Improving ocean reanalysis with the offline ensemble Kalman smoother

Yiguo Wang[1], François Counillon[1,2], Yue Ying[1], Sébastien Barthélémy[2], and Geir Evensen[1,3]

[1]Nansen Environmental and Remote Sensing Centre and Bjerknes Centre for Climate Research, Bergen, Norway
[2]Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway
[3]Norwegian Research Center, Bergen, Norway

Abstract. The Ensemble Kalman Smoother (EnKS), an extension of the Ensemble Kalman Filter, can improve the accuracy of the state estimate by assimilating 'future' observations. We propose to use the EnKS algorithm to enhance the accuracy and reliability of pre-existing reanalyses produced with a fully coupled Earth system model. The offline EnKS is applied to two reanalyses of the Norwegian Climate Prediction Model (NorCPM) to update sea surface height, mixed layer depth, and temperature and salinity for all depth levels of the reanalyses. In an idealized framework, we tune temporal localization parameters and reveal that the optimal temporal localization parameter is 0.1, corresponding to a time delay of about 13 days. In a real framework, we find that observation error variance has to be inflated by a factor of four to account for the autocorrelation of the gridded observational product and avoid overfitting. In both frameworks, the offline EnKS improves the accuracy for the top 300 m temperature, sea surface height, and mixed layer depth, but yields limited improvements in the top 300 m salinity and the water properties below 300 m. Also, it enhances the reliability of the reanalysis. Mostly due to the lack of high-quality and independent datasets for proper validation, the improvement is notably lower in a real framework than in an idealized framework. Overall, this study demonstrates that the offline EnKS has the potential for efficiently improving pre-existing reanalyses.

1 Introduction

Much of our understanding of historical climate comes from reanalyses, which provide a complete, continuous, and dynamically consistent reconstruction of the past state of the climate system (Saha et al., 2006; Dee et al., 2011; Zuo et al., 2019; Laloyaux et al., 2018; Hersbach et al., 2020; O'Kane et al., 2021; Wang et al., 2025). Reanalyses are constructed by assimilating data into a numerical model (Kalnay et al., 1996; Carrassi et al., 2018) and can broadly be categorized into uncoupled and coupled reanalyses, depending on the numerical model used in their production. Uncoupled reanalyses are generated using numerical models that simulate only certain components of the Earth system, e.g, atmosphere reanalyses (Dee et al., 2011) and ocean reanalyses (Zuo et al., 2019). Coupled reanalyses produced with coupled systems better account for coupled processes (Counillon et al., 2016a; Penny et al., 2017; Laloyaux et al., 2018; Brune et al., 2015; Zhang et al., 2007; O'Kane et al., 2021).

Filtering techniques involve recursive analysis steps and model integration, where the model state is updated based on previous observations. The ensemble Kalman filter (EnKF, Evensen, 2003) is widely used in Earth system models for producing

reanalysis (Zhang et al., 2007; Karspeck et al., 2013; Brune et al., 2015; Counillon et al., 2016a; O'Kane et al., 2021; Wang et al., 2025) because the ensemble can be used to quantify the uncertainty of the system and because the forecast error covariances are flow-dependent, i.e., they evolve in time with the dynamic of the system (e.g., Counillon et al., 2016a). While filtering techniques are well suited for forecasting, smoother techniques can use 'future' observations to reduce uncertainty and enhance the reliability of reanalysis.

The ensemble Kalman smoother (EnKS, Evensen and van Leeuwen, 2000; Ravela and McLaughlin, 2007; Khare et al., 2008; Grudzien and Bocquet, 2022) is derived from the EnKF, where the analysis makes also use of 'future' observations within a fixed time window (i.e., time lag) via the cross-time error covariances constructed from the Monte-Carlo ensembles. In such a way, the EnKS would provide a more accurate state than the EnKF (e.g., Ngodock et al., 2006; Bocquet and Sakov, 2014). However, the EnKS has been mostly demonstrated in toy models (Evensen and van Leeuwen, 2000; Ngodock et al., 2006; Ravela and McLaughlin, 2007; Khare et al., 2008; Grudzien and Bocquet, 2022; Dong et al., 2023). For instance, Dong et al. (2023) proposed in Lorenz 1963 to use the EnKS as a post-processing of reanalysis. There have been very few applications of the EnKS in high-dimensional systems, e.g., a simplified Atmospheric General Circulation model (Khare et al., 2008) and an ocean circulation model (Cosme et al., 2010).

The primary objective of this study is to investigate whether using the EnKS as a post-processing approach (or the offline EnKS) can improve pre-existing coupled reanalyses while preserving its reliability. We utilize the Norwegian Climate Prediction Model (NorCPM, Counillon et al., 2014), an advanced fully coupled Earth system model known for providing comprehensive long-term reanalyses (Counillon et al., 2016a; Bethke et al., 2021). We first work in identical twin experiments, where the truth is known, and observations are constructed with the same model to fulfill hypotheses about uncorrelated observation error. The key advantage of the twin experiment (Halem and Dlouhy, 1984) is that it allows controlled testing and evaluation of the offline EnKS. Since the true state is known, the method's performance can be directly assessed by comparing the estimated state with the known truth. We assess the offline EnKS's sensitivity to localization in time (i.e., temporal localization). Second, we test the offline smoother method in a real-world framework on a pre-existing 60-year reanalysis of NorCPM.

2 Methodology

2.1 Ensemble Kalman smoother

The EnKF (Evensen, 2003) is a recursive ensemble-based data assimilation method. The ensemble of realizations of the model allows for a flow-dependent forecast error covariance estimate. A deterministic EnKF (DEnKF) variant has been developed by Sakov and Oke (2008). It uses a linear approximation of the theoretical error covariances to update the ensemble perturbations. This approximation yields a deterministic form of the conventional EnKF. In addition, compared to the traditional EnKF, the DEnKF inherently inflates the analysis error covariances by construction (Sakov and Oke, 2008).

Let the ensemble of model states at analysis time step $\mathbf{X}_\mathrm{f} = [\mathbf{x}_\mathrm{f}^1, \mathbf{x}_\mathrm{f}^2, ..., \mathbf{x}_\mathrm{f}^m] \in \mathbb{R}^{n \times m}$, the ensemble mean be $\overline{\mathbf{x}}_\mathrm{f} \in \mathbb{R}^n$ and the ensemble anomalies or perturbations $\mathbf{A}_\mathrm{f} = \mathbf{X}_\mathrm{f} - \overline{\mathbf{x}}_\mathrm{f}\mathbf{1}_m^\mathrm{T} \in \mathbb{R}^{n \times m}$, where the subscript 'f' denotes forecast, n is the size of

model states, m is the ensemble size, and $\mathbf{1}_m = [1, 1, ..., 1]^{\mathrm{T}} \in \mathbb{R}^m$. The DEnKF update can be written as follows:

$$\overline{\mathbf{x}}_{\mathrm{a}} = \overline{\mathbf{x}}_{\mathrm{f}} + \mathbf{K}\left(\mathbf{y} - H(\overline{\mathbf{x}}_{\mathrm{f}})\right), \tag{1}$$

$$\mathbf{A}_{\mathrm{a}} = \mathbf{A}_{\mathrm{f}} - \frac{1}{2}\mathbf{K}\mathbf{H}\mathbf{A}_{\mathrm{f}}, \tag{2}$$

$$\mathbf{X}_{\mathrm{a}} = \overline{\mathbf{x}}_{\mathrm{a}}\mathbf{1}_m^{\mathrm{T}} + \mathbf{A}_{\mathrm{a}}, \tag{3}$$

where the subscript 'a' denotes analysis, $\mathbf{y}$ is the observation vector, H is the observation operator that maps model states to the observation space, and $\mathbf{H}$ is the tangent linear operator of H. The Kalman gain $\mathbf{K}$ is defined by

$$\mathbf{K} = \mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}, \tag{4}$$

where $\mathbf{P}_{\mathrm{f}}$ is the forecast error covariance matrix estimated by the ensemble perturbations:

$$\mathbf{P}_{\mathrm{f}} = \frac{1}{m-1}\mathbf{A}_{\mathrm{f}}\mathbf{A}_{\mathrm{f}}^{\mathrm{T}} \tag{5}$$

and $\mathbf{R}$ is the observation error covariance matrix.

We can rewrite Eqs. 1-3 into a form of the Ensemble Transform Kalman Filter as follows:

$$\mathbf{X}_{\mathrm{a}} = \mathbf{X}_{\mathrm{f}}\mathbf{T}, \tag{6}$$

$$\mathbf{T} = \mathbf{I} + \mathbf{A}_{\mathrm{f}}^{\mathrm{T}}\mathbf{H}\left(\mathbf{H}\mathbf{A}_{\mathrm{f}}\mathbf{A}_{\mathrm{f}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} + (m-1)\mathbf{R}\right)^{-1}\left(\left(\mathbf{y} - H\left(\overline{\mathbf{x}}_{\mathrm{f}}\right)\right)\mathbf{1}_m^{\mathrm{T}} - \frac{1}{2}\mathbf{K}\mathbf{H}\mathbf{A}_{\mathrm{f}}\right), \tag{7}$$

where $\mathbf{T}$ is the transform matrix (also known as $\mathbf{X}_5$ in Evensen (2003)).

The (online/standard) EnKS is an EnKF extension (Evensen and van Leeuwen, 2000). It extends the analysis by adding the model state vector at the previous steps into $\mathbf{X}_{\mathrm{f}}$. In such a way, it updates the instantaneous model state that served as the initial conditions for the subsequent model integration and the model outputs in previous time steps.

The EnKS developed in this study is similar to the online/standard EnKS (Evensen and van Leeuwen, 2000) to incorporate future observations not used in the analysis, but does not require a model integration (black arrows in Figure 1). Thus, it is numerically cheap. In this way, we can consider the EnKS method developed in this study as an offline approach (Raanes, 2016) or a post-processing approach to the pre-existing reanalysis (green dots in Figure 1).

We propose to carry out the EnKS forward in time (i.e., sequentially from the first analysis to the last analysis) and consider a fixed lag of L time steps, meaning that the offline EnKS assimilates 'future' observations at L time steps ahead of time t. Please refer to the dotted shape in Figure 1.

We define the concatenated observations as

$$\mathbf{y}_{\mathrm{s}} = \left[\mathbf{y}_1^{\mathrm{T}}, \mathbf{y}_2^{\mathrm{T}}, ..., \mathbf{y}_L^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{8}$$

where $\mathbf{y}_k$ are observations at k time steps ahead of the existing analysis $\mathbf{X}_a$ at time t (Eq. 3) and $1 \leq k \leq L$. Note that $\mathbf{y}_0$ (i.e., $\mathbf{y}$ in Eq. 1) has been assimilated during the production of the filter analysis $\mathbf{X}_a$ and is not included in $\mathbf{y}_s$ for the smoother

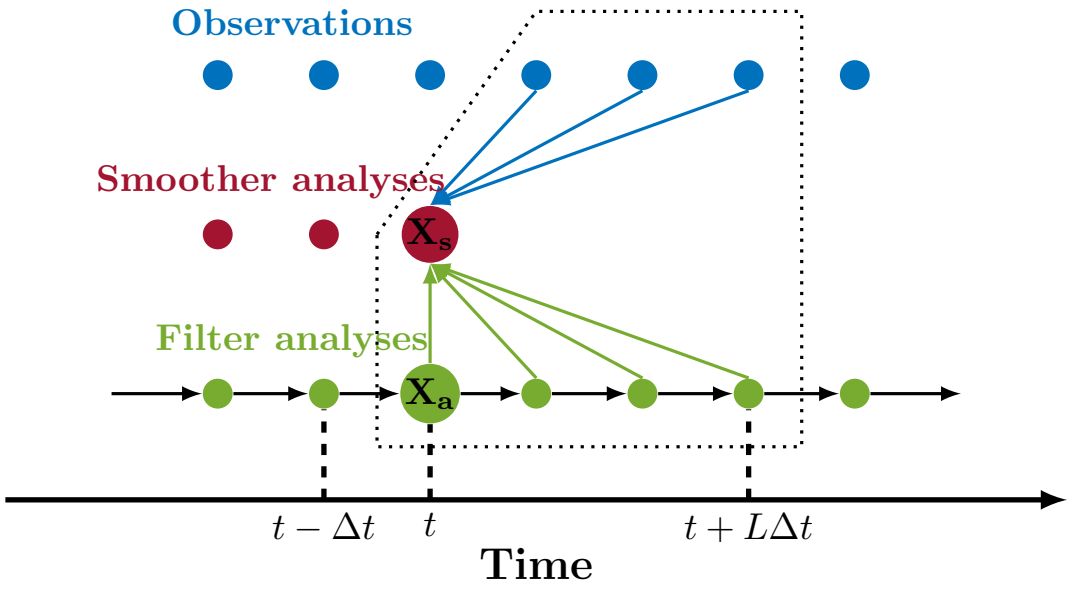

Figure 1. Schematic of the offline EnKS. The green dots represent the filter analysis after a model integration of Δt schematized by the black arrows. The blue dots depict the observations, and the red dots represent the smoother analyses of the offline EnKS. The dotted shape contains 'future' observations and filter analyses from t to $t + L\Delta t$ to generate the smoother analysis $\mathbf{X_s}$ at t, in which the blue and green arrows are information flows

analysis. The concentrated observation error covariance matrix becomes a block diagonal matrix as follows:

$$\mathbf{R}_\mathrm{s} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_L \end{bmatrix}, \tag{9}$$

where observations from different times are assumed to be independent. The ensemble mean of the filter analyses within L time steps ahead of the model state $\mathbf{X}_\mathrm{a}$ mapped to the observation space is written as

$$H_\mathrm{s}(\overline{\mathbf{x}}_{\mathrm{a},1:L}) = \left[H_1(\overline{\mathbf{x}}_{\mathrm{a},1})^\mathrm{T}, H_2(\overline{\mathbf{x}}_{\mathrm{a},2})^\mathrm{T}, ..., H_L(\overline{\mathbf{x}}_{\mathrm{a},L})^\mathrm{T}\right]^\mathrm{T}. \tag{10}$$

The tangent linear operator of the operator H_s is

$$\mathbf{H}_\mathrm{s} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_L \end{bmatrix}. \tag{11}$$

The concatenated ensemble anomaly is defined as

$$\mathbf{A}_{1:L} = \left[\mathbf{A}_1^{\mathrm{T}}, \mathbf{A}_2^{\mathrm{T}}, ..., \mathbf{A}_L^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{12}$$

The Kalman gain of the offline EnKS is written as

$$\mathbf{K}_{\mathrm{s}} = \mathbf{A}_{\mathrm{a}}\mathbf{A}_{1:L}^{\mathrm{T}}\mathbf{H}_{\mathrm{s}}^{\mathrm{T}}\left(\mathbf{H}_{\mathrm{s}}\mathbf{A}_{1:L}\mathbf{A}_{1:L}^{\mathrm{T}}\mathbf{H}_{\mathrm{s}}^{\mathrm{T}} + (m-1)\mathbf{R}_{\mathrm{s}}\right)^{-1}. \tag{13}$$

The offline EnKS can be written as follows:

$$\overline{\mathbf{x}}_{\mathrm{s}} = \overline{\mathbf{x}}_{\mathrm{a}} + \mathbf{K}_{\mathrm{s}}\left(\mathbf{y}_{\mathrm{s}} - H_{\mathrm{s}}\left(\overline{\mathbf{x}}_{\mathrm{a},1:L}\right)\right), \tag{14}$$

$$\mathbf{A}_{\mathrm{s}} = \mathbf{A}_{\mathrm{a}} - \frac{1}{2}\mathbf{K}_{\mathrm{s}}\mathbf{H}_{\mathrm{s}}\mathbf{A}_{1:L}, \tag{15}$$

$$\mathbf{X}_{\mathrm{s}} = \overline{\mathbf{x}}_{\mathrm{s}}\mathbf{1}_m^{\mathrm{T}} + \mathbf{A}_{\mathrm{s}}. \tag{16}$$

For simplification, we can rewrite Eq. 16 as follows:

$$\mathbf{X}_{\mathrm{s}} = \mathbf{X}_{\mathrm{a}}\mathbf{T}_{\mathrm{s}}, \tag{17}$$

$$\mathbf{T}_{\mathrm{s}} = \mathbf{I} + \mathbf{A}_{1:L}^{\mathrm{T}}\mathbf{H}_{\mathrm{s}}\left(\mathbf{H}_{\mathrm{s}}\mathbf{A}_{1:L}\mathbf{A}_{1:L}^{\mathrm{T}}\mathbf{H}_{\mathrm{s}}^{\mathrm{T}} + (m-1)\mathbf{R}_{\mathrm{s}}\right)^{-1}\left(\left(\mathbf{y}_{\mathrm{s}} - H_{\mathrm{s}}\left(\overline{\mathbf{x}}_{\mathrm{a},1:L}\right)\right)\mathbf{1}_m^{\mathrm{T}} - \frac{1}{2}\mathbf{K}_{\mathrm{s}}\mathbf{H}_{\mathrm{s}}\mathbf{A}_{1:L}\right), \tag{18}$$

where $\mathbf{T}_{\mathrm{s}}$ is the transform matrix (Evensen, 2003).

Note that the online/standard EnKS (Evensen and van Leeuwen, 2000) computes the transform matrix $\mathbf{T}$ (Eq. 7) based on the instantaneous model state $\mathbf{X}_{\mathrm{f}}$. We here compute the transform matrix $\mathbf{T}_{\mathrm{s}}$ based on time-averaged model outputs of the reanalysis since it is costly to store for all instantaneous model states. In such a way, using the online/standard EnKS and our offline EnKS would lead to different solutions.

2.2 Test datasets

We apply the offline EnKS in two pre-existing NorCPM reanalysis datasets in an identical twin experiment or the real framework. NorCPM combines the Norwegian Earth System Model (NorESM, Bentsen et al., 2013) with the EnKF (Evensen, 2003) and has been developed to provide long-term coupled reanalysis (Counillon et al., 2016a; Bethke et al., 2021; Wang et al., 2022) and seasonal-to-decadal climate predictions (Counillon et al., 2014; Wang et al., 2019; Bethke et al., 2021). NorESM is a state-of-the-art global fully–coupled Earth system model for climate simulations and participates in the Coupled Model Intercomparison Project (Taylor et al., 2012; Eyring et al., 2015). It is based on the Community Earth System Model version 1.0.3 (Vertenstein et al., 2012), a successor to the Community Climate System Model version 4 (Gent et al., 2011). Its ocean component is the Bergen Layered Ocean Model (BLOM, Bentsen et al., 2013). Its sea ice component is the Los Alamos sea ice model (Gent et al., 2011; Holland et al., 2012). Its atmosphere component is a version of the Community Atmosphere Model (Kirkevåg et al., 2013). Its land component is the Community Land Model (Oleson et al., 2010; Lawrence et al., 2011). The version 7 coupler (Craig et al., 2012) is used. CAM4 and CLM4 have a horizontal resolution of 1.9° at latitude and 2.5° at longitude. BLOM and CICE4 have a horizontal resolution of approximately 1°. NorCPM comprises 30 ensemble members ($m = 30$ in section 2.1).

In the identical twin experiment (Halem and Dlouhy, 1984), the 'truth' is known and generated from the same model. We take advantage of the existence of simulations of twin experiments in Wang et al. (2022):

- The 'truth' has been spawned from the NorESM historical simulation by perturbing the sea surface temperature (SST) of the initial conditions in January 1960 with spatially uncorrelated white noise with a standard deviation of 10^{-6} °C and then integrating it up to 2010.
- Synthetic SST observations have been generated from monthly outputs of the truth with white noise. The noise amplitude has been set equal to the time-varying standard deviation of HadISST2 (Rayner et al., 2003) computed from its ten realizations.
- The synthetic reanalysis of the NorCPM over the period 1980–2010 (Wang et al., 2022, its STEPmax simulation of twin experiments) has been produced by monthly assimilation of synthetic SST observations. This reanalysis has been produced with a vertical localization function updating vertical layers down to the last layer for which a significant correlation with SST is found.

Please refer to Wang et al. (2022) for more details. These data (including the synthetic reanalysis, the truth, and synthetic SST observations) can be downloaded at Wang (2023).

In the real-world framework, we employ the offline EnKS technique on a 60-year coupled reanalysis dataset generated by the NorCPM model (Counillon et al., 2016b) and documented in Counillon et al. (2016a). This dataset spans from 1950 to 2010 and assimilates monthly SST observations from the HadISST2 dataset (Rayner et al., 2003) through the EnKF (Counillon et al., 2016a). It uses an anomaly assimilation framework, which assimilates observational anomalies for the climatology period from 1950 to 2009. The observation uncertainties are defined as the standard deviation of ten realizations of HadISST2. It is essential to highlight that the reanalysis employed here has been rigorously validated. As detailed in Counillon et al. (2016a), this reanalysis has demonstrated considerable skill compared to independent measurements, including assessments of sea surface height (SSH), ocean heat content, and ocean salt content. Furthermore, this reanalysis is reliable in providing accurate estimations of critical oceanic indicators, such as the North Atlantic Subtropical Gyre index and vertical temperature variability. Additionally, it has shown promising results in simulating the Atlantic Meridional Overturning Circulation. Please refer to Counillon et al. (2016a) for more details.

Note that the data assimilated in the offline EnKS (section 2.1) are the same monthly SST data used in the pre-existing reanalyses in Wang et al. (2022) and Counillon et al. (2016a).

2.3 Practical implementation and experiment setups

The EnKS, like many other ensemble-based data assimilation methods, relies on background error covariances to incorporate observed information into the model state (section 2.1). To mitigate the impact of spurious covariances due to sampling errors and small ensemble size, one commonly uses localization (Hamill et al., 2001; Houtekamer and Mitchell, 2001).

The localization is crucial for achieving optimal results of the EnKS (e.g., Khare et al., 2008). The appropriate choice of localization parameters, such as the spatial extent of the localization function or 'localization radius', depends on the specific

characteristics of the model and the observations. Overly aggressive localization, such as setting a very short localization radius, can lead to underutilizing available observations. In this case, the assimilation process becomes excessively conservative and fails to effectively integrate observed data into the model, resulting in less accurate state estimates. Conversely, an overly large localization radius allows spurious variability to introduce artificial signals into the assimilation process. This can lead to the incorporation of noise or irrelevant information, potentially deteriorating the quality of the state estimates.

We use a local analysis framework (Evensen, 2003; Hunt et al., 2007). The smoother analysis uses the neighboring local observations (in space and time). To effectively account for the influence of distant observations in space, we adjust the observation error variance by applying the reciprocal of the Gaspari and Cohn localization function (Gaspari and Cohn, 1999) as in NorCPM. As demonstrated by Sakov and Bertino (2010), this approach is equivalent to utilizing the localization function to filter the background error covariances. The spatial localization radius is a bimodal Gaussian distribution that varies with latitude (Wang et al., 2017). It is smaller in high latitudes compared to mid-latitudes to account for the Coriolis force's effects. The localization radius reaches a local maximum of approximately 2300 km at mid-latitudes and decreases to around 1500 km near the Equator to account for the anisotropy of the covariance there.

For the temporal localization, we employ the reciprocal of an exponential function γ^l for tapering observation error variances, such that γ accounts for the decay rate per month (representing the assimilation cycle), and l denotes the time lag in number of months within three months (i.e., $L = 3$ in section 2.1 and Figure 1). This temporal localization function has demonstrated its utility in prior studies, such as Dong et al. (2021), where it contributed to the enhancement of daily high-resolution ocean reanalysis, as well as in the Lorenz 1963 idealized system examined by Dong et al. (2023). In our twin experiments, we perform 6 experiments, with temporal parameters $\gamma = 0.01, 0.05, 0.1, 0.2, 0.3,$ or 0.5. To provide some insight on the delay timescale, we make use of the formulation $-\frac{1}{\ln(\gamma)}$ from Dong et al. (2023). The chosen temporal parameters correspond to a delay timescale of 6.5, 10.0, 13.0, 18.6, 25.0, and 43.0 days.

Since the process of producing a gridded product from sparse observations introduces correlation in the observation error, in the real framework, we multiply observation error variance by a factor of 1, 2, 4, 6, 8, 10, 16, or 25 to take into account observation correlation as suggested in Stewart et al. (2008). We conduct $8 * 6$ (i.e., 48) experiments in the real framework, considering both correlated observations and temporal parameters.

In all experiments, the offline EnKS updates monthly data for SSH, mixed layer depth (MLD), and temperature and salinity for all depth levels of the reanalysis.

2.4 Validation datasets

We use the truth to validate the smoother analyses in the twin experiment. In the real framework, we use the EN4.2.1 objective analysis dataset (Good et al., 2013, https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html) to validate anomalies of top 300 m temperature (T300) over the period from 1950 to 2010. For SSH, we estimate anomalies from 1993 to 2010 using the gridded dataset ARMOR-3D L4, which can be accessed at https://doi.org/10.48670/moi-00148. The MLD dataset is available from 1958 to 2010 and taken from the ORAS5 global ocean reanalysis at https://doi.org/10.24381/cds.67e8eeb7. All these datasets are independent of the SST dataset assimilated into the system (i.e., HadISST2).

2.5 Evaluation metric

We use the mean squared skill score (MSSS, Murphy, 1988) to evaluate the smoother analysis. We take the original (or filter) analysis as the reference to define the MSSS as follows:

$$\textbf{MSSS} = 1 - \frac{\mathrm{MSE_s}}{\mathrm{MSE_f}} = \frac{\mathrm{MSE_f} - \mathrm{MSE_s}}{\mathrm{MSE_f}}, \tag{19}$$

where $\mathrm{MSE_f}$ ($\mathrm{MSE_s}$) is the mean squared error of the filter (smoother) analysis against the truth or independent observations. A positive (negative) MSSS value indicates that the smoother analysis is more (less) accurate than the filter analysis. If $\mathrm{MSE_s}$ is equal to zero (i.e., the smoother analysis is perfect), the MSSS is equal to one. It is worth noting that the MSSS is equivalent to the squared error reduction ratio.

Before the MSSS computation, the reanalysis and the (independent) validation data are interpolated to a common regular $5^\circ \times 5^\circ$ grid. We evaluate the top 300 m temperature (T300), SSH, and MLD, which are highly relevant to climate studies. We also verified other quantities, such as the top 300 m salinity (S300) and temperature below 300 m, but improvements are either weaker or none (please refer to Figure 1 in Supplementary Information).

To check the reliability of the reanalysis, we make use of the assimilation diagnostics proposed by Desroziers et al. (2005) and used in many applications (e.g., Bethke et al., 2018; Counillon et al., 2016a; Sakov et al., 2012; Slivinski et al., 2021). We define global statistics as follows:

$$d = \sqrt{\sum_i w_i d_i^2}, \tag{20}$$

$$\overline{\sigma^r} = \sqrt{\sum_i w_i \left(\sigma_i^r\right)^2}, \tag{21}$$

$$\overline{\sigma^o} = \sqrt{\sum_i w_i \left(\sigma_i^o\right)^2}, \tag{22}$$

$$\sigma = \sqrt{\left(\overline{\sigma^r}\right)^2 + \left(\overline{\sigma^o}\right)^2}, \tag{23}$$

where w_i is the area of grid cell i relative to the global ocean area, d_i is the difference between observation and reanalysis, σ_i^r is the standard deviation of the ensemble of reanalysis representing reanalysis error, and σ_i^o is observation error. d represents the RMSE computed over the difference between the observations and the reanalysis. $\overline{\sigma^r}$ and $\overline{\sigma^o}$ represent the globally-averaged reanalysis and observation errors, respectively. According to Desroziers et al. (2005), in the case where the observation and reanalysis errors are uncorrelated and unbiased, the RMSE d is expected to be equal to the combined error σ, meaning the reanalysis is reliable. Since the reliability budget relies on the observation errors or uncertainties often not provided in the datasets, we only check the reliability of the reanalysis for SST (i.e., the assimilated quantity).

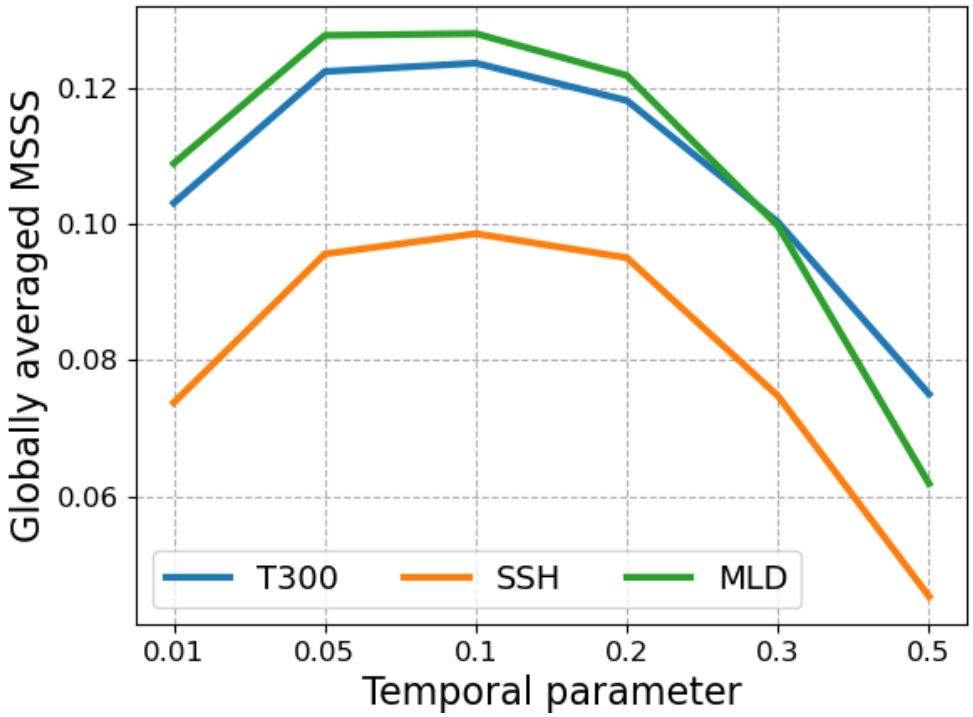

Figure 2. Globally averaged MSSS scores for different experiments for T300 (blue line), SSH (orange line), and MLD (green line).

3 Results

3.1 Twin experiments

All temporal parameters yield positive MSSS values for monthly T300, SSH, and MLD (Figure 2), underscoring the promising potential of the offline EnKS method. However, the improvements vary with the temporal localization parameter γ. The optimal results are obtained with $\gamma = 0.1$ for all three quantities, corresponding to a delay time of 13 days. Higher or lower temporal localization degrades performance, revealing that smoother timescale at the ocean surface is less than one month – in close agreement with the findings in Dong et al. (2021). Assimilating future SST data is slightly more efficient for T300 (13%) and MLD (12%) than SSH (10%, Figure 2). We show only the experiment results with $\gamma = 0.1$ in the following of this section.

For monthly T300, we find larger errors near the Equator and western boundary currents (Figures 3a), where the improvements (or error reduction) from the offline EnKS with the optimal parameter are evident. We also find prominent improvements in the Southern Ocean along the Antarctic Circumpolar Current, the largest wind-driven current on Earth. In these regions, dynamics are chaotic, and errors evolve quickly. Future observations can help constrain state errors effectively.

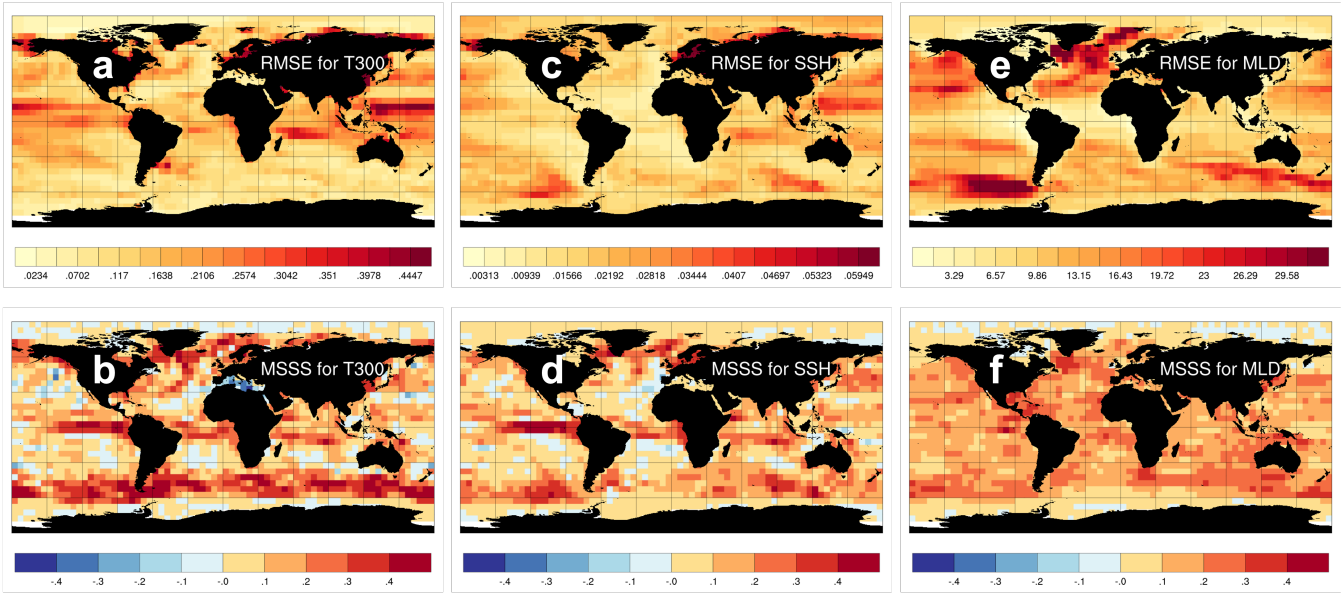

Figure 3. The upper row: the global maps of RMSE for monthly T300 (unit: °C), SSH (unit: meter), and MLD (unit: meter) in the filter analysis (reference). The lower row: the global map of MSSS for monthly T300, SSH, and MLD.

SSH variability encompasses a broad range of factors, making it a comprehensive metric for capturing changes in ocean conditions. It reflects both steric and dynamic alterations within the oceanic system. SSH variations are closely related to the thickness of the upper layer and thermocline-depth variations in the tropical regions (Wyrtki and Kendall, 1967). The patterns of RMSE in monthly SSH are similar to T300's (Figure 3c) due to the contribution of thermosteric changes, except for the Southern Ocean, where dynamic changes play a key role in SSH variability. SSH exhibits MSSS scores slightly lower than T300, but their spatial patterns exhibit remarkable consistency (Figures 3b,d).

Ocean stratification is mainly driven by temperature in most places, and salinity plays a role in the high latitudes. We find large RMSEs in MLD in the North Pacific, North Atlantic, and Southern Ocean (Figure 3e). These regions extend into higher latitudes with a greater contrast in temperature between the surface waters and the deeper ocean. This contrast promotes vertical mixing, especially in winter, and contributes to a deeper MLD. In addition, ocean currents and gyres, such as the Antarctic Circumpolar Current and the North Atlantic Subtropical Gyre, lead to changes in the ocean's vertical structure, deepening the mixed layer. Figure 3f shows that the offline EnKS assimilating future SST data significantly reduces the error in monthly MLD. The MSSS score is globally homogeneous, except for sea ice cover regions where SST data is not used.

The offline EnKS significantly improves the reliability of the reanalysis for monthly SST. Figure 4 shows that the filter analysis has a larger combined error σ^f than the RMSE d^f. The offline EnKS reduces both the combined error σ and the RMSE d, but the reduction of the combined error is more prominent. It makes the two quantities, the combined error and the RMSE, closer in the smoother analysis, meaning the reanalysis becomes more reliable (Desroziers et al., 2005).

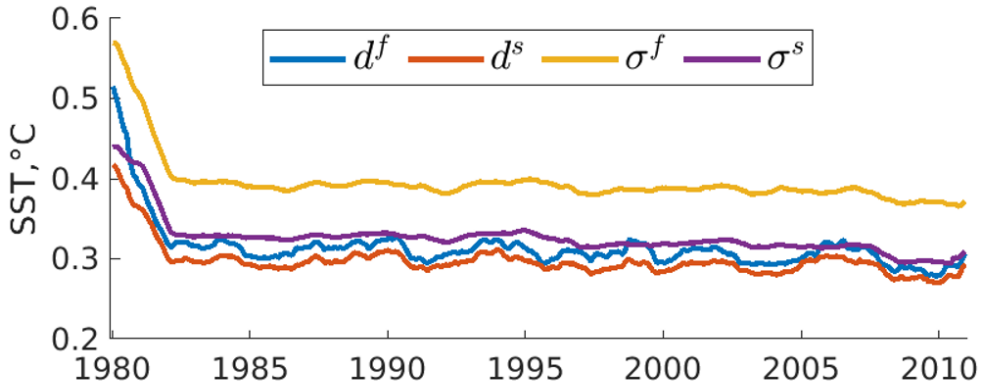

Figure 4. Global assimilation statistics for monthly SST (i.e., assimilated quantity): RMSE for the filter analysis (blue line), RMSE for the smoother analysis (red line), combined error for the filter analysis (orange line), and combined error for the smoother analysis (purple line). The time series is smoothed with a time window of 12 months.

Essentially, the offline EnKS uses an ensemble of pre-existing filter analyses to capture and represent the temporal relationship (i.e., covariance) between model states at different time steps and to correct and improve the filter analyses with future observational data. This approach enhances the accuracy of the reanalysis, particularly in regions where ocean currents are active and the error is large. Also, the reliability of the reanalysis is improved.

3.2 Real framework

It is important to note that, unlike in the above twin experiments, where the model was perfect, the Earth system model often has a large model bias. We calculate the scores in the anomaly field within the real framework, meaning the validation process does not consider model biases.

We tested different temporal localization parameters for the twin experiments (section 2.3). Several combinations of temporal localization parameters and observation error inflation factors lead to a similar performance (not shown in this paper). The larger the temporal localization parameter γ, the more we need to inflate the observation error. In the following, we use the temporal localization parameter $\gamma = 0.1$ to be consistent with section 3.1.

When validating on the monthly timescale in the real-world framework, we find the positive MSSS values for monthly T300, SSH, and MLD anomalies in most grids (please refer to Supplementary Information) but notably lower than those in twin experiments. Potential factors causing this inconsistency will be discussed in section 4. In the following discussion, we will show the results on the yearly timescale, meaning the validation for yearly T300, SSH, and MLD anomalies.

Overall, the offline EnKS yields positive globally averaged MSSS values in all realistic experiments for the three quantities (Figure 5). The improvement (or error reduction) is larger for MLD and lower for SSH, which is consistent with the results in the twin experiments (Figure 2). However, it is moderate for T300, which is inconsistent with the twin experiment results

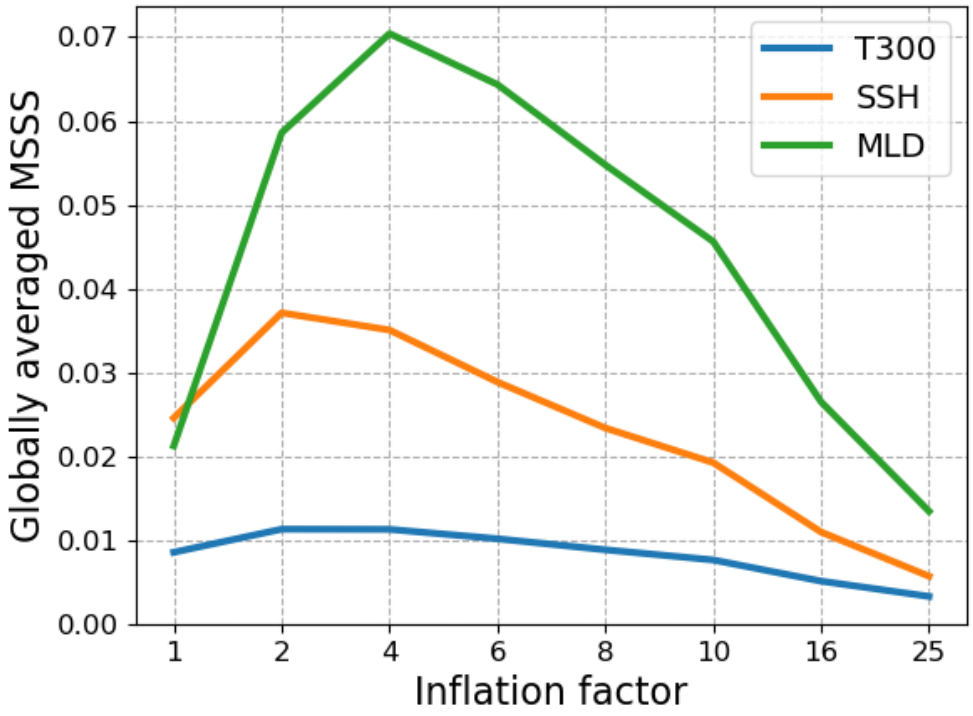

Figure 5. Globally averaged MSSS scores for different real-world experiments for yearly T300 (blue line), SSH (orange line), and MLD (green line) anomalies. In this experiment, $\gamma = 0.1$.

(Figure 2). It may be because sparse hydrographic profile measurements before the Argo float period (2000s) led to high uncertainties in the EN4 Objective Analysis (i.e., the reference dataset).

Note that the inflation of observation error variance (Stewart et al., 2008) was designed to neutralize that the errors of actual observations are correlated. According to overall MSSS values in Figure 5, we have to inflate observation error variance by four times to reach the optimal performance of the offline EnKS.

For yearly T300 anomalies, the large RMSE is found in the regions aligning with the ocean currents, e.g., the Gulf Stream and Agulhas current (Figure 6a). The offline EnKS leads to positive MSSS values in the regions with larger RMSE (Cosme et al., 2010), but negative MSSS values in some other regions (Figure 6b).

For yearly SSH anomalies, the spatial pattern of RMSE (Figure 6c) is similar to T300 anomalies (Figure 6a), except for the Southern Ocean, where ocean dynamic is a key factor for SSH variability and the quality of EN4 Objective Analysis is questionable. The magnitude of the MSSS values is larger than T300 anomalies (Figure 6). Many other factors influencing SSH variability, in addition to T300 anomalies, are improved by the offline EnKS. Furthermore, SSH observations have only

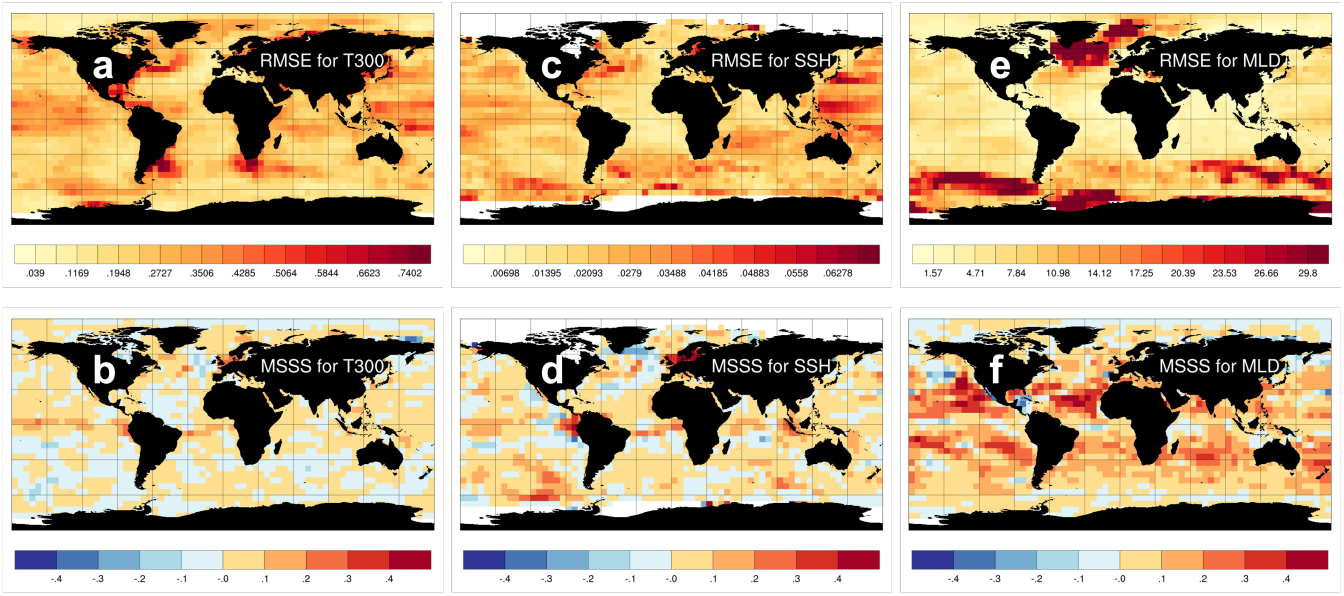

Figure 6. The upper row: the global maps of RMSE for yearly T300 anomalies (unit: °C), SSH anomalies (unit: meter), and MLD anomalies (unit: meter) in the filter analysis (reference). The lower row: the global map of MSSS for yearly T300, SSH, and MLD anomalies. In this experiment, SST observation error variance is inflated by a factor of four.

been available since 1993, and the MSSS of T300 anomalies was computed from 1950 to 2010. Their validation periods are different.

For yearly MLD anomalies (Figures 6e-f), significant errors are found in the North Atlantic SPG region, Norwegian Sea, Southern Ocean, and the Weddell Sea where deep convection occurs (Courtois et al., 2017). The positive MSSS scores are found in most regions, particularly subtropical areas. Overall, the improvement in MLD is more prominent than in T300 and SSH, consistent with Figure 5.

The offline EnKS improves the reliability of the reanalysis for monthly SST anomalies. Figure 7 shows the time series of the RMSE d and combined error σ for monthly SST anomalies in the filter and smoother analyses. The discontinuity in the combined error σ in the 1980s is due to the availability of satellite data dramatically increasing the quality of observations. While the offline EnKS reduces the RMSE d, the combined error σ decreases. Regarding reliability, in the satellite era, d and σ match better in the smoother analysis than in the filter analysis.

4 Conclusions, discussions and perspectives

In this study, we proposed implementing the EnKS as a post-processing method for long-term coupled reanalysis. Our approach involves formulating the EnKS offline and developing a dedicated algorithm for the offline EnKS. Notably, the offline EnKS algorithm's computational cost is significantly lower than the original production of the reanalysis.

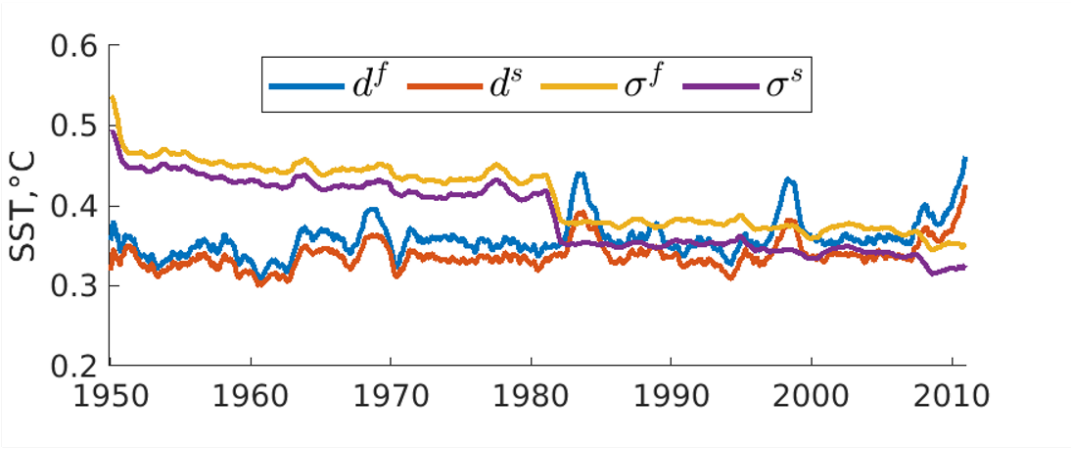

Figure 7. Global assimilation statistics for monthly SST anomalies (i.e., assimilated quantity): RMSE for the filter analysis (blue line), RMSE for the smoother analysis (red line), combined error for the filter analysis (orange line), and combined error for the smoother analysis (purple line). In this experiment, $\gamma = 0.1$, and the observation error variance is inflated by four. The time series is smoothed with a time window of 12 months.

We applied the offline algorithm to assimilate 'future' SST observations and update monthly SSH, MLD, temperature, and salinity at all ocean depths in the reanalyses of the Norwegian Climate Prediction Model (NorCPM) in an idealized framework, where the truth is known, and a real-world framework. Tuning temporal localization parameters reveals that excessively smaller or larger temporal parameters can compromise the offline EnKS's performance. The optimal temporal localization parameter is $\gamma = 0.1$, corresponding to a delay time of 13 days. Inflating the observation error variance can dampen the impact of the correlated observation errors in the real-world framework (Stewart et al., 2008), and an inflation of the factor of four leads to the best performance. In both cases, the offline EnKS improves the accuracy and reliability of the reanalyses, particularly in regions where ocean currents are active and the error is large. In these regions, ocean currents facilitated the effective use of 'future' SST observations to update analyses in a backward temporal manner.

Compared to the twin experiments, the accuracy improvement is notably lower in the real-world framework but still positive. In the real world, the true state of the Earth system remains unknown. Observational data is susceptible to diverse sources of error, including instrument inaccuracies, sampling inconsistencies, and measurement drift over time. Importantly, real-world observations exhibit spatial and temporal correlations, a fact omitted in the twin experiments. Furthermore, for proper validation, the lack of a definitive truth requires reliance on alternative measurements and datasets, each bearing its uncertainties. These uncertainties, coupled with potential discrepancies between assimilated and validation datasets, can introduce confounding factors that influence the validation of the offline EnKS in the real framework.

Our findings indicate that when solely assimilating SST data, the proposed offline method yields minimal improvements and even degradation in the ocean below 300 m (please refer to Supplementary Information). This is primarily because the covariances between SST and the subsurface are usually small in most places, and sampling error due to the finite ensemble

size can degrade interior ocean temperature variability in the North Atlantic Subpolar Gyre region (Counillon et al., 2016a) and contribute to drift in decadal climate predictions (Bethke et al., 2018). In future work, we could enhance the offline EnKS by incorporating the vertical localization approach developed in Wang et al. (2022) to improve the accuracy of the smoothed reanalysis in the interior ocean.

In this study, the proposed offline method assimilates SST data and updates only the ocean component of the coupled reanalysis. In future work, building on the concept of coupled data assimilation (Penny et al., 2017), the offline EnKS could be extended to update several components of the coupled Earth system reanalysis. For example, assimilating SST data could also update surface air temperature and sea ice concentration fields, which are strongly correlated with SST through air–sea and ocean–ice interactions.

In this study, the proposed offline method uses a dynamic ensemble to construct model error covariances, facilitating the correction of unobserved variables with observed ones. A crucial prerequisite for the method is that the pre-existing reanalysis should be stochastic and possess multiple realizations. When applying the proposed method to a deterministic reanalysis, one may turn to the ensemble Optimal Interpolation (EnOI) technique, a simplified variant of the EnKF (Evensen, 2003; Counillon and Bertino, 2009). Likewise, the offline EnKS could use a historical ensemble and be employed in deterministic reanalyses.

The proposed offline smoother algorithm primarily integrates 'future' observations not considered in the filter analysis (Figure 1). Notably, this offline approach can assimilate 'novel' observations not originally part of the reanalysis. To exemplify, the proposed algorithm can be applied to perform offline assimilation of hydrographic profile data or SSH observations in the reanalysis presented in Counillon et al. (2016a), which initially utilized only SST observations.

Code and data availability. The software code of the offline EnKS is available at https://doi.org/10.5281/zenodo.10169021. The coupled reanalysis from 1950 to 2010 is accessed at https://doi.org/10.11582/2016.00002. The EN4.2.1 objective analysis data is downloaded from https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html. The gridded dataset ARMOR-3D L4 is accessed at https://doi.org/10.48670/moi-00148. The MLD data is downloaded from https://doi.org/10.24381/cds.67e8eeb7.

Competing interests. The authors declare that no competing interests are present.

Acknowledgements. This study was supported by the Research Council of Norway (Grant Nos. 301396, 270061, and 350390), the Trond Mohn Foundation under project number BFS2018TMT01, and the Bjerknes Centre for Climate Research (BCCR) – Centre of Climate Dynamics (SKD) Strategic Project PARCIM (Proxy Assimilation for Reconstructing Climate and Improving Model). We thank Sigma2 for computing and storage resources (nn9039k and NS9039K).

References

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, O., Drange, H., Roelandt, C., Seierstad, I. a., Hoose, C., and Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1- Part 1: Description and basic evaluation of the physical climate, Geoscientific Model Development, 6, 687–720, https://doi.org/10.5194/gmd-6-687-2013, 2013.

Bethke, I., Wang, Y., Counillon, F., Kimmritz, M., Langehaug, H., Bentsen, M., and Keenlyside, N.: Subtropical North Atlantic preconditioning key to skillful subpolar gyre prediction, https://www.wcrp-climate.org/images/WCRP_conferences/S2S_S2D_2018/pdf/Programme/orals/presentations/B2-10_Ingo-Bethke.pdf, second International Conference on Seasonal to Decadal Prediction, 2018.

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., Samuelsen, A., Langehaug, H., Svendsen, L., Chiu, P.-G., Passos, L., Bentsen, M., Guo, C., Gupta, A., Tjiputra, J., Kirkevåg, A., Olivié, D., Seland, Ø., Solsvik Vågane, J., Fan, Y., and Eldevik, T.: NorCPM1 and its contribution to CMIP6 DCPP, Geoscientific Model Development, 14, 7073–7116, https://doi.org/10.5194/gmd-14-7073-2021, 2021.

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Quarterly Journal of the Royal Meteorological Society, 140, 1521–1535, https://doi.org/https://doi.org/10.1002/qj.2236, 2014.

Brune, S., Nerger, L., and Baehr, J.: Assimilation of oceanic observations in a global coupled Earth system model with the SEIK filter, Ocean Modelling, 96, 254–264, https://doi.org/10.1016/j.ocemod.2015.09.011, 2015.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives, Wiley Interdisciplinary Reviews: Climate Change, 9, e535, https://doi.org/10.1002/wcc.535, 2018.

Cosme, E., Brankart, J.-M., Verron, J., Brasseur, P., and Krysta, M.: Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation, Ocean Modelling, 33, 87–100, https://doi.org/https://doi.org/10.1016/j.ocemod.2009.12.004, 2010.

Counillon, F. and Bertino, L.: Ensemble Optimal Interpolation: Multivariate properties in the Gulf of Mexico, Tellus, Series A: Dynamic Meteorology and Oceanography, 61, 296–308, https://doi.org/10.1111/j.1600-0870.2008.00383.x, 2009.

Counillon, F., Bethke, I., Keenlyside, N., Bentsen, M., Bertino, L., and Zheng, F.: Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: a twin experiment, Tellus A, 66, 1–21, https://doi.org/10.3402/tellusa.v66.21074, 2014.

Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., and Bentsen, M.: Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model, Tellus A, 68, 1–17, https://doi.org/10.3402/tellusa.v68.32437, 2016a.

Counillon, F., Keenlyside, N., Wang, Y., and Bethke, I.: Norwegian Climate Prediction Model reanalysis with assimilation of SST anomaly: 1950-2010 [Data set], https://doi.org/10.11582/2016.00002, 2016b.

Courtois, P., Hu, X., Pennelly, C., Spence, P., and Myers, P. G.: Mixed layer depth calculation in deep convection regions in ocean numerical models, Ocean Modelling, 120, 60–78, https://doi.org/https://doi.org/10.1016/j.ocemod.2017.10.007, 2017.

Craig, A. P., Vertenstein, M., and Jacob, R.: A new flexible coupler for earth system modeling developed for CCSM4 and CESM1, Int. J. High Perform. Comput. Appl., 26, 31–42, https://doi.org/10.1177/1094342011428141, 2012.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., H?lm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis:

configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, https://doi.org/https://doi.org/10.1256/qj.05.108, 2005.

Dong, B., Haines, K., and Martin, M.: Improved High Resolution Ocean Reanalyses Using a Simple Smoother Algorithm, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 626, https://doi.org/https://doi.org/10.1029/2021MS002626, e2021MS002626 2021MS002626, 2021.

Dong, B., Bannister, R., Chen, Y., Fowler, A., and Haines, K.: Simplified Kalman smoother and ensemble Kalman smoother for improving reanalyses, Geoscientific Model Development, 16, 4233–4247, https://doi.org/10.5194/gmd-16-4233-2023, 2023.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynamics, 53, 343–367, https://doi.org/10.1007/s10236-003-0036-9, 2003.

Evensen, G. and van Leeuwen, P. J.: An Ensemble Kalman Smoother for Nonlinear Dynamics, Monthly Weather Review, 128, 1852–1867, https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2, 2000.

Eyring, V., Bony, S., Meehl, G. A., Senior, C., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation, Geoscientific Model Development Discussions, 8, 10 539–10 583, https://doi.org/10.5194/gmdd-8-10539-2015, 2015.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Q.J.R. Meteorol. Soc., 125, 723–757, https://doi.org/10.1002/qj.49712555417, 1999.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z. L., and Zhang, M.: The community climate system model version 4, Journal of Climate, 24, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1, 2011.

Good, S. a., Martin, M. J., and Rayner, N. a.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, Journal of Geophysical Research: Oceans, 118, 6704–6716, https://doi.org/10.1002/2013JC009067, 2013.

Grudzien, C. and Bocquet, M.: A fast, single-iteration ensemble Kalman smoother for sequential data assimilation, Geoscientific Model Development, 15, 7641–7681, https://doi.org/10.5194/gmd-15-7641-2022, 2022.

Halem, M. and Dlouhy, R.: Observing system simulation experiments related to space-borne Lidar wind profiling. Part 1: Forecast impacts of highly idealized observing systems, Res. Rev., 1983, 1984.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter, Monthly Weather Review, 129, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2, 2001.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Holland, M. M., Bailey, D. A., Briegleb, B. P., Light, B., and Hunke, E.: Improved sea ice shortwave radiation physics in CCSM4: the impact of melt ponds and aerosols on arctic sea ice, J. Climate, 25, 1413–1430, https://doi.org/10.1175/JCLI-D-11-00078.1, 2012.

Houtekamer, P. L. and Mitchell, H. L.: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, Monthly Weather Review, 129, 123 – 137, https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2, 2001.

Hunt, B., Kostelich, E., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D: Nonlinear Phenomena, 230, 112–126, https://doi.org/10.1016/j.physd.2006.11.008, 2007.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437–471, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Karspeck, A. R., Yeager, S., Danabasoglu, G., Hoar, T., Collins, N., Raeder, K., Anderson, J., and Tribbia, J.: An ensemble adjustment kalman filter for the CCSM4 ocean component, Journal of Climate, 26, 7392–7413, https://doi.org/10.1175/JCLI-D-12-00402.1, 2013.

Khare, S. P., Anderson, J. L., Hoar, T. J., and Nychka, D.: An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems, Tellus A, https://doi.org/10.1111/j.1600-0870.2007.00281.x, 2008.

Kirkevåg, A., Iversen, T., Seland, Ø., Hoose, C., Kristjánsson, J. E., Struthers, H., Ekman, A. M. L., Ghan, S., Griesfeller, J., Nilsson, E. D., and Schulz, M.: Aerosol-climate interactions in the Norwegian Earth System – NorESM1-M, Geoscientific Model Development, 6, 207–244, https://doi.org/10.5194/gmd-6-207-2013, 2013.

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Broennimann, S., Buizza, R., Dalhgren, P., Dee, D., Haimberger, L., Hersbach, H., Kosaka, Y., Martin, M., Poli, P., Rayner, N., Rustemeier, E., and Schepers, D.: CERA-20C: A Coupled Reanalysis of the Twentieth Century, Journal of Advances in Modeling Earth Systems, 10, 1172–1195, https://doi.org/10.1029/2018MS001273, 2018.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in version 4 of the community land model, J. Adv. Model. Earth Syst., 3, M03001, https://doi.org/10.1029/2011MS000045, 2011.

Murphy, A. H.: Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient, Monthly Weather Review, 116, 2417 – 2424, https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2, 1988.

Ngodock, H. E., Jacobs, G. A., and Chen, M.: The representer method, the ensemble Kalman filter and the ensemble Kalman smoother: A comparison study using a nonlinear reduced gravity ocean model, Ocean Modelling, 12, 378–400, https://doi.org/https://doi.org/10.1016/j.ocemod.2005.08.001, 2006.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., Levis, S., Swenson, S. C., Thornton, P. E., Dai, A., Decker, M., Dickinson, R., Feddema, J., Heald, C. L., Hoffman, F., Lamarque, J.-F., Mahowald, N., Niu, G.-Y., Qian, T., Randerson, J., Running, S., Sakaguchi, K., Slater, A., Stöckli, R., Wang, A., Yang, Z.-L., Zeng, X., and Zeng, X.: Technical Description of version 4.0 of the Community Land Model (CLM), Tech. Rep. NCAR/TN-478+STR, National Center for Atmospheric Research, Boulder, Colorado, USA, 2010.

O'Kane, T. J., Sandery, P. A., Kitsios, V., Sakov, P., Chamberlain, M. A., Collier, M. A., Fiedler, R., Moore, T. S., Chapman, C. C., Sloyan, B. M., and Matear, R. J.: CAFE60v1: A 60-Year Large Ensemble Climate Reanalysis. Part I: System Design, Model Configuration, and Data Assimilation, Journal of Climate, 34, 5153 – 5169, https://doi.org/10.1175/JCLI-D-20-0974.1, 2021.

Penny, S., Akella, S., Alves, O., Bishop, C., Buehner, M., Chevallier, M., Counillon, F., Draper, C., Frolov, S., Fujii, Y., Karspeck, A., Kumar, A., Laloyaux, P., Mahfouf, J.-F., Martin, M., Peña, M., de Rosnay, P., Subramanian, A., Tardif, R., Wang, Y., and Wu, X.: Coupled Data Assimilation for Integrated Earth System Analysis and Prediction: Goals, Challenges and Recommendations, Tech. Rep. WWRP 2017-3, World Meteorol. Org. (WMO), https://library.wmo.int/idurl/4/57666, 2017.

Raanes, P. N.: On the ensemble Rauch-Tung-Striebel smoother and its equivalence to the ensemble Kalman smoother, Quarterly Journal of the Royal Meteorological Society, 142, 1259–1264, https://doi.org/https://doi.org/10.1002/qj.2728, 2016.

Ravela, S. and McLaughlin, D.: Fast ensemble smoothing, Ocean Dynamics, 57, 123–134, https://doi.org/https://doi.org/10.1007/s10236-006-0098-6, 2007.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD002670, 2003.

Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., Van den Dool, H. M., Pan, H.-L., Moorthi, S., Behringer, D., Stokes, D., Peña, M., Lord, S., White, G., Ebisuzaki, W., Peng, P., and Xie, P.: The NCEP Climate Forecast System, Journal of Climate, 19, 3483–3517, https://doi.org/10.1175/JCLI3812.1, 2006.

Sakov, P. and Bertino, L.: Relation between two common localisation methods for the EnKF, Computational Geosciences, 15, 225–237, https://doi.org/10.1007/s10596-010-9202-6, 2010.

Sakov, P. and Oke, P. R.: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters, Tellus A, 60, 361–371, https://doi.org/10.1111/j.1600-0870.2007.00299.x, 2008.

Sakov, P., Counillon, F., Bertino, L., Lisæ ter, K. A., Oke, P. R., and Korablev, A.: TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Science, 8, 633–656, https://doi.org/10.5194/os-8-633-2012, 2012.

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., Brohan, P., Yin, X., Smith, C. A., Spencer, L. J., Vose, R. S., Rohrer, M., Conroy, R. P., Schuster, D. C., Kennedy, J. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Kubota, H., Lee, T. C., Lorrey, A. M., Luterbacher, J., Mock, C. J., Przybylak, R. K., Pudmenzky, C., Slonosky, V. C., Tinz, B., Trewin, B., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: An Evaluation of the Performance of the Twentieth Century Reanalysis Version 3, Journal of Climate, 34, 1417 – 1438, https://doi.org/10.1175/JCLI-D-20-0505.1, 2021.

Stewart, L. M., Dance, S. L., and Nichols, N. K.: Correlated observation errors in data assimilation, International Journal for Numerical Methods in Fluids, 56, 1521–1527, https://doi.org/https://doi.org/10.1002/fld.1636, 2008.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Vertenstein, M., Craig, T., Middleton, A., Feddema, D., and Fischer, C.: CESM1.0.3 User Guide, http://www.cesm.ucar.edu/models/cesm1.0/cesm/cesm_doc_1_0_4/ug.pdf,accessed:2015-01-23, 2012.

Wang, Y.: NorCPM's twin experiments over 1980-2010 [Data set], https://doi.org/10.11582/2023.00137, 2023.

Wang, Y., Counillon, F., Bethke, I., Keenlyside, N., Bocquet, M., and Shen, M.-l.: Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation, Ocean Modelling, 114, 33–44, https://doi.org/10.1016/j.ocemod.2017.04.007, 2017.

Wang, Y., Counillon, F., Keenlyside, N., Svendsen, L., Gleixner, S., Kimmritz, M., Dai, P., and Gao, Y.: Seasonal predictions initialised by assimilating sea surface temperature observations with the EnKF, Clim Dyn, 53, 5777–5797, https://doi.org/10.1007/s00382-019-04897-9, 2019.

Wang, Y., Counillon, F., Barthélémy, S., and Barth, A.: Benefit of vertical localization for sea surface temperature assimilation in isopycnal coordinate model, Frontiers in Climate, 4, https://doi.org/10.3389/fclim.2022.918572, 2022.

Wang, Y., Counillon, F., Svendsen, L., Chiu, P.-G., Keenlyside, N., Laloyaux, P., Koseki, M., and de Boisseson, E.: An ensemble-based coupled reanalysis of the climate from 1860 to the present (CoRea1860+), Earth System Science Data Discussions, 2025, 1–38, https://doi.org/10.5194/essd-2025-127, 2025.

Wyrtki, K. and Kendall, R.: Transports of the Pacific Equatorial Countercurrent, Journal of Geophysical Research (1896-1977), 72, 2073–2076, https://doi.org/https://doi.org/10.1029/JZ072i008p02073, 1967.

Zhang, S., Harrison, M. J., Rosati, a., and Wittenberg, a.: System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies, Monthly Weather Review, 135, 3541–3564, https://doi.org/10.1175/MWR3466.1, 2007.

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M.: The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment, Ocean Science, 15, 779–808, https://doi.org/10.5194/os-15-779-2019, 2019.