Tropical Observability and Predictability

by

Timothy Robert Whitcomb

Submitted to the Department of Earth, Atmospheric, and Planetary Sciences
in partial fulfillment of the requirements for the degree of

Master of Science in Atmospheric Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2008

Author .....
Department of Earth, Atmospheric, and Planetary Sciences
October 31, 2007

Certified by .....
Kerry A. Emanuel
Professor of Atmospheric Science
Thesis Supervisor

Accepted by .....
Maria T. Zuber
E. A. Griswold Professor of Geophysics
Head, Department of Earth, Atmospheric, and Planetary Sciences

Tropical Observability and Predictability

by

Timothy Robert Whitcomb

Submitted to the Department of Earth, Atmospheric, and Planetary Sciences
on October 31, 2007, in partial fulfillment of the
requirements for the degree of
Master of Science in Atmospheric Science

Abstract

Many studies have investigated tropical data assimilation in the context of global models or specifically for tropical cyclones, but relatively few have focused on the mesoscale predictability and observability of the general tropical environment. This work constructs an ensemble data assimilation system for the tropics using a state of the science mesoscale prediction model, and tests the effect of a sparse observational network of wind and moisture in constraining the estimate of the state. A perfect model framework is used as a necessary first step to ease interpretation of results. Ensemble assimilation allows for state-dependent error covariances, foregoing pre-derived balances and correlations and allowing for the use of the full nonlinear model.

Boundary conditions are necessary for limited-area models, and the perturbed lateral boundaries and initial conditions are taken from a global ensemble using a non-perturbed sea surface temperature analysis. In the mesoscale model, this uniform surface had a profound effect on moisture levels in the lower levels, rapidly bringing the spread of vapor mixing ratio to near zero. Comparing the mesoscale forecast with a downscaled global model forecast showed that the interior solution was not completely dependent on the boundary conditions.

Observing system experiments that assimilated synthetic moisture and wind component observations in the boundary layer and in the free atmosphere had a small effect on the state estimate when compared with an unconstrained control case. The largest improvement was in the upper troposphere obtained by observing upper-level moisture, but several analyses were degraded by the data, due in part to the sparse network and small localization radius.

Thesis Supervisor: Kerry A. Emanuel
Title: Professor of Atmospheric Science

Acknowledgments

Jim Hansen, in his roles at MIT and at NRL, has been an excellent host, advocate, and resource. His patience, understanding, promptness, teaching, and attention to detail have helped make the completion of this project possible.

Kerry Emanuel graciously agreed to take me as a student when Jim left MIT and has provided welcome and patient advice and support. I am also grateful for the late-term assistance from Alan Plumb, who agreed with short notice to serve on my committee.

The staff at NRL Monterey, particularly Craig Bishop, Sue Chen, Jim Doyle, Saša Gaberšek, and Justin McLay were incredibly helpful in dealing with the intricacies of the COAMPS model, aiding in the configuration and debugging of idealized model runs and providing data from the NOGAPS global ensemble.

Ensemble forecasting and data assimilation are possible only through a judicious application of raw computing power. Greg Shomo at TechSquare and the rest of the ACES team maintained the computing clusters that made this work feasible.

No graduate program would function without the foundation of a skilled administrative staff. Roberta Allard and Jacqui Taylor in the EAPS front office were always ready with answers and conversation. Mary Elliff constantly did her best to make the graduate students feel welcome and was always ready with a smile. She was always willing to give us whatever assistance she could, whether from providing cookies on her desk to helping use the fax machine. Carol Sprague in the Education Office gave guidance and helped navigate me through the intricacies of MIT paperwork and was incredibly helpful in dealing with procedural matters. Kristen Barilaro was always ready to help with grant finances and logistics.

Dan Gombos, Kelly Klima, Brian Tang, and Angela Zalucha worked together with me to help us through our classwork, and provided welcome interaction outside of the Green Building. I have greatly valued my time with them and many other fellow students. Brian in particular has demonstrated the qualities of kind & gracious friendship and has provided an example that I would do well to emulate.

While not officially connected to MIT, Jack Bartley and Chris Schlesinger have both furnished welcome oases for many of my guests and colleagues through their Cambridge establishments.

This work was funded by ONR Grant N00014-06-1-0500, and my enjoyment was funded in part by the Houghton Fund. I am grateful for both the financial support and travel opportunities afforded by the AMS/NASA Graduate Fellowship.

NCAR's Data Assimilation Research Testbed formed the foundation of this work. I'm thankful for the quick answers and support from Jeff Anderson, Tim Hoar, and Nancy Collins.

Throughout my entire tenure at MIT, both my parents have been extremely supportive through all the ups and downs that are ubiquitous to the graduate experience - their continued care and kindness was a welcome anchor when so many other things around me were changing.

Finally, I am chiefly indebted to my wife Kaarin, whose support and steadfastness strengthened both my faith and determination and helped make the completion of this thesis possible. She endured many late, frustrating nights and yet provided encouragement for me to continue to keep pushing forward and giving my best.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In 1911, Norwegian meteorologist Vilhelm Bjerknes posited forecasting as the ultimate problem in meteorology: it requires both means and methods of estimating the current atmospheric state as well as a knowledge of the complete set of laws governing the time evolution of that state. Lewis Richardson was the first to attempt such a dynamical forecast in 1922, with poor results due in part to unbalanced initial conditions. As computer technology progressed from the mid-1950s to the present day, numerical models became more and more complicated - the barotropic vorticity model used by Charney evolved into modern high-resolution packages accounting for everything from aerosols to surface vegetation. Procedures for state estimation, or data assimilation, have grown as well. Straightforward polynomial interpolations using only radiosondes and surface observations have matured into sophisticated variational procedures incorporating data from weather radar, aircraft, and satellites.

Despite these advances, Bjerknes's problems still remain. Even though grid resolution is becoming ever finer, the continuous spectrum of atmospheric motions means that there will always be unresolved scales. The use of instruments and digital computers means that there will always be errors in the observations. These errors, no matter how small, will eventually grow and render "precise very-long-range forecast[s] ... non-existent" (Lorenz, 1963). The problem of prediction, then, is to realize these limitations and improve understanding, modeling, and assimilation methods to give the greatest utility for forecasts before the limit is reached.

The tropics represent a particular challenge for both atmospheric modeling and data assimilation. Even early numerical models for midlatitudes were successful because they were able to rely on quasi-geostrophic (QG) models that are remarkably accurate in predicting planetary-scale flows. In the tropics, though, the assumptions (such as small Rossby number) that give QG theory its power in midlatitudes begin to disintegrate. The large percentage of population and land mass in the Northern Hemisphere led to dense observational networks placed where they would maximally improve forecasts in the midlatitudes - tropical observing networks were left relatively sparse. Finally, while global models have moved to finer and finer resolution over the past decades, they are still far above the scale of the convection that dominates the tropics. This scale separation requires parameterizations, which introduces errors both in the parameterizations themselves and in the model's ability to accurately represent the environment in which the convection is taking place, whether it be in the surrounding air or the sea surface below. Forcing in the tropics also tends to be relatively weak.

Accurate forecasting in the tropics is important both for the tropics and midlatitudes. Long-term signals like the El Niño-Southern Oscillation (ENSO) and the Madden-Julian Oscillation (MJO) have been connected to changes in worldwide weather patterns, but general circulation models have still been hard-pressed to accurately represent MJO signals. Diabatic heating and divergent flow associated with a tropical convective response to sea surface temperature (SST) anomalies can give rise to Rossby waves that propagate into midlatitudes (Trenberth et al., 1998). On shorter scales, tropical cyclones can pose a direct threat to life and property as they move into populated regions. Even if they remain at sea, they can transition through the extratropics and interact with mid-latitude systems leading to strong storms.

1.1 Previous Work

While data assimilation issues in the tropics have been tackled in the past, most projects have either focused specifically on assimilation for tropical cyclones or for

global models. There has been a good deal of work in understanding the large-scale tropical environmental development, but the goal here is to build a framework to attempt to understand the mesoscale development details.

There are many previous studies that contribute pieces to the current work. Outlined below are some applicable works that show some important results from tropical data assimilation on the mesoscale and in global models, as well as the promise of mesoscale ensemble data assimilation.

1.1.1 Tropical Data Assimilation in Global Models

Undén (1989) found that if the nondivergent wind constraint on the European Centre for Medium Range Weather Forecasts (ECMWF) wind analyses was relaxed in the tropics, tropical flow patterns appeared more realistic. Tsuyuki (1996, 1997) carried out an extensive series of experiments using four-dimensional variational assimilation (4DVAR), which made model dynamics an important part of the assimilation. He found that a significant difficulty of using variational assimilation in the tropics is the importance of nonlinear physical parameterizations. The vertical distribution of analysis error showed innovations without specific structure functions, even when observing winds at only two levels in the tropics.

1.1.2 Tropical Mesoscale Data Assimilation

Ramamurthy and Carr (1987, 1988) is one of the few studies concerned with mesoscale data assimilation in the general tropical environment. They used data from the 1979 Summer Monsoon Experiment (MONEX; Fein and Kuettner, 1980), and combined them with their limited-area semi-Lagrangian primitive equation hydrostatic model using a modified successive correction method similar to Cressman analysis (see, e.g. Daley, 1991, chap. 3). They were able to assimilate wind, temperature, and moisture observations obtained at asynoptic times, and found that temperature data provided the least utility, while the combination of wind and moisture observations gave the best analyses.

1.1.3 Ensemble Data Assimilation in Mesoscale Models

Studies using mesoscale ensembles have been limited mostly to midlatitudes. Zhang et al. (2006) used a perfect model scenario to simulate an explosive cyclogenesis, assimilating sounding and surface observation data. They found that although the filter kept the analysis close to the known true value, the greatest error reduction was in large scales and less so in the marginally resolvable scales. Dirren et al. (2007) added simulated aircraft observations to the typical radiosonde and surface observations for a domain over the Pacific Northwest United States. An example of the technique showed that an error in a 500 hPa short-wave trough was corrected using only surface pressure observations, demonstrating the use of an ensemble-estimated flow-dependent background error covariance.

1.2 Present Work

This study begins to bridge the gap between data assimilation in global models and tropical-cyclone-specific models by constructing an ensemble data assimilation system for a mesoscale model of the tropical environment, bringing together the pieces in §1.1.

Ensemble-based filtering is chosen since it incorporates model dynamics into the assimilation process and allows for continuous assimilation of asynoptic observations. State-dependent background error covariances allow unobserved variables to be updated without the use of static correlation fields: these can be difficult to construct in the tropics, due to the lack of large-scale balances like geostrophy. While 4DVAR can accomplish these goals, it requires the use of linearizations and adjoints: ensemble methods can use the full nonlinear forward model.

While the model resolution used for these experiments is only slightly smaller than that used by global models, mesoscale models offer several advantages over global models. They have the capability to be run with much smaller grid spacing, and can nest in regions where even finer detail is desired. They also, as is used here, use the non-hydrostatic form of the governing equations. One drawback of this approach, though, is that these limited-area models require boundary forcings from

global models.

Although mesoscale models can provide very fine detail and coherent structures (depending, of course, on the resolution used), this is no guarantee that the details shown actually exist in reality. For these initial steps, a perfect model framework is used: this uses a reference model integration as the truth, and draws observations from it. This defines the real world as that described by the model, eliminating the consideration of model error from the results and allowing the specification of observational networks and densities that are useful, but not necessarily available in actuality. This limits the conclusions that can be drawn from the results, and actual observations will be included in future work.

This work tests whether observations of wind or moisture at a single level in either the boundary layer or the free atmosphere are able to constrain a 50-member tropical mesoscale ensemble integrated out to 24 hours, measured by both analysis error and ensemble spread. Chapter 2 describes the data and assimilation methods used in carrying out the experiments. Chapter 3 describes the results those experiments. Chapter 4 concludes and suggests further directions of research, and Appendix C describes the implementation of the data assimilation system.

Chapter 2

Data and Methods

This chapter describes the theory and tools used for the observability experiments and the data sets used in their construction. The ensemble is generated using NCAR's Data Assimilation Research Testbed (DART) - details on the implementation of COAMPS within the DART framework are provided in Appendix C.1.

2.1 Ensemble Filters

A filter is designed to give the best estimate of a state based on information available up to and including the present time. For NWP models, this corresponds to using a history of analyses, forecasts, and observations. In addition to a quality state estimate, ensemble methods allow for correlative analysis, like Ensemble Synoptic Analysis (ESA; Hakim and Torn, 2006) or statistical potential vorticity inversion (Gombos and Hansen, 2007), that can improve understanding of the system in question. Ensemble filters produce initial conditions for ensemble forecasts that outperform other common operational perturbation methods (Descamps and Talagrand, 2007). This section describes the background and implementation of the ensemble-based filter used for these experiments.

2.1.1 General Background

One of the most common methods of defining a "best" estimate is that that minimizes the squared error. If the error $\boldsymbol{\varepsilon}$ in a state estimate $\mathbf{x}$ is defined as

$$\boldsymbol{\varepsilon} = \mathbf{x}^t - \mathbf{x}$$

with $\mathbf{x}^t$ as the true state vector, a cost function can be written as:

$$J = \boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{S} \boldsymbol{\varepsilon}.$$

The choice of the state estimate $\mathbf{x}$ that minimizes J for any positive definite $\mathbf{S}$ is the minimum variance or conditional mean estimate

$$\mathbf{x} = \mathrm{E}\left[\mathbf{x}^t | \mathbf{Y}^o\right], \tag{2.1}$$

where $\mathbf{Y}^o$ is all observations available up to the current time. Provided that the dynamics are linear and model and observation errors are Gaussian and not state-dependent, the Kalman filter provides this minimum variance solution (Cohn, 1997). The Kalman filter equations (Kalman, 1960; Kalman and Bucy, 1961), using the Ide et al. (1997) notation (where the subscript i refers to the value of a time-dependent quantity at time t_i) are:

$$\begin{aligned}
\mathbf{x}^f(t_i) &= \mathbf{M}_{i-1}\left[\mathbf{x}^a(t_{i-1})\right], & (2.2)\\
\mathbf{P}^f(t_i) &= \mathbf{M}_{i-1}\mathbf{P}^a(t_{i-1})\mathbf{M}_{i-1}^{\mathrm{T}} + \mathbf{Q}(t_{i-1}), & (2.3)\\
\mathbf{x}^a(t_i) &= \mathbf{x}^f(t_i) + \mathbf{K}_i\left(\mathbf{y}_i^o - \mathbf{H}_i\mathbf{x}^f(t_i)\right), & (2.4)\\
\mathbf{P}^a(t_i) &= \left(\mathbf{I} - \mathbf{K}_i\mathbf{H}_i\right)\mathbf{P}^f(t_i). & (2.5)
\end{aligned}$$

The superscript f refers to prior, or forecast, estimates and a refers to posterior, or analysis, estimates. Computing (2.2) - (2.3) is the forecast step, advancing the state $\mathbf{x}$ and state error covariance $\mathbf{P}$ from time t_{i-1} to t_i under the model $\mathbf{M}$ with known

model error covariance $\mathbf{Q}$. The analysis step ((2.4)-(2.5)) combines the forecasted state with observations $\mathbf{y}^o$ (with $\mathbf{H}$ mapping model state space to observation space) using the weighting given by the Kalman gain $\mathbf{K}$,

$$\mathbf{K}_i = \mathbf{P}^f(t_i)\mathbf{H}_i^{\mathrm{T}} \left[\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^{\mathrm{T}} + \mathbf{R}_i\right]^{-1}, \tag{2.6}$$

where $\mathbf{R}$ is the error covariance of the observations ($\mathbf{y}^o$).

One of the primary barriers to using this formulation for systems with large state dimension is computational: for a state vector of length n, the error covariance matrix $\mathbf{P}$ has size n^2 – this implies significant expenses in computation and storage. The stipulation of linear dynamics is also unrealistic for many geophysical models. The linear dynamics limitation can be dealt with by using the extended Kalman filter (EKF; Gelb, 1974), which advances the state using a nonlinear model $\mathcal{M}$ in (2.2), and redefines $\mathbf{M}$ in (2.3) to be a linearization of $\mathcal{M}$ about the model trajectory. Still, though, it is difficult to calculate $\mathbf{P}$ given its size. The computational difficulty with the traditional and extended Kalman filters was alleviated with the development of the ensemble Kalman filter (EnKF; Evensen, 1994).

The EnKF replaces the full representation of the covariance matrix in (2.3) and (2.5) with an estimate of the error covariance based on an ensemble of states that sample the full error distribution:

$$\mathbf{P}^f(t_i) = \frac{1}{N-1}\tilde{\mathbf{X}}^f(t_i)\tilde{\mathbf{X}}^{f\mathrm{T}}(t_i) \tag{2.7}$$

$$\mathbf{P}^a(t_i) = \frac{1}{N-1}\tilde{\mathbf{X}}^a(t_i)\tilde{\mathbf{X}}^{a\mathrm{T}}(t_i). \tag{2.8}$$

The zero-mean state matrix $\tilde{\mathbf{X}}$ contains N columns, each consisting of a single ensemble member's state vector $\mathbf{x}^{(i)}$ with the ensemble mean $\overline{\mathbf{x}}$ removed:

$$\tilde{\mathbf{X}} = \left[\mathbf{x}^{(1)} - \overline{\mathbf{x}} \quad \mathbf{x}^{(2)} - \overline{\mathbf{x}} \quad \mathbf{x}^{(3)} - \overline{\mathbf{x}} \quad \mathbf{x}^{(4)} - \overline{\mathbf{x}} \quad \ldots \quad \mathbf{x}^{(n)} - \overline{\mathbf{x}}\right].$$

Depending on the size N of the ensemble this can be computationally feasible, even for large models. The reduced-rank representation of the covariance matrix (essentially

its matrix square root) can be easier to store. The work required to derive and specify $\mathbf{M}$ and $\mathbf{M}^{\mathrm{T}}$ is eliminated in favor of multiple runs of the full nonlinear model.

This method also allows use of an explicitly nonlinear observation operator $\mathcal{H}$ in the $\mathbf{PH}^{\mathrm{T}}$ terms: using the definition of $\mathbf{P}$ in (2.7)/(2.8),

$$\mathbf{PH}^{\mathrm{T}} \propto \tilde{\mathbf{X}}\tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} = \tilde{\mathbf{X}}\left(\mathcal{H}\left[\tilde{\mathbf{X}}\right]\right)^{\mathrm{T}}, \tag{2.9}$$

which is the covariance between the ensemble in model space and the ensemble mapped to observation space by $\mathcal{H}$. Another method of incorporating nonlinear observation operators is perform the assimilation in a joint model-observation state space, defined as $\mathbf{z} = [\mathbf{x}\ \mathcal{H}(\mathbf{x})]^{\mathrm{T}}$ and using $\mathbf{z}$ in lieu of $\mathbf{x}$ in the filter equations (Anderson, 2001). Although no explicit linearization is necessary, the effect of the update step in ensemble Kalman filters is a statistical linearization of $\mathcal{H}$, mapping analysis increments in observation space to analysis increments in state space (Anderson, 2003), so strongly nonlinear operators can still be a significant source of error.

The implementation of an EnKF gives advantages beyond computational tractability. Since $\mathbf{P}^f$ is used in the update step both directly and indirectly through the Kalman gain, using an ensemble means that the error covariances will be state-dependent rather than the carefully constructed static covariance fields used by other data assimilation methods. This allows the assimilation to be effective even in the absence of large-scale balances like geostrophy, given a sufficiently large ensemble.

Despite the computational advantages of the EnKF, there are multiple shortcomings. The small size of the ensemble relative to the size of the state vector gives an error covariance matrix that is severely rank-deficient, perhaps attenuating or completely missing important directions of variability in state space. Poor sampling underestimates the variance. In order to compute $\mathbf{P}^f$, the ensemble mean is assumed to be the best state estimate and the error spread is about the ensemble mean, rather than the true state (Evensen, 2003; Burgers et al., 1998).

2.1.2 The Ensemble Adjustment Filter

The ensemble filter implementation used in this work is the Ensemble Adjustment Filter (EAKF; Anderson, 2001), which is a member of a larger class of ensemble square-root filters (EnSRFs; see Tippett et al., 2003; Bishop et al., 2001; Whitaker and Hamill, 2002). These filters distinguish themselves by updating the matrix square root of the covariance matrix, rather than the entire covariance matrix itself. This update is deterministic (i.e., if the prior ensemble is known, the posterior ensemble is completely known as well), which eliminates the sampling errors from stochastic filters that use random number generators to provide a proper estimate of $\mathbf{P}^a$ (such as the perturbed observation method used by Houtekamer and Mitchell (1998); Burgers et al. (1998))

To simplify the use of nonlinear observation operators, the EAKF is cast in the joint state space defined in the previous section. This means that the joint ensemble state $\mathbf{z}$ for member k at time t_i is

$$\mathbf{z}_{i,k} = \left[\mathbf{x}_{i,k}\ \mathcal{H}(\mathbf{x}_{i,k}, t_i)\right]^{\mathrm{T}}. \tag{2.10}$$

The ensemble mean is updated using the observations;

$$\overline{\mathbf{z}}^a(t_i) = \overline{\mathbf{z}}^f(t_i) + \mathbf{K}_i\left(\mathbf{y}_i^o - \mathbf{H}_i\overline{\mathbf{z}}^f(t_i)\right). \tag{2.11}$$

Since we are working in a joint space, $\mathbf{H}$ is linear and easily specified. For a state dimension n with observation space dimension p, the observations are given by

$$\mathbf{y}_i = \begin{bmatrix}\mathbf{0}_{p,n} & \mathbf{I}_{p,p}\end{bmatrix}\mathbf{z}_i. \tag{2.12}$$

so $\mathbf{H}$ consists of two submatrices: a p-by-n block of zeroes and a p-by-p identity matrix. This serves to "pick out" the observation subspace of the joint state-observation space.

To update the forecast covariance, the EAKF applies a linear transformation to the ensemble $\mathbf{Z}^f$ (the matrix square root of the forecast covariance matrix $\mathbf{P}^f$) that gives an analysis covariance $\mathbf{P}^a$ consistent with (2.5). The EAKF solves for a linear

operator $\mathbf{A}$ that satisfies

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\,\mathbf{P}^f = \mathbf{A}\mathbf{P}^f\mathbf{A}^{\mathrm{T}}, \tag{2.13}$$

then uses that $\mathbf{A}$ to adjust individual ensemble members to give the proper spread about the mean:

$$\mathbf{Z}^f = \mathbf{A}\left(\mathbf{Z}^f - \overline{\mathbf{Z}}^f\right) + \overline{\mathbf{Z}}^a, \tag{2.14}$$

where $\overline{\mathbf{Z}}^a$ is generated by copying the updated ensemble mean $\mathbf{z}^a$ from (2.11) into N columns.

2.1.3 Ensemble Size

Since the estimate of state error covariance is based on sampling by the ensemble members, adequate ensemble size is crucial in determining a realistic error distribution. Ideally, the ensemble should be much larger than the dimension of the state subspace where the dynamics evolve - while this is likely smaller than the state dimension, it is still an unreachable goal for any model used for numerical weather prediction. An ensemble that is smaller than the state dimension leads to an error covariance matrix $\mathbf{P}$ that is not full-rank, and may have a null space that includes important directions of variability. Covariance localization (see §2.1.4) can make $\mathbf{P}$ full rank, though not necessarily correct.

The singular value decomposition (SVD) can be used to get a general feel for the impact of ensemble size on the error covariance matrix $\mathbf{P}$. The full SVD of a matrix $\mathbf{L}$ is defined as

$$\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}, \tag{2.15}$$

with orthogonal $\mathbf{U}$ and $\mathbf{V}$ and diagonal $\mathbf{\Sigma}$. By considering only the first r columns of $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{S}$, the result is the best approximation to $\mathbf{L}$ of rank r. The effect of the ensemble size on $\mathbf{P}$ is shown by the full covariance matrix and plotting the singular values σ_i to show their decay rate.

While this is helpful as a cursory estimate, the SVD method has its drawbacks.

First, it only measures the effect of truncation on the full ensemble-estimated error covariance matrix. If the initial ensemble is deficient and does not point in important directions, these shortcomings will be carried through - the truncation is only valid for the subspace that is spanned by the original ensemble. Second, it shows the rank r approximation to $\mathbf{P}$, but does not show which r members of the original ensemble to pick. Significant modes may only project onto a few ensemble members, and the omission of those members may greatly decrease the filter performance. Finally, it presumes that the important directions are the ones with the highest variance, which is not always the case in nonlinear systems like the atmosphere.

2.1.4 Implementation Details

While the EAKF and the perturbed-observation formulation of the EnKF can be implemented as shown, several adjustments are made to accommodate the small ensemble size which will be used in this work.

If the ensemble size is small, the true covariance between points is small, or both, sampling errors can give spuriously large covariances: this can be alleviated by limiting covariance computation to points within a specified radius (Houtekamer and Mitchell, 1998; Hamill et al., 2001). Results from idealized model runs showed that correlations in moisture and horizontal wind fields become indistinguishable from zero on a scale of about 100 km, so this is used as the width of the localization function, a Gaussian-like fifth-order polynomial after Gaspari and Cohn (1999) - this weighted taper method gives better results than just using a cutoff radius (Anderson and Anderson, 1999).

The errors in both the numerical model and sample size lead to an estimated prior covariance $\mathbf{P}^f$ that is too small. As seen in the formulation of the gain term in (2.6), this underestimation will give too little weight to incoming observations. This can yield to filter divergence, where the model trajectory can depart significantly from the "true" trajectory. This can be dealt with by inflating the prior covariance by some small amount (Anderson and Anderson, 1999; Anderson, 2001). While there are many methods of choosing an inflation factor, here the adaptive inflation technique

of Anderson (2007) is used. A filter estimates the inflation factor λ based on the observations in the same manner that the full state vector is estimated[1]. The filter is uses a naïve "model" for the inflation factor and assumes that the distribution of $\lambda^f(t_i)$ is equal to $\lambda^a(t_{i-1})$: only the observations have an impact on the distribution.

2.2 Numerical Model

The Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS; Hodur, 1997) is used as the numerical model for these experiments. Although COAMPS is capable of running in either an idealized or real-data mode, simulating tropical convection over an idealized domain gave unphysical results, so only the real-world mode is used here. Initial conditions are generated by interpolating a global NOGAPS forecast onto the limited-area domain and integrating forward for 6 hours to allow time for the interpolated state to adjust to the COAMPS model dynamics.

The specification of boundary conditions can have a profound effect on mesoscale predictability (Warner et al., 1997), and can affect ensemble spread (Nutter et al., 2004a,b). The boundary conditions for each ensemble member are drawn from a member of a NOGAPS ensemble (which was constructed using the ensemble transform (Bishop and Toth, 1999) technique) and are updated every six hours using the flow relaxation method of Davies (1976).

The domain chosen is from 0°-13°N latitude and 180°E to 220°E longitude, which is located over the ocean in the mid-Pacific Ocean and shown in Figure B-1. The base time of the analysis is 00Z on 22 June 2005. The computational grid uses 30 vertical levels and 25 km horizontal spacing, with a timestep of 100 seconds. Convection is parameterized using the Kain-Fritsch cumulus parameterization (Kain, 2004). Ice and graupel physics, along with surface fluxes, are enabled. Turbulence is parameterized using the Richardson-number dependent method of Mellor and Yamada (1982).

[1]The estimate uses the observations to update $p(\lambda, t_i) \propto p(y^o|\lambda)p(\lambda, t_{i-1})$ - a product of the likelihood that the observation y^o would be observed given an inflation factor λ and the previous distribution for λ. While both distributions are Gaussian, their product is not - the filter, though, assumes a Gaussian and calculates the mean and variance as if it were (Anderson, 2007).

2.3 Verification

The metrics used to quantify ensemble performance follow those used in Dirren et al. (2007) and will evaluate the ensemble based on RMS error and ensemble spread. The spatial-averaged RMS error in the ensemble mean $\overline{X}$ from truth X^{true} is

$$\text{RMS}_{\text{xy}}(t) = \sqrt{\frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \left(\overline{X}(x_i, y_j, t) - X^{\text{true}}(x_i, y_j, t)\right)^2}. \tag{2.16}$$

The ensemble spread for an ensemble of size N is defined as

$$\sigma_{\text{ens}}(x, y, t) = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left(X^{(i)}(x, y, t) - \overline{X}^{(i)}(x, y, t)\right)^2}, \tag{2.17}$$

where the superscripted index indicates the ensemble member.

The primary variable for verification is the relative humidity, due to its sensitivity (Emanuel, 1994).

Chapter 3

Results

This chapter demonstrates some of the results from the ensemble system. Due to limited computational power, the ensemble size is reduced from 128 members to 50 members, and the approximate effect is shown. Several observing system experiments are run and compared with a no-assimilation control for analysis error and ensemble spread.

3.1 Ensemble Size

Using the SVD method outlined in 2.1.3, the decay of normalized singular values with ensemble size for the subset of the NOGAPS model output over the experimental domain is shown in Figure B-2. These estimates are based on an initial 128-member ensemble of relative humidity at 200mb, 50mb, 700mb, 850mb, and 100mb, taken from the forecast at 24 and 48 hours lead time. The values decay rapidly, and drop below 0.1 around 40-60 ensemble members for most of the levels. A truncation to 50 members is much more computationally feasible and includes a good portion of the variability. Contributions from many small singular values is usually indicative of noise, so the shape and evolution of the 200mb curve may be indicative of higher variability due to small absolute vapor content and distance from the moisture source at the surface. Although the ensemble spread increases with time, the effective rank is expected to fall as time advances - an initial isotropic error stretches along the

model attractor under the influence of the model dynamics. The lack of this behavior is most likely due to the consideration of only a small part of the entire NOGAPS domain.

The results are similar with the 128-member COAMPS ensemble, shown in Figure B-3, using the water vapor mixing ratio field instead of the relative humidity field. This shows that 50 members still is not extremely low, but now the top curves are the lower levels instead of upper levels. This may be due to the increased resolution showing more of the effects of the surface on smaller scales. This decay pattern is roughly the same for other variables, but is not shown. Unlike the case of the global model subset, the effective rank decreases with time. The choice of 50 members is consistent with other mesoscale ensembles: Zhang et al. (2006) used 40 members and Snyder and Zhang (2003) used 50.

3.2 Lateral Boundary Conditions

Warner et al. (1997) predicted that given a regional forecast domain with length scale 2000 km, the approximate range of a useful forecast in the tropics is about 36 hours. After this time, the mesoscale model continues to run and produce small-scale structures, but they are completely determined by the boundary conditions. Although this can be useful by providing a dynamically consistent downscaling, it represents a significant limit to the mesoscale predictability.

To test the influence of lateral boundary conditions on the limited-area solution, a 10-member ensemble is integrated using non-perturbed boundary conditions. This allows the ensemble members to collapse to a single solution soley determined by the boundary condition. Figure B-4 shows the spread for the water vapor mixing ratio field - spread grows with time in the domain interior, but along the edges it quickly decreases. While this presents difficulty for the general consideration of tropical mesoscale predictability, the integrations here are short enough to avoid most adverse effects.

To test the effect of an actual mesoscale forecast versus a simple downscaling

of the global model forecast, the global forecast at 6, 18, and 24 hour lead time is interpolated onto the model grid and briefly integrated forward to generate fine-scale detail rather than the interpolated state. The COAMPS forecast and the downscaled NOGAPS forecast of 500 hPa relative humidity, together with the absolute difference between the two at $\tau = 6$ hours are shown in Figure B-5. Many of the large-scale details are similar, but the COAMPS forecast shows more than simply sharpening the NOGAPS forecast: near 186°E and 9°N, there is a region of high RH that does not appear at all in the downscaled forecast - there are large differences as well in the NE region of the domain.

3.3 Lower Boundary Conditions

One of the significant results seen in all the assimilation cases is the importance of the *lower* boundary condition (i.e. the sea surface). The low-level moisture ensemble spread for one of the experiments is shown in Figure B-6 - the other assimilation experiments showed similar results. The domain interior has a near-zero spread, with the maximum value occurring along the boundaries, driven by the encroachment of coarse-domain moisture into the domain from the perturbed lateral boundaries. The NOGAPS ensemble used to provide initial and boundary conditions was run with an unperturbed sea-surface temperature based on an SST analysis (J. McLay, personal communication), so no varying lower boundary consistent with the varying lateral boundaries was available. In the global model, though, there is considerable spread in the 1000 hPa relative humidity field throughout the integration, suggesting that the mesoscale model offers a stronger surface coupling than found in NOGAPS. The entropy[1] budget for the subcloud layer in a convecting atmosphere contains contributions from surface sensible and latent heat fluxes, convective downdrafts, entrainment at the layer top, and radiative cooling (Emanuel et al., 1994): all of these can be configured and parameterized very differently between a global and mesoscale

[1]Moist entropy includes contributions from temperature, pressure, and the moisture content of the air and is conserved under reversible moist adiabatic transformations (Emanuel, 1994, p. 120)

model. The capability of COAMPS to be a coupled mesoscale model may prove to be powerful for including the strong effects of air-sea interaction in the tropics.

3.4 Evolution of the True State

This section briefly describes the evolution of the model trajectory from which observations are drawn and which is used as the true state for verification of the analyses.

3.4.1 Coarse Domain

The time evolution of the NOGAPS control member's relative humidity and wind field is shown in Figures B-7, B-8, and B-9. The domain encompasses a small region of the ITCZ. In the lower troposphere, easterly flow with a small meridional component dominates the domain, with a westerly shear. At 500mb, relative humidity levels drop, consistent with the mid-tropospheric drying that is observed in the tropics. The flow remains easterly, with a convergence zone midway through the domain. Higher up at 200mb, the higher relative humidity consistent with convective anvils is evident. The wind is now westerly and there is a region of confluence on the western edge. There is a strong jet just to the north of the domain.

3.4.2 Fine Domain

Results from the mesoscale domain are similar to the global domain. There is a strong influence on boundaries where the large-scale flow is into the mesoscale domain. Figure B-10 - B-12 shows the responses for 6, 12, 18, and 24 hours at 200, 500, and 700 hPa. The strengthening of the anticyclonic 200 hPa flow is much more evident in the fine data than the global data, and the relative humidities are different thanks to different parameterizations and rounding errors used in the interpolation (since these are taken directly from height levels and the global files were available on pressure levels).

3.5 Unconstrained Ensemble

As an experimental control, the ensemble is integrated forward in the absence of any constraining observations[2]. With the exception of the boundary layer moisture noted in 3.3, the spread of the ensemble is concentrated in the northern half of the domain. Figure B-13 shows the spread in the zonal wind field at approximately 850mb concomitant with the relative humidity maximum, so the spread is being largely influenced by uncertainty in the convective region.

3.6 Observing System Experiments

These experiments look at the influence of observations on analysis error and ensemble spread. As a first effort, the simplest forward operators to construct are for wind, moisture, and temperature variables. Ramamurthy and Carr (1988) found that for the 1979 Monsoon Experiment (MONEX), the inclusion of wind and moisture in their objective analysis provided the greatest utility, so they will be used first.

Instead of using the DART/COAMPS support for interpolation and arbitrary observation operators, observations of wind and moisture are made directly from the state vector. This allows the assimilation to run much faster, and eliminates projection to observation space from contributing to the analysis error. Observations are made on a 3-by-2 grid evenly spaced throughout the domain, for a total of 6 observations per level.

The error characteristics of the synthetic wind- and moisture-measuring instruments were chosen to reflect a mix of operational characteristics. The observation error variance for wind variables is defined to be 8 $m^2\ s^{-2}$. This is chosen to be between the radiosonde errors used by Mitchell et al. (2002) and the cloud-track wind errors used by Tsuyuki (1996). The observation error for the moisture field is defined to be 5% of the average field value at the initial time (shown in Table A.1). This was chosen since it yields a relative humidity error of a few percent, which is consistent

[2]In DART, this is done by observing a single state variable with a very large variance (such as 9999999999.0), ensuring that the gain term will be zero.

with a noisy operational instrument.

The vertical distribution of relative humidity analysis RMSE is shown in Figure B-14 - for comparison, the RMSE of the control (unconstrained) ensemble mean is subtracted from each experiment's results. The two levels where the observations are taken are denoted by black horizontal lines. The values are not all negative, where values less than zero indicate that assimilating the observations gives an analysis closer to truth. Counting points shows that over the course of the assimilations, approximately half the realizations (each level in each experiment) are further from truth than the unconstrained case. The greatest improvement was in the upper troposphere when assimilating water vapor mixing ratio on the upper level.

At observation locations, the difference between the first-guess field and the analyzed field is a function of the observation value and the first-guess value only (combined with their respective uncertainties). Away from observation locations, the increment relies on that difference in conjunction with the ensemble-estimated error covariance. If this covariance estimate is poor, the analysis may be degraded regardless of the quality of the observation. However, looking at the vertical distribution of error shows that while there are some improvements on unobserved levels, there are positive relative errors even on the observed levels.

A demonstration of the increments in observation space is shown in Figure B-15. These show that the filter is performing its expected function - it pulls the estimate toward the observation and reduces the uncertainty of the final estimate. In the first figure, B-15(a) (from observing low-level moisture), this results in an analysis that is closer to the true value than before. However, in the second, B-15(b) (from observing upper-level wind), this results in pulling the analysis further from truth. The observation is not drawn from a remote tail of the distribution, but still results in being further from truth than the model – this is one of the dangers of a sparse observing network. In the presence of a sufficient number of observation realizations, the random errors in nearby observations in space and time should partially cancel, reducing the impact of a single observation. The choice of a small localization radius exacerbates this problem, since the sparse network is prohibited from updating any

of the other observation locations. The adaptive inflation, meant to account for the overconfidence in the first guess, gave an inflation factor of approximately 1.08, which spreads out the prior probability density slightly more than shown.

The localization used for these experiments is quite strong, but based on results from idealized runs for other tropical experiments (G. J. Hakim, personal communications). Nevertheless, it may be too small for real-world conditions. Static localization radii and inflation factors can be chosen, but require a large initial investment of computational time.

The observation error distribution may appear much broader than it should be, but the values chosen exemplify the difficulty in finding quality in situ observations in the tropics. Tsuyuki (1996) noted that the NCEP error standard deviations for cloud track winds were 6.1 m/s at the higher level and 3.9 m/s on the lower level - both much higher than the value used here (but note that satellite products would offer many more than 6 points, somewhat compensating for the broader error distribution). Lower errors have also been used - Leidner et al. (2003) used a more accurate wind error of 2 m/s when assimilating satellite scatterometer winds, but that is only slightly less than the error here.

Chapter 4

Conclusions and Future Work

This work set out to construct and use an ensemble data assimilation system for a mesoscale model in the tropics.

While a larger ensemble is preferable, examination of the forecast error covariance matrices show that truncating the ensemble to 50 members retains enough of the variability to give credence to the results.

Lateral boundary conditions were an important limitation to the solution, since as the integration time continues the solution becomes merely a high-resolution reflection of the boundaries. The mesoscale model added detail not found in a downscaled global model forecast valid at the same time. An unperturbed sea-surface temperature field was used as a lower boundary condition, and the boundary layer moisture spread in the interior of the domain quickly dropped to zero, emphasizing the importance of accurately capturing air-sea interaction in the tropics.

Most of the ensemble spread is in the region of active convection. Convection is parameterized, so statistical effects rather than specific cells are simulated - still, though, it remains a mechanism for generating strong uncertainty when compared to non-convective regions.

Observing system experiments were carried out using synthetic observations of wind and moisture taken directly from the model state vector to minimize the error from an inaccurate observation operator. The observations had small effects on the state estimate, but actually degraded the analysis relative to an unconstrained

ensemble in several locations due in part to the sparseness of the network. Future experiments will incorporate denser networks, as well as non-identity observations, which should give some stronger constraints for how dense a network must be.

It should be noted that since these experiments were run in a perfect model framework, these conclusions apply only insofar as the model is correct. Also, the use of actual data as the basis for these experiments may limit their validity as well. Several methods for dealing with these shortcomings and expanding the interpretation of these results are presented below.

This work touches on several preliminary experiments, but also has developed a framework that allows for simple expansion of the research presented herein. While the possibilities are many, there are several natural extensions that would be quite insightful.

Since the tropics represent an area of relatively sparse observations compared to midlatitudes (especially terrestrial areas), the impact of observational network density can be profound. While satellite observations can provide information over a larger area, a network of various in situ observations may be more useful in conjunction with the ensemble data assimilation procedures used here.

Experiments with satellites observations and radar observations require the definition of more advanced forward operators to give an ensemble estimate of the observation. These will require more computation (e.g. the use of a radiative transfer code), but can show the response to the type of observations that are common for the tropical atmosphere.

The success of ensemble methods depend largely on the validity of the Monte Carlo methods to sample the forecast and analysis probability distributions. More members can increase the dimension of the state subspace spanned by the ensemble, and so provide a more accurate estimate of the error statistics. Although these results were limited by the number of global ensemble members available, Torn et al. (2006) give a method for perturbing boundary conditions that would allow an ensemble of arbitrary size based on a single global model run. Since air-sea interaction is so important in the tropics, the surface boundary condition may need to be perturbed

as well, or even provided by a coupled ocean model.

The results here for a single case of a real-world model run. Although more cases can provide more applicable results, experiments with an idealized model can shed light on important relations and processes free of constraints like boundary conditions. These idealized studies can be constructed using a method similar to Bretherton et al. (2005), with a long integration over a small area to provide a radiative-convective equilibrium solution that is used as an initial condition for a larger area. Running with doubly periodic boundary conditions with a static, uniform sea surface temperature can provide a baseline for ensemble perturbation and subsequent assimilation experiments. Since moist convection plays such an important role in the tropical atmosphere, model configurations run with convection-permitting resolutions and parameterized convection can assist in evaluation of physical parameterizations.

The use of parameterizations for moist convection, microphysics, turbulence, radiation, etc., together with the assumptions and approximations inherent in the development in any numerical model, can introduce significant errors in the analysis and forecasts. The experiments presented here assume a perfect model, since the observations are derived from a reference model run, but any system used in the real world must deal with the inadequacy of numerical models in simulating all scales of motion and all physical processes. Although using multi-model ensembles can produce estimated states that bound the true attractor, care must be taken to ensure that the multi-model information does not pull the constituent models off their respective attractors (Hansen, 2002). Dealing with model error is an active area of ongoing research, and is essential to the success of these data assimilation methods.

The Kalman filter is a simplification of the generalized nonlinear Bayesian filtering problem assuming that the error statistics are completely described by their first and second moment. The effects of non-Gaussian distributions can be tested by both normalizing these results using the stochastic EnKF, which can affect the modification of higher moments differently than the EAKF (Lawson and Hansen, 2004), and by using more advanced methods such as a kernel filter (Anderson and Anderson, 1999, see Fig. 2) that uses Gaussian kernels to form an arbitrary probability distribution.

These are few of many future directions, but the framework and foundation developed in the present work can facilitate the implementation and investigation of many of them, furthering progress toward Bjerknes's challenge: greater understanding of mesoscale tropical physics and dynamics, and an improved ability to estimate the state of the tropical atmosphere with an eye toward improving predictions that can have global impact.

Appendix A

Tables

Index	Height (m)	$\overline{p}$ (hPa)	$\overline{T}$ (°C)	$\overline{q_v}$ (g kg^{-1})	$\lvert\overline{u}\rvert$ (m s^{-1})	$\lvert\overline{v}\rvert$ (m s^{-1})	σ^o_{uv}(m s^{-1})	σ^o_q (g kg^{-1})
1	31,050	9.88	-47.25	0.00	25.29	24.00	–	–
2	24,400	27.12	-55.39	0.01	16.29	15.69	–	–
3	19,400	61.11	-69.72	0.01	1.10	2.08	–	–
4	16,050	108.89	-78.31	0.00	6.66	0.85	–	–
5	14,300	147.07	-68.94	0.01	8.86	4.76	–	–
6	13,300	173.18	-61.15	0.03	8.65	5.67	–	–
7	12,425	199.20	-54.15	0.07	8.25	6.33	–	–
8	11,675	223.33	-48.08	0.13	7.74	5.50	2.83	.0065
9	10,925	250.26	-42.03	0.23	7.20	4.57	2.83	.0115
10	10,175	279.13	-36.10	0.37	6.04	2.99	2.83	.0185
11	9,425	310.56	-30.33	0.52	4.78	1.66	2.83	.0260
12	8,675	344.52	-24.76	0.66	3.33	0.88	2.83	.0330
13	7,800	388.20	-18.36	0.81	1.64	0.03	2.83	.0405
14	6,800	442.99	-11.74	1.21	0.25	1.07	2.83	.0605
15	5,800	504.15	-5.45	1.75	0.93	2.08	2.83	.0875
16	4,800	571.21	-0.03	3.36	1.36	2.55	2.83	.1680
17	3,900	638.35	4.78	4.82	1.75	2.98	2.83	.2410
18	3,100	703.76	8.95	6.18	2.17	3.39	2.83	.3090
19	2,300	774.10	12.15	9.00	3.75	4.28	2.83	.4500
20	1,600	840.71	14.91	11.34	5.15	5.07	2.83	.5670
21	1,100	891.05	18.11	12.38	5.95	5.21	2.83	.6190
22	750	927.80	20.53	12.90	6.42	5.19	2.83	.6450
23	500	954.60	22.08	13.99	6.26	4.69	2.83	.6995
24	330	973.22	23.13	14.71	6.16	4.35	2.83	.7355
25	215	986.01	23.84	15.17	6.09	4.13	2.83	.7585
26	140	994.42	24.30	15.46	6.04	3.98	2.83	.7730
27	90	1000.07	24.61	15.63	6.00	3.88	2.83	.7815
28	55	1004.06	24.84	15.66	5.99	3.87	2.83	.7830
29	30	1006.92	25.00	15.67	5.99	3.87	2.83	.7835
30	10	1009.22	25.13	15.68	5.99	3.87	2.83	.7840

Table A.1: Information on model levels used in constructing the observational networks. The level index and the geometric height of the level are static throughout the run. The mean pressure, vapor mixing ratio, and wind speed components are taken from the initial analysis. The observational error standard deviation σ^o is set to $\sqrt{8}$ m s^{-1} for the wind components and to 5% of the mean vapor mixing ratio for the moisture field.

Appendix B

Figures

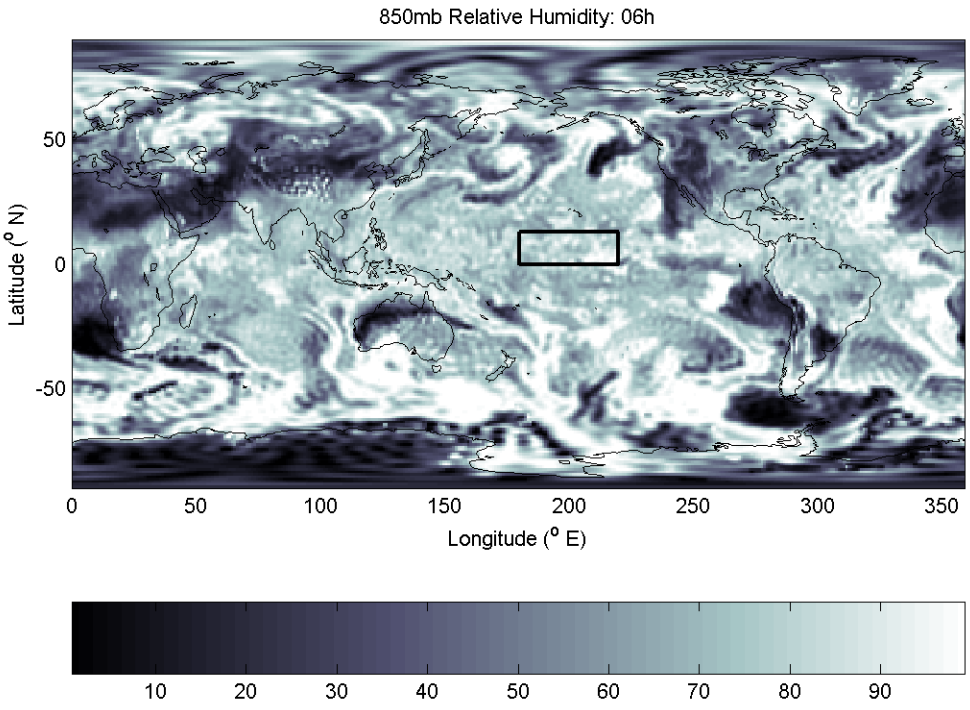
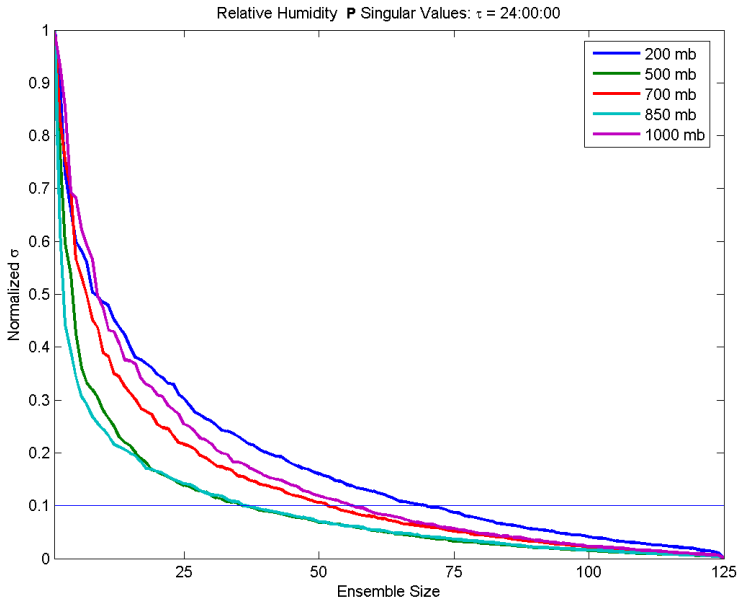
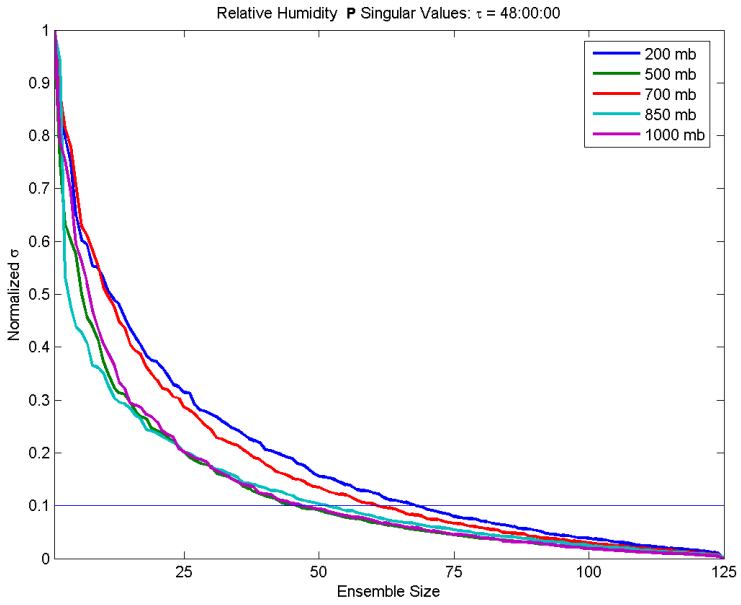


Figure B-1: Location of the model domain. The shaded box shows the limits of 0° to 15°N latitude and 190°E to 215°E longitude.

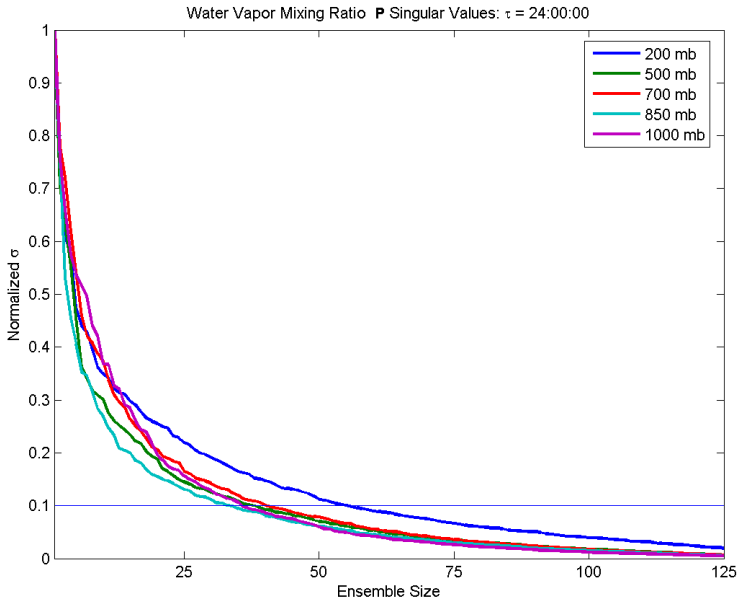


(a) $\tau = 24$ hours

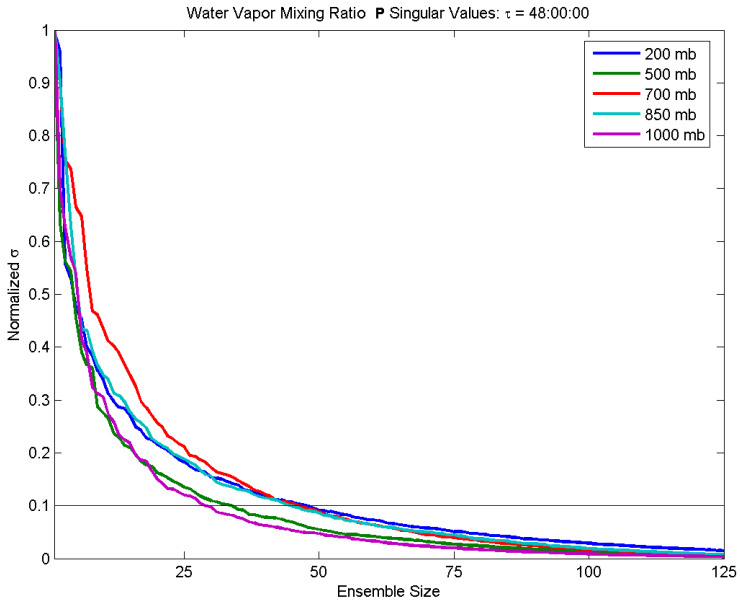


(b) $\tau = 48$ hours

Figure B-2: NOGAPS forecast relative humidity covariance matrix singular values for multiple pressure levels, showing the effect of using fewer ensemble members on the accuracy of the covariance matrix representation.

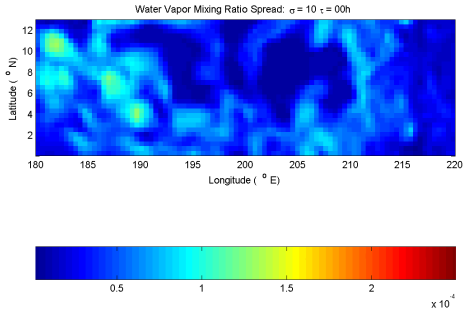


(a) $\tau = 24$ hours

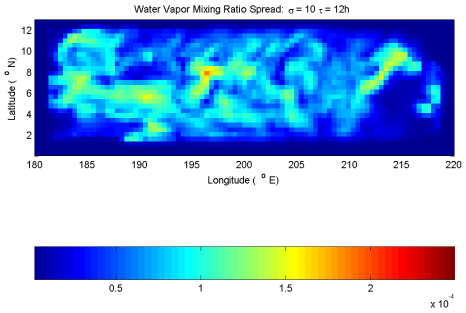


(b) $\tau = 48$ hours

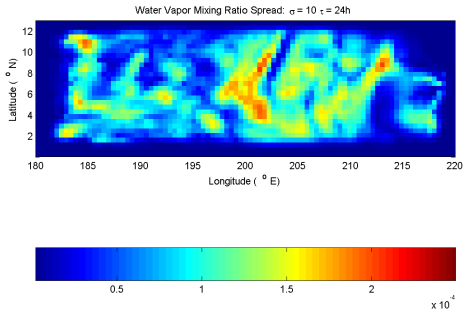
Figure B-3: COAMPS forecast water vapor mixing ratio covariance matrix singular values for multiple pressure levels, showing the effect of using fewer ensemble members on the accuracy of the covariance matrix representation.



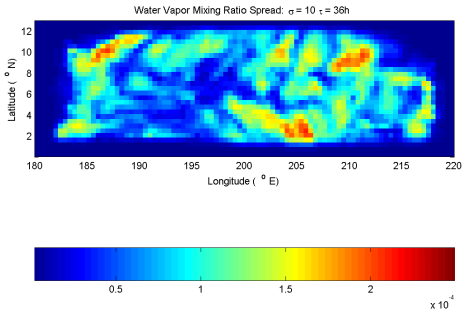
(a) $\tau = 00$ hours



(b) $\tau = 12$ hours

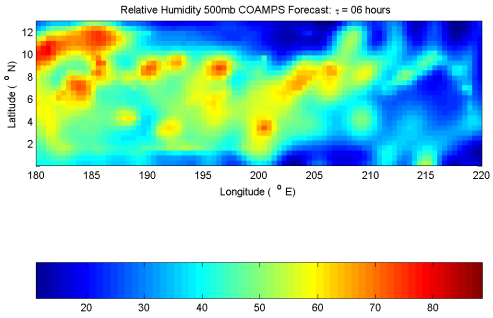


(c) $\tau = 24$ hours

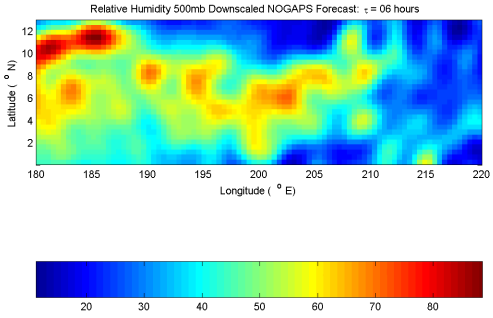


(d) $\tau = 36$ hours

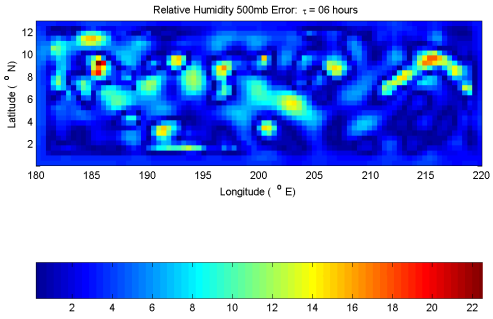
Figure B-4: Ensemble spread of water vapor mixing ratio (kg/kg) at level $\sigma = 10$ (approximately 300 hPa) integrated with non-perturbed boundary conditions. This ensemble consisted of 10 members to show the quick decay of ensemble spread near the edges of the domain and moving inward with time.



(a) COAMPS Forecast

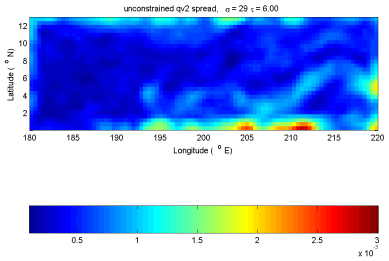


(b) Downscaled NOGAPS forecast

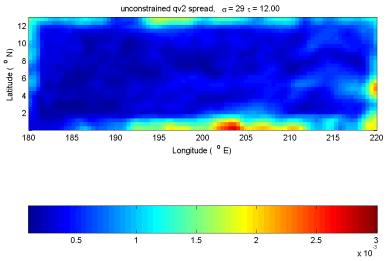


(c) Difference ($\sqrt{(\mathrm{RH}_{\mathrm{forecast}} - \mathrm{RH}_{\mathrm{downscaled}})^2}$)

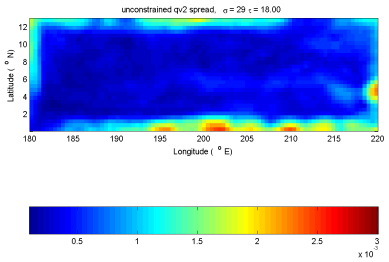
Figure B-5: Difference in 500 hPa relative humidity between a 6-hour COAMPS forecast based on NOGAPS initial conditions and a COAMPS forecast based on the NOGAPS 6-hour forecast, showing the added value of the mesoscale forecast vs. a simple downscaling of the global model.



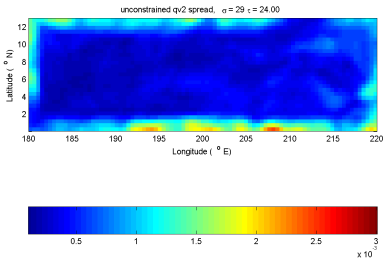
(a) $\sigma = 29$, $\tau = 06$ hours



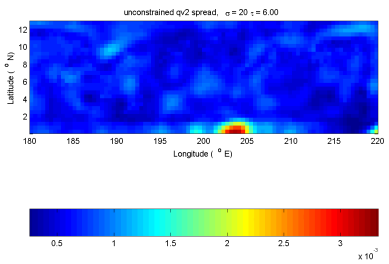
(b) $\sigma = 29$, $\tau = 12$ hours



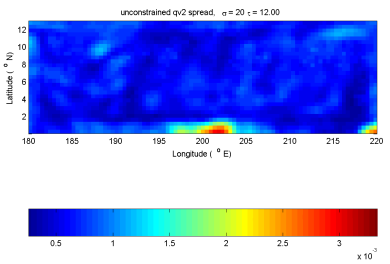
(c) $\sigma = 29$, $\tau = 18$ hours



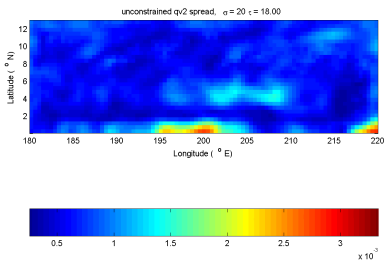
(d) $\sigma = 29$, $\tau = 24$ hours



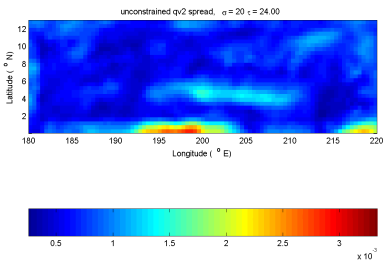
(e) $\sigma = 20$, $\tau = 06$ hours



(f) $\sigma = 20$, $\tau = 12$ hours

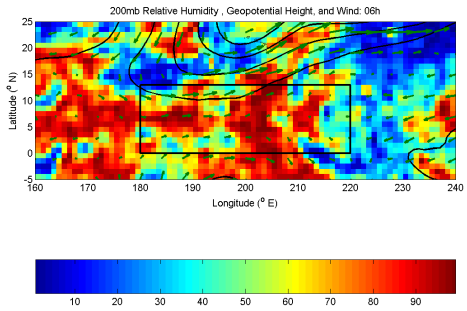


(g) $\sigma = 20$, $\tau = 18$ hours

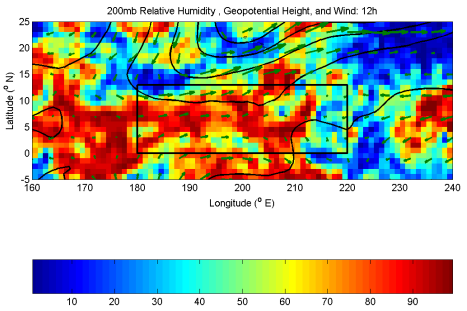


(h) $\sigma = 20$, $\tau = 24$ hours

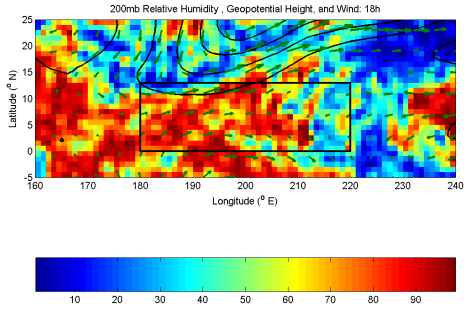
Figure B-6: Ensemble spread for low-level moisture (kg/kg) showing the very low levels of spread due to the lower boundary conditions.



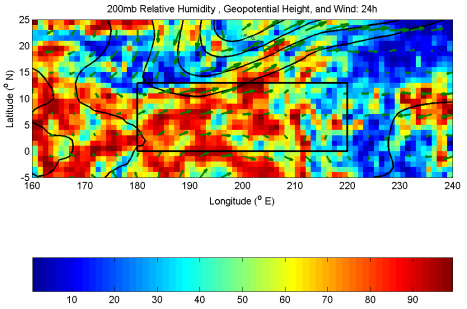
(a) $\tau = 06$ hours: $\max(|\boldsymbol{u}| = 38.06$ m/s



(b) $\tau = 12$ hours: $\max(|\boldsymbol{u}| = 39.71$ m/s

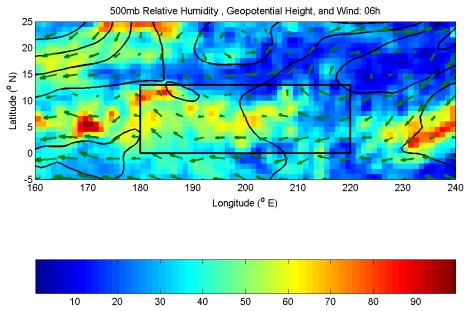


(c) $\tau = 18$ hours: $\max(|\boldsymbol{u}| = 38.31$ m/s

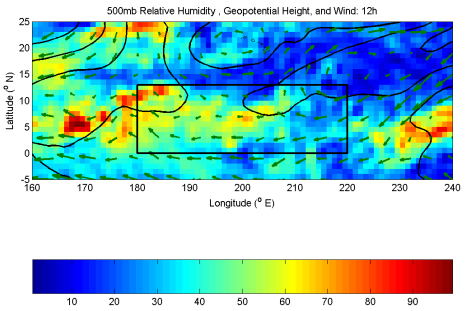


(d) $\tau = 24$ hours: $\max(|\boldsymbol{u}| = 37.39$ m/s

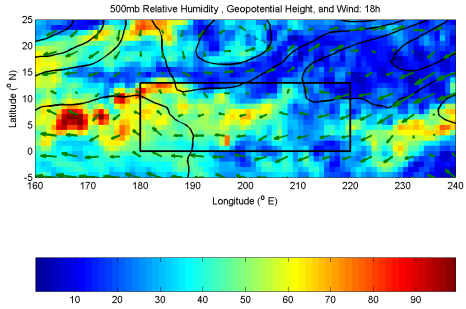
Figure B-7: NOGAPS control forecast, showing the time evolution of 200 hPa relative humidity (in percent), geopotential height, and wind. Maximum wind speed is given for scale.



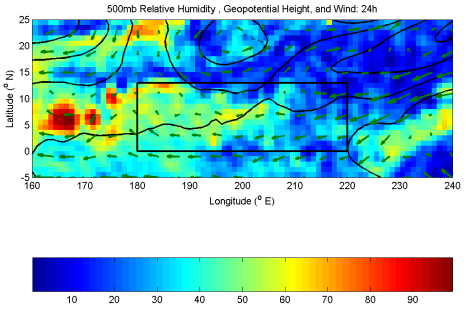
(a) $\tau = 06$ hours: $\max(|\boldsymbol{u}| = 15.23$ m/s



(b) $\tau = 12$ hours: $\max(|\boldsymbol{u}| = 16.10$ m/s

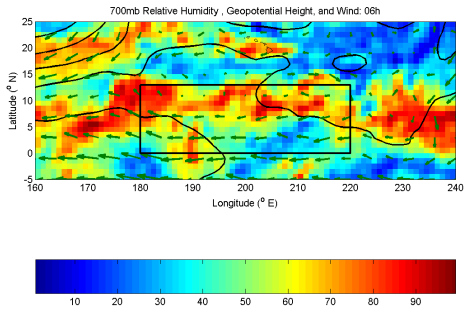


(c) $\tau = 18$ hours: $\max(|\boldsymbol{u}| = 16.90$ m/s

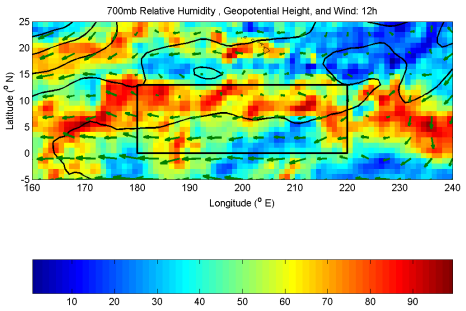


(d) $\tau = 24$ hours: $\max(|\boldsymbol{u}| = 17.19$ m/s

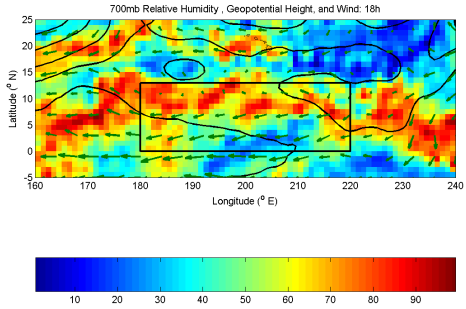
Figure B-8: NOGAPS control forecast, showing the time evolution of 500 hPa relative humidity (in percent), geopotential height, and wind. Maximum wind speed is given for scale.



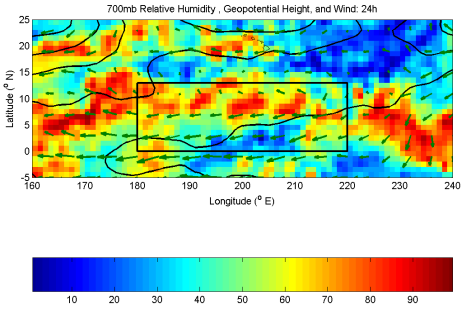
(a) $\tau = 06$ hours: $\max(|\boldsymbol{u}| = 14.10$ m/s)



(b) $\tau = 12$ hours: $\max(|\boldsymbol{u}| = 15.62$ m/s)

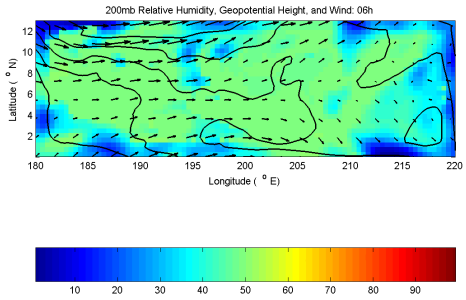


(c) $\tau = 18$ hours: $\max(|\boldsymbol{u}| = 14.77$ m/s)

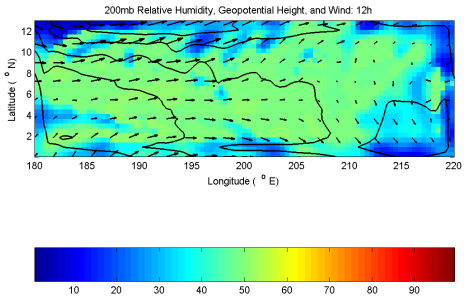


(d) $\tau = 24$ hours: $\max(|\boldsymbol{u}| = 14.22$ m/s)

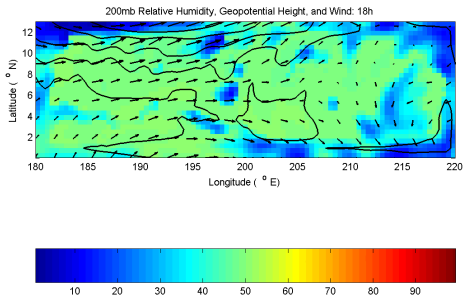
Figure B-9: NOGAPS control forecast, showing the time evolution of 700 hPa relative humidity (in percent), geopotential height, and wind. Maximum wind speed is given for scale.



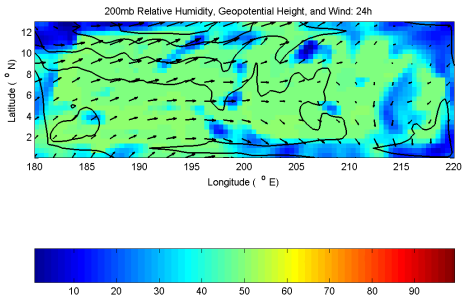
(a) $\tau = 06$ hours: $\max(|\boldsymbol{u}| = 26.90$ m/s



(b) $\tau = 12$ hours: $\max(|\boldsymbol{u}| = 28.99$ m/s

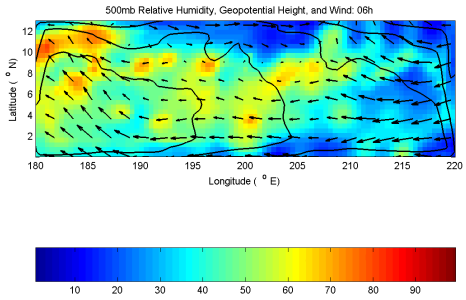


(c) $\tau = 18$ hours: $\max(|\boldsymbol{u}| = 30.14$ m/s

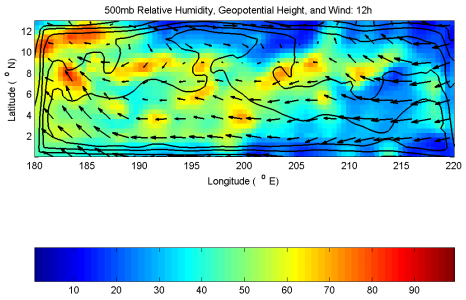


(d) $\tau = 24$ hours: $\max(|\boldsymbol{u}| = 31.56$ m/s

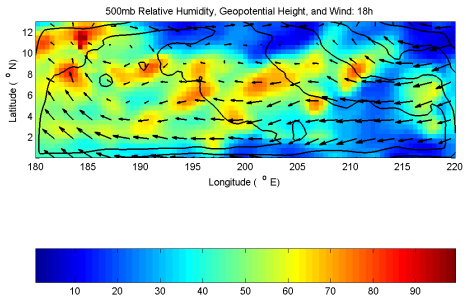
Figure B-10: COAMPS control forecast, showing the time evolution of 200 hPa relative humidity (in percent), geopotential height, and wind. Maximum wind speed is given for scale.



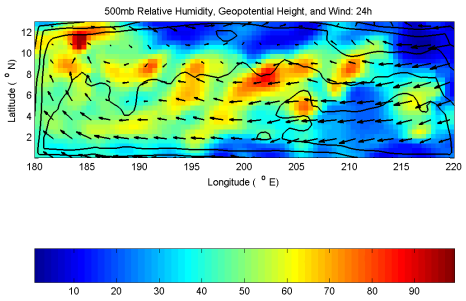
(a) $\tau = 06$ hours: $\max(|\boldsymbol{u}| = 10.02$ m/s



(b) $\tau = 12$ hours: $\max(|\boldsymbol{u}| = 10.22$ m/s

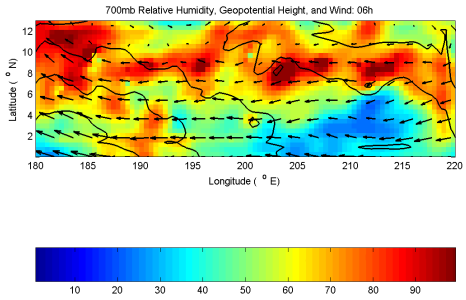


(c) $\tau = 18$ hours: $\max(|\boldsymbol{u}| = 11.00$ m/s

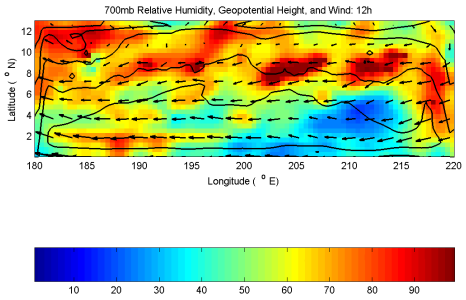


(d) $\tau = 24$ hours: $\max(|\boldsymbol{u}| = 12.06$ m/s

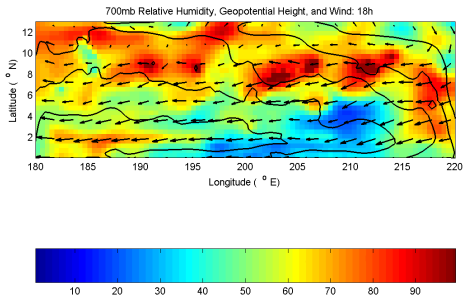
Figure B-11: COAMPS control forecast, showing the time evolution of 500 hPa relative humidity (in percent), geopotential height, and wind. Maximum wind speed is given for scale.



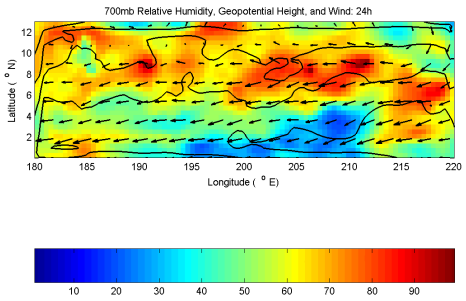
(a) $\tau = 06$ hours: $\max(|\boldsymbol{u}| = 11.84$ m/s)



(b) $\tau = 12$ hours: $\max(|\boldsymbol{u}| = 12.57$ m/s)

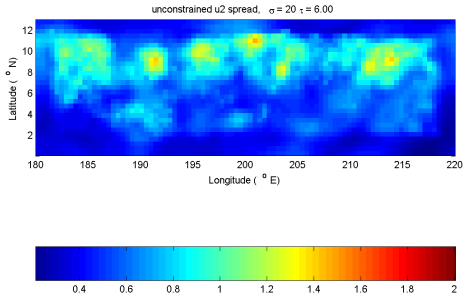


(c) $\tau = 18$ hours: $\max(|\boldsymbol{u}| = 12.38$ m/s)

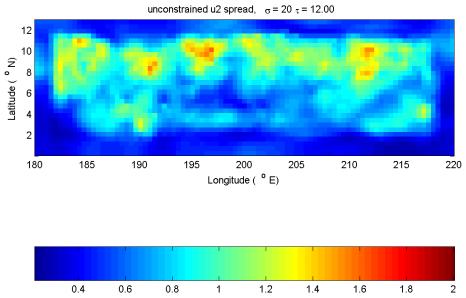


(d) $\tau = 24$ hours: $\max(|\boldsymbol{u}| = 12.90$ m/s)

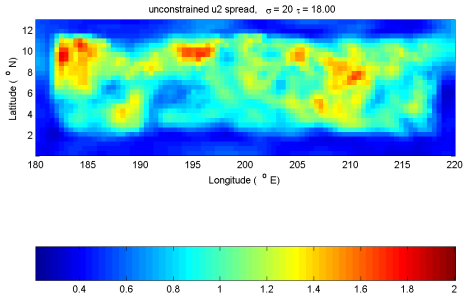
Figure B-12: COAMPS control forecast, showing the time evolution of 700 hPa relative humidity (in percent), geopotential height, and wind. Maximum wind speed is given for scale.



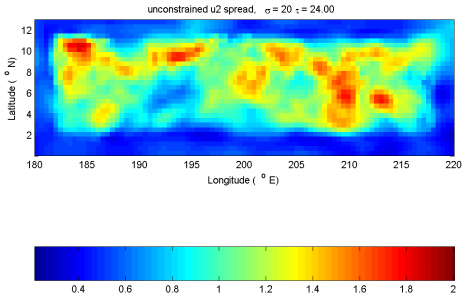
(a) $\tau = 06$ hours



(b) $\tau = 12$ hours

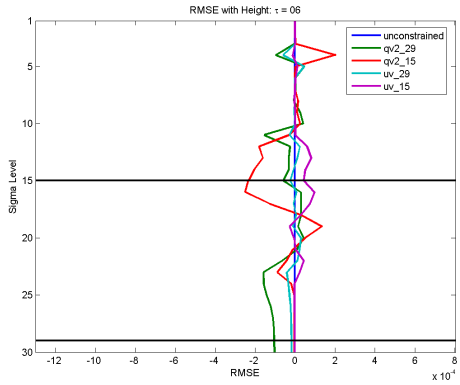


(c) $\tau = 18$ hours

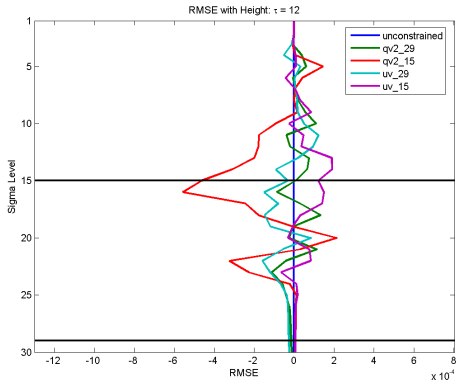


(d) $\tau = 24$ hours

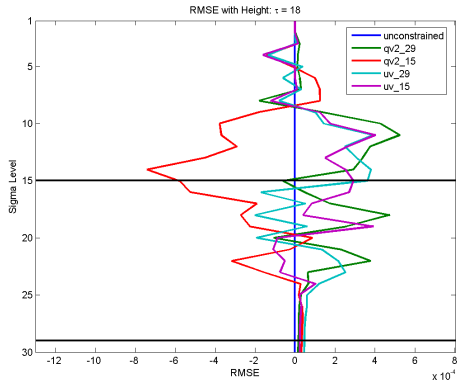
Figure B-13: COAMPS unconstrained ensemble spread, showing the time evolution of $\sigma = 20$ (approximately 850 hPa) zonal wind component (m/s).



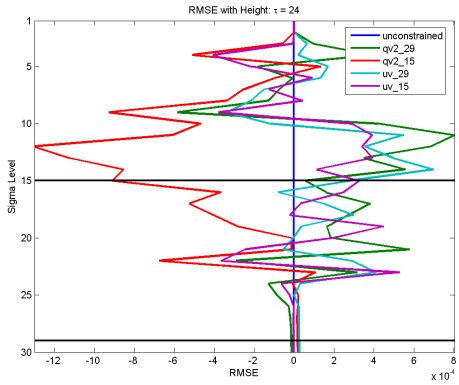
(a) $\tau = 06$ hours



(b) $\tau = 12$ hours

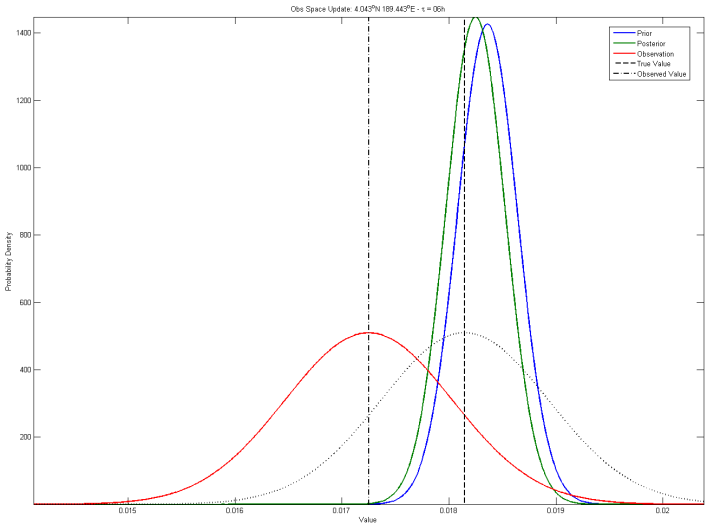


(c) $\tau = 18$ hours

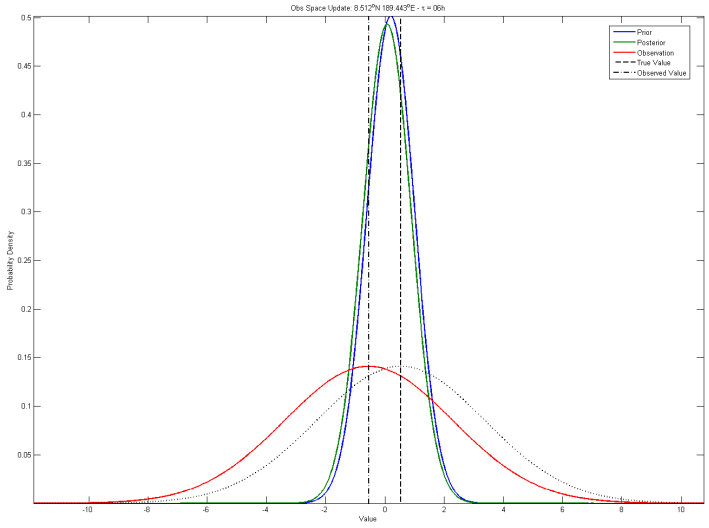


(d) $\tau = 24$ hours

Figure B-14: Vertical distribution of COAMPS relative humidity (unitless) ensemble mean RMSE for all four observing experiments at the four assimilation times. Observations are taken at the levels shown by the horizontal lines.



(a) Lower-Level Water Vapor Mixing Ratio



(b) Upper-Level Wind Component

Figure B-15: Analysis increments in observation space showing the estimate moving closer to the observation and reducing the uncertainty. The first-guess (prior) estimate PDF is shown in green, the analysis (posterior) is shown in blue, and the observation PDF is in red. The actual value of the observation is shown by the dot-dash line for comparison with the true value, shown by the vertical dotted line. The PDF that the synthetic observation was drawn from is given by the dotted PDF.

Appendix C

Software Implementation

This appendix describes in more detail the actual implementation of the model in the ensemble data assimilation system.

C.1 The Data Assimilation Research Testbed

The Data Assimilation Research Testbed (DART) is a project by the National Center for Atmospheric Research's Data Assimilation Research Section (DAReS) designed to "facilitate the combination of assimilation algorithms, models, and ... observations to allow increased understanding of all three." (DART documentation). The DART framework provides an easy interface to add new models and new observation types, scales from simple two-variable models to global climate circulation models, and provides several implementations of various ensemble filter algorithms.

One of DART's main strengths is the extensive ability to deal with both real and synthetic observations. The user can define observation types from simple identity observations (i.e. observing elements of the state vector) to complex observations like satellite radiances or radio occulation data, provided that a suitable (possibly non-linear) observation operator exists to map the model state vector to the observation. These observation definitions, combined with appropriate error statistics, are used to define complete observational networks for the assimilation. If synthetic observations are necessary (such as in "perfect model" experiments), DART integrates the model

forward, sampling values at each observation time to yield both observations and a reference model trajectory.

Once the observations are defined, the user can choose among several different ensemble filters, options for covariance localization and inflation, easily change the ensemble size and output characteristics, and allow for either existing perturbations to be used or for the software to perturb a single model state to generate the ensemble.

For the case of simple models, the model code can be embedded entirely in the DART implementation of the model interface - however, for most full numerical weather prediction models, the assimilation software hands off the state vector to the model, which is run externally to advance to the next observation time.

C.2 DART interface to COAMPS

The COAMPS model is a modern mesoscale numerical weather prediction model. While this provides powerful opportunities for experiments in a variety of idealized and operational configurations, it requires a dynamic, adaptable interface to communicate with the DART assimilation system. As described in Section C.1, complex models require an asynchronous integration for advancing the model to the next observation time. It is this asynchronous configuration that is the cornerstone of the DART COAMPS interface.

C.2.1 State Vector Definition

The main task if the model is run in an asynchronous mode is to translate the state vector from the form handled by DART into a state vector recognized by the model. For large models like COAMPS, the model state vector is stored in a restart file that contains information that is necessary for the model (e.g. timestepping data), but not needed for assimilation. The translation from a COAMPS restart file to a DART state vector file, then, requires the selection and consolidation out of disjoint pieces of the full restart file.

The multiple configuration possibilities of the COAMPS model means that the

translation method must be flexible as well. Each variable in the model consists of one or more levels of two-dimensional fields. This leads to the use of a state vector definition file, which contains a list of each field to include in the DART state vector. For example, to assimilate temperature in a model with a 50x50 horizontal grid and 30 vertical levels requires 30 entries, representing one 50x50 grid at every level. Each entry in the state vector definition file contains information not only about the variable name and the vertical level, but perturbation size and structure, the type of vertical level on which the variable is defined, the DART observation type of the variable, whether the variable should be assimilated (since some entries are necessary in the state vector for computation of observation operators, but should not be changed), whether the field is a mean field (since COAMPS operates on mean and perturbation fields separately), and whether the variable is constrained to be non-negative (e.g. moisture mixing ratios).

Once the state vector definition is in hand, it is processed to provide a list of those fields' locations in the COAMPS restart file. If the conversion is from COAMPS to DART, the fields are picked out and packed together. For the conversion from DART to COAMPS, if the variable is to be assimilated it is unpacked and placed at the proper locations within the larger restart file.

Computations like observation operators and covariance localization require information about the model domain. Data like domain size, latitude, longitude, and terrain heights are read directly from the model analysis files for maximum flexibility.

C.2.2 Ensemble Integration

Running COAMPS as an ensemble requires a separate directory for each ensemble member, each with a complete copy of the model data (e.g. restart files and boundary conditions). When control is passed to the ensemble integration scripts from within the assimilation program, the start and stop times are adjusted in the model's namelist based on information in the DART restart file, and the forecast model is run in each member directory. When the integration is complete, the restart file is consolidated into a state vector as in C.2.1 and control is passed back to the assimilation program.

Bibliography

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Anderson, J. L., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642.

Anderson, J. L., 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, **59**, 210–224.

Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.

Bishop, C. H., B. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.

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

Bretherton, C. S., P. N. Blossey, and M. Khairoutdinov, 2005: An energy-balance analysis of deep convective self-aggregation above uniform SST. *J. Atmos. Sci*, **62**, 4273–4292.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*, **75**, 257–288.

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

Davies, H. C., 1976: A lateral boundary formulation for multi-level prediction models. *Quart. J. R. Met. Soc.*, **102**, 405–418.

Descamps, L., and O. Talagrand, 2007: On some aspects of the definition of initial conditions for ensemble prediction. *Mon. Wea. Rev.*, **135**, 3260–3272.

Dirren, S., R. D. Torn, and G. J. Hakim, 2007: A data assimilation case study using a limited-area ensemble Kalman filter. *Mon. Wea. Rev.*, **135**, 1455–1473.

Emanuel, K. A., 1994: *Atmospheric Convection*. Oxford University Press, 580 pp.

Emanuel, K. A., J. D. Neelin, and C. S. Bretherton, 1994: On large-scale circulations in convecting atmospheres. *Quart. J. Roy. Meteor. Soc.*, **120**, 1111–1143.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162.

Evensen, G., 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.

Fein, J. S., and J. P. Kuettner, 1980: Report on the summer MONEX field phase. *Bull. Amer. Met. Soc.*, **61**, 461–474.

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

Gelb, A., Ed., 1974: *Applied Optimal Estimation*. MIT Press, 374 pp.

Gombos, D., and J. A. Hansen, 2007: On potential vorticity regression and its relationship to dynamical piecewise inversion. *Mon. Wea. Rev.*. Submitted.

Hakim, G. J., and R. D. Torn, 2006: Ensemble synoptic analysis. *Fred Sanders Monograph*, American Meteorological Society. Accepted.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.

Hansen, J. A., 2002: Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Wea. Rev.*, **130**, 2373–2391.

Hodur, R. M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.*, **125**, 1414–1430.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan*, **75**, 181–189.

Kain, J. S., 2004: The Kain-Fritsch convective parameterization: An update. *J. App. Meteo.*, **43**, 170–181.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, **82**, 35–45.

Kalman, R. E., and R. C. Bucy, 1961: New results in linear filtering and prediction theory. *Transactions of the ASME–Journal of Basic Engineering*, **83**, 95–107.

Lawson, W. G., and J. A. Hansen, 2004: Implications of stochastic and deterministic filters as an ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Wea. Rev.*, **132**, 1966–1981.

Leidner, S. M., L. Isaksen, and R. N. Hoffman, 2003: Impact of NSCAT winds on tropical cyclones in the ECMWF 4DVAR assimilation system. *Mon. Wea. Rev.*, **131**, 3–26.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20**, 851–875.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791–2808.

Nutter, P., D. Stensrud, and M. Xue, 2004a: Effects of coarsely resolved and temporally interpolated lateral boundary conditions on the dispersion of limited-area ensemble forecasts. *Mon. Wea. Rev.*, **132**, 2358–2377.

Nutter, P., M. Xue, and D. Stensrud, 2004b: Application of lateral boundary condition perturbations to help restore dispersion in limited-area ensemble forecasts. *Mon. Wea. Rev.*, **132**, 2378–2390.

Ramamurthy, M. K., and F. H. Carr, 1987: Four-dimensional data assimilation in the monsoon region. Part I: Experiments with wind data. *Mon. Wea. Rev.*, **115**, 1678–1706.

Ramamurthy, M. K., and F. H. Carr, 1988: Four-dimensional data assimilation in the monsoon region. Part II: Role of temperature and moisture data. *Mon. Wea. Rev.*, **116**, 1986–1913.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for a limited-area ensemble Kalman filter. *Mon. Wea. Rev.*, **134**, 2490–2502.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geo. Res.*, **103**, 14 291–14 324.

Tsuyuki, T., 1996: Variational data assimilation in the tropics using precipitation data. Part II: 3D model. *Mon. Wea. Rev.*, **124**, 2545–2561.

Tsuyuki, T., 1997: Variational data assimilation in the tropics using precipitation data. Part III: Assimilation of SSM/I precipitation rates. *Mon. Wea. Rev.*, **125**, 1447–1464.

Undén, P., 1989: Tropical data assimilation and analysis of divergence. *Mon. Wea. Rev.*, **117**, 2495–2517.

Warner, T. T., R. A. Peterson, and R. E. Treadon, 1997: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599–2617.

Whitaker, J., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Zhang, F., Z. Meng, and A. Aksoy, 2006: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect model experiments. *Mon. Wea. Rev.*, **134**, 722–736.