CURRENT PRACTICES AND PERSPECTIVES IN MEDIUM RANGE FORECAST

F. Atger
Météo-France
Toulouse, France

1. INTRODUCTION

Many things have changed during the last fifteen years in medium range forecast, and forecasters have had to change their usual practices. We are currently living a revolution in medium range forecast, because of the high resolution of the new Center's model and the perspectives of the experimental Ensemble Prediction System (EPS).

2. HISTORY

2.1 70's: prehistory

French medium range forecast was a very special activity: a staff of specialists used various empiric methods, based on climatology and analogues, to make a very rough medium range forecast.

2.2 80's: delusion and disillusion

Since August 1979, the Center's model made numerical forecast until D+10 routinely available to forecasters. Medium range forecast was regarded as a very promising activity. It was thought the progress of numerical forecast would be fast and it would be soon possible to perfectly know the weather five days before. The T63, then the T106 Center's model had yet relatively bad results: forecasted fields were very smoothed beyond D+2 and were often little realistic.

The forecaster had to try to correct the numerical products, increasing the dynamics of synoptic features. For instance, the forecaster could see a cut-off where the model only showed a trough; or increase the activity of a wave when its contrast was underestimated. The aim was to emphasize the weather phenomena in end forecasts: temperature drop, strong wind, rain, etc. Forecasters had a too deterministic attitude, they were always looking for details and often tempted to see a strong trough where the model only showed a light weakness in a western flow for instance.

As the fields were smoothed, the model was quite day to day consistent and the skill was never very high, but seldom very low: it was generally possible to find similarities between forecast and reality, regarding the forecast as a smooth of reality.

Thus, forecasters began to wonder if they really could improve the numerical forecast, or if it would be better to follow the smoothed pattern: result would be wrong, but enough close to a medium scheme to not cause strong errors in end forecast. On the contrary, forecasters always risked to damage the numerical forecast modifying it.

At this moment, temptation was great to use automatical systems to make medium range end forecasts, using statistical adaptations of numerical products.

2.3 90's: perplexity

In 1991, the Center's model was changed from T106 to T213. This change perplexed forecasters. From this moment they had to change all their habits, since the new model had a contrary behaviour to the previous one: contrasts were exaggerated instead to be underestimated, synoptic features were too much dynamical instead to be smoothed.

Forecasters noticed the new model was sometimes able to forecast synoptic features with a high skill, even beyond D+4, what was quite impossible with the previous model.

But some forecasts were also very wrong, much worse than it was usual with the previous model. Spurious features might make forecasters mistake a lot in end forecasts. Above all, the realism of the fields might make forecasters self-confident in their forecasts.

<u>These problems showed numerical forecasts could absolutely not be directly used to issue end forecasts but had to be interpreted by forecasters</u>.

Forecasters found again serenity: the nightmare of automatical forecast had disappeared. But how can they achieve the interpretation of numerical products?

3. CURRENT PROBLEMS AND PRACTICES

3.1 Specificity of medium range forecast

Forecaster's responsability is: interpretation of numerical products in order to make the best weather forecast as possible. That means (in brief!): deciding if a numerical forecast is right or wrong. We know it is relatively easy in short range. First because the future evolution is very depending of the current situation. Next because the high skill of the model allows forecasters to detect errors or inaccuracies comparing various predictive elements. But in medium range it becomes very difficult, especially because of the low skill. Finally, <u>medium range forecasters never know wether a forecast is right or not</u>.

3.2 Inconsistency

3.2.1 *Day to day inconsistency*

One time out of three, three consecutive runs are inconsistent over France. Fig. 3a, 3b and 3c show an example of the model's jumpiness. No comment.

During a long time French forecasters had been using the empiric rule which is: as day to day consistent the numerical forecast is, as good it is. Many studies showed it is wrong: since the skill is decreasing in medium range, inconsistency must exist in order to correct the previous run and to improve the last one.

But forecasters are pigheaded. In spite of many examples they have seen during last years, even now they can absolutely not accept that there is no relation between consistency and skill. In France we are maybe more pigheaded than elsewhere: we succeeded in proving that when three consecutive runs are consistent, then the skill is a little higher (T. Lefort and B. Mornet, Synthèse des contrôles subjectifs du modèle haute résolution du CEPMMT, mars 1993). Actually this result only concerns 20% of the cases, and most of them are blocked situations.

When inconsistency is important, forecasters know the last run is statistically more skilful than the previous. But it is not an absolute rule, sometimes the oldest run is more skilful.

3.2.2 *Differences between models*

Inconsistency also exists between models. In France we use only the French model "Arpege" (wich is only a short range model) and the Center's one, so our problems only concern short range, until 72 hours. In most cases, differences between both models are light and forecasters choose an intermediary forecast for the D+2/D+3 turning point. But in 25% of cases, differences are too important and the two numerical forecasts are almost incompatible. Forecasters must make an absolute choice, they must partly reject one of both forecasts. French forecasters generally prefer the French model, not only because of patriotism, but also because its analysis is more recent than the Center's one, and they can easilier find reasons of the difference.

Fig 4a et 4b show a great difference between the models at 84 hours. We can see the reason of this difference comparing the analysis of the French model and the Center's 12 hours forecast (fig 5a and 5b): the trough is too much emphasized on the Center's forecast and it is too early cutting-off. Actually the French model was better.

Such a choice lead the forecaster to a big problem: how can he make a medium range forecast when he rejects the Center's model in the short range? He must imagine the next, since the French model is only a short range model. Fig 7a and 7b show such a problem. The forecaster had chosen the French solution, with a low near Portugal, not over Britanny. He was right (fig 7c). For his medium range forecast he imagined the low staying near Portugal, since it was "independent" of the large scale flow. Fig 2b shows the result at +120 hours. Bravo!

3.3 **How to make a choice?**

These problems of inconsistency always lead the forecaster to be left in the alternative: wich run must he choose, wich model must he choose?

3.3.1 *Objective methods*

Actually, two models (or two runs) are seldom really incompatible. Experience taught forecasters to see similarities rather than differences, and an intermediary forecast generally is more skilful than the both ones. The forecaster know some systematic defects of the models -"the Center's model is too meridian" for instance- and he can make a combination with the "best" things of each model to obtain the "best" forecast. He also know looking for differences in different analysis to see which run might be wrong.

3.3.2 *Realism*

When the evolution is not natural, not realistic -but I don't know exactly what it means- the forecaster naturally tends to reject it and to choose an other one. In 1991, when the model changed, forecasts always seemed not realistic, because forecasters were used to see the smoothed fields of T106 model. And the new one is more realistic than reality. After two years of use, forecasters think they can see when something is "too much" in the model. It is probably true for small scale spurious features which appear without reason, although it is generally impossible to find the reason of error in the analysis, because of the small scale. Fig. 8a, 8b and 8c show a good example of such an "anomaly". Fortunately it is not frequent. This low was rejected by the forecaster. He was right (fig 9a and 9b).

But for large scale features it is quite dangerous to reject a numerical forecast only because it is supposed unnatural, when it is not a consequence of an obvious error in analysis. A good forecast may be completely rejected, as fig. 3b and 10 show: the forecaster completely rejected a very unatoural -and also very day-to-day inconsistent- evolution... which was right! Because of the high activity of the model, it is better to distinguish between what is essential and what is secondary. For instance, the model often roughly cut-off troughs. Generally the idea of cut-off is good, but forecasters feel it is excessive and they want decrease the cyclonism and maybe delay the cut-off. The problem is that the position of a low of course depends of the moment of the cut-off. Thus, forecasters never exactly know where lows will take place.

3.4 Beyond D+4...

With such an uncertainty, we must wonder if it is reasonable to make weather forecasts beyond D+4. We know we cannot say four days before if it will rain or not. But we hope we can have an idea of the synoptic pattern. Subjective French controls showed that more than 50% of numerical forecasts are "good" until +108 hours, what is an improvement of 24 hours in five years, since the last control in 1987. But at 156 hours (D+6/D+7) only 15% of numerical forecasts are "good" (Lefort and Mornet, Synthèse des contrôles subjectifs du modèle haute résolution du CEPMMT, mars 1993). A forecast is "good" when the synoptic pattern is sufficiently right to give right outlines of the weather over France. So beyond 120 hours we should have to forget synoptic features and to study only the large scale pattern. But it is not easy, and forecasters are always tempted to have confidence in the synoptic flow pattern when the model is realistic.

3.5 Predictability

Eventually, forecasters got use to mistake in medium range forecasting. But they know that the error is very variable and they would like be able to forecast the importance of this error. They would want to know the predictability of atmosphere to inform some end users of the expected quality of the forecast: they would improve quality of their forecast just by informing about the quality. Forecasters intuitively use the confidence they have in the model as a predictability tool: when this confidence is high, they are very affirmative in weather reports; when this confidence is low, they use the conditional mood and do not give details. But that is very subjective and, at present, forecasters are not provided with a real predictability tool which would allow them to give probabilistic indications in end forecasts.

3.6 In brief

We meet two problems in medium range forecasting. First: what can we do in the alternative, when we must choose between two forecasts, one as probable as the other one. Next: how can we forecast the quality of our forecast. For almost one year Ensemble Prediction System (EPS) have been an attempt to provide forecasters with a predictability tool and a new method of investigation in medium range.

4. ENSEMBLE FORECAST

For more details about experimental results, see in the report of Expert Meeting on EPS, july 1993: F. Atger and T. Lefort, ECMWF's Ensemble Forecast preliminary results; and: P. Lamboley and E. Legrand, Investigation and Valorisation of the Ensemble Forecast Experiment at Météo-France.

4.1 Usefulness

Ensemble Prediction System (EPS) was first described as a predictability tool for medium range forecasting. Since the 33 runs have the same probability, their spread, their dispersion, must be theoretically in relation with predictability. Infortunately it does not practically work. In France, forecasters did not find any relation between spread and skill. Statisticians found something, but only beyond D+5, and with a very large confidence interval. Fortunately, predictability is not the only use of EPS.

Forecasters must be little masochist to use EPS. The main difficulty they have encountered for many years is to choose a forecast between two models or different successive runs. EPS increases the number of these alternatives with its 33 different runs. If you ask a forecaster what he does prefer, only one model or many different ones, his answer probably is ambiguous: he prefers only one perfect model, but many poor ones. It is the same thing as day to day inconsistency: every day forecasters wait for the new run, hoping it will be consistent with the previous one, but also it will be better; what is dramatically impossible. So, we already know what the forecaster's problem with EPS will be: how can he choose a forecast among the 33 ones? And, reasonably, must he choose? That is the question, and that is an essential difference between a deterministic approach of EPS and a probabilistic one.

4.2 **Deterministic approach**

Clustering is the technic used in this approach. Since the probability of the 33 runs is the same, the probability of each cluster depends on the number of its runs. So it would be easy: the largest cluster would lead to the best forecast. But practically it is not so easy to use.

4.2.1 *Spread*

First, the spread is too weak. Maybe the number of runs must be increased; or initial perturbations must be modified. Runs are frequently too close together and too close to the operational one. When some runs are quite different of the operational one, they are generally in minority and EPS only confirm the operational forecast. <u>In 80% of cases, EPS results are useless</u>. For instance fig.12a and 12b show some differences between clusters, but a forecaster should not make different end forecasts over France.

4.2.2 *Clustering*

Then, in 20% of cases, only one time out of five, significant differences appear between the 33 runs. But in these cases, French forecasters often consider the clustering is unsuited. The classification does not separate runs with significant differences; on the contrary individual runs from different clusters are very close together. But the gravity of the problem depends on the geographical area and on the range, we shall see that point later. Currently, forecasters cannot work directly with the 33 runs. So they try to use provided clusters, sometimes trying to group some of them showing similarities. But differences generally appear beyond 108 hours, just when the scale of these differences must be large to lead to significant differences in the end forecast. Eventually, there are only few cases when EPS is useful.

4.3 **Examples**

4.3.1 *The good choice?*

First example is shown in fig. 13 and 14a, 14b. Cluster 1 was not a real alternative, but first it confirmed the cut-off, next its low was a little more south than the operational one, further over Lion's Gulf than over Center of France. Considering the number of runs in the other clusters, other hypothesis could be prefered, as an Iberic cut-off, or an Italian one. Actually, it was only a good intuition to choose the most probable one. Unfortunately this example is very representative of the cases of possible usefulness of EPS.

4.3.2 *Clustering clusters*

In this second example, very interesting, most of runs had a common difference with the operational one: they maintained a blocked flow with cold air on West Europe, whereas the operational run developped a ridge (fig 15). And it was the right forecast (fig 16a and 16b). It is the best example we have seen, because EPS really provided an alternative forecast for the large scale pattern, in middle medium range.

4.3.3 *Truncation's weight*

This last choice could have been dangerous if the scale had been smaller, because the difference might have been consequent on the lower truncation T63. It is a common problem when the forecaster must interpret a difference between EPS runs and the operational one: are differences the consequence of initial perturbations or only of the truncation? Fortunately forecasters can partly resolve this problem studying the non-perturbed T63 run ("control"). Experience showed us that when the use of EPS leads to damage the operational one, what it means when the operational run is better than majority of EPS ones, it is generally situations in which the weight of truncation is particularly important, with small scale features. This third example shows that: only a high resolution model could provide a correct forecast, EPS runs were too smoothed (fig 12a, 12b and 17a, 17b).

4.5 **Clustering improvement**

Eventually, although there were only few cases of usefulness, EPS generally brought an improvement, and many cases of damage could be avoided in "sensitive to truncation" situations. Except the too weak dispersion, the main obstacle to a real EPS usefulness seems the inappropriate clustering.

4.5.1 *Distance*

The chosen distance is a norm, a mathematical distance, not a meteorological one: for instance two similar, but out of phase, patterns will be mathematically distant. The forecasters' intuition must be the best distance and it is possible to approach it experimenting various distances with them.

4.5.2 *Domain*

Essential problem is the domain of clustering: the geographical domain but also the chosen period. Runs are currently clustered by the Center on a European domain, between 120 and 168 hours. Until 120 hours this domain seems too large in order to use the clustering to make a forecast on a small area, as an European country. A synoptic scale difference in a distant place may cause an unsuited clustering, hiding smaller scale differences over the country, which might have important consequences on the weather. The current period of clustering is not either satisfactory in this perspective because +168h is too far.

On the contrary the current domain seems too small beyond 120/132 hours, since synoptic scale differences must not be considered at this range. Only differences in the long wave pattern would have to be used. It might be necessary to make two clusterings, one in a large domain, for the middle medium range, an other for the early medium range, in a smaller domain, as a country or a group of countries.

We experimented a clustering on a "French" domain, with current distance and period. Fig. 20 and 21 show respectively a representation of the Center's clustering and the French one: it is the projection of the interdistance matrix (the points cloud of the 33 runs) on the two axes of maximum explained variance. We can see that distances between runs are very different on the European domain and on the French one.

4.5.3 *An interactive clustering?*

This representation also allows us to wonder what is the best clustering for a given distance and a given domain. What rules? How many clusters? etc. The center's clustering seems logical on European domain (Fig 20), but maybe a forecaster would have differently clustered. Probably the manual human choice would be the best, not directly from the fields, but for instance with this projection. Moreover, it would be interesting to change the domain and the period of clustering according to the meteorological situation. At last, the location of the initial perturbation may be a precious information to evaluate the intrinsic reliability of an independent run, and to cluster it. An interactive system on workstation would allow the forecaster to achieve various subjective clusterings, considering the relative positions of runs, the meteorological situation, the places where initial perturbations growed, etc.

4.6 Probabilistic approach

In the deterministic approach we were trying to find the most probable forecast. Now, we give up knowing how the weather will be, and we accept all the possible forecasts. The ensemble of all different forecasts is regarded as one probabilistic forecast.

4.6.1 *Current products*

Currently, only the T850 hPa and Z500 hPa plumes and the 33 runs rainfall are available. The two first parameters don't seem very interesting, because the weather doesn't depend on Z or T value, but rather on flow, especially in medium range. T850 could give an idea of average temperature, what could be interesting in medium range, but there are some problems with the too rough orography of T63 model. And there are many places near moutains all over Europe... Rainfall seems more interesting, even if the rain forecast is not very skilful in a T63 model.

4.6.2 *Probability charts*

For the moment we have not systematically controlled these documents, but we have noticed a drawback: they give probabilistic indications in particular points, what is not really useful in medium range because of the quite important uncertainty, and the large scale of features we can forecast. It would be interesting to provide forecasters with probability charts, on which different fields could be contoured, for each range: probability areas, for instance probability that rainfall will exceed 10 mm/12 hours; caracteristic contours of the 33 EPS runs, for instance 0°C at 850 hPa; caracteristic points of the 33 EPS runs, for instance lows and highs. According to the range, the domain could be extended and parameters smoothed. For instance the 33 smoothed hemispheric jet-streams could be useful beyond D+4.

4.7 Phrasing

4.7.1 *Today*

Probabilistic approach is not currently practiced in France, but we are thinking about the problem, especially about this question: how can we phrase uncertainty or probability in end forecasts? This problem already exists a little, since forecasters are used to give more or less details about the forecast, according to the confidence they have. But they never -or seldom- really mentionned their confidence and the uncertainty about the forecast. The end user only notices the forecast is not very accurate and thinks the forecaster is not self-confident. For instance the phrase "week end forecast is very uncertain" is common in technical reports for meteorologists, but much less frequent in end forecasts.

4.7.2 *Tomorrow*

When we shall be able to forecast a rainy period with 30% probability, what shall we have to write in our reports? Will users be able to understand what it means? It may rain? Or it must rain over 30% of country? or 30% of people? Perhaps it will be better to reserve this probabilistic information for our professional users who will be able to use it as a risk. And we shall tell "the man in the street": weather will be fine, with a very low risk of rain. In the same way it seems easy to forecast a temperature between 2 and 7 degrees.

4.7.3 *Shipping forecast*

Here is a domain where it is very difficult to phrase the uncertainty of the forecast: we must forecast the wind whereas the position of lows is very uncertain. What information about the wind can we give to sailors in a medium range forecast? In a deterministic approach, giving accurate values would be a swindle. Maybe in a probabilistic approach should we be able to indicate some areas where strong depressionary winds could occur. But only until D+4. In middle medium range we can only forecast the main wind direction, and maybe an activity indicator of atmosphere which could inform sailors about the risk of strong winds.

4.8 **Practical remark**

Practically, there is not a real contradiction between a deterministic approach and a probabilistic one. Indeed, when spread is high it is better to choose one alternative, one cluster, before to use statistical products. On the contrary, if spread is low all solutions, all runs, may be considered as an only one cluster, an only one statistical solution. Moreover, a probabilistic approach is absolutely necessary in middle and late medium range, beyond D+4, since it is impossible to give some accurate indications about weather or parameters as wind. Actually, even if it doesn't clearly appear in end forecasts, a medium range forecast must result from a global investigation, in both deterministic and probabilistic approach.

5. CONCLUSIONS

In spite of the high resolution of the Center's model, our medium range forecasts do not seem to have improve for the last years. Forecasters need specific tools, adapted to medium range, in order to make a real interpretation of numerical products. Ensemble Prediction System may give means of resolving their problems: forecasters are optimistic, but also realistic. They regard EPS as a very promising method, but quite useless for the moment. They are waiting for an improvement and thinking about new problems which will appear when EPS will be really in an operational use.

fig. 3a: EC +144h from 16.10.93
valid 22.10.93

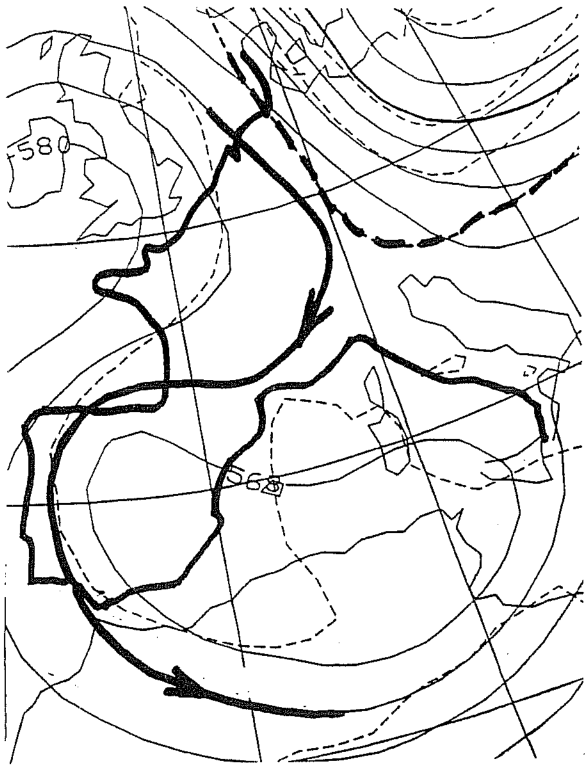


fig. 3c : EC +96h from 18.10.93
valid 22.10.93

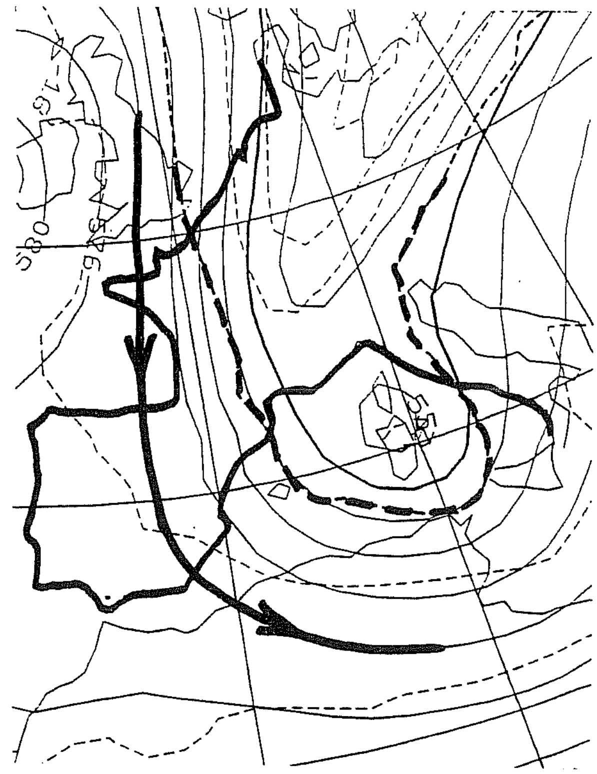


fig. 3b: EC +120h from 17.10.93
valid 22.10.93

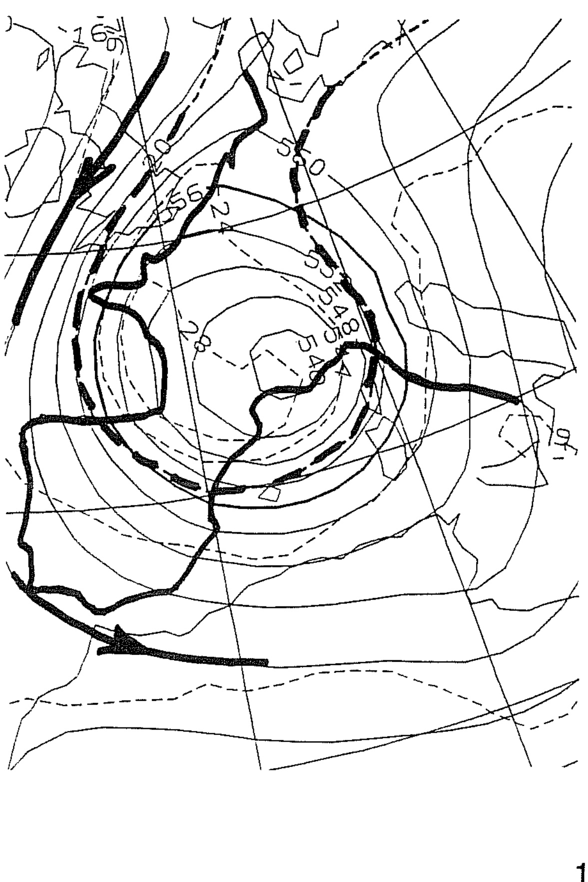
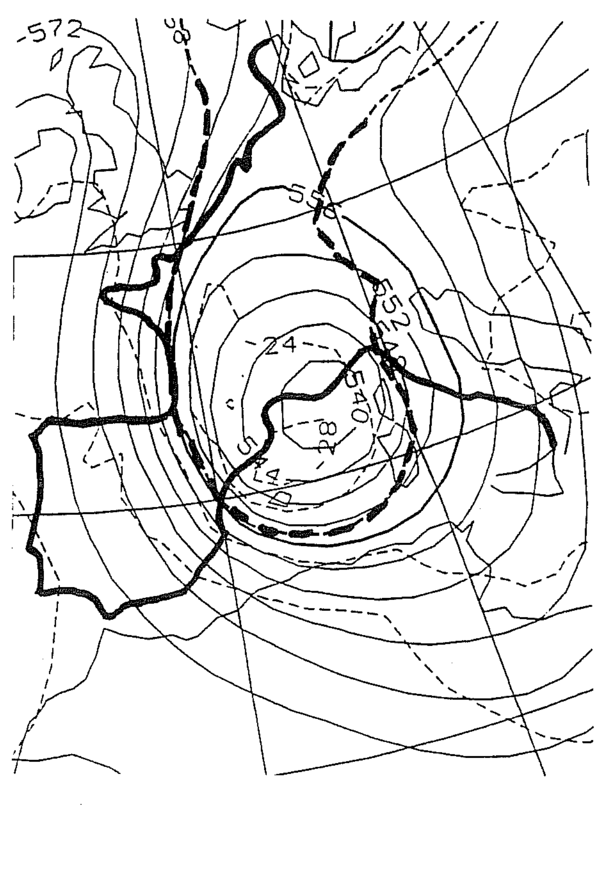


fig. 10 : analysis 22.10.93



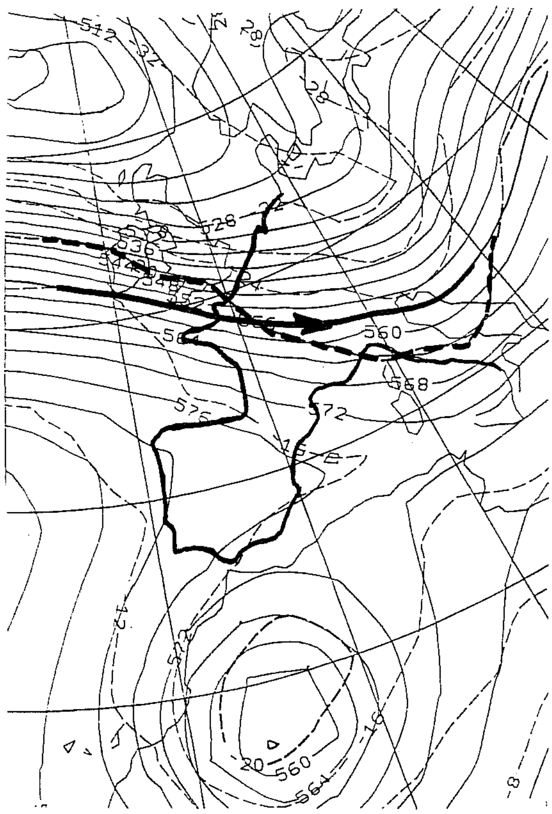


fig. 4a : EC +72h from 14.11.92
valid 17.11.92

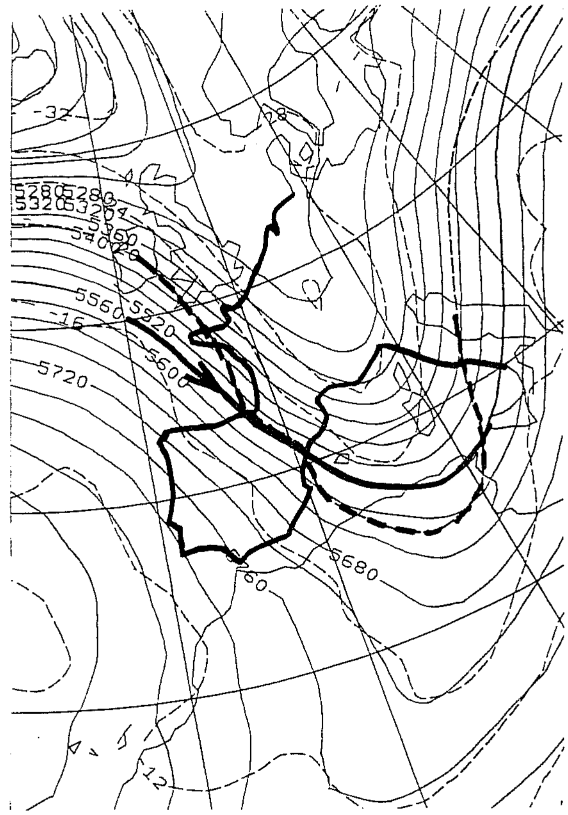


fig. 4b : ARP +60h from
15.11.92 (00UTC) valid 17.11.92

fig. 5a : EC +12h from 14.11.92
valid 15.11.92

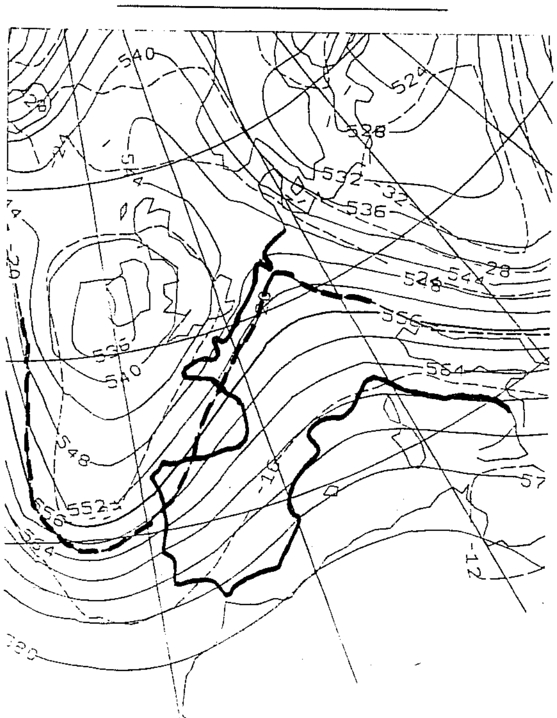


fig. 5b : analysis 15.11.92
(00UTC)

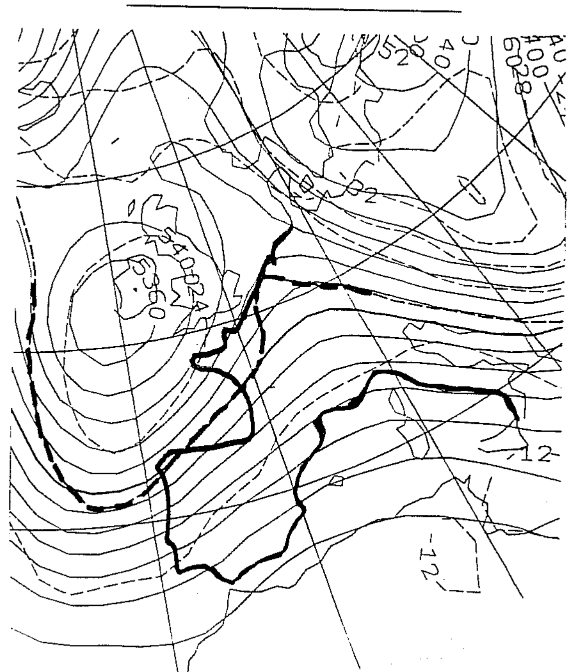


fig. 7a : EC +84h from 13.08.93
valid 17.08.93

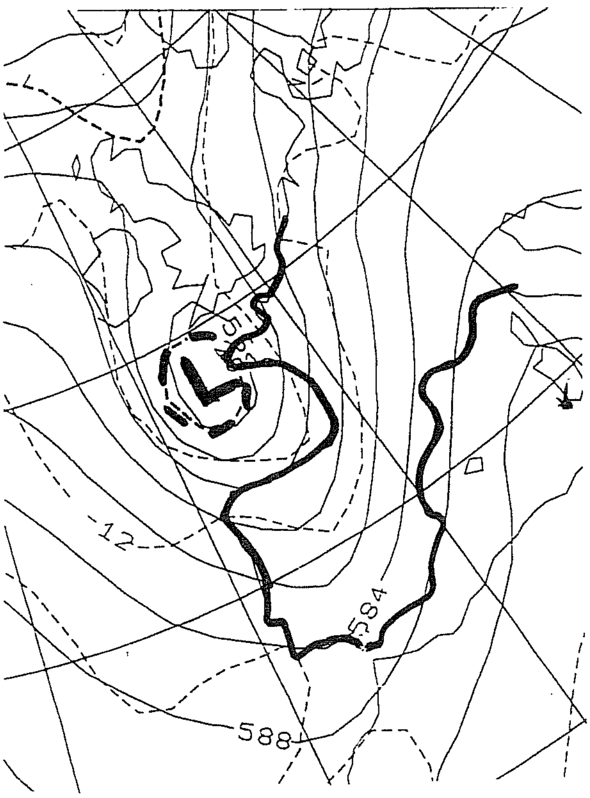


fig. 7b : ARP +72h from
14.08.93 (00UTC) valid 17.08.93

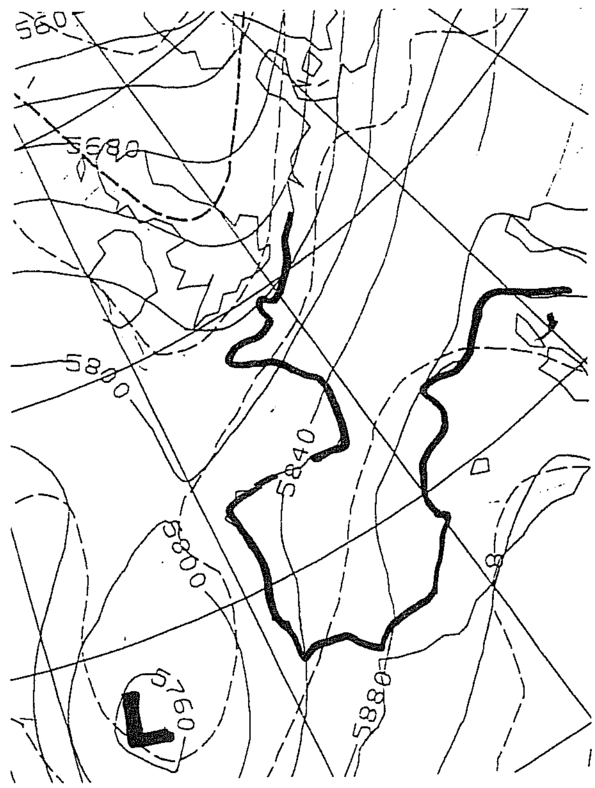


fig. 7c : analysis 17.08.93

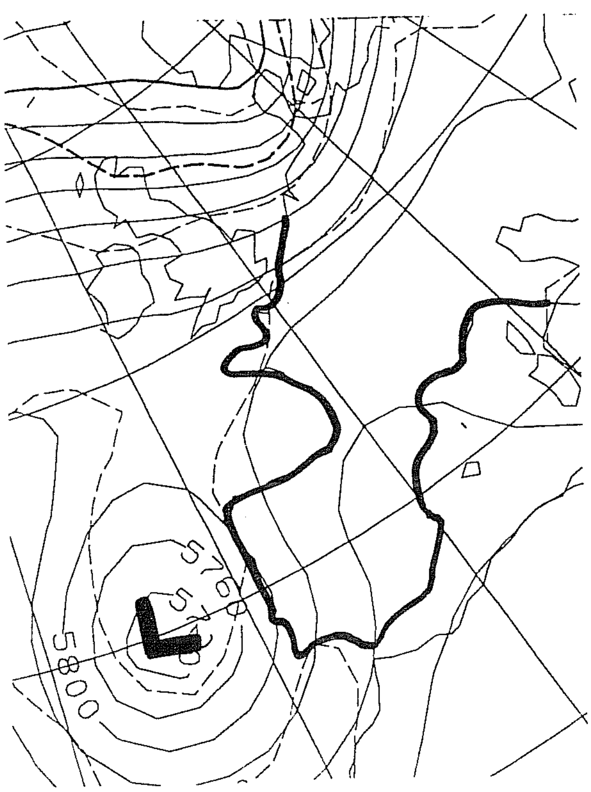


fig. 2b : analysis 18.08.93

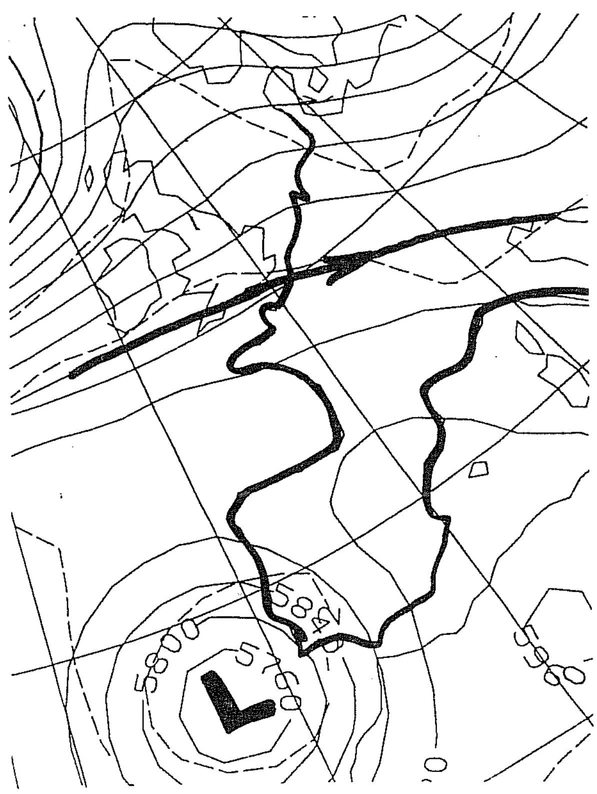


fig. 8b : EC +72h from 24.08.93

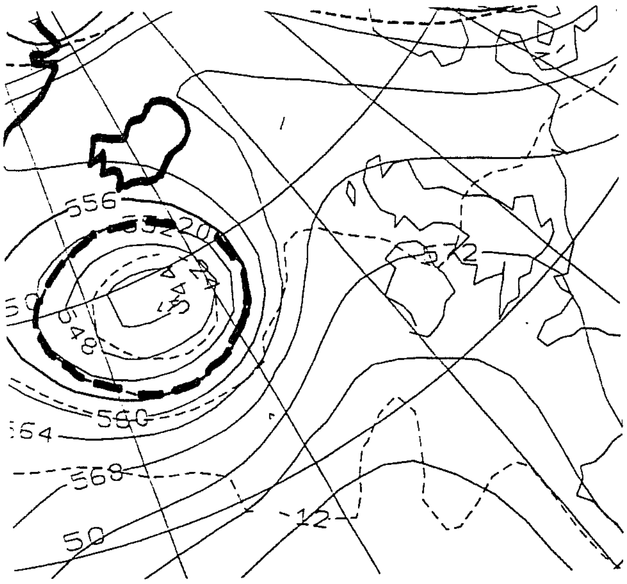


fig. 8c : EC +96h from 24.08.93

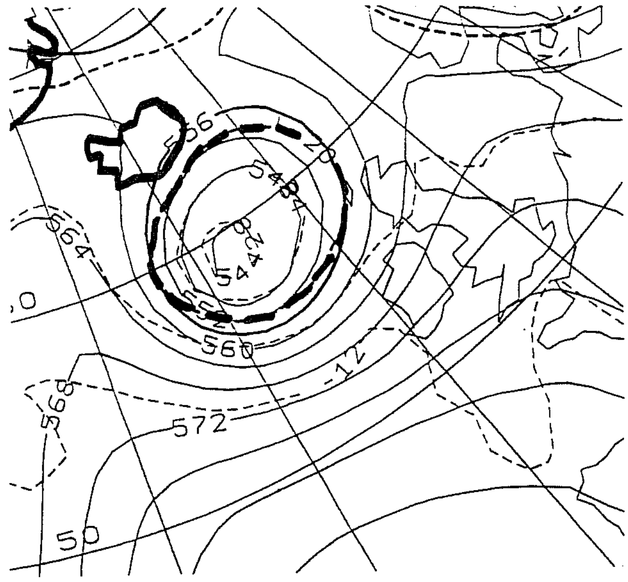


fig. 9a : EC +120h from 24.08.93 valid 29.08.93

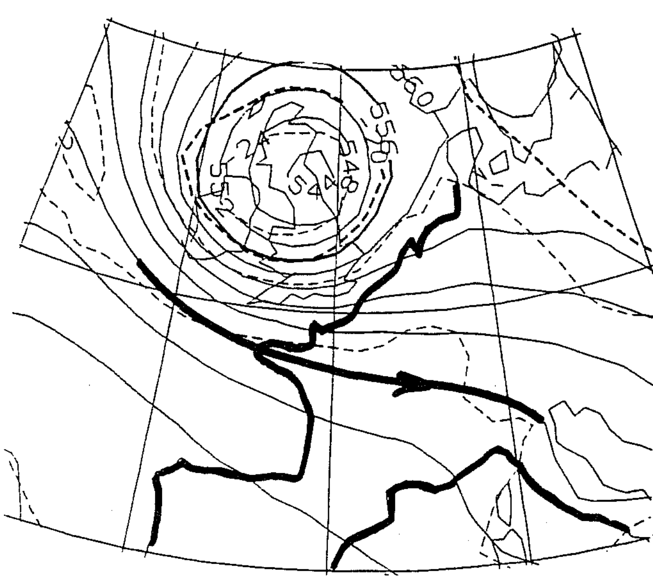


fig. 9b : analysis 29.08.93

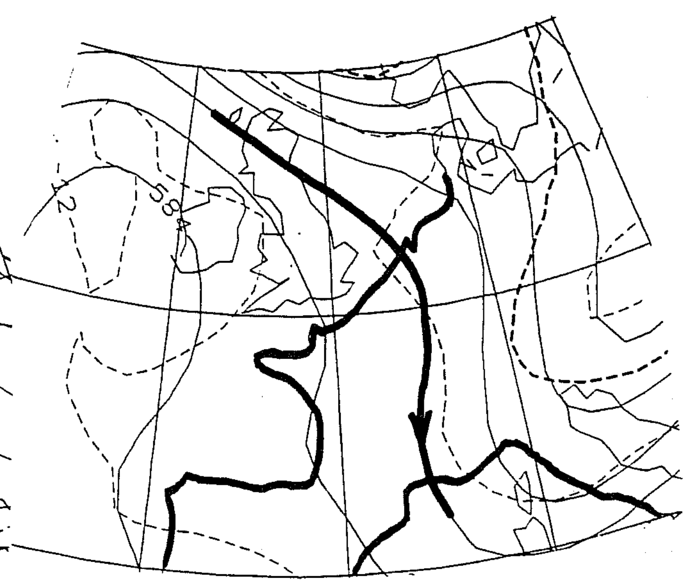


fig. 12a : EF +72h +96h +120h from 15.05.93 cluster 1 (20 runs)

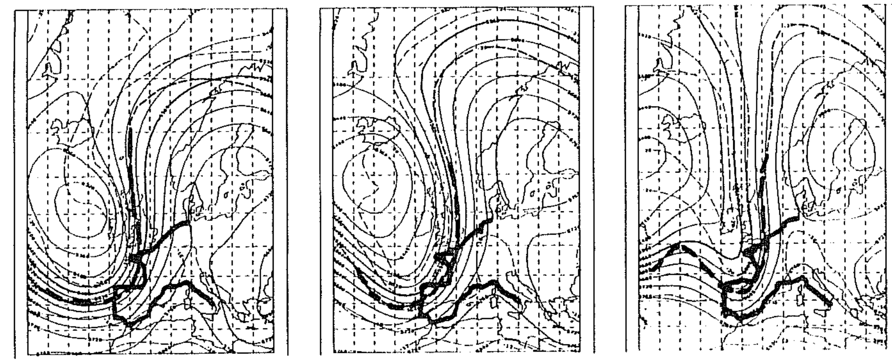


fig. 12b : EF +72h +96h +120h from 15.05.93 cluster 2 (13 runs)

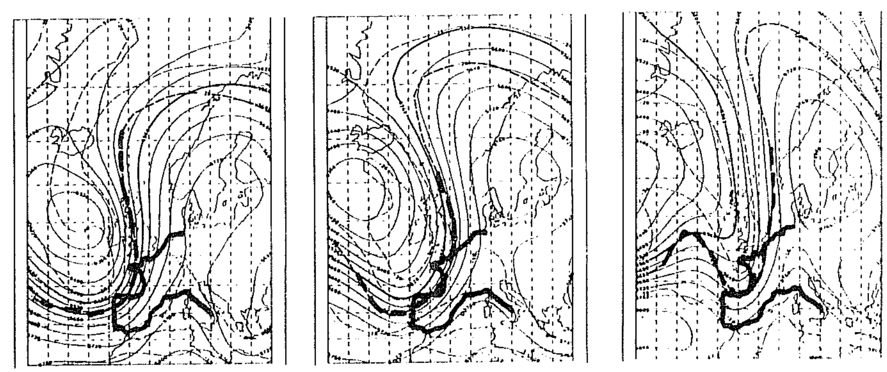


fig. 17a : EC +120h from 15.05.93 valid 20.05.93

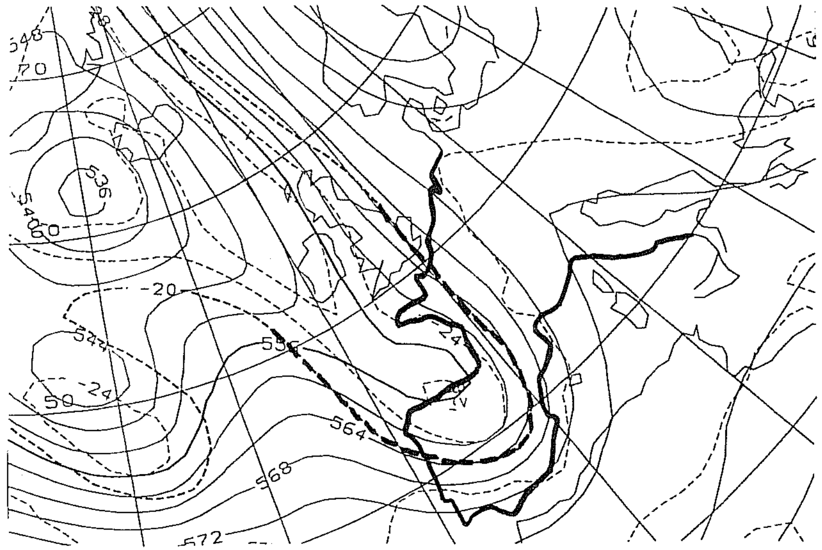


fig. 17b : analysis 20.05.93

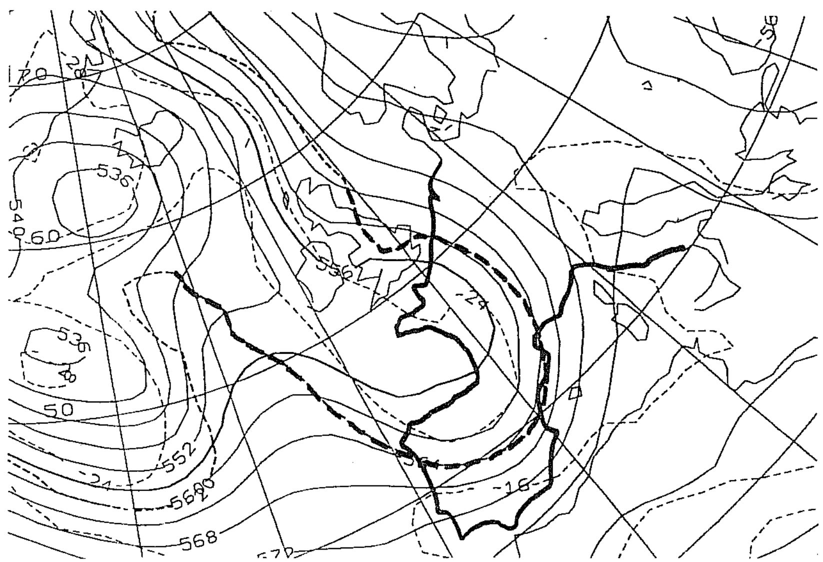


fig. 14a : EC +96h from 20.09.93 valid 24.09.93

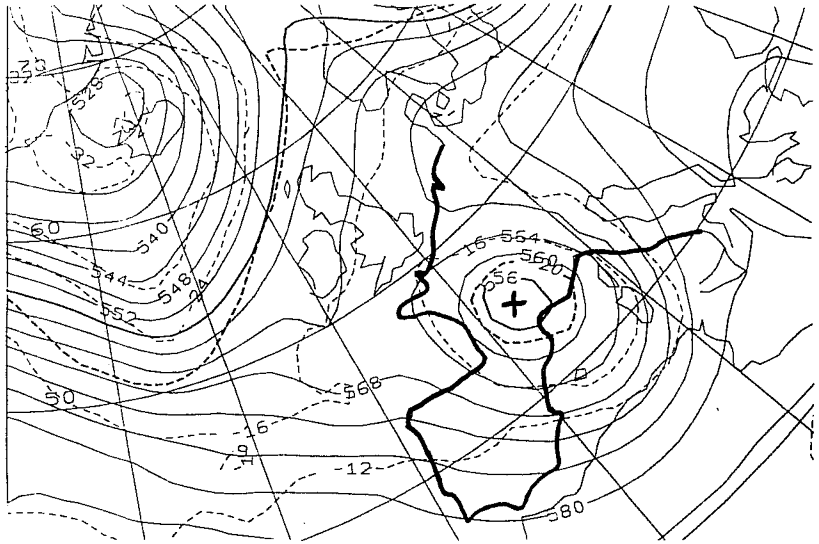


fig. 14b : analysis 24.09.93

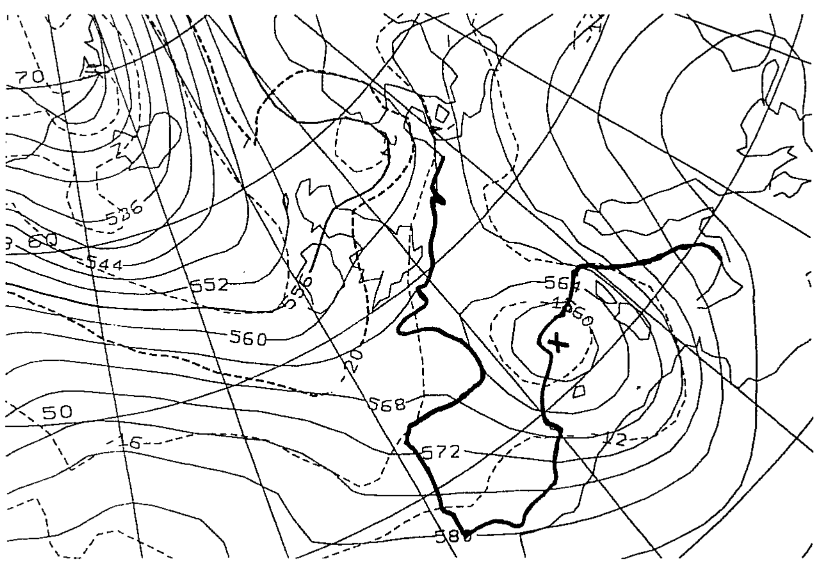


fig. 13 : EF +96h from 20.09.93 valid 24.09.93 clusters 1 2 3 4

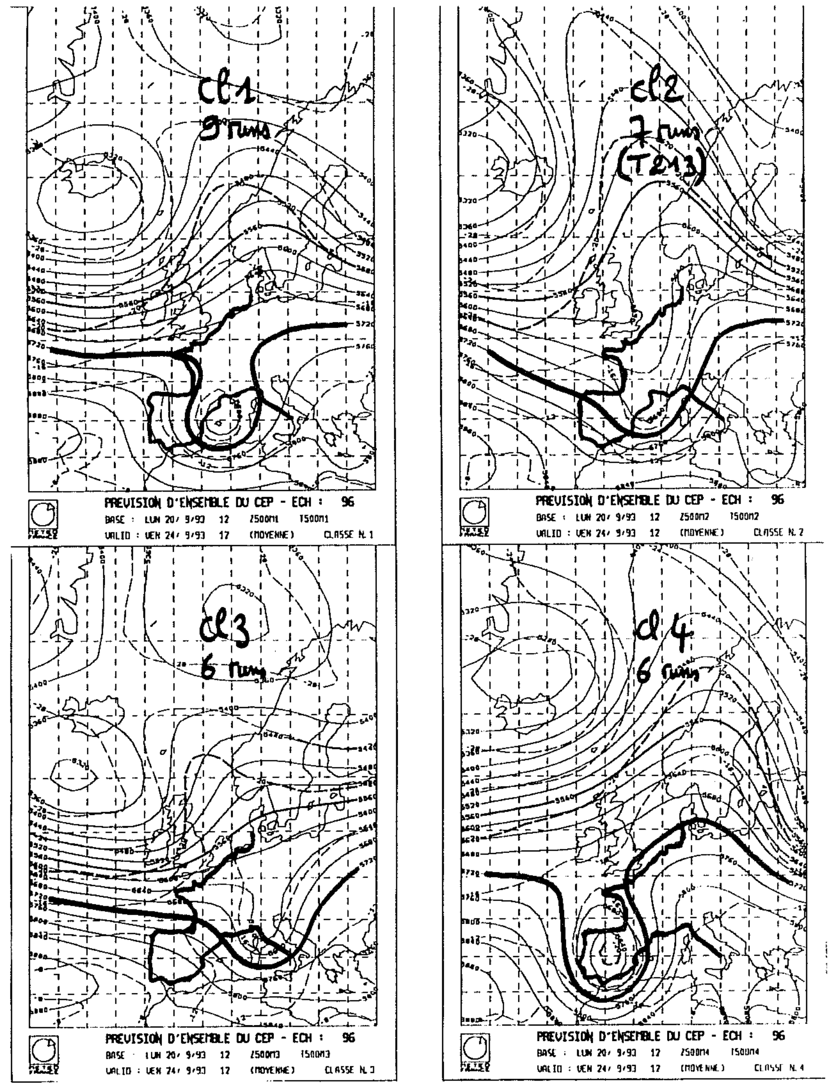


fig. 16a : EC +168h from 26.09.93 valid 03.10.93

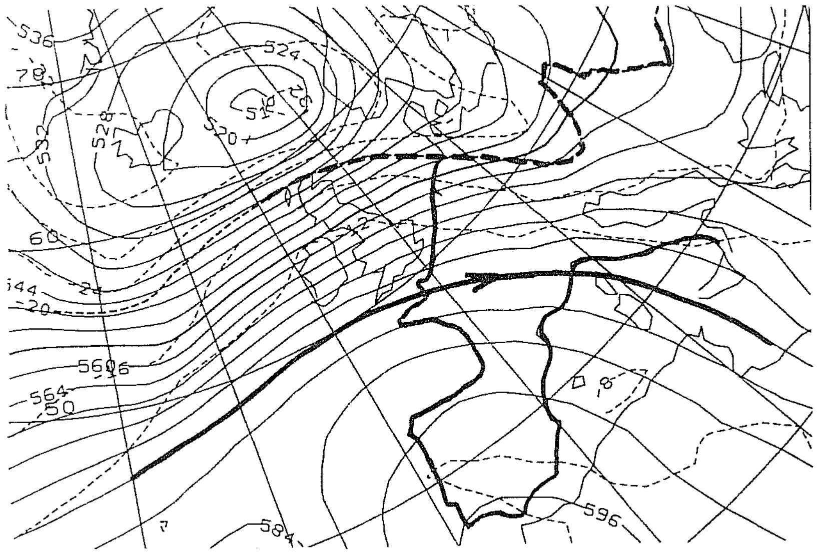


fig. 16b : analysis 03.10.93

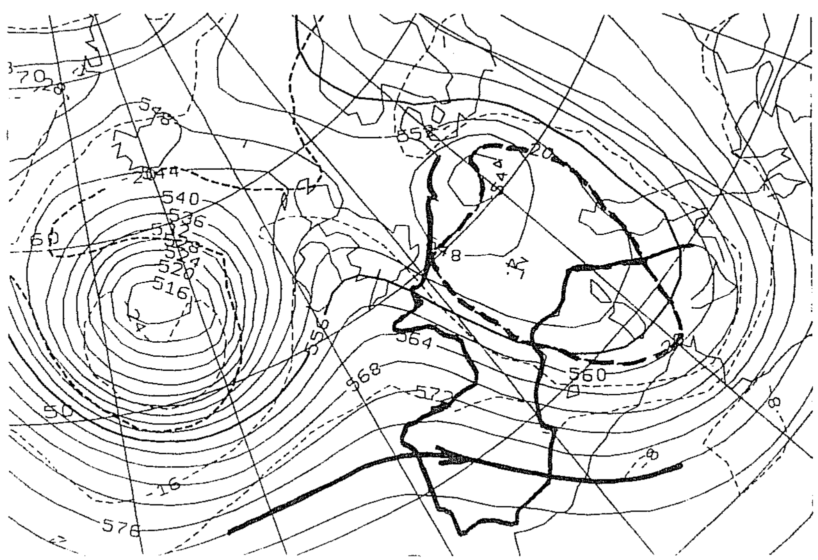


fig. 15 : EF +168h from 26.09.93 valid 03.10.93 clusters 1 2 3 4 5 6

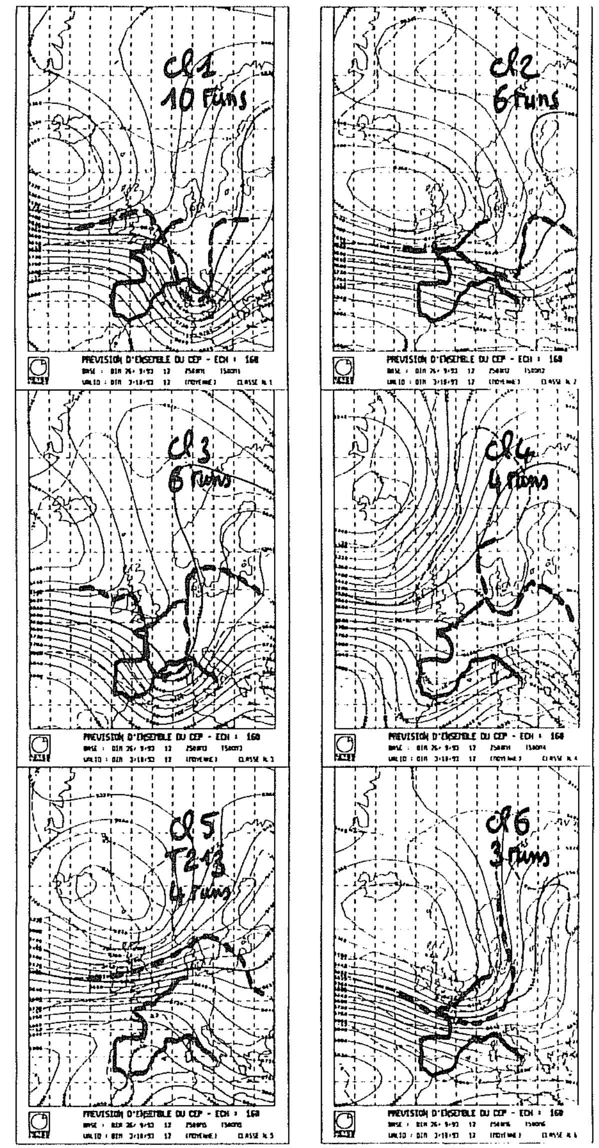


fig. 20 : EF from 26.09.93 ECMWF's interdistance matrix and ECMWF's clusters

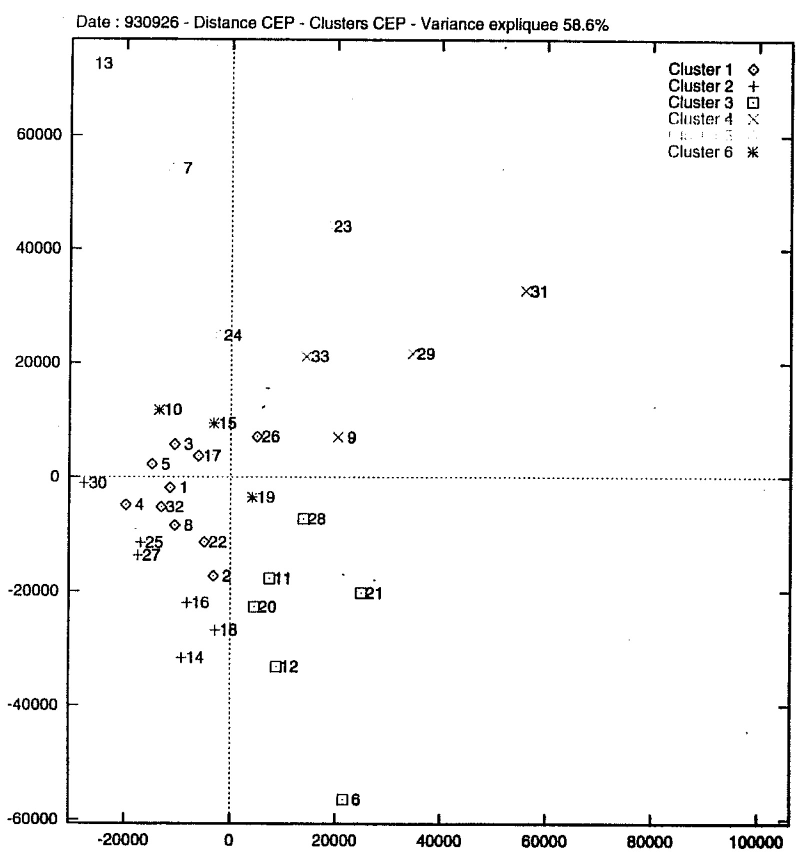


fig. 21 : EF from 26.09.93 METEO-FRANCE's interdidtance matrix and ECMWF's clusters