What Constrains Spread Growth in Forecasts Initialized from Ensemble Kalman Filters?

Thomas M. Hamill and Jeffrey S. Whitaker

NOAA Earth System Research Lab, Physical Sciences Division

Boulder, Colorado

Submitted to *Monthly Weather Review*

a contribution to the special collection for the Buenos Aires workshop

on 4D-Var and Ensemble Kalman Filter Intercomparisons (Herschel Mitchell, editor)

24 August 2010

Corresponding author address:

Dr. Thomas M. Hamill
NOAA ESRL / PSD 1
325 Broadway, Boulder, CO 80305-3337
E-mail: tom.hamill@noaa.gov
Phone: (303) 497-3060
Fax: (303) 497-6449

ABSTRACT

The spread of an ensemble of weather predictions initialized from an ensemble Kalman filter may grow slowly relative to other methods for initializing ensemble predictions, degrading its skill. Several possible causes of the slow spread growth were evaluated in perfect- and imperfect-model experiments with a 2-layer primitive equation spectral model of the atmosphere. The causes examined were covariance localization, the additive noise used to stabilize the assimilation method and parameterize the system error, and model error itself. In these experiments, the flow-independent additive noise was the biggest factor in constraining spread growth. Pre-evolving additive noise perturbations was tested as a way to make the additive noise more flow dependent. This modestly improved the data assimilation and ensemble predictions both in the 2-layer model results and in a brief test of the assimilation of real observations into a global multi-level spectral primitive equation model. More generally, these results suggest that methods for treating model error in ensemble Kalman filters that greatly reduce the flow dependency of the background-error covariances may increase the filter analysis error and decrease the rate of forecast spread growth.

1. Introduction

The ensemble Kalman filter, or "EnKF" (Evensen 1994; Houtekamer and Mitchell 1998) and its variants (e.g., Hamill and Snyder 2000, Anderson 2001; Whitaker and Hamill 2002; Hunt et al. 2006) are being explored for their use in improving the accuracy of initial conditions and for initializing ensemble weather predictions. The EnKF produces an ensemble of parallel short-term forecasts and analyses; background-error covariances from the ensemble are used in the data assimilation step. Introductions to the EnKF are provided in Evensen (2006), Hamill (2006), and Ehrendorfer (2007). The technology behind the EnKF has matured to the point where it is used operationally for atmospheric data assimilation and ensemble predictions (Houtekamer and Mitchell 2005; Houtekamer et al. 2009) or is being tested actively with real data (e.g., Whitaker et al. 2004, 2008, 2009; Houtekamer et al. 2005; Compo et al. 2006; Miyoshi and Yamane 2007; Meng and Zhang 2008ab; Torn and Hakim 2008, 2009; Wang et al. 2008; Szunyogh et al. 2008; Zhang et al. 2009, Aksoy et al. 2009; Buehner et al. 2010ab, Hamill et al. 2010).

The EnKF is now becoming a viable alternative to or complement of other advanced data assimilation schemes such as 4-dimensional variational data assimilation (4D-Var; Le Dimet and Talagrand 1986; Courtier et al. 1994; Rabier et al. 2000). A potential advantage that the EnKF may have for ensemble prediction is that an ensemble of initial conditions is automatically generated that, theoretically at least, have the proper characteristics for initializing ensemble forecasts (Kalnay et al. 2006). In comparison, an additional step is needed to create the ensemble of initial conditions when using the

standard 4D-Var for the data assimilation[1]. Hybridizations of these two methods are possible (Buehner et al. 2010ab).

Only a modest amount of experimentation has been performed on the characteristics of ensemble predictions initialized from EnKFs with real observations. Of particular concern is ensuring that the spread (the standard deviation of ensemble perturbations about the mean) of ensemble forecast perturbations are consistent with the ensemble-mean forecast error; commonly, spread growth is smaller than error growth. The spread growth in forecasts from operational EnKFs is likely to be affected in part by the choice of methods for dealing with the model uncertainty during the ensemble forecasts. There are now a variety of techniques for addressing model uncertainty, such as stochastically perturbed parameterization tendencies (Buizza et al. 1999; Palmer et al. 2009), stochastic backscatter (Shutts 2005, Berner et al. 2009), and the use of multi-model or multi-center ensembles (e.g., Bougeault et al. 2010 and references therein). While the methods for dealing with model uncertainty are certainly relevant for the growth of ensemble forecast spread, here we are interested in *what characteristics of the EnKF alone affect the spread growth*.

Among the research that has been performed on spread growth with EnKFs, Houtekamer et al. (2005) showed that in an earlier implementation of their EnKF, spread actually decreased during the first 12-24 h of the forecast. They attributed this in part to

[1] The European Centre for Medium-Range Weather Forecasts (ECMWF) has experimented recently with an ensemble of 4D-Var analyses assimilating perturbed observations and including stochastic backscatter (Buizza et al. 2008); their method would permit the initialization of an ensemble directly using 4D-Var. However, their method must be performed at reduced resolution to make the computational expense tractable.

the dynamical structure of the noise added to each member used to address "system error;" this noise inflated the prior spread so it was consistent with innovation statistics. In their case, the noise consisted of a sample that was consistent with the 3D-Var background-error statistics but which was not related to the meteorological situation of the day. They suggested also that their use of an overly diffusive forecast model, especially near the model top, unrealistically constrained spread growth. More recently, Charron et al. (2010) reported greater spread growth when the use of the excessively diffusive model was eliminated.

Previously, Mitchell et al. (2002) had also demonstrated that the "covariance localization" applied in the EnKF to mute spurious long-distance covariances in the ensemble estimates (Houtekamer and Mitchell 2001; Hamill et al. 2001) introduced imbalances into the ensemble of analyzed states, which may also constrain spread growth. Subsequently, Lorenc (2003), Buehner and Charron (2007), Bishop and Hodyss (2009ab) and Kepert (2009) have also discussed this effect and have suggested possible algorithmic modifications to remedy this.

Are there other mechanisms that constrain spread growth in EnKFs? In addition to covariance localization and the additive noise, the forecast model may have a very different chaotic attractor (Lorenz 1993) than that of the natural atmosphere, an effect we shall refer to simply as "model error." The data assimilation and short-range forecasts may produce an oscillation of the model state back and forth, toward the observations and the atmosphere's attractor during the update step and back toward the model attractor during the forecast step (Judd and Smith 2001). It is possible that this results in less

projection of the perturbations onto the model's unstable manifold, and hence constrains perturbation growth. Another possibility for the slow spread growth is the nature of effective data assimilation, which adjusts the background more toward the observations in the directions in phase space where background errors are large (and presumably spread growth is large). The analysis process naturally whitens the analysis-error spectrum relative to the background error spectrum (Daley 1991, Fig 5.9; Hamill et al. 2002, Fig. 9), decreasing the projection onto the growing modes.

In this manuscript we seek to understand some of the mechanisms for slow spread growth in ensemble Kalman filters. In particular, we examine the effects of covariance localization, additive error noise, and model error. We perform simulation experiments with a simple, two-level primitive equation model, a model that hopefully is a realistic enough analog to shed light on approaches to be tried in modern-day numerical weather prediction models yet simple enough to permit the generation of a very large number of tests and many cases. To isolate the effect of the data assimilation on spread growth, these experiments are performed with a deterministic forecast model, i.e., none of the stochastic model aspects discussed earlier are included.

The rest of the manuscript is organized as follows. Section 2 provides a brief review of the model and the data assimilation system used. Section 3 shows the results of simulation experiments under perfect-model conditions, and section 4 with an imperfect model. Section 5 provides results from experiments with a global numerical weather prediction model and real observations, and section 6 provides conclusions.

2. **The forecast model, data assimilation system, and experimental design**.

a. Forecast model.

The forecast model used in these experiments was virtually identical to the two-level spectral model of Lee and Held (1993), and a version of it with hemispheric symmetry was used for the ensemble data assimilation experiments in Whitaker and Hamill (2002). No hemispheric symmetry was imposed for these experiments. Here, the data assimilation experiments were run at T31 horizontal resolution, though for imperfect-model data assimilation experiments the nature run was computed at T42 resolution. The prognostic variables of the forecast model are baroclinic and barotropic vorticity, baroclinic divergence, and interface barotropic potential temperature. Barotropic divergence was set to zero, and baroclinic potential temperature was set to 10K. Lower-level winds were mechanically damped with an e-folding timescale of 4 days (4.5 days for T42 nature run). The baroclinic potential temperature was relaxed back to a radiative equilibrium state with a pole-to-equator temperature difference of 80K (74K for T42 nature run) with a timescale of 20 days. The radiative equilibrium profile of Lee and Held (1993; eq. 3) was used. ∇^8 diffusion was applied to all the prognostic variables, with the smallest resolvable scale damped with an e-folding timescale of 3 h. Time integration proceeded with a 4th – order Runge-Kutta scheme with 18 time steps per day (64 for the T42 nature run). The error doubling time of the T31 model was approximately 2.4 days.

This model is obviously much simpler than the operational numerical weather prediction models currently in use; the resolution is lower, there is no terrain, no land and

water, no atmospheric moisture. In fact, while this model is capable of supporting internal gravity waves, it does not produce an external mode. These simplifications should be kept in mind while interpreting the results and their implications for operational numerical weather prediction.

b. Data assimilation methodology.

The ensemble square-root filter (EnSRF) of Whitaker and Hamill (2002) was used for the data assimilation. The EnSRF, like other EnKF algorithms, consists of two steps that are repeated, a set of short-range ensemble forecasts, and a data assimilation that uses the short-range forecasts to estimate the background-error covariances for the ensemble update. Assume the availability of an ensemble of forecasts estimating the state at a particular time when new observations are ready to be assimilated. The EnSRF algorithm separates the EnKF update into an update of the mean and an update of the perturbations around the mean. Observations are assimilated serially, so that the updated mean and perturbations after the assimilation of the first observation are used as the background for the second observation, and so on. Let $\overline{\mathbf{x}}^a$ denote the mean analysis state at the current time, $\overline{\mathbf{x}}^b$ denote the mean background state, $\mathbf{y}^o$ denote the current observation, and $\mathbf{H}$ denote the observation operator that converts the background state to the observation location and type; here this operator is linear. The EnSRF update equations applied to this simplified model and simplified observations are:

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}}^b\right) \quad , \qquad (1)$$

where the Kalman gain $\mathbf{K}$ is

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}\right)^{-1}. \tag{2}$$

Here $\mathbf{R}$ denotes the observation-error variance and $\mathbf{P}^b$ the estimate of the background-error covariance from the ensemble. This covariance matrix was not explicitly calculated, but instead $\mathbf{P}^b\mathbf{H}^T$ was calculated in the EnSRF as a product

$$\mathbf{P}^b\mathbf{H}^T = \frac{1}{n-1}\sum_{i=1}^{n}\rho_{\mathbf{H}} \circ \left[\mathbf{x}_i'^b\left(\mathbf{H}\mathbf{x}_i'^b\right)^T\right] \quad , \tag{3}$$

where $\mathbf{x}_i'^b$ is the ith of n member's deviation from the ensemble mean and $\rho_{\mathbf{H}}$ denotes the Gaspari and Cohn (1999) quasi-Gaussian, compactly supported correlation vector, 1.0 at the observation location and tapering to zero at and beyond a user-specified distance; the $_{\mathbf{H}}$ subscript is intended to remind the reader that this is an "observation-space" localization, and the localization is not directly applied to the ensemble estimate of $\mathbf{P}^b$ (as it should be, ideally; Campbell et al. 2010), but instead to the product $\mathbf{P}^b\mathbf{H}^T$ as an approximation. Similarly, $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$ is constructed without ever explicitly computing $\mathbf{P}^b$:

$$\mathbf{H}\mathbf{P}^b\mathbf{H}^T = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{H}x_i'^b\right)\left(\mathbf{H}x_i'^b\right)^T \quad . \tag{4}$$

Since all of the observations were point observations (see section 2c below), no localization is included in eq. (4) since it would have no effect. Equations (1) – (4) indicate how this implementation of the EnSRF updates the mean state to a new observation. Perturbations around the mean used a slightly different update, following Whitaker and Hamill (2002). Let $\mathbf{x}_i'^a$ denote the updated analysis perturbation for the ith

member around the analyzed mean state. Then the update of the perturbations proceeded according to

$$\mathbf{x}_i'^a = \mathbf{x}_i'^b - \tilde{\mathbf{K}}\mathbf{H}\mathbf{x}_i'^b \quad , \tag{5}$$

where $\tilde{\mathbf{K}}$, the "reduced" Kalman gain, was calculated according to

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

By adding $\mathbf{x}_i'^a$ to $\overline{\mathbf{x}}^a$ for each member, an ensemble of analyzed states are reconstructed, and the full nonlinear forecast model is used to integrate each member forward to the next time when observations are available. This process is then repeated for the duration when observations are assimilated. If desired, at any time the ensemble of analysis states can be integrated forward for a longer period of time to produce an ensemble of weather forecasts.

In some ensemble Kalman filters, particularly deterministic formulations, covariance estimates from the ensemble may be modified to stabilize the system and account for "system errors" such as (1) model error, (2) the underestimation of ensemble spread by using the ensemble information both to calculate the Kalman gain and to update the ensemble (Houtekamer and Mitchell 1998, section 2e and Mitchell and Houtekamer 2009),[2] or (3) the development of inappropriate non-Gaussianity (Lawson

[2] There are other approaches to deal with the systematic mis-estimation of error covariances due to the underestimation of spread with finite ensemble size, including the double EnKF of Houtekamer and Mitchell (1998) or the approach in Hamill and Snyder

and Hansen 2004; Sakov and Oke 2008). Commonly it is assumed that the system error has zero mean and covariance **Q**. If in fact the system error does not have zero mean, this should be corrected beforehand, if possible (Dee 2005, Danforth et al. 2007, Li et al. 2009).

The covariance localization applied in eq. (3) is in fact one form of stabilization that can implicitly account for system error. Two others were tested here. The first was "covariance inflation" (Anderson and Anderson 1999), whereby before the assimilation of the first observation, the deviation of every member of the ensemble around its mean was inflated by a constant:

$$\mathbf{x}_i'^b \leftarrow (1-\alpha)\mathbf{x}_i'^b \quad , \tag{7}$$

where $\leftarrow$ denotes a replacement of the previous state perturbation on the right-hand side. Covariance inflation assumes that the system error is purely an underestimate of ensemble spread; the directions spanned by the background ensemble are appropriate.

Another potential method of stabilization is "additive noise." Either before or after the update, structured noise is added to each ensemble member (e.g., Mitchell and Houtekamer 2000; Houtekamer et al. 2005, Hamill and Whitaker 2005). In perfect-model experiments, for the *i*th member, additive noise ε_i was added to the *i*th background forecast ensemble member:

$$\mathbf{x}_i^b \leftarrow \mathbf{x}_i^b + \varepsilon_i \ . \tag{8}$$

(2000) of updating the *i*th member using a covariance estimated without that *i*th member. Mitchell and Houtekamer (2009) explore these issues in much greater depth.

In the imperfect-model experiments, it was more convenient to add noise to the analysis ensemble (Houtekamer and Mitchell 2005; see discussion on pp. 3284-3285). How imperfect-model samples of additive noise were generated will be explained in the following section.

c. Experimental design.

Two sets of experiments were conducted, perfect-model and imperfect-model experiments. In each experiment the ensemble-mean error, ensemble spread, and ensemble spread growth were examined for a variety of stabilization techniques (localization, additive noise, covariance inflation). Unless mentioned otherwise, the ensemble size was $n = 50$, and the same forecast model dynamics was used for each member; the model incorporated no stochastic physics, nor did it use multiple models. In all experiments, the ensemble was initialized with a random draw from the forecast model climatology.

In both the perfect and imperfect-model experiments, an observation network with 490 nearly equally spaced observation locations was used. The observations were located at the nodes of a spherical geodesic grid, approximately 2000 km apart. At each location, observations were created for the barotropic potential temperature and the u- and v- wind components at the two model levels, 250 and 750 hPa. Observations were created by interpolating the true state to the observation location and adding random, independent, normally distributed observation errors. Errors had zero mean and variances of 1 K^2 and 1.0 m^2s^{-2} for potential temperature and winds, respectively. The

nature run for generating the true state was produced by starting the forecast model from a random perturbation superimposed on a resting state, integrating for 500 days and discarding the first 200 days. Observational data was assimilated over the 300 days, with an update to new observations every 12 h. In the computation of assimilation statistics, the first 25 days of data assimilation were discarded due to transient effects, leaving 275 days to calculate statistics.

A time series of the root-mean square (RMS) errors, spread, and spread growth were calculated using the total-energy norm:

$$\|\bullet\| = \left[\frac{1}{2} \left(\frac{\iint\limits_{A,L} \left(u^2 + v^2\right) dA\, dL}{\iint\limits_{A,L} dA\, dL} + \frac{\int\limits_{A} \frac{c_p}{\theta_{ref}} \theta^2\, dA}{\int\limits_{A} dA} \right) \right]^{1/2}, \tag{9}$$

where c_p is the specific heat capacity of dry air at constant pressure (1004 J kg^{-1} K^{-1}), θ_{ref} = 300 K, and the integrals were performed over the earth's surface area (A) and, in the case of winds, over the two model levels (L). The error, spread, and spread growth statistics presented in this paper represent the average of the 12-hourly samples over the 275 days.

Different types of additive noise were used for the perfect- and imperfect-model simulation experiments. In the perfect-model experiments, the additive noise was generated to (a) have a structure that was consistent with the assumed system error covariance structure, and (b) to ensure a consistency in the innovation statistics. Regarding (b), it can be readily shown that a calibrated ensemble should have the following approximate equivalence of expected value (Houtekamer et al. 2005):

$$diag\left(\left[\mathbf{y}-\mathbf{H}\overline{\mathbf{x}}^{b}\right]\left[\mathbf{y}-\mathbf{H}\overline{\mathbf{x}}^{b}\right]^{\mathrm{T}}\right)\cong diag\left(\mathbf{R}\right)+diag\left(\frac{1}{n-1}\sum_{i=1}^{n}\left[\mathbf{H}\mathbf{x}_{i}'^{b}\right]\left[\mathbf{H}\mathbf{x}_{i}'^{b}\right]^{\mathrm{T}}\right), \quad (10)$$

where $diag\,(\bullet)$ denotes the sum of the diagonal elements. The question then, is how to increase the background spread if the right-hand side of eq. (10) is smaller than the left-hand side. Regarding (a), the procedure we used assumed that the system error could be reasonably estimated using scaled differences between model forecast states that differed in time by a small amount (Hamill and Whitaker 2005). Such perfect-model additive noise typically exhibits larger amplitudes in the mid-latitude storm tracks and much smaller amplitudes elsewhere (not shown).

The overall perfect-model additive noise procedure was thus as follows: a time series of differences between the model nature run at time t and time t + 24 h was calculated from the truth run. At any particular assimilation time, 50 random samples of these differences were chosen from the time series, without replacement. The mean state of the 50 samples was computed and subtracted from the 50 random samples. Denote the ith noise sample as $\mathbf{x}_{i}'^{n}$. These samples were then scaled by a constant β and added to the ensemble of *background forecast states*, i.e., in eq. (8), $\varepsilon_i = \beta\mathbf{x}_{i}'^{n}$. The magnitude of β was chosen so that

$$\begin{aligned} &diag\left(\left[\mathbf{y}-\mathbf{H}\overline{\mathbf{x}}^{b}\right]\left[\mathbf{y}-\mathbf{H}\overline{\mathbf{x}}^{b}\right]^{\mathrm{T}}\right)= \\ &\quad diag\left(\mathbf{R}\right)+diag\left(\frac{1}{n-1}\sum_{i=1}^{n}\left[\mathbf{H}\left(\mathbf{x}_{i}'^{b}+\beta\mathbf{x}_{i}'^{n}\right)\right]\left[\mathbf{H}\left(\mathbf{x}_{i}'^{b}+\beta\mathbf{x}_{i}'^{n}\right)\right]^{\mathrm{T}}\right). \end{aligned} \quad (11)$$

When the right-hand side of eq. (10) was larger than the left before the addition of any noise, β was set to 0.0.

The imperfect-model nature run used a higher-resolution model (T42), a different pole-to-equator temperature difference (74K), and a different mechanical damping time scale (4.5 days) than the forecast model. Hence, one would expect systematic differences between the forecast model's and the nature run's climatologies beyond the storm track. Here, we did not try to correct for potential systematic biases (Dee 2005, Danforth et al. 2007, Li et al. 2009) but instead we assumed that the system error covariance **Q** could be estimated by drawing random samples and then by adding them to the *analyzed states*, i.e., $\mathbf{x}_i^a \leftarrow \mathbf{x}_i^a + \beta \mathbf{x}_i'^n$. Implicitly what is assumed is that a reasonable estimate $\hat{\mathbf{Q}}$ of the system-error covariance (not including localization) is

$$\hat{\mathbf{Q}} = \left[\mathbf{M}\left(\beta \mathbf{x}_i'^n\right)\right]\left[\mathbf{M}\left(\beta \mathbf{x}_i'^n\right)\right]^T \quad , \qquad (12)$$

where **M** denotes the linear tangent of the model forward operator. In practice, eq. (12) is never actually computed; the perturbations are simply added to the ensemble of analysis states, these are integrated forward with the fully nonlinear model, and the (observation-space localized) perturbations of those states about the mean are used to estimate the Kalman gain, eqs. (2) – (4).

The challenge with such a system-error method is to generate $\beta \mathbf{x}_i'^n$ so that the later background forecasts realistically sample these possible differences in forecast and true model states. For this imperfect-model scenario, we assumed some knowledge of the system error was reasonable, that while the true pole-to-equator temperature

difference and damping time scales were not known, at least it was known that these two model parameters were sources of uncertainty. Accordingly, to produce such additive noise samples, multiple T31 nature runs were created, each using a different pole-to-equator temperature difference and different damping time scale. Pole-to-equator temperature differences ranged from 74 to 83 K, and damping timescales ranged from 3 to 5 days. Figure 1 shows the zonal-mean profiles of the upper- and lower-level u-wind component, and interface potential temperature for the forecast model nature run, the T42 nature run, and the set of perturbed T31 nature runs. 400 random model states were extracted from the set of perturbed T31 nature runs. The 50 samples of additive noise at any particular update time during a data assimilation experiment were drawn randomly from the 400 perturbed states, without replacement. The mean state of these 50 members was then calculated and subtracted from each to create 50 perturbed states from the various model climatologies. These samples of additive noise were typically scaled down, the magnitude of the scaling β specified in the experiment, and then added to the ensembles of analyses rather than the background (for more rationale on adding noise to the analyses, see Houtekamer and Mitchell 2005; especially discussion on pp. 3284-3285).

3. **Perfect-model experiment results**.

Figure 2 shows error and spread for perfect-model experiments. Here, multiple parallel cycles of the EnSRF were conducted, varying the covariance localization across a range of length scales and stabilizing the data assimilation either with 2% covariance inflation or adaptive additive noise. The magnitude of the additive noise was determined

adaptively each update step using the procedure described in section 2c, and varied moderately with time. For example, the scaling factor β varied between 0.0 and 0.034 for the simulation with a localization length scale of 10,000 km; at smaller localization radii, zeros became more common, and at larger radii positive numbers were more common, with decreasing variation from one update cycle to the next. Examining Fig. 2, several characteristics of the spread and error were notable. First, errors were strongly affected by the covariance localization length scale. Very small length scales produced analyses with larger errors, and analysis-error minima were found with localization length scales of approximately 7,000 km for the additive-noise and 17,000 km for the covariance inflation. Similar effects of localization on error were previously demonstrated in Houtekamer and Mitchell (1998), Houtekamer and Mitchell (2001), and Hamill et al. (2001).

Overall, the covariance inflation simulations had much less error and a greater consistency between spread and error than the adaptive additive error simulations. This raises two questions: first, why did the adaptive additive error simulations have more spread than error at small localization radii? And why did they have larger analysis errors? As for why there was an inconsistency in spread, a likely reason for this was that the amount of adaptive additive error was chosen to ensure a consistency between spread and error *at the observation locations*. However, the spread and error shown in Fig. 2 were calculated globally, at points both near to and relatively far from the observations. As the localization length scale was shortened, the potential corrective effect of an observation was confined to a smaller and smaller neighborhood of the observation, while away from it the analysis continued to reflect the influence of the prior and

preserved the prior's spread. With a broader localization, the observation had a larger effect on the analysis and produced a larger reduction in the analysis spread further away from the observations.

Why did the adaptive additive error simulations have more error? Figure 3 shows the growth rate of spread during the 12-h between updates to the observations. Spread generally grew more slowly in the adaptive additive error simulations; the adaptive noise was not dynamically conditioned to the flow of the day, while covariance inflation preserved the flow-dependent structures. This most likely explains the larger error; the error covariance model was less accurate in describing the situationally dependent background errors, reducing the efficacy of the data assimilation. The exceptions to the higher errors with additive noise were at the smallest localization length scales. For these parameter values, the additive-noise EnSRF simulations were stabilized purely by the covariance localization, and the adaptive additive noise typically consisted of no noise at all. Consequently, the errors and growth rates were more similar to those from the covariance-inflation simulations.

Other characteristics of spread growth are also evident in Fig. 3. At large localization length scales, the rate of growth of spread for the covariance inflation simulations was approximately equal to that from a 400-member simulation that was stabilized by 1% inflation but which utilized no localization whatsoever. Based on this, the assumption is that the growth rate of ~1.22 was taken to represent an approximate upper-limit of the possible spread growth rate in this model with an EnSRF. The use of a short localization radius did decrease the rate of growth of spread somewhat,

approximately 2 percent for the covariance inflation simulations relative to the large localization radii. Spread growth was smaller for the adaptive additive error simulations, but spread growth did not increase with increasing localization radius, as did the covariance inflation. The greater growth rate from less localization was counteracted by the slower growth from the application of greater amounts of flow-independent additive noise to stabilize the filter.

Overall, in this model covariance localization had only a small effect on the rate of growth of spread, decreasing growth by a few percent; introducing non-dynamically conditioned additive noise had a larger effect. However, the magnitudes of these changes in growth rate may have been an artifact of this simple model and may not be fully realistic of what may occur in real numerical weather prediction models. Such models may support additional unbalanced modes (e.g., external gravity waves) and may both generate noise and organize it much more readily due to the presence of moist convection (Zhang et al. 2003).

4. **Imperfect-model results**.

Figure 4 provides the average RMS error, spread, and spread-growth statistics for experiments with a variety of combinations of globally constant covariance inflation magnitude and covariance localization length scale[3]. Errors were much higher than the perfect-model results, and the minimum error occurred with much more inflation (50%) and at a much narrower localization radius (3000 km). However, the globally averaged

[3] Anderson (2009) discusses spatially and temporally varying covariance inflation; such methods were not tried here.

analysis spread was smaller than the RMS error for this length scale/inflation. Errors increased dramatically as the covariance inflation amount was lowered, while spread decreased and spread growth increased; this was a sign that filter divergence was occurring. There were some combinations of large inflation and narrow localization radii where the filter was numerically unstable; here, model-state perturbations away from the observations occasionally experienced an uncontrolled growth of spread that led to numerical instabilities.

Aside from when very small inflation was applied, spread growth was decreased significantly relative to the maximum determined in perfect-model experiments. In fact, spread growth was nearly nonexistent on average during the 12 h for the length scale and inflation that produced the minimum of error. Figure 5 illustrates the challenges of tuning a globally constant covariance inflation to produce spread consistent with error at all locations. Here, zonal and time-averaged spread, RMS error, and bias (ensemble-mean forecast minus truth) are plotted for the minimum-error inflation rate/length scale, the dot in Fig. 4. Spread was generally smaller than error, but was greater than error for tropical temperatures. When spread was further increased, temperature and low-level wind errors increased in the tropics (not shown), indicating that the drastic inflation was degrading the correlation structures in the ensemble.

Unlike the perfect-model experiments, the imperfect-model experiments produced slightly smaller analysis errors when the EnSRF was stabilized with additive noise instead of covariance inflation (Fig. 6). When more additive noise was applied to stabilize the EnSRF, the analysis spread increased, as expected. Unexpectedly, the

spread growth rate varied only slightly no matter how much additive noise was applied. Relative to the perfect-model experiments where spread growth rate could exceed 1.2 per 12 h, the spread growth for these imperfect model experiments was lower, with a maximum of approximately 1.13. This magnitude varied little with the localization length scale and the amount of additive noise. Figure 7 shows that there was a greater consistency between wind spreads and errors across latitudes and variables than for the covariance inflation in Fig. 5, though temperature spread in the mid-latitude storm track was too large. There was also a substantial warm bias in the tropics and cold bias slightly poleward of the storm track.

With the results presented thus far, it is difficult to determine whether the decrease in spread growth relative to perfect-model experiments can be attributed primarily to the additive noise or to the effects of model error. To better understand the potential effects of additive noise, an additive noise perturbation was added to a nature run from the T31 forecast model, and the size of the perturbation was calculated in the energy norm as the control and perturbed forecasts were integrated to 4 days lead. This process was repeated over 23 different case days, equally spaced every 12.5 days during the nature run. The zonal and sample-average growth of perturbation magnitude is shown in Fig. 8. It takes approximately 12 h for the spread to increase by a factor of 1.1, consistent with the spread growth for the 12-h data assimilation cycle shown in Fig 7. Between 24 and 48 h, latitudinally averaged spread grew from ~1.3 to ~2.0, with even greater growth between 48 and 72 h. This is consistent with the concept of a randomly oriented, small perturbation projecting more and more on the leading Lyapunov vectors

as the control and perturbed are integrated forward (Toth and Kalnay 1993, Vannitsem and Nicolis 1997, Snyder and Hamill 2003).

Perhaps a deficiency of the additive noise perturbations that were used in the imperfect-model experiments was that they were not dynamically conditioned, i.e., they had no relevance to the "flow of the day." Palmer (2002) has previously argued that the component of model error that is most important is the component that projects onto the growing forecast structures. Following a similar rationale, perhaps additive noise that was both consistent with model error statistics and dynamically conditioned to project to a greater extent onto the leading Lyapunov vectors of the system would produce analyses with less error and greater spread growth. To test this, another additive noise experiment was performed with the imperfect model. In this experiment, instead of adding the noise samples directly at the time of the update, a slightly modified process was followed. First, the ensemble-mean analysis from 24 h prior was extracted. Additive noise perturbations were applied to the ensemble mean analysis, and 24-h forecasts were conducted. The ensemble-mean forecast was subtracted to yield a set of evolved perturbations. After a latitudinally dependent rescaling so their magnitude was consistent with that of the original additive noise perturbations, these evolved perturbations were used as the additive noise in the data assimilation.

Figure 9 shows the error, spread, and spread growth using these evolved perturbations. The minimum-error analysis was now ~2.08, compared to the ~2.16 previously in Fig. 6, an ~3.7% decrease in error. The new minimum error now occured at a slightly larger additive noise amount, 15%. Most notably, as Fig. 9c shows, the spread

growth in the subsequent forecast increased. For the parameter combination with the minimum error, spread growth was ~1.21 / 12-h cycle, which was actually larger than the spread growth for the perfect-model experiments at the same localization radius. This suggests that for this model and experimental design, the structure of the additive errors and not model error was the primary cause of a deficiency of spread growth; if model error were the underlying cause, it would not have been ameliorated by changes in the type of additive noise. Figure 10 illustrates, however, that the evolved additive noise did increase temperature bias somewhat, though wind errors were slightly lower. Why would the evolved additive noise increase temperature bias? Figure 11 shows a map of ensemble temperature correlations with the ensemble temperatures at a point in the tropics, averaged over 23 cases. The original additive noise that was introduced tended to have a zonally symmetric structure and thus exhibited large zonal correlations relative to the evolved additive noise. Consequently, temperature observations in the original additive simulations produced larger and more spatially extended analysis increments, even after the application of covariance localization, somewhat more effectively correcting the bias over wider swaths in the tropics. Thus, while evolved additive noise produced more rapidly growing, flow-dependent structures (most evident in the mid-latitudes), it had the unintended consequence of decreasing the realistic zonal correlations of temperature errors and increasing bias. In a global norm, however, the evolved additive noise reduced error and increased spread growth.

Did the evolved additive noise have a positive impact on longer-lead forecast error and spread? Figure 12 shows that it may have had a modest beneficial effect. Two sets of 50-member ensemble forecast spread and error curves are shown, averaged over

23 case days, each case day separated by 12.5 days between samples to ensure independence of the errors between samples. The first set was for the data assimilation experiments shown in Fig. 7, with a 4000-km localization and 12% additive noise; call this "Add4000-12." The second used the evolved perturbations and the same 4000-km localization and 12% additive noise; call this "Evo4000-12." To test the statistical significance of changes in error, a 1000-sample paired block bootstrap was performed for each forecast lead using the 23 daily global RMSE and spread statistics, following Hamill (1999). The 5th and 95th percentiles of the resampled distribution are plotted atop the Evo4000-12 RMSE and spread. As shown in Fig 12, there was a small positive impact on spread growth for Evo4000-12 relative to Add4000-12; the two started with nearly equal spread, but spread growth was faster during the first 1.5 days of the forecast, and that extra spread was mostly preserved through 10-days lead. Additionally, at the longer leads the forecast error was slightly reduced, perhaps because the larger spread results in a more effective averaging of the ensemble. This difference was generally not statistically significant. We note that even after the use of evolved error covariances, there was still a large inconsistency between spread and error, indicating the presence of bias and the potential importance of treating model-related uncertainty in actual ensemble prediction systems.

5. **Experiments with a T62 global forecast model**.

To determine whether the increased spread growth with evolved additive noise may occur with realistic weather prediction systems, experiments were also conducted using the EnSRF with a T62, 28-level version of National Centers for Environmental

Prediction (NCEP) Global Forecast System (GFS) model. Further details on the model and the data assimilation methodology were provided in Whitaker et al. (2008), with the following recent changes to the algorithm. An adaptive radiance bias correction algorithm developed by T. Miyoshi (personal communication, 2009) was included, which allows satellite radiances to be assimilated. The algorithm mimics what is done in the NCEP Grid-point statistical interpolation (GSI) variational system (Wu et al. 2002) and uses the same air-mass predictors in the bias calculation. Additionally, the fast parallel algorithm of Anderson and Collins (2007) was used to calculate the EnSRF increment.

The EnSRF assimilations were started on 0000 UTC 1 December 2007 and ended on 0000 UTC 10 January 2008. As in Whitaker et al. (2008), the NCEP GSI system was used for the forward operator calculation, and all conventional, satellite wind, and global positioning system radio-occultation data are assimilated, as well as Advanced Microwave Sounding Unit (AMSU) and High-Resolution Infrared Radiation Sounder (HIRS) radiances and Solar Backscatter Ultra-Violet Instrument (SBUV) ozone retrievals. Covariance localization using the Gaspari and Cohn (1999) compactly supported quasi-Gaussian function that tapered to zero at 1500 km in the horizontal and 1.1 scale heights (in units of –ln(pressure)) was employed. Updates occurred every 6 h.

The EnSRF was run in two parallel cycles, the first employing scaled additive noise generated with 48-h minus 24-h forecast differences (the "NMC method;" Parrish and Derber 1992), and the second using the same scaled additive noise pre-evolved over the prior 24-h period. The scaling in both was a globally constant 0.5. Ten-member ensemble forecasts were conducted from each once daily from 0000 UTC initial

conditions and integrated to 7 days lead for every day between 10 December 2007 and 10 January 2008. As with the two-level model, no methods of dealing with model uncertainty such as stochastic physics were utilized.

Figure 13 provides the results. The evolved additive noise started with slightly higher spread, and that spread grew much faster during the first 24 hours of the forecast, so that at all subsequent leads the spread was significantly larger with the evolved additive noise. The ensemble-mean error was decreased slightly at the longest leads. There was a notable inconsistency between spread and error at the longest leads, due presumably to the low model resolution, the strong model diffusion at this resolution, and the lack of any treatment of model error in this ensemble prediction system. Nonetheless, the T62 GFS results suggest that the evolved additive noise will have a modest beneficial impact on spread growth during the early hours of the forecast, and it may provide some decrease in ensemble-mean error, especially at the longest leads.

6. **Conclusions**

While the EnKF has been demonstrated to be an advanced data assimilation method that can produce initial conditions with errors that are competitive with variational methods (Whitaker et al. 2008; Buehner et al. 2010ab), to date little experimentation has been performed on the characteristics of forecasts. The only center that currently runs the EnKF operationally, Environment Canada, has previously been concerned with the relatively slow growth of spread from their ensemble of initial conditions. This study attempted to determine whether the covariance localization, additive noise, or model error played the lead role in limiting spread growth from

ensemble Kalman filters. The model chosen for these experiments was a T31, 2-level dry primitive equation global model. A uniform network of wind and temperature observations were assimilated using an ensemble square-root filter (EnSRF). This model is much simpler and the observations are more sparse than are used in operational weather prediction. Still, this simplicity permitted a wide range of experiments to be conducted, and the model had some of the essential characteristics of more complex models, such as the ability to support internal gravity wave activity as well as baroclinic modes.

In perfect-model experiments, covariance localization was found to have a relatively modest effect on the growth of forecast spread. In experiments where the EnSRF was also stabilized with covariance inflation, the localization reduced the growth of spread in a global energy norm from approximately 1.21 / 12-h cycle with long localization scales to 1.19 when using very short localization length scales. In comparison, at the longest localization length scale, changing from stabilizing the filter with covariance inflation to stabilizing it with additive noise reduced the spread growth from 1.21 to ~1.11. The use of additive noise also increased the ensemble-mean analysis error substantially in the perfect-model experiment, with an error of ~0.29 m/s vs. ~0.15 for covariance inflation.

Next, a set of imperfect-model experiments was conducted using a T42 nature run with a different pole-to-equator temperature gradient and different mechanical damping time scale. Globally constant covariance inflation proved less useful for stabilizing perfect-model simulations, as previously discussed in Hamill and Whitaker (2005).

Spread growth was also much smaller than in perfect-model experiments under stabilization by covariance inflation. Additive noise successfully stabilized the EnKF, but spread growth was also much smaller, approximately 1.13 / 12-h cycle.

An examination of the characteristics of additive-noise perturbations showed that they typically grew very slowly during the first few hours of the forecast, but thereafter much more rapidly. This suggested a possible improved ad-hoc procedure: instead of adding random additive noise samples, back up some period of time (in our study, 24 h), add the noise to an earlier ensemble mean, evolve the forecasts forward in time to the current update time, and use the rescaled, re-centered perturbations with their dynamically conditioned additive noise. When this was done, this resulted in a modest (3.7%) decrease in analysis error, it reduced the forecast error slightly, it increased forecast spread growth during the first day or two of the forecast, and consequently it resulted in a modest improvement in spread-error consistency at longer leads. A disadvantage was an increase in tropical temperature bias. L. Magnusson and M. Leutbecher (personal communication, 2009) are also exploring the use of evolved additive noise for initializing ensemble predictions.

Results with a T62 version of the NCEP GFS provided confirmatory evidence that evolved additive noise could improve the rate of spread growth in the early hours of the ensemble forecasts, and possibly provide some reduction in ensemble-mean error, especially at the longest leads.

The application of evolved additive noise may appear at first glance somewhat impractical for operational numerical weather prediction, for costs of the EnSRF go up

significantly, as evolving the additive noise increases the effective number of ensemble members that must be integrated forward in time during each data assimilation cycle. In higher-resolution, operational models, evolving the ensemble forward in time is the predominant computational expense. However, perhaps the additive noise could be evolved with a lower-resolution version of the forecast model, reducing its computational expense.

We propose two possible theoretical justifications for flow-dependent additive noise samples. First, assume that system error is introduced at a constant rate during the x hours between EnKF updates. The system error introduced during the first hours will have a component that will project onto the system's leading Lyapunov vectors, and that part will grow like any other perturbation and thus be better represented by short-term evolved additive noise. Second, perhaps the actual system error was related to, say, an inappropriate estimate of mountain drag (Klinker and Sardeshmukh 1992). The standard additive noise perturbation may introduce noise over topography, regardless of whether there was strong flow in the region. Evolved additive noise will at least be more likely to decrease the amplitude of perturbations when the flow is weak and increase it when the flow is stronger. To the extent that model error is larger when and where the dynamics are more active, evolved additive noise should provide some improvement.

Abstracting more generally, methods used to stabilize EnKFs to treat sampling and model error should be designed in a way that they do not dramatically reduce the flow dependency of the covariance estimate, for this is the very property of the EnKF that has led to its widespread usage.

ACKNOWLEDGMENTS

This work was stimulated by discussions during the November 2008 World Meteorological Organization Buenos Aires workshop on 4D-Var and EnKF inter-comparisons, and in conversations thereafter. In particular, the authors would like to thank Chris Snyder and Jeff Anderson (NCAR), Mark Buehner and Herschel Mitchell (Environment Canada), Jeff Kepert (Bureau of Meteorology, Australia), and Ron Gelaro (NASA).

References

Aksoy, A., D. C. Dowell, C. Snyder, 2009: A multi-case comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805-1824.

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.

----------, 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev*., **129**, 2884-2903.

----------, and Collins, N., 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Ocean. Tech*., **24**, 1452-1463.

---------- , 2009: Spatially and temporally varying adaptive covariance inflation for enemble filters. *Tellus*, **61**, 72-83.

Berner J., G. Shutts, M. Leutbecher, and T.N. Palmer, 2009: A Spectral stochastic Kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci*., **66**, 603-626.

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

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

Bougeault, P., Toth, Z., and others, 2010: The THORPEX Interactive Grand Global Ensemble (TIGGE). *Bull. Amer. Meteor. Soc*., in press. Available from philippe.bougeault@meteo.fr

Buehner, M., and M. Charron, 2007: Spectral and spatial localization of background-error correlations for data assimilation. *Quart. J. Royal. Meteor. Soc*., **133**, 615-630.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: description and single-observation experiments. *Mon. Wea. Rev*., **138**, 1550-1566.

----------, ----------, ----------, ----------, and ----------, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: 1-month experiments with real observations. *Mon. Wea. Rev*., **138**, 1567-1586.

Buizza, R., M. Miller, and T. N. Palmer (1999), Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System, *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.

Buizza, R., M. Leutbecher, and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quart. J. Royal Meteor. Soc.,* **134**, 2051-2066.

Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Wea. Rev*., **138**, 282-290.

Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Towards random sampling of model error in the Canadian ensemble prediction system. *Mon. Wea. Rev*., **138**, 1877-1901.

Compo, G.P., J.S. Whitaker, and P.D. Sardeshmukh, 2006: Feasibility of a 100 year reanalysis using only surface pressure data. *Bull. Amer. Met. Soc.*, **87**, 175-190.

Courtier, P., J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Royal Meteor. Soc*., **120**, 1367-1387.

Danforth, C. M., E. Kalnay, and T. Miyoshi, 2007: Estimating and correcting global weather model error. *Mon. Wea. Rev*., **135**, 281-299.

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

Dee, D. P., 2005: Bias and data assimilation. *Quart. J. Royal Meteor. Soc*., **131**, 3323-3343.

Ehrendorfer, M., 2007: A review of issues in ensemble Kalman filtering. *Meteor. Zeitschrift*, **16**, 795-818.

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

Evensen, G., 2006: *Data Assimilation, the Ensemble Kalman Filter*. Springer Press, 285 pp.

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

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155-167.

----------, and C. Snyder, 2000: A hybrid ensemble Kalman filter / 3d-variational analysis scheme. *Mon. Wea. Rev*., **128**, 2905-2919.

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

----------, C. Snyder, and R. Morss, 2002: Analysis-error statistics of a quasigeostrophic model using three-dimensional variational assimilation. *Mon. Wea. Rev.*, **130**, 2777-2790.

----------, and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Mon. Wea. Rev.*, **133**, 3132-3147.

----------, 2006: Ensemble-based atmospheric data assimilation. Chapter 6 in *Predictability of Weather and Climate*, T. N. Palmer and R. Hagedorn, Eds. Cambridge Press, 124-156.

---------- , J. S. Whitaker, M Fiorino, and S. J. Benjamin, 2010: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.*, in review. Available at http://tinyurl.com/2d56uuq .

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

----------, ----------, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

----------, ----------, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. *Mon. Wea. Rev.*, **133**, 604-620.

----------, ----------, 2005: Ensemble Kalman filtering. *Quart. J. Roy.*

Meteor. Soc., **131**, 3269–3289.

----------, ----------, and X. Deng, 2009: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2126-2143.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2006: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112-126. http://dx.doi.org/10.1016/j.physd.2006.11.008

Judd, K., and L. Smith, 2001: Indistinguishable states II: the imperfect model scenario. *Physica D*, **196**, 224-242.

Kalnay, E., B. Hunt, E. Ott, and I. Szunyogh, 2006: Ensemble forecasting and data assimilation: two problems with the same solution? Chapter 7 of *Predictability of Weather and Climate*, T. N. Palmer and R. Hagedorn, Eds., Cambridge Press, 157-180.

Kepert, J. D., 2009: Covariance localization and balance in an ensemble Kalman filter. *Quart. J. Royal. Meteor. Soc.*, **135**, 1157-1176.

Klinker, E., and P. D. Sardeshmukh, 1992: The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, **49**, 608-627.

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

Le-Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations. Theoretical aspects. *Tellus*, **38A**, 97-110.

Lee, S., and I. M. Held, 1993: Baroclinic wave packets in models and observations. *J. Atmos. Sci.*, **50**, 1413-1428.

Li, H., E. Kalnay, T. Miyoshi, C. M. Danforth, 2009. Accounting for model errors in ensemble data assimilation. *Mon. Wea. Rev*., **137**, 3407-3419.

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

Lorenz, E. N., 1993: *The Essence of Chao*s. University of Washington Press, 227 pp.

Meng, Z., and F. Zhang, 2008a: Test of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVar in a real-data case study. *Mon. Wea. Rev.*, **136**, 522-540.

----------, and ----------, 2008b: Test of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVar in a month-long experiment. *Mon. Wea. Rev.*, **136**, 3671-3682.

Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.,* **128**, 416-433.

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

----------, and ----------, 2009: Ensemble Kalman filter configurations and their performance with the logistic map. *Mon. Wea. Rev*., **137**, 4325-4343.

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

Palmer, T. N., 2002: Predicting uncertainty in numerical forecasts. Chapter 1 of *Meteorology at the Millennium, Volume 83, International Geophysics*. Academic Press, 450 pp.

Palmer, T.N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G.J. Shutts, M. Steinheimer and A Weisheimer, 2009: Stochastic parametrization and model uncertainty. *ECMWF Tech Memo 598*, 42 pp. Available at http://www.ecmwf.int/publications/library/do/references/list/14

Parrish D., and J. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.

Rabier, F., H. Jarvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation: I: experimental results with simplified physics. *Quart. J. Royal Meteor. Soc*., **126**, 1143-1170.

Sakov, P., and P. R. Oke, 2008: Implications of the form of the ensemble transformation in the ensemble square-root filters. *Mon. Wea. Rev*., **136**, 1042-1053.

Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Royal Meteor. Soc*., **132**, 3079-3102.

Snyder, C., and T. M. Hamill, 2003: Leading Lyapunov vectors of a turbulent baroclinic jet in a quasigeostrophic model. *J. Atmos. Sci*., **60**, 683-688.

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

Torn, R. D., and G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Wea.Rev.,* **136,** 3947–3963.

----------, and ----------, 2009: Ensemble data assimilation applied to RAINEX observations of hurricane Katrina (2005). *Mon. Wea. Rev*., **137**, 2817-2829.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Amer. Meteor. Soc*., **74**, 2317-2330.

Vannitsem, S. and C. Nicolis, 1997: Lyapunov vectors and error growth patterns in a T21L3 quasigeostrophic model. *J. Atmos. Sci*., **54**, 347-361.

Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: real observation experiments. *Mon. Wea. Rev.*, **13**6, 5116-5131.

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

----------, G.P. Compo, X. Wei and and T.M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190-1200.

----------, T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. R*ev., **136**, 463–482.

----------, G. P. Compo, and J.-N. Thépaut, 2009: A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Mon. Wea. Rev*., **137**, 1991-1999.

Wu, W.S., R.J. Purser, and D.F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci*., **60**, 1173-1185.

----------, Y. Weng, Z. Meng, J. A. Sippel and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman Filter: Humberto (2007). *Mon. Wea. Rev*., **137**, 2105-2125.

LIST OF FIGURE CAPTIONS

Figure 1: Zonal-mean averages for (a) upper-level *u*-wind component, (b) lower-level *u*-wind component, and (c) interface barotropic potential temperature in nature runs. Red line indicates the T31 model used for data assimilation experiments, dashed blue line the T42 model nature run used in imperfect-model experiments, and thin solid lines the various T31 models run with perturbed pole-to-equator temperature difference and perturbed mechanical damping time scales.

Figure 2: Spread and error for perfect-model experiments when ensemble is stabilized by covariance inflation (black curves) and adaptive additive noise (grey curves).

Figure 3: Growth of ensemble forecast spread during the 12-h between update steps for perfect-model experiments.

Figure 4. (a) Ensemble-mean RMS error in the energy norm for imperfect-model experiments with the EnSRF when filter is stabilized by covariance inflation. Data is plotted as a function of the covariance localization radius (x axis) and the amount of covariance inflation (y axis). The black dot indicates the (localization radius, inflation amount) pair used in Fig. 5. (b) As in (a), but for ensemble spread in the energy norm. (c) As in (a), but for the ensemble spread growth rate over the 12-h period between data assimilation updates.

Figure 5. (a) Zonal and time average of the upper-level *u*-wind component analysis RMS error, spread, and bias for the 3000-km localization and 50% inflation rate data point shown in Fig. 4. (b) As in (a), but for lower-level *u*-wind component. (c) As in (a), but for interface potential temperature.

Figure 6: As in Fig. 4, but for imperfect-model experiments stabilized by additive noise. Dot indicates the localization radius and additive noise amount where error was approximately at a minimum and spread was approximately consistent with error. Spread and error for this combination is examined more in Fig. 7.

Figure 7: As in Fig. 5, but for the 4000-km localization and 12% additive noise experiment data shown in Fig. 6.

Figure 8: Growth of small model-error additive noise perturbations around a state sampled from the T31 forecast model nature run. Dashed line indicates the relative proportion of total energy in the perturbations at the initial time. Perturbations are scaled so that their average magnitude at each latitude is 1.0 at the initial time.

Figure 9. As in Fig. 6, but where evolved additive noise is used to stabilize the filter. Black dot indicates the localization radius/additive-noise amount that was approximately a minimum in Fig. 7.

Figure 10: As in Fig. 5, but for the 4000-km localization and 12% evolved additive noise experiment data shown in Fig. 9.

Figure 11: Map of multi-case average background correlations without covariance localization (colors and dashed lines) between the ensemble interface potential temperature at the location with the dot and the ensemble temperature at every other grid point. The average background potential temperature is also plotted (solid lines, every 10K). Panel (a) is for the additive noise, panel (b) for the evolved additive noise.

Figure 12: Ensemble forecast spread and RMS error in the energy norm, initialized from imperfect-model additive-noise forecasts with 4000 km localization and 12% additive noise scaling (Add4000-12, black lines); from evolved additive noise with 4000 km localization and 12% additive noise scaling (Evo4000-12, grey lines); Error bars represent the 5th and 95th percentiles from a paired block bootstrap.

Figure 13: Ensemble forecast spread of mean-sea level pressure (dashed lines) and error (solid lines) from T62 GFS experiments. Grey color denotes evolved additive noise and black denotes conventional additive noise. 5th and 95th percentiles of a block bootstrap assuming independence of samples on each day are over-plotted on the evolved additive noise lines. Error is measured with respect to the EnSRF ensemble-mean analysis.

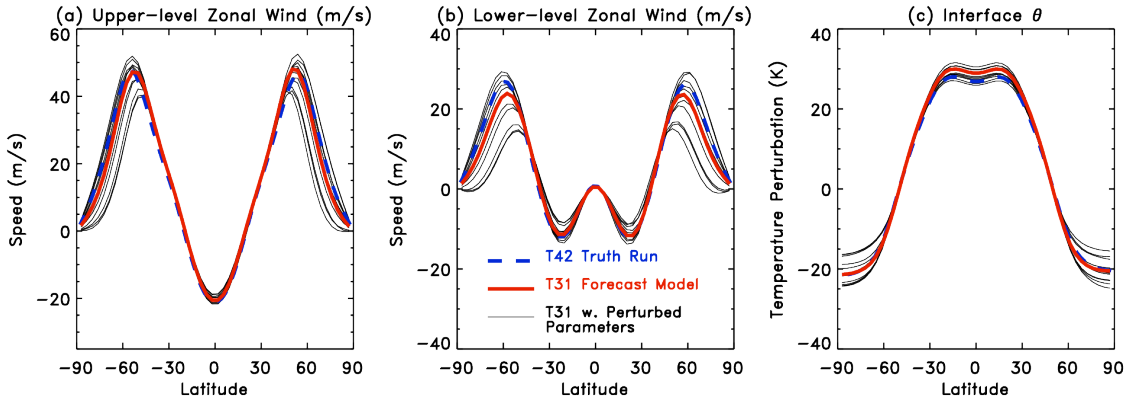

Figure 1: Zonal-mean averages for (a) upper-level *u*-wind component, (b) lower-level *u*-wind component, and (c) interface barotropic potential temperature in nature runs. Red line indicates the T31 model used for data assimilation experiments, dashed blue line the T42 model nature run used in imperfect-model experiments, and thin solid lines the various T31 models run with perturbed pole-to-equator temperature difference and perturbed mechanical damping time scales.

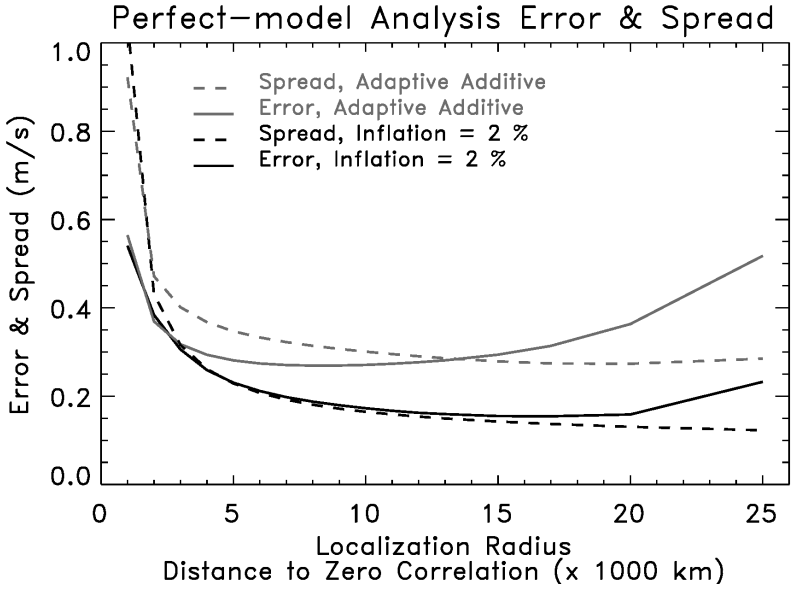

Figure 2: Spread and error for perfect-model experiments when ensemble is stabilized by covariance inflation (black curves) and adaptive additive noise (grey curves).

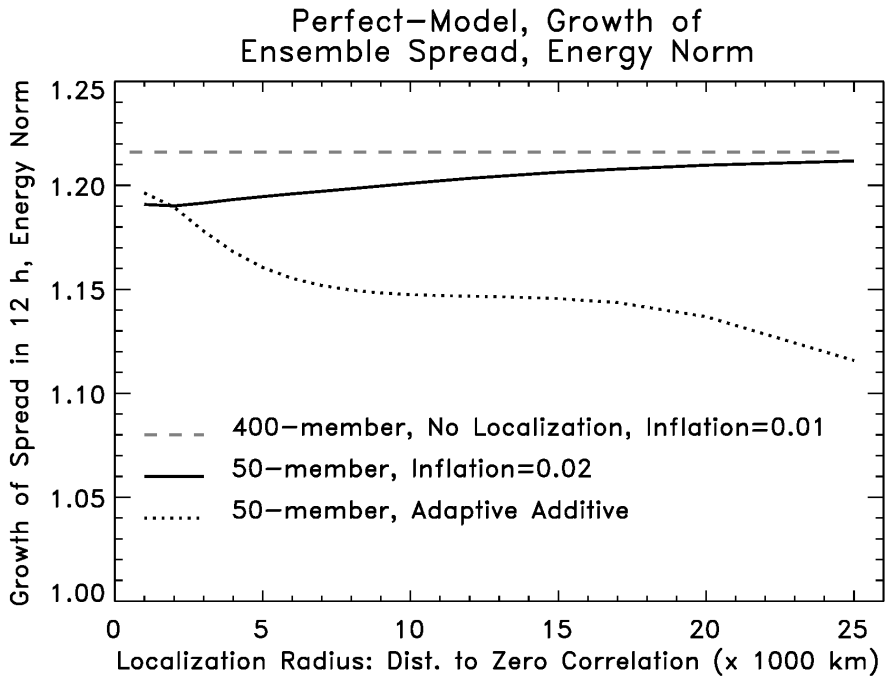

Figure 3: Growth of ensemble forecast spread during the 12-h between update steps for perfect-model experiments.

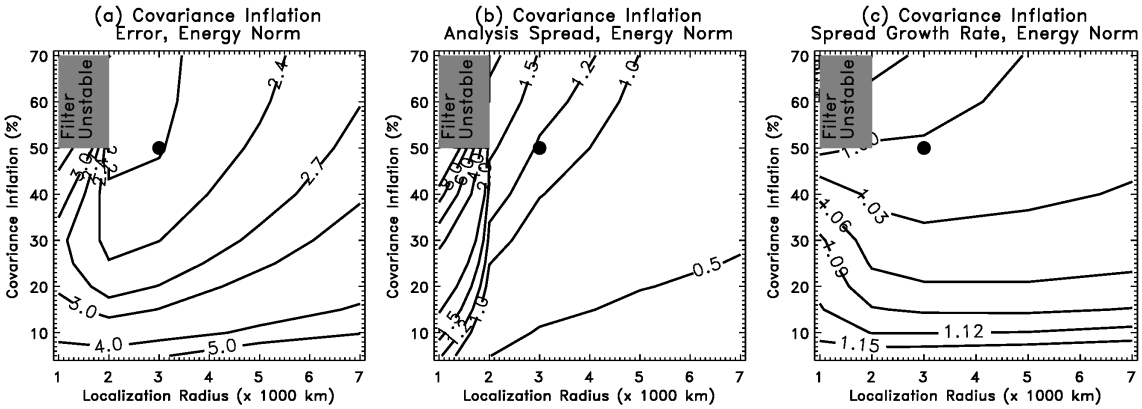

Figure 4. (a) Ensemble-mean RMS error in the energy norm for imperfect-model experiments with the EnSRF when filter is stabilized by covariance inflation. Data is plotted as a function of the covariance localization radius (x axis) and the amount of covariance inflation (y axis). The black dot indicates the (localization radius, inflation amount) pair used in Fig. 5. (b) As in (a), but for ensemble spread in the energy norm. (c) As in (a), but for the ensemble spread growth rate over the 12-h period between data assimilation updates.

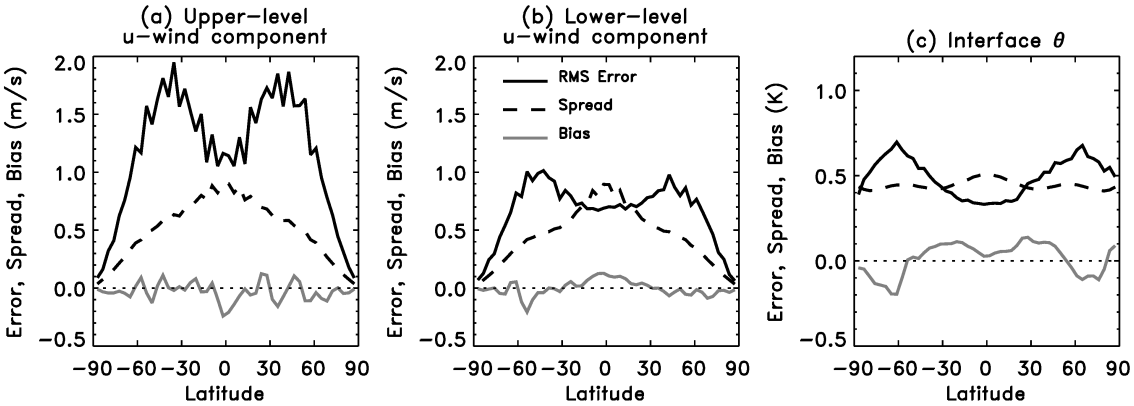

Figure 5. (a) Zonal and time average of the upper-level *u*-wind component analysis RMS error, spread, and bias for the 3000-km localization and 50% inflation rate data point shown in Fig. 4. (b) As in (a), but for lower-level *u*-wind component. (c) As in (a), but for interface potential temperature.

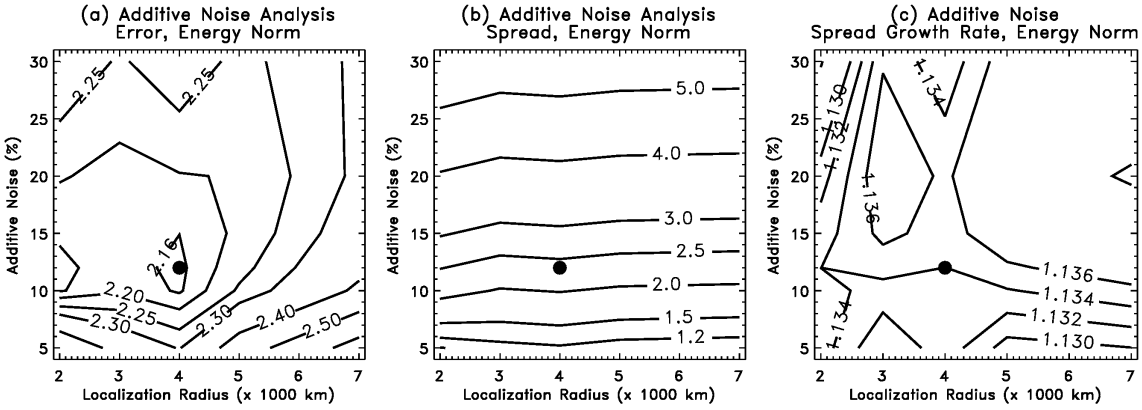

Figure 6: As in Fig. 4, but for imperfect-model experiments stabilized by additive noise. Dot indicates the localization radius and additive noise amount where error was approximately at a minimum and spread was approximately consistent with error. Spread and error for this combination is examined more in Fig. 7.

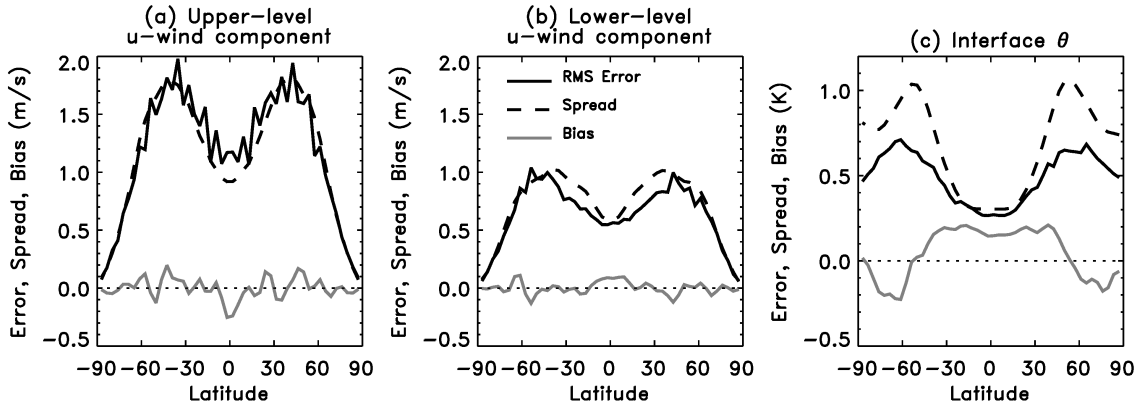

Figure 7: As in Fig. 5, but for the 4000-km localization and 12% additive noise experiment data shown in Fig. 6.

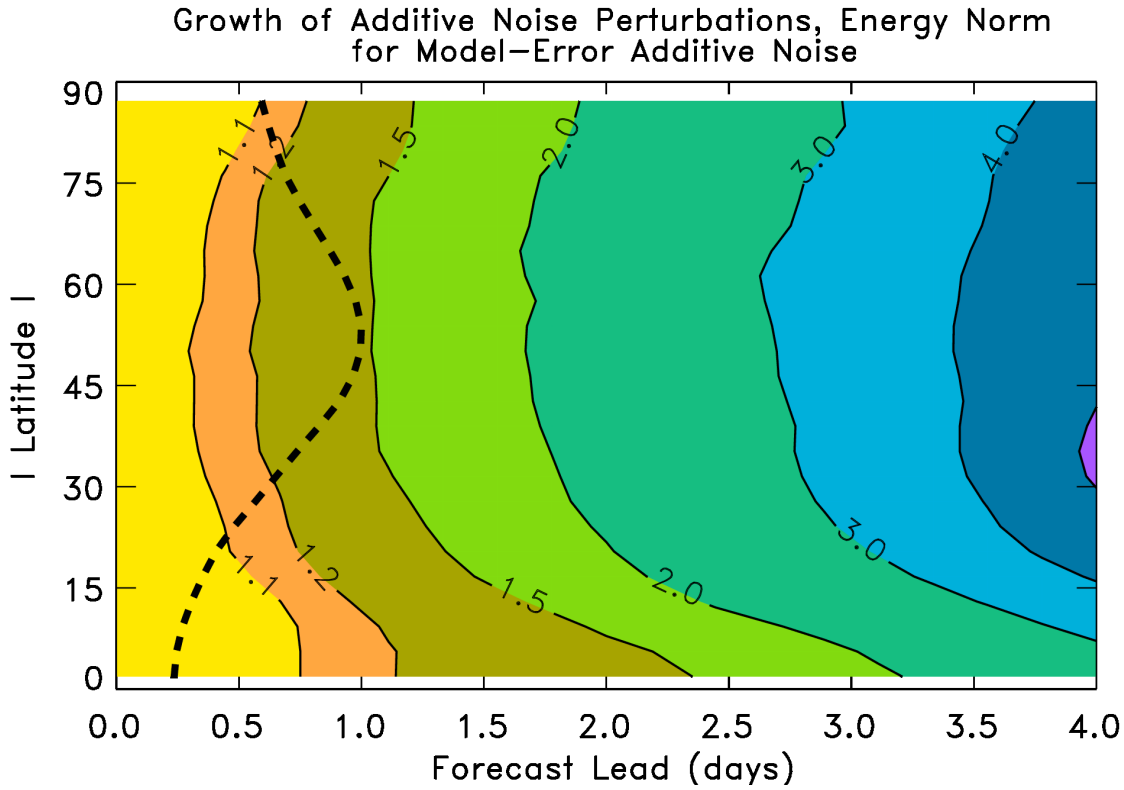

Figure 8: Growth of small model-error additive noise perturbations around a state sampled from the T31 forecast model nature run. Dashed line indicates the relative proportion of total energy in the perturbations at the initial time. Perturbations are scaled so that their average magnitude at each latitude is 1.0 at the initial time.

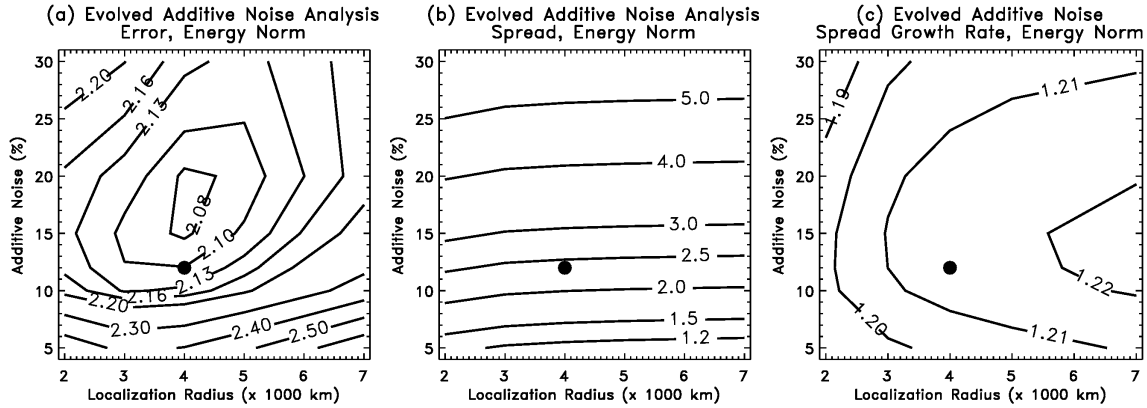

Figure 9. As in Fig. 6, but where evolved additive noise is used to stabilize the filter. Black dot indicates the localization radius/additive-noise amount that was approximately a minimum in Fig. 7.

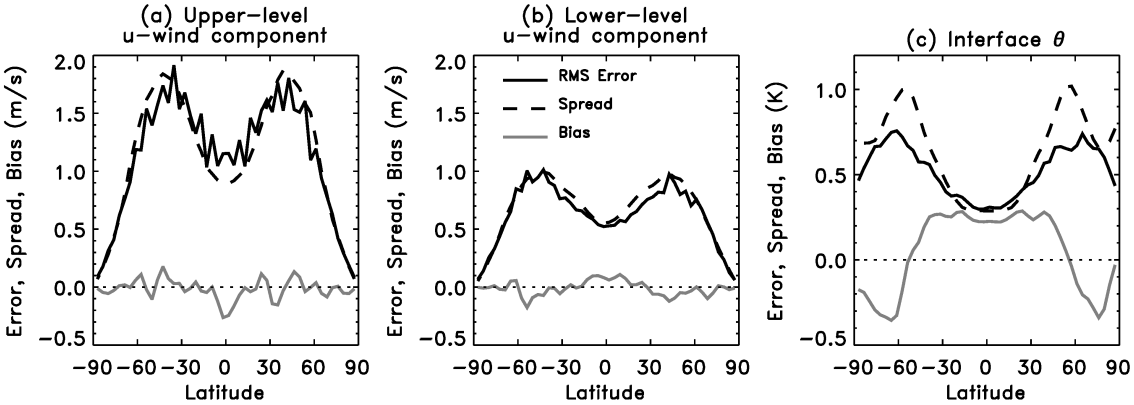

Figure 10: As in Fig. 5, but for the 4000-km localization and 12% evolved additive noise experiment data shown in Fig. 9.

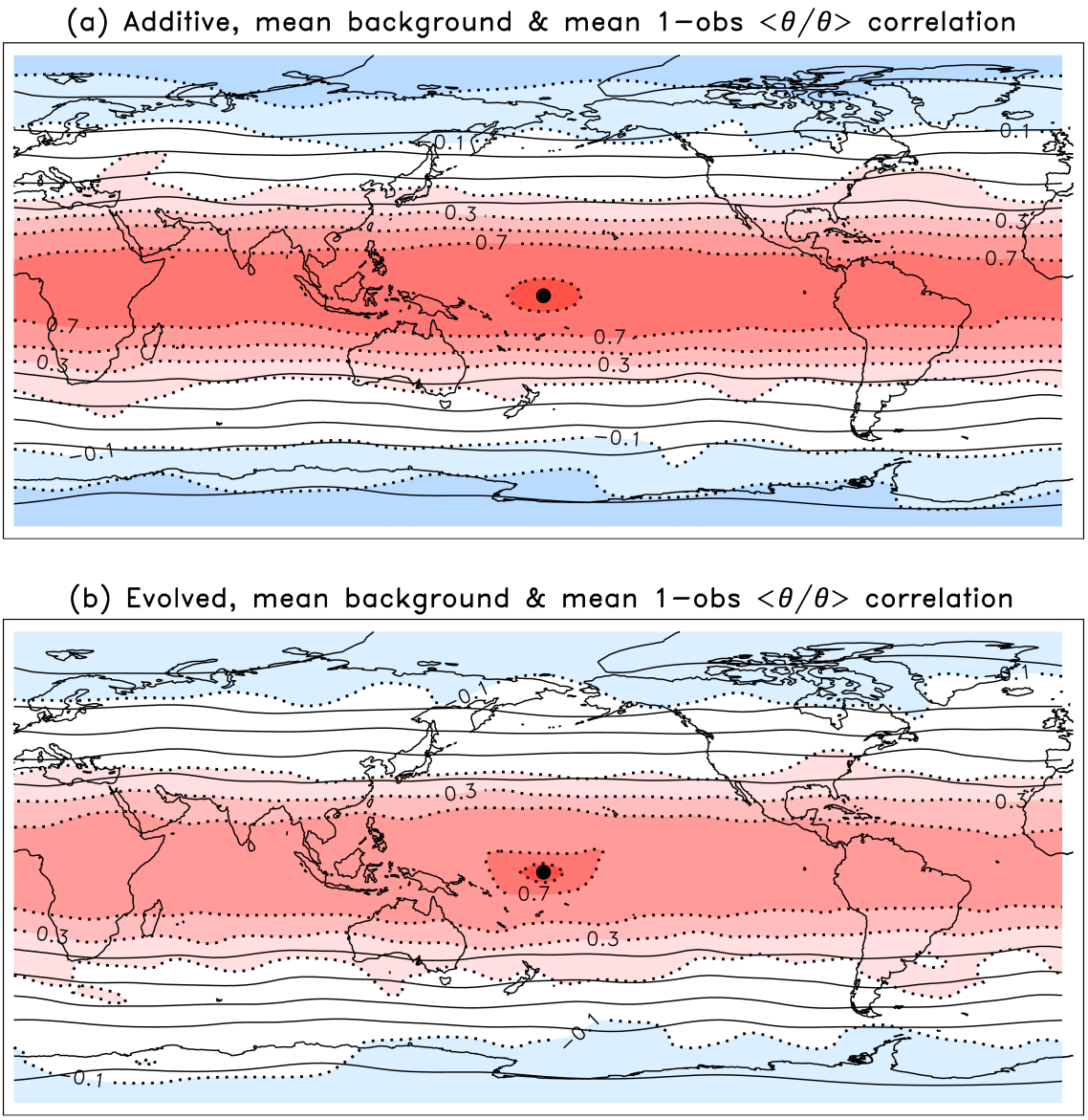

Figure 11: Map of multi-case average background correlations without covariance localization (colors and dashed lines) between the ensemble interface potential temperature at the location with the dot and the ensemble temperature at every other grid point. The average background potential temperature is also plotted (solid lines, every 10K). Panel (a) is for the additive noise, panel (b) for the evolved additive noise.

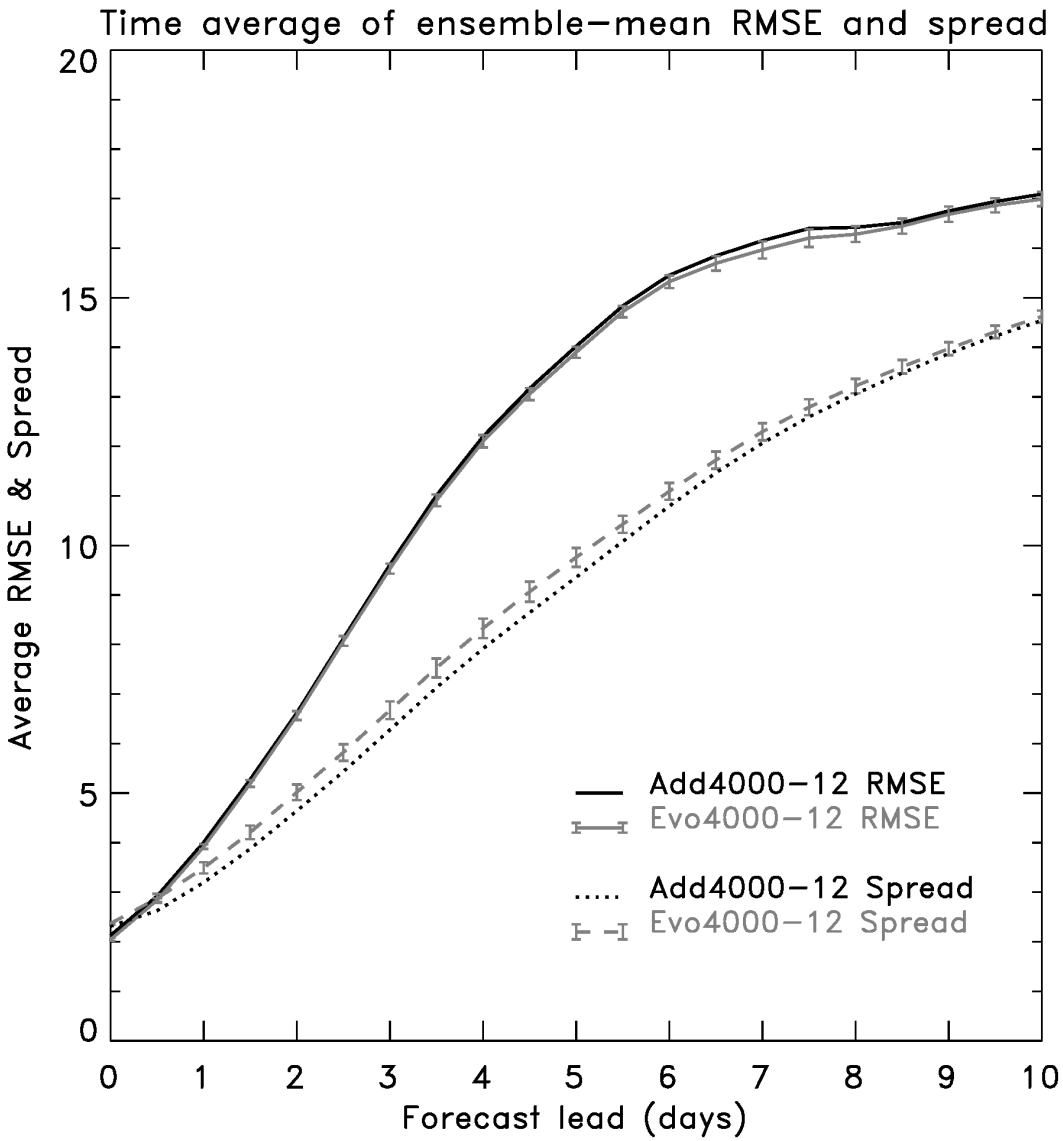

Figure 12: Ensemble forecast spread and RMS error in the energy norm, initialized from imperfect-model additive-noise forecasts with 4000 km localization and 12% additive noise scaling (Add4000-12, black lines); from evolved additive noise with 4000 km localization and 12% additive noise scaling (Evo4000-12, grey lines); Error bars represent the 5th and 95th percentiles from a paired block bootstrap.

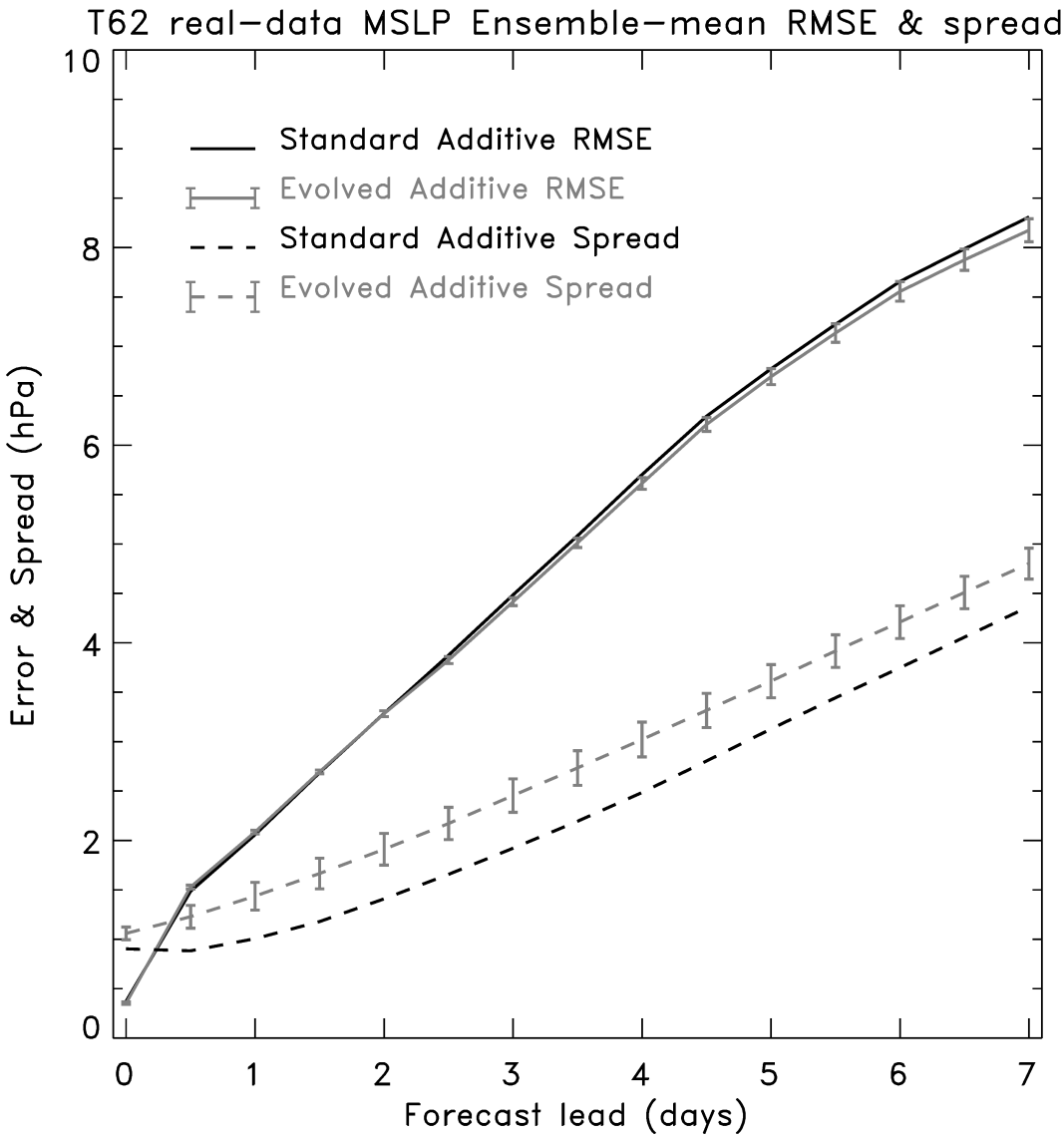

Figure 13: Ensemble forecast spread of mean-sea level pressure (dashed lines) and error (solid lines) from T62 GFS experiments. Grey color denotes evolved additive noise and black denotes conventional additive noise. 5th and 95th percentiles of a block bootstrap assuming independence of samples on each day are over-plotted on the evolved additive noise lines. Error is measured with respect to the EnSRF ensemble-mean analysis.