Tellus (1993), 45A, 424–434
Printed in Belgium – all rights reserved

TELLUS
ISSN 0280–6495

Validation of a skill prediction method

By JAN BARKMEIJER*, PETER HOUTEKAMER‡ and XUELI WANG, *Royal Netherlands Meteorological Institute, PO Box 201, 3730 AE De Bilt, The Netherlands*

(Manuscript received 3 November 1992; in final form 11 February 1993)

ABSTRACT

A large experiment is performed to validate two predictors for the quality of ECMWF forecasts over Western Europe. One predictor yields the spread of the probability distribution for the error in the predicted 500 hPa geopotential height. It is determined by the trace of the covariance matrix for the geographically local forecast error. In addition the spread for the 500 hPa vorticity error at a location near the Netherlands is computed. The local covariance matrix, necessary for determining both predictors, is computed for 607 days, using the tangent linear and an adjoint version of a quasi-geostrophic 3-level model with truncation T21. We assume linear error growth and the absence of model errors. The forward reference orbit is obtained by interpolating actual ECMWF forecasts with the 3-level model. Small values of the predictor imply small error growth, and therefore accurate forecasts. Large values may or may not be associated with large actual forecast errors, depending on whether the initial error strongly projects on the fastest growing modes. How the uncertainty in the structure of the initial error influences the performance of the skill predictor is studied by considering three different covariance matrices for the initial error. Validation of the predicted variance with the 2-day and 3-day ECMWF forecast error shows that for all initial covariance matrices, both predictors provide significant information about the quality of the forecast. In case of small and large predicted variance, the probabilities for small and large prediction errors are 10% higher than the climatological probabilities. Projection of the observed forecast error onto the eigenvectors of the local covariance matrix indicates that a few eigenvectors already describe a large portion of the forecast error.

1. Introduction

One of the intriguing aspects of present weather forecasting is the variability in the quality of the forecast. Sometimes a forecast is accurate for up to a week but it also occurs that a forecast has lost all value after one or two days.

In general, it is difficult to decide whether forecast errors originate from initial errors or model errors (Tribbia and Baumhefner, 1988). Studies by Lorenz (1982, 1990) for the ECMWF model indicate that between 1981 and 1987, both the initial error and model error have decreased simultaneously. Between 1987 and 1989, the major improvements were in the model. At some point, it will become difficult to improve the model further, because the forecast error will be more and more caused by errors in the initial state. This makes it difficult to detect errors in the model. On the other hand, there is no point in reducing the initial error, while significant imperfections in the model still exist (Boer, 1984). Therefore, it is natural to expect that model errors and initial errors will remain of comparable magnitude (Lorenz, 1990) and, consequently, skill predictions should involve the study of initial errors as well as model errors. Part of the variability in skill can be explained as a purely

* Corresponding author.
‡ Present affiliation: Recherche en Prévision Numérique, 2121 Route Trans-Canadienne, Dorval, Québec H9P 1J3, Canada.

statistical effect (Kruizinga and Kok, 1988). This is because the initial error partly results from random observational errors. The forecast error will therefore also have a random component. Thus, one particular forecast must be considered as an arbitrary member of a probability distribution.

Lorenz (1965) proposed the Monte Carlo method for studying the quality of forecasts. A small ensemble of randomly-chosen initial states is integrated with the forecast model. The spread between the ensemble members is considered as a measure of the atmospheric predictability. However, the complexity of the current forecast models greatly obstructs the operational use of the Monte Carlo method. A more fundamental problem here is, how to obtain a good statistical description of the initial errors. Such a description is necessary for selecting a representative sample. All current skill prediction methods suffer from the lack of knowledge on the statistics of the initial error. The ECMWF ensemble prediction scheme (Palmer et al., 1992) uses analysis error variance fields.

As an alternative to Monte Carlo forecasting, Hoffman and Kalnay (1983) introduced lagged average forecasting. Here, the forecast ensemble consists of the latest operational forecast together with forecasts based on previous analyses but with the same verification time. By weighting the members, the lagged average forecast is obtained. This procedure may yield a more accurate forecast and because all forecasts in the ensemble are already available, there are almost no additional computational costs. Unfortunately, several studies, e.g., Dalcher et al. (1988) and Branković et al. (1990), indicate that the correlation between the forecast skill and the ensemble spread is not very high. Possible explanations for this may be the small size of the ensemble, unrealistic ensemble properties or too large verification areas. In the latter case, regional differences in predictability are not properly taken into account.

A major development in meteorology was the introduction of adjoint models (Marchuk, 1974). It took some time before this technique received serious attention. Early applications of adjoint equations are Hall et al. (1982), on a sensitivity study of model output to many parameters, and Talagrand and Courtier (1987), on the use of the adjoint method in data assimilation. The great advantage of adjoint models is that one can easily compute gradients of multi-variable functions. In predictability research, this property is employed to study error growth in an efficient manner (Errico and Vukićević, 1992; Barkmeijer, 1992). Here, one is interested in determining the sensitivity of one aspect of the forecast, such as local pressure, to uncertainties in the initial conditions. Using an adjoint model it is possible to retrieve this sensitivity pattern in the initial condition by integrating the adjoint model only once backwards in time. Thus, it is not necessary to integrate the forecast model with a perturbation once for each coordinate in order to obtain the impact of an arbitrary initial perturbation. From this, it is clear that adjoint methods are useful in computing the patterns that give the maximal forecast error in a certain area at a pre-chosen forecast time. These sensitivity patterns are usually referred to as optimal modes (Farrell, 1989; Zhang, 1988). At a fixed forecast time, they show a larger error growth than the exponentially growing normal modes. Molteni and Palmer (1992) and Mureau et al. (1992) employ these optimal modes for Monte Carlo forecasting. In one of the presented cases, some members of the ensemble captured the development of a ridge, which was not predicted in the operational forecast.

A basic assumption in using adjoint models is that the error dynamics is linear. Studies by Lacarra and Talagrand (1988) and Vukićević (1991) indicate that this is the case for forecast periods up to 48 hours. In Barkmeijer and Opsteegh (1992), a predictor for regional forecast skill is suggested which can benefit from the use of adjoint models. The predictor of the skill is the maximum likely amplitude of the squared error in the streamfunction field at 500 hPa for a pre-chosen area and forecast period. It is given by the largest eigenvalue of the covariance matrix of the local forecast error. Results indicate that this predictor gives significant information about the quality of the local forecast.

In view of this, we presently want to perform a similar study for a much larger number of cases. Again, we are concerned with the quality of short-range forecasts, so we will assume that the error growth is linear. As a further simplification, we omit the contribution of model errors to the forecast error, so the forecast error entirely originates from uncertainties in the initial state. Houtekamer (1992) showed, for a very simple model, the importance of an accurate description of the initial error statistics. Unfortunately, the present knowledge on this subject is not very good. We investigate the impact of three different covariance matrices of the initial error on the performance of the skill predictor. In the pilot experiment (Barkmeijer and Opsteegh, 1992), we were able to describe a large portion of the forecast error with only the fastest growing mode. In the present study, we will give skill forecasts for a larger area. We therefore use as a predictor of the forecast skill, the sum of the eigenvalues of the covariance matrix for the local forecast error thus including all modes. This predictor is equal to the sum of the variances in geopotential height for each gridpoint of the area in consideration. We also present a predictor of the vorticity error near the Netherlands. In Section 2, we give an outline of the techniques we use. The validation of the method as a possible skill predictor is discussed in Secion 3. Finally, in Section 4 we give some concluding remarks.

2. The model and the method

In validating a skill prediction method, it is necessary to perform a large experiment. The experiment that we will describe is based on 607 forecasts. Therefore, it is important that the numerical costs of the method are as low as possible. For the prediction of the skill, we use a 3-level quasi-geostrophic model triangularly truncated at wavenumber 21. For a description of the model, we refer to Molteni and Palmer (1992). We will denote this model by T21QG. In order to perform the experiment, we developed the tangent linear and adjoint T21QG models. The levels of the model are at 200 hPa, 500 hPa and 800 hPa. Since the numerical costs of running these models are low, we are able to perform a large skill experiment.

The period for which we validate the predictors of skill consists of 5 extended winters: 15 November–15 March of the years 1987-88 to 1991-92. This gives a total of 607 days. For each day, we retrieve from the ECMWF archives the analysis at 1200 UTC and the forecast each 12 h up to a forecast period of 72 h. The data for the 800 hPa model level is obtained by linearly interpolating 700 hPa and 850 hPa data.

The quality of the 12 hours T21QG forecast is such that we may use it to define a forward orbit which closely follows the ECMWF forecast orbit. To that end, we interpolate the archived ECMWF forecast, made for a particular day, with T21QG. Starting with an analysis at 1200 UTC, the model T21QG is integrated for 12 h. Then in the final time step, the obtained field is replaced by the corresponding 12 h ECMWF forecast. With this forecast as new initial field, we again integrate T21QG for 12 h replacing in the final time step the field by the 24-h ECMWF forecast. This procedure is repeated until the forecast period is reached. The forward orbit thus obtained may capture some of the ECMWF model dynamics. It is along this reference orbit that we will study the evolution of the initial error.

The contribution of model errors to the forecast error is not included in the skill prediction method. As a further simplification, we assume that the error dynamics is linear. This is reasonable for forecast periods up to 48 h. The integration of an initial error $\varepsilon(0)$ to time T of the forecast is done with the tangent linear T21QG model. This gives rise to a linear operator $R(0, T)$ such that:

$$\varepsilon(T) = R(0, T)\,\varepsilon(0). \tag{1}$$

For convenience, we leave from now on the time dependence of R from the notation. We point out that R has to be determined each day because it depends on the daily varying reference orbit.

2.1. Structure of the initial error

We have not yet specified the structure of the initial error $\varepsilon(0)$. One would like to use the full covariance matrix of the initial error which can be determined with a Kalman filter. However, Kalman filters are not yet used operationally. It is not clear whether optimal interpolation schemes give an initial covariance matrix which is good enough for the purpose of making skill predictions (Bouttier, 1992). In the present paper, we validate the performance of the skill prediction method for three different covariance matrices Q of the initial error.

Given Q, an ensemble of random perturbations can be formed easily. For that one performs an eigenvector analysis of the matrix Q:

$$Q = BB^{t}. \tag{2}$$

The columns of B contain the eigenvectors of the matrix Q multiplied with the square root of the corresponding eigenvalue. An ensemble member can be generated by a multiplication of a random vector d with the matrix B:

$$\varepsilon(0) = Bd. \tag{3}$$

The elements of the vector d have independent gaussian probability distributions with expectation value zero and unit variance. In the following, such a distribution is denoted with $N(0, 1)$. The use of a Gaussian distribution would be implied by Gaussian statistics of the observational errors and a linear response of the data assimilation to these errors.

As a 1st choice for the covariance matrix, we let Q be the identity matrix. This means that the errors in the different spectral modes are initially uncorrelated. Their squared amplitudes have the same expectation value. This constraint on $\varepsilon(0)$ can be interpreted as a white noise error distribution. A surface of equal probability density is now defined by the sphere $\{\|\varepsilon(0)\|_0 = \text{constant}\}$, where

$$\|\varepsilon\|_n^2 = \int \nabla^n \varepsilon \cdot \nabla^n \varepsilon \, \mathrm{d}s. \tag{4}$$

We will refer to this first choice for the initial error statistics with SN.

As a 2nd choice, we assume that the error in the different spectral modes have the same expectation value for their kinetic energy. A surface of equal probability density is now given by the ellipsoid $\{\|\varepsilon(0)\|_1 = \text{constant}\}$. Also in this case, the different spectral modes are initially uncorrelated, so the matrices Q and B are diagonal. This case will be referred to with KE. The kinetic energy constraint states that the initial error is in the large scales rather than in the small scales. This type of errors may be called red noise.

As a final choice we use the covariance matrix Q which was previously determined by one of the authors (Houtekamer, 1993). This matrix is obtained using a variational method, a simulated radiosonde network, background error statistics and the T21QG model for the data assimilation. The error statistics are corrected for model error after the assimilation. In this, the method differs from a Kalman filter. Covariances were determined for the entire month of December 1989. It appeared that the day-to-day variability in this matrix Q was not very large. We therefore decided to use the matrix for 1200 UTC 8 December 1989 for the entire 607 day period. Both matrices Q and B are now full matrices. We refer to this last case with COV. It is characterized by an equivalent barotropic structure of the errors with large amplitudes over the oceans.

2.2. *Local forecast error*

Apart from using a time-independent covariance matrix for the initial error, the computational costs can be reduced still further, by performing skill forecasts for a restricted area. In the experiment, we are interested in the quality of forecasts in an area at 500 hPa, consisting of 23 gridpoints.

These gridpoints, defined on a Gaussian grid, are separated thus far that the local errors are not very much correlated, see Fig. 1. We reduce the dimension of the forecast error with a projection operator P defined by:

$$\eta(T) = P\varepsilon(T). \tag{5}$$

The projection operator P depends on the meteorological quantity one is interested in. Here, $\varepsilon(T)$ is the global error field in streamfunction coordinates at the various model levels for forecast time T and $\eta_i(T)$ denotes the error in geopotential height in one of the 23 gridpoints at 500 hPa. The conversion from streamfunction field to geopotential height field is done with the linear balance equation.

2.3. *The predictors*

By using the linear operators R, B and P, we can write the forecast error (m^2) for the 23 gridpoint area in the following form:

$$\langle PRBd, PRBd \rangle = \langle B^* R^* P^* PRBd, d \rangle, \tag{6}$$

where $\langle\, , \rangle$ is the Euclidean inner product $\langle x, y \rangle = \sum x_i \cdot y_i$. Further, B^*, for instance, denotes the adjoint of B with respect to the Euclidean inner product. Since the adjoint of this operator B with respect to this inner product is equal to the transpose B^t, we have:

$$PRB = (B^t R^t P^t)^t = (B^* R^* P^*)^t. \tag{7}$$

The 23 non-zero eigenvalues of the matrix $B^* R^* P^* PRB$ are identical to the eigenvalues of the 23 × 23 matrix $Q_{\mathrm{loc}} = PRQR^*P^*$. For the Euclidean coordinate system, Q_{loc} is equal to the covariance matrix for the local forecast error as can be seen from:

$$\begin{aligned} Q_{\mathrm{loc}} &= \overline{P\varepsilon(t)(P\varepsilon(t))^{\mathrm{t}}} \\ &= \overline{PR\varepsilon(0)(PR\varepsilon(0))^{\mathrm{t}}} \\ &= PRQR^{\mathrm{t}}P^{\mathrm{t}}. \end{aligned} \tag{8}$$

We denote the mean for an ensemble of randomly chosen initial errors $\varepsilon(0) = Bd$ by a bar, where the elements d_i of d are $N(0, 1)$ distributed.

The matrix Q_{loc} can be obtained by applying $B^*R^*P^*$ to the 23 unit vectors in the low dimensional space and making use of eq. (7). Provided that B and P are known, the numerical costs to compute Q_{loc} are determined by integrating the adjoint model 23 times backwards in time. The eigenvalues $\{\mu_i^2,\ i = 1, \ldots, 23\}$ of $B^*R^*P^*PRB$ correspond with eigenvectors $\{v_i,\ i = 1, \ldots, 23\}$. Initial error fields Bv_i contain information with respect to areas that are sensitive to error growth. The corresponding spatial patterns of the forecast error for the restricted area is given by $PRBv_i$.

As one predictor for the quality of the local forecast we propose the trace of the local covariance matrix Q_{loc}. Note that the predictor is equal to the sum of the eigenvalues of Q_{loc}. Each element of the trace of Q_{loc} yields the spread in geopotential height for the corresponding grid point. Using a finite difference approximation we also compute from this local covariance matrix the spread of the vorticity at a location near the Netherlands. This method yields an estimate for the vorticity of a gridpoint based on the streamfunction values in the gridpoint itself and in four surrounding points. This estimated vorticity is highly correlated with the estimate obtained when using the geopotential values for these points. Therefore, we decided to use geopotential values. Let $\{g_i,\ i = 1, 2\}$ and $\{g_i,\ i = 3, 4\}$ be the geopotential height error in the surrounding points respectively in lateral direction and in the meridional direction and g_5 the geopotential height error in the middle point, see Fig. 1. Then

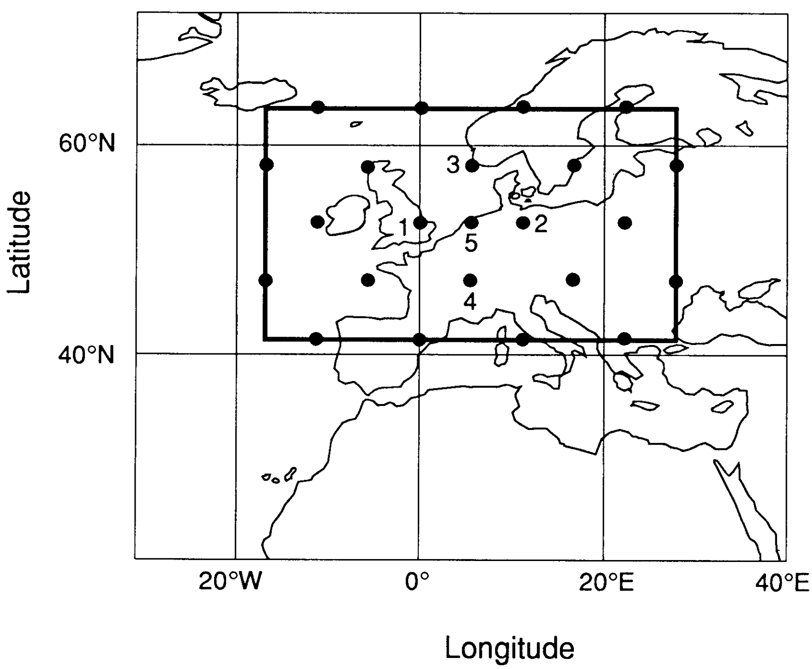

Fig. 1. Heavy lines denote the 23 gridpoint area for which the skill predictions are performed. The numbered gridpoints refer to the use of a finite difference approximation for computing the vorticity.

the estimated vorticity error $\tilde{\zeta}$, except for a fixed multiplication factor, is given by:

$$\tilde{\zeta} = \frac{g_1 - 2g_5 + g_2}{d_l^2} + \frac{g_3 - 2g_5 + g_4}{d_m^2}, \tag{9}$$

with d_l and d_m the gridpoint distance in lateral and meridional direction, respectively. Given Q_{loc}, the covariance matrix for the local forecast error in the geopotential height, one can extract the predicted variance in the vorticity error by using a projection P_1 acting on the 23 gridpoint area. This projection P_1 is simply the one defined by eq. (9): $P_1 \eta = \tilde{\zeta}$, where η is a vector of length 23 containing the error in geopotential height for each gridpoint. The 1×1 matrix $P_1 Q_{\text{loc}} P_1^*$ equals the variance of the error in the vorticity.

3. The experiment

The approach as described in the previous section enables us to perform a large validation experiment. For the 607 days, we compute 2-day and 3-day local skill forecasts starting from the 3 different covariance matrices for the initial error. We use contingency tables to investigate the performance of the two predictors.

For each day, we determine the interpolated forward orbit. Along this orbit, the adjoint model is integrated backwards in time. This eventually yields the covariance matrix $Q_{\text{loc}} = PRQR^*P^*$ of the local forecast error, with Q the covariance matrix for the initial error. The predicted variance (m^2) is equal to the trace of Q_{loc}. The real error is the forecast error in geopotential height squared for the 23 gridpoint area. It is obtained from the ECMWF forecast for this day and its verifying ECMWF analysis.

The first result, as depicted in Fig. 2a, shows the real error squared plotted against the predicted variance for the COV case. The forecast period is 2 days. For a forecast period of 3 days, the result is shown in Fig. 2b. Clearly, Figs. 2a, b reveal an increase in density of points towards the lower left corner. For larger values of the predicted variance, the distribution of the real error becomes broader. This is in accordance with the interpretation of the predictor as an upper limit to error growth.

We investigate whether the distribution of the predicted variance indeed gives significant information about the distribution of the real error by applying the χ^2-method to a contingency table. For each SN, KE and COV experiment, we construct a 3×3-contingency table. Let ε_3 be the maximal value of the predicted variance. Now choose ε_1 and ε_2, with $0 = \varepsilon_0 < \varepsilon_1 < \varepsilon_2 < \varepsilon_3$, such that each set $P_k = \{$days with $\varepsilon_k <$ predicted variance $< \varepsilon_{k+1}\}$ contains an equal number of days. The same is done for the set of real errors. This gives the 3 sets E_0, E_1 and E_2 for the real errors. The sets of predicted variances and the sets

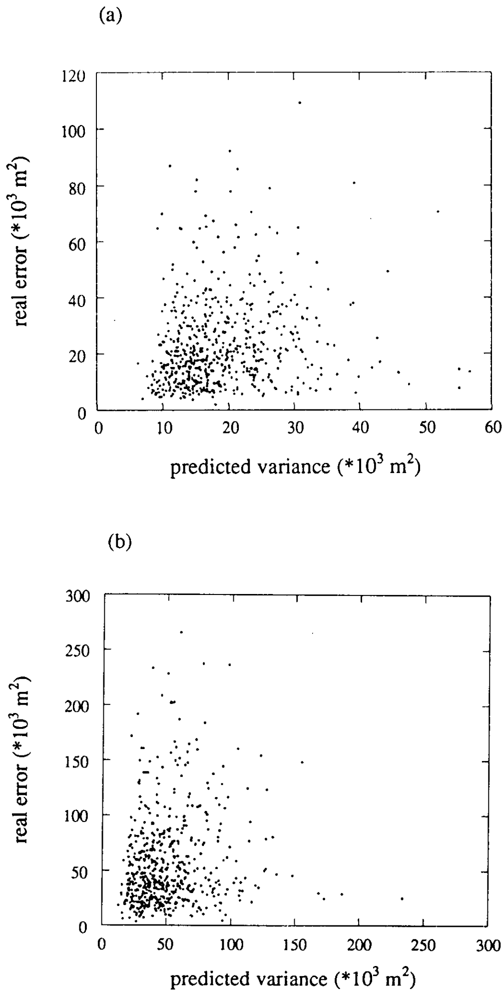

Fig. 2. For 607 days, the actual geopotential height error squared is plotted against the predicted variance for this error, with a prediction period of (a) 2 days and (b) 3 days.

Table 1. *The 3 × 3-contingency table for the COV experiment (trace).*

Actual forecast error	Predicted variance 2 days P_0	P_1	P_2	3 days P_0	P_1	P_2
E_2	24	33	43	28	29	42
E_1	32	35	32	29	38	31
E_0	44	31	24	42	31	25

of real errors define a 3 × 3-contingency table. Table 1 shows the contingency table for COV in case of 2-day and 3-day forecasts; numbers are given in %s. In the absence of round-off errors, both row and columns would add up to 100%.

If the predicted variance would tell us nothing about the real error, then the points in Fig. 2 would be homogeneously distributed and all entries in the table would be 33%. Fortunately, this is not the case. Especially in the extreme situations, the use of the predicted variance as predictor of the forecast quality results in an improvement of 10%. The χ^2 statistic may tell us in an objective manner whether the predicted variance indeed gives information about the quality of the real error. If the two distributions of predicted variance and real error have no association, then the expected number for all entries N_{ij} in the contingency table would be $n_{ij} = 607/9$. The χ^2 statistic is given by eq. (10) where the summation is over all entries in the contingency table.

$$\chi^2 = \sum_{i,j} \frac{(N_{ij} - n_{ij})^2}{n_{ij}}. \tag{10}$$

In Table 2, we present for all experiments the results of the χ^2 test with 4 degrees of freedom. The

Table 2. *The χ^2-test for the predicted variance (trace) for the cases SN, KE and COV. Between brackets the significance level is given.*

Forecast period	2 days	3 days
SN	$12.2(1.6 \cdot 10^{-2})$	$26.3(2.8 \cdot 10^{-5})$
KE	$9.9(4.3 \cdot 10^{-2})$	$16.6(2.3 \cdot 10^{-3})$
COV	$24.5(6.5 \cdot 10^{-5})$	$18.7(8.8 \cdot 10^{-4})$

significance level is given in parantheses. This is the probability to obtain higher values for χ^2 with random values of N_{ij}.

The χ^2-test shows in all 6 cases a significant performance of the skill prediction scheme. These results indicate that linear methods are still suitable for forecast periods up to 3 days. For short-range skill predictions, knowledge about the structure of the initial error is important, so we expect the best performance for the COV case and for 2-day forecasts it is. But all in all, it is difficult to decide which covariance matrix is the best choice for the skill predicion method. We also want to point out that the correlation between the predicted variances for SN, KE and COV yields values over 0.9 both for 2-day and 3-day forecasts. Apparently, variations in the predicted variance are mainly caused by the forecast orbit. The precise form of the initial covariance matrix seems to be of less importance for this prediction range.

In order to have some indication of the optimal performance of the skill prediction method, we assume now that the matrix Q_{loc} is a perfect prediction of the covariance matrix for the local forecast error. Let $\{E_i, i = 1, ..., 23\}$ be the set of eigenvectors with unit Euclidean norm and $\{\mu_i^2, i = 1, ..., 23\}$ the set of corresponding eigenvalues. Unfortunately, we do not know how the real error projects along the eigenvectors E_i. A particular forecast must be considered as a draw from a statistical distribution. From the predicted Q_{loc} we simulate a forecast error using random projections $d_i \mu_i$ on the eigenvectors. The random numbers $\{d_i, i = 1, ..., 23\}$ have zero mean and unit variance. Thus, the simulated forecast error is of the form $\sum_{i=1}^{23} d_i \mu_i E_i$. The squared length of the simulated forecast error is $\sum_{i=1}^{23} d_i^2 \mu_i^2$. We assume that the 607 computed matrices Q_{loc} are typical for the variability in the local covariances. In order to get reliable statistics, we compute for each of the 607 days 100 sets of random projections $\{d_i, i = 1, ..., 23\}$. This simulates 100 forecast errors that might have occurred. As before we use 3 × 3-contingency tables to study the association between the distributions of predicted variance and the thus obtained real errors. The contingency table for the COV case is listed in Table 3.

An interpretation of these optimal results is that the performance of the skill prediction method as shown in Table 1 is at an intermediate level. In computing the correlation between the predicted variance and the real error for these optimal experiments, we obtained for all cases and forecast periods a value of about 0.4. This correlation is 0.17 and 0.13 for the sets of points shown in Figs. 2a, b, respectively. It is not yet clear whether the assumption, that the projection coefficients d_i have independent $N(0, 1)$ distributions, is realistic.

The corresponding results for vorticity at a grid point near the Netherlands are listed in Table 4. The variance in the vorticity is computed from $Q_{\rm loc}$ using a finite difference approximation (see Subsection 2.3).

In comparing Table 4c with Table 3, we conclude that the variability of spread in the distribution of the vorticity error is smaller than for the geopotential error in case of an optimal skill forecast. This has consequences for the validation experiment. With smaller variations in the predicted variance, the skill forecast becomes more difficult to validate. The results for vorticity may also depend more on the smaller scales than is the case for geopotential height, and on that the smaller scales are poorly represented with the model used. Table 4b shows that, although the predictor still significantly informs about the vorticity error, the performance has decreased compared with the performance of the predictor for the geopotential error (see Table 2). We have no explanation as yet for the sudden drop in skill for the 3-day COV case.

In the above, we were only interested in two aspects of the forecast error. The eigenvectors $\{E_k,\ k = 1, ..., 23\}$ of $Q_{\rm loc}$ may tell us more about the structure of the forecast error. As before, we assume that the observed forecast error has the form $\sum_{k=1}^{23} d_k \mu_k E_k$, where d_k is $N(0, 1)$ distributed. We obtain the projection coefficients $d_k \mu_k$ from the observed ECMWF forecast error. The

Table 3. *The 3 × 3-contingency table for the optimal COV experiment (trace).*

Actual forecast error	Predicted variance: 2 days P_0	P_1	P_2	3 days P_0	P_1	P_2
E_2	18	29	53	17	29	55
E_1	32	36	32	31	37	32
E_0	49	35	16	52	34	14

Table 4a. *As Table 1 but now for the vorticity*

Actual forecast error	Predicted variance: 2 days P_0	P_1	P_2	3 days P_0	P_1	P_2
E_2	25	33	42	28	35	36
E_1	38	32	29	33	32	34
E_0	37	34	27	37	31	29

Table 4b. *Table 2 but now for the vorticity*

Forecast period	2 days	3 days
SN	$11.7(2.0 \cdot 10^{-2})$	$12.0(1.7 \cdot 10^{-2})$
KE	$9.9(4.3 \cdot 10^{-2})$	$12.4(1.4 \cdot 10^{-2})$
COV	$14.6(5.5 \cdot 10^{-3})$	$4.4(3.5 \cdot 10^{-1})$

Table 4c. *As Table 3 but now for the vorticity*

Actual forecast error	Predicted variance: 2 days P_0	P_1	P_2	3 days P_0	P_1	P_2
E_2	20	33	47	17	33	50
E_1	38	34	28	39	34	27
E_0	42	33	24	44	33	23

average projections OBS(k), $k = 1, ..., 23$ are then defined as follows:

$$\mathrm{OBS}^2(k) = \sum_{j=1}^{607} d_{kj}^2 \mu_{kj}^2 / 607. \tag{11}$$

This observed spectrum can be compared with the predicted spectrum PR(k), $k = 1, ..., 23$, given by:

$$\mathrm{PR}^2(k) = \sum_{j=1}^{607} \mu_{kj}^2 / 607. \tag{12}$$

If the projection coefficients d_k have indeed independent $N(0, 1)$ distributions, then the observed spectrum should resemble the predicted spectrum:

$$\begin{aligned}\mathrm{OBS}^2(k) &= \sum_{j=1}^{607} d_{kj}^2 \mu_{kj}^2/607 \\ &\approx \sum_{j=1}^{607} d_{kj}^2/607 \sum_{j=1}^{607} \mu_{kj}^2/607 \\ &\approx \sum_{j=1}^{607} \mu_{kj}^2/607 = \mathrm{PR}^2(k). \end{aligned} \tag{13}$$

Fig. 3 shows for 2-day and 3-day forecasts the predicted spectrum PR and the observed spectrum OBS for the COV case. We normalized the eigenvalues of Q_{loc} so that $\sum_{j=1}^{607} \sum_{k=1}^{23} \mu_{kj}^2/607 = 100$. For all SN, KE and COV experiments, the observed spectrum and the predicted spectrum differ in the same manner. It is worth noting that for all cases, the observed spectrum OBS decreases monotonically as the eigenvalues become smaller. There is a tendency for the observed error to have a smaller projection onto the fastest growing modes as predicted and a larger projection onto the contracting modes. It means that we do not perfectly predict the error patterns. Several causes may be due to this, like neglected processes such as model errors or imperfect description of the ECMWF errors with T21QG. Finally we remark that both spectra are quite steep. This suggests that a few eigenvectors are sufficient to describe a large portion of the forecast error.

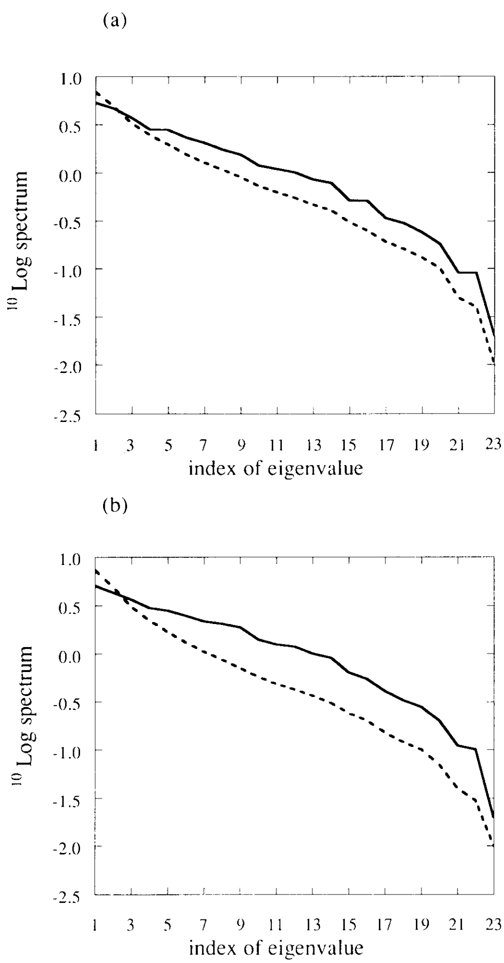

Fig. 3. The predicted spectrum PR (dashed line) and the observed spectrum OBS (solid line) are shown for COV, in case of (a) 2-day and (b) 3-day predictions. The 23 eigenvalues are sorted in decreasing order.

4. Concluding remarks

The numerical costs of running weather prediction models is keeping pace with the computational capacity. Operationally, this means that a weather prediction is based on only one model run. In such circumstances, some kind of predictor for the quality of the prediction may be useful for the forecaster.

It was our goal to develop a skill prediction scheme for the ECMWF model without the need for large computational investments. We therefore focused on the short-range predictions so that we could assume linear error growth. As a further simplification, we only considered the forecast error in a small area. This, together with the use of a relatively simple 3-level quasi-geostrophic model, its tangent linear and adjoint version, enabled us to determine the covariance matrix for the local forecast error in an efficient manner. From this matrix, we derived a predictor for the geopotential height error and a predictor for the vorticity at a certain location. We did not consider the contribution of model errors to the forecast error. We assumed that the internal error growth of the 3-level model not only resembled the internal error growth of the ECMWF model, but also determined the forecast error to a large extent.

In a large experiment, made possible by the above assumptions, we validated the skill prediction scheme. The results show that the method gives information about the quality of the

ECMWF forecast. In this paper, we used three different covariance matrices for the initial error. The idea is that the structure of the initial error becomes more and more important for decreasing forecast periods. The improvement that is gained when using a more realistic matrix (COV) seems to support this. One would expect that the structure of the initial error depends on the atmospheric circulation. We do not yet know the impact of a time- (or flow-) dependent covariance matrix. One of the authors (Houtekamer, 1992) demonstrated with a very simple model, that it is important to properly describe the changing statistics of the initial error. This result was not supported by a later study with the T21QG model (Houtekamer, 1993). Even the difference as large as between the 3 covariance matrices used in this study do not have the impact we had expected. So far, an explanation for this is still lacking.

Fig. 3 suggests that only a few modes are responsible for the forecast error. Thus, instead of determining the complete local covariance matrix, it may be sufficient to compute the fastest growing modes with the Lanczos algorithm. Finally, this experiment illustrates that the linear error growth assumption can still be used for ranges up to 3 days. At the moment, The Royal Netherlands Meteorological Institute (KNMI) is in the process of testing this method in an operational environment.

5. Acknowledgments

We would like to thank the members of the predictability group at KNMI, in particular Theo Opsteegh, for assistance during the course of this work. One of the authors (PLH) acknowledges the support of the working group on Meteorology and Physical Oceanography (MFO) with financial aid from The Netherlands Organization for the advancement of Research (NWO). Also thanks are due to two anonymous reviewers for their useful comments.

REFERENCES

Barkmeijer, J. 1992. Local error growth in a barotropic model. *Tellus 44A*, 314–323.

Barkmeijer, J. and Opsteegh, J. D. 1992. Local skill predictions for the ECMWF model using adjoint techniques. Proceedings of the Workshop on *New developments in predictability*, 1991. Reading, England, ECMWF, 55–64.

Boer, G. J. 1984. A spectral analysis of predictability and error in an operational forecast system. *Mon. Wea. Rev. 112*, 1183–1197.

Bouttier, F. 1992. *The dynamics of error covariances in a barotropic model.* Preprint Météo-France/CNRM, Toulouse, France, 45 pp.

Branković, C., Palmer, T. N., Molteni, F., Tibaldi, S. and Cubasch, U. 1990. Extended-range predictions with ECMWF models: time-lagged ensemble forecasting. *Q. J. R. Meteor. Soc. 166*, 867-912.

Dalcher, A., Kalnay, E. and Hoffman, R. N. 1988. Medium-range lagged average forecasts. *Mon. Wea. Rev. 116*, 402–416.

Errico, R. and Vukićević, T. 1992. Sensitivity analysis using ad adjoint of the PSU/NCAR mesoscale model. *Mon. Wea. Rev. 120*, 1644–1660.

Farrell, B. 1989. Optimal excitation of baroclinic waves. *J. Atmos. Sci. 46*, 1193–1206.

Hall, M. C. G., Cacuci, D. G. and Schlesinger, M. E. 1982. Sensitivity analysis of a radiative convective model by the adjoint method. *J. Atmos. Sci.* **39**, 2038–2050.

Hoffman, R. N. and Kalnay, E. 1983. Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus 35A*, 100–118.

Houtekamer, P. L. 1992. The quality of skill forecasts for a low-order spectral model. *Mon. Wea. Rev. 120*, 2993–3002.

Houtekamer, P. L. 1993. Global and local skill forecasts. *Mon. Wea. Rev.*, in press.

Kruizinga, S. and Kok, C. J. 1988. Evaluation of the ECMWF experimental skill prediction scheme and a statistical analysis of forecast errors. Proceedings of the workshop on *Predictability in the medium and extended range*, 1987. Reading, England, ECMWF, 403–415.

Lacarra, J.-F. and Talagrand, O. 1988. Short-range evolution of small perturbations in a barotropic model. *Tellus 40A*, 81–95.

Lorenz, E. N. 1965. A study of the predictability of a 28-variable atmospheric model. *Tellus 27*, 321–333.

Lorenz, E. N. 1982. Atmospheric predictability experiments with a large numerical model. *Tellus 34*, 505–513.

Lorenz, E. N. 1990. Effects of analysis and model error on routine weather forecasting. Proceedings of the seminar on *Ten years of medium-range weather forecasting*, 1989. Reading, England, ECMWF, 115–128.

Marchuk, G. I. 1974. *Numerical solution of the problems of the dynamics of the atmosphere and the ocean.* Academic Press, New York (Russian version 1967), 387 pp.

Molteni, F. and Palmer, T. N. 1992. Predictability and finite-time instability of the Northern Winter Circulation. Proceedings of the Workshop on *New developments in predictability*, 1991. Reading, England, ECMWF, 101–142.

Mureau, R., Molteni, F. and Palmer, T. N. 1992. Ensemble predictions using dynamically-conditioned perturbations. Proceedings of the workshop on *New developments in predictability*, 1991. Reading, England, ECMWF, 143–174.

Palmer, T., Molteni, F., Mureau, R., Buizza, R., Chapelet, P. and Tribbia, J. 1992. Ensemble prediction, *ECMWF Technical Memorandum no. 188*, 45 pp.

Talagrand, O. and Courtier, P. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. (I). Theory. *Q. J. R. Meteorol. Soc. 113*, 1311–1328.

Tribbia, J. J. and Baumhefner, D. P. 1988. The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiences. *Mon. Wea. Rev. 116*, 2276–2288.

Vukićević, T. 1991. Nonlinear and linear evolution of initial forecast errors. *Mon. Wea. Rev. 119*, 1460–1482.

Zhang, Z. 1988. *The linear study of zonally asymmetric barotropic flows*, PhD thesis. University of Reading, England, 163 pp.