

# Optimal verification of ECMWF surface temperature forecasts at two locations in Europe

H. Böttger and S. Grønås

Research Department

July 1982

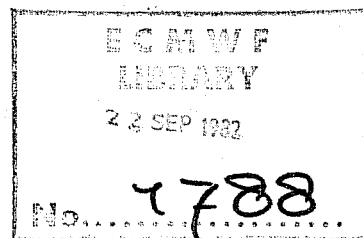
This paper has not been published and should be regarded as an Internal Report from ECMWF.  
Permission to quote from it should be obtained from the ECMWF.



European Centre for Medium-Range Weather Forecasts  
Europäisches Zentrum für mittelfristige Wettervorhersage  
Centre européen pour les prévisions météorologiques à moyen

CONTENTS

1. Introduction
2. The data
3. Verification of surface temperature for De Bilt and Hannover
  - 3.1 Temperature as a predictand
  - 3.2 The predictors
  - 3.3 Results
    - 3.3.1 1-day mean temperature
    - 3.3.2 3-day and 5-day mean temperature: the impact of time and space averaging
    - 3.3.3 Timegraphs of temperature
    - 3.3.4 Optimal verification: the multiple predictor case
4. Summary
5. Appendix A: The optimal regression technique
6. References



## 1. INTRODUCTION

ECMWF has been producing medium range forecast products on an operational basis since August 1979. All numerical products become available to the Member States as grid point values on standard pressure levels for the upper air data and on the model's surface in an equidistant horizontal latitude and longitude grid. The model runs with a 1.875 degree grid interval but, for the user and for archiving purposes, the values may be interpolated to other grid intervals. The catalogue of products available directly from the model contains a wide range of parameters, including the traditional upper air fields like wind, height, temperature, humidity. Additionally, there are surface fields available, such as temperature at 2m above the model surface, wind at 10m, precipitation and model-predicted cloud. These can be used directly as guidance to predict near surface weather conditions.

After three years of operational forecasting, the Centre and the Member States have gained some experience with the ECMWF products. Initially the objective and subjective evaluation of the forecast was focussed on the large and synoptic scale features of the circulation. This traditional approach of assessing the forecast quality in terms of the differences between predicted fields and analysed fields gives valuable information related to the model's ability to forecast circulation patterns.

There is a further need to assess the model's ability to forecast surface or near surface weather parameters, and regional and local verification of such direct model output has been performed (Åkesson, Johannessen, Åkesson et al). The results show that forecasts of many parameters are subject to systematic biases and other errors which could be reduced substantially through the use of statistical interpretation models.

Such methods are widely used in short range weather prediction. They quantify the uncertainties in the forecast and thereby increase the information content and usefulness of the products. Further, the methods could be used to estimate weather parameters not currently predicted by the model and thus the requirements of the ultimate users of forecasts could be taken into account.

There is so far little experience with interpretation methods in medium range forecasting. The results of the regional and local verification studies already mentioned indicate that different kinds of errors, including systematic errors like phase errors, contaminate the verification results when a typical short range approach is assumed in the verification. Useful forecast information will still be present, but new forms and modes of presentation are needed. In terms of statistical interpretations, this means that useful predictands and predictors have to be defined.

The present study may be considered as a pilot study in using interpretation methods in medium range forecasting. Compared to many previous models, the ECMWF model contains a reasonable resolution in the boundary level with four sigma levels between sea level and 700mb and a comprehensive physical package is used. One objective is therefore to evaluate the usefulness of boundary layer parameters as predictors for surface weather forecasting. As a by-product, verification of direct model output at the local site will enable modellers to check the impact of model changes on parameters which are widely used as guidance in local and regional weather forecasting.

The local temperature forecasts at two different locations in central Europe are investigated and only data from two and three month periods respectively are used. By an optimal regression technique, which is explained in Appendix A, the linear relationship between the observed surface temperature and several model predictors is explored. All the model predictors which are used are either temperature values at different levels or are parameters which are known to affect locally observed temperatures, including precipitation, cloud amount and surface wind. We found that at the two locations chosen for this study, Hannover and De Bilt, the model temperature in the boundary layer is highly correlated to the observed value and that other parameters contribute very little to improve this correlation. We therefore put our emphasis on possible corrections for phase errors in the model by using time offset predictors. This is described in the first paragraph of the results. In the next chapter, the question is then raised of whether time and space averaging have any beneficial effect on the guidance that can be obtained from the direct model output for predicting the local temperature in the medium range. We then finally outline in the results what is potentially achievable, when using multiple predictor equations. It should be stressed, however, that this is a dependent data sample. Derivation of stable prognostic equations to be used on independent data was not intended.

## 2. THE DATA

The ECMWF operational numerical weather prediction system uses the observations exchanged world-wide via the Global Telecommunications System. The data assimilation at ECMWF runs in a cycle providing four analyses per day for 00, 06, 12 and 18Z. Once a day, starting from 12Z, a 10-day forecast is run operationally. The model computations are performed on 15  $\sigma$ -coordinate levels in the vertical and on a staggered horizontal grid of 1.875 x 1.875 degrees of latitude and longitude. The predicted free atmospheric variables are interpolated in a postprocessing to standard pressure levels and to a non-staggered horizontal grid of 1.5 x 1.5 degrees, in which form they become available to the Member States. All surface parameters are interpolated to the same grid. Out to 84 hours of forecast time every 6 hours and thereafter out to 240 hours every 12 hours, the model output is postprocessed but only the 12-hourly steps are available from the archive. Further information on ECMWF products is available through the ECMWF Meteorological Bulletins.

All observations and fields are stored in the ECMWF reports and fields data base, an archive which is comprehensive, global and accessible via the GETDATA retrieval facility. However, the quality of the observational data is not satisfactory for detailed verification and interpretation studies. It is real-time synoptic data which has undergone only crude quality control checks against climatological limits before being stored in the reports data base. No spatial cross checks or time consistency checks are carried out at that stage.

It is, therefore, recommended that studies, using the observational data from ECMWF, are confined to stations known for their reliability.

In this preliminary study, we used two stations, De Bilt in the Netherlands and Hannover in the Federal Republic of Germany. They are both situated in flat homogeneous terrain with a strong maritime influence at De Bilt which decreases towards Hannover, situated at the southern part of the North German plains. For De Bilt, we retrieved 2 winter months, January and February 1982, of surface observations (SYNOPS), the corresponding ECMWF forecasts of near-surface parameters<sup>(1)</sup> and the forecasts of upper air parameters for the four levels 1000, 850, 700 and 500 mb.<sup>(2)</sup> In the De Bilt study, we limited ourselves to forecasts every 12 hours out to 168 hours and to only one grid point which is a few kilometres away from the geographical location of the observing site. For the Hannover study, we extracted the same set of observation and forecast parameters, but for 3 winter months, January to March 1981, and for the complete length of 10 day forecasts out to 240 hours. Further, an array of 11 x 11 grid points was extracted around the location of Hannover instead of using only one gridpoint, thus allowing various space averages to be taken in the course of the investigation. At the time of this study, a new European archive was under development but was not yet available. All computations were carried out in the original model grid interval of 1.875 degrees.

- 
1. Temperature at 2m and wind at 10m above the model surface (for their definition, see the ECMWF Meteorological Bulletin M3.4/1(3)).
  2. It should be pointed out that the 1000mb parameters are extrapolated below the model's surface, if the surface pressure is below 1000mb.

### 3. VERIFICATION OF SURFACE TEMPERATURE FOR HANNOVER AND DE BILT

#### 3.1 Temperature as a predictand

We are primarily interested in investigating a verification of ECMWF temperature forecasts. As the diurnal cycle at present is not simulated in the ECMWF forecasting system, it seems reasonable to verify temperature forecasts against observed daily mean temperatures. In the early stage of the forecast, the one-day mean temperature will be of prime interest. After day 4 of the forecast, the model errors become substantial and any spot verification in space and time will be contaminated by errors in phase and amplitude of the migratory waves (wavenumber  $\geq 7$ ). All verification results show (e.g. ECMWF Forecast Reports) that a higher predictability is achieved in the long wave domain. This broad scale flow pattern determines the prevailing weather type, providing valuable forecast information in the medium range.

Phase errors can be cancelled out by averaging the predictand in time or space. This will result in a loss of information for day to day forecasting at given locations but, by compressing the information from several gridpoints and forecast steps, the reliability and, for some purposes, the usefulness of the forecast will be increased.

In this study, we have chosen to verify the three-day and five-day mean temperature together with the daily mean temperature at two locations. No averaging in space of the predictand is applied. This seems to be more important for discontinuous parameters like precipitation (Åkesson, 1981), where the inhomogeneity in the observations has to be smoothed in order to achieve meaningful verification results in accordance with the model's resolution.

#### 3.2 The predictors

With today's high speed computers, the derivation of statistical relationships between numerical prediction products and near-surface weather parameters is a straight forward procedure. Any reasonable amount of predictors can be offered and the statistical forecast model will emerge from the computer within the preset limits of the significance level. In order to establish a statistically significant relationship between direct model output and observed weather parameters (Model Output Statistics technique) numerical forecasts covering at least two years of data are needed. Even then, the selection of predictors can depend very much on events typical of this relatively short period and the derived prognostic equation might become unstable when applied to independent data.

Given the limitations of the short data sample used in this study, it is not intended to derive any prognostic equations for prediction of local temperature conditions. Instead, by verifying different physically meaningful parameters related to surface temperature conditions and predicted directly by the large scale model, the attempt is made to find the best single predictor to forecast the surface temperature.

In a previous verification study of direct model output of near-surface weather parameters carried out for 17 locations in Europe (Åkesson et al, 1981), it was shown that the model develops a strong bias for the temperature and wind in the boundary layer and, when using the forecast as spot values in space and time, hardly any skill is left after day 3 or 4 of the forecast time. Meanwhile, this study for the 17 stations has been extended and time averaging proved to be a useful tool to eliminate some of the effects of phase errors in migratory short waves. Additionally, the verification of predicted tendencies gave promising results and eliminated largely the bias effects. In this current study, the use of averaging in space and time will be evaluated and, by offering the predicted values from all 12 hourly forecast steps in one specific verification, the effect of phase errors in the model is explored.

### 3.3 Results

#### 3.3.1 1-day mean temperature

Verification of the temperature forecast for De Bilt for January and February 1981 is given in Figure 1. The direct model output of temperature at 2m above the model surface is compared to the observed 1-day mean temperature calculated from the four synoptic times 00Z, 06Z, 12Z and 18Z.

Figure 1 shows the explained variance of the observed temperature variation at De Bilt for the day 1, day 2, etc. out to the day 6 forecast. Each of the six graphs gives the fraction of variance explained from all 12 hourly forecast steps out to 144 hours for the specific time of verification.

For the day 1 and the day 2 forecast, the 2m temperature predicted for the De Bilt gridpoint for 24 and 48 hours of forecast time respectively gives the best explanation of the observed variance, while the previous and the following forecast steps clearly contain less information. For the subsequent forecast times, day 3, day 4, day 5 and day 6, the curves get flatter and, apart from some irregularities which might be specific to this particular data sample, there is a tendency that towards day 5 and day 6, the forecasts from earlier timesteps than those valid at the time of the verification contain as much or even more information and thus explain a larger fraction of the observed variance. In other words, when forecasting

# 1-DAY MEAN TEMPERATURE DE BILT

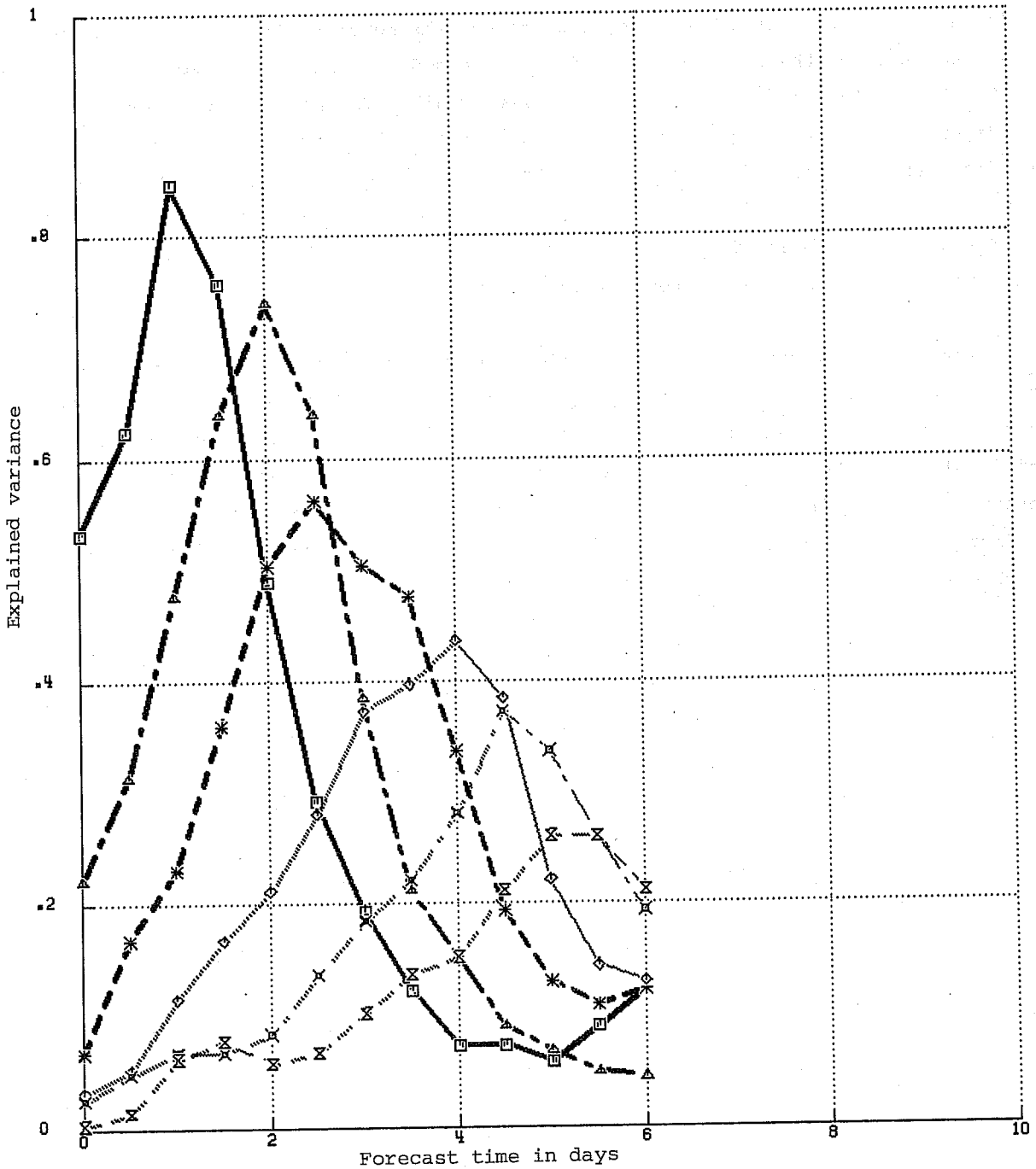
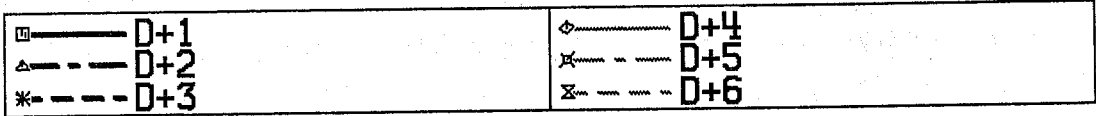


Fig. 1. Verification of 1-day mean temperature forecasts for De Bilt, January and February 1981, valid for day 1 to day 6. The predictor is the temperature at 2m above the model surface.



beyond day 4-5, persistence taken from the model forecast of day 4½ and day 5 seems to be a good predictor. The overall quality of the forecast drops sharply towards day 6 when only about 25% of the variance can be explained by the predicted 2m temperature. The forecast quality can be compared with a persistence forecast quality indicated in the figure for forecast day 0. It indicates the amount of variance that can be explained by using the observed daily mean temperature as a forecast for each of the six verifying times. As can be expected, the persistence gives high scores as a predictor for the first and even for the second day, but for forecasts beyond day 2, the persistence hardly shows any skill at all. The persistence forecast for day 1 and day 2 is used as a standard of comparison to assess the quality of the forecast. In the short range, the model output shows considerable skill compared to the day 1 persistence. By day 3, the quality of the forecast corresponds to that of a persistence forecast for day 1, and by day 6, the forecast quality drops to the level of a persistence forecast for day 2.

When correlating the predicted temperature at 1000mb with the observed 1-day mean temperature at De Bilt, the results are very similar. In Figure 2, the verification results for both predictors respectively are given but for the forecast days 2, 4 and 6 only. In the short range out to day 3, the 2m temperature gives slightly better results, while for day 4 and thereafter, the 1000mb temperature is the predictor to be preferred. During both months, the average mean sea level pressure at De Bilt was close to 1020mb. As the elevations of both the observing site and the gridpoint in question are close to the mean sea level, the height of the 1000mb surface was on average above the model ground. It might therefore be suggested that, towards the medium range forecast period, predictors from even higher up in the atmosphere, e.g. 850mb or thickness values, might give more useful results in explaining near-surface temperature variations than the model values from these levels themselves. This, however, did not verify as later results will show.

The temperature verification at De Bilt was limited to using the predicted values at one gridpoint only. When verifying the temperature at Hannover, all the surrounding gridpoints over an area of 20 degrees latitude by 20 degrees longitude were available. It turned out that an average in space over the four gridpoints next to the observing site yields the best correlation with the observed daily mean temperature. Only for the first 48 hours of the forecast may the predicted value, interpolated linearly to the location of the station, have some advantage over the space average, but the difference is negligible and in the order of one or two per cent of explained variance.

The verification for Hannover, using the 4 point mean temperature as a predictor, is given in Figure 3. For the first six days of the forecast, the results are in good

# 1-DAY MEAN TEMPERATURE DE BILT

□ ——— D+2 △ ——— D+2 * ——— D+4	T AT 2M T (1000MB) T AT 2M	◇ ——— D+4 × ——— D+6 ⊗ ——— D+6	T (1000MB) T AT 2M T (1000MB)
-------------------------------------	----------------------------------	-------------------------------------	-------------------------------------

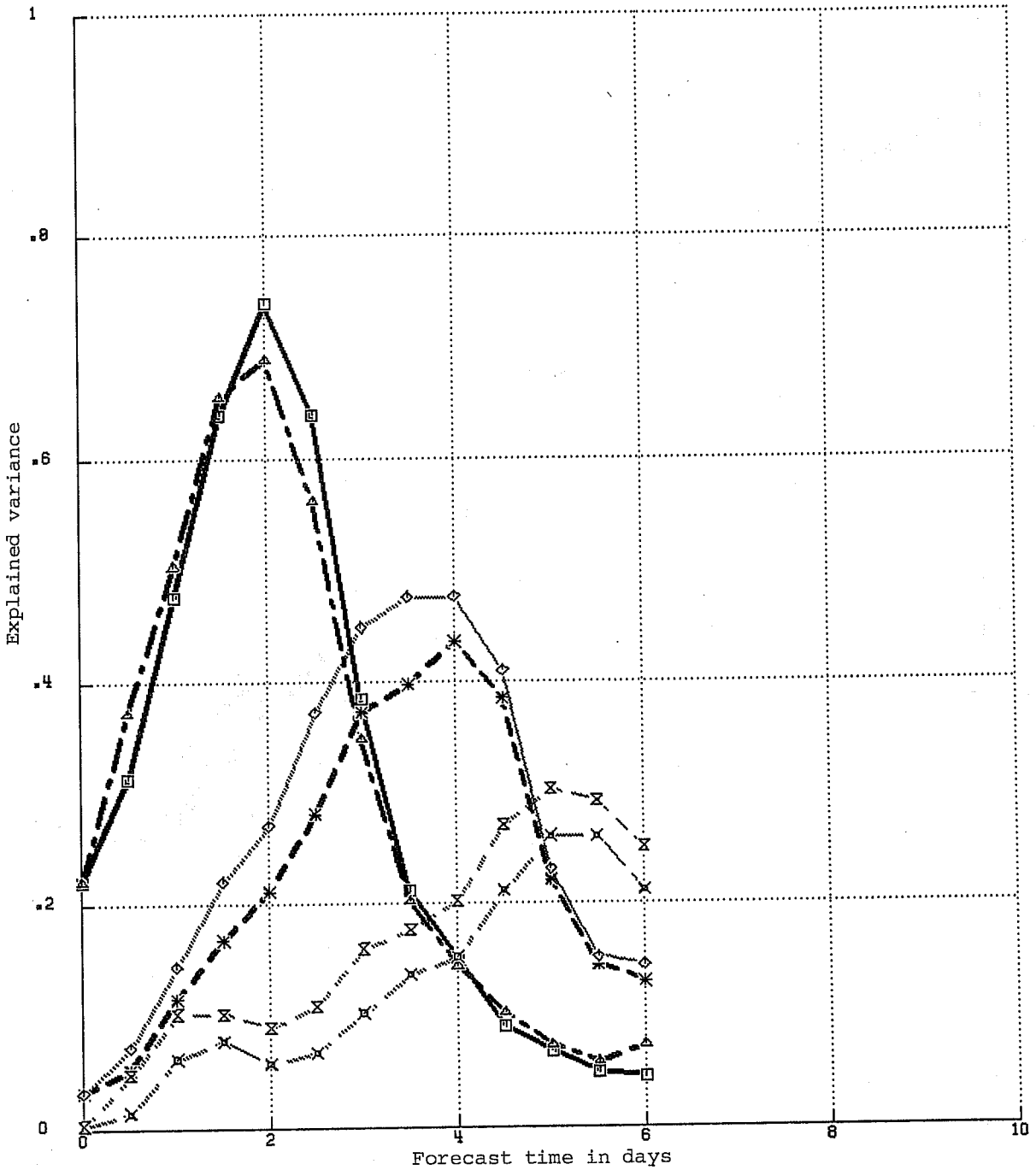


Fig. 2 Verification of 1-day mean temperature forecasts for De Bilt, January and February 1981, valid for day 2, day 4 and day 6. Comparison of two predictors, temperature at 2m and temperature at 1000mb.

# 1-DAY MEAN TEMPERATURE HANNOVER

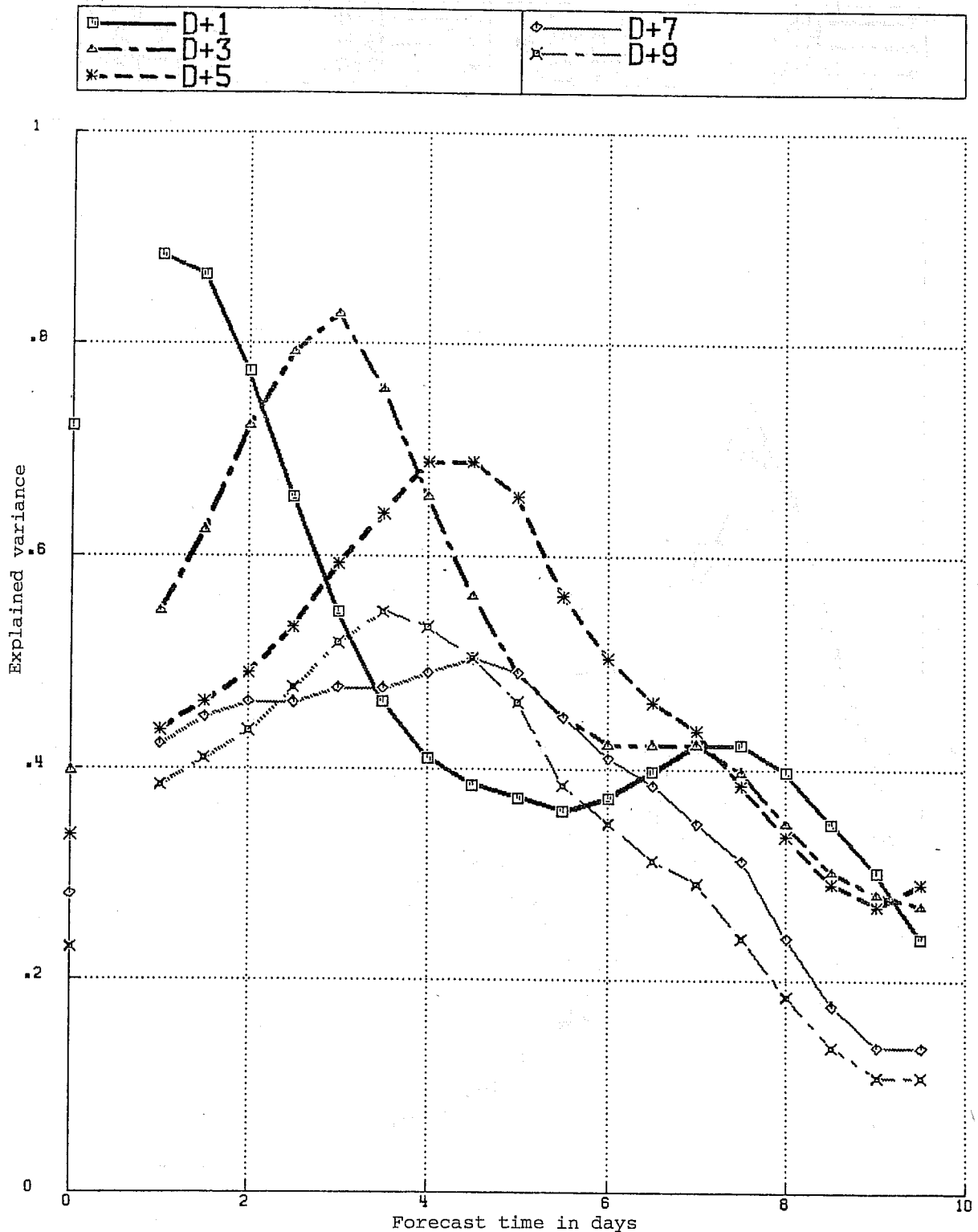


Fig. 3 Verification of 1-day mean temperature forecasts for Hannover, January to March 1981, valid for days 1, 3, 5, 7 and 9. The predictor is the temperature at 2m above the model surface, averaged over four surrounding gridpoints.

agreement with those obtained for De Bilt. In general, a larger fraction of variance can be explained by the forecast but this is due to the fact that, between January and March 1981, the seasonal trend becomes very pronounced in the data set. Persistence alone explains a larger part of the variance than during January and February at De Bilt. Still, using persistence as a standard of comparison, the quality of the D+5 forecast at Hannover is just below the level of a persistence forecast for D+1. This is an improvement against the De Bilt verification and it is generally supported by a more comprehensive verification study (Pümpel, Böttger and Geleyn, 1982) at 17 stations in Europe, where the best verification results are achieved at central European stations under continental influence and in homogeneous terrain.

The verification of the 120 hour forecast in Fig. 3 confirms the result from De Bilt that previous timesteps may explain a larger fraction of the observed temperature variance than the 120 hour forecast itself. After D+5, the quality of the forecast drops sharply. The results seem to indicate that only the predicted temperature of the first five days of the forecast period should be used and kept as persistence forecasts for the later stage of the forecast period.

A secondary peak can be detected in some curves, but it is most pronounced in the graph, showing the D+1 correlation. It improves significantly after another six days of forecast time. This upswing can also be detected in the results for De Bilt (Fig. 1). It seems to indicate a weather type cycle repeating itself after a period of around six days, a phenomenon which can just be particular to this data set. This cycle gives an explanation of the peculiar verification result for the D+9 forecast where the best correlation is obtained between day 3 and day 4, a time lag of approximately six days to the actual verification time of D+9.

### 3.3.2 3-day and 5-day mean temperature: the impact of time and space averaging

The verification of the daily mean temperature at two different locations (both over homogeneous terrain) gives evidence that the ECMWF model captures the temperature variations in the atmospheric layers close to the surface quite satisfactorily during the first 96 to 120 hours of forecast time. Both the temperature at 2m above the model surface and the 1000mb temperature are predictors of nearly equal quality in the early stage of the forecast. In the medium range forecast period around day 5, instantaneous forecast values verified against the daily mean temperature exhibit a time lag and an overall drop in the quality of the forecast. With increasing forecast errors, the spot values in time and space might not be the optimum choice of predictors, and further, upper air predictors like the 850mb temperature or thickness values might give better results than predictors from the model boundary layer.

Figure 4 gives the verification of the 3-day mean temperature observed at Hannover in January to March 1981 using the 1000mb temperature averaged over 16 surrounding gridpoints and over 3 forecast days as a predictor. After applying the time averaging, a corresponding space averaging over a larger area than previously gives the best result, but the difference to the 4 point average is not significant. Compared to the 1-day mean temperature, the curves are smoother. For day 3 and day 5, an increase in explained variance is achieved, but the quality of the forecast still drops rapidly for day 7 and day 9. The use of day 3 or day 4 as a persistence forecast gives the best indication of the temperature conditions towards the end of the forecast period.

In Figure 5, the verification result of the 3-day mean temperature for one forecast step, 96 hours, but for various predictors, is displayed. The temperature at 2m above the model surface is averaged over 4 and 16 gridpoints respectively but no averaging in time is applied, while for the predictor temperature at 1000mb, both averaging in time and space is used. As noted before, the differences to space averaging over 4 or 16 gridpoints are minor but, in general, one can state from various trials that, in the medium range forecast period around days 4, 5 and 6, the best results are achieved with an area averaged predictor for the 3-day and the 5-day mean temperature forecast. Further, for the prediction of a time averaged temperature, the predictors should be averaged over consecutive forecast days, accordingly

Figure 5 also shows that the temperature at 850mb and the thickness 700/1000mb are clearly not as useful as those near surface predictors. The best correlation is achieved 12 to 24 hours before the actual forecast time and there is a significant gap between the low level predictors and those representing the free atmospheric flow. This is underlined by the results given in Figure 6 where the explained variance is displayed after correlating the 5-day mean temperature at Hannover with the ECMWF model forecast of T at 1000mb and 850mb, both averaged over 16 gridpoints and 5 forecast days. Again, the temperature close to the model surface is to be preferred as a predictor. The 120 hour forecast is better than a day 3 persistence forecast, and the 168 hour forecast improves on persistence used for day 5. The 850mb temperature does not reach the quality of a persistence forecast. In common with all predictors is the fact that the best use of them will be made by applying earlier timesteps than the actual verification time to predict the temperature conditions five or seven days ahead.

# 3-DAY MEAN TEMPERATURE HANNOVER

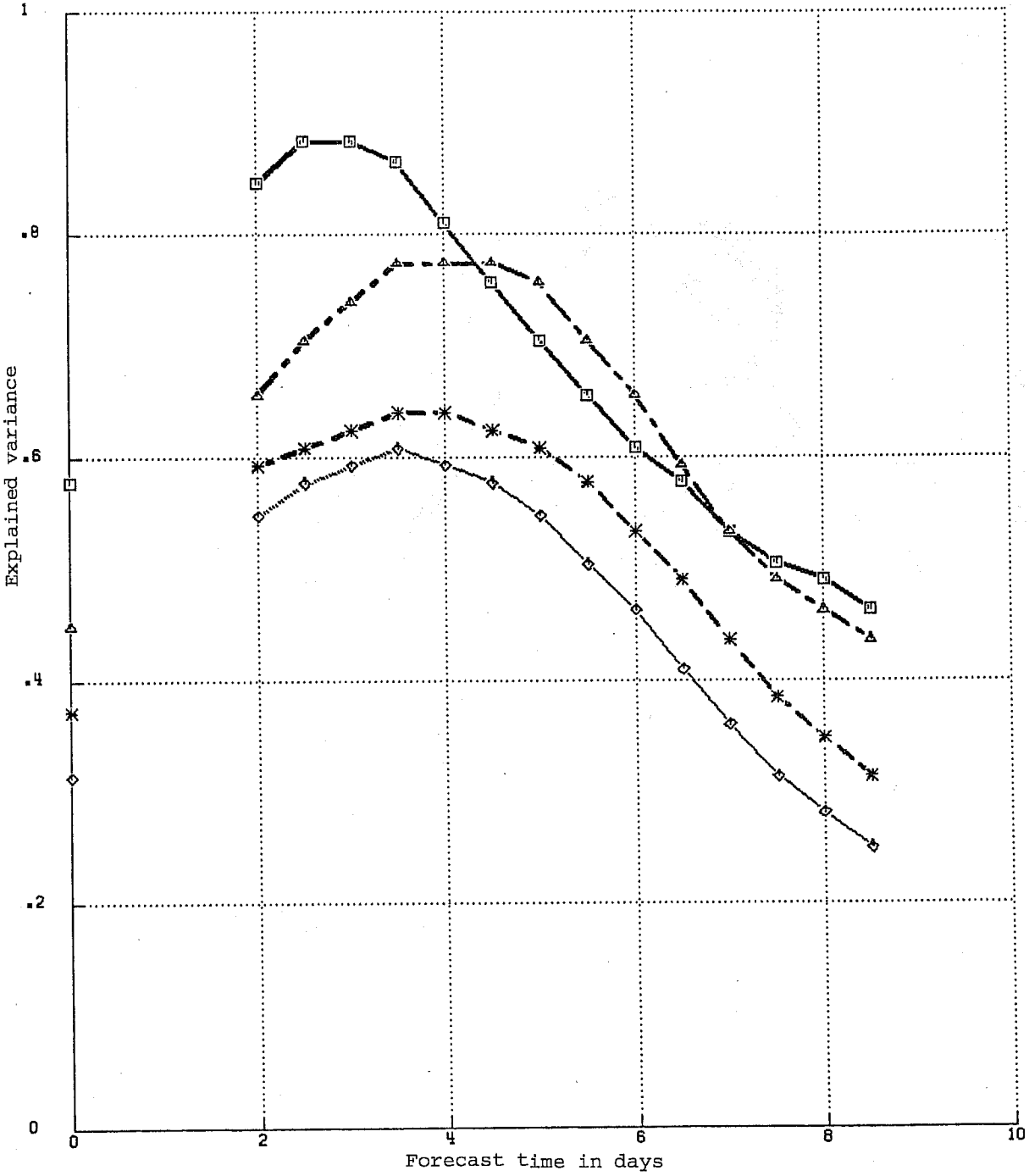
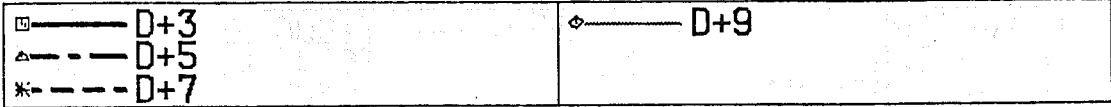


Fig. 4 Verification of 3-day mean temperature forecasts for Hannover, January to March 1981, valid for days 3, 5, 7 and 9. The predictor is the temperature at 1000mb, averaged over 16 surrounding gridpoints.

# 3-DAY MEAN TEMPERATURE HANNOVER

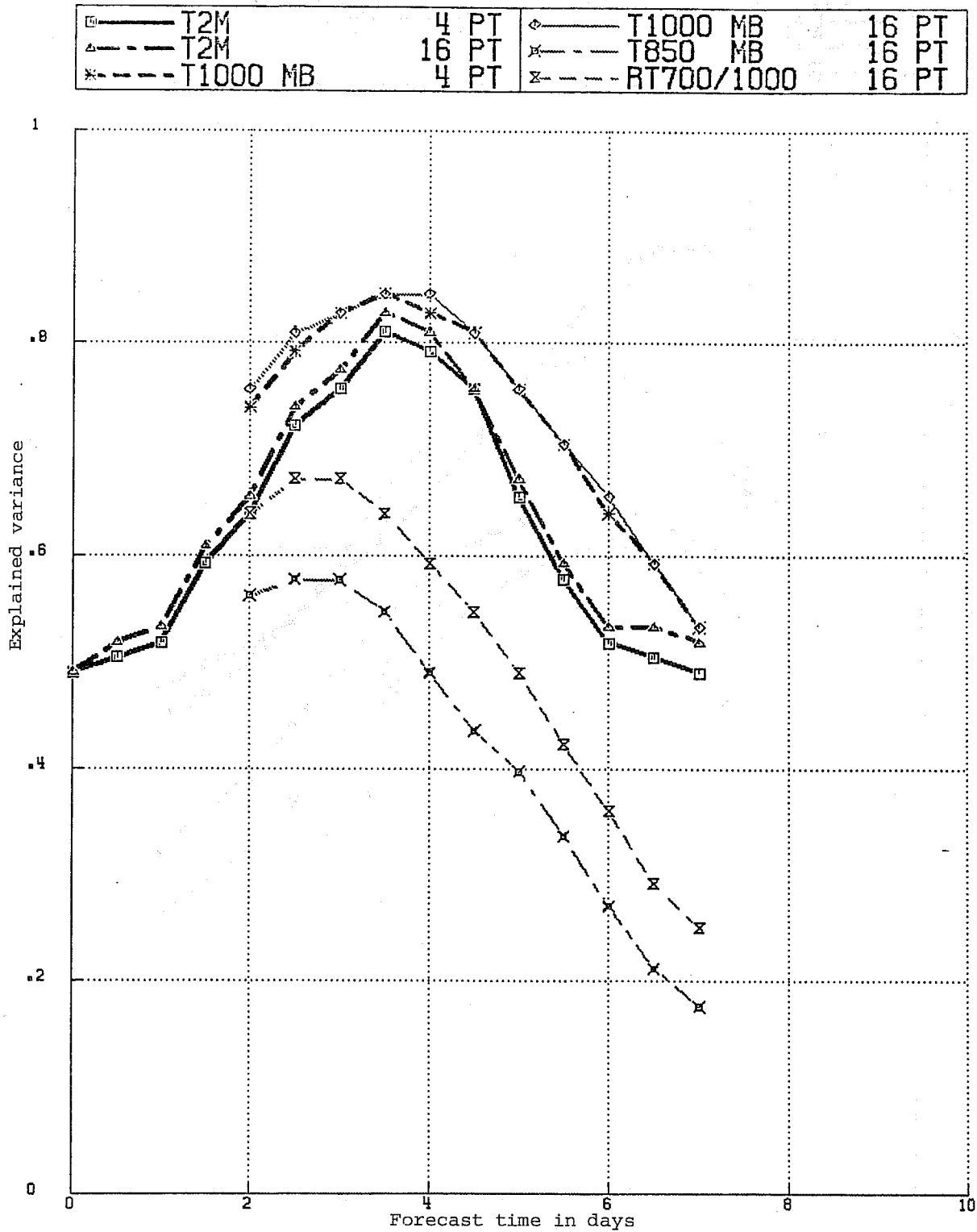


Fig. 5 Verification of 3-day mean temperature forecasts for Hannover, January to March 1981, valid on day 4. Comparison of various predictors as illustrated in the figure.

# 5-DAY MEAN TEMPERATURE HANNOVER

□ — D+3 △ — D+3 * — D+5	1000 MB 850 MB 1000 MB	◇ — D+5 × — D+7 × — D+7	850 MB 1000 MB 850 MB
-------------------------------	------------------------------	-------------------------------	-----------------------------

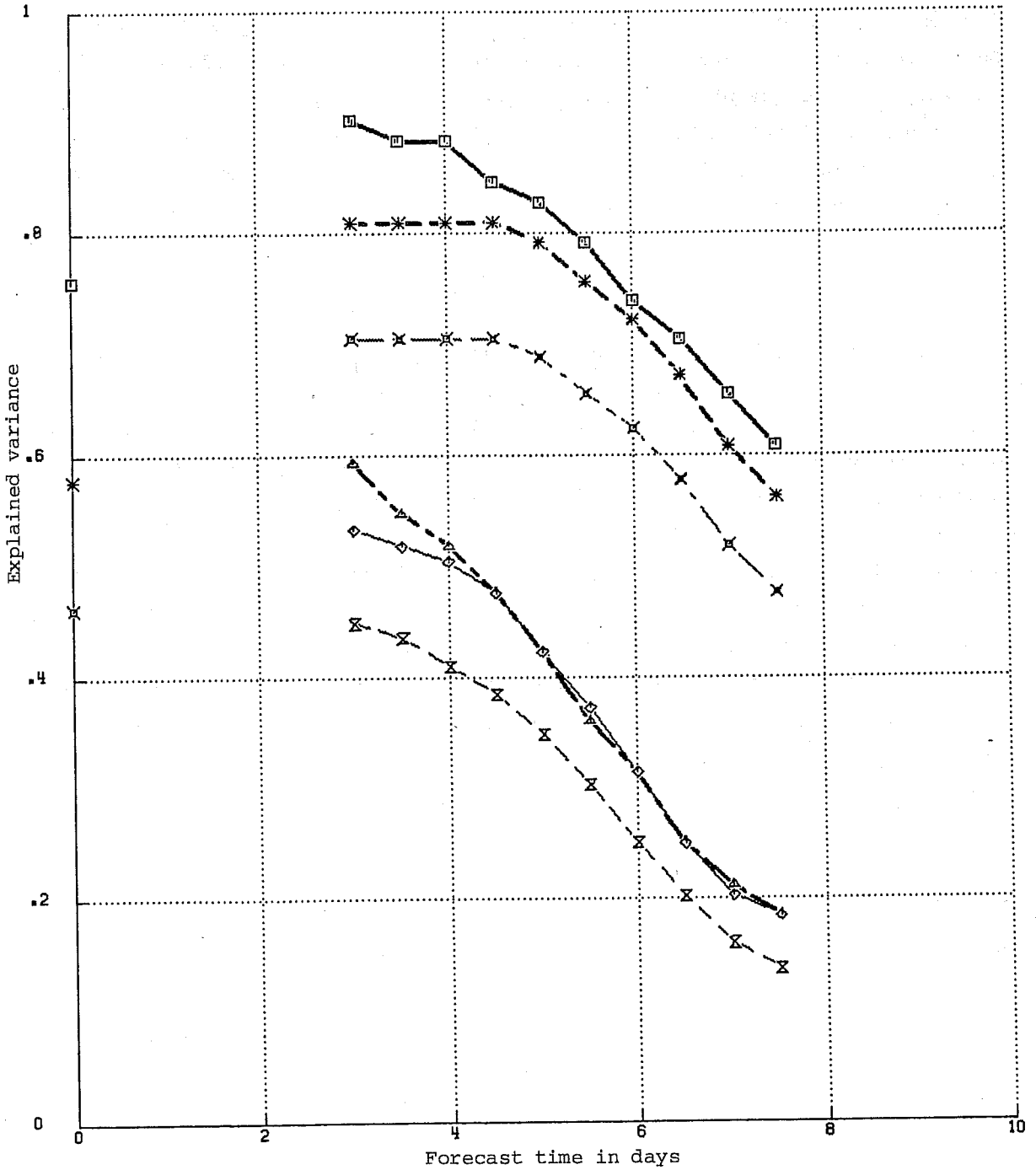


Fig. 6 Verification of 5-day mean temperature forecasts for Hannover, January to March 1981, valid on days 3, 5 and 7. Two predictors, temperature at 100mb and at 850mb, are compared; both are averaged over 16 gridpoints.



### 3.3.3 Meteograms of temperature

The quality of the temperature forecast can best be demonstrated by considering meteograms of predicted and observed values. In Figure 7, the 120 hour forecast of the 2m temperatures averaged over four gridpoints around the location of the observing site and averaged over 3 days (direct model output). Warm and cold spells are captured as well as the seasonal trend. Note the predicted temperatures below 0°C for a period of 10 days or so at the beginning of February and the verification. Such information five days in advance and the predicted warming afterwards is of extreme value for example to energy, construction, transport or the tourist industry. The meteogram exhibits a tendency of the model to be too slow in capturing extreme changes; instead, it maintains the extreme situations. The dashed curve in the graph is the interpreted temperature forecast  $\hat{Y}$  by just using predictor X.

$$\hat{Y} = a + bX$$

a and b are determined through linear regression and the best relation is chosen. In this case, the predictor turns out to be the 1000mb temperature valid at 96 hours averaged over 16 gridpoints and 3 forecast days (one predictor case). The variance of this interpreted temperature forecast is reduced but, as it can be expected in one predictor case, it follows closely the model forecast.

Figure 8 gives the same meteogram for the day 7 forecast (168 hours). The model errors are clearly growing and especially in January, when northern Germany was snow covered, the model forecast at 2m above the surface is positively biased. Using the one predictor equation with the 1000mb temperature valid at 108 hours, the bias is reduced but it cannot improve on model forecast errors in March. As can be seen from Figure 9, averaging over 5 forecast days gives an improvement but the tendency of the model to be slow in predicting major changes in temperature conditions remains.

### 3.3.3 Optimal verification: the multiple prediction case

So far, temperature forecasts have been verified using the best single predictors. In the multiple predictor case, which will be considered next, many predictors are offered to the multiple linear regression program (see Appendix) and several optimal equations for up to 4 predictors are obtained. The application of the optimal regression program is documented in Technical Memorandum No. 55 (Grønaas, 1982).

The addition of more predictors will improve the result in terms of explained variance but, with this small sample, it is difficult to obtain stable solutions and it is necessary to test the equations on an independent data set.

# HANNOVER TEMPERATURE FORECAST

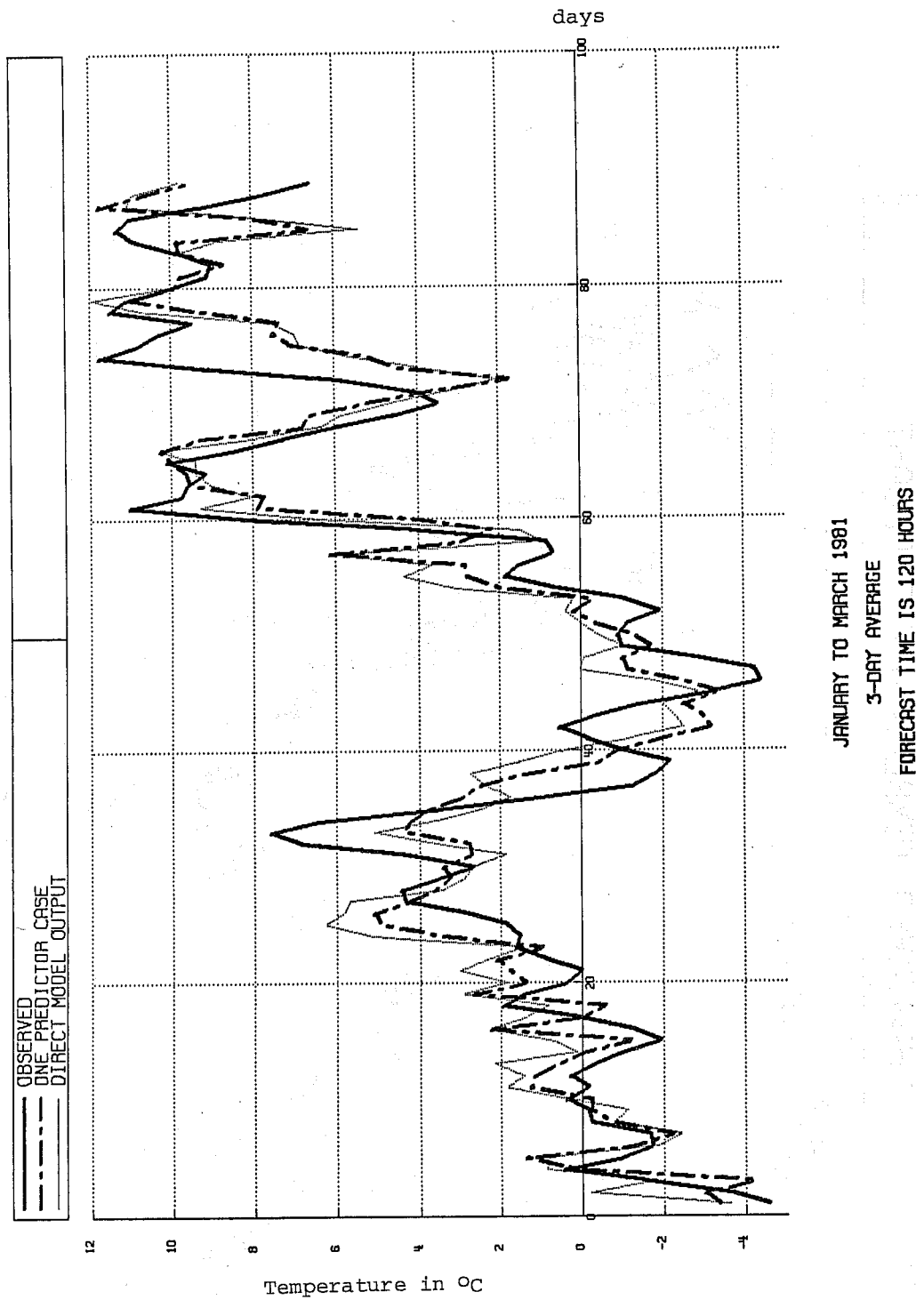


Fig. 7

Meteogram of observed and predicted 3-day mean temperature for Hannover, January to March 1981. The forecast is valid on day 5 and averaged over four gridpoints. See text for further details.

# HANNOVER TEMPERATURE FORECAST

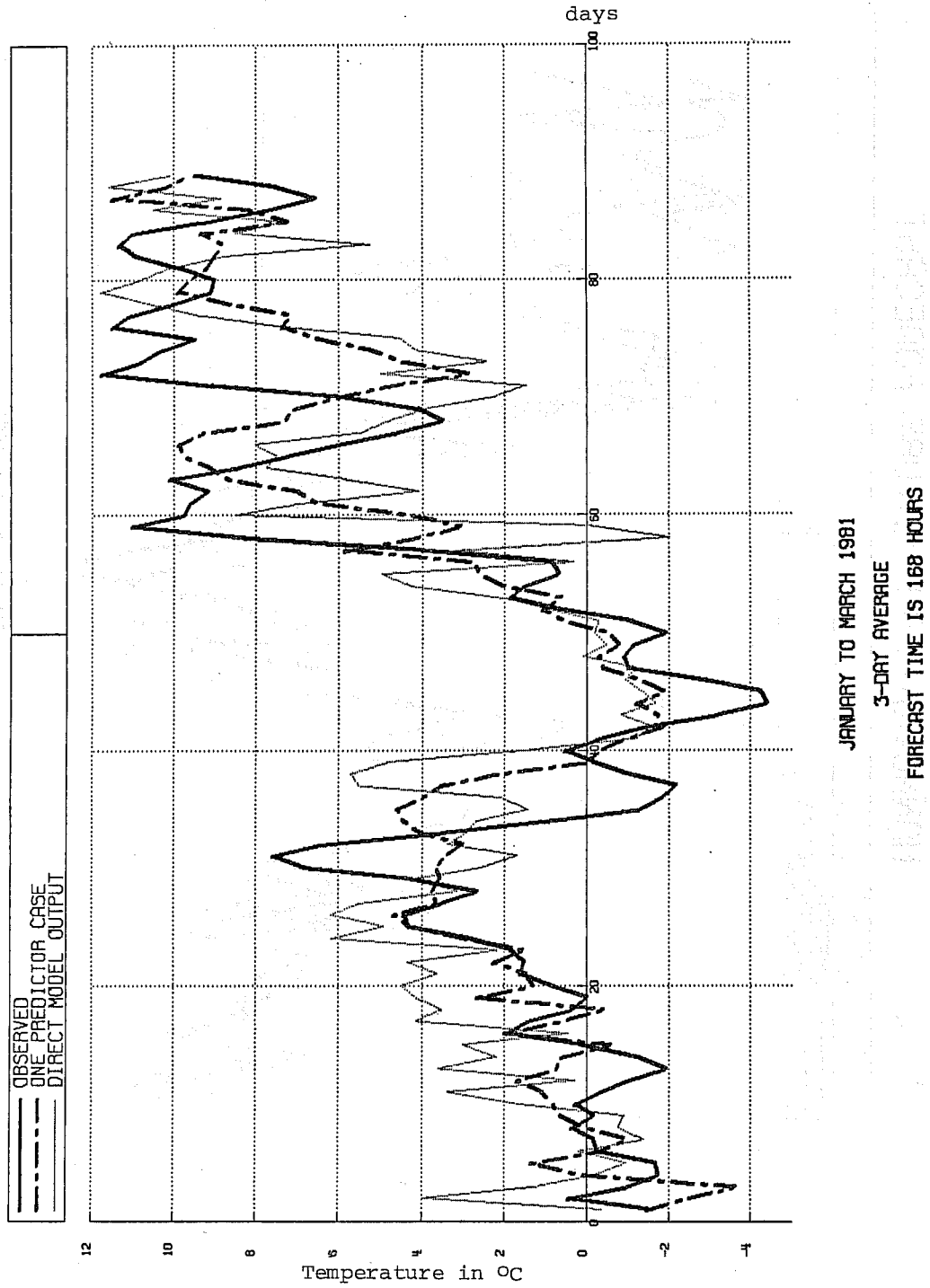


Fig. 8 Meteogram of observed and predicted 3-day mean temperature for Hannover, January to March 1981. The forecast is valid on day 7 and averaged over four gridpoints. See text for further details.

# HANNOVER TEMPERATURE FORECAST

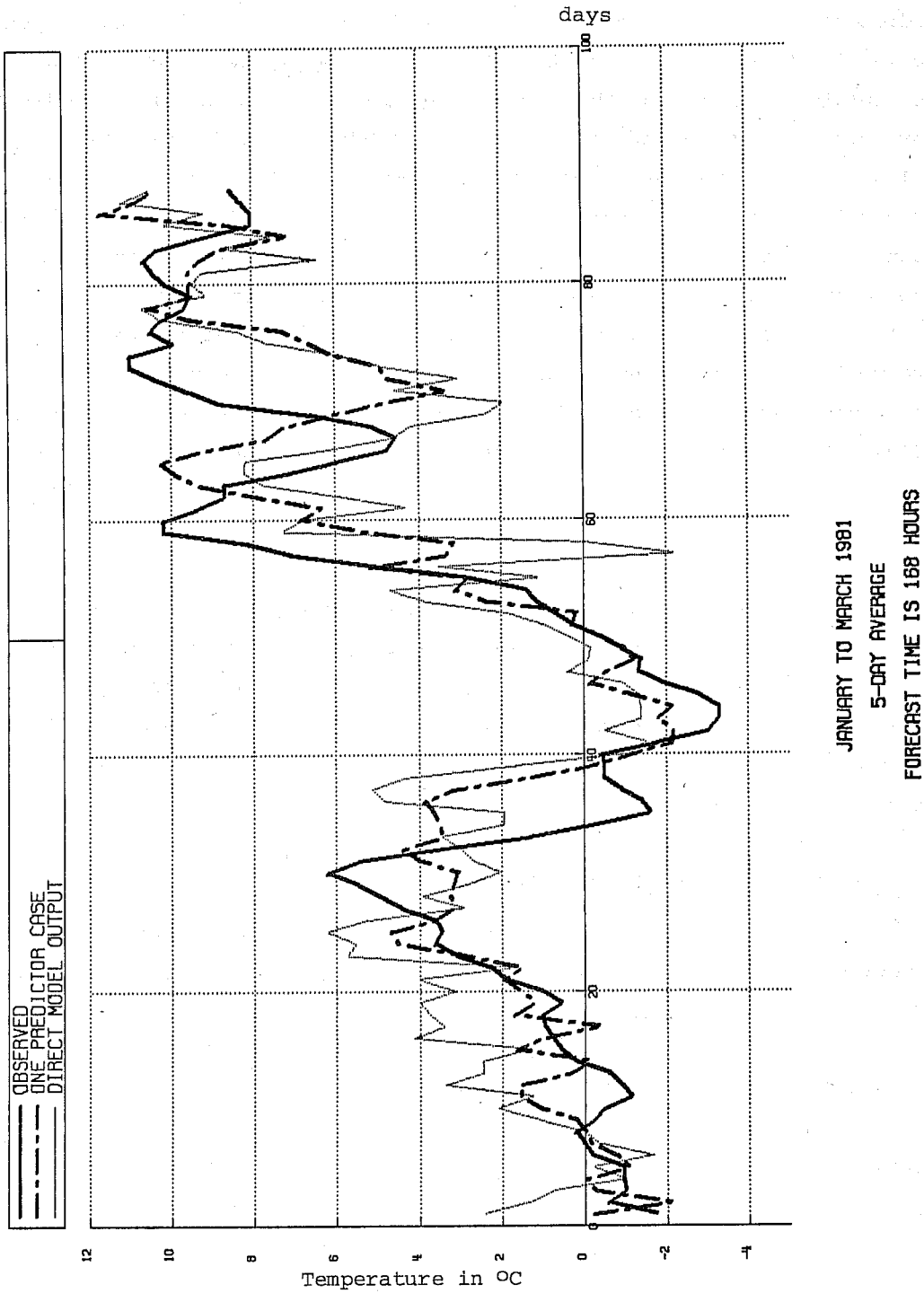


Fig. 9 Meteogram of observed and predicted 5-day mean temperature for Hannover, January to March 1981. The forecast is valid on day 7 and averaged over four gridpoints. See text for further details.

The following example demonstrates what can be obtained by using more than one predictor in our data sample. Figure 10 shows the result derived for Hannover in the verification of the three day mean surface temperature. Four curves are shown, (i) the explained variance of the persistence forecast, (ii) the direct model output of the four point average 1000mb temperature (T at 1000mb), (iii) the best single predictor (T(1000)time off set) revealing, what we have called the model persistence temperature forecast, and (iv) the result using four predictors (RMAX). It is clearly shown that, after day 5, the model persistence gives a better score than the direct model output. The curve for the explained variance using four predictors runs nearly parallel to the curve for the model persistence and only approximately 5% more variance is explained for all forecasts up to nine days. The choice of predictors seems to indicate that the regression equations try to correct for typical flow errors but it is otherwise hard to interpret the results any further. One can look upon the upper curve as a potential that might be achieved using regression methods.

A potential increase of 5% in the explained variance using 4 predictors compared with the simple result using one predictor seems to be modest and indicates that model correction will be difficult to do using regression methods. In more complicated terrain, more than one predictor will often be necessary just to correct for sub-scale influences but, otherwise, good results should be achievable using one or two predictors.

#### 4. SUMMARY AND CONCLUDING REMARKS

In this study, local temperature forecasting in the medium range by the ECMWF operational model has been investigated for two central European stations, De Bilt and Hannover. Linear regression technique has been used to find optimal predictors to forecast the daily mean temperature, the 3-day mean temperature and the 5-day mean temperature. As potential predictors, forecasts of all surface parameters and upper air forecasts of temperature, wind, height and humidity for the pressure levels 1000, 850, 700 and 500 mb have been offered. For De Bilt, only forecasts from the closest gridpoint have been used, while for Hannover, averaged over 4 and 16 surrounding gridpoints have been offered additionally. The data samples were small, comprising just two months, January and February 1981 for De Bilt and three months, including March 1981, for Hannover.

Temperature forecasts from the model's boundary layer were found to be the best predictors for observed surface temperatures. As expected at both stations, the temperature conditions are closely related to the synoptic flow, and just one predictor explains a large fraction of the observed variance of the temperature. The predicted surface temperature and the 1000mb temperature are by

# 3-DAY MEAN TEMPERATURE HANNOVER

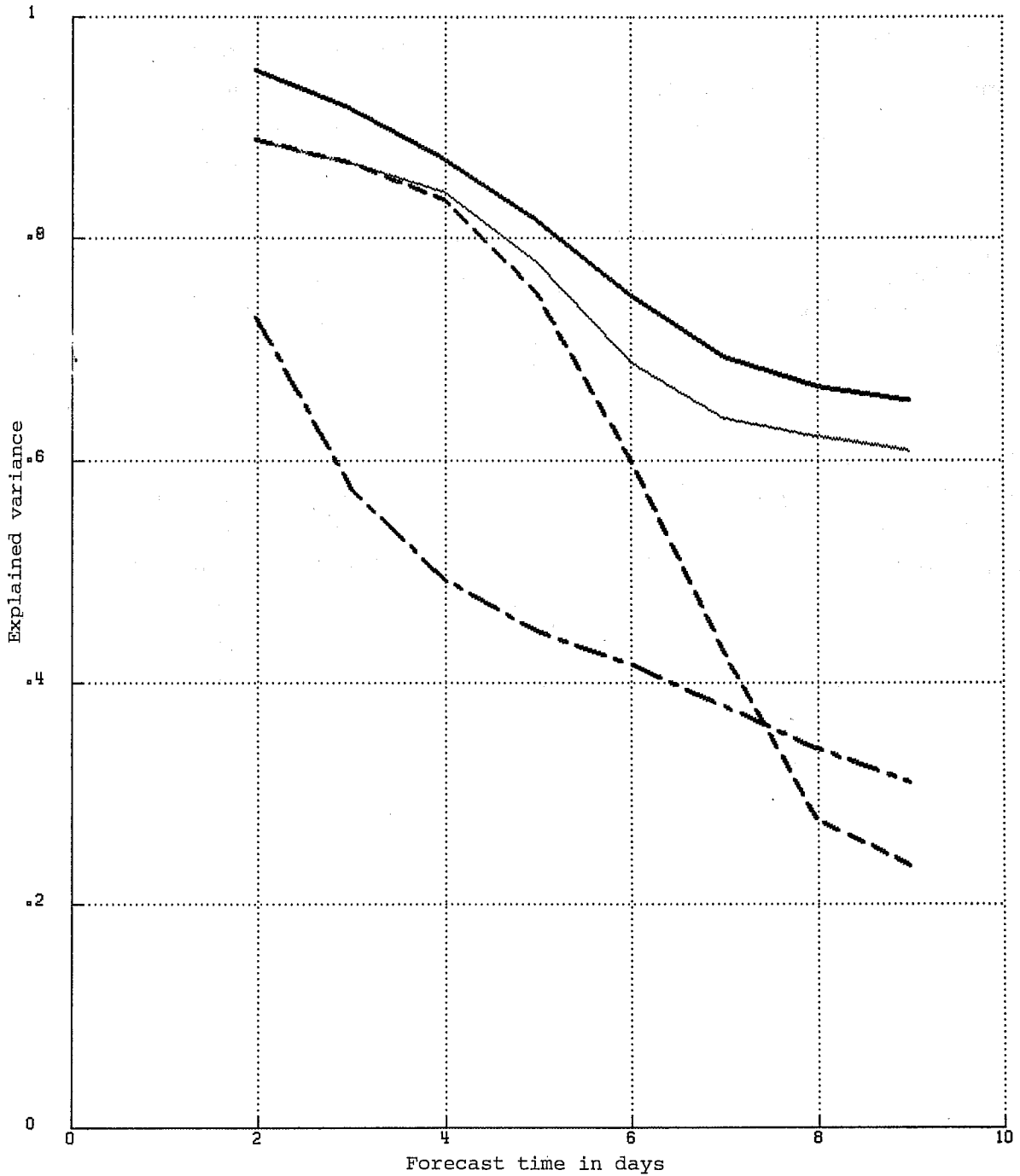
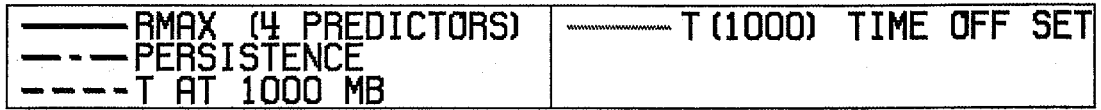


Fig. 10 Optimal verification of 3-day mean temperature forecasts for Hannover, January to March 1981. See text for further details.

far the best single predictors and better than any of the other upper air temperature predictors that were offered. The 1000mb temperature proves to be the somewhat better predictor in the medium range, while the model's 2m temperature has some advantage in the short range. The use of area mean predictors has relatively little effect on the forecast, compared to the benefits from the use of time averaged predictors.

In the medium range, time averaging reduces the effect of phase errors in the migratory waves. The 3-day mean temperature has shown to be a suitable predictand for medium range forecasting, and the model revealed considerable skill in predicting this 3-day mean temperature at the two stations even out to forecast day 7.

For temperature forecasts beyond day 5, best results are achieved by using the model persistence from day 4½ or day 5. These 3-day mean time off-set predictors give reasonably good results even out to day 9. This does not necessarily mean that there is no useful information left in the model after day 5 or day 6. The data sample was not large enough to come to any general conclusions.

Optimal regression equations were derived with up to four predictors.

Little seems to be gained by using more than one predictor for these stations and equations with three or four predictors were difficult to interpret. With data from short time periods, any regression will tend to overfit the data and predictors might just be selected at random and the equations will be valid only for the dependent sample. The verification results for Hannover and De Bilt show that statistical interpretation of ECMWF model output can be kept relatively simple and should mainly be confined to correction of biases and some sub-grid scale influences.

## Appendix A

### The optimal regression method

The starting point for this method is the theory of linear multiple regression analyses which is thoroughly described in textbooks on statistics.

It deals with data containing two or more variables measured for a set of events or set of objects. For example, we may have variables consisting of observed and forecast temperatures at a station and for a set of each day in a season. In general, we have  $p$  variables and  $n$  events, so that the data matrix contains  $n \times p$  pieces of information. The method of least squares is used to establish the statistically best linear relation between one of the variables, called the predictand, and the remaining variables  $x_i$ , called the predictors. From the method, weights are computed giving the relation

$$y = \sum_{i=1}^{p-1} a_i x_i + e,$$

where  $e$  is the residual not explained by the predictors. A multiple correlation coefficient can be computed, giving the fraction  $R^2$  of explained variance of the predictand by the predictors in the relation

$$\text{var}(y) = R^2 \text{var}(y) + \text{var}(e) \equiv R^2 \text{var}(y) + (1-R^2) \text{var}(y)$$

Very often, the number of predictors is high, and we are interested in a small subset which explains an optimal amount of the variance of the predictand. For this purpose, certain screening methods are available and the optimal regression method is one of them. It may be considered as an extension of the most common screening procedure, the stepwise regression procedure.

In the stepwise regression procedure, the best single predictor is first found. Then a second predictor, which together with the first predictor, explains the highest amount of the variance, will be picked out. The procedure will then continue to find a third predictor which, together with the first two, gives the best result and so on, until a regression equation of  $n$  predictors is found.

In the optimal regression method, all possible linear combinations on one, two, three, four, ... variables are investigated and several of the best relations for each step are picked out. In this way, we may keep the number of predictors on a low level and get the simplest possible relations.



Let, for instance, the amount of precipitation at a synoptic station in a mountainous area be the predictand, and the surface pressure at that station and the surrounding stations be the predictors. The cloudiness at the station itself might possibly be the best single predictor, but the best pair of predictors could be properly scaled east-west and north-south pressure gradients given by the surrounding stations, indicating a wind component towards the mountain slope. This would be found by the optimal method, whereas the stepwise regression procedure will keep the cloudiness as the first predictor.

The optimal method will, in addition, give several equations, which may be important when we try to understand the relations. However, the method is, as explained later, relatively expensive.

A computer program for the optimal regression technique was first written by J. Nordø at the Norwegian Meteorological Institute, and the application of the method is described by J. Nordø (1966). A description of a computer program available at ECMWF is written by S. Grønnaas (1982).

The theoretical foundations may be found in textbooks on statistics, for instance, G. Udny Yule (1949).

As mentioned, the explained variance, expressed by the square of the multiple correlation coefficient, is computed for each combination of  $1, 2, \dots, n$  predictors, and a certain number of the very best combinations are selected on each level. In order to compute the variance, a recursion formula is used. It gives the impact in terms of explained variance by adding a new predictor. By introducing a notation, which expresses the effect of predictor  $m$  on the predictand 1 when the predictors 2 to  $m-1$  are already taken into consideration, the following orthogonal recursion formula can be written:

$$R_{1,23\dots m}^2 = 1 - (1 - r_{12}^2) (1 - r_{13,2}^2) (1 - r_{14,23}^2) \dots (1 - r_{1m,23\dots m-1}^2)$$

$$= R_{1,234\dots m-1}^2 + (1 - R_{1,234\dots m-1}^2) \cdot r_{1m,234\dots m-1}^2$$

where

$$r_{1m,234\dots m-1}^2 = \frac{r_{1m,23\dots m-2}^2 - r_{1m-1,23\dots m-2} \cdot r_{mm-1,23\dots m-2}}{\sqrt{(1 - r_{1m-1,23\dots m-2}^2) (1 - r_{mm-1,23\dots m-2}^2)}}$$

The corresponding regression equations can easily be computed once the correlations are present.

In the program used in this study, combinations of 1, 2, 3 and 4 predictors are investigated and a list of the five best relations at each step is given.

For  $p$  predictors, 
$$\frac{p(p-1)(p-2)(p-3)}{1.2.3.4}$$

different residuals have to be inspected and, therefore, the computer time increases considerably with the amount of predictors. The program allows for 100 predictors and smaller subsets can be selected for each run.

## References

- Åkesson, O., 1981: Evaluation of the ECMWF operational model precipitation forecast in October-November 1980, ECMWF Operations Department Technical Memorandum No. 24.
- Åkesson, O., Böttger, H., Pümpel, H., 1982: First results of direct model output verification of near-surface weather parameters at 17 locations in Europe, Operations Department Technical Memorandum No. 47.
- Grønaas, S., 1982: A program for optimal regression analysis, ECMWF Operations Department Technical Memorandum No. 55.
- Johannessen, K.R., 1982: Verification of ECMWF quantitative precipitation forecasts over Europe from January 1980 to April 1981, ECMWF Research Department Technical Memorandum No. 51.
- Nordø, J., 1966: On empirical deduction of laws from geophysical records, Scientific Report No. 15, The Norwegian Meteorological Institute.
- Pümpel, H., Böttger, H., Geleyn, J-F., 1982: Local and regional temperature forecasting in the medium range derived from direct model output (in preparation).
- Udny Yule, G., 1949: An introduction to the theory of statistics, Charles Griffin & Co. Ltd., London, Ch. 14.