Ocean Dynamics (2017) 67:433–449
DOI 10.1007/s10236-017-1030-y

An effective method based on dynamic sampling for data assimilation in a global wave model

Meng Sun[1,2] · **Xunqiang Yin**[2,3] · **Yongzeng Yang**[2,3] · **Kejian Wu**[1]

Received: 19 September 2016 / Accepted: 2 January 2017 / Published online: 24 January 2017

Abstract The ensemble Kalman filter (EnKF) performs well because that the covariance of background error is varying along time. It provides a dynamic estimate of background error and represents the reasonable statistic characters of background error. However, high computational cost due to model ensemble in EnKF is employed. In this study, two methods referred as static and dynamic sampling methods are proposed to obtain a good performance and reduce the computation cost. Ensemble adjustment Kalman filter (EAKF) method is used in a global surface wave model to examine the performance of EnKF. The 24-h interval difference of simulated significant wave height (SWH) within 1 year is used to compose the static samples for ensemble errors, and these errors are used to construct the ensemble states at each time the observations are available. And then, the same method of updating the model states in the EAKF is applied for the ensemble states constructed by a static sampling method. The dynamic sampling method employs a similar method to construct the ensemble states, but the period of the simulated SWH is changing with time. Here, 7 days before and after the observation time is used as this period. To examine the performance of three schemes, EAKF, static, or dynamic sampling method, observations from satellite Jason-2 in 2014 are assimilated into a global wave model, and observations from satellite Saral are used for validation. The results indicate that the EAKF performs best, while the static sampling method is relatively worse. The dynamic sampling method improves an assimilation effect dramatically compared to the static sampling method, and its overall performance is closed to the EAKF. In low latitudes, the dynamic sampling method has a slight advantage over the EAKF. In the dynamic or static sampling methods, only one wave model is required to run and their computational cost is reduced sharply. According to the performance of these three methods, the dynamic sampling method can treated as an effective alternative of EnKF, which could reduce the computational cost and provide a good performance of data assimilation.

Keywords Dynamic/static ensemble · Sampling method · EAKF · Computational cost

This article is part of the Topical Collection on the *8th International Workshop on Modeling the Ocean (IWMO), Bologna, Italy, 7-10 June 2016*

Responsible Editor: Tal Ezer

✉ Xunqiang Yin
yinxq@fio.org.cn

[1] College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao 266100, China

[2] The First Institute of Oceanography, State Oceanic Administration (SOA), Qingdao 266061, China

[3] Laboratory for Regional Oceanography and Numerical Modeling, National Laboratory for Marine Science and Technology, Qingdao 266071, China

1 Introduction

The combination of using model results and observations to create an optimal estimate of the sea state is called data assimilation or analysis (Komen et al. 1996). The availability of near real-time global wave data from the all weather microwave sensors on board satellites such as Jason-1/2, ERS-1/2, and Saral has provided a strong impetus for the development of wave data assimilation schemes. Considering both an assimilation effect and computational cost, methods for wave data assimilation such as optimal interpolation (OI) and EnKF have their

own advantages and disadvantages. EnKF with high computational cost performs better than OI with low computational cost, because covariance of background error varies along time.

There are many studies of wave data assimilation with OI to optimize wave parameters (Lionello et al. 1992, 1995; Greenslade 2001) and ocean wave spectra (Hasselmann et al. 1997; Voorrips et al. 1997; Waters 2013). Nowadays, operational wave data assimilation systems at the European Centre for Medium-Range Weather Forecasts (ECMWF), the BoM, and Fleet Numerical Meteorology and Oceanography Center (FNMOC) use an OI method (Lorenc 1981) with low computation cost to obtain analyzed wave fields (Greenslade 2004). One of the main limitations to current OI schemes is the lack of an accurate representation of the structure of the background error. In the OI method, the background error covariance matrix is supposed as non-flow-dependent and insufficient to describe the real structure of the background error. In contrast, EnKF (Evensen 1994; Burgers et al. 1998) can reveal the probability distribution of numerical models through ensemble statistical analysis and provide a dynamic estimate of background error. An EAKF method is an important representation of EnKF methods. There are many successful implementations of the EAKF method in ocean/atmosphere data assimilation (Zhang et al. 2005; Zhang et al. 2007; Karspeck et al. 2007; Yin et al. 2010, 2011). The spread of an ensemble of short-range forecasts can be analyzed to calculate spatially and temporally varying background error statistics. Thus, a flow-dependent background error correlation matrix can be produced. However, computational cost of EnKF is huge for running a wave model ensemble simultaneously.

Focusing on the problems mentioned above, we try to find a solution with a favorable assimilation effect and low computational cost. Ensemble errors constructed by 24-h interval difference between simulated SWH from long-term history model results are used to compose ensemble states by superposing ensemble errors to an SWH field at time window for assimilation. The same method of updating model states in the EAKF is applied for ensemble states constructed by ensemble errors. In this study, two sampling methods for ensemble errors are employed. One is a static sampling method collecting one sample each day for a year, and these static samples for ensemble errors (also referred as a static ensemble) maintain unchangeable during an assimilation process. Another is a dynamic sampling method collecting 24 samples each day on interval ($t-7d$, $t+7d$) at each time window t for assimilation, and these dynamic samples for ensemble errors (also referred as a dynamic ensemble) are updated over time. To examine the assimilation effect of these two methods, we compare them with the EAKF method. In this study, a wave model ensemble prepared for the EAKF is constructed by wind fields with random field perturbation (Sun et al. 2016). Three assimilation experiments with EAKF, static, and dynamic sampling methods respectively will be carried out to assimilate observations from satellite Jason-2 into a global wave model over the period 2014. Observations from satellite Saral will be used for validation.

The paper is organized as follows. Section 2 describes EAKF, static, and dynamic sampling methods. Section 3 describes designing of data assimilation experiments, implementing process, and wave model and data. Section 4 presents analysis of assimilation results. Finally, summary and discussions are given in Section 5.

2 Methodology

2.1 Ensemble adjustment Kalman filter

For the sake of clarity of the overall procedure, it is convenient to provide here a brief description of the EAKF method. Anderson (2001) proposed this method denoted the ensemble adjustment Kalman filter, where the analysis is computed without adding perturbations to the observations. It retains the nonlinear information of high-order moments from prior probability distribution function as much as possible. In this method, a linear operator which replaces the traditional gain matrix was introduced. This method makes it possible to obtain reliable results for small ensembles (10–20 members) (Evensen 2003). Anderson (2003) derived the ensemble filter methods that made a (local) least squares assumption about the relation between prior distributions of an observation variable and model state variables. Thus, the update procedure can be described into two parts. First, an update increment is computed for each prior ensemble estimate of the observation variable by applying a scalar ensemble filter. Second, a linear regression of the prior ensemble sample of each state variable on the observation variable is performed to compute update increments for each state variable ensemble member from corresponding observation variable increments. The two-part filter method which will be used in this study can provide a computationally efficient implementation of ensemble filters.

The construction of a wave model ensemble which should be representative is crucial for wave data assimilation with the EAKF. In fact, wave simulation is in many aspects not very sensitive to an initial spectrum; the wave field tends to lose the memory of the initial state; thus, a perturbation ensemble is superposed to a wind forcing field rather than an initial field to construct a wave model ensemble. In this study, to construct a wind forcing field

ensemble, a random field scheme (Sun et al. 2016) is used. The random field scheme can be written as follows.

$$W^a_{i,j,n} = W^b_{i,j} + \alpha_{B_n} \cdot \lambda_{i,j,n} \quad (1)$$

Here, $W^b_{i,j}$ and $W^a_{i,j,n}$ represent the wind forcing field before and after the perturbation, respectively; subscripts i and j are grid indexes on a horizontal space; subscript $n = 1,2,\ldots,N$ is the ensemble index where N is ensemble size; λ represents the random field; α_{B_n} represents a parameter which is used to control the amplitude of the perturbation. α_{B_n} is the ratio of standard deviation between the wind field and the random field. The method for generating the random field given by Evensen (1994) is used here to perturb wind fields. A spatial correlation scale used for generating the random field is 5°. The random field is smooth spatially, and their spatial correlation decreases with increasing distance. Therefore, this kind of perturbation will not break the smoothness of the integration itself.

Based on simulated results of a wave model ensemble driven by a wind field ensemble, correlation coefficient between a referred location and the grid points around it can be calculated, which reflects the weight information during a wave data assimilation process. This is a key indicator to evaluate the rationality of the construction of a wave model ensemble. Based on the results of an experiment with the random field scheme in which the global wind field from ECMWF over the period of January 2014 was applied to construct a wave model ensemble (Sun et al. 2016), here four grid points in different latitudes and oceans are selected to calculate the monthly mean of SWH spatial correlation coefficient. The result is shown in Fig. 1. In Fig. 1, correlation coefficient decreases with increasing distance and a spatial correlation scale of SWH is reasonable; especially, swell propagation from southern hemisphere westerlies plays an important role in East Indian Ocean and Southeast Pacific Ocean. In low latitudes, a spatial correlation scale is larger than that in mid-high latitudes, and that is consistent with the result by Greenslade (2004).

2.2 Static sampling method

As an approximation of background error, ensemble errors which consist of 24-h interval difference between simulated SWH are used by Ren (2010) to assimilate Envisat ASAR data with the OI method. Sun et al. (2014) proposed a new concept referred as a static sample ensemble which is defined as the difference between several hours of interval forecasts (6, 12, 24, 36, 48, 72 h) of simulated SWH and found that 24 h is the optimal temporal interval through comparison of six groups of the static sample ensemble to the difference between model results and satellite altimeter-merged observation data. In addition, based on the optimal static sample ensemble, a hindcast experiment with a two-part filter method (see Section 2.1) is implemented to assimilate observations from satellite Jason-1.

Based on the studies mentioned above, a static sampling method is suggested in this study. A static ensemble collecting one sample each day for 1 year from simulated SWH time series could be used to estimate background error covariance. The approach can be written as follows:

$$Cov(X,Y) = E\left\{\begin{matrix}\left[\left(x^t - x^{t-24h}\right) - E\left(x^t - x^{t-24h}\right)\right] \\ \left[\left(y^t - y^{t-24h}\right) - E\left(y^t - y^{t-24h}\right)\right]\end{matrix}\right\} \quad (2)$$

where Cov(X, Y) represents covariance between vectors X and Y, $E(\ldots)$ is the expected value, vectors X and Y consist of SWH time series at two different locations, superscript t or $t-24$ h represents time.

In the static sampling method, ensemble states are constructed by superposing the static ensemble unchangeable over time to the SWH field at assimilation time rather than the wave model ensemble. Then, the two-part filter method (see Section 2.1) is used to update ensemble states. Thus, only one wave model is carried out during the whole data assimilation process.

In this work, the static ensemble used in a subsequent assimilation experiment is calculated from wave model results without assimilation over the period 2014. An error sample at 00:00:00 UTC for each day was collected. Thus, size of the static ensemble is 365. Spatial distribution of ensemble spread in the static sampling method is shown in Fig. 2. Background error in mid-high latitudes is larger than that in low latitude. In this study, ensemble spread is defined as follows.

$$S(A) = \frac{1}{N}\sum_{i=1}^{N}\left|A_i - \overline{A}\right| \quad (3)$$

where $S(A)$ represents spread of ensemble A, A_i represents the ith ensemble member, $\overline{A}$ represents an ensemble mean, N is ensemble size.

2.3 Dynamic sampling method

In an EnKF method, background error is calculated by a model forecast ensemble at each time window of data assimilation and it is flow-dependent. This is the advantage of the EnKF for estimating background error compared to OI. Thus, we suggest a dynamic sampling method using a dynamic ensemble to estimate background error. Similar to the static ensemble, the dynamic ensemble consists of 24-h interval difference between simulated SWH as well, but it is updated over time during the data assimilation process.

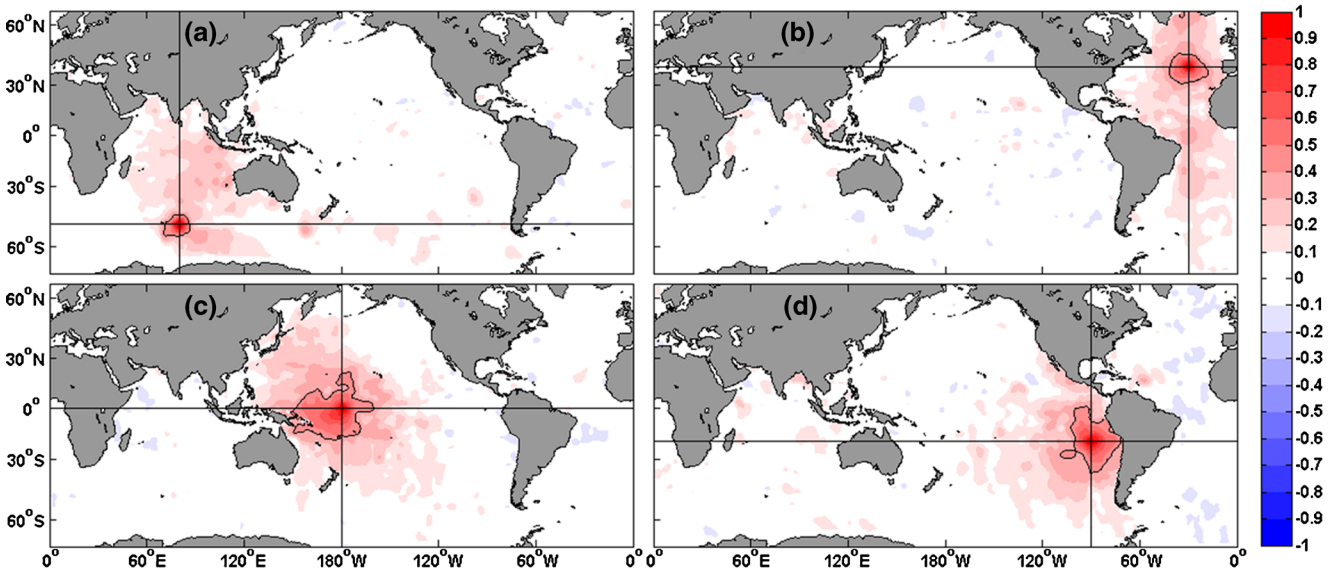

Fig. 1 Spatial distributions of monthly mean correlation coefficient in SWH background error referred to different positions in January 2014. For data assimilation experiment with the EAKF method, a wave model ensemble is forced by wind field-introduced random field perturbation. Based on the result of a wave model ensemble without assimilation in January 2014, correlation coefficients in SWH background error referred to different positions are calculated to illustrate the reasonability of a wave model ensemble. The referred positions for panels **a**, **b**, **c**, and **d** are 80°E, 50°S; 30°W, 40°N; 180°E, 0°; and 90°W, 20°S, respectively, and these locations which are surrounded by *black contour* of 0.4 are marked by *black crosses* in these panels. The contour interval is 0.1 in these panels

It is known that a temporal scale of weather variations is about 1 week. In this study, at each time window t for data assimilation, the dynamic ensemble with 360 members is constructed by collecting 24 samples each day on an interval ($t-7d$, $t+7d$) from simulated SWH time series without assimilation in 2014. Spatial distributions of ensemble spread in the dynamic sampling method at 15 January 2014 00:00:00 UTC and 15 July 2014 00:00:00 UTC are shown in Fig. 3. It is suggested that there is a dramatic seasonal variation characteristic of background error. In boreal winter, background error is larger in latitudes extending from 30°N to 60°N than in the rest of the region. On the contrary, in austral winter, background error is larger in latitudes extending from 30°S to 60°S than in the rest of the region. Thus, the dynamic ensemble is more reasonable than the static ensemble in this sense.

A flowchart of the dynamic sampling method is shown in Fig. 4. Wave data assimilation is implemented every 6 h with the integration of the wave model. At each time window for

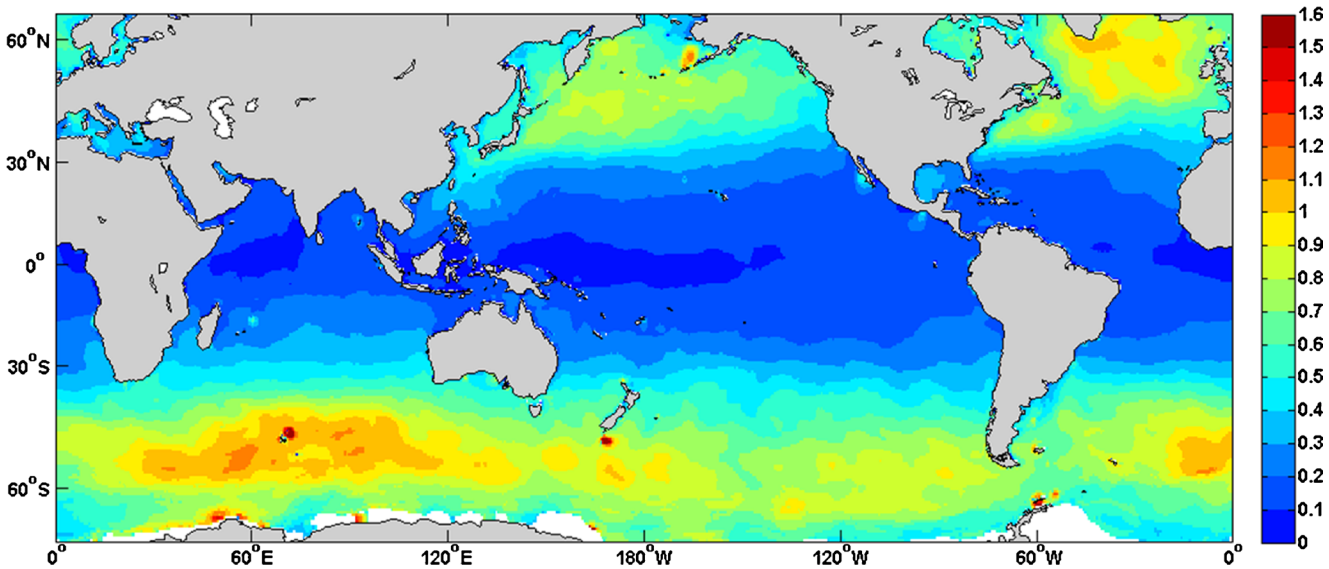

Fig. 2 Spatial distribution of ensemble spread in the static sampling method. For data assimilation experiment with the static sampling method, a static ensemble collecting one sample each day for 1 year from wave model results without assimilation in 2014 is used to compose ensemble states, and it maintains unchangeable during the whole assimilation process. Ensemble spread is larger in mid-high latitudes than in low latitudes. The contour interval is 0.1 m. The *white region* represents sea ice

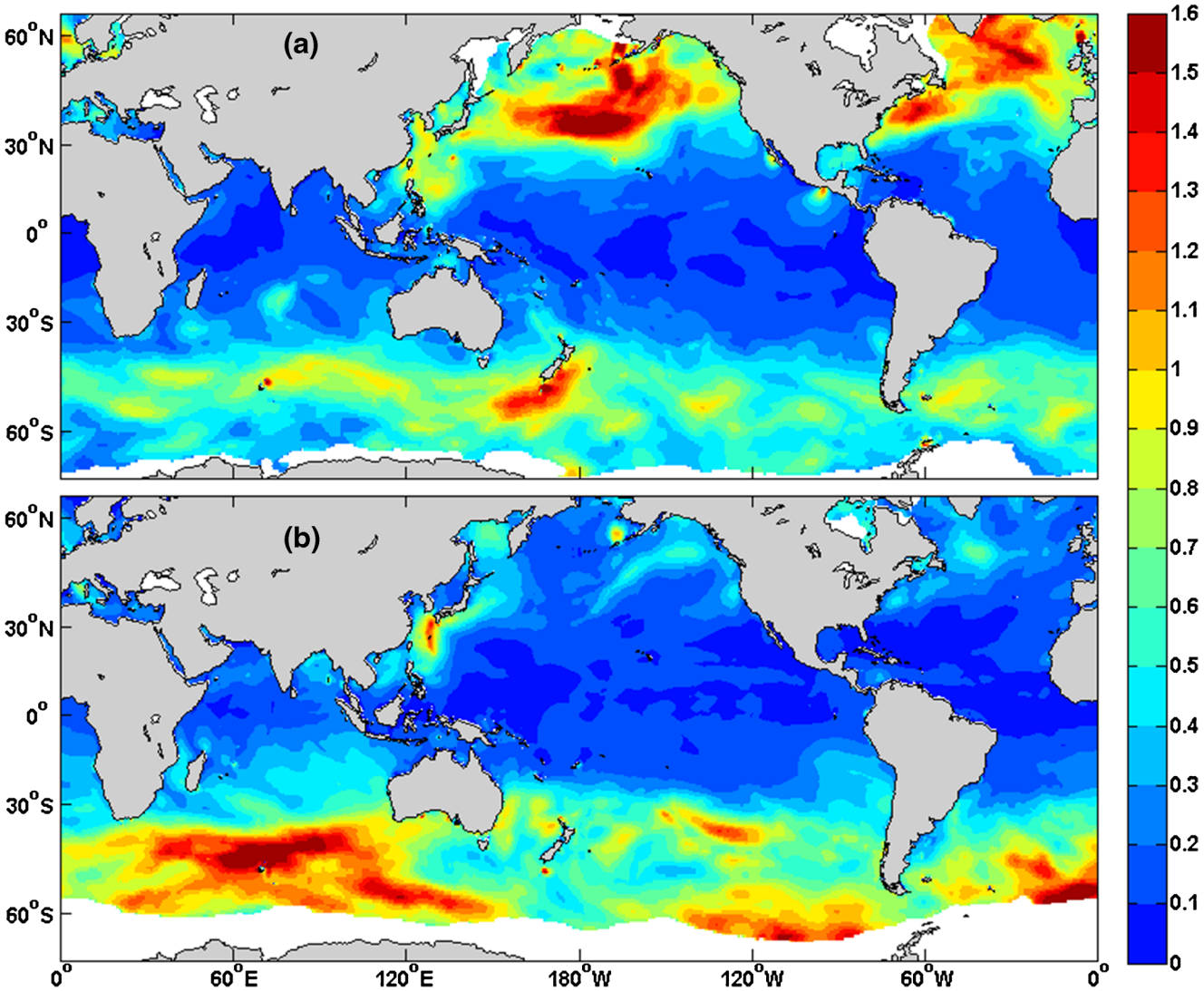

Fig. 3 Spatial distributions of ensemble spread in the dynamic sampling method. For data assimilation experiment with the dynamic sampling method, a dynamic ensemble as an approximation to background error is constructed at each time window t for data assimilation and is used to compose ensemble states. The dynamic ensemble consists of 24-h interval difference between simulated SWH and is constructed by collecting 24 samples each day on interval ($t-7d$, $t+7d$) from wave model results without assimilation in 2014. Thus, it is updated over time during the whole assimilation process. Spatial distributions of ensemble spread at **a** 15 January 2014 00:00:00 UTC and **b** 15 July 2014 00:00:00 UTC are quite different, especially in mid-high latitudes. Thus, the dynamic ensemble containing seasonal variation information of background error is more reasonable than the static ensemble. The contour interval is 0.1 m. The *white region* represents sea ice

assimilation, an energy spectrum is scaled by a ratio of analyzed SWH and forecast SWH. The procedure to obtain the analyzed SWH field is shown in dashed box. It is divided into four steps as follows:

Step 1: Construction of the dynamic ensemble

Data source for the dynamic ensemble is global model results without assimilation in 2014. Time series of SWH on an interval ($t-7d$, $t+7d$) are collected to construct ensemble A. Each member of ensemble A is difference between simulated SWH fields at t' and $t'-24$; here, $t' \in (t-7d,\ t+7d)$. Then, ensemble $A' = A-\overline{A}$; here, $\overline{A}$ represents the ensemble mean of A. Ensemble A' is the dynamic ensemble. It is known that output fields of integrated wave parameters are archived at 1-hourly intervals. Thus, size of the dynamic ensemble is 360 = 15d × 24 which is consistent with size of the static ensemble.

Step 2: Construction of ensemble states

The model state ensemble is the key of the EnKF. The information contained by a full probability density function can

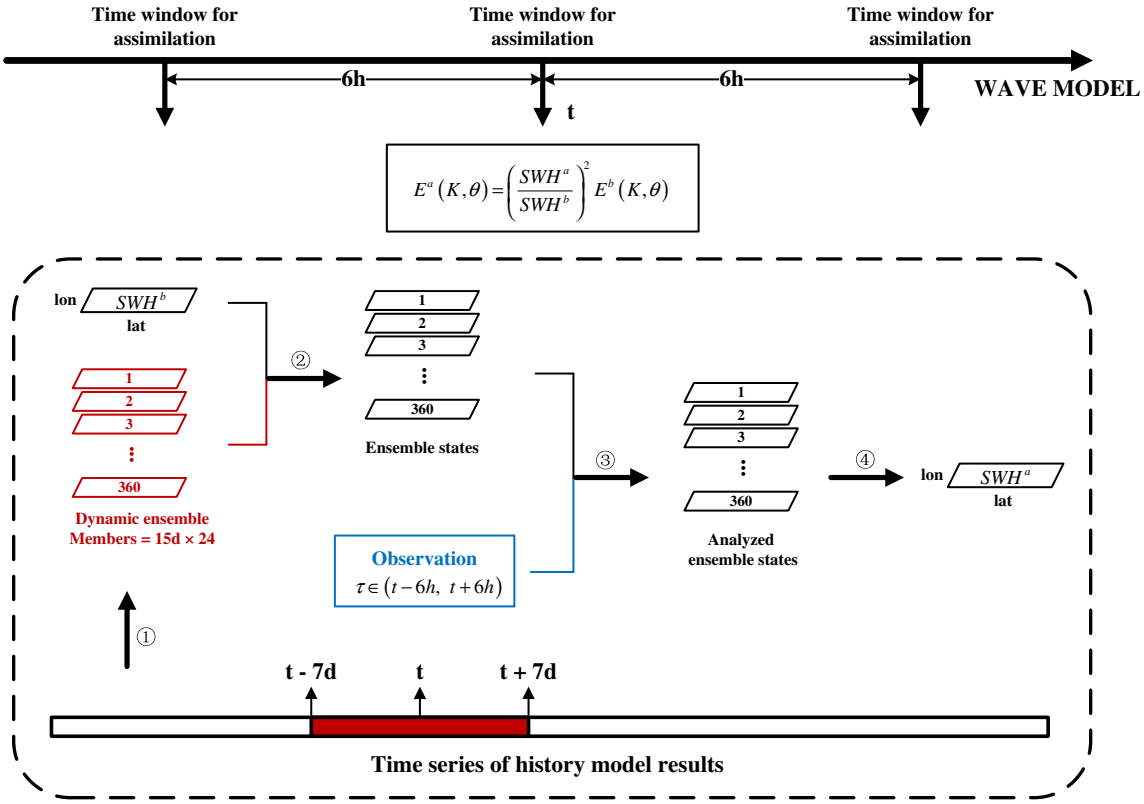

Fig. 4 Flowchart of the dynamic sampling method. With integration of the wave model, data assimilation is implemented every 6 h and the wave spectrum is scaled by a ratio between analyzed SWH and forecast SWH. Four steps are defined to obtain analyzed SWH. The first step is construction of the dynamic ensemble. Ensemble states are constructed by superposing the dynamic ensemble to a forecast SWH field in the second step. In the third step, a two-part filter method is used to achieve analyzed ensemble states by assimilating observations from satellite Jason-2. In the fourth step, analyzed SWH is calculated by taking the average of analyzed ensemble states

be exactly represented by an infinite ensemble of model states (Evensen 2003). The dynamic ensemble as an approximation to background error is superposed to the simulated SWH field at assimilation time t to obtain ensemble states with 360 members.

Step 3: Update ensemble states

The two-part filter method (see Section 2.1) is applied to update ensemble states. Information of each observation was spread within a radius 2° with a localization weight (Yin et al. 2010). Observation at time $\tau\in(t-6h,\ t+6h)$ is assimilated into the wave model by turn at time window t for assimilation. Analyzed ensemble states will not achieve until all observations are assimilated.

Step 4: Output-analyzed SWH field

The analyzed SWH field is the mean of analyzed ensemble states.

3 Experiment design

In this work, comparison among three assimilation schemes, the traditional EAKF, static, and dynamic sampling method, is carried out. There are four experiments that covered the global region (longitude 0°–360°, latitude 70°S–65°N) over the period 2014. Model simulation without data assimilation is referred as a control run. The other three assimilation experiments are using the EAKF method, static sampling method, and dynamic sampling method, respectively. The same observations from satellite Jason-2 are assimilated into a third-generation wave model-Marine Science and Numerical Modeling (MASNUM). For validation, independent observations from satellite Saral are used here. To provide the same initial condition for four experiments at the start of model simulation, a model without assimilation was run over the period December 2013 in advance.

3.1 Implementing process

A flowchart of three data assimilation schemes is shown in Fig. 5. In a wave data assimilation system, there are two parts. One part is integration of the wave model and another part is the procedure of data assimilation. These two parts will be started simultaneously. A control file is employed here to control the running of the wave model(s) and the assimilation process. There are 10 models for experiment with the

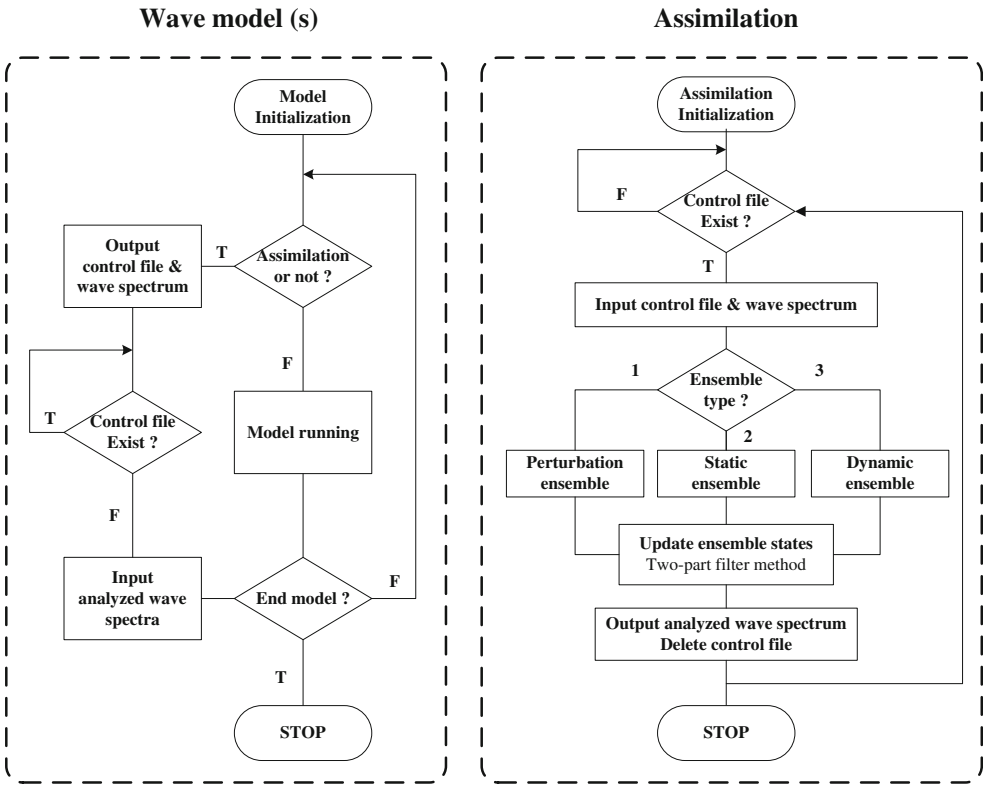

Fig. 5 Flowchart of the data assimilation system for three schemes EAKF, static sampling method and dynamic sampling method. There are two parts in the data assimilation system. Part I (on the *left*) represents the wave model for the static/dynamic sampling method or wave models with 10 members for the EAKF method. Part II (on the *right*) represents an assimilation module in which three approaches are available to construct ensemble states, a perturbation ensemble for EAKF, and a static/dynamic ensemble for the static/dynamic sampling method. At each time window of assimilation, wave model(s) outputs a control file and wave spectrum, and it continues to integrate forward till the control file is deleted by the assimilation module

EAKF, and there is only one model for experiment with the static/dynamic sampling method.

Part I: wave model(s)

The wave model begins to integrate forward after initialization. Judgment statement of calling an assimilation module is executed at each time step (15 min). A temporal interval between two adjacent data assimilation windows is 6 h. At each time window for assimilation, the wave model(s) will output a control file and wave spectrum and will wait until the control file is deleted by the assimilation module. Then, the wave model(s) will input an analyzed wave spectrum and will continue to integrate forward.

Part II: Assimilation procedure

There is a cycle of assimilation after initialization in this part. This cycle will begin when a control file exists. At the beginning of this cycle, input the control file and wave spectrum. Then, construct ensemble states which are essential for ensemble Kalman filter methods. Three modules for construction of ensemble states are available during the data assimilation process, which are the perturbation ensemble of the wave model for the EAKF, static ensemble, and dynamic ensemble, respectively. Then, the ensemble states will be updated by the two-part filter method (see Section 3.1). At the end of this cycle, output-analyzed the wave spectrum and delete the control file.

The assimilation of SWH into the directional spectrum was done by scaling the forecasted spectral components with a ratio (Esteva 1988), which can be written as.

$$E^{a}_{i,j}(K,\theta) = \left(\frac{H^{a}_{i,j}}{H^{b}_{i,j}}\right)^{2} E^{b}_{i,j}(K,\theta) \tag{4}$$

Here, subscripts i and j are grid indexes on a horizontal space; K, θ, and $E(K,\theta)$ represent wavenumber, direction, and wavenumber spectrum, respectively; $H^{a}_{i,j}$ and $H^{b}_{i,j}$ are analyzed and forecasted SWH, respectively. The total energy under the modified spectrum will correspond to the analyzed SWH, while the spectrum shape remains unchanged. But

indeed, it still varies gradually due to driving and dissipation source functions, especially the nonlinear wave-wave interaction. This approach assumes that the model correctly predicts the proportion of energy in wind-sea and swell portions of the spectrum.

3.2 Wave model and data

The model used in this study is a third-generation global wave model-MASNUM (Yuan et al. 1992a, b; Yang et al. 2005). The complicated characteristic inlaid method is applied to integrate the wave energy spectrum balance equation. In the model, the wave energy spectrum balance equation and its complicated characteristic equations are derived in a wave-number space. The breaking dissipation source function adopted a theoretical result based on a statistical study of breaking waves (Yuan et al. 1986).

The MASNUM wave model was implemented on a 0.5° × 0.5° latitude-longitude grid that covers the global domain extending from 70°S to 65°N and from 180°W to 180°E. The model simulates the 2D spectrum of wave energy discretized into 24 directional bands, 15° wide, and 25 frequency bands spaced from 0.042 to 0.413 Hz. The propagation time step is 15 min. Output fields of integrated wave parameters (i.e., H_s, T_p) are archived at 1-hourly intervals. The model topography is interpolated from a gridded bathymetric dataset of $5'\times5'$ resolution (ETOP5 1986) with the maximum depth set to 5500 m. Wind forcing fields for the global wave model are wind velocities at 10 m above sea level. These are obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF). Spatial and temporal resolutions of the winds are 0.5° × 0.5° and 6 h.

The data used for data assimilation in this study were Jason-2 Fast Delivery altimeter SWH. Jason-2/OSTM takes over and continues Topex/Poseidon and Jason-1 missions in June 2008, in the frame of cooperation between CNES, EUMETSAT, NASA, and NOAA. It carries the same kind of payload than its two predecessors, for a high-accuracy altimetry mission: a Poseidon-class altimeter and a radiometer and three location systems. The orbit is also the same. It has a repeat period of 10 days.

The data used for validation of assimilation experiments were Saral Fast Delivery altimeter SWH. An ISRO (Indian Space Research Organization) satellite, Saral (satellite with ARgos and ALtika), embarks the AltiKa altimeter (working in Ka-band, 35 GHz), built by CNES. The mission was launched in February 2013 and has a repeat period of 35 days. The Saral mission is an essential component of the altimetry constellation from 2013 onwards, re-occupying the long-term ERS and Envisat ground track.

Along track SWH data of altimeter products, the final geophysical data records of Jason-2 and Saral satellites are produced and distributed by Archiving, Validation, and Interpolation of Satellite Oceanographic data (AVISO, http://www.aviso.altimetry.fr/en/home.html).

4 Analysis of assimilation results

The evaluation of assimilation results will be clarified in three parts. In Section 4.1, comparison between simulated SWH and observed SWH along satellite track will be given. In Section 4.2, overall performance of three assimilation schemes will be presented through a statistical analysis of model results against satellite data over the full period. In Section 4.3, through maps and time series, difference of spatial and temporal distribution characteristics among three assimilation methods will be illustrated. To evaluate assimilation results of experiments effectively, model results that covered the global region over the period 2014 were collocated with Saral satellite altimeter data which is along track firstly.

4.1 Local error along track

Comparison between simulated SWH and observed SWH along track in January and July 2014 is shown in Figs. 6 and 7, respectively. Four satellite tracks in different oceans are chosen here. Figure 6 shows that in boreal winter, the dynamic sampling method has a performance closed to the EAKF scheme. In addition, there is a remarkable improvement of the dynamic sampling method compared with the static sampling method in high sea state. Similar conclusion could be obtained from Fig. 7. In austral winter, wind speed is larger than other seasons and the wind field changes rapidly in southern hemisphere westerlies. In spite of the huge difference of ensemble size between the EAKF and static sampling method, the EAKF containing instantaneous information of background error still has a natural advantage compared with the static sampling method, especially in high sea state. The dynamic sampling method improved significantly compared with the static sampling method for updating ensemble errors over time during the assimilation process. For the dynamic sampling method having both statistical and seasonal information of background error, it is comparable with the EAKF.

4.2 Overall error statistics

The results of four experiments have been validated against measurements. There are four statistical regions. One is the global region. The other three statistical regions are defined in Fig. 7: region I covers the region extending from 30°S to 70°S, region II covers the region extending from 30°S to 30°N, and region III covers the region extending from 30°N to 65°N.

Table 1 compares the control run with three assimilation runs in terms of mean absolute error (MAE) of SWH. MAE reduction percentage of assimilation experiments

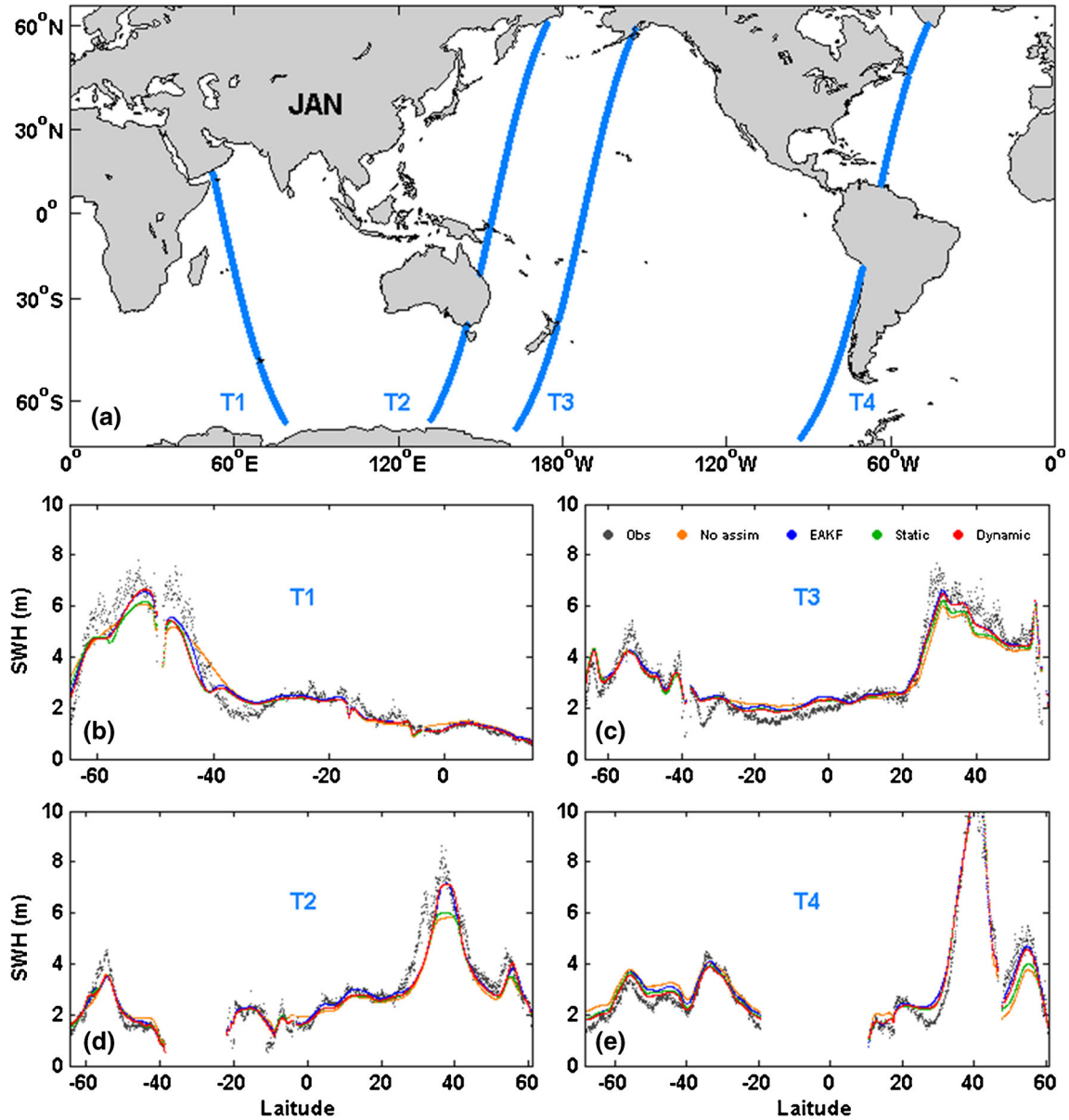

Fig. 6 Comparison between simulated SWH and observed SWH along track in January 2014. Simulated SWH from a control run without assimilation and three assimilation experiments with the EAKF, static sampling method and dynamic sampling method are compared with observed SWH along track from satellite Saral. Four tracks in January 2014 are T1: Cycle 10, Pass 197; T2: Cycle 9, Pass 776; T3: Cycle 9; Pass 802; and T4: Cycle 9, Pass 450. In boreal winter, the EAKF and dynamic sampling method have a remarkable advantage over the static sampling method in northern hemisphere westerlies

compared with the control run was also presented. MAE for SWH is given by:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|H_i^{\mathrm{mod}} - H_i^{obs}\right| \tag{5}$$

Here, N is the number of observations.

Each of assimilation schemes manages to produce a more accurate analyzed wave field. MAE that covered the global region has been reduced by about 25% in EAKF and static sampling assimilation runs compared with the control run, while it has reduced by about 28% in the dynamic sampling assimilation run. In regions I and III, MAE of the EAKF scheme is less than that of the static sampling method, while it is reverse in region II. The dynamic sampling method has a better performance than the static sampling method in each statistical region. Moreover, MAE of the dynamic sampling method is closed to that of the EAKF in region I and region III, even better in region II.

Figure 8 compares the control run with three assimilation runs in terms of correlation coefficient. Scattered points

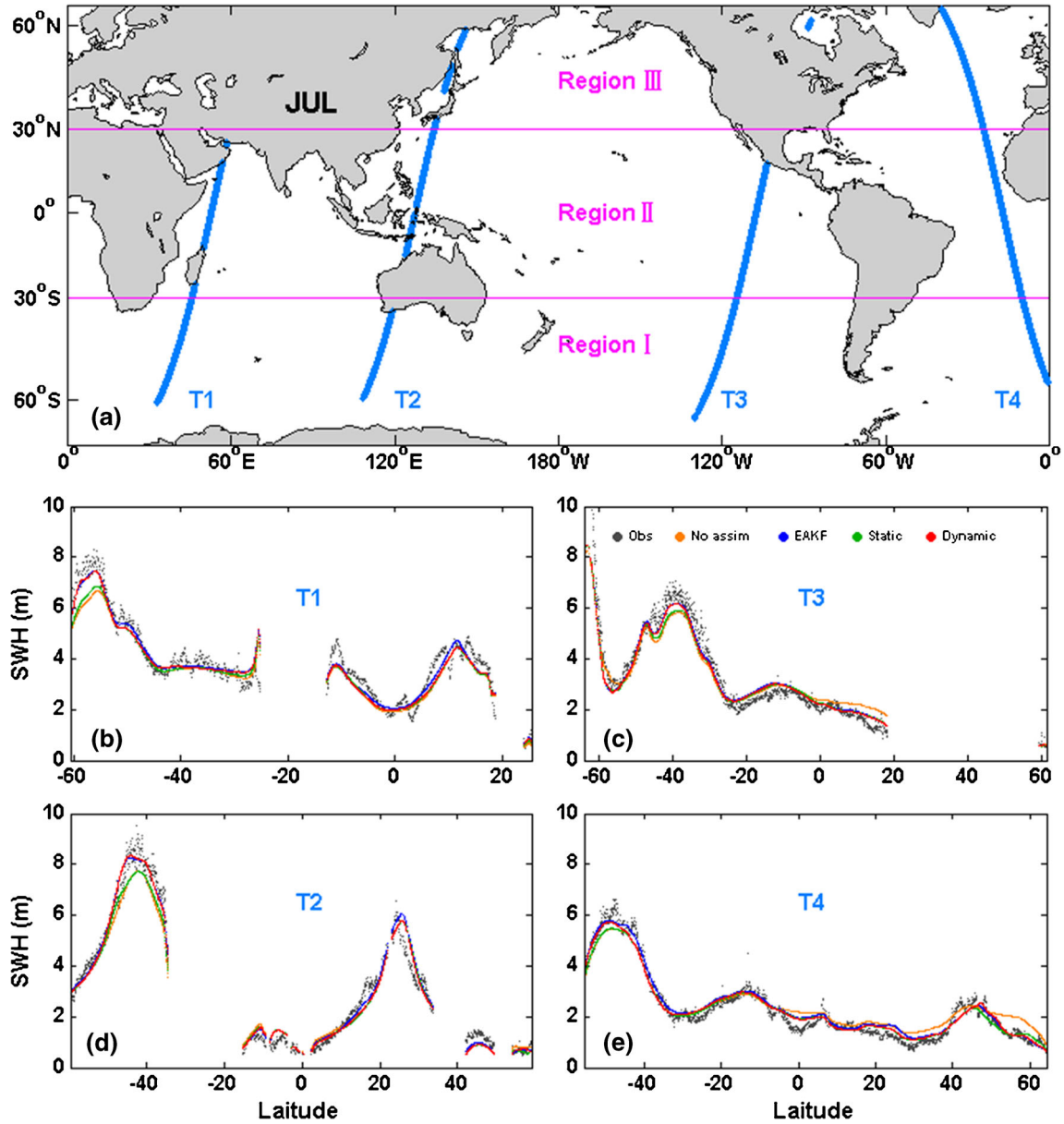

Fig. 7 Comparison between simulated SWH and observed SWH along track in July 2014. Simulated SWH from a control run without assimilation and three assimilation experiments with the EAKF, static sampling method and dynamic sampling method are compared with observed SWH along track from satellite Saral. Four tracks in July 2014 are T1: Cycle 14, Pass 870; T2: Cycle 15, Pass 406; T3: Cycle 14; Pass 940; and T4: Cycle 15, Pass 31. In austral winter, the EAKF and dynamic sampling method have a remarkable advantage over the static sampling method in southern hemisphere westerlies.

Table 1 Mean absolute error comparing simulated SWH with observations from satellite Saral

Experiment	Region I		Region II		Region III		Global	
	MAE	Reduction (%)	MAE	Reduction (%)	MAE	Reduction (%)	MAE	Reduction (%)
No assim	0.444	–	0.400	–	0.447	–	0.418	–
EAKF	0.356	19.8	0.276	31.0	0.344	23.0	0.311	25.6
Static	0.369	16.9	0.259	35.3	0.351	21.5	0.309	26.1
Dynamic	0.361	18.7	0.250	37.5	0.342	23.5	0.301	28.0

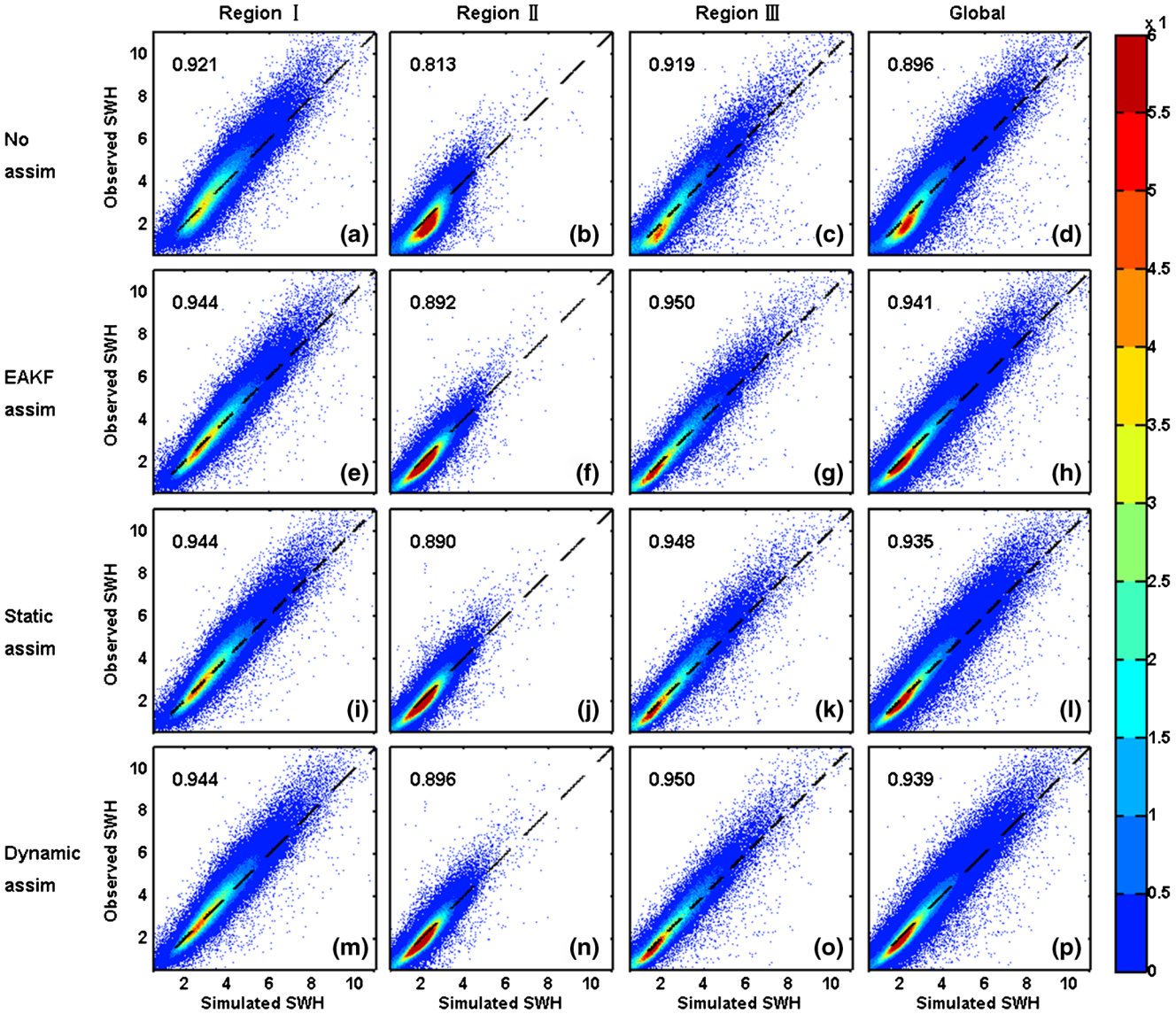

Fig. 8 Scatter diagrams comparing simulated SWH with observed SWH in 2014. Simulated SWH from four experiments and observed SWH from satellite Saral are compared through correlation analysis. Four statistical regions are defined as in Fig.7 region I: 70°S–30°S; region II: 30°S–30°N; region III: 30°N–65°N; and global region: 70°S–65°N. *Color* means the ratio between number of points located at each box of 0.01 m by 0.01 m and the total number of points in each corresponding panel. *Numbers* in the *upper left corner* denote correlation coefficient. Overall, EAKF performs best; the dynamic sampling method is suboptimal

concentrate near the black line mostly, which indicates that the simulated SWH is close to the observed SWH. Fig. 8d, h, l, p shows that correlation coefficient of three assimilation experiments increased by 5.0, 4.4, and 4.8, respectively, compared with the control run. The EAKF scheme has an advantage over the static sampling method, while the dynamic sampling method implemented as an amelioration of the static sampling method performs closed to the EAKF. In addition, model results are less than observations generally in case that observed SWH is higher than 4 m, and it is reverse in case that observed SWH is less than 4 m. Hence, probability density distributions of simulation error of four experiments are statistic within the range of 0.5~2 m, 2~4 m, and higher than 4 m to investigate the difference among EAKF, static, and dynamic sampling methods. The result is shown in Fig. 9. In Fig. 9a, observed SWH is between 0.5~2 m and simulated SWH is larger than observed SWH for the peak of curves which is on the right of zero value. Namely, the wave model without data assimilation overestimates SWH in case of low sea state. On the contrary, Fig. 9c in which observed SWH is higher than 4 m implies that the wave model without data assimilation underestimates SWH in case of high sea state. Fig. 9b shows that the wave model is consistent with observation in general in case that observed SWH ranges from 2 to 4 m. Three experiments with data assimilation reduced simulation error dramatically, especially in low sea state. In addition, under condition of high sea

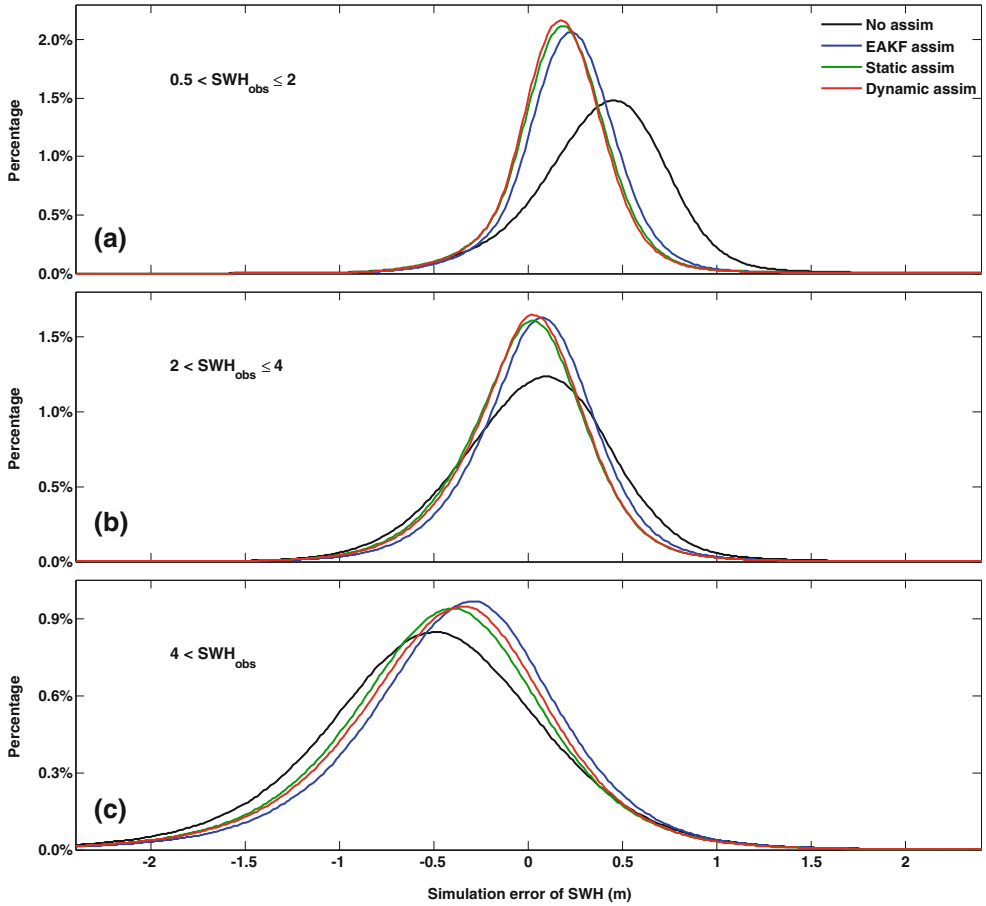

Fig. 9 Probability density distributions of simulation error in the global region over the period 2014. Probability density distributions of SWH simulation error from four experiments compared with observed SWH from satellite Saral are statistic on three intervals for observed SWH: **a** 0.5–2 m, **b** 2–4 m, and **c** higher than 4 m. An interval of axis-x for statistics of probability density distribution is 0.012 m. The static/dynamic sampling method has a better performance than the EAKF method in case that observed SWH is less than 4 m. In case that observed SWH is larger than 4 m, the dynamic sampling method improves the effect of assimilation significantly compared with the static sampling method, and its performance is closed to the EAKF method which is optimal

state, the EAKF has a remarkable advantage over the other two assimilation schemes and the dynamic sampling method is the suboptimal one. Under condition of low sea state, the dynamic sampling method is the optimal scheme.

As a whole, three data assimilation schemes improved the ability of the numerical model significantly. The EAKF scheme performs better than the static/dynamic sampling method in the high sea state, and the dynamic sampling method improves the static sampling method dramatically and performs closed to the EAKF. The reason for that will be further discussed in the next subsection.

4.3 Error distribution characteristics

4.3.1 Error spatial distribution

Spatial distributions of SWH simulation error (MAE) of four experiments with or without data assimilation are shown in Fig. 10. Figure 10a shows distribution of MAE of the control run. It implied that simulation error of the wave model without assimilation is between 0.4~0.8 m mostly. Higher simulation error regions are located in the equator and domain of sea ice. The lower simulation error region is located in Indian Ocean. Distribution of MAE reduction of the EAKF compared with that of thecontrol run is shown in Fig. 10b. It implied that distribution of MAE reduction is consistent with that of MAE of the control run over the global region except westerlies. The reason for that is associated with the proportion of swell to wind waves at specific locations. It is recognized that assimilation of wave observations can improve both the analyzed sea states, especially in swell-dominated situations (Lionello et al. 1995). The proportion is larger in low latitudes than in mid-high latitudes (Chen et al. 2002). During the assimilation process, the whole wave energy spectrum including wind waves and swell is scaled with a ratio. However, wind waves vary rapidly for the impact of the wind forcing field. As a result, it has little ability to remain the information from data assimilation for a long time. Conversely, swell could propagate for a longer distance

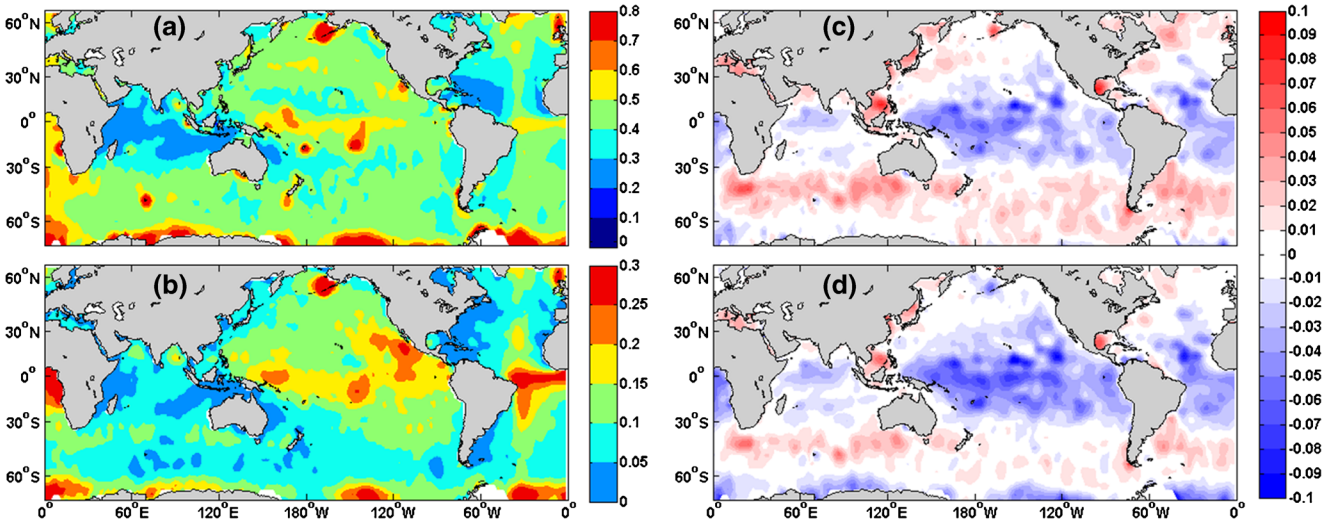

Fig. 10 Spatial distributions of SWH simulation error of experiments with or without data assimilation over the period 2014 (m). Panel **a** represents spatial distribution of MAE of SWH from a control run. The contour interval is 0.1 m. Panel **b** shows spatial distribution of MAE reduction of SWH from experiment with EAKF compared with a control run. The contour interval is 0.05 m. Panels **c** and **d** represent spatial distribution of MAE difference of SWH between the static/dynamic sampling experiment and the EAKF experiment. The contour interval is 0.01 m. The *red region* represents MAE of the static/dynamic sampling method is larger than that of EAKF, the *blue region* represents MAE of the static/dynamic sampling method is less than that of EAKF, the *white region* represents MAE of the static/dynamic sampling method and EAKF is equal. The dynamic sampling method improves the effect of assimilation compared with the static sampling method over the global region, and it performs closed to the EAKF method in westerlies, but better in low latitudes

with assimilation information. Thus, the assimilation effect is more remarkable in low latitudes than in mid-high latitudes. Ensemble spread in westerlies is not small. The wave model is forced by wind dynamically, and assimilation information cannot be remained for a long time as the wind field changes rapidly in westerlies. That is the reason for that EAKF to perform better in low latitude compared to in mid-high latitudes. Larger MAE reduction of SWH in high latitudes around 60°S and 60°N is probably related to energy convergence due to great circle propagation of waves.

Comparison between the static/dynamic sampling method and the EAKF is presented in Fig. 10c, d. The range of difference between the static/dynamic sampling method and the EAKF is about 0.1 m. Namely, difference of results between the static/dynamic sampling method and the EAKF is not very large. The reason for that is the static/dynamic ensemble having comparable characteristic information of background error. The static/dynamic ensemble regarded as an approximation of the background error consisted of 24-h interval difference between simulated SWH from long-term history model results. Namely, the static/dynamic ensemble contains characteristic information of background error in a certain period. Thus, there is characteristic information of background error at an arbitrary location in the static/dynamic ensemble.

In spite of the resemblance among results of three experiments with data assimilation, there is still remarkable difference among them. In Fig. 10c, d, blue regions represent MAE of the static/dynamic sampling method which is less than the EAKF, while red regions represent MAE of the static/dynamic sampling method which is larger than the EAKF. White regions represent MAE of two compared assimilation schemes which are equal. In mid-high latitudes, the EAKF has an advantage over the other two methods, while it is reverse in low latitudes. The reason for that is a spatial correlation scale of the EAKF which is relative localization. The wave model ensemble is constructed by the wind field with perturbation of the random field. However, the influence of perturbation on wind wave and swell is different. The former is more remarkable. Thus, a spatial correlation scale of EAKF is relative localization compared with static/dynamic sampling method, especially in swell-dominated situation. Figure 11 show distributions of spatial correlation of the static/dynamic sampling method (Fig. 11a, b) and EAKF (Fig. 11c), and it verifies the analysis above. On the other hand, a small number of ensembles can appear spurious correlations for random perturbation. The perturbed result of 100 members without assimilation is shown in Fig. 11d. Compared with the perturbed result of 10 members (Fig. 11c), it has less spurious correlation and is more anisotropic. However, in spite of sufficient number of ensembles, the perturbed result (Fig. 11d) is still relatively locally distributed compared with the result of static/dynamic sampling (Fig. 11a, b). Only perturbing wind forcing may not enough for the wave model in some regions, especially where the swell is important. Although the EAKF represents the dynamical evolution of the error covariance matrix, the static/dynamic sampling method outperforms the EAKF in low latitudes. In addition, the static ensemble contains annual mean characteristic information of background error, and it is

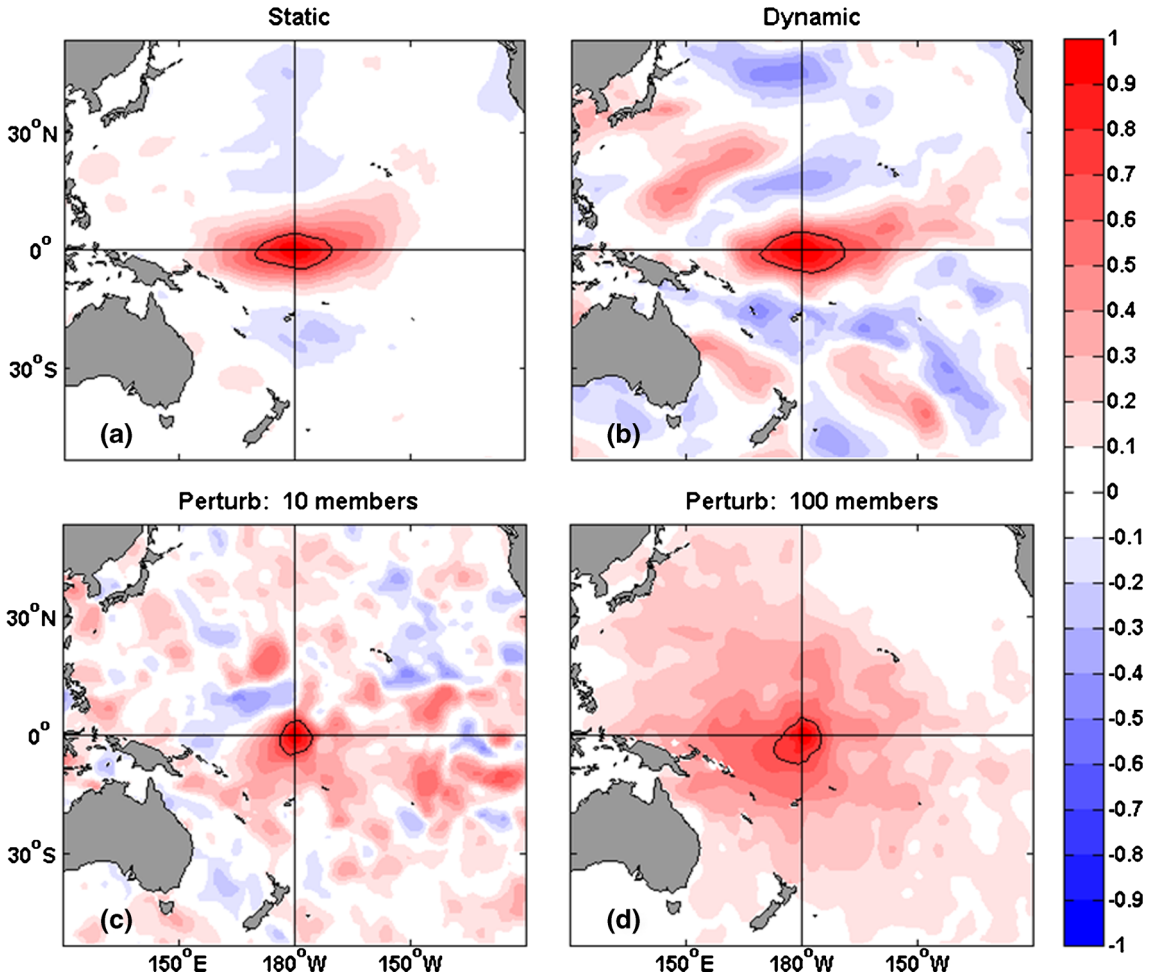

Fig. 11 Spatial distributions of monthly mean correlation coefficient in January 2014. Referred location (180°W, 0°) which is surrounded by *black contour* of 0.7 is marked by *black crosses* in these panels. The contour interval is 0.1

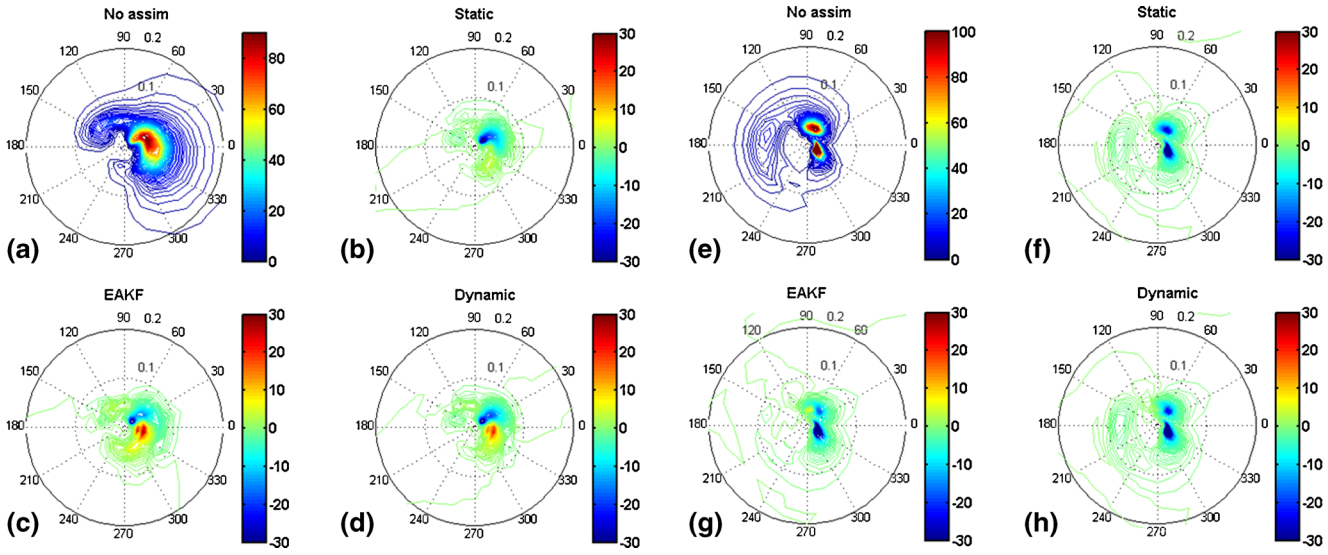

Fig. 12 Spatial distributions of a wave spectrum (m^4). Panels **a**–**d** and **e**–**h** represent wave spectra at 31 December 2014 23:00:00 UTC at position 80°E, 50°S and position 90°W, 20°S, respectively. Panels **a** and **e** represent a wave spectrum without assimilation. The rest panels represent variation of a wave spectrum with assimilation in the EAKF or static/dynamic sampling method, compared to a wave spectrum without assimilation

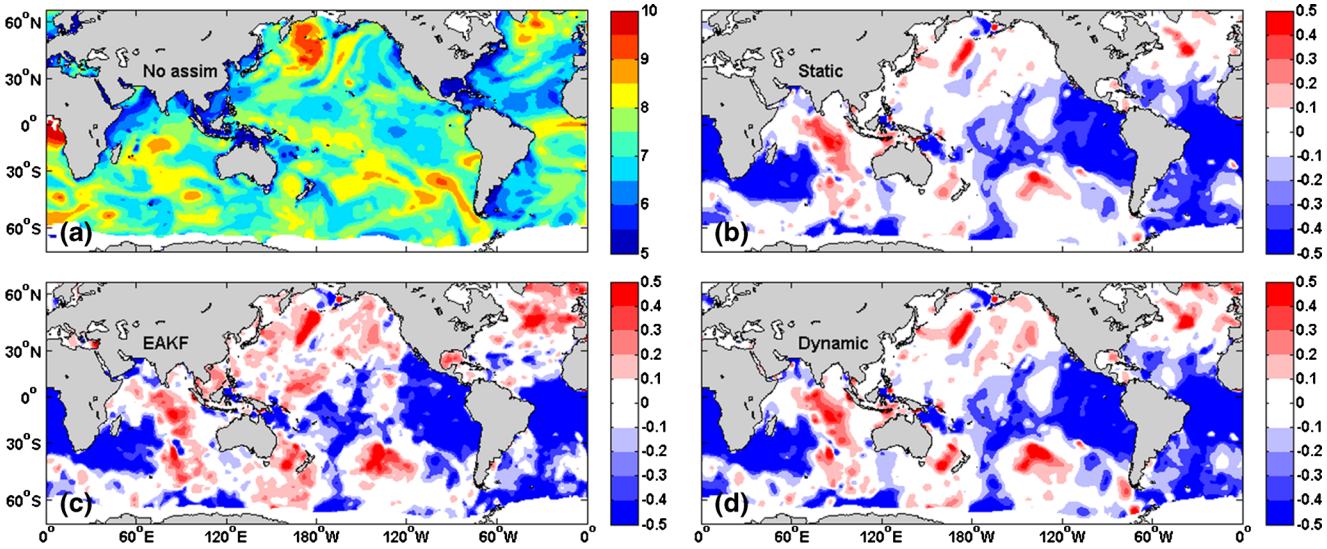

Fig. 13 Spatial distributions of a zero-crossing wave period (s). Panel **a** represents spatial distribution of the zero-crossing wave period without assimilation at 31 December 2014 23:00:00 UTC. The contour interval is 0.5. The rest panels represent variation of the zero-crossing wave period with assimilation in the EAKF or static/dynamic sampling method, compared to the zero-crossing wave period without assimilation. The contour interval is 0.1

unchangeable during the data assimilation process. However, the dynamic ensemble contains seasonal characteristic information of background error and it is changes over time. Thus, compared with the static sampling method, the dynamic sampling method improved the effect of data assimilation on a global scale dramatically. That is the reason the dynamic sampling method is more closed to the EAKF, especially in mid-high latitudes where sea states vary rapidly.

Difference in the wave spectrum among the simulation without data assimilation, the EAKF and static/dynamic method on 31 December 2014, is shown in Fig. 12. Difference in a zero-crossing wave period among the simulations without data assimilation, the EAKF and static/dynamic method at the same time with the wave spectrum, is shown in Fig. 13. Patterns in assimilation results of three methods are comparable and similar. Variation of the wave spectrum or the zero-crossing wave period in the EAKF is the largest among three assimilation schemes. Compared to the static sampling method, the dynamic sampling method is more closed to the EAKF in terms of the wave spectrum or zero-crossing wave period.

4.3.2 Error temporal distribution

Time-dependent curves of MAE reduction percentage of three assimilation experiments compared with the control run in different statistical regions over the period 2014 are shown in Fig. 14. In Fig. 14a, compared with the control run, MAE reduction percentage of SWH of three data assimilation schemes is about 25~30% over the global region. The blue curve representing the EAKF scheme and the green curve representing the static sampling method are almost overlapping sometimes, while the red curve representing the dynamic sampling method is higher than the other two curves over the period 2014. Overall, the dynamic sampling method is the optimal assimilation scheme among them.

The statistical result in low latitudes from 30°S to 30°N is shown in Fig. 14c. MAE reduction percentage of three assimilation experiments changes smoothly around 30~40%. The static sampling method has a slight advantage over the EAKF scheme in region II. The dynamic sampling method performed best in low sea state. The statistical result in mid-high latitudes extending from 30°N to 65°N has a seasonal variation which is shown in Fig. 14b. In boreal winter, MAE reduction percentage of the EAKF scheme containing instantaneous information of background error and the dynamic sampling method containing dynamic information of background error is larger than that of the static sampling method, as the wind forcing field changes rapidly in region III. There is a similar conclusion in southern mid-high latitudes extending from 30°S to 70°S which is shown in Fig. 14d. Namely, in high sea state, the dynamic sampling method can achieve a similar assimilation effect compared with the EAKF.

5 Summary and discussions

Background error which is estimated once in the OI method is a key factor determining the effect of data assimilation. The EnKF method which can update background error over time by using the model variable ensemble is more reasonable. On the basis of EnKF, Anderson (2001, 2003) suggested a method referred as the EAKF. This method which avoided perturbing observations introduced a linear operator as a substitution for traditional gain matrix. The

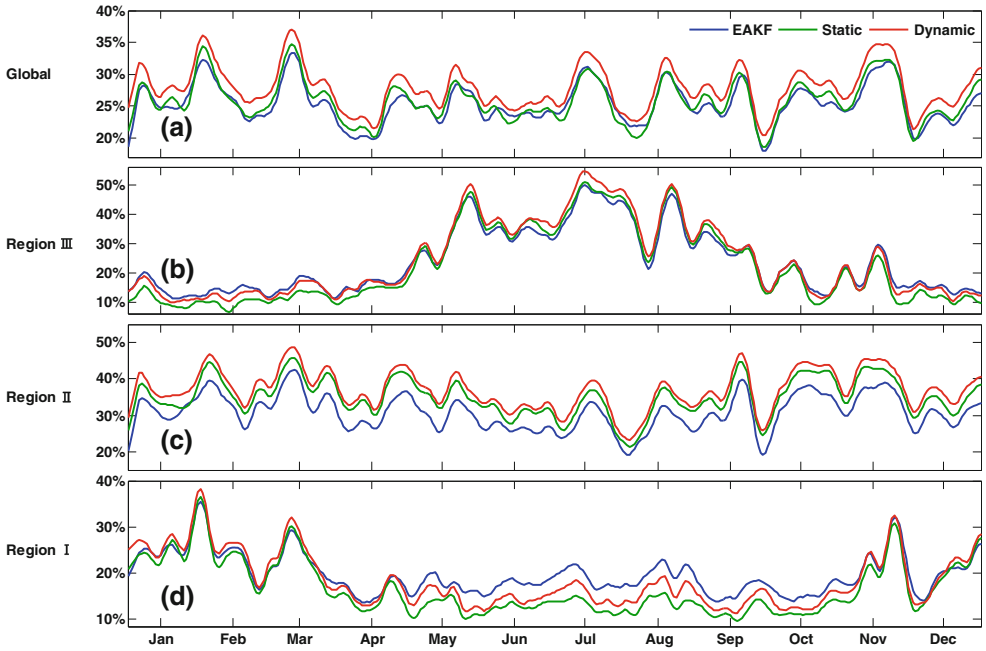

Fig. 14 Time-dependent curves of MAE reduction percentage of SWH of three assimilation experiments compared with a control run over the period 2014. Four statistical regions are defined in Fig. 7. MAE reduction percentage of SWH is calculated for each day in 2014. In low latitude, the static/dynamic sampling method with 365/360 members has an advantage over the EAKF method with 10 members. In mid-high latitude where a wind field changes rapidly, the EAKF method containing instantaneous information of background error performs best and the dynamic sampling method containing statistical significance and seasonal variation information of background error performs better than the static sampling method and closed to the EAKF method especially in the winter of the corresponding region

performance of the EAKF is better than the EnKF in case that the size of the ensemble is relatively small (10~20 members). To achieve "dynamic" statistics of background error and reduce computation cost, we suggested static and dynamic sampling methods for wave data assimilation, in which ensemble errors constructed by 24–h interval difference between simulated SWH from long-term history model results, are used to compose ensemble states when observation is available. The same method of updating model states in the EAKF is applied for ensemble states constructed by ensemble errors. In the static sampling method, the static ensemble which is collected each day for a year from time series of model results without assimilation is used to construct ensemble states and it is not updated over time. In the dynamic sampling method, the dynamic ensemble which is collected by using time series of model results without data assimilation on an interval $(t-7d,\ \ t+7d)$, 24 members for each day; here, t represents time window for data assimilation. Thus, the dynamic ensemble changes over time.

To examine the effect of static and dynamic sampling methods, four global experiments over the period 2014 with a third-generation wave model-MASNUM were carried out. One of them is the control run without data assimilation. The other three experiments with EAKF, static, and dynamic sampling methods respectively assimilated observations from satellite Jason-2. In the EAKF scheme, the random field (Evensen 1994) perturbation is introduced into the wind forcing field to construct the wind field ensemble and then to drive the wave model ensemble. Observations from satellite Saral are used for validation. The results imply that the EAKF method performs best. The static sampling method is relatively worse. The dynamic sampling method improves the assimilation effect compared with the static sampling method. In addition, the dynamic sampling method performs closed to the EAKF in mid-high latitudes, and it has a slight advantage over the EAKF in low latitudes.

The dynamic sampling method has low computation cost for that one wave model is implemented during the data assimilation process. Furthermore, the result of the dynamic sampling method is quite closed to the result of the traditional EAKF scheme in high sea state, and the former is better in low sea state. Thus, it is an efficient and affordable approach for wave data assimilation. In the subsequent work, the dynamic sampling method will be applied into wave operational forecast and construction of history reanalyzed data set. Based on the dynamic sampling method, preliminary experiments with modification of wave spectrum variables will be carried out in further studies by assimilating other observations such as the mean period and SAR data.

Acknowledgments The work is supported by the National High Technology Research and Development Program (2013AA09A506), Public Science and Technology Research Funds Projects of Ocean (201505013), National Programme on Global Change and Air-Sea Interaction (GASI-IPOVAI-06), and the National Key Research and Development Program of China (2016YFC1402000). Thanks go to two anonymous reviewers for their thorough examination and comments that were very useful for improving the manuscript.

References

Anderson JL (2001) An ensemble adjustment Kalman filter for data assimilation. Mon Weather Rev 129(12):2884–2903

Anderson JL (2003) A local least squares framework for ensemble filtering. Mon Weather Rev 131(4):634–642

Burgers G, Jan van Leeuwen P, Evensen G (1998) Analysis scheme in the ensemble Kalman filter. Mon Weather Rev 126(6):1719–1724

Chen G, Chapron B, Ezraty R et al (2002) A global view of swell and wind sea climate in the ocean by satellite altimeter and scatterometer. J Atmos Ocean Technol 19(11):1849–1859

Esteva D (1988) Evaluation of preliminary experiments assimilating Seasat significant wave heights into a spectral wave model. J Geophys Res 93(C11):14099–14105

ETOP5 (1986) 5′ × 5′ topography and elevation. Marine Geology and Geophysics Division, National Geophysical Data Center. (Available from National Geophysical Data Center, NOAA, Code E/GC3, Boulder, CO 80303)

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans 99(C5): 10143–10162

Evensen G (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn 53(4):343–367

Greenslade DJM (2001) The assimilation of ERS-2 significant wave height data in the Australian region. J Mar Syst 28(1):141–160

Greenslade D J M, 2004. The structure of the background errors in a global wave model[D].

Hasselmann S, Lionello P, Hasselmann K (1997) An optimal interpolation scheme for the assimilation of spectral wave data. Journal of Geophysical Research: Oceans 102(C7):15823–15836

Karspeck AR, Anderson JL (2007) Experimental implementation of an ensemble adjustment filter for an intermediate ENSO model. J Clim 20(18):4638–4658

Komen GJ, Cavaleri L, Donelan M et al (1996) Dynamics and modelling of ocean waves[M]. Cambridge university press, Cambridge

Lionello P, Günther H, Janssen PAEM (1992) Assimilation of altimeter data in a global third-generation wave model. Journal of Geophysical Research: Oceans 97(C9):14453–14474

Lionello P, Günther H, Hansen B (1995) A sequential assimilation scheme applied to global wave analysis and prediction. J Mar Syst 6(1):87–107

Lorenc AC (1981) A global three-dimensional multivariate statistical interpolation scheme. Mon Weather Rev 109(4):701–721

Ren QF (2010) Statistical analysis of background error covariance structure and Envisat ASAR wave spectrum data assimilation research. Institute of Oceanology, Chinese Academy of Sciences, Qingdao (in Chinese)

Sun M, Yin XQ, Yang YZ (2014) Construction and application in global wave data assimilation of static sample set. Oceanologia et Limnologia Sinica 45(5):918–927

Sun M, Yin X Q, Yang Y Z, et al., 2016. Research on EnKF data assimilation based on MASNUM-WAM-I. Influence on wave simulation of ensemble-disturbance wind field. Oceanologia et Limnologia Sinica. (accepted for publication)

Voorrips AC, Makin VK, Hasselmann S (1997) Assimilation of wave spectra from pitch-and-roll buoys in a North Sea wave model. Journal of Geophysical Research: Oceans 102(C3):5829–5849

Waters J, Wyatt LR, Wolf J et al (2013) Data assimilation of partitioned HF radar wave data into Wavewatch III. Ocean Model 72:17–31

Yang YZ, Qiao FL, Zhao W et al (2005) MASNUM ocean wave numerical model in spherical coordinates and its application. Acta Oceanol Sin 27(2):1–7

Yin XQ, Qiao FL, Yang YZ et al (2010) An ensemble adjustment Kalman filter study for Argo data. Chin J Oceanol Limnol 28:626–635

Yin XQ, Qiao FL, Shu Q (2011) Using ensemble adjustment Kalman filter to assimilate Argo profiles in a global OGCM. Ocean Dyn 61(7):1017–1031

Yuan YL, Tung CC, Huang NE (1986) Statistical characteristics of breaking waves[M]//wave dynamics and radio probing of the ocean surface. Springer US:265–272

Yuan YL, Pan ZD, Hua F et al (1992a) LAGFD-WAM wave numerical model (I), the basic physical model. Acta Oceanol Sin 14(5):1–7

Yuan YL, Hua F, Pan ZD et al (1992b) LAGFD-WAM numerical wave model-II. Characteristics inlaid scheme and its application. Acta Oceanol Sin 14(6):12–24

Zhang S, Harrison MJ, Wittenberg AT et al (2005) Initialization of an ENSO forecast system using a parallelized ensemble filter. Mon Weather Rev 133(11):3176–3201

Zhang S, Harrison MJ, Rosati A et al (2007) System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. Mon Weather Rev 135(10):3541–3564