STATISTICAL INTERPRETATION OF ECMWF PRODUCTS IN THE DUTCH WEATHER SERVICE

S. Kruizinga

Royal Netherlands Meteorological Institute
De Bilt, The Netherlands

1. INTRODUCTION

Statistical interpretation of medium range numerical weather prediction (NWP) products started at the Royal Netherlands Meteorological Institute (KNMI) in 1970. At that time the forecasters in the Operational Division used a technique based on manually selected analogues for the interpretation of the 72 hour 500 mbar prognostic charts. These NWP products were received from the National Meteorological Centre (NMC) of the United States of America. In 1975 de Jongh developed a numerical procedure for selecting the analogues and in the subsequent years the whole procedure was transformed to a completely objective scheme for the interpretation of 500 mbar prognoses. When the European Centre for Medium Range Weather Forecasts (ECMWF) began operational forecasting this scheme was rewritten for use on the ECMWF products. In section 2 a description of the analogue selection scheme is given as well as the verification results obtained for ECMWF forecasts.

Inspired by the results obtained by Klein and Glahn in the USA, research on the interpretation with regression techniques was started in 1978; this resulted in guidance being based on NMC-NWP products from 1979. Although only 500 mbar data were used as input in a Perfect Prog system the results were encouraging. As a result some procedures based on European Centre products were developed. In section 3 a description of the current Perfect Prog system is given. It must be emphasized, however, that the description given in this section should not be seen as a static one but just as a description of a rapidly changing state of affairs. The recent growth in the number of products received by us from the Centre will lead to new procedures. Furthermore our growing archive will soon allow us to start research on MOS-procedures. In section 4 some preliminary research on this subject is described and the first results certainly are encouraging.

2. ANALOGUES

The use of analogues as a tool in interpreting the output of numerical models is fairly well known. For instance Wilson and Yacowar (1980) and Woodcock (1980) reported recently about the use of analogues. Also Yacowar (1975) and Agnew and

Alexander (1980) report on the successful use of analogues. Although the principal approach is similar, the selection of analogues can be based on quite different procedures. The procedure described here was designed to simulate a particular manual selection as much as possible, see de Jongh and Kruizinga (1975).

2.1 Analogue selection procedure

The selection of analogues is based on the 500 mbar forecasts of the European Centre for Medium Range Weather Forecasts (ECMWF). The forecast times used are +24, +48 up to +144 (indicated by day 1, 2 etc) valid for 12 GMT. For each forecast chart a set of 30 analogues is selected in the following way. From each field the 500 mbar heights on the 58 gridpoints indicated in figure 1 are used. The historical data-set scanned for analogues consists of 500 mbar heights at the same gridpoints daily at 0 GMT, for the period January 1, 1949 up to December 31, 1979. Before the similarity of two 500 mbar patterns is computed a simple selection based on the date is performed. A possible analogue must be in the same period of the year, a difference of 20 days between the dates is allowed. When this criterion is met the similarity S of patterns is computed from

$$S = \sum_{n=1}^{58} w_n \left((F_n - \overline{F}_n) - (A_n - \overline{A}_n)\right)^2 \qquad (2.1)$$

where F_n and A_n denote gridpoint values of forecast and possible analogue respectively. Furthermore w_n denotes the weights assigned to the gridpoints (see figure 1). These weights are also used in the computation of the averages $\overline{F}_n$ and $\overline{I}$

$$\overline{F}_n = \sum_{n=1}^{58} w_n F_n \Big/ \sum_{n=1}^{58} w_n \qquad (2.2)$$

The similarity measure S becomes zero in case of a perfect match. When scanning through the historical data the dates and the similarities of the 30 most similar patterns are retained. The weather associated with these analogues is used to produce a "guidance" forecast.

2.2 Verification results of analogues

The elements mainly used to express the medium and long range forecast at our institute are minimum- and maximum-temperature (T_n and T_x) and the occurrence of a measurable amount of precipitation (>.3mm) in the periods 06-18 GMT (daytime) and 18-18 GMT-POP12 and POP24; all elements being observed at De Bilt. In this paper temperatures are expressed as deviations from the pentad normal in whole degrees Celsius. Forecasts for maximum- and minimum-temperatures are obtained by averaging the maximum and miminum temperatures of the analogues. These averages are rounded to the nearest degree. The fractions of analogues (in %) with precipitation > .3mm during daytime (POP12) and during the period 18-18 GMT (POP24) respectively are interpreted as the probabilities of precipitation. The verification was performed over a 20 month period from December 1980 up to July 1982 inclusive. Temperatures are verified with the Mean Absolute Error and with a score, in current use at our Institute, - a Performance Index (PI). The PI yields 0 in the case of no-skill. The result related to perfect skill is dependent on the element studied. The Brier score and also the PI were applied to POP12 and POP24. For the sake of clarity the Mean Absolute Error and Brier score were transformed to skill scores with a scale from 0 (no-skill) to 100 (perfect skill), indicated by MSS and BRS respectively. These transformations are discussed in the appendix.

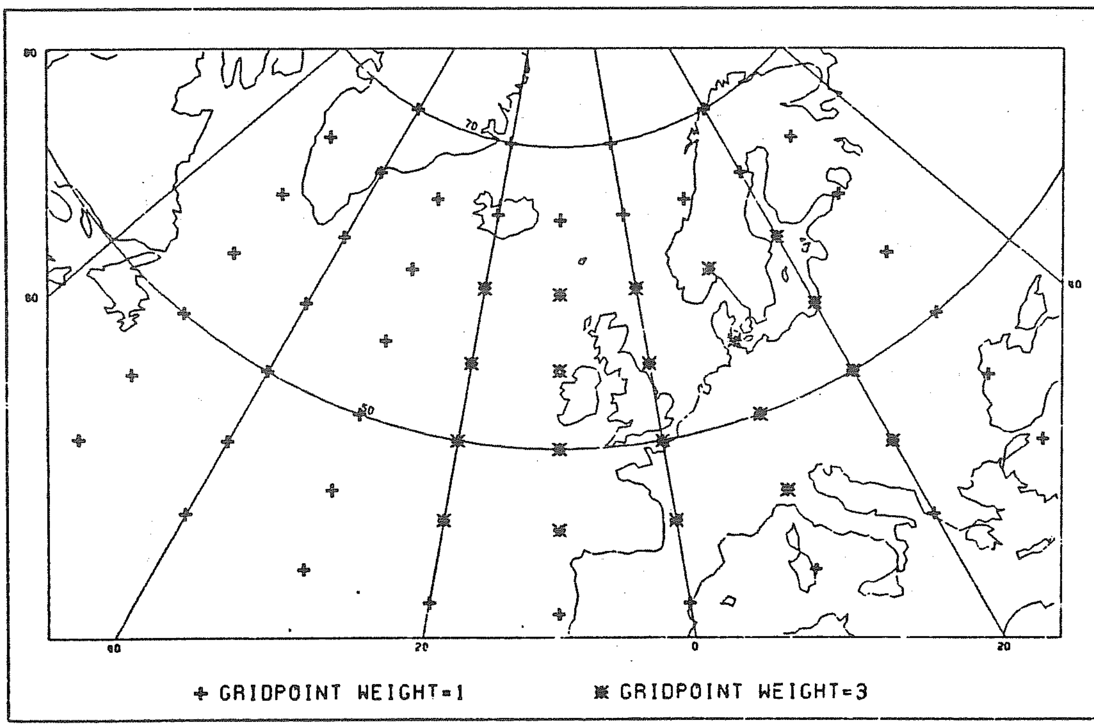

Figure 1: Grid on which the similarity is computed.

The results obtained with skill scores are given in table 2.1. The Performance Indices are given in table 2.2. Both tables lead to the same conclusion: up to day six a positive score is obtained with the analogue interpretation of the 500 mbar ECMWF forecasts. The minimum temperature is clearly the worst element treated in this way. As is seen, the skills are rather stable during the first days but later on mostly decrease with forecast time. This suggests that during the first days the total skill is limited by the quality of the interpretaion scheme, whereas on davs 5 and 6 the quality of the numerical model is the limiting factor.

In table 2.3 the bias of the forecasts is given. As is seen this bias is very small for the temperatures. The probability of precipitation is slightly overestimated.

Table 2.1: Skill scores of T_n and T_x (MSS) and of POP12 and POP24 (BRS) versus forecast time in days

	1	2	3	4	5	6
T_n	10	10	10	9	3	3
T_x	36	32	32	26	19	19
POP12	19	16	17	9	9	4
POP24	29	29	27	22	24	16

Table 2.2: Performance Index (PI) of T_n, T_x, POP12 and POP24.

	1	2	3	4	5	6
T_n	16	15	15	12	11	9
T_x	34	33	29	27	21	20
POP12	18	18	18	14	16	13
POP24	26	26	22	21	23	19

Table 2.3: Bias of T_n and T_x (°C) of POP12 and POP24 (%).

	1	2	3	4	5	6
T_n	.2	.2	.2	.1	.2	.2
T_x	.3	.2	.2	.2	.2	.1
POP12	2	2	3	5	6	6
POP24	1	3	3	4	4	5

3. PERFECT PROG GUIDANCE

The interpretation of NWP-products through the Perfect Prog (PP) approach was introduced by Klein et al (1959). In the PP-system a local variable (predictand) is forecast through a regression equation from predictors obtained from a numerical forecast. The statistical regression equations are developed with a historical data set containing observations of the predictand as well as the predictors. When the equations are applied, the model predictor values are assumed to be perfect. Local climatology has been accounted for in the regression equation.

3.1 Statistical Techniques

The forecast equations for maximum- and minimum-temperature are derived with a standard multiple linear regression scheme:

$$T_x = a_0 + \sum_n a_n \cdot x_n \qquad (3.1)$$

where T_x is the forecast temperature and then the x_n are parameters predicted by the model. The selection of the input parameters x_n (predictors) is performed with a forward stepwise regression procedure. This technique has been discussed frequently in meteorological as well as statistical literature. Dempster (1969) contains a thorough description.

For the elements POP12, POP24 and the probability of thunderstorms in the Netherlands (POT) use is made of the logit model to formulate the forecast equations, Glahn and Bocchieri (1975). The probability of occurrence is related to the input parameters (predictors) in the following way:

$$P = 100/(1 + \exp(f)) \qquad (3.2)$$

with f:

$$f = a_o + \sum_n a_n x_n \qquad (3.3)$$

The values of the coefficients a_0, a_n are estimated with an iterative maximum likelihood procedure. For the statistical background for this procedure we refer to Anderson (1972). The predictors were selected in advance with a forward stepwise regression scheme.

3.2 Predictors

The data set used to derive the forecast equations consists of the daily 12 GMT analyses of 500 and 1000 mbar geopotential height in the period 1-1-1972 up to 31-12-1979. These analyses were obtained from NCAR (USA).

All gridded input fields were transformed to the grid depicted in figure 2. With the grid values on this grid the following predictors related to the height, geostrophic wind and vorticity above De Bilt, were defined:

$$H = H_0 \qquad (3.4)$$

$$VNZ = H_1 - H_3$$

$$VOW = H_2 - H_4$$

$$VORT = (H_1 + H_2 + H_3 + H_4 - 4H_0)$$

with H_n indicating the height of 500 mbar or 1000 mbar at the points indicated in fig.2. These predictors were defined at both levels resulting in 8 possible predictors. For POP12 and POP24 these 8 predictors were used only.

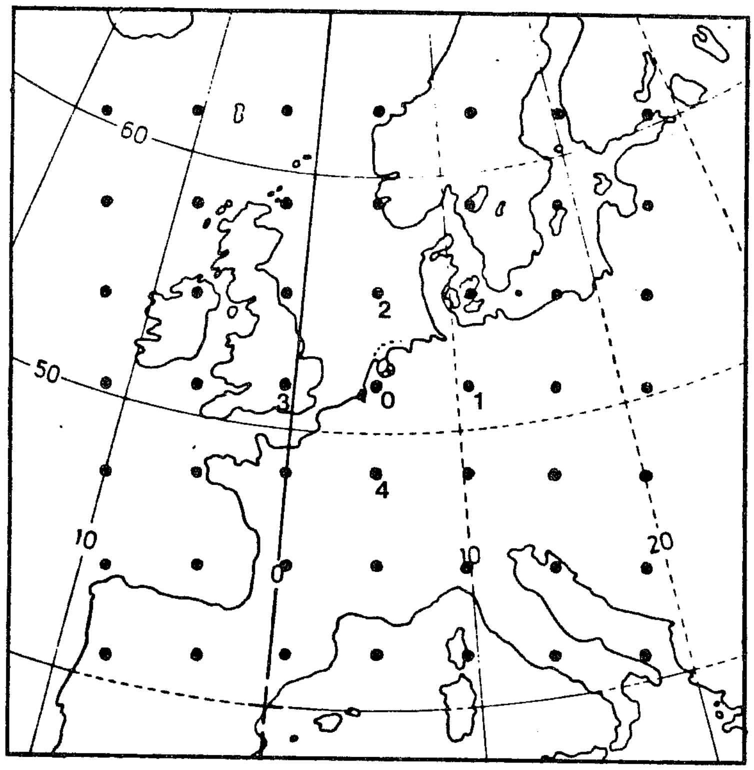

Fig. 2: Grid used with Perfect Prog Guidance. Gridspacing = 400 km.

For T_n and T_x more predictors were available. In order to have some predictors related to the larger scales of atmospheric motion 10 predictors defined on the whole grid of figure 2 were added. These predictors were constructed in the following way:

For each daily grid the central gridpoint height at a given level was subtracted from all the other gridpoint heights. From the resulting data set the Empirical Orthongonal Functions (EOF) were computed for each level separately. The daily scores on the first five EOF's of each level were added to the set of possible predictors.

Specific predictors for T_n and T_x respectively were defined with the help of the geostrophic wind at the 1000 mbar level. For this wind 8 direction classes and a calm class were defined. For each class the climatological mean of T_n and T_x was computed per month. These climatological means indicated by the daily wind were also entered into the regression. Furthermore the thickness of the layer 500-1000 mb was available.

The prediction equation for the predictand, POT, probability of thunderstorms in the Netherlands in the period 0-24 GMT, was developed at a later stage. At that time grids at 850 mbar were available so predictors of this level were also entered. However, firstly one additional special predictor was developed. Hanssen (1965) has shown that the thickness difference ΔD is a good predictor of thunderstorms in the Netherlands. This thickness difference ΔD is defined by

$$\Delta D = (H_{700} - H_{1000}) - (H_{500} - H_{700}) \qquad (3.5)$$

with H_i the height of the i mbar level. With the 1000, 850 and 500 mbar heights we developed through regression a predictor DDH:

$$DDH = -.53\ H_{500} + 2.12\ H_{850} - 1.82\ H_{1000} \qquad (3.6)$$

In the POP and POT equations only 4 predictors were allowed. The temperature equations contained 5 to 10 predictors.

3.3 Verification Results

For the elements T_n, T_x, POP12 and POP24 forecast equations were developed for each of the four seasons. The thunderstorm probability POT is only forecast in the warm season, one equation for the 5 month period May to September inclusive was developed. Skill scores obtained on the dependent data set, about 720 days in each season, are given in table 3.1. Here only skill scores calculated with either mean absolute error or Brier score are given.

Table 3.1: Skill scores for T_n and T_x (MSS) and for POP12, POP24 and POT (BRS) obtained on the dependent data set.

	Winter	Spring	Summer	Fall	Year
T_n	23	40	43	27	33
T_x	32	39	45	45	40
POP12	27	31	40	34	33
POP24	36	40	42	42	40
POT			38		

Comparison of these results with the results given in table 2.1 leads to the conclusion that the regression equations have at least the potential to perform better than the analogues.

The verification results with forecast data were obtained over the same 20 month period as used with the analogues except for the thunderstorm probability. Only 8 months of verification data were available for this element. Table 3.2 gives the skill scores measured with mean absolute error and Brier score respectively.

Table 3.2: Skill scores of T_n and T_x (MSS) and of POP12, POP24 and POT versus forecast time in days.

	1	2	3	4	5	6
T_n	30	27	23	17	13	10
T_x	39	32	32	26	19	16
POP12	36	27	13	0	-7	-17
POP24	42	35	24	10	9	0
POT	35	31	24	6	-1	-21

Table 3.3: PI of T_n, T_x, POP12, POP24 and POT versus forecast time in days.

	1	2	3	4	5	6
T_n	31	28	25	19	17	14
T_x	38	35	31	28	25	22
POP12	25	22	19	16	13	12
POP24	29	26	22	19	20	15
POT	22	19	22	15	15	10

Table 3.4: Bias of T_n and T_x (°C) and of POP12, POP24 and POT (%)

	1	2	3	4	5	6
T_n	.4	.4	.3	.3	.2	.1
T_x	.2	.2	.2	.2	.2	.1
POP12	2	5	5	8	9	9
POP24	0	2	3	4	5	6
POT	8	8	7	8	12	16

The day 1 skills given in this table are near to the maxima given in table 3.1 and much better than the skill obtained with the analogues, table 2.1. However, the skills of POP12, POP24 and POT show a much sharper decline with forecast time than the skill of the analogues. From day 3 onward the analogues tend to be superior.

For temperature forecasts the Perfect Prog guidance is better. Apparently the temperature equations use predictors which are forecast better by the European Centre Model.

In table 3.3 the Performance Indices of the Perfect Prog Guidance are given. The results given here for day 1 and 2 are in accord with the results given in table 3.2. However, the results given for day 4, 5 and 6, especially for POP12, POP24 and POT are in sharp contrast. According to this table these elements show skill even out to day 6, because all PI's given in this table are significantly different from zero (a PI > 4 is significant at a 95% level). The results of table 3.2 and 3.3 combined lead to the conclusion that forecasts and observations are related to each other out to day 6 but there are other reasons than pure randomness of forecasts that damage the skill scores. Partly these poor results can be explained by the bias given in table 3.4. The probabilities of precipitation and thunderstorms are clearly too high.

4. MOS-EXPERIMENTS

The term Model Output Statistics (MOS) was coined by Glahn and Lowry (1972). Since then this approach has been used by several authors. Just as PP the MOS approach is based on statistical regression techniques. However, with MOS, equations are developed with model forecast predictors and observed predictands. The well-known advantages of MOS relative to PP are:

a. Model bias and random errors are accounted for in the regression.

b. More predictors are available. For instance special predictors such as vertical velocities which are only available from models.

c. Recent observations can easily be used as predictors.

Since we are interested mainly in forecast times ranging from 2 to 5 days it is thought that the first advantage will be the most important. A second reason for applying MOS is derived from the requirements for the forecasts. The requirement "accuracy" measured by a score or skill score is widely known. However, "reliability", introduced by Sanders, see Murphy (1973) should be just as important. The meaning of "reliability" is illustrated with the following examples:

1. After a set of forecasts "the probability of precipitation is p%" the event should occur in p% of the cases.

2. After a set of forecasts "maximum temperature 25^{o}C" the average of observed values should be 25^{o}C.

The PP approach will usually fail on this aspect except with short forecast times.

With MOS, forecasts are known to be reliable. In our preliminary MOS study we have performed two experiments.

I Corrected PP (CPP)

PP-POP24 forecasts were offered as sole predictors to a MOS procedure. This results in bias correction and smoothing of the predictors.

II Limited MOS (LMOS)

The same 8 predictors used with the development of the PP-POP24 equations but now with model predicted parameters were offered to a forward stepwise regression scheme. The first four selected were used to develop the seasonal MOS-equations for day 3, 4, 5 and 6.

The available data consisted of approximately 600 runs in the period March 1980 up to April 1982 inclusive. The period December 1980 up to November 1981 inclusive was used to test the equations. It was expected that the amount of data in the dependent-set was too small to develop stable equations. So we took the following precautions:

I CPP

For each forecast time two yearly equations were developed. The first one based on all the data outside the test period. The second one on all available data. The verification results are indicated with CPP1 and CPP2. We expect that the average of both will be a good estimate of the results obtained with stable equations.

II LMOS

First the amount of data was artificially enlarged by using 9 other stations as statistically independent replicas of the Bilt. So with the same values of predictors we generated 10 input records. After that two equations were developed per season and per forecast time in the same way as with CPP. The verification results given under LMOS1 and LMOS2 are based on De Bilt solely.

Both experiments were limited to the forecast times of 3, 4, 5 and 6 days. In table 4.1 the Brier skill scores of the experiments as well as skill scores of analogues and PP over the same period are given.

Table 4.1: Brier skill scores of POP24 obtained with different methods.
Verification period December 1980 upto November 1981 inclusive.

	FORECAST TIME IN DAYS			
	3	4	5	6
Analogues	27	22	25	18
PP-Guidance	22	3	8	6
CPP1	24	10	9	8
CPP2	25	10	14	12
LMOS1	21	11	16	10
LMOS2	27	22	23	21

This table clearly demonstrates that MOS and correction of PP can lead to improved forecasts. However, the analogues, (a PP method), still out-perform the other methods. It may be expected that MOS with more predictors available could be superior to an analogue approach.

With the data for the same verification period the reliability has also been studied. However, to study reliability a large number of forecasts is required, so we combined the data for day 5 and 6 (as well as day 3 and 4). Thereafter the set of forecasts was grouped into subsets. In our case we defined 10 classes, the first class contained all forecasts of POP24 between 0 and 10%, the second class all forecasts from 10 to 20% etc. The expected frequency of the occurrence of precipitation in each class is 5%, 15%, 25% etc. In figure 3 the observed frequency of occurrence in each class is plotted (+) against these expected frequencies. The number of forecast in each class is also indicated. For the ideal reliability the plots should "fall on" the straight dotted line.

From figure 3 we see that the analogues behave very well whereas PP is the most unreliable of the four sets shown. The reliability can be expressed as a number by computing the weighted average of the squared vertical distances from the plots to the ideal dotted line. The weights used are the number of cases in each class. These computations lead to the results given in table 4.2., these results are computed with vertical distances expressed in percent. Just as with the Brier score a small figure indicates a higher reliability.

Perfect reliability results in 0. However, due to sampling effects this limit will not be reached even in the case of perfect reliability. It can be shown that with the number of cases used here the lower limit will be about 24 so the analogues are very near to perfect. Furthermore, it can be seen that MOS as well as correction of PP results in an improved reliability.

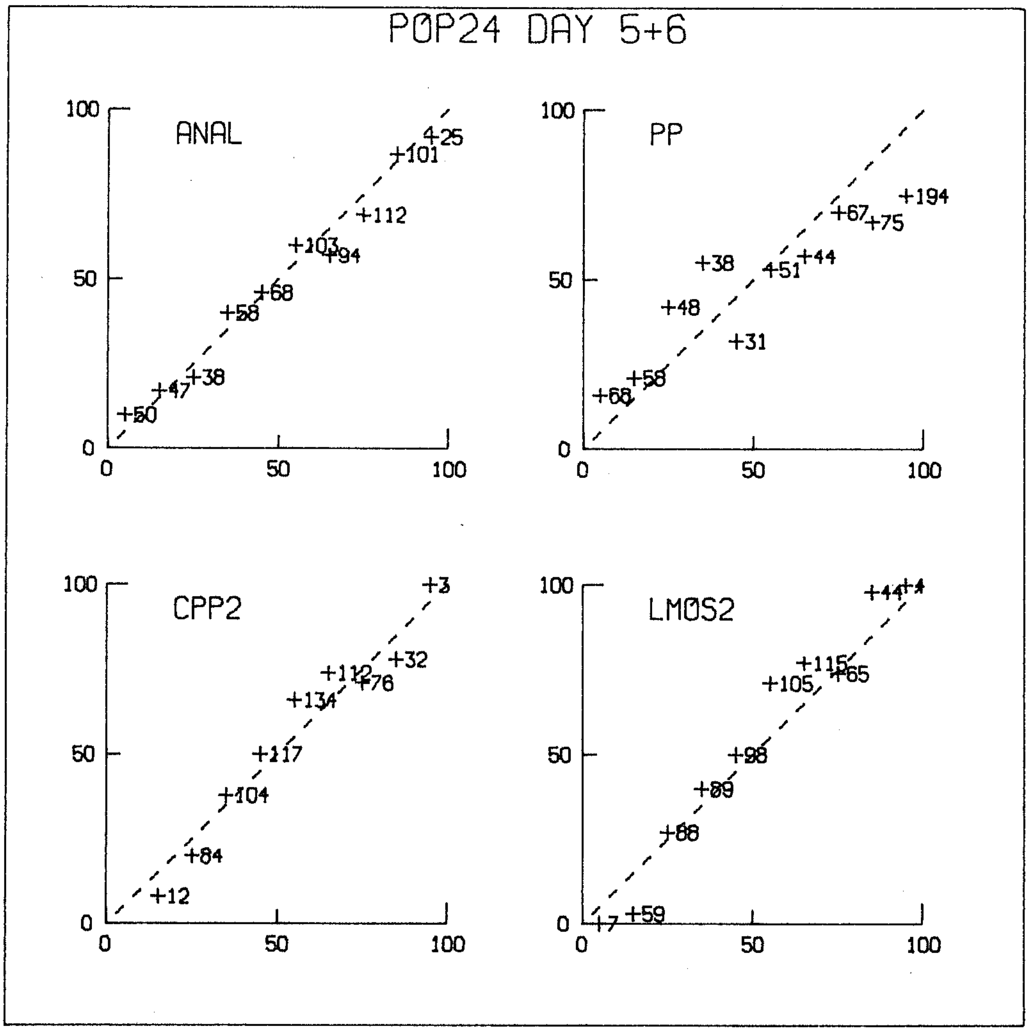

Figure 3: Reliability diagrams (see text)

Table 4.2: Reliabilities

	3 + 4	5 + 6
ANAL	14	24
PP	160	224
CPP1	70	104
CPP2	45	51
LMOS1	120	106
LMOS2	47	96

5. CONCLUDING REMARKS

As from June 1, 1981, the Operational Division of the KNMI has been issuing 5-day forecasts to the general public. Analogues and PP-guidance as well as directly plotted output from the ECMWF are used by the forecaster to make up his forecast. The verification results given in this paper are based solely on the output of the schemes. So the scores given here are not representative for the scores obtained by the Operational Division. The differences between the Performance Indices, however, are generally small, except for day 1. The skill scores can differ substantially. For instance the skill score of POP24 of the PP-guidance is clearly lower than the skill scores of the analogues and the forecaster. For the temperatures the scores of PP are better than the analogue scores, while the forecaster is able to improve on PP by about 5 points. For POP24 and POP12 analogues have better scores on day 3, 4 and 5. With POP24 forecaster and analogues score almost equal to POP12, the forecaster scoring higher than the analogues. It is important to note that due to the inevitable delay between model initial time and forecast issuing time the day 5 forecast must be based on day 6 model output. So when comparing verification results of forecasters and objective schemes a shift of one day must be applied.

The MOS-experiment described in section 4 has no operational status at the moment. It is generally felt, however, that it is undesirable to use more than one interpretation scheme. So we plan to replace both schemes by one MOS-based scheme. In section 4 it has been shown, however, that analogues can compete with a MOS scheme so we plan to use analogue output as a predictor to a MOS scheme in order to retain the benefits of both schemes. Additional research is planned with regard to the elements of sunshine, as well as to the amount and phase of precipitation.

Appendix : Verification Scores

Mean Absolute Error

This score is used frequently for the verification of point forecasts for elements on a continuous scale e.g. temperature, sunshine duration. If f_n indicates the successive forecasts and o_n the related observations then MAE is defined by

$$MAE = \frac{1}{N} \sum_{n=1}^{N} ABS(f_n - o_n). \tag{A1}$$

(N = number of forecasts)

Clearly the optimum forecast will yield MAE = 0. The no-skill-level is chosen equal to the MAE of climatology MAE_{CLIM}. The transformation to the skill-score MSS is obtained by

$$MSS = 100 \frac{MAE_{CLIM} - MAE}{MAE_{CLIM}} \tag{A2}$$

In the preceding paper the skill score is used for maximum- and minimum temperature. The climatological values of MAE for these elements at the Bilt are 3.1 C^o and 3.0 C^o respectively.

Brier Score (BS)

This score which was developed by Brier (1950), is applied to the verification of probabilistic forecasts. In our case it is used for the verification of the forecast probability of a measureable amount of precipitation (>.3mm) in 12 or 24 hour periods and the occurrence of thunderstorms in the Netherlands respectively (POP12, POP24 and POT). When applied to yes/no variables the score is equal to twice the mean square error. If the observations o_n are recorded as a row of 1's and 0's for yes and no respectively and the forecast probabilities are denoted by p_n $(0 < p_n < 1)$ then

$$BS = \frac{2}{N} \sum_{n=1}^{N} (o_n - p_n)^2 \tag{A3}$$

Again perfect forecasts lead to BS = 0. The no-skill reference point is usually chosen equal to the score obtained by climatology in the long run BS_{CLIM}. Transformation to the skill score BRS is obtained by

$$BRS = 100 \cdot \frac{BS_{CLIM} - BS}{BS_{CLIM}} \tag{A4}$$

the climatological values of the Brier-score for respectively 12 hour and 24 hour precipitation at De Bilt and the occurrence of thunderstorm in the Netherlands are

.43, .49 and .45.

Performance Index (PI)

The verification score was developed by Kuipers about 1958 but has recently been described in Kuipers (1980). This score can be used with continuous variables as well as with categorical variables. However, this score requires an adapted way of formulating the forecast. For continuous variables the forecast must be an interval and for categorical variables one or more classes must be chosen as forecast. The Performance Index (PI) of an individual forecast is given by the expression

$$PI = \begin{Bmatrix} 100 \text{ in case of hit} \\ 0 \text{ otherwise} \end{Bmatrix} - p_c \qquad (A5)$$

where a hit means an observation within the interval forecast or occurrence of one of the classes forecast and p_c is the climatological probability (%) that a hit will occur. The PI of a set of forecasts is simply the average of the individual PI's. For continuous variables the PI has a scale running from 0 (no-skill) to 100 (perfect skill). With categorical forecasts again 0 indicates no skill, however, in general, a result of 100 will not be reached even with perfect forecasts. The maximum that can be reached is defined by climatology. In this paper we will apply this score to temperatures which are given as point forecast and expressed as deviation ΔT from pentad normal. So we need a rule to transform the point forecast to an interval forecast. This is done with the help of table A1, which defines forecast intervals dependent on the point forecast. The fact that the intervals become one-sided at the extremes is related to the low climatological probabilities for the extremes.

The other elements to which the PI is applied are the occurrence of precipitation and of thunderstorms. The probability forecasts we have are translated to a categorical forecast by forecasting the occurrence if the forecast probability is higher than the climatological probability. The maximum PI's obtainable are 43 for POP12, 49 for POP24 and 46 for POT.

Table A.1: Transformation of a point forecast to an interval used with maximum and minimum temperature.

ΔT	INTERVAL	ΔT	INTERVAL	ΔT	INTERVAL
-7	$T<-5$	-2	$-5<T<0$	3	$1<T$
-6	$T<-4$	-1	$-3<T<1$	4	$2<T$
-5	$T<-3$	0	$-2<T<2$	5	$3<T$
-4	$T<-2$	1	$-1<T<3$	6	$4<T$
-3	$T<-1$	2	$0<T<5$	7	$5<T$

REFERENCES

Agnew,T. and Alexander, J. (1980) An Operational Evaluation of probability of precipitation amount forecasts generated from statistical and analogue techniques. Preprints 8th Conference on Weather Forecasting and Analysis, Amer. Meteor. Soc., Denver, Colo.

Anderson, J.A. (1972) Separate Sample Logistic Discrimination. Biometrika, 59, 19-35.

Brier, G.W. (1950) Verification of Forecasts Expressed in terms of Probability, Monthly Weather Review, 78, 1-3.

Dempster, A.P. (1969) Elements of Continous Multivariate Analysis, Addison-Wesley.

Glahn, H.R. and Lowry, D.A. (1972) The use of model output statistics in objective weather forecasting. J. App. Meteor., 11, 1203-1211.

Glahn, H.R. and Bocchieri, J.R. (1975) Objective estimation of the Conditional Probability on Frozen Precipitation. Monthly Weather Review 103, 3-15.

Hanssen, A.W. (1965) An objective method for forecasting thunderstorms in the Netherlands. Journ. Appl. Meteor., 4, 172-177.

De Jongh, J.P. and Kruizinga, S. (1975) Empirisch orthogonale eigen velden. KNMI, de Bilt, WR 75-10 (in Dutch).

Klein, W.H., Lewis, B.M. and Enger, I. (1959) Objective prediction of 5-day mean temperatures during winter. J. of Meteor., 16, 672-682.

Kuipers, W.J.A. (1980) 25 years of objective evaluation of weather forecasts in the Netherlands. WMO Symposium on Probabilistic and Statistical Methods in Weather Forecasting, Nice, 1980.

Murphy, A.H. (1973). A new vector partition of the Probability Score. J. Appl. Meteor, 12, 595-600.

Wilson, L.W. and Yacowar, N. (1980) Statistical Weather Element Forecasting in the Canadian Weather Service. WMO Symposium on Probabilistic and

Statistical Methods in Weather Forecasting, Nice, 1980.

Woodcock, F. (1980) On the use of analogues to improve regression forecasts. Mon. Wea.Rev., 108, 292-297.

Yacowar, N. (1975) Probability forecasts using finely tuned analogs. Fourth Conf. on Probability and Statistics in Atmos. Sci., Tallahassee, Florida, Preprints 49-50.