OPEARATIONAL USE OF MEDIUM-RANGE WEATHER PRODUCTS AND PREDICTIVE SKILL TRIALS AT SMHI

by R.Joelsson and O.Åkesson
Svedish Meteorological and Hydrological Institute
Norrköping, Sweden

1. INTRODUCTION

SMHI makes extensive use of ECMWF products, especially in the medium range. Also in the short range the ECMWF forecasts are used in parallell with products from a primitive limited area model of our own, which is run every 12 hour on local computers. Numerical forecasts from NMC, DWD and UKMO are also considered as well in the medium range.

These models are all verified from the user's point of view, i.e. we compare ECMWF (+84) with NMC/DWD/UKMO (+72) and so on. We also compare the forecasts aiming to draw conclusions about predictability of the ECMWF forecast both in a subjective way as well as with an objective method (section 4).

Forecast fields are presented on maps plotted locally. In addition parameters from direct model output (DMO) and various statistical interpretations are used.

In general ECMWF forecasts give useful information out to 6.5 days at 500 hPa over the European area. This predictability range has been more or less constant during the last 3 years. When this report is being finalized (Jan 1992), we have more than 3 months of experience of the new T213 forecasting system. and some comparative verification results will be given from the T106 and T213 models. Note that in more extensive information can be found in Joelsson (1992) on verification results of the T213 model and of its medium range guidance.

2. OBJECTIVE VERIFICATION

2.1 Direct ECMWF model output at 500 hPa.

The circulation at 500 hPa (and at the surface) has been verified. The verification area covers central and northern Europe. The verification parameters are RMS error of forecast height, the anomaly correlation and also S1-skill score. Note that forecast lead times are the same for all models. Some comments can be found in section 2.2.

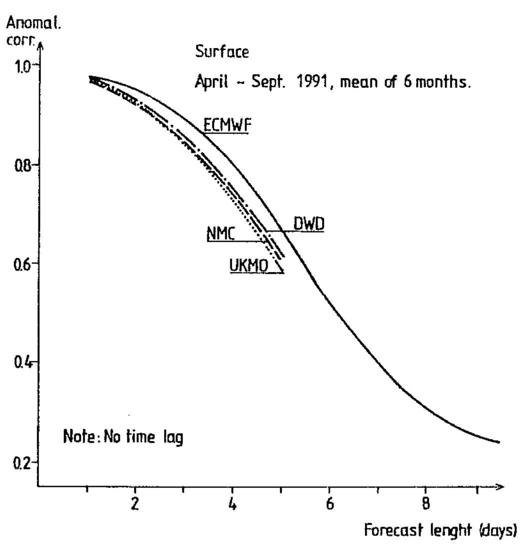

FIG. 1. The mean anomaly correlation of the surface forecasts from ECMWF, NMC, DWD and UKMO during the period Apr-Sep 1991.

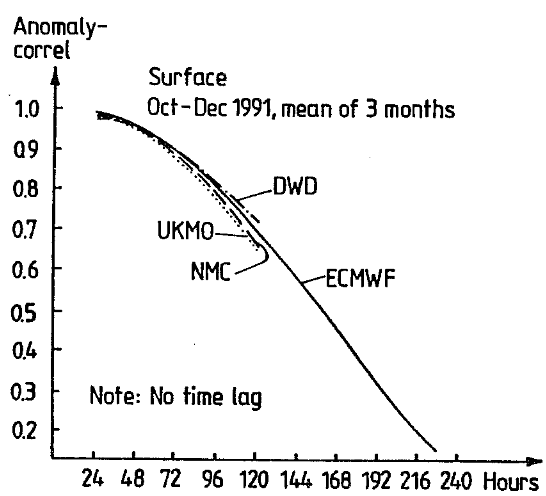

FIG. 2. Same as Fig 1 but for the period Oct-Dec 1991.

2.2 Near surface parameters

These are: temperature at 2m, total precipitation, wind at 10 m and cloud cover. These are available as direct model output from the ECMWF operational forecast model and serve as guidance to the forecasters. To get an idea of the skill of DMO near surface parameters, a comparison with chance and persistence is made. Additionally we have also verified the output from statistical interpretation schemes applied to the ECMWF products (PPM, MOS and Kalman filtering).

Our impression is that (with no illustrations) the DMO 2m (maximum) temperatures in northern Sweden are still too low, especially during spring and over snowcovered ground. The negative bias has however has decreased with the new model. Despite a slight negative bias also in southern Sweden the DMO 2m temperatures give better guidance than in the northern part.

The DMO precipitation normally gives a good guidance to the forecasters, both of the distibution and the amounts, albeit with a small positive bias. Results from the statistical interpretation schemes are of similar standard but are normally biassed with a negative sign.

The ECMWF DMO cloud amount generally shows a negative bias, mainly due to a lack of low clouds. The skill over chance and persistence is often positive out to day 3, but not significant. For day 4 and onwards, there is normally no positive skill.

The DMO has a tendency (in the new model) to forecast too low wind speed over mountainous areas above the tree line, which is found at altitudes exceeding 500-900 m in the Scandinavian mountains

2.3 Comparison with output from several models

The graphs in the Figs 1 and 2 give an idea of predictability as a function of forecast lead time for each model.

The ECMWF T106 model was superior to the other models and significantly better than UKMO and NMC (confirmed objectively by a t-test to the monthly scores). This was not the case with DWD, however, for the period Apr-Sep 1991).

The ECMWF T213 model has so far not shown the superiority over the other models. In fact the DWD T106 model has in general performed best and the UKMO forecast scored the lowest.

3 SUBJECTIVE VERIFICATION

In Figs 3 and 4 comparative subjective scores for forecasts from ECMWF, NMC, DWD and UKMO are displayed. The scoring method is: 5: excellent, 4: good, 3: useful, 2: poor and 1: misleading.

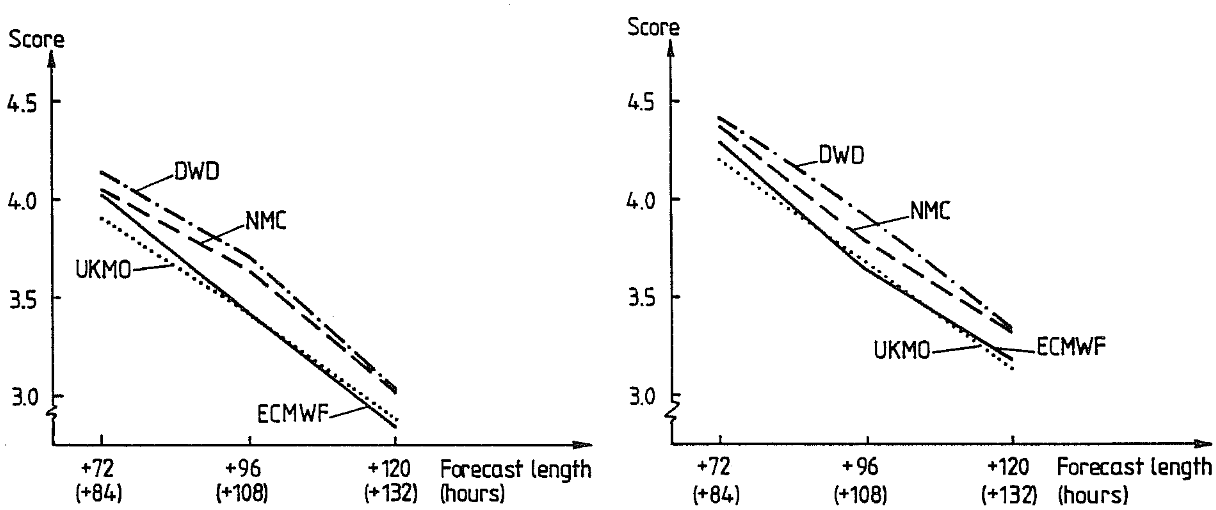

FIG. 3. Subjective assessment of surface (left) and 500 hPa (right) forecasts from ECMWF (+84/+108/+132h), NMC (+72/+96/+120h), UKMO (+72/+96/+120h) and DWD (+72/+96/+120h). Period: Oct-Dec 1991

The subjective assessment in Fig 3 mainly supports the objective scores of Fig 2.

Figure 4 shows that during the last 3 years there has been no definite skill trend of ECMWF and ECMWF has on average been assigned the highest scores, but during 1991 DWD has improved compared with the years before. The UKMO has in general been assessed lower during 1991, compared with ealier years.

3.1 Good and poor forecasts

In Figs 5 and 6 the frequency of good and poor forecasts from ECMWF, UKMO, NMC and DWD for 3, 4 and day 5 are shown. Note that ECMWF forecasts are based on 12 hours older data than those from the other three centres.

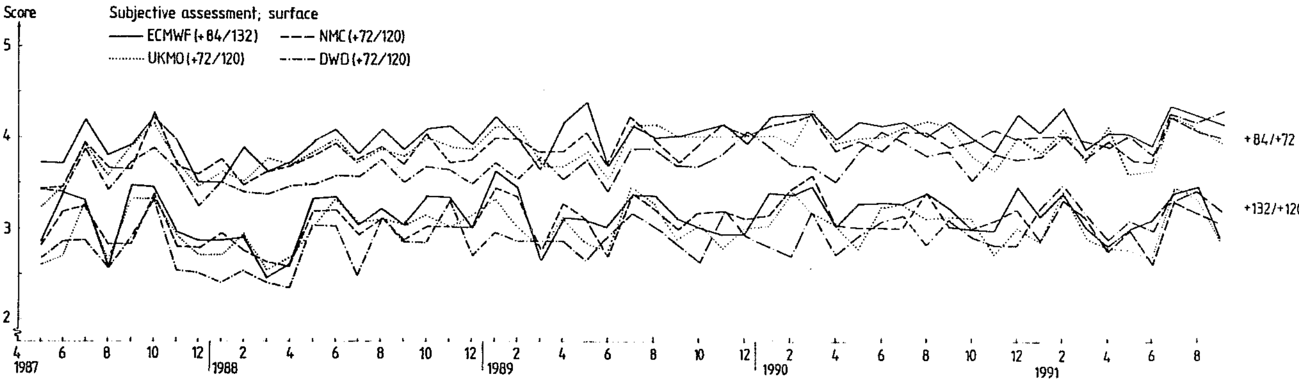

FIG. 4. Subjective assessment of surface forecasts during the last 5 years.

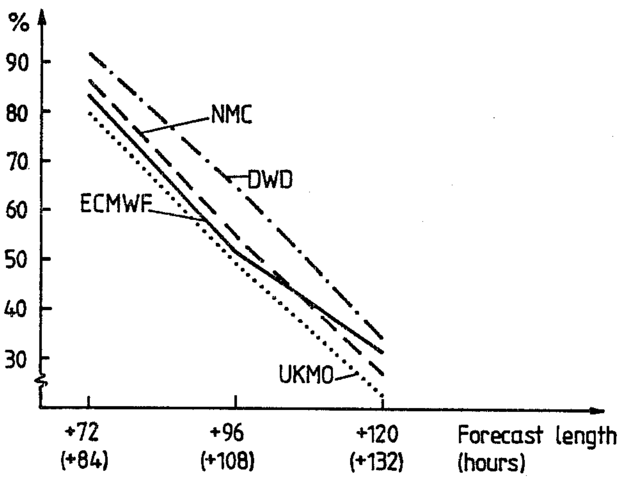

FIG. 5. Percent of good forecasts (score 4).
500 hPa. Northernmost Europe.
Period: Oct-Dec 1991

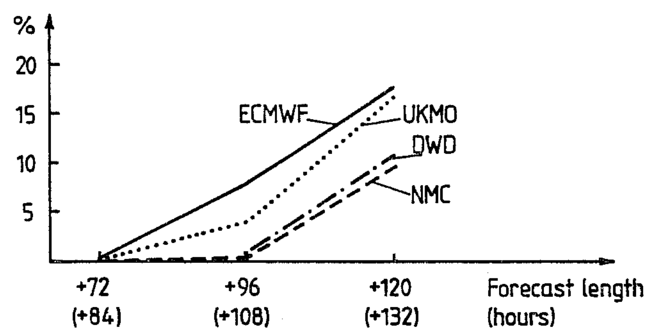

FIG. 6. Percent of poor forecasts (score 2).
500 hPa. Northernmost Europe.
Period: Oct- Dec 1991

Noticable in Fig 5 is that DWD in general provides the highest number of good forecasts, compared with the other models and UKMO the lowest number. ECMWF has produced the highest number of poor forecasts, especially compared with the information from DWD and NMC.

The predictability curves of ECMWF 500 and 1000 hPa height forecasts have been extended until the end of 1991 (Fig 7).

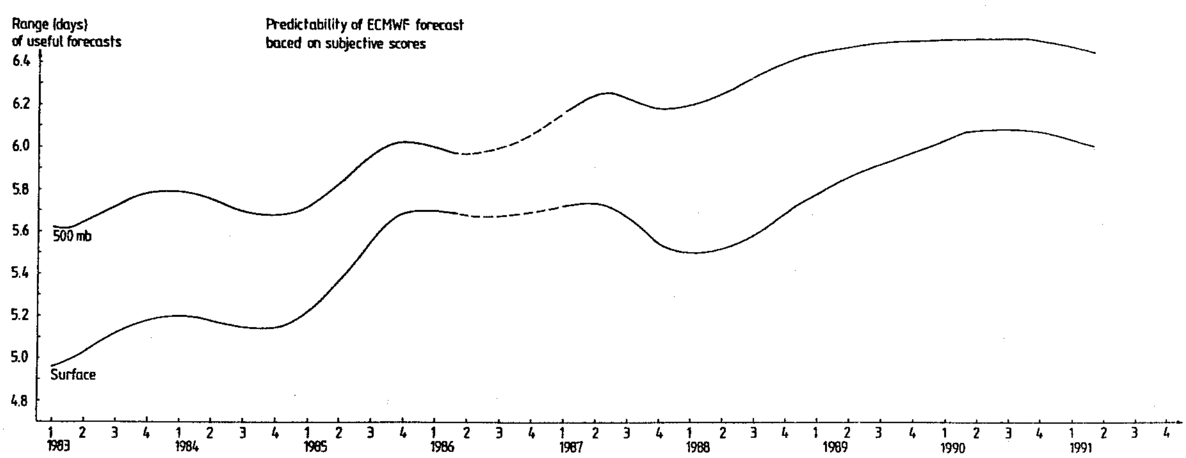

FIG. 7. Time evolution of predictability of ECMWF forecasts expressed as forecast days of useful information. Curves have been smoothed by a 12 month running average.

The predictability is determined by computing 12 month running means of subjective scores, where score 3 has been defined as the limit of usefulness. From Fig 7 the predictability is 6.5 days at 500 hPa at the beginning of 1991. The corresponding surface predictability is 6.1 days, with a slight positive trend from the beginning of 1988. For 500 hPa, however, there is no obvious trend in the subjective scores, whereas using anomaly correlation (0.6) as the limit of predictability, there has been a decrease by almost half a day.

3.2 Evaluation of the behavior of the model

The day to day consistency of consecutive forecasts have been investigated as well as divergence from forecasts of other Centres. We have used the results both from subjective assessment and objective verification (anomaly correlation). The consistency has been investigated betwen today's D+3 and D+4 from the previous forecast. Similarly D+4 and D+5 and D+3 and D+5 forecasts have been compared, valid at the same time. The result from the subjective scores is strikingly similar to that of objective verification.

Our first objective for this investigation has been to investigate if the forecast information from UKMO and from NMC might add some additional information to that given by ECMWF. As for consistency between consecutive forecasts from ECMWF, would there be cases when the older information should be used instead of that from the latest forecast and how frequent are these cases?

The second objective is to determine which model has the highest day to day forecast variability (jumpiness).

Here some preliminary results will be given, but only from D+4 and D+5 consistency (50 cases). The results are from subjective evaluations, for the period 1 Jan to 21 March 1991. See the table below.

	D+4 better than D+5	D+4 equal to D+5	D+4 worse than D+5	D+4 better than D+5	D+4 equal to D+5	D+4 worse than D+5
ECMWF	50 %	34 %	16 %	58 %	29 %	13 %
UKMO	52 %	40 %	8 %	46 %	44 %	10 %
NMC	46 %	40 %	14 %	57 %	33 %	10 %
DWD	-	-	-	61 %	32 %	7 %

The first 3 columns are for the period 1 Jan- 21 Mar 1991, the last 3 for 1 Oct-31 Dec 1991, also including DWD.

ECMWF has the highest "jumpiness", which is also supported by objective verification (not shown). The variability seems to have increased somewhat with the introduction of the new model.

The UKMO is the least variable with strengthened tendency with time during 1991.

Furthermore we have found some other interesting facts regarding the consistency of forecasts:

Supposing that ECMWF D+6 (from 2 days ago) and D+5 (from yesterdays forecasts), valid at the same time, shows little spread, but D+4 (from the latest forecast) diverge significantly, then the probability that D+6/D+5 are superior, is somewhat higher than the D+4 forecast being correct (or better).

In case all 4 models give the same proposal for D+5, the probability is higher than 90 % for this forecast to be more than useful (score 4).

Suppose that the ECMWF forecast for D+5 differs substantially from the other model forecasts, all of which all have little spread in the circulation pattern, then there is a high likelyhood that ECMWF forecast should be abandoned in favour of the other models.

4. A STATISTICAL TECHNIQUE FOR SKILL PREDICTION

4.1 Introduction

The skill of medium range forecasts sometimes vary substantially from one forecasts run to the next. In some cases forecasts are still useful 10 days after initiation (D+10), whereas at other times forecasts are misleading already from D+3 and onwards. An a priory estimate of the quality of the latest forecast at different forecast lengths would be most useful and in particular it is important to be warned of potentially misleading forecasts in advance.

Medium range forecasters have for long compared forecasts from different forecast centres in order to subjectively estimate the confidence in the evolution of the ECMWF prediction, but also comparing consecutive forecasts for consistency; if the various predictions diverge substantially from the base forecast, usually ECMWF for that time step, less confidence is placed in the ECMWF forecast than when normal spread is found. This is done in a subjective way from experience. These spread measures can be quantified and objective schemes developed using these spread quantities. Ensemble forecast schemes have been used by Hoffman and Kalnay (1983), with the lagged average forecast method for medium range skill prediction. Leslie and Holland (1991) have used single and multiple forecast approaches for short range skill forecasts.

In this section the model and the results are only briefly documented. For a full description see Akesson (1991).

4.2 Data

A data base containing 500 hPa hemispheric ECMWF, UKMO and NMC forecast fields in 5 by 5 degree resolution was used for the predictor creation. The data base consisted of data from 3 winter seasons (Dec trough Feb) from 1986/87, 1987/88 and 1988/89, altogether 271 days with medium range forecasts from the 3 forecast centres.

4.3 The Model

This statistical model is a modified version of the skill prediction model tested during the winter 1988/89 (Molteni and Palmer, 1991), which used single forecast predictors only, such as the 12 h error field. However, in this version predictors were created from the spread (divergence) between ECMWF and UKMO/NMC forecasts from 00 UTC initiated 12 hours later than the ECMWF forecast and the spread between consecutive ECMWF forecasts (consistency), expressed as either RMS of the differences or as ACC between the various forecasts. The corresponding predictands were RMS of the error and ACC for the latest ECMWF forecast run at D+3 (84h) and D+5 (132h) respectively and there were separate sets of predictors for RMS and ACC using only RMS predictors for the RMS predictand and only ACC predictors for the ACC predictand.

The spread was calculated after having merged the UKMO and NMC forecasts into an ensemble using D+3 and D+5 separately. In cases when one of UKMO and NMC forecasts were missing the other one was used alone with a separate set of regression coefficients. In rare events when both UKMO as well as NMC forecasts from 00 UTC were missing the UKMO forecast from 12 UTC on the previous day was used, again with a new set of coefficients, but this happened only twice during the test period.

A number of spread predictors were calculated and tested from various time steps of the forecasts. Furthermore additional predictors were tested between the latest and the previous ECMWF forecasts, both valid at the same time. Initially a large number of predictors were tested, mixing all possible combinations of time steps. However, after screening a set of 22 remaining potential predictors were tested. They are listed in Table 1. Significance tests reduced the final number of predictors to only 2 for forecast day 3 and day 5 and 4 predictors each for forecast day 7 and day 9. They are listed in Table 2.

1. RMS consistency between 2 consecutive ECMWF forecasts at appropriate time steps.
2. RMS spread between ECMWF 12 UTC+48 h and UKMO/NMC 12 UTC+48 h
3. RMS spread between ECMWF 12 UTC+60 h and UKMO/NMC 00 UTC+72 h (previous)
4. RMS spread between ECMWF 12 UTC+72 h and UKMO 12 UTC+72 h
5. RMS spread between ECMWF 12 UTC+84 h and UKMO/NMC 00 UTC+72 h (latest)
6. RMS spread between ECMWF 12 UTC+96 h and UKMO 12 UTC+96 h
7. RMS spread between ECMWF 12 UTC+108 h and UKMO 00 UTC +120 h (previous)
8. RMS spread between ECMWF 12 UTC+120 h and UKMO 12 UTC+120 h
9. RMS spread between ECMWF 12 UTC+132 h and UKMO 00 UTC+120 h (latest)

10 - 18. Same as 1-9 above but for ACC

19. Transition index
20. Amplification index
21. Results from previous RMS regression step
22. Results from previous ACC regression step

TABLE 1. Total number of predictors tested.

D+3		D+5		D+7		D+9	
RMS	ACC	RMS	ACC	RMS	ACC	RMS	ACC
1	10	1	10	1	10	1	10
5	14	9	18	19	19	19	19
-	-	-	-	20	20	20	20
-	-	-	-	21	22(fr d+5)	21	22(fr d+7)

TABLE 2. Predictors selected for RMS and ACC at various forecast times. Numbers refer to predictor number in TABLE 1.

4.4 Results from the training sample

Results (correlation coefficients %) from the dependent sample is summarized in Table 4 for Europe (EU), North America (NA) and the Northern Hemisphere (NH).

	D+3			D+5			D+7			D+9		
	EU	NA	NH	EU	NA	NH	EU	NA	NH	EU	NA	NH
RMS	22	33	30	28	37	40	26	44	36	16	40	25
ACC	42	42	43	45	41	46	31	44	42	19	28	29

TABLE 3. Correlation coefficients (%) for Europe (EU), North America (NA) and the Northern Hemisphere (NH) at different forecast days for RMS (upper row) and ACC (lower row).

Except for ACC at forecast days 3 and 5 scores for Europe are generally lower than those for North America and the Northern Hemisphere and below 40 %. At forecast days 7 and 9 the scores for North America are above 40 % and higher than for the other two areas except for ACC at day 9.

4.5 Results from winter 1990/91.

Contingency tables are derived for RMS and ACC separately as for D+3 and D+5. These are shown in tables 5a and b respectively. Results for D+7 and D+9 and not available, since hemispheric fields were not retrieved during that time period.

	(OBS) LOW	NORM	HIGH	
LOW	12	14	11	62%
(FC) NORM	4	2	3	15%
HIGH	3	6	5	23%
	32%	37%	32%	32%

TABLE 4A. RMS D+3.

	(OBS) LOW	NORM	HIGH	
LOW	16	8	1	42%
NORM	6	7	5	31%
HIGH	1	4	11	27%
	39%	32%	29%	58%

TABLE 4B. RMS D+5.

Tables 5a and 5b show that, using RMS as a measure of skill, D+3 has a large bias towards predicting too many high skill cases, whereas this is not the case for D+5, where the number of cases are much more evenly distributed among the forecast categories. Skill expressed in terms of anomaly correlation is shown in Tables 6a and 6b for D+3 and D+5 respectively. For D+3, contrary to RMS, there is instead a strong bias towards forecast low or normal skill and only one event with higher than normal skill. For D+5 there is also a bias to predict low skill.

	(OBS) LOW	NORM	HIGH	
LOW	10	5	8	38%
(FC) NORM	5	19	12	60%
HIGH	1	0	0	2%
	27%	40%	33%	48%

TABLE 5A. ACC D+3.

	(OBS) LOW	NORM	HIGH	
LOW	11	10	10	53%
NORM	3	7	8	30%
HIGH	0	2	8	17%
	24%	32%	44%	44%

TABLE 5B. ACC D+5.

Summarizing the percent correct in table 6, the near failures and the severe failures showes the RMS at D+5 to be most and RMS at D+3 to be the least successful.

	PERCENT CORRECT	NEAR FAILURES	SEVERE FAILURES
D+3 RMS:	32 %	45 %	23 %
D+5 RMS:	58 %	39 %	3 %
D+3 ACC:	48 %	37 %	15 %
D+5 ACC:	44 %	39 %	17 %

TABLE 6.

The following table summarize Heidke skill score with chance as reference and also the "chi-square" significance test value:

	HEIDKE SKILL SCORE:	CHI-SQUARE (5 % LIMIT: 9.49):	CORRELATION COEFF
D+3 RMS	-0.01	1.73	0.13
D+5 RMS	0.36	23.27	0.64
D+3 ACC	0.21	10.79	0.36
D+5 ACC	0.20	9.18	0.56

TABLE 7.

From table 7 it is evident that only RMS for D+5 is well above the significance level and D+3 RMS well below.

Figure 1 shows the day to day variability of forecast and the corresponding observed skill for RMS D+5. The correlation coefficient is 0.64, which is higher than for RMS D+3 and for ACC as indicator of skill. For more detailed and comprehensive results see Akesson (1991).

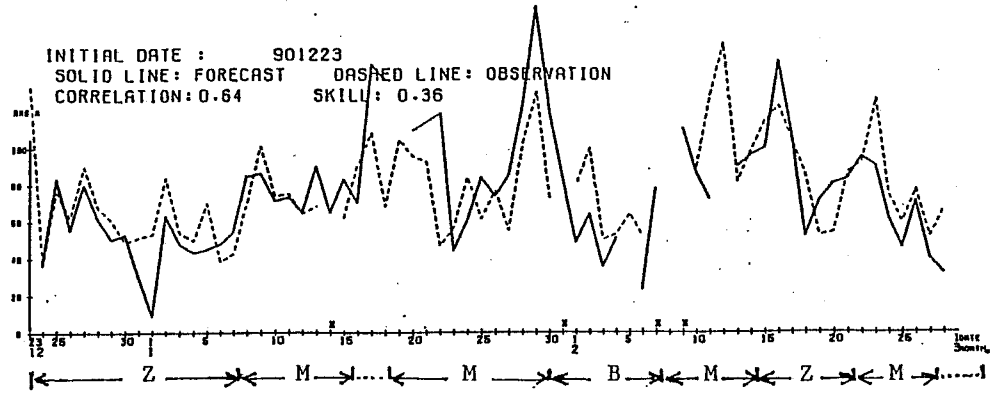

FIG. 1. Daily forecast skill values expressed as RMS difference for ECMWF D+5 (+84h) forecast (solid line) and verification (dashed line) as RMSE. Unit: m. M=meridional flow type, B=blocking, Z=zonal flow type.

Based on 500 hPa analysis charts for Europe the period was subjectively divided into zonal (Z) and meridional (M) cases, including blocking (B), indicated in the figures. There is no obvious relationship between flow and model behaviour, but possibly the day to day variability is generally larger for meridional cases than for the zonal period during the end of December and beginning of January.

Comparison with results from the 1988/89 experiment indicates similar skill, with the exception of the D+5 RMS, which is significantly better than that from the 1988/89 experiment. On the other hand the D+3 ACC scores from 1988/89 were higher than those obtained in this latest experiment.

Leslie and Holland (1991) calculated the Heidke skill score for a subregion to the south-east of Australia from a 3-category contingency table for 36 hour skill forecasts and obtained a score of 0.28, which is of the same order of magnitude as our scores (Table 7).

This supports Palmer and Tibaldi (1990) in their statement that predictive skill in the medium range is not necessarily more difficult than in the short range. However, Akesson (1983), obtained higher skill using ACC with spread between D+4 and D+5 than from spread between D+5 and D+6.

4.6 Comments

Until the Monte_Carlo method comes into operation experiments with simple regression models should continue and ideally be expanded to other seasons and to smaller areas as well. For the winter season 1991/92 at SMHI we will be able to receive hemispheric ECMWF and NMC fields but not UKMO fields. This will enable us to run the scheme for D+7 and D+9 as well. However, recent upgrading of all models calls for new statistics and new regression equations, and soon there will be enough data available to recalculate the equations. Sub-regions of Europe should also be considered, alternatively calculating the spatial distribution of the predicted skill in a new real-time skill prediction scheme.

5 REFERENCES

Akesson, O., 1983: Evaluation and Use of ECMWF Forecasts. ECMWF Proceedings of Seminar/Workshop, 1982 on Interpretation of Numerical Weather Prediction Products. 237-262.

Akesson, O., 1991: A Statistical Technique for Skill Prediction. ECMWF Proceedings of Workshop on Predictability 1991. In preparation.

Hoffman, R. N. and E. Kalnay, 1983: Lagged-averaged forecasting, an alternative to Monte-Carlo forecasting. Tellus, 35A, 100-118.

Joelsson, R., 1992: SMHI Feed-back on the Characteristics of the Forecasts from ECMWF T213 Model for the period Oct. through Dec. 1991. ECMWF report on experiences from member states, 1992. In preparation.

Leslie, L. M. and G. J. Holland, 1991: Predicting Regional Forecast Skill Using Single and Ensemble Forecast Techniques. Mon. Wea. Rev., 119, 425-435.

Molteni, F. and T. N. Palmer, 1991: A Real_Time Scheme for the Prediction of Forecast Skill. Mon. Wea. Rev., 119, 1088-1097.

Molteni, F., S. Tibaldi and T. N. Palmer, 1990: Regimes in the wintertime circulation over northern extratropics. Part 1: Observational evidence. Quart. J. Roy. Meteor. Soc., 116, 31-67.

___, and ______, 1990: Regimes in the wintertime circulation over northern extratropics. Part II: Consequences on dynamical predictability. Quart.J. Roy. Meteor. Soc., 116, 1263-1288.