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

Tellus (2009), 61A, 97–111

TELLUS

Ensemble covariances adaptively localized with ECO-RAP. Part 2: a strategy for the atmosphere

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

Part 1's localization method, Ensemble COrrelations Raised to A Power (ECO-RAP), is incorporated into a Local Ensemble Transform Kalman Filter (LETKF). Because brute force incorporation would be too expensive, we demonstrate a factorization property for Part 1's Covariances Adaptively Localized with ECO-rap (CALECO) forecast error covariance matrix that, together with other simplifications, reduces the cost. The property inexpensively provides a large CALECO ensemble whose covariance is the CALECO matrix. Each member of the CALECO ensemble is an element-wise product between one raw ensemble member and one column of the square root of the ECO-RAP matrix. The LETKF is applied to the CALECO ensemble rather than the raw ensemble. The approach enables the update of large numbers of variables within each observation volume at little additional computational cost. Under plausible assumptions, this makes the CALECO and standard LETKF costs similar. The CALECO LETKF does not require artificial observation error inflation or vertically confined observation volumes both of which confound the assimilation of non-local observations such as satellite observations. Using a 27 member ensemble from a global Numerical Weather Prediction (NWP) system, we depict four-dimensional (4-D) flow-adaptive error covariance localization and test the ability of the CALECO LETKF to reduce analysis error.

1. Introduction

Bishop and Hodyss (2009; hereafter Part 1, this issue) indicates that the adaptive ensemble covariance localization provided by ECO-RAP is superior to non-adaptive localization when error correlation length scales vary and/or when errors propagate significant distances over the data assimilation (DA) time window. To realize these benefits in an ensemble DA scheme for a sophisticated NWP model, a practical method for the incorporation of ECO-RAP in ensemble DA schemes must be outlined. This is the prime objective of this paper.

Currently, ensemble DA schemes fall into two categories: schemes that use serial observation processing and schemes that use observation volumes. Serial observation processing schemes sequentially process small batches of observations; each batch of observations is used to update the minimum error variance state estimate and the ensemble based estimate of its error covariance (Houtekamer et al., 2005; Whitaker et al., 2004, 2008; Khare et al., 2008). On the other hand, Local Ensemble Transform Kalman Filters (LETKFs) (Bowler, 2006; Hunt et al., 2007; Miyoshi and Yamane, 2007; Szunyogh et al., 2008) feature overlapping observation volumes in which all observations within the observation volume are used to estimate the state and error covariance of variables near the centre of the volume. The calculations required for each LETKF observation volume are independent and can be performed in parallel. While one could envisage ways of incorporating ECO-RAP localization into serial observation processing schemes, here we focus on incorporating ECO-RAP into the LETKF.

Hunt et al. (2007); Miyoshi and Yamane (2007) and Szunyogh et al. (2008) have implemented LETKF schemes in which observation error variance is (artificially) made to be an increasing function of distance from the centre of the observation volume. The technique described here provides localization without recourse to the artificial inflation of observation error variances.

Miyoshi and Yamane (2007) and Szunyogh et al. (2008) have both employed vertical covariance localization schemes that require judgments to be made about the extent to which observations at some distance above or below a grid point should be allowed to influence the analysis of variables at that grid point. Such distance dependent judgments are problematic with observations that represent vertical integrals of the atmospheric state because some of the variables that contribute to the integral are close to the grid point of interest while other variables that contribute to the integral are distant from the grid point (Fertig et al., 2007). Radiance observations from satellites are of this type.

*Corresponding author.
e-mail: bishop@nrlmry.navy.mil
DOI: 10.1111/j.1600-0870.2008.00372.x

More recently, Miyoshi and Sato (2007) appear to recommend that the observation error variance that is assigned to satellite observations in the LETKF be inversely proportional to the value of the normalized-satellite-observation-operator-sensitivity-function at the location of the grid point being updated. We think that this approach is also unsatisfactory because, for example, satellite observations with very shallow sensitivity functions would produce corrections more localized in the vertical than conventional observations whereas a statistically consistent treatment would not exhibit such behaviour.

Since radiance observations from space are the most numerous of all observation types, the development of ensemble methods that treat them in a statistically consistent way is of some importance. The technique described here provides statistical consistency turning an ensemble with tens of members into an ensemble with thousands of members. With such a large ensemble, the LETKF does not require vertical localization beyond what has been implicitly imparted by the ensemble enlargement. Each member of the enlarged ensemble can be put into radiance space and hence the covariance between first guess errors of radiances and model variables can be handled in a way that is perfectly consistent with the CALECO forecast error covariance model.

Standard LETKF observation volumes small enough to filter spurious ensemble perturbations are also problematic for four-dimensional (4-D) DA because errors propagate. (Errors moving at a typical jet level wind speed of 30 ms^{-1} travel 650 km every 6 h). The longer the assimilation time window, the further they propagate. Ensembles can predict strong correlations through time as accurately as they predict strong correlations through space and across variables. Hence, observation volumes small enough to remove spurious ensemble correlations may also remove observations that, because of error propagation, have a non-spurious correlation with an analysis variable. The larger observation volumes of the CALECO LETKF provide room for error propagation while suppressing spurious correlations with the propagating localization functions provided by ECO-RAP.

Section 2 shows that the costs of preparing an unfactorized CALECO forecast error covariance model for use in the LETKF would be greater than the cost of performing the LETKF update. Section 3 introduces a factorization theorem for CALECO and other simplifications that reduce the cost of the preparation steps to a small fraction of the cost required for the LETKF update. Section 4 shows how the ability of the CALECO LETKF to accurately and simultaneously update a larger number of grid points per observation volume than the standard LETKF might make the CALECO LETKF faster than the standard LETKF. Section 5 applies the CALECO LETKF to a 27 member ensemble generated using the Navy Operational Global Atmospheric Prediction System (NOGAPS) (Hogan and Rosmond, 1991). It illustrates 4-D adaptive localization functions and demonstrates the ability of the CALECO LETKF to reduce analysis error. Conclusions follow in Section 6.

2. The high cost of a brute force CALECO LETKF

The LETKF requires the square root of the forecast error covariance matrix. The columns of this square root correspond to ensemble members. The LETKF creates correction vectors for the ensemble mean and the ensemble members by applying Bishop et al.'s (2001) ETKF to the ensemble members within local observation volumes. Hence, a brute force CALECO LETKF involves evaluating the CALECO matrix within the observation volume and then finding its square root. The first part of this section outlines the computational costs of such preparatory steps (P1–P3). The second part of this section gives the costs of the steps (E1–E5) involved in using the ETKF to compute the posterior mean and covariance of the ensemble.

(P1) Define a 4-D observation rectangular volume in space–time with sides n_x,n_y,n_z and n_t grid points long in the longitudinal, latitudinal, vertical and temporal directions. Let $n_g = n_x n_y n_z n_t$ denote the number of grid points in each volume. Let n_v denote the number of variables at each grid point and let $n = n_g n_v$ denote the number of variables within each observation volume. Let n_u denote the number of grid points near the centre of the observation volume that are to be updated. Let p denote the number of observations within each observation volume and K the number of ensemble members. The values assigned to these variables for the example given in Section 5 are given in Table 1

(P2) Compute the CALECO matrix $\mathbf{P}^f_{\mathrm{CALECO}} = \mathbf{P}^f_K \odot \mathbf{C}_{\mathrm{ECORAP}} = \mathbf{P}^f_K \odot \widetilde{\mathbf{C}^{\odot q}\mathbf{S}}\widetilde{\mathbf{S}^T\mathbf{C}^{\odot q}}$ in each observation volume where $\mathbf{P}^f_K$ is the sample covariance matrix of a raw K member ensemble and $\mathbf{C}_{\mathrm{ECORAP}} = \widetilde{\mathbf{C}^{\odot q}\mathbf{S}}\widetilde{\mathbf{S}^T\mathbf{C}^{\odot q}}$ gives the ECO-RAP localization matrix. $\mathbf{C}^{\odot q}$ is the correlation matrix obtained by raising every sample correlation of the raw ensemble to the qth power. $\mathbf{S}\mathbf{S}^T$ is a smoothing operator; $\mathbf{S}^T$ is the forward transformation

Table 1. List of parameters discussed in Section 2 and the values assigned to them for the NOGAPS experiments discussed in Section 5

Parameter	Meaning	Chosen value
n_x	number of grid points in x-direction	7
n_y	number of grid points in y-direction	7
n_z	number of grid points in z-direction	30
n_t	number of grid pts in t-direction	3
n_g	number of grid points in observation volume	4410
n_v	number of variables at each grid point	3
n	number of variables in observation volume	13230
n_u	number of grid points updated in obs volume	90
p	number of observations in each obs volume	7350
K	number of ensemble members	27
L	number of rows left in $\mathbf{S}^T\mathbf{C}^{\circ q}$ after truncation	126
$J = (K-1)L$	number of non-zero eigenvalues of $\mathbf{P}^f_{\mathrm{CALECO}}$	3276
KL	number of members in CALECO ensemble	3402

Table 2. Summary of primary computational costs associated with the preparation of CALECO forecast error covariance matrix for use in the LETKF. The estimated order of magnitude of the number of operations is based on the parameter values given in Table 1. The 2nd and 3rd columns give the costs associated with the brute force and factorized forms of the CALECO LETKF, respectively

Task	Operations: brute force	Operations: factorized
$(\mathbf{P}^f_{\mathrm{CALECO}})^{1/2}$	$2n^3/3 \sim 10^{12}$	$nLK \cong nJ \sim 10^7$
Evaluate $\mathbf{S}^T\mathbf{C}^{\odot q}$	$n^2[\log(n_x)+\log(n_y)+\log(n_z)] \sim 10^9$	$nn^t\ [\ n_x \log(n_x) + n_y \log(n_y) + n_y \log(n_y)] \sim 10^7$
$\mathbf{C}^{\odot q}$	$n^2K \sim 10^{10}$	$nn^t(n_x+n_y+n_z)K \sim 10^7$

that maps physical space state vectors into spectral space and then rescales the spectral coefficients such that small scale modes are suppressed (see Part 1 for details). Producing $\mathbf{C}^{\odot q}$ costs order n^2K $[O(n^2K)]$ operations. The cost of $\mathbf{S}^T\mathbf{C}^{\odot q}$ depends on the type of transform employed. If Fourier transforms are used in each spatial direction, operating $\mathbf{S}^T$ on one column of $\mathbf{C}^{\odot q}$ costs $n_t n_v[n_z n_y n_x \log(n_x) + n_z n_x n_y \log(n_y) + n_x n_y n_z \log(n_z)]$ operations. Hence, $\mathbf{S}^T\mathbf{C}^{\odot q}$ costs order $n^2[\log(n_x)+\log(n_y)+\log(n_z)]$ operations. Assuming that $\mathbf{S}^T\mathbf{C}^{\odot q}$ was truncated to L rows then formation of $\widetilde{\mathbf{C}^{\odot q}\mathbf{S}}\widetilde{\mathbf{S}^T\mathbf{C}^{\odot q}}$ is order n^2L operations. Finally, the element-wise product $\mathbf{P}^f_K \odot \widetilde{\mathbf{C}^{\odot q}\mathbf{S}}\widetilde{\mathbf{S}^T\mathbf{C}^{\odot q}}$ costs n^2 operations.

(P3) Essentially, the LETKF uses Bishop et al.'s (2001) Ensemble Transform Kalman Filter (ETKF) to update the state and error covariance estimate within each of its observation volumes. Ideally, the ETKF is applied to a concise left square root $\mathbf{Z}$ of $\mathbf{P}^f_{\mathrm{CALECO}}$ that has about the same number of columns as the effective rank of $\mathbf{P}^f_{\mathrm{CALECO}}$. The brute force eigenvalue decomposition $\mathbf{P}^f_{\mathrm{CALECO}} = \mathbf{V}\,\Lambda\mathbf{V}^T$ where $\mathbf{V}$ contains $\mathbf{P}^f_{\mathrm{CALECO}}$'s eigenvectors and the diagonal matrix Λcontains its eigenvalue provides the concise square root $\mathbf{Z} = \mathbf{V}_J\ \Lambda_J^{1/2}$ where the $J \times J$ diagonal matrix $\Lambda_J^{1/2}$ contains the J significant eigenvalues of $\mathbf{P}^f_{\mathrm{CALECO}}$ and where the $n \times J$ matrix $\mathbf{V}_J$ contains the corresponding eigenvectors. The cost of obtaining this concise square root is $O(2n^3/3)$ operations (Press et al., 2007). Table 2, above, outlines the primary costs of preparatory steps P1–P3.

Having obtained a concise square root $\mathbf{Z}$ of $\mathbf{P}^f_{\mathrm{CALECO}}$, one can treat the columns of the square root as pseudo-ensemble members or CALECO ensemble members. We refer to DA methods based on the application of the LETKF to these CALECO ensemble members as the CALECO LETKF. The costs for the CALECO LETKF are as follows.

(E1) The first step of Bishop et al.'s (2001) ETKF is to compute a $p \times J$ matrix $\mathbf{HZ}$ whose columns give the ensemble perturbations in observation space. The jth column $(\mathbf{HZ})_j$ of $\mathbf{HZ}$ is given by $(\mathbf{HZ})_j = H\left(\bar{\mathbf{x}} + \mathbf{z}_j\right) - \frac{1}{J}\sum_{i=1}^{J} H\left(\bar{\mathbf{x}} + \mathbf{z}_i\right)$ where $\bar{\mathbf{x}}$ is the ensemble mean and where H is the (possibly) non-linear observation operator such that $\mathbf{y}^t = H(\mathbf{x}^t)$ where $\mathbf{x}^t$ is the true value of the state and $\mathbf{y}^t$ is the true value of the observation. The most computationally expensive mappings are generally associated with observations that represent weighted vertical integrals of the state such as satellite observations. The cost of mapping the state to an observation that represents a vertical integral of the state at some random horizontal location is around $10n_z n_v$ operations. Hence, if all p observations were of this expensive type, the cost of forming the column $(\mathbf{HZ})_j$ would be about $10n_z n_v p$ while the cost of forming $\mathbf{HZ}$ in its entirety would be around $10n_z n_v pJ$.

(E2) The evaluation of the $J \times J$ matrix $\mathbf{Z}^T\ \mathbf{H}^T\ \mathbf{R}^{-1}\ \mathbf{HZ}$ required by eq (11) of Bishop et al.'s (2001) description of the ETKF is $O(J^2\ p/2)$ operations where $\mathbf{R}$ gives the observation error covariance matrix for the observations. The eigenvalue decomposition of this matrix $\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HZ} = \mathbf{C}\Gamma\mathbf{C}^T$($\mathbf{C}$ lists the eigenvectors, diagonal matrix Γ gives the corresponding eigenvalues) also required by eq. (11) of Bishop et al. (2001) costs about $O(2J^3/3)$ operations. For observation volumes that fill the entire depth of the atmosphere and a dense observation network, these two steps are usually the most expensive parts of the ETKF algorithm. Both of them could be avoided if a variational approach were taken to updating the ensemble mean and a global ensemble update along the lines of Wang and Bishop (2003), Wang et al. (2004) or McLay et al. (2008) were used. Details of a variational approach to updating the ensemble mean and the corresponding costs are given in the Appendix S1 in supplementary material. With the variational approach, the cost of each iteration of the minimization algorithm is order $2Jp$. While this approach might be attractive for extremely large observation volumes (very large p and J) we will not pursue it here.

(E3) Compute $\bar{\mathbf{x}}^a_u$ the mean of the corrected or posterior ensemble members for the $n_u n_v$ variables near the centre of the observation volume using $\bar{\mathbf{x}}^a_u = \bar{\mathbf{x}}^f_u + \mathbf{Z}^f_u\mathbf{C}(\Gamma+\mathbf{I})^{-1}\mathbf{C}^T\mathbf{Z}^{fT}\mathbf{H}^T\mathbf{R}^{-1/2}\left(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f\right)$ where $\bar{\mathbf{x}}^f_u$ is the mean of the prior or forecast ensemble members at the $n_u n_v$ variables, $\mathbf{Z}^f_u$ is the $(n_u n_v)xK$ matrix obtained by removing all of the rows of the $n \times J$ matrix $\mathbf{Z}$ except those that pertain to the $n_u n_v$ variables that are to be updated, $\mathbf{y}$ is the p-vector of observations within the observation volume. The cost of this operation is about $Jp + J^2 + J + J^2 + n_u n_v J \approx J(p + 2J + n_u n_v)$ operations.

(E4) Create K ensemble perturbations $\mathbf{x}^a_i$ whose covariance estimates the error covariance of the $n_u n_v$ analysed variables. There are at least three options available for this step. The first option involves creating an $(n_u n_v)xK$ matrix of posterior ensemble perturbations $\mathbf{X}^a_u$ using $\mathbf{X}^a_u = \mathbf{Z}^a_u\mathbf{A}_{JxK}$ where $\mathbf{Z}^a_u$ is obtained

Table 3. Summary of primary computational costs for the ETKF steps in each observation volume with J CALECO ensemble members, p observations and n_u update variables. The estimated order of magnitude of the number of operations is based on the parameter values given in Table 1

Task	Operations
$\mathbf{HZ}$	$10n_z n_v pJ \sim 10^{10}$
$\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HZ}$	$J^2p/2 \sim 10^{11}$
$\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HZ} = \mathbf{C}\Gamma\mathbf{C}^T$	$2J^3/3 \sim 10^{10}$
Compute $\bar{\mathbf{x}}_u^a$	$J(p + 2J + n_u n_v) \sim 5x10^7$
$\mathbf{x}_i^a, i = 1, \ldots, K$	$KJ(p + 2J + n_u n_v) \sim 10^9$

Table 4. Summary of primary computational costs associated with a variational approach to computing the update within each observation volume with J ensemble members, p observations and n_u update variables. The estimated order of magnitude of the number of operations is based on the parameter values given in Table 1

Task	Operations
$\mathbf{HZ}$	$10n_z n_v pJ \sim 10^{10}$
Compute $\bar{\mathbf{x}}_u^a$	$100Jp \sim 10^{10}$
$\mathbf{x}_i^a, i = 1, \ldots, K$	$K100Jp \sim 10^{11}$

using the precise square root solution $\mathbf{Z}_u^a = \mathbf{Z}_u^f\mathbf{C}(\Gamma + \mathbf{I})^{-1/2}\mathbf{C}^T$ of the equation relating analysis error covariance to forecast error covariance for an optimal DA scheme (Bishop et al., 2001) at a cost of about $n_v n_u(J^2 + J + J^2) \approx 2n_v n_u J^2$operations. Each column $\mathbf{a}_i$ of the K columns of the $J \times K$ matrix $\mathbf{A}_{JxK}$ is a random vector with mean zero and covariance matrix equal to the identity matrix $\mathbf{a}_i \sim N(\mathbf{0}, \mathbf{I})$. The operation $\mathbf{X}_u^a = \mathbf{Z}_u^a\mathbf{A}_{JxK}$ costs about $n_v n_u JK$ operations. In order to facilitate similarity between the analysis perturbations of one volume and the next, one could let $\mathbf{A}_{JxK}$ be identical for all volumes. The second option uses perturbed observations $\mathbf{y}_i = \mathbf{y} + \varepsilon_i^o$ and the equation $\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{Z}_u^f\mathbf{C}(\Gamma + \mathbf{I})^{-1}\mathbf{C}^T\mathbf{Z}_u^{fT}\mathbf{H}^T\mathbf{R}^{-1/2}(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f)$, $i = 1, 2, \ldots, K$ where ε_i^o is a random normal variable with covariance $\mathbf{R}$ and $\mathbf{x}_i^f$ gives the ith member of the ensemble forecast. This Monte-Carlo approach would incur costs around $KJ(p + 2J + n_u n_v)$ operations. Table 3 gives the cost of this approach for the parameter values listed in Table 1. For these parameters, this second option costs about half as much as the first option. In addition to this computational advantage, in practice, the Monte-Carlo approach may be more accurate because it does not rely on the assumption of an optimal DA scheme. The third option is to update the ensemble using a global ensemble update algorithm such as those suggested in (Wang and Bishop, 2003; Wang et al., 2004; McLay et al., 2008) that do not use localization.

(E5) Create the K member posterior ensemble by adding the perturbations $\mathbf{X}_u^a$ to the updated ensemble $\bar{\mathbf{x}}_u^a$ at a cost of order Kn_u operations. After this step, the posterior ensemble is ready for the next forecast/assimilation cycle.

Comparison of the costs (Table 2) of the brute force CALECO LETKF with Table 3 shows that the costs of preparing CALECO for use in an LETKF observation volume using the brute force method exceeds the costs associated with the ETKF steps for the parameter values given in Table 1. In the next section, we introduce methods that make the costs of these preparatory steps orders of magnitude cheaper than the ETKF costs.

3. Reducing the costs of preparing CALECO for the LETKF

3.1. A factorization property for the CALECO matrix

Table 2 shows that the most expensive CALECO preparation cost is that of finding a concise square root of CALECO via the decomposition $\mathbf{P}_{\text{CALECO}}^f = \mathbf{V}\Lambda\mathbf{V}^T$. Let $\mathbf{W} = \widetilde{\mathbf{C}^{\odot q}\mathbf{S}}$ so that $\mathbf{C}_{\text{ECO-RAP}} = \mathbf{W}\mathbf{W}^T$. Let $\mathbf{Z}_K = [\mathbf{x}_1 - \bar{\mathbf{x}}, \mathbf{x}_2 - \bar{\mathbf{x}}, \ldots, \mathbf{x}_K - \bar{\mathbf{x}}]\big/\sqrt{K-1}$ give the square root of the sample covariance matrix $\mathbf{P}_K^f$ of the raw K-member ensemble ($\mathbf{P}_K^f = \mathbf{Z}_K\mathbf{Z}_K^T$). Let $\mathbf{z}_i$ and $\mathbf{w}_j$ denote the ith and jth columns of $\mathbf{Z}_K$ and $\mathbf{W}$, respectively. Let z_{ki} and w_{kj} denote the kth row element of the ith and jth columns of $\mathbf{Z}_K$ and $\mathbf{W}$, respectively. Now note that the element $[\sum_{i=1}^K\sum_{j=1}^L(\mathbf{z}_i \odot \mathbf{w}_j)(\mathbf{z}_i \odot \mathbf{w}_j)^T]_{kl}$on the kth row and lth column of the matrix $\sum_{i=1}^K\sum_{j=1}^L(\mathbf{z}_i \odot \mathbf{w}_j)(\mathbf{z}_i \odot \mathbf{w}_j)^T$ is given by

$$
\begin{aligned}
\left[\sum_{i=1}^K\sum_{j=1}^L\left(\mathbf{z}_i \odot \mathbf{w}_j\right)\left(\mathbf{z}_i \odot \mathbf{w}_j\right)^T\right]_{kl} &= \sum_{i=1}^K\sum_{j=1}^L z_{ki}w_{kj}z_{li}w_{lj} \\
&= \sum_{i=1}^K\left(\sum_{j=1}^L w_{kj}w_{lj}\right)z_{ki}z_{li} \\
&= \left(\sum_{j=1}^L w_{kj}w_{lj}\right)\left(\sum_{i=1}^K z_{ki}z_{li}\right) = \left(\mathbf{W}\mathbf{W}^T\right)_{kl}\left(\mathbf{Z}_K\mathbf{Z}_K^T\right)_{kl},
\end{aligned}
\tag{1}
$$

where the symbol $\odot$ indicates the element-wise matrix product that is often called the Schur or Hadamard product. Equation (1) implies the factorization property

$$
\begin{aligned}
\mathbf{Z}_K\mathbf{Z}_K^T \odot \mathbf{W}\mathbf{W}^T &= \sum_{i=1}^K\sum_{j=1}^L\left(\mathbf{z}_i \odot \mathbf{w}_j\right)\left(\mathbf{z}_i \odot \mathbf{w}_j\right)^T \\
&= (\mathbf{Z}_K \vartriangle \mathbf{W})(\mathbf{Z}_K \vartriangle \mathbf{W})^T \\
&= \mathbf{P}_K^f \odot \mathbf{C}_{\text{ECORAP}} = \mathbf{P}_{\text{CALECO}}^f, \text{ where} \\
\mathbf{Z}_K \vartriangle \mathbf{W} &= [[\mathbf{z}_1 \odot \mathbf{w}_1, \mathbf{z}_1 \odot \mathbf{w}_2, \ldots, \mathbf{z}_1 \odot \mathbf{w}_L], \\
&[\mathbf{z}_2 \odot \mathbf{w}_1, \ldots, \mathbf{z}_2 \odot \mathbf{w}_L], \ldots, [\mathbf{z}_K \odot \mathbf{w}_1, \ldots, \mathbf{z}_K \odot \mathbf{w}_L]].
\end{aligned}
\tag{2}
$$

Each of the KL columns of $\mathbf{Z}_K \vartriangle \mathbf{W}$ is given by the element-wise product of a column of $\mathbf{Z}_K$ with a column of $\mathbf{W}$. In the same way that a musician can modulate sound by varying its amplitude (volume), phase (timing) or frequency (pitch), we think of the element-wise products $\mathbf{z}_i \odot \mathbf{w}_j$ as causing $\mathbf{w}_j$ to modulate $\mathbf{z}_i$. For this reason, we refer to $\mathbf{Z}_K \vartriangle \mathbf{W}$ as the modulation matrix product between $\mathbf{Z}_K$ and $\mathbf{W}$. Equations (1) and (2) show that $\mathbf{Z}_K \vartriangle \mathbf{W}$ is a concise square root $\mathbf{Z}$ (having only KL columns) of $\mathbf{P}_{\text{CALECO}}^f$ and hence we can set $\mathbf{Z} = \mathbf{Z}_K \triangle \mathbf{W}$. The cost of computing the $KL \cong J$ columns of $\mathbf{Z}$ from $\mathbf{Z}_K$ and $\mathbf{W}$ is order $KLn \cong Jn$ operations. For the observation volume parameters listed in Table 1, the last column of Table 2 shows that this cost is five orders of magnitude less than the cost of finding the square root of $\mathbf{P}^f_{\mathrm{CALECO}}$ using an eigenvalue decomposition and that it is insignificant compared with the ETKF costs given in Table 3.

Judicious use of the factorization property (2) greatly reduces the memory requirements of a CALECO ETKF. Without this property, one would have to store in memory all nJ numbers defining the $J \cong KL$ member CALECO ensemble. However, eq. (2) shows how any member of the CALECO ensemble can be computed as needed from the element-wise product of one of the columns of $\mathbf{Z}_K$ with one of the columns of $\mathbf{W}$. Consequently, one only needs to hold $\mathbf{Z}_K$ and $\mathbf{W}$ in memory. Only $n(K + L)$ numbers need to be held in memory to precisely define $\mathbf{Z}_K$ and $\mathbf{W}$. For the observation volume parameter values given in Table 1, $n(K + L) = 2 \times 10^6$ whereas $nJ = nKL = 4.5 \times 10^7$ and hence (2) enables an order of magnitude reduction in memory storage requirements.

3.2. *Approximate* $\mathbf{C}^{\odot q}$ *using separable functions*

Table 2 shows that the evaluation of $\mathbf{C}^{\odot q}$ is the second most expensive step in the brute force approach to preparing CALECO for use in the LETKF. The jth column of the matrix $\mathbf{C}^{\odot q}$ lists the correlations of all the variables with the jth variable. If there are n_t model output times in the assimilation window and m variables at each time, then the jth column $\mathbf{c}_j$ of $\mathbf{C}^{\odot q}$ represents the concatenation of n_t m-vectors $\mathbf{c}_{\tau j}$, $\tau = 1, 2, \ldots, n_t$ each of which lists correlations with all of the variables at the τ time level with the jth variable; in other words, $\mathbf{c}_j^T = [\mathbf{c}_{1j}^T, \mathbf{c}_{2j}^T, \ldots, \mathbf{c}_{n_t j}^T]$. The cost of computing this vector can be greatly reduced by assuming that all of the elements in the m-vector $\mathbf{c}_{\tau j}$ can be approximated by a separable function of x, y and z. We found that the following recipe produced useful separable approximations to $\mathbf{c}_{\tau j}$, $\tau = 1, 2, \ldots, n_t$ and its elementwise powers.

(i) Find the maximum $c_{\tau\,\max}$ of the vector $\mathbf{c}_{\tau j}$. If τ indicates the same time level as that of the jth variable then the maximum will be at the jth variable. At other time levels, the maximum will have propagated some distance away from the physical location of the jth variable. Hence, information on the propagation speed of significant correlations such as its maximum possible magnitude enables an efficient search algorithm for the maximum at other times.

(ii) Compute the n_x correlations along the line parallel to the x-axis that intersects the maximum value and let these values define the function $f_\tau(i_x)$, $i_x = 1, 2, \ldots, n_x$.

(iii) Compute the n_y correlations along the line parallel to the y-axis that intersects the maximum value and let these values define the function $g^*_\tau(i_y)$, $i_y = 1, \ldots, n_y$ then normalize this function by the maximum correlation $c_{\tau\,\max} = \max[g^*_\tau(i_y)]$ to obtain $g_\tau(i_y) = g^*_\tau(i_y)/c_{\tau\,\max}$, $i_y = 1, \ldots, n_y$.

(iv) Repeat step (iii) but this time for a line parallel to the z-axis thus obtaining the normalized correlation function $h_\tau(i_z)$, $i_z = 1, \ldots, n_z$.

(v) Approximate the element of $\mathbf{c}_{\tau j}$ corresponding to the (i_x, i_y, i_z) grid point by $f_\tau(i_x) g_\tau\left(i_y\right) h_\tau\left(i_z\right)$.

(vi) Approximate the qth power of the elements of $\mathbf{c}_{\tau j}$ using $f^q_\tau(i_x) g^q_\tau\left(i_y\right) h^q_\tau\left(i_z\right)$

(vii) Repeat steps (i)–(vi) for all τ values ($\tau = 1, 2, \ldots, n_t$.)

Thus, using steps (i)–(vii), the qth Hadamard power of the jth column $\mathbf{c}_j^{\odot q}$ can be approximated by adaptive separable functions for $\tau = 1, n_t$. The computational expense of this approximation to $\mathbf{c}_j^{\odot q}$ is the relatively small cost of identifying the maximum value of $\mathbf{c}_{\tau j}$ at each time plus about $Kn_t(n_x + n_y + n_z)$ operations required to create the separable functions defining the approximation $\mathbf{c}_j^{\odot q}$. For large observation volumes, this cost is many orders of magnitude less than the roughly Kn operations that would be required to evaluate all the correlations that precisely define the jth column $\mathbf{c}_j^{\odot q}$.

The matrix obtained by approximating each column of $\mathbf{C}^{\odot q}$ by a separable function in this manner will be denoted by $\hat{\mathbf{C}}^{\odot q}$. Since there are n columns in $\hat{\mathbf{C}}^{\odot q}$, the cost of defining all these columns is about $nn_t(n_x + n_y + n_z)K$. Table 2 shows that for the parameter settings of Table 1, the cost of computing $\hat{\mathbf{C}}^{\odot q}$ is 2 orders of magnitude smaller than the cost of computing $\mathbf{C}^{\odot q}$. In addition, while n^2 numbers are needed to define $\mathbf{C}^{\odot q}$ only $nn_t(n_x + n_y + n_z)$ numbers are needed to define $\hat{\mathbf{C}}^{\odot q}$. For the parameters given in Table 1, $n^2 = 2 \times 10^8$ while $nn_t(n_x + n_y + n_z) = 2 \times 10^6$. Hence, $\hat{\mathbf{C}}^{\odot q}$ requires 2 orders of magnitude less computer memory than $\mathbf{C}^{\odot q}$.

Replacing $\mathbf{S}^T\mathbf{C}^{\odot q}$ by the separable form $\mathbf{S}^T\hat{\mathbf{C}}^{\odot q}$ greatly reduces the cost of estimating $\mathbf{S}^T\mathbf{C}^{\odot q}$. This is because the transform of a separable function $f^q_\tau(i_x) g^q_\tau\left(i_y\right) h^q_\tau\left(i_z\right)$ is just $F_\tau(\nu_k) G_\tau(\nu_l) H_\tau(\nu_m)$ where $F_\tau(\nu_k)$, $G_\tau(\nu_l)$ and $H_\tau(\nu_m)$ are the spectral transforms of $f^q_\tau(i_x)$, $g^q_\tau\left(i_y\right)$ and $h^q_\tau\left(i_z\right)$, respectively. Thus, the separability assumption reduces the cost of taking the transform of $\mathbf{c}_j$ from order $n^2[\log(n_x) + \log(n_y) + \log(n_z)]$ operations to $nn_t[n_x \ln(n_x) + n_y \ln(n_y) + n_z \ln(n_z)]$ operations. Table 2 shows that for the parameters given in Table 1, this approach reduces the cost of estimating $\mathbf{S}^T\mathbf{C}^{\odot q}$ by two orders of magnitude.

Note that while the columns of $\mathbf{C}^{\odot q}$ are replaced by separable functions, the rows of $\mathbf{C}^{\odot q}$ are not. Hence, the columns of $\mathbf{C}_{\mathrm{ECO-RAP}} = (\widehat{\mathbf{C}^{\odot q}})^T\mathbf{S}\mathbf{S}^T\hat{\mathbf{C}}^{\odot q}$ will not, in general, be describable by separable functions. This non-separability is evident in Figs. 2–6, which display various columns of $\mathbf{C}_{\mathrm{ECO-RAP}}$.

3.3. *Use univariate ECORAP localization rather than multivariate localization*

In principle ECO-RAP provides separate functions for the localization of covariances between distinct variables. However, Ensemble Kalman Filters that use the same non-adaptive localization function for all ensemble covariances have been shown to work reasonably well (e.g. Houtekamer et al., 2005). This suggests that computing the adaptive ECO-RAP localization matrix for the univariate ensemble covariances associated with a single variable type and then applying it to all ensemble correlations (including multivariate correlations) would also work reasonably well.

For the experiments to be reported on in this paper, we found that ECO-RAP localization functions based on equivalent potential temperature θ^e gave reasonable looking localization functions and it is these functions that were used for localization in this paper. To implement this univariate localization within the computationally efficient framework provided by (2), we alter the way we compute the correlations from which the square root $\mathbf{W}$ is derived. To be specific, we (a) replace all univariate correlations between wind, temperature and humidity variables by the corresponding univariate correlations of θ^e (b) we replace all multivariate correlations by the corresponding univariate correlations from θ^e. This is equivalent to replacing multivariate perturbation state vectors by univariate perturbation state vectors in which the perturbation variables associated with the wind, temperature and humidity variables have been overwritten by θ^e variables. When this is done, it can be shown that each column of the $\mathbf{W}$ matrix is composed of concatenated identical vectors each of which would be identical to the corresponding column of the $\mathbf{W}$ matrix one would obtain if one was only attempting a univariate analysis of θ^e. This fact means that this univariate approximation allows further memory and computational savings.

A rigorous comparison of the quality of univariate ECO-RAP localization functions derived from, say, streamfunction, with those from θ^e is beyond the scope of this paper. This univariate approximation reduces the costs listed in the third column of Table 2 by the factor $(1/n_v)$ where n_v is the number of variables at each grid point.

4. Costs of CALECO LETKF versus standard LETKF

Assuming that there are more observations than ensemble members in each observation volume, the dominating cost of both the CALECO LETKF and the standard ETKF is the computation of $\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HZ}$. The cost is about $J^2\ p/2 = L^2K^2p/2$ operations for the CALECO LETKF and about $K^2p_s/2$ operations for the standard LETKF, where p_s gives the typical number of observations in a standard LETKF observation volume. While the standard LETKF updates variables at the single grid point that lies far enough away from the edges of its small observation volume, the CALECO LETKF can simultaneously update the many n_u grid points that lie sufficiently far from the edges of its larger observation volume with little additional computational cost.

To see the additional cost, recall that (following Bishop et al., 2001) the ETKF analysis update equation is $\bar{\mathbf{x}}_u^a = \bar{\mathbf{x}}_u^f + \mathbf{Z}_u^f\mathbf{C}(\Gamma+\mathbf{I})^{-1}\mathbf{C}^T\mathbf{Z}^{fT}\mathbf{H}^T\mathbf{R}^{-1/2}\left(\mathbf{y}-\mathbf{H}\bar{\mathbf{x}}^f\right)$ and that the cost of this operation is order $Jp + J^2 + J + J^2 + n_un_vJ$ operations where the number of grid points to be simultaneously updated is n_u. Hence, provided n_un_vJ is much smaller than $J^2p/2$ (the cost of $\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HZ}$), the cost of updating n_u grid points will be about the same as updating a single grid point. If we insist that the cost n_un_vJ of updating n_u grid points be no greater than 10% of the cost $J^2p/2$ of computing $\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HZ}$, we arrive at the requirement that $n_u \leq 0.1Jp/(2n_v)$. In the experiment of Section 5, $J = 3402$, $p = 7350$ and $n_v = 3$. From these numbers we can deduce that the simultaneous update of $n_u = 0.1Jp/(2n_v) = 340.2 \times 7350/6 = 4.2 \times 10^5$ grid points would only be 10% more than the cost of updating a single grid point. Thus, if there where 42 grid points in each vertical column, we could update a 100×100 square of vertical columns and only incur a 10% increase in cost.

Recall that the standard LETKF can only update one grid point per observation volume because spurious correlations associated with its K member ensemble are suppressed either by (1) letting observation error variance (artificially) increase with distance from the grid point to be updated (2) shrinking the observation volume until the effect of spurious correlations is minimized. In contrast, the $J = KL$ member ensemble of the CALECO LETKF is largely free of spurious correlations and the only role of the observation volume is to break the data assimilation problem into manageable computational chunks. Hence, since there is little additional computational cost associated with simultaneously updating large numbers of grid points, one can choose much larger observation volumes for the LETKF and allow the scheme to simultaneously update all of the grid points that lie far enough away from the edges of the observation volume. Here, we explore the conditions under which the higher cost of the CALECO LETKF ($L^2K^2p > K^2p_s$) for a single observation volume could be offset by the fact that a large number of grid points can be simultaneously updated in the CALECO LETKF whereas only a single grid point can be updated in the standard LETKF.

Let us suppose that the parameters m_x, m_y and m_z define the number of grid points in the *x, y* or *z* directions one must go in order for ensemble correlations to be ignored without negatively impacting the DA scheme. In this case, the ideal standard LETKF observation volume would have $(2m_x+1)\left(2m_y+1\right)(2m_z+1)$ grid points and the corresponding average number of observations p_s in the standard LETKF volume would be given by $p_s = \rho_o(2m_x+1)\left(2m_y+1\right)(2m_z+1)$ where the observation density ρ_o gives the average number of observations per LETKF observation volume.

Let the CALECO LETKF observation volumes cover the entire n_z vertical levels of the model and let their x and y horizontal boundaries be respectively $m_x + n_e$ and $m_y + n_e$ grid points from the centre of the volume such that each volume contains $[2(m_x+n_e)+1]\left[2(m_y+n_e)+1\right]n_z$ grid points. These larger observation volumes provide $n_u = (2n_e+1)^2\ n_z$ grid points that are sufficiently distant from the boundaries to be simultaneously updated using the $p = \rho_o[2(m_x + n_e) + 1]\left[2(m_y + n_e) + 1\right] n_z$ observations within the volume. The standard LETKF cannot update these n_u grid points simultaneously. Hence, it costs the standard LETKF roughly ($n_u K^2\, p_s$ operations) to update these grid points.

In order for the cost of the CALECO LETKF to be less than or equal to the standard LETKF, the cost ($n_u K^2 p_s$ operations) of updating the n_u grid points with the standard LETKF must be greater than or equal to the cost ($L^2 K^2 p$ operations) of updating the n_u grid points using the CALECO LETKF. In other words, the condition

$$n_u K^2 p_s \geq L^2 K^2 p \tag{3}$$

must be satisfied. Substituting for p_s, p and n_u in (3) using the above relations gives

$$\begin{aligned}&(2n_e + 1)^2\, n_z\, (2m_x + 1)\left(2m_y + 1\right)(2m_z + 1)\\ &\quad \geq L^2\, [2(m_x + n_e) + 1]\left[2(m_y + n_e) + 1\right] n_z.\end{aligned} \tag{4}$$

The significance of (4) depends on the relationship between L, the rank of the ECO-RAP localization matrix and $2n_e$, the grid point length by which the sides of the CALECO LETKF observation volume exceeds that of the standard LETKF observation volume.

To infer this relation, we recall that a keystone of the standard LETKF (Ott et al., 2004; Patil et al., 2001) is the assumption that the dimension of the error space within the standard LETKF observation volume is small enough to be spanned, in large measure, by K *raw* ensemble perturbations. Here we shall assume that aK CALECO ensemble perturbations are sufficient to analyse the error at the centre of each independent $(2m_x + 1)\left(2m_y + 1\right)(2m_z + 1)$ standard LETKF observation volume where a is an unknown constant. If Ott et al.'s assumption applies to members of the CALECO ensemble then a will be near unity (between 0.5 and 5, say). This suggests that the minimum size J of the CALECO ensemble required for a CALECO observation volume with $[2(m_x + n_e) + 1]\left[2(m_y + n_e) + 1\right] n_z$ grid points is aK multiplied by the ratio of the CALECO observation volume divided by the LETKF observation volume. Hence,

$$\begin{aligned}J = LK = &\left[\frac{aK}{(2m_x + 1)\left(2m_y + 1\right)(2m_z + 1)}\right]\\ &[2(m_x + n_e) + 1]\left[2(m_y + n_e) + 1\right] n_z.\end{aligned} \tag{5}$$

Consequently,

$$L = a\frac{[2(m_x + n_e) + 1]\left[2(m_y + n_e) + 1\right] n_z}{(2m_x + 1)\left(2m_y + 1\right)(2m_z + 1)}. \tag{6}$$

Using (6) in (4) gives

$$(2n_e + 1)^2\, n_z \geq a^2 \left\{\frac{[2(m_x + n_e) + 1]\left[2(m_y + n_e) + 1\right] n_z}{(2m_x + 1)\left(2m_y + 1\right)(2m_z + 1)}\right\}^3 \tag{7}$$

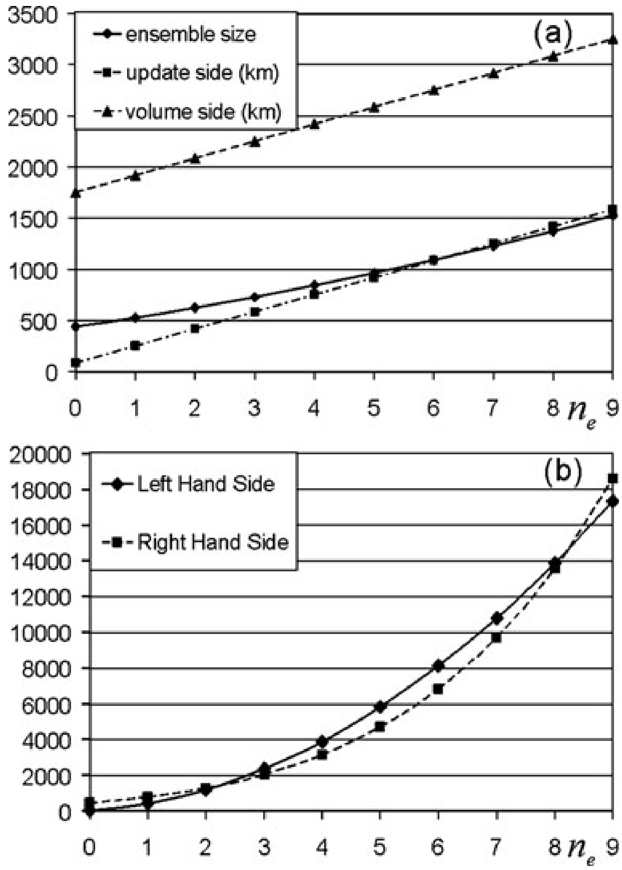

Fig. 1. In (a) the solid line shows the CALECO ensemble size as a function of n_e. [Recall that the number of grid points to be simultaneously updated by the CALECO LETKF is $n_u = (2n_e + 1)^2\, n_z$]. The dashed line (triangles) and the dash–dotted line (squares) give the longest edge of the rectangles defined by the entire CALECO observation volume and the region that is simultaneously updated by the CALECO LETKF, respectively. In (b) the solid and dashed lines give, as a function of n_e, the value of the left and right hand sides of eq. (7), respectively.

as the condition for the CALECO LETKF to be more efficient than the standard LETKF.

For standard LETKF experiments with real and simulated observations Miyoshi and Yamane's (2007, p. 3855) work used $m_x = m_y = 10$, $m_z = 6$, $K = 40$ and $n_z = 48$. Miyoshi and Yamane's (2007) horizontal grid spacing was 0.75 degrees (~83.4 km along a longitude line). Hence, the longest side of their observation volumes was about 1750 km. To associate some precise numbers with (6) and (7), we set $a = 3$ and assume the aforementioned parameter values of Miyoshi and Yamane[1] . For

[1]The grid point based definitions of observation volumes, although convenient for the current analysis, are problematic at the poles. Miyoshi et al. (2007) describe observation volumes based on distance alone that avoid this problem. It would be a simple matter to extend our analysis to these types of observation volumes.

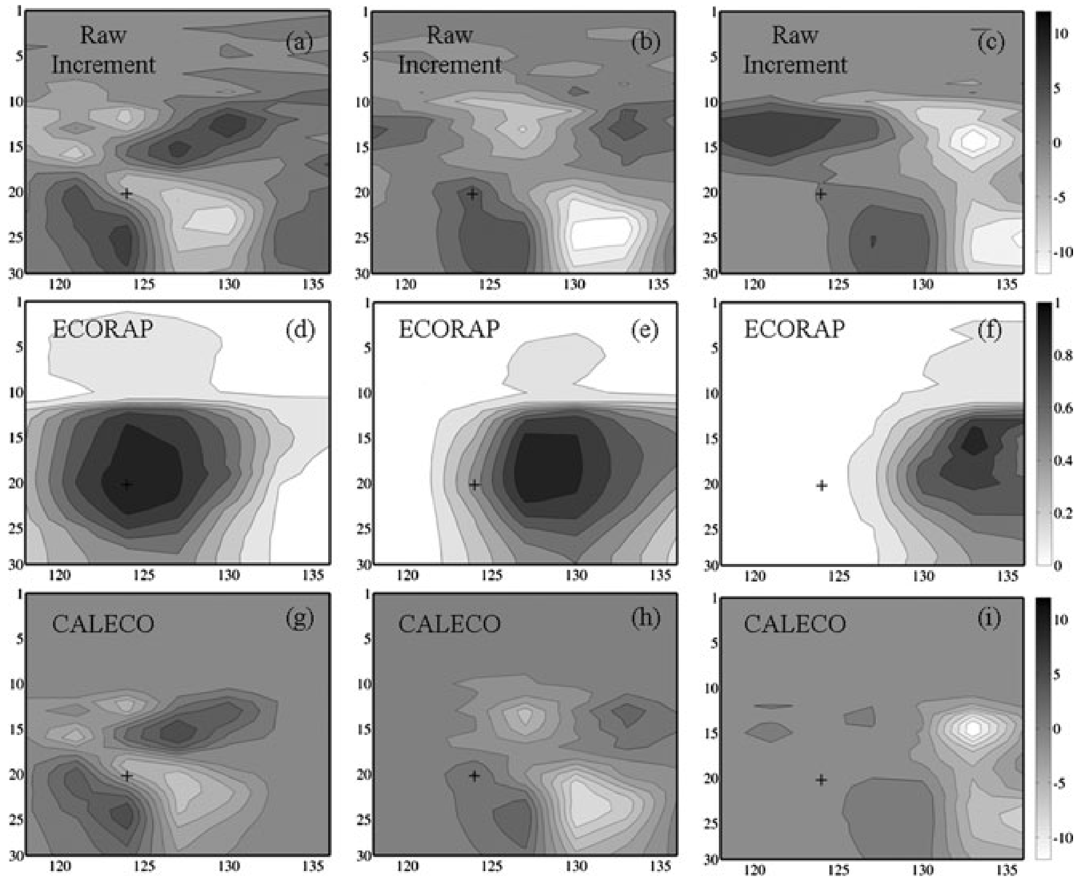

Fig. 2. In a, b and c is shown a vertical cross-section of the evolution of the unlocalized meridional wind increment from a single temperature observation at 3Z and (56S, 124E, $\sigma = 20$) denoted by the '+'; (the abscissa and ordinate give longitude and σ-level, respectively). In d, e and f is shown the corresponding vertical cross-section of the evolution of the ECORAP localization function associated with this observation. In g, h and i the corresponding CALECO increment is shown. Bars give the relation between 'shade' and 'value'. For (a)–(c) and (g)–(i) values have units of Kelvin.

these parameters, Fig. 1a plots the relationship between n_e and the size $J = KL = 40L$ of the CALECO ensemble implied by (6). As a function of n_e, it also gives the length of the longest side (in kilometres) of the rectangle of grid points that would be simultaneously updated by CALECO together with the longest side of the rectangle defining the CALECO observation volume. In Fig. 1b is plotted the corresponding left and right hand sides of eq. (7) as a function of n_e. It shows that the necessary condition (7) for the CALECO LETKF to be more efficient than the regular LETKF is satisfied $3 \leq n_e \leq 8$.

Figure 1 is based on the assumption that $a = 3$. The appropriate value for a is unknown but could be determined for specific systems by tuning CALECO performance for various ensemble sizes. Calculations using (7) show that it is satisfied for an even larger range of n_e values for $a < 3$; however, for $a > 3$ it isn't satisfied for any n_e values. Might the CALECO LETKF outperform the standard LETKF at a lower cost?

5. Illustration of CALECO LETKF in an atmospheric model

5.1. *NOGAPS, the ET ensemble, and the size of observation volumes*

Determining whether the CALECO LETKF can or cannot outperform the standard LETKF in a sophisticated model and at what cost is beyond the scope of this paper. Here, we illustrate the 4-D mobility and adaptivity of ECO-RAP localization functions, CALECO covariance functions and test their ability to reduce analysis error. The experiments performed are qualitatively similar to those described in Bishop and Hodyss (2007) except here, due to the computational efficiency of ECO-RAP, we are able to assimilate a higher number of observations per observation volume (7350 versus 1666). This higher observation density is more in line with the numbers of observations currently being assimilated in operational numerical weather prediction. Apart from the computational superiority of ECO-RAP, we were

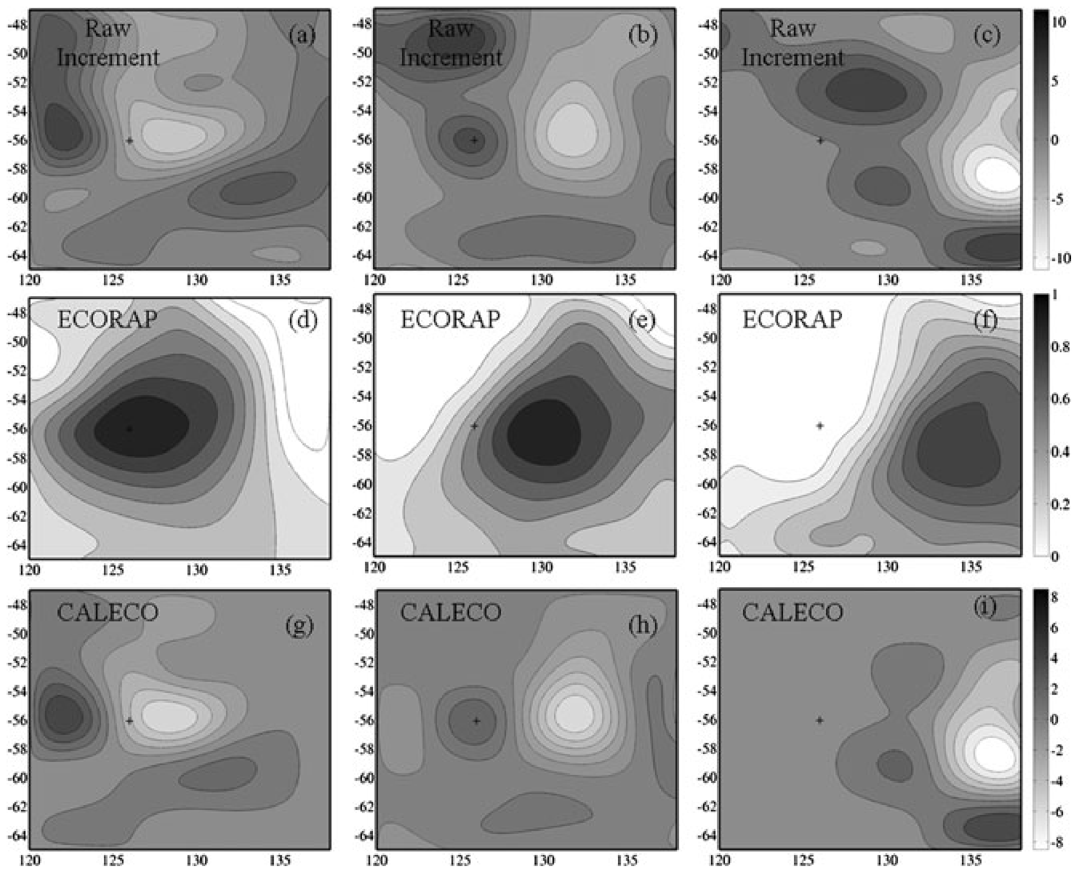

Fig. 3. As in figure 2, but for a horizontal cross-section on σ-level = 20. The abscissa gives degrees longitude, the ordinate gives degrees latitude.

unable to detect any significant performance difference between the SENCORP approach used in Bishop and Hodyss (2007) and ECO-RAP.

The forecast model used is the Navy Operational Global Atmospheric Prediction System (NOGAPS) (Hogan and Rosmond, 1991). NOGAPS is run at horizontal spectral resolution T119 with 30 vertical levels. To simplify our calculations (that were all performed on a single processor computer running Matlab), we degrade the horizontal resolution of the ensemble output to a $3^\circ \times 3^\circ$ latitude–longitude grid. This resolution is comparable to that used by Szunyogh et al. (2005, 2008). The ensemble is generated using the ensemble system developed at our laboratory and described in McLay et al. (2007, 2008). Initial perturbations are based on the Ensemble Transform (ET) technique of Bishop and Toth (1999). All results pertain to 27 ensemble members taken from 17 July 2005. Each 4-D CALECO observation volume considered in this section spans a 6 h assimilation window with output every 3 hrs, covers all 30 vertical levels of the model and a 21×21 degree area (7×7 grid point box) in the horizontal.

To generate synthetic observations for our experiment, we define "truth" to be the state of a forecast initialized at 6 UTC on 16 July 2005 at 3, 6 and 9 UTC on the 17 July 2005. In other words, we let the truth over our 6 h window be given by the last 6 h of a 27 h forecast. The 4-D first guess for our DA experiment is the 3, 6 and 9 h forecast initialized at 0 UTC on the 17 July 2005. The difference between the first guess and the truth is the combined effect of Naval research laboratory Atmospheric Variational DA System (NAVDAS) (Daley and Barker, 2001) analysis corrections at 6, 12, 18 and 24 UTC on the 16 July 2005. The approach yields a difference that is global and informed by real observations. Where needed, pseudo-observations are created by adding normally distributed noise to 'truth' with variances equal to 1 m^2s^{-2} and 2 K^2 for wind and temperature observations, respectively.

5.2. *Parameter settings for ECO-RAP*

As mentioned in subsection 2d of Part 1, the extent to which ECO-RAP localization preserves ensemble correlations through time can sometimes be increased by spatially smoothing the ensemble perturbations upon which ECO-RAP is based. Experimentation showed that this was the case for NOGAPS ensemble perturbations. We smoothed θ_e ensemble perturbations using the smoother $\mathbf{E}\Lambda_s\mathbf{E}^T$ where the orthonormal matrix $\mathbf{E}$ was a

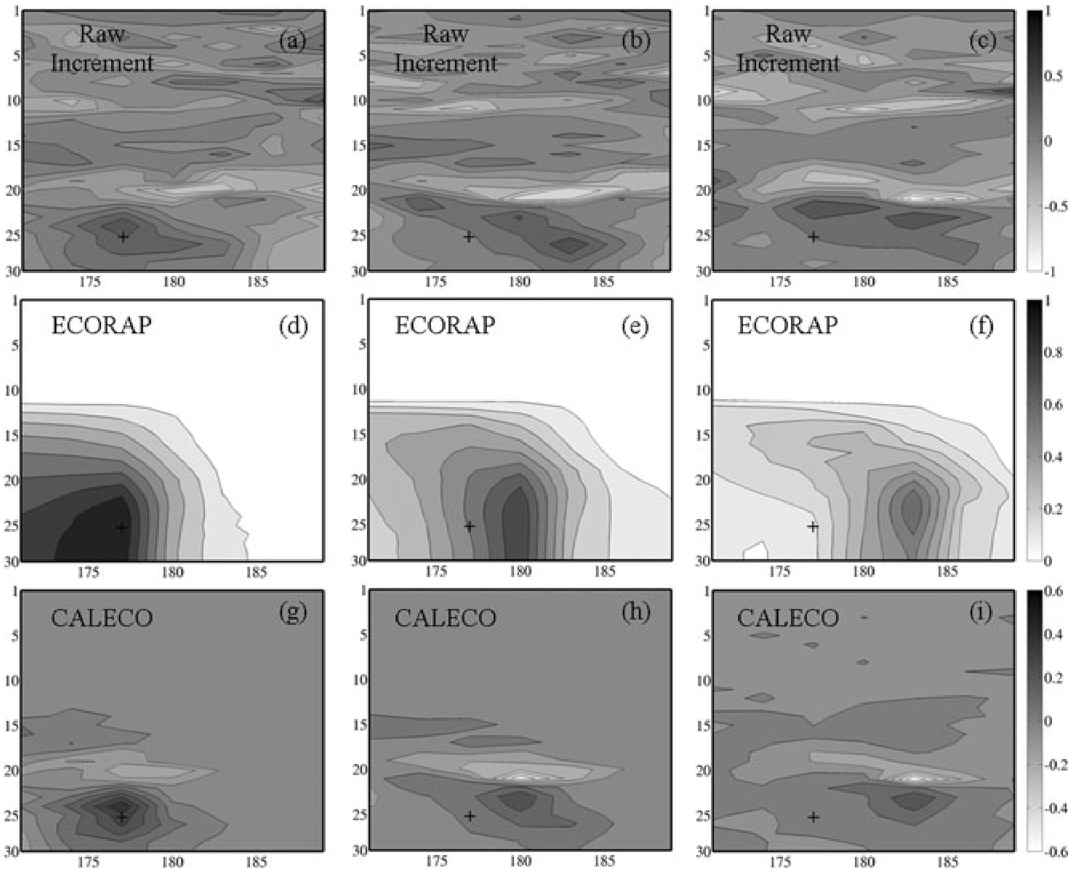

Fig. 4. As in figure 2, but for a temperature-increment from a single meridional wind observation at 3Z and (60S, 177E, $\sigma = 25$).

half-cosine basis such that the element on the kth row and jth column is

$$e_{kj} = a_j \cos\left(\frac{\pi j}{2N}\,[2k+1]\right), \tag{8}$$

where N is the number of variables, $k = 0,1, \ldots N{-}1$, and $a_1 = 1/\sqrt{N}$, $a_j = \sqrt{2/N}$ for $j \neq 1$ and the elements of the diagonal matrix Λ_s were made proportional to $\exp[-(\frac{j}{d_{h/v}})^2]$ This smoothing was accomplished using the parameters of $d_h = 2.2$ and $d_v = 3.2$ in the horizontal and vertical directions, respectively. This same spatial smoothing was performed on θ_e perturbations at 3, 6 and 9 UTC. The half-cosine basis we employ is suitable for non-periodic phenomena. We denote the 4-D correlation matrix obtained from the smoothed perturbations $\mathbf{C}_s$ and its separable approximation with $\hat{\mathbf{C}}_s$.

Experimentation showed that setting the Hadamard power $q = 6$ produced an effective reduction in spurious ensemble correlations. The operation $\mathbf{S}^T\hat{\mathbf{C}}_s^{\odot q}$ that transforms the functions $f_\tau^q(i_x), g_\tau^q(i_y)$ and $h_\tau^q(i_z)$ defining the columns of $\hat{\mathbf{C}}_s^{\odot q}$ into $F_\tau(\nu_k), G_\tau(\nu_l)$ and $H_\tau(\nu_m)$ was achieved using the half cosine transform. The operation $\mathbf{S}^T\hat{\mathbf{C}}_s^{\odot q}$ requires that the spectral coefficients $F_\tau(\nu_k), G_\tau(\nu_l)$ and $H_\tau(\nu_m)$ be multiplied by wavenumber dependent coefficients such that small scale modes are suppressed. This high wavenumber suppression was again achieved using coefficients proportional to $\exp[-(\frac{j}{d_{h/v}})^2]$ where j denotes the relevant wavenumber. For this operation, we employed the parameters of $d_h = 4.5$ and $d_v = 3.2$ in the horizontal and vertical directions, respectively. Inspection of the filtered coefficients showed that many of them were of negligible amplitude. This led us to perform a triangular truncation[2] of the smoothed spectral representation $\mathbf{S}^T\hat{\mathbf{C}}_s^{\odot q}$ to at most 5 horizontal modes in each horizontal direction and at most seven vertical modes. The triangular truncation reduced the number of rows in $\mathbf{S}^T\hat{\mathbf{C}}_s^{\odot q}$ to $L = 126$. The normalized matrix $\mathbf{W}^T = \mathbf{S}^T\hat{\mathbf{C}}_s^{\odot q}$ was obtained from $\mathbf{S}^T\hat{\mathbf{C}}_s^{\odot q}$ by dividing each column by its magnitude. Since $\mathbf{C}_{\mathrm{ECORAP}} = \mathbf{W}\mathbf{W}^T = \hat{\mathbf{C}}_s^{\odot 6}\mathbf{S}\mathbf{S}^T\hat{\mathbf{C}}_s^{\odot 6}$, (2) implies that the square root $\mathbf{Z}$ of the CALECO matrix is given by $\mathbf{Z} = [\mathbf{Z}_K \triangle \mathbf{W}]$, where $\mathbf{P}^f_{\mathrm{CALECO}} = \mathbf{Z}\mathbf{Z}^T$. Equation (2) shows that $\mathbf{Z}_K\mathbf{W}$ is a matrix whose LK columns list the LK possible element-wise products of the K columns of $\mathbf{Z}_K$ and the L columns of $\mathbf{W}$. Since we have $K = 27$ raw ensemble members, 26 of which are likely to be linearly independent and since we have $L = 126$ potentially linearly independent columns in $\mathbf{W}$, this means that the rank of the CALECO

[2]Triangular truncation involves the systematic removal of products of high wavenumber coefficients. "Triangular truncation" is obtained if the sum of the coefficients contributing to some coefficient-product exceeds some critical value.

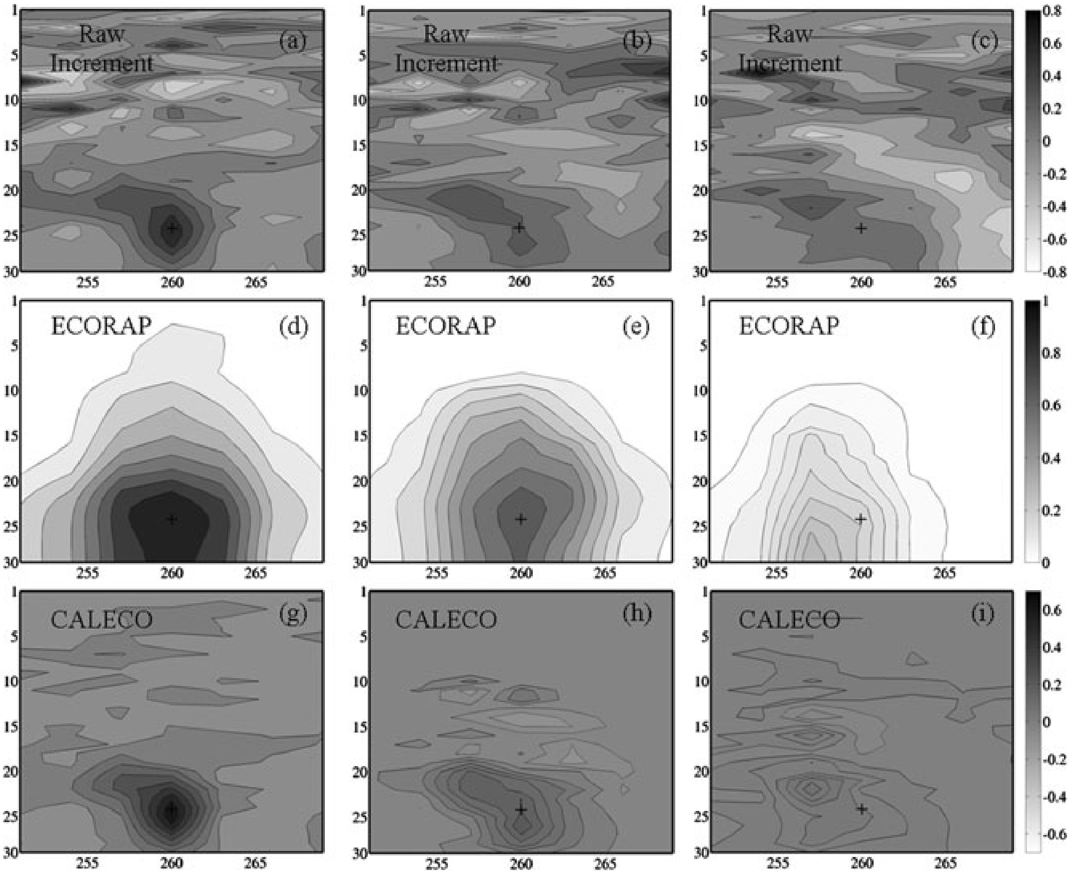

Fig. 5. As in figure 2, but for a meridional wind increment given a single meridional wind observation at 3Z in the tropics (20N, 260E, $\sigma = 24$). Bars give the relation between'shade' and 'value'. For (a)–(c) and (g)–(i) values have units of ms^{-1}.

forecast error covariance matrix is no greater than $26\times126 = 3276$. The number of columns defining $\mathbf{Z}$ as defined by (2) is $27\times126 = 3402$. While we could have reduced this number to 3276 by first performing singular vector decompositions of $\mathbf{Z}_K$ and then discarding the singular vectors associated with zero eigenvalues, we did not bother with this here.

As noted in Bishop et al. (2001) and Hunt et al. (2007), when the number of columns defining the square root of the forecast error covariance matrix (3402 in our case) is less than the number of observations then there is a computational advantage in using the ETKF form of the minimum-error variance estimation equations. The advent of hyper-spectral satellite data has increased the number of data available for assimilation. Most operational centres plan to assimilate between 10^6 and 10^8 observations each assimilation cycle in the very near future. This number of global observations translates into order $10^4 - 10^6$ observations for each of our 21×21 degree observation volumes. Hence, one expects the number of observations in each observation volume to exceed the 3402 columns of $\mathbf{Z}$ defining the CALECO matrix in each volume. (Our assimilation experiment in section 5d features 7350 observations per volume). Under these circumstances, projecting the estimation equations onto the ensemble perturbation subspace as in the ETKF is faster than solving the equations in observation space.

5.3. *Examples of 4-D covariance functions of ECO-RAP and CALECO*

Figure 2 illustrates the key fields defining the CALECO analysis increment (or correction) at 3, 6 and 9 UTC of the meridional wind from a single observation of temperature at 128E, 56S, σ-level 20 at 3 UTC. σ-level 20 corresponds to about 700 hPa. Figures 2a–c gives the vertical cross-section of the 4-D single observation increment that would be obtained if the ensemble covariances were not localized. It shows how the raw 27 member increment propagates East across the observation volume during the six hour DA window and that it tilts westward with height in a manner somewhat reminiscent of a growing baroclinic wave (Holton, 2004). Note the existence of apparently spurious corrections in the stratosphere. Figures 2d–f shows how the ECO-RAP localization function corresponding to this grid point also propagates East across the LETKF volume during the same 6 h DA window. The analysis increment obtained from the CALECO forecast error covariance model is shown in Figures 2g–i. It is

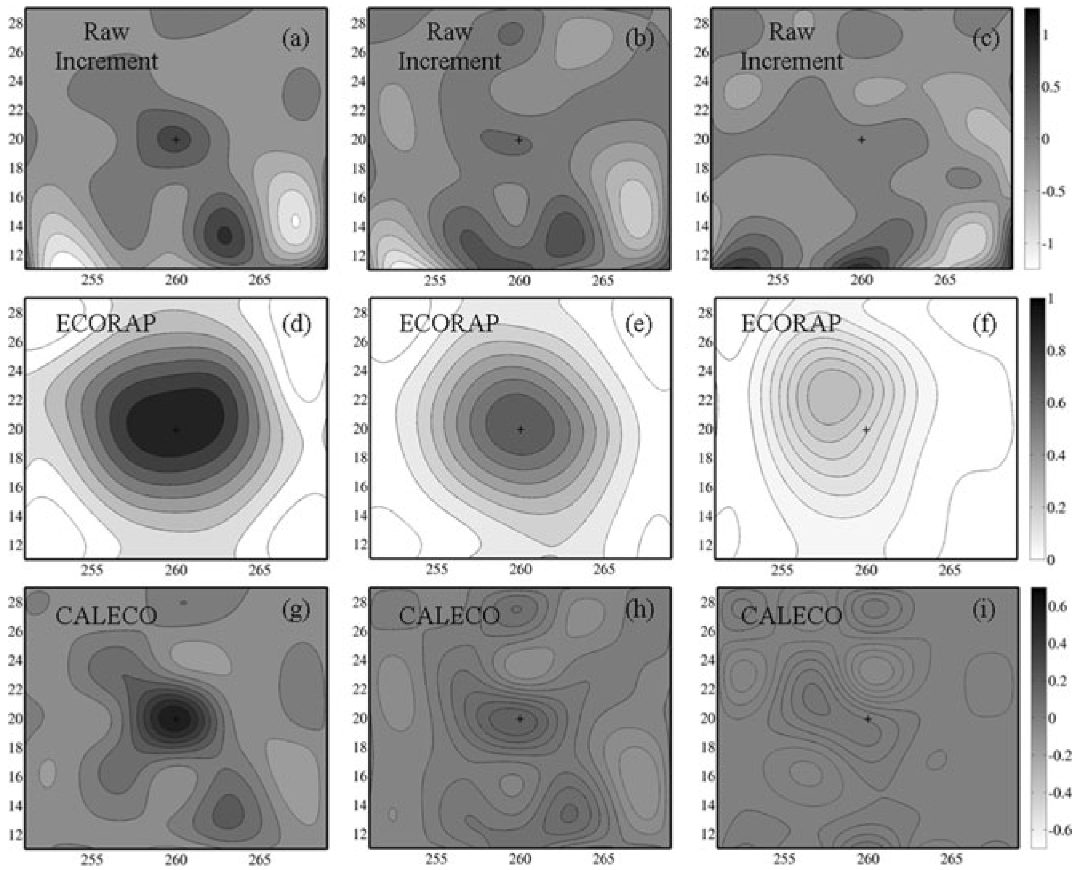

Fig. 6. As in Fig. 5, but for a horizontal cross-section on $\sigma = 24$.

simply the product of the function depicted in the middle panels (d–f) with the corresponding upper panels (a–c). Note that the CALECO increment is more focused around the observation location than the unlocalized increment but nevertheless moves Eastward in a very similar way to the unlocalized increment. The stratosphere is largely unaffected by the CALECO increment.

Figure 3 illustrates the horizontal structure of the increments shown in Fig. 1 on the $\sigma = 20$ surface. It focuses the analysis correction around the observation while moving the analysis correction downstream over the assimilation window.

While the unlocalized analysis correction shown in Figs. 2 and 3 is fairly broad scale and looks qualitatively similar to the CALECO correction, this is not always the case. Figure 4 compares CALECO localized and unlocalized temperature corrections associated with a meridional wind observation at 77E, 60S, σ-level 25 at 3 UTC. σ-level 25 is at about 850 hPa. Here, the unlocalized increment is extremely noisy while the CALECO increment is focused but nevertheless moving with the unlocalized increment.

Figures 5 and 6 give an example of ECO-RAP localization closer to the tropics. The figures show the meridional wind increment associated with a meridional wind observation at (3Z, 20N, 260E, σ-level 25). In this case, the maximum value of the ECO-RAP localization function drops from a value of 1 at 3 UTC to a value of about 0.3 at 9 UTC. Part 1 suggested that such drops in the magnitude of the ECO-RAP localization function can be due to dispersion or to processes that have little correlation through time. We found that this property of ECO-RAP localization increased as the location of single observation increments was moved closer to the equator.

5.4. *Global assimilation experiment*

Here, we compare the performance of the LETKF with and without ECORAP localization. Assimilation is performed in 153 distinct observation volumes that are equally distributed across the globe at 20° intervals from 20E to 340E in longitude and 80S to 80N in latitude. The zonal and meridional winds are observed at every grid point at 3 UTC and 9 UTC (5880 observations). Temperature is observed at every grid point at 6 UTC (1470 observations). These 7350 observations are then used to estimate the 6Z state of u, v and T at the 30 vertical levels at the centre of the observation volume.

To the extent that each observation volume is sufficiently separated from its neighbours to ensure statistical independence, the

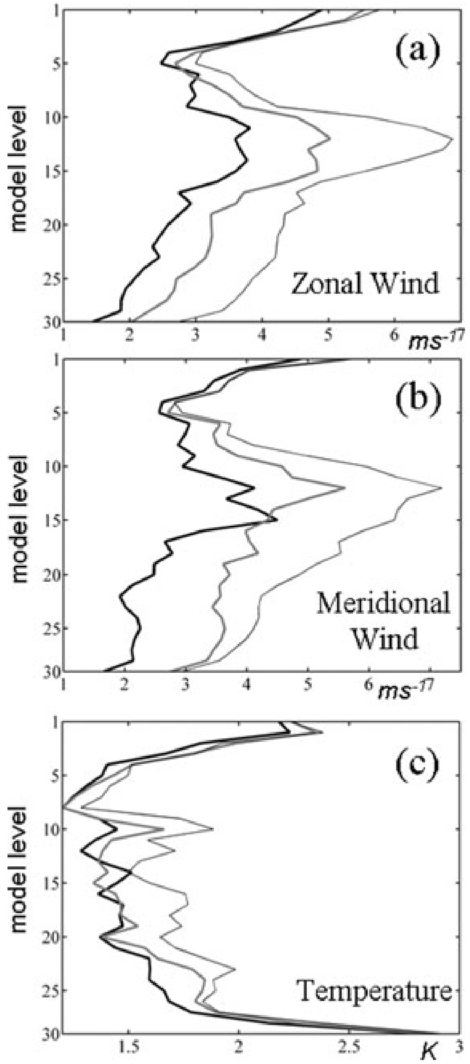

Fig. 7. Globally averaged RMS error (abscissa) as a function of model-level (ordinate). In each panel is the rms forecast error (thin black line), rms analysis error from the LETKF (thick grey line), and the rms analysis error from the CALECO LETKF (thick black line). Units are m s^{-1} for u and v and Kelvin for T.

153 observation volumes provide 153 statistically independent events. These 153 distinct observation volumes are ordered into 9 groups of 17. Each group lies on 1 of 9 latitude bands. We compute the average analysis and forecast error in each of the 9 bands for the LETKF with and without ECO-RAP localization. If there were no underlying differences in DA performance with or without ECO-RAP, and the forecast errors within each band were independent, then the distribution of wins and losses would be the same as the distribution of outcomes associated with 9 flips of a coin. If, for example, we found that ECO-RAP localization produced smaller analysis errors in all 9 bands then the odds of this happening by chance would be $1/2^9$ and the null hypothesis could be rejected with a confidence of 99.8%.

Figure 7 shows the globally averaged RMS error as a function of σ-level at 6Z. ECO-RAP localization produces much smaller analysis estimates for both the zonal and meridional winds and slightly better estimates for temperature. Since there are no wind observations at 6 UTC, the improvements in wind estimates at 6 UTC are attributable to CALECO's covariances turning wind observations at 3 and 9 UTC into wind analysis improvements at 6 UTC and/or CALECO's covariances turning the temperature observations at 6 UTC into wind analysis improvements at 6 UTC.

The globally vertically averaged results for root mean square analysis and forecast errors are given in Fig. 8a. The LETKF with ECO-RAP localization achieves approximately double the improvement in the RMS analysis error as the LETKF without localization for all variables. It was found that CALECO LETKF beat LETKF in 8 out of the 9 latitude bands for the meridional wind and for the zonal wind. Hence, for these fields, the null hypothesis 'that CALECO LETKF analyses were no better than LETKF analyses' could be rejected with a confidence of 98%. For temperature, CALECO LETKF beat LETKF in 7 out of the 9 bands and hence the null hypothesis for temperature improvements could be rejected with 93% confidence.

Figures 8b–d show that the largest forecast errors and the greatest improvements due to DA occurred at mid-latitudes in both hemispheres. They also show that the smallest difference between the raw LETKF and the ECO-RAP enhanced LETKF occurred for temperature in the southern hemisphere. Figure 8c shows that both schemes gave relatively small improvements in the tropics. While CALECO LETKF improved temperature in the tropics, standard LETKF did not.

6. Conclusions

A cost effective method for incorporating the CALECO adaptively localized ensemble covariance model in a LETKF type DA scheme has been presented. Without the cost saving procedures introduced in this paper, the cost of preparing the CALECO forecast error covariance matrix for use in the LETKF would likely exceed the cost of the LETKF itself. With the cost saving procedures, the preparation cost is trivial in comparison to the LETKF cost. The key cost saving is associated with a factorization property of the CALECO matrix that allows a large ensemble, whose covariance is the CALECO matrix, to be inexpensively generated from element-wise products between raw ensemble members and columns of the square root of the ECO-RAP localization matrix.

In the standard LETKF, observation volumes and distance dependent inflation of observation error variances are used to prevent spurious ensemble correlations degrading the quality of the analysis. In the CALECO LETKF, spurious ensemble correlations are suppressed using ECO-RAP. Hence, CALECO LETKF observation volumes can be much larger than those of the standard LETKF.

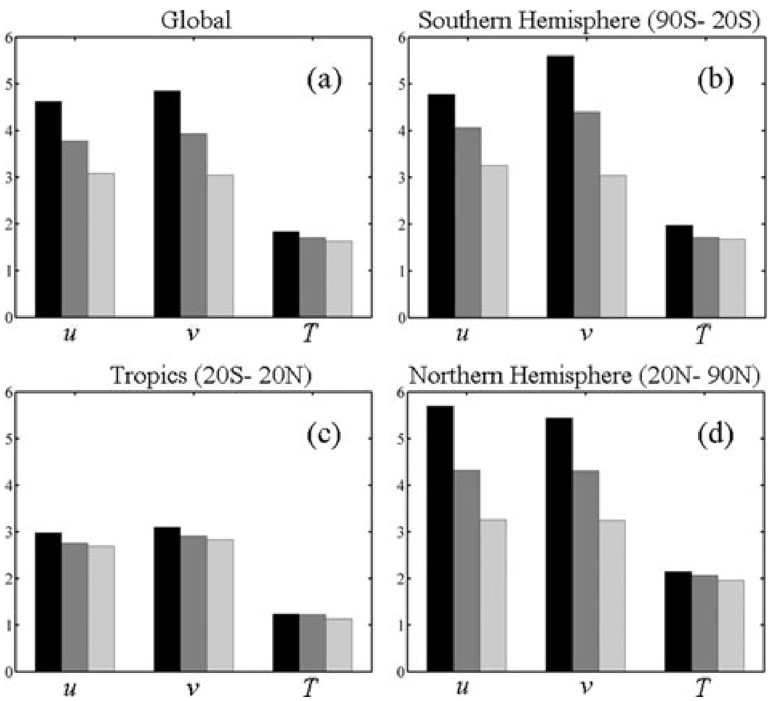

Fig. 8. Averaged rms error binned according to region. Black, grey and light-grey bars give the vertically averaged rms of the forecast error, LETKF analysis error, and CALECO LETKF analysis error, respectively. Units are m s^{-1} for u and v and Kelvin for T.

The large observation volumes enabled by the CALECO LETKF mean that each CALECO observation volume can be used to update many grid points simultaneously. The potential cost savings associated with this feature were analysed using the concept of error dimension density. The analysis indicated that when the number of CALECO ensemble members required to accurately analyse Miyoshi and Yamane's (2007) standard observation volume was less than three times the size of Miyoshi and Yamane's ensemble, then a range of configurations existed in which the CALECO LETKF would be able to update the state more efficiently than the standard LETKF (Fig. 1). However, for higher error dimension densities, the CALECO LETKF would always be more expensive than the standard LETKF.

This two part paper does not settle the question of whether, for sophisticated weather models, CALECO LETKF DA with large observation volumes is more accurate than standard LETKF DA but it does provide a basis for experiments that could. Part 1 showed that adaptive localization is better than non-adaptive localization for simple error systems where either the scale of error changes in space–time and/or the error propagates a significant distance compared with the scale of localization. Part 2 shows that such error propagation occurs in the atmosphere and that ECO-RAP localization functions adapt to it (Figs. 2–6). The localization functions associated with these figures looked reasonable and they reduced analysis error relative to a LETKF that did not use localization but used the same observation volume as the CALECO LETKF (Figs. 7 and 8). Another potential performance advantage of the CALECO LETKF over the standard LETKF is that it is better suited for non-local satellite observations because (1) its observation volumes are not confined in the vertical and (2) it does not require distance dependent observation error variance inflation.

7. Acknowledgments

CHB and DH acknowledge support from ONR Project Element 0602435N, Project Number BE-435-003, ONR Grant number N0001407WX30012 and from NOAA THORPEX grant NA04AANRG0233. The research was performed while DH held a National Research Council Research Associateship Award at the Naval Research Laboratory.

Appendix A: Conjugate gradient approach (Supporting Information)

For the Monte-Carlo form, note that the update equation for the ith ensemble member satisfies the following relations

$$\begin{aligned}\mathbf{x}_{ui}^{a} &= \mathbf{x}_{ui}^{f} + \mathbf{Z}_u \mathbf{C}\,(\Gamma + \mathbf{I})^{-1}\,\mathbf{C}^T \mathbf{Z}_u^T \mathbf{H}^T \mathbf{R}^{-1/2}(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f) \\ &= \mathbf{x}_{ui}^{f} + \mathbf{Z}_u (\mathbf{C}\Gamma\mathbf{C}^T + \mathbf{I})^{-1} \mathbf{Z}_u^T \mathbf{H}^T \mathbf{R}^{-1/2}(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f), \\ &\qquad \text{by orthonormality of } \mathbf{C} \\ &= \mathbf{x}_{ui}^{f} + \mathbf{Z}_u (\mathbf{Z}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\mathbf{Z} + \mathbf{I})^{-1} \mathbf{Z}_u^T \mathbf{H}^T \mathbf{R}^{-1/2} \\ &\qquad (\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f), \text{ because } \mathbf{Z}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\mathbf{Z} = \mathbf{C}\Gamma\mathbf{C}^T. \end{aligned} \tag{A1}$$

Hence, if the J-vector $\mathbf{w}$ satisfies

$$(\mathbf{Z}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\mathbf{Z} + \mathbf{I})\mathbf{w} = \mathbf{Z}_u^T \mathbf{H}^T \mathbf{R}^{-1/2}(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f) \tag{A2}$$

then

$$\mathbf{x}_{ui}^{a} = \mathbf{x}_{ui}^{f} + \mathbf{Z}_u \mathbf{w}. \tag{A3}$$

Thus, if one defines a cost function F such that

$$F = \frac{1}{2}\mathbf{w}^T\left(\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{Z} + \mathbf{I}\right)\mathbf{w} - \mathbf{w}^T\mathbf{Z}_u^T\mathbf{H}^T\mathbf{R}^{-1/2}\left(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f\right) \quad \text{(A4)}$$

then because

$$\frac{\partial F}{\partial \mathbf{w}} = \left(\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{Z} + \mathbf{I}\right)\mathbf{w} - \mathbf{Z}_u^T\mathbf{H}^T\mathbf{R}^{-1/2}\left(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f\right) \quad \text{(A5)}$$

it follows that F is minimized when (A2) is satisfied. Thus, (A3) can be used to update the ensemble members, provided that $\mathbf{w}$ minimizes F. Assuming that a descent algorithm such as the conjugate gradient method could minimize F in 10–100 iterations, the cost of finding the $\mathbf{w}$ value that minimizes F would be between 10 and 100 times the cost of performing the operation $(\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{Z} + \mathbf{I})\mathbf{w}$.

The cost of this operation depends on how the term $\mathbf{HZ}$ is handled. If handled in the manner described in Section 2, $\mathbf{HZ}$ would be turned into a $p \times J$ matrix using only the non-linear observation operator at a cost of about $10n_z n_v pJ$ operations. Having formed $\mathbf{HZ}$ in this way, the cost of $(\mathbf{Z}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{Z} + \mathbf{I})\mathbf{w}$ would be between $10Jp$ and $100Jp$ operations. In this case, the $\mathbf{HZ}$ once formed can be used for all ensemble members and hence, the total cost of this approach to update the entire ensemble is $10n_z n_v pJ + K\ 100\ J^2$.

References

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

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

Bishop, C. H. and Toth, Z. 1999. Ensemble transformation and adaptive observations. *J. Atmos. Sci.* **56**, 1748–1765.

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.

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

Daley, R. and Barker, E. 2001. NAVDAS: Formulation and diagnostics. *Mon. Wea. Rev.* **129**, 869–883.

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

Hogan, T. F. and Rosmond, T. E. 1991. The description of the Navy Operational Global Atmospheric Prediction System's spectral forecast model. *Mon. Wea. Rev.* **119**, 1786–1815.

Holton, J. R. 2004. *An Introduction to Dynamic Meteorology*. 4th Edition, Academic Press, San Diego, pp. 511.

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., Kostelich, E. J. and Szunyogh, I. 2007. Efficient data assimilation for a spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D* **230**, 112–126.

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

McLay, J. G., Bishop, C. H. and Reynolds, C. A. 2007. The ensemble transform scheme adapted for the generation of stochastic perturbations. *Quart. J. Roy. Meteor. Soc.* **133**, 1257–1266.

McLay, J. G., Bishop, C. H. and Reynolds, C. A. 2008. Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.* **136**, 1093–1108.

Miyoshi, T. and Sato, Y. 2007. Assimilating satellite radiances with a local ensemble transform kalman filter (LETKF) applied to the JMA global model (GMS). *SOLA 2007*, Vol. 3, 37–40. Available at http://www.jstage.jst.go.jp/article/sola/3/0/3_37/_article

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.

Miyoshi, T., Yamane, S. and Enomoto, T. 2007. Localizing the error covariance by physical distances within a local ensemble transform kalman filter (LETKF). *SOLA 2007*, Vol. 3, 37–40. Available at http://www.jstage.jst.go.jp/article/sola/3/0/3_37/_article

Ott, E., Hunt, B. H., 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.

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A. and Ott, E. 2001. Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.* **86**, 5878– 5881.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. 2007. *Numerical Recipes: The Art of Scientific Computing*, 3rd Edition. Cambridge University Press, Cambridge, pp. 1256.

Reynolds, C. A., Teixeira, J. A. and McLay, J. G. 2008. Impact of stochastic convection on the Ensemble Transform. *Mon. Wea. Rev.* **136**, 4517–4526.

Szunyogh, I., Kostelich, E. J., Gyarmati, G., Patil, D. J., Hunt, B. R. and co-authors. 2005. Assessing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus* **57A**, 528–545.

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* **60**A, 113–130.

Wang, X. and Bishop, C. H. 2003. A comparison of breeding and ensemble transform kalman filter forecast schemes. *J. Atmos. Sci.* **60**, 1140–1158.

Wang, X., Bishop, C. H. and Julier, S. J. 2004. Which is better, an ensemble of positive/negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.* **132**, 1590–1605.

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., 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.