REGIONAL AND LOCAL VERIFICATION OF NEAR SURFACE WEATHER PARAMETERS

H. Pümpel

European Centre for Medium Range Weather Forecasts,

Reading, U.K.

1. INTRODUCTION

The inclusion of most of the physical processes involving radiation, moisture and turbulence, albeit in parameterised form, in the forecasting system presently in use at ECMWF has, together with a reasonable resolution in the planetary boundary layer, made it possible to predict several "weather parameters" directly. Previously, forecasters had in general to rely on experience or objective statistical interpretation methods to derive actual weather parameters as desired by the public or special customers from height, wind and temperature fields. The "weather parameters" referred to are: temperature near the earth's surface (hereafter called 2m temperature), near-surface wind speed ("10m wind"), cloudiness and precipitation. These parameters are available in post-processed form either as charts giving a field distribution at a particular forecast time or as values interpolated to one point and shown in a time-graph for 7-10 days ahead called "meteograms". The way in which these parameters are derived from the model output is described by Louis (1983).

After these parameters had been available in "Meteogram" form for some time, it was considered useful to verify the "direct model output weather forecasts" against synoptic information stored in ECMWF's operational database. A set of 17 locations in Europe (one for each of ECMWF's Member States) was selected, and the interpolated values of these parameters for these locations for analyses and forecasts out to 168 hours at intervals of 12 hours were stored daily from 1 December 1980 onwards. The position of these 17 locations is shown in Fig. 1. An example of a forecast meteogram and its observed counterpart for the station of Copenhagen is given in Fig. 2. One limitation of the present forecasting system becomes immediately evident when comparing the forecast and observed temperature time-graph: at the time of this study a diurnal cycle of solar radiation was not included in the parameterisation package, and consequently, there is no daily temperature cycle near the surface.

The verification has been carried out on a monthly basis for the winter 1980/81 and summer 1981, and on a seasonal basis (3 months seasons) for the winters 1980/81 and 1981/82, and for the summer 1981. In addition to verification using the reports of single synoptic stations, an experiment has been carried

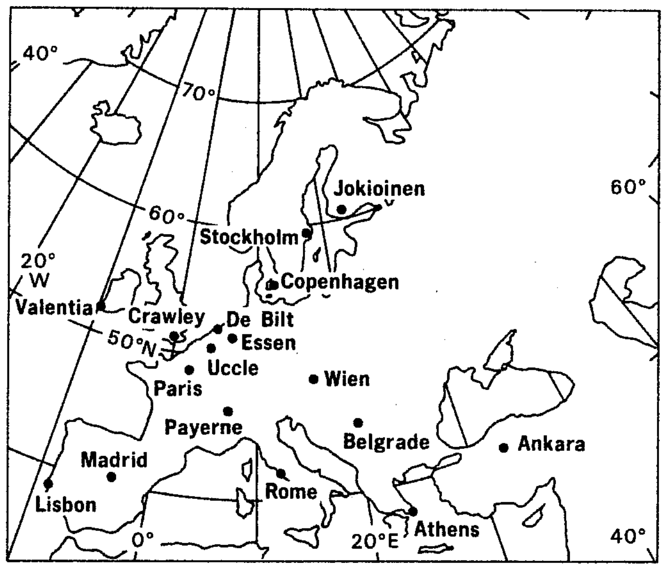

Fig. 1 Stations used in direct model output verification.

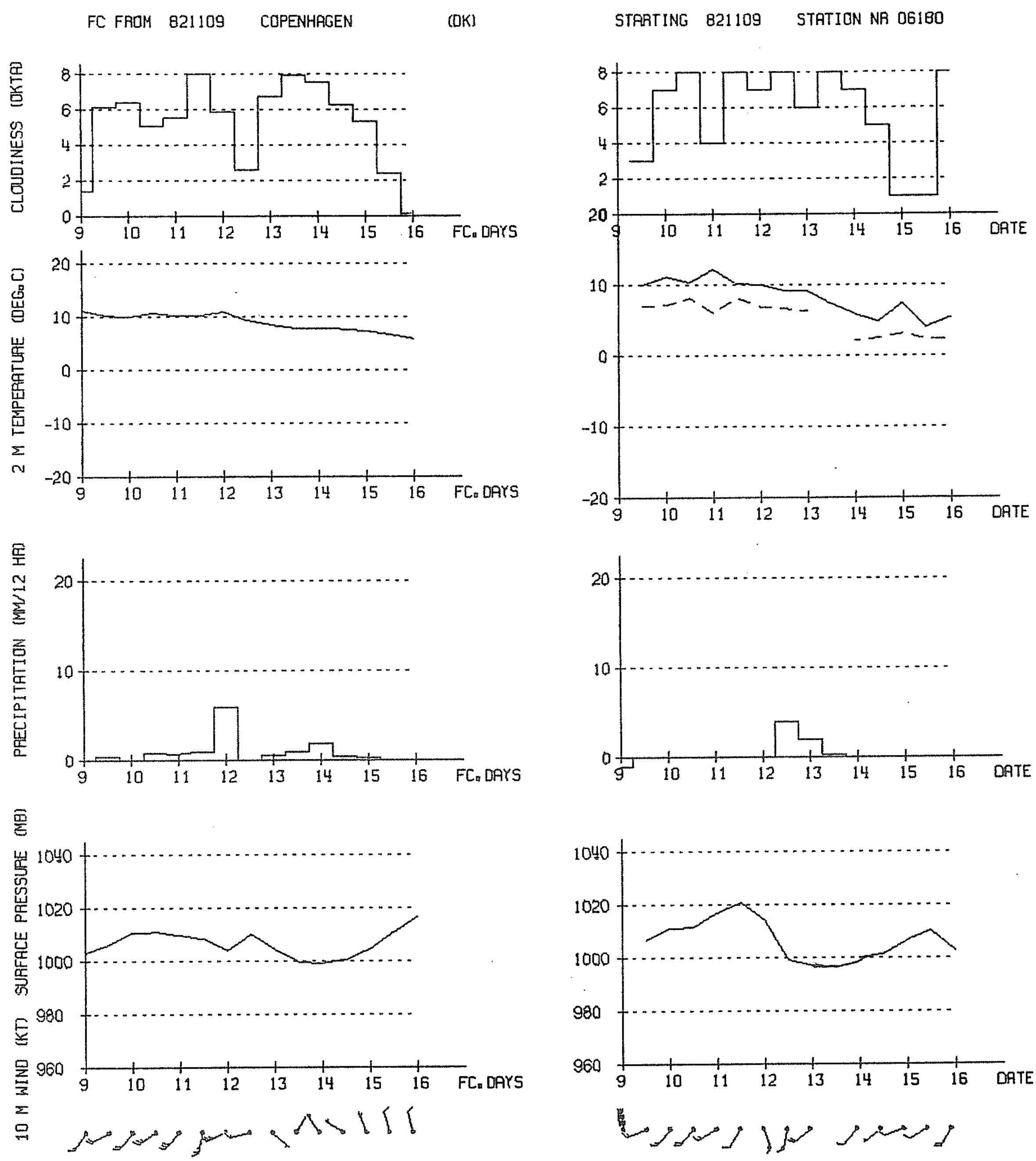

Fig. 2 Example of forecast (left) and observed (right) meteogram for Copenhagen.

The scoring systems for a parameter was chosen to reflect the characteristic behaviour of this parameter. Correlation coefficients can be considered meaningful for smoothly varying fields like temperature and wind, but are rather meaningless for typically discontinuous parameters like precipitation and cloudiness. The calculation of mean errors and standard deviations separately allows both random and systematic errors of the forecasts to be considered and it was hoped that this would help modellers to find the causes of deficiencies. All verified parameters, values of the parameters or deviations from the sample climatology (in the case of 2m temperature) were categorised according to climatologically meaningful limits, and contingency tables for the total of all stations and for ensembles of stations were constructed. Heidke skill scores were calculated against control forecasts of chance and persistence but these scores should only be used to compare skills over different months, seasons or ensembles of stations as their absolute values depend heavily on the number and boundaries of the categories selected.

It may be helpful to consider the possible limits of skill to be expected from a global model, as used at ECMWF, in local forecasting before assessing the relative merits of the forecasts under consideration. These limitations for the assessed parameters could be summarised as follows:

a. 2m temperature: No diurnal cycle of radiation was included. Soil conditions and snow cover start from climatology, but are altered during the forecast. Consecutive interpolations and the limited resolution of the land-sea-mask used lead to some smoothing of land-sea contrasts, and model topography does not accurately reflect local conditions especially, for example, near high ground or over irregular terrain.

b. 10m wind: Again the absence of a diurnal cycle of radiation in connection with an alternating mixed or stable boundary layer is noticeable, and local topographic conditions affecting the roughness length are not reflected.

c. Cloudiness: As already mentioned, the time and space scale of synoptic cloudiness observations in partly cloudy situations is too small to be compared realistically to the model output. Comparing the fairly modest scores obtained from this verification with subjective evaluations of the cloud pattern in forecasts and satellite images, the true skill of cloudiness forecasts appears to be higher than indicated in the objective scores. Digitised cloudiness information derived from geostationary satellites would be very helpful for a fair and objective verification, but this was not available for the present study.

d. Precipitation: Rainfall climatologies show very clearly the dominant influence of local topography on precipitation amounts and, consequently, we concentrated on stations in fairly homogeneous terrain. Interpolation from four surrounding gridpoints (only one of which may have a non-zero forecast of precipitation) to specific sites apparently enhances a tendency to produce small amounts of precipitation too often, and the inherent randomness of convective precipitation renders purely dynamical forecasts for specific sites practically impossible in the case of scattered observed precipitation. More skill, however, could be expected in the case of widespread frontal precipitation or large-scale convective systems, e.g. in cut-off lows.

3. Results

3.1 2m temperature

Single station verification on a monthly basis for the winter 1980/81 revealed large differences in skill for different stations, the importance of mean errors and the generally very clear advantage over persistence as a standard of comparison. The examples given in Figs. 3 and 4 show some of the typical features for specific climatic regions for January 1981. The northernmost station Jokioinen (Finland) suffers from high standard deviation of error from the analysis onward, and a rapid decline of correlation coefficient with forecast time. Coming further south and towards a more maritime climatic regime, the errors become gradually smaller for Stockholm, Copenhagen, De Bilt and Valentia. At the same time, the trend to larger negative mean errors, which are growing with forecast time, found at Jokioinen and Stockholm is reversed and systematic positive mean errors are found at the more maritime stations like Valentia (Ireland) or De Bilt.

These findings seem to be in agreement with the results of objective field verification (Nieminen, 1983) where a trend to a southward shift of the polar jet is found over Europe in the later stages of the forecasts.

Central European stations show very good results (Essen, Vienna) with the exception of Payerne (Switzerland), where extremely large mean errors are probably caused by a difference between the real topographic height of the station and its height in the model's orography (Fig. 4).

The verification for southern European stations is less clear cut, with different error characteristics for each station. Large mean errors are found for the two high-level stations with a continental character, Madrid and Ankara, whereas Athens or Rome feature a more regular behaviour. The mild maritime character of Lisbon's winter is revealed in the very low figures for the standard deviation of error both for the forecast and persistence, only affected by a rather large, but constant mean error probably related to the effects of climatological sea-surface temperatures (Fig. 5). For the ensemble of all 17 Stations, a warm bias of the forecasts can be detected from contingency tables of temperature deviations from the sample climatology (Tables 1 and 2) which is increasing with forecast time. This bias is most pronounced around the middle categories (near normal temperatures). At 24 hours forecast time, 92 cases of near normal temperature are forecast correctly, whereas 76 are forecast as rather warm and only 25 as rather cool. At 72 hours (Table 2), this deteriorates to 58 correct forecasts with 102 one class too warm and only 20 one too cool. The numbers in the top right and bottom left corners of the table, however, representing totally wrong forecasts, remain very low throughout. Heidke skill scores for December 1980 and January 1981 derived from these contingency tables are given in Table 3, both for chance and persistence as standards of comparison. Positive skill can usually be expected up to about 120 hours, with generally higher scores against chance. This is mainly due to the fact that the "chance" table is derived from the forecast table and reflects the model bias, whereas persistence is free of bias.

In order to reduce the effect of local topographic conditions, a verification study has been undertaken on a seasonal basis using observational values averaged over 4-5 neighbouring stations around the interpolation point covering about 1-2 grid squares. For this, we concentrated on 5-day forecasts and, taking into account the inherent tendency to minor phase errors at this forecast range, 3-day mean temperatures in forecasts and observations were used (i.e. means of forecast days 4, 5 and 6).

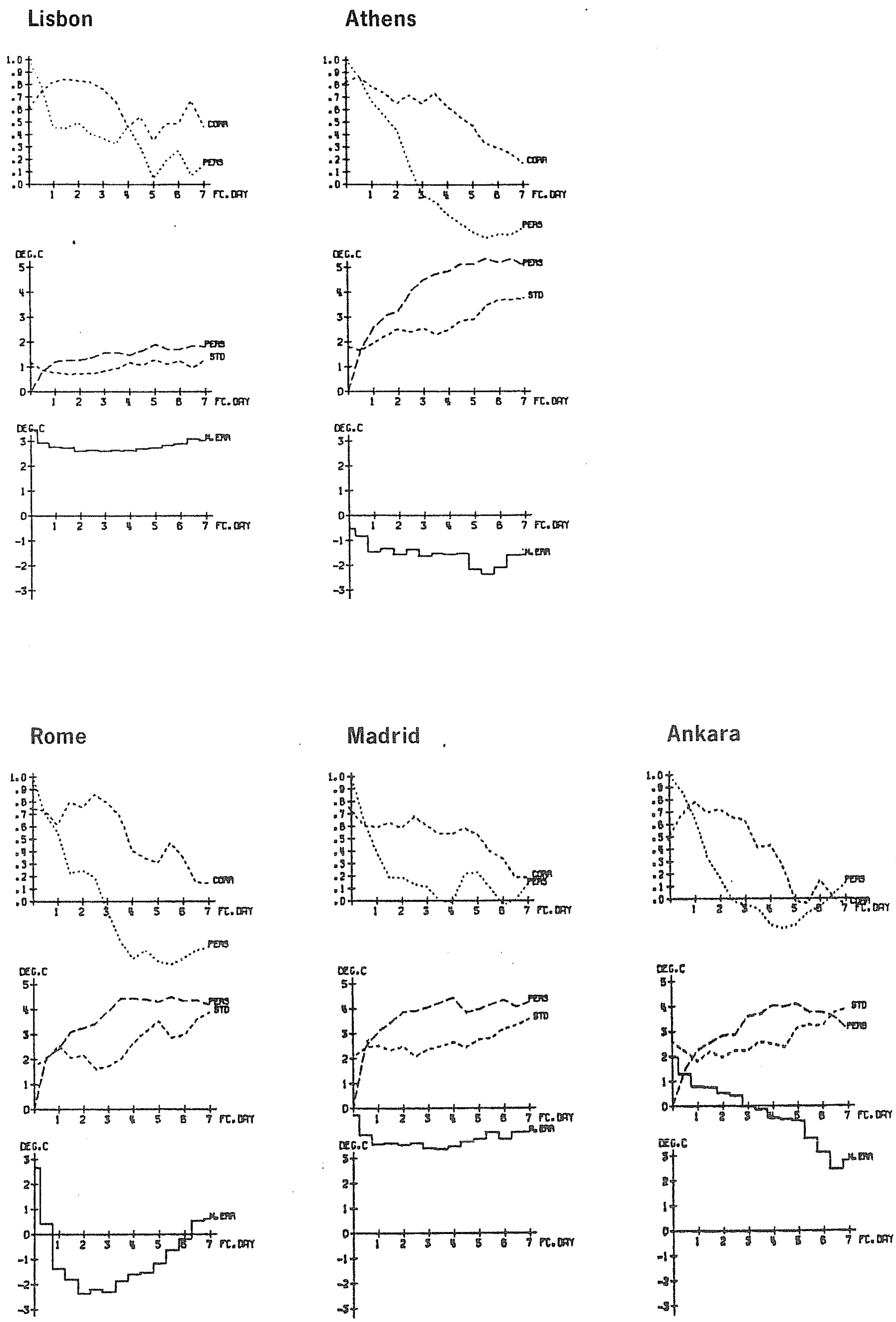

Fig. 5 As for Fig. 3 but for southern European sites.

Forecast time / Standard of comparison	Hours 12	24	36	48	60	72	84	96	108	120	
Chance	.34	.27	.24	.27	.19	.14	.10	.08	.07	.05	December 1980
persistence	-.29	-.01	.13	.08	.18	.12	.09	.09	.09	.07	December 1980
Chance	.24	.19	.17	.18	.13	.10	.11	.12	.07	.06	January 1981
persistence	-.29	.02	.03	.09	.06	.03	.08	.09	.01	-.02	January 1981

Table 3 : Heidke skill score for 2m temperature anomaly forecasts from 12 to 120 hours for 17 stations in Europe, using chance and persistence as standards of comparison

General trends in temperature can realistically be expected to be forecast up to 5 days ahead, whereas the timing of e.g. frontal passages, would be expected to be less accurate.

The experiences gained with the monthly verification results are reflected here insofar as central European stations again show better skill than stations with extreme climates. The presentation of 5-day forecast and observed values in timegraphs over 90 days, however, revealed new aspects of these forecasts. The example for Stockholm (Fig. 6) shows two of the typical deficiencies found at northern locations during the winter 1980/81. The observed values show more pronounced extremes, particularly on the cold side, and a delay of about 1 day can be found in the prediction of some of the extreme events. This problem, however, has been shown to exist already in the analyses of 2m temperature values, see Böttger (1983). More southerly stations show better results, as can be seen in the example of Copenhagen (Fig. 7) or Vienna (Fig. 8). In central Europe, we still find the extremes of temperature underestimated by the forecasts, but the 1-day lag has almost disappeared. Mean errors are still obvious in the results for the Alpine station Payerne (Fig. 9). A simple way to compensate for mean errors which are constant with forecast time can be found by predicting the temperature trend over the next 5 days instead of the absolute values. Figs. 10-11 show a substantial improvement for Copenhagen and Payerne for this verification of temperature tendency.

The results for the summer season 1981 reveal a quite different characteristic. The problems with the forecasts of extremely low temperatures in winter, which were caused both by problems with snow cover and a trend to bring the westerlies too far into the European continent, have practically disappeared. Well-mixed boundary layers and a smaller systematic error of the flow forecasts in summer probably account for this improvement. The forecast and observed temperature tendencies for Payerne (Fig. 12) or Essen (Fig. 13) indicate a very high skill of these forecasts, which is reflected in the correlation coefficients of over .9 for both stations.

3.2 10m wind speed

This parameter has been verified on a monthly basis for the winter 1980/81 for the 17 stations and, for a 3 month period in spring 1982, for the ocean weather ship LIMA at 57N 20W. Local topographic conditions play an important rôle for landstations as expected, and the smaller scale of features in the windfield also contribute to a reduction of the skill that can be achieved by the forecasts. This can be seen in the extremely low skill of the persistence

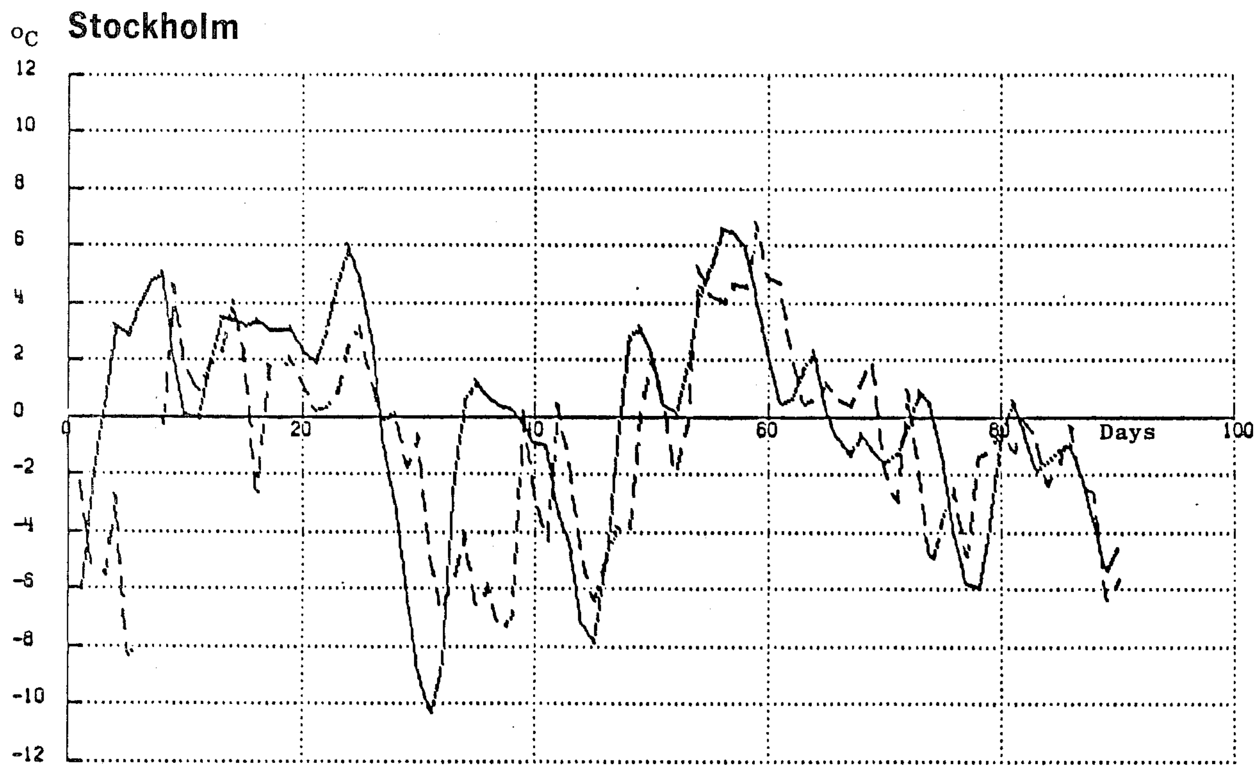

Fig. 6 3 day mean 2m-temperature deviations from sample climatology, centred on 120 hour forecast time for Stockholm, winter 1980-81. (Observed ———; Direct model output ------)

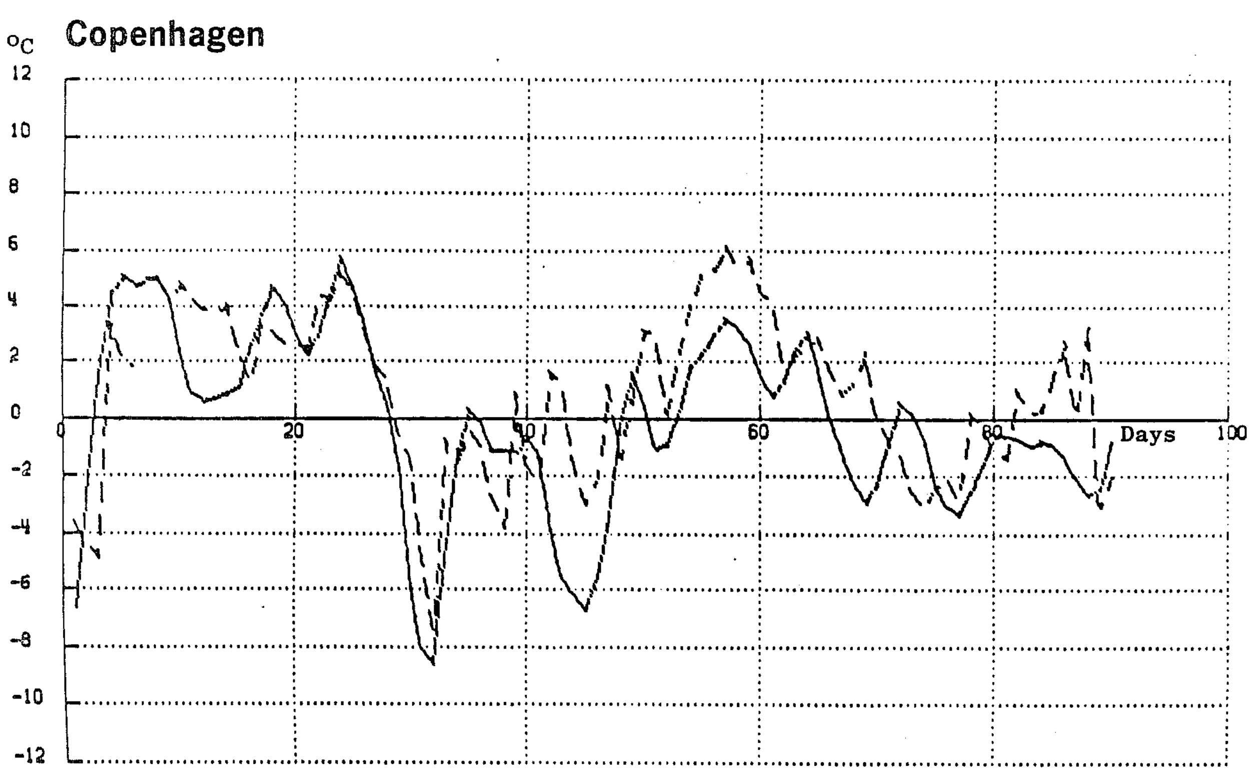

Fig. 7 As for Fig. 6 but for Copenhagen.

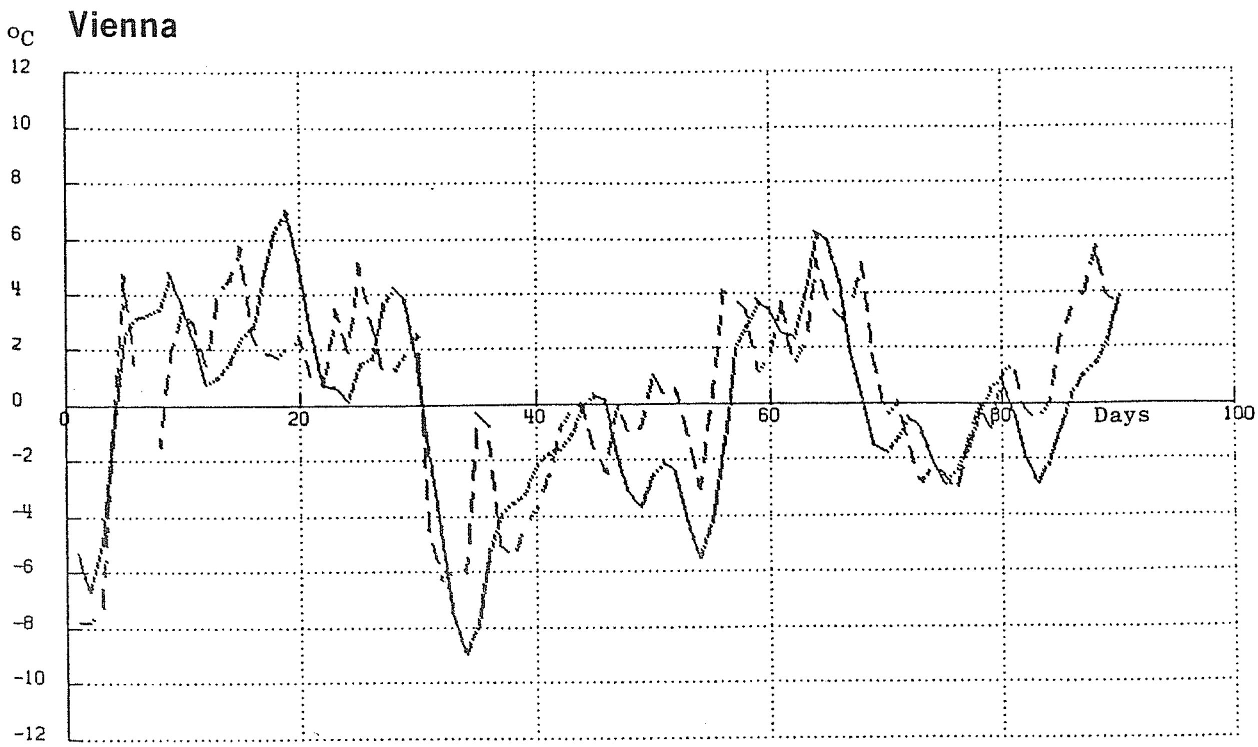

Fig. 8 As for Fig. 6 but for Vienna.

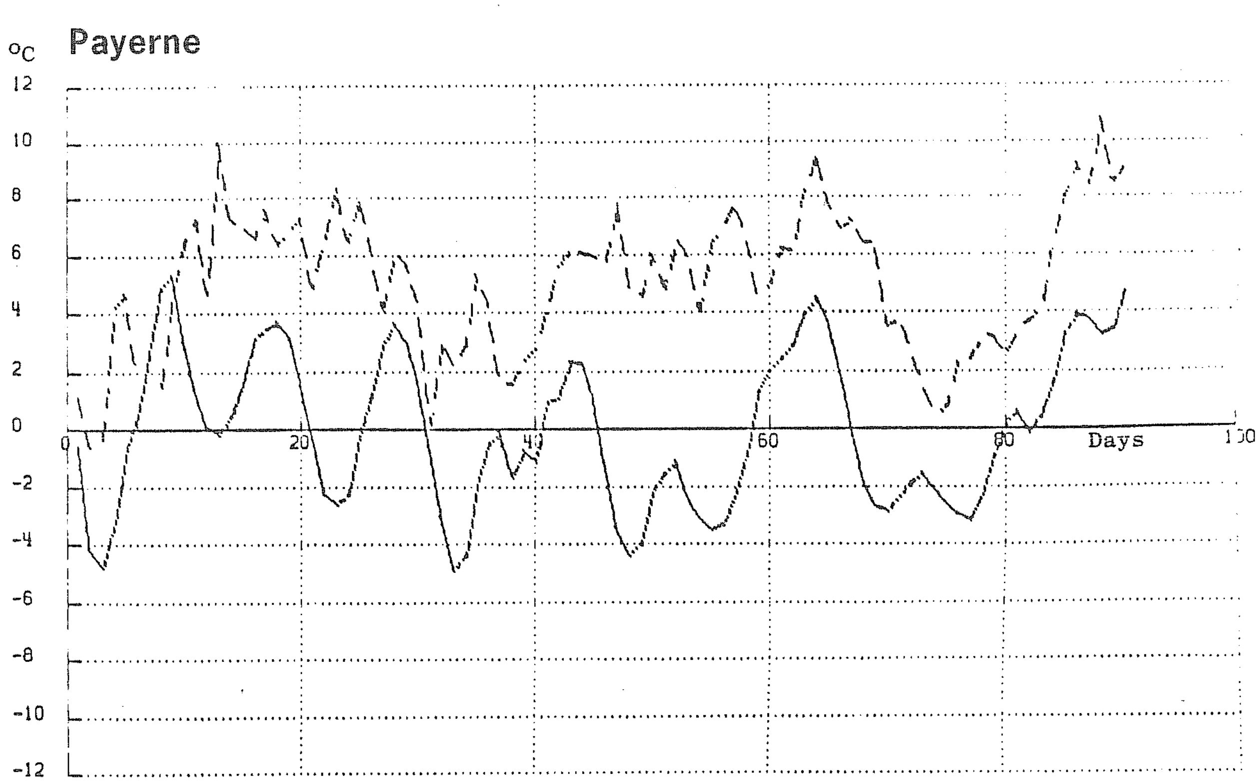

Fig. 9 As for Fig. 6 but for Payerne.

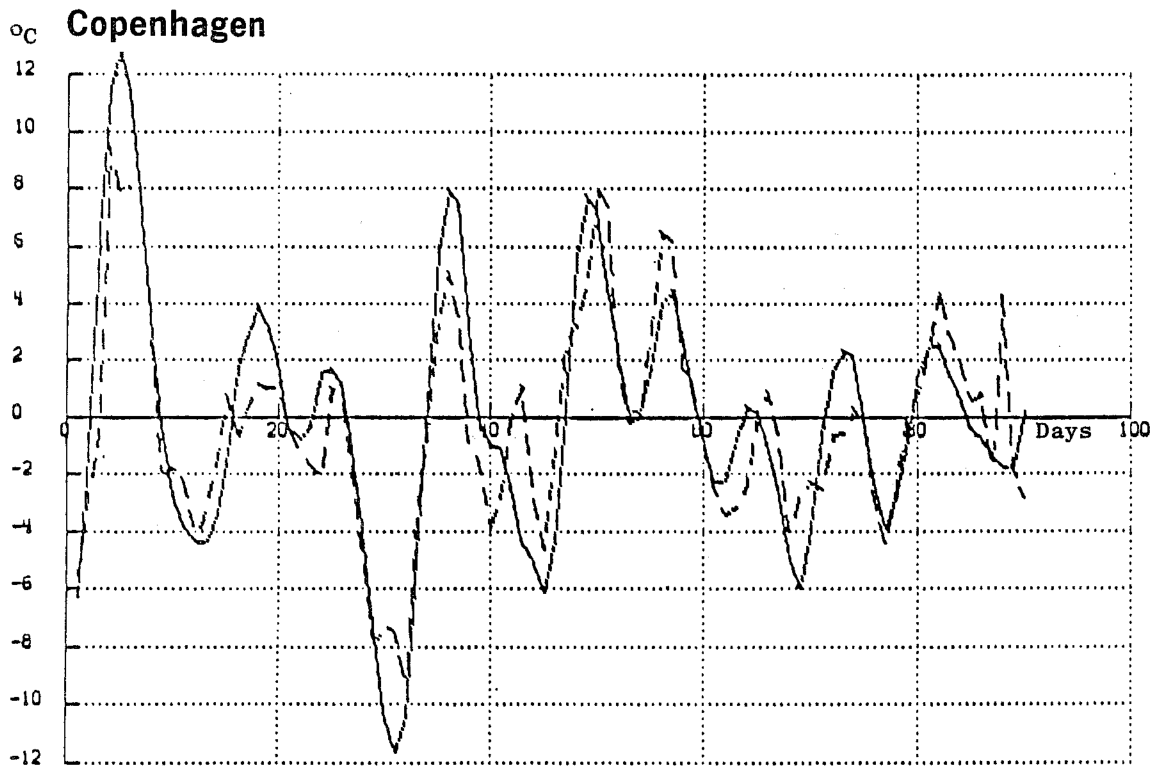

Fig. 10 Forecast and observed 3 day mean 2m-temperature changes to 120 hour forecasts for Copenhagen, winter 1980/81.
(Observed ———; Direct model output ------)

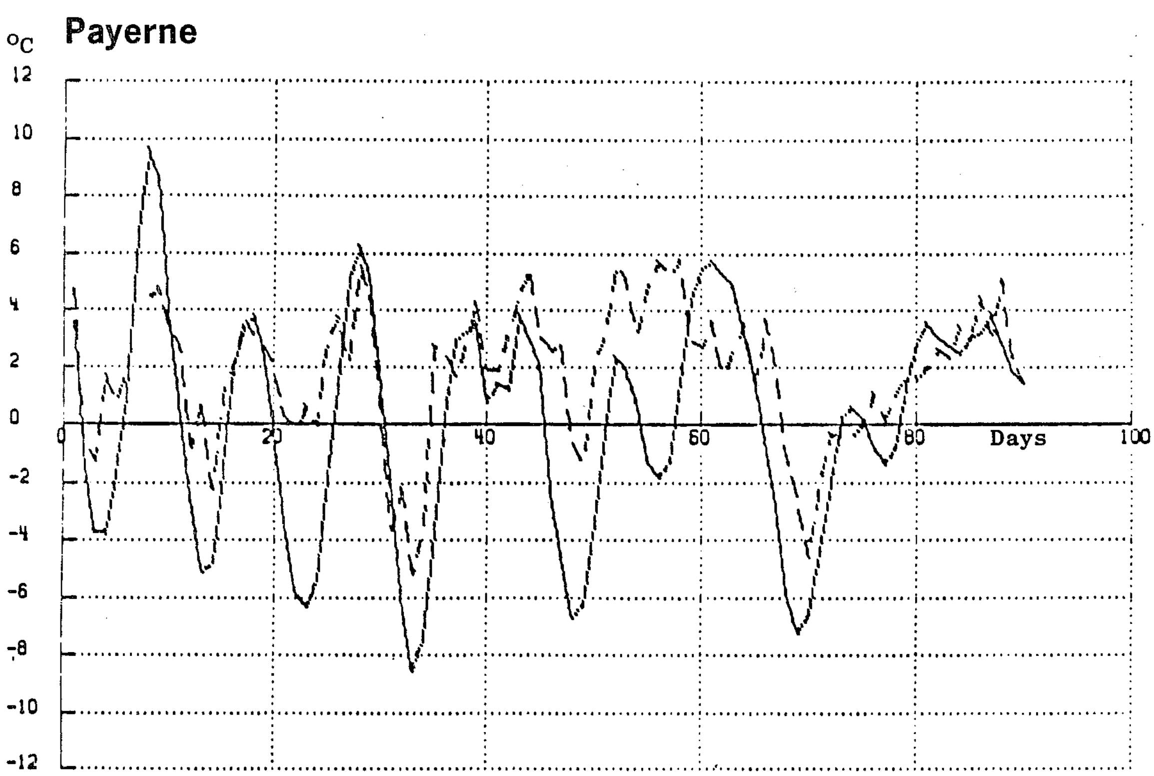

Fig. 11 As for Fig. 10 but for Payerne.

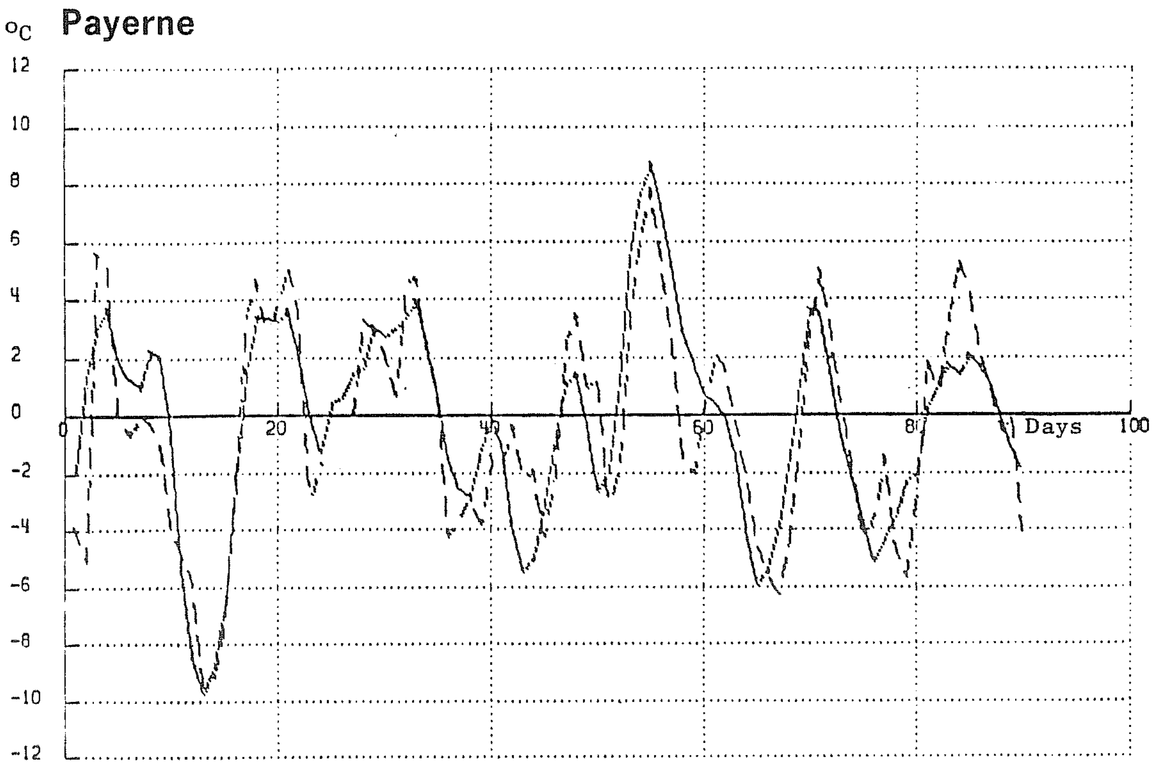

Fig. 12 As for Fig. 10 but for Payerne, summer 1981.

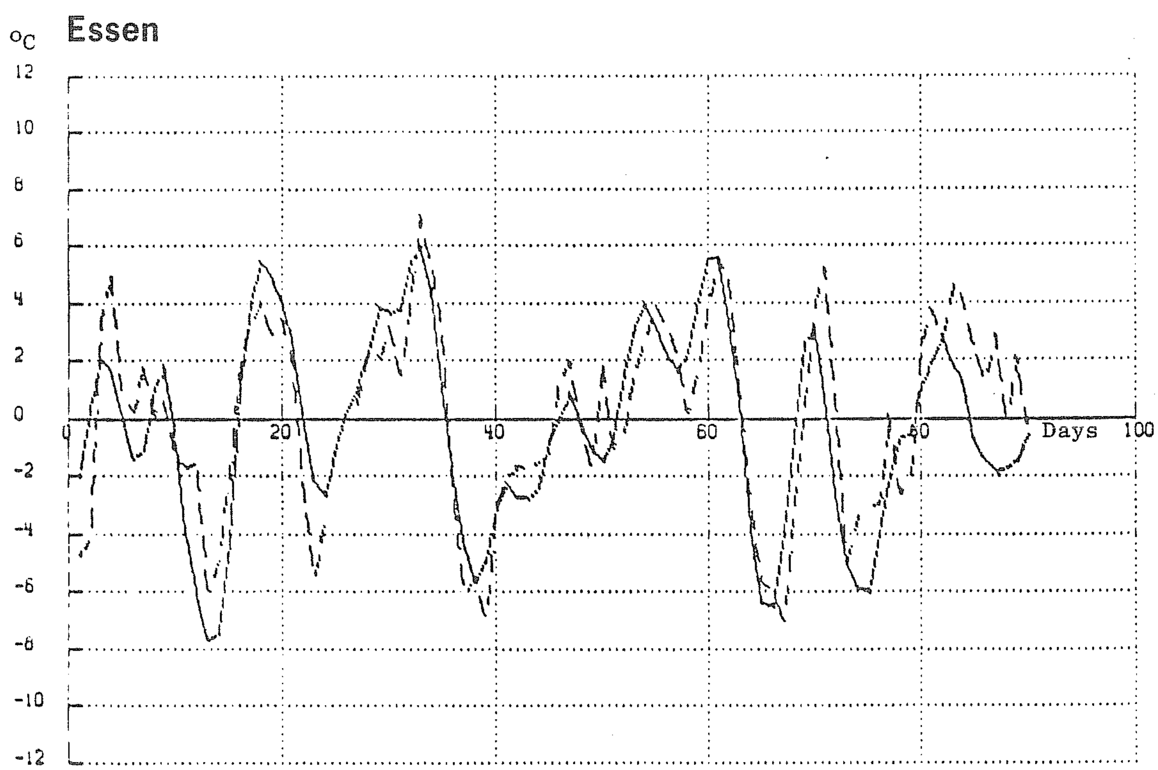

Fig. 13 As for Fig. 10 but for Essen, summer 1981.

forecast, reaching 0% correlation coefficient in some cases after 24 hours (see Essen, Uccle or Paris in Fig. 14. For stations in reasonably homogeneous terrain, quite reliable forecasts can be expected, however, for up to 72-96 hours, whereas for e.g. Belgrade, a limit of usefulness (correlation coefficient ⩾ 0.6) is not even guaranteed for the 24 hour forecast. With very few exceptions, a positive bias in the forecasts of 1-3 m/s is found, with a maximum for western European stations near the model's coastline (Crawley, Valentia, de Bilt) and almost unbiased forecasts for the Scandinavian stations and Wien.

The absence of an diurnal cycle of radiation is again manifested in the alternating values for the mean errors, as the nightly stabilisation of the boundary layer is missing in the forecasts.

In the case of the Ocean Weather Ship LIMA, the problems of local topography and shallow inversions are eliminated, and it can be used to test the model's ability to predict windspeed in the lower atmosphere. Fig. 15 gives a graph for a 4-month period in spring 1982 (missing data are omitted, so that strictly speaking, the abscissa is not a time-axis, but an event axis), for 24 hour forecasts. Peaks in the observed winds (solid line) tend to exceed the forecast peaks (dashed) but good agreement is found in the phase of events, with a few notable exceptions. After 72 hours, however, many events are out of phase (Fig. 16). This phase problem is more evident for the Ocean Weather Ship than for central European landstations, since LIMA is situated on a climatological cyclone-track and rapidly-moving disturbances can pass this area usually at high speed at this time of the year. The typical passing time of a frontal system with peaks in windspeed is in the order of a few hours, and small phase errors in the forecast can affect the results considerably.

The positive forecast bias observed for the landstations is reversed for the ocean verification, and the 1000mb wind field fits the ocean observations better than the values interpolated to 10m above the surface. Part of this effect could be due to the way wind observations are made at ships: anemometer heights are usually around 20m above the surface, and reported wind speeds should therefore be higher than for landstations.

3.3 Cloudiness

A verification in 4 categories has been chosen for this parameter according to a widely used definition in Europe, "clear" is below 1/8, "fair" 1/8 to 3/8, "cloudy" 4/8 to 7/8 and "overcast" over 7/8. This terminology, following that widely used in Europe, is commonly employed in worded forecasts but, for the

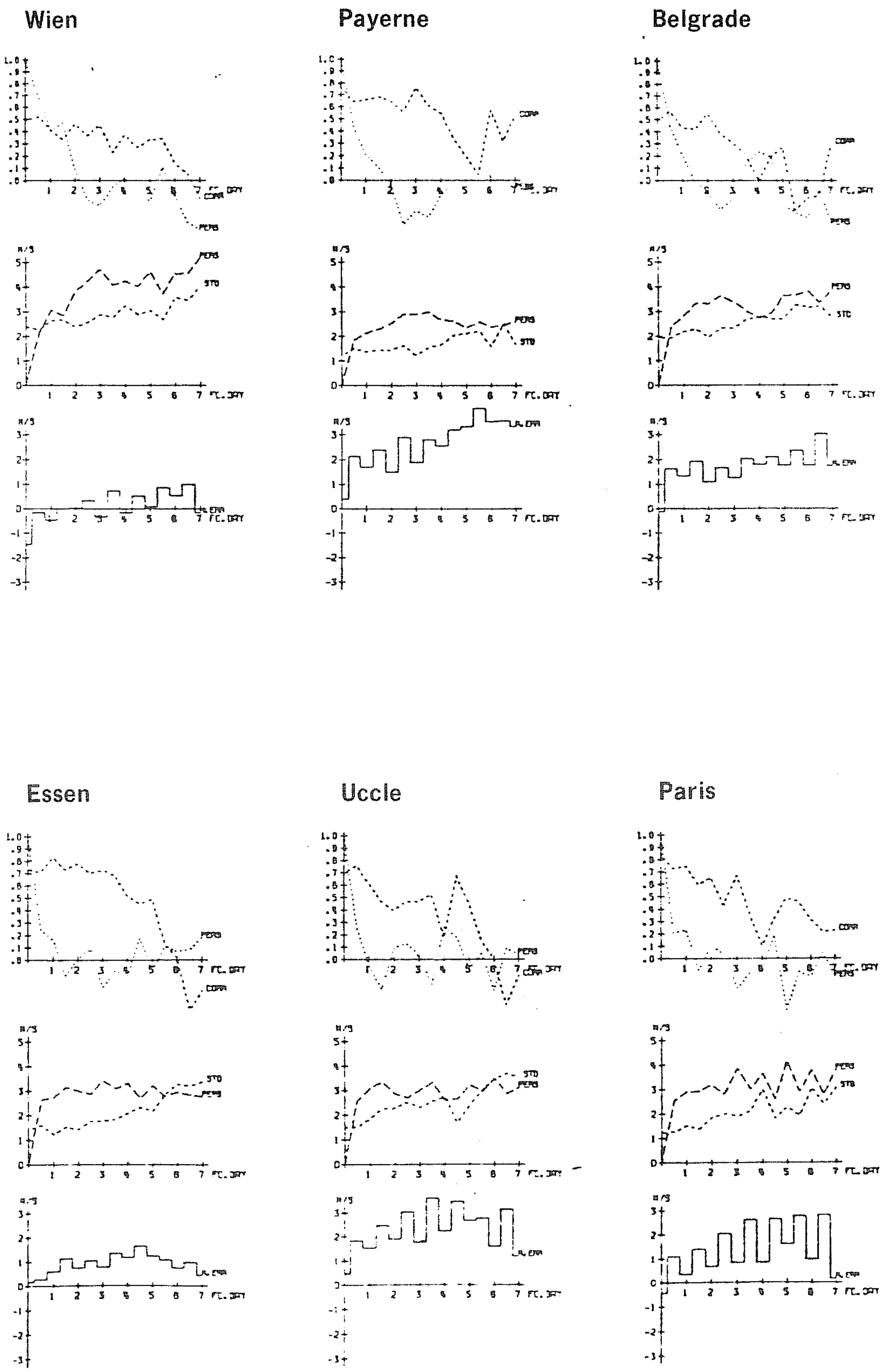

Fig. 14 10m-windspeed verification: correlation coefficient, standard deviation of error and mean error (top to bottom) for forecast and persistence, January 1981.

Lima 24 hour

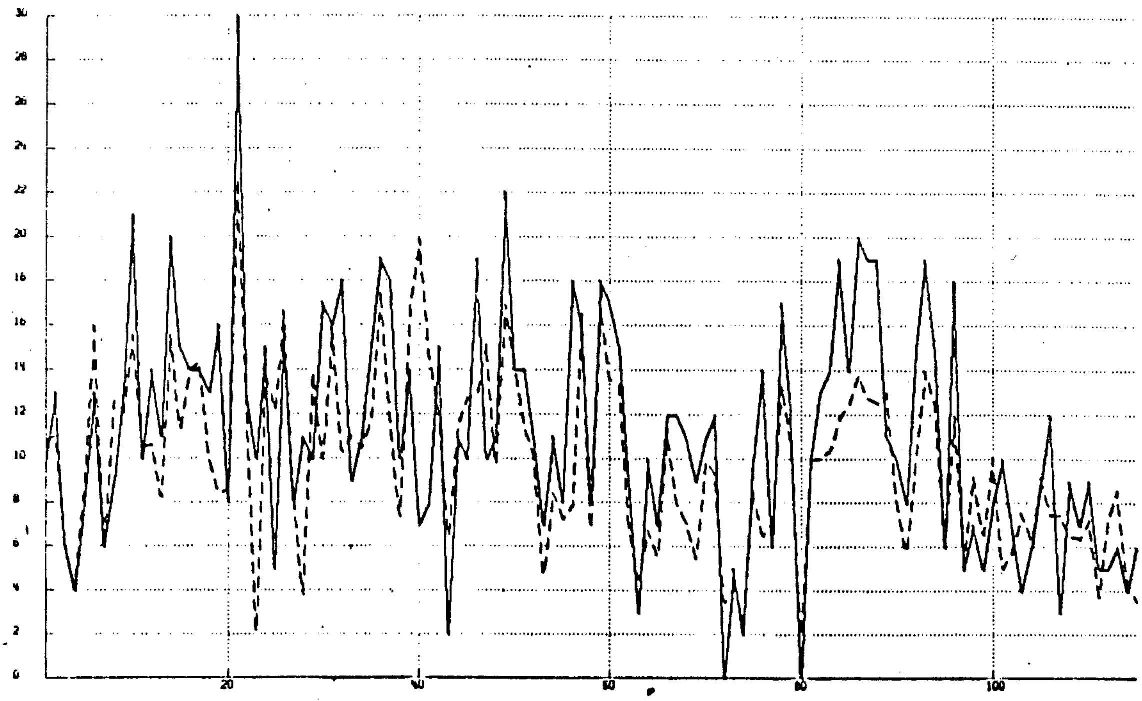

Fig. 15 Forecast and observed 10m-windspeed at Ocean Weathership Lima, spring 1982 for 24 hour forecast time.

Forecast winds (dashed line)
Observed winds (solid line)

Lima 72 hour

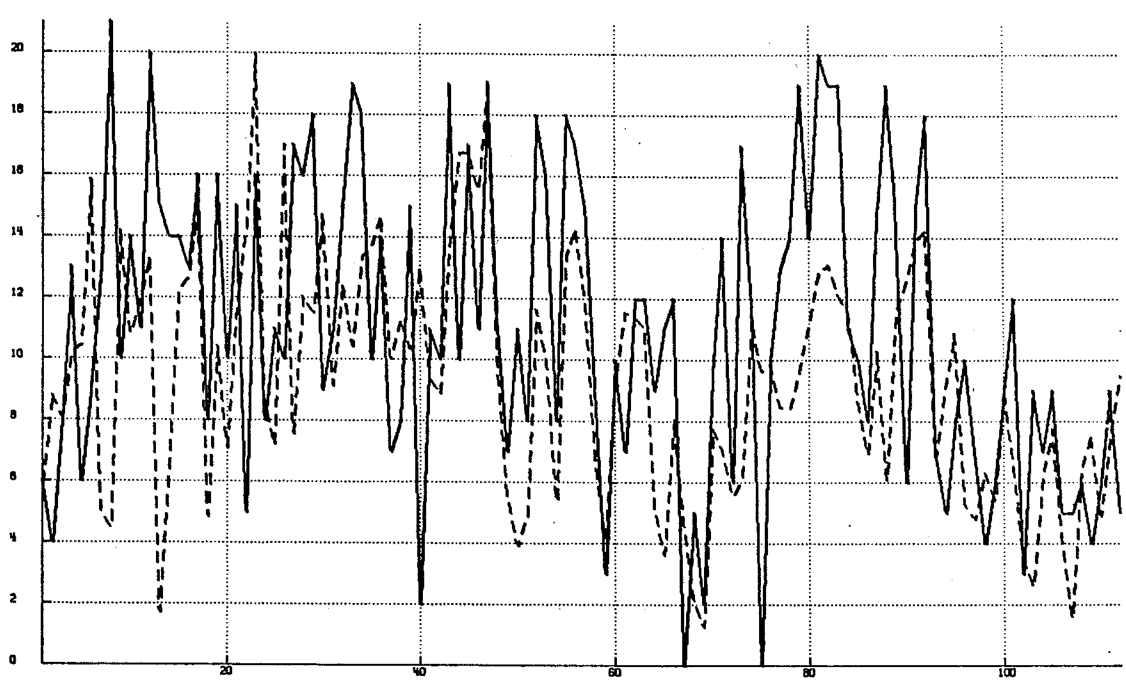

Fig. 16 As for Fig. 15 but for 72 hour forecast time.

test sample, did not produce an equal distribution of cases in the four categories. Central and northern European stations especially, show a large number of "overcast" cases, followed in frequency by "cloudy", "clear" and quite a few "fair" cases.

The model tends to overpredict the number of cloudy cases at the expense of overcast cases (see Table 4) but the sum of categories 3 and 4 is almost exactly the same as for the observed cases. The observed preference for "clear" over "fair" is also not reflected in the forecast.

Most skill can be found, so far only subjectively, in comparing infrared satellite images to model "cloud pictures", as seen in Fig. 17 a,b. This example of a 1-day cloud forecast shows that frontal cloud systems over central Europe, the Alps and the Mediterranean are well predicted, and also the system approaching Ireland is well captured. Note also the thinning out of the cold front over Poland. Digitised satellite cloud information, can be used to corroborate in an objective way the subjective impression of quite adequate skill in forecasting large-scale cloud systems; this was not available for the present study.

3.4 Precipitation

This parameter is very difficult to forecast, and almost as difficult to verify without a dense network of climatological data. Reporting practices of precipitation vary from country to country and, as mentioned in paragraph 2, most reports give 12 hour precipitation at 06 and 18z, whereas this verification database used postprocessed data at 00z and 12z. Precipitation forecasts are also very sensitive to changes in the model formulation. A new and steeper orography, giving overall better forecasts, was introduced in April 1981, but led to unrealistic forecasts of convective precipitation over mountainous terrain. In October 1981, a corrective measure was taken to eliminate this problem to a large extent. We are including here results based on 15 stations for the winter 1980/81 (before the model change) and a second result for an area around Schleswig in northern Germany for the autumn 1981 (after the correction of the precipitation problem).

The forecasts for the ensemble of 15 stations are summarised in the contingency tables in Table 5. A marked tendency to overpredict small amounts of precipitation is found from the 24 hour forecasts onward and increasing with forecast time. At the same time, the "dry" category (≤.2mm) is being depleted. Most skill is found in the light to moderate categories (≤5mm), where about 50% of the observed cases are correctly predicted. For the dry cases, persistence is a very strong competitor, so that the advantage of the model forecast in the

Table 4a

24 hr

Observed

Σ	Forecast	< 1/8	1/8-3/8	4/8-7/8	⩾ 7/8
61	< 1/8	35	6	6	14
62	1/8-3/8	18	10	12	22
175	4/8-7/8	23	18	45	89
209	⩾ 7/8	5	10	45	149
	Σ	81	44	108	274

Table 4b

72 hr

Observed

Σ	Forecast	< 1/8	1/8-3/8	4/8-7/8	⩾ 7/8
58	< 1/8	31	5	6	16
70	1/8-3/8	21	12	9	28
163	4/8-7/8	15	20	42	86
218	⩾ 7/8	18	11	48	141
	Σ	85	48	105	271

Table 4: Contingency tables of cloudiness for ensemble of 17 stations, January 1981. Top: 24 hour forecast time, bottom: 72 hour forecast time.

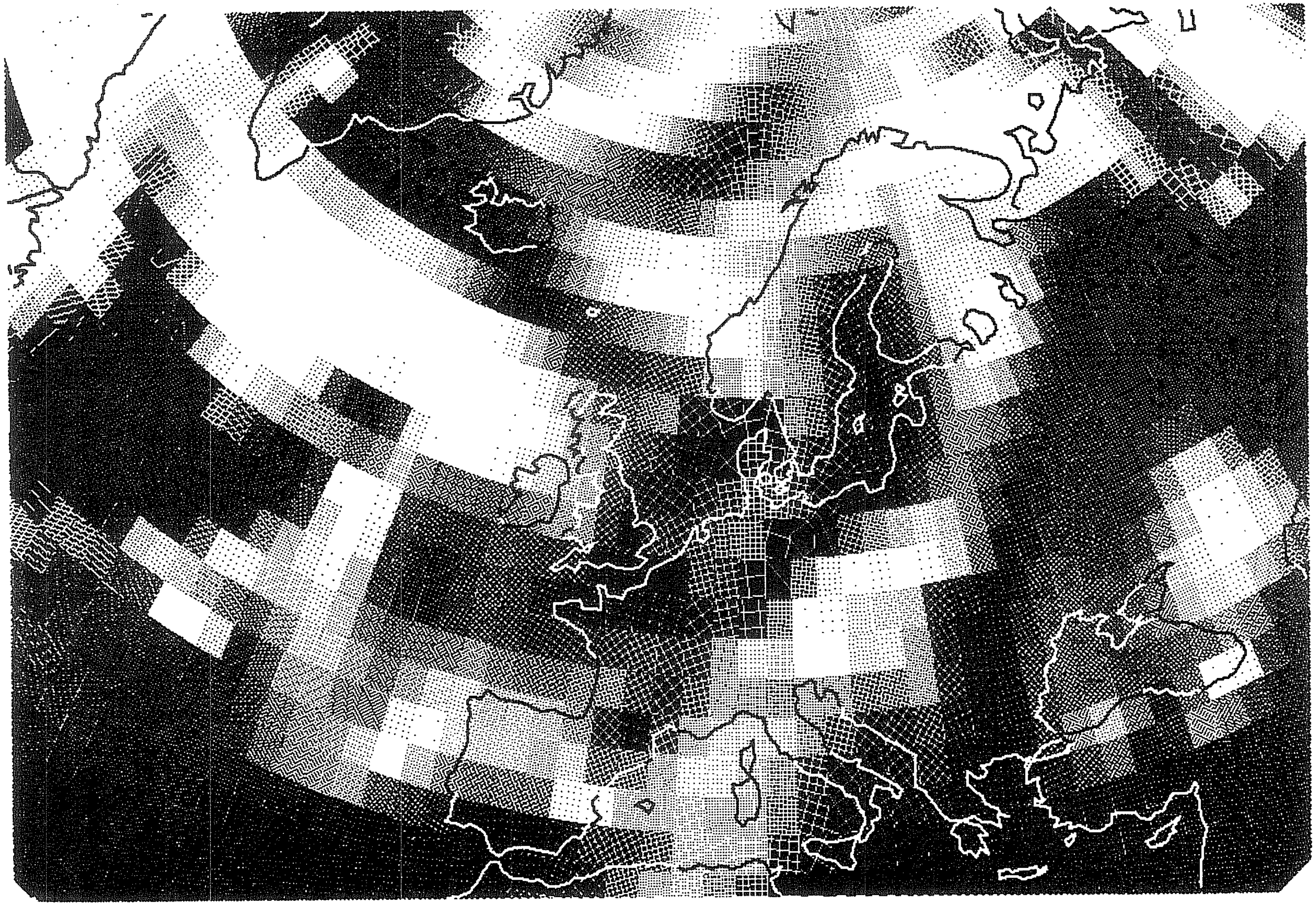

Fig. 17a 24 hour model predictor of cloudiness for 27 August 1982.

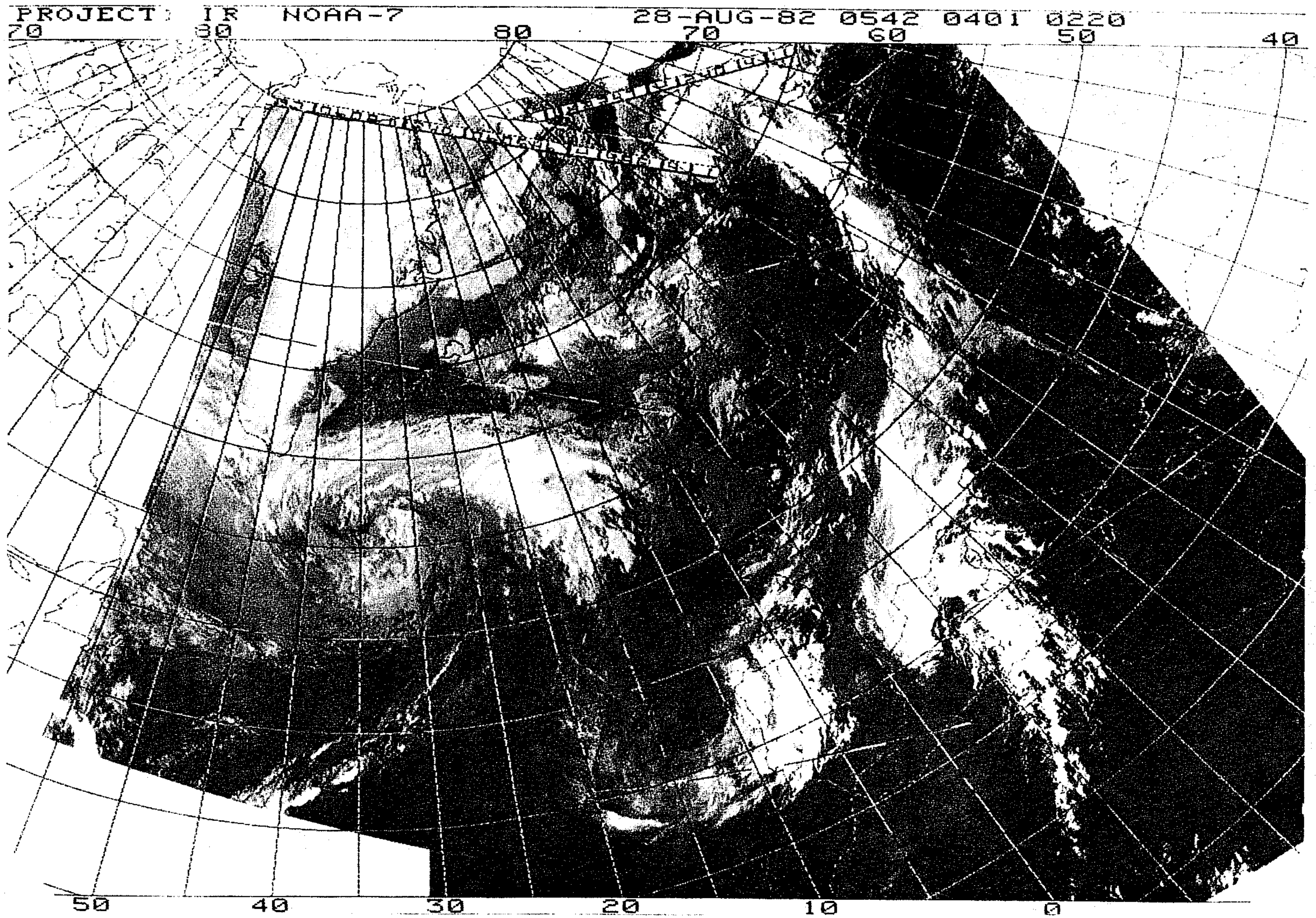

Fig. 17b Verifying IR-cloud picture (NOAA-7) for Fig. 17a.

Model Forecast (rows) vs. **Observed** (columns)

	0-0.2	0.3-2.0	2.1-5.0	5.1-10.0	10.1-
0-0.2	148	14	1	0	0
0.3-2.0	78	30	15	8	4
2.1-5.0	23	15	19	6	1
5.1-10.0	21	7	6	1	1
10.0-	2	1	1	3	2

Persistence Forecast (rows) vs. Observed (columns)

	0-0.2	0.3-2.0	2.1-5.0	5.1-10.0	10.1-
0-0.2	182	39	20	12	5
0.3-2.0	43	7	10	3	0
2.1-5.0	23	13	5	1	0
5.1-10.0	8	1	5	0	1
10.0-	2	1	1	1	1

Table 5: Contingency tables for 72 hour forecasts (top) and persistence (bottom) of precipitation for the ensemble of 15 European stations for January 1981.

middle categories is usually overcompensated in terms of percent correct forecasts.

The data from October-December 1981 are again shown in time-graph form in Figs. 18 and 19 for 48 and 72 hours forecast time. These results indicate better skill for flat, homogeneous terrain in central Europe than do the overall results for the ensembl of stations from all climatic areas of Europe. At 48 hours (Fig. 18), most significant precipitation events during this fairly wet period are predicted with quite accurate timing, although amounts often appear overpredicted. The tendency to produce spurious small amounts in dry situations was not obvious for this very wet observed sample. After 72 hours (Fig. 19), errors become more obvious, main events appear out of phase and overprediction is more widespread. These findings seem to corroborate results from previous studies (Akesson, 1981) that the predictability of single precipitation events at specific sites at the present level of skill will be not more than 72 hours.

4. SUMMARY AND CONCLUSIONS

The results of this first and very limited exercise in verifying so-called "near surface weather parameters" against synoptic observations in Europe and their consequences for future work could be summarised as follows:

- Since "near surface weather parameters" as derived from the postprocessed model output are shown to depend to a large extent on the formulation of the boundary layer and radiation parameterisation scheme, biases have to be expected arising from the discrepancies between actual conditions and the climatological fields used in the model. This is likely to be the case until a full analysis of surface conditions (e.g. soilwetness, snowcover) is implemented.

- For parameters which vary strongly on a small space- and time-scale such as cloudiness and precipitation, and are subject to local topographic influences, the possible skill of a global model with a resolution of approximately 200km^2 is limited. The predictability is strongly dependent on the location and usually does not exceed 72 hours.

- Temperature and wind near the ground are subject to a daily variation depending on the solar radiation. This daily cycle of radiation had not been included in the version of the model under verification and it was therefore decided to verify a daily mean temperature. Apart from problems with extreme climates (northern winters), where low level inversions seem to play an important rôle, considerable skill of forecasting temperature trends well into the medium range has been found, especially for most central European areas.

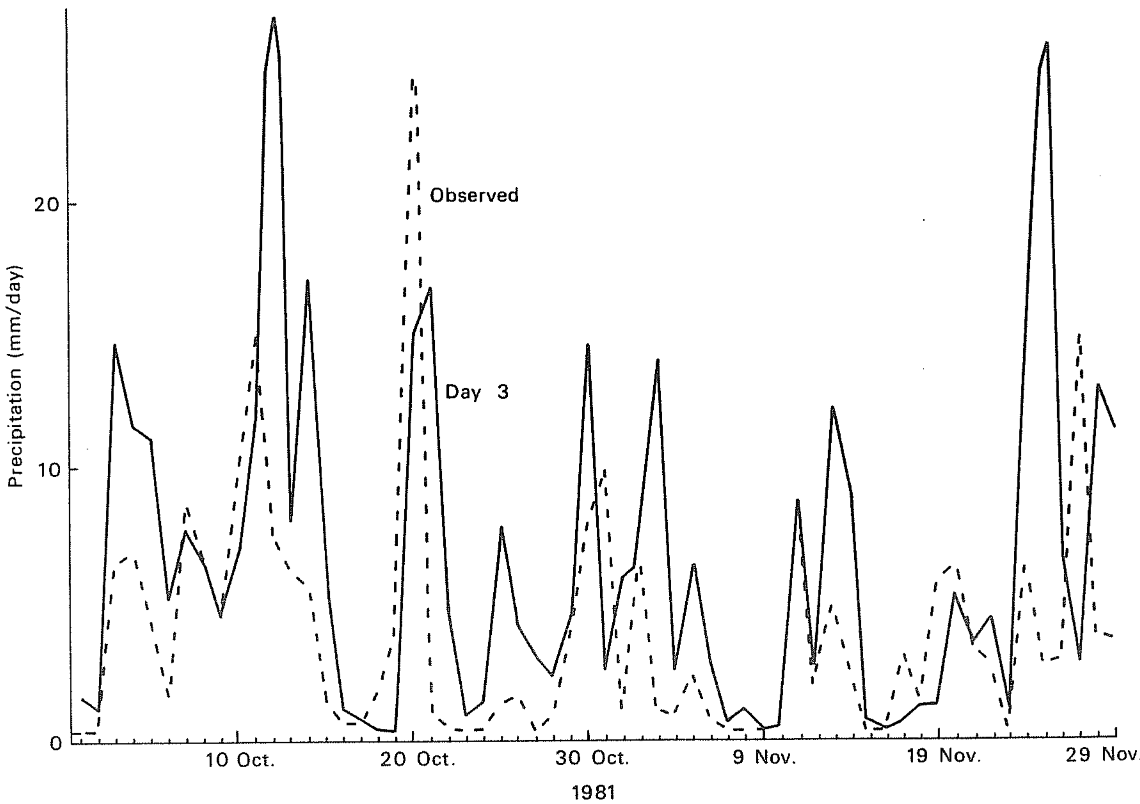

Fig. 18 Forecast and observed precipitation at an area near Schleswig (northern Germany), autumn 1981.

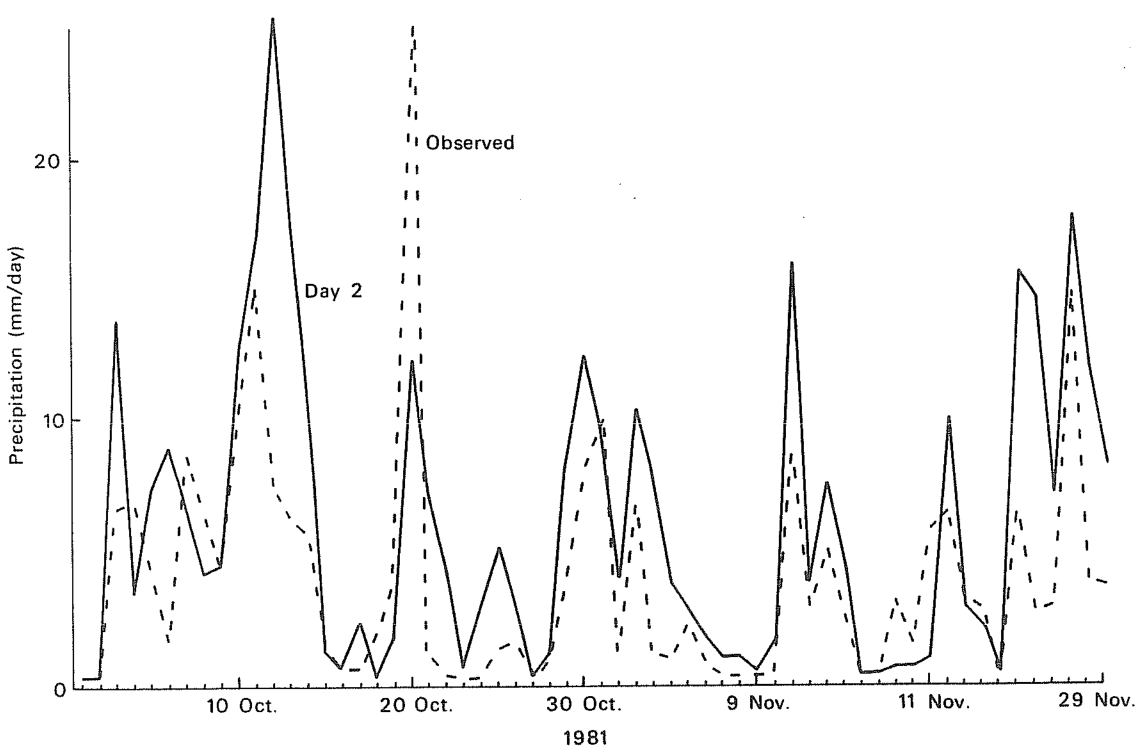

Fig. 19 As for Fig. 18 but for 72 hour forecast time.

- Apart from the benefit to forecasters receiving guidelines for the reliance to be placed in these forecasts, the results of verification of these parameters are useful to modellers too as they reveal some weaknesses in the parameterisation schemes. Used within their limits of reliability, these parameters could give good guidance for forecasters in most areas of Europe.

References

Akesson, O., 1981: Evaluation of the ECMWF operational model precipitation forecasts in October, November 1980. Operations Department Techical Memorandum No. 24.

Akesson, O., Böttger, H. and 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.

Böttger, H., 1982: Local weather element guidance from the ECMWF forecasting system in the medium range. ECMWF Seminar Proceedings, 1982.

Louis, J-F., 1982: Parameterisation of sub-grid scale processes. ECMWF Seminar Proceedings, 1982.

Murphy, A., 1982: Verification of weather forecasts. ECMWF Seminar Proceedings, 1982.

Nieminen, R., 1983: Operational verification of ECMWF forecast fields and results for 1980-1981. Operations Department Technical Report (in print), 1983.

Acknowledgements

I wish to express my thanks to K. Arpe, who kindly allowed me to make use of the results of his verification of precipitation over Germany (Figs. 18 and 19). The efforts of Linda Baker for the typing of the manuscript, Rosemarie Shambrook for the artwork and the Print Room of ECMWF also deserve my sincerest gratitude.