Independent Estimations of the Asymptotic Variability in an Ensemble Forecast System

LISA K. BENGTSSON*

Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

LINUS MAGNUSSON AND ERLAND KÄLLÉN

Department of Meteorology, Stockholm University, Stockholm, Sweden

(Manuscript received 14 January 2008, in final form 26 March 2008)

ABSTRACT

One desirable property within an ensemble forecast system is to have a one-to-one ratio between the root-mean-square error (rmse) of the ensemble mean and the standard deviation of the ensemble (spread). The ensemble spread and forecast error within the ECMWF ensemble prediction system has been extrapolated beyond 10 forecast days using a simple model for error growth. The behavior of the ensemble spread and the rmse at the time of the deterministic predictability are compared with derived relations of rmse at the infinite forecast length and the characteristic variability of the atmosphere in the limit of deterministic predictability. Utilizing this methodology suggests that the forecast model and the atmosphere do not have the same variability, which raises the question of how to obtain a perfect ensemble.

1. Introduction

Forecast errors in numerical weather prediction models grow in time because of the model error and the inaccuracy of the initial state. Because of the nonlinearity of the governing equations the forecast error will saturate after some time period, which on average is about two weeks (Lorenz 1969). Ensemble prediction systems (EPS) are used in order to estimate forecast uncertainties. However, in present-day EPS there is a lack of spread around the ensemble mean (i.e., the ensemble systems today are in general underdispersive; Buizza et al. 1999). This may be due to one of two things; either there is a lack of variability in the forecast model, or inadequate sampling of initial perturbations. In this study we will investigate how well the forecast model is able to reproduce the characteristic variability of the atmosphere in the limit of deterministic predictability. This will be done by using a simplified model for forecast error growth (Lorenz 1982; Savijärvi 1995) and fitting the parameters of this model to results from an EPS. Accurate climatological properties of an ensemble prediction system are important, for instance, if the ensemble succeeds to capture the atmosphere's characteristic variability, then its potential for use in seasonal prediction may be realized.

With a perfect prediction model, initial inaccuracy in some spatial scale will eventually contaminate all other scales through nonlinear interactions (Lorenz 1969). On the other hand, even with a perfect initial state, forecast errors are inevitable because of model imperfections. For instance, truncation error and physical parameterization are components that contribute to forecast error. One way of representing model errors resulting from physical parameterizations in the ensemble system is to account for uncertainties in the parameterized fluxes using stochastic physics (Shutts 2004). We will also investigate the effects of using stochastic physics in the EPS on the ensemble spread.

One member within an EPS uses the analysis of the deterministic system as its initial state (i.e., it is not initially perturbed). This is commonly called the control forecast. Its error should initially be less than the aver-

* Additional affiliation: Department of Meteorology, Stockholm University, Stockholm, Sweden.

Corresponding author address: Lisa K. Bengtsson, Swedish Meteorological and Hydrological Institute, SMHI SE-601 76, Norrköping, Sweden.
E-mail: lisa.bengtsson@smhi.se

DOI: 10.1175/2008MWR2526.1

age error of the forecast ensemble members, but after some time the errors of the control forecasts will be equal to that of the individual ensemble members due to error saturation (error growth with time equals zero). This will be true for a perfect model, and certainly so when the forecasts have reached the limit of the deterministic predictability range. At this time, the RMSE of the ensemble members and the control forecast should be of the same magnitude as the difference between two randomly chosen states in the atmosphere (Toth 1991). This may be referred to as the atmospheric variability (i.e., the characteristic variability of the atmosphere in the limit of deterministic predictability).

In the same fashion, the RMSE of the ensemble mean and the ensemble spread should saturate toward the same asymptotic value after a sufficiently long forecast time (Palmer 1999). In an underdispersive model the error saturation limit of the ensemble members and the control forecast will be lower than the asymptotic limit of the atmospheric variability. Furthermore, the asymptotic limit of the ensemble spread will be lower than the error saturation limit of the ensemble mean. We will investigate these relationships by utilizing some simple relations of the root-mean-square error (rmse) and compare them with our extrapolated values of the rmse of the ensemble members, ensemble mean, control forecast, and ensemble spread.

In the next section we describe the forecast model and EPS used for this study, in section 3 we introduce some relationship between the ensemble mean rmse and the ensemble spread, which can be expected at infinite forecast length. We also introduce the error growth parameterization described by Savijärvi (1995). Our results are presented in section 4 and the paper is concluded in section 5.

2. Model description

Since 19 December 1992 the European Centre for Medium-Range Weather Forecasts (ECMWF) has been conducting operational global ensemble forecasts in order to estimate the probability distribution function of forecast states as a complement to the deterministic forecasting system. Initially, the ECMWF EPS only used initial perturbations for the analysis, but since October 1998 the EPS also includes a stochastic scheme designed to simulate the random model errors due to parameterized physical processes (Buizza et al. 2005). The EPS used for this study consists of 51 members and has a resolution of TL255L40 (spectral truncation at total wavenumber 255 and 40 vertical levels). The data for this study is retrieved on a regular longitude–latitude grid with 1° resolution. All the experimentation has been performed using geopotential height at 500 hPa. The time period used for this study includes the period 1 December 2004–28 February 2005 for forecasts valid at 1200 UTC each day.

To generate initial perturbations for the EPS, ECMWF use the singular vector (SV) technique (Barkmeijer et al. 1999; Leutbecher and Palmer 2008). Singular vectors are designed to identify directions in phase space that give the largest amplifications of perturbations during a finite-time interval (optimization time), in order to find possible extreme events in the atmosphere. To calculate the singular vectors a tangent linear forward and adjoint model with simplified physics compared to the ordinary model is used. The perturbation growth is maximized in terms of total energy norm and over 48 h.

To address forecast error resulting from model imperfection, such as the finite resolution of the model grid or simplified physical parameterization, the ECMWF EPS has introduced stochastic physics. For each ensemble member, the stochastic physics perturbs tendencies of parameterized physical processes. For the data used in this study the stochastic physics was implemented by multiplying a total subgrid-scale parameterization tendency at all points within the predefined longitude–latitude boxes by a number selected randomly, with uniform probability, from the range 0.5 to 1.5. A uniform latitude–longitude grid with 10° resolution was used operationally to define these boxes and the numbers are updated every 6 h (Shutts 2004). The whole globe is perturbed, including the tropics. The control forecast is a member within the EPS that uses the analysis of the deterministic system as its initial state (i.e., it is not initially perturbed and is run without stochastic physics).

3. Methodology

A commonly used measure of the error in a numerical weather prediction system is the rmse. Mathematically rmse for a forecast f evaluated against an analysis a can be written as

$$E^2 = \overline{(f - a)^2}, \tag{1}$$

where the overbar represents the mean over all grid points and a number of forecasts.

This notation can be centered round the climate c:

$$E^2 = \overline{(f - c)^2} + \overline{(a - c)^2} - 2\overline{(f - c)(a - c)}. \tag{2}$$

The first term is the forecast variability around the climate, the second is the analysis variability around the climate, and the third is the correlation between the forecast and analysis anomalies with respect to climate. For short forecast times, the correlation term is comparable to the sum of the other two, and yields small rmses. For very long forecast times the forecast and analysis become uncorrelated and the rmse becomes (Simmons et al. 1995):

$$E^2 = \overline{(f - c)^2} + \overline{(a - c)^2}. \quad (3)$$

Therefore, if the numerical model has the same variability as the atmosphere, the forecast variability around the climate will saturate with a factor of $\sqrt{2}$ larger than the analysis variability around the climate for long forecast lead times (beyond the limit of deterministic predictability).

In ensemble prediction systems the initial state of a forecast model is perturbed around the most probable analysis state, and an ensemble forecast is produced to obtain an estimate of the forecast uncertainty. By calculating the mean of all forecasts included in an ensemble forecast (ensemble members), it is possible to get a forecast that filters unpredictable scales. For long lead times, the rmse for the ensemble mean (EM) becomes

$$E_{\mathrm{EM}}^2 = \overline{(\mathrm{EM} - a)^2} = \overline{(a - c)^2} + \overline{(\mathrm{EM} - c)^2} - 2\overline{(\mathrm{EM} - c)(a - c)}. \quad (4)$$

From Jung (2005) we find that the systematic error component is an order of magnitude smaller than the total error at the medium range. Since our estimates of the asymptotic error limit are based on an extrapolation from the medium range we can assume that the affect of the systematic model error component is sufficiently small, thus the model's climate can be assumed to equal the climate of the atmosphere. For very long lead times the ensemble mean approaches climate and we can therefore eliminate the last term in Eq. (4), and the $\overline{(\mathrm{EM} - c)^2}$ term will be negligible. In this case the error for the ensemble mean will on average never exceed the variance of the analysis around the climate, and we get (as $t \to \infty$)

$$E_{\mathrm{EM}}^2 = \overline{(\mathrm{EM} - a)^2} = \overline{(a - c)^2}. \quad (5)$$

Using the same technique as above, the spread of the ensemble can be written as

$$S^2 = \overline{(f - \mathrm{EM})^2} = \overline{(f - c)^2} + \overline{(\mathrm{EM} - c)^2} - 2\overline{(f - c)(\mathrm{EM} - c)}. \quad (6)$$

For a sufficiently large ensemble, the correlation between the members and ensemble mean goes to zero for a long forecast times. When $t \to \infty$, $EM \to c$ and spread goes to

$$S^2 = \overline{(f - \mathrm{EM})^2} = \overline{(f - c)^2}. \quad (7)$$

Using the derived relations in Eqs. (5) and (7), it is clear that the variability of the atmosphere and the numerical model have to be equal to obtain a one to one ratio between the rmse and spread for the ensemble. The average ensemble forecast error E_{ENS} is defined as the average of the rmse of all individual ensemble members. As for the forecast error in Eq. (3) we find that E_{ENS} goes to E if $(a - c) = (f - c)$. Therefore, E_{ENS} can be written as

$$E_{\mathrm{ENS}}^2 = E_{\mathrm{EM}}^2 + S^2. \quad (8)$$

The main result of these derivations can thus be summarized as follows: the rmse of the ensemble members can be described in the same way as the forecast rmse at infinite forecast time if $(a - c) = (f - c)$. Therefore, the asymptotic error limit of the control forecast should be equal to that of the ensemble members. Furthermore, some information is given about how the EPS behaves at infinite forecast length; that is, the rmse of the ensemble members should saturate toward $\sqrt{E_{\mathrm{EM}}^2 + S^2}$ and also equal $\sqrt{2}E_{\mathrm{EM}}$. We will investigate how will the EPS is able to reproduce these statistics in the limit of deterministic predictability, this will be done by using a simplified model for forecast error growth (Lorenz 1982; Savijärvi 1995) described below.

Error growth in the EPS

Lorenz (1982) suggested a way to parameterize the growth of small initial errors in a perfect model using the simple quadratic equation:

$$\frac{dE}{dt} = \alpha E\left(1 - \frac{E}{\beta}\right). \quad (9)$$

Here, E represents the rmse between the analysis and the forecast of the geopotential height at 500 hPa. The constant α governs the growth rate of small errors. After the errors have growth large they reach a finite amplitude and asymptote toward a saturation value β. The asymptotic constant β can be defined as the average differences between two randomly chosen atmospheric states (Toth 1991).

Savijärvi (1995) recognized that the systematic error (model's climate drift) is an essential, multiplicative,

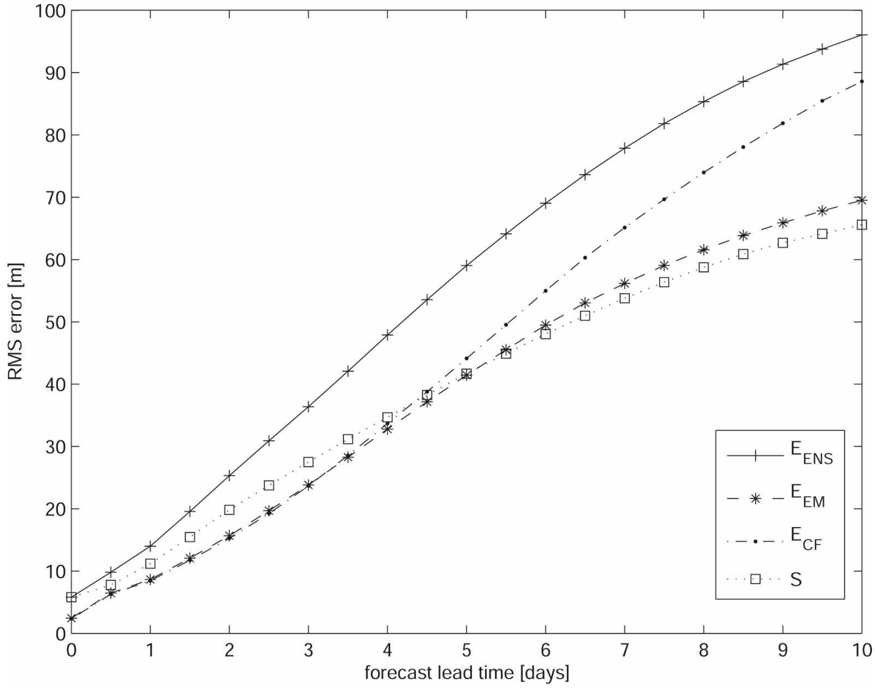

FIG. 1. Rmse of E_{ENS}, E_{EM}, E_{CF}, and S as a function of lead time for the period DJF 2004/05.

part of the model error, and included an extra term s representing the model error is

$$\frac{dE}{dt} = (\alpha E + s)\left(1 - \frac{E}{\beta}\right). \tag{10}$$

By plotting a finite-difference estimate of dE/dt against E, we can find β for the ensemble forecast. If the error growth were governed by Eq. (10), the relation between E and dE/dt will lie on a parabola and can be fitted with a quadratic polynomial:

$$\frac{dE}{dt} = p_1 E - p_2 E^2 + p_3. \tag{11}$$

Using Eqs. (10) and (11) the coefficients p_1, p_2, and p_3 may be written as

$$p_1 = \alpha - \frac{s}{\beta}, \quad p_2 = \frac{\alpha}{\beta}, \quad p_3 = s. \tag{12}$$

We apply this methodology to the control forecast and E_{ENS} to determine when these two error converge. We also apply Eq. (8) and the relation $E_{ENS} = \sqrt{2}E_{EM}$ at $dE/dt = 0$ in order to investigate if these relationships hold true.

In all of our experiments we have used the analysis to compute the rmse. The analysis is based on a background forecast from the same model. The results may therefore suggest that the ensemble performs better than it actually does if we knew the truth, and this problem is likely to be larger over ocean than over land areas.

4. Result

Globally averaged rmses computed for the winter months December–February (DJF) 2004/05 are presented in Fig. 1 for ensemble members (solid), control forecast (dash–dotted), ensemble mean (dashed), and the ensemble spread (dotted). The error of the ensemble mean and the control forecast is the square root of the calculated rmse squared plus the analysis error (from the ECMWF archive system) squared since the forecast error at day 1 is calculated assuming that the verifying analyses at day 0 are errorless. However, it was found that the average global inaccuracy in geopotential height at day 0 is roughly 3 m at 500 hPa. Initially the error growth is exponential, but after about a week the error asymptotes toward a saturation value. For short lead times it can be seen that the rmse of the ensemble mean forecast is similar to the control forecast, but after some days mainly due to filtering of unpredictable scales, the ensemble mean outperforms the control forecast in terms of rmse.

To investigate the error saturation value of the ensemble members, we must extrapolate the error growth

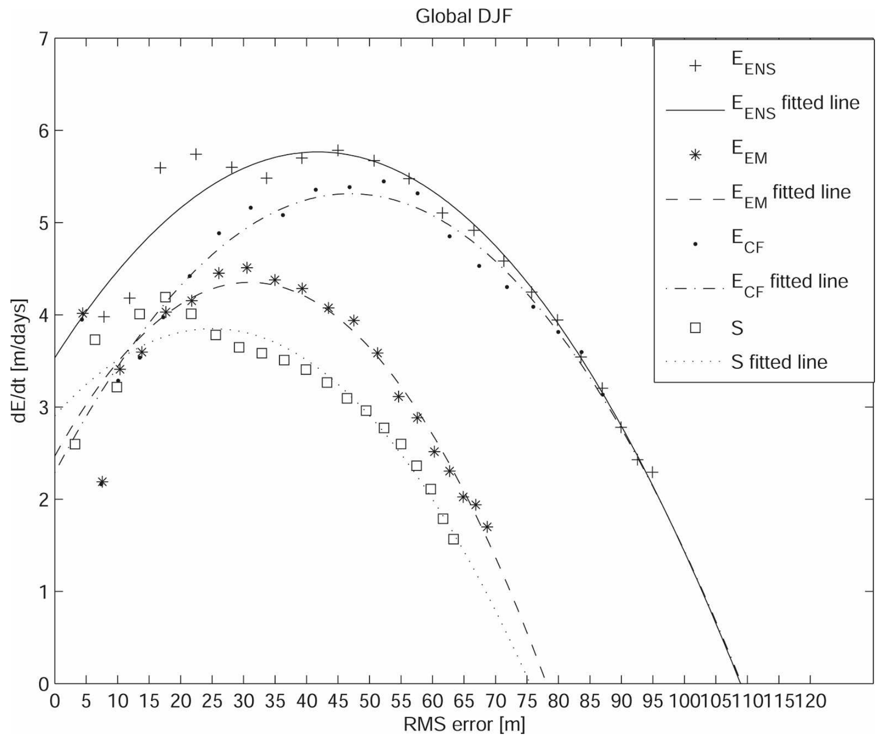

FIG. 2. Change in rmse against the deterministic forecast of the average rmse for E_{ENS}, E_{EM}, E_{CF}. Also, the increase in S against the average S of the EPS for the period DJF 2004/05.

beyond 10 days. We use the parameterization described in section 3 in which the tendencies of the parameters plotted in Fig. 1 are plotted as a function of the rmse and spread. If the error growth were governed by Eq. (10), Fig. 2's dots would lie on a parabola and fit well the quadratic polynomial [i.e., Eq. (11)]. These polynomials are also plotted in Fig. 2.

The polynomial gives a poor fit for the first few ensemble spread points. The derivative of the spread is too large compared to the polynomial. This could be related to the superexponential growth characteristics of singular vectors (Trevisan et al. 2001; Wei and Frederiksen 2004). Having too large a spread at short lead times is also reflected in the rmse of the ensemble members, and gives a faster error growth in the beginning; this is also evident in Fig. 1. After the optimization time (the first 4 points in Fig. 2, 48 h) of the singular vectors the quadratic polynomial [i.e., Eq. (11)] supplies an accurate representation for the three rmse variables (i.e., the perturbed ensemble members, the ensemble mean, and the control forecast).

From the fitted polynomials it is possible to estimate the saturation level of the rmse ($dE/dt = 0$) and the spread ($dS/dt = 0$). We found the error saturation limit for the ensemble members to be 108.9 m, the control forecast 109.0 m, and 78.1 m for the ensemble mean. The spread saturates at 74.1 m. The fact that the rmse for the ensemble mean and the spread does not saturate at the same level indicates that the model and the atmosphere do not have the same variability.

We can theoretically compute the asymptotic error limit for the ensemble members using Eq. (3) in section 3. Using $(a - c)$ for RMSE for ensemble mean and $(f - c)$ for the spread:

$$E = \sqrt{S^2 + E_{EM}^2}, \qquad (13)$$

which gives a value for rmse for the ensemble members of 108.6 m, which is close to the value obtained at $dE/dt = 0$ (108.9 m). Because of the arguments presented in section 3, we can find the growth of the rmse of the ensemble members by taking $\sqrt{2}E_{EM}$ if the variability of the model and the atmosphere is equal. It may be seen however that the rmse of the ensemble members in this dataset is not represented by $\sqrt{2}E_{EM} = 110.5$ since it has a lower asymptotic error saturation value. The fact that the relationship between the rmse for ensemble mean and ensemble spread can be employed to calculate a good value for the rmse of the ensemble members that nicely matches the extrapolated value; it indicates that the use of the relations are accurate.

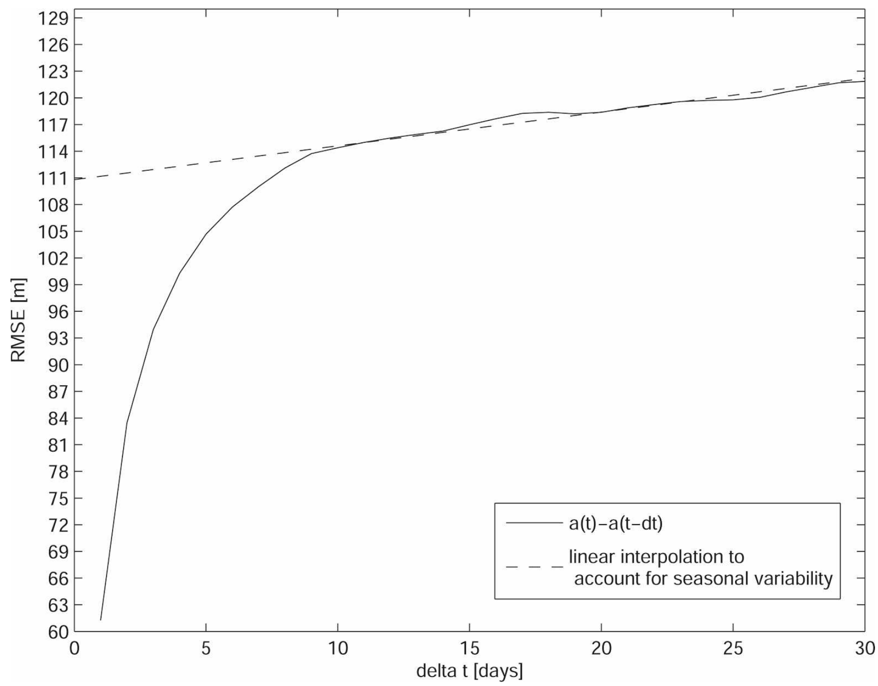

FIG. 3. The rmse of two analyses δt apart, where δt is between 1 and 30 days over the winter months DJF of 1997–2005 (solid line) and linearly extrapolated in order to account for seasonal variability (dashed line). See text for more information.

So far we have estimated the variability of the atmosphere using arguments presented in section 3, using $\sqrt{2}E_{\mathrm{EM}} = 110.5$. Another method in which we can compute the atmospheric variability is to look at the difference between two randomly chosen atmospheric states (Toth 1991) by using

$$E^2 = \overline{(a_t - a_{t-\delta t})^2} = 2\overline{(a - c)^2}. \qquad (14)$$

Letting t vary between 1 and 90 days, and δt between 1 and 30 days the error saturation limit was calculated for analyses δt days apart. The average error growth with respect to the winter months DJF of years 1997–2005 are presented in Fig. 3. It can be seen that the error grows rapidly for small δt, but for δt greater than 10 days the error grows linearly with less than a meter per day. We assume that this growth beyond 10 δt is due to a bias introduced by seasonal variability. By taking this linear error growth due to seasonal variability into account, we can linearly extrapolate the atmospheric variability to be 110.8 m. This value is in close agreement with the asymptotic error level found from $\sqrt{2}E_{\mathrm{EM}} = 110.5$ m. The agreement between these two methods to find that atmospheric variability indicates that the assumption $\overline{(\mathrm{EM} - c)^2} = 0$ holds true, and that linearly extrapolating the curve showing the difference between arbitrary atmospheric states as a technique for eliminating the effects of a seasonal cycle works.

We can also observe in Fig. 2 the error growth of the control forecast and ensemble members coverage when the error has become sufficiently large. By finding the solution to Eq. (10), and finding the coefficients p_1, p_2, and p_3 in the polynomial given by Eq. (11), we can extrapolate the error growth beyond 10 days and find that the error of the control forecast days converges with the error of the ensemble members after roughly 21 forecast days (Fig. 4). According to (Lorenz 1969; Daley 1993), and as seen in Fig. 3, 21 forecast days is well beyond the limit of deterministic predictability, therefore, the rmse growth of the ensemble members is too rapid, or the growth of the rmse of the control forecast is too slow. A likely reason for this behavior is the representation of model error in the EPS, by the inclusion of stochastic physics. As an example, Fig. 5 shows the same extrapolation for the dataset of DJF 1997–98, which was utilized before stochastic physics was implemented in the EPS. At that time the control forecast converges with the ensemble members and the control forecast for the time period DJF 1997/98 have a higher asymptotic limit than the rmse of ensemble members and the control forecast during the time period DJF 2004–05. This can be caused by several fac-

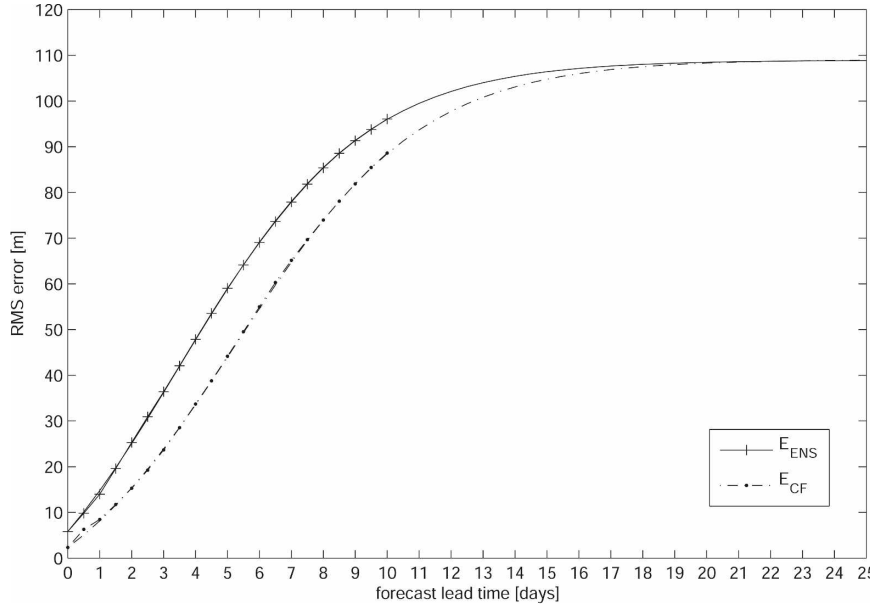

FIG. 4. Rmse of E_{ENS} and E_{CF} as a function of lead time extrapolated beyond 10 days for the time period DJF 2004/05.

tors, for example the horizontal and vertical resolution has been updated since 1998, or the weather was simply more difficult to forecast resulting in poorer model results that particular year. However, such differences would affect the error growth of both the control forecast and the ensemble members' stochastic physics is only reflected in the error growth of the ensemble members. Applying stochastic physics therefore appears to give a more rapid error growth of the ensemble members (the solid curve only in Fig. 3), causing a convergence with the control forecast beyond the limit of deterministic predictability. However, despite the inclusion of stochastic physics, the EPS is still underdispersive in comparison with the atmosphere's characteristic variability and $\sqrt{2}E_{\mathrm{EM}}$. This suggests that the forecast model itself is indeed lacking in variability.

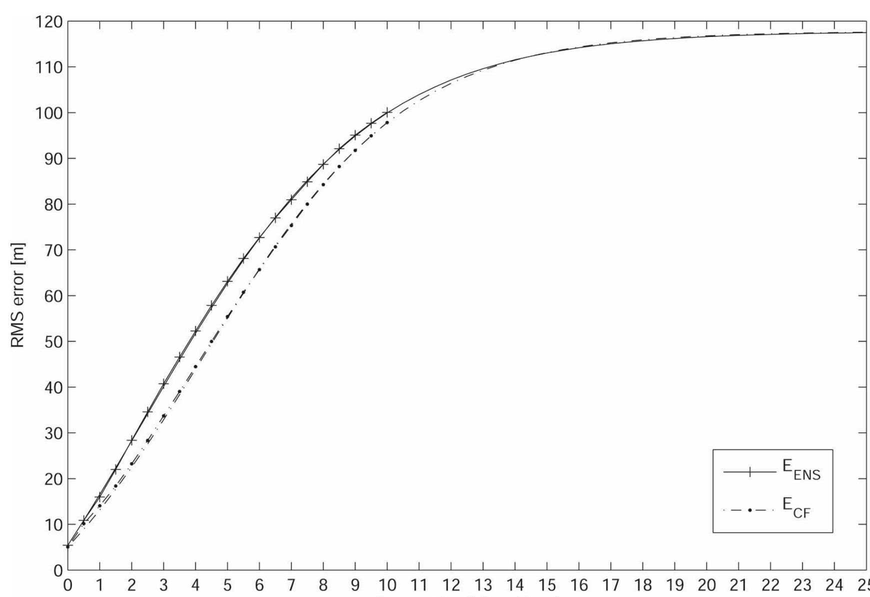

FIG. 5. Same as Fig. 4, but for the time period DJF 1997/98.

5. Conclusions

In this study we have focused on the climatology of ensemble forecasting; what characteristics can we expect at infinitely long forecast lengths? By using a simple model of error growth, as first completed by Lorenz 1982 and later described by (Savijärvi 1995) in which a term is included in order to represent the model error, we can extrapolate the forecasts beyond 10 days. By utilizing some simple relations of the rmse and investigating how well the EPS is able to reproduce these relations at the limit of the deterministic predictability, we may draw some conclusions about the ensemble prediction system at hand. More specifically, we have compared the rmse of the ensemble spread at the time of deterministic predictability, and explored how this relationship is affected by the variability of the forecast model and the variability of the atmosphere. Doing so, it was found that the rmse for the ensemble mean and the spread does not saturate toward the same asymptotic error limit, this indicates that the model and the atmosphere do not have the same variability. This is also indicated by the fact that the rmse for the ensemble members does not asymptote toward $\sqrt{2(a-c)^2}$, and by the fact that $\sqrt{2(a-c)^2}$, is in agreement with atmosphere's characteristic variability computed from two randomly chosen analyses in the atmosphere.

We have also tried to investigate where the error of the control forecast converges with the error of the ensemble members using Savijärvi's (1995) parameterization of error growth in NWP. It was found that the rmse of the control forecast converge with the rmse of the ensemble members at an asymptotic error level lower than that of the atmosphere's characteristic variability. This indicates that the model is not variable enough. However, it was also found that the rmse of the control forecast and rmse of the ensemble members converge after 21 forecast days, which is well beyond the deterministic range of predictability that is roughly of the order of two weeks (Daley 1993). Applying stochastic physics appears to give a more rapid error growth to the ensemble members, although, the asymptotic error level remains the same; this results in a convergence with the control forecast beyond the limit of deterministic predictability. However, despite the inclusion of stochastic physics, the EPS is still underdispersive in comparison with the atmosphere's characteristic variability and $\sqrt{2}E_{\mathrm{EM}}$.

Our results that the sufficient variability in the model makes it impossible to obtain sufficient ensemble spread raise the question of how to obtain a perfect ensemble from only looking at the initial perturbation technique. Perhaps it is more interesting to investigate the behavior of the atmospheric model itself?

Acknowledgments. The authors thank ECMWF for providing ensemble forecasts and analysis data for this study. We also thank Anders Persson for his helpful comments.

REFERENCES

Barkmeijer, J., R. Buizza, and T. N. Palmer, 1999: 3D-var Hessian singular vectors and their potential use in ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.,* **125,** 2333–2351.

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

——, P. L. Houtekamer, Z. Toth, G. Pelerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.,* **133,** 1067–1097.

Daley, R., 1993: *Atmospheric Data Analysis.* Cambridge University Press, 471 pp.

Jung, T., 2005: Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.,* **131,** 1045–1073.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.,* **227,** 3515–3539.

Lorenz, E. N., 1969: The predictability of a flow which posses many scales of motion. *Tellus,* **21,** 289–307.

——, 1982: Atmospheric predictability experiments with a large numerical model. *Tellus,* **34,** 505–513.

Palmer, T. N., 1999: Predicting uncertainty in forecasts of weather and climate. Tech Memo. 294, ECMWF, 64 pp.

Savijärvi, H., 1995: Error growth in a large numerical forecast system. *Mon. Wea. Rev.,* **123,** 212–221.

Shutts, G., 2004: A stochastic kinetic energy backscatter algorithm for use in ensemble prediction systems. Tech. Memo. 449, ECMWF, 28 pp.

Simmons, A., R. Mureau, and T. Petroliagis, 1995: Error growth and estimates of predictability from the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.,* **121,** 1739–1771.

Toth, Z., 1991: Estimation of atmospheric predictability by circulation analogs. *Mon. Wea. Rev.,* **119,** 65–72.

Trevisan, A., F. Pancotti, and F. Molteni, 2001: Ensemble prediction in a model with flow regimes. *Quart. J. Roy. Meteor. Soc.,* **127,** 343–358.

Wei, M., and J. Frederiksen, 2004: Error growth and dynamical vectors during southern hemisphere blocking. *Nonlinear Processes Geophys.,* **11,** 99–118.