Nonlinear Processes in Geophysics, 12, 1021–1032, 2005
SRef-ID: 1607-7946/npg/2005-12-1021
European Geosciences Union

Nonlinear Processes in Geophysics

A comparison of predictors of the error of weather forecasts

M. S. Roulston

Department of Meteorology, Pennsylvania State University, University Park, USA

Received: 29 August 2005 – Revised: 22 November 2005 – Accepted: 3 December 2005 – Published: 20 December 2005

Abstract. Three different potential predictors of forecast error – ensemble spread, mean errors of recent forecasts and the local gradient of the predicted field – were compared. The comparison was performed using the forecasts of 500 hPa geopotential and 2-m temperature of the ECMWF ensemble prediction system at lead times of 96, 168 and 240 h, over North America for each day in 2004. Ensemble spread was found to be the best overall predictor of absolute forecast error. The mean absolute error of recent forecasts (past 30 days) was found to contain some information, however, and the local gradient of the geopotential also provided some information about the error in the prediction of this variable.

Ensemble spatial error covariance and the mean spatial error covariance of recent forecasts (past 30 days) were also compared as predictors of actual spatial error covariance. Both were found to provide some predictive information, although the ensemble error covariance was found to provide substantially more information for both variables tested at all three lead times.

The results of the study suggest that past errors and local field gradients should not be ignored as predictors of forecast error as they can be computed cheaply from single forecasts when an ensemble is not available. Alternatively, in some cases, they could be used to supplement the information about forecast error provided by an ensemble to provide a better prediction of forecast skill.

1 Introduction

It is now accepted that the potential value of weather forecasts can be greatly enhanced if some information concerning the uncertainty of the prediction is also available (Zhu et al., 2002; Palmer, 2002). This information may take the form of a classification of forecasts into “low” and “high” predictability (Ziehmann, 2001), or a quantitative estimate of the likely error in the forecast of a continuous quantity, or a fully probabilistic prediction that provides an estimate of the probability distribution of a weather variable. Whatever the exact form of the information, the meteorological community now increasingly subscribes to the mantra of Tennekes et al. (1987) that *“No forecast is complete without a forecast of forecast skill!”* There is now a widespread appreciation that providing forecasts without acknowledgement, let alone quantification, of uncertainty is not tenable.

Probably the most straightforward method to estimate the likely error of a quantitative forecast is to base the estimate on past forecast errors. This method was used by Anders Ångström to produce probabilistic frost forecasts in the 1920s (Liljas and Murphy, 1994). This approach can be refined by conditioning the error statistics on such things as time of year, or even meteorological conditions. Such conditioning cannot be carried too far, however, or one ends up calculating the statistics based on too few cases.

In the 1960s, it was recognized that the atmosphere is a chaotic system – its evolution is sensitive to its initial state – and also that the predictability of the atmosphere is *state dependent*, changing from day-to-day depending on atmospheric conditions (Lorenz, 1963, 1965, 1968). This discovery motivated the development of *ensemble forecasting* in which multiple simulations are made using a numerical weather prediction model (Epstein, 1969; Leith, 1974). The simulations differ in their initial conditions and/or the details of the model used. These perturbations are made in an effort to estimate the impact of errors in the initial condition or in the model on the accuracy of the final forecast (Palmer, 2000). Model error can make a significant contribution to forecast error (Orrell et al., 2001) and this has motivated the development of *multi-model ensembles* (Palmer et al., 2005). Several different ensemble systems have been shown to have “spread-skill relationships”, that is the spread of the ensemble defined using a measure such as standard deviation or mean absolute deviation, is correlated with the variance of error between the verification and the mean of the ensemble (e.g. Whitaker and Loughe, 1998).

Correspondence to: M. S. Roulston
(roulston@met.psu.edu)

While ensemble forecasts have become a standard tool at several operational forecasting centres (Houtekamer and Derome, 1995; Molteni et al., 1996; Toth and Kalnay, 1997; Stensrud et al., 1999), other potential predictors of forecast error should not be neglected. This paper presents a comparison of three predictors: the spread of an ensemble, the mean error of past forecasts, and the local gradient of the predicted field in the ensemble control member. The latter two predictors are, in terms of computation, much cheaper than calculating the ensemble spread as they do not require multiple integrations of an NWP model. Also, the mean error of recent forecasts provides a baseline level of error predictability as it should capture persistent spatial structure in the error field as well as smoothly varying temporal structure – such as any seasonal variation in errors. This baseline error predictability should also be captured by the ensembles but in addition the ensembles should also capture day-to-day temporal variations in predictability if the extra computational expense is to be justified. This analysis is then extended to the problem of predicting spatial error covariances. A prediction of how the forecast errors at two different locations are likely to be correlated is potentially valuable information to a forecast-user whose risk or utility depends upon the weather at both locations. Ensemble error covariances and past error covariances are compared as potential predictors of error covariance.

2 Forecast data

The forecast data used in this study were provided by the European Centre for Medium Range Weather Forecasting. The ensemble forecasts were produced by the ECMWF ensemble prediction system (EPS) (Molteni et al., 1996). Each ensemble consisted of a control member, initialized with the best estimate of the current state of the atmosphere obtained using a 4D-var data assimilation scheme, and 50 other members initialized with their initial conditions perturbed in the space of the leading 25 singular vectors. Two fields, the 500 hPa geopotential and the 2-m temperature, were studied at three forecast lead times: 96 h, 168 h and 240 h. Although the ECMWF EPS forecasts are global only the forecasts over North America were analysed (30° N–70° N, 130° W–60° W). The forecasts issued on the 356 days starting with 1 January 2004 were used in the study. The numerical model used in the ECMWF EPS is a T255 spectral model but all the calculations described in this paper were conducted on forecasts and analyses projected onto a $1^\circ \times 1^\circ$ latitude-longitude grid.

The ECMWF operational analysis was used as the verification. In practice the purpose of numerical weather prediction is to predict actual observed weather conditions at specific locations. However, even under the assumption that an NWP model is "perfect" it still only predicts values of weather variables at scales commensurate with the resolution of the model. Some type of downscaling scheme, therefore, should be applied before the model is compared with observations. To avoid this additional complexity, and the possible ambiguity as to whether the origin of forecast errors lies in the numerical model or in the downscaling scheme, the model's own analysis was used for forecast verification.

3 Predictors of forecast error

The forecast error in this study is defined as the absolute error of the control (unperturbed) member of the ensemble. When using ensemble forecasts the forecast error is often defined as the error of the *ensemble mean*. This definition was not used in this study because part of the motivation for the investigation was to determine the usefulness of other predictors of forecast error that are available even in the absence of an ensemble forecast, and obviously if no ensemble is available then an ensemble mean cannot be calculated.

The three predictors of forecast error analysed in this study are described below.

– *Ensemble spread*: If ϕ_k is the value of the forecast variable in the k^{th} member of the ensemble then the ensemble spread of an N-member ensemble is defined, in this study, as the standard deviation of the ensemble, given by

$$\mathrm{SPREAD} = \frac{1}{\mathrm{N}-1}\sqrt{\sum_{\mathrm{k}=1}^{\mathrm{N}}\left(\phi_{\mathrm{k}} - \frac{1}{\mathrm{N}}\sum_{\mathrm{k}=1}^{\mathrm{N}}\phi_{\mathrm{k}}\right)^2} \tag{1}$$

"Spread-skill" relationships normally refer to the relationship between ensemble spread and the error of the ensemble mean, not the control forecast which is a single ensemble member. However, since the ensemble control, or indeed any ensemble member, is an estimator of the ensemble mean, a relationship between the ensemble mean and the error of any single member should also exist.

– *Mean past error*: The average past error used was the mean absolute error of the control forecast averaged over the most recent 30 daily forecasts for which verifications would be available. The average was calculated for each gridpoint. This predictor should reflect any persistent dependencies of error on location and also to some extent, the seasonality of forecast errors. For the forecast issued on day T with a lead time of L days, the mean past error is defined as

$$\mathrm{PAST} = \frac{1}{30}\sum_{\mathrm{t}=\mathrm{T}-\mathrm{L}-30}^{\mathrm{T}-\mathrm{L}-1} |\mathrm{verification}_{\mathrm{t}} - \mathrm{control}_{\mathrm{t}}| \tag{2}$$

Note that the most recent forecast used to calculate this quantity is the one issued $L+1$ days ago, as this is the most recent forecast for which a verification is available. The choice of 30 days was made as a trade off to obtain a reasonably representative estimate of the mean error while still allowing smoothly varying temporal variations in error variation that might be associated with the seasonal cycle to be captured. Error histories ranging from 10 to 60 days were also tried as predictors. Histories of less than 30 days yielded inferior predictions – a 10 day history typically explains about one-third less variability in error magnitudes, while histories of 40 to 60 days offered no improvement over 30 days.

– *Local gradient*: If $\phi_0(x, y)$ is the forecast variable in the control member of the ensemble then the local gradient of the predicted field is given by

$$\mathrm{GRADIENT} = \sqrt{\left(\frac{\partial\phi_0}{\partial \mathrm{x}}\right)^2 + \left(\frac{\partial\phi_0}{\partial \mathrm{y}}\right)^2} \quad (3)$$

The rationale behind this predictor is that some forecast errors can be characterized as distortions or displacements of the forecast field (Hoffman et al., 1995). In such cases, the error at a given location is likely to be larger if the local spatial gradient of the predicted field is large. This predictor shares a similar philosophy to the idea of "neighbourhood ensembles" used by Theis et al. (2005) for precipitation forecasting. The neighbourhood ensemble approach essentially creates an ensemble by using predictions from a neighbourhood of gridpoints defined in both space and time. This approach is also an attempt to account for spatial displacements and timing errors in the forecast. The numerical value of the GRADIENT predictor was estimated using midpoint finite differencing on a 1° grid, taking into account latitudinal differences in gridpoint size.

Figure 1 is a density estimate of absolute forecast errors of the ECMWF control member of 500 hPa geopotential, at a lead time of 96 h, plotted against each of the three predictors described above. The density plot was calculated by aggregating each of the 41×71 gridpoints over North America, for each of the 322 days in the dataset (the first 34 days of 2004 were required to calculate the mean past error of 30 verified 96 h forecasts leading up to February 4). The coefficients of linear correlation between the forecast error and the ensemble spread, the past errors and the local gradient respectively are 0.43, 0.27 and 0.25. These correlations may not appear to be particularly high but it must be remembered that the predictors are not predicting the actual magnitude of the forecast error but *the width of the distribution from which the error will be drawn*. As Houtekamer (1993) has pointed out, even under highly idealized circumstances spread-skill relationships are unlikely to show correlations greater than about 0.8.

A better idea of the relationship between the predictors of forecast error and the forecast error itself can be obtained if the mean absolute errors of many forecasts with similar values of the relevant predictor are calculated and plotted against the value of the predictor. Such plots are shown in Figs. 2–4 for the 500 hPa geopotential at lead times of 96, 168 and 240 h and in Figs. 5–7 for 2-m temperature, also at lead times of 96, 168 and 240 h. These plots were constructed by sorting the forecasts at each of the 2911 gridpoints, and for each day of the 326-L days starting on 31 January 2004 + L days (where L is the lead time of the forecast in days) into ascending order of the value of the predictor variable under examination. The mean absolute forecast errors of successive blocks of 1000 forecasts were then calculated. These mean errors are plotted against the mean value of the predictor for the 1000 forecasts in the block. The horizontal bars in Figs. 2–7 denote the range of the predictor values in each block of 1000 forecasts. The vertical bars represent the standard error in the mean forecast error for those 1000 forecasts.

Examination of Figs. 2–7 indicates that all three predictors provide some information about the expected magnitude of the forecast error. The ensemble spread generally exhibits the cleanest, most linear, relationship over the largest range of forecast error. This ability to predict a larger range of forecast error suggests that is has the most predictive power, although it is difficult to assess relative predictive power from graphical plots alone. A more quantitative analysis is provided in the next section. The slope of the ensemble spread-mean error line is sometimes close to, or even less than, unity. This is not necessarily an indication that the ensemble is overdispersive. If the errors are normally distributed the expected value of the *absolute error* will be $\sqrt{2/\pi}$ times the standard deviation of the distribution, so even if the ensembles were perfect the slope of the lines in Figs. 2–7 would be about 0.80. In addition, the spread of the ensemble is not exactly equal to the spread of the underlying distribution from which the ensemble is drawn, sometimes it will be less than the spread of the "true" distribution. To produce the plots the ensemble spreads are sorted into ascending order and thus, by construction, those forecasts where the ensemble spread underestimated the true spread of the ensemble distribution will tend to be at the low end, while the cases when the ensemble spread overestimated the true spread will tend to be at the high end.

The mean past error also has a reasonably linear relationship with forecast error, although over a smaller range which suggests it is less able to predict very large or small forecast errors. The mean forecast error is also approximately linear in the local field gradient for the 500 hPa geopotential fields although not for the 2-m temperature fields. For 2-m temperature the local field gradient is only a reasonable predictor for relatively small forecast errors. In the case of the 500 hPa geopotential the slope of the line represents a rough estimate of the horizontal scale of field displacements and distortions that are contributing to the local error. For the lead times of 96 h, 168 h and 240 h the slopes are approximately 60 km, 75 km and 100 km, respectively.

4 Multi-predictor analysis

To better quantify the relative amounts of information about forecast uncertainty provided by the three predictors a simple linear model was used,

$$\hat{e} = a[\mathrm{SPREAD}] + \mathrm{b}[\mathrm{PAST}] + \mathrm{c}[\mathrm{GRADIENT}] + \mathrm{d} \quad (4)$$

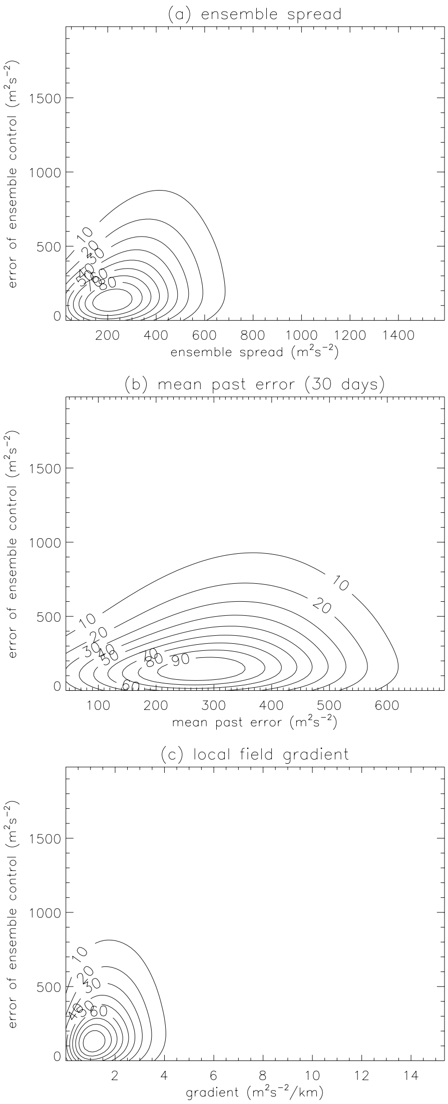

Fig. 1. Density estimates of "spread-skill" relationships of the actual forecast error, |verification−control| against the three predictors of forecast error: **(a)** ensemble spread, **(b)** average error of the most recent 30 daily forecasts for which the verification would be available and **(c)** local field gradient. The forecasts used were the ECMWF ensemble forecasts of the 500 hPa geopotential over North America at a lead time of 96 h initialized on each of the 322 days starting with 4 February 2004. Density is given in arbitrary units.

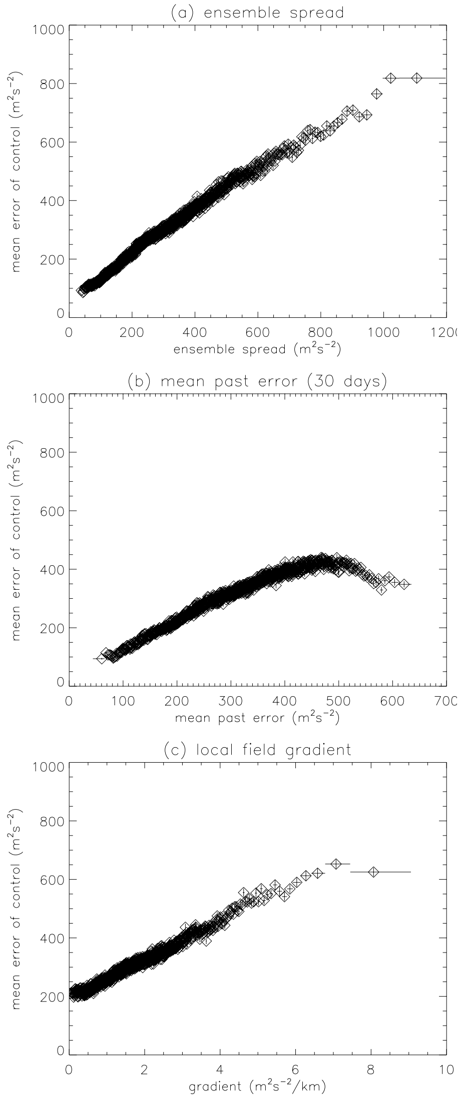

Fig. 2. Plots of the mean forecast error, ⟨|verification−control|⟩, of sets of 1000 forecasts with similar values of the three predictors of forecast error: **(a)** ensemble spread, **(b)** average error of the most recent daily forecasts for which the verification would be available and **(c)** local field gradient. The horizontal bars represent the ranges of the values of these predictors for each of the 1000 forecasts used to calculate each mean error (for lower values of the predictors these ranges are too small to yield visible bars). The vertical bars represent the standard error in the estimate of the mean forecast error. The forecasts used were the ECMWF ensemble forecasts of the 500 hPa geopotential over North America at a lead time of 96 h initialized on each of the 322 days starting with 4 February 2004.

where $\hat{e}$ is the estimated magnitude of the forecast error. Equation (4) is essentially an extension of the models used by Gneiting et al. (2005) and Jewson et al. (2004) to predict error as a linear function of ensemble spread[1]. In addition a simple linear model that used only the mean past error and the local gradient as predictors was also used,

$$\hat{e}' = b'[\mathrm{PAST}] + \mathrm{c}'[\mathrm{GRADIENT}] + \mathrm{d}' \tag{5}$$

The model coefficients were determined using a least-squares fit. Table 1 shows the percentages of variance of absolute forecast error explained by linear fits to each predictor individually and also to the linear combination of mean past error and local gradient (Eq. 5) and to all three predictors

[1] In the present study, ensemble standard deviation was used to predict error magnitude whereas Gneiting et al. (2005) used ensemble variance to predict error variance, so even without the extra predictors Eq. (4) is slightly different from their equation. Jewson et al. (2004) used a linear function of standard deviation, as in this study.

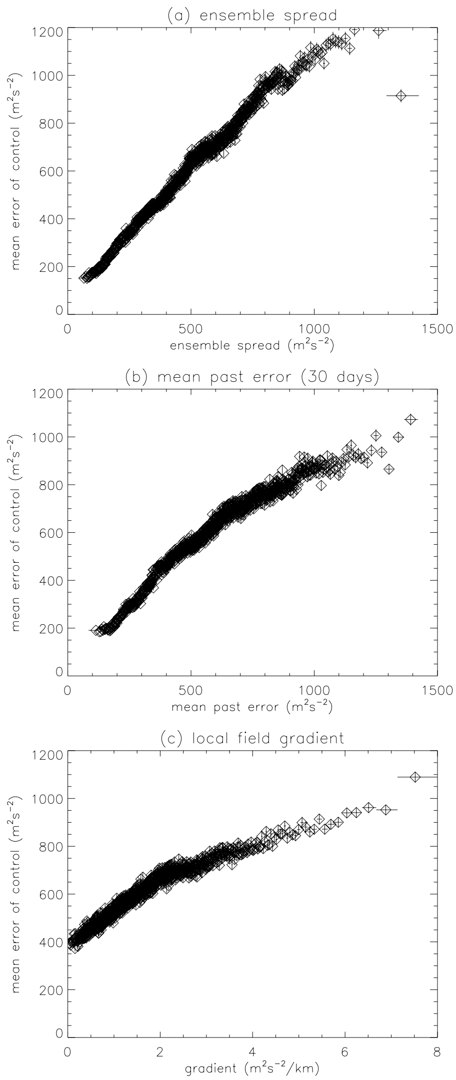

Fig. 3. As Fig. 2 but for a lead time of 168 h. First forecast was issued on 7 February 2004.

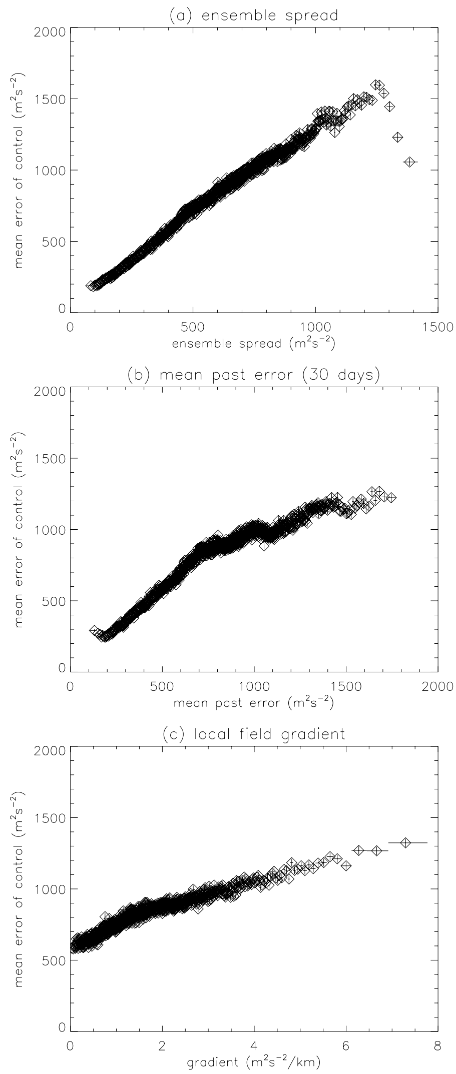

Fig. 4. As Fig. 2 but for lead time of 240 h. First forecast was issued on 10 February 2004.

(Eq. 4). These percentages are given by the squares of the coefficients of linear correlation between the error magnitude and the predictors, or the combination of predictors. The percentages given in Table 1 are the means obtained from 50 datasets constructed by bootstrap resampling the original dataset. For the purposes of resampling the dataset was divided into blocks 15° in longitude ×15° in latitude ×5 days in time. The blocks were then randomly resampled with replacement to create the 50 datasets. The standard deviations given in Table 1 were also estimated from the 50 resampled datasets. This bootstrap resampling in blocks was done to take account of spatial and temporal correlations in the forecast error.

From Table 1 it can be seen that, in the case of the 500 hPa geopotential, the ensemble spread alone predicts the error magnitude almost as well as all three predictors combined at all three lead time times examined. However, both the mean past error and the local field gradient contain some information about the error magnitude. Furthermore, the combination of these latter two predictors contains an appreciable amount of information. While the ensemble spread provides more information than the mean past error and the local gradient it must be remembered that these latter two predictors are computationally much cheaper and will be available even when multiple simulations from an NWP model are not. The ensemble spread is also the best predictor of 2-m temperature error, but for this variable the mean past error provides an appreciable fraction of information, while the local gradient is a poor predictor of error magnitude. Although the ensemble spread explains a larger amount of the variation in the magnitude of the forecast error than the mean past error or the local gradient, the fact that all three predictors explain more than either predictor alone indicates that the predictors are, to some degree, providing independent information about the magnitude of the error.

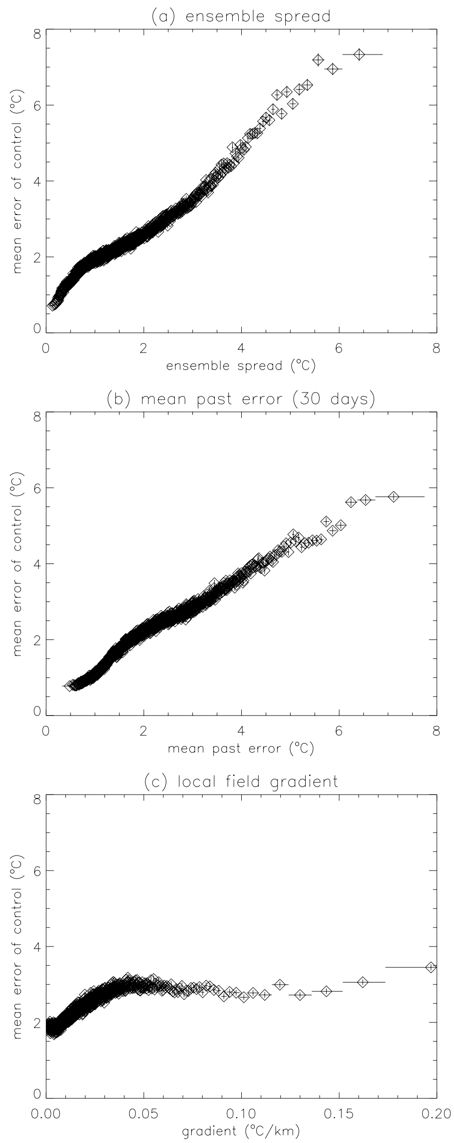

Fig. 5. Plots of the mean forecast error, ⟨|verification−control|⟩, of sets of 1000 forecasts with similar values of the three predictors of forecast error: **(a)** ensemble spread, **(b)** average error of the 30 most recent daily forecasts for which verifications would be available and **(c)** local field gradient. The horizontal bars represent the ranges of the values of these predictors for each of the 1000 forecasts used to calculate each mean error (for lower values of the predictors these ranges are too small to yield visible bars). The vertical bars represent the standard error in the estimate of the mean forecast error. The forecasts used were the ECMWF ensemble forecasts of the 2-m temperature over North America at a lead time of 96 h initialized on each of the 322 days starting with 4 February 2004.

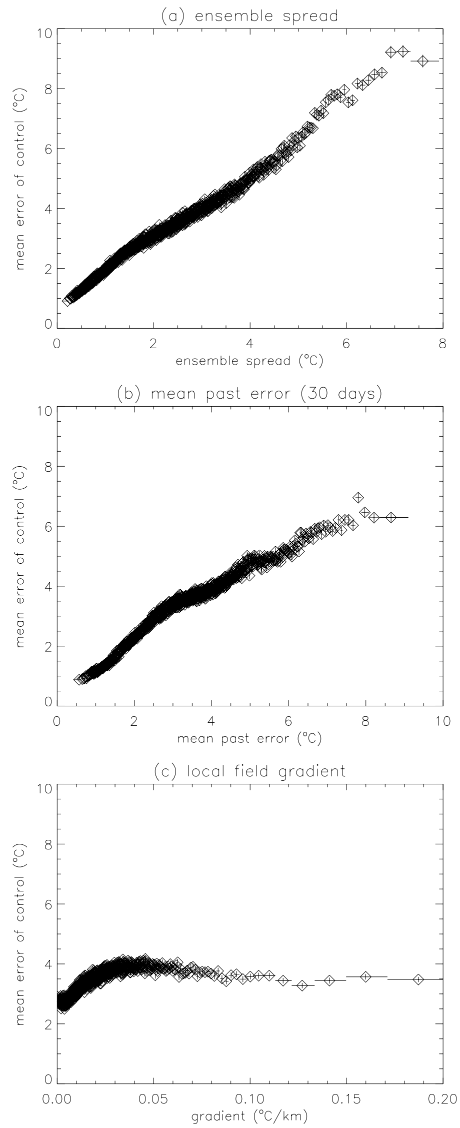

Fig. 6. As Fig. 5 but for a lead time of 168 h. First forecast was issued on 7 February 2004.

Table 1 was calculated by aggregating all gridpoints. Therefore it describes the ability of the predictors to explain the *spatio-temporal* variation in forecast error. It is conceivable that a predictor might only explain spatial variation – that the errors at one gridpoint are, on average, smaller than at another gridpoint – without providing any information about the *temporal* variation – how the error at a given gridpoint changes from forecast to forecast. While a good ensemble prediction system should explain such spatial variability in errors it should also be able to explain temporal variability because other predictors, particularly the average past errors, can capture purely spatial variability far more cheaply. The variability in error explained by the ensemble beyond that explained by the mean past error provides an approximate indication of how much of the error variability is temporal, rather than spatial. In this sense, the mean past error provides a control that accounts for spatial variability and also, to some extent, smoothly varying temporal variability, such as that associated with the seasonal cycle. A more explicit way of assessing how much purely *temporal* variability in errors is explained by the different predictors is to examine the correlations between the different predictors and the magnitude of the error at each gridpoint individually. This is done in Table 2. The percentages of variation explained were obtained by averaging the percentages of temporal variability explained at each gridpoint. Therefore in this case spatial error variation has been removed from the analysis although the past mean error predictor should still provide information about smoothly varying temporal error variability.

An examination of Table 2 indicates that the ensemble spread explains significantly more temporal variability in the forecast error than the other two predictors for both variables at all three lead times. This suggests that the relatively high percentage of the error variability explained by the mean past error in Table 1 is due to purely spatial variability in the forecast errors.

5 Predicting spatial error covariance

Forecast products are now available that provide quantitative information about the uncertainty of predictions of particular variables at given locations for different lead times. Such information tells forecast users about *univariate* errors, that is, one dimensional probability distributions. For some decisions, however, more sophisticated decision-making strategies require information about *multivariate* errors, or information about error covariances. For example, if a power company has two major power consuming cities in its region it would be helpful to know the likely magnitude of the errors in the temperature forecasts for these cities as these errors will translate into errors in predicted energy demand. Furthermore, knowing the *covariance* of these errors would be useful. If the errors are positively correlated then if it is colder than expected in one city it is also likely to be colder than expected in the other city, implying greater overall demand in the region. On the other hand, if the errors are negatively correlated then if it is colder than expected in one city it is likely to be warmer than expected in the other city, suggesting that the demand errors for each city will, at least partly, compensate for each other and result in a smaller error in the overall demand. Forecast users whose utility depends on the weather at multiple locations may benefit from information concerning the spatial covariance of forecast errors.

The relative usefulness of the ECMWF ensemble, and of recent errors of the control forecast, for predicting error covariances was compared. The use of recent error covariances essentially provides a baseline predictor which should be capable of capturing persistent structure in the error covariance field, including the fact that error covariances of widely separated points is likely to be zero. If the ensemble covariance outperforms the historical predictor then it must be predicting more than relatively stable structures in the error covariance field and must also be predicting features in the error covariance field that vary from forecast to forecast. The local gradient of the predicted field was not used as a predictor for this part of the study as it is not clear how this relatively simple quantity might be related to spatial error covariances.

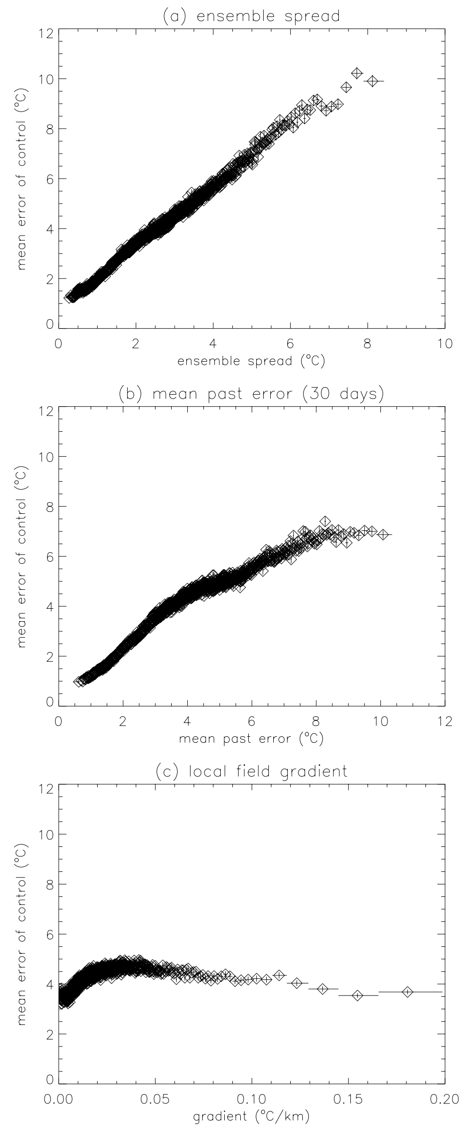

Fig. 7. As Fig. 5 but for a lead time of 240 h. First forecast was issued on 10 February 2004.

Let c_i be the value of the control forecast at gridpoint i, and let v_i be the value of the verification at gridpoint i. Furthermore, let $e_{i,k}$ be the value of the k^{th} ensemble member at gridpoint i. The actual error covariance between gridpoints i and j can be defined as

$$\mathrm{COV(i,j)_{act}} = (\mathrm{v_i} - \mathrm{c_i}) \cdot (\mathrm{v_j} - \mathrm{c_j}) \tag{6}$$

The error covariance of an N-member ensemble can be defined as

$$\mathrm{COV(i,j)_{ens}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{k=1}}^{\mathrm{N}} (\mathrm{e_{i,k}} - \mathrm{c_i}) \cdot (\mathrm{e_{j,k}} - \mathrm{c_j}) \tag{7}$$

The mean past error covariance was defined as the mean value of $\mathrm{COV(i,j)_{act}}$ averaged over the most recent 30 forecasts for which verifications would be available.

Table 1. The percentages of *spatio-temporal* variation in forecast error magnitudes explained by individual linear fits to the three predictors used in this study, and to a linear combination of past error and local gradient and to all three predictors combined. These percentages were determined across all gridpoints and thus include both spatial and temporal variations. The standard deviations were estimated by bootstrap resampling of the data, with replacement. The data were resampled in blocks of 15° in longitude ×15° in latitude and 5 days in time, this was to account for spatial and temporal correlations in the forecast error.

forecast variable	lead time	ensemble spread	mean past error	local gradient	2 predictors	3 predictors
500 hPa GP	96 h	18.4 ± 0.5	7.2 ± 0.1	6.1 ± 0.1	11.2 ± 0.2	19.1 ± 0.5
500 hPa GP	168 h	16.3 ± 0.6	9.8 ± 0.5	3.7 ± 0.2	11.6 ± 0.4	16.9 ± 0.5
500 hPa GP	240 h	16.0 ± 0.7	8.4 ± 0.3	2.6 ± 0.3	9.7 ± 0.4	16.1 ± 0.7
2 m TMP	96 h	14.7 ± 0.7	13.1 ± 0.4	1.7 ± 0.1	13.9 ± 0.4	20.7 ± 0.1
2 m TMP	168 h	17.6 ± 0.2	12.7 ± 0.6	0.7 ± 0.04	13.0 ± 0.6	19.7 ± 0.4
2 m TMP	240 h	19.2 ± 1.1	12.2 ± 0.5	0.3 ± 0.01	12.4 ± 0.5	19.9 ± 1.0

Table 2. The percentages of *temporal* variation in forecast error magnitudes explained by individual linear fits to the three predictors used in this study, and to a linear combination of past error and local gradient and to all three predictors combined. The percentage of variation at each gridpoint was determined separately and the mean of these percentages is given. This means that purely spatial error variation (the fact that some gridpoints have smaller errors than other gridpoints) is not included in these estimates. The standard deviations were estimated by bootstrap resampling of the data, with replacement. The data were resampled in blocks of 15° in longitude × 15° in latitude and 5 days in time, this was to account for spatial and temporal correlations in the forecast error.

forecast variable	lead time	ensemble spread	mean past error	local gradient	2 predictors	3 predictors
500 hPa GP	96 h	15.9 ± 0.6	3.5 ± 0.3	7.7 ± 0.3	4.4 ± 0.3	17.3 ± 0.6
500 hPa GP	168 h	12.2 ± 0.5	4.7 ± 0.4	5.6 ± 0.3	5.1 ± 0.3	13.6 ± 0.5
500 hPa GP	240 h	11.5 ± 0.6	4.2 ± 0.3	4.9 ± 0.1	4.7 ± 0.3	13.1 ± 0.6
2 m TMP	96 h	10.8 ± 0.6	6.2 ± 0.3	3.6 ± 0.2	8.6 ± 0.3	15.5 ± 0.5
2 m TMP	168 h	11.9 ± 0.6	5.6 ± 0.4	2.2 ± 0.1	6.8 ± 0.4	13.9 ± 0.5
2 m TMP	240 h	12.3 ± 0.5	5.6 ± 0.4	1.6 ± 0.05	6.3 ± 0.4	13.6 ± 0.5

As with conventional, univariate, spread-skill relationships there is large amount of scatter when actual error covariances are plotted against ensemble covariances, or mean past covariances. As before, the mean actual covariances of 1000 forecasts with similar values of the relevant predictor were plotted against the mean value of the predictor. These plots are shown in Figs. 8–13. As in Figs. 2–7 the horizontal bars indicate the ranges of predictor values of the forecasts that contributed to each mean. The plots in Figs. 8–13 where constructed using a random sample of 1000 pairs of gridpoints.

The results shown in Figs. 8–13 suggest that the ensemble error covariance is a good predictor of the likely actual error covariance. This is the case for all three lead times examined and both the 500 hPa geopotential and the 2-m temperature. The mean past error covariance provides some information concerning the actual error covariance, but in general this predictor does not span as large a range as the ensemble error covariance and thus does not provide as much predictive information. In particular, the mean past error covariance has little ability to predict strongly negative error covariances in the same way as the ensemble error covariance. This suggests that the expectation of negative error covariances is somewhat transient, and therefore averaging over 30 previous forecasts does not lead to strongly negative error covariances. The ensembles, however, are better able to capture the possibility of negative error covariances based upon the specific meteorological conditions.

To quantify the relative amount of information concerning error covariances in the ensemble and in past error covariances, a linear model was used. This model is given by

$$\widehat{\mathrm{COV}}(i,\, j)_{act} = A \cdot \mathrm{COV(i, j)_{ens}} + \mathrm{B} \cdot \mathrm{COV(i, j)_{past}} + \mathrm{C} \quad (8)$$

where $\mathrm{COV(i, j)_{past}}$ represents the mean error covariance of the 30 most recent daily forecasts for which verifications would be available.

The percentages of variation in the error covariance explained by the two predictors used in Eq. (8) are shown in

Table 3. The percentages of variation in forecast error covariance explained by linear fits to the ensemble covariance and the mean error covariance of the same two gridpoints averaged over the most recent 30 forecasts for which verifications would be available. The estimates of the percentages and the standard deviations were estimated by bootstrap resampling 1000 random pairs of points at 5 day intervals. This was done to account for temporal correlations in forecast errors.

forecast variable	lead time	ensemble covariance	mean past covariance	both predictors
500 hPa GP	96 h	7.9 ± 0.2	2.5 ± 0.1	8.3 ± 0.2
500 hPa GP	168 h	16.6 ± 0.3	3.9 ± 0.1	16.7 ± 0.3
500 hPa GP	240 h	22.8 ± 0.3	5.0 ± 0.1	22.9 ± 0.3
2 m TMP	96 h	4.8 ± 0.3	4.1 ± 0.1	7.9 ± 0.3
2 m TMP	168 h	9.4 ± 0.4	2.9 ± 0.1	10.4 ± 0.4
2 m TMP	240 h	15.2 ± 0.3	3.1 ± 0.1	15.7 ± 0.3

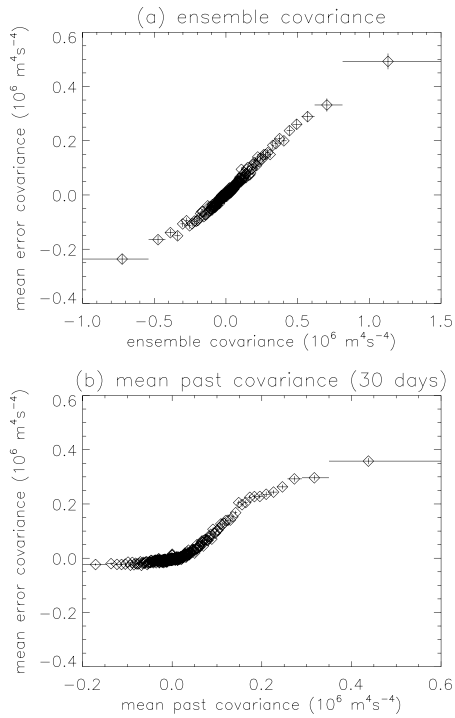

Fig. 8. Plots of the mean error covariance, $\langle (v_i - c_i) \cdot (v_j - c_j) \rangle$, where v and c denote the verification and control, and i and j denote gridpoints. The averages were calculated over blocks of 1000 forecasts with similar values of the predictor variables: **(a)** the error covariance between points i and j averaged over the ECMWF ensemble, and **(b)** the error covariance between points i and j averaged over the most recent 30 forecasts for which verifications would be available. The horizontal bars represent the ranges of the values of the predictors for each of the 1000 forecasts used to calculate each mean error, The vertical bars represent the standard error in the estimate of the mean forecast error. The forecasts used were for 500 hPa geopotential over North America at a lead time of 96 h initialized on each of the 322 days starting with 4 February 2004.

Table 3. The mean percentages and their standard deviations were estimated by bootstrap resampling, with replacement, the 1000 random pairs of gridpoints and only using forecasts at 5 day intervals to account for possible temporal correlations in forecast error covariances. Due to spatial correlations in error covariance, if all pairs of gridpoints had been used this would have resulted in an underestimate of the uncertainties in Table 3.

The results in Table 3 indicate that in the case of 500 hPa geopotential the ensemble error covariance is a significantly better predictor than the mean past error covariance. The ensemble error covariance alone is almost as good a predictor as a linear combination of both predictors. Mean past error covariances do relatively better at predicting spatial error covariances in 2-m temperature, but they are still outperformed by the ensemble error covariances. For both variables, the extra information that the ensemble provides over what is provided by past control forecasts and verifications is greater for spatial error covariances than for the magnitude of forecast error.

6 Summary and conclusions

This paper has compared three potential predictors of weather forecast error: the spread of an ensemble of multiple NWP simulations, the mean past error over the previous month at the location in question, and the local spatial gradient of the predicted variable in the control forecast. The ensemble spread is quite an "expensive" predictor in that it requires multiple runs of a numerical forecast model. The other two predictors can be cheaply calculated even when no ensemble is available.

It was found that, for the ECMWF medium-range ensemble forecasts of 500 hPa geopotential and 2-m temperature over North America, the spread of the ensemble was the best single predictor of the magnitude of forecast error. The mean past error, however, also explained a substantial amount of the variation in forecast error, especially in the case of 2-m temperature where it was almost as good a predictor as ensemble spread. As the mean past error was calculated for each gridpoint individually, it incorporates information concerning the spatial dependence of forecast error, as well some information about the seasonal variation in forecast error magnitudes. When discussing the value of ensembles for predicting forecast error it is important to consider how much of the variation in predictability is spatial or seasonal as such variation can be estimated using methods that are computationally cheaper than ensemble forecasting. The local gradient of the forecast variable in the control forecast was found to contain some information about forecast errors. This particular predictor is likely to be most effective when the forecast errors are the result of distortions or displacements of weather patterns in the forecast relative to the verification. The gradient was a better error predictor for the

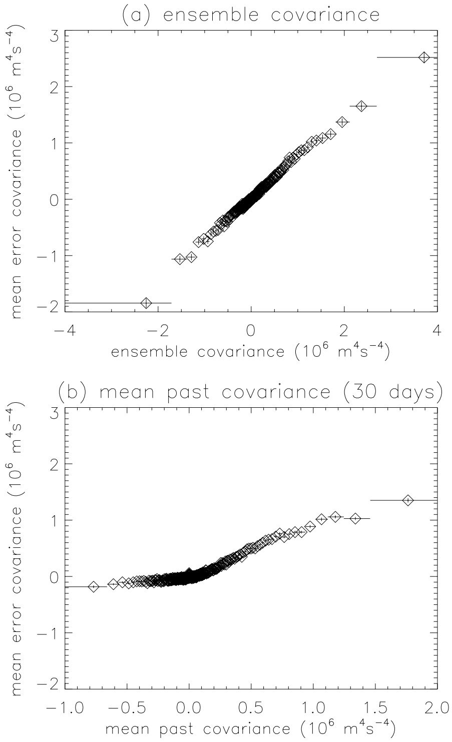

Fig. 9. As Fig. 8 but for a lead time of 168 h. First forecast was issued on 7 February 2004.

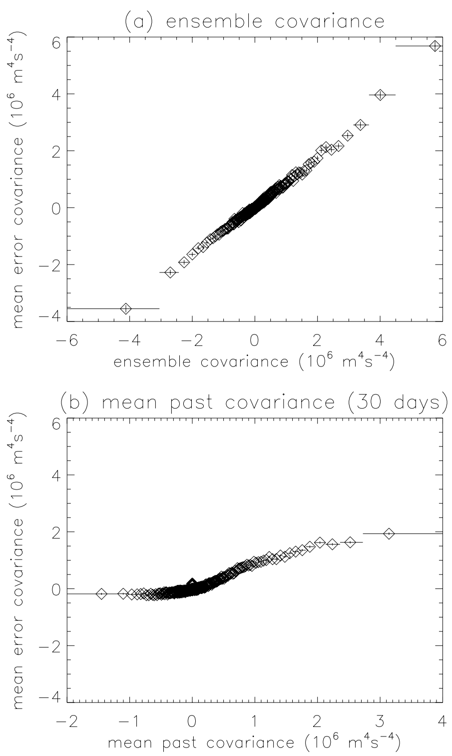

Fig. 10. As Fig. 8 but for a lead time of 240 h. First forecast was issued on 10 February 2004.

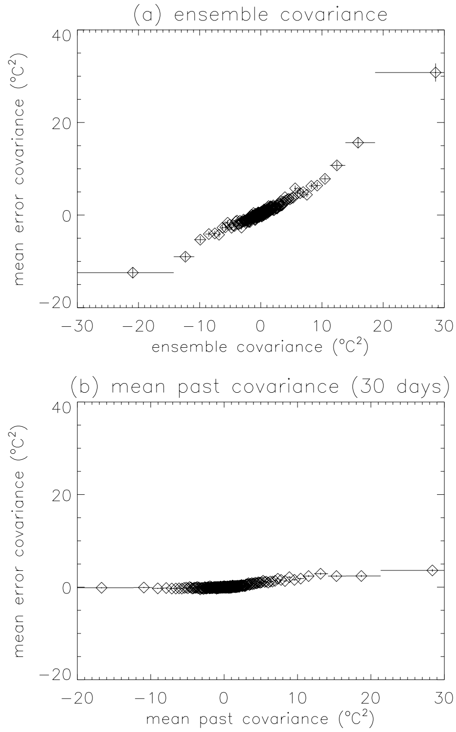

Fig. 11. Plots of the mean error covariance, $\langle (v_i - c_i)\cdot(v_j - c_j)\rangle$, where v and c denote the verification and control, and i and j denote gridpoints. The averages were calculated over blocks of 1000 forecasts with similar values of the predictor variables: **(a)** the error covariance between points i and j averaged over the ECMWF ensemble, and **(b)** the error covariance between points i and j averaged over the most recent 30 forecasts for which verifications would be available. The horizontal bars represent the ranges of the values of the predictors for each of the 1000 forecasts used to calculate each mean error, The vertical bars represent the standard error in the estimate of the mean forecast error. The forecasts used were for 2-m temperature over North America at a lead time of 96 h initialized on each of the 322 days starting with 4 February 2004.

500 hPa geopotential than the 2-m temperature, for which it contained very little information. It is possible that refinement of this predictor could improve its performance. For example, the spatial structure of distortions and displacements could be characterized and used to generate relatively computationally cheap ensembles by deforming the fields in a single NWP simulation.

This paper has also examined the potential value of ensemble error covariance and mean past error covariances as predictors of spatial error covariance. While the mean error covariance between two given gridpoints over the previous 30 forecasts was found to provide some information about likely forecast error covariance, the ensemble error covariance was found to be a significantly better predictor.

In this paper, it has been demonstrated, that the ECMWF medium-range ensemble provides predictive information about the likely magnitude of forecast errors, more so than the two predictors with which it was compared. Furthermore,

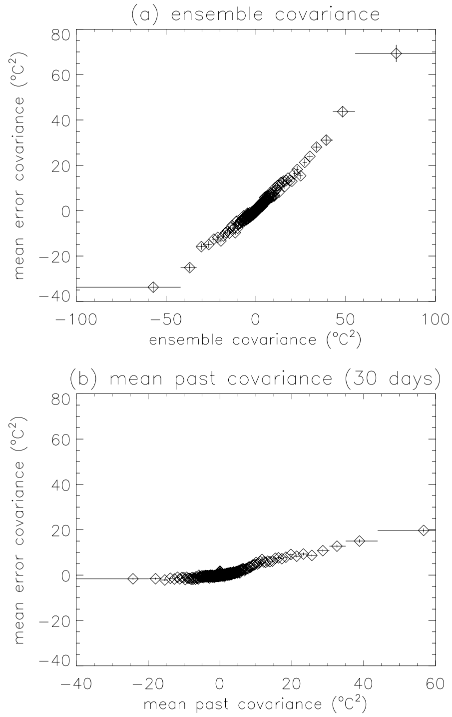

Fig. 12. As Fig. 11 but for a lead time of 168 h. The first forecast was issued on 7 February 2004.

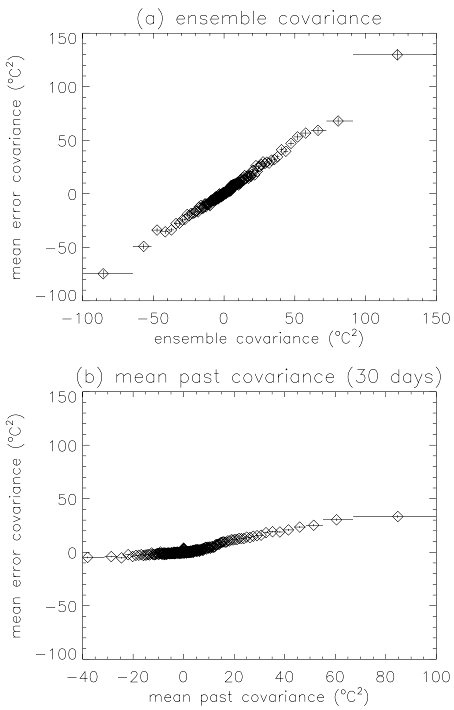

Fig. 13. As Fig. 11 but for a lead time of 240 h. The first forecast was issued on 10 February 2004.

the ECMWF ensemble also provides predictive information concerning likely spatial error covariances, and in this respect its marginal information content compared to a predictor based on past errors is even greater. The other predictors of forecast error – mean past errors and local field gradients – do in some cases, however, provide some predictive information. This information can be exploited when an ensemble is not availble, or it can be used to supplement the information provided by the ensemble to yield a better prediction of likely forecast error.

Acknowledgements. The author wishes to thank ECMWF for providing the forecast data and R. Hagedorn for extracting it from the ECMWF archive and for providing useful comments. Two anonymous reviewers also provided suggestions that significantly improved this paper.

Edited by: O. Talagrand
Reviewed by: two referees

References

Epstein, E. S.: Stochastic dynamic prediction, Tellus, 21, 739–759, 1969.

Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldman, T.: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation, Mon. Wea. Rev., 133, 1098–1118, 2005.

Hoffman, R. N., Liu, Z., Louis, J.-F., and Grassotti, C.: Distortion representation of forecast errors, Mon. Wea. Rev., 123, 2758–2770, 1995.

Houtekamer, P. L.: Global and local skill forecasts, Mon. Wea. Rev., 121, 1834–1847, 1993.

Houtekamer, P. L. and Derome, J.: Methods for ensemble prediction, Mon. Wea. Rev., 123, 2181–2196, 1995.

Jewson, S., Brix, A., and Ziehmann, C.: A new parametric model for the assessment and calibration of medium-range ensemble temperature forecasts, Atmos. Sci. Lett., 5, 96–102, 2004.

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

Liljas, E. and Murphy, A. H.: Anders Ångström and his early papers on probability forecasting and the use/value of weather forecasts, Bull. Amer. Meteor. Soc., 75, 1227–1236, 1994.

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

Lorenz, E. N.: A study of the predictability of a 28-variable atmospheric model, Tellus, 17, 321–333, 1965.

Lorenz, E. N.: The predictability of a flow which possesses many scales of motion, Tellus, 21, 289–307, 1968.

Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T.: The ECMWF ensemble prediction system: Methodology and validation, Q. J. R. Meteorol. Soc., 122, 73–119, 1996.

Orrell, D., Smith, L., Barkmeijer, J., and Palmer, T. N.: Model error in weather forecasting, Nonlin. Processes Geophys., 8, 357–371, 2001,
SRef-ID: 1607-7946/npg/2001-8-357.

Palmer, T. N.: Predicting uncertainty in forecasts of weather and climate, Rep. Prog. Phys., 63, 71–116, 2000.

Palmer, T. N.: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades, Q. J. R. Meteorol. Soc., 128, 747–774, 2002.

Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblas-Reyes, E., Jung, T., and Leutbecher, M.: Representing model uncertainty in weather and climate prediction, Ann. Rev. of Earth and Plan. Sci., 33, 163–193, 2005.

Stensrud, D. J., Brooks, H. E., Du, J., Tracton, M. S., and Rogers, E.: Using ensembles for short-range forecasting, Mon. Wea. Rev., 127, 433–446, 1999.

Tennekes, H., Baede, A., and Opsteegh, J.: Proc. ECMWF Workshop on predictability, chap. Forecasting forecast skill, ECMWF, Reading, UK, 1987.

Theis, S. E., Hense, A., and Damrath, U.: Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach, Meteorol. Appl., 12, 257–268, 2005.

Toth, Z. and Kalnay, E.: Ensemble forecasting at NCEP and the breeding method, Mon. Wea. Rev., 125, 3297–3319, 1997.

Whitaker, J. S. and Loughe, A. F.: The relationship between ensemble spread and ensemble mean skill, Mon. Wea. Rev., 126, 3292–3302, 1998.

Zhu, Y., Toth, Z., Wobus, R., Richardson, D., and Mylne, K.: The economic value of ensemble-based weather forecasts, Bull. Amer. Meteor. Soc., 83, 73–83, 2002.

Ziehmann, C.: Skill prediction of local weather forecasts based on the ECMWF ensemble, Nonlin. Processes Geophys., 8, 419–428, 2001,
SRef-ID: 1607-7946/npg/2001-8-419.