1.3 ASSESSING PREDICTABILITY OF ATMOSPHERIC PREDICTABILITY WITH AN ENSEMBLE KALMAN FILTER

Elizabeth Satterfield* and Istvan Szunyogh
University of Maryland, College Park, Maryland

1. INTRODUCTION

In this paper, the spatio-temporally changing nature of predictability is studied 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. Uncertain initial conditions (analyses) are obtained by assimilating noisy simulated observations of the hypothetical "true" state and observations of the real atmosphere with the Local Ensemble Transform Kalman Filter (LETKF) data assimilation system of the University of Maryland. This data assimilation system also provides initial conditions for an ensemble of forecasts. Predictability of atmospheric predictability is assessed by investigating the dependence of the performance of the ensemble forecast system on the atmospheric flow in capturing the forecast uncertainties. We show that the larger the forecast error, the more certain that the ensemble can fully capture the space in which forecast errors evolve. We explain this behavior using the E-dimension, a diagnostic that was developed at the University of Maryland (Patil et al. 2001).

The E-dimension characterizes the local complexity of dynamics in a high dimensional prediction model. Patil et al. (2001) found that the lowest dimensional regions are often the regions of largest forecast uncertainty. For these low dimensional regions, prediction for short and mid range forecasts could be improved by improving prediction in a few phase space directions (Oczkowski et al. 2005). Szunyogh et al. (2005) showed that for lower values of E-dimension, the ensemble more certainly captures the error in the forecasts that serve as the background (first guess) for the analysis. Kuhl et al. (2007) found that, in the extratropics, fast error growth always leads to low E-dimension and therefore, to increased certainty that a greater portion of the forecast error is captured by the ensemble. This study assumed a perfect model and used the Local Ensemble Kalman Filter (LEKF) scheme to assimilate randomly distributed simulated observations that provided a 10% observational coverage of the model grid points. Thus the geographical distribution of forecast errors reflected the spatio-temporally changing nature of predictability in the model and not inhomogeneities of the observational coverage. In this paper, we build on the findings of Kuhl et al. (2007) by gradually adding realistic features to the observing network. First, we account for the effects of an inhomogeneous observing system and then we add the impact of model errors by assimilating observations of the real atmosphere. In what follows, uncertain initial conditions (analyses) are obtained by assimilating noisy simulated observations of the hypothetical "true" state and observations of the real atmosphere using the Local Ensemble Transform Kalman Filter (LETKF).

The first formulation of the LETKF, the Local Ensemble Kalman Filter (LEKF) was proposed by Ott et al. (2002, 2004). The LEKF solves the Kalman filter equations locally in model grid space. An analysis is obtained independently for each point, by considering the observations and background state from a local region centered at the given point. For each point, the LEKF scheme provides an estimate of analysis uncertainty and generates an ensemble of analysis perturbations representative of the uncertainty at the given point. The LETKF introduced by Hunt et al. (2007) includes changes to improve computational efficiency of the algorithm and to add flexibility when non-local observations are assimilated (Szunyogh et al. 2007). In this study, similar to our previous work (Szunyogh et al. 2008, and Kuhl et al. 2007) we carry out experiments with an implementation of the LETKF algorithm 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).

2. EXPERIMENT DESIGN

In the first step of our experiments, we assimilate simulated observations of the hypothetical “true” atmospheric state at the locations of conventional 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 surface pressure, temperature, and two horizontal components of the wind vector. The “true” states are generated by an integration of the GFS model at T62L28 resolution starting from an operational NCEP analysis truncated to T62L28 resolution. The location and type of observations is obtained from a database that includes all observations operationally assimilated at NCEP between 000UTC 1 January 2004 – 000UTC 15 February 2004, with the exception of satellite radiances, but including satellite derived winds. We also excluded surface observations, except for the surface pressure. In the second step, the observations

* *Corresponding author address*: Elizabeth A. Satterfield, Univ. of Maryland, Dept. of Atmospheric and Oceanic Science, College Park, MD 20742, email: esatterf@atmos.umd.edu

of the real atmosphere, which were used to obtain the type and location for the simulated observations, were assimilated.

For each data set, a state estimate is obtained at the native model resolution every 6hrs. Diagnostics are computed at a reduced 2.5º x 2.5º 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 800km horizontal radius of the grid point, where the state is estimated.
- Observations have equal weight within a 500km radius of the given grid point, beyond which the weight of the observations tapers linearly to zero at 800km.
- 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 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%-30% in the NH extratropics, and changes smoothly in the tropics (between 25ºS and 25ºN) from the values of the SH extratropics to the values of the NH extratropics
- For both 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.

Deterministic forecasts are started from the ensemble mean twice daily and output every 12 hours. In addition to the state estimate, this analysis scheme also generates an ensemble of analysis perturbations that represent the estimated uncertainty in the state estimate. These analysis perturbations serve as initial conditions for an ensemble of forecasts. Ensemble forecasts are started at 0000UTC daily and output every 12 hours. Both the deterministic forecast and the ensemble forecasts are carried out to a two week lead time.

3. FORECASTS ERRORS

Forecast error statistics are computed by comparing the deterministic forecasts to the “true” states. For the forecasts started from analyses generated by assimilating simulated observations in realistic locations, the “truth” is taken to be an integration of the GFS model starting from the operational NCEP analysis at 0000 UTC 1 January 2004. Forecasts started from analyses generated by assimilating conventional data are verified using high resolution operational NCEP analyses truncated to 2.5º x 2.5º resolution. The NCEP analyses were generated using Spectral Statistical Interpolation (SSI) with the version of the GFS model at T62L28 resolution and the observational data set used in our LETKF experiments. Forecast error statistics are generated for the 36-day period, 0000 UTC 11 January 2004 – 0000 UTC 15 February 2004.

First, we compare the absolute error in the meridional wind component at 500 hPa for each forecast lead time. Plots of the absolute error are obtained by plotting the time averaged distance between the forecast and the “true” states at each grid point. Figure 1 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 (Kuhl et al. 2007) show that the largest analysis errors are in the Tropics and smallest analysis errors are found in regions of mid-latitude storm tracks, 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 analysis errors which tend to reflect observation density. The lowest errors are found over continents in the Northern Hemisphere and highest errors over Antarctica and in the oceanic region between Cape Horn and the Antarctic Peninsula. As forecast lead time increases, we see error growth in the poles and extratropics. We see strong similarities in the structure of errors at analysis time and for short term forecasts in both experiments that assimilate observations in realistic locations. These results indicate that observation density plays a more important role than model error in determining the spatial distribution of short term forecast errors, in agreement with Szunyogh et al. (2007). The results obtained by assimilating conventional 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 72hr lead time.

4. E-DIMENSION, EXPLAINED VARIANCE, AND FORECAST ERROR

In order to explore the performance of the ensemble in capturing the forecast uncertainty, the explained variance diagnostic is used. Explained variance is the fraction of forecast error that is contained in the space spanned by the ensemble, after projecting the forecast

error on ensemble space. Formally, it is calculated by taking the square of the projection of forecast error on ensemble space and then normalizing by the square of the forecast error (Kuhl et al. 2007). Explained variance has a maximum value of one, when the forecast error projects entirely onto the space spanned by the ensemble, and a minimum value of zero, when the forecast error falls completely outside of the ensemble space.

E-dimension is a spatiotemporally evolving measure of the effective number of spatial degrees of freedom in a set of ensemble perturbations (Oczkowski et al. 2005). In other words, the E-dimension measures the effective dimension of the space spanned by a set of ensemble perturbations (Oczkowski et al. 2005). To calculate the E-dimension statistic, we follow the method presented in Oczkowski et al. (2005) and Kuhl et al. (2007), and transform the ensemble perturbations such that the square of the Euclidean norm of the transformed perturbations has dimensions of energy. We then define k local vectors consisting of ensemble perturbations for each dynamic variable in a local region. Here, the local region is defined, as in Kuhl et al. (2007), by all grid point variables in a 5x5 horizontal grid (at 2.5º x 2.5º grid resolution) and the entire column of the model atmosphere. A (Vxk) matrix **B**, whose columns consist of k local vectors, is then formed (V is the number of grid point variables in each region). A singular value decomposition is performed on that matrix. The E-dimension statistic is computed as a statistic on the singular values (Oczkowski et al. 2005). Numerically, the formula for E-dimension is:

$$\text{E-dimension} = \frac{\left(\sum_{i=1}^{k} \sigma_i\right)^2}{\sum_{i=1}^{k} \sigma_i^2}, \qquad (1)$$

Where σ_i, i=1,…,k are the singular values of the matrix **B** and $\sigma_i = \sqrt{\lambda_i}$, where λ_i, i=1…k are nonnegative eigen values of the covariance matrix, $\mathbf{C}=\mathbf{B}^T\mathbf{B}$. A low value of E-dimension would mean that the uncertainty captured by the ensemble is contained within only a few independently varying spatial patterns. The maximum value of E-dimension, which is equal to the number of ensemble perturbations, would occur if the uncertainty was evenly distributed in independently varying spatial patterns.

All three experiments show a similar geographic structure and evolution of E-dimension (Figure 2). E-dimension is higher in the Tropics than in the extratropics and decreases with forecast time throughout the entire 5-day forecast period. The decrease is most rapid in the storm track region in all three experiments. Not surprisingly, the results for the two experiments which use realistically placed observations are more similar to each other than to the results from the randomly placed simulated observations and show a lower E-dimension at analysis time. As for forecast errors, observation density has a greater impact on the structure and evolution of E-dimension than does model error.

E-dimension and explained variance have a strong negative correlation (Szunyogh et al. 2005). The lower the E-dimension, the greater the probability that explained variance is high (Kuhl et al. 2007). This relationship is illustrated by the joint probability distribution function (JPDF) shown in Figure 3. The JPDF 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, 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. We find that, independent of experiment, lead time and geographic region, the lower the E-dimension, the more likely that explained variance is high. As forecast lead time increases, low values of E-dimension have a greater probability of corresponding to high value of explained variance. For the experiments which use realistically placed observations, we find lower values of E-dimension at analysis time, which correspond to higher values of explained variance. In addition, the values of E-dimension and explained variance have a slower evolution with lead time than for the experiment that uses randomly placed simulated observations. The main difference for the experiments which use real observations is that values of explained variance never reach their maximum of 1 and instead saturate around 0.9.

The relationship between explained variance, E-dimension and forecast error is illustrated in Figures 4 and 5. Figure 4 shows the joint probability distribution function for forecast error and explained variance for the 500 hPa meridional wind component in the Northern Hemisphere extratropics. The top panels show results obtained by assimilating randomly distributed simulated observations, the middle panels are for simulated observations in realistic locations, and the bottom panels are for conventional observations of the real atmosphere. Two panels are shown for each experiment: the left panel shows results at analysis time and the right panel shows results at 5-day forecast lead time. When forecast error is large, we expect, based on earlier results, the ensemble to do a better job of capturing forecast error. Both experiments that use realistically placed observations show higher lower and upper bounds of explained variance at analysis time than the third experiment that uses randomly placed observations. For all experiments, the values of explained variance increase with forecast time, although more slowly for the experiment which uses observations of the real atmosphere. Again, for the experiment which

uses real observations we find a lower saturation point for explained variance.

The relationship between explained variance and forecast error is explained by Figure 5, which shows the mean E-dimension for the bins in the JPDF for forecast error and explained variance. Large forecast error leads to a greater likelihood of low E-dimension, and therefore to higher explained variance. In the extratropics, we expect the ensemble to do a better job of capturing the space of uncertainty when forecast error is large. Interestingly, we see that the distribution of E-dimension with explained variance at analysis time is more similar for the experiments which assimilate realistically distributed observations, even though forecast and analysis error is almost twice as large for the experiment that assimilated observations of the real atmosphere.

5. THE SPREAD-SKILL RELATIONSHIP

The relationship between ensemble spread, which is measured by the ensemble standard deviation, and skill, measured by the absolute error of the ensemble mean forecast, is often used to evaluate ensemble performance. The ability of the ensemble spread to predict mean forecast error is typically quantified by linear correlation (Whitaker and Loughe 1998). It is widely believed that for a well designed ensemble prediction system this correlation should be high. However, even in "perfect model" experiments, the correlation between spread and skill is disappointingly small (Whitaker and Loughe 1998). If the error field is dominated by random errors, the expected correlation is close to zero (Szunyogh et al 2007).

To explore the relationship between the ensemble spread and the skill of the ensemble mean forecast, we compute the Pearson correlation coefficient:

$$R(x,y) = \frac{C(x,y)}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{(\sum_{i=1}^{n}(x_i-\bar{x})(\sqrt{(\sum_{i=1}^{n}(y_i-\bar{y})} \quad (2)$$

Here, n is equal to the number of spatial points in each region multiplied by the number of points in time. Figure 6 shows the correlation for ensemble spread and the ensemble mean forecast error calculated for the meridional wind component at 500hPa in three geographical regions. The top panel shows the results obtained by assimilating randomly distributed simulated observations, the middle panel is for simulated observations in realistic locations, and the bottom panel is for conventional observations of the real atmosphere. The top panel of Figure 6 shows analysis time correlations near zero for the extratropics in both hemispheres. Here, 10% observational coverage is sufficient to remove errors with well defined structures. As errors grow, well defined structures develop, increasing the correlation. The correlation reaches a maximum of 0.46 around the 96 hour lead time in the Northern Hemisphere extratropics. In the Tropics and Southern Hemisphere extratropics, the errors continue to grow throughout the 120 hour period. In the Tropics, there is a high correlation at analysis time, due to the insufficient data coverage to suppress the small scale errors in that region. When simulated observations are placed in the locations of conventional observations, the correlation is much higher in the extratropics at analysis time, indicating that data coverage is insufficient to suppress errors correctly identified by the ensemble. The correlation at analysis time is the highest for the Southern Hemisphere extratropics, where data is sparse. The correlation in the Northern Hemisphere extratropics continues to grow throughout the 120 hour period (maximum is reached at 144 hour lead time). For conventional observations of the real atmosphere, the main difference is lower correlations at longer lead times, which is most likely due to the impact of model errors.

6. DISCUSSION

In this paper, we study 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 experiment design addresses the issues of determining the influence of observational density and model error on predictability.

- 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 initially high correlations between ensemble spread and skill. In contrast, a 10% randomly distributed observational coverage (Kuhl et al. 2007) shows initial correlations in the extratropics that are close to zero.
- The main impact of model error on the correlation between ensemble spread and skill is lower correlation at longer lead times.

Finally, we note that there is a discrepancy between the time periods used in this paper and in the results presented in Kuhl et al (2007). The data used in this study come from the winter season of 2004, whereas the data in Kuhl et al. (2007) come from the winter season of 2000. In order to make a more accurate comparison, we plan to repeat the experiment described in Kuhl et al. (2007) using data from 2004.

REFERENCES:

Houtekamer, P. L., 1993: Global and local skill forecasts. *Mon. Wea. Rev.,* 121, 1834–1846.

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

Kuhl, D., I. Szunyogh, E.J. Kostelich, G. Gyarmati, D.J. Patil, M. Oczkowski, B.R. Hunt, E. Kalnay, E. Ott, and J.A. Yorke, 2007: Assessing Predictability with a Local Ensemble Kalman Filter. *J. Atmos. Sci.*, 64, 1116–1140.

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., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, J. A. Yorke, 2004: A local ensemble Kalman Filter for atmospheric data assimilation. *Tellus* 56A , 415-428.

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.

Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems, Proceedings of a Workshop held at ECMWF on Predictability, 20-22 October 1997, Reading, United Kingdom: European Centre for Medium-Range Weather Forecasts, 1-25.

Szunyogh, I., E. Kostelich, G. Gyarmati, D. Patil, B. Hunt, E. Kalnay, E. Ott, and J. Yorke, 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus*, 57A, 528–545.

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

Szunyogh, I., E. A. Satterfield, J. A. Aravequia, E. J. Fertig, G. Gyarmati, E. Kalnay, B. R. Hunt, E. J. Kostelich, D. D. Kuhl, E. Ott, and J. A. Yorke, 2007: The Local Ensemble Transform Kalman Filter and it implementation on the NCEP global model at the University of Maryland. Workshop Proceedings, Flow Dependent Aspects of Data Assimilation, June 11-13, 2007, 47-63.

Whitaker, J.S., and A.F. Loughe, 1998: The Relationship between Ensemble Spread and Ensemble Mean Skill. *Mon. Wea. Rev.*, 126, 3292–3302.

Wilks, D. S., 2006. Statistical Methods in the Atmospheric Sciences, 2nd ed. Academic Press, Burlington, MA

Zimin, A.V., I. Szunyogh, D.J. Patil, B.R. Hunt, and E. Ott, 2003: Extracting Envelopes of Rossby Wave Packets. *Mon. Wea. Rev.*, 131, 1011–1017.

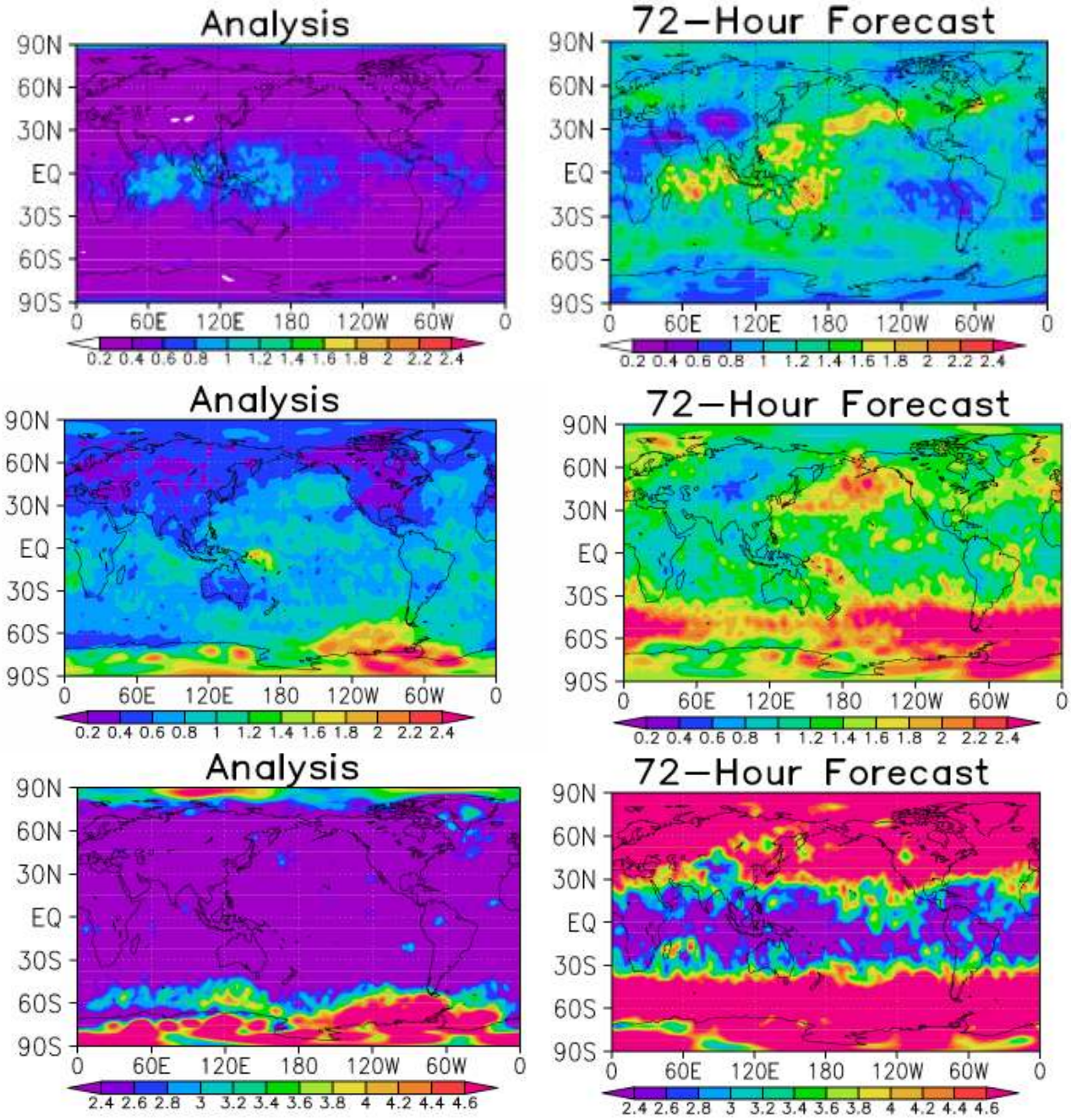

Figure 1: Time-mean absolute analysis/forecast error of the meridional wind component at the 500 hPa pressure level. Results are shown for the analysis (left) and the 72-hour forecast (right) for experiments that assimilate randomly distributed simulated observations (top panel), simulated observations 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.

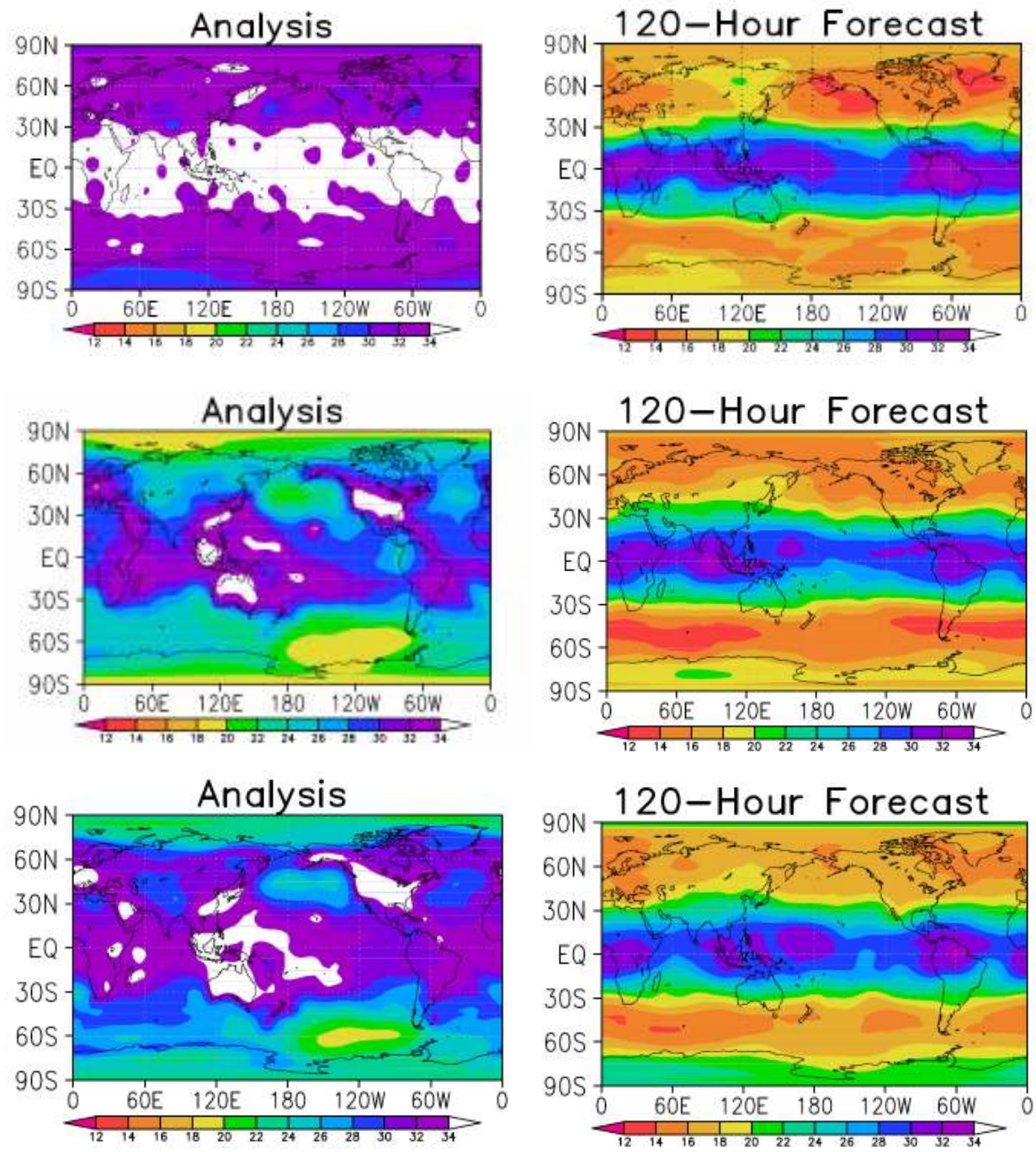

Figure 2: Time-mean E-dimension. Results are shown for the analysis (left) and the 120-hour forecast lead time (right) for experiments that assimilate randomly distributed simulated observations (top panel), simulated observations at the locations of conventional observations (middle panel), and conventional observations of the real atmosphere (bottom panel). The average is taken over all forecasts started between 11 January 2004 0000UTC and 15 February 2004 0000UTC.

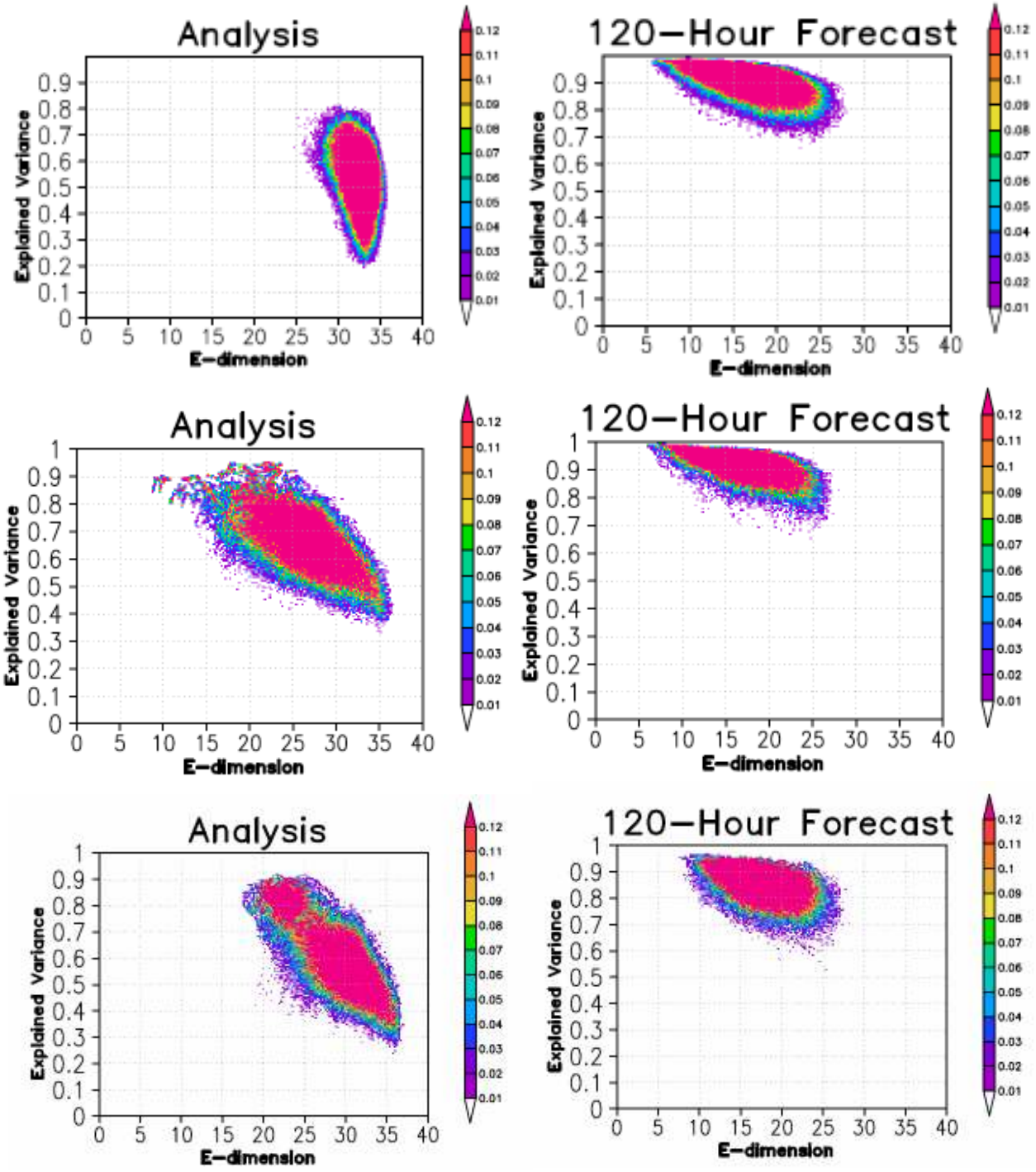

Figure 3: Joint probability distribution of the E-dimension and the explained variance in the NH extratropics. The bin increments are 0.005 for the explained variance and 0.2 for the E-dimension. Shown are the distributions for the analysis (left) and the 120-hour forecast lead time (right) for experiments 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).

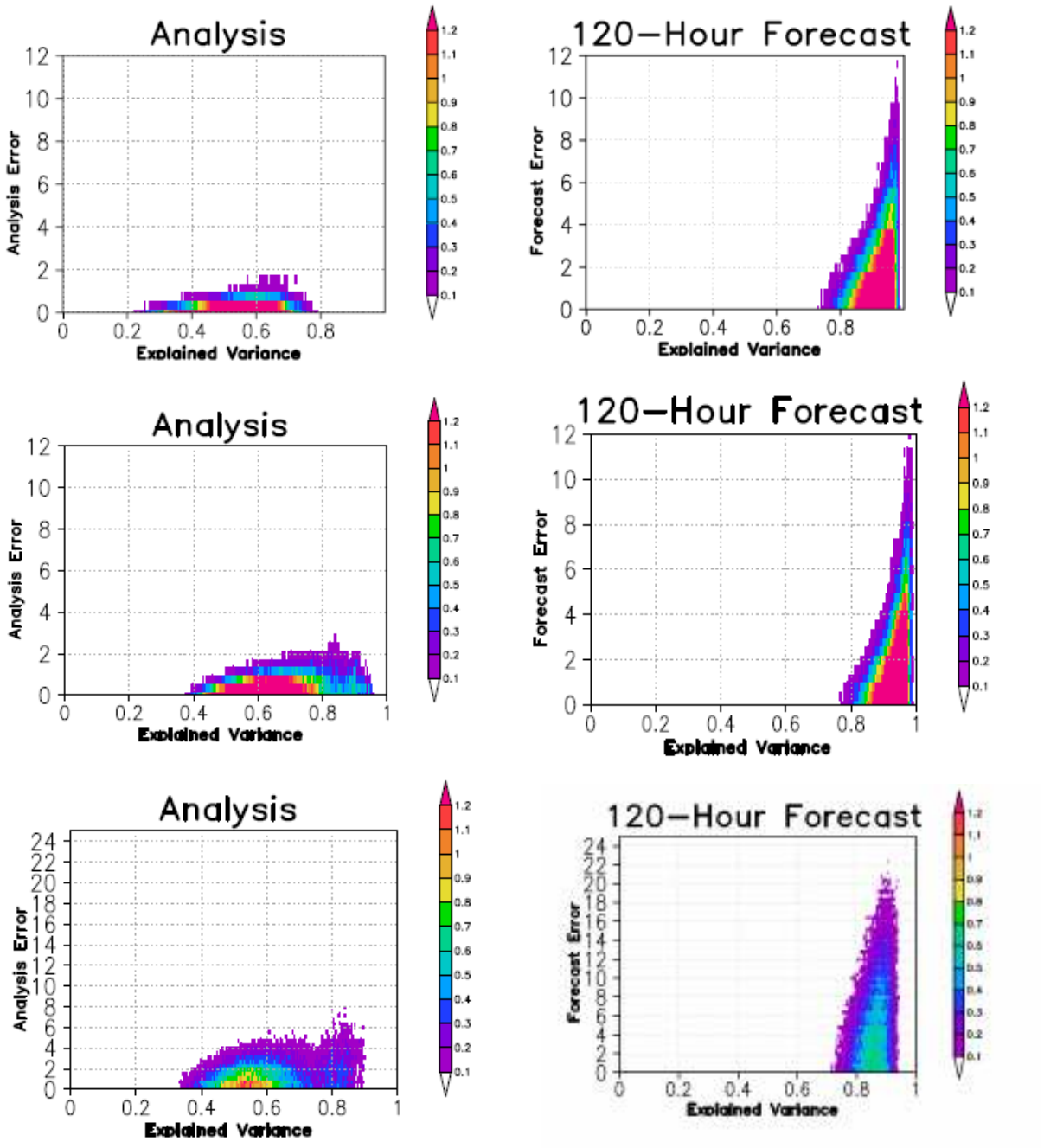

Figure 4: Joint probability distribution of the analysis/forecast errors and the explained variance. The bin increments are 0.005 for the explained variance and 0.4 for the forecast error. Shown are the distributions for experiments that assimilate randomly distributed simulated observations (top panel), 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 panels.

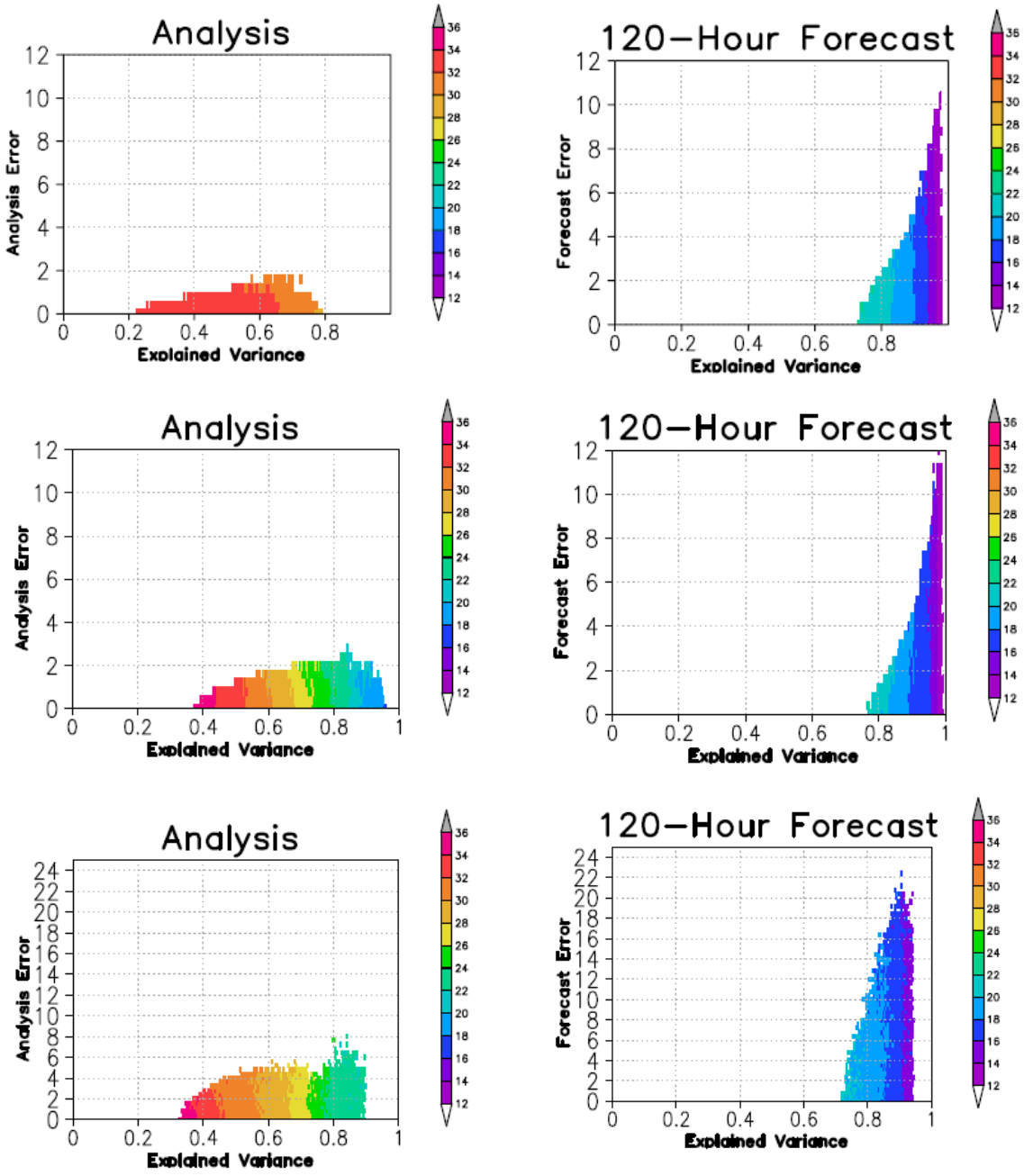

Figure 5: Color shades indicate the mean E-dimension for each nonempty bin in Figure 1. Shown are the distributions for experiments that assimilate randomly distributed simulated observations (top panel), simulated observations at the locations of conventional observations (middle panel), and conventional observations of the real atmosphere (bottom panel). Note the different scale for the forecast errors in the bottom panels.

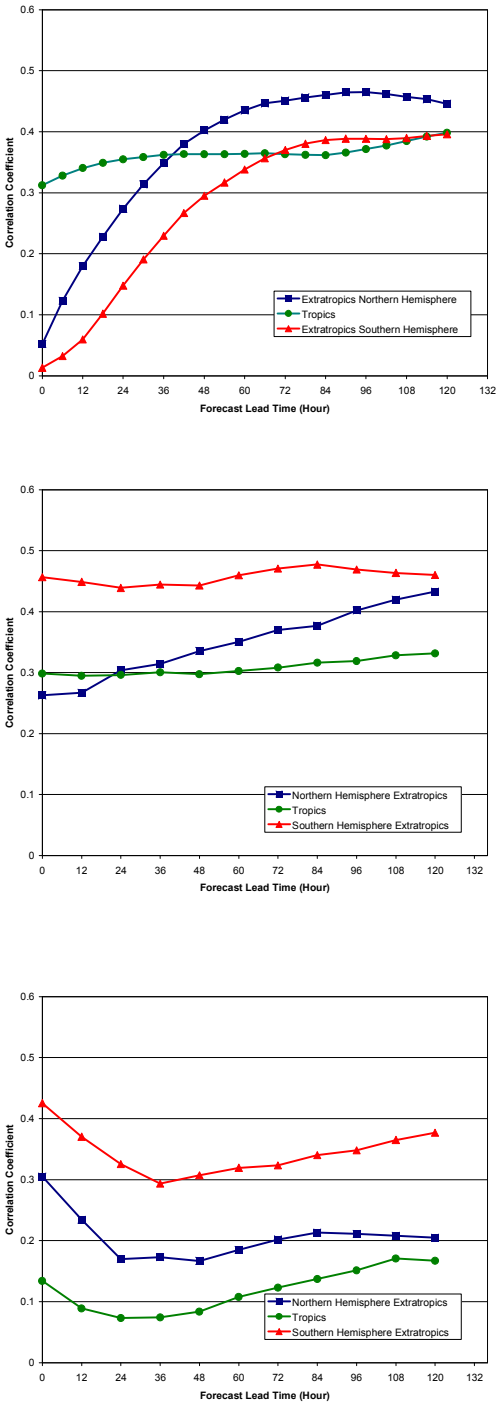

Figure 6: Correlation between the absolute error of the ensemble mean forecast and the ensemble spread for the meridional wind component at 500 hPa. The analyses are obtained by assimilating simulated observations at random locations (top panel), simulated observations at the location of the conventional observations of the real atmosphere (middle panel), and conventional observations of the real atmosphere (bottom panel). Results are shown for three different geographical regions: Northern Hemisphere extratropics (blue), Southern Hemisphere extratropics (red), and the tropics (green). The average is taken over all forecasts started between 11 January 2004 0000UTC and 15 February 2004 0000UTC.