Forecasting Research

Forecasting Research Division
Technical Report No. 153

MEDIUM-RANGE ENSEMBLES USING BOTH THE ECMWF T63 AND UNIFIED MODELS - AN INITIAL REPORT

by

M.S.J. Harrison, D.S. Richardson, K. Robertson and A. Woodcock

March 1995

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

1. INTRODUCTION

One fundamental objective of the technique used at the European Centre for Medium-Range Weather Forecasts (ECMWF) for generating initial perturbations for medium-range ensembles is that the perturbations be concentrated around the most unstable errors in the analysis (in terms of error development over the first two days of the integrations) as discernable at forecast time. Additionally it is desirable that all unstable errors be tested within the ensemble as efficiently as possible, using the minimum number of members feasible, thus reducing computing costs. First ECMWF ensembles, constructed of 33 members, were generated with the T63 model but using a T21 adjoint. This selection was based on results from preliminary tests but also on computing limitations (Molteni *et al.*, 1994).

Experience with the early ensembles indicated that the ensemble spread was insufficiently large at medium ranges, an aspect commonly noted subjectively, commented upon by Member State forecasters (ECMWF, 1993, 1994), and confirmed by objective analyses. As an initial response to this problem the amplitude of the perturbations was increased by a factor of $\sqrt{2}$ on 23 August 1994 pending further investigation. Using adjoint techniques it has been possible to trace sources of errors back to their origins, work which has indicated that the initial error perturbations are frequently not resolved at T21. Tests with the computationally more expensive T42 adjoint have indicated that improved spread is produced within the ensemble and ECMWF are proposing to introduce T42 adjoints in the near future.

One potential problem with increasing the spread of an ensemble is that each increase tends closer to the background climatological spread, thus introducing a potential problem of decreasing information content with increasing spread. All models have a tendency to drift towards their own climates, with spreads limited in comparison to observed climatology, and thus there may be a limit to the maximum desirable spread within an ensemble for any given model in order to balance information content with cover of all possible future states. An alternative approach to generating an ensemble which partially lifts the restrictions associated with use of a single model is to build ensembles with multiple models, with the expectation that the limitations of the individual models might counterbalance themselves. One clear

justification for this approach is the difference in forecasts routinely produced by operational models, although there is little doubt that differences in analyses contribute in addition to differences in model formulations. But, in practice, combination of ensembles from two models increases predicted spreads by joining forecast distributions from the various models. Whether or not this approach modifies the information content of an ensemble remains to be determined.

Preliminary results are reported below from a set of two-model medium-range ensembles run during autumn 1994. All ensembles were generated from T21 adjoints with perturbation amplitudes increased by $\sqrt{2}$. Half of the ensemble was provided by the regular 33 members of the ECMWF T63 ensemble, the other half by a 33-member Unified Model ensemble at 1.67° by 2.5° resolution (this resolution is the standard, but infrequently-used, resolution one step above climate resolution). Unified Model ensembles were based on the ECMWF perturbations but added to the equivalent 1200Z Meteorological Office analysis. Joint ensembles have been run regularly since Saturday 29 October 1994 in this configuration, although the Unified Model is only run from Saturday and Sunday initial conditions and is not run operationally (and so, unlike the ECMWF ensemble, is not available to operational forecasters). Results are presented below from ten joint ensembles up to Sunday 27 November 1994. All diagnostics are for 500 hPa height fields verified against the UKMO 'Superfile' archive of 0000Z operational analyses, on which data are stored at 5° by 10° resolution from 15°N northwards. Most results refer only to a subset of 99 points distributed across the North Atlantic Ocean and Europe, an area bound at 30° to 70°N and 60°W to 40°E.

2. RESULTS

a. Anomaly correlations

Anomaly correlations over Europe and the North Atlantic from the two ensemble sets tend to be similar overall (Fig. 1), although there are occasions, such as for ensembles initialised on 6 November and 19 November, when one model is clearly the more skilful, at least for some of the forecast period, according to this measure. The most substantial difference

between ensembles is for those initialised on 26 and 27 November, in this particular case with the Unified Model the more skilful at longer ranges. Differences in the forecasts were related to the possible development of a block over north-western Europe; operational models were also divided on forecasts for the block. The majority of ECMWF members predicted maintenance of a block out to day 10 whereas relatively few Unified Models members did. In the event the ridge of high pressure displaced eastward, leaving Western Europe in a mobile westerly flow by day 7. These are two occasions on which anomaly correlations unequivocally demonstrate the value of the joint ensemble.

Also included in each diagram of Figure 1 is the anomaly correlation of the ensemble mean from each model's ensemble and that from the grand ensemble mean of 66 members. As expected theoretically, ensemble mean scores tend to be in the upper ranges of scores from individual members. Grand ensemble mean scores, while not necessarily always greater than those of the individual ensemble means, do tend to be closer to the mean of the more skilful model, suggesting that averaging joint ensembles tends to extract the skill of the more skilful, but unknown (at forecast time), model. In fact, averaging across the grand ensemble also increases predictability by the order of half a day over the Northern Hemisphere, if skill is defined in terms of anomaly correlations greater than 0.6 (Fig.2). There are, however, spatial variations in the skill between the two models over the 10 cases. Over North America, for example, the ECMWF model was the more skilful whereas in the European area the Unified Model was the more skilful at the longer ranges. Note that this latter result is biased by the ensembles of 26 and 27 November.

b. Local phase space

An important aspect of multi-model ensembles to be assessed is the distribution of variance across the ensembles, both within and between individual models. There are numerous manners in which this problem might be addressed, but only a single simple approach is discussed here. Eigenvector analysis has been applied to the North Atlantic/Europe 500 hPa fields of all 66 members at T+00, to create 'local' phase spaces for each case. Eigenvectors were calculated from correlation matrices and subjected to VARIMAX rotation. Most of the

variance, of the order of 90%, is captured by the first ten eigenvectors, but only the first four of the rotated eigenvectors have been studied in any detail to date.

At T+00 the first eigenvector is the most important as this fundamentally defines the direction between the basic analyses, provided these analyses differ significantly by comparison to the amplitudes of the perturbations. Comparison of the spatial fields of the first eigenvectors with fields of the analysis differences confirms, in all cases, that these eigenvectors are directed towards the analysis differences (Fig. 3). Lower eigenvectors capture the variance distribution across the perturbations themselves and, with commonality of this variance between the two models, have similar patterns for each model. Variance captured by the first eigenvectors for the ten cases varies between 22% and 44%, although it generally lies between 30% and 40%. Distributions of loadings on the first eigenvectors for a few selected cases, representative of all ensembles, are shown in Figure 4. In only one case, that of 29 October and the occasion with lowest variance on the first eigenvector, was there any, and then only marginal, overlap between distributions for the two models; otherwise distributions were entirely separate.

Apparently there is limited variance in the perturbations directed towards the analysis differences in all of these cases, despite which anomaly correlations were similar later in the forecasts for both models except in the last two cases. These results suggest *either* that the analysis differences do not define the most unstable growing errors in the initial conditions and therefore it might be correct that the perturbations are not pointed in the direction of the analysis differences *or* that the perturbations are not adequately exploring all unstable directions. Most of the cases appear supportive of the first of these hypotheses but the cases of 26 and 27 November illustrate forecasts in which different areas of phase space were explored by the models later in the integrations, a result which could only have originated in one or both of model and analysis differences.

No detailed analyses have been carried out so far on aspects of local phase spaces of the joint ensembles beyond T+00. Initial analyses of some preliminary joint ensembles, likely to be representative of these ten cases, indicated that the analysis differences could always be traced at least to T+48, and later in some cases. Also it was possible to trace development

of patterns of variance which resulted from differences between the models; normally these differences could only be recognised from several days into the integrations. It is expected that local phase space analyses will yield valuable information on the behaviour of multi-model ensembles.

c. Global phase space

One of the most important questions to be addressed in study of joint ensembles is whether the second model provides valuable useful additional information over the first. Although it has been demonstrated that the differences between analyses are not covered by the perturbations, only in two cases do these differences appear to translate into substantially different sets of predictions; otherwise anomaly correlations tend to be similar across the two models' ensembles. It must be remembered, however, that an infinite number of fields may produce a given anomaly correlation using given climatologies and observations and, therefore, that simple diagnostics, such as anomaly correlations, may not adequately elucidate details of phase space explorations.

A full investigation of the information content of the joint ensemble requires more than just 500 hPa fields, but some indicators can be obtained by projecting ensemble members into a predefined global phase space. With no adequate phase space available, a 'poor man's' phase space has been developed using 30 years of Superfile daily data across the North Atlantic and Europe. Eigenvectors have been calculated from correlation matrices of 5-day means (means over other periods produced minimal differences in final results) for November and December and VARIMAX rotation applied to define the phase space, model fields being projected onto these eigenvectors. Only the first ten eigenvectors have been retained, which account for about 90% of the variance. No details of the phase space will be provided here.

Only a few of the more interesting results are shown in Figure 5; many of the phase space directions may not be of importance at a given time in any prediction. Anomaly correlations for the ensembles initialised on 29 October were similar throughout the first nine days; only at D+10 was one ensemble set markedly more skilful than the other (Fig. 1). Both models performed similarly in the direction of the first eigenvector (E1), although neither adequately

simulated the large positive observed departures later in the forecast period. On E5, however, the Unified Model was the more successful in simulating changes in the real atmosphere, whereas the opposite was true on E6. The net effect of the different model behaviours along E5 and E6 was to approximately equalise anomaly correlations but with different solutions within the two ensembles. Further work is required to investigate whether such differences in phase space properties equate to important differences in information provided to forecasters, but similar differences in phase space explorations between the models are found in all 10 cases.

Phase space diagrams are included in Figure 5 for all of E1 to E10 for the 26 November case, one of the occasions on which the two models provided rather different predictions. Overall there are rather different patterns of phase space exploration between the two models in most, but not all, directions. By and large the Unified Model was the more successful, as would be expected from the anomaly correlations, but the ECMWF model provided better predictions on occasion in certain directions (e.g. D+5 to D+8 on E2). E10 provides a fine example of continuously-changing large displacements in phase space successfully captured by both models. It may be tentatively concluded that, despite the overall more skilful forecast from the Unified Model (in anomaly correlation terms), there was probably beneficial information in the joint ensemble even in this case.

In conclusion it has been demonstrated that the two models may explore different areas of phase space even when simple one-dimensional indicators suggest that the models are performing similarly. Whether these differences result from dissimilar model climatologies, dissimilar abilities of the models to explore different regions of phase space or simply from dissimilar analyses remains to be determined. Equally, differences in information content to forecasters cannot be determined from these data.

d. 'Talagrand' diagrams

A new method of examining the spread of an ensemble has been suggested by Olivier Talagrand. The distribution of predictions at a given time define regions with equal probability of occurrence, assuming a perfect model. In other words, the probability that the

observation will lie on either side of the ensemble is the same as that it will lie between any two adjacent predictions within the ensemble. Hence, if the number of occurrences of the observation lying between any two forecasts are counted and then a distribution drawn, this distribution should be flat for a correctly-formulated ensemble.

Talagrand diagrams for 500 hPa heights over the North Atlantic and Europe at D+7 for individual model ensembles (Fig. 6) are similar in that the observation falls too frequently outside the ensemble. In terms of this measure there is probably little to choose between the two models, except, perhaps, for the bias in the Unified Model for the observation to lie on the 'too high' rather than the 'too low' side of the ensemble, a fact consistent with the cold bias of the model. Joining the two ensembles together does not entirely eliminate the higher-than-expected frequencies with which observations lie outside the joint distribution. However, for both individual models about 25% of observations lie outside the ensemble, a figure reduced to about 11% for the joint ensemble, with the expected value reduced from about 6% to 3%.

Use of two models does not eliminate the problems of inadequate ensemble spread but does appear to improve the situation. It should be noted, however, that in its current form the Talagrand diagram represents a measure that can be 'played'. A climatological distribution, or even a random distribution based on climatology, will produce, in time, a flat Talagrand distribution, but one with no information content. Improved methods of developing these diagrams are possible by producing diagrams for specific phase space projections or for limited cases (such as only observations above normal), but further work is necessary to achieve these ends. In the meantime current results are, at the least, consistent with the hypothesis that use of two models improves forecast spread.

e. Reliability; Brier scores

If ensembles are to be used to provide probabilistic predictions then one of the more important diagnostic tests of the forecasts is the reliability, a measure of the ability of any forecast system to provide probabilities in the correct context (i.e. an event should happen

on x% of those occasions where it is predicted with x% probability). Reliability diagrams have been prepared for forecasts over the North Atlantic and Europe of both positive and negative 500 hPa departures in excess of 10 dam (Fig. 7). This does not provide an ideal measure of reliability as the climatological probability of departures of this magnitude varies across the region. However no consistent climatology of 500 hPa height variances is currently available and so use of a fixed anomaly is all that can be attempted at present.

Predictions of both positive and negative anomalies tend to be overconfident for both models and for the joint ensemble, at least for the higher probability categories. It appears reasonable to state that, for these 10 cases, both the Unified Model and the joint ensembles were more reliable in predicting positive anomalies than was the ECMWF model; differences are less clear for negative anomalies. Brier scores have also been calculated for both positive and negative anomaly predictions (greater skill is indicated by lower Brier scores - unfortunately climatological Brier scores cannot be estimated at present). In both cases lower scores are provided by the Unified Model rather than the ECMWF model, but lower scores again by the joint ensemble (Table 1). Reliability plots and Brier scores have also been calculated for the first 8 cases in order to eliminate any possible bias from the final two cases. Overall results were essentially similar, although Brier scores tended to be uniformly somewhat higher. Within the limitations of this analysis, then, there are clear benefits for probabilistic predictions from joint ensembles over those from ensembles from either individual model.

	ECMWF	UM	Joint ensemble
>10 dam above normal	0.148	0.135	0.127
>10 dam below normal	0.172	0.151	0.145

Table 1. Brier scores over all 10 cases for prediction of positive and negative 500 hPa anomalies of greater than 10 dam.

f. Relative operating characteristics

The relative operative characteristic curve originated in signal detection theory and is finding increasing use in meteorology (Stanski *et al.*, 1989, Swets and Pickett, 1982). Details of the relative operating characteristic (ROC) curve are provided in Annex A. Briefly, the curves indicate the performance of a system in predicting a particular event in terms of the hit and false alarm rates (stratified by the observations). Each point on the curves is located by the total number of hits and false alarms achieved with probabilities at or greater than a specific value. Curves lie closer to the upper left-hand corner for the more skilful systems.

As for reliability, ROC's have been calculated across the North Atlantic and Europe at D+7 for both positive and negative anomalies in excess of 10 dam (Fig. 8). Results are consistent with the Brier scores in that, overall, best performances were achieved by the joint ensemble, although the Unified Model ensembles do not have substantially reduced skill compared to the joint ensemble. Results have also been calculated for the first 8 cases, thus removing any bias that might occur in favour of the Unified Model ensembles from the final two sets. Overall results (not illustrated) were essentially similar.

3. DISCUSSION

Any results based on a limited set of experiments suffer from possible problems of restricted representativeness, but it is apparent that all of the results presented above are consistent with the concept that joint model ensembles provide improved information over single model ensembles. In general, results for the joint ensemble tend to lie close to, although on occasions perhaps with lower skill than, those for the more skilful of the single model ensembles (of course the most skilful individual model varies from case to case and cannot be determined *a priori*). Thus, at the least, running joint ensembles might provide a filter towards selecting the more skilful options from the two models. But there do appear to be additional benefits of the joint ensemble in terms of phase space exploration, in terms of spread across the ensemble and in terms of providing improved probabilistic information. These benefits appear to extend well beyond the two cases when the individual models are

well split in terms of their predictions to the eight remaining cases when simple one-dimensional analyses suggest the models are performing similarly.

Several fundamental questions have been posed in relation to the efficacy of joint ensembles. No complete answers to any of these questions may be supplied on the basis of the current results, but preliminary responses may be attempted. The questions, with responses based on the above results, are:

1. **Is it beneficial in practice to use different analyses and models in an ensemble?** The evidence presented above suggests that this is the case. However the results are based on a limited set of analyses and on only a single field. Practical use of ensembles will be based on other fields, such as temperature and rainfall fields, and a complete response to this question will require analyses of these;

2. **Can the perturbations used in these ensembles capture the key differences between analyses?** In none of the current cases was there extensive variance in the perturbations in the direction of the analysis differences, indicating that current singular vectors do not capture these differences. The situation will need re-evaluation once ECMWF transfers to T42 adjoints. Of course it may be argued, with validity, that the analysis differences may not be critical in terms of the fastest growing errors in the models, but the end of November cases, in which the two models diverged, suggest otherwise. However, differences in these two cases might be critically dependent on model formulation rather than analysis differences;

3. **Do model or analysis differences explain the variability between forecast centres?** No response to this query can be given on the basis of the above results. 'Hybrid' experiments are being planned in which the Unified Model will be run from ECMWF analyses and *vice versa*. Further, the Unified and ECMWF models have certain critical low-frequency performance differences, at least for operational resolution model versions; for example, the ECMWF model has an improved blocking climatology compared to the Unified Model (Van der Wal, 1995). The extent to which these differences are applicable to the models at the resolutions used here remains to be determined, but the phase space analyses suggest there may be differences in low-frequency variability performances of the two models.

4. CONCLUSIONS

Based on 10 cases of joint ensembles using a) the ECMWF T63 model initialised off the ECMWF analysis together with 32 singular vector perturbations derived from a T21 adjoint and b) the Unified Model at reduced resolution initialised off the Meteorological Office analysis together with the same ECMWF perturbations, it has been demonstrated using 500 hPa fields across the North Atlantic and Europe that:

a) in 8 cases anomaly correlations tended to be similar across both individual ensembles, although either model may have been the more skilful at a particular time;

b) in 2 cases anomaly correlations for one model were substantial higher than for the other, indicating rather different sets of solutions between the two ensembles;

c) anomaly correlations of the ensemble mean for the joint ensemble were higher than for means for individual ensembles, and indicated a gain in predictability of the order of half a day over the individual ensemble means;

d) the perturbations are not, in general, directed towards the differences between the analyses and do not span these differences regardless of whether the models produce similar or diverging predictions;

e) phase space exploration can be different between the models, even when anomaly correlations are similar;

f) the joint ensemble provides an apparently improved spread of future states in comparison to either individual ensemble;

g) there was an improvement in the reliability of the joint ensemble in predicting positive anomalies in excess of 10 dam, but the differences were not so clear for equivalent negative anomalies;

h) Brier scores (unreliable because of the small number of cases) were lowest (indicating higher skill) for the joint ensemble for both positive and negative 10 dam anomalies:

i) relative operating characteristics of the joint ensemble were superior to those for either individual ensemble for both positive and negative 10 dam anomalies.

ANNEX A. RELATIVE OPERATING CHARACTERISTICS

Relative operating characteristics, derived from signal detection theory, are intended to provide information of the characteristics of systems upon which management decisions can be taken. In the case of weather forecasts, the decision might relate to the most appropriate manner in which to use a forecast system for a given purpose. ROC's are useful in contrasting characteristics of deterministic and probabilistic systems but here they are used only for comparison of probabilistic systems.

Take the following 2x2 contingency table for any yes/no forecast:

		Forecast		
		Yes	No	
Observed	Yes	Hits (H)	Misses (M)	H + M
	No	False alarms (FA)	Correct rejections (CR)	FA + CR
		H + FA	M + CR	

Using stratification by observed (rather than by forecast) the following can be defined:

$$\text{Hit Rate} = H/(H + M)$$

$$\text{False Alarm Rate} = FA/(FA + CR)$$

A probabilistic forecast can be converted into a 2x2 table as follows. Tabulate probabilities in, say, 10% ranges stratified against observations, i.e.:

Probability Range	Number of Observed events for each probability range	Number of Non-Observed events for each probability range
90-100%	O_{10}	NO_{10}
80-90%	O_9	NO_9
70-80%	O_8	NO_8
60-70%	O_7	NO_7
50-60%	O_6	NO_6
40-50%	O_5	NO_5
30-40%	O_4	NO_4
20-30%	O_3	NO_3
10-20%	O_2	NO_2
0-10%	O_1	NO_1
Total	ΣO_i	ΣNO_i

For any threshold, such as 50%, the Hit Rate (False Alarm Rate) can be calculated by the sum of O's (NO's) at and above the threshold value divided by ΣO_i (ΣNO_i). So for the above case

$$\text{Hit Rate} = (O_{10} + O_9 + O_8 + O_7 + O_6) / \Sigma O_i$$

$$\text{False Alarm Rate} = (NO_{10} + NO_9 + NO_8 + NO_7 + NO_6) / \Sigma NO_i$$

This calculation can be repeated at each threshold and the points plotted to produce the ROC curve, which, by definition, must pass through the points (0,0) and (100,100). The further the curve lies towards the upper left-hand corner the better; no-skill forecasts are indicated by a diagonal line. It is possible to provide further statistics based on the ROC, such as the area under the curve (which can be used to contrast two or more curves), but this has not been done in this paper.

REFERENCES

ECMWF, 1993: Report on expert meeting on ensemble prediction system, 6-7 July 1993, ECMWF, Reading, UK. 133pp.

ECMWF, 1994: Report on expert meeting on ensemble prediction system, 27-28 June 1994, ECMWF, Reading, UK. 53pp.

Molteni, F, R Buizza, TN Palmer, T Petroliagis, 1994: The ECMWF ensemble prediction system: methodology and validation. Submitted to *Quart. J. Royal Meteorol Soc.* (see also Proceedings of Fourth Workshop on Meteorological Operational Systems, 22-26 November 1993, ECMWF, Reading, UK.)

Stanski, HR, LJ Wilson, R Burrows, 1989: Survey of common verification methods in Meteorology. *WWW Technical report no. 8, WMO/TD 358*, 114pp.

Swets, JA and RM Pickett, 1982: Evaluation of diagnostic systems - methods from signal detection theory. Academic Press, 253pp.

Van der Wal, A: Blocking in the global Unified Model - it's characteristics and predictability. Forecasting Research Technical Report No. 152. Meteorological Office, Bracknell, UK.

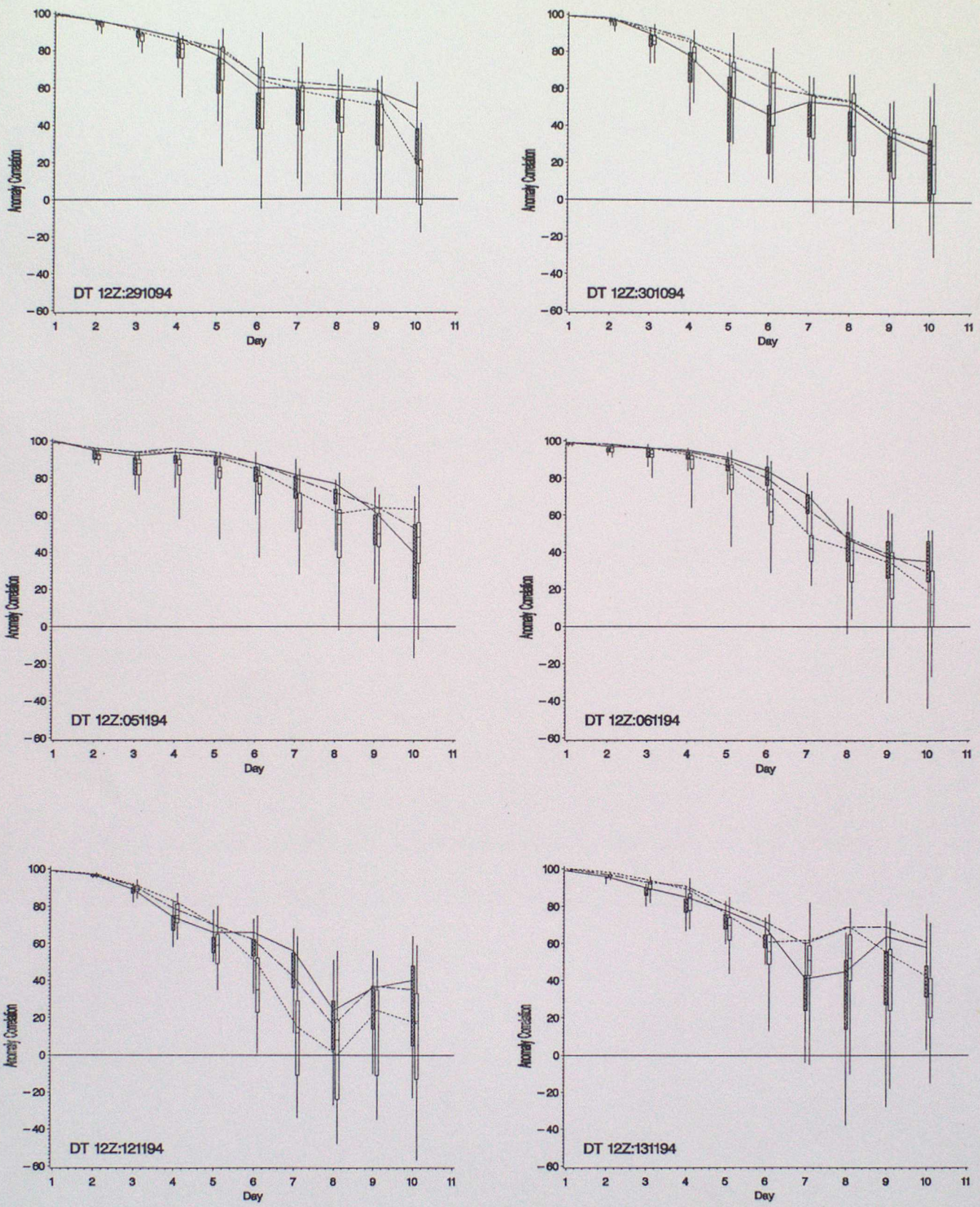
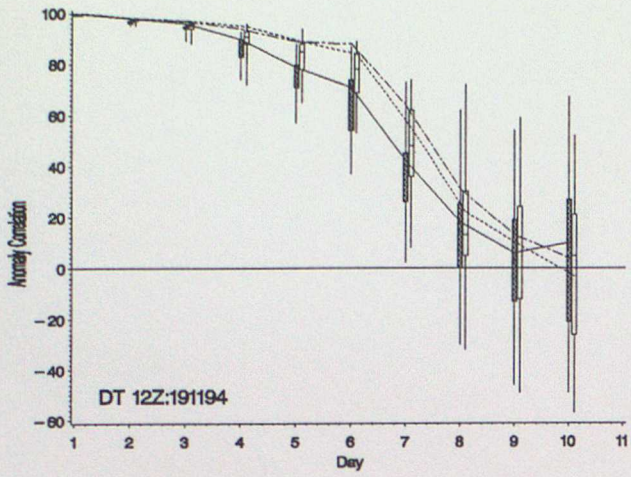
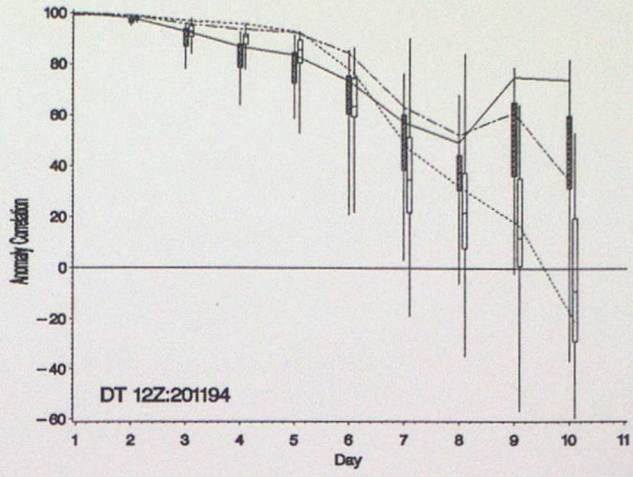
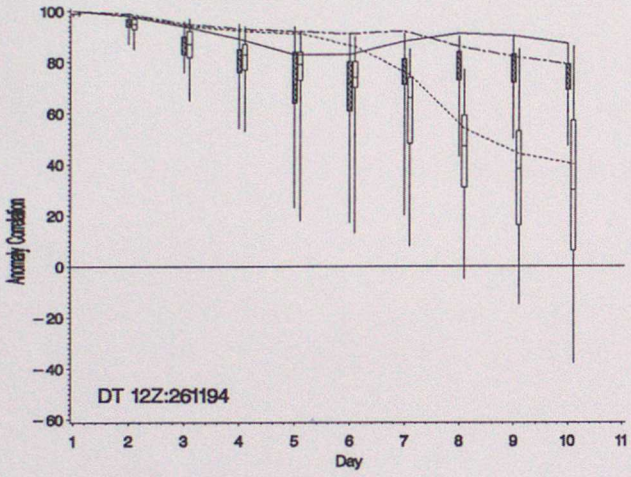


Figure 1. 500 hPa anomaly correlations for the ten cases over the North Atlantic and Europe. Boxes show median, 25th and 75th percentile and range of scores for UM (filled boxes) and ECMWF ensembles. Lines show scores for UM (solid), ECMWF (dotted) and grand (dash-dot) ensemble means.







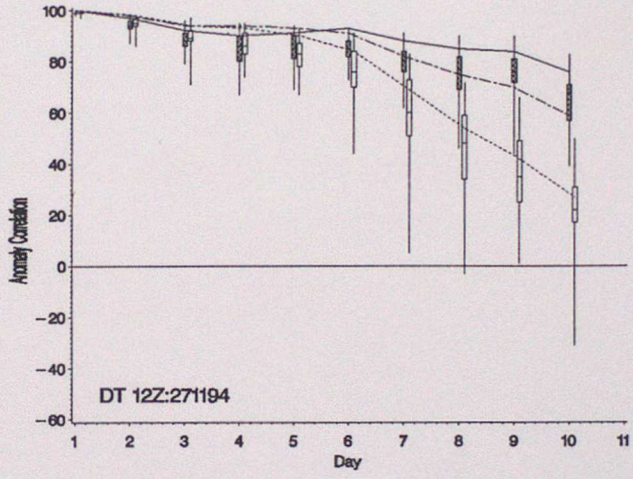


Figure 1. (continued)

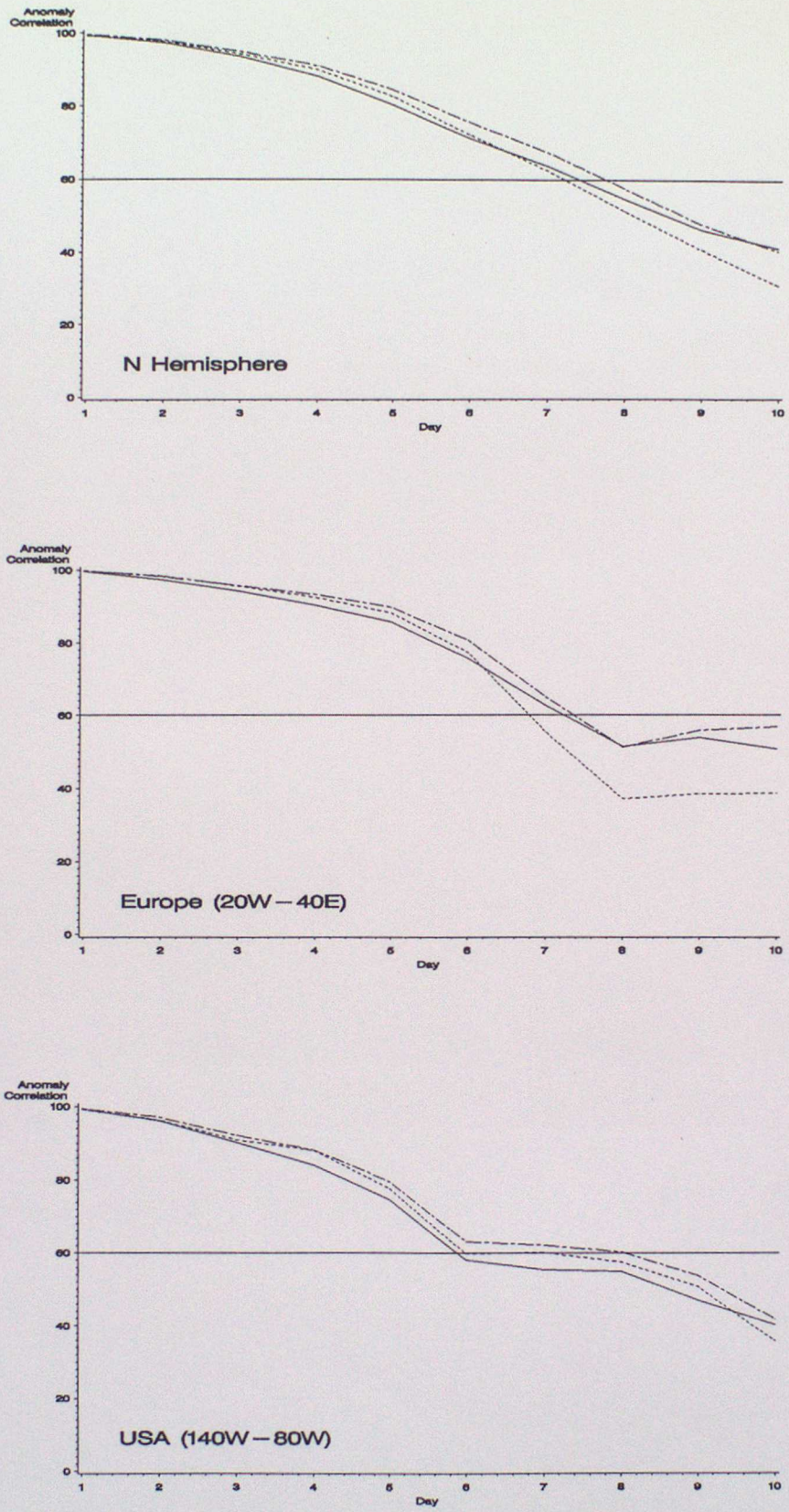


Figure 2. Average 500 hPa ensemble mean anomaly correlations over all ten cases for the Northern Hemisphere, Europe and North America; UM (solid line), ECMWF (dotted) and grand (dash-dot) ensemble mean.

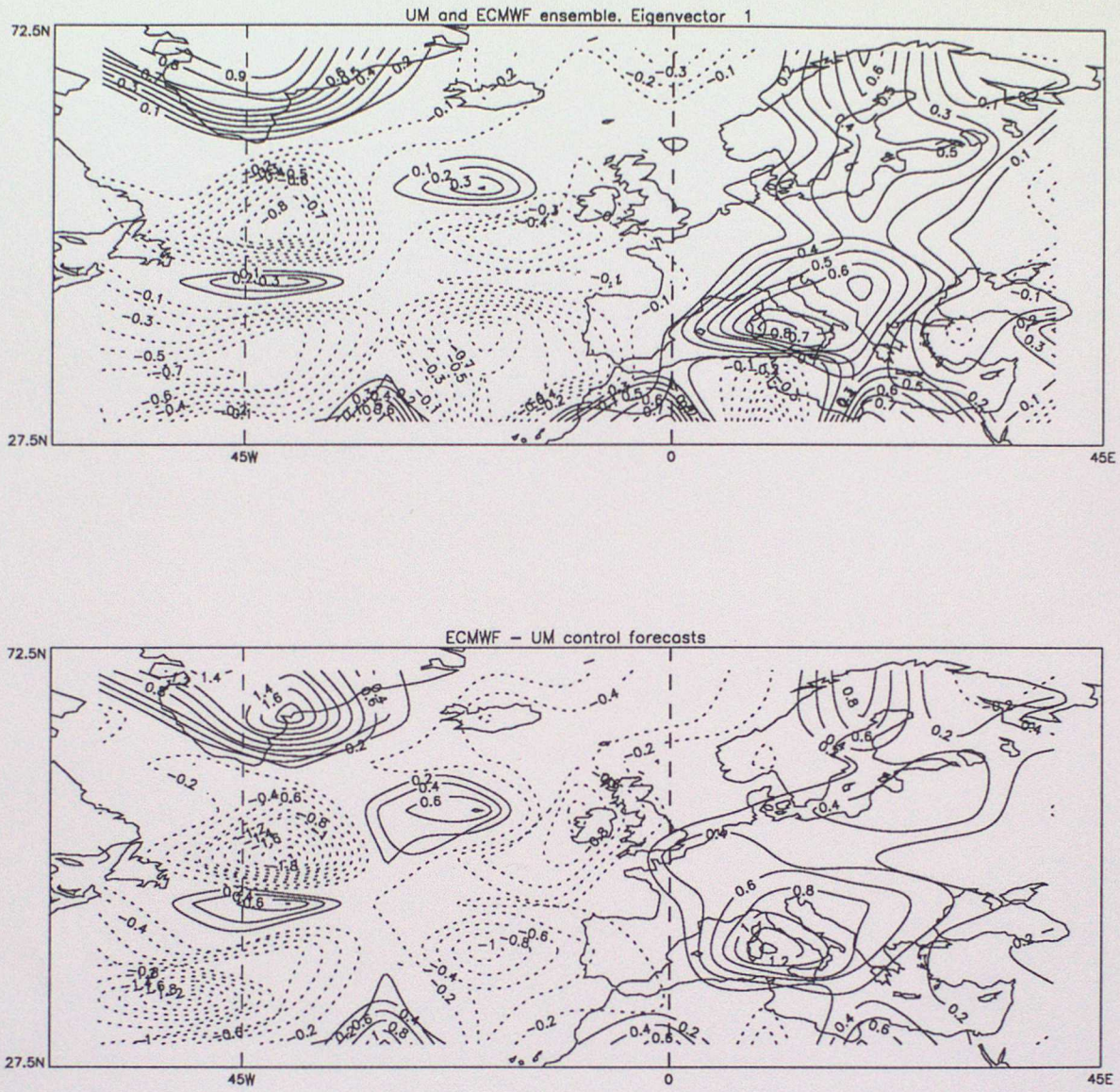


Figure 3. First eigenvector of the joint ensemble at T+00 for the North Atlantic and Europe (upper panel) and difference between ECMWF and UM analyses (lower panel) for 29 October 1994. Equivalent correspondence between first eigenvector and analysis difference is found in all cases.

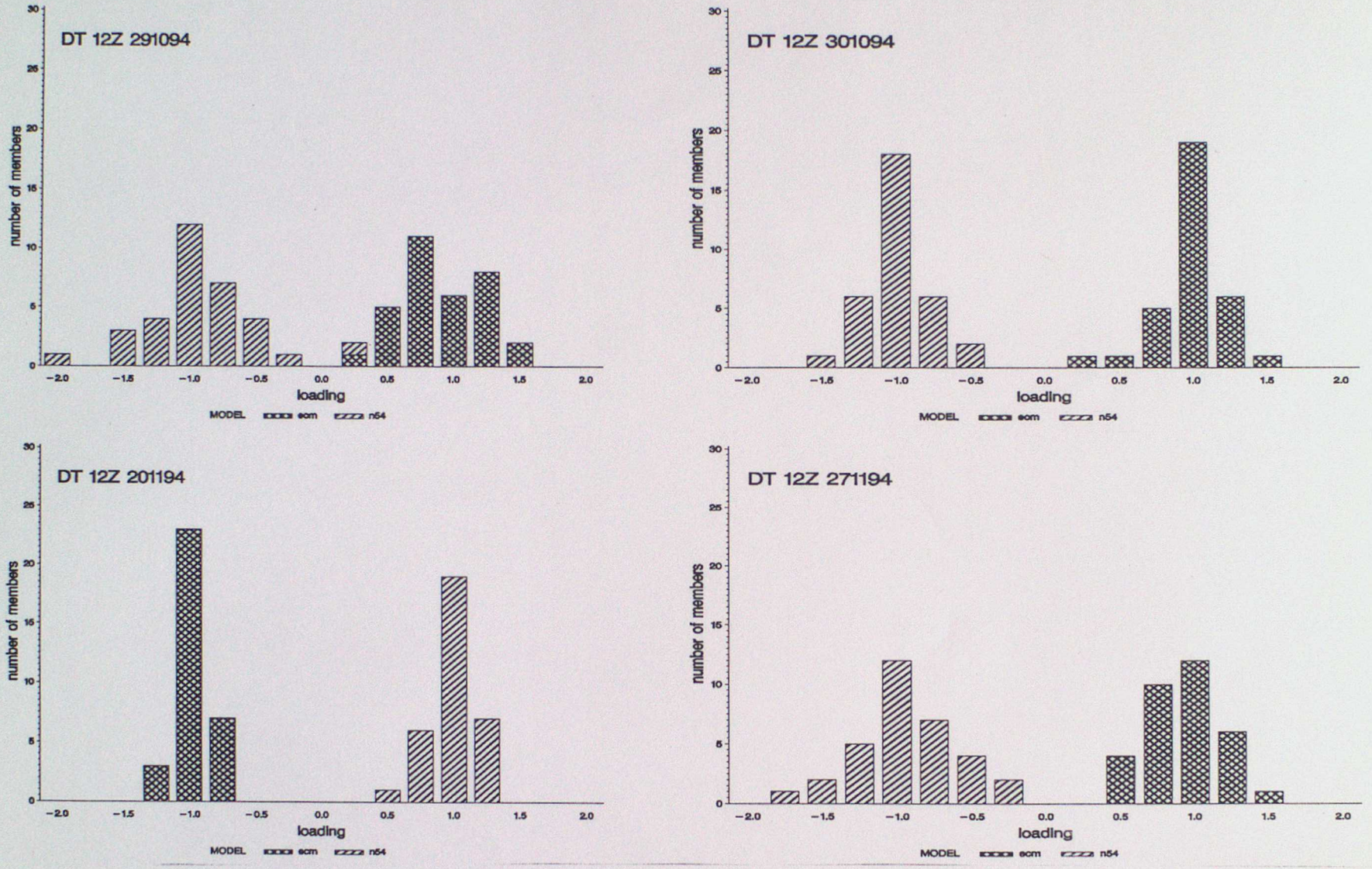
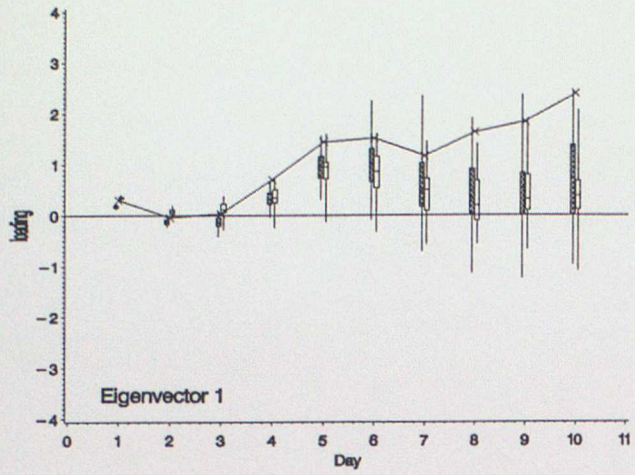
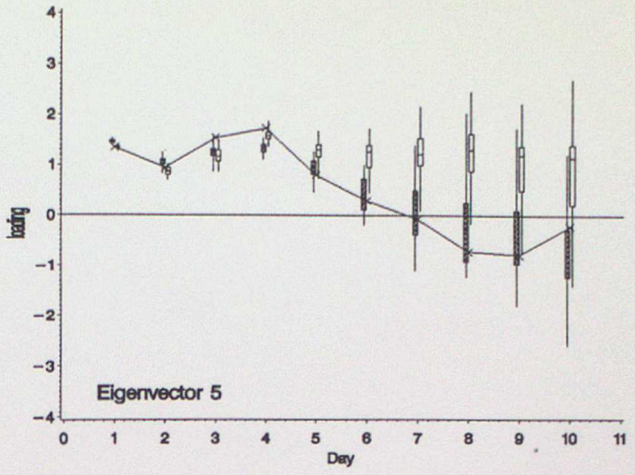


Figure 4. Distribution of loadings on the first eigenvector of the joint ensemble at T+00 for the North Atlantic and Europe (ECMWF members cross-hatched).





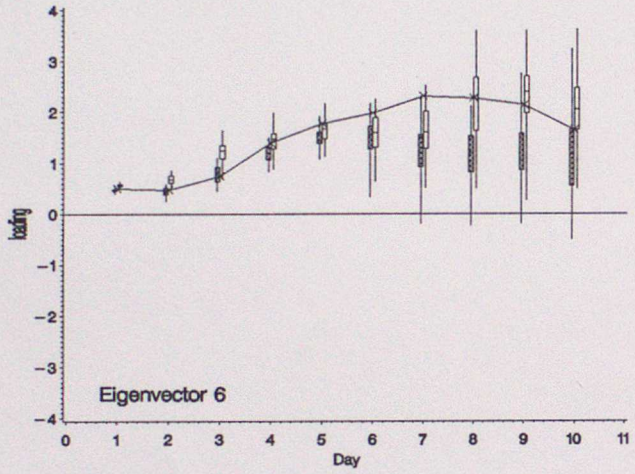


Figure 5(a). Timeseries of loadings on eigenvectors of the observed global phase space for the verifying analysis (solid line), UM (filled boxes) and ECMWF ensembles initialised at 12Z on 29 October 1994. Boxes show median, 25th and 75th percentile and range of loadings for each ensemble. Only loadings on eigenvectors 1,5 and 6 are shown.

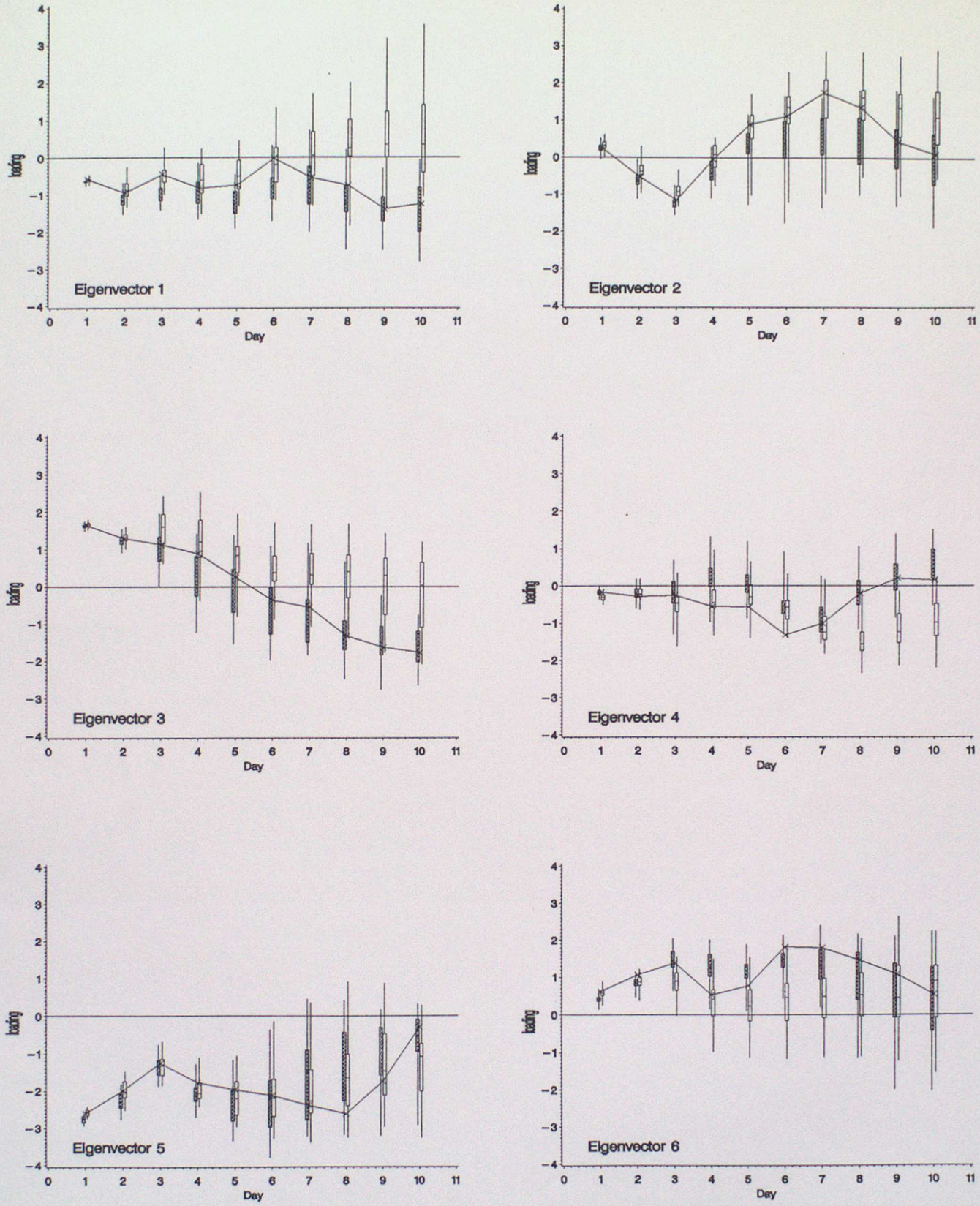
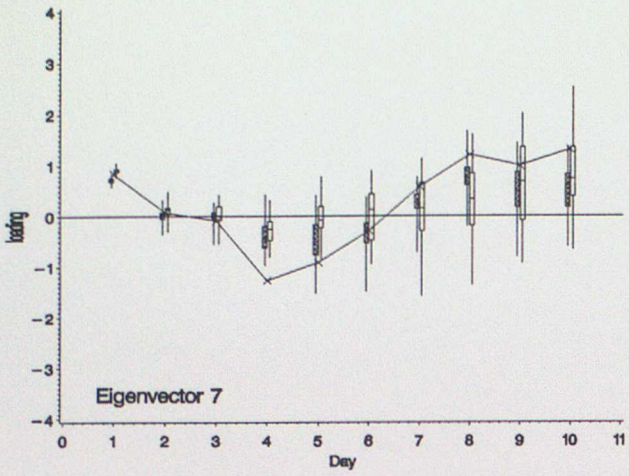
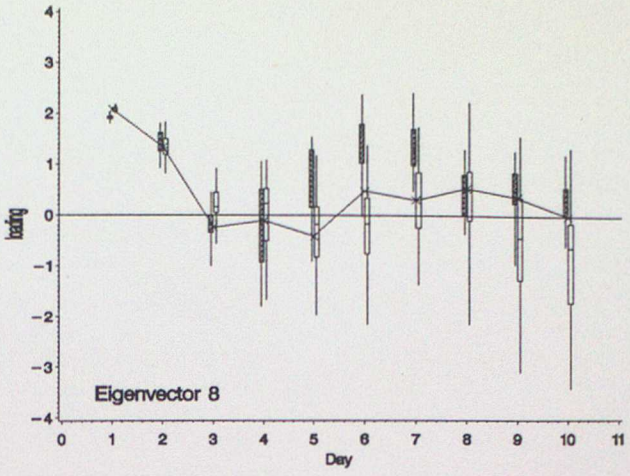
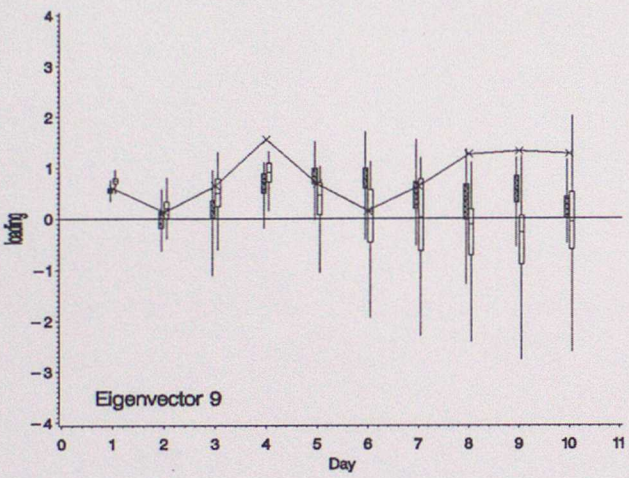


Figure 5(b). As figure 5(a) for ensembles initialised at 12Z on 26 November 1994. All eigenvectors are shown.







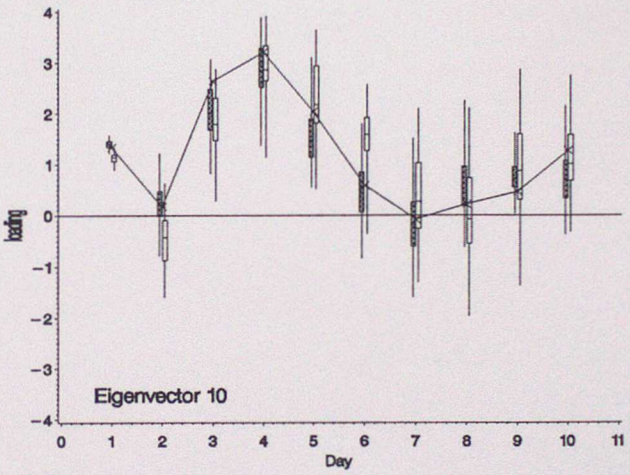


Figure 5(b). (continued)

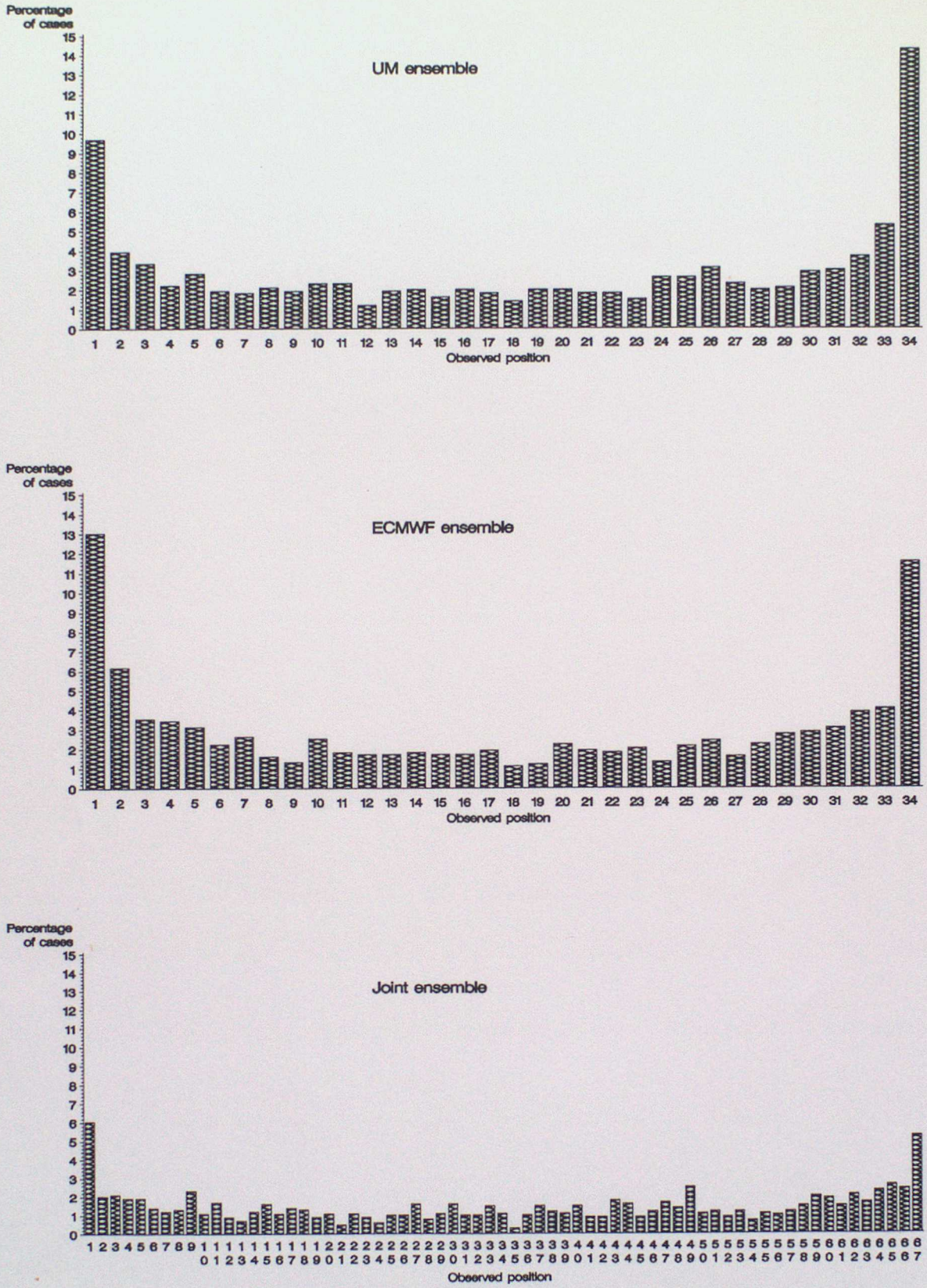


Figure 6. Talagrand diagrams for 500 hPa height over the North Atlantic and Europe at day 7 over all ten cases. The extreme left-hand (right-hand) bar indicates the percentage of occasions where the observed height was below (above) all the ensemble members, i.e. the observation lay outside the ensemble distribution.

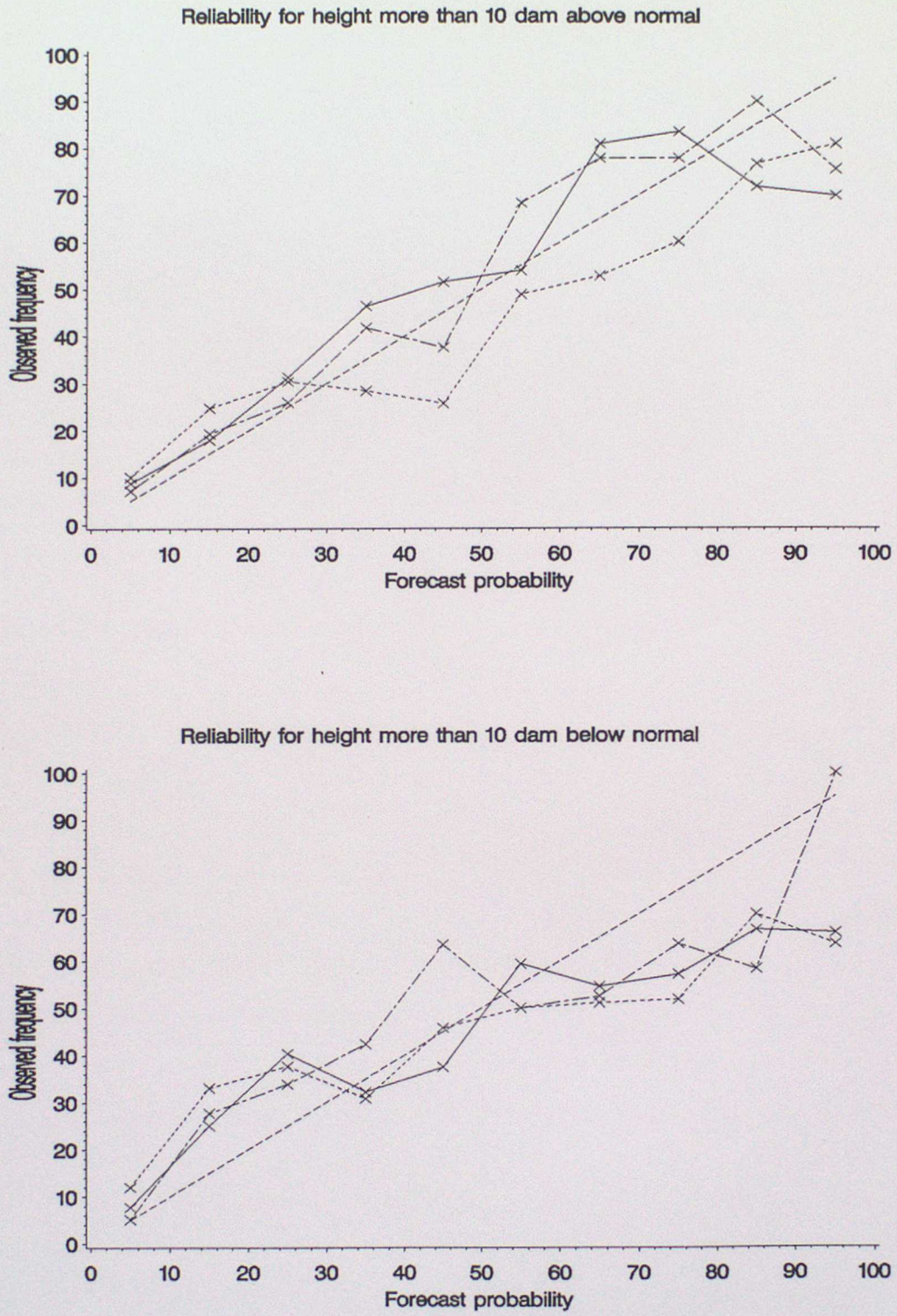


Figure 7. Reliability over all ten cases for forecasts of 500 hPa height anomalies greater than 10 dam above or below normal over the North Atlantic and Europe; UM (solid line), ECMWF (dashed line) and joint ensemble (dash-dotted line).

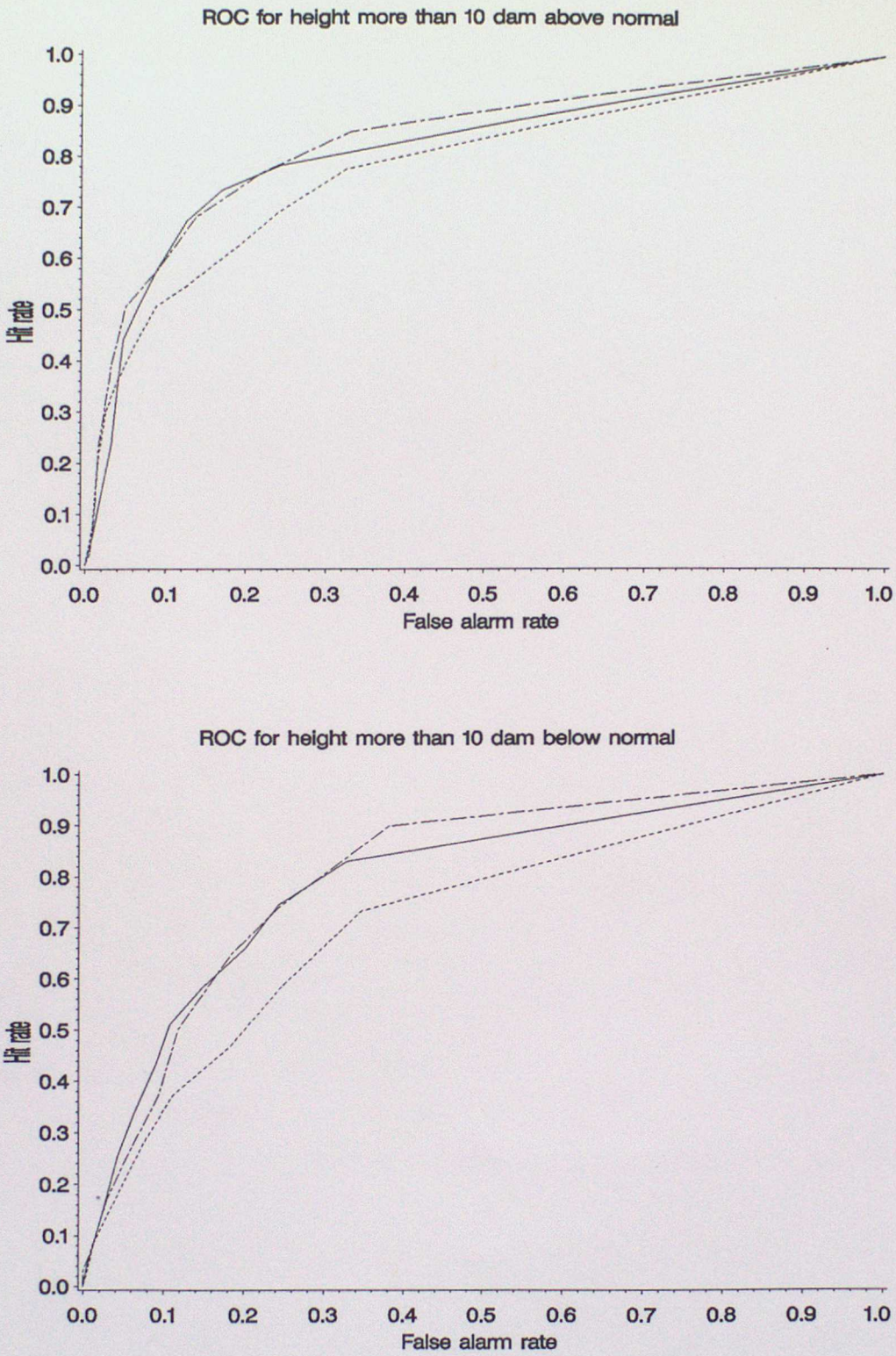


Figure 8. ROC over all ten cases for forecasts of 500 hPa height anomalies greater than 10 dam above or below normal over the North Atlantic and Europe; UM (solid line), ECMWF (dashed line) and joint ensemble (dash-dotted line).