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

Tellus (2009), 61A, 84–96
Printed in Singapore. All rights reserved

TELLUS

Ensemble covariances adaptively localized with ECO-RAP. Part 1: tests on simple error models

By CRAIG H. BISHOP* and DANIEL HODYSS, *Naval Research Laboratory, Marine Meteorology Division, 7 Grace Hopper Ave, Stop 2, Building 702, Room 212, Monterey, CA 93943-5502, USA*

(Manuscript received 17 March 2008; in final form 30 September 2008)

ABSTRACT

In atmospheric data assimilation (DA), observations over a 6–12-h time window are used to estimate the state. Non-adaptive moderation or localization functions are widely used in ensemble DA to reduce the amplitude of spurious ensemble correlations. These functions are inappropriate (1) if true error correlation functions move a comparable distance to the localization length scale over the time window and/or (2) if the widths of true error correlation functions are highly flow dependent. A method for generating localization functions that move with the true error correlation functions and that also adapt to the width of the true error correlation function is given. The method uses ensemble correlations raised to a power (ECO-RAP). A gallery of periodic one-dimensional error models is used to show how the method uses error propagation information and error correlation width information retained by powers of raw ensemble correlations to propagate and adaptively adjust the width of the localization function. It is found that ECO-RAP localization outperforms non-adaptive localization when the true errors are propagating or the error correlation length scale is varying and is as good as non-adaptive localization when such variations in error covariance structure are absent.

1. Introduction

Accurate estimates of the state of the ocean and atmosphere are a key part of environmental forecasting and climate monitoring. To make such estimates, one typically combines information from a short-range forecast of the state with information from observations. For some purposes, it is of interest to use a forecast and observations over a time window to either estimate the four-dimensional trajectory of the atmospheric state over the window or to estimate a single state within the time window. In either case, minimum error variance and maximal likelihood state estimation schemes require the specification of the forecast and observation error covariance matrices (Daley, 1991). The forecast error covariance matrices define the covariance of forecast errors of variables that may have differing space–time locations and differing variable types.

Ensemble forecasts are collections of forecasts whose initial condition differences sample analysis error and, possibly, whose model formulations sample model uncertainty. The sample covariance matrix of an ensemble forecast approximates the forecast error covariance matrix. The literature on ensemble based atmospheric and oceanic data assimilation (DA) is vast and the reader is referred to Evensen (2003) for a review of many of the pre-2003 papers describing how ensemble covariances can be used in atmospheric and oceanic DA. Because of the computational expense of generating ensemble members, the number of ensemble members K used for atmospheric DA has generally been less than 100. Consequently, it is often found that ensemble sample covariance matrices spuriously assign non-zero correlations to variables that are uncorrelated. Furthermore, the rank of the ensemble sample covariance matrix is likely to be much smaller than the rank of the true forecast error covariance matrix. To suppress spurious ensemble correlations and increase the rank of the ensemble based forecast error covariance model, most ensemble DA schemes incorporate a localization procedure that zeros out many of these spurious correlations (Evensen, 2003). Expositions on the need for localization in ensemble based DA can be found in Houtekamer and Mitchell (1998), Hamill et al. (2001) and Lorenc (2003). Generally, ensemble DA schemes that serially process observations such as those of Houtekamer et al. (2005), Khare et al. (2008) and Whitaker et al. (2008), have used broader localization functions (correlations set to zero around 2000–3000 km) than the observation volume based schemes used in Miyoshi and Yamane (2007) and Szunyogh et al. (2008) (correlations set to zero around 800 km) that allow parallel processing of observations.

In the studies of Houtekamer et al. (2005), Miyoshi and Yamane (2007), Khare et al. (2008) and Whitaker et al. (2008)

*Corresponding author.
e-mail: bishop@nrlmry.navy.mil
DOI: 10.1111/j.1600-0870.2008.00371.x

observations distributed over a 6-h time window were used to estimate the state at the centre of the time window.[1] In all of these studies, the localization functions employed were immobile and their widths were flow independent. Two problems associated with such non-adaptive localization functions are that (1) the spatial 'area' containing variables whose errors are strongly correlated with the error variables at the initial time may be 'moving'[2] and (2) the horizontal and/or vertical width of the optimal localization function may change depending on the flow. While the localization width adaptivity issue is relevant to all atmospheric ensemble Kalman filters (EnKFs), the error movement issue is crucially important for problems involving correlations between variables widely separated in time such as the adaptive sampling problem (Bishop et al., 2001; Khare and Anderson, 2006) or DA over time windows in which error movement is comparable to the half-width of the localization function. This paper addresses these two problems by giving a method (ensemble correlations raised to a power, ECO-RAP) for constructing flow adaptive localization functions/matrices.

Previous work on propagating and/or adaptive localization functions includes Anderson (2007) who showed how flow-dependent localization functions for ensembles of size K/N can be derived by (1) splitting an ensemble of size K into N subensembles of size K/N and (2) basing the moderation function on the degree of disagreement between correlation estimates from the independent subensembles. In Anderson's hierarchical approach, the moderation functions are intended to optimize the performance of K/N member subensembles rather than K member ensembles. Thus, in the context of three-dimensional DA with ensembles of size K and no variation in the true error correlation functions, the upper bound on the performance of the hierarchical approach is that of an optimally tuned non-adaptive localization of a K/N-member ensemble. Here we demonstrate that, in such circumstances, ECO-RAP can deliver performance, that is, statistically indistinguishable from that of an optimally tuned non-adaptive localization of a K-member ensemble.

Closely related to the current work is the localization method smoothed ensemble correlations raised to a power (SENCORP) described in Bishop and Hodyss (2007). Simple calculations show that SENCORP is more than a thousand times slower than ECO-RAP for the observation volumes considered in Bishop and Hodyss (2009; hereafter Part 2, this issue). Non-adaptive localization is a special case of ECO-RAP (see Section 2.5) but no such relationship is evident for SENCORP. In both schemes, powers of sample correlations play a role in attenuating spurious ensemble covariances. However, the role of the smoothing operator in ECO-RAP is very different to its role in SENCORP. In ECO-RAP, the rows of the matrix containing some element-wise power of the ensemble correlation matrix are smoothed whereas, in SENCORP, it is the ensemble perturbations themselves that are smoothed. Given these differences, the localization characteristics of ECO-RAP cannot be predicted from the performance characteristics of SENCORP. Hence, to achieve the primary aim of Part 1 (this paper) – an understanding of ECO-RAP's localization characteristics – it is necessary to test it on a wide range of error models.

Simple linear error models in which the true forecast error covariance is known and in which the movement, dispersion, growth and dimension of error can easily be controlled are used in Part 1. For the purposes of Part 1, these systems are preferable to idealized non-linear dynamical models such as the elegant system described in Lorenz (1996) because (1) the true forecast error covariance matrices associated with non-linear systems can only be obtained from very expensive non-linear DA schemes and (2) the authors know of no simple method to precisely control the movement, dispersion, growth and dimension of errors associated with such non-linear systems. Nevertheless, atmospheric DA schemes must be tested on non-linear systems so in Part 2, we consider a non-linear system provided by a numerical weather prediction model.

In Part 1, we explore the performance characteristics of ECO-RAP localization using a wide variety of one-dimensional periodic error models. We show that the row smoothing of ECO-RAP is associated with a spectral transform that allows the user to control the number of columns required to represent the square root of the localization matrix. In Part 2, we show that this feature of the ECO-RAP square root enables a highly efficient, large observation volume local ensemble transform Kalman filter (LETKF) similar to those described in Miyoshi and Yamane (2007) and Szunyogh et al. (2008) and related to that of Bowler (2006).

Section 2 describes and illustrates ECO-RAP and ensemble covariances adaptively localized with ECO-RAP (CALECO) using a variety of error models. In Section 3, we compare the DA performance of the CALECO forecast error covariance model against that based on non-adaptive localization for propagating error and for error whose correlation length scale varies. Conclusions follow in Section 4.

2. ECO-RAP

2.1. Error models

Many fundamental DA studies have used dynamical models to generate distributions of forecast errors. We chose not to use this approach because it would limit us to the error distributions associated with the one or two dynamical models we were able to implement. By choosing to directly describe time-dependent forecast error distributions using the methods described below and in the appendices, we could test just about any error distribution we could imagine. In what follows we

[1]Hunt et al. (2004) discussed four-dimensional Kalman filtering in an idealized model.

[2]Errors moving at a jet level wind speed of 30 ms^{-1} would travel about 650 km in 6 h.

have created a gallery of physically relevant error distributions that any meteorologically sound DA system should be able to assimilate.

All of the error distributions considered pertained to a system consisting of M variables evenly spaced along a circle with circumference 2π. The propagating error systems considered feature initial time and later time errors related by some assigned set of rules. If, say, n_t time levels were assumed, then n the state dimension is given by $n = n_t M$. The unifying concept for the error distributions considered is that if the true $n \times n$ forecast error covariance matrix $\mathbf{P}^{\mathrm{f}}$ can be written in the form

$$\mathbf{P}^{\mathrm{f}} = \mathbf{V}\beta\mathbf{V}^{T} \tag{1}$$

then this matrix is precisely equal to the covariance of random forecast errors ε^{f} modelled as

$$\varepsilon^{\mathrm{f}} = \mathbf{V}\mathbf{a}, \quad \text{where } \mathbf{a} \sim N(\mathbf{0}, \beta). \tag{2}$$

where $\mathbf{V}$ is a $n \times L$ matrix whose L columns list the basis functions and $\mathbf{a}$ is a random L-vector with mean zero and covariance given by the diagonal matrix β (The reader may prove this connection between (2) and (1) by right multiplying (2) by its transpose and taking the expectation of the resulting equation). Hence, once the $\mathbf{V}$ and the β defining $\mathbf{P}^{\mathrm{f}}$ have been obtained, it is trivial to generate random forecast errors using eq. (2). Throughout this paper, it is assumed that ensemble perturbations are drawn from the same distribution as forecast error. Under this assumption, eq. (2) is not only used to generate random forecast or first guess errors, it is also used to generate K member ensembles. Each ensemble member is obtained from an independent application of (2).

The basis functions chosen for the columns of $\mathbf{V}$ corresponded to either (1) sine and cosine functions possibly concatenated with propagated sine and cosine functions (see Sections 1–3 of the appendix) or (2) a non-orthogonal, non-centred wavelet basis (see Section 4 of the appendix).[3] The non-orthogonal wavelet basis enabled us to produce non-local error correlation functions in which errors were strongly correlated with nearby errors and with errors a half circumference away while being uncorrelated with errors at intermediate distances. Note that for the sine and cosine function basis, broader error correlation functions can be obtained by assigning larger values to the elements of β corresponding to larger scale waves. However, for the non-orthogonal wavelet basis, β was chosen to be proportional to the identity matrix because error correlation width information was implicit in the width of the functions used as wavelets.

[3]Some authors apparently believe that the word 'wavelet' should only be applied to a set of functions satisfying a list of properties including orthogonality and/or having a mean of zero. Our apologies to these authors. By 'wavelet' we just mean 'little wave'.

2.2. ECO-RAP and the localization of an Eady wave packet

Figure 1a gives the correlation function for the first error distribution to be considered. It features a wave packet in the centre of the domain at the initial time and an expanded wave packet centred on the right-hand side of the domain at the final time. As discussed in subsection A.3 of Appendix A, the dispersion relation that produced Fig. 1a is that of Eady (1949) for idealized baroclinic waves. Note that the wave packet at the final time is broader than that at the initial time. If typical mid-latitude parameters are used in the Eady model, the associated time difference between the initial time wave packet and the final time wave packet would be approximately 6 d. Hence, the wave packet expansion depicted in Fig. 1 is crudely similar to that associated with the adaptive sampling problems discussed in Majumdar et al. (2001). It is also similar to the wave packet expansion that would be associated with four-dimensional state estimation over a 6-d time window.

Even if it were possible to design an ensemble forecasting system that truly sampled the distribution of possible truths given the forecast, finite ensemble size would render error covariance and error correlation estimates inaccurate. Figure 1b illustrates the sample correlation function for a 64-member ensemble of random draws from the true error distribution. It illustrates how small ensemble size leads to spurious ensemble correlations. The expected value of the square of the sample correlation between two uncorrelated variables x and y for large ensemble size K is

$$\begin{aligned}
\left\langle c_{xy}^{2}\right\rangle &= \left\langle \frac{\left(\frac{1}{K-1}\sum_{i=1}^{K} x_i' y_i'\right)^2}{\frac{1}{K-1}\sum_{i=1}^{K} x_i'^2 \frac{1}{K-1}\sum_{i=1}^{K} y_i'^2} \right\rangle \\
&\approx \frac{\left(\frac{1}{K-1}\right)^2 \sum_{i=1}^{K} \left\langle x_i'^2\right\rangle \left\langle y_i'^2\right\rangle}{\left\langle \frac{1}{K-1}\sum_{i=1}^{K} x_i'^2 \right\rangle \left\langle \frac{1}{K-1}\sum_{i=1}^{K} y_i'^2 \right\rangle} \text{ for large } K \\
&= \frac{\left(\frac{1}{K-1}\right)^2 \sum_{i=1}^{K} \left\langle x_i'^2\right\rangle \left\langle y_i'^2\right\rangle}{\left\langle x_i'^2\right\rangle \left\langle y_i'^2\right\rangle} = \frac{K}{(K-1)^2} \approx \frac{1}{K} \text{ for large } K,
\end{aligned} \tag{3}$$

where x_i' and y_i' denote perturbations about the ensemble mean.

Houtekamer and Mitchell (1998) and Lorenc (2003), amongst others, have shown that these spurious correlations can severely compromise the performance of a DA scheme that uses them. The prime objective of error covariance localization in ensemble based DA is to remove these spurious smaller correlations while keeping the larger correlations that correspond to non-zero true correlations.

In Bishop and Hodyss' (2007) SENCORP adaptive localization technique, one began by smoothing the ensemble perturbations and then raising the correlations of the smoothed ensemble correlations to a power. In ECO-RAP, the initial smoothing of perturbations is optional. To give our explanation of the role of each of the steps involved in constructing the ECO-RAP

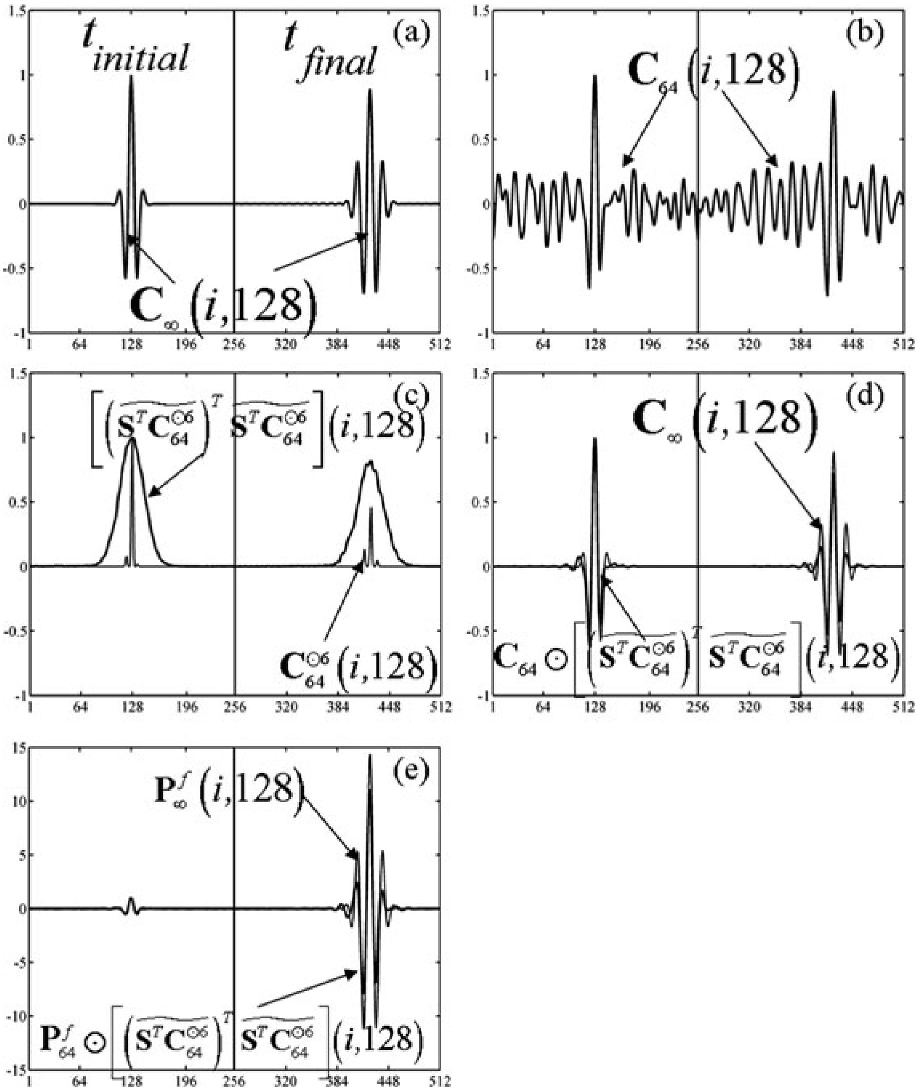

Fig. 1. In (a) the true error correlations with the 128th initial time variable of the 256 initial time variables (left-hand side of domain) and 256 final time variables (right-hand side of domain) associated with a packet of amplifying Eady waves is given. In (b) the corresponding sample correlation function associated with a 64-member ensemble is given. The grey line in (c) shows the 6th power of the raw sample correlation function while the black line shows the ECO-RAP localization function. The black line in (d) gives the correlation function obtained by localizing the sample correlation function in (b) with the ECO-RAP localization function shown in (c). The corresponding Covariances Adaptively Localized with ECO-RAP (CALECO) covariance function is shown in (e) along side the true covariance.

matrix a context, we first summarize the general method by which ECO-RAP localization matrices are obtained:

(1) Compute the sample correlation matrix $\mathbf{C}_K$ of the (possibly smoothed) K member ensemble (Fig. 1b depicts a column of this matrix for an unsmoothed 64 member ensemble).

(2) Raise the elements of the sample correlation matrix to the qth power to diminish small correlations (likely to be noise) and emphasize large correlations (likely to be real). Denote this matrix $\mathbf{C}_K^{\odot q}$.

(3) The ECO-RAP correlation matrix is then given by $\widetilde{\mathbf{C}_K^{\odot q}\mathbf{S}}\,\widetilde{\mathbf{S}^T\mathbf{C}_K^{\odot q}}$ where $\mathbf{S}\mathbf{S}^T$ is a smoothing operator. Note that the matrix $\widetilde{\mathbf{C}_K^{\odot q}\mathbf{S}} = \mathbf{G}^{-1}(\mathbf{C}_K^{\odot q}\mathbf{S})$ where $\mathbf{G}^{-1}$ is a diagonal matrix that makes the length of each row of $\widetilde{\mathbf{C}_K^{\odot q}\mathbf{S}}$ equal to unity. Also note that $\widetilde{\mathbf{S}^T\mathbf{C}_K^{\odot q}} = (\widetilde{\mathbf{C}_K^{\odot q}\mathbf{S}})^T$. The smoothing operator makes the correlation length scales associated with $\widetilde{\mathbf{C}_K^{\odot q}\mathbf{S}}\,\widetilde{\mathbf{S}^T\mathbf{C}_K^{\odot q}}$ longer than those associated with $\mathbf{C}_K^{\odot q}$.

Figure 1 highlights the role of these steps in building the ECO-RAP matrix. The grey line in Fig. 1c depicts the role of step (2). It shows the one point correlation function of the correlation matrix $\mathbf{C}_{64}^{\odot 6}$ whose elements are equal to the 6th element-wise power of the sample correlation matrix from a 64-member ensemble.

Note that the elements of the columns of $\mathbf{C}_{64}^{\odot 6}$ plotted in Fig. 1c corresponding to true zero correlations are very close to zero. Note also that no negative correlations are present because an even power was taken. Finally, observe that the peaks of the function describing a column of $\mathbf{C}_{64}^{\odot 6}$ are not broad enough to embrace all of the crests and troughs in the wave packet, and that the peak at the later time only reaches to 0.5.

The third step is to form $\widetilde{\mathbf{C}_K^{\odot q}}\mathbf{S}\mathbf{S}^T\widetilde{\mathbf{C}_K^{\odot q}}$. In Part 2, we show that computational advantages can be realized if one lets $\mathbf{S}$ be a spectrum weighting matrix given by

$$\mathbf{S}_{n\times(Jn_t)} = \begin{bmatrix} \mathbf{E}_{M\times J}\Lambda_f^{1/2} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{M\times J}\Lambda_f^{1/2} & & \\ . & & . & \\ \mathbf{0} & & & \mathbf{E}_{M\times J}\Lambda_f^{1/2} \end{bmatrix}, \quad (4)$$

where $\Lambda_f^{1/2}$ is a $J \times J$ diagonal matrix giving weights to the J columns listed in $\mathbf{E}_{M\times J}$. The J orthonormal vectors contained in $\mathbf{E}_{M\times J}$ are a subset of the M orthonormal vectors $\mathbf{E}_{M\times M}$ that define an orthonormal basis for the M-dimensional gridpoint space. In (4), M gives the number of gridpoints on the circle, $n = n_t M$ gives the total number of state variables. The number of columns J required to define (4) needs only be as large as the number of non-zero elements in the diagonal matrix $\Lambda_f^{1/2}$. Hence, the stronger the smoother, the smaller the number of columns J that are required to define $\mathbf{S}$.

If the J columns of $\mathbf{E}_{M\times J}$ were a sinusoid basis then the operation $\Lambda_f^{1/2}(\mathbf{E}_{M\times J})^T\mathbf{x}_M$ on some M-vector $\mathbf{x}_M$ would be equivalent to taking the discrete Fourier transform of $\mathbf{x}_M$ and then weighting the resulting coefficient vector with the corresponding non-zero elements of the diagonal matrix $\Lambda_f^{1/2}$. The operation $(\mathbf{S}_{(Mn_t)\times(Jn_t)})^T\mathbf{x}_{(Mn_t)}$ would do the same but for all n_t subvectors of length M in the Mn_t-vector $\mathbf{x}_{(Mn_t)}$.

The right square root of ECO-RAP is obtained from $\mathbf{C}_K^{\odot q}$ and $\mathbf{S}$ using

$$\mathbf{W}^T = \widetilde{\mathbf{S}^T\mathbf{C}_K^{\odot q}}, \quad (5)$$

where $\mathbf{W}^T$ is a $(Jn_t) \times n$ matrix whose columns list the weighted spectral coefficients of $\mathbf{C}_K^{\odot q}$. The curved bar over the matrix on the right-hand side of (6) indicates that, as required by step (3), its column vectors have been normalized so that each column is a unit vector. We perform this column normalization so that the matrix

$$\mathbf{C}_{\text{ECO-RAP}} = \mathbf{W}\mathbf{W}^T = \left(\widetilde{\mathbf{S}^T\mathbf{C}_K^{\odot q}}\right)^T \widetilde{\mathbf{S}^T\mathbf{C}_K^{\odot q}} \quad (6)$$

is a correlation matrix. This is the ECO-RAP localization matrix. Note that $\mathbf{W}$ is a $n \times (Jn_t)$ matrix and hence has (Jn_t) columns. In general, the stronger the smoothing applied, the smaller the number J of basis functions required to represent the smoothing operator.

It is of interest to compare and contrast (6) with the SENCORP localization. Eq. 10 of Bishop and Hodyss (2007) gives $\mathbf{C}_{\text{SENCORP}} = [(\widetilde{\mathbf{C}_s^{\odot q}})^m]^{\odot r}$, where q and r are integers defining element-wise powers, m is an integer defining the order of a regular matrix product and $\mathbf{C}_s$ is the correlation matrix from a smoothed K member ensemble. One difference is the role of the smoother. In the SENCORP localization, the smoother is only ever applied to the raw ensemble perturbations whereas in ECO-RAP it is applied to powers of correlation matrices. The effect of this difference on localization properties was unclear to the authors and motivated the tests described in this paper. ECO-RAP has fewer tunable parameters than SENCORP. Another difference is that in the ECO-RAP approach, the finding of a concise square root of the localization matrix (5) is an integral part of the construction of the localization matrix whereas in the SENCORP approach one would have to perform additional calculations to find a concise square root.

The thick black line on Fig. 1c shows a column of the ECO-RAP localization matrix for the case where the jth diagonal elements $\lambda_f(j)$of Λ_f were chosen to be exponentially decreasing functions $\exp[-(\frac{k(j)}{d_f})^2]$ of the wavenumber $k(j)$ of the jth sinusoidal basis function. For the problem considered in Fig. 1, the parameter setting $d_f = 4$ gave a suitable localization function with $q = 6$. In general, the parameter d_f is a tunable parameter that partially controls the width of ECO-RAP localization. For fixed q, the smaller the d_f, the broader the ECO-RAP localization function.

Figure 1c illustrates how the columns of the ECO-RAP correlation matrix correspond to broad 'propagating' localization functions. It involves two tunable parameters, q, the power to which the raw ensemble correlations are raised and the parameter d_f controlling the spectral smoothing. Neither q nor d_f contain any propagation information. All of the propagation information has been imparted to the ECO-RAP matrix by the $\mathbf{C}_{64}^{\odot 6}$ matrix that has been used in its construction.

While the column of $\mathbf{C}^{\odot 6}$ displayed in Fig. 1c has a maximum value of 0.5 at the final time, the corresponding column of the ECO-RAP matrix has a maximum value of 0.8 at the final time. We attribute this useful ability of ECO-RAP to boost small but significant correlations in $\mathbf{C}^{\odot q}$ to the matrix product smoothing effect discussed in Bishop and Hodyss (2007; 2032 pp.). Experiments show that increasing the strength of the smoother $\mathbf{S}$ increases the boosting effect of the matrix product $(\widetilde{\mathbf{S}^T\mathbf{C}_K^{\odot q}})^T\widetilde{\mathbf{S}^T\mathbf{C}_K^{\odot q}}$ that defines ECO-RAP.

Figure 1 shows how the functions described by the columns of the ECO-RAP correlation matrix envelop the portions of the ensemble correlation function shown in Fig. 1b corresponding to small but true non-zero error correlations. The thick black line in Fig. 1d gives the correlation function that results from the element-wise product of the raw ensemble correlation function depicted in Fig. 1b with the ECO-RAP correlation function. In other words, it plots the relevant column of $\mathbf{C}_{64} \odot \mathbf{C}_{\text{ECO-RAP}}$.

Figure 1e shows the relevant column of the CALECO forecast error covariance matrix

$$\mathbf{P}^{\mathrm{f}}_{\mathrm{CALECO}} = \mathbf{P}^{\mathrm{f}}_{K} \odot \mathbf{C}_{\mathrm{ECO-RAP}}, \tag{7}$$

where $\mathbf{P}^{\mathrm{f}}_{K} = \sum_{i=1}^{K} \frac{\mathbf{x}_i \mathbf{x}_i^T}{K-1}$ is the sample covariance matrix. Note the very large amplitude of the covariances at the later time associated with the exponential growth of the Eady waves. Otherwise, the curves in Fig. 1e are similar to those in Fig. 1d.

Note that, in general, the size of the power q required to usefully suppress spurious non-zero sample correlations will be a decreasing function of the ensemble size K because the size of spurious correlations is a decreasing function of K. Indeed, if the distribution of spurious non-zero sample correlations was precisely known one could compute the power required to ensure that, say, 99% of spurious non-zero correlations were smaller than some threshold, for example, 0.01. Random number generators are easily employed to give approximations to distributions of spurious non-zero sample correlations and these too could be used to inform the choice of q. However, since the matrix product in step (2) can amplify spurious correlations as much as it can amplify real correlations, at present, we recommend 'trial and error' as the primary means of determining the power q.

Localized ensemble correlations have shorter correlation length scales than that of the sample correlation matrix of an infinite member ensemble. Furthermore, localization damages the dynamical balance of the ensemble covariance matrix (Lorenc, 2003). The broader the localization, the less the dynamical balance of the covariance model is damaged and the less the 'inherent' ensemble correlation length scale is shortened. As the ensemble size is increased, there is less need for tight localization as the spurious non-zero correlations are not as large. Broader localization is achieved in the ECO-RAP localization model by (1) taking a smaller element-wise power q, (2) using a stronger smoother $\mathbf{S}\mathbf{S}^T$ by increasing the weighting of large length scales relative to small length scales by decreasing d_f and/or (3) taking the option of smoothing the ensemble perturbations before forming $\mathbf{C}_K$ in step (1) of the ECO-RAP procedure.

The key computational difference between ECO-RAP and SENCORP is that ECO-RAP is expressed in terms of a square root $\mathbf{W}$ that may have far fewer columns $(Jn/M) = Jn_t$ than the total number of variables n. As shown in Part 2, this feature can yield profound computational advantages in atmospheric DA. A key functional difference to be demonstrated in Section 2.5 is that, under certain conditions, ECO-RAP has a limiting case that is equivalent to non-adaptive localization whereas SENCORP does not have this limit. This feature is theoretically attractive because it shows that ECO-RAP allows a spectrum of localizations ranging from the highly adaptive localizations obtained when q is small to the less adaptive localizations obtained when q is large. It is also practically attractive when the ensemble size is small and a large q value is required (see the experiment associated with Fig. 4).

2.3. Non-local 'localization'

Non-adaptive moderation of spurious ensemble correlations is primarily based on the assumption that the correlation between forecast errors of variables is a monotonically decreasing function of separation distance. In particular, it is normally assumed that the correlation between variables separated by more than a few 1000 km is zero. However, consider model error associated with, say, stratus clouds. Stratus clouds cover vast subtropical regions of the eastern portions of ocean basins and the arctic. Might not model error associated with stratus clouds be correlated over distances as large as the stratus regions themselves? Might not model error associated with one patch of stratus be correlated with the model error associated with a distant patch of stratus but not with the model error associated with the part of the atmosphere lying between the two patches? These considerations indicate the desirability of a method of moderating spurious ensemble correlations that do not depend on variable separation distance.

Figure 2a gives an example of a non-local error correlation function. The function features two broad peaks separated by half the circumference of the domain. To produce the error distribution associated with this non-local error distribution, a wavelet basis $\mathbf{V}$ comprised of two peaked functions was used. Thus, in this case, it was not the coefficients of spectral coefficients of initial errors that were uncorrelated, it was the coefficients of the differently positioned two peaked functions that were uncorrelated. Figures 2b–d show that ECO-RAP is just as capable of moderating spurious correlations in error distributions with non-local error correlation functions as it is with local error correlation functions.

2.4. Smoothing to decrease the tendency of ECO-RAP to localize error correlations in time

ECO-RAP temporal localization is firmly based on correlations of ensemble perturbations through time. When these correlations are weak then the ECO-RAP localization function will suppress these correlations even further. In some instances, ECO-RAP might overly suppress true but small correlations through time. One mechanism likely to weaken correlations through time is the rapid growth and non-linear saturation of small-scale processes such as convection. The tendency of ECO-RAP localization to suppress covariances at later times when such small-scale noise is present can be reduced by applying the ECO-RAP methodology to spatially smoothed ensemble perturbations. The idea is that the spatial smoothing filters out small-scale perturbations that have a weak correlation through time. To illustrate this point, we considered a very simple error distribution in which the final time error ε^{F} was equal to the initial time error ε^{I} plus white noise; in other words, $\varepsilon^{\mathrm{F}} = \varepsilon^{\mathrm{I}} + \zeta$ where ζ is a white noise perturbation. Figure 3 compares the time evolution of a sample correlation function from an unsmoothed ensemble

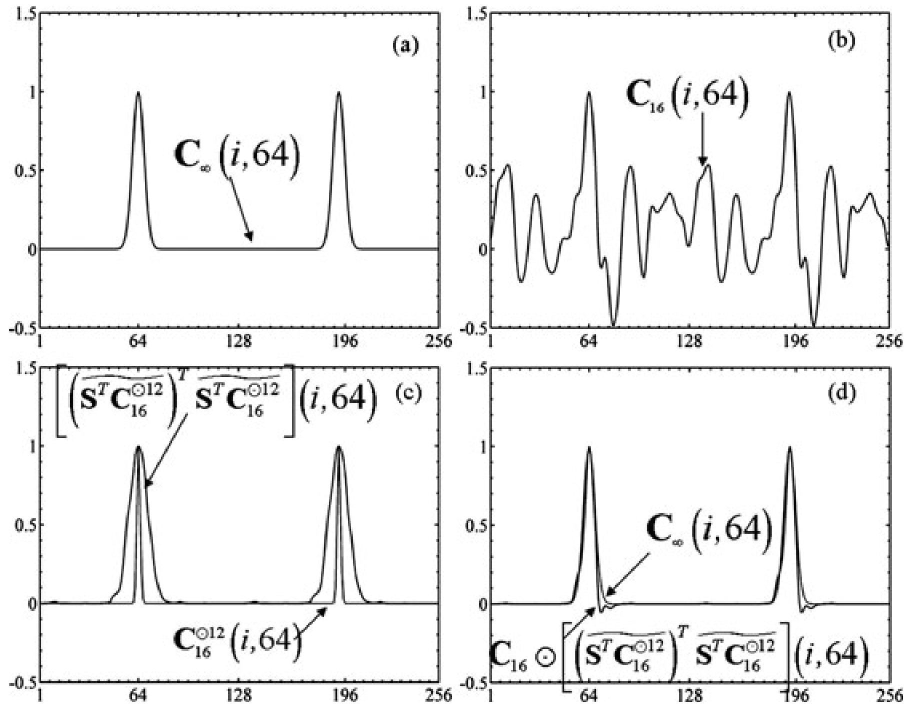

Fig. 2. In (a) is shown the true forecast error correlations with the 64th gridpoint variable for an error distribution in which significant correlations exist not only in the near neighbourhood of the 64th gridpoint but also in the neighbourhood of the 195th gridpoint. In (b) is shown the corresponding sample correlation function from a 16-member ensemble. The two thin peaks in (c) give the function shown in (b) raised to the element-wise 12th power. The two broader peaks in (c) give the ECO-RAP localization function. In (d) is shown the element-wise product of the ECO-RAP localization function shown in (c) with the sample correlation function shown in (b) together with the true correlation function $\mathbf{C}_\infty$.

with that of a smoothed ensemble for this error distribution. It shows that smoothing greatly increases the correlation between the initial time variable (128) and the final time variables (257–512). We have found this strategy to be helpful when working with ensembles of numerical weather forecasts (see Part 2 for details).

2.5. *Non-adaptive localization is a special case of ECO-RAP*

To see that non-adaptive localization functions are a special case of ECO-RAP note that if variables at only one time are considered then as the number of Hadamard powers q in (6) goes to infinity $\mathbf{C}^{\odot q}$ tends to the identity matrix provided that only autocorrelations are precisely equal to unity. Hence,

$$\lim_{q\to\infty}\mathbf{C}_{\mathrm{ECO-RAP}} = \left(\widetilde{\mathbf{S}^T\mathbf{C}_K^{\odot q}}\right)^T \widetilde{\mathbf{S}^T\mathbf{C}_K^{\odot q}} = \widetilde{\mathbf{S}}\widetilde{\mathbf{S}}^T. \tag{8}$$

Consequently, if $\widetilde{\mathbf{S}}$ was the square root of the correlation matrix associated with Gaspari and Cohn's (1999) function of compact support, then ECO-RAP would deliver localization identical to that provided by the Gaspari and Cohn localization function.

3. DA performance of the CALECO covariance model

3.1. *Idealized DA experiments*

Using idealized error models of the type described in Appendix A, we aim to compare the DA performance of the CALECO covariance model against the DA performance of the non-adaptively localized covariance model

$$\mathbf{P}^{\mathrm{f}}_{\mathrm{NA}} = \mathbf{P}^{\mathrm{f}}_K \odot \mathbf{C}_{\mathrm{NA}}, \tag{9}$$

where

$$\mathbf{C}_{\mathrm{NA}} = \begin{bmatrix} \mathbf{E}_{\mathrm{s}}\Lambda_G\mathbf{E}_{\mathrm{s}}^T & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{\mathrm{s}}\Lambda_G\mathbf{E}_{\mathrm{s}}^T & & \\ . & . & & \\ \mathbf{0} & & & \mathbf{E}_{\mathrm{s}}\Lambda_G\mathbf{E}_{\mathrm{s}}^T \end{bmatrix}, \tag{10}$$

where $\mathbf{E}_{\mathrm{s}}$ gives the sinusoidal orthonormal basis for the periodic domain and the jth element λ_{Gj} of the diagonal matrix Λ_G was defined by

$$\lambda_{Gj} = \frac{M\exp\left[-\frac{k(j)^2}{d_G^2}\right]}{\sum_{j=1}^{M}\exp\left[-\frac{k(j)^2}{d_G^2}\right]}. \tag{11}$$

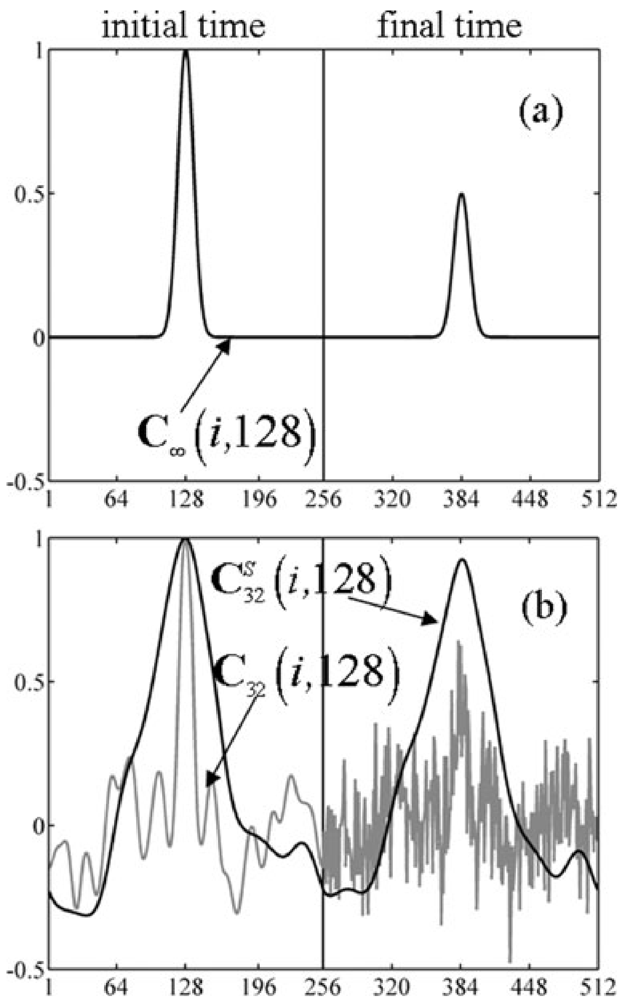

Fig. 3. The abscissa indicates the index of variables defining a one-dimensional state at two distinct times. In (a) is shown the true correlation function between initial and final time forecast errors with the error of the (initial time) 128th variable. Variables 1–256 pertain to initial time variables whereas variables 257–512 pertain to final time variables. This correlation function was obtained by assuming that final time error is equal to the initial time error plus white noise. In (b) is shown a comparison between the correlation from a raw 32 member ensemble $\mathbf{C}_{32}$ (grey line) and a smoothed 32 member ensemble $\mathbf{C}^s_{32}$ (black line).

The correlation matrix given by (9)–(11) provides a Gaussian localization function that does not propagate and whose width is invariant. When used as a localization correlation matrix, it has a qualitatively similar effect as the correlation matrix based on Gaspari and Cohn's (1999) compact polynomial that is often used in ensemble DA.

The estimate of a state that is distributed through time $\mathbf{x}^{\mathrm{a}}$ can be obtained using

$$\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{f}} = \mathbf{P}^{\mathrm{f}}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{f}}\right), \tag{12}$$

where $\mathbf{x}^{\mathrm{f}}$ is the first guess of the state at all times defining the assimilation window with error covariance $\mathbf{P}^{\mathrm{f}}$, $\mathbf{y}$ lists p observations (distributed through time) with error covariance $\mathbf{R}$, $\mathbf{H}\mathbf{x}^{\mathrm{f}}$ gives model forecasts of observations with estimated error covariance $\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^T$ and estimated covariance with forecast errors of $\mathbf{P}^{\mathrm{f}}\mathbf{H}^T$. It can be shown [Daley, 1991 (section 4.2), Cohn, 1997] that if the estimates of $\mathbf{P}^{\mathrm{f}}$, $\mathbf{R}$ and $\mathbf{H}$ are accurate then (12) gives the minimum error variance state estimate.

We can assess the DA performance of CALECO by comparing the error of state estimates $\mathbf{x}^{\mathrm{a}}$ made by replacing $\mathbf{P}^{\mathrm{f}}$ in (12) with either $\mathbf{P}^{\mathrm{f}}_{\mathrm{CALECO}}$ or $\mathbf{P}^{\mathrm{f}}_{\mathrm{NA}}$. Each independent test of DA performance requires the generation of an independent forecast error vector and an independent observation error vector. The term $\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{f}}$ in (12) is called the 'innovation vector'. To form the innovation vector, we first noted that

$$\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{f}} = \left(\mathbf{y} - \mathbf{y}^t\right) - \left(\mathbf{H}\mathbf{x}^{\mathrm{f}} - \mathbf{H}\mathbf{x}^t\right) = \varepsilon^{\mathrm{o}} - \mathbf{H}\varepsilon^{\mathrm{f}}, \tag{13}$$

where we assume that $\mathbf{y}^t = \mathbf{H}\mathbf{x}^t$ gives the true value of the observed variables and $\mathbf{x}^t$ gives the true state. The vectors ε^{o} and ε^{f} denote the observation error and forecast error vectors, respectively. Following (13), we formed random innovation vectors by subtracting randomly generated forecast errors interpolated to observation space from randomly generated observation error vectors. Observation error vectors were formed by making a random draw from a normal distribution with covariance $\mathbf{R}$ while forecast error vectors were formed by making a random draw from a normal distribution with covariance $\mathbf{P}^{\mathrm{f}}$ (see Appendix A for more details; The random number generator used to produce the errors was the LAPACK subroutine 'dlarnv').

When approximations to $\mathbf{P}^{\mathrm{f}}$ such as $\mathbf{P}^{\mathrm{f}}_{\mathrm{CALECO}}$ or $\mathbf{P}^{\mathrm{f}}_{\mathrm{NA}}$ are used in (12), suboptimal analysis corrections are obtained. It is the efficacy of these suboptimal corrections that we wish to compare and contrast. One measure of efficacy is the root mean square distance of a suboptimal correction from the optimal correction. Another is the root mean square analysis error.

To compute the analysis error, we noted that since the true state is given by $\mathbf{x}^t = \mathbf{x}^{\mathrm{f}} - \varepsilon^{\mathrm{f}}$ and since $\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{f}} = \varepsilon^{\mathrm{a}} - \varepsilon^{\mathrm{f}}$ where ε^{a} is the analysis error vector, it followed that the analysis error is the sum of the correction $\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{f}}$ and the forecast error; in other words,

$$\varepsilon^{\mathrm{a}} = \left(\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{f}}\right) + \varepsilon^{\mathrm{f}}. \tag{14}$$

While the specific analysis error associated with specific forecast and observation errors is a random variable, its covariance over a large number of independent trials can be obtained by performing a large number of independent trials and then computing the covariance of the analysis errors obtained from (14).

3.2. *Single observation DA experiment*

Consider a periodic 32 variable system. Suppose that the dynamics of the system are such that the error at each gridpoint is advected one gridpoint to its right each nominal 'time step'. In

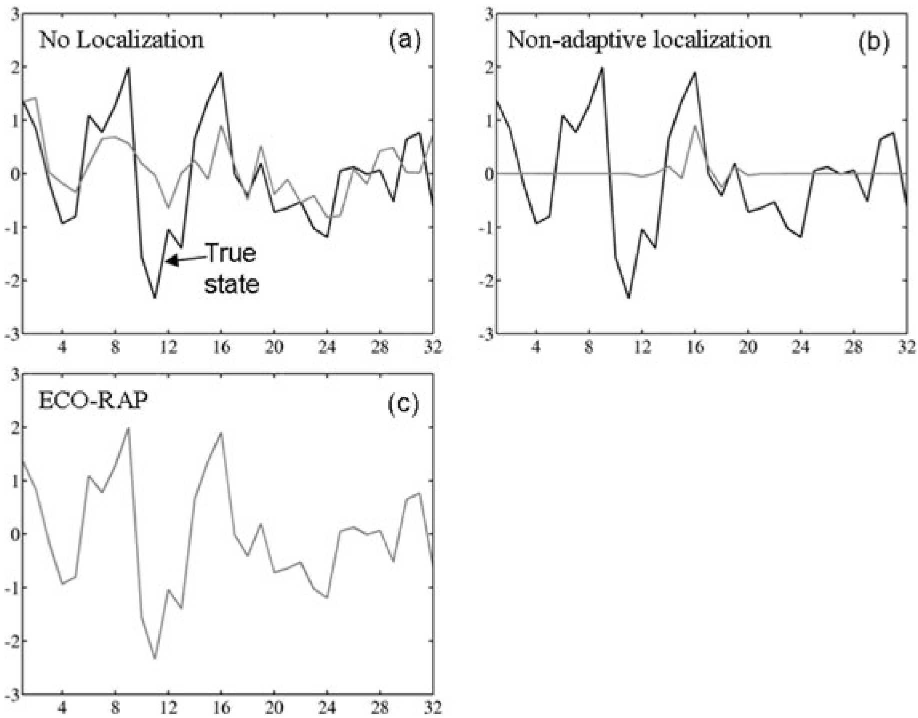

Fig. 4. The black line in (a) and (b) shows the true initial state. The grey lines in (a–c) give 16 member ensemble Kalman filter state estimates of the initial state from 32 consecutive observations of variable 16 when (a) no covariance localization is used, (b) when non-adaptive covariance localization is used and (c) when ECO-RAP localization is used. The black line cannot be seen in (c) because it is covered by the grey line.

this case, taking a single perfect observation at 32 consecutive time steps at the same location provides enough information to perfectly define the state of the system at the beginning of the time-series of observations. Hence, a good model of temporal error covariances ought to enable the reconstruction of the entire initial state from 32 consecutive observations of a single variable. Figure 4 shows initial state estimates obtained from three different ensemble DA schemes after the assimilation of 32 consecutive error free observations of gridpoint 16 using a 16-member ensemble.

Figure 4a shows that when no localization is used, a very noisy estimate of the initial state is obtained. Figure 4b shows that when non-adaptive error covariance localization is used, the state estimate is poor. In contrast, Fig. 4c shows that when the ensemble correlations are adaptively localized with ECO-RAP, the true initial state is recovered. The ECO-RAP tuning parameters used to obtain Fig. 4c were $q = 40$ and $d_f = 8$. The optimally tuned non-adaptive localization used the correlation function given by $d_G = 6$. All of these parameters were obtained by trial and error.

Previously it was mentioned that the power q could be estimated directly from knowledge of the distribution of spurious correlations associated with a K-member ensemble. In this error system, however, the maximum correlation between a variable and variables at other times was equal to 1. Hence, in this case, even an infinite value of q would not entirely eliminate correlations through time. For this reason, there was a wide range of high values of q and d_f that resulted in systems perfectly able to reconstruct the initial condition. The parameters $q = 40$ and $d_f = 8$ gave us a localization function that, at any instant, looked like a Gaussian non-adaptive localization function. However, unlike stationary non-adaptive Gaussian localization functions, the ECO-RAP localization function was moving with the flow.

3.3. *Variable correlation length scale experiment*

Error correlation length scale is likely to depend on observation density (Daley, 1992, fig. 6) and the length scale and the spectrum of instabilities supported by the true state. It is conceivable that the error correlation length scale in a particular region might be broad under the influence of large-scale descent but narrow when under the influence of large-scale ascent and/or convectively unstable conditions.

For a 256 variable periodic system in which every second variable was observed (128 observations in all), we tested the ability of ECO-RAP localization to adapt to such changes in the error correlation length scale by tuning it for one error correlation length scale and then documenting its performance for error correlation length scales that it was not tuned for. To be specific, we tuned the ECO-RAP parameters q and d_f to minimize analysis error variance when the error correlation length scale of the true forecast error distribution was given by the wavenumber scale

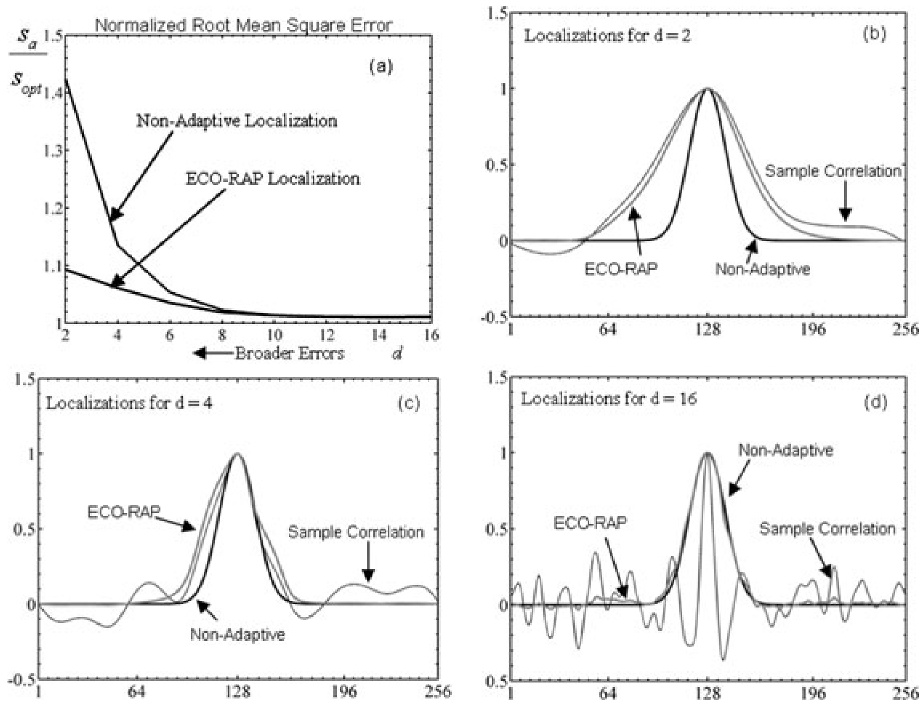

Fig. 5. In (a) the ordinate is s_a/s_{opt} the ratio of root mean square analysis error from the suboptimal scheme s_a divided by the root mean square error from the optimal scheme s_{opt} while the abscissa gives the wavenumber scale of the correlation functions of the true forecast error covariance matrix. The thin grey lines in (b–d) give 64 member ensemble correlation functions corresponding to $d = 2$, $d = 4$ and $d = 16$, respectively. The medium grey lines in (b–d) give the corresponding ECO-RAP localization functions while the black lines give the non-adaptive localization function.

$d = 16$. A 64-member ensemble was used. Similarly, we tuned a non-adaptive Gaussian localization function to minimize analysis error variance at this very same error correlation length scale. Optimization via trial and error for ECO-RAP was obtained with the parameters[4] $q = 3$ and $d_f = 5$ while non-adaptive Gaussian localization was optimized with $d_G = 5$. Without changing these tuning parameters, we then evaluated the performance of both non-adaptive and ECO-RAP localizations for forecast error distributions with $d \neq 16$.

The performance was measured by computing the domain averaged root mean square error of analyses produced (1) with ECO-RAP localization, (2) with non-adaptive localization and (3) with optimal DA (The optimal scheme uses the true forecast error covariance matrix in 12). The statistical significance of all results at all d values was tested by repeating each experiment 16 times with differing seeds for the random number generators. It was found that the performances of ECO-RAP and non-adaptive localizations were statistically indistinguishable for $d \geq 10$; however, for $d < 8$ (longer correlation length scales in physical space) it was found that ECO-RAP localization outperformed non-adaptive localization. Figure 5a summarizes the results of this experiment and clearly demonstrates the superiority of ECO-RAP localization over non-adaptive localization for small d values. These smaller d values correspond to longer error correlation length scales in physical space.

The reason for the superior DA performance of ECO-RAP with small d values is evident in Figs. 5b–d. The thin grey line in Fig. 5b gives the very broad 64 ensemble member sample correlation function for $d = 2$, that is, indicative of the true error correlation length scale. The thick black line in Fig. 5b gives the non-adaptive localization function that was optimal for $d = 16$. It is narrower than the sample correlation function. Hence, the product of the non-adaptive localization function with the ensemble sample correlation function yields a correlation function much narrower than the true error correlation function. In contrast, the ECO-RAP correlation function given by the medium grey line in Fig. 5b is much broader than the non-adaptive error correlation function. Consequently, it yields a localized correlation function closer to the true error correlation function than that provided by non-adaptive localization.

Optimal localization functions are generally broader than the true error correlation function. Figure 5b shows that for $d = 2$, the ECO-RAP localization function is not broader than the true error correlation function. As such, it failed to adapt as much as it needed to for $d = 2$. However, Fig. 5c shows that this aspect is improved for $d = 4$ because at $d = 4$ the ECO-RAP localization

[4]Note that the $q = 3$ used in this experiment is smaller than that used in the propagating error experiment. This is primarily because the ensemble size is four times larger than in the propagating error experiment.

function is broader than the true error correlation function. Figure 5d shows that the ECO-RAP and the non-adaptive localization functions are very similar for the d value that they were both tuned for ($d = 16$) as expected.

4. Conclusions

A new method for ensemble DA has been described that uses ECO-RAP for generating localization functions that move with the true error correlation functions and that adapt to the width of the true error correlation function. The following summarizes our findings for the gallery of one-dimensional error models considered in this paper.

(1) ECO-RAP localization functions propagated at the same speed as the true error correlation functions in the examples considered. These examples included both dispersive and non-dispersive error dynamics.

(2) Because the width and location of ECO-RAP functions are directly linked to the location and width of larger amplitude ensemble correlations, ECO-RAP correlation functions are not necessarily monotonically decreasing functions of separation distance and hence do not necessarily 'localize' correlations. It was shown that ECO-RAP is well suited to error distributions in which the true error correlation function is non-local and features two distinct peaks separated by a large area of zero correlation. Such error distributions require moderation of spurious correlations not localization—ECO-RAP meets this requirement.

(3) ECO-RAP localization enables the initial state of a 32-variable one-dimensional periodic system featuring uniform advection of a passive scalar to be perfectly reconstructed from a time-series of observations of a single variable using ensemble DA with a 16-member ensemble. Attempts to do this without localization or with non-adaptive localization failed.

(4) Even if ECO-RAP has been tuned for a fixed error correlation length scale, it readily adapts to changes in the true error correlation length scale and consequently outperforms non-adaptive localization in situations of fluctuating error correlation length scale.

(5) In situations where the true error correlation length scale is fixed, the DA performance of ECO-RAP is statistically indistinguishable to that of non-adaptive localization.

We conclude by noting that, in a large range of experiments not reported here, we attempted to find an error model in which ECO-RAP would outperform or underperform Bishop and Hodyss' (2007) SENCORP method. However, we were always able to find some tuning of each scheme that rendered the differences between the performances of each scheme insignificant. Thus, the advantage of ECO-RAP over SENCORP is that it enables the profound computational cost saving procedures described in Part 2.

5. Acknowledgments

Discussions with Dr. James A. Hansen improved this work. CHB and DH acknowledge support from ONR Project Element 0602435N, Project Number BE-435-003, ONR Grant number N0001407WX30012 and from NOAA THORPEX Grant number NA04AANRG0233 is acknowledged. This research was performed while DH held a National Research Council Research Associateship Award at the Naval Research Laboratory.

6. Appendix A: Error distributions

A.1. Stationary error distributions using a sinusoidal basis

To begin, consider a non-propagating system whose state is defined by the value of M temperature (T) values collocated with each of M gridpoints. To generate random T errors that were spatially correlated, we used

$$\varepsilon = \Sigma \mathbf{E}_{\mathrm{s}} \mathbf{c}, \tag{A1}$$

where Σ is a diagonal matrix listing the standard deviations of the T errors at each gridpoint. $\mathbf{E}_{\mathrm{s}}$ is a $M \times M$ orthonormal matrix whose M columns list the sinusoidal vectors spanning the M-dimensional gridpoint space. The ith row $(\mathbf{E}_{\mathrm{s}})_i$ of $\mathbf{E}_{\mathrm{s}}$ is given by

$$(\mathbf{E}_{\mathrm{s}})_i = \sqrt{\frac{2}{M}}\left[\sqrt{\frac{1}{2}}, \sin(z_i), \cos(z_i), \sin(2z_i), \cos(2z_i), \ldots, \sqrt{\frac{1}{2}} \cos\left[\left(\frac{M}{2}\right) z_i\right]\right], \tag{A2}$$

where $z_i = 2\pi i / M$ gives the position of the ith gridpoint on a periodic M-dimensional domain of non-dimensional length 2π. The random M-vector $\mathbf{c}$ is a normally distributed vector whose covariance is given by the diagonal matrix Γ; in other words $\mathbf{c} \sim N(\mathbf{0}, \Gamma)$. Note that the statement that Σ represents the standard deviations of temperature error is only true if the diagonal matrix Γ is defined in a way that ensures that $\mathbf{C} = \langle \mathbf{E}_{\mathrm{s}} \mathbf{c} \mathbf{c}^T \mathbf{E}_{\mathrm{s}}^T \rangle = \mathbf{E}_{\mathrm{s}} \Gamma \mathbf{E}_{\mathrm{s}}^T$ is a correlation matrix. For $\mathbf{C}$ to be a correlation matrix all of its diagonal elements must equal unity. This is achieved by (1) making sure that trace(Γ)/$M = 1$ and (2) ensuring that the eigenvalues associated with cosine and sine functions of the same wavenumber have precisely the same value. We met these conditions by letting the eigenvalues be given by

$$\gamma_j = \frac{M \exp\left[-\frac{k(j)^2}{d^2}\right]}{\sum_{j=1}^{M} \exp\left[-\frac{k(j)^2}{d^2}\right]}, \tag{A3}$$

where k is the wavenumber of the jth sinusoidal eigenvector corresponding to the jth eigenvalue. $k(j)$ = (integer part of the quotient j/2). d is a scalar constant.

A.2. Propagating error distributions using a sinusoidal basis

Propagating error distributions were defined by letting the M-vectors $\mathbf{T}^{\mathrm{I}} = [T_1^{\mathrm{I}}, T_2^{\mathrm{I}}, ..., T_M^{\mathrm{I}}]^T$ and $\mathbf{T}^{\mathrm{F}} = [T_1^{\mathrm{F}}, T_2^{\mathrm{F}}, ..., T_M^{\mathrm{F}}]^T$ list the forecast values of the T variables at initial and final times, respectively. Let $\mathbf{T}^{\mathrm{I}t}$ and $\mathbf{T}^{\mathrm{F}t}$ denote the M-vectors listing the true values of the T variables at the initial and final times, respectively, so that

$$\varepsilon^{\mathrm{I}} = \mathbf{T}^{\mathrm{I}} - \mathbf{T}^{\mathrm{I}t} \text{ and } \varepsilon^{\mathrm{F}} = \mathbf{T}^{\mathrm{F}} - \mathbf{T}^{\mathrm{F}t} \tag{A4}$$

give the forecast errors associated with the initial and final times, respectively. We used (A1) to generate random T errors that were spatially correlated at the initial time such that $\varepsilon^{\mathrm{I}} = \Sigma^{\mathrm{I}}\mathbf{E}_{\mathrm{s}}\mathbf{c}^{\mathrm{I}}$ where $\mathbf{c}^{\mathrm{I}} \sim N(\mathbf{0}, \Gamma^{\mathrm{I}})$.

Having generated a random error at the initial time, it remained to create a random error at the final time that represented error propagation through space. Wave propagation from the initial time to the final time can be simulated as follows. First, note that the double angle formulae for sines and cosines implies that

$$\begin{aligned}[\sin(kz_j), \cos(kz_j)]\mathbf{B}_k = [&\sin[k(z_j - z_c(k))],\\ &\cos[k(z_j - z_c(k))]],\end{aligned} \tag{A5}$$

where $z_c(k)$ is the propagation distance at wavenumber k and $\mathbf{B}_k$ is the 2×2 matrix defined by

$$\mathbf{B}_k = \begin{bmatrix} \cos(kz_c), & \sin(kz_c) \\ -\sin(kz_c), & \cos(kz_c) \end{bmatrix}. \tag{A6}$$

Consequently, propagated final time error takes the form

$$\varepsilon^{\mathrm{F}} = \mathbf{D}\Sigma^{\mathrm{I}}\mathbf{E}_{\mathrm{s}}\mathbf{B}\mathbf{c}^{\mathrm{I}}, \tag{A7}$$

where

$$\mathbf{B} = \begin{bmatrix} 1 \; ... & & 0 \\ 0 \; \mathbf{B}_1 \; \mathbf{0} & & \\ . \;\; \mathbf{0} \;\; \mathbf{B}_2 & & \\ . & . & \\ & \mathbf{B}_{\frac{m-2}{2}} & \\ 0 & & 1 \end{bmatrix} \tag{A8}$$

is an orthogonal matrix that propagates wave components. $\mathbf{D}$ is a diagonal matrix.

Letting

$$\left\langle \varepsilon^{\mathrm{F}}\varepsilon^{\mathrm{F}T} \right\rangle = \Sigma^{\mathrm{F}}\mathbf{C}^{\mathrm{FF}}\Sigma^{\mathrm{F}}, \tag{A9}$$

where Σ^{F} and $\mathbf{C}^{\mathrm{FF}}$ are standard deviation and correlation matrices, respectively, and taking the expectation of the outer product of (A7) gives

$$\Sigma^{\mathrm{F}}\mathbf{C}^{\mathrm{FF}}\Sigma^{\mathrm{F}} = \mathbf{D}\Sigma^{\mathrm{I}}\mathbf{E}_{\mathrm{s}}\mathbf{B}\Gamma^{\mathrm{I}}\mathbf{B}^T\mathbf{E}_{\mathrm{s}}^T\Sigma^{\mathrm{I}}\mathbf{D}. \tag{A10}$$

Since $\mathbf{B}$ and $\mathbf{E}_{\mathrm{s}}$ are both orthogonal matrices and since $\mathrm{trace}(\Gamma^I)/M = 1$ and the sine and cosine columns of $\mathbf{E}_{\mathrm{s}}\mathbf{B}$ having the same wavenumber are associated with the same eigenvalue, it follows that the diagonal elements of $\mathbf{E}_{\mathrm{s}}\mathbf{B}\Gamma^{\mathrm{I}}\mathbf{B}^T\mathbf{E}_{\mathrm{s}}^T$ are all equal to 1. Hence, the diagonal elements of (A10) are defined by

$$\left(\Sigma^{\mathrm{F}}\right)^2 = \mathbf{D}\left(\Sigma^{\mathrm{I}}\right)^2\mathbf{D} = \mathbf{D}^2\left(\Sigma^{\mathrm{I}}\right)^2. \tag{A11}$$

The forecast error correlation matrix $\mathbf{C}_{\mathrm{FI}}$ between initial time and final time forecast errors is given by $\mathbf{C}_{\mathrm{FI}} = (\Sigma^{\mathrm{F}})^{-1}\langle\varepsilon^{\mathrm{F}}(\varepsilon^{\mathrm{I}})^T\rangle(\Sigma^{\mathrm{I}})^{-1} = (\mathbf{D}\Sigma^{\mathrm{I}})^{-1}\langle\varepsilon^{\mathrm{F}}(\varepsilon^{\mathrm{I}})^T\rangle(\Sigma^{\mathrm{I}})^{-1}$, right multiplying (A7) by $(\varepsilon^{\mathrm{I}})^T = (\mathbf{c}^{\mathrm{I}}t)^T\mathbf{E}_{\mathrm{s}}^T\Sigma^{\mathrm{I}}$ taking the expectation and then right multiplying by $(\mathbf{D}\Sigma^{\mathrm{I}})^{-1}$ and left multiplying by $(\Sigma^{\mathrm{I}})^{-1}$ yields

$$\begin{aligned}(\mathbf{D}\Sigma^{\mathrm{I}})^{-1}\langle\varepsilon^{\mathrm{F}}(\varepsilon^{\mathrm{I}})^T\rangle(\Sigma^{\mathrm{I}})^{-1} &= (\mathbf{D}\Sigma^{\mathrm{I}})^{-1}\mathbf{D}\Sigma^{\mathrm{I}}\mathbf{E}_{\mathrm{s}}\mathbf{B}\Gamma\mathbf{E}_{\mathrm{s}}^T\Sigma^{\mathrm{I}}(\Sigma^{\mathrm{I}})^{-1}\\ &= \mathbf{E}_{\mathrm{s}}\mathbf{B}\Gamma\mathbf{E}_{\mathrm{s}}^T = \mathbf{C}_{\mathrm{FI}}.\end{aligned} \tag{A12}$$

Finally, note that the aforementioned error generation mechanisms imply that the full forecast error covariance matrix $\mathbf{P}^{\mathrm{f}}$ is given by

$$\begin{aligned}\mathbf{P}^{\mathrm{f}} = \left\langle \begin{bmatrix} \varepsilon^{\mathrm{I}} \\ \varepsilon^{\mathrm{F}} \end{bmatrix} \left[\left(\varepsilon^{\mathrm{I}}\right)^T \; \left(\varepsilon^{\mathrm{F}}\right)^T\right] \right\rangle = \begin{bmatrix} \Sigma^{\mathrm{I}} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}\Sigma^{\mathrm{I}} \end{bmatrix}\\ \times \begin{bmatrix} \mathbf{C}_{\mathrm{II}} & \mathbf{C}_{\mathrm{IF}} \\ \mathbf{C}_{\mathrm{FI}} & \mathbf{C}_{\mathrm{FF}} \end{bmatrix} \begin{bmatrix} \Sigma^{\mathrm{I}} & \mathbf{0} \\ \mathbf{0} & \Sigma^{\mathrm{I}}\mathbf{D} \end{bmatrix}.\end{aligned} \tag{A.13}$$

A.3. Propagating Eady waves error distribution

The dispersion relation for Eady waves takes the form (Pedlosky 1987; pp. 523–531):

$$c = \frac{1}{2} \pm \frac{1}{\nu}\sqrt{\left(\frac{\nu}{2} - \coth\left(\frac{\nu}{2}\right)\right)\left(\frac{\nu}{2} - \tanh\left(\frac{\nu}{2}\right)\right)}, \tag{A14}$$

where ν is the wavenumber non-dimensionalized by the Rossby radius of deformation. The exponential growth rate of these waves is proportional to $\nu(k)\mathrm{Im}\,(c)$. The imaginary part of (A14) is zero for $\nu > 2.4$ but is maximized at $\nu = 1.6$. We choose the positive branch of the square root in (A14) so that (1) decaying modes are excluded (see Bishop et al., 2003 for motivation) and (2) the neutral modes ($\nu > 2.4$) are associated with the lower troposphere rather than the upper troposphere. For Fig. 1, we scale the wavenumber with $\nu = bk$, where $b = 0.1$. This value of b makes $k = 16$ the most unstable wavenumber. A wavepacket centred at the most unstable wave is defined by choosing the diagonal matrix $\Gamma = \mathrm{diag}(\gamma_1, \gamma_2, ..., \gamma_m)$ such that

$$\gamma_j \propto \exp\left(-\frac{[k(j) - 16]^2}{d_w^2}\right), \tag{A15}$$

where for Fig. 1, $d_w = 8$. This wavepacket is propagated forward in time using $z_c(k) = \mathrm{Re}(c)\Delta t$ and

$$(\mathbf{D})_{jj} = \exp\left\{\nu\,[k(j)]\,\mathrm{Im}\,[c\,(\nu\,[k(j)])]\,\Delta t\right\}. \tag{A16}$$

in (A6)–(A7), where we set the non-dimensional time $\Delta t = 20$. Assuming the parameters $U = 30$ m s^{-1} and

$L = 1000$ km, $\Delta t = 20$ corresponds to a non-dimensional time of $L\Delta t/U \approx 7$ d. (Recall that j indexes the jth sinusoidal eigenfunction).

A.4. Non-local error correlation function error distribution

To create the non-local error distribution shown in Fig. 2 we let the ith element of the jth column v_{ij} of the matrix $\mathbf{V}$ of error basis functions in eq. (1) be given by

$$v_{ij} = b\left[\exp\left(-\frac{l_{ij}^2}{L^2}\right) + \exp\left(-\frac{l_{i(j+0.5M)}^2}{L^2}\right)\right]. \quad \text{(A17)}$$

l_{ij} and $l_{i(j+0.5M)}$, respectively give the distances between the ith and jth and the ith and $(j + 0.5M)$th gridpoints. Since the domain is periodic with length M, we interpret it as a line of points on a circle so that distance between gridpoints always corresponds to an arc length around a circle. L gives a length scale. The constant b is used to control the magnitude of the variance of error realizations produced by the stochastic equation

$$\varepsilon = \mathbf{Va}, \text{ where } \mathbf{a} \sim N(\mathbf{0}, \mathbf{I}). \quad \text{(A18)}$$

For Fig. 2, $L = 3.65$, $M = 256$ and $b = 0.11$.

References

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. 2009. Ensemble covariances adaptively localized with ECO-RAP. Part 2: a strategy for the atmosphere. *Tellus* **61A**, 97–111.

Bishop, C. H., Etherton, B. J. and Majumdar, S. J. 2001. Adaptive sampling with the ensemble transform Kalman Filter, part I: theoretical aspects. *Mon. Wea. Rev.* **129**, 420–436.

Bishop, C. H., Reynolds, C. A. and Tippett, M. K. 2003. Optimization of the fixed observing network in a simple model. *J. Atmos. Sci.* **60**, 1471–1489.

Bowler, N. E. 2006. Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus* **58A**, 538–548.

Cohn, S. E. 1997. An introduction to estimation theory. *J. Meteorol. Soc. Jpn.* **1B**, 257–288.

Daley, R. 1991. *Atmospheric Data Analysis*. Cambridge University Press, Cambridge, 457 pp.

Daley, R. 1992. Forecast-error statistics for homogeneous and inhomogeneous observation networks. *Mon. Wea. Rev.* **120**, 627–643.

Eady, E. T. 1949. Long waves and cyclone waves. *Tellus* **1**, 17–31.

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

Gaspari, G. and Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. *Quart. J. R. Meteorol. 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., Mitchell, H. L., Pellerin, G., Buehner, M. Charron, M., and co-authors. 2005. Atmospheric data assimilation with an ensemble Kalman Filter: results with real observations. *Mon. Wea. Rev.* **133**, 604–620.

Hunt, B. R., Kalnay, E., Kostelich, E., Ott, E., Patil, D., and co-authors. 2004. Four-dimensional ensemble Kalman filtering. *Tellus A* **56**, 273–277.

Khare, S. P. and Anderson, J. L. 2006. An examination of ensemble filter based adaptive observation methodologies. *Tellus* **58A**, 179–195.

Khare, S. P., Anderson, J. L., Hoar, T. J. and Nychka, D. 2008. An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems. *Tellus* **60A**, 97–112.

Lorenc, A. C. 2003. The potential of the ensemble Kalman filter for NWP—a comparison with 4D-VAR. *Quart. J. R. Meteorol. Soc.* **129**, 3183–3203.

Lorenz, E. N. 1996. *ECMWF Seminar on Predictability* Volume 1, Reading, United Kingdom , 1 pp.

Majumdar, S. J., Bishop, C. H., Szunyogh, I. and Toth, Z. 2001. Can an Ensemble Transform Kalman Filter predict the reduction in forecast error variance produced bytargeted observations? *Quart. J. R. Meteorol. Soc.* **127**, 2803–2820.

Miyoshi, T. and Yamane, S. 2007. Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.* **135**, 3841–3861.

Pedlosky, J. 1987. *Geophysical Fluid Dynamics*. Springer-Verlag, New York , 710 pp.

Szunyogh, I., Kostelic, E. J., Gyarmati, G., Kalnay, E., Hunt, B. R., and co-authors. 2008. A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus* **60A**, 113–130.

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