TECHNICAL MEMORANDUM

47

First results of direct model output verification of near surface weather parameters at 17 locations in Europe

O. Åkesson, H. Böttger and H. Pümpel

Research Department

January 1982

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

C O N T E N T

I. INTRODUCTION

Since 1 December 1980 the following additional global fields of near-surface weather parameters are available within ECMWF as direct model output from ECMWF operational forecasts:

(i)	Temperature at 2m	above the model surface
(ii)	Dewpoint at 2m	
(iii)	u and v component of the wind at 10m	

Together with cloud amount and precipiation they form a basis to extract local weather element forecasts from operational model runs which can be made available in e.g. the form of meteograms (examples for Copenhagen and Stockholm, Rome and Paris are given in Figures 1 and 2) and which can be verified against single station observations.

The assessment of near-surface weather parameters will exhibit deficiencies in forecasting boundary layer conditions, and is primarily meant to complement the evaluation of the model performance by standardized methods such as correlation coefficients, RMS scores and bias score. Local weather element forecast guidance could, in principle, be made available directly to the field forecaster. He should use the information not independently, but in connection with and in addition to standard forecast fields like geopotential height, temperature and wind. In this sense meteograms can be regarded as a useful tool to compress the enormous amount of data being produced by each operational forecast.

In chapter V. paras. 1 to 4 we assess ECMWF model forecasts of temperature, precipitation, cloud amount and wind speed for 17 specific sites in Europe for December 1980 and January 1981. The spot forecasts were linearly interpolated from the nearest surrounding four gridpoints (the grid interval being 1.875 degrees latitude and longitude) to the location of the following 17 Synop stations which were used for verification, the WMO code number is given in brackets (see also Figure 3):

Jokioinen	(02963)	Essen	(10410)	Rome	(16242)
Stockholm	(02464)	Uccle	(06447)	Ankara	(17128)
Copenhagen	(06180)	Wien	(11035)	Madrid	(08221)
Crawley	(03776)	Payerne	(06610)	Lisbon	(08536)
De Bilt	(02260)	Paris	(07147)	Athens	(16716)
Valentia	(03953)	Belgrade	(13272)		

These stations report regularly at least at each main synoptic hour. They were selected partly because they are situated in a region of reasonably homogeneous terrain, thus avoiding a priori systematic errors through orographic effects.

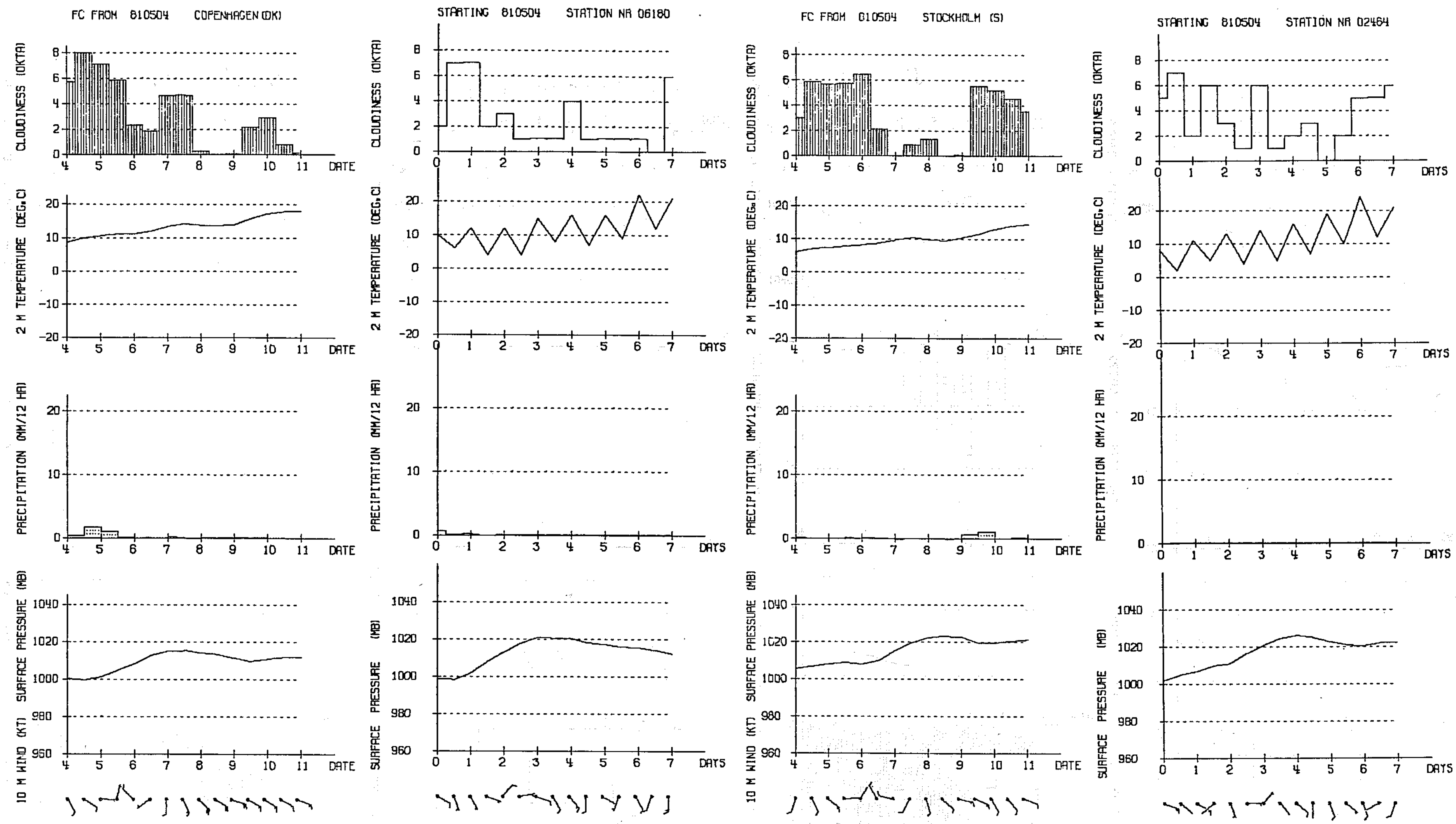


Fig. 1 Local cloud amount. 2m temperature, precipitation, surface pressure and 10m wind forecast out to 168h for Copenhagen and Stockholm and verification

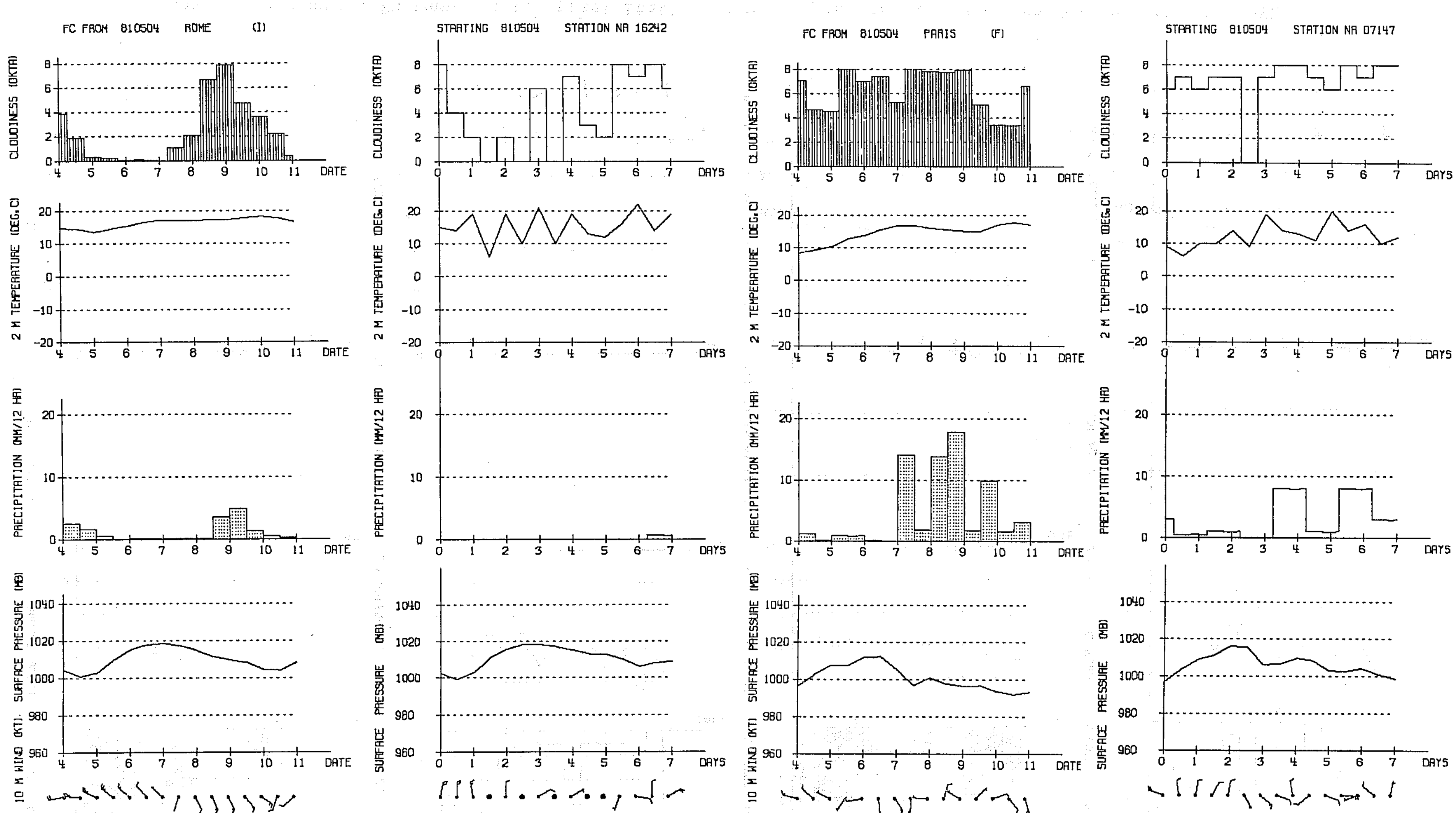
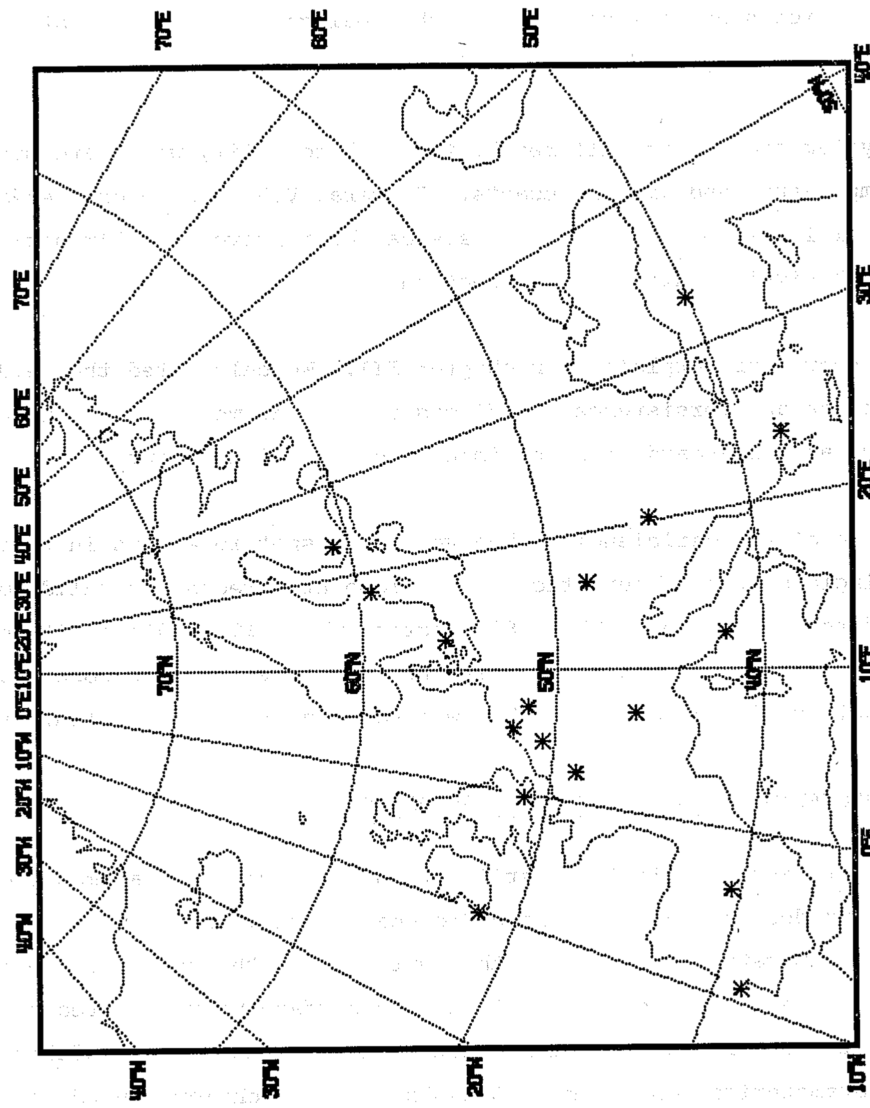


Fig. 2 Same as Figure 1 for Rome and Paris



STATIONS USED IN DIRECT MODEL OUTPUT VERIFICATION

Fig. 3

The derivation of near-surface weather parameters is briefly outlined in para.III and for the following assessment some important facts should be remembered:

(i) The daily cycle is not yet included in the ECMWF model.

(ii) The sea surface temperature is held constant at the climatological monthly mean value.

(iii) The model orography is smooth and thus in some essential features does not reflect the real geographical elevation of the ground or the coastlines.

Various systematic errors will result from (i) to (iii), which are most pronounced in the temperature and wind forecasts. In para. V.5 we show one example of how direct model local temperature forecasts can be improved by eliminating part of the bias through simple statistical correction.

The skill scores are explained in chapter III. We calculated the Heidke scores against chance and persistence. A discussion of the methods and the choice of persistence as a standard of comparison is contained in chapter IV.

We are aware of the deficiencies when verifying spot forecasts in time and space against single station observations. A gridpoint forecast is valid for the complete forecast area of 1.875*1.875 degrees which at middle latitudes is about 200*200 km. For verification purposes one should average the observed data over the available verification box. This was not done for this preliminary exercise.

II. COMPUTATION OF LOCAL WEATHER PARAMETERS

The parameters assessed in this verification exercise are "raw model output", i.e. they are a product of the operational forecast model. The two parameters, wind at 10m and temperature at 2m above the surface, are the result of an interpolation between the lowest model sigma level and the surface values. Precipitation and cloud amount are calculated from parameters used in the ECMWF model's interactive physical parameterization scheme. Details of the techniques used for the derivation of the cloud amount can be found in the Technical Newsletter No.2, February 1980, in an article by J.F. Geleyn. The computation of the precipitation is documented in ECMWF forecast Model Documentation Manual, Vol 1.

Wind and temperature near the surface are interpolated using the models drag coefficient for any physical parameter Ψ

$$C_{D\Psi} = f\left(Ri, \frac{z15}{z0}\right)$$

Where Ri denotes the Richardson number, z_{15} the height of the lowest model level and z_o the roughness length. The function f is highly complicated so that no inversion of the profiles can be obtained easily, however this drag coefficient can formally be obtained from another relationship where $\varkappa$ is the Karman constant

$$C_{D\Psi} = \left(\frac{\varkappa}{\ln \frac{z_{15}+z'_o}{z'_o}} \right)^2$$

and z'_o is a modified roughness length varying with the stability conditions. From this relation z'_o can be determined and a logarithmic profile of the parameter Ψ between the surface and the lowest model level can be derived thus allowing the calculation of the temperature and humidity at 2m above the surface and the wind at 10m above the surface to compare them with synoptic observations.

$$\Psi(z) = \Psi surf + (\Psi 15 - \Psi surf) \frac{\ln \frac{z+z'_o}{z'_o}}{\ln \frac{z_{15}+z'_o}{z'_o}}$$

III. VERIFICATION SCORES

Classical scores like RMS error, correlation coefficient etc. are widely used when forecast elements are expressed on a continuous scale. For verification of temperature and windspeed, we chose the anomaly correlation, the standard deviation and the mean error as a standard set of verification parameters. Temperature and wind forecast and observed anomalies are computed in relation to the observed mean of the month. The standard deviation was chosen in preference to the RMS error as one can expect to correct for large parts of the mean forecast error in future.

Forecast elements like cloud amount and precipitation amount which we expressed categorically cannot be verified in a meaningful way by correlation coefficients. For these elements contingency tables seem to be a useful tool to present the forecast results and evaluate them, and the same method can be applied to temperature and wind speed when they are expressed in categories.

Different verification scores can be obtained from these contingency tables. We applied those suggested e.g. by Brier and Allen (1951) and Klein (1978). To explain the verification scores we consider the following two-category contingency table:

		Observed category 1	Observed category 2	Total
Forecast category	1	A	B	E
	2	C	D	F
	total	G	H	I

Schematic contingency table

Percent correct is given by (A+D)/I*100.

The bias, BS, of category 1 is

$$BS_1 = \frac{E}{G}$$

The Heidke skill score, HS, can be computed from the table directly, when chance is chosen as a standard of comparison.

$$HS_{chance} = \frac{(A+D)-(chance)}{I-(chance)}$$

Chance is the number of forecasts expected to be correct at random and it is based on the margins of the contingency table:

$$chance = \frac{(E*G)+(F*H)}{I}$$

If persistence is chosen as a standard of comparison, persistence forecasts are being made by predicting a continuation of the weather at the time of the analysis and a contingency table for the persistence forecast can be derived in the same way as for the actual forecast. The skill score is then computed by

$$HS_{persistence} = \frac{(A+D)-(a+d)}{I-(a+d)}$$

with small letters a and d denoting the resulting numbers along the diagonal in the persistemce contingency table.

The Heidke score only uses the values along the diagonal to evaluate the success of the forecast and otherwise ignores the remaining distribution of the values in the contingency table. So one does not get penalised for extreme errors as they would appear at both ends of the contingency table orthogonal to the correct diagonal.

Fig. 4 shows a nomographic display of the Heidke score. Along the ordinate one can use any standard of comparison such as chance, persistence or climatology and the graph then shows what portion of correct forecasts with the computed skill score is explained by chance or persistence.

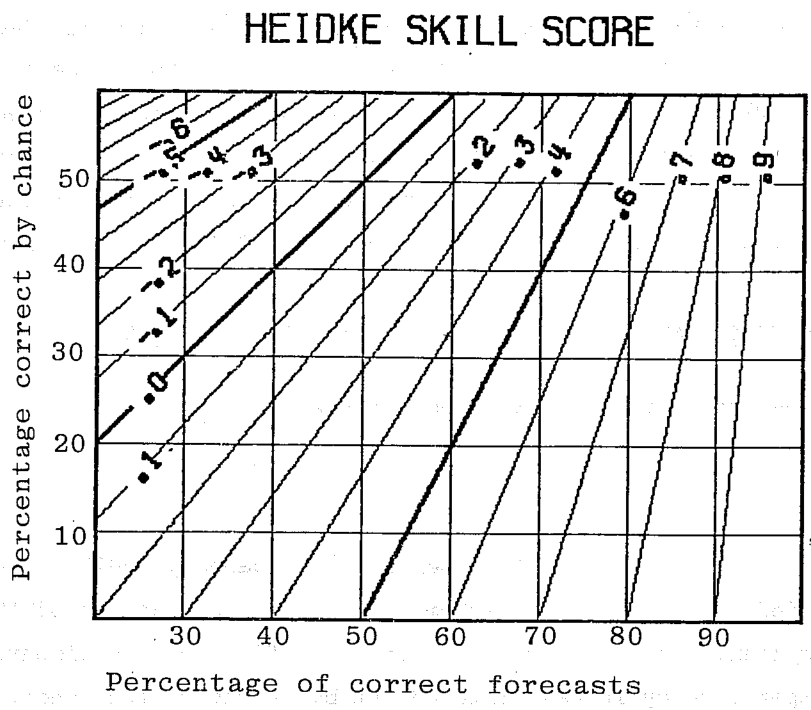


Fig. 4: **Nomograph showing Heidke skill score as a function of the numbers of correct forecasts and the number of correct forecasts by chance as a standard of comparison.**

IV. <u>TWO DIFFERENT STANDARDS OF COMPARISON: PERSISTENCE VERSUS CHANCE</u>

Skill scores can only be used to compare two different forecasts with each other, whereby one of them is considered as the standard of comparison. This standard has to be chosen carefully in order to give meaningful results. Obvious candidates are climatology, chance or random distribution, and persistence. Climatology was not readily available for all the 17 stations and would not have been suitable for precipitation verification. Chance tables derived from forecast contingency tables are widely used and recommended in the literature (Brier and Allen, 1951, and Klein, 1978) and Heidke skill scores can be computed. The examples given in table 1, which is for the 72 hour precipitation forecast, exhibits clearly a serious shortcoming in the choice of chance. The systematic error of the forecast in the chance table is reproduced and even exaggerated, while persistence is a model-independent standard of comparison, easily available and unbiased. Skill scores for persistence are usually lower (see tables 5 and 10) than those for chance, but since we consider them to be more meaningful we preferred to use them in this study. In the example given in table 1 the chance forecast fails to maintain the high number of correct cases in the dry category, reducing it from 148 to 109, while persistence gives 182 correct forecasts in this category. This behaviour in the dry category, in reducing the percent correct in the chance forecast, partly accounts for the higher skill score when choosing chance as a standard of comparison. The importance of the skill score should not be over-estimated because it only reflects the result of the comparison of the numbers of correct forecasts along the diagonal and does not reveal extreme forecast errors and the structure of systematic model errors. One should always assess the overall structure of a contingency table and only use the skill score in addition to other means of evaluation.

V. LOCAL VERIFICATION OF NEAR-SURFACE WEATHER PARAMETERS

V.1 Temperature at 2m above the model surface

Method

During the verification period (December 1980 and January 1981) a diurnal cycle was not included in the radiation scheme of the physics package of the ECMWF forecasting system. For the purpose of this verification experiment, therefore, the daily temperature cycle was filtered out by averaging the synoptic temperature observations in six-hour intervals centred close to the verification time, e.g. for a forecast valid at 00z on day D the observations at 12z and 18z on day D-1 and 00z and 06z on day D were averaged. In cases where one or more data out of these 4 were missing, the verification for this time-step was abandoned to avoid any bias. Fortunately, this happened on average only once or twice per station and month so that the total number of cases was hardly affected.

As measures for the quality of a temperature forecast the mean error, standard deviation and anomaly correlation coefficient were used together with Heidke Skill-scores derived from contingency tables. The classes are defined in terms of deviation from the observed monthly mean for each station, using seven categories ranging from "very cold" (less than -7^{o}), "cold" (-6.9 to -4.5), "rather cold" (-4.4 to -1.5), "normal" (-1.4 to +1.5), "rather warm" (1.6 to 4.5), "warm" (4.6 to 7.0) to "very warm" (over +7.1)

Results

Time graphs for correlation coefficients, standard deviation and mean error for the 15 timesteps (Analysis to day 7 in 12 hr intervals) for individual stations are shown in Figs. 5 to 10 for December and January. Persistence is used as a standard of comparison. The main trends in error growth with time are similar for December and January, with noticeably large mean errors for some stations. These mean errors could be related to differences in true topographic versus model elevation of the surface, coastal effects or systematic errors of the model (both in the

Model Forecast

Model Forecast \ Observed	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

Persistence Forecast \ Observed	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

Chance Forecast

Chance Forecast \ Observed	0-0.2	0.3-2.0	2.1-5.0	5.1-10.0	10.1-
0-0.2	109	27	17	7	3
0.3-2.0	90	22	14	6	3
2.1-5.0	43	11	7	3	1
5.1-10.0	24	6	4	2	1
10.0-	6	2	1	0	0

Table 1: Contingency table for precipitation forecast (top), persistence (centre), and chance (bottom) for ensemble of 17 European stations, 72 hours forecast time, December 1980.

physics and dynamics) in the cases of mean errors with a dominant time evolution. Almost without exception, the anomaly correlations shows a definitive improvement over persistence, which usually reaches the 0% correlation between days 2½ and 4. The limit of .6 which is considered to determine the usefulness of a forecast, is generally reached by the forecasts around days 3 to 5 and around days 1 to 2 by persistence. Correlation reduces more slowly after this and an asymptotic value between 0.3 and 0.4 is maintained at most locations until the end of the forecast.

The trends in standard deviation seem to reflect very much the local climates, with a low, nearly time-constant value for maritime locations (see for example Valentia) and rather rapid increases for some Scandinavian stations (example Jokioinen). Persistence is hardly ever reached before day 7, which is rising sharply at the beginning and levelling off usually around day 5 to approach the forecast value from above.

A more detailed insight into the characteristics of the temperature forecasts can be gained by looking at the contingency tables which summarize the results for all 17 stations, tables 2 and 3. These tables again show the tendency to large mean errors in the forecast, specially in the categories around the climatological average, where the forecasts exhibit a clear positive bias. For the "cold" category, a more balanced picture is found, whereas in the warm observed cases a trend to underforecast the temperature can be observed.

The total number of temperature forecasts in the "warm" and "very warm" categories, however, is higher than the number of observed, but most of these "warm" forecasts occur when "normal" or "rather warm" are observed.

For the 24-hr forecast the percentage of correct forecasts (40%) is very similar to the 38.8% correct for the persistence forecast. Allowing a much wider margin including the two diagonals in the table adjacent to the correct one (+/-4.5°C) 88% of the forecasts and 84% for the persistence are included. The main improvement over the persistence can be found in the correct forecasts for the four extreme categories (warm, very warm, cold, very cold), where the forecasts score 27 cases against 17 for the persistence).

The resulting skill scores seem to be very low, but one has to take into account the mean errors of the forecast and the contribution of stations with a small daily variation in temperature, where persistence scores high for 24 hours.

The general picture for the 48hr forecasts appears very similar to the one for 24 hrs for the forecasts, but the persistence starts to become less valuable at this

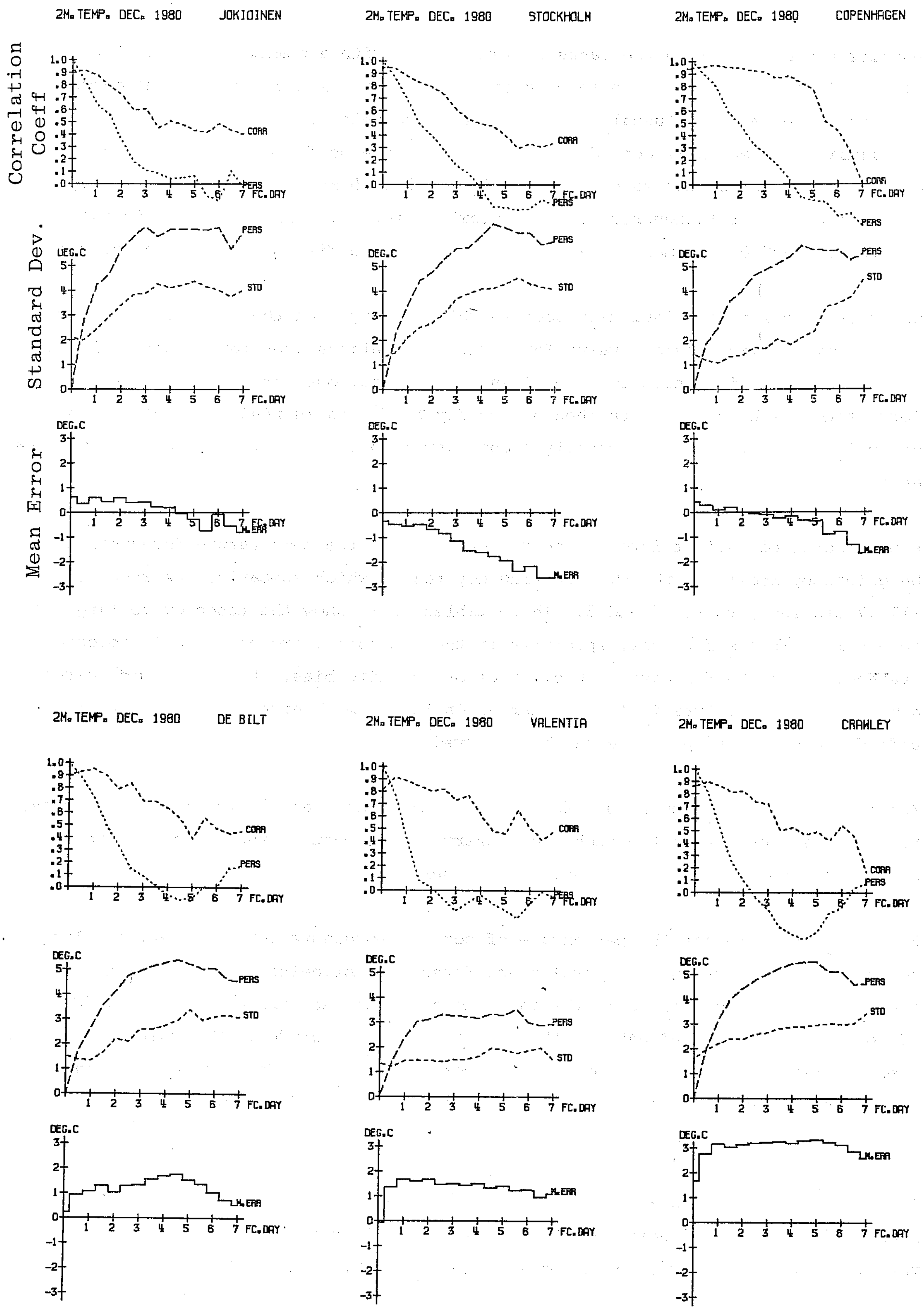


Fig. 5 2m Temperature December 1980

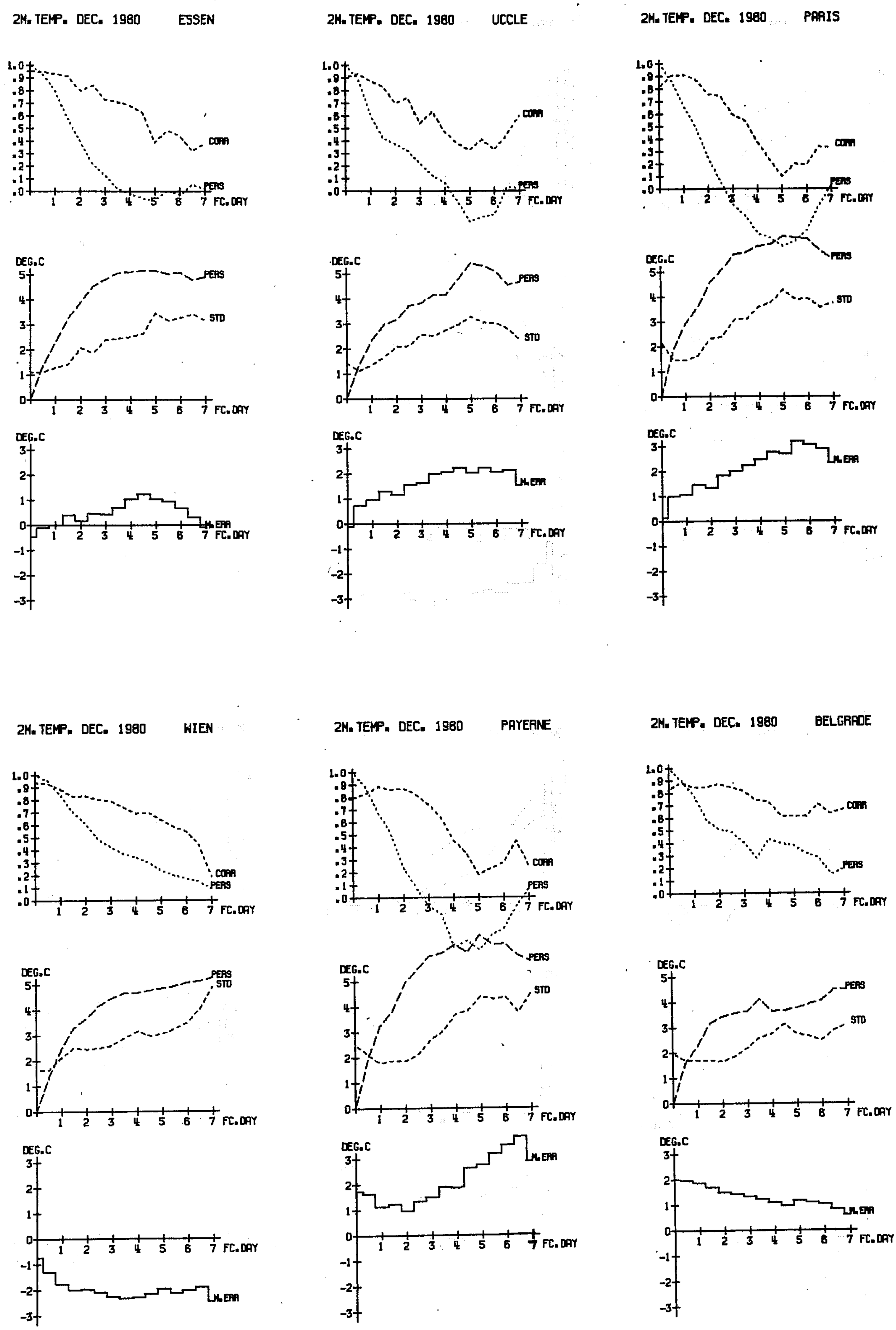


Fig. 6 2m Temperature December 1980

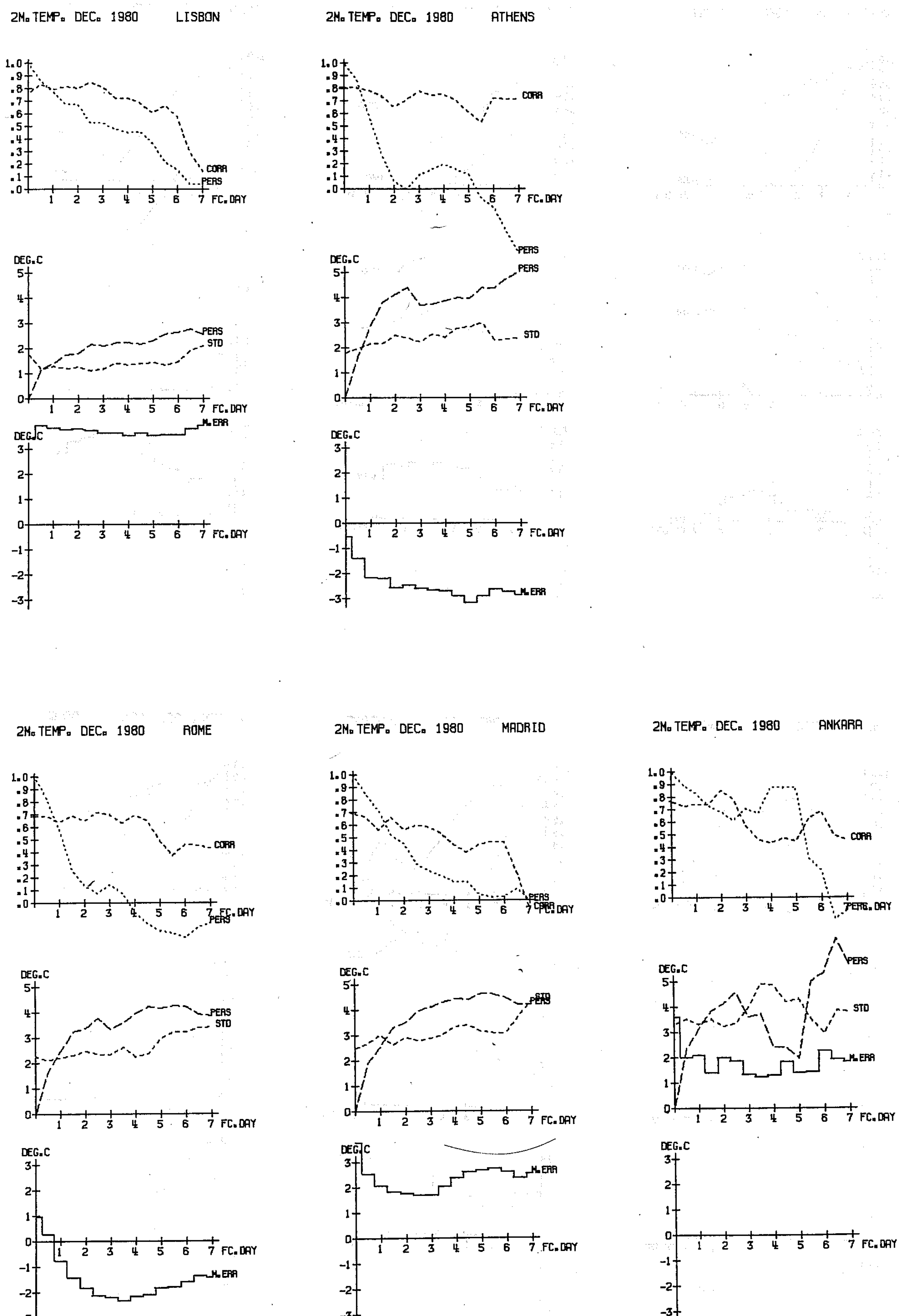


Fig. 7 2m Temperature December 1980

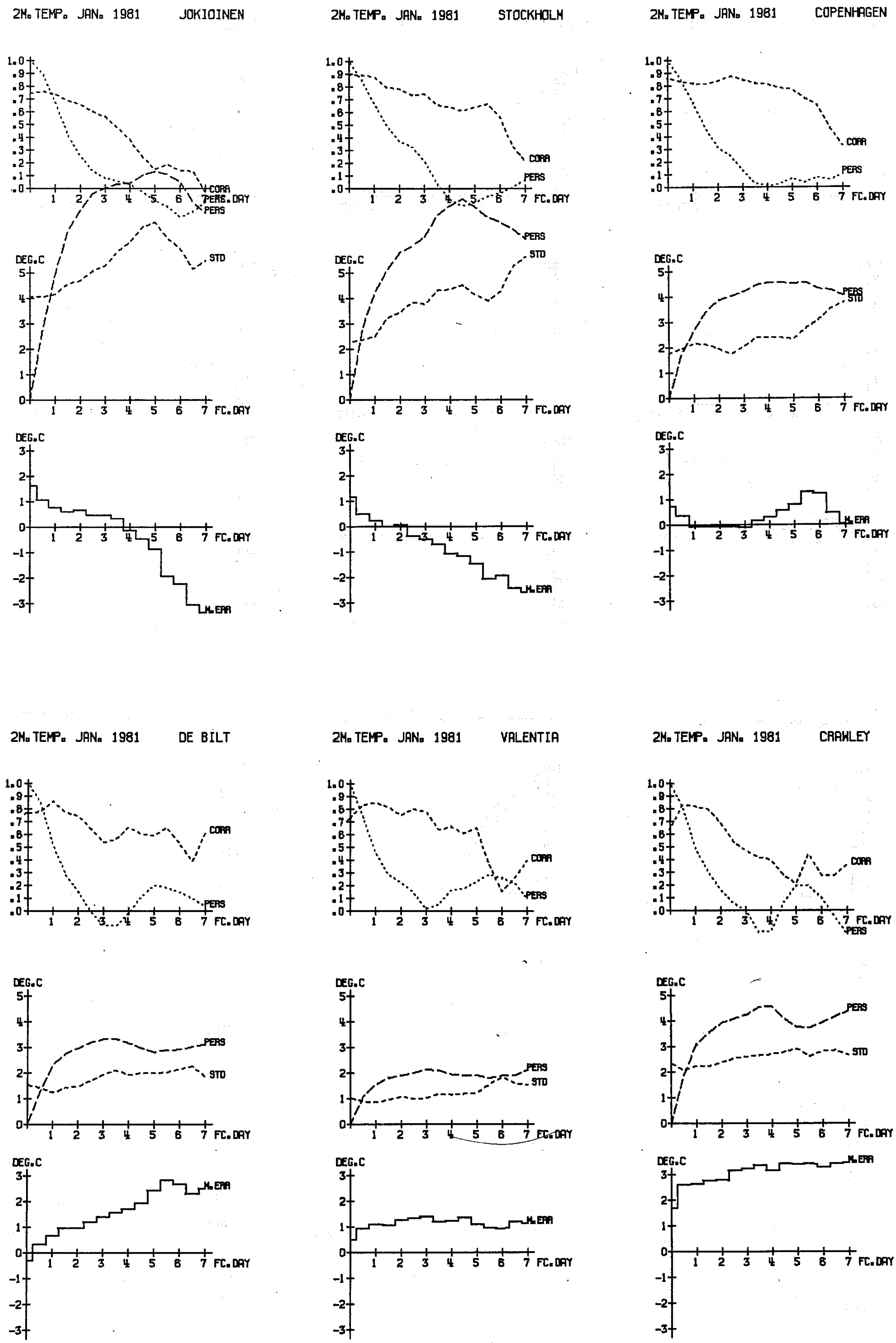


Fig. 8 2m Temperature January 1981

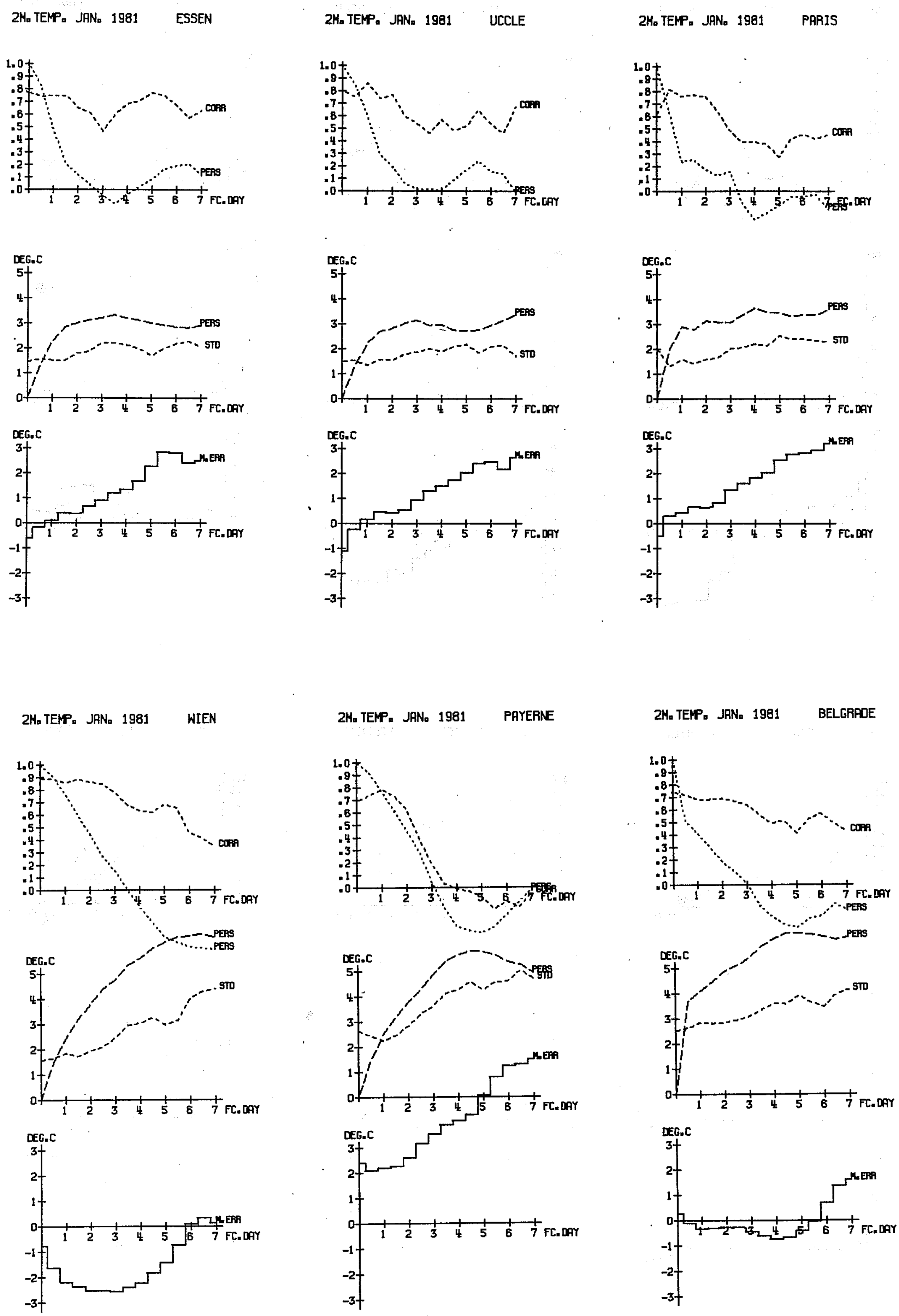


Fig. 9 2m Temperature January 1981

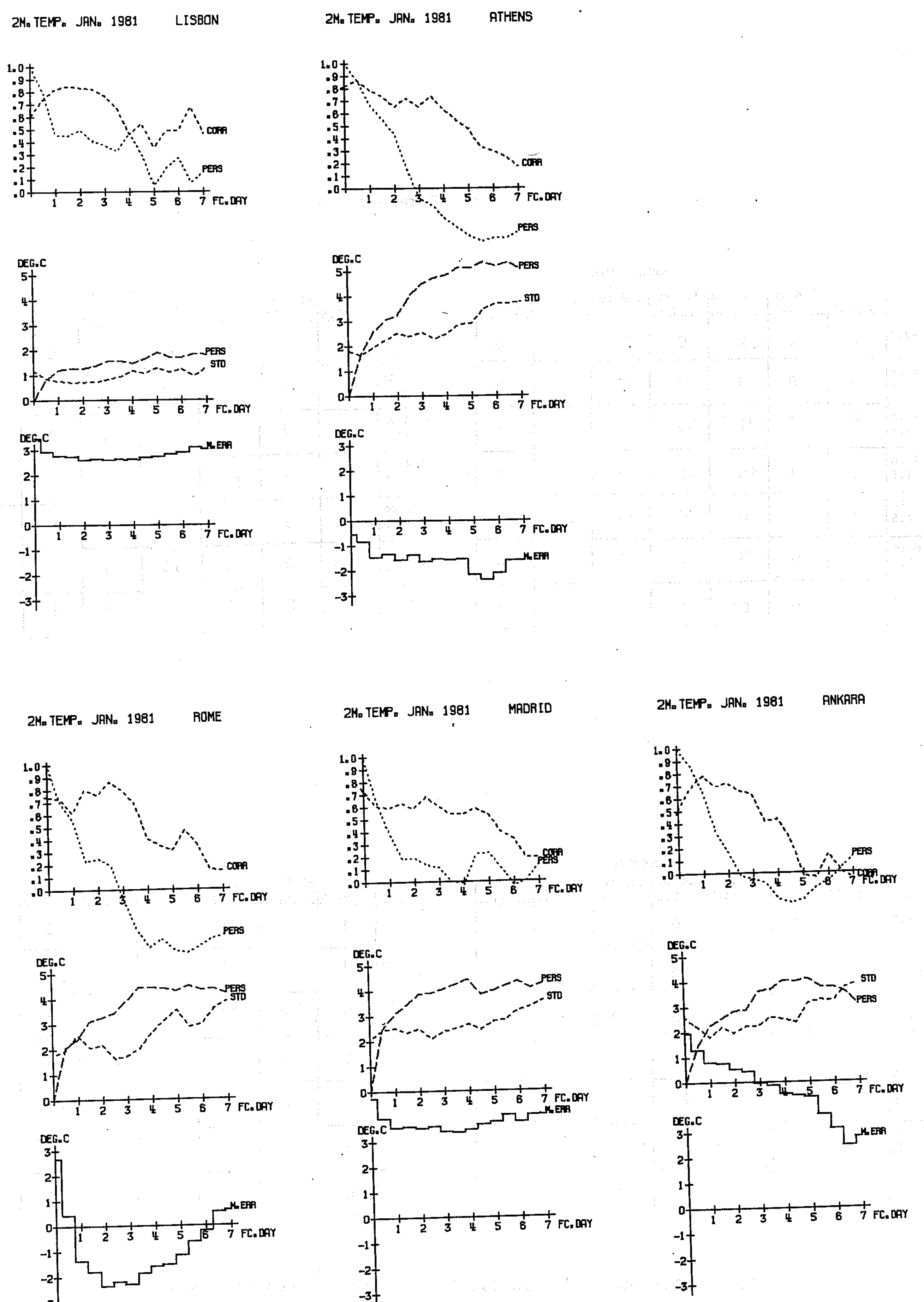


Fig. 10 2m Temperature January 1981

Model Forecast

(a) 24 hour forecast time

	Observed						
	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	4	4	2	0	0	0	0
-7.4 to -4.5	5	8	8	2	8	8	8
-4.4 to -1.5	0	7	31	25	9	0	1
-1.4 to 1.5	0	4	55	92	27	3	0
1.6 to 4.5	1	3	10	76	50	15	0
4.6 to 7.5	0	0	3	11	14	13	2
≥7.6	0	0	0	3	8	2	2

(a) 48 hour forecast time

	Observed						
	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	4	5	3	0	0	0	0
-7.4 to -4.5	5	5	8	5	0	0	0
-4.4 to -1.5	1	7	28	24	6	2	0
-1.4 to 1.5	0	4	52	89	25	5	1
1.6 to 4.5	1	3	19	79	55	13	1
4.6 to 7.5	0	0	3	13	12	13	2
≥7.6	0	0	0	1	5	3	2

Persistence Forecast

(b) 24 hour forecast time

	Observed						
	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	3	4	3	0	0	0	0
-7.4 to -4.5	3	6	12	3	2	0	0
-4.4 to -1.5	3	14	31	40	9	3	1
-1.4 to 1.5	0	2	45	109	38	5	0
1.6 to 4.5	0	0	11	43	36	13	2
4.6 to 7.5	0	0	1	4	17	8	1
≥7.6	0	0	1	1	0	0	1

(b) 48 hour forecast time

	Observed						
	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	1	2	5	0	1	1	0
-7.4 to -4.5	1	3	7	7	2	2	0
-4.4 to -1.5	6	8	31	40	19	3	0
-1.4 to 1.5	2	8	37	96	39	6	3
1.6 to 4.5	0	2	22	42	30	14	1
4.6 to 7.5	0	0	5	13	6	4	2
≥7.6	0	0	0	2	1	1	0

Table 2: Contingency table for 2m temperature anomaly forecast (a) and for persistence (b) for ensemble of 17 European stations, January 1981, left; 24 hour forecast time, right; 48 hour forecast time.

Model Forecast

(a)

Forecast \ Observed	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	2	3	3	0	0	1	0
-7.4 to -4.5	6	6	10	5	2	0	0
-4.4 to -1.5	3	7	22	28	8	1	1
-1.4 to 1.5	1	1	44	69	27	4	0
1.6 to 4.5	0	2	25	100	56	11	3
4.6 to 7.5	0	3	7	9	15	9	1
≥7.6	0	0	0	3	5	4	1

Forecast \ Observed	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	0	1	2	1	1	1	0
-7.4 to -4.5	6	8	9	6	3	1	0
-4.4 to -1.5	2	2	15	20	13	4	1
-1.4 to 1.5	1	5	42	58	16	5	0
1.6 to 4.5	0	5	26	102	47	10	2
4.6 to 7.5	0	5	9	30	26	10	4
≥7.6	0	0	2	3	1	0	0

Persistence Forecast

(b)

Forecast \ Observed	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	0	0	5	2	2	2	0
-7.4 to -4.5	0	1	2	7	8	3	0
-4.4 to -1.5	7	6	29	46	18	3	2
-1.4 to 1.5	2	4	40	91	40	9	2
1.6 to 4.5	1	10	23	45	29	6	2
4.6 to 7.5	0	0	5	12	7	5	0
≥7.6	0	0	0	1	2	0	0

Forecast \ Observed	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	0	0	2	3	6	1	0
-7.4 to -4.5	0	0	2	15	5	4	0
-4.4 to -1.5	1	5	30	40	20	8	4
-1.4 to 1.5	1	5	37	94	45	11	2
1.6 to 4.5	5	10	19	50	19	4	0
4.6 to 7.5	2	3	8	7	7	1	0
≥7.6	0	0	0	1	0	1	1

Table 3: same as Table 2 for 72 hour forecast time (left) and 120 hour forecast time (right)

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

Table 4: **Heidke skill score for 2m temperature anomaly forecasts from 12 to 120 hours for 17 stations in Europe, using chance and persistence as a standards of comparison.**

stage. We still find 38.9. of the forecasts in the correct diagonal as compared to 32.7% for persistence, and in only 11 cases (January) the forecast is out by more than two categories, none being wrong by more than 4. This result is similar for the December cases. The cases of misleading forecasts (more than two cateogries out) may, to a large extent, be attributed to stations with large mean errors, which can be shown in chapter V.5.

After 72 and 120 hours into the forecast, a wider spread of cases away from the correct diagonal can be found (Table 3), with the bias towards warmer forecasts near the observed climatological mean becoming more and more pronounced. Only 33% of the forecasts are now found in the correct category.

For the "normal" and "rather cold" categories, the forecasts fall behind persistence in January and are only equivalent in December. The improvement over persistence, however, remains for the extreme categories. As one would expect, the corners of the table far away from the correct diagonal now contain a few cases – 3 in December and 5 in January are now 4 or more categories out for 72 hours of forecast time, but even for 120 hour forecast time these numbers still stay below 2% of the total number of cases.

V.2 Cloud amount

The direct model output parameter, cloud amount, is a global field giving the total amount of clouds at each individual gridpoint at discrete forecast timesteps, presently 00z and 12z. No information about cloud types or cloud amount at different levels is available from the model at present. We therefore verify the cloud forecast against the total cloud cover as it is reported in synoptic observations. This is a verification of a spot forecast value in time and in space against a spot observation and no averaging in time has been applied.

The results for all 17 stations are summarized in contingency tables from which skill scores were derived, again using persistence as a standard of comparison. When defining the classes we followed the widely used terminology of forecasters and chose four classes: "cloud free or clear" (1/8), "fair" (1/8 to 3/8), "cloudy" (4/8 to 7/8), and "overcast" (7/8).

Tables 5 and 6 show the results from the 24, 48 and 72 hour forecast for the 17 European stations summarized in contingency tables. In general, as can be seen from the skill scores (given in table 7), the forecast was slightly more successful in December than it was in January, but the results are very similar for both months. Persistence is very difficult to beat over Europe in winter as "overcast" sky is

Model Forecast (a)

Left (24 hour):

Model Forecast \ Observed	< 1/8	1/8-3/8	4/8-7/8	≥ 7/8
< 1/8	31	7	13	10
1/8-3/8	20	8	7	16
4/8-7/8	24	16	41	77
≥ 7/8	4	9	23	186

Centre (48 hour):

Model Forecast \ Observed	< 1/8	1/8-3/8	4/8-7/8	≥ 7/8
< 1/8	34	5	10	10
1/8-3/8	15	7	5	18
4/8-7/8	22	20	41	76
≥ 7/8	10	8	32	177

Right (72 hour):

Model Forecast \ Observed	< 1/8	1/8-3/8	4/8-7/8	≥ 7/8
< 1/8	33	3	8	12
1/8-3/8	17	11	11	14
4/8-7/8	25	16	38	82
≥ 7/8	8	9	31	172

Persistence Forecast

Left (24 hour):

Persistence Forecast \ Observed	< 1/8	1/8-3/8	4/8-7/8	≥ 7/8
1/8	28	6	21	26
1/8-3/8	9	6	6	17
3/8-7/8	17	7	10	47
≥ 7/8	27	21	45	197

Centre (48 hour):

Persistence Forecast \ Observed	< 1/8	1/8-3/8	4/8-7/8	≥ 7/8
< 1/8	20	11	17	32
1/8-3/8	8	2	8	20
4/8-7/8	14	3	16	47
≥ 7/8	40	25	43	182

Right (72 hour):

Persistence Forecast \ Observed	< 1/8	1/8-3/8	4/8-7/8	≥ 7/8
< 1/8	23	10	12	35
1/8-3/8	5	2	10	22
4/8-7/8	12	7	12	51
≥ 7/8	43	15	50	178

Table 5: Contingency tables for cloud amount forecast (a) and for persistence (b) for ensemble of 17 European stations, December 1980, left: 24 hour, Centre: 48 hour, right: 72 hour forecast time.

(a) Model Forecast

Model Forecast \ Observed	<1/8	1/8-3/8	4/8-7/8	≥ 7/8
< 1/8	35	6	6	14
1/8-3/8	18	10	12	22
4/8-7/8	23	18	45	89
≥7/8	5	10	45	149

Model Forecast \ Observed	<1/8	1/8-3/8	4/8-7/8	≥ 7/8
< 1/8	30	11	7	16
1/8-3/8	20	3	12	35
4/8-7/8	21	17	47	80
≥ 7/8	14	16	41	138

Model Forecast \ Observed	< 1/8	1/8-3/8	4/8-7/8	≥ 7/8
< 1/8	31	5	6	16
1/8-3/8	21	12	9	28
4/8-7/8	15	20	42	86
≥ 7/8	18	11	48	141

(a)

(b) Persistence Forecast

Persistence Forecast \ Observed	< 1/8	1/8-3/8	4/8-7/8	≥7/8
<1/8	36	7	6	20
1/8-3/8	10	3	7	23
4/8-7/8	13	10	26	60
≥ 7/8	19	21	62	166

Persistence Forecast \ Observed	< 1/8	1/8-3/8	4/8-7/8	≥7/8
< 1/8	31	12	12	16
1/8-3/8	11	1	4	24
4/8-7/8	10	9	25	64
≥ 7/8	26	22	72	160

Persistence Forecast \ Observed	< 1/8	1/8-3/8	4/8-7/8	≥7/8
<1/8	30	8	10	23
1/8-3/8	13	2	9	19
4/8-7/8	11	11	19	66
≥7/8	25	25	66	153

(b)

Table 6: Same as table 5 for January 1981

Forecast time / Standard of comparison	Hours						
	12	24	36	48	60	72	
Chance	.27	.30	.22	.27	.20	.27	December 1980
persistence	-.07	.10	-.03	.14	-.03	.14	
Chance	.20	.22	.25	.16	.17	.17	January 1981
persistence	-.22	.03	.07	.00	.03	.07	

Table 7: **Heidke skill score for cloud amount forecasts from 12 to 72 hours for 17 stations in Europe using chance and persistence as standards of comparison.**

observed in more than 50% of the cases, and good results can be achieved by just forecasting this state of the sky. Although the percent correct score is only slightly above 50% for all three forecast steps in December and even below that mark in January the model still shows some skill in forecasting the cloud amount. Only 15% of the 24 hour forecasts in December and 16% in January are wrong by more than one class and these figures only come close to 20% for the 48 and 72 hour forecasts while for persistence the number exceeds 20% for the 24 hour forecast and soon reaches 30% thereafter. Less than 50% of the "clear" sky cases are correctly predicted by the model in 48 and 72 hour forecasts. In this category a strong bias to overestimate the cloud amount at each of the three forecast steps is observed. In the other three categories the forecast is more balanced. Spot value cloud forecasts are, of course, very sensitive to minor errors in the predicted flow pattern and sub-synoptic processes cannot be expected to be represented locally in a realistic way by a global model. A field verification of the predicted cloud amount against digitized satellite images would be preferable to a point verification.

V.3 Wind speed at 10m above the model surface

The u and v component of the wind are directly forecast at 15 sigma levels in the operational forecast model and the wind at 10m above the model surface is interpolated from the lowest sigma level using a logarithmic vertical profile as described in para II. We verified the wind speed at 00z and at 12z against the observed wind speed as it is reported in synoptic observations i.e. against the 10 minute average wind speed. Although we tried to find stations which are located in homogeneous terrain one cannot exclude the local effects completely and that is reflected in the mean errors for the wind speed at most locations as they exhibit the same error pattern in December and January. In Fig. 11 to 14, from top to bottom, the anomaly correlation coefficients, the standard deviation and the mean error for the wind speed in December and January for a selection of stations is given. The mean error for most stations is fairly high and positive, in the order 2 to 4 m/s.

Some systematic model errors become very obvious in the graphs of the mean error of the 10m wind speed. As the daily cycle is not included in the model the mean error exhibits a pronounced daily cycle between midnight and midday verification steps. The night time stabilisation in the boundary layer with its drop in the wind speed is not simulated by the model at present. For exposed stations like Copenhagen and Wien (the station is situated on a hill) the mean error is low and close to zero, but larger errors are observed for Valentia (Ireland) and Crawley (England). Both stations are, of course, land stations and Crawley is not even a coastal station, but in the ECMWF model they were both situated over sea as

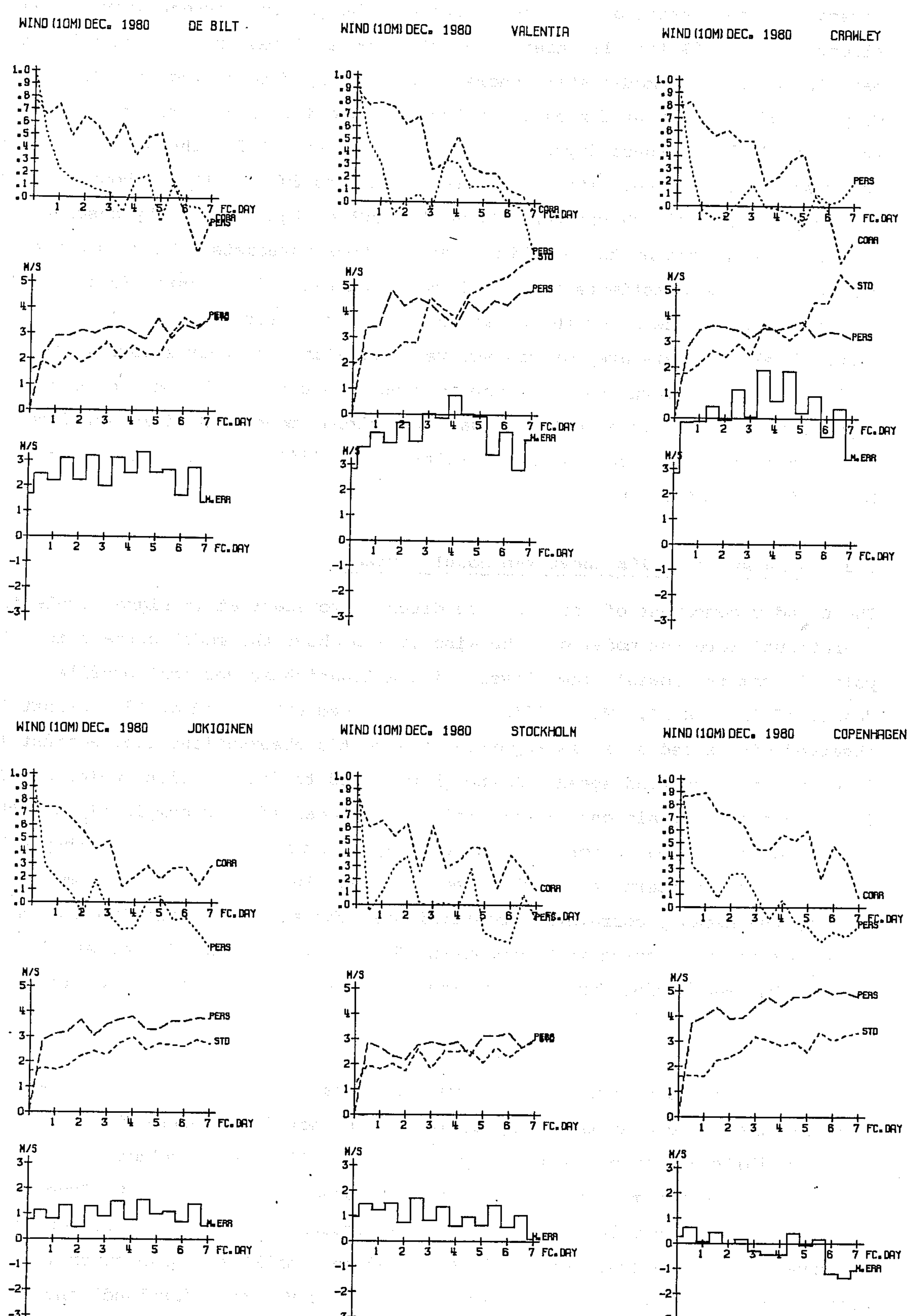


Fig. 11 10m Wind Speed December 1980

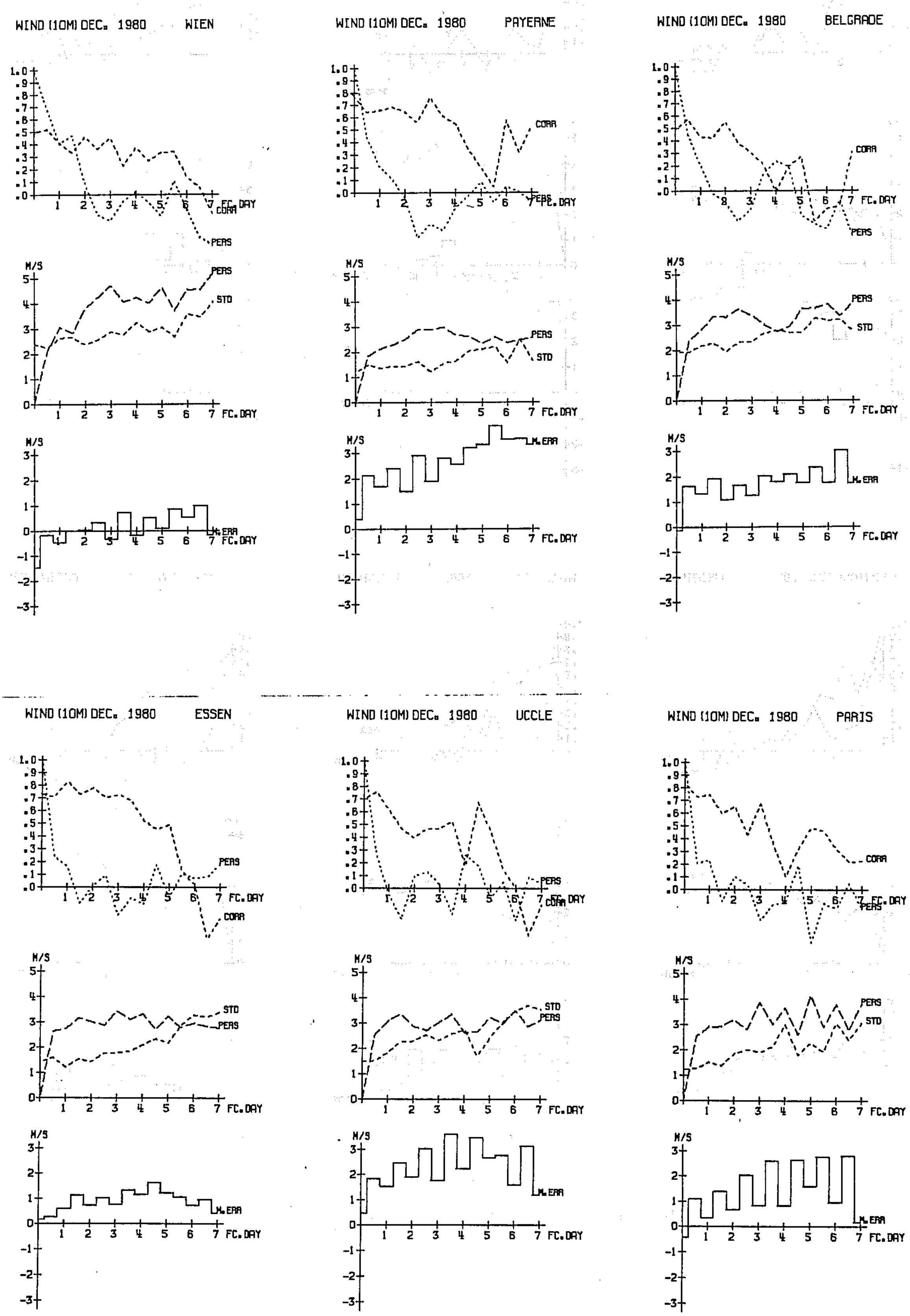


Fig. 12 10m Wind Speed Dec. 1980

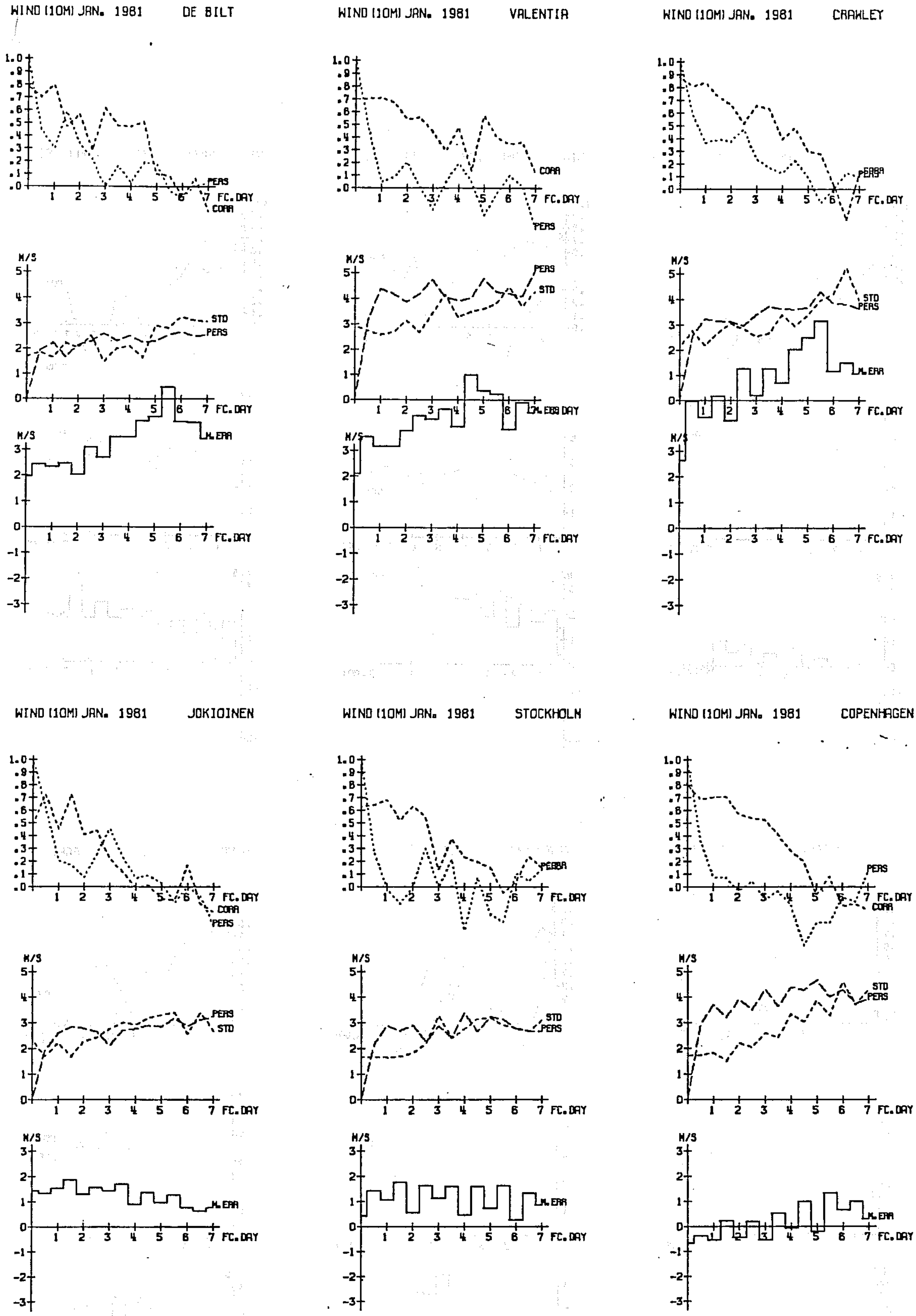


Fig. 13 10m Wind Speed January 1981

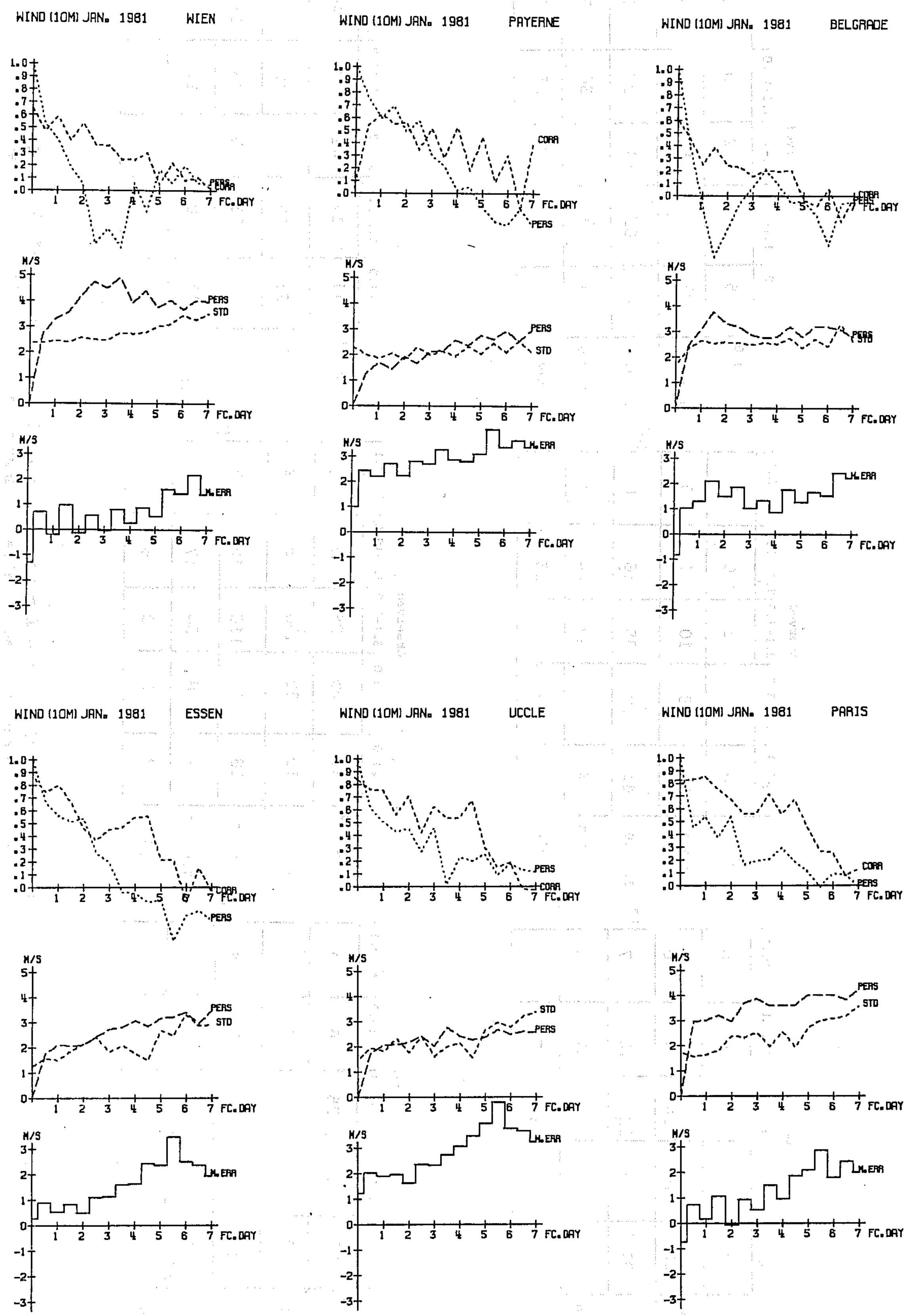


Fig. 14 10m Wind Speed January 1981

Model Forecast — 24 hours (a)

Model Forecast \ Observed	≤1.0	1.1-3.0	3.1÷7.0	7.1-10.0	≥ 10.1
≤ 1.0	1	1	1	0	0
1.1-3.0	35	16	14	1	0
3.1-7.0	37	79	127	10	7
7.1-10.0	0	4	68	32	6
≥10.1	0	0	26	27	7

Model Forecast — 48 hours (a)

Model Forecast \ Observed	≤1.0	1.1-3.0	3.1÷7.0	7.1-10.0	≥ 10.1
≤ 1.0	0	2	1	0	0
1.1-3.0	31	18	10	1	0
3.1-7.0	41	70	135	16	8
7.1-10.0	2	9	53	32	7
≥10.1	0	0	33	25	4

Model Forecast — 72 hours (a)

Model Forecast \ Observed	≤1.0	1.1-3.0	3.1÷7.0	7.1-10.0	≥ 10.1
≤ 1.0	5	1	2	0	0
1.1-3.0	22	23	10	1	2
3.1-7.0	43	63	131	22	8
7.1-10.0	4	9	55	30	7
≥10.1	0	1	28	26	5

Persistence Forecast — 24 hours (b)

Persistence Forecast \ Observed	≤1.0	1.1-3.0	3.1÷7.0	7.1-10.0	≥ 10.1
≤ 1.0	22	16	25	6	4
1.1-3.0	19	27	46	7	2
3.1-7.0	24	52	130	30	5
7.1-10.0	5	3	32	24	7
≥10.1	0	2	9	6	2

Persistence Forecast — 48 hours (b)

Persistence Forecast \ Observed	≤1.0	1.1-3.0	3.1÷7.0	7.1-10.0	≥ 10.1
≤ 1.0	15	20	25	9	4
1.1-3.0	19	22	50	5	2
3.1-7.0	28	47	116	42	9
7.1-10.0	7	8	34	17	6
≥10.1	2	2	8	6	0

Persistence Forecast — 72 hours (b)

Persistence Forecast \ Observed	≤1.0	1.1-3.0	3.1÷7.0	7.1-10.0	≥ 10.1
≤ 1.0	11	14	41	5	4
1.1-3.0	21	19	46	10	2
3.1-7.0	30	51	105	43	11
7.1-10.0	11	8	32	17	3
≥10.1	3	3	9	4	0

Table 8: Contingency table for 10m wind speed forecast (a) and for persistence (b) for ensemble of 17 European stations, December 1980, left: 24 hours, forecast time, centre: 48 hours and right: 72 hours forecast times.

Model Forecast

Forecast \ Observed	≤1.0	1.1-3.0	3.1-7.0	7.1-10.0	≥ 10.1
≤ 1.0	1	1	1	0	0
1.1-3.0	28	29	13	2	0
3.1-7.0	49	84	132	25	3
7.1-10.0	8	14	53	27	4
≥10.1	0	2	17	18	7

(a)

Forecast \ Observed	≤1.0	1.1-3.0	3.1-7.0	7.1-10.0	≥ 10.1
≤ 1.0	1	1	1	0	0
1.1-3.0	29	25	24	2	0
3.1-7.0	49	87	127	27	4
7.1-10.0	6	15	54	24	5
≥10.1	1	4	13	17	4

(a)

Forecast \ Observed	≤1.0	1.1-3.0	3.1-7.0	7.1-10.0	≥ 10.1
≤ 1.0	2	2	3	0	0
1.1-3.0	22	21	13	2	1
3.1-7.0	51	88	124	28	5
7.1-10.0	10	21	64	20	2
≥10.1	3	3	13	19	4

(a)

Persistence Forecast

Forecast \ Observed	≤1.0	1.1-3.0	3.1-7.0	7.1-10.0	≥ 10.1
≤ 1.0	33	27	16	5	0
1.1-3.0	28	36	51	9	1
3.1-7.0	18	53	108	29	6
7.1-10.0	4	11	33	20	5
≥10.1	1	1	4	8	1

(b)

Forecast \ Observed	≤1.0	1.1-3.0	3.1-7.0	7.1-10.0	≥ 10.1
≤ 1.0	26	26	22	6	2
1.1-3.0	26	38	52	9	0
3.1-7.0	26	52	98	33	5
7.1-10.0	3	12	37	17	4
≥10.1	3	2	6	3	1

(b)

Forecast \ Observed	≤1.0	1.1-3.0	3.1-7.0	7.1-10.0	≥ 10.1
≤ 1.0	27	18	22	10	4
1.1-3.0	23	44	50	8	1
3.1-7.0	24	51	101	36	4
7.1-10.0	12	14	31	12	4
≥10.1	0	3	10	2	0

(b)

Table 9: Same as Table 8 January 1981

Britain and Ireland were treated as sea areas in the surface fields in operational use in December 1980 and January 1981.

When comparing analyzed wind speeds to observations, a standard deviation of about 2m/s is reached for most stations and this figure then only increases slowly with forecast time. In many cases the standard deviation is very close (but seldom exceeds) the values achieved by persistence forecasts. The anomaly correlation of the wind speed drops rapidly during the first three days and after that in most cases lies below 0.5. This certainly is an advantage against a persistence forecast where in most cases zero correlation is reached within 24 or 36 hours.

The strong mean error in wind speed forecasts is also reflected in the contingency tables (Tables 8 and 9) which summarize the result of the verification of the wind speed for all 17 stations for December and January. All forecasts are highly biased towards overestimation of the wind speed in each class and only the December forecast shows some skill against persistence when applying the Heidke score. However, as it can clearly be seen from the contingency tables, the forecast beats persistence in forecasting strong winds and hardly any error occurs in the extreme corners of the contingency tables orthogonal to the diagonal of the correct forecasts.

In para V.5 we give an example of how the strong bias can be reduced by simple statistical methods, thus improving the forecast and raising the skill score.

V.4 Precipitation

The forecast accumulated precipitation at 12 hourly intervals has been verified against synoptic observations for December 1980 and January 1981. Observations of precipitation are available at 00z and 12z, in addition to the main observations at 06z and 18z, for all the pertinent stations except Jokioinen and Ankara, which have not been included in this study, as the verification data were not available.

A more comprehensive verification study has been performed for the months of October and November 1980 for selected points and areas over Euope (Åkesson, 1980). The results from that study are consistent with the results of the present paper.

Results

Monthly values of mean error (bias) and standard deviation for the 15 stations are shown in meteograms in Figs. 15 for December 1980 and for January 1981. When comparing December and January it is obvious, at least for some stations, that they have certain characteristics especially in the mean error and its trend, e.g. for Paris with a rising 12 hourly bias up to day 7, Payerne with almost no

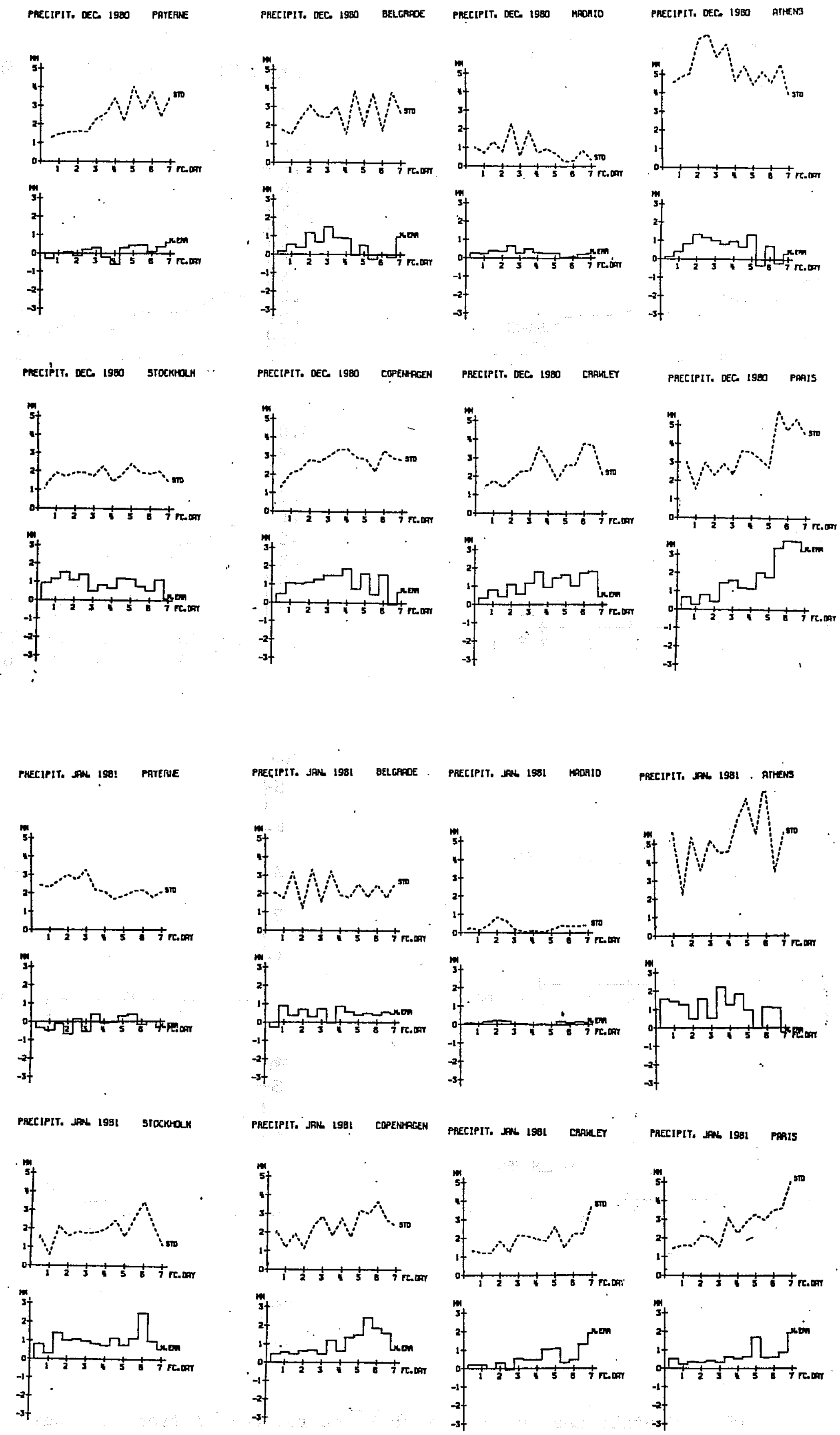


Figure 15. Monthly mean values of standard deviation and mean error for Payerne, Belgrade, Madrid, Athens, Stockholm, Copenhagen, Crawley and Paris. Top: December 1980, bottom: January 1981.

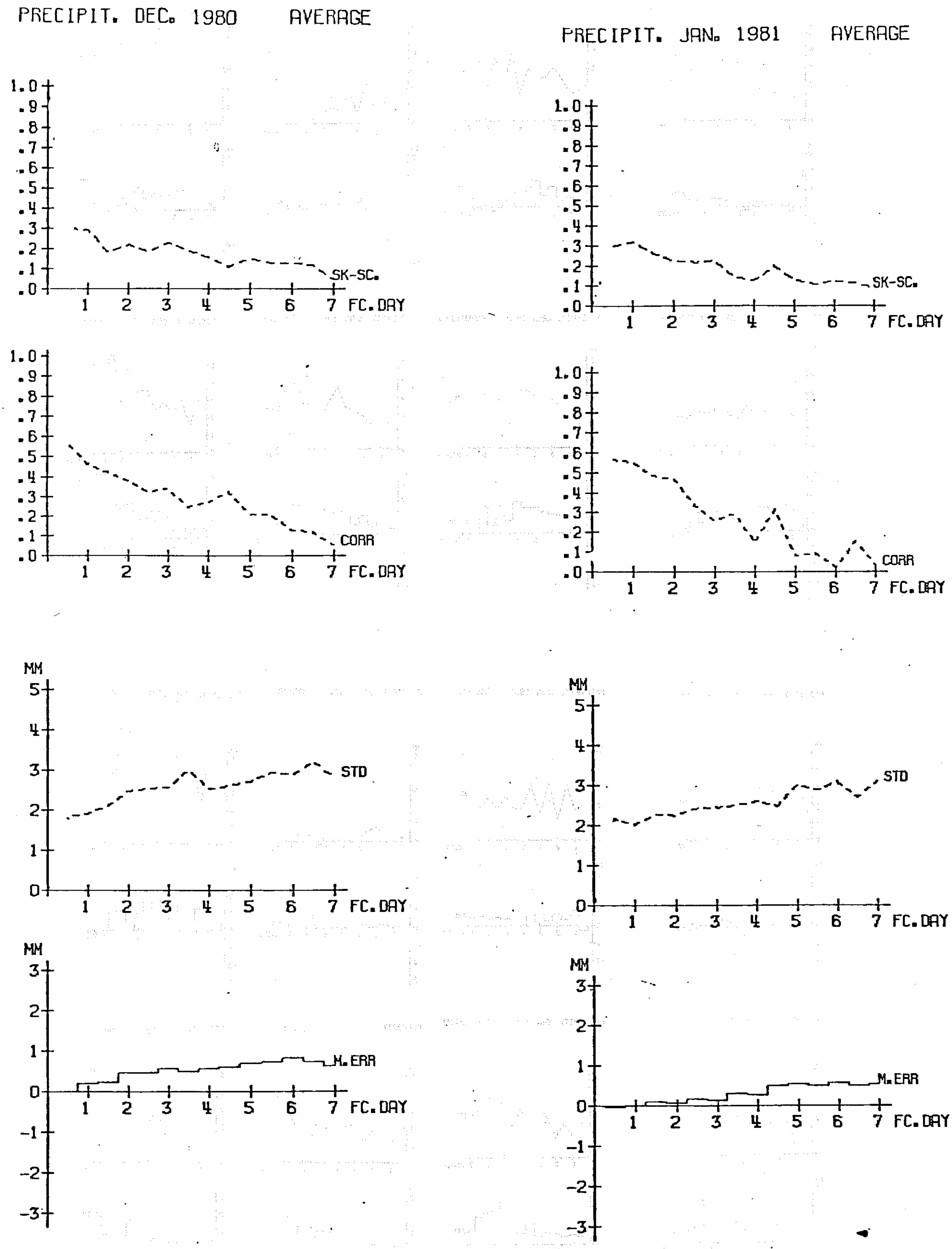


Figure 16: Monthly mean values of skill score, correlation, standard deviation and mean error averaged over 15 European stations. Left: December 1980, right: January 1981.

bias throughout the 7 day period and Athens with high over-prediction culminating around day 2-4 and with a slight falling tendency thereafter. Note also the relatively high standard deviation of error for Athens related to high forecast (and observed) amounts in contrast to Lisbon where only small amounts were forecast. The standard deviation of error is thus a function of the location, the season, but above all of the forecast amount. An indication of a daily cycle can also be seen from low latitude stations, both in standard deviation or error and mean error. Fig. 16 shows the 15 station average verification statistics for both months. In these the local influence for individual stations have been largely removed and over-prediction of precipitation is shown to occur from +24 hours onwards with a slow and steady increase with time. The correlation drops from about 0.55 for the +12 hour forecast to near 0.0 for the +168 hour forecast whereas the skill score shows only a minor drop with time. The standard deviation of the errors is also rather constant with time and with an error of around 2mm for the 12 hour forecast rising to around 3mm for the 168 hour forecast.

It is in the nature of precipitation that stochastic processes are important and time or area-averaging will thus improve the verification statistics without obscuring the information. This has been shown by Åkesson (1980) by averaging the 6 gridpoints with 1.875 degrees resolution. In particular when the observations were averaged there was a significant reduction in standard error. A small improvement can also be expected to result from extending the time interval from 12 to 24 hours. Furthermore the bias could be reduced by statistical means.

Based on contingency tables with five categories: 0-0.2, 0.21-2.0, 2.0-5.0, 5.01-10.0 and more than 10 mm, Heidke skill scores based on chance and persistence have been computed for the ensemble of the 15 stations. Contingency tables with 5 categories for the ensemble up to 72 hours are shown in Table 10a.

In Table 10a contingency tables for +24, +48 and +72 hour forecasts for December 1980 and January 1981 respectively, are shown. It is clear that the dry events (0-0.2 mm) dominate throughout this two month period. There is a certain bias in the forecasts showing many more predicted than observed events of small amounts of precipitation, especially on D+3 of the forecast. Also, at the upper end of the categories there are more events of high amounts of observed precipitation compared to forecast precipitation. These events are common in precipitation episodes of convective character where the forecast model does not resolve the detailed structure of the convective clouds and the consequent horizontal variation of precipitation intensity. For this reason significantly higher skills are obtained by averaging the forecast, and in particular the observed values, over an area e.g. over 4 or 6 gridpoints. The different biases for small and large amounts respectively

FC MW \ OBS MW	0-0.2	0.3-2.0	2.1-5.0	5.1-10.0	10.1-
0-0.2	167	16	1	2	0
0.3-2.0	80	34	13	3	4
2.1-5.0	18	15	23	9	0
5.1-10.0	3	6	4	3	2
10.0-	0	0	2	0	1

FC MW \ OBS MW	0-0.2	0.3-2.0	2.1-5.0	5.1-10.0	10.1-
0-0.2	199	17	2	1	0
0.3-2.0	94	32	13	7	1
2.1-5.0	22	16	19	6	3
5.1-10.0	5	4	7	2	1
10.0-	3	1	2	0	1

FC MW \ OBS MW	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

FC MW \ OBS MW	0-0.2	0.3-2.0	2.1-5.0	5.1-10.0	10.1-
0-0.2	207	18	4	0	2
0.3-2.0	83	47	6	3	2
2.1-5.0	12	9	13	5	3
5.1-10.0	6	2	3	2	4
10.0-	1	0	1	1	0

FC MW \ OBS MW	0-0.2	0.3-2.0	2.1-5.0	5.1-10.0	10.1-
0-0.2	197	21	7	0	0
0.3-2.0	89	38	5	4	5
2.1-5.0	23	13	9	4	4
5.1-10.0	6	3	2	1	2
10.0-	1	0	1	2	0

FC MW \ OBS MW	0-0.2	0.3-2.0	2.1-5.0	5.1-10.0	10.1-
0-0.2	189	13	6	2	3
0.3-2.0	95	42	6	3	5
2.1-5.0	27	16	6	2	2
5.1-10.0	6	3	6	4	2
10.0-	2	0	0	0	0

Table 10a: Contingency tables of precipitation forecasts for the ensemble of 15 European stations.
Top row: December 1980, bottom row: January 1981

Hrs.	12	24	36	48	60	72	84	96	108	120	132	144	156	168
CH.	.30	.32	.26	.22	.22	.22	.15	.12	.19	.13	.11	.12	.11	.09
PER.	-.02	.13	.01	-.02	-.11	-.08	-.07	-.12	-.07	-.12	-.13	-.13	-.15	-.17

Table 10b. Heidke skill scores based on Chance (CH) and persistence (PER) for January 1981.

FC. HOURS	12	24	36	48	60	72	84	96
PERCENT CORRECT FORECAST	62	62	57	56	55	55	51	49

Table 10c. Percent correct forecasts up to +96 hours for January 1981.

counter-balance each other to yield a reasonably low overall bias over a month, as seen in Figure 16. There is, however, a gradual shift to more events of forecast precipitation with time reflected in Table 10a.

Table 10b summarizes Heidke skill scores based on chance and persistence for January 1981. Scores based on persistence are significantly lower and mostly negative for all time steps in the forecast. Both scores drop as the forecast proceeds but level out somewhat around day D+4.

Table 10c shows the percent correct numbers for the first 4 days of the January sample. These numbers are closely related to the skill scores. An alternative score would be to consider percent totally wrong events. From Table 10a it is clear that measuring success or failure in this manner by, for instance, counting only the events in the 3 upper right classes and in the 3 lower left classes and comparing these events to those so obtained from the persistence table, the forecast is significantly better than persistence up to +72 hours. This shows the limited usefulness of the skill score, particularly when used with persistence, and that complementary information could be yielded by a "failure score". The superiority of the forecast over persistence continues till around +108 hours into the forecast, when events of failure are about the same in both forecast and persistence.

V.5 A test of forecast improvement by means of simple statistical correction

Mean temperature errors in the forecasts are similar for December and January, compare Figures 5,6 and 7 with Figures 8,9 and 10 respectively. This indicates that a large part of these mean errors are associated with differences between real and model topography, or are due to local climatic effects. The same is true for the mean wind errors, compare Figures 11 and 12 with Figures 13 and 14 respectively. Hence an experiment has been carried out in order to test the degree of improvement which can be achieved by means of simple statistical correction, taking into account the dominant role of the mean errors. Thus, the error found for each time-step at individual stations for the December-data was subtracted from the January-forecast values.

Given the short period of data available, a further differentiation into errors specific to the different categories, although desirable in principle, was not undertaken. The results of this first step are quite encouraging: The Heidke Skill score for temperature increased from 0.02 to 0.20 for the 24 hour forecasts (Table 12), the percentage of correct forecasts from 40% to 51% amd 94% of all forecasts are now within 1 category of the correct one; only 3 cases out of 500 are now out by more than two categories, and none more than 4 categories (Table 11a).

Model Forecast / Observed

	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	4	4	2	0	0	0	0
-7.4 to -4.5	5	9	4	2	0	0	0
-4.4 to -1.5	1	7	59	39	6	1	1
-1.4 to 1.5	0	6	42	138	63	4	0
1.6 to 4.5	0	0	1	27	30	11	1
4.6 to 7.5	0	0	1	3	8	14	1
≥7.6	0	0	0	0	1	3	2

Observed

	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	3	2	3	0	0	1	0
-7.4 to -4.5	6	5	4	3	2	0	0
-4.4 to -1.5	3	8	43	31	6	2	1
-1.4 to 1.5	0	5	52	141	62	4	0
1.6 to 4.5	0	1	9	37	30	8	3
4.6 to 7.5	0	1	0	2	11	12	1
≥7.6	0	0	0	0	2	3	1

Persistence Forecast / Observed

	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	4	4	3	0	0	0	0
-7.4 to -4.5	5	7	4	2	0	0	0
-4.4 to -1.5	2	6	53	30	6	1	1
-1.4 to 1.5	0	6	44	143	60	8	0
1.6 to 4.5	0	1	8	34	26	10	2
4.6 to 7.5	0	0	1	2	9	16	1
≥7.6	0	0	0	0	2	1	2

Observed

	≤ -7.5	-7.4 to -4.5	-4.4 to -1.5	-1.4 to 1.5	1.6 to 4.5	4.6 to 7.5	≥ 7.6
≤-7.5	0	2	1	0	1	1	0
-7.4 to -4.5	5	3	8	4	5	0	0
-4.4 to -1.5	3	9	29	37	9	3	1
-1.4 to 1.5	0	9	46	125	39	5	0
1.6 to 4.5	2	3	18	46	39	12	3
4.6 to 7.5	0	1	3	7	13	10	3
≥7.6	0	0	0	1	1	0	0

Table: 11

a) Contingency table for 2m temperature anomaly forecast for ensemble of 17 European stations, January 1981, corrected by the mean error of December 1980 forecasts, top: 24 hour forecast time, bottom: 48 hour forecast time.

b) same as a), top: 72 hour forecast time, bottom: 120 hour forecast time.

Forecast time / Standard of comparison	Hours 12	24	36	48	60	72	84	96	108	120	
Persistence	-.29	.02	.03	.09	.06	.03	.08	.09	.01	-.02	JANUARY 1981
Persistence with mean error corrected	-.17	.20	.14	.25	.22	.23	.22	.19	.15	.17	

Table 12: Heidke skill score for 2m temperature anomaly forecast from 12 to 120 hours for 17 stations in Europe, January 1981. Persistence was used in both forecast evaluations as a standard of comparison, the scores in the bottom row are for forecasts which were corrected for the bias using the mean error of the previous month.

	Observed				
	≤1.0	1.1-3.0	3.1-7.0	7.1-10.0	≥ 10.1
≤ 1.0	17	18	7	0	0
1.1-3.0	37	44	30	1	1
3.1-7.0	31	64	141	31	3
7.1-10.0	1	4	35	33	5
≥10.1	0	0	3	7	5

	Observed				
	≤1.0	1.1-3.0	3.1-7.0	7.1-10.0	≥ 10.1
≤ 1.0	15	11	17	0	0
1.1-3.0	37	47	40	3	2
3.1-7.0	33	66	132	35	4
7.1-10.0	1	7	27	26	5
≥10.1	0	1	3	6	2

a): Contingency table for 10m wind speed forecast for ensemble of 17 European stations, January 1981, corrected by the mean error for December 1980, top: 24 hours forecast time, bottom: 48 hours forecast time.

	Observed				
	≤1.0	1.1-3.0	3.1-7.0	7.1-10.0	≥ 10.1
≤ 1.0	13	14	6	0	0
1.1-3.0	31	37	41	2	2
3.1-7.0	38	71	139	34	5
7.1-10.0	5	11	29	27	2
≥10.1	1	2	2	6	3

b): same as a) 72 hour forecast time.

TABLE 13

This rather heartening picture remains virtually unchanged for the 48hr forecasts, (Table 11a), the only major drawback being found in the observed "rather warm" category, where the modified forecasts show an increased tendency to predict the lower, "normal" class.

The improvement over persistence is now very clear, at 49.8% correct forecasts compared to 32.7 for persistence. We also find still over 90% of all the forecasts within one category from correct, and the number of failures (more than 2 categories out) hardly increased at all to 4 out of 500. Even at 72 hours, the forecasts are to 46% correct, but more cases are now found in the two diagonals adjacent to the correct one (Table 11b). The number of real failures, however, is still very low at 6 out of 506. The improvement over persistence is now found in all categories, especially in the "normal" category (141 compared to 91 cases), where the unmodified forecast was already beaten by persistence. This positive trend is now conserved up to at least 120 hours, where a skill score of 0.17 still indicates a clear advantage over persistence and a large improvement over the unmodified forecasts, whose best score in January is reached at 0.09 at 48 hrs (0.15 for December).

The results of the corrected wind forecasts show a similar trend (Table 13), but for the reasons mentioned in para. V.3, the forecasts remain strongly biased towards an overestimation of the wind in the low wind speed classes. The skill scores (not tabled) using persistence as a standard of comparison increased against the uncorrected forecast from a level of no skill to .13 for the 24 hour forecast and only drops to .10 for the 72 hour forecast, where the number of the extreme wrong forecasts is now halved and only accounts for less than 2% of the total number of forecasts.

VI. SUMMARY AND CONCLUSIONS

ECMWF direct model output of near-surface weather parameters - 2m temperature, cloud amount, 10m wind speed and precipitation - are verified against synoptic observations at 17 locations in Europe for December 1980 and January 1981 forecasts out to 168 hours.

The results are summarised in graphs of anomaly correlation, standard deviation, and mean error for individual stations (temperature and wind only) and in contingency tables for the ensemble of the 17 stations.

The results can be summarised as follows:

1. The temperature at 2m above model surface:
   Given the absence of a diurnal cycle in the model it is reasonable at this time to verify the forecast values for every twelve hours against an average of four six-hourly temperature observations centred around the forecast time. Furthermore, the parameterization scheme of a global model does not reflect local climatic conditions. The mean errors show, therefore, large variations from station to station, and cause the bias exhibited in the contingency table where the results are summarized for the ensemble of the 17 stations. Despite this bias the forecast shows considerable skill in predicting extreme events with a noticeably small number of extreme errors even out to 120 hours of forecast time. As the mean error pattern was found to be similar in December and January, an experiment was carried out to improve the January forecasts by subtracting at each location the bias observed in December. This correction improved the skill of the forecasts significantly, as can be seen by the Heidke scores.

2. Cloud cover:
   When interpreting the results of the verification of forecast cloud amount against local, spot observations, several points should be borne in mind:

   The lack of a diurnal cycle of radiation in the model limits the skill in predicting convective cloud.

   Cloud cover in synoptic observation and in the forecast are given as instantaneous values.

Fog and low level clouds are not well represented in the model. However, the verification study shows that overcast skies, which are frequently observed in winter, are well forecast, and there is some skill in predicting clear skies, but misleading forecasts account for about 15 to 20% of the cases for forecast time out to 72 hours.

3. Windspeed at 10m above model surface:

Again, the lack of diurnal cycle in the model accounts for a large proportion of the forecast error, especially the failure to predict calms in stable night time inversion layers.

This gives a strong positive bias at low wind speeds whereas high windspeeds are adequately captured. Mean errors show a large station to station variability, reflecting the influence of local topographic and climatic conditions. A statistical correction applied in the same way as for the 2m temperature extends the predictability of the wind forecast, although the quality of the corrected temperature forecast is not reached.

4. The station verification of precipitation shows improvement over persistence in cases of more than 0.2mm. observed. The dry category, however, is better described by the persistence forecast since the model tends to produce more and more spurious precipitation with increasing forecast time.

The number of misleading forecasts (more than 2 categories wrong) are lower than for persistence until about 108 hours into the forecast.

The mean bias for all stations exhibits an increasing tendency with forecast time to overpredict total monthly precipitation with almost no bias during the first 24 hours. It should be noted, however, that the high spatial variability of observed precipitation is limiting the scope of direct model output of precipitation.

The result of this investigation indicated that the direct model output of near-surface weather parameters are potentially useful as predictors in local weather element forecasting. Temperature and wind site forecasts promise to give good results after application of some statistical bias correction, whereas for cloud amount and precipitation the forecast field should be compared to area averaged observation. Within the given limitations of a global model for local forecasting this appears to be a realistic approach.

References

Åkesson, O. 1981, Evaluation of the ECMWF operational model precipitation forecast in October-November 1980, ECMWF, Operations Department Technical Memorandum No. 24.

Brier, G.W. and Allen, R.A. 1951, Verification of weather forecasts, Compendium of Metoeorlogy, AMS, p841-848.

Geleyn, J-F. 1980, Pseudo "satellite picture" presentation of model results, ECMWF Newsletter No.1, February 1980, p2.

Klein, W. 1978, Statistical Forecasts of local weather by means of model output statistics. ECMWF seminars, p186-220.

\- 1981, ECMWF Forecast Model Documentation Manual