

Ocean satellite data assimilation using the implicit equal-weights variational particle smoother

Pinqiang Wang ^a, Mengbin Zhu ^{b,1}, Yan Chen ^a, Weimin Zhang ^{a,c,*}, Yi Yu ^a

^a College of Meteorology and Oceanology, National University of Defense Technology, Deya Road, 109, Changsha, 410073, China

^b Beijing Institute of Applied Meteorology, Beijing 100029, China

^c Key Laboratory of Software Engineering for Complex Systems, Changsha 410073, China

ARTICLE INFO

Keywords:

Implicit equal-weights
4D-var
Particle smoother
Ocean data assimilation

ABSTRACT

The implicit equal-weights variational particle smoother (IEWVPS) is a combination of the particle filter (PF) and weak-constraint 4-dimensional variational (4D-Var) method, that inherits the merits of both. The IEWVPS avoids the filter degeneracy of particle filters through an implicit equal-weights scheme and reduces the root mean square deviations (RMSDs) by introducing the 4D-Var method. This method has been tested using the Lorenz 96 model in a previous study, and we now implement it in the Regional Ocean Model System (ROMS), which is a realistic and complex ocean model. Two key problems, the representation of the analysis error covariance and the choice of the parameter α , were solved during this implementation. With an eddy-permitting model, satellite-based sea surface height (SSH) and sea surface temperature (SST) observations were assimilated with a set of 40 particles IEWVPS scheme. Compared with the ensemble 4D-Var method, the IEWVPS can reduce the bias introduced by perturbed atmospheric forcing, effectively improving temperature simulations in the upper 50 m while maintaining the RMSD of SSH at the same level. Therefore, the cooling effect caused by typhoons in the upper ocean is better characterized under the IEWVPS scheme than with previously used method. The ratio of RMSD to the ensemble spread indicates that the ensemble quality of the IEWVPS is much better than that of the ensemble 4D-Var. In addition, the computational cost of the IEWVPS is only slightly larger than that of the ensemble 4D-Var. One additional tangent linear model integration, one additional nonlinear model integration, and perturbation fields inputs/outputs are still needed.

1. Introduction

The particle filter (PF) is a nonlinear data assimilation method that approximates posterior probability density functions (PDFs) using a set of weighted particles. The main problem associated with applying the PF to a real geophysical system is filter degeneracy in high-dimensional situations. Filter degeneracy means that one particle has a weight close to 1, while the weights of the other particles tend toward zero. van Leeuwen et al. (2019) reviewed four categories of methods to prevent filter degeneracy, including the proposal density method, localization, transformation and hybridization. Although there are still many technical obstacles to make PFs operational in numerical weather prediction centers, many researchers and scientists have made their own efforts to implement PFs into real geophysical models. Ades and van Leeuwen (2015) tested the equivalent-weights particle filter (EWPF) (van Leeuwen, 2010; Ades and van Leeuwen, 2013) using a reduced-gravity system and found that the EWPF would not disrupt dynamical balances. Browne and van Leeuwen (2015) further tested the

EWPF using a coupled HadCM3 model with synthetic observations, and filter degeneracy was not observed, but the performance of the scheme was dependent on a nudging term (Browne, 2016). The localized particle filter (LPF) proposed by Poterjoy (2016) outperformed the EnKF method in the atmospheric Weather Research and Forecasting (WRF) model with a series of nonlinear observation operators (Poterjoy and Anderson, 2016; Poterjoy et al., 2019). In contrast from the LPF used by Poterjoy (2016), the local particle filter introduced by Penny and Miyoshi (2016) was based on state-domain localization, a method that is quite similar to the LETKF (local ensemble transform Kalman filter). When implemented in the COSMO model, the PF and LETKF hybrid method improved the forecasting of non-Gaussian variables (Robert et al., 2018). The localized adaptive particle filter (LAPF) was the first particle filter to be implemented in the global operational system at Deutscher Wetterdienst (DWD) (Potthast et al., 2019), and its performance was comparable with that of the LETKF method. Under the framework of a LPF, Chen et al. (2020b) proposed a localized

* Corresponding author at: College of Meteorology and Oceanology, National University of Defense Technology, Deya Road, 109, Changsha, 410073, China.

E-mail addresses: wangpinqiang16@nudt.edu.cn (P. Wang), zhumengbin@nudt.edu.cn (M. Zhu), chenyan16@nudt.edu.cn (Y. Chen), weiminzhang@nudt.edu.cn (W. Zhang), yuyi2019@nudt.edu.cn (Y. Yu).

¹ Co-first author.

weighted ensemble Kalman filter (LWEnKF), and applied it to a real ocean data assimilation scheme in the Regional Ocean Modeling System (ROMS). Comparisons have been made among the LWEnKF, EnKF and LPF, indicating that the LWEnKF is the most accurate method for unobserved ocean state variables, while the accuracies of observed variables are maintained at the same levels as those resulting from the EnKF method (Chen et al., 2020a).

The 4-dimensional variational (4D-Var) method has been widely used in numerical weather predictions due to the benefits of its dynamical constraints and that it allows for additional terms in the cost function, e.g., variational bias correction, variational quality control, digital filter initialization, weak constraint term, etc. (Bonavita et al., 2017). Since the minimization process required by the implicit particle filter method is similar to that used in the variational method, Atkins et al. (2013) explored the inner equivalent connection between the implicit particle filter (Chorin and Tu, 2009) and 4D-Var methods. Morzfeld et al. (2018) introduced a variational particle smoother method that eliminated the impact of filter degeneracy through localization. Wang et al. (2020) illustrated a scheme called the implicit equal-weights variational particle smoother (IEWVPS) that avoided filter degeneracy through the implicit equal-weights scheme proposed by Zhu et al. (2016) and combined with this scheme the weak-constraint 4D-Var method. The Lorenz96 experimental results showed that this new scheme outperformed a simple ensemble 4D-Var scheme in the ensemble quality and reduced the root mean square error of the ensemble members when compared with outputs of the implicit equal-weights particle filter (IEWPF) and LETKF methods.

Considering the success of the IEWVPS in the Lorenz96 model, we further implemented the IEWVPS scheme in the weak-constraint 4D-Var framework of ROMS and conducted real ocean satellite data assimilation experiments. Compared with the simple Lorenz96 model, ROMS has more complicated dynamic and thermodynamic constraints and much larger model dimensions. The major challenge of the implementation of the IEWVPS method in a real geophysical system is the representation of the analysis error covariance matrix of the proposal density, which is part of the calculation of the equal-weights adjustment component. Furthermore, dimensions of the analysis error matrix are dependent not only on model dimensions but also on the length of the time window. It is impossible to calculate the analysis error matrix explicitly in a high dimensional system. In this study, we introduce a random method to solve the problem of calculating the analysis error covariance matrix of this new scheme. The response of the upper ocean to a typhoon is used as the test object, and satellite-based sea surface height (SSH) and sea surface temperature (SST) observations are assimilated in the experiments.

This paper is organized as follows. In Section 2, we describe the core algorithm and the implementation of the IEWVPS in the ROMS weak-constraint 4D-Var framework in detail, as well as the random method of deriving the analysis error covariance matrix. Section 3 introduces detailed experimental information. Section 4 illustrates the performance when assimilating the SSH and SST observations. The conclusions and discussions are included in Section 5.

2. Implementation of implicit equal-weights variational particle smoother in ROMS

The IEWVPS is an extension of the IEWP (Zhu et al., 2016), which is based on the ideas of proposal density, implicit sampling (Chorin and Tu, 2009) and equal-weights (van Leeuwen, 2010). Weak-constraint 4D-Var is chosen as a proposal density. The implicit part of the scheme follows from drawing samples implicitly from a standard Gaussian distributed proposal density instead of the original one. The equal-weights property is achieved by setting the weight of each particle to a target weight. Combining these ideas, the position of each particle can be expressed as the mode of the chosen proposal density plus a scaled random perturbation (denoted as the equal-weights adjustment component in this study). The core algorithm of the IEWVPS

is introduced in Section 2.1. However, the equal-weights adjustment component requires the square root of the covariance matrix of the chosen proposal density, and this matrix is impossible to calculate explicitly due to the high dimension of the system. Thus, we have to approximate the equal-weights adjustment component, and a random method is used for this purpose, which is described in Section 2.2. To obtain the position of each particle, the equal-weights adjustment component operates on each model time step though a nonlinear model operator.

2.1. The core algorithm of the IEWVPS

According to Bayes' theorem, the posterior PDF of model state $\mathbf{x}^{0:n}$, given the observations $\mathbf{y}^{1:n}$ during the time window $[0, n]$, can be written as follows:

$$p(\mathbf{x}^{0:n}|\mathbf{y}^{1:n}) = \frac{p(\mathbf{x}^{0:n})p(\mathbf{y}^{1:n}|\mathbf{x}^{0:n})}{p(\mathbf{y}^{1:n})} \quad (1)$$

where $\mathbf{x}^{0:n} = (\mathbf{x}^0, \dots, \mathbf{x}^n)^T$ and $\mathbf{y}^{0:n} = (\mathbf{y}^0, \dots, \mathbf{y}^n)^T$.

Applying the idea of proposal density, we multiply and divide the right-hand side in Eq. (1) by the same factor $q(\mathbf{x}^{0:n}|\mathbf{y}^{1:n})$, which is the so-called proposal density, leading to the following equation.

$$p(\mathbf{x}^{0:n}|\mathbf{y}^{1:n}) = \frac{p(\mathbf{x}^{0:n})p(\mathbf{y}^{1:n}|\mathbf{x}^{0:n})}{p(\mathbf{y}^{1:n})q(\mathbf{x}^{0:n}|\mathbf{y}^{1:n})} q(\mathbf{x}^{0:n}|\mathbf{y}^{1:n}) \quad (2)$$

Drawing samples from the proposal transition density $q(\mathbf{x}^{0:n}|\mathbf{y}^{1:n})$ leads the posterior PDF to be expressed as:

$$\begin{aligned} p(\mathbf{x}^{0:n}|\mathbf{y}^{1:n}) &= \frac{1}{N_e} \sum_{i=1}^{N_e} \frac{p(\mathbf{y}^{1:n}|\mathbf{x}_i^{0:n})p(\mathbf{x}_i^{0:n})}{p(\mathbf{y}^{1:n})q(\mathbf{x}_i^{0:n}|\mathbf{y}^{1:n})} \delta(\mathbf{x}^{0:n} - \mathbf{x}_i^{0:n}) \\ &= \frac{1}{N_e} \sum_{i=1}^{N_e} w_i \delta(\mathbf{x}^{0:n} - \mathbf{x}_i^{0:n}) \end{aligned} \quad (3)$$

where N_e is the number of particles and w_i is the weight of particle i . The term w_i can be calculated using the equation below.

$$w_i = \frac{p(\mathbf{y}^{1:n}|\mathbf{x}_i^{0:n})}{p(\mathbf{y}^{1:n})} \frac{p(\mathbf{x}_i^{0:n})}{q(\mathbf{x}_i^{0:n}|\mathbf{y}^{1:n})} \quad (4)$$

Here, the assumptions are as follows: (1) the errors between \mathbf{x}^0 and the initial condition of the background \mathbf{x}^b are Gaussian-distributed with covariance \mathbf{B} ; (2) the observation errors have a Gaussian distribution with covariance \mathbf{R} , and the observations are temporally independent; and (3) the model errors are Gaussian-distributed with covariance \mathbf{Q} . Under these assumptions, the numerator of the weight of each particle can be expressed in terms of the cost function of the weak-constraint 4D-Var:

$$p(\mathbf{y}^{1:n}|\mathbf{x}_i^{0:n})p(\mathbf{x}_i^{0:n}) \propto \exp[-J_i(\mathbf{x}_i^{0:n})] \quad (5)$$

where $J_i(\mathbf{x}_i^{0:n})$ is the cost function of the weak-constraint 4D-Var for particle i . The term $J_i(\mathbf{x}_i^{0:n})$ is expressed as follows:

$$\begin{aligned} J_i(\mathbf{x}_i^{0:n}) &= \frac{1}{2} \left\| \mathbf{x}_i^0 - \mathbf{x}_i^b \right\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \sum_{k=1}^n \left\| \mathbf{y}^k - \mathcal{H}^k(\mathbf{x}_i^k) \right\|_{\mathbf{R}_k^{-1}}^2 \\ &\quad + \frac{1}{2} \sum_{k=1}^n \left\| \mathbf{x}_i^k - \mathcal{M}^k(\mathbf{x}_i^{k-1}) \right\|_{\mathbf{Q}_k^{-1}}^2 \end{aligned} \quad (6)$$

in which \mathcal{H} and \mathcal{M} are the nonlinear observation and model operator, respectively. The notation of $\left\| \mathbf{x}_i^0 - \mathbf{x}_i^b \right\|_{\mathbf{B}^{-1}}^2$ represents $(\mathbf{x}_i^0 - \mathbf{x}_i^b)^T \mathbf{B}^{-1} (\mathbf{x}_i^0 - \mathbf{x}_i^b)$, and the same notation is used for the other terms in Eq. (6). The numerator of the weight for each particle, calculated with Eq. (4), can be further simplified as follows:

$$p(\mathbf{y}^{1:n}|\mathbf{x}_i^{0:n})p(\mathbf{x}_i^{0:n}) = N(\mathbf{x}_i^{a,0:n}, \mathbf{P}_i) \exp(-\frac{1}{2} \phi_i) \quad (7)$$

where $N(\mathbf{x}_i^{a,0:n}, \mathbf{P}) = C \exp[-\frac{1}{2} (\mathbf{x}^{0:n} - \mathbf{x}^{a,0:n})^T \mathbf{P}^{-1} (\mathbf{x}^{0:n} - \mathbf{x}^{a,0:n})]$, C is a constant, $\mathbf{x}_i^{a,0:n} = \arg \min J_i(\mathbf{x}_i^{0:n})$ is the minimizer of the weak-constraint 4D-Var cost function, $\frac{1}{2} \phi_i = \min J_i(\mathbf{x}_i^{0:n})$ is the minimum

value of the cost function, \mathbf{P}_i is the inverse of the Hessian of the cost function $J_i(\mathbf{x}_i^{0:n})$, and the superscript “a” in $\mathbf{x}_i^{a,0:n}$ means “analysis”.

The proposal transition density is now chosen to be Gaussian with a mean $\mathbf{x}_i^{a,0:n}$ and covariance \mathbf{P}_i for each particle, such that the following expression can be constructed.

$$q(\mathbf{x}_i^{0:n} | \mathbf{y}^{1:n}) = N(\mathbf{x}_i^{a,0:n}, \mathbf{P}_i) \quad (8)$$

The implicit part of the scheme involves implicitly drawing samples from a standard Gaussian proposal density, $q(\xi^{0:n}) = N(0, \mathbf{I})$, instead of from the original proposal density. These two proposal densities are related by the following expression:

$$q(\mathbf{x}^{0:n} | \mathbf{y}^{1:n}) = \frac{q(\xi^{0:n})}{\left\| \frac{d\mathbf{x}^{0:n}}{d\xi^{0:n}} \right\|} \quad (9)$$

where $\left\| \frac{d\mathbf{x}^{0:n}}{d\xi^{0:n}} \right\|$ denotes the absolute value of the determinant of the Jacobian matrix of the transformation $\mathbf{x}^{0:n} = g(\xi^{0:n})$. In the IEWVPS, the transformation is defined as follows:

$$\begin{aligned} \mathbf{x}_i^{e,0:n} &= g(\xi_i^{0:n}) \\ &= \mathbf{x}_i^{a,0:n} + \alpha_i^{1/2} \mathbf{P}_i^{1/2} \xi_i^{0:n} \end{aligned} \quad (10)$$

where α_i is a scalar that is to be determined for each particle, and the superscript “e” in $\mathbf{x}_i^{e,0:n}$ means “equal-weights”.

According to Eq. (9), the weight of each particle in Eq. (4) is now given by the following formula.

$$w_i = \frac{p(\mathbf{y}^{1:n} | \mathbf{x}_i^{0:n}) p(\mathbf{x}_i^{0:n})}{p(\mathbf{y}^{1:n}) q(\xi_i^{0:n})} \left\| \frac{d\mathbf{x}^{0:n}}{d\xi^{0:n}} \right\| \quad (11)$$

Neglecting $p(\mathbf{y}^{1:n})$ and taking the $-2 \log$ of both sides gives the following expression.

$$-2 \log w_i = -2 \log p(\mathbf{y}^{1:n} | \mathbf{x}_i^{0:n}) p(\mathbf{x}_i^{0:n}) + 2 \log q(\xi_i^{0:n}) - 2 \log \left(\left\| \frac{d\mathbf{x}^{0:n}}{d\xi^{0:n}} \right\| \right) \quad (12)$$

Using Sylvester's determinant lemma and the random map of Eq. (10), $\left\| \frac{d\mathbf{x}^{0:n}}{d\xi^{0:n}} \right\|$ can be reduced to the following term.

$$\left\| \frac{d\mathbf{x}^{0:n}}{d\xi^{0:n}} \right\| = \alpha_i^{N_z/2} \left\| \mathbf{P}_i^{1/2} \right\| \left| 1 + \frac{\xi_i^{0:n}}{\alpha_i^{1/2}} \frac{\partial \alpha_i^{1/2}}{\partial \xi_i^{0:n}} \right| \quad (13)$$

where N_z is the dimension of $\mathbf{x}^{0:n}$ and is depending on the model dimension and the length of the time window.

Using Eq. (7), the $-2 \log w_i$ becomes the following expression.

$$\begin{aligned} -2 \log w_i &= (\mathbf{x}_i^{0:n} - \mathbf{x}_i^{a,0:n})^T \mathbf{P}_i^{-1} (\mathbf{x}_i^{0:n} - \mathbf{x}_i^{a,0:n}) + \phi_i \\ &\quad - (\xi_i^{0:n})^T \xi_i^{0:n} - 2 \log \left(\left\| \frac{d\mathbf{x}^{0:n}}{d\xi^{0:n}} \right\| \right) \\ &= (\alpha_i - 1)(\xi_i^{0:n})^T \xi_i^{0:n} + \phi_i - 2 \log \left(\left\| \frac{d\mathbf{x}^{0:n}}{d\xi^{0:n}} \right\| \right) \\ &= (\alpha_i - 1)(\xi_i^{0:n})^T \xi_i^{0:n} + \phi_i - 2 \log \alpha_i^{N_z/2} \\ &\quad - 2 \log \left(\left| 1 + \frac{\xi_i^{0:n}}{\alpha_i^{1/2}} \frac{\partial \alpha_i^{1/2}}{\partial \xi_i^{0:n}} \right| \right) - 2 \log \left(\left\| \mathbf{P}_i^{1/2} \right\| \right) \end{aligned} \quad (14)$$

To satisfy the equal-weights property, the weight of each particle is set to a target weight $w_i(\alpha_i) = w_{target}$. For this purpose, the α_i of each particle must satisfy the following equation.

$$(\alpha_i - 1)(\xi_i^{0:n})^T \xi_i^{0:n} - 2 \log \alpha_i^{N_z/2} - 2 \log \left(\left| 1 + \frac{\xi_i^{0:n}}{\alpha_i^{1/2}} \frac{\partial \alpha_i^{1/2}}{\partial \xi_i^{0:n}} \right| \right) = c_i \quad (15)$$

where $c_i = C - \phi_i$, C is a constant and $\frac{1}{2}\phi_i$ is the minimum of the weak-constraint 4D-Var cost function. If we write $\xi_i^{0:n}$ in a simplified

form as ξ_i , Eq. (15) has the same form as that used in Zhu et al. (2016) and Skauvold et al. (2019):

$$(\alpha_i - 1)\xi_i^T \xi_i - 2 \log \alpha_i^{N_z/2} - 2 \log \left(\left| 1 + \frac{\xi_i}{\alpha_i^{1/2}} \frac{\partial \alpha_i^{1/2}}{\partial \xi_i} \right| \right) = c_i \quad (16)$$

The detailed solution of α_i can be found in Zhu et al. (2016) or Skauvold et al. (2019). Eq. (16) has analytical solutions in terms of the Lambert W function, but it is solved using the Newton iterative method for practical reasons.

If $c_i = 0$, then the value of α becomes a single constant solution with a value of 1. When $c_i \neq 0$, there are two branches of α that exist: > 1 and < 1 . As explored by Zhu et al. (2016), sampling on the > 1 branch gives a large ensemble spread, while sampling on the < 1 branch yields the opposite effect. A 50% random sampling scheme on the two branches achieves the most stable results in the IEWPF, but the results are under-dispersed in the IEWVPS (Wang et al., 2020). Thus, α is sampled on the > 1 branch in this study. It should be noted that the proposal density distribution has a gap as discussed in Zhu et al. (2016), that leads to a systematic bias. Skauvold et al. (2019) eliminated this gap by sampling α only on the < 1 branch and moving the particles twice. In this study, we followed the implementation of the IEWPF of Zhu et al. (2016) for two reasons. One reason is that no parameter tuning is required in the method of Zhu et al. (2016), so that it is easy to implement. The other is that the sampling ratio between the two branches is treated as a way to tune the ensemble spread.

2.2. Estimation of the equal-weights adjustment component

In Eq. (10), the posterior states are divided into two components: the first, $\mathbf{x}_i^{a,0:n}$, is the deterministic part, and the second, $\alpha_i^{1/2} \mathbf{P}_i^{1/2} \xi_i^{0:n}$, which we refer to is the equal-weights adjustment part. Once α_i is solved, $\mathbf{x}_i^{a,0:n}$ that satisfy the equal-weights property can be found using Eq. (10). However, $\mathbf{P}_i^{1/2}$ is impossible to calculate directly due to the high dimensions. Thus, we have to estimate $\mathbf{P}_i^{1/2} \xi_i^{0:n}$ indirectly, meaning that we need to first know the structure of \mathbf{P} in the ROMS 4D-Var system. The ROMS weak-constraint 4D-Var method is an incremental 4D-Var method that is solved in the observation space (Moore et al., 2011b). The entire cost function of the ROMS weak-constraint 4D-Var method (for a time window $[0, n]$) can be written as follows:

$$J(\delta\mathbf{x}^0, \boldsymbol{\eta}^1, \dots, \boldsymbol{\eta}^n) = \frac{1}{2} \left\| \delta\mathbf{x}^0 \right\|_{\mathbf{B}_x^{-1}}^2 + \frac{1}{2} \sum_{k=1}^n \left\| \mathbf{d}^k - \mathbf{H} \delta\mathbf{x}^k \right\|_{\mathbf{R}_k^{-1}}^2 + \frac{1}{2} \sum_{k=1}^n \left\| \boldsymbol{\eta}^k \right\|_{\mathbf{Q}_k^{-1}}^2 \quad (17)$$

where \mathbf{x} represents the model states of ROMS, which includes the potential temperature, salinity, sea surface height and velocities, $\delta\mathbf{x}$ means the increments of state vector \mathbf{x} , $\boldsymbol{\eta}$ is the model error, \mathbf{B}_x represents the background error covariance matrix of the initial condition, \mathbf{Q}_k is the model error covariance matrix at time k , and \mathbf{R}_k is the observation error covariance at time k . The vector $\mathbf{d}^k = \mathbf{y} - \mathcal{H}(\mathbf{x}^k)$ is the innovation vector at time k . \mathcal{H} is an observation operator that transforms the model state vector to the observed variables in the observation space, and \mathbf{H} is the tangent linearization of \mathcal{H} .

To obtain the expression of \mathbf{P} , we introduce the control vector, $\delta\mathbf{z} = ((\delta\mathbf{x}^0)^T, \dots, (\boldsymbol{\eta}^n)^T, \dots)^T$, where $k = 1, \dots, n$ and rewrite the cost function in a compact form, as follows (Moore et al., 2011b):

$$J(\delta\mathbf{z}) = \frac{1}{2} \delta\mathbf{z}^T \mathbf{D}^{-1} \delta\mathbf{z} + \frac{1}{2} (\mathbf{d} - \mathbf{G} \delta\mathbf{z})^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{G} \delta\mathbf{z}) \quad (18)$$

where $\mathbf{G} = (\dots, (\mathbf{H}\mathbf{M}^k)^T, \dots)^T$ is considered as a combined forward model and observation operator, which propagates the initial state forward in time to match the observation time through tangent linear model (\mathbf{M}). The vector $\mathbf{d} = (\dots, (\mathbf{d}^k)^T, \dots)^T$ is the innovation vector

over the assimilation window. \mathbf{D} and \mathbf{R} are block diagonal matrices, as expressed below.

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & & \\ & \mathbf{R}_2 & \cdots \\ & & \ddots \\ & & & \mathbf{R}_n \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} \mathbf{B}_x & & \\ & \mathbf{Q}_1 & \\ & & \ddots \\ & & & \mathbf{Q}_n \end{pmatrix}$$

In this study, R_k is considered to be a diagonal matrix at time k . B_x is assumed to be factorized as $B_x = K_b \Sigma_x C \Sigma_x^T K_b^T$, where K_b is the balance operator, C is a univariate correlation matrix and Σ_x is the matrix of prior standard deviations for all control variables (as shown in Fig. 3d and f for SSH and SST, respectively). Q_k is assumed to have the same structure as B_x (Moore et al., 2011a), but $\Sigma_q = 0.05 * \Sigma_x$.

In the formulation of Eq. (18), the Hessian matrix S_H is the second derivative of the cost function $J(\delta z)$.

$$S_H = \mathbf{D}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \quad (19)$$

The analysis error covariance matrix is the inverse of S_H (Fisher and Courtier, 1995).

$$\mathbf{P} = (\mathbf{D}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G})^{-1} \quad (20)$$

where time dimension is included in \mathbf{D} , \mathbf{G} , \mathbf{R} and \mathbf{P} . If the model dimension is N_x and the length of time window is n , the dimension of the control vector is $N_z = N_x \times (n+1)$. Thus, the dimension of \mathbf{P} is $N_z \times N_z$, which is the same as that of \mathbf{D} . If the dimension of observations is N_y over the time window, the dimensions of \mathbf{R} and \mathbf{G} are $N_y \times N_y$ and $N_y \times N_z$, respectively. In a high dimensional system, \mathbf{P} and \mathbf{D} cannot be calculated and inverted explicitly. Therefore, we have to estimate \mathbf{P} indirectly and use it implicitly. We only need $\mathbf{P}^{1/2} \xi_i^{0:n}$ for each particle; the dimension of $\mathbf{P}^{1/2} \xi_i^{0:n}$, $N_z \times N_e$, is much smaller than those of \mathbf{P} and $\mathbf{P}^{1/2}$. We can estimate $\mathbf{P}^{1/2} \xi_i^{0:n}$ using a random method that was reported in Wang et al. (2020) as follows.

Extracting $\mathbf{D}^{-1/2}$ out of the parentheses, we get:

$$\begin{aligned} \mathbf{P} &= (\mathbf{D}^{-T/2} \mathbf{D}^{-1/2} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G})^{-1} \\ &= \{\mathbf{D}^{-T/2} [I + \mathbf{D}^{T/2} \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \mathbf{D}^{1/2}] \mathbf{D}^{-1/2}\}^{-1} \\ &= \mathbf{D}^{1/2} [I + (\mathbf{R}^{-1/2} \mathbf{G} \mathbf{D}^{1/2})^T (\mathbf{R}^{-1/2} \mathbf{G} \mathbf{D}^{1/2})]^{-1} \mathbf{D}^{T/2} \end{aligned} \quad (21)$$

Performing a singular value decomposition (SVD) of $\mathbf{R}^{-1/2} \mathbf{G} \mathbf{D}^{1/2}$:

$$\mathbf{R}^{-1/2} \mathbf{G} \mathbf{D}^{1/2} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (22)$$

where \mathbf{U} and \mathbf{V} are orthogonal unitary matrices containing the left and right singular vectors respectively, and \mathbf{S} contains the singular values ordered in descending order of magnitude.

We then obtain the following expression for $\mathbf{P}^{1/2}$.

$$\begin{aligned} \mathbf{P} &= \mathbf{D}^{1/2} [\mathbf{I} + \mathbf{V} \mathbf{S} \mathbf{U}^T \mathbf{U} \mathbf{S}^T \mathbf{V}^T]^{-1} \mathbf{D}^{T/2} \\ &= \mathbf{D}^{1/2} \mathbf{V} (\mathbf{I} + \mathbf{S}^2)^{-1} \mathbf{V}^T \mathbf{D}^{T/2} \\ \mathbf{P}^{1/2} &= \mathbf{D}^{1/2} \mathbf{V} (\mathbf{I} + \mathbf{S}^2)^{-1/2} \end{aligned} \quad (23)$$

We cannot calculate $\mathbf{P}^{1/2}$ directly using Eq. (23), since large matrix operations are needed and \mathbf{G} is difficult to be expressed explicitly. Instead of using Eq. (23), we use a random method.

For convenience, we define the ensemble of $\xi^{0:n}$ as $\xi = [\xi_1^{0:n}, \dots, \xi_{N_e}^{0:n}]$, and introduce $\mathbf{q}_0(i, :) = [\mathbf{G} \mathbf{D}^{1/2} \xi_i - \mathbf{G} \mathbf{D}^{1/2} \xi], i = 1, \dots, N_e$, where $N_e = 40$ is the ensemble size used in our study. The overline represents the ensemble mean. Instead of performing SVD on $\mathbf{R}^{-1/2} \mathbf{G} \mathbf{D}^{1/2}$, we perform SVD on $\mathbf{R}^{-1/2} \mathbf{q}_0 = \mathbf{U} \mathbf{S} \mathbf{V}^T$. The random component, $\mathbf{P}^{1/2} \xi$, can be expressed as follows.

$$\mathbf{P}^{1/2} \xi = \mathbf{D}^{1/2} \xi \mathbf{V} (\mathbf{I} + \mathbf{S}^2)^{-1/2} \quad (24)$$

The dimensions of $\mathbf{R}^{-1/2} \mathbf{G} \mathbf{D}^{1/2}$ are $N_y \times N_z$ in Eq. (23), and the dimension of $\mathbf{R}^{-1/2} \mathbf{q}_0$ are $N_y \times N_e$ in Eq. (24); the latter is much smaller than the dimensions of \mathbf{D} or \mathbf{P} since the observation dimensions is much smaller than the model dimension. This kind of situation usually occurs in ocean data assimilation. In Eq. (24), $\mathbf{D}^{1/2}$, \mathbf{G} and \mathbf{R} perform

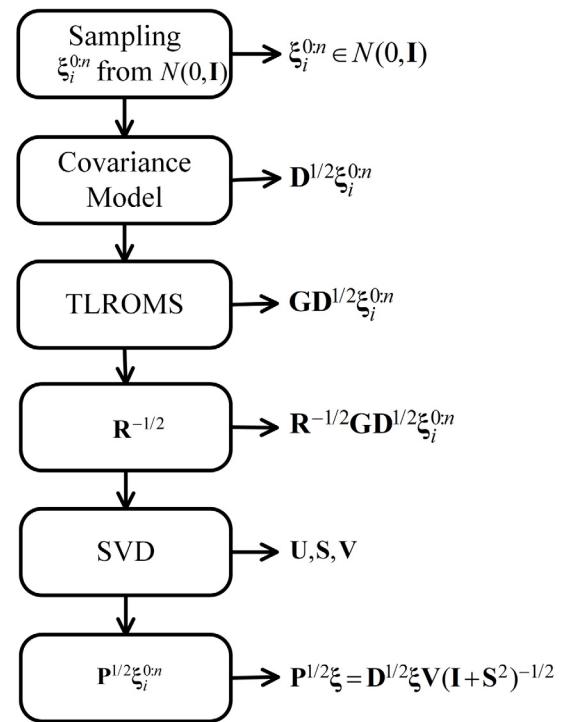


Fig. 1. Estimation of the analysis error covariance square root $P^{1/2} \xi$. The subscript of ξ denotes the particle, and the superscript denotes the time index. The TLROMS represents the ROMS tangent linear model.

as operators and there is no need to explicitly calculate these values, so large matrix operations can be avoided. The calculation process flow chart is shown in Fig. 1. As shown, $\mathbf{D}^{1/2} \xi_i^{0:n}$ appears twice; therefore, we store it in a netcdf file after the first use. Overall, the whole IEWPS process includes 4 steps:

- (1) Run a weak-constraint 4D-Var for each particle, and then store the minimum of the cost function ($\min J_i$), as well as the initial condition x_i^0 at the beginning of time window;
- (2) Calculate α in Eq. (15) using the Newton iterative method.
- (3) Estimate the random part of each particle, $\mathbf{P}^{1/2} \xi_i$, according to the process shown in Fig. 1, then multiply it with $\alpha^{1/2}$ to generate the equal-weights adjustment component;
- (4) Move each particle from the mode of the proposal density to the equal-weights position according to Eq. (10). This is done by integrating the nonlinear model from x^0 to time n using the equal-weights adjustment component as a forcing term.

We know that the optimization of 4D-Var is across the entire time window, so that particles sampled in the IEWPS are from a set of 4D trajectories that are valid at each time step throughout the window, not only at the beginning of the time window. Therefore, the equal-weights adjustment component operates on each model time step practically and theoretically. Although we only mentioned the initial condition x^0 at the beginning of the time window in step (1) above, the 4D trajectories of the 4D-Var minimization can be obtained by integrating the nonlinear model \mathcal{M} with the model error η as follows,

$$\mathbf{x}^k = \mathcal{M}^{k,k-1}(\mathbf{x}^{k-1}) + \boldsymbol{\eta}^k \quad (25)$$

It should be noted that the model error is assumed to be additive and not state-dependent in the ROMS weak-constraint 4D-Var, as shown in Eq. (25).

In step (4), the equal-weights adjustment component across the whole time window are not directly added to the 4D trajectories of the 4D-Var. The 4D trajectories of the 4D-Var minimization are adjusted using the equal-weights adjustment component as follows:

$$\mathbf{x}^k = \mathcal{M}^{k,k-1}(\mathbf{x}^{k-1}) + \boldsymbol{\eta}^k + \alpha^{1/2} \mathbf{P}^{1/2} \xi \quad (26)$$

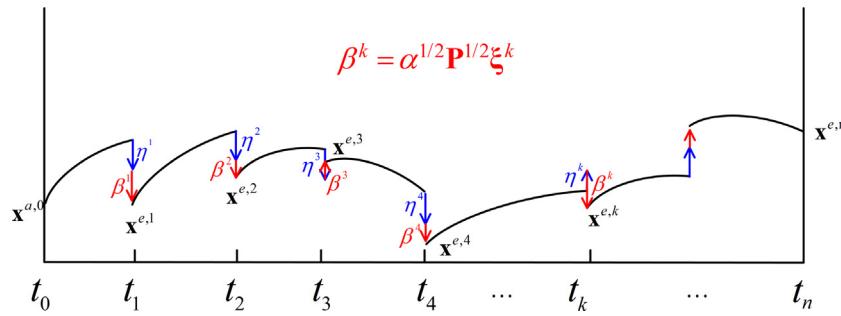


Fig. 2. Schematic diagram of the IEWVPS process on each time step for each particle. “a” means “analysis”, “e” mean “equal-weights”. Blue arrows denote the modifications made by the model error and red arrow denotes the modifications made by the equal-weights adjustment component. In this figure, $\beta^k = \alpha^{1/2} P^{1/2} \xi^k$.

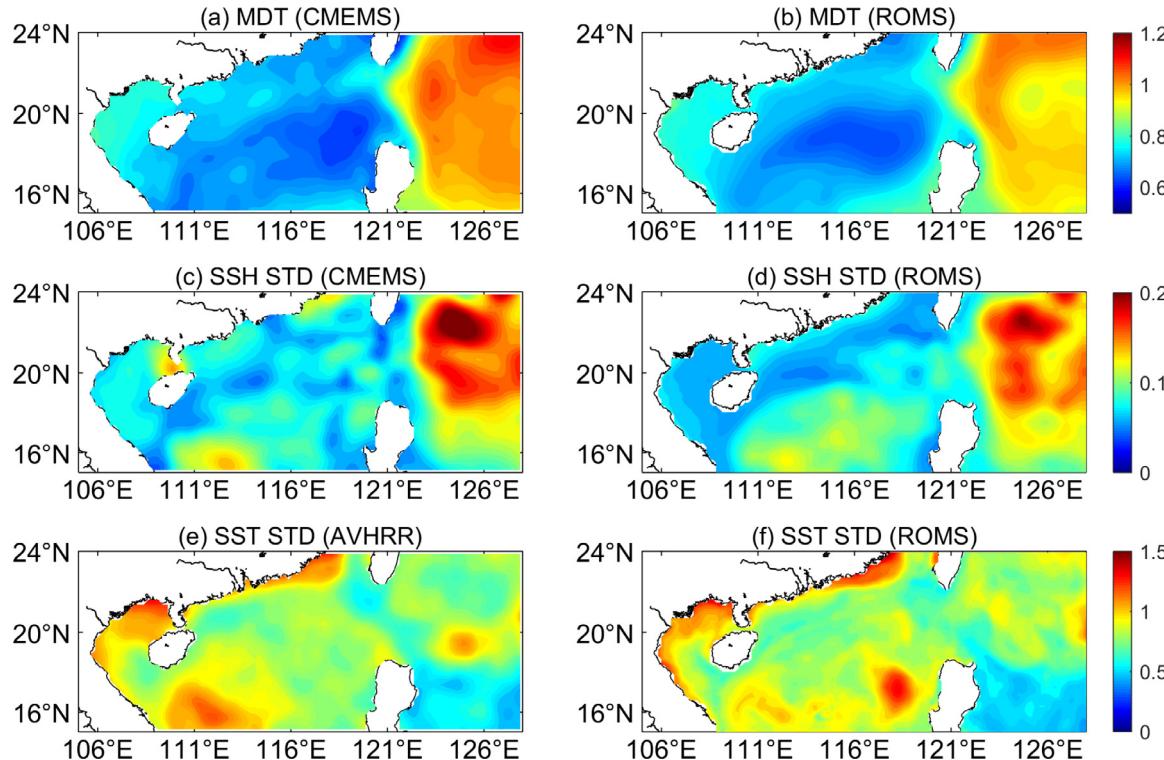


Fig. 3. Validation of the sea surface height (SSH) and sea surface temperature (SST) using a 9-year (2008–2016) model free run. (a) The mean dynamic topography obtained from CMEMS with an offset removed (unit: m); (b) the mean sea surface height obtained from ROMS free run without any data assimilation (unit: m); (c) the climatic monthly standard deviations of SSH for October obtained from CMEMS (unit: m); (d) the climatic monthly standard deviations of SSH for October obtained from the ROMS free run (unit: m); (e) the climatic monthly standard deviations of SST for October obtained from AVHRR (unit: °C); (f) the climatic monthly standard deviations of SST for October obtained from the ROMS free run (unit: °C).

The IEWVPS process on each time step for each particle is shown in Fig. 2. It should be noted that the adjustment of η^k and $\alpha^{1/2} P^{1/2} \xi^k$ could be in opposite direction, e.g., the t_3 and t_k in Fig. 2.

3. Experiment setups

3.1. ROMS configuration

The model domain spans from 105 °E to 128 °E and from 15 °N to 24 °N with a horizontal resolution of 1/6° and 24 vertical levels. This is an eddy-permitting resolution that can resolve both large scale circulation and mesoscale eddies. The surface layers in the model are weakly nonlinear, which fits the linear and Gaussian assumptions of the EnKF and variational methods. As mentioned by Chen et al. (2020b), this relatively low resolution is more feasible for the EnKF and variational methods but may be challenging for particle filters due to the high dimension of the system. Thus, we can use this resolution to test the performance of the IEWVPS in weakly nonlinear situations.

A model free run was initialized using the initial conditions obtained from HYCOM-NCODA (Hybrid Coordinate Ocean Model–Navy Coupled Ocean Data Assimilation; Metzger et al. (2014)) product on 1 Jan, 2007, and the model was integrated for 10 years (without any data assimilation), with atmospheric forcings obtained from ECMWF ERA-interim datasets (include windstress, heat flux and freshwater flux) and open boundary conditions obtained from the HYCOM-NCODA product. There are 4 purposes for this real simulation: (1) to validate the model before data assimilation (Fig. 3); (2) to derive statistics regarding the climatic background error standard deviation, which is important for 4D-Var (Fig. 3d and f for SSH and SST); (3) to provide a mean sea surface height (Fig. 3b) for satellite sea level anomaly (SLA) assimilations; and (4) to provide a dynamical balanced initial condition for assimilation experiments.

The climatic monthly mean values and standard deviations of the sea surface height (SSH) and sea surface temperature (SST) (Fig. 3) were used for the model validation (without any data assimilation). As shown in Fig. 3, ROMS successfully captured the large scale SSH

Table 1
Experimental design.

Experiment	Ensemble size	Observations
IEWVPS	40	SST+SLA
En4DVAR	40	SST+SLA
WCPSAS	1	SST+SLA

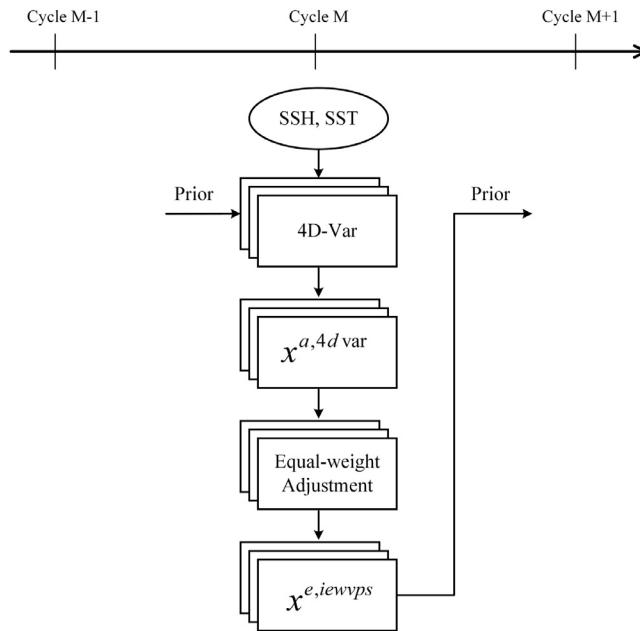


Fig. 4. The process of the continuous data assimilation using the IEWVPS. The overlapping boxes represent the ensemble members. The left panel is the assimilation cycle $M - 1$, the middle panel is the assimilation cycle M , and the right panel is the assimilation cycle $M + 1$. The prior at the current assimilation cycle M are from the states at the end of previous assimilation cycle $M - 1$.

pattern, as well as the variabilities in SSH and SST in the northern South China Sea (NSCS); this pattern and these variabilities are coincident with the satellite observations.

3.2. Data assimilation experiments

To test performances of the IEWVPS, satellite-derived SSH and SST observations are assimilated using the IEWVPS method, and two additional experiments were designed (Table 1). En4DVAR is an ensemble of perturbed weak-constraint 4D-Var similar to the EDA strategy of ECMWF (Bonavita et al., 2017). The WCPSAS experiment uses a single weak-constraint 4D-Var. En4DVAR and IEWVPS are ensemble methods, but WCPSAS is not. The difference between IEWVPS and En4DVAR is that an equal-weights adjustment step exists in the IEWVPS, making the whole system a practical particle filter data assimilation system. From a practical perspective, the model error cannot be neglected; thus, we do not use the strong-constraint 4D-Var (strong-constraint assumes that the model error is perfect and without error).

The SSH data were delayed time and gridded maps of sea level anomaly (MLSA) from Copernicus Marine Environment Monitoring Service (CMEMS). The SST data were gridded version 2 AVHRR-only products obtained from NOAA; these data were developed using optimum interpolation and included a large-scale adjustments of satellite biases with respect to the in situ data collected from ships and buoys (Reynolds et al., 2007). Both the SSHs and SSTs are Level 4 products and are available daily with a horizontal resolution of 1/4°. Although the observation errors in gridded products are never uncorrelated in the actual nature, we still assume that observation errors are uncorrelated in space and time for simplicity, which leads to the

fact that the observation error covariance, \mathbf{R} , is a diagonal matrix. The observation error is assumed to be a combination of the measurement error and the representation error, which are additive. Measurement errors are considered to be independent of the data source and have the following standard deviations (Moore et al., 2011a): 0.4 °C for SST and 2 cm for SSH. The representativeness error is measured by the standard deviation of the observations that contribute to each super observation. To diminish long-distance false dependencies, the decorrelation length scales of the covariance matrix were set to 50 km in the horizontal direction and 30 m in the vertical direction (Moore et al., 2011a).

Assimilation experiments started from 1 Oct, 2015 and lasted until 30 Nov, 2015, with a 1-day assimilation time window. In the IEWVPS and En4DVAR experiments, the initial and atmospheric forcing ensemble members were generated through an exact second-order sampling scheme (Hoteit et al., 2013; Chen et al., 2020b):

$$X_i = \bar{X} + \beta \sqrt{N} \mathbf{E} \epsilon_i^T \quad (27)$$

where \bar{X} represents the ensemble mean of the initial conditions (a 9-year averaged model state) or atmospheric forcings (windstress and heat flux in this study). \mathbf{E} is a matrix whose columns consist of $N - 1$ EOFs (empirical orthogonal functions). The term ϵ_i is the i th row of a $N \times (N - 1)$ random matrix with orthogonal columns, and the sum of columns is zero. β is an inflation parameter with a value of 1.0 for initial conditions, 0.2 for the windstress, and 100 for the heat fluxes. The ensemble spread of surface net heat flux generated with a factor 100 is approximately 10 W m^{-2} . In all assimilation experiments, the observations were cycled assimilated into the model (Fig. 4).

4. Results

The northern South China Sea (NSCS) is strongly influenced by monsoons and the Kuroshio; as a result, the large-scale circulation is cyclonic with active mesoscale processes. Moreover, the NSCS is also an area where typhoons generate or pass through frequently. During the period from 1 Oct to 10 Oct in 2015, Typhoon Mujigae passed through the NSCS, and the temperature of the upper layers was lowered by Ekman pumping. In this research, the upper layer states are investigated and compared among different assimilation experiments. The surface states at the end of the time window are used for the comparisons, while the error statistics of forecasts for November, 2015 are also compared.

4.1. Upper ocean states

The SLA simulations are shown in Fig. 5. For the duration of the typhoon, the SLA was high in the NSCS (especially along the coast) and low in the western Pacific. The mesoscale eddies in the three assimilation experiments are consistent with the observations. The SLA differences between the three experiments and satellite products are small (approximately 4 cm), as shown in Fig. 5. Generally, the three assimilation experiments performed similarly in modeling the SLAs.

When a typhoon crosses the sea, the Ekman pumping induced by the typhoon can cause dramatic changes in upper-level seawater temperatures, especially at the sea surface. This SST cooling is shown in Fig. 6 command exceeds 1 °C around the studied typhoon track. The cooling patterns output in the three experiments are similar during the cooling period (before 5 Oct). Large differences appeared on 5 Oct, and the SST cooling was strongest in the IEWVPS experiment and remained consistent with the AVHRR SSTs, while a warm bias existed in the En4DVAR experimental outputs. The WCPSAS performed similarly to the IEWVPS in modeling SST cooling; however, the SSTs tended to be overly high in the northeastern part of the modeled region (Fig. 6b). After 5 Oct, the SSTs began to recover, and large biases can be seen in the western Philippines in the WCPSAS experiment (Fig. 7b). The En4DVAR experiment performed better than the WCPSAS experiment,

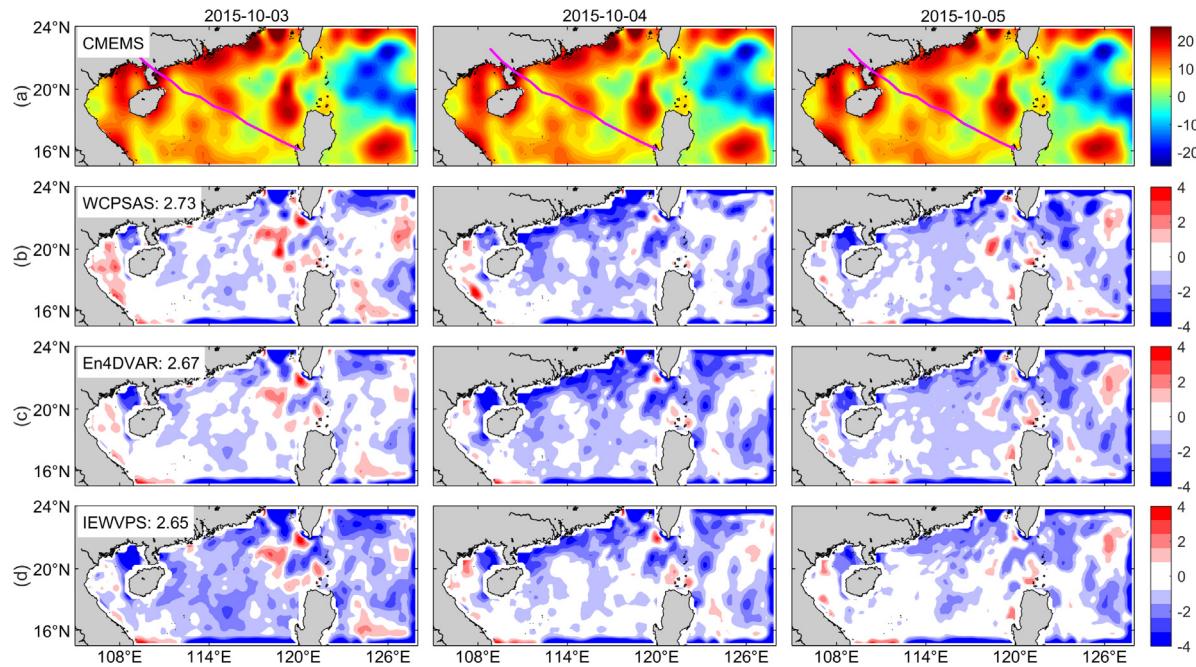


Fig. 5. Analysis SLAs at the end of the time window during 3 Oct and 5 Oct, 2015. Each column represents a different date. Panel (a) shows SLAs from CMEMS; panel (b) shows the differences in the SLAs between the WCPAS and CMEMS; panel (c) shows the difference in the SLAs between the En4DVAR and CMEMS; and panel (d) shows the difference in the SLAs between the IEWVPS and CMEMS. The solid line indicates the best track of Typhoon Mujigae. The numbers are mean RMSDs (unit: cm).

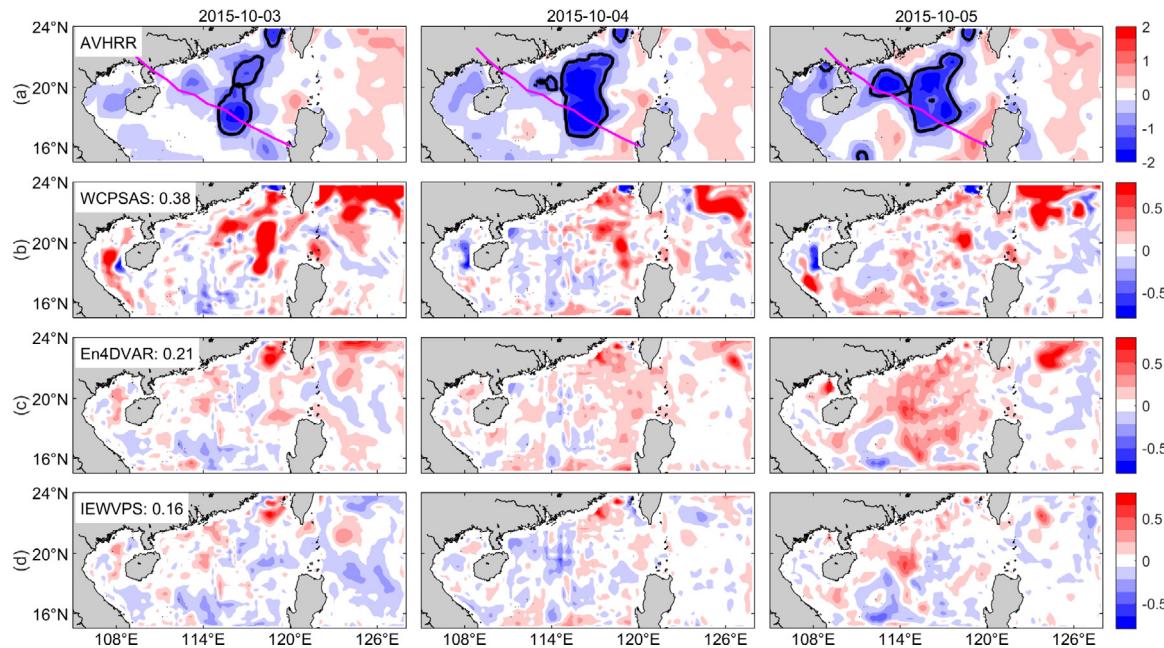


Fig. 6. Analysis SST cooling at the end of the time window relative to 1 Oct 2015. Each column represents a different date with panel (a) reflecting the AVHRR product; (b) reflecting the differences in SST cooling between the WCPAS outputs and AVHRR product; (c) showing the differences in SST cooling between the En4DVAR outputs and AVHRR product; (d) showing the differences in SST cooling between the IEWVPS outputs and AVHRR product. The black line represents the -1°C isotherm. The purple solid line shows the best track of Typhoon Mujigae between 2 Oct and 4 Oct 2015. The numbers are mean RMSDs (unit: $^{\circ}\text{C}$).

but the outputs were still overly high according to the AVHRR products (Fig. 7c).

The influences of satellite data are not limited to the surface and can propagate vertically through the cross-correlations of the background error covariance matrix and move forward through the model dynamics. The temperatures along Typhoon Mujigae track are shown in Fig. 8. Although the SST cooling reflected in the WCPAS outputs was similar to that output by the En4DVAR and IEWVPS before 5 Oct, the temperature changes output by the WCPAS in the upper 100

m were worse than those of the En4DVAR and IEWVPS. Compared with En4DVAR, the temperatures obtained from IEWVPS were cooler in upper 50 m and warmer below 100 m (Fig. 9), indicating that the Ekman pumping effect was more significant in the IEWVPS experiment than in the En4DVAR experiment. Compared with the observed SSTs, the temperatures near 119°E tended to be overly high in the WCPAS and En4DVAR outputs; these high temperatures also existed in the deterministic part of the IEWVPS and were improved after an equal-weights adjustment in the IEWVPS (Fig. 9). As a result, the

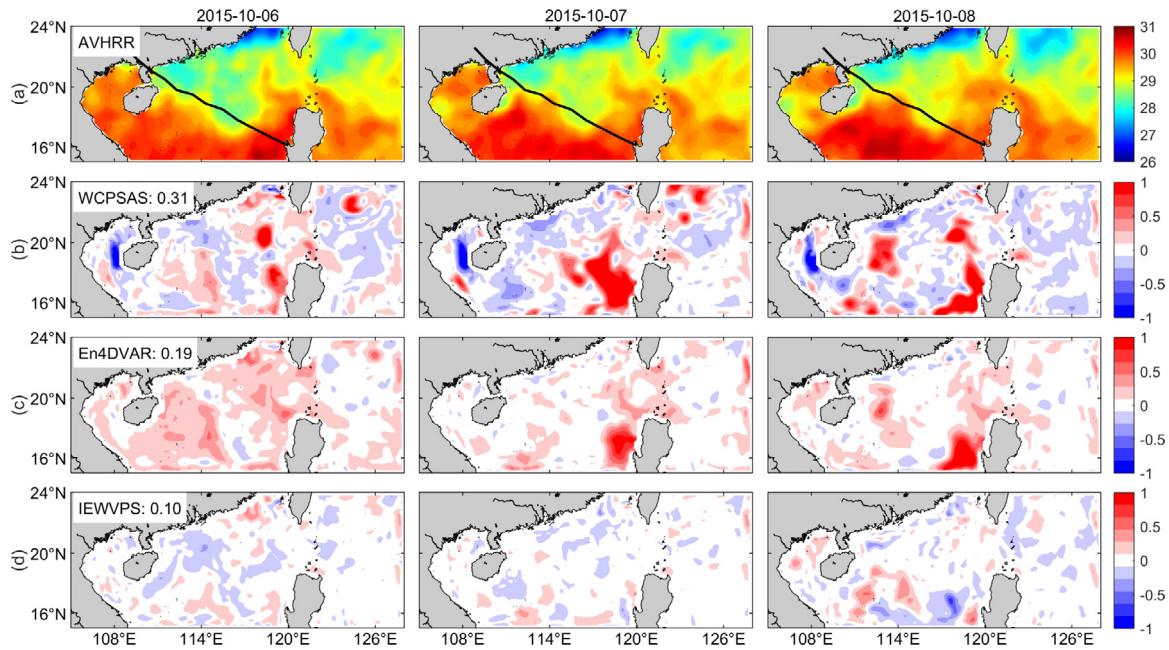


Fig. 7. Analysis SSTs at the end of the time window from 6 Oct to 8 Oct 2015. Each column represents a different date with row (a) representing the AVHRR products; (b) revealing the differences between the WCPAS outputs and AVHRR products; (c) showing the differences between the En4DVAR outputs and AVHRR products; (d) indicating the differences between the IEWVPS outputs and AVHRR products. The black line shows the best track of Typhoon Mujigae between 2 Oct and 4 Oct 2015. The numbers are mean RMSEs (unit: °C).

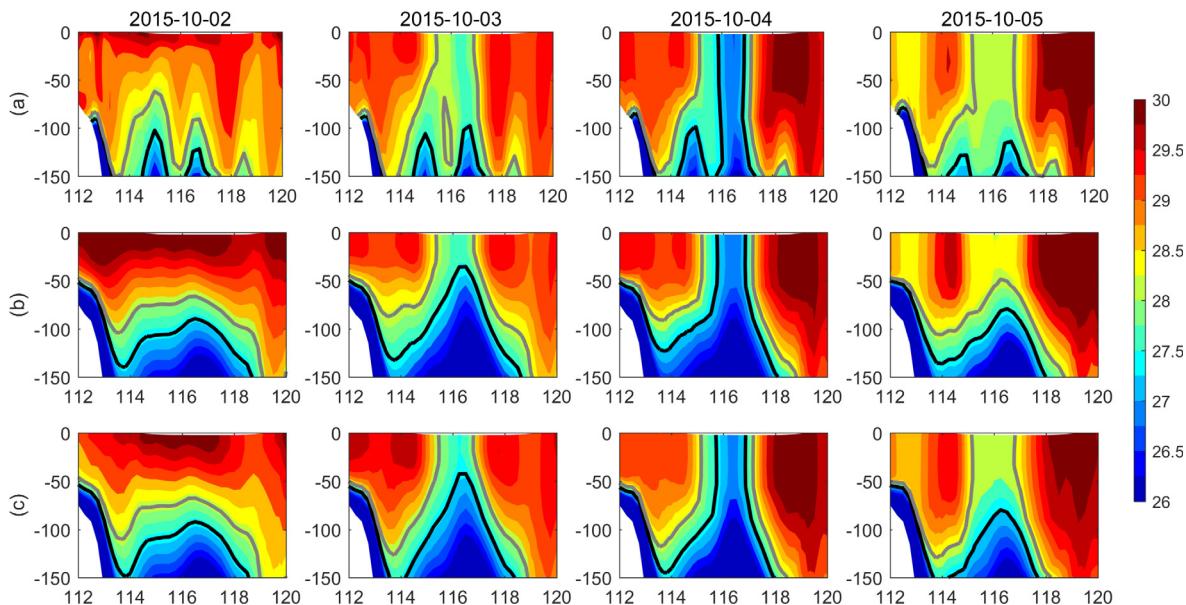


Fig. 8. Temperature cross-sections along the typhoon track (unit: °C). Each column represents a different date and each row represents a different method: (a) WCPAS; (b) En4DVAR; (c) IEWVPS. The black line is the 27.5 °C isotherm, and the gray line is the 28.3 °C isotherm.

temperatures near 119°E in the IEWVPS outputs were cooler than those output by the En4DVAR (Fig. 9). It should be noted that the large biases on 2 Oct are mainly caused by the initial ensemble generation scheme calculated through Eq. (27). When assimilating SSTs, the error tends to be largest at the thermocline (at a depth of approximately 100 m) (Luo et al., 2017); this is caused by the net heat input when correcting the cold bias in the mixed layer (Shu et al., 2009).

4.2. Error statistics

The above analyses are qualitative; we also quantitatively compared the three experiments. To make the comparisons more effective and

convincing, the assimilation period was extended to 2 months (from 1 Oct to 30 Nov 2015). However, only the results obtained in November 2015 were compared to exclude the influence of the initial ensemble generated using Eq. (27). The root mean square deviations (RMSEs) of the analysis states at the end of the assimilation window (that is t_n for a window $[0, n]$) and the 1-day forecast states were calculated, as shown in Fig. 10. The RMSEs of the SSHs output in the IEWVPS and En4DVAR experiment were equivalent and slightly smaller than those output in the WCPAS experiment (Fig. 10a and c). For the SSTs, the RMSE at the end of the assimilation window in the IEWVPS experiment was 0.1 °C (Fig. 10b), which was smallest among the three experiments and corresponded to a reduction of 52.6% (relative to the En4DVAR

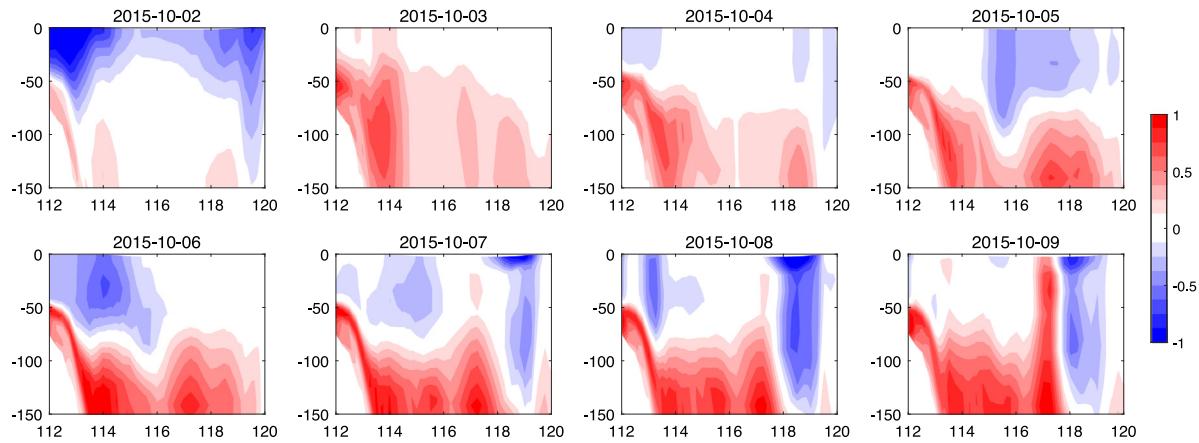


Fig. 9. Temperature differences between the IEWVPS and En4DVAR outputs along the track of Typhoon Mujigae from 2 Oct to 9 Oct 2015 (unit: °C).

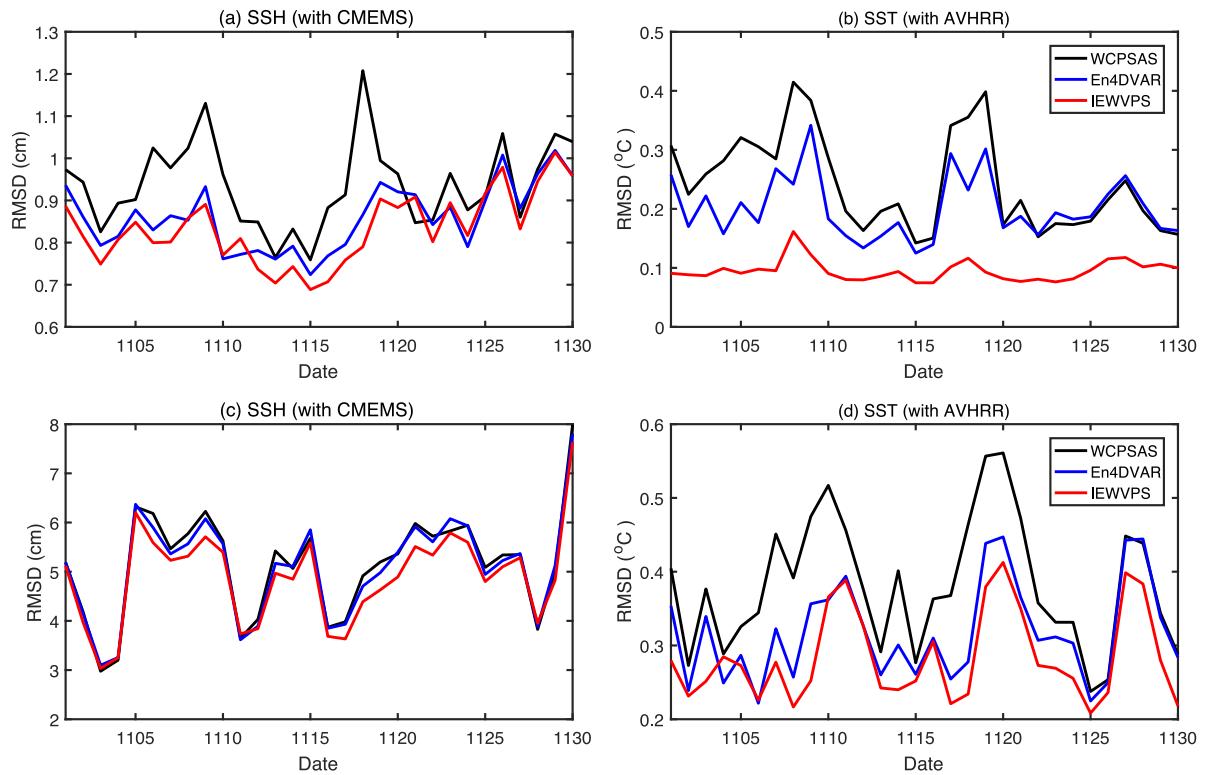


Fig. 10. RMSDs of the output SSHs (left panel) and SSTs (right panel). Subgraphs (a) and (b) show the analysis states at the end of the assimilation time window; subgraphs (c) and (d) show the 1-day forecast states.

RMSD). The SST RMSD of the 1-day forecast (Fig. 10d) was larger than that of the analysis states at the beginning of the assimilation window, as was expected. Compared with the En4DVAR experimental outputs, the SST RMSD of the 1-day forecast in the IEWVPS experiment was reduced by approximately 10.5%.

To compare the three experiments more clearly, the RMSD time series shown in Fig. 10 were used to draw box plots (Fig. 11). Consistent with Fig. 10, the SSH RMSDs in the three experiments are at the same level. Among the three methods, the SST RMSD of the analysis and 1-day forecast are the smallest in the outputs of the IEWVPS experiment.

To be more objective, we also compared the output SSTs with ESACCI (European Space Agency Sea Surface Temperature Climate Change Initiative; Merchant et al. (2019)) and UKMO SST (Good et al., 2020) products, as shown in Fig. 12. Since the horizontal resolutions of the ESACCI and UKMO SST products are $1/20^\circ$ and our model

resolution was $1/6^\circ$, we interpolated the ESACCI and UKMO SSTs to our model grids. The RMSDs relative to the ESACCI and UKMO SSTs were larger than those relative to the AVHRR product. Despite the differences in the SST products, the IEWVPS produced the smallest RMSD among the three assimilation experiments. At the end of the assimilation window, the SST RMSD of the IEWVPS was reduced by 12.1% (ESACCI) and 12.3% (UKMO) compared to the En4DVAR values (Fig. 12). For the 1-day forecast, the SST RMSD of the IEWVPS was reduced by 6.7% (ESACCI) and 6.2% (UKMO) compared to the En4DVAR values.

In general, the SSH RMSDs of the three assimilation experiments were similar and the SST RMSDs in the IEWVPS experiment were the smallest, especially for the analysis states at the end of the time window. During the 1-day forecast step, the RMSDs increased for both SSHs and SSTs. One factor affecting this result is that large biases appear at open boundaries, as shown in Figs. 5 to 7. The other factor

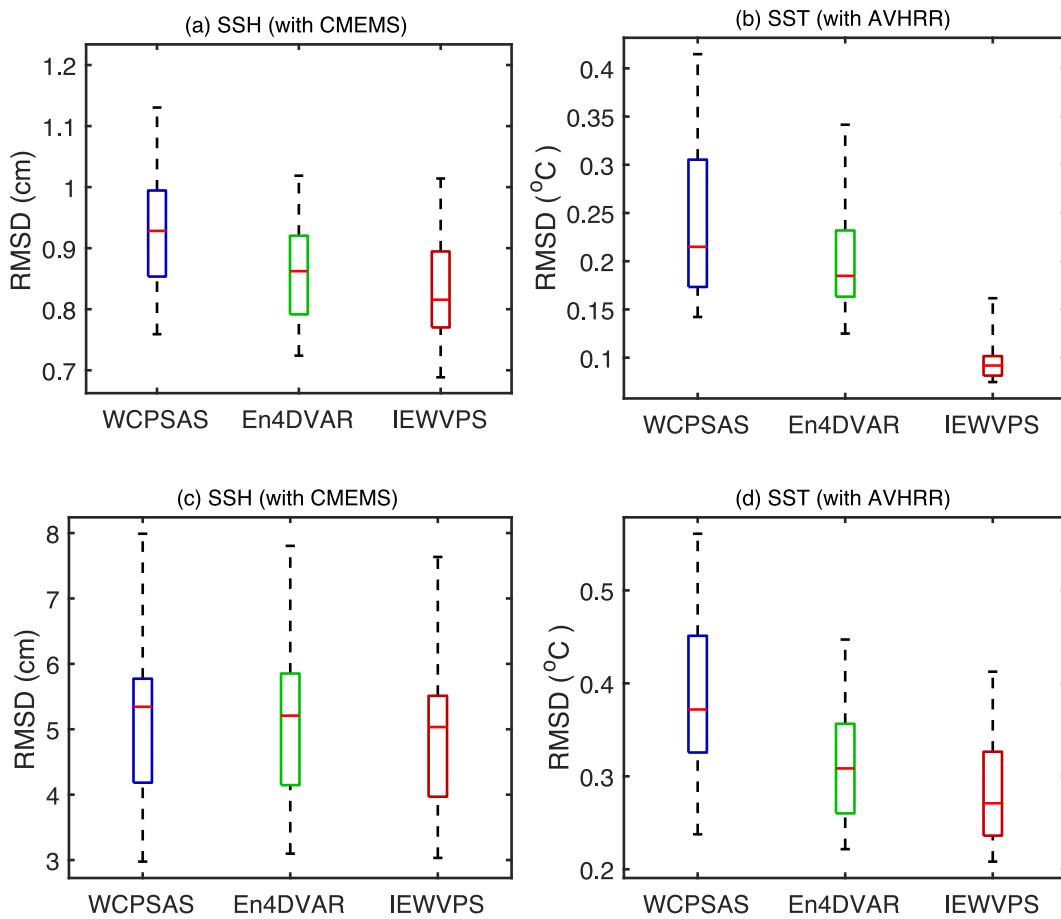


Fig. 11. Box plots of the RMSDs of the SSHs (left panel) and SSTs (right panel). Subgraphs (a) and (b) show the analysis states at the end of the assimilation time window; subgraphs (c) and (d) show the 1-day forecast states.

is that the model error is not considered during the forecast step. Given the same atmospheric forcing and open boundary conditions, the RMSDs of the forecasts are dependent on the initial conditions which are characterized by the analysis states at the end of the time window (Fig. 10c and d).

For the purpose of understanding how the equal-weights adjustment component works, we compared the biases of the SSTs; the bias is the mean value of the distances between observations and the model-simulated values, $(y - H(x))$, and is shown in Fig. 13. There are at least four sources accounting for the biases in our experiments: the initial conditions, perturbed atmospheric forcing, open boundary conditions and model errors. As shown in Fig. 13a, the bias in the initial conditions is very small (a mean value of approximately 0.002 °C), so the initial conditions can be excluded from the sources of biases. However, as the model is forced by perturbed atmospheric forcings, open boundary conditions and model errors, the bias becomes larger when the model propagates over time. Thus, the output SSTs were biased at the end of the time window in the En4DVAR experiment (with a mean value of 0.1 °C), and a similar pattern was found in the WCPSAS experiment. We can see that the bias in the IEWVPS experiment was very small even at the end of the time window (Fig. 13b), and the mean value was approximately 0.007 °C. The SST bias in the En4DVAR experiment is approximately 0.093 °C greater than that of the IEWVPS experiment. This bias deviation was close to the 0.1 °C deviation in the RMSDs calculated between the En4DVAR and IEWVPS experiments.

The rank histograms of the one-model run experiment conducted for November 2015 were also compared, as shown in Fig. 14. The rank histograms were generated by ranking the observations in the set of the state ensemble members from lowest to highest. A left-extreme pattern was found in the En4DVAR experiment (Fig. 14a), indicating

a positive bias (Hamill, 2001). In the IEWVPS experiment, slightly U-shaped rank histograms indicated that the ensemble spread was slightly smaller than the RMSD. Together with Fig. 13, we can conclude that the equal-weights adjustment components can reduce the bias at the end of the time window. We know that the full particle weight of each particle consists of the proposal weight and the likelihood weight. The implicit equal-weights scheme ensures that the full weights are equal for all particles, which introducing more freedom in the IEWVPS than the En4DVAR to fit the observations.

4.3. An additional experiment

The above analyses are based on Typhoon Mujigae, which occurred in 2015 and whose track was almost a straight line. In reality, some typhoons may turn to the north, exhibiting complex air-sea interactions, Typhoon Megi, which occurred in 2010, was one of these. One trial is not enough to show the advantages of the IEWVPS. Thus, we performed the same experiments as those described above for Typhoon Megi in 2010. The largest SST cooling occurred on 22 and 23 Oct, as shown in Fig. 15. The SST cooling in the IEWVPS experiment was closer to the AVHRR products than those in the En4DVAR experiment, especially the -2 °C isotherm. The RMSDs of the SSHs and SSTs are shown in Fig. 16. The RMSD of SSHs in the IEWVPS experiment was equivalent to that in the En4DVAR, but the RMSD of SST in the IEWVPS experiment was much smaller than that in the En4DVAR experiment. Since the ensemble spread is a measure of the uncertainty in the data assimilation method it is better to have a slightly too large spread than a too small spread. The ratios of the RMSDs to the ensemble spreads for the SSHs and SSTs output by the IEWVPS were slightly larger than 1.0, indicating that the ensemble is under-dispersed. Similar to the case of Typhoon

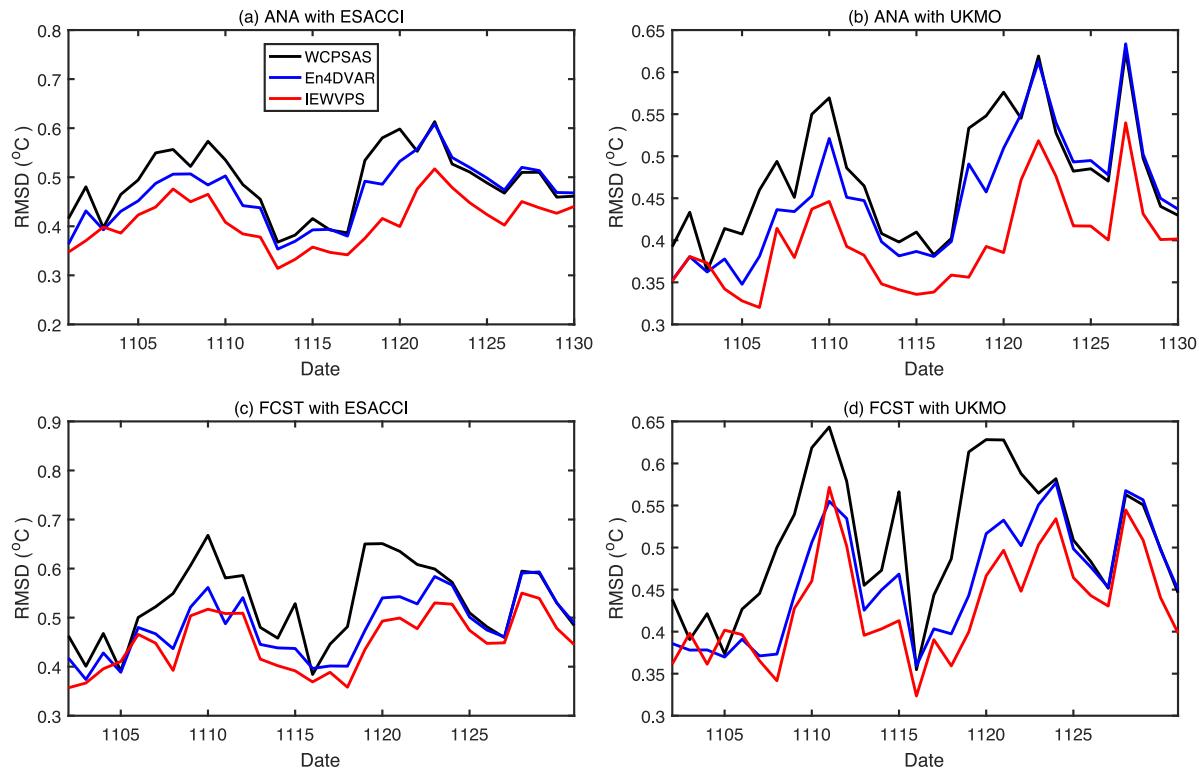


Fig. 12. SST RMSDs relative to the ESACCI (left panel) and UKMO (right panel) SST products. ESACCI: European Space Agency Sea Surface Temperature Climate Change Initiative (Merchant et al., 2019); UKMO: United Kingdom Meteorological Office (Good et al., 2020). Subgraphs (a) and (b) show the analysis states at the end of the assimilation time window; subgraphs (c) and (d) show the 1-day forecast states.

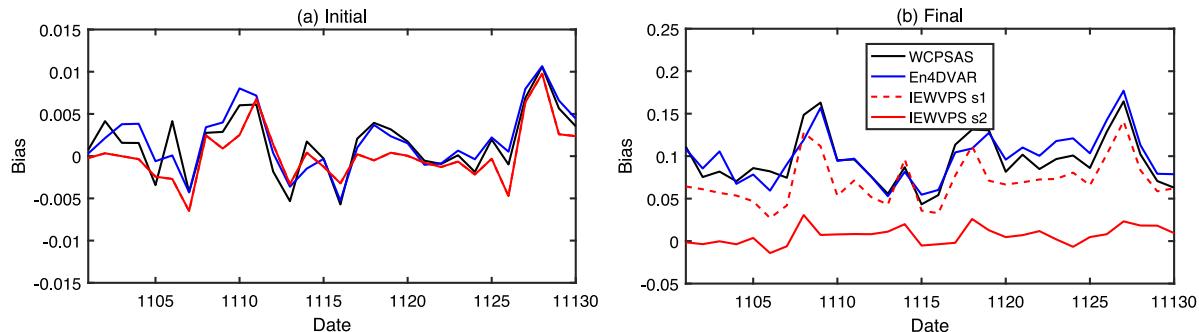


Fig. 13. SST biases in the observation space. (a) Biases of SSTs at the beginning of the time window; (b) Biases of SSTs at the end of the time window. IEWVPS s1 and s2 represent the ocean states before and after the equal-weights adjustment, respectively. Note that IEWVPS s1 is same as IEWVPS s2 in the panel (a).

Mujigae, a warm bias existed at open boundary and propagated to the interior ocean (Fig. 6b and Fig. 7b).

5. Discussions and conclusions

In this study, the implicit equal-weights variational particle smoother (IEWVPS) method was implemented in the Regional Ocean Modeling System (ROMS) and was applied to the assimilation of ocean satellite data. The key problem we solved in the implementation of the IEWVPS into ROMS is the estimation of the equal-weights stochastic adjustment component. We used an implicit random mapping method to estimate the perturbation component which was represented through the tangent linear model, and then combined this component with the equal-weights parameter α . The IEWVPS uses a modified weak-constraint 4D-Var framework as the proposal transition density in a standard implicit equal-weights particle filter (IEWPF). This method combines the merits of the weak-constraint 4D-Var and the IEWPF: it reduces the RMSD by applying the weak-constraint 4D-Var as the

proposal density, and prevents filter degeneracy by using an implicit equal-weights sampling scheme.

To verify the effects of the IEWVPS method, we tested the responses of the upper ocean variables during a typhoon case study before and after the assimilation of ocean satellite data. Satellite remote sensing SST and SSH observations in the northern South China Sea area were assimilated in the experiments, and the results of the IEWVPS were compared with those of the En4DVar method and ROMS weak-constraint 4D-Var (WCPAS). The SSH RMSDs were equivalent among three assimilation experiments, while the SST RMSDs in the IEWVPS experiment was the smallest for both the analysis (at the end of the time window) and 1-day forecasts. Because the RMSDs were reduced at the end of the time window, the SST RMSDs of the 1-day forecast derived from the IEWVPS experiment were reduced by 10.5% (AVHRR), 6.7% (ESACCI) and 6.2% (UKMO) relative to the results of En4DVAR experiment. Therefore, the temperature cooling induced by Ekman pumping in the upper 50 m was best characterized in the IEWVPS experiment. However, the largest improvement did not occur during typhoon activity over the

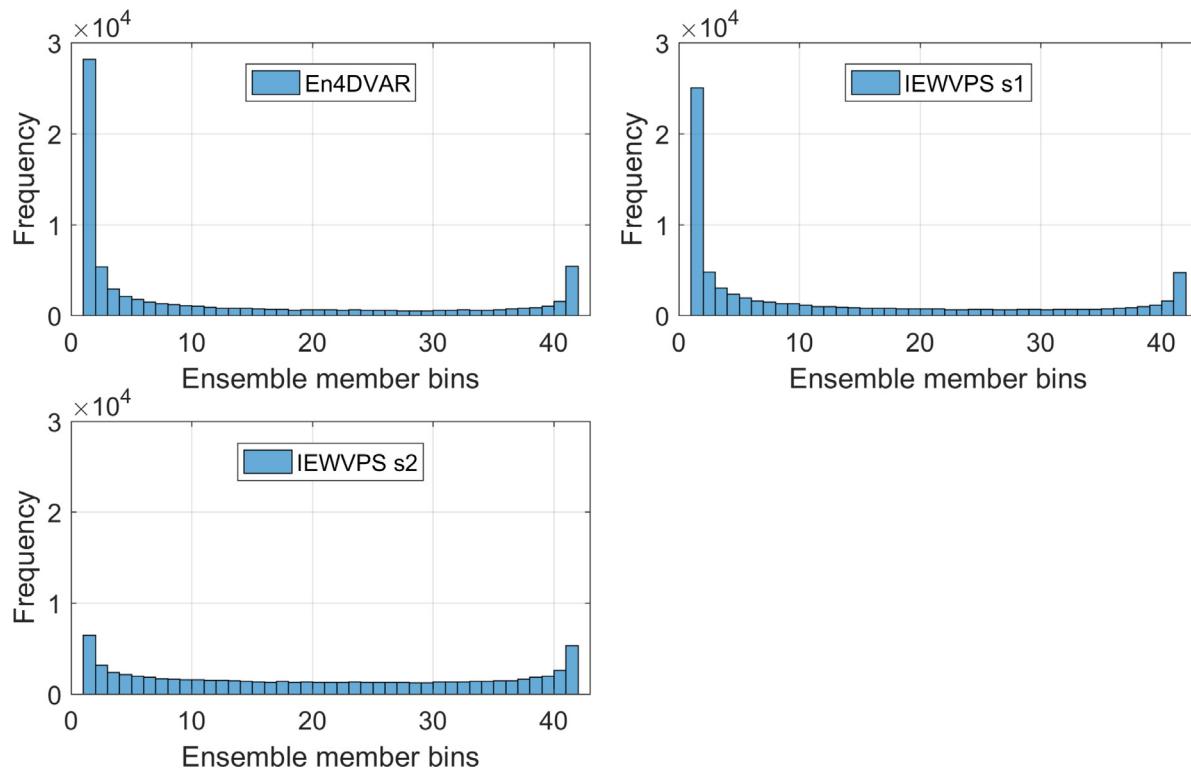


Fig. 14. The rank histograms of SSTs obtained from one-model run from 1 Nov to 30 Nov 2015. IEWVPS s1 and s2 represent the ocean states before and after the equal-weights adjustment, respectively.

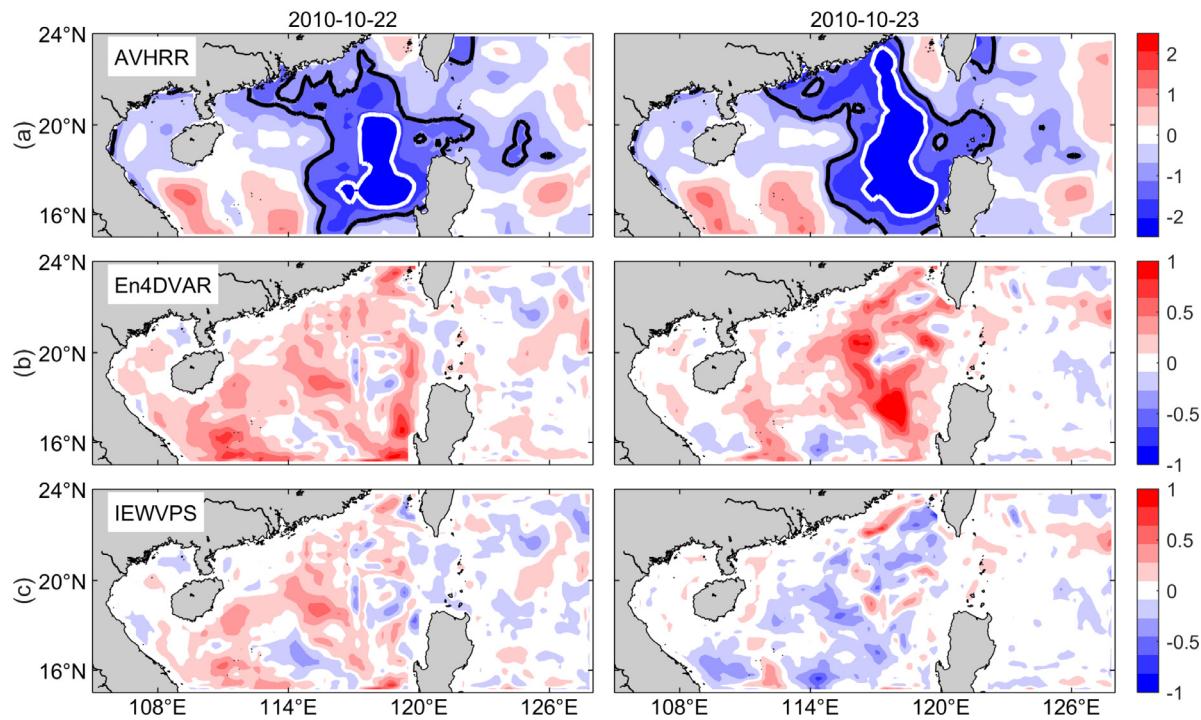


Fig. 15. SST cooling relative to that on 18 Oct, 2010 (unit: $^{\circ}\text{C}$). The black line is the $-1\text{ }^{\circ}\text{C}$ isotherm, and the white line is the $-2\text{ }^{\circ}\text{C}$ isotherm. (a) AVHRR SST product; (b) the SST cooling difference between the En4DVAR outputs and AVHRR product; (c) the SST cooling difference between the IEWVPS outputs and AVHRR product.

sea, but occurred after the typhoon made landfall over China. Excessive temperature arose at open boundaries in the En4DVAR and WCPAS experiments; this situation was relieved in the IEWVPS experiment. The bias existed at open boundaries is one of the reasons why the RMSDs were increased in the 1-day forecasts. Another reason is that the

model errors are considered during the assimilations but are neglected during the forecasts. To improve the forecasts, model error should be accounted for, such as by using the stochastically perturbed parameterization tendencies (SPPT) or stochastic kinetic energy backscatter (SKEB) schemes in the model as done in ECMWF (Bonavita, 2011).

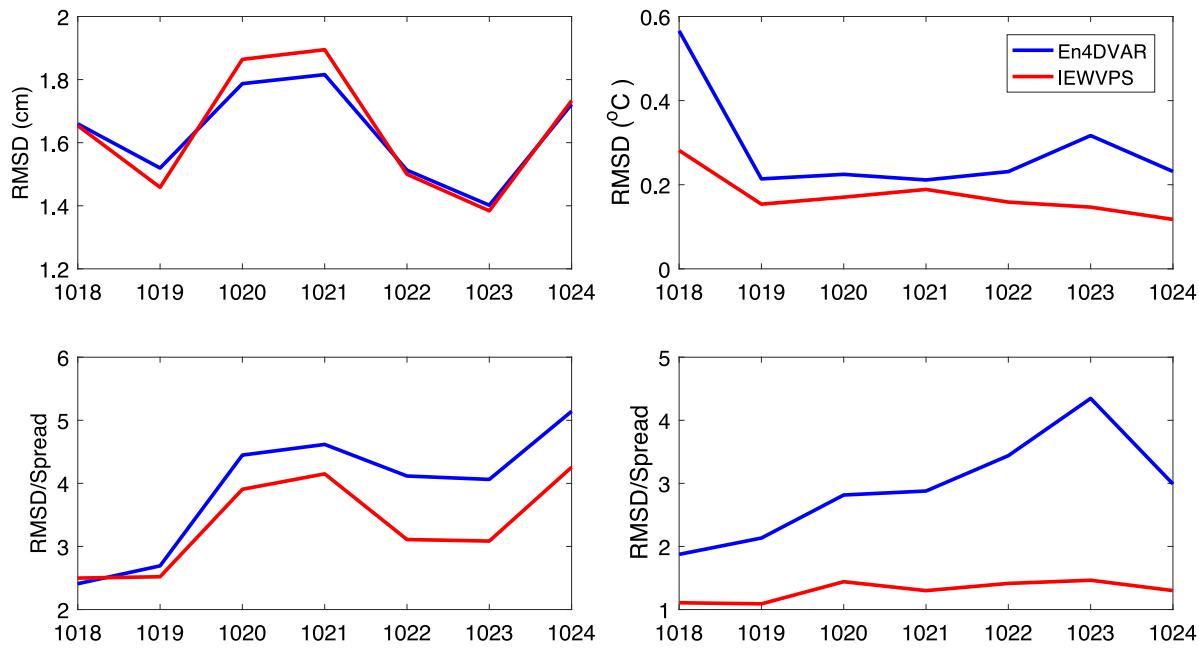


Fig. 16. Error statistics of SSH (left panel) and SST (right panel). Top panel: RMSD; Bottom panel: the ratio of RMSD to ensemble spread.

To increase the ensemble spread, an exact second-order sampling scheme (Hoteit et al., 2013; Chen et al., 2020b) was adopted for the atmospheric forcings. One drawback of this scheme is that the choice of inflation parameter is artificial. The inflation parameter was 0.2 for the windstress and 100 for the heat fluxes; these values are not optimal. The other drawback is that the independent choices of inflation parameters for different variables (windstress and heat fluxes in our study) may lead to inconsistent statistical results among variables. Another drawback of the perturbation scheme used in our study is that the covariance of atmospheric forcing is static and cannot reflect the high frequency variabilities in the atmosphere. To improve the ensemble quality, the perturbation scheme for atmospheric forcing should be studied further, and the uncertainty at open boundaries should also be considered. The ensemble spread is also influenced by the sampling ratio of α on the > 1 and < 1 branches (Zhu et al., 2016). However, the $\alpha > 1$ solution is spurious, which is leading to a systematic bias. This systematic bias can be eliminated by the two-stage IEWPF proposed by Skauvold et al. (2019). Then, the ensemble spread will be dependent on a different parameter β . We will replace the implicit equal-weights scheme according to Skauvold et al. (2019) and investigate an adaptive scheme for β to improve the ensemble quality.

The computational cost of the IEWVPS is slightly larger than that of the En4DVAR, since the IEWVPS is based on the En4DVAR (a set of weak-constraint 4D-Var). Additional costs for each particle are needed to calculate of the equal-weights adjustment component, including α_i and $\mathbf{P}^{1/2}\xi_i^{0:n}$. In total, an additional TL (tangent linear) model integration is needed for the calculation of the equal-weights adjustment component and a NL (nonlinear) model integration is needed to move each proposal particle to its equal-weight position for each particle; these computational costs are relatively small compared to the total cost of each 4D-Var (2 NL integrations, 30 TL integrations and 30 adjoint model integrations). As shown in Fig. 1, we need to calculate and deliver $\mathbf{D}^{1/2}\xi^{0:n}$. The assimilation window in our study is 1 day, which equals 120 model steps (one time step is 720 s); thus, we need to deliver $n = 120 \mathbf{D}^{1/2}\xi$ for each particle as outputs. If the assimilation window is very long or the time step is very small, this cost will become larger. One method used to reduce this cost involves delivering $\mathbf{D}^{1/2}\xi$ outputs every M model steps; which in our study, $M = 15$ (equivalent to 3 h in the simulation). The total computational cost of the IEWVPS scheme was divided into Fortran and MATLAB

components. The MATLAB components include the calculation of α_i and SVD on $\mathbf{R}^{-1/2}\mathbf{q}_0$. The computational cost of the Fortran component is 910.83 seconds/particle in one cycle and the cost of the MATLAB component is less than 60 s, while the computational cost of each 4D-Var is 852.18 s. Therefore, the additional computational cost is approximately 120 s for each particle, an increment of approximately 14% (relative to cost of the 4D-Var). The MATLAB components will be rewritten in Fortran language, thus further reducing the computational cost.

One advantage of the IEWVPS is that the equal-weights stochastic adjustment component maintains the model dynamic balance because of the balance operator in 4D-Var and the balanced theoretical implicit equal-weights sampling framework. We know that the major advantage of the EnKF is the flow-dependent background error covariance matrix. Therefore, one drawback in our study is that the background error covariance matrix (\mathbf{B}) is climatological and static. Furthermore, a flow dependent \mathbf{B} matrix based on the PF and 4D-Var method will be investigated in the future.

CRediT authorship contribution statement

Pinqiang Wang: Conceptualization, Methodology, Software, Writing - original draft, Writing- review and editing, Formal analysis, Investigation, Validation, Visualization, Data curation. **Mengbin Zhu:** Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review and editing, Validation, Supervision, Funding acquisition. **Yan Chen:** Conceptualization, Methodology, Formal analysis, Investigation, Funding acquisition. **Weimin Zhang:** Conceptualization, Methodology, Writing - review and editing, Project administration, Supervision, Funding acquisition. **Yi Yu:** Visualization, Formal analysis, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The weak-constraint 4D-Var codes in this study is downloaded from <http://www.myroms.org>. Satellite SSHs are downloaded from <http://marine.copernicus.eu>. Satellite SSTs are downloaded from <ftp://eclipse.ncdc.noaa.gov>.

This paper is supported by the National Key R&D Program of China (2018YFC1406202), the National Natural Science Foundation of China (NSFC, Grant Nos. 42005113, 41675097, 41830964, 42075149).

References

- Ades, M., van Leeuwen, P.J., 2013. An exploration of the equivalent weights particle filter. *Q. J. R. Meteorol. Soc.* 139 (672), 820–840.
- Ades, M., van Leeuwen, P.J., 2015. The effect of the equivalent-weights particle filter on dynamical balance in a primitive equation model. *Mon. Weather Rev.* 143 (2), 581–596.
- Atkins, E., Morzfeld, M., Chorin, A.J., 2013. Implicit particle methods and their connection with variational data assimilation. *Mon. Weather Rev.* 141 (6), 1786–1803.
- Bonavita, M., 2011. Impact and diagnosis of model error in the ECMWF ensemble of data assimilations. In: Proc. ECMWF Workshop on Representing Model Uncertainty and Error in Numerical Weather and Climate Prediction Models, pp. 303–318.
- Bonavita, M., Trémolié, Y., Holm, E.e.a., 2017. A strategy for data assimilation. *ECMWF Tech. Mem.* 2017 (800).
- Browne, P.A., 2016. A comparison of the equivalent weights particle filter and the local ensemble transform Kalman filter in application to the barotropic vorticity equation. *Tellus A* 68 (1), 30466.
- Browne, P.A., van Leeuwen, P.J., 2015. Twin experiments with the equivalent weights particle filter and hadcm3. *Q. J. R. Meteorol. Soc.* 141 (693), 3399–3414.
- Chen, Y., Zhang, W., Wang, P., 2020a. An application of the localized weighted ensemble Kalman filter for ocean data assimilation. *Q. J. R. Meteorol. Soc.*
- Chen, Y., Zhang, W., Zhu, M., 2020b. A localized weighted ensemble Kalman filter for high-dimensional systems. *Q. J. R. Meteorol. Soc.* 146 (726), 438–453.
- Chorin, A.J., Tu, X., 2009. Implicit sampling for particle filters. *Proc. Natl. Acad. Sci.* 106 (41), 17249–17254.
- Fisher, M., Courtier, P., 1995. Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *ECMWF Tech. Mem.* 1995 (220).
- Good, S., Fiedler, E., Mao, C., Martin, M.J., Maycock, A., Reid, R., Roberts-Jones, J., Searle, T., Waters, J., While, J., et al., 2020. The current configuration of the ostia system for operational production of foundation sea surface temperature and ice concentration analyses. *Remote Sens.* 12 (4), 720.
- Hamill, T.M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weather Rev.* 129 (3), 550–560.
- Hoteit, I., Hoar, T., Gopalakrishnan, G., Collins, N., Anderson, J., Cornuelle, B., Köhl, A., Heimbach, P., 2013. A mitgcm/dart ensemble analysis and prediction system with application to the gulf of mexico. *Dynam. Atmos. Oceans* 63, 1–23.
- van Leeuwen, P.J., 2010. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Q. J. R. Meteorol. Soc.* 136 (653), 1991–1999.
- van Leeuwen, P.J., Künsch, H.R., Nerger, L., Potthast, R., Reich, S., 2019. Particle filters for high-dimensional geoscience applications: A review. *Q. J. R. Meteorol. Soc.* 145 (723), 2335–2365, URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6774339/>.
- Luo, H., Zheng, F., Zhu, J., 2017. Evaluation of oceanic surface observation for reproducing the upper ocean structure in echam5/mpi-om. *J. Geophys. Res.: Oceans* 122 (12), 9695–9711.
- Merchant, C.J., Embury, O., Bulgin, C.E., Block, T., Corlett, G.K., Fiedler, E., Good, S.A., Mittaz, J., Rayner, N.A., Berry, D., et al., 2019. Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Sci. Data* 6 (1), 1–18.
- Metzger, E.J., Smedstad, O.M., Thoppil, P., Hurlburt, H., Cummings, J., Walcraft, A., Zambudio, L., Franklin, D., Posey, P., Phelps, M., Hogan, P., Bub, F., DeHaan, C., 2014. US Navy operational global ocean and arctic ice prediction systems. *Oceanography* 27 (3), 32–43.
- Moore, A.M., Arango, H.G., Broquet, G., Edwards, C., Veneziani, M., Powell, B., Foley, D., Doyle, J.D., Costa, D., Robinson, P., 2011a. The regional ocean modeling system (roms) 4-dimensional variational data assimilation systems: part II—performance and application to the california current system. *Prog. Oceanogr.* 91 (1), 50–73.
- Moore, A.M., Arango, H.G., Broquet, G., Powell, B.S., Weaver, A.T., Zavala-Garay, J., 2011b. The regional ocean modeling system (roms) 4-dimensional variational data assimilation systems: Part I—system overview and formulation. *Prog. Oceanogr.* 91 (1), 34–49.
- Morzfeld, M., Hodys, D., Poterjoy, J., 2018. Variational particle smoothers and their localization. *Q. J. R. Meteorol. Soc.* 144 (712), 806–825.
- Penny, S.G., Miyoshi, T., 2016. A local particle filter for high-dimensional geophysical systems. *Nonlinear Processes Geophys.* 23 (6), 391.
- Poterjoy, J., 2016. A localized particle filter for high-dimensional nonlinear systems. *Mon. Weather Rev.* 144 (1), 59–76.
- Poterjoy, J., Anderson, J.L., 2016. Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter. *Mon. Weather Rev.* 144 (5), 2007–2020.
- Poterjoy, J., Wicker, L., Buehner, M., 2019. Progress toward the application of a localized particle filter for numerical weather prediction. *Mon. Weather Rev.* 147 (4), 1107–1126.
- Potthast, R., Walter, A., Rhodin, A., 2019. A localized adaptive particle filter within an operational nwp framework. *Mon. Weather Rev.* 147 (1), 345–362.
- Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., Schlax, M.G., 2007. Daily high-resolution-blended analyses for sea surface temperature. *J. Clim.* 20 (22), 5473–5496.
- Robert, S., Leuenberger, D., Künsch, H.R., 2018. A local ensemble transform Kalman particle filter for convective-scale data assimilation. *Q. J. R. Meteorol. Soc.* 144 (713), 1279–1296.
- Shu, Y., Zhu, J., Wang, D., Yan, C., Xiao, X., 2009. Performance of four sea surface temperature assimilation schemes in the south China sea. *Cont. Shelf Res.* 29 (11–12), 1489–1501.
- Skaauvold, J., Eidsvik, J., van Leeuwen, P.J., Amezcua, J., 2019. A revised implicit equal-weights particle filter. *Q. J. R. Meteorol. Soc.* 145 (721), 1490–1502.
- Wang, P., Zhu, M., Chen, Y., et al., 2020. Implicit equal-weights variational particle smoother. *Atmosphere* 11 (4), 338.
- Zhu, M., van Leeuwen, P.J., Amezcua, J., 2016. Implicit equal-weights particle filter. *Q. J. R. Meteorol. Soc.* 142 (698), 1904–1919.