

An application of the A-4DEnVar to coupled parameter optimization

Yantian Gong¹, Kangzhuang Liang¹, Xinrong Wu², Qi Shao¹, Wei Li^{1,3*}, Siyuan Liu¹, Guijun Han^{1*}, Hanyu Liu¹

¹School of Marine Science and Technology, Tianjin University, Tianjin 300072, China

²Key Laboratory of Marine Environmental Information Technology, National Marine Data and Information Service, Ministry of Natural Resources, Tianjin 300171, China

³Tianjin Key Laboratory for Oceanic Meteorology, Tianjin 300074, China

Received 21 July 2021; accepted 20 December 2021

© Chinese Society for Oceanography and Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

In variational methods, coupled parameter optimization (CPO) often needs a long minimization time window (MTW) to fully incorporate observational information, but the optimal MTW somehow depends on the model nonlinearity. The analytical four-dimensional ensemble-variational (A-4DEnVar) considers model nonlinearity well and avoids adjoint model. It can theoretically be applied to CPO. To verify the feasibility and the ability of the A-4DEnVar in CPO, “twin” experiments based on A-4DEnVar CPO are conducted for the first time with the comparison of four-dimensional variational (4D-Var). Two algorithms use the same background error covariance matrix and optimization algorithm to control variates. The experiments are based on a simple coupled ocean-atmosphere model, in which the atmospheric part is the highly nonlinear Lorenz-63 model, and the oceanic part is a slab ocean model. The results show that both A-4DEnVar and 4D-Var can effectively reduce the error of state variables through CPO. Besides, two methods produce almost the same results in most cases when the MTW is less than 560 time steps. The results are similar when the MTW is larger than 560 time steps and less than 880 time steps. The largest MTW of 4D-Var and A-4DEnVar are 1 200 time steps. Moreover, A-4DEnVar is not sensitive to ensemble size when the MTW is less than 720 time steps. A-4DEnVar obtains satisfactory results in the case of highly nonlinear model and long MTW, suggesting that it has the potential to be widely applied to realistic CPO.

Key words: 4D-Var, A-4DEnVar, coupled parameter optimization, “twin” experiments

Citation: Gong Yantian, Liang Kangzhuang, Wu Xinrong, Shao Qi, Li Wei, Liu Siyuan, Han Guijun, Liu Hanyu. 2022. An application of the A-4DEnVar to coupled parameter optimization. *Acta Oceanologica Sinica*, 41(9): 60–70, doi: 10.1007/s13131-022-1997-1

1 Introduction

Coupled model plays an increasingly important role in numerical weather prediction (NWP) and climate prediction. Two key factors affecting the forecast skill of NWP are the error of the initial condition and the coupled model bias. Generally, initial condition error can be reduced by assimilating observations to model states using the weakly or strongly coupled data assimilation (CDA). In contrast, model bias contains multiple sources, including imperfect dynamical core, uncertainty of physical parameterization and the involving empirical parameters (Zhang et al., 2012). Among them, the empirical parameters can be optimized by coupled parameter optimization (CPO) with data assimilation technique. For coupled models, CPO can effectively reduce model biases (Han et al., 2014). However, CPO is a challenging task due to the different characteristic time scales from different medium state variables and the existence of sampling errors in some ensemble-based data assimilation methods (Han et al., 2013; Zhang et al., 2020). Thus, appropriate data assimilation methods are required in CPO.

Currently, the main advanced data assimilation methods are four-dimensional variational (4D-Var) (Lewis and Derber, 1985; Le Dimet and Talagrand, 1986; Courtier et al., 1994), ensemble

Kalman filter (EnKF) (Evensen, 1994) and particle filter (PF) (Doucet et al., 2001). Previous studies about CPO are mainly based on 4D-Var (Lu and Hsieh, 1998; Du et al., 2009; Ito et al., 2010) and EnKF (Annan et al., 2005; Zhang et al., 2012; Han et al., 2013), while PF CPO is still under development. Here, we are concerned with 4D-Var. Compared with EnKF, 4D-Var can obtain the global optimal analysis value in the minimization time window (MTW) and assimilate unconventional data through nonlinear observation operator (Lindskog et al., 2004). Moreover, 4D-Var has the ability to consider leading and lagging interactions, which may exist in coupled model and influence CPO. However, 4D-Var also has some disadvantages, for example, the static background error covariance matrix (the \mathbf{B} matrix in brief) and the requirement of adjoint model that is difficult to generate.

To ease the above two problems of 4D-Var, the ensemble idea is introduced. First, the use of ensemble makes it possible for 4D-Var to construct a flow-dependent \mathbf{B} matrix, which has been demonstrated to be better than the climatological \mathbf{B} matrix (Buehner et al., 2010a, b; Fairbairn et al., 2014). Second, ensemble members can also be used to compute the gradient and therefore avoid the adjoint model (e.g. the four-dimensional ensemble-variational, 4DEnVar; Liu et al., 2008; Tian et al., 2008).

Foundation item: The National Key Research and Development Program under contract No. 2021YFC3101501; the National Natural Science Foundation of China under contract No. 41876014.

*Corresponding author, E-mail: liwe1978@tju.edu.cn; guijun_han@tju.edu.cn

Compared to another type of ensemble four-dimensional variational method (En4DVar; Lorenc, 2003; Zupanski, 2005) that uses the adjoint model to compute the gradient, the accuracy of 4DEnVar is lower. If the accuracy loss is small enough to be ignored, the 4DEnVar might be very attractive, especially for complex coupled models like coupled general circulation models (CGCM).

However, model nonlinearity is a tricky problem and always influences the accuracy of 4DEnVar when it is applied to a coupled model. On the one hand, the linear correlation between state perturbation and observation perturbation assumed by 4DEnVar causes large errors for highly nonlinear models (Tian and Feng, 2015). On the other hand, the optimization of low-frequency variable with large time scales needs long MTW that might introduce impact of the model nonlinearity. To partly address these issues, Liang et al. (2021) proposed the Analytical 4DEnVar (A-4DEnVar) that has the ability to adapt to model nonlinearity well. Unlike previous 4DEnVar methods which imitate EnKF to generate the ensemble, the A-4DEnVar adds a very small perturbation to the initial ensemble and updates the ensemble after every iteration, which effectively reduces the impact of model nonlinearity. However, seldom research about A-4DEnVar is targeted at application. The practicability and the potential of A-4DEnVar remains to be studied. Hence, considering the potential of A-4DEnVar application in CDA, we apply the A-4DEnVar to CPO for the first time in this study. Based on a coupled model with strong nonlinearity, we set up an observation system simulation experiment (OSSE) to investigate the performance of A-4DEnVar.

The paper is organized as follows. Section 2 introduce the methodologies of 4D-Var and A-4DEnVar. In Section 3 we introduce the setup of OSSE and variational algorithm. Then, we compare the results of 4D-Var and A-4DEnVar in Section 4, and the result of A-4DEnVar under different situations is also shown in this section. Finally, we summarize the paper in Section 5.

2 Methodology

2.1 4D-Var

4D-Var uses dynamic equations as constraints to build the cost function. It searches the minimum value of the cost function with respect to the control variable to obtain the optimal analysis value. The control variable is the initial model state when only the initial condition is optimized. The parameter can be regarded as the special state variable that does not change with time and has no observation when the initial condition and the parameter need to be optimized.

The general form of the cost function is

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}_c^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2} \sum_{i=P}^{i=0} [\mathbf{H}_i \mathbf{M}_{0 \rightarrow i}(\mathbf{x}_0) - \mathbf{y}_i]^T \mathbf{R}_i^{-1} [\mathbf{H}_i \mathbf{M}_{0 \rightarrow i}(\mathbf{x}_0) - \mathbf{y}_i], \quad (1)$$

where \mathbf{x}_0 represents the state variable (including the parameter which has been regarded as the special state variable, similarly hereinafter) at time t_0 ; \mathbf{x}_b represents the background value of state variable at time t_0 ; \mathbf{B}_c^{-1} denotes the inverse of climatological \mathbf{B} matrix; \mathbf{R}_i denotes the observation error covariance matrix at time t_i ; \mathbf{H}_i denotes the observational operator at time t_i ; \mathbf{M} is the nonlinear evolution operator; \mathbf{y}_i is the observation at time t_i . The gradient of the cost function with respect to \mathbf{x}_0 is

$$\nabla J(\mathbf{x}_0) = \mathbf{B}_c^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \sum_{i=P}^{i=0} \mathbf{M}_{0 \rightarrow i}^T \mathbf{H}_i^T \mathbf{R}_i^{-1} [\mathbf{H}_i \mathbf{M}_{0 \rightarrow i}(\mathbf{x}_0) - \mathbf{y}_i], \quad (2)$$

where \mathbf{M} is the tangent linear model and \mathbf{M}^T is the adjoint model. After the gradient is obtained, an appropriate optimization algorithm is selected for iterative search. For example, the formula of gradient descent method is

$$\mathbf{x}_{a,j+1} = \mathbf{x}_{a,j} - \alpha_1 \nabla J(\mathbf{x}_{a,j}), \quad (3)$$

where $\mathbf{x}_{a,j}$ is the analysis value of the state variable at time t_0 in the j th iteration; α_1 is a relaxation factor, and the optimal analysis value can be obtained after iteration convergence.

2.2 A-4DEnVar

The main idea of A-4DEnVar is to use the ensemble to accurately estimate the adjoint matrix \mathbf{M}^T , which is the transpose of the Jacobian matrix of the \mathbf{M} with respect to \mathbf{x}_0 . We denote the integral of the state variable from time t_0 to time t_i with a nonlinear model as

$$\mathbf{x}_i = \mathbf{M}_{0 \rightarrow i}(\mathbf{x}_0). \quad (4)$$

We set the reference state of the state variable at time t_0 as \mathbf{x}_0^* and add a perturbation $\tilde{\mathbf{x}}_0$ to \mathbf{x}_0^* ,

$$\mathbf{x}_i^* + \tilde{\mathbf{x}}_i = \mathbf{M}_{0 \rightarrow i}(\mathbf{x}_0^* + \tilde{\mathbf{x}}_0). \quad (5)$$

The first-order Taylor expansion expression of Eq. (5) at \mathbf{x}_0^* is

$$\mathbf{x}_i^* + \tilde{\mathbf{x}}_i = \mathbf{M}_{0 \rightarrow i}(\mathbf{x}_0^*) + \mathbf{M}_{0 \rightarrow i} \tilde{\mathbf{x}}_0 + o(\tilde{\mathbf{x}}_0). \quad (6)$$

With the neglect of $o(\tilde{\mathbf{x}}_0)$, the perturbation at time t_i due to the perturbation at time t_0 can be expressed as

$$\tilde{\mathbf{x}}_i \approx \mathbf{M}_{0 \rightarrow i} \tilde{\mathbf{x}}_0. \quad (7)$$

The perturbation ensemble $\tilde{\mathbf{X}}_0$ is generated for the initial condition

$$\tilde{\mathbf{X}}_0 = (\tilde{\mathbf{x}}_0^1, \tilde{\mathbf{x}}_0^2, \dots, \tilde{\mathbf{x}}_0^N), \quad (8)$$

where N is the ensemble size. According to Eq. (7), the perturbation ensemble $\tilde{\mathbf{X}}_i$ at time t_i is

$$\tilde{\mathbf{X}}_i \approx \mathbf{M}_{0 \rightarrow i} \tilde{\mathbf{X}}_0. \quad (9)$$

Notice that the perturbation ensemble $\tilde{\mathbf{X}}_0$ must be small enough to ensure that the high order infinitesimals in Eq. (6) can be ignored, which will reduce the error from Eq. (9).

A temporal cross error covariance matrix is defined as

$$\mathbf{B}_{ij} = \frac{1}{N-1} \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_j^T. \quad (10)$$

Then, Eq. (9) can be expressed as

$$\mathbf{B}_{i0} \approx \mathbf{M}_{0 \rightarrow i} \mathbf{B}_{00}, \quad (11)$$

while the adjoint matrix $M_{0 \rightarrow i}^T$ satisfies

$$B_{0i} \approx B_{00} M_{0 \rightarrow i}^T. \quad (12)$$

In EnKF or previous 4DEnVar, B_{00} is a flow-dependent B matrix. In A-4DEnVar, B_{00} is computed by perturbation ensemble \tilde{X}_0 , which is too small to be the B matrix. To consider the accuracy of the adjoint model estimation and the proportion of the background term in the cost function, a deflation factor μ is introduced to reduce the initial perturbation by $\sqrt{\mu}$ times. Meanwhile, the flow-dependent B matrix is approximated as

$$B_e = \frac{B_{00}}{\mu}, \quad (13)$$

where B_e represents the flow-dependent B matrix and has the same order of magnitude as B_c . Notice that it is not necessary for A-4DEnVar to use flow-dependent B matrix, static B matrix or hybrid B matrix can also be used for some special reasons. For example, the “twin” experiments in Section 3 and Section 4 use the same static B matrix as 4D-Var for a more convincing comparison. In the following formulas, we continue to use the flow-dependent B matrix.

The increment \tilde{x}_0 is regarded as the control variable and B_e is introduced. According to Eq. (4) and Eq. (5), the cost function can be rewritten as

$$J(\tilde{x}_0) = \frac{1}{2} (x_0^* + \tilde{x}_0 - x_b)^T B_e^{-1} (x_0^* + \tilde{x}_0 - x_b) + \frac{1}{2} \sum_{i=p}^{i=0} [H_i M_{0 \rightarrow i} (x_0^*) + H_i M_{0 \rightarrow i} \tilde{x}_0 - y_i]^T \times R_i^{-1} [H_i M_{0 \rightarrow i} (x_0^*) + H_i M_{0 \rightarrow i} \tilde{x}_0 - y_i]. \quad (14)$$

The gradient of the cost function with respect to \tilde{x}_0 is

$$\nabla J(\tilde{x}_0) = B_e^{-1} (x_0^* + \tilde{x}_0 - x_b) + \sum_{i=p}^{i=0} M_{0 \rightarrow i}^T H_i^T R_i^{-1} \times [H_i M_{0 \rightarrow i} (x_0^*) + H_i M_{0 \rightarrow i} \tilde{x}_0 - y_i]. \quad (15)$$

Let the gradient of Eq. (15) be 0. Substituting Eqs (10)–(13), the analytical solution form of the optimal analysis increment can be obtained

$$\tilde{x}_0 = \mu B_{00} \left(\mu B_{00} + \sum_{i=p}^{i=0} B_{0i} H_i^T R_i^{-1} H_i B_{i0} \right)^{-1} (x_b - x_0^*) + B_{00} \left(\mu B_{00} + \sum_{i=p}^{i=0} B_{0i} H_i^T R_i^{-1} H_i B_{i0} \right)^{-1} \times \left[\sum_{i=p}^{i=0} B_{0i} H_i^T R_i^{-1} (y_i - H_i x_0^*) \right]. \quad (16)$$

A relaxation factor α_2 is introduced in the iterative search process imitating the traditional 4D-Var, and the iterative formula is

$$x_{a,j+1} = x_{a,j} + \alpha_2 \tilde{x}_{0,j}, \quad (17)$$

where $\tilde{x}_{0,j}$ is the optimal analysis increment obtained in the j th iteration. Notice that the reference state x_0^* needs to be adjusted to

$x_{a,j}$ in the j th iteration.

A-4DEnVar makes efforts to overcome the model nonlinearity by two measures. On the one hand, the reference state x_0^* is adjusted to $x_{a,j}$ in the j th iteration, which means that the ensemble needs to be updated after every iteration. It is beneficial for A-4DEnVar to track the model nonlinearity. On the other hand, the assumption of linear correlation between state perturbation and observation perturbation (Tian and Feng, 2015) still exists in A-4DEnVar, which is similar to the approximation after ignoring the high order infinitesimals in Eq. (9). A-4DEnVar generates a small initial perturbation to reduce the errors caused by the assumption effectively. Due to the two measures mentioned above, A-4DEnVar is approximately equivalent to 4D-Var in adapting nonlinearity. Moreover, the feasible MTW will extend because of the ability to adapt to nonlinearity, which means more observations of the low-frequency variable can be contained in the MTW in CPO.

3 Experiment setup

To investigate whether A-4DEnVar can be applied to CPO in highly nonlinear coupled models and attain approximate equivalent precision to 4D-Var, “twin” experiments are established.

3.1 Simple coupled ocean-atmosphere model

We use a simple coupled ocean-atmosphere model, in which the atmospheric part is a Lorenz three-variable chaotic model (Lorenz, 1963), and the oceanic part is a slab ocean model (Zhang, 2011a, b; Zhang et al., 2012). The model equation is

$$\begin{aligned} \frac{dx_1}{dt} &= -\sigma x_1 + \sigma x_2, \\ \frac{dx_2}{dt} &= (1+c_1\omega)\kappa x_1 - x_2 - x_1 x_3, \\ \frac{dx_3}{dt} &= x_1 x_2 - b x_3, \\ O_m \frac{d\omega}{dt} &= c_2 x_2 - O_d \omega + S_m + S_s \cos\left(\frac{2\pi t}{S_{pd}}\right), \end{aligned} \quad (18)$$

where x_1 , x_2 and x_3 are three high-frequency atmospheric variables, ω is a low-frequency ocean variable. σ , κ and b are three parameters for the atmospheric part, with the standard values set to 9.95, 29, and 8/3. O_m and O_d are two ocean parameters used to define time scales. The standard values are set to 10 and 1, indicating that the time scale of the ocean is 10 times that of the atmosphere. S_m , S_s and S_{pd} are three forced parameters of ocean, which make the ocean variables supplemented when they are damped. S_m represents constant forcing with a standard value of 10, S_s defines the amplitude of seasonal forcing with a standard value of 1, and S_{pd} defines the frequency of seasonal forcing with a standard value of 10, so that the force is identical to the time scale of ocean variables. c_1 and c_2 are two coupled parameters, with standard values of 0.1 and 1, to ensure the chaotic properties of atmospheric variables and the stability of the coupled system. A total of 10 parameters of standard values set to $(\sigma, \kappa, b, c_1, c_2, O_m, O_d, S_m, S_s, S_{pd}) = (9.95, 29, 8/3, 0.1, 1, 10, 1, 10, 1, 10)$. The fourth-order Runge-Kutta method is selected as the time difference scheme, and the time step is 0.01 time unit (TU). It is proved that in this model, c_1 and κ are the most sensitive parameters among the four atmospheric parameters (σ, κ, b, c_1) , while c_2 is the most sensitive parameter among the three oceanic parameters (c_2, O_m, O_d) (Han et al., 2015). Therefore, c_1 and c_2 are selected as the parameters to be optimized.

3.2 OSSE setup

In the truth model, starting from the state $(x_1, x_2, x_3, \omega) = (0, 1, 0, 0)$, the true initial condition of the MTW is obtained by free integration of 10^6 time steps with the standard values of parameters. Observations are sampled from the truth model and are added by Gaussian random errors. In order to simulate the fact that ocean observation is sparse than atmospheric observation, the ocean observation interval is 4 times of atmospheric observation interval. The standard deviation of atmospheric (oceanic) observation error is 2 (0.2).

In the assimilation model, it is assumed that model biases only come from biases of c_1 and c_2 . The first guess value is set to 0.11 and 1.1 for c_1 and c_2 , respectively. Starting from the state $(x_1, x_2, x_3, \omega) = (0, 1, 0, 0)$, the initial condition of the MTW is obtained by free integration of 10^6 time steps with the first guess values of c_1 and c_2 .

Figure 1 shows the time series of x_2 and ω in the MTW. We can see that the atmosphere variable x_2 varies much more quickly than the oceanic variable ω , which reflects the characteristics of coupled model. Besides, the model control run without CPO is far away from the truth, indicating the importance of CPO.

We design the following four experiments: (1) EXP-1: Observations are only used to optimize initial state variables; (2) EXP-2: Observations are used to optimize initial state variables and c_2 ; (3) EXP-3: Observations are used to optimize initial state variables and c_1 ; (4) EXP-4: Observations are used to optimize initial state variables, c_1 and c_2 . Note that all experiments use observed data from all media for assimilation. The four experiments can not only study the effect of CPO, but also explore the different performance of 4D-Var and A-4DnVar under different conditions.

3.3 Setup of 4D-Var and A-4DnVar

We use 4D-Var and A-4DnVar to implement each of the above four experiments. We focus on whether A-4DnVar can es-

timate the adjoint model accurately in CPO, so A-4DnVar imitates the implementation of 4D-Var. Both algorithms use the same static background error covariance matrix B_c and optimization algorithm. The only difference is that A-4DnVar uses the ensemble to compute the gradient while 4D-Var uses the adjoint model to compute the gradient.

We use the same method as Yang et al. (2006) to generate of B_c for EXP-1. For EXP-2 to EXP-4, the background error covariance between initial state variables uses B_c of EXP-1. The background error covariance between the initial state and parameter as well as between c_1 and c_2 are set to 0. The background error variance of parameter c_1 (c_2) is 0.000 1 (0.01). μ is set to a value between 10^{-20} and 10^{-24} . It needs to be adjusted according to the specific situation.

Cost function can be calculated after the B matrix is determined. Figure 2 plots the variation of cost function in EXP-4 with respect to parameter c_1 (Fig. 2a) and c_2 (Fig. 2b) when the initial state variable and another parameter is truth. As the MTW increases, the number of local minimum points of the cost function also increases, indicating that the model nonlinearity is gradually introduced into the cost function. We can see that the variation of the cost function with respect to c_1 (c_2) becomes very complicated when the MTW is 480 (720) time steps, which brings a huge challenge to both 4D-Var and A-4DnVar.

Due to the fact that many local minimums exist when the MTW is long, the standard quasi-Newton method like L-BFGS (Liu and Nocedal, 1989) often fails to converge to the global minimum point. The limited memory bundle method (LMBM) (Haarala et al., 2004, 2007; Steward et al., 2012) can solve this problem to a certain extent (Han et al., 2015), so we use it to optimize the cost function. The LMBM algorithm enhances the probability of finding the global minimum of the cost function, but it also needs to reduce the difficulty in searching for the global minimum of the cost function due to the strong nonlinearity of the model. Thus, the quasi static variational assimilation

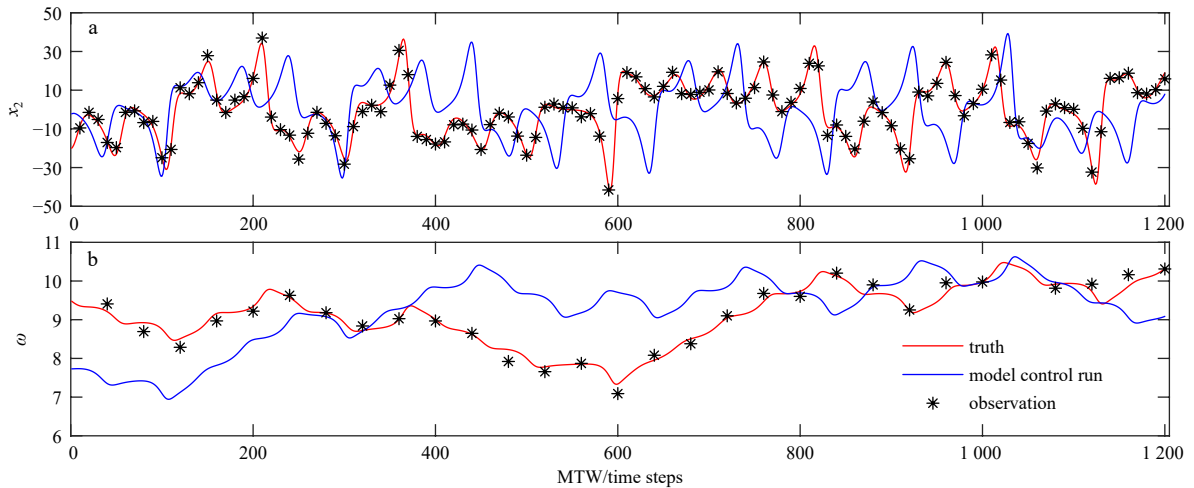


Fig. 1. Time series of x_2 (a) and ω (b) in the MTW for the truth (red), model control run (blue) and observation (star). The observation interval of the atmosphere (ocean) is 10 (40) steps.

Table 1. List of “twin” experiments

Experiment	Observations	Adjusted variables and parameters
EXP-1	x_1, x_2, x_3, ω	x_1, x_2, x_3, ω
EXP-2	x_1, x_2, x_3, ω	$x_1, x_2, x_3, \omega, c_2$
EXP-3	x_1, x_2, x_3, ω	$x_1, x_2, x_3, \omega, c_1$
EXP-4	x_1, x_2, x_3, ω	$x_1, x_2, x_3, \omega, c_1, c_2$

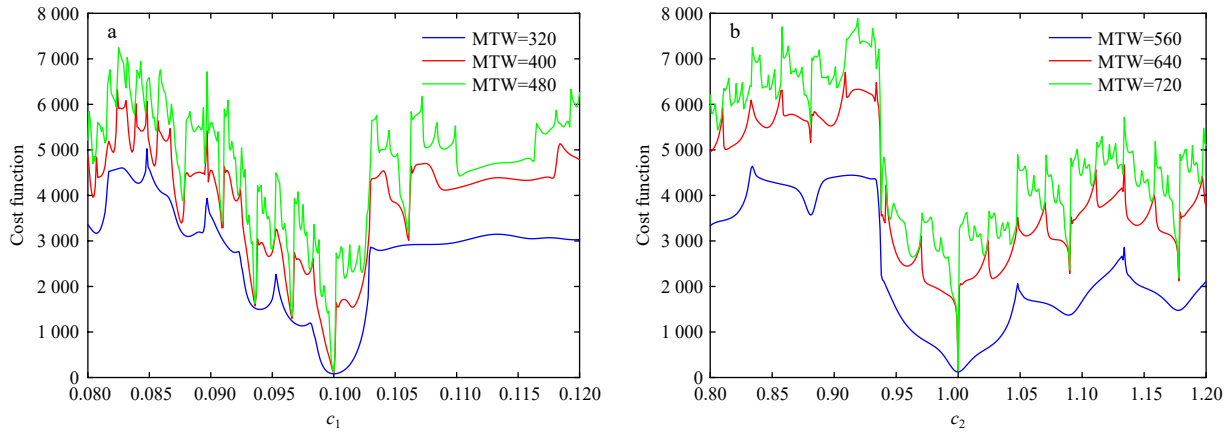


Fig. 2. Variation of cost function in EXP-4 with respect to parameter c_1 (a) and c_2 (b) when the initial state variable and another parameter is truth. The blue, red and green curves respectively correspond to the minimization time window (MTW) of 320, 400 and 480 time steps (560, 640 and 720 time steps).

algorithm (QSVA) (Pires et al., 1996) is used. QSVA means that the assimilation starts from a short MTW, and then gradually increases the MTW. The assimilation result of the previous short window acts as the background value for next long window.

4 Results

Results of the two methods under different conditions is compared in this section. Two indexes are used: one is the cost function, i.e., Eq. (1); the other is root mean square error (RMSE) that is computed by

$$\text{RMSE}(x) = \sqrt{\frac{1}{T} \sum_{t=1}^T (x_{t,\text{analysis}} - x_{t,\text{truth}})^2}, \quad (19)$$

where T is the length of MTW, $x_{t,\text{analysis}}$ and $x_{t,\text{truth}}$ are the analysis and truth of state variables at time t .

4.1 Impacts of observation interval and MTW

The observation interval and MTW are expected to have a great influence on 4D-Var. The smaller the observation interval

is, the more observation data will be used. For MTW, when it is very short, it is easy to find the global minimum value of the cost function, although optimal value might have large error. For example, the response of state variables to the parameters of low-frequency variables is slow, and a short MTW cannot accurately optimize such parameters. If the MTW is too long, the nonlinearity of model will make the cost function non-convex (Fig. 2) that is difficult to be optimized. A-4DEnVar has the same problems and suffers from the gradient calculated by ensemble.

For the four experiments in Section 3.2, the ensemble size is set to 20 for the A-4DEnVar. To examine the impact of observation interval and MTW, the atmospheric observation interval is set to 5, 10 and 20 time steps (the ocean observation interval is 4 times that of the atmosphere observation interval), and 13 values of MTW between 240 and 1 200 time steps with the even interval of 80 time steps are considered. Figures 3–6 show the RMSE of x_2 with respect to MTW and observation interval in EXP-1 to EXP-4. Figures 7–10 shows the RMSE of ω with respect to MTW and observation interval in EXP-1 to EXP-4.

Compared with EXP-1 (Figs 3 and 7) and EXP-3 (Figs 5 and 9), EXP-2 (Figs 4 and 8) and EXP-4 (Figs 6 and 10), the importance of

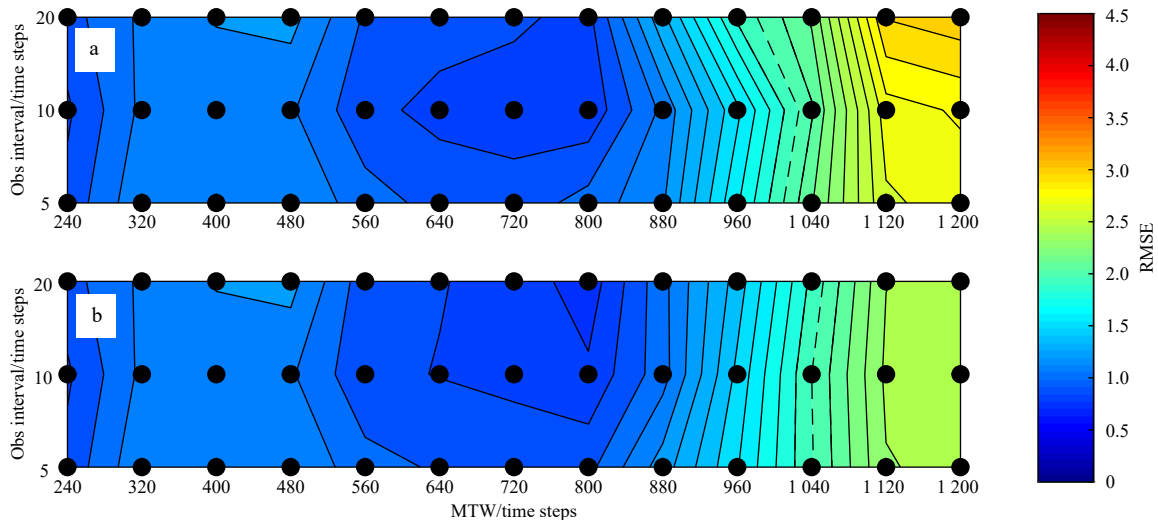


Fig. 3. RMSE of x_2 in A-4DEnVar (a) and 4D-Var (b) in EXP-1. The black solid circle is the data point of the contour, and the dash curve is the contour of 2.0 (observation error standard deviation).

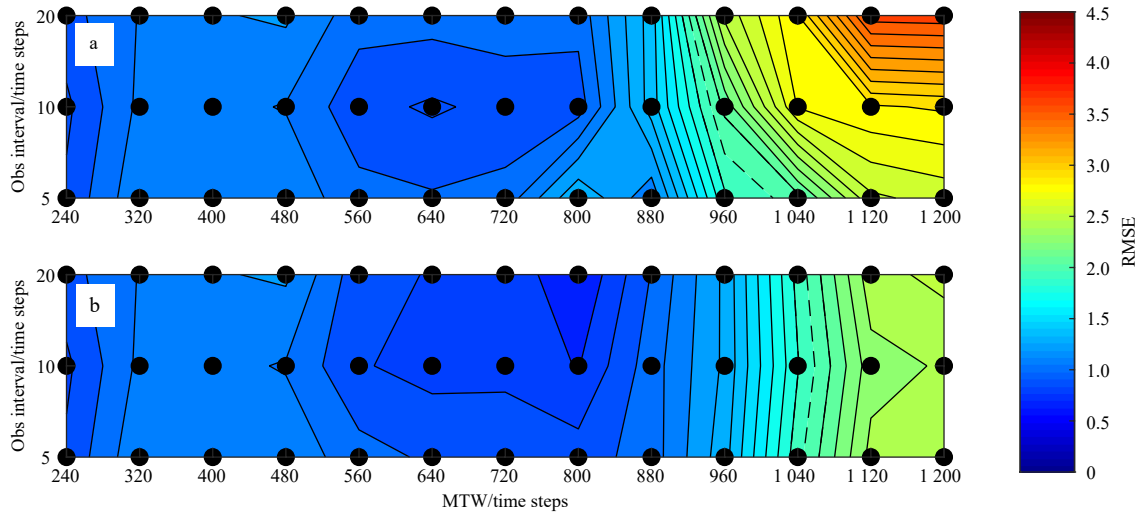


Fig. 4. RMSE of x_2 in A-4D-EnVar (a) and 4D-Var (b) in EXP-2. The black solid circle is the data point of the contour, and the dash curve is the contour of 2.0 (observation error standard deviation).

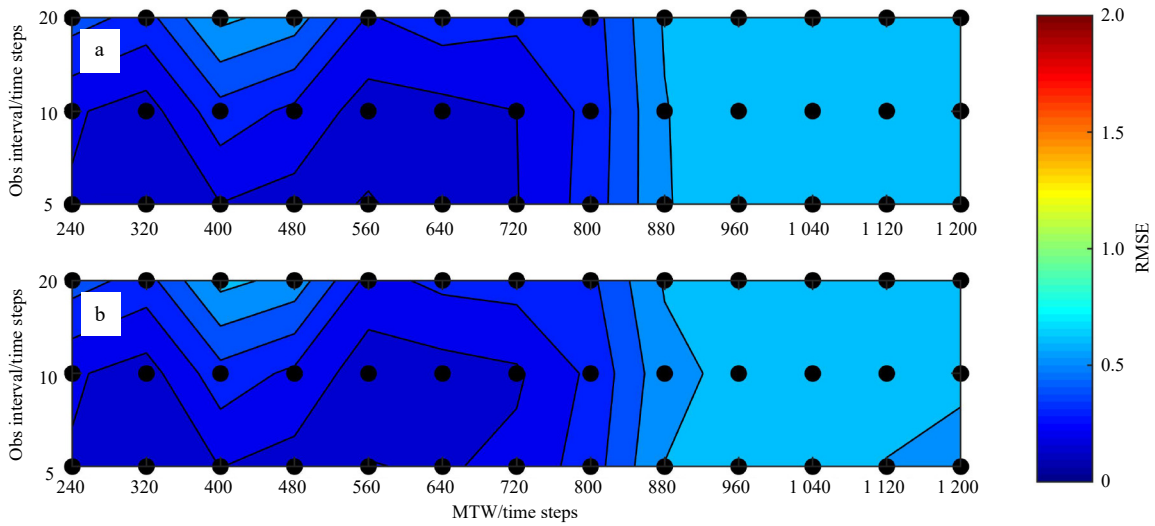


Fig. 5. RMSE of x_2 in A-4D-EnVar (a) and 4D-Var (b) in EXP-3. The black solid circle is the data point of the contour.

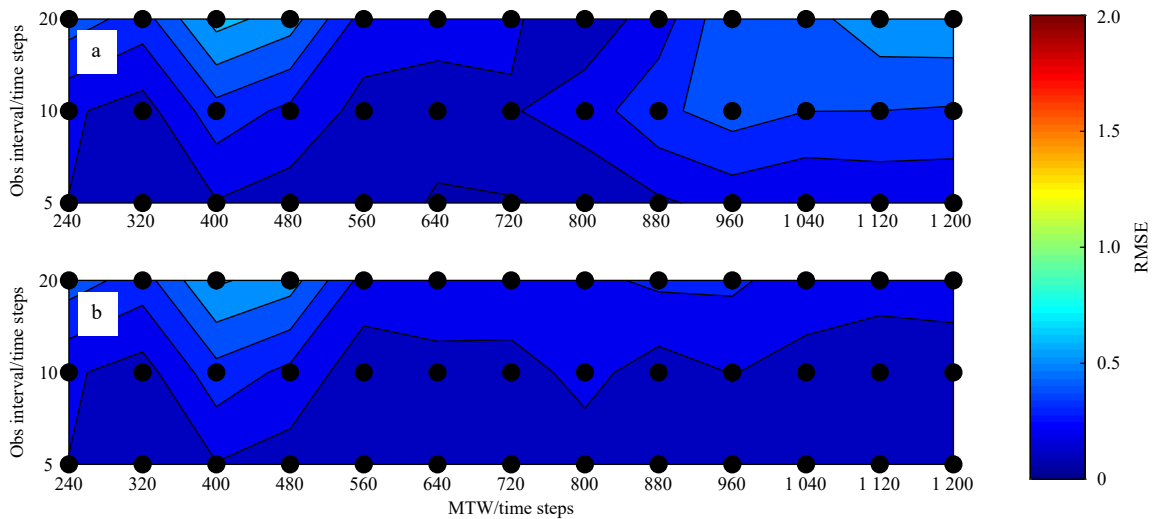


Fig. 6. RMSE of x_2 in A-4D-EnVar (a) and 4D-Var (b) in EXP-4. The black solid circle is the data point of the contour.

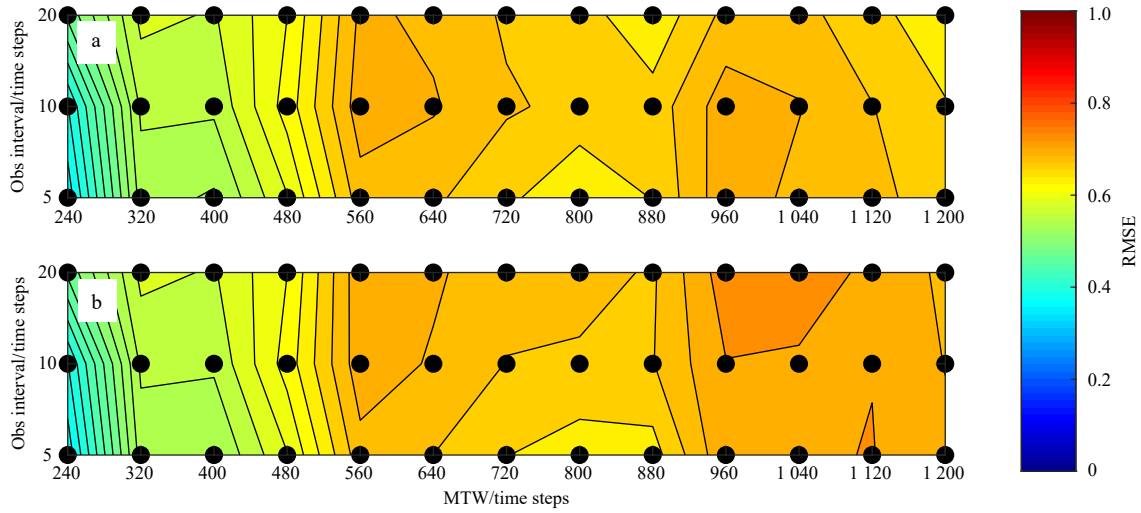


Fig. 7. RMSE of ω in A-4DEnVar (a) and 4D-Var (b) in EXP-1. The black solid circle is the data point of the contour.

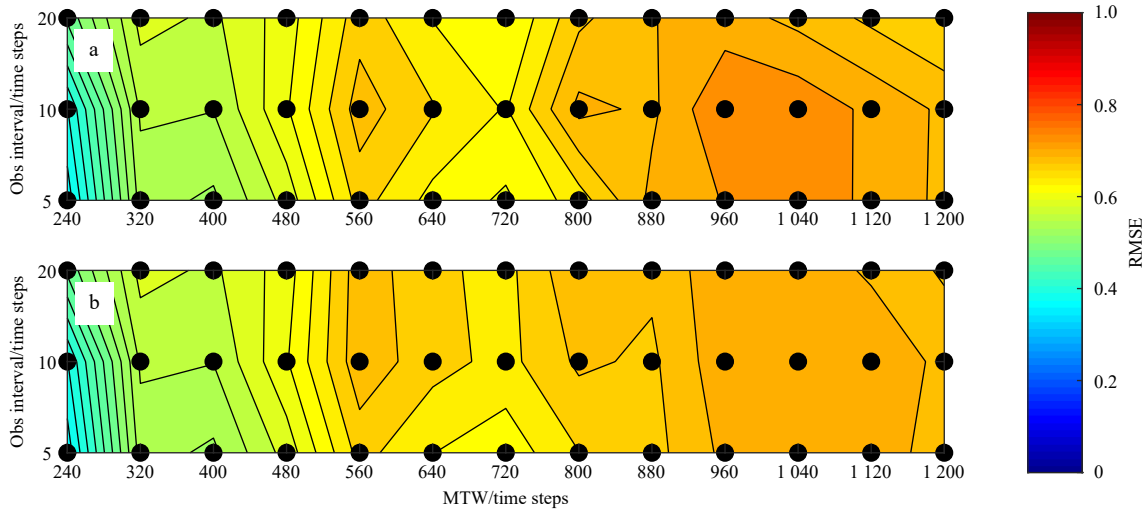


Fig. 8. RMSE of ω in A-4DEnVar (a) and 4D-Var (b) in EXP-2. The black solid circle is the data point of the contour.

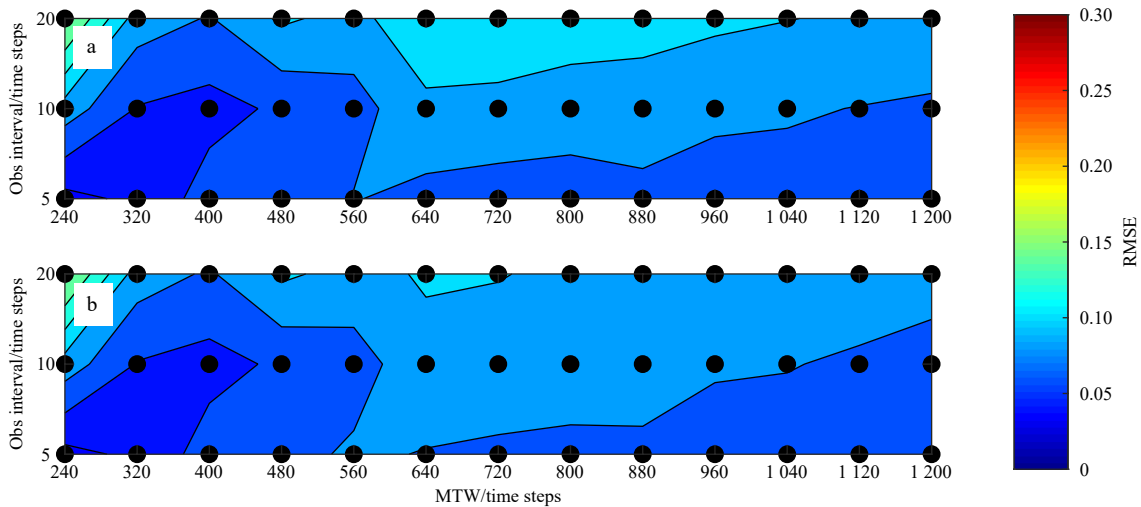


Fig. 9. RMSE of ω in A-4DEnVar (a) and 4D-Var (b) in EXP-3. The black solid circle is the data point of the contour.

optimizing c_1 is evident. The RMSE of x_2 and ω is significantly large without optimizing c_1 . RMSE of ω in EXP-1 and EXP-2 is even larger than the standard deviation of ocean observation er-

ror 0.2 in all cases. Comparison between EXP-1 (Figs 3 and 7) and EXP-2 (Figs 4 and 8) shows no obvious differences, indicating that optimization of c_2 has no obvious effect when c_1 is not op-

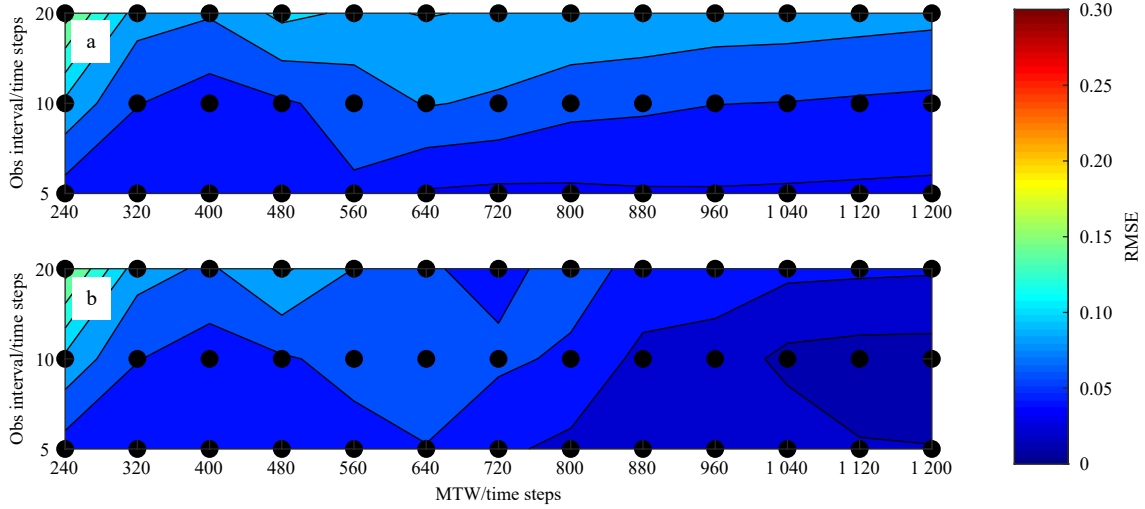


Fig. 10. RMSE of ω in A-4D-EnVar (a) and 4D-Var (b) in EXP-4. The black solid circle is the data point of the contour.

timized. Comparison between EXP-3 (Figs 5 and 9) and EXP-4 (Figs 6 and 10) shows that the optimization effect in a short MTW is almost the same, but for a long MTW, EXP-4 is obviously better than EXP-3, demonstrating that the optimization of low-frequency parameter c_2 can further improve the optimization of c_1 only for long MTWs. Besides, comparison between Figs 3, 4, 7, 8 and Figs 5, 6, 9, 10 indicates that the impact of observation interval (MTW) is relatively obvious in EXP-3 and EXP-4 (EXP-1 and EXP-2).

As mentioned in Section 3, the only difference between the two methods in the four experiments is the way to obtain the gradient. The gradient solved by 4D-Var is certainly accurate. Thus, the result of 4D-Var is seen as the truth of A-4D-EnVar. Comparison between a and b in Figs 3–10, the two methods are mainly affected by the MTW rather than the observation interval. The results are almost the same when the MTW is less than 560 time steps. For MTWs in [560, 880] time steps, results are similar. When MTW is larger than 880 time steps, there are some differences. The largest MTW of 4D-Var and A-4D-EnVar are 1 200 time steps.

There are mainly two reasons which lead to the difference between A-4D-EnVar and 4D-Var for long MTWs. For one thing, the high order infinitesimals ignored in Eq. (9) will increase as MTW increases. Thus, the long MTW may cause large gradient error. For another, note that the cost function with respect to c_1 (c_2) becomes complicated when the MTW is 480 (720) time steps. Even if the gradient error is the same, it will have a great influence on optimization under long MTWs.

Figure 11 shows the difference (A-4D-EnVar minus 4D-Var) of the cost function with respect to the MTW with the atmospheric observation interval of 5, 10 and 20 time steps in EXP-1 to EXP-4. We can see that cost function of A-4D-EnVar is larger than that of 4D-Var in the most cases. When the MTW is short, there is no difference between A-4D-EnVar and 4D-Var. When the MTW is long, the difference between the two methods is obvious in EXP-1 and EXP-2, which might be caused by the error of c_1 . Besides, we can also see that the difference between the two methods is not significantly correlated with the observed interval.

4.2 Impacts of ensemble size

In general, the larger the ensemble size, the smaller the sampling error. However, due to the limitation of computational

cost, the ensemble size is often limited. Therefore, it is necessary to discuss the influence of the ensemble size on the accuracy. The observation interval of the atmosphere (ocean) is 10 (40) steps in this section. Figure 12 shows the difference (A-4D-EnVar minus 4D-Var) of cost function with respect to ensemble size in EXP-1 to EXP-4 with the MTW being 720 and 960 time steps. 720-time steps is the largest MTW with almost no difference in cost function. We find that the ensemble size nearly has no influence when the MTW is 720 time steps, even 10 members can get similar cost function as 4D-Var. When the MTW is 960 time steps, we can see that the difference of cost function converges to a small value as ensemble size increases. This is because that a large ensemble size reduces the sampling error and subsequently improves the accuracy of cost function.

5 Conclusions

CPO is an effective way to reduce model error in coupled model. In the realization of variational CPO, the optimization of low-frequency variables always needs a long MTW which might introduce the model nonlinearity. As a kind of 4D-EnVar method, A-4D-EnVar is inevitably influenced by the model nonlinearity. But the influence can be reduced by giving a small initial perturbation to the ensemble and updating the ensemble after every iteration. To verify the feasibility and the ability of A-4D-EnVar in CPO, we conduct “twin” experiments using A-4D-EnVar CPO for the first time.

Based on a simple coupled ocean-atmosphere model, whose atmospheric part is the strongly nonlinear Lorenz 1963 model, and the oceanic part is a slab ocean model, the effect of CPO between A-4D-EnVar and 4D-Var is compared. Results show that A-4D-EnVar and 4D-Var can achieve good results in CPO. Results of A-4D-EnVar are almost the same as those of 4D-Var for different observation intervals and small ensemble sizes when the MTW is less than 560 time steps. For MTWs in [560, 880], results are similar. There are some differences when the MTW is larger than 880 time steps. A-4D-EnVar and 4D-Var still work for 1 200 time steps-MTW. The main reasons of difference are the gradient error and the huge difficulty in searching minimum. Besides, the difference of the cost function for a large MTW is affected by model biases, while it is not significantly correlated with the observation interval. Moreover, large ensemble size has some improvement in results only if the MTW is longer than 720 time steps.

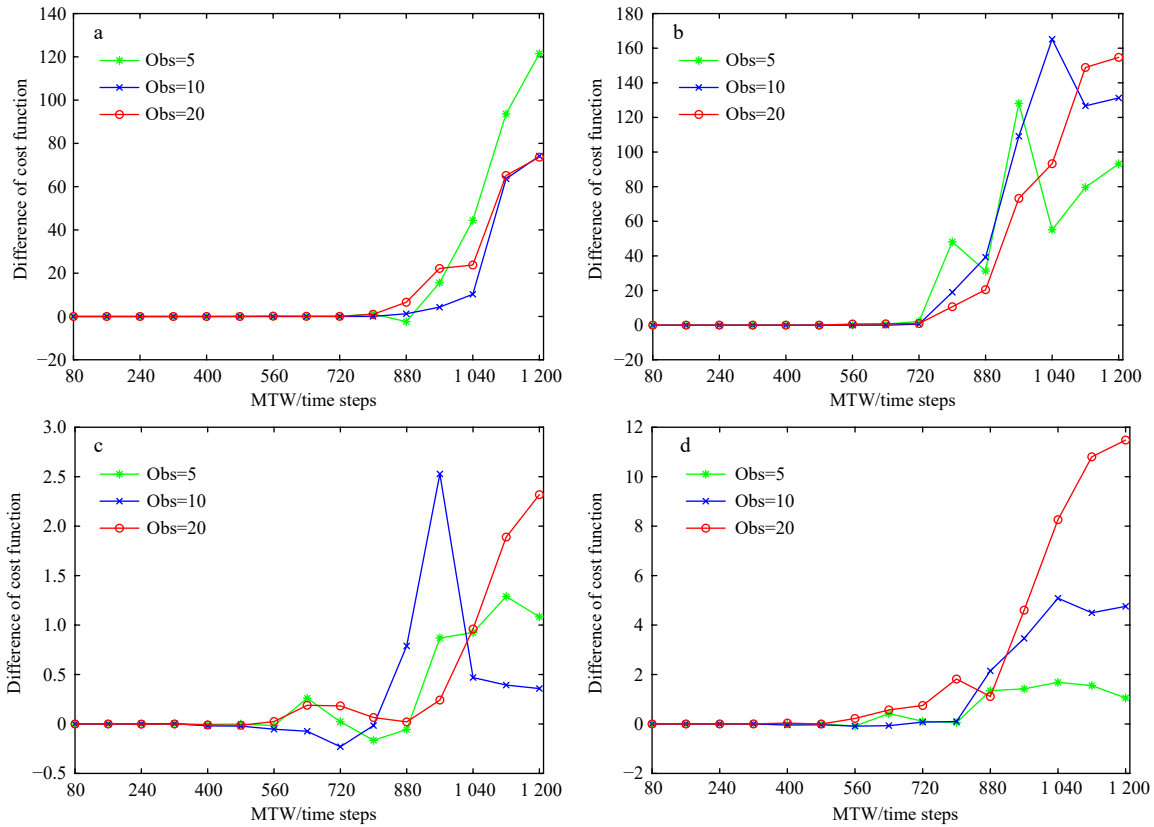


Fig. 11. Variation of the difference (A-4DEnVar minus 4D-Var) of the cost function with respect to the minimization time window (MTW) with the atmospheric observation interval of 5 (green), 10 (blue) and 20 (red) time steps in EXP-1 (a), EXP-2 (b), EXP-3 (c) and EXP-4 (d).

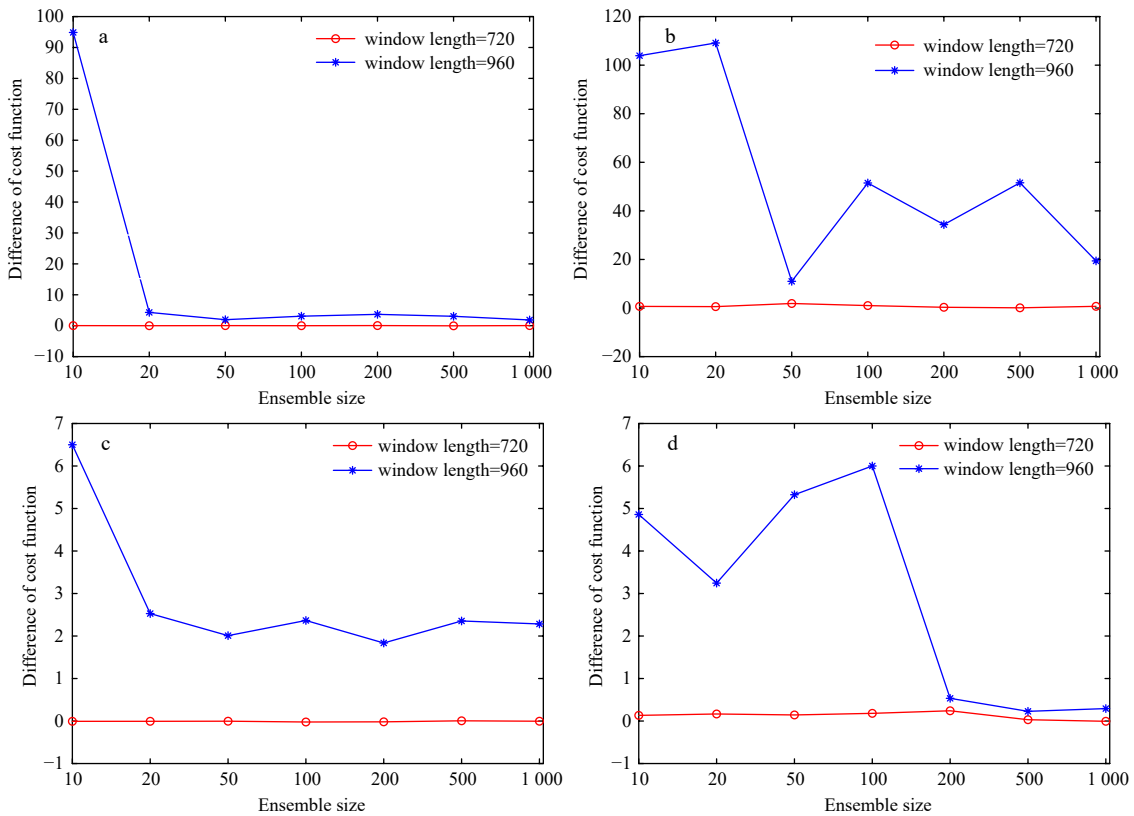


Fig. 12. Variation of the difference (A-4DEnVar minus 4D-Var) of cost function with respect to ensemble size in EXP-1 (a), EXP-2 (b), EXP-3 (c) and EXP-4 (d) with the MTW being 720 (red) and 960 (blue) time steps.

A-4DEnVar CPO has the potential to be applied in realistic CGCM, but there are still several challenges. First, the CGCM with high dimensionality, multi time scales and complex model bias will influence A-4DEnVar CPO. A probably way to partly improve it is introducing the conditional nonlinear optimal perturbation related to the parameter perturbations (Duan and Zhang, 2010; Mu et al., 2010), which is an effective method to select the especially important coupled parameters in the model. Second, the computational efficiency and the consumption of computation resources of the A-4DEnVar CPO applied to a CGCM are also issues to be investigated. It is important that we need to calculate the pseudo-inverse of high dimensionality matrix instead of calculating the matrix inversion directly (Liang et al., 2021), which makes that only ensemble-size matrix inversion is needed and reduces the computation effectively. Third, A-4DEnVar needs appropriate localization methods. One idea is making a Schur product between B matrix and a coefficient matrix, which imitates the EnKF (Hamill et al., 2001). More superior prediction skills are needed to realize the implementation of A-4DEnVar to a realistic CPO, which will be discussed in future works.

References

- Annan J D, Hargreaves J C, Edwards N R, et al. 2005. Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Modelling*, 8(1–2): 135–154, doi: [10.1016/j.ocemod.2003.12.004](https://doi.org/10.1016/j.ocemod.2003.12.004)
- Buehner M, Houtekamer P L, Charette C, et al. 2010a. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: description and single-observation experiments. *Monthly Weather Review*, 138(5): 1550–1566, doi: [10.1175/2009MWR3157.1](https://doi.org/10.1175/2009MWR3157.1)
- Buehner M, Houtekamer P L, Charette C, et al. 2010b. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: one-month experiments with real observations. *Monthly Weather Review*, 138(5): 1567–1586, doi: [10.1175/2009MWR3158.1](https://doi.org/10.1175/2009MWR3158.1)
- Courtier P, Thépaut J N, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519): 1367–1387, doi: [10.1002/qj.49712051912](https://doi.org/10.1002/qj.49712051912)
- Doucet A, de Freitas N, Gordon N. 2001. *Sequential Monte Carlo Methods in Practice*. New York: Springer, 582
- Du Huadong, Huang Sixun, Cai Qifa, et al. 2009. Studies of variational assimilation for the inversion of the coupled air-sea model. *Marine Science Bulletin*, 11(2): 13–22
- Duan Wansuo, Zhang Rui. 2010. Is model parameter error related to a significant spring predictability barrier for El Niño events? Results from a theoretical model. *Advances in Atmospheric Sciences*, 27(5): 1003–1013, doi: [10.1007/s00376-009-9166-4](https://doi.org/10.1007/s00376-009-9166-4)
- Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5): 10143–10162, doi: [10.1029/94JC00572](https://doi.org/10.1029/94JC00572)
- Fairbairn D, Pring S R, Lorenc A C, et al. 2014. A comparison of 4DVar with ensemble data assimilation methods. *Quarterly Journal of the Royal Meteorological Society*, 140(678): 281–294, doi: [10.1002/qj.2135](https://doi.org/10.1002/qj.2135)
- Haarala M, Miettinen K, Mäkelä M M. 2004. New limited memory bundle method for large-scale nonsmooth optimization. *Optimization Methods and Software*, 19(6): 673–692, doi: [10.1080/10556780410001689225](https://doi.org/10.1080/10556780410001689225)
- Haarala N, Miettinen K, Mäkelä M M. 2007. Globally convergent limited memory bundle method for large-scale nonsmooth optimization. *Mathematical Programming*, 109(1): 181–205, doi: [10.1007/s10107-006-0728-2](https://doi.org/10.1007/s10107-006-0728-2)
- Hamill T M, Whitaker J S, Snyder C. 2001. Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11): 2776–2790, doi: [10.1175/1520-0493\(2001\)129<2776:DDFOBE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2)
- Han Guijun, Wu Xinrong, Zhang Shaoqing, et al. 2013. Error covariance estimation for coupled data assimilation using a Lorenz atmosphere and a simple pycnocline ocean model. *Journal of Climate*, 26(24): 10218–10231, doi: [10.1175/JCLI-D-13-00236.1](https://doi.org/10.1175/JCLI-D-13-00236.1)
- Han Guijun, Wu Xinrong, Zhang Shaoqing, et al. 2015. A study of coupling parameter estimation implemented by 4D-Var and EnKF with a simple coupled system. *Advances in Meteorology*, 2015: 530764, doi: [10.1155/2015/530764](https://doi.org/10.1155/2015/530764)
- Han Guijun, Zhang Xuefeng, Zhang Shaoqing, et al. 2014. Mitigation of coupled model biases induced by dynamical core misfitting through parameter optimization: simulation with a simple pycnocline prediction model. *Nonlinear Processes in Geophysics*, 21(2): 357–366, doi: [10.5194/npg-21-357-2014](https://doi.org/10.5194/npg-21-357-2014)
- Ito K, Ishikawa Y, Awaji T. 2010. Specifying air-sea exchange coefficients in the high-wind regime of a mature tropical cyclone by an adjoint data assimilation method. *SOLA*, 6: 13–16, doi: [10.2151/sola.2010-004](https://doi.org/10.2151/sola.2010-004)
- Le Dimet F X, Talagrand O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A: Dynamic Meteorology and Oceanography*, 38(2): 97–110, doi: [10.3402/tellusa.v38i2.11706](https://doi.org/10.3402/tellusa.v38i2.11706)
- Lewis J M, Derber J C. 1985. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus A: Dynamic Meteorology and Oceanography*, 37(4): 309–322, doi: [10.3402/tellusa.v37i4.11675](https://doi.org/10.3402/tellusa.v37i4.11675)
- Liang Kangzhuang, Li Wei, Han Guijun, et al. 2021. An analytical four-dimensional ensemble-variational data assimilation scheme. *Journal of Advances in Modeling Earth Systems*, 13(1): e2020MS002314, doi: [10.1029/2020MS002314](https://doi.org/10.1029/2020MS002314)
- Lindskog M, Salonen K, Järvinen H, et al. 2004. Doppler radar wind data assimilation with HIRLAM 3DVAR. *Monthly Weather Review*, 132(5): 1081–1092, doi: [10.1175/1520-0493\(2004\)132<1081:DRWDAW>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1081:DRWDAW>2.0.CO;2)
- Liu D C, Nocedal J. 1989. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1): 503–528, doi: [10.1007/bf01589116](https://doi.org/10.1007/bf01589116)
- Liu Chengsi, Xiao Qingnong, Wang Bin. 2008. An ensemble-based four-dimensional variational data assimilation scheme. Part I: technical formulation and preliminary test. *Monthly Weather Review*, 136(9): 3363–3373, doi: [10.1175/2008MWR2312.1](https://doi.org/10.1175/2008MWR2312.1)
- Lorenc A C. 2003. The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 129(595): 3183–3203, doi: [10.1256/qj.02.132](https://doi.org/10.1256/qj.02.132)
- Lorenz E N. 1963. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2): 130–141, doi: [10.1175/1520-0469\(1963\)020<0130:Dnf>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:Dnf>2.0.CO;2)
- Lu Jingxi, Hsieh W W. 1998. On determining initial conditions and parameters in a simple coupled atmosphere-ocean model by adjoint data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 50(4): 534–544, doi: [10.3402/tellusa.v50i4.14531](https://doi.org/10.3402/tellusa.v50i4.14531)
- Mu Mu, Duan Wansuo, Wang Qiang, et al. 2010. An extension of conditional nonlinear optimal perturbation approach and its applications. *Nonlinear Processes in Geophysics*, 17(2): 211–220, doi: [10.5194/npg-17-211-2010](https://doi.org/10.5194/npg-17-211-2010)
- Pires C, Vautard R, Talagrand O. 1996. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus A: Dynamic Meteorology and Oceanography*, 48(1): 96–121, doi: [10.3402/tellusa.v48i1.11634](https://doi.org/10.3402/tellusa.v48i1.11634)
- Steward J L, Navon I M, Zupanski M, et al. 2012. Impact of nonsmooth observation operators on variational and sequential data assimilation for a limited-area shallow-water equation model. *Quarterly Journal of the Royal Meteorological Society*, 138(663): 323–339, doi: [10.1002/qj.935](https://doi.org/10.1002/qj.935)
- Tian Xiangjun, Feng Xiaobing. 2015. A non-linear least squares enhanced POD-4DVar algorithm for data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 67(1): 25340, doi: [10.3402/tellusa.v67.25340](https://doi.org/10.3402/tellusa.v67.25340)
- Tian Xiangjun, Xie Zhenghui, Dai Aiguo. 2008. An ensemble-based explicit four-dimensional variational assimilation method.

- Journal of Geophysical Research: Atmospheres, 113(D21): D21124, doi: [10.1029/2008JD010358](https://doi.org/10.1029/2008JD010358)
- Yang S C, Baker D, Li Hong, et al. 2006. Data assimilation as synchronization of truth and model: experiments with the three-variable Lorenz system. *Journal of the Atmospheric Sciences*, 63(9): 2340–2354, doi: [10.1175/jas3739.1](https://doi.org/10.1175/jas3739.1)
- Zhang Shaoqing. 2011a. Impact of observation-optimized model parameters on decadal predictions: simulation with a simple pycnocline prediction model. *Geophysical Research Letters*, 38(2): L02702, doi: [10.1029/2010GL046133](https://doi.org/10.1029/2010GL046133)
- Zhang Shaoqing. 2011b. A study of impacts of coupled model initial shocks and state-parameter optimization on climate predictions using a simple pycnocline prediction model. *Journal of Climate*, 24(23): 6210–6226, doi: [10.1175/JCLI-D-10-05003.1](https://doi.org/10.1175/JCLI-D-10-05003.1)
- Zhang Shaoqing, Liu Zhengyu, Rosati A, et al. 2012. A study of enhance parameter correction with coupled data assimilation for climate estimation and prediction using a simple coupled model. *Tellus A: Dynamic Meteorology and Oceanography*, 64(1): 10963, doi: [10.3402/tellusa.v64i0.10963](https://doi.org/10.3402/tellusa.v64i0.10963)
- Zhang Shaoqing, Liu Zhengyu, Zhang Xuefeng, et al. 2020. Coupled data assimilation and parameter estimation in coupled ocean-atmosphere models: a review. *Climate Dynamics*, 54(11–12): 5127–5144, doi: [10.1007/s00382-020-05275-6](https://doi.org/10.1007/s00382-020-05275-6)
- Zupanski M. 2005. Maximum likelihood ensemble filter: theoretical aspects. *Monthly Weather Review*, 133(6): 1710–1726, doi: [10.1175/MWR2946.1](https://doi.org/10.1175/MWR2946.1)