

# **Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter**

Shu-Chih Yang<sup>1,2</sup>, Eugenia Kalnay<sup>1,3</sup>, Brian Hunt<sup>1,3</sup> and Neill E. Bowler<sup>4</sup>

<sup>1</sup> Department of Atmospheric and Oceanic Science, University of Maryland

<sup>2</sup> Department of Atmospheric Sciences, National Central University

<sup>3</sup> Institute for Physical Science and Technology, University of Maryland

<sup>4</sup> Met Office, FitzRoy Road, Exeter, UK

Corresponding author:

Shu-Chih Yang

Email: [shuchih.yang@atm.ncu.edu.tw](mailto:shuchih.yang@atm.ncu.edu.tw)

Department of Atmospheric Sciences,  
National Central University, Jhung-Li, Taiwan, 320

## **Abstract**

We have investigated a method to substantially reduce the analysis computations within the Local Ensemble Transform Kalman Filter (LETKF) framework. Instead of computing the LETKF analysis at every model grid point, we compute the analysis on a coarser grid and interpolate onto a high-resolution grid by interpolating the analysis weights of the ensemble forecast members derived from the LETKF.

Because the weights vary on larger scales than the analysis increments, there is little degradation in the quality of the weight-interpolated analyses compared to the analyses derived with the high-resolution grid. The weight-interpolated analyses are more accurate than the ones derived by interpolating the analysis increments. Additional benefit from the weight-interpolation method includes improving the analysis accuracy in the data-void regions, where the standard LEKTF with the high-resolution grid gives no analysis corrections due to a lack of available observations.

## 1. Background

The resolutions of modern numerical weather/ocean models and observation density have greatly increased in recent years in order to resolve the dynamic processes in smaller scales such as convective scales in the atmosphere and eddies in the ocean. As a consequence, the computational cost required for data assimilation (DA) has also increased. To reduce the computational cost, methods like variational analysis (3DVar and 4D-Var) focus on reducing the computation during the minimization process of the cost function. For example, 4D-Var is solved using the incremental form (Courtier et al. 1994) in which the so-called inner loop is carried out by running the adjoint model at a lower resolution with simplified physics.

For the ensemble methods such as Ensemble Kalman Filter, the computational cost can be alleviated by allowing the analysis to be computed in parallel in local regions (Keppenne et al. 2002, Ott et al. 2004, Hunt et al., 2007). However, the computational burden for such local analyses is still constrained by the ensemble size and the total number of local regions in a high-resolution model. It is possible to reduce the computation further by carrying out the ensemble analyses on a coarser resolution, as done with incremental variational analyses, and then interpolating to the finer resolution. We will show that such interpolation step degrades the accuracy of the analysis, compared to a full-resolution analysis.

In this study, we investigate the feasibility of a method to reduce the computational cost in the assimilation procedure. This method is developed following a suggestion of Bowler (2006, Section 6.4), who computed the transform matrix derived from the solution of Ensemble Transform Kalman Filter (Bishop et al. 2001) on a coarse grid and then interpolated it onto the high-resolution grid. The idea is to use the model output of ensemble forecasts in full resolution without sacrificing the accuracy of the analysis. A similar idea of weight interpolation (Section 3.2 in Keppenne et al. 2008) is used to speed up the analysis computation of an ensemble Kalman filter (EnKF ocean, Keppenne et al. 2002). However, in these studies the properties of the weights for linearly combining the ensemble perturbations and how this relates to maintain the analysis accuracy are not discussed in detail. For example, Keppenne et al. (2008) compares the computational time and the performance loss for one analysis cycle with different sparseness of the analysis grid, not mentioning the impact on the overall analysis accuracy by recycling these analyses. In addition, Bowler (2006) investigates the weight-interpolation method with a simple Lorenz95 model (Lorenz, 1995; Lorenz and Emanuel, 1998) under a perfectly observed framework (observation is available at every grid point). Therefore, implications in a more realistic setup with under-observed regions, are still unclear.

In this study, the weight-interpolation method is adapted to the framework of the Local Ensemble Transform Kalman Filter (LETKF, Ott et al., 2004 and Hunt et al. 2007). This method has the advantage of being configured for local analyses, where the analysis ensemble is expressed with weights that linearly combine the background (12-hour forecast) ensemble. The interpolation is done through these weights for the background

ensemble in order to create an analysis at full resolution. As we will show, this method allows us to substantially reduce the required number of local analyses for the LETKF assimilation without sacrificing analysis accuracy.

This paper is organized as follows: Section 2 briefly describes the implementation of the LETKF in a Quasi-Geostrophic model and the setup of the assimilation experiments. In Section 3, we explain how the sparse analysis is done within the LETKF framework. The results of the interpolated data assimilation experiments are discussed in the Section 4. A summary and discussion is given in Section 5.

## **2. Data assimilation in a Quasi-Geostrophic model**

### **2.1 Local Ensemble Transform Kalman Filter (LETKF)**

The LETKF scheme, described in detail in Hunt et al. (2007), is an ensemble-based Kalman Filter that analyzes the current model state within a local region. Similar to other ensemble Kalman filters the error statistics are sampled by the ensemble except that they are evaluated locally. The local analysis is then derived based on the local error statistics and the observations collected in the same local region.

The LETKF scheme is a more efficient version of the Local Ensemble Kalman Filter proposed in Ott et al. (2004). This scheme has been shown to have similar accuracy as other sequential ensemble-based Kalman Filters implemented in the same numerical weather prediction model (Whitaker et al. 2008). Its parallel implementation, made

possible because the analysis at each grid point is independent of other grid points, becomes more efficient as the number of processors grows.

We briefly discuss the LETKF scheme, where the data assimilation is performed at the central grid point of a local region and sequentially loops for all the analysis grids.

Let  $\mathbf{x}_f$  denote a matrix whose columns are the background ensemble in a local region evolved from a set of perturbed initial conditions. The ensemble states can be represented by (1):

$$\mathbf{x}_f = \bar{\mathbf{x}}_f + \mathbf{X}_f, \quad (1)$$

where  $\bar{\mathbf{x}}_f$  is a column vector containing the mean of the ensemble and  $\mathbf{X}_f$  is a matrix whose columns are the background ensemble perturbations from the ensemble mean.

With  $K$  denoting the ensemble size, the background error covariance matrix is defined as

$$\frac{1}{K-1} \mathbf{X}_f \mathbf{X}_f^T.$$

Similar definitions are given for the analysis ensemble mean and the perturbations:  $\bar{\mathbf{x}}_a$  and  $\mathbf{X}_a$  respectively.

The LETKF determines a transform matrix that converts the local background ensemble perturbations into the analysis ensemble perturbations. The local analysis error covariance can be written as (2), where  $\mathbf{P}_a$  is the analysis error covariance matrix and  $\tilde{\mathbf{P}}_a$ , the analysis error covariance matrix in ensemble space.

$$\mathbf{P}_a = \frac{1}{K-1} \mathbf{X}_a \mathbf{X}_a^T = \mathbf{X}_f \tilde{\mathbf{P}}_a \mathbf{X}_f^T, \quad (2)$$

The transform matrix (Hunt et al., 2007), is computed as:

$$\tilde{\mathbf{P}}_a = \left[ (K-1)\mathbf{I} + \mathbf{Y}_f^T \mathbf{R}^{-1} \mathbf{Y}_f \right]^{-1} \quad (3)$$

, where  $\mathbf{Y}_f$ , the background perturbations in observation space,  $\mathbf{R}$  is the observation error covariance assumed to be diagonal. With the observation operator  $H$ , the background ensemble is converted from the model space to observation space. The background perturbations in observation space are then approximated by computing the mean and the deviations in observation space, denoting as  $\bar{\mathbf{y}}_f$  and  $\mathbf{Y}_f$  in (4).

$$H(\mathbf{x}_f) = H(\bar{\mathbf{x}}_f + \mathbf{X}_f) \approx \bar{\mathbf{y}}_f + \mathbf{Y}_f \quad (4)$$

Given that the ensemble size is much smaller than the model dimension, the  $K \times K$  matrix  $\tilde{\mathbf{P}}_a$  is efficiently computed within the ensemble space.

After  $\tilde{\mathbf{P}}_a$  is obtained, the mean analysis at the central grid point of the local region is computed from the background ensemble mean according to (5)

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_f + \mathbf{X}_f \tilde{\mathbf{P}}_a \mathbf{Y}_f^T \mathbf{R}^{-1} (\mathbf{y}_o - \bar{\mathbf{y}}_f) = \bar{\mathbf{x}}_f + \mathbf{X}_f \bar{\mathbf{w}}_a \quad (5)$$

In (5), the  $K \times 1$  vector of weight,  $\bar{\mathbf{w}}_a$  is derived with the information about observational increments,  $\mathbf{y}_o - \bar{\mathbf{y}}_f$ . In the final step, the analysis ensemble perturbations at the central grid point are derived by multiplying the background ensemble perturbations by the symmetric square root of  $(K-1)\tilde{\mathbf{P}}_a$ :

$$\mathbf{X}_a = \mathbf{X}_f [(K-1)\tilde{\mathbf{P}}_a]^{1/2} = \mathbf{X}_f \mathbf{W}_a. \quad (6)$$

In (6),  $\mathbf{W}_a$  is a multiple of the symmetric square root of the local analysis error covariance matrix in ensemble space. It is computed by singular vector decomposition (SVD):

$$\mathbf{W}_a = \mathbf{U}\mathbf{S}^{1/2}\mathbf{U}^T \quad (7)$$

where  $\mathbf{U}$  is the matrix whose columns are the left singular vectors of  $(K-1)\tilde{\mathbf{P}}_a$  and  $\mathbf{S}$  is a diagonal matrix whose diagonal elements are the singular values.

The use of a symmetric square root matrix ensures that the sum of the analysis ensemble perturbations is zero and depends continuously on  $\tilde{\mathbf{P}}_a$  (Hunt et al. 2007). Adjacent analysis points, whose corresponding local background ensemble has small differences, will have slightly different  $\tilde{\mathbf{P}}_a$ . The derived symmetric square root matrix can carry such characteristics and thus yield similar analysis ensemble perturbations at adjacent points, necessary to ensure the smoothness of the analysis. This property of the symmetric square root matrix also ensures that the analysis ensemble perturbations are consistent with the background ensemble perturbation since the symmetric square root matrix makes  $\mathbf{W}_a$  the matrix closest to the identity matrix<sup>1</sup>, given the constraint of the analysis error covariance matrix (Ott et al. 2004). Harlim (2006) demonstrated that the symmetric solution has better performance than those obtained with a non-symmetric square root, given the same ensemble size.

Eqs. (5) and (6) show that the analysis ensemble at each grid point is simply derived through a linear combination of the background ensemble, with weighting coefficients

---

<sup>1</sup> The distance between  $\mathbf{W}_a$  and the identity matrix is measured by the Frobenius norm.



given by  $\bar{\mathbf{w}}_a$  (a  $K \times 1$  vector) for the mean analysis, and  $\mathbf{W}_a$  (a  $K \times K$  matrix) for the analysis perturbations. Thus, the  $k$ th analysis ensemble member is given by

$$\mathbf{x}_{a,k} = \bar{\mathbf{x}}_f + \mathbf{X}_{f,k} [\bar{\mathbf{w}}_a + \mathbf{W}_{a,k}] \quad (8)$$

where  $\mathbf{W}_{a,k}$  is the  $k$ th column of  $\mathbf{W}_a$ .

Let  $\mathbf{u}$  denotes a vector of  $K$  ones,  $\mathbf{u} = (1,1,\dots,1)^T$ ,  $\mathbf{u}$  is an eigenvector of  $\mathbf{X}_f$  because of the sum of the perturbations has zero mean ( $\mathbf{X}_f \mathbf{u} = 0$ ). Since the sum of the columns of  $\mathbf{Y}_f$  is equal to zero,  $\mathbf{v}$  is also an eigenvector of  $\tilde{\mathbf{P}}_a^{-1}$  and, thus, of  $\mathbf{W}_a$  as well (Hunt et al. 2007):

$$\mathbf{W}_a \mathbf{W}_a \mathbf{u} = (K - 1) \tilde{\mathbf{P}}_a \mathbf{u} = \mathbf{u} \quad (10)$$

With  $\mathbf{X}_f \mathbf{u} = 0$ , the analysis perturbations also have a zero mean (11).

$$\mathbf{X}_a \mathbf{u} = \mathbf{X}_f \mathbf{W}_a \mathbf{u} = 0, \quad (11)$$

## 2.2 The Quasi-Geostrophic model

In this study, the LETKF is implemented on a quasi-geostrophic (QG) model (Rotunno and Bao 1996). The dynamic flow described by this model is confined to a zonally periodic channel and bounded meridionally at top and bottom by rigid surfaces. The model is forced by relaxation to a specific zonal mean state at all levels and an Ekman pumping at the bottom level. A fourth-order horizontal diffusion provides dissipation. The model variables are non-dimensionalised. They are pseudo-potential vorticity ( $q$ ) in

the five internal levels, governed by (11a) and non-dimensional temperatures ( $\theta$ ) at the bottom and top levels ( $z = 0,1$ ), governed by Eq (11b).

$$\frac{\partial q}{\partial t} + \mathbf{v} \cdot \nabla q = (-\tau^{-1} + D)(q - q_{ref}) \quad (11a)$$

$$\frac{\partial \theta}{\partial t} + \mathbf{v} \cdot \nabla \theta = \begin{cases} -\Gamma \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \phi + (-\tau^{-1} + D)(\theta - \theta_{ref}), & at\ z = 0 \\ (-\tau^{-1} + D)(\theta - \theta_{ref}), & at\ z = 1 \end{cases} \quad (11b)$$

$q$  is defined through the streamfunction,  $\phi$ , as  $q = \beta y + \partial^2 \phi / \partial z^2 + \nabla^2 \phi$  and the velocity and the temperature are  $(u, v, \theta) = (-\partial \phi / \partial y, \partial \phi / \partial x, \partial \phi / \partial z)$ . In (11a,b),  $q_{ref}, \theta_{ref}$  are the zonal mean reference state for potential vorticity and temperature and  $\mathbf{v} = (u, v)$  is the horizontal velocity.  $\tau$  is the relaxation time and  $D$  denotes the fourth-order horizontal diffusion and  $\Gamma$  controls the Ekman pumping at the bottom level ( $z=0$ ). Further details on the mathematical formulations and the numerical schemes used for advecting and inverting the potential vorticity are described in Rotunno and Bao (1996). The forcing and dissipation included in the model are specified in Snyder et al. (2003).

In the following experiments, the dimensional values for the model parameters are the same as in Morss (1999) and Snyder et al. (2003, see Table 1). There are 64 grid-points in the zonal direction, 33 grid-points in the meridional direction and a total of 7 levels in vertical.

For this QG model, its error-doubling rate is 2.5-4 days, on average slower than the real atmosphere. Their characteristics of the dynamical errors can be found in Corrazza et al. (2003) and Snyder et al. (2003). Yang et al. (2008) discuss the analysis errors from different assimilation schemes implemented in this model.

### **2.3 Setup for data assimilation experiments**

The data assimilation framework has a 12-hour analysis cycle and the analysis variables are the model variables. The analyses are verified against the true states, generated from a long model integration with the perfect model setting. Observations are generated by adding random Gaussian errors to the true profiles of zonal and meridional winds and potential temperatures. The observation error is  $0.8 \text{ ms}^{-1}$  for the zonal wind,  $0.5 \text{ ms}^{-1}$  for the meridional wind and  $0.8^\circ\text{C}$  for potential temperature. These amplitudes are about 20%, 17% and 11% of the climate variability for temperature, zonal and meridional wind, respectively. For the data assimilation experiments, 128 “rawinsonde” observations are available every 12-hour. Their locations are determined at the beginning of the experiments by random selection with a minimum spacing between observations. In this study, we allow the forecast model to be slightly imperfect by underestimating the amount of Ekman pumping. This is done by decreasing the vertical eddy diffusion from  $5 \text{ m}^2/\text{s}$  (in the true run) to  $4.75 \text{ m}^2/\text{s}$ . With such choice, the imperfect model causes slightly smaller climate variability. Compared to the perfect climate variability, it reduces 0.7% of the climate variability of the potential temperature at the bottom level.

The implementation of the LETKF for this QG model has been discussed in Yang et al. (2008), who compared the performance of LETKF with other DA schemes in this model. They showed that with a perfect model configuration (no model errors), the analysis derived from the LETKF was more accurate than the analyses from the 3D-Var or from the 4D-Var with a short assimilation window<sup>2</sup> (12-hour), but with 24-hour windows 4D-Var was more accurate. A similar DA framework as in Yang et al. (2008) is used here to test the method of interpolated LETKF analyses.

To initialize the ensemble members, the first initial condition of the ensemble is centered at an analysis derived from the three-dimensional variational data assimilation system (3D-Var, developed by Morss 1999). The 3D-Var background error covariance is time-independent and is constructed following the assumptions in Parrish and Derber (1992)<sup>3</sup>. Note that the 3D-Var analysis is performed 100 days before the starting time of the LETKF experiments to obtain a initial condition not too far from the true state. This choice for the first initial condition is because that the spin-up time for the ensemble-based assimilation schemes depends on the accuracy of the background state in the first analysis cycle (Caya et al. 2005, Kalnay and Yang, 2008). The first initial ensemble perturbations are Gaussian random perturbations of 0.05 variance. This amplitude is an order larger than the mean analysis error obtained with the LETKF with 128 observations.

---

<sup>2</sup> In this QG model, 12-hour is a relatively short assimilation window, according to the error growth rate (Yang et al. 2008).

<sup>3</sup> The 3D-Var background error covariance matrix is diagonal in the spectral space with separable horizontal and vertical structures. It is also assumed that the structure of the covariance is insensitive to the observation density. Details about the 3D-Var system in the QG model can be found in Morss (1999).

For the LETKF analysis, the local patch for performing local analysis is a cuboid whose center is the analysis grid point, For our experiments, we use 20 ensemble members and choose the local patch to be  $7 \times 7$  grid-points in the horizontal and to include the whole column (7 levels). For each local analysis applied on each analysis grid,  $\bar{\mathbf{x}}_{a/f}$  is a vector with a size of 1029, containing all the model variables within each local domain. Similarly,  $\mathbf{X}_{a/f}$  is the perturbation matrix with a size of  $1029 \times 20$ . The local analysis ensemble is then updated by (5-6).

To optimize the LETKF performance, three procedures are carried out additionally. First, a Gaussian function with a decorrelation length of 5 grid points is applied to the observation error covariance to reduce the impact of more distant observations on the analysis (Miyoshi, 2005). Second, the multiplicative inflation values are arranged to vary in vertical due to the mean error profile and remain constant in time. From the bottom to top levels, they are 8.4%, 5.8%, 6.7%, 8%, 7.5%, 7.4% and 13.8 %. Finally, an additive perturbation with a very small amount of random perturbations (2% of the observation error) is added onto the analysis ensemble perturbations (Corazza et al., 2007). This procedure encourages the ensemble to capture more sub-growing directions and avoids the ensemble tendency to collapse into similar directions.

We note that for local patches without available observations, there is no update in the standard LETKF for the background ensemble. The analysis ensemble is directly taken

from the background ensemble in these local patches, giving  $\bar{\mathbf{w}}_a$  a vector of zeros and  $\mathbf{W}_a$  equals to the identity matrix.

### 3. Interpolation of the analysis weights

The local analysis error covariance matrix in the LETKF is estimated by combining the contributions from each member of background perturbations and each available observation in the local region. Such contributions are represented in the local ensemble space by the weighting coefficients ( $\bar{\mathbf{w}}_a$ ,  $\mathbf{W}_a$  in (4-5)). The same information from observations and background states is used over several regions due to the use of overlapping local regions. This ensures that the weights vary slowly and allows the local analysis to be performed on a limited number of grid points (a coarse-resolution grid), and to spread out the information of the error statistics to the higher-resolution grid through the interpolation of the weights.

Figure 1 illustrates how the sparse analysis is done within the LETKF configuration. The background ensemble is available at all the grid points of high resolution denoted as dots and crosses in Figure 1. The dots, arranged on a coarser grid, denote the grid points where the LETKF analyses and the weighting coefficients ( $\bar{\mathbf{w}}_a$  vector and  $\mathbf{W}_a$  matrix) are computed. In this example, the analysis is only computed every three grid-points in both zonal and meridional. Therefore, considering any  $3 \times 3$  grid-box, the coverage of the analysis grid, defined as the number of analysis grid points divided by the total number of

grid points is 11%. After the weight coefficients are collected on the coarse grids, we interpolate  $\bar{w}_a$  and  $\mathbf{W}_a$  onto the high-resolution grid points where no analysis has been computed, in order to generate the weight fields.

This will result in one map of weight coefficients associated with the mean analysis increment, and  $K$  maps of weights associated with the analysis ensemble perturbations. As discussed in Section 2, the first map of weights ( $\bar{w}_a$ ) represents the observational impact in correcting the mean state, and the latter  $K$  maps ( $\mathbf{W}_a$ ) apply “errors of the day” structures<sup>4</sup> based on the dynamical evolution of the ensemble perturbations to generate the analysis perturbations. Yang et al. (2008) showed that the accuracy of the ensemble mean state and the flow-dependent perturbations related to “errors of the day” are both important in improving the accuracy of the LETKF analysis. By projecting the corrections onto the local dynamical instabilities estimated by the ensemble, the LETKF background error covariance is able to properly correct the background ensemble mean with the available observations. At the same time, the accuracy of the mean state also determines the effectiveness of ensemble perturbations.

For the interpolation of the weights, we apply a smooth bi-variate interpolation scheme (Akima, 1978) based on locally fitting quintic polynomials as the function of the zonal and meridional positions of the analysis grid. The chosen interpolation scheme is linear in the interpolated values. This ensures that the vector of  $K$  ones is still an eigenvector of

---

<sup>4</sup> The “error of the day” is the time-dependent, dynamical error related to the dynamically evolving instability (Kalnay, 2003). Corazza et al. (2003) provides detail discussions about the characteristics of the “error of the day” from this QG model.

the interpolated  $\mathbf{W}_a$ , so that the analysis ensemble perturbations calculated from the interpolated weights will maintain the property of zero mean as in (10).

In addition to the analysis grid arrangement shown in Figure 1, we also conducted assimilation experiments with even coarser analysis grids. In the following, we show the results of three weight-interpolation experiments that the LETKF analysis is computed every 3, 5 or 7 grid-points zonally and meridinally. In terms of the sparseness, the corresponding analysis coverage is 11%, 4% and 2% respectively (one local analysis is computed every  $3\times 3$ ,  $5\times 5$  and  $7\times 7$  horizontal grid-box). We will show in the next section that even though the number of total local analyses is substantially reduced, the analyses derived by interpolating the weights to full-resolution show little degradation.

Bowler (2006) indicated that the spatial consistency of the weighting coefficients is the advantage of the interpolation of the weights for performing analyses on a coarse grid. As an example, Figure 2(a) shows the weight coefficient derived from the full analysis for observational correction of the background ensemble mean. With the use of the symmetric square root matrix, the weighting coefficients derived at adjacent points are typically close to each other. This example is the contribution from the 4<sup>th</sup> ensemble member (an element of the  $\bar{\mathbf{w}}_a$  vector of maps) to the analysis mean at Day 48 00Z. The time and ensemble member for this example are arbitrarily chosen for illustration. The empty spots in Figure 2(a) are areas where there are no observations available in the corresponding local patch and the background ensemble is therefore not updated in these



regions. We note that there are some blockiness shown in Figure 2(a) due to the local analysis procedure. This does not generate sharp gradients in the analysis increments since the patterns of the weights are not the same among ensemble members at the same analysis time (Figure 4a shows the corresponding analysis increment for Figure 2a).

Figures 2(b)-(d) are the interpolated maps of this weighting coefficient with coarser analysis grids for the same ensemble member. Because the weights tend to be consistent at adjacent points, the interpolated weight structures can represent the original structures reasonably well. Although the structure of the interpolated weights is smoothed out with 2% analysis grid coverage (Figure 2d), the overall large-scale features are retained after interpolation. In section 4.1, we will give further discussion about representing the large-scale features from the weight-interpolation.

We now examine the weights applied to form the analysis ensemble perturbations, based on a diagonal element of  $\mathbf{W}_a$  and an off-diagonal element. Figures 2(e)-(h) show the maps of the weight coefficients of the first element in the first column of the matrix  $\mathbf{W}_a$ , representing the contribution from the first background ensemble perturbation to the first analysis ensemble perturbation. The values derived from the full analysis coverage are shown in Figure 2(e). Figure 2(f)-(h) are the same weights after interpolation on coarser analysis grids. Since  $\mathbf{W}_a$  is close to the identity matrix (as discussed in Section 3) the property gives the consistency between the analysis and background perturbations so that the main contribution for constructing an analysis ensemble perturbation is taken from its own background perturbation. Thus in Figure 2(e), the first ensemble forecast

perturbation has the most influence in determining the first analysis perturbation, with weights ranging from 0.7 to 1.0. Regions with lower values in these maps indicate that the contributions from other ensemble perturbations are important. The main features of the weight map derived from full analysis coverage are well recovered in the interpolated maps, except for the 2% case, in which the patterns are heavily smoothed.

As an example of the off-diagonal elements, Figures 2(i)-(l) show the maps of the weight coefficients of the fourth element in the first column of the matrix  $\mathbf{W}_a$ , representing the contribution from the fourth background ensemble perturbation to the first analysis ensemble perturbation. The large values in Figures 2(i)-(l) correspond to small values appearing in Figures 2(e)-(h). This again indicates that at this location, the other ensemble perturbations contribute more. Overall, their amplitude is much smaller compared to Figures 2(e)-(h), also because of the property that  $\mathbf{W}_a$  is close to the identity matrix. The main features shown in Figure 2(i) are well recovered in Figures 2(j) –(l) after interpolation.

Besides reducing the analysis computation, the interpolated weights also provide an additional benefit in obtaining reasonable weights for those local patches without available observations (the empty points in the interior domain of Figure 2 (a), (e) and (i)). For these points, the analysis ensemble is equal to the background ensemble in the original full analysis setting ( $\bar{\mathbf{w}}_a = (0,0,\dots,0)^T$  and  $\mathbf{W}_a = \mathbf{I}$ ). This interpolation provides a feasible way to update the background ensemble in the regions that have not been

observed, or that are under-observed compared to neighboring regions (more discussions in Section 4.2).

#### 4. Interpolated Analysis Results

In this section, we compare the LETKF sparse analyses constructed using weight interpolation with the analysis derived at full resolution. For comparison, we also conduct sparse analyses using interpolated analysis increments, after running the full-resolution LETKF for 20 days. The analysis increment is the difference between the ensemble mean of the analysis and background ( $\bar{\mathbf{x}}_f - \bar{\mathbf{x}}_a$ ), representing the correction made to the background ensemble mean. The increment interpolation is a traditional method used to convert a coarse resolution analysis into a full-resolution analysis. In this study, the increment interpolation uses the same interpolation method as used in the weight interpolation.

The time series of the root mean square error for the analysis ensemble mean of the potential vorticity is shown in Figure 3. The LETKF analyses derived from full resolution or weights interpolation from the sparse grids outperform the 3D-Var analysis derived at the high (full) resolution. The 3D-Var analysis provides a reference when a background error covariance consists of the statistically-averaged structures. Results show that the sparse analyses with the interpolated weights at different analysis coverage (blue lines) have a quality similar to or even slightly better than the one obtained at full resolution. This result suggests that the interpolated weights are very useful in retaining the quality

of the analysis. Furthermore, they are supported by a large reduction in their computational costs. Table 1 summarizes the average computing time required for finishing one analysis cycle for different analysis grid coverage, including the time for looping for all the analysis grids and for interpolating the weights. Results show that the total computational time required for total local analyses can be reduced if the LETKF is not performed at full resolution. However, with the 11% analysis-grid coverage, the time spent on interpolation makes the overall computation cost slightly higher than the original full resolution. The computational time required for interpolation is reduced as the analysis coverage decreases below 4%, almost having half of computational time used to perform the full resolution LETKF.

The analysis accuracy from using interpolated weights becomes worse only when the analysis grid is so coarse (2% or less) that the local patches of each analysis do not overlap. The sparse analyses using interpolated increments (shown as the blue lines in Figure 3b) are more sensitive to the sparseness of the analysis grid-point and have much lower analysis accuracy than the ones derived from the interpolated weights.

The analysis accuracy from the interpolated increments remains satisfactory only when using a high analysis coverage of 50%, where the analysis grid-point is arranged as a staggered grid available every other grid-point. Once the coverage decreases, the accuracy degrades quickly. With 25% analysis coverage, the analysis increments are smoothed out and stretch isotropically and thus the analysis has accuracy similar to the 3D-Var analysis. With a lower analysis coverage of 11%, the LETKF with the

interpolated increments diverges (the dashed blue line in Figure 3b). These results show that interpolating the analysis increments leads to a serious degradation of the analysis, whereas interpolating the weights retains the analysis accuracy and the advantages of the LETKF with the efficiency of low-resolution analysis.

The results shown in Figure 3 can be further understood from the structures of the analysis increments obtained from the sparse analyses constructed with interpolated weights. The analysis increments, as shown for potential vorticity at the mid-level in Figure 4, represent forecast errors stretched by the flow, and they have elongated structures and scales similar to that of bred vectors (Corazza et al. 2002).

We found that the analysis increments obtained from interpolated weights at different analysis coverage (11%, 4% and 2%) are very similar, as shown in an example in Figure 4(a)-(d). It is clear that interpolating the weights succeeds in recovering quite well the full analysis increments and the obtained analysis increments are insensitive to the analysis coverage. This is because the weights vary on a larger scale than the analysis increments (this is discussed further in Section 4.1). By contrast, if the interpolation is done on the analysis increments, as in Figures 4(e)-(h), the structures of the analysis increments are quickly smoothed out as the analysis grid becomes coarser. Only with a high analysis-coverage of 50% can the local characteristics in the analysis increment at full resolution be retained and thus maintain an advantage over 3D-Var. In Figure 4(h), the pattern of interpolated analysis increment with 2% analysis coverage has an unrealistic large-scale feature. As could be expected,

this will impose false corrections to the background state and lose the advantage of using the time-dependent error statistics in the LETKF. Thus, the analysis accuracy will be quickly degraded once these increment-interpolated analyses are recycled in a continuous analysis cycles).

#### 4.1 Spectral analysis

The impact of the different interpolation schemes can be illustrated through the spectral analysis of the distribution of errors and analysis increments, shown for potential temperature on the bottom level. This variable is chosen to show because, in this QG model, the perturbation growths at the boundaries account most of the variances in wind, temperature or streamfunction (Snyder et al. 2003).

Figure 5 shows the time-averaged power spectra for the weights,  $\bar{w}_a$  used to update the background ensemble mean, the analysis increment and the forecast and analysis errors obtained from the full resolution LETKF. The time averaged spectrum of the true potential temperature is also included in each figure 5 in grey. The spectrum of the true state shows that the model solution varies on a large scale, with most of its power in global wavenumbers smaller than 15 and little amplitude in the small scales (global wavenumbers larger than 50). Comparing the spectral slopes of the weights and the analysis increments (Figures 5a, b) shows that the weights derived from LETKF vary on much larger scales than the analysis increments. From Figure 5(c), the analysis increments correct background errors for the features of scales smaller than the wavenumber 50, but the analysis errors in the short waves are an order of magnitude

larger than the forecast errors. This indicates that the analysis increment from the full-resolution LETKF introduces unnecessary error structures at small scales, even though their amplitudes are insignificant compared to the large scale. Also, such structures are quickly damped being evolved by the model.

Figure 6 shows the corresponding results of the spectral analysis for weight and increment interpolations. For weight interpolation (Figure 6a), the amplitudes of the large and medium scales in the interpolated weights are similar to those of the full resolution analysis. However, the spectral amplitudes of the small-scale structures are reduced by an order of magnitude as the analysis-grid coverage is reduced. These features are also reflected on the analysis increments created by using these weights (the blue lines in Figure 6b). For the analysis increments derived from the weight-interpolation, the spectral amplitudes of the larger scales are not sensitive to the analysis-grid coverage while the amplitudes of the small scales decrease as the coverage reduces. Therefore, the small-scale error structures appearing in the full resolution analysis, are avoided. By contrast, when using increment interpolation for the 50% analysis-grid coverage, the small-scale structures of the increments are amplified although the large-scale features remain similar to the full resolution analysis. The small-scale errors are largely increased by increment interpolation because the analysis grids cannot resolve the small-scale features of the increments. Since these small-scale structures have relatively small amplitudes, with a 50% grid coverage this does not affect the overall analysis accuracy significantly. However, as the coverage reduces to 25%, the structures of the increments are smoothed out (the red dashed line in Figure 6b). This degrades the analysis accuracy

at all scales, as indicated by the larger analysis error (the red dashed line in Figure 6c). This also indicates that the increment interpolation is more sensitive to the analysis grid coverage.

## **4.2 Impact on the data-void region**

In the standard LETKF, the local analysis for the region without observations is directly taken from the corresponding background states. Therefore, the forecast skill of such region can only be improved from dynamical propagations through model integration. The weight-interpolation method could have an advantage at these regions by providing values for the weight coefficients to construct the analysis. To investigate this, 14 observations were removed from the original 128-observation distribution. This create a data-void region in the centre of the model domain with a size of  $17 \times 9$  grid-points.

Figure 7 compares the RMS analysis errors in the data-void region for the full and the weight-interpolated LETKF. After removing 14 observations, the analysis accuracy from derived with the full-resolution is much degraded (the black line) compared to the one derived from 128 observations (the grey line). With the same 114 observations, the weight-interpolated analyses with either 11% or 2% analysis grid coverage are both better than the full-resolution one, especially at the times with large errors. Averagely, the weight interpolation can reduce as much as 18% of the analysis errors in this region.



The impact can be illustrated by regarding the corresponding analysis increments. Figure 8(a) shows the increment structures of potential vorticity at the 3<sup>rd</sup> level required to correct the background ensemble mean if all observations are available in the full-resolution LETKF. With the existence of the data-void region, the number of locally available observations is greatly reduced since observation distribution is not regular. This even affects the local analysis outside but near the data-void region, increasing the amount of zero corrections outside the data-void region, as shown in Figure 8(b). This degrades the analysis accuracy near this region. In contrast, the analysis increments derived from the weight interpolation are able to provide useful corrections for the background (Figure 8c,d). In this example, the increments in the northeast part of this region are retrieved with either the 11% or the 2% analysis-grid coverage. Our results confirm that for the under-observed region the analysis accuracy can benefit from the weight-interpolation method.

## 5. Summary

In this study, we investigated an efficient method to reduce the analysis computational cost within the framework of the LETKF following a suggestion by Bowler (2006). The LETKF analyses are computed on coarse grids, but the weights used to update the background ensemble are interpolated onto the high-resolution grids. Instead of repeatedly using the observations and the background ensemble to perform the LETKF, the analysis at the high-resolution is derived through estimating the interpolated weights. In the LETKF, the weights of the analysis ensemble represent two sources of information: one is associated with the observations contribution to the mean background

state and the other is associated with the dynamically evolving error structures obtained from background ensemble perturbations. Interpolations are done separately on these two components of the weights by taking advantage of the symmetric square root solution of the transform matrix used in the LETKF.

We showed that the weights derived from the LETKF vary on larger scales than the analysis increment and they are consistent among nearby points. The interpolated weight maps for a coarse analysis-grid can represent evolving features very well. For the same reason, there is little degradation in the quality of the weight-interpolated analyses compared to the analyses derived with the high-resolution grid. Therefore, the corresponding analysis accuracy is still high even when the analysis grid is reduced to just 2% of the grid-points of the full resolution grid. The results are insensitive to the analysis coverage in this study.

For comparison, we show that the weight-interpolated analysis are much more accurate than the analysis constructed from interpolating the analysis increments. Also, the analysis quality from the weight-interpolation is insensitive to the sparseness of the analysis-grid while the quality from the increment-interpolation is much more sensitive to the analysis grid coverage due to the stretching of dynamical scales.

The analysis accuracies by interpolating the weights or the increments are related to their corresponding spectral characteristics. Spectral analysis shows that the weights vary on a

larger scale than the analysis increment. For the weights, the spectral power at larger scales is not sensitive to the analysis grid converge but the amplitudes of short waves are damped after interpolation. Given that the background errors have very little amplitude in on small-scales, the analysis accuracy can be well retained by the weight-interpolation. This implies that the weight-interpolation may bring benefits on damping the unwanted small scales features such as the gravity waves. For increment-interpolation the spectral power is very sensitive to the sparseness of the analysis grids. Not only the amplitudes of the small scales of the increments are amplified with interpolation and the analysis errors are degraded at all scales by reducing the analysis grid coverage.

Besides performing sparse analyses, the weight-interpolation can also be used to provide analysis weights for regions without local observations, so that instead of returning the background ensemble values without updating them, as in the original LETKF procedure, the analysis ensemble can still be computed for these under-observed regions. This advantage of smoothing the weights in handling under-observed regions and the fact that the spectral power of the weights is dominated by the large scales explains why the interpolated weight results are not just comparable but even slightly better than the full resolution analysis. As a result, smoothing the weights may be advantageous even when using the full resolution analysis grid.

For each local analysis, the analysis ensemble is the linear combination of the background ensemble and any conserved quantity that is a linear function of the model state will be equally conserved in the original LETKF analysis. With the method of

weight interpolation, the linear property of  $\mathbf{W}_a \mathbf{v} = 0$  is well maintained at the non-analysis grid point. This can keep the analysis ensemble centered at the mean state ( $\mathbf{X}_a \mathbf{v} = 0$ ). Additionally, the analysis remains in the subspace of the forecast ensemble, so that properties such as conservation of total mass and balance, satisfied by each ensemble member, are also satisfied in the analysis.

As the ensemble-based data assimilation becomes feasible for operational use (Houtekamer et al. 2005, Keppenne et al. 2002, Miyoshi et al. 2008 and Whitaker et al. 2008), this study provides guideline for performing sparse analyses with interpolated weights. We note that the results obtained in this study may be over-optimistic since the quasi-geostrophic model used in the experiments was only slightly imperfect.

## 6. References

- Akima, H. (1978). A Method of Bivariate Interpolation and Smooth Surface Fitting for Irregularly Distributed Data Points. *ACM Transactions on Mathematical Software*, **4**, 148-164.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420-436.
- Bowler, N. 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus A*, **58**, 538-548.
- Courtier, P., J. N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational

implementation of 4DVAR, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**,1367-1387.

Corazza, M., E. Kalnay, D. J. Patil, E. Ott, J. Yorke, I. Szunyogh, M. Cai, 2002: Use of the breeding technique in the estimation of the background error covariance matrix for a quasigeostrophic model. AMS Symposium on Observations, Data Assimilation and Probabilistic Prediction, Orland, Florida, 154-157.

-----, E. Kalnay, and S.-C. Yang, 2007: An implementation of the Local Ensemble Kalman Filter for a simple quasi-geostrophic model: Results and comparison with a 3D-Var data assimilation system. *Nonlinear Processes in Geophysics*, **14**, 89-101.

Harlim, H., 2006: Errors in the initial conditions for numerical weather prediction: A study of error growth patterns and error reduction with ensemble filtering. *PhD dissertation*, University of Maryland, 86 pages.

Houtekamer, P. L. , H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. *Mon. Wea. Rev.*,133,604–620.

Hunt, B., E. Kostelich, I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: a Local Ensemble Transform Kalman Filter. *Physica D*, 230, **112-126**.

Lorenz, E. N., 1995: Predictability: a problem partly solved. In: *In Proceedings of the seminar on predictability*. Volume I, ECMWF, Reading, Berkshire, UK., pp. 1–18.

Lorenz, E. N. and Emanuel, K. A., 1998: Optimal sites for supplementary weather observations: simulation with a small model. *J. Atmos. Sci.* **55**, 399–414.

Kalnay, E. and S.-C. Yang, 2008: Accelerating the spin-up of Ensemble Kalman Filtering.

*Quart. J. Roy. Meteor. Soc.*,(submitted)

Keppenne, C.L., and M.M. Rienecker, 2002: Development and initial testing of a parallel

Ensemble Kalman filter for the Poseidon isopycnal ocean general circulation model.

*Mon. Wea. Rev.*, **130**, 2951-2965.

-----, M. M. Rienecker, J. P. Jacob and R. Kovach, 2007: Error covariance

modeling in the GMAO ocean ensemble Kalman filter. *Mon. Wea. Rev.*, (in press).

Miyoshi, T., 2005: Ensemble Kalman filter experiments with a Primitive-Equation global

model. Doctoral dissertation, University of Maryland, College Park, 197pp. Available

at <http://hdl.handle.net/1903/3046>.

----- and S. Yamane, 2008: Local ensemble transform Kalman filtering with an

AGCM at a T159/L48 resolution. *Mon. Wea. Rev.*, **135**, 3841-3861.

Morss, R. E. 1998: Adaptive observations: Idealized sampling strategies for improving

numerical weather prediction. PhD thesis, Massachusetts Institute of Technology.

225 pp.

Parrish, D. and J. Derber, 1992: The National Meteorology Center's spectral

statisticalinterpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D.

J. Patil, and J. A. Yorke, 2004: A local ensemble Kalman filter for atmospheric data  
assimilation. *Tellus A*, **56**, 415-428.

Rotunno, R. and Bao, J. W., 1996: A case study of cyclogenesis using a model

hierarchy. *Mon. Wea. Rev.*, **124**, 1051-1066.

Wang, X., and S. J. Julier, 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. and T. M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations. *Mon. Wea. Rev.*, **130**, 1913-1924.

-----, T. M. Hamill, X. Wei, Y. Song and Z. Toth, 2007: Ensemble Data Assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, (in press).

Yang, S-C., M. Corazza, A. Carrassi, E. Kalnay and T. Miyoshi, 2008: Comparisons of ensemble-based and variational-based data assimilation schemes in a Quasi-Geostrophic model. *Mon. Wea. Rev.* (accepted).

## **7. Acknowledgements**

We are grateful for helpful interactions with Jeff Whitaker, Istvan Szunyogh, Eric Kostelich, and Milija Zupanski. Rebecca Morss and Matteo Corazza provided the original 3D-Var code. S.-C. Yang was supported by NASA grants NNG004GK78A and NNG06GB77G.

**Table caption**

**Table 1** Computational time required for one analysis cycle for different analysis-grid coverage.



## Figure Captions

**Figure 1** The grid arrangement for an 11% analysis-grid coverage. The dots are the grids where the LETKF analysis is actually performed. The crosses indicate high-resolution grid points whose analysis will be derived by weight-interpolation or increment interpolation (see the explanation in the text).

**Figure 2** (a) The weighting coefficients corresponding to the fourth element of the weighting vector ( $\bar{\mathbf{W}}^a$ ) for updating the background ensemble mean, derived with 100% analysis grid-point coverage (full resolution); (b)-(d) the same as (a), except that the weight is interpolated at the 11%, 4% and 2% analysis grid coverage. (e)-(h) are the first element of the first column of the weighting matrix ( $\mathbf{W}^a$ ) for constructing the first analysis ensemble perturbation derived at the same resolutions used in (a)-(d). (i)-(l) are the same as (e)-(h) except for the fourth element of the first column of  $\mathbf{W}^a$ . In (a), (e) and (i), the empty spots denote the local regions with no available observations: they are zero for (a) and (i) and one for (e).

**Figure 3** (a) The time series of the RMS analysis error in terms of the potential vorticity from different DA experiments. The LETKF analysis from the full-resolution is denoted as the black line and the 3D-Var derived at the same resolution is denoted as the grey line. The LETKF analyses derived from weight-interpolation with different analysis coverage are indicated with blue lines, (b) the same as (a), except for the blue lines, they are the LETKF analyses derived after the first 20 days from increment-interpolation with different analysis coverage.

**Figure 4** The analysis increment for potential vorticity at the 3rd level at Day 48 00Z from the LETKF with (a) a full resolution, (b) obtained through interpolated weights with 11% analysis coverage, (c) the same as (b) except for 4% analysis coverage, (d) the same as (b), except for 2% analysis-grid coverage. (e)-(h) are the interpolated analysis increments computed taking 50%, 11%, 4% and 2% analysis coverage of analysis increment obtained at the full-resolution and interpolating back to the full-resolution.

**Figure 5** (a) Power spectra of weights derived from the full-resolution LETKF, (b) same as (a) but for analysis increment for potential temperature at the bottom level, and (c) same as (b) but for analysis and forecast errors. Power spectra of the true potential temperature at the bottom level is overplotted with the grey line, adjusted by a constant factor of  $10^{-4}$ .

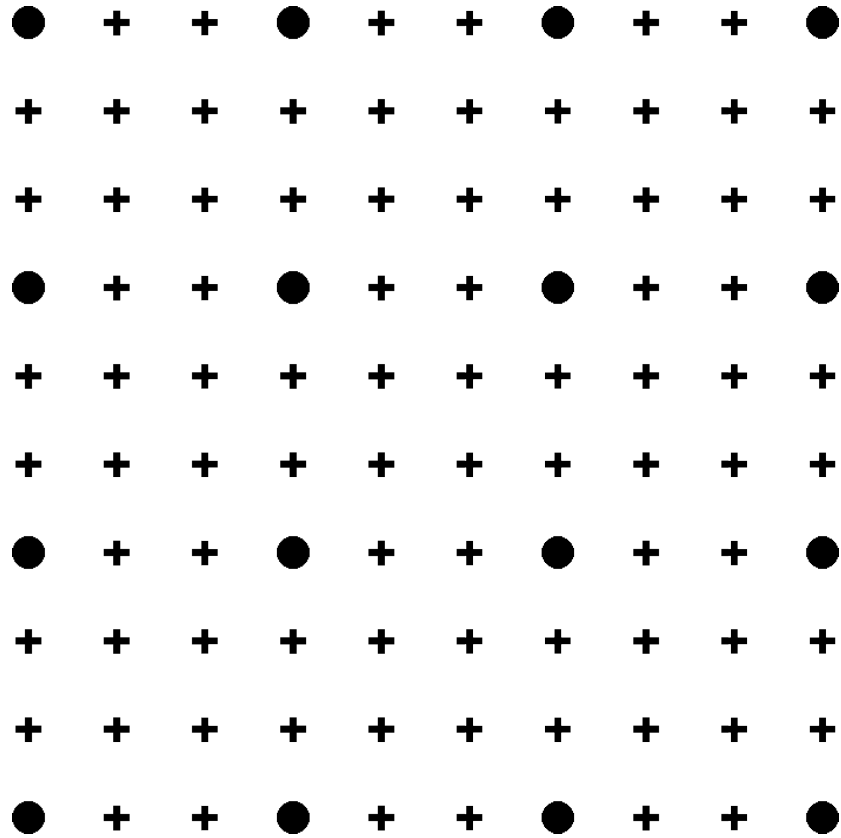
**Figure 6** (a) Power spectra of weights derived from the LETKF with full resolution and interpolated weights with different analysis grid coverage, (b) same as (a) but for analysis increment for potential temperature at the bottom level and the ones obtained from interpolated increment, and (c) same as (b) but for analysis errors.

**Figure 7** The time series of the RMS analysis error in terms of the potential vorticity from different DA experiments for the domain that the x-grid points are between 23 and 41 and the y grid points are between 12 and 22 (the data-void region defined in the text). The LETKF analysis from the full-resolution with 114 observations is denoted as the black line. The analyses derived from weight-interpolated LETKF with 11% and 2% analysis coverage are indicated with blue lines. The analysis derived with 128 observations from full-resolution LETKF is denoted as the grey line.

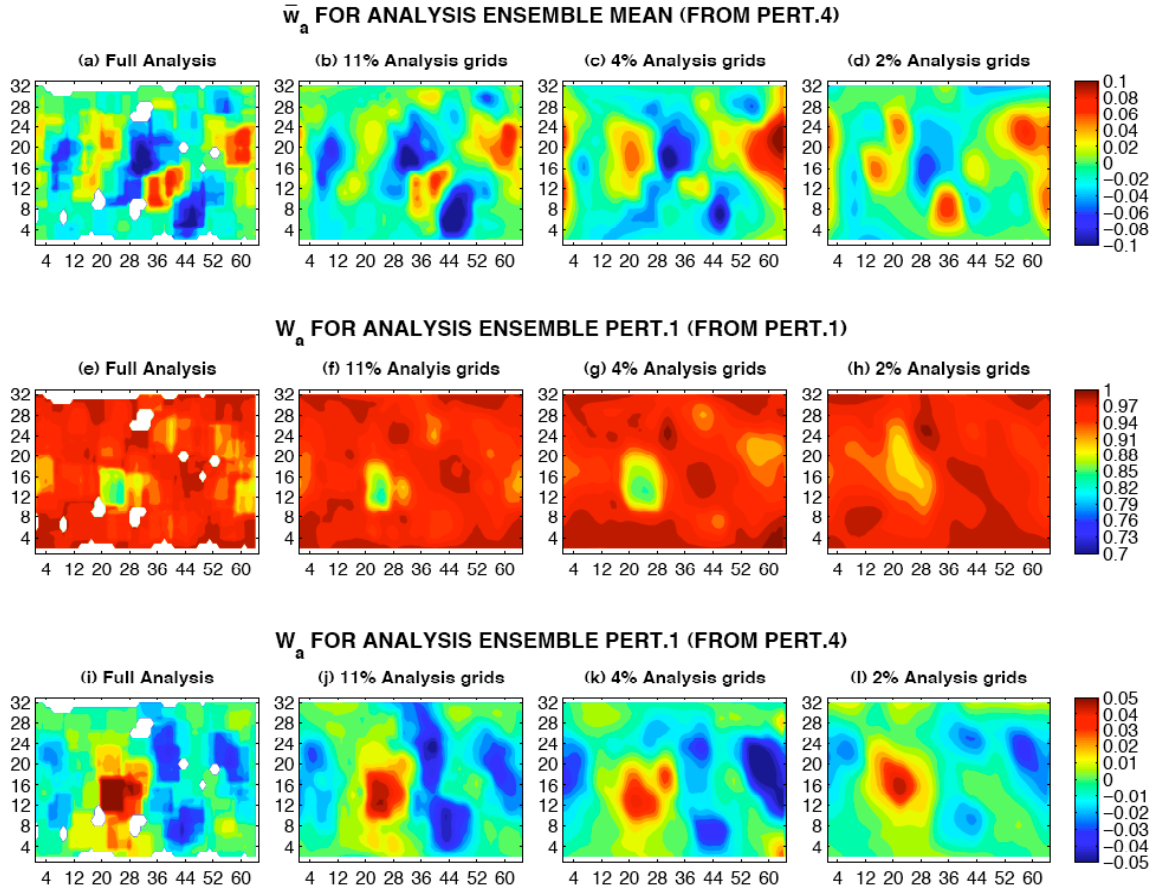
**Figure 8** The analysis increment for potential vorticity at the 3<sup>rd</sup> level at Day 77 from the LETKF with (a) full resolution and 128-observation network, (b) full resolution and 114-observation network that has a data-void region indicated by the dashed box (c) obtained through interpolated weights with 11% analysis coverage, (c) the same as (b), except for 2% analysis-grid coverage.

**Table 2** Computational time required for one analysis cycle for different analysis-grid coverage

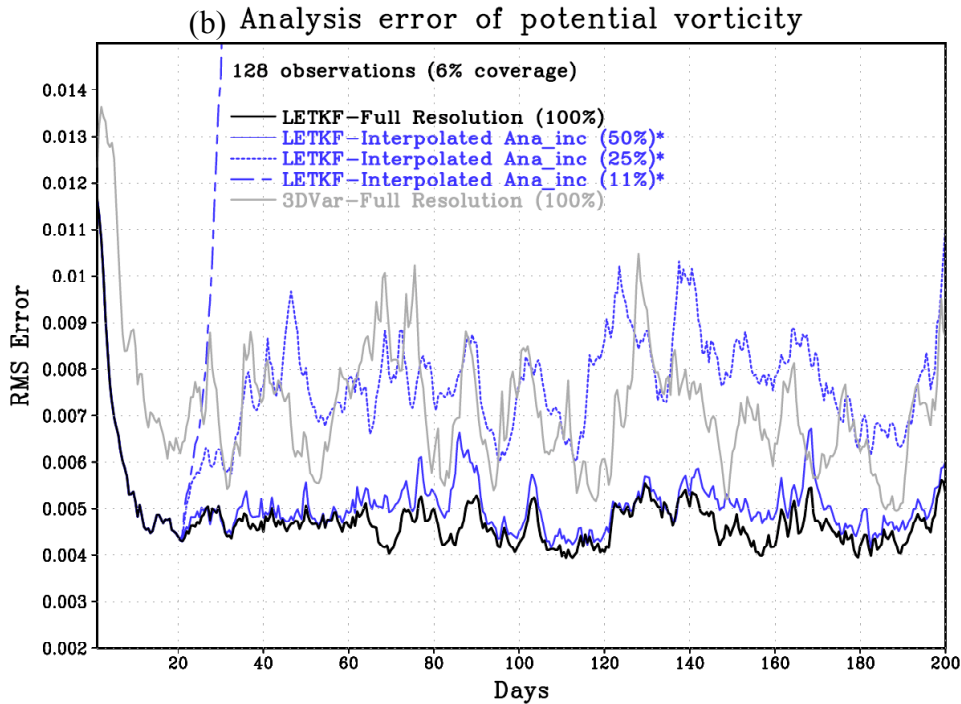
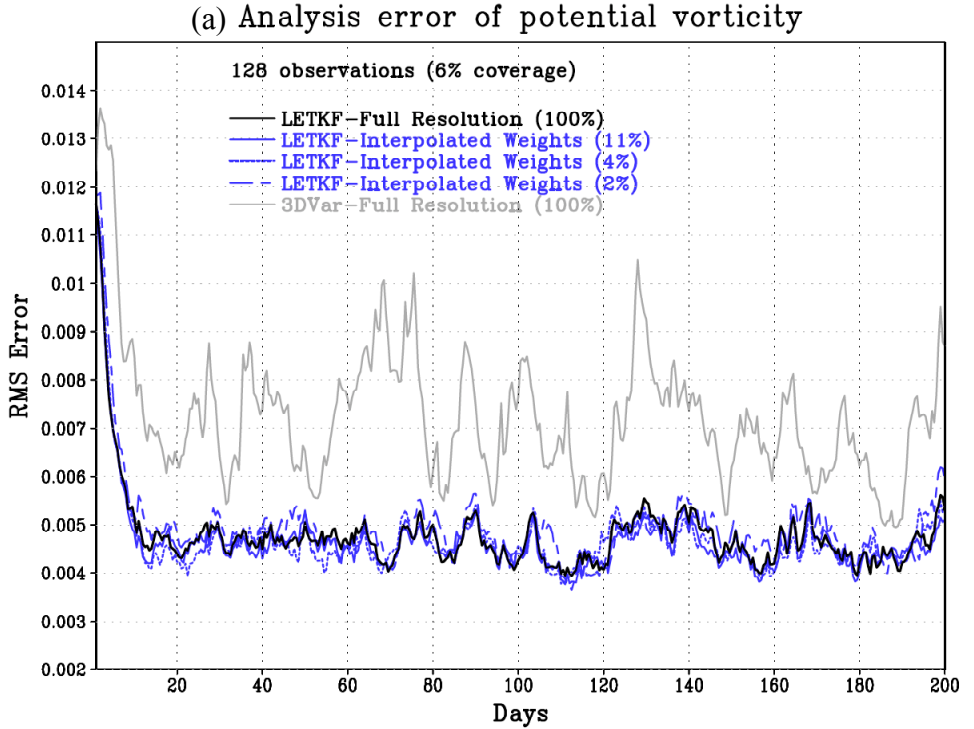
Analysis grid coverage \ Computing time (sec)	100%	11%	4%	2%
Total	5.1	6.7	2.9	2.1
local analysis	4.6	0.5	0.2	0.1
interpolation		4.9	1.6	1.0



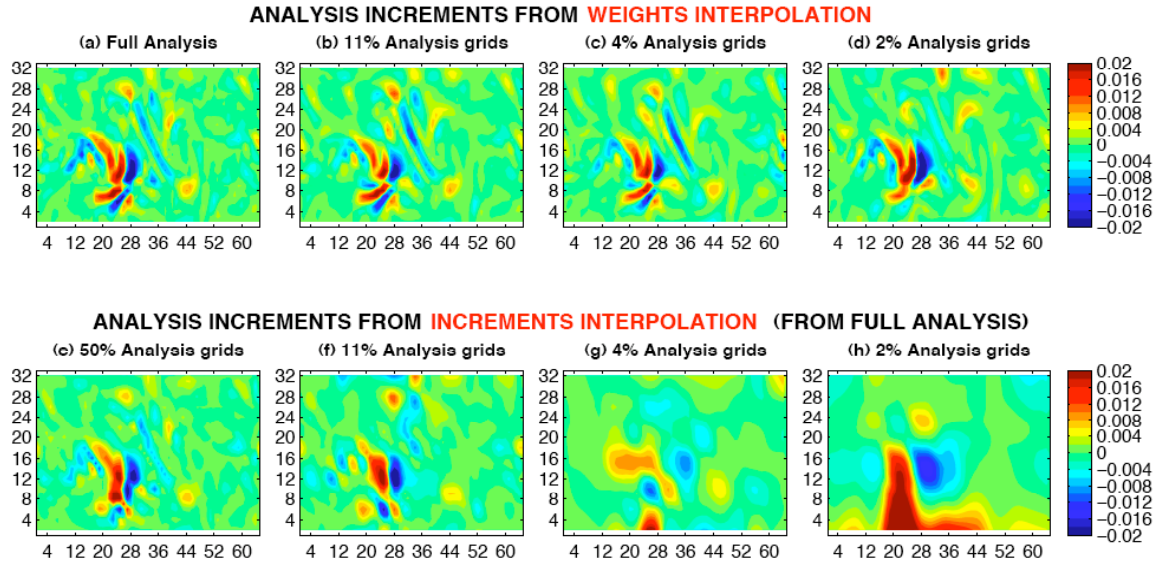
**Figure 1** The grid arrangement for an 11% analysis-grid coverage. The dots are the grids where the LETKF analysis is actually performed. The crosses indicate high-resolution grid points whose analysis will be derived by weight-interpolation or increment interpolation (see the explanation in the text).



**Figure 2** (a) The weighting coefficients corresponding to the fourth element of the weighting vector ( $\bar{w}_a$ ) for updating the background ensemble mean, derived with 100% analysis grid-point coverage (full resolution); (b)-(d) the same as (a), except that the weight is interpolated at the 11%, 4% and 2% analysis grid coverage. (e)-(h) are the first element of the first column of the weighting matrix ( $W_a$ ) for constructing the first analysis ensemble perturbation derived at the same resolutions used in (a)-(d). (i)-(l) are the same as (e)-(h) except for the fourth element of the first column of  $W_a$ . In (a), (e) and (i), the empty spots denote the local regions with no available observations: they are zero for (a) and (i) and one for (e).

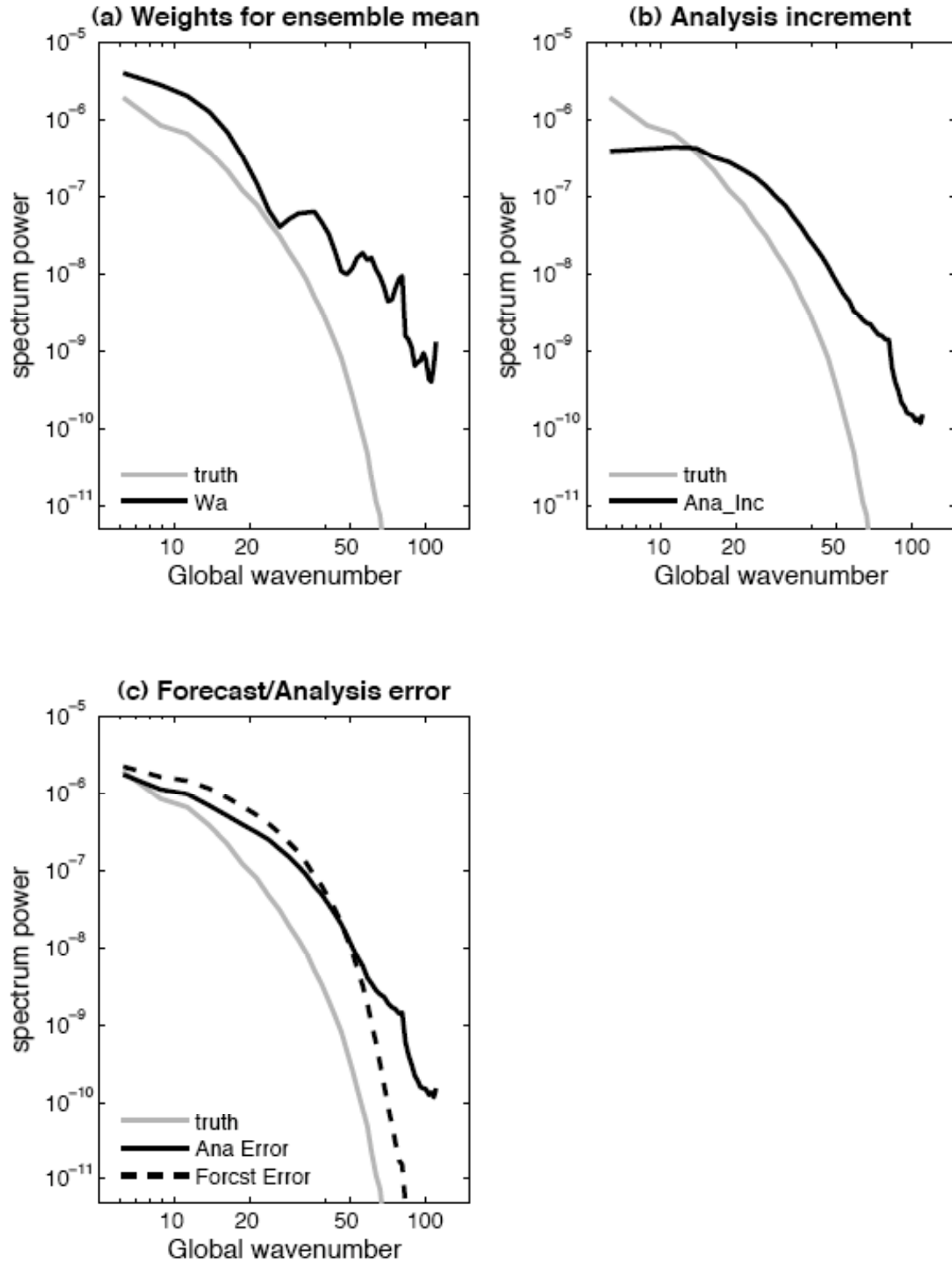


**Figure 3 (a)** The time series of the RMS analysis error in terms of the potential vorticity from different DA experiments. The LETKF analysis from the full-resolution is denoted as the black line and the 3D-Var derived at the same resolution is denoted as the grey line. The LETKF analyses derived from weight-interpolation with different analysis coverage are indicated with blue lines, (b) the same as (a), except for the blue lines, they are the LETKF analyses derived after the first 20 days from increment-interpolation with different analysis coverage.

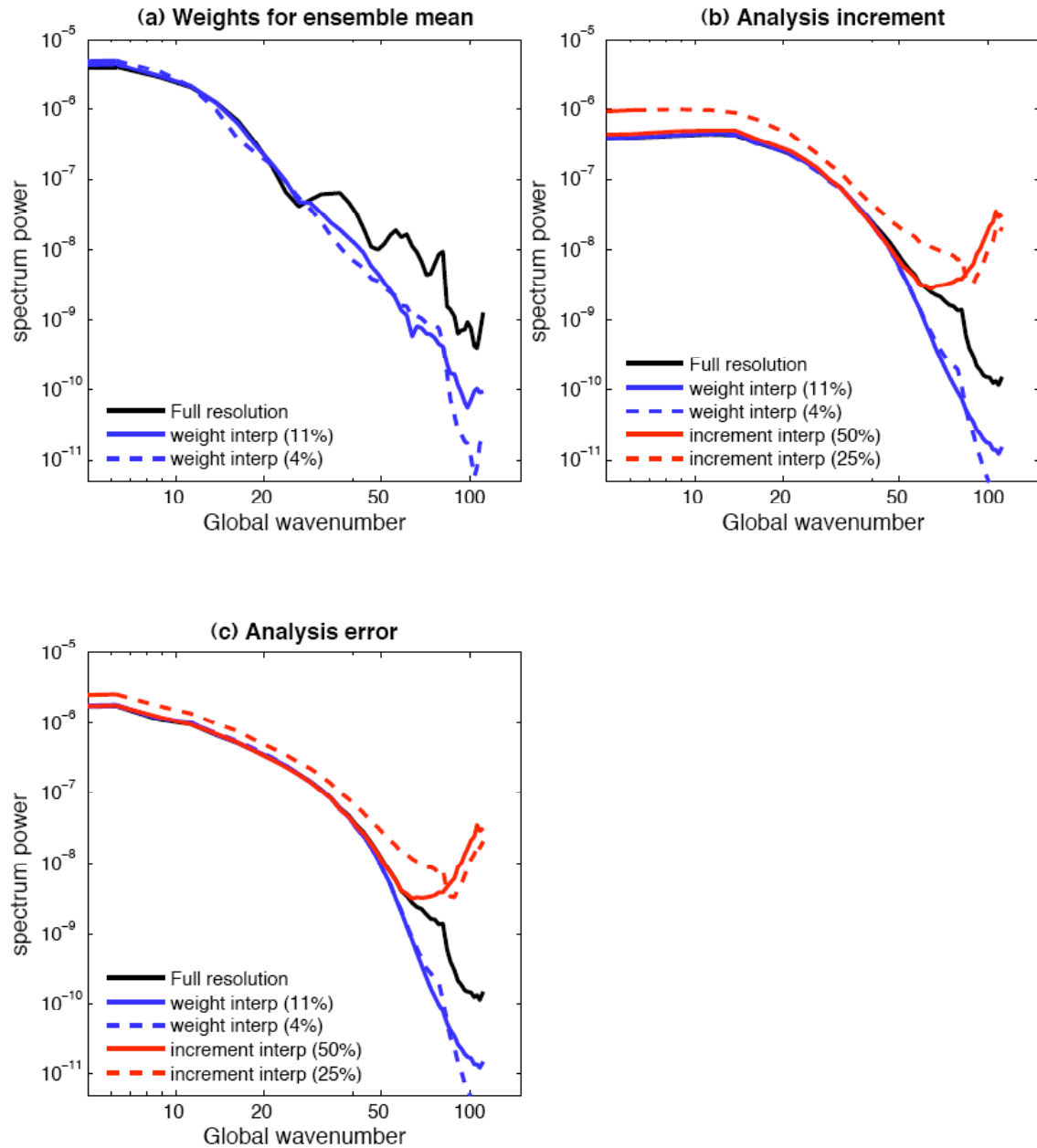


**Figure 4** The analysis increment for potential vorticity at the 3<sup>rd</sup> level at Day 48 00Z from the LETKF with (a) a full resolution, (b) obtained through interpolated weights with 11% analysis coverage, (c) the same as (b) except for 4% analysis coverage, (d) the same as (b), except for 2% analysis-grid coverage. (e)-(h) are the interpolated analysis increments computed by taking 50%, 11%, 4% and 2% analysis coverage of analysis increment obtained at the full-resolution and interpolating back to the full-resolution.

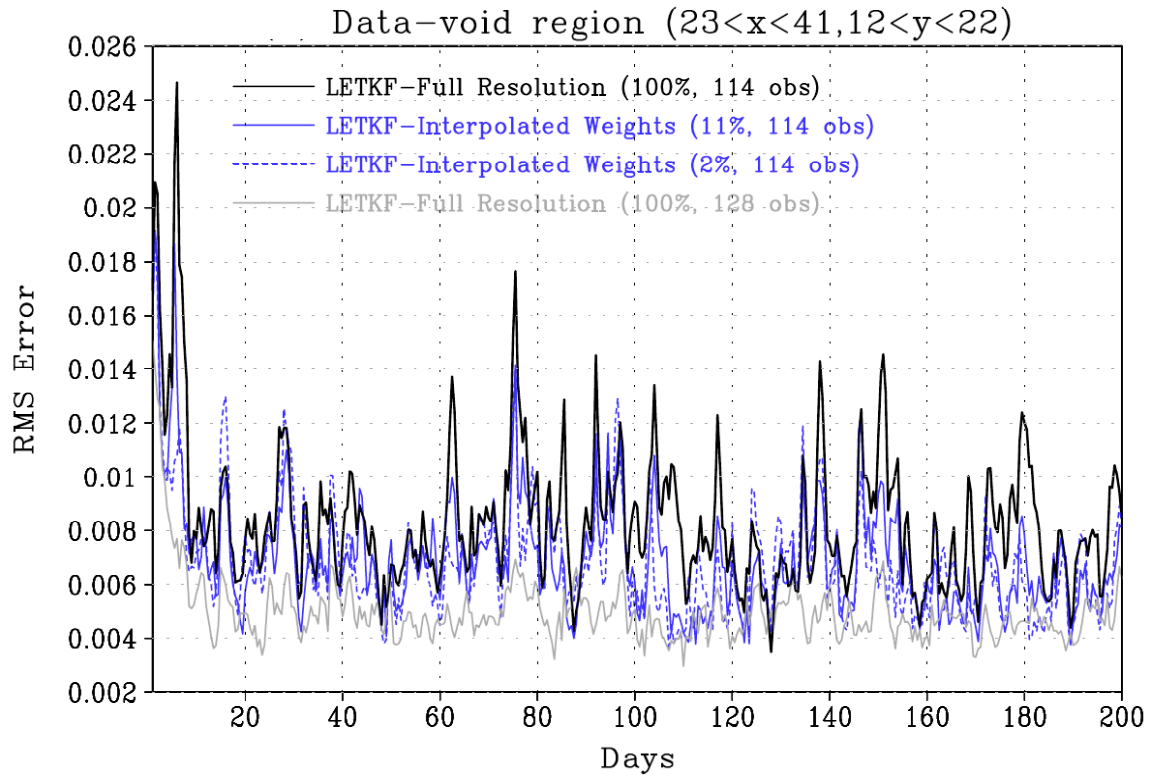




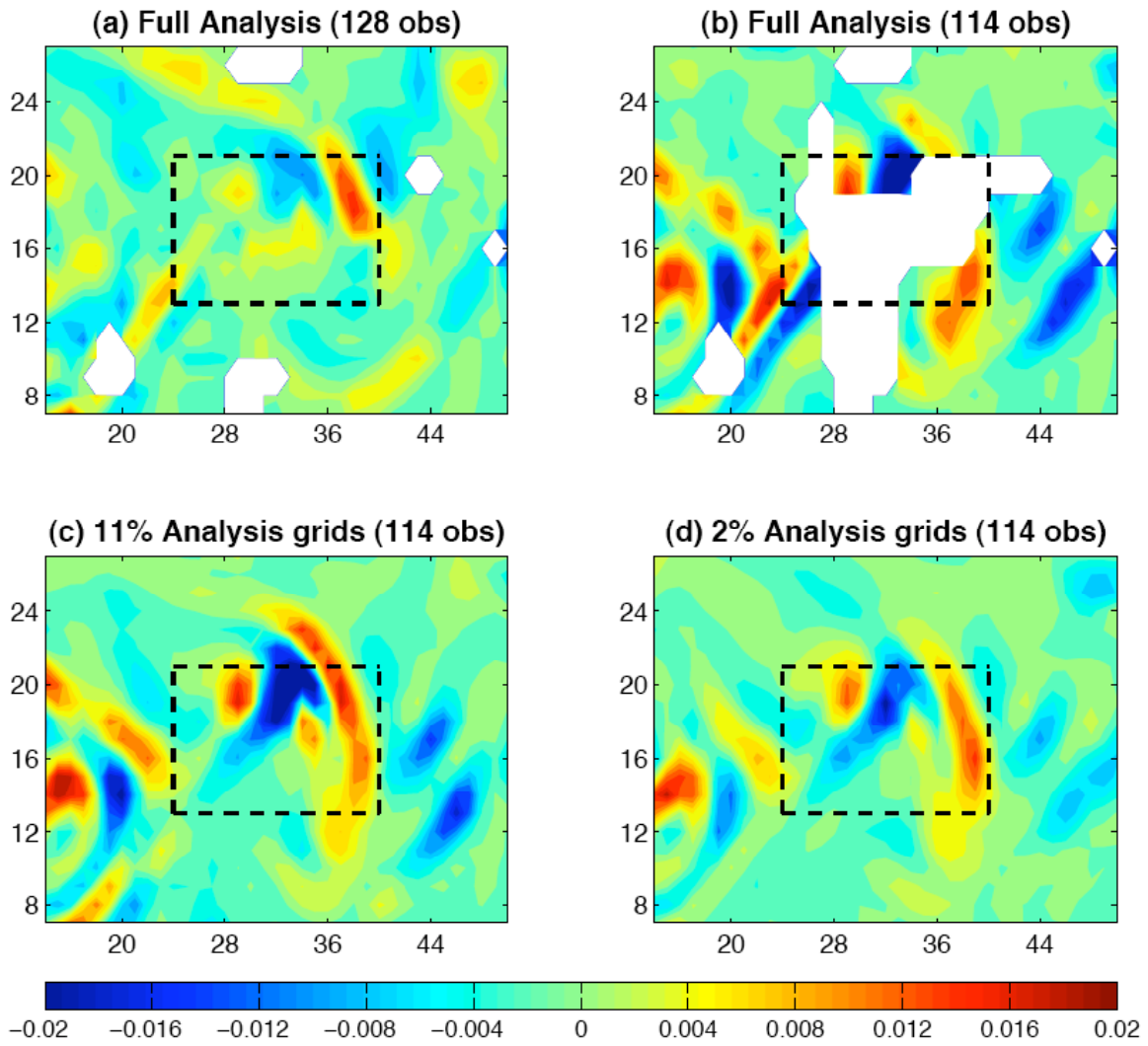
**Figure 5** (a) Power spectra of weights derived from the full-resolution LETKF, (b) same as (a) but for analysis increment for potential temperature at the bottom level, and (c) same as (b) but for analysis and forecast errors. Power spectra of the true potential temperature at the bottom level is overplotted with the grey line, adjusted by a constant factor of  $10^{-4}$ .



**Figure 6** (a) Power spectra of weights derived from the LETKF with full resolution and interpolated weights with different analysis grid coverage, (b) same as (a) but for analysis increment for potential temperature at the bottom level and the ones obtained from interpolated increment, and (c) same as (b) but for analysis errors.



**Figure 7** The time series of the RMS analysis error in terms of the potential vorticity from different DA experiments for the domain that the x-grid points are between 23 and 41 and the y grid points are between 12 and 22 (the data-void region defined in the text). The LETKF analysis from the full-resolution with 114 observations is denoted as the black line. The analyses derived from weight-interpolated LETKF with 11% and 2% analysis coverage are indicated with blue lines. The analysis derived with 128 observations from full-resolution LETKF is denoted as the grey line.



**Figure 8** The analysis increment for potential vorticity at the 3<sup>rd</sup> level at Day 77 from the LETKF with (a) full resolution and 128-observation network, (b) full resolution and 114-observation network that has a data-void region indicated by the dashed box (c) obtained through interpolated weights with 11% analysis coverage, (d) the same as (b), except for 2% analysis-grid coverage.