TECHNICAL MEMORANDUM

317

Classifications and validation of the ECMWF EPS perturbed forecasts using pre-defined weather regimes

Piero A. Chessa

Operations Department

October 2000

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

Classifications and validation of the ECMWF EPS perturbed forecasts using pre-defined weather regimes

Piero A. Chessa

European Centre for Medium-Range Weather Forecasts, Reading, Berkshire, England

October 2000

Abstract

ABSTRACT An alternative method to classify the perturbed forecasts of the Ensemble Prediction System (EPS), issued daily by the ECMWF, is presented. A fixed number of pre-defined patterns for a chosen area and a certain meteorological variable have been determined and the EPS members (and any other forecast and analysis available) have been assigned to the closest pattern for each forecast time step. Form the resulting homogeneous group empirical forecast probabilities have been computed. This approach allows an appropriate clustering for each time step, a consistent comparison with other methods like the so called poor man's ensembles and a consistent verification of probabilistic forecasts. Deciding how many flow-patterns have to be used is a key issue and it is strongly user-dependent. The non-hierarchical algorithm applied here, although effective and reliable, has to be considered as one of many possible options.

1. Introduction

ECMWF is producing Ensemble forecasts since 1992 and many operational applications have been devised in a number of Meteorological Services across Europe. ECMWF also supplies several by-products, like clusters and tubes, as guidance for further applications. Nonetheless probabilistic forecasts made by using EPS are still relatively rare and even rarer are their verifications. Moreover, while several studies about the EPS performance have been accomplished (see e.g. *Richardson* 2000, *Atger* 1999a, *Buizza* 1997, *Molteni et al* 1996, *Strauss and Lanzinger* 1995), applications involving weather regimes are still relatively infrequent (see e.g. *Chessa*, 1999b and c and reference therein). Though many targeted application now directly focus on weather parameters, the use of alternative synoptic scenarios is as valuable as the direct information on variables as precipitation or 2m temperature, especially for long term forecasts. Yet, products like clusters or tubes (*Atger* 1999b, *Molteni* 1996) are often used without any underlying verification and this because of the very nature of their set up. For instance, clustering is accomplished using the Ward algorithm applied to the 500 hPa geopotential height for the 120h to 168h forecasts considered together. The clusters are kept fixed through all the forecast time range and for all the other parameters. This technique might create inconsistencies for time steps not close to those used for the aggregation and even simple verifications (like, for instance, determining how many times the most populated cluster verifies) face subtle problems due to the daily change of synoptic scenarios.

Some problems of the present EPS classification can be overcome using predefined patterns to which assign the perturbed forecasts for every single time step. Then elements belonging to the same regime can be clustered , for instance using a suitable threshold (e.g. 10% or 5 members with the present EPS configuration). In this way, non-classifiable members can be ruled out and treated similarly to the tube-extremes (*Atger* 1999b). Moreover, probabilistic forecasts can be based either on the relative frequency at each time step or combining consecutive forecasts, all verifying at the same time, with weights depending on the consistency in time. Advantages for verifications and consistency checks have already been examined in two previous memoranda (*Chessa* 1999 a and b) and the use of a similar method in the Swiss Met-Service has proved valuable for operational purposes.

In this work the alternative method for an operational classification of the EPS is applied to the daily forecasts of the years 1997 to 2000. The approach used is similar to that implemented operationally in Switzerland (*Eckert et al.* 1996) but relies on a non-hierarchical recursive clustering method instead of an Artificial Neural Network (ANN). Formally the outcome of this method is the same as for the non-hierarchical clustering apart from the topological ordering typical of the Kohonen method used for ANN. On the other hand, clustering is definitely less expensive in terms of computation time and easier to implement and interpret.

In section 1 the clustering method is described and in section 2 an application to the ERA15 re-analysis (*Gibson* 1997) is performed. Section 3 illustrates the determination of the fixed flow patterns, while in section 4 some operational applications are reported. A preliminary statistical verification of EPS forecasts in terms of Brier scores and Brier skill scores is illustrates in section 5 and some conclusion is drawn in section 6.

2. The clustering algorithm

A large number of clustering algorithms has been used in many meteorological applications. All these method depend on the particular type of investigation that is carried out (*Gong and Richman* 1996, *Chessa et al.* 1999). For an overview of clustering algorithms and their applications see *Anderberg* 1973.

In this work a recursive non-hierarchical method has been applied which is known as Kmeans in the MacQueen convergent variant (*Anderberg* 1973). Some other constraint can be added to allow the number of clusters varying. The method, applied to a set of N data can be described in five steps:

i) Set an initial number of clusters k

ii) Take k random elements as initial clusters (seed points) of one element each.

iii) Assign the N-k remaining data to the nearest centroid (the average value of cluster's elements) upgrading the centroids after every assignment.

iv) After the partition is completed take the final k centroids as new seed points and, considering all the data in sequence, check whether they belong to the parent cluster. If not, re-assign the element and upgrade both the loosing and the gaining cluster.

v) Repeat the point 4 until the process converges and during a complete cycle no elements are moved.

While the convergence of this method can be formally proved (a hint of the very complex proof can be found in *Anderberg* 1973), a limitation is that the number of clusters has to be fixed a priori. This minor shortcoming can be partially overcome using some objective parameter to be maximised or minimised when the number of clusters is allow to vary. However, objective criteria have generally to be supported by subjective interventions that are useful to assess the effectiveness of the algorithm with regard to the particular aspects under investigation.

One advantage of non-hierarchical methods is the flexibility due to the recursive rearrangement. In fact the successive aggregation or partition typical of the hierarchical methods is more sensitive to the starting point. Instead, a possible dependence on the initial choice of the seed points can be easily checked in the non-hierarchical algorithms and it is in any case minimised by the recursive rearrangement.

Also important is the choice of the metrics used to calculate the distances among data units and centroids. For sake of simplicity an Euclidean metric (scaled with the cosine of latitude) has been adopted, but results have been tested also for a metric based on the correlation coefficient. It has to be born in mind that this could be not the case with other variables; for instance for precipitation the differences can be noticeable (see e.g. *Chessa et al.* 1999).

3. The synoptic patterns

The technique just described has been applied to the daily re-analysis taken from the ERA15 data for 1979 to 1993. The study has been focused on the 500 hPa geopotential height but the technique used can be applied to different parameters and levels. For every day four analyses (at 00, 06, 12 and 18 UTC) have been considered and, overall, 21916 fields have been used.

The area selected encompasses the Euro-Atlantic region (30N to 76N and 76W to 40E) and the grid points are taken every two degrees both in latitude and in longitude. It means that each field contains 1416 point. The size of the area selected has been chosen in order to take into account a region dynamically and synoptically consistent with the forecast range considered.

Clusters have been computed considering the field anomalies with respect to the monthly means calculated from the ERA15 sample, for the standard three-month seasons autumn (SON), winter (DJF), spring (MAM) and summer (JJA). Same computations can be easily applied to longer seasons as the periods from October to March (extended winter) and from April to September (extended summer).

Initially the K-means algorithm has been applied allowing the variation of the number of clusters in order to have an appreciation of possible natural data partitions. The number of clusters generally stabilised around 5 to 9 without showing any strong dependence on the starting point. Eventually, a subjective evaluation of the different combinations brought to the choice of 6 clusters whose centroids both for anomalies and full fields are shown in figs 1 to 8 for the standard seasons. It is important to note that the order among clusters in the different seasons is not the same. This is a consequence of the K-means random sampling in the initial step. Ordering them a posteriori is not straightforward since the regime structure changes from one season to the other.

The aggregation has been operated without any underlying dynamical assumption and the raw data have been used without any filtering (see e.g. *Vautard* 1990). Nonetheless it is interesting to note that when a partition with four winter clusters is considered, their centroids are very close to the quasi-stationary regimes found by *Vautard* (1990) for the 700 hPa geopotential height using an utterly different data set and for a longer winter season. This can be seen for instance comparing figs 9 and 10 to the four quasi-stationary regimes described in that paper. When the number of clusters is allowed to increase the K-means finds some extra-regimes (not shown) that correspond to those Vautard called preferred precursor and successor of the four quasi-stationary patterns. A common feature among the four regimes is their relation with the North Atlantic Oscillation

(*Wallace and Gutzler*, 1981) as thoroughly described in *Vautard*, 1990. This gives some confidence in the dynamical meaning of the used flow patterns and confirms the ability of this non-hierarchical algorithm of extracting sensible information from meteorological data-set.

4. Operational applications

The classification of the daily EPS and of other available forecasts (and analyses) has been accomplished calculating the distance of each element from the cluster's centroids and assigning it to the closest one. In this classification, like for the determination of the fixed regimes, anomalies with respect to the climatological monthly means have been used.

After the classification is completed for a certain time step, the average of the EPS members belonging to the same cluster is calculated. To obtain a significant mean it is preferable to set a minimum number of elements all assigned to the same cluster In this work this number has been set to 5, which corresponds to 10% of the present EPS members. When less than 5 elements are in the same cluster they are considered as non-classifiable and treated like the tube extremes (*Atger* 1999b). Forecast probabilities for each regime represented by the fixed clusters can be calculated using the empirical frequencies.

Deterministic forecasts as well as analyses can be assigned to the closest regime. This also allows simple ensembles to be worked out using models from different Meteorological Centres. In this work this has been accomplished when DWD, UKMO, NCEP and ECMWF (T319) models were all available assigning to each of them a 25% probability.

The EPS classification through fixed pattern can be used for several operational applications. A simple table can be produced daily to monitor the distribution of the EPS and deterministic forecasts for each forecast time step. In Table 1 an example for the 16th of March 2000 is reported. The columns refer to the fixed pattern, as reported in figs 5 and 6 for spring, and the rows represent forecasts from different sources. Figures in italic represent the number of members in each EPS cluster and the corresponding synoptic patterns as reported in fig 11 for the 120h to 168h time steps (interval 24h).

The fixed reference frame for weather patterns can also be used for consistency checks among consecutive forecasts. For example probabilistic forecasts for the medium-range, obtained employing the Ensemble Prediction System, would benefit from a careful analysis of the consistency between consecutive forecasts all verifying at the same time. In fact, if the system is working properly, while approaching the verification time the cluster with the right solution will most likely increase its population. This property can be used to calculate a different probability composed by a suitable (weighted) combination of the probabilities coming from a few consecutive EPS forecasts.

Two examples for a 96h forecast are reported in Table 2 while in figs 12 and 13 the corresponding cluster means and the verifying analysis are represented. In the first case (verification on the 2000-06-20) looking only to the last forecast (96h) there is quite a clear signal towards the right solution (the sixth cluster), but a non negligible probability is associated to the fourth cluster as well. However, more confidence can be attributed to the corresponding probabilistic forecast if the forecast evolution is taken into account. In fact the sixth cluster is the only one whose population increases consistently in the last two time steps. For the same case the ECMWF high-resolution model gave less consistent information from day to day (see fig. 14) and the support from the EPS would have been very useful. The second case is even clearer and ends up with an almost deterministic outcome.

Warnings about significant or extreme events can also be obtained, especially if surface parameters are considered. These warnings can be attained by checking if the forecast to be classified has a distance from the cluster centroid greater than the most far away among the cluster elements. In that case it can be inferred that the pattern examined is liable to be rare or even extreme. This kind of classification is of course dependent on the partition chosen.

5. Verification of probabilistic forecasts

5.1 Probabilistic Scores

Probabilistic forecast for the fixed clusters have been verified using the Brier Score and Brier Skill Score (*Wilks* 1995, *Brier* 1950) and an index related to the cluster's population. As well known the Brier Score is defined as

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2$$

where p_i is the probability assigned to a certain regime and o_i is 1 if the analysis belongs to the same regime and 0 otherwise. The *BS* is negatively oriented and the lower it is the better the forecast.

The Brier Skill Score is instead defined as

$$BSS = 1 - \frac{BS_{for}}{BS_{ref}}$$

Here BS_{for} is the Brier Score for the forecast to be verified and BS_{ref} is for the forecast used as reference. The closer *BSS* to 1 the better the gain obtainable with respect to the reference forecast (see *Wilks* 1995 for more details). Negative values mean that the forecast tested is worse than the reference.

Another verification measure has been used in the present analysis. It can be called Cluster Score (CS) and pertains to the number of times the analyses verify in the most populated EPS daily cluster.

A statistical assessment has been carried out for the 72, 120, 168 and 240 hours time ranges for the years 1997 to 2000 (up to August). In the following discussion the 120h step will be examined in detail while only a summary will be provided for the other forecast time steps.

As for the 120h forecast data for the NCEP model, used in the already mentioned Poor Man Ensemble (PME), the 00UTC run and then the 132h forecast have been used. In terms of the scores just described, the EPS (composed by the 50 perturbed forecasts) has been compared with PME, the Ensemble mean (EME), the EPS unperturbed forecast (CTE) and the sample climate (CLI). Such comparisons are statistically awkward since two systems, CTE and EME, are forced to be probabilistic with probabilities that can attain only zero or one.

On the other hand the PME is an ensemble built on much less elements than the EPS and this has some implication for a fair juxtaposition of the systems (see for instance *Ziehmann*, 2000). However, though problematic this assessment is, it is of primarily importance to gain some insight about its operational values. As a matter of fact the EPS has to be evaluated against the available alternatives that often are, especially from a statistical point of view, very different from it.

All calculations have been accomplished without any distinction among regimes in order to rely on a bigger statistical sample. It will be anyway valuable to make this distinction as soon as a significant sample will become available.

5.2 Verification Results

Let us consider the 120h forecast lead time first. The corresponding Brier Skill Scores of EPS versus PME and the Cluster Score for EPS are reported in table 3 and 4 for every season and the whole year for 1997 to 2000. Summer scores for the current year refer to June and July only. Tables 5 to 8 report the BSS for EPS and PME when compared with EME, CTE and CLI. For EPS the BSS with respect to PME is reported again to facilitate comparison with the other values.

Table 3 clearly shows that PME regularly outperforms EPS in summer. Another clear fact is the bad EPS performance in 1998 with the exception of the Spring which, vice-versa, has got a negative BSS for 2000. Apparently there are no indication for positive or negative trends and the values oscillate from year to year. This seems to point more to a flow dependent behaviour rather than to some specific differences among the ensembles (as, for instance, number of members, resolution, etc.).

Tables 5 to 8 indicate that from 1997 to 1999 both EPS and PME had a positive trend in summer with respect to the other reference forecasts, but PME went at a faster pace. Unfortunately, the situation for the first two months of the summer 2000 is worse. The performance of both EPS and PME versus the Ensemble mean plunged becoming even negative for EPS. Bad periods for T319, like spring and summer 1999, have got different characters also in probabilistic terms. In fact in that particular spring there seems to have been a clear link to predictability issues as the BSS of EPS versus CLIM were much lower with respect to the other years, whilst in summer they were comparable. Similar information but of opposite sign can be found for the last winter (1999-2000) when the scores of the deterministic forecast were excellent.

Table 4, which refers to the CL score, gives similar indications and normal situations are apparently characterised by values close to 70% while very predictable (unpredictable) situations have higher (lower) values. It is interesting to note that the average population of the most populated cluster (number in parenthesis) roughly corresponds to the frequency of occurrence, thus giving a measure of the system's reliability. Now, for spring 1999, reliability seems quite poor while no remarkable differences show up for the same year's Summer. This could be taken as a further indication on the possible different nature of those problematic periods. Similar considerations applies to 1998 when only in Spring EPS had positive Brier Skill Scores.

In Table 9 to 12 a summary for all time steps analysed is reported considering the entire period from 1997 to August 2000 and making distinction only among seasons. The evolution of the Brier Score with the forecast time is showed in Table 9. Table 10 and 11 report the Brier Skill Scores versus the PME and the sample climate respectively. The former can be assessed only for the 72 and 120 hours forecasts since the DWD, UKMO and NCEP models are not available for later time steps. Finally Table 12 reports the Cluster Score.

There is a clear degradation in time of the Brier Scores and the BSS versus climate, but the EPS usefulness appears to be quite valuable also in the late medium range (ten days forecast). Not so clear cut is the situation when PME is considered. As a matter of fact the gain of EPS with respect to the used PME is quite limited and apparently very much flow dependent since it strongly varies with the season.

As for the Cluster score, Table 12, it shows the same negative trend in time and an increasing loss of reliability as was to be expected due to the general high EPS spread in the late medium range.

6. Conclusion

An alternative method to classify and verify the daily outcome of the ECMWF Ensemble Prediction System has been described in this work. Some of the advantages of using it in a operational environment have been highlighted and a preliminary verification of the last four year has been carried out in terms of Brier Score, Skill Score and a simple index related to the most populated EPS cluster.

On of the most significant part of the verification is the comparison, for the 120 hour forecast, of the EPS with a poor man's ensemble composed by four deterministic models. Results indicate that summer is in general a critical period for EPS. Similar considerations can be draw for Autumn and Winter 1998 and for Spring 2000. In the other seasons and for the 72 hours forecast too, the EPS has been only marginally better than PME.

However, these results have to be taken with a pinch of salt and put into a wider context. In fact the two ensembles have a remarkable difference in resolution and size. The latter allows much weight to be put onto single forecasts in the PME. Other combinations should be tested in order to have a fair comparison, but they are likely to bring to a never ending story where each time a suitable combination scoring better than EPS can be found. In fact this does not give any real information about a sensible alternative to the Ensemble Prediction System. Difference in size has another implication since small ensembles are not suitable for application where a refined set of probability classes is needed. Applications to decision making processes, whether or not they involve economical gain or losses, benefit more from systems like EPS that are able to sample forecast distributions better than ensembles with just a few members (e.g., *Richardson*, 2000).

Another point is that in this work only Brier Scores have been considered and they account just for a specific and limited part of the quality of probabilistic forecasts. More complete conclusion can be draw only when an extensive set of verifications, investigating different aspects of the EPS forecast, will be carried out.

The purpose of the analysis presented in this paper was to give some preliminary indications about a possible alternative to the operational ECMWF clustering. For this reason both the operational applications and the verification were aimed to prove the usefulness of the method. Therefore this paper describes only some preliminary results and it is believed that deeper and more extensive analysis should be accomplished in order to properly use and evaluate the Ensemble Prediction System.

The entire forecast range should be considered and possible flow-dependent model-errors should be investigated in connection with the regimes actually used for the classification. Applications to longer forecast ranges, as for instance those going to produced with a coupled model, should be worked out. A critical analysis about the optimal number of pre-defined regimes should be accomplished and application to important parameters like precipitation would result beneficial for an extension of the EPS use in operational contexts.

References

Anderberg, M. R. 1973. Cluster Analysis for Applications. Academic Press, 359 pp. Atger, F., 1999a. The Skill of Ensemble Prediction Systems. *Mon. Wea. Rev.*, **127**, 1941-1953.

Atger, F., 1999b. Tubing: An Alternative to Clustering for the Classification of Ensemble Forecasts. W*eather and Forecasting*, **14**, 741-757.

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

Buizza, R., 1997. Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction System. *Mon. Wea. Rev.*, **126**, 2503-2518.

Cacciamani, C., Nanni, S., Tibaldi, S. 1994. Mesoclimatology of winter temperature and precipitation in the Po valley of Northern Italy. *J. Climatol.,* **14**, 777-814.

Chessa, P.A., Cesari, D., Delitala, A.M. 1999a. Mesoscale precipitation and temperature Regimes in Sardinia (Italy) and the related synoptic circulations. *Theor. Appl. Climatol.*, **63**, 195-222.

Chessa, P. A., 1999b. Verification of the probabilistic forecast of synoptic flow patterns by means the ECMWF Ensemble Prediction System. *ECMWF Operation Department Memorandum.*

Chessa, P. A., 1999c. Preliminary result on the EPS performances using fixed synoptic patterns obtained through of a neural network. *ECMWF Memorandum.*

Eckert, P., Cattani, D. and Ambuhl, J. 1996: Classification of ensemble forecast by means of an artificial neural network. *Meteorol. Appl*, **3**, 169-178.

Gibson, J.K., et al. 1997. ECMWF Re-Analysis. 1. ERA description. *ECMWF Re-analysis Project Report Series,* pp 72.

Gong, X.-F., and M.B. Richman, 1995: An examination of cluster analysis methodology in geophysical research. Part I: Precipitation data. *J. Climate*, **8**, 897-931.

Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis T. 1996. The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119.

Richardson, D. S, 2000. Skill and economic value of the ECMWF ensemble prediction system, *Quart. J. Roy. Meteorol. Soc.*, **126**, 649-668.

Strauss, B. and Lanzinger, A., 1995, Validation of the ECMWF Ensemble Prediction System. *ECMWF seminar proceedings on Predictability*, Vol 2. Reading, U.K, ECMWF, pp 157-166.

Vautard, R. 1990. Multiple weather Regimes over the north Atlantic: Analysis of precursors and successors. *Mon. Wea. Rev.*, **118**, 2056-2081.

Wallace,J.M. and D.S. Gutzler, 1981. Teleconnection in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784-812

Wilks, D.S., 1995. Statistical methods in the Atmospheric Sciences. Academic Press, 467 pp.

Ziehmann, C. 2000. Comparison of a single-model EPS with a multimodel ensemble consisting of a few operational models. *Tellus*, **52A**, 280-299.

Regime	1	2	3	4	5	6	1	2	3	4	5	6
	48h						**72h**					
CONTROL	-	-	1	-	-	-	-	-	1	-	-	-
ECMWF(T319)	-	-	1	-	-	-	-	-	1	-	-	-
EPS mean	-	-	1	-	-	-	-	-	-	1	-	-
DWD	-	-	1	-	-	-	-	-	1	-	-	-
UKMO	-	-	1	-	-	-	-	-	1	-	-	-
NCEP	-	-	1	-	-	-	-	-	1	-	-	-
EPS	-	-	*47*	3	-	-	-	-	*23*	*24*	-	3
	96h						**120h**					
CONTROL	-	-	-	1	-	-	-	-	-	-	-	1
ECMWF(T319)	-	-	-	-	-	1	-	-	-	-	-	1
EPS mean	-	-	-	1	-	-	-	-	-	-	-	1
DWD	-	-	1	-	-	-	-	-	-	-	-	1
UKMO	-	-	1	-	-	-	-	-	-	-	-	1
NCEP	-	-	-	-	-	1	-	-	-	-	-	1
EPS	-	-	4	*27*	1	*18*	3	-	1	*18*	1	*27*
	144h						**168h**					
CONTROL	-	-	-	-	-	1	-	-	-	-	-	1
ECMWF(T319)	-	-	-	-	-	1	-	-	-	-	-	1
EPS mean	-	-	-	-	-	1	-	-	-	-	-	1
DWD	-	-	-	-	-	1	-	-	-	-	-	1
UKMO	1	-	-	-	-	-	-	-	-	-	-	-
NCEP	-	-	-	-	-	-	-	-	-	-	-	-
EPS	*5*	-	-	*15*	1	*29*	*6*	-	1	*9*	-	*34*
	192h						**216h**					
CONTROL	-	-	-	-	-	1	-	-	-	-	-	1
ECMWF(T319)	-	-	-	-	-	1	-	-	-	-	-	1
EPS mean	-	-	-	-	-	1	-	-	-	-	-	1
DWD	-	-	-	-	-	-	-	-	-	-	-	-
UKMO	-	-	-	-	-	-	-	-	-	-	-	-
NCEP	-	-	-	-	-	-	-	-	-	-	-	-
EPS	*5*	*7*	1	*5*	-	*32*	*6*	*8*	1	4	2	*29*
	240h											
CONTROL	-	-	-	-	-	1						
ECMWF(T319)	-	-	-	-	-	1						
EPS mean	-	-	-	-	-	1						
DWD	-	-	-	-	-	-						
UKMO	-	-	-	-	-	-						
NCEP	-	-	-	-	-	-						
EPS	*9*	*8*	3	3	*6*	*21*						

Table 1: Ten days forecasts issued on the 16th of March 2000. The 2nd and 3rd column report the time steps every 24h. The forecast considered are reported in each row in the 1st column. For the EPS the number indicated in italic refer to the groups taken as clusters (that is with more than 4 elements). In each cell the six columns refer to the regimes as reported in Fig 5 and 6.

Regime	Forecasts verifying on the 20th of June 2000						Forecasts verifying on the 22nd of March 2000					
	1	2	3	4	5	6	1	2	3	4	5	6
	168h						**168h**					
CONTROL	-	-	-	-	-	1	-	-	-	-	-	1
ECMWF(T319)	-	-	-	1	-	-	-	-	-	-	-	1
EPS mean	-	-	-	-	-	1	-	-	-	-	-	1
DWD	-	-	1	-	-	-	-	-	-	-	1	-
UKMO	-	-	-	-	-	-	-	-	-	-	-	-
NCEP	-	-	-	-	-	-	-	-	-	-	-	-
ANALYSIS	-	-	-	-	-	1	-	-	-	-	-	1
EPS	1	*15*	4	*10*	*5*	*15*	1	1	2	*13*	*5*	*28*
	144h						**144h**					
CONTROL	-	-	-	-	-	1	-	-	-	-	-	1
ECMWF(T319)	-	1	-	-	-	-	-	-	-	-	-	1
EPS mean	-	-	-	-	-	1	-	-	-	-	-	1
DWD	-	1	-	-	-	-	-	-	-	-	-	1
UKMO	-	1	-	-	-	-	1	-	-	-	-	-
NCEP	-	-	-	-	-	-	-	-	-	-	-	-
ANALYSIS	-	-	-	-	-	1	-	-	-	-	-	1
EPS	1	*15*	*10*	*9*	1	*14*	*5*	-	-	*15*	1	*29*
	120h						**120h**					
CONTROL	-	-	-	-	-	1	-	-	-	-	-	1
ECMWF(T319)	-	-	-	1	-	-	-	-	-	-	-	1
EPS mean	-	-	-	-	-	1	-	-	-	-	-	1
DWD	-	-	-	-	-	1	-	-	-	-	-	1
UKMO	-	-	-	-	-	1	-	-	-	-	-	1
NCEP	-	-	-	-	-	1	1	-	-	-	-	-
ANALYSIS	-	-	-	-	-	1	-	-	-	-	-	1
EPS	1	*5*	*7*	*14*	-	*23*	3	1	-	1	-	*45*
	96h						**96h**					
CONTROL	-	-	-	-	-	1	-	-	-	-	-	1
ECMWF(T319)	-	-	-	-	-	1	-	-	-	-	-	1
EPS mean	-	-	-	-	-	1	-	-	-	-	-	1
DWD	-	-	-	-	-	1	-	-	-	-	-	1
UKMO	-	-	-	-	-	1	1	-	-	-	-	-
NCEP	-	-	-	-	-	1	-	-	-	-	-	1
ANALYSIS	-	-	-	-	-	1	-	-	-	-	-	1
EPS	-	1	3	*12*	-	*34*	*7*	-	-	-	-	*43*

Table 2: As in Table 1, but in the 2nd and 3rd column and from the 2nd row a series of forecasts issued on consecutive days, all verifying on the day indicated in the 1st row, are reported.

Forecast Range: 120h					
Year	**Autumn**	**Winter**	**Spring**	**Summer**	**Whole year**
1997	0.160	0.097	0.136	-0.018	0.092
1998	-0.029	-0.040	0.114	-0.076	-0.007
1999	0.056	0.016	0.113	-0.120	0.032
2000		0.082	-0.044	-0.075	-0.013

Table 3: Brier Skill Score of EPS versus PME (see text for details) for the years from 1997 to 2000 and forecast time step 120h. For year the 2000 autumn is not available.

Forecast Range: 120h					
Year	**Autumn**	**Winter**	**Spring**	**Summer**	**Whole year**
1997	0.74 *(0.74)*	0.69 *(0.73)*	0.75 *(0.72)*	0.68 *(0.71)*	0.72 *(0.73)*
1998	0.65 *(0.72)*	0.68 *(0.70)*	0.72 *(0.80)*	0.63 *(0.71)*	0.67 *(0.73)*
1999	0.77 *(0.74)*	0.73 (0.72)	0.53 (0.68)	0.67 (0.67)	0.67 (0.70)
2000		0.82 *(0.78)*	0.63 *(0.66)*	0.77 *(0.71)*	0.74 *(0.72)*

Table 4: Cluster Score for EPS (see text for details) for the years from 1997 to 2000. For year the 2000 autumn is not available. Between parentheses the corresponding relative frequencies are reported.

AUTUMN Forecast Range: 120h			
Brier Skill Score	**1997**	**1998**	**1999**
EPS vs PME	0.160	-0.029	0.056
EPS vs EME	0.325	0.358	0.261
EPS vs CTE	0.273	0.298	0.379
EPS vs CLI	0.510	0.416	0.584
PME vs EME	0.196	0.377	0.217
PME vs CTE	0.135	0.318	0.342
PME vd CLI	0.417	0.432	0.560

Table 5: Seasonal Brier Skill Scores of EPS and PME, versus the other forecasts used as a reference. In the 2nd row the BSS of EPS vs PME is reported again for a sake of simplicity. As for the other forecasts CTE stands for Control, EME for Ensemble mean and CLI for Climate. For the year 2000 autumn is not available and summer refers to June and July only.

WINTER Forecast Range: 120h				
Brier Skill Score	**1997**	**1998**	**1999**	**2000**
EPS vs PME	0.097	-0.040	0.016	0.082
EPS vs EME	0.281	0.243	0.134	0.229
EPS vs CTE	0.331	0.269	0.283	0.274
EPS vs CLI	0.488	0.435	0.429	0.668
PME vs EME	0.204	0.272	0.120	0.160
PME vs CTE	0.259	0.297	0.272	0.210
PME vd CLI	0.433	0.457	0.419	0.638

Table 6: Seasonal Brier Skill Scores of EPS and PME, versus the other forecasts used as a reference. In the 2nd row the BSS of EPS vs PME is reported again for a sake of simplicity. As for the other forecasts CTE stands for Control, EME for Ensemble mean and CLI for Climate. For the year 2000 autumn is not available and summer refers to June and July only.

SPRING Forecast Range: 120h				
Brier Skill Score	**1997**	**1998**	**1999**	**2000**
EPS vs PME	0.136	0.114	0.113	-0.044
EPS vs EME	0.333	0.312	0.279	0.361
EPS vs CTE	0.441	0.259	0.331	0.341
EPS vs CLI	0.553	0.498	0.265	0.451
PME vs EME	0.228	0.223	0.188	0.388
PME vs CTE	0.353	0.163	0.246	0.369
PME vd CLI	0.483	0.433	0.172	0.474

Table 7: Seasonal Brier Skill Scores of EPS and PME, versus the other forecasts used as a reference. In the 2nd row the BSS of EPS vs PME is reported again for a sake of simplicity. As for the other forecasts CTE stands for Control, EME for Ensemble mean and CLI for Climate. For the year 2000 autumn is not available and summer refers to June and July only.

SUMMER Forecast Range: 120h				
Brier Skill Score	**1997**	**1998**	**1999**	**2000**
EPS vs PME	-0.018	-0.076	-0.120	-0.075
EPS vs EME	0.177	0.289	0.287	-0.002
EPS vs CTE	0.302	0.331	0.415	0.388
EPS vs CLI	0.396	0.424	0.491	0.538
PME vs EME	0.192	0.340	0.364	0.068
PME vs CTE	0.314	0.379	0.478	0.431
PME vd CLI	0.406	0.464	0.546	0.570

Table 8: Seasonal Brier Skill Scores of EPS and PME, versus the other forecasts used as a reference. In the 2nd row the BSS of EPS vs PME is reported again for a sake of simplicity. As for the other forecasts CTE stands for Control, EME for Ensemble mean and CLI for Climate. For the year 2000 autumn is not available and summer refers to June and July only.

Brier Score					
Forecast range	**Autumn**	**Winter**	**Spring**	**Summer**	**Whole year**
72h	0.094	0.037	0.041	0.040	0.050
120h	0.070	0.076	0.089	0.074	0.074
168h	0.102	0.114	0.118	0.130	0.105
240h	0.128	0.125	0.133	0.131	0.129

Table 9: Seasonal Brier Scores for the EPS for all period from 1997 to August 2000

Brier Skill Score vs Poor Man's Ensemble					
Forecast range	**Autumn**	**Winter**	**Spring**	**Summer**	**Whole year**
72h	0.040	0.076	-0.033	0.025	0.029
120h	0.061	0.025	0.119	-0.067	0.038

Table 10: Seasonal Brier Skill Scores for the EPS versus PME for all period from 1997 to August 2000 and for 72 and 120 hours forecast ranges.

Brier Skill Score vs Climate				
Forecast range	**Autumn**	**Winter**	**Spring**	**Summer**
72h	0.541	0.737	0.706	0.711
120h	0.502	0.461	0.440	0.435
168h	0.269	0.337	0.201	0.189
240h	0.084	0.099	0.047	0.065

Table 11: Seasonal Brier Skill Scores for the EPS versus the sample Climate for all period from 1997 to August 2000.

Cluster Score					
Forecast range	**Autumn**	**Winter**	**Spring**	**Summer**	**Whole year**
72h	0.86 *(0.84)*	0.84 *(0.83)*	0.81 *(0.83)*	0.82 *(0.84)*	0.83 *(0.84)*
120h	0.71 *(0.72)*	0.70 *(0.72)*	0.67 *(0.74)*	0.66 *(0.70)*	0.69 *(0.72)*
168h	0.52 *(0.59)*	0.59 *(0.63)*	0.49 *(0.62)*	0.45 *(0.57)*	0.51 *(0.60)*
240h	0.35 *(0.44)*	0.41 *(0.49)*	0.35 *(0.47)*	0.33 *(0.43)*	0.36 *(0.46)*

Table 12: As in Table 4 but for different time steps and all period from 1997 to August 2000.

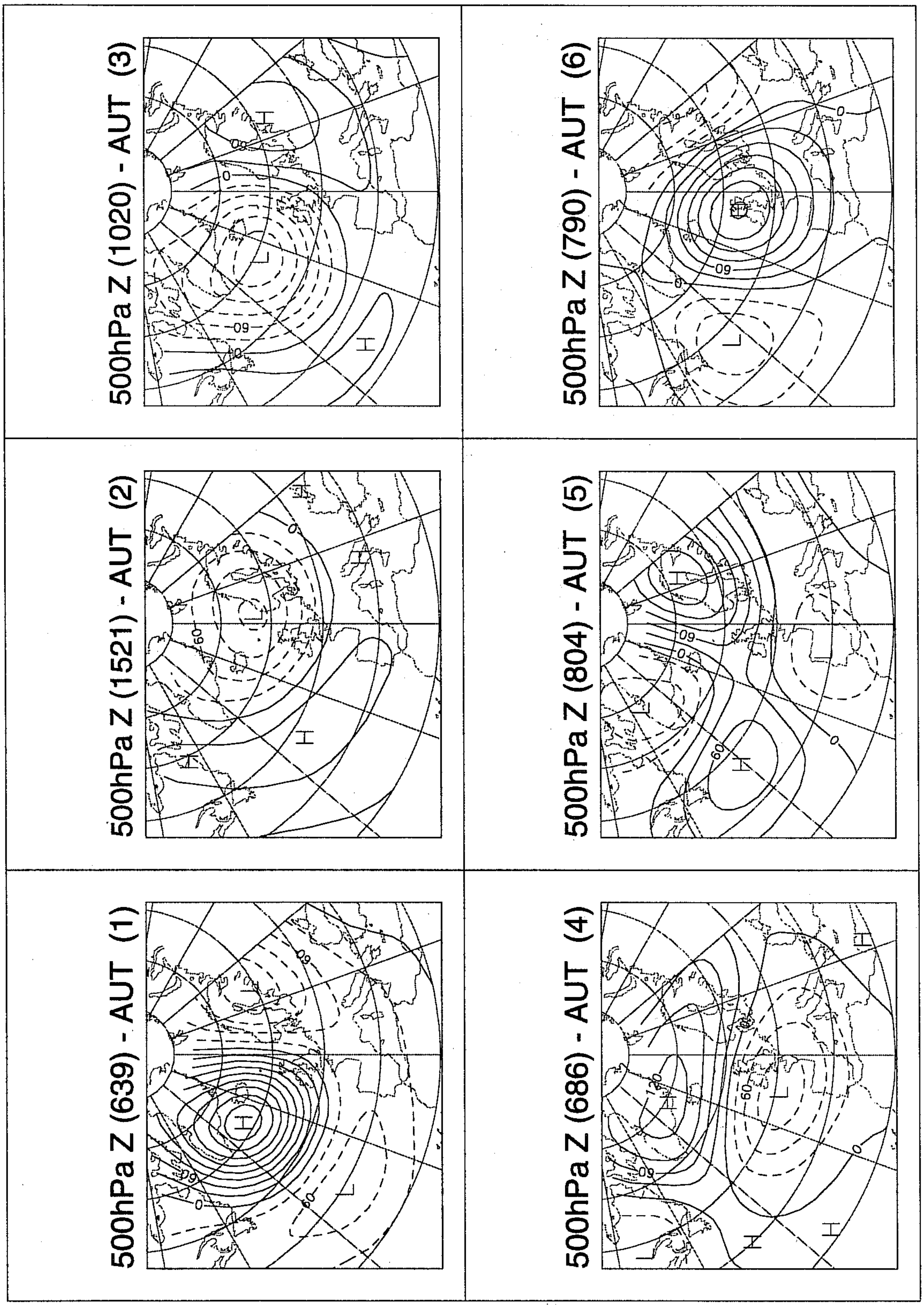

Figure 1: The six regimes used for the classification. Both the anomalies with respect to the climatological monthly means and the corresponding full fields are reported. Each panel reports the level and the parameter used, the number (between parenthesis) of analysis used to obtain the cluster and the season. *au* stands for autumn, *wi* for winter, *sp* for spring and *su* for summer. The regimes are ordered from the left to the right, starting from the upper row

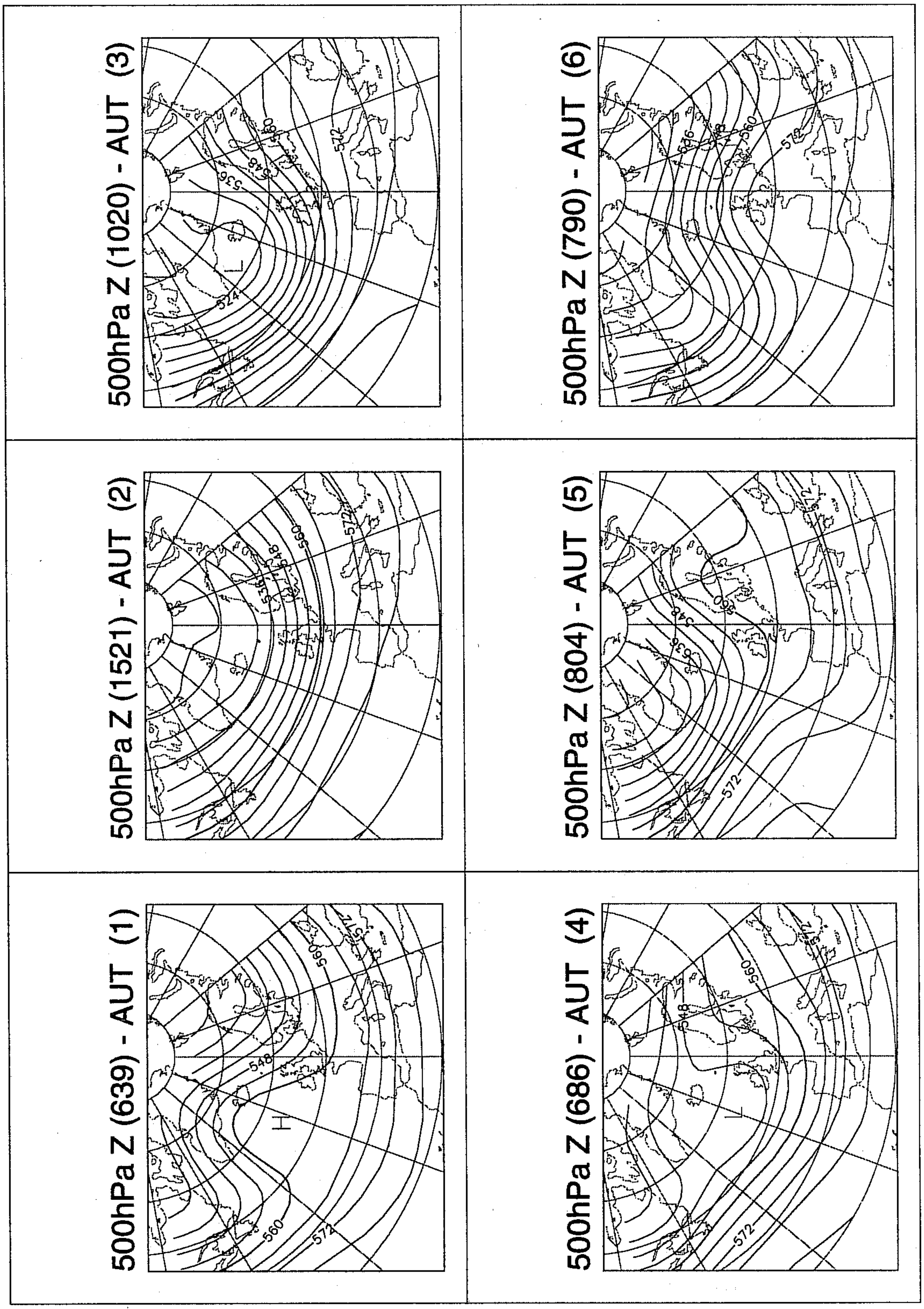

Figure 2: See Fig 1

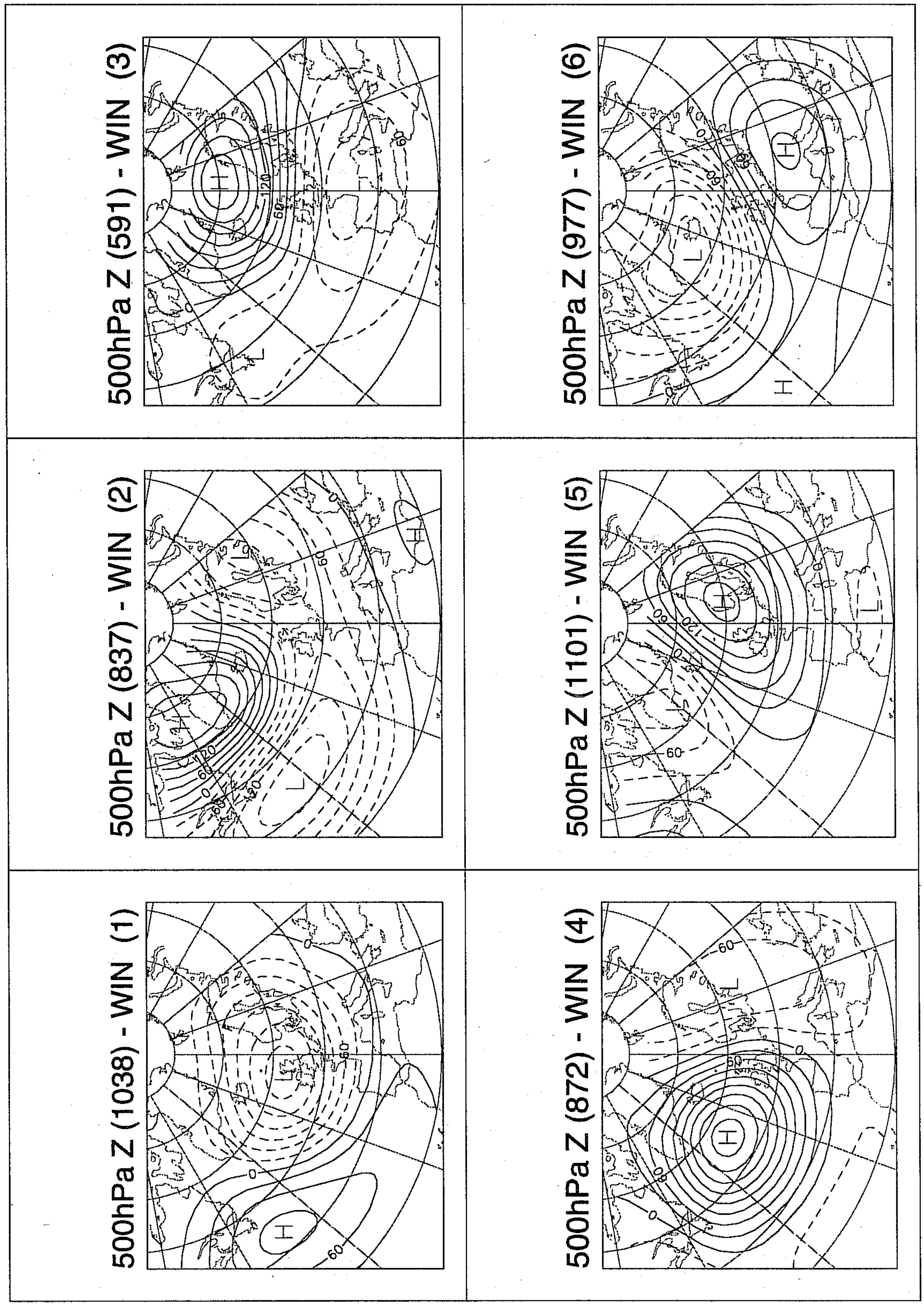

Figure 3: See Fig 1

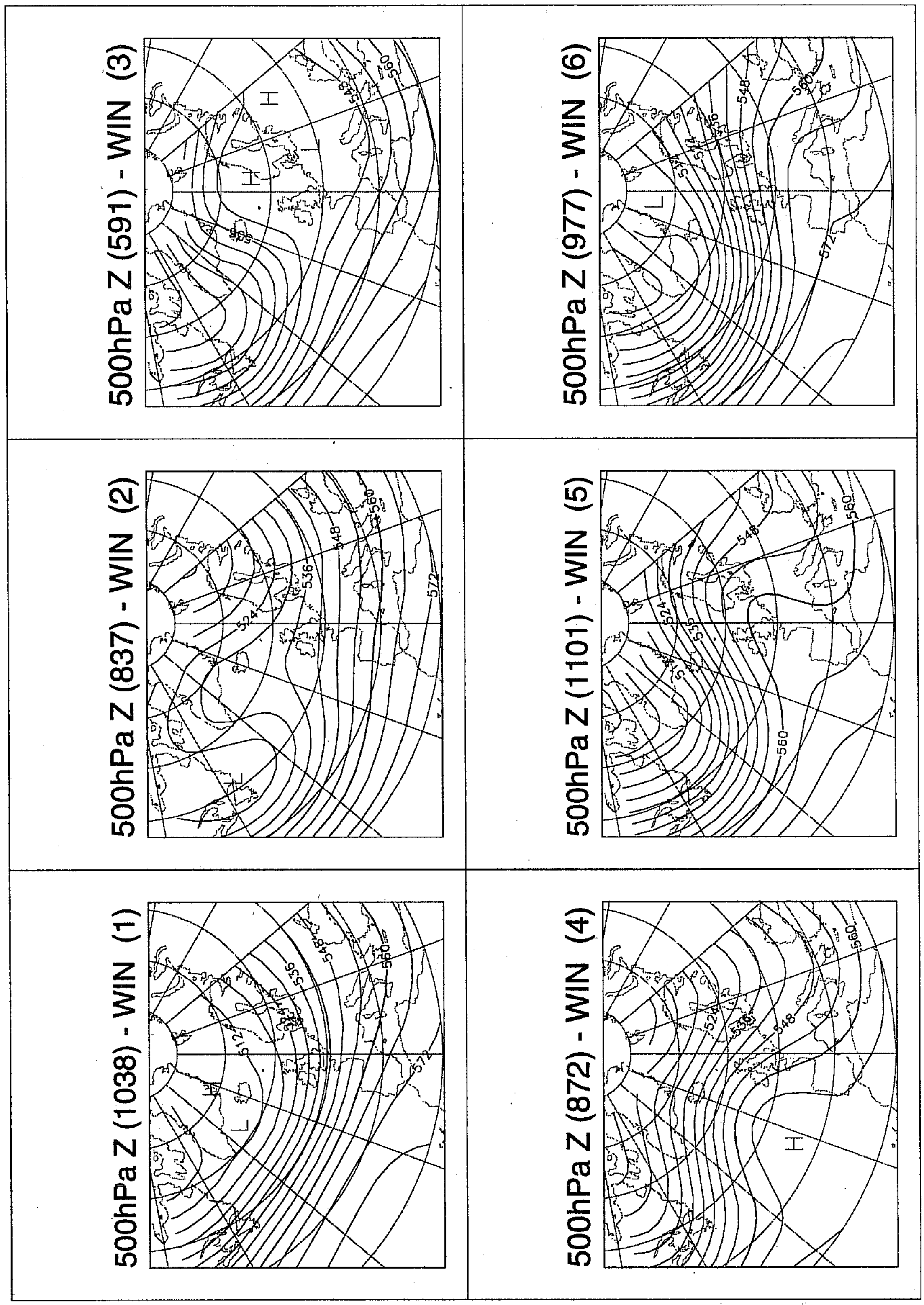

Figure 4: See Fig 1

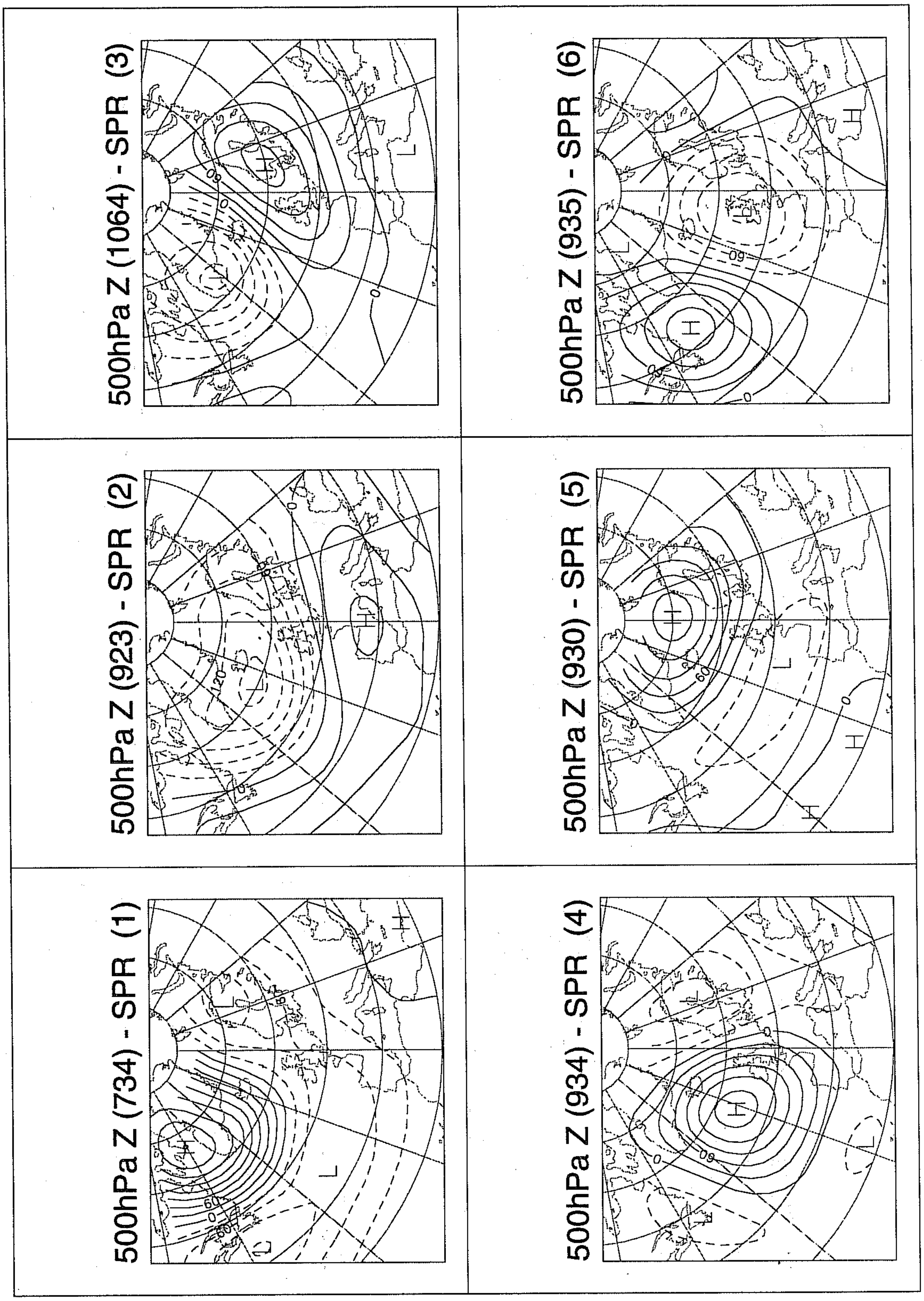

Figure 5: See Fig 1

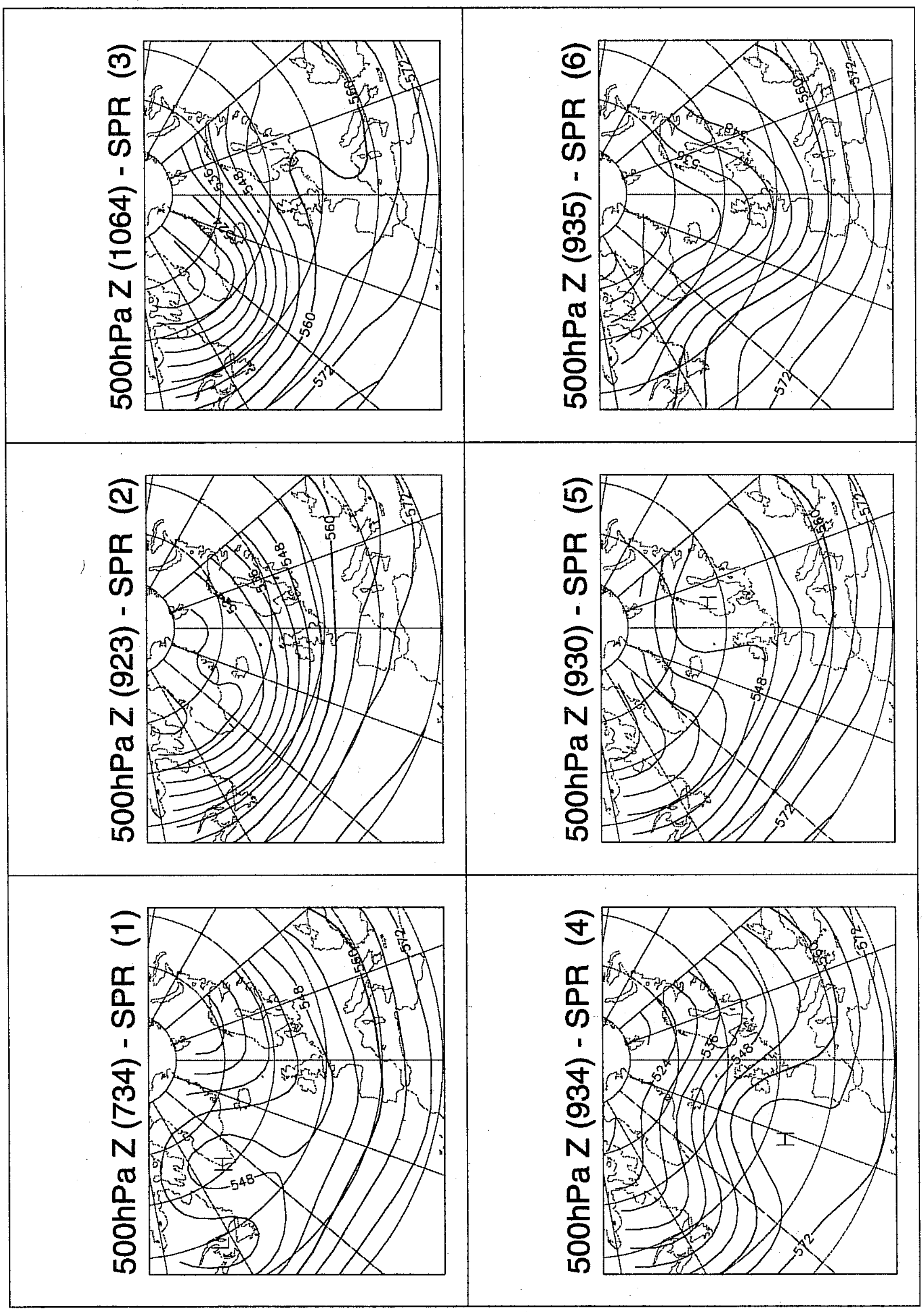

Figure 6: See Fig 1

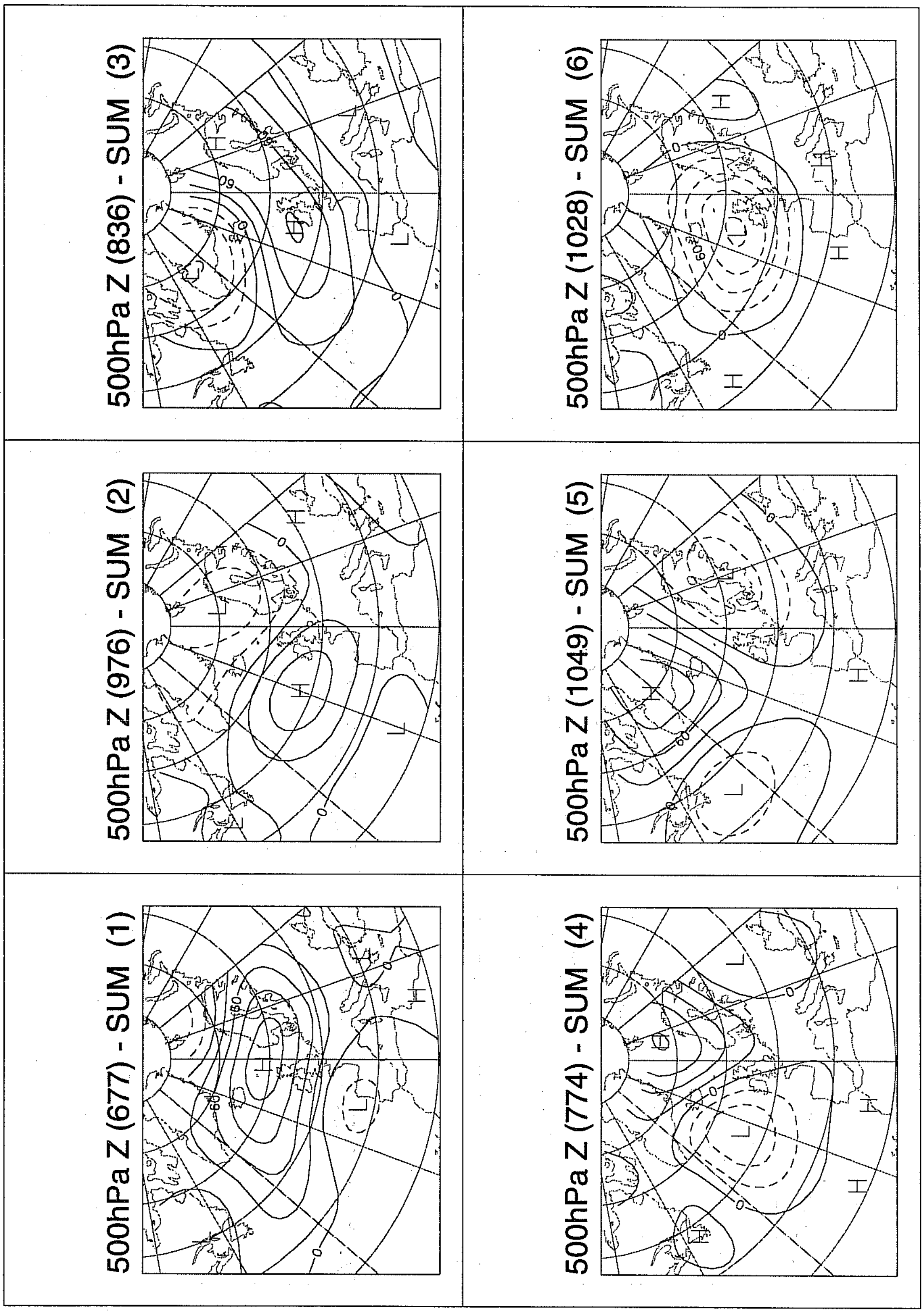

Figure 7: See Fig 1

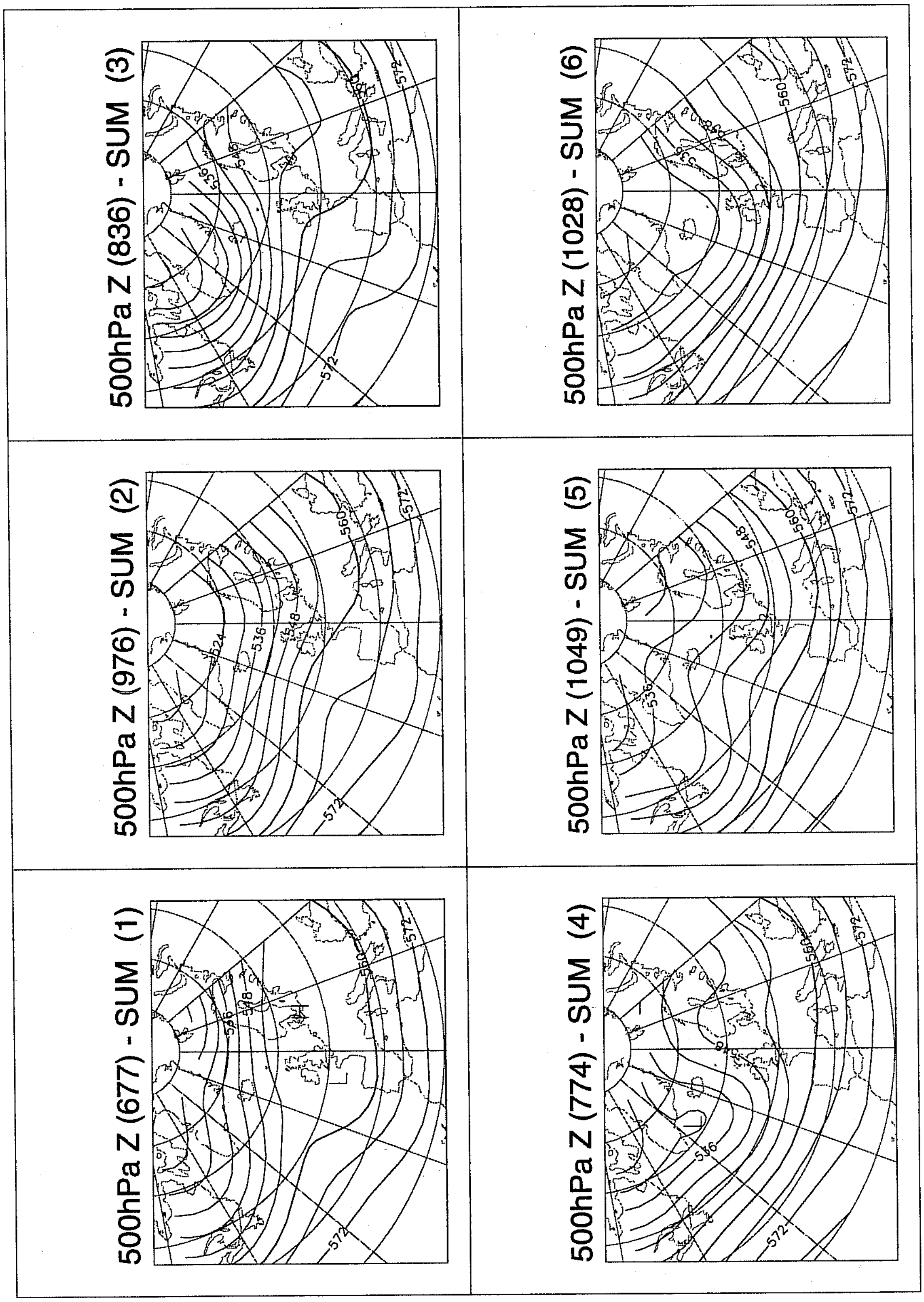

Figure 8: See Fig 1

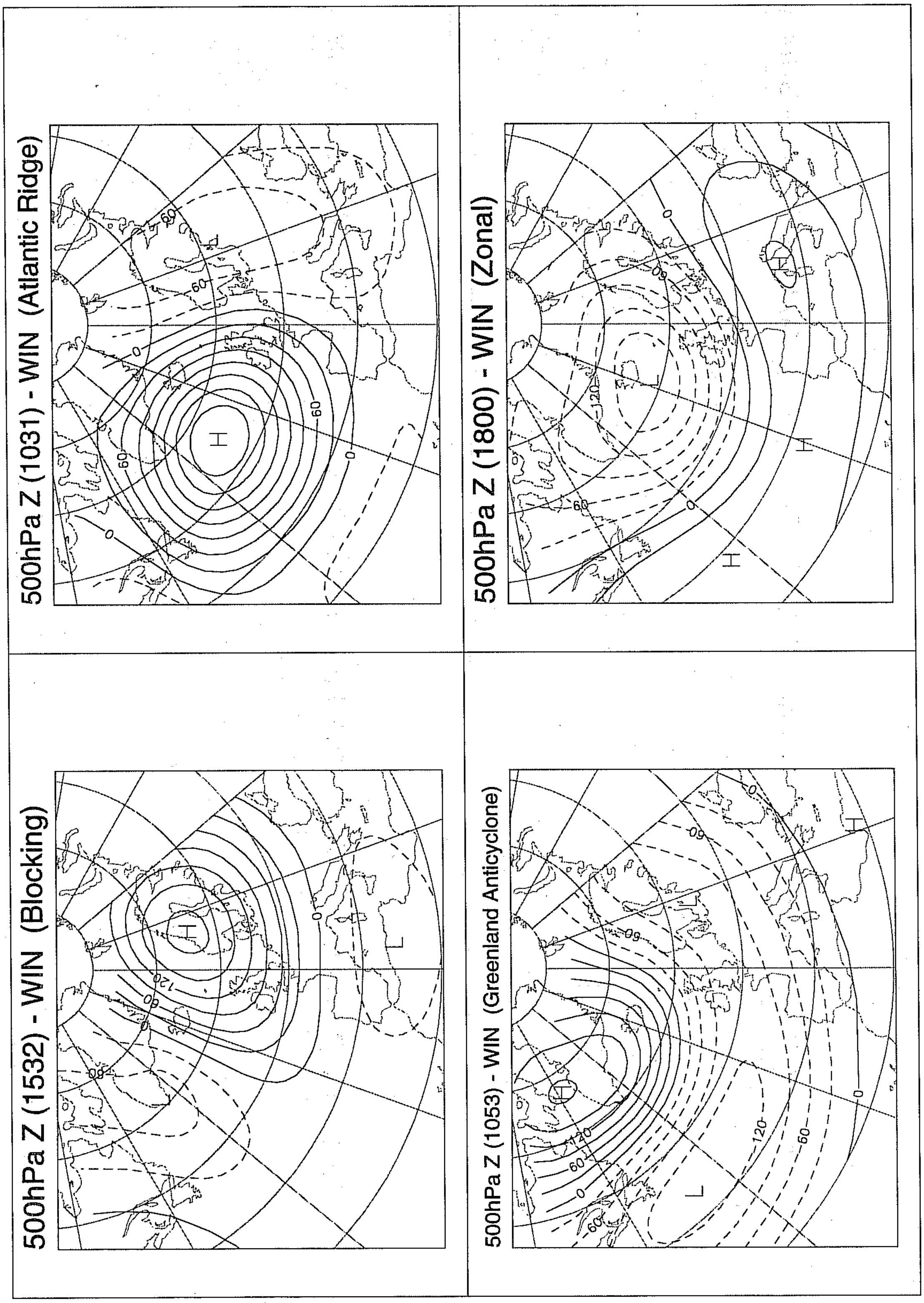

Figure 9: As for the previous figures but for four regimes in winter.

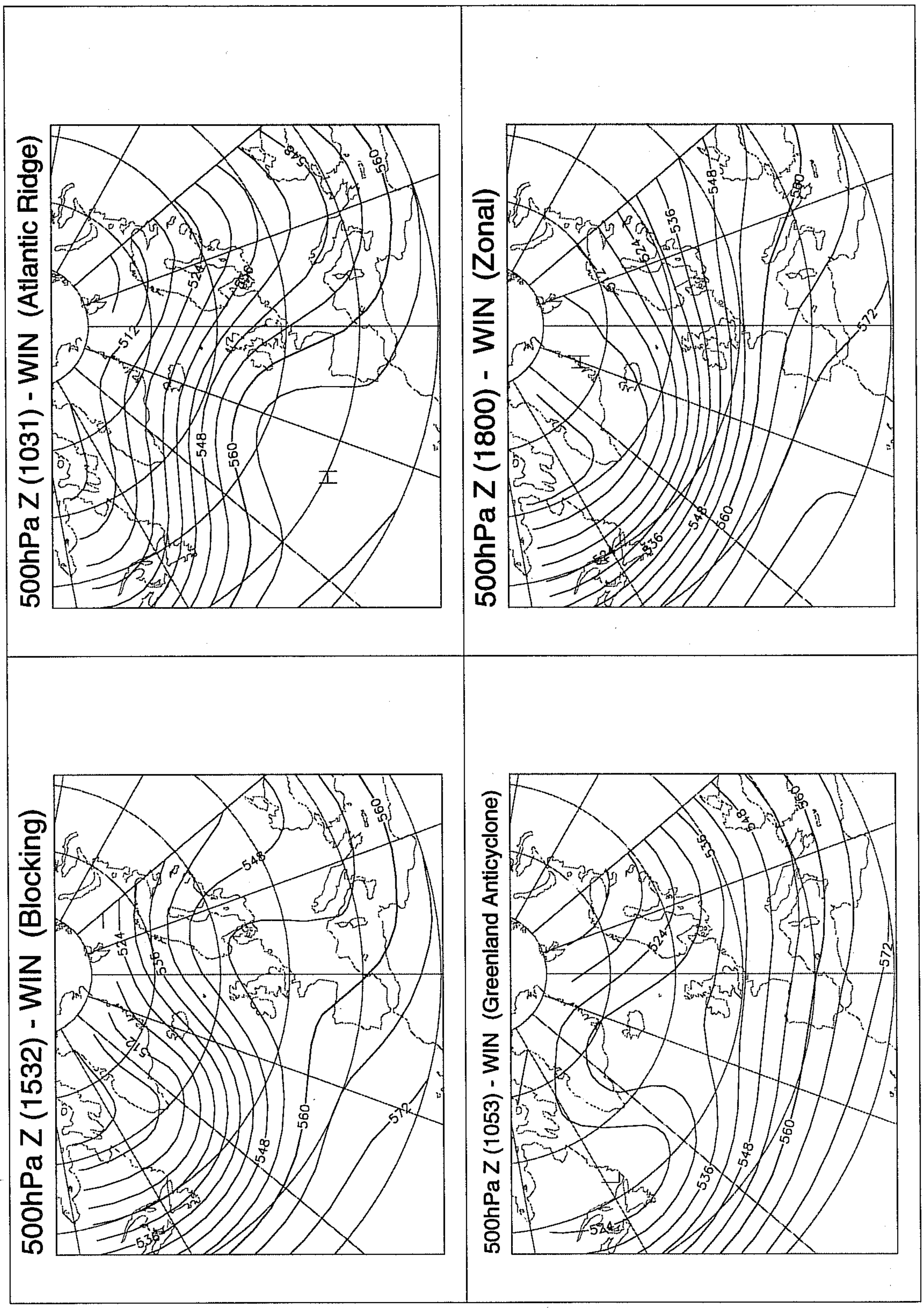

Figure 10: As for the previous figures but for four regimes in winter.

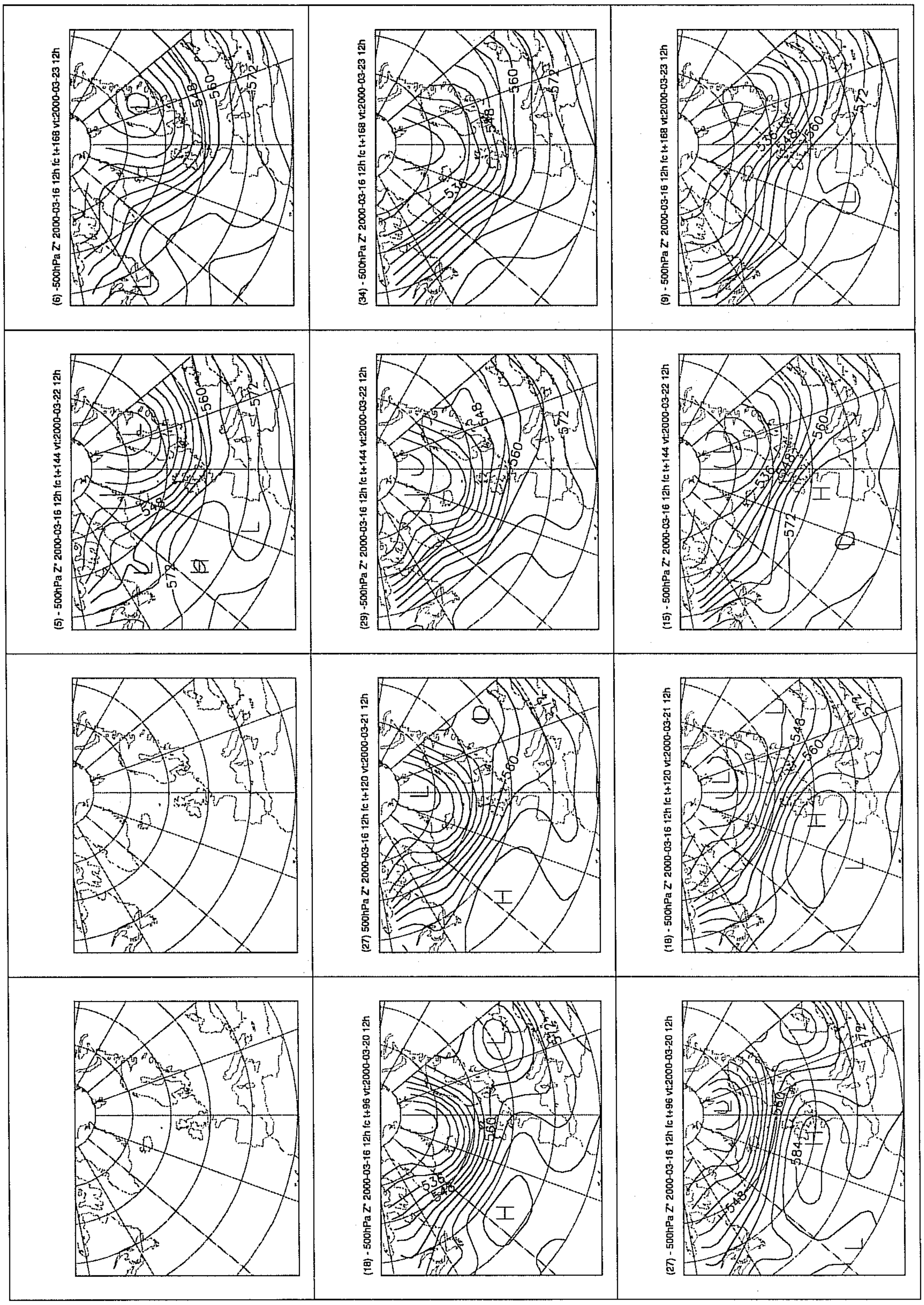

Figure 11: Cluster means for the 120h to 168h forecasts (interval 24h) obtained classifying the EPS outcome of the 16th of March 2000 Classification has been accomplished for each time step and the means are ordered according to the reference regimes. In parenthesis the population of each cluster is reported and each row corresponds to a particular spring regime. Namely the first, the forth and the sixth.

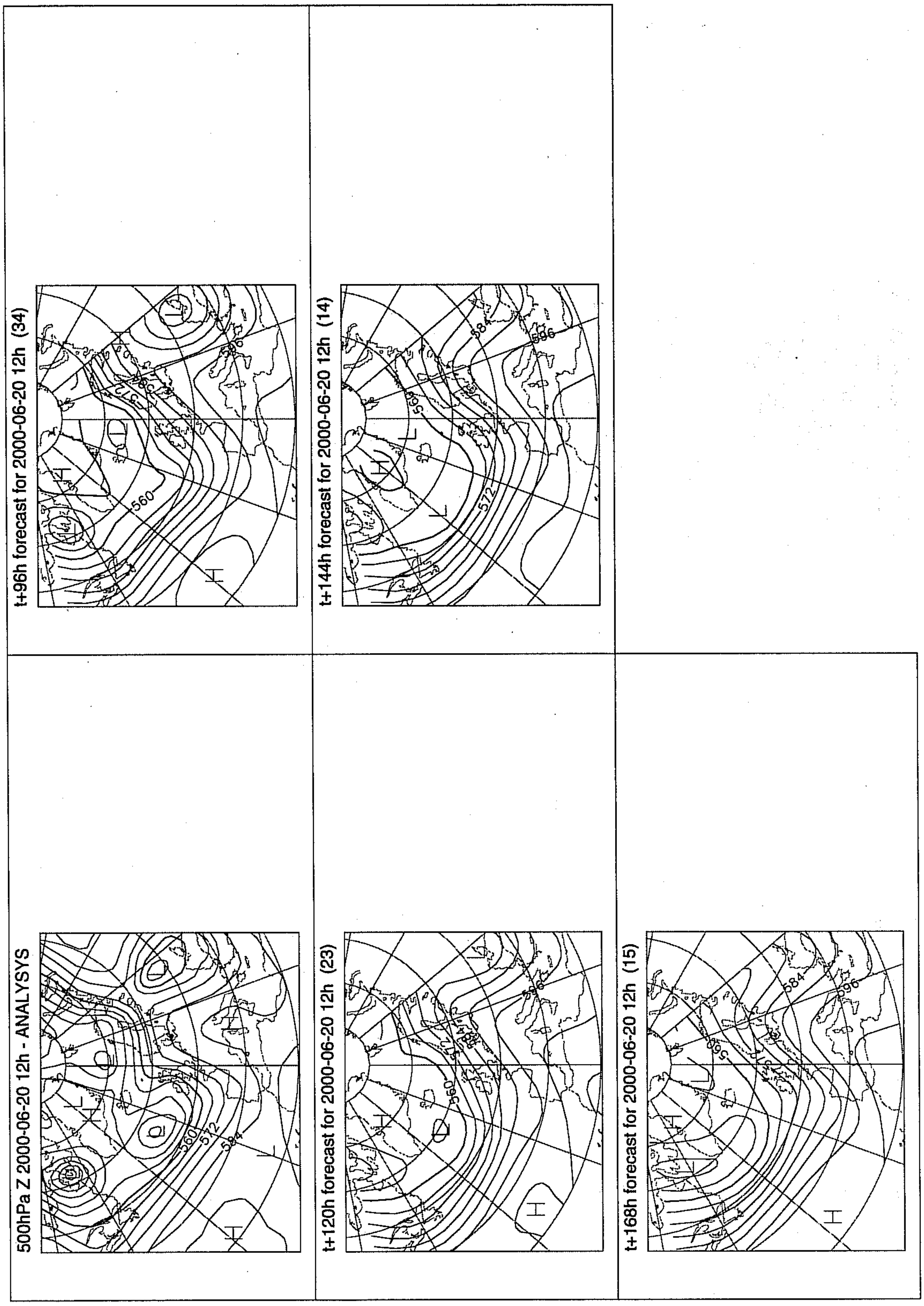

Figure 12: Consecutive forecasts verifying on the 20th of June 2000. The first panel represents the analysis for those days and in the other panel both the time step and the cluster population are reported.

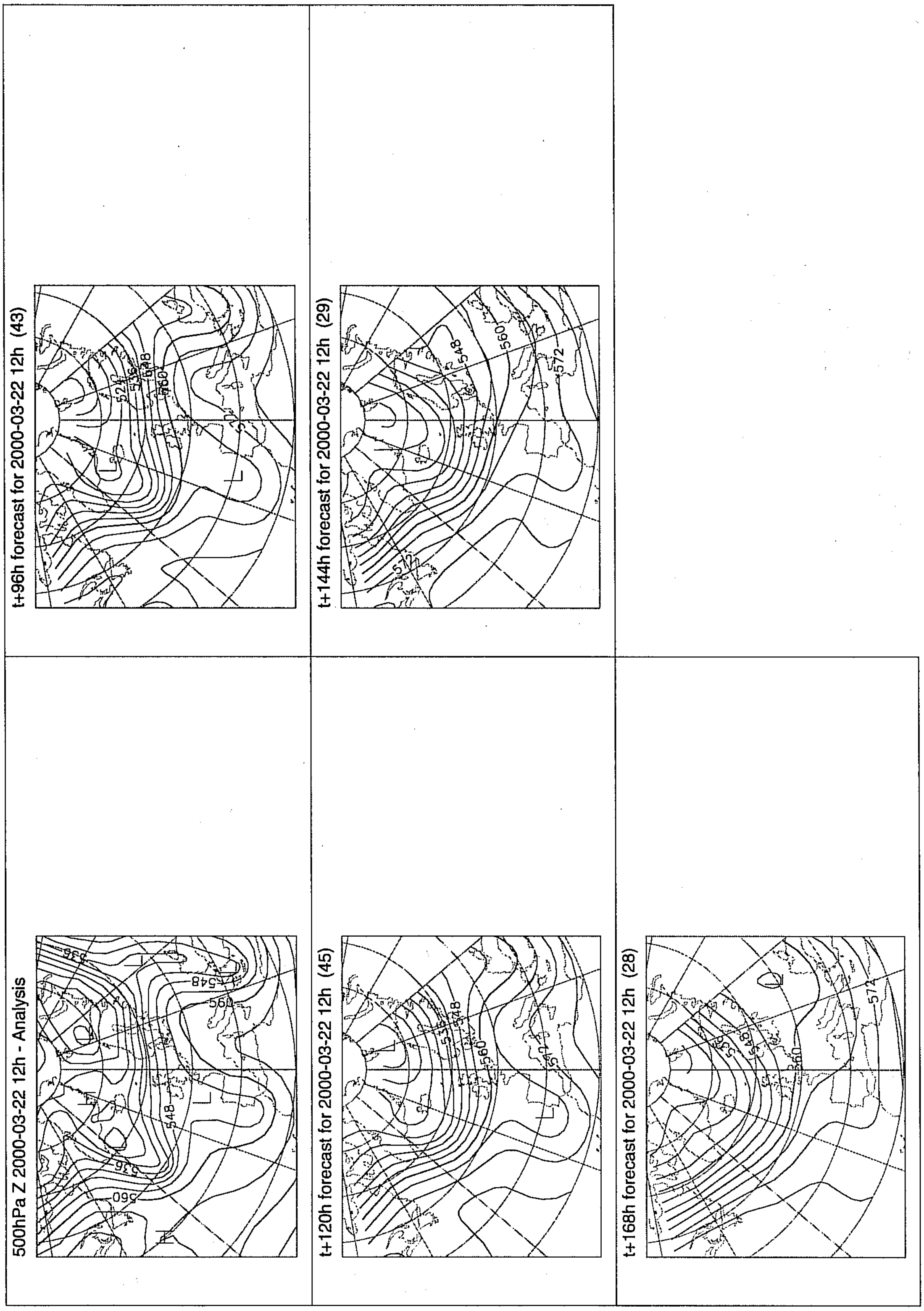

Figure 13: Consecutive forecasts verifying on the 22nd of March 2000. The first panel represents the analysis for those days and in the other panel both the time step and the cluster population are reported.

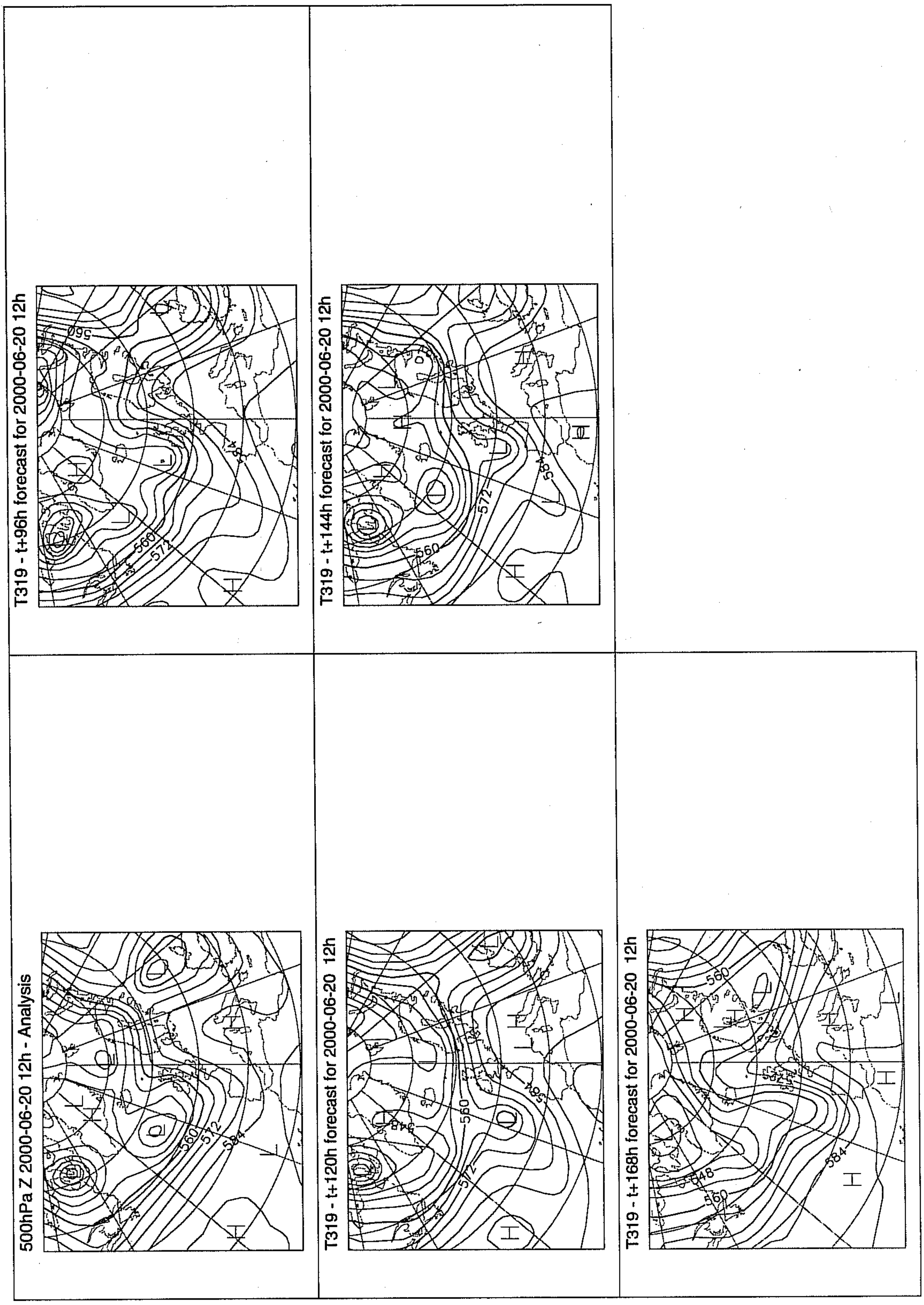

Figure 14: Consecutive forecasts verifying on the 20th of June 2000 for the high resolution model (spectral truncation T319). As in the previous figures the first panel represents the analysis for that day.