TECHNICAL MEMORANDUM

227

Extreme rainfall prediction using the ECMWF Ensemble Prediction System

T. Petroliagis, R. Buizza,
A. Lanzinger and T. Palmer

Research Department

May 1996

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

Extreme Rainfall Prediction using the ECMWF Ensemble Prediction System

T Petroliagis, R Buizza, A Lanzinger and T N Palmer

European Centre for Medium-Range Weather Forecasts

Shinfield Park, Reading

United Kingdom

ABSTRACT

The combined use of the European Centre for Medium-Range Weather Forecasts (ECMWF) high-resolution at T213 spectral triangular truncation and with 31 vertical levels (T213L31) operational model and Ensemble Prediction System (EPS), during cases of intense Mediterranean storms, is studied. In particular, it is discussed how EPS products can be used to provide a measure of confidence in the high-resolution precipitation forecast.

Three case studies (two extreme events plus one false-alarm case) are analyzed. The first event took place between the 21st and the 22nd of October 1994 over Greece where heavy rainfall led to local flash-floods in many areas that cost the loss of twelve lives, and caused significant property damage. The second event occurred in northwest Italy (and some parts of southern France) exactly two weeks later. More than 60 lives were lost and the cost of damage was enormous. For both these cases, the EPS probability values for precipitation occurrence supported the medium-range T213L31 prediction, which proved to be successful.

A third case is also investigated, where the high-resolution forecast suggested heavy rainfall over northern Italy but was not supported by the EPS. The T213L31 prediction for this case was poor.

Finally, the reliability of EPS probability predictions is discussed. It is shown that probability forecasts of low-level temperature and wind show considerable skill. Reliability statistics of clusters of large-scale flow show considerable skill as well. Probabilistic precipitation predictions are less skilful, possibly due to the EPS model resolution.

EPS forecasts of extreme weather events are necessarily compromised by the moderate resolution of the T63L19 model (a version of the operational high-resolution T213L31 model, at T63 spectral triangular truncation and with 19 levels) used to generate the ensembles. In future studies, ensembles will be made using at least T106L31 resolution combined with an increase in ensemble size.

1. INTRODUCTION

Since December 1992, the European Centre for Medium-Range Weather Forecasts (ECMWF) operational 10-day predictions have included both a high-resolution deterministic forecast and an Ensemble Prediction System (EPS). The development of the ECMWF EPS follows the work of *Epstein* (1969), *Gleeson* (1970), *Fleming* (1971a-b) and *Leith* (1974), who laid the theoretical and numerical foundations of a probabilistic approach to weather forecasting.

At the time of writing, the high-resolution model has a T213 spectral triangular truncation and 31 vertical levels (*Simmons et al*, 1989, *Courtier et al*, 1991), whilst the EPS consists of 33 lower resolution (T63L19) forecasts (*Molteni et al*, 1996), made by perturbing the operational analyses using dynamical singular vectors (*Buizza and Palmer*, 1995).

A major responsibility of operational meteorological offices is to issue warnings when severe weather events are predicted. But false alarms have also to be minimised, since they can cause a loss of credibility in the meteorological operational centre itself. Accurate probability forecasts for such severe events will in principle reduce the number of false alarms. By definition, a single deterministic forecast cannot give such probabilities.

The purpose of this study is to show how the EPS probability predictions can be used to estimate the degree of confidence that could be associated with a deterministic high-resolution prediction. The discussion focuses particularly on precipitation for three different cases. The study is performed using the type of products available to the ECMWF Member States' Meteorological Operational Services.

The first two cases were characterized by very intense rainfall, which led to floods, loss of lives and enormous damage. For both cases the deterministic prediction of precipitation proved to be very successful and well supported by EPS forecasts of the probability of precipitation exceeding predefined thresholds. By contrast, the EPS did not support the deterministic prediction of heavy rainfall for the third case, for which the operational high-resolution prediction proved to be wrong.

The degree of support of the operational forecast by the EPS is related to the overall ability of the EPS to provide reliable probability forecasts. A probabilistic forecast is considered to be reliable if the observed relative frequency of the occurrence of a predefined event has a value similar to the forecast probability. The reliability of the EPS is assessed quantitatively in this way using the so-called reliability diagrams.

The paper is organized as follows. In Section 2, a brief description of the ECMWF EPS is given, in Section 3, a synoptic overview of the three analyzed cases is reported, while Sections 4, 5 and 6 discuss the deterministic and probabilistic predictions for the three cases. The validation of the EPS probability products in terms of reliability diagnostics is presented in Section 7. Conclusions are drawn in Section 8.

2. ECMWF Ensemble Prediction System

The ECMWF EPS comprises, at the moment of writing, 32 perturbed and one unperturbed (control) non-linear integrations of a version of the ECMWF model (*Simmons et al*, 1989, and *Courtier et al*, 1991) with spectral truncation T63 and 19 vertical levels. The initial conditions are created by adding and subtracting 16 perturbations to the control initial conditions. The initial perturbations are defined using the singular vectors (*Buizza and Palmer*, 1995) of a linear approximation of the ECMWF model. The singular vectors (SVs) identify the most unstable directions of the system growing over a finite time interval (48 hours) named the optimisation time interval.

Once the perturbations have been constructed (as linear combinations of 16 selected singular vectors), they are added and subtracted to the control initial conditions to define 32 perturbed initial conditions. Then, 32 plus 1 (control) 10-day T63L19 non-linear integrations are performed. With the current ECMWF computer facilities (CRAY C90 with 16 processors), the total elapsed time is approximately 2.7 hours, which is about 1.3 times the elapsed time needed to performed the 10-day T213L31 ECMWF operational forecast.

Schematically, the main steps of the ECMWF EPS can be described as follows:

Step (1) *Computation of 30 to 35 SVs*

Step (2) *Selection of 16 SVs*

Step (3) *Construction (Rotation) of Initial Perturbations*

Step (4) *32 + 1 Non Linear T63L19 Integrations*

Step (5) *Post-Processing - Dissemination of EPS Products*

The main ensemble products (Step 5) which are routinely disseminated to the Meteorological Services of the ECMWF Member States are listed hereafter.

i. "Stamp" Maps

The set of 500 hPa geopotential height forecast maps over Europe can be plotted on a single sheet of paper of A4 size or similar. Although the size of an individual map is clearly minimal, it conveys the principal features of the synoptic-scale flow. An experienced forecaster can readily spot whether the synoptic development of one or two individual ensemble members is unusual, and therefore worthy of further investigation.

ii. Clusters of 500 hPa Height Trajectories

To condense the number of flow patterns predicted by the ensemble members into more basic varieties, a cluster analysis on the 500 hPa height fields produced by the 33 individual forecasts is performed. As in *Brankovic et al* (1990) and *Palmer et al* (1993), we have used Ward's hierarchical clustering algorithm (e.g. *Anderberg,* 1973).

For operational implementation, it was felt that information for more than one forecast time should be provided, while avoiding potentially confusing situations in which the grouping of ensemble members varied at different forecast ranges. It was therefore decided to cluster portions of forecast trajectories rather than instantaneous fields. This was done by defining the 'distance' between two ensemble members as the root mean square (rms) difference (see Annex for precise definition) between height fields in the time interval from day 5 to day 7.

Of course, each clustering option (such as the choice of the clustering area and the time

window, or the criterion for the 'best' number of clusters) has advantages and disadvantages; the usefulness of objective clustering would be greatly increased if these choices were made at the 'consumer' (i.e., operational forecaster) level rather than at the 'producer' (i.e., ECMWF) level. For operational purposes the maximum number of clusters has been set equal to six while the minimum equal to two. The clusters are ranked with respect to the number of elements they comprise of (i.e. descending order). Cluster-1 Centroid (the mean field averaged over all the cluster-1 members) always represents the most populated cluster (examples of EPS clusters are shown in Figs 2 and 7).

iii. Probability "Plumes"

In order to give an assessment of the ensemble dispersion occurring throughout the forecast range at a particular location, "plumes" showing the time evolving probability that the 850 hPa temperature lies within intervals of 1 degree width are disseminated. The probabilities are computed assuming that each ensemble member is equally likely, and are expressed as percentages of the maximum possible value (which corresponds to all 33 ensemble forecasts being in a 1 degree interval). A Gaussian smoother is applied to the sample frequencies to produce smooth probability estimates.

iv. Probability maps

Two dimensional fields representing probabilities of rainfall, 10 m wind speed and 850 hPa temperature anomalies for specific forecast days are also post-processed and disseminated. As with the plumes, the probabilities are calculated on the basis that each ensemble member is equally likely. The rainfall categories are: >1mm/day, >5mm/day, >10mm/day and >20mm/day. The wind speed categories are: >10m/s and >20m/s. Finally the temperature anomaly categories are: <8K, <4K, >4K and >8K. Examples of probability fields for rainfall and temperature categories are given in Figs. 3c, 4, 8c, 9, 10c and 11).

3. SYNOPTIC OVERVIEW

3.1 The Italian storm: 4-6 November 1994

During the first week of November 1994, heavy rainfall caused catastrophic floods and land

slides over southern France and northern Italy. The north-western Italian Piedmont was the region which suffered the most, with the loss of more than 60 lives.

A detailed description of the evolution and spatial variability of the observed precipitation can be found in *Buzzi et al* (1995), who identified the period between the 12 UTC of the 5th and 6th of November as the most critical (also see *Athanassiadou and Thorpe,* 1995, *Binder and Rossa,* 1995, *Buzzi and Tartaglione,* 1995, and *Dorninger and Hantel,* 1995). Our analysis will concentrate on this time interval.

Between the 2nd and the 4th of November 1994 an upper-level cyclonic circulation system was propagating across the northern Atlantic with an associated trough deepening east of Iberia. The associated low-level advection of warm and moist Mediterranean air-masses led to moderate and heavy rain-falls over Provence, the French and Italian Alps, and the western Po valley from the 2nd of November.

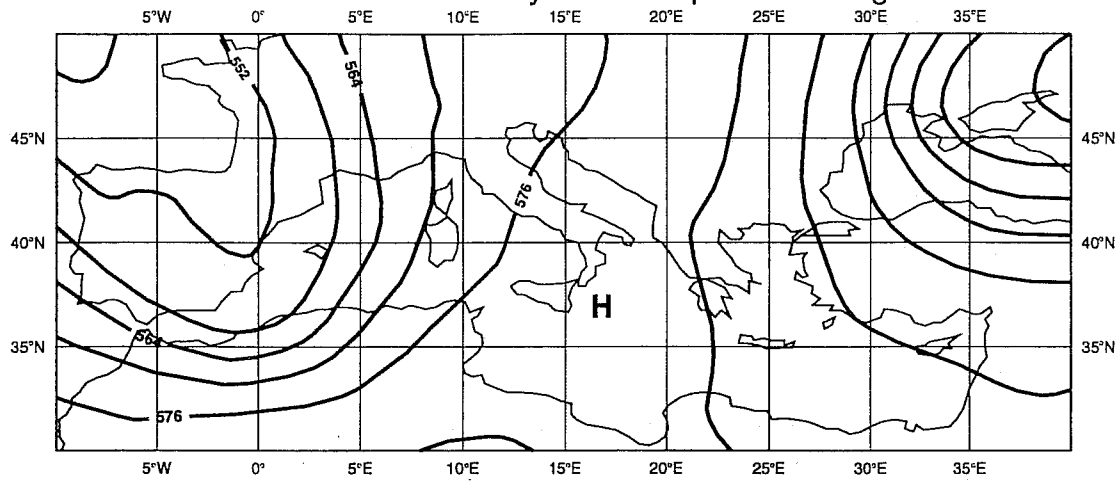
On the 5th of November 1994, while the eastward progression of the upper-level low slowed and eventually became quasi-stationary off Ireland, the upper-level trough moved across Iberia into the western Mediterranean (Fig. 1a). Meanwhile, low-level advection of warm moist air continued. On the 6th of November, a cut-off formed over the Baleares and moved across Corsica and northern Italy to the eastern Alps. As a result of the progression of cold air into the western Mediterranean, of the sharpening of air-mass gradients and of the formation of the cut-off low, frontal activity and associated rainfalls were substantially enhanced.

3.2 The Hellenic storm: 20-22 October 1994

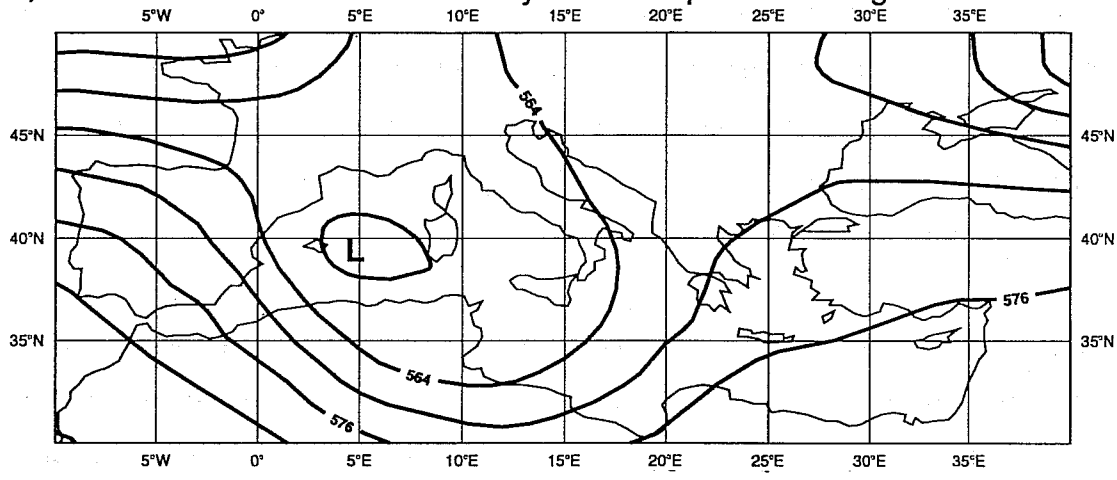
Between the 21st and the 22nd of October 1994, heavy rainfall caused major disruption in many areas of the Hellenic peninsula and more specifically in the greater area of Athens. The catastrophe caused the loss of twelve lives and significant damage to property (for more details see *Lagouvardos et al,* 1995).

On the 20th of October 1994, an upper level trough (with a NW-SE tilt) associated with a

a) 5 November 1994 12 UTC - Analysis of Geopotential Height 500 hPa



b) 21 October 1994 12 UTC - Analysis of Geopotential Height 500 hPa



c) 15 September 1994 12 UTC - Analysis of Geopotential Height 500 hPa

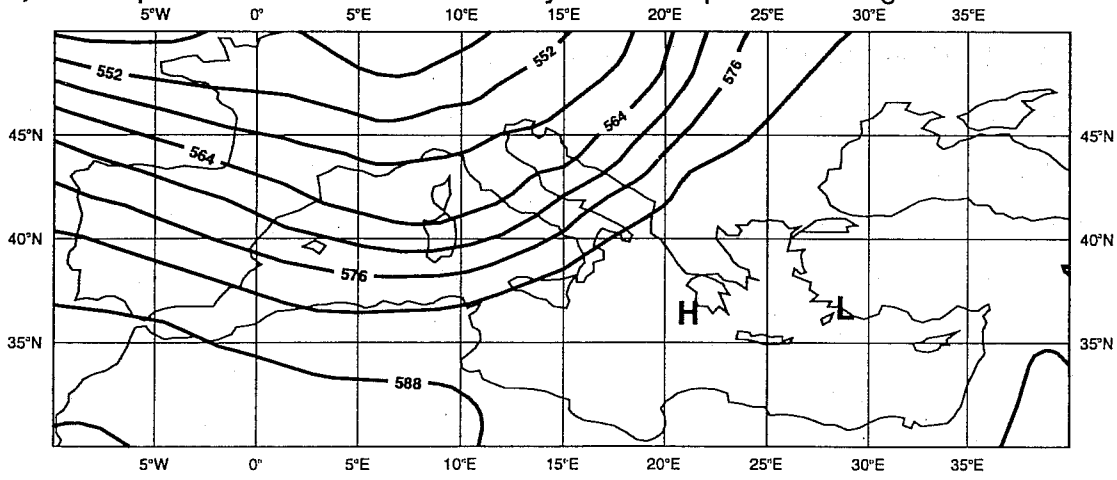


Fig 1 a) ECMWF 500 hPa geopotential height analysis for 94.11.05 12.00 UTC. b) as a) but for 94.10.21. c) as a) but for 94.09.15. Contour interval: 60 m.

deep low off Ireland was penetrating south-ward over western Europe, leading to a cut-off low over the western Mediterranean on the 21st of October 1994 (Fig. 1b). On the forward side of this system low-level warm moist air was advected over the Hellenic peninsula. During the following days the low circulation system moved slowly east-wards and the associated continuous convergence over the Greek mainland of air-masses originating from the central and southeastern Mediterranean was enhanced. The combined effect of the convergence of the warm moist air-masses and their interaction with colder air-masses converging from the northeast (being advected on the southern side of a continental high pressure system), resulted in extreme rainfalls over many areas of Greece.

3.3 The false-alarm case: 14-16 September 1994

On 14 September 1994, a deep low was centred over north-western France. On its southern side Mediterranean air-masses were advected over Italy towards central Europe. During the following 24 hours the low circulation system deepened and moved east-wards with its centre over Denmark starting to fill slowly. Westerly, slightly cyclonic flow prevailed over the western Mediterranean area (Fig. 1c).

Considering the 24 hour time interval (between 12 UTC of the 15th and the 16th of September 1994), no intense precipitation was observed over the area of northern Italy, where the operational high-resolution day-5 forecast had predicted significant amount of precipitation.

4. PREDICTION OF THE ITALIAN STORM

The area of interest for all three cases is taken to be the greater Mediterranean region (GMR) with latitude between 30 and 50 degrees north, and longitude between 10 degrees west and 40 degrees east. In all three cases we concentrate on the ECMWF T213L31 and EPS predictions 5 days prior to the event. The forecast of both the 500 hPa geopotential height field on 05.11.94 (day-5 forecast initiated on 31.10.94), and the accumulated precipitation between 12 UTC of the 5th and the 6th of November 1994 (day-5 precipitation forecast initiated on 01.11.94) are verified.

i) Synoptic flow

The T213L31 day-5 500 hPa geopotential height forecast (Fig. 2a) is generally skilful though the trough is positioned too far south. The anomaly correlation skill score (ACC) is 67%, above 60%, i.e. the so-called limit of useful skill (more about the ACC skill score and the useful threshold limit can be found in the Annex).

The deterministic forecast is supported by the principal EPS clusters centroids. Figure 2b shows the day-5 forecast given by the cluster-1 centroid. It is worth recalling that the clustering algorithm produces a maximum of 6 and a minimum of 2 clusters from the 33-member ensemble. The ECMWF operational clustering algorithm has been used in this paper, but the clustering has been performed using only day-5 fields (since we concentrate on predictions 5 days prior the main impact), i.e., without considering the forecast trajectories between day 5 and day 7.

The cluster-1 centroid (Fig. 2b) contains 15 EPS members and has an anomaly correlation skill score of 65%. Although its predicted field is smooth, due both to the T63L19 EPS (lower) resolution and to averaging over (15) cluster members, it provides a skilful forecast of the large-scale flow. In particular, it supports the T213L31 day-5 forecast verifying on the same day. The ratio between the number of ensemble members included in the cluster and the number of the EPS members (45%) can be used as an indication of the probability that this type of large-scale flow can occur (this probability product is verified in section 7).

Figure 2c shows cluster-2, the cluster characterized by the highest anomaly correlation skill score, 85%. Its main difference from cluster-1 is a better prediction of the trough over the Black sea. Cluster-2 comprises 10 EPS members. As far as the circulation over the western and the central Mediterranean region is concerned, cluster-2 does not give an alternative scenario to cluster-1, and thus one could estimate the probability that the T213L31 day-5 prediction will occur to be about 75%. By contrast, cluster-2 gives an alternative (synoptic) scenario for the area of eastern Mediterranean and Black sea.

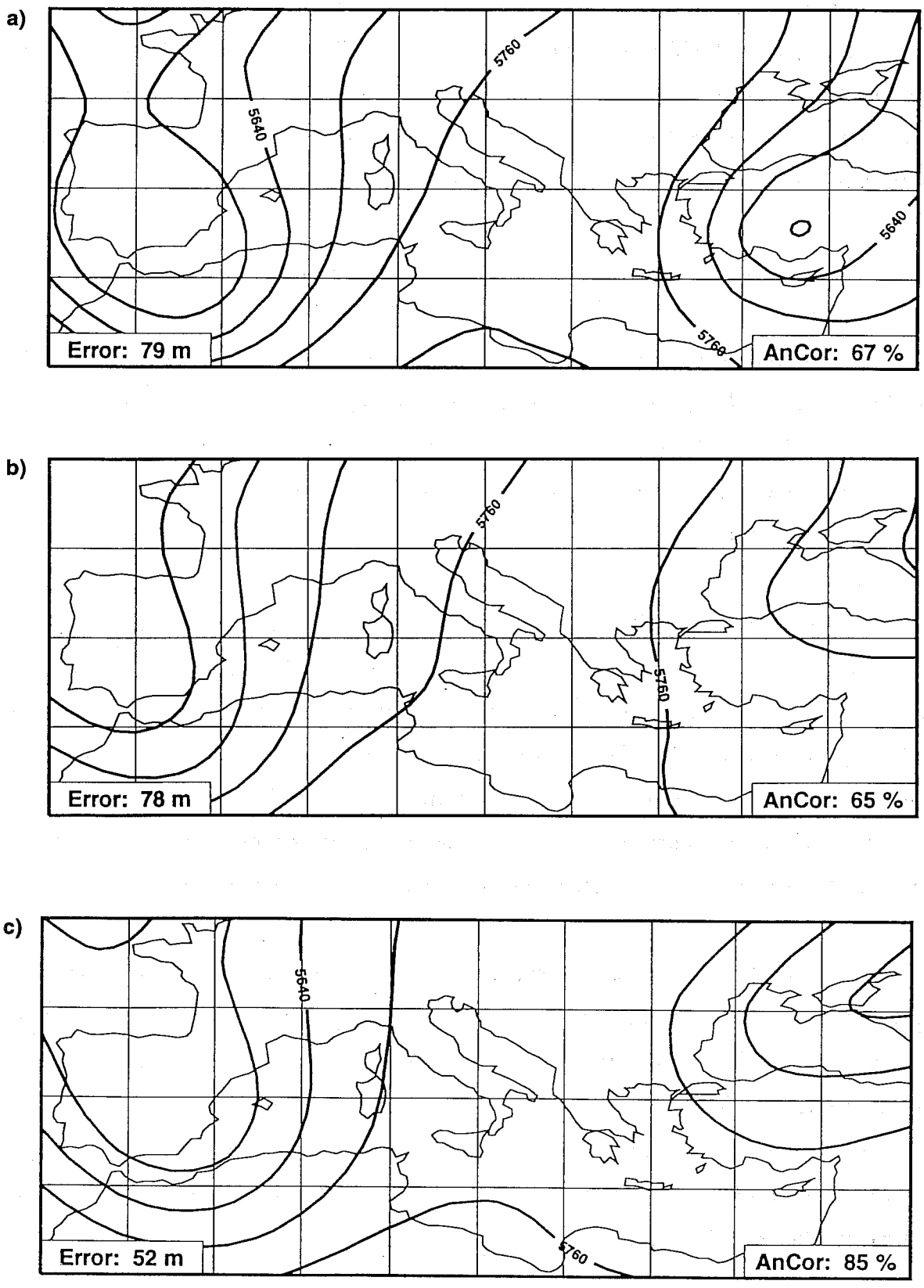


Fig 2 a) T213L31 day-5 forecast for 500 hPa geopotenUal height initiated from 94.10.31. b) as a) but for the EPS cluster-1 centroid. c) as a) but for the EPS cluster-2 (best-cluster) centroid. Contour interval: 60 m.

ii) Precipitation

A detailed picture of the total precipitation observed over the area of interest between 12 UTC of the 5th and the 6th of November 1994 can be found in *Buzzi et al,* (1995). For this interval, values consistently exceeding 100 mm/day were reported. Here (and for the other two remaining cases) we consider the 24h precipitation forecast from T63 control to highlight areas where the observed precipitation exceeded the 20 mm/day (Fig. 3a). These areas agree well with areas obtained from a similar set of analyses made using direct (observed) rainfall.

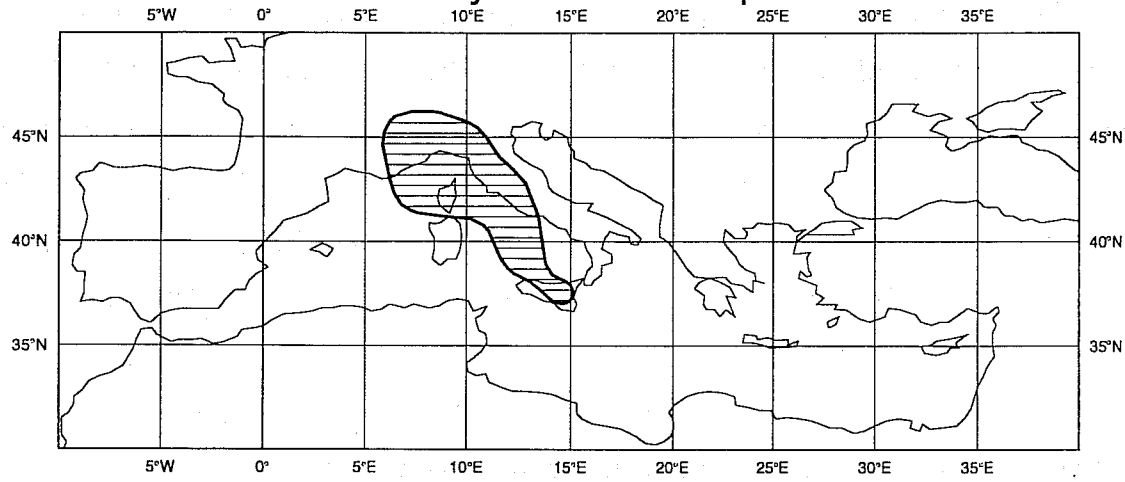
The T213L31 day-5 precipitation forecast is very skilful (Fig. 3b). The predicted pattern is positioned almost correctly (as indicated by the shaded areas corresponding to values exceeding the 20 mm/day), and the rainfall amounts (a peak value of 135 mm/day is forecasted) give a clear warning of severe weather over the region of interest.

Figure 3c shows the day 4 to day 5 probability forecast initiated on 01.11.94 for the event 'total precipitation exceeding 20 mm/day' to occur. Note that a 10% probability means that about 3 ensemble members support the occurrence of the event. Also note that since the EPS resolution is T63L19, the relatively poor resolution can be a substantial handicap especially in orographic regions. However, overall the probability map strongly supports the T213L31 prediction, and thus provides the forecaster with additional confidence in the high-resolution forecast.

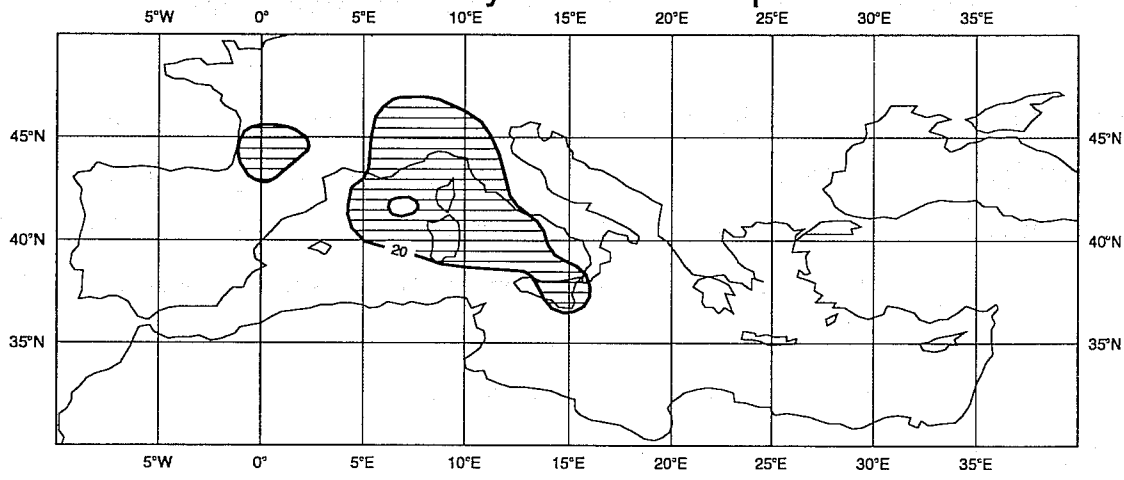
Additional confidence can be obtained by studying the consistency of the EPS from previous days. Figure 4 shows the evolution of the 2-dimensional probability maps from the EPS started 8, 7, .., 3 days before the event, all verifying over the same time interval. Significant possibility of an extreme event is already detectable in the day-6 forecast from the EPS initiated on 31.10.94 (Fig. 4c). Starting from the EPS initiated on 01.11.94, the signal of an extreme event becomes clear and consistent (Fig. 4d-f). The strong consistency between the three more recent EPS forecasts should add confidence in the T213L31 forecast.

Forecasters are frequently asked to issue local predictions together with forecasts valid for

a) 94.11.05 - T63L19 Day-1 Total Precipitation Forecast



b) 94.11.01 - T213L31 Day-5 Total Precipitation Forecast



c) 94.11.01 - Probabilities for 'precipitation exceeding 20 mm/day'

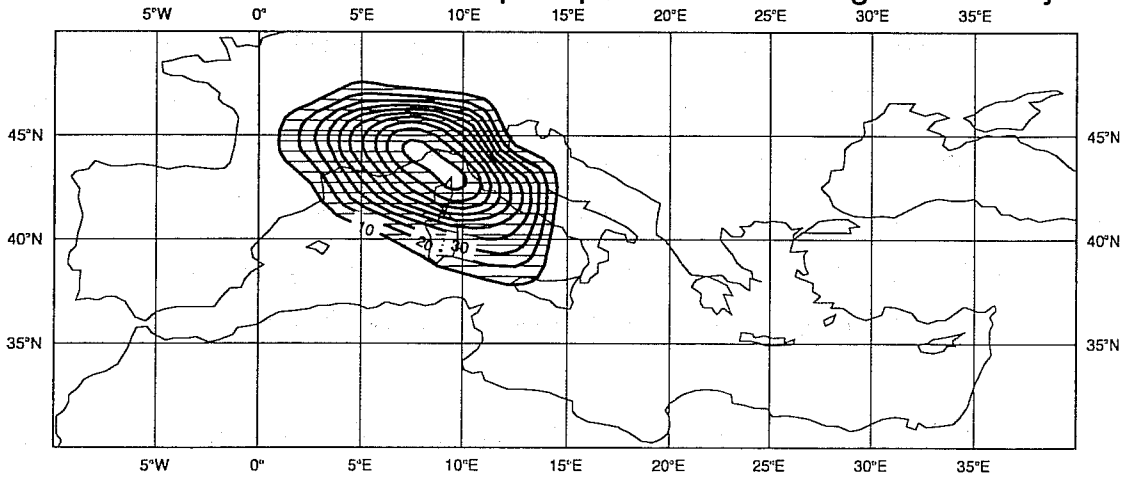


Fig 3 a) T63L19 precipitation prediction of the forecast initiated on 94.11.05, accumulated between day 0 and day 1 (areas over 20 mm/day are shaded). b) T213L31 precipitation prediction of the forecast initiated on 94.11.01, accumulated between day 4 and day 5 (areas over 20 mm/day are shaded). c) EPS probability values for the event 'total precipitation exceeding 20 mm/day' for the EPS started on 94.11.01 (precipitation accumulated between day 4 and day 5). Contour isolines every 10%.

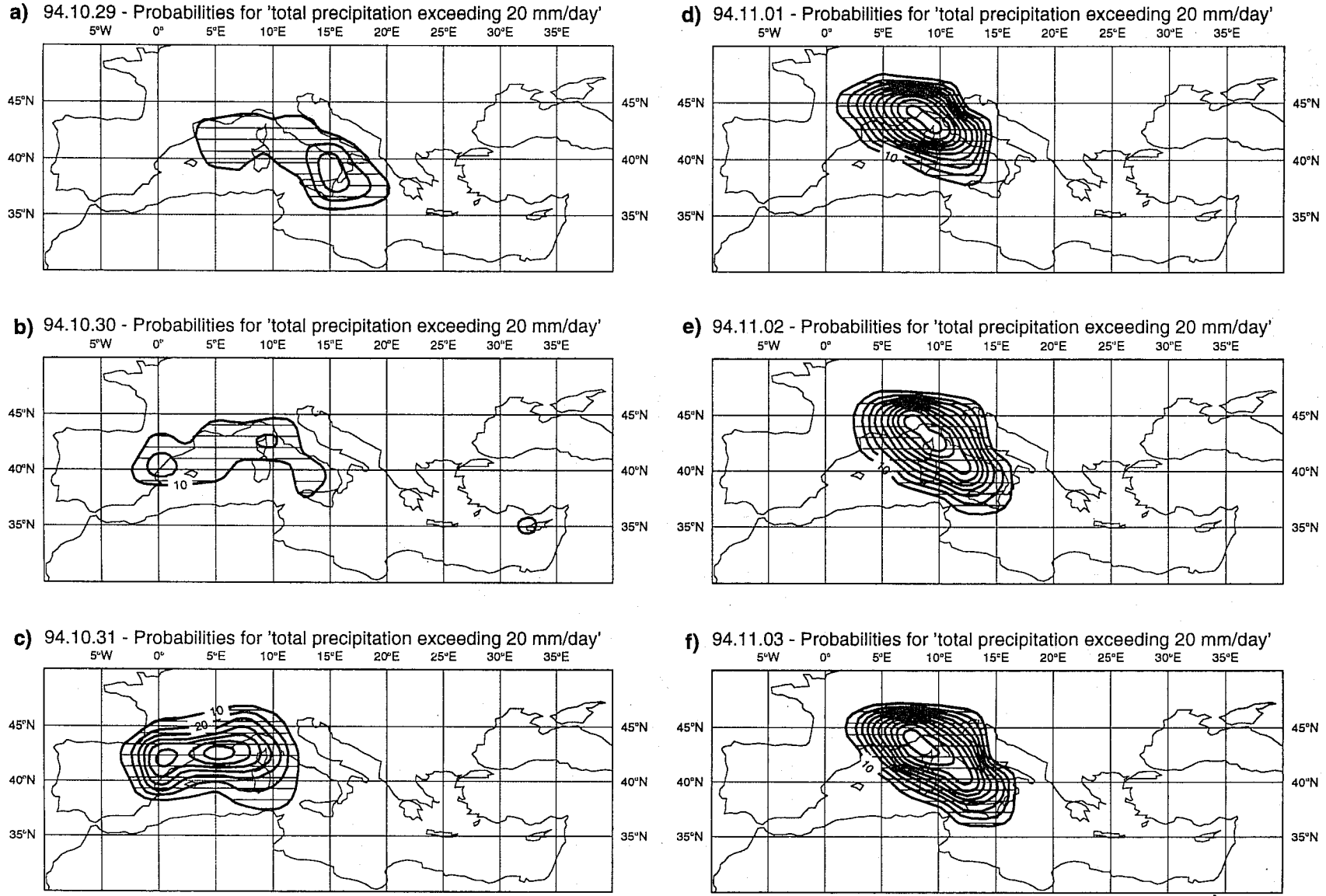


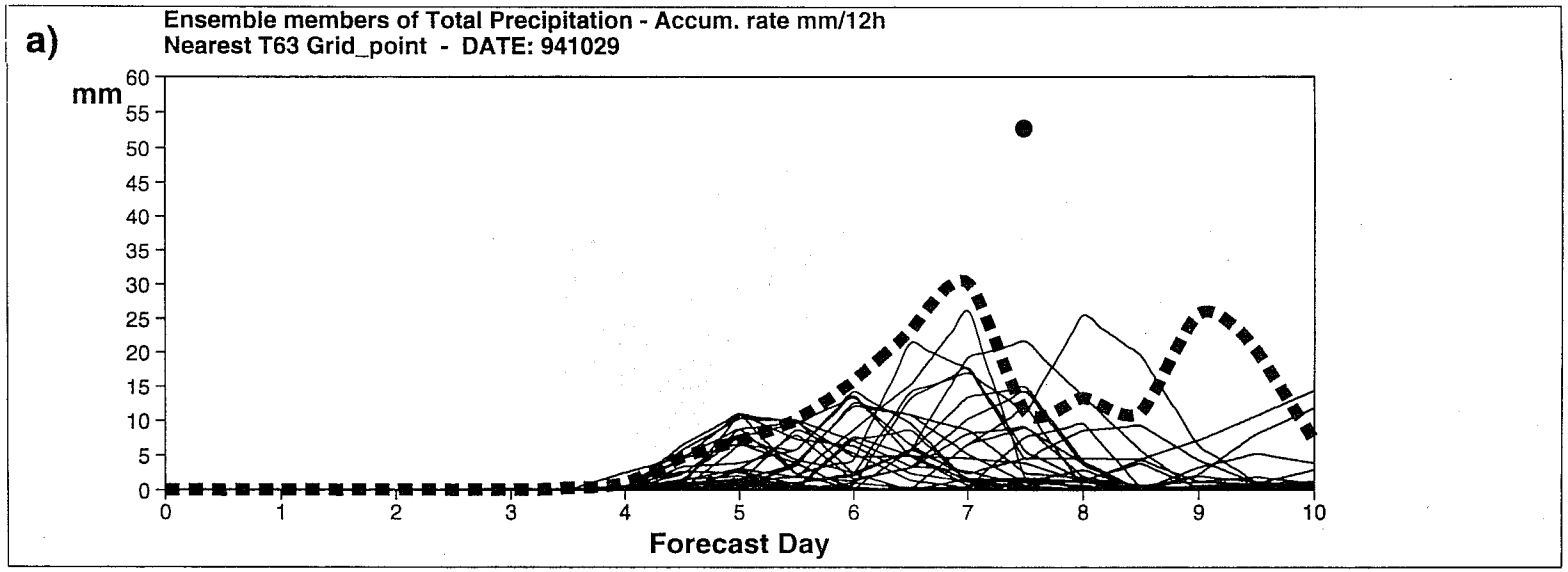
Fig 4 Probability prediction for the event 'total precipitation exceeding 20 mm/day' to occur between 12 UTC of 5 and 6 November 1994, forthe EPS initiated on a) 94.10.29 (precipitation accumulated between day 7 and day 8), b) 94.10.30 (precipitation accumulated between day 6 and day 7), c) 94.10.31 (precipitation accumulated between day 5 and day 6), d) g4.11.01 (as Fig. 3c), e) 94.11.02 (precipitation accumulated between day 3 and day 4), and f) 94.11.02 (precipitation accumulated between day 2 and day 3). Contour isolines every 10%.

larger areas. Figure 5 shows the 10-day precipitation prediction for the T63 grid-point (land-point) closest to the Piedmont area (with coordinates: lat 43.8 N - lon 7.5 E), given by the EPS started 8, 7, .., 3 days before the event (panels a, b, .., f, respectively of Figure 5). Note that values are reported every 12 hours, and thus represent total precipitation accumulated over a 12 hour time interval. A mark (dot) on each panel identifies the 12 hour accumulated precipitation from observations (obtained by averaging four observations surrounding the selected point, as being reported by *Buzzi et al,* 1995) valid for 00 UTC 06.11.94. Note how the dispersion of the predicted rainfall decreases as the EPS starting date gets closer to the verifying time interval, and how this corresponds to a better agreement between the predicted and observed value.

iii) Spread and skill

Figure 6 shows the skill (taken as the RMSE distance between the forecast field and verifying analysis, for 500 geopotential height values) of the T213L31 deterministic forecast, and the spread (taken as the RMSE distance between any ensemble member and the control forecast, for 500 hPa geopotential values) and skill of the 33 EPS members started on 31.10.94, computed for the area defined at the beginning of this section (i.e. area GMR). The T213L31 is extremely skilful up to day 4. Nevertheless, at forecast day 5 which corresponds to the extreme precipitation event, the T213L31 retains a high anomaly correlation skill score. The spread of the EPS reflects the T213L31 skill: it is very small up to forecast day 4, and then starts increasing. This should be considered as an indication of loss of predictability.

The comparison between the skill of the T63L19 control and the T213L31 operational forecast gives an estimate of the impact of resolution on the forecast skill. The T213L31 gives the best forecast during the first 1.5 days. However even at forecast day 2, 11 EPS (lower resolution) members have a higher anomaly-correlation skill score than the T213L31, (more about for both the RMSE and ACC skill scores can be found in the Annex). At forecast day 5, 20 EPS members are more skilful than the T63L19 control, and 12 are better than the T213L31 forecast. It should be noted, however, that despite a high correlation score, none of the T63L19 members were able to forecast the intensity of precipitation of the T213L31



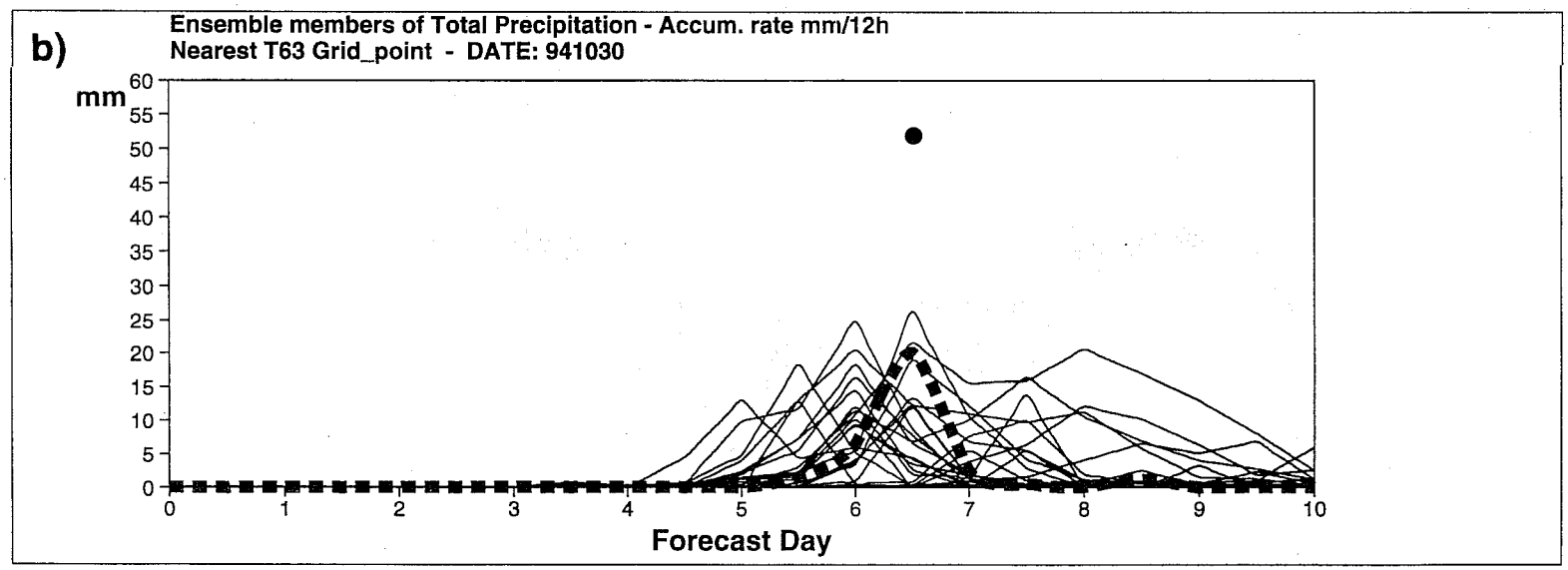
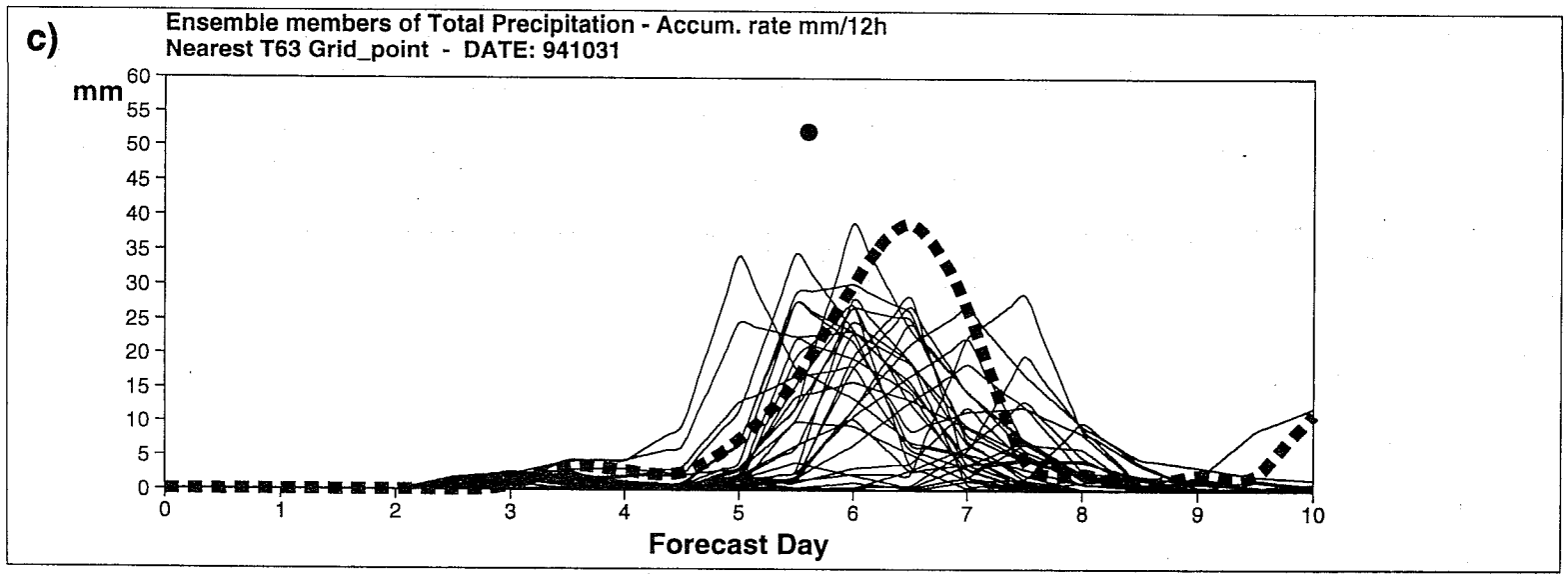


Fig 5 10-day EPS plumes of the T63 land grid-point closest to the Piedmont area (lat: 43.8 N - lon: 7.5 E), of total precipitation predicted by the T213L31 (bold dash) and by the 33 EPS members (thin solid), for the predictions started on a) 94.10.29, b) 94.10.30, c) 94.10.31, d) 94.11.01, e) 94.11.02 and f) 94.11.03. Values are reported every 12 hours. For each time *t* the corresponding value represents the precipitation cumulated between *t- 12h* and *t.* A mark on each panel identifies the amount of the 12 hour accumulated precipitation from observations valid for 00 UTC 94.11.06.



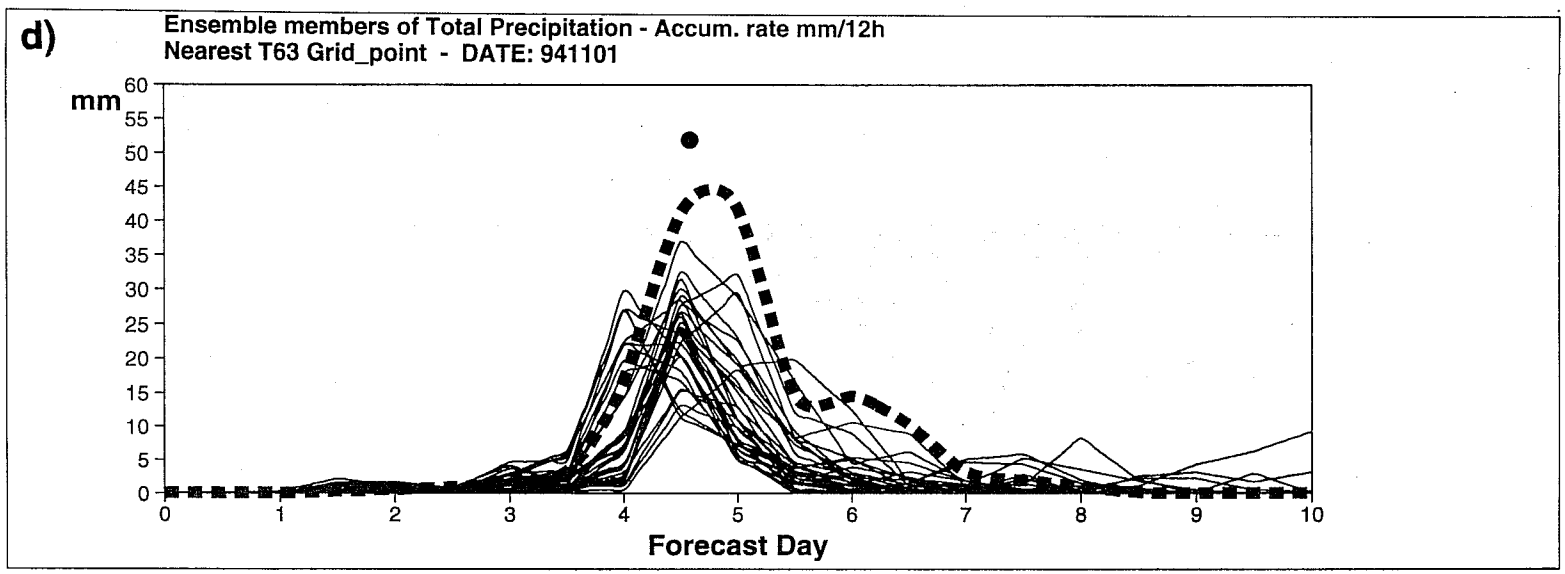
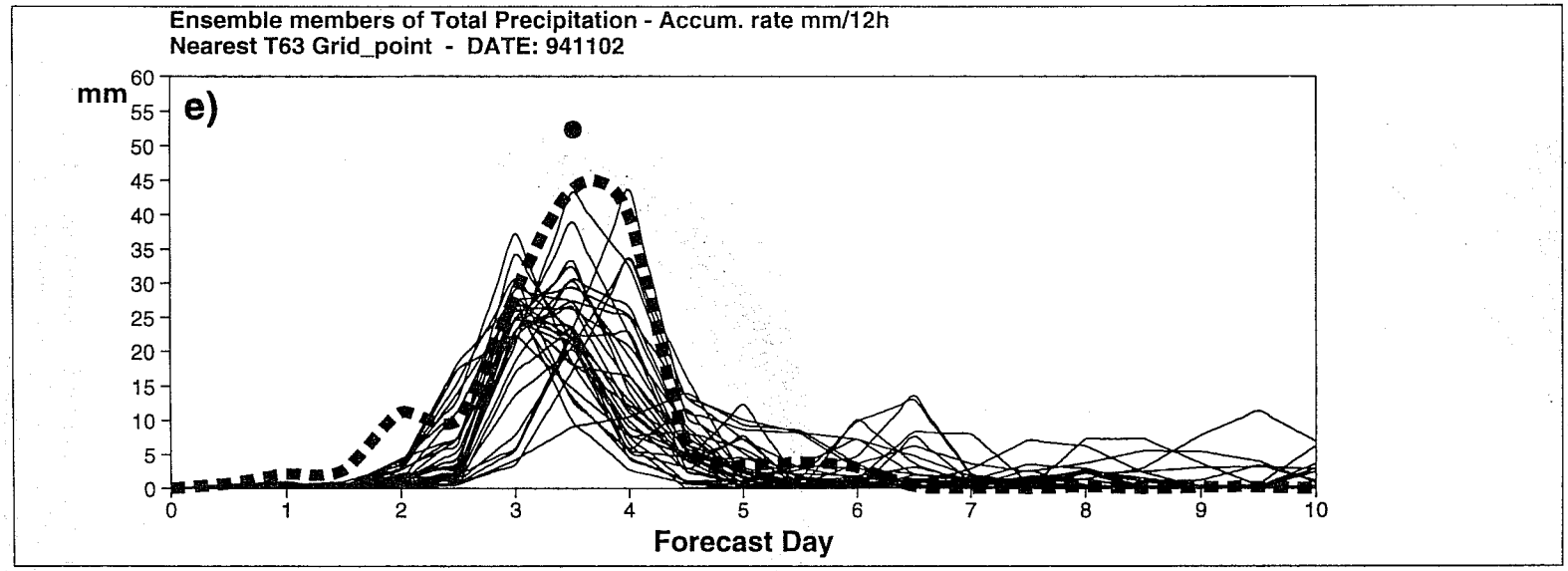


Fig 5 continued



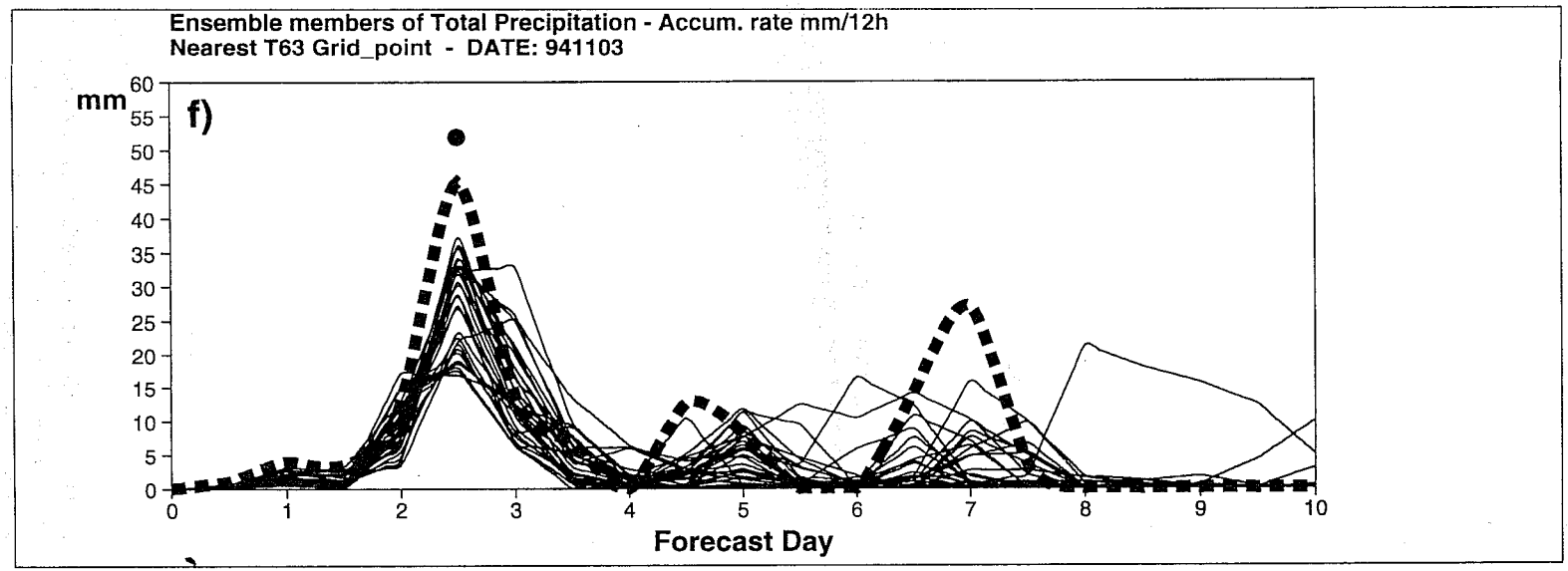


Fig 5 Continued

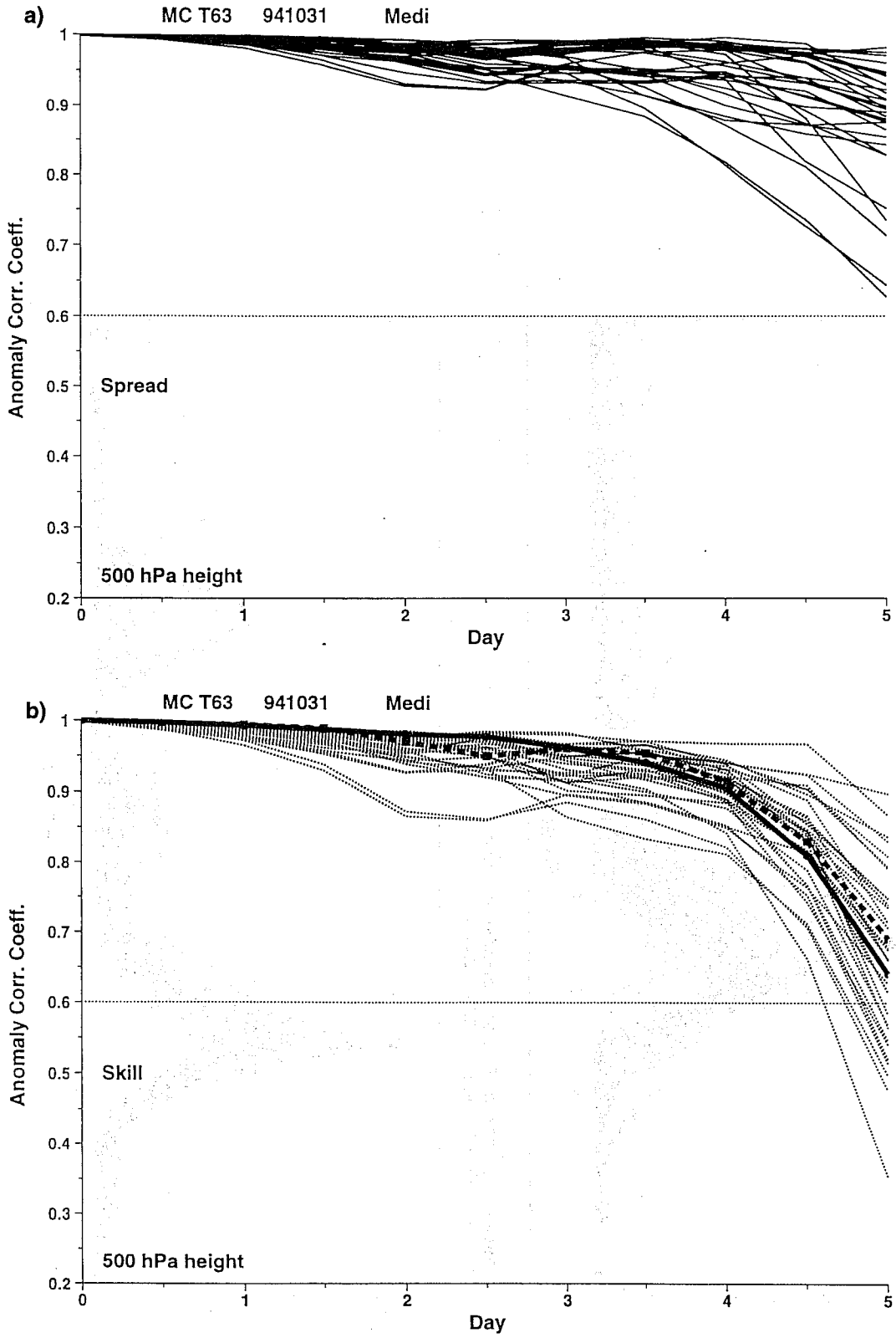


Fig 6 a) Spread of each perturbed ensemble member, computed as the distance between the perturbed and the control members in temms of anomaly correlation coefficient (ACC), over the GMR area for the ensemble forecast initiated on 94.10.31. b) Skill of the T213L31 (bold dash line), the control T63L19 (bold solid line), and each perturbed ensemble member, computed as the distance between the forecast and the verifying analysis, in terms of anomaly correlation coefficient, over the same region as a).

model.

5. PREDICTION OF THE HELLENIC STORM

As in the previous case (Section 2) we concentrate on the operational T213L31 and EPS predictions 5 days prior to the event. The large-scale flow prediction is verified from the day-5 forecast of 500 hPa geopotential height field valid for 21.10.94. The precipitation is verified from the observed accumulated precipitation between 12 UTC of the 21st and the 22nd of October 1994.

i) Synoptic flow

The T213L31 day-5 forecast (Fig. 7a) originating from 16.10.94 is considered to provide useful information to the forecaster with an anomaly correlation skill score of 78%. The day-5 cluster-1 centroid (Fig. 7b) consists of 12 members and it is also considered to contain useful synoptic elements with an anomaly correlation skill score of 73%. It is clear that the operational day-5 forecast is supported from the day-5 cluster-1 prognostic field centroid.

Figure 7c shows the day-5 cluster-5 centroid which scored best reaching an anomaly correlation skill score of 75%. The day-5 best-cluster is synoptically close to cluster-1 and possessing a lot of similarities with the T213L31 day-5 forecast. Cluster-5 consists of 4 members which in probability terms defines the probability of its occurrence at around 12%.

ii) Precipitation

A detailed picture of the total precipitation observed over the area of interest between 12 UTC of the 21st and 22nd can be found in *Lagouvardos et al,* (1995). For the same reasons as discussed for the Italian case we use the 24h precipitation forecast from the T63 control to highlight areas where precipitation exceeds 20 mm/day (Fig. 8a).

Figure 8b shows the day-5 T213L31 forecast originating from 17.10.94. The operational day-5 forecast positions the precipitation maxima correctly (over the centre of the Hellenic pensisula), reaching a maximum of 125 mm/day (a clear indication that an extreme event is

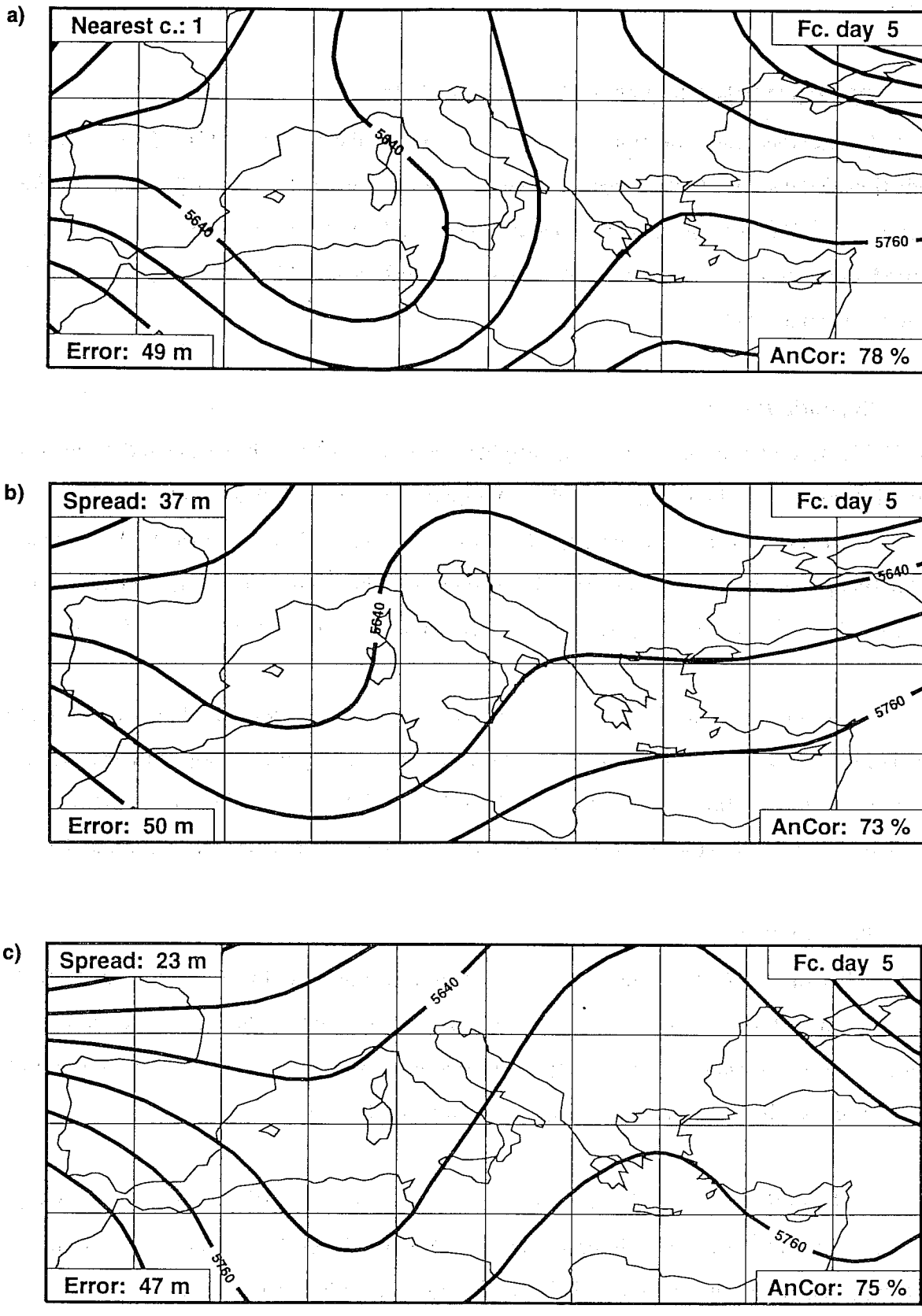
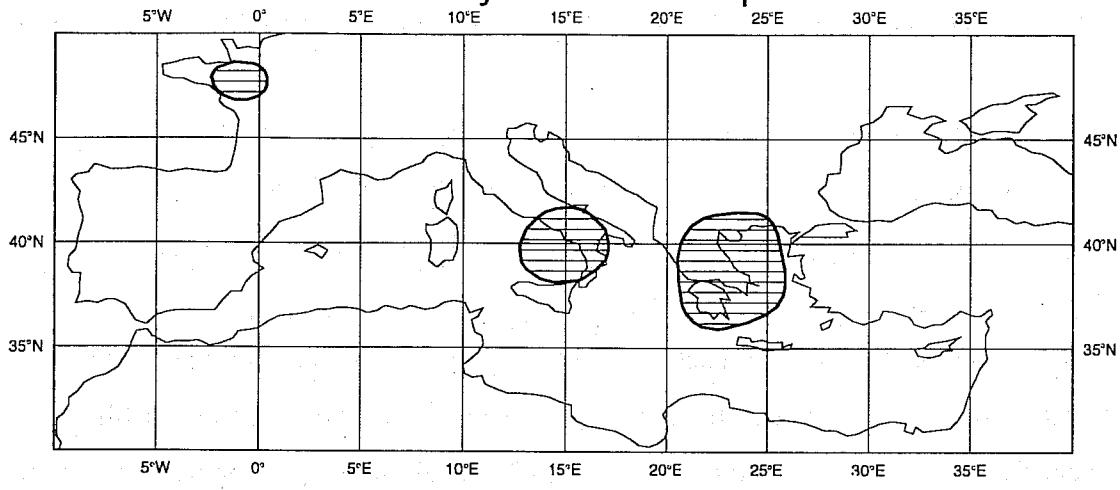
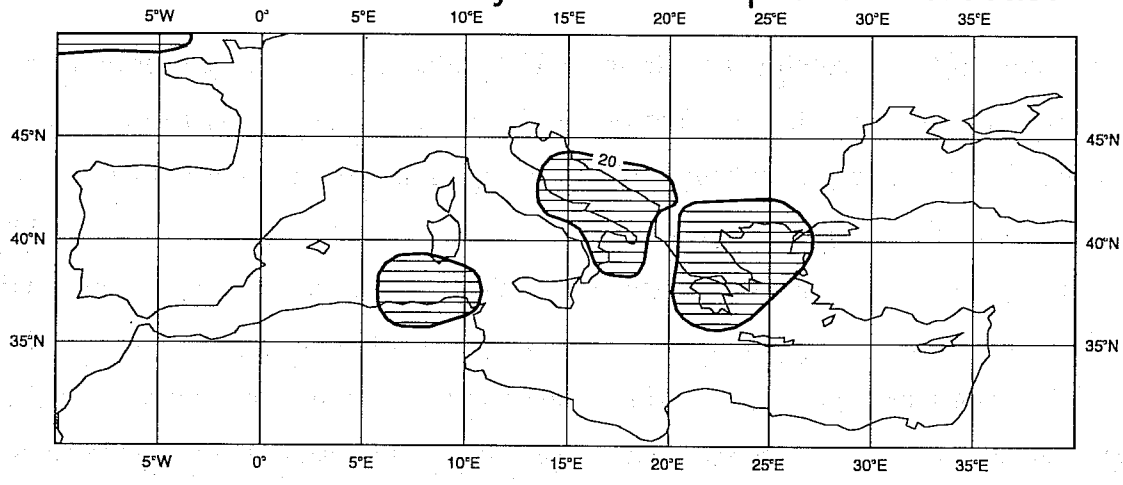


Fig 7 a) T213L31 day-5 forecast for 500 hPa geopotential height initiated on 94.10.16. b) as a) but for the EPS cluster-1 centroid. c) as a) but for the EPS cluster-5 (best-cluster) centroid. Contour interval: 60 m.

a) 94.10.21 - T63L19 Day-1 total Precipitation Forecast



b) 94.10.17 - T213L31 Day-5 total Precipitation Forecast



c) 94.10.17 - Probabilities for 'precipitation exceeding 20 mm/day'

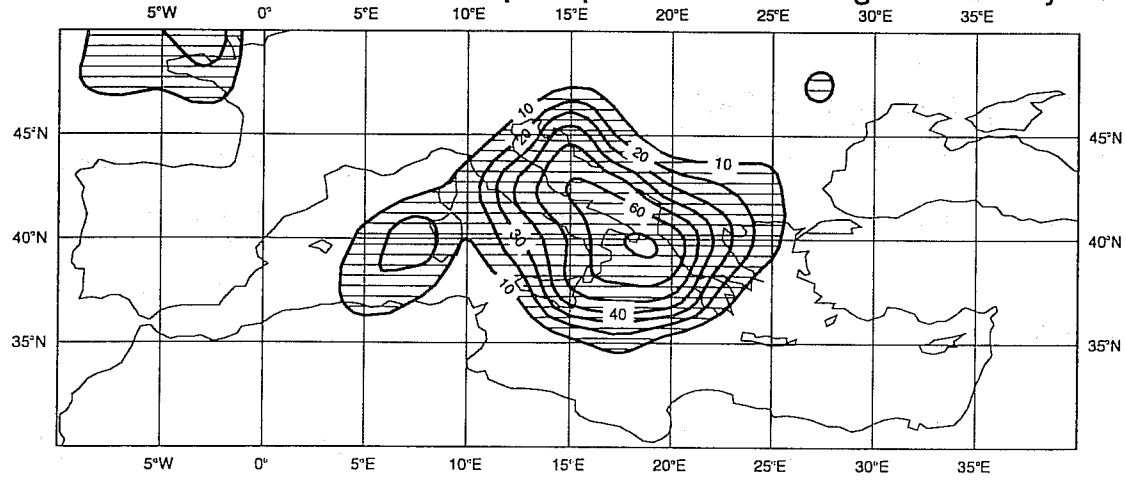


Fig 8 a) T63L19 precipitation prediction of the forecast initiated on 94.10.21, accumulated between day 0 and day 1 (areas over 20 mm/day are shaded). b) T213L31 precipitation prediction of the forecast initiated on 94.10.16, accumulated between day 4 and day 5 (areas over 20 mm/day are shaded). c) EPS probability values for the event 'total precipitation exceeding 20 mm/day' for the EPS started on 94.10.16 (preciptation accumulated between day 4 and day 5). Contour isolines every 10%.

likely to take place). There are suggestions of two other precipitation maxima the first over the southern Adriatic sea and the second over north Tunisia, but their intensity is much lower.

In Figure 8c the day-5 probability values for the event 'total precipitation exceeding 20 mm/day' to take place are shown, as predicted by the EPS initiated on 17.10.94 (i.e., constructed using the total precipitation ensemble member predictions between day 4 and day 5). The day-5 probability map reveals a high chance of an extreme event. Although the area of maximum probability (values reaching a value of 70%) is shifted over the northern Ionion sea (western sea area of Greece), the area of the main impact shows high probability values. It is clear that the EPS as it is expressed by the 2-D probability maps supports the operational precipitation forecast.

The evolution of the probability maps from the EPS initiated 8, 7, ..., 3 days prior the event can be seen in Figure 9. It is evident that a significant warning of an extreme event is becoming visible from day 6 (although shifted over the southern Italy - west Ionion sea) for the ensemble originated from 16.10.94 (Fig. 9c). In the previous two days low probability values are present over the wrong area (Figs. 9a-b). For the next three consecutive probability forecasts (forecast day-5, -4 and -3) the probabilities become consistent with the area of the main impact gradually shifting to the east and to a more correct position (Figs. 9d-f). It should be noted that the area of absolute maxima of probability is placed to the west of the main impact area. However, overall the EPS members support reasonably well the operational forecast.

6. THE FALSE-ALARM CASE

We try to assess the information provided from EPS for the false-alarm case comparing the day-5 T213L31 deterministic forecast with the corresponding precipitation probability maps. We also study the evolution of the probability maps for the period 8 to 3 days prior the false-alarm event.

In Figure 10a we use the 24h precipitation forecast from the T63 control run to highlight

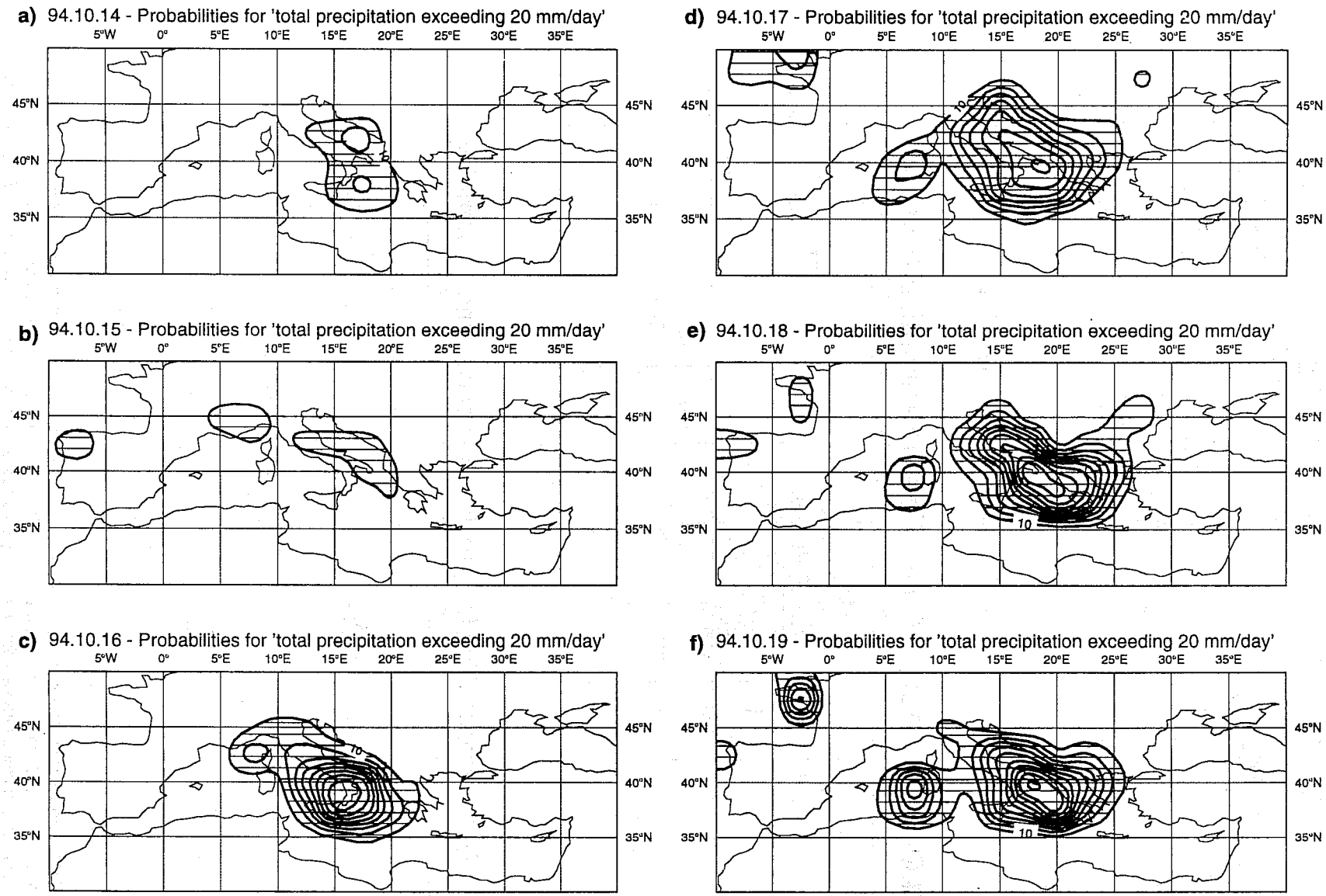


Fig 9 Probability prediction for the event 'total precipitation exceeding 20 mm/day' to occur between 12 UTC of 21and 22 October 1994, for the EPS initiated on a) 94.10.14 (precipitation accumulated between day 7 and day 8), b) 94.10.15 (precipitation accumulated between day 6 and day 7), c) 94.10.16 (precipitation accumulated between day 5 and day 6), d) 94.10.17 (as Fig 8c), e) 94.10.18 (precipitation accumulated between day 3 and day 4), and f) 94.10.19 (precipitation accumulated between day 2 and day 3). Contour isolines every 10%.

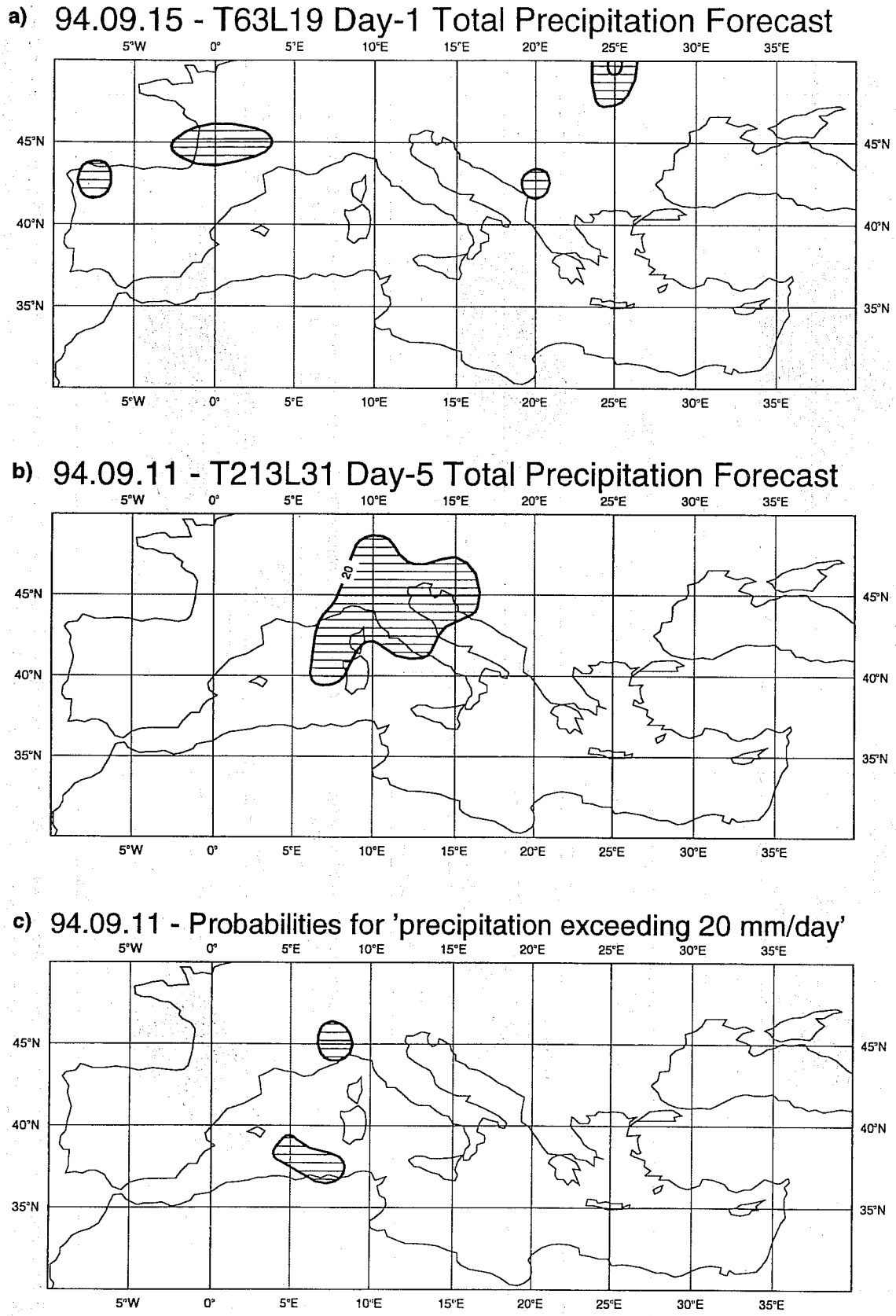


Fig 10 a) T63L19 precipitation prediction of the forecast initiated on 94.09.15, accumulated between day 0 and day 1 (areas over 20 mm/day are shaded). b) T213L31 precipitation prediction of the forecast initiated on 94.09.11, accumulated between day 4 and day 5 (areas over 20 mm/day are shaded). c) EPS probability values for the event 'total precipitation exceeding 20 mm/day' for the EPS started on 94.09.11 (precipitation accumulated between day 4 and day 5). Contour isolines every 5%.

areas where precipitation exceeded the 20 mm/day (in well agreement with the corresponding observations of total precipitation for the same area and period).

The 120 hour operational T213L31 precipitation prediction (total precipitation between day 4 and day 5) verifying for the time interval between 12 UTC of the 15th and the 16th of September, is shown in Figure 10b. The operational day-5 precipitation forecast originating from 12 UTC of 11.09.94 is considered to be poor. The operational precipitation forecast maxima (reaching a value of 65 mm/day) is wrongly positioned over northeastern Italy.

In Figure 10c the 120 hour probability values (two-dimensional probability maps) are shown for the event 'precipitation exceeding 20 mm/day' using precipitation fields from the EPS started on 11 September 1994 12 UTC. The day-5 probabilities do not support the operational T231L31 forecast at all. Contour interval of only 5% (half of the contour interval used in all other probability maps presented in this paper) had to be used to see at least one isoline (i.e. the isopleth of 5%).

The same interval (5%) is used in the next figure (Fig. 11) where the evolution of probability maps from the EPS initiated 8, 7, ..., 3 days prior the event is shown. All forecasts verify between the interval 12 UTC of the 15th and 16th of September 1994. It is evident that the EPS probability values are minimal over the area of interest (i.e. main impact area as suggested by the operational day-5 precipitation forecast) and they do not provide any support to the operational forecast throughout the whole interval.

7. RELIABILITY OF THE EPS PROBABILITY PREDICTION

i) Spread and Skill

There should be a clear relationship between ensemble skill and ensemble spread in a skilful ensemble system. When the spread of the ensemble is large then the forecast of any one member of the ensemble is unreliable. On the other hand when the spread is small then the forecast by any one member should be reliable.

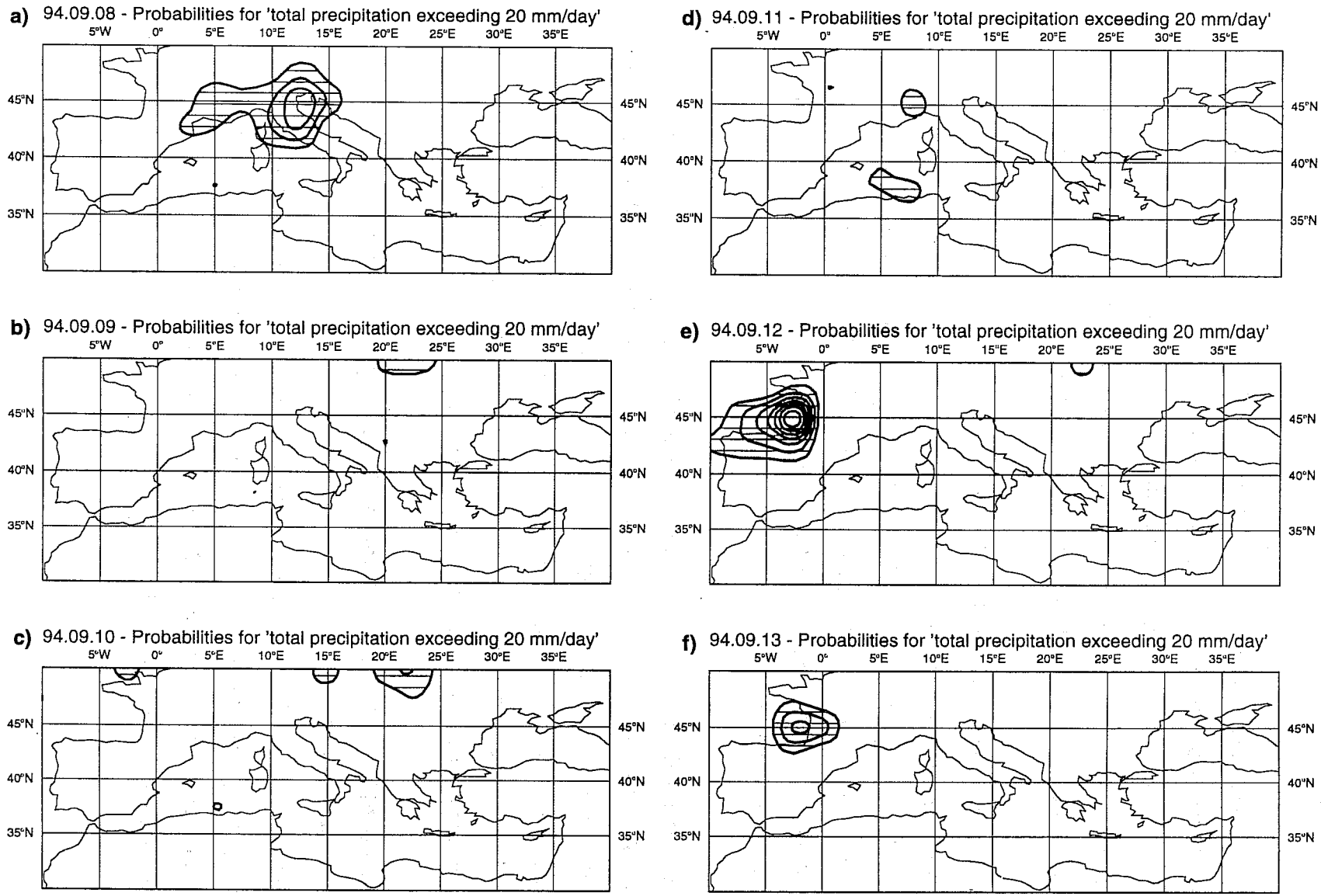


Fig 11 Probability prediction for the event 'total precipitation exceeding 20 mm/day' to occur between 12 UTC of 15 and 16 September 1994, for the EPS initiated on a) 94.09.08 (precipitation accumulated between day 7 and day 8), b) 94.09.09 (precipitation accumulated between day 6 and day 7), c) 94.09.10 (precipitation accumulated between day 5 and day 6), d) 94.09.11 (as Fig 10c), e) 94.09.12 (precipitation accumulated between day 3 and day 4), and f) 94.09.13 (precipitation accumulated between day 2 and day 3). Contour isolines every 5%.

Figures 12a and 12b show the seasonal (Autumn 1994) scatter diagrams of skill and spread for Europe (75.0 N - 30.0 N - 20.0 W - 45.0 E) for day 5 and day 7 respectively using RMS error and distance values. Here the ensemble spread is taken as the 75th percentile of the distribution of the RMS 500 hPa geopotential height difference (distance) between the perturbed members and the unperturbed control. The RMS 500 hPa geopotential height error (distance of the analysis) of the control is taken as the skill value. For each diagram the distribution is divided by the (seasonal) median value and the number of elements is shown in each quadrant corner. These numbers provide us with 2 x 2 contingency table for high/low spread, high/low skill. It is interesting to see that in both panels the diagonal entries are notably more populated than the off-diagonal entries. The off-diagonal entries are not negligible but we have to keep in mind that even in a perfect-model environment we should expect the off-diagonal elements to be non-zero (see *Murphy,* 1988, *Barker,* 1991, *Molteni et al.,* 1996, *Buizza,* 1995).

We have also estimated 'perfect-model' contingency tables by taking , at random, one member of each ensemble to be a verifying analysis and averaging the results over several possible realizations. The contingency tables for this 'perfect-model' is given below:

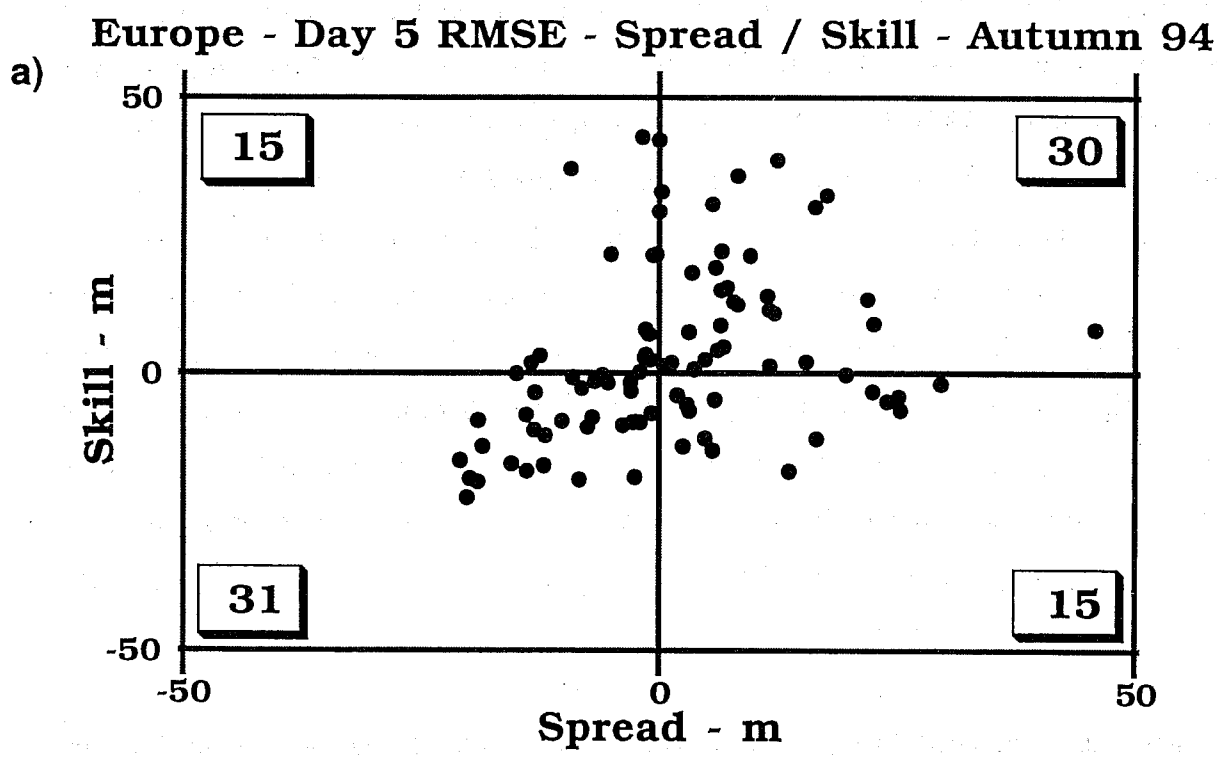
Europe - Day 5 RMSE - Spread / Skill - Autumn 1994 (total of 91 cases)

	Low Spread	High Spread
Low Skill	**13**	**32**
High Skill	**33**	**13**

Europe - Day 7 RMSE - Spread / Skill - Autumn 1994 (total of 91 cases)

	Low Spread	High Spread
Low Skill	**16**	**29**
High Skill	**30**	**16**

The comparison of the actual (observed) contingency tables (Fig. 14) with the 'perfect-model' simulations gives a better indication of the strength of the spread-skill relationship than any



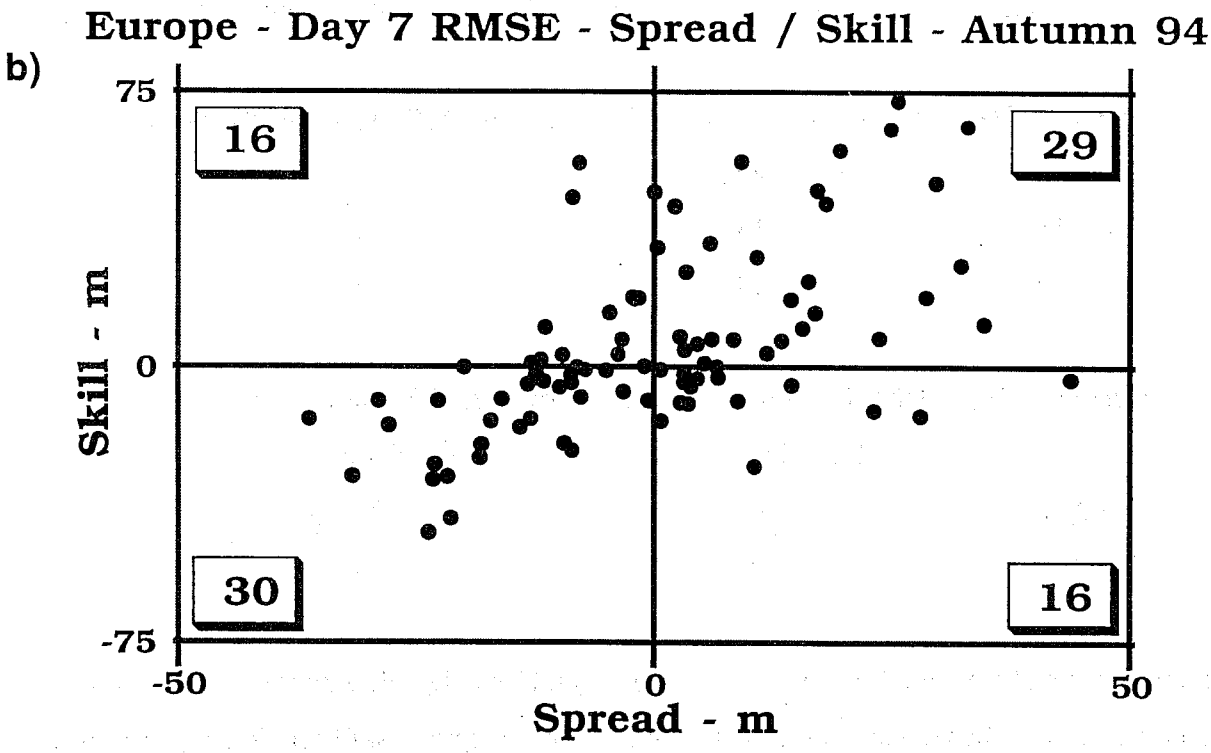


Fig 12 a) Scatter diagram between day-5 European RMS error of control forecast (ordinate) versus day-5 European RMS spread (abscissa) for Autumn 1994. b) as in a) but for day-7. Data are plotted as deviations from the median (seasonal) value.

other 'absolute' index. Overall, the actual contingency table values compare well with the 'perfect-model' estimates. In addition we made an effort to validate the statistical significance of the observed (contingency table) values using a chi-square test. For both the cases (Day 5 and Day 7 RMSE values) the H_0 hypothesis (the criteria are independent) was not accepted at a level of significance of 0.05.

Following this overall assessment for Autumn 1994 we present in Figure 13 10-day precipitation plumes initiating 5 days prior the events for locations (T63 grid-points) being close to the main impact area. For the Italian precipitation plume the grid-point close to the Piedmont area (already used, i.e. Fig. 5) is selected once more. For the Hellenic case a land grid-point near to the central Hellenic area (where the maxima of the observed precipitation have reported) is selected, whilst a grid-point close to the area of the maxima of the day-5 T213L31 forecast is selected for the false-alarm case. From panel a (Fig. 13) it becomes clear that almost all the members of the EPS support the operational maxima (which reaches a value of 45 mm/12h) a clear signal that the event was predictable. The underprediction of magnitude probably reflects the resolution of the EPS model.

In panel b of the same figure (Fig. 13) the majority of EPS members support the operational suggestion (although EPS values are noticeably lower than the operational maximum of 60 mm/12h). In addition there are two important points to notice. Firstly the timing is incorrect since the second operational maximum (precipitation between day 5 to day 6) does not verify; surprisingly the EPS makes an effort to cover all the possible operational suggestions (for this critical time interval) and this seems to be the reason behind its spreading between day 4 and 6. Secondly there is another maximum in the operational precipitation forecast valid for the period between day 6.5 and 7.5 which is wrong, and correctly not supported by the EPS.

From panel c (Fig. 13) it becomes evident that the EPS does not support the operational prediction of precipitation for the period between day 4 and day 5. There is also additional disagreement in the early medium-range where almost every member of the EPS forecasts higher amounts of rainfall than the high-resolution model precipitation forecast. Finally a

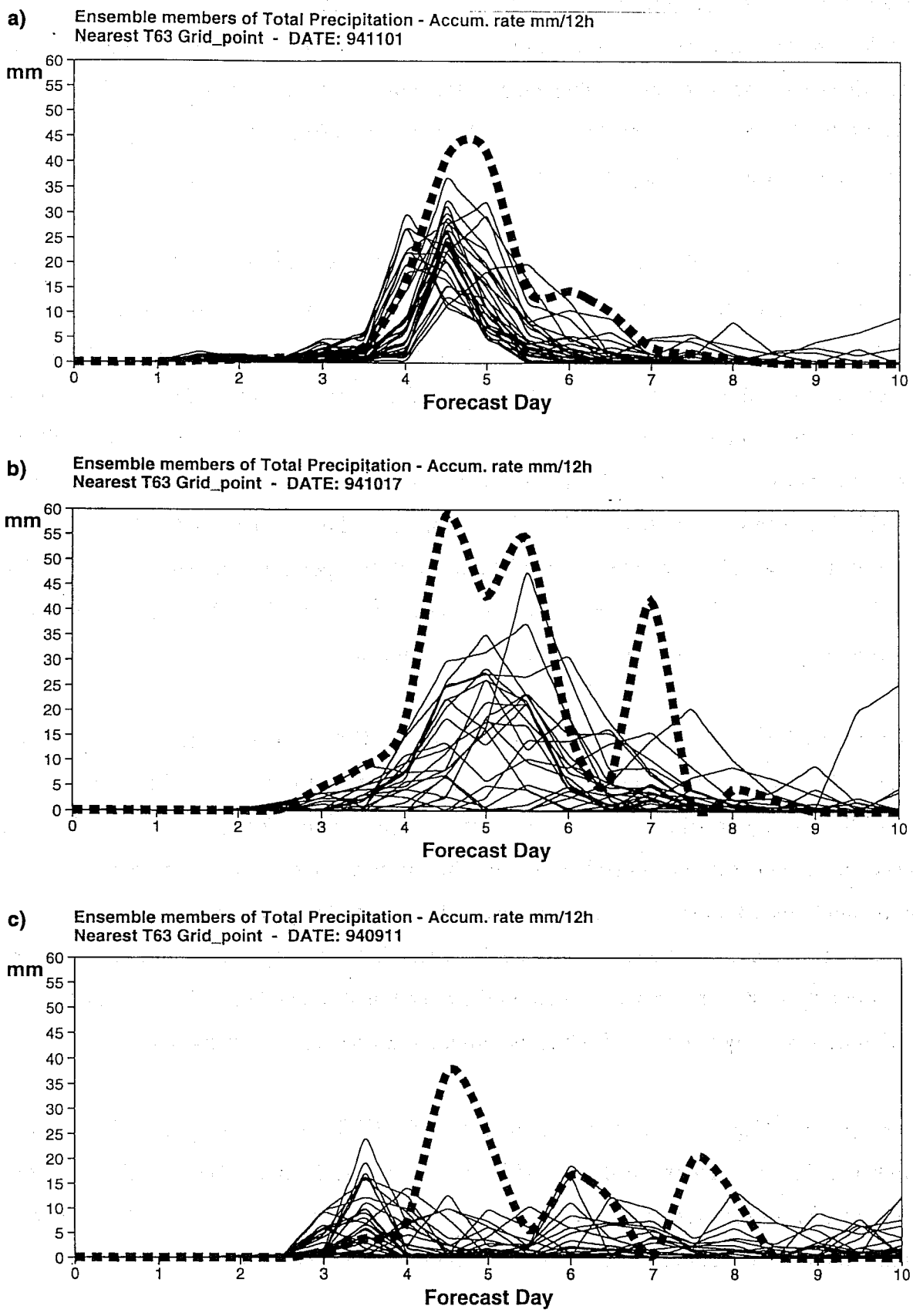


Fig 13 Precipitation plumes 5 days prior each event (for all study cases) for the nearest (to the main impact area) T63 land grid-point for a) Italian case (lat: 43.8 N - lon: 7.5 E), b) Hellenic case (lat: 40.1 N - lon: 22.5 E) and c) False Alarm case (lat: 43.8 N - lon: 11.3 E).

better agreement between the operational and the EPS precipitation forecasts seems to exist after day 5.5.

ii) Reliability of Weather Elements

There exists a variety of verification measures and tools for probabilistic forecasts. Descriptions, derivations and overviews can be found, e.g., *Epstein* (1967), *Glahn et. al* (1991), *Hsu and Murphy* (1986), *Murphy* (1988), *Murphy and Winkler* (1992), *Stanski et. al.* (1989), and *Wilks* (1995).

An important aspect of the EPS is to provide reliable probabilities of predefined events. This is assessed using the so-called reliability diagrams. The reliability diagram is constructed by dividing the range of forecast probabilities into discrete intervals, counting the occurrence of the event in each of these intervals and plotting the relative frequency of occurrence in each interval against the interval centre. For perfect reliability one should expect that all the points of this curve lie on the diagonal. Points below the diagonal indicate that probabilities are over-forecasted while points above the diagonal mean that probabilities are under-forecasted. Below we assess the reliability of a selection of the probability products.

For the European area the frequency of the event 'precipitation exceeding 20 mm/day' is found to occur too infrequently to permit construction of a meaningful reliability diagram. For this reason we select and present the seasonal (Autumn 1994) reliability diagrams for the following events:

a. 850 hPa Temperature anomaly below -4 degrees
b. 850 hPa Temperature anomaly above 4 degrees
c. Wind Speed exceeding 10 m/s
d. 24 hour Precipitation exceeding 10 mm

For all the events (panels a to d of Fig. 14), the area of verification is taken to be the European area represented by 144 grid-points defining 13,104 values for the period of the 91

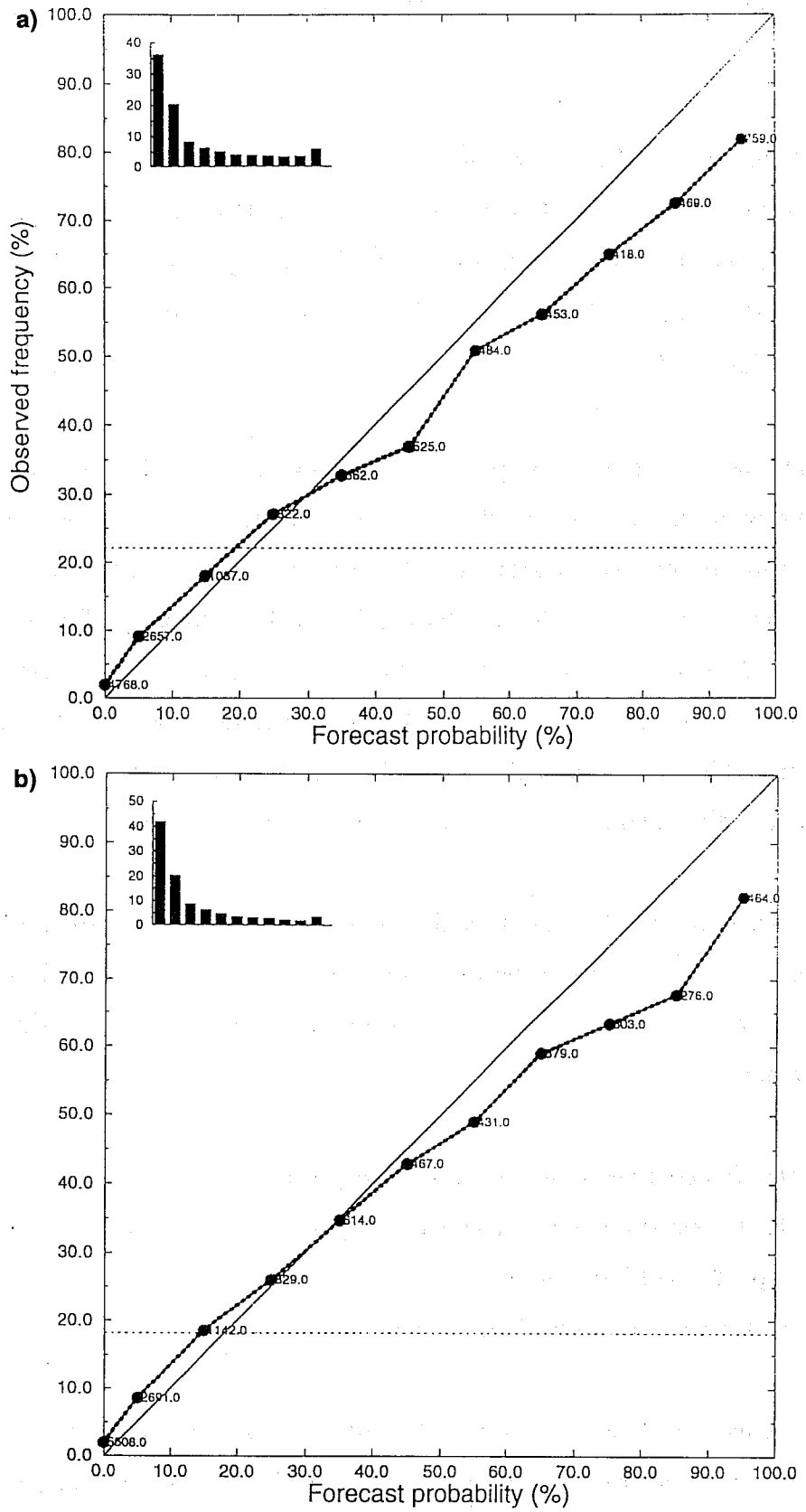


Fig 14 Seasonal (Autumn 1994) reliability diagrams of EPS day-5 probabilities for a) 850 hPa temperature anomaly colder than 4 degrees, b) as a) but for anomaly warmer than 4 degrees, c) wind speed (at z=10m) exceeding 10 m/s and d) 24 hour total precipitation exceeding 10 mm.

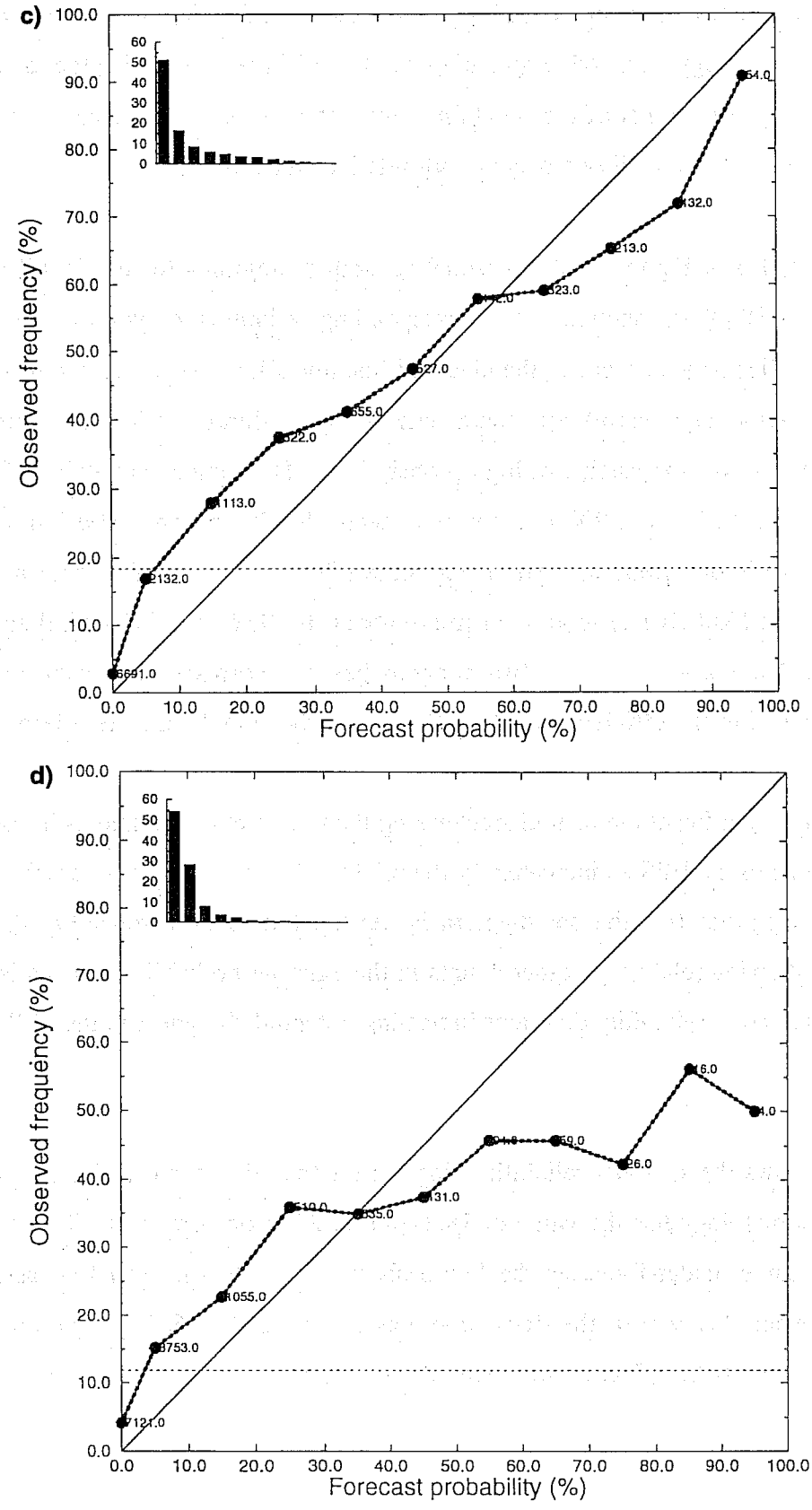


Fig 14 continued

days of Autumn 1994. The bold dash line represents the reliability curve while the thin dash the sample climatology. The 45 degree diagonal line (thin solid line) is plotted as well. For perfect reliability the points of the reliability curve should lie on the diagonal meaning that an event assigned a probability f occurs with relative frequency f.

In panels a and b of Figure 14 the seasonal reliability diagrams for the 'cold' and 'warm' anomalies (of T850) are shown. The corresponding reliability curves for the 'cold' and 'warm' anomalies lie very close to the diagonal line revealing considerable skill for the EPS. Both events (cold and warm) are characterized by a slightly under-forecasting of low-probabilities and over-forecasting of high-probabilities. The sample climatology for the 'cold anomalies' is 22.13% (i.e. 2900 occurrences), while for 'warm anomalies' has a value of 18.15% (i.e. 2378 occurrences). The Brier score (*Brier,* 1950) for the 'cold anomalies' is 0.11295 and the Skill (Brier) score 0.34 (more about the Brier and Brier Skill scores can be found in the Annex), while for the 'warm anomalies' the corresponding values are 0.10191 and 0.31 (related to the efficiency of the EPS to beat the sample climate reference).

One additional significant element characterizing the temperature anomalies is the sharpness (or refinement) of the EPS as indicated by the relative distribution of the sample sizes in the probability categories (i.e. the bar diagram in the top left of each reliability diagram). A distribution showing relatively higher counts in the extreme probability classes is evident in both the temperature reliability diagrams indicating potential sharpness in the EPS probability distribution.

Figure 14c shows the seasonal reliability diagram for the wind events. For the wind events the sample climatology has the value of 18.44% (i.e. 2417 occurrences). Results show that the EPS somehow under-forecasts the low-probability events and over-forecasts the high-probability events. However, the Brier score has a value of 0.11683 and the corresponding Skill (Brier) score 0.22 reflecting that overall the EPS probability forecasts wind events are skilful.

The seasonal reliability diagram for precipitation is shown in Figure 14d and is less skilful than for any other variable. For the precipitation events the sample climatology has the value of 11.85% (i.e. 1553 occurrences). As for the wind events the position of the reliability curve reveals problems of under-forecasting low-probability events. The problem of over-forecasting high-probability events is more pronounced than in the 'temperature anomalies' or the 'wind' events. All probabilities over 35% are over-forecast. In addition the reliability curve becomes more or less horizontal above 55% meaning that all forecasts over 55% are bound to achieve about the same rate (i.e. 45%). The Brier score has a value of 0.09984 and the corresponding Skill (Brier) score 0.04. It becomes clear that both of them are only marginally skilful.

Overall Brier and Skill (Brier) scores for temperature anomalies and wind speed show that the EPS has considerable usefulness. Problems with precipitation probabilities can be accounted for by the limited resolution of the T63L19 model.

iii) Reliability of EPS Synoptic Flow Clusters

Reliability diagrams for the clusters provided by the EPS are made by interpreting the cluster population as forecast probabilities for the circulation of the cluster centroid to occur. Thus the reliability of the clusters can be evaluated in a similar way to the reliability for other predefined events such as already discussed above.

Figure 15 shows the seasonal (Autumn 1994) reliability diagram for the EPS clusters. These clusters are produced for the GMR area following the operational approach, i.e. clustering forecast trajectories from day 5 to 7. For brevity we show the mean field (mean centroid) over day 5 to 7 (similar results are found when the reliability for day 5, 6, 7 is examined separately). The cluster reliability diagram is reasonably skilful. The anomalous 'zig-zag' structure is associated with large clusters (i.e. clusters with many members) having small sample size.

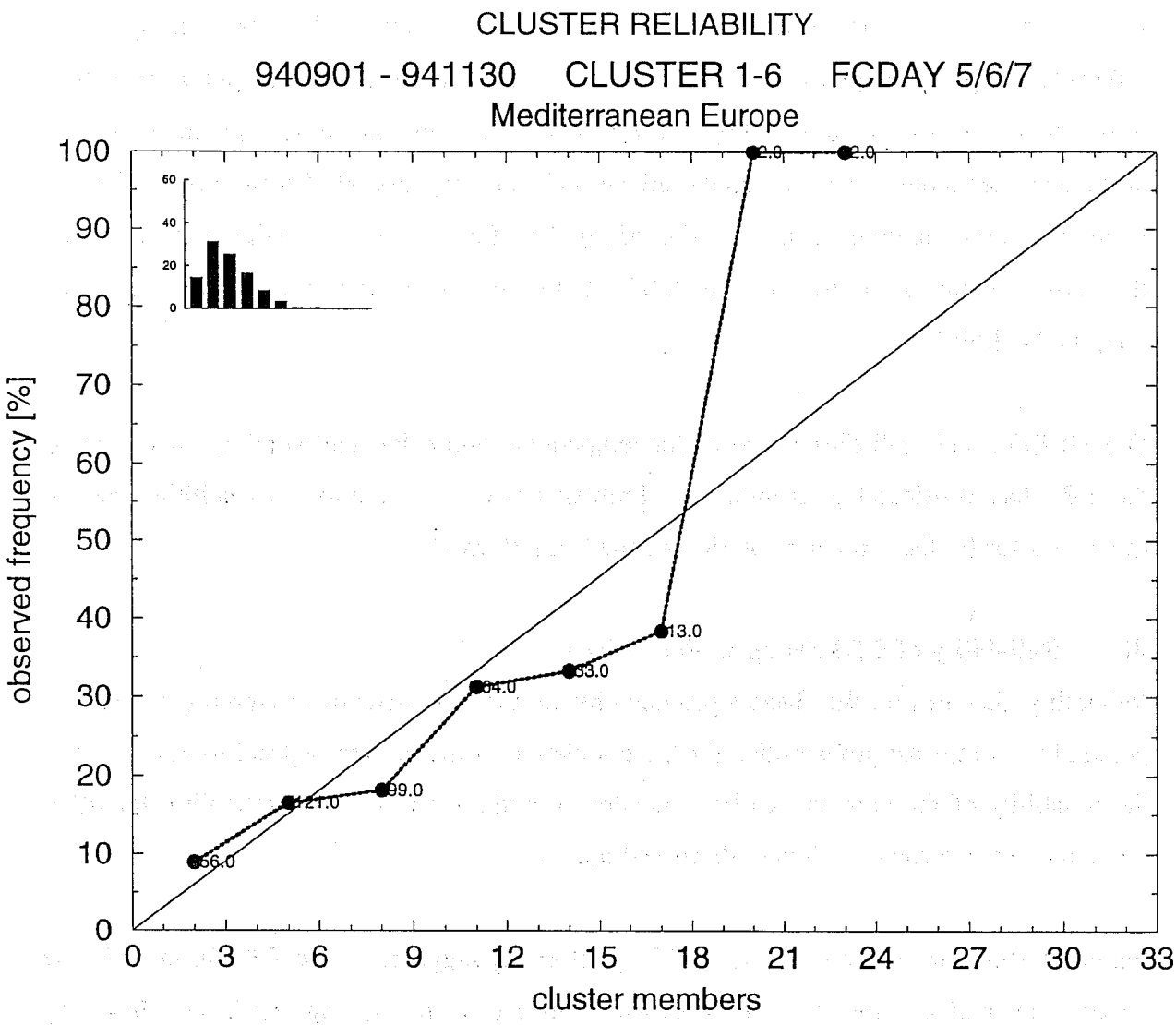


Fig 15 Seasonal reliability diagram of the EPS clusters (mean centroid fields) for Autumn 1994. Clusters were produced by clustering forecast trajectories and averaging from day 5 to day 7 over the GMR area.

8. CONCLUSIONS

In this paper, the ability of the ECMWF EPS to provide a measure of confidence in the high resolution operational forecast, has been studied. Three specific case studies were presented, corresponding to synoptic situations in which the high-resolution operational deterministic forecast predicted heavy rainfall in the medium range. In the first two cases, the EPS supported the high resolution model, and the deterministic forecast verified correctly. In the third case, the EPS did not support the high resolution model forecast, and this forecast did not verify correctly.

Some statistics of reliability of the EPS products were presented. These showed that probability forecasts of low-level temperature and wind were more reliable than those of precipitation. Nevertheless, probability forecasts of all three variables showed some level of skill. In addition, reliability statistics of clusters (of large-scale flow) showed considerable skill.

EPS forecasts of extreme weather events are necessarily compromised by the moderate resolution of the T63L19 model used to generate the ensembles. In future studies, ensembles will be made using at least T106L31 resolution. The benefits of such increases in model resolution will be compared with the benefits of increases in ensemble size. However, the ability of the higher resolution model to simulate heavy rainfall, strong winds, and extreme temperatures, suggests that high resolution will be necessary for the production of reliable probability forecasts of extreme weather.

ACKNOWLEDGEMENTS

The implementation of the Ensemble Prediction System has been the result of the work of many ECMWF staff and consultants in both the Operations and Research Department. Our appreciation goes to all of them. We also thank Pascal Mailier for his valuable help during the preparation of the reliability diagrams.

Annex: Skill Scores

(i) Root-mean-square (rms) error

The rms error E_j of the day j forecast is defined by

$$E_j = \sqrt{\overline{(f_j - \alpha)^2}},$$

where f_j is the day j forecast of the height field at a particular pressure, here taken to be 500 hPa, and α is the corresponding verifying analysis. The overbar denotes an average over area and over a set of forecasts. The rms error is initially small, and generally grows in time and reaching an asymptotic level.

(ii) Anomaly correlation coefficient - (ACC)

The anomaly correlation coefficient measures the correlation between forecast and analyzed deviations from climatology:

$$ACC_j = \frac{\overline{(f_j - c)(\alpha - c)} \;\; \overline{(f_j - c)(\alpha - c)}}{\sqrt{[\overline{(f_j - c)^2}\;\overline{(f_j - c)^2}]\;[\overline{(\alpha - c)^2}\;\overline{(\alpha - c)^2}]}}.$$

Here c is the climatological value of the height field for the verifying day.

Over the years an acceptable value of ACC characterizing a useful prognostic field has been established in both the communities of modellers and forecasters. This limit is set to be the 60% threshold. Prognostic fields scoring above this limit are considered as useful (skilful), while forecasts with ACC skill less than 60% are not considered to give any usefull synoptic

information.

(iii) Brier score - (BS)

The most common measure of accuracy of probabilistic forecasts is the Brier score (*Brier*, 1950), which can be written for the two category case (event occurs or doesn't occur), here in form of the half Brier score:

$$BS = \frac{1}{N} \sum_{i=1}^{N} (f_i - O_i)^2,$$

where f_i is the forecast probability of the event to occur and O_i equals 1 if the event occurred and 0 if it did not occurred. BS is simply the mean square error of the probabilistic forecast and thus is a measure for its accuracy. BS = 0 is a perfect score, BS = 1 is the worst possible. It should be noted that both BS = 0 and BS = 1 can only be achieved by a categorical (non-probabilistic) system.

A natural reference for the Brier score is the Brier score for climatology (i.e., a forecast always using the observed climatological frequency $\overline{O}$ as forecast probability).

$$BS_{cl} = \frac{1}{N} \sum_{i=1}^{N} (\overline{O} - O_i)^2,$$

or simply

$$BS_{cl} = (1 - \overline{O})\ \overline{O}^2 + \overline{O}\ (1 - \overline{O})^2 = \overline{O}\ (1 - \overline{O}).$$

In the middle expression of the last equation, the first term is for the cases when the event did not occur [relative frequency $(1 - \overline{o})$ and square error $(\overline{o} - o)^2]$, and the second term is for all occurrences of the event [relative frequency $\overline{o}$ and square error $(\overline{o} - 1)^2]$.

A better reference is based on long term climatology. If $\overline{o}$ is the sample climatology and μ is the long term climatology, the respective reference scores are related as follows:

$$BS_{lcl} = BS_{cl} + (\mu - \overline{o})^2,$$

the index "lcl" denoting long term climatology. It is immediately evident that the reference based on long term climate yields a higher, i.e. worse, value, given the sample and long term climatologies differ. This means that such a reference is easier to beat than the sample climate reference (which is unknown in advance).

(iv) Brier Skill Score - BSS

A Brier skill score can be defined as

$$BSS_{BS} = \frac{BS_{lcl} - BS}{BS_{lcl}}.$$

This skill score (BSS) is 1 for a perfect forecast (BS = 0), 0 for a probabilistic forecast which is no more accurate than a trivial forecast using climatology, and negative for even worse forecasts.

References

Anderberg, M R, 1973. Cluster analysis for applications. *Academic Press,* New York.

Athanassiadou M and A J Thorpe, 1995. The ANOMALIA project and MAP. *MAP Newsletter*, **3**, 48-50.

Barker, T W, 1991. The relationship between spread and forecast error in extended-range forecasts. *J. Climate,* **4**, 733-742.

Binder, P and A Rossa, 1995. The Piedmont flood: Operational prediction by the Swiss Model. *MAP Newsletter,* **2**, 12-16.

Brankovic, C, T N Palmer, F Molteni, S Tibaldi and U Cubasch, 1990. Extended-range predictions with ECMWF models: Time-lagged ensemble forecasting. *Q. J. R. Meteor. Soc.,* **116,** 867-912.

Brier G W, 1950. Verification of forecasts expressed in terms of probability. *Mon Wea. Rev.* **78**, 1-3.

Buizza, R, 1995. Potential Forecast Skill of ensemble prediction, and spred and skill distributions of the ECMWF Ensemble Prediction System. Proceedings of ECMWF Seminar on Predictability, ECMWF, Shinfield Park, Reading, RG2-9AX, UK, 4-8 September 1995, in press.

Buizza, R and T N Palmer, 1995. The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 9, 1647-1681.

Buzzi, A, C Cacciamani, T Paccagnella, P Patruno and N Tartaglione, 1995. Preliminary meteorological analysis of the Piedmont flood of November 1994. *MAP Newsletter*, **2**, 2-6.

Buzzi A and N Tartaglione, 1995. Meteorological modelling aspects of the Piedmont 1994 flood. *MAP Newsletter,* **3**, 27-28.

Courtier, P, C Freydier, J F Geleyn, F Rabier and M Rochas, 1991. The Arpege project at Meteo France. Proceedings of ECMWF Seminar on Numerical methods in atmospheric models, Shinfield Park, Reading RG2-9AX, UK, 9-13 September 1991, vol. 2, 193-231.

Dorninger M and M Hantel, 1995. Diagnosis of convection for the Piedmont flood of November 1994. *MAP Newsletter,* **3**, 40-42.

Epstein, E S, 1967. A scoring system of probability forecasts of ranked categories. *J. Appl. Meteor.,* **8**, 985-987.

Epstein, E S, 1969. Stochastic dynamic prediction. *Tellus*, **21**, 739-759.

Fleming, R J, 1971a. On stochastic dynamic prediction. I: the energetics of uncertainty and the question of closure. *Mon. Wea. Rev.*, **99**, 851-872.

Fleming, R J, 1971b. On stochastic dynamic prediction. II: predictability and utility. *Mon. Wea. Rev.*, **99**, 927.938.

Glahn H R, A H Murphy, L J Wilson and J S Jensenius Jr, 1991. Lectures presented at the WMO Training Workshop on the interprtation of NWP products in terms of local weather phenomena and their verification. *WMO PSMP* Rep. No **34**.

Gleeson, T A, 1970. Statistical-dynamical predictions. *J. Appl. Meteorol.*, **9**, 333-344.

Hsu W and A H Murphy, 1968. The attributes diagram. A geometrical framework for assessing the quality of probability forecasts. *International Journal of Forecasting*, **2**, 258-293.

Lagouvardos, K., V. Kotroni, S. Dobrivic, S. Nickovic and G. Kallos, 1995. The Storm of 21-22 October 1994 over Greece: Observations and Model Results. Submitted to *J. G. R.*

Leith, C E, 1974. Theoretical skill of Montecarlo forecasts. *Mon. Wea. Rev.*, **102**, 409-418.

Molteni F, R Buizza, T N Palmer and T Petroliagis, 1996. The ECMWF Ensemble Prediction System: methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73-119.

Murphy, J M, 1988. The impact of ensemble forecasts on predicatbility. *Q. J. R. Meteorol. Soc.*, **114**, 463-493.

Murphy A H and R L Winkler, 1992. Diagnostic verification of probability forecasts. *International Journal of Forecasting*, **7**, 435-455.

Palmer, T N, F Molteni, R Mureau, R Buizza, P Chapelet and J Tribbia, 1993. 'Ensemble Prediction'. ECMWF seminar proceedings 'Validation of models over Europe: Vol 1'. ECMWF, Shinfield Park, Reading, UK, 7-11 September 1992, vol. 1, 21-66

Simmons, A J, D M Burridge, M Jarraud, C Girard and W Wergen, 1989. The ECMWF medium-range prediction models: development of the numerical formulations and the impact of increased resolution. *Meteorol. Atmos. Phys.*, **40**, 28-60.

Stanski, H R, L J Wilson, W R Burrows, 1989. Survey of common verification methods in meteorology. Second edition, published as *WMO WWW* Rep. No **8**.

Wilks D S, 1995. Statistical Methods in Atmospheric Sciences. New York, *Academic Press,* 464 pp.