbf{v}}_q$ is the temporal mean of the verification data at grid point q. A CE value close to 1.0 indicates a highly accurate reconstruction, a value of 0 means the reconstruction is only as accurate as the climatological mean, and negative values suggest lower skill.

Beyond pointwise accuracy metrics, we also examine the ability of the reconstructions to capture the dominant large-scale modes of climate variability. For this purpose, we apply Empirical Orthogonal Function (EOF) analysis (Lorenz, 1956) to the reconstructed spatial fields The EOF method decomposes a spatiotemporal field $\mathbf{X}(s,t)$ into a set of orthogonal spatial patterns (EOFs) and their associated principal component (PC) time series. The leading EOF mode explains the largest possible fraction of the total variance, with each subsequent mode explaining the maximum remaining variance under the constraint of orthogonality to all previous modes. The corresponding PC represent the temporal evolution of these spatial patterns. In the following analysis, we compare the leading EOF modes – their spatial patterns, explained variance – obtained from different reconstruction methods against those derived from the verification dataset, to assess which method better recovers the key climate teleconnections and oscillatory signals.

As validation data, the spatially completed version of the near-surface air temperature and sea-surface temperature analyses product HadCRUT4.6 (Vaccaro et al., 2021) is employed.

3.2 Results

3.2.1 Adaptive observation error inflation

Figure 3 shows the RMSE for the surface temperature analyses in experiment E_20000, as well as the RMSE and percentage RMSE differences between E_20000 and E_AOEI, E_HAOEI(0.67), E_HAOEI(0.38), and E_HAOEI(0.13). 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 (Fig. 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. Hence, the choice of the threshold δ in HAOEI involves a fundamental trade-off: a value that is too small (e.g., 0.13) makes the method overly sensitive, flagging even moderate model-observation mismatches as outliers and down-weighting them, which leads to excessive loss of useful information and degraded analysis quality; conversely, a value that is too large (e.g., 0.67) makes the method overly conservative, affecting only the most extreme observations and failing to adequately control many genuinely large errors, thereby yielding only limited improvements. The optimal value (e.g., 0.38) strikes the best balance – maximizing the correction of harmful errors while minimizing the unnecessary loss of informative observations.

Figure 5 compares the spatial pattern and PC time series of the leading EOF mode derived from the observations with those from the reconstructed analyses of E_AOEI, E_HAOEI($\delta = 0.67$), E_HAOEI($\delta = 0.38$) and E_HAOEI($\delta = 0.13$). All experiments display a similar spatial structure to the observations, yet they show markedly

Table 1. Experimental setups, which are divided into two groups, sensitivity experiments on observation error inflation and on localization.

Group	EXP	Localization	Observation error inflation	Estimation of correlations in Eq. (14)
Adaptive observation error inflation				
	E_20000	20 000 km	No	–
	E_AOEI	20 000 km	AOEI	–
	E_HAOEI0.67	20 000 km	HAOEI, $\delta = 0.67$	–
	E_HAOEI0.38	20 000 km	HAOEI, $\delta = 0.38$	–
	E_HAOEI0.13	20 000 km	HAOEI, $\delta = 0.13$	–
Adaptive localization				
	E_20000	20 000 km	No	–
	E_10000	10 000 km	No	–
	E_5000	5000 km	No	–
	E_1000	1000 km	No	–
	E_AL1	(Adaptive)	No	Reanalyses
	E_AL2	(Adaptive)	No	Model priors
	E_HAOEI0.38_AL1	(Adaptive)	HAOEI, $\delta = 0.38$	Reanalyses

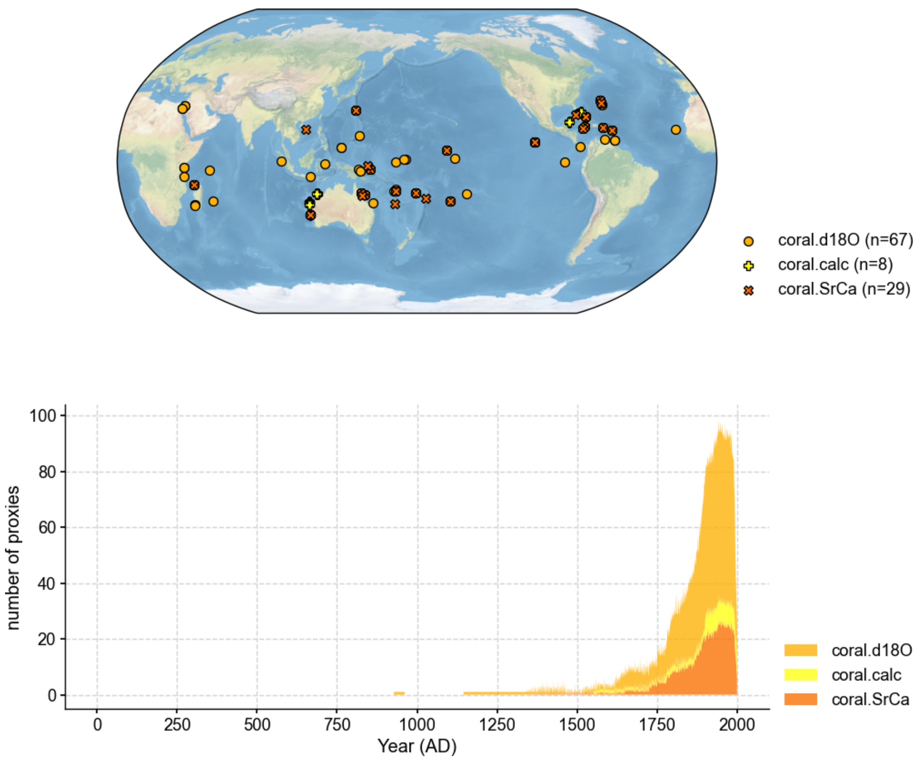

Figure 2. Spatial and temporal distributions of proxy records from the PAGES 2k dataset, visualized with cfr (Zhu et al., 2024b).

stronger variability in the ENSO region and produce a substantially higher explained variance. This discrepancy suggests that within the current assimilation framework, the reconstructed temperature field may be overly focused on a single dominant spatial mode, thereby failing to capture the more dispersed, multi-scale variability present in the actual observations. This effect may be partly due to the use of an excessively large localization radius, a point that will be further discussed in the following section.

3.2.2 Adaptive localization

Figure 6 presents the time series of RMSE for surface temperature analyses across experiments using different fixed localization radii (1000–20 000 km). 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 Janjić, 2016). More importantly, regarding the leading EOF mode as shown in Fig. 7, smaller radii significantly dis-

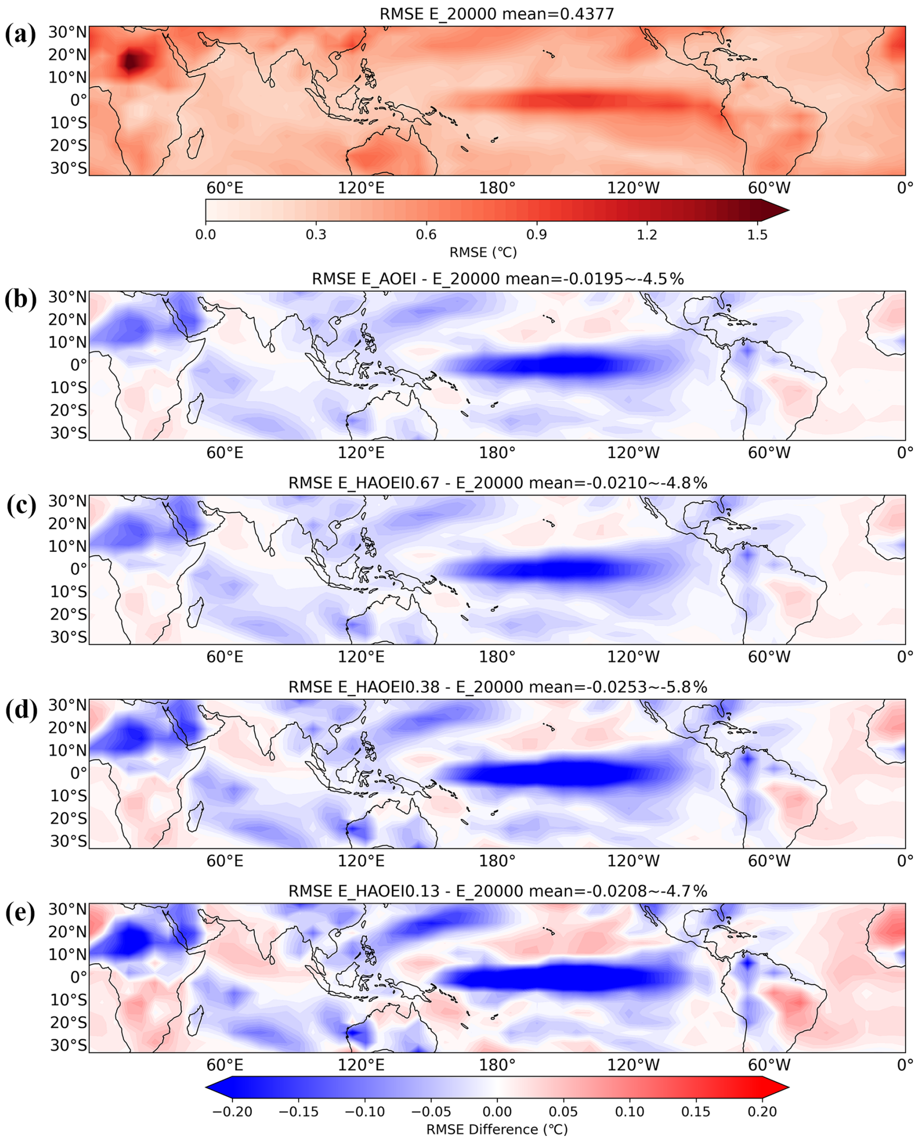

Figure 3. **(a)** The RMSE for the surface temperature analyses in E_20000, **(b)** the RMSE differences and the percentage RMSE differences between E_AOEI and E_20000, **(c)** between E_HAOEI0.67 and E_20000, **(d)** between E_HAOEI0.38 and E_20000, and **(e)** between E_HAOEI0.13 and E_20000.

tort 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., 5000 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., 1000 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.

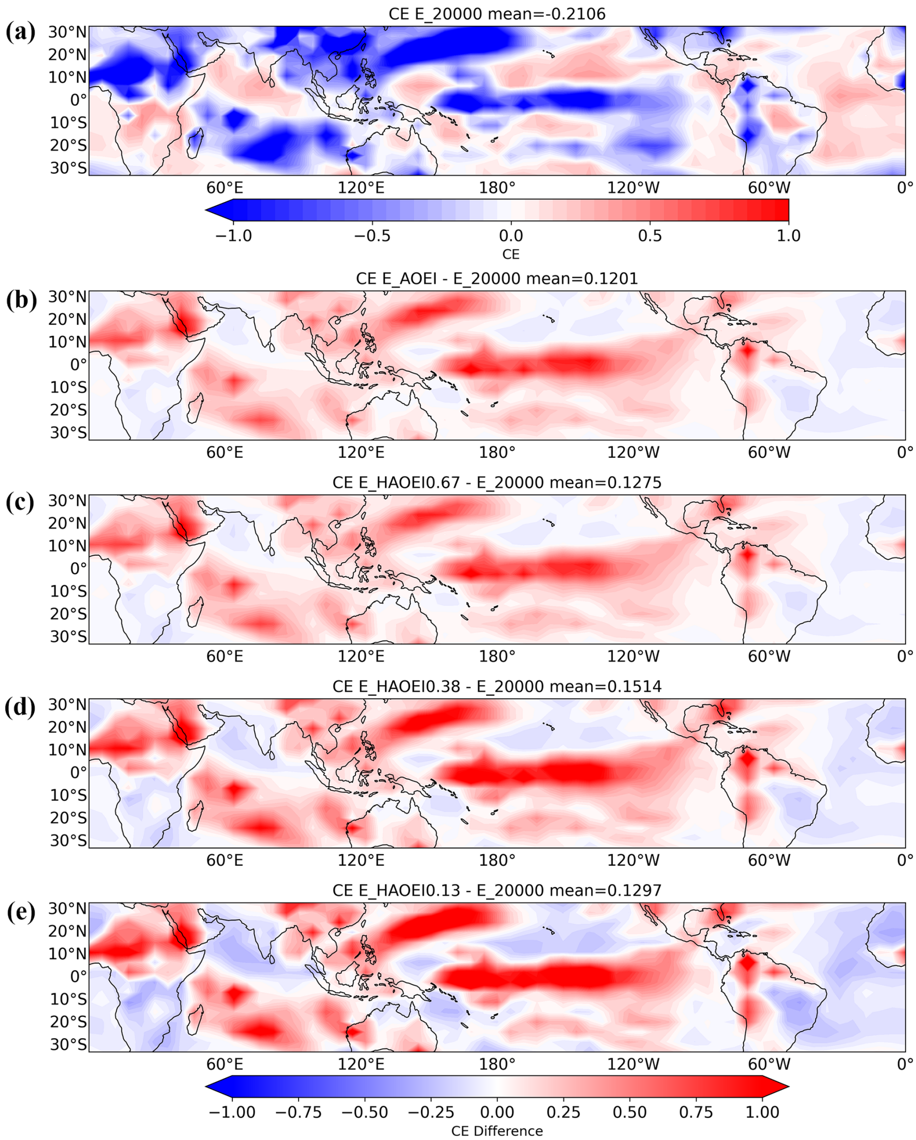

Figure 4. Same as Fig. 3 but for the CE.

To implement the adaptive localization method described in Sect. 2.4, the spatial density of proxy records is first estimated using kernel density estimation (KDE), as visualized in Fig. 8. While this density varies from year to year, the time-averaged distribution reveals distinct regional patterns: Australia and Central America are data-rich, the equatorial ENSO region exhibits medium proxy density, and areas such as South America remain relatively data-sparse. Furthermore, proxy-state correlations given in Eq. (14) are given by using the reanalyses and model priors, respectively. Figure 9 illustrates the correlations for one representative proxy record (ID: Ocn_090). Both estimated correlations reflect similar spatial structures but with detailed differences, e.g., the former one shows generally higher values and distinct discrepancies over Africa and the adjacent Indian Ocean region.

As shown in Figs. 10 and 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

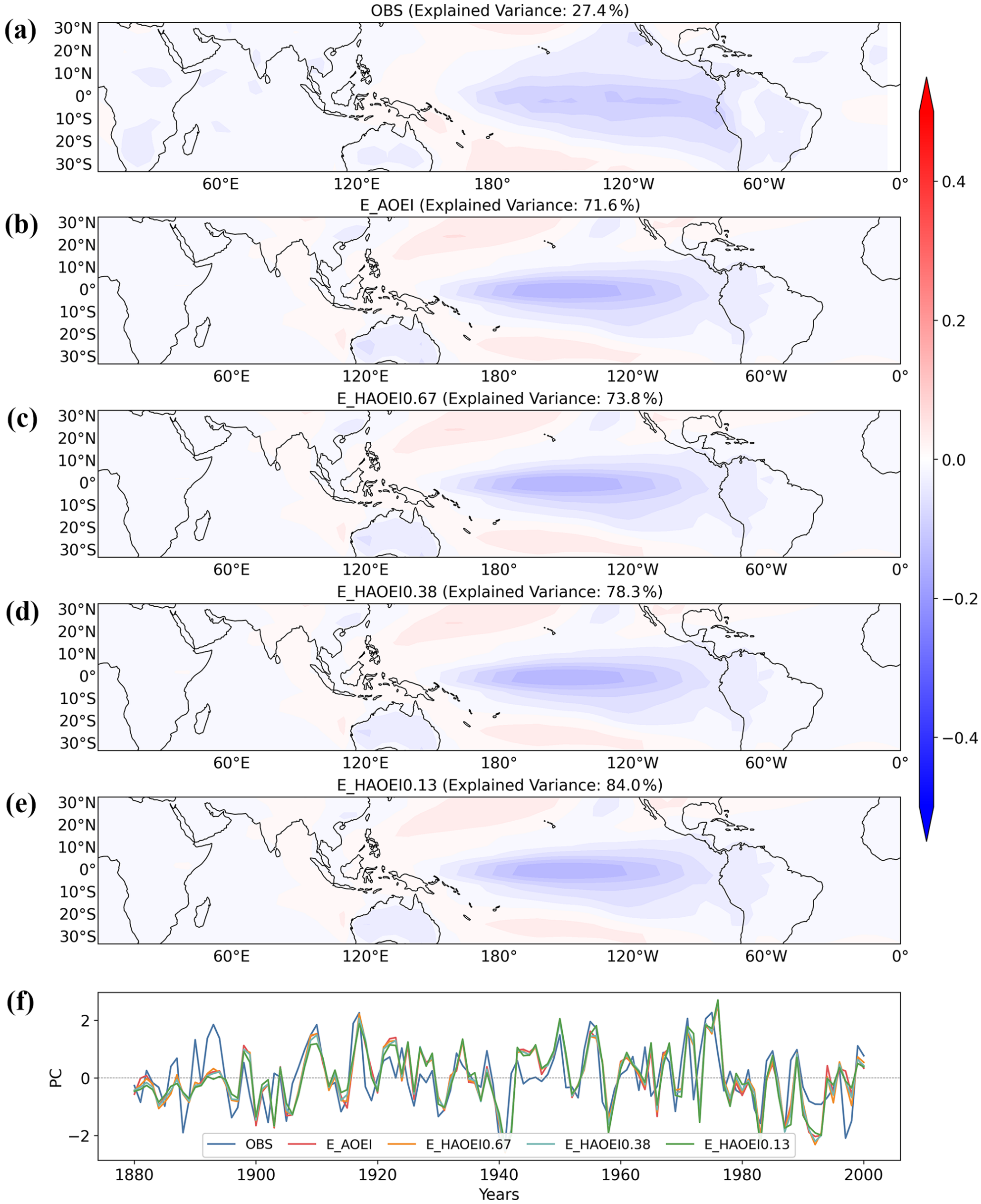

Figure 5. The leading EOF mode of **(a)** observations **(b)** E_AOEI analyses **(c)** E_HAOEI ($\delta = 0.67$) analyses **(d)** E_HAOEI ($\delta = 0.38$) analyses **(e)** E_HAOEI ($\delta = 0.13$) analyses **(f)** corresponding PC time series for all experiments.

adaptive localization yields further improvements. Furthermore, Fig. 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 %, Fig. 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.

Finally, Fig. 13 illustrates that adaptive localization using model-prior correlations E_AL2 improves upon the fixed-radius baseline E_20000, as evidenced by its lower RMSEs and higher CEs, even if it does not match the performance of the reanalysis-driven version E_AL1. This demonstrates the intrinsic value of the model prior: its physically constrained correlations offer a viable and effective guide for adaptive localization in the absence of reanalysis data. Consequently, this approach ensures the method's applicability to pre-industrial and deeper-time paleoclimate periods, where observational constraints are otherwise unavailable.

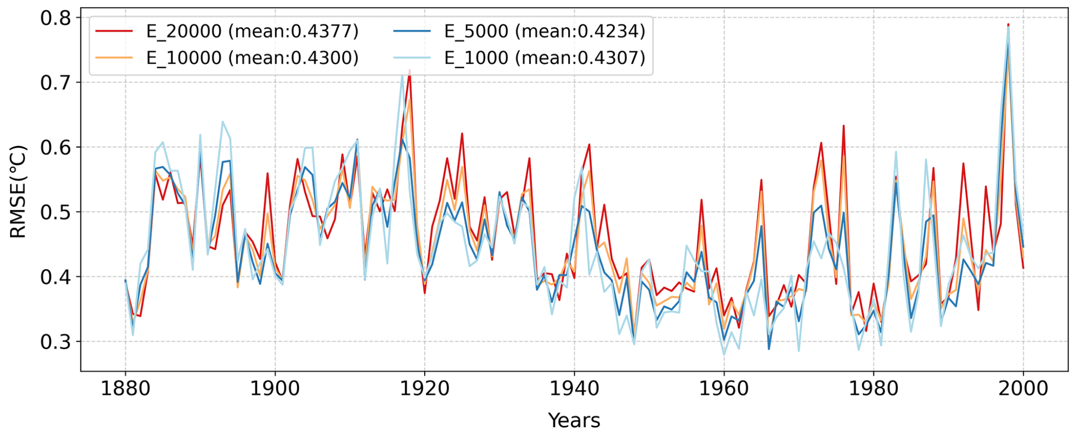

Figure 6. The time series of the RMSE for the surface temperature analyses in experiments with varying fixed localization radii.

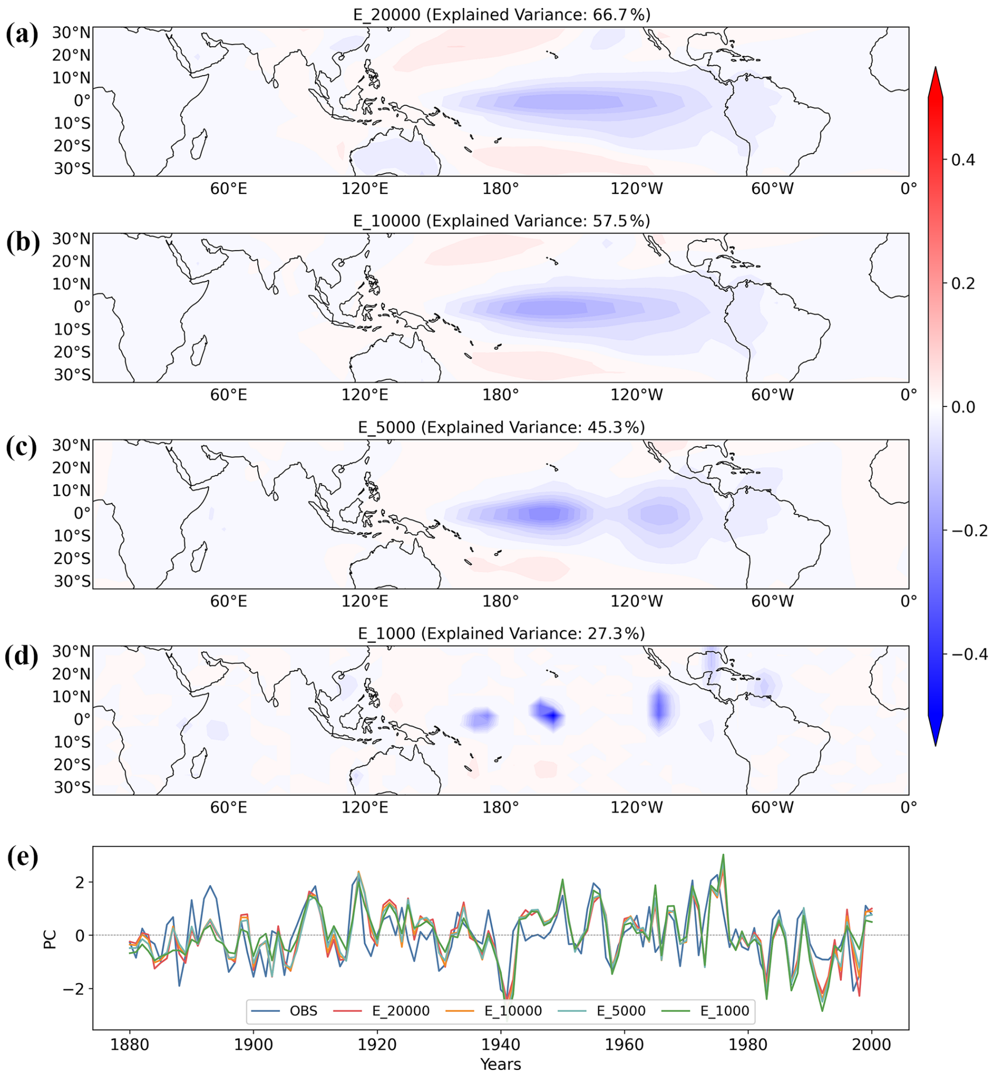

Figure 7. The leading EOF mode of the tropics: **(a)** reconstructed EOF field with 20 000 km localization radius; **(b)** reconstructed EOF field with 10 000 km localization radius; **(c)** reconstructed EOF field with 5000 km localization radius; **(d)** reconstructed EOF field with 1000 km localization radius; **(e)** corresponding PC time series for all experiments. All results are based on the period 1880–2000. EOF units are °C.

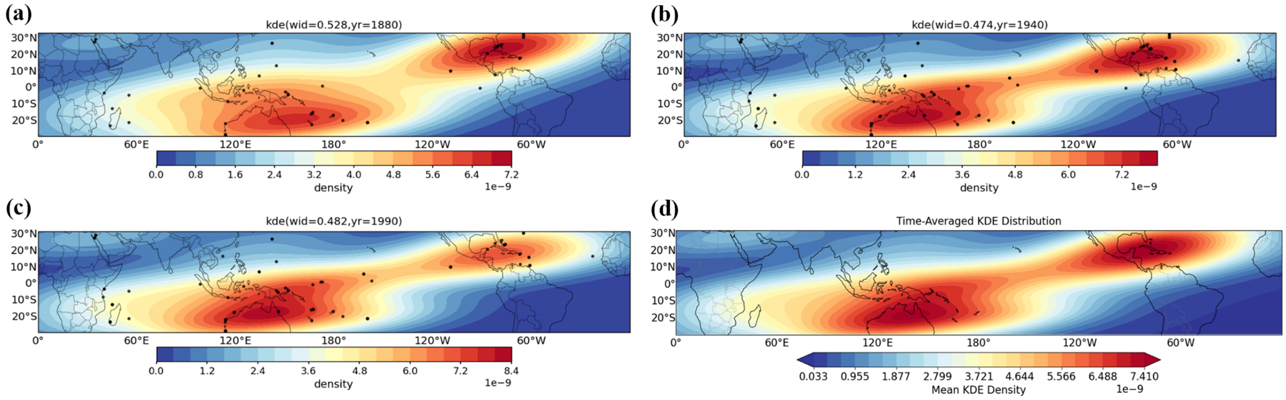

Figure 8. Gaussian kernel density estimation maps for different years: **(a)** 1880; **(b)** 1940; **(c)** 1990; **(d)** time averaged.

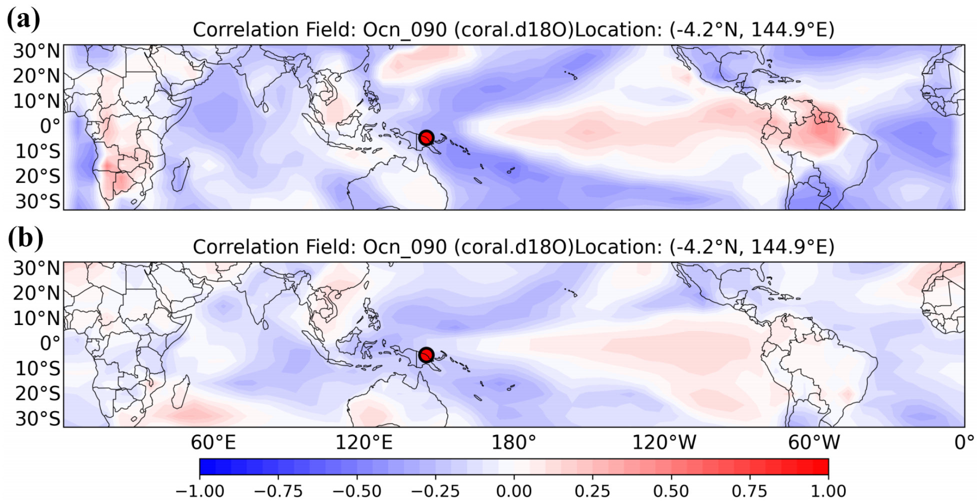

Figure 9. Spatial correlation maps between coral proxy data (ID: Ocn_090) located at 4.2° S, 144° E and **(a)** instrumental observations, **(b)** prior data.

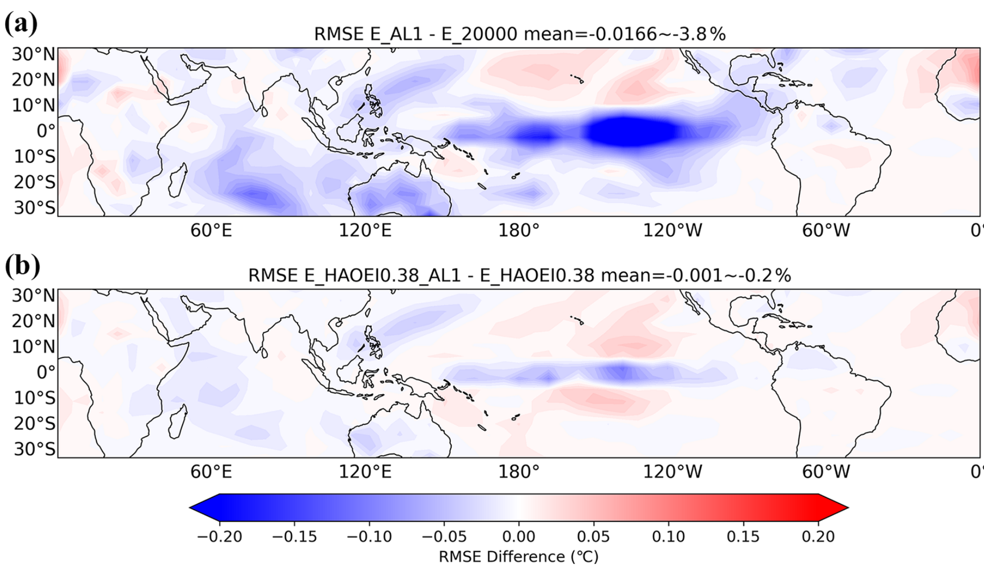

Figure 10. **(a)** The RMSE differences and the percentage RMSE differences between E_AL1 and E_20000 **(b)** between E_HAOEI0.38_AL1 and E_HAOEI0.38.

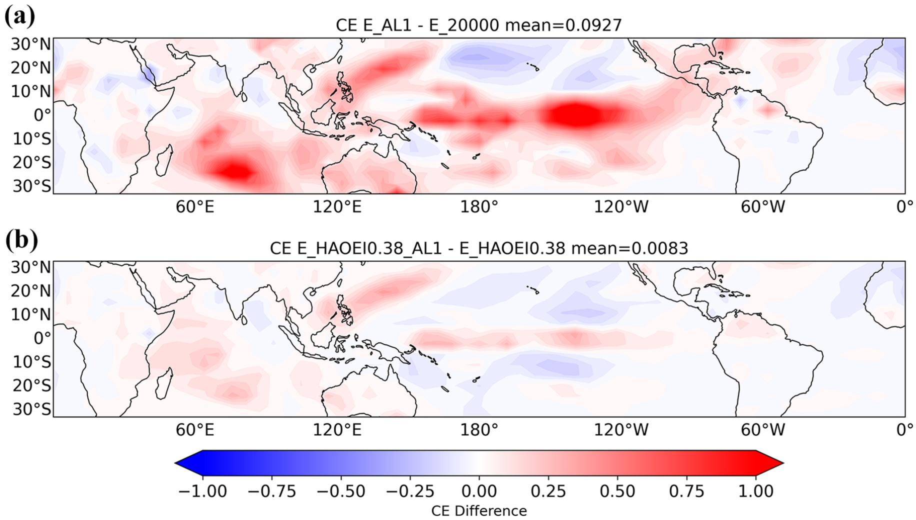

Figure 11. Same as Fig. 10 but for the CE.

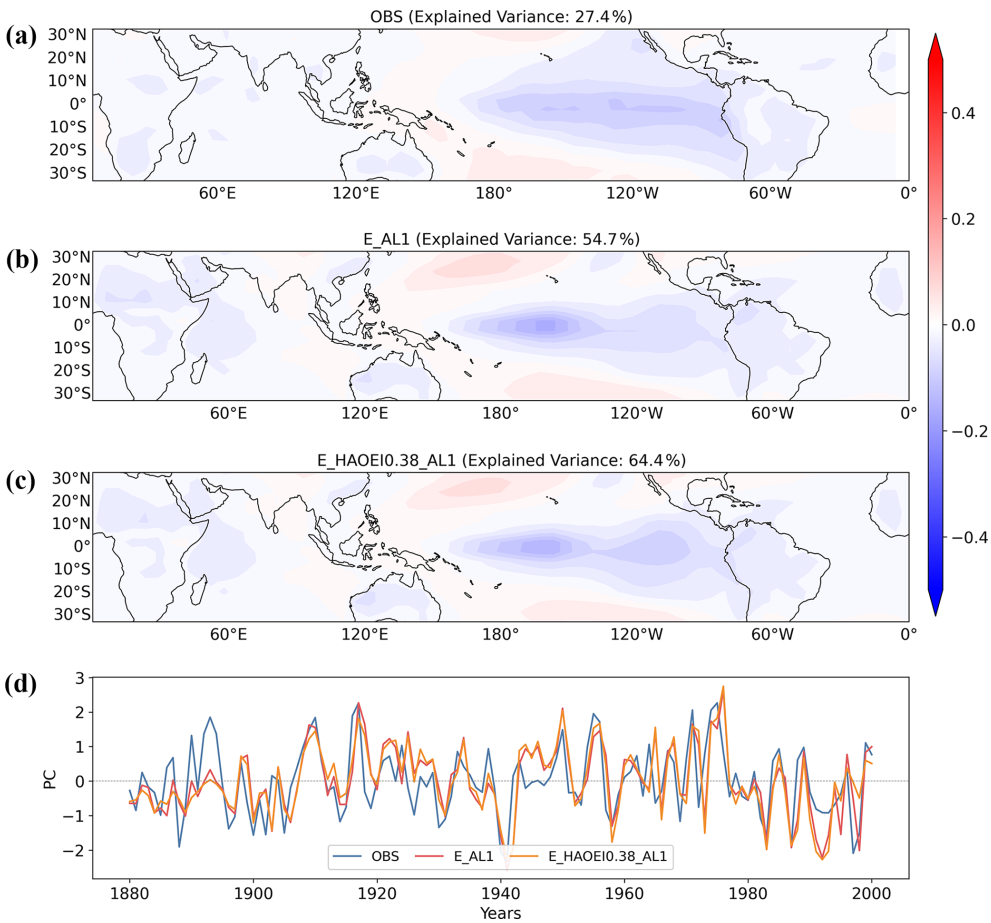

Figure 12. The leading EOF mode of **(a)** observations **(b)** adaptive localization analyses **(c)** adaptive localization and HAOEI ($\delta = 0.38$) analyses **(d)** corresponding PC time series for all experiments.

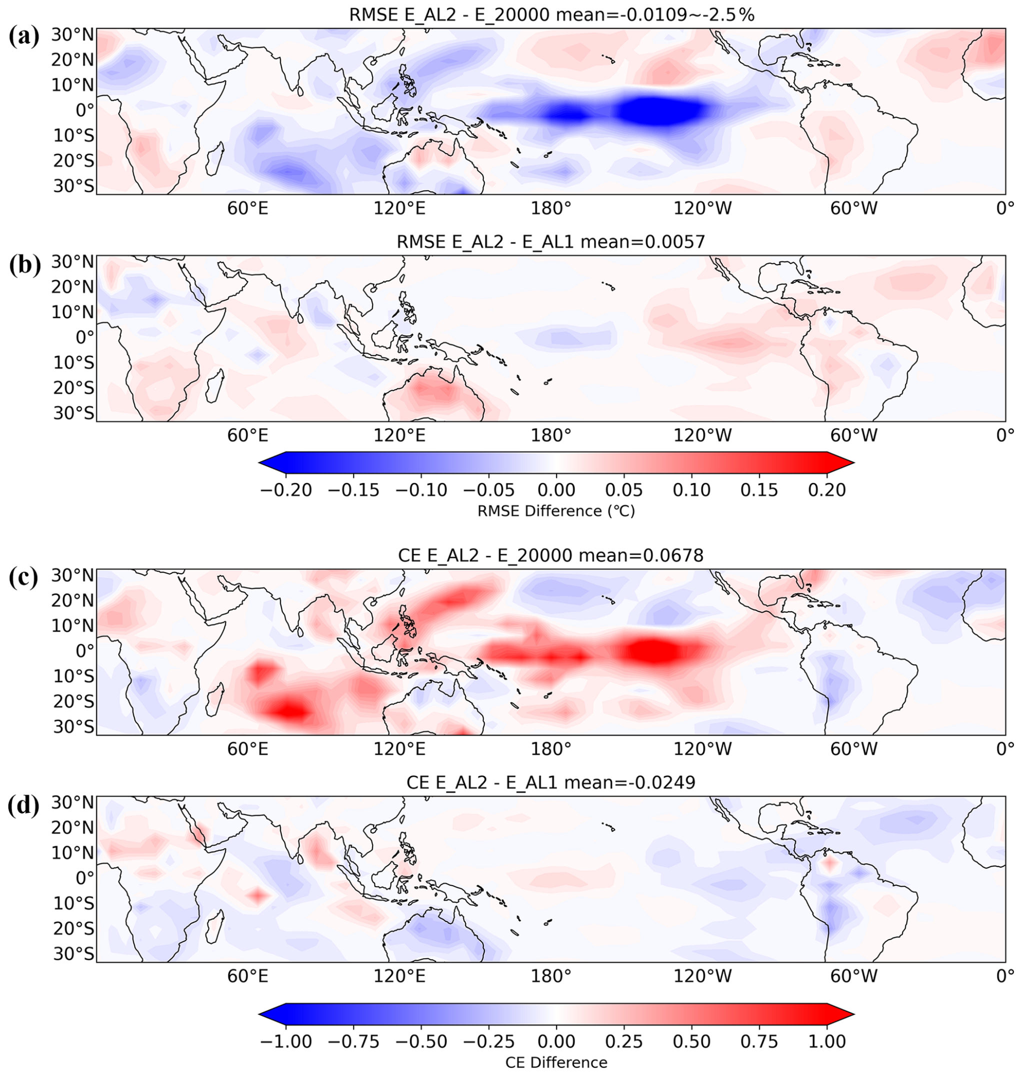

Figure 13. **(a)** The RMSE differences and the percentage RMSE differences between E_AL2 and E_20000 **(b)** between E_AL2 and E_AL1 **(c)** the CE differences between E_AL2 and E_20000 **(d)** between E_AL2 and E_AL1.

4 Conclusion and outlook

This study has developed and systematically evaluated two adaptive strategies within an ensemble-based paleoclimate data assimilation framework: adaptive observation error inflation and adaptive covariance localization. Through sensitivity experiments using coral-based proxy records over the tropical region (1880–2000), we have quantified their impacts on reconstruction quality in terms of pointwise accuracy (RMSE and CE) and fidelity of large-scale climate modes (EOF analysis).

For adaptive observation error inflation, the aggressive AOEI method yields moderate improvements but introduces local degradations where accurate observations are mistakenly down-weighted. The HAOEI method, with its threshold-based linear inflation, provides more balanced and robust performance. Systematic experiments with three δ values (0.67, 0.38, and 0.13) reveal a clear non-monotonic behavior, with $\delta = 0.38$ achieving the largest RMSE reduction (approximately 5.8 %) and the highest CE among all inflation experiments. The existence of an optimal δ indicates a fundamental trade-off: a threshold that is too small over-penalizes moderately inconsistent observations, while a threshold that is too large fails to adequately control genuinely large errors. HAOEI with optimal tuning delivers spatially consistent improvements across most of the tropical domain with substantially fewer local degradations than AOEI.

For covariance localization, the fixed-radius experiments demonstrate that the localization radius critically controls the balance between sampling error and spatial coherence. A large radius preserves large-scale structures but introduces spurious correlations with over-amplified EOF variance, while a small radius distorts the spatial pattern of the leading mode despite yielding a coincidentally close variance percentage. Our proposed adaptive localization method addresses these limitations by dynamically adjusting the radius based on local data density and incorporating correlation-based weights. Compared to the fixed large-radius baseline, adaptive localization reduces RMSE by approximately 3.8 % and substantially improves CE. When combined with HAOEI, the full adaptive system achieves the best overall performance, with the lowest RMSE and highest CE among all experiments. Moreover, the leading EOF modes of the adaptive experiments exhibit spatial structures that closely resemble observations, with explained variances substantially lower than the fixed large-radius experiment and much closer to the observed value. Importantly, the adaptive method remains effective even when correlation information is derived solely from the model prior rather than reanalysis data, confirming its applicability to pre-industrial periods.

In summary, the optimal combination – HAOEI with $\delta = 0.38$ and adaptive localization – yields meaningful improvements in both statistical accuracy and the fidelity of dominant climate modes, reducing RMSE by approximately 5 %–6 % and substantially correcting the over-amplified EOF variance. These improvements, while modest in magnitude, are statistically robust, spatially consistent, and physically meaningful. They demonstrate that the adaptive strategies developed here offer a practical and effective pathway for improving paleoclimate reconstructions, particularly in data-sparse regions and for capturing large-scale climate variability modes such as ENSO.

While this study demonstrates the considerable promise of adaptive strategies in paleoclimate data assimilation, several important avenues merit further investigation to advance the methodology and expand its applications. For the adaptive localization, one could explore time-varying correlation estimates to account for non-stationary teleconnections. Moreover, this study focused on coral-based temperature reconstructions within the tropics; extending the framework to a global scale and to multiple proxy types and climate variables would provide a rigorous test of its generalizability and enable more holistic climate field reconstructions. The methods should be also tested during periods of abrupt climate change (e.g., the Last Glacial Termination, Dansgaard-Oeschger events) or past warm periods (e.g., the Last Interglacial, the Pliocene), where data constraints are particularly challenging but scientific stakes are high.

Code and data availability. The code and data that support the findings of this study are openly available. All input datasets actually used in this study (including the PAGES 2k Phase 2 global multiproxy database, the "iCESM1" last millennium simulation, the NASA Goddard's Global Surface Temperature Analysis, as well as the spatially completed version of the near-surface air temperature and sea-surface temperature analysis product HadCRUT4.6) are hosted within the cfr and can be accessed at https://doi.org/10.5281/zenodo.10575537 (Zhu et al., 2024a). The exact version of the code, the source code of the assimilation system, and the validation results of the reconstructed estimates used in this study are also archived in a trusted permanent repository at Zenodo under the following DOI: https://doi.org/10.5281/zenodo.19015635 (Luo et al., 2026).

Author contributions. G. Luo conducted the experiments and wrote the first draft of the manuscript. Y. Zeng and F. Zhu proposed the ideas of adaptive observation error inflation and adaptive localization, and implemented them. J. Zhao contributed to the analysis of the results. All authors contributed to revising the text and defining the structure of the paper.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Geoscientific Model Development*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

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

Financial support. This research has been supported by the National Key Research and Development Program of China (grant no. 2023YFF0804703).

Review statement. This paper was edited by Benjamin Gaubert and reviewed by three anonymous referees.

References

Anderson, J. L.: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, Physica D, 230, 99–111, 2007.

Anderson, J. L.: Localization and sampling error correction in ensemble Kalman filter data assimilation, Mon. Weather Rev., 140, 2359–2371, 2012.

Bishop, C. and Hodyss, D.: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models, Tellus A, 61, 84–96, 2009a.

Bishop, C. and Hodyss, D.: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere, Tellus A, 61, 97–111, 2009b.

Brady, E., Stevenson, S., Bailey, D., Liu, Z., Noone, D., Nusbaumer, J., Otto-Bliesner, B. L., Tabor, C., Tomas, R., Wong, T., Zhang, J., and Zhu, J.: The connected isotopic water cycle in the Community Earth System Model version 1, J. Adv. Model. Earth Sy., 11, 2547–2566, 2019.

Dee, S., Emile-Geay, J., Evans, M., Allam, A., Steig, E., and Thompson, D.: PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, J. Adv. Model. Earth Sy., 7, 1220–1247, 2015.

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, Quaternary Sci. Rev., 76, 16–28, 2013.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res.-Oceans, 99, 10143–10162, 1994.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Q. J. Roy. Meteor. Soc., 125, 723–757, 1999.

Goosse, H.: Reconstructed and simulated temperature asymmetry between continents in both hemispheres over the last centuries, Clim. Dynam., 48, 1483–1501, 2017.

Guillot, D., Rajaratnam, B., and Emile-Geay, J.: Statistical paleoclimate reconstructions via Markov random fields, Ann. Appl. Stat., 9, 324–352, 2015.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The last millennium climate reanalysis project: Framework and first results, J. Geophys. Res.-Atmos., 121, 6745–6764, 2016.

Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev., 126, 796–811, 1998.

Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 129, 123–137, 2001.

Huber, P. J.: Robust estimation of a location parameter, in: Breakthroughs in statistics: Methodology and distribution, Springer, 492–518, https://doi.org/10.1007/978-1-4612-4380-9_35, 1992.

Kageyama, M., Harrison, S. P., Kapsch, M.-L., Lofverstrom, M., Lora, J. M., Mikolajewicz, U., Sherriff-Tadano, S., Vadsaria, T., Abe-Ouchi, A., Bouttes, N., Chandan, D., Gregoire, L. J., Ivanovic, R. F., Izumi, K., LeGrande, A. N., Lhardy, F., Lohmann, G., Morozova, P. A., Ohgaito, R., Paul, A., Peltier, W. R., Poulsen, C. J., Quiquet, A., Roche, D. M., Shi, X., Tierney, J. E., Valdes, P. J., Volodin, E., and Zhu, J.: The PMIP4 Last Glacial Maximum experiments: preliminary results and comparison with the PMIP3 simulations, Clim. Past, 17, 1065–1089, https://doi.org/10.5194/cp-17-1065-2021, 2021.

Lei, L. and Anderson, J. L.: Comparisons of empirical localization techniques for serial ensemble Kalman filters in a simple atmospheric general circulation model, Mon. Weather Rev., 142, 739–754, 2014a.

Lei, L. and Anderson, J. L.: Empirical localization of observations for serial ensemble Kalman filter data assimilation in an atmospheric general circulation model, Mon. Weather Rev., 142, 1835–1851, 2014b.

Lenssen, N. J., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP uncertainty model, J. Geophys. Res.-Atmos., 124, 6307–6326, 2019.

Liu, Z., Zhu, J., Rosenthal, Y., Zhang, X., Otto-Bliesner, B. L., Timmermann, A., Smith, R. S., Lohmann, G., Zheng, W., and Elison Timm, O.: The Holocene temperature conundrum, P. Natl. Acad. Sci. USA, 111, E3501–E3505, 2014.

Lorenz, E. N.: Empirical orthogonal functions and statistical weather prediction, Vol. 1, Department of Meteorology, Massachusetts Institute of Technology, Cambridge, 1956.

Luo, G., Zeng, Y., Zhu, F., and Zhao, J.: Paleoclimate data assimilation with adaptive observation error inflation and adaptive localization, Version v2, Zenodo [data set/code], https://doi.org/10.5281/zenodo.19015635, 2026.

Mann, M. E., Bradley, R. S., and Hughes, M. K.: Global-scale temperature patterns and climate forcing over the past six centuries, Nature, 392, 779–787, 1998.

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, P. Natl. Acad. Sci. USA, 105, 13252–13257, 2008.

Ménétrier, B., Montmerle, T., Michel, Y., and Berre, L.: Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization, Mon. Weather Rev., 143, 1622–1643, 2015a.

Ménétrier, B., Montmerle, T., Michel, Y., and Berre, L.: Linear filtering of sample covariances for ensemble-based data assimilation. Part II: Application to a convective-scale NWP model, Mon. Weather Rev., 143, 1644–1664, 2015b.

Minamide, M. and Zhang, F.: Adaptive observation error inflation for assimilating all-sky satellite radiance, Mon. Weather Rev., 145, 1063–1081, 2017.

Osman, M. B., Tierney, J. E., Zhu, J., Tardif, R., Hakim, G. J., King, J., and Poulsen, C. J.: Globally resolved surface temperatures since the Last Glacial Maximum, Nature, 599, 239–244, 2021.

PAGES 2k Consortium: A global multiproxy database for temperature reconstructions of the Common Era, Scientific Data, 4, 170088, https://doi.org/10.1038/sdata.2017.88, 2017.

Parzen, E.: On estimation of a probability density function and mode, Ann. Math. Stat., 33, 1065–1076, 1962.

Phipps, S. J., McGregor, H. V., Gergis, J., Gallant, A. J., Neukom, R., Stevenson, S., Ackerley, D., Brown, J. R., Fischer, M. J., and Van Ommen, T. D.: Paleoclimate data–model comparison and the role of climate forcings over the past 1500 years, J. Climate, 26, 6915–6936, 2013.

Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H.: Assimilation of time-averaged pseudoproxies for climate reconstruction, J. Climate, 27, 426–441, 2014.

Sun, H., Lei, L., Liu, Z., Ning, L., and Tan, Z.-M.: An online paleoclimate data assimilation with a deep learning-based network, J. Adv. Model. Earth Sy., 17, e2024MS004675, https://doi.org/10.1029/2024MS004675, 2025.

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling, Clim. Past, 15, 1251–1273, https://doi.org/10.5194/cp-15-1251-2019, 2019.

Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B.: Piecing together the past: statistical insights into paleoclimatic reconstructions, Quaternary Sci. Rev., 35, 1–22, 2012.

Vaccaro, A., Emile-Geay, J., Guillot, D., Verna, R., Morice, C., Kennedy, J., and Rajaratnam, B.: Climate field completion via Markov random fields: Application to the HadCRUT4. 6 temperature dataset, J. Climate, 34, 4169–4188, 2021.

von Storch, H., Cubasch, U., Gonzalez-Rouco, J. F., Jones, J. M., Voss, R., Widmann, M., and Zorita, E.: Combining paleoclimatic evidence and GCMs by means of data assimilation through upscaling and nudging (DATUN), in: Proceedings of the 11th symposium on global change studies, 28–31, 2000.

Wang, J., Emile-Geay, J., Guillot, D., McKay, N. P., and Rajaratnam, B.: Fragility of reconstructed temperature patterns over the Common Era: Implications for model evaluation, Geophys. Res. Lett., 42, 7162–7170, 2015.

Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, 2002.

Widmann, M., Goosse, H., van der Schrier, G., Schnur, R., and Barkmeijer, J.: Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium, Clim. Past, 6, 627–644, https://doi.org/10.5194/cp-6-627-2010, 2010.

Zeng, Y. and Janjić, T.: Study of conservation laws with the local ensemble transform kalman filter, Q. J. Roy. Meteor. Soc., 699, 2359–2372, 2016.

Zhu, F., Emile-Geay, J., Anchukaitis, K. J., McKay, N. P., Stevenson, S., and Meng, Z.: A pseudoproxy emulation of the PAGES 2k database using a hierarchy of proxy system models, Scientific Data, 10, 624, https://doi.org/10.1038/s41597-023-02489-1, 2023.

Zhu, F., Emile-Geay, J., Hakim, G. J., Guillot, D., Khider, D., Tardif, R., and Perkins, W. A.: cfr: a Python package for Climate Field Reconstruction, Version v2024.1.26, Zenodo [code], https://doi.org/10.5281/zenodo.10575537, 2024a.

Zhu, F., Emile-Geay, J., Hakim, G. J., Guillot, D., Khider, D., Tardif, R., and Perkins, W. A.: cfr (v2024.1.26): a Python package for climate field reconstruction, Geosci. Model Dev., 17, 3409–3431, https://doi.org/10.5194/gmd-17-3409-2024, 2024b.