KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT

TECHNISCHE RAPPORTEN

T.R. - 54

C. Lemcke en S. Kruizinga

Development of guidance forecasts at KNMI

1. General information

De Bilt, 1984

Publikatienummer: K.N.M.I. T.R. 54 (DM)

Koninklijk Nederlands Meteorologisch Instituut,
Dynamische Meteorologie,
Postbus 201,
3730 AE De Bilt,
Nederland.

U.D.C.: 551.509.313

Contents

1. Introduction

The introduction of numerical weather prediction (NWP) models led to a demand for objective interpretation or translation of these synoptic scale forecasts to local weather forecasts. In the U.S.A. the development of such objective interpretation schemes started in the sixties and remarkable results were obtained since then by Klein, Glahn and others using statistical schemes. At KNMI the development of such schemes started in the mid-seventies. Initially these interpretation schemes were applied to the output of the National Meteorological Center (NMC-USA) model; later on the output of the European Centre on Medium Range Weather Forecasts (ECMWF) model was used. In general these schemes focused on local weather forecasting for lead times ranging from 48 hours upwards.

In this report a general introduction to this interpretation work is given. The evolution of these methods at KNMI since about 1970 is described. Furthermore some data sets and commonly used predictors are described. Results of research on specific weather elements will be published separately.

Finally we will discuss some scoring rules which will be frequently used in our studies.

2. Objective weather forecasting and guidance forecasts

In the pre-computer age almost all weather forecasts were based on forecasters' experience and their subjective judgement. Objective methods played a minor role in the forecasting procedure. By "objective methods" we mean procedures as defined by Allen and Vernon (1951):
"An objective system is one which can produce one and only one forecast from a specific set of data".
Glahn (1965) also used this definition but added:
"... these restrictions are sometimes relaxed slightly or some subjectivity creeps into the definition of a specific set of data".
For instance when some input variables (predictors) to an objective scheme are estimated by the forecaster the scheme will still be named objective, even though the results will depend on the subjective input of the forecaster.
The foregoing definitions include numerical modelling as well as statistical methods. In this report we will limit ourselves to statistical methods which are applied to the output of numerical models. We will indicate such a method as an "Objective Interpretation Scheme" which is subsequently defined as:
"A procedure which computes from the output of numerical weather model (and recent observations) a numerical forecast value of a given weather element at a given location or area. This forecast value can be either a category or a point-value or even the probability that a given category or weather event will occur".
From this definition we derive the definition of the guidance forecast (Dutch: gidsverwachting):
"The presentation of the results of the various objective interpretation schemes ordered according to the forecast time".

3. PP versus MOS

These two acronyms indicate the two most often used approaches to the interpretation problem. Most interpretation schemes are based on statistical models relating the predictors (produced by the numerical model) and the predictands. Unlike physical models, which are strongly dependent on the variables studied, statistical models are very general and almost independent of the nature of the variables. The adaptation to the specific problem is performed by estimating the coefficients of the model. This estimation is based on a rather large set of observations of the predictands and the concurrent predictors, the dependent data set. After the estimation of the coefficients the model can be used for the translation from predictors to predictands. There are, however, two possible ways of assembling a dependent data set. The oldest method (Klein, 1959) is called Perfect Prog (PP), the most recent method, developed by Glahn and Lowry (1972), is called Model Output Statistics (MOS).

MOS: In this approach a series of model-forecast predictors with concurrent predictands is used as dependent data set. Generally also time-lagged (local) observations are added to the predictors but even in that case the method is still called MOS.

PP: Most of the model-forecast predictors are also (indirectly) observable and in this approach the dependent data set consists of observed predictors and concurrent predictands. Subsequently it is assumed that the numerical model forecasts the predictors perfectly. Adding time lagged observations is uncommon.

These different approaches are illustrated in Figure 1.

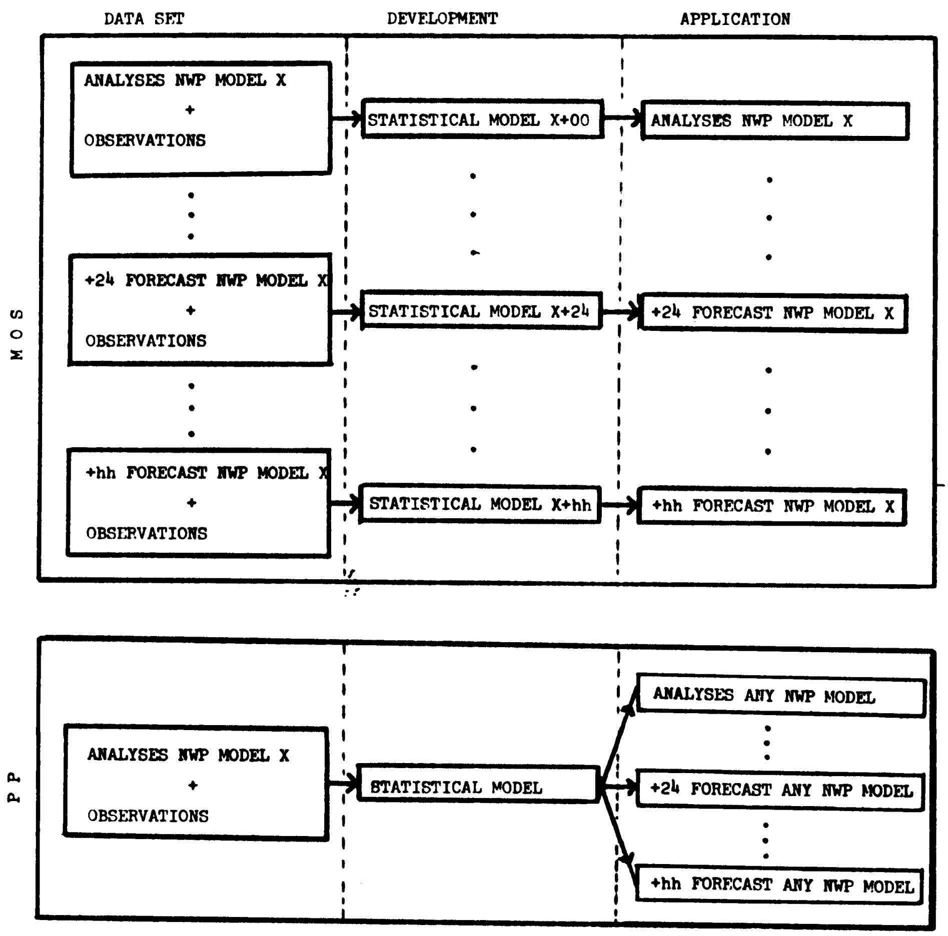

Figure 1

The use of the MOS-approach seems more logical but the PP-approach has some advantages which makes it a worthwhile alternative. An extensive comparison of both approaches lies outside the scope of this report. We will only present some comparative characteristics of both methods:
Simplicity: PP-developed schemes can be applied to any numerical model and any lead-time. MOS-schemes are specific for model (version) and lead time.
Availability of data: PP-schemes can be developed on large historical data sets. For MOS large data sets are scarce and due to model changes satisfactory data sets will probably never be available.
Availability of predictors: With MOS generally more and better predictors are available than with PP.
Skill and reliability: MOS-forecasts are more skillful than PP-forecasts except at very short ranges. Also the MOS approach generally warrants a better reliability (Sanders, 1958; Murphy, 1973) of the forecasts.

4. History of objective methods at KNMI

Statistical interpretation of NWP-products started at KNMI in 1970. At that time the forecasters in the Operational Division used a technique based on manually selected historical 500 mb maps which looked more or less similar to the current 72 hour forecast map of NMC (U.S.A.). These historical maps were called "analogues". The observed weather of these analogues served as a guidance. In 1975 De Jongh and Kruizinga (1975) developed a numerical procedure for selecting the analogues and in the subsequent years the whole procedure was transformed into a completely objective scheme for the interpretation of 500 mbar prognoses. In 1980 the dissemination of NWP-products from ECMWF to the Netherlands started and the analogues scheme was rewritten. More details will be given in section 4.

Inspired by the results obtained by Klein (1959) and Glahn (1972) in the U.S.A., research on the interpretation with regression techniques was started at KNMI in 1978 by Kruizinga and Hofstee. This resulted in a guidance forecast based on NMC (U.S.A.) products in 1979. With Perfect Prog interpretation a forecast could be made for maximum and minimum temperature, probability of precipitation, and probability of thunderstorms. Although only 500 mbar data were used as input the results were encouraging. This resulted in the development of a Perfect Prog guidance by Kruizinga and Lemcke (1981) with ECMWF 1000 and 500 mbar data as input in 1981. The forecast period was extended up to +144 hours. The ECMWF products, based on 12Z observations, are received during the night. Subsequently the forecasts up to +144 hour inclusive are used to issue a five day forecast next morning (see Fig. 2).

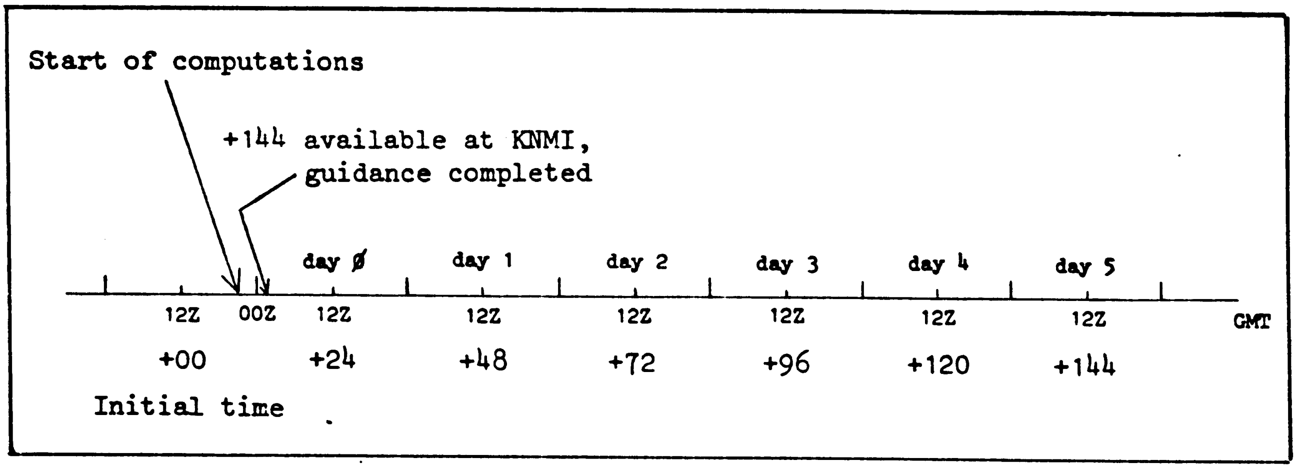

Figure 2 Time scheme ECMWF forecasts. The model run starts at plus minus 20.30 GMT, based on 1200 GMT observations. The +144 forecasts are available at KNMI at about 0100 GMT. Shortly thereafter the objective forecasts are available to the forecaster.

In 1982 two years of ECMWF forecasts had been archived and the first experiments with Model Output Statistics started. This resulted in the introduction of the third generation of the operational guidance forecast at KNMI in November 1983 (Kruizinga and Lemcke). In this guidance forecast also output from the analogues method served as input for the objective interpretation scheme. The following elements are forecast: percentage of sunshine, probability of sunshine for four classes, probability of precipitation, minimum and maximum temperature, the daily extreme wind speed at IJmuiden (coastal station) and the probability of three wind speed classes. The conditional probability of frozen precipitation and the probability of thunderstorms are planned to be included.
Also in 1983 Lemcke and Brouwer introduced a PP-guidance forecast of the maximum temperature and the probability of precipitation during daytime

for about 100 locations in Europe. This guidance is used operationally for the preparation of a forecast for Dutch holiday-makers in Europe. This "European guidance" is an example of a guidance for special users. In the future probably there will be an increasing interest for guidance forecasts for special users.

5. Statistical Models

As said before the interpretation schemes are mostly based on statistical models. Currently three statistical models or methods are in use at KNMI. The linear multiple regression model for point forecasts, the logistic or logit model for probability forecasts and the analogues method for either point forecasts or probability forecasts. The first two methods can be applied with PP as well as MOS-approach. The analogues are limited to the PP-approach due to the large data-base that is required.

5.1 Analogues

The general idea behind the analogues method is very simple: pick up your forecast map(s) and compare them with the map(s) in the historical data base. The weather associated with the maps which are fairly similar to the forecast map is used as a guidance to forecast the weather. In this sense the analogues method was used, for instance, by Yacowar (1975), Balzer (1976), Wilson and Yacowar (1980) and Woodcock (1980). However large differences exist in the choice of maps to be compared with and the judgement of the similarity of the maps. The analogue selection procedure used at the KNMI (and described hereafter) was initially developed to imitate the manual analogue selection procedure.

The analogue selection procedure is based on the 500 mb pattern. The forecast maps are compared with the historical maps from the period 1-1-49 up to 31-12-79 inclusive. For this comparison the 58 gridpoints depicted in Figure 3 are used. Before any comparison is made however a preselection is performed. Only maps from the same period of the year as the validation date of the forecast map are allowed as potential analogues. Specifically a difference of 20 days is allowed between the two dates. For the potential analogues the similarities S are computed. This similarity is defined by

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

with F_n and A_n the gridpoint values of forecast and analogue respectively. The spatial average $\overline{F}_n$ is defined by

$$F_n = \sum_{n=1}^{58} \omega_n F_n / \left(\sum_{n=1}^{58} \omega_n\right)$$

and $\overline{A}_n$ is defined in a similar way. The weights ω_n used in both formulas are indicated in Figure 3.

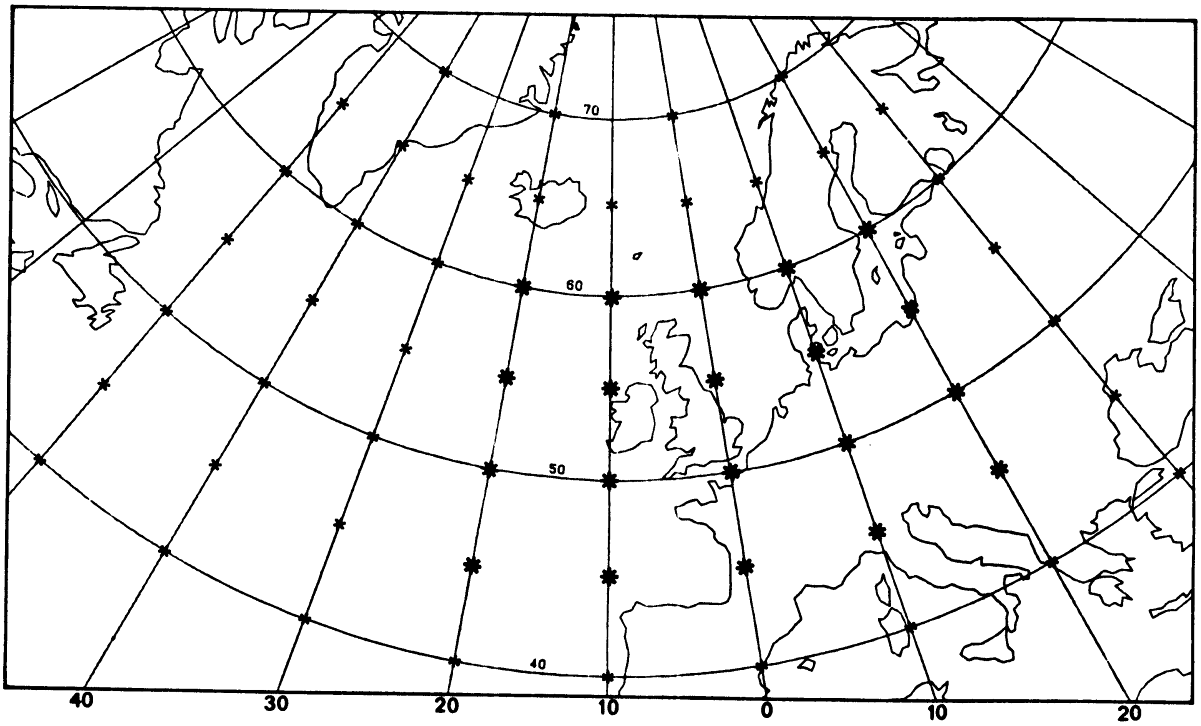

Figure 3 Grid on which the similarity is computed.
* GRID POINT WEIGHT = 1 ✱ GRID POINT WEIGHT = 3

When scanning the historical data set the dates and similarities of the best 30 analogues are retained. The weather observations corresponding to the selected analogues are used to make up a forecast. In which way these observations are used depends on the specific element studied.

5.2 Multiple Linear Regression

With point forecasts it is often possible to use a simple linear equation to compute a forecast value y from variable values x_n extracted from the numerical forecast. This linear equation reads

$$y = a_o + a_1 x_1 + a_2 x_2 + \ldots$$

In this equation the coefficients a_n are estimated with statistical methods from a dependent data set. Often it is also unknown which predictors x and how many predictors should be used. In that case the predictors are selected by a stepwise regression technique which selects in any step the best next predictor. Multiple linear regression and stepwise selection are thoroughly dealt with in Dempster (1969).

5.3 Logistic Discrimination

For probabilistic forecasts mostly the logit-model (Brelsford and Jones, 1967) or logistic model is used. This model relates the probability of an event with the predictors. Examples of such events are: the occurrence of precipitation; observing a continuous predictand within a prescribed interval and so on. When the problem is a yes/no problem e.g. rain/no rain the model equation is fairly simple:

$$P\ (yes) = 100/(1 + \exp(fx))\ (\%)$$

with $fx = a_o + a_1x_1 + a_2x_2 + \ldots$
and x_1, x_2 ... are the predictors and a_o, a_1 ... are the coefficients. In the case that the probabilities of more than two mutually exclusive events have to be forecast the extended model (Anderson, 1972) can be used. In the case of K classed the model reads:

$$P(k=1) = \exp(fk)\ .\ P(k=K)$$

$$P(k=K) = 100/(1 + \sum_{k=1}^{K-1} \exp(fk))$$

with $fk = a_{o,k} + a_{1,k} \cdot x_1 + a_{2,k} \cdot x_2 \ldots$
($P(k=1) \equiv$ the probability that class 1 will occur).

The coefficients $a_{n,k}$ are estimated on the basis of the historical (dependent) data set. However, the estimation of the coefficients is performed with a numerical iterative procedure. This hampers the stepwise inclusion of new predictors. Therefore we normally select the predictors beforehand with stepwise multiple regression.

6. Data sets

As said before historical or dependent data sets are essential for the estimation of the coefficients in a statistical model. Weather observations are available at KNMI over a long period. The data sets with analyzed maps, needed for PP-equations, are mostly obtained from other weather offices. The data set with forecast maps, needed for MOS-equations is currently being assembled from ECMWF model output.

6.1 data set DWD

period	: January 1, 1949 - December 31, 1979
analyses	: 00Z
forecasts	: -
grid	: geographical, see Figure 3
parameters	: height 500 mbar
used for	: analogues method first version guidance (PP)
obtained from	: Deutsche Wetterdienst.

6.2 data set NCAR

period	: January 1, 1972 - December 31, 1979
analyses	: 00Z, 12Z
forecasts	: -
grid	: stereographic, see Figure 4 (interpolated from the original NMC grid)
parameters	: heights 1000, 850 and 500 mbar temperature 850 and 500 mbar
used for	: second version guidance forecast (PP)
obtained from	: NCAR

6.3 data set ECMWF

period	: December 31, 1980 - present

analyses	: 00Z, 12Z	
forecasts	: +12, +24, +36,....., +120, +132, +144	
grid	: stereographic, see Figure 4	
	(interpolated from ECMWF gridsystems 14, 17, 86)	
parameters	: heights 1000, 850 and 500 mbar temperature 850 mbar vertical velocity 500 mbar wind at 10 m, wind 1000 mbar	up to +144
	heights 700, 250 mbar temperature 700, 500 mbar mixing ratio 850, 700 and 500 mbar relative humidity 850, 700 and 500 mbar cloudiness precipitation	up to +72
	This archive can easily be expanded with other parameters and time steps, using the archives at ECMWF	
used for	: third version guidance forecast (MOS)	

The grid used for the NCAR and ECMWF data sets has been defined especially for the development of the interpretation methods. It is a stereographic grid of 7x7 gridpoints with De Bilt exactly in the centre and an orientation North-South at De Bilt.

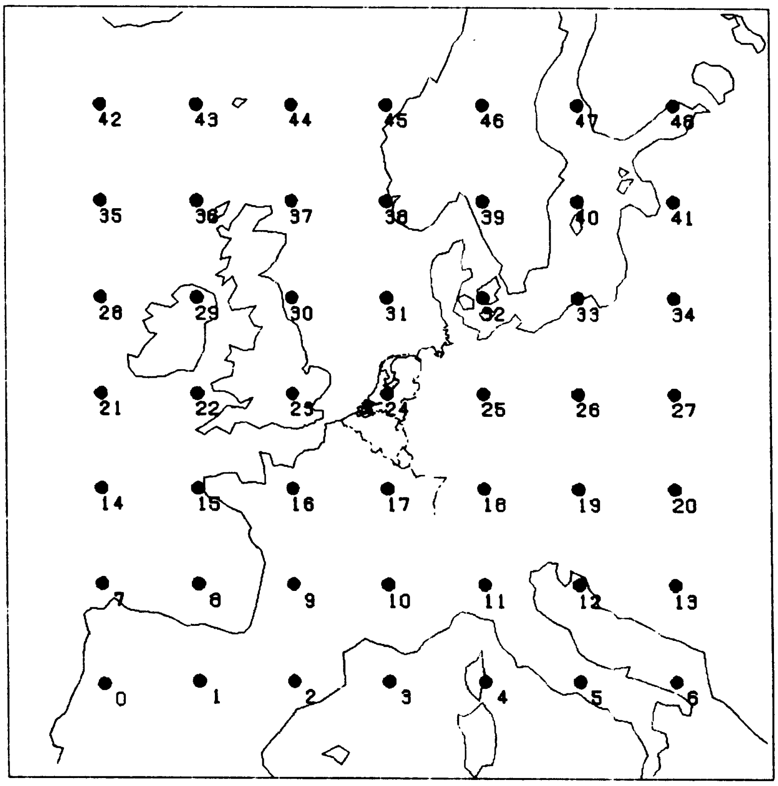

Figure 4
Stereographic grid,
grid spacing 400 km.

7. Frequently used predictors

In this section we will pay some attention to the predictors used for the development of the second and third version of the guidance forecast. For a number of predictors the value at gridpoint De Bilt can be used, for instance the temperature at 850 mbar, the thickness 1000/850 mbar etc. Sometimes not the value itself will be used, but the departure from climatology. In 7.2 more information about climatology will be presented.
Other predictors are computed from several grid points, such as the geostrophic wind components, the relative vorticity and the relative vorticity advection. In 7.1 the formulae for these derived quantities can be found. All these predictors represent rather small scales. It is well known that the quality of the forecast of small scale phenomena will decrease with time even though there may still be valuable information in the large scale. A method to extract this large scale information from the fields is to expand the fields in their principle components (Glahn, 1982) and to use the coefficients of only those components that represent the large scales. In the second and third version of the guidance forecast these coefficients are used. The idea of the principle components will be illustrated with some examples in 7.1.

7.1 Formulae of used predictors

The indices used in the formulae refer to Figure 4.
The geostrophic wind components are:

$$u = -\frac{g}{f}\frac{\partial Z}{\partial y} \qquad (1)$$

and

$$v = \frac{g}{f}\frac{\partial Z}{\partial x} \qquad (2)$$

For grid point De Bilt these equations can be expressed in finite differences as:

$$u_{24} = -\frac{g}{2f.\Delta y}(Z_{25} - Z_{23}) \qquad (3)$$

and

$$v_{24} = -\frac{g}{2f.\Delta x}(Z_{31} - Z_{17}) \qquad (4)$$

$f = 2\Omega \sin\phi$
$\phi = 52.1^{o}$
$\Omega = 7.292\ 10^{-5}$ rad s^{-1}
$g = 9.81$ m s^{-2}
$\Delta x = \Delta y = 4\ 10^{5}$ m

substitution of the values of g, f,Δx respectively Δy results in

$$u_{24} = 0.106\ (Z_{25} - Z_{23})\ (m/s) \tag{5}$$

and

$$v_{24} = 0.106\ (Z_{17} - Z_{31})\ (m/s) \tag{6}$$

The formula for the relative vorticity is:

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{7}$$

In finite differences:

$$\zeta_{24} = 5.3\ 10^{-7}(Z_{25} + Z_{23} + Z_{17} + Z_{31} - 4Z_{24})\quad (s^{-1}) \tag{8}$$

For the relative vorticity advection we have:

$$RVA = -\overline{V}.\overline{\nabla}\zeta = -u\frac{\partial\zeta}{\partial x} - v\frac{\partial\zeta}{\partial y} \tag{9}$$

$$RVA_{24} = -1.25\ 10^{-6}\ ((\zeta_{31} - \zeta_{17}).v_{24} + (\zeta_{25} - \zeta_{23})\ u_{24}))\ (s^{-2}) \tag{10}$$

principal components

For the 1000, 850 and 500 mbar heights principal components are derived based on the NCAR data set. The height of gridpoint De Bilt is subtracted from each field, resulting in a difference pattern from the De Bilt value, and then the mean value for each gridpoint is computed over the 8 years. This mean is subtracted from each pattern. On these resulting patterns the principal components analysis has been applied. We can now approximate each field as the sum of the mean flow and a finite expansion of principal components and the De Bilt value. The height Z at gridpoint i is now:

$$Z_i = \langle Z_i \rangle + Z_{24} + \sum_{n=1}^{N} C_n\ P_i^{w} \tag{11}$$

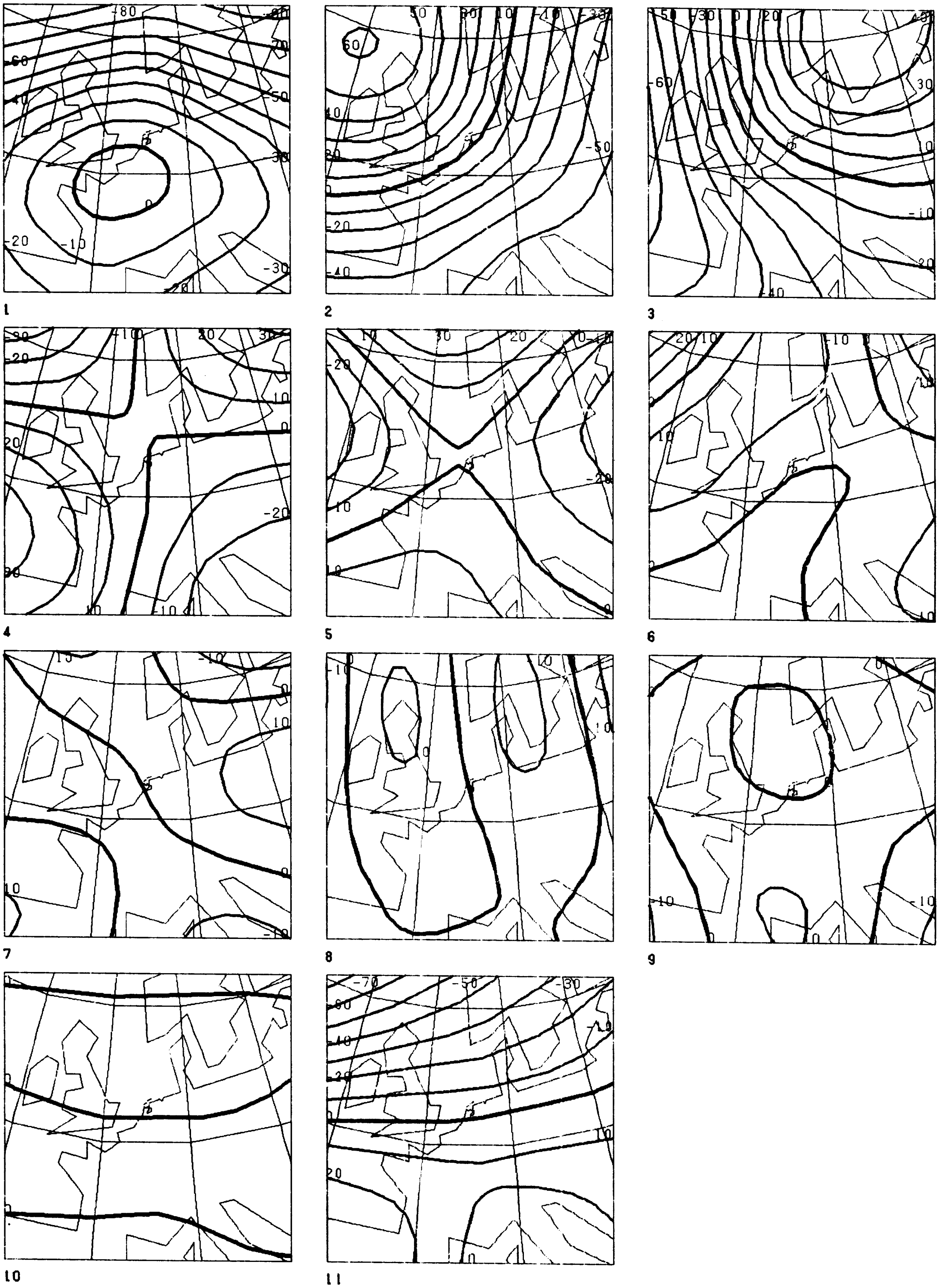

Fig. 5   The first 10 principal components of 1000 mbar height, and the height with respect to height at gridpoint De Bilt, averaged over 1972-1979.

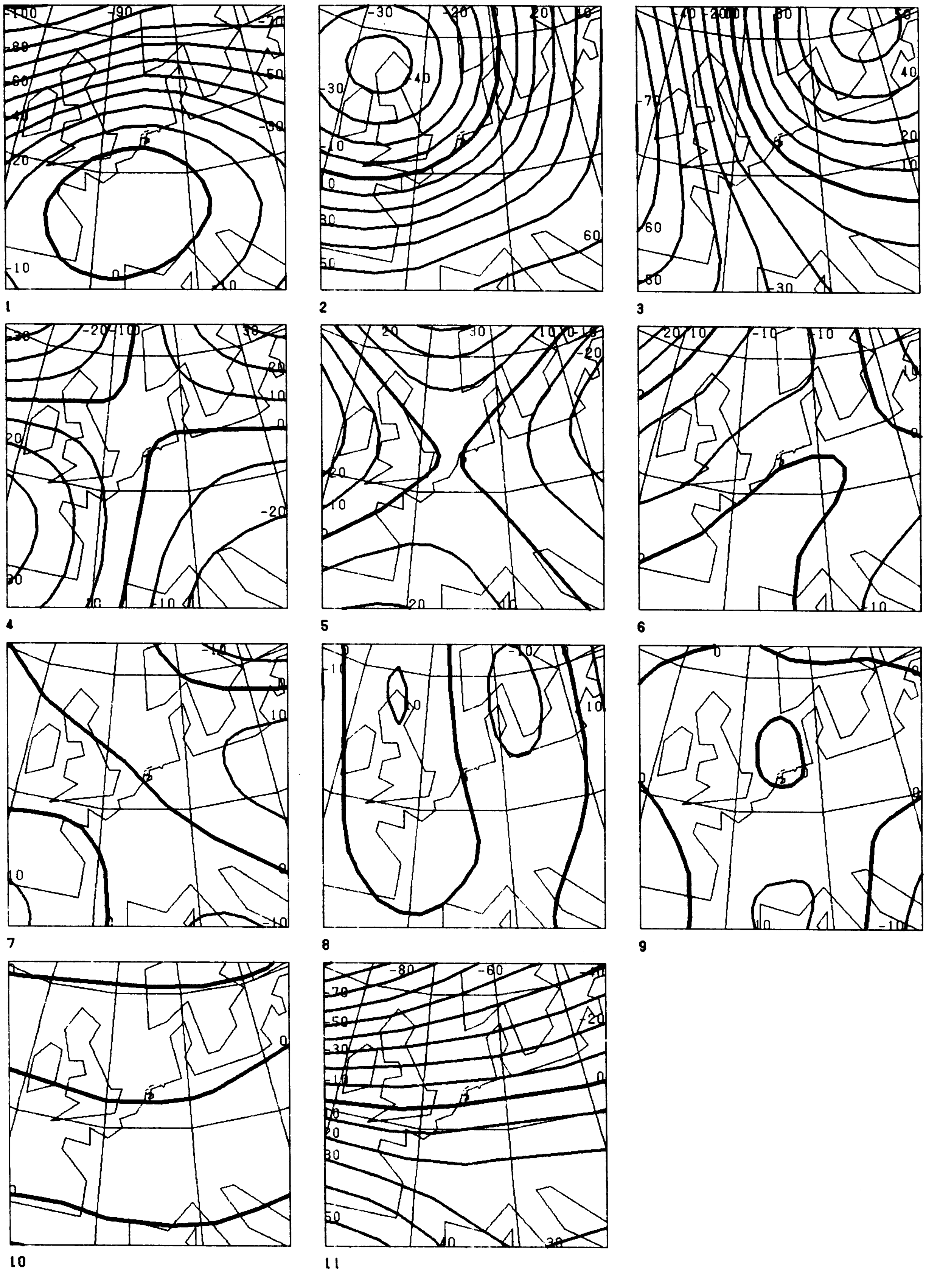

Fig. 6   The same as Fig. 5, however for 850 mbar height

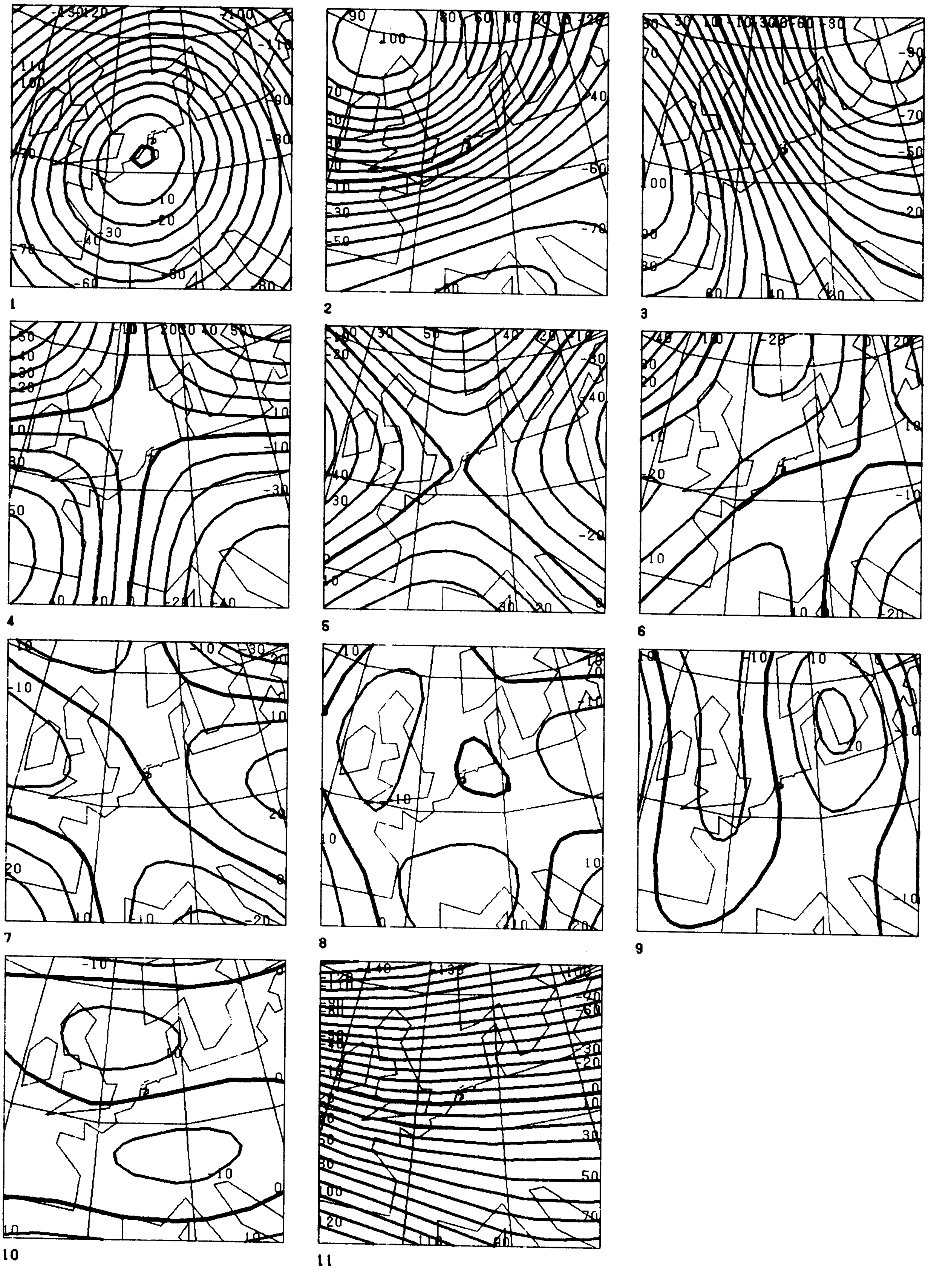

Fig. 7 The same as Fig. 5, however for 500 mbar height.

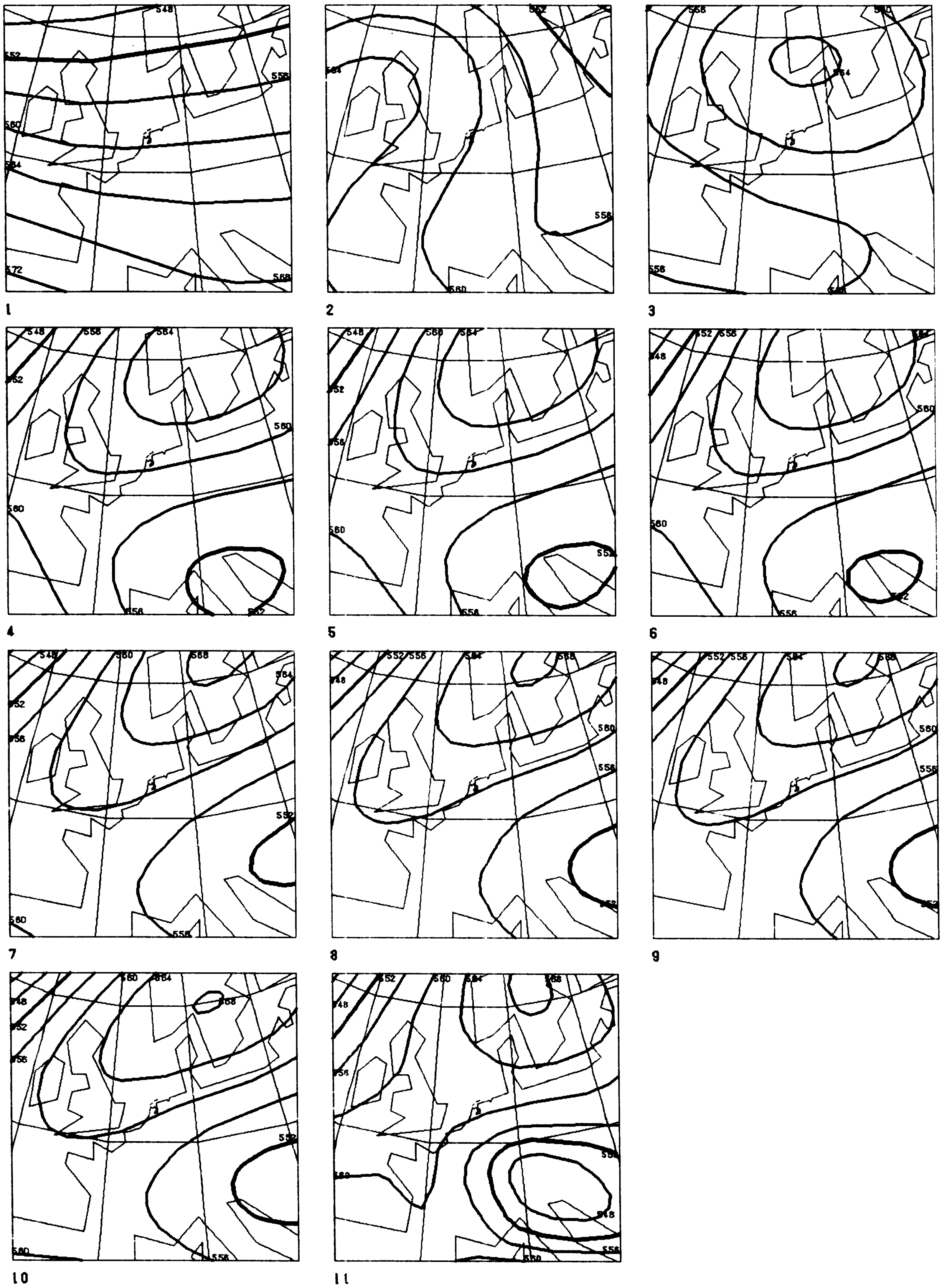

Fig. 8 In picture 11 the ECMWF 500 mbar analysis of April 29, 1984 is presented. Picture 1 represents the mean flow plus the first component, picture 2 the mean flow plus the first 2 components and so on. The coefficients are: -0.01, 1.28, -1.01, 1.13, 0.28, -0.15, -1.01, -0.55, -0.03, 0.69.

$\langle Z_i \rangle = Z_i - Z_{24}$ averaged over 8 years
P_i^n = value of n^{th} principal component
C_n = coefficient of n^{th} principal component
N = number of gridpoints in the used grid
The values of P_i^n are computed from the NCAR-data set. Each field can be expressed in these principal components and will have its own unique coefficients C. It is these coefficients that are used as predictors.
Using only the first five components, describing the large scales, already 95% of the variance could be explained, so we use only these five as predictors.
In Figures 5, 6 and 7 the first 10 principal components of 1000, 850 and 500 mbar are presented. In Figure 8 an example of a 500 mbar field with respectively 1, 2, 3, ..., 10 principal components is shown.

7.2 Climatology

For climatology of the predictands see the concerning reports. Many parameters have an annual variation which can be represented in terms of a mean value plus a Fourier series of sinusoidal components:

$$Y(I) = \overline{Y} + \sum_{n=1}^{N} \left(A_n \sin(n.I.\frac{2\pi}{365}) + B_n \cos(n.I.\frac{2\pi}{365})\right)$$

I = day of the year, January 1 = 0, February 5 = 35 and so on.

Based on the NCAR data set the coefficients are established for the 850 and 500 mbar height and the 850 mbar temperature. In Table I the coefficients are presented. The number of components N is two, which gives a good approximation as can be seen from Fig. 9.

Table I

	$\overline{Y}$	A_1	A_2	B_1	B_2
850 mbar temperature	2.7	- 3.2	+ 0.6	- 4.6	+ 0.6
850 mbar height	1459	-23.7	+ 2.2	- 40.2	+ 3.6
500 mbar height	5568	- 76.8	+11.6	-110.7	+14.3

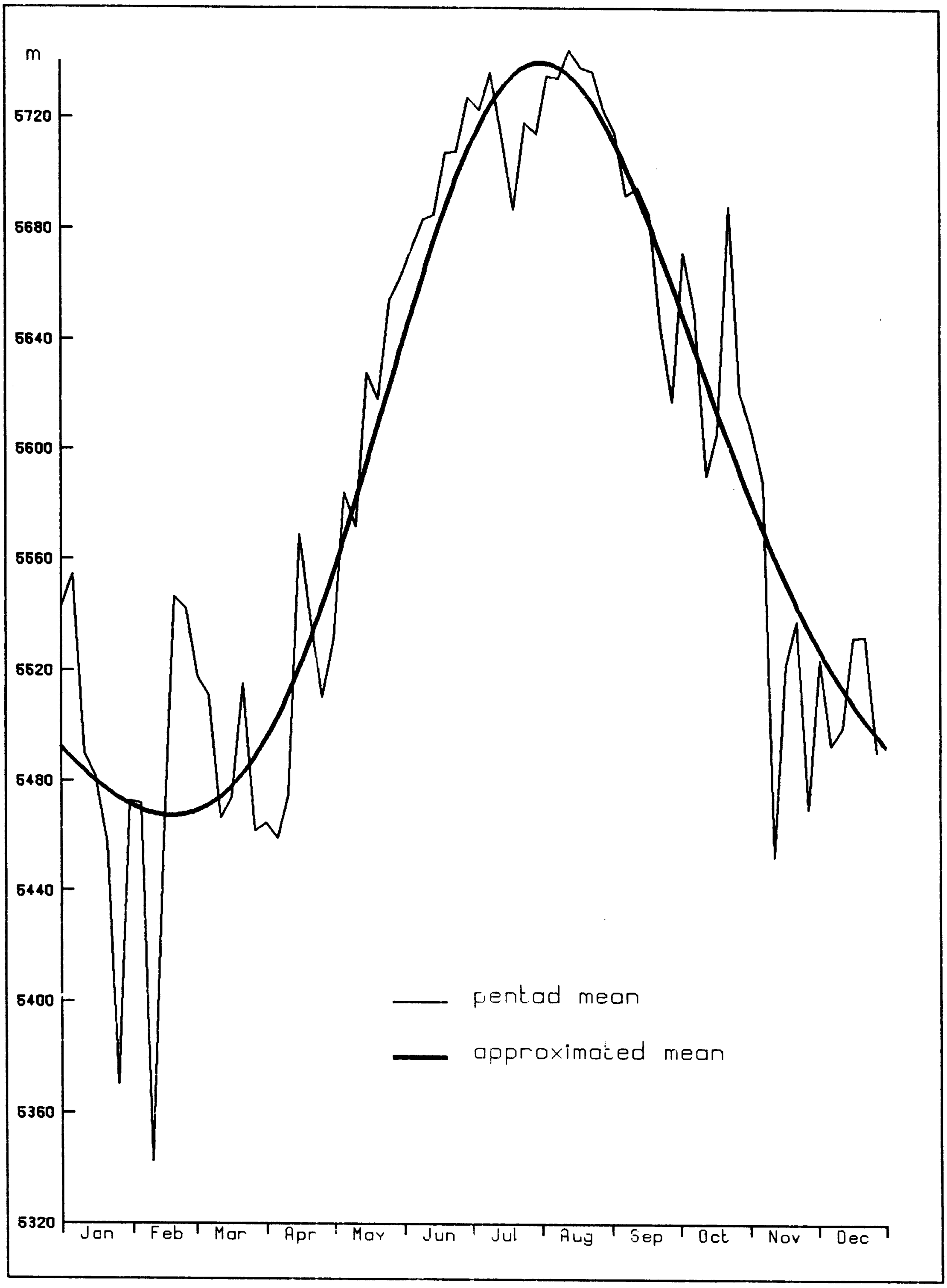

Fig.9. Height of 500 mb averaged over 1972-1979.

8. Verification of forecasts

There exists a wide range of scoring rules for all types of weather elements. In this Chapter we will give a short description of the scoring rules in common use at the KNMI. We will distinguish between the score and the skill of a time series of forecasts. With a score we will indicate a numerical value giving the average "distance" between the series of forecasts and corresponding observations. This distance will be zero in the case of a perfect forecast system. The skill is subsequently defined to be the relative improvement obtained with a forecast technique relative to some reference forecast system, usually climatology. The skill is computed from the score by

$$\text{SKILL} = 100 * \frac{\text{SCORE}_{\text{CLIMATOLOGY}} - \text{SCORE}_{\text{FORECAST}}}{\text{SCORE}_{\text{CLIMATOLOGY}}} \qquad (1)$$

A perfect forecast system scores 100 percent whereas climatology or an equivalent forecast system gets zero skill. Usually the reference $\text{SCORE}_{\text{CLIMATOLOGY}}$ will be the score obtained by climatology applied to the same period as the one for which the forecast score is computed. In theory it is not allowed to average skills obtained in such a way. In practice, however, the error will be small. The used climatology is at least a monthly climatology unless specified otherwise.

First some scoring rules will be given related with the type of forecast, probabilistic forecast versus point forecast. Thereafter a skill score called "Performance Index" (PI) is described. This PI can be used in any case after a minor transformation of the forecast.

Finally the reliability of probabilistic forecasts is discussed briefly.

8.1 Probabilistic forecasts

Giving the probability that a certain category will occur is a very useful way of presenting forecasts. The categories used in such a case are usually mutually exclusive and cover the complete range of possible outcomes. The categories used can be either natural e.g. rain, snow, hail or artificial e.g. below normal, normal or above normal with temperature forecasting. However in all cases the determination of which category has occurred must be unambiguous. It is worthwhile to make a distinction between the two-category case and the multi-category case.

8.1.1 The two-category probability forecast

A simple example of such a forecast is the precipitation/no precipitation forecast. The forecast only needs to specify the probability of one category. In this case we use the adapted Brier-score (Brier, 1951). If we denote the forecast time series of one category as p_n (in %) and the occurence of that category as o_n (occurence denoted by 100, no occurence by 0) then the Brier-score BS of N forecasts is defined:

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

8.1.2 The multi-category probability forecast

In the case that the categories have a natural ordering we use the average Rank Probability Score ($\overline{RPS}$) (Epstein, 1969; Daan and Murphy, 1982). Natural ordering mostly occurs when a continuous scale of a weather variable has been subdivided into categories. For instance with temperature the categories below normal and above have a natural ordering. With no natural ordering e.g. rain, hail and snow we use the original Brier Score. We denote the n^{th} forecast of a K category forecast as $(p_{1,n}, \ldots, p_{K,n})$ and the subsequent observation as $(o_{1,n}, \ldots, o_{K,n})$, with one of the o_n's = 100 all the other 0. Now the Brier Score of N forecasts follows from

$$BS = \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{k=1}^{K} (p_{k,n} - o_{k,n})^2 \right) \tag{3}$$

and the RPS of N forecasts

$$RPS = \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{k=1}^{K} (P_{k,n} - O_{k,n})^2 \right). \tag{4}$$

where $P_{k,n} = \sum_{k'=1}^{k} p_{k',n}$

and $O_{k,n} = \sum_{k'=1}^{k} o_{k',n}$.

8.2 Point forecasts for continuous elements

For elements such as temperature and wind-velocity the forecast is frequently expressed as a single value (point-forecast) in whose neighbourhood the observed value is expected. Given such a forecast we usually use the mean of the time series of absolute differences as a scoring rule. Incidentally the (root) mean square error will be used. The skills are computed according to eq. 1.

8.3 Performance Index (PI)

This scoring rule, introduced by Hanssen and Kuipers (1965), leads directly to a measure of skill. The PI is used at our Institute for all relevant weather elements and records of monthly averages covering the last 25 years are available. In order to use this score the forecast must be transformed into one of two standard formats:

1. The forecast is expressed as one or more categories in which the observation is expected or
2. The forecast is expressed as one or more intervals on a continuous scale. The observation is expected within one of the intervals.

After that the PI is obtained in the following way. First the climatological probability, P_C, of a hit of each forecast is computed. Secondly the relative frequency P_H of hits is ascertained. (Both expressed in percent.) The PI is defined to be

$$PI = P_H - \overline{P_c}$$

the overbar indicates averaging over the time series. The PI of a random forecast from climatology scores zero. The maximum PI depends on the nature of the forecast.

When the PI is used to verify forecasters results the forecaster is requested to indicate on a form the categories or intervals (see Kuipers, 1980). When applied to forecasts resulting from an objective scheme it is necessary to use a transformation rule which translates the objective forecasts to the requested formats.

The transformation of the forecasts to the standard format is very simple in the case of probabilistic forecasts: Naming all the categories with a forecast probability higher than the climatological probability gives an optimum PI when the probabilities are reliable. For categorial forecasts there is no problem. However, for point-forecasts on a continuous scale an optimum rule is not easily found. In the practical case this problem is solved by assigning an interval to each possible forecast value. The boundaries of this interval are based on experience obtained with similar forecasts in the past. When used the chosen intervals are given. In this case one must be careful when comparing two Performance Indices. Differences may be due to different transformation schemes instead of different skills of the forecast systems.

8.4 Reliability of probabilities

Apart from having a good skill it is often required that forecast probabilities are "reliable" (Sanders, 1958; Murphy, 1973). This means that if we select from a set of forecasts all forecasts of say 25% the event is expected to occur in 25% of all cases after those forecasts. Usually the reliability is studied even in the multi-category case for each of the categories independently. We will limit our discussion about reliability to the one event case (two categories) and the forecast probability is the probability that the event will occur. Reliability is mostly studied by presenting a reliability diagram. Such a diagram is constructed in the following way. First the forecasts are grouped in intervals of forecast probability, for instance in ten groups covering the intervals 0-10%, 10-20% and so on. The relative frequency of occurence of the event after a forecast in a given group is plotted versus the midpoint of the intervals. For reliable forecasts this relative frequency is expected near to this midpoint. So the plotted points should cluster near to the diagonal running from (0, 0) to (100, 100). In Figure 10 the reliability diagram of the 48 hour forecast probability on precipitation at De Bilt in a 12 hour period obtained with the analogue technique is given as an illustration. However, reliability can also be interpreted as a part of the Brier Score (Daan and Murphy, 1982). In terms of the diagram the reliability defined by Murphy is the weighted average of the squared vertical distances between the plotted points and the diagonal. The weights must be chosen equal to the number of forecasts assiociated with the plotted point. If we denoted the midpoint of the interval by P_s, the observed frequency by O_s and the number of forecasts in each interval by N_s, then the reliability RLB can be found by

$$RLB = \frac{1}{N} \sum_{s=1}^{S} N_s \, (O_s - P_s)^2$$

with N the total number of observations and S the number of intervals. In the ideal case RLB will be zero. However, the computed RLB is the result of a random process and therefore almost never will be zero even when the forecasts are strictly reliable. The expected value of RLB depends strongly on the total number of forecasts. The expected value can be estimated from the binomial frequency distribution in the following way:
assuming the forecasts probabilities to be reliable the observed frequency O_s will be the result of N_s trials in a binomial process with probability P_s. Therefore the expected value of O_s will be P_s and the variance of O_s will be $P_s\ (100 - P_s)/N_s$. Now the variance is equal to the expected value of $(O_s - P_s)^2$ and the expected value of RLB is found by replacing this term by its expected value

$$E(RLB) = \frac{1}{N} \sum_{s=1}^{S} N_s \frac{P_s(100 - P_s)}{N_s} .$$

When the 10 interval grouping described before is chosen these equation simplifies to

$$E(RLB) = \frac{16750}{N} .$$

where P_s and O_s are expressed in percent. The reliability diagram in Figure 10 results in an RLB of 37,3. The expected reliability is 16,0. The diagram in Figure 10 clearly illustrates that the analogues overestimate the probability of precipitation at all levels.

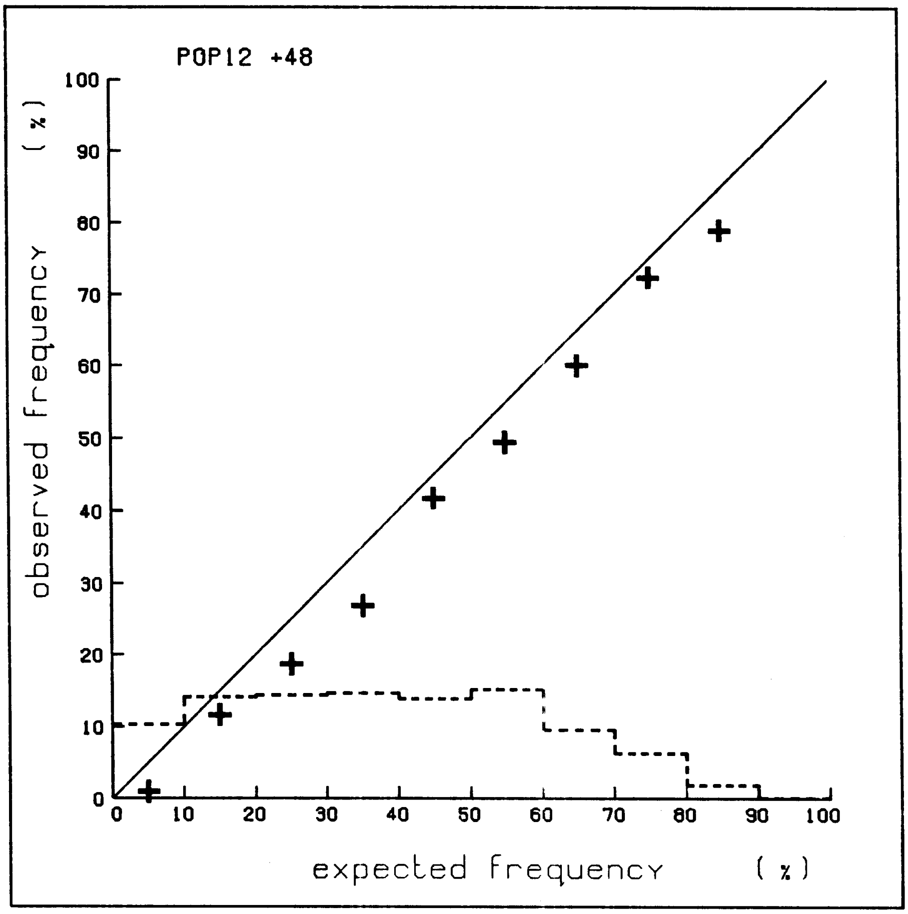

Fig. 10 The reliability diagram of the forecast probability of precipitation in a twelve hour period at De Bilt. Lead time 48 hours, verification period 1-12-1980 to 1-12-1983, 1048 forecasts. The dotted histogram gives the relative frequency of forecasts in each group.

9. Literature

Allen, R.A. and Vernon E.M., 1951: Objective weather forecasting. Compendium of meteorology.

Anderson, J.A., 1972: Separate sample logistic discrimination. Biometrika (1972), 59, 1, 19-35.

Balzer, K., 1976: Method and results of interpretation of geopotential forecasts, WMO symposium on the interpretation of broad-scale NWP products for local forecasting purposes. Warsaw, 11-16 October, 1976.

Dempster, A.P., 1969: Elements of continuous multivariate analysis. Addison-wesley publishing company.

Glahn, H.R., 1965: Objective Weather forecasting by statistical methods. The Statistician, Vol. 15, No. 2.

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

Glahn, H.R., 1982: Statistical Weather forecasting. ECMWF Seminar on interpretation of numerical weather prediction products.

Hanssen, A.W. and W.J.A. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. KNMI, Mededelingen en Verhandelingen nr. 81.

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

Kruizinga, S. and J.P. de Jongh, 1975: Analogenselectie met behulp van de computer voor de meerdaagse verwachting. KNMI verslag V-267. Dutch only.

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, 8-12 September 1980).

Lowry, D.A., 1980: How to use and not use MOS guidance. Proceedings of Eight Conference on Weather Forecasting and Analysis, Denver, Colo. Publ. by the Am. Meteor. Soc.

Murphy, A.H., 1973: A new vector partition of the probability Score. J. Appl. Meteor., 12, 595-600.

Sanders, F., 1958: The evaluation of subjective probability forecasts. Cambridge, Massachusetts Institute of Technology. Dept. of Meteor., Sci. Rept. No. 5 [Contract AF 19 (604-1305)], 62 pp.

Staver, A.E., 1980: More light on MOS. Proceedings of Eight Conference on Weather Forecasting and Analysis, Denver, Colo. Publ. by the Am. Meteor. Soc.

<u>KNMI memoranda</u>, unpublished, Dutch only.

Kruizinga, S. and E. Hofstee, 1979: Introductie van de objectieve gidsverwachting, memo 79-222.

Lemcke, C. and S. Kruizinga, 1981: De gidsverwachting op ECMWF-produkten (1) Tekst & Uitleg 81-1.

Kruizinga, S. and C. Lemcke, 1981: De gidsverwachting op ECMWF-produkten (2) Tekst & Uitleg 81-2.

Kruizinga, S. and C. Lemcke, 1983: De gidsverwachting op ECMWF-produkten (3) Tekst & Uitleg 83-2.