Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter

Journal:	*QJRMS*
Manuscript ID:	Draft
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Holland, Brian; University of Oklahoma, School of Meteorology Wang, Xuguang; University of Oklahoma, School of Meteorology
Keywords:	covariance localization, serial assimilation, balance

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter

Brian Holland and Xuguang Wang*
School of Meteorology and Center for Analysis and Prediction of Storms
The University of Oklahoma, OK, USA

*Corresponding author: Dr. Xuguang Wang, University of Oklahoma, School of Meteorology, Suite 5900, 120 David L. Boren Blvd., Norman, OK 73072, USA. E-mail: xuguang.wang@ou.edu

Abstract

The effect of using sequential or simultaneous assimilation of observations and the effect of applying localization to the observation error covariance matrix (R-localization) or the background error covariance matrix (B-localization) in the ensemble Kalman filter (EnKF) were examined. A B-localized sequential scheme, a B-localized simultaneous scheme, an R-localized sequential scheme, and an R-localized simultaneous scheme were compared using a primitive equation two-layer model with simulated observations and an imperfect model assumption. The R-localized simultaneous scheme produced the least analysis and forecast error and imbalance, while the B-localized simultaneous scheme produced the most analysis and forecast error and imbalance. The B-localized and R-localized sequential schemes produced similar results to each other that were intermediate between the performances of the two simultaneous schemes. The schemes were further compared in a non-cycling experiment where the same background ensemble was used as input to each scheme. In this experiment, the root mean square analysis error of the schemes was similar, while the same pattern of imbalance differences was observed as in the original cycling experiment. This result suggests that the imbalance differences among the schemes were translated into error differences during the forecast step and were accumulated during the cycling experiment. To further understand the effects of the choice of the observation assimilation pattern and the localization method on the performance of the EnKF, the original cycling experiment was repeated under a variety of circumstances. The differences between the schemes were found to be smaller when digital filter initialization was used, when both wind and mass observations were assimilated, when the number of observations was decreased, when the

ensemble size was increased, and when the ratio of forecast error to observation error in

the system was decreased.

Keywords: covariance localization, serial assimilation, balance

1. Introduction

In recent years, ensemble Kalman filter (EnKF) data assimilation (Evensen, 1994) has been used in climate applications (e.g., Whitaker et al., 2004; Compo et al., 2011), prediction of convective storms (e.g., Snyder and Zhang, 2003; Tong and Xue, 2005; Aksoy et al., 2009; Dowell and Wicker, 2009), prediction of tropical systems (e.g., Chen and Snyder, 2007; Torn and Hakim, 2009; Zhang et al., 2009; Hamill et al., 2011; Wang, 2011), and general numerical weather prediction (NWP) using global models (e.g., Whitaker et al., 2008; Miyoshi et al., 2010) and limited-area models (e.g., Bonavita et al., 2008; Meng and Zhang, 2008; Wang et al., 2008a; 2008b; Bonavita et al., 2010). Several operational centers have implemented or plan to implement the EnKF operationally (e.g., Houtekamer and Mitchell, 2005; Charron et al., 2010). While all of these implementations of the EnKF share the same spirit, in which the ensemble covariance is used to estimate the flow-dependent background error covariance during the assimilation, the specific implementations of the EnKF can be different in several ways. Some examples of these differences include whether observations are assimilated simultaneously or sequentially, how covariance localization is applied, whether a stochastic or deterministic approach is used, and, in the deterministic approach, how the analysis perturbations are generated. This study focuses on examining the impacts of two of these choices: simultaneous or sequential assimilation of observations and the method used to apply localization, on the performance of the EnKF.

In the original formulation of the EnKF, all p observations available for a given analysis are assimilated simultaneously (Evensen, 1994). This approach requires

4

inversion of a p x p matrix, which can be computationally prohibitive when dealing with realistic atmospheric observation sets. One means of reducing the computational burden is the use of a sequential assimilation pattern (Houtekamer and Mitchell, 2001). In this method, observations are divided up into groups, which can be as small as one observation each in the "serial" assimilation case (e.g., Whitaker and Hamill, 2002). Each group of observations is assimilated sequentially, and the resulting analysis after assimilation of one group is used as the background for assimilation of the next group. The computational demands can also be reduced while still using simultaneous assimilation by generating local analyses rather than a global analysis (Ott et al., 2004; Hunt et al., 2007). In this method, the analysis for each gridpoint is calculated independently by simultaneously assimilating those observations located within a "patch" which is centered on the gridpoint and bounded by a specified cutoff radius. Whitaker et al. (2008) showed that sequential assimilation can also be applied within the patch framework described above. Ehrendorfer (2007) demonstrated mathematically that the simultaneous and sequential assimilation patterns are not equivalent when covariance localization is applied, as is the case with all practical implementations of the EnKF. Throughout this paper, the choice between a sequential or simultaneous assimilation method will be referred to simply as the choice of "assimilation pattern" for brevity.

Covariance localization (Hamill et al., 2001; Houtekamer and Mitchell, 2001) is used in the EnKF to ameliorate the impact on the analysis of spurious error correlations estimated between distant points by the ensemble covariances. While various covariance localizations have been proposed, in practice two approaches are most common. The first, which will be referred to as "B-localization," applies a localization function with

5

local support to the background error covariances. The second, which will be referred to as "R-localization," applies the inverse of the localization function to the observation error covariances, in effect increasing the error of observations as their distance from the gridpoint being updated increases.

The effect of using the B-localization or R-localization method is illustrated in Figure 1 for assimilation of a single observation. When equal cutoff distances at which the correlation function goes to zero are used for the two methods, the B-localization method yields less observation influence (as measured by the magnitude of the Kalman gain) than the R-localization method, as also shown by Miyoshi and Yamane (2007). However, when a cutoff distance 25% longer than that used for the R-localization is used for the B-localization, the observations are given less influence when using the B-localization than when using the R-localization for points near the observation location, but more influence when using the B-localization than when using the R-localization for points far from the observation location. This is consistent with Janjic et al. (2011), who noted that there are no two localization functions depending only on distance that can cause the B- and R-localization methods to produce the same increment.

Previous studies, using different models and experiment configurations, relevant to the effects of the sequential or simultaneous assimilation and/or covariance localization method on the performance of the EnKF include the following examples. Houtekamer and Mitchell (2001) found that a sequential scheme using the B-localization produced less accurate analyses than a simultaneous scheme using the B-localization, especially with a small ensemble size. Whitaker et al. (2008) obtained similar performance from a simultaneous scheme using the R-localization and a sequential

6

scheme using the B-localization. Kepert (2009) found that a sequential scheme using the

B-localization yielded better balanced analyses than a simultaneous scheme using the B-

localization, but found little difference in the analysis error of the sequential and

simultaneous schemes. Greybush et al. (2011) found similar analysis error and balance

performance from a sequential scheme using the B-localization, a sequential scheme

using the R-localization, and a simultaneous scheme using the R-localization. Janjic et

al. (2011) obtained the best analysis error performance from a simultaneous scheme using

the B-localization, intermediate performance from a sequential scheme using the B-

localization, and the worst performance from a simultaneous scheme using the R-

localization.

These previous studies did not reach a consensus, and the investigation of

interactions between the observation assimilation patterns and the localization methods

were limited to two or three schemes; the parameters for each scheme may not be tuned

optimally; or the studies were limited to a particular configuration. The goal of this study

was to conduct a more rigorous study of the effects of observation assimilation pattern

and localization methods and their interactions following these guidelines: First, six

schemes were used or designed. This allowed effects of assimilation pattern and

localization method were studied both in isolation and together. Second, all parameters

were optimally tuned to ensure that the best possible performance of each scheme was

being compared. Third, all schemes were compared in a variety of contexts in order to

allow a more complete understanding of the effect of the assimilation pattern and

localization method on the performance of the EnKF. A two-layer primitive equation

model and a simplified observation network were used in order to make the large number

of experiments needed to follow these guidelines computationally feasible. Future work

is needed to compare the schemes using a real NWP model and observation network.

The rest of the paper is organized as follows. Section 2 describes the design of

the several variations of the EnKF algorithm used herein and the experiment setup.

Section 3 presents the comparison of a B-localized serial, a B-localized simultaneous, an

R-localized serial, and two R-localized simultaneous schemes in an experiment using

forecast and assimilation cycles. Section 4 further examines the differences between these

schemes by using a non-cycling experiment, in which identical background ensembles

are given as input to each scheme. Section 5 compares the schemes in various contexts

including with and without digital filter initialization, with different observation

networks, with different ensemble sizes, and with different ratios of background error to

observation error. Section 6 provides a summary and draws together conclusions.

2. Methods

In order to allow the interaction of the choice of assimilation pattern and the

choice of localization method to be examined while also examining the effect of each

choice in isolation, the following schemes were used and designed: a B-localized serial

scheme (Bserial), an R-localized serial scheme (Rserial), a B-localized simultaneous

scheme (Bsimult), and an R-localized simultaneous scheme (Rsimult). The R-localized

simultaneous local ensemble transform Kalman filter (LETKF, Hunt et al., 2007) and the

"patch serial" scheme of Whitaker et al. (2008) were also examined. The main

properties of these schemes are summarized in Table 1 and described in more detail

below. The localization function used for all schemes in this study was the fifth-order

piecewise rational function given in Eq, (4.10) of Gaspari and Cohn (1999).

2.1 Bserial Scheme

The serial version of the ensemble square root filter (EnSRF, Whitaker and

Hamill, 2002) where observations are assimilated one at a time, is used as the Bserial

scheme. In this scheme, the analysis mean is generated using

$$\overline{\mathbf{x}}^{a} = \overline{\mathbf{x}}^{b} + \mathbf{K}(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^{b}) \, , \tag{1}$$

where $\overline{\mathbf{x}}^{a}$ is the analysis ensemble mean, $\overline{\mathbf{x}}^{b}$ is the background ensemble mean, $\mathbf{y}$

represents the observations, $\mathbf{H}$ is the observation operator used to convert from the model

space to the observation space, and $\mathbf{K}$ is the Kalman gain. For the Bserial scheme, the

Kalman gain is

$$\mathbf{K} = \boldsymbol{\rho}_{\mathbf{Bser}} \circ \mathbf{P}^{b}\mathbf{H}^{T}(\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T} + \mathbf{R})^{-1} \, , \tag{2}$$

where $\mathbf{P}^{b}$ is the background error covariance matrix estimated by the sample ensemble

covariance, $\mathbf{R}$ is the observation error covariance matrix, $\circ$ represents a Schur product, or

element-by-element multiplication, and $\boldsymbol{\rho}_{\mathbf{Bser}}$ is the covariance localization matrix,

calculated based on the distance between the gridpoints and the single observation being

assimilated. When generating the analysis ensemble perturbations, the Bserial scheme

uses

$$\mathbf{X}^a = \mathbf{X}^b - \tilde{\mathbf{K}}(\mathbf{H}\mathbf{X}^b),$$ (3)

where $\mathbf{X}^a$ is the analysis ensemble perturbation matrix, $\mathbf{X}^b$ is the background ensemble

perturbation matrix, and $\tilde{\mathbf{K}}$ is the "reduced" Kalman gain. For the Bserial scheme, where

observations are assimilated one at a time, this reduced Kalman gain is

$$\tilde{\mathbf{K}} = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}}}\right)^{-1}\mathbf{K} .$$ (4)

In the Bserial scheme, for the assimilation of the next observation, the $\bar{\mathbf{x}}^a$ and $\mathbf{X}^a$

resulting from assimilation of the previous observation become the $\bar{\mathbf{x}}^b$ and $\mathbf{X}^b$ in Eqs. (1)

and (3), and $\mathbf{X}^a$ is used to produce a new estimate of $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$ and $\mathbf{P}^b\mathbf{H}^T$. This process is

repeated until all observations have been assimilated.

2.2 Rserial Scheme

The Rserial scheme is created by modifying the Bserial as in Greybush et al.

(2011). Rather than applying the localization matrix to the background error covariance

as in Eq. (3), the inverse of the localization function computed based on the distance

between the gridpoint and observation, ρ_{Rser}, is applied to $\mathbf{R}$ in both the Kalman gain

10

$$\mathbf{K} = \mathbf{P}^{b}\mathbf{H}^{T}(\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T} + \rho_{\text{Rser}}^{-1} \circ \mathbf{R})^{-1} \tag{5}$$

and the reduced Kalman gain

$$\tilde{\mathbf{K}} = \left(1 + \sqrt{\frac{\rho_{\text{Rser}}^{-1} \circ \mathbf{R}}{\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T} + \rho_{\text{Rser}}^{-1} \circ \mathbf{R}}}\right)^{-1} \mathbf{K}. \tag{6}$$

In other words, Eq. (1) and Eq. (3) are used to update the ensemble mean and the

ensemble perturbations as in the Bserial scheme, except Eq. (5) is used in place of Eq. (2)

and Eq. (6) is used in place of Eq. (4).

2.3 Bsimult Scheme

In the Bsimult scheme, the patch framework used in the LETKF is adopted:

observations within the specified cutoff radius of a gridpoint are assimilated

simultaneously to update the state variable at the central grid point of the patch, and the

model state variables at each grid point are updated independently. In order to allow for

a B-localization, the simultaneous version of the EnSRF update equations (Eqs. (2), (4),

(5), and (10) of Whitaker and Hamill, 2002) are used, in which the background error

covariance matrices are explicitly calculated. The ensemble mean is updated using Eq.

(1) with the Kalman gain **K** given by

$$\mathbf{K} = \rho_{\mathbf{Bsim}} \circ \mathbf{P}^{b}\mathbf{H}^{T}(\gamma \circ \mathbf{HP}^{b}\mathbf{H}^{T} + \mathbf{R})^{-1}. \tag{7}$$

The values of the elements of $\rho_{\mathbf{Bsim}}$ in Eq. (7) are calculated based on the distance

between the gridpoint and the observations. The diagonal elements of the localization

matrix γ are 1, while the off-diagonal elements of γ are calculated based on the distance

between the observations. The ensemble perturbations are updated using Eq. (3) except

that $\tilde{\mathbf{K}}$ for the Bsimult scheme is

$$\tilde{\mathbf{K}} = \rho_{\mathbf{Bsim}} \circ \mathbf{P}^{b}\mathbf{H}^{T} \{(\sqrt{\gamma \circ \mathbf{HP}^{b}\mathbf{H}^{T} + \mathbf{R}})^{-1}\}^{T} \{\sqrt{\gamma \circ \mathbf{HP}^{b}\mathbf{H}^{T} + \mathbf{R}} + \sqrt{\mathbf{R}}\}^{-1}. \tag{8}$$

The square roots in Eq. (8) are calculated using the Cholesky decomposition method.

Note that the original serial EnSRF update method is derived from Eq. (7) and (8) for a

single observation.

2.4 Rsimult Scheme

Different from the Bsimult, the Rsimult applies the localization to $\mathbf{R}$ instead of to

the background error covariance matrices. Thus, the equivalents of Eq. (7) and (8) for the

Rsimult scheme are

$$\mathbf{K} = \mathbf{P}^{b}\mathbf{H}^{T}(\mathbf{HP}^{b}\mathbf{H}^{T} + \rho_{\mathbf{Rsim}} \circ \mathbf{R})^{-1}$$

12

and

$$\tilde{\mathbf{K}} = \mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}\{(\sqrt{\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \boldsymbol{\rho}_{\mathbf{Rsim}} \circ \mathbf{R}})^{-1}\}^{\mathrm{T}}\{\sqrt{\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \boldsymbol{\rho}_{\mathbf{Rsim}} \circ \mathbf{R}} + \sqrt{\boldsymbol{\rho}_{\mathbf{Rsim}} \circ \mathbf{R}}\}^{-1}, \quad (9)$$

in which $\boldsymbol{\rho}_{\mathbf{Rsim}}$ is the diagonal localization matrix with diagonal elements equal to the

element-wise inverse of the elements of $\boldsymbol{\rho}_{\mathbf{Bsim}}$ used in Eq. (7) and (8), and off-diagonal

elements equal to zero.

2.5 LETKF Scheme

Like the Rsimult scheme, the LETKF scheme (Hunt et al., 2007) utilizes the R-

localization and simultaneously assimilates all the observations in a patch. However, the

LETKF scheme uses a different set of equations than the Rsimult scheme to produce the

analyses. The mean update, which is equivalent to the Rsimult mean update, is given by

$$\overline{\mathbf{x}}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{b}} + \mathbf{X}^{\mathrm{b}}\overline{\mathbf{w}}^{\mathrm{a}},$$

where $\overline{\mathbf{w}}^{\mathrm{a}}$ is

$$\{(k-1)\mathbf{I} + (\mathbf{H}\mathbf{X}^{\mathrm{b}})^{\mathrm{T}}(\boldsymbol{\rho}_{\mathbf{Rsim}} \circ \mathbf{R})^{-1}(\mathbf{H}\mathbf{X}^{\mathrm{b}})\}^{-1}(\mathbf{H}\mathbf{X}^{\mathrm{b}})^{\mathrm{T}}(\boldsymbol{\rho}_{\mathbf{Rsim}} \circ \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^{\mathrm{b}}).$$

The perturbation update is given by

13

$$\mathbf{X}^{a} = \mathbf{X}^{b}\mathbf{W}^{a},$$

where

$$\mathbf{W}^{a} = [(k-1)\{(k-1)\mathbf{I} + (\mathbf{H}\mathbf{X}^{b})^{T}(\boldsymbol{\rho}_{\mathbf{Rsim}} \circ \mathbf{R})^{-1}(\mathbf{H}\mathbf{X}^{b})\}^{-1}]^{1/2}.$$

The matrix square roots in the LETKF are calculated using a symmetric square root method. Hunt et al. (2007) speculated that this symmetric approach might be superior to the Cholesky decomposition matrix square root method used in Eqs. (8) and (9). The performance of the Rsimult and LETKF schemes are compared here to explore whether these differences in the square root calculation create a performance difference between the two schemes.

2.6 Patch Serial scheme

Whitaker et al. (2008) used an EnSRF scheme that updated each model state variable and gridpoint independently using the patch framework of the LETKF. Within each patch a serial assimilation pattern was adopted. A scheme using this approach and R-localization (denoted as "patch Rserial") was designed, and its performance was compared to that of the Rserial scheme described above. No statistically significant performance differences were found between this patch Rserial scheme and the Rserial scheme, and so patch Rserial results are not shown in this paper.

14

2.7 Experimental Design

        The dry, two-layer, primitive-equation global spectral forecast model used for all

experiments herein is described in Zou et al. (1993). Originally used for studies of

atmospheric blocking, it has since been used in several ensemble-based data assimilation

experiments (e.g., Hamill et al., 2001; Hamill and Whitaker, 2005; Wang et al., 2007;

Wang et al., 2009). As mentioned in the introduction, a simplified model was chosen for

this study because its relatively low computational demands allowed all parameters to be

fully tuned and a large number of experiments to be run. The model state included

coefficients of three variables – vorticity, divergence, and layer thickness ($\Delta\pi$, where π is

the Exner function) – for each layer. Radiative heating and surface drag were treated

with simple parameterizations, and a wavenumber 2 topography was used. The same

model parameters were used as in Hamill and Whitaker (2005) and Wang et al. (2009).

Model integration was done using a fourth-order Runge-Kutta scheme, and the model

was run at T31 resolution.

        The performance of the schemes was first compared using data assimilation and

forecast cycling experiments. In these experiments, 250 cycles were conducted. A 24-

hour assimilation interval was chosen due to the relatively long (3.78 days at T31

truncation) error doubling time of the model following Hamill and Whitaker (2005). The

first 100 assimilation cycles were used to allow the system to stabilize, and only the final

150 cycles were used to compute the statistics reported here. The same ensemble of states

was used for each scheme in the initial cycle, chosen as a random draw from the states in

15

the truth run. To further understand the differences between the schemes, non-cycling

experiments were also conducted, as described in section 4.1.

For all experiments described here, following Hamill and Whitaker (2005) and

Wang et al. (2009), imperfect model observation system simulation experiments were

conducted. A model run using T127 resolution was used as the truth state. The

truncation error between the T127 truth state and T31 model resolution at which the data

assimilation experiments were run provided a source of model error. Observations were

generated from the truth state by adding errors drawn from a distribution with zero mean

and a fixed standard deviation defined as one-fourth of the climatological standard

deviation of the observed variable in the truth run following Wang et al. (2007). For the

experiments presented in section 3, 362 observations of interface height with root mean

square observation error of 250 m and 362 observations of surface π with observation

error of 0.875 J kg^{-1} K^{-1} were used. In later experiments, observations of the u-

component of the wind were used, with standard deviations of 1.25 m s^{-1} for lower-level

wind observations and 3.0 m s^{-1} for upper-level wind observations. The observations

were approximately evenly spaced over the globe on a geodesic grid (see Figure 2 of

Wang et al., 2007), and the errors were uncorrelated in time and space.

In order to address the proclivity of all EnKF schemes to underestimate

background error due to both a limited number of ensemble members and model errors,

both a multiplicative and an additive inflation were implemented. A multiplicative

covariance inflation factor was applied to the analysis ensemble perturbations at the end

of each assimilation step. A multiplicative adaptive inflation scheme (Wang and Bishop,

2003; Wang et al., 2007; Wang et al., 2009), which aims to keep consistent the

16

background ensemble variance and the ensemble mean background error variance in the

observation space on average, was used. A wide range of fixed inflation factors was also

tested to ensure that the adaptive inflation produced optimal results for each scheme. In

addition to the multiplicative inflation factor, following Wang et al. (2009) (see Eq. (4)

therein), an additive inflation was used, where a weighted combination of the forecast

ensemble perturbations and random draws from a historical inventory of 24-hour forecast

errors was used as $\mathbf{X}^b$. As noted by Wang et al. (2009), this additive inflation method

accounts for the typical non-uniform spatial distribution of forecast errors in the system,

which a globally uniform multiplicative inflation factor cannot. For each scheme, the

weighting between forecast perturbations and additive errors was tuned to produce the

smallest values of root-mean-square (rms) analysis error. All schemes in the current

study have the same optimal values of 40% weight on the additive error and 60% weight

on the forecast perturbations. 50 member ensembles were used throughout, with the

exception of Section 5.3, which explores the sensitivity of the differences among the

schemes to the variation in ensemble sizes.

In order to assess the performance of the various schemes and parameter settings,

root-mean-square (rms) analysis and forecast errors of π at the surface (analogous to

surface pressure) and the layer interface height were calculated from the ensemble mean.

Surface π tendency of the ensemble mean was also calculated, for use as a measure of the

imbalance of the analyses (Wang et al., 2009). Tendency was calculated as the absolute

value of the change in surface π over a 1-hour forecast initialized by each analysis.

Statistical significance of differences for all norms was calculated using results at

different times as replicates and a paired t-test that accounts for temporal correlations

17

between samples (Wilks, 2006). Differences at the 95% confidence level were considered statistically significant. Statistical significance was also presented as error bars in some figures. The error bars represent standard error that was calculated using a boostrap resampling technique that accounted for temporal correlation between samples, following Roulston and Smith (2003) and Wang and Bishop (2005). 100 resamples with replacement were used for these calculations.

3. Cycling Experiment Results

3.1 Analysis Error

Figure 2 shows the rms analysis error for the cycling experiment for the first five schemes described in Table 1. The localization cutoff distance that minimized rms error for all three R-localized schemes (Rsimult, Rserial, LETKF) was 4000 km for both surface π and interface height, while the optimal cutoff distances for the two B-localized schemes (Bsimult, Bserial) were 6000 km for surface π and 5000 km for interface height. The difference in the rms errors of the Rsimult and LETKF schemes was not statistically significant at the localization scales considered. The rms errors are further compared at the optimal localization scales: For both surface π and interface height, the Rsimult and LETKF schemes had lower rms error than the other three schemes when using the optimal localization cutoff distance for each. The Bsimult scheme rms error was statistically significantly higher than that of the other schemes: 26.4% higher than the Rsimult scheme rms error for the surface π norm and 6.8% higher for the interface height

18

norm. The rms errors of the two schemes using serial assimilation were intermediate

between the performance of the schemes using simultaneous assimilation, with the rms

error of the Bserial and Rserial being lower than that of the Bsimult scheme and higher

than the Rsimult scheme. The difference between the rms errors of the Bserial and the

Rserial schemes were statistically insignificant. In all cases in this paper, the wind field

rms error performance was similar to the interface height rms error performance, and so

wind field results are not shown.

The choice between the serial and simultaneous assimilation patterns produced

differences in the rms error. This difference was dependent on which localization method

was used, with the serial assimilation pattern producing lower rms error than the

simultaneous assimilation pattern when the B-localization was used and the simultaneous

assimilation pattern having lower rms error than the serial assimilation pattern when the

R-localization was used. The choice between using the B-localization or the R-

localization method produced differences in the rms error when a simultaneous

assimilation pattern was used. When the serial assimilation pattern was used, the choice

of localization method had a small, statistically insignificant impact on the rms error.

3.2 Imbalance

The presence of unrealistic imbalances between the mass and wind fields in the

analysis ensemble can produce gravity wave noise that negatively impacts the accuracy

of subsequent forecasts and analyses. As in Wang et al. (2009), the magnitude of the

surface π tendency was used here to assess the degree of imbalance introduced by each

19

scheme. Figure 3 shows the average 1-hour surface π tendency, calculated as described in section 2.7. Results shown are using the localization cutoff distance that produced the smallest surface π rms analysis error for each scheme in section 3.1. The Rsimult and LETKF schemes produced smaller surface π tendency than the other three schemes. The surface π tendency of the Bsimult scheme was the highest among the five schemes: 102% greater than that of the Rsimult scheme. The surface π tendency of the Rserial and the Bserial were intermediate between the Bsimult and Rsimult. The order of the schemes from the least to the most imbalance matched the order from the smallest to the largest surface π rms error in section 3.1. All schemes were less balanced than the truth state.

Similar to the rms analysis error comparison, the imbalance of the analyses was affected both by the choice of assimilation pattern and the choice of localization method and their interaction. The serial assimilation pattern produced more imbalance than the simultaneous method when R-localization was used, but less imbalance than the simultaneous method when B-localization was used. The R-localization method produced statistically significantly less imbalance than the B-localization method only when the simultaneous assimilation pattern was used. As was the case for rms errors, the difference between the B- and R-localized schemes was much greater when the simultaneous assimilation pattern was used than when the serial assimilation pattern was used.

3.3 Forecast

20

In order to further assess the differences between different schemes, the means of the analysis ensembles were used to initialize 10-day forecasts. Figure 4 shows the rms forecast error for the surface π norm averaged over the 150 forecasts. The corresponding result for the forecast launched from the truth state is also shown. The result for the LETKF is not shown because the Rsimult and LETKF scheme were similar. The difference of the rms error between the schemes increased early in the forecast and then decreased. The rms forecast error was the highest when the Bsimult scheme was used and the lowest when the Rsimult scheme was used. The rms error of both the Bserial and Rserial schemes was smaller than the rms error of the Rsimult scheme and larger than that of the Bsimult scheme at all times. The difference of the rms forecast error between the Bserial and Rserial schemes was not statistically significant.

The contributions of the choice of assimilation pattern and the choice of localization method to the forecast error were examined by comparing the forecast error differences between the four schemes to the contribution of model error to the forecast error. The contribution of the model error to the forecast error was represented by the rms error of the truth-initialized forecast. The contribution of initial condition error to forecast error was larger than the contribution of model error until day 2, after which the model error contribution was larger. The magnitude of the rms error differences between the serial and simultaneous schemes was less than 10% of the size of the model error contribution to forecast error at lead times longer than 4-days. The magnitude of the rms error differences between the B- and R-localized simultaneous schemes was less than 10% of the size of the model error contribution to forecast error at lead times longer than 6-days.

21

4. Understanding the differences through non-cycling experiments

4.1 Analysis error and imbalance

A non-cycling experiment was set up in order to better understand the differences observed in section 3. In this experiment, the assimilation was performed using the same background ensemble input for all four schemes. It was hoped that removing the complication of feedbacks from the forecast/analysis cycle would simplify the process of exploring why the four schemes generated different analyses. 150 non-cycling data assimilations were performed for each scheme using the background ensemble of the Bserial cycling experiment. Repetition of the experiments using the background ensemble of the Rsimult cycling experiments produced similar results (not shown). The optimal localization cutoff distances of each scheme from the cycling experiment were used for this experiment.

As in the cycling experiment, the accuracy and imbalance of the analyses were compared using rms error and surface π tendency, respectively. The comparison of the rms analysis error of the four schemes for the surface π and interface height norms are shown in Figure 5. Unlike the cycling experiment, in the non-cycling case, there were no statistically significant differences in the rms analysis error among the schemes.

Figure 6 shows the 1-hour surface π tendency for the four schemes in the non-cycling experiments, averaged over 150 cases. The Rsimult scheme produced the smallest surface π tendency, followed in order of increasing surface π tendency by the

22

two serial schemes, and Bsimult schemes. The difference between Bserial and Rserial is not statistically significant. This order matches the order from the smallest to the largest surface π tendency in the cycling experiment (Figure 3). The differences in imbalance between the schemes in the non-cycling experiment were smaller than those in the cycling experiment, which suggests that the differences in the imbalance were accumulating during the cycling experiment. Given that the rms analysis error differences in the non-cycling experiment were smaller than in the cycling experiment, and that the order of the schemes from the least to the most imbalanced in the non-cycling experiment matches the order from smallest to largest surface π rms analysis error seen in the cycling experiment, the results suggest that the rms error differences in the cycling experiment in section 3.1 are associated with imbalance differences that accumulate during the data assimilation cycling.

4.2 Further exploration of balance problems

To better understand the differences in the imbalance of the analyses shown in Figure 6, the mass and wind analyses of two pairs of the schemes from the non-cycling experiments were examined further. The mass and wind analyses of the Rserial and Rsimult schemes were compared to study the possible contribution of the choice of assimilation pattern to the differences in the imbalance, and those of the Bsimult and Rsimult schemes were compared to study the possible contribution of the choice of localization method. Bserial and Rserial were not further compared here in section 4.2

because results in previous sections show their difference is much smaller or not statistically significant.

Inconsistency of the height gradient and wind increments has been identified by previous studies as a potential cause of imbalance in the EnKF schemes with covariance localization (Lorenc, 2003; Miyoshi and Yamane, 2007; Kepert, 2009). For example, steepening of height gradients and reduction of wind fields can contribute to creating sub-geostrophic flow in the analysis. Figure 7 compares the surface π tendency, wind, and height gradient analysis increments from the non-cycled Bsimult and Rsimult schemes and the non-cycled Rserial and Rsimult experiments. The "surface π tendency increment" is defined as the magnitude of the surface π tendency of a 1-hour forecast initialized using the analysis ensemble mean minus the magnitude of the surface π tendency of a 1-hour forecast initialized using the background ensemble mean. In other words, the surface π tendency increment is used to measure the change in imbalance due to the assimilation. The "wind increment" is defined as the absolute value of the v-component of analysis ensemble mean wind minus the absolute value of the v-component of the background ensemble mean wind, while the "height gradient increment" is defined as the magnitude of the longitudinal gradient of the analysis ensemble mean layer thickness field minus the magnitude of the longitudinal gradient of the background ensemble mean layer thickness field.

Figure 7a shows the difference of the surface π tendency increment between the Bsimult and Rsimult as a function of latitude for the southern latitudes. Results for northern latitudes are similar in all cases. Most of the difference in the surface π tendency increment between the two schemes occurred in the mid- and high latitudes, where the

24

Bsimult scheme surface π tendency increment was higher than that of the Rsimult

scheme. Figures 7b and 7c show the difference in the wind and height gradient analysis

increments between the Bsimult and Rsimult schemes at the upper and lower levels,

respectively. The Bsimult scheme had a weaker wind increment than the Rsimult scheme

and a stronger height gradient increment in the mid- and high latitudes, which is

consistent with the larger surface π tendency increment found in the Bsimult at mid- and

high latitudes (Figure 7a). This result was found at both upper and lower levels.

Similar diagnostics were performed comparing the Rserial and Rsimult schemes

(Figure 7d-f). Figure 7d shows that the Rserial scheme surface π tendency increment was

larger than that of the Rsimult scheme in the mid- and high latitudes. This difference was

not as large as the difference between the Bsimult and Rsimult schemes shown in Figure

7a, consistent with the result in the cycling experiment where rms analysis error

differences seen in Figure 3 between the Rserial and Rsimult schemes were smaller than

those between the Bsimult and Rsimult schemes. Similar to Figure 7b, Figure 7e shows

that at the lower level, the wind increment of the Rserial scheme was weaker than the

Rsimult, while the height gradient increment of the Rserial was stronger than the Rsimult.

Different from Figure 7c, Figure 7f shows that in the upper level, the Rserial scheme had

both stronger mid- and high-latitude wind increments and stronger height gradient

increments than the Rsimult scheme. This result suggests that while both levels

contributed to Bsimult being more imbalanced than Rsimult, the higher level was less of

a contributor to Rserial being more imbalanced than Rsimult than the lower level. The

greater imbalance observed in the Bsimult than in the Rserial (Figures 3 and 6) is also

consistent with this result.

25

In an effort to gain more insight into why the increments of the wind and height gradient fields responded differently to the choice of observation assimilation pattern and the choice of localization method, experiments were conducted using a simple simulation in which parameters such as the distance from an observation at which the observation influence was largest, the rate at which the correlation between points changed with distance, and the ratio of $\mathbf{HP^bH^T}$ to $\mathbf{R}$ could be easily manipulated. These experiments (not shown) suggested that while all of those factors can shape whether the choice of assimilation pattern and localization method will affect different model fields in different ways, the peak observation influence on the model state occurred at different distances from the observation location for the wind and height fields might have been the primary reason why those model fields were affected differently by the choice of assimilation pattern and localization method.

5. Comparison of schemes in different situations

To further understand the impacts of the sequential and simultaneous of assimilation of observations and the localization methods on the performance of the EnKF, the comparison of the Bserial, Bsimult, Rserial, and Rsimult schemes was repeated in a variety of contexts by modifying the setup of the cycling experiment described in section 3.

5.1 Digital Filter Initialization

Digital filter initialization (DFI, Lynch and Huang, 1992), in which a low-pass digital filter is applied to a series of model integration time steps, has been used in several studies (e.g., Houtekamer and Mitchell, 2005; Whitaker et al., 2008) to remove the unrealistic imbalance and small-scale noise in the analyses that could damage the subsequent forecast. Given that differences in imbalance appear to be a cause of the accuracy differences described in section 3.1, it was hypothesized that the use of the DFI might decrease the performance differences introduced by choice of assimilation pattern and localization method. To test this, the cycling experiment from section 3 was repeated with the DFI implemented. The cutoff period, which determines the frequencies that are removed by the filter, was optimally tuned for each scheme, and a 6-hour period was found to produce the smallest rms errors for all cases. As in Fillion et al. (1995), rather than using a 3 hour forecast and 3 hour hindcast to produce a filtered state at the analysis time, the model was integrated forward for 6-hours and the filter was applied to the time steps in that period, producing a filtered state at analysis time plus 3 hours.

Figure 8 shows that the surface π and interface height rms error were reduced for all of the schemes when the DFI was used. For both the surface π and the interface height, the schemes with more rms error in the no-DFI experiment showed more reduction in rms error when the DFI was applied than those schemes with less rms error in the no-DFI experiments. Thus, the differences between all four of the schemes were smaller when the DFI was applied, with the differences in the interface height and wind rms errors being statistically insignificant.

5.2 Type and number of observations

27

Several additional experiments were performed to explore the impact of the

assimilation pattern and localization methods on the performance of the EnKF when

different types and numbers of observations were used.   First, in addition to the

experiment in section 3 where interface height and surface π observations were

assimilated, more experiments were performed where in addition to interface height and

surface π observations, upper- and lower-level wind observations were also assimilated.

Table 2 shows the surface π and interface height rms analysis errors of the original

reference experiment and the experiment that also assimilated wind observations.

Compared to the reference experiment, when wind observations were assimilated in

addition to the interface and surface π observations, the relative rms error differences

between the schemes were reduced. It is suggested that with the wind observations, rather

than solely relying on the mass-wind cross variable covariance, the wind fields were

updated directly by wind observations, which reduced the imbalance differences among

the schemes than in the reference experiment when only mass observations are

assimilated.

Another experiment was conducted to study the impact of the choice of

assimilation pattern and localization method on the performance of the EnKF when the

number of observations was reduced. Compared to the reference experiment in section 3

where 362 interface height and 362 surface π observations were assimilated, in this

experiment, 161 observations of each type were assimilated. The number and location of

the observations in the new experiment were chosen in order to produce a geodesic grid

with nearly uniform observation spacing as in the reference experiment. Figure 9 shows

that for both the surface π and interface height norms, the rms error of all of the schemes

was larger when fewer observations were assimilated than when more observations were

assimilated. The differences between the schemes were in general smaller when fewer

observations were assimilated, especially for the schemes (e.g., Bsimult vs. Rsimult,

Rserial vs. Rsimult) that show relatively large differences in the reference experiment.

The reduction of the differences among the schemes is presumably because of the

reduction of sampling error when fewer, same type of observations are assimilated for a

given ensemble size. In such a case, the localization, designed to ameliorate the sampling

error has less of an effect.

5.3 Ensemble size

        As discussed in section 1, covariance localization provides a way to ameliorate

the sampling error of the EnKF due to the use of a limited number of ensemble members.

The imbalance caused by the covariance localization can be dependent on the size of the

ensemble. In order to examine the effect of ensemble size on the differences between the

schemes that use different assimilation patterns or localization methods, cycling

experiments with 100 and 200 member ensembles were conducted in addition to the

reference experiment described in section 3.1 that used a 50 member ensemble. Figure

10 shows that as ensemble size increased, the rms error of all of the schemes was

reduced, and the relative difference in rms error between all of the schemes decreased.

The optimal localization cutoff distance for each of the schemes also increased as

ensemble size increased (not shown). This result indicates that the effect of the sequential

29

and simultaneous assimilation pattern and the localization method choices decrease as

ensemble size increases, and vice versa.

5.4 The ratio of background error variance to observation error variance

If the same localization function is used for both the B-localization and the R-

localization methods, then for a case of a single observation updating a single grid point,

Eqs. (2) and (5) can be used to show that

$$\frac{\mathbf{K}_r}{\mathbf{K}_b} = \frac{\dfrac{\mathbf{HP}^b\mathbf{H}^T}{\mathbf{R}} + 1}{\rho \circ \left(\dfrac{\mathbf{HP}^b\mathbf{H}^T}{\mathbf{R}}\right) + 1} \,, \tag{10}$$

where $\mathbf{K}_r$ is the Kalman gain resulting from use of the R-localization method and $\mathbf{K}_b$ is

the Kalman gain resulting from use of the B-localization method. In other words, Eq.

(10) shows that the difference between the observation influence produced by using the

B- or R-localization is a function of the value of the localization function, ρ, and the ratio

of $\mathbf{HP}^b\mathbf{H}^T$ to $\mathbf{R}$. Consistent with the two solid lines in Figure 1, Eq. (10) indicates that

unless $\rho = 1$, use of the R-localization method will yield more observation influence than

use of the B-localization method. Note also that this tendency for the R-localization to

yield more observation influence will become stronger as the ratio of $\mathbf{HP}^b\mathbf{H}^T$ to $\mathbf{R}$

increases. In reality, of course, the assumption of Eq. (10) that the same localization

function is used for both the B- and R-localization methods need not be true. Figure 11

30

shows the relative observation influence produced by the B-localization and the R-localization methods for a case in which a longer localization cutoff distance is used for the B-localization method than for the R-localization method, as in the experiment in section 3. As in the same-localization-scale case of Eq. (10), Figure 11 suggests that the observation influence depends on the ratio of $\mathbf{HP^bH^T}$ to $\mathbf{R}$. Figure 11 also shows that the difference between the observation influence when using the B- or R-localization method is larger for a grid point further away (solid line) from the observation. To further explore the influence of the $\mathbf{HP^bH^T}$ to $\mathbf{R}$ ratio on the effects of the choice of the localization method, an experiment was conducted in which the $\mathbf{HP^bH^T}$ to $\mathbf{R}$ ratio was effectively modified by using different observation errors. In this experiment, compared to the reference experiment in section 3, observations with errors twice as large were used, causing a decrease in the average ratio of $\mathbf{HP^bH^T}$ to $\mathbf{R}$. Table 3 shows that decreasing the ratio of $\mathbf{HP^bH^T}$ to $\mathbf{R}$ caused a decrease in the relative rms analysis error difference between the Bsimult scheme and the Rsimult scheme analyses for both the surface π and interface height.

The dependence of the difference between the Bsimult and Rsimult schemes on the ratio of $\mathbf{HP^bH^T}$ to $\mathbf{R}$ may explain the latitudinal variation of the difference in the wind increments between the Bsimult and Rsimult schemes seen in Figures 7b and 7c. In Figure 11, when the ratio of $\mathbf{HP^bH^T}$ to $\mathbf{R}$ is sufficiently small, the observation influence using the B-localization method is larger than that using the R-localization, with the reverse being true when the ratio is sufficiently large. In Figures 7b and 7c, the wind increments of the Bsimult scheme were weaker than the Rsimult scheme in the mid- and upper latitudes, but were stronger than the Rsimult scheme increments in the tropics. The

ratio of $\mathbf{HP}^b\mathbf{H}^T$ to $\mathbf{R}$ in the results shown in Figure 7 varied from a maximum of

approximately 2.8 in the mid-latitudes, which Figure 11 suggests would lead to a larger

update by the R-localized scheme, to an average of approximately 0.8 in the tropics,

which Figure 11 suggests would lead to a larger update by the B-localized scheme. The

latitudinal change in the ratio of $\mathbf{HP}^b\mathbf{H}^T$ to $\mathbf{R}$ was due to the fact that the background

ensemble error variance was much larger in the mid-latitudes than in the tropics, while

the observation error variance was the same everywhere.

The ratio of $\mathbf{HP}^b\mathbf{H}^T$ to $\mathbf{R}$ may also explain why the difference in performance

between the B- and R-localization methods was smaller in the serial case than in the

simultaneous case. With the assimilation of each additional observation in a serial

assimilation case, the ratio of $\mathbf{HP}^b\mathbf{H}^T$ to $\mathbf{R}$ is incrementally decreased as ensemble spread

is reduced. Thus, later observations are assimilated with a smaller ratio of $\mathbf{HP}^b\mathbf{H}^T$ to $\mathbf{R}$,

which previous studies and our own experiments suggest may decrease some of the

differences between the B-localization and the R-localization methods.

6. Conclusions

The main goal of this study was to explore the effect of two common variations

among EnKF schemes and their interactions: assimilating observations either sequentially

or simultaneously, and applying covariance localization either to the background error

covariance matrices (B-localization) or to the observation error covariance matrix (R-

localization), on the performance of the EnKF. In order to examine the impact of these

two choices both separately and in combination with each other, six schemes were used

32

and designed: a B-localized sequential EnSRF, an R-localized sequential EnSRF, a B-localized simultaneous EnSRF, an R-localized simultaneous EnSRF, the R-localized simultaneous LETKF scheme, and an R-localized sequential scheme that used an LETKF-style patch framework. These schemes were compared using a data assimilation cycling experiment using a primitive equation two-layer model with simulated observations and an imperfect model assumption. In this first set of experiments, the R-localized simultaneous scheme produced the least analysis error, forecast error, and imbalance, while the B-localized simultaneous scheme produced the most analysis error, forecast error, and imbalance. The B-localized and R-localized sequential schemes produced similar results to each other that were intermediate between the performances of the two simultaneous schemes. Neither the algorithmic difference between the LETKF and the R-localized simultaneous EnSRF, nor whether or not a "patch" assimilation method was used in the sequential case had a statistically significant impact on EnKF performance.

Next, non-cycling experiments were conducted in which the same background ensemble was used as input for each scheme. In this case, there were no statistically significant differences between the schemes in terms of the rms analysis errors, but the differences in the imbalance, as measured by surface π tendency, were still observed. The order of the schemes from the least to the most imbalanced in the non-cycling experiment was the same as the cycling experiment. These results suggested that the differences in the analysis accuracy between the schemes in the cycling experiments were associated with the feedback of the amount of imbalance introduced by the different schemes over time. Further examination of the analysis increments in the non-cycling

33

experiments showed that the schemes that generated more imbalance had stronger height gradient increments and weaker wind increments than those that generated less imbalance.

To further understand the differences caused by the choice of sequential or simultaneous observation assimilation and the covariance localization method, several cycling experiments were conducted that compared the schemes in different contexts. The differences between all schemes were reduced when a digitial filter initialization step was used to reduce analysis imbalance, when wind observations were assimilated in conjunction with mass observations, when the number of observations of the same type within the range considered was decreased, and when the ensemble size was increased. The difference between the B- and R-localized simultaneous schemes was reduced when the characteristic ratio of $\mathbf{HP^bH^T}$ to $\mathbf{R}$ in the system was reduced by increasing observation errors.

In this study, a much simpler forecast system and observation network than would be found in real-world applications were used in order to allow many experiments to be performed and all parameters to be optimally tuned for each scheme. While various sensitivity experiments were conducted to understand the effects of choice of assimilation pattern and localization method in a range of situations, caution is warranted when attempting to extrapolate the results to real-world applications of the EnKF. For example, this study showed that the differences caused by the choice of assimilation and localization methods may be dependent on observation type and number, characteristic $\mathbf{HP^bH^T}$ to $\mathbf{R}$ ratio, and whether digital filter localization is used. The effects of these

34

choices and the root causes of the differences of the schemes may differ between global,

regional, and storm-scale applications.

Acknowledgements

References

Aksoy A, Dowell DC and Snyder C. 2009. A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part I: Storm-Scale Analyses. *Mon Weather Rev* 137: 1805-1824.

Bonavita M, Torrisi L and Marcucci F. 2008. The ensemble Kalman filter in an operational regional NWP system: Preliminary results with real observations. *Q J Roy Meteor Soc* 134: 1733-1744.

Bonavita M, Torrisi L and Marcucci F. 2010. Ensemble data assimilation with the CNMCA regional forecasting system. *Q J Roy Meteor Soc* 136: 132-145.

Campbell WF, Bishop CH and Hodyss D. 2010. Vertical Covariance Localization for Satellite Radiances in Ensemble Kalman Filters. *Mon Weather Rev* 138: 282-290.

Charron M, Pellerin G, Spacek L, Houtekamer PL, Gagnon N, Mitchell HL and Michelin L. 2010. Toward Random Sampling of Model Error in the Canadian Ensemble Prediction System. *Mon Weather Rev* 138: 1877-1901.

Chen YS and Snyder C. 2007. Assimilating vortex position with an ensemble Kalman filter. *Mon Weather Rev* 135: 1828-1845.

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Bronnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli O, Ross TF, Trigo RM, Wang XL, Woodruff SD and Worley SJ. 2011. The Twentieth Century Reanalysis Project. *Q J Roy Meteor Soc* 137: 1-28.

Dowell DC and Wicker LJ. 2009. Additive Noise for Storm-Scale Ensemble Data Assimilation. *J Atmos Ocean Tech* 26: 911-927.

Ehrendorfer M. 2007. A review of issues in ensemble-based Kalman filtering. *Meteorol Z* 16: 795-818.

Evensen G. 1994. Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistics. *J Geophys Res-Oceans* 99: 10143-10162.

Fillion L, Mitchell HL, Ritchie H and Staniforth A. 1995. The Impact of a Digital-Filter Finalization Technique in a Global Data Assimilation System. *Tellus A* 47: 304-323.

Gaspari G and Cohn SE. 1999. Construction of correlation functions in two and three dimensions. *Q J Roy Meteor Soc* 125: 723-757.

Greybush SJ, Kalnay E, Miyoshi T, Ide K and Hunt BR. 2011. Balance and Ensemble Kalman Filter Localization Techniques. *Mon Weather Rev* 139: 511-522.

Hamill TM and Whitaker JS. 2005. Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon Weather Rev* 133: 3132-3147.

Hamill TM, Whitaker JS, Fiorino M and Benjamin SG. 2011. Global Ensemble Predictions of 2009's Tropical Cyclones Initialized with an Ensemble Kalman Filter. *Mon Weather Rev* 139: 668-688.

Hamill TM, Whitaker JS and Snyder C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon Weather Rev* 129: 2776-2790.

Houtekamer PL and Mitchell HL. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon Weather Rev* 129: 123-137.

Houtekamer PL and Mitchell HL. 2005. Ensemble Kalman filtering. *Q J Roy Meteor Soc* 131: 3269-3289.

Hunt BR, Kostelich EJ and Szunyogh I. 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D* 230: 112-126.

Janjic T, Nerger L, Albertella A, Schroter J and Skachko S. 2011. On Domain Localization in Ensemble-Based Kalman Filter Algorithms. *Mon Weather Rev* 139: 2046-2060.

Kepert JD. 2009. Covariance localisation and balance in an Ensemble Kalman Filter. *Q J Roy Meteor Soc* 135: 1157-1176.

Lorenc AC. 2003. The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var. *Q J Roy Meteor Soc* 129: 3183-3203.

Lynch P and Huang XY. 1992. Initialization of the Hirlam Model Using a Digital-Filter. *Mon Weather Rev* 120: 1019-1034.

Meng ZY and Zhang FQ. 2008. Tests of an Ensemble Kalman Filter for Mesoscale and Regional-Scale Data Assimilation. Part IV: Comparison with 3DVAR in a Month-Long Experiment. *Mon Weather Rev* 136: 3671-3682.

Miyoshi T, Sato Y and Kadowaki T. 2010. Ensemble Kalman Filter and 4D-Var Intercomparison with the Japanese Operational Global Analysis and Prediction System. *Mon Weather Rev* 138: 2846-2866.

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

Ott E, Hunt BR, Szunyogh I, Zimin AV, Kostelich EJ, Corazza M, Kalnay E, Patil DJ and Yorke JA. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A* 56: 415-428.

Roulston MS and Smith LA. 2003. Combining dynamical and statistical ensembles. *Tellus A* 55: 16-30.

Snyder C and Zhang FQ. 2003. Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon Weather Rev* 131: 1663-1677.

Tong MJ and Xue M. 2005. Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon Weather Rev* 133: 1789-1807.

Torn RD and Hakim GJ. 2009. Ensemble Data Assimilation Applied to RAINEX Observations of Hurricane Katrina (2005). *Mon Weather Rev* 137: 2817-2829.

Wang X. 2011. Application of the WRF hybrid ETKF-3DVAR data assimilation system for hurricane track forecasts. *Weather Forecast*.

Wang XG, Barker DM, Snyder C and Hamill TM. 2008a. A Hybrid ETKF-3DVAR Data Assimilation Scheme for the WRF Model. Part I: Observing System Simulation Experiment. *Mon Weather Rev* 136: 5116-5131.

Wang XG, Barker DM, Snyder C and Hamill TM. 2008b. A Hybrid ETKF-3DVAR Data Assimilation Scheme for the WRF Model. Part II: Real Observation Experiments. *Mon Weather Rev* 136: 5132-5147.

Wang XG and Bishop CH. 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J Atmos Sci* 60: 1140-1158.

Wang XG and Bishop CH. 2005. Improvement of ensemble reliability with a new dressing kernel. *Q J Roy Meteor Soc* 131: 965-986.

Wang XG, Hamill TA, Whitaker JS and Bishop CH. 2007. A comparison of hybrid ensemble transform Kalman filter-optimum interpolation and ensemble square root filter analysis schemes. *Mon Weather Rev* 135: 1055-1076.

Wang XG, Hamill TM, Whitaker JS and Bishop CH. 2009. A Comparison of the Hybrid and EnSRF Analysis Schemes in the Presence of Model Errors due to Unresolved Scales. *Mon Weather Rev* 137: 3219-3232.

Whitaker JS, Compo GP, Wei X and Hamill TM. 2004. Reanalysis without radiosondes using ensemble data assimilation. *Mon Weather Rev* 132: 1190-1200.

Whitaker JS and Hamill TM. 2002. Ensemble data assimilation without perturbed observations. *Mon Weather Rev* 130: 1913-1924.

Whitaker JS, Hamill TM, Wei X, Song YC and Toth Z. 2008. Ensemble data assimilation with the NCEP Global Forecast System. *Mon Weather Rev* 136: 463-482.

Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*. Academic Press.

Zhang F, Weng Y, Sippel JA, Meng Z and Bishop CA. 2009. Cloud-resolving Hurricane Initialization and Prediciton through Assimilation of Doppler Radar Observations with an Ensemble Kalman Filter: Humberto (2007). *Mon Weather Rev* Submitted: Submitted.

Zou X, Barcilon A, Navon IM, Whitaker J and Cacuci DG. 1993. An Adjoint Sensitivity Study of Blocking in a 2-Layer Isentropic Model. *Mon Weather Rev* 121: 2833-2857.

List of Tables

List of Figures

localization scale in each case are shown, and averages are over 150 cases. Error
bars are standard error calculated using a bootstrap resampling method.

Figure 10. Average rms analysis error from the baseline cycling experiment of section 3.1
using 50 members and experiments using 100-member and 200-member ensembles.
(a) Surface π and (b) interface height results are shown. The results using the
optimal localization scale in each case are shown, and averages are over 150 cases.
Error bars are standard error calculated using a bootstrap resampling method.

Figure 11. Ratio of the observation influences produced using R-localization and B-
localization in a single observation case, shown as a function of the ratio of the
background error variance to the observation error variance. Ordinate values greater
than 1 indicate that the R-localization method yields more observation influence,
and ordinate values less than one indicate that the B-localization yields more
observation influence. Optimal localization cutoff distances for the surface π rms
error in the cycling experiment are used (4000 km for R-localization, 6000 km for
B-localization). Two cases are shown: one in which the model grid point being
updated is 1000 km from the observation location and one in which the model grid
point is 1500 km from the observation location.

41

Tables

Scheme	Assimilation Pattern	Localization	Algorithm
Bserial	Serial	B	EnSRF
Rserial	Serial	R	EnSRF
Bsimult	Simultaneous	B	EnSRF
Rsimult	Simultaneous	R	EnSRF
LETKF	Simultaneous	R	LETKF
Patch Rserial	Serial	R	EnSRF

Table 1. Summary of schemes used in this study, as described in the text.

Observations Used	Rserial vs. Rsimult		Bsimult vs. Bserial		Bsimult vs. Rsimult	
	Surface π	Interface Height	Surface π	Interface Height	Surface π	Interface Height
Interface height, surface π	**12.9%**	**2.8%**	**11.1%**	**3.0%**	**26.4%**	**6.8%**
Interface height, surface π, wind	**7.2%**	0.9%	**10.8%**	**2.1%**	**18.4%**	**3.0%**

Table 2. Percent surface π and interface height rms error difference between three pairs of schemes using either interface height and surface π observations (original baseline experiment, first row) or interface height, surface π, and upper and lower level meridional wind observations (second row). Positive numbers indicate that the first scheme listed had higher rms error. Results shown used the optimal localization scale for each scheme. Significant differences are shown in bold.

	Original experiment	Reduced $\mathbf{HP^bH^T/R}$ ratio
Surface π rms error	20.9%	17.5%
Interface height rms error	7.1%	2.3%

Table 3. Percent rms analysis error difference between the Bsimult and Rsimult schemes using either a large or small ratio of background error variance to observation error variance. The first column is the result of the original baseline experiment and the second column is the result after reducing the ratio of the background error variance to observation error variance. Positive numbers indicate that the Bsimult scheme had higher error. All differences were statistically significant. See text for more details.

Figures
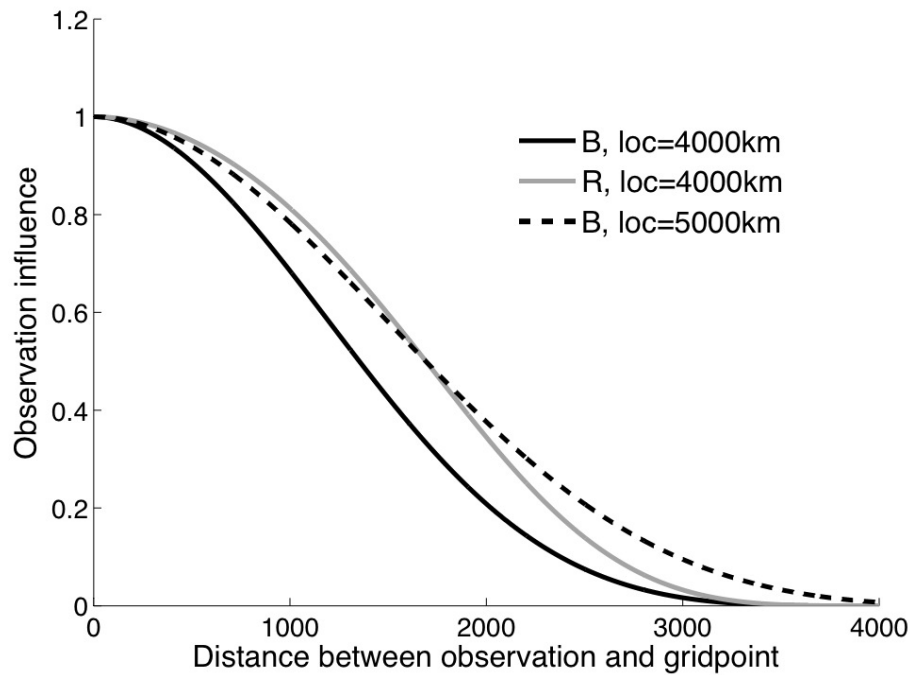
Figure 1. Illustration of the B-/R-localization difference using a simple one-observation example. The observation influence plotted on the y-axis is the Kalman gain normalized to have a maximum value of 1. Solid lines show B-localization (black) and R-localization (gray) using the same 4000km localization length scale. Dotted line shows B-localization using a localization length scale 25% longer than in the other two cases.
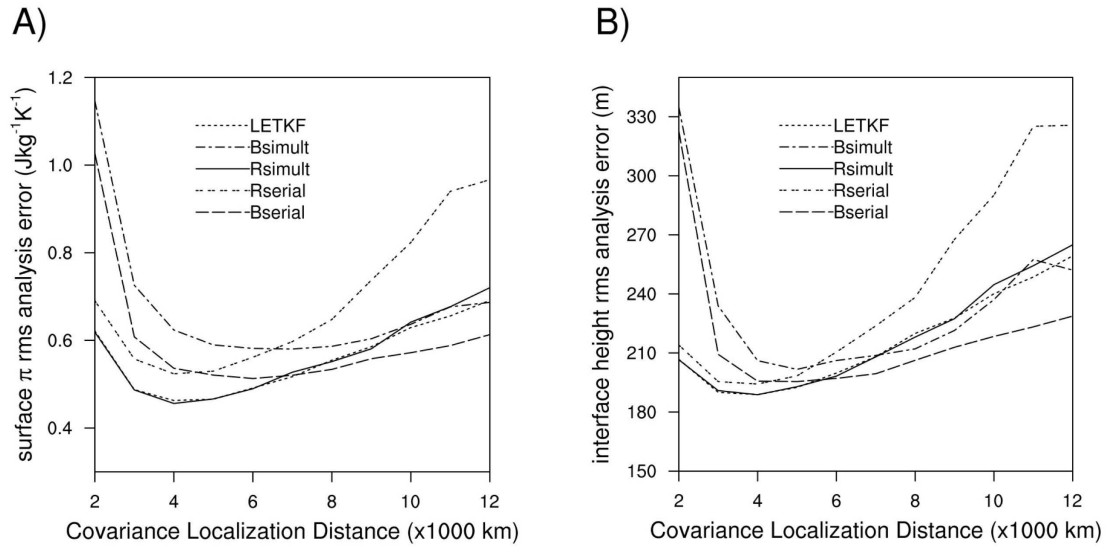
Figure 2. (a) Surface π and (b) interface height analysis error for the four combinations of serial/simultaneous assimilation and B-/R-localization and the LETKF, plotted as a function of covariance localization cutoff distance. See text for more details.
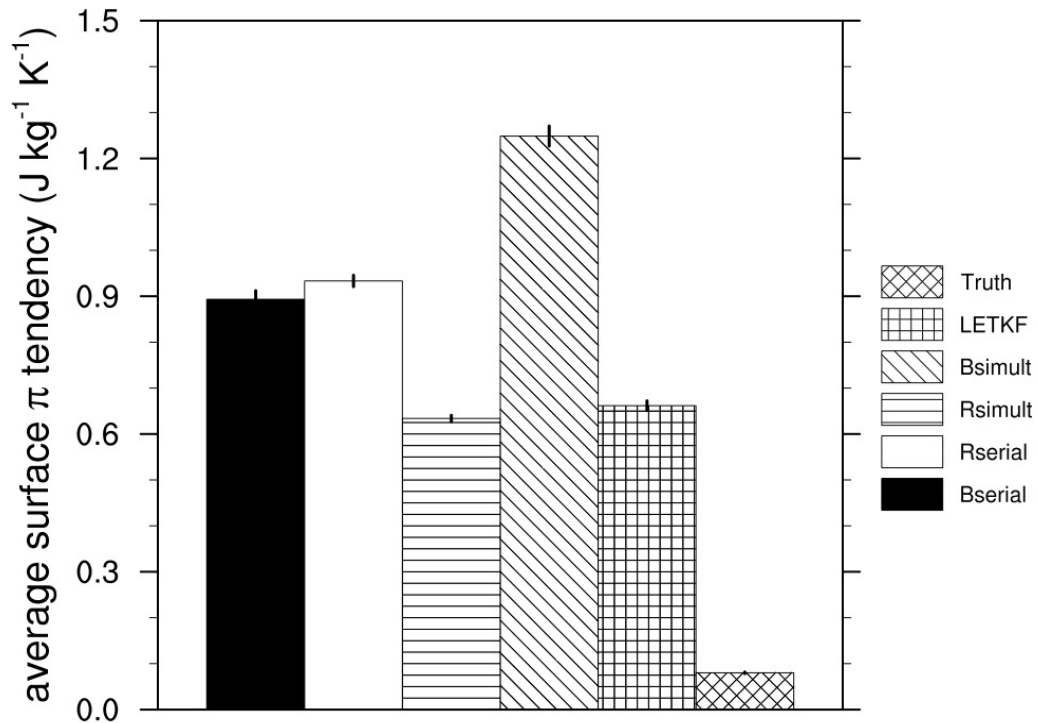
Figure 3. Average of absolute 1-hour surface π tendency values for the five schemes depicted in Figure 2. Truth tendency is the hourly tendency of the truth run. Averages shown are of 150 analyses and error bars show standard error using a bootstrap resampling method. See text for details of tendency calculation.
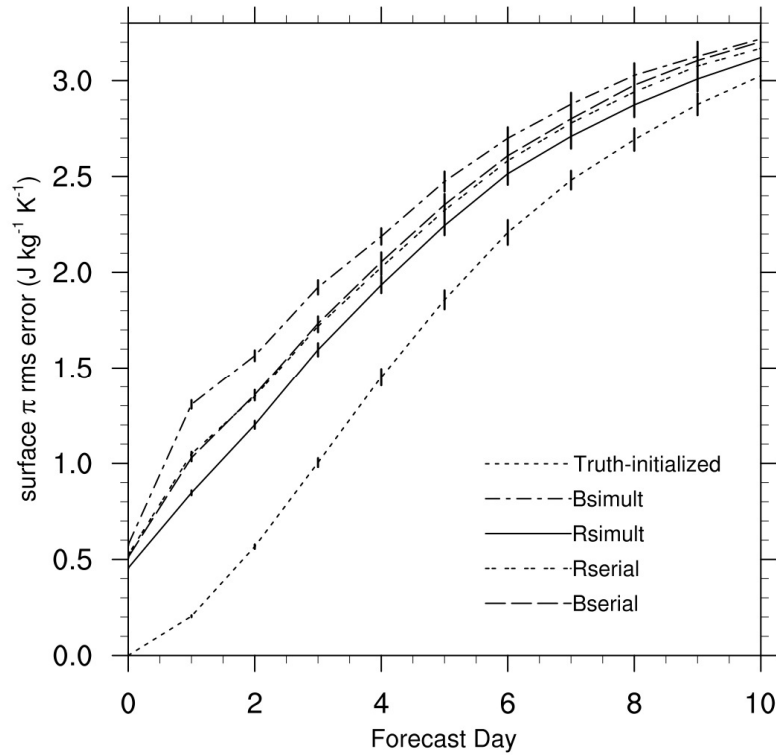
Figure 4. Average surface π rms error for forecasts initialized from the ensemble analysis mean of the Bserial, Rserial, Rsimult and Bsimult schemes. The average error of the truth-initialized forecasts is also presented as an indication of the contribution of the model error to the forecast error. Averages are over 150 cases and error bars show standard error calculated using a bootstrap resampling method.
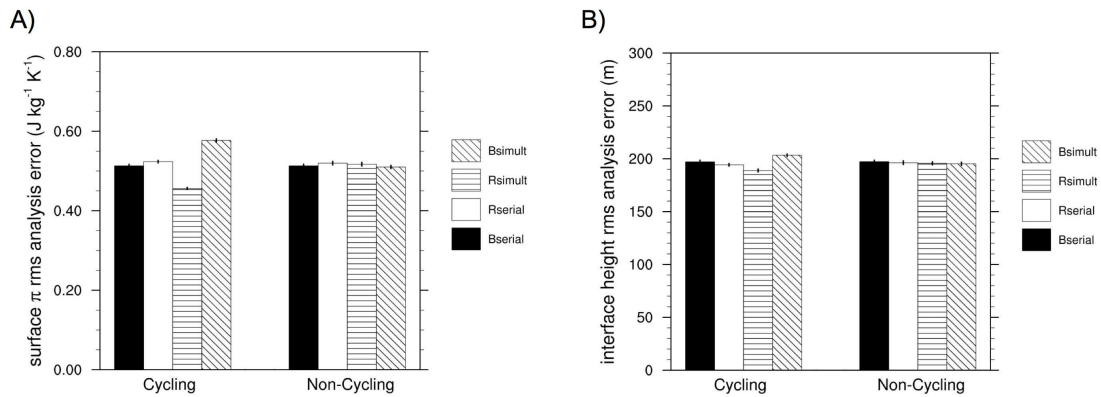
Figure 5. Comparison of average rms analysis error of four schemes in a data assimilation and forecast cycling experiment and a non-cycling experiment (see text for a description of the non-cycling experiment). (a) Surface π and (b) interface height results are shown. Averages are over 150 cases. Error bars are standard error calculated using a bootstrap resampling method.
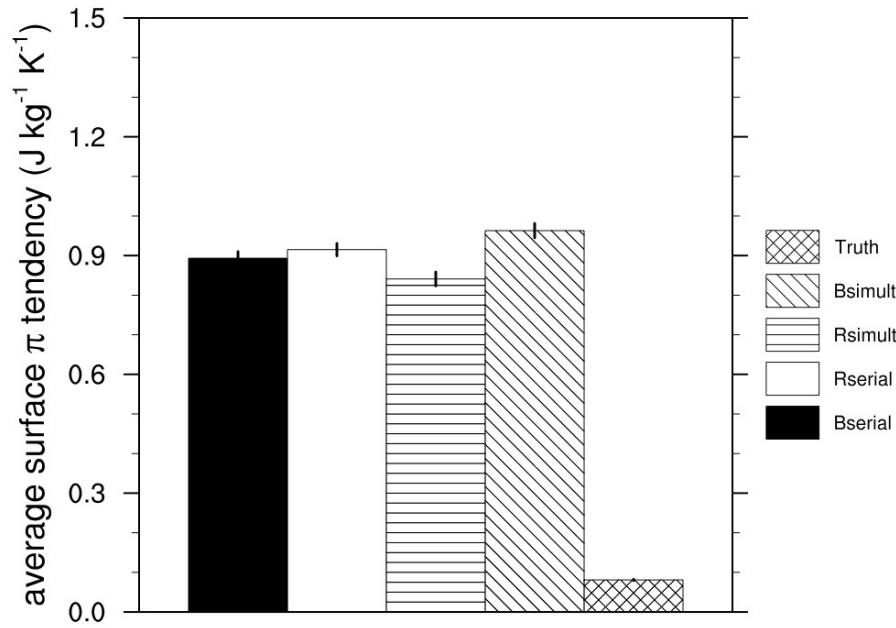
Figure 6. Average absolute 1-hour surface π tendency in the non-cycling experiment for the four schemes compared in Figure 5. The truth state surface π tendency is the average surface π tendency in the truth run. Averages are over 150 cases. Error bars are standard error using a bootstrap resampling method. See text for details of the surface π tendency calculation.
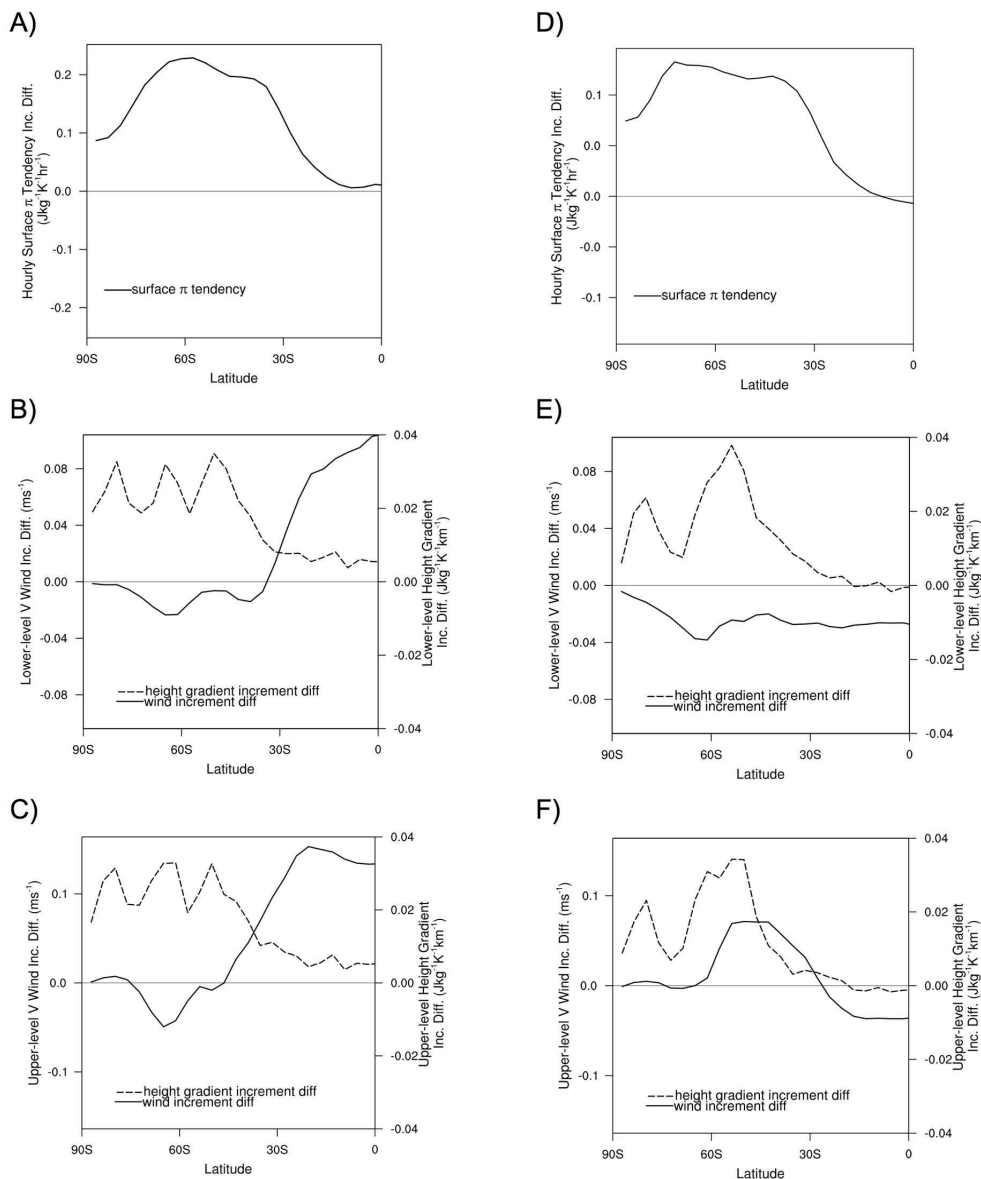
Figure 7. Differences in increment magnitude between (a-c) the Bsimult and Rsimult schemes and (d-f) the Rserial and Rsimult schemes. Positive values indicate the Bsimult or Rserial scheme increment was larger than the Rsimult increment. (a,d) Hourly surface π tendency increment difference, (b,e) lower-level meridional wind increment difference and longitudinal height gradient increment difference, and (c,f) upper-level meridional wind increment and longitudinal height gradient increment difference are shown. See text for further explanation of the quantities shown. Averages are over 150 cases.
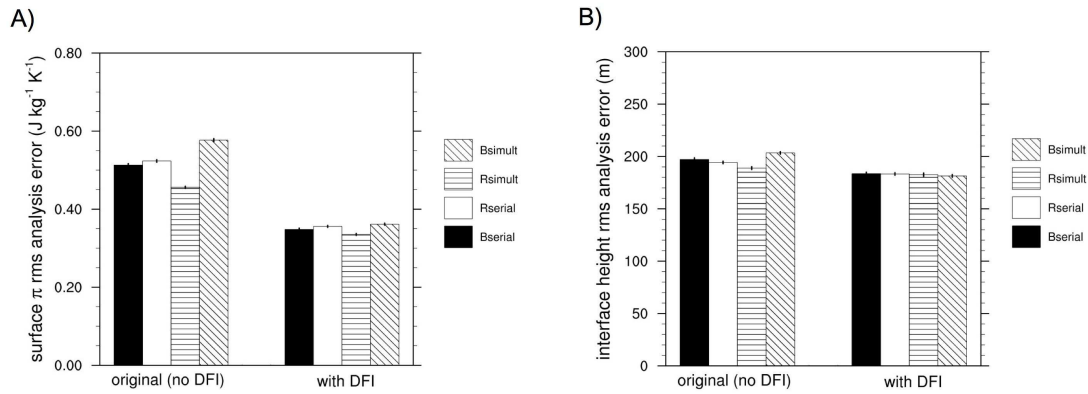
A)

B)

Figure 8. Average rms analysis error from the cycling experiments of section 3.1 and experiments using digital filter initialization (DFI). (a) Shows surface π and (b) shows interface height results. The results shown are using the optimal localization scale in each case, and are averages over 150 cases. Error bars are standard error calculated using a bootstrap resampling method.
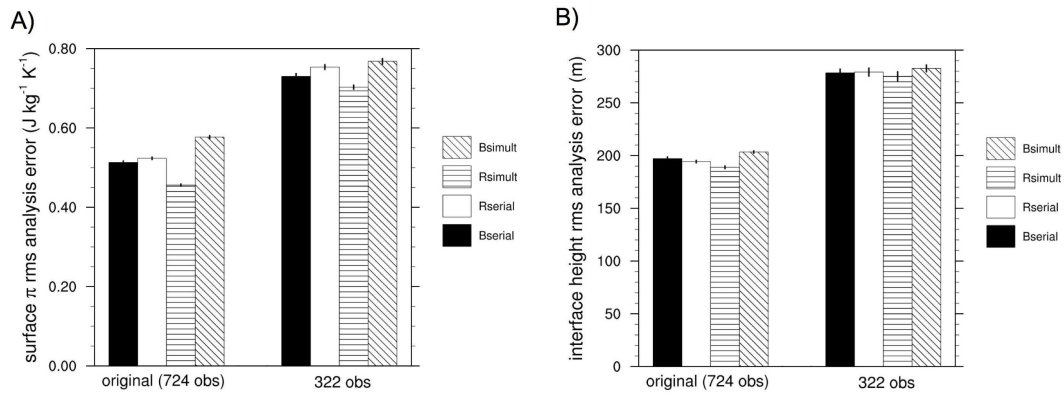
Figure 9. Average rms analysis error from the baseline cycling experiments of section 3.1 and experiments reducing the number of observations from 724 to 322. (a) Surface π and (b) interface height results are shown. The results using the optimal localization scale in each case are shown, and averages are over 150 cases. Error bars are standard error calculated using a bootstrap resampling method.
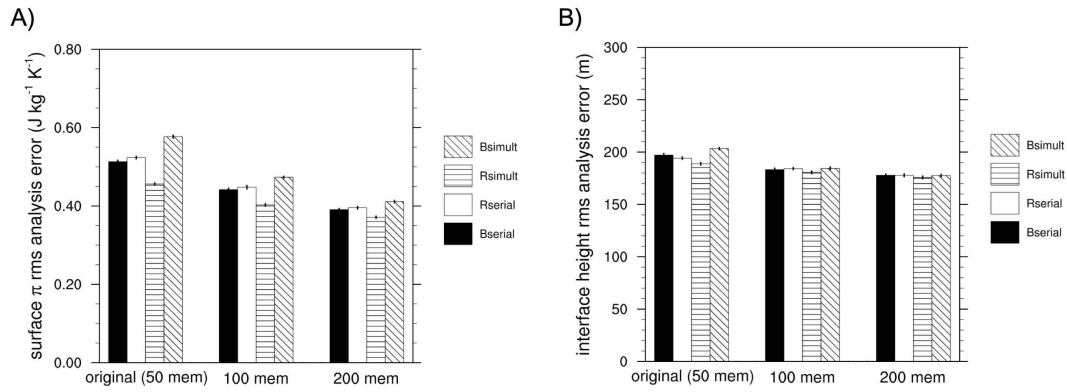
Figure 10. Average rms analysis error from the baseline cycling experiment of section 3.1 using 50 members and experiments using 100-member and 200-member ensembles. (a) Surface π and (b) interface height results are shown. The results using the optimal localization scale in each case are shown, and averages are over 150 cases. Error bars are standard error calculated using a bootstrap resampling method.
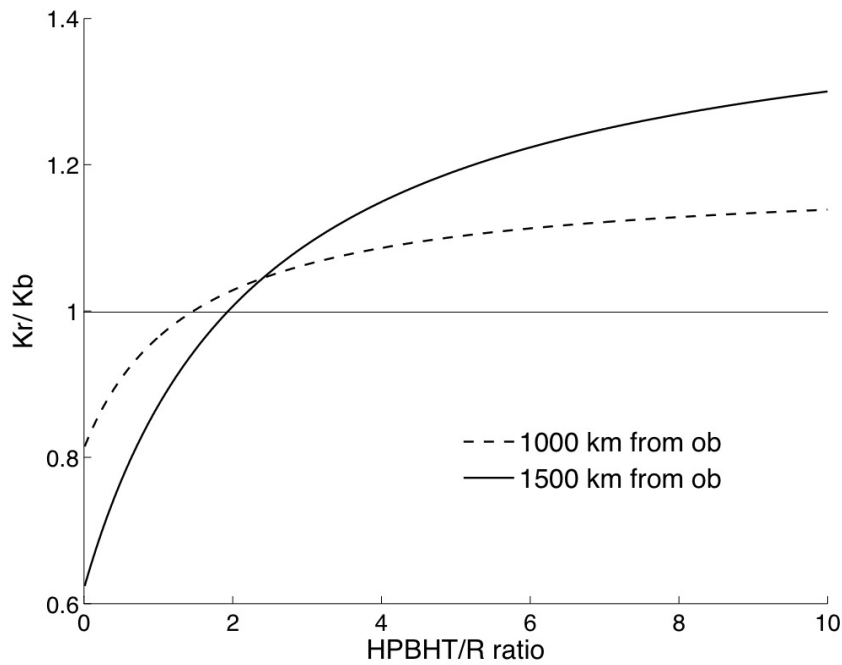
Figure 11. Ratio of the observation influences produced using R-localization and B-localization in a single observation case, shown as a function of the ratio of the background error variance to the observation error variance. Ordinate values greater than 1 indicate that the R-localization method yields more observation influence, and ordinate values less than one indicate that the B-localization yields more observation influence. Optimal localization cutoff distances for the surface π rms error in the cycling experiment are used (4000 km for R-localization, 6000 km for B-localization). Two cases are shown: one in which the model grid point being updated is 1000 km from the observation location and one in which the model grid point is 1500 km from the observation location.