MET O 11 TECHNICAL NOTE NO. 147

Subjective assessments of medium-range numerical forecasts produced by ECMWF and the Meteorological Office

by

B.A. Hall

Meteorological Office, Bracknell, UK.

Second "Annual Forecasters' Meeting" to discuss the subjective views of users in the Member States of ECMWF products.

ECMWF, Shinfield Park, Reading, Berkshire, UK.

18-19 May 1981.

Met O 11 (Forecasting Research)
Meteorological Office
London Road
Bracknell
Berkshire
UK.

May 1981.

Contents

Subjective assessments of medium-range numerical forecasts produced by ECMWF and the Meteorological Office

by

B.A. Hall

Meteorological Office, Bracknell, UK.

1. Introduction

The European Centre for Medium-range Weather Forecasts (ECMWF) has issued forecasts for up to seven days ahead since August 1979; the forecasts are produced daily, based on data (analysis) time of 1200 GMT but prior to August 1980, forecasts were not issued on Fridays and Saturdays. Charts of the 1000 mb and 500 mb contour prognoses have been available at the Meteorological Office in Bracknell, where the medium-range forecasters in the Central Forecast Office (CFO) make use of them alongside numerical guidance from other models.

Starting in September 1979, the Forecasting Research Branch of the Meteorological Office has made detailed subjective assessments of the guidance value of the ECMWF forecasts together with similar assessments for the Meteorological Office's own operational 10-level (Octagon) model. In addition, since September 1980, subjective assessments of the guidance value of the new 15-level Meteorological Office model (still under development) have been made but it is too soon to include results from these assessments in this report. Work has also begun on accumulating data from selected grid-points around the northern hemisphere for the ECMWF model, covering forecasts for up to seven days ahead. These data are being archived daily; it is intended to produce similar output for the Meteorological Office's current operational model and for the new model and to summarise a comparison of the results in due course.

The subjective assessments described in this report are carried out for two areas, one covering eastern USA, the north Atlantic and Europe (a sector north of 30°N, from 90°W to 40°E) and the other, a subset of the first, covering the United Kingdom and Republic of Ireland (49°-61°N, 13°W-4°E). Findlater (1980) and Jones and Findlater (1980) have already described the main results arising out of the first six months (September 1979-February 1980) of the assessments. Dutton and Hall (1980) updated those reports by presenting the results of the first complete year of the project (September 1979-August 1980). Modifications to the ECMWF model have been introduced from time to time (see e.g. ECMWF (1979)) and consequently important changes in the characteristics of the model have affected its performance since the results of the first six months were summarised. For that reason, this report is based mainly on results of assessments during the period March 1980 to March 1981. Another reason for choosing this period is that assessments, except in a very few cases, were made by the same forecaster (B. Hall), whilst the assessments before March 1980 were carried out by J. Findlater. Caution should therefore be exercised in comparing statistics arising out of the first six months with those for the subsequent twelve months, although the performances of the models relative to each other are probably not significantly affected by the change of assessor.

2. The operational use of medium-range products

General details of the two numerical models being compared in this report, together with references, have been previously presented (Dutton and Hall, 1980). Modifications continue to be made to the ECMWF model; for instance, at the time of writing this report, changes in the topography dataset are being incorporated. However, the Octagon model has remained largely unchanged.

In the Central Forecasting Office at the Meteorological Office Headquarters, Bracknell, forecast fields of surface pressure and 1000-500 mb thickness are produced daily for up to five days ahead, together with a review of the expected synoptic developments and associated weather. These are for general operational forecasting purposes. The subjectively drawn forecast charts are based mainly on the Octagon model output but are influenced by consideration of numerical forecasts from the United States, West Germany and ECMWF, as well as the personal ideas of the forecaster and other synoptic information. The reliance placed on the various models depends not just on their known characteristics/defects in particular meteorological situations but on the times that forecast charts become available in relation to the tight schedules which must be adhered to in the production of forecasts. The twice daily T + 48 h and T + 72 h forecasts, prepared between 04 and 06 GMT and between 16 and 18 GMT, are based on output from the Octagon model run from 00 and 12 GMT data, respectively. Output from the USA and Germany does not become available until 10 to 13 GMT (based on 00 GMT data, once daily) whilst output from the ECMWF model, based on 12 GMT data, once daily, is not normally available until 01 or 02 GMT the following day. The daily 4-day and 5-day forecasts (T + 96 h, T + 120 h), prepared between 18 and 20 GMT, are based on Octagon forecasts from 12 GMT data.

The Octagon model is run daily out to six days ahead; in addition, since 14 January 1980, a 7-day forecast based on 12 GMT has been produced every Monday. The ECMWF forecasts are not used directly for forecasting rainfall, temperature or wind for specific days, but for assisting with the general synoptic-scale evolution.

3. Characteristics of the models

During recent years, great strides have been made in the development of numerical models for use in weather forecasting. In particular, forecasts made for T + 72 h now show the same degree of accuracy that T + 48 h forecasts did a few years ago. Looking further into the medium-range, forecasts for six and seven days ahead have certainly improved, but with limitations; comparing a 6-day forecast chart with a verification chart one may find a certain amount of merit over parts of the area considered whilst over other parts the forecast may be entirely misleading. In between the two time-scales, forecasts for 4 and 5 days ahead have proved to be extremely variable in quality, depending very much upon season/degree of mobility or whether a few large-scale features or a greater number of rapidly developing/declining smaller features predominate.

All numerical weather prediction models exhibit tendencies or biases and these may be particularly evident in certain synoptic situations or may be associated with well-defined fixed geographical features. Throughout this project of subjective assessments, attempts are made to identify any such idiosyncracies of the respective models. Several years of regular operational experience with Octagon products has helped to establish a well-founded knowledge of the characteristic behaviour of that model in a wide range of synoptic situations. Initially, output from the ECMWF model was treated with some reservation, especially in certain synoptic situations; for instance, depressions tended to be over-developed and were often incorrectly forecast to become large slow-moving features near Iceland by days 5-7. Sometimes two large depressions would be predicted to co-exist in rather close proximity, and this is seldom observed to occur (Findlater, 1980). These, and other faults, showed up in the results of the subjective assessments carried out during the first six months of the project. However, it has become obvious during the past year that, as improvements have continued to be made to the model, many of the major faults appear to have been corrected or reduced, particularly for forecasts up to 4 days ahead.

Although the ECMWF model has shown a distinct improvement since the report by Jones and Findlater (1980), some characteristic errors remain. These are re-iterated below with some elaboration, qualifications and additions stemming out of the experience with the model during the past year; these points are contrasted or compared with the well-known characteristics of the Octagon model when appropriate.

It has emerged during the assessments that both models are sometimes capable of following individual features such as fronts (particularly cold fronts) or troughs through to day 6 or 7, and to predict their development in a realistic fashion. On the other hand, although on many occasions a certain amount of merit may be found in the forecasts up to day 7, the practical value of the model output to a forecaster may not be as great as the theoretical value perceived by the numerical modeller. The type of error may also be important; for instance, a moderate longitudinal error in a zonal situation is less serious than an error of similar magnitude in latitudinal position. Up to days 3 or 4 the development of new depressions in cases of frontogenesis is often well indicated but if substantial developments occur at or after day 4, they are often poorly modelled or missed altogether.

The ECMWF model is generally very active, maintaining the strength of the main upper flows well, both zonal and meridional components, whereas the Octagon has a distinct tendency to weaken them with time, particularly their meridional components. The ECMWF model's development of new depressions from shallow wave features moving into regions of frontogenesis, such as that along the eastern coast of USA in winter, often appears convincing; more often than not the rate of deepening and phase speed of new depressions in the early development is quite well predicted, especially when this occurs before day 4 in the forecast period (see, for example, Fig A1 in the Appendix) although, as with most numerical models, there is frequently a tendency for the phase and development of such features to lag behind the real atmospheric evolution (example: Fig A2).

Although the early stages of the development of new depressions were often adequately predicted by the ECMWF model, there was ample evidence that, in the case of mature depressions, the model frequently failed to turn them to the left of their original tracks (example: Fig A3), as frequently happens in the real atmosphere; in fact the tracks of major depressions and anticyclones were often predicted to be further south than the actual tracks. This was associated with the tendency for the ECMWF model's jet stream axes to be too far south. In the past 12 months, however, these faults have occurred less frequently, and when they have occurred the magnitude of the error has usually been considerably less.

In contrast the Octagon has a distinct bias towards zonal flow; meridional developments tend to be underforecast with the typical result that the central pressures of surface depressions, particularly those in the developing stage, are not predicted low enough. In spite of this, the timing of events and positions of the main synoptic systems were often good, particularly during the summer months or periods of low activity. One interesting feature noted whilst assessing output from the Octagon model was a tendency for a short-wave phase error to occur at days 3 and 4, with a fortuitous improvement evident at days 5 or 6 as the model returned to approximately the correct phase.

There seems little evidence to suggest which model, if any, is better at indicating changes of type. For instance, in a change from a quiet, anticyclonic spell to a mobile westerly the ECMWF model will often predict this change earlier than the Octagon. Conversely, given the situation of a mobile westerly flow giving way to a blocking pattern, the Octagon model may be better, especially if the block develops over Europe. If the block builds out of a developing anticyclone in mid-Atlantic, the ECMWF model is usually more emphatic (example: Fig A4). The extent to which these features reflect characteristic errors of the models is uncertain.

Both models are capable of producing reasonable predictions of the disruption of upper troughs and the subsequent formation of cut-off vortices, the ECMWF model more frequently being the superior in this respect (example: Fig A5), particularly, it seems, when the cut-off vortex forms south of about 40°N, in the Mediterranean for example. (This is not inconsistent with the ECMWF model's propensity for meridional developments). However, the process of disruption usually occurs more rapidly than predicted by either model with the typical consequence that the low-latitude axis of the disrupting trough is carried too far east in the forecast.

Indications, during the first six months of the assessments, that the Octagon model performed better than the ECMWF model in handling developments in the Mediterranean were not upheld. The ECMWF model predicts the development of lows moving ENE'wards from the Sahara, across the Mediterranean, into the Black Sea region better than the Octagon. Once having forecast the central pressures of the lows reasonably accurately when at their deepest, the ECMWF model does not usually fill the lows quickly enough. It is of synoptic interest to note here that, on most occasions when this type of development has been observed during the past year, low latitude upper troughs or vortices associated with the features mentioned have phased in with troughs extending southwards from the main flow at higher latitudes across USSR.

During the first six months both models tended to predict 850 mb temperatures too low south of 40°N in the Atlantic, Iberia and N Africa regions; in particular the Octagon model typically forecast 850 mb temperatures to be 4-8 deg C too low in these areas by D + 6. Insufficient allowance for warning by subsidence and convection, especially that due to the release of latent heat in the convection process, is possibly the cause. During the past year this tendency has persisted with the Octagon, particularly in association with the Azores anticyclone, but has not been so much in evidence in the ECMWF 850 mb temperature fields.

A fact well known of all numerical models is their inability to cope with hurricanes and their poor handling of ex-hurricanes, ex-tropical depressions etc. There were a number of occasions during the autumn of 1980 when such features significantly affected the larger scale pressure distribution (e.g. the northward advection of large quantities of air of tropical origin on the western side of an Atlantic high, thereby helping to build the high into an important block). Neither of the models being assessed was helpful in that respect.

4. Assessment scheme

The main aim of the project is to assess the information content of the numerical forecasts (D + 1 to D + 7) and hence their value as guidance to forecasters. In an attempt to overcome the difficulties noted by Findlater (1980), associated with subjective assessment, a very simple and general scheme for the assessment of the forecasts was adopted; it was based on a scheme in the Central Forecast Office at Bracknell for the subjective assessment of forecasts (by experienced forecasters). There were three "guidance value" categories used:-

A - Good guidance
B - Did not lead to any major error
C - Misleading in some important respect

However, it was the general feeling of forecasters at the first annual Meeting of Forecasters (held at ECMWF in Shinfield Park on 2-3 June 1980) that a scheme with a larger number of categories should be adopted; the scheme favoured was one suggested by Pirkko Saarikivi of the Finnish Meteorological Institute, using six categories. The problem then faced by the forecaster, based at the Meteorological Office Headquarters, responsible for continuing the assessment project was how to adjust to the new scheme but at the same time enable comparisons using the original three-category scheme to continue. This problem was overcome by assigning 'pluses and minuses' to the A's, B's and C's, the resultant 6 categories being:-

A+ Excellent guidance
A- Good guidance
B+ Quite useful guidance
B- Some errors but not too misleading
C+ Somewhat misleading
C- Completely misleading

Although at first sight it may appear that there could be problems with absolute standards, it will be shown in section 5 that this is not the case. This latter scheme was introduced from 1 September 1980, although the basic method of assessment was unaltered. For each weekly sequence of forecasts (based on data time 12 GMT Mondays) the charts for D + 1 to D + 7 at 500 mb and surface/1000 mb were individually assessed by a meteorologist with good experience of synoptic forecasting. Separate assessments are made for the two areas:-

(i) "WHOLE AREA" - Covering a sector north of 30°N, from 90°W-40°E.

(ii) "UK AREA" - 49°-61°N, 13°W-4°E.

The assessment attempts to take account of the various constituent elements of the weather, as inferred from the predicted pressure distribution; qualitative impressions of temperature, wind (speed and direction), general weather type (including likelihood of precipitation) can all be inferred. In addition the sequence of events indicated by the forecast is taken into account. Within the scheme adopted, the guidance value frequently depends to some extent on the forecast period; for instance a moderate displacement of a particular feature from its actual position at D + 1 or D + 2 is generally regarded as a more serious error than the same displacement at D + 5 or D + 6, when the forecast should be taken to give only a general indication of the positions and tracks of synoptic-scale systems.

Apart from the subjective assessment described above, actual and predicted zonal and meridional indices at 500 mb for the UK area were logged for each of the weekly sequences. Some results of the analysis of these data are shown in section 5.

5. Comparison of model forecasts

In this section the content of the Figures which present the main results will be described, together with discussions of the main points arising.

It is thought to be important that the original classifications of the ECMWF and Octagon forecasts (that is, omitting the plus and minus subdivisions) for the past year should be included; these appear in Figure 1. Jones and Findlater (1980) produced a similar figure in their report covering the first six months of the project but, as mentioned in the introduction to this report, caution should be exercised when making direct comparisons between the two periods. Within each figure the results for each area (WHOLE AREA and UK AREA) and level (SURFACE/1000 mb and 500 mb) are presented separately. Inspection of Figure 1 reveals that generally for the N Atlantic/Europe (whole) area, a) the ECMWF model received a greater number of 'A' markings than the Octagon model and b) the Octagon model received a greater number of 'C' markings than the ECMWF model. At 500 mb, on days 1 and 2, there was little significant difference between the total marks; on days 3 and 4 marks for the ECMWF model were considerably better than those for the Octagon; at day 5 EMCWF forecasts were "misleading" on fewer occasions than forecasts by the Octagon model, but on days 6 and 7 results were poor for both models. At the surface (whole area), the superiority of the ECMWF forecasts widened to cover the period of days 2 to 5; it must be emphasised, however, that the total number of misleading forecasts by both models for days 5 to 7 was high. Results for the UK area were slightly different: at 500 mb ECMWF achieved more 'A''s and fewer 'C''s than the Octagon model for the forecasts for days 2 to 6 inclusive, but at the surface the forecasts for days 3 and 4 were the only ones in which the ECMWF model showed any significant advantage.

A set of graphs is provided in Figure 2, showing the percentage of occasions when A/B marks were retained up to and including day N; A/B may be interpreted here as meaning 'not misleading'. The arrows superimposed on the diagrams correspond to the positions at which the percentages for the respective models cross the 50% ordinate. From these graphs it can be seen that, in each of the four assessment divisions, the ECMWF model achieved the same 'degree of usefulness' out to around three-quarters of a day (15 to 21 hours) longer than the Octagon model, on average, during the past year. However, an interesting point to consider here is the amount of CPU time required by the computers to run their respective models: the Octagon model takes approximately $3\frac{1}{2}$ minutes to run a 24-hour forecast on the IBM 360/195 whilst the ECMWF model takes approximately 12 minutes for the same forecast time on the Cray-1 (a computer which is some 5 to 10 times faster than the IBM machine). A further factor in considering the reliance placed on the models is the "cut-off" time. Cut-off times for the two models are T + $3\frac{1}{3}$ h for the Octagon and T + 6 h* for the ECMWF model.

Although no objective hemispheric statistical comparisons between the two models are yet available, a small sub-project has been carried out in order to compare forecast and actual zonal and meridional indices (in decametres) at 500 mb for each of the models, measured across the UK** using the respective model output charts and verifying against the Octagon analyses. The results are shown in Figures 3 and 4 (zonal and meridional indices, respectively), each presenting the root-mean-square errors (dm) for days 1 to 7 separately for each quarter of the project. Although each graph represents a fairly small sample (13 values for each day), it can be seen that there is generally less deviation in the indices for the spring and summer of 1980 than during the first six months covering the previous autumn and winter. The zonal figures for the summer of 1980 show

---

* From 29 March 1981 this has been brought foward to T + 5 h.

** Zonal indices measured between 50° and 60°N at 5°W and meridional indices measured between 13°W and 4°E at 55°N over an 18-month period September 1979 to March 1981.

a distinct advantage to the Octagon model for days 4 and 5. Both the zonal and meridional graphs for the autumn of 1980 are remarkably similar for the two models. During the winter of 1980-81, however, the ECMWF model proved the better for days 2 to 4 but for days 5 to 7 the errors were somewhat erratic and certainly on the large side - rather like those for the previous winter. Dutton and Hall (1980) confirmed the commonly observed tendency of most numerical models to overforecast zonality in weak flow and to underestimate it in strong zonal flow; this certainly appears to continue to be the case.

At the 1980 annual Forecasters Meeting some lively discussion took place over the relative merits and faults of the various marking schemes used by the Member Countries for their assessments. As described in section 4, two marking schemes have been run concurrently in order to try to justify the original simple 'A B C' scheme started by Findlater (1980). Figures 5 and 6 compare results from the two schemes, covering the period September 1980 to March 1981. In the 3-point scheme, A is taken as +3, B as +1 and C as -1; in the 6-point scheme A+ is taken as 3.5, A- as 2.5, B+ as 1.5 and so on to -1.5 for C-. Figure 5 shows the results for the N Atlantic and Europe area and Figure 6 those for the UK area. Apart from a little smoothing, as would be expected using a finer scheme, it can readily be seen that there is very little difference between the results from the two schemes. In the opinion of the author, who was responsible for carrying out the task of making the assessments during the past year, the extra time and thought required to produce marks under the 6-point scheme is not necessary if assessments are carried out over a long period for the purpose of subjective statistical comparison. For individual case-studies, however, the six-point scheme is justifiable; indeed, arguments have also been put forward for further dividing the area under consideration but it is felt that the resultant complications in the presentation of results would be unacceptable.

5. Conclusions

This programme of subjective assessment of ECMWF and Meteorological Office operational numerical forecasts has shown that during the year March 1980 to March 1981:

a) A noticeable improvement in the performance of the ECMWF model relative to the Octagon has occurred compared with the first six months of the project (September 1979 to February 1980), when neither model proved to be superior in any consistent fashion.

b) The general guidance of forecasts by both models for days 1 and 2 was good, with ECMWF marginally better at the surface on day 2. Marks for the ECMWF model were markedly superior to those for the Octagon during days 3 to 5 but the total number of misleading forecasts by both models for days 5 to 7 was high.

c) On average, the ECMWF model achieved the same 'degree of usefulness' out to around 18 hours longer than the Octagon. This is not necessarily a true measure of the difference between the models as the ECMWF model operates with a later cut-off time.

d) During periods of weak zonal flow (e.g. summer 1980 over UK), statistics show that the Octagon model proved to be less misleading than the ECMWF model at days 4 and 5.

e) Little significant difference is apparent between overall results from a simple 3-point marking scheme and those from a 6-point marking scheme for assessing model products.

References

Burridge, D.M. and Gadd, A.J. 1977 The Meteorological Office operational 10-level numerical weather prediction model (December 1975). Sci Pap No 34, HMSO, London.

Burridge, D.M. and Hascler, J. 1977 A model for medium-range weather forecasts - adiabatic formulation. ECMWF Tech Rep No 4.

Dutton, M.J.O. and Hall, B.A. 1980 Subjective assessments of medium-range forecasts produced by ECMWF and Meteorological Office operational numerical models. Met O 11 Tech Note No 146.

Findlater, J. 1980 A preliminary report on the subjective comparison of medium-range forecasts produced by ECMWF and the Meteorological Office. Met O 11 Tech Note No 137.

Gadd, A.J. 1978 A split-explicit integration scheme for numerical weather prediction. QJR Met Soc, 104, pp 569-582.

Jones, D.E. and Findlater, J. 1980 Subjective comparisons of medium-range forecasts produced by ECMWF and the United Kingdom Meteorological Office. Meeting on synoptic quality of ECMWF forecasts, ECMWF, June 1980.

Tiedtke, M., Geleyn, J.F., Hollingsworth, A., and Louis, J.F. 1979 ECMWF model - parameterization of sub-grid scale processes. ECMWF Tech Report No 10.

ECMWF 1979 Annual Report 1979.

Figures and Appendices

N. ATLANTIC/EUROPE

SURFACE

DAY

	1	2	3	4	5	6	7
A	50/50	44/35	21/9	10/4	2/0	2/1	1/1
B	2/2	8/17	25/29	21/17	17/10	7/7	7/7
C	0/0	0/0	6/14	21/31	33/42	43/44	44/44

500 MB

DAY

	1	2	3	4	5	6	7
A	52/52	48/46	30/15	13/5	2/2	2/2	1/1
B	0/0	4/6	19/33	25/21	25/12	15/12	8/10
C	0/0	0/0	3/4	14/26	25/38	35/38	43/41

U.K. AREA

SURFACE

DAY

	1	2	3	4	5	6	7
A	49/49	37/37	30/16	17/11	9/8	5/7	2/6
B	3/3	15/12	16/25	19/14	19/20	18/17	14/12
C	0/0	0/3	6/11	16/27	24/24	29/28	36/34

500 MB

DAY

	1	2	3	4	5	6	7
A	51/50	47/42	37/27	21/16	16/5	10/7	4/6
B	1/2	5/9	12/21	19/21	19/21	17/13	11/12
C	0/0	0/1	3/4	12/15	17/26	25/32	37/34

Key: EC/MO (upper left of cell = EC, lower right = MO)

EC – ECMWF model
MO – Met.Office model

Fig. 1 . Classification of ECMWF and Meteorological Office (Octagon) forecasts for the period 3/3/80 to 9/3/81. (52 forecast sequences).

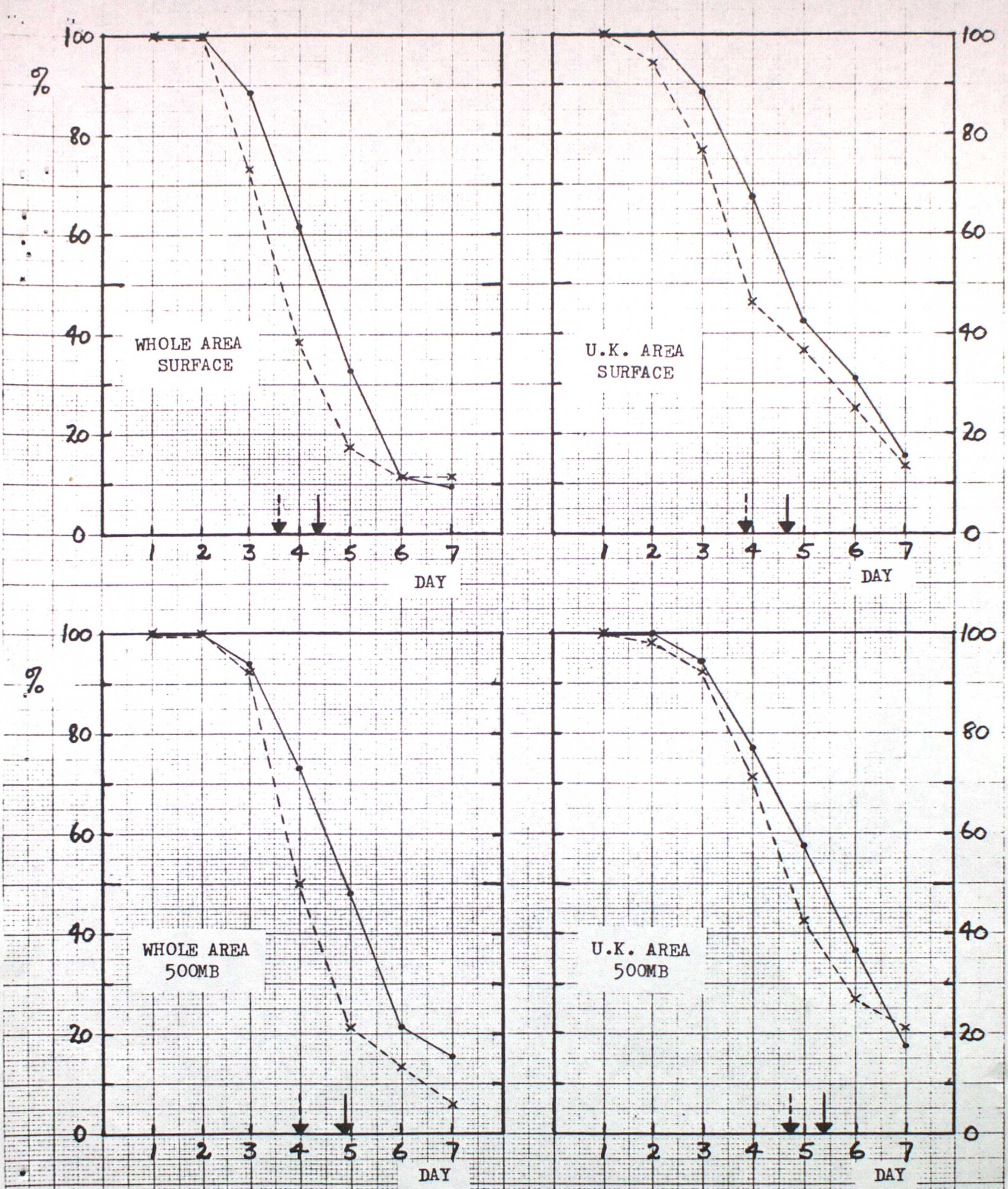


Fig. 2. Percentage of occasions when A/B marks were retained up to and including day N, for the period 3/3/80 to 9/3/81 (52 weeks).

●——● ECMWF Model

×---× Met O Model

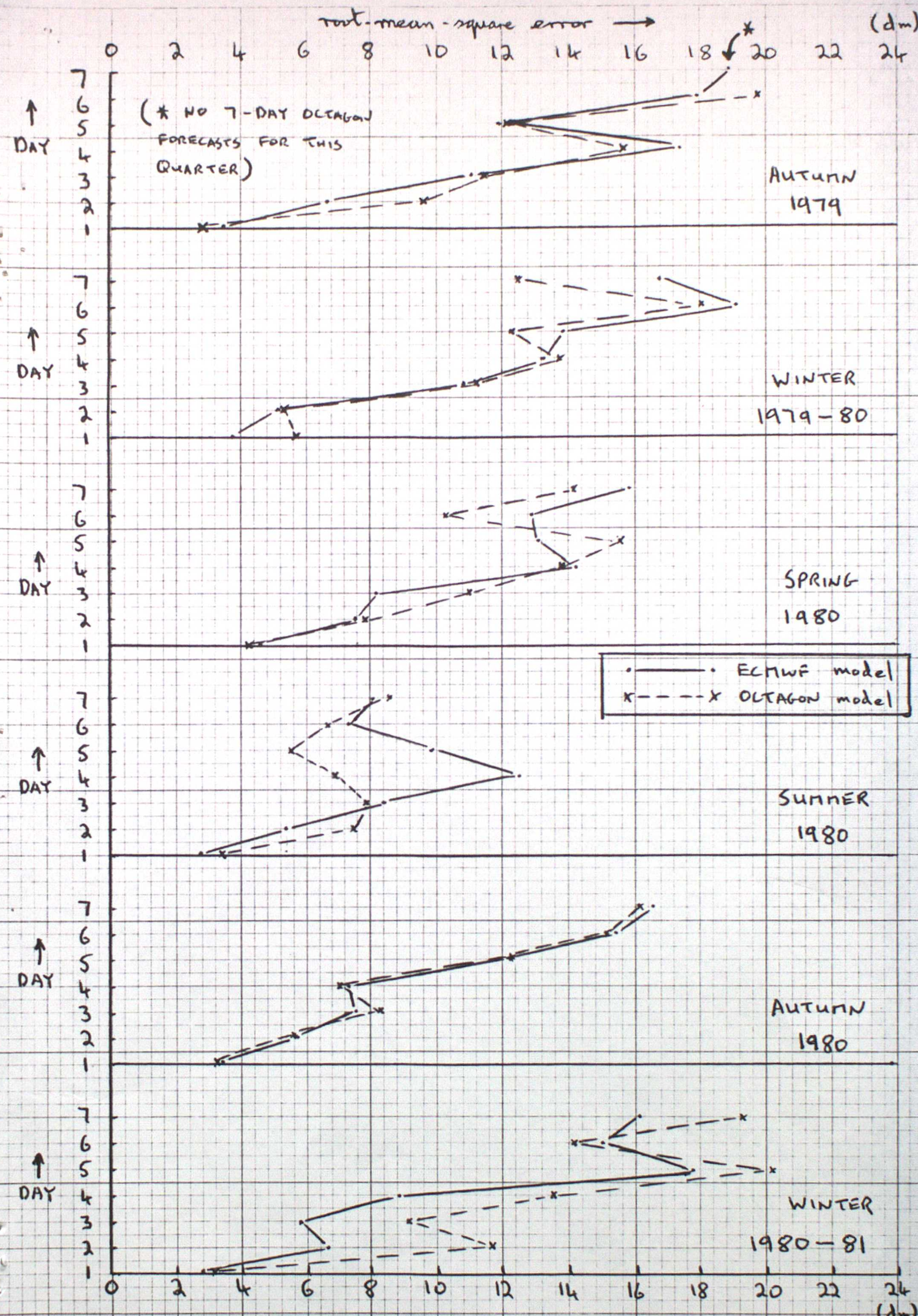


Fig. 3. Seasonal variations in root-mean-square error (dm) of zonal indices at 500 mb measured (forecast minus actual) between 50° and 60°N at 5°W, a total of 78 weekly sequences.

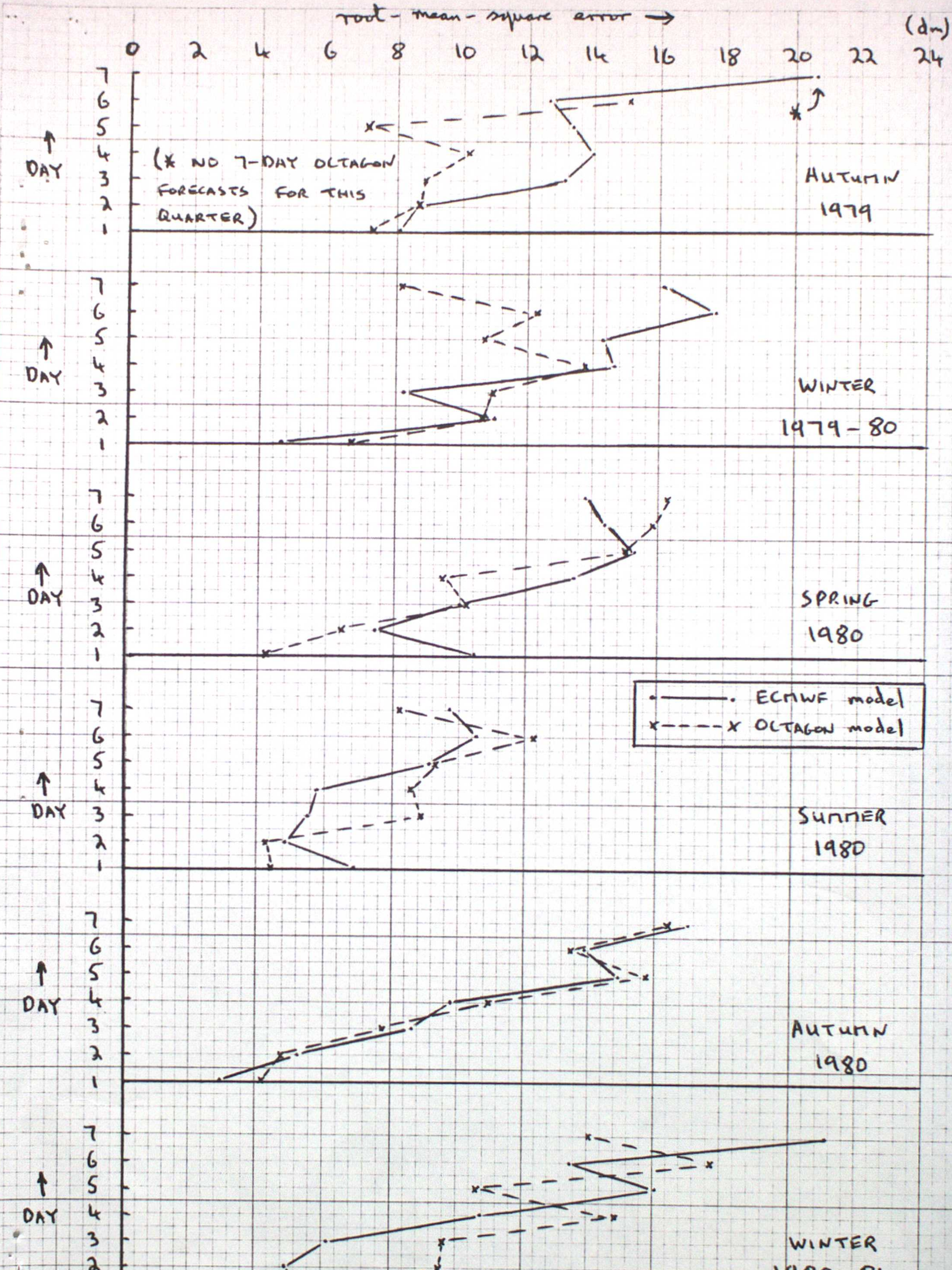


Fig. 4. Seasonal variations in root-mean-square error (dm) of meridional indices at 500mb measured (forecast minus actual) between 13°W and 4°E at 55°N, a total of 78 weekly sequences.

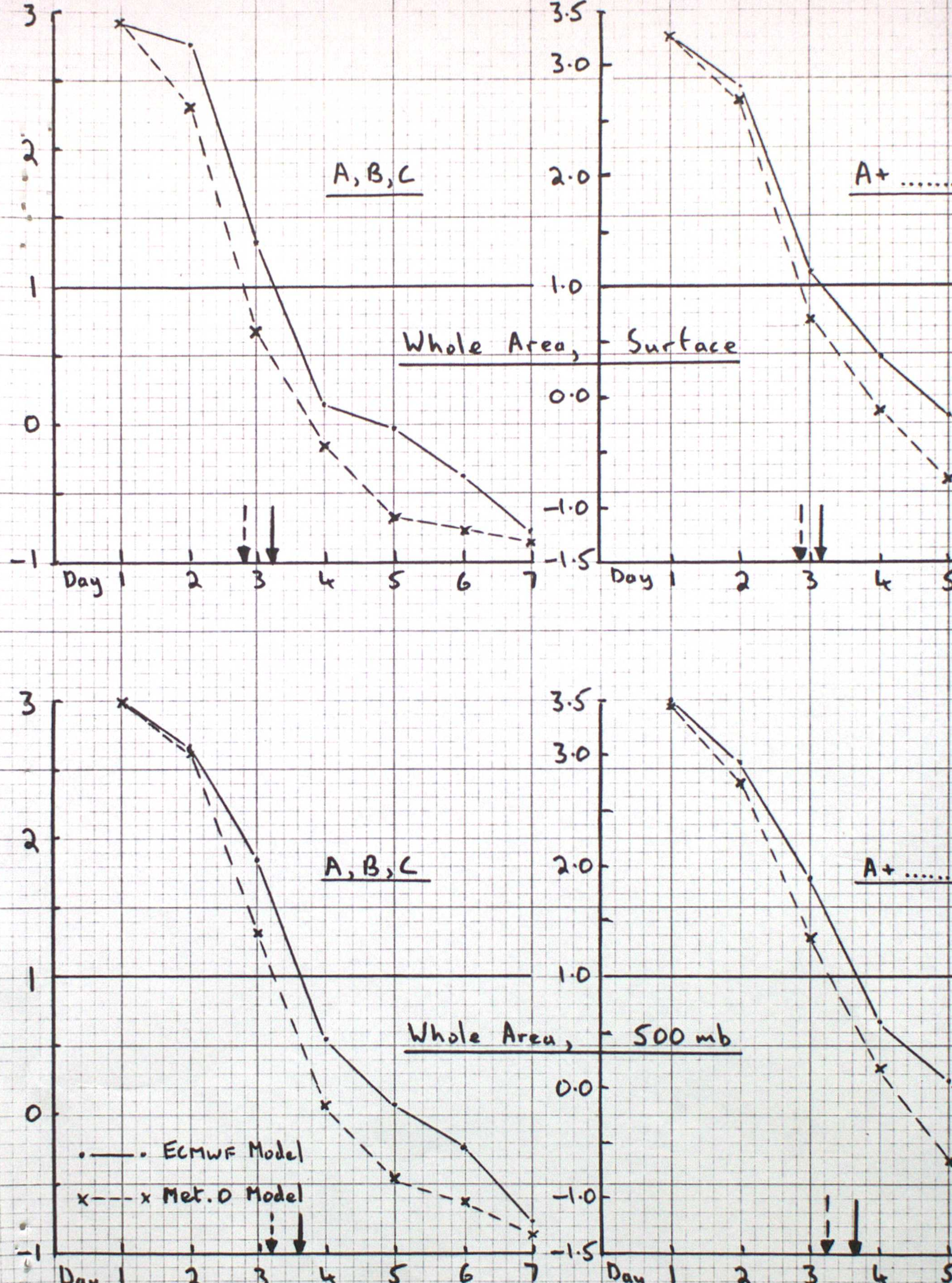


Fig. 5. Comparison of 2 marking schemes - a 3-point (A, B, C) and a 6-point scheme (A+, A-, B+, B-, C+, C-) the period 1/9/80 to 9/3/81 (26 weeks), whole area.

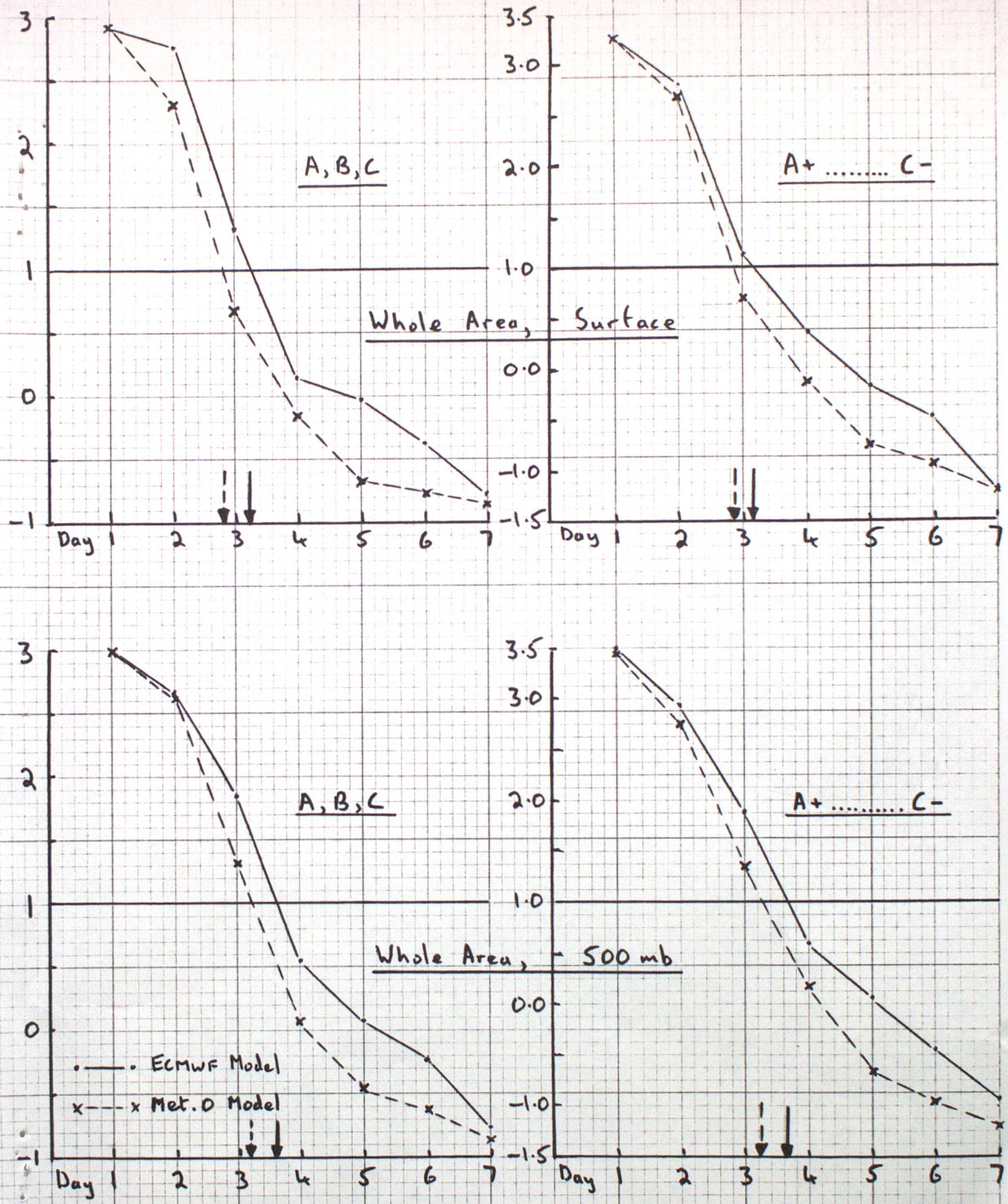


Fig. 5. Comparison of 2 marking schemes – a 3-point scheme (A, B, C) and a 6-point scheme (A+, A−, B+, B−, C+, C−) for the period 1/9/80 to 9/3/81 (26 weeks), whole area.

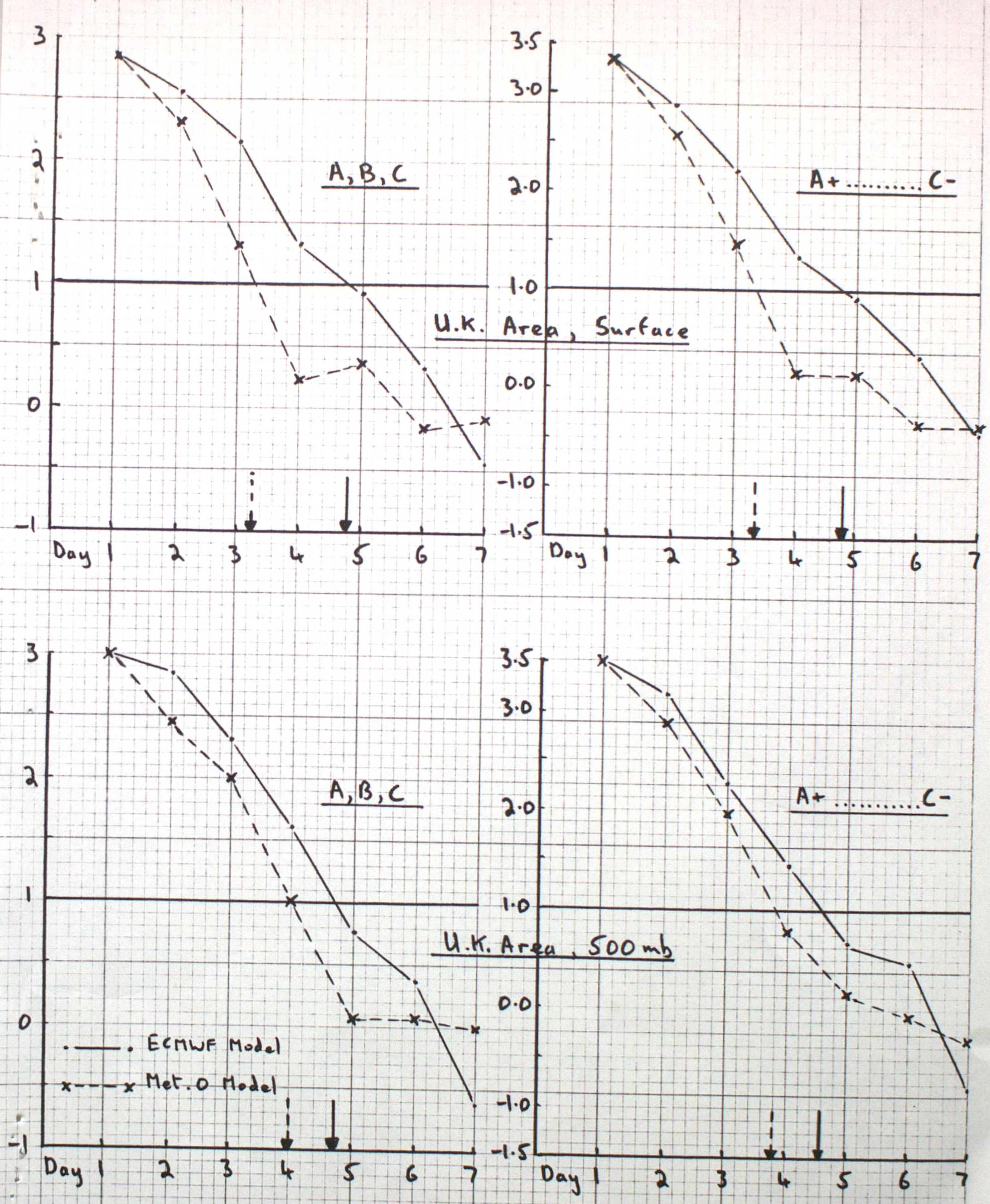
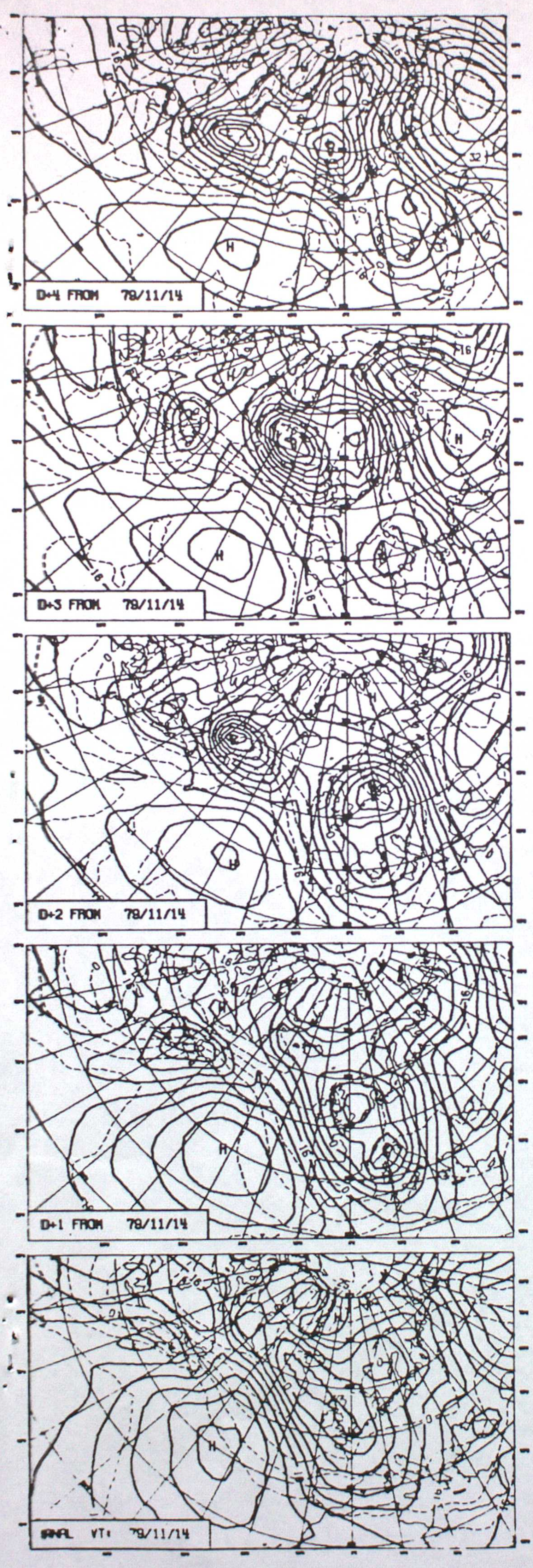


Fig. 6. Comparison of 2 marking schemes – a 3-point scheme (A, B, C) and a 6-point scheme (A+, A–, B+, B–, C+, C–) for the period 1/9/80 to 9/3/81 (26 weeks), U.K. area.



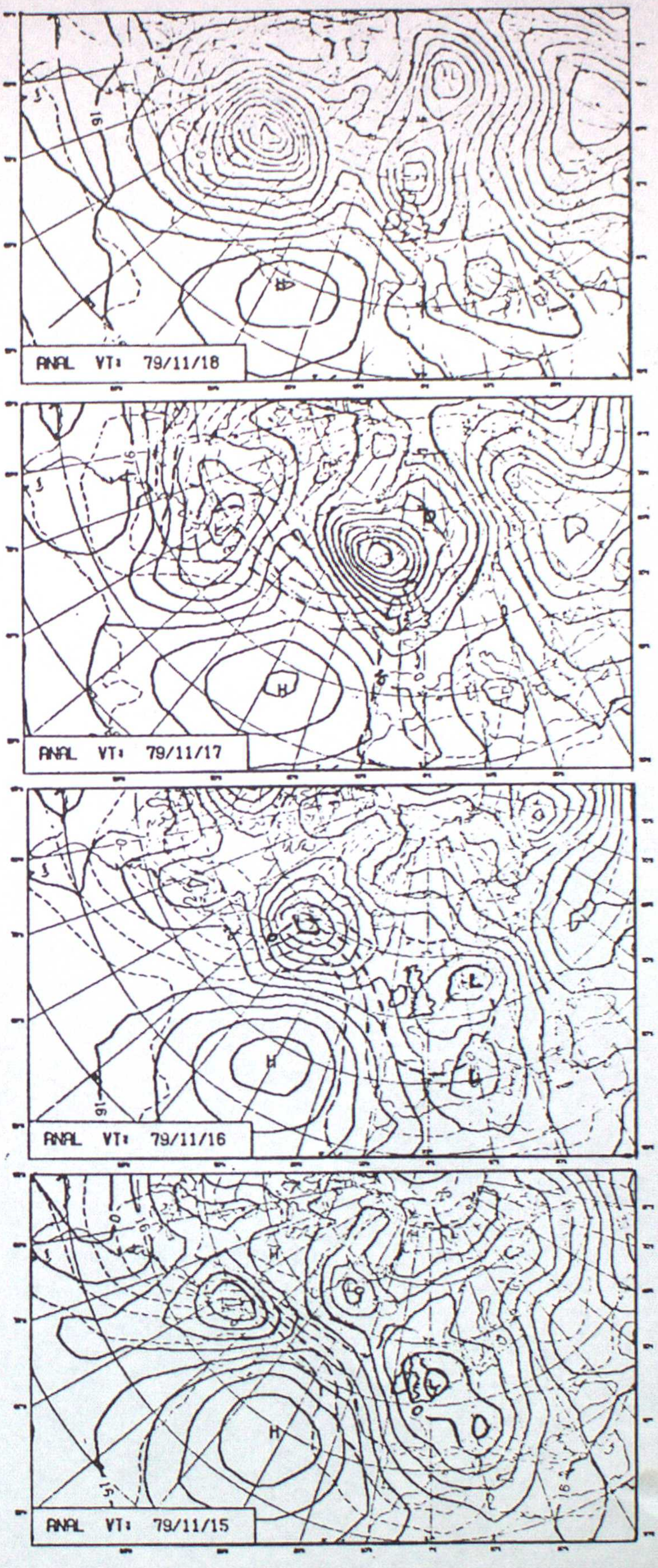


Fig. A1. ECMWF 1000mb analysis and forecast fields from 12 GMT on 14/11/79.

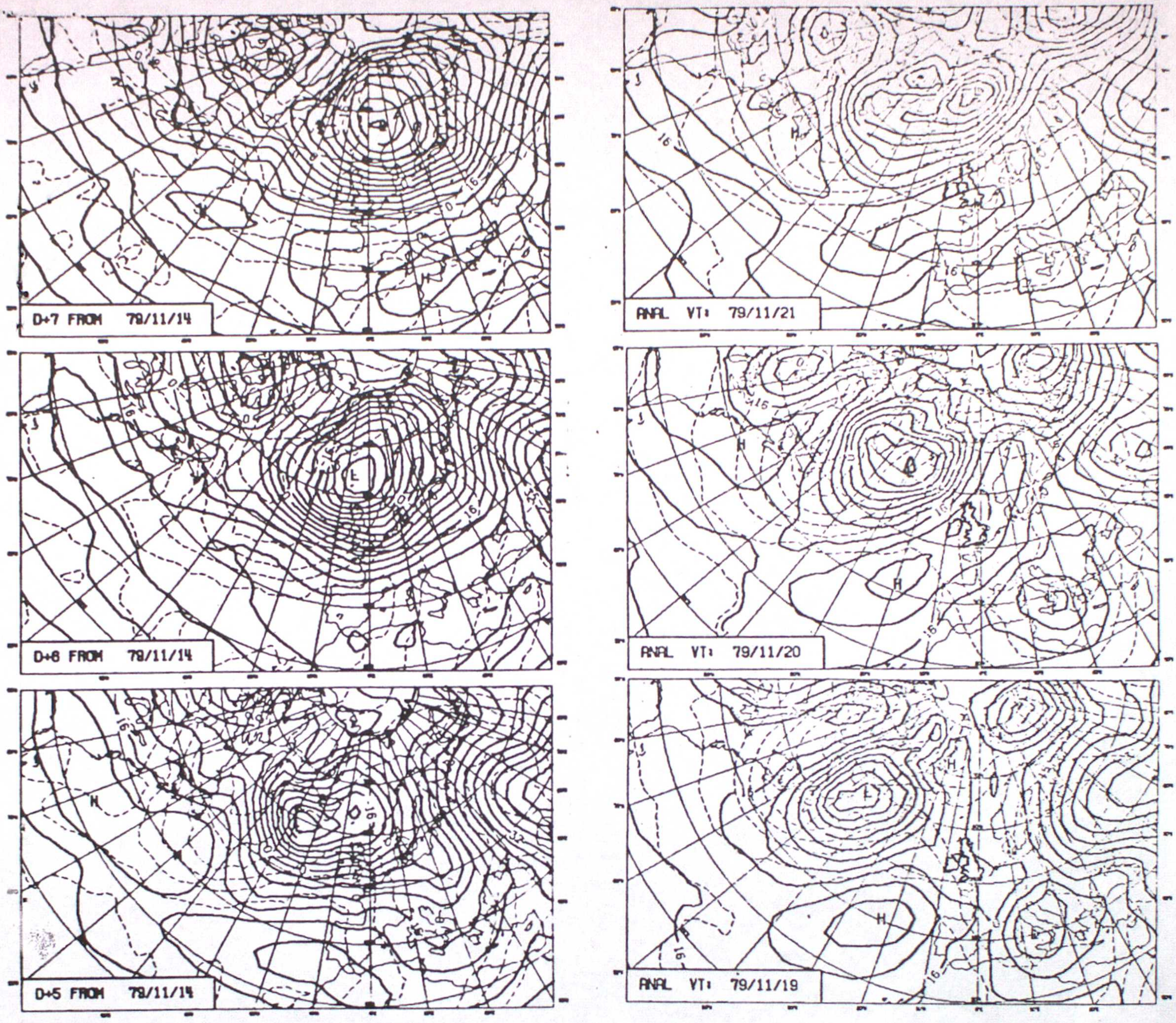


Fig. A1. (continued)

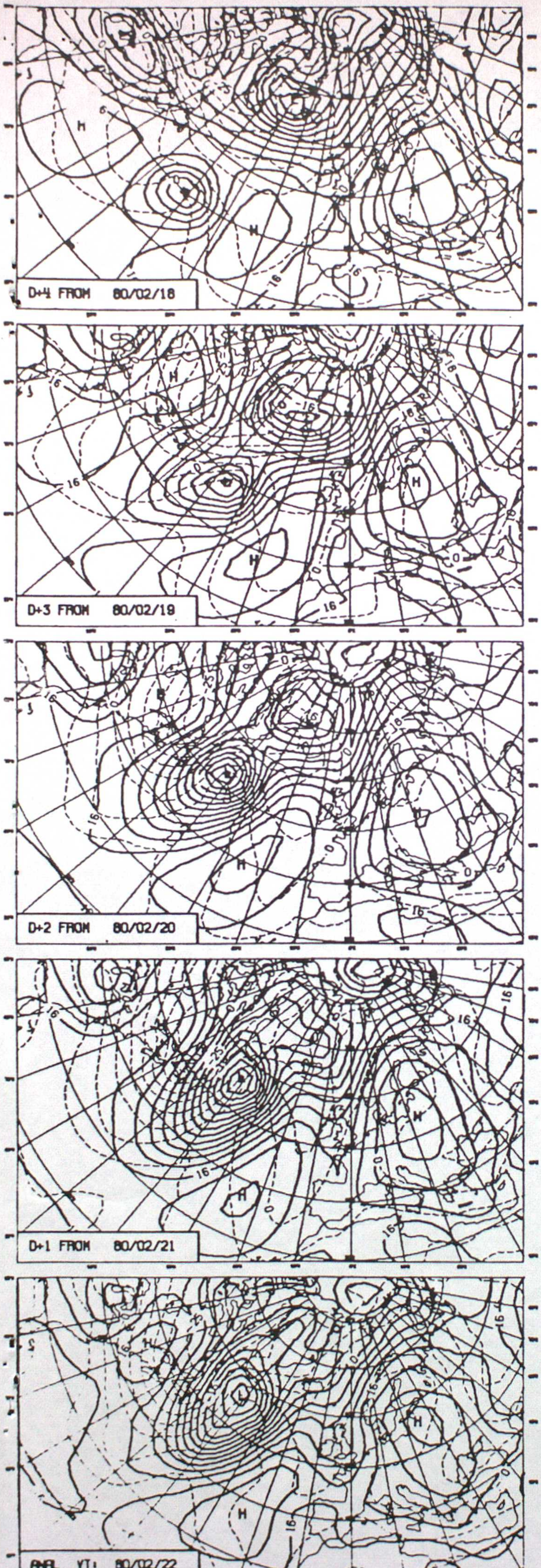
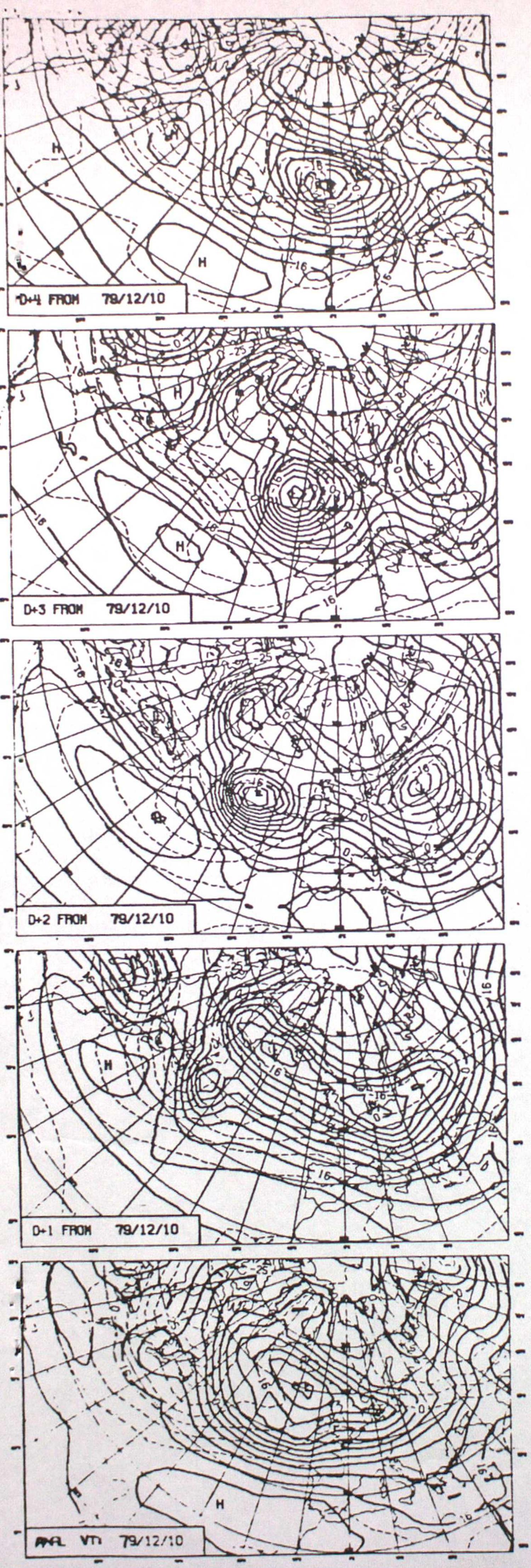


Fig. A2. ECMWF 1000mb analysis for 12 GMT on 22/2/80, and forecast fields verifying at that time.



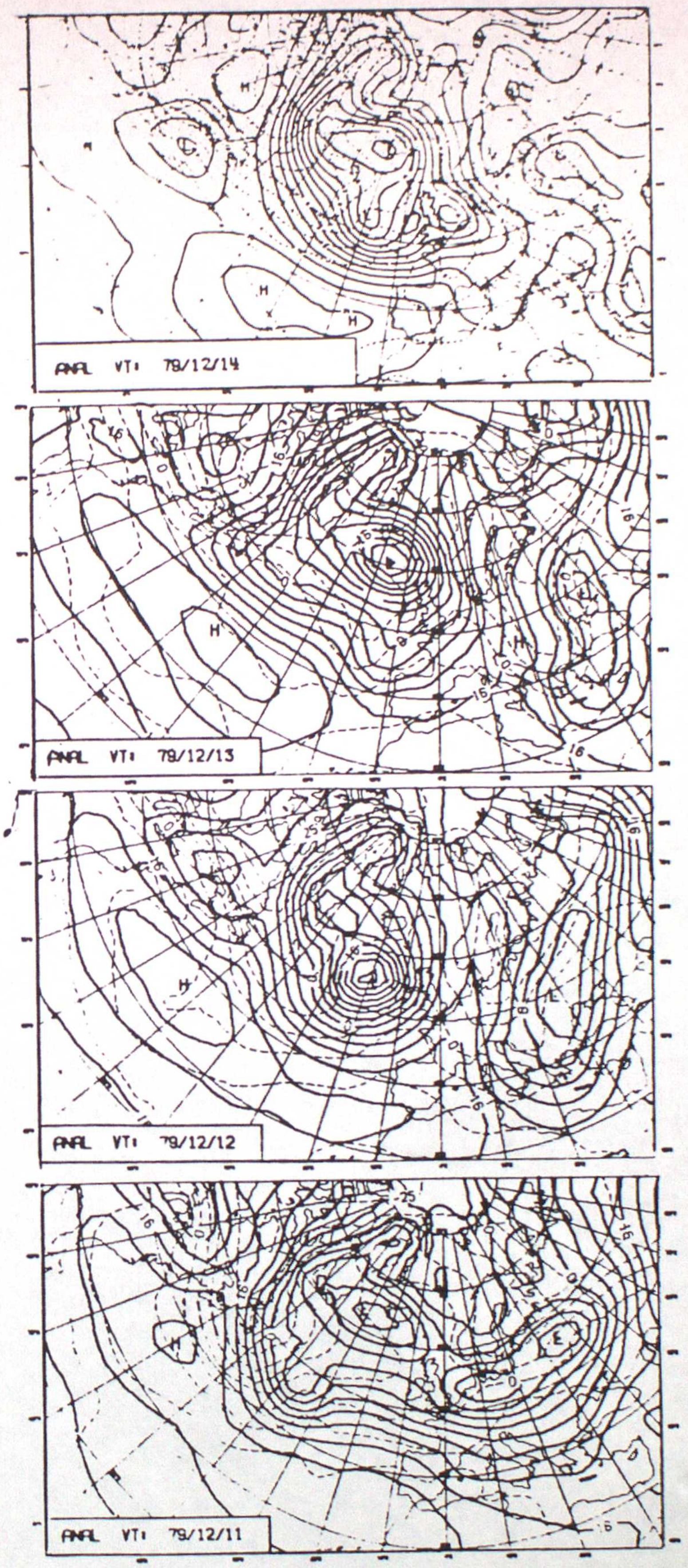
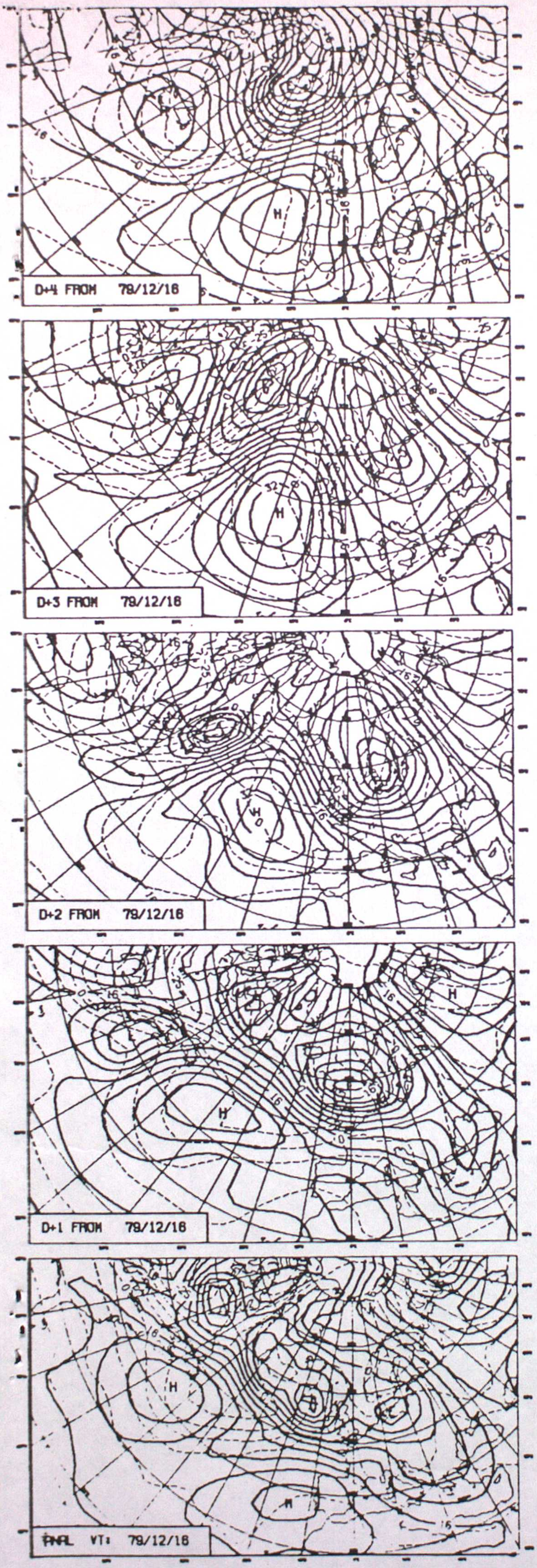


Fig. A3. ECMWF 1000mb analysis and forecast fields from 12 GMT on 10/12/79.



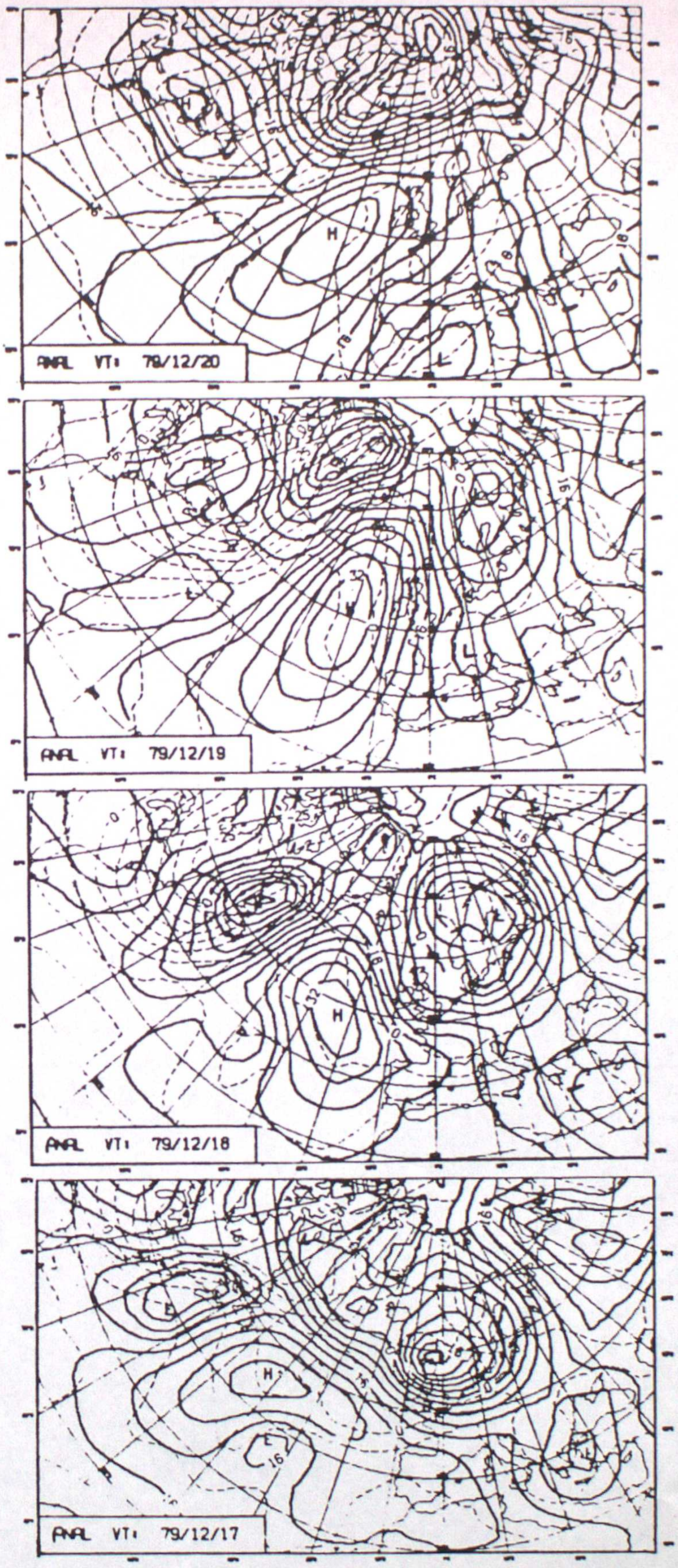


Fig. A4. ECMWF 1000mb analysis and forecast fields from 12 GMT on 16/12/79.

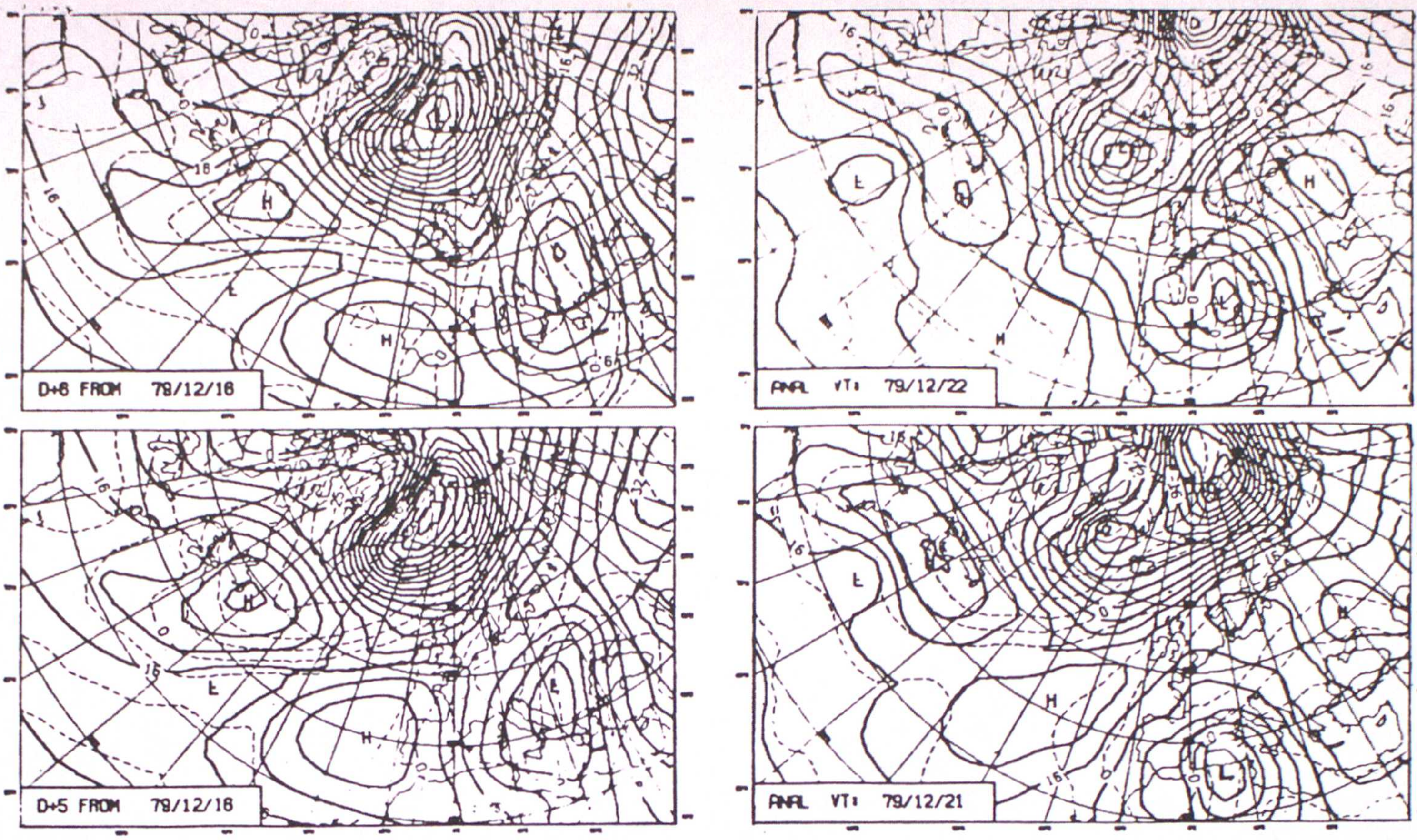


Fig. A4. (continued)

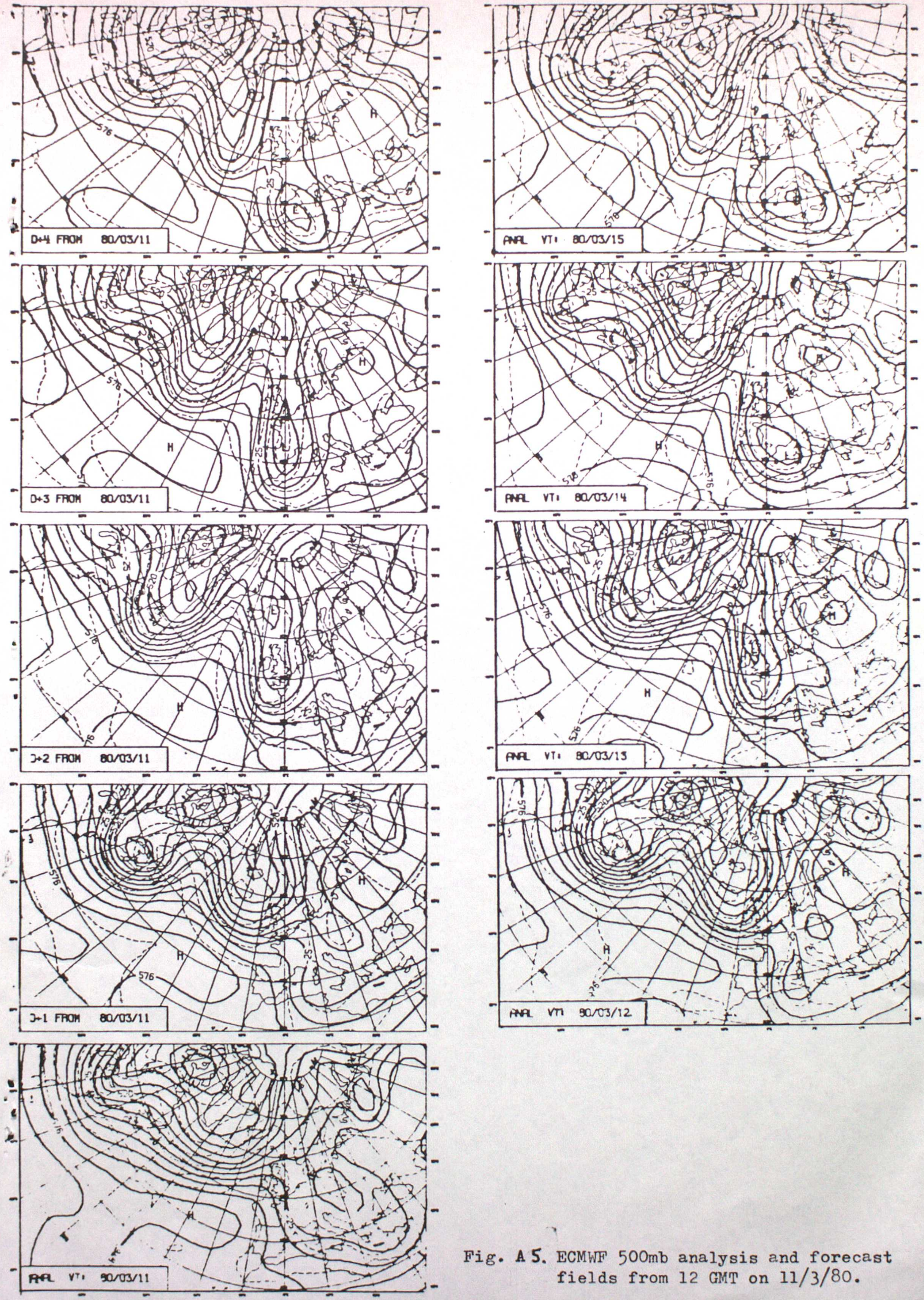


Fig. A5. ECMWF 500mb analysis and forecast fields from 12 GMT on 11/3/80.

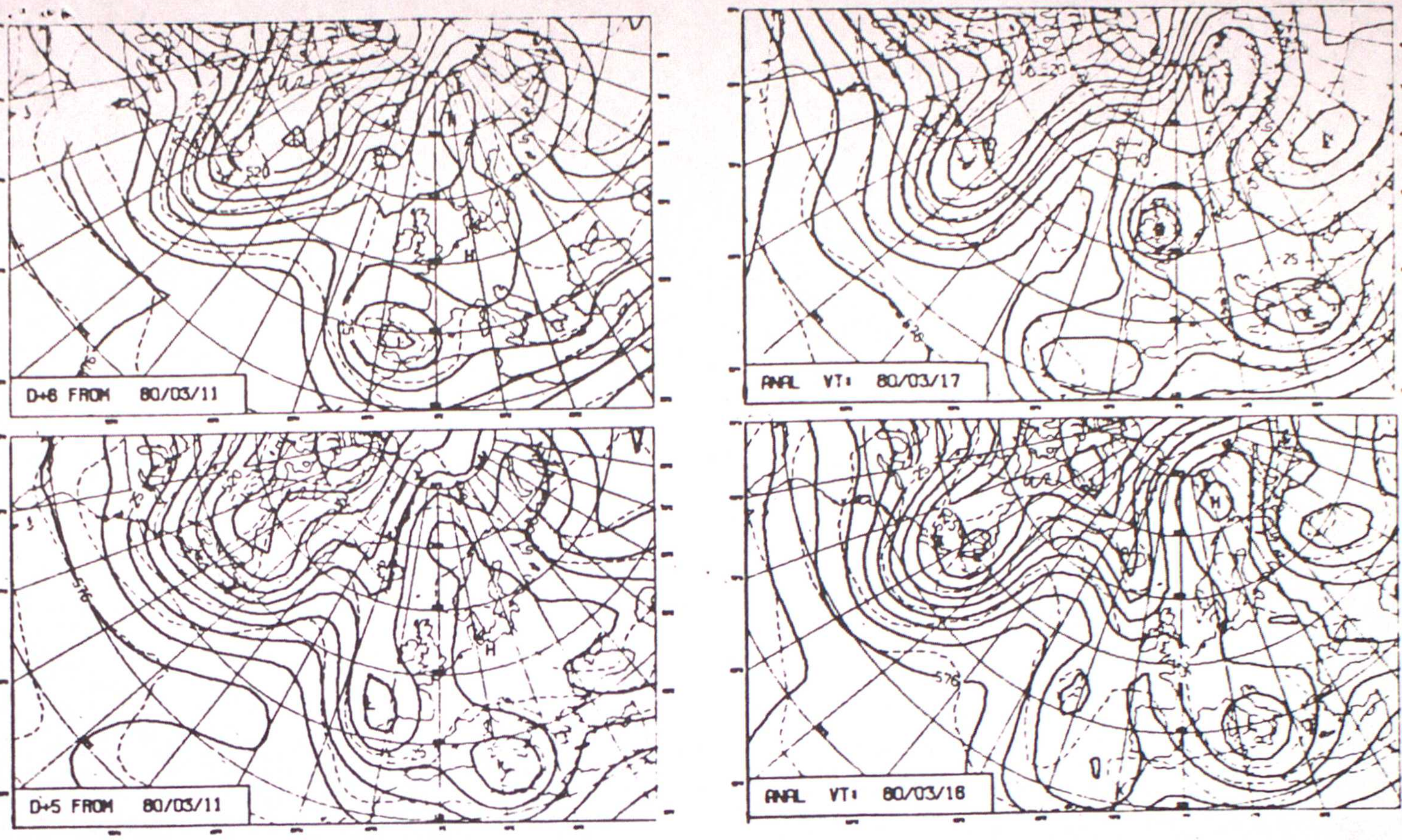


Fig. A 5. (continued)