TECHNICAL MEMORANDUM

138

30 day integrations using the operational ECMWF spectral model

S. Tibaldi, Č. Branković,
and U. Cubasch

Research Department

September 1987

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

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

1. INTRODUCTION

As a result of the ongoing efforts to improve the ECMWF's operational model climate, the Centre's model mean errors are now lower in amplitude and their growth is slower. It has therefore become necessary to diagnose them over a longer forecast range and, in particular, to diagnose their dependence upon season and model resolution. An experimental programme was therefore started to provide a suitable database of extended (30-day) integrations during different seasons and with a variety of model resolutions. Such a database provides not only a source of information for mean error studies, but it also affords the opportunity to examine the forecasting ability of the Centre's operational system in the extended range, and the dependence of forecast skill in this range upon different weather regimes and model characteristics.

This report describes preliminary results from an analysis of the first year of integrations. Section 2 describes the database, Section 3 is devoted to the analysis of the mean errors, while Section 4 discusses the objective skill scores of the extended range integrations. Section 5 attempts to summarize the main results and draw some tentative conclusions to form the basis for future work.

2. THE DATABASE

The set of experiments consists essentially of a monthly series of integration "pairs", started from two consecutive initial conditions, separated by 24 hours, around the mid-month period. The first month considered was April 1985 and the experiment is still ongoing. This report will be confined almost exclusively to the first year of integrations, from April 1985 to March 1986, inclusive. All integrations have been performed at four different model resolutions: T21, T42, T63 and T106 and with the operational 16-level, envelope orography version of the model, with the physical parameterization package as of March 1986. The shallow convection and modified Kuo convection parameterisations (Tiedtke et al, 1987) are included, but gravity wave drag (GWD, Miller et al, 1987) and modified surface exchanges (Blondin, 1987) are not. The second year of integrations (May 1986 to April 1987, T63 and T106 only) has recently been completed and will be evaluated in the future. The second year of experiments was run with the GWD 19-level version of the model. When completed, the comparison between the first two years will be of particular interest since it will provide an assessment of the impact of GWD on the model's climate drift. Some results from the first winter season of T106 runs (Oct 85 - March 86) were also discussed in Hollingsworth et al (1987).

3. THE MODEL MEAN ERRORS

This section will be devoted to the diagnosis of the model's mean errors as a function of forecast time (the so-called "climate drift") for all model resolutions and for two "extended" seasons, the October to March (OM) period (Northern Hemisphere "extended" winter), and the April to September (AS) period (Northern Hemisphere "extended" summer). The analysis discusses the zonally averaged fields, the mean mid-latitude fields (of both Northern and Southern Hemispheres (NH and SH)) and some fields of particular relevance to tropical regions. For a more comprehensive discussion on the climate drift of the operational model during the first ten days of integration, the reader is referred to Arpe (1987) and Brankovic (1986).

3.1 The zonally averaged mean errors

Fig 1 shows latitude-height cross-sections of the zonal mean of zonal wind for the ensemble means of the last ten days (days 21-30) of model integrations at the four different resolutions (top to bottom), plus analysed values for comparison. The left panels refer to the OM period, while the right panels

refer to the AS period. At first sight the model-simulated mean zonal flow field seems quite satisfactory. A closer examination of the difference fields for wind and temperature (Figs 2 and 3) shows, however, several important discrepancies. Although jet maxima and the general shape of the isolines of [U] and [T] are fairly well reproduced, their north-south positioning shows significant poleward shifts, with a fairly equivalent-barotropic structure, at least in extratropical regions.

Regarding the dependence of zonal mean errors upon model resolution, the T21 model stands out from the others. Its zonal climate is in some respect worse and in others better than those produced by the higher resolution models. The NH winter (OM) jet is better positioned and the tropospheric errors are smaller in the T21 runs, but the SH zonal flow and the tropical low-level easterlies are far too weak. Possible reasons for such discrepancies have been discussed by Miller et al (1987) who suggest that they are probably connected with a compensation of resolution-related errors between the eddy momentum flux convergence term and the orographically-related zonal-flow dissipation mechanisms (e.g. form drag). The T42 model climate has a somewhat "transitional" nature, with features resembling more the T63-T106 family, but also with features typical of the T21 behaviour, eg. in the SH stratospheric [U] error during the OM period, and in the intensity of the extratropical stratospheric [T] errors, in both hemispheres and in both seasons. The [T] and [U] errors for the highest resolution models are very similar both in amplitude and pattern, with the T63 model showing marginally smaller amplitude mean [U] errors.

We now discuss the T106 errors in some detail; the discussion applies equally to the T63 model and, to a lesser extent, to the T42 model as well.

Fig 4 shows the day 21-30 mean cross sections (for the OM period) of both the observed mean fields (top) and the T106 mean errors (bottom) of [U] (left panels) and [T] (right panels). For purposes of discussion, the zonally averaged atmosphere has been divided into three main areas: area 1, the tropical regions; area 2, the NH extratropics and area 3, the SH extratropics. In the troposphere the subtropical regions belong to the "extratropics" areas, while in the stratosphere they belong to the "tropics"

area. This is because in areas 2 and 3 the [U] mean errors are essentially equivalent-barotropic, as can be seen from the figures, while in area 1 they show marked changes with height. A comparison between the error fields and the observed fields shows that, in area 1 where the errors show a vertically varying structure, their sign is consistent with the hypothesis that their main effect is to smooth out the vertical shear of the zonal flow (positive errors where the mean flow is easterly and negative errors where the mean flow is mainly westerly). This is consistent with recent results of Miller and Viterbo (pers. comm.) that show that systematic errors (SEs) in tropical regions are sensitive to the intensity of the vertical diffusion used in the model, and that the tropical SEs can be reduced if the vertical diffusion algorithm is not applied above the planetary boundary layer. Fig 5 shows latitude-height cross-sections of the [U] error for two 10-day forecast experiments where free-atmosphere vertical diffusion was omitted and the two corresponding control experiments. Mean day 1-10 fields are compared and the reduction of [U] error in the tropical regions and, at upper levels, at all latitudes is evident. In areas 2 and 3, the vertical structure of the [U] errors is essentially equivalent barotropic, while the [T] errors are practically negligible in areas 2a and 3a (mainly the troposphere, see Fig 4) while they become sizeable (and mostly negative) in areas 2b and 3b (the stratosphere). The main characteristic of the [U] error in areas 2 and 3 is not so much an amplitude error as a poleward shift of the main jet maxima. To discuss the possible origin of this it is better to distinguish between the stratosphere, where both the radiative equilibrium towards which the model tends, and the top boundary condition, are likely to play a predominant role, and the troposphere, where we have to look for a mechanism capable of shifting the jet cores poleward.

We already know that GWD and an improved stratospheric vertical resolution (19 levels compared to 16 levels) can alleviate part of the problem (they are both absent in the experiments discussed here) but are not able to change the essential pattern (and, therefore, probably nature) of the errors. The treatment of mid-latitude baroclinic eddies and their essentially barotropic feedback on the mean zonal flow (via convergence of eddy momentum flux) is a likely candidate for investigation. The eddy kinetic energy decreases progressively in midlatitudes during the model integration, in both seasons and for all wavelengths, while the zonal KE marginally increases in the NH

winter for T63 and T106 models; Fig 6 shows the kinetic energy as a function of forecast time for the NH mid-latitudes - SH is not shown. Indeed, the changes in the eddy momentum flux are consistent with these ideas, showing a poleward shift of the convergence zone in the OM period, see Fig 7. Another interesting consequence of the elimination of the free-atmosphere vertical diffusion reported by Miller and Viterbo is to raise the level of eddy KE during the model integration, and to reduce (albeit to a much lesser extent than in tropical regions) the mid-latitude [U] mean errors as well. One possible link between vertical diffusion and the life cycle of baroclinic eddies is via its effect on atmospheric stability, which in turn affects the growth rate of the baroclinic waves.

Regarding the time evolution of the zonally averaged climate drift, Fig 8 shows the time-latitude diagrams of [U] error, vertically integrated between 1000 and 500 mb and between 300 and 30 mb for the OM period. The figure shows that the error growth is nearly linear in time and takes place mainly during the first 20 days of integration. During the last ten days of the integrations the error growth seems to approach an asymptotic level. The T21 model behaves differently from all other three resolutions, with smaller errors in the NH and larger errors in the SH (growing much faster in the first ten days).

Fig 9 shows the same quantities for the same vertical slabs as in Fig 8 but for the AS period (NH extended summer period). The [U] errors for the NH troposphere show a marked seasonal cycle and are at a minimum in this season. In the stratosphere, we observe an equally large seasonal cycle in the NH error, with a faster growth of the error during the first 10 days, and more marked latitudinal shifts of the error features.

A similar diagram (Fig 10) highlights a large seasonal cycle in the SH [T] stratospheric errors (positive in the AS period and negative in the OM period) and an almost complete absence of seasonal cycle in the errors in the NH and in the tropical stratosphere. In the troposphere the [T] errors are small (tropospheric diagrams are not shown, but see Fig 3). Also worthy of note are the similarities between the T21 and T42 [T] errors in both seasons, with the absence of the SH positive error area during the AS period that characterises the T63 and T106 integrations. This is probably due to an interaction between the radiation scheme and the model's resolution.

Returning to the idea of the life cycle of baroclinic waves affecting (and in turn being affected by) the onset of the SE and its growth rate, Fig 11 shows a Lorenz-type box diagram for the Northern Hemisphere energy cycle of both the observed atmosphere and the T106 model integrations for the OM period. The energy cycle is displayed separately for the first, second, and third ten day periods of the integrations. It is evident that the energy cycle of the baroclinic eddies is too intense in the first ten days. Athough the model still shows some drift in the second and third ten-day periods, most of the disruption of the energy cycle takes place during the first ten days, consistent with the duration of the life cycle of the baroclinic eddies.

3.2 The mean errors in the extratropical eddy fields

This section will be devoted to the extratropical mean errors of the model integrations and their dependence upon resolution and season. For the sake of brevity, the discussion will be limited to 500 mb height fields.

We first consider full fields. Fig 12 shows mean 500 mb height maps for the last 10 days of integration (day 21-30), for the OM period, for all four model resolutions (T21, T42, T63 and T106) together with the observed maps. The overall picture that emerges is a familiar one: the T21 model shows a much better simulation of the NH winter climate than the other (higher) resolutions but, conversely, heavily damps the strength of the westerlies in the SH during the same period, and also during the AS period (not shown). A possible explanation for this, connected to the representation of orography (and associated drag) and the convergence of eddy momentum flux has already been extensively discussed in Miller et al (1987).

The T42 model shows, again, a "transitional" behaviour, while the T63 and T106 models behave very similarly. We therefore concentrate the discussion mainly on the T106 model maps from here on. The main characteristic is the loss of amplitude of the large-scale planetary waves, especially in the NH and during the OM period, when their observed amplitude is larger. Figs 13 to 16 show the actual errors for all resolutions and for the three averaging periods day 1-10, day 11-20 and day 21-30. Errors for both OM and AS seasons and for both hemispheres are shown. Comparisons with similar maps characteristic of older versions of the ECMWF operational model (e.g. Wallace et al., 1983) demonstrate that the rate of growth of such errors has been reduced

considerably, so much so that typical 10-day SE amplitudes of the 1981-type operational model are now to be found around days 20 to 30. The general character and structure of the errors, however, has changed less. We still find a large north-south negative-positive dipole over the Central Pacific and a similar, weaker one, over the E-Atlantic and W-Europe, all corresponding to weakening of the planetary waves and strengthening of the westerly jet in mid-latitudes.

While, the (mainly negative) error maxima at high latitudes in the earlier operational integrations were dominated by a combination of wavenumbers 2 and 3, now wavenumber 1 and 2 seem to dominate. The low-latitude positive maximum over the Central Pacific (see Fig 1 of Wallace et al, 1983) is now broken into two weaker features. Since one still cannot discount the representation of orography as a possible contributor to such errors, it will be interesting to compare these results with those of the second year of experimentation, carried out with a model that includes parametrisation of GWD and with 19 levels. Results of diagnostic work on operational 10-day medium-range forecasts indicate that the general character of SEs is left unchanged by increased vertical resolution (Arpe, 1987).

The NH extratropical 500 mb mean errors show a very large dependence upon season, which is larger than the dependence upon model resolution (excluding T21); compare Fig 13 with 14. For the SH, although the seasonal dependence is much less marked, it is still larger than the effect of resolution, if we exclude again T21. The T21 climate of the SH suffers from particularly large deficiencies, as has been indicated above. The pole-to-equator geopotential height gradient is severely underestimated, together with the strength of the zonal flow. For a discussion of the possible causes, see again Miller et al (1987). It is, however, interesting to notice how the mid-latitude SH time-mean errors for the last 10-day period of the AS season seem to have somewhat decreased with respect to the preceding ten-day interval, indicating a much faster levelling-off time. This could be related to the much lower planetary-scale wave activity of the SH and to the consequently lower large-scale eddy mean-flow interactions.

Such differences can be put into perspective by comparing the error maps with the full field maps which shows that the error centres move with the seasonal (and hemispherical) displacements of the main troughs and ridges of the stationary eddies. This large dependence of the mean errors upon flow regime is reflected in the comparison between the mean errors and the RMS errors shown in Fig 17 (for the T106 and T63 model, day 21-30 only), together with the ratio of mean error variance to RMS error variance. The contribution of the mean error variance to the total error variance is quite modest in mid latitudes during the earlier forecast periods (not shown) but it is fairly large during the last ten-day period. This shows that, even within the same season, the time-mean errors depend considerably on the single forecasts realization and therefore, presumably, on the particular large-scale synoptic situation. This lends support to the idea that prognostic information about the forecast skill can be derived from the Groβwetterlage of both initial conditions and forecast evolution (Palmer and Tibaldi, 1987).

The similarity of the mean (and, to a lesser extent RMS) errors at T63 and T106 model resolutions is an important result, but it is not yet fully understood. Previous experience suggests that extrapolation of such behaviour to higher model resolutions is quite unjustified. On the other hand, if the speculations on the role of baroclinic eddies on the climate drift have any basis, this probably suggests that either further resolution increases or a fundamental change in physical parametrisation (further to envelope orography, GWD and even a possible reduction in free-atmospheric vertical diffusion) are needed to reduce the patterns shown in Figs 13 to 16.

3.3 The mean errors in the tropical regions

This subsection will be devoted to a diagnosis of the mean model errors in the tropical regions. Since the tropical dynamics are dominated by moist processes with approximate balance between heating and vertial velocity, we will start by looking at the field of velocity potential.

Fig 18 and 19 show the velocity potential (χ) for the AS period for the first ten days and the last ten days of integration. Both 850 mb and 200 mb levels are shown. Each figure shows the full χ field for T106 and T21 model resolutions and for the observed analysis. The T63 and T42 fields are very similar to T106. It is evident that all model resolutions represent the

low-level convergence pattern fairly satisfactorily during the first ten days. The same is not true for the divergence pattern at upper levels (Fig 18), where the general shape of the Walker circulation dipole is satisfactory but its strength is already severely underestimated during the first ten days of integration. During the last ten days (Fig 19) the tendency shown during the first ten days has been maintained and intensified, showing a low-level convergence pattern with an underestimated intensity (the excessively concentrated and intense maxima are mostly due to values extrapolated under high ground). The upper level divergence, furthermore, is very severely under-estimated, with maxima of the velocity potential less than 40% of the observed values (see also Heckley, 1985).

Since the low-level mass convergence in the model is not as poorly represented as the upper-level divergence, while mass is conserved, we must deduce that the upper level model divergence is not only underrepresented, but must have an erroneous vertical structure.

Fig 20shows the vertical profile of divergence integrated over an area containing the Indonesian maritime continent, as shown in the inset. This figure was derived from operational 10-day forecasts for the June 1986-August 1986 period and shows instantaneous profiles for day 10 forecasts and verifying analyses. Similar figures will be available for our set of 30-day forecasts. The present figures show that the day 10 upper-level divergence is not only underestimated, but has an erroneous vertical profile, with the zero level divergence as low as 850 mb (compared to the analysed 450 mb) and with a downward displacement of the upper-level maximum of 50 mb, together with a 40% reduction of its intensity, thus confirming our diagnosis.

The general weakening of the convectively driven Hadley (and Walker) circulations shown by the model is confirmed by the latitude-height cross sections of Fig 21 which show the $[\bar{\omega}]$ and $[\bar{v}]$ full fields for the T106 model and for the verifying analyses. The main characteristics of the model mean errors in these fields are: the circulation is substantially weaker, the cells are displaced polewards in both hemispheres and the vertical structure of the upper-level Hadley return flow is spread out, instead of being

concentrated around the 200 mb level (see above). The poleward shifts of the meridional circulation cells (consistent with the poleward shift of the jetstreams) is much larger during the OM period (for both hemispheres) and weaker in the AS period (and NH only).

The model's tropical atmosphere progressively dries out during the 30 days of forecast time. Fig 22 shows latitude-time cross-sections of [q] error, vertically integrated in the 1000-700 mb layer. The drying effect is quite evident for both seasons and is larger in the AS period. It is very similar for the T106, T63 and T42 model resolutions (T106 and T21 only shown); for the T21 it is somewhat smaller in both periods while a weak moistening of the NH mid-latitudes becomes evident during the latter part of the T21 integrations. The general tendency to dry out the tropical troposphere could be due to a variety of reasons, not excluding possible inadequacies of the moisture analyses. Since 1985, considerable improvements have been made to the moisture analysis scheme (Illari, 1985 and Lönnberg, personal communication) and it will be interesting to see whether measurable changes will be evident in the second year of integrations. The most natural candidate for investigation is however the moist convection scheme. It should, furthermore, be remembered that the tropical SEs have recently been shown (as mentioned in Section 3.1) to be sensitive to the level of free-atmospheric vertical diffusion (currently considered to be excessive). This could account for a proportion of the overall model weaknesses in representing the tropical water cycle and for part of the loss of eddy kinetic energy that is typical of the model's tropical dynamics, see Fig 23. Exactly how the process is started and how much is caused by either the vertical diffusion or the moist parametrisation scheme (or on some other model component, e.g. the radiation scheme) is currently under investigation.

4. OBJECTIVE SKILL SCORES AT THE EXTENDED RANGE

We now turn to the objective evaluation of the skill of the model integrations, considered as deterministic extended range forecasts. We start by briefly reviewing the skill scores used. We then show the daily skill scores and the ten-day mean scores. This work is still in progress and at the time of writing only the T106 and the T63 model integrations had been evaluated in this way. It is hoped that, by the time of the SAC meeting,

results from all four resolutions will be available. It can, however, be expected (and it is being confirmed by the ongoing work) that the T21 model stands out as the poorest forecasting model at all time scales up to a month, and that the T42 performance resembles more closely those of the T63 and T106.

Only hemispheric scores will be discussed here, but work is underway to also evaluate the model's performance on smaller limited areas of the NH. For a more comprehensive comparison of these results with similar sets of experiments (e.g. Miyakoda et al, 1986 and Mansfield, 1986) the reader is referred to Hollingsworth et al, (1987).

4.1 Definition of the skill scores

Two quantities, anomaly correlation coefficient (ACC) and RMS error, are usually employed to evaluate objectively the success of numerical forecasts. We have calculated both such quantities for all resolutions considered in our 30-day forecast set and for the following meteorological parameters: geopotential height at 1000, 500 and 300 mb, and temperature at 850, 500 and 300 mb. We will, however, limit our discussion to 500 mb height and 850 mb temperature.

Both skill scores have been computed on T42 spectrally truncated fields evaluated on a 3.75° regular latitude/longitude grid. Hemispheric scores are defined over the area enclosed between 22.50° and 86.25°N. In the following subsections the scores are computed on daily fields, 10-day mean fields or 30-day mean fields and the performance of the model will be discussed in terms of NH skill scores and their dependence on resolution and season.

4.2 The daily scores

Fig 24 shows RMS error and Fig 25 ACC daily scores for 850 mb temperature (bottom diagrams) and for 500 mb height (top diagrams) and for both OM (left panels) and AS (right panels) periods. In addition to the T63 and T106 RMS error (or ACC) scores, the diagrams also show the same scores for persistence of full fields. The model scores, representing the average of 12 model integrations for each period, are computed over the Northern Hemisphere and

are averaged within the sample of 12; the Fisher z-transform (e.g. Morrison, 1983) is used to average the correlations. Two further curves show the departures between successive T106 or T63 forecasts, with initial conditions separated by 24 hours (the "forecast spread"). These curves are, therefore, the mean over 6 pairs and are computed between forecast verifying at the same time. In the RMS error diagrams, the horizontal dashed line represents the climate norm for the corresponding field. In the ACC error diagrams, the 60% level is indicated by the dashed line. It should be noted that, with a perfect model, not only persistence but also forecast RMS error (and RMS forecast spread) should asymptote to $\sqrt{2}$ times the climate norm (e.g. Miyakoda et al., 1972). Indeed persistence does asymptote to approximately the expected value, but both forecast error and, even more so, forecast spread asymptote to a much lower level. The fact that the model has systematic errors will tend to increase the RMS forecast error asymptote but will have no effect on the RMS forecast spread. Conversely, the fact that the model has less time variability than the real atmosphere will affect both RMS error and RMS spread, by lowering both their asymptotic values (e.g. Hollingsworth et al., 1987). Simple considerations about the asymptotic levels reached by the various quantities plotted in the diagram of Fig 24a (see Appendix) suggests that, for daily fields, the variance of the "winter" 500 mb height mean model error is about $(50m)^2$ and that the transient variance of the models (both T63 and T106) is approximately 35% lower than observed. By and large, this means that the lack of transience is the dominating effect (over the mean model error) in determining the asymptotic level of the model RMS error, for both daily and 10-day mean fields. We will see later how such considerations carry through to 10-day mean RMS errors.

An analysis of Fig 24 reveals that the complete loss of predictability of 500 mb geopotential height (measured as the time it takes to the RMS error to reach its asymptotic value) is reached by both models around day 12, with little or no difference due to either model resolution or season. The T106 model appears marginally worse than the T63 in the OM period and marginally better during the AS period. There is also some evidence that the T106 model has slightly higher mean errors (see also Section 3.2) and slightly lower loss of transience than the T63. The time at which the model RMS error reaches the climate error (around day 7 to 8) also coincides fairly well with the time it

takes for the ACC to reach .6 (around day 6 to 7) - consistent with medium-range forecasting experience (see Fig 25). The daily instantaneous 850 mb temperature also shows predictability characteristics very similar to the 500 mb height.

We should try to put such results on objective skill scores into perspective. Since we know from other, more extensive, independent comparison studies (e.g. Simmons, 1987) that T106 is measurably more skillfull than T63 in the medium range, irrespective of season, we must conclude that our sampling is too small. An indirect confirmation that there is a sampling problem is to be found when comparing the T106 RMS spread curve of Fig 24a with a similar curve evaluated on a much larger sample of 100 medium range forecasts performed in the same period and using the same model, see Hollingsworth et al (1987). The two curves are quite different, the one shown here growing much faster. Work is also underway to establish whether the forecast spread between the two successive forecasts contains any indication of the likely forecast skill (see Palmer and Tibaldi, 1987).

4.3 The ten-day mean scores

Figs 26 and 27 are the exact counterparts of Figs 24 and 25, but for 10-day mean fields. Some interesting aspects of model behaviour may be discerned. Firstly, the predictability or, more precisely, the "forecastability" of the low-frequency part of the atmospheric behaviour is only marginally higher than that of the total fields. The 500 mb height RMS error crosses the climate norm only 1 to 2 days later than for instantaneous fields. What this implies for atmospheric predictability is open to speculation. The substantial equivalence of the two T63 and T106 models in terms of extended-range objective skill scores is also confirmed, together with their relative opposite behaviour in the two "seasons".

An obvious change of behaviour can be seen in the relative differences between the T106 and T63 forecast spread curves. The T106 now shows a considerably larger variability than the T63 model, indicating that most of the difference in variability shown in the total curves of Fig 24 was due to low frequency rather than day-to-day variability. This is only true, however, for the OM

period and not for the AS period. Moreover the effect disappears towards the end of the integration period. It is, therefore, likely that sampling again has to be held responsible for such a difference (in this case the sample has only six members).

Interestingly, the 850 mb temperature ten day means appear to be measurably more predictable than 500 mb height (ten days to cross the climate norm, as opposed to eight for 500 mb height). This effect is only present in the OM period (extended NH winter) and does not occur during the AS extended NH summer. The mid-latitude ocean surface thermal forcing is much larger in winter than in summer and so could be a possible reason for the increased predictability of the low-frequency variability of 850 mb temperature. Forcing from the time-evolving land surface temperature could also contribute to this effect since it is a model prognostic variable. One might expect, however, such effect to be larger in summer than in winter. It is worth noticing that the asymptotic values for 850 mb temperature ACC (for both OM and AS periods) is considerably higher than zero, unlike, for example for 500 mb height. This confirms similar results reported in Molteni, et al (1987). It should, however, be noticed that the same advantage of 850 mb temperature over 500 mb height ACC is also noticeable for persistence. It cannot be seen for RMS error because persistence of full field has been used (new persistence scores using persistence of anomalies, showing much lower persistence errors, will be soon available for all experiments).

Using the results of the Appendix, and the asymptotic values of Fig 26a, it can be seen that the estimated SE variance is $(38m)^2$ and the low-frequency model transient variance is 44% lower than observed. This indicates that the loss of transient eddy activity experienced by the model affects low-frequency variability even more than high-frequency variability.

4.4 The variability of the forecast skill

The variability about the mean skill scores described above is high. Fig 28 shows the time series of the ACC throughout the year of the experiments, for both T63 and T106 models and for persistence. Four diagrams are shown, for the day 6-15, 11-20, 21-30 and 1-30 mean field scores respectively. Single dots indicate the skill of single experiments, while the curves are drawn

through the mean, for each month, of the two 30 day forecasts initiated 24 hours apart.

Together with the progressive degradation of the forecast skill with time, the figures show wide fluctuations in skill from day to day and from month to month. There is little or no evidence of a seasonal cycle in the skill in contrast to the experience with medium range forecasting. Up to day 11-20 the models' skill is above persistence (see Figs 27a and 27a) as also the 30-day mean scores. How much this is due to the contribution to this skill coming from the first 10 to 20 days of integration is currently under investigation and it is quite possible that by taking the first 15 or 20 days of forecast as an indication of the entire 30 day period will yield considerably improved 30 day forecasts (Roads, 1986).

There are some particularly good forecasts (with 30-day mean ACC in excess of .6, e.g. July, October, November, February and March), but also some particularly poor forecasts (e.g. January 1986). In particular, the major failure of January 1986 corresponds to the onset of the Euro-Atlantic block during early February 1986. Figs 29 and 30 show, as examples, the observed and forecast 500 mb height full fields and corresponding anomalies from two examples, the "good" October case and the "bad" January case. There is no doubt that if one could identify good cases a-priori with some degree of success then forecasts such as the one for October (a T106 case) could provide exceptionally useful guidance to operational forecasters.

The practical value of 30-day forecasts would be vastly enhanced if we could determine a-priori which forecasts are likely to display above average skill. Investigations already under way for medium-range forecasts, Palmer and Tibaldi (1987), will be extended to monthly forecasts in the near future employing a variety of methods, including the Lagged-Average-Forecasting Technique to relate skill with forecast spread (e.g. Molteni et al., 1987). It will be of particular interest to see whether the flow-type skill relationship, found for the medium-range by Palmer and Tibaldi (1987), can be successfully extended to longer forecasting periods. It would also be desirable to find ways to take explicit account of the probability of formation of quasi-stable structures, such as blocking, that are still often not captured by the model.

5. PRELIMINARY CONCLUSIONS AND PLANS FOR FUTURE WORK

The dataset of 30-day integrations at various model resolutions being accumulated at the Centre has allowed us to evaluate the climate drift of the operational model as a function of resolution and season. Furthermore, the same database can be used to investigate the forecast skill of the model beyond the currently accepted average limit for deterministic predictability. A few results are available to date, but most of the diagnostic and evaluation work is still in progress.

The early results reported here are fairly consistent with previous work on both systematic errors and on extended range predictability and can be summarized as follows:

- The zonal mean and mid-latitude eddy systematic errors show a strong dependence upon resolution at the lower end of the resolution spectrum (T21-T42). At the upper end (T63-T106) the dependence upon resolution is weak and shows indications of levelling off.

- Mid-latitude SEs (and time mean errors of a single realisation) are largely dependent upon season and large scale Groβwetterlage, indicating that some information about the initial conditions is retained almost throughout the integration-time, although not necessarily of a nature that can be exploited in practical forecasting. This shows promise for predictability (forecast skill) studies.

- The zonal mean tropospheric SEs in mid-latitudes can be mainly described as a poleward displacement of the main jet maxima, with an equivalent barotropic structure.

- The tropical SEs (and the mid-latitude stratospheric SEs) can be mainly described as a progressive loss of the observed vertical shear of the zonal wind (and thermal wind balance related temperature errors). Although the main contributions to such errors are probably to be found in the moist convection and radiation schemes, the structure of the errors is also consistent with the hypothesis that too much free-atmosphere vertical diffusion of momentum (and heat) is taking place.

- The mid-latitude SEs of the eddy flow can be mainly described as a progressive loss of stationary and transient eddy activity, with the model 500 mb height time mean errors having a winter asymptotic variance of about $(50\ m)^2$ and a transient variance more than 35% lower than observed for daily fields and 44% lower than observed for 10-day mean fields.

- Both the time-mean circulation and the eddy activity are much weaker than observed in tropical regions and the vertical structure of the (mainly diabatically forced) divergence field is erroneously represented by the model. Because of the strong feedbacks between the mean flow and the eddies it is very difficult to partition, even qualitatively, the responsibility for such loss between the moist convection parameterisation schemes and the other model characteristics.

- The model tropical atmosphere appears to dry out considerably during the extended integrations, for all resolutions and all seasons, but due to uncertainties on tropical moisture analysis, more in-depth studies are needed to put this result into proper dynamical perspective.

- The T106 daily 500 mb height RMS error during NH "extended winter" reaches it asymptotic value around day 12 ("complete" loss of instantaneous deterministic skill). Very similar values (relative to the respective climatic variances and, therefore, asymptotic values) are attained by both 500 mb geopotential height and 850 mb temperature RMS error, for both seasons. The effect of taking 10-day means of forecast field is to increase very marginally (1-2 days) the perceived "forecastability" of 500 mb height but of 2-3 day that of 850 mb temperature, possibly indicating a measurable beneficial effect of surface thermal forcing on atmospheric predictability.

- RMS forecast error of both daily 500 mb height and 850 mb temperature reaches the climate norm around day 7-8, by which time the anomaly correlation coefficient has dropped to about 60%. These values are only slightly increased by 10-day mean averaging the fields with a somewhat better impact on 850 mb temperature. This, together with the

comparisons with persistence forecasts (persistence, however, of full fields and not of the anomaly, which is a better zero-cost forecast) indicates obviously useful information on average until day 7-10, with some potentially useful information in ten-day mean fields until the day 6 to 15 averaging period.

- The 30-day mean forecast fields are almost consistently better than persistence, but how much this depends on the first 10 to 15 days of integration remains to be more accurately evaluated.

- The variability of extended-range forecast skill is large; larger even than medium-range skill. A major difficulty is undoubtedly the forecasting of blocking onset beyond the first few days of model integration time. The possibility of estimating a priori the skill of a forecast might be vital if extended range forecasts are to move to the forefront of operational weather prediction.

Many areas remain to be explored. More diagnostic work is needed to identify possible causes of the climate drift and to document the climate drift in model parameters such as precipitation. More effort needs to be put into diagnosing the drift and its possible causes in the Southern Hemisphere and in comparing the two hemispheres. In the area of objective skill scores, the full set of resolutions needs to be documented and the sample needs to be enlarged, the scores specialized to limited areas (for which even larger samples are required) and the performance of persistence reassessed (using persistence of anomalies) in the light of the importance of the seasonal cycle in 30-day integrations. The skill in quantitative precipitation forecasting also needs to be explored. On the topic of forecast skill variability, much more work needs to be done to assess whether extended-range forecast skill is predictable. To this end a number of T63 Lagged-Average-Forecast experiments have been run (ensembles of time integrations started from initial conditions 6 hours apart) and remain to be evaluated.

Furthermore, analysis of the second year of integrations will provide us with an interesting comparison between 16 levels, no gravity wave drag and 19 levels, gravity wave drag T106 models, and their impact on model climate drift and extended forecasting skill, although interannual variability effects will be difficult to evaluate.

Acknowledgements

Frequent and stimulating discussions with F Molteni, particularly on the content of the Appendix, and with K Arpe are gratefully acknowledged.
A Simmons, A Hollingsworth and T Palmer carefully read the manuscript and contributed to improving it considerably.

References

Arpe, K, 1987: Impact of changes in the ECMWF analysis-forecasting scheme on the systematic error of the model and on analysis data. To be published.

Blondin, C and H Böttger, 1987: Surface processes. ECMWF Tech Memo No 135. ECMWF, Reading, UK.

Brankovic, C, 1986: Zonal diagnostic of the ECMWF 1984-85 operational analyses and forecasts. ECMWF Tech Rep No 57, 72 pp.

Heckley, W A, 1985: Systematic errors of the ECMWF operational forecasting model in tropical regions. Quart J R Met Soc, 111, 709-738.

Hollingsworth, A, U Cubasch, S Tibaldi, C Brankovic, T N Palmer and L Campbell, 1987: Mid-latitude atmospheric prediction on time scales of 10-30 days. In "Variability in the atmosphere and oceans", Roy Met Soc Monograph.

Illari, L, 1985: The quality of ECMWF humidity analysis. ECMWF Workshop on High Resolution Analyses. 24-27 June 1985. Reding, UK. Available from ECMWF.

Mansfield, D. A., 1986: The skill of dynamical long-range forecasts, including the effect of sea surface temperature anomalies, Quart J R Met Soc, 112, 1145-1176.

Miller, M., T.N. Palmer and R. Swinbank, 1987: Orographic gravity wave drag: its parametrization and influence in general circulation and numerical weather prediction models. Submitted to Met and Atmos Phys.

Miyakoda, K., G.D. Hembree, R.F. Strickler and I. Shulman, 1972: Cumulative results of extended forecast experiments. I. Models performance for winter cases. Mon Wea Rev, 100, 836-855.

Miyakoda, K., J. Sirutis and J. Ploshay, 1986: One-Month Forecast Experiments. Mon Wea Rev, 114, 2363-2401.

Molteni, F., U. Cubasch and S. Tibaldi, 1987: Monthly forecast experiments with the ECMWF spectral models. Proceedings of the ECMWF Workshop on predictability in the medium and extended range. ECWMF, Shinfield Park, Reading, U.K.

Morrison, D.F., 1983: Applied linear statistical methods. Prentice-Hall, Englewood Cliffs, N.J., 562 pp.

Palmer, T.N. and S. Tibaldi, 1987: Predictability in the medium and extended range. ECMWF Tech Memo No 139.

Roads, J O 1986: Forecast of time averages with a numerical weather prediction model. J Atmos Sci, 43, 871-892.

Simmons, A. J., 1987: High resolution experiments. ECMWF Tech Memo No 136.

Tiedtke, M, W A Heckley, J Slingo, 1987: Tropical forecasting at ECMWF: on the influence of physical parametrisation on the mean structure of forecast and analyses. To appear in Quart J R Met Soc.

Wallace, J.M., S. Tibaldi and A.J. Simmons, 1983: Prediction of systematic forecast errors in the ECMWF model through the introduction of an envelope orography. Quart J R Met Soc, 109, 683-717.

Appendix: A quantitative estimate of the model's transient variance

An estimate of the asymptotic values of RMS forecast error and RMS forecast spread (as can, for example, be obtained from Figs 24a and 26a for daily fields and 10-day mean fields respectively) together with the knowledge of the climate norm, makes it possible to estimate the mean hemispheric variance of the model's systematic error and the proportion of observed transient variance represented by the model (e.g. Hollingsworth et al, 1987).

If we define

A: analysed anomaly field

F: forecast anomaly field

$E \equiv F-A$: forecast error field

$S \equiv F_1 - F_2$: forecast spread field

$\overline{} \equiv$ expectation operator (ensemble mean)

$v^2 \equiv \overline{\|A\|^2}$: total climatic variance

and consider a large enough ensemble, for which

$\bar{A} = 0.$

If the model has a systematic error,

$\bar{F} \neq 0,$

and we can define

$F' = F - \bar{F}.$

Let us also consider a forecast time at which there is no more predictability (i.e. the RMS has reached its asymptotic value). Then

$$\overline{A \cdot F'} = 0 \; ; \quad \overline{F_1' \cdot F_2'} = 0$$

We then define forecast error and forecast spread variances respectively as:

$$\varepsilon^2 = \overline{\|E\|^2} = \overline{(F-A) \cdot (F-A)} = \overline{\|F\|^2} + \overline{\|A\|^2} + \overline{2F \cdot A}$$

$$\sigma^2 = \overline{\|S\|^2} = \overline{\|F_1 - F_2\|^2} = \overline{\|F_1' - F_2'\|^2} =$$

$$= \overline{\|F_1'\|^2} + \overline{\|F_2'\|^2} - 2\,\overline{F_1' \cdot F_2'} = \overline{2\,\|F'\|^2}$$

But

$$\overline{\|F\|^2} = \overline{(\bar{F} + F') \cdot (\bar{F} + F')} = \overline{\|\bar{F}\|^2} + \overline{\|F'\|^2} + 2\overline{\bar{F} \cdot F'} =$$

$$= \|\bar{F}\|^2 + \overline{\|F'\|^2}$$

that is, the perceived model variance is the sum of a "true" part, $\overline{\|F'\|^2}$, that represents real time-transience, and a "false" part, $\|\bar{F}\|^2$, that only reflects the existence of a model mean error. Of course, we also know that the "true" model variance is likely to be smaller than the observed one. It is, therefore, interesting to evaluate their ratio.

We can see that

$$\overline{F \cdot A} = \overline{(\bar{F} + F') \cdot A} = \bar{F} \cdot \bar{A} + \overline{F' \cdot A} = 0$$

So, if we define

$\alpha^2 = \overline{\|F'\|^2} \,/\, \overline{\|A\|^2}$ ratio between "true" model variance and observed variance

and

$\beta^2 = \|\bar{F}\|^2 \,/\, \overline{\|A\|^2}$ ratio between mean square amplitude of systematic error ("false" model variance) and observed variance,

we have

$$\varepsilon^2 = v^2 (1 + \alpha^2 + \beta^2)$$

$$\sigma^2 = 2v^2\alpha^2$$

Concluding:

$$\alpha^2 = \frac{\sigma^2}{2v^2} \qquad \text{and} \qquad \beta^2 = \frac{\varepsilon^2}{v^2} - 1 - \alpha^2$$

If we now evaluate approximately ε and σ from Fig 24a (500 mb height RMS NH "winter" errors for daily maps) and from Fig 26a (500 mb height RMS NH "winter" errors for 10 day mean maps) we have, for daily values:

$v = 110$m
$\varepsilon = 150$m
$\sigma = 125$m

and therefore:

$\alpha^2 = .65$
$\beta^2 = .21$
$\|\bar{F}\| = \beta v = 50$m

and, for 10-day mean values

$v = 80m$		$\alpha^2 = .56$
$\varepsilon = 107m$	and therefore:	$\beta^2 = .23$
$\sigma = 90m$		$\|\bar{F}\| = \beta v = 38m$

β^2 is statistically significant at the 95% confidence level if it exceeds $2/n$, where n is the sample size, in our case 12. β^2 has, therefore, to exceed .17 to be highly significant.

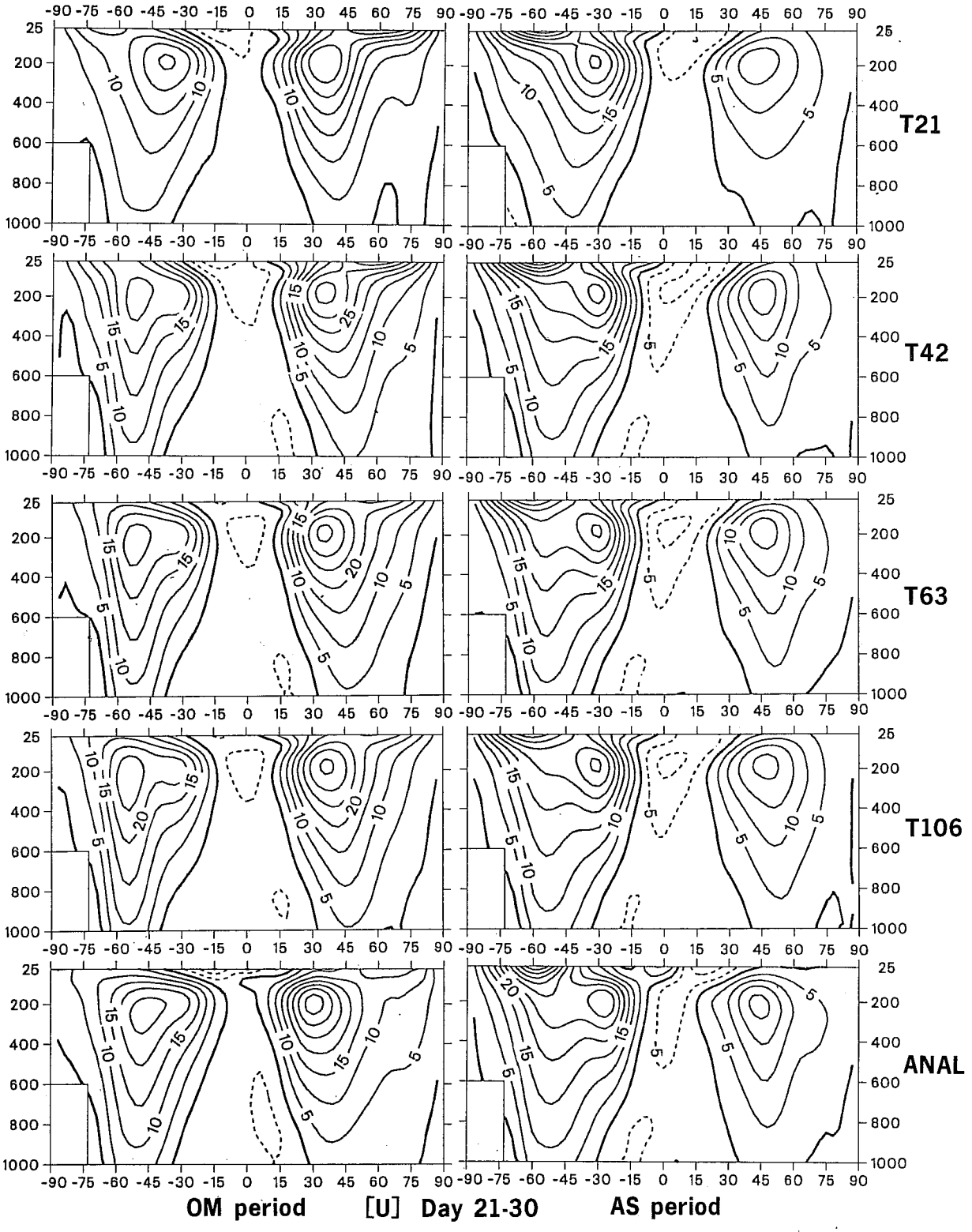

Fig 1 Latitude-height cross-sections of zonal mean of zonal wind [U]. Day 21-30 time mean. Top to bottom: T21, T42, T63, T106, analysis. Left: OM period; right: AS period. 12 cases in each period. Contours every 5 ms^{-1}.

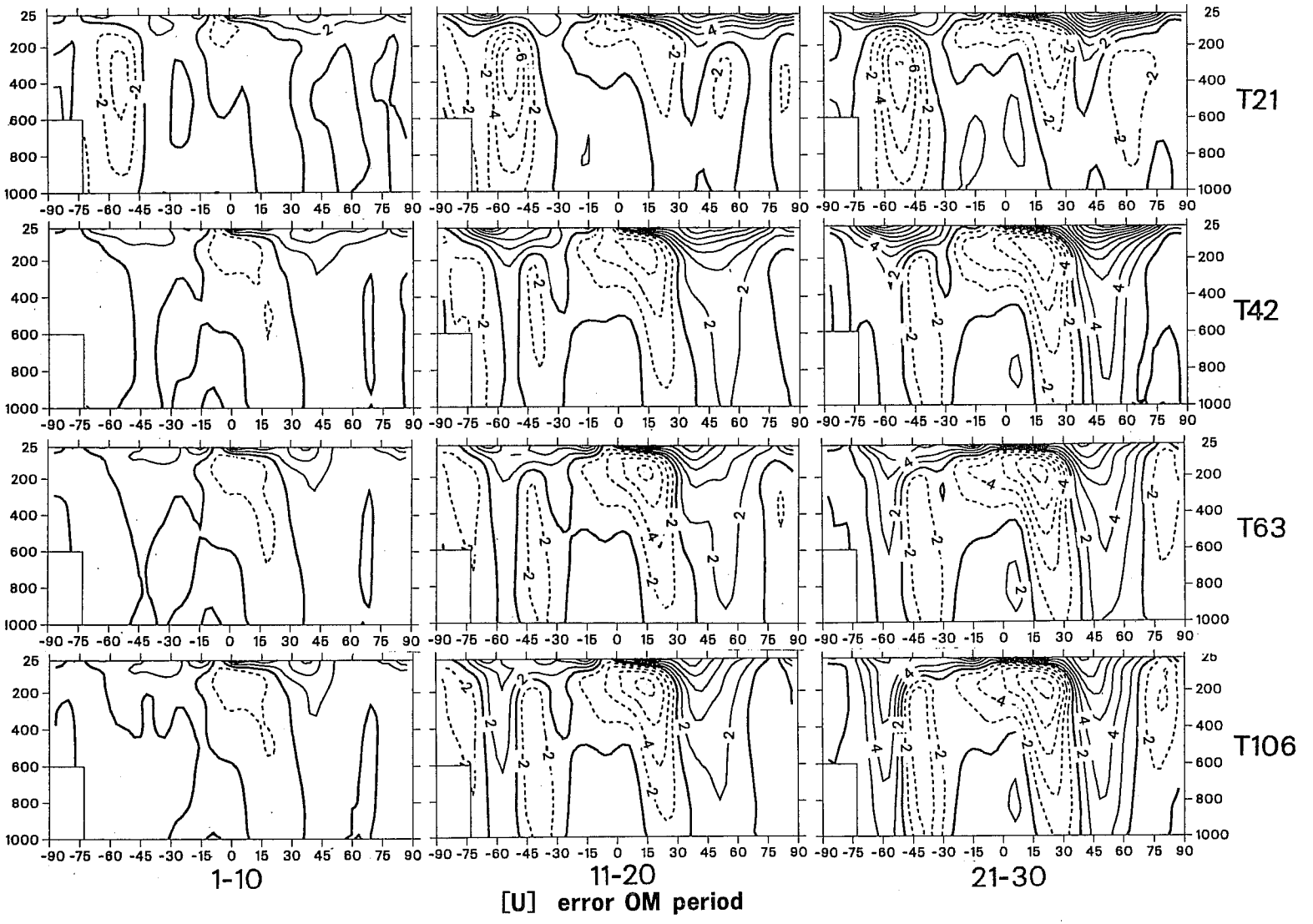

Fig 2 Latitude-height cross-sections of zonal mean of zonal wind errors for the OM period. Top to bottom: T21, T42, T63, T106. Left: day 1-10; centre: day 11-20; right: day 21-30. Contours every 2ms^{-1}.

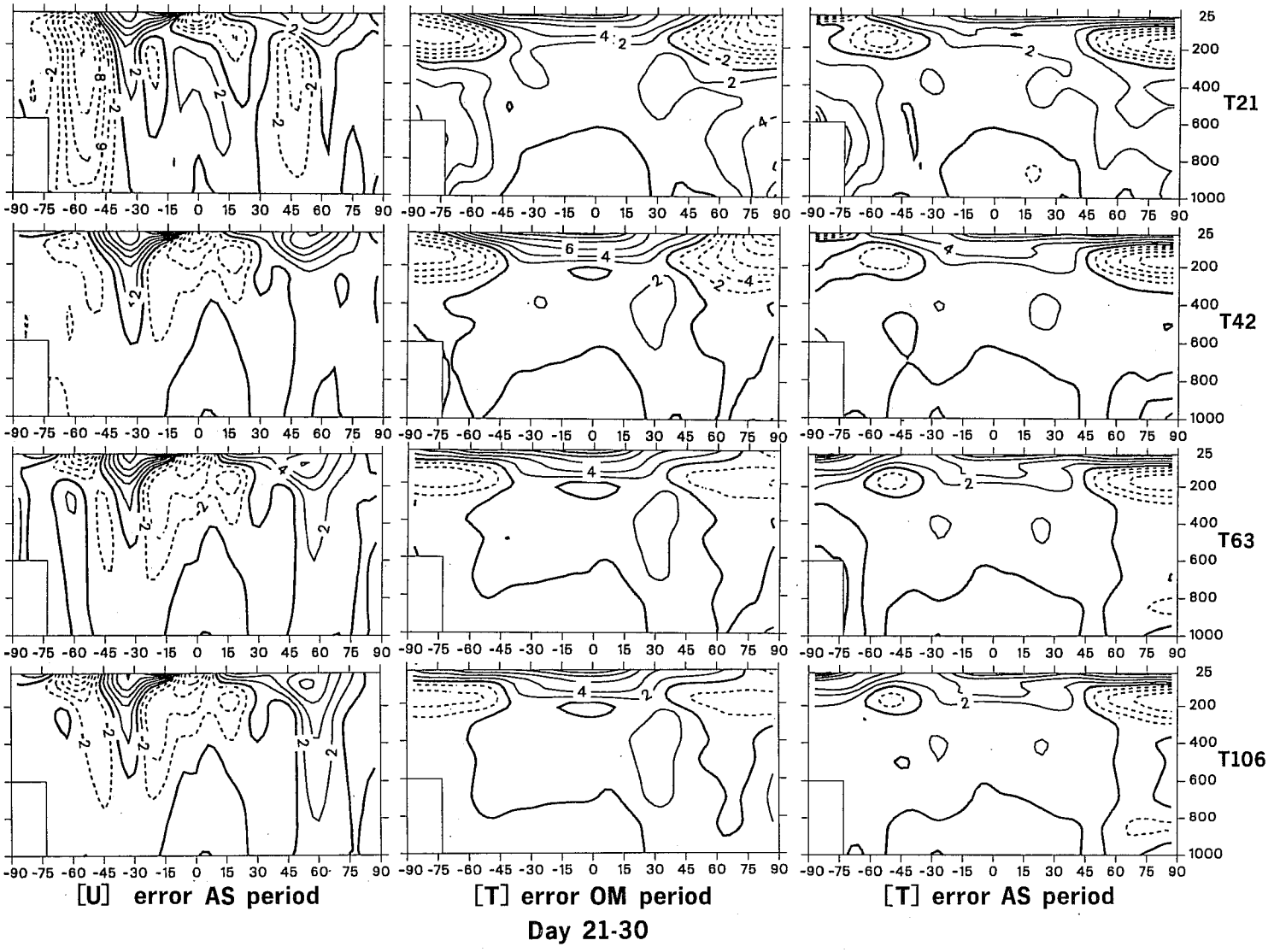

Fig 3 Latitude-height cross-sections of day 21-30 zonal mean errors. Top to bottom: T21, T42, T63, T106. Left: [U] AS period; centre: [T] OM period; right: [T] AS period. Contours every 2 ms^{-1} or 2°K.

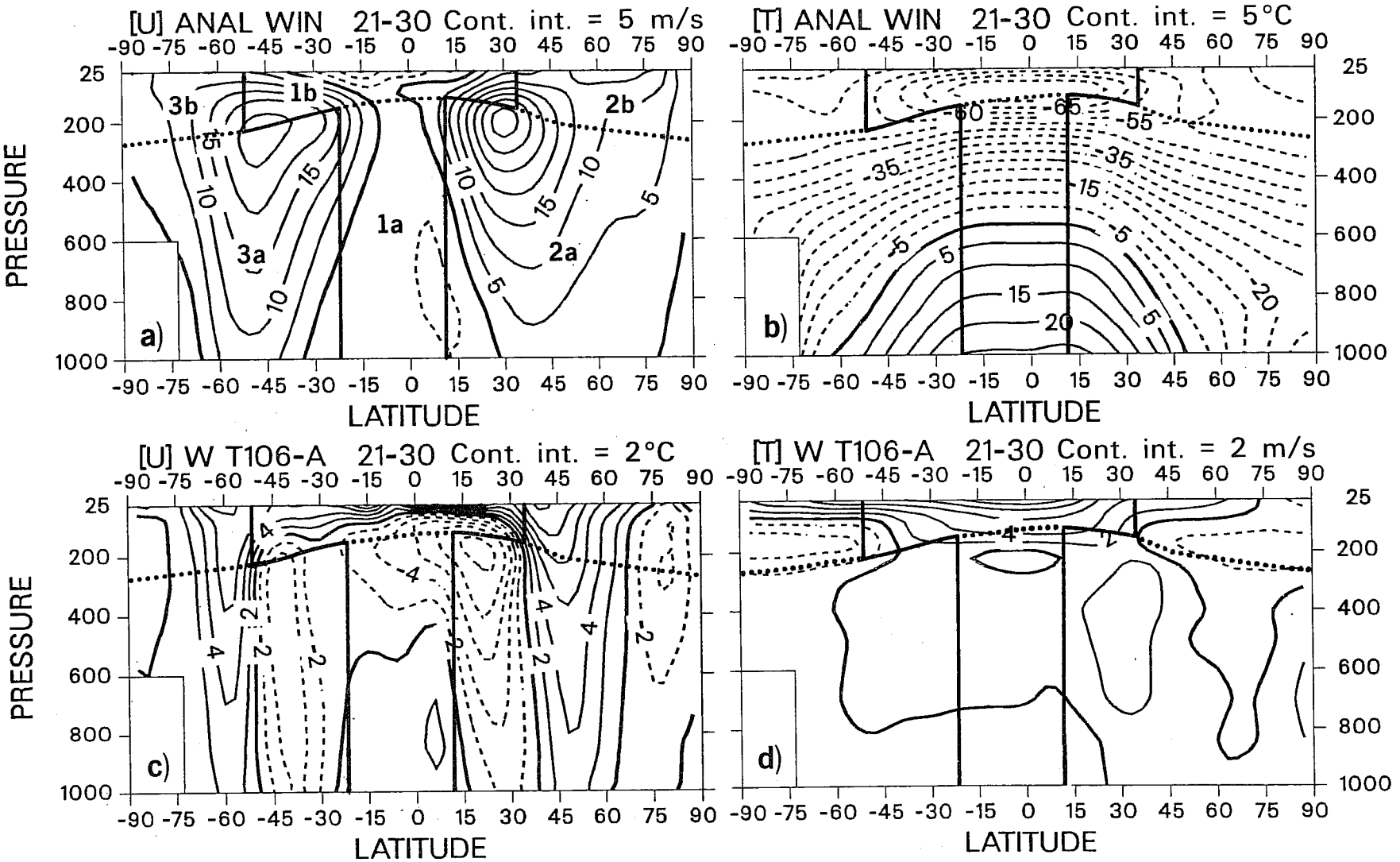

Fig 4 Latitude-height cross sections of day 21-30 zonal mean analysed fields (top) and forecast errors (bottom) of zonal wind (left) and temperature (right). For meaning of area division, see text. Contours every 5 ms^{-1} and 5°K for full fields and 2 ms^{-1} and 2°K for error fields.

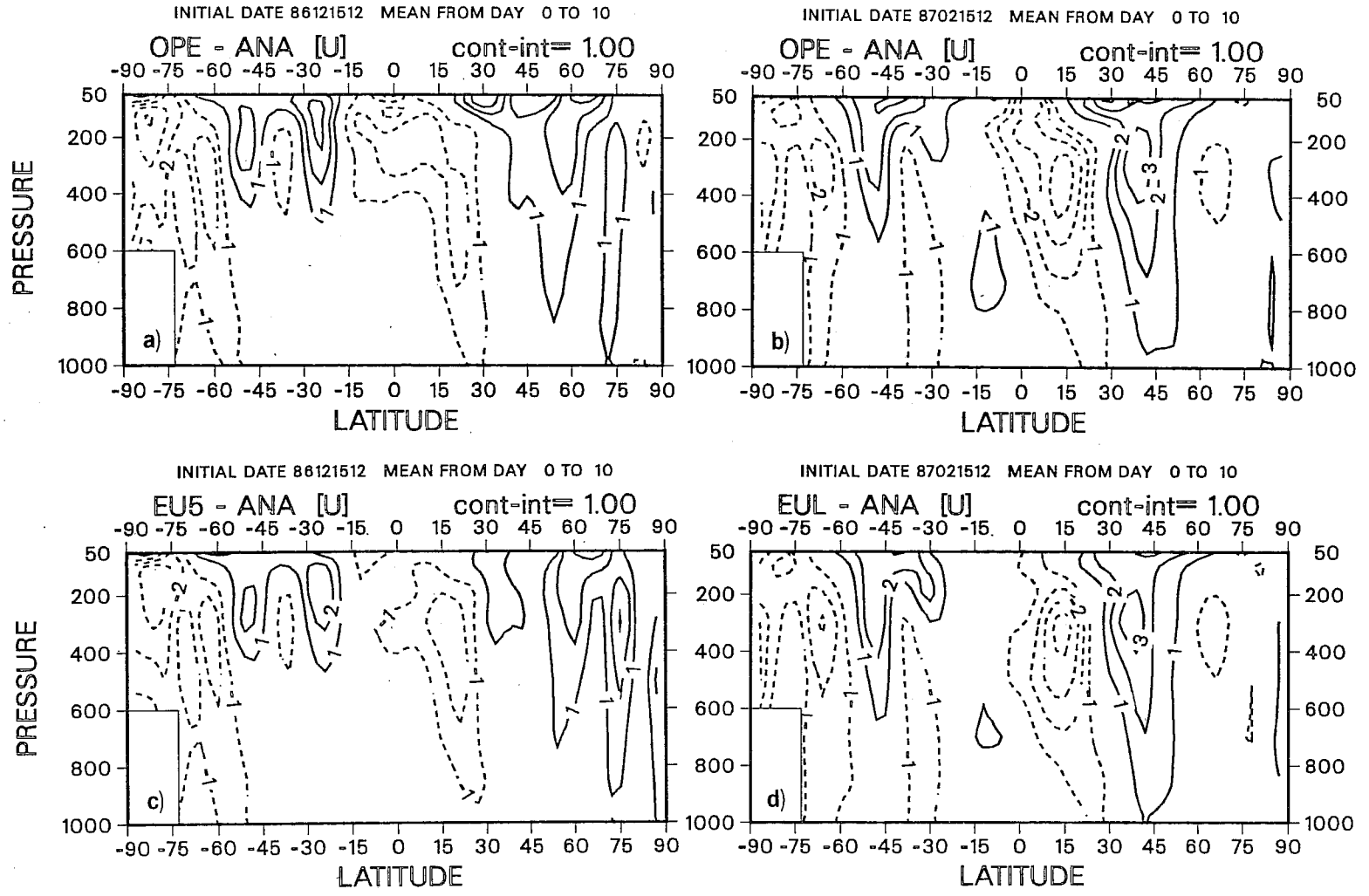

Fig. 5 Latitude-height cross-sections of day 1-10 zonal mean of zonal wind errors for two control 10-day forecasts and two experimental forecasts where free-atmosphere vertical diffusion had been eliminated. Top: control runs; bottom: experimental (no vertical diffusion) runs; right: experiment started from 15 February 1987; left: experiment started from 15 December 1986, (courtesy of Miller and Viterbo).

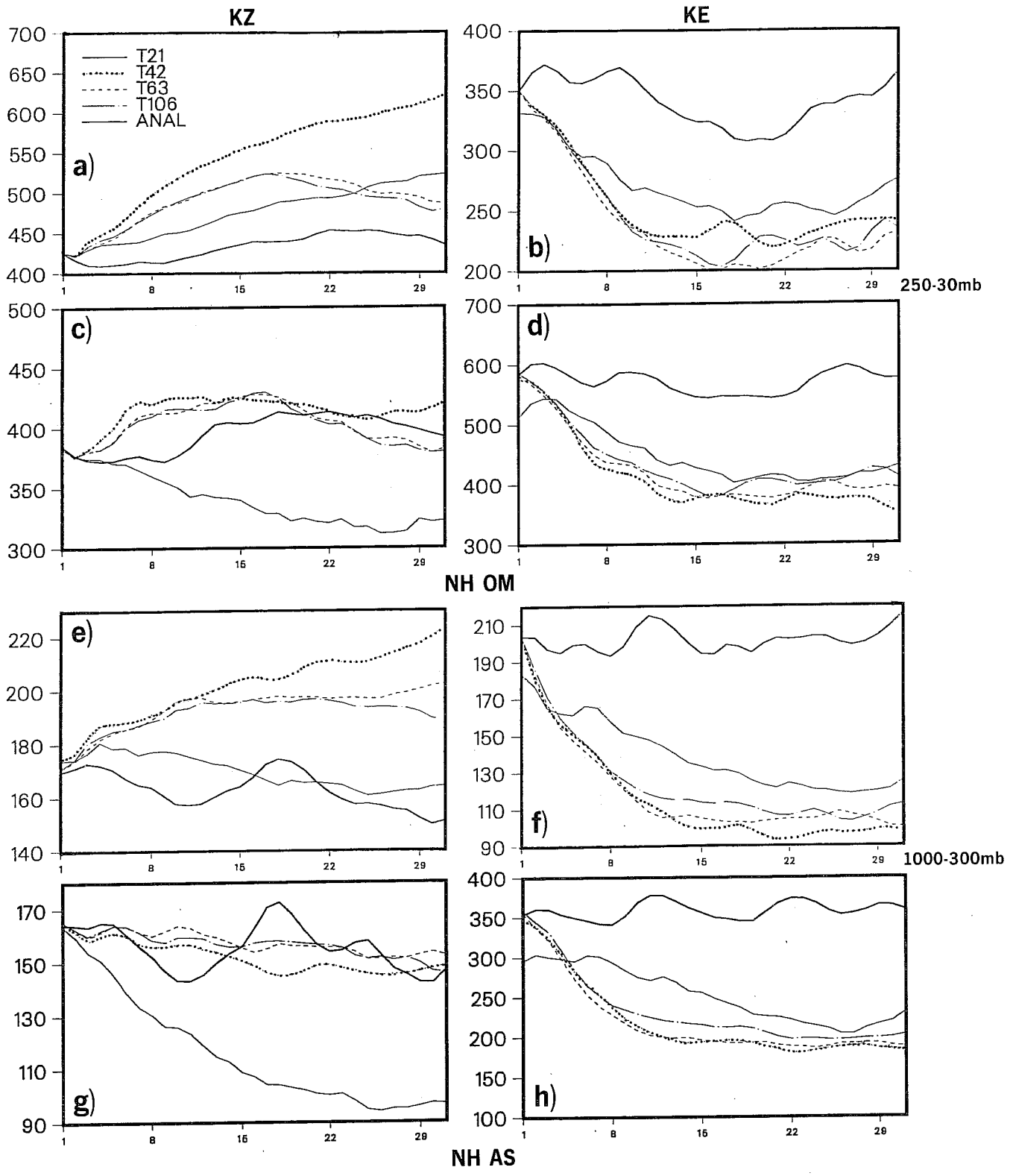

Fig 6 Time evolution (day 1 to day 30) of zonal (left) and eddy (right) kinetic energy for northern hemispheric forecasts. T21, T42, T63 and T106 forecasts are shown, together with analysed values.
a) to d) OM period; e) to h) AS period. a), b), e) and f) vertical integral from 250 mb to 30 mb; c), d) g) and h) from 1000 to 300 mb.

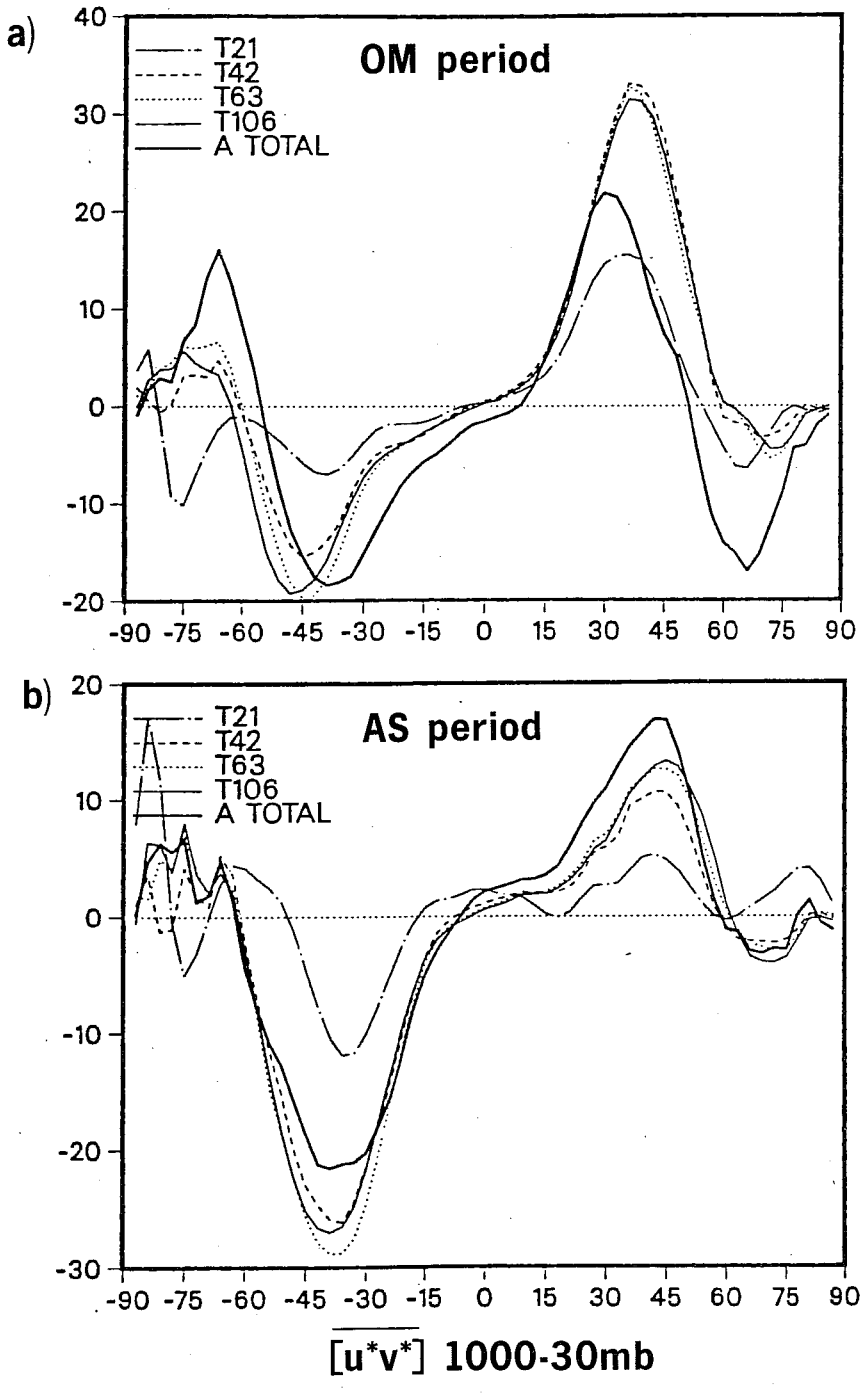

Fig 7 Latitudinal profile of 30 day mean integral of $\overline{[u^*v^*]}$ between 1000 and 30 mb. Top: OM period; bottom: AS period.

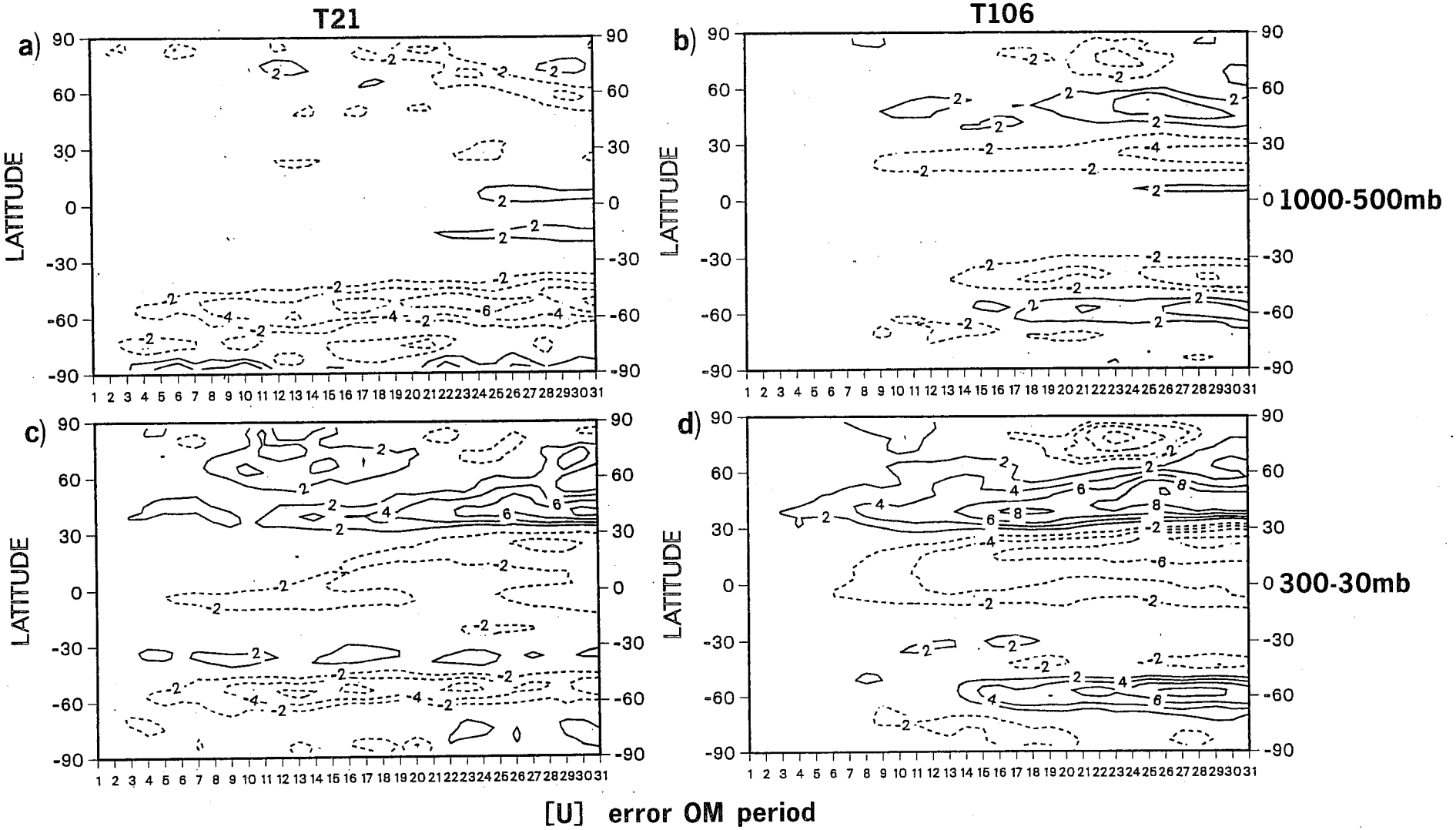

Fig 8 Latitude-time diagrams of vertically integrated [U] error for the OM period. Left: T21; right: T106. Top: 1000-500 mb; bottom: 300-30 mb.

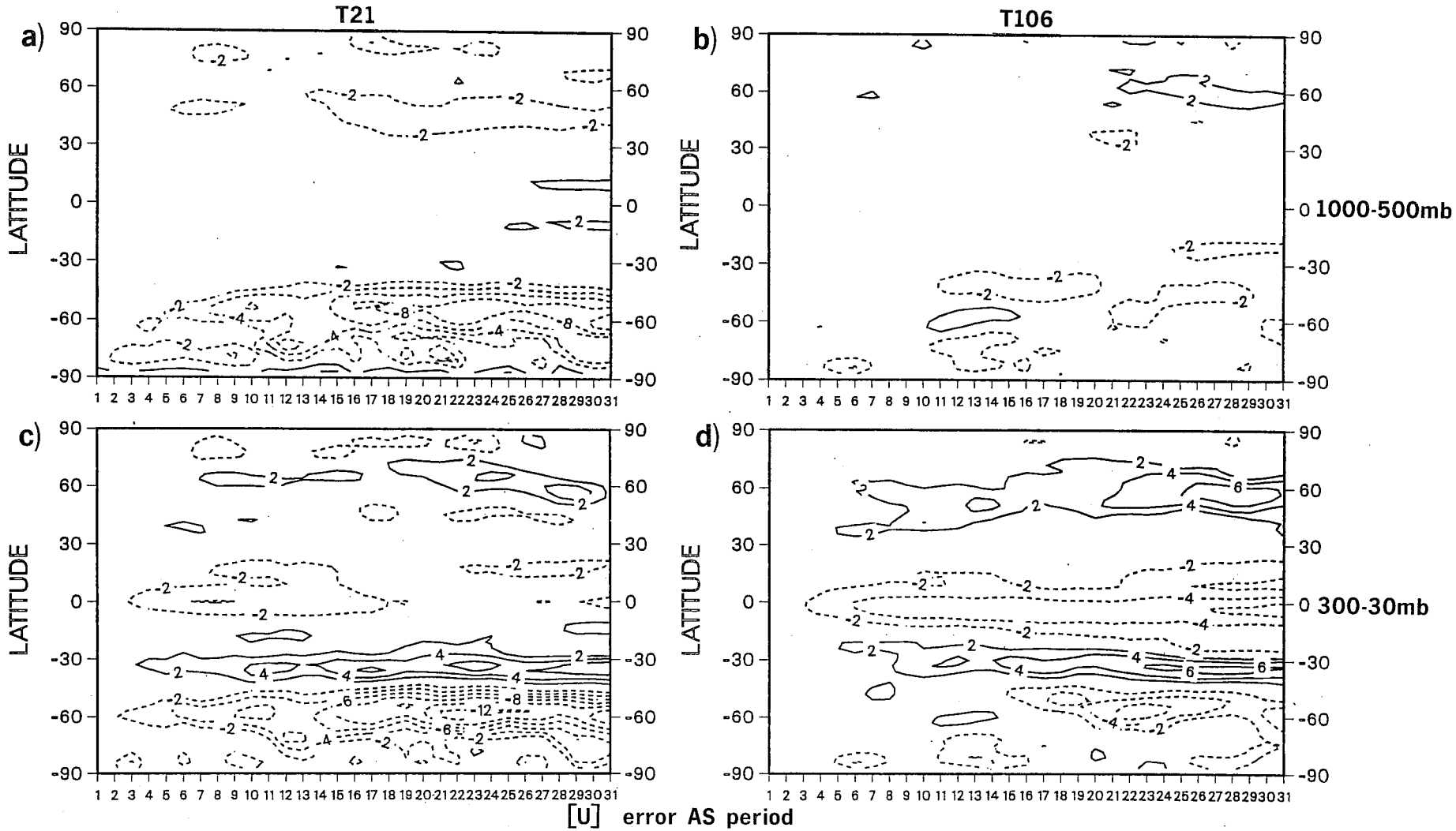

Fig 9 As Fig 7, but for the AS period.

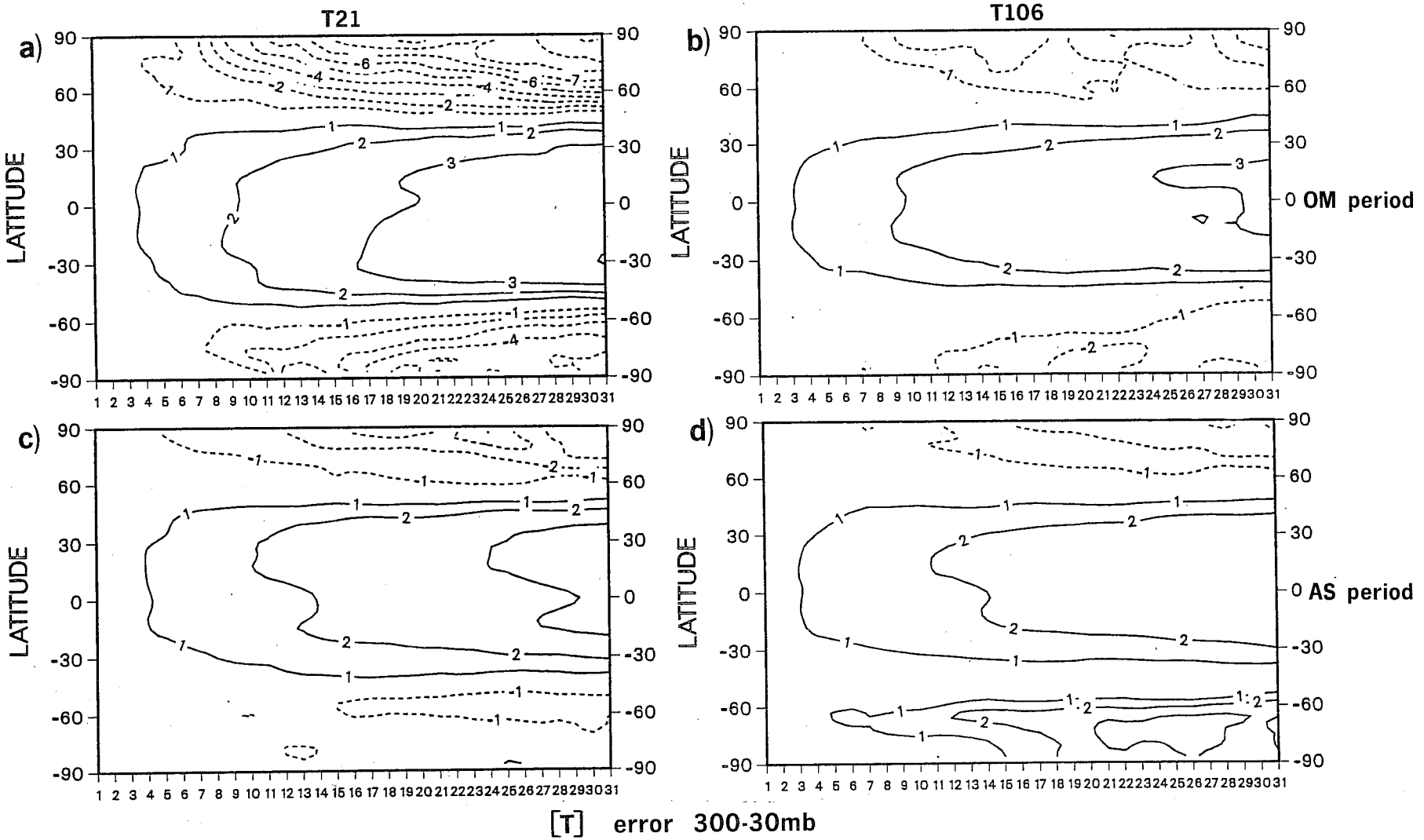

Fig 10 Latitude-time diagram of vertically integrated [T] error between 300 and 30 mb. Left: T21; right: T106. Top: OM period; bottom AS period.

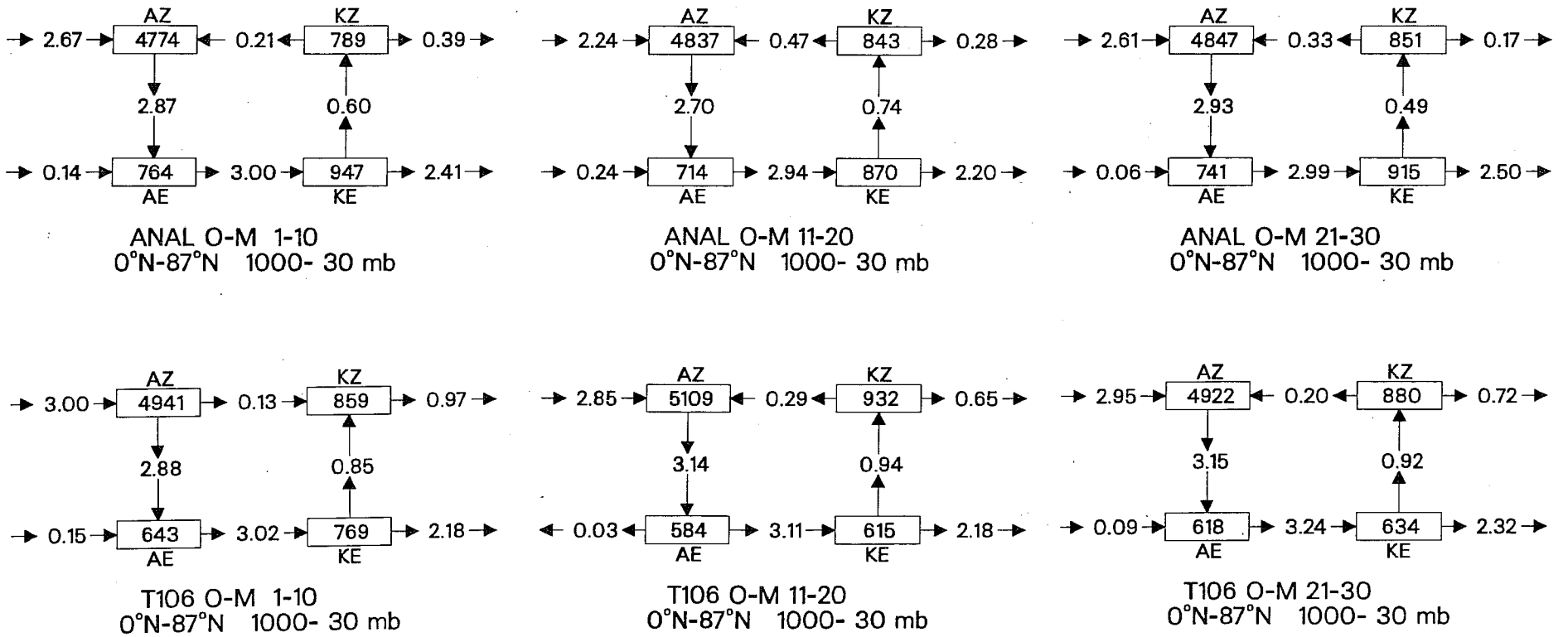

Fig 11 Box-diagrams Lorenz-type energetic conversions. AZ: zonal available. KZ: zonal kinetic. Left: dd 1-10; centre: dd 11-20; right: dd 21-30. Top: analyzed values; Bottom: T106 model. Units: kJm^{-2} and Wm^{-2}

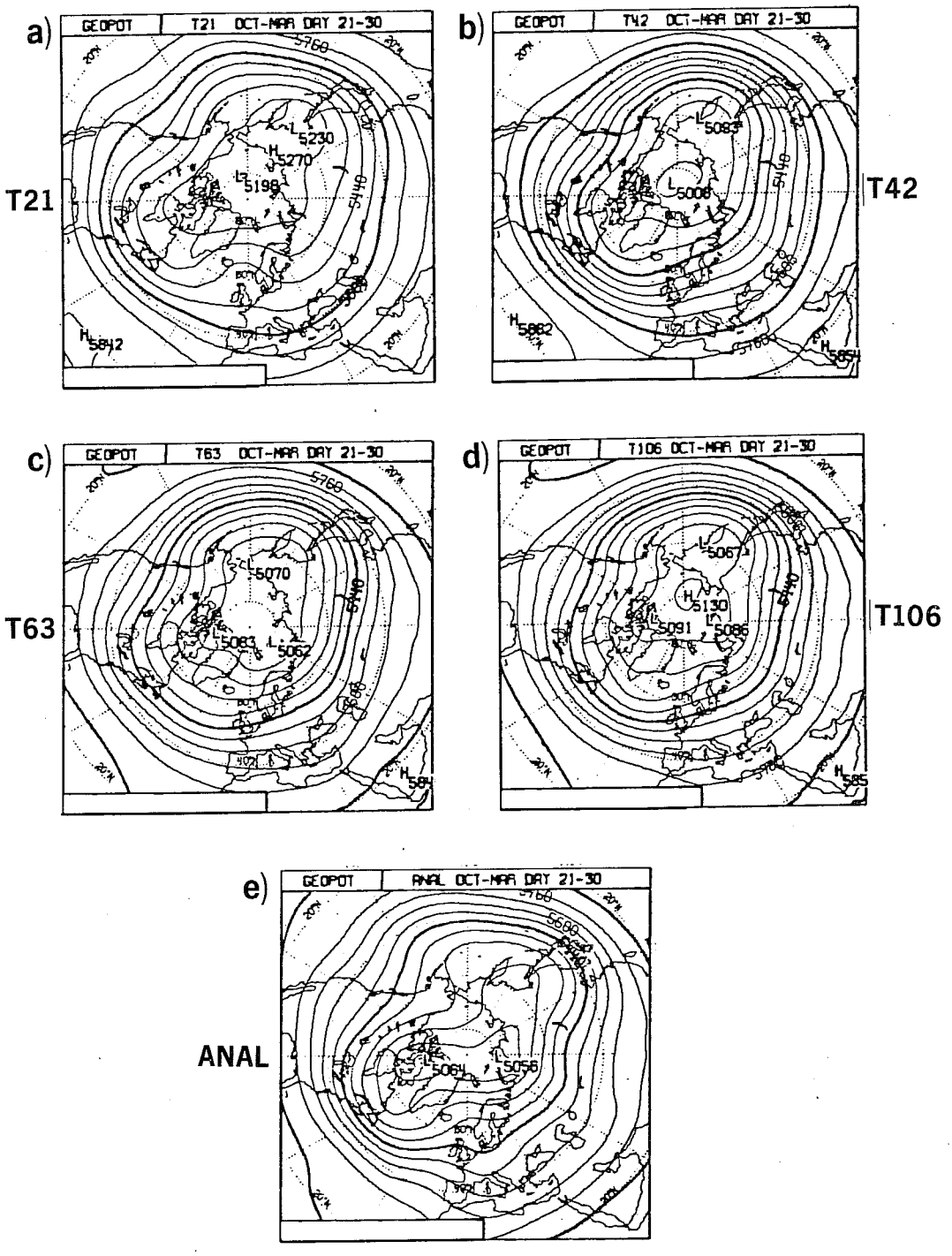

Fig 12 500 mb mean geopotential field for the 12 integrations in the OM period. a) T21; b) T42; c) T63; d) T106; e) analysis.

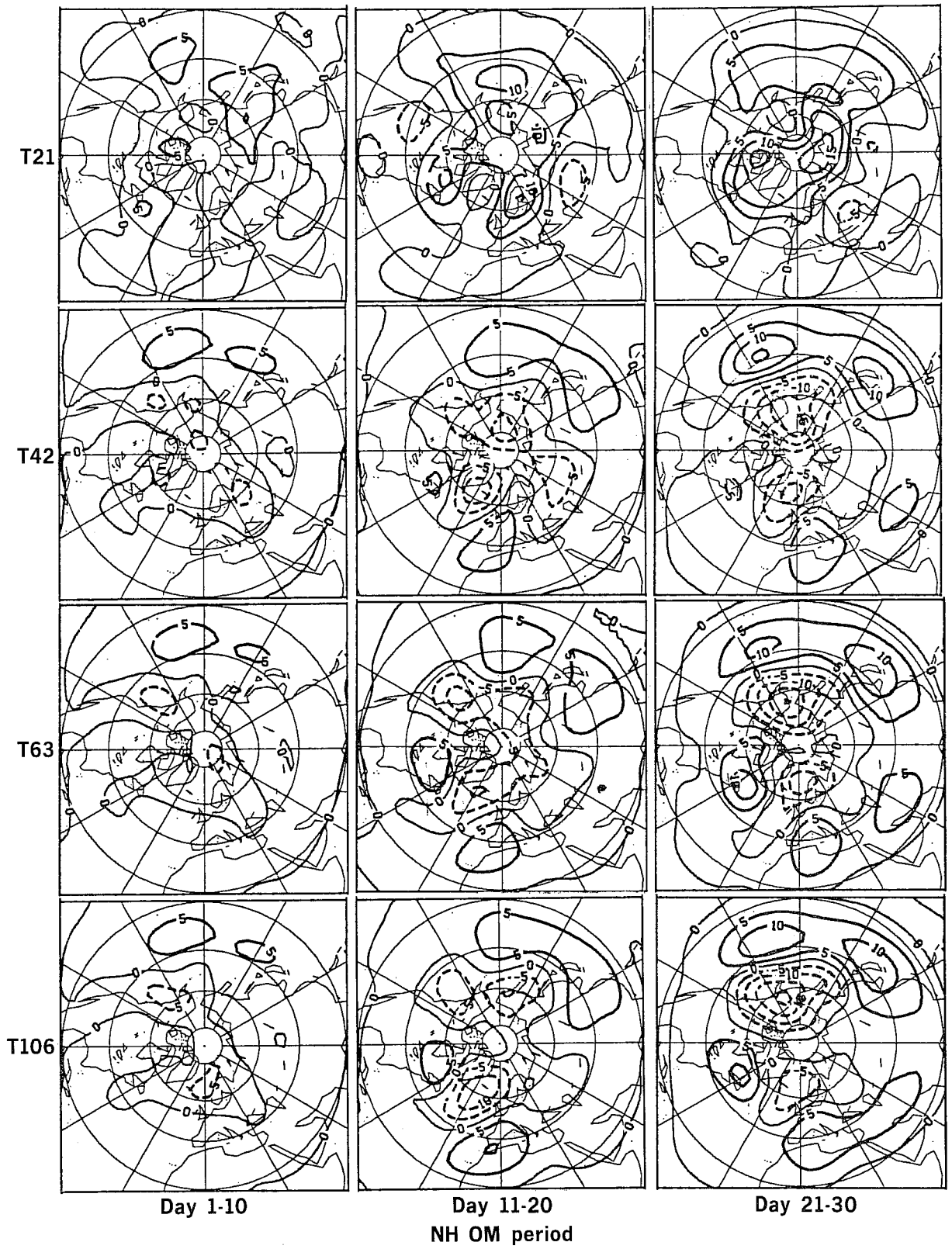

Fig 13 500 mb geopotential height northern hemispheric mean errors for the OM period. Top to bottom: T21, T42, T63 and T106. Left to right: dd 1-10, dd 11-20; dd 21-30. Contour intervals every 5 dam.

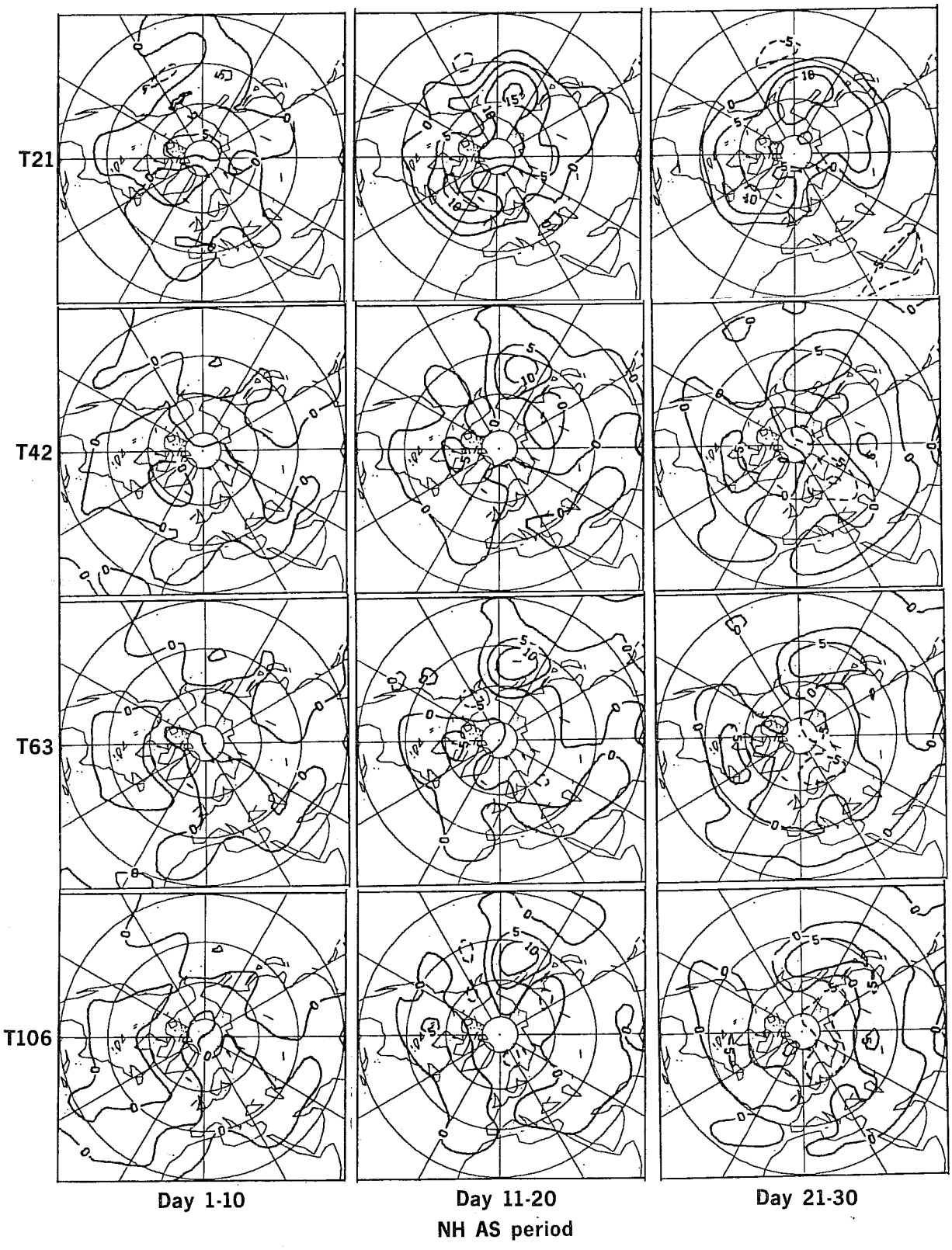

Fig 14 As Fig 12, but for the AS period.

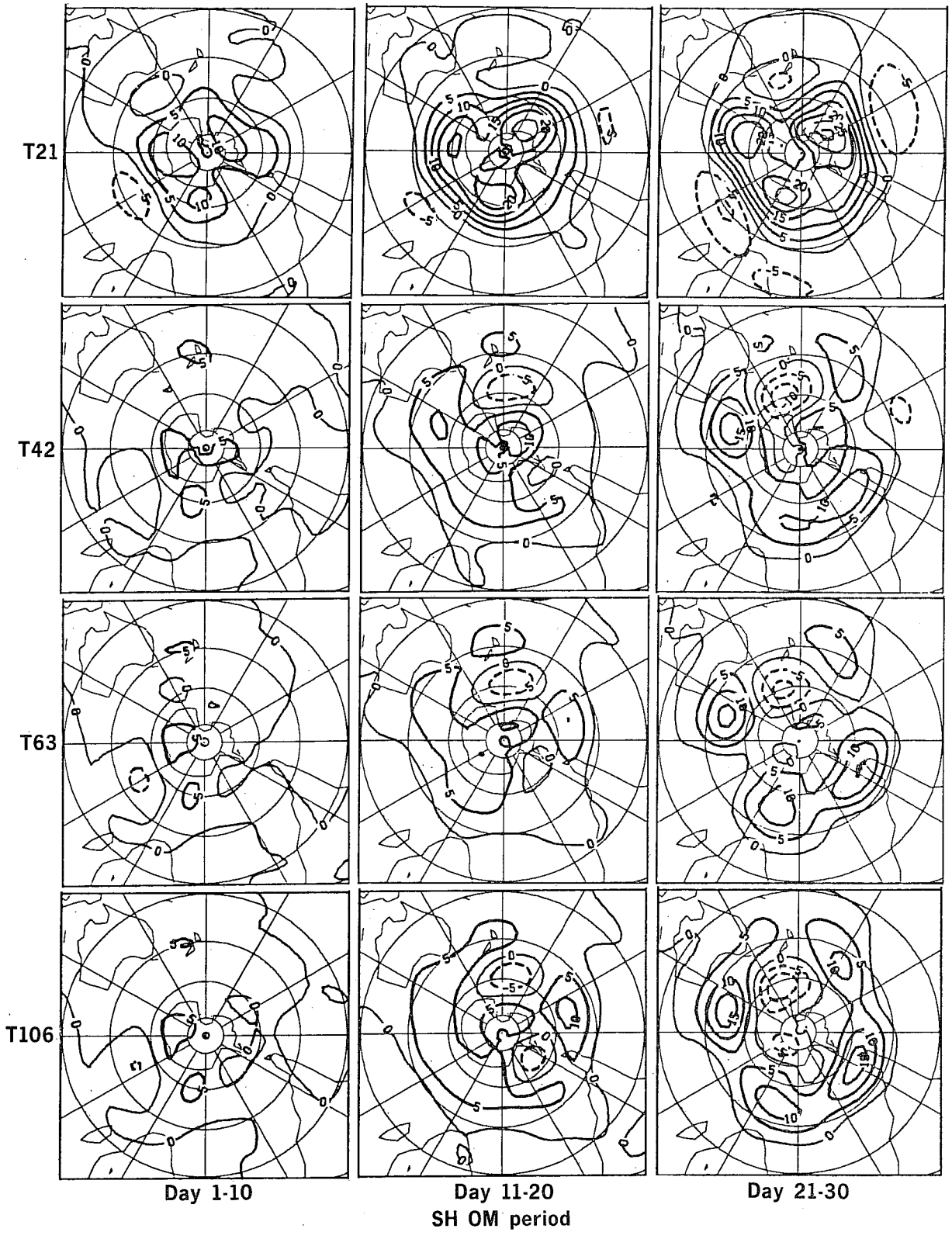

Fig 15 As Fig 12, but for the Southern Hemisphere.

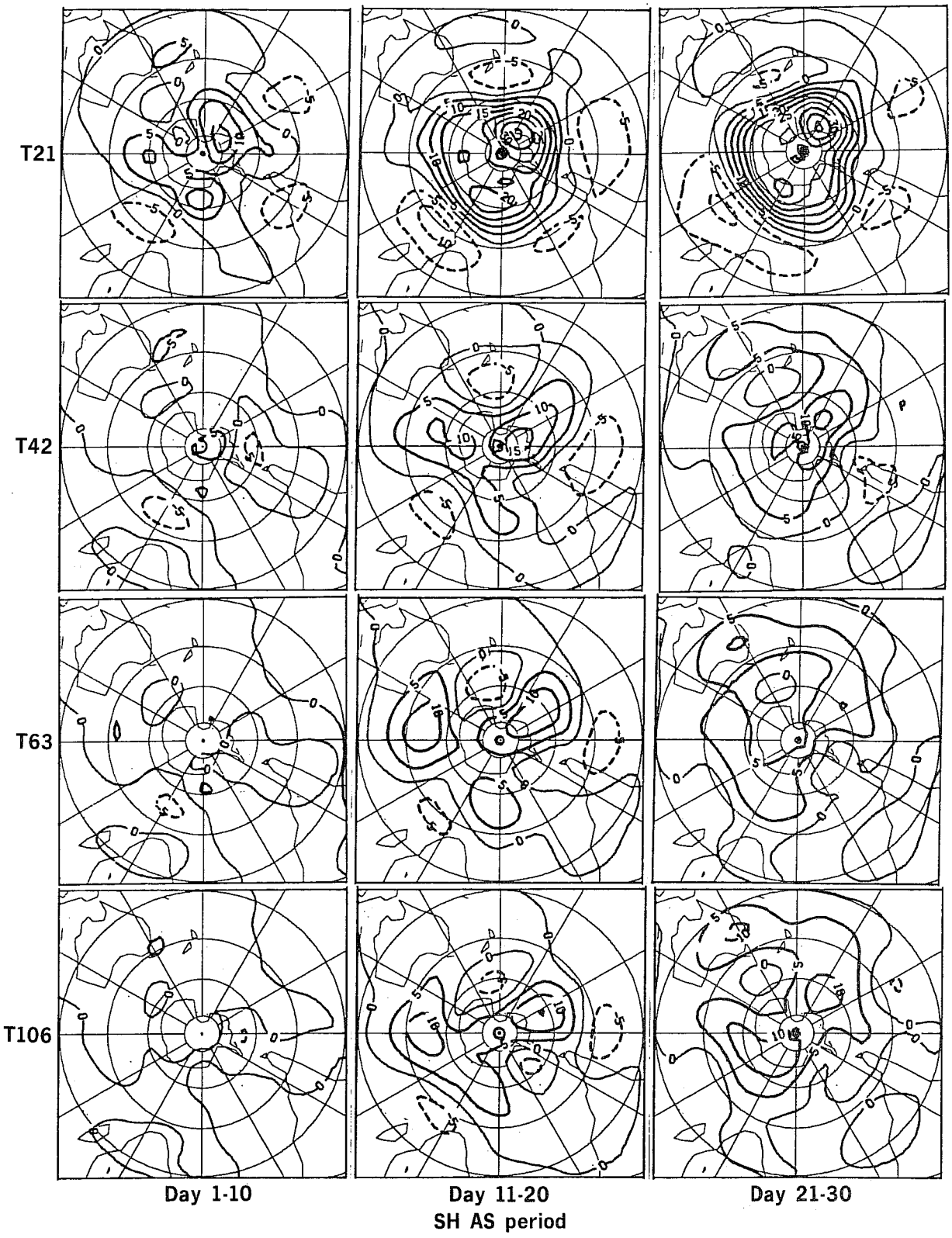

Fig 16 As Fig 12, but for the Southern Hemisphere and the AS period.

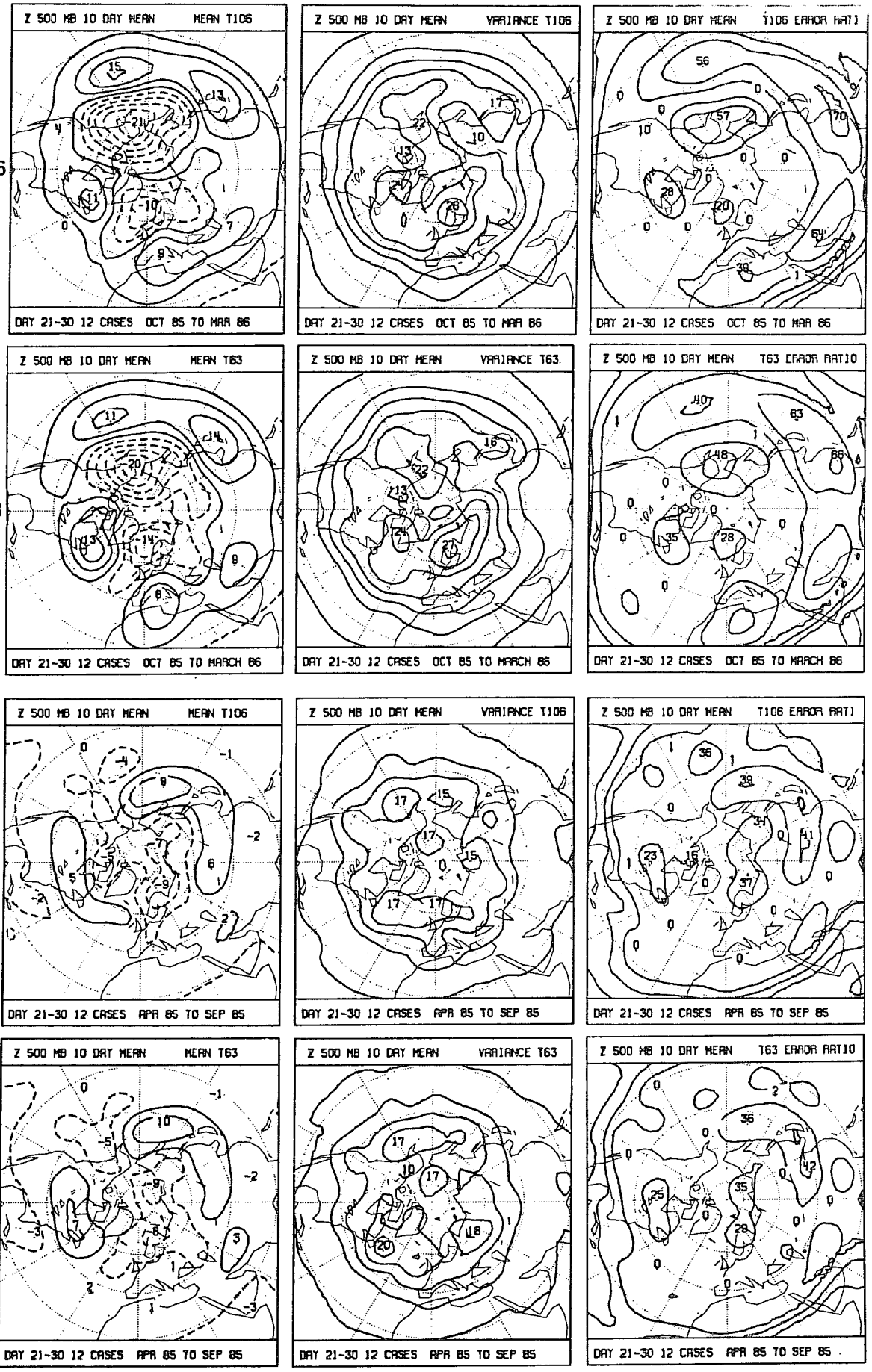

Fig 17 Day 21-30 500 mb geopotential height errors for the T106 and T63 models. Top six panels: OM period. Bottom six panels: AS period. Left: mean errors. Centre: RMS errors. Right: percent ratio of mean error variance to total error variance. Contour interval is 5 dam for left and centre panels and 20% for right panels.

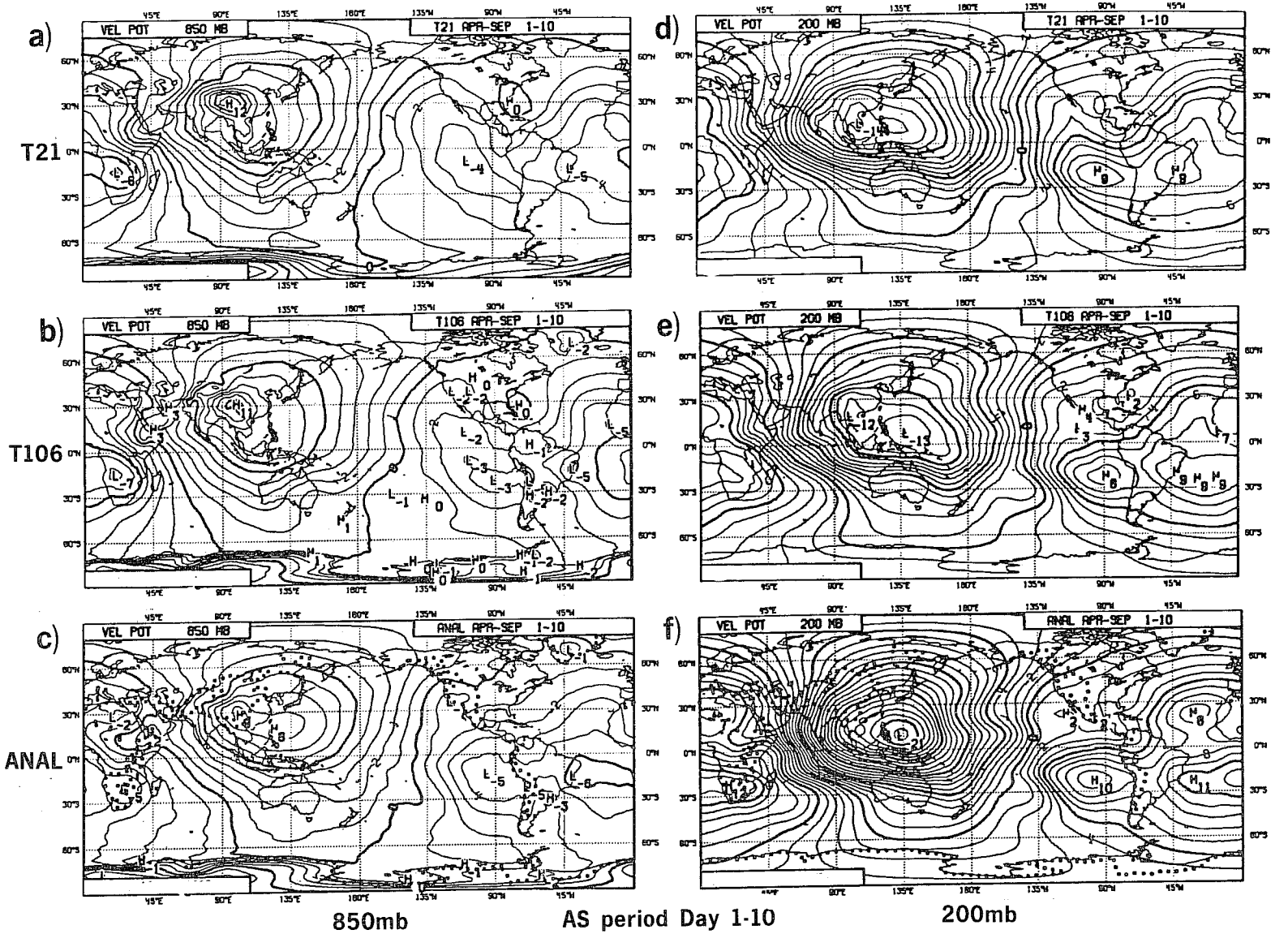

Fig 18 Mean dd 1-10 fields of velocity potential (χ) for the AS period. Top: T21. Centre: T106. Bottom: analysis. Left: 850mb. Right: 200 mb.

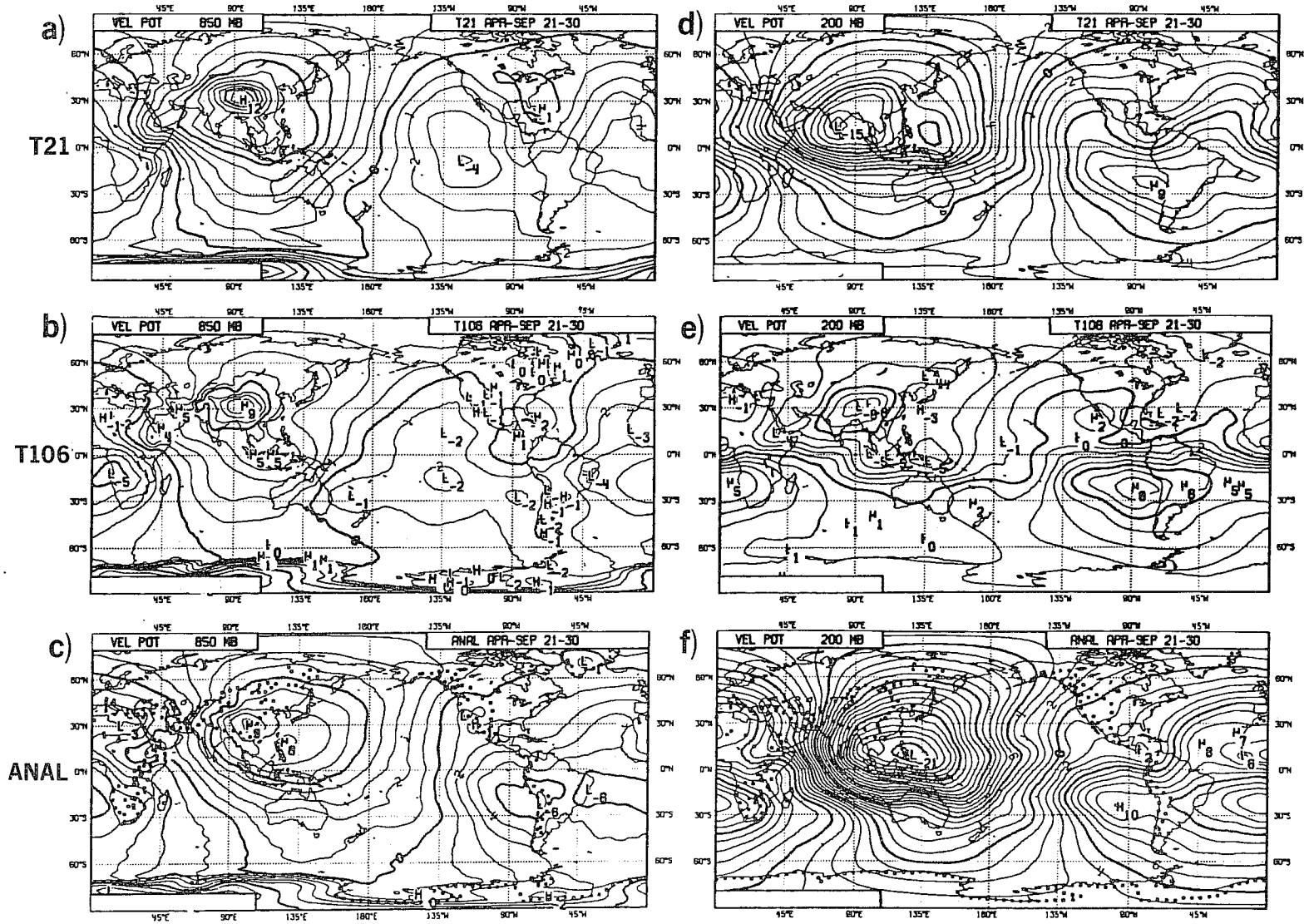

Fig 19 As Fig 17, but for dd 21-30.

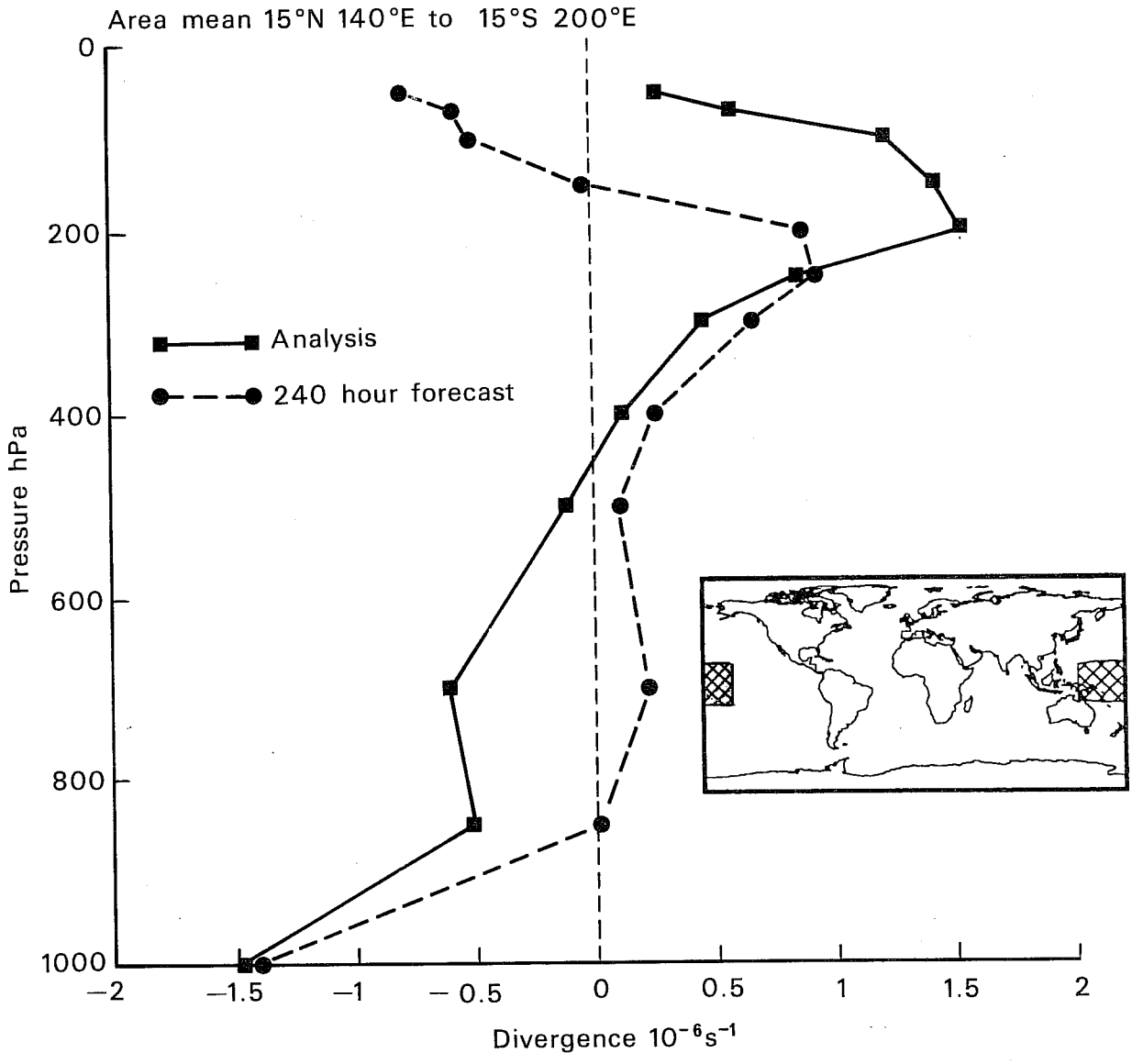

Fig 20 Vertical profile of divergence, integrated in the hatched area in the west Tropical Pacific and shown in the small panel. Full line and squares: analysed values. Dashed lines and dots: 10-day forecast values. Averaged over 3 months of operational forecasts, June to August 1986.

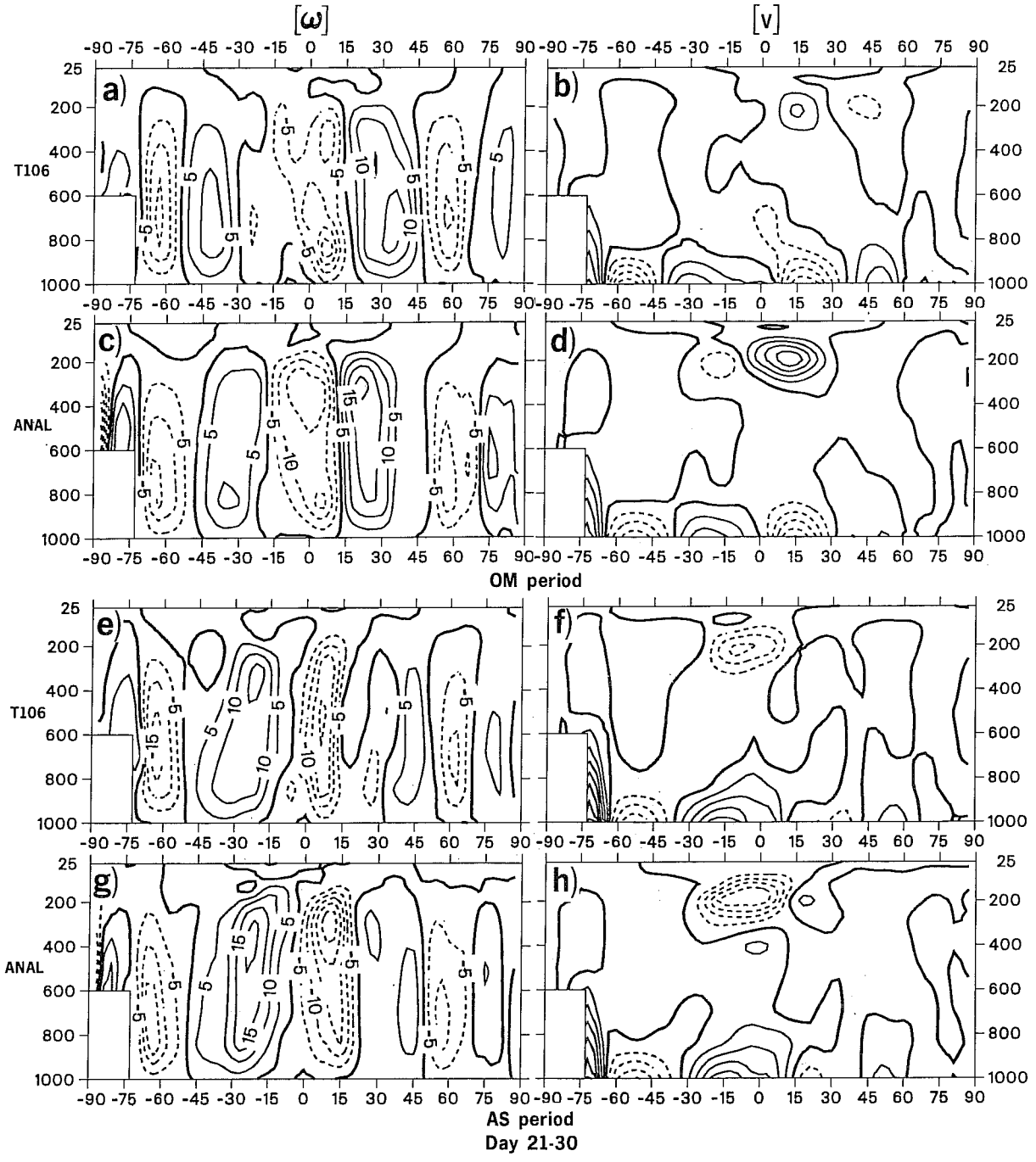

Fig 21 Latitude-height cross sections of $[\bar{\omega}]$ (left panels) and $[\bar{v}]$ (right panels). Top to bottom: T106 OM period; analysis OM period; T106 AS period; analysis AS period. Units: $mPas^{-1}$ and ms^{-1}

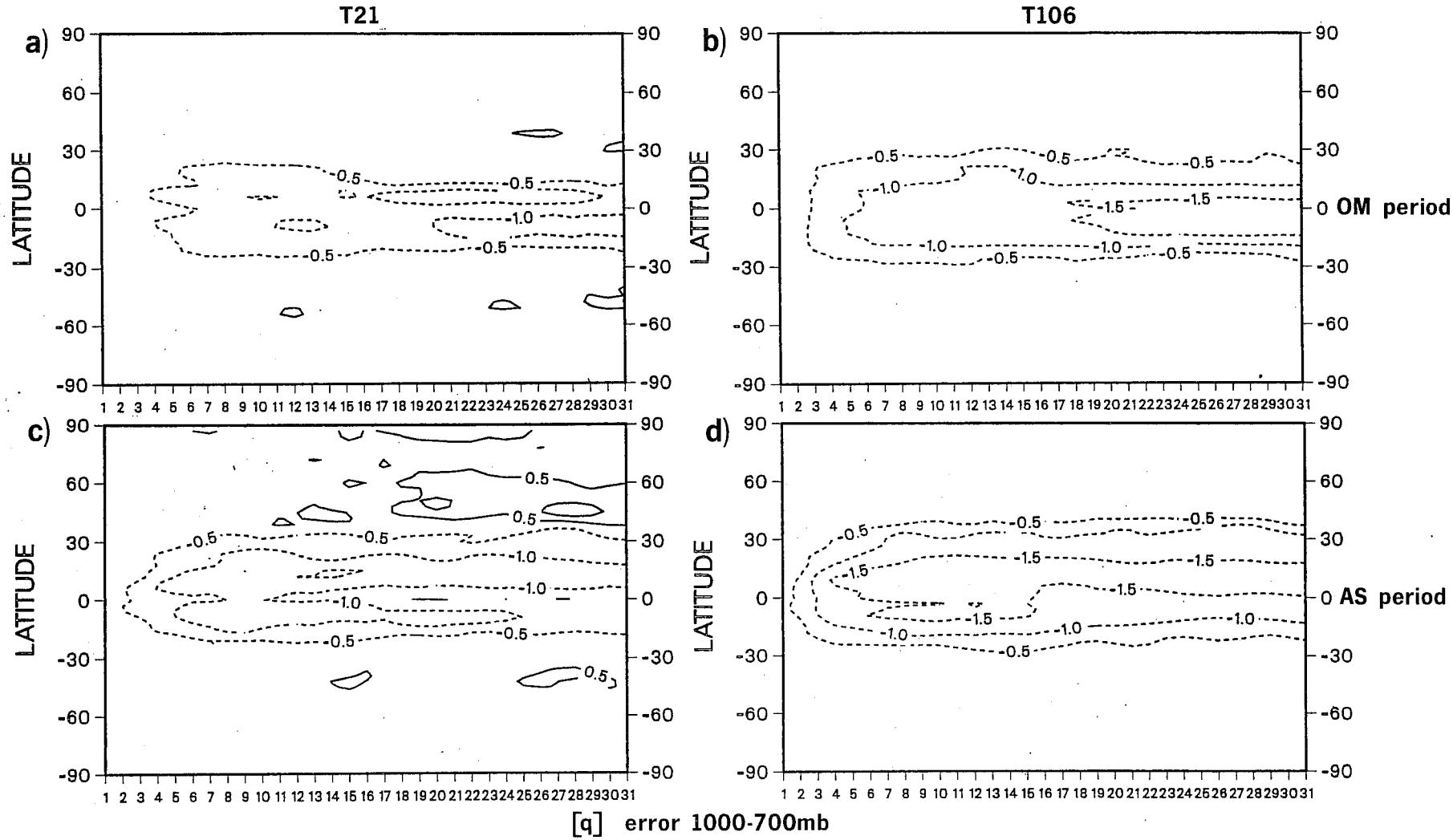

Fig 22 Latitude-time diagrams of vertically integrated [q] error between 1000 and 700 mb. Left: T21; right: T106. Top: OM period; bottom: AS period.

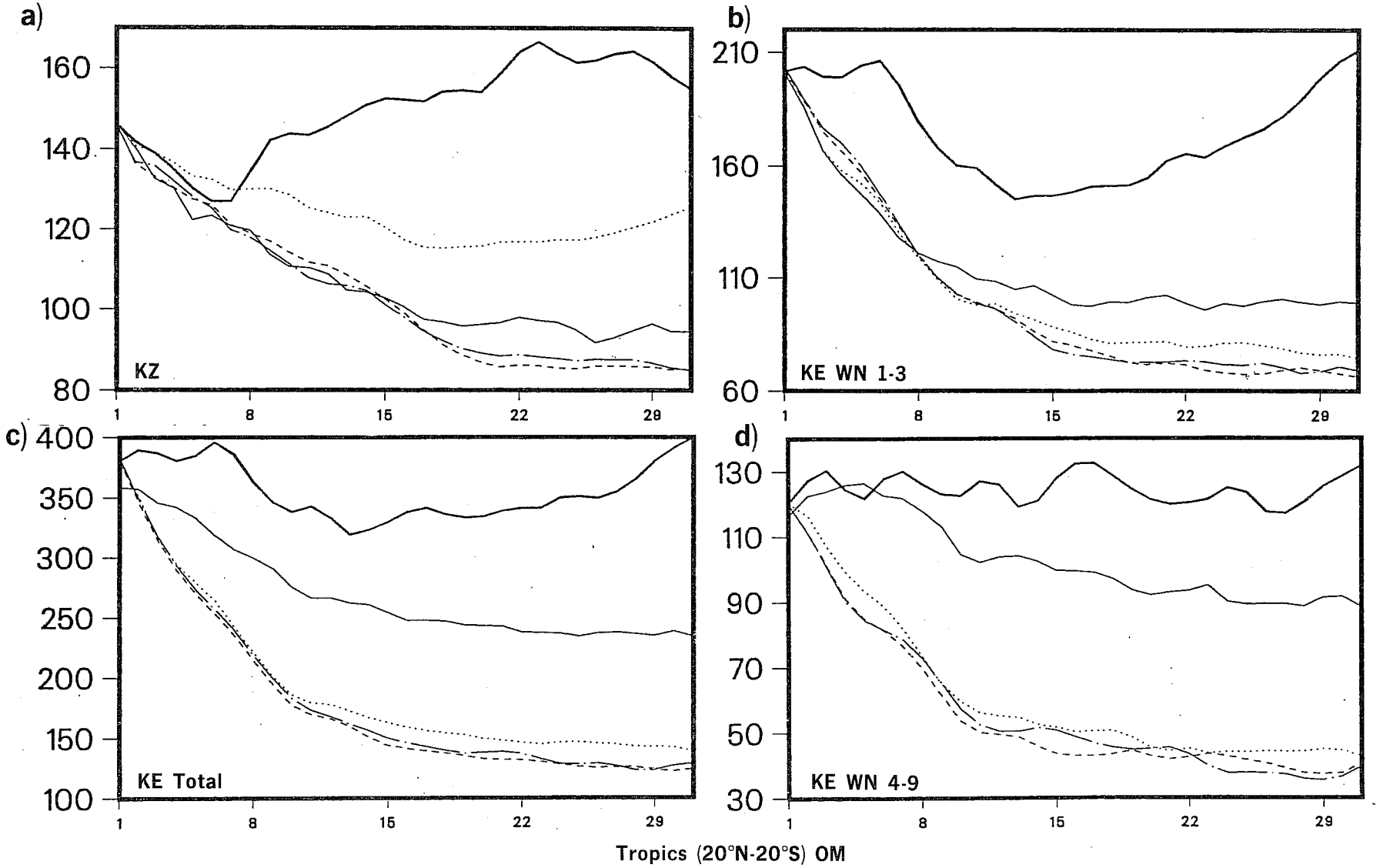

Fig 23 Time evolution (day 1 to day 30) of zonal (left) and eddy (right) kinetic energy for the forecasts of the OM period, integrated over the tropical region (20°N-20°S). T21, T42, T63 and T106 forecasts are shown, together with analysed values.

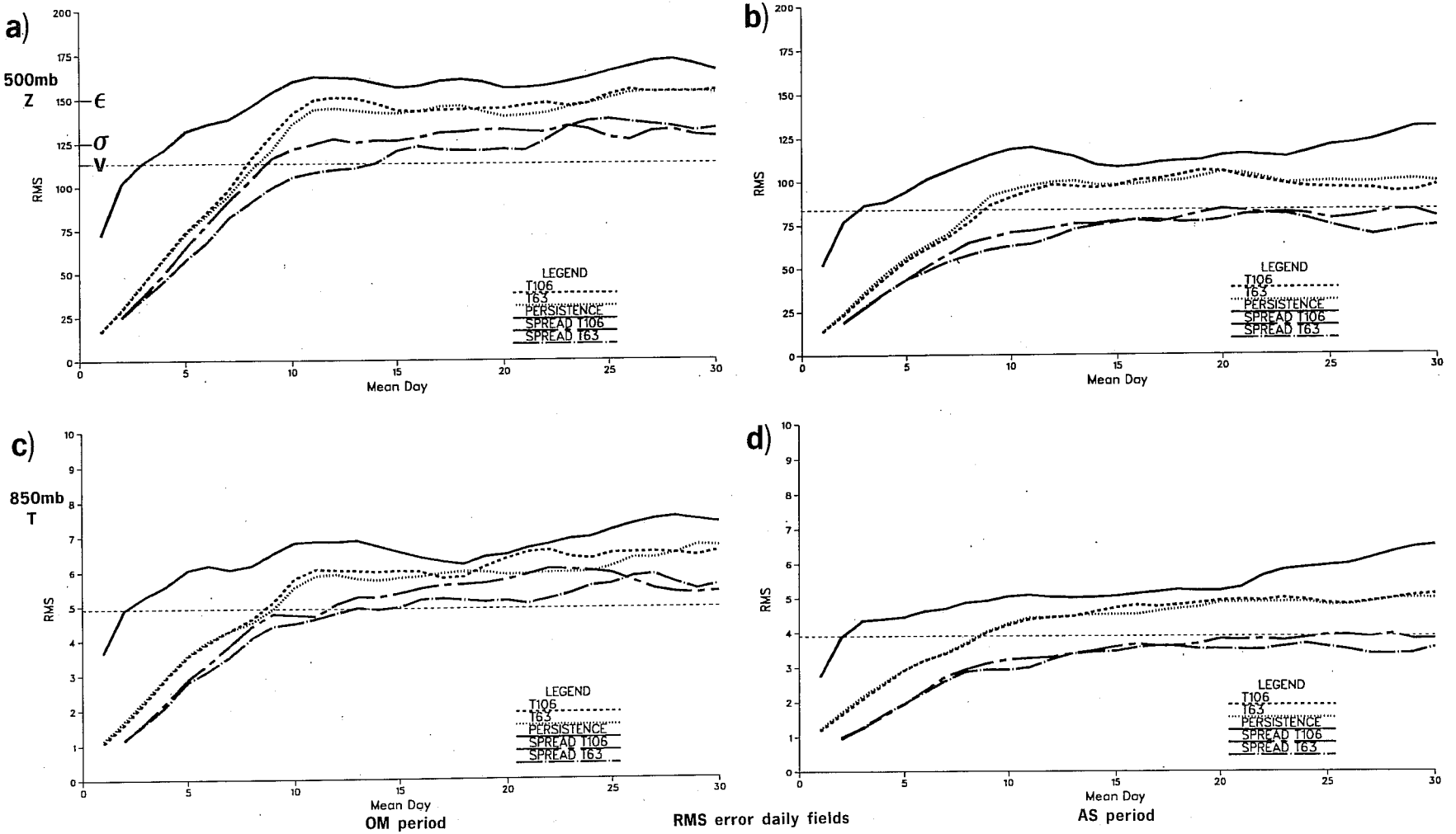

Fig 24 RMS error of daily northern hemispheric forecast fields. Solid line: persistence. Dashed: T106. Dotted: T63. Chain: T106 forecast spread. Dash-dot: T63 forecast spread. Spreads are computed between successive forecasts, initiated 24 hours apart (six pairs per six-month period). Left: OM period. Right: AS period. Top: 500 mb height. Bottom: 850 mb temperature. (For the meaning and use of ε, σ and v, see the Appendix.)

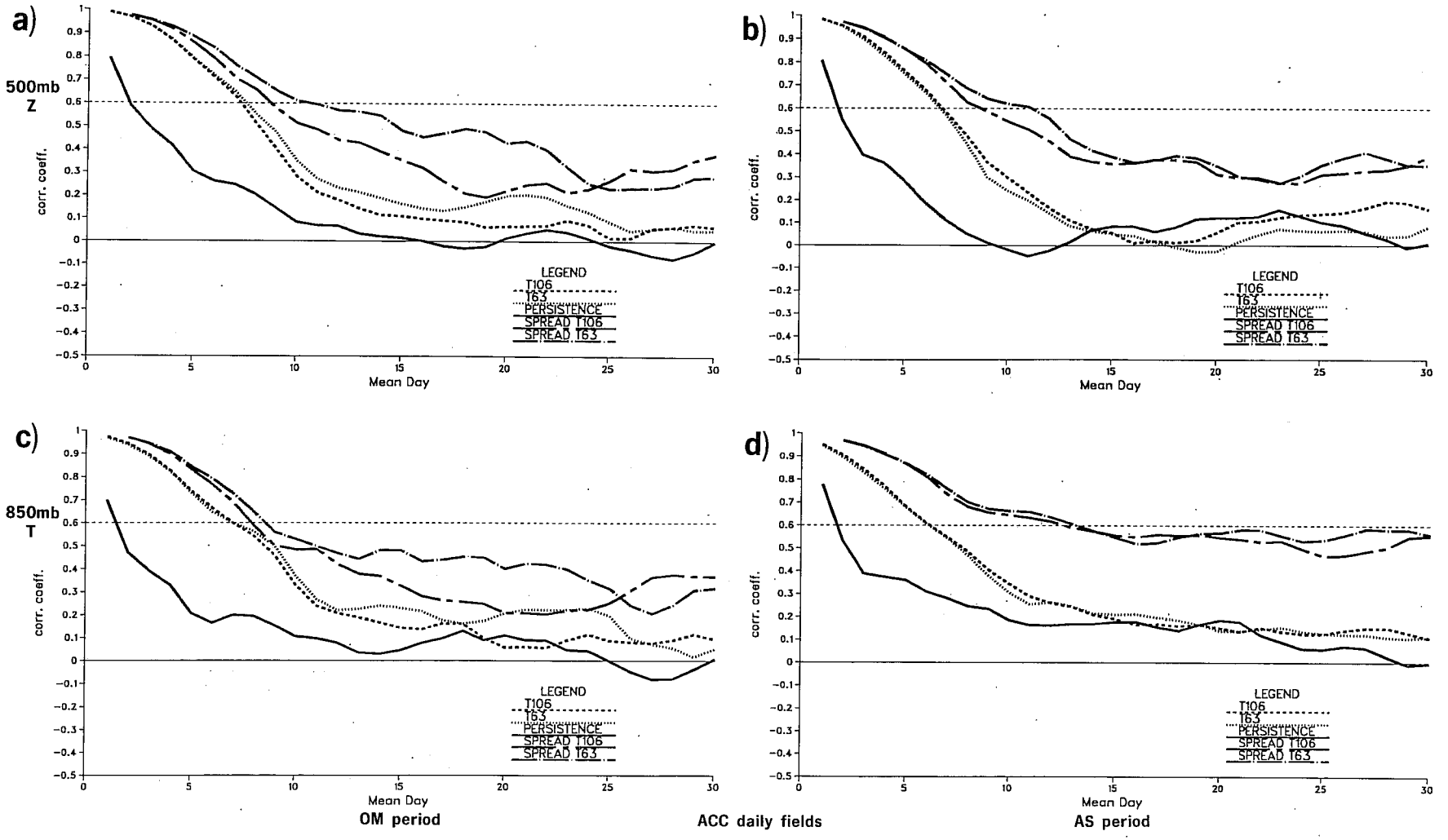

Fig 25 As Fig 23 but for ACC instead of RMS error.

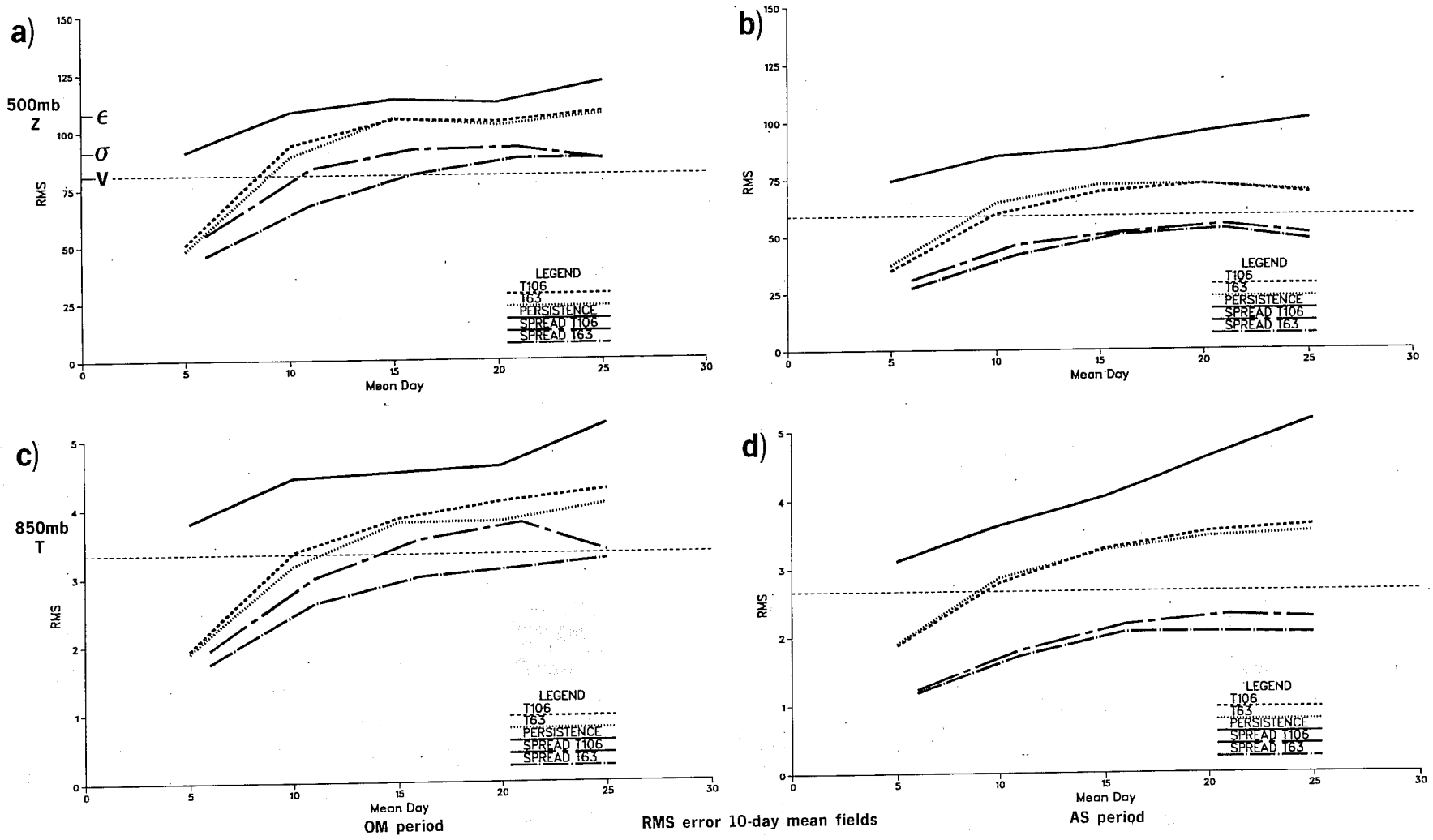

Fig 26 As Fig 23 but for 10-day running mean fields.

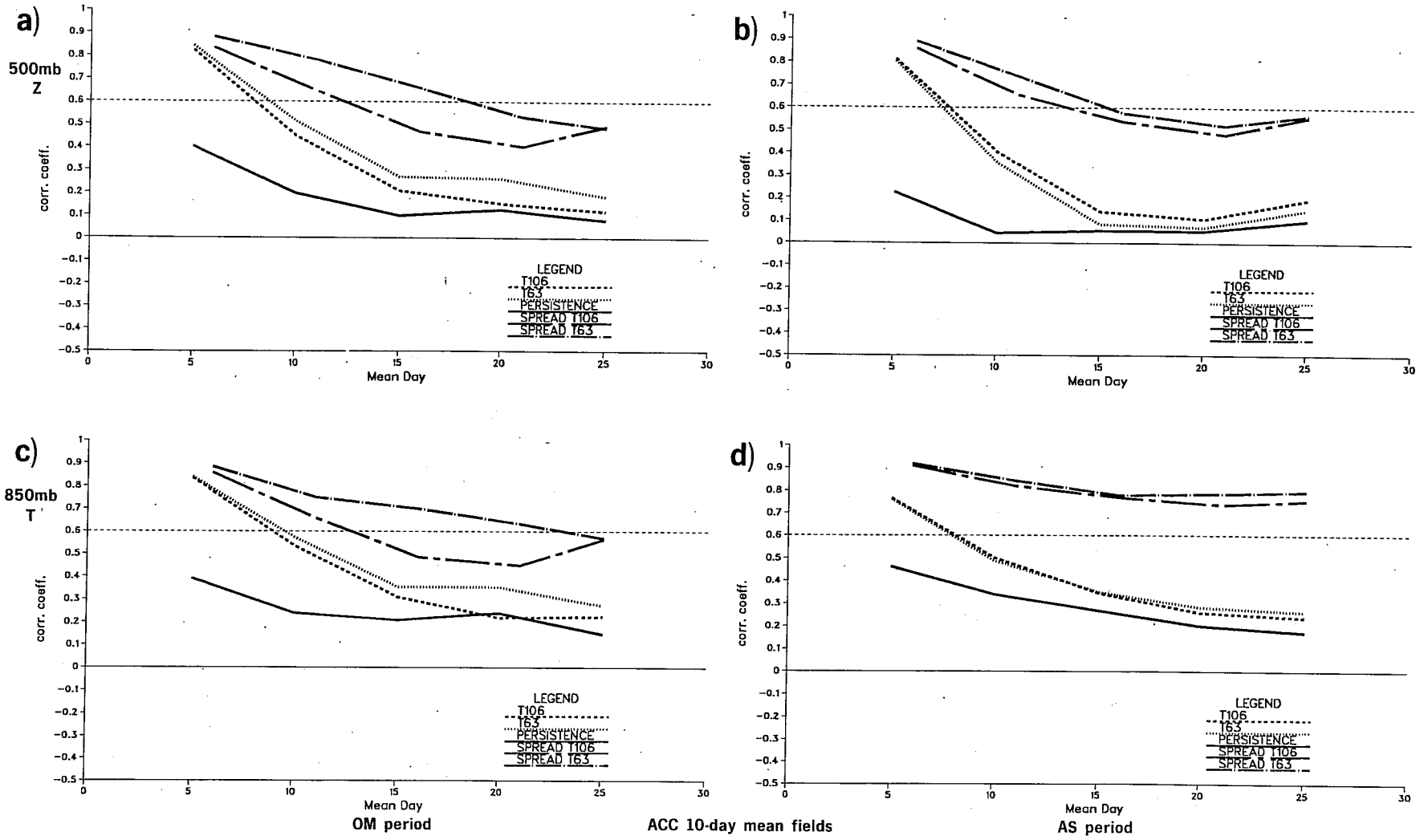

Fig 27 As Fig 23 but for 10-day running mean fields and ACC instead of RMS error.

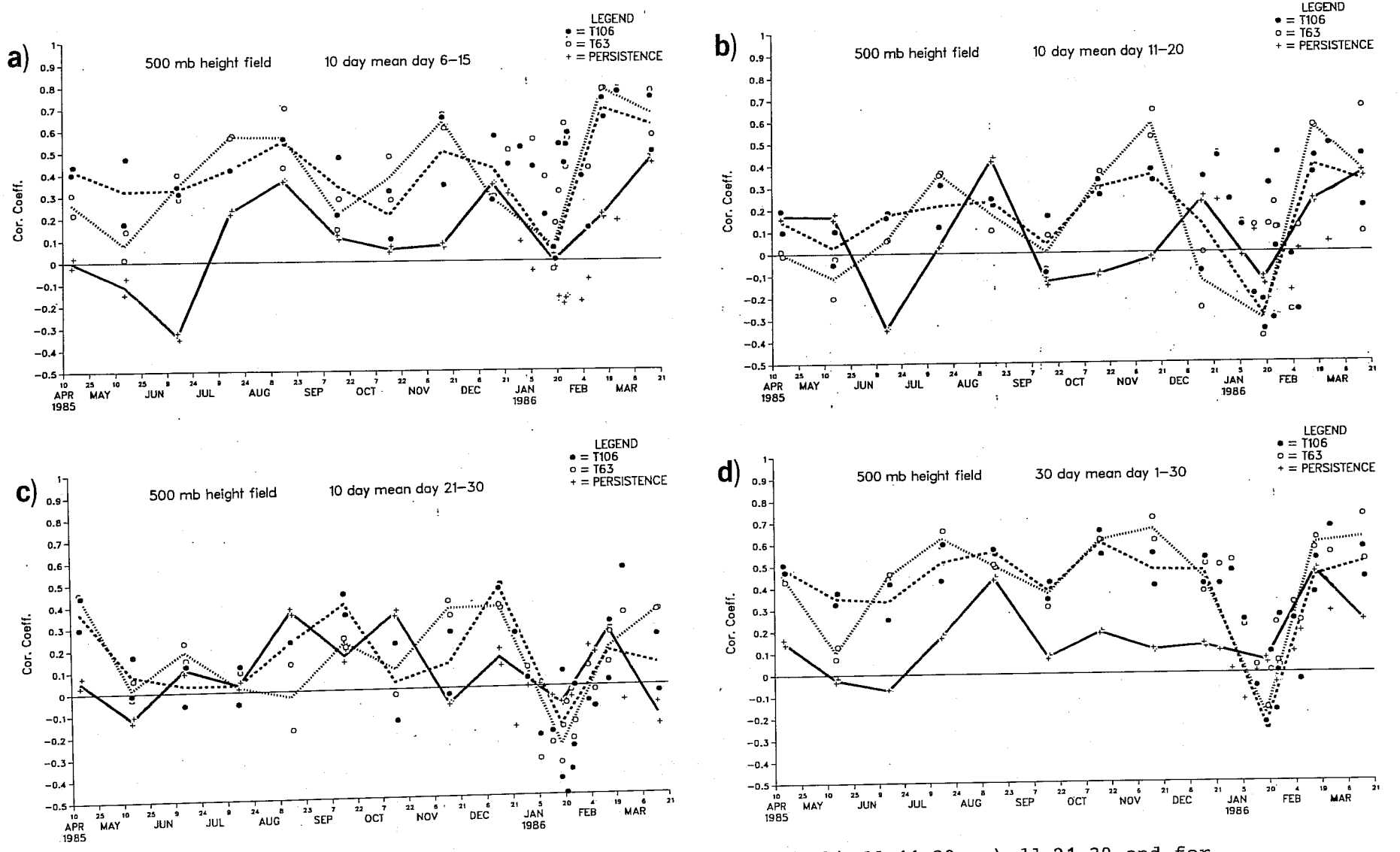

Fig 28 500 mb height ACC for 10-day mean fields: a) dd 6-15, b) dd 11-20, c) dd 21-30 and for 30-day mean fields (d). The lines join the average scores for the monthly pair. Solid: persistence. Dashed: T106. Dotted: T63. All 24 experiments are shown.

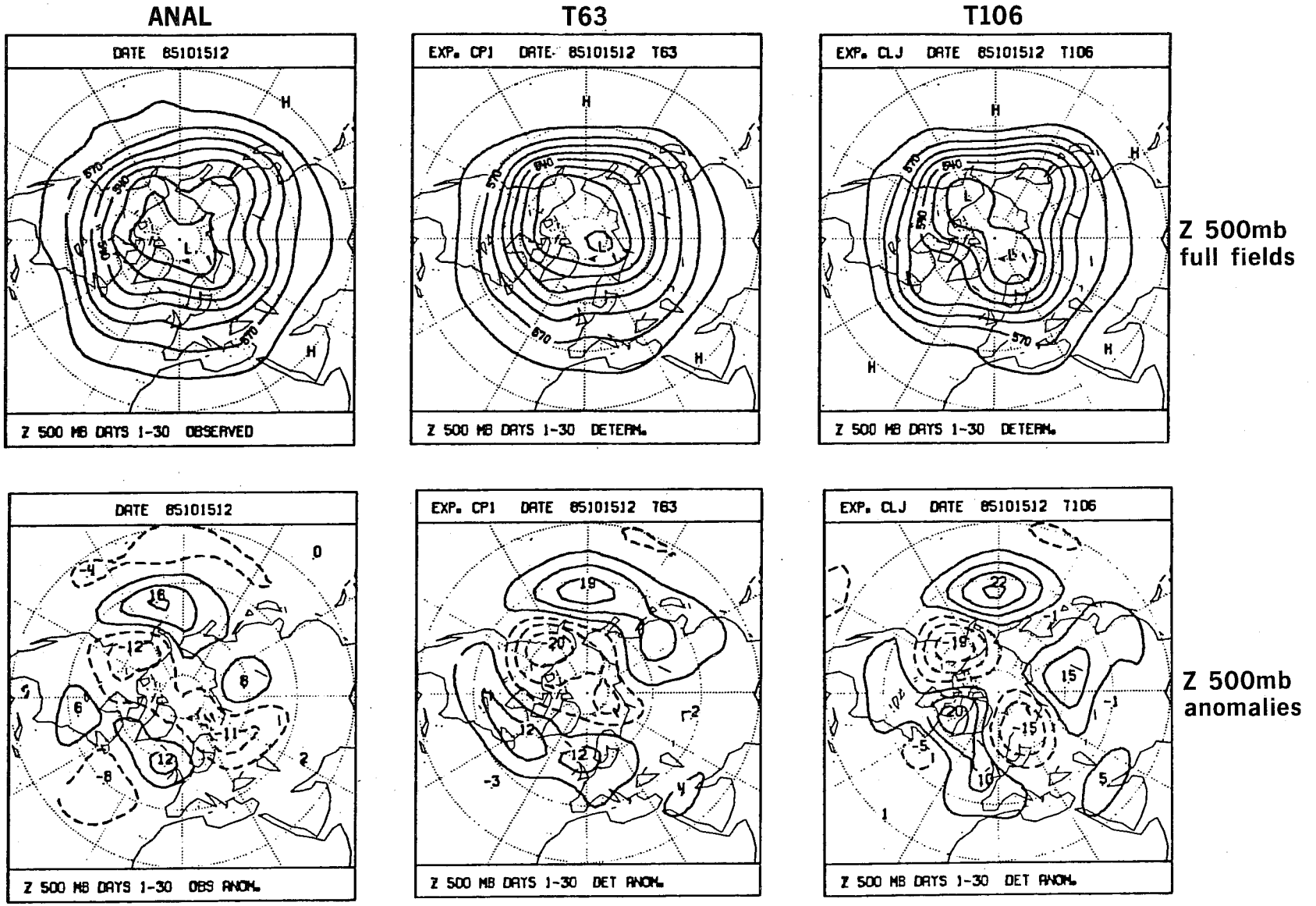

Fig 29 30-day mean 500 mb height fields for the "good" case of October 15, 1985.
Top: full fields. Bottom: anomalies. Left: analysis. Centre: T63. Right: T106.

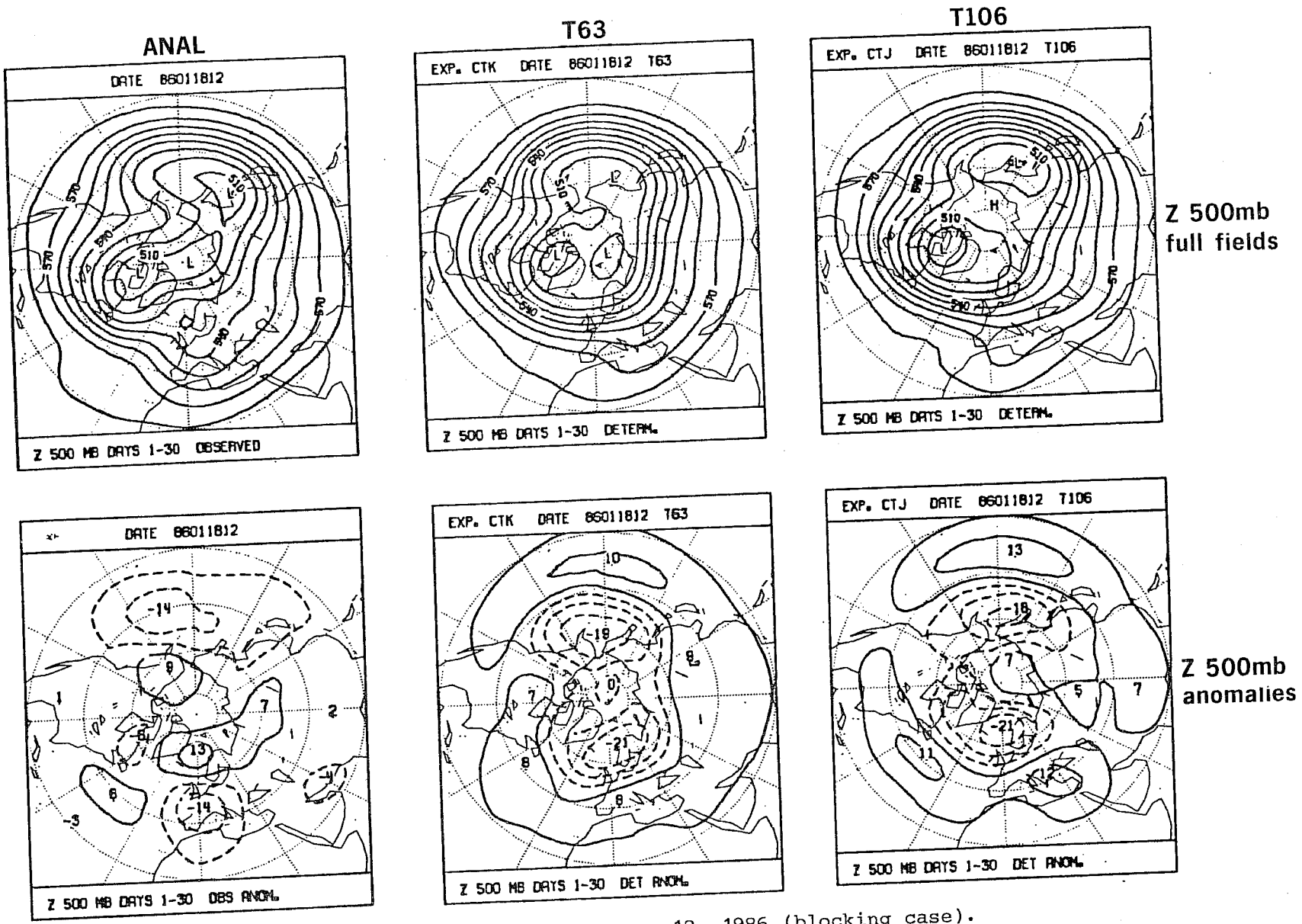

Fig 30 As Fig 28 but for the "bad" case of January 18, 1986 (blocking case).