VALIDATION OF WEATHER FORECASTS DERIVED FROM LARGE SCALE SIMULATIONS USING DIRECT MODEL OUTPUT OR STATISTICAL INTERPRETATION

S. Kruizinga
Royal Netherlands Meteorological Institute
De Bilt, The Netherlands

1. INTRODUCTION

Typically a weather forecast for a given location or area consists of a general description of the weather events to be expected and a set of forecast values for various observable weather parameters at that location or in the area. Since the start of numerical weather forecasting these forecast values have been derived from the output of the numerical models either by subjective judgement of the forecaster or by objective statistical interpretation. In the last twenty years the importance of statistical interpretation has grown considerably and is nowadays used in many meteorological offices.

In the early days of numerical modelling when only 500 hPa maps were available interpretation was a necessity for all surface weather parameters. Nowadays the output of numerical models also contains fields with surface parameters. So it seems that statistical interpretation is only needed for very specific parameters like visibility etc. In this paper we wish to show that although Direct Model Output (DMO) for surface parameters already produces high quality forecasts it is still possible to improve these forecasts by statistical interpretation.

In this paper we will give a brief resumé of some characteristics of statistical interpretation in section 2, a very extensive introduction to statistical interpretation and verification is given in *Glahn et al.* 1991. In the sections 3 and 4 we will compare forecast skills of DMO and statistical interpretation for two very common weather parameters namely maximum temperature and occurrence of precipitation. In section 5 our conclusions will be given.

2. CHARACTERISTICS OF STATISTICAL INTERPRETATION

In a statistical interpretation system the forecast value for a specific surface parameter at a given location is computed by a statistically developed equation. The input values for the parameters are retrieved from the forecast of a numerical model. The derivation of the statistical forecast equation requires a time series of observations of the surface parameter in question the "predictand" together with a concurrent time series of either analyzed or forecast parameters retrieved from the numerical model the "predictors". The type of equation used depends strongly on the characteristics of the predictand. The coefficients of these equations are optimized on the basis of the available time series (the dependent data set).

2.1. Statistical Techniques

The most common statistical technique used for developing forecast equations is Multiple Linear Regression (MLR), *Klein et al.* (1959). This technique can be used for weather parameters with an almost normal probability distribution like temperature, wind components and wind velocity.

However, a lot of weather parameters have probability distributions which differ very much from the normal distribution. Examples are cloudiness with often high probabilities at both extremes or precipitation amount. In most cases these weather parameters are categorized before forecast equations are developed. The forecast equations for this type of weather parameters mostly result in a probability for a given category.

For the development of probabilistic equations several techniques are available. Examples are Logistic Regression (*Brelsford and Jones,* 1967), Regression Estimation of Event Probabilities (*Miller,* 1964) and Multiple Discriminant Analysis (*Miller,* 1962).

A very general technique is the analogue technique (*Kruizinga and Murphy,* 1983). In this technique a forecast map, e.g. 500 hPa height map, is retrieved from the model output and compared with a set of historical maps. The statistical summary of the weather of the most similar maps is used as forecast.

2.2. Perfect Prog and Model Output Statistics

For the derivation of forecast equations a time series of predictors and predictands is needed. In general two approaches are used for the development of statistical interpretation eqautions, Perfect Prog (*Klein et al.* 1959) and Model Output Statistics (*Glahn and Lowry,* 1972).

- Perfect Prog (PP)

  In this approach predictor values which are needed for the development of the equations are retrieved from the analyses of a numerical model or from observations. In the operational mode these equations are applied to the forecasts of a numerical model implicitly assuming that the forecasts are perfect. This type of equation can be applied to any numerical model and any forecast time
- Model Output Statistics (MOS)

  In this case the values for the predictors in the time series are retrieved from forecasts from the numerical model. The equations are applied to the same model output at the same forecast time. So for each model and each forecast time a separate equation is needed.

2.3. Applications of Statistical Interpretation

Statistical interpretation can be used for a lot of purposes. Here some of these objectives will be discussed:

- correction of bias and range of DMO

  In many cases DMO has a severe bias or a too large range of values. Simple linear regression or Kalman filtering (*Persson* or *Simonsen* in *Glahn et al.* 1991) can be used to correct these errors. The objective in this case is not to improve the forecast skill but to make the application of the forecasts more direct.

- combining several parameters into one forecast

  In this case the objective is to improve the forecast skill by combining several output parameters into one forecast of a surface weather parameter. This is especially useful at medium- and long-range forecasting.

- probabilistic forecasting

  Often the end-user wants to know if a certain critical threshold will be passed e.g. will frost occur or not. In such a case a probabilistic forecast giving the probability that the threshold will be passed is very useful for risk evaluation by the end-user.

- blending recent observations with model results

  Typically the forecaster has to make his forecast on the basis of model output that uses initial conditions of several hours earlier. Forecasts for the first few hours should be based on recent observations in combination with model output, *Unger* (1987)

2.4. Verification of Statistical Forecasts

The verification of statistical forecasts requires some special care. In contrast to field verification of numerical models which is mostly based on the comparison of forecast and analyzed maps at a given time the verification of statistical forecasts and DMO is based on the comparison of time series of forecasts and observations at a given site or small area. The verification results are strongly dependent on the "climate" of the observations in the verification period. Especially when two forecast systems are compared it is neccessary to verify both forecast systems on the same period and to filter out all days on which one of the forecasts is missing.

3. VERIFICATION RESULTS OF DMO AND MOS FOR TEMPERATURE

In this section we will compare the skill of the operational MOS forecast (based on the ECMWF model) for the daytime (06-18 GMT) maximum temperature at De Bilt with the skill of the 12 GMT two meter temperature T2M retrieved from the ECMWF model output at the nearest gridpoint. Both forecasts will be verified against the observed 12 GMT temperature (T12) as well as the observed daytime maximum temperature at De Bilt (Tx).

The MOS forecast uses different equations, based on linear regression, for different seasons. In winter (December to February) the main predictors are the 1000 hPa temperature at 12 GMT, the average maximum temperature of the analogues of the 500 hPa field (*Kruizinga and Murphy*, 1983), the east-west component of the 1000 hPa wind and the height of the 1000 hPa level. In summer (June to August) the 850 hPa temperature replaces the east-west component of the 1000 hPa wind. In both seasons the 1000 hPa temperature is the most important predictor at lead times up to about 96 hours. Thereafter the analogue temperature becomes very important. In figure 1 the yearly skill of the MOS maximum temperature forecast over the period 1981 upto 1991 inclusive is plotted. The MOS equations used in the period that will be verified were implemented in 1987, before that year mixed MOS/PP equations were used (*Lemcke and Kruizinga*, 1988).

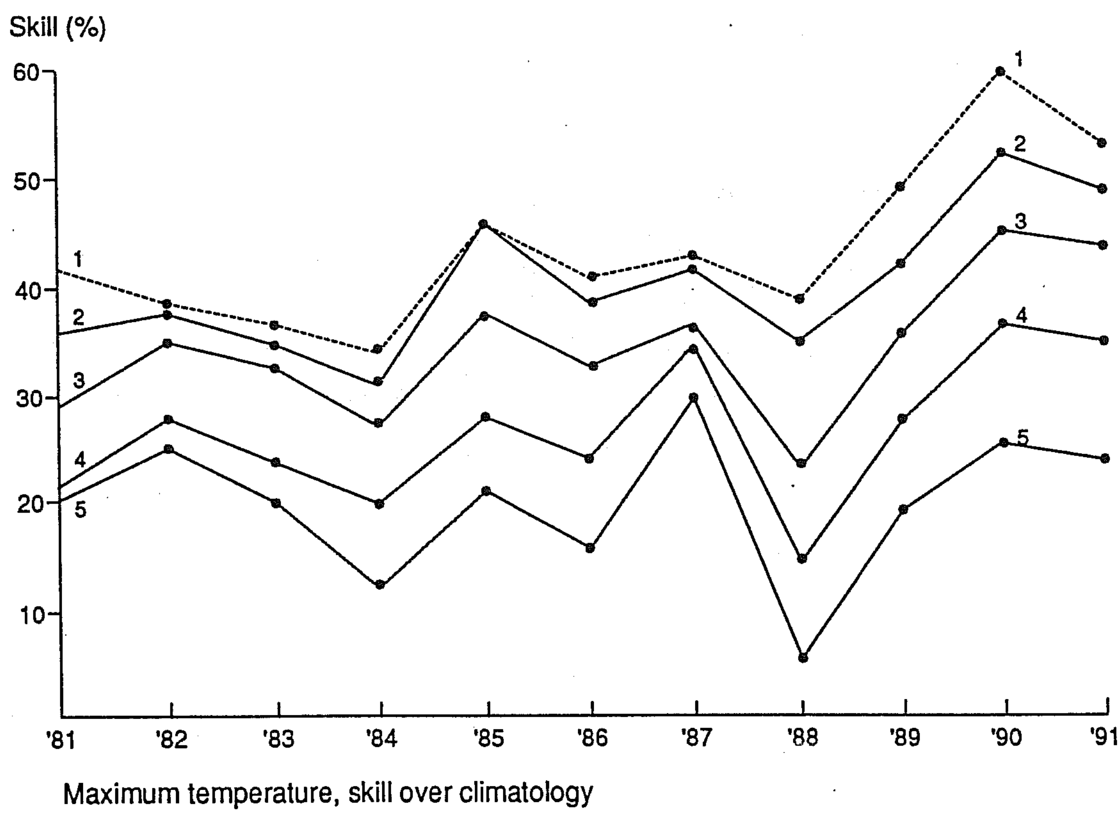

Fig. 1 Yearly skill scores for MOS maximum temperature forecasts for the years 1981 to 1991. Lead time in days, day 1 is based on +48 hours and so on.

The T2M DMO forecasts were retrieved from the MARS system at ECMWF on a Gaussian N80 grid. During winter 1991 and summer 1991 this was the basic grid of the ECMWF T106 model. In figure 2 some gridpoints in the neighboorhood of De Bilt are depicted as well as the location of De Bilt. The distance between De Bilt and the nearest gridpoint is about 35 km.

3.1. Verification score

Usually for parameters like maximum temperature the verification score used is Mean Absolute Error (MAE) or Root Mean Square Error (RMSE). Both these scores are consistent scores (*Daan*, 1984) but both scores are influenced by the bias of the forecasts as well as by differences in the standard deviation of forecast and observation. For this study we preferred the correlation coefficient as verification score. This score is not a consistent score for final forecasts but gives in our opinion a better measure for the applicability of the forecast parameter to an experienced forecaster. The use of this scoring rule made it possible to verify both forecasts, MOS and T2M, against both observed weather parameters, Tx and T12.

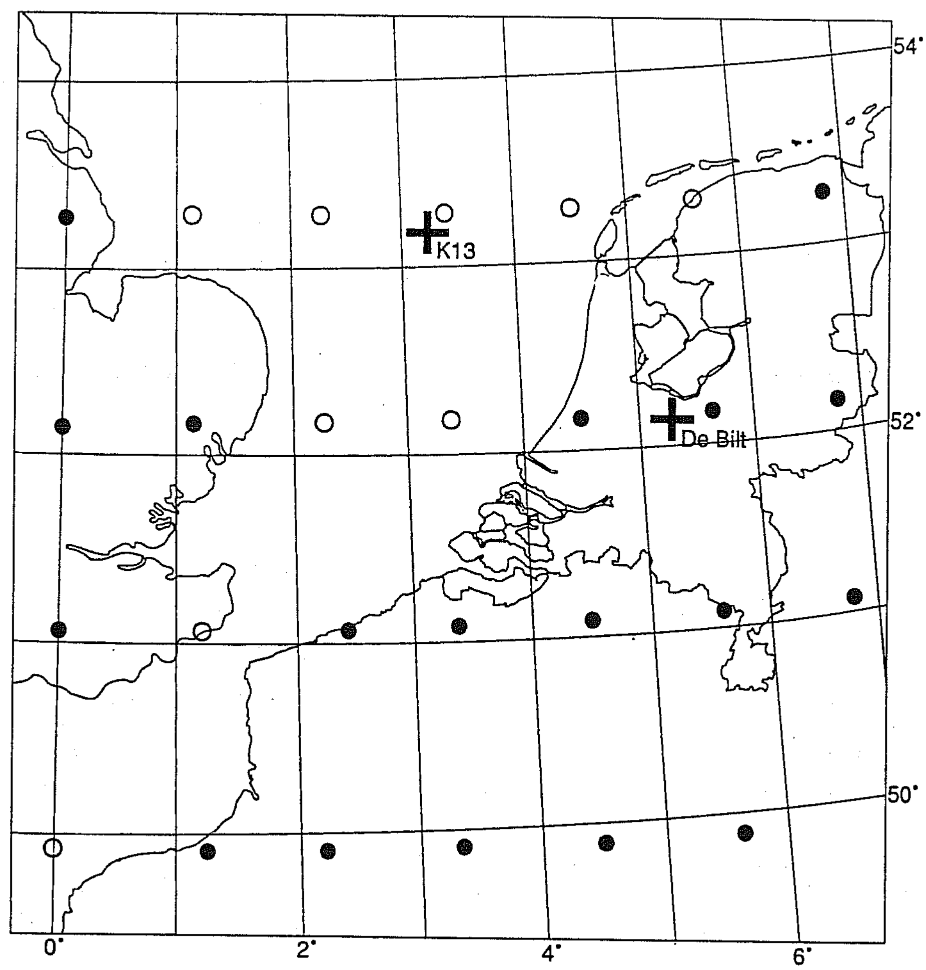

Fig. 2 Gridpoints of N80 Gaussian grid near to De Bilt. Open dots are sea points, filled dots are land points.

3.2. Results

The verification period covered one year of data, 1 December 1990 to 30 November 1991. In order to remove the seasonal change in all parameters the climatological normal of the maximum temperature was subtracted from all elements. Moreover, because the MOS-system uses four seasonal equations the data set was split into four seasonal datasets. In figure 3 the correlation coefficients (CC)

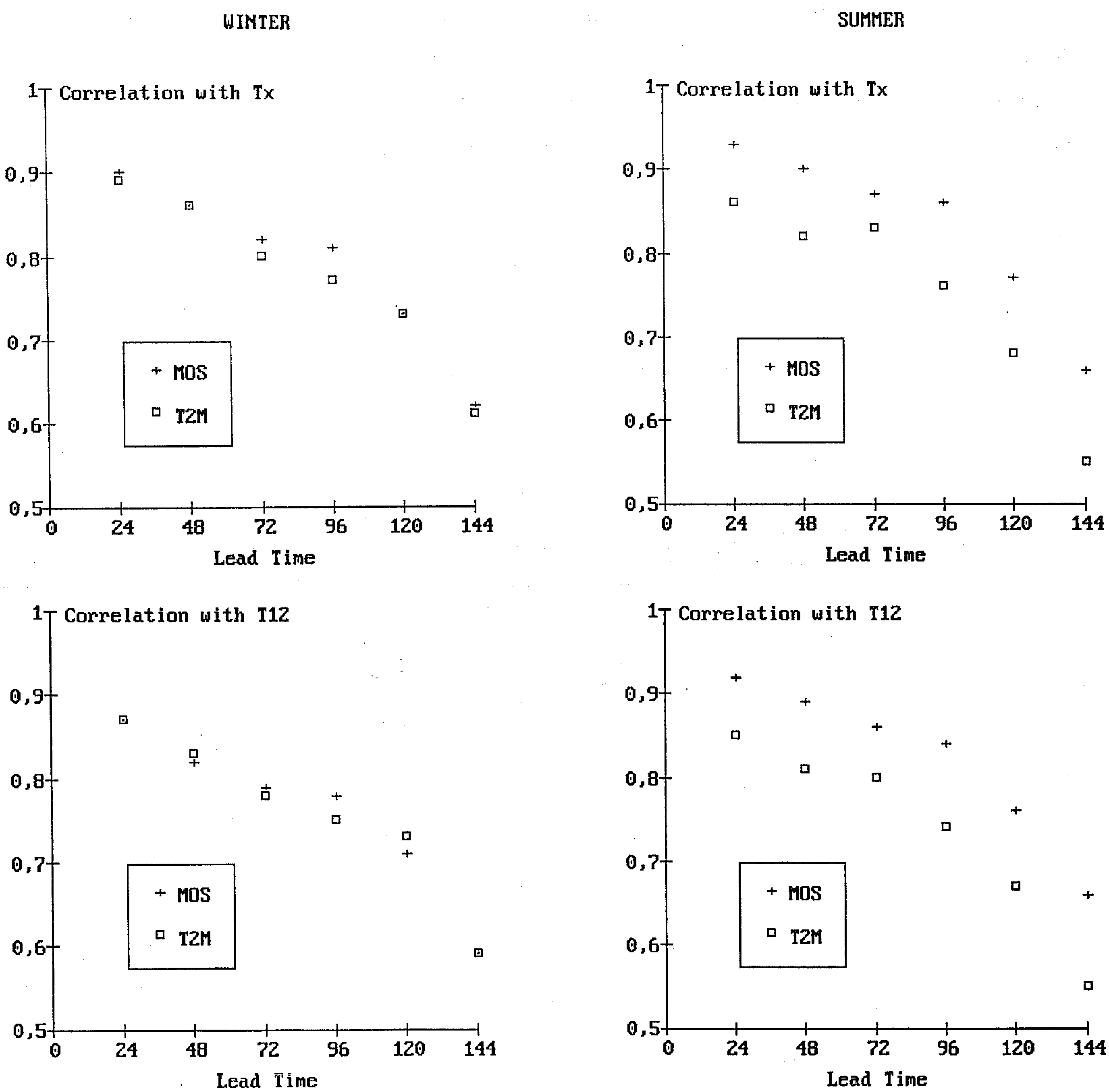

Fig. 3 Correlation coefficients of MOS and T2M with Tx and T12 respectively as function of lead time for winter and summer.

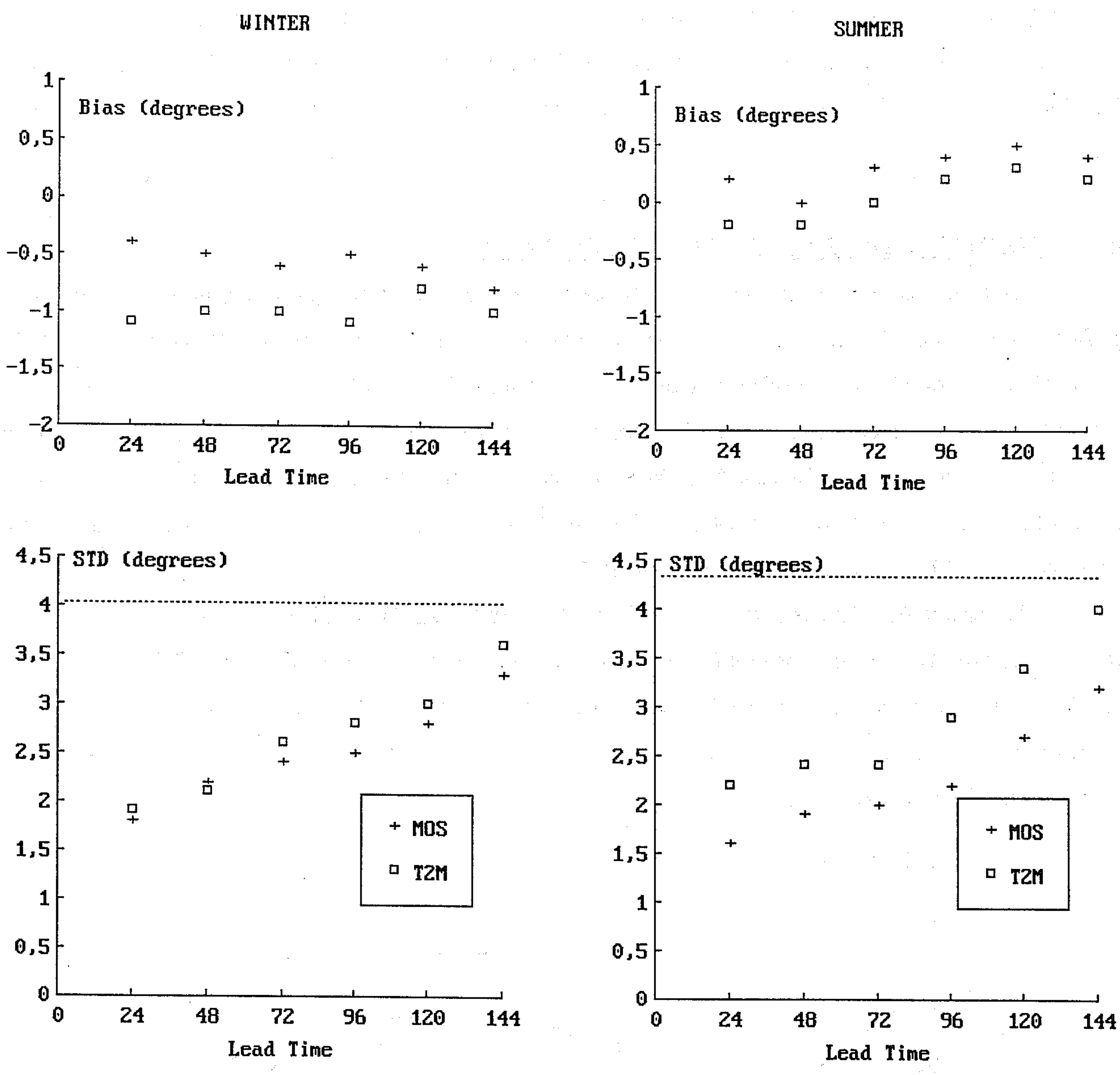

Fig. 4 Mean and standard deviation of forecast errors T2M-Tx and MOS-Tx. The dotted line gives the standard deviation of climatology.

obtained are depicted versus lead time. All these correlations were computed for those cases in which both forecasts for the appropiate lead time were available. (There were no missing observations).

In figure 3 the panels on the left show the winter correlations with temperature at 12 GMT (T12) and daytime maximum temperature Tx respectively. In the panels on the right the summer correlations are depicted. Although one would expect T2M to correlate better with T12 than with Tx this is not the case. Furthermore it is clear that T2M and MOS are of equal quality in winter whereas MOS outperforms DMO in summer. The latter holds also for spring and autumn, results of these seasons are not shown.

In figure 4 the mean and the standard deviation of the forecast error (forecast minus observed) for both forecasts MOS and T2M versus Tx only are given. In winter the mean error of T2M is in agreement with the mean difference between T12 and Tx, MOS shows a small bias of about 0,5 degrees. In summer both biasses are low. The results for the standard deviation confirm the results found with the correlation coefficient.

4. VERIFICATION RESULTS FOR PRECIPITATION

In this second experiment we will compare the precipitation forecasts of the model with the MOS probability forecasts. Again the precipitation forecasts were retrieved from the MARS system at ECMWF at the N80 Gaussgrid. At the gridpoint nearest to De Bilt the large scale precipitation and the convective precipitation were retrieved and added. The amount accumulated between +12 and +36 hours was used as day 1 (or +24) precipitation forecast RR, the accumulated amount between +36 and +60 was used as day 2 (or +48) forecast and so on. The observation used to verify this forecast was the precipitation amount observed in De Bilt between 00 GMT and 24 GMT.

The probability forecasts were taken from the operational statistical forecast system of the Dutch weather service. The forecast probability POP24 refers to the period 18 GMT to 18 GMT the next day and it is verified against the observed rainfall over the same period. For POP24 one logistic regression equation is used for the whole year and the most important predictor is the number of analogues with precipitation (*Lemcke and Kruizinga,* 1988).

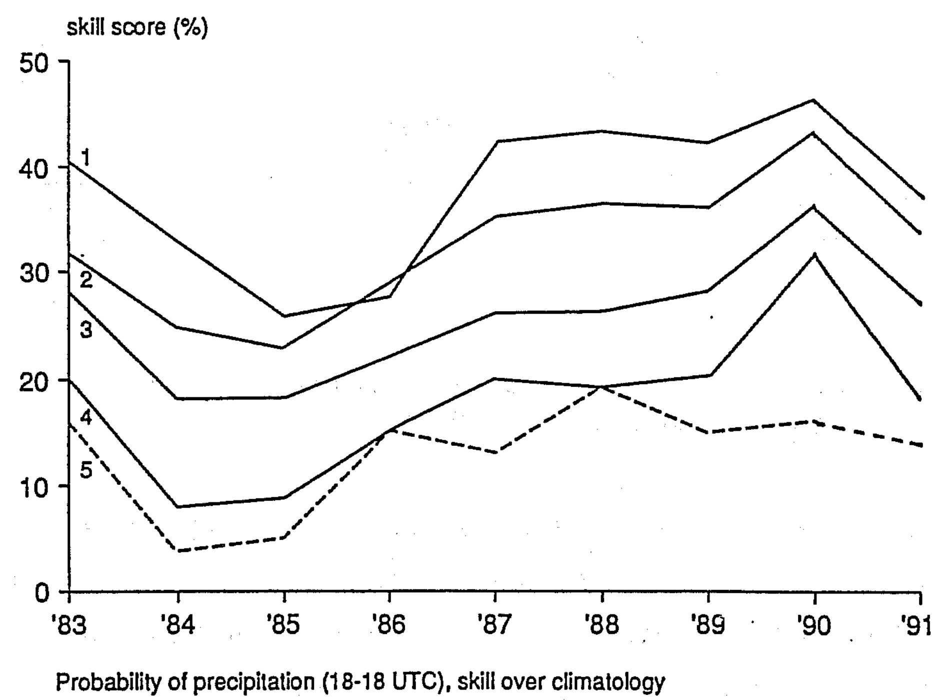

Fig. 5 Yearly skill scores of MOS POP24 forecast for the period 1983 to 1991.

4.1. Verification score

It is not possible to compare and verify the two forecasts directly. First we have to transform both forecasts to a common type of forecast. Therefore both forecasts were transformed to a yes/no forecast (whether precipitation will occur or not). The observations were also converted to yes/no by selecting yes if the observed precipitation was .3 mm or more. For the transformation of RR (the model precipitation) the same decision rule was used. The POP24 was transformed to yes/no by chosing yes if the probability is higher than 50 percent.

The verification of yes/no forecasts requires special scores. In this case we choose the Hanssen-Kuipers skill score PI (*Hanssen and Kuipers,* 1965). This score is computed from the two by two contingency table in the following way:

		Forecast		
		No	Yes	Total
Observation	No	N00	N01	N0t
	Yes	N10	N11	N1t
	Total	Nt0	Nt1	Ntt

$$PI = \frac{N00}{N0t} + \frac{N11}{N1t} - 1$$

4.2. Verification results

In the Netherlands the probability of precipitation in 24 hours (POP24) shows only a moderate seasonal change, ranging from 38 to 59 procent. Furthermore the MOS equation is valid throughout the whole year. Therefore we do not show seasonal verification results but only for the complete period from 1 December 1990 upto 30 November 1991 inclusive. This means that since ECMWF changed model resolution that the DMO data from 17 September 1991 onwards are interpolated (to a N80 Gaussian grid) data from the T213 model of ECMWF, before that date the data come from the T106 model on the N80 Gaussian grid.

In figure 6 the bias and the PI of both transformed forecasts are depicted. The DMO slightly overforecasts the occurrence of precipitation except at the 72 hour forecast. The MOS system overforecasts the occurrence of precipitation at short lead times and underforecasts at longer ranges. The panel with the PI shows a high skill for DMO at a lead time of 24 hours. For longer lead times the MOS system has an equal or higher skill than the DMO output. It should be emphasized that the verification shown here only refers to a part of the information given by both the types of forecasts. In the case of POP24 the probabilities give additional information for risk evaluation. In the case of DMO additional information is contained in the amounts given.

5. Conclusion

Current forecasting practices require forecast values for a wide range of weather parameters. For some of the parameters these forecast values can be derived directly from the numerical model (DMO) others have to be computed with the help of a statistical interpretation scheme. For other parameters only the last option is possible. In this paper it has been shown that for commonly used parameters available in DMO, statistical interpretation still can improve the overall skill of the forecast. From these results we conclude that in a local weather office two programmes should run in parallel:

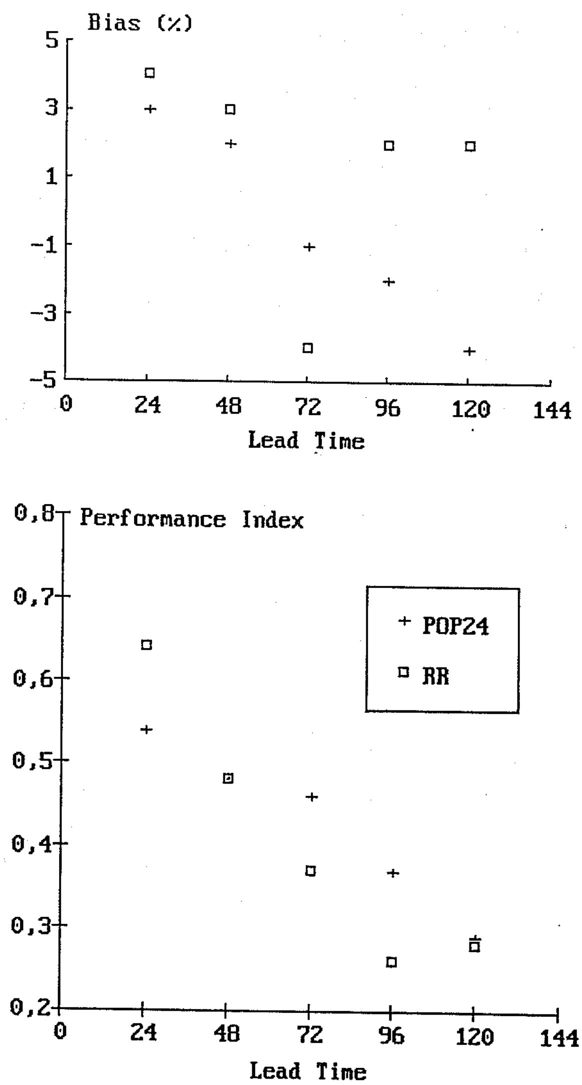

Fig. 6 Bias and Performance Index of MOS and DMO versus lead time.

- a DMO programme: In this programme forecast values are retrieved from a numerical model with high quality interpolation schemes. The resulting forecasts are routinely verified versus observational data at station sites. From these results the skill of DMO forecasts at non station sites can be estimated.
- a statistical interpretation programme: In this programme forecast values for DMO parameters, for non-DMO parameters and probabilistic forecasts are computed with statistical interpretation equations. Again the resulting forecasts are verified against the observational data. However, these results are strictly local but can support the need for statistical interpretation at more stations.

Forecasters must have access to both types of objective weather forecasts and should be informed on the skill of both types of weather forecasts on a regular basis.

References

Daan, H.,1984:Scoring rules in forecast verification. PSMP Report No. 4, WMO Geneva 1984

Brelsford, W.M. and R.H. Jones, 1967: Estimating probabilities, Mon. Wea. Rev.,95, 570-576

Glahn, H.R. and D.A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. J. Appl. Meteor.,11, 1203-1211

Glahn, H.R., A.H. Murphy, L.J. Wilson and J.J. Jensenius Jr.,1991:Lectures presented at the WMO training workshop on interpretation of NWP products in terms of local weather phenomena and their verification, PSMP Report No. 34, WMO/TD No. 421, Geneva 1991.

Hanssen, A.W. and W.J.A. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. Mededelingen en Verhandelingen No. 81, KNMI, De Bilt.

Klein, W.H., B.M. Lewis and I. Enger, 1959: Objective prediction of five-day mean temperature during winter. J. Meteor.,16, 672-682.

Kruizinga, S. and A. H. Murphy, 1983: Use of an Analogue Procedure to Formulate Objective Probabilistic Temperature Forecasts in The Netherlands, Mon. Wea. Rev.,111, 2244-2254

Lemcke, C. and S. Kruizinga, 1988: Model Output Statistics Forecasts: Three years of operational experience in The Netherlands, Mon. Wea. Rev.,116,1077-1090

Miller, R.G., 1962: Statistical Prediction by Discriminant Analysis, Meteorological Monograph Vol. 4, No. 25, Amer. Meteor. Soc. Boston, 53 pp.

Miller, R.G., 1964: Regression estimation of event probabilities. Travelers Research Center, Hartford, Contract CWB-10704, Tech. Rep. No. 1, 153 pp.

Unger, D.A., 1987: Updated MOS ceiling height and visibility forecasts from the local AFOS MOS program, Preprints Tenth Conference on Probability and Statistics on Atmospheric Sciences, Edmonton, Alta., Amer. Meteor. Soc., 18-25.