TECHNICAL MEMORANDUM

51

Verification of ECMWF quantitative precipitation forecasts over Europe January 1980 to April 1981

K.R. Johannessen

Research Department

January 1982

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

TABLE OF CONTENTS

1. INTRODUCTION

As an element of weather forecasts, precipitation is probably the most important and also, as has been pointed out many times, the most difficult to get right. As an output of numerical weather prediction, precipitation was added relatively recently and there are few published statistics available which can serve to establish the quality and usefulness of objective quantitative numerical precipitation forecasts.

At ECMWF, precipitation has been an integral part of the output since operational forecasts began in September of 1979. The fields of precipitation are defined globally at grid points with mesh length 1.875 degrees of longitude and latitude and are archived in the same format for the entire length of the forecast period of 10 days. Here is an excellent basis on which to assess the state of numerical precipitation forecasting. ECMWF also predicts the kind of precipitation, either liquid or frozen. This aspect has not been pursued in this report and would also be more difficult to verify than amount of precipitation. This should be the subject of a later investigation.

Verification statistics are necessary to monitor the progress of modelling the atmosphere for the purpose of numerical weather prediction. They are equally important for learning how to use the products of a forecast centre in many fields of application such as agriculture, energy, industry, transportation, and construction, just to mention a few of the areas where planning and operations have reached such levels of development that systematic use of weather forecasts can greatly increase efficiency.

The founding fathers of the World Weather Watch of the WMO spoke in the initial declaration of the enormous benefits that would derive to mankind if forecasts of precipitation for agriculture could be extended to weeks instead of a day or two. ECMWF and other forecast centres around the world are now reaching into this extension of forecasts, and the progress should be measured by appropriate verification programs.

It is the purpose of this report to set a few benchmarks for precipitation forecasts. It presents, in graphical form, verification results during 16 months from January 1980 to April 1981, using scores and comparison control forecasts that have become accepted standards.

2. ECMWF MODEL OROGRAPHY

This report is concerned with verifying and measuring the skill of ECMWF forecasts of quantitative precipitation during the 16 months from January 1980 to April 1981. The verification area is essentially Europe with a bit of North Africa included to give border continuity for the member states on the Mediterranean.

The orography used by the model during the first 15 months of this period is depicted in Fig. 1a. It has several notable deficiencies, such as flat country over British Isles, the high country over Scandinavia displaced to the Baltic Sea, and in general, greatly reduced values of effective height. In April 1981 the new orography depicted in Fig. 1b was introduced from the first of the month and would have affected the verification for that month.

Figure 1 may be useful for the interpretation of the results of this verification.

3. THE FORECAST PRECIPITATION FIELDS

The present ECMWF operational forecast model is a global grid point model on a mesh of 1.875 degrees longitude and latitude. Large-scale and convective precipitation is computed separately at all grid points and is output as cumulative fields at 20 intervals of 12 hours out to 10 days or 240 hours. One 10-day forecast is made per day with 12Z as initial time. Prior to August 1980, five forecast runs were made per week, Sunday through Thursday. In August 1980, this was stepped up to one forecast per day.

The verifications in this report extend from January 1980 through April 1981. They include 16 months and 421 forecasts, each for 10 days, as shown in Table 1.

TABLE 1: Number of Forecasts Verified per Month

Jan 1980	22	Sep 1980	29
Feb 1980	20	Oct 1980	31
Mar 1980	22	Nov 1980	30
Apr 1980	22	Dec 1980	30
May 1980	20	Jan 1981	31
Jun 1980	21	Feb 1981	28
Jul 1980	23	Mar 1981	31
Aug 1980	31	Apr 1981	30

TOTAL 421

3.1 Sample Forecasts

To illustrate the forecast fields, a forecast run from January 3, 12Z, 1981, is reproduced in Figs. 2a to 2f. The maps 2a to 2c depict the total precipitation for D1, D2, and D3, where D1 indicates the first day of the forecast run, D2 the second, etc. The day used in this verification is the period from 18Z to the next 18Z, for reasons explained later, so the forecast for D1 is a 30-hour forecast and the forecast for D9, a 222-hour forecast. The unit used throughout the report is 1/10 of a mm. The forecast fields are contoured at 50-unit intervals for 1-day durations, 100-unit intervals for 5-day durations, and 200-unit intervals for 9-day durations. The numbers plotted at the regularly spaced grid points are the verifying average amounts of

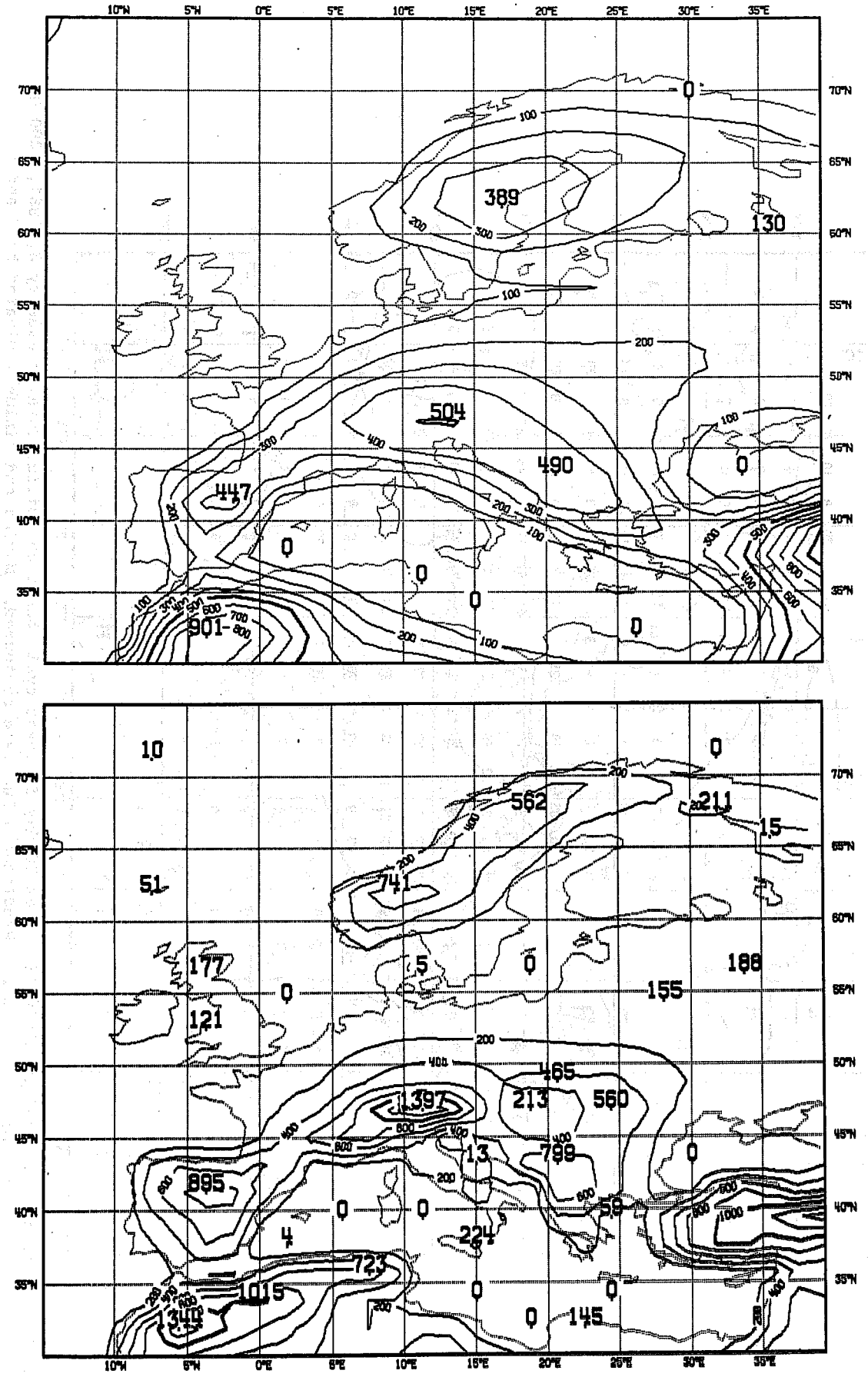

Fig. 1 ECMWF model orography over verification area, meters. Maxima and minima indicated. Top: Prior to April 1, 1981. Contour interval 100m. Bottom: From April 1, 1981. Contour interval 200m.

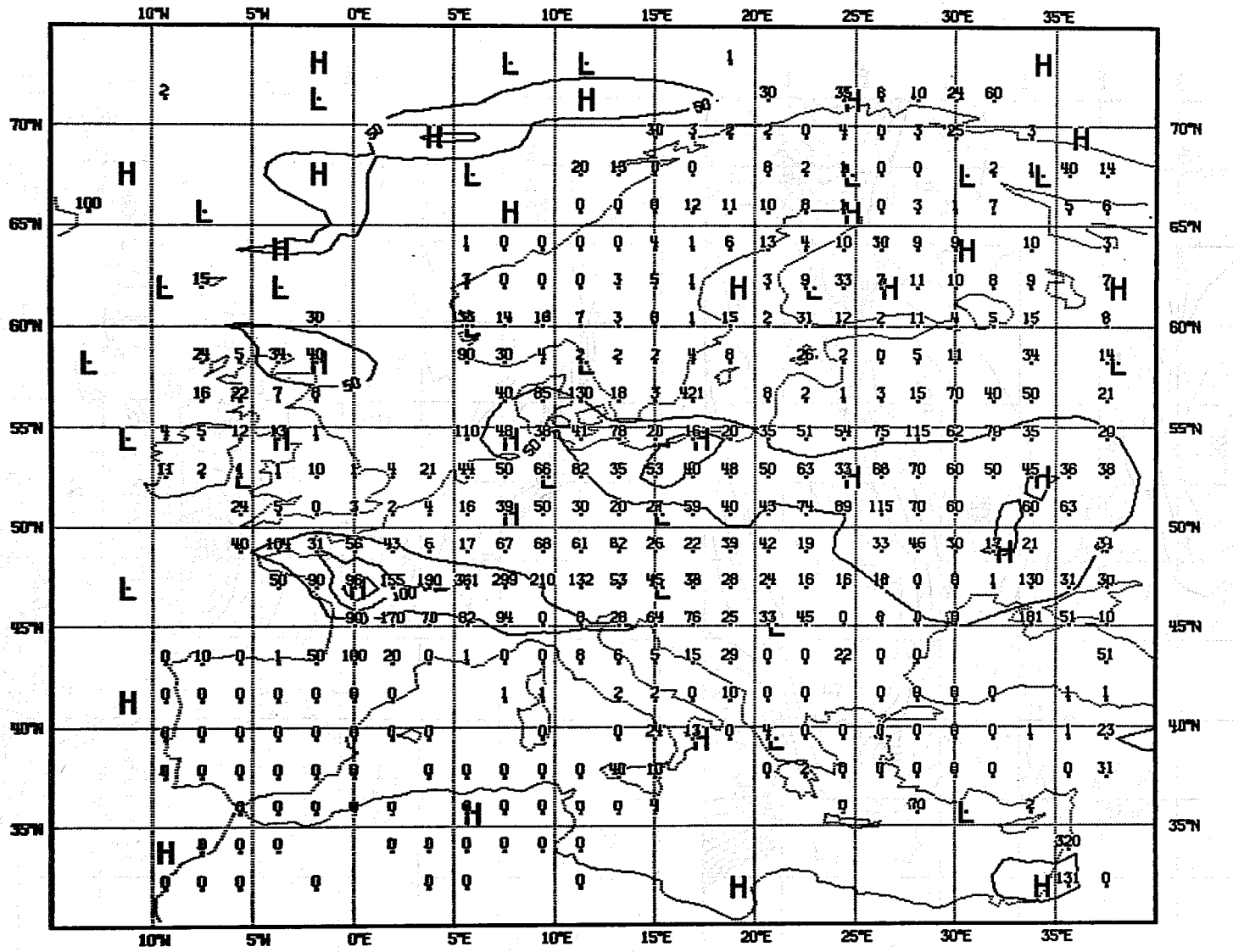

Fig. 2a Predicted and observed precipitation from 10-day forecast cycle based on initial data on January 3, 1981, 1200 GMT. Unit is one-tenth of a millimeter. Predicted values are contoured. Maxima and minima indicated by H and L. Observed values are plotted at grid points, where available. 24-hour interval 1800Z on D0 to 1800Z D1. Contour interval 50.

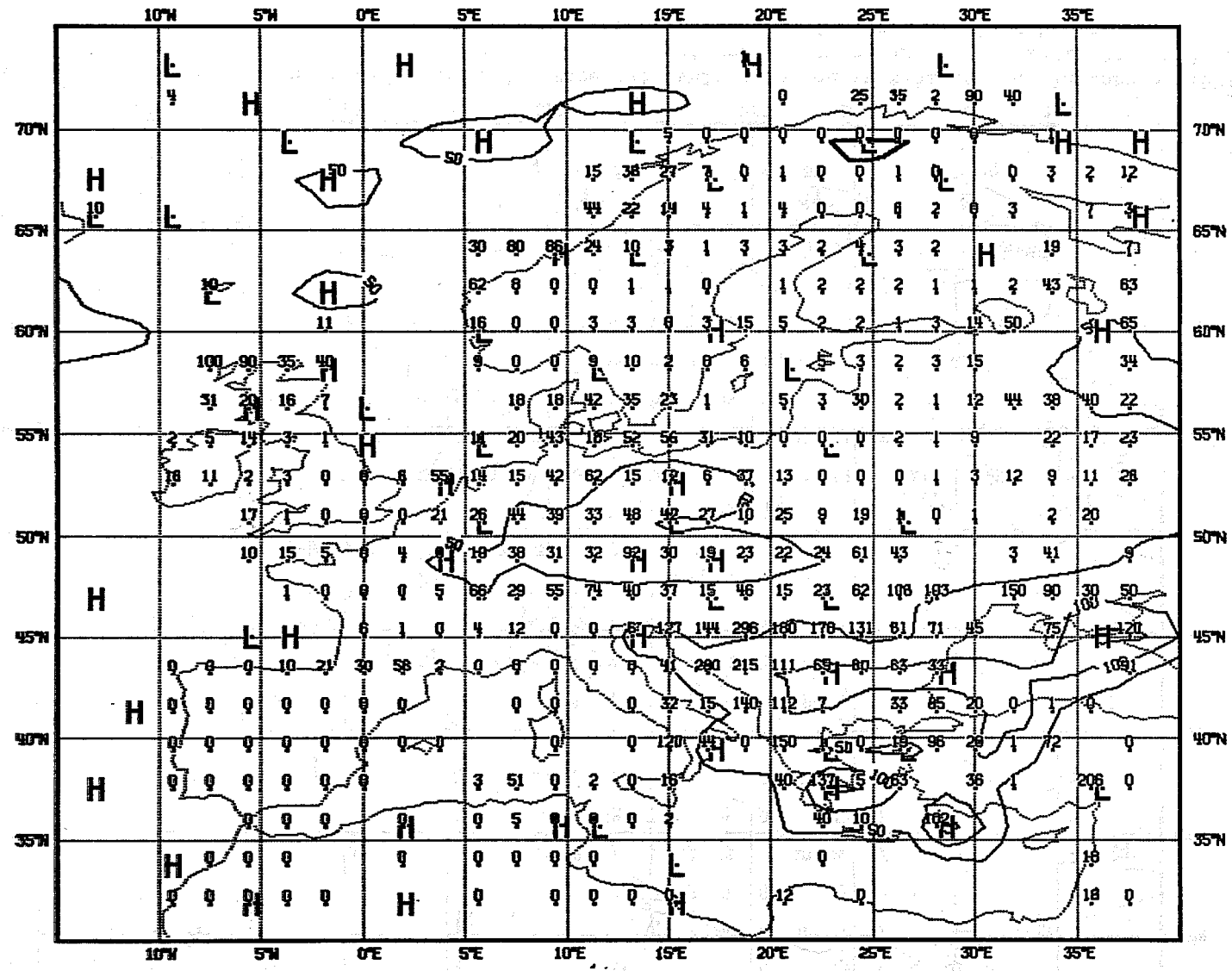

Fig. 2b Predicted and observed precipitation from 10-day forecast cycle based on initial data on January 3, 1981, 1200 GMT. Unit is one-tenth of a millimeter. Predicted values are contoured. Maxima and minima indicated by H and L. Observed values are plotted at grid points, where available. 24-hour interval 1800Z on D1 to 1800Z on D2. Contour interval 50.

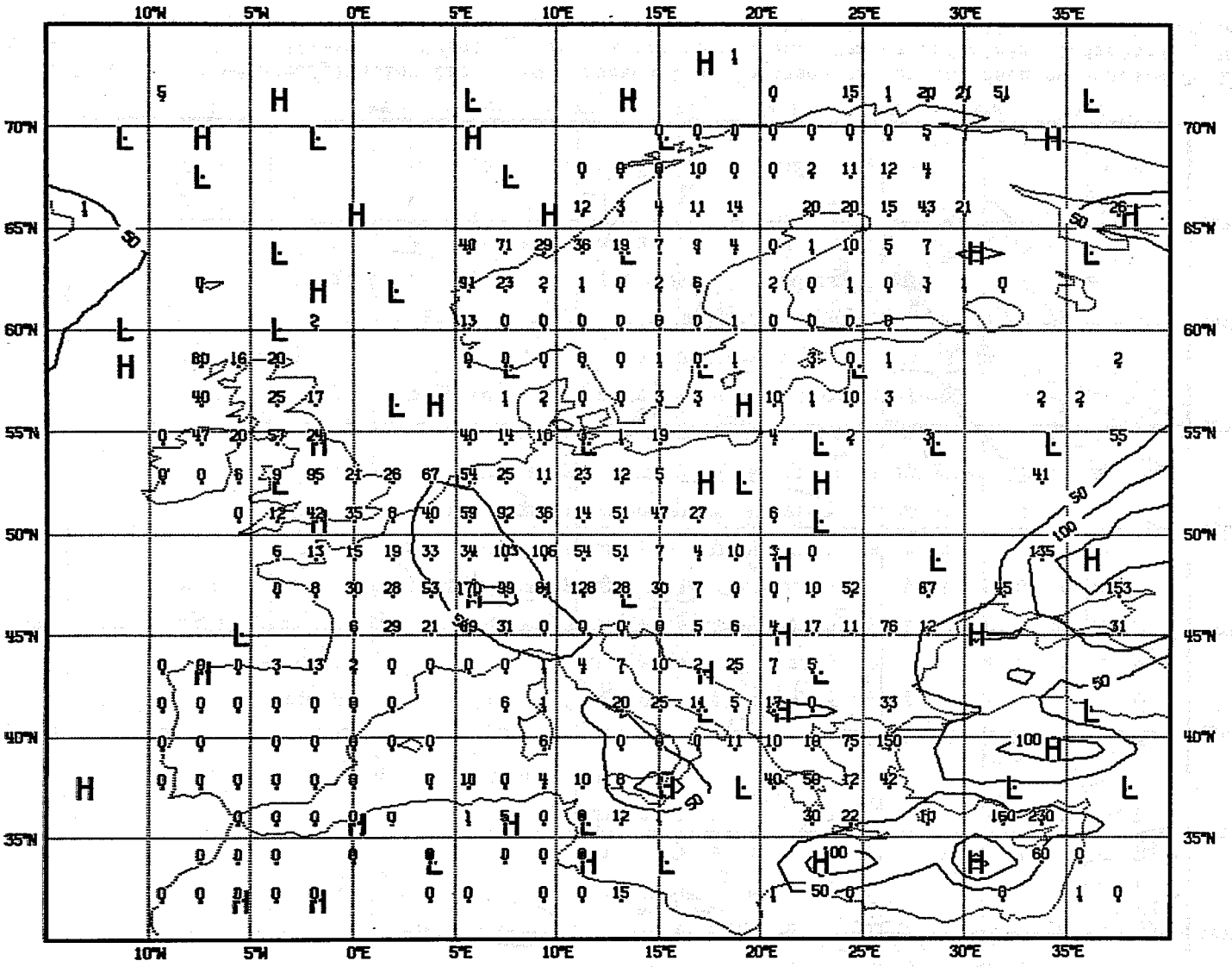

Fig. 2c Predicted and observed precipitation from 10-day forecast cycle based on initial data on January 3, 1981, 1200 GMT. Unit is one-tenth of a millimeter. Predicted values are contoured. Maxima and Minima indicated by H and L. Observed values are plotted at grid points, where available. 24-hour interval 1800Z on D2 to 1800Z D3. Contour interval 50.

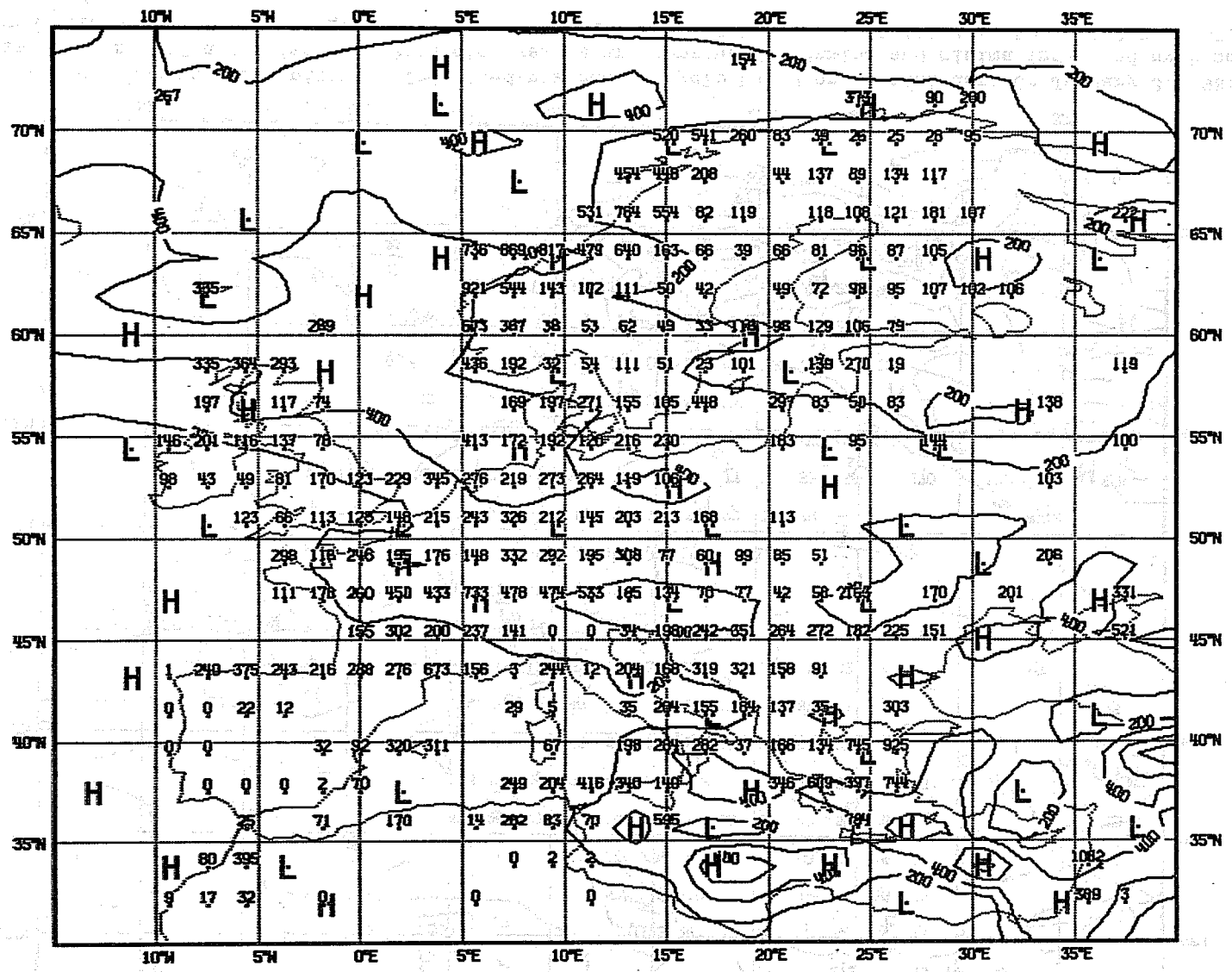

Fig. 2d Predicted and observed precipitation from 10-day forecast cycle based on initial data on January 3, 1981, 1200 GMT. Unit is one-tenth of a millimeter. Predicted values are contoured. Maxima and minima indicated by H and L. Observed values are plotted at grid points, where available. 9-day interval 1800Z on D0 to 1800Z on D9. Contour interval 200.

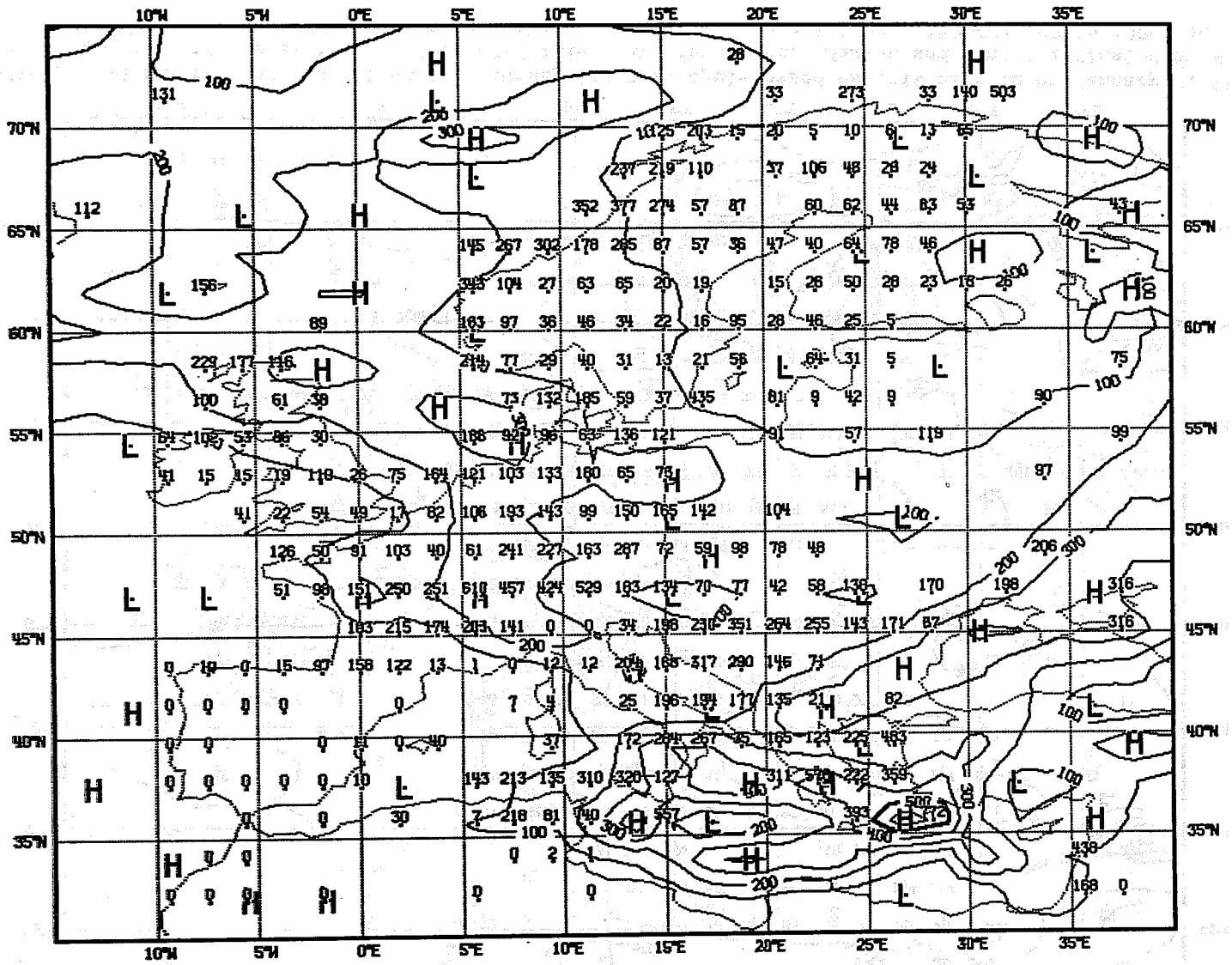

Fig. 2e Predicted and observed precipitation from 10-day forecast cycle based on initial data on January 3, 1981, 1200 GMT. Unit is one-tenth of a millimeter. Predicted values are contoured. Maxima and minima indicated by H and L. Observed values are plotted at grid points, where available. 5-day interval 1800Z on D0 to 1800Z on D5. Contour interval 100.

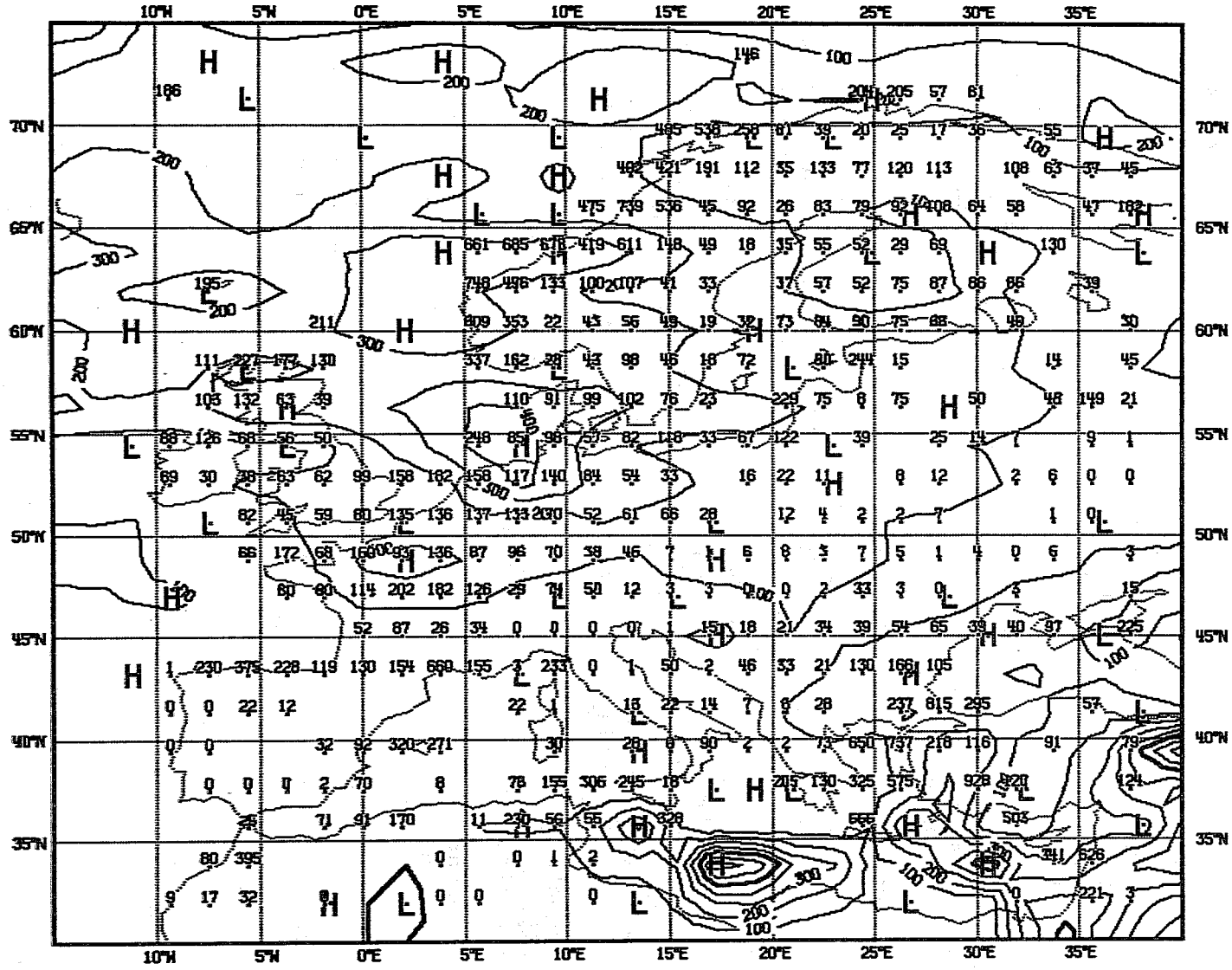

Fig. 2f Predicted and observed precipitation from 10-day forecast cycle based on initial data on January 3, 1981, 1200 GMT. Unit is one-tenth of a millimeter. Predicted values are contoured. Maxima and minima indicated by H and L. Observed values are plotted at grid points, where available. 5-day interval 1800Z on D4 to 1800Z on D9. Contour interval 100.

precipitation in each grid square, as will be explained more fully later.

The model generates readily new centres of precipitation. For instance, in the sequence shown, the activity over the Balkans developing on D2, intensifies, breaks up into cells, and moves south to North Africa on D6, then decays.

Figure 2d shows forecasts and verifying observations for the entire 9-day time interval from 18Z on D0 to 18Z on D9. Contour interval is 200 units. Figures 2e and 2f show the 5-day intervals D0 to D5 and D4 to D9.

3.2 Geographical Characteristics

The geographic features of the forecast fields can be brought out by averaging all forecasts for a month and applying a 5x5-point smoother to them. Figures 3a to 3d show the resultant fields for a winter month, January 1981, and a summer month, August 1980. The frame used is the verification area, 15°W to 39.375°E, 30°N to 75°N, containing 30x25 = 750 grid points. The 5x5 point smoother loses the 2-grid point wide border of the frame in Figs. 3a-d. The smoothed values are shown in the centre of each 5x5 grid point square. The number above the centre point represents the daily total amount of precipitation in units of 1/10 mm. The smaller, slanted number below the centre point represents large-scale precipitation as a precent of total precipitation. The full lines contour the former field, the dotted lines the latter. Contour interval is 5 for both fields.

January of 1981 had a persistent cold upper low over the eastern Mediterranean, giving higher than normal amounts of precipitation there. The contours for D1 in Fig. 3a reflect this. The contours for D9 in Fig. 3b show that the forecasting process tries to dissipate the cold low, or rather, the precipitation associated with it. In a later section, maps show the corresponding systematic errors for the month and show indeed that the area of the cold low in January 1981 developed an increasing negative bias (underforecasting precipitation) during the forecast cycle from D1 to D9. While D1 shows a near-zero bias, D9 shows underforecasting of 20-30 units per day in this area. This points to problems in sustaining precipitation in Mediterranean cold lows during the forecast cycle. The month of August 1980, in Figs. 3c-d, does not show significant dissimilarities between D1 and D9, neither do other summer months.

Since averaging occurs over 28 to 31 forecasts in any month (20 to 23 before August 1980), some differences will occur between maps showing D1 and D9 due to the changing synoptic situations. The D9 forecasts are for conditions which prevail 8 days later than conditions for the D1 forecasts. For January we are considering the changes in a 31-day moving average over a 9-day time span. In magnitude, this should be on the order of one-third of the climatic change from one month to the next,

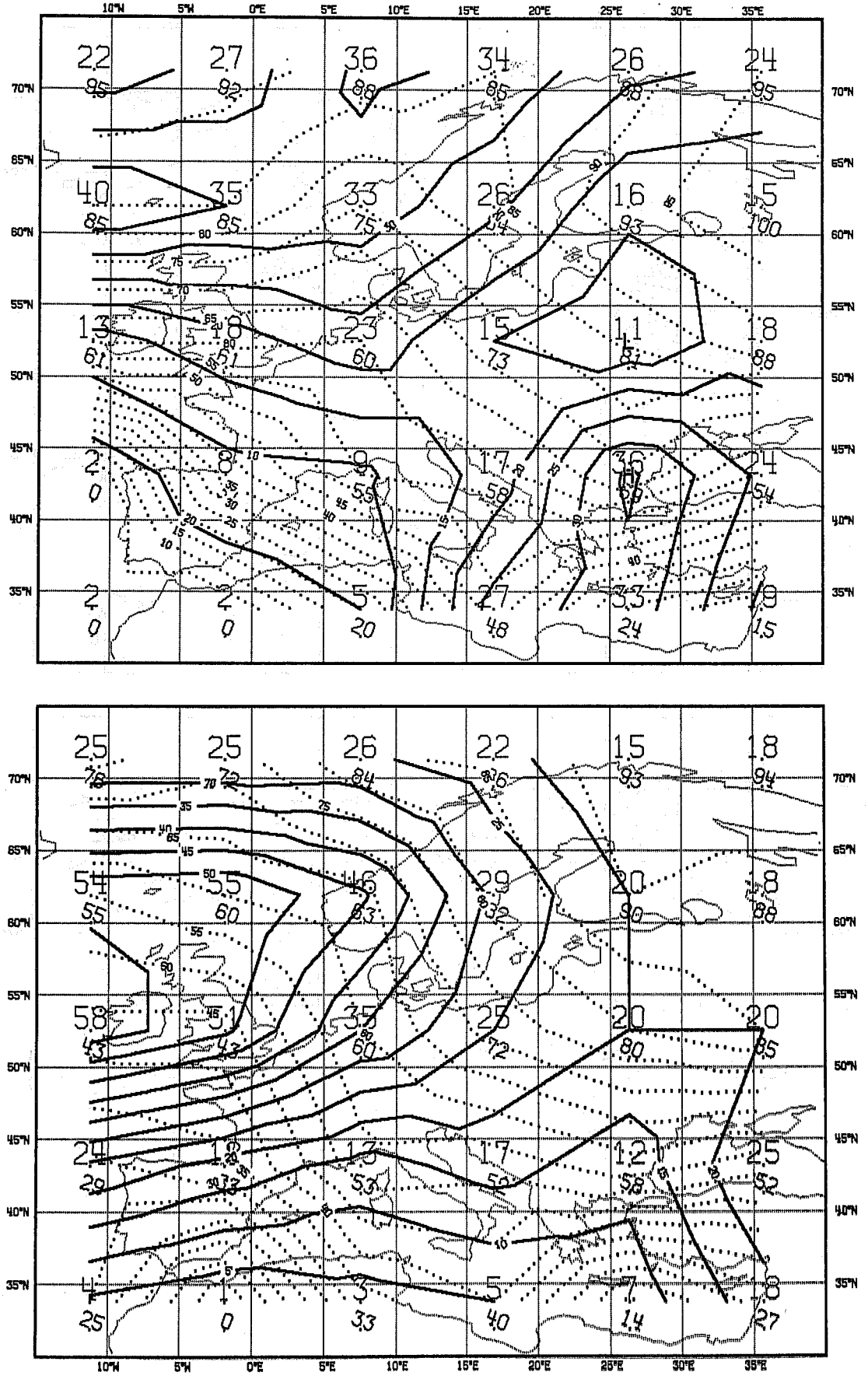

Fig. 3(a) & (b) Monthly average 24-hour precipitation. Full lines: Total precipitation = large scale + convective precipitation. Unit = 1/10 mm. Dotted lines: Large scale precipitation only as per cent of total. Contour intervals 5. January 1981, 31 forecasts. a) Top: D1. b) Bottom: D9.

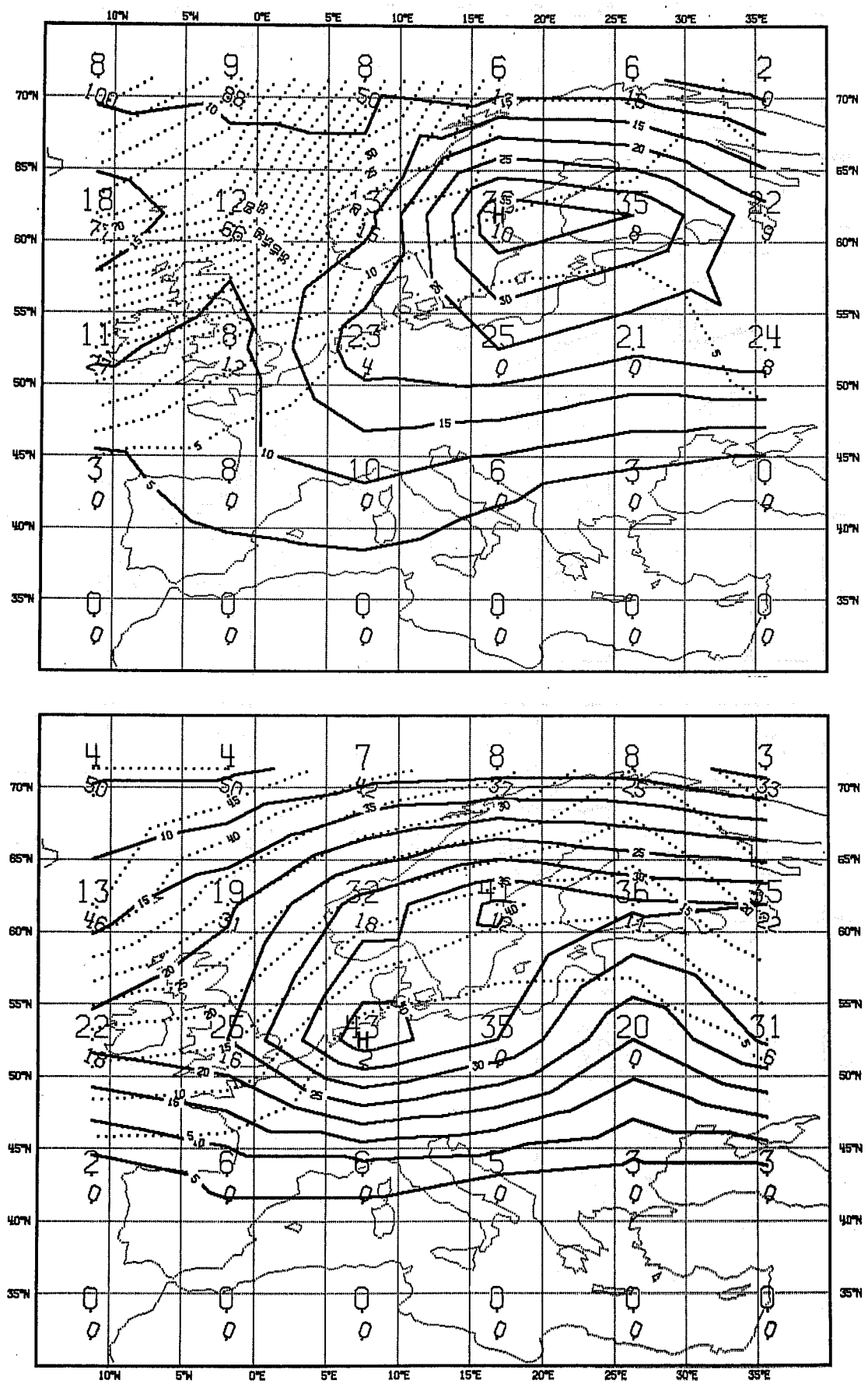

Fig. 3(c) & (d) Monthly average 24-hour precipitation. Full lines: Total precipitation = large scale + convective precipitation. Unit = 1/10 mm. Dotted lines: Large scale precipitation only, as per cent of total. Contour intervals 5. August 1980, 31 forecasts. c) Top: D1. d) Bottom: D9.

though we know that in individual years, the climatic changes may be well exceeded. The changes should not be as large as in Figs. 3a-b.

This kind of comparison of the space-and-forecast averaged prognostic fields for D1 and D9 might prove to be a useful tool in pointing out the systematic errors that develop during the forecast cycle and are tied to geography or to typical synoptic features.

Turning to the percent large-scale precipitation contoured by the dotted lines in Figs. 3a-d, they look very similar for D1 (3a) and D9 (3b) in winter. It should be kept in mind that where total precipitation is low, the percent figure for large-scale precipitation may be erratic, as it will be based on only the few cases where precipitation was forecast.

In January 1981, the percent large-scale precipitation remains high, near 100 percent, from D1 to D9 over the Arctic Ocean and over the winter time surface inversion covering the NE part of the verification area.

During August 1980, the percent large-scale precipitation for D1 (3c) starts the forecast cycle with near 100 percent in the NW corner. There is little penetration of large-scale precipitation into the continent. The 10 percent contour goes from south England, up the Skagerak, across south Sweden to the Bothnian Bay. Practically all the precipitation over the continent is convective, 80-90 percent over Scandinavia, 50-90 percent over the British Isles.

At the end of the forecast cycle, at D9, convection has developed over the NW oceanic area, making up about half of the total precipitation. The 90 percent convection contour (10 percent large scale) remains in the same position during the forecast cycle, except for some southward displacement over the Kola Peninsula.

3.3 Seasonal Changes, Forecast Cycle Changes

There is a tendency for increase in the forecast amounts of precipitation during the forecast cycle from D1 to D9. In Figs. 4a-b the abscissa is time from D1 to D9, the ordinate is total precipitation averaged over all 750 grid points of the verification area and over all forecasts of the months of January 1980 (Fig. 4a) and August 1980 (Fig. 4b). There is a systematic increase in both January and August 1980.

Averaged over the grid and the forecasts, the percent large-scale precipitation shows no such systematic change. As can be seen in Figs. 4c and 4d, which are for the same 2 months, the changes in this parameter are insignificant and appear random. They are typical for most months. They are no bigger than can be expected from normal changes in the synoptic situation during the month.

The tendency for increasing forecast amounts from D1 to D9 is present in almost all months. Fig. 4e sums up the essential characteristics of the area- and forecast-averaged cycle from D1 to D9 for 16 months from January 1980 to April 1981.

The forecasts for D1, averaged over all forecasts in a month and over all 750 grid points, show a steady amount in most months. The curve for D9 in Fig. 4e lies systematically above the curve for D1. Exceptions are June 1980 and March 1981. The average amount of increase from D1 to D9 is 33 percent. The increase is getting smaller towards the end of the verification period, starting with an increase of 100 percent in January 1980 and ending with an increase of 24 percent in April 1981. In April 1981, a new orography was introduced in the ECMWF model.

The systematic increase during the forecast cycle points to an imbalance in the model physics and dynamics which controls precipitation. One can only speculate what the causes might be and which one of the several factors -- evaporation, vertical boundary flux of humidity, vertical turbulent transport of humidity, vertical motion, convection scheme, condensation, and fall-out processes, or others -- are the causes of the imbalance. Lengthy trials are needed to isolate the causes.

The four other curves in Fig. 4e show the month-to-month changes in the forecast- and area-averaged large-scale percent. They show a large seasonal change from about 65 percent in winter to a minimum of about 15 percent in summer. The four curves represent D1, D9, the maximum value from D1 to D9, and the minimum value from D1 to D9. The four points cluster well together for any one month indicating little change during the forecast cycle. All-in-all, the partitioning into large-scale and convective precipitation appears to be stable with no tendency for systematic changes during the forecast cycle. The changes that do occur are of a magnitude that can be explained by the randomness of synoptic sampling.

Further aspects of the forecast fields will be discussed in later sections where they are examined in light of observational evidence.

4. OBSERVATIONAL DATA BASE

The verifications in this report are based on synoptic surface reports received over GTS at ECMWF. The ECMWF archiving system stores all such reports in the Reports Data Base (RDB).

The verification area was selected by two main criteria, (1) interest spheres of the ECMWF member nations and (2) data availability in the RDB. Europe is one of the few areas where data in the RDB are dense enough and quality of reports high enough to do a decent job of verifying quantitative precipitation forecasts on an extended area basis.

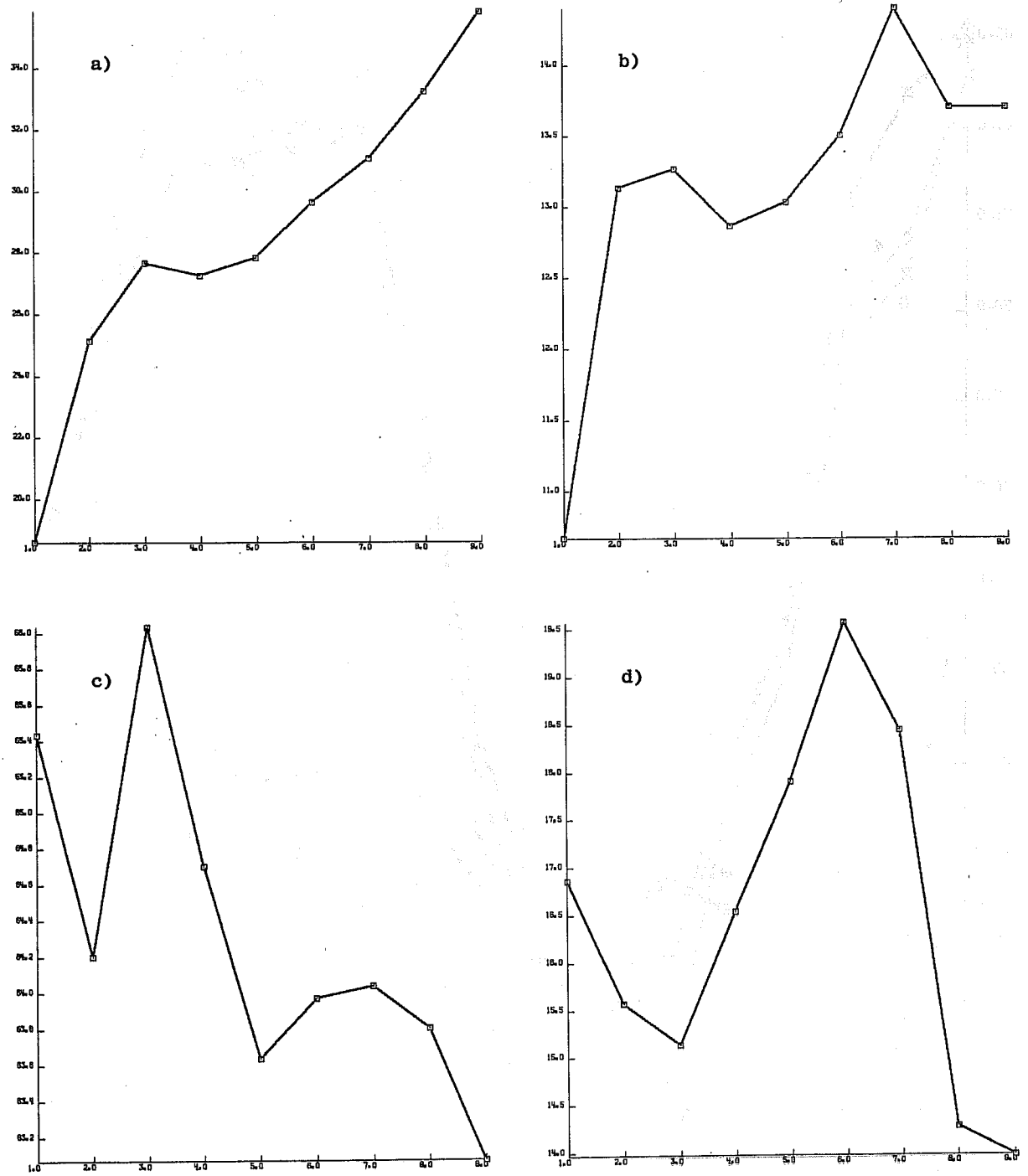

Fig. 4(a)-(d) Changes during forecast cycle from D1 to D9 of average 24-hour values of:
(a) & (b) forecast total precipitation, units 1/10 mm per day, January and August 1980, respectively
(c) & (d) forecast large-scale precipitation as per cent of total, January and August 1980, respectively.

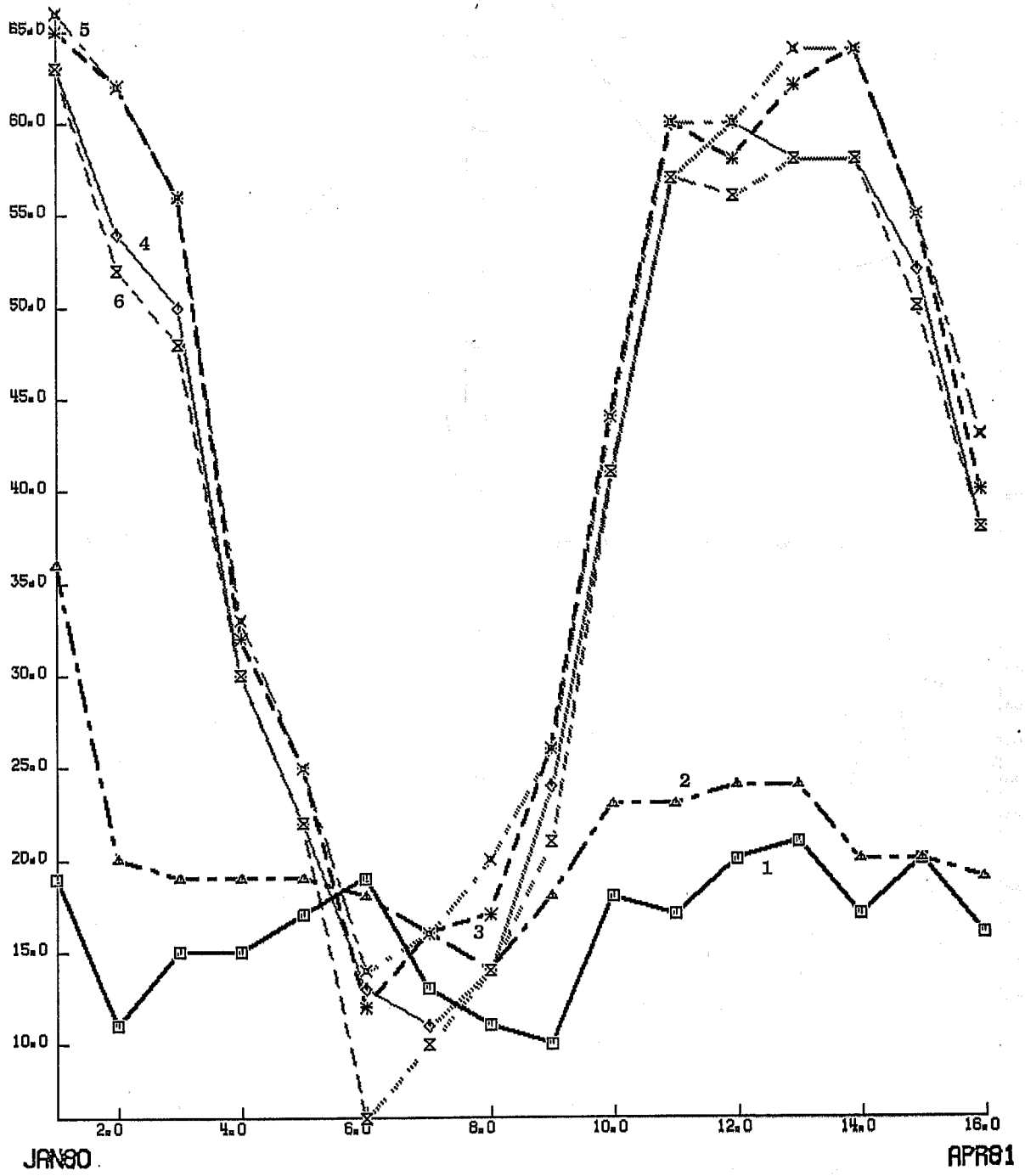

Fig. 4 (e) Monthly average of forecast 24-hour precipitation for the 16-month period January 1980 to April 1981, average of all 750 grid points. Curves are for: (1) Total precip for D1 (1/10 mm per day). (2) Same for D9. (3) Large scale precip as per cent of total for D1. (4) Same for D9. (5) Maximum of same during forecast cycle. (6) Minimum of same during forecast cycle. Curve symbols explained on diagram. Abscissa: Number of month starting with January 1980.

4.1 Masterlist of Reporting Stations

The RDB was polled for reports inside the verification frame during a 7-day period starting January 1, 1980. Each station which appeared at least once during this period at 00, 06, 12, or 18Z was included in a masterlist of stations and is shown in Fig. 5. There are 1318 stations altogether. Fig. 6 shows the number of stations in each of the 1.875 degrees longitude by 1.875 degrees latitude boxes centered on the 644 internal grid points. The box area shrinks systematically towards the north. Of the 644 boxes, 221 are empty, mostly oceanic boxes. The remaining 423 boxes are used in verification. A few stations lie in the half-grid-length wide border. They are assigned to the nearest box, so each border box has observations from a larger area than others. On the average, a box contains 3.12 stations. One hundred, fifteen boxes contain only one station. The maximum density is around the grid point located in northeast Switzerland where there are 21 stations in the box.

4.2 Report Rate

Data reception at ECMWF is very good. Table 2 shows statistics of the synoptic reports received during the 38-day period from September 2 to October 9, 1980, which were used in verifying the September 1980 forecasts. Of the 1318 stations in the masterlist 70, 90, 90, and 90 percent were received at 00, 06, 12, and 18Z, respectively. Of these, only 2 percent had missing precipitation groups, resulting in an overall availability of precipitation reports of 88 percent of the masterlist. The precipitation groups used are sent at 06 and 18Z. The verification area falls mostly in WMO Region VI (Europe) and the rest in Region I (Africa). There are regional and national options in reporting precipitation. All stations in Region VI report 12-hour precipitation at 06Z and 18Z. Some nations in addition report 6-hour precipitation at 00Z and 12Z. At these stations, verification time periods can be constructed to coincide with the times 00Z and 12Z which are the times of the archived ECMWF forecast fields. However, if we limit the verification to these stations, the data base is reduced. For this reason the time period chosen for verification is the day from 18Z to the next 18Z. The associated forecast fields are linearly interpolated from the archived 00Z and 12Z fields.

TABLE 2: Data Reception During 38 Days, September 2-October 9, 1980

	SYNOP REPORTS	PRECIP REPORTS
00Z	34939 (69.8%)	-
06Z	45108 (90.0%)	44007 (87.9%)
12Z	45214 (90.3%)	-
18Z	45187 (90.2%)	44095 (88.0%)

Region I stations also report at 06Z and 18Z, but the reports at 06Z are 24-hour preceding precipitation, while the 18Z

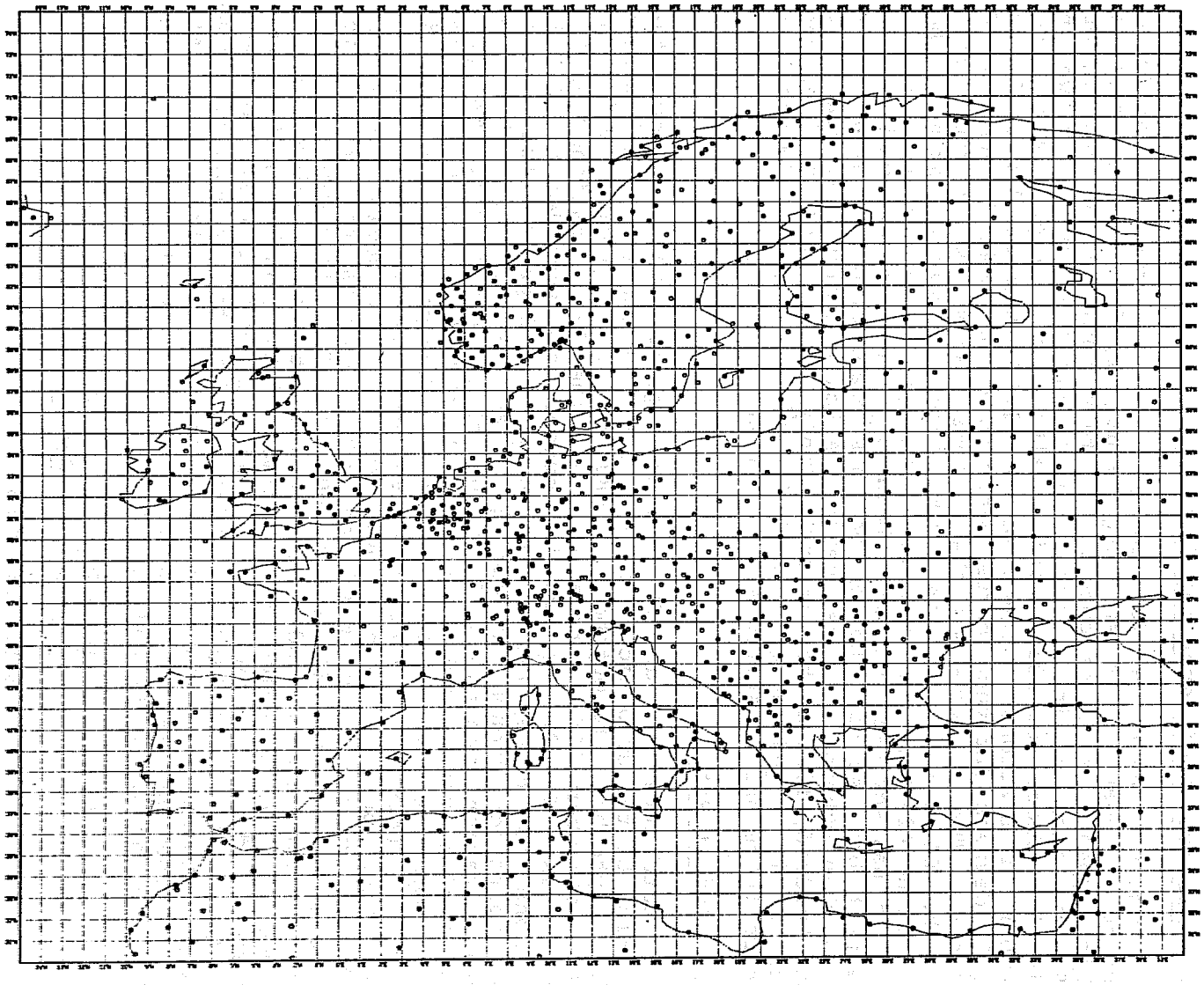

Fig. 5 1318 reporting stations used in verification.

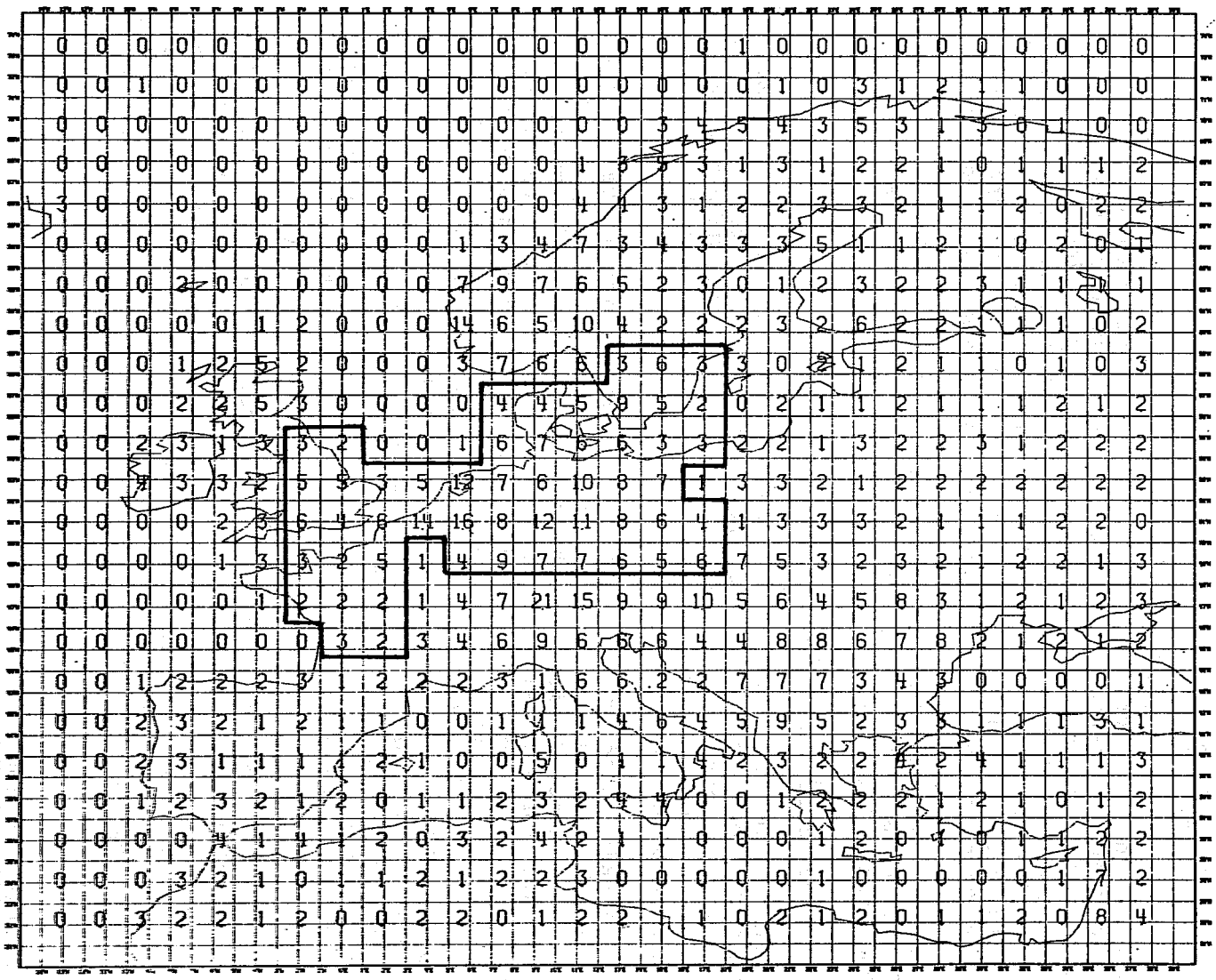

Fig. 6 Reporting stations partitioned in 1.875° longitude by 1.875° latitude boxes centered on grid points. Number of stations per box. Heavy Line: Window of flat, well observed boxes.

reports are 12-hour preceding precipitation as in Region VI. There are some national exceptions to this schedule.

4.3 Errors in Reports

The archived precipitation data have not been quality checked. An error check of sorts can be done without referring back to original national records. Most erroneous reports consist of unreasonably large amounts of precipitation as a result of coding or telecommunication errors. All reports of precipitation of 30 mm or more were examined by the processing program and compared with the associated synoptic elements in the report, such as temperature, present weather, past weather, and type of low cloud. In a typical verification month there would be about 200 such examinations, i.e., about 0.2 to 0.3 percent of 90,000 reports. Of these, only 10-20 would have to be rejected on the basis of element comparison.

5. PROJECTION OF OBSERVATIONS TO FORECASTS

There are 115 boxes that contain only one station. Much depends on how representative those stations are for the area. We know that orographic exposure is important. Unfortunately, there is no simple way in which a normalization can be effected based on orography.

The value used in this report, as observation to be compared with the forecast value at the grid point, is the simple average of the reporting stations in the associated box. The reports at 06Z and 18Z for each station are first added to give the 24-hour amount. If one or both the synoptic hours are missing for a station, that station is rejected for the averaging process for that day. For instance, in a one-station box, where either the 06Z or the 18Z is missing, no comparison is made. The remaining stations are averaged, and the resulting amount is compared with the forecast amount valid from 18Z to the next 18Z. The latter amount is obtained by a linear interpolation between the three consecutive, 12-hourly spaced forecast fields which bracket the verification period in question.

To denote the time period of a forecast, two notations are used interchangeably in this report. D1 and D0-D1 both mean the same. D1 (and D0-D1) is the time period from 18Z on the day of the forecast to 18Z the following day. Since initial time is at 12Z, a forecast for D1 (or D0-D1) is a 30-hour forecast. Similarly, D9 (or D8-D9) is the time period from 18Z on the 8th day after initial time to 18Z on the following day, and represents a 24x9+6 = 222-hour forecast.

6. SCORING METHODS

Interest in precipitation forecasts takes many forms. There is the farmer who is interested in knowing if and when he is going to have a 3-day dry spell in which to dry his hay during the next 10 days. There is the farmer who wants to know if he

should use costly irrigation water or wait for ample natural rain in the coming week. There is the administrator of a hydro-electric power plant who wants to know the prospects of replenishing the reservoirs in the near future. There are the arrangers of sports events and other outdoor functions who just want to know whether it will rain at a specific time. Whatever the interest, there is a verification format which is appropriate to it, and none that serves all purposes.

With the quantitative forecasts that are available at ECMWF out to 10 days, it is possible to be both general and specific and develop verification statistics that can be applied to many problems.

The measures of skill in forecasting used in the following are intended for general use. It may be necessary to go back to the primary product of the verification process, the contingency matrix, in order to derive verification statistics tailored for specific applications. Further, it may be necessary to develop corresponding subsets of the contingency matrix for the whole European area in order to define better the skill of forecasts for specific geographic areas.

6.1 Class Intervals and Contingency Tables

For the purpose of verification, predicted and observed precipitation amounts have been divided into eight categories as shown in Table 3.

TABLE 3: Classification of Precipitation Amounts

Category	Limits: Unit, 1/10 mm
1	<1 (not measurable)
2	1-6
3	7-20
4	21-50
5	51-100
6	101-150
7	151-200
8	>200

The class intervals have been chosen mainly by the options offered by the reporting practices. The reports are coded as a 2-digit number according to a pseudo-logarithmic scheme. At the lower end of the spectrum the reporting interval is 1/10 of a mm, at the upper end 1 cm. From 0.1 mm to 0.6 mm the increment is 0.1 mm. In the next interval from 1 mm to 55 mm the increment is 1 mm. The last interval from 6 cm to 40 cm has an increment of 1 cm. One can assume that the coding practices by and large reflect user requirements, so Table 3 reflects operationally significant classes.

This is important for application of any verification statistics to the problem of planning how to use meteorological forecasts.

The basic products of verification are contingency tables of forecast categories cross-referenced to their associated observed categories. Fig. 7 shows as an example the contingency table for ECMWF forecasts of precipitation for D1 (or D0 to D1) for the month of September 1980. The numbers have been normalized by the total number of cases, 11440. A case is a forecast for a specific box. Since there were 29 forecasts verified in September 1980 and 423 boxes, ideally the total should be their product, 12267. Because of missing observations, 7 percent of cases were lost. Column 9 represents the total frequency of forecasts in each category and row 9 total frequency of observations in each category.

Similar contingency matrices have to be constructed for the other one-day forecast intervals in the range D1 to D9, and for multi-day intervals, such as D0 to D5 and D4 to D9. Similar sets have been recorded or computed for each of the control forecasts, such as climate forecasts, random forecasts, and dry forecasts. These forecasts are defined in the following.

While there is no substitute for the contingency table as a basic documentary of the verification process, an 8x8 matrix is unwieldy to use when comparing forecast systems. There is a requirement to summarize the matrix to single numbers for several purposes, such as comparing performance of forecasting systems or models, long-term monitoring of performance, or making simple assessment of usefulness of forecasts.

In many cases it will be found that it is necessary to turn to the contingency tables to gain assessment of usefulness. Summary indices can be deceptive.

Three scoring methods are used in this report, the RMS Class Error (RCE), the Heidke Skill Score (HSS), and the Threat Score (TS). The RCE is the equivalent of the RMS error of a continuous variable. Precipitation amounts can of course be treated as a continuous variable, but because of the masking influence which large amounts of precipitation would have on the statistics, it is preferred to express the statistics in terms of the classes in Table 3. The HSS concerns itself with the number of correct forecasts only (hits) and disregards the distribution of the rest of the forecasts in the matrix. It is a relative score that comes out positive or negative according to whether the rated forecasts have more or fewer correct forecasts than the comparison forecasts. The TS is concerned with two-category events, such as rain -- no rain, less than -- more than 15 mm of precipitation, etc. The TS accounts for the number of cases when the correct choice is made, but guards against achieving high scores by overforecasting.

6.2 RMS Class Error (RCE)

Turning to Fig. 7, the frequencies of the correct forecasts are tabulated along the diagonal from the upper left corner of the 8x8 matrix. Drawing lines parallel to the diagonal on both

CONTINGENCY TABLE OF OBSERVED (HORIZONTAL) VERSUS FORECAST (VERTICAL) PRECIP FROM D0 TO D1 18Z IN 8 CLASS INTERVALS

29 FORECASTS 9/1980 EUR UNITS: FRACTION OF TOTAL NUMBER OF BOX FORECASTS 11440

	.LE. 0	.LE. 6	.LE. 20	.LE. 50	.LE. 100	.LE. 150	.LE. 200	.LE.2046	TOTAL FORECAST
.LE. 0	.4565	.0640	.0349	.0231	.0108	.0030	.0012	.0011	.5945
.LE. 6	.0544	.0272	.0229	.0170	.0084	.0023	.0011	.0011	.1344
.LE. 20	.0306	.0212	.0209	.0176	.0108	.0032	.0015	.0011	.1068
.LE. 50	.0150	.0156	.0187	.0234	.0128	.0045	.0015	.0016	.0930
.LE. 100	.0032	.0057	.0091	.0150	.0117	.0042	.0017	.0010	.0517
.LE. 150	.0005	.0012	.0024	.0039	.0035	.0016	.0004	.0008	.0143
.LE. 200	.0001	.0003	.0003	.0006	.0009	.0011	.0003	.0003	.0039
.LE.2046	.0000	.0001	.0001	.0001	.0003	.0005	.0002	.0001	.0014
TOTAL OBSERVED	.5603	.1851	.1093	.1007	.0591	.0204	.0080	.0072	1.0000

Fig. 7 Example of contingency tables as basic product of verification process. The 8 class intervals of precipitation amounts are given in Table 3. Forecasts are arranged vertically, observations horizontally. Each case of a forecast for a grid point and verifying observations is entered in the appropriate location of the 8 x 8 matrix. Cases are summed for the month and normalized to fractions of unity. Column 9 has totals of each forecast class, row 9 has totals of each observed class, forming a 9 x 9 matrix.

sides and summing the 14 frequencies along them, we obtain the total frequency of the forecasts that miss by one class interval. The next pair of lines has 12 frequencies and this sum gives total frequency of forecasts that miss by 2 class intervals, and so on. Each sum is weighted by the square of the class departure, i.e., 1, 4, 9, 16, 25, 36, 49. The square root of the sum of the weighted sums is the RCE. The range of the RCE is 0 to 7, 0 for perfect forecasts.

6.3 Heidke Skill Score (HSS)

This score measures the correct forecast only, not as an absolute value but in relation to the correct forecasts of a competing forecast system. We shall refer to the competing forecast system as the control system. If we express the sum of the frequencies of the correct forecasts of the system we want to verify, by C and the corresponding sum for the control forecasts as C*, the Heidke Skill Score is:

$$HSS = \frac{C-C^*}{1-C^*}$$

Perfect forecasts give HSS=1. HSS becomes negative when there are fewer correct forecast than in the control forecasts.

The frequency of correct forecasts can be obtained from the contingency tables by summing along the diagonal of Fig. 7.

$$C = \sum_{i=1}^{8} f(i,i)$$

where f(r,c) is the frequency element in the r-th row and c-th column of the 9x9 frequency matrix.

6.4 Threat Score (TS)

The Threat Score (TS) is another summation of the contingency table, sometimes used in verification (Charba and Klein, 1980). For our purpose it is defined as:

$$TS(N) = \frac{H(N)}{F(N) + O(N) - H(N)}$$

where TS(N) is the Threat Score of forecasting precipitation exceeding the upper limit of class N, F(N) is number of forecasts of precipitation exceeding the same limit, O(N) the number of observations exceeding the same limit and H(N) the number of hits or cases common to set F(N) and set O(N).

Referring to Table 3, TS(1) is the Threat Score of forecasting precipitation of any measurable amount (rain/no rain),

TS(4) of forecasting precipitation of more than 5 mm, and TS(6) of forecasting precipitation exceeding 15 mm.

Using the previous notation the quantities can be computed from the contingency tables as:

$$F(N) = \sum_{r=N+1}^{8} f(r,9)$$

$$O(N) = \sum_{c=N+1}^{8} f(9,c)$$

$$H(N) = \sum_{r=N+1}^{8} \sum_{c=N+1}^{8} f(r,c)$$

No hits gives TS=0, perfect forecasts gives TS=1.

The Threat Score is a useful measure of skill for operations which require knowledge of when precipitation will exceed certain limits. This includes the limit zero which is the case of forecasting "precipitation" or "no precipitation."

6.5 Control Forecasts

It is convenient to use forecasts produced by simple systems as control forecasts. Examples are climate forecasts, random forecasts, persistence forecasts, and forecasts of "zero precipitation."

Common to all these is the ease and low cost with which they can be generated. One might say that a forecast centre, in order to deliver useful forecasts, should at least show better results than these simple control forecasts. As later examples will demonstrate, useful skill by this criterion depends very much on which score is used.

Each of the control forecasts needs a closer definition, spelling out how they are generated.

6.5.1 Climate Forecasts

As used in this report, a climate forecast for a time interval is equal to the proportional part of the normal rainfall for that month. For instance, if the time interval for which the forecast is made is a day, the climate forecast is the climatological monthly mean precipitation for that month and area, divided by the number of days in the month. The same value is used whether the time interval occurs early or late in the month. Forecasts for a 5-day interval would be 5 times this amount.

While the definition is straightforward, when it comes to practical application interpretations have to be made. The observations used for verification are from a few synoptic stations, sometimes only one per grid point. In order to make a climate forecast as strong a competitor as possible, the climate for the box should be based on those same stations. Long-term precipitation normals for the stations in the masterlist have not been available. Instead, a substitute box climatology has been constructed from the data in the RDB.

The seasonal average of precipitation for each of the 423 boxes of Fig. 6b was computed from the archived data for 1980. Each of the seasons consisted of 3 months, starting with January 1980. The climate forecast for a month is taken as the seasonal value. Seasonal statistics were preferred to monthly statistics to avoid the effects of large month-to-month fluctuations as may occur in one-year statistics and also to eliminate as much as possible the effect of dependency. The "climate" had to be established from the same year of data which was used for verification. If monthly statistics had been used, climate forecasts would have had an unfair advantage. A much broader basis for the climate should be established as the archives expand. Five years would form a more satisfactory basis.

The values used in this report for the 24-hour interval climate forecasts are shown in Figs. 8a-d.

6.5.2 Random Forecasts

Random forecasts are forecasts picked at random out of a given frequency distribution of the 8 classes. The strongest competition from a random forecast would probably come from a climatological distribution particular to each of the 423 boxes and based on the masterlist. As the data base is presently inadequate to establish a climatological distribution independent of the data used in verification, the distribution used in this report is the distribution common to all the 423 boxes in the month being verified.

This gives the random forecast the advantage of using the distribution for the verified month (which really is not available at the time of the forecast), but the disadvantage of being valid for the ensemble of 423 boxes but not necessarily for a particular box. The two factors may cancel each other.

Again, the longer term distribution should be established as data become available.

The frequency of correct forecast obtained by our random experiment can be written as:

$$\sum_{c=1}^{8} [f(9,c)]^2$$

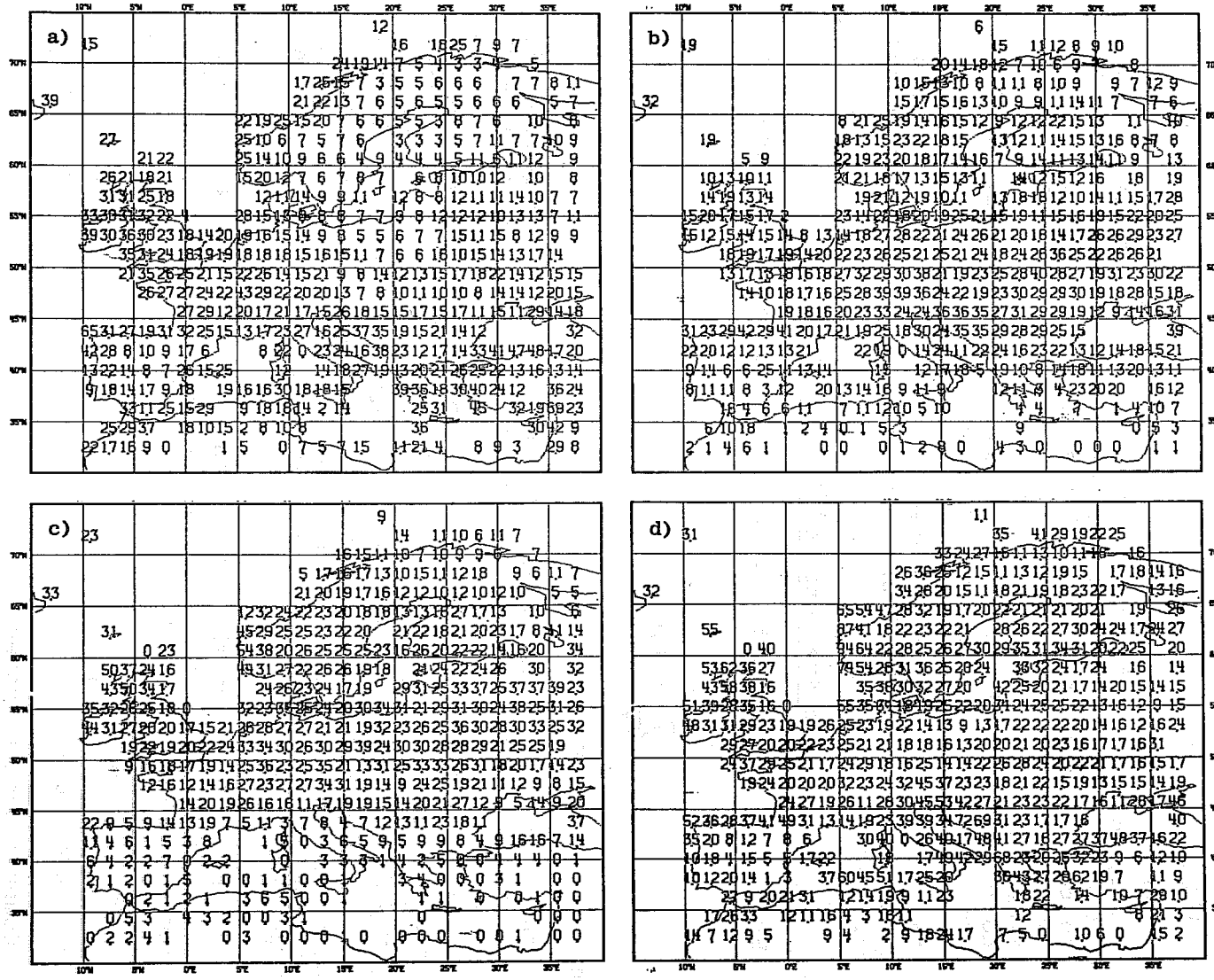

Fig. 8 Average daily precipitation used as climate forecasts. Unit = 1/10 mm per 24 hours. Derived from ECMWF archives for 1980 for stations in Fig.5. (a) Jan/Mar. (b) Apr/Jan. (c) Jul/sep. (d) Oct/Dec.

where f(r,c) is the frequency element in row r and column c in the matrix depicted in Fig. 7.

6.5.3 Dry Forecasts

These denote forecasts of "zero precipitation" at all times. They verify well in a dry month or in areas with a dry climate. Measured by the HSS, the dry forecasts for a one-day duration are hard to beat in most months and boxes, as 50 percent or more of the days are often dry and hence give over 50 percent correct forecasts. Dry forecasts for an interval of several days duration do not do so well as dry forecasts for a one-day duration since dry spells lasting several days are rarer.

The number of correct forecasts by this forecast method is simply the frequency f(9,1) using the same notation as above and referring to Fig. 7.

6.5.4 Brier's Chance Frequency of Correct Forecasts

Brier and Allen (1951) give as the expected frequency of correct forecasts by chance:

$$\sum_{i=1}^{8} f(i,9)\ f(9,i)$$

using the above notation.

This frequency, called here the Brier Chance Frequency, is different from the one obtained in the random experiment described in 6.5.2. The Brier Chance Frequency is the one that would result if the forecasts were unchanged (not random) but the observations were picked out at random from their own total frequency distribution. One might say it represents the frequency obtained if the observations of the month were randomly distributed (uncorrelated with the forecasts) with a total distribution equal to the observed one. This is not a control forecast in the usual sense of the word. The observations are randomized rather than the forecasts.

The HSS's based on the Brier Chance Frequency have been computed and are shown in some of the score comparisons. The monthly verifications show that both random forecasts and dry forecasts have more correct forecasts than the Brier Chance Frequency. In other words, the HSS of a forecast system looks better when based on Brier Chance Frequency than when based on dry or random forecasts.

6.6 Computation of Scores

In order to compute the RCE, HSS, and TS for comparison with climate forecasts, it is necessary to construct a separate contingency table for the climate forecasts as the climate forecasts vary from box to box.

The contingency table for random forecasts and dry forecasts are given by their definitions, in both cases using the total distribution of the observations.

Returning to Fig. 7, the contingency table for dry forecasts for the same month would have row 9 of Fig. 7 as row 1 and row 9, and all other elements would be zero.

The contingency table of the random forecasts for the same month would have elements:

$$FR\ (r,c) = f(9,r)\ f(9,c)$$

where FR (r,c) is the frequency of the element in row r and column c of the random contingency table and f (9,i) is the total frequency of observations in class i or the element in row 9 and column i of Fig. 7.

7. RESULTS

The immediate products of the verification process are maps of observed and forecast values of precipitation, such as those shown in Figs. 2a-f. The results can be summed box by box for all the forecasts of the month and separately for D1, D2,...D9. Figs. 9a-h show the outcome of this averaging process for a winter month and a summer month.

7.1 Systematic Errors

Fig. 9a shows the monthly average for the 31 forecasts for D1 made in January 1981. The average daily values of forecast precipitation for the month are given by the larger numbers printed above the grid points. The smaller numbers below are the verifying observed values. Their differences, the biases, are plotted on Fig. 9b.

Similarly, Figs. 9c-d give forecast, observation, and bias for the 423 boxes for D9.

The main features of the systematic error maps for the month are the areas of underforecasting (negative bias) of precipitation over and to the windward of mountainous areas, such as in Scandinavia, Scotland, Wales, Spain, Morocco, and the Alps.

Averaged over all the boxes, forecasts and observations are well balanced during the forecast cycle. The bias is relatively small and typically starts out as slight underforecasting for D1 and ends up as slight overforecasting for D9. Systematically growing bias is the result of an increase in the forecast value. This can be seen in Fig. 10 which depicts the month-to-month values of forecast, observation, and bias for D1 and D9. There is no tendency for the observations to differ much from D1 to D9 except such vagaries as one would expect from the fact that on

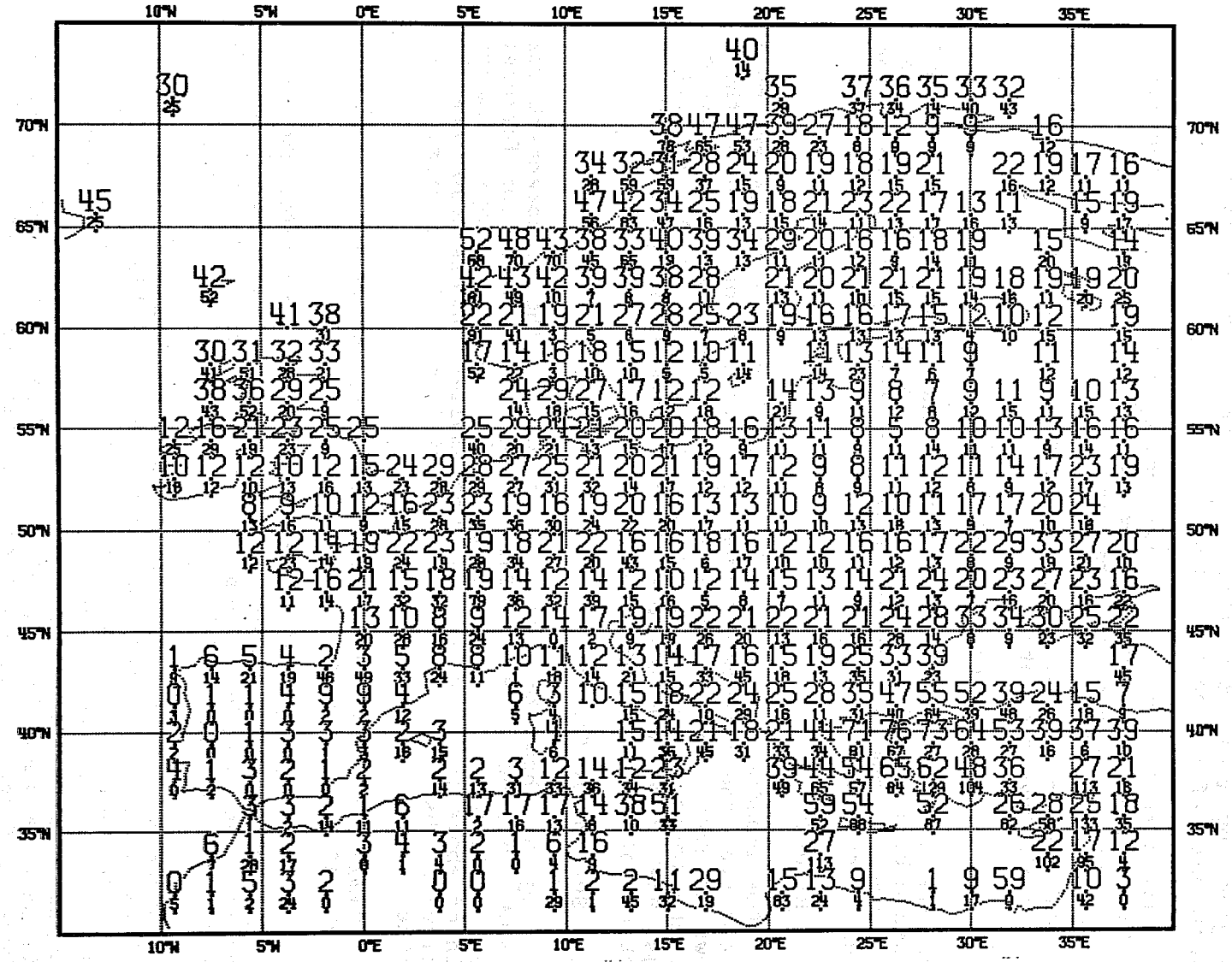

Fig. 9a Monthly mean of daily values at grid points of forecast observation. Unit 1/10 mm per 24 hours. Large number above grid point is forecasts, smaller number below grid point is observation.
D1, January 1980, 31 forecasts.

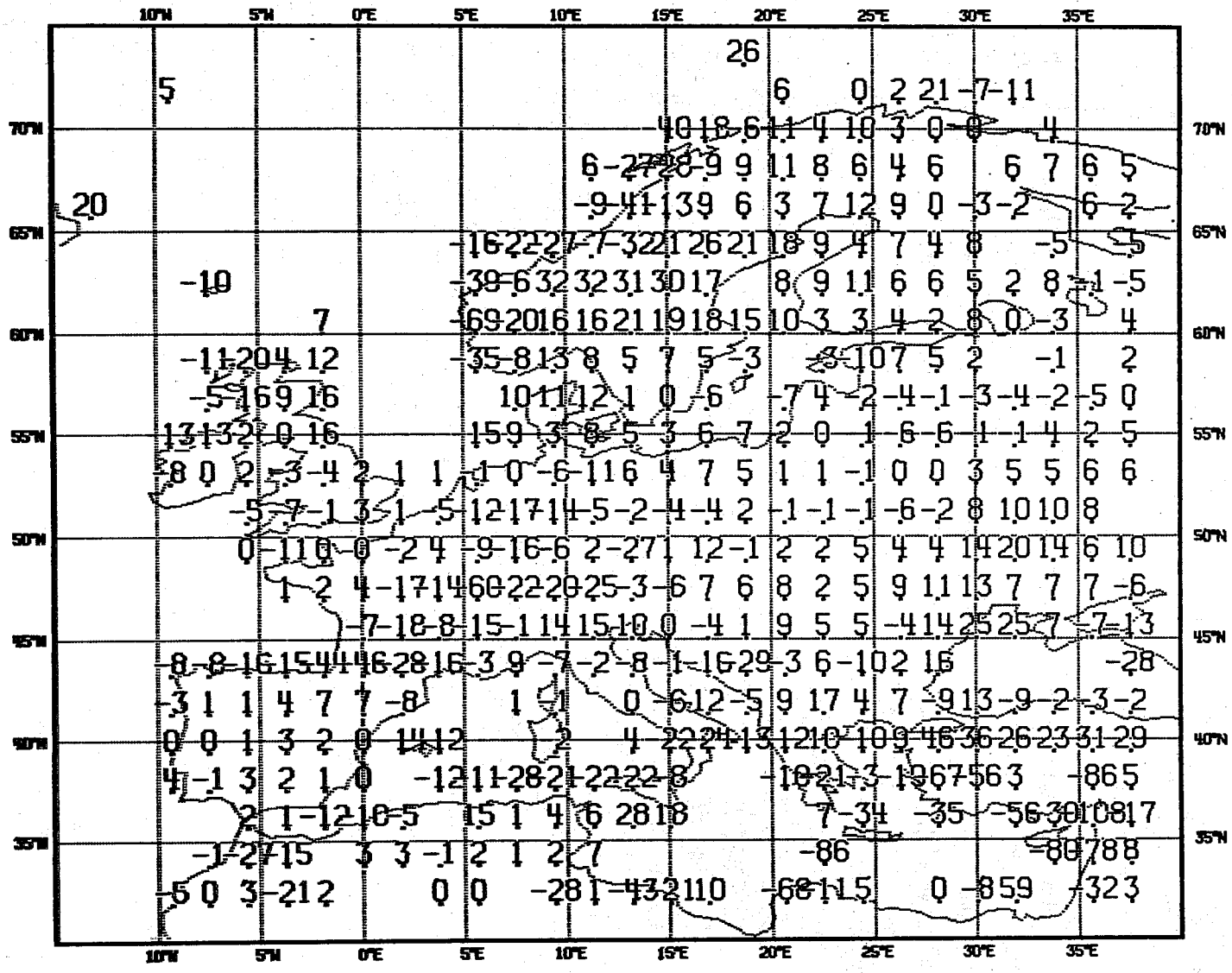

Fig. 9b Bias for Fig. 9a.

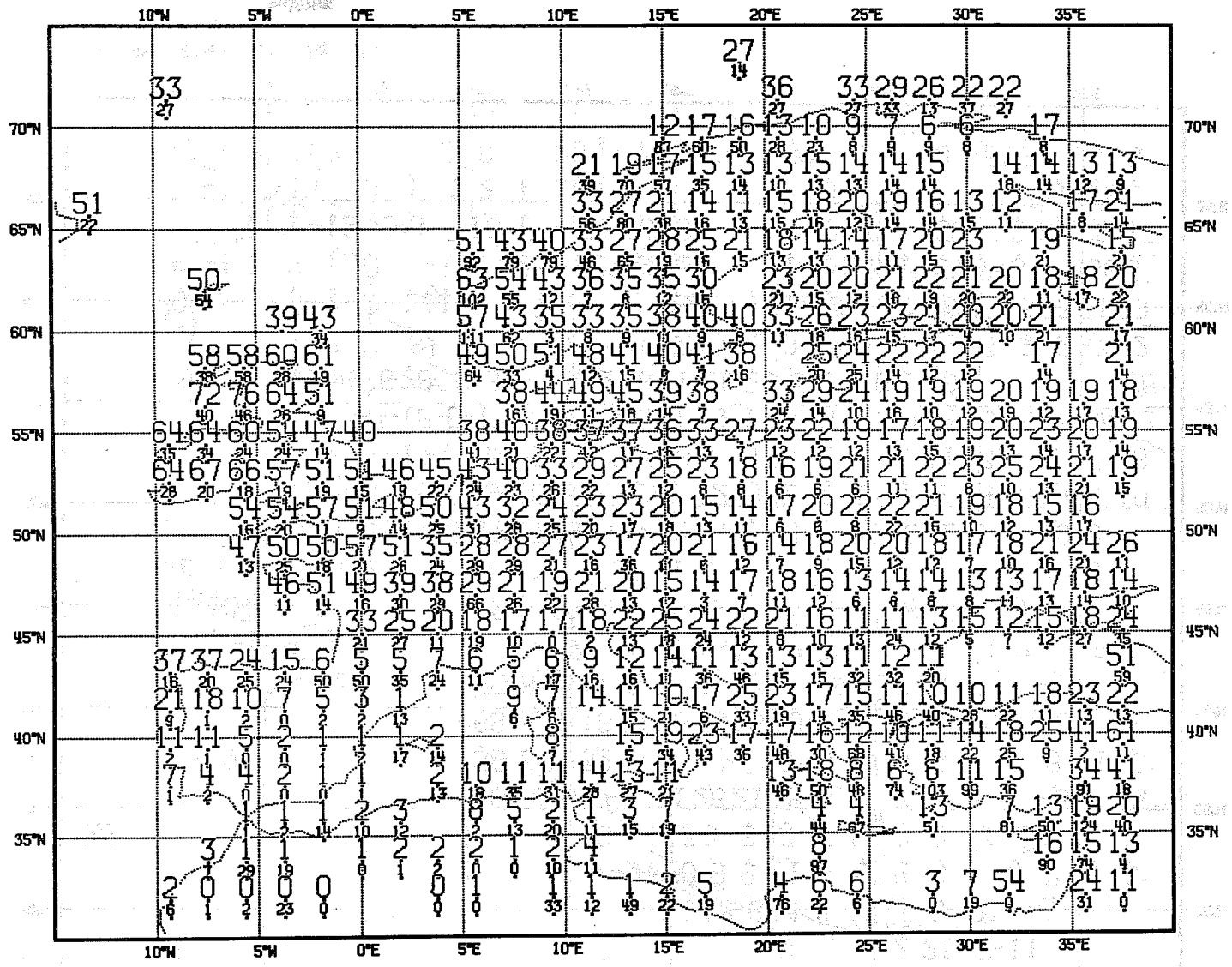

Fig. 9c Monthly mean of daily values at grid points of forecast and observation. Unit 1/10 mm per 24 hours. Large number above grid point is forecasts, smaller number below grid point is observation.
D9, January 1980, 31 forecasts.

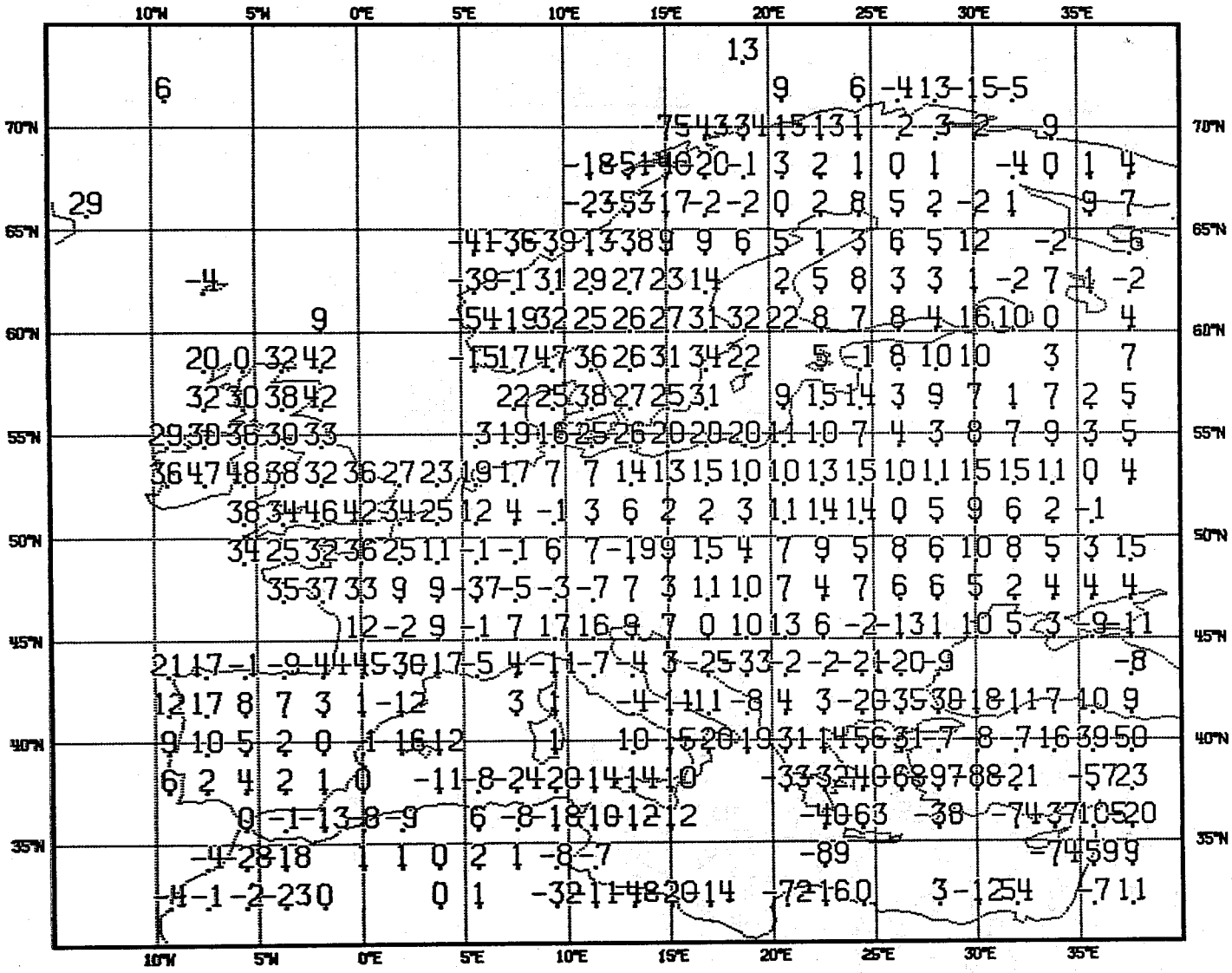

Fig. 9d Bias for Fig. 9c.

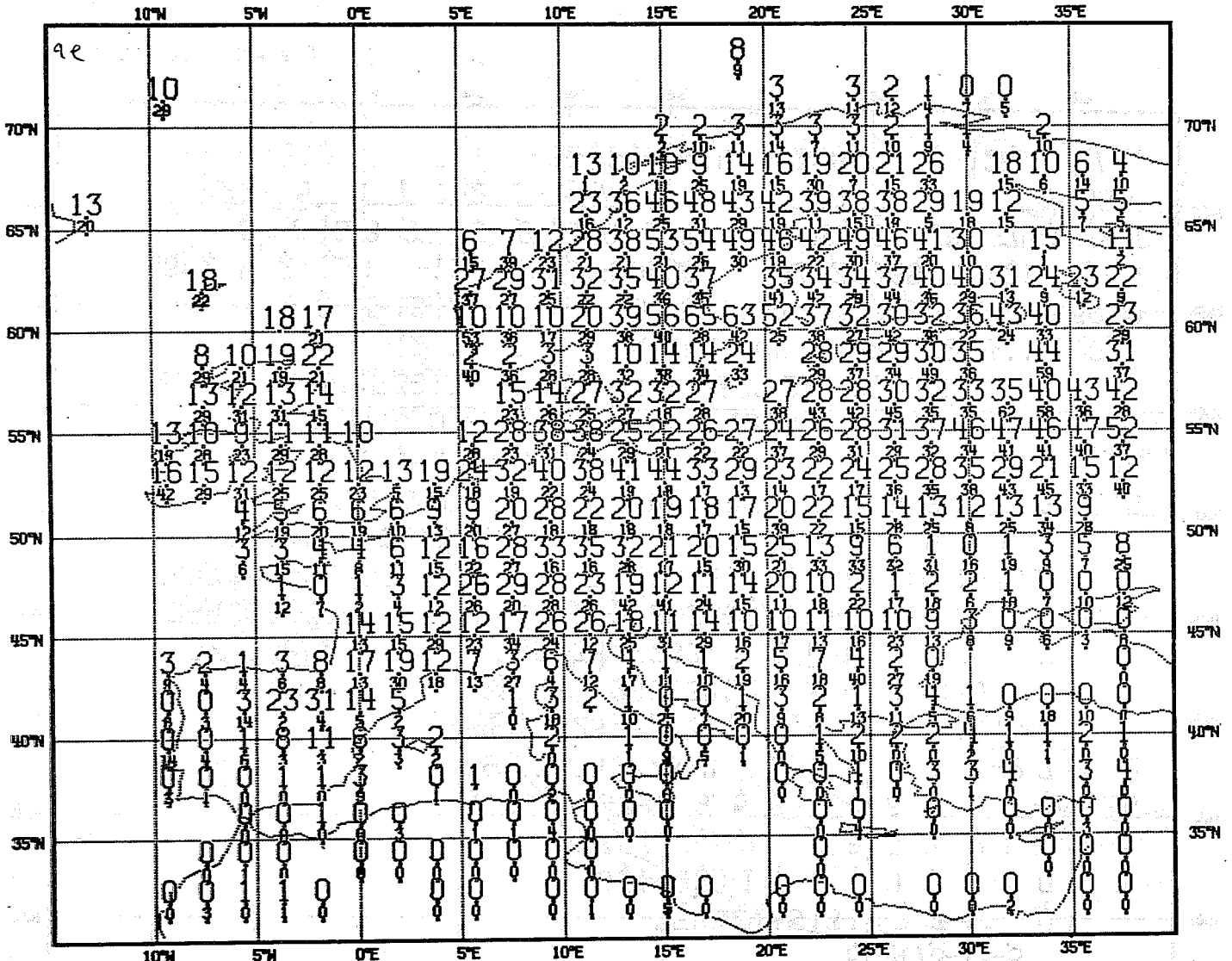

Fig. 9e Monthly mean of daily values at grid points of forecast and observation. Unit 1/10 mm per 24 hours. Large number above grid point is forecasts, smaller number below grid point is observation.
D1, August 1980, 31 forecasts.

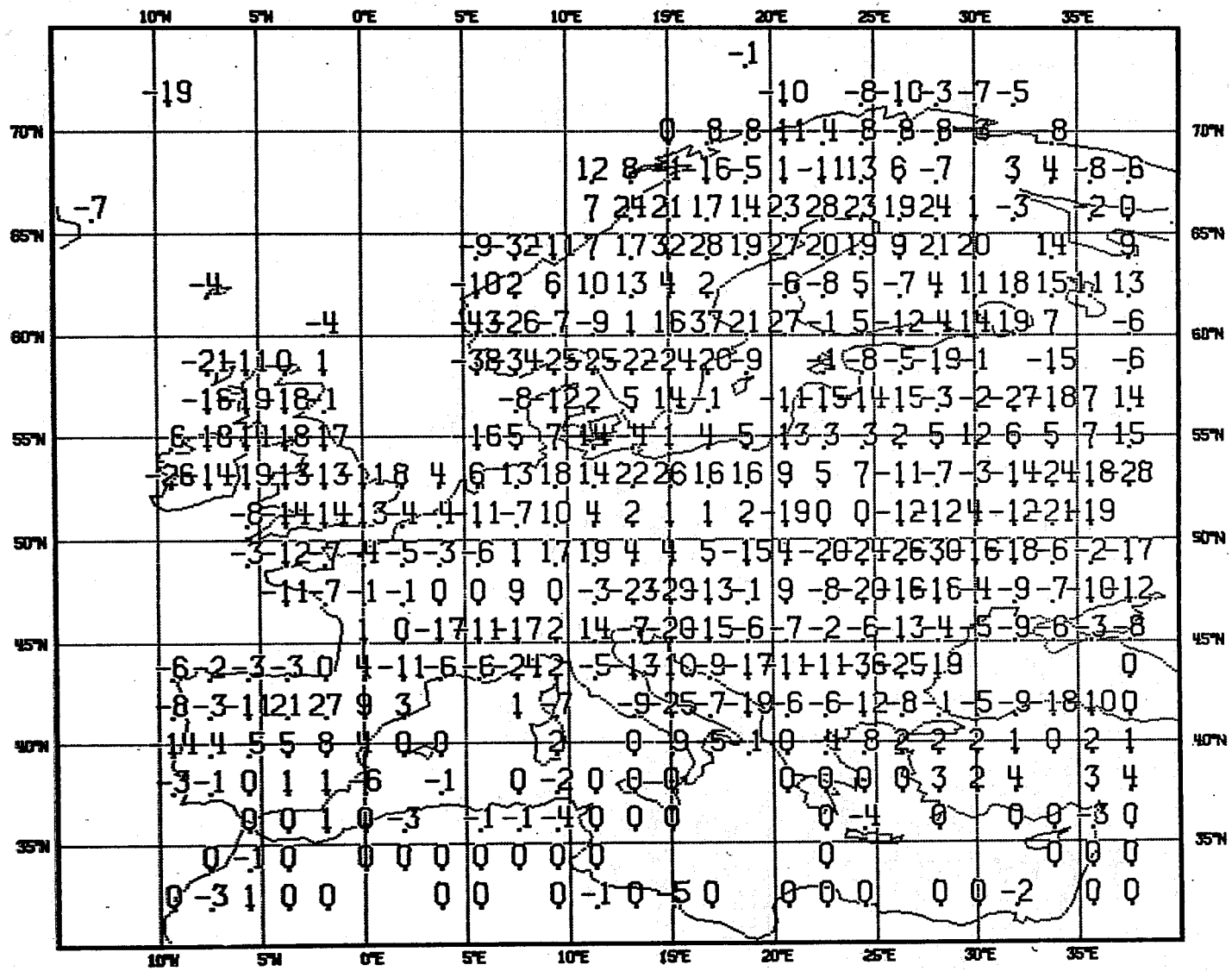

Fig. 9f Bias for Fig. 9e.

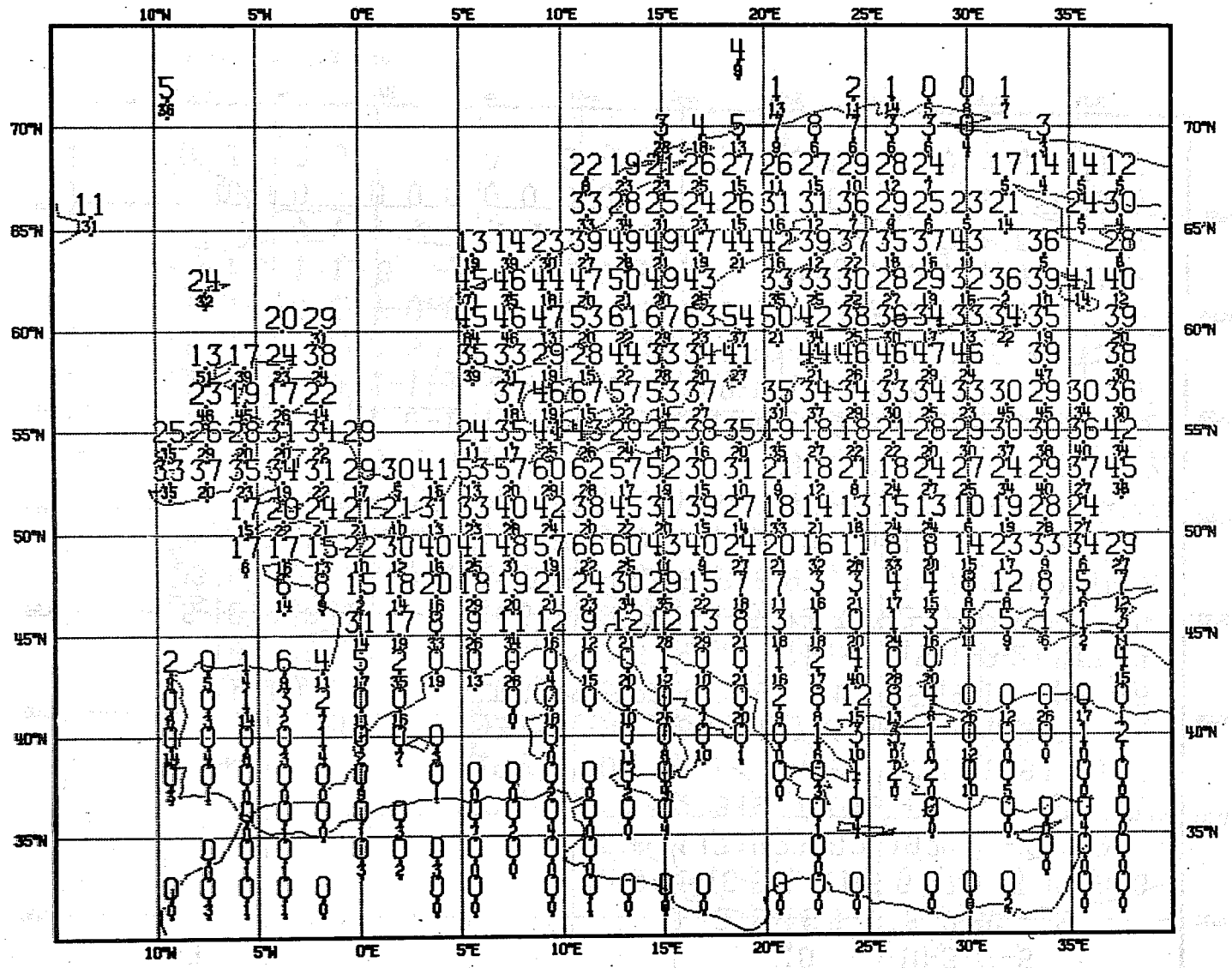

Fig. 9g Monthly mean of daily values at grid points of forecast and observation. Unit 1/10 mm per 24 hours. Large number above grid point is forecasts, smaller number below grid point is observation.
D9, August 1980, 31 forecasts.

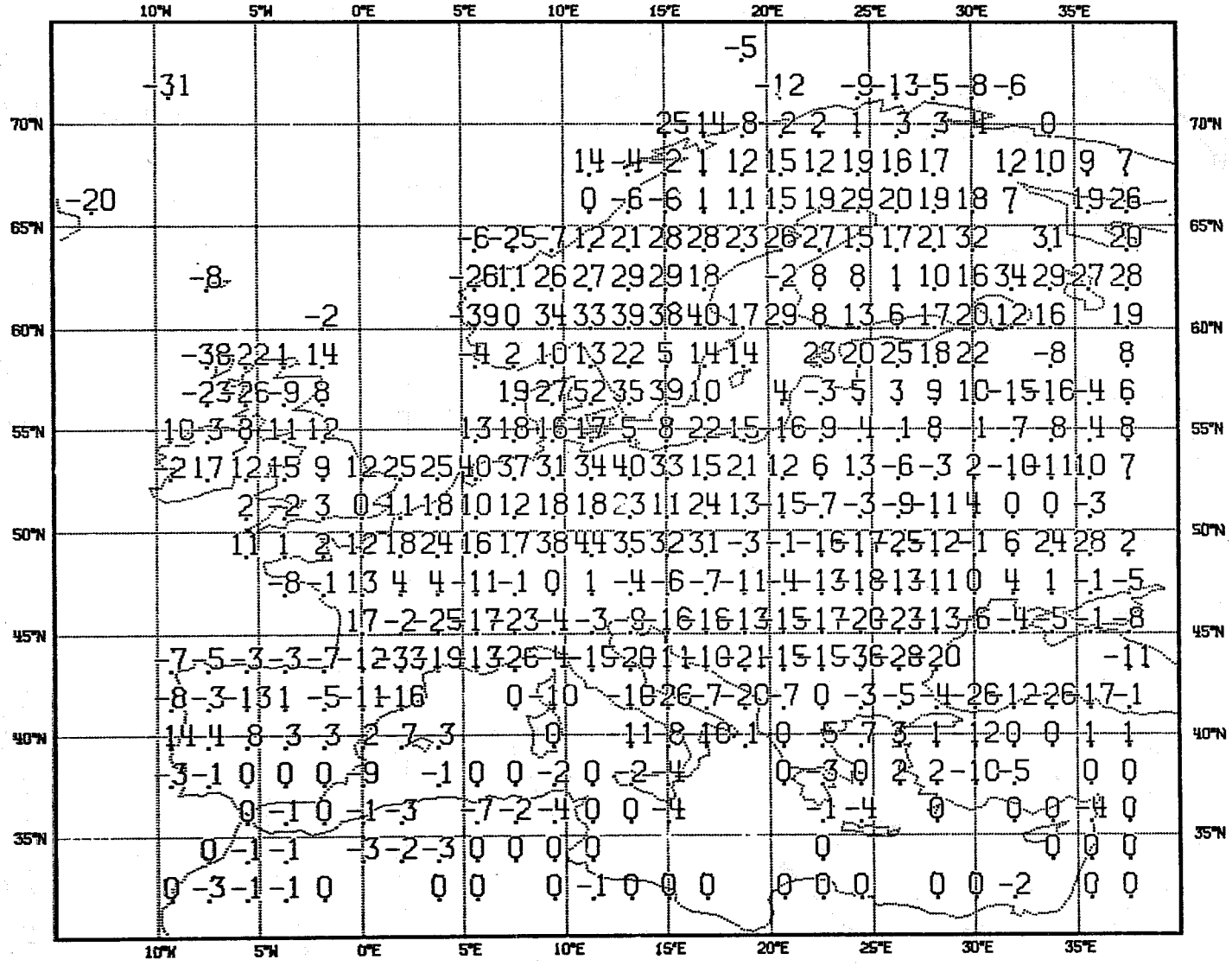

Fig. 9h Bias for Fig. 9g.

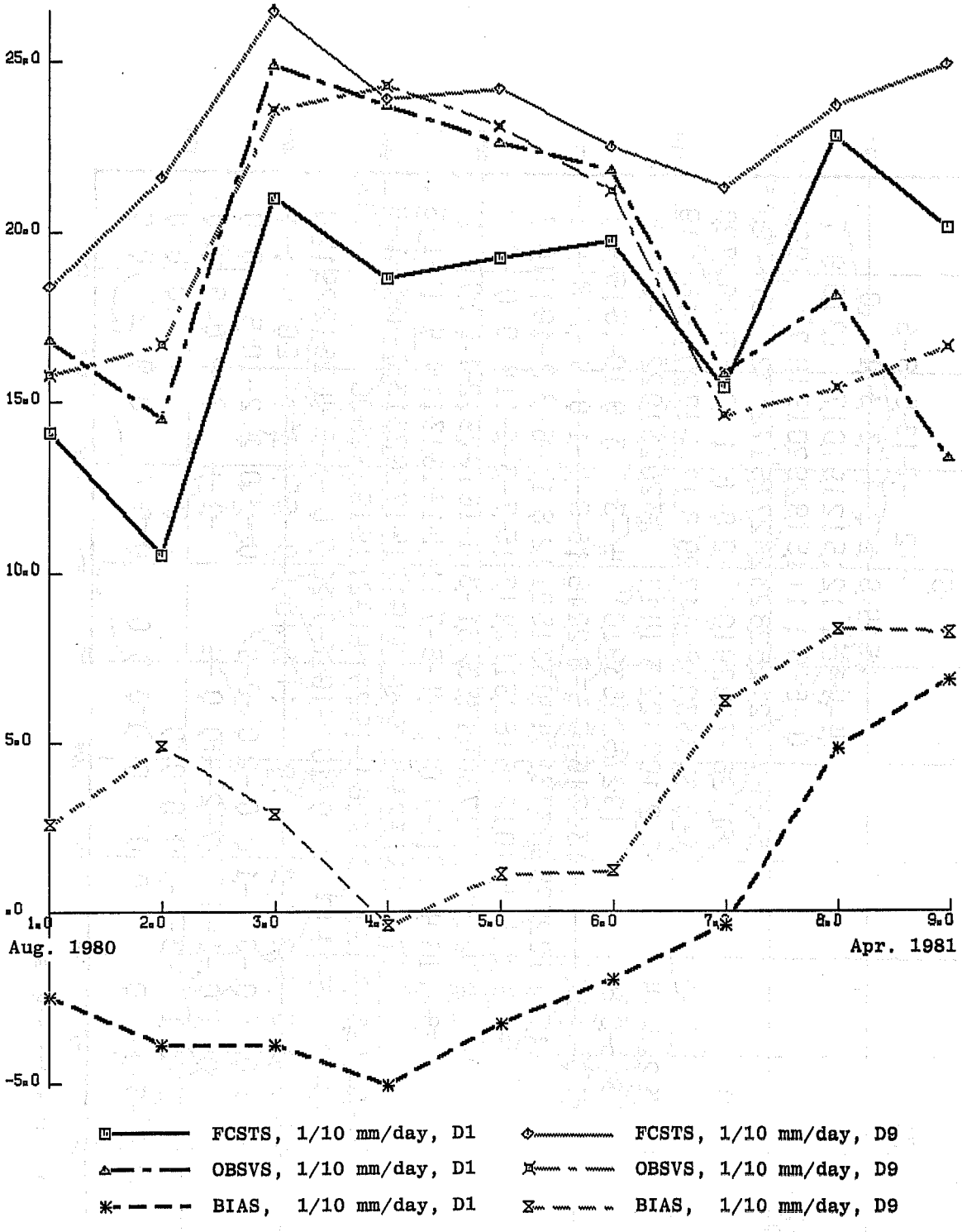

Fig. 10 Changes in forecast and bias during forecast cycle from D1 to D9. Shown are monthly means of forecast, observation, and bias of 24-hour precipitation averaged over all 423 grid points during 9 months from August 1980 to April 1981. Abscissa: Number of month starting with August 1980.

the average the sampling periods for D1 and D9 are displaced 8 days in time.

The increase in forecast values during the forecast cycle was already discussed in Section 3 which summed up for all 750 grid points over land and sea. We notice here that it is equally true of the 423 grid points over land. Comparing the curves in Fig. 10 for 423 land points with the corresponding forecast curves for D1 and D9 over the total 750 land and sea points in Fig. 4e, we notice they are almost identical for the corresponding months, indicating that land-induced and sea-induced physical conditions do not seem to have different influence on the systematic increase in precipitation during the forecast cycle.

The increase is quite modest. The average increase in forecast precipitation from D1 to D9 over the 9 months, August 1980 to April 1981, was 33 percent of the D1 forecast. This modest increase is spread over 8 days and does not make forecasts unrealistic, particularly as they start with a modest negative bias and end up with a modest positive bias, averaging to zero somewhere in mid-cycle.

7.2 Class Frequency Distributions

The frequency distributions of the 8 classes of amounts into which forecast and observation have been divided, can be monitored and are valuable in diagnosing the model physics of the precipitation process. One should expect seasonal changes, as indeed is found. In addition, one finds marked changes in the distributions occurring during the forecast cycle.

Figs. 11a-h show the class frequency distribution for D1, D2, D3, D6, and D9 as well as observed distribution for every other month of the 16 months of verification, starting with February 1980 and ending with April 1981. The chief characteristics are:

(1) The forecast frequency of class 1 (zero precipitation) diminishes steadily during the forecast cycle. It is not a one-day start-up effect but lasts all through the 9-day cycle, even if the biggest change is from D1 to D2 and subsequent changes are progressively smaller.

(2) Zero precipitation frequencies were greater for observations than for forecasts, with the exception of forecasts for D1 during a few months (February 1980, August 1980, October 1980). The feature can be described as a systematic shrinking of precipitation-free areas during the forecast cycle, or conversely, a systematic spreading of precipitation areas during the cycle.

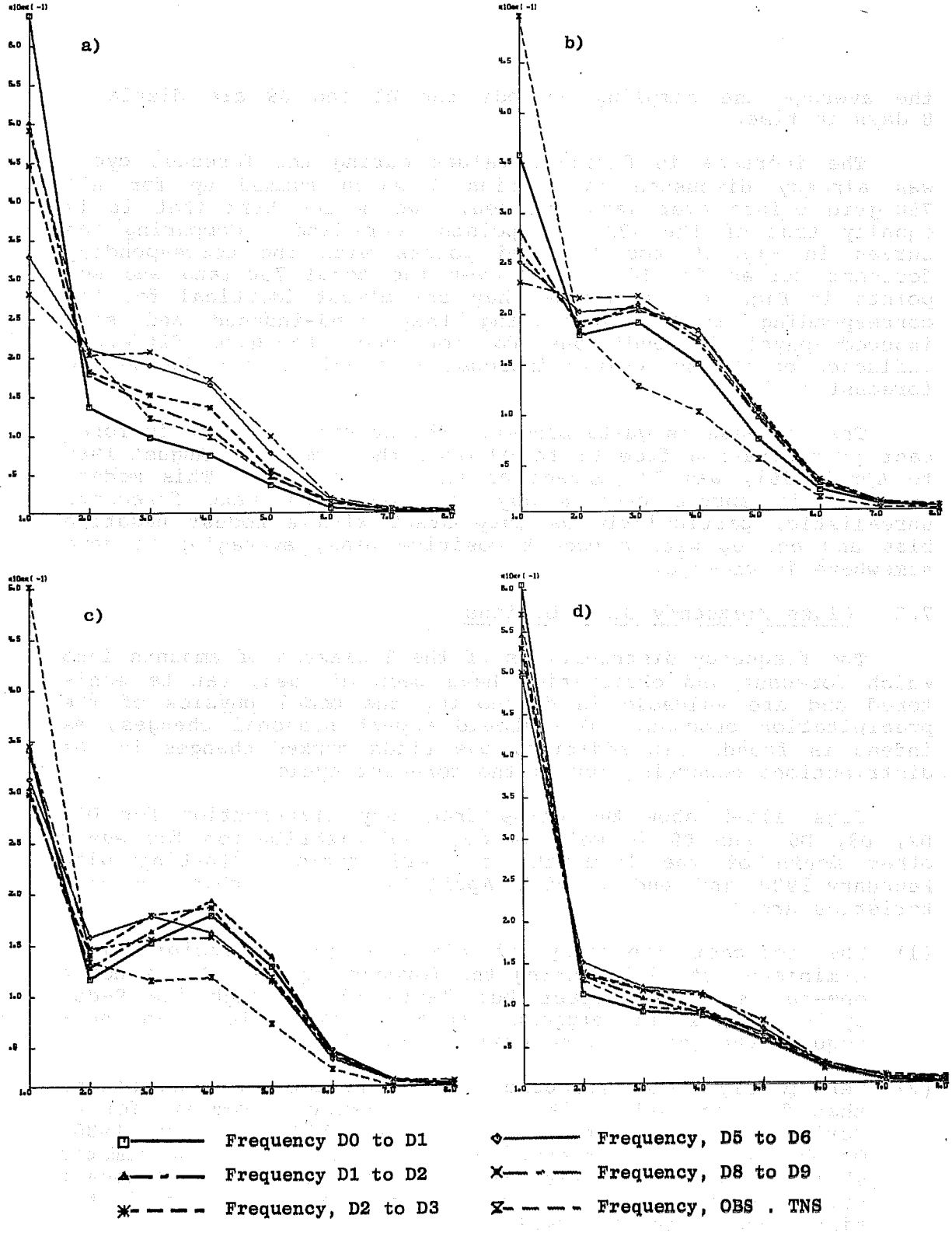

Fig. 11 (a)-(d) Frequency distribution of precipitation amounts to 8 classes (Table 3) for observations and forecasts for D1, D2, D3, D6 and D9. Shown are every other month from February 1980 to April 1981. Abscissa: Class 1 to 8. (a) Feb.80 (b) Apr.80 (c) Jun.80 (d) Aug.80.

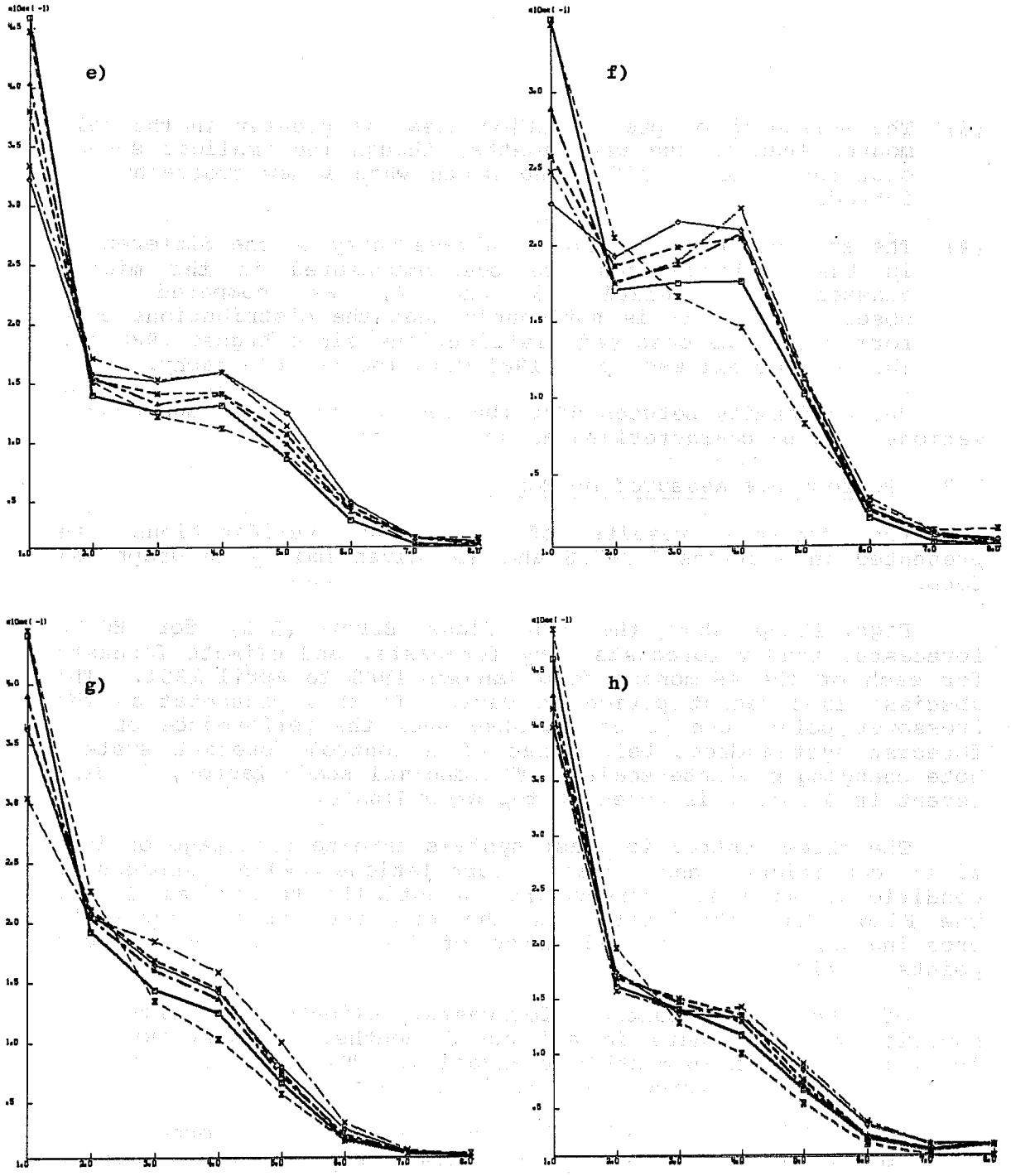

Fig. 11(e)-(h) Frequency distribution of precipitation amounts in 8 classes (Table 3) for observations and forecasts for D1, D2, D3, D6 and D9. Shown are every other month from February 1980 to April 1981. Abscissa: Class 1 to 8. (e) Oct.80 (f) Dec.80 (g) Feb.81 (h) Apr.81.

(3) The spreading of precipitation areas is greater in the cold months than in the warm months, though the smallest spread occurred in April 1981, the month when a new orography was introduced.

(4) The area under each curve is always unity so the differences in the lower classes are overcompensated in the middle classes, particularly 3 and 4, as compared with observations. It is noticeable that the distributions conform better to observed distributions since August 1980 and the best of all was April 1981 with the new orography.

The comformity between distributions of forecasts and observations must be characterized as good overall.

7.3 Performance Measured by RCE

The concrete results of the score verifications are presented in Sections 7.3-7.6 and are given mainly in graphical form.

Figs. 12a-p show the RMS Class Error (RCE) for ECMWF forecasts, random forecasts, dry forecasts, and climate foreasts for each of the 16 months from January 1980 to April 1981. The abscissa is forecast period in days. In this presentation the crossover points can be established when the performance of the forecast system drops below that of a control forecast system. Note changing ordinate scales. The decimal scale factor, if different from unity, is noted at top of ordinate.

The three control forecast systems take no advantage of initial conditions and their correlations with subsequent conditions, so their curves are essentially horizontal lines. The ECMWF RCE deteriorates as the forecast period increases, crossing one or two or all three of the rival curves at some points in time.

Of the three control forecasts, climate forecasts are superior to the others in all but 2 months, February 1980 and April 1981 out of 16 month's evaluation. The next best are dry forecasts. Random forecasts lose in all months.

ECMWF forecasts have smaller RCE than climate forecasts up to a point in time which varies from month to month from as little as one day (30 hours) in June 1980 to 5.4 days in December 1980. The average time span before ECMWF forecasts lose to climate forecasts is 3.4 days.

Fig. 14a shows the monthly variation in the RCE for D1 and D3 for ECMWF, climate, and random forecasts. The peaks and valleys on the curves occur nearly always in the same month. One would expect the random forecasts and the climate forecasts for D1 and D3 to have almost identical curves. This is confirmed by Fig. 14a. The largest RCE for ECMWF D1 forecasts occurred in

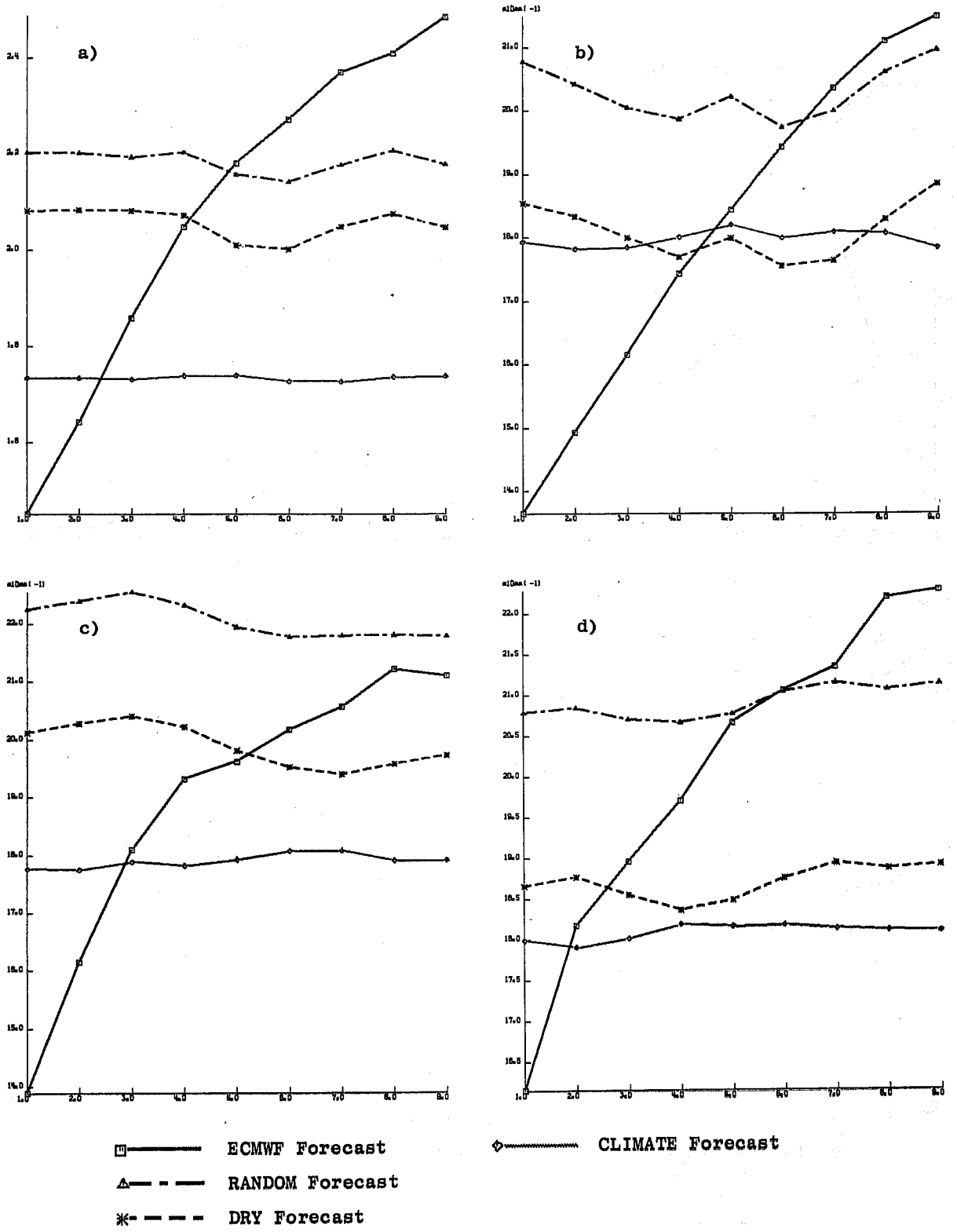

Fig. 12(a)-(d) RMS Class Error of ECMWF forecasts, random forecasts, dry forecasts, and climate forecasts, D1 to D9 averaged over all forecasts of month. Abscissa: D1 to D9. (a) Jan.80 (b) Feb.80 (c) Mar.80 (d) Apr.80.

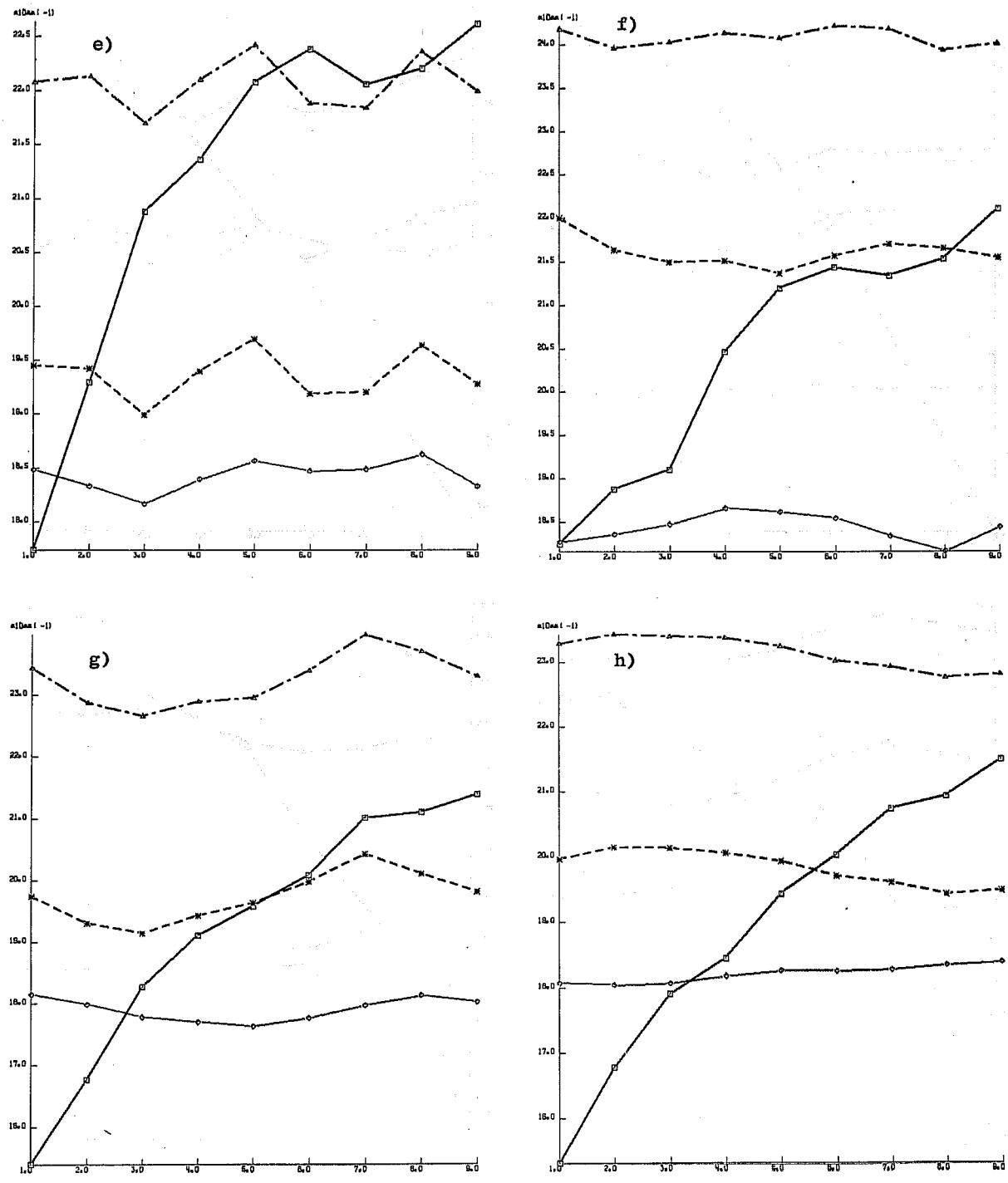

Fig. 12(e)-(h) RMS Class Error og ECMWF forecasts, random forecasts, dry forecasts, and climate forecasts, D1 to D9 averaged over all forecasts of month. Abscissa: D1 to D9. (e) May 80 (f) Jun.80 (g) Jul.80 (h) Aug.80.

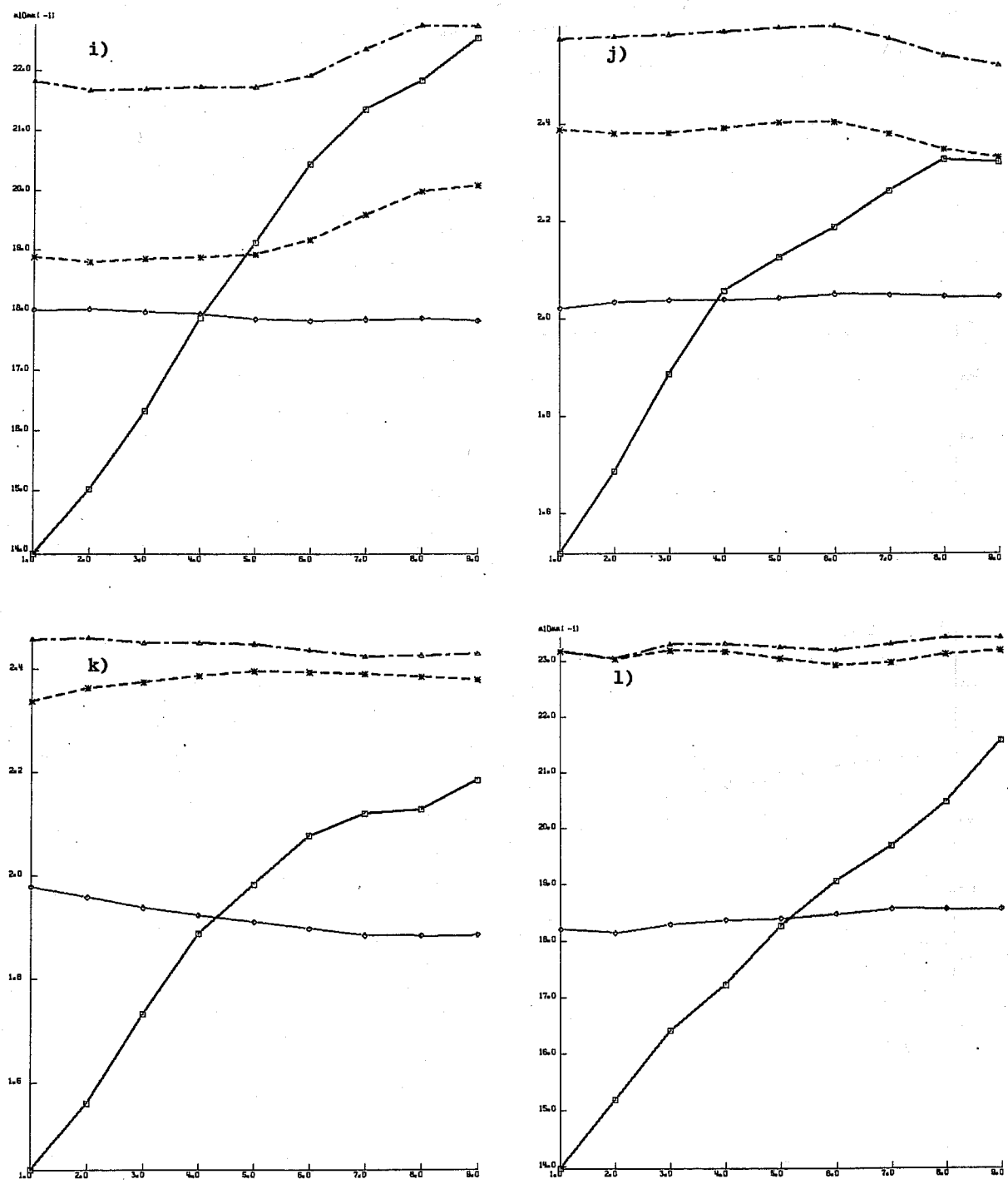

Fig. 12(i)-(l) RMS Class Error of ECMWF forecasts, random forecasts, dry forecasts, and climate forecasts, D1 to D9 averageed over all forecasts of month. Abscissa: D1 to D9. (i) Sep.80 (j) Oct.80 (k) Nov.80 (l) Dec.80.

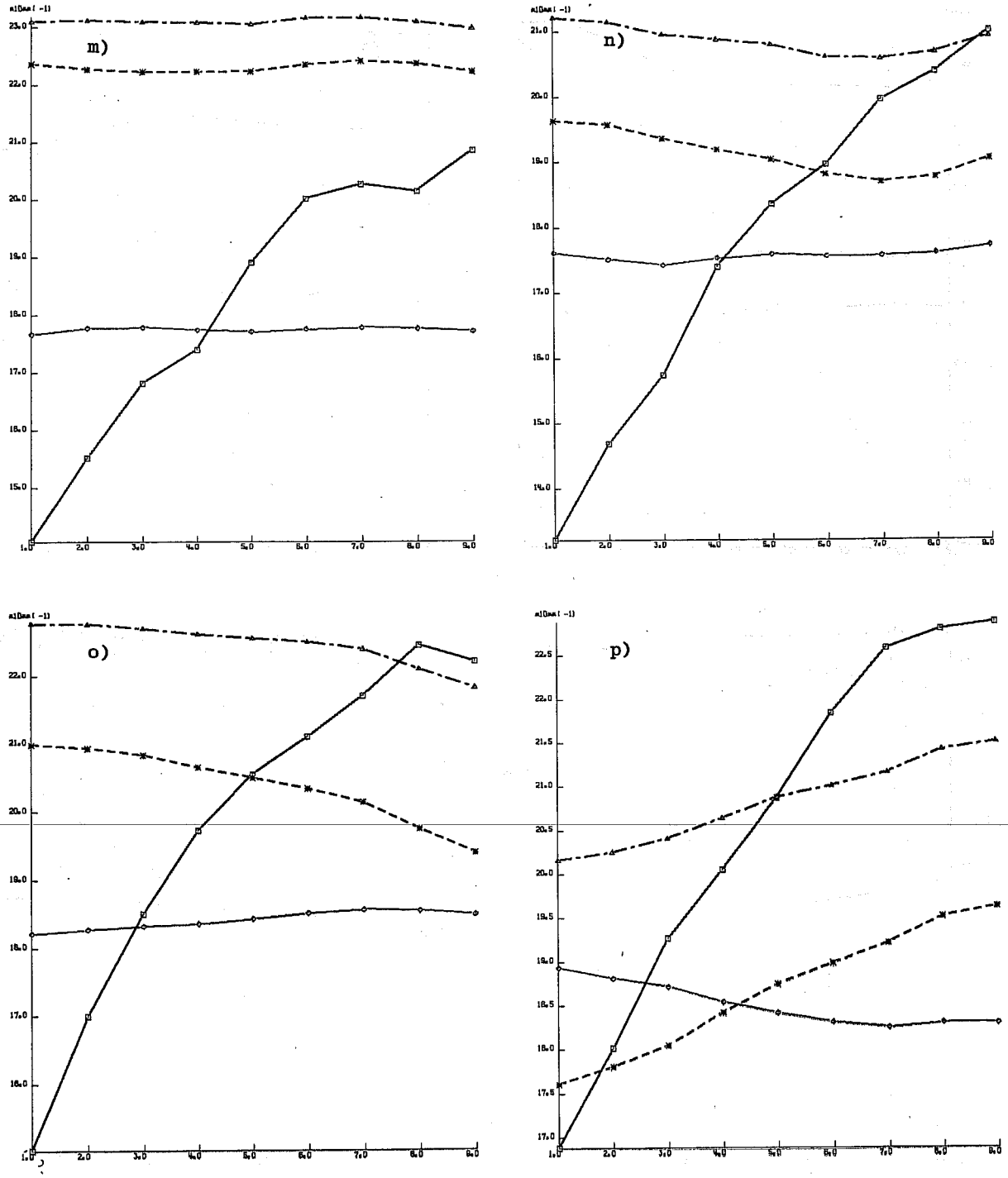

Fig. 12(m)-(p) RMS Class Error of ECMWF forecasts, random forecasts dry forecasts, and climate forecasts, D1 to D9 averaged over all forecasts of month. Abscissa: D1 to D9. (m) Jan.81 (n) Feb.81 (o) Mar.81 (p) Apr.81.

June 1980, in May 1980 for D3 forecasts, and in October 1980 for both climate and random forecasts. As can be seen from Fig. 10, October 1980 was the wettest month on the diagram, and we would expect climate forecasts and random forecasts to do poorly in abnormally wet months. ECMWF did relatively well in October 1980 as can be seen from Figs. 12j and 14a.

Fig. 14b shows the performance for 5-day interval forecasts, D0 to D5, and D4-D9. Again, random and climate forecasts have curves for the first and the last interval which almost coincide.

Climate forecasts are hard to beat for 5-day intervals when measured by the RCE, as Fig. 14b shows. For the first 5-day interval, D0 to D5, ECMWF and climate forecasts run neck and neck, ECMWF winning in 6 months, climate in 9 months, and one month undecided. For the last interval, D5 to D9, climate forecasts beat ECMWF forecasts by an uncomfortable margin in all months, as measured by the RCE.

To sum up, ECMWF forecasts are better than all control forecasts to about 3 days, but the best control forecast, the climate forecast, takes the lead from there, being slightly better for the first 5-day period and decidedly better for the last 5-day period.

7.4 Performance Measured by HSS

Figs. 13a-p show the HSS of the ECMWF forecasts. The abscissa is time from D1 to D9. The HSS is measured relative to random, dry, and climate forecasts. Comparison with Brier's "chance observation" experiment is also shown. Note changing scale along ordinate. Decimal scale factor, if different from unity, is noted at top of ordinate.

While the RCE is better the smaller it is, the HSS is better the larger it is since it is a measure of the number of correct forecasts relative to the competing forecast. A negative HSS means losing to the control forecast.

In all months dry forecasts for D1 to D9 have the greatest number of correct forecasts with exception of 5 months (January 1980, February 1980, October 1980, January 1981, February 1981) where ECMWF beat dry forecasts for D1 only, 2 months (November 1980 and December 1980) where ECMWF forecasts beat dry forecasts for D1 to D2 and for D1 to D3, respectively.

The HSS based on dry forecasts is a tough competition because all cases of observed zero precipitation become correct forecasts. As shown in Section 7.2 the verification area had observed dry conditions in time-area space ranging from 35 percent in December 1980 to 60 percent in July 1980. December 1980 was also ECMWF's best HSS against dry forecasts and July 1980 one of the worst.

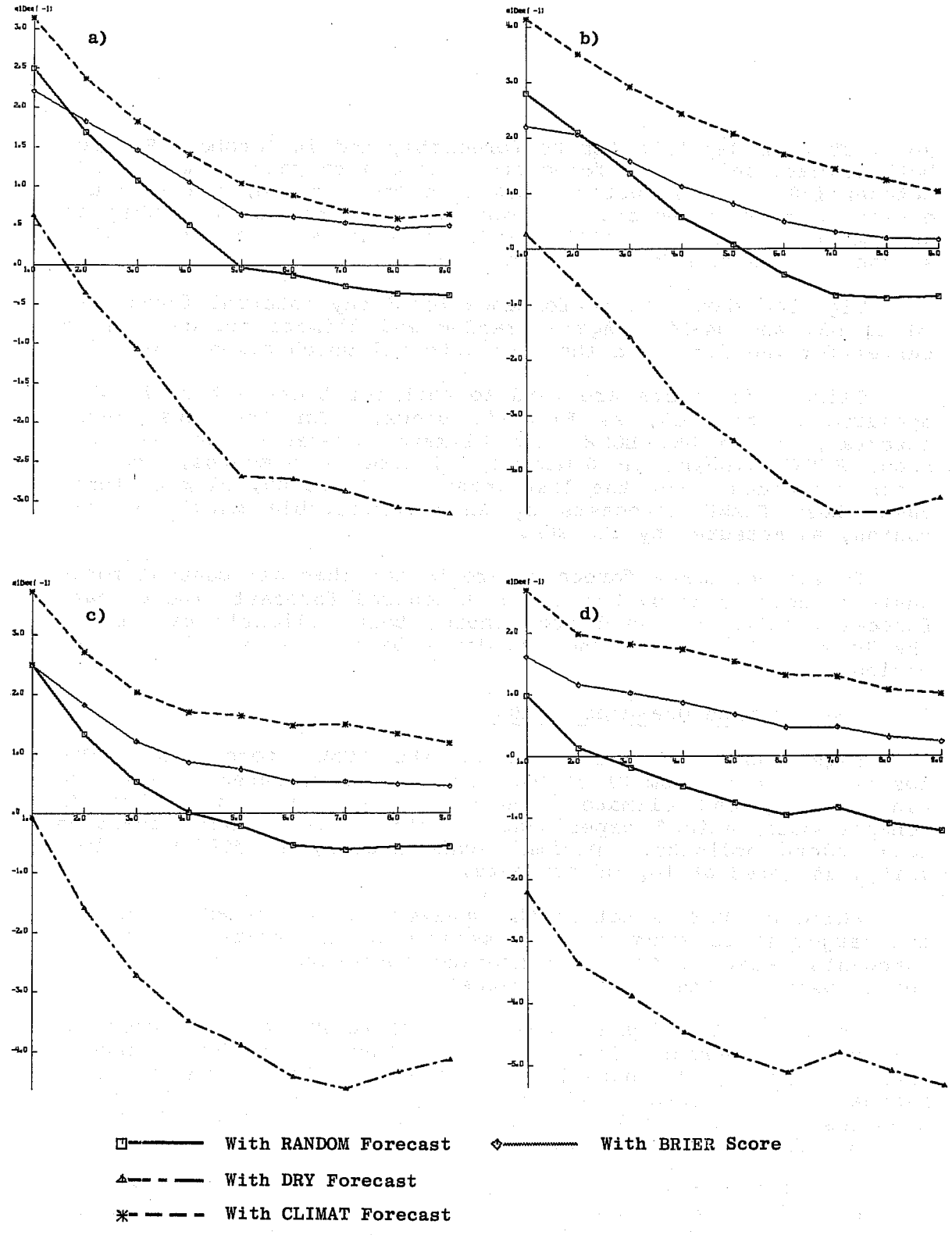

Fig. 13(a)-(d) Heidke Skill Score of ECMWF forecasts in relation to random forecasts, dry forecasts, climate forecasts, and compared to Brier's chance forecasts. Abscissa: D1 to D9. (a) Jan.80 (b) Feb.80 (c) Mar.80 (d) Apr.80.

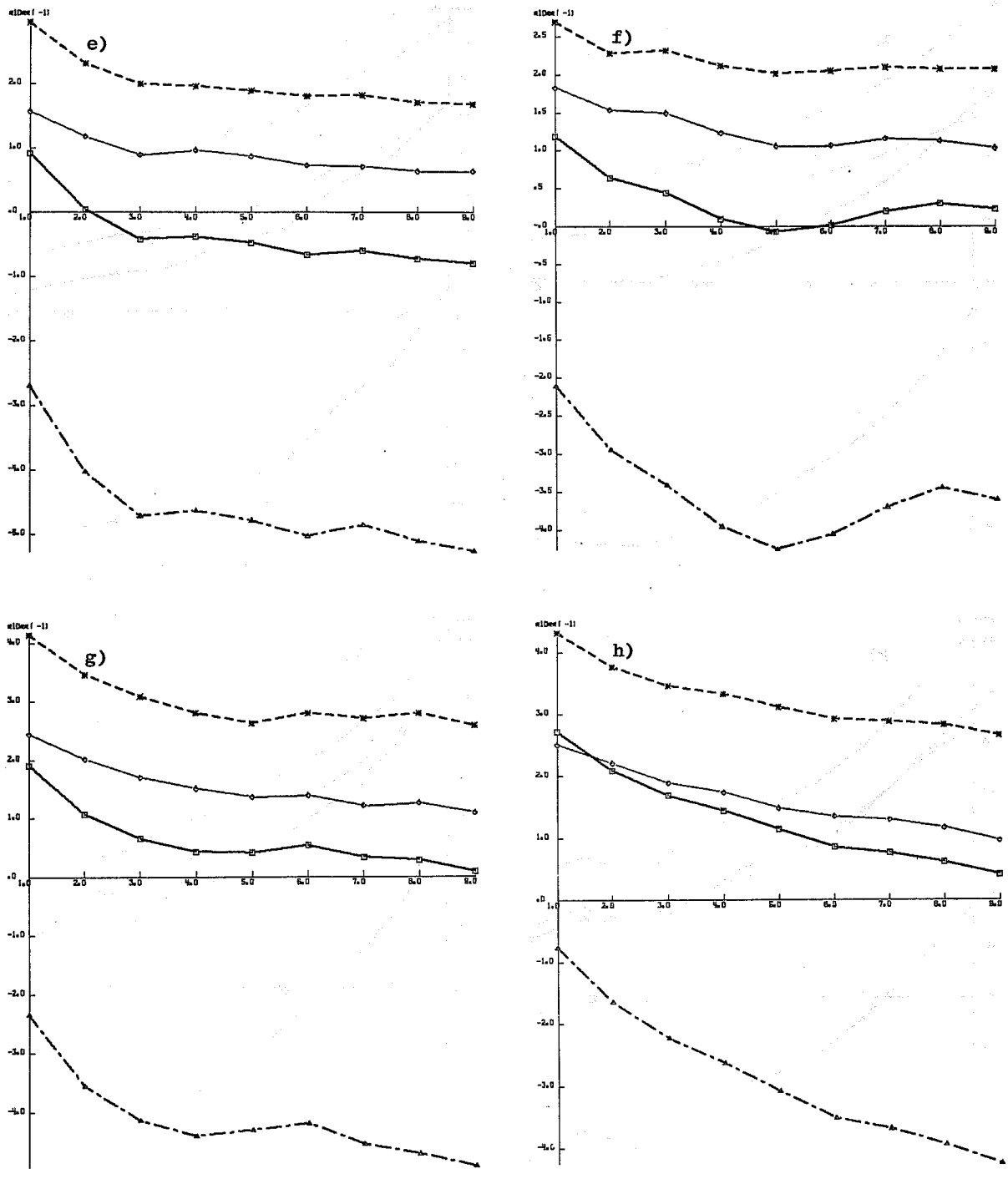

Fig. 13(e)-(h) Heidke Skill Score of ECMWF Forecasts in relation to random forecasts, dry forecasts, climate forecasts, and compared to Brier's chance forecasts. Abscissa: D1 to D9. (e) May 80 (f) Jun.80 (g) Jul.80 (h) Aug.80.

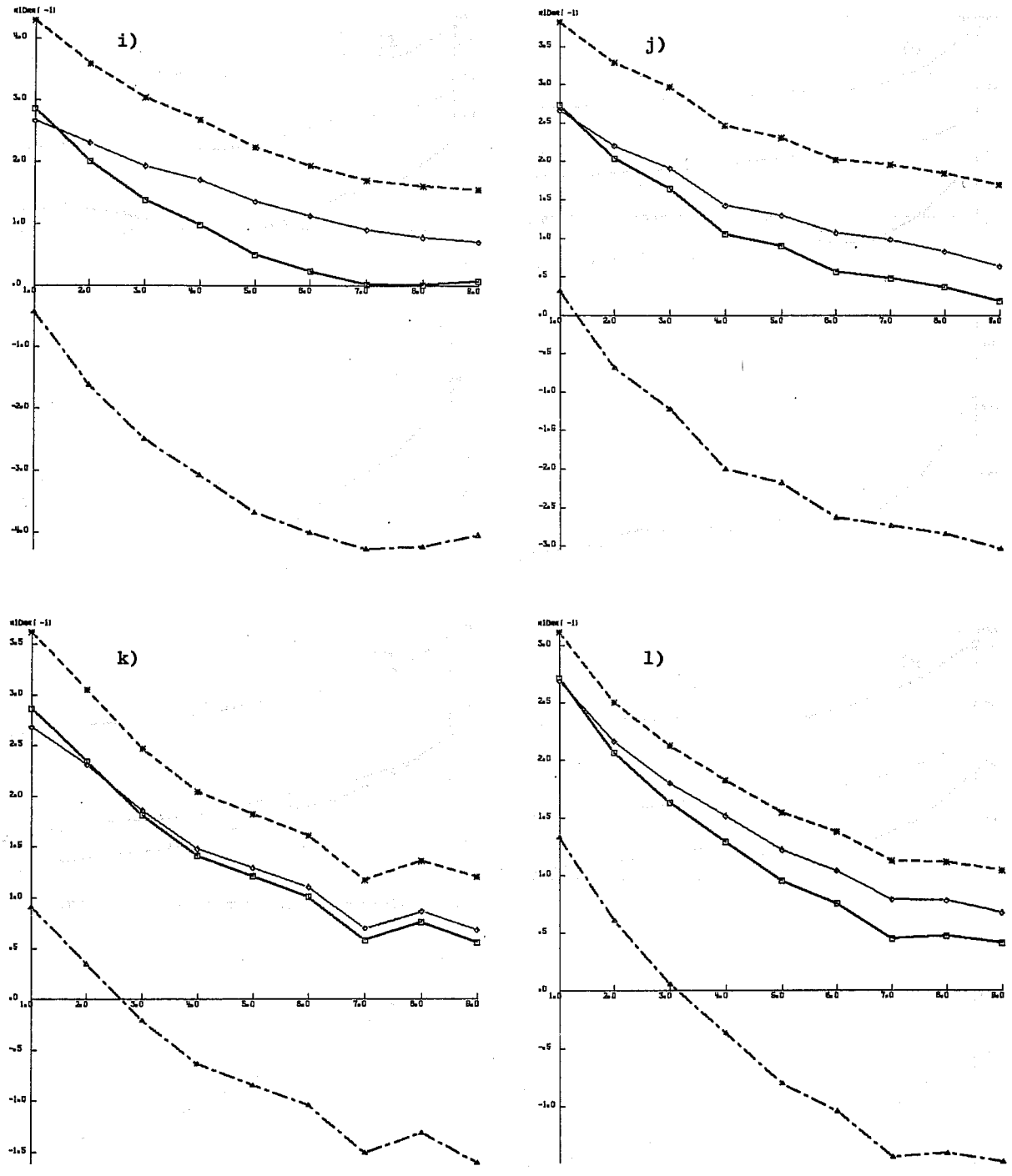

Fig. 13(i)-(l) Heidke Skill Score of ECMWF forecasts in relation to random forecasts, dry forecasts, climate forecasts, and compared to Brier's chance forecasts. Abscissa: D1 to D9. (i) Sep.80 (j) Oct.80 (k) Nov.80 (l) Dec.80.

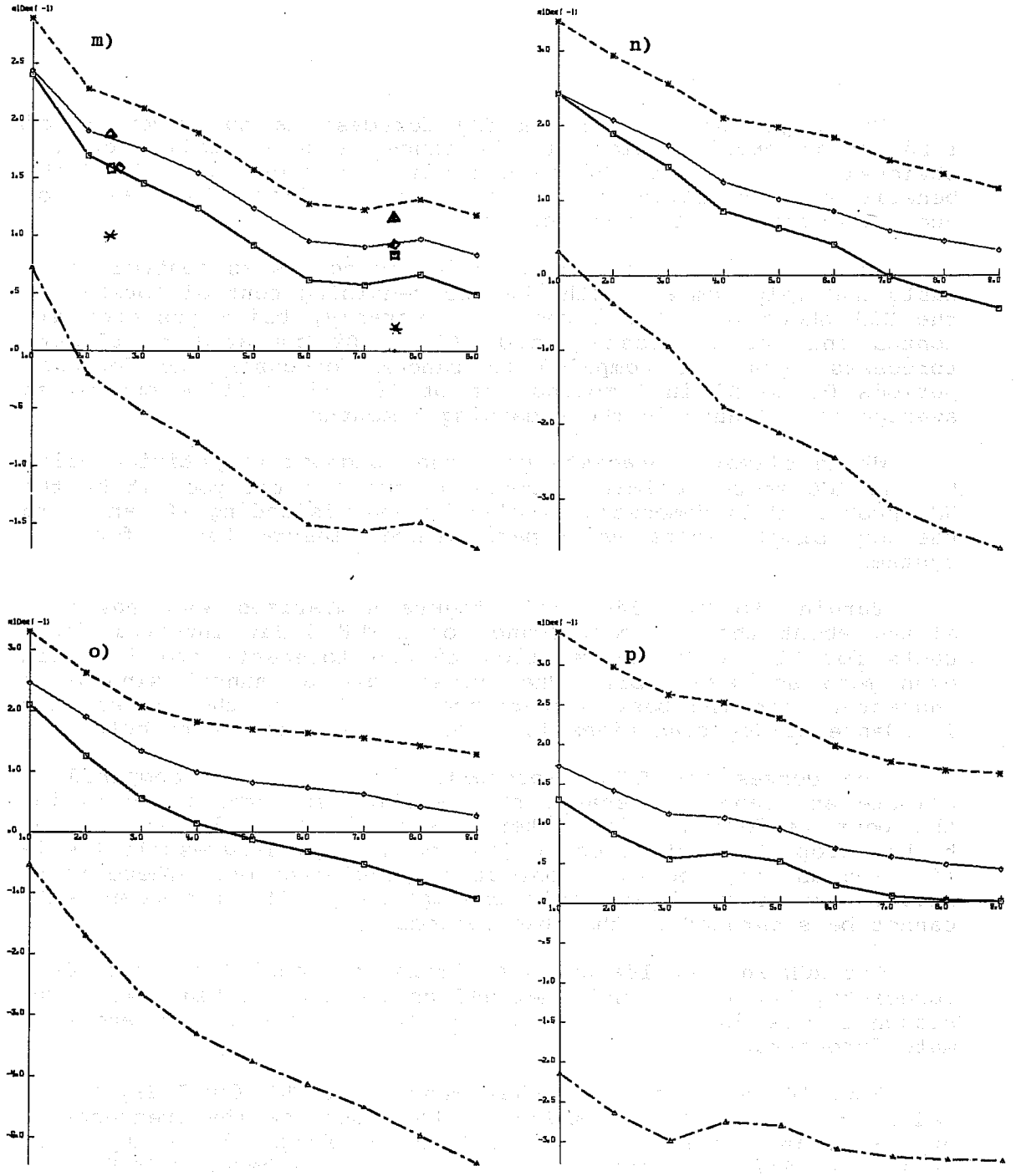

Fig. 13(m)-(p) Heidke Skill Score of ECMWF forecasts in relation to random forecasts, dry forecasts, climate forecasts, and compared to Brier's chance forecasts. Abscissa: D1 to D9. (m) Jan.81 (n) Feb.81 (o) Mar.81 (p) Apr.81.

One might contend that a dry forecast is not a method of forecasting precipitation at all since it never calls for it. Obviously, the method cannot be applied to operations where the benefit of forecasting would come from calling the cases of occurring precipitation correctly.

If we on this ground dismiss dry forecasts as control forecasts and only compare with the two remaining control forecasts, the HSS shows up ECMWF forecasts favorably, being positive all months and for forecast periods D1 to D9 compared to climate forecasts, positive compared to random forecasts and forecast periods D1 to D9 in 9 months out of 16 and positive out to an average of 4.4 days in the remaining 7 months.

While climate forecasts gave the toughest competition going by the RCE score, climate forecasts come out the poorest by the HSS score. This demonstrates clearly how misleading it can be to use any single score as a performance measure for a forecast system.

Turning to Fig. 14c, this figure summarizes what has been stated about the HSS performance of ECMWF 1-day interval forecasts for D1 and D3 in relation to dry forecasts and is valid even more so beyond D3. The curves have an annual sinusoidal character with the best performance of ECMWF in the winter when incidence of dry conditions is lower. This could be expected.

The curves for ECMWF forecasts for D1 and D3 compared to climate and random forecasts show similar seasonal waves giving the best performance in winter. This is probably due to the higher proportion of forecast (and occurring) large-scale precipitation as compared with convective precipitation. There is no simple way of verifying this assumption as the two components cannot be separated in the observations.

The RCE in Fig. 14a shows a similar seasonal trend for ECMWF forecasts, but not so well defined as for HSS in Fig. 14c. The seasonal wave is not evident in Fig 14a for the random and climate forecasts.

Fig. 14d shows the HSS performance of ECMWF for 5-day interval forecasts, D0 to D5, and D4 to D9. Because the frequency of dry 5-day intervals is much less than the frequency of dry 1-day intervals, dry forecasts are not as successful here. ECMWF forecasts show superiority over all control forecasts for the first interval, D0 to D5.

For the second 5-day interval, D4 to D9, there is skill over dry forecasts in all months but most in winter. Climate forecasts are more competitive over 5-day intervals than over 1-day intervals and run very close to ECMWF in the winter months. In the summer months, June-August, there is resultant skill over climate forecasts for the second 5-day interval.

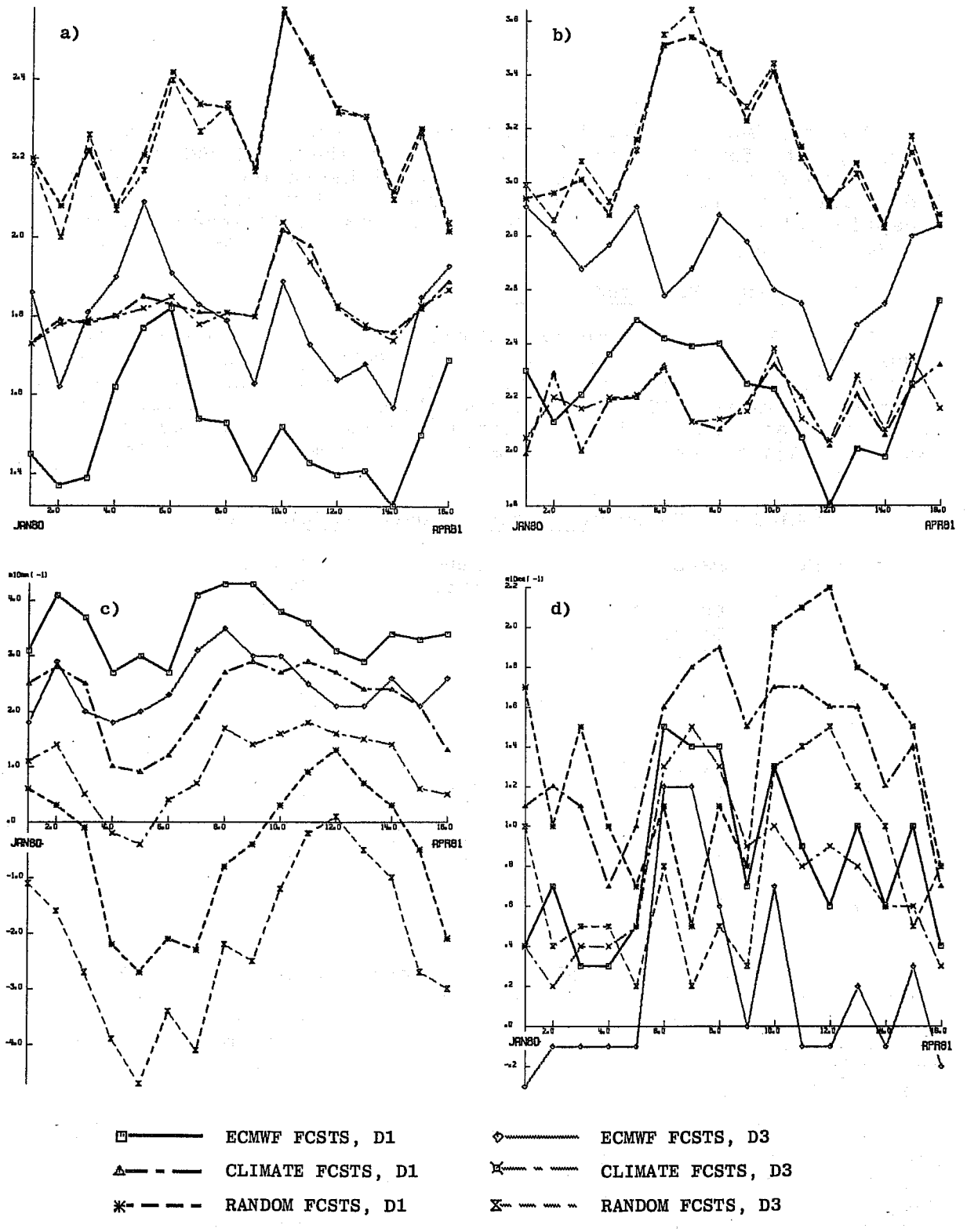

Fig. 14 Comparison of scores of ECMWF forecasts, climate forecasts, and random forecasts over 16 months from January 1980 to April 1981. Abscissa: Number of month starting with January 1980.
(a) RMS Class Error, D1 and D3.
(b) RMS Class Error, D0 to D5, D4 to D9.
(c) Heidke Skill Score, D1 and D3.
(d) Heidke Skill Score, D0 to D5, D4-D9.

One may sum up Fig. 14d as follows: Averaging over 5-day periods, both for D0-D5 and D5-D9, improves the ECMWF HSS as compared with dry forecasts. ECMWF has a higher number of correct 5-day interval forecasts than all control forecasts with the exception of climate forecasts which tie the ECMWF forecasts for the second 5-day period D4-D9 during the winter half of the year.

7.5 Performance Measured by TS

Figs. 15-23 show Threat Scores for the months from August 1980 to April 1981. The TS for ECMWF forecasts and for climate forecasts and for class 1, 4, and 6 as thresholds are given by 6 curves. (See Table 3 for class limits.) The abscissa is a forecast interval code ranging from 1 to 15 given in Table 4. Again, note varying scales and decimal scale factor at top of ordinate.

TABLE 4: Forecast Intervals Used in Figs. 17-25

Interval No.	Days (18Z-18Z)	Length of Interval	Length of Forecast
1	D0 to D1	24 hours	30 hours
2	D1 to D2	24	54
3	D2 to D3	24	78
4	D3 to D4	24	102
5	D4 to D5	24	126
6	D5 to D6	24	150
7	D6 to D7	24	174
8	D7 to D8	24	198
9	D8 to D9	24	222
10	D0 to D3	72	78
11	D2 to D6	96	150
12	D0 to D5	120	126
13	D4 to D9	120	222
14	D1 to D4	72	102
15	D0 to D9	216	222

The first nine intervals are 24-hour intervals. The ECMWF scores drop gradually out to D9 and diminish as the threshold value of precipitation increases, as one might expect. For instance, the TS for forecasting more than 5 mm of precipitation is only about half of the TS for forecasting measurable precipitation.

As the interval lengthens, the TS goes up since the likelihood of precipitation exceeding a certain value increases with the length of the interval. The intervals 10 to 15 are of varying lengths, from 3 days to 9 days. As can be seen from the figures, for the lowest class limit of 1 (zero precipitation) the TS of ECMWF and climate are very close for the multi-day intervals 10-15 with the exception of one month, April 1981. In April 1981, the ECMWF Threat Scores for forecasting measurable precipitation were definitely inferior to climate forecasts for multi-day intervals, the only month where this was the case.

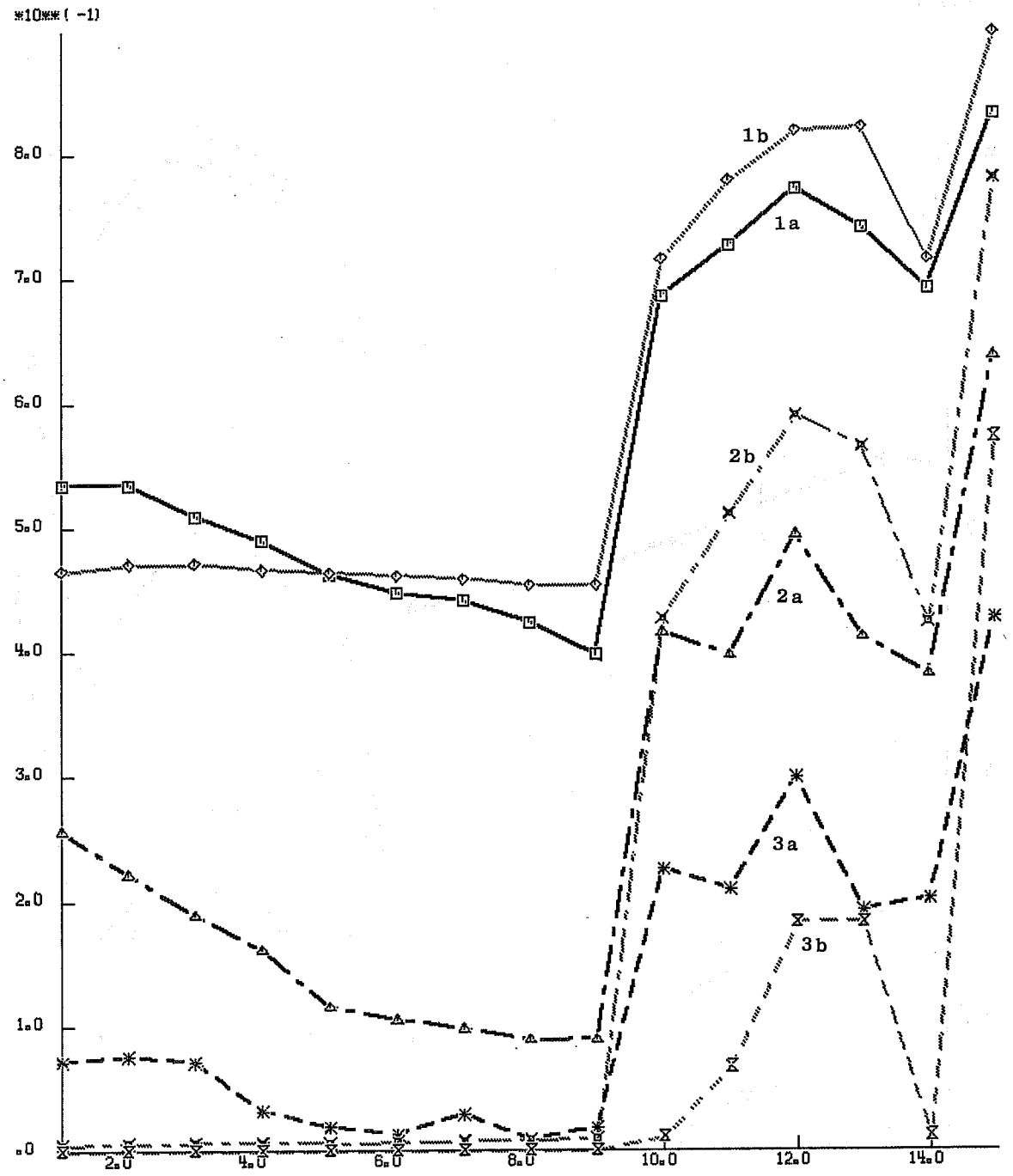

Fig. 15 Threat Score for exceeding limits of precipitation, for ECMWF forecasts and climate forecasts. Comparison of ECMWF and climate forecasts of precipitation exceeding (1) non-measurable amounts, (2) 5 mm, (3) 15 mm. Abscissa: Code for forecast interval (See Table 4). August 1980.

Dark lines (a) ECMWF forecasts
Grey lines (b) Climate forecasts

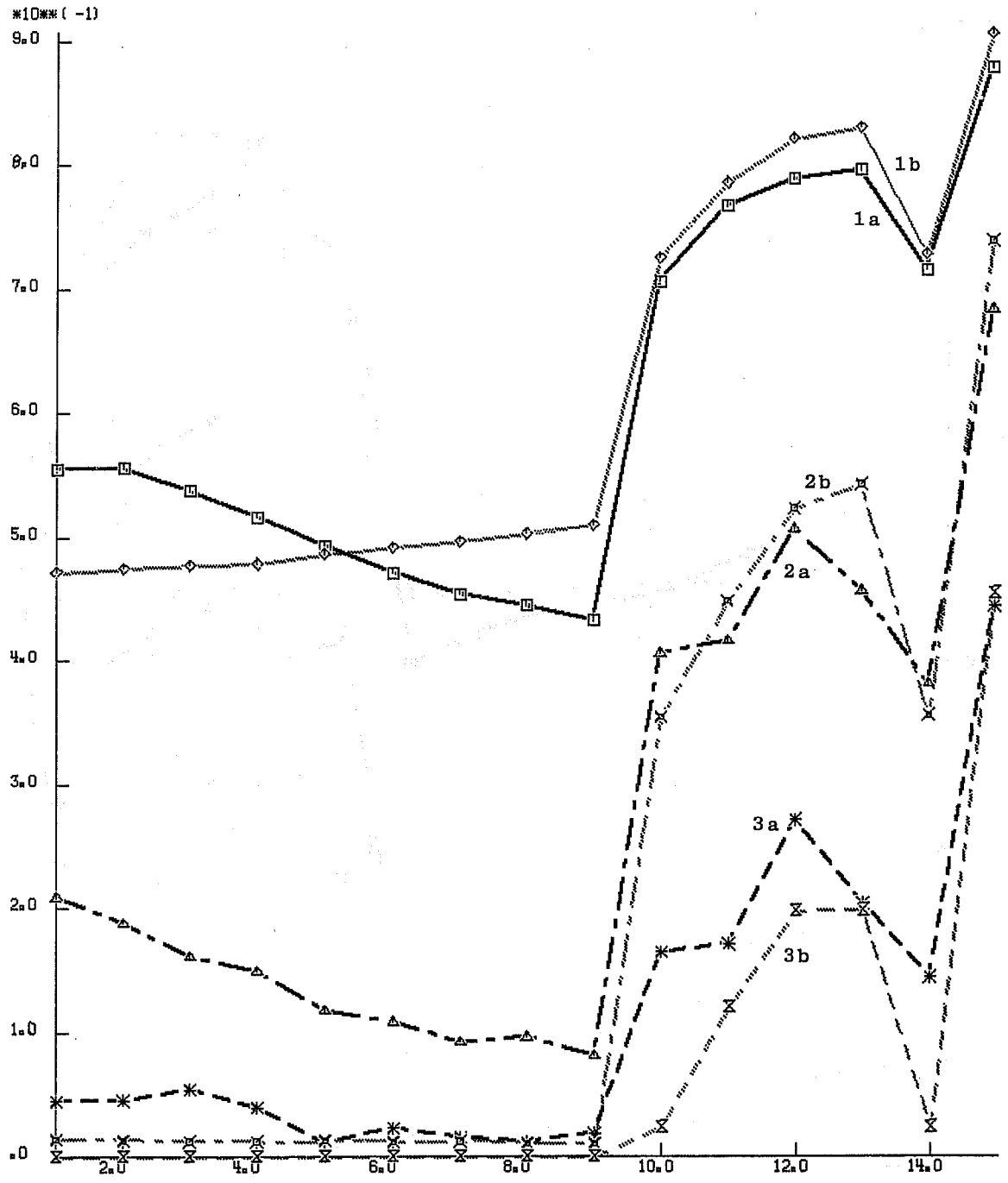

Fig. 16 Threat Score for exceeding limits of precipitation, for ECMWF forecasts and climate forecasts. Comparison of ECMWF and climate forecasts of precipitation exceeding (1) non-measurable amounts, (2) 5 mm, (3) 15 mm. Abscissa: Code for forecast interval (See Table 4). September 1980.

Dark lines (a) ECMWF forecasts
Grey lines (b) Climate forecasts

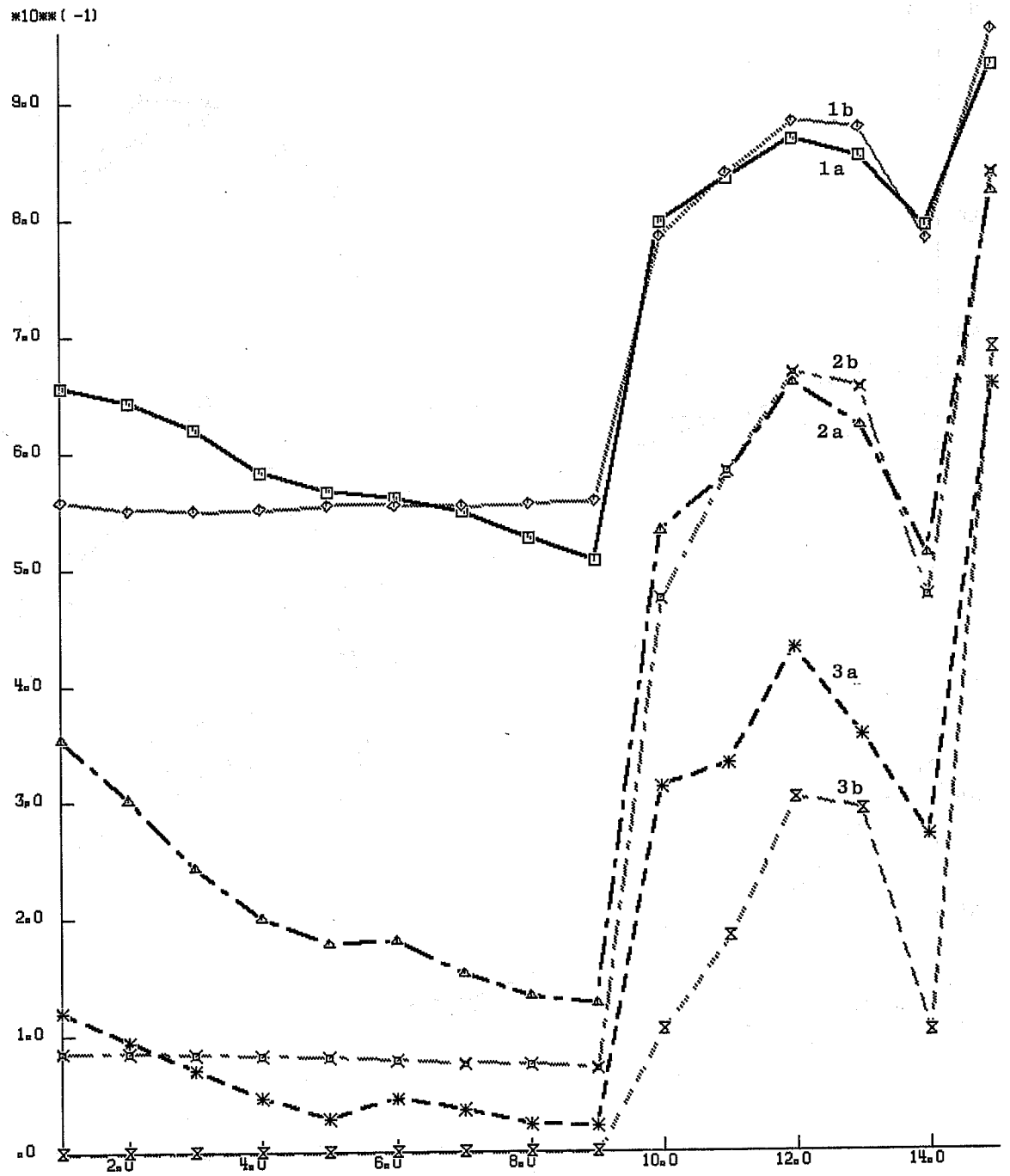

Fig. 17 Threat Score for exceeding limits of precipitation, for ECMWF forecasts and climate forecasts. Comparison of ECMWF and climate forecasts of precipitation exceeding (1) non-measurable amounts, (2) 5 mm, (3) 15 mm. Abscissa: Code for forecast interval (See Table 4). October 1980.

Dark lines (a) ECMWF forecasts
Grey lines (b) Climate forecasts

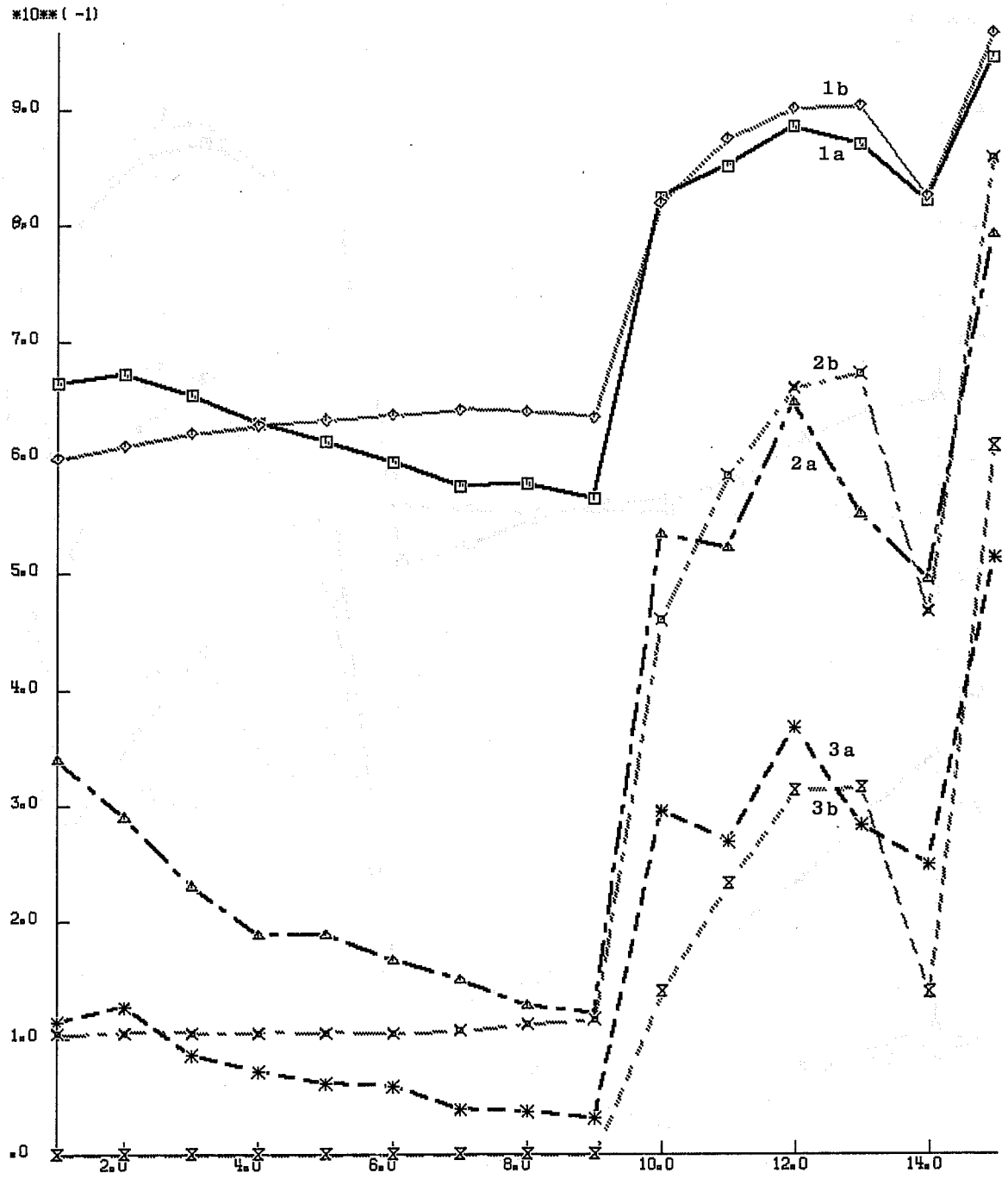

Fig. 18 Threat Score for exceeding limits of precipitation, for ECMWF forecasts and climate forecasts. Comparison of ECMWF and climate forecasts of precipitation exceeding (1) non-measurable amounts, (2) 5 mm, (3) 15 mm. Abscissa: Code for forecast interval (See Table 4). November 1980.

Dark lines (a) ECMWF forecasts
Grey lines (b) Climate forecasts

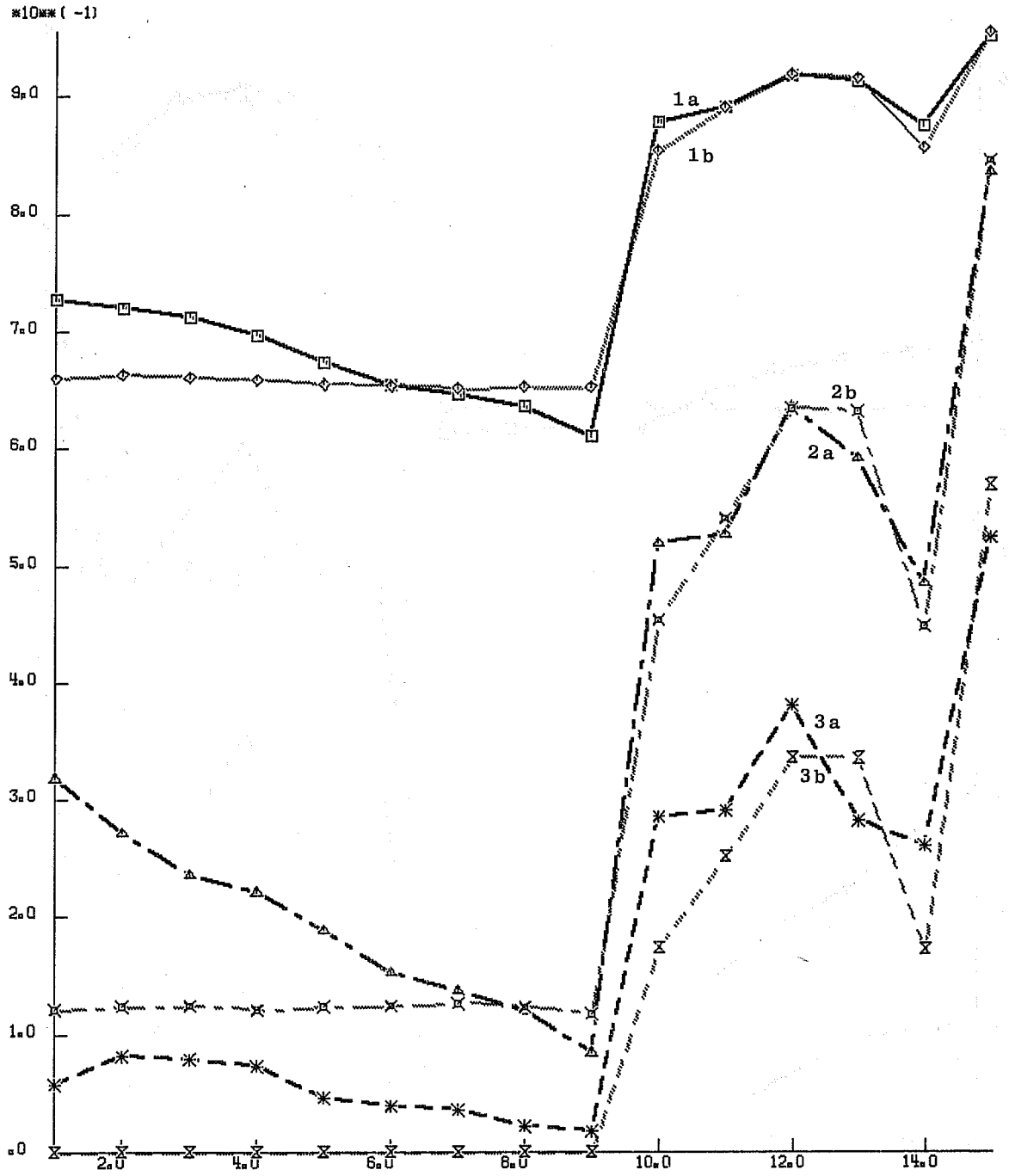

Fig. 19 Threat Score for exceeding limits of precipitation, for ECMWF forecasts and climate forecasts. Comparison of ECMWF and climate forecasts of precipitation exceeding (1) non-measurable amounts, (2) 5 mm, (3) 15 mm. Abscissa: Code for forecast interval (See Table 4). December 1980.

Dark lines (a) ECMWF forecasts
Grey lines (b) Climate forecasts

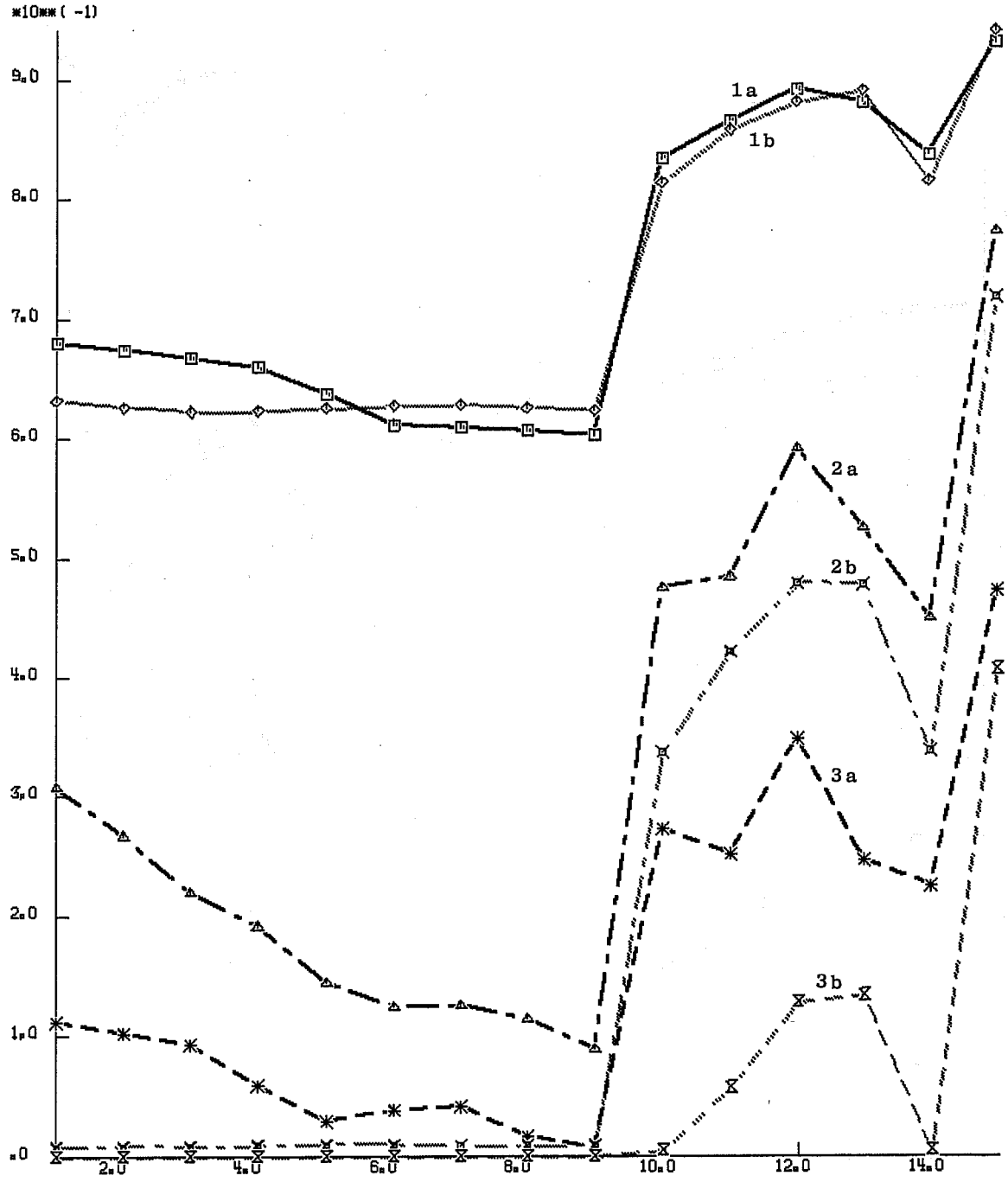

Fig. 20 Threat Score for exceeding limits of precipitation, for ECMWF forecasts and climate forecasts. Comparison of ECMWF and climate forecasts of precipitation exceeding (1) non-measurable mounts, (2) 5 mm, (3) 15 mm. Abscissa: Code for forecast interval (See Table 4). January 1981.

Dark lines (a) ECMWF forecasts
Grey lines (b) Climate forecasts

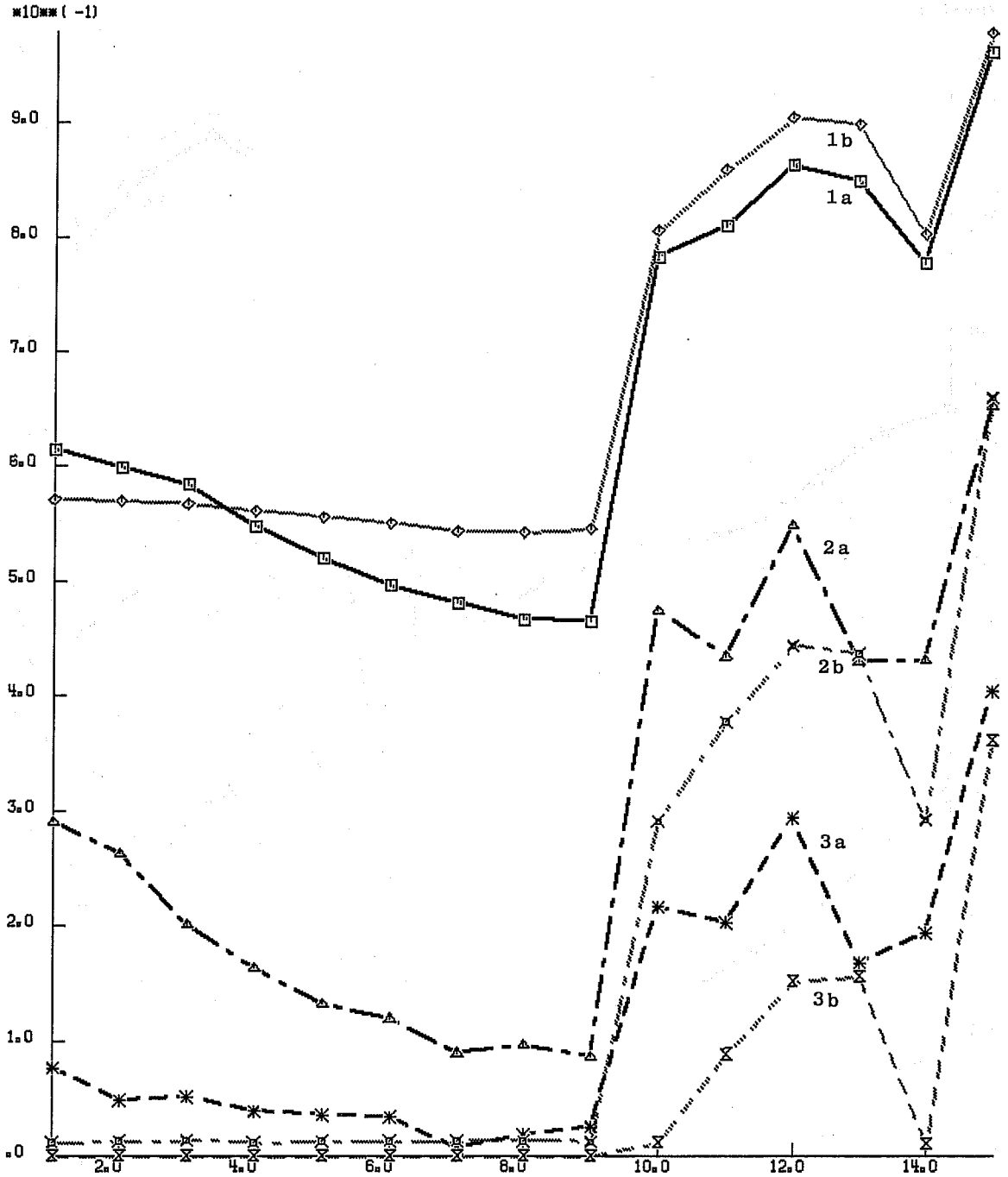

Fig. 21 Threat Score for exceeding limits of precipitation, for ECMWF forecasts and climate forecasts. Comparison of ECMWF and climate forecasts of precipitation exceeding (1) non-measurable amounts, (2) 5 mm, (3) 15 mm. Abscissa: Code for forecast interval (See Table 4). February 1981.

Dark lines (a) ECMWF forecasts
Grey lines (b) Climate forecasts

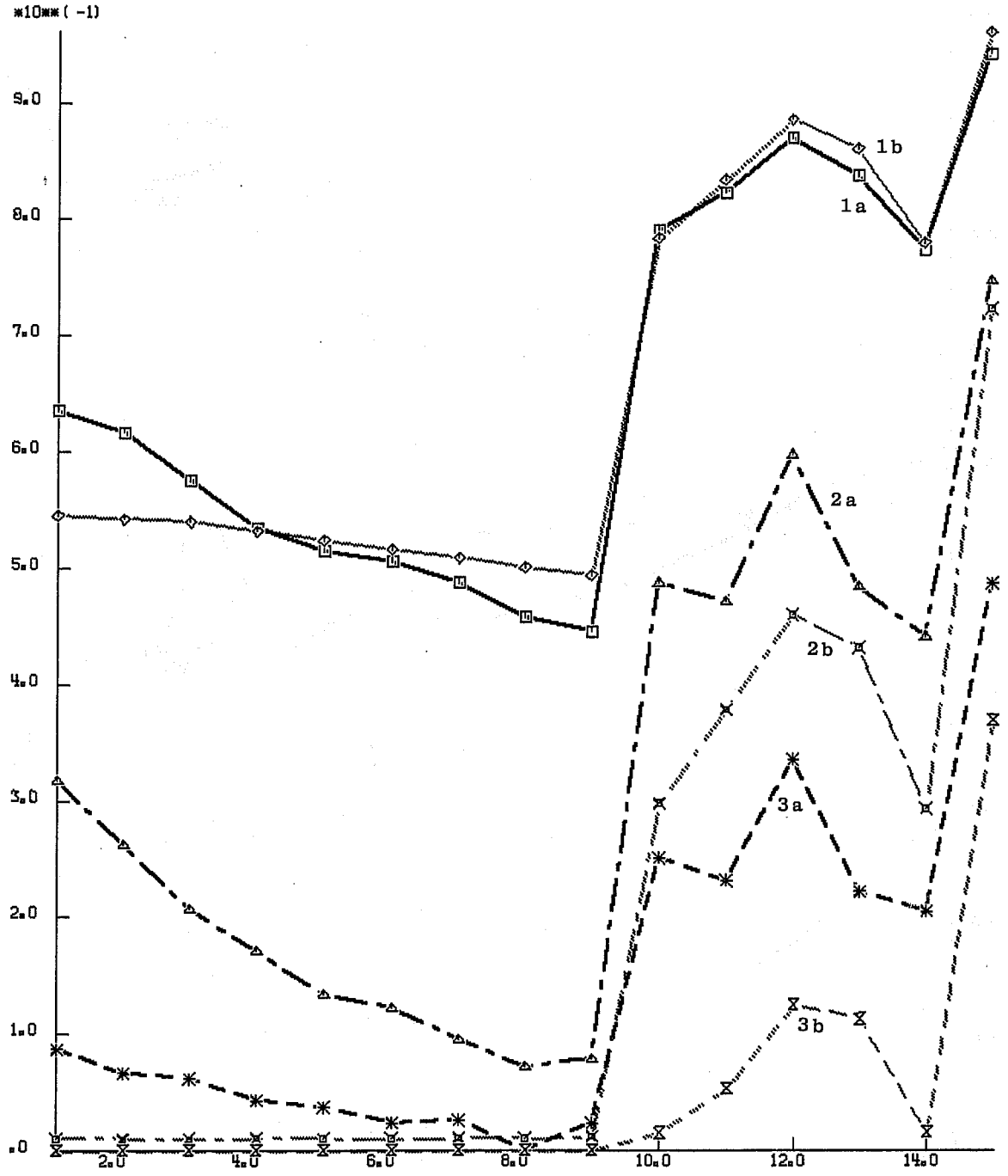

Fig. 22 Threat Score for exceeding limite of precipitation, for ECMWF forecasts and climate forecasts. Comparison of ECMWF and climate forecasts of precipitation exceeding (1) non-measurable amounts (2) 5 mm, (3) 15 mm. Abscissa: Code for forecast interval (See Table 4). March 1981.

Dark lines (a) ECMWF forecasts
Grey lines (b) Climate forecasts

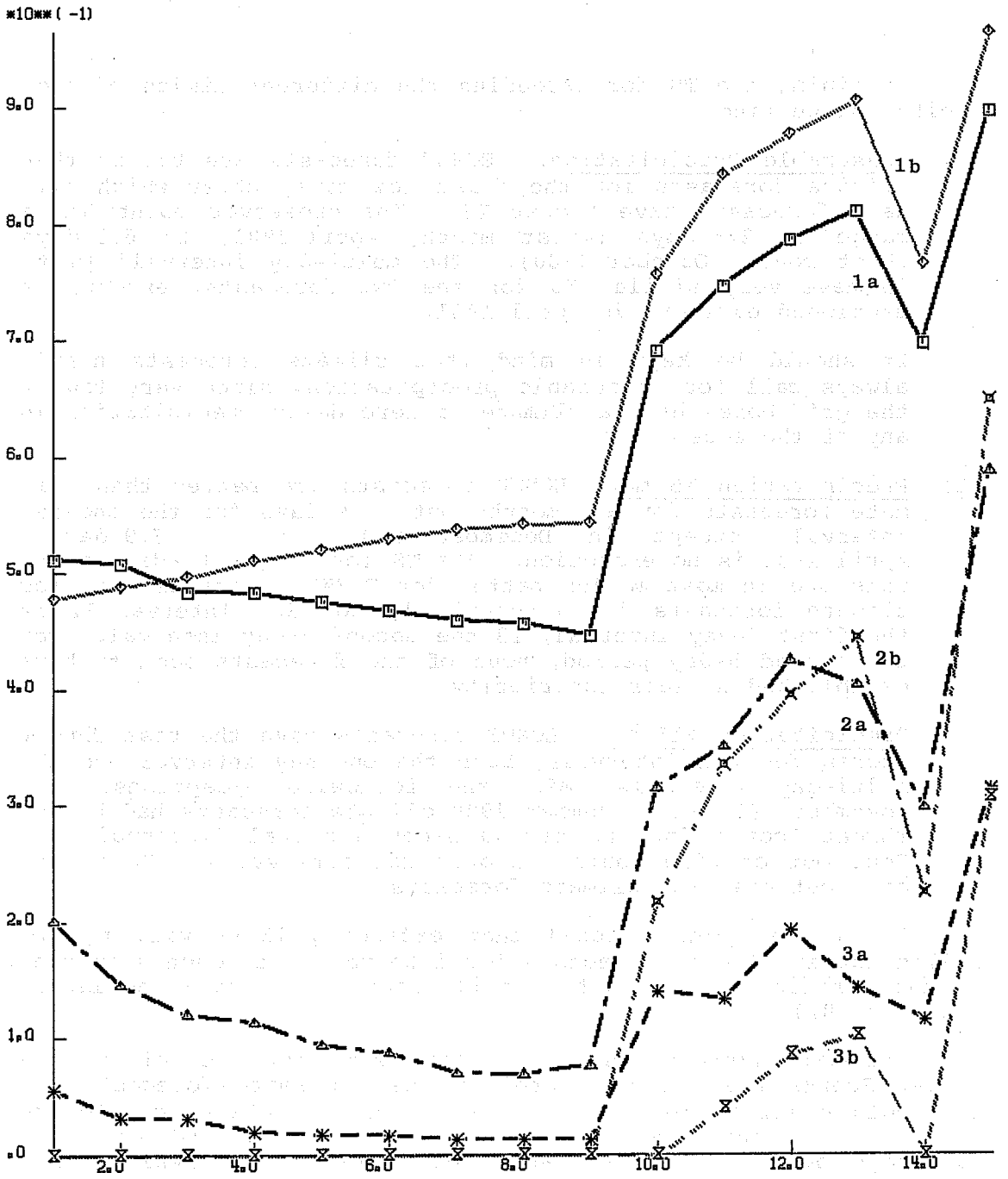

Fig. 23 Threat Score for exceeding limits of precipitation, for ECMWF forecasts and climate forecasts. Comparison of ECMWF and climate forecasts of precipitation exceeding (1) non-measurable amounts, (2) 5 mm, (3) 15 mm. Abscissa: Code for forecast interval (See Table 4). April 1981.

Dark lines (a) ECMWF forecasts
Grey lines (b) Climate forecasts

Examining the TS for exceeding the different limits of precipitation we find:

(1) Measurable Precipitation. ECMWF forecasts are better than climate forecasts for the first few days, after which climate forecasts have higher TS. The crossover point has a range of 2.5 days (worst month, April 1981) to 6.5 days (best month, October 1980). The multi-day intervals 10 to 15 have very similar TS for the two forecasts, except, as mentioned earlier, in April 1981.

It should be kept in mind that climate forecasts nearly always call for measurable precipitation, since very few of the grid boxes have a climate of zero daily precipitation in any of the seasons.

(2) Precipitation >5 mm. ECMWF forecasts are better than climate forecasts for all months out to 9 days for the one-day interval, except in December 1980 (out to 7.9 days). April 1981 is no exception. The TS for the multi-day intervals are in most months better for ECMWF forecasts than for climate forecasts for intervals 10 to 12. Interval 12 is the first 5-day interval, 13 the second 5-day interval. For the second 5-day period, none of the forecasts seem to have established a clear superiority.

(3) Precipitation >15 mm. ECMWF forecasts have the best Threat Scores for all intervals, both the one-day interval and the multi-day intervals, with the following exceptions. In November 1980 and December 1980 climate forecasts had higher Threat Scores for the second 5-day interval (interval 13). Four out of nine months' scores for interval 15 (D0 to D9) came out best for climate forecasts.

Again, one might recall that exceeding 15 mm will not be called for in climate forecasts for intervals 1 through 9 as none of the verification boxes have daily seasonal averages as large. (See Fig. 8.)

The main impressions one is left with after examining the Threat Scores are: ECMWF forecasts beat climate forecasts for threshold equal or larger than 5 mm of precipitation, out to the end of the forecast period for 24-hour intervals, and have some advantage over climate forecasts for multi-day intervals within the first 5 days. Further, ECMWF beat climate in forecasting measurable precipitation during the first 3-6 one-day intervals.

7.6 Effect of New Orography

The new orography of Fig. 1b was introduced on April 1, 1981. It departed quite radically from the old one (Fig. 1a), and it is of interest to note the effect on the verifications for April 1981, even if one month is too short a period to establish lasting conclusions.

As shown in Fig. 10, the bias went up for D1, but the D9 bias remained constant from the previous month, making April 1981 the month with the smallest forecast-cycle-induced bias increase.

The RCE (Figs. 12p and 14a-b) showed an increase and the HSS (Figs. 13p and 14c-d) a decrease from March 1981, but similar changes happened from March to April 1980 and might just reflect seasonal changes.

The Threat Score for measurable precipitation was the poorest on record, losing to climate forecasts after 2.6 days and staying well below climate for multi-day intervals. The reason cannot be that the enhanced topography caused precipitation to be called for in too many places. It will be recalled that climate forecasts under our definition call for measurable precipitation practically everywhere. The correct conclusion is probably that the new orography puts the precipitation in the wrong places.

Fig. 24 shows the biases in April 1981 for D2 and D4. The enhanced orography has not removed the deficiencies of forecast precipitation to the windward of mountainous areas. Precipitation is still underforecast here. Also, overforecasting which is greater than before, takes place to the leeward of the same areas. The mountain-induced precipitation in the model seems to be displaced and out of phase with observations more in April than in earlier months with the much flatter orography. Again, this is only one month's evidence and further months must be examined.

For thresholds 5 mm and 15 mm, ECMWF forecasts beat climate forecasts in April 1981 as in other months, both for all the 24-hour intervals to D9 and for the first threee multi-day intervals (Fig. 23c). For the 5 mm threshold, the two forecasts were about equal for the 5-day and the 9-day intervals. For the 15 mm threshold, ECMWF forecasts had highest TS in all intervals.

7.7 Performance Away from Mountains

One might expect forecasts to be better where mountains do not complicate the precipitation process in a way which is difficult to model. Also, we might expect that in boxes with many observing stations the verifying observed average would be less erratic than in boxes with just one or two stations, neither of which may be representative of the true box mean. We can test the combined effect of these hypotheses by carrying out a special verification for selected boxes remote from mountains and with higher average density of observing stations. A window in the observing network with these qualifications is shown in Fig. 6 by the heavy outline. The average station density is 6.2 per box, as against 3.2 for all 423 boxes.

Results for the window have been compared to results for the full verification area in Fig. 25a-d and 26a-d.

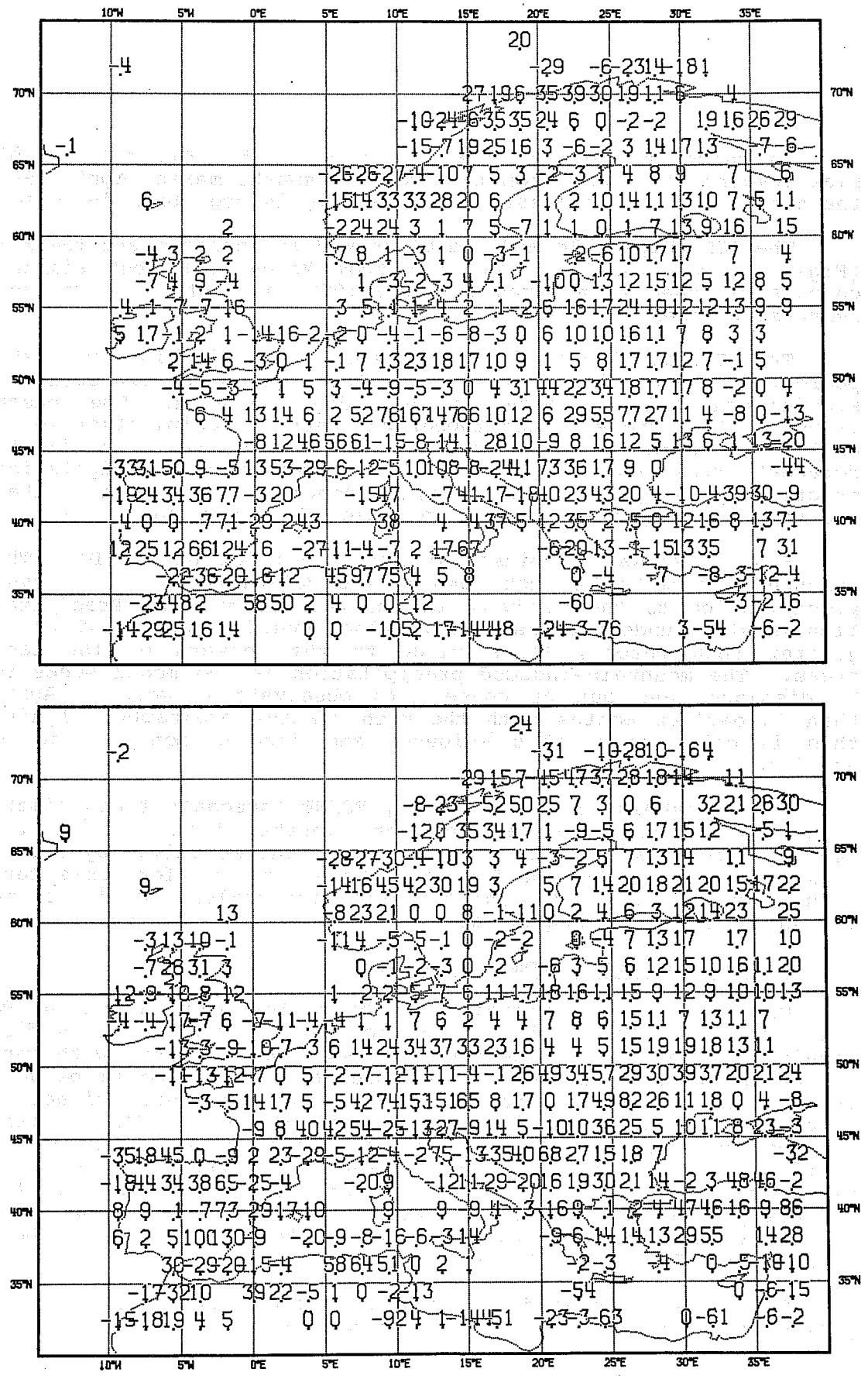

Fig. 24 Bias of precipitation forecasts April 1981, 30 forecasts. Units, 1/10 mm per day. Top: D2 (54-hour forecasts). Bottom: D4 (102-hour forecasts).

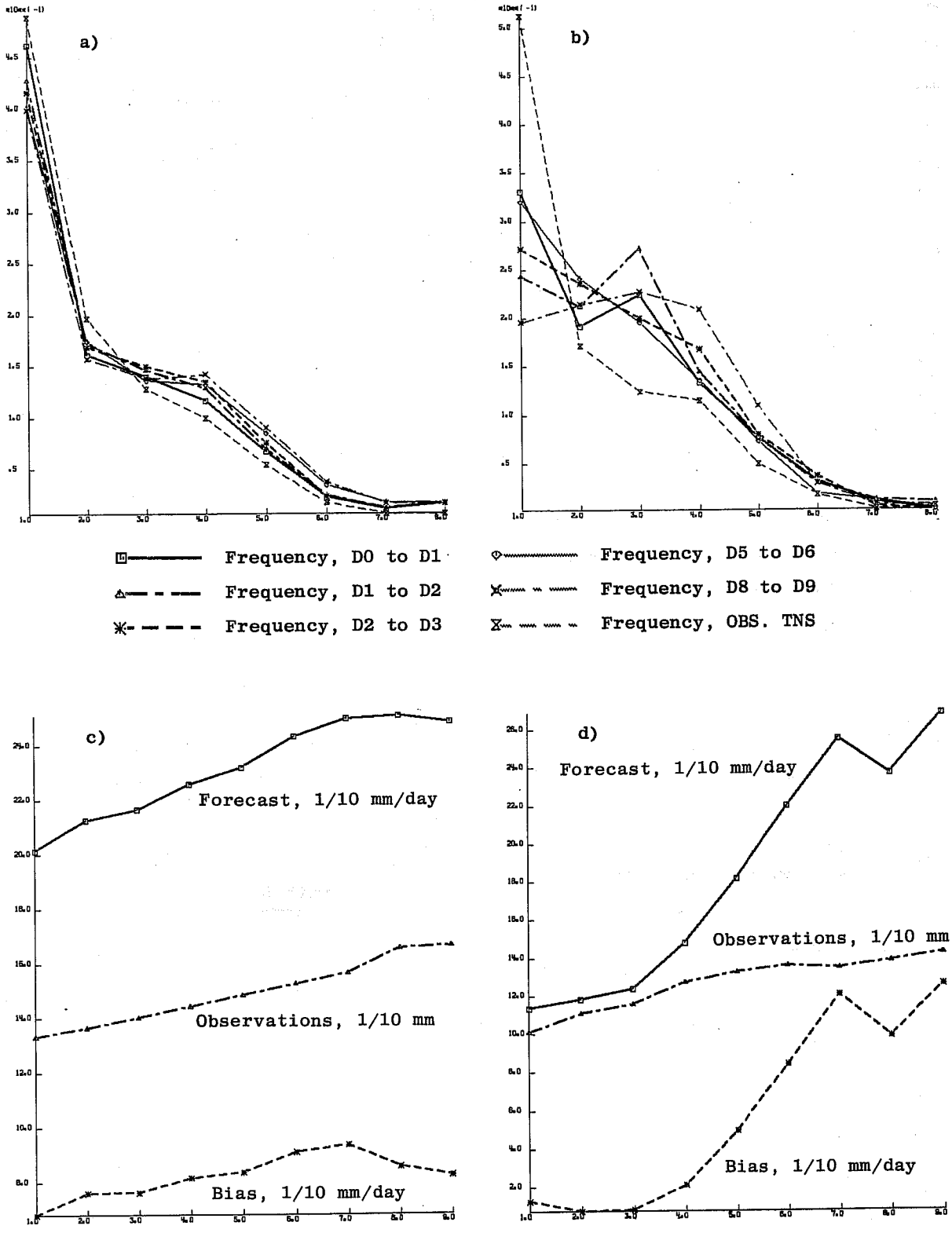

Fig. 25 Comparison of result in window of 53 grid points in flat well observed terrain with result over entire verification area. April 1981 (new orography).

(a) 8-class frequency distribution of observations and forecasts for D1, D2, D3, D6, D9 for all grid points.

(b) Same as (a) for window.

(c) Average of forecasts, observations, and bias of 24-hour precipitation, D1 to D9, for all 423 grid points.

(d) Same as (c) for window.

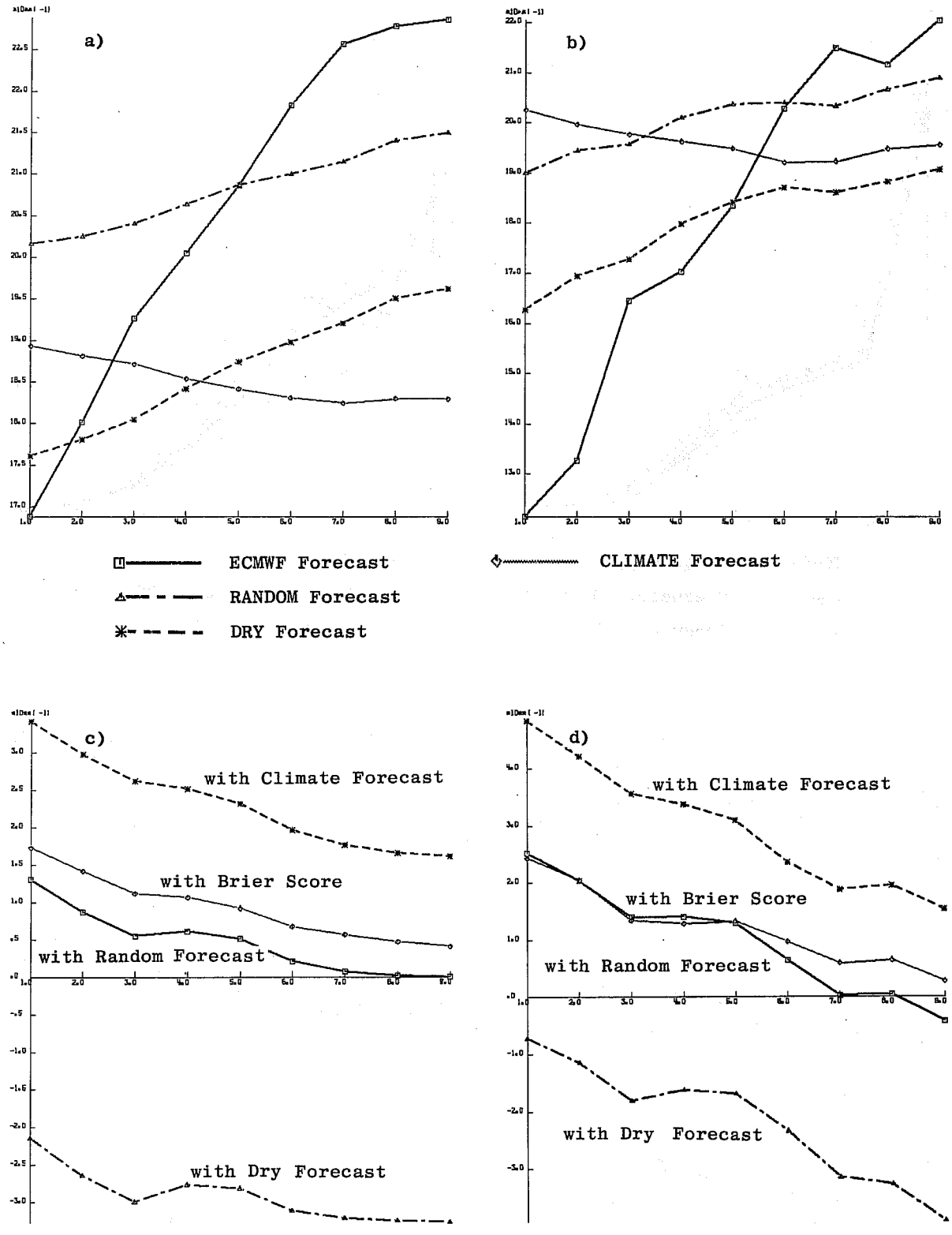

Fig. 26 Comparison of scores of ECMWF forecasts and control forecasts in window (right) with entire area (left). April 1981.
(a) & (c) RMS Class Error and Heidke Skill Score, respectively over entire area.
(b) & (d) Same over window.

In Fig. 25c forecasts, observations, and bias are averaged over the whole area and in Fig. 25d over the window. The two sets of curves are quite different. The month shown is April 1981, so the new orography is at play as well as the geographic division into a flat area and the whole area. In the window, the forecast amounts increase rapidly after the third day, doubling by the ninth day, and resulting in positive bias almost equal to the observed quantities by the end of the forecast period. On the left the positive bias is sizeable but stays almost constant during the forecast cycle at 50 percent of the observed values. It appears the problem of the forecast cycle increase in precipitation is worse over flat country than over the area as a whole. The reason for this is obscure. It may be connected with the new orography. Future months will show whether this is a fortuitous effect of synoptic situation or a typical behavior. The window is small compared to the whole area, so the chance that the synoptic situations of the month may have affected the D1 to D9 statistics is greater for the window than for the whole area.

The class distributions in Fig. 25a and b show closer affinity to the observed distribution in the whole area (a) than in the window (b), and we note that the bias developing during the cycle derives from overforecasting the middle classes 4 and 5 (2 to 10 mm).

In spite of the bias problem of the window, the scores are much better in the window than for the whole area, as can be seen in Fig. 26. Measured by the RCE the window forecasts (b) beat all control forecasts out to D5, but only to 1.7 days for the whole area (a). On D1 the RCE was 1.21 for the window and 1.68 overall.

It appears that, as surmised at the beginning of this section, the ECMWF precipitation forecasts score better over flat country than overall.

8. CONCLUSIONS

Over 16 months the ECMWF quantitative precipitation forecasts have been compared to some 1300 synoptic reporting stations over an area covering Europe and the adjacent coast of Africa. The comparisons included monthly means of forecasts versus observations for each of 423 grid points of the ECMWF operational model and 15 monthly 8x8 class contingency matrices for each of the forecast systems involved in the verification, i.e., ECMWF forecasts, random forecasts, climate forecasts, etc. The contingency matrices were processed to produce various scores customary to verification objectives. Each of these scores brings out only one particular aspect of the forecast performance. This aspect may or may not be pertinent to the answers one wants from verification. Only the matrices themselves hold all the answers.

In order to evaluate the usefulness of the ECMWF precipitation forecasts for a particular purpose, one must decide which

score is best for the purpose. It may be that none of the scores are very applicable, in which case, it is necessary to go back to the contingency tables to calculate the mode of application and the usefulness.

For these reasons, the conclusions that can be drawn from a verification program are quite general when it comes to evaluation of usefulness.

The main findings of this verification study may be summarized as follows:

(1) Some of the biases of the precipitation forecasts are systematically connected with geography. Forecasts can be improved at the local level by being aware of these biases. This can for instance be done by bringing monthly bias maps to the attention of field forecasters. The main features of the bias pattern are the areas of underforecasting to the windward of mountainous areas and overforecasting to the lee of the same area.

(2) The overall bias of the verification area increases during the forecast cycle by about 30 percent of observed values. It is a modest increase spread over the forecast period of 10 days. In itself the systematic increase does not cause unrealistic forecasts.

(3) Taken over the whole verification area, forecasts and observations are well balanced, starting with about 15 percent underforecasting on D1 and ending with about 15 percent overforecasting on D9.

(4) Areas of precipitation tend to spread during the cycle, or conversely, dry areas tend to shrink. This leads on the average to overforecasting of areas of measurable precipitation already early in the forecast cycle.

(5) By and large the ECMWF precipitation forecasts have skill above control forecast level for the first 3 days of the forecast cycle. In some months this is increased to 5 days. This is a very encouraging result. It is thought by some that precipitation, because of its strong mesoscale component, is not amenable to numerical prediction beyond quite short time frames. Our verifications show that at ECMWF precipitation as an element of routine forecast guidance has entered medium-range skill levels. It should be kept in mind that the lower limit of the scales considered in this study are about 25000 square kilometers and 24 hours.

(6) Beyond the 3-day range the skill becomes selective of the skill score used, in other words, the forecasts, used as they come, may be useful for some purposes but not for others. For some operations forecasts for the entire 9-day cycle may be useful.

References

Brier, Glenn W. and Allen, Roger A. "Verification of Weather Forecasts." Compendium of Meteorology, American Meteorological Society, Boston, Mass., 1951, pp. 841-848

Charba, Jerome P. and Klein, William H. "Skill in Precipitation Forecasting in the National Weather Service." Bulletin: American Meteorological Society, Vol. 61, No. 12, December 1980, pp. 1546-1555.