PROBLEMS OF USING DMO FOR FORECASTING IN SWEDEN AND SOME COMMENTS ABOUT THE SYNOPTIC PERFORMANCE OF NWP MODELS

Rune Joelsson
SMHI
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 parallel 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 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 forecasts both in a subjective way as well with an objective, statistical technique for skill prediction. The last method so far mostly experimentally.

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

In section 2 below the synoptic performance of the operational numerical models used at SMHI is commented as well some remarks about the relation between consistency and skill. Results from subjective assessment and objective verification are used.

In section 3 the DMO near surface parameters will be discussed: which of them are most useful and normally give the best guidance and so on. In the last section some conclusions will be drawn.

2. SYNOPTIC PERFORMANCE

From the end of 1979 when ECMWF forecast products became operational, SMHI has had the ambition to try to determine the reliability of the model outputs by objective verification and subjective assessment and has also compared the guidances with those from other main operational models.

2.1 Objective verification

The verification parameters usually used for the circulation at 500 hPa and surface are correlation between predicted and observed anomaly and also S1-skill score; now and then also RMS-error. The verification area is depicted in Fig 1 below, where also the area for subjective assessment is included.

Figs 2, 3 and 4 show some examples for the period April 1991 and March 1992. Note that forecast ranges are the same for all models.

The figures are followed by some comments, as well the development during the period April and July this year which are depicted in Figs. 5 and 6.

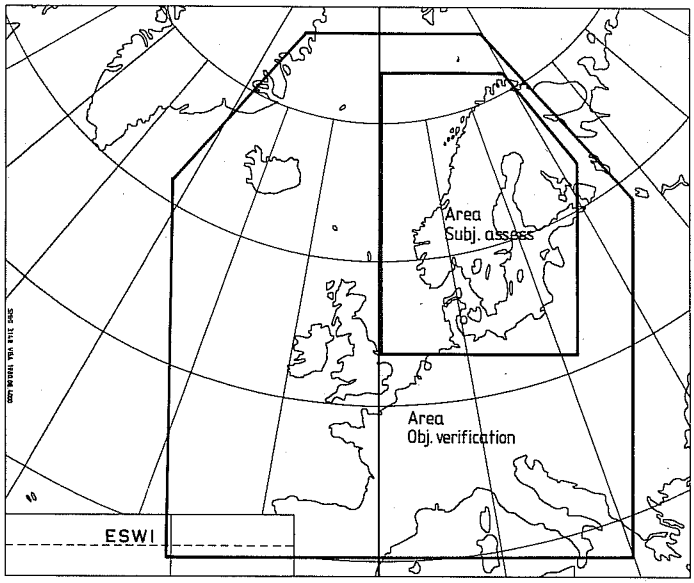


Fig 1 Areas for objective verification and subjective assessment

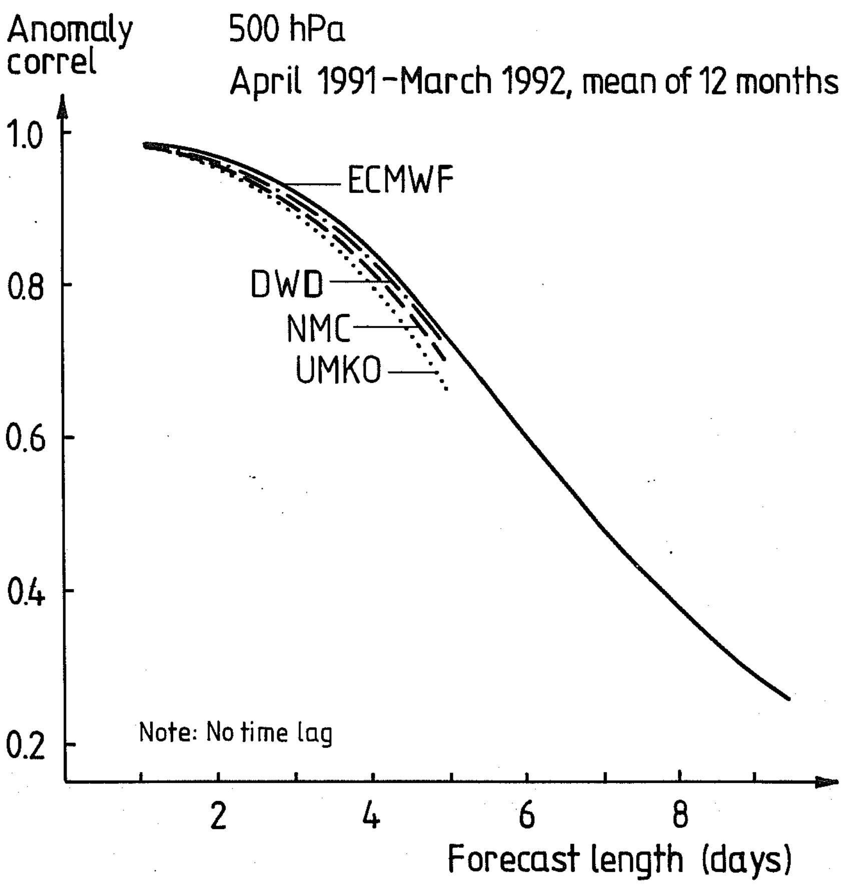


Fig 2 The mean anomaly correlation of the 500 hPa forecast height for ECMWF, NMC, DWD and UKMO during the period April 1991 - March 1992

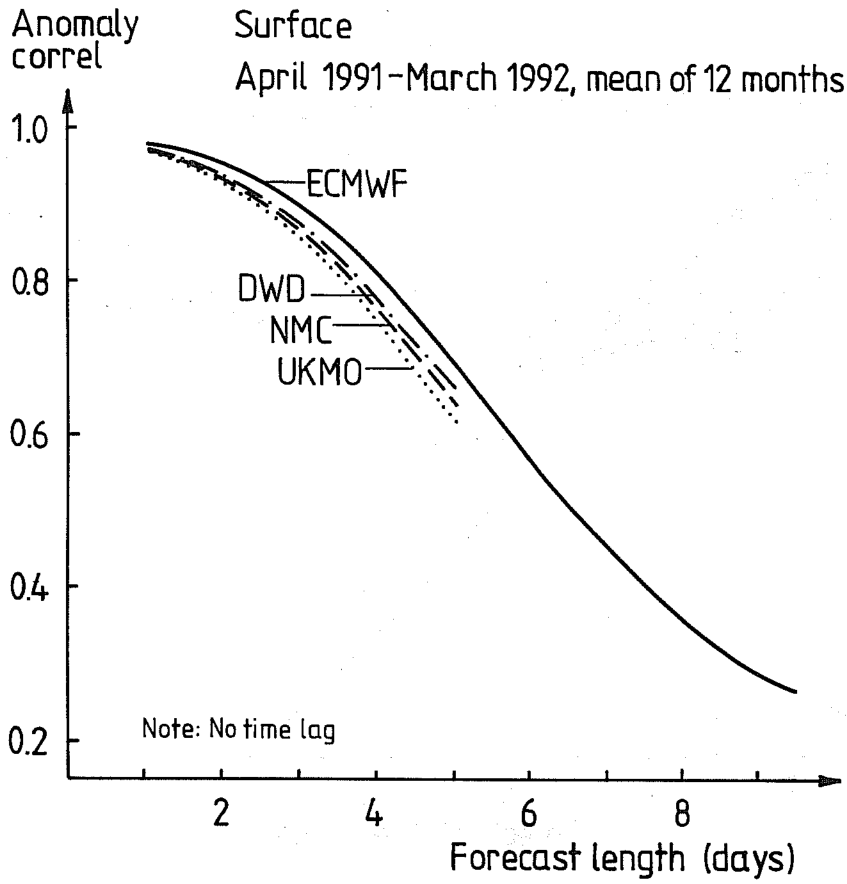


Fig 3 The same as Fig 2, but for surface pressure

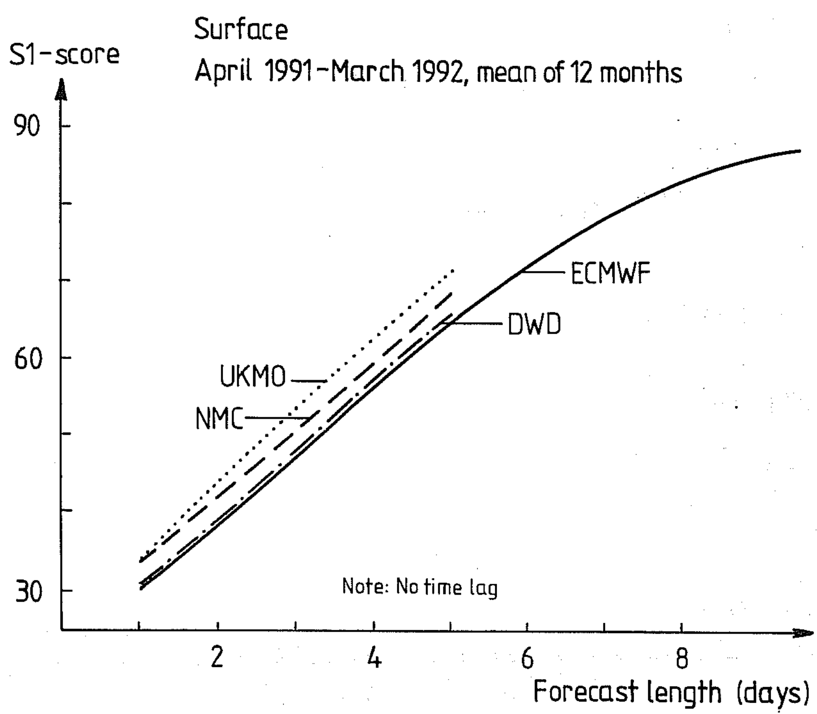


Fig 4 The S1-skill score for pressure. Forecasts from ECMWF, NMC, DWD and UKMO. Averages for the period April 1991 - March 1992

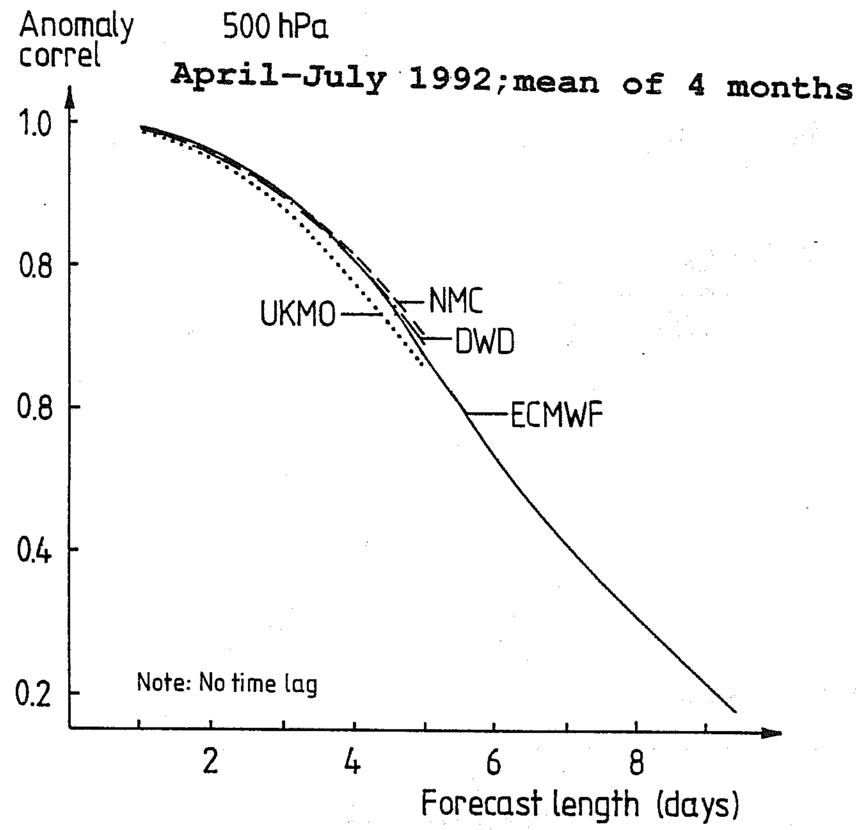


Fig 5 The mean anomaly correlation of the 500 hPa forecast height for ECMWF, NMC, DWD and UKMO during the period April - July 1992

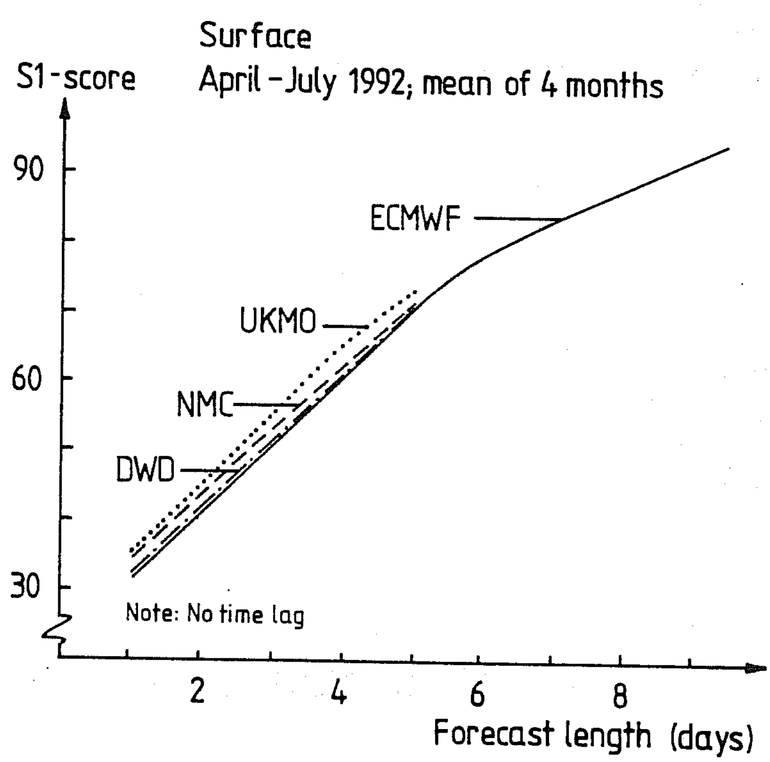


Fig 6 The S1-skill score for pressure. Forecasts from ECMWF, NMC and UKMO. Averages for the period April - July 1992

Comments:

The graphs in the figures 2, 3 and 4 give an idea of predictability as a function of forecast range for each model. We notice that the ECMWF products have a higher skill than the outputs from the other models for all ranges. In the medium range (4 to 5 days) the differences (to the advantage of ECMWF) in predictability are, with:

a) Anomaly correlation

	500 hPa	Surface
ECMWF/UKMO	11 hours	14 hours
ECMWF/NMC	8 hours	11 hours
ECMWF/DWD	3 hours	7 hours

b) S1-skill score

	Surface
ECMWF/UKMO	17 hours
ECMWF/NMC	10 hours
ECMWF/DWD	3 hours

Furthermore we notice:

* The better performance of ECMWF compared with UKMO appears to be systematic and it is confirmed objectively by applying a significance test (t-test; level .05) to the monthly scores. It is valid for both surface and 500 hPa, using both the scores.

* Comparing ECMWF forecasts with those from DWD (T 106) we find the difference not significant.

* The better performance of ECMWF compared with NMC we find significantly by using the t-test; but it is not as apparent as comparing with UKMO.

Objective verification for the period April - July 1992:

For this period we have used the same score to see if there are some differences in the performance of each model; especially compared with the preceding 12 months.

As an example see figures 5 and 6 below.

Comments to figures 5 and 6:

* In the medium range the guidance from ECMWF, DWD and NMC has about the same skill. That means that the guidance from NMC seems to have improved most significant during this period (depicted with these scores).

* For the "short range" (up to day 3) the ECMWF has the best guidance especially for surface.

* UKMO has in general the lowest scores.

2.2 Subjective assessment.

To visualize the skill of forecasts by subjective assessment we use a five degree scale, where score 5 is excellent; 4 good; 3 useful; 2 poor and score 1 very poor and misleading.

In figure 7 below is depicted scores for forecasts from ECMWF, NMC, DWD, and UKMO over Scandinavia and nearest surrounding.

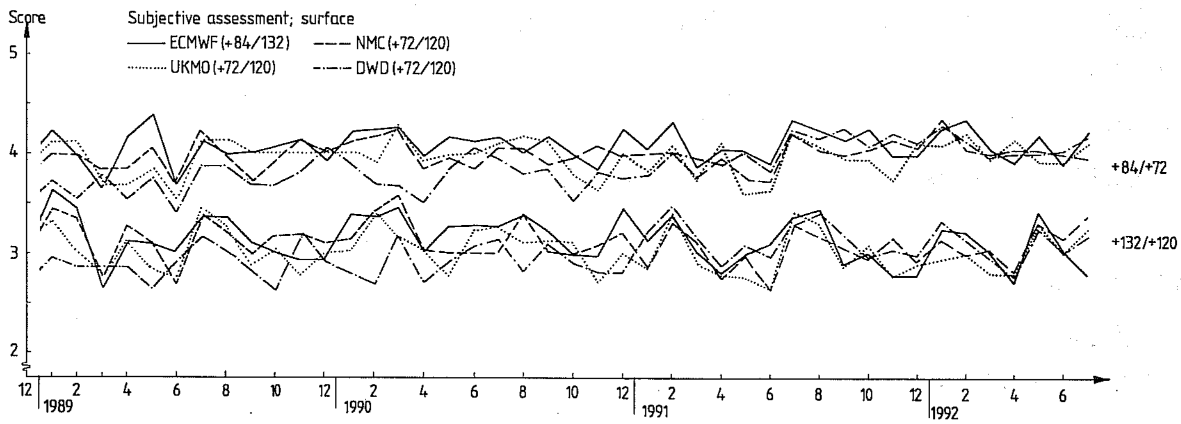


Fig 7 Subjective assessment of surface forecasts from ECMWF (+84/+132), NMC (+72/+120) UKMO (+72/+120) and DWD (+72/+120)

Note that ECMWF forecasts are based on 12 hours older data than those from the other three centres.

Comments: *3 day forecasts:*

* The ECMWF forecasts in general have a somewhat higher reliability than the information from the other models, though the scores during the last 5 months have about the same level for all the models and that this level is somewhat lower compared with the 9-month period before.

* the quality of forecasts from DWD have become significantly higher with the new model.

* The forecasts from NMC are in general the most consistent in terms of quality from month to month.

* The quality of forecasts from UKMO has reached a somewhat higher level, compared with a year ago.

5 day forecasts:

* During the last year the scores for forecasts from DWD, in general have the highest level.

The scores from UKMO are the lowest, with exception of the last 4 months when the guidance from the four models are about the same. For July, however, there are a great spread of the scores, with the lowest value from ECMWF.

* During the last 5 months we also notice great variations of the quality from month to month for all the models.

2.2.1 Verification results for the period October 1991 - July 1992

For this period we have specially assessed the performance of the new upgraded ECMWF model compared with the other operational numerical models. The results we see in the figures 8 to 10 below.

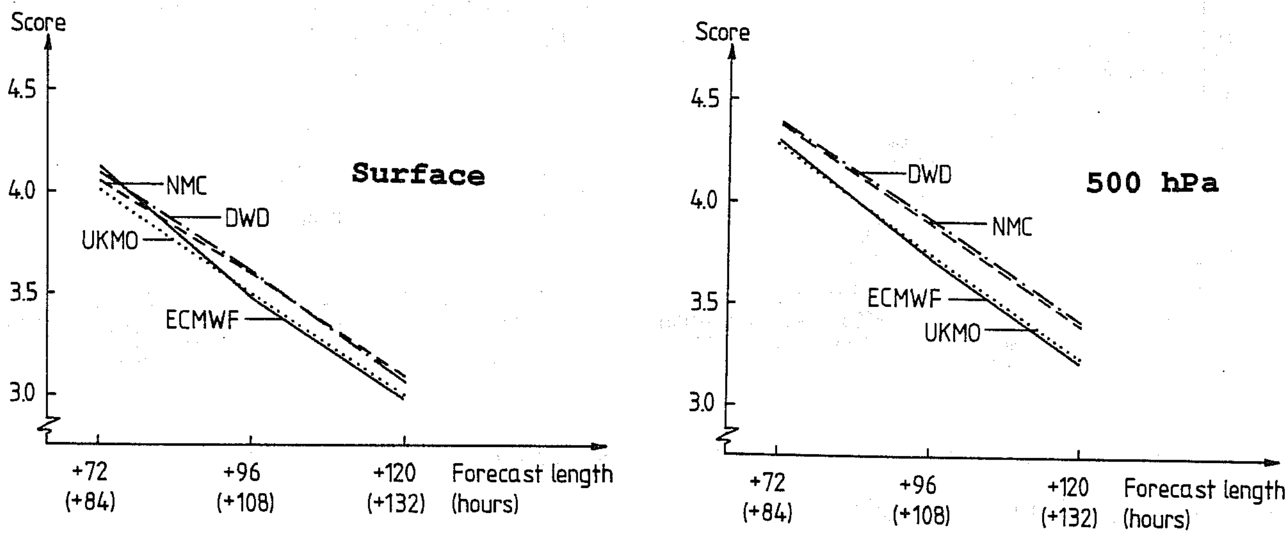


Fig 8 Subjective assessment of forecasts from ECMWF (+84/+108/+132), NMC (+72/+96/+120), UKMO (+72/+96/+120) and DWD (+72/+96/+120) during the period Oct 1991 - July 1992
Surface to the left. 500 hPa to the right.

Note that ECMWF forecasts are based on 12 hours older data than those from the other three centres.

Comments to figure 8:

* For both surface and 500 hPa we notice that the scores for DWD and NMC are close to each other and that the forecasts from these centres in general have been judged to have a higher reliability (from an operational point of view; that is the numerical guidance is presented to the forecasters at the same time) than the information from ECMWF and UKMO.

Note, however, that for surface and day 3 the scores from ECMWF (+84) are the highest together with those from NMC.

* The scores from ECMWF and UKMO are about the same and in general lower than those from DWD and NMC, with exception of day 3 for surface where ECMWF scored high.

Good and poor forecasts

We have also determined the distribution of good and poor forecasts from ECMWF, UKMO, NMC and DWD for day 3, 4 and 5, which are illustrated in Figures 9 and 10 below

Note that ECMWF forecasts are based on 12 hours older data than those from the other three centres.

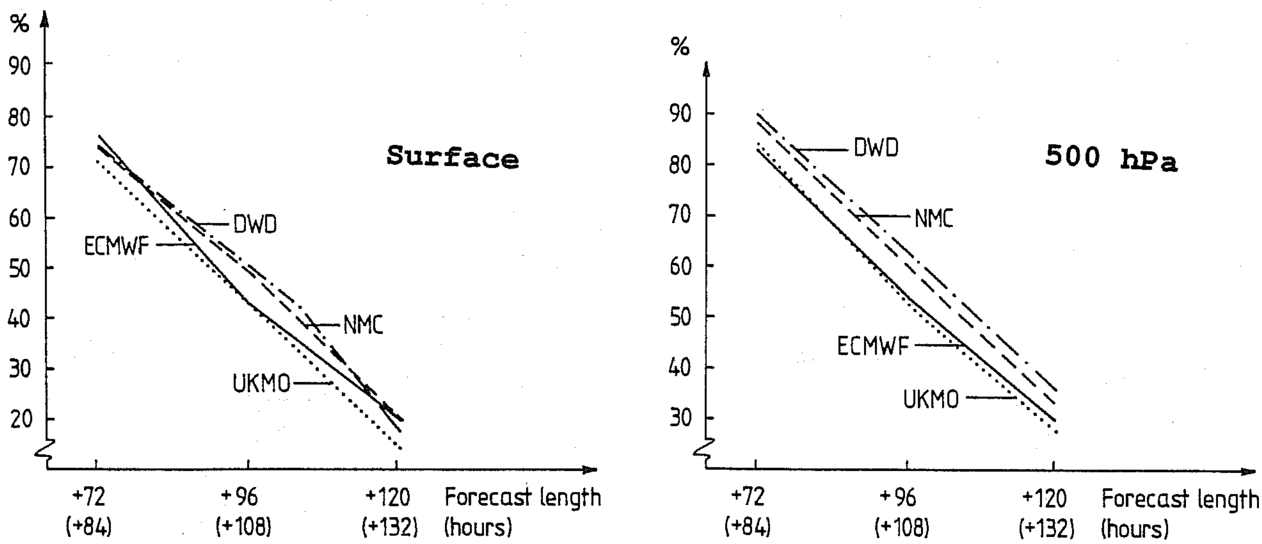


Fig 9 Percentage of good forecasts (score≥4). Average of 10 months (Oct 1991-July 1992) Surface to the left. 500 to the right.

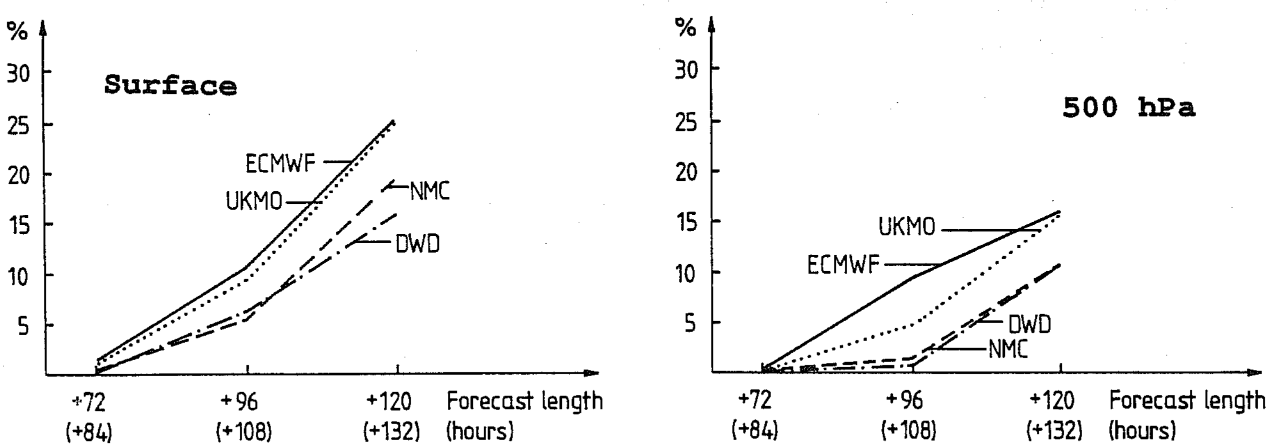


Fig 10 Percentage of poor forecasts (score≤2). Average of 10 months (Oct 1991-July 1992). Surface to the left. 500 to the right.

Comments to figures 9 and 10:

* For both surface and 500 hPa DWD and NMC in general give the largest numbers of good forecasts compared with ECMWF and UKMO, which curves have the same performance.

We, however, notice that for surface day 3 the ECMWF has been judged the same level as NMC (the highest).

* ECMWF in general produces the largest numbers of poor forecasts; especially for 500 hPa and day 4.

* For both surface and 500 hPa DWD and NMC in general gives the fewest numbers of poor forecasts.

Regarding the predictability of ECMWF 500 and 1000 hPa height forecasts, see the curves in Figure 11 below:

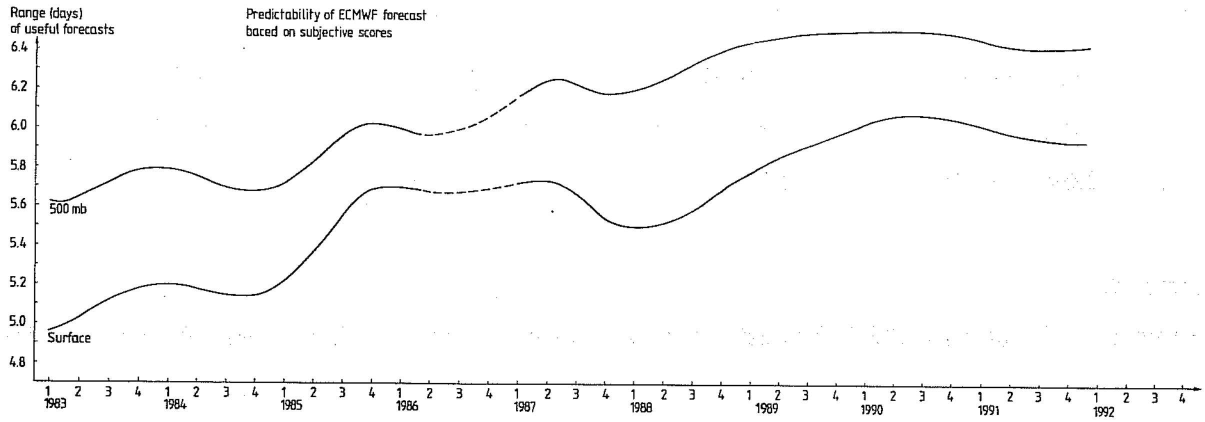


Fig 11 Development of predictability of ECMWF forecasts expressed in range (days) of useful forecasts. Smoothed curves. (Dashed lines for a short period 1986/87 indicates uncertainty because of an occasional interruption in verification procedure; changing to a new central computer system)

The predictability is determined by computing 12 months running means of subjective scores, where score has been defined as the limit of useful predictability. The last month included in the statistics is July 1992. The curves in Fig.11 indicate useful information to day 6.5 for 500 hPa. The corresponding number for surface is about 6.0. We notice that the levels in principle have been unchanged during the last 2 to 3 years. Note: using anomaly correlation (0.6) as the limit of predictability, the range of useful forecasts will be somewhat lower (around half a day)

2.2.2 Behavior of the model; some comments about inconsistency/consistency and skill

The day to day consistency has shown to be an important parameter for the forecaster during the last years. A rather common opinion is that high consistency is a sign for high skill. In another word: if for instance to day's D+5 is consistent with yesterdays D+6, then D+5 is more reliable than normal.
And, on the contrary, that inconsistency or "jumpiness" indicates less reliable guidance from the last run. Another rather common opinion is that ECMWF is more "jumpy" than the other models used in the operational environment, to a great part depending on its over activity.
The question is, if such statements are correct?

We have made some investigations based on the scores from synoptic assassment. Mostly we have compared D+5 with D+6 to determine the day to day variability; where score ≥ 4 means consistency by definition and score ≤ 2 means inconsistency. To get information about skill, D+5 has been compared with the verifying analysis.

Among the results, we can mention:

a) for **April 1992** and for northern part of Europe (500hPa)

ECMWF		UKMO		NMC		DWD	
consis.	skill	consis.	skill	consis.	skill	consis.	skill
3.69	3.75	3.58	3.56	3.55	3.66	3.69	3.62

Comments:

* for this special month ECMWF forecasts are more consistent than the other models (the same as DWD).

* the ECMWF has also the highest score

* the cases where consistency is high (score ≥ 4), we find that the scores have increased to:
ECMWF: to 4.12 (from 3.75; 50 % of all cases)
UKMO: to 3.75 (from 3.56; 43 % of all cases)
NMC: to 3.88 (from 3.66; 48 % of all cases)
DWD: to 3.75 (from 3.62; 47 % of all cases)

* the correlation between skill and consistency for ECMWF is 0.38.

b) for **May 1992** we found that for ECMWF correlation between skill and consistency (D+5/D+6) is 0.52, which is surprisingly high. A possible explanation is the often persistent weather situations during this month.

c) on the other hand for the period **15 February to 15 April** we found:

* the correlation between skill and consistency (D+5/D+6) for ECMWF is 0.24; that is lower than corresponding figures for both April and May

* the corresponding correlation for DWD is zero.

For this period (15 February to 15 April), the following can also be depicted, which can serve as illustrations of the great variations from month to month and between different models:

* the skill for ECMWF was scored 3.74 and consistency 4.06

* the corresponding figures for DWD is 3.77 and 4.15

* for both models the percentage number of consistent forecasts is around 50%

* the percentage number with skill ≥4 (good forecasts) - when the consistency is ≥4 - is: for ECMWF: 68 % and for DWD: 59 %. Compare this with the below:
* the percentage number with skill ≤2 (poor forecasts) - when the consistency is ≥4 - is: for ECMWF: 13 % and for DWD: 3 %

d) To further illustrate the relationship between consistency and skill we have for the period between 15 February and 31 July 1992 compared D+4 with D+5 and D+5 with D+6 respectively to pick out the cases when both forecasts are poor at the same time and also determine the percentage number of inconsistency and consistency between the two consecutive forecasts. We have utilized two values of the limit score for consistency to see the differences.
The **result** became: In the special situations with two poor consecutive forecasts verifying on the same time the probability that they are *consistent is 3 to 4 times greater than an inconsistent state.*
Reasonable numbers are around 15 % for inconsistency and 50 to 60 % for consistency, about 30 % "ambiguous". For details see the table below!

	D+4/D+5	D+5/D+6
Percentage number of cases when both forecasts were bad	14 %	14 %
From these special cases, percentage number of inconsistency (score≤ 2)	11 %	19 %
The percentage number of consistency, using the score 4(minus) as the limit	48 %	52 %
The percentage number of consistency, using the score 3(plus) as the limit	67 %	56 %

Note: plus and minus has both the value .25 in the 5 degree scale used in the subjective assessment.

e) We have found a sort of **forecasting rules** by comparing ECMWF consecutive forecasts and also to compare outputs from the different models:

* Suppose for instance that ECMWF D+7 gives the same solution as D+6 - valid at the same time - but D+5 (from the last run) has another proposal for the same time. In these cases the probability that D+5 gives a better guidance, is greater than D+7/D+6 forecasts is the most correct.

* The same conclusions you can draw when comparing D+6/D+5 with D+4. That is, it is a greater probability that the last forecast run gives better guidance than the yesterday's with a different solution, though this one is alike the run from the day before.

* If all four models give the same proposal for day D+5, the probability is more than 80 % that the forecast is good.

* If the ECMWF D+7, D+6 and D+5 - valid at the same time - gives the same solution, the probability is more than 90 % that the forecast is good.

f) **discussions:**

The ECMWF T213 model is more energetic than the T106 which means that the new model is able to develop intense systems during all lead times and thus take greater risks in the medium range which often cause larger RMS error and also sometimes great variations in the consecutive runs.
Our investigations, however, indicate that the forecasts from ECMWF are not significantly more inconsistent than the other models. We also conclude that consistent forecasts in general give somewhat higher skill than normally. This circumstance is not, however, significant and should not be exaggerated.
On the other hand it is important to state that deterministic forecasting always gives variations in predictive skill and that all predictive systems show more or less inconsistency.
It is also worth mentioning that all discussions about consistency and skill is most relevant for large scale atmospheric circulation pattern. Reasonable numbers are around 15 % for inconsistency and 50 to 70 % for consistency.

3. DMO NEAR SURFACE PARAMETERS

3.1 Background

Forecasted values of near surface parameters such as 2-m temperatures, 10-m wind, precipitation and cloudiness we have for many years utilized in the operational weather service as useful guidance.

When using a grid point from the original - gaussian - grid to represent a station or an area, we have with great care chosed the most suitable with consideration of the distance, height, land or sea and so on.

Statistical interpretations of output from numerical weather models (e. g. ECMWF) has been developed to catch the impact of the local conditions (such as topography and surface characteristics) on DMO near surface parameters. The techniques we used are PPM, MOS and Kalmanfiltering.

Below is depicted some verification results. To get an idea of the skill a comparison with chance and persistence is made.

3.2 2-m temperature

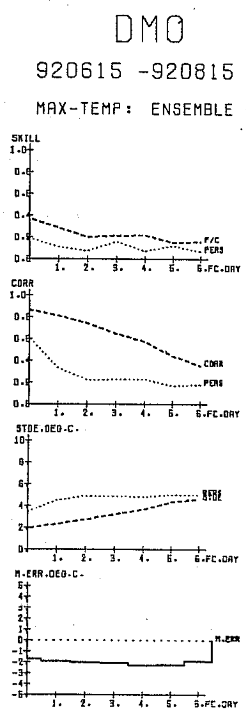


Fig 12 Verification of DMO 2-m maxtemp 15 June-15 August 1992. 4 Sites in northern Sweden (inland). From top: Skill of DMO and persistence compared with chance; correlation; standard dev. of mean error and at the bottom mean error.

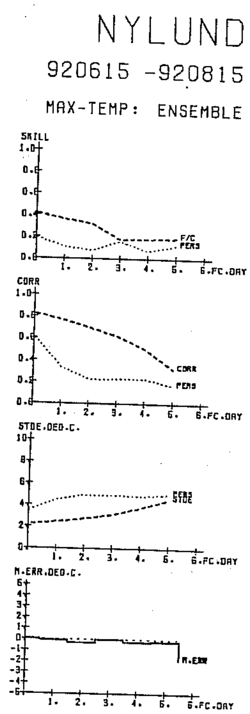


Fig 13 As fig 12 but verification of a statistical interpretation (PPM)

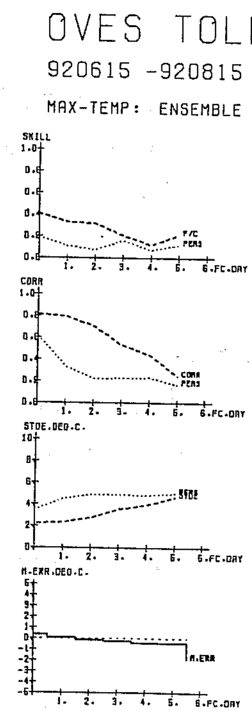


Fig 14 As fig 13 but verification of another statistical interpretation (MOS)

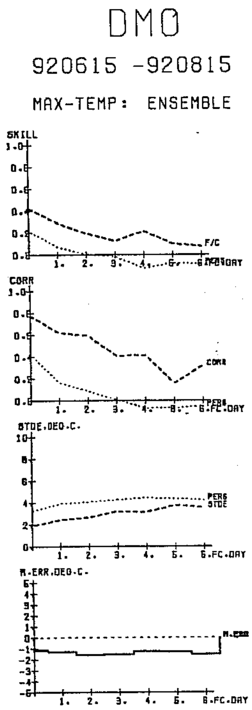


Fig 15 As fig 12 but for 4 sites in southern part of Sweden

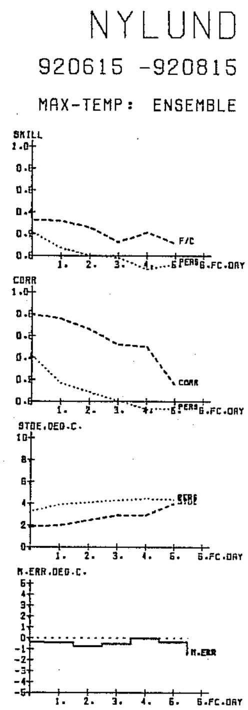


Fig 16 As fig 13 but for 4 sites in southern part of Sweden

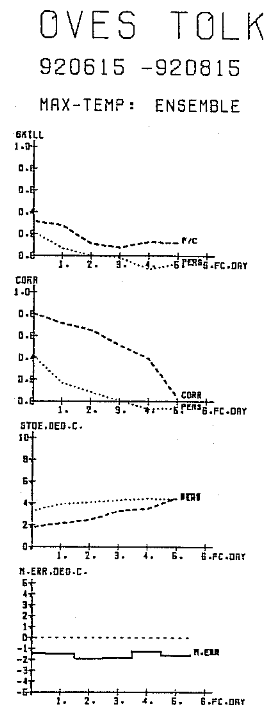


Fig 17 As fig 14 but for 4 sites in southern part of Sweden

Comments:

* The DMO 2-m (maximum) temperatures in northern Sweden are too low. This tendency is less pronounced in the statistical interpretation (both PPM and MOS).

* In southern Sweden the DMO 2m-temperatures often give better guidance than in the northern part of our country (less negative bias). We notice that MOS interpretation show a negative bias in southern Sweden.

* Furthermore we can conclude (without illustrations however):
- The negative bias in northern Sweden is most expressed during the spring and over snow covered ground.
- The DMO minimum temperature often has less negative bias compared with the maximum temperature.
- The DMO mean temperature is in most cases the best guidance compared for instance with maximum temperature.

3.3 Precipitation

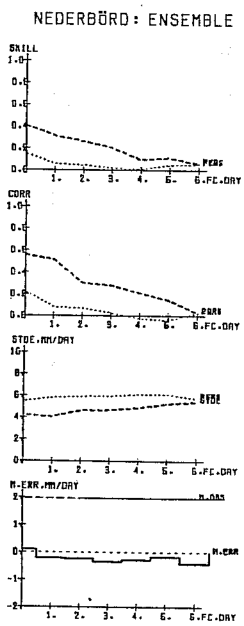


Fig 18 As fig 12 but for precipitation and for 14 stations through Sweden

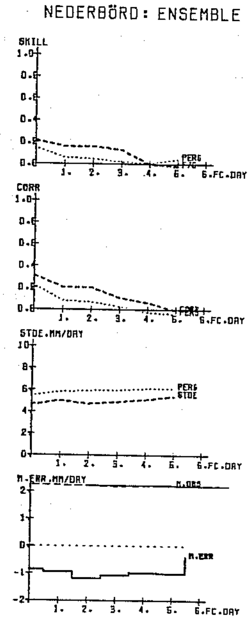


Fig 19 As fig 18 but verification of PPM

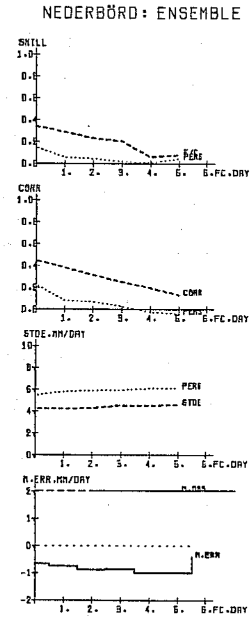
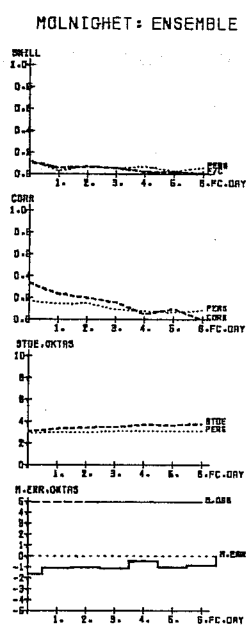


Fig 20 As fig 19 but verification of MOS

Comments:

The DMO precipitation normally gives good guidance to the forecasters with almost no bias.
We notice that both PPM and MOS has a noticeable negative bias.

3.4 Cloudiness



Comments to figure 21:

The ECMWF DMO cloud amount in general has a negative bias.
The skill against chance has a small value, but isn't different from the skill of persistence in a significant way.
The correlation coefficient, however, is greater than persistence out to day 3.

Fig 21 As fig 12 but for total cloudiness (14 sites)

Subjectively assessed ECMWF DMO cloudiness:

To get a subjective view of the usefulness of the output from the model cloud scheme we have, on an experimental basis, depicted convective cloud and the three layer clouds on four different charts with proper traditional meteorological fields to show the stream pattern for respective levels. We have begun to compare the forecasts of clouds with multispectral classified clouds from satellite images and in that way try to determine the quality in the 5 degree scale and also utilized the technique with contingency tables. From our so far restricted material (summer), we can conclude:

- convective cloud scheme in general gives good guidance
- also forecasted high clouds seem to be fair reliable. In some cases, however, we have noticed clouds connected with a frontal system, somewhat too southerly or/and easterly.
- low clouds are sometimes underestimated (around 25 % of the cases)
- to a certain degree the middle clouds also are underestimated.

3.5 10-m wind

We conclude (without illustrations) that DMO has a tendency to forecast too low wind speed over sea. Over land there are also a negative bias as well; but smaller.

In the output from statistical interpretation (PPM), however, there is no bias.

3.6 Post processed products

Kalmanfiltering of temperature, wind, cloudiness and precipitation from ECMWF DMO delivered to customers, for instance energy production companies, has been used during recent years.

Verification and test normally shows improvement of quality, especially for temperature and wind for all lead times.

As an example see the table below, which shows the verification results of the mean temperature forecasts for the energy production (average of lead times from 6 to 30 hours). The verification period is May 1991 to April 1992 for 5 sites in Sweden. The verification score is Mean Absolute Error, MAE.

	Mean Absolute Error
DMO	2.6
Kalman	2.0
Forecaster	1.8

4 SUMMARY AND CONCLUSIONS

The product from ECMWF analysis and forecast system are for all ranges used extensively at SMHI; firstly the traditional meteorological fields but also in an increased extent DMO near surface parameters.
We try, by routine, to get information of the quality of the guidance by verification activities.
From that we can summarize:
* The new model,T213, has the ability to catch intense synoptic systems in medium range; often better than the older model,T106, and other models (UKMO, NMC, DWD). That ability (increased realism), however, entails at the same time greater risks; for instance good guidance regarding the amplitude of a low but error in the phase, which will show up negatively in the scores.

* Thus, in fact, we have noticed that the traditional scores - anomaly correlation, RMS error and S1-skill score - hasn't improved with the new model. The levels are about the same compared with the former model, regarding the stream pattern

* DMO near surface patterns have shown to be valuable in the forecasting procedure.
Precipitation perhaps gives the best guidance, though it is a difficult parameter to forecast.
Forcasts of maximum temperature often show a negative bias, especially in northern Sweden during the spring and over snowcovered area. The skill of the forecasts of total cloudiness in general has a low skill; not much above persistence.

* Our impression is, however, though problems in the beginning, that the new model has a high potential and with continued improvements in the future will give yet better information.

5. REFERENCES

Blondin C., Böttger H. 1987: The surface and sub-surface parametrisation scheme in the ECMWF forecasting system. Technical Memorandum No. 135.

ECMWF, all Newsletters: Changes to the opertional forecasting system and other articles covering meterological topics; for instance: " Good prediction of a severe storm over southern Sweden " June 1990.

ECMWF, all yearly reports: "Verification of ECMWF products in Member states"; see for instance the contributions from Sweden.

ECMWF: Meteorological Bulletin, Research Manuals.

ECMWF: User Guide to ECMWF products.