AMERICAN METEOROLOGICAL SOCIETY

Monthly Weather Review

EARLY ONLINE RELEASE

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/2010MWR3328.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

Balance and Ensemble Kalman Filter Localization Techniques

Steven J. Greybush [1]

Eugenia Kalnay [1,2,3]

Takemasa Miyoshi [1]

Kayo Ide [1,2,3,4]

Brian R. Hunt [3,5]

[1] Department of Atmospheric and Oceanic Science
[2] Earth System Science Interdisciplinary Center
[3] Institute for Physical Science and Technology
[4] Center for Scientific Computation and Mathematical Modeling
[5] Department of Mathematics

University of Maryland, College Park, Maryland

_____

Corresponding author address: Steven Greybush, Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742, USA.
E-mail: greybush@atmos.umd.edu

Abstract

In Ensemble Kalman Filter data assimilation, localization modifies the error covariance matrices to suppress the influence of distant observations, removing spurious long distance correlations. In addition to allowing efficient parallel implementation, this takes advantage of the atmosphere's lower dimensionality in local regions. There are two primary methods for localization. In B-localization, the background error covariance matrix elements are reduced by a Schur product so that correlations between grid points that are far apart are removed. In R-localization, the observation error covariance matrix is multiplied by a distance-dependent function, so that far away observations are considered to have infinite error. Successful numerical weather prediction depends upon well-balanced initial conditions to avoid spurious propagation of inertial-gravity waves. Previous studies note that B-localization can disrupt the relationship between the height gradient and the wind speed of the analysis increments, resulting in an analysis that can be significantly ageostrophic.

This study begins with a comparison of the accuracy and geostrophic balance of EnKF analyses using no localization, B-localization, and R-localization with simple one-dimensional balanced waves derived from the shallow water equations, indicating that the optimal length scale for R-localization is shorter than for B-localization, and that for the same length scale R-localization is more balanced. The comparison of localization techniques is then expanded to the SPEEDY global atmospheric model. Here, natural imbalance of the slow manifold must be contrasted with undesired imbalance introduced by data assimilation. Performance of the two techniques is comparable, also with a shorter optimal localization distance for R-localization than for B-localization.

1. Introduction

The Ensemble Kalman Filter (EnKF) (Evensen, 1994) is a Monte-Carlo approximation to the traditional filter of Kalman (1960) that is suitable for high-dimensional problems such as Numerical Weather Prediction (NWP). One of the strengths of Ensemble Kalman Filters is the ability to evolve in time estimates of forecast error covariance, using the flow-dependent information inherent in an ensemble of model runs.

Localization is a technique by which the impact of observations from distant regions upon an analysis is suppressed. There are two categories of localization techniques (discussed in detail in section 2b): those that operate on background error covariances B, which we call B-localization, and those that operate on observation error covariances R, which we call R-localization. Adaptive localization techniques, such as the hierarchical filter of Anderson (2007) and ECO-RAP of Bishop and Hodyss (2009a, 2009b), are beyond the scope of this work.

It is the error covariances between model variables, along with the observation error characteristics, that ultimately describe the impact pattern of an observation upon the analysis via the Kalman gain K. In practice, the accuracy of the background error covariance estimate is limited by the size of the ensemble, which must be kept small for computational feasibility (typically of order 20-100 for NWP). Empirically, at larger geographical distances background error covariance estimates tend to be dominated by noise rather than signal (Hamill et al., 2001); it is this "distance-dependent assumption"

that motivates the technique of (non-adaptive) localization to eliminate correlations that are deemed to be spurious.

The background error covariance determined from an ensemble of P members has at most $P-1$ degrees of freedom to express uncertainty. However, in local regions of large error growth the atmosphere has been shown to exhibit low dimensionality (Patil et al., 2001). When using localization, the ensemble needs to account for the instabilities in a local region. Additionally, if local analyses can choose different linear combinations of ensemble members in different regions, this allows the analysis to greatly reduce the previously noted dimensionality limitation (Hunt et al., 2007). Lorenc (2003) notes that the assimilation of a perfect observation removes a degree of freedom from the ensemble, but that localization with a Schur product allows for extra degrees of freedom in the analysis.

Localization can also lead to significant savings in computational resources. The analysis at each grid point only needs to consider local observations and the values at nearby model grid points that are linked to these observations by the observation operator. Analyses for local regions can thus be considered independently, allowing for more efficient parallelization of the code (Hunt et al., 2007; Szunyogh et al., 2008).

Successful NWP depends upon well-balanced initial conditions to avoid the generation of spurious inertial gravity waves such as those that ruined the 1922 Richardson forecast. By balanced, we mean an atmospheric state in the slow manifold that approximately follows physical balance equations appropriate to the scale and location, such as the geostrophic relationship. In practice, there are initialization techniques for improving the balance of an analysis, such as nonlinear normal mode

initialization and digital filters (Lynch and Huang, 1992). However, once an analysis is filtered the resulting atmospheric state cannot be guaranteed to be optimal. Daley (1991, chapter 6) notes that there is no unique balanced state corresponding to a given unbalanced state; a filter may merely ignore the increment and move the solution back toward the balanced background state! Thus an ideal data assimilation system should avoid or reduce the initialization by filtering and try to create a well-balanced analysis.

The impact of localization on the balance of an analysis is discussed in Cohn et al. (1998) who noted an unrealistically high ratio of divergence to vorticity as a consequence of local observation selection. Mitchell et al. (2002) show that the optimum localization distance (in terms of improving analysis error) grows with ensemble size, and that balance is improved with longer localization distances. Lorenc (2003) provides an example of how localization produces imbalance. Consider the assimilation of a single height observation located at the origin (x=0) of Figure 1. The solid lines in Figure 1 represent a perfect scenario where the height h and meridional wind v are in geostrophic balance in the context of the shallow water equations (see Section 2 for details). The black line is proportional to the error covariances between h at the location x and h at the origin, while the gray line is proportional to the error covariances between v at x and h at the origin. In the assimilation of a single h observation, these lines are also proportional to the respective elements of the Kalman Gain matrix K, and therefore the analysis increments. Localization is then applied to these error covariances by multiplying them by a Gaussian function with length scale 250 km based upon distance from the observation, so that the error covariances decay to zero for larger x (dashed lines). In the region of $x = 250$ km, the analysis increment of v is reduced by localization. If

geostrophic balance is to be maintained, then the magnitude of the height gradient with respect to x should also be smaller. However, the height gradient is actually increased by localization and therefore the wind becomes significantly ageostrophic in this region (dash-dot line). In general, EnKF covariance localization modifies the elements of either the B matrix or the R matrix, which in turn reduces the elements of K as one moves further from the observation. Thus, as in this example, the analysis increments asymptote to zero as the analysis converges to the background in the absence of observation information. During this transition the geostrophic balance of the analysis increment is disrupted.

Kepert (2009) demonstrates how assimilation of wind and height observations with localized covariances produce imbalanced analyses with excess divergence, and proposes assimilation in terms of streamfunction ψ and velocity potential χ rather than u and v wind components. This technique results in a smaller (and more natural) ratio of divergence to rotation in the analysis, and hence balance is improved, but these improvements are less noticeable after initialization.

The purpose of this paper is to compare the B- and R-localizations and their impact on balance. Following a description of the EnKF and localization techniques (Section 2), we first compare the localizations using a simple model (Section 3), and then apply them to a global atmospheric model (Section 4).

2. Methods

a. Ensemble Kalman Filter Data Assimilation

The data assimilation cycle consists of a forecast stage, where the estimate of the state is evolved in time using a model, and an analysis stage, where the estimate of the state x_a is improved through optimal combination of forecast x_b and observations y_o.

$$x_a = x_b + K (y_o - h_{op}(x_b)) \qquad (1)$$

The optimal weight matrix K, or Kalman Gain, is given by

$$K = B H^T (HBH^T + R)^{-1} \qquad (2)$$

where B is the background error covariance matrix, R the observation error covariance matrix, and H the linearization of the observation operator h_{op}. In ensemble data assimilation methods, the background error covariance matrix is estimated using an ensemble of P forecasts:

$$B = \frac{1}{P-1} X_b X_b^T \qquad (3)$$

where X_b is the matrix of background ensemble perturbations from the ensemble mean with each row referring to a model variable, and each column to an ensemble member. The exact technique for updating the analysis ensemble members depends on the version of EnKF.

b. Localization Techniques

For B-localization, the B matrix is multiplied elementwise (i.e., through a Schur product) by another matrix C whose elements represent some localization function f_{loc} of distance d between grid points i and j (Houtekamer and Mitchell, 2001). Gaspari and Cohn (1999) describe a Gaussian localization function:

$$f_{Bloc} = \exp\left[\frac{-d(i,j)^2}{2L^2}\right] \qquad (4)$$

where L is a localization distance used for scaling the width of the localization. Gaspari and Cohn (1999) also introduced a piecewise polynomial approximation of a Gaussian localization function with compact support (this means it becomes zero beyond some finite distance, in this case at about 3.65 times L). Physically, this means that the background errors at model grid points that are far apart should have no statistical relationship.

With R-localization, modifications are made to the observation information. The simplest technique is through observation selection, by excluding observations that lie beyond a cutoff radius from the analysis (as in Houtekamer and Mitchell, 1998). However, abrupt localization cutoff can result in a noisy analysis. Hunt et al. (2007) proposed a gradual R localization by multiplying the elements of R by an increasing function of distance from the analysis grid point. Here we use the positive exponential function:

$$f_{Rloc} = \exp\left[\frac{+d(i,j)^2}{2L^2}\right] \qquad (5)$$

With uncorrelated observation error (which is a reasonable assumption for many instruments), R is diagonal. Then in (5), d is the distance between observation i and model grid point j. Since d varies depending upon which grid point the analysis is being performed at, the rows of K (in Equation 2) must be computed independently because the $(HBH^T + R)$ term will be different at each grid point location. Physically, this means

that far-away observations can be considered to have infinite error, and thus do not impact the analysis.

R-localization rather than B-localization is necessary for the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007), because as the calculations are done in ensemble space, the B matrix is not represented explicitly in physical space. The formulation of the Kalman gain (2) can be stated for the LETKF as

$$\mathbf{K} = \mathbf{X}_b \left[(P\text{-}1)\mathbf{I}_p + (\mathbf{HX}_b)^T \mathbf{R}^{-1} (\mathbf{HX}_b) \right]^{-1} (\mathbf{HX}_b)^T \mathbf{R}^{-1} \quad (6)$$

where I_p is the PxP identity matrix. For this study, we employ Gaussian localization (equations 4 and 5) with a cutoff distance of approximately 3.65 times L beyond which there is no observation impact (the localization function is set to zero). The application of (5) to a diagonal R using an observation cutoff radius of 3.65 L puts an upper bound on the conditioning number for R at 10^3 for the case of uniform observation errors. Localization can also be applied by dividing the diagonal elements of R^{-1} in (6) by f_{Rloc}. This reduces the size of the rightmost term of the bracketed expression in (6); as this smaller term is then added to the identity matrix, the inversion of the bracketed expression remains a stable calculation. Note that some studies (i.e., Houtekamer and Mitchell, 2005) report localization values in terms of cutoff distance rather than L.

For NWP applications, B (N x N, where N is the dimension of x) is too large to be represented explicitly, therefore the BH^T and HBH^T terms of Equation 2 are calculated directly from the ensemble, as in Houtekamer and Mitchell (2001). For the serial EnSRF (Whitaker and Hamill, 2002), localization by a distance-dependent function is performed upon BH^T, where each element represents the covariance between a model grid point and

observation. Because $\boldsymbol{HBH^T}$ is a scalar, it does not require localization. In the case of observations on grid points (which is the case used in this study), this form of localization (on $\boldsymbol{BH^T}$) is equivalent to B-localization. When observations are located off grid points, or relate to more than one grid point, this technique exhibits hybrid properties of B-localization and R-localization. The problem of defining distance for vertically integrated measurements, such as satellite observations (Campbell et al., 2010), is equally challenging for $\boldsymbol{BH^T}$ and R localization techniques, as both require a distance between an observation and model grid point, and this issue is a motivation for adaptive localization (Anderson 2007; Bishop and Hodyss, 2009). This study focuses on horizontal localization with point observations; vertical localization in the LETKF is addressed in Miyoshi and Sato (2007).

3. Simple Model Experiments

The goal of this section is to demonstrate the impact of EnKF localization on balance using a simple model consisting of one-dimensional balanced waveforms. These initially balanced wave solutions (which are not integrated forward in time) serve as truth and background ensemble states for identical twin data assimilation experiments; any disruption to the balance of the resulting analysis is thus easily detectable and attributable to the properties of the EnKF technique.

a. Simple Model Description

Consider the shallow water momentum equation in the *x*-direction for a rotating (constant Coriolis parameter *f*), inviscid fluid:

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + fv - g \frac{\partial h}{\partial x} \qquad (7)$$

The geostrophic balance between the pressure gradient and Coriolis terms can thus be stated:

$$fv_g = g\frac{\partial h}{\partial x} \qquad\qquad (8)$$

Here v_g is the geostrophic wind. Assuming that the wave structure is uniform in the y-direction, harmonic form is applied to the perturbation variables to achieve a wave solution for h, with h_{depth} being the mean depth of the fluid, h_{amp} the amplitude of the height perturbation, k the wavenumber, and x_{ps} a wave phase shift:

$$h = h_{depth} + h_{amp}\cos(k(x - x_{ps})) \qquad\qquad (9)$$

Assuming geostrophically balanced wind field, we arrive at the wave solution for v:

$$v = -\frac{g}{f}kh_{amp}\sin(k(x - x_{ps})) \qquad\qquad (10)$$

For the simple model, consider a one-dimensional non-periodic domain of 5000 km along the x-axis, with model grid points spaced regularly at 50-km intervals. The Coriolis parameter f was selected to be $10^{-4}\,\text{s}^{-1}$, a reasonable value for the mid-latitudes.

b. Experiment Design

The truth state and 5 background ensemble members, plotted in Figure 2, are defined for both height and v-component of the wind. Each ensemble member is generated by randomly selecting a height perturbation amplitude from a uniform distribution of [9, 11] m, a wavelength from [1950, 2050] km and phase shift from [-50, 50] km. The truth waveform (amplitude = 10m, wavelength = 2100 km, offset = -100 km) is fixed in order to avoid having a mean background state too close to the ensemble mean. This would be undesirable, as an analysis that moves further from the background toward an observation would be overly penalized, whereas one that remained close to the

background would be falsely rewarded. The meridional wind waveform is then generated to be in geostrophic balance with the height waveform. These waves are represented discretely as height and meridional wind values at each of the 101 model grid points. Observations of both h and v at regularly spaced grid points 250 km apart are chosen based upon the truth value at the corresponding model grid point plus a random observation error equal to 10% of the wave amplitude.

Ensemble mean analyses resulting from assimilation using no localization, B-localization, and R-localization using various localization distances L are compared. As the wind can be partitioned into geostrophic and ageostrophic components ($v = v_g + v_a$), the RMS value of v_a over all grid points is used as a summary metric of imbalance; accuracy is also assessed as the RMS difference from the truth. To obtain significant results not dependent upon the peculiarities of a specific random configuration of ensemble members and observation errors, each configuration is repeated 100 times in a Monte Carlo experiment. Note that the model is not advanced in time, so boundary conditions are not needed.

c. Simple Model Results

Figure 3a shows the dependence of RMSE for each analysis as a function of localization distance L. LETKF rather than the generic EnKF formula is used for R-localization; the differences in accuracy and balance metrics between LETKF and EnSRF R-localization for this experiment (not shown) are on the order of 1%, so the comparison is fair. R-localization has an optimal scale of $L = 500$ km, whereas B-localization is close to optimal at around $L = 1000$ km and larger for 5 ensemble members. A scenario using 40 ensemble members and no localization is also plotted as a best-case performance

scenario to which the localized 5-ensemble member analyses aspire. Note that results for *v*-wind error (not shown) are similar. An explanation for the disparity in optimal length scales is provided in the Appendix.

Figure 3b shows the dependence of RMS imbalance (ageostrophic wind) for each analysis as a function of localization distance L. Analyses without localization show no ageostrophic wind, which is to be expected from the design of the experiment. For the localized cases, as the localization distance increases, the analysis becomes more balanced. R-localization is always more balanced than B-localization for the same localization distance L, although the levels of imbalance are comparable when considering the optimal configuration of each method.

4. SPEEDY Model Experiments

a. Measuring Balance in a Realistic Model

In a realistic atmospheric model we can no longer assume that the background state is initially balanced, since an atmosphere with purely geostrophic flow would not allow for interesting weather such as intense baroclinic development and the vertical motion associated with heavy precipitation. Therefore, although much of the energy in the atmosphere is associated with the slow mode (Daley 1991), there is a natural level of imbalance in the atmosphere. The challenge is to differentiate between this background amount of imbalance, and additional spurious amounts introduced as an artifact of data assimilation.

There are several metrics for evaluating atmospheric imbalance. Section 3 (and Lorenc 2003) uses the magnitude of the ageostrophic wind. While this metric is

straightforward to compute, it is not applicable at all latitudes; there are also more sophisticated balance equations, such as nonlinear balance (Raymond, 1992), to consider. High frequency oscillations can be diagnosed directly by examining the second derivative of the surface pressure field in time (Houtekamer and Mitchell, 2005). Finally, the analysis can be compared to an initialized (filtered) version of itself using a Lynch and Huang (1992) Lanczos digital filter (as in Mitchell et al., 2002) that removes high frequency oscillations, and thus inertial-gravity waves, from the model time series (not included in this study). Similarly, Kepert (2009) used the magnitude of the NNMI increment as a measure of balance. The surface pressure and digital filter metrics require model output from several time steps at a relatively fine temporal resolution (smaller than one hour).

b. Experiment Design

The Simplified Parametrizations, primitivE-Equation DYnamics, or SPEEDY, model (Molteni, 2003) is an atmospheric global circulation model of intermediate complexity designed for climate experiments. While containing many of the physics components found in larger models (including convection, condensation, cloud, radiation, and surface flux parameterizations), it is computationally inexpensive so it can be run on a single processor. There are seven vertical levels using the sigma coordinate system, with a horizontal spectral resolution of T30, which corresponds to a standard Gaussian grid of 96 by 48 points. The time scheme is leapfrog. There are five dynamical variables included in the output: zonal wind (u), meridional wind (v), temperature (T), specific humidity, and surface pressure (p_s). Miyoshi (2005) modified the SPEEDY model for weather forecasting by creating output every six hours, and implemented several data

assimilation techniques on the SPEEDY model. Horizontal diffusion (of vorticity, divergence, temperature, specific humidity) in the SPEEDY model is done with the fourth power of the Laplacian, and is applied on the sigma surfaces. Maximum damping time is 18 hours for temperature and vorticity, and 9 hours for divergence, with an additional 12 hours applied at the top level (representing the stratosphere). There is also vertical diffusion that simulates shallow convection in regions with conditional instability, as well as water vapor and static energy vertical diffusion (Molteni, 2003). Frequency damping with a Robert-Asselin filter (with filter parameter = 0.05) is included in the SPEEDY model to suppress the spurious computational mode. Amezcua et al. (2010) has examined the use of a Robert-Asselin-Williams (RAW) filter (which successfully dampens the computational mode without damping the physical solution; Williams, 2009) with the SPEEDY model, and found that there are very few changes to the model climatology that pass a field significance test, and the quality of the forecasts was slightly improved. This change in the high frequency damping did not seem to affect the model balance. Note that the RAW filter is not employed in the experiments presented in this paper.

The ultimate goal of using the SPEEDY model is a realistic comparison of B-localization and R-localization in terms of balance and accuracy. Here, B-localization is employed with the EnSRF algorithm (Whitaker and Hamill, 2002), whereas R-localization is used with LETKF (Hunt et al., 2007). In addition, a third configuration using the EnSRF with R-localization is employed to investigate whether any differences between the first two configurations are primarily due to variation in localization technique rather than assimilation algorithm (serial versus simultaneous, etc.); see

Holland and Wang (2010) for an independent comparison of EnSRF and LETKF. All systems use identical observations, which are generated as random perturbations from the nature run, or true state, in an identical twin experiment. The observation network used for this study approximately follows the rawinsonde locations (Figure 7), with all observations located on model grid points. Observations are located at each of the seven model levels. Observation error is 1K for temperature, 1 m/s for u and v wind magnitudes, 1 g/kg for specific humidity, and 1 mb for surface pressure. Multiplicative inflation of 2 % is applied to the background ensemble spread. Vertical localization is by model level so that an observation corresponding to one of the model's seven levels does not impact any other level; previous experience with the SPEEDY model has shown that vertical correlations for wind and temperature errors are minimal. The ensembles are comprised of 20 members, with initial conditions taken from consecutive dates in January 1982.

The forecast-assimilation cycle is every 6 hours over a period of 48 days from Feb 1 to Mar 20, 1982. The assessment of accuracy is made by comparing the ensemble mean analysis of wind magnitude to the truth at each 6-hour period. Balance is assessed through the magnitude of the ageostrophic wind, as well as the second derivative of surface pressure. These metrics are applied during the month-long period of Feb 20 to Mar 20 following 20 days of spinup. Wind metrics are obtained from model level 4 (~500 hPa). Results are reported as an areal mean, either globally or over mid-latitude bands (~30 to 60 deg) separately for the northern hemisphere (NH) and southern hemisphere (SH).

c. SPEEDY Model Results

Figure 4 shows the accuracy of analyses (measured by mean absolute wind error at ~500 hPa) for the EnSRF B-localization and LETKF R-localization relative to the true state as a function of localization distance parameter L (see the discussion surrounding equations 4 and 5). The performance of the system is highly dependent upon the choice of localization parameter. Too long a localization distance and the system is dominated by spurious observation increments that prevent it from converging to the truth, whereas too short a localization distance and observations introduce imbalanced increments, as well as fail to adequately impact their neighborhood of grid points. An optimal localization distance parameter L with respect to accuracy is 500 km for R-localization, and 750 km for B-localization. Error is higher and the optimal length scale is slightly longer for the SH compared to the NH (not shown), as the former has a relative paucity of observations. The performance for R-localization in both LETKF and EnSRF is similar, particularly for L< 500 km where the results are essentially identical. The results for wind error at other vertical levels (not shown) reveal a similar dependence on localization, with slightly higher errors as altitude increases. Note that the areal mean ensemble spread (not shown) is also highly sensitive to L, with shorter L corresponding to greater ensemble spread. Observation information reduces the uncertainty of an analysis; for shorter localization distances this reduction in analysis spread takes place over smaller regions (nearest to the observations), and thus the areal mean ensemble spread remains high.

Figure 5 reveals the performance of the two systems with respect to balance, measured by the mean magnitude of the ageostrophic wind at ~500 hPa as a function of the localization distance parameter L. There exists a larger natural state of geostrophic imbalance in the NH (~ 3 m/s) compared to the SH (~ 2 m/s) due to the presence of the

Himalayan plateau protruding into the mid-latitude belt as well as the fact that the experiment occurred in the NH winter with its stronger wind speeds. In all cases, the imbalance of the analyses is larger than that of the true state, indicating that data assimilation has introduced artificial imbalance. Although the magnitudes of the mean ageostrophic winds are higher for the NH, the difference in imbalance between the nature run and assimilation runs (assimilation-induced imbalance) is greater for the SH. Short localization distances ($L < 300$ km) are detrimental to balance, which agrees with the results of Section 3 using a simple model. For very long localization distances (L=2000 km), presumed spurious correlations can lead to larger values of both error and imbalance. Examination of performance time series reveal that values of imbalance tend to stabilize, along with the error, after 20 days of spin-up, although there are day-to-day fluctuations on the order of 0.5 m/s that are reflected in both the nature run and assimilation analyses.

Figure 6 also depicts imbalance, but measured by the second derivative of surface pressure at each model time step. As in Figure 5, short localization distances (L<300 km) are very harmful to balance. Here, the NH is significantly more balanced than the SH, which agrees with the result for assimilation-induced imbalance in Figure 5. Optimal values of L are slightly larger using this metric compared to Figure 5; averaging the optimal L values for both metrics of imbalance results in an optimal L that agrees with the results for accuracy in Figure 4. The occasional lack of smoothness in the relationship curves between imbalance and L in Figures 5-6 reveal that an evaluation time period of at least one month is required to overcome sampling error for these techniques.

Figure 7 reveals the spatial distribution of imbalance as a time mean over the period from Feb. 20 – Mar. 20. For short localization distances, imbalance is large in the immediate vicinity of observations. For long localization distances, imbalance is smaller and spread over broader areas. This finding agrees with the Lorenc (2003) explanation using Figure 1 in that imbalance can be introduced in the region where the impact of an observation moves toward zero. The circular patterns of imbalance surrounding the Southern Ocean islands in the case of L=250 km demonstrate the detrimental impact of strong localization resulting from a sharp transition between a region with strong observation impact and a region with little observation impact. Imbalance is greatest along the Pacific coast of South America; the lack of observations in the South Pacific leads to large observation increments in the region. Inaccurate background fields, which require larger subsequent analysis increments resulting in greater potential for imbalance introduced by data assimilation, may explain the somewhat unexpected increase in imbalance for large L in Figures 5 and 6.

5. Conclusions

This study has examined the impact of EnKF localization techniques upon the accuracy and balance of analyses. Localization is used to combat spurious correlations due to sampling error from finite ensemble size, to take advantage of low dimensionality in local regions, and for efficient computation. Localization techniques can be classified into two methods: B-localization, where the background error covariance is modified by a distant-dependent localization function, and R-localization, where observation error variances are increased as distance from the analysis grid point increases. Variations of the B-localization technique are appropriate for EnSRF where the entire domain is

updated with each observation, whereas R-localization is used for LETKF as the background error covariances are specified in ensemble space and each model grid point is updated independently. In addition to accurately depicting the state of the system, atmospheric data assimilation should produce a balanced analysis so that information is not lost through spurious inertial-gravity wave propagation.

We first described experiments with simple, one-dimensional waveforms based upon the shallow water equations. As the background ensemble is initially balanced, imbalance introduced by data assimilation is easy to measure as the magnitude of the ageostrophic wind. The two techniques have differing optimal localization distances L with respect to analysis accuracy; approximately 500 km for R-localization, and 1000 km or larger for B-localization. For the same localization length R-localization is more balanced than B-localization but the balance of both techniques improves as L grows larger.

We then made a more realistic comparison between EnSRF B-localization and LETKF R-localization involving the global SPEEDY model in identical twin experiments. Here, the background state can no longer be assumed to be in balance. Two methods for evaluating imbalance are used: the magnitude of the ageostrophic wind, and the second derivative of surface pressure. The two localization techniques are roughly comparable in performance with respect to localization and balance when the optimal length scale of L is selected: 500 km for R-localization, and 750 km for B-localization. This result is consistent with the discussion in the Appendix, which demonstrates that B-localization is more severe than R-localization for the same L. We conclude that the differences in data assimilation algorithm (LETKF vs. EnSRF) are

smaller than differences in localization technique when identifying the optimal localization distance L.

Both types of localization introduce imbalance; as the solution reverts toward the background at long distances from observations, the damping of the height and wind increments results in a smaller wind increment but a larger height gradient, which does not satisfy the geostrophic relationship. Localization can also introduce excess divergence to an analysis (Kepert, 2009). The localization parameter L should be tuned depending on the particular scale and application of data assimilation, as well as the size of the ensemble. Tuning inflation values for each localization parameter L may result in improved performance. Future studies should consider balance in the context of the adaptive localization methods, as these techniques do not necessarily require a specification of L.

APPENDIX

Mathematical Analysis of B- and R-Localizations

The relative strength of B-localization and R-localization techniques are verified mathematically using a simple example with model variable x_1 and x_2 at grid points 1 and 2, respectively. Consider a single observation of x_1, with $\boldsymbol{H}$=[1,0]. Using (2), the Kalman gain matrix (without localization) can be specified as follows:

$$\begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} B_{11} \\ B_{12} \end{bmatrix} \left(B_{11} + R_1 \right)^{-1} \tag{A1}$$

where B_{ij} is the background covariance between x_i and x_j, and R_1 is the observation covariance.

Consider the application of the B-localization function f_{Bloc} (4) to (A1). Using $f_{Bloc}\,(d_{ii}) = 1$ where d_{ij} is the distance between grid points i and j, K_1 remains the same but K_2 becomes:

$$K_2 = f_{Bloc}(d_{12})B_{12}\,(f_{Bloc}(d_{11})B_{11} + R_1)^{-1} = f_{Bloc}(d_{12})B_{12}\,(B_{11} + R_1)^{-1} \tag{A2}$$

Note that since we are assimilating a single observation located on a grid point, (A2) is identical for both B-localization and the $\boldsymbol{BH}^{\mathrm{T}}$ localization described at the end of Section 2. Now we apply the R-localization function f_{Rloc} (5). Again, K_1 remains the same as in (A1). Using the fact that $f_{Bloc} = f_{Rloc}^{-1}$, K_2 becomes:

$$K_2 = B_{12}\,(B_{11} + f_{Rloc}(d_{12})R_1)^{-1} = f_{Bloc}(d_{12})B_{12}\,(f_{Bloc}(d_{12})B_{11} + R_1)^{-1} \tag{A3}$$

Comparing (A2) and (A3), the R-localization (A3) has an extra localization term in the denominator. The localization function f_{Bloc} ranges from 1 to 0. Therefore, the

amplitude of K_2 (and hence the corresponding analysis increment) will be larger at grid point 2 for R-localization than for B-localization. This means that with B-localization, the analysis reverts to the background (ignores observation information) more quickly with distance compared to R-localization. In this respect, B-localization can be considered more "severe" than R-localization for the same localization distance parameter L; see discussion of (11) and (12) in Miyoshi and Yamane, (2007).

Now consider the same example but with two observations (one at each of the grid points) with uncorrelated errors, i.e., H is a 2-dimensional identity matrix. The Kalman gain would be written as

$$\begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \left(\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1} \tag{A4}$$

where R_1 and R_2 represent the error variances of the two observations. Because the analysis process is the same for x_1 and x_2 by permuting the indices 1 and 2, we consider the impact of the localizations on x_1 (*i.e., K_1*) only. The application of the B-localization function leads to

$$[K_1] = \begin{bmatrix} f_{Bloc}(d_{11})B_{11} & f_{Bloc}(d_{12})B_{12} \end{bmatrix} \left(\begin{bmatrix} f_{Bloc}(d_{11})B_{11} & f_{Bloc}(d_{12})B_{12} \\ f_{Bloc}(d_{21})B_{21} & f_{Bloc}(d_{22})B_{22} \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} B_{11} & f_{Bloc}(d_{12})B_{12} \end{bmatrix} \left(\begin{bmatrix} B_{11} & f_{Bloc}(d_{12})B_{12} \\ f_{Bloc}(d_{12})B_{21} & B_{22} \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1} \tag{A5}$$

The application of the R-localization function with $f_{Bloc} = f_{Rloc}^{-1}$ gives

$$[K_1] = \begin{bmatrix} B_{11} & B_{12} \end{bmatrix} \left(\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} + \begin{bmatrix} f_{Rloc}(d_{11})R_1 & 0 \\ 0 & f_{Rloc}(d_{12})R_2 \end{bmatrix} \right)^{-1} \tag{A6}$$

$$= \begin{bmatrix} B_{11} & B_{12} \end{bmatrix} \left(\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \left(\begin{bmatrix} f_{Bloc}(d_{11}) & 0 \\ 0 & f_{Bloc}(d_{12}) \end{bmatrix} \right)^{-1} \right)^{-1}$$

$$= \begin{bmatrix} B_{11} & B_{12} \end{bmatrix} \begin{bmatrix} f_{Bloc}(d_{11}) & 0 \\ 0 & f_{Bloc}(d_{12}) \end{bmatrix} \left(\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} f_{Bloc}(d_{11}) & 0 \\ 0 & f_{Bloc}(d_{12}) \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} B_{11} & f_{Bloc}(d_{12})B_{12} \end{bmatrix} \left(\begin{bmatrix} B_{11} & f_{Bloc}(d_{12})B_{12} \\ B_{21} & f_{Bloc}(d_{12})B_{22} \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1} \qquad \text{(A7)}$$

Comparing (A5) and (A7) in terms of the B-localization function f_{Bloc}, we note that the $\boldsymbol{BH^T}$ terms are identical. However, the $\boldsymbol{HBH^T}$ terms differ. Using this formulation, we arrive at an $\boldsymbol{HBH^T}$ matrix for R-localization in (A7) that is no longer symmetric, although the original formulation of R-localization in terms of the R-localization function had symmetric covariance matrices (A6). Consequently, it is not straightforward to compute a priori the quantitative difference in localization strength between the techniques in the case of multiple observations. With localized serial EnSRF, the resulting analysis depends upon the order in which the observations are assimilated; this is not true for the simultaneous assimilation of LETKF. For this study we focus on R-localization with the LETKF algorithm, performing EnSRF R-localization in order to confirm that differences in the results are primarily due to difference in localization technique rather than algorithm. Note that EnSRF R-localization requires a unique $\boldsymbol{R}$ (and hence a separate computation) for every gridpoint-observation pair.

References

Amezcua, J., E. Kalnay, and P. Williams, 2010: The effects of the RAW filter on the climatology and forecast skill of the SPEEDY model. *Mon. Wea. Rev.*, Accepted.

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

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

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

Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical Covariance Localization for Satellite Radiances in Ensemble Kalman Filters. *Mon. Wea. Rev.*, **138**, 282-290.

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the Effects of Data Selection with the DAO Physical-Space Statistical Analysis System. *Mon. Wea. Rev.*, **126**, 2913-2926.

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

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10143-10162.

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

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Mon. Wea. Rev.*, **129**, 2776-2790.

Holland, B. W. and X. Wang, 2010: A comparison of the local ensemble transform Kalman filter and ensemble square root filter data assimilation schemes. *14th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, Atlanta, GA, Amer. Meteor. Soc.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Mon. Wea. Rev.*, **126**, 796-811.

Houtekamer, P. L. and H. L. Mitchell, 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

Houtekamer, P. L. and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteorol. Soc.*, **131**, 3269-3289.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman Filter. *Physica D*, **230**, 112-126.

Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng., Trans. ASME*, 35-45.

Kepert, J., 2009: Covariance localisation and balance in an ensemble Kalman filter. *Quart. J. Roy. Meteorol. Soc.*, **135**, 1157-1176.

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

Lynch, P. and X.-Y. Huang, 1992: Initialization of the HIRLAM Model Using a Digital Filter. *Mon. Wea. Rev.*, **120**, 1019-1034.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble Size, Balance, and Model-Error Representation in an Ensemble Kalman Filter. *Mon. Wea. Rev.*, **130**, 2791-2808.

Miyoshi, T., 2005: Ensemble Kalman Filter Experiments with a Primitive-Equation Global Model. Ph.D. dissertation, University of Maryland, 226 pp.

Miyoshi, T., and Y. Sato, 2007: Assimilating Satellite Radiances with a Local Ensemble Transform Kalman Filter (LETKF) Applied to the JMA Global Model (GSM). *SOLA*, **3**, 37-40.

Miyoshi, T. and S. Yamane, 2007: Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 Resolution. *Mon. Wea. Rev.*, **135**, 3841-3861.

Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: model climatology and variability in multi-decadal experiments. *Climate Dynamics*, **20**, 175-191.

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

Raymond, D. J., 1992: Nonlinear balance and potential-vorticity thinking at large Rossby number. *Quart. J. Roy. Meteor. Soc.,* **118**, 987–1015.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus A*, **60**, 113-130.

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

Williams, P. D., 2009: A Proposed Modification to the Robert-Asselin Time Filter. *Mon. Wea. Rev.,* **137**, 2538-2546.
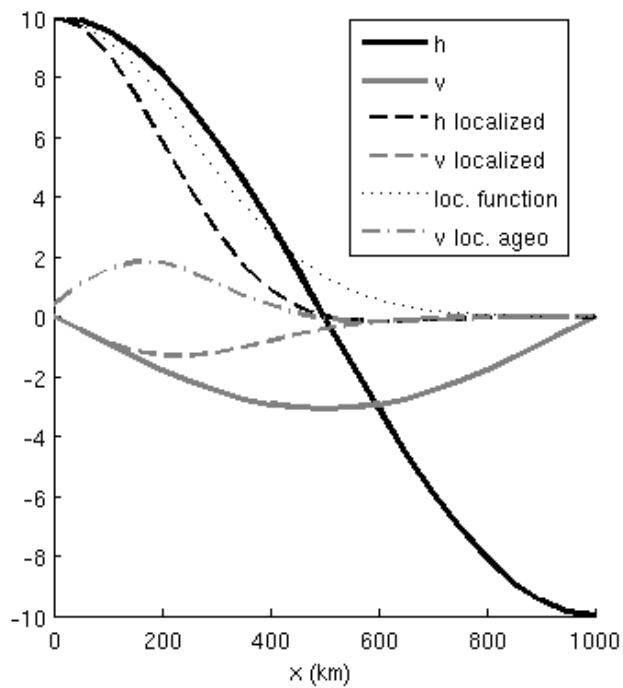
List of Figures

FIG. 1. Example showing the introduction of imbalance by localization (after Lorenc, 2003). Waveforms of height (black) and meridional wind (gray) before (solid) and after (dashed) multiplication by a Gaussian localization function (dotted). Values on the y-axis denote the size of the analysis increment (m; m s^{-1}) from the assimilation of a height observation located at the origin. The ageostrophic portion of the wind increment after localization is dash-dotted.

FIG. 2. Sample experimental setup for the simple model experiment. The black curves represent the height waveform, while the gray represent the meridional wind. Thick solid lines depict the truth waveforms, whereas dashed lines are used for the ensemble members. Black circles are height observations, whereas gray diamonds are wind observations.

FIG. 3. RMS error of the analysis from the truth for height (m; left panel) and RMS ageostrophic wind (m/s, right panel) using no localization, B localization, and R localization for 5 ensemble members and a variety of localization distances L. For comparison, an analysis with no localization and 40 ensemble members is also plotted. Arrows depict optimum values of L.

FIG. 4. Summary of SPEEDY accuracy statistics for B-localization vs. R-localization. Error bars denote standard deviation over time. Arrows denote optimal values of localization distance L. For L < 500km, EnSRF R-localization and LETKF R-localization give essentially identical results.

FIG. 5. Summary of SPEEDY imbalance statistics for B-localization vs. R-localization as measured by the ageostrophic wind (m/s). Natural levels of imbalance are noted as horizontal lines. Error bars denote standard deviation over time. Arrows denote optimal values of localization distance *L*.

FIG. 6. Summary of SPEEDY imbalance statistics for B-localization vs. R-localization as measured by the second derivative of surface pressure (Pa s^{-2}). Arrows denote optimal values of localization distance *L*.

FIG. 7. Time average spatial distribution of imbalance measured by the second derivative of surface pressure (Pa s^{-2}) for short (100 km, left panels) and long (2000 km, right panels) localization distances using EnSRF B-localization (top panels) and LETKF R-localization (bottom panels). Observation locations are depicted by black dots.

FIG. 1. Example showing the introduction of imbalance by localization (after Lorenc, 2003). Waveforms of height (black) and meridional wind (gray) before (solid) and after (dashed) multiplication by a Gaussian localization function (dotted). Values on the y-axis denote the size of the analysis increment (m; m s^{-1}) from the assimilation of a height observation located at the origin. The ageostrophic portion of the wind increment after localization is dash-dotted.
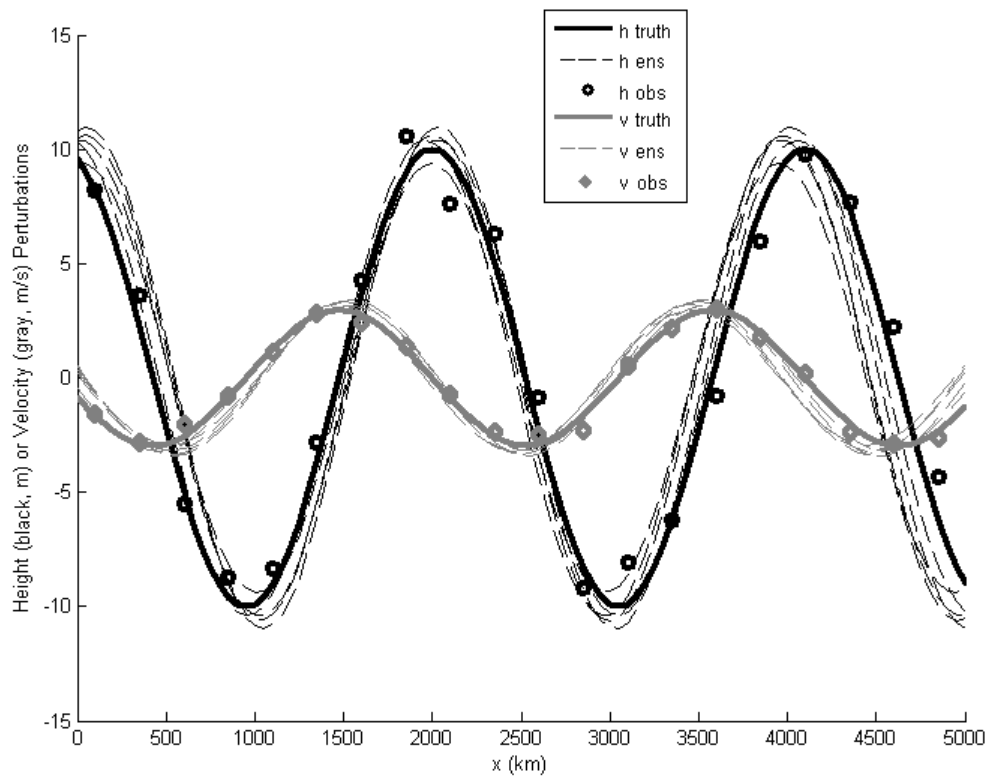
FIG. 2. Sample experimental setup for the simple model experiment. The black curves

represent the height waveform, while the gray represent the meridional wind. Thick solid

lines depict the truth waveforms, whereas dashed lines are used for the ensemble

members. Black circles are height observations, whereas gray diamonds are wind
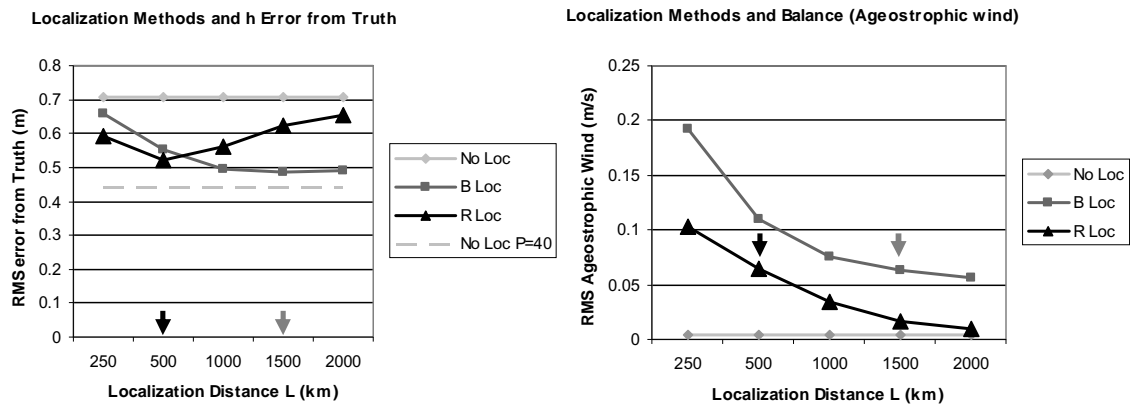
observations.

Localization Methods and h Error from Truth

Localization Methods and Balance (Ageostrophic wind)

FIG. 3. RMS error of the analysis from the truth for height (m; left panel) and RMS

ageostrophic wind (m/s, right panel) using no localization, B localization, and R

localization for 5 ensemble members and a variety of localization distances *L*. For

comparison, an analysis with no localization and 40 ensemble members is also plotted.
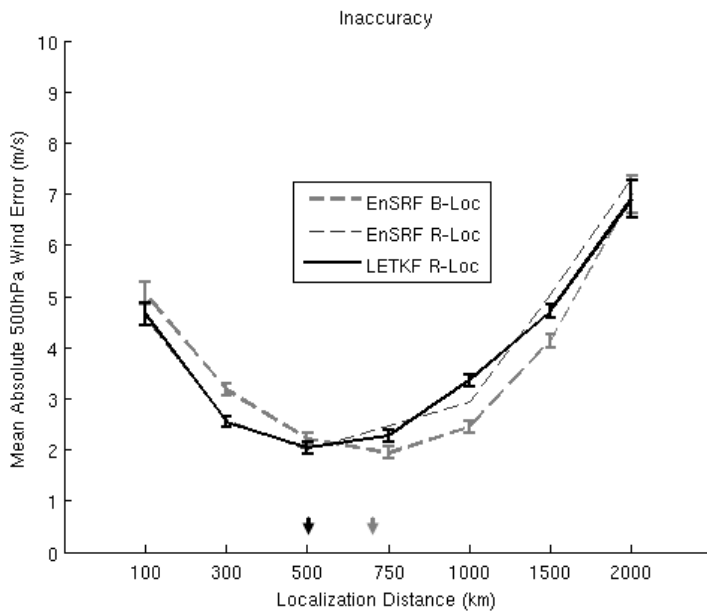
Arrows depict optimum values of *L*.

FIG. 4. Summary of SPEEDY accuracy statistics for B-localization vs. R-localization. Error bars denote standard deviation over time. Arrows denote optimal values of localization distance *L*. For L < 500km, EnSRF R-localization and LETKF R-localization give essentially identical results.
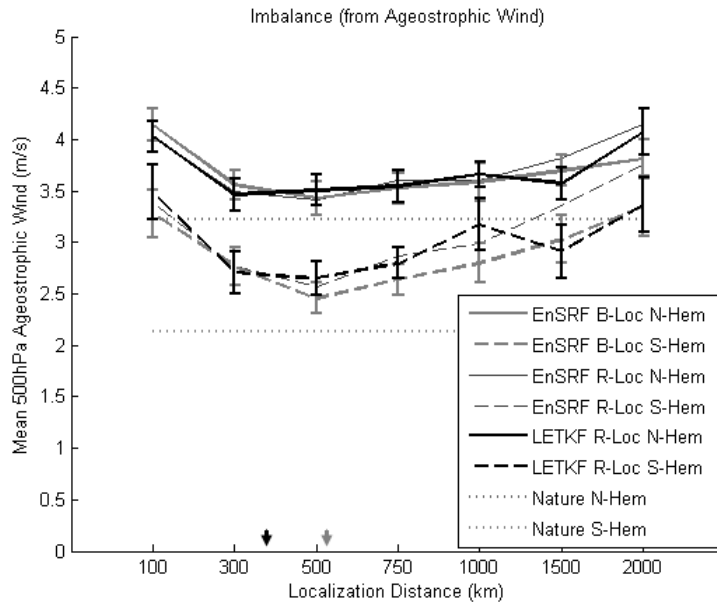
FIG. 5. Summary of SPEEDY imbalance statistics for B-localization vs. R-localization as measured by the ageostrophic wind (m/s). Natural levels of imbalance are noted as horizontal lines. Error bars denote standard deviation over time. Arrows denote optimal values of localization distance *L*.
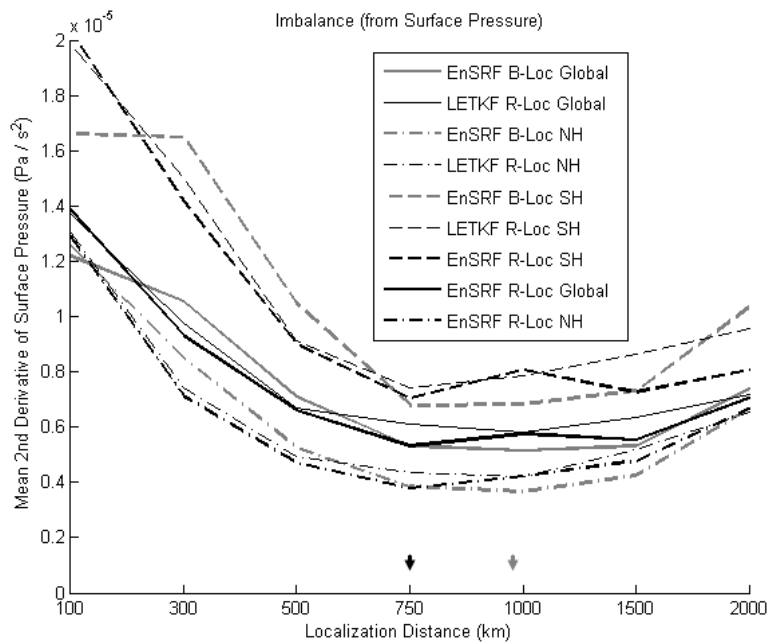
FIG. 6. Summary of SPEEDY imbalance statistics for B-localization vs. R-localization as measured by the second derivative of surface pressure (Pa s^{-2}). Arrows denote optimal values of localization distance *L*.
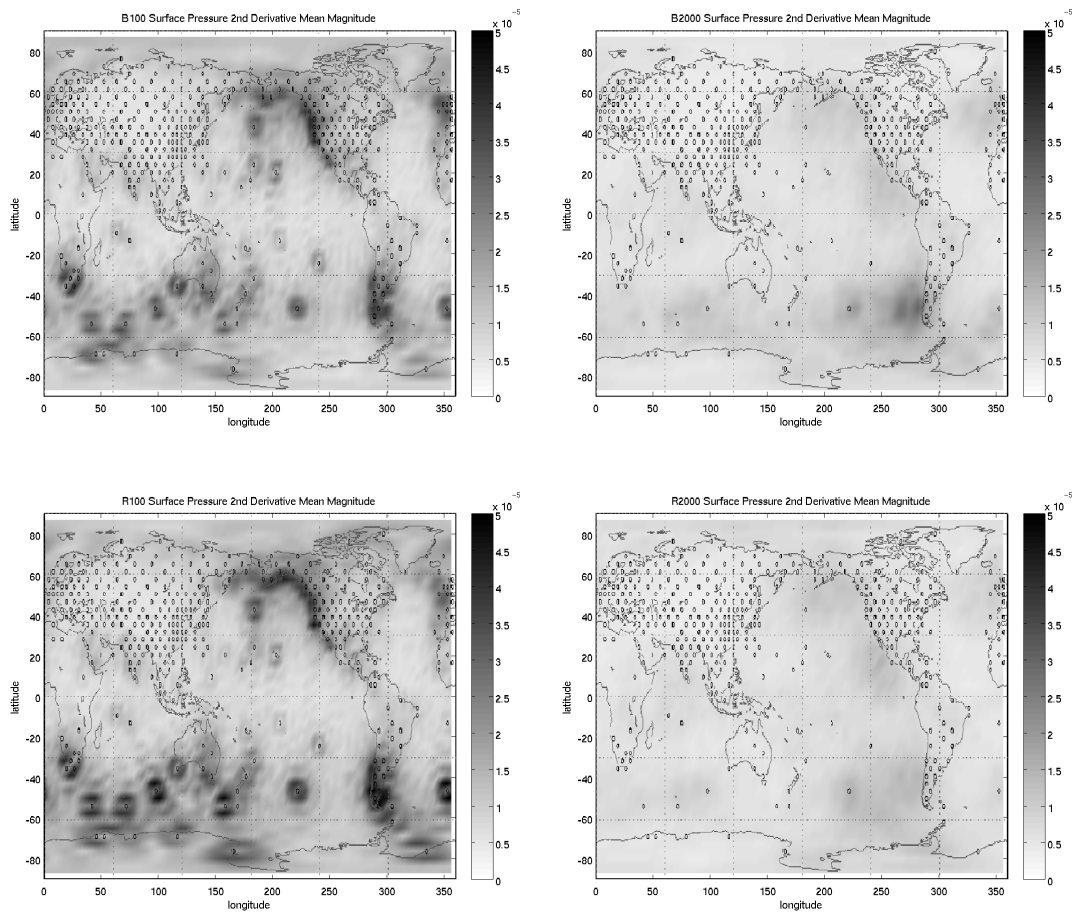
FIG. 7. Time average spatial distribution of imbalance measured by the second derivative of surface pressure (Pa s^{-2}) for short (100 km, left panels) and long (2000 km, right panels) localization distances using EnSRF B-localization (top panels) and LETKF R-localization (bottom panels). Observation locations are depicted by black dots.