Tellus
SERIES A
DYNAMIC
METEOROLOGY
AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2011), 63A, 244–255

Printed in Singapore. All rights reserved

TELLUS

A new localization implementation scheme for ensemble data assimilation of non-local observations

By JIANG ZHU[1], FEI ZHENG[2]* and XICHEN LI[2], [1]*LAPC, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China;* [2]*ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China*

(Manuscript received 17 December 2009; in final form 27 August 2010)

ABSTRACT

Localization technique is commonly used in ensemble data assimilation of small-size ensemble members. It effectively eliminates the spurious correlations of the background and increases the rank of the system. However, one disadvantage in current localization schemes is that it is difficult to implement the assimilation of non-local observations. In this paper, we test a new localized implementation scheme that can directly assimilate non-local observations without pinpointing them. A classical local support correlation function matrix is first sampled by a set of local correlation function ensemble members (the size is M). Then, the dynamical ensemble (the size is N) is combined with the local correlation function ensemble to form an $N \times M$ ensemble by multiplying each dynamical member with each local correlation function member using the Schur product. The covariance matrix constructed by the $N \times M$ members is proved to approximate the Schur product of the local support correlation matrix and the dynamical covariance matrix. This scheme is verified through assimilating both local and non-local observations with a linear advection model and an intermediate coupled model. The analysis results show that this scheme is feasible and effective in providing reasonable and high-quality analysis fields with a relatively small dynamical ensemble size.

1. Introduction

The accuracy of the background-error covariance matrix is a crucial factor of systems used to assimilate data for weather and climate predictions, and the ensemble-based assimilation methods depend on a set of ensemble forecasts to calculate the background-error covariances. However, the ensemble size is always significantly smaller than the dimension of the system for realistic numerical predictions, and then the imperfectly estimated covariances provide insufficient background-error information to result in a relatively poor performance or even the occurrence of filter divergence for the ensemble-based assimilation system. A possible solution to perform a successful ensemble-based assimilation when only a small-sized ensemble is feasible is the use of a technique called localization (e.g. Houtekamer and Mitchell, 1998; Keppenne, 2000; Anderson, 2001; Hamill et al., 2001; Houtekamer and Mitchell, 2001, 2005; Whitaker and Hamill, 2002; Ott et al., 2004; Anderson, 2007; Oke et al., 2007; Bishop and Hodyss, 2007, 2009a,b; Liu et al., 2009). Localization is a simple technique to 'localize' the impact of an observation to a subset of the model state variables. The subset is usually defined as those spatially close to the location of the observation. One of the implementations of localization is the distance-dependent covariance localization (Houtekamer and Mitchell, 2001; Hamill et al., 2001), which is done by updating the analysis at all grid points within a predefined correlation length from each observation. Another implementation is the local ensemble transform Kalman filter (LETKF; Ott et al., 2004). In LETKF, the analysis is performed at each grid point simultaneously using the state variables and all observations in the local region centred at that point. The localization in spectral space has also been tested by Buehner and Carron (2007).

There are two benefits of localization demonstrated by previous research works. One is that localization can eliminate spurious correlations in the background ensemble between distant state variables due to using only a limited number of ensemble members. The other is that localization can effectively increase the rank of the system by increasing the effective number of independent ensemble members (Ott et al., 2004; Oke et al., 2007). However, localization also has some disadvantages. Physical balances between various physical variables present in the unmodified ensemble may be disturbed by localization (Oke et al., 2007). Assimilation of non-local observations (e.g. satellite radiance data) cannot be readily implemented in the presence of localization, although this problem can be partly solved by updating the state at a given location by assimilating non-local

*Corresponding author.
e-mail: zhengfei@mail.iap.ac.cn

DOI: 10.1111/j.1600-0870.2010.00486.x

observations that are strongly correlated to the model state there (Fertig et al., 2007) or by assimilating the retrieved local observations from the non-local observations.

Here, we test a new localization implementation scheme of which the basic idea is similar to that of Liu et al. (2009) with focusing on its performance on assimilating non-local observations. We describe the scheme in Section 2. The scheme is tested in Section 3 by a linear advection (LA) model and by some realistic sea surface temperature (SST) data assimilation experiments using an intermediate coupled model (ICM) of the tropical Pacific. The non-local 'observations' used are pseudo, that is generated by averaging local observations over some spatial sub-domains. Conclusions are given in Section 4. An important implementation issue is the handling of the large size of the combined ensemble (i.e. 1000–2000). Here we utilize a memory-saving algorithm at the update stage without storing the whole ensemble and the memory requirement can be kept within a practical level. This algorithm is described in the Appendix B.

2. Scheme

In this section, we illustrate the details of the new localization scheme and its implementation. Table 1 describes several notations used in this paper.

2.1. Distance-dependent covariance localization

Write **x,** a column vector, for a model state vector with a dimension n. In the ensemble data assimilation algorithm, the background covariance matrix $\mathbf{P}^b$ is calculated from a set of dynamical ensemble members $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)}$ (with their mean denoted by $\bar{x}$) as follows:

$$\mathbf{P}^b = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}^{(i)}-\bar{\mathbf{x}})(\mathbf{x}^{(i)}-\bar{\mathbf{x}})^{\mathrm{T}} = \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{x}'^{(i)}\mathbf{x}'^{(i)\mathrm{T}}. \tag{1}$$

The gain matrix is then

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}. \tag{2}$$

Here, $\mathbf{H}$ is the observation operator (for the linear case, an $m \times n$ matrix if the observation space is m-dimensional) and $\mathbf{R}$ is the observational error covariance (an $m \times m$ matrix).

Table 1. Glossary of each notation

Notation	Description
n	Model dimension
m	Observation dimension
N	Size of dynamical ensemble
M	Size of local correlation ensemble
L	De-correlation scale

When N is small (e.g. $N = 20$ for typical atmospheric and oceanic applications), the covariance defined by (1) can suffer from sampling errors. The current standard distance-dependent covariance localization replaces the background error covariance by a Schur product (an element-by-element multiplication), $\rho_{\mathrm{s}} \circ \mathbf{P}^b$. Here, ρ_{s} is an $n \times n$ local support correlation matrix. Gaspari and Cohn (1999) gave various examples of such functions. Then the gain matrix becomes

$$\mathbf{K} = (\rho_{\mathrm{s}} \circ \mathbf{P}^b)\mathbf{H}^{\mathrm{T}}\left[\mathbf{H}(\rho_{\mathrm{s}} \circ \mathbf{P}^b)\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right]^{-1}. \tag{3}$$

As demonstrated in Houtekamer and Mitchell (2001), the observation operator $\mathbf{H}$ involves operations on vertical columns of grid points and/or interpolations or finite differences on the horizontal analysis grid, but ρ_{s} is a relatively broad function with some compactly supported correlation functions presented in Gaspari and Cohn (1999). Thus, one strategy is to rewrite eq. (3) using the following gain matrix equation (Houtekamer and Mitchell, 2001, 2005):

$$\mathbf{K} = \rho_{\mathrm{s}}' \circ (\mathbf{P}^b\mathbf{H}^{\mathrm{T}})\left[\rho_{\mathrm{s}}'' \circ (\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}}) + \mathbf{R}\right]^{-1}. \tag{4}$$

Here, ρ_{s}' is an $n \times m$ local support correlation matrix in which each column represents correlations at an observation location with all model grid points. Similar to ρ_{s}, ρ_{s}'' is an $m \times m$ local support correlation matrix in which each column represents correlations at an observation location with all other observation locations. ρ_{s}' and ρ_{s}'' share the same spatial function with ρ_s. Formally, they should be defined as $\rho_{\mathrm{s}}' = \rho_{\mathrm{s}}\mathbf{H}^{\mathrm{T}}$ and $\rho_{\mathrm{s}}'' = \mathbf{H}\rho_{\mathrm{s}}\mathbf{H}^{\mathrm{T}}$.

The reason for modifying eq. (3) by eq. (4) was pointed out by Houtekamer and Mitchell (2001), in which the order of the observation operator $\mathbf{H}$ and the Schur product can be changed because the observation operator $\mathbf{H}$ involves operations on vertical columns of grid points and/or interpolations or finite differences on the horizontal analysis grid, while ρ_{s} is a relatively broad function. Therefore, the essential underlying assumption is that $\mathbf{H}$ can only be a 'local' operation on vertical columns of grid points comparing to ρ_{s}. In other words, if the observations are not local, the gain matrix given by eq. (4) is no longer a valid equivalence of eq. (3), and an example of such is given in Appendix A.

For local observations, based on eq. (4), one can carry out the serial processing algorithm as described by Whitaker et al. (2004) or Whitaker et al. (2008). In serial processing algorithms, one processes observations serially (one-by-one or batch-by-batch) to update the analysis based on the assumptions that observational errors are uncorrelated and there is independence between background and observational errors.

2.2. Sampling the local support correlation matrix

The local support correlation matrix ρ_{s} can be sampled by a set of ensemble members. Through adopting the fast Fourier transform (FFT) approach, Evensen (2003) developed a Monte Carlo sampling algorithm that can be used to produce a set of smooth pseudorandom fields, which were used for generating the initial ensemble, the model noise, and the observation perturbations (Evensen, 2003, 2004). With a predefined specific de-correlation length L, an ensemble of pseudorandom fields can be created with a mean equal to zero, a variance equal to one and a specified covariance. More details and explanations of this approach are given in Appendix E of Evensen (2003). Evensen (2007) further systematically discussed efficient ways to generate a set of ensemble to present the model/observation error covariance defined by a Gaussian function with an e-folding scale of L.

In this study, we also use such an approach to present the local support correlation matrix ρ_s. The random ensemble with zero mean generated using the method of Evensen (2003, 2007) are denoted by $\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \ldots, \mathbf{s}^{(M)}$ and are defined in the model space. Therefore, ρ_s can be approximated by

$$\rho_s \approx \frac{1}{M-1} \sum_{j=1}^{M} \mathbf{s}^{(j)} \mathbf{s}^{(j)\mathrm{T}}. \tag{5}$$

The size of the local correlation ensemble M can be relatively small, as pointed out by Evensen (2007), if a sophisticated sampling strategy is applied. We will illustrate this in Section 3.

2.3. *Combining the dynamical ensemble and the local correlation ensemble*

Liu et al. (2009) showed that the localized covariance matrix $\rho_s \circ \mathbf{P}^b$ can be approximated by combining the dynamical ensemble $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)}$ and some local correlation ensemble $\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \ldots, \mathbf{s}^{(M)}$. Similarly to the Schur product of two matrices (i.e. an element-by-element multiplication, is also meaningful for two vectors of the same dimension). In the following we show that the $N \times M$ ensemble,

$$\begin{aligned}\mathbf{C} &= (\mathbf{s}^{(j)} \circ \mathbf{x}'^{(i)};\ \ i = 1, \ldots, N;\ \ j = 1, \ldots, M) \\ &= (\mathbf{s}^{(1)}\mathbf{x}'^{(1)}, \ldots, \mathbf{s}^{(1)}\mathbf{x}'^{(N)}, \mathbf{s}^{(2)}\mathbf{x}'^{(1)}, \ldots, \mathbf{s}^{(2)}\mathbf{x}'^{(N)}, \ldots, \\ &\qquad \mathbf{s}^{(M)}\mathbf{x}'^{(1)}, \ldots, \mathbf{s}^{(M)}\mathbf{x}'^{(N)}),\end{aligned} \tag{6}$$

can approximate the localized covariance matrix $\rho_s \circ \mathbf{P}^b$ through

$$\begin{aligned}\mathbf{CC}^{\mathrm{T}} &= \sum_{i=1}^{N}\sum_{j=1}^{M} (\mathbf{s}^{(j)} \circ \mathbf{x}'^{(i)})(\mathbf{s}^{(j)} \circ \mathbf{x}'^{(i)})^{\mathrm{T}} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{M} (\mathbf{s}^{(j)}\mathbf{s}^{(j)\mathrm{T}}) \circ (\mathbf{x}'^{(i)}\mathbf{x}'^{(i)\mathrm{T}}) \\ &= \left(\sum_{j=1}^{M} \mathbf{s}^{(j)}\mathbf{s}^{(j)\mathrm{T}}\right) \circ \left(\sum_{i=1}^{N} \mathbf{x}'^{(i)}\mathbf{x}'^{(i)\mathrm{T}}\right) \\ &\approx (M-1)(N-1)\rho_s \circ \mathbf{P}^b\end{aligned} \tag{7}$$

Here, we used the property of $(\mathbf{y} \circ \mathbf{z})(\mathbf{y} \circ \mathbf{z})^{\mathbf{T}} = (\mathbf{yy}^{\mathbf{T}}) \circ (\mathbf{zz}^{\mathbf{T}})$ for any two vectors $\mathbf{y}$ and $\mathbf{z}$ of the same dimension. This property can be proved straightforwardly by writing out both sides.

Using eq. (8), we can calculate the gain matrix as follows:

$$\mathbf{K} = \mathbf{C}\,(\mathbf{HC})^{\mathrm{T}} \left[\mathbf{HC}\,(\mathbf{HC})^{\mathrm{T}} + (N-1)(M-1)\mathbf{R}\right]^{-1}. \tag{8}$$

The increased rank of the system made by the localization can be easily seen from eq. (7). If one chooses N dynamical ensemble members and M local correlation function ensemble members, the independent number of ensemble members from the resulting $N \times M$ ensemble members is the rank made by the localization. The sampling algorithm of local support correlation matrix ρ_s in this study as described earlier is slightly different from Liu et al. (2009) in which an empirical orthogonal function decomposed correlation function operator was introduced to modify the background error covariance.

2.4. *Implementation and memory saving*

In practical implementation for ensemble data assimilation, first, we use the dynamical forecast ensemble and the local correlation ensemble to construct the matrix $\mathbf{C}^f$ and $\mathbf{HC}^f$:

$$\begin{aligned}\mathbf{C}^f &= (\mathbf{s}^{(j)} \circ \mathbf{x}'^{f(i)};\ \ i = 1, \ldots, N;\ \ j = 1, \ldots, M) \\ \mathbf{HC}^f &= (\mathbf{H}(\mathbf{s}^{(j)} \circ \mathbf{x}'^{f(i)});\ \ i = 1, \ldots, N;\ \ j = 1, \ldots, M).\end{aligned} \tag{9}$$

Here, the ensemble state is $\mathbf{X} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)})$, and the ensemble perturbation matrix was defined as $\mathbf{X}' = \mathbf{X} - \overline{\mathbf{X}}$. The superscript f denotes forecast.

Then the analysis equation for traditional EnKF (e.g. Evensen, 2003) can be easily rewritten as

$$\begin{aligned}\mathbf{X}^a &= \mathbf{X}^f + \mathbf{C}^f(\mathbf{HC}^f)^{\mathrm{T}} \\ &\quad \times [\mathbf{HC}^f(\mathbf{HC}^f)^{\mathrm{T}} + (N-1)(M-1)\mathbf{R}]^{-1}(\mathbf{Y} - \mathbf{HX}^f).\end{aligned} \tag{10}$$

Here, $\mathbf{Y}$ represents the ensemble of perturbed measurements. The superscripts f and a denote the forecast and analysis, respectively.

Meanwhile, for the ensemble square root filter (EnSRF; e.g. Evensen, 2004), the ensemble mean and the anomalies are updated separately by

$$\begin{aligned}\overline{\mathbf{X}}^a &= \overline{\mathbf{X}}^f + \mathbf{C}^f(\mathbf{HC}^f)^{\mathrm{T}} \\ &\quad \times [\mathbf{HC}^f(\mathbf{HC}^f)^{\mathrm{T}} + (N-1)(M-1)\mathbf{R}]^{-1}(\mathbf{y} - \mathbf{H}\overline{\mathbf{X}}^f).\end{aligned} \tag{11}$$

Here, $\mathbf{y}$ is the unperturbed measurement vector.

$$\begin{aligned}\mathbf{X}'^a &= \mathbf{C}^f\mathbf{Z}\sqrt{\mathbf{I} - \Lambda}\,\mathbf{\Theta}^{\mathrm{T}} \\ \mathbf{Z}\Lambda^{-1}\mathbf{Z}^{\mathrm{T}} &= (\mathbf{HC}^f)^{\mathrm{T}}(\mathbf{HC}^f(\mathbf{HC}^f)^{\mathrm{T}} + (N-1)(M-1)\mathbf{R})^{-1}\mathbf{HC}^f.\end{aligned} \tag{12}$$

Here, the eigenvectors $\mathbf{Z}$ and eigenvalues $\mathbf{\Lambda}$ were defined in Evensen (2004), $\mathbf{\Theta}^{\mathrm{T}}$ is a random orthogonal matrix ($\mathbf{I} = \mathbf{\Theta}^{\mathrm{T}}\mathbf{\Theta}$), and multiplication with $\mathbf{\Theta}^{\mathrm{T}}$ is equivalent to a random rotation of the eigenvectors about $\mathbf{Z}$.

However, using eq. (10) or eqs. (11)–(12) for the analysis update, the memory requirement to store the entire ensemble matrix **C** ($n \times M \times N$-dimensional) can be large. Appendix B provides a memory-saving algorithm at the update stage without storing the whole ensemble matrix **C**.

3. Experiments

In this section, we compare the performance of the new localization implementation scheme for two models, namely the LA model of Evensen (2004) and the ICM (Zhang et al., 2005). We first give a detailed description of the models and the configurations used in the tests, and then we describe the results. We compare the new localization implementation scheme with the traditional covariance localization scheme by the LA model with local observations. In addition, we further test the new localization scheme by assimilating some realistic (local) and artificial (non-local) SST data into the ICM model. Both the local and non-local data assimilation experiment results are presented. We implement all experiments with the LA and ICM models by using the ensemble square root filter (EnSRF) scheme without perturbations of observations (e.g. Evensen, 2004), and the scheme was upgraded to use mean preserving transformations in the square root scheme (Sakov and Oke, 2008a).

3.1. LA model

The LA model, described later, is based on that of Evensen (2004) and Sakov and Oke (2008b). The dimension of the state vector **x** is 1000; the signal propagates (advects) in the positive direction by one element at each time-step without changing its shape, and the model domain is periodic:

$$\mathbf{x}(t) = [\mathbf{x}_1(t), \ldots, \mathbf{x}_{1000}(t)], \quad t = 1, 2, \ldots;$$

$$\mathbf{x}_i(t+1) = \begin{cases} \mathbf{x}_{i-1}(t), & i = 2, \ldots, 1000; \\ \mathbf{x}_{1000}(t), & i = 1. \end{cases} \tag{13}$$

Here, $\mathbf{x}_i(t)$ is the ith component of the state vector at the tth time step.

The true initial state is sampled from a standard normal distribution, with a mean equal to 0, a variance equal to 1 and a spatial de-correlation length of 20. The first guess solution is generated by drawing another sample from the standard normal distribution and adding this to the true state. The initial ensemble is then generated by adding samples drawn from the standard normal distribution to the first guess solution. Thus, the initial state is assumed to have an error variance equal to 1.

Four observations of the true field are conducted and assimilated into the model at every fifth time step, $t = 1, 6, 11, \ldots,$ at equidistant locations $i = \{125, 375, 625, 875\}$. Each observation is contaminated with random normally distributed uncorrelated noise with a variance of 0.01.

3.2. ICM model

The more complex model we used here is an ICM that was first developed by Keenlyside and Kleeman (2002) and Zhang et al. (2005). Its dynamical component consists of both linear and non-linear components. The former was essentially a McCreary-type (1981) modal model but was extended to include horizontally varying background stratification. In addition, 10 baroclinic modes, along with a parameterization of the local Ekman-driven upwelling, were included. A SST anomaly model was embedded within this dynamical framework. The governing equation described the evolution of the mixed-layer temperature anomalies. As demonstrated by Zhang et al. (2005), having a realistic parameterization for the temperature of the subsurface water entrained into the mixed-layer (T_e) is crucial to the performance of SST simulations in the equatorial Pacific. An empirical T_e model was constructed from historical data and was demonstrated to be effective in improving the SST simulations. The ocean model was coupled with a statistical atmospheric model that specifically relates wind stress (τ) to SST anomaly fields. All coupled-model components exchange simulated anomaly fields. Information concerning the interactions between the atmosphere (τ) and the ocean (SST) was exchanged once a day.

An ensemble data assimilation system for this model was firstly developed by Zheng et al. (2006, 2007). In the system, the ensemble data assimilation is implemented using EnSRF algorithm without perturbations of observations (e.g. Evensen, 2004). The assimilation scheme was further improved by using a balanced multivariate model-error approach (Zheng and Zhu, 2008), and upgraded to use mean preserving transformations in the square root scheme (Sakov and Oke, 2008b). This balanced model-error approach was verified that it not only can generate accurate and dynamically consistent initial conditions for the model, but also provide reasonable initial stochastic uncertainties by combining both background and observation errors during the assimilation cycles.

The data used in this study includes the daily SST observations from the Tropical Atmosphere Ocean (TAO) array and the monthly OI.v2 SST data (Reynolds et al., 2002).

3.3. Assimilation experiments of the LA model with local observations

The results using the LA model are shown in Fig. 1. The behavior of the root mean square (RMS) errors of the analyses and the ensemble spread during the initial stage of one particular realization of the system are shown. Here, we define the ensemble spread as the ensemble mean of the root mean squared deviation of the anomalies. The LA model runs are conducted with a series of experiments with different ensemble configurations. Four experiments with different combinations of N and M: (i.e. $N = 20$ & $M = 20$, $N = 20$ & $M = 40$, $N = 40$ & $M = 20$, and

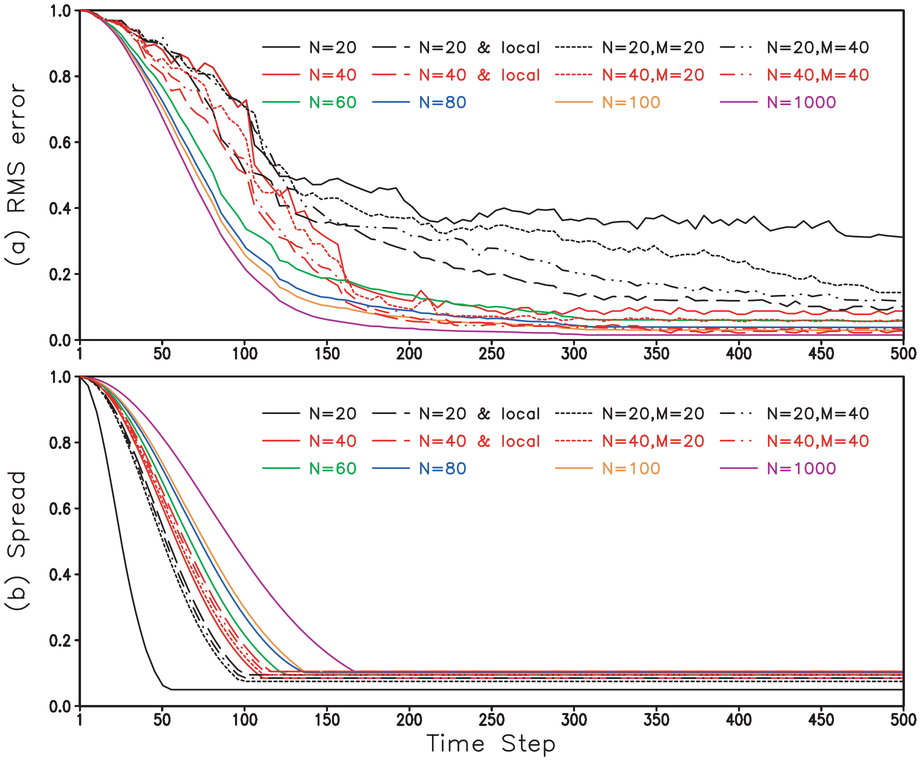

Fig. 1. Time series of the (a) RMS errors and (b) spread of EnSRF assimilation experiments for the LA model.

$N = 40$ & $M = 40$) are carried out to verify the capabilities of the new localization scheme. Two experiments with 20 and 40 dynamical members are performed with the traditional covariance localization scheme to compare the performance of the new localization scheme during the assimilation process. The EnSRF also runs with 20, 40, 60, 80, 100 and 1000 members and without localization. Also, the de-correlation scale in both the local correlation ensemble (i.e. see Section 2.2) and the local covariance matrix (i.e. Gaussian localization function) is set as 10 grids. At the same time, inflation is not considered in these experiments, so these experiments should result in a relatively small residual. And in order to reduce the effect of random sampling error, we perform all experiments of LA model five times.

When the EnSRF runs with 20 dynamical ensemble members, the experiment without localization can observe the filter collapse, whereas the experiments with two different localization schemes can somewhat alleviate the filter divergence. Besides these experiments with 20 dynamical ensemble members, after the initial transient period, for all other experiments with or without applying localization, the RMS errors and the ensemble spread become approximately equal. Also, the two kinds of localization experiments with 40 dynamical ensemble members can achieve similar assimilation results as the traditional EnSRF experiments with greater than 60 dynamical ensemble members. This demonstrates a consistency in this particular case between the actual analysis error covariance (represented by RMS error) and its representation by the ensemble for relatively large dynamical ensemble members or considering the localization approach. In particular, based on the LA model and local observations, the performances of the EnSRF with covariance localization and the EnSKF with the combined ensemble are very comparable, with a slightly smaller residual achieved by the EnSRF with covariance localization for the LA model. This is due to the sampling errors introduced by the local correlation ensemble with small size, and this will be further illustrated by Fig. 3.

Table 2 shows the mean value of the RMS errors over the time interval $t = [1, 500]$ versus the dynamical ensemble size N and local correlation ensemble size M. For small size N, as M increases, the performance of the combined ensemble improves. The performance of the combined ensemble $N = 40$ & $M = 40$ is almost identical to that of the traditional EnSRF with 80 dynamical ensemble, but the experiment of the combined ensemble $N = 40$ & $M = 40$ can reduce the model integration time by almost half and will also not induce a significantly additional memory requirement, through adopting the memory-saving algorithm in Appendix B. This will be further discussed in Section 3.4 for the ICM model, and the ability of the new localization scheme to assimilate "non-local" observations will be shown in Section 3.5.

Table 2. Time-averaged RMS errors of the assimilation results for the LA model

		Dynamical ensemble					
		$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$	$N = 1000$
Traditional EnSRF		0.491	0.294	0.245	0.221	0.196	0.172
EnSRF with covariance localization		0.329	0.224	–	–	–	–
EnSRF combined with the local correlation ensemble	$M = 20$	0.426	0.266	–	–	–	–
	$M = 40$	0.346	0.235	–	–	–	–

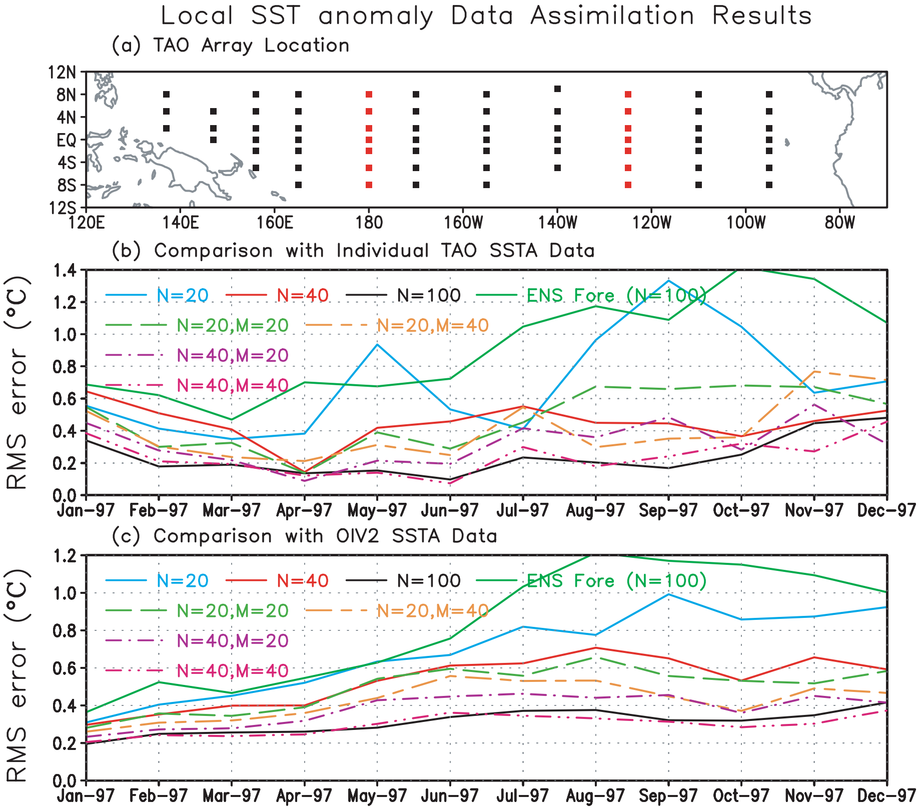

Fig. 2. (a) The locations of the *in situ* TAO SST data. Observations at the black squares are assimilated, while observations at the red squares are used to verify assimilation results as the independent data; (b) RMS errors of experiments verified using the independent TAO SST data. The green solid line denoted by 'ENS Fore ($N = 100$)' is the RMS error of the mean of the free run ensemble forecasts without data assimilation. See text for details; (c) the same as (b), but the RMS errors are calculated using the OI.v2 SST data set over the entire model region.

3.4. Assimilation experiments of the ICM model with local observations

The ICM adopted here further verifies the capability of the new localization scheme in a more complex model. Several local data assimilation experiments are performed over the period January 1997–December 1997 with assimilated monthly averaged *in situ* TAO SST observations (black square points in Fig. 2a). The experiments are performed in a SGI Origin 2000 server with 8 CPUs and 4-GB memories, and the memory-saving algorithm is also applied for the assimilation experiments with the combined ensemble. These experiments are carried out with different configurations of N and M (i.e. $N = 20$ & $M = 20$, $N = 20$ & $M = 40$, $N = 40$ & $M = 20$ and $N = 40$ & $M = 40$). At the same time, the EnSRF runs with 20, 40 and 100 members, and a free run without data assimilation is also performed. The free run is an ensemble forecast of 100 members with model error perturbations described by Zheng et al. (2009). For the EnSRF runs with 20, 40 and 100 members, there is no localization applied. All of these experiments are verified here by only using their ensemble means. The assimilation updates the model forecast once per month. Figure 2b displays the RMS errors compared with the independent TAO SST data (their locations are shown by the red square points in Fig. 2a) of the ensemble-mean analysis results

Table 3. CPU time consumption and memory requirements of the assimilation experiments with local observations for the ICM model

		Dynamical ensemble			Combined ensemble			
					$N = 20$	$N = 20$	$N = 40$	$N = 40$
		$N = 20$	$N = 40$	$N = 100$	$M = 20$	$M = 40$	$M = 20$	$M = 40$
CPU time consumption (min)	12-Month forecast process	128.0	254.0	643.0	128.0	128.0	254.0	254.0
	12 Times analysis process	26.0	43.0	112.0	37.0	54.0	52.0	71.0
	Total	154.0	297.0	755.0	165.0	182.0	308.0	325.0
Memory requirement (Mbytes)	12-Month forecast process	265.0	349.0	627.0	265.0	265.0	349.0	349.0
	12 Times analysis process	17.0	33.0	81.0	23.0	37.0	42.0	64.0

of these experiments. A further comparison is also performed by comparing all assimilation results with the OI.v2 SST anomaly data in all of the model grids, and this is presented in Fig. 2c. To provide a thorough benchmarking of the performance for the implementation of the new localization scheme, a summary and comparison of the CPU time consumption and the memory requirements of the different assimilation experiments is given in Table 3.

In general, the RMS errors of the assimilation results with different M and N are all smaller than those of the free run, indicating that the assimilation experiments are all effective. The small-size EnSRF runs combined with the local correlation function ensemble are better than those of the small-size EnSRF runs not combined with the local correlation function ensemble (e.g. the assimilation run with $N = 20$ & $M = 20$ vs. the assimilation run with $N = 20$), and both of the two experiments have similar CPU times and memory costs. For small-size N (e.g. $N =$ 20 and 40), as M increases, the performance is more improved. This means that the combined ensemble is able to obtain more reasonable assimilation results than the same dynamical ensemble size when N is small, while their CPU time consumptions and memory requirements increase only a little. The RMS errors of the combined ensemble with 40 dynamical members and 40 local correlation function members (i.e. $N = 40$ & $M = 40$) are very close to those of the 100-member dynamical ensemble and are even smaller than those of the 100-member dynamical ensemble at some assimilation steps, while the experiment of the combined ensemble (i.e. $N = 40\&M = 40$) reduces almost half of the CPU time consumption and memory cost compared with the 100-member dynamical ensemble during the 12-month ensemble forecast process (shown in Table 3). These all indicate that the new localization scheme with a relatively small dynamical size is capable of providing high assimilation quality through assimilating local observations, without increasing its price too much.

Next, we look at the impacts of the combined ensemble on the horizontal correlations of the point in the middle of the equatorial Pacific with other points at the first analysis step. Here, we further perform some additional experiments of an EnSRF with $N = 200$ ensemble members and an EnSRF with $N = 20$ & $M = 100$ at only the first analysis step. As displayed in Fig. 3, the left column shows the horizontal correlations estimated from the random ensemble sampled from a local correlation function with 20, 40, 100, 400 and 1000 members, respectively. The local correlation function has a Gaussian form with length scales of $L_x = 2000$ km and $L_y = 1000$ km. The shapes of the correlations are converging to a normal ellipse as the size of the ensemble increases. The middle column of Fig. 3 shows the correlations estimated from different sizes of dynamical ensemble of the above experiments. The long-distance correlations are obvious for $N =$ 20 and 40, and the horizontal distributions of the correlations are reasonable until the dynamical ensemble size exceeds 100. For the combined ensemble shown in the right column of Fig. 3, the long-distance correlations are effectively truncated even with 20-member dynamical ensemble when combining with a set of small-size (e.g. $M = 20$) local correlation function ensemble. When $N = 40$ & $M = 40$, the horizontal correlation distribution is similar to that of the 100-member dynamical ensemble and is comparable to that of the 400-member local correlation ensemble or 200-member dynamical ensemble.

3.5. Assimilation experiments of the ICM model with non-local 'observations'

As described in the Introduction, the assimilation of non-local observations (e.g. satellite radiance data) cannot be readily implemented in the presence of localization, while our new localization scheme can overcome this disadvantage. To verify the ability of the new localization scheme to directly assimilate non-local observations, similar to the design of the local observation assimilation experiments, some 1-year non-local 'observation' assimilation experiments are performed with generated pseudo non-local 'observations' by averaging the OIv2 SST observations over some meridional or zonal model lines of model grids, as shown in Fig. 4a (11 meridional lines and 9 zonal lines). Except for non-local 'observations', the other configurations of the experiments are the same as the previous local observation experiments for the ICM model.

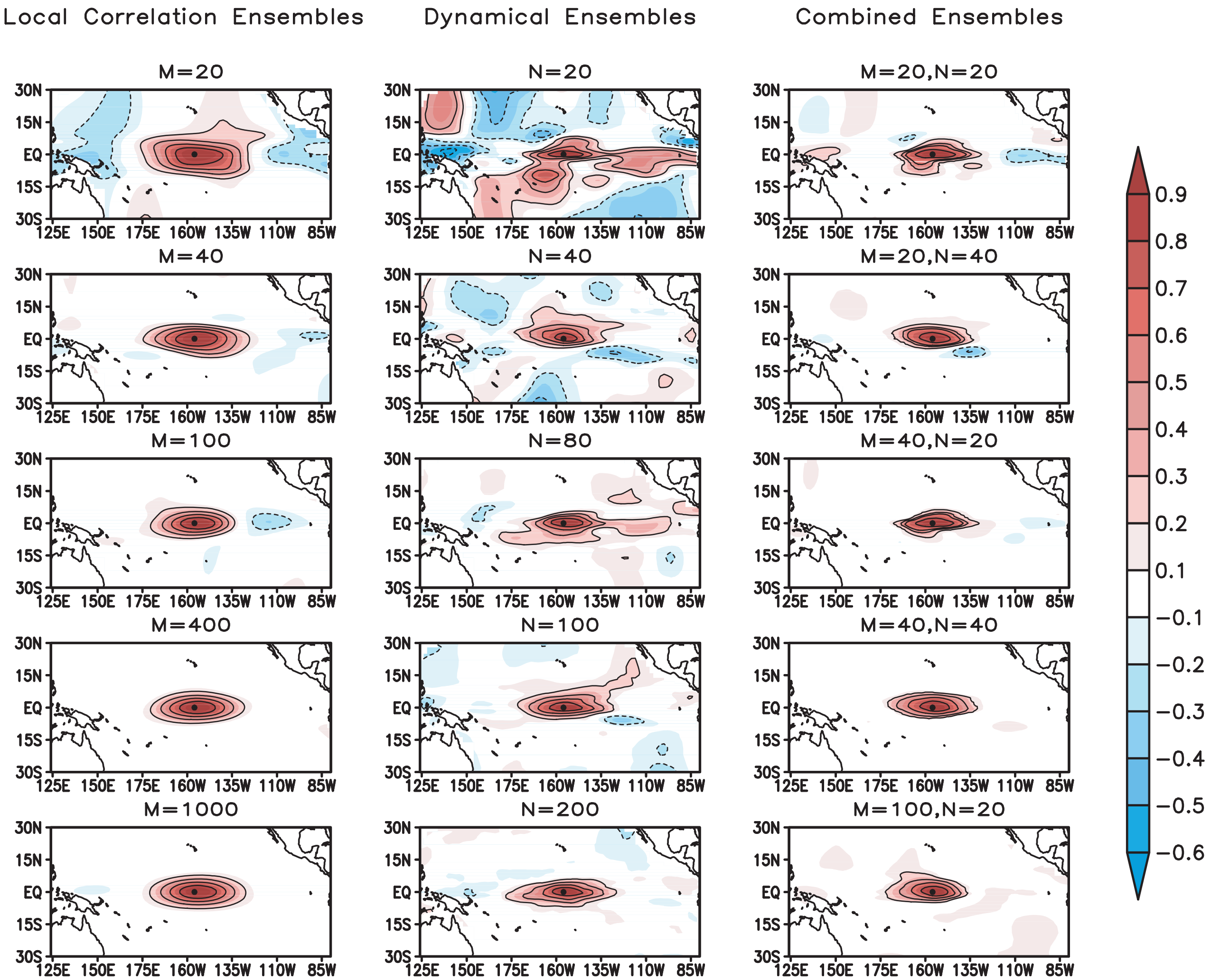

Fig. 3. Correlations of the SST anomaly at every model grid points to the SST anomaly at the middle point of the tropic Pacific (denoted by the black dot) estimated from different experiments. Left column: Correlations estimated from the random local correlation function ensemble with 20, 40, 100, 400 and 1000 members, respectively. Middle column: Estimated from the dynamical ensemble with 20, 40, 80,100 and 200 members, respectively. Right column: Estimated from the combined ensemble of the new localization scheme.

Figure 4b shows the RMS errors of different experiments that are validated by the OIv2 SST data over all model grid points. Similar to the comparison results for the local observation experiments, the assimilation qualities for different combined ensemble are all better than those of the experiments with 20- and 40-member dynamical ensemble. Also, the combined ensemble with 40-member dynamical ensemble and 40-member local correlation function ensemble has similar qualities of analysis as the 100-member dynamical ensemble.

To check how the analysis updates the model simulation from a single non-local 'observation' based on different localizations with different N and M, we pick up analysis increments made by assimilating only one non-local 'observation' from one analysis step. Two such examples are displayed in Figs 5 and 6, which show the location (i.e. green line in Figs 5 and 6) of the non-local 'observations' and the analysis increments from different dynamical ensemble and combined ensemble. For all experiments, the analysis increments are mainly along the non-local 'observation' line, but there are still some long-distance increments from the lines when using only small-size dynamical ensemble. The combined ensemble can obtain similar analysis results along the non-local 'observation' lines compared with the large-size dynamical ensemble ($N = 100$), and they are also able to eliminate the long-distance false increments far away from the non-local 'observation' lines, especially for the combined ensemble with 40-member dynamical ensemble and 40-member local correlation function ensemble. The new location scheme thus limits the impact of the non-local 'observations' within the length scale reach of the 'observation' lines, while it enables information propagation along the 'observation' lines.

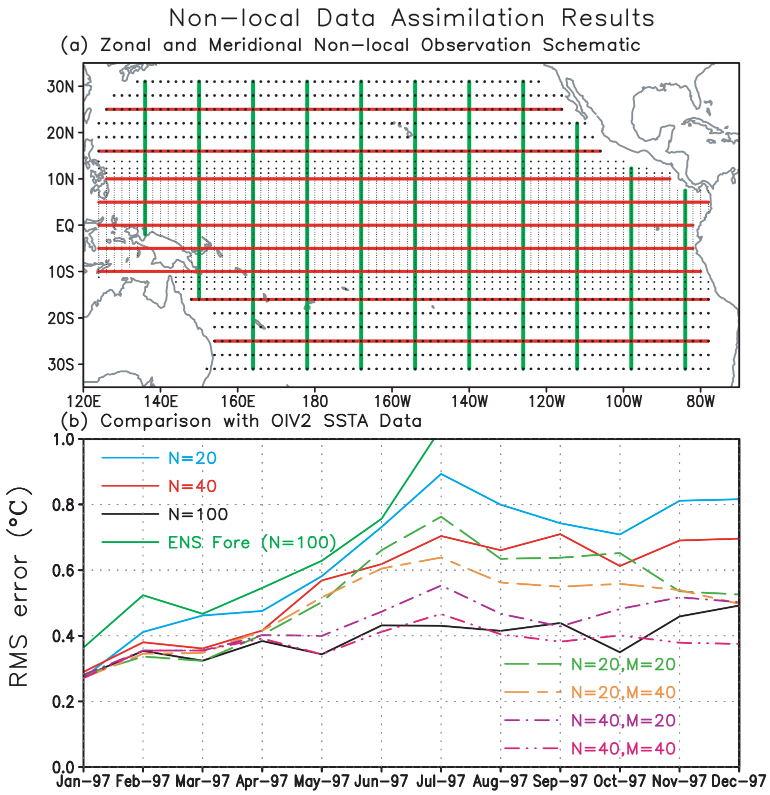

Fig. 4. (a) The lines of the artificial non-local 'observation'. Eleven meridional lines are shown in green, while nine zonal lines are in red. The non-local 'observations' are made by averaging the OI.v2 SST data over these lines. The black dots present the locations of the model grid points. (b) The same as Fig. 2c, but for non-local 'observation' experiments.

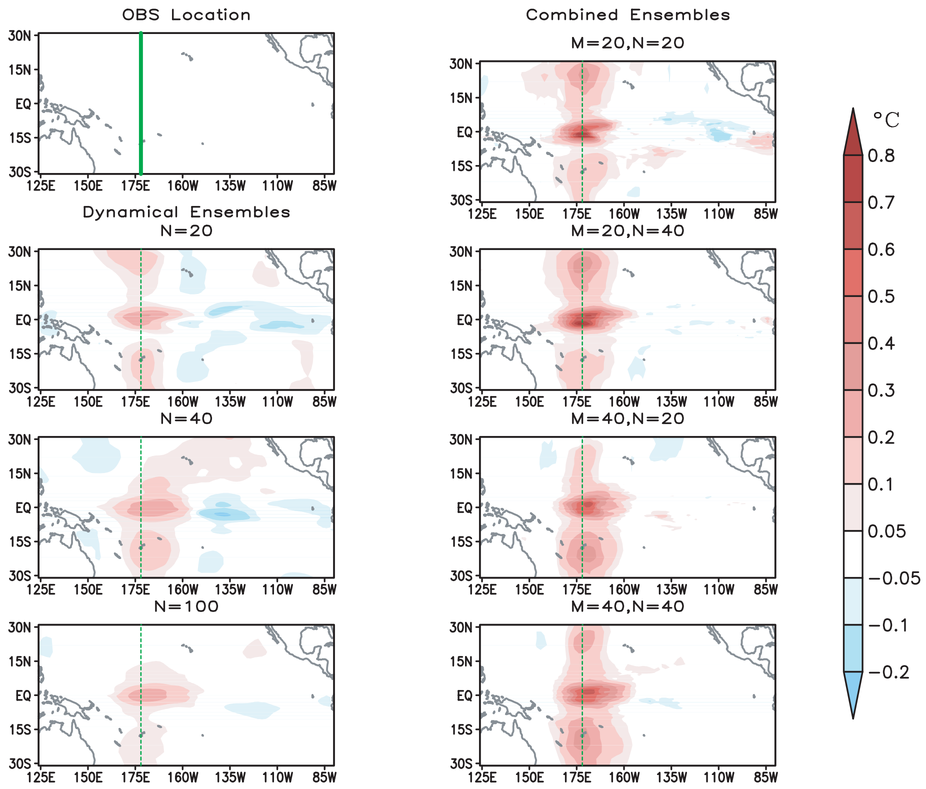

Fig. 5. The analysis increments of a single non-local 'observation' along the green line in the data assimilation experiments.

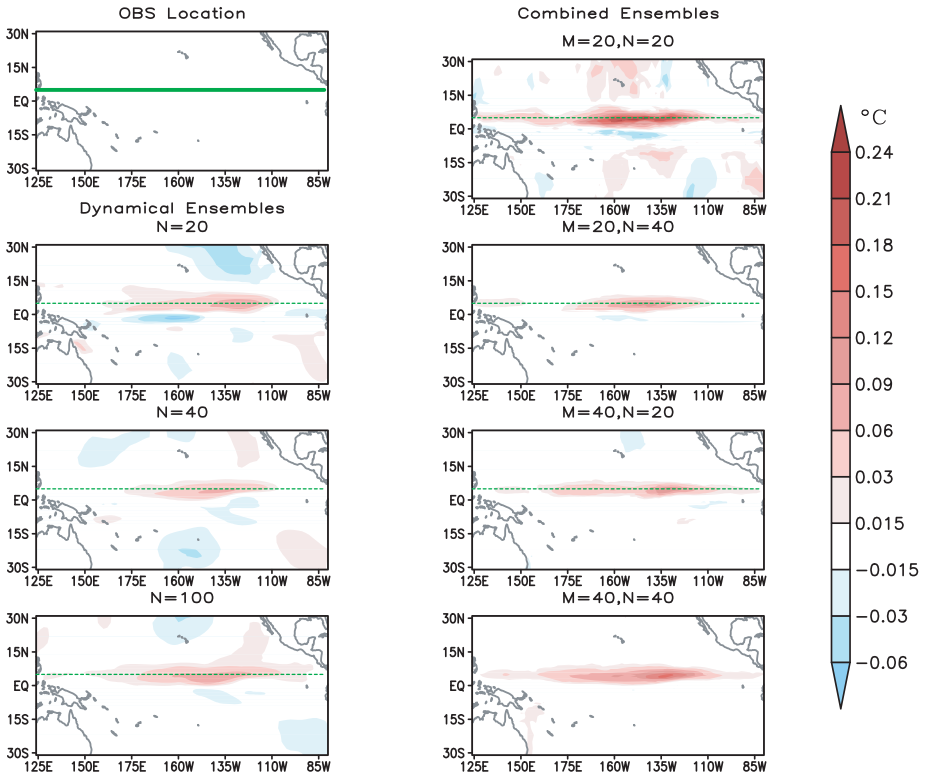

Fig. 6. The same as in Fig. 5, but for another non-local 'observation' line.

The results indicate that the new localization scheme is able to directly assimilate the non-local observations without any need to artificially pinpoint the non-local observations, which would reduce the impact of the non-local observations.

4. Conclusions

A new localization implementation scheme based on covariance localization for ensemble data assimilation of non-local observations by combining ensemble sampling the local correlation function matrix and the small-size dynamical ensemble is tested with focusing on assimilating non-local observations. Although the size of the combined ensemble is large (i.e. 1000–2000), the memory requirement can be kept within a practical level using a memory-saving algorithm at the update stage without storing the whole ensemble. The performance of the new scheme is first verified by an LA model through the assimilation of local observations, and then the scheme is further verified in both local and non-local SST anomaly data assimilation experiments by using an ICM in the tropical Pacific. The analysis results show that the new localization scheme is feasible and effective. The assimilation experiments with the combined ensemble can successfully assimilate both the local and non-local observations, and they can also provide reasonable and high-quality analysis fields with a small dynamical ensemble size.

The results described in this paper represent our preliminary attempts to develop and improve the implementation of localization within the ensemble data assimilation process. Based on our experience in this work, a combined ensemble with 40 dynamical ensemble members and 40 local correlation ensemble members might be the relatively balanced combination between the computational requirement and the performance. However, as demonstrated in many previous studies (e.g. Houtekamer and Mitchell, 2005; Whitaker et al., 2008), localization may disturb the physical balances between various physical variables. In the localization implementation scheme here, the imbalance problem associated with localization is still there. Further applications of the assimilation of the radiance observations into a global weather forecast model are currently under way.

5. Acknowledgments

The authors wish to thank two anonymous reviewers for their comments. This work was supported by the Knowledge Innovation Program of the Chinese Academy of Sciences (Grant No. KZCX1-YW-12-03), National Basic Research Program of China (Grant No. 2010CB951901) and the Natural Science Foundation of China (Grant No. 40805033). The authors would like to thank Geir Evensen for providing the EnKF analysis codes at http://enkf.nersc.no/.

Appendix A

Generally, the gain matrices given by eqs. (3) and (4) do not equal each other. Inserting $\rho'_{\mathrm{s}} = \rho_{\mathrm{s}}\mathbf{H}^{\mathrm{T}}$ and$\rho''_{\mathrm{s}} = \mathbf{H}\rho_{\mathrm{s}}\mathbf{H}^{\mathrm{T}}$ into the right-hand side of eq. (4), the equation becomes

$$\mathbf{K} = (\rho_{\mathrm{s}}\mathbf{H}^{\mathrm{T}}) \circ (\mathbf{P}^b\mathbf{H}^{\mathrm{T}})[(\mathbf{H}\rho_{\mathrm{s}}\mathbf{H}^{\mathrm{T}}) \circ (\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}}) + \mathbf{R}]^{-1}. \qquad \text{(A1)}$$

The following gives an example with a ‘non-local’ observation operator that the right-hand side of eq. (A1) does not equal that of eq. (3) when $\mathbf{H}$ is dimensionless.

For a three-dimension state vector $(\, x_1 \;\; x_2 \;\; x_3 \,)^{\mathrm{T}}$, if $\mathbf{H}$ is defined by

$$\mathbf{H}(\, x_1 \;\; x_2 \;\; x_3 \,)^{\mathrm{T}} = \frac{1}{3}(x_1 + x_2 + x_3), \quad \mathbf{H} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}, \qquad \text{(A2)}$$

we also define

$$\rho_{\mathrm{s}} = \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix}, \quad \mathbf{P}^b = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{12} & p_{22} & p_{23} \\ p_{13} & p_{23} & p_{33} \end{pmatrix}. \qquad \text{(A3)}$$

Then

$$\begin{aligned} &\left(\rho_{\mathrm{s}}\mathbf{H}^{\mathrm{T}}\right) \circ \left(\mathbf{P}^b\mathbf{H}^{\mathrm{T}}\right) \\ &= \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} \circ \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{12} & p_{22} & p_{23} \\ p_{13} & p_{23} & p_{33} \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} \\ &= \begin{pmatrix} 1/6(p_{11} + p_{12} + p_{13}) \\ 1/3(p_{12} + p_{22} + p_{23}) \\ 1/6(p_{13} + p_{23} + p_{33}) \end{pmatrix}, \end{aligned} \qquad \text{(A4)}$$

$$\begin{aligned} \left(\rho_{\mathrm{s}} \circ \mathbf{P}^b\right)\mathbf{H}^{\mathrm{T}} &= \begin{pmatrix} p_{11} & 0.5p_{12} & 0 \\ 0.5p_{12} & p_{22} & 0.5p_{23} \\ 0 & 0.5p_{23} & p_{33} \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} \\ &= \begin{pmatrix} 1/3p_{11} + 1/6p_{12} \\ 1/6p_{12} + 1/3p_{22} + 1/6p_{23} \\ 1/6p_{23} + 1/3p_{33} \end{pmatrix}. \end{aligned} \qquad \text{(A5)}$$

So,

$$\left(\rho_{\mathrm{s}}\mathbf{H}^{\mathrm{T}}\right) \circ \left(\mathbf{P}^b\mathbf{H}^{\mathrm{T}}\right) \neq \left(\rho_{\mathrm{s}} \circ \mathbf{P}^b\right)\mathbf{H}^{\mathrm{T}}. \qquad \text{(A6)}$$

Appendix B

In practical implementation, based on the traditional EnKF scheme (e.g. Evensen, 2003), an example algorithm is provided here to show how to save memory at the analysis step. In detail, the matrices $\mathbf{HC}$ are calculated with single ensemble members iteratively and without storing the whole matrix $\mathbf{C}$. Using $\mathbf{HC}$, the matrix $(\mathbf{HC}^f)^{\mathrm{T}}[\mathbf{HC}^f(\mathbf{HC}^f)^{\mathrm{T}} + (N-1)(M-1)\mathbf{R}]^{-1}(\mathbf{Y} - \mathbf{HX}^f)$ in Eq. (10) is an $(M \times N)$-dimensional *column vector*, and it can be written as

$$\begin{pmatrix} l_{11} \\ l_{21} \\ \bullet \\ \bullet \\ \bullet \\ l_{ij} \\ \bullet \\ \bullet \\ \bullet \\ l_{MN} \end{pmatrix}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, M. \qquad \text{(B1)}$$

Then following the analysis incremental equation in eq. (10):

$$\begin{aligned} &\mathbf{C}^f(\mathbf{HC}^f)^{\mathrm{T}}[\mathbf{HC}^f(\mathbf{HC}^f)^{\mathrm{T}} + (N-1)(M-1)\mathbf{R}]^{-1}(\mathbf{Y} - \mathbf{HX}^f) \\ &\quad = \sum_{i,j} (\mathbf{s}^{(j)} \circ \mathbf{x}'^{(i)}) l_{ij}, \end{aligned} \qquad \text{(B2)}$$

enables an iterative summing operation to calculate the analysis incremental. So the memory requirement is reduced to store $\mathbf{HC}$ ($m \times M \times N$-dimensional) plus a single ensemble member (n-dimensional). The update analysis equation in EnSRF [i.e. eqs. (11)–(12)] can also use this algorithm to reduce the memory requirement.

Table 4 clearly shows the significant improvements for both of the CPU time consumptions and the memory requirements,

Table 4. Comparisons of CPU time consumptions and memory requirements for using the memory-saving algorithm and without using the memory-saving algorithm in the assimilation experiments with local observations for the ICM model

		Using memory-saving algorithm		Without using memory-saving algorithm	
		$N = 20$ M = 20	$N = 40$ M = 40	$N = 20$ M = 20	$N = 40$ M = 40
CPU-time consumption (min)	12-Month forecast process	128.0	254.0	128.0	254.0
	12 Times analysis process	37.0	71.0	57.0	134.0
	Total	165.0	325.0	205.0	388.0
Memory requirement (Mbytes)	12-Month forecast process	265.0	349.0	265.0	349.0
	12 Times analysis process	23.0	64.0	283.0	1125.0

especially for the memory requirements, when adopting the memory-saving algorithm in the assimilation experiments with local observations for the ICM (i.e. Section 3.4).

References

Anderson, J. L. 2001. An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–2903.

Anderson, J. L. 2007. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D* **230**, 99–111.

Bishop, C. H. and Hodyss, D. 2007. Flow adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quart. J. R. Meteor. Soc.* **133**, 2029–2044.

Bishop, C. H. and Hodyss, D. 2009a. Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus* **61A**, 84–96.

Bishop, C. H. and Hodyss, D. 2009b. Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus* **61A**, 97–111.

Buehner M. and Charron, M. 2007. Spectral and spatial localization of background-error correlations for data assimilation. *Quart. J. R. Meteor. Soc.* **133**, 615–630.

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

Evensen, G. 2004. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn.* **54**, 539–560.

Evensen, G. 2007. *Data Assimilation—The Ensemble Kalman Filter*. Springer, Berlin, 279 pp.

Fertig, E. J., Hunt, B. R., Ott, E. and Szunyogh, I. 2007. Assimilating non-local observations with a local ensemble Kalman filter. *Tellus* **59A**, 719–730.

Gaspari, G. and Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. *Quart. J. R. Meteor. Soc.* **125**, 723–757.

Hamill, T. M., Whitaker, J. S. and Snyder, C. 2001. Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**, 2776–2790.

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811.

Houtekamer, P. L. and Mitchell, H. L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **129**, 123–137.

Houtekamer, P. L. and Mitchell, H. L. 2005. Ensemble Kalman filtering. *Quart. J. R. Meteor. Soc.* **131**, 3269–3289.

Keenlyside, N. and Kleeman, R. 2002. Annual cycle of equatorial zonal currents in the Pacific. *J. Geophys. Res.* **107**, doi: 10.1029/2000JC0007111.

Keppenne, C. L. 2000. Data assimilation into a primitive equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.* **128**, 1971–1981.

Liu, C., Xiao, Q. and Wang, B. 2009. An ensemble-based four-dimensional variational data assimilation scheme. part II: Observing system simulation experiments with advanced research WRF (ARW). *Mon. Wea. Rev.* **137**, 1687–1704.

McCreary, J. P. 1981. A linear stratified ocean model of the equatorial undercurrent. *Philos. Trans. R. Soc. (Lond.)* **298**, 603–635.

Oke, P. R., Sakov, P. and Corney, S. P. 2007. Impacts of localisation in the EnKF and EnOI: experiments with a small model. *Ocean Dyn.* **57**, 32–45.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A.V., Kostelich, E. J. and co-authors. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**, 415–428.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. and Wang, W. 2002. An improved in situ and satellite SST analysis for climate. *J. Climate* **15**, 1609–1625.

Sakov, P. and Oke, P. R. 2008a. Implications of the form of the ensemble transformations in the ensemble square root filters. *Mon. Wea. Rev.* **136**, 1042–1053.

Sakov, P. and Oke, P. R. 2008b. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus* **60A**, 361–371.

Whitaker, J. S., Compo, G. P., Wei, X. and Hamill, T. M. 2004. Reanalysis without Radiosondes Using Ensemble Data Assimilation. *Mon. Wea. Rev.* **132**, 1190–1200.

Whitaker, J. S. and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913–1924.

Whitaker, J. S., Hamill, T. M. and Wei, X. 2008. Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.* **136**, 463-482.

Zhang, R.-H., Zebiak, S. E., Kleeman, R. and Keenlyside, N. 2005. Retrospective El Niño forecast using an improved intermediate coupled model. *Mon. Wea. Rev.* **133**, 2777–2802.

Zheng, F. and Zhu, J. 2008. Balanced multivariate model error in the ensemble kalman filter data assimilation for an intermediate coupled model. *J. Geophys. Res.* **C113**, C07002, doi: 10.1029/2007JC004621.

Zheng, F., Zhu, J., Wang, H. and Zhang, R.-H. 2009. Ensemble hindcasts of ENSO events over the past 120 years using a large number of ensembles. *Adv. Atmos. Sci*. **26**, 359–372.

Zheng, F., Zhu, J., Zhang, R.-H. and Zhou, G. Q. 2006. Ensemble hindcasts of SST anomalies in the tropical Pacific using an intermediate coupled model. *Geophys. Res. Lett.* **33**, L19604, doi: 10.1029/2006GL026994.

Zheng, F., Zhu, J. and Zhang, R.-H. 2007. The impact of altimetry data on ENSO ensemble initializations and predictions. *Geophys. Res. Lett.* **34**, L13611, doi: 10.1029/2007GL030451.