2015

Diffusion Filters for Variational Data Assimilation of Sea Surface Temperature in an Intermediate Climate Model

Chu, Peter C.

Research Article

Diffusion Filters for Variational Data Assimilation of Sea Surface Temperature in an Intermediate Climate Model

Xuefeng Zhang,[1,2] Dong Li,[2] Peter C. Chu,[3] Lianxin Zhang,[2,4] and Wei Li[2]

[1]*College of Automation, Harbin Engineering University, Harbin 150001, China*
[2]*Key Laboratory of Marine Environmental Information Technology, National Marine Data and Information Service,
State Oceanic Administration, Tianjin 300171, China*
[3]*Naval Ocean Analysis and Prediction Laboratory, Department of Oceanography, Naval Postgraduate School,
Monterey, CA 93943, USA*
[4]*College of Physical and Environmental Oceanography, Ocean University of China, Qingdao 266100, China*

Correspondence should be addressed to Xuefeng Zhang; xfz_nmdis@126.com

Sequential, adaptive, and gradient diffusion filters are implemented into spatial multiscale three-dimensional variational data assimilation (3DVAR) as alternative schemes to model background error covariance matrix for the commonly used correction scale method, recursive filter method, and sequential 3DVAR. The gradient diffusion filter (GDF) is verified by a two-dimensional sea surface temperature (SST) assimilation experiment. Compared to the existing DF, the new GDF scheme shows a superior performance in the assimilation experiment due to its success in extracting the spatial multiscale information. The GDF can retrieve successfully the longwave information over the whole analysis domain and the shortwave information over data-dense regions. After that, a perfect twin data assimilation experiment framework is designed to study the effect of the GDF on the state estimation based on an intermediate coupled model. In this framework, the assimilation model is subject to "biased" initial fields from the "truth" model. While the GDF reduces the model bias in general, it can enhance the accuracy of the state estimation in the region that the observations are removed, especially in the South Ocean. In addition, the higher forecast skill can be obtained through the better initial state fields produced by the GDF.

1. Introduction

In general, standard three-dimensional variational data assimilation (3DVAR) can be formulated as the minimization of the following cost function [1, 2]:

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1} (x - x_b)$$
$$+ \frac{1}{2}(Hx - y)^T R^{-1}(Hx - y) = J_b + J_o, \quad (1)$$

where x is the analysis vector, x_b is the background vector, y is the observation vector, H is an interpolation operator from model space to observation space, R is the observational error covariance matrix, $(\cdot)^T$ indicates transpose, and $(\cdot)^{-1}$ indicates inversion. B is the background error covariance matrix.

It is a challenge to determine B in any data assimilation including 3DVar. The spatial structure and the magnitude of the correction for the state variables being estimated are determined completely by B.

Two common approaches are used to prescribe B. The first approach is the correlation scale method (CSM) [3], in which B is represented by the Gaussian function:

$$B = A(x, y) \exp\left(-\frac{r_x^2}{L_x^2} - \frac{r_y^2}{L_y^2}\right), \quad (2)$$

where A is an estimate of the magnitude of the background error, r_x and r_y are the distances between two grid points, and L_x and L_y are the characteristic length scales reflecting the extent of spatial correction of the background error in the x and y directions, respectively. A more general anisotropic

shape with an ellipsoid about the spatial covariance can also be found [4]. It is noted that B is explicitly generated statistically using the correlation scales [3, 5–7]. However, limitations of the CSM are (1) positive value for each element in B (which is not always true), (2) nonexistence of B^{-1} unless using sufficiently small correction scales, and (3) requirement of large computer memory to store B since every element in B is calculated explicitly. To avoid the inversion of B and to speed up the convergence of descent algorithms such as the steepest descent and conjugate-gradient methods, a new vector w is introduced by Lorenc [8] and Derber and Rosati [3], defined as

$$w = B^{-1} (x - x_b). \tag{3}$$

Then the cost function J can now be rewritten as

$$J(w) = \frac{1}{2} w^T B^T w + \frac{1}{2} (HBw - d)^T R^{-1} (HBw - d), \tag{4}$$

where $d = y - Hx_b$ is the "innovation" vector, and for simplification, hereafter, we will call it observation.

Considering $B^T = B$, (4) is equal to

$$J(w) = \frac{1}{2} w^T Bw + \frac{1}{2} (HBw - d)^T R^{-1} (HBw - d). \tag{5}$$

The effect of Bw in (5) can be modeled by applying an equivalent spatial filter on w.

The second approach to prescribe B is the recursive filter method (RFM) [9],

$$\begin{aligned} Y_i &= \alpha Y_{i-1} + (1 - \alpha) X_i \quad \text{for } i = 1, \dots, n, \\ Z_i &= \alpha Z_{i+1} + (1 - \alpha) Y_i \quad \text{for } i = n, \dots, 1, \end{aligned} \tag{6}$$

where X_i is the initial value at grid point i, Y_i is the value after filtering for $i = 1$ to n, Z is the initial value after one pass of the filter in each direction, and α is the filter coefficient, which determines the extent of spreading of observational information over the analysis domain. Multipass filter can be built up by repeated application of (6). Multidimensional filter can be constructed by applying this one-dimensional filter in each direction. It can be shown [10] that such multidimensional filter, when applied with several passes, can accurately model isotropic Gaussian error correlations. The implementation using recursive filter to model B has been widely used due to its relatively computational inexpensiveness [10–13].

An outstanding issue of either CSM or RFM is its inefficiency in capturing the spatial multiscale information by observations. A difficulty in practice is how to properly choose the characteristic length scales L_x and L_y in the CSM or the filter coefficient α in the RFM. Observational studies show that (L_x, L_y) change with location, depth, and time [14–16]. If they are too large, the analysis is too smooth and shortwave information is lost. If they are too small, the analysis lacks coherent structure in data sparse regions because the longwave information cannot be properly corrected. Thus, in the past it has been thought that the characteristic length scales (L_x, L_y) in the CSM or the filter coefficient α in the RFM is responsible for the unsatisfactory analysis in the 3DVAR.

To avoid empirically or statistically setting the characteristic length scales and to correctly minimize the longwave and shortwave errors in turn, a sequential 3DVAR (S3DVar) method was developed [17] to assimilate sea surface temperature (SST) in a global ocean model [18]. The S3DVar method is simply composed of a series of 3DVars, each of which uses recursive filters with different filter coefficients. These 3DVars sweep through all resolvable scales by observational networks from longwaves to shortwaves. In addition, a multigrid data assimilation scheme was also introduced to extract the resolvable information from longwave to shortwave in an observational system [19]. Recently, a sequential variational approach based on the multigrid data assimilation method was proposed to accurately retrieve the multiscale information from available observation systems [20].

Since the matrix B is treated as the Gaussian type in the CSM and modeled as the diffusion process (or Gaussian filtering process) in the RFM, the spread of the information from the analysis point to the entire region is interpreted as the diffusion phenomenon [21]. The diffusion filter (DF) was developed on the base of the Gaussian diffusion process and therefore can be used directly to model B. Several spatial multiscale variational analysis schemes, based on the modification to the standard DF scheme, are proposed in this study. As a pilot study, one of the spatial multiscale variational analysis schemes, the gradient DF (GDF), is used to assimilate SST observations into an intermediate coupled model within a perfect "twin" experiment framework.

The paper is organized as follows. The methodology of the standard DF scheme is described in Section 2. Several spatial multiscale DF schemes are presented in Section 3. In Sections 4 and 5, simple observing/assimilation system simulation experiments and global SST simulation with an intermediate coupled atmosphere-ocean-land model are conducted to evaluate one of the new DF schemes, that is, the gradient DF (GDF), on the model estimation and forecast. The conclusions are summarized in Section 6.

2. Diffusion Filter

The DF is in fact a Gaussian filter. Given the following initial value problem for one-dimensional diffusion equation

$$\begin{aligned} \frac{\partial u}{\partial t} &= a \frac{\partial^2 u}{\partial x^2}, \\ u &= w(x), \\ t &= 0, \end{aligned} \tag{7}$$

where $a > 0$ is the diffusion coefficient, assumed to be constant. Its solution can be formulated by the convolution of $w(x)$ with a Gaussian kernel $G(x, t)$:

$$u(x, t) = w(x) * G(x, t), \tag{8}$$

where $(*)$ indicates convolution, $G(x, t) = (1/\sqrt{2\pi}\sigma)e^{-x^2/2\sigma^2}$, $\sigma = \sqrt{2at}$. That is, $u(x, t)$ is equivalent to applying a Gaussian filter on initial value $w(x)$. The second moment of the filter kernel is $\sigma^2 = 2at$, which characterizes the intrinsic spatial

scale. And σ^2 is only determined by diffusion coefficient a when "time" duration t is set to be constant, which implies that the larger the value of a is, the lower the frequency information of $w(x)$ would be acquired by $u(x, t)$.

Generally, in a two-dimensional finite domain, the diffusion model can be written by

$$\frac{\partial u}{\partial t} = a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial^2 u}{\partial y^2} \quad (x, y) \in \Omega, \ t \in (0, S],$$

$$u = w(x, y) \quad (x, y) \in \overline{\Omega}, \ t = 0, \tag{9}$$

$$\frac{\partial u}{\partial n} = 0 \quad (x, y) \in \Gamma,$$

where $\overline{\Omega} = [0, D] \times [0, H]$, $\Omega = \overline{\Omega} \backslash \Gamma$ is the interior domain of $\overline{\Omega}$, Γ is the boundary of $\overline{\Omega}$, n is the outer normal direction of Γ, and a and b are the diffusion coefficients in x and y directions, respectively.

If $u^S(w)$ denotes $u(w)|_{t=S}$, the cost function (5) then becomes

$$J(w) = \frac{1}{2}w^T u^S(w)$$

$$+ \frac{1}{2}\left(Hu^S(w) - d\right)^T R^{-1}\left(Hu^S(w) - d\right). \tag{10}$$

Now the analysis is converted to the problem of optimizing the initial value of the diffusion equation (9). To do so, we need the gradient of the cost function, which can be derived by using adjoint methods, just as four-dimensional variational (4DVAR) data assimilation usually does.

For convenience of illustration, a continuous adjoint system is considered and J_b is omitted. It is also assumed that the observations are located at analysis points and H is the identity matrix. Then the adjoint of the tangential linear model of (9) takes the following form:

$$-\frac{\partial R}{\partial t} - a\frac{\partial^2 R}{\partial x^2} - b\frac{\partial^2 R}{\partial y^2} = f \quad (x, z) \in \Omega, \ 0 < t \leq S,$$

$$R(x, y, S) = 0, \quad (x, z) \in \overline{\Omega}, \tag{11}$$

$$\frac{\partial R}{\partial n} = 0 \quad (x, y) \in \Gamma,$$

where

$$f = \begin{cases} d_{\text{res}}(w) = d - u^S(w) & t = S, \\ 0 & t < S. \end{cases} \tag{12}$$

Note that $d_{\text{res}}(w)$ is the observation residue, which characterizes the remaining observational signals after the abstracted information at current solution w, $u^S(w)$, has been removed form observations d, and $d_{\text{res}}(w)$ is set to be zero at the grid points with no observations.

The gradient of J with respect to w is $g(w) = -R^0(w)$, where $R^0(w)$ is the initial value of the adjoint variables. Once the adjoint model is available, the analysis can be performed in the following steps.

(1) Choose an appropriate diffusion coefficient a; give the initial guess of w ($w = 0$, for instance).

(2) Integrate the diffusion model (9) from "time" $t = 0$ to S to obtain $u^S(w)$.

(3) Calculate f according to (12).

(4) Integrate the adjoint model (11) from "time" $t = S$ to 0 to obtain $R^0(w)$; then the gradient $g(w)$ of the cost function J is $-R^0(w)$.

(5) Use descent algorithms to adjust w.

(6) Loop from step (2) until the convergence criterion is met.

Use of DF for determining the matrix B is called the DF method (DFM), which has the same computation loads as the RFM if the ADI difference scheme (or the other operator splitting scheme; see Appendix) is applied to calculate the diffusion equation (9). The diffusion filter scheme has the same problem as the recursive filter scheme in extracting observational information. As the extent of spatial dispersion is only determined by diffusion coefficient a when "time" duration t is set to be constant, if a is large, the shortwave information will be lost. Conversely, if a is small, the longwave information will not be properly captured. Obviously, the diffusion coefficient a plays the same role as the filter coefficient α does in the recursive filter scheme.

3. Spatial Multiscale Diffusion Filters

To retrieve longwave information over the whole domain and shortwave information over data-dense regions, three spatial multiscale variational analysis schemes, based on the diffusion filter, are proposed.

3.1. Sequential Diffusion Filter (SDF). The sequential diffusion filter (SDF) scheme is similar to the S3DVar method derived by Xie et al. [17]. The SDF scheme uses a sequence of 3DVars to obtain the final estimation to retrieve information from all wavelengths from long- to shortwaves in turn. The matrix B is modeled by applying the diffusion filter sequentially in x and y direction, respectively. SDF begins its sequence with a big value of the diffusion coefficient a; then an initial estimation is obtained through analyzing the observed data. After that, a S3DVar is solved using the diffusion filter with a smaller a than before. For the S3DVars, observations to be assimilated are produced by subtracting the previously analyzed values from the observations assimilated by the previous 3DVar until the diffusion coefficient a is small enough. The final estimation is the summation of all the previous 3Dvar analyses based on the diffusion filter.

From the above description, it is noted that the SDF scheme is a simple extension of the DF, in which information is retrieved step by step from long- to shortwaves. During the process of the SDF, B is changed gradually with the different diffusion coefficient a and thus becomes flow dependent and anisotropic following the multiscale information of the observation.

3.2. Adaptive Diffusion Filter (ADF). Due to the introduction of the heat diffusion equation, the gradient of the cost function with respect to the state variables can be obtained using the adjoint method with 4DVar. In general, the diffusion coefficients $a(x, y)$ and $b(x, y)$ are not constants but are space dependent. Therefore, it is possible to optimize not only the state variables but also the diffusion coefficients using 4DVar. State variables and diffusion coefficients are used together as control variables, so values of $a(x, y)$ and $b(x, y)$ will change adaptively according to the distribution of observations.

Set $\overline{I_{h_1}}, \overline{I_{h_2}}, \overline{I_\tau}$ as

$$\overline{I_{h_1}} = \left\{ x_i = ih_1 \mid i = 0, 1, \ldots, I, h_1 = \frac{D}{I} \right\},$$

$$\overline{I_{h_2}} = \left\{ y_i = jh_2 \mid j = 0, 1, \ldots, J, h_2 = \frac{H}{J} \right\},$$

$$\overline{I_\tau} = \left\{ t_n = n\tau \mid n = 0, 1, \ldots, N, \tau = \frac{S}{N} \right\},$$

$$\overline{\Omega_h} = \overline{I_{h_1}} \times \overline{I_{h_2}}.$$

(13)

The cost function is transferred to the following form:

$$J(w, a, b)$$

$$= \frac{1}{2} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} w_{i,j}^2$$

$$+ \frac{1}{2} \sum_{m=1}^{M} \left\{ p^{(m)} \left[\sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \left(q_{i,j}^{(m)} u_{i,j}^N \right) - d^{(m)} \right]^2 \right\},$$

(14)

where M is the number of observations and $q_{i,j}^{(m)}$ is the interpolation coefficient of the grid point (i, j) with respect to the mth observation. $p^{(m)}$ is the mth element of the diagonal matrix R^{-1}. For calculating the gradients of the cost function J with respect to w, a, b in (12), the discrete adjoint models of (A.1)–(A.11) should be deduced firstly according to the Lagrange multiplier method,

$$\frac{R_{i,j}^n - R_{i,j}^{n-1/2}}{\tau/2} + \Delta_y \left(b_{i,j} \overline{\Delta}_y \left(R_{i,j}^n + R_{i,j}^{n-1/2} \right) \right) = f,$$

$$n \in [1, N]; \ i \in [1, I-1]; \ j \in [1, J-1],$$

$$\frac{R_{i,j}^{n+1/2} - R_{i,j}^n}{\tau/2} + \Delta_x \left(a_{i,j} \overline{\Delta}_x \left(R_{i,j}^{n+1/2} + R_{i,j}^n \right) \right) = 0,$$

$$n \in [0, N-1]; \ i \in [1, I-1]; \ j \in [1, J-1],$$

$$R_{i,j}^N = 0,$$

$$i \in [0, I]; \ j \in [0, J],$$

$$\Delta_x R_{0,j}^n = 0,$$

$$\overline{\Delta}_x R_{I,j}^n = 0,$$

$$\Delta_y R_{i,0}^n = 0,$$

$$\overline{\Delta}_y R_{i,J}^n = 0,$$

$$n \in [0, N-1],$$

$$\Delta_x R_{0,j}^{n+1/2} = 0,$$

$$\overline{\Delta}_x R_{I,j}^{n+1/2} = 0,$$

$$\Delta_y R_{i,0}^{n+1/2} = 0,$$

$$\overline{\Delta}_y R_{i,J}^{n+1/2} = 0,$$

$$n \in [0, N-1],$$

(15)

where

$$f$$

$$= \begin{cases} 2 \sum_{m=1}^{M} \left\{ p^{(m)} q_{i,j}^{(m)} \left[\sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \left(q_{i,j}^{(m)} u_{i,j}^N \right) - d^{(m)} \right] \right\} & n = N, \\ 0 & n < N. \end{cases}$$

(16)

The gradients of the cost function J with respect to w, a, b can be expressed as follows:

$$\frac{\partial J}{\partial w_{i,j}} = w_{i,j} - \frac{R_{i,j}^0}{\tau} - \frac{1}{2} \Delta_y \left(b_{i,j} \overline{\Delta}_y R_{i,j}^0 \right),$$

$$\frac{\partial J}{\partial a_{i,j}} = \frac{1}{2} \sum_{n=1}^{N-1} \left[\overline{\Delta}_x u_{i,j}^{n+1/2} \cdot \overline{\Delta}_x \left(R_{i,j}^n + R_{i,j}^{n+1/2} \right) \right],$$

(17)

$$\frac{\partial J}{\partial b_{i,j}} = \frac{1}{2} \sum_{n=1}^{N-1} \left(\overline{\Delta}_y u_{i,j}^n \overline{\Delta}_y R_{i,j}^n + \overline{\Delta}_y u_{i,j}^{n+1} \overline{\Delta}_y R_{i,j}^{n+1/2} \right).$$

The process for the state estimation with the 4DVar is outlined as follows. (a) Begin with the initial w, a, b. (b) Integrate the model equations (A.1)–(A.11) forward into a fixed time window and calculate the value of the cost function $J(w, a, b)$ using (14). (c) Integrate the adjoint model (15) backward in time and calculate the values of the gradient of the cost function with respect to the control variables ∇J using (17). (d) With the values of the cost function $J(w, a, b)$ and the gradient ∇J, use the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton minimization algorithm to obtain the new values of the control variables, namely, the two diffusion coefficients a, b and the state variables w. (e) With the updated control variables from process (d), repeat processes (b), (c), and (d) until the convergence criterion for the minimization is satisfied.

3.3. Gradient Diffusion Filter (GDF). The algorithm is a variant of the spatial multiscale recursive filter [22]. For small diffusion coefficients a, b, the gradient contains not only all the observational signals from longer to shorter wavelengths, but also a lot of erroneous signals in data sparse regions, which causes lack of coherent longwave structure in space. If this gradient is simply introduced into the minimization algorithm without careful considerations, the analysis departs far from reality. Thus, a prerequisite for the minimization algorithm used in 3DVAR is needed to extract the longwave information from the gradient and at the same time to preserve the valuable shortwave signals.

However, the longwave information implied in the gradient cannot be made best use of to construct a reasonable descent direction in general minimization algorithms. Take the steepest descent algorithm as an example, in which the descent direction is simply chosen as $-g(w)$. Suppose the initial guess of w (i.e., w_0) is equal to zero. Then at the ith iteration, the new solution $w_i = w_{i-1} + l_{i-1} * (-g(w_{i-1}))$ is obtained by using a line search algorithm to find an appropriate step size l_{i-1}. According to what have been indicated, the gradient $g(w_0)$ actually represents certain scales of observations d, and these scales will be extracted by the line search at the first iteration and incorporated into a new solution w_1. However, if the diffusion coefficients a, b are small, the gradient $g(w_0)$ will lack coherent structure in data sparse regions though it actually carries all observational signals. And since the new solution w_1 is simply obtained along the descent direction, $-g(w_0)$, the same problem will also exist in w_1, which indicates that the longwave information of observations d is not effectively extracted from the gradient $g(w_0)$ at the first iteration. Similarly, at the second iteration, the longwave information of the observation residue after the first iteration will not be extracted from the gradient $g(w_1)$ and incorporated into the new solution w_2, and so on. As a consequence, in data sparse regions, the final analysis will also lose the longwave structure of observations. The same problem also exists for other minimization algorithms such as BFGS and the conjugate gradient method, for the same reason.

The GDF scheme is designed to effectively retrieve the longwave information over the whole domain and shortwave information over data-dense regions. Since the gradient carries all observational information, the main idea of this new scheme is to apply the diffusion filter on the gradient to extract the implied longwave signal. While the diffusion coefficient decreases continuously with iteration, the multiscale information, from long to short wavelengths, can be extracted successively. The algorithm is designed as follows:

(1) Give an initial guess of w (i.e., w_0) which equals zero. Then select diffusion coefficients a, b as small ones and give a large enough value to an extra diffusion coefficient denoted as β.

(2) Use the diffusion filter with coefficient a, b to calculate Bw in (5).

(3) Calculate the difference between observations d and HBw, namely, the observation residue.

(4) Calculate the gradient g of the cost function J with respect to w using the DF through the adjoint model.

(5) Apply the diffusion filter with coefficient β on $-g$ to calculate the descent direction $E(-g)$, where E represents a positive definite operator.

(6) Select $E(-g)$ as the descent direction, and use line search algorithm to find the step size, l; then w is adjusted to $w = w + l * E(-g)$.

(7) The value of β diminishes.

(8) Loop from step (2) until the convergence criterion is met.

If the background term J_b is involved in the cost function J, the same procedure is performed except that g calculated in step (4) is the gradient of cost function J, which includes both J_b and J_o.

4. Observing/Assimilation System Simulation Experiments

Observing/assimilation system simulation experiments are performed to evaluate the spatial multiscale variational analysis. The "truth" field in these experiments is represented by an analytic temperature field defined over the area of 100°E–110°E and 30°N–40°N. The "truth" field of the temperature is plotted in Figure 1(b), whose high nonlinearity can be seen from Figure 1(a). The grid resolution is set to 1/8° × 1/8°, and the total numbers of the grid are 80 × 80. The observational dataset is generated using the analytic solution. Observational error is simulated by adding a sample of white noise with a standard deviation of 0.2 to the "truth." Three experiments are conducted in which different configurations of numbers of observations are employed.

4.1. Experiment 1. In this experiment, the number of observations is set to 2000 at first, and the observations are randomly and uniformly distributed in the whole domain, which can be seen from the black dots in Figure 1(b). In the experiment with DF, several values of the diffusion coefficient are used to verify the impacts on the analyzed field. In the experiment with GDF, the processes (1)–(8) described in Section 3.3 are conducted. The diffusion coefficients a, b are set to a small value, of 0.1, which suggests almost all the observational signals, from long to short wavelengths, can be retrieved. However, a large enough value, 1.0, is given to the extra diffusion coefficient β of the gradient at the first step. For the subsequent steps, β is reduced by 0.1 from the previous step. At the last step, β becomes 0.1, which is small enough for the case. The limited memory BFGS quasi-Newton minimization algorithm [23] is used during the minimizing procedure.

The major scales of the truth field are reconstructed by 2000 observations almost fully using the GDF (Figure 2(a)), but not well reconstructed using the DF with different diffusion coefficients (a, b): 1.0 (Figure 2(b)), 0.5 (Figure 2(c)), and 0.1 (Figure 2(d)). The small scale features begin to dominate the analyzed fields when the diffusion coefficients are reduced gradually, while the large scale signals are contaminated dramatically by an abundance of small scale features.
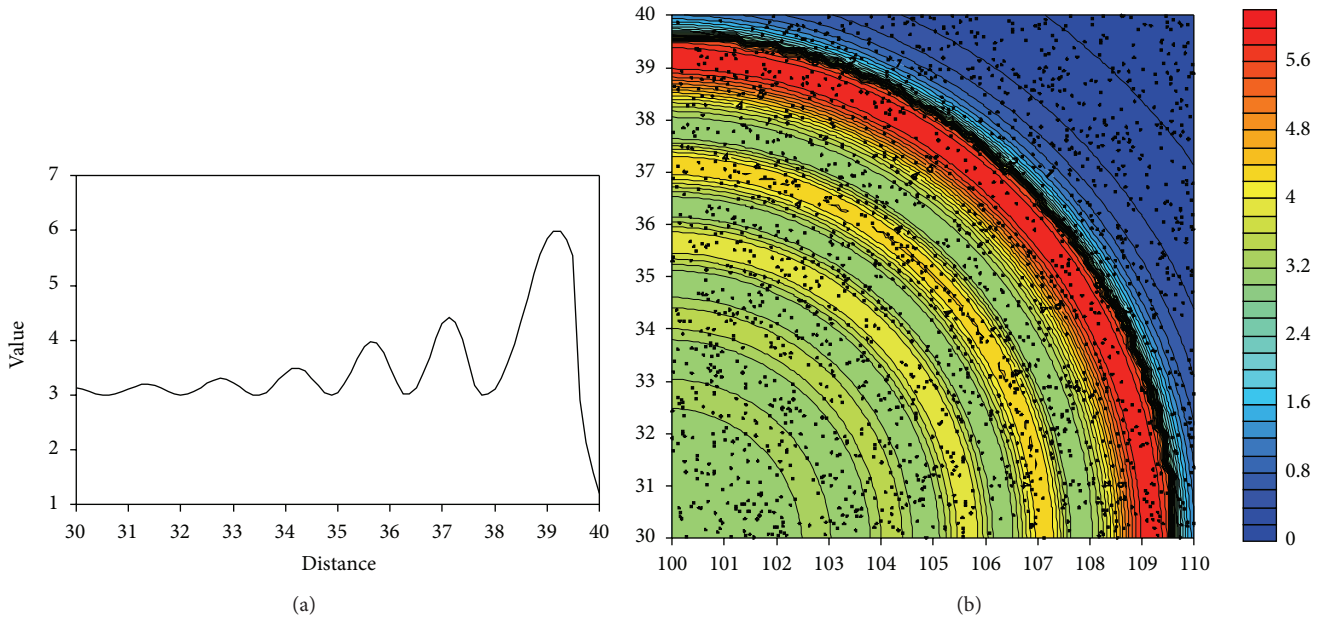
FIGURE 1: The true temperature field to be analyzed (unit: °C): (a) latitudinal variation along 100°E and (b) ichnography image. Black dots in the panel (b) show the distribution of 2000 random observations.

As He et al. [18] indicated, artificial signals can be produced during the data assimilation if the chosen diffusion coefficient cannot represent the actual scale. In contrast, the GDF can handle spatial multiscale analysis pretty well compared to the simple DF with a fixed diffusion coefficient. In addition, the GDF is easy to avoid in empirical selection of the diffusion coefficient.

Figure 3 shows the performance of GDF and DF when the number of observations decreases from 2000 to 500. The GDF (Figure 3(a)) can retrieve large scale information from observations and leave the unresolved scale as errors on top of the resolvable scales. These errors are smaller than those generated by DF with a fixed diffusion coefficient (Figure 3(b)) in the condition of the sparseness of the observations and the lack of information.

4.2. Experiment 2. The second experiment is conducted with removal of observations in the area of 103°E~107°E and 35°N~40°N (Figure 4) to further evaluate the GDF capability in retrieving the multiscale information from observations. The analyzed field of the GDF (Figure 5(a)) performs much better in the data void region than that of the DF with $(a, b) = 0.8$ (Figure 5(b)) and 0.5 (Figure 5(c)). The GDF can reconstruct the temperature field (Figure 5(a)) reasonably well despite the absence of the observations in the region as shown in Figure 4. The spatial pattern of the whole temperature field can be captured roughly according to the large scale information derived from all the observations in the whole analyzed region. However, the DF fails to reconstruct the temperature field and produces false features especially in the data void region. For example, a strong cold tongue is produced for $(a, b) = 0.8$ (Figure 5(b)), and large scale temperature field is distorted with displacement of the thermal front in the data void region for $(a, b) = 0.5$ (Figure 5(c)). Little information of

the observations can be extracted from data rich area to the data void region using DF.

Such capabilities make the GDF invaluable to get well represented values for the data void (or insufficiently covered) areas such as a typhoon-affected area during typhoon passage or the Southern Hemisphere Oceans (compared to other ocean basins). The GDF can reconstruct the analyzed field roughly according to the longwave information of the observations beyond the data void area such as typhoon-affected region or the Southern Ocean. On the other hand, both the standard DF and the traditional RF may lead to false results in the data void region, as shown in Figures 5(b)-5(c); an improper analysis is also likely to be produced, which will affect the analysis/forecast accuracy seriously.

In addition, several classical geostatistical tools, such as inverse distance to a power, triangulation with linear interpolation, and Kriging method are used to interpolate such observations (no white noise is imposed on the observations). Compared to the other two geostatistical tools, the Kriging method is able to accurately fill in the hidden information (Figures 6(a) and 6(b) versus 6(c)). However, compared with the variational method, the geostatistical tools have a limited application and cannot handle corrections between different analysis variables or physical balances and other constraints [20].

5. Global SST Assimilation Using GDF

In this section, we apply the GDF to assimilate the SST into an intermediate climate model to improve the climate representation and forecast.

5.1. Brief Description of an Intermediate Atmosphere-Ocean-Land Coupled Model. An intermediate atmosphere-ocean-land coupled model [24] is employed as the first step to
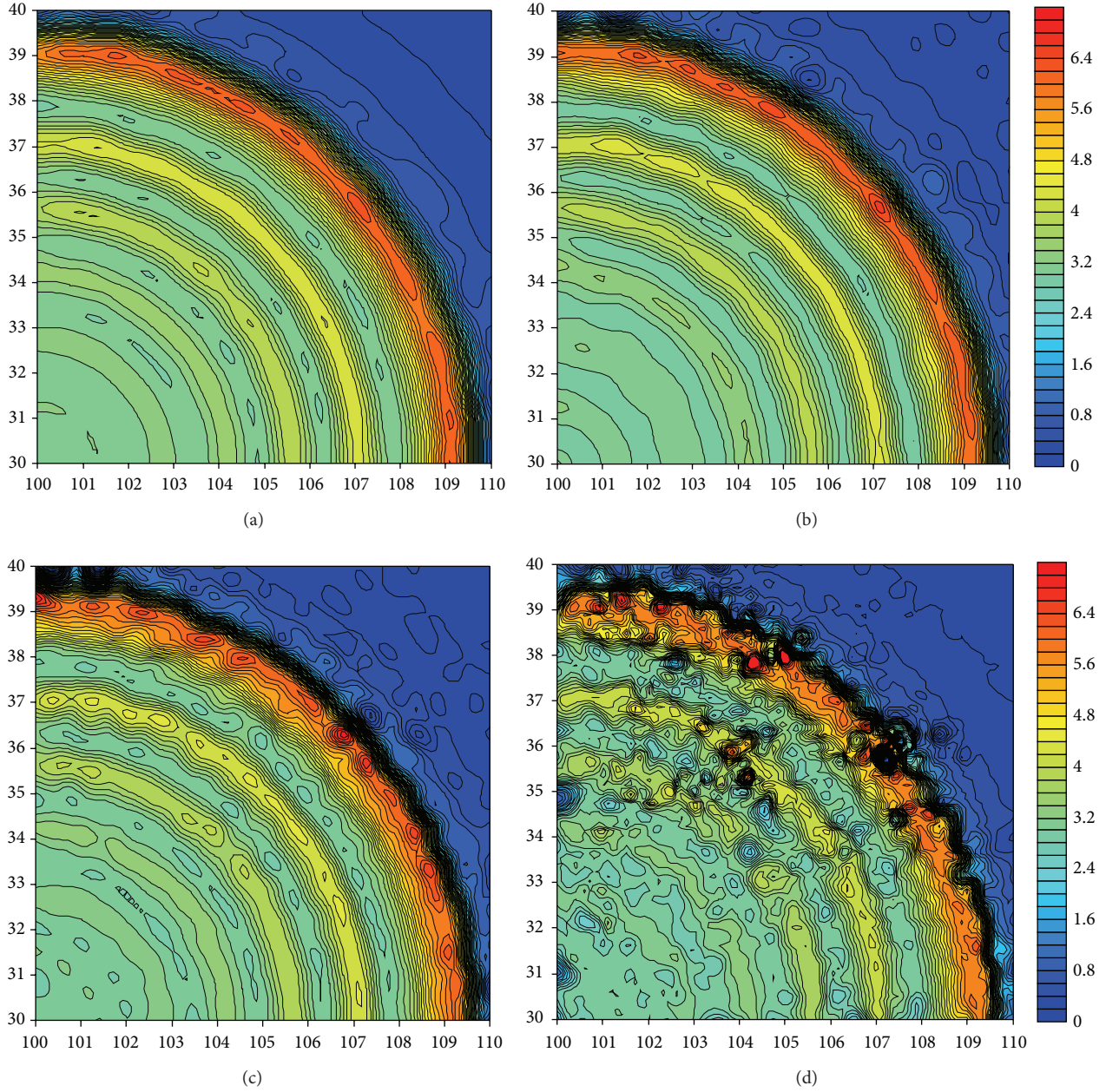
(a)

(b)

(c)

(d)

FIGURE 2: The analyzed temperature fields (unit: °C) using (a) GDF, (b) DF with $a, b = 1.0$, (c) DF with $a, b = 0.5$, and (d) DF with $a, b = 0.1$.

examine the GDF. Despite limitations in the representations of some basic physical processes such as the absence of ENSO dynamical mechanism, the model is of sufficient mathematical complexity for the purposes of this study. The intermediate coupled model has some successful applications in coupled data assimilation fields recently. For example, Wu et al. [25] investigated the impact of the geographic dependence of observing system on parameter estimation, and Zhang et al. [26] studied parameter optimization when the assimilation model contains biased physics within a biased assimilation experiment framework. The configuration of the model is presented here. The atmosphere is represented by

a global barotropic spectral model based on the potential vorticity conservation:

$$\frac{\partial q}{\partial t} + J(\psi, q) = \begin{cases} \lambda (T_o - \mu\psi) & \text{ocean} \cdot \text{surface} \\ \lambda (T_l - \mu\psi) & \text{land} \cdot \text{surface}, \end{cases} \quad (18)$$

where $q = \beta y + \nabla^2\psi$, $\beta = df/dy$, f is the Coriolis parameter, y is the meridional distance from the equator (northward positive), and ψ is the geostrophic atmosphere stream function. μ is a scale factor which converts stream function to temperature. λ is the flux coefficient from the ocean (land) to the atmosphere. T_o and T_l denote SST
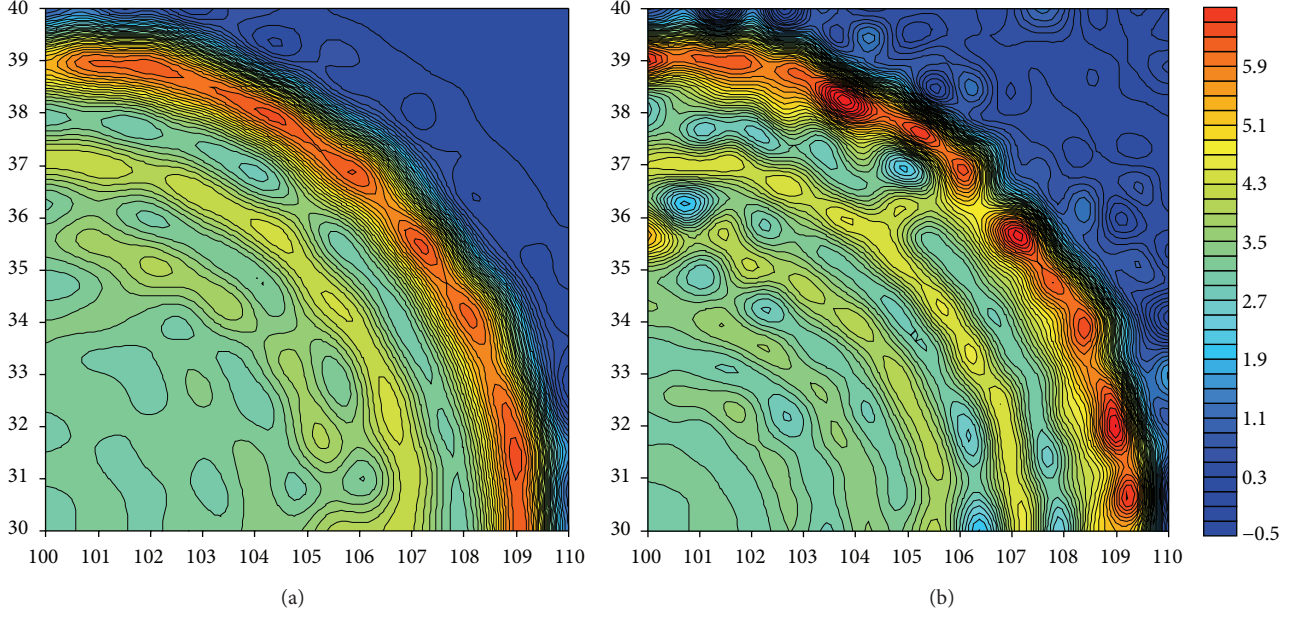
FIGURE 3: The analyzed temperature fields (unit: °C) with 500 observations: analyzed fields using (a) GDF and (b) DF with $a, b = 1.0$.

and land surface temperature (LST), respectively. Wu et al. [27] used the nonlinear atmospheric model to develop a compensatory approach of the fixed localization in EnKF analysis to improve short-term weather forecasts.

The ocean is composed of a 1.5-layer baroclinic ocean with a slab mixed layer [28] as

$$\frac{\partial}{\partial t}\left(-\frac{\phi}{L_0^2}\right) + \beta\frac{\partial}{\partial x}\phi = \gamma\nabla^2\psi - K_q\nabla^2\phi,$$

$$\frac{\partial}{\partial t}T_o + u\frac{\partial T_o}{\partial x} + v\frac{\partial T_o}{\partial y} - K_h\phi \qquad (19)$$

$$= -K_T T_o + A_T\nabla^2 T_o + s(\tau, t) + C_o(T_o - \mu\psi),$$

where ϕ is the oceanic stream function and $L_0^2 = g'h_0/f^2$ is the oceanic deformation radius, with g' and h_0 being the reduced gravity and mean thermocline depth. γ denotes momentum coupling coefficient between the atmosphere and ocean. K_q is the horizontal diffusive coefficient of ϕ. K_T and A_T are the damping coefficient and horizontal diffusive coefficient of T_o; $K_h = K_T \times \kappa \times f/g'$ [29], where κ is the ratio of upwelling to damping. C_o is the flux coefficient from the atmosphere to the ocean. $s(\tau, t)$ is the solar forcing which introduces the seasonal cycle.

The evolution of land surface temperature (LST) is given by

$$m\frac{\partial}{\partial t}T_l = -K_L T_l + A_L\nabla^2 T_l + s(\tau, t) + C_l(T_l - \mu\psi), \quad (20)$$

where m represents the ratio of heat capacity between the land and the ocean mixed layer, K_L and A_L are damping and diffusive coefficients of T_l, respectively, and C_l denotes the flux coefficient from the atmosphere to the land.
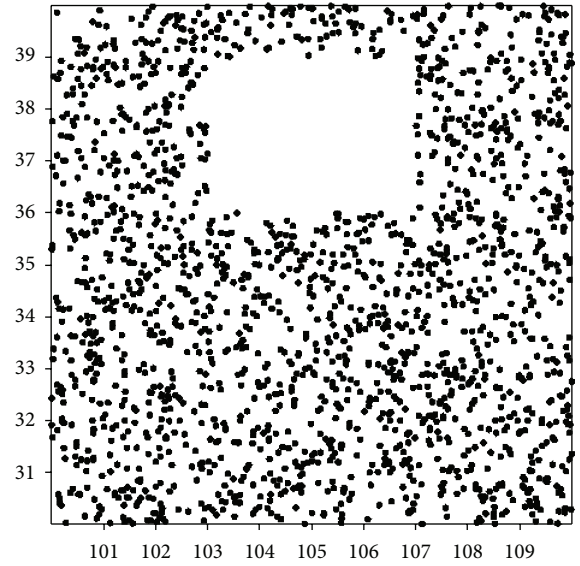

FIGURE 4: The distribution of 2000 random observations, but removal of the observations located in the range of 103°N~107°N and 35°E~40°E.

All the three model components adopt 64×54 Gaussian grid and are forwarded by a leap frog time stepping with a half hour integration step size. There are 2287 and 1169 grid points over the ocean and land, respectively. An Asselin-Robert time filter [30, 31] is introduced to damp spurious computational modes in the leap frog time integration. Default values of all parameters are listed in Table 1 in Wu et al. [24].

Starting from initial conditions $Z0 = (\psi^0, \phi^0, T_o^0, T_l^0)$, where ψ^0, ϕ^0, T_o^0, and T_l^0 are zonal mean values of corresponding climatological fields, the coupled model is run for 60 years
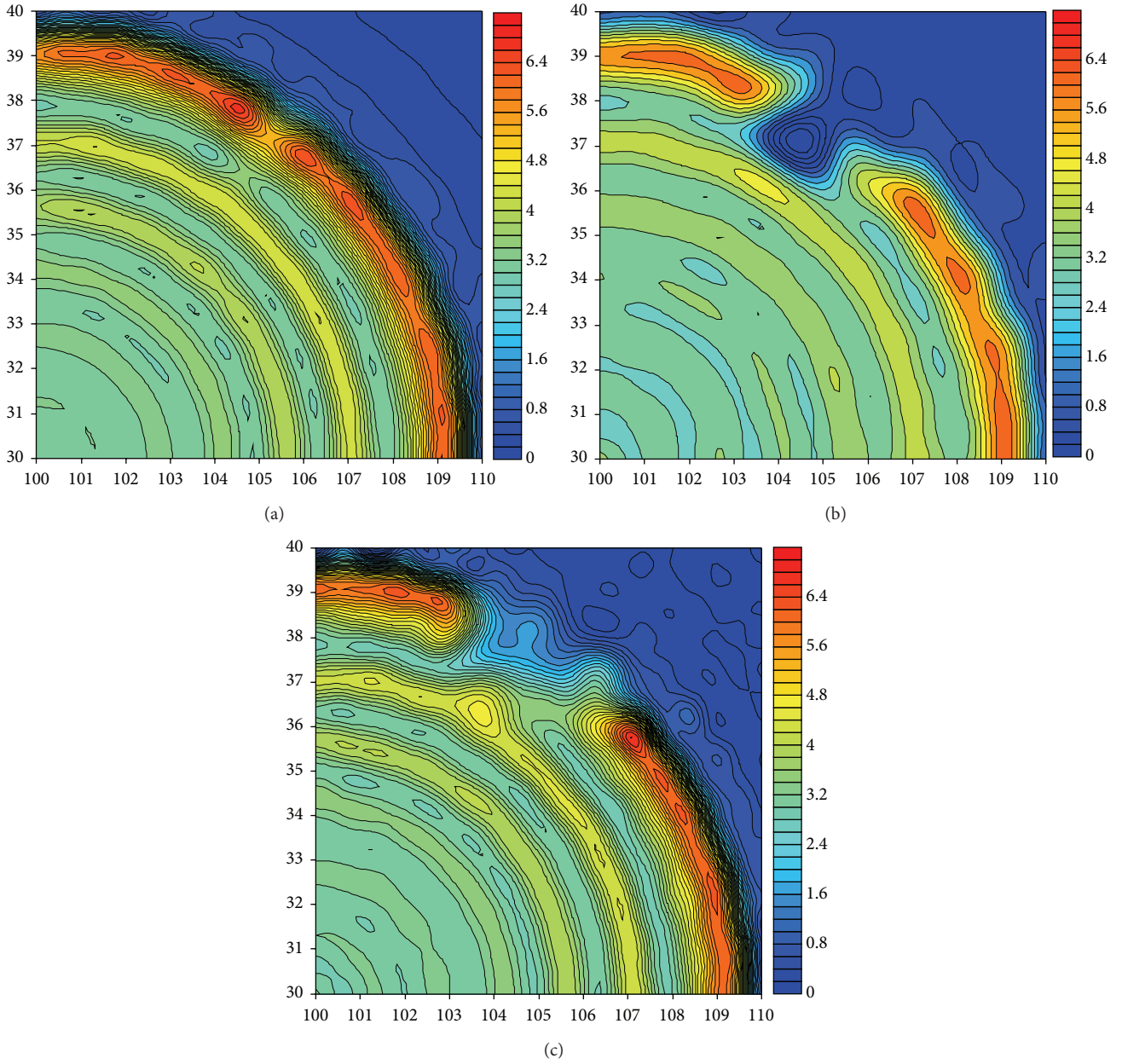
(a)

(b)

(c)

FIGURE 5: The analyzed temperature fields (unit: °C) using (a) GDF, (b) DF with $a, b = 0.8$, (c) DF with $a, b = 0.5$.

to generate the model states $Z1 = (\psi^1, \phi^1, T_o^1, T_l^1)$. The last 10 years' model states ($Z1$) are used as the "truth" fields. Figure 7 shows the annual mean of ψ (Figure 7(a)), ϕ (Figure 7(b)), T_o and T_l (Figure 7(c)), where the associated wave trains in the ψ field are observed. For ϕ, one can see the distinct pattern of the western boundary currents, gyre systems and the Antarctic Circumpolar Current (ACC). For T_o and T_l, reasonable temperature gradients are also produced. Note that the low temperature in tropical lands can be attributed to the linear damping of K_T in the solar forcing. The above model configuration is called the "truth" model, which has reasonable but rough representation for the basic climate characteristics of the atmosphere, land, and ocean.

5.2. Model "Bias" Arising from the Initial States. However, starting also from the same initial conditions $Z0$, the Gaussian random numbers are added to ψ^0 and ϕ^0, with standard deviations of $10^7 \text{ m}^2 \text{s}^{-1}$ (for ψ^0) and $10^5 \text{ m}^2 \text{s}^{-1}$ (for ϕ^0), respectively. The coupled model is also run for 60 years to generate the model states $Z2 = (\psi^2, \phi^2, T_o^2, T_l^2)$. The last 10 years' model states are used for analysis. This model configuration is called the biased model.

The model "biases" induced by perturbed initial fields are examined. Figure 8 shows time series of the spatial averaged root mean square errors (RMSEs) of ψ, ϕ, T_o, and T_l for the assimilation model, which are calculated according to the difference in the assimilation model and the "truth" model.
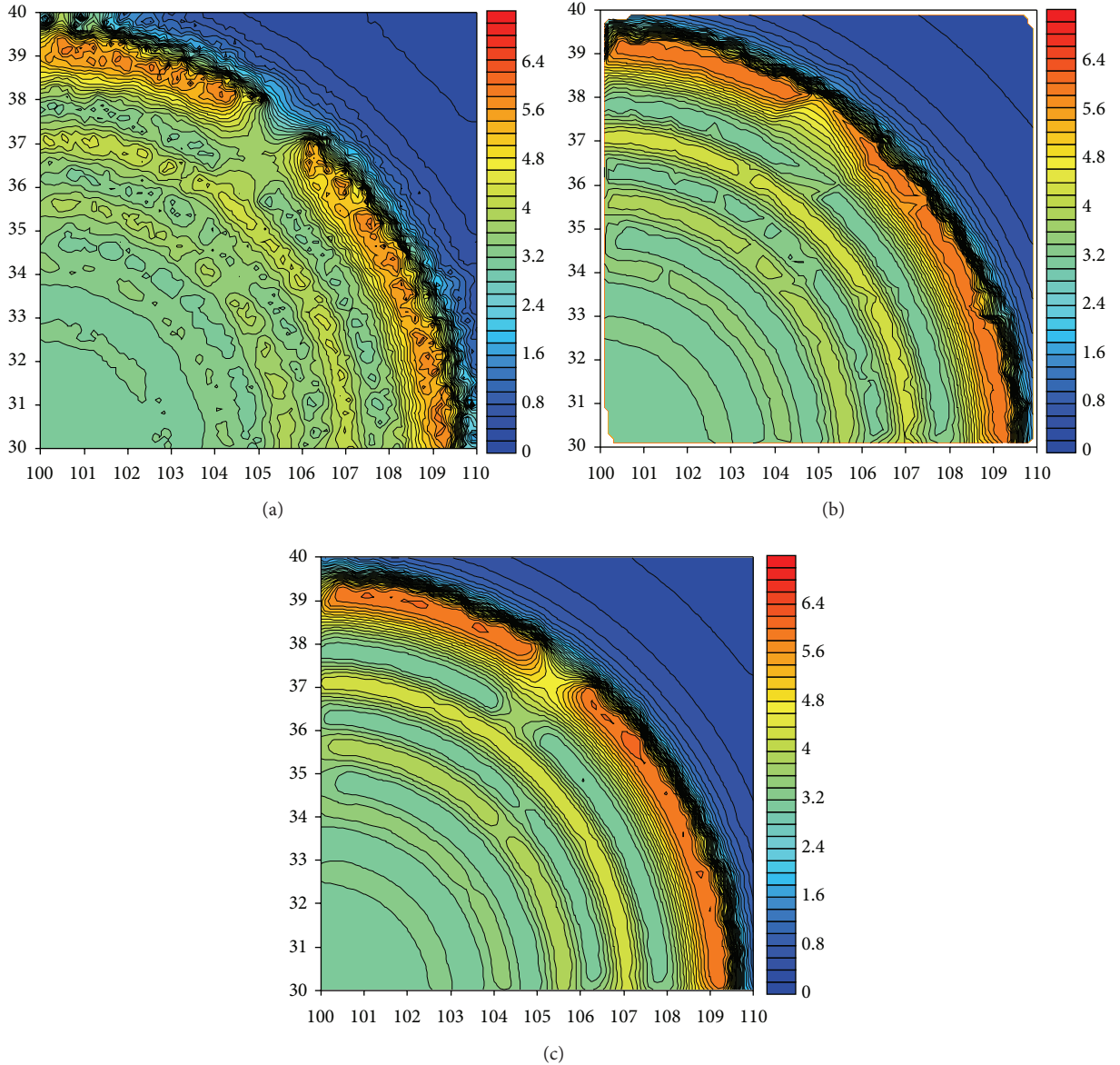
FIGURE 6: Analyzed temperature results (unit: °C) from classical geostatistical tools such as (a) inverse distance to a power, (b) triangulation with linear interpolation, and (c) Kriging method.

The obvious difference about all the four components can be seen from Figure 8. The RMSE of ψ reaches about $1.6 \times 10^7 \, \mathrm{m^2 s^{-1}}$ with a high frequent oscillation (see Figure 8(a)). In contrast, the RMSE of ϕ performs smoothly and rapidly decreases within the first year and gradually reaches a low and stable value about $10^4 \, \mathrm{m^2 s^{-1}}$ (see Figure 8(b)). The RMSEs of T_o (Figure 8(c)) and T_l (Figure 8(d)) increase rapidly in the first year, which are generated by the initially perturbed ψ^0 and ϕ^0 through the coupling. High frequency oscillation is noted in the time series of the RMSE of T_l, which indicates that the land surface temperature T_l is dominated by the atmospheric motion (ψ). However, time series of the RMSE of T_o is much smoother than that of T_l, indicating that T_o is modulated by the oceanic motion (ϕ).

The spatial distribution of RMSE of T_o for the biased model (Figure 9) shows a notable bias in the ACC region of the Southern Ocean with a maximum value over 15 K near the southern tip of Africa. Besides, obvious biases also exist in the west boundaries of the ocean in the subtropical regions.

5.3. Twin Experiment Design. In this section, with the intermediate model and DF/GDF data assimilation scheme described above, a perfect twin experiment framework is designed with the assumption that the errors of initial model states are the only source of assimilation model biases. Starting from the model states, $Z1$, described in Section 5.1, the "truth" model is run for 1 year to generate the time series of the "truth" states. Only synthetic observations of T_o
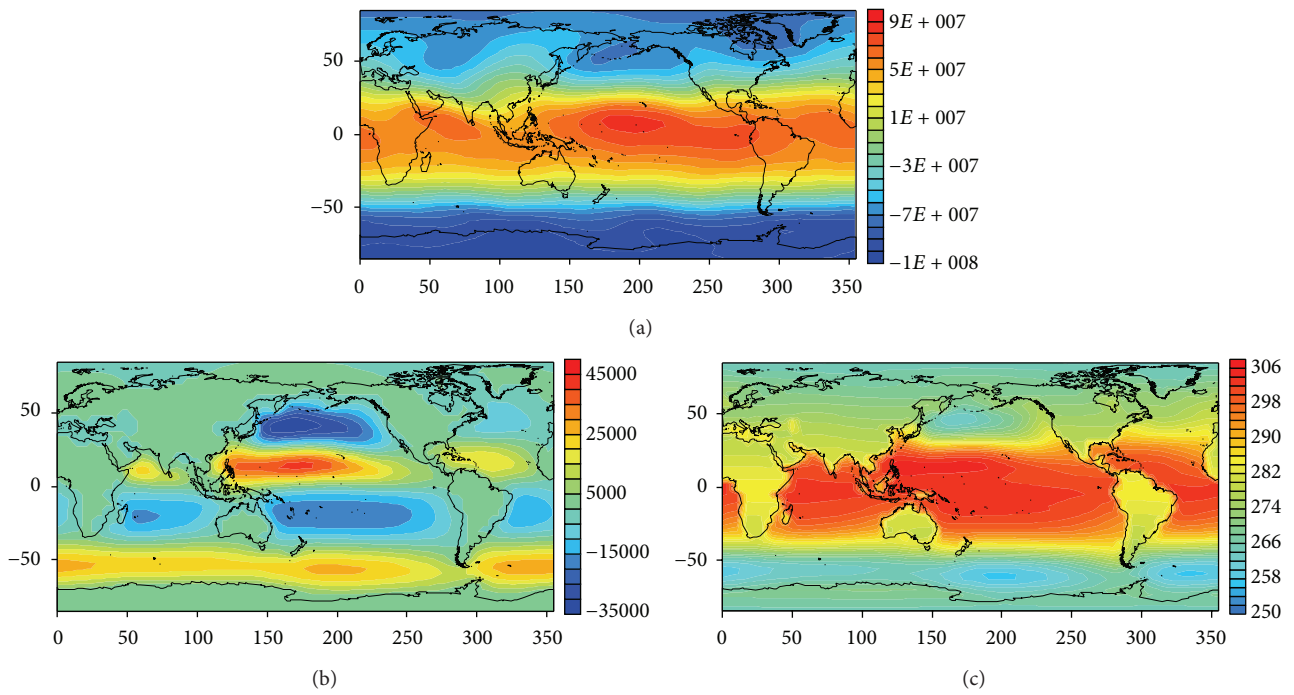
FIGURE 7: Annual mean of (a) atmospheric stream function (unit: $m^2 s^{-1}$), (b) oceanic stream function (unit: $m^2 s^{-1}$), and (c) sea surface temperature and land surface temperature (unit: K).
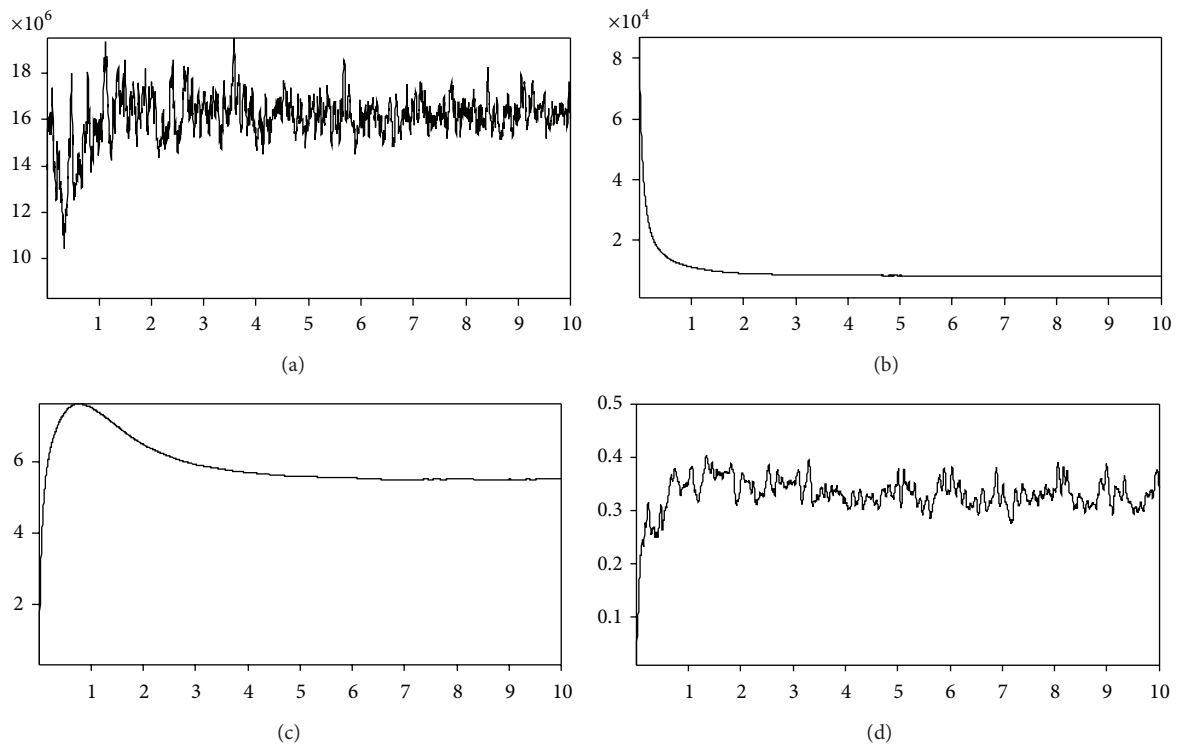

FIGURE 8: Time series of RMSEs of (a) atmospheric stream function, (b) oceanic stream function, (c) sea surface temperature, and (d) land surface temperature for the assimilation model.
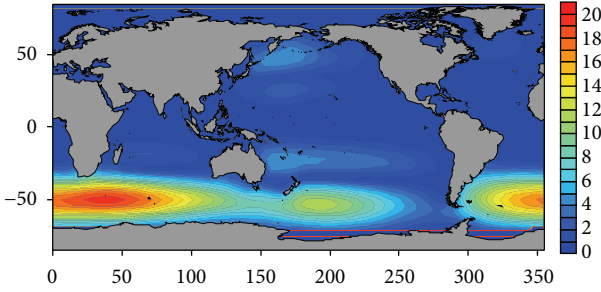
FIGURE 9: The spatial distribution of RMSE of sea surface temperature for the biased model.

are produced through sampling the "truth" states at specific observational frequencies. A Gaussian white noise is added for simulating observational errors. The standard deviations of observational errors are 0.5°C for T_o. The sampling period is 24 hours. The "observation" locations of T_o are global randomly distributed with the same density of the ocean model grid points.

The biased model uses the biased initial fields depicted in Section 5.2. Starting from the biased model states, $Z2$, the experiment E_GDF consists on assimilating observations into model states using the GDF scheme. In comparison, the experiment E_DF is carried out, where the standard DF scheme is used with the diffusion coefficients $a, b = 0.5$. In addition, a control run without any observational constraint, called CTRL, serves as a reference for the evaluation of assimilation experiments.

5.4. Impact of the GDF on the Estimate of the States. The performance of GDF is investigated. Figures 10(a)–10(d) show time series of RMSEs of ψ, ϕ, T_o, and T_l for the CTRL (solid line) and the GDF (dash line). Compared to the CTRL (solid line in Figure 10(c)), T_o of the GDF has significant improvement (dash line in Figure 10(c)), in which the RMSE decreases to approximately 0.5 K. Figure 11 presents the spatial distributions of RMSEs of T_o using GDF. The RMSE of T_o over ocean is obviously reduced compared with that of the CTRL (see Figure 9), especially in the Southern Ocean, the subtropical and the subpolar regions. In particular, the reduction of RMSE is much significant in the ACC region, in which the RMSE decreases from above 15 K to below 3 K.

Unlike the RMSEs of T_o, there is no direct observations constraint for ψ, ϕ, and T_l; therefore, their RMSEs decrease gradually owing to the effect of the coupling. The RMSEs of ψ for the GDF are reduced significantly from about $1.6 \times 10^7 \, m^2 s^{-1}$ to about $1.1 \times 10^7 \, m^2 s^{-1}$ with a high frequent oscillation (see Figure 10(a)). The ϕ in GDF is also improved significantly comparing to CTRL (solid versus dash lines in Figure 10(b)), whose RMSE decreases gradually and smoothly, but it does not reach a stable value within the experimental period, indicating that the low frequency signal needs a much longer time to reach equilibrium compared to the high frequency signal. For T_l, the GDF reduces the error by approximately 60%. Note that T_o has no direct effect on T_l, which can be realized according to the framework

of the coupling model (see (18)–(20)). Instead, T_o affects T_l indirectly via ψ. The improved T_o by the observational constraint increases the quality of ψ over land through the dynamical constraint. Then, the improved ψ ameliorates T_l through the process of the external forcing.

5.5. Removal of Observational Data in the Southern Ocean. In the real ocean, the observations are scarce in the southern polar region. Therefore, another set of data assimilation experiment is carried out, which is the same as the experiment in Section 5.3, but in which the observations, south of 50°S and 50°E~300°E, are removed completely.

Figures 12(a)–12(d) show the time series of RMSEs of ψ, ϕ, T_o, and T_l with the GDF (black line) and the standard DF with $a, b = 0.5$ (red line). The RMSE of T_o for the DF increases persistently during the experimental period, while the RMSE for the GDF begins to descend after 0.2 years and converges after 0.6 years (Figure 12(c)). When the diffusion coefficients are set to different values in the DF experiment (e.g., $a, b = 0.2, 0.8, 1.0$), similar results as the ones presented in Figure 12 are obtained. Results indicate that DF cannot correct the model bias in the data void region. However, the GDF is able to mitigate the model bias to some degree through extracting the spatial multiscale information from the available observations to the data void region. Figure 13 presents the spatial distributions of RMSEs of T_o for the GDF and the standard DF with $a, b = 0.5$. Compared to the DF, the GDF produces a significant improvement within the data void region in the Southern Ocean (compare Figures 13(a) and 13(b)).

The RMSE of ψ in the GDF is not always smaller than the DF owing to the strong nonlinear nature of the high frequent atmosphere (red line versus black line in Figure 12(a)). In contrast, the evolution of T_l in the model (see (20)) is rather simple (i.e., linear); the RMSE in the GDF is almost always smaller than that for the DF (red line versus black line in Figure 12(d)). For the low frequent component ϕ (see Figure 12(b)), the RMSEs of both the GDF and the DF decrease gradually, indicating that the effect of the data void region on the low frequent signal is small in the given time scale.

5.6. Impact of the GDF on the Forecast. From a more practical point of view, the role of the GDF should be judged from the model forecast. In this section, two forecast experiments without any observational constraint are integrated for 1 year, respectively, starting from the final analyzed states of the above two assimilation experiments (the GDF and the DF).

Figures 14(a)–14(d) show the forecasted time series of RMSEs of ψ, ϕ, T_o, and T_l for the GDF (black line) and the DF with $a, b = 0.5$ (red line). The GDF performs much better than the DF in 1 year's forecast lead time of all the state variables such as the high frequent component ψ and the low frequent component ϕ (black versus red curves in Figures 14(a) and 14(b)). It is interesting that the forecasted RMSE of ϕ still decreases inertially owing to the longer adjustment time of the low frequent signal, but whose trend becomes mildly with the increase of the forecasted lead time. For
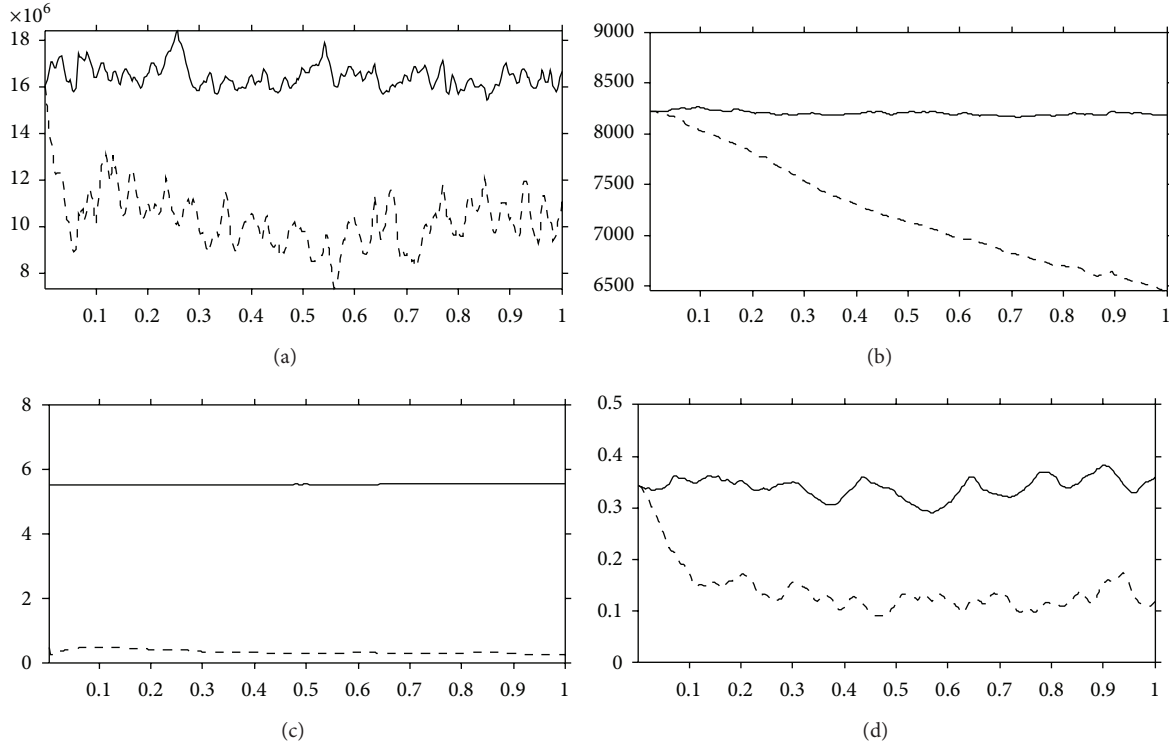
FIGURE 10: Time series of RMSEs of (a) atmospheric stream function, (b) oceanic stream function, (c) sea surface temperature, and (d) land surface temperature for CTRL (solid curve) and GDF (dashed curve).
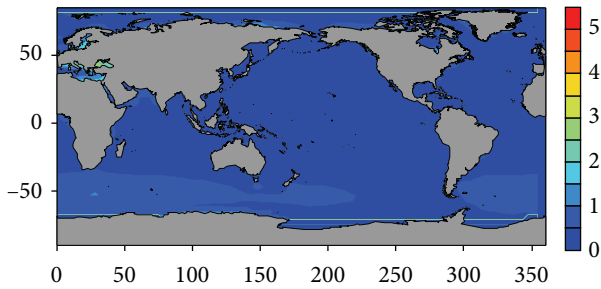

FIGURE 11: The spatial distribution of RMSE of sea surface temperature using GDF.

T_o, because of the absence of the observational constraint, the forecasted RMSE has an obvious positive trend (see Figure 14(c)), indicating that the forecasted state is gradually drifting away from the truth. Anyway, the GDF retains its superiority relative to the DF during the entire forecasted lead time. The forecasted RMSEs of T_l have similar patterns to those of T_o (see Figure 14(d)).

6. Conclusions and Discussions

In this study, the diffusion filter (DF) is introduced as a concrete implementation of the 3DVAR scheme. Similar to the recursive filter (RF), the outstanding issue of DF is its inefficiency in capturing the spatial multiscale information resolved by observations. Therefore, several spatial multiscale variational analysis schemes based on the DF are proposed to retrieve the spatial multiscale information from longwaves to shortwaves. As one of the spatial multiscale variational analysis schemes, the gradient diffusion filter (GDF) scheme is proposed and verified through a set of observing/assimilation system simulation experiments, where the "truth" field of the sea surface temperature is represented by a high nonlinear analytic function in a given sea region, and the observations are sampled randomly and uniformly in the whole domain. Results of the assimilation experiments indicate that the GDF has noticeable advantages over the standard RF and DF schemes, especially in the data void region. The GDF can retrieve the longwave information over the whole domain and the shortwave information over data-dense regions.

After that, a perfect twin experiment framework is designed to study the effect of the GDF on the state estimation based on an intermediate atmosphere-ocean-land coupled model. In this framework, the assimilation model is subject to "biased" initial fields from the "truth" model. The RMSE of the sea surface temperature can be reduced significantly through the observational constraint via the GDF. At the same time, the RMSEs of the other model components, such as the land surface temperature and the atmospheric and oceanic stream functions can also be mitigated by the dynamical constraint and the external constraint through the ocean-atmosphere-land coupled process. For simulating the real
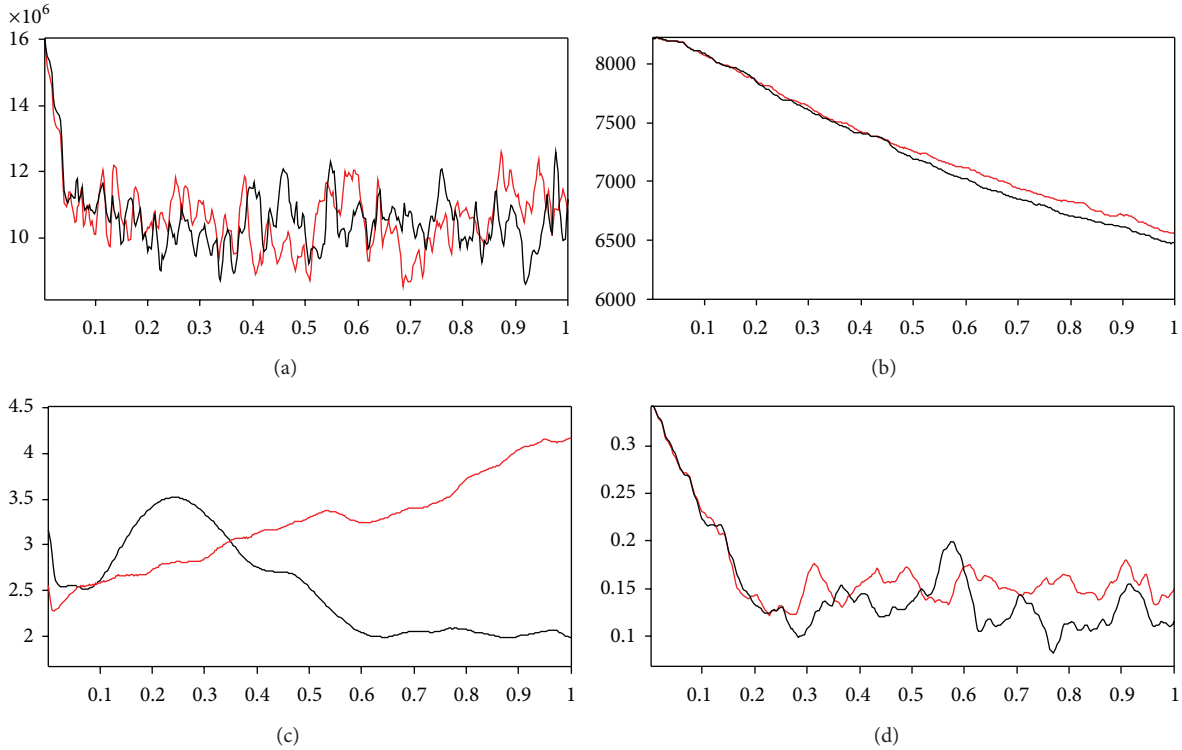
Figure 12: Time series of RMSEs of (a) atmospheric stream function, (b) oceanic stream function, (c) sea surface temperature, and (d) land surface temperature for GDF (black curve) and DF with $a, b = 0.5$ (red curve).
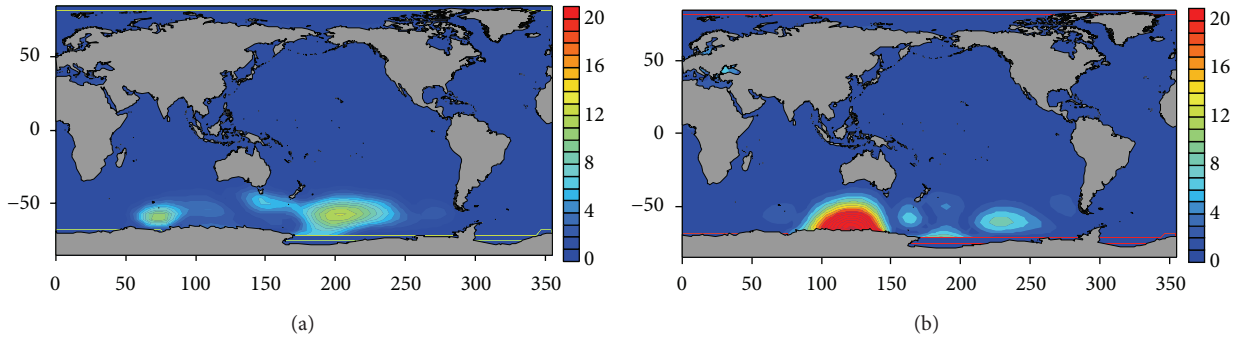

Figure 13: The spatial distributions of RMSEs of SST using (a) GDF and (b) DF with $a, b = 0.5$.

observational networks in the world ocean roughly, the observations locating in the Southern Ocean are removed to investigate the role of the GDF in retrieving the multi-scale information from observations. While the standard DF hardly removes the model bias in the data void region, the GDF may mitigate the model bias to some degree through extracting the multiscale information from the observations beyond the data void region. In addition, the higher forecast skill can also be obtained through the better initial state fields produced by the GDF.

It should be noted that the background term J_b is omitted in the above assimilation experiments. When high-density, accurate, resolvable information is available in observational datasets, it is much essential to extract the multiscale information from the observations with deterministic data assimilation approaches, as this study does. High-quality

background fields can be obtained firstly when deterministic data assimilation approaches are carried out. Next, the statistical data assimilation approaches, such as traditional 3DVar and 4Dvar, can be used to treat observations as random variables, in which J_b will be included to extract the information that cannot be resolved by the observation networks.

In spite of the promising results produced by the GDF in the intermediate climate model, much work is needed to explore the impact of the multiscale variational analysis schemes on the state estimation and forecast in real applications using general circulation models (GCMs). In addition, other spatial multiscale variational analysis schemes based on the DF, such as the adaptive diffusion filter (ADF) scheme, should also be studied to further improve the convergence speed and accuracy.
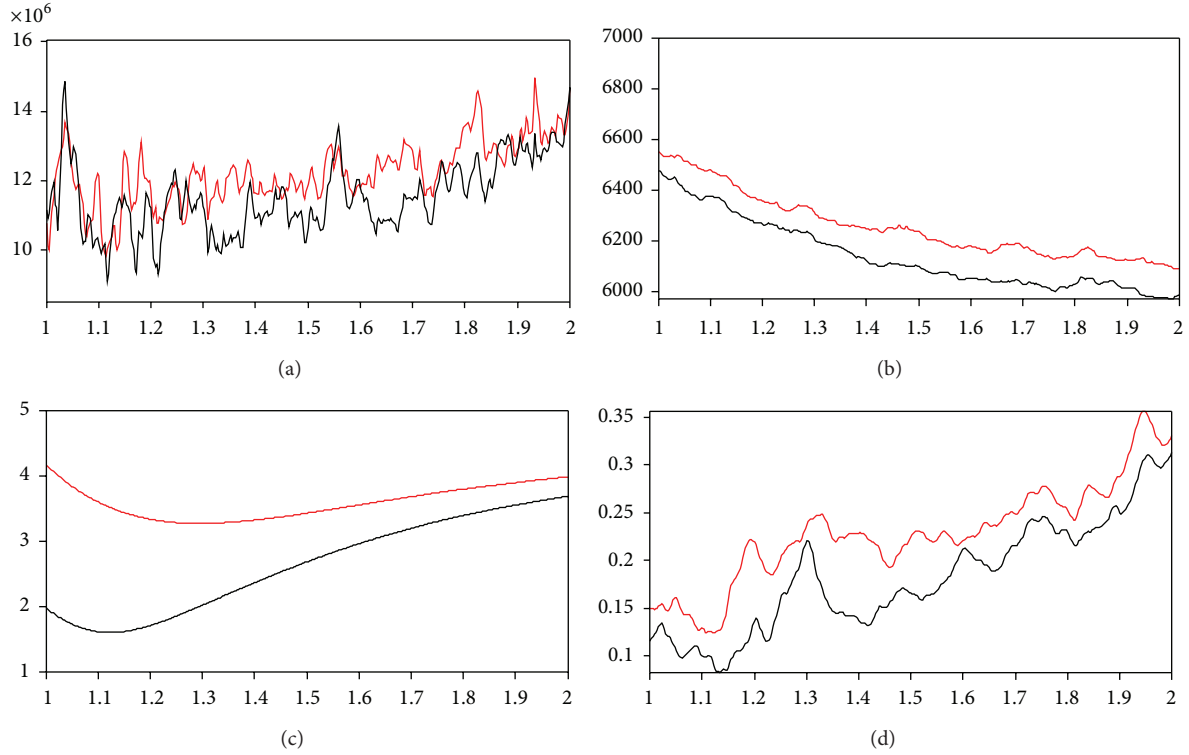
FIGURE 14: Forecasted time series of RMSEs of (a) atmospheric stream function, (b) oceanic stream function, (c) sea surface temperature, and (d) land surface temperature for GDF (black curve) and DF with $a, b = 0.5$ (red curve).

Appendix

Equivalence between the RF and DF Methods

Using ADI scheme, (9) can be discretized as follows:

$$
\frac{u_{i,j}^{n+1/2} - u_{i,j}^{n}}{\tau/2} - \Delta_x \left(a_{i,j} \overline{\Delta}_x u_{i,j}^{n+1/2} \right)
$$
$$
- \Delta_y \left(b_{i,j} \overline{\Delta}_y u_{i,j}^{n} \right) = 0, \tag{A.1}
$$
$$
n \in [0, N-1]; \; i \in [1, I-1]; \; j \in [1, J-1],
$$

$$
\frac{u_{i,j}^{n+1} - u_{i,j}^{n+1/2}}{\tau/2} - \Delta_x \left(a_{i,j} \overline{\Delta}_x u_{i,j}^{n+1/2} \right)
$$
$$
- \Delta_y \left(b_{i,j} \overline{\Delta}_y u_{i,j}^{n+1} \right) = 0, \tag{A.2}
$$
$$
n \in [0, N-1]; \; i \in [1, I-1]; \; j \in [1, J-1],
$$

$$
u_{i,j}^{0} = w_{i,j} \quad i \in [0, I]; \; j \in [0, J], \tag{A.3}
$$

$$
\Delta_x u_{0,j}^{n} = 0, \tag{A.4}
$$

$$
\overline{\Delta}_x u_{I,j}^{n} = 0, \tag{A.5}
$$

$$
\Delta_y u_{i,0}^{n} = 0, \tag{A.6}
$$

$$
\overline{\Delta}_y u_{i,J}^{n} = 0 \quad n \in [0, N], \tag{A.7}
$$

$$
\Delta_x u_{0,j}^{n+1/2} = 0, \tag{A.8}
$$

$$
\overline{\Delta}_x u_{I,j}^{n+1/2} = 0, \tag{A.9}
$$

$$
\Delta_y u_{i,0}^{n+1/2} = 0, \tag{A.10}
$$

$$
\overline{\Delta}_y u_{i,J}^{n+1/2} = 0 \quad n \in [0, N-1], \tag{A.11}
$$

where Δ_x, Δ_y, $\overline{\Delta}_x$, $\overline{\Delta}_y$ are forward and backward difference operator in x and y direction, respectively. τ is the time step, i, j and I, J are grid index and grid numbers in x and y direction, respectively, and n and N are the time index and the total time step numbers. The common tridiagonal matrix algorithm (TDMA) can be used to solve both (A.1) and (A.2). For example, the tridiagonal equation (A.2) in the jth row can be written as follows:

$$
Au = \gamma, \tag{A.12}
$$

where

$$
A = \begin{bmatrix} m_1 & l_1 & & \cdots & 0 \\ l_1 & m_2 & l_2 & & \vdots \\ & l_2 & m_3 & & \\ \vdots & & \ddots & \ddots & l_{I-2} \\ 0 & \cdots & & l_{I-2} & m_{I-1} \end{bmatrix},
$$

$$u = \begin{bmatrix} u_{1,j}^{n+1/2} \\ u_{2,j}^{n+1/2} \\ u_{3,j}^{n+1/2} \\ \vdots \\ u_{I-1,j}^{n+1/2} \end{bmatrix},$$

$$\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_{I-1} \end{bmatrix},$$

$$l_i = -\frac{\tau}{2h_1^2} a_{i+1,j}, \quad i \in [1, I-2],$$

$$m_i = \begin{cases} 1 - l_{i-1} - l_i, & i \in [2, I-2] \\ 1 - l_1, & i = 1 \\ 1 - l_{I-2}, & i = I-1, \end{cases}$$

$$\gamma_i = u_{i,j}^n + \frac{\tau}{2} \Delta_y \left(b_{i,j} \overline{\Delta}_y u_{i,j}^n \right), \quad i \in [1, I-1].$$

$$(\text{A.13})$$

It is easily testified that A is a positive definite and symmetrical matrix. Therefore, the Cholesky decomposition of A can be processed as follows:

$$A$$

$$= LL^T$$

$$= \begin{bmatrix} p_1 & & \cdots & 0 \\ q_1 & p_2 & & \vdots \\ & q_2 & p_3 & \\ \vdots & & \ddots & \ddots \\ 0 & \cdots & & q_{I-2} & p_{I-1} \end{bmatrix} \begin{bmatrix} p_1 & q_1 & \cdots & 0 \\ & p_2 & q_2 & \vdots \\ & & p_3 & \ddots \\ \vdots & & & \ddots & q_{I-2} \\ 0 & \cdots & & & p_{I-1} \end{bmatrix} \quad (\text{A.14})$$

which leads to

$$L\varphi = \gamma,$$
$$L^T u = \varphi,$$
$$(\text{A.15})$$

where

$$\varphi_{i+1} = -\frac{q_i}{p_{i+1}} \varphi_i + \frac{1}{p_{i+1}} \gamma_{i+1}, \quad i = 1, 2, \ldots, I-2,$$

$$\varphi_1 = \frac{1}{p_1} \gamma_1,$$

$$u_i = -\frac{q_i}{p_i} u_{i+1} + \frac{1}{p_i} \varphi_i, \quad i = I-2, I-1, \ldots, 1,$$

$$u_{I-1} = \frac{1}{p_{I-1}} \varphi_{I-1}.$$

$$(\text{A.16})$$

Specially, if $a_{i,j}$ is a constant, which is equivalent to an isotropic filter, we know that

$$p_i = p, \quad i \in [1, I-2],$$
$$q_i = q, \quad i \in [1, I-2], \quad (\text{A.17})$$
$$p + q = 1.$$

Set $\alpha = -q/p$, then (18) and (19) can be formulated as

$$\varphi_{i+1} = \alpha \varphi_i + (1 - \alpha) \gamma_{i+1}, \quad i = 1, 2, \ldots, I-2,$$
$$u_i = \alpha u_{i+1} + (1 - \alpha) \varphi_i, \quad i = I-2, I-1, \ldots, 1.$$
$$(\text{A.18})$$

Equation (A.18) has the same form as (6) in the RFM.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] A. C. Lorenc, "Analysis methods for numerical weather prediction," *Quarterly Journal of the Royal Meteorological Society*, vol. 112, no. 474, pp. 1177–1194, 1986.

[2] P. Courtier, "Variational methods," *Journal of the Meteorological Society of Japan*, vol. 75, no. 1, pp. 211–218, 1997.

[3] J. Derber and A. Rosati, "A global oceanic data assimilation system," *Journal of Physical Oceanography*, vol. 19, no. 9, pp. 1333–1347, 1989.

[4] P. Tandeo, E. Autret, B. Chapron, R. Fablet, and R. Garello, "SST spatial anisotropic covariances from METOP-AVHRR data," *Remote Sensing of Environment*, vol. 141, pp. 144–148, 2014.

[5] D. W. Behringer, M. Ji, and A. Leetmaa, "An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: the ocean data assimilation system," *Monthly Weather Review*, vol. 126, no. 4, pp. 1013–1021, 1998.

[6] S. Masina and N. Pinardi, "A global ocean temperature and altimeter data assimilation system for studies of climate variability," *Climate Dynamics*, vol. 17, no. 9, pp. 687–700, 2001.

[7] B. Huang, J. L. Kinter III, and P. S. Schopf, "Ocean data assimilation using intermittent analyses and continuous model error correction," *Advances in Atmospheric Sciences*, vol. 19, no. 6, pp. 965–992, 2002.

[8] A. C. Lorenc, "Optimal nonlinear objective analysis," *Quarterly Journal—Royal Meteorological Society*, vol. 114, no. 479, pp. 205–240, 1988.

[9] C. M. Hayden and R. J. Purser, "Recursive filter objective analysis of meteorological fields: applications to NESDIS operational processing," *Journal of Applied Meteorology*, vol. 34, no. 1, pp. 3–15, 1995.

[10] R. J. Purser, W.-S. Wu, D. F. Parrish, and N. M. Roberts, "Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: spatially homogeneous and isotropic Gaussian covariances," *Monthly Weather Review*, vol. 131, no. 8, pp. 1524–1535, 2003.

[11] A. Lorenc, "Iterative analysis using covariance functions and filters," *Quarterly Journal of the Royal Meteorological Society*, vol. 118, no. 505, pp. 569–591, 1992.

[12] X.-Y. Huang, "Variational analysis using spatial filters," *Monthly Weather Review*, vol. 128, no. 7, pp. 2588–2600, 2000.

[13] J. Gao, M. Xue, K. Brewster, and K. K. Droegemeier, "A three-dimensional variational data analysis method with recursive filter for Doppler radars," *Journal of Atmospheric and Oceanic Technology*, vol. 21, no. 3, pp. 457–469, 2004.

[14] P. C. Chu, S. K. Wells, S. D. Haeger, C. Szczechowski, and M. Carron, "Temporal and spatial scales of the Yellow Sea thermal variability," *Journal of Geophysical Research C: Oceans*, vol. 102, no. 3, pp. 5655–5667, 1997.

[15] P. C. Chu, W. Guihua, and Y. Chen, "Japan Sea thermohaline structure and circulation. Part III: autocorrelation functions," *Journal of Physical Oceanography*, vol. 32, no. 12, pp. 3596–3615, 2002.

[16] K.-A. Park and J. Y. Chung, "Spatial and temporal scale variations of sea surface temperature in the East Sea using NOAA/AVHRR data," *Journal of Oceanography*, vol. 55, no. 2, pp. 271–288, 1999.

[17] Y. Xie, S. E. Koch, J. A. McGinley, S. Albers, and N. Wang, "A sequential variational analysis approach for mesoscale data assimilation," in *Proceedings of the 21st Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction*, 15B.7, American Meteorological Society, Washington, DC, USA, 2005, http://ams.confex.com/ams/pdfpapers/93468.pdf.

[18] Z. He, Y. Xie, W. Li et al., "Application of the sequential three-dimensional variational method to assimilating SST in a global ocean model," *Journal of Atmospheric and Oceanic Technology*, vol. 25, no. 6, pp. 1018–1033, 2008.

[19] W. Li, Y. Xie, Z. He et al., "Application of the multigrid data assimilation scheme to the China seas' temperature forecast," *Journal of Atmospheric and Oceanic Technology*, vol. 25, no. 11, pp. 2106–2116, 2008.

[20] Y. Xie, S. E. Koch, J. A. McGinley et al., "A space–time multiscale analysis system: a sequential variational analysis approach," *Monthly Weather Review*, vol. 139, no. 4, pp. 1224–1240, 2011.

[21] A. Weaver and P. Courtier, "Correlation modeling on the sphere using a generalized diffusion equation," *Quarterly Journal of the Royal Meteorological Society*, vol. 122, pp. 535–561, 2001.

[22] D. Li, X. Zhang, H. L. Fu, L. Zhang, X. Wu, and G. Han, "A spatial multi-scale three-dimensional variational analysis based on recursive filter algorithm," *Journal of Atmospheric and Oceanic Technology*. In press.

[23] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1–3, pp. 503–528, 1989.

[24] X. Wu, S. Zhang, Z. Liu, A. Rosati, T. L. Delworth, and Y. Liu, "Impact of geographic-dependent parameter optimization on climate estimation and prediction: simulation with an intermediate coupled model," *Monthly Weather Review*, vol. 140, no. 12, pp. 3956–3971, 2012.

[25] X. Wu, S. Zhang, Z. Liu, A. Rosati, and T. L. Delworth, "A study of impact of the geographic dependence of observing system on parameter estimation with an intermediate coupled model," *Climate Dynamics*, vol. 40, no. 7-8, pp. 1789–1798, 2013.

[26] X. Zhang, S. Zhang, Z. Liu, X. Wu, and G. Han, "Parameter optimization in an intermediate coupled climate model with biased physics," *Journal of Climate*, vol. 28, no. 3, pp. 1227–1247, 2015.

[27] X. Wu, W. Li, G. Han, S. Zhang, and X. Wang, "A compensatory approach of the fixed localization in EnKF," *Monthly Weather Review*, vol. 142, no. 10, pp. 3713–3733, 2014.

[28] Z. Liu, "Interannual positive feedbacks in a simple extratropical air-sea coupling system," *Journal of the Atmospheric Sciences*, vol. 50, pp. 3022–3028, 1993.

[29] S. G. H. Philander, T. Yamagata, and R. C. Pacanowski, "Unstable air-sea interactions in the tropics," *Journal of the Atmospheric Sciences*, vol. 41, no. 4, pp. 604–613, 1984.

[30] R. Asselin, "Frequency filter for time integrations," *Monthly Weather Review*, vol. 100, no. 6, pp. 487–490, 1972.

[31] A. Robert, "The integration of a spectral model of the atmosphere by the implicit method," in *Proceedings of the WMO/IUGG Symposium on NWP*, pp. 19–24, Japan Meteorological Society, Tokyo, Japan, 1969.