EFFECTS OF SEQUENTIAL OR SIMULTANEOUS ASSIMILATION OF
OBSERVATIONS AND LOCALIZATION METHODS ON THE PERFORMANCE
OF THE ENSEMBLE KALMAN FILTER

A THESIS
SUBMITTED TO THE GRADUATE FACULTY
In partial fulfillment of the requirements for the
Degree of
MASTER OF SCIENCE IN METEOROLOGY

By
BRIAN HOLLAND
Norman, Oklahoma
2011
EFFECTS OF SEQUENTIAL OR SIMULTANEOUS ASSIMILATION OF OBSERVATIONS AND LOCALIZATION METHOD ON THE PERFORMANCE OF THE ENSEMBLE KALMAN FILTER

A THESIS APPROVED FOR THE SCHOOL OF METEOROLOGY

BY

Dr. Xuguang Wang, Chair

Dr. Ming Xue

Dr. Louis Wicker
Acknowledgements

I would like to thank my committee chair, Dr. Xuguang Wang, for her guidance, knowledge, and provision of funding. I would also like to thank my other committee members, Dr. Louis Wicker and Dr. Ming Xue, for their guidance and suggestions. This thesis work was supported by University of Oklahoma faculty award 122-792100, NOAA THORPEX funds NA08OAR4320904 and NASA New Investigator Program Award NNX10AQ78G.

Thanks are also due to the many other people who have helped guide me through my graduate education: Celia Jones, Marcia Pallutto, Nancy Campbell, Don Rodgers, Dr. David Parsons, Dr. Alan Shapiro, Dr. Kenneth Crawford, and all of my officemates, particularly Alex Schenkman. The staff of the OU Supercomputing Center for Education & Research, particularly Henry Neeman, and Scott Hill at the Center for the Analysis and Prediction of Storms, are also owed a great thank you for their resources, patience, and incredible willingness to help.

Above all I must thank my family, Mike, Mary, Jamie, and Debbie Holland, and my girlfriend, Alice Ann Spurgin. Their support and encouragement have been more important than they can know. Finally, thanks to the many friends in the School of Meteorology and at St. Thomas More Catholic Church who helped keep me sane and entertained.
# Table of Contents

Acknowledgements ........................................................................................................ iv

Table of Contents ............................................................................................................ v

List of Tables .................................................................................................................. vi

List of Figures ............................................................................................................... vii

Abstract .......................................................................................................................... ix

Chapter 1: Introduction ................................................................................................. 1

Chapter 2: Methods ........................................................................................................ 6
  2.1 Bserial Scheme ....................................................................................................... 6
  2.2 Rserial Scheme ....................................................................................................... 8
  2.3 Bsimult Scheme .................................................................................................... 9
  2.4 Rsimult Scheme ................................................................................................... 10
  2.5 LETKF Scheme .................................................................................................. 10
  2.6 Patch Serial scheme ......................................................................................... 12
  2.7 Experimental Design ....................................................................................... 12

Chapter 3: Cycling Experiment Results ..................................................................... 16
  3.1 Analysis Error ...................................................................................................... 16
  3.2 Imbalance ............................................................................................................ 17
  3.3 Forecast ............................................................................................................... 18

Chapter 4: Understanding the Differences Through Non-Cycling Experiments .... 20
  4.1 Root Mean Square Error and Imbalance .............................................................. 20
  4.2 Further Exploration of Balance Problems: Increment Differences ....................... 21
  4.3 Further Exploration of Balance Problems: Simulation of Updates with a Simple System .............................................................................................................. 23
  4.4 Ensemble Spread Differences Between Schemes ............................................... 28

Chapter 5: Comparison of Schemes in Different Situations .................................... 30
  5.1 Digital Filter Initialization .................................................................................... 30
  5.2 Type and Number of Observations ..................................................................... 31
  5.3 Ensemble Size .................................................................................................... 32
  5.4 The Ratio of Background Error Variance to Observation Error Variance ............ 33
  5.5 Localization before and after application of observation operator ..................... 36

Chapter 6: Summary and Conclusions ..................................................................... 38

Tables ............................................................................................................................. 41

Figures ........................................................................................................................ 45

References ..................................................................................................................... 61
List of Tables

Table 2.1: Summary of schemes used in this study, as described in text................................. 41
Table 5.1. Percent surface $\pi$ and interface height rms error difference between three pairs of
schemes using either interface height and surface $\pi$ observations (original baseline
experiment, first row) or interface height, surface $\pi$, 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. ........................................................................... 42
Table 5.2. 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............... 43
Table 5.3. Comparison of rms error results when applying covariance localization before and
after application of the observation operator. Results shown are using the Bsimult scheme,
and the differences are not statistically significant.......................................................... 44
List of Figures

Figure 1.1. 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 length scale. Dotted line shows B-localization using a length scale 25% longer than in the other two cases. ..........................................................................

Figure 3.1. (a) Surface $\pi$ 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 3.2. Average of 1-hour surface $\pi$ tendency values for the five schemes depicted in Figure 3.1. 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 3.3. Average surface $\pi$ rms error for forecasts initialized from the ensemble analysis mean of the four EnSRF-based schemes. The average error of the truth-initialized forecasts is also presented as an indication of the contribution of model error to the forecast error. Averages are over 150 cases and error bars show standard error using a bootstrap resampling method. ..........................................................

Figure 4.1. 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 $\pi$ and (b) interface height results are shown. Averages are over 150 cases. Error bars are standard error calculated using a bootstrap resampling method. ..........................................................

Figure 4.2. Average 1-hour surface $\pi$ tendency in the non-cycling experiment for the four schemes compared in Figure 4.1. The truth state surface $\pi$ tendency is the average surface $\pi$ 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 $\pi$ tendency calculation. ..........................................................

Figure 4.3. Differences in increment magnitude between the Bsimult and Rsimult schemes. Positive values indicate the Bsimult scheme increment was larger. (a) Hourly surface $\pi$ tendency increment difference, (b) lower-level meridional wind increment difference and longitudinal height gradient increment difference, and (c) 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. ..........................

Figure 4.4. Differences in increment magnitude between the Rserial and Rsimult schemes. Positive values indicate the Rserial scheme increment was larger. (a) Hourly surface $\pi$ tendency increment difference, (b) lower-level meridional wind increment difference and longitudinal height gradient increment difference, and (c) 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. ..........................

Figure 4.5: Setup of the simple simulation used to explore why Bsimult and Rserial produce weaker wind updates and stronger height updates than Rsimult. The bold line represents the line of gridpoints whose updates were computed in Figures 10 and 11. Each X represents the location of a simulated observation, and the large X represents the observation whose impact is shown in Figure 10. ..........................................................

Figure 4.6: Simulated influence of a surface $\pi$ observation located at the large X in Figure 4.5 on a line of gridpoints using either the Bsimult or Rsimult scheme. Observation influence is equivalent to the Kalman gain. (a) Shows a mid-latitude update of the height field, (b) shows a mid-latitude update of the wind field, and (c) shows an update of the wind field in the tropics. Figure annotations are referenced in the text. .............................
Figure 4.7: Average simulated influence of all observations shown in Figure 4.5, using the Rsimult or the Rserial scheme. For the wind, Y-axis is the sum of the magnitudes of the Kalman gains corresponding to the update by the seven observations in the center of the grid in Figure 4.5. For the height gradient, Y-axis is the sum of the gradients of the Kalman gains, with both bars scaled by the same factor. The sums at all gridpoints not near the edge of the domain were averaged to produce these data. ........................................ 55

Figure 4.8: Average analysis spread for four schemes in the non-cycling experiment. The same localization scales are shown as in Figures 4.1 and 4.2. Average is over 150 cases, and error bars are standard deviation of the mean calculated using a bootstrap resampling method. ............................................................................................................................... 56

Figure 5.1. Average rms analysis error from the cycling experiments of section 3.1 and experiments using digital filter initialization (DFI). (a) Shows surface \( \pi \) 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. ............................................................................................. 57

Figure 5.2. 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 \( \pi \) 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. ................................................................................ 58

Figure 5.3. 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 \( \pi \) 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. ........................................ 59

Figure 5.4. Ratio of the observation influences produced using B-localization and R-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 \( \pi \) 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. .......... 60
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 an 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. An additional cycling experiment showed that the method used to calculate the matrix square roots in the ensemble perturbation update, as well as whether or not observations were assimilated in a local “patch,” did not significantly affect performance. The four 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 analysis error of all four 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 four schemes were found to be smaller when digital filter initialization was used, when surface Exner function observations were not assimilated or assimilated in conjunction with wind observations, 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.
Chapter 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 inversion of a $p \times p$ 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 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 1.1 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 related to the effects of the sequential or simultaneous assimilation and covariance localization method on the performance of the EnKF include the following. 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 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 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, and also to eliminate factors
such as uncertainties in observation quality and physical parameterizations that could complicate comparisons. Future work is needed to compare the schemes using a real 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 an R-localized simultaneous scheme in an experiment using repeated 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, different observation networks, different ensemble sizes, and different ratios of background error to observation error. Section 6 provides a summary and draws together conclusions.
Chapter 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 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

\[
\bar{x}^a = \bar{x}^b + K(y - H\bar{x}^b),
\]

(1)

where \( \bar{x}^a \) is the analysis ensemble mean, \( \bar{x}^b \) is the background ensemble mean, \( y \) represents the observations, \( H \) is the observation operator used to convert from the
model space to the observation space, and $K$ is the Kalman gain. For the Bserial scheme, the Kalman gain is

$$K = \rho_{\text{Bser}} \circ P^b H^T (HP^b H^T + R)^{-1},$$  \hfill (2)$$

where $P^b$ is the background error covariance matrix estimated by the sample ensemble covariance, $R$ is the observation error covariance matrix, $\circ$ represents a Schur product, or element-by-element multiplication, and $\rho_{\text{Bser}}$ 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

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

where $X^a$ is the analysis ensemble perturbation matrix, $X^b$ is the background ensemble perturbation matrix, and $\tilde{K}$ is the “reduced” Kalman gain. For the Bserial scheme, where observations are assimilated one at a time, this reduced Kalman gain is

$$\tilde{K} = \left(1 + R \sqrt{HP^b H^T + R}^{-1} K \right)^{-1} K.$$ \hfill (4)
In the Bserial scheme, for the assimilation of the next observation, the $\mathbf{x}^a$ and $\mathbf{X}^a$ resulting from assimilation of the previous observation become the $\mathbf{x}^b$ 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, $\rho_{Rser}$, is applied to R in both the Kalman gain

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

and the reduced Kalman gain

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

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 Bs simultaneous Scheme

In the Bs simultaneous 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

\[
K = \rho_{Bsim} \cdot P^b H^T (\gamma \cdot H P^b H^T + R)^{-1}.
\]  

(7)

The values of the elements of \( \rho_{Bsim} \) in Eq. (7) are calculated based on the distance between the gridpoint and the observations. The diagonal elements of the localization matrix \( \gamma \) are 1, while the off-diagonal elements of \( \gamma \) are calculated based on the distance between the observations. The ensemble perturbations are updated using Eq. (3) except that \( \tilde{K} \) for the Bs simultaneous scheme is

\[
\tilde{K} = \rho_{Bsim} \cdot P^b H^T \{ (\sqrt{\gamma \cdot H P^b H^T + R})^{-1} \}^T \{ (\sqrt{\gamma \cdot H P^b H^T + R + \sqrt{R}})^{-1} \}.
\]

(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

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

and

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

in which $\rho_{Rsim}$ is the diagonal localization matrix with diagonal elements equal to the element-wise inverse of the elements of $\rho_{Bsim}$ 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

$$\bar{x}^a = \bar{x}^b + X^b \bar{w}^a,$$

where $\bar{w}^a$ is

$$\{(k-1)I + (HX^b)^T(\rho_{\text{Rsim}} \circ R)^{-1}(HX^b)^{-1}(HX^b)^T(\rho_{\text{Rsim}} \circ R)^{-1}(y - Hx^b).$$

The perturbation update is given by

$$X^a = X^b W^a,$$

where

$$W^a = [(k-1)\{(k-1)I + (HX^b)^T(\rho_{\text{Rsim}} \circ R)^{-1}(HX^b)\}^{-1}]^{1/2}.$$

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.

### 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 $\pi$ 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 such 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 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 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 $\pi$ at the surface (analogous to surface pressure) and the layer interface height were calculated from the ensemble mean. Surface $\pi$ 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 $\pi$ 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 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 bootstrap 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.
Chapter 3: Cycling Experiment Results

3.1 Analysis Error

Figure 2.1 shows the rms analysis error for the cycling experiment for the first five schemes described in Table 2.1. The localization cutoff distance that minimized rms error for all three R-localized schemes (Rsimult, Rserial, LETKF) was 4000 km for both surface $\pi$ and interface height, while the optimal cutoff distances for the two B-localized schemes (Bsimult, Bserial) were 6000 km for surface $\pi$ 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 $\pi$ 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 higher than that of the other schemes: 26.4% higher than the Rsimult scheme rms error for the surface $\pi$ norm and 6.8% higher for the interface height norm. The rms errors of the two schemes using serial assimilation were intermediate between the performance of the schemes using simultaneous assimilation, with the surface $\pi$ rms error of the Bserial being 11.1% lower than that of the Bsimult scheme and 12.5% higher than the Rsimult scheme surface $\pi$ error. The difference between the rms errors of the Bserial and the Rserial schemes was smaller: the rms error of the Bserial scheme was 2.0% lower than that of the Rserial scheme for the surface $\pi$ norm and 0.7% higher for the interface height norm. 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 especially when a simultaneous assimilation pattern was used. When the serial assimilation pattern was used, the choice of localization method had a small 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 $\pi$ tendency was used here to assess the degree of imbalance introduced by each scheme. Figure 3.2 shows the average 1-hour surface $\pi$ tendency, calculated as described in section 2.7. Results shown are using the localization cutoff distance that produced the smallest surface $\pi$ rms analysis error for each scheme in section 3.1. The Rsimult and LETKF schemes produced smaller surface $\pi$ tendency than the other three schemes. The surface $\pi$ tendency of the Bsimult scheme was the highest among the five schemes: 102% greater than that of the Rsimult scheme. The difference in the surface $\pi$
tendency between the Rserial and the Bserial was smaller, with the Rserial scheme surface $\pi$ tendency being 10.6% higher than the Bserial scheme surface $\pi$ tendency. The Rserial scheme surface $\pi$ tendency was 25.6% lower than the Bsimult scheme surface $\pi$ tendency and 33.4% higher than the Rsimult scheme surface $\pi$ tendency. The order of the schemes from the least to the most imbalance matched the order from the smallest to the largest surface $\pi$ 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 less imbalance than the B-localization method when the simultaneous assimilation pattern was used, but somewhat more imbalance than the B-localization method when the serial assimilation pattern was used. As was the case for rms error, 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

In order to further assess the differences between different schemes, the means of the analysis ensembles were used to initialize 10-day forecasts. Figure 3.3 shows the
rms forecast error for the surface $\pi$ 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.
Chapter 4: Understanding the Differences Through Non-Cycling Experiments

4.1 Root Mean Square 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 experiment 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 $\pi$ tendency, respectively. The comparison of the rms analysis error of the four schemes for the surface $\pi$ and interface height norms are shown in Figure 4.1. Unlike the cycling experiment, in the non-cycling case, there were no statistically significant differences in the rms analysis error between the schemes.

Figure 4.2 shows the 1-hour surface $\pi$ tendency for the four schemes in the non-cycling experiments, averaged over 150 cases. The Rsimult scheme produced the smallest surface $\pi$ tendency, followed in order of increasing surface $\pi$ tendency by the Bserial, Rserial, and Bsimult schemes. This order matches the order from the smallest
to the largest surface $\pi$ tendency in the cycling experiment (Figure 3.2). 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 $\pi$ 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: Increment Differences

To better understand the differences in the imbalance of the analyses shown in Figure 4.2, 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.

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. Figures 4.3 and 4.4 compare the surface $\pi$ tendency, wind, and height gradient analysis increments from the non-cycled Bsimult and Rsimult schemes and the non-cycled Rserial and Rsimult experiments, respectively. The “surface $\pi$ tendency increment” is defined as the magnitude of the surface $\pi$ tendency of a 1-hour forecast initialized using the analysis ensemble mean minus the magnitude of the surface $\pi$ tendency of a 1-hour forecast initialized using the background ensemble mean. In other words, the surface $\pi$ 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 4.3a shows the difference of the surface $\pi$ 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 $\pi$ tendency increment between the two schemes occurred in the mid- and high latitudes, where the Bsimult scheme surface $\pi$ tendency increment was higher than that of the Rsimult scheme. Figures 4.3b and 4.3c 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 $\pi$ tendency increment found in the Bsimult (Figure 4.3a). This result was found at both upper and lower levels. The wind and height gradient increment patterns in the tropics are explored further in section 5.4.

Similar diagnostics were performed comparing the Rserial and Rsimult schemes (Figure 4.4). Figure 4.4a shows that the Rserial scheme surface $\pi$ 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 4.3a, consistent with the result in the cycling experiment where rms analysis error differences seen in Figure 3.2 between the Rserial and Rsimult schemes were smaller than those between the Bsimult and Rsimult schemes. Similar to Figure 4.3b, Figure 4.4b 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 4.3c, Figure 4.4c 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 3.2 and 4.2) is also consistent with this result.

4.3 Further Exploration of Balance Problems: Simulation of Updates with a Simple System
To attempt to better understand why differences in the choice of localization method produced differences in the analysis increments of the schemes, a simple simulation was designed that would eliminate many of the complexities of the primitive-equation model environment while still providing a good mathematical approximation of the EnKF in the primitive-equation experiments. The Bsimult and Rsimult schemes were coded up for use with a grid of model points and 23 observations arranged as shown in Figure 4.5. As much as possible, realistic values were taken from the non-cycling experiments of section 4.1 and 4.2: the typical observation errors, the typical observation-space model error variances (i.e., the diagonal elements of $H P^b H^T$), the typical rate at which observation-space forecast error covariances decrease with distance (i.e., the off-diagonal elements of $H P^b H^T$), the typical distance from the observation location at which maximum correlation of an observation and a model variable for a given observation type and model field was found, and the typical rate at which that correlation decreased with distance (i.e., the elements of $P^b H^T$) were all taken from the non-cycling experiment.

Using these parameters, the Kalman gain corresponding to the update of the gridpoints located on the bold line in Figure 4.5 by the central observation in the system (the large X in Figure 4.5) was calculated. The results when simulating assimilation of a surface $\pi$ observation, which was the observation type that produced the greatest differences in balance and error between Bsimult and Rsimult, are shown in Figure 4.6. When simulating the update of the height field in the mid-latitudes (Figure 4.6a), the Bsimult update was almost universally larger than the Rsimult update, and the gradient of the Bsimult height update appeared to be steeper than the Bsimult update for points
between 800 and 2200 km away from the observation (denoted by the brackets at the top of Figure 4.6a). When simulating the mid-latitude update of the wind field (Figure 4.6b), the Bsimult update was weaker than the Rsimult update over most of the domain, including the area in which the stronger height gradient was observed, denoted again by the brackets at the top of the figure. Both of these findings were consistent with what was seen in the primitive equation model experiments (Figures 4.3b and 4.3c).

Additionally, the effect of changing the $H P^b H^T$ to $R$ ratio can be seen by comparing Figure 4.6b to a simulation of the wind field update in the tropics (Figure 4.6c)—the Bsimult increment was larger than the Rsimult increment in the tropics case, simulated by using a small $H P^b H^T$ to $R$ ratio (0.8), but not in the mid-latitude case, simulated by using a ratio of 2.8. This phenomenon is discussed in more depth in section 5.4.

While this experiment does not definitively explain the imbalance (and therefore rms error) differences observed between Bsimult and Rsimult, it does suggest a hypothesis for the cause of the observed differences: The characteristic way that a given observation type updates the height and wind fields may determine whether or not it will yield analyses of equal quality when used with B- and R-localization. Informal experimentation with this simple system showed that many parameters can determine this “characteristic update” and whether it is different when using B- or R-localization. Among these factors are the ratio of $H P^b H^T$ to $R$ (see section 5.4), the observation spacing, how rapidly correlations between points drop off with distance, and, perhaps more than anything, the distance from the observation at which the wind and height fields are typically updated the most by that type of observation.
It was not possible to compare serial and simultaneous schemes using exactly the same methodology used above because the magnitude of the influence of a single observation in the serial scheme will vary depending on where in the assimilation order it is processed. However, summing the influence of all observations in the simple case outlined in Figure 4.5 rather than looking at the influence of only one can achieve a similar comparison. One complicating factor in this is that $\mathbf{HP}^b\mathbf{H}^T$ and $\mathbf{P}^b\mathbf{H}^T$ must be updated after the assimilation of each observation in a serial scheme. Because in this simple simulation the spatial correlations contained in the covariance matrices were parameterized with Gaussian functions rather than being explicitly calculated from the ensemble, the equation for optimal analysis error covariance,

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^b$$

was used to calculate $\mathbf{HP}^a\mathbf{H}^T$ and $\mathbf{P}^a\mathbf{H}^T$ after the assimilation of each observation. Another complication was the fact that this simple simulation estimated observation influence using only Kalman gain magnitude, without consideration of any effects related to the innovation, $\mathbf{y} - \mathbf{Hx}^b$. In practice, each iteration in a serial assimilation would reduce the magnitude of the innovation, an effect which cannot be directly modeled in this simulation. Without correction, serial updates would generally appear to be larger than if innovation changes were considered. Because the goal of this experiment was simply to demonstrate that the serial/simultaneous differences seen in Figure 4.4 are reasonable, the innovation effect was corrected for by simply multiplying the serial Kalman gain by 0.8, a value that was found to produce relative differences
between serial and simultaneous that were comparable to that seen in earlier experiments.

The results of this comparison for Rserial and Rsimult are shown in Figure 4.7. Because the earlier primitive equation model experiments were done globally (meaning no observation or gridpoint was near a lateral boundary), only observations and gridpoints near the center of the domain in Figure 4.5 were used to calculate the statistics in Figure 4.7. The Rserial mid-latitude height field update was stronger than the Rsimult update. In the real model system, upper- and lower-level height updates were nearly identical (aside from the sign of the update being opposite at each level), and thus one case was used to simulate both upper- and lower-level height updates here. The Rserial lower-level mid-latitude wind field update was weaker than the Rsimult update, but the upper-level wind field update, characterized by more rapid decrease of correlations between points with distance than in the lower-level case, showed the opposite trend. As with the comparison of Bsimult and Rsimult described above, experimentation with this simple system suggested that the patterns that caused Rserial to have worse balance than Rsimult could be changed by modifying any of a number of the characteristics of the system, including the ratio of the diagonal elements of $\mathbf{H}^\mathbf{b} \mathbf{H}^\mathbf{T}$ to $\mathbf{R}$, the observation spacing, how rapidly correlations between points decrease with distance, and the distance from the observation at which the wind and height fields are typically updated the most by that type of observation. Thus, as in the case of B-localization and R-localization, it seems likely that the differences between serial and simultaneous processing of observations will vary from system to system and, even in the same system, will vary depending on the type of observations being assimilated.
4.4 Ensemble Spread Differences Between Schemes

While the most obvious contributor to the performance differences seen between the schemes in the cycling experiment of section 3 appeared to be imbalance differences, another possible contributing factor could be differences in ensemble spread among the schemes. In the non-cycling experiment, the B-localized schemes, particularly the Bsimult scheme, were found to reduce ensemble spread more than the R-localized schemes (Figure 4.8), which could over time cause the analyses produced by the B-localized schemes to be of lower quality due to the problems created in EnKF by underspread ensembles (Ehrendorfer, 2007 and references therein). The combination of the effects of imbalance differences and spread differences may better explain some of the results of this study than imbalance differences alone. The rms error differences between the Rserial and the Bserial schemes in the cycling experiment are one example. The non-cycling experiment showed the Bserial scheme analyses to be better balanced than the Rserial scheme analyses, but in that experiment the Bserial scheme also reduced the ensemble spread more than the Rserial scheme. The Bserial scheme performed better in the cycling experiment than the Rserial scheme for the surface $\pi$ norm, which is very sensitive to imbalance, but the Bserial scheme performed worse than the Rserial scheme for the interface height norm, which is less sensitive to imbalance. This suggests that feedback from both balance and ensemble spread differences may have been at work in the cycling experiment, with the effect of ensemble spread differences being predominant for variables such as interface height.
that are not as affected by imbalance, and the effect of balance differences being predominant for variables such as surface $\pi$ that are strongly affected by imbalance.
Chapter 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 5.1 shows that the surface $\pi$ and interface height rms error were reduced for all of the schemes when the DFI was used. For both the surface $\pi$ 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 all norms except the surface $\pi$ norm being statistically insignificant.

5.2 Type and Number of Observations

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 $\pi$ observations were assimilated, one more experiment was performed where in addition to interface height and surface $\pi$ observations, upper- and lower-level wind observations were also assimilated. Table 5.1 shows the surface $\pi$ 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 $\pi$ observations, the relative rms error differences between the schemes were reduced. It is hypothesized that the assimilation of wind observations, by providing an additional source of information with which to
update the wind field, produced an analysis with better wind-mass balance than in the reference experiment.

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 5.2 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 presumvably because of the reduction of sampling error when fewer, but the same type, of observations are assimilated. In such a case, the localization, designed to ameliorate the sampling 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 5.3 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 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

\[
\frac{K_r}{K_b} = \frac{\frac{H^b H^T}{R} + 1}{\rho \left( \frac{H^b H^T}{R} \right) + 1},
\]

where \(K_r\) is the Kalman gain resulting from use of the R-localization method and \(K_b\) 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, \( \rho \), and the ratio of \( \mathbf{H} \mathbf{P}^{b} \mathbf{H}^{T} \) to \( \mathbf{R} \). Consistent with the two solid lines in Figure 1.1, 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 \( \mathbf{H} \mathbf{P}^{b} \mathbf{H}^{T} \) 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 5.4 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 5.4 suggests that the observation influence depends on the ratio of \( \mathbf{H} \mathbf{P}^{b} \mathbf{H}^{T} \) to \( \mathbf{R} \). Figure 5.4 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 \( \mathbf{H} \mathbf{P}^{b} \mathbf{H}^{T} \) to \( \mathbf{R} \) ratio on the effects of the choice of the localization method, an experiment was conducted in which the \( \mathbf{H} \mathbf{P}^{b} \mathbf{H}^{T} \) 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 \( \mathbf{H} \mathbf{P}^{b} \mathbf{H}^{T} \) to \( \mathbf{R} \). Table 5.2 shows that decreasing the ratio of \( \mathbf{H} \mathbf{P}^{b} \mathbf{H}^{T} \) 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 \( \pi \) and interface height.
The dependence of the difference between the Bsimult and Rsimult schemes on the ratio of $HP^bH^T$ to $R$ may explain the latitudinal variation of the difference in the wind increments between the Bsimult and Rsimult schemes seen in Figures 4.3b and 4.3c. In Figure 5.4, when the ratio of $HP^bH^T$ to $R$ 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 4.3b and 4.3c, 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 $HP^bH^T$ to $R$ in the results shown in Figure 4.3 varied from a maximum of approximately 2.8 in the mid-latitudes, which Figure 5.4 suggests would lead to a larger update by the R-localized scheme, to an average of approximately 0.8 in the tropics, which Figure 5.4 suggests would lead to a larger update by the B-localized scheme. The latitudinal change in the ratio of $HP^bH^T$ to $R$ 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 $HP^bH^T$ to $R$ 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 $HP^bH^T$ to $R$ is incrementally decreased as ensemble spread is reduced. Thus, later observations are assimilated with a smaller ratio of $HP^bH^T$ to $R$, which previous studies and our own experiments suggest may decrease or even reverse some of the differences between the B-localization and the R-localization methods.
5.5 Localization before and after application of observation operator

Campbell et al. (2010) found that when an observation operator was used that relates an observation to the weighted average of multiple model gridpoints or variables, e.g., a satellite radiance observation influenced by temperatures at several levels of the model, covariance localization performed before the application of the observation operator (i.e. model space localization) gave more accurate results than covariance localization performed after applying the observation operator (observation space localization). In this study the observation operator used spectral interpolation, which involves using multiple model gridpoints, in order to generate values at the observation locations. In the B-localized schemes in the current study, observation space localization defined in Campbell et al. (2010) is more explicitly used than in the R-localized schemes. To study if such explicit observation space localization in the B-localized schemes could degrade their performances (e.g., Bsimult being the worst among the schemes tested), the Bsimult cycling experiment of section 3 was repeated where the model space localization was adopted using

\[
K = (\rho \circ P^b)H^T \{H(\rho \circ P^b)H^T + R\}^{-1}
\]

and

\[
\tilde{K} = (\rho \circ P^b)H^T \{\sqrt{H(\rho \circ P^b)H^T + R}\}^{-1} \{\sqrt{H(\rho \circ P^b)H^T + R + \sqrt{R}}\}^{-1},
\]
where $\rho$ is the localization function matrix computed using the distance between model points, in place of Eqs. (7) and (8). Applying the localization in the model space rather than the observation space did not produce a statistically significant difference in the Bsimult scheme surface $\pi$ and interface height rms errors (Table 5.3).
Chapter 6: Summary and Conclusions

The main goal of this study was to explore the effect of two common variations among EnKF schemes: 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 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 simultaneous EnSRF, nor whether or not a “patch” assimilation method was used in the sequential case had a 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 significant differences between the schemes in terms of the analysis accuracy, but the differences in the imbalance, as measured by surface $\pi$ 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 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 digital filter initialization step was used to reduce analysis imbalance, when mass observations were assimilated in conjunction with wind observations, when the number of observations of the same type 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 $\mathbf{H}^\mathbf{b}\mathbf{H}^\mathsf{T}$ to $\mathbf{R}$ in the system was reduced by using observations with larger error. Application of a B-localization function in the model space rather than the observation space appeared to have no impact on EnKF performance.
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 $\mathbf{HP}^b \mathbf{H}^T$ to $\mathbf{R}$ ratio, and whether digital filter localization is used. Therefore, the effects of these choices and the root causes of the differences of the schemes may differ between global, regional, and storm-scale applications.
## Tables

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Assimilation Pattern</th>
<th>Localization</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bserial</td>
<td>Serial</td>
<td>B</td>
<td>EnSRF</td>
</tr>
<tr>
<td>Rserial</td>
<td>Serial</td>
<td>R</td>
<td>EnSRF</td>
</tr>
<tr>
<td>Bsimult</td>
<td>Simultaneous</td>
<td>B</td>
<td>EnSRF</td>
</tr>
<tr>
<td>Rsimult</td>
<td>Simultaneous</td>
<td>R</td>
<td>EnSRF</td>
</tr>
<tr>
<td>LETKF</td>
<td>Simultaneous</td>
<td>R</td>
<td>LETKF</td>
</tr>
<tr>
<td>Patch Rserial</td>
<td>Serial</td>
<td>R</td>
<td>EnSRF</td>
</tr>
</tbody>
</table>

Table 2.1: Summary of schemes used in this study, as described in text.
<table>
<thead>
<tr>
<th>Observations Used</th>
<th>Rserial vs. Rsimult</th>
<th>Bserial vs. Bsimult</th>
<th>Bsimult vs. Rsimult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface height, surface π</td>
<td>12.9% 2.8%</td>
<td>11.1% 3.0%</td>
<td>26.4% 6.8%</td>
</tr>
<tr>
<td>Interface height, surface π, wind</td>
<td>7.2% 0.9%</td>
<td>10.8% 2.1%</td>
<td>18.4% 3.0%</td>
</tr>
</tbody>
</table>

Table 5.1. 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.
Table 5.2. 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.

<table>
<thead>
<tr>
<th></th>
<th>Original experiment</th>
<th>Reduced $\mathbf{H}\mathbf{p}\mathbf{H}^\top/\mathbf{R}$ ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface $\pi$ rms error</td>
<td>20.9%</td>
<td>17.5%</td>
</tr>
<tr>
<td>Interface height rms error</td>
<td>7.1%</td>
<td>2.3%</td>
</tr>
<tr>
<td></td>
<td>Original experiment (localization before $\mathbf{H}$)</td>
<td>Localization After $\mathbf{H}$</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>------------------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Surface $\pi$ rms error (J kg$^{-1}$K$^{-1}$)</td>
<td>0.5805</td>
<td>0.5770</td>
</tr>
<tr>
<td>Interface height rms error (m)</td>
<td>204.9</td>
<td>203.4</td>
</tr>
</tbody>
</table>

Table 5.3. Comparison of rms error results when applying covariance localization before and after application of the observation operator. Results shown are using the Bsimult scheme, and the differences are not statistically significant.
Figures

Figure 1.1. 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 length scale. Dotted line shows B-localization using a length scale 25% longer than in the other two cases.
Figure 3.1. (a) Surface $\pi$ 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 3.2. Average of 1-hour surface $\pi$ tendency values for the five schemes depicted in Figure 3.1. 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 3.3. Average surface $\pi$ rms error for forecasts initialized from the ensemble analysis mean of the four EnSRF-based schemes. The average error of the truth-initialized forecasts is also presented as an indication of the contribution of model error to the forecast error. Averages are over 150 cases and error bars show standard error calculated using a bootstrap resampling method.
Figure 4.1. 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 4.2. Average 1-hour surface $\pi$ tendency in the non-cycling experiment for the four schemes compared in Figure 4.1. The truth state surface $\pi$ tendency is the average surface $\pi$ 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 $\pi$ tendency calculation.
Figure 4.3. Differences in increment magnitude between the Bsimult and Rsimult schemes. Positive values indicate the Bsimult scheme increment was larger. (a) Hourly surface $\pi$ tendency increment difference, (b) lower-level meridional wind increment difference and longitudinal height gradient increment difference, and (c) 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.
Figure 4.4. Differences in increment magnitude between the Rserial and Rsimult schemes. Positive values indicate the Rserial scheme increment was larger. (a) Hourly surface $\pi$ tendency increment difference, (b) lower-level meridional wind increment difference and longitudinal height gradient increment difference, and (c) 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.
Figure 4.5: Setup of the simple simulation used to explore why Bsimult and Rserial produce weaker wind updates and stronger height updates than Rsimult. The bold line represents the line of gridpoints whose updates were computed in Figures 10 and 11. Each X represents the location of a simulated observation, and the large X represents the observation whose impact is shown in Figure 10.
Figure 4.6: Simulated influence of a surface $\pi$ observation located at the large X in Figure 4.5 on a line of gridpoints using either the Bsimult or Rsimult scheme. Observation influence is equivalent to the Kalman gain. (a) Shows a mid-latitude update of the height field, (b) shows a mid-latitude update of the wind field, and (c) shows an update of the wind field in the tropics. Figure annotations are referenced in the text.
Figure 4.7: Average simulated influence of all observations shown in Figure 4.5, using the Rsimult or the Rserial scheme. For the wind, Y-axis is the sum of the magnitudes of the Kalman gains corresponding to the update by the seven observations in the center of the grid in Figure 4.5. For the height gradient, Y-axis is the sum of the gradients of the Kalman gains, with both bars scaled by the same factor. The sums at all gridpoints not near the edge of the domain were averaged to produce these data.
Figure 4.8: Average analysis spread for four schemes in the non-cycling experiment. The same localization scales are shown as in Figures 4.1 and 4.2. Average is over 150 cases, and error bars are standard deviation of the mean calculated using a bootstrap resampling method.
Figure 5.1. Average rms analysis error from the cycling experiments of section 3.1 and experiments using digital filter initialization (DFI). (a) Shows surface $\pi$ 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 5.2. 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 $\pi$ 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 5.3. 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 5.4. Ratio of the observation influences produced using B-localization and R-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 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.
References


