

A Model-Based Observation-Thinning Scheme for the Assimilation of High-Resolution SST in the Shelf and Coastal Seas around China

XICHEN LI, JIANG ZHU, YIGUO XIAO, AND RUIWEN WANG

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

(Manuscript received 20 April 2009, in final form 25 December 2009)

ABSTRACT

The use of high-density remote sensing buoys and ship-based observations play an increasingly crucial role in the operational assimilation and forecast of oceans. With the recent release of several high-resolution observation datasets, such as the Global Ocean Data Assimilation Experiment (GODAE) high-resolution SST (GHRSSST) datasets, the development of observation-thinning schemes becomes important in the process of data assimilation because the huge quantity and dense spatial-temporal distributions of these datasets might make it expensive to assimilate the full dataset into ocean models or even decay the assimilation result. In this paper, an objective model simulation ensemble-based observation-thinning scheme is proposed and applied to a Chinese shelf-coastal seas eddy-resolving model. A successful thinning scheme should select a subset of observations yielding a small analysis error variance (AEV) while keeping the number of observations to as few as possible. In this study, the background error covariance (BEC) is estimated using the historical ensemble and then the subset of observations to minimize the AEV is selected, which is estimated from the Kalman theory. The authors used this method in the GHRSSST product to cover the shelf and coastal seas around China and then verified the result with an estimation function and assimilation-forecast systems.

1. Introduction

Recent observations, such as those from satellites, are characterized by their huge quantity and dense spatial distributions, which are often denser than the model grids. For example, the density of high-resolution SST datasets provided by the Global Ocean Data Assimilation Experiment (GODAE) high-resolution SST (GHRSSST) project exceeds $1/20^\circ$ every 6 h (Donlon et al. 2002, 2004, 2007; Donlon 2003). However, for many numerical ocean models it is unnecessary or even infeasible to assimilate such a huge dataset into numeric ocean models. The excessively high resolution of the observations occupies too much of the computational resources during the process of assimilation for some data assimilation schemes.

Data assimilation schemes can be formulated to solve the analysis in ensemble space (e.g., Testut et al. 2003; Sakov et al. 2009) Liu and Rabier (2002) established a relationship between the observation density and the resolution of the model grid with a theoretical model.

Their study shows that exorbitant density of observation may decrease the quality of analysis states if error correlations are neglected between different observational points. The disadvantages and restrictions of the high observation density mentioned earlier motivates us to establish and develop observation thinning methods and schemes to reduce the computational costs, as well as to avoid impairing analysis results in the assimilation process.

Several so-called super observation schemes have been proposed to thin densely distributed observations, and to help us confirm the representative error of the thinned observations. For example, Ochotta et al. (2005, 2007) proposed an estimation error analysis (EEA) scheme. In this method, a continuous estimation function (EF) is established in the model domain, where the value at any location is calculated by a weighted average of the observation values in the neighboring locations. The EEA method constructs a thinned dataset by iteratively removing observations from the full dataset and minimizing the degradation of the continuous EF in each step. In this way, the information of the full dataset can be maximally preserved in the subset. In another respect, Sakov and Oke (2008) proposed a method to estimate the representation error (RE) of the thinned observations. They assumed that the main source of the RE is unresolved

Corresponding author address: Jiang Zhu, ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China 100029.

E-mail: jzhu@mail.iap.ac.cn

processes and scales. The observations are averaged according to the resolution of the model grid, and the RE is considered as the deviation of the original data from the averaged field. Several other methods and principles were also established to help us determine the total number and concentration of observations in “optimal” assimilations, such as the method of relative entropy and Shannon entropy (Xu 2007) and the theoretical one-dimensional study of the interaction between model resolution and observation density (Liu and Rabier 2002). These studies supported that the subset of observation in an optimal assimilation should not only subject to containing the maximum information of the whole observation but also be associated with the model resolution as well as the covariance between model grids and their adjacent points. However, the factors of the model resolution and the assimilation scheme are rarely considered in current studies on developing practical high-resolution SST observation-thinning schemes.

Ensemble-based methods for optimal array design have been widely used in the last few years (e.g., Bishop et al. 2001; Tippett et al. 2003; Langland 2005; Khare and Anderson 2006; Oke and Sakov 2008). In this paper, an ensemble-based method is employed to thin the high-resolution SST, for the assimilation with an eddy-resolving ocean model of Chinese shelf/coastal seas (CSCS). The thinning scheme is verified based on Ochotta et al.’s (2005) EEA method and a high-resolution SST dataset.

This paper is organized as follows: a brief description of the method is presented in section 2; followed by the description of the model and observation dataset used in this study in section 3; next, the results and their verification are presented in section 4 and section 5; and finally, section 6 is the discussion and conclusions.

2. Method

a. Principles and procedure of the thinning scheme

In assimilation schemes, the state space of a model can be defined as a column vector called the state vector, written as \mathbf{x} , whose components depend on the choice of model’s discretization. The dimension n of \mathbf{x} is often large and can be 10^6 – 10^8 in realistic ocean models (e.g., 10^6 in this study). For a given data assimilation scheme we use a number of observed values that are gathered into an m -dimensional observation vector \mathbf{y} . For high-resolution observations such as the GHRSSST products m can be larger than n . Current data assimilation schemes require an *observation operator*, which relates the state vector to the observation vector. For conventional observations that are measurements of model variables, the observation operator is often defined as a linear interpolation operator (an $m \times n$ matrix) as the following:

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_m \end{pmatrix}. \quad (1)$$

In each of the components, \mathbf{h}_i is an n -dimensional row vector and relates to the location of the i th observed value in \mathbf{y} . For a given interpolation scheme the locations of the observations can be exclusively determined by \mathbf{H} .

For high-resolution SST observations, three steps are involved in the observation-thinning procedure. In the first step, the complete observation dataset (with resolution of 0.1°), whose observation operator is written as \mathbf{H}_c , was thinned into several regular coarse-grid subsets (such as 0.5° by 0.5° , 0.3° by 0.3° , etc.), whose observation operators are defined as \mathbf{H}_r . For each of these subsets of observations, the analysis error variance, denoted by AEV_r , is the trace of the analysis error covariance (AEC) matrix \mathbf{P}^a with the covariance updated equation:

$$\mathbf{P}_r^a = [\mathbf{I} - \mathbf{P}^b \mathbf{H}_r^T (\mathbf{H}_r^T \mathbf{P}^b \mathbf{H}_r^T + \mathbf{R}_r)^{-1} \mathbf{H}_r] \mathbf{P}^b. \quad (2)$$

Here, \mathbf{I} is the identity matrix, \mathbf{R}_r is the observational error covariance matrix of the subset, and the superscript T denotes matrix transposition. The \mathbf{P}^b is the covariance matrix constructed from the historical ensemble, as shown in Eq. (4), and \mathbf{R}_r is the observation covariance matrix. Because we can hardly estimate the measurement error of each single observation, we cannot accurately evaluate the observation covariance. In this study, the observation covariance matrix is defined as a diagonal matrix that can dramatically induce computational cost.

However, these regular grid subsets of observations are often overthinned and do not associate with the model resolution and the covariance between model grids and their adjacent points. Therefore, we need to add more observations to them to further reduce their AEVs in some optimal manner. In the second step, we added a group of optimal observational points iteratively into the coarse regular-grid subsets to form augmented observation subsets. The observation operator of the augmented observation subsets is defined as \mathbf{H}_a , and the analysis error variance of these augmented subsets is denoted as AEV_a . To reduce the calculation expenditure, we adopted a simplified scheme of Oke et al. (2008) and Oke and Sakov (2008, and the equation) that applied the scheme to estimate and improve the objective array observation system in the tropical Indian Ocean. In this scheme, we select only one optimal observation point during each of the iterations, with the augmented observation operator matrix \mathbf{H}_a defined as follows:

$$\mathbf{H}_a = \begin{pmatrix} \mathbf{h}_r \\ \mathbf{h}_{a,1} \\ \vdots \\ \mathbf{h}_{a,i} \end{pmatrix}. \quad (3)$$

Each $\mathbf{h}_{a,i}$ associates with a certain location of an additional observation. The location of each additional location is selected iteratively, following the principle that this additional observation location can maximally reduce the AEV_a in each of the iterations. In this procedure, the AEV_a is continuously updated to evaluate the augmented subsets. This simplified scheme will be further described in section 2c.

In the third step, we validated these augmented subsets by evaluating their capability of reserving the information of the complete datasets with the Fleet Numerical Meteorology and Oceanography Center (FNMOC) High-Resolution SST/Sea Ice Analysis for GHRSS (FSTIA) SST dataset. The validation method and procedure will be described in section 5 in detail. A proper augmented subset is then finally selected based on their validation results.

b. Estimating AEV_r of the regular-grid subsets

The large dimension of the state vector makes it impractical to explicitly manage the n by n background error covariance (BEC) matrix. Instead, we can store and manipulate the BEC matrix by establishing a relationship between the covariance matrix \mathbf{P} and a representative ensemble \mathbf{A} of the system state anomalies, $\mathbf{A}^{n \times k} = [\delta\mathbf{x}^{(1)}, \dots, \delta\mathbf{x}^{(k)}]$, where k is the sample size, $\delta\mathbf{x}^{(i)} = \mathbf{x}^{(i)} - \bar{\mathbf{x}}$ and the overbar denotes the ensemble average. Therefore, the error BEC matrix \mathbf{P} associated with the ensemble \mathbf{A} can be calculated by the function

$$\mathbf{P} = \frac{1}{m-1} \mathbf{A}\mathbf{A}^T. \quad (4)$$

In fact, the function (2) can be further simplified by processing each single observation point iteratively. Bishop et al. (2001) proposed a method to calculate the matrix \mathbf{P}^a [cited as Eq. (4)–Eq. (8)], by updating the background ensembles $\mathbf{A}^b \rightarrow \mathbf{A}^a$. In this method, the new ensembles \mathbf{A}^a with the single observational point assimilated were calculated primarily, and then the AEC matrix \mathbf{P}^a was updated via function (4). Ensembles \mathbf{A}^a can be obtained by the ensemble transform function:

$$\mathbf{A}^a = \mathbf{A}^b \mathbf{T} \quad \text{and} \quad (5)$$

$$\mathbf{T} = \left[\mathbf{I} + \frac{1}{m-1} (\mathbf{H}\mathbf{A}^b)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{A}^b \right]^{1/2}. \quad (6)$$

In Eq. (6), we have to calculate the inverse square root of a matrix in the ensemble space of dimension m . This calculation can be operated by using an eigenvalue decomposition of the matrix in (7). Because this matrix is symmetric, it can be decomposed as

$$\mathbf{I} + \frac{1}{m-1} (\mathbf{H}\mathbf{A}^b)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{A}^b = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad (7)$$

where \mathbf{U} is an orthonormal matrix ($\mathbf{U}\mathbf{U}^T = \mathbf{I}$) and $\mathbf{\Lambda}$ is a diagonal matrix. The inverse square root of the matrix can be calculated as

$$\mathbf{T} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^T. \quad (8)$$

Given a certain resolution, we can easily update the ensembles and then calculate the AEC_r matrix by the functions mentioned earlier [Eq. (4)]. The AEV_r was considered as an index, which can help us estimate the resolutions of different thinned grids.

c. Method of adding optimal observations and estimating of AEV_a

The observation operator of the augmented observation subsets \mathbf{H}_a can be determined by solving the following:

$$\begin{aligned} \mathbf{H}_a &= \arg \min \text{trace}(\mathbf{P}^a) \\ &= \arg \min \text{trace}\{[\mathbf{I} - \mathbf{P}^b \mathbf{H}^T (\mathbf{H}^T \mathbf{P}^b \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H}]\mathbf{P}^b\}. \end{aligned} \quad (9)$$

This augmented operator \mathbf{H}_a can minimize the trace of the AEC matrix \mathbf{P}^a , which indicates the sum of the AEV_a throughout the model domain. According to the relationship that $\text{trace}(\mathbf{A}\mathbf{B}) = \text{trace}(\mathbf{B}\mathbf{A})$, the solution of Eq. (9) can be written as (Oke and Sakov 2008)

$$\mathbf{H}_a = \arg \max \text{trace} \frac{\mathbf{H}_a \mathbf{P} (\mathbf{H}_a \mathbf{P})^T}{\mathbf{H}_a \mathbf{P} \mathbf{H}_a^T + \mathbf{R}_a}. \quad (10)$$

Here, \mathbf{R}_a indicates the observation error covariance matrix of the augmented observation subsets. Equation (10) may include multiple observations. In a simpler case, if only one additional single observation location is selected in each step, the solution can be simplified to the following functions:

$$\mathbf{h}_{a,i} = \arg \max(E) \quad \text{and} \quad (11)$$

$$E = \frac{1}{\mathbf{h}_{a,i} \mathbf{P} \mathbf{h}_{a,i}^T + \mathbf{r}(i)} \sum_{j=1}^n (\mathbf{h}_{a,i} \mathbf{P}) (\mathbf{h}_{a,i} \mathbf{P})^T. \quad (12)$$

In this solution, both of the matrix $\mathbf{H}_a \mathbf{P} \mathbf{H}_a^T$ and the observation covariance matrix \mathbf{R} becomes a scalar: $\mathbf{H}_a \mathbf{P} \mathbf{H}_a^T = \mathbf{h}_{a,i} \mathbf{P} \mathbf{h}_{a,i}^T$ and $\mathbf{R} = \mathbf{r}(i)$, which makes it unnecessary for us

to calculate the inverse of the matrix $\mathbf{H}_a \mathbf{P} \mathbf{H}_a^T + \mathbf{R}$. This function helps us to estimate which point can maximally reduce the trace of the AEC matrix; in each step, all available observation locations are estimated by calculating their evaluation function E according to Eq. (12) and the optimal interpolation operator $\mathbf{h}_{a,i}$ that maximizes the function E will be selected to add into the augmented operator \mathbf{H}_a . Besides, the scheme mentioned in section 2a was employed to update the ensembles every step after a certain optimal observation point was confirmed.

In this study, the optimal scheme is only optimal if the following assumptions hold. First, this scheme assumes that the background ensemble is unbiased, just like in most of the ensemble assimilation schemes. Second, we assume that the temporal variable can be represented by a group of static ensembles and construct the background covariance matrix with the historical ensemble (seasonal cycle removed, without localization). This is a strong assumption that will lead to inaccurate estimation of the covariance matrix. Nonetheless, the covariance matrix constructed by the stationary historical ensemble is currently widely used in data assimilation. Third, we do not introduce localization to this thinning scheme but use a global strategy because localization dramatically changes the form of Eq. (6) in the thinning process, which makes it impossible to update the ensemble. This may influence the result if we use the subset in some localized assimilation system. We plan to solve this problem by conducting a simplified localization scheme in the thinning process. Finally, in this scheme, we employed an iterative process to obtain the most suitable observation one by one. During this process, the reduction of AEV tends to be overestimated all over the modal. It is the distribution and structure of AEV, rather than the absolute value of the variance that helps us to better define the positions of additional observations. These assumptions only approximate to the real world and result in a much easier implementation. To remind readers of these assumptions, we use the term optimal through out this paper.

d. Computational cost of this method

In this observation thinning scheme, the system selects additional observations and adds them into the subset through an iterative procedure. Therefore, it is necessary for us to discuss its computational cost.

Here, Let M denote the dimension of the observation, m the number of observation in the subset we selected, n the dimension of background vector, and k the number of the ensemble, which equals 120. For each cycle there are two independent steps. First, we select the “most suitable” observation by calculating the function E for every M observation and obtaining the maximum value. Here, E is defined as (12)

$$E = \frac{1}{\mathbf{h}_{a,i} \mathbf{P} \mathbf{h}_{a,i}^T + \mathbf{r}(i)} \sum_{j=1}^k (\mathbf{h}_{a,i} \mathbf{P})(\mathbf{h}_{a,i} \mathbf{P})^T$$

$$= \frac{1}{\mathbf{h}_{a,i} \mathbf{P} \mathbf{h}_{a,i}^T + \mathbf{r}(i)} \sum_{j=1}^k (\mathbf{h}_{a,i} \mathbf{A})(\mathbf{A}^T \mathbf{A})(\mathbf{h}_{a,i} \mathbf{A})^T.$$

Matrix $(\mathbf{A}^T \mathbf{A})$ is the first calculated and remains constant during the whole cycle, whose computational cost is on the order of $k^2 \times n$. Then, we calculate E for each observation using the calculated matrix $(\mathbf{A}^T \mathbf{A})$ and the cost of this process is approximately $k \times m$. As a result, the computational cost of the first step is $\sim k \times (k \times n + m)$. Second, we update the ensemble according to Eq. (7) (cost $\sim k^2 \times n + k^3$), Eq. (8) (cost $\sim k^3$), and Eq. (5) (cost $\sim k^2 \times n$), successively. The cost of each cycle (each additional observation) can be easily calculate as follows: $k \times (k \times n + m) + k^2 \times n + k^3 + k^3 + k^2 \times n$. Note that in most forecast systems, $k \times n \gg M$ and $k^2 \times n \gg k^3$, we can conclude that the cost of each cycle $\sim k^2 \times n$, and the cost of the whole process (adding m observations) $\sim k^2 \times n \times m$. This value can be primarily attributed to the dimension of the background, the ensemble size, and the number of additional observations (rather than the total number of original observations dataset). In this study, the ensemble size is 120 and the dimension of model vector is $\sim 10^6$. It costs about 2–3 s for each additional observation and the whole procedure costs several hours. Considering the computational cost can be primarily attributed to matrix multiplication (rather than matrix inverting), the time of computation can be easily reduced by parallel computation.

3. Model and datasets

a. Ocean model

A Chinese shelf/coastal seas (CSCS) model based on a three-dimensional Hybrid Coordinate Ocean Model (HYCOM; Bleck 2002; Chassignet et al. 2003, 2007) is used to provide a simulation as realistic as possible. A curvilinear horizontal grid is utilized with an average spatial resolution of about 13 km. There are 22 layers in the vertical coordinate. Using the bottom topography obtained from the 2-minute gridded elevations/bathymetry for the world (ETOPO2; $1/30^\circ$), the model domain includes the whole CSCS and part of the west Pacific Ocean. Figure 1 concisely displays the model domain and topography. The model was initialized with the *World Ocean Atlas* (Boyer et al. 2002) dataset (*WOA01*) and was spunup for 5 years. Then, the model was forced by the Comprehensive Ocean–Atmosphere Dataset (COADS) climatological cloud amount and radiation dataset and the European Centre for Medium-Range Weather Forecasts

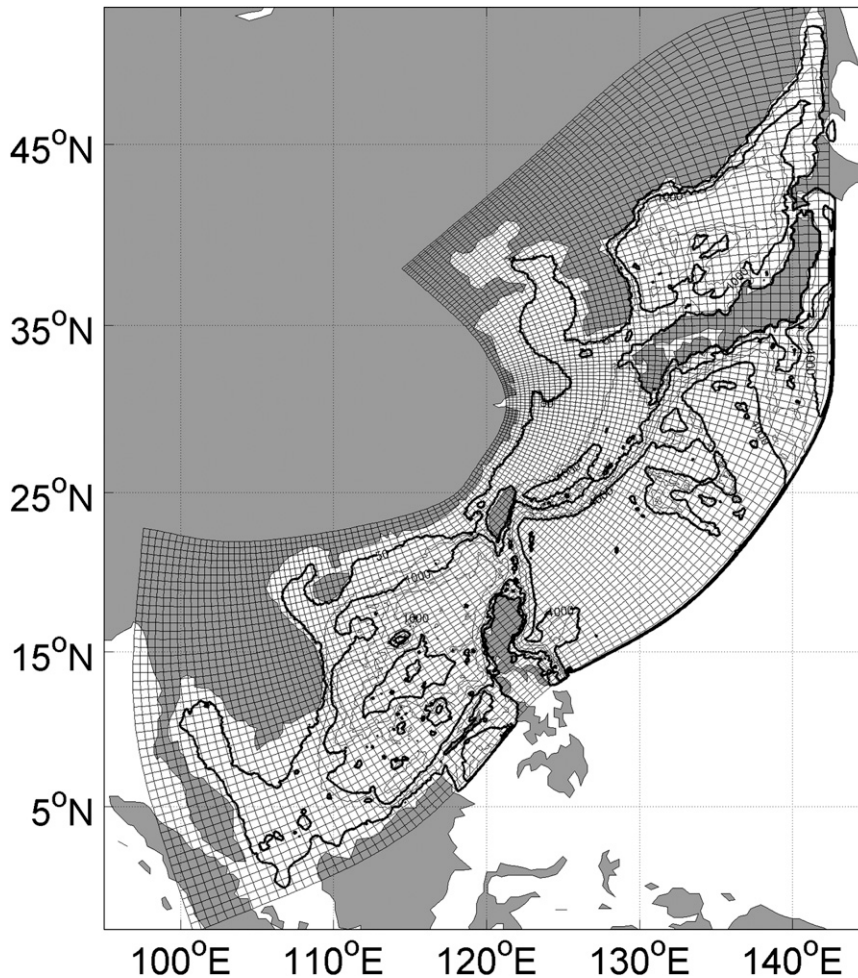


FIG. 1. The model's curvilinear grid and topography contour. One-fifth of the horizontal grid point is shown. The 50-, 1000-, and 4000-m isobaths are shown as thicker dark lines, while the 200-, 2000-, and 3000-m isobaths are denoted by gray contours.

(ECMWF) 6-hourly reanalysis dataset (Uppala et al. 2005) from 1997 to 2006. And we use a one-way nesting to an India–Pacific domain HYCOM simulation ($1/4^\circ$ resolution; Yan et al. 2007) as a sponge boundary condition. The surface temperature and salinity are relaxed to the climate on a time scale of 100 days.

The seasonal circulation and dynamic processes of the CSCS are mainly controlled by the monsoonal wind force and the impact of the western boundary currents. The mean seasonal sea surface temperature (SST) and surface velocity vectors of the two monsoon seasons in the CSCS model are illustrated in Fig. 2. The temperature front along the Kuroshio mainstream is well reproduced in this model. In winter, the SST distribution of the South China Sea (SCS) is characterized by a bifrontal structure (Chu et al. 2002), where distribution corresponds to the current along the Chinese coastline and the east of Vietnam. This pattern is well illustrated by solid lines (23° and 26.5° C

isothermals) in the control run SST panels of Fig. 2. Meanwhile, the circulation pattern was successfully simulated and reproduced in most regions, characterized by the strong Kuroshio in the Northwest Pacific and several gyres in SCS. The axis of the Kuroshio and its extension are indicated with labels A and B, respectively. The winter circulation in SCS is characterized by two cyclonic eddies (E and F in Fig. 2a) located to the west of the Philippines and to the southeast of the southern Vietnam shore, respectively (Qu et al. 2000). In summer, the Vietnam offshore stream (marked by C in Fig. 2b) and the anticyclonic eddy (D in Fig. 2b) located to the southeast of Hainan, China, are clearly visible (Gan et al. 2006). However, the cyclonic eddy to the west of the Philippines observed in summer is too weak to be identified in this model.

To calculate the BEC, a historical ensemble was constructed quasirandomly from the 10-yr model run. We

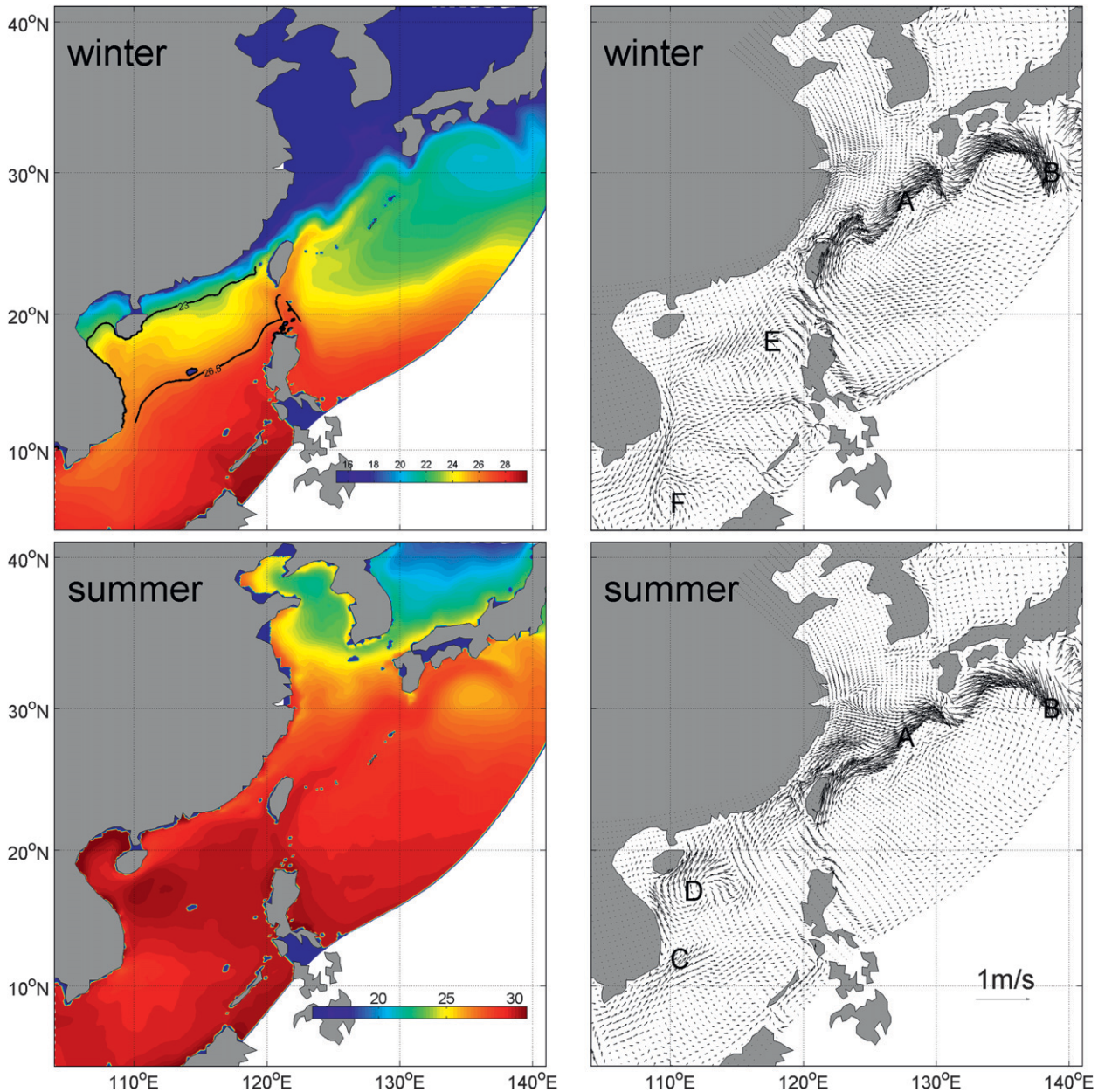


FIG. 2. (left) Mean seasonal SST and (right) surface velocity vectors (m s^{-1}) in (top) summer and (bottom) winter. Major jets and eddies are labeled.

selected a sample from each month during the 10 yr and constructed a 120 ensemble, although there is no fixed rule in selecting the “date” of the sample. The spread of these ensembles is illustrated in Fig. 3. The regions with large spread are mainly concentrated to the Northwest of the Kuroshio axis and along the continental coast. In addition, the monthly mean component was eliminated from these ensembles to construct a representative ensemble associated with the intraseasonal variability of the model SST fields.

b. High-resolution SST dataset

GHRSSST-PP provides several SST products based on satellite remote sensing observations (Donlon et al. 2002, 2004, 2007; Donlon 2003). In this study, we used FSTIA provided by U.S. GODAE. The FSTIA dataset was available from October 2005 until now, with the spatial resolution of up to 10 km and the temporal resolution of 6 h. The FSTIA dataset, along with several other GHRSSST products, were assessed by Xie et al. (2008) for

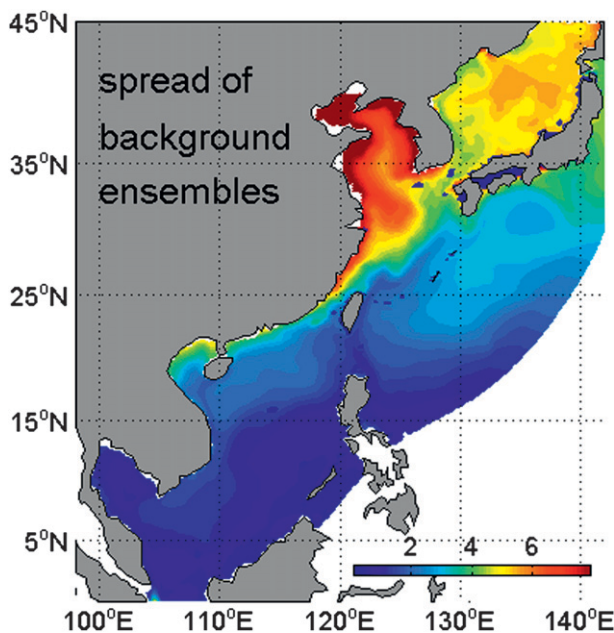


FIG. 3. Spread (variance) of model ensembles taken randomly from a 10-yr model simulation run (unit: $^{\circ}\text{C}$). It shows patterns of large-mesoscale and interseasonal–interannual variability of the model SST fields and is used as a proxy of the SST background error variance in this study.

the shelf–coastal seas around China using drifter observations and ship reports. The FSTIA dataset have good quality for the studied region.

In this paper, the observational error of the FSTIA dataset is assumed uncorrelated and the covariance matrix of observation is defined as a diagonal matrix, in that it is hard to estimate the covariance between the positions of different observations. We estimate the observation error variance according to Xie et al. (2008), which evaluated the FSTIA observation with float observations and other satellite-based datasets and indicated that an RMSE ranging from 0.3° to 1.2° exists in the FSTIA dataset in different regions over and around the model domain. The variance was also increased with an estimated representative error all over the model domain (about 0.3°).

4. Application to SST thinning in the coastal and shelf seas around China

Different resolutions of coarse-grid subsets of FSTIA are estimated in section 4a to find out an approximately appropriate resolution for further improvement; and in section 4b, the subsets of different original resolutions were concentrated in several special regions following the principle of maximally reducing the sum of AEV iteratively. Considering the multi-initial resolutions of this

scheme and the complex distributions of these concentrated subsets, this section mainly focuses on the issue of what an adaptive number of additional optimal observations is and how to define a well-restricted AEV field.

In the analysis procedure, it is found that the results of our observation-thinning performances are significantly impacted by two factors: the background variances of the points and the covariance (or correlation) between different points. It is supported that in some circumstances the latter is as important as or even more important than the former in terms of determining the analysis variance of certain positions. That is because the influence range of an observation is determined by the correlation length scale (between the observation site and other model grids) in the background error covariance in data assimilation schemes. The larger the length scale is the larger impact/weight the observation has. We further explain this phenomenon in mathematical terms as follows.

Consider the classic ensemble assimilation equation $\mathbf{X}_a = \mathbf{X}_b + \mathbf{A}\mathbf{A}^T\mathbf{H}^T(\mathbf{H}\mathbf{A}\mathbf{A}^T\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{D}$, where \mathbf{X}_a is analysis vector, \mathbf{X}_b is the first-guess field, \mathbf{A} is the background ensemble, \mathbf{H} is the observational operator, \mathbf{R} is the observation covariance matrix, and \mathbf{D} is the innovation. The matrix $\mathbf{C} = \mathbf{A}\mathbf{A}^T\mathbf{H}^T$ defines the covariance between model grids and observation positions. If $\mathbf{C}_{i,j}$ is large, the j th observation has a large weight in the increment of the i th element of the analysis vector. This process draws the analysis field closer to observation. For a single model grid, if it has high covariance with a large area, a comparatively low observation density can insure that there are enough observations in this area to restrict the analysis value of this model grid. In contrast, if the high-covariance area of a model grid is small, we need to increase the observation density in this area to sufficiently match the analysis field to observation. That is why the observation density is not only associated with the background variance but also results from the covariance. (The matrix $\mathbf{H}\mathbf{A}\mathbf{A}^T\mathbf{H}^T$ also contributes to these processes, but it is too complicated to explain mathematically because of the inversion.

As a result, a point will be influenced and restricted by many more observations under the condition of large-scale high correlation and vice versa. This issue will be addressed when it is associated with the model result in the following section (4a) to help us explain some physical processes and phenomena.

a. AEV_r of four regular coarse-grid subsets

Using the 120 model ensembles and method described in section 2a, the AEV_r of four resolutions of coarse-grid subsets with resolution of $0.2^{\circ} \times 0.2^{\circ}$, $0.3^{\circ} \times 0.3^{\circ}$, $0.4^{\circ} \times 0.4^{\circ}$, and $0.5^{\circ} \times 0.5^{\circ}$, respectively, were estimated and displayed in Fig. 4. For the $0.2^{\circ} \times 0.2^{\circ}$ and $0.3^{\circ} \times 0.3^{\circ}$

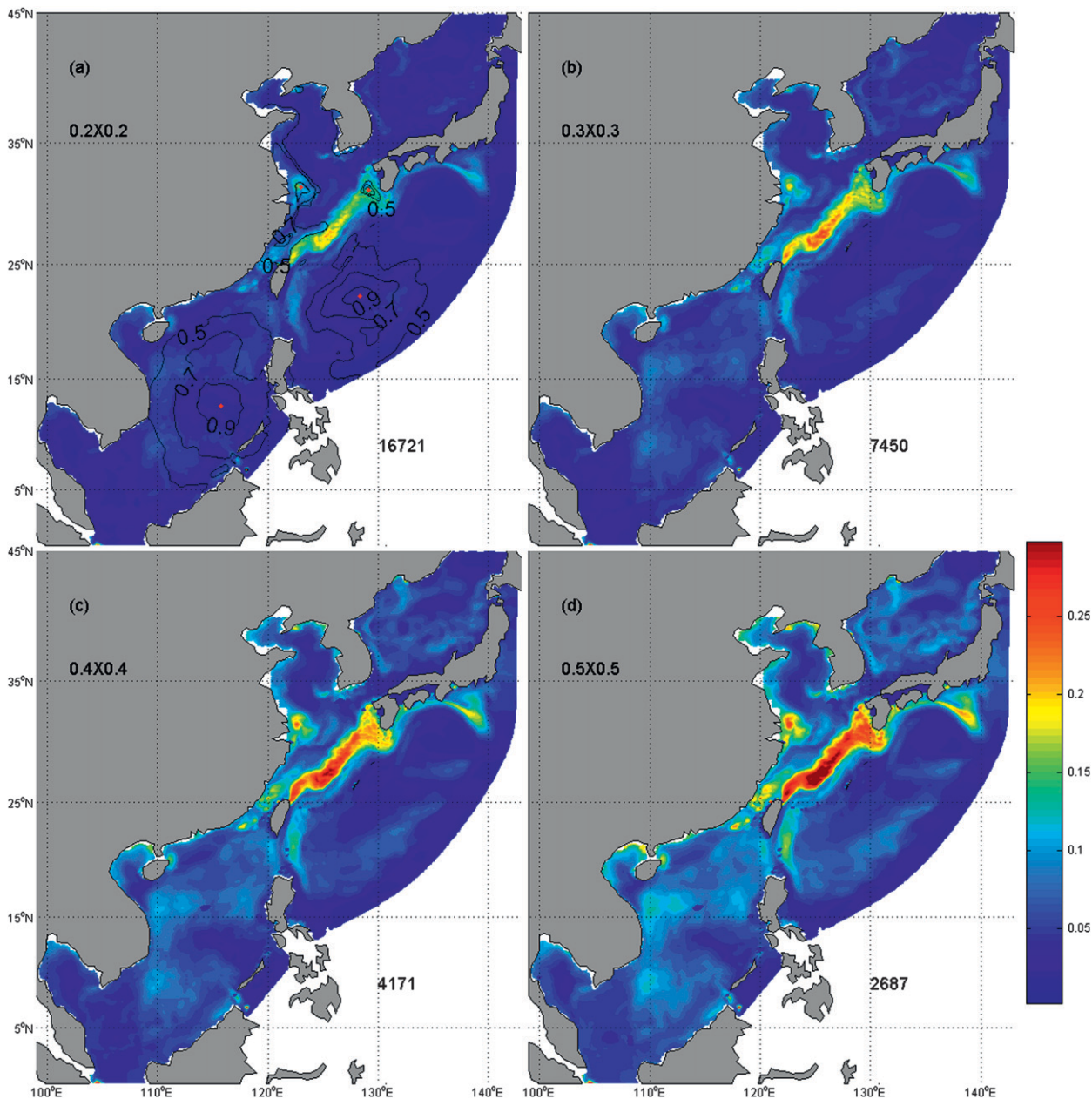


FIG. 4. AEVs (equal to the diagonals of \mathbf{P}^d) of four coarse-grid subsets. The coarse-grid resolutions are indicated in panel (a), with the total number in the networks shown in panel (d). (a) The correlation distributions around the four red points are indicated with a solid contour.

subsets, the AEV_r values are well below 0.1°C throughout the model domain; except for the mainstream of the Kuroshio and the estuary of Changjiang River, where the maximum variance extends to 0.25°C (it is necessary to mention that these AEVs are overly reduced by the iterative procedure. We can hardly expect that the analysis result will also be reduced to this level with reasonable assimilation parameters. However, the assimilation field has the same pattern with the updated AEV, as illustrated

in section 5b. In sharp contrast to this phenomenon, according to the background spread illustrated in Fig. 3, the Kuroshio is not characterized by the most significant background variances. The Kuroshio transports warm water northeastward from the west Pacific warm pool and constructs a sharp front along its mainstream. The high temperature gradient along with the strong current reduces the covariance around the mainstream and restricts the correlated regions of high covariance within

a small region (according to the correlation distributions of the red points illustrated in Fig. 4a). As a result, the observations around the Kuroshio are given with tiny weights and can hardly contribute to reduce the error variances along the Kuroshio. The situation is similar for the Changjiang River's outflow. Because of the complicated topography and huge amounts of freshwater inflow in this area, the characteristics of its water mass are dramatically distinguished from the surrounding water masses. The correlation between its freshwater and that of the adjacent locations is even negligible; therefore, the state vector is rarely impacted by the surrounding observations. The correlations around four points are illustrated in Fig. 4a, in which the spatial scale of correlated regions around SCS points and the west Pacific points extends to 500 km. In contrast, the correlated regions around the other two points (located in the freshwater outflow and the Kuroshio mainstream, respectively) are restricted within a small range.

For the $0.4^\circ \times 0.4^\circ$ and $0.5^\circ \times 0.5^\circ$ subsets, the AEV_r in SCS exceeds 0.1°C , whereas a part of the Kuroshio mainstream is characterized by a high variance of more than 0.4°C . However, for several other regions over the model domain, a horizontal resolution of 0.5° is sufficient enough to reduce the uncertainty, such as in the Northwestern Pacific Ocean area to the southeast of the Kuroshio mainstream in this model. The subset with $0.5^\circ \times 0.5^\circ$ resolution is a little rough to sufficiently restrict the AEV_r all over the model domain. However, the analysis variances have been successfully reduced to around 0.1°C in most regions and the gaps with large AEV_r can be filled by some optimally selected observations that will be presented next.

b. Locations and AEV_a of augmented subsets

In this section, we discuss adding some additional locations to each of the $0.3^\circ \times 0.3^\circ$, $0.4^\circ \times 0.4^\circ$, and $0.5^\circ \times 0.5^\circ$ subsets, respectively. As described in section 2c, these additional optimal locations are determined one by one iteratively. The space mean and the maximum of AEV_a s were calculated in each iteration and plotted in Fig. 5, with the x label indicating the total number of the observation locations (including those in each coarse-resolution subset). The spatial mean of AEV_a is significantly reduced by additional optimal observations, decreasing from 0.1°C and approaching 0.05°C (these AEV_a s are also overly reduced by iteratively strategy), whereas the maximum is reduced from 0.4°C to nearly 0.25°C . The reduction of these two indices slows down after the number of additional observation locations exceeds 4000. According to these curves, the augmented subsets from the $0.5^\circ \times 0.5^\circ$ subsets are obviously superior compared to the other two subsets (i.e., $0.3^\circ \times 0.3^\circ$ and $0.4^\circ \times 0.4^\circ$) when added

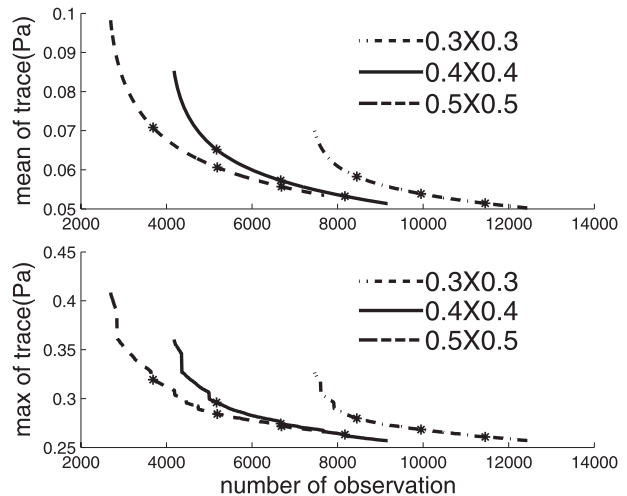


FIG. 5. Decreasing curves of AEV s associated with the increase of additional optimal observations of the four augmented coarse-grid subsets. (top) The x label is the total number of observation, and (bottom) the y labels indicate the space mean and the maximum of variances (equal to the mean of \mathbf{P}^a trace and the maximum of \mathbf{P}^a diagonal). Three points marked with an asterisk in each curve represent +1000, +2500, and +4000 additional observations, respectively.

with the same number of points. The asterisks in each curve represent +1000, +2500, and +4000 observation points, respectively.

The locations of these augmented subsets are illustrated in Fig. 6. It is clearly seen that additional observation locations are concentrated in several special areas, which can be associated with certain physical processes. With 1000 observation locations added, the additional observation locations are mainly concentrated along the Kuroshio mainstream and around the outflow of the Changjiang River (marked A and B in Fig. 6a). With the added locations reaching 2500, the density in the Taiwan Strait and the Kuroshio extension are also dramatically increased. This can be explained by their complicated current structures and the high variability of SST states (marked C and D in Fig. 6b). Finally, when further increasing the additional locations to 4000, three centers of high location concentration appear, which can also be related to specific physical processes. The center to the west of the Northern Philippines (marked E in Fig. 6c) corresponds to the cyclonic warm eddy in the winter monsoon season, described in section 3 and marked with label E in Fig. 2. The second center (marked F in Fig. 6c), located to the east of the Vietnam shore and is associated with the warm eddy and the Vietnam offshore jet in summer. And the center marked with "G" can be associated with the winter anticyclonic cold eddy located to the southeast of Vietnam. These eddies and jets disturb the structure of circulation and increase the variability

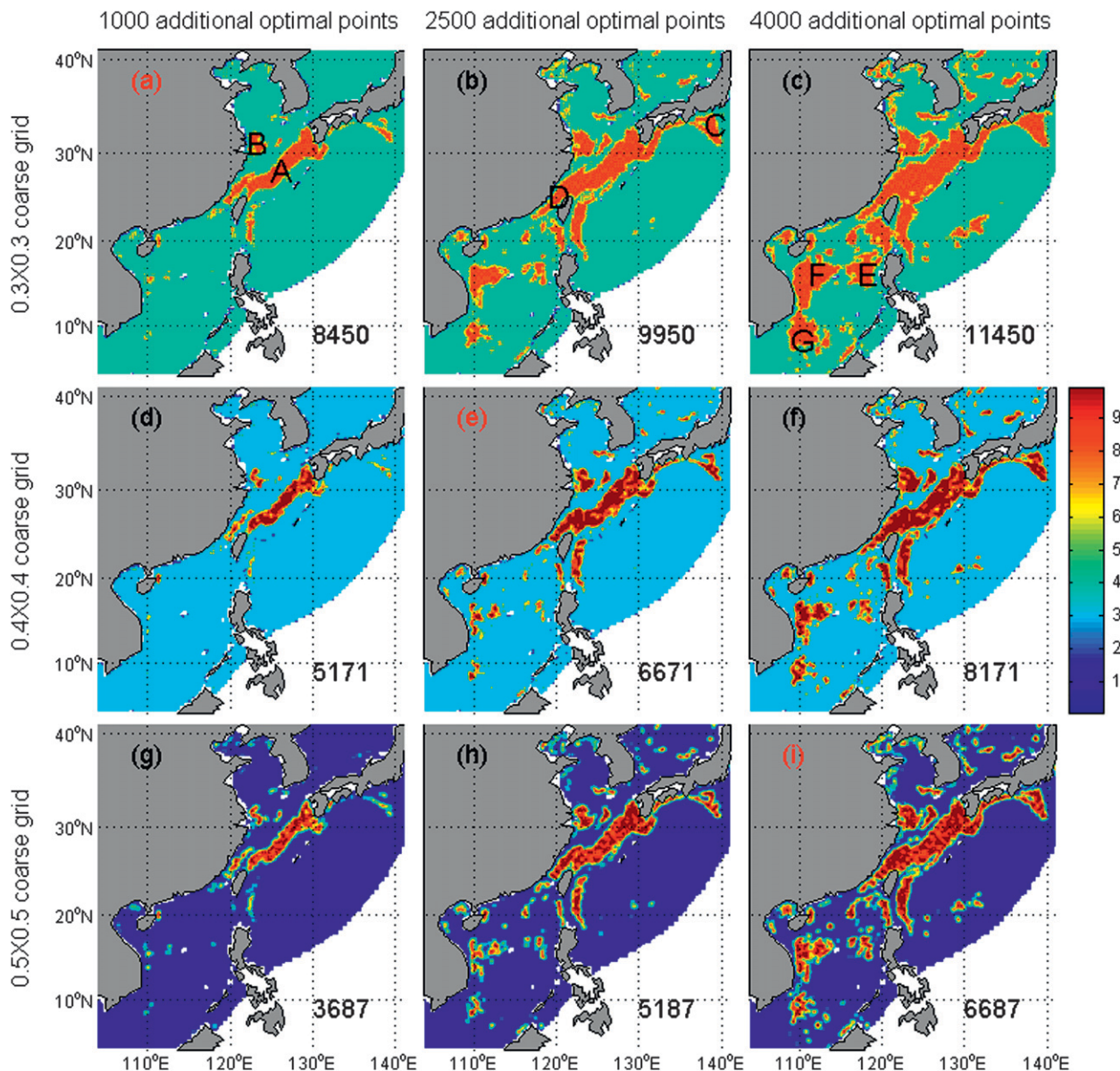


FIG. 6. The density of observation locations of the three augmented subsets with (left to right) different numbers (1000, 2500, and 4000) of additional optimally selected locations. The color bar indicates the number of locations within the $30 \text{ km} \times 30 \text{ km}$ boxes. (top to bottom) The augmented subsets of $0.3^\circ \times 0.3^\circ$, $0.4^\circ \times 0.4^\circ$, and $0.5^\circ \times 0.5^\circ$. The total number of observations in each scheme is indicated in the right corner, respectively.

of sea surface temperature, thus, requiring denser observations to reduce the AEV_a at these regions. High observation location concentration is also found at several other discrete areas such as the Japan Sea (JS).

The distributions of AEV_a of these augmented subsets are compared and illustrated in Fig. 7. We choose three augmented observation subsets: 1) a $0.3^\circ \times 0.3^\circ$ subset plus the first 1000 optimal locations (Fig. 7a); 2) a $0.4^\circ \times 0.4^\circ$ subset plus the first 2500 optimal locations (Fig. 7e); and 3) a $0.5^\circ \times 0.5^\circ$ subset plus the first 4000

optimal locations (Fig. 7i), with the total number of locations 8459, 6671, and 6687, respectively. They produce similar efficiency, although the sizes of these subsets are different. In these augmented subsets, the AEV_a of the Kuroshio and Changjiang River outflow are successfully restricted below 0.2°C and around 0.05°C for most of the other areas. Among these three augmented subsets, the thinning scheme of $0.5^\circ \times 0.5^\circ + 4000$ is superior in both the analysis variances and number of observations; therefore, it is considered as the optimal thinning scheme

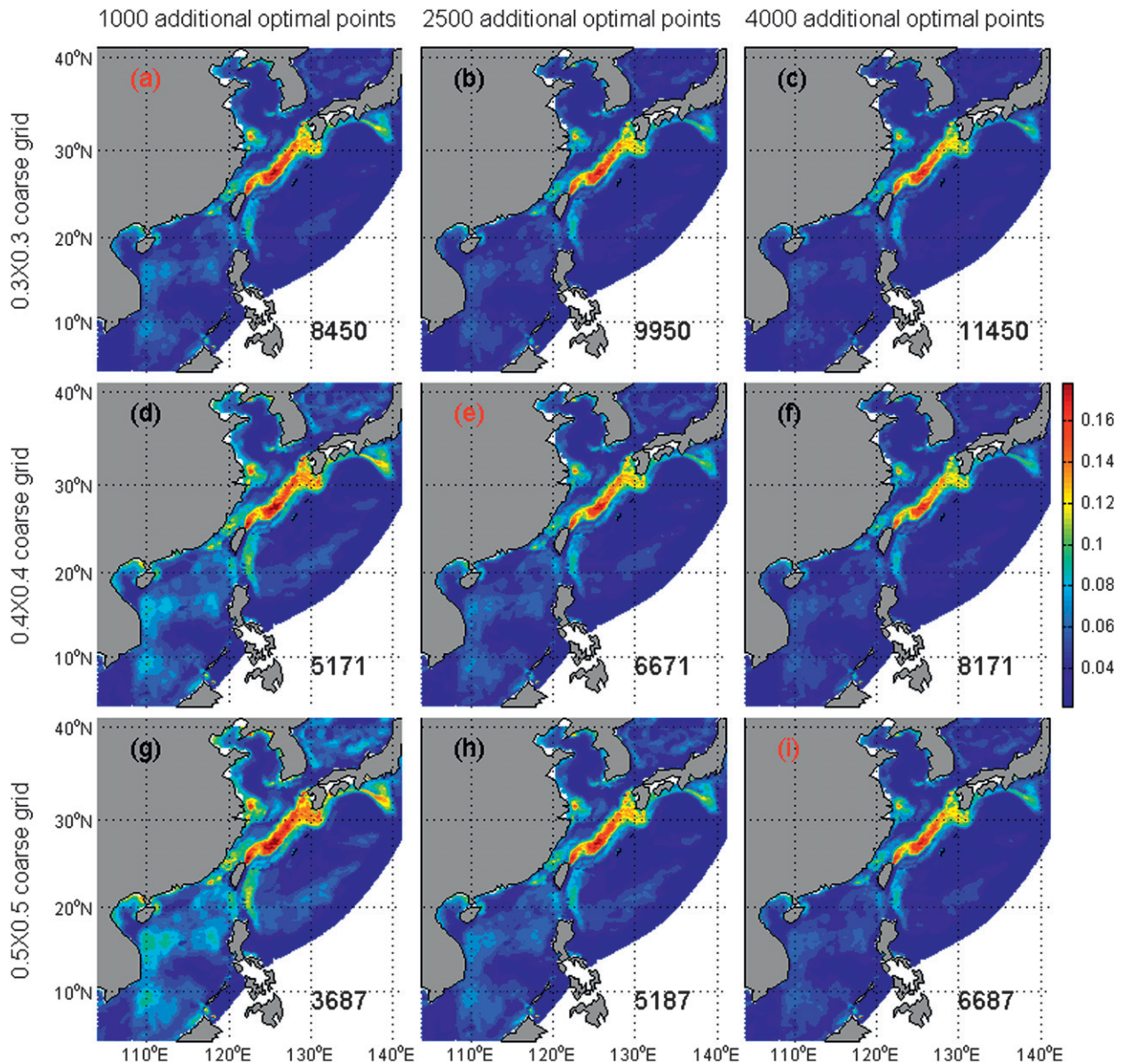


FIG. 7. As in Fig. 6, but for the AEVs of augmented subsets ($^{\circ}\text{C}$).

in this study. Through comparison between Figs. 7a and 7b, or between Figs. 7e and 7f, it is clearly seen that the situation is only slightly improved by further increasing the number of observation locations.

5. Verifications of thinning results

a. Verification with the EEA method

As previously mentioned in the introduction, the capability of a thinned subset to preserve the information of the whole dataset also plays a crucial role in assessing an observation-thinning scheme. However, this capability

is associated with neither of the two steps described in section 2. Therefore, it is necessary for us to verify the results of the previous sections using the real, full-resolution dataset. One year (2006) of the daily FSTIA dataset is used.

The EEA scheme (Ochotta et al. 2005) is a method that attempts to fit the states with rough grids and then estimates the error of the fitting result. An estimation function (EF) is established to obtain the value of any arbitrary location of the model domain with a given dataset \mathbf{P}_i , by calculating the weighted average of surrounding observations as follows:

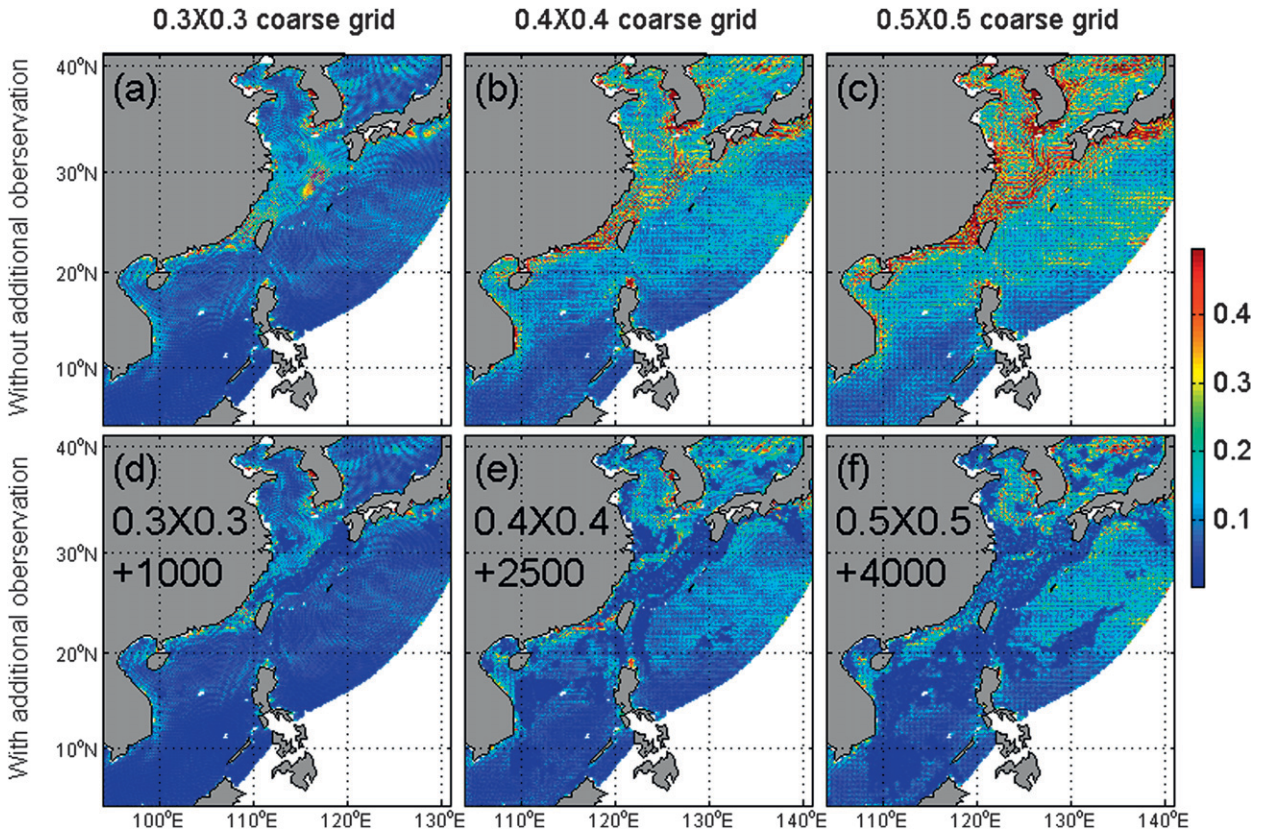


FIG. 8. AEVs [(a)–(c) of 3 coarse-grid subsets and of (d)–(f) 3 augmented subsets] calculated based on the FSTIA dataset during 2006.

$$f_b(x) = \frac{\sum_{p \in \mathbf{P}} f(p)w_h(\|x - p\|)}{\sum_{p \in \mathbf{P}} w_h(\|x - p\|)}, \quad (13)$$

$$EV(\mathbf{P}_i) = \frac{1}{N} \sum_{t=1}^N \|f_{p,0}(p, t) - f_{p,i}(p, t)\|^2. \quad (14)$$

where $w_h(s) = e^{-s^2/h^2}$ is a positive, exponentially decreasing the weighting function with points closer to x assigned with larger weights. The parameter h defines the spatial scale of w_h . And the function $f(p)$ shows the value of any appointed observation points.

Given an observation dataset \mathbf{P}_0 and its thinned subset \mathbf{P}_i , the EF $f_{p,i}$ serves as an approximation of $f_{p,0}$. The EF of the subset, the EF of the whole dataset, and the differences between the two functions were calculated continuously for 1 yr and were described as function $f_{p,i}(p, t)$ and $f_{p,0}(p, t)$, where $t = 1 \dots 365$ presents the time (every day of 2006) of the dataset and p represents the location of the point. To verify the results of our thinning results, we defined a function in every location as the temporal mean of the root of the differences between the EF of the full dataset and a thinned subset during the whole year, by the following equation:

Here, N is the range of verification time whose value is about 1 yr. The EV function reflects the error of the approximation produced by a thinned subset from the full dataset, which serves as a criterion for evaluating the quality of a subset’s resolution and distribution, and to demonstrate the capability of the subset to preserve the original information of the full dataset on an intraseasonal–interseasonal time scale. The EF of a certain subset is obtained to fit the full dataset all over the model domain.

The three regular coarse-grid subsets: $0.3^\circ \times 0.3^\circ$, $0.4^\circ \times 0.4^\circ$, and $0.5^\circ \times 0.5^\circ$ are first verified using the EV function. The EV functions of augmented subsets $0.3^\circ \times 0.3^\circ + 1000$, $0.4^\circ \times 0.4^\circ + 2500$, and $0.5^\circ \times 0.5^\circ + 4000$ are also calculated and the results are shown in Fig. 8.

For the $0.3^\circ \times 0.3^\circ$, $0.4^\circ \times 0.4^\circ$, and $0.5^\circ \times 0.5^\circ$ subsets, the EV functions increase with the dilution of the resolutions. The error variance extends to 0.4°C in several areas, such as the Kuroshio mainstream and its extension, the Changjiang River estuary, the Taiwan Strait, and the JS, indicating that the SST states cannot be well

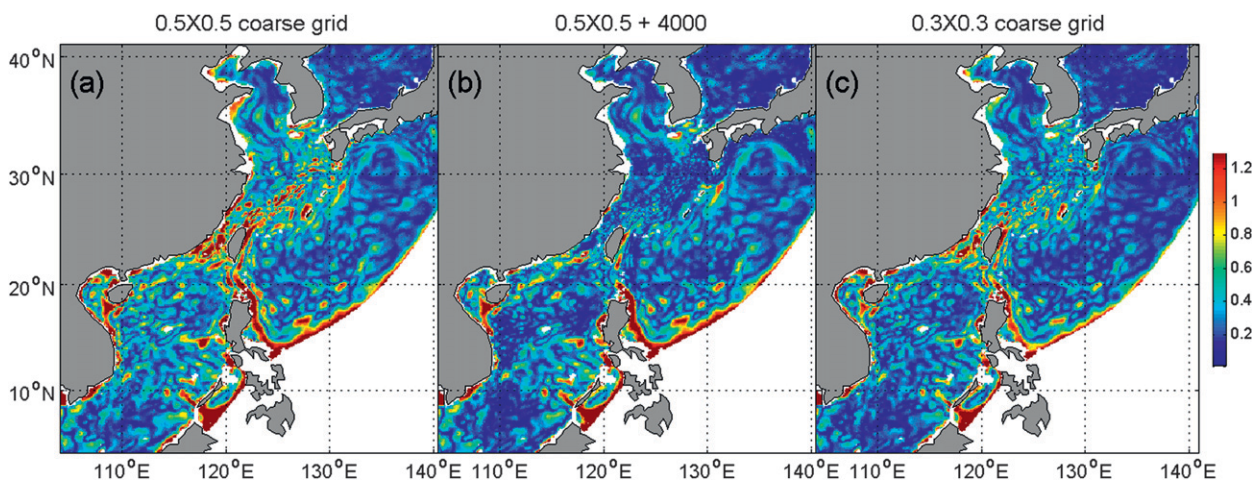


FIG. 9. Time-mean RMSE of SST analysis fields of 3 different observation subsets in 2006: (a) $0.5^{\circ} \times 0.5^{\circ}$ coarse grid, (b) $0.5^{\circ} \times 0.5^{\circ}$ plus 4000 additional observations, and (c) $0.3^{\circ} \times 0.3^{\circ}$ grid.

fitted by a coarse-grid subset in these regions. In sharp contrast to the regular coarse-grid subsets, the augmented subsets can all successfully fit the FSTIA dataset, with EV functions having values below 0.1°C over the model domain, suggesting that the optimally augmented subsets can well preserve the information of the whole dataset.

b. Impacts on data assimilation and forecast

To further examine the impacts of differently thinned observation subsets on data assimilation, we also performed a series of assimilation experiments. A localized ensemble optimal interpolation scheme was used to compare the assimilation results of different observation subsets ($0.3^{\circ} \times 0.3^{\circ}$, $0.5^{\circ} \times 0.5^{\circ}$, and $0.5^{\circ} \times 0.5^{\circ}$ plus 4000 additional observations). In the assimilation scheme, a Gaussian local correlation function with the length scale of 250 km is used. The background error ensembles for the data assimilation are the same as used in observation thinning. The model used in these assimilation experiments is the same as described in section 3. We assimilate the three different subsets every day during the whole 2006 separately. The time-mean SST RMSE of the analysis field is displayed in Fig. 9, which indicates that the analysis result of the coarse-grid observation subset ($0.5^{\circ} \times 0.5^{\circ}$) is not unsatisfying along the Kuroshio and Taiwan Strait. The optimally densified subset ($0.5^{\circ} \times 0.5^{\circ}$ plus 4000 additional observations) significantly contrasts to the homogeneous one ($0.3^{\circ} \times 0.3^{\circ}$); the former successfully reduces the analysis variance over these two high-uncertainty regions, whereas the latter reduces the RMSE uniformly all over the model domain, although it still retains a high variance along the Kuroshio and the Taiwan Strait. In another respect, we also notice two phenomena. First, an extremely high RMSE appears along the boundary of the

model domain in all of the three results, which are not successfully reduced by additional observations. The reason is that the sponge boundary condition leads to a high bias along the model boundary, which can hardly be captured and represented by the ensemble. Second, the high RMSE along the Kuroshio has not been perfectly reduced in the analysis field of the optimal-thinned observations, especially along the southern boundary of the Kuroshio and its extension. One reason could be that the stationary ensemble cannot perfectly represent the variable of the dynamical system. Using thinning observations constructed by a global strategy in a localized assimilation scheme could be another reason. In this localized assimilation scheme, the covariance constructed by the ensemble will be further restricted by the localization operator, which puts the AEV under constraint. However, introducing localization into the thinning system will dramatically increase the computational cost, even making the procedure unpractical. We will make further attempts to improve the algorithm to utilize the dynamic-dependent ensemble and localization scheme in this thinning method in the future work.

The thinning strategy should be justified ultimately by the associated forecast errors. However, the forecast errors are sensitive to many other factors, such as the inflation coefficient, the localization length scales in assimilation schemes, the control of fast gravity waves created by the imbalance between model variables immediately after the assimilation of observations, errors in the atmospheric forecast, as well as the model bias. As a result, it is difficult to handle a verification based on the forecast errors. We here performed three 10-day hindcasts starting from 1 March 2009 with the assimilation of three thinned subsets of the SST—that is, $0.3^{\circ} \times 0.3^{\circ}$, $0.5^{\circ} \times 0.5^{\circ}$, and

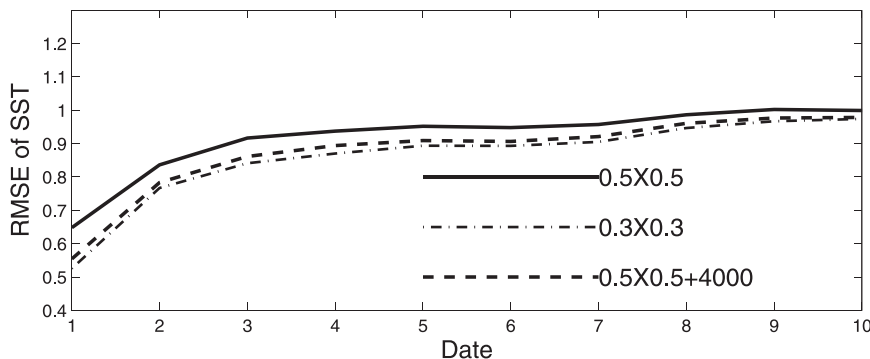


FIG. 10. Space-mean RMSE of 3 parallel hindcast experiments from 1 to 10 Mar, beginning with the assimilation results of these 3 observation subsets, respectively.

$0.5^{\circ} \times 0.5^{\circ}$ plus 4000 additional observations, respectively. The ECMWF reanalysis is used as the atmosphere forcing to minimize the errors in the “forecasted” atmospheric field. The setups for the three hindcast experiments are the same, except for the initial inputs of SST data that are thinned differently. The space-mean RMSE was indicated in Fig. 10, which partly supports the impact of different observation subsets on the forecast results. The deviations among different subsets are small, and the RMSE increased rapidly in the first several days, which could be caused by the adjustment of gravity waves in the first several days and the model bias. We find that the SST along the boundary area decayed very fast. In addition, a significant warm bias appears along the Kuroshio after the 5 day hindcast because of an overestimated transportation along the Kuroshio in the model.

6. Conclusions and discussion

In this study, an observation-thinning scheme is proposed to thin high-resolution SST observations such as the GHRSSST products for a CSCS eddy-resolving model and applied to its assimilation system. The scheme is established following the objective principle of maximally reducing the analysis error with a limited number of observations. The analysis variances are obtained using the AEC matrix update equation before assimilation is performed. The procedure of the thinning scheme is verified by EEA using 1 yr of high-resolution SST observations and is also verified with an assimilation forecast system. The main conclusions are as follows:

- 1) For the Chinese shelf/coastal eddy-resolving model, the $0.3^{\circ} \times 0.3^{\circ}$ resolution SST observations can already successfully reduce the analysis variance in an assimilation system. With the additional optimally located observations taken into consideration, the
- 2) subset of $0.5^{\circ} \times 0.5^{\circ}$ resolution plus 4000 additional optimal locations is selected to be the optimal thinning scheme. After being updated with the thinned grid observations, the AEV are effectively restricted below 0.1°C all over the model domain.
- 3) The additional optimal locations mainly concentrate within several special areas: the Kuroshio mainstream and its extension, the outflow of the Changjiang River, the Taiwan Strait, and the three centers in SCS. These areas produced by the optimal observation thinning scheme can always be associated with physical processes, and the refined subsets can efficiently restrict the high variances around these regions.
- 4) The optimally thinned subsets are verified using the full-resolution dataset and the result is satisfying—that is, the optimal subset can maximally restrict the analysis variances with a limited number of observations, and the full dataset is well fitted by the optimal subsets constructed by the thinning scheme.
- 5) The optimal thinned subsets are also verified with a series of assimilation–hindcast experiments. The results indicate that this thinning scheme is efficient in improving the assimilation results.

The purpose of this study is to develop an objective scheme to help us locally reduce the resolution of observations before conducting data assimilation. Despite the fact that this observation-thinning scheme was applied only to the SST dataset here, it has the potential for application to other dense observations and further investigation is necessary to evaluate the effects. The results of the experiments are meaningful in that the concentrated regions can always be well explained by the particular physical phenomena and processes. Further study will focus on improving this method by localized scheme, as well as by improving the arithmetic to implement a time-dependent/dynamic-dependent ensemble optimal thinning scheme.

Acknowledgments. This research was supported by the Chinese Academy of Science (Contract KZCX1-YW-12-03), the National Basic Research Program of China (2006CB403600), and the Natural Science Foundation of China (Contract 40437017 and 40221503).

REFERENCES

- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Bleck, R., 2002: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Modell.*, **4**, 55–88.
- Boyer, T. P., C. Stephens, J. I. Antonov, M. E. Conkright, R. A. Locarnini, T. D. O'Brien, and H. E. Garcia, 2002: *Salinity*. Vol. 2, *World Ocean Atlas 2001*, S. Levitus, Ed., NOAA Atlas NESDIS 50, 165 pp.
- Chassignet, E. P., L. T. Smith, G. R. Halliwell, and R. Bleck, 2003: North Atlantic simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity. *J. Phys. Oceanogr.*, **33**, 2504–2526.
- , H. E. Hurlburt, O. M. Smedstad, G. R. Halliwell, P. J. Hogan, A. J. Wallcraft, R. Baraille, and R. Bleck, 2007: The HYCOM (Hybrid Coordinate Ocean Model) data assimilative system. *J. Mar. Syst.*, **65**, 60–83.
- Chu, P. C., B. Ma, and Y. Chen, 2002: The South China Sea thermohaline structure and circulation. *Acta Oceanol. Sin.*, **21**, 227–261.
- Donlon, C. J., 2003: *Proceedings from the Third GODAE High-Resolution SST Pilot Project Workshop*. International GHRSSST-PP Project Office, 141 pp.
- , P. J. Minnett, C. L. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and M. J. Murray, 2002: Toward improved validation of satellite sea surface skin temperature measurements for climate research. *J. Climate*, **15**, 353–369.
- , and Coauthors, 2004: The recommended GHRSSST-PP data processing specification gDS (version 1 revision 1.5). GHRSSST-PP International Project Office, Met Office, 241 pp. [Available online at <http://www.ghrsst-pp.org/>]
- , and Coauthors, 2007: The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bull. Amer. Meteor. Soc.*, **88**, 1197–1213.
- Gan, J. P., H. Li, E. N. Curchitser, and D. B. Haidvogel, 2006: Modeling South China Sea circulation: Response to seasonal forcing regimes. *J. Geophys. Res.*, **111**, C06034, doi:10.1029/2005JC003298.
- Khare, S. P., and J. L. Anderson, 2006: An examination of ensemble filters-based adaptive observation methodologies. *Tellus*, **58A**, 179–195.
- Langland, R. H., 2005: Issues in targeted observations. *Quart. J. Roy. Meteor. Soc.*, **131**, 3409–3425.
- Levitus, S., and T. P. Boyer, 1994: *Temperature*. Vol. 4, *World Ocean Atlas 1994*, NOAA Atlas NESDIS 4, 117 pp.
- , R. Burgett, and T. P. Boyer, 1994: *Salinity*. Vol. 3, *World Ocean Atlas 1994*, NOAA Atlas NESDIS 3, 99 pp.
- Liu, Z.-Q., and F. Rabier, 2002: The interaction between model resolution, observation resolution, and observation density in data assimilation: A one-dimensional study. *Quart. J. Roy. Meteor. Soc.*, **128**, 1367–1386.
- Ochotta, T., C. Gebhardt, D. Saupe, and W. Wergen, 2005: Adaptive thinning of atmospheric observations in data assimilation with vector quantization and filtering methods. *Quart. J. Roy. Meteor. Soc.*, **131**, 3427–3437.
- , —, V. Bondarenko, D. Saupe, and W. Wergen, 2007: On thinning methods for data assimilation of satellite observations. Preprints, *23rd Int. Conf. on Interactive Information Processing Systems (IIPS)*, San Antonio, TX, Amer. Meteor. Soc., 2B.3. [Available online at http://ams.confex.com/ams/87ANNUAL/techprogram/paper_118511.htm.]
- Oke, P. R., and P. Sakov, 2008: Representation error of oceanic observations for data assimilation. *J. Atmos. Oceanic Technol.*, **25**, 1004–1017.
- , G. B. Brassington, D. A. Griffin, and A. Schiller, 2008: The Bluelink ocean data assimilation system (BODAS). *Ocean Modell.*, **21**, 46–70.
- Qu, T., H. Mitsudera, and T. Yamagata, 2000: Intrusion of the North Pacific waters into the South China Sea. *J. Geophys. Res.*, **105**, 6415–6424.
- Sakov, P., and P. R. Oke, 2008: Objective array design: Application to the tropical Indian Ocean. *J. Atmos. Oceanic Technol.*, **25**, 794–807.
- , G. Evensen, and L. Bertino, 2009: Asynchronous data assimilation with the EnKF. *Tellus*, **62A**, 24–29.
- Testut, C., P. Brasseur, J. Brankart, and J. Verron, 2003: Assimilation of sea surface temperature and altimetric observations during 1992–1993 into an eddy permitting primitive equation model of the North Atlantic Ocean. *J. Mar. Syst.*, **40–41**, 291–316.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- Xie, J. P., J. Zhu, and L. Yan, 2008: Assessment and intercomparison of five high-resolution sea surface temperature products in the shelf and coastal seas around China. *Cont. Shelf Res.*, **28**, 1286–1293.
- Xu, Q., 2007: Measuring information content from observations for data assimilation: Relative entropy versus Shannon entropy difference. *Tellus*, **59A**, 198–209.
- Yan, C. X., J. Zhu, and G. Q. Zhou, 2007: Impacts of XBT, TAO, altimetry, and ARGO observations on the tropic Pacific Ocean data assimilation. *Adv. Atmos. Sci.*, **24**, 383–398.