EVALUATION OF THE ECMWF EXPERIMENTAL SKILL PREDICTION SCHEME AND A STATISTICAL ANALYSIS OF FORECAST ERRORS

S. KRUIZINGA AND C.J. KOK
Royal Netherlands Meteorological Institute
De Bilt, The Netherlands

1. INTRODUCTION

The forecasts of the atmospheric state produced by a numerical model will slowly diverge from the observed atmospheric states. This is caused by the small differences between the model initial state and the atmospheric initial state. On the average a steady increase of forecast errors with increasing lead time is observed. However, in the last two decades, it has become clear that the growth rate of the errors depends on the state of the atmosphere. Due to this effect the skill of a numerical model varies from day to day. Skill Prediction aims at forecasting these variations in model performance. Recent research in this field has concentrated on two approaches:

i) Monte Carlo; several runs of the numerical model with slightly disturbed initial states are used to derive the statistical parameters of the error distribution

ii) Regression techniques. Predictors derived from the initial state and successive model runs are correlated with the observed error growth.

Last year the European Centre for Medium Range Weather Forecasting (ECMWF) announced a semi-operational experiment in skill forecasting based on a statistical regression approach. In this paper we will give some verification results of this experiment. These verification results will be based on field verification data obtained from ECMWF. Additionally some results based on local (KNMI) field verification data will be given. The results obtained initiated a statistical study on model errors and skill prediction. A statistical model of the growth rate variations will be presented and from this model the evaluation results related to skill prediction are explained.

2. VERIFICATION RESULTS OF THE ECMWF EXPERIMENT

The ECMWF experiment aimed at forecasting the field verification results expressed in either Root Mean Square (RMS) error or Anomaly Correlation Coefficient (ACC). The forecasts were expressed in probabilities. To that end the ranges of RMS and ACC were subdivided into 5 equally probable classes. The skill forecast assigned probabilities to each of these classes. The forecasts were given on a daily basis for different lead times (day 3, 5, 7 and 9) and for different areas. Later on the observed RMS and ACC values were distributed as well. For a more comprehensive description of the experiment reference is made to Palmer and Molteni (1987).

In table 1 some verification results based on data obtained from the ECMWF are given. In this table a skill score based on the Ranked Probability Score (RPS) (Epstein, 1969) and the hit frequency obtained when the highest probability class was selected as forecast. The RPS was transformed to a skill over climatology by

$$RPSS = \frac{RPS_{CLIMATOLOGY} - RPS_{FORECAST}}{RPS_{CLIMATOLOGY}}$$

RMS as well as ACC results are presented for two areas. The selected areas (3 and 7) were most interesting for our local purposes. The results obtained for the RMS indicate almost no skill whereas the ACC results indicate a marginal skill. Additionally the RMS forecasts were correlated

Table 1 Verification results on the skill prediction for RMS and ACC for area 3 and area 7 for day 3, 5, 7 and 9. Verification period winter 87/88.

Ranked Probability Skill Score (%)

		Day 3	Day 5	Day 7	Day9
RMS	Area 3	0	2	0	0
	Area 7	0	-1	-2	1
ACC	Area 3	12	7	-1	-4
	Area 7	14	16	-1	0

Hit frequency (chance = 20)

RMS	Area 3	23	20	22	22
	Area 7	28	21	18	22
ACC	Area 3	32	32	20	21
	Area 7	29	33	21	31

Area 3 = (71.25 N, 33.75 N), (-11.25 E, 41.25 E)
Area 7 = (56.25 N, 45.00 N), (-11.25 E, 15.00 E)

with the field verification results obtained on a locally defined grid (see fig. 1). First the probability forecast was transformed to a skill forecast by

$$S = 2 * (P_1 - P_5) + (P_2 - P_4)$$

with P_1 to P_5 the probabilities assigned to the lowest upto the highest RMS class for day 3 on area 7. This formula was proposed on an expert meeting at ECMWF in February 1988. From the locally archived +72 forecasts and analyses the following daily statistics were computed on the given grid: the average error, the RMS error and the standard deviation (STD) of the errors. This was done for 500 hPa, 700 hPa and 1000 hPa height forecasts. In table 2 the correlation coefficients with S are given for verification period December 1, 1987 upto February 28, 1988. This table also indicates marginal skill for RMS as well as STD. Remarkable is the high correlation between mean error and predicted skill. In table 3 similar results for day 5 are given.

3. STATISTICAL MODEL OF SKILL VARIATIONS

The results obtained in the last paragraph initiated a study on the statistical properties of errors of a numerical forecast. In order to be able to interprete the results it is necessary to develop a statistical model. First we will describe the statistical model, next we will present the experimental results. Finally we will use the statistical model to construct distributions which can be compared to the observed ones.
First we assume that when the numerical model is run several times starting from nearly the same initial state the resulting forecast errors at a given gridpoint will be normally distributed with zero mean and a fixed standard

Table 2 Correlation coefficients between predicted skill S and the daily average error ME, the daily RMS error and the daily standard deviation of the forecast errors STD at different pressure levels (RMS/EC is the ECMWF-RMS at area 7)

	ME	STD	RMS	RMS/EC
500 hPa	.43	-.14	-.15	-.19
700 hPa	.35	-.17	-.13	
850 hPa	.28	-.19	-.14	
1000 hPa	.20	-.17	-.11	

Note: Correlation between skill and RMS is negative because low RMS indicates high skill

Table 3 As table 2 but now for day 5

	ME	STD	RMS	RMS/EC
500 hPa	.20	-.16	-.06	-.07
850 hPa	.10	-.17	-.01	
1000 hPa	.03	-.14	-.15	

Table 4 Mean (ME) and Standard Deviations (STD) of observed and generated RMS-distributions

	ME	STD
Day 3	1.84	0.78
Day 5	1.85	0.76
$\lambda = 0$	1.89	0.69
$\lambda = 0,1$	1.88	0.71
$\lambda = 0,2$	1.88	0.79
$\lambda = 0,3$	1.88	0.91

deviation (Bias is neglected). Furthermore we assume that variations in forecast skill, due to different initial states, will lead to variations in the standard deviation of the errors. So the overall distribution at a given gridpoint will be a mixture of normal distributions with different standard deviations. Now we will study forecast quality on the basis of RMS on a given area. Therefore we also assume that we can find within this area a number of gridpoints sufficiently apart for the errors to be uncorrelated. This means that the sum of squares of the errors will have a χ^2 distribution for a given model skill. Assuming next that skill variations have nearly the same effect on the whole area we can conclude that the overall distribution of the RMS will deviate from the χ^2 distribution. The variations in standard deviations will be enhanced in the RMS values.

The data set used to study the forecast errors consisted of seven winters (80/81 - 86/87) with daily forecasts and analyses from the ECMWF model archived locally on the grid given in figure 1. From the data set the day 3 and day 5 500 hPa forecasts and 500 hPa analyses were extracted. For these forecasts the daily errors at each of the gridpoints were computed. However due to model changes this dataset cannot be considered to be homogeneous. This is shown in figure 2 where the seasonal mean errors and standard deviations for the central gridpoint are plotted versus time. Furthermore the standard deviation and the bias vary considerably over the area. Therefore we decided to standardize the errors at the gridpoints. The standardized error d at day i in a given season at gridpoint k is defined by:

$$d(i,k) = (e(i,k) - me(k))/std(k)$$

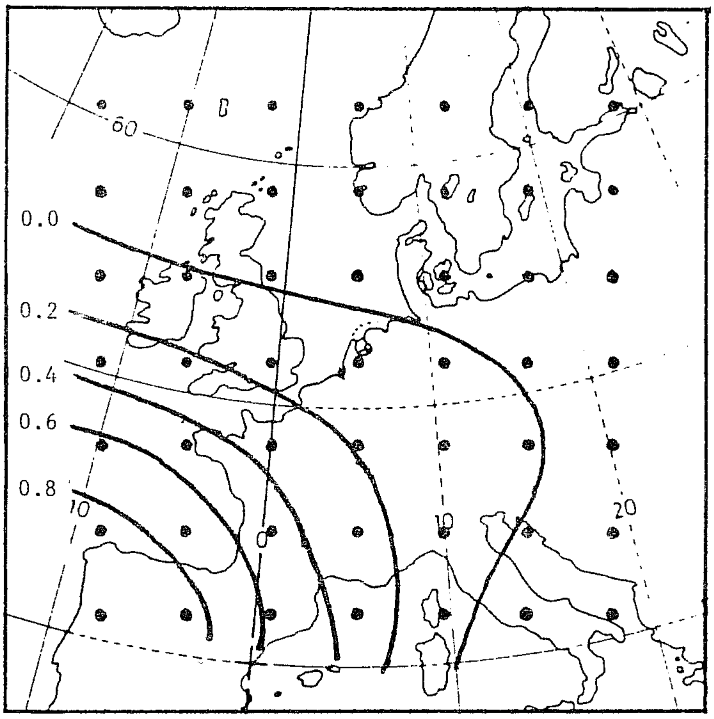

Figure 1. Grid used for archiving ECMWF forecasts and analyses. Solid lines depict the correlation pattern from left lower gridpoint.

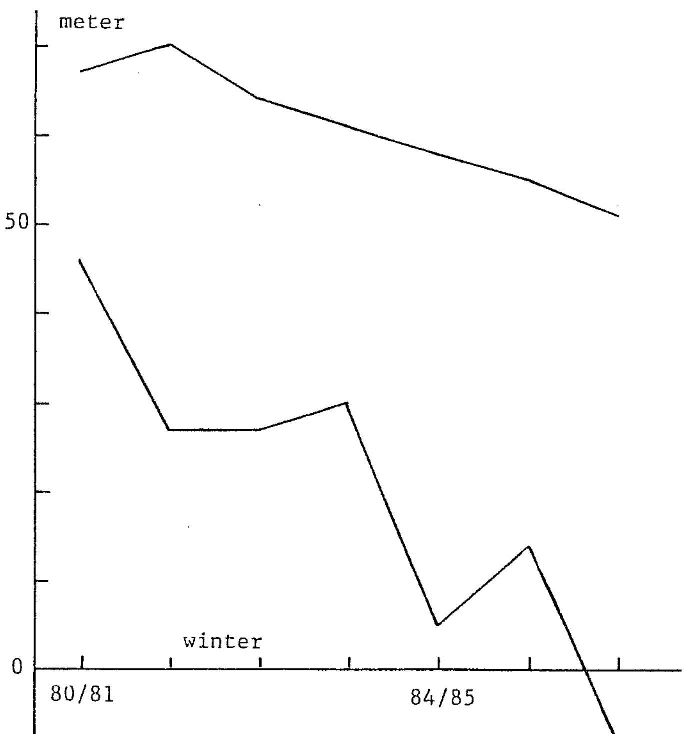

Figure 2. Seasonal mean forecast error and standard deviation at day 3, 500 hPa, central gridpoint.

where $e(i,k)$ = forecast error at day i at gridpoint k

$me(k)$ = seasonal mean error at gridpoint k

$std(k)$ = seasonal standard deviation at gridpoint k.

After this standardization the seven seasons were combined. First we studied the distribution of the errors at the central gridpoint. In figure 3 the cumulative distribution for day 5 is plotted on normal probability paper. The solid line in this figure represents the normal distribution with zero mean and unit standard deviation. As can be seen the plotted points are close to this line. Next we studied the RMS errors. We selected the four gridpoints at the corners of the grid and assumed that the errors at these points were nearly uncorrelated. This assumption is supported by the correlation pattern, from the lower left corner point, plotted in figure 1 which was computed from this data set. In figures 4 and 5 the cumulative distributions for day 3 and 5 are plotted. The solid lines in these figures represent the distribution derived from the χ^2 distribution with four degrees of freedom assuming no variation in the standard deviation of the errors. Clearly the observed curves deviate from the theoretical curve and moreover the deviations agree with what we expect from skill variations: Too many low values and too many high values.
Next we want to compare these distributions with the distributions related to the described model. These distributions were constructed with Monte Carlo techniques. The daily variation in the forecast skill was simulated by drawing a random number e from a normal distribution with $\mu = 0$ and $\sigma = \lambda$. With the parameter λ the overall skill variation is described. Thereafter the daily standard deviation of the errors was set equal to 1+e. This value is related to the daily skill. In the next step four random numbers were generated from a normal distribution with $\mu = 0$ and $\sigma = 1+e$.

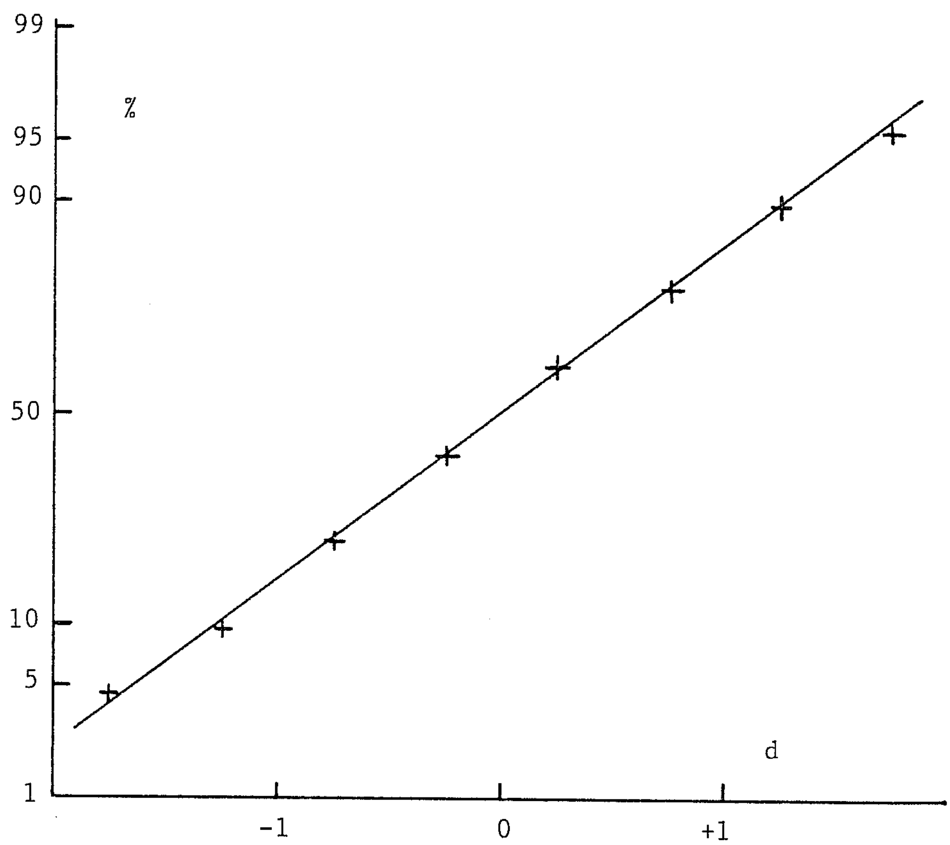

Figure 3. Cumulative distribution of standardized errors at the central gridpoint, day 5 forecast.

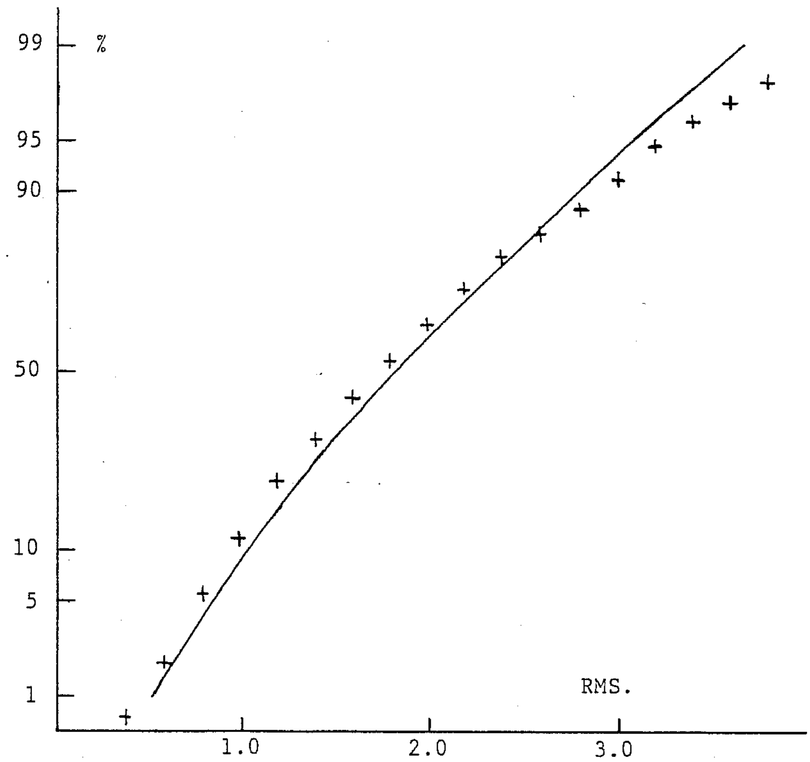

Figure 4. Cumulative distribution of RMS values at day 3. (4 gridpoints, standardized errors)

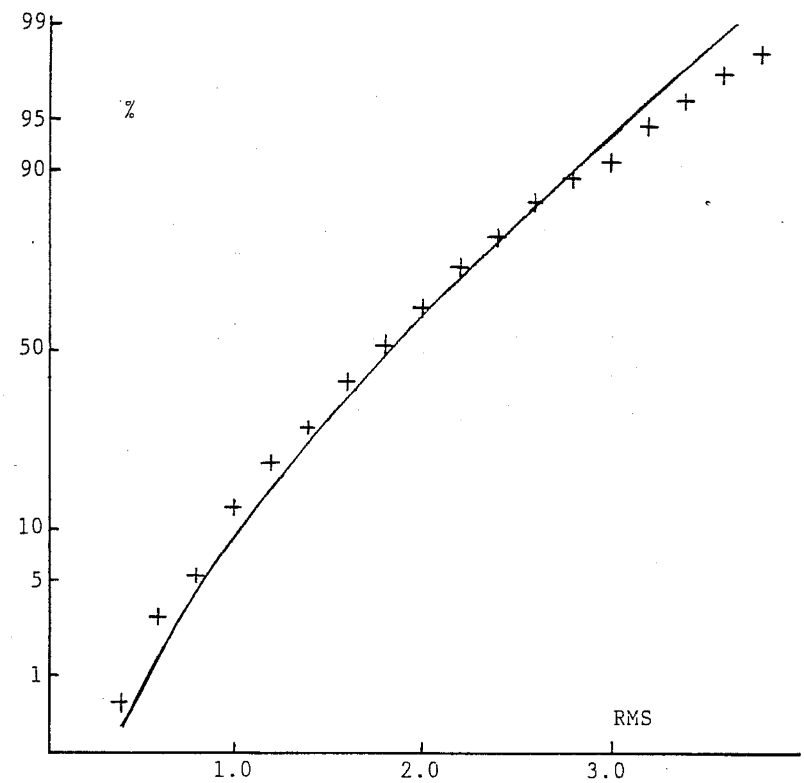

Figure 5. Cumulative distribution of RMS values at day 5. (4 gridpoints, standardized errors)

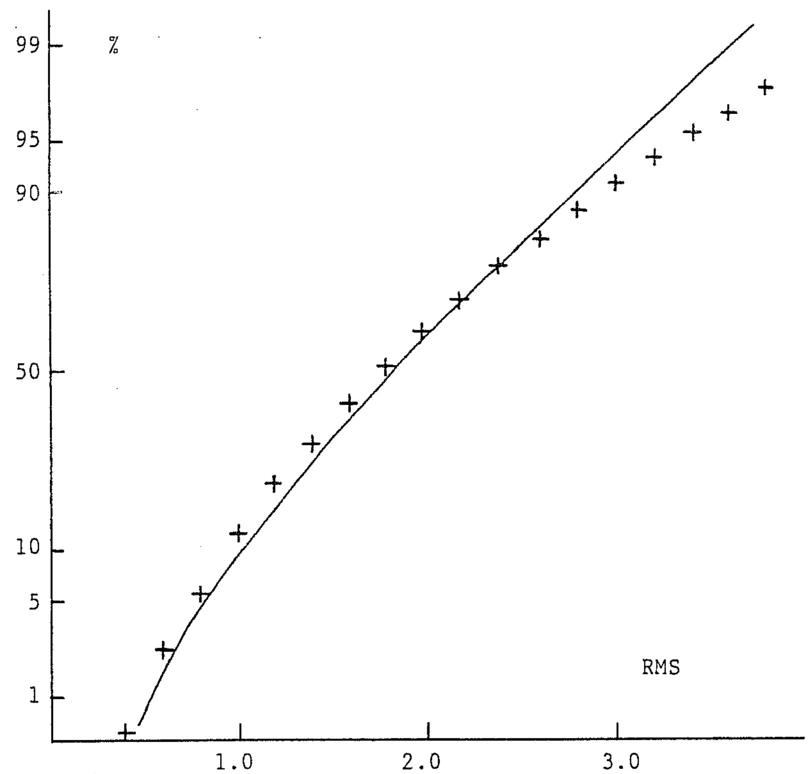

Figure 6. Cumulative distribution of generated RMS values, λ =0.2, 4 degrees of freedom.

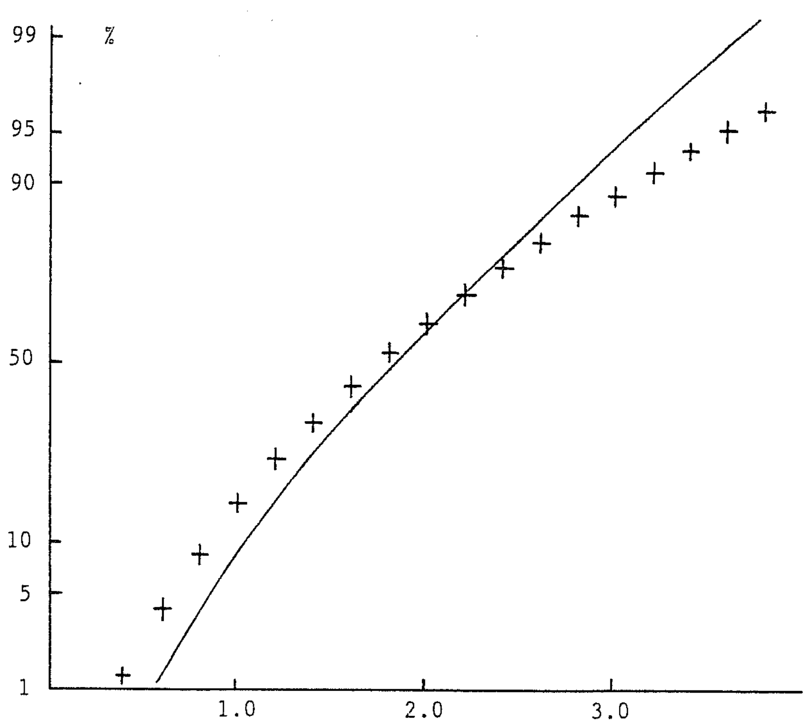

Figure 7. Cumulative distribution of generated RMS values, λ =0.3, 4 degrees of freedom.

From these four numbers, representing the actual errors, the RMS was computed. This process was repeated 5000 times and for the errors as well as the RMS values frequency distributions were constructed. In figures 6 and 7 the cumulative distributions of the RMS are shown for $\lambda = 0.2$ and for $\lambda = 0.3$. The resemblance to the observed distributions is very clear and $\lambda = 0.2$ was selected as a good estimate for the parameter of the observed RMS distributions (see also table 4). The distribution of the errors was very close to the normal distribution just as with the observations. This means that based on this model the variation of the standard deviation of the forecast error will be about 20%. This implies that about 95% of the standard deviations will be in the range from 0.6 up to 1.4 on a relative scale.

In the above model the term 1+e is directly related to the model skill and the computed RMS represent the observed skill. So it was possible to correlate model skill with observed skill. For $\lambda = 0.2$ the correlation coefficient was 0.47. This means that even with perfect knowledge about the model skill only 22% of the variance in the RMS can be predicted, leaving 78% unexplained. If more gridpoints (more degrees of freedom) are used in the computation of the RMS the correlation will improve and for instance at 8 degrees of freedom the explained variance will be 36%. However, regarding the correlation pattern in figure 1, eight degrees of freedom seems rather high.

4. CONCLUSIONS

In this paper a statistical analysis of prediction errors was presented. The simple statistical model that was used allowed some general conclusions with regard to skill prediction. The analysis showed that the correlation

between forecast skill and the observed RMS will be about 0.5. This means that large datasets will be needed to develop stable forecast equations. Moreover the results obtained with these equations will be masked by the unavoidable residual variance in the process. So for verification large datasets are needed as well. These results are important for the Monte Carlo approach too. At this moment this study indicates that at least one hundred parallel runs are needed for a Monte Carlo estimate of the standard deviation. For with a variation of only 20% in the standard deviations an individual standard deviation has to be estimated with a relative error of 5%. However, the relative error of a standard deviation computed from a sample of size N is $1/\sqrt{2N}$. So we need about 200 independent runs.
In our opinion further development of the statistical model will be very useful. Especially a study of RMS errors based on more gridpoints is needed. Furthermore the analysis should be applied to more areas and other seasons.

REFERENCES

Epstein, E.S., 1969: A scoring system for probability forecasts of ranked categories. J. Appl. Meteor., 8, 985-987.

Palmer, T.N. and F. Molteni, 1987: An experimental scheme to predict forecast skill. ECMWF Tech. Memo. 141, 13 pp.