Local Predictability of the Performance of an Ensemble Forecast System

Elizabeth Satterfield *

Istvan Szunyogh

University of Maryland, College Park, Maryland

To be submitted to JAS

* *Corresponding author address:* Elizabeth Satterfield, Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, MD 20742-2425.

E-mail: esatterf@atmos.umd.edu

ABSTRACT

The main goal of the present study is to lay the theoretical foundation of a practical approach to predict the spatio-temporal changes in the performance of an ensemble prediction system. The motivation to develop such an approach is the recognition that the performance of an ensemble prediction system is inherently flow dependent.

Linear diagnostics applied to the ensemble perturbations in a small local neighborhood of each model grid point are used to explore the spatio-temporally changing predictive qualities of the ensemble. In particular, a local state vector and the associated local covariance matrix is defined to represent the state and the uncertainty in the state estimate at each grid point. A set of local diagnostics based on the eigen-solution of the local covariance matrix is introduced.

Numerical experiments are carried out with an implementation of the Local Ensemble Transform Kalman Filter (LETKF) data assimilation system on a reduced resolution (T62L28) version of the National Centers of Environmental Prediction (NCEP) Global Forecast System (GFS). It is found that the ensemble dimension (E-dimension) diagnostic is a good predictor of the ensemble performance, in the sense that a low value of the E-dimension indicates that the ensemble efficiently captures the uncertain forecast features. The significance of this result is that the location of large forecast errors in the 1-5 days are also the locations of low E-dimension.

1. Introduction

The purpose of an ensemble prediction system is to account for the influence of the spatio-temporal changes in predictability on the forecasts [e.g., Kalnay (2002); Palmer and Hagedorn (2006)]. It is often assumed that an ensemble prediction system can provide a closure of the predictability problem: failures of the system in predicting the forecast error statistics are usually attributed to design flaws of the system. Kuhl et al. (2007), KEA07 hereafter, argued that this view should be reconsidered by showing that the performance of an ensemble prediction system is inherently flow dependent. We refer to this dependence as the local predictability of the ensemble performance.

The study of KEA07 was based on assimilating randomly located simulated observations under the perfect model scenario with an implementation of the Local Ensemble Transform Kalman Filter data assimilation system (Hunt et al. 2006; Szunyogh et al. 2008) on a reduced resolution (T62 and 28 vertical levels) version of the model component of the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS). The purpose of the present study is to extend the investigation of KEA07 building to a more realistic setting by first assimilating simulated observations in realistic locations under a perfect model scenario and then assimilating an operationally used set of observations of the real atmosphere. The main goal of our study is to lay the theoretical foundation of a practical approach to predict the spatio-temporal changes in the performance of an ensemble prediction system.

The structure of the paper is as follows. In section 2, we introduce the diagnostics we use to assess and explain the performance of the ensemble prediction system at the different locations and times. In section 3, we describe the design of the numerical experiments.

In section 4, we examine the spatio-temporal evolution of the forecast errors, which is our preferred way to assess the spatio-temporal evolution of predictability, and we analyze the relationship between predictability and our diagnostics. In section 5, we summarize our conclusions.

2. Diagnostics

We use linear diagnostics applied to the ensemble perturbations in a small local neighborhood of each model grid point to explore the spatio-temporally changing predictive qualities of the ensemble. In particular, we define a local state vector and the associated local covariance matrix to represent the state and the uncertainty in the state estimate at each grid point. In addition, we introduce a set of local diagnostics based on the eigen-solution of the local covariance matrix and a measure of nonlinearity in the evolution of the local state vectors.

a. Local vectors and their covariance

We define a local state vector $\mathbf{x}(\ell)$ with all N state variables of the model within a local volume centered at location (grid point) ℓ. For the rest of this article, we will discuss what to do at an arbitrary location ℓ, and so we now drop the argument ℓ. The mathematical model we adopt to predict the evolution of uncertainty in a local state estimate (analysis or forecast), $\mathbf{x}^{a,f}$, is based on the assumption that the error in the state estimate,

$$\boldsymbol{\xi} = \mathbf{x}^{a,f} - \mathbf{z}, \tag{1}$$

is a random variable. In Equation (1) $\mathbf{z}$ is the, in practice unknown, true state of the atmosphere. In addition, we assume that $\boldsymbol{\xi}$ is normally distributed with a zero mean. Under this assumption the error in the state estimate can be fully described by an error covariance matrix, $\mathbf{P}$, that defines the covariance between the different components of $\boldsymbol{\xi}$.

We employ a K-member ensemble of local state vectors, $\mathbf{x}^{(k)}$, $k = 1 \dots K$, to represent the uncertainty in the knowledge of the local state. The ensemble-based estimate of the covariance matrix $\mathbf{P}$ is,

$$\hat{\mathbf{P}} = (K-1)^{-1} \sum_{k=1}^{K} \delta\mathbf{x}^{(k)} \left(\delta\mathbf{x}^{(k)}\right)^T, \tag{2}$$

where the ensemble perturbations $\delta\mathbf{x}^{(k)}$, k=1...K, are defined by the difference,

$$\delta\mathbf{x}^{(k)} = \mathbf{x}^{(k)} - \bar{\mathbf{x}},\ k = 1 \dots K, \tag{3}$$

between the ensemble members $\mathbf{x}^{(k)}$, $k = 1 \dots K$, and the ensemble mean,

$$\bar{\mathbf{x}} = K^{-1} \sum_{k=1}^{K} \mathbf{x}^{(k)}. \tag{4}$$

In Equation (2), T denotes the matrix transpose. Based on Equations (3) and (4) the sum of the ensemble perturbations is zero at all forecast lead times, that is,

$$\sum_{k=1}^{K} \delta\mathbf{x}^{(k)} = \mathbf{0}. \tag{5}$$

Equation (5) indicates that the K ensemble perturbations are not linearly independent. Thus, the dimension of the linear space $\mathbb{S}$ spanned by the ensemble perturbations cannot be larger than *K-1*.

Since $\hat{\mathbf{P}}$ is a nonnegative definite and symmetric N by N matrix, it has N non-negative eigenvalues, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \dots \geq \lambda_N \geq 0$, and the N associated eigenvectors, $\mathbf{u}_n$,

$n = 1, \ldots, N$, are orthogonal with respect to the Euclidean inner product. That is, when the eigenvectors are chosen to be of unit length with respect to the Euclidean vector norm,

$$\left(\mathbf{u}_i\right)^T \mathbf{u}_j = \delta_{ij}, \tag{6}$$

where $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$. When the number of components of the local state vector is larger than the number of ensemble members ($N > K$), only the first $K - 1$ eigenvalues can be larger than zero (In what follows, $N > K$ is assumed unless noted otherwise). In this case, the normalized eigenvectors associated with the first $K - 1$ eigenvalues, $\mathbf{u}_k$, $k = 1, \ldots, K - 1$, define an orthonormal basis in $\mathbb{S}$. The physical interpretation of the N-vectors $\mathbf{u}_k$, $k = 1, \ldots, K - 1$, is that they represent linearly independent uncertain analysis and forecast patterns in the ensemble in the local region.

An arbitrary local state vector $\mathbf{x}$ can be decomposed as

$$\mathbf{x} = \bar{\mathbf{x}} + \delta\mathbf{x}, \tag{7}$$

where $\delta\mathbf{x}$ is the deviation of $\mathbf{x}$ from its ensemble mean $\bar{\mathbf{x}}$. The perturbation vector $\delta\mathbf{x}$ can be further decomposed as,

$$\delta\mathbf{x} = \delta\mathbf{x}^{(\|)} + \delta\mathbf{x}^{(\perp)} \tag{8}$$

where $\delta\mathbf{x}^{(\|)}$ is the component that projects into $\mathbb{S}$ that is,

$$\delta\mathbf{x}^{(\|)} = \sum_{k=1}^{K-1} \delta x_k \mathbf{u}_k \tag{9}$$

where the coordinate δx_k, $k = 1 \ldots K - 1$, can be computed by

$$\delta x_k = \delta\mathbf{x}^{(\|)T} \mathbf{u}_k \tag{10}$$

The vector $\delta\mathbf{x}^{(\perp)}$ denotes the component of $\delta\mathbf{x}$ that projects into the null space of $\hat{\mathbf{P}}$, that is $\delta\mathbf{x}^{(\perp)}$ is that component of the perturbation, which cannot be represented by the ensemble

perturbations. Using this notation, the error $\boldsymbol{\xi}$ in the state estimate $\mathbf{x}^{a,f}$ can be decomposed as

$$\boldsymbol{\xi} = \mathbf{x}^{a,f} - \mathbf{z} = \sum_{k=1}^{K-1} \big(\delta x_k^{a,f} - \delta z_k\big)\mathbf{u}_k + \big(\delta \mathbf{x}^{a,f(\perp)} - \delta \mathbf{z}^{(\perp)}\big) = \delta\boldsymbol{\xi}^{(\|)} + \delta\boldsymbol{\xi}^{(\perp)}, \tag{11}$$

where,

$$\delta z_k = \mathbf{z}^T\mathbf{u}_k, \quad \delta\boldsymbol{\xi}^{(\|)} = \sum_{k=1}^{K-1} \big(\delta x_k^{a,f} - \delta z_k\big)\mathbf{u}_k, \quad \delta\boldsymbol{\xi}^{(\perp)} = \delta\mathbf{x}^{a,f(\perp)} - \delta\mathbf{z}^{(\perp)} \tag{12}$$

Although the ensemble mean, or the error in the ensemble mean, does not appear directly in the local decomposition of the error (the rhs of Eq 11), the ensemble mean provides the reference point for the definition of the basis vectors that span the space $\mathbb{S}$.

The main focus of our investigation in this paper is the linear space $\mathbb{S}$. We choose diagnostics and design numerical experiments to identify the conditions under which the evolution of the forecast uncertainties can be efficiently described in $\mathbb{S}$. We emphasize that an ensemble forecast system can, in principle, describe the evolution of the forecast uncertainty even if it follows a probability distribution much more complex than normal. Restricting our attention to $\mathbb{S}$, however, has a number of practical advantages. First, studies that attempt to directly verify the probability distribution forecast face the problem that a probabilistic forecast can be verified only in a statistical sense, which means that one cannot validate the ensemble forecast at a given time and location. Also, extreme weather events are rare and it is often impossible to collect a large enough sample of them to verify the performance of the ensemble in their prediction by probabilistic verification scores (e.g., Toth et al. (2006)). In contrast, $\mathbb{S}$ is a deterministic forecast of the space of uncertain forecast features, thus, it can be verified on a case by case basis. Also, one can verify whether the error in the prediction of an extreme weather event projects into $\mathbb{S}$. Of course, $\mathbb{S}$ carries less information

than a quantitative probability forecast in cases where the probability distribution is more complex than normal, but an accurate prediction of the potential error patterns is still valuable forecast information. Second, because $\mathbb{S}$ is a linear space, it justifies the use of linear statistical post-processing techniques, such as Hamill and Whitaker (2007), to enhance the raw probability forecasts provided by the ensemble.

b. Explained variance

In order to quantify the performance of $\mathbb{S}$ in capturing the uncertain components of the flow in a state estimate, we use the explained variance diagnostic. Formally, it is calculated as

$$EV = \frac{\|\delta\boldsymbol{\xi}^{(\|)}\|}{\|\boldsymbol{\xi}\|} = \frac{\|\delta\boldsymbol{\xi}^{(\|)}\|}{\|\delta\boldsymbol{\xi}^{(\|)} + \delta\boldsymbol{\xi}^{(\perp)}\|}. \tag{13}$$

Here $\|\cdot\|$ is the Euclidean vector norm on the space of the local state vectors. (Since $\mathbb{S}$ is a subspace of the space of the local state vectors, this norm can be used to measure the magnitude of both the error and its projection into $\mathbb{S}$.)

The larger EV, the more efficient is $\mathbb{S}$ in capturing the uncertain components of the analysis and forecast fields. EV takes its maximum value of one when the entire forecast error projects into $\mathbb{S}$ ($\delta\boldsymbol{\xi}^{(\|)} = \boldsymbol{\xi}$ and $\delta\boldsymbol{\xi}^{(\perp)} = \mathbf{0}$), and takes its minimum value of zero when the forecast error does not have projection into $\mathbb{S}$ ($\delta\boldsymbol{\xi}^{(\|)} = \mathbf{0}$ and $\delta\boldsymbol{\xi}^{(\perp)} = \boldsymbol{\xi}$).

c. E-dimension

The ensemble dimension (E-dimension),

$$E = \frac{\left[\sum_{i=1}^{K} \sqrt{\lambda_i}\right]^2}{\sum_{i=1}^{K} \lambda_i}, \tag{14}$$

which was introduced by Patil et al. (2001) and discussed in details in Oczkowski et al. (2005), characterizes the local complexity of dynamics. E is a spatio-temporally evolving measure of the steepness of the eigenvalue spectrum, $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_r \ldots \geq \lambda_K$, having smaller values for a steeper spectrum (Szunyogh et al. 2007). For our choice of the perturbations, where the K perturbations are linearly dependent, $\lambda_K = 0$, the largest possible value of E is *K-1*. For a set of linearly independent ensemble perturbations, the maximum value of E is equal to the number of ensemble perturbations, K, which occurs when the uncertainty predicted by the ensemble is evenly distributed between K linear spatial patterns in $\mathbb{S}$.

d. Spectrum of the d-ratio

While the explained variance diagnostic quantifies the efficiency of the space $\mathbb{S}$ in capturing the space of the uncertainty in the state estimate, it does not measure the performance of the ensemble in distinguishing between the more and less important directions within $\mathbb{S}$. To introduce a diagnostic that can measure the performance of the ensemble in quantifying the importance of the different state space directions within $\mathbb{S}$, we first recall that the eigenvalues $\lambda_1, \lambda_2 \ldots \lambda_{K-1}$ measure the relative importance of the eigen-directions $\mathbf{u}_k$, $k = 1, \ldots, K-1$, in a statistical sense. More precisely, λ_k is the ensemble-based prediction of the variance of

the uncertainty in the k-th eigen-direction[1]. Our goal is to verify this prediction.

Since the prediction of a variance is a probabilistic prediction, it cannot be verified on a case by case basis. Instead, the verification of such a prediction requires a probabilistic verification score and a sufficiently large sample of the prediction of the variance. Our choice for the probabilistic verification score is the *d-ratio*, which was first introduced in Ott et al. (2002). Since the d-ratio

$$d_k = \sqrt{\frac{\left(\delta\boldsymbol{\xi}_k\right)^2}{\lambda_k}}, \tag{15}$$

is defined independently for each eigen-direction, it is more appropriate to talk about a *spectrum of the d-ratio*. If the ensemble, on average, correctly predicts the uncertainty in the k-th direction, the mean of d_k over a large sample is one. If the ensemble overestimates the uncertainty in the k-th direction, the mean of d_k is smaller than one, while a mean value of d_k larger than one indicates an underestimation of the uncertainty in the k-th direction.

3. Experiment Design

We carry out numerical experiments both under the perfect model scenario and in a realistic NWP setting. In the perfect model experiments, we generate simulated observations of the hypothetical "true" trajectory of the atmospheric state, where the time series of "true" states, $\mathbf{z}$, is generated by a 60-day model integration of the GFS model at T62L28 resolution starting from an operational NCEP analysis truncated to T62L28 resolution. We first repeat the experiment of KEA07 to verify that the findings of that paper remain valid for the

[1]Graphically, the vectors $\sqrt{\lambda_k}\mathbf{u}_k$ are the principal axes of the ellipsoid defined by $(\mathbf{x}-\overline{\mathbf{x}})^T(\mathbf{P})^{-1}(\mathbf{x}-\overline{\mathbf{x}}) = 1$. This ellipsoid represents states of equal probability in $\mathbb{S}$.

different time period investigated here (January and February of 2004 instead of January and February of 2000).[2] Then, we build to the realistic NWP setting in two steps: first we replace the randomly located simulated observations by simulated observations taken by a realistic observing network, then we replace the simulated observations with observations of the real atmosphere.

a. Observational data sets

1) Randomly placed simulated observations

The "truth", $\mathbf{z}$, is taken to be an integration of the GFS model starting from the operational NCEP analysis at 0000 UTC 1 January 2004. At each grid point and model level, we generate simulated observations of the two horizontal components of the wind, the temperature, and the surface pressure by perturbing the "true" states with normally distributed, zero mean assumed observational errors with standard deviations of 1 K, 1.1 m/s, and 0.6 hPa for temperature, wind, and surface pressure, respectively. Next, similar to Szunyogh et al. (2005) and KEA07, we randomly choose 2000 soundings, to reflect a 10% observational coverage of the model grid. By choosing observations randomly, we ensure that the simulated observing network has little systematic impact on the geographical distribution of analysis and forecast errors.

[2]Another important difference between the experiment design of the two studies is that we use a later version of the LETKF. Most importantly, the LETKF used in this study provides more accurate analyses in the polar regions.

2) Simulated observations at realistic locations

In the second set of experiments, we assimilate simulated observations at the locations of routine non-radiance observations of the real atmosphere. These simulated observations are generated by adding random observational noise, created by using the standard deviation of the estimated observational error provided with each observation by NCEP, to the "true" grid point values of the surface pressure, the temperature, and the two horizontal components of the wind vector. The location and type of observations is obtained from a database that includes all observations operationally assimilated at NCEP between 000UTC 1 January 2004 and 000UTC 15 February 2004, with the exception of satellite radiances, but including satellite derived winds. We also exclude all surface observations, except for the surface pressure and the scatterometer wind measurements over oceans.

3) Observations of the real atmosphere

Finally, the observations of the real atmosphere, which are used to obtain the type and location for the simulated observations at realistic locations, are assimilated.

b. Selection of the LETKF parameters

For each observational data set, an analysis is obtained at the native model resolution every 6 hours. Diagnostics are computed at a reduced $2.5^\circ \times 2.5^\circ$ grid resolution. We assimilate observations between 1 January 2004 0000 UTC and 15 February 2004 0000 UTC. In these experiments, multiplicative covariance inflation is used at each analysis step to

increase the estimated analysis uncertainty to compensate for the loss of ensemble variance due to sampling errors, the effects of nonlinearities and model errors. The parameters of the LETKF used in this experiment are the following:

- The ensemble has $K = 40$ members.
- Observations are considered in a 800 km horizontal radius of the grid point, where the state is estimated.
- Observations have equal weight within a 500 km radius of the given grid point, beyond which the weight of the observations tapers linearly to zero at 800 km.
- Observations are considered in a vertical patch radius centered at the grid point. This layer has depth 0.35 scale height between model levels 1 to 15 and gradually increases to 2 at the top of the model atmosphere.
- For simulated randomly distributed observations, we use a 10% covariance inflation at all vertical levels in all geographic regions.
- For the simulated observations taken at realistic locations, the covariance inflation is 2.5% at all vertical levels in the SH extratropics and 10% in the NH extratropics. In the Tropics, the covariance inflation varies from 2.5% to 7.5%.
- For the conventional observations of the real atmosphere, the covariance inflation tapers from 25% at the surface to 20% at the top of the model atmosphere in the SH extratropics and from 50% to 30% in the NH extratropics, and changes smoothly in the tropics (between 25°S and 25°N) from the values of the SH extratropics to the values of the NH extratropics.

- For all data sets, surface pressure is assimilated at the first model level and temperature, and zonal and meridional winds are assimilated at all 28 model levels.

The variance inflation factors were determined by numerical experimentation, searching for values that minimized the analysis root-mean-square errors.

c. Initialization

In the two sets of experiments which assimilate observations in realistic locations, high-frequency oscillations (typically associated with gravity waves) are filtered from all background ensemble members with a digital filter scheme (Huang and Lynch 1993), which is part of the NCEP GFS model and can be turned on or off by choice. (Unlike in the original formulation of the digital filter algorithm, where a filtered analysis is produced, the NCEP filter provides only a filtered background field.) We use the filter with a 3 hr cutoff frequency. We find that turning the digital filter on in these two sets of experiments leads to a major improvement of the analyses.

In the experiments with randomly placed observations, turning the digital filter on degrades the analysis in the Tropics (Figure 1). More precisely, the surface pressure errors with the digital filter turned on (top panel of Figure 1) have a clear wavenumber two pattern in the tropics. This suggests that the digital filter leads to the distortion of an equatorial wave type motion in the forecasts that serve as the background ensemble in the LETKF calculations. A more careful examination of the structure of the error fields reveals that the distortion is in the amplitude of the *semidiurnal tidal wave* [3]. We illustrate this distortion

[3]The possibility that the error field in the top panel of Figure 1 may be associated with the semidiurnal

by showing the spectrum of Fourier amplitudes of the time series of surface pressure at the location 0°N, 160°W for the nature run and the two analyses (Figure 2). The 12-hour frequency oscillation characteristic of the semidiurnal tidal wave is not present in the run that uses the digital filter, even though this oscillation is the dominant signal in the nature run and in the analysis cycles that do not use the digital filter. (We also note that the digital filter initialization has a negative effect on the analysis of the semi diurnal tidal wave, but this problem gets exposed only in the experiment with the randomly placed observations, where the analysis does not generate spurious gravity wave oscillations elsewhere.)

The semidiurnal tidal wave is primarily caused by the absorption of solar radiation by ozone in the stratosphere and the atmosphere. The response to this stratospheric excitation propagates downward in the form of an inertia-gravity wave (Chapman and Lindzen 1970). Our conjecture is that the digital filter affects this inertia-gravity wave. We also suspect that applying a digital filter initialization to the analysis increment instead of filtering the six-hour background forecast, would eliminate the negative effect of the filter on the semi-diurnal tidal wave [4].

d. Forecasts

We prepare the deterministic forecasts daily, started from the mean analysis at 0000UTC and 1200UTC, and output every 12 hours. These model integrations provide the state estimate $\mathbf{x}^{a,f}$. At analysis time and at short forecast lead times (while the time evolution of the ensemble perturbations stays linear), this state estimate provides our best deterministic

tidal wave signal was first pointed out to us by Nedjeljka Zagar of the University of Ljubljana.

[4]We are currently in the process of developing such an initialization algorithm for the LETKF scheme.

estimate of the state. At longer lead times $\mathbf{x}^{a,f}$ simply represents a forecast for which the analysis was drawn from a probability distribution that is consistent with our estimate of the analysis uncertainty.

In addition to the state estimate, the LETKF also generates the ensemble of analyses that estimate the uncertainty in the state estimate. These analyses serve as initial conditions for the ensemble of forecasts. Ensemble forecasts are obtained once daily, started from the ensemble of analyses, at 0000UTC and output every 12 hours. Both the deterministic forecast and the ensemble forecasts are carried out to a five day lead time. Unlike the experiments which use realistically placed observations, forecasts for the experiment that assimilates observations in random locations are run without the use of the digital filter. We note that turning the digital filter off in this experiment slightly increases the forecast error up to 12-hour lead times, after which the filter has no effect on the forecast errors.

Forecast error statistics are computed by comparing the deterministic forecasts, $\mathbf{x}^{a,f}$, to the "true" states, $\mathbf{z}$. Forecasts started from analyses generated by assimilating conventional observations of the real atmosphere are verified using the high (T254L64) resolution operational NCEP analyses truncated to 2.5° x 2.5° resolution as proxy for the "true" state. These operational analyses were obtained by NCEP assimilating a large number of satellite radiance observations in addition to the conventional observations used in our experiments. Forecast error statistics are generated for the 36-day period, 0000 UTC 11 January 2004 - 0000 UTC 15 February 2004.

To calculate the E-dimension, we follow the approach of Oczkowski et al. (2005) and KEA07: we transform the ensemble perturbations such that the square of the Euclidean norm of the transformed perturbations has dimension of energy. For this calculation, the

local volume is defined, as in KEA07, by all temperature, wind, and surface pressure grid point variables in a cube that is defined by 5x5 horizontal grid points (at 2.5° x 2.5° grid resolution) and the entire column of the model atmosphere. For this definition of local region, the dimension of the local state vector, N, is 1975 (Since K=40, K<N, as we assumed in Section 2.)

4. Numerical experiments

a. Forecast Errors

First, to illustrate the general spatial distribution of the errors in the $\mathbf{x}_g^{a,f}$ state estimate we examine the absolute error in the analysis and forecasts of the meridional wind component at 500 hPa. We choose the meridional wind instead of the more commonly used geopotential height, because this way we can use the same quantity to characterize the errors in the Tropics and the extratropics. Plots of the absolute error are obtained by computing the time average of $\|\boldsymbol{\xi}\|$ at each location (grid point). Figure 3 shows the time mean absolute error at analysis time and at the 72-hour forecast lead time for all three experiments. The results obtained by assimilating simulated observations in randomly placed locations show that the largest analysis errors are in the Tropics and the smallest analysis errors are in mid-latitude storm track regions, in agreement with Szunyogh et al. (2005). Forecast errors become dominant in the storm track regions within 48-72 hours. In comparison, when simulated observations are placed in realistic locations, the results show that the distribution of the magnitude of the analysis errors is strongly modulated by the observation density: the lowest errors are

over continents in the Northern Hemisphere and the highest errors are over Antarctica and in the oceanic region between Cape Horn and the Antarctic Peninsula.

We see strong similarities in the spatial distribution of the errors at analysis time and for short term forecasts in both experiments that assimilate observations in realistic locations. This similarity indicates that observation density plays a more dominant role than model error in determining the spatial variation of the analysis and the short term forecast errors. Nevertheless, the results obtained by assimilating observations of the real atmosphere show that the magnitude of the forecast error is almost double the forecast error found in the experiments which used simulated observations. In all three experiments, we find rapid growth of forecast errors in the mid-latitude storm track regions, which become the dominant region of forecast error by the 72-hour lead time.

b. E-Dimension, Explained Variance, and Forecast Error

Szunyogh et al. (2005) showed that for lower values of E-dimension, the ensemble more certainly captured the structure of the background error. KEA07 extended the E-dimension diagnostic to study predictability of the performance of ensemble forecasts and found that, in the extratropics, fast error growth led to low E-dimension and, therefore, to increased certainty that a greater portion of the forecast error was efficiently captured by the ensemble.

We investigate the relationship between the E-dimension, explained variance, and forecast errors with the help of joint probability distribution functions (JPDFs). The JPDF shown in Figure 4 is obtained by calculating the number of occurrences in each bin defined by $\Delta E \times \Delta EV$, where ΔE denotes the bin increment for E-dimension and ΔEV denotes

the bin increment for the explained variance. The number of occurrences is then normalized by $\Delta E \times \Delta EV \times n$, where n is the total sample size, which is equal to the total number of grid points in a geographic region multiplied by the total number of verification times. This normalization ensures that the integral of the plotted values over all bins is equal to one. In good agreement with KEA07, we find that, *the lower the E-dimension, the greater the probability that explained variance is high. We find this relationship independent of experiment, lead time and geographic region.* As forecast lead time increases, lower values of E-dimension have a greater probability of corresponding to high value of explained variance. At analysis time, we find lower values of E-dimension corresponding to higher values of explained variance for the experiments which use realistically placed observations (two lower left panels of Figure 4) than for the experiment that uses randomly placed simulated observations.

A unique feature of the results for the experiments which use real observations (bottom two panels of Figure 4) is that the largest value of explained variance is about 0.9 instead of the theoretical maximum of one. This reduced maximum is most likely due to the effects of the model errors, as we cannot observe such reduction of the maximum in the two experiments that use simulated observations. We cannot determine, however, based on the information provided by our experiments, whether this reduction in the maximum of the explained variance occurs because some of the forecast errors are orthogonal to the model attractor, thus an ensemble of model forecasts cannot capture them, or because a flaw in the generation of ensemble perturbations does not allow the members of the forecast ensemble to explore that part of the model attractor that includes the true system state.

Figure 5 shows the JPDF for the explained variance and the state estimation error in the NH extratropics. As in KEA07, *the ensemble captures the patterns associated with*

larger forecast errors more efficiently. In addition, both the minimum and the maximum of explained variance increase with forecast time in all three experiments.

Figure 6 shows the mean E-dimension for the bins in the JPDF for forecast error and explained variance. The findings of KEA07 extend to the two more realistic settings: larger forecast errors lead to an on average lower E-dimension, and therefore to higher explained variance. Interestingly, the distribution of E-dimension with explained variance at analysis time is more similar for the two experiments which assimilate realistically distributed observations. For these two experiments, we find locations where the explained variance is high and the E-dimension is low, but the analysis error is relatively large. These are locations where the ensemble efficiently captures the space of uncertainties, but there are no observations available to take advantage of this information. Such locations do not exist for the experiment that assimilates randomly placed observations, as in that experiment the observational coverage is sufficient to effectively remove the background errors at locations of high explained variance (low E-dimension).

Figure 7 shows the evolution of the mean of the NH average of the sum of the squared forecast error, $ERV = \|\boldsymbol{\xi}_i\|^2$, the sum of the squared forecast error in $\mathbb{S}$, $ERV_S = \sum_{i=1}^{K} \boldsymbol{\xi}_i^{(\|)2}$, and the total variance in the ensemble $V_S = \sum_{i=1}^{K} \lambda_i$. We also show the curve $2V_S$ since the initial condition of $\mathbf{x}^{a,f}$ is defined by the ensemble mean. For a perfect ensemble, the initial value of ERV, ERV_S, and V_S would be equal, and with increasing forecast time ERV and ERV_S would drift toward $2V_S$. (This is because the asymptotic value of ERV, which is reached at the forecast time where predictability is completely lost, is twice the variance of the error in the climatological forecast, while the asymptotic value of V_S is equal to the variance of the error of the climatological forecast [e.g., Leith (1974); Szunyogh and Toth (2002)].) We

find that the ensemble variance underestimates the total forecast error in all experiments. For the perfect model experiment that assimilates realistically placed observations, at short lead times ERV_S is in good agreement with V_S. In this case, further inflating the variance to make $V_S = ERV$ would not lead to improved analyses and forecasts because part of ERV is due to errors, $\boldsymbol{\xi}^{(\perp)}$, that lie in directions orthogonal to the space spanned by ensemble perturbations. The forecast error in the space $\mathbb{S}$ better approximates the total forecast error for the two perfect model experiments. For the experiment that assimilates conventional observations we find that the variance is grossly underestimated, most likely due to the effects of model errors. When we fit the curves in Figure 7 to exponential functions, $V(t) = e^{\sigma t}$ we find that the total variance V_S grows more slowly for the experiment that uses conventional observations $\sigma = 0.0184$ than for the experiments that use randomly placed simulated observations $\sigma = 0.0224$ or realistically placed simulated observations $\sigma = 0.0283$. This slower growth rate is most likely due to the effects of nonlinearity, which play a more important role in this experiment than in the others, because the initial magnitude of the errors and the ensemble perturbations are much larger in this experiment.

For the experiment that uses conventional observations, the initial value of ERV is largest, but the growth rate is slowest, $\sigma = 0.0204$ ($\sigma = 0.0262$ for ERV_S). The forecast error growth is fastest for the perfect model experiment with randomly placed observations, $\sigma = 0.0266$ ($\sigma = 0.0337$ for ERV_S). The perfect model experiment with realistically placed observations has a forecast error growth rate of $\sigma = 0.0243$ ($\sigma = 0.0297$ for ERV_S).

c. Spectrum of the d-ratio

So far we have shown that $\mathbb{S}$ provides a representation of the space of forecast uncertainty, which becomes increasingly more efficient with increasing forecast time (Figure 5 and 6.) We have also shown that $\mathbb{S}$ is more certain to capture a larger portion of the forecast error where the error in the deterministic forecast is larger. Now we turn our attention to investigating the efficiency of the ensemble in distinguishing between the importance of the error patterns (eigen directions) in $\mathbb{S}$.

We first compute the spectrum of d-ratio d_k using the same definition of the local volume as in our calculations of E-dimension and explained variance. On average, the ensemble underestimates the error in the forecast in all directions it captures (Figure 8). The exceptions to this underestimation are the first few directions, that explain a larger portion of ensemble variance, for the perfect model experiment that uses realistic observation coverage at short lead times (in agreement with figure 7). The underestimation is typically smaller in directions that explain a larger portion of the ensemble variance. The problem is most pronounced for the experiment that assimilates observations of the real atmosphere and less serious for the perfect model experiments.

In order to obtain d-ratio figures whose meteorological (physical) meaning is easier to interpret, we change the definition of the local volume: we investigate a single variable at a single level using 5 by 5 horizontal grid points. In these calculations N=25 ($N < K$), hence, the upper bound for the dimension of $\mathbb{S}$ is 25. The variable and levels we choose for this analysis are the surface pressure, the temperature at 850 hPa, the two horizontal wind components at 500 hPa, and the geopotential height at 500 hPa.

Figure 9 shows the time mean of this ratio in the leading direction, d_1, for the temperature at 850 hPa. This figure shows that initially (at 12 hr lead time) d_1 tends to be higher in regions of high observation density, such as Western Europe and the United States, than in regions of lower observational density, such as the Southern Hemisphere and the oceanic regions (panels c and e of Figure 9). The contrast between regions of low and high d_1 ratio is most striking for the experiment that assimilates observations of the real atmosphere: the ensemble grossly underestimates the forecast uncertainty in regions of high observation density. This result is an indication that our zonally constant covariance inflation strategy in the LETKF cannot be optimal when there are zonal changes in the observation density. Thus we conjecture that implementing a spatially varying adaptive covariance inflation technique, such as described in Anderson (2007), would lead to an improvement of the analyses and the short term ensemble forecasts. The time averaged spectrum of the d-ratio for a particular grid point (Figure 10) over the ocean, at the 12-hour lead time shows that the ensemble, on average, underestimates the uncertainty in all directions in $\mathbb{S}$, most severely for directions that explain a smaller portion of the variance. For the same grid point at the 5-day lead time, the underestimation is less severe, especially for the leading directions. Figure 9 and Figure 10 together suggest that in the NH extratropics, where the observation density is high our ensembles underestimate the variance in the space, $\mathbb{S}$, that they capture. The exception is the experiment that assimilates simulated observations at the location of real observations. This suggests that the variance inflation coefficients that we obtained by tuning, monitoring changes in the analysis errors, may be significantly smaller than optimal.

5. Conclusions

In this paper, we studied the spatio-temporally changing nature of predictability in a reduced resolution version of the model component of the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS), a state-of-the-art numerical weather prediction model using the LETKF data assimilation scheme. Our conclusions fall into two categories: conclusion that point to the suboptimality of our current implementation of the LETKF analysis scheme and conclusions that we believe to identify inherent properties of the local predictability of the performance of an ensemble forecast system.

- We find that observational density has a greater impact on the structure of analysis and forecast error than does model error. Including the impact of model error has a greater influence on the magnitude of error than the structure of error.

- Independent of experiment, lead time, and geographic region, the lower the E-dimension, the more likely the explained variance is high. Further, as forecast lead time increases, smaller values of E-dimension more certainly predict high explained variance.

- In the extratropics, the ensemble does a better job of capturing forecast error when forecast error is high. This behavior can be explained by the fact that high forecast error leads to low E-dimension. We find this result to hold for both perfect model and the real atmosphere.

- Realistic observation coverage, when only conventional (non-radiance) observations are considered, is not adequate to remove errors correctly identified by the ensemble at analysis time in the extratropics, leading to the presence of regions of high explained

variance and low E-dimension at analysis time.

- Conclusions related to the suboptimality of our implementation of the LETKF:
  - At analysis time, we find that the ensemble typically underestimates uncertainty more severely in regions of high observation density than for regions of low observation density. This result indicates that implementing a spatially varying adaptive covariance inflation technique may improve analyses.
  - We find that the variance inflation coefficient used in the current implementation of the LETKF on the NCEP GFS (Szunyogh et al. 2008) may be smaller than optimal in the NH extratropics. We conjecture that even if a spatially varying adaptive covariance inflation technique is not implemented, further tuning of the current coefficients of the variance inflation scheme may lead to major improvements of the analysis in the NH extratropics.

REFERENCES

Anderson, J., 2007: Spatially-varying adaptive covariance inflation for an ensemble filter. *Submitted to Tellus A*.

Chapman, S. and R. Lindzen, 1970: *Atmospheric Tides*, 200. D. Reidel Press, Dordrecht, Holland.

Hamill, T. M. and J. Whitaker, 2007: Ensemble calibration of 500 hPa geopotential height and 850 hPa and 2-meter temperatures using reforecasts. *Mon. Wea. Rev.*, **135**, 3273–3280.

Huang, X.-Y. and P. Lynch, 1993: Diabatic digital filtering initialization: Application to the HIRLAM Model. *Mon. Weather Rev.*, **121**, 589–603.

Hunt, B., E. Kostelich, and I. Szunyogh, 2006: Efficient data assimilation for spatio-temporal chaos: A local ensemble transform Kalman filter. *Physica D*, submitted.

Kalnay, E., 2002: *Atmospheric Modelling, Data Assimilation and Predictability*, 364. Cambridge University Press, Cambridge.

Kuhl, D., et al., 2007: Assessing Predictability with a Local Ensemble Kalman Filter. *J. Atmos. Sci.*, **64**, 1116–1140.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Weather Rev.*, **102**, 409–418.

Oczkowski, M., I. Szunyogh, and D. J. Patil, 2005: Mechanisms for the development of locally low dimensional atmospheric dynamics. *J. Atmos. Sci.*, 1135–1156.

Ott, E., et al., 2002: A Local Ensemble Kalman Filter for Atmospheric Data Assimilation. *arXiv.org:physics/0203058*.

Palmer, T. and R. Hagedorn, (Eds.), 2006: *Predictability of Weather and Climate*, 718. Cambridge University Press, Cambridge.

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

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

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Ott, E. Kalnay, and J. A. York, 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus*.

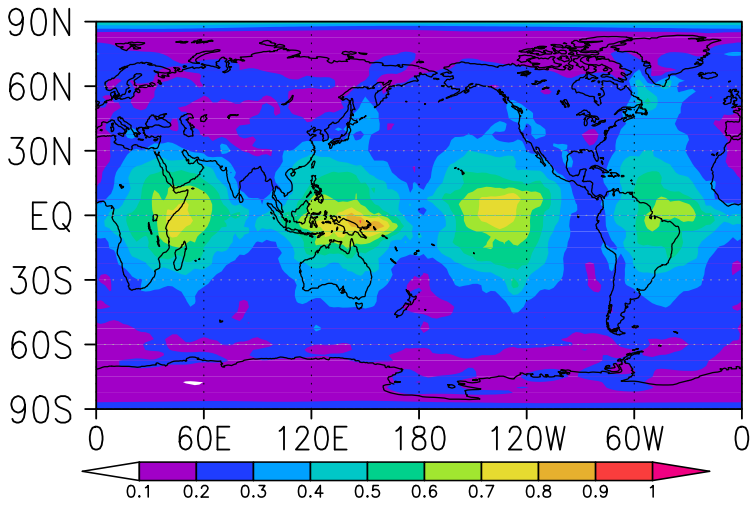
Szunyogh, I. and Z. Toth, 2002: The Effect of Increased Horizontal Resolution on the NCEP Global Ensemble Mean Forecasts. *Mon. Weather Rev.*, **130**, 11251143.

Szunyogh, I., et al., 2007: The Local Ensemble Transform Kalman Filter and its implementation on the NCEP global model at the University of Maryland. *Workshop Proceedings, Flow Dependent Aspects of Data Assimilation, June 11-13, 2007*, ECMWF, 47–63.

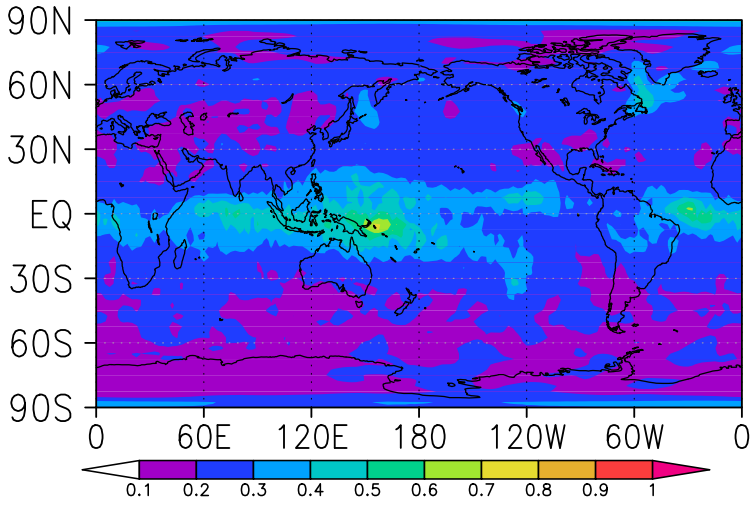
Toth, Z., O. Talagrand, and Y. Zhu, 2006: The attributes of forecast systems: a general

framework for the evaluation and calibration of weather forecasts. *In Palmer and Hagedorn, 2006*, 584–595.

List of Figures



(a)



(b)

FIG. 1. The time mean absolute error of the surface pressure for simulated observations at random locations for analyses generated with digital filter initialization (top) and without digital filter initialization (bottom). The average is taken over all analyses between 01 January 2004 0000UTC and 29 February 2004 1800UTC

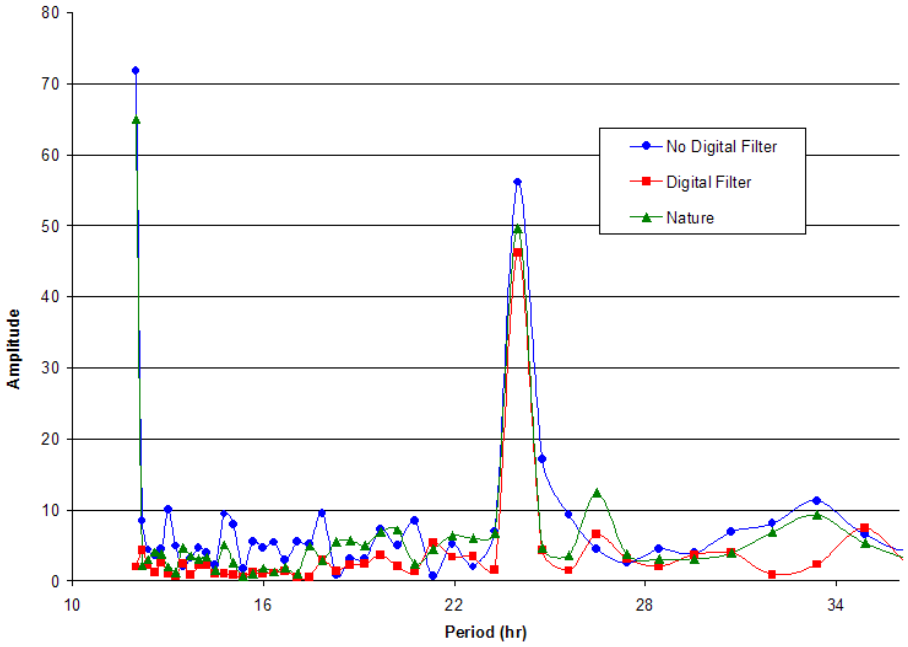


FIG. 2. The Fourier analysis at the location 0°S, 160°W between 14 January 2004 0600UTC and 15 February 2004 0000UTC shown for simulated observations at random locations without digital filter initialization, with digital filter initialization, and for the nature run

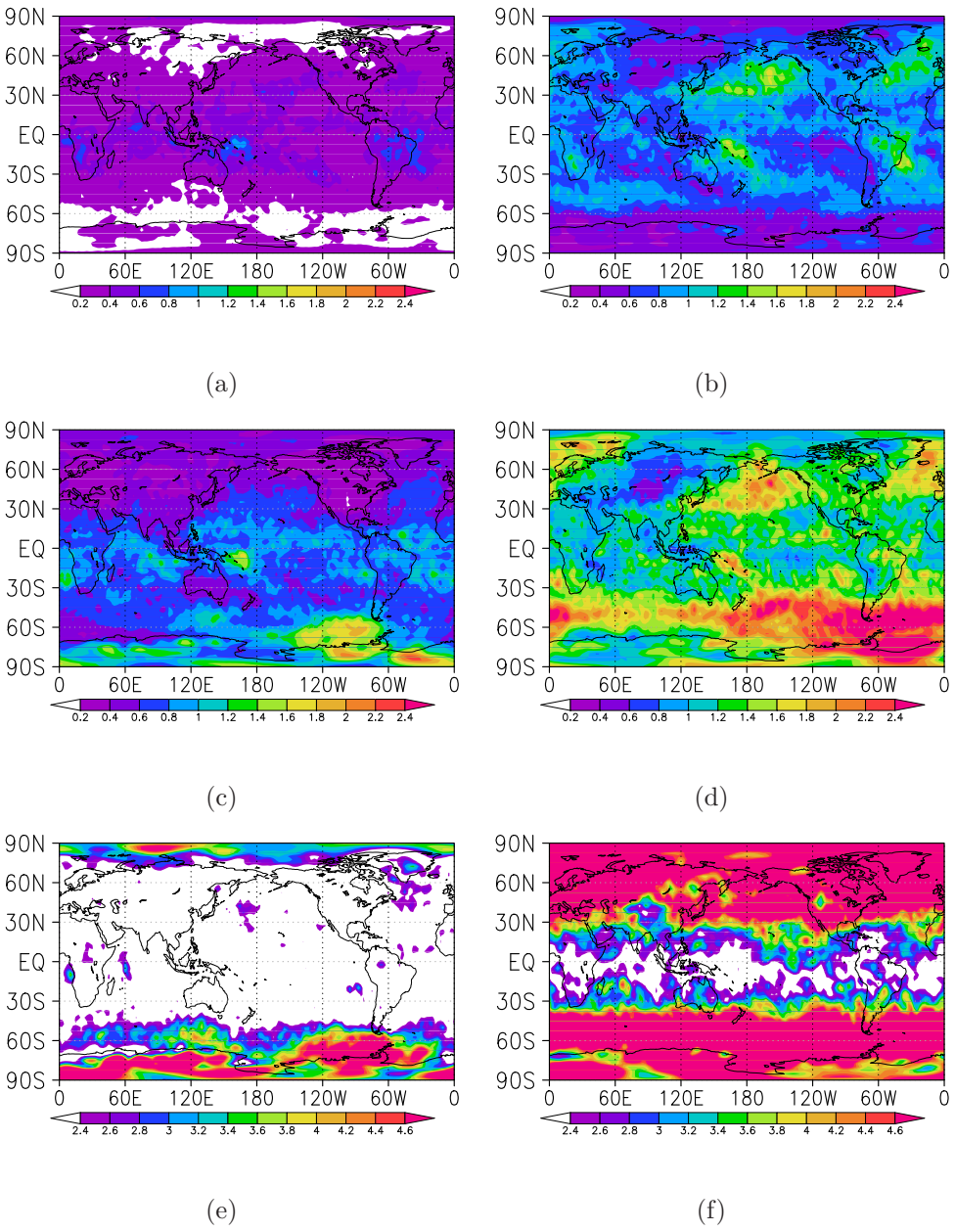


FIG. 3. Time-mean absolute analysis/forecast error of the meridional wind component at the 500 hPa pressure level. Results are shown for the analysis (left) and the 72-hour forecast (right) for experiments that assimilate randomly distributed simulated observations (top panel), simulated observations at the locations of conventional observations (middle panel), and conventional observations of the real atmosphere (bottom panel). The average is taken over all forecasts started between 11 January 2004 0000UTC and 15 February 2004 0000UTC. Note the different scale for the forecast errors in the bottom panels.

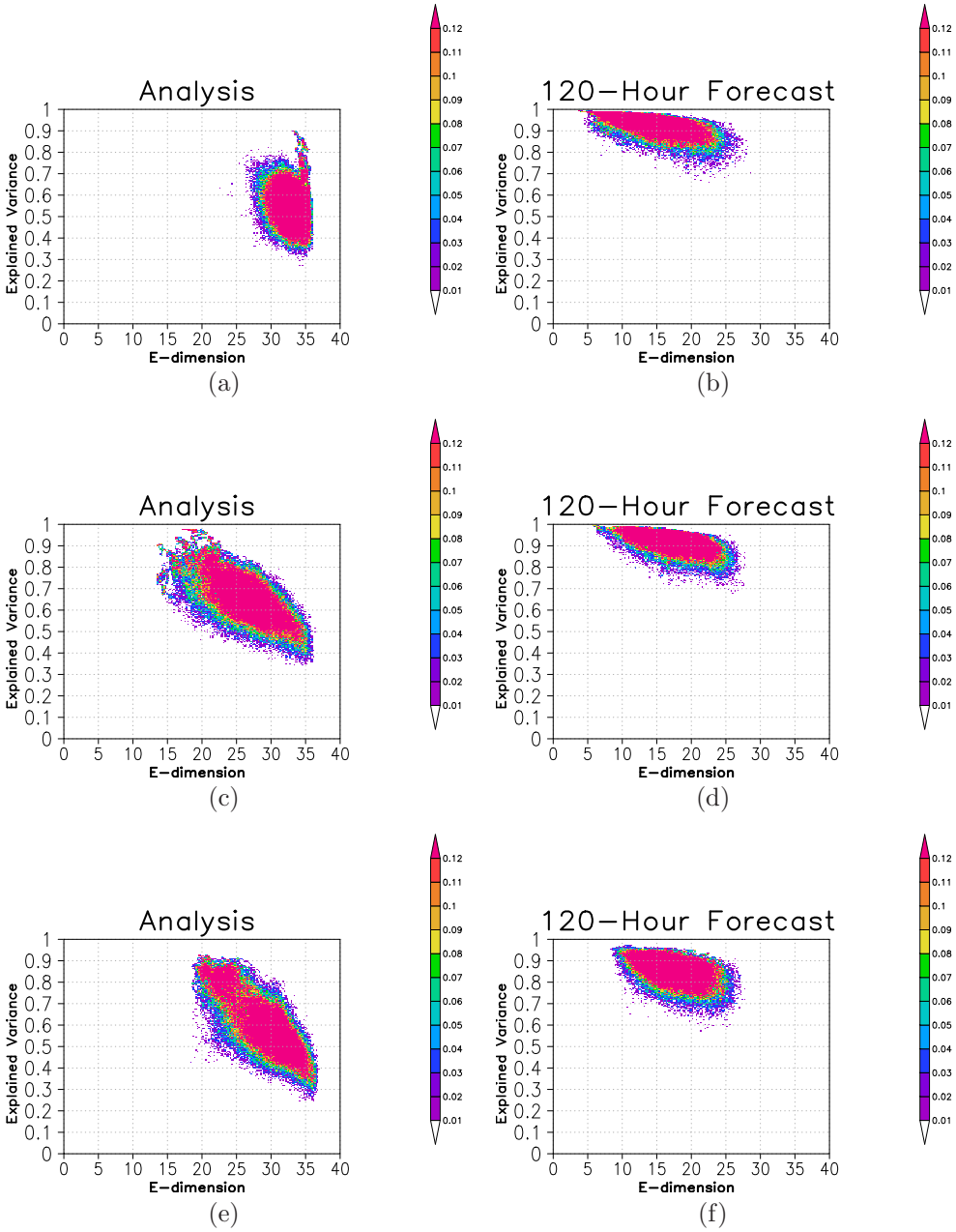


FIG. 4. Joint probability distribution of the E-dimension and the explained variance in the NH extratropics. The bin increments are 0.005 for the explained variance and 0.2 for the E-dimension. Shown are the distributions for the analysis (left) and the 5-day forecast lead time (right) for experiments that assimilate randomly distributed simulated observations (top panel), simulated observations at the locations of conventional observations (middle panel), and conventional observations of the real atmosphere (bottom panel).

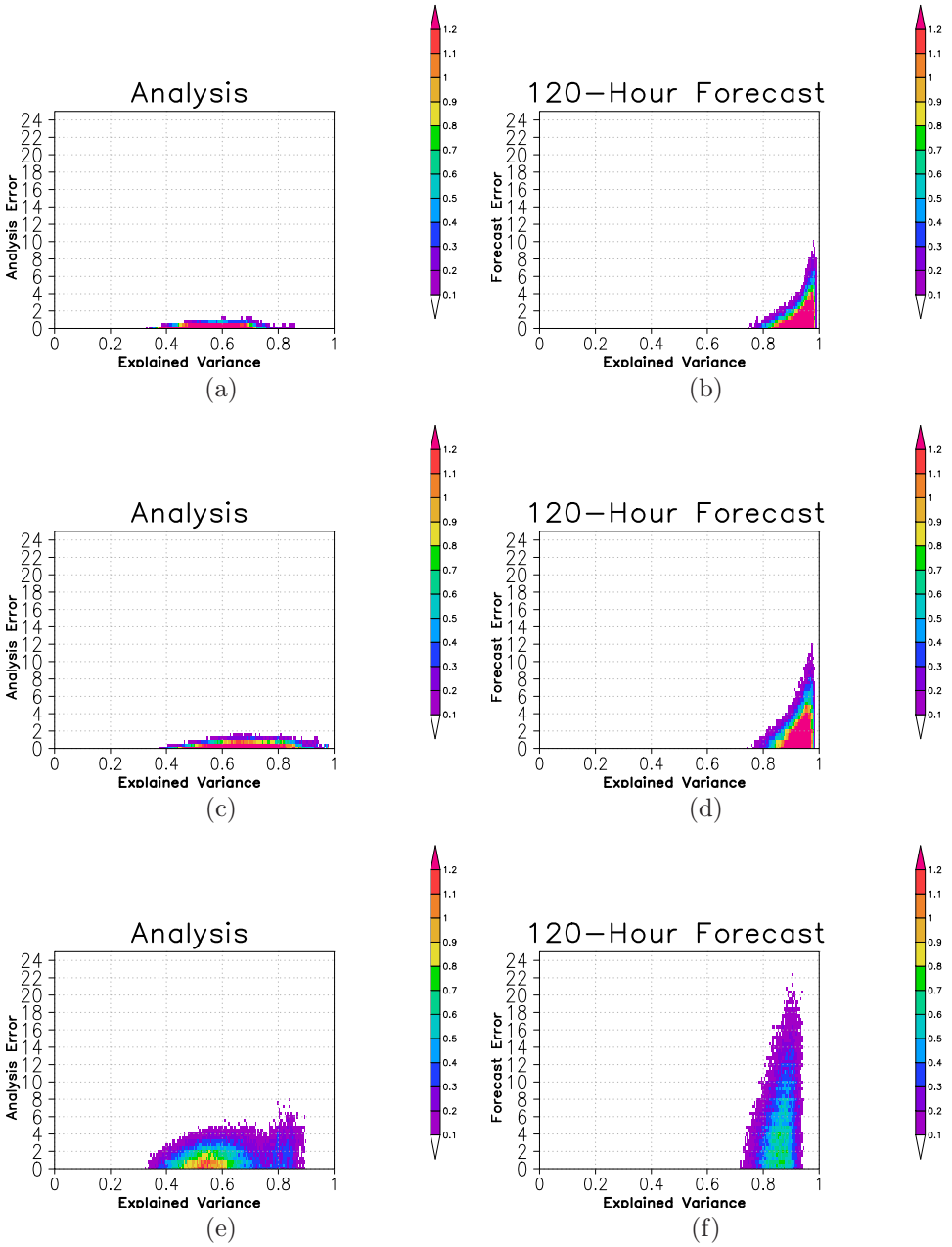


FIG. 5. Joint probability distribution of the analysis/forecast errors and the explained variance. The bin increments are 0.005 for the explained variance and 0.4 for the forecast error. Shown are the distributions for experiments that assimilate randomly distributed simulated observations (top panel), simulated observations at the locations of conventional observations (middle panel), and conventional observations of the real atmosphere (bottom panel). Note the different scale for the forecast errors in the bottom panel.

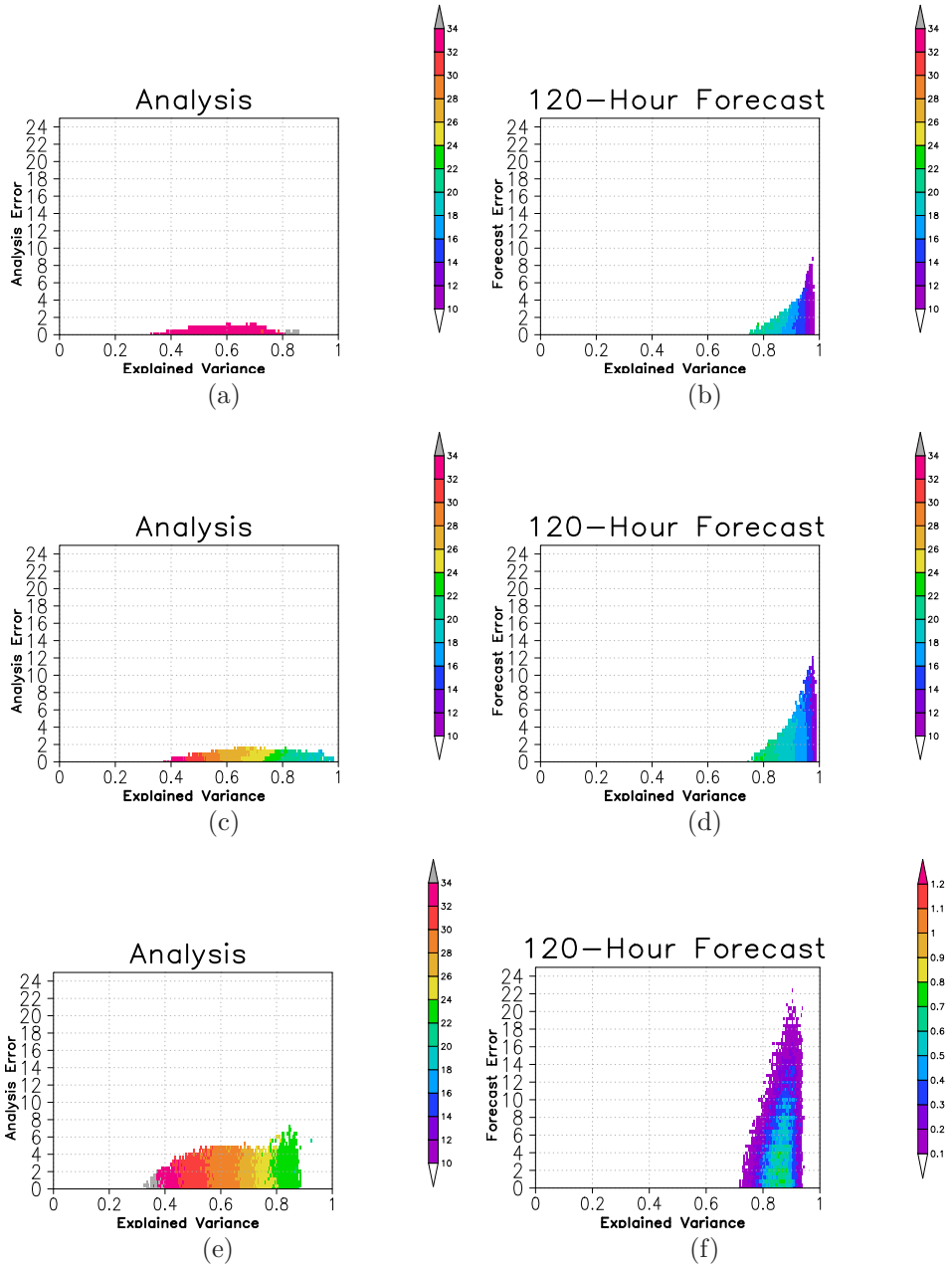


FIG. 6. Color shades indicate the mean E-dimension for each nonempty bin in Figure 5. Shown are the distributions for experiments that assimilate randomly distributed simulated observations (top panel), simulated observations at the locations of conventional observations (middle panel), and conventional observations of the real atmosphere (bottom panel). Note the different scale for the forecast errors in the bottom panel.

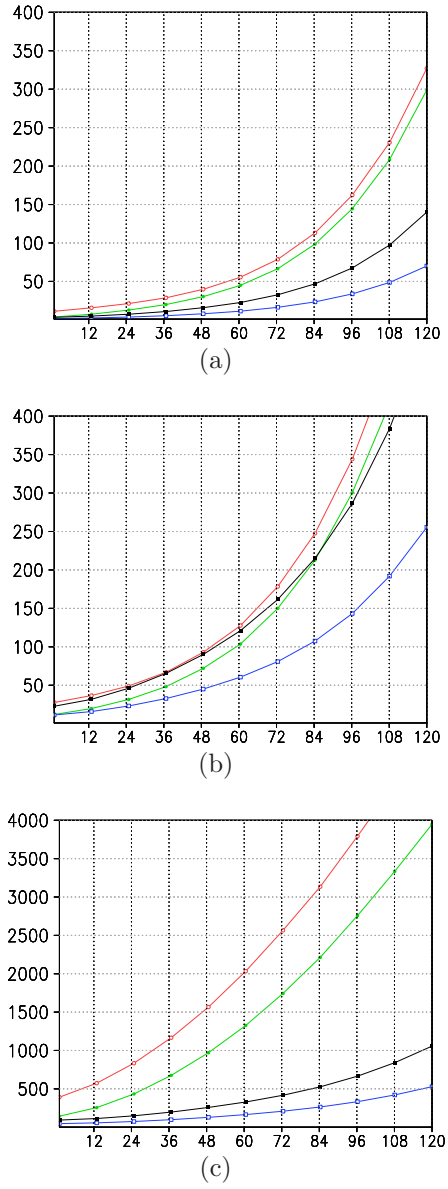


FIG. 7. The mean of the Northern Hemisphere average of ERV (red), ERV_S (green), V_S (blue), and $2V_S$ (black) calculated for all assimilated variables in the local regions with energy rescaling. Results are shown for each forecast lead time the for experiments that assimilate randomly distributed simulated observations (top panel), simulated observations at the locations of conventional observations (middle panel), and conventional observations of the real atmosphere (bottom panel). The average is taken over all forecasts started between 11 January 2004 0000UTC and 15 February 2004 0000UTC. Note the different scale for the forecast errors in the bottom panel.

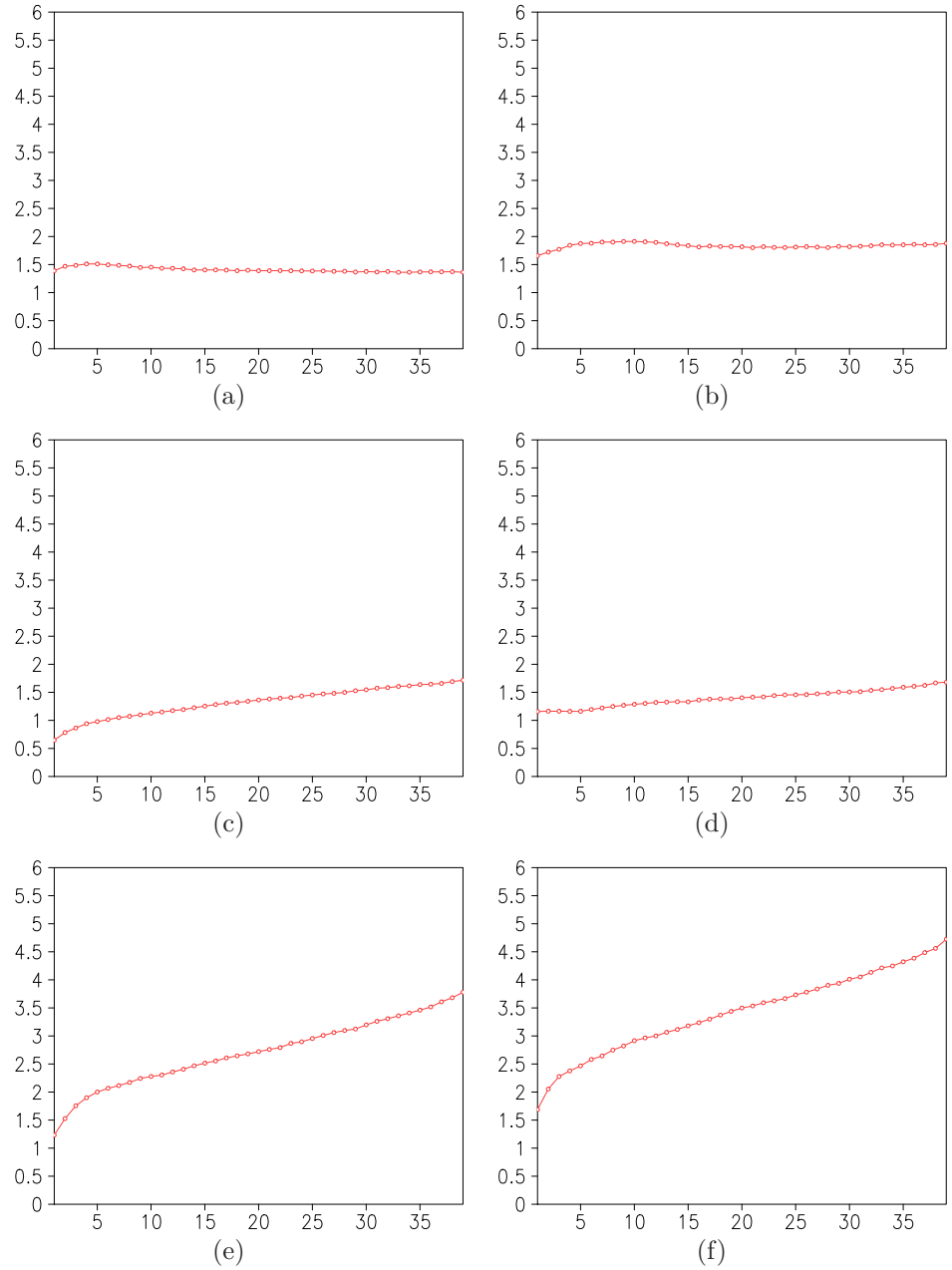


FIG. 8. The time mean of the Northern Hemisphere average spectrum of the ratio d_i, calculated for all assimilated variables in the local regions with energy rescaling. Results are shown for 12-hour lead time (right) and the 5-day lead time (left) for experiments that assimilate randomly distributed simulated observations (top panels), simulated observations at the locations of conventional observations (middle panels), and conventional observations of the real atmosphere (bottom panels). The average is taken over all forecasts started between 11 January 2004 0000UTC and 15 February 2004 0000UTC.

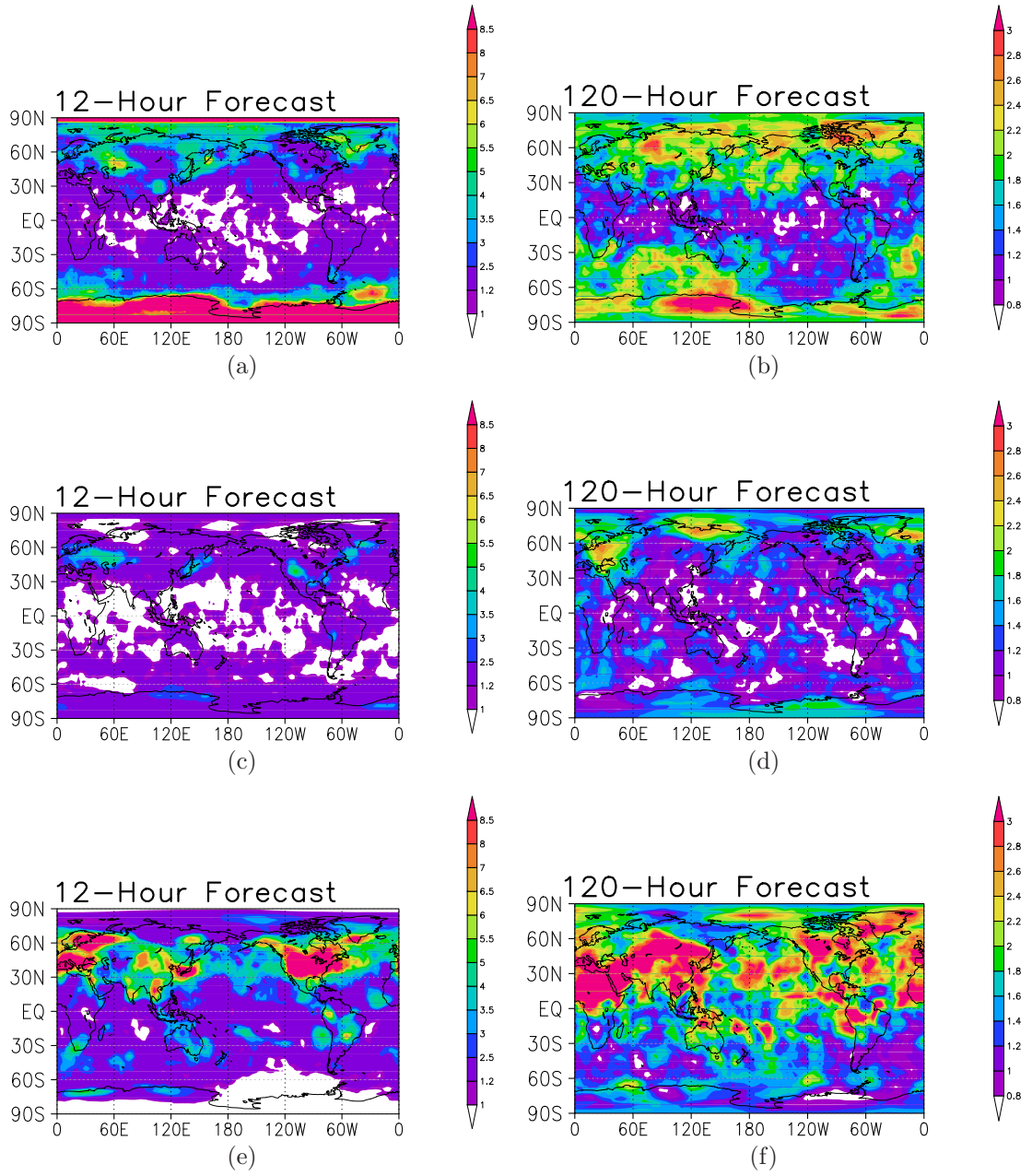


FIG. 9. The time average of the ratio d_i in the leading direction for the temperature at 850 hPa. Results are shown for the 12-hour forecast (left) and the 5-day forecast (right) for experiments that assimilate randomly distributed simulated observations (top panel), simulated observations at the locations of conventional observations (middle panel), and conventional observations of the real atmosphere (bottom panel). The average is taken over all forecasts started between 11 January 2004 0000UTC and 15 February 2004 0000UTC. Note the different scale for the d-ratio for the 5-day lead time.

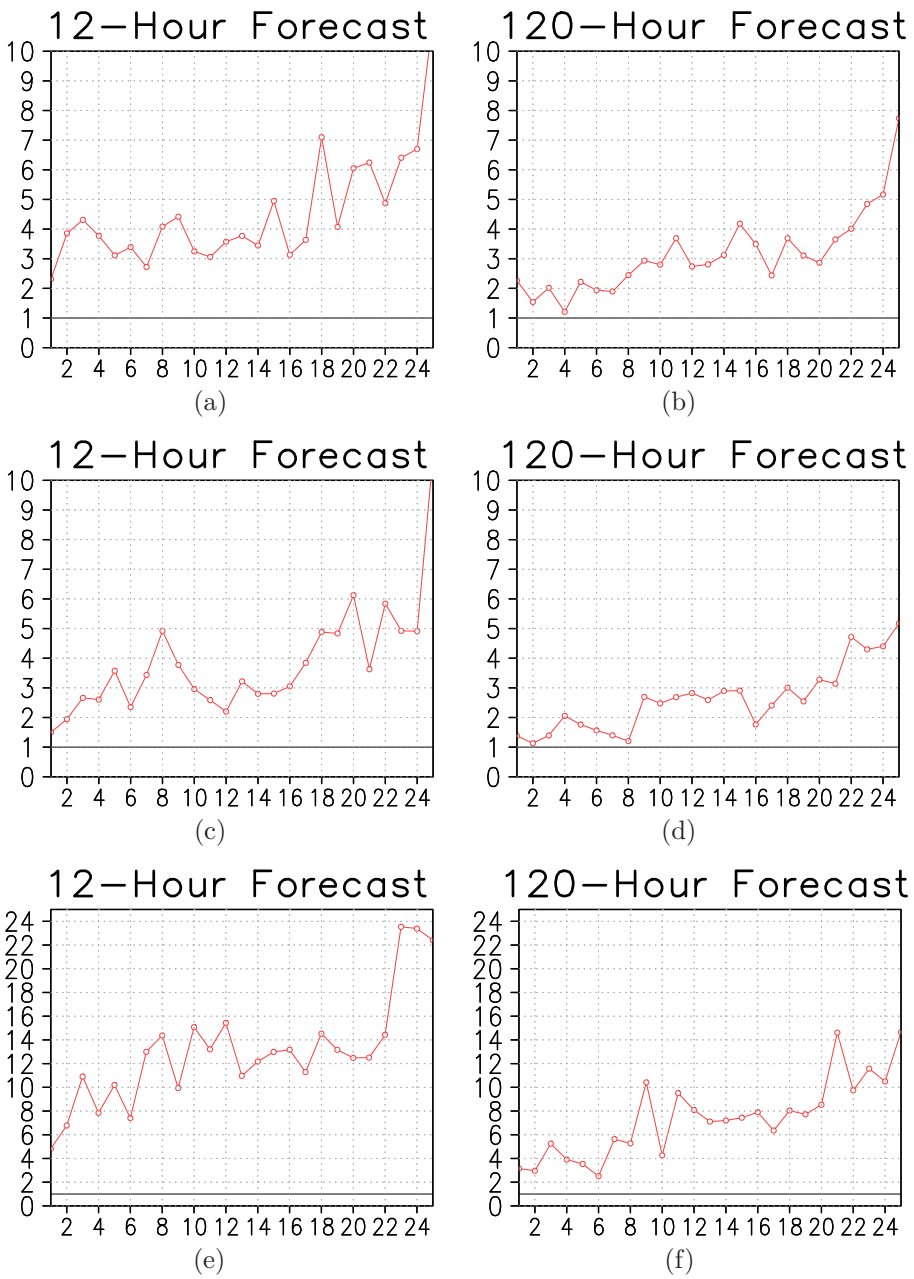


FIG. 10. The spectrum of the ratio d_i at the point 60°N 120°W for the temperature at 850 hPa. Results are shown for the 12-hour forecast (left) and the 5-day forecast (right) for experiments that assimilate randomly distributed simulated observations (top panel), simulated observations at the locations of conventional observations (middle panel), and conventional observations of the real atmosphere (bottom panel). The average is taken over all forecasts started between 11 January 2004 0000UTC and 15 February 2004 0000UTC. Note the different scale for the d-ratio in the bottom panel.