TECHNICAL MEMORANDUM

201

Error growth and predictability estimates for the ECMWF forecasting system

A.J. Simmons, R. Mureau
and T. Petroliagis

Research Department

September 1994 (SAC Paper)

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

Error growth and predictability estimates for the ECMWF forecasting system

A J Simmons, R Mureau[1] and T Petroliagis

European Centre for Medium-Range Weather Forecasts,
Shinfield Park, Reading, UK

Abstract

Examination has been made of the skill of ECMWF forecasts of the 500 hPa height field produced daily out to ten days ahead, verifying in the period from 1 December 1980 to 31 May 1994. Over this time accuracy has been improved substantially over the first half of the forecast range. The systematic (seasonal-mean) component of the error has been greatly reduced at all forecast times, but there has been little reduction in the non-systematic component later in the range.

The simple model proposed by Lorenz for the intrinsic growth of forecast error has been applied to the evolution of differences between consecutive forecasts. The implied growth rates of small forecast errors have increased significantly since 1981. They do not show much variation with season, and are a little lower in the Southern than in the Northern Hemisphere. The most recent error-doubling times are around 1.5 days for the Northern Hemisphere and 1.7 days for the Southern Hemisphere. Error saturation levels are presently similar to or greater than those of the 1981 version of the model, having been significantly lower in intermediate years. The accuracy of recent short- and early medium-range forecasts and realism of the climatology of the forecast model support the view that estimates of error-growth parameters from the current forecasting system are more reliable than those obtained earlier. Forecast accuracy later in the medium range may thus not have benefitted fully from improvements earlier in the range because of the faster error growth associated with a more active, though more realistic, forecast model. Overprediction of variance may nevertheless detrimentally affect present skill levels and predictability estimates in all seasons other than summer.

The error-growth model currently indicates that it is possible in principle to make deterministic medium-range forecasts for the extratropical 500 hPa height field of the Northern Hemisphere that are as accurate five days ahead as present forecasts are three days ahead, provided the one-day forecast error can be reduced by the same factor in the future as has actually been achieved over the years since 1981. The level of error currently reached at day seven would then be reached at around day ten. The scope for improvement of forecasts over the Southern Hemisphere appears to be rather larger. Improvements seem to be possible throughout the spectral range studied, up to total wavenumber forty. This is found also for the rotational and divergent wind components at 850 and 200 hPa. For these components, particularly the divergent component, there is a quite pronounced error in the representation of the largest scales.

[1] Current affiliation: KNMI, De Bilt, The Netherlands

1. INTRODUCTION

Lorenz (1982) examined the prospects for more accurate numerical weather predictions. He studied the evolution of differences between successive forecasts of the 500 hPa height field produced daily by ECMWF, and verifying in the 100-day period beginning 1 December 1980. He argued that if the forecast model in operational use at the time was realistic enough for small differences in initial conditions to amplify at a rate close to that at which separate but similar atmospheric states diverge, then the rate of growth of differences between consecutive forecasts valid at the same time provided a limit to the potential accuracy of the forecast which could not be surpassed without reduction of the one-day forecast error. Lorenz also calculated a "best fit" to a simple analytical model of this limiting error growth, from which it was possible to estimate the further reduction in forecast error beyond the one-day range that could be expected to result from a given reduction in the one-day forecast error. He concluded that it appeared to be possible to make predictions that would be at least as skilful at ten days ahead as the 1980/81 forecasts were at seven days ahead, and that additional improvements might be realized if the one-day forecast was capable of being improved significantly.

The ECMWF forecasting system has been changed substantially over the more than thirteen years that have elapsed since the forecasts analyzed by Lorenz were made. *Simmons et al* (1989) have reviewed the development of the numerical formulations up to 1987, discussing the replacement of the original 1.875° finite-difference model by a spectral model with T63 horizontal resolution in 1983, the increase in horizontal resolution to T106 in 1985 and the increase in stratospheric resolution in 1986. In 1991 both horizontal and vertical resolutions were increased to their current values of T213 and 31 levels, and a semi-Lagrangian treatment of advection was also introduced (*Ritchie et al*, 1994). Over the same period there were numerous changes to the physical parametrization schemes. Those believed to be the most significant for the evolution of the accuracy of the 500 hPa height forecasts include revised treatments of long-wave radiation, cloud, convection and condensation in late-1984 and 1985 (*Tiedtke et al*, 1988), reduction in vertical diffusion in the free atmosphere in 1988 (*Miller*, 1988), and new schemes for convection (*Tiedtke*, 1989) and radiation (*Morcrette*, 1990) introduced in 1989. The representation of mountain effects has been changed (apart from resolution changes) both by introduction in 1983 (and revision in 1985) of "envelope" orography (*Wallace et al*, 1983; *Jarraud et al*, 1988) and by introduction of a parametrization of gravity-wave drag in 1986 (*Miller et al*, 1989). Throughout the period covered by this study the data assimilation system used the optimal interpolation method (*Lorenc*, 1981); major revisions of the system have been reported by *Shaw et al* (1987) and *Lönnberg* (1988). Recent research effort in data assimilation has concentrated on variational methods, and a one-dimensional scheme for the processing of satellite radiance measurements was introduced for the Northern Hemisphere in 1992 (*Eyre et al*, 1993).

The net effect of these changes (and of changes in the type, coverage and external processing of the observational data available operationally) has been a large improvement in the accuracy of forecasts in the short and early medium ranges, as will be illustrated in this paper. However, conventional skill scores do not show as substantial an improvement in the latter half of the ten-day forecast range. It is important to investigate the reasons for this. Improved understanding is needed to guide decisions regarding the extent to which research effort and computational resources should be allocated between the areas of modelling and data assimilation, and between deterministic and probabilistic (ensemble) forecasting for the medium range.

Lorenz' study was facilitated by extracting relevant fields from the archive of operational ECMWF forecasts to form a purpose-designed 100-day dataset starting 1 December 1980. This dataset also proved useful in

study of the relationship between systematic forecast error and the model's representation of orography (*Wallace et al*, 1983). It was used (together with a similar dataset beginning 1 June 1981) by *Savijärvi* (1984) and by *Dalcher and Kalnay* (1987) in further examinations of predictability, in which dependence on scale and the contribution of systematic errors were discussed. Datasets of the same form have subsequently been produced regularly for 100-day periods commencing at the beginning of each March, June, September and December. Calculations of the type carried out by Lorenz have been repeated from time to time on the newer datasets, and have been reported in several internal ECMWF publications (e.g. *Lorenz*, 1990; *Mureau*, 1990). The datasets have also been used in more specific studies of the practical predictability of blocking (*Tibaldi and Molteni*, 1990) and of the relationship between variations in predictability and fluctuations in the Pacific/North American (PNA) mode of low-frequency atmospheric variability (*Palmer*, 1988).

In this study we utilize the full set of "Lorenz" datasets to examine past and present levels of forecast skill, and to estimate the potential for future improvements. A brief description of the database and calculations is given in the following section. In section 3, errors computed over the extratropical Northern Hemisphere are compared for the years of 1981, 1987 and 1993/94. The component due to systematic (seasonal-mean) error is identified. The following section discusses estimates of predictability based on applying the analysis of Lorenz to the most recent datasets for the Northern Hemisphere. The parameters determined for the simple model of error growth are compared with those calculated for earlier years. In section 5 asymptotic limits are discussed, and some statistics on transient activity are presented in section 6. Here the aim is to illustrate some improved characteristics of recent model versions and to enable discussion of the reliance that can be placed on the more recent predictability estimates. Trends in performance of the forecasting system are also discussed further. A few results for the extratropical Southern Hemisphere are given in section 7, and section 8 contains a spectral breakdown of results for the globe as a whole. A concluding discussion is given in section 9.

2. THE DATABASE AND CALCULATIONS

Each of the fifty-four datasets that form the basis for this study contains data for a different 100-day period. For each day of each period, the 500 hPa height analysis for 12UTC is stored together with the ten operational forecasts that verified at this time. The forecasts were produced daily from earlier 12UTC analyses, and their range thus spans from one to ten days ahead. Fields are represented by the coefficients of spherical harmonic expansions, truncated triangularly at total wavenumber forty.

The datasets cover analysis dates from 1 December 1980 to 8 June 1994. Individual sets begin on the first day of either March, June, September or December. The combined datasets thus contain overlaps of between eight and ten days which have been removed in the calculations to be presented here. We work with three-month seasons: December-February, March-May, June-August and September-November. We refer to these respectively as winter, spring, summer and autumn, except when dealing explicitly with the Southern Hemisphere in section 7. "Winter 1994" will be used to refer to the period 1 December 1993 to 28 February 1994. Data for the three 29 Februaries that occur in the period of investigation are omitted.

We compute a number of measures relating to forecast skill or differences. These are set out in the Annex to this paper. The calculations involve computing temporal and areal averages of a number of quadratic quantities. When the areal average is a global one, this mean can be simply calculated from the spherical harmonic coefficients. If the field X is represented by the spectral expansion

$$\sum_{m=-40}^{m=40} \sum_{n=|m|}^{n=40} X_n^m \, Y_n^m$$

in terms of the spherical harmonics Y_n^m, then the global mean of X^2 is given by

$$\sum_{n=0}^{n=40} S_n$$

where $S_n = |X_n^0|^2 + 2\sum_{m=1}^{m=n} |X_n^m|^2$ for an appropriate normalization of the Y_n^m. The quantity S_n represents the contribution to the global mean of X^2 from the components of X with total wavenumber n, and such quantities are shown separately for each n in the spectral breakdown of results presented in section 8. Here X will in most cases represent either errors or differences of height forecasts. If instead X represents either errors or differences of forecasts of vorticity or divergence, then the corresponding global-mean squared error or difference of the rotational or divergent wind forecasts is given by

$$a^2 \sum_{n=1}^{n=40} \frac{S_n}{n(n+1)}$$

where a is the radius of the earth.

Most of the results presented here are not for global means, but for means over the extratropical hemispheres. These have been calculated numerically from field values computed for a regular 3.75° grid. The accuracy of this calculation has been checked by computing global root mean square errors of 500 hPa height directly from the spherical harmonic coefficients, and from the 3.75° grid values, for the winter of 1994. The largest difference was a negligible 0.02m at day nine.

Several of the calculations require knowledge of a climatology of the 500 hPa height. We have used analyses from the whole database to produce a climatological field for each day of the year, and these fields were used in calculating the results presented here. They represent averages over only either thirteen or fourteen years. To check that this restricted sampling period was not a serious limitation, some of the calculations were repeated with the daily climatological fields smoothed by applying a seven-day running mean. No substantial differences were found. For example, the anomaly correlations presented in Table 1 in the following section change by a maximum of 0.4% for winter and 0.7% for summer.

3. TOTAL AND SYSTEMATIC ERRORS FOR THE EXTRATROPICAL NORTHERN HEMISPHERE

Root mean square errors computed for the earliest and latest years available (1981 and 1993/1994) are presented in Fig 1 for each of the four seasons. Means are over the extratropical Northern Hemisphere. In each case both the total error and the non-systematic error are plotted. For 1981 these can be distinguished as the upper and lower dashed curves separated by stippling. The upper and lower limits of the thick black lines denote the results for 1993/94. The non-systematic (or transient) error was computed using daily analysis and forecast fields from which corresponding seasonal-mean values had been subtracted.

It is clear from Fig 1 that there has been a substantial reduction in forecast errors over the past twelve or thirteen years. Improvement is seen across the whole ten-day range, but is most apparent earlier rather than later in the range. The reduction in the systematic component of the error over the period is evident. By

day 10, most of the reduction of total error is accounted for by reduction of systematic error. In 1981 error grew approximately as a linear function of forecast range out to five days ahead; today the error curves have more of the exponential form expected from simple models of the growth of small analysis errors (e.g. *Lorenz*, 1982; *Trevisan*, 1993).

Fig 2 is a repeat of Fig 1, but with the results for 1987 replacing those for 1981. It shows that much of the reduction in systematic error since 1981 came from changes made to the forecasting system prior to 1987. Since 1987 there has been a continued reduction in root mean square errors in the first half of the forecast range. At day ten, however, the error has, on the whole, actually increased since 1987. Discussion of the significance of this will be given later.

We have seen that the systematic component of forecast error is now small when computed on a seasonal basis for the extratropical Northern Hemisphere. This is found also when the calculations are carried out for smaller domains such as Europe. The earlier studies by *Palmer* (1988) and *Tibaldi and Molteni* (1990) examined the dependence of error on flow regime. The current version of the forecasting system is much more successful at simulating blocking in the medium range (*Tibaldi et al*, 1994), but the time-mean component of forecast error is nevertheless evident when the averaging period covers a few weeks, rather than a season. To illustrate this we have computed total and transient (non-systematic) errors for separate 20-day periods within the most recent 100-day winter dataset. Results for the extratropical Northern Hemisphere are presented in Fig 3, along with the corresponding result for the full 100-day period. There is clearly variability in the 20-day mean error from spell to spell, but this error is generally several times larger than the seasonal-mean error. Comparison with previous years shows that typical 20-day mean errors have been substantially reduced since the early days of operational forecasting at ECMWF, though not by as large a factor as for the seasonal-mean error.

Examination has also been made of anomaly correlations: the correlations between forecast and actual deviations from climatology. Here, results are presented in tabular form to emphasize the similarity between the exact form of the anomaly correlation coefficient as formally defined in the Annex and an approximate form computed neglecting the areal- and time-mean deviations of analyses and forecasts from climatology. This approximate form can be related simply to the mean square forecast error normalized by the sum of the mean square deviations of analyses and forecasts from climatology (see *Boer*, 1994, or Annex). Results for the winters and summers of 1981, 1987 and 1993/94 are presented in Table 1. Three numbers are tabulated for each of the seasons shown. The left-hand column shows the exact form of anomaly correlation (in %) and the central column the approximate form. The third column shows the mean deviation of the analyses and forecasts from climatology.

The interannual and interseasonal differences in anomaly correlations are much larger than the differences between the exact and approximate forms of anomaly correlation given in Table 1. These measures of forecast skill confirm the substantial overall improvement since 1981 seen already for the root mean square errors. Anomaly correlation gives a rather more favourable view of the improvement since 1987 towards the end of the forecast range. The relationship between the approximate form of anomaly correlation and the mean square error can be used to show that this more favourable view is a consequence of a recent increase in the level of variance simulated by the forecast model, as will be illustrated later in this paper.

The mean deviations from climatology for 1981 show a substantial net lowering of height as the forecast range increases, associated with a systematic error pattern dominated by underestimation of the height field over the Atlantic and Pacific Oceans. Neglecting this in the calculation of the anomaly correlation lowers the correlation by more than one percentage point in the second half of the forecast range. The systematic lowering of heights was reversed by the various model changes made in late 1984 and 1985 (*Tiedtke et al*, 1988), so the results for 1987 show a gradual increase in mean height as the forecast proceeds. The most recent results indicate a small lowering of heights in winter, and the opposite trend in summer. It should be noted, however, that the areal-mean climatology used in these calculations is subject to errors of a few metres; *Simmons and Chen* (1991) showed that the post-processing in use up to May 1990 could introduce a negative bias of almost five metres in the 500 hPa height.

4. POTENTIAL IMPROVEMENTS IN FORECAST SKILL

Results relating to estimates of the potential improvement in forecast skill are presented in Fig 4. These are for the extratropical Northern Hemisphere and are derived for each season from the latest year for which data are available. The solid lines show root mean square forecast errors, and simply repeat for reference purposes the plots of total error shown in Figs.1 and 2.

The dashed curves in Fig 4 display the root mean square differences between consecutive forecasts, a measure of what is sometimes referred to as forecast inconsistency. The result plotted for day j is the difference between the day j forecast made j days previously and the day $j-1$ forecast made one day later. The points for day one coincide with the one-day forecast errors. It is these dashed curves that, following Lorenz' arguments, provide an estimate of the limit of forecast skill that would be reached by a model which behaved perfectly beyond day one, given the current one-day forecast error. This is under the assumption that the current model, though evidently not perfect, is sufficiently accurate to represent realistically the divergence of two forecasts which start with relatively small initial differences. In practice, of course, model developments which improve performance beyond day one are also likely to lower the day-one error, both through a direct reduction of error growth during the first day, and from use in data assimilation to lower analysis error.

The scope for improvements of forecast skill from model improvements beyond day one can be measured in terms of the extension of the forecast range at which a certain error level is reached. This scope appears from Fig 4 to be largest in summer. For this season it corresponds to an extension of range by about one day at the skill level currently reached around day four or five, and over two days at today's seven- or eight-day level. The calculations shown here for winter are not exactly as carried out by Lorenz for the winter 1981 dataset, but when applied for 1981 they give very similar results to those quoted by Lorenz. For each season, the separation between the actual error growth and the "perfect-model" error growth has approximately halved since 1981. There thus now appears to be a significantly lower scope for improvement than found earlier. The model is presumably now closer to being perfect beyond day one than it was in 1981, but there are also indications that the early operational forecasts provided an underestimate of the intrinsic error doubling time.

Lorenz (1982) proposed use of a simple model of the dependence of the root mean square error, E, of a sequence of "perfect-model" forecasts on their range, t. The growth of error is given by:

$$\frac{\partial E}{\partial t} = \alpha E \left(1-\frac{E}{E_\infty}\right) \tag{1}$$

Parameters of the model are the rate α at which small error grows early in the range, and the asymptotic level E_∞ at which error saturates. The doubling time of small errors is $ln2/\alpha$.

The parameters α and E_∞ have been determined as in Lorenz' study by a weighted least-squares fit of root mean square differences between consecutive forecasts to a finite-difference approximation of equation (1). The approximation is:

$$\frac{\Delta E}{\Delta t} = \alpha \overline{E} \left(1-\frac{\overline{E}}{E_\infty}\right) \tag{2}$$

where Δt = 1 day, and the data points to be fitted are given by:

$$\Delta E_j = D_{j+1} - D_j$$

and

$$\overline{E_j} = \tfrac{1}{2}(D_{j+1} + D_j)$$

Here j = 1,2,3, ... 10 days, and D_j is the root mean square forecast difference between day j and day$j$–1 forecasts verifying on the same day. D_1 is equal to the root mean square error of the one-day forecast, E_1.

The resulting approximations to the growth of forecast differences for the most recent four seasons are shown by the dotted curves in Fig 4. These curves are in fact extremely close to the corresponding dashed curves which they approximate, agreement being at least as good as when the calculation is carried out for the original 1981 data. Indeed, the fits might be regarded as surprisingly good, since *Stroe and Royer* (1993) found a better fit to a somewhat different form of error-growth equation than to (1), for a set of extended-range winter forecasts. However, Stroe and Royer carried out least square fits of *mean square* forecast differences, whereas here (and in *Lorenz*, 1982) the fit is of *root mean square* differences.

Stroe and Royer (loc. cit.) examined an error-growth equation of the more general form:

$$\frac{\partial V}{\partial t} = \frac{\alpha}{p} V \left(1 - \left(\frac{V}{V_\infty} \right)^p \right) \tag{3}$$

examining sensitivity to the value of the parameter p. Here V represents the squared error E^2. Equation (3) is analytically equivalent to Lorenz' error-growth equation (1) in the case p=0.5, but does not give exactly the same results as illustrated in Fig 4 when α and V_∞ are determined by a least-squares fit to (3) of squared forecast differences. p=1 corresponds to the calculations of *Dalcher and Kalnay* (1987). We have evaluated fits to (3) of squared forecast differences for several values of p, for each of the most recent seasons. For winter, our results agree with those of Stroe and Royer in that we find a better fit for a value of p smaller than 0.5, although the optimal value of around 0.3 is rather larger than found by them. In this case the

quality of fit is comparable with that shown in Fig 4 for the fit of the Lorenz model to root mean square differences. However, to obtain the same degree of fit for the other three seasons we need to use values of p around 0.7 in summer and 0.5 in spring and autumn.

We now use Lorenz' representation of perfect-model error growth, with the latest determinations of the error growth rate and asymptotic limit (from the root mean square fits), to estimate the level of forecast error that would result beyond day one from a particular reduction in the one-day error. These estimates are shown for the four seasons by the dash-dotted curves in Fig 4. The assumed reductions in one-day errors are not arbitrary. Rather, these errors are reduced by the same factors by which the non-systematic components of one-day error have actually been reduced over the period since 1981. We have neither a guarantee that a further such reduction is possible, nor an estimate of the period of time over which such an improvement might be achieved. However, if such an improvement is indeed achievable, then the scope for extending the forecast range increases to about two days at the level of skill now reached at day three, and to between two and a half and three and a half days towards the end of the ten-day range.

Fig 5 shows results from the error-growth models plotted for a forecast range extended to twenty days. The starting points are the actual and assumed future one-day error levels as in Fig 4. Results are shown for the Lorenz model as presented to day 10 in Fig 4, and for the model of Stroe and Royer using the optimal values of p indicated above. The two curves starting from the higher, actual day-one errors are derived from fits to the actual forecast differences out to day ten, and thus agree well over this time range. For winter and summer they diverge rapidly beyond day ten, as the two error-growth models give distinctly different asymptotic levels. For all seasons, however, the two models are in quite close agreement out to day ten when started from the lower day-one errors. Of the two models, Lorenz' gives a slightly more optimistic picture in winter, throughout the forecast range. The Stroe and Royer model with values of p less than 0.5 (as used for winter) gives, however, a slower approach to the asymptotic limit. Generally, the apparent scope for error reduction drops quite quickly beyond day ten, and is extremely small beyond the two-week range. Further comment on asymptotic limits is given in the following section.

The parameters of Lorenz' error-growth model have been evaluated for each season since 1981. Doubling times and asymptotic limits are shown for the extratropical Northern Hemisphere in Table 2. The doubling times decrease from around two days or a little less early in the period to current values of about one and a half days. There is no substantial seasonal variation, contrary to an expectation expressed by *Lorenz* (1982). Dynamical instabilities might be expected to be stronger for the more intense flows that occur during the winter months, with diabatic processes contributing more to forecast divergence in the summer months. If anything, error growth rates are largest in autumn.

The asymptotic error limits decline substantially from their 1981 values to reach minimum values around the years 1985 to 1987. They increase thereafter, so that current values for spring, summer and autumn are rather close to their values in 1981, and recent winter values are higher than ever before. Clearly there is no simple direct relationship between the error growth rate and asymptotic limit, a conclusion reached also by *Stroe and Royer* (1993). Dynamical processes causing error growth include baroclinic and Rossby-wave instabilities (*Lorenz*, 1972). However, idealized calculations show that the amplitude reached by mature baroclinic waves cannot always be simply related to initial growth rates, and that the potentially rapid

barotropic decay process depends on the scale as well as the amplitude of the mature wave, and on the ambient large-scale flow (e.g. *Simmons and Hoskins*, 1980).

Variations from one year to the next in monthly-mean forecast skill can be quite large, even when no substantial change has been made to the forecasting system over the intervening period (*Simmons*, 1986). Thus some fluctuations in error-growth parameters may be related to changes in circulation regime rather than to changes in the forecasting system. Nevertheless, the large changes from year to year seen in Table 2 are believed to represent primarily the effect of changes to the forecasting model. In particular, it is known that a number of changes made since 1987, notably the reduction in vertical diffusion in the free atmosphere, the new parametrizations of radiation and convection, and the increased resolution (and associated reduced horizontal diffusion), have undoubtedly increased the degree of eddy activity in the model. More accurate short- and early medium-range forecasts, and more realistic extended-range simulations have resulted, but a side-effect seems to have been a faster error doubling time (a possibility anticipated by *Lorenz*, 1982) and a higher error saturation level, leading to root mean square errors at day ten that are higher today than in 1987.

Some evidence that greater reliance can be placed on current rather than earlier predictability estimates will be given in following sections, and some of the remaining uncertainties will also be mentioned. For the present, we note that the particularly high values of the asymptotic error limit for spring and summer 1992 are associated with a known tendency of the operational system at the time to predict excessive eddy activity. This was due partly to a deficiency of the version of the semi-Lagrangian advection scheme in use at the time (*Ritchie et al*, 1994), and partly to a deficiency in cloud/radiation interaction when the model was run with the 31-level vertical resolution introduced operationally the previous autumn.

The indication from the simple error-growth model is that an assumed future reduction of the one-day forecast error will result in reduction of error across the rest of the ten-day forecast range if the model is unchanged. Just such a result is found if past forecasts are sorted according to the magnitude of the one-day error. Forecasts from each season for the most recent two years have been categorized on the basis that the one-day error is above or below its average value for the season, and averages of the two categories have been computed. For each day of the two-year period, mean square errors or forecast differences have been scaled by dividing by the climatological variance for that particular day of the year, and multiplying by the seasonal mean climatological variance, to compensate for bias due to the annual cycle in variance of the 500 hPa height field. For these calculations the seasonal average has to be with respect to the starting date of the forecast rather than the verifying date. Because of model changes during the period, the categorization has been done separately for the two years, before formation of the averages.

Results are presented in Fig 6 for both total forecast errors and differences between consecutive forecasts. Across the whole forecast range the root mean square errors and differences of forecasts that have below average one-day errors are smaller than the errors and differences of forecasts which have above-average one-day errors. Towards the end of the forecast range the separation appears to be generally rather larger for the forecast differences (the "perfect-model" errors) than for the actual errors.

The predicted reduction in short-range forecast error due to reduced initial error has been checked against a dataset with twelve-hourly rather than daily frequency of analyses and forecasts. The ECMWF model is currently run routinely to three days ahead from 00UTC analyses produced with a cut-off time for data receipt

that is shorter than used for the 12UTC analyses. The purpose is to provide timely boundary conditions for the limited-area prediction models of a number of weather services. The model version differs from that employed for the 12UTC forecasts in its use of a longer timestep and modified semi-implicit time differencing (*Ritchie et al*, 1994). However, despite the differences in data cut-off and model parameters, root mean square errors of the 00UTC forecasts are very close to those of the 12UTC forecasts for the extratropical Northern Hemisphere. A dataset of analyses and forecasts to three days ahead has thus been produced for winter 1994 combining the 12 and 00UTC analyses and forecasts.

The solid curve in Fig 7 shows the root mean square error of the combined set of forecasts, and the dashed curves the corresponding differences between consecutive forecasts started twelve hours apart. The solid circles denote the errors computed for the 12UTC forecasts and analyses alone. They are centred only just below the solid line, confirming that the errors of the 00UTC forecasts are only slightly larger than those of the 12UTC forecasts.

The growth of forecast differences shown in Fig 7 is matched quite well by the dotted curve included in the figure. This curve was computed from the Lorenz model using the error growth rate and asymptotic limit derived from the daily data. The agreement indicates that the Lorenz model gives a reasonable estimate of the growth of differences starting from the smaller twelve-hour forecast errors, despite having been fitted to differences which grew from the larger one-day errors.

5. ASYMPTOTIC LIMITS

Information relating to the asymptotic limits of the root mean square errors and differences is presented in Fig 8. Here the solid and heavy dashed curves simply repeat curves shown in Fig 4, showing for reference purposes the forecast errors and differences between consecutive forecasts for the latest seasons. The dotted curves denote the square root of the sum of the mean square deviations from climatology of the analyses and forecasts. Under assumptions specified in the Annex, the solid curves representing forecast error would asymptote to these dotted curves at a sufficiently long forecast range and for a sufficiently large sample of forecasts. The dash-dotted curves represent the square root of twice the mean square deviation of the forecasts from climatology, and (also under conditions discussed in the Annex) they provide the asymptotic limits for the heavy dashed curves representing forecast differences. The short, lighter dashed lines show the asymptotic limits calculated from fitting Lorenz' model of error growth to the forecast differences.

Perhaps the most striking aspect of Fig 8 is the discrepancy between the asymptotic limit computed simply from statistics of the forecast deviation from climatology and the asymptotic limit computed from Lorenz' model. The results shown in Fig 5 suggest that factors which could contribute to this are the choice of asymptotic term in the error-growth equation, and the way the data out to day ten are fitted. Two other related factors are extended-range predictability and regime-dependent systematic error. If some component of the extratropical 500 hPa height field is predictable on a time-scale substantially longer than ten days, due for example to long-lived anomalies in sea-surface temperature or soil moisture, then it is unlikely that the net error of perfect-model forecasts would be described by an equation of as simple a form as (1). Instead, there would be both a rapid element of error growth due to relatively short time-scale dynamical instabilities, and a slower element of error growth associated with the more predictable low-frequency component of the atmospheric evolution. Separate error-growth equations were used for large and smaller scales by *Schubert and Suarez* (1989) to improve the representation of the growth of small perturbations in a simple two-layer general circulation model.

Although systematic error is now small when calculated on a seasonal basis, it has been shown in Fig 3 that it is much more obvious when computed for 20-day means. If consecutive forecasts tend to lie preferentially in the same regime of low-frequency variability at day ten (which itself suggests some extended-range predictability or model bias), then any regime-dependent systematic error would result in a correlation between the deviations from climatology of consecutive forecasts. As shown in the Annex, mean square differences between consecutive forecasts would then tend to be smaller than expected from the mean square differences between forecasts and climatology.

Comment must also be made on the differences between the dotted and dash-dotted curves in Fig 8, and the upward slopes of these curves. These are seen in all seasons other than summer. Outside this season, the spatial variance of the forecast 500 hPa height field tends to increase as the forecast range increases. It is likely that this is a consequence of a defect of the model, rather than an artefact of an analysis which underestimates the true variance, as there has been relatively little fluctuation in analysis variance from year to year since 1987, despite significant changes to the forecasting system. If the model is indeed in error, its excessive variance contributes to root mean square forecast error in the latter half of the forecast range. Moreover, this model defect contributes more to the difference between consecutive forecasts than to the forecast error, at least in the asymptotic limit. Thus if the defect were to be removed, it is expected that root mean square errors would be reduced, and estimates of the potential benefit of further model improvement beyond day one would increase, for the autumn, winter and spring seasons. The apparent scope seen earlier for a greater improvement in summer than in other seasons may thus be illusory.

Error in the level of forecast variance is compared for the years of 1981, 1987 and 1993/94 in Fig 9. The differences between the root mean square amplitudes of forecast deviations from climatology and of analyzed deviations are presented. For summer, the result for 1993 lies close to the zero line, with just a slight underestimation of variance by the forecast model that is barely perceptible on the scale of Fig 8. Variance was much more substantially underestimated in 1987, and overestimated in 1981. For the other seasons, variance is overestimated at the end of the forecast range in each year, though by least in 1987, and the error in the level of variance for this year is very small earlier in the range. There is little to choose between 1981 and 1993/94 near the end of the range, but earlier in the range there is a much more substantial overestimate of variance in 1981.

It is important to recall that the results we have been discussing are for averages over the extratropical Northern Hemisphere. The performance of the current model compared to that of 1987 is seen in a better light if the spatial distributions of error in the prediction of variance are examined. Maps are shown for three-day and seven-day forecasts from the winters of 1981, 1987 and 1994 in Fig 10. The quite substantial overestimation of variance in the 1981 forecasts is evident, particularly at the earlier range. However, the comparison between 1987 and 1994 does not so obviously favour the earlier year. Erroneous levels of variance are quite widespread in the 1987 forecasts, but there are extensive regions of both positive and negative values, so that the integral over the extratropical hemisphere yields a relatively low net error. Error is lower over much of the hemisphere in 1994, but is predominantly positive, contributing to a much larger areal-mean error than in 1987. The relatively large localized overpredictions of variance peaking over Norway at day three and to the west of the British Isles at day seven are alarming features of the 1994 results from a European viewpoint. There is, however, a very substantial interannual variability in the patterns of error. The predominantly positive nature of the error has, nevertheless, been a feature of all but one of the

winters since 1988. The poorest results were in 1992, due to the previously mentioned problems with the new high resolution semi-Lagrangian forecast model.

6. TRANSIENT ACTIVITY AND ADJUSTED FORECAST ERRORS

As an overall measure of transient activity in the model we have computed the root mean square change from one day to the next of the analyses and forecasts. Here the forecast result for day j refers to the root mean square difference between day $j-1$ and day j forecasts which start from the same analysis. Fig 11 plots the difference between forecast and analyzed values for each season for the years 1981, 1987 and 1993/94.

The performance of the forecasting system over the past year is superior to that of 1987 for all seasons and time ranges. Results are poorer than 1981 only towards the end of the range for winter. The forecasting system in 1981 and 1987 evidently underestimated the mean square rate of change in time of the 500 hPa height field, whereas the most recent system produces about the right level of change in autumn and for the first half of the range in winter. Over the past year the rate of change was still underestimated in summer and for much of the range in spring. It was overestimated in the second half of the forecast range in winter and near the end of the range in spring. Examination of maps of the geographical distribution of error confirms that the best results are those obtained recently.

These results indicate that the shorter "perfect-model" error doubling times computed from recent forecast results are not arising from an obviously excessive transient eddy activity in recent versions of the model. Moreover, the longer doubling times computed earlier came from model versions which underestimated temporal change, despite an overprediction of spatial variance by the 1981 version. Given the lower short and early medium-range forecast errors and better climatology of the current forecasting model, we have no reason from these results to suppose that recent error doubling times are unrealistically short.

We now reconsider the evolution in forecast skill over the years covered by the "Lorenz" datasets. The upper panel of Fig 12 plots the root mean square error of the 500 hPa height field calculated over the extratropical Northern Hemisphere for each season at forecast ranges of one, three, five, seven and ten days. Running means of four consecutive seasons are plotted to smooth out the annual cycle. The plots show a steady decline of error at the one-day and three-day ranges. Errors declined also at longer forecast ranges until 1986. They have fallen again recently at these ranges. In between, root mean square errors changed little at the five-day range, and grew at days seven and ten.

The middle panel of Fig 12 shows the evolution of root mean square error normalized by the square root of the sum of the mean square deviations of analyses and forecasts from climatology. It is shown in the Annex that this score is closely related to the anomaly correlation coefficient. This normalized error has evolved largely as the root mean square error itself, but a gradual decline can be seen in the years following 1986 at days five and seven, and the ten-day score does not show the rise during this period seen for the un-normalized error.

Normalizing the root mean square error in the above way ensures that the measure of forecast skill does not, in the asymptotic limit, favour a model version with a relatively low level of variance (whether or not this level of variance is more realistic than that of other model versions). However, it does not compensate for differences in the growth rate of small errors. A larger growth rate implies an earlier approach to the asymptotic limit. The normalized errors have thus been adjusted to compensate for differences in the error

growth rate from year to year. The premise is that the faster growth rates of recent model versions are more reasonable than those of earlier versions, and thus that the earlier forecasts exhibited fortuitously low errors because the model at the time underestimated transient activity. The adjustment applied is multiplication by a factor deduced from Lorenz' model of error growth. Two error growth curves are computed for each season and year, both using the one-day error and asymptotic limit for that season and year. One curve is computed using the error growth rate calculated for the particular season and year, and the other using the growth rate calculated for the corresponding season from the most recent year. The ratio of the two provides the scaling factor used to adjust the normalized root mean square errors.

The adjusted curves are shown in the lower panel of Fig 12. The method of adjustment is such that the day-one curve is unchanged for all years, and the right-hand end points are unchanged for each forecast range. Otherwise, errors are generally increased. The adjusted curves show a more substantial improvement over the years in the latter half of the forecast range. They also show a larger impact of the changes made to the forecasting system over the past three years.

7. THE EXTRATROPICAL SOUTHERN HEMISPHERE

The calculations reported in preceding sections have also been carried out for the extratropical Southern Hemisphere. Fig 13 shows root mean square errors, differences between consecutive forecasts, and curves from Lorenz' model of error growth, in the same form as presented in Fig 4 for the Northern Hemisphere. The errors are generally larger for the Southern Hemisphere, though towards the end of the forecast range this appears to reflect largely the higher natural level of variance in this hemisphere. The reduction of error in the early part of the forecast range since 1981 has in fact been larger for the Southern Hemisphere, so there is a greater separation between the dash-dotted and dotted curves in Fig 13 than in Fig 4. The systematic component of the total error in 1981 was much less for the Southern than for the Northern Hemisphere, so there has not been the scope for as striking a reduction in this component.

Lorenz (1982) found that the doubling time for small errors computed for the Northern Hemisphere alone was shorter than the value found when the calculation was carried out for the whole globe. He believed this was because the data he was examining related to winter in the Northern Hemisphere and summer in the Southern Hemisphere. In fact, we have already seen that error doubling times for the Northern Hemisphere differ little across the seasons of the year. Moreover, we find doubling times to be generally longer for the Southern than for the Northern Hemisphere. For the most recent seasons they lie in the range 1.4 to 1.6 days for the Northern Hemisphere and 1.6 to 1.8 days for the Southern Hemisphere. Asymptotic limits computed for Lorenz' model of error growth are higher for the Southern Hemisphere, again reflecting its higher natural variance.

The solid and dashed curves in Fig 13 are generally more widely separated than the corresponding curves shown in Fig 4. This suggests a greater scope for model improvement in the Southern than in the Northern Hemisphere. This result is achieved despite the fact that the current version of the forecast model produces levels of variance in the Southern Hemisphere that exceed those of the analyses in all seasons, and by amounts larger than shown in Fig 9 for the Northern Hemisphere. The potential benefit from model improvement deduced from Figs. 4 and 13 may thus be underestimated more for the Southern than for the Northern Hemisphere.

The differences between the root mean square rates of change of the forecasts and analyses from one day to the next are presented in Fig 14 for the Southern Hemisphere. Comparison should be made with Fig 11, which shows the corresponding results for the Northern Hemisphere. In (austral) summer the forecast model currently produces about the same mean square rate of change as analyzed, an improvement over earlier years. However, for the other seasons the rate of change appears now to be overestimated, and this aspect of model performance seems to be poorer than in the earlier years shown.

The possibility that the verifying analyses underestimate the true variance of the atmosphere cannot, however, be discounted in interpreting these results for the Southern Hemisphere. Satellite radiance data are a more critical component of the observing system in the Southern than in the Northern Hemisphere, and *Eyre* (1987) has noted that the retrieval method used for them may result in analyses with erroneously low spatial and temporal variances. Research aircraft measurements made in the Southern Hemisphere have shown that both analyses and forecasts underestimated wave amplitudes in late winter 1987 (*Tuck et al*, 1989). Recently, the experimental application of a one-dimensional variational analysis of satellite data (*Eyre et al*, 1993) to the Southern Hemisphere resulted in a significant increase in the eddy kinetic energy of initial states (McNally, personal communication).

8. SPECTRAL PROPERTIES

Savijärvi (1984) and *Dalcher and Kalnay* (1987) examined the spectral properties of squared global ECMWF forecast errors and differences for winter and summer 1981. A spectral analysis of errors of short-range forecasts from the Canadian Meteorological Centre was reported by *Boer* (1984), and extended-range forecast experiments using a variant of the French operational model provided the data on which the spectral study by *Stroe and Royer* (1993) was based. In this section we illustrate how some of the spectral properties of the ECMWF forecasts have changed over the years of operational prediction since 1981. The spectral breakdown is with respect to total wavenumber n, as indicated in section 2. Results are presented only for the December to February period. No significant differences in these spectra of global-mean quantities have been found for other seasons. Here we are limited to the spectral range up to total wavenumber forty because of the truncation of these datasets. Results for the complete spectral range of the current model are presented by *Boer* (1994) for the month of February 1993, including a split into stationary and transient components.

The left-hand panels of Fig 15 show the spectra of squared forecast errors averaged for the winters of 1981 and 1994, for several days out of the ten-day range. Corresponding spectra of differences between consecutive forecasts are presented in the right-hand panels; as in section 4, the results plotted for day j are the spectra of mean squared differences between the day j forecasts made j days previously and the day $j-1$ forecasts made one day later. Results are presented on log/linear plots to emphasise higher wavenumber behaviour, and in particular the approach to saturation. Moreover, forecast errors and differences in the early part of the forecast range in fact have a dependence on wavenumber for large n that is closer to linear in a log/linear plot than in a log/log plot. This is true also of later forecast ranges for 1981. It is only at later ranges for 1994 that a simple power law appears to apply, with a high-wavenumber dependence of around $n^{-4.4}$ for these squared errors and differences of the 500 hPa height field.

The forecast errors for 1981 show clearly that the higher wavenumber components of error saturate sooner than the lower wavenumber components, as noted in previous studies. Qualitatively similar behaviour is seen for 1994, but saturation occurs noticeably later in the forecast range in 1994 than in 1981. The higher

wavenumber components of the error are however of larger amplitude for 1994 than 1981. The relatively large global-mean (n=0) component of the error for 1981 is partly due to the systematic lowering of heights over the extratropical Northern Hemisphere noted in our earlier discussion of Table 1, but a major contribution came from the tropics. This component of the error did not become substantially smaller until after the model changes in May 1985 discussed by *Tiedtke et al* (1988).

As also discussed earlier, the growth of the forecast differences provides an estimate of the potential reduction in error due to model improvement, for fixed one-day error. As might be hoped, the results for 1981 and 1994 are more consistent as regards saturation of forecast differences. For both years the day-five curve lies clearly separate from the day-ten curve out to wavenumber forty, and the day-seven curve is just distinguishable from the day-ten curve for some wavenumbers beyond thirty. The actual error curve for 1994 shows the day-five curve merging with the day-seven and day-ten curves at a wavenumber close to forty. There thus appears to be at least some scope for model improvement alone to bring about a reduction of the five-day forecast error across the whole spectral range studied.

It should be noted, however, that a more pessimistic result is found when spectra of forecast differences are examined for some intermediate years. In particular, for 1985 the day-five curve merges with the day-seven and day-ten curves at around wavenumber thirty. Thus although hemispheric and global-mean estimates of the potential for forecast improvement were rather encouraging around that time, due to relatively low error growth rates and saturation limits, the outlook for improved medium-range prediction of the shorter scales was poorer than at present.

The spectra of forecast errors and differences peak quite sharply around wavenumber seven or eight, at about day five and beyond. The spectrum of one-day error has a particularly broad maximum for 1994, and for this year the day-three spectrum of forecast differences peaks at wavenumber ten. There is generally a rather weak tendency for maximum amplitude to shift to longer scales with increasing forecast range, which is seen more for forecast differences than for forecast errors, and more for recent than for earlier years. Since also the shorter scales saturate sooner than the longer scales, there is a marked increase in the characteristic length scale of forecast errors and differences as the forecast range increases.

Spectra of the one- and three-day errors are presented in Fig 16 for several years. The upper panels show the basic spectra, and the lower panels show spectra normalized by dividing each spectral component by twice the corresponding spectral component of the squared deviation of the analysis from climatology. This normalizing factor is the asymptotic error limit for a perfect model, under the assumptions discussed in the Annex. The basic spectra show that for the most part the one-day error has been reduced across the whole spectral range over the years since 1981. The main exception is the curve for 1981 itself, which indicates a much sharper decline in error with increasing wavenumber for the medium and short scales. However, the spectra of the deviations of analysis from climatology also show anomalously low amplitudes of the high-wavenumber components for 1981, as will be seen explicitly later. The scaled spectra in fact show the 1981 results to be the poorest at all wavenumbers. The basic spectra show the day-one error of all other years to be at about the same level by wavenumber forty, whereas the scaled spectra indicate more of an improvement in recent years. The poor treatment of the global-mean component in 1981 is again evident.

For day three the error spectra from different years coincide at around wavenumber twenty-eight. The longer scales still exhibit a general reduction in error over the years. For shorter wavelengths there appears to have

been a rather steady increase of error. The scaled spectra show the 1981 results in a very poor light. Of the other years, the 1994 results are clearly superior in the middle spectral range. The scaled errors for 1985 are the lowest for the shorter scales, but it will be seen shortly that this year stands out as one where error saturated at less than the perfect-model asymptotic limit of twice the variance of the analysis, the scaling factor used here. The scaled global-mean error components are not plotted for 1981 and 1985 as they are substantially greater than one by day three.

Fig 17 shows the spectra of day-ten errors and of twice the squared deviations of the analyses from climatology[2], for the same sample of years. The day-ten curves indicate the levels at which error saturates for the shorter scales, as shown already for 1981 and 1994. Progressively higher levels are seen for the later years shown. Much less interannual change is seen in the higher-wavenumber components of the analysis variance, indicating that the analyzed observations rather than the assimilating model are principally determining the spectral distribution for this range of wavenumbers. The first year, 1981, stands out as having rather low levels of variance at high wavenumber, and the 1994 values are slightly above those of intervening years for the highest few wavenumbers. Monitoring each season and year shows that larger short-wave values such as seen for 1994 originated with the introduction of the higher resolution model in 1991. As day-ten levels increased by even more at the same time, it is not clear whether the higher level shown for the 1994 analyses represents a better analysis of shorter scales, or a bias introduced by the assimilating model.

Subject to reservations as to the accuracy of the smaller-scale components of the analyses, the two sets of curves in Fig 17 can be compared in order to identify differences between actual and perfect-model error saturation levels. The underestimation of higher-wavenumber components of variance by the 1985 version of the forecasting model is the largest discrepancy. These components of variance were slightly underestimated by the 1988 model, and captured at about the right level by the 1991 system. The most recent version of the model produces a small-scale variance that is higher than seen in the analyses.

Error growth rates (not shown) have been computed for each wavenumber other than the gravest few by applying the Lorenz model to the growth of individual spectral components of the global-mean forecast differences. The model is evidently inapplicable for the smallest wavenumbers, as may be seen in Fig 15, where for 1994 the $n=0$ component of the differences first decreases and then increases as the forecast range increases. The calculation was carried out for comparison with the results for 1981 discussed by *Dalcher and Kalnay* (1987), who found that the error growth rate increased sharply with reducing scale. For the more recent years we find larger growth rates in the dominant scales centred around wavenumber eight or so, consistent with the larger growth rates found in the calculations based on means for the extratropics of the two hemispheres. We also find lower growth rates for higher wavenumber components, so that overall there is now a more gradual increase in growth rate with increasing wavenumber. However, a question must be raised as to the applicability to the spectral domain of such an error growth model, which does not allow for inter-scale transfer of error as shown to be important by *Boer* (1984).

Limited study has also been made of recent analyses and forecasts of vorticity and divergence at the 850 hPa and 200 hPa levels. Calculations were first made of regional means of streamfunction errors and differences,

[2] High-wavenumber components are dominated by the contribution from the analyses rather than the climatology.

for comparison with the extratropical 500 hPa height data discussed in earlier sections. However, for the purposes of a limited discussion here, we show in Fig 18 only spectra of the global-mean squared forecast errors and differences, for the rotational and divergent wind components at 850 hPa, and for winter 1994. Mostly similar results have been found for the 200 hPa level and for summer 1993.

Comparing the left-hand panels of Fig 18 with the lower left panel of Fig 15, we see a broadly similar evolution of the error spectra. As expected, the rotational wind spectra exhibit shallower slopes for large wavenumbers and steeper slopes for small wavenumbers than found for the height spectra. The divergent wind spectra are more different at shorter forecast ranges, with a peak at wavenumber five. Close inspection of the higher-wavenumber regime for the later forecast ranges suggests that saturation occurs slightly later for the wind fields than for the height field.

The forecast-difference spectra of the wind components show more clearly the slower approach to saturation for these components. For the 500 hPa height field, the day-seven and day-ten curves in the lower-right panel of Fig 15 are almost indistinguishable towards the end of the wavenumber range shown. In contrast, for both wind components, particularly the divergent one, the day-seven and day-ten curves can be seen to be clearly separate across the whole spectral range presented in Fig 18. For higher wavenumbers and later forecast ranges the dependence on wavenumber approximates an $n^{-3.7}$ form for the squared rotational wind and an $n^{-1.9}$ form for the squared divergent wind, for this 850 hPa level. Corresponding dependences are $n^{-2.3}$ and $n^{-1.6}$ at 200 hPa.

A more obvious feature of the difference spectra is seen at low wavenumbers. Early in the forecast range differences decay rather than grow. This is particularly marked for the divergent wind component, for which there is a rapid reduction of the wavenumber-five peak seen in the one-day forecast error. The decay is, however, large enough in the rotational component for the effect to be quite marked in streamfunction spectra, which can be obtained from the rotational wind spectra by division by $n(n+1)$. In physical space the effect is most marked in the tropics. The tropical mean of the squared differences between consecutive forecasts of the 850 hPa streamfunction hardly changes between day one and day four, only later growing, and then at a rate rather slower than that of forecast error. The corresponding differences at 200 hPa decrease to a minimum that occurs between day three and four.

These results point to a major deficiency of the forecasting system in its treatment of the largest scales of motion. There is evidently an inconsistency of structure between the analyses and the forecast model. One possibility is that the forecast model has a substantial error which manifests itself in a rapid adjustment of some realistic structure of the analyzed fields towards an erroneous structure preferred by the model. Alternatively, the planetary-scale analysis itself could have significant errors (*Daley et al*, 1986), and these might be rapidly removed in the early stages of the forecast. A significant part of what we have referred to as one-day forecast error could in fact be analysis error. Either way, and notwithstanding the fact that the problem appears to have been worse still in earlier years, this unsatisfactory feature of the performance of the forecasting system merits further investigation.

9. CONCLUDING DISCUSSION

We have examined the skill of 500 hPa height forecasts made by ECMWF over a period of thirteen and a half years. During this time there has been a considerable increase in the accuracy of forecasts for a few days ahead. Error has, however, grown more rapidly in the medium range in recent years, so that by day ten there has been little net improvement, particularly if the quite substantial reduction in systematic error is discounted. The improved skill of shorter-range forecasts and the generally improved climatology of the forecasts throughout the range, in particular relating to root mean square daily changes, suggest that the larger error growth may be a consequence of a more active forecast model. This model is more realistic in its portrayal of actual changes given accurate initial conditions, but also amplifies error more rapidly. A broadly similar conclusion has recently been reported by *Savijärvi* (1994) in a study of the performance of the medium-range forecasting system of NMC, Washington, for the years 1988 to 1993.

Reduction of the root mean square error of the 500 hPa height field is neither the sole nor the primary objective of medium-range weather prediction. If it was, we could, for example, simply blend the numerical forecast and a forecast of climatology in an optimal way to reduce the error, as illustrated in Fig 19. Smoothing the numerical forecast in this way brings virtually no reduction in error out to about four days ahead. Moreover, only towards the end of the ten-day range does this "improvement" reduce error by as much as can be gained through improvement of the numerical forecasting system, according to the arguments made in discussing Fig 4. Nevertheless, if it is considered desirable to smooth unpredictable scales, then it seems preferable to achieve this through adaptation of the output of the numerical model rather than through an underactive model. A realistic model climatology of short-scale transient-eddy activity is believed to be necessary for the accurate deterministic prediction of larger scales, and is needed for accurate probabilistic forecasts from an ensemble forecasting system, for example indicating likelihoods of extreme events.

We have also briefly looked at other upper-air fields. Forecast errors and differences of the 850 hPa and 200 hPa wind components indicate at least as much scope for benefit from model improvements beyond the one-day range as for the 500 hPa height field. We have not, however, considered direct model predictions of weather elements. These generally exhibit less skill than the predictions of upper-air fields. Error in the forecasting of variables such as temperature at two metres or precipitation may arise either from an inaccurate prediction of the synoptic situation or from an inaccurate parametrization of the relevant local physical processes (or from both). Experience shows that model changes may be found which have only a rather small influence on conventional skill scores for the 500 hPa height, but which nevertheless greatly improve some aspect of weather-element prediction.

Results have been presented for averages over the extratropical hemispheres. We have also carried out calculations for smaller domains such as Europe. Although skill levels at specific points or over small areas are what is often relevant to the user of forecasts, we have chosen not to present such results. This is partly because they are much more subject to interannual fluctuations than hemispheric results, making it more difficult to distinguish consequences of natural fluctuations from effects of changes to the forecasting system. Also, improvement of medium-range forecasts for a particular limited area may follow from improvements to analyses or short-range forecasts some considerable distance upstream. For example, *Hollingsworth et al* (1985) illustrated how differences in initial analyses over the North Pacific can subsequently evolve downstream to influence the forecast over Europe at the range of five to six days. For relatively small domains we cannot confidently apply a simple error-growth model such as that of Lorenz to estimate the reduction in medium-range forecast error that would result from a reduction in short-range error. We can,

nevertheless, fit data to such a model merely as a way of summarizing error growth characteristics. Thus for Europe we find both error growth rates and asymptotic limits from Lorenz' model that are higher than for the extratropical Northern Hemisphere as a whole, with an error doubling time of around 1.3 days. We have not tried to diagnose how much if any of this larger error growth is due to intrinsically faster amplification mechanisms over Europe, and how much is due to propagation of error into the region.

Lorenz (1982) noted that there were both optimistic and pessimistic views that could be taken of the results he had obtained. This remains true for the present study. Now, the optimistic view is that the penalty of developing a more realistic and active model has been paid in terms of limited reduction of error in the later medium range. Future improvements in data assimilation and modelling which lead to lower short-range forecast errors should consequently lead also to lower errors across the later forecast range. In addition, the current forecast model actually appears to be overactive, so there is a gain to be achieved also from correction of this deficiency.

The pessimistic view is that the error growth rate will continue to increase. This could happen, as argued by Lorenz, if future, more realistic, versions of the forecast model amplify small differences in initial conditions even more rapidly than at present. Alternatively, *Palmer* (personal communication) has pointed out that the error growth rate would increase if an overall reduction in analysis error were not to be accompanied by a proportionate reduction in the projection of the error onto those structures that grow fastest over the first few days of the forecast range. A recent discussion of these "singular vector" structures has been given by *Buizza and Palmer* (1994). The perturbations initially are highly baroclinic, with largest amplitude in the lower troposphere, and a predominant horizontal length scale that increases as the perturbation develops. The growth in time of amplitude is initially faster than exponential. A rapid initial "superexponential" growth of small perturbations was reported by *Schubert and Suarez* (1989), and has been found also for simpler models (e.g. *Trevisan*, 1993). *Nicolis et al* (1994) show that the mechanisms involved may nevertheless be complex.

Sensitivity studies with the ECMWF model have shown how individual forecasts which exhibit anomalously rapid error growth can be strikingly improved by perturbing initial conditions with structures similar to those of the singular vectors (Klinker and Rabier, personal communication). The three-dimensional analysis method currently used at ECMWF is poorly suited to elimination of erroneous small-scale baroclinic structures present in the background forecasts from the assimilating model. There is, however, good reason to believe that significant improvements will result from the planned introduction of four-dimensional variational assimilation, with its implicit use of baroclinic, flow-dependent structure functions (*Thépaut et al*, 1993).

Acknowledgements
We have an obvious debt of gratitude to Ed Lorenz for initiating this line of work, and thank Reinhard Strüfing and Jean-Michel Hoyer for establishment and early maintenance of the database. We thank also Anders Persson for suggesting diagnosis of the rate of change of the height field, and Tim Palmer, Tony Hollingsworth and Philippe Courtier for helpful comments.

References

Boer, G J, 1984: A spectral analysis of predictability and error in an operational forecast system. Mon Wea Rev, 112, 1183-1197.

Boer, G J, 1994: Predictability regimes in atmospheric flow. To appear in Mon Wea Rev.

Buizza, R, and T N Palmer, 1994: The singular-vector structure of the atmospheric general circulation. Submitted to J Atmos Sci.

Dalcher, A and E Kalnay, 1987: Error growth and predictability in operational ECMWF forecasts. Tellus, 39, 474-491.

Daley, R, W Wergen and G Cats, 1986: The objective analysis of planetary-scale flow. Mon Wea Rev, 114, 1892-1908.

Eyre, J R, 1987: On systematic errors in satellite sounding products and their climatological mean values. Q J R Meteorol Soc, 113, 279-292.

Eyre, J R, G A Kelly, A P McNally, E Andersson and A Persson, 1993: Assimilation of TOVS radiance information through one-dimensional variational analysis. Q J R Meteorol Soc, 119, 1427-1463.

Hollingsworth, A, A C Lorenc, M S Tracton, K Arpe, G Cats, S Uppala and P Kållberg, 1985: The response of numerical weather prediction systems to FGGE level IIb data. Part I: Analyses. Q J R Meteorol Soc, 111, 1-66.

Jarraud, M, A J Simmons and M Kanamitsu, 1988: Sensitivity of medium-range weather forecasts to the use of an envelope orography. Q J R Meteorol Soc, 114, 989-1025.

Lorenc, A C, 1981: A global three-dimensional multivariate statistical interpolation scheme. Mon Wea Rev, 109, 701-721.

Lorenz, E N, 1972: Barotropic instability of Rossby wave motion. J Atmos Sci, 29, 258-264.

Lorenz, E N, 1982: Atmospheric predictability experiments with a large numerical model. Tellus, 34, 505-513.

Lorenz, E N, 1990: Effects of analysis and model errors on routine weather forecasts. Proceedings of 1989 ECMWF Seminar on Ten Years of Medium-range Weather Forecasting, I, 115-128.

Lönnberg, P, 1988: Developments in the ECMWF analysis system. Proceedings of 1988 ECMWF Seminar on Data Assimilation and the Use of Satellite Data, I, 75-119.

Miller, M J, 1988: The sensitivity of systematic errors of the ECMWF forecast model to parametrized processes. Workshop on systematic errors in models of the atmosphere, Report 12, CAS/JSC WGNE, WMO/TD No. 273, 289-296.

Miller, M J, T N Palmer and R Swinbank, 1989: Parametrization and influence of subgrid-scale orography in general circulation an numerical weather prediction models. Meteor Atmos Phys, 40, 84-109.

Morcrette, J-J, 1990: Impact of changes to the radiation transfer parametrizations plus cloud optical properties in the ECMWF model. Mon Wea Rev, 118, 847-873.

Mureau, R, 1990: The decrease of the systematic error and the increased predictability of the long waves in the ECMWF model. ECMWF Tech Memo, 167, 37pp.

Nicolis, C, S Vannitsem and J-F Royer, 1994: Short range predictability of the atmosphere: mechanisms for superexponential error growth. Submitted to Q J R Meteorol Soc.

Palmer, T N, 1988: Medium and extended range predictability and stability of the Pacific-North American mode. Q J R Meteorol Soc, 114, 691-713.

Ritchie, H C Temperton, A J Simmons, M Hortal, T Davies, D Dent and M Hamrud, 1994: Implementation of the semi-Lagrangian method in a high resolution version of the ECMWF forecast model. To appear in Mon Wea Rev.

Savijärvi, H, 1984: Spectral properties of analyzed and forecast global 500 mb fields. J Atmos Sci, 41, 1745-1754.

Savijärvi, H, 1994: Error growth in a large numerical forecast system. Submitted for publication.

Schubert, S D and M Suarez 1989: Dynamical predictability in a simple general circulation model: average error growth. J Atmos Sci, 46, 353-370.

Shaw, D B, P Lönnberg, A Hollingsworth and P Undén, 1987: The ECMWF mass and wind analysis. Q J R Meteorol Soc, 113, 533-566.

Simmons, A J, 1986: Numerical prediction: Some results from operational forecasting at ECMWF. Advances in Geophysics, 29, 305-338.

Simmons, A J, D M Burridge, M Jarraud, C Girard and W Wergen, 1989: The ECMWF medium-range prediction models: Development of the numerical formulations and the impact of increased resolution. Meteor Atmos Phys, 40, 28-60.

Simmons, A J and Chen Jiabin, 1991: The calculation of geopotential and the pressure gradient in the ECMWF atmospheric model: Influence on the simulation of the polar atmosphere and on temperature analyses. Q J R Meteorol Soc, 117, 29-58.

Simmons, A J and B J Hoskins, 1980: Barotropic influences on the growth and decay of non-linear baroclinic waves. J Atmos Sci, 37, 1679-1684.

Stroe, R and J F Royer, 1993: Comparison of different error growth formulas and predictability estimation in numerical extended-range forecasts. Ann Geophysicae, 11, 296-316.

Thépaut, J-N, R N Hoffman and P Courtier, 1993: Interactions of dynamics and observations in a four-dimensional variational assimilation. Mon Wea Rev, 121, 3393-3414.

Tibaldi, S and F Molteni, 1990: On the operational predictability of blocking. Tellus, 42A, 343-365.

Tibaldi, S, P Ruti, E Tosi and M Maruca, 1994: Operational predictability of winter blocking at ECMWF: An update. To appear in Ann Geophysicae.

Tiedtke, M, 1989: A comprehensive mass flux scheme for cumulus parametrization in large-scale models. Mon Wea Rev, **117**, 1779-1800.

Tiedtke, M, W A Heckley and J Slingo, 1988: Tropical forecasting at ECMWF: The influence of physical parametrization on the mean structure of forecasts and analyses. Q J R Meteorol Soc, 114, 639-664.

Trevisan, A, 1993: Impact of transient error growth on global average predictability measures. J Atmos Sci, 50, 1016-1028.

Tuck, A F, R T Watson, E P Condon, J J Margitan and O B Toon, 1989: The planning and execution of ER-2 and DC-8 aircraft flights over Antarctica, August and September 1987. J Geophys Res, 94, 11181-11222.

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

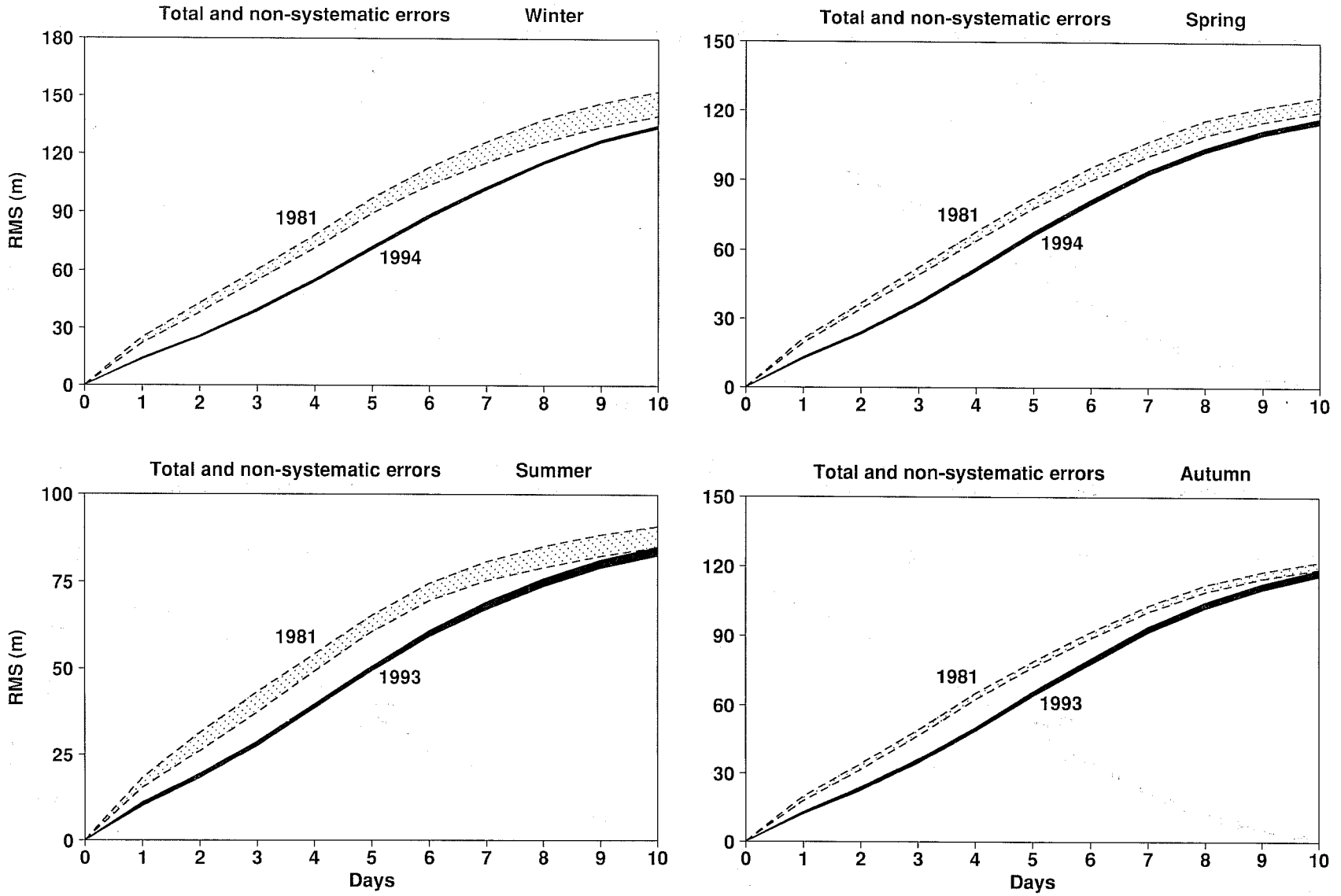


Fig. 1 Root mean square errors (m) of 500hPa height forecasts for the extratropical Northern Hemisphere, shown for each of the four seasons as functions of the forecast range (days). Results are shown for 1981 and 1993/94, and for total errors and the non-systematic components of these errors, as indicated in the text.

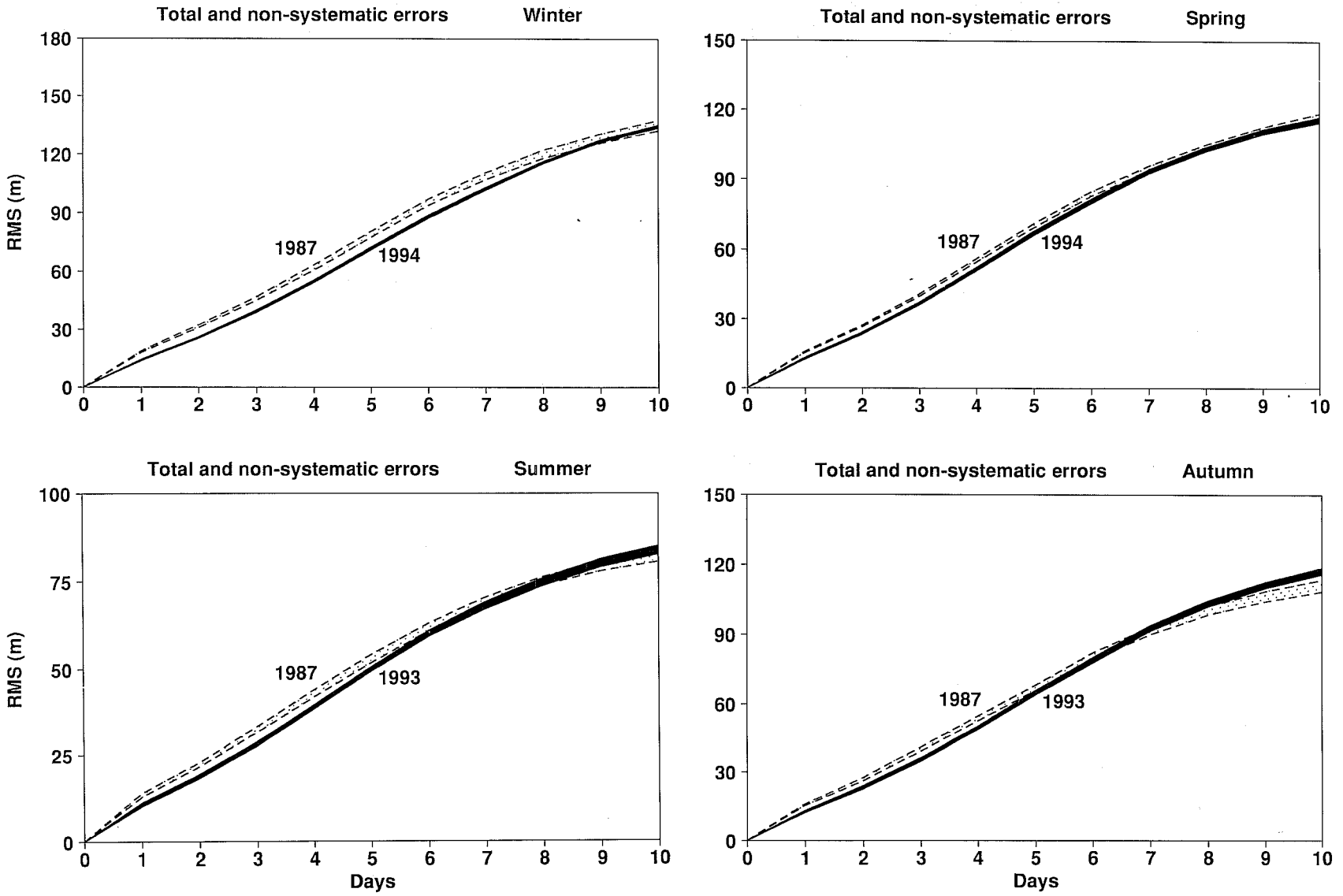


Fig. 2 Root mean square errors (m) of 500hPa height forecasts for the extratropical Northern Hemisphere, as in Fig. 1, but comparing results for 1987 with those for 1993/94.

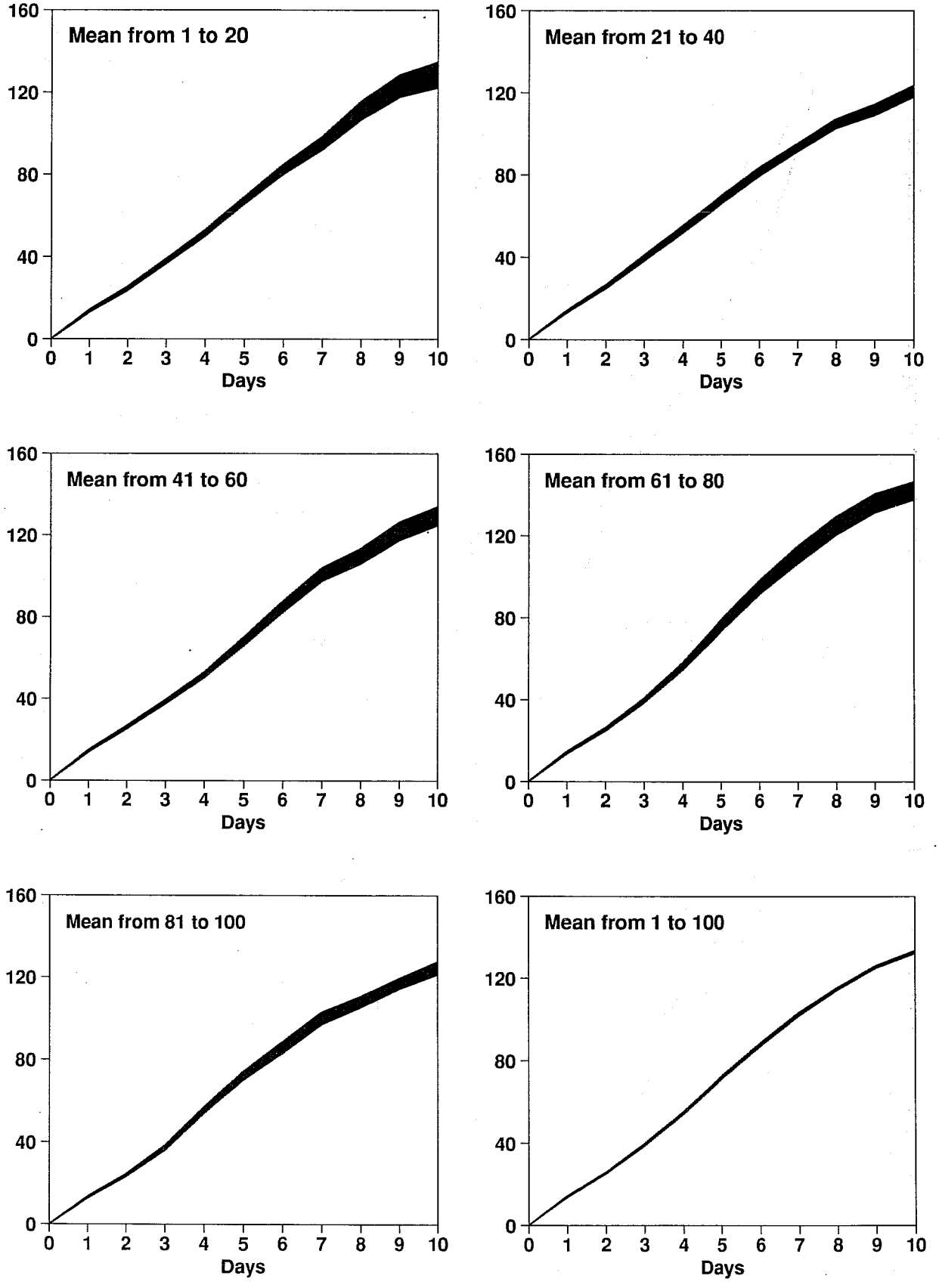


Fig. 3 Total and non-systematic Northern Hemispheric root mean square forecast errors (m) of 500hPa height computed for five consecutive 20-day periods starting 1 December 1993, and for the corresponding 100-day period.

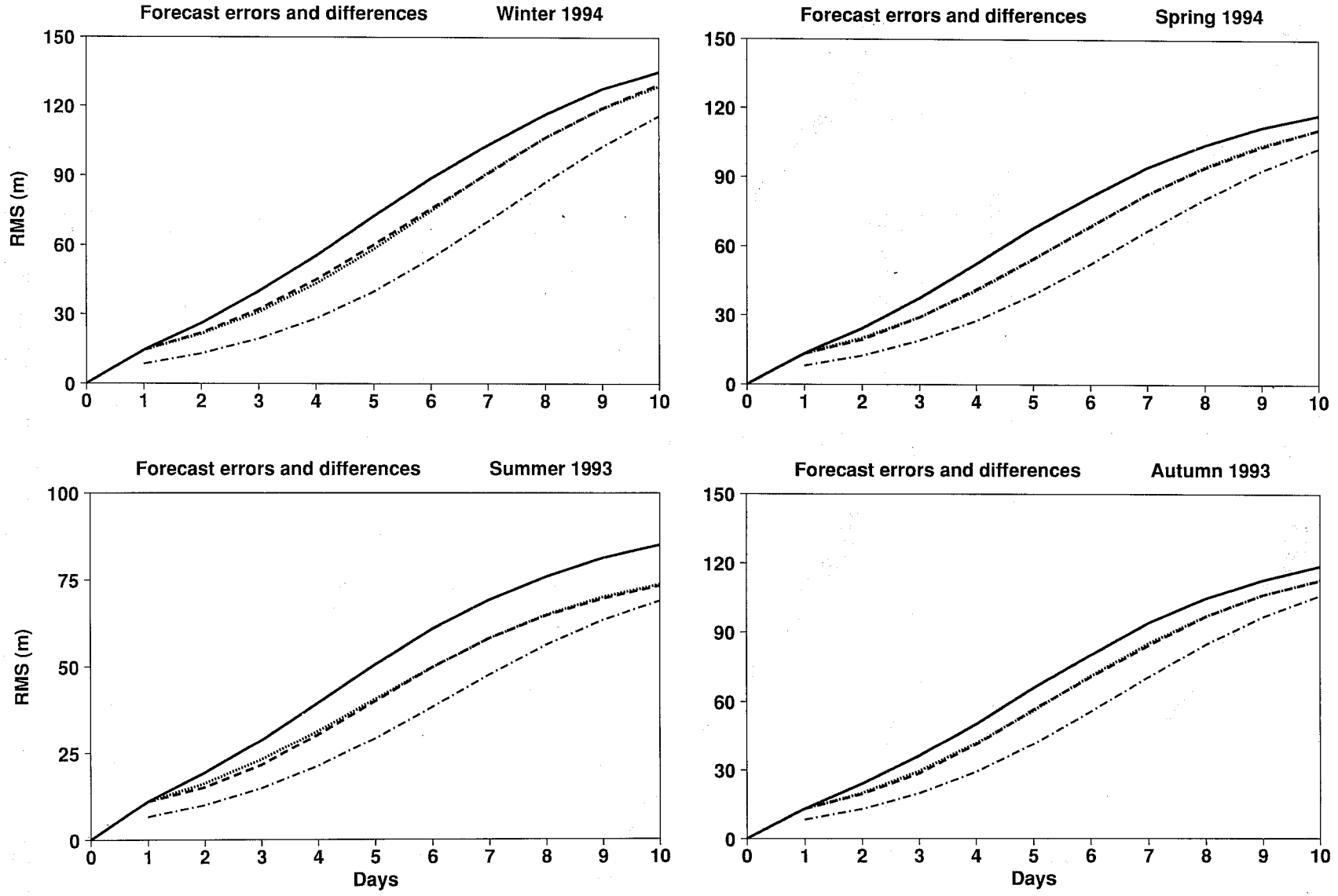


Fig. 4 Actual (solid curves) and potential root mean square forecast errors (m) of 500hPa height for the latest four seasons over the extratropical Northern Hemisphere. See text for details.

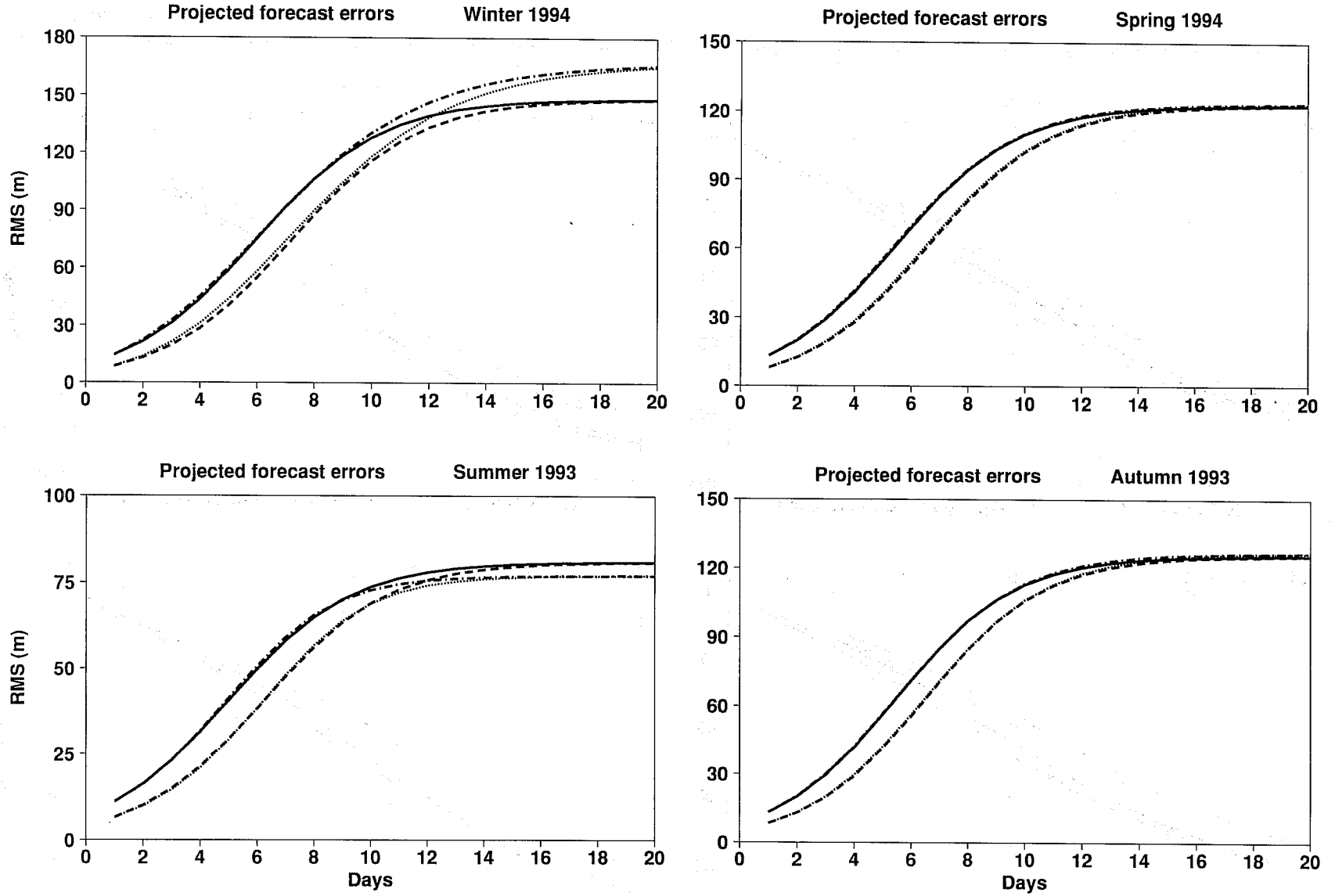


Fig. 5 Projected root mean square forecast errors (m) of 500hPa height out to twenty days for the latest four seasons over the extratropical Northern Hemisphere. Solid and dash-dotted curves start from current one-day forecast errors, and other curves from reduced one-day errors as in Fig. 4. Solid and dashed curves are from Lorenz' model and other curves are from the model of Stroe and Royer.

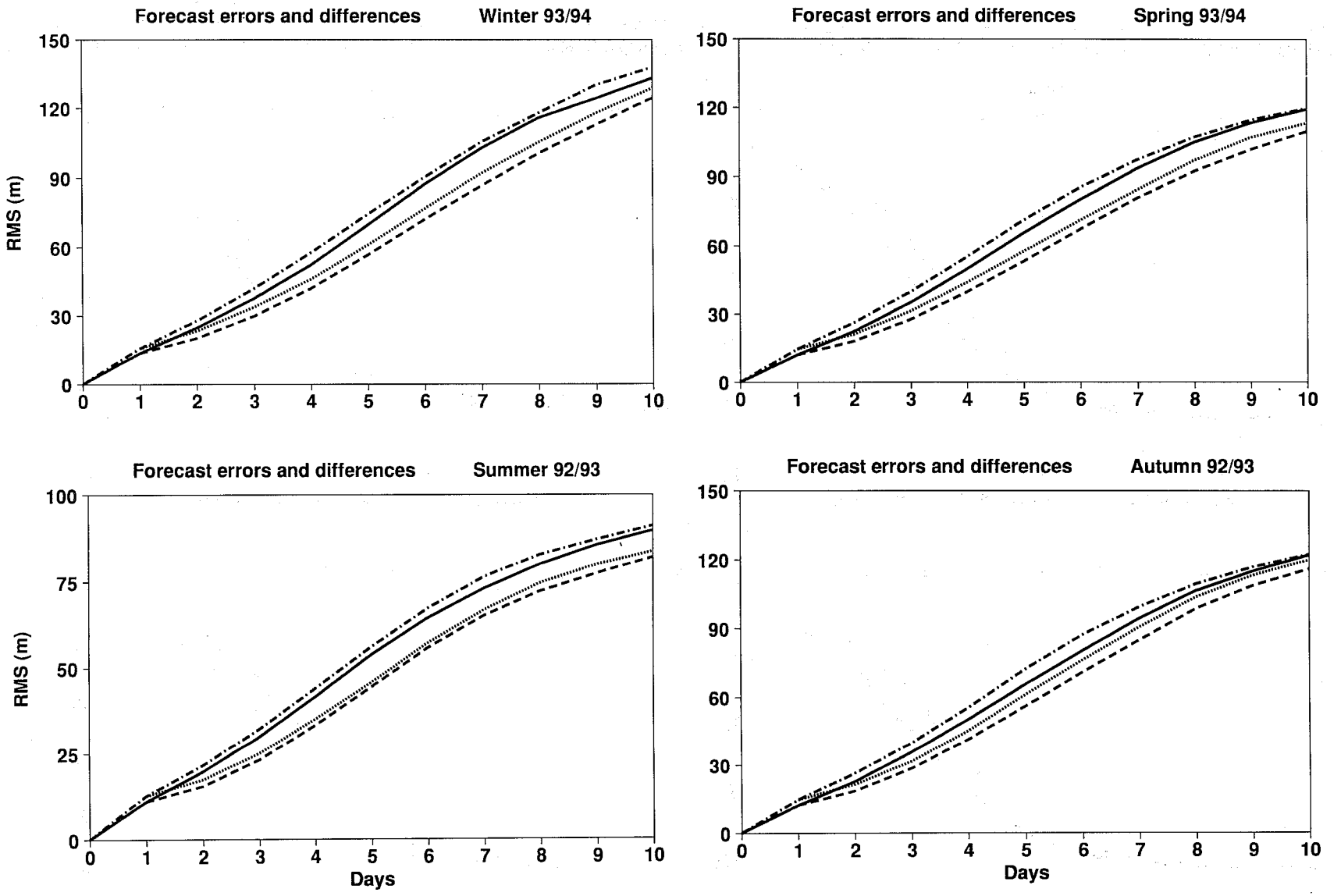


Fig. 6 Root mean square forecast errors and differences between consecutive forecasts (m) of 500hPa height over the extratropical Northern Hemisphere, stratified on the basis of above or below average one-day forecast errors.

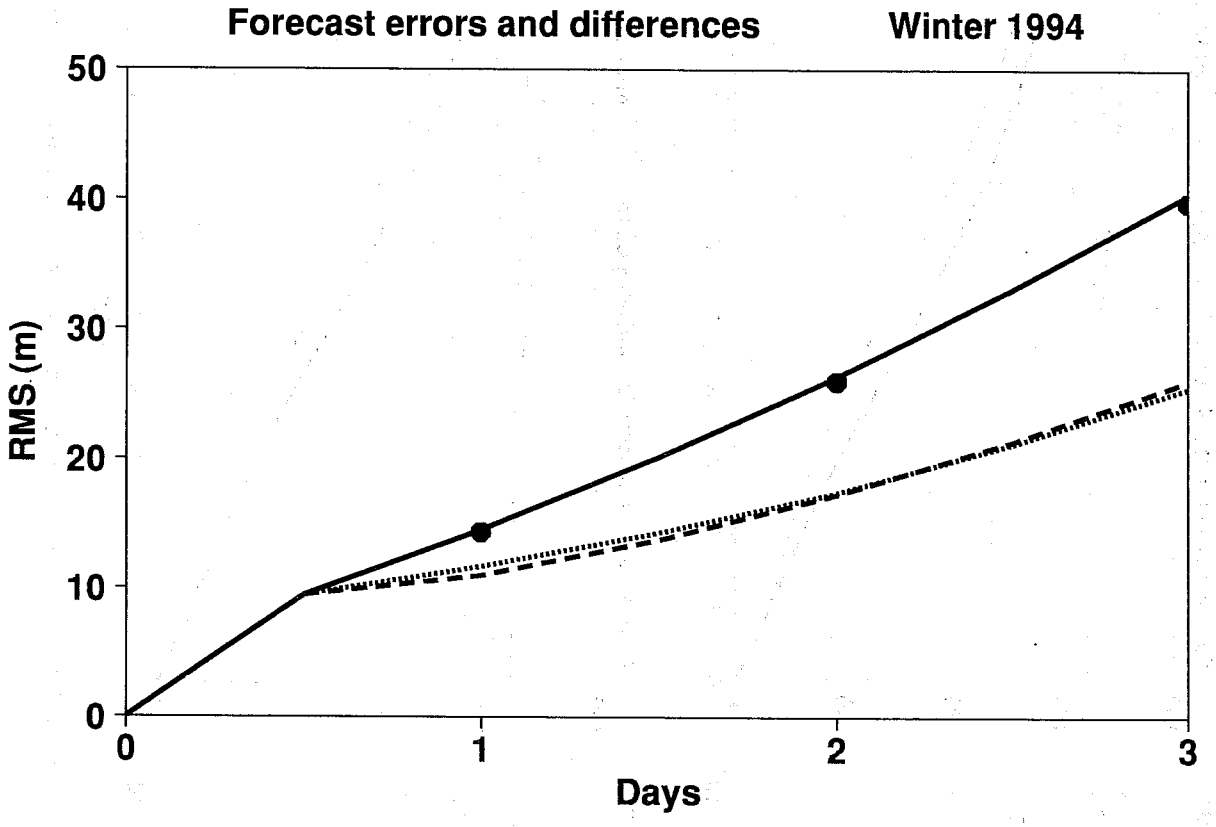


Fig. 7 Root mean square forecast errors (solid) and differences between consecutive forecasts (dashed) of 500hPa height over the extratropical Northern Hemisphere, based on forecasts carried out 12-hourly to day three. Circles denote the errors of the subset of forecasts from 12UTC. The dotted curve is the evolution of 12-hourly forecast differences predicted by the Lorenz model on the basis of the daily results, given the actual 12-hour forecast error.

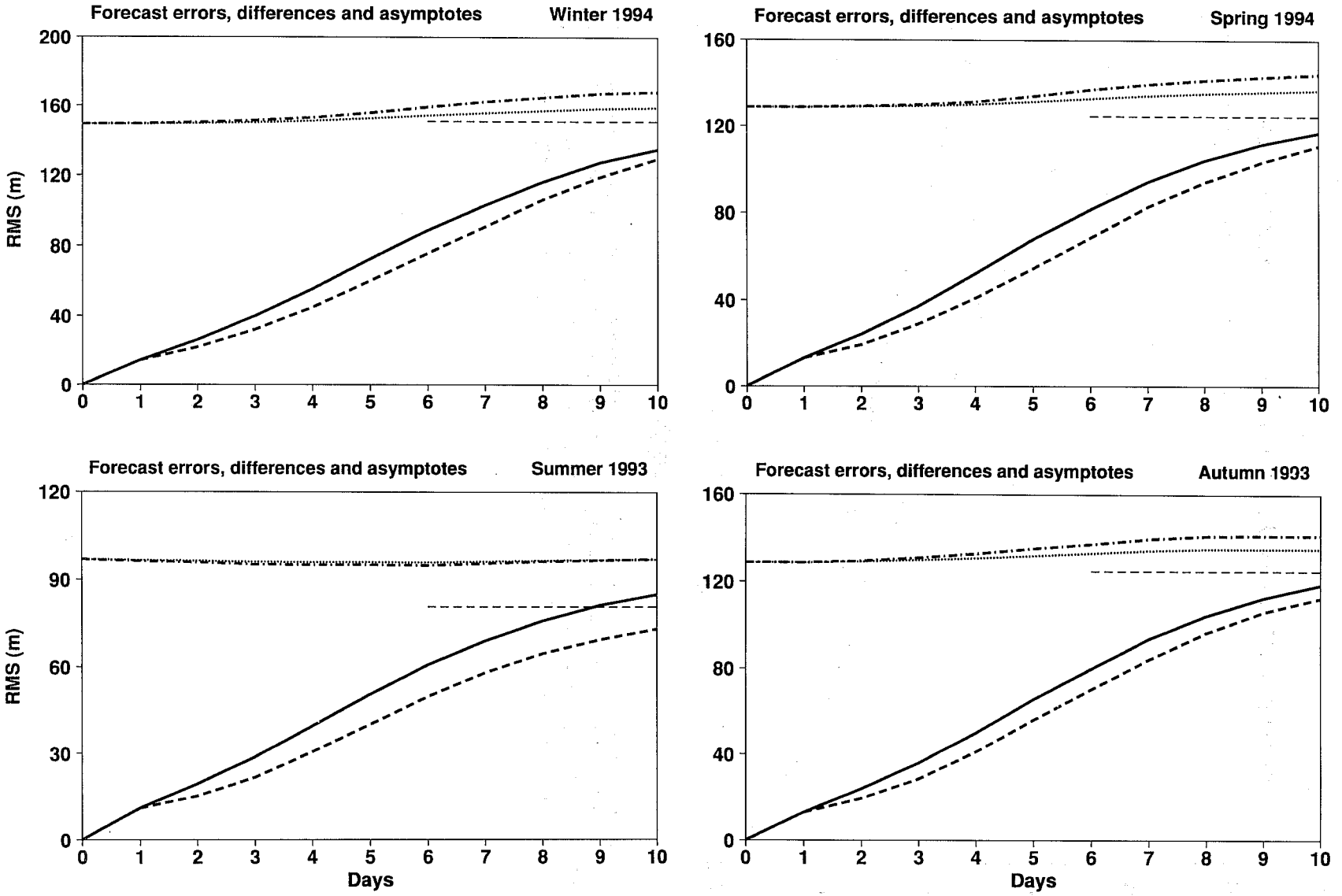


Fig. 8 Root mean square forecast errors and differences between consecutive forecasts (m) of 500hPa height over the extratropical Northern Hemisphere, and estimates of asymptotic limits.

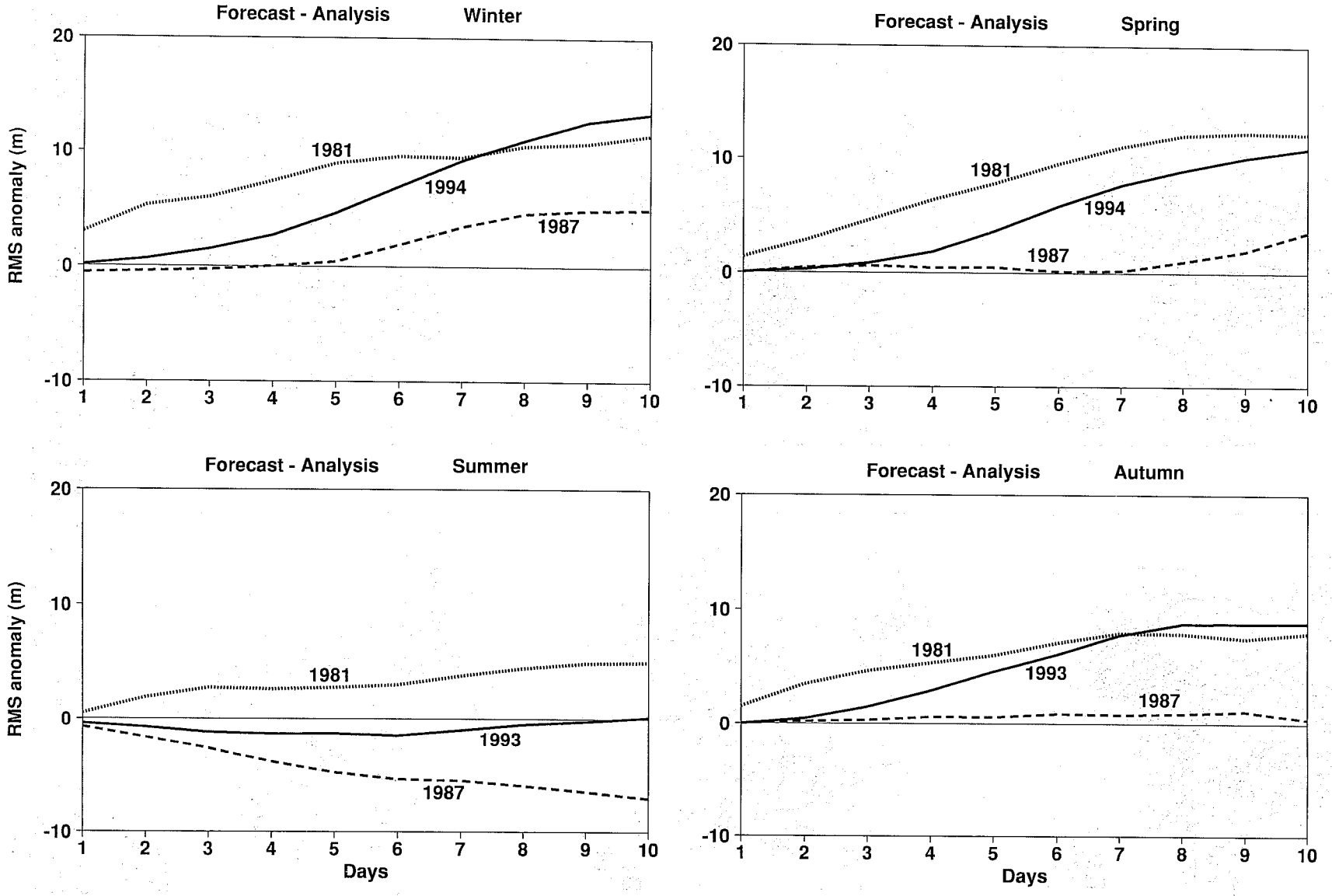


Fig. 9 Differences between forecast and analyzed values of the root mean square deviations (m) of the 500hPa height field from climatology, computed for each season of 1981, 1987 and 1993/94, over the extratropical Northern Hemisphere.

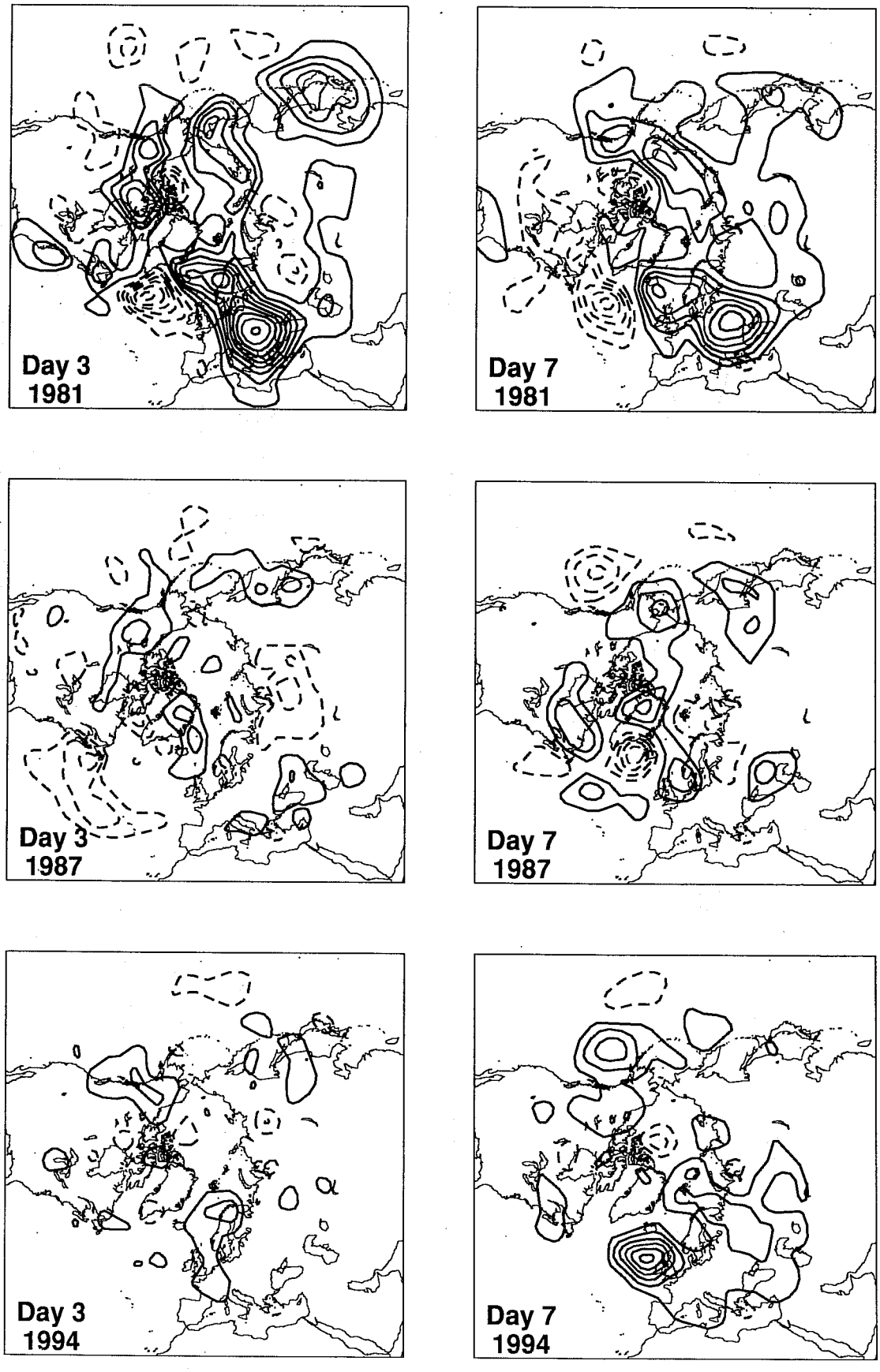


Fig. 10 Maps of the differences between forecast and analyzed values of the mean square deviations of the 500hPa height field from climatology, for day-three (contour interval 2000m^2) and day-seven forecasts (contour interval 4000m^2) for the winters of 1981, 1987 and 1994. Solid contours show where variance is overestimated, and dashed contours where it is underestimated.

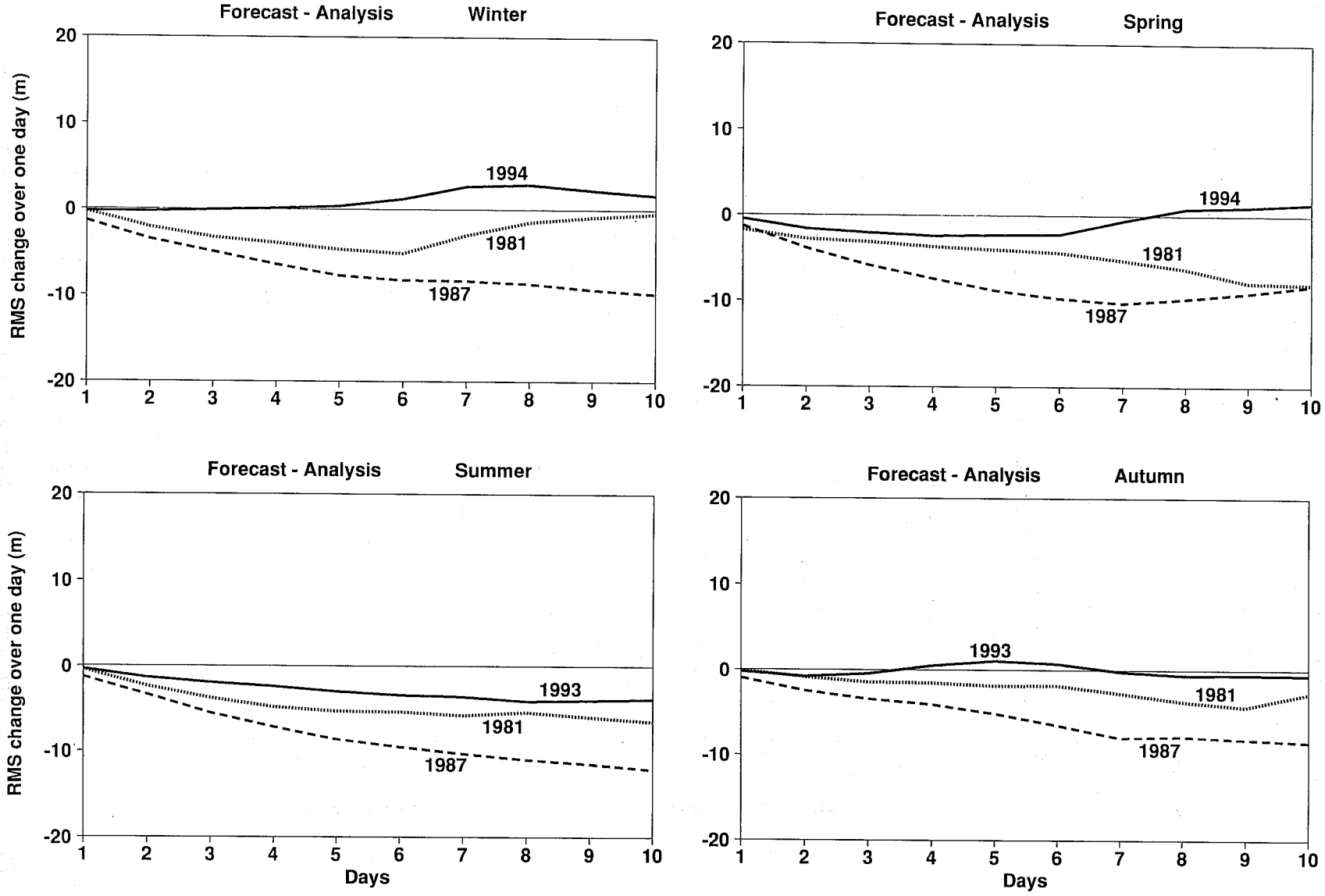
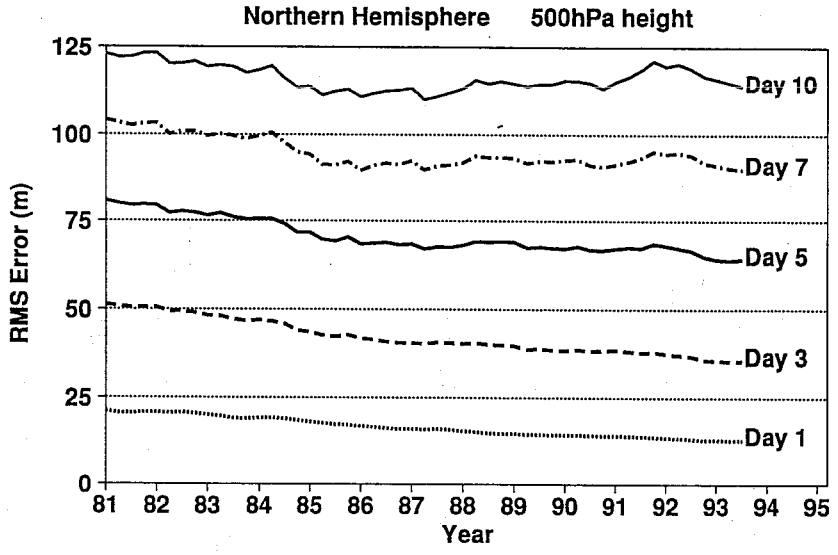
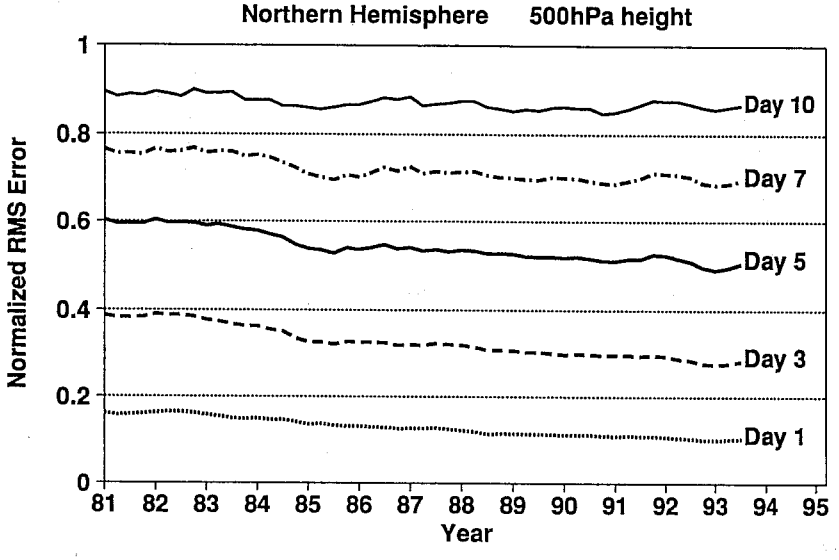


Fig. 11 Differences between forecast and analyzed values of the root mean square changes (m) of the 500hPa height field from one day to the next, computed for each season of 1981, 1987 and 1993/94, over the extratropical Northern Hemisphere.





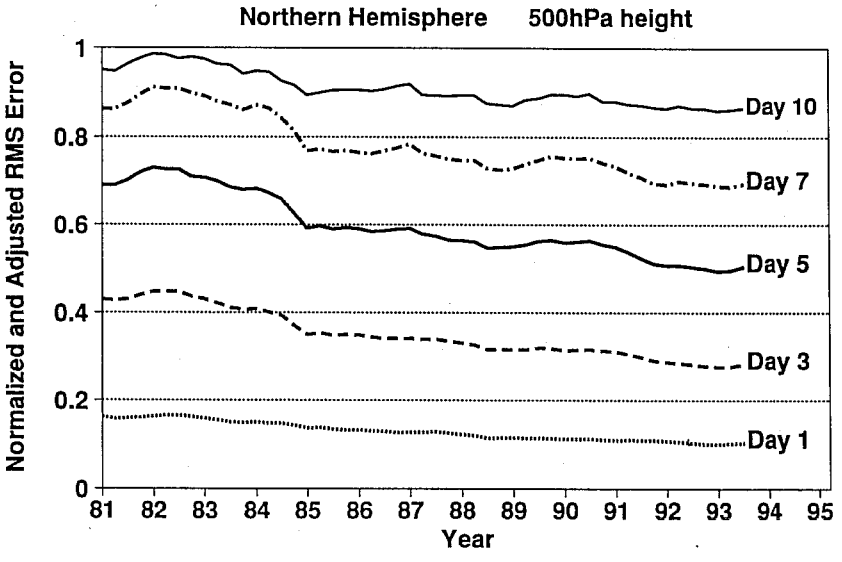


Fig.12 Evolution over the years since 1981 of measures of skill of forecasts of the 500hPa height over the extratropical Northern Hemisphere, at five forecast ranges from one to ten days. See text for details.

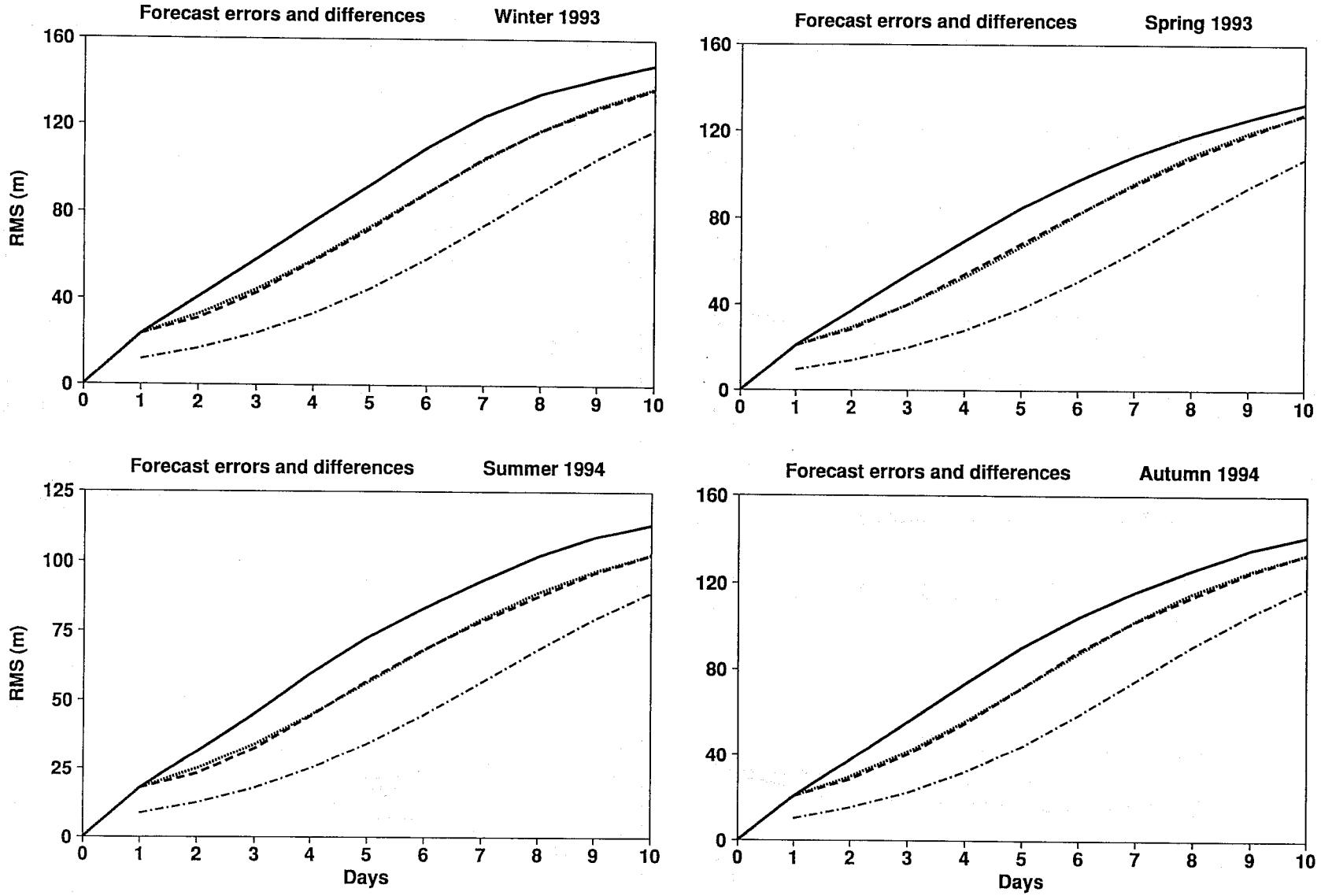


Fig. 13 Actual (solid curves) and potential root mean square forecast errors (m) of 500hPa height for the latest four seasons, as in Fig. 4, but for the extratropical Southern Hemisphere.

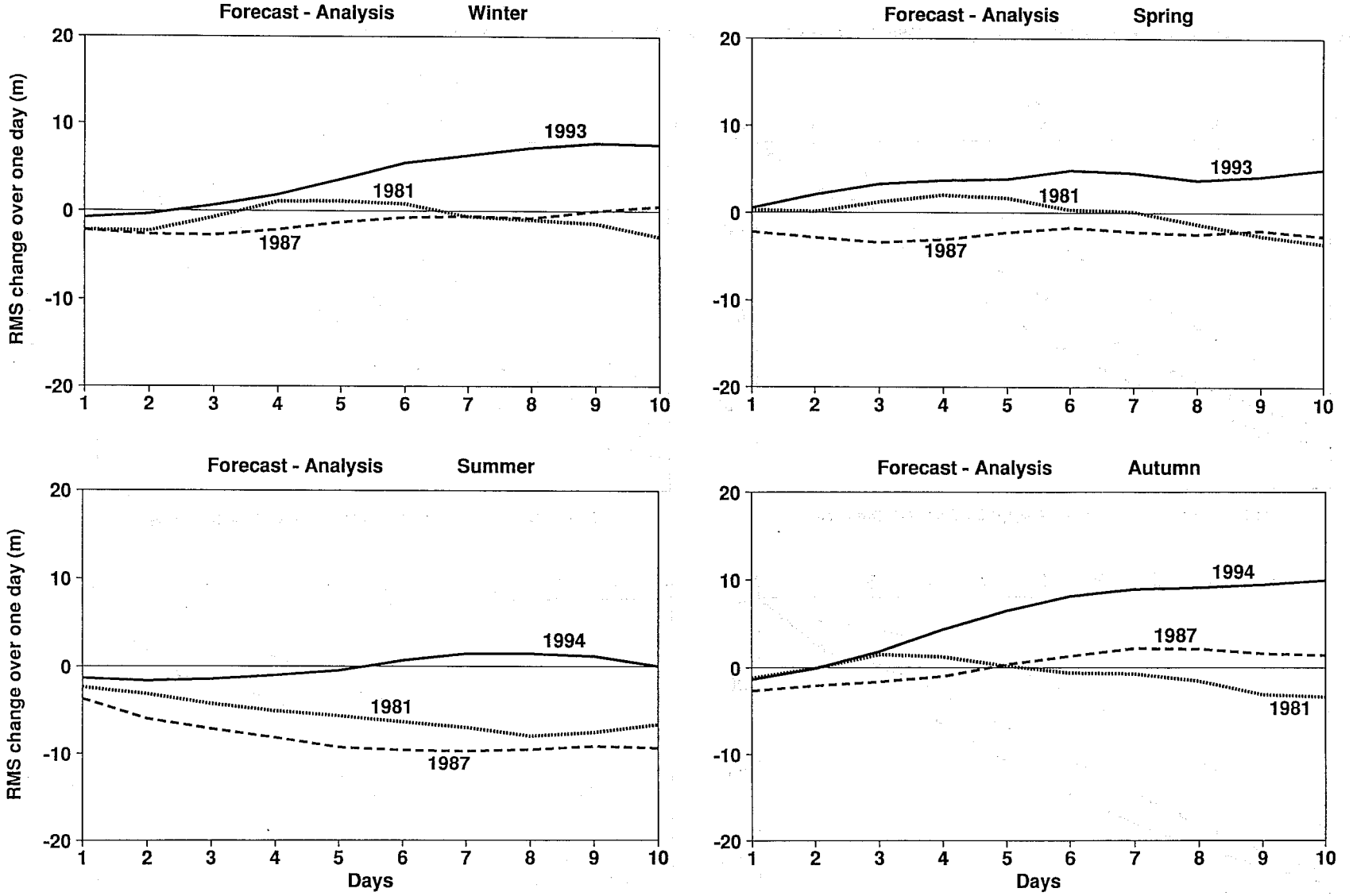


Fig. 14 Differences between forecast and analyzed values of the root mean square changes (m) of the 500hPa height field from one day to the next, computed for each season of 1981, 1987 and 1993/94, as in Fig. 9, but for the extratropical Southern Hemisphere.

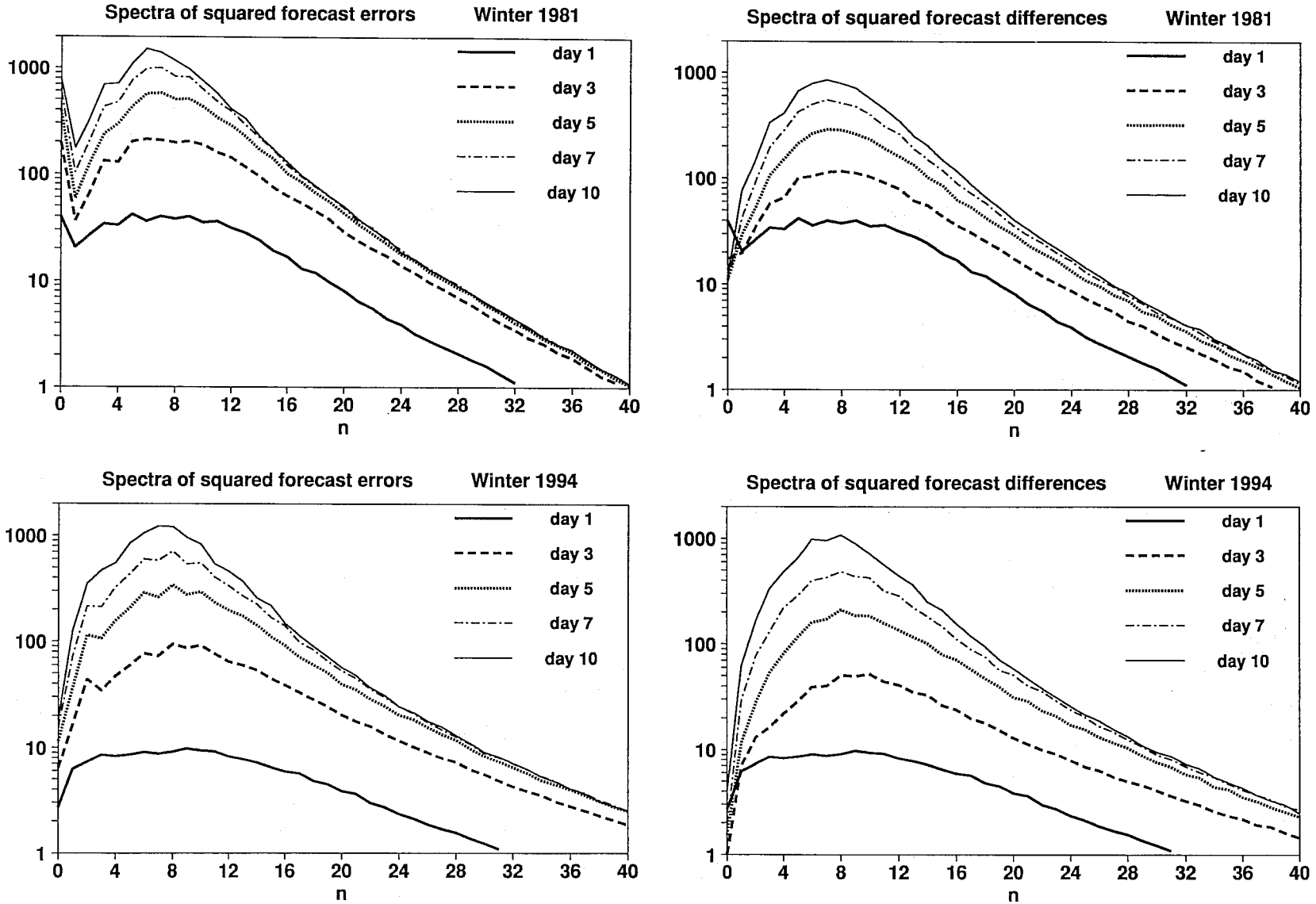


Fig. 15 Total-wavenumber spectra of global-mean squared forecast errors and differences between consecutive forecasts, for the 500hPa height field (m^2) at five forecast ranges from one to ten days, for the (Northern Hemisphere) winter periods of 1981 and 1994.

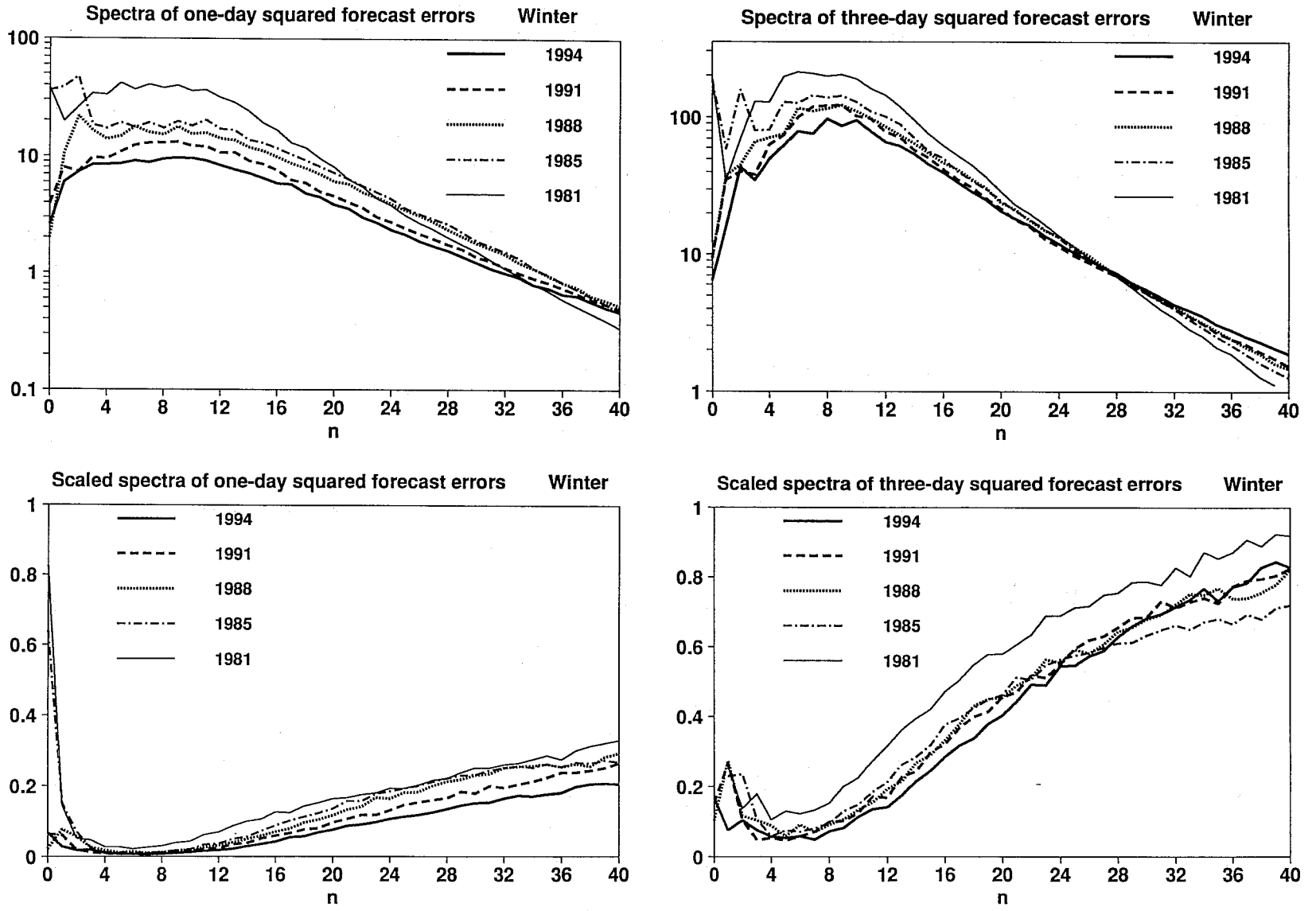


Fig. 16 Total-wavenumber spectra of global-mean squared errors of one- and three-day forecasts (m^2), and spectra normalized as described in the text, for the 500hPa height field for five winter periods selected between 1981 and 1994.

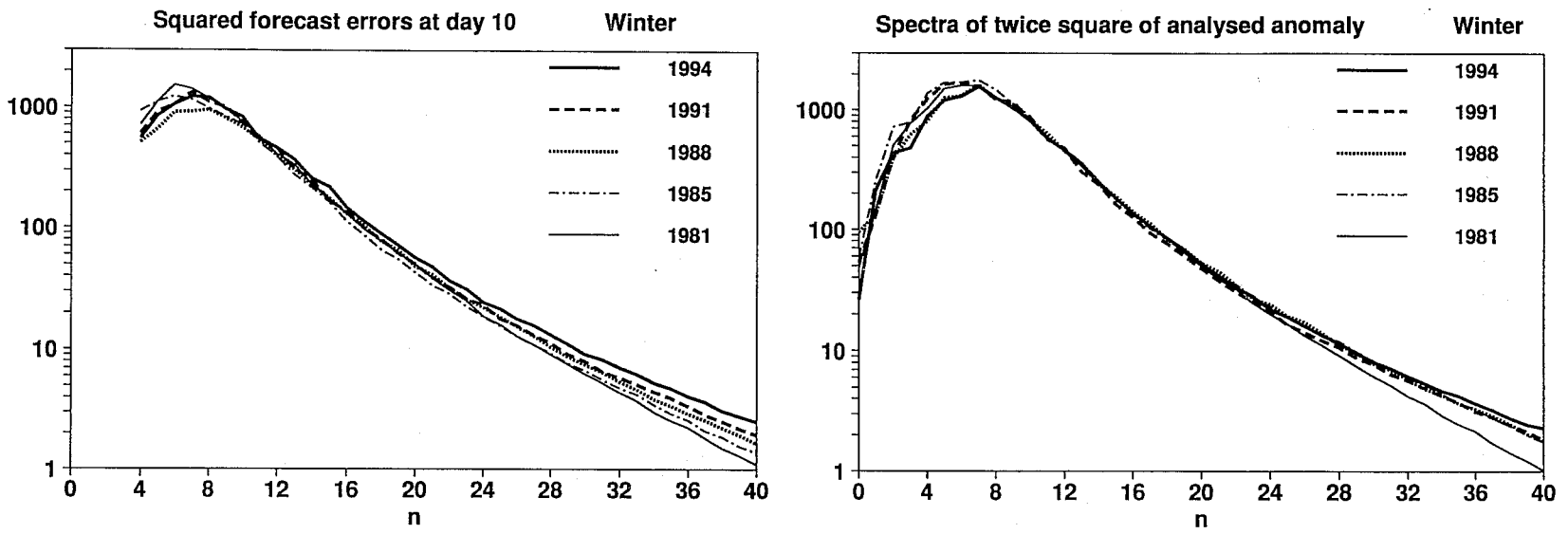


Fig. 17 Total-wavenumber spectra of global-mean squared errors of day-ten forecasts of the 500hPa height field (m^2), for the five selected winter periods between 1981 and 1994, and corresponding spectra of twice the squared deviation of the analyses from climatology.

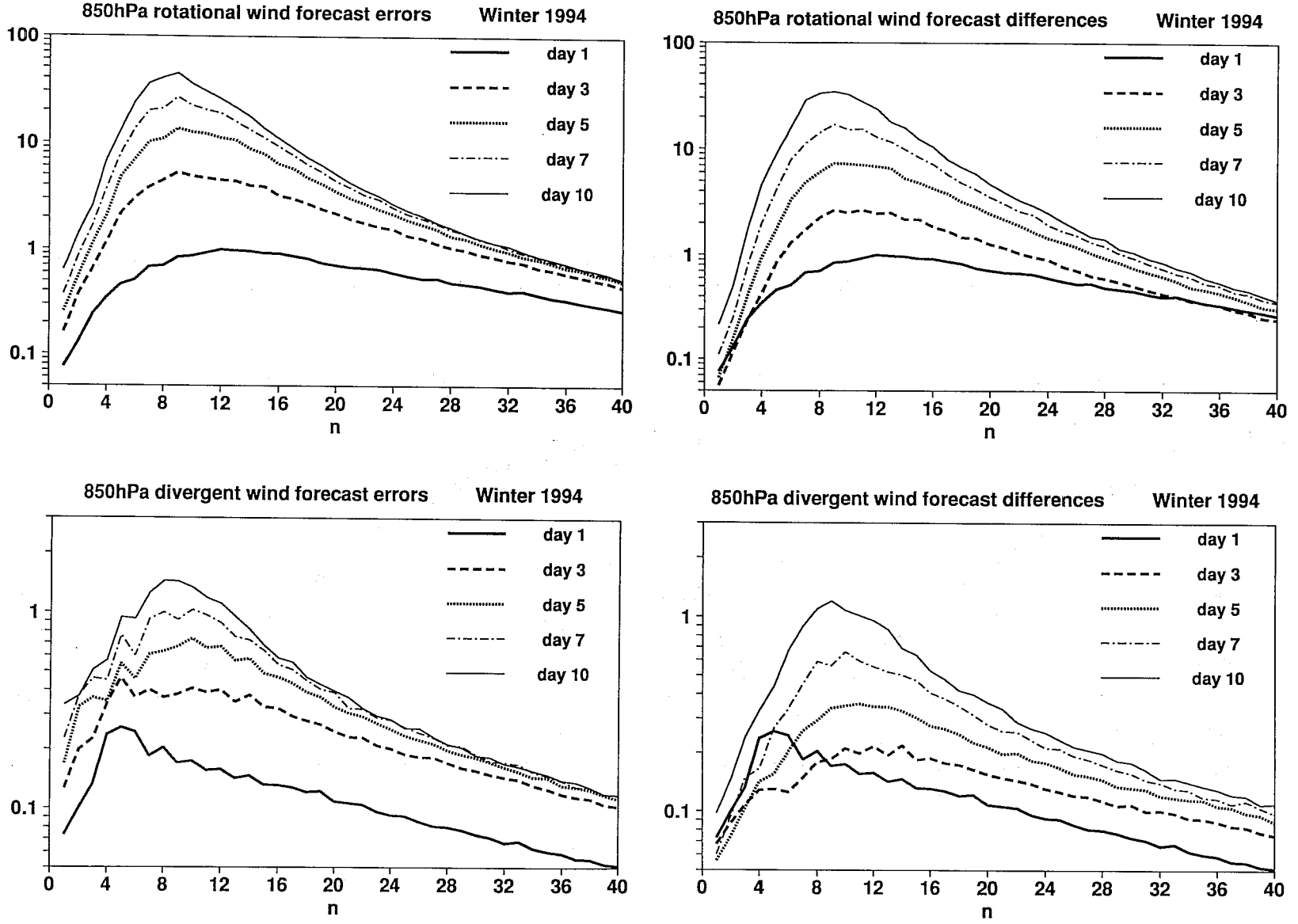


Fig. 18 Total-wavenumber spectra of global-mean squared forecast errors and differences, for the 850hPa rotational and divergent wind components (m^2s^{-2}) at five forecast ranges from one to ten days, for winter 1994.

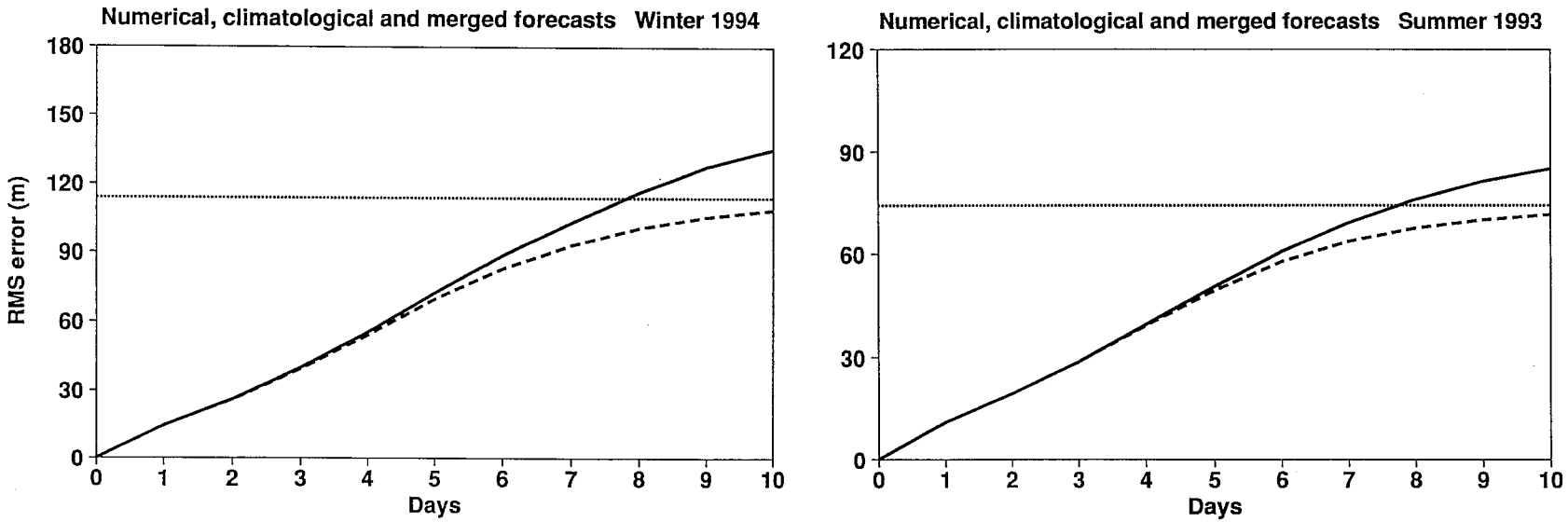


Fig. 19 Root mean square errors (m) of 500hPa height forecasts for the extratropical Northern Hemisphere, for summer 1993 and winter 1994. The solid curves are the errors of the operational ECMWF forecasts and the dotted curves the errors of climatological forecasts. The dashed curves show the results for an optimal blend of numerical and climatological forecasts, determined as in the Annex, with weights calculated using statistics from the forecasts carried out during the corresponding season of the preceding year.

Annex: Measures of forecast skill and asymptotic limits

Let f_j denote the forecast value of a particular field produced j days prior to the verifying time,
a denote the corresponding value of the verifying analysis, and
c denote the corresponding climatological value.

(i) Root mean square error

The root mean square error of the day-j forecast, E_j, is defined by

$$E_j = \sqrt{\overline{(f_j - a)^2}}$$

where the overbar denotes an average over area and over all forecasts carried out in a particular period.

We can write

$$\begin{aligned}(E_j)^2 &= \overline{((f_j - c) - (a - c))^2} \\ &= \overline{(f_j - c)^2} + \overline{(a - c)^2} - 2\,\overline{(f_j - c)(a - c)} \\ &= (A_j)^2 + (A_a)^2 - 2\,\overline{(f_j - c)(a - c)} \end{aligned} \tag{A1}$$

where

A_j denotes the root mean square anomaly of the day-j forecast, and
A_a denotes the root mean square anomaly of the verifying analysis.

As the forecast range and averaging period increase, the forecast and analyzed anomalies will tend to become uncorrelated:

$$\overline{(f_j - c)(a - c)} \to 0$$

This assumes there to be no systematic correlation such as would occur if forecasts and analyses shared a common error, or if the estimate of the climatological field itself were in error. Such a correlation could also occur for a limited (e.g. seasonal) averaging period due to the presence in data assimilation and forecasts of a common anomaly, due for example to anomalous sea-surface temperatures or soil moisture. Such effects will not be completely absent in practice, but assuming them to be negligible,

$$E_j \to \sqrt{(A_j)^2 + (A_a)^2}$$

for large j and averaging period. It is important here that the climatology c is defined on a daily basis, or on the basis of averaging for just a few days. If c is a seasonal mean climatology, or the sample average, there will be a correlation between $(f_j - c)$ and $(a - c)$ due simply to the annual cycle. This will be particularly marked for spring and autumn.

The smaller the asymptotic value of A_j, the smaller will be the asymptotic value of root mean square forecast error, for a given level of analysis variance. The value is a minimum if the forecast model ultimately loses all variance about the climatological mean, in which case

$$E_j \to A_a$$

whereas for a perfect model

$$A_j \to A_a$$

and

$$E_j \to \sqrt{2}\, A_a$$

(ii) Root mean square forecast differences

The root mean square difference between consecutive (day-j and day-$(j-1)$) forecasts, D_j, is given by

$$D_j = \sqrt{\overline{(f_j - f_{j-1})^2}}$$

This can be written:

$$\begin{aligned}(D_j)^2 &= \overline{((f_j - c) - (f_{j-1} - c))^2} \\ &= \overline{(f_j - c)^2} + \overline{(f_{j-1} - c)^2} - 2\,\overline{(f_j - c)(f_{j-1} - c)} \\ &= (A_j)^2 + (A_{j-1})^2 - 2\,\overline{(f_j - c)(f_{j-1} - c)}\end{aligned}$$

The consecutive forecast anomalies will tend to become uncorrelated as the forecast range and averaging period increase:

$$\overline{(f_j - c)(f_{j-1} - c)} \to 0$$

This assumes no systematic error in the forecast fields, no error in the estimation of climatology, and again no common forecast anomaly such as may occur for limited averaging period due to anomalous surface conditions. In this case,

$$D_j \to \sqrt{2}\, A_j$$

for large j and ensemble size.

The smaller the asymptotic value of A_j, the smaller will be the asymptotic value of root mean square difference between consecutive forecasts. For a perfect model

$$A_j \to A_a$$

and D_j tends to the same limit ($\sqrt{2}\, A_a$) as E_j

(iii) Normalized mean square error and anomaly correlation

The normalized mean square forecast error

$$\frac{(E_j)^2}{(A_a)^2 + (A_j)^2}$$

has an initial value of zero (provided the initial and verifying analyses are the same), and under the assumptions discussed above concerning the long-term correlation of forecast and analyzed anomalies, tends to one as the forecast range and averaging period increase. If the anomaly correlation coefficient is approximated by AC_j, defined by

$$\begin{aligned}AC_j &= \frac{\overline{(f_j - c)(a - c)}}{\sqrt{\overline{(f_j - c)^2}\;\overline{(a - c)^2}}} \\ &= \frac{\overline{(f_j - c)(a - c)}}{A_j\, A_a}\end{aligned}$$

equation (A1) can be used to show that the normalized mean square error is equal to

$$1 - \frac{2\, A_j\, A_a}{(A_j)^2 + (A_a)^2}\, AC_j$$

The factor scaling the approximate anomaly correlation in this expression has a maximum value of one when

$$A_j = A_a$$

or, in words, when the forecast model exhibits the same level of deviation about climatology as the atmosphere itself.

We can also write

$$AC_j = \frac{1}{2}\left(\frac{A_a}{A_j}\right)\left(1 + \left(\frac{A_j}{A_a}\right)^2 - \left(\frac{E_j}{A_a}\right)^2\right)$$

from which it can be shown that AC_j increases as A_j increases, for A_j sufficiently close to A_a. The approximate anomaly correlation will thus increase if a model change increases the variance of the forecast, but leaves the root mean square error unchanged.

The formal definition of anomaly correlation coefficient is

$$\frac{\overline{(f_j-c)\,(a-c)} - \overline{(f_j-c)}\;\overline{(a-c)}}{\sqrt{\left[\,\overline{(f_j-c)^2}-\overline{(f_j-c)}^{\,2}\,\right]\left[\,\overline{(a-c)^2}-\overline{(a-c)}^{\,2}\,\right]}}$$

In practice, when the coefficient is calculated over an area such as the extratropical Northern Hemisphere, and the averaging period covers a season, the mean deviation of analysis and forecast from climatology, $\overline{(a-c)}$ and $\overline{(f_j-c)}$, are small. It is shown in section 3 that AC_j is a good approximation to the true anomaly correlation coefficient.

(iv) Smoothing the forecast to reduce root mean square error

We may readily determine, for each day of the forecast range, the weighting factor, ϵ_j, which minimizes the root mean square error of a forecast comprising a blend of the numerical forecast, $\epsilon_j f_j$, and climatology , $(1-\epsilon_j)\,c$. We seek to minimize

$$\overline{(\epsilon_j f_j + (1-\epsilon_j)\,c - a)^2}$$

which is achieved by

$$\epsilon_j = \frac{\overline{(f_j-c)(a-c)}}{\overline{(f_j-c)^2}}$$

ϵ_j is equal to the approximate anomaly correlation AC_j scaled by the ratio A_a / A_j

Winter

Day	1981			1987			1994		
0	100.0	100.0	4.2	100.0	100.0	-2.1	100.0	100.0	3.2
1	97.7	97.6	-2.2	98.6	98.5	0.5	99.1	99.1	2.3
2	93.4	93.1	-5.8	95.8	95.6	3.5	97.0	97.0	2.6
3	86.7	86.2	-8.4	91.0	90.8	5.6	93.0	93.0	2.5
4	77.9	77.1	-11.4	83.6	83.3	7.2	86.7	86.7	2.2
5	66.1	65.1	-14.3	73.4	73.0	8.0	77.6	77.6	1.8
6	54.1	52.9	-17.2	61.7	61.4	7.9	67.1	67.1	1.1
7	42.3	41.1	-20.2	50.7	50.4	7.7	56.6	56.6	0.6
8	31.5	30.2	-23.5	40.4	40.2	8.3	45.5	45.5	0.1
9	22.8	21.4	-25.8	32.1	31.9	8.3	35.5	35.5	0.0
10	16.5	15.2	-27.5	24.5	24.3	8.4	28.1	28.1	-0.6

Summer

Day	1981			1987			1993		
0	100.0	100.0	3.9	100.0	100.0	0.4	100.0	100.0	2.4
1	96.7	96.5	-0.6	98.0	98.0	2.2	98.7	98.7	3.5
2	90.4	89.8	-3.8	94.6	94.4	4.4	96.0	96.0	3.5
3	82.2	81.0	-7.2	88.2	88.0	5.5	91.1	91.1	3.7
4	71.4	70.0	-8.6	78.9	78.6	6.6	83.0	83.0	4.8
5	58.1	56.6	-9.9	67.6	67.2	7.6	72.3	72.2	6.1
6	45.2	43.5	-11.9	55.4	55.1	7.6	59.8	59.8	6.7
7	36.2	34.4	-13.3	43.9	43.6	7.7	48.4	48.4	7.5
8	29.5	27.9	-13.7	33.7	33.5	7.8	38.1	38.2	8.2
9	24.1	22.5	-13.8	25.7	25.6	7.7	29.3	29.5	8.7
10	19.9	18.3	-14.7	19.3	19.3	7.2	23.0	23.3	8.6

Table 1 Anomaly correlations (%) and mean deviations from climate (m) for 500 hPa height analyses (Day 0) and forecasts (Days 1 to 10) for sample seasons, evaluated for the extratropical Northern Hemisphere. See text and Annex for discussion of the two sets of correlations.

	t_d (days)	E_∞ (m)	t_d (days)	E_∞ (m)	t_d (days)	E_∞ (m)	t_d (days)	E_∞ (m)
	Winter		Spring		Summer		Autumn	
1981	1.8	135	1.8	122	1.9	81	1.7	123
1982	1.9	139	1.9	129	2.1	92	1.7	119
1983	2.0	138	1.9	124	1.9	82	1.8	107
1984	1.8	118	1.9	117	1.9	79	1.9	109
1985	1.8	112	1.8	100	1.7	65	1.6	117
1986	1.9	126	1.8	106	1.6	64	1.6	97
1987	1.7	117	1.7	109	1.8	70	1.6	98
1988	1.6	121	1.6	120	1.7	75	1.5	108
1989	1.6	134	1.5	116	1.7	92	1.5	116
1990	1.7	131	1.6	125	1.7	89	1.5	120
1991	1.7	135	1.6	128	1.7	76	1.4	133
1992	1.5	147	1.5	138	1.4	98	1.4	133
1993	1.6	148	1.5	126	1.5	81	1.4	125
1994	1.6	151	1.5	125				

Table 2 Doubling times (t_d) and asymptotic limits (E_∞) of "perfect-model" error growth of the 500 hPa height field computed over the extratropical Northern Hemisphere for each season since the winter of 1981.