USE AND VERIFICATION OF EPS PRODUCTS

Seijo Kruizinga
Royal Netherlands Meteorological Institute (KNMI)
De Bilt, The Netherlands

1. Introduction.

Shortly after the introduction of the Ensemble Prediction System (EPS) forecasters at KNMI started with experimental internal forecasts for days 6 to 9. After the substantial enhancement of EPS introduced in December 1996 these experiments were intensified and forecasts were issued on a limited scale. Especially the successful support given to the large skating event "Elfstedentocht" stimulated the use of EPS products. In this paper the applications of EPS are briefly described and furthermore some verification results are reported.

2. Use of products.

The products of the Ensemble Prediction System of ECMWF are applied in operational mode in the following ways:

- Objective interpretation of 500 hPa maps.
  Each 500 hPa map of the ensemble of a given forecast step is classified according to the objective classification system P27 (Kruizinga, 1979). The forecast probability distribution of maximum temperature and minimum temperature is computed by averaging the conditional climatologies of the classes selected. The ranked probability score shows skill for both elements ranging from 31% (day 6) to 15% (day 9) and 25% to 10% for max and min respectively (Mureau and Plomp, 1997 and ECMWF, Verification report 1997).

- Subjective interpretation of cluster maps.
  The 500 hPa maps of the ensemble are clustered into a small number of clusters. The cluster maps are subjectively classified into the Gross Wetterlagen. The conditional climatologies of the selected classes are averaged and a forecast for the average maximum temperature for the next week (Monday to Sunday) is issued on Thursday each week. The skill score (based on the mean absolute error) over climatology is 12% and over persistence 27%.

- Support for large skating events.
  The "Elfstedentocht", a skating event with 20.000 participants and approximatedly 500.000 local spectators, takes a long preparation. The go/no go decision has to be taken 3 days ahead and the preparations start already earlier. Last year the organizing committee was supported by forecasts from EPS. The information given consisted of plume diagrams for 2-meter temperature, wind speed and precipitation and additonally a plume diagram for ice-thickness computed with a local model

(ECMWF Newsletter 1996/97).

- Dissemination to users of simplified plumes.
  In September 1997 KNMI started a fax service with simplified plume diagrams for temperature, wind and precipitation. In the plume diagrams the 10%, 25%, 50%, 75% and 90% percentiles are plotted versus lead time.

3. Verification of flow pattern indices.

The 500 hPa forecasts of ECMWF are verified at KNMI using an objective flow pattern classification system, P27. This classification system is based on three flow pattern indices Zonality, Meridionality and Cyclonality. These indices were derived using the EOF-technique (Kruizinga, 1979). The performance of the Ensemble Prediction System of ECMWF was studied by comparing the time series of forecast values of these indices with time series of observed values. In this paper the results for Cyclonality are reported. The period studied for EPS covered the first four months of 1997. For reference, however, we first verified the operational (deterministic) model over the same period in the two previous years. The results are shown in Fig. 1. In this figure the standard deviations (Std) of the observed and forecast time series are plotted as well as the bias (ME) and the root mean square error (RMS) of the forecast series. As can be seen the bias is almost negligible and the standard deviation is nearly constant over the whole forecast range. This means that the forecasts of the ECMWF model are consistent with climatology. Furthermore the RMSE is, as expected, very small for early forecast ranges and shows skill up to day 7.

In Figs. 2 and 3 the verification results for the operational model and EPS for the first four months of 1997 are depicted. Fig. 2 shows as a function of lead time:

RMS.Oper: The RMSE of the deterministic operational model. These results show great resemblance with the results found for the previous years.

RMS.EPS: The RMSE of the ensemble mean Cyclonality.

SPR.EPS: The average spread of the ensemble. This average spread was computed by averaging the daily variances of the ensembles.

RMS.MOS: The RMSE of a MOS corrected forecast of Cyclonality. The MOS equations were derived on the basis of the data of the 2 previous years and contained only one predictor: the Cyclonality predicted by the ECMWF model.

As can be seen the ensemble mean clearly improves on the operational model and, moreover, the RMSE of the ensemble mean is close to the average spread of the ensemble which is required for a correct Ensemble Prediction System. However, in terms of RMSE the simple MOS scheme performs just as well, suggesting that with such a scheme the same result can be obtained. Fig. 3 shows that EPS has additional qualities. In this figure the standard deviations of the time series of the Cyclonality of the operational model, the ensemble mean and the MOS scheme respectively are plotted versus lead time. The standard deviation of the observations is added for reference. As can be seen the operational model overforecasts the Cyclonality during this period. In contrast, the standard deviation of the ensemble mean decreases with increasing lead time. The MOS forecast shows a very strong decrease. Therefore, although MOS and ensemble mean are almost equal in RMSE sense the ensemble mean should be preferred from day 6 onwards.

Apart from the improvement in quality of the forecast it is expected that an EPS system can produce information about the day to day variations in the predictability of the atmosphere or in other words, that the EPS system can forecast its own skill. In order to study this property we defined the credibility interval W77 as the interval between the 6th and 46th value of the ordered values of Cyclonality of the 51 members (including the control run) of the ensemble. When the ensemble represents the "real" distribution of the verifying observations 77% of these observations should be in that interval and, moreover, the hit frequency should not depend on the interval width. In order to verify this requirement we divided the dataset associated with a given forecast step into 5 subsets according to the width of W77. The lowest 20% into the first class and so on. In Fig. 4 the hit frequency of these five classes is plotted versus the average width for the +228 forecasts. As can be seen the hit frequency hardly depends on the interval width. This holds also for shorter forecast ranges down to about day 7. This means that in this forecast range a smaller W77 implies a higher forecast skill.

4. Verification of surface parameters.

We studied only two surface parameters: the temperature at 2 meter at 12 UTC at De Bilt (T12) and the 24 hour probability of precipitation at the same location (POP24). For the occurrence of precipitation a threshold value of 0.3 mm is applied. The forecast data were extracted from the so-called weather parameter files available at ECMWF which contain interpolated values of some surface parameters for several observation stations in Europe. These files contain forecast values from each of the members of the ensemble as well as from the control and the operational run.

4.1 Temperature at 2 meters at 12 UTC.

For the T12 we compared the skill of the average of the ensembles with the skill of the operational model and the skill of the local MOS forecast for Tx, the maximum temperature between 06 UTC and 18 UTC. It was assumed that Tx is strongly related to T12 and that Tx shows a bias when verified against T12. In order to remove the influence of this bias from the verification results we used the correlation coefficient as a measure for skill. Furthermore, in order to remove effects related to seasonal change the daily climatological mean of T12 is subtracted from all temperatures. In Fig. 5 the correlation between the time series of forecast values and observed values is plotted versus forecast time. From this figure it can be concluded that the MOS scheme improves slightly over the operational model but the average of the temperature of of the ensemble members shows clearly the highest skill. In Fig. 6 the standard deviation of these forecasts is plotted versus lead time. Both the ensemble mean and Tx tend to lower standard deviations at longer ranges. The standard deviaton of the forecast temperatures of the operational model is close to the standard deviation of the observations.

The relation between the spread in the ensemble and the skill of the forecast was studied in the same way as with the Cyclonality. In Figures 7 and 8 the hit frequency for credibility intervals of varying width is plotted versus the average width of these intervals for the +168 and +240 hour forecast respectively. Both figures show that the hit frequency depends on the interval width and at +168 the inserted regression line is significant at a 5% level. This means that there has to be some serious doubt about the quality of the skill forecast for this parameter.

4.2 Probability of Precipitation.

The forecast of the probability of precipitation was derived from the ensemble output by counting the number of members with precipitation amounts over 0.3 mm and dividing this number by the total number of ensemble members. In Fig. 9 the Brier scores of these forecasts are plotted for lead times from 24 hour up to 240 hour. In the same figure also the Brier scores of the current MOS scheme and the Brier scores of climatology are given. As can be seen the forecast derived from the ensemble has a very bad performance. In Fig. 10 the biases of each of these forecasts is depicted. The very large bias related to the forecasts derived from the ensemble is probably responsible for the bad Brier scores. This last assumption was tested by offering the forecast probability of both, the MOS scheme probability as well as the ensemble derived probability, to a logistic regression scheme. Fig. 11 shows the results in terms of Brier skill score over climatology. Clearly now both forecasts are comparable in performance. Note that although the regression results are tested on dependent material the comparison is a valid one because both are applied and tested under the same conditions.

5. Summary of conclusions.

The results shown is this paper lead to the following conclusions:

- the climatology of forecast Cyclonality from the operational model is consistent with the climatology of the observations over the whole forecast range.

- the RMS of the forecast error of the average of the Cyclonalities of the ensemble members is clearly smaller than the RMS error of the operational model and slightly smaller than the RMS error of a simple MOS derived forecast for Cyclonality. The standard deviation of the forecasts derived from the ensemble mean is clearly larger than the standard deviation of the MOS forecast indicating that EPS produces more pronounced forecasts with equal performance.

- for Cyclonality, a parameter in the free atmosphere, the spread in the ensemble is related to the skill of the forecast. This could not be established for the 2 meter temperature directly derived from EPS.

- for some surface parameters, 2 meter temperature and POP24, the skill of the forecasts derived from "averages" of the ensemble members show improved or equal skill in the early medium range compared to other systems currently used at KNMI.

In general it can be concluded that EPS products are very useful for operational medium range forecasting.

Acknowledgement. Kees Kok provided valuable comments on earlier versions of this paper.

References.

Mureau R. and H. Plomp 1997: Use of P27 Circulation Regimes for EPS clustering, Report of the expert meeting on ensemble prediction system, ECMWF, Reading, 17-18 June 1997.

Skating on EPS, ECMWF Newsletter, Number 74, Winter 1996/97.

Kruizinga S.,1979: Objective classification of daily 500 mbar patters, Sixth Conference on Probability and Statistics in Atmospheric Sciences, Banff, Alberta, Canada

Verification of ECMWF products in Member States and Co-operating states, ECMWF Report 1997.

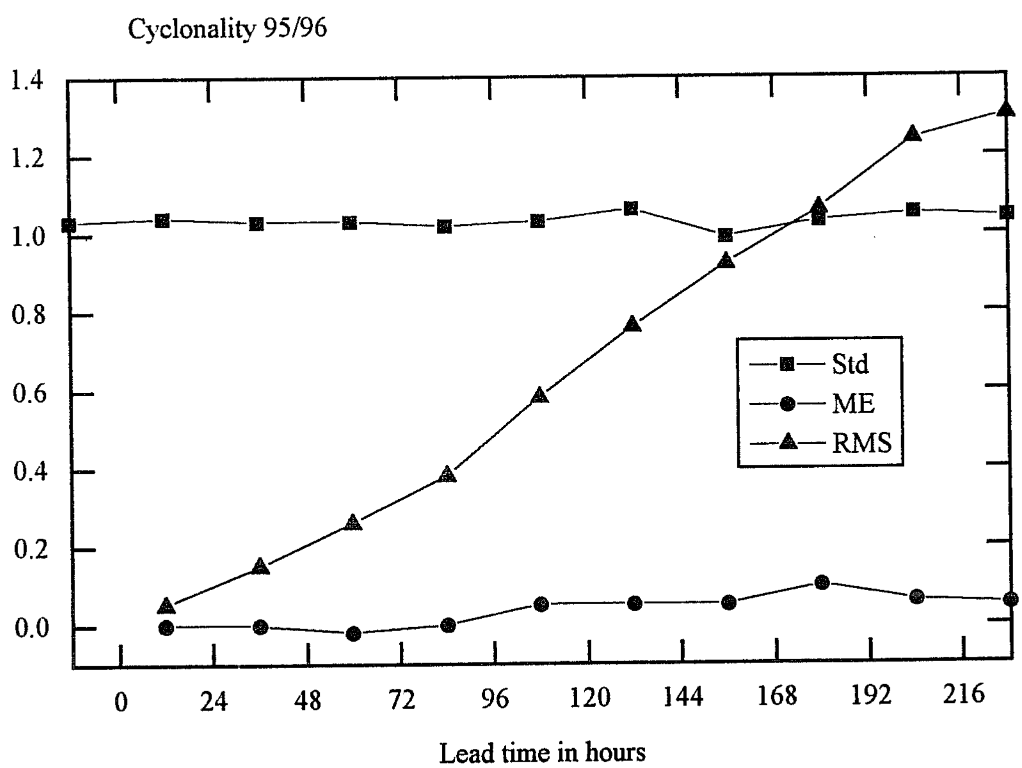

Figure 1: Standard deviation (Std) of observed and forecast cyclonality. Bias (ME) and root mean square error (RMS) of forecast cyclonality of the operational deterministic model.

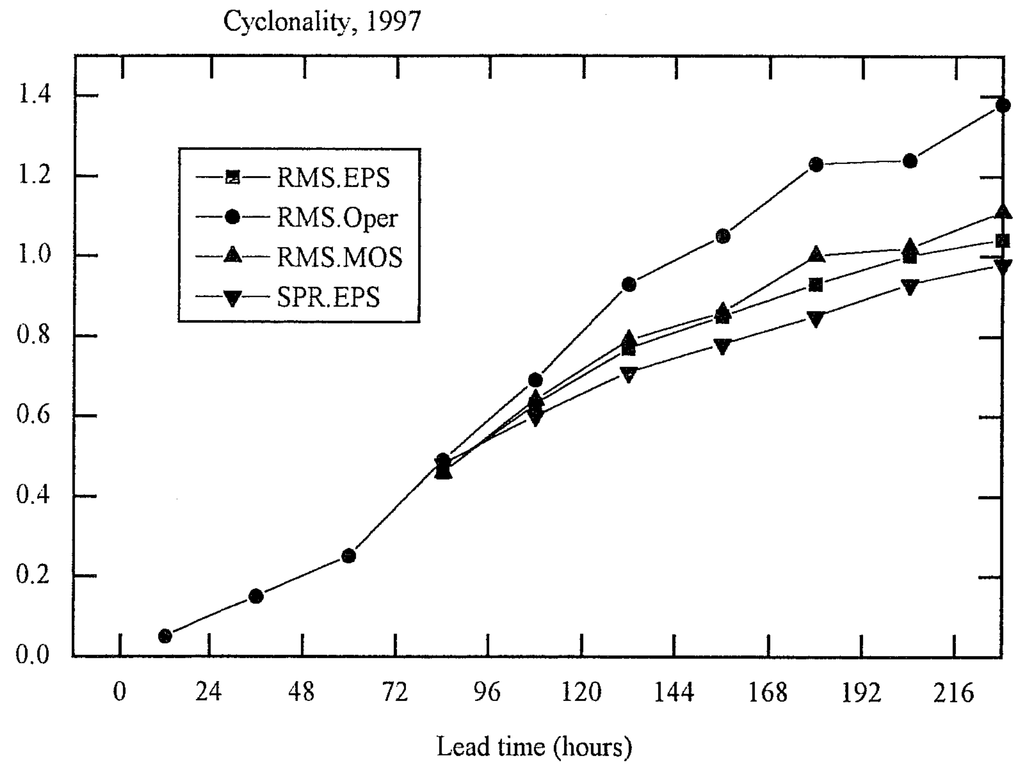

Figure 2: Root mean square errors versus lead time of EPS (RMS.EPS), operational deterministic model (RMS.Oper) and MOS (RMS.MOS) respectively in the first four months of 1997. SPR.EPS is the average variance of the ensembles over the same period.

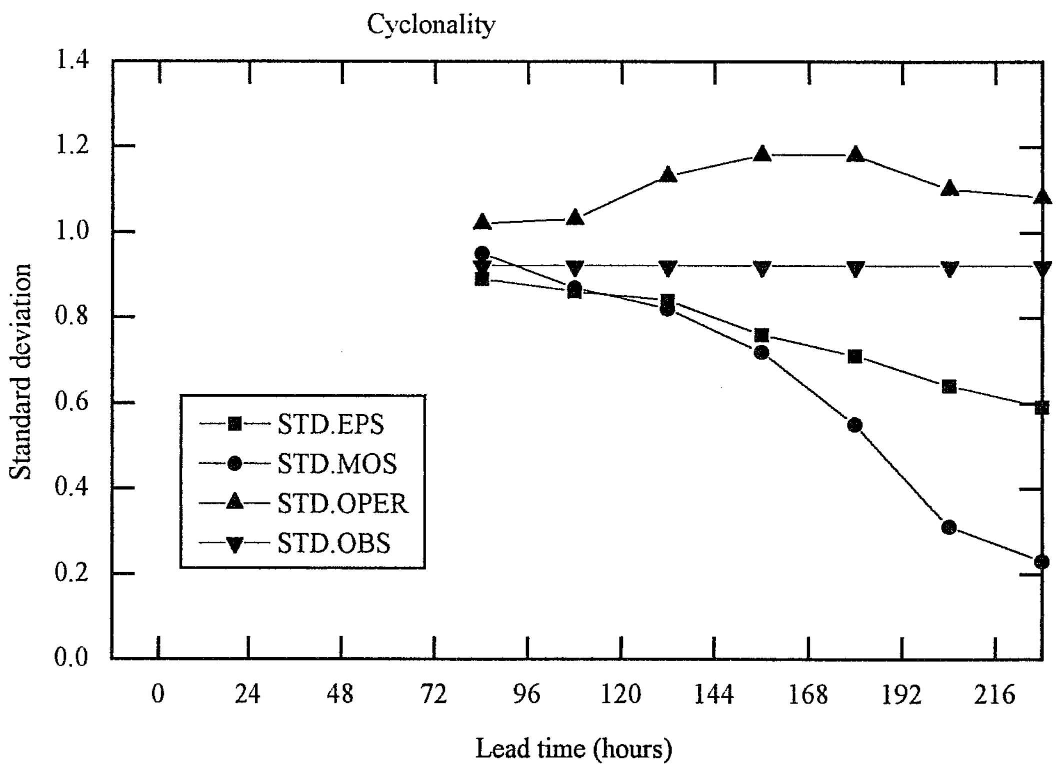

Figure 3: Standard deviation versus lead time of observed cyclonality and of forecast cyclonality of operational model, ensemble mean and MOS respectively.

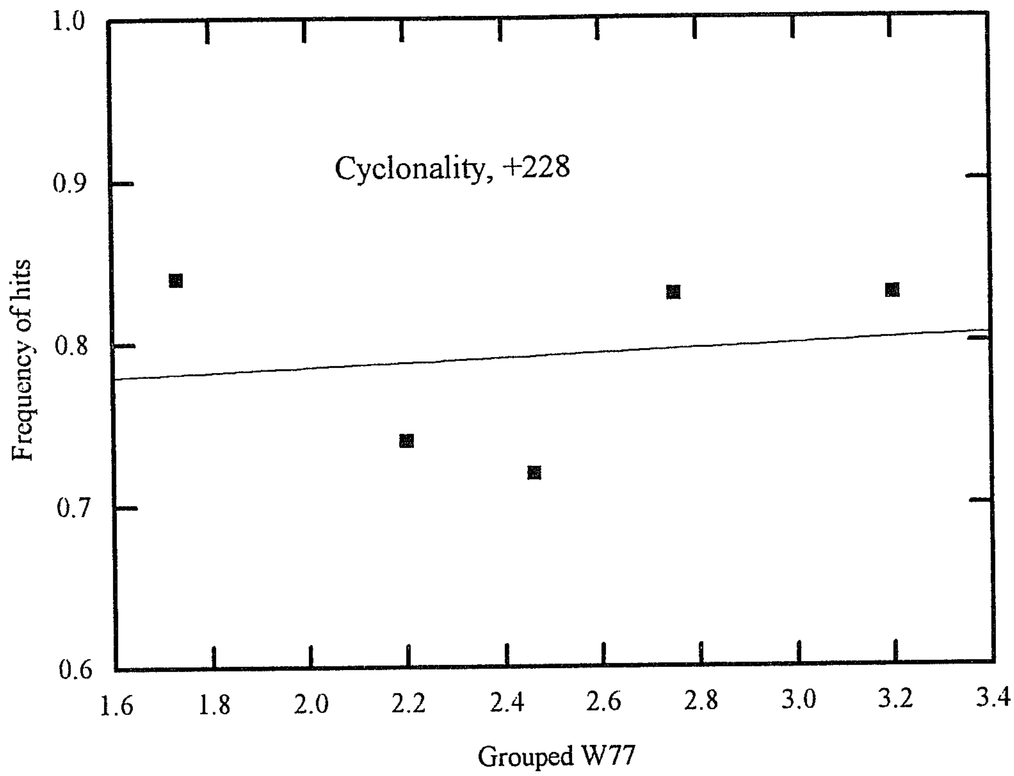

Figure 4: Hit frequency versus width of credibility interval W77.

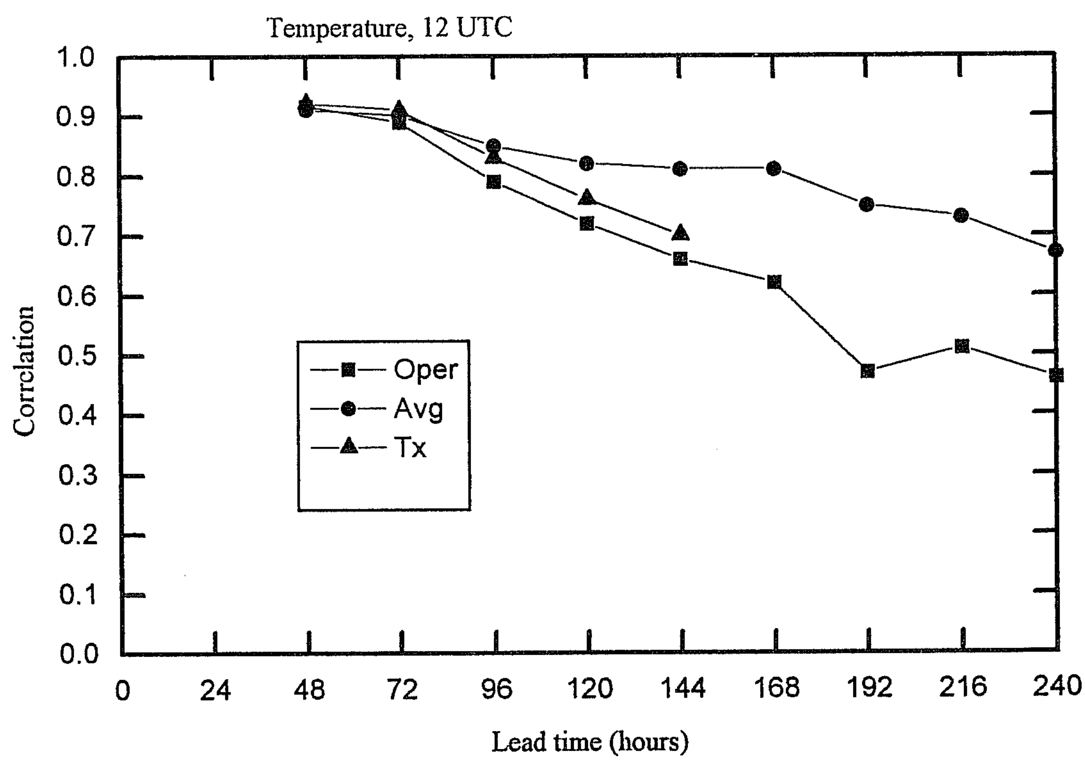

Figure 5: Correlation with observed temperature at 12 UTC, versus lead time, of the forecast time series of operational model (Oper), ensemble mean temperature (Avg) and Tx from a local MOS scheme respectively.

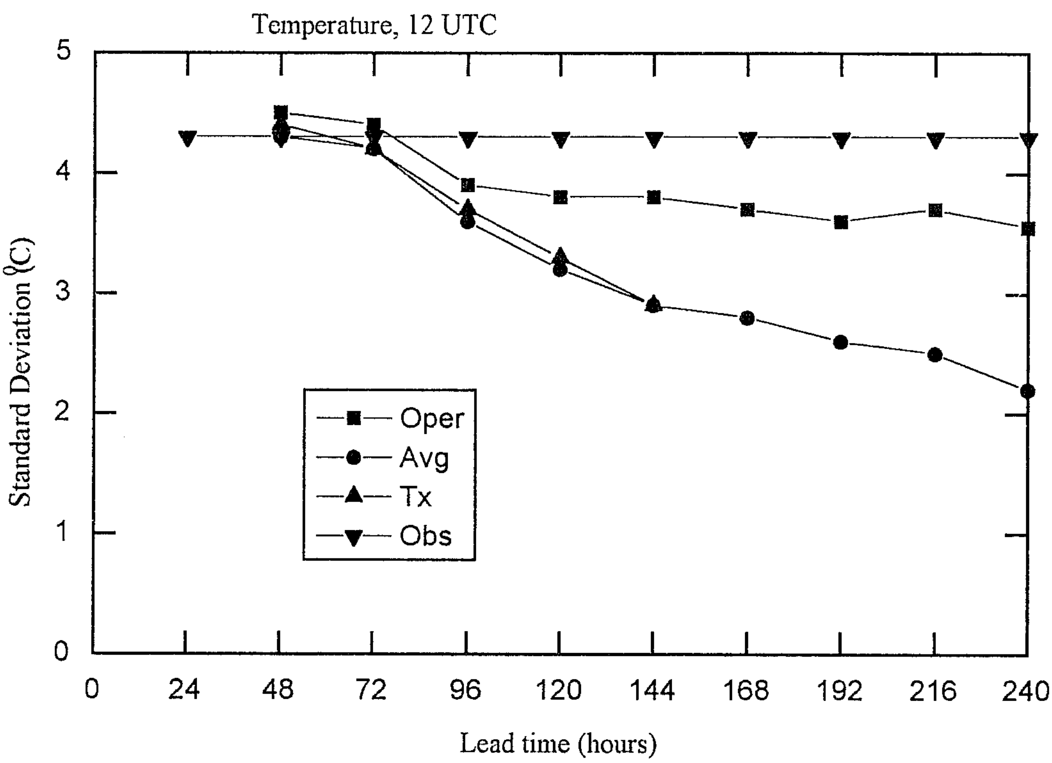

Figure 6: Standard deviation versus lead time of forecast temperature time series of operational model (Oper), ensemble mean (Avg), Tx from a local MOS scheme and observations respectively. Seasonal change was removed first.

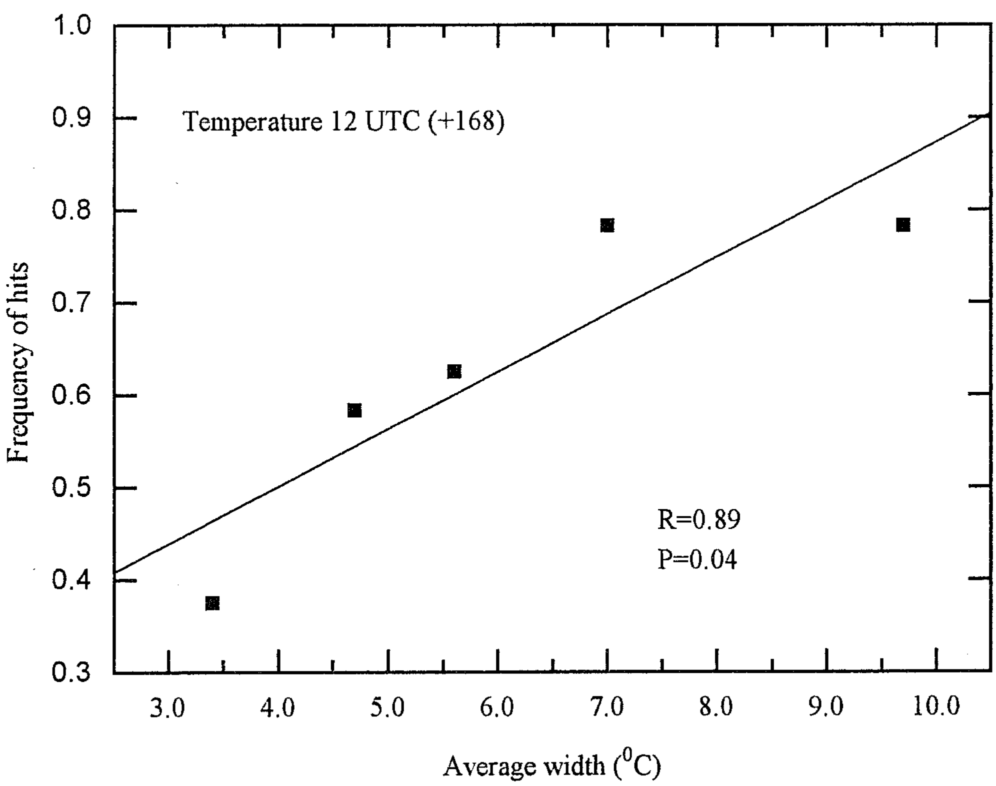

Figure 7: Hit frequency versus credibility interval width for ensemble 2 meter temperature forecasts with a lead of +168 hour.

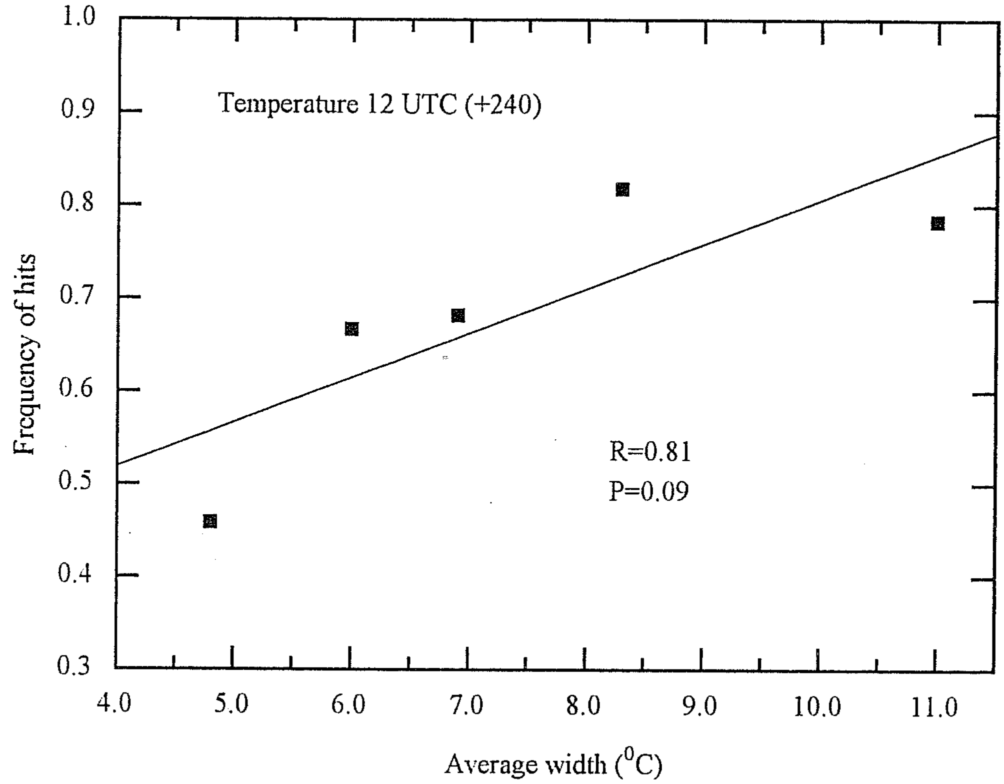

Figure 8: Hit frequency versus credibility interval width for ensemble 2 meter temperature forecasts at a lead time of +240 hours.

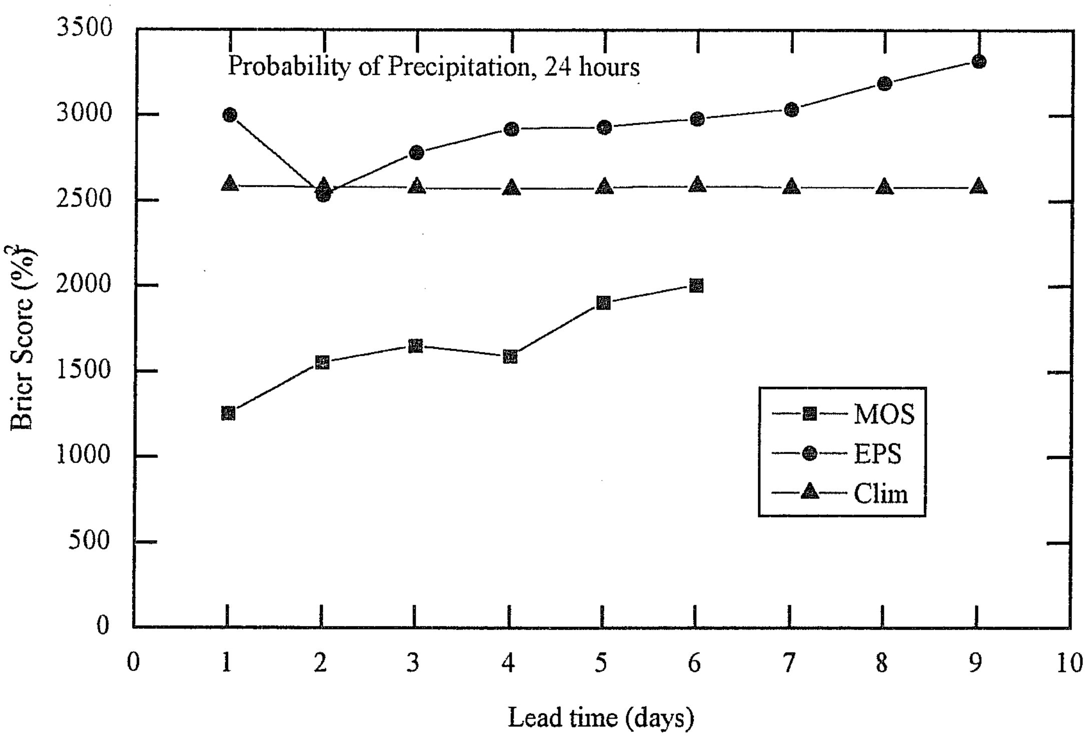

Figure 9: Brier score versus lead time of forecasts of the probability of precipitation from EPS, a local MOS scheme and climatology (Clim) respectively.

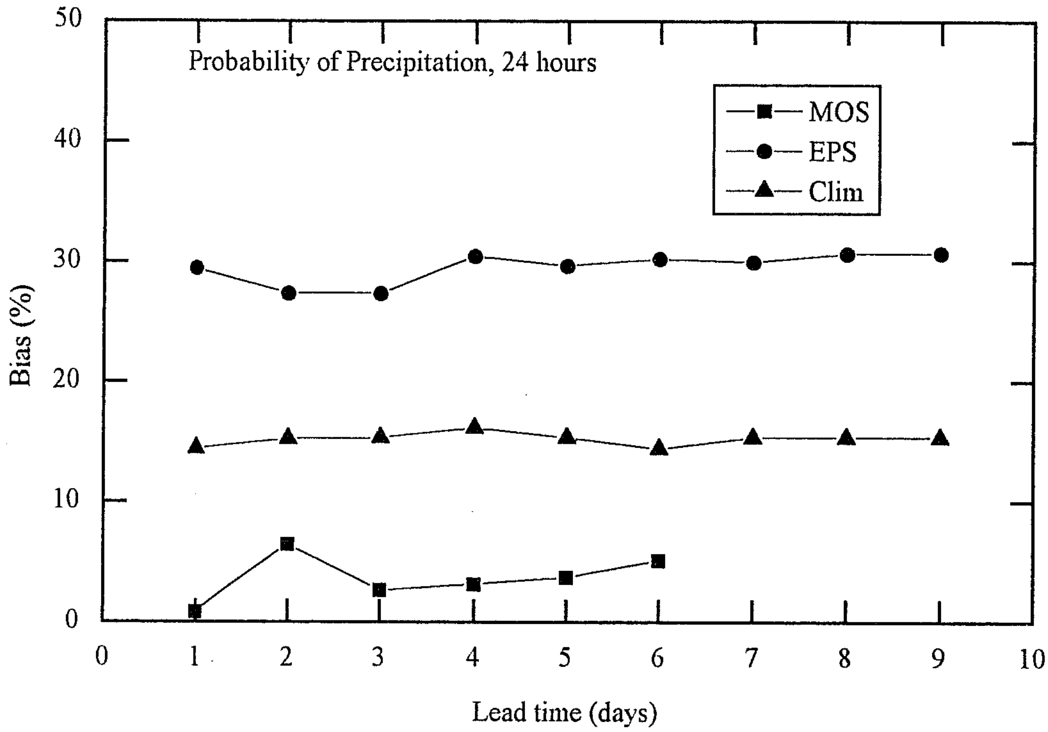

Figure 10: Bias versus lead time of forecasts of the probability of precipitation from EPS, a local MOS scheme and climatology (Clim) respectively.

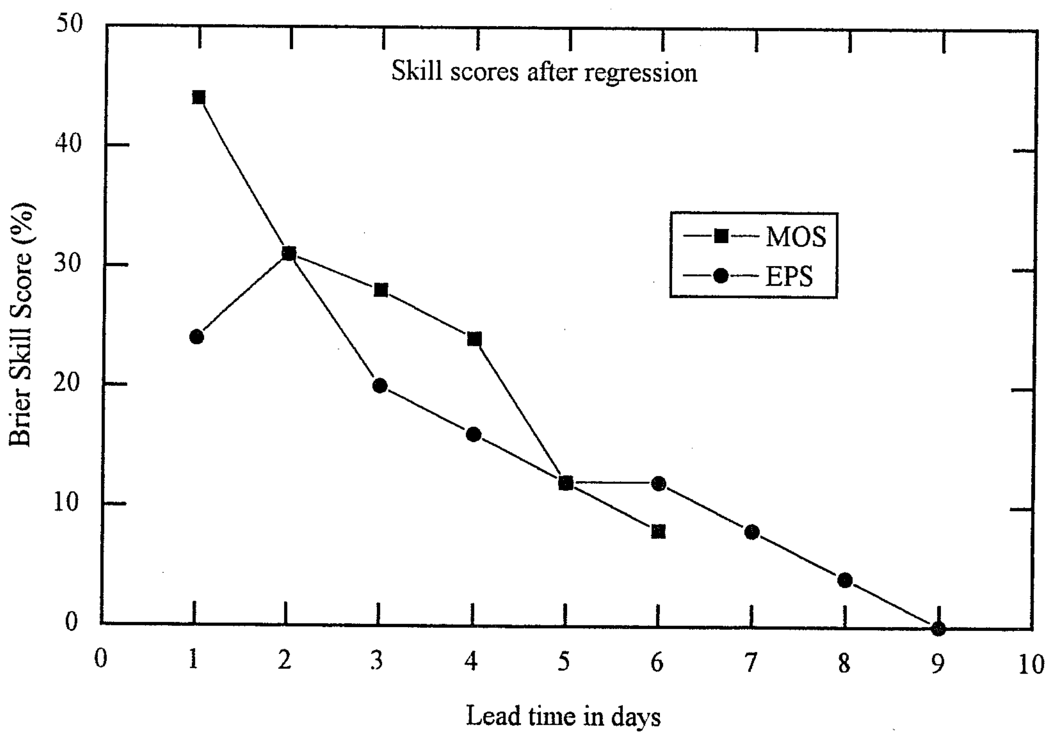

Figure 11: Brier skill scores versus lead time of logistic regression equations using either MOS or EPS probabilities as predictor.