Q. J. R. Meteorol. Soc. (2002), **128**, pp. 647–677

Some aspects of the improvement in skill of numerical weather prediction

By A. J. SIMMONS* and A. HOLLINGSWORTH
European Centre for Medium-Range Weather Forecasts, UK

(Received 30 May 2001; revised 18 October 2001)

SUMMARY

Recent verification statistics show a considerable improvement in the accuracy of forecasts from three global numerical weather prediction systems. The improvement amounts to about a 1-day gain in predictability of mean-sea-level pressure and 500 hPa height over the last decade in the northern hemisphere, with a similar gain over the last 3 years in the southern hemisphere. Differences between the initial analyses from the three systems have been substantially reduced.

Detailed study of the European Centre for Medium-Range Weather Forecasts verifications shows that identifiable improvements in the data assimilation, model and observing systems have significantly increased the accuracy of both short- and medium-range forecasts, although interannual (flow-dependent) variations in error-growth characteristics complicate the picture. The implied r.m.s. error of 500 hPa height analyses has fallen well below the 10 m level typical of radiosonde measurement error. Intrinsic error-doubling times, computed from the divergence of northern hemisphere forecasts started 1 day apart, exhibit a small overall reduction over the past 10 years at day two and beyond, and a much larger reduction at day one. Error-doubling times for the southern hemisphere have become generally shorter and are now similar to those for the northern hemisphere. One-day forecast errors have been reduced so much in the southern hemisphere that medium-range forecasts for the region have become almost as skilful as those for the northern hemisphere.

The approach to saturation of forecast error beyond the 10-day range has been examined for sets of 21-day forecasts. When the systematic (sample-mean) component of the error is subtracted, forecast errors and the differences between successive forecasts both appear to level out near the end of the 21-day range at values close to the limit set by the natural level of variance of the atmosphere for the northern hemisphere. A number of features of the model 500 hPa height fields remain quite realistic at the three-week range. The most obvious discrepancy in mean climate is in the Pacific/North-American sector, and variance is too high in the southern hemisphere.

KEYWORDS: Forecast errors Medium-range weather forecasting Predictability

1. INTRODUCTION

A substantial increase in the skill of global numerical weather prediction (NWP) has been achieved by several forecasting centres in recent years. This is seen in more accurate synoptic forecasts as judged by objective verification of high-resolution deterministic forecasting systems, in the improved realism of the weather-element forecasts produced by these systems, and in the increasing utility of probabilistic forecast information produced by lower-resolution ensemble prediction systems. In this paper some aspects of the first of these types of improvement are considered.

Simmons *et al.* (1995) examined the evolution of skill of 500 hPa height forecasts produced daily by the European Centre for Medium-Range Weather Forecasts (ECMWF) from December 1980 to May 1994. Results were discussed in the context of an earlier predictability study by Lorenz (1982) that was based on forecasts for the first 100 days of the period. Simmons *et al.* showed that between 1981 and 1994 there had been a considerable increase in the accuracy of forecasts for a few days ahead, but also a significant increase in the rate at which incipient error was amplified by the (improving) forecast model. Forecast accuracy beyond a few days ahead thus appeared not to have benefited as fully as might have been expected given the reduction in error in the shorter forecast range. Broadly similar findings were reported by Savijärvi (1995) for the US National Centers for Environmental Prediction (NCEP) forecasting system. Simmons *et al.* expressed an optimistic view that the forecast model had reached a

* Corresponding author: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK. e-mail: adrian.simmons@ecmwf.int

stage of development at which its error-amplification rate was realistic, and as a consequence future improvements in data assimilation and modelling that reduced short-range forecast error would be reflected more fully in accompanying improvements in the accuracy of forecasts later in the range. A pessimistic alternative view that short-range error-amplification rates could continue to increase despite overall improvements in the accuracy of initial analyses was also noted.

Some examples of objective verification statistics illustrating the recent increase in forecast skill are presented in the following section of this paper. Section 3 discusses the representation of the intrinsic growth of forecast error as measured by the growth of differences between ECMWF forecasts started 1 day apart and valid for the same time. In section 4 the recent behaviour of the forecasting system is shown to be increasingly difficult to represent accurately by simple exponential growth of small, short-range errors as assumed in the error-growth model utilized by Lorenz (1982) and Simmons *et al.* (1995). A more general error-growth model is adopted (Dalcher and Kalnay 1987; Reynolds *et al.* 1994), and the evolution of parameters of this model is presented over more than 20 years of operational 10-day forecasting at ECMWF. Results for forecasts out to 21 days, made through extension of the range of the control forecasts of ECMWF's ensemble prediction system, are discussed in section 5 for winter 2000–2001 and summer 2001. Analysis accuracy is considered in section 6 and a concluding discussion follows in section 7.

2. Improvements in operational forecasts

Figure 1 presents r.m.s. errors of forecasts of 500 hPa height and mean-sea-level pressure for the extratropical northern and southern hemispheres. Time series from 1990 onwards are shown for 3- and 5-day forecasts from three global prediction systems, that of ECMWF, and those of the Met Office and NCEP, the two national centres that come closest to matching ECMWF's performance according to these measures of forecast accuracy. ECMWF results are also presented for the 4-day range. The plots show annual running means derived from the verification statistics that forecasting centres exchange monthly under the auspices of the World Meteorological Organization, following a pilot study by Lange and Hellsten (1984). Each centre's forecasts are verified by comparison with its own analyses. Results are presented for initial forecast times of 12 UTC for ECMWF and the Met Office, and 00 UTC for NCEP. Additional Met Office results from 00 UTC have been available since 1995, but are not plotted as they are very similar to those presented for 12 UTC. The ECMWF forecasts are routinely produced with a cut-off time for data reception prior to the final analysis cycle that is several hours later than used by the other centres, but evidence from ECMWF forecasts produced with earlier cut-off times indicates that differences in the forecasting systems rather than in data reception are the primary cause of the differences in forecast accuracy illustrated here.

Figure 1 shows a general trend towards lower forecast errors in both hemispheres, for both 500 hPa height and mean-sea-level pressure. The improvement between 1990 and 2001 in ECMWF forecasts for the northern hemisphere amounts to around a 1-day extension of the forecast range at which a given level of error is reached, today's 4- and 5-day forecasts being, respectively, about as accurate on average as the 3- and 4-day forecasts of 10 years ago. The rate of improvement has recently been especially rapid in forecasts for the southern hemisphere, amounting to a 1-day gain in predictability in just 3 years.

The starting point for the rapid recent improvement in ECMWF forecasts shown in Fig. 1 was the operational introduction of four-dimensional variational (4D-Var) data

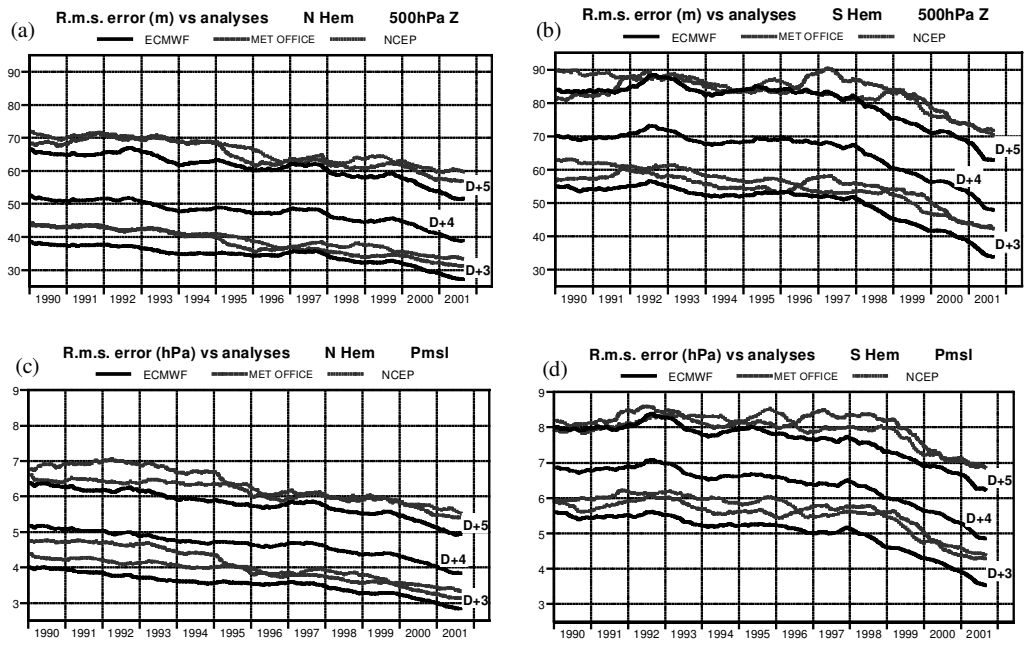

Figure 1. (a) R.m.s. errors of 3- and 5-day forecasts of 500 hPa height (m) for the extratropical northern hemisphere. Results from ECMWF, the Met Office and NCEP are plotted in the form of annual running means of all monthly data exchanged by the centres from January 1989 to August 2001. ECMWF 4-day forecast errors are also shown. Values plotted for a particular month are averages over that month and the 11 preceding months, so that the effect of a forecasting-system change introduced in that month is seen from then onwards. (b) As (a) but for the southern hemisphere; (c) and (d) as (a) and (b), respectively, but for mean-sea-level pressure (hPa).

assimilation (Mahfouf and Rabier 2000, and references therein) in late November 1997. Subsequent data assimilation changes include improved utilization of surface (Järvinen *et al.* 1999) and radiosonde data, assimilation of raw microwave radiances from the TOVS* and new ATOVS (Advanced TOVS) satellite-borne instruments (McNally *et al.* 1999), assimilation of retrievals of humidity (Gérard and Saunders 1999) and surface wind speed from the SSM/I† satellite-borne instrument, and general refinements and extensions of the 4D-Var analysis and use of raw radiances. The atmospheric forecast model has been coupled with an ocean-wave model (Janssen *et al.* 2002) and improved in a number of other ways, including increased vertical resolution in the stratosphere (Untch and Simmons 1999) and planetary boundary layer (Teixeira 1999), revisions to the representations of clouds and convection (Jakob and Klein 2000; Gregory *et al.* 2000) and new schemes for long-wave radiation (Morcrette *et al.* 2001) and for the land surface and sea-ice (van den Hurk *et al.* 2000). Significant increases in the horizontal resolutions of the model and 4D-Var analysis were introduced in November 2000.

Also noteworthy in Fig. 1 are the substantial recent improvements in the forecasts for the southern hemisphere produced by the Met Office and NCEP. Both these centres have reported benefits from use of 3D-Var analysis and direct assimilation of TOVS and ATOVS radiances (Parrish and Derber 1992; English *et al.* 2000; Lorenc *et al.* 2000; McNally *et al.* 2000).

* Television Infrared Observation Satellite Operational Vertical Sounder.

† Special Sensor Microwave/Imager.

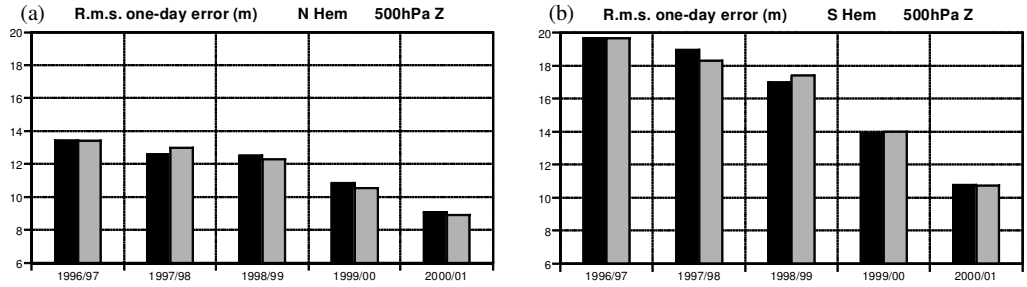

Figure 2. R.m.s. errors of ECMWF 1-day forecasts of 500 hPa height (m) for the extratropical: (a) northern, and (b) southern hemispheres. Twelve-month averages from September to August are plotted for the five years up to August 2001. The black bars denote the actual annual-mean errors, and the grey bars denote the errors that would have occurred had these operational forecasts been improved by exactly the average amounts as measured in the pre-operational trials of the forecasting-system changes introduced between November 1997 and November 2000.

Recent improvements in short-range ECMWF forecasts can be linked very directly to the forecasting-system changes summarized above. The appendix shows when the principal changes were implemented and the impacts on 1-day forecast errors measured during pre-operational trials. Figure 2 shows the actual annual-mean r.m.s. errors of 1-day 500 hPa height forecasts for the past 5 years, together with the errors that would have occurred had the changes introduced between November 1997 and November 2000 given exactly the same average forecast improvements in operational use as were measured in the pre-operational trials. The agreement is remarkable, and indicates that the overall recent improvement in short-range forecasts is indeed due overwhelmingly to changes to the forecasting system rather than to circulation regimes that were unusually easy to predict in the last year or two.

The extent of the reduction in 1-day forecast errors shown in Fig. 2 is also noteworthy. The error has been reduced by almost a third from 13.4 to 9.1 m over 4 years for the northern hemisphere, and almost halved from 19.7 to 10.7 m over the same period for the southern hemisphere. The accuracy of such measures of error is discussed later in section 6.

Verification of forecasts by comparison with radiosonde observations provides a more independent validation than verification by comparison with a centre's own analyses. The observations are, however, mostly located over land, and where sparsely distributed this can give rise to difficulties in interpretation of verification statistics due to variations over time in the number of stations reporting. This inhibits the straightforward comparison of ECMWF and NCEP verifications against radiosondes over the southern hemisphere in particular, since the internationally exchanged verification statistics are for different forecast starting times, and the verification is quite sensitive to differences in radiosonde coverage between 00 and 12 UTC. For example, the 5-day southern hemisphere errors for the year to August 2001 are 72 m for both 00 and 12 UTC Met Office forecasts when verified against analyses, but are 53 and 57 m, respectively, for the 00 and 12 UTC forecasts when verified against radiosondes. Verification against radiosondes generally gives lower values than verification against analyses for this hemisphere, because the observations are predominantly located away from the main band of variance over the Southern Ocean.

Figure 3 presents r.m.s. errors of 3-, 4- and 5-day 12 UTC ECMWF 500 hPa height forecasts verified against radiosondes from 1995 onwards. Values for the northern hemisphere are quite similar to those from verification against analyses at these forecast

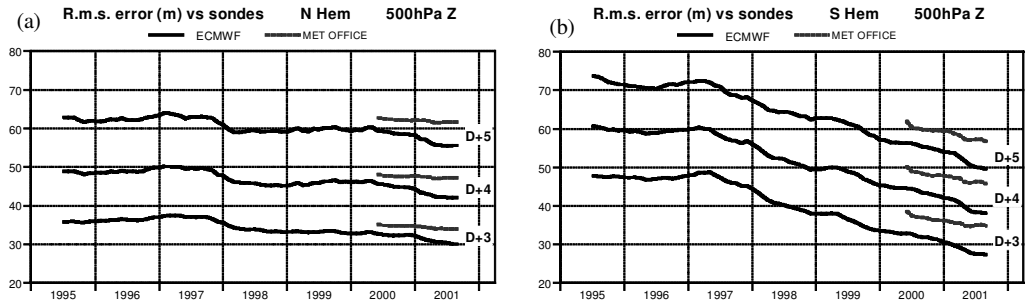

Figure 3. R.m.s. errors of 3-, 4- and 5-day ECMWF 500 hPa height (m) forecasts for the extratropical: (a) northern, and (b) southern hemispheres, plotted in the form of annual running means of monthly data for verification against radiosondes from July 1994 to August 2001. Recent Met Office forecast errors are also shown. Values plotted for a particular month are averages over that month and the 11 preceding months.

ranges. The only difference of note for the southern hemisphere is that the rapid reduction in forecast error is seen to begin some 6 months earlier in 1997 in Fig. 3 than in Fig. 1. This earlier fall in error measured by comparison with radiosonde data appears to be associated with the introduction in May 1997 of a new formulation for the background error constraint in the then operational 3D-Var data assimilation system (Derber and Bouttier 1999). This change brought a marked improvement in tropical forecasts, and may thus have given a greater improvement in verification against those radiosondes located in the southern subtropics than in verification against analyses over the whole southern extratropics.

Figure 3 includes recently available r.m.s. errors of 12 UTC Met Office forecasts verified against radiosondes. Current Met Office errors for the southern hemisphere are distinctly larger than current ECMWF errors, but are much smaller than the ECMWF errors of 4 years ago. A major element of recent changes to the forecasting systems involved wider utilization and new methods of assimilation of satellite data, so it is particularly reassuring to see a large reduction in forecast errors as measured against the southern hemisphere radiosonde network. Evidence that better usage of satellite data has also improved forecasts for the northern hemisphere is provided by recent observing-system experiments reported by Bouttier and Kelly (2001). These indicate a more substantial benefit from the satellite component of the observing system for this hemisphere than reported previously by Kelly (1997).

The levels of skill of northern and southern hemisphere forecasts cannot be compared simply in terms of r.m.s. errors because of inter-hemispheric differences in the levels of variance of the fields. Comparison can, however, be made directly in terms of anomaly correlation coefficients, which are closely related to mean-square errors normalized by corresponding variances (see e.g. Simmons *et al.* 1995). Figure 4 presents anomaly correlations of 500 hPa height based on ECMWF's operational 3-, 5- and 7-day forecasts from January 1980 to August 2001. Running annual means of the monthly-mean skill scores archived routinely over the years are plotted for the two hemispheres.

Figure 4 shows a higher overall rate of improvement in the forecasts for the southern hemisphere. In the early 1980s, the skill levels of the 3- and 5-day forecasts for this hemisphere were only a little better than those of the 5- and 7-day northern hemisphere forecasts. At the time this was not surprising in view of the sparsity of conventional ground-based and aircraft observations in the southern hemisphere (Bengtsson and Simmons 1983). Today, however, the skill at a particular forecast range in the southern hemisphere is only a little lower than that at the same range in the northern hemisphere.

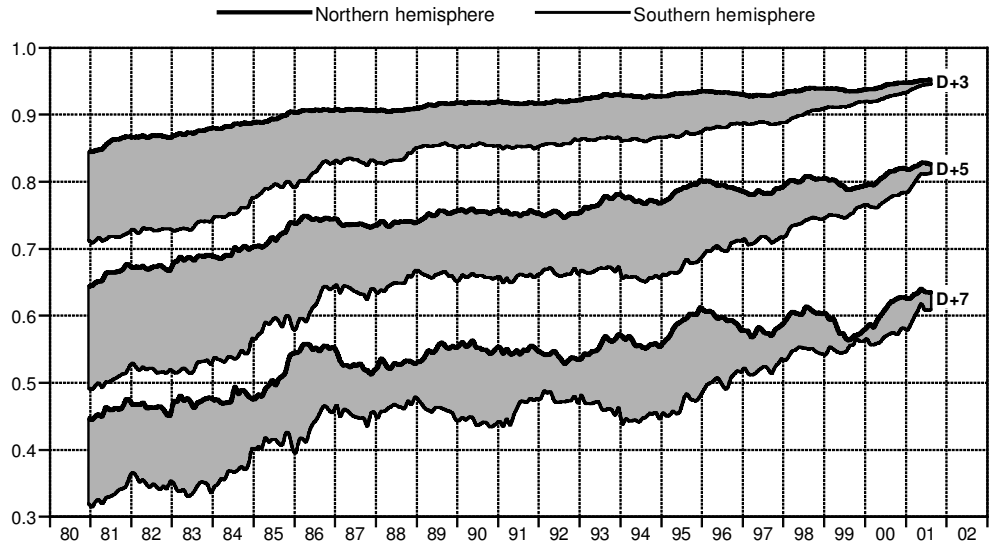

Figure 4. Anomaly correlation coefficients of 3-, 5- and 7-day ECMWF 500 hPa height forecasts for the extratropical northern and southern hemispheres, plotted in the form of annual running means of archived monthly-mean scores for the period from January 1980 to August 2001. Values plotted for a particular month are averages over that month and the 11 preceding months. The shading shows the differences in scores between the two hemispheres at the forecast ranges indicated.

There is little doubt that improvements in the availability, accuracy and assimilation of satellite data have been major factors contributing to the relative improvement in forecast skill. In addition to the changes referred to earlier, information on marine winds has also come from scatterometers on the Earth Resources Satellites (ERS, Tomassini *et al.* 1998), while ERS altimeter data are used in the analysis of ocean wave heights (Janssen 1999). There have been evolutionary improvements in the wind estimates derived by tracking features in successive images from geostationary satellites (Tomassini *et al.* 1999; Rohn *et al.* 2001). Moreover, the newly assimilated raw ATOVS radiances are used more comprehensively over the oceans (where surface radiative properties are easier to characterize) than over land. All these developments would be expected to improve forecasts more in the southern than in the northern hemisphere, both because satellite data provide a more important component of the observing system in the southern hemisphere and because of the greater extent of the oceans in that hemisphere.

Interannual variations in skill are also evident in Fig. 4, especially for the northern hemisphere at the 5- and 7-day time ranges. In particular, there is a pronounced minimum in the northern hemisphere scores arising from relatively poor performance over the year to August 1999. A corresponding maximum can be seen in the time series of r.m.s. errors shown in Figs. 1(a) and (c). This is evident for the Met Office forecasts as well as for those of ECMWF.

An alternative representation of the performance of the ECMWF system since the beginning of daily forecasting is given in Fig. 5. It is based on statistics of 401-day sequences of anomaly correlations of day-five 500 hPa height forecasts. Results are shown for the extratropical northern hemisphere and for Europe (defined as lying from 35 to 75°N and 12.5°W to 42.5°E). The central black lines in each panel denote the

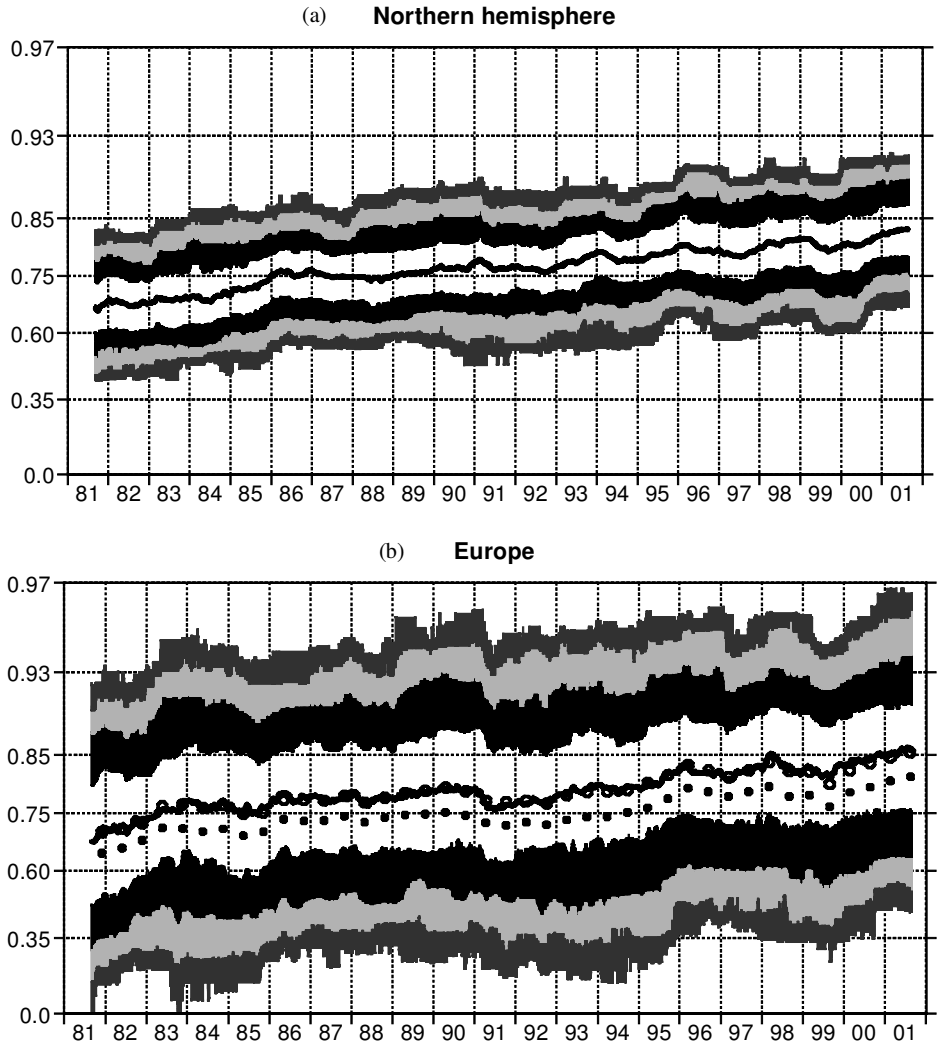

Figure 5. Time series of day-five 500 hPa height anomaly correlation coefficients based on operational ECMWF forecasts with initial dates from 27 July 1980 to 31 August 2001 for: (a) the northern hemisphere, and (b) Europe. The black lines show the median of running 401-day sequences of daily correlations. The shaded bands indicate the ranges of correlations which lie between the 2nd and 4th, the 4th and 10th, and the 10th and 25th percentiles of the 401-day distributions, and also between the 75th and 90th, the 90th and 96th, and the 96th and 98th percentiles. The ordinates are linear in the Fisher transform of the anomaly correlations. The black dots and open circles plotted for Europe denote the running means discussed in the text. Values plotted for a particular day relate to the 401-day period ending on that day.

medians of running 401-day distributions of scores. Comparison with Fig. 4 shows the median for the northern hemisphere to be similar to the annual running mean for the hemisphere. A similar result holds for the southern hemisphere (not shown). Larger differences are found for smaller domains such as Europe, where the mean (shown by the sample of black dots in Fig. 5(b)) is distinctly lower than the median, due to a more skewed distribution of correlations for this region. The Fisher transform, $0.5\{\ln(1+ac) - \ln(1-ac)\}$ where ac is the anomaly correlation coefficient, is the

standard way of transforming a set of correlations for statistical analysis*. The inverse transform of the running mean of the Fisher transforms of the daily correlations is generally close to the median, as indicated by the open circles plotted for Europe in Fig. 5.

The main purpose of Fig. 5 is to demonstrate the variability of forecast skill within a period of a year or so, and the improvement that has been achieved across the range of forecast skill. This is shown by shading the ranges of the correlations lying between the 2nd and 4th percentiles, the 4th and 10th percentiles and the 10th and 25th percentiles of the distribution, and also the ranges between the 75th and 90th, the 90th and 96th, and the 96th and 98th percentiles. The ordinates of Fig. 5 are linear in the Fisher transform, which makes the upper and lower shaded bands of similar width in the plot for one or other region. The much larger variability of correlations over the smaller European region is evident from the wider shaded bands of the lower plot. Overall, improvement has indeed been achieved across the skill range, with no general indication that advances have been concentrated on near- or above-average cases at the expense of not improving the poorer ones. Some variations can occur within shorter periods, as seen for example in the early 1990s, when the median forecast performance for Europe was rising but the skill of the poorer forecasts in the sample was declining. A sharp improvement in these relatively poor forecasts occurred in 1995, probably related to important model changes in April 1995 (Miller *et al.* 1995). Performance in the year to August 1999 was disappointing for Europe in particular, with a marked drop in skill of the better forecasts in the sample. Performance more recently has been the best ever across the range of skill. In the twelve months to 31 August 2001 only 6% of the day-five forecasts for Europe scored below the 0.6 level of correlation that is often regarded as an indication of the limit of usefulness of forecasts (Hollingsworth *et al.* 1980; Simmons 1986), compared with 36% in the first year of daily forecasting, and 19% in the year to 31 August 1991.

3. Forecast errors and the differences between successive forecasts

Lorenz (1982) discussed the comparison of r.m.s. forecast errors with r.m.s. differences between forecasts started a day apart and verifying at the same time. He argued that if the forecast model in operational use at the time was realistic enough for small differences in initial conditions to cause forecasts to diverge at a rate close to that at which separate but similar atmospheric states diverge, then the rate of growth of the forecast differences would provide a limit to the potential accuracy of the forecast that could not be surpassed without analysis or model changes which reduced the 1-day forecast error. The evolution of the forecast differences (or 'perfect-model' errors) would in particular provide a basis for estimating the intrinsic rate of growth of initially small forecast errors.

Lorenz illustrated his discussion with results derived from a purpose-built dataset of ECMWF 500 hPa height analyses and forecasts for the period from 1 December 1980 to 10 March 1981. Such datasets have been produced for every subsequent season, and they provide a convenient basis for the study of forecast performance over more than 20 years. They have been used in the calculations reported in this and the following section. Computational details are as given by Simmons *et al.* (1995).

* The Fisher transform is not ideally suited for the correlations of predicted and analysed anomalies, as it gives large weight to correlations close to -1 as well as to correlations close to $+1$. This is not important here, however, as very few forecasts give negative correlations at the 5-day range.

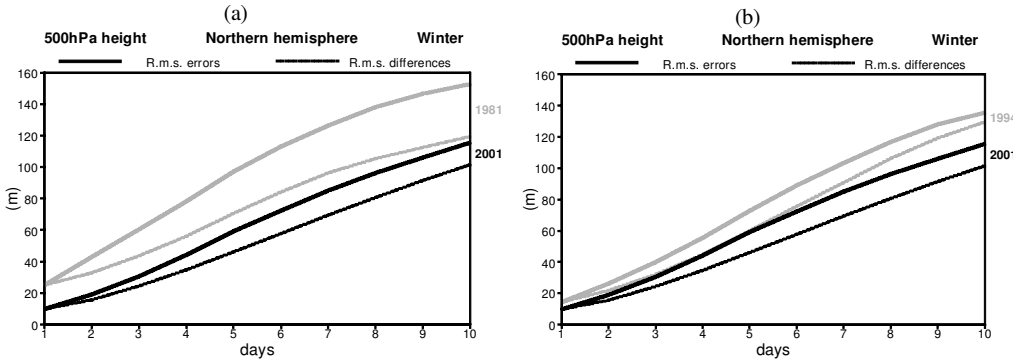

Figure 6. (a) R.m.s 500 hPa height forecast errors (solid) and differences between successive forecasts verifying at the same time (dashed) as functions of the forecast range, computed over the extratropical northern hemisphere, and shown for the winters of 1981 (grey) and 2001 (black). (b) As (a) but for the winters of 1994 (grey) and 2001 (black).

Figure 6(a) presents r.m.s. forecast errors and forecast differences for the extratropical northern hemisphere as a function of forecast range for the first and latest winters* for which results are available, 1981 and 2001. The r.m.s. errors are very much lower in 2001, across the whole forecast range. The reduction since 1981 is about 60% at day one, 50% at day three and 25% at day ten. The gap between the forecast-error and forecast-difference curves is much smaller in 2001 than 1981, indicative of model improvement since 1981. This is known to include a significant reduction in the systematic component of forecast error (Simmons *et al.* 1995; Ferranti *et al.* 2002).

Figure 6(b) compares results for winter 2001 with those for winter 1994, the latest winter for which Simmons *et al.* presented results. The r.m.s. forecast errors for 2001 are some 30% lower at day one and some 15% lower at day ten. In contrast to the comparison of 2001 with 1981, the forecast-difference curves for 1994 and 2001 are farther apart than the corresponding forecast-error curves. Interannual variations in the general circulation can cause variations in the inherent predictability of the atmosphere from one year to another, making it difficult to draw conclusions based on a few sample years. Nevertheless, inspection of results for each winter does not indicate a general closing of the gap between the forecast-error and forecast-difference curves in recent years. Taking the past eight northern winters, the separation between the times at which the two curves reach the 60 m level ranges from close to 1 day for 1997, 2000 and 2001 to as short as 12 hours for 1999. The latter year is thus characterized not only by the relatively large forecast errors noted earlier, but also by a rapid intrinsic error amplification as measured by the divergence of forecasts started a day apart.

It is instructive to compare further the winters of 1999 and 2001. Figure 7 presents maps of mean 500 hPa height over the northern hemisphere for the two winters, and corresponding maps of the mean-square differences between 2- and 1-day forecasts, and between 6- and 5-day forecasts. The short-range forecast differences are generally much smaller in 2001 than in 1999. Examining such plots for each forecast day and many winters reveals a picture of the typical spread of differences. One-day forecast errors and the differences between the analyses from different centres (see Fig. 13) exhibit a maximum over the central/eastern Pacific Ocean. As in Fig. 7, the

* Here winter refers to the period with forecast verification dates from 1 December to 28 February, and the season from 1 December 1980 to 28 February 1981 is referred to as the (northern hemisphere) winter of 1981, and so forth for other years.

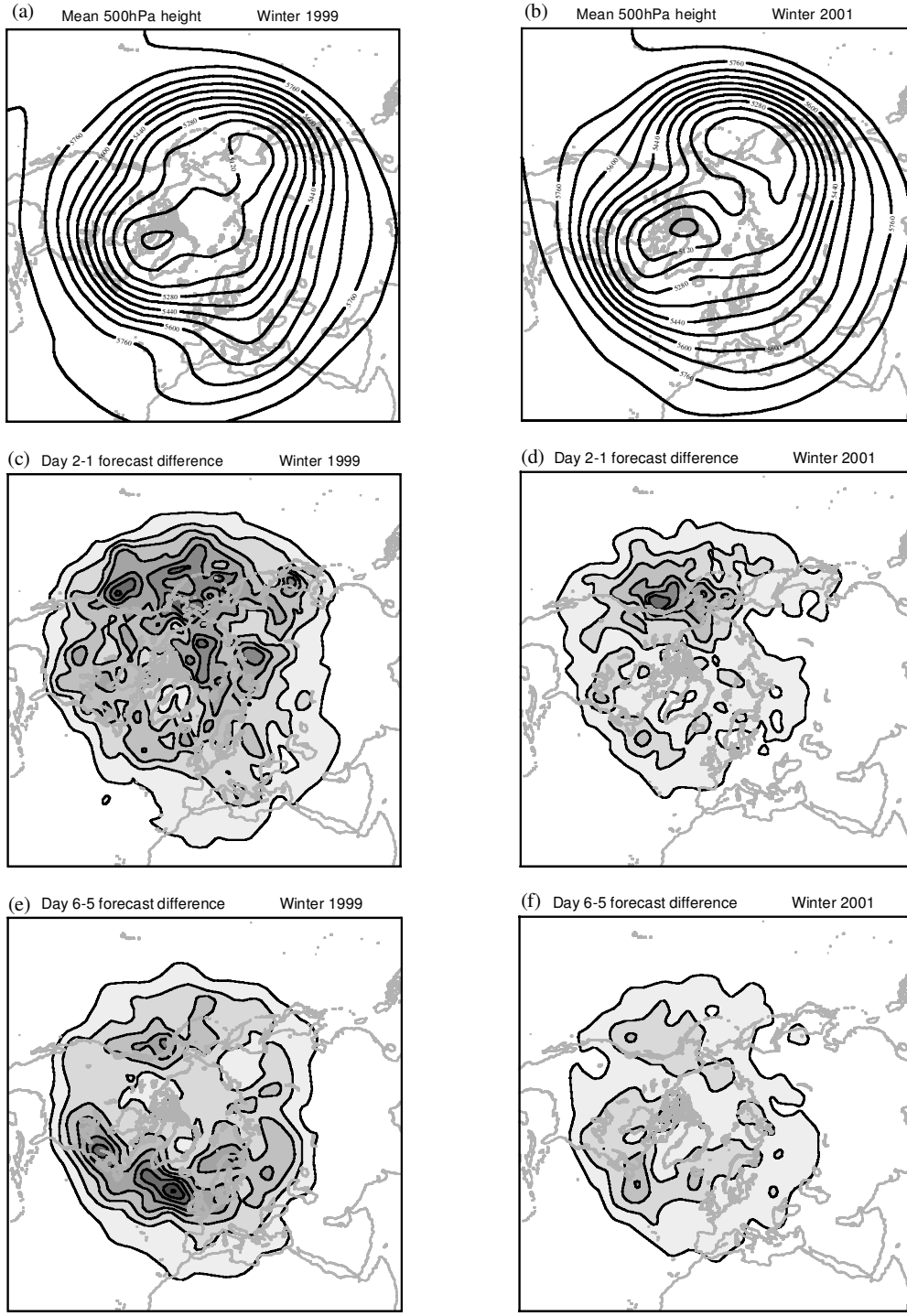

Figure 7. (a) Mean 500 hPa height field (contour interval 80 m) for the winter of 1999; (b) as (a) but for the winter of 2001; (c) and (d) as (a) and (b), respectively, but for mean-square differences between day-two and day-one forecasts (contour interval 200 m^2); (e) and (f) as (c) and (d), respectively, but for mean square differences between day-six and day-five forecasts (contour interval 3000 m^2).

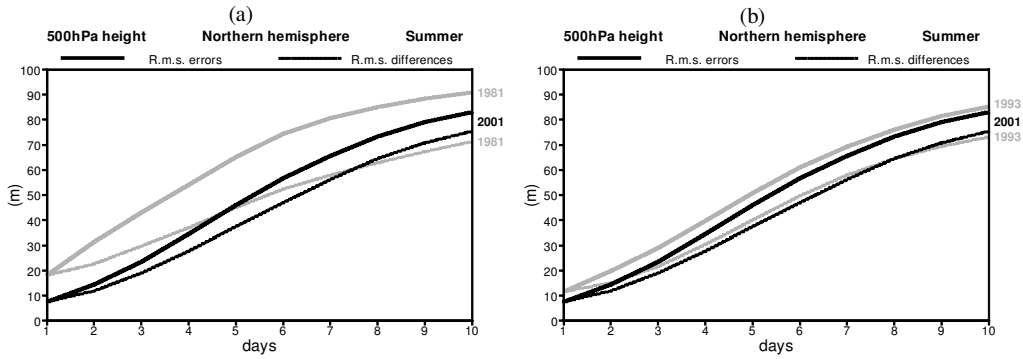

Figure 8. (a) R.m.s 500 hPa height forecast errors (solid) and differences between successive forecasts verifying at the same time (dashed) as functions of the forecast range, computed over the extratropical northern hemisphere, for the summers of 1981 (grey) and 2001 (black); (b) as (a) but for the summers of 1993 (grey) and 2001 (black).

differences between 2- and 1-day forecasts are generally largest over the eastern Pacific. Differences subsequently propagate downstream and amplify in the Atlantic storm-track region. At around day six they are larger over the Atlantic than over the Pacific. This is in accord with synoptic tracking of differences in individual cases and with numerical experiments in which initial analyses are modified over the Pacific, either by withholding observations or by transplanting locally the analysis produced by a different forecasting system (Hollingsworth *et al.* 1985; Rabier *et al.* 1996; Persson 2000). The medium-range forecast differences shown in Fig. 7 differ markedly in pattern and amplitude over the Atlantic between 1999 and 2001, suggesting that the strong mean flow that extended from the eastern seaboard of North America across to the British Isles in 1999 gave a much stronger amplification of differences of upstream origin than occurred in 2001. This is reflected in a stronger intrinsic medium-range error growth as measured by the evolution of r.m.s. forecast differences, illustrated in Fig. 9.

Figure 8 presents r.m.s. northern hemisphere forecast errors and differences for the summers (1 June to 31 August) of 1981, 1993 and 2001. A considerable reduction in the day-one errors can again be seen: by more than 50% between 1981 and 2001 and by more than 30% between 1993 and 2001. By day ten, however, the errors are much more similar to each other than shown for winter in Fig. 6. Moreover, the forecast-difference curve for summer 2001 rises above those for 1993 and 1981 at about day eight, and is closer to the corresponding forecast-error curve than is the case for the two earlier years. The results for 1981 are strongly influenced by the underactive nature of the model at the time (Simmons *et al.* 1995), but the comparison of 1993 and 2001 is not straightforward to interpret. The rate of growth of northern hemisphere forecast differences in the summers of 1999, 2000 and 2001 was larger than in preceding summers, and this may be due either to an inherent lower predictability of the northern summertime circulation for the past 3 years or to a specific deficiency of the current ECMWF forecasting system in summer. In this context, it should be noted that the differences in accuracy between ECMWF forecasts and those of the Met Office and NCEP have, on average, been much smaller in summer than winter in recent years.

The errors and differences of medium-range forecasts were unusually large over Europe in the summer of 1999. There was again substantial amplification of perturbations in relatively strong Atlantic flow, but the errors and differences originated predominantly over the Canadian sector of the Arctic rather than the Pacific (Klinker and Ferranti 2001). There was also an unusually large number of ECMWF forecasts that

were poorer than Met Office forecasts during this period, and the analyses from the two centres tended to differ principally over the Arctic. In the pre-operational trial of changes introduced into the ECMWF system in October 1999 (see appendix), the new version of the system produced analyses that were on average closer to those of the Met Office at high northern latitudes, and forecasts for Europe were much improved in the medium range (Simmons *et al.* 2001). Evidence of a more general reduction of differences between ECMWF and Met Office analyses is presented in section 6.

4. Error-growth models and the evolution of estimates of intrinsic forecast skill

Growth of forecast errors or differences arises from a variety of processes and has much geographical variability, as illustrated above. Simple error-growth models can, nevertheless, be useful as a means of characterizing gross features of forecast performance, although they should not be over-imbued with physical significance.

Lorenz (1982) proposed the use of a two-parameter model of the dependence of r.m.s. error E of a sequence of 'perfect-model' forecasts on their range t:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \alpha E - \beta E^2. \tag{1}$$

The parameters of the model can be expressed in terms of the doubling time of small errors, $(\ln 2)/\alpha$, and the asymptotic level at which error saturates, α/β. These, plus the initial error, determine how the error evolves.

In Lorenz' original study (and in Simmons *et al.* 1995), (1) is replaced by a finite-difference form for the evolution of error from one day to the next:

$$\frac{\Delta E}{\Delta t} = \alpha \overline{E} - \beta \overline{E}^2. \tag{2}$$

Here Δt is one day, $\Delta E_j = E_{j+1} - E_j$ and $\overline{E}_j = 0.5(E_{j+1} + E_j)$, where E_j denotes the error at day j. The parameters α and β are determined by a least-squares fit of the r.m.s. differences between successive real-model forecasts (with weighting inversely proportional to $\overline{E}$). The values of ΔE and $\overline{E}$ to be fitted are given by $(D_{j+1} - D_j)$ and $0.5\ (D_{j+1} - D_j)$ respectively, where D_j is the r.m.s. difference between forecasts with ranges j and $(j-1)$ days, verifying on the same day, and $j = 1, 2, 3, \ldots, 10$ for forecasts carried out to 10 days ahead. D_1 is the r.m.s. error of the 1-day forecast measured against the analysis which initiates the next forecast in the sequence, and E, the initial value for the modelled errors, is set equal to D. With E given, the fitted perfect-model error at day j is derived sequentially in terms of E from the discrete form (2):

$$E_j = \frac{\alpha - 2}{\beta} - E_{j-1} + \left\{ \left(\frac{\alpha - 2}{\beta} \right)^2 + \frac{8E_{j-1}}{\beta} \right\}^{\frac{1}{2}}.$$

Simmons *et al.* (1995) illustrated how this two-parameter error-growth model provided a good fit to forecast differences throughout the 10-day range for the most recent data available at the time. Since then, however, the fit has become poorer in general, although there are exceptions. This is illustrated in Fig. 9, which plots the evolution of northern hemisphere forecast differences out to day seven for the winters of 1999 and 2001. The actual differences are denoted by the black continuous lines, and the results of fitting the two-parameter model are shown by the grey solid lines with

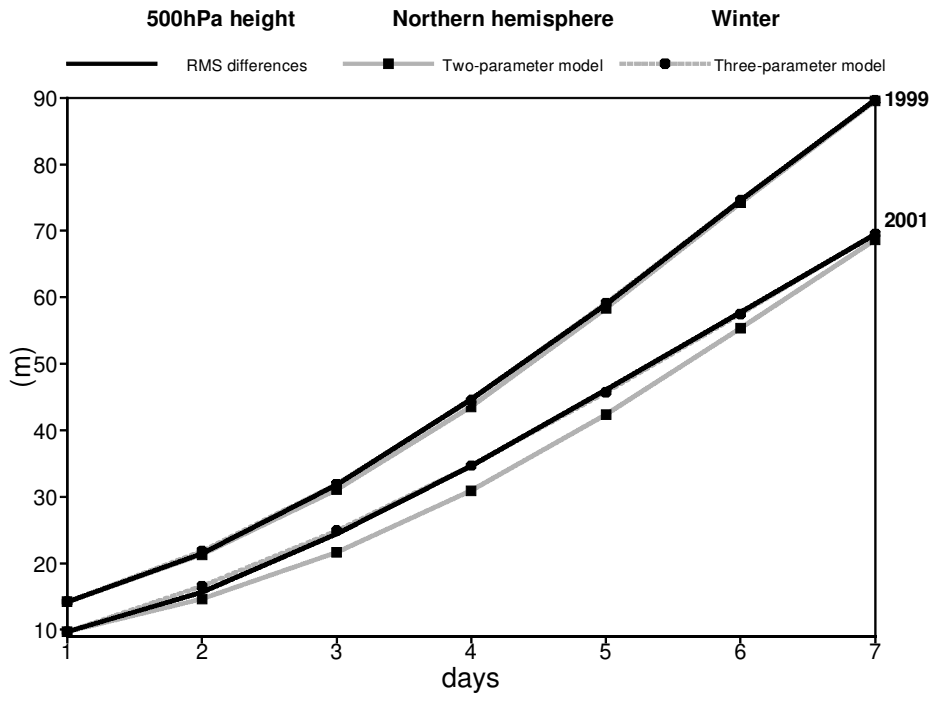

Figure 9. R.m.s. differences between successive 500 hPa height forecasts verifying at the same time ('perfect-model' errors) as functions of the forecast range, for the northern hemisphere winters of 1999 and 2001. The solid black lines denote the actual differences, the solid grey lines with square marker symbols denote the results of fitting the two-parameter error-growth model to these differences, and the grey dashed curves with circular marker symbols denote the results of fitting the three-parameter model. Results are shown only to day seven so that the closeness of fit can be more easily seen.

square marker symbols. The fit is close for 1999 but much poorer for 2001. As discussed earlier, the growth of forecast differences is much more rapid in the medium range for 1999.

The growth of r.m.s. forecast differences for winter 2001 is in fact more linear in time than is given by the fit to the two-parameter model. A modified functional fit including a linear growth term (Dalcher and Kalnay 1987; Reynolds *et al.* 1994) has thus been considered:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \gamma + \alpha E - \beta E^2. \tag{3}$$

Dalcher and Kalnay (1987) and, in effect, Reynolds *et al.* (1994) applied (3) to the variance of forecast errors rather than r.m.s forecast differences. They followed Leith (1978) in viewing the linear term γ as describing the growth of forecast error due to model deficiencies, and the exponential term primarily as describing the growth of error originating from erroneous initial conditions. Justification for this view was based on their finding (supported by Simmons *et al.* 1995) that the short-range growth of differences between successive forecasts could be modelled quite accurately by the exponential growth term αE, and that the linear term was needed only to obtain a good representation of actual forecast errors.

This justification clearly does not apply to current ECMWF forecasts. The linear term in the present case allows some representation of a growth of analysis error that is more rapid over the first day or two of the forecast than further into the forecast

range. This fast initial growth could result from amplification of components of the analysis error with the shallow structures identified by singular-vector studies as having the greatest potential for short-range growth (Buizza and Palmer 1995; Rabier *et al.* 1996). This would occur if the data assimilation system is unable to recognize and correct background errors in such structures as effectively as it recognizes and corrects background error in general, either because of lack of appropriate observations or because of inadequate formulation of the system itself.

Results from the 'three-parameter' error-growth model will be presented in terms of the small error-doubling times at 1- and 2-day ranges, $(\ln 2)/(\alpha + \gamma/E_1)$ and $(\ln 2)/(\alpha + \gamma/E_2)$, and the asymptotic limit, E_∞ where

$$E_\infty = \frac{\alpha}{2\beta} + \left(\frac{\alpha^2}{4\beta^2} + \frac{\gamma}{\beta}\right)^{\frac{1}{2}}.$$

For continuity with the studies of Lorenz (1982) and Simmons *et al.* (1995), the three parameters α, β and γ are derived by a least-squares fit to forecast differences as in the case of the two-parameter model. E_j is now given in terms of E_{j-1} by

$$E_j = \frac{\alpha - 2}{\beta} - E_{j-1} + \left\{\left(\frac{\alpha - 2}{\beta}\right)^2 + \frac{8E_{j-1}}{\beta} + \frac{4\gamma}{\beta}\right\}^{\frac{1}{2}}.$$

The grey dashed curves with circular marker symbols included in Fig. 9 show that the three-parameter model provides a good fit to the r.m.s. forecast differences for both 1999 and 2001. This has also been found for other seasons. It should be noted, though, that the fit tends to be poorest at day two, as can be seen for 2001 in Fig. 9. The three-parameter model tends to overestimate the growth-rate at day one, in contrast to the underestimation provided by the two-parameter Lorenz model for recent years. In general, the doubling times calculated at day two in terms of the two growth terms of this three-parameter model, $(\ln 2)/(\alpha + \gamma/E_2)$, are quite close to the doubling times given by the Lorenz model.

Figure 10 presents: (a) the evolution from 1981 to 2001 of the 1-day forecast errors of 500 hPa height for the two hemispheres; (b) the corresponding 1- and 2-day intrinsic error doubling times from the three-parameter error-growth model; and (c) estimates of the asymptotic error limits. The 1-day errors are plotted for each season, but for clarity the growth rates and asymptotic limits are plotted as annual running means of values computed for each season.

The 1-day forecast errors show a general decline during the 1980s in the northern hemisphere, a more substantial decline in the mid-to-late 1980s in the southern hemisphere, a relatively slow decline in the early to mid 1990s in both hemispheres, and a recent sharp decline, especially in the southern hemisphere. This pattern is generally reflected in the time evolution of anomaly correlations of the medium-range forecasts for the two hemispheres shown in Fig. 4. For the season from 1 December 2000 to 28 February 2001, the 1-day error for both hemispheres is below the 10 m level typical of radiosonde measurement error (Lönnberg and Hollingsworth 1986). The error for the (summer) southern hemisphere has become lower than that for the (winter) northern hemisphere.

The 1-day error for the northern hemisphere winter of 1999 plotted in Fig. 10 is rather larger than a smooth downward trend would indicate, suggesting that the flow pattern may have made this a difficult winter for numerical weather prediction even in the short range. It has been noted in reference to Fig. 9 that the evolution

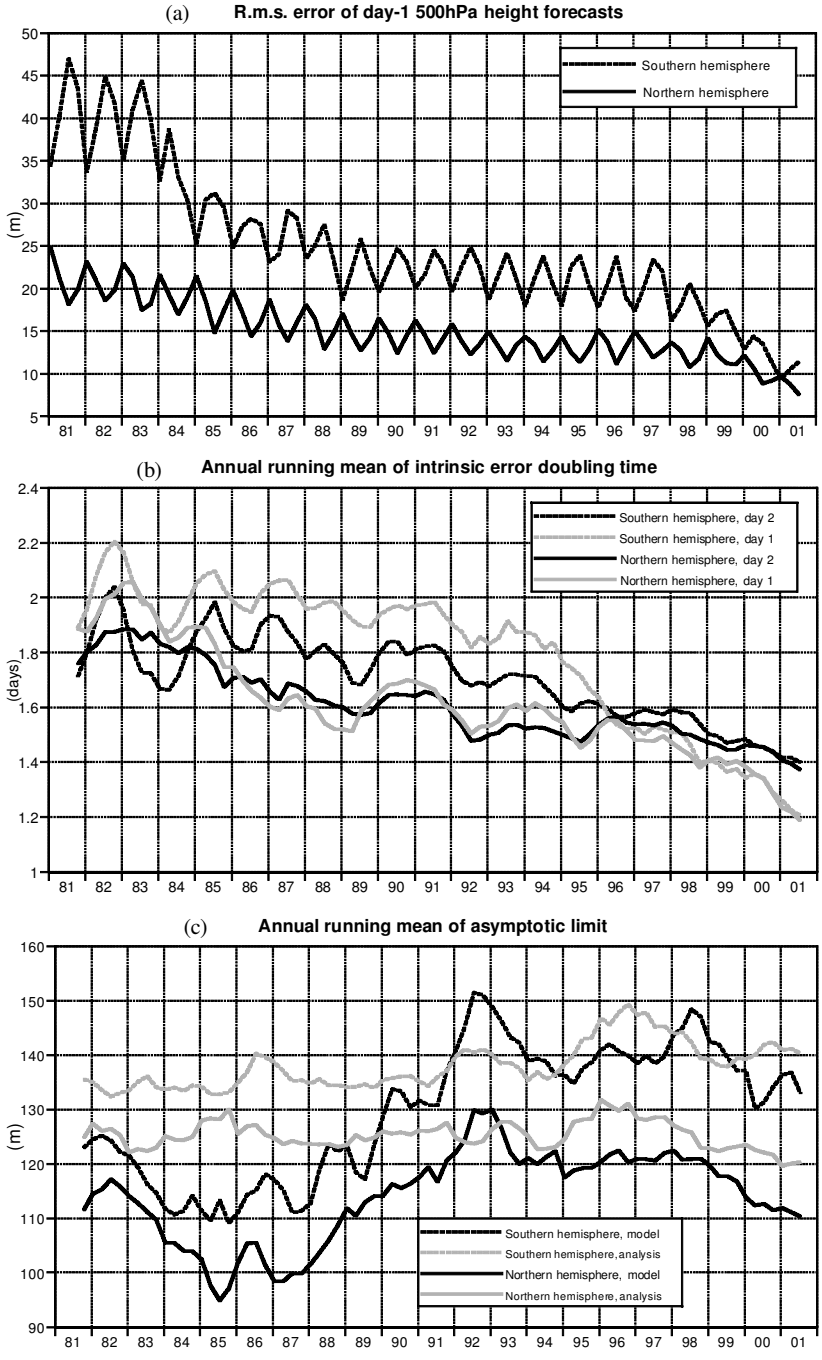

Figure 10. (a) R.m.s. errors of 1-day 500 hPa height forecasts for the extratropical northern (solid) and southern (dashed) hemispheres plotted for each season since 1981; (b) corresponding running annual means of the error-doubling times at days one (grey) and two (black); and (c) the asymptotic limits (black) from the three-parameter error-growth model, and also the asymptotic limits given by the root-mean-square over the hemispheres of twice the intraseasonal variance of analyses (grey).

TABLE 1. DOUBLING TIMES (DAYS) FROM THE TWO GROWTH TERMS OF THE THREE-PARAMETER MODEL FOR THE NORTHERN HEMISPHERE, FOR FORECAST RANGES OF 1, 2, 3 AND 4 DAYS

Season	Day 1	Day 2	Day 3	Day 4	Season	Day 1	Day 2	Day 3	Day 4
DJF 1996/97	1.42	1.59	1.71	1.79	JJA 1997	1.62	1.53	1.48	1.45
DJF 1997/98	1.32	1.54	1.70	1.81	JJA 1998	1.53	1.51	1.50	1.49
DJF 1998/99	1.43	1.49	1.53	1.56	JJA 1999	1.43	1.44	1.45	1.46
DJF 1999/00	1.35	1.54	1.68	1.77	JJA 2000	1.36	1.42	1.46	1.48
DJF 2000/01	1.14	1.42	1.62	1.76	JJA 2001	1.23	1.33	1.39	1.42

DJF is December, January and February, JJA is June, July and August.

of r.m.s. forecast differences was much closer to exponential in form in winter 1999 than in winter 2001. Table 1 shows that, of the past five winters, 1999 was unusual in this respect. The table also indicates that the recent evolution of short-range forecast differences has been generally closer to exponential in summer than in winter, as the summer doubling times show less variation with forecast range.

Several features of the error-doubling-time curves in Fig. 10 are worthy of note. The first is a general reduction of doubling times in both hemispheres up to the early 1990s, as noted by Simmons *et al.* (1995) for the Lorenz error-growth model. During this period the doubling times computed for the northern hemisphere are generally similar for days one and two, indicating that fitting the three-parameter error-growth model to the forecast differences yields rather small values for the linear growth parameter γ, consistent with the two-parameter Lorenz model providing a reasonable fit. Also as noted by Simmons *et al.*, there is a marked gap between the doubling times for the two hemispheres for the 10 years up to 1994, with slower growth in the southern hemisphere.

The day-two error-doubling times have shortened a little over recent years for the northern hemisphere, more so than the doubling times computed for days three and four (Table 1). Error-doubling times have shortened more substantially for the southern hemisphere, to values that are now very similar to those of the northern hemisphere. Thus recent forecast behaviour does not support the inference of Simmons *et al.* (1995) that the general circulation of the southern hemisphere is inherently more predictable than that of the northern hemisphere.

A novel feature of the most recent verifications is that error-doubling times in both hemispheres have become distinctly shorter at day one than at day two. The reduction at day one for the southern hemisphere is substantial. Although the three-parameter model tends to overestimate the growth of forecast differences from day one to day two, a marked increase in diagnosed initial growth rates is confirmed by examining simple direct measures of the rate at which differences grow from day one to two and day two to three, such as given by $\ln (D_2/D_1)$ and $\ln (D_3/D_2)$. A plausible explanation for this behaviour is that recent analysis and model changes have been more effective in reducing the larger scales of analysis and short-range forecast error than they have been in reducing rapidly amplifying smaller-scale analysis errors. Supporting evidence is provided by r.m.s. 500 hPa wind errors at 1-day range. Annual-mean values have been reduced by 11% and 27% respectively for the northern and southern hemispheres over the past 4 years. These reductions in wind errors are significantly smaller than the corresponding reductions of 32% and 45% in 500 hPa height errors shown in Fig. 2.

Figure 10(c) shows not only the asymptotic limits computed for the error-growth model but also the limits given by the r.m.s. of twice the intraseasonal variance of the analyses. To remove the effect of the seasonal cycle, the intraseasonal variance is defined as the variance about the seasonal mean of the analysed anomaly with respect to climate. The analysed anomaly for a particular day and year is computed as the difference

between the analysis for that day and year and the average analysis (the climate) for that day taken over all years in the period shown. This provides the true error limit for a large sample of long-range forecasts from a perfect model, assuming error in the estimation of the intraseasonal variance is negligible.

The asymptotic limits derived from forecast differences can be seen in Fig. 10 to be particularly low in the 1980s. This is because of the underactive nature of the forecast model at the time (Simmons *et al.* 1995). They are high around 1992 due to initial problems of overactivity of the higher-resolution semi-Lagrangian model introduced in September 1991 (Ritchie *et al.* 1995). For recent years, the limits from the error-growth model are generally lower than those derived from the intraseasonal variance of analyses, rather more so for the northern than for the southern hemisphere. Further discussion is given in the following section.

5. FORECASTS TO A RANGE OF 21 DAYS

Since 21 November 2000, the ECMWF ensemble prediction system (EPS) has been based on a model with T255 horizontal resolution (Buizza and Hollingsworth 2002), and T511 resolution has been used for the deterministic forecasting system, results from which have been presented in preceding sections. The 'T255L40' EPS model has 40 levels in the vertical. Its vertical resolution is similar in the troposphere to that of the 60-level T511L60 deterministic forecasting model and is degraded in the stratosphere. The perturbed EPS forecasts are carried out over the same 10-day range as the deterministic forecast, but the unperturbed EPS control forecast is carried out to 21 days ahead, to gain information on predictability and model error beyond the 10-day range. It should be noted that this 21-day forecast, like the 10-day forecasts, is carried out with fixed rather than evolving modelled sea-surface temperatures. This may be one factor limiting its skill towards the end of the time range. The initial analysis for the T255L40 forecast is interpolated from the operational T511L60 analysis.

The errors and differences of the 10-day T511L60 and 21-day T255L40 forecasts have been analysed for the first 90 days for which 21-day T255L40 forecasts are available, 12 December 2000 to 11 March 2001, and for the period from 1 June to 31 August 2001. To facilitate comparison with results from other centres, the calculations in this and the following section use a 2.5° computational grid and T63 truncation of operationally archived spherical-harmonic analysis and forecast fields rather than the specialized T40 datasets and 3.75° grid used for the results shown in sections 3 and 4. Climatological information for the winter and summer periods is derived from a 22-year set of analyses formed by combining ERA-15 analyses from 12 December 1979 to 31 August 1993 (Gibson *et al.* 1997) and operational analyses for subsequent seasons up to 31 August 2001.

Figure 11(a) shows r.m.s. errors and differences computed for winter 2001 over the extratropical northern hemisphere from the two models. The r.m.s. errors are fairly similar out to day ten, though somewhat smaller for the higher-resolution model. The r.m.s differences are almost indistinguishable, but grow slightly more slowly beyond day seven for T511L60. The approach to saturation of T255L40 errors and differences can be seen beyond 10 days, with r.m.s. differences ceasing to grow during the final 2 days of the forecast range, having reached a level some 10 m or so lower than that of the r.m.s. errors. This mismatch could result either from a 'systematic' (seasonal-mean) component of the forecast error, which would inflate the errors but hardly affect the differences, or from an underactive model producing too little variance, which would reduce the differences more than the errors.

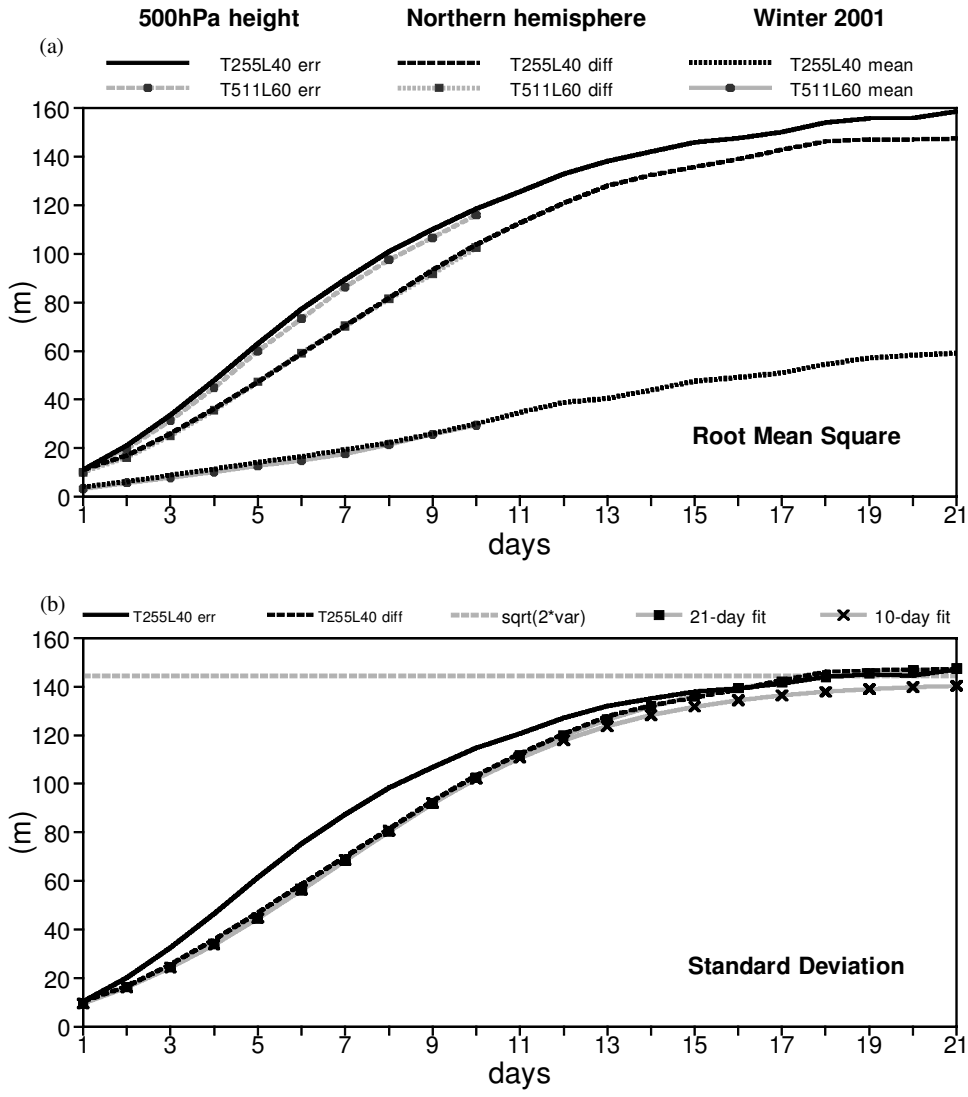

Figure 11. (a) R.m.s., and (b) standard deviation of T255L40 forecast errors (solid, black) and differences between successive forecasts (dashed, black) for the forecast range up to 21 days, for northern hemisphere forecasts verifying in the period 12 December 2000 to 11 March 2001. (a) Also shows errors and differences (grey, dashed and dotted) for corresponding T511L60 forecasts to 10 days, and the root-mean-square for the hemisphere of the sample-mean T255L40 (black, dotted) and T511L60 (grey, solid) forecast errors. (b) Includes the curves (solid, grey) that result from fitting the three-parameter error-growth model to the differences over 21 days (square marker symbols) and 10 days (crosses) and the asymptotic limit derived from the variance of analyses (dashed, grey).

It is the systematic error that appears to be responsible for the mismatch between the forecast errors and differences. Figure 11(a) includes the evolution of the r.m.s. of the systematic component of the error. Its magnitude of 55 m at day 21 is sufficient when squared to account for the difference between the mean-square errors and differences at this range. This is confirmed in Fig. 11(b), which shows that the standard deviations*

* Standard deviation is taken here to refer to the r.m.s. value computed over a region for a seasonal sample of fields from each of which the seasonal average but not the regional average has been subtracted.

TABLE 2. ASYMPTOTIC LIMITS (m) DERIVED FROM T255L40 FORECAST DIFFERENCES OUT TO 10 AND 21 DAYS AHEAD, AND AS GIVEN BY THE R.M.S. OF TWICE THE INTRASEASONAL VARIANCE OF ANALYSES

	12 December 2000 to 11 March 2001			1 June to 31 August 2001		
Hemisphere	From 10-day forecasts	From 21-day forecasts	R.m.s. of twice intraseasonal variance	From 10-day forecasts	From 21-day forecasts	R.m.s. of twice intraseasonal variance
Northern	142	150	144	83	90	94
Southern	123	126	118	144	170	157

of forecast errors and differences become very close to each other towards the end of the 21-day forecast range. Moreover, the two curves reach a level very close to the asymptotic limit derived from the intraseasonal variance of analyses, with an indication of a very slightly higher level of activity in the forecasts than in the analyses after day 17.

Also shown in Fig. 11(b) is the representation of the forecast differences given by the three-parameter error-growth model, with parameters determined firstly using all 21 available daily values of the standard deviations of forecast differences, and secondly using only forecast differences out to day ten. Fitting only the 10-day forecast differences can be seen to underestimate the asymptotic limit. This is found also for summer and for both seasons in the southern hemisphere, and is consistent with the results shown in Fig. 10(c). The asymptotic limits derived from both 10- and 21-day forecast differences and also from the intraseasonal variance of analyses are presented in Table 2 for the two hemispheres and seasons. Forecast differences for the summer northern hemisphere level out a little below the natural limit derived from the variance of analyses, suggesting that the forecast model is slightly underactive for this region and season. Conversely, forecast differences exceed the natural limit beyond about day 15 in the southern hemisphere, for both summer and winter. The variance of the forecast fields remains close to that of the analyses over the first week or so of the forecast range in the southern hemisphere, but the model becomes overactive thereafter.

The nature of the systematic, or seasonal-mean, component of the forecast error merits discussion. Systematic forecast error has typically been ascribed to error in the forecast model causing the 'climate' of forecasts (the mean of a large sample of forecasts) to drift away from the true climate as the forecast range increases. Such an attribution of error was certainly justified in the early years of forecasting at ECMWF when the systematic component of the error was especially large. This component has, however, been substantially reduced in amplitude since then, and three sources of mean forecast error must now be considered. One source remains the model error that would cause a long integration to drift to a climate different to that of the atmosphere. The second, which is likely to be weak in general, is due to error in the analysis of prescribed or slowly varying fields such as sea-surface temperature, soil moisture or stratospheric humidity.

The third source of systematic forecast error is a consequence of the nature and predictability of seasonal-mean anomalies. The mean of a set of daily day-21 forecasts for a single 90-day season is an average of 90 realizations of the atmosphere whose deviations from climatology are to a large extent uncorrelated from one verification day to the next in the absence of strongly anomalous and persistent forcing such as from anomalous sea-surface temperatures in an extreme El Niño. The mean of the analyses for the season is, in contrast, a mean over 90 consecutive days, and the field for a particular day may be quite strongly correlated with fields for neighbouring days, particularly during persistent blocking events for example. The intraseasonal variability

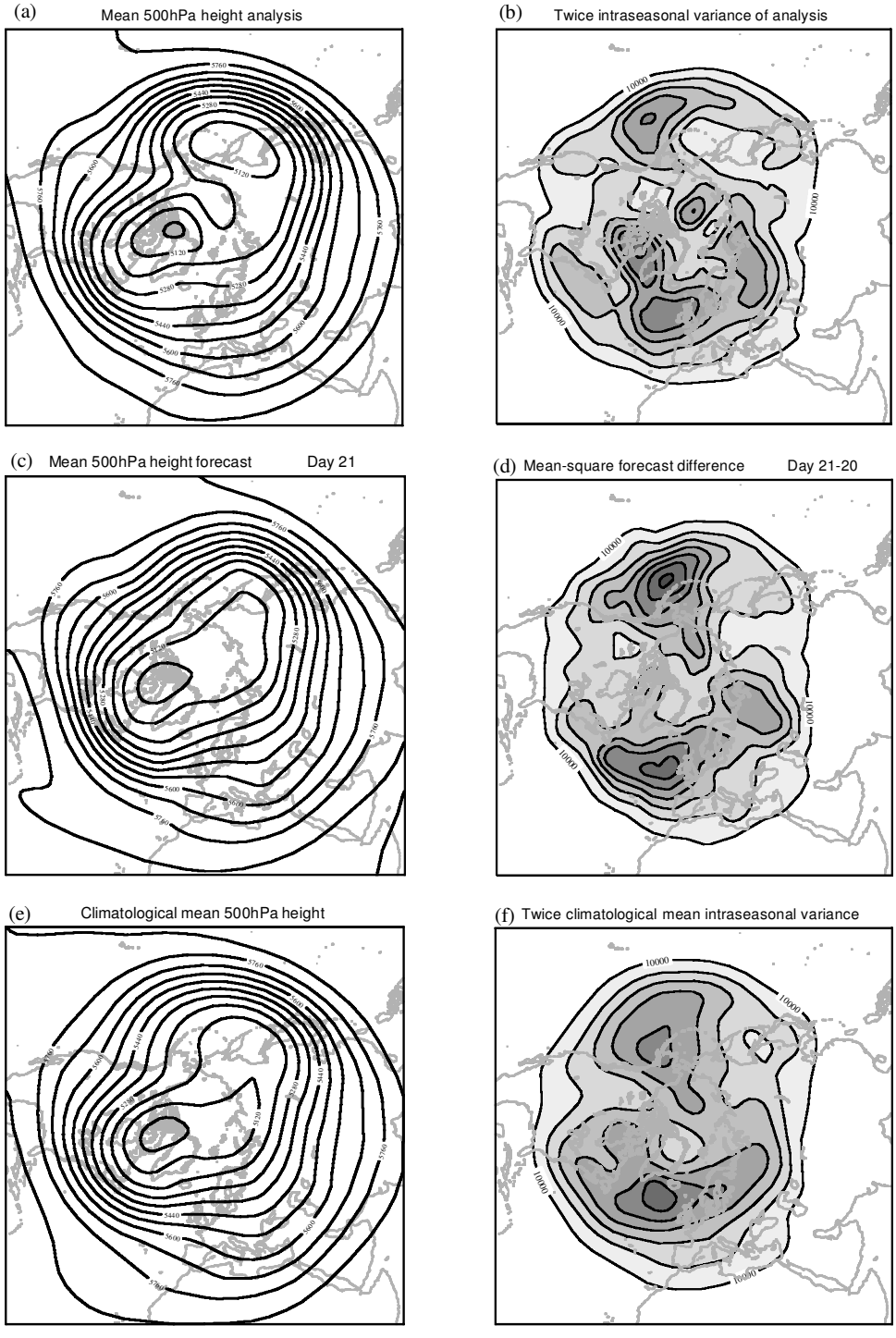

Figure 12. (a) Mean 500 hPa height field from analyses for the period from 12 December 2000 to 11 March 2001, and (b) twice the intraseasonal variance of analyses for the period (contour interval 10 000 m^2). (c) Mean 500 hPa height field from 21-day T255L40 forecasts verifying in the period, and (d) the mean-square differences between 21-day and 20-day forecasts verifying in the period. (e) Mean 500 hPa height field from a corresponding 22-year climatology, and (f) twice the corresponding mean intraseasonal variance from the 22-year climatology. (500 hPa contour intervals are 80 m.)

of the general circulation on time-scales longer than the few days typical of baroclinic-wave growth and barotropic-wave decay may be viewed as comprising a sequence of regimes with stochastic transitions from one to another, associated with individual baroclinic developments, for example. The seasonal anomaly may in turn be viewed as the residual from averaging over the set of regimes that occur during the season, with anomalous forcing biasing the frequency of occurrence of particular regimes (Corti *et al.* 1999). If the predictability of regime transitions is low in the second half of the 21-day forecast range, the 90 21-day forecasts would be expected to provide a sampling of the climatological frequencies of occurrence of regimes that is more complete than that provided by the 90 consecutive daily analyses. The forecast means and variances would thus be expected to be closer to climatological values than are the means and variances of the analyses for the season.

Figures 12(a), (c) and (e) present northern hemisphere maps of 500 hPa height from 12 December 2000 to 11 March 2001: Fig. 12(a) shows the mean analysis, Fig. 12(c) the mean of the 21-day forecasts verifying in the period, and Fig. 12(e) the climatological mean. Several features of the mean forecast height field in the sector stretching eastwards from eastern North America to Asia are indeed closer to climatology than to the mean analysis for winter 2001. It is not straightforward to separate genuine climate drift due to model imperfection from apparent drift due to regime sampling in these results for a single winter season, but the mean forecast pattern over the central/eastern Pacific and western North America is more typical of a seasonal extreme of the Pacific/North-American teleconnection pattern (Wallace and Gutzler 1981) than of the climate mean, and is thus indicative of a systematic model error.

Figures 12(b), (d) and (f) show maps of twice the intraseasonal variance of the analyses for winter 2001, the mean-square difference between the 21- and 20-day forecasts for the period, and twice the climatological mean of the intraseasonal variance, respectively. The pattern of the mean-square forecast differences at 3-week range is more similar to the pattern of the climatological intraseasonal variance than it is to the pattern of the intraseasonal variance for winter 2001. The overall magnitude of the mean-square differences is less than twice the climatological mean intraseasonal variance, 2.2×10^4 m^2 rather than 2.4×10^4 m^2 in the extratropical average, but slightly larger than the figure of 2.1×10^4 m^2 for twice the intraseasonal variance for winter 2001. The variance for the northern hemisphere winter of 2001 was, in fact, lower than for any other northern winter in the 22-year record, and the tendency of the extended-range forecasts to exhibit a slightly higher level of activity than the analyses may thus not be indicative of a general model trait.

6. ANALYSIS ACCURACY

Comparison of the analyses and short-range forecasts produced by different centres can be used to assess the accuracy of these analyses and the reliance that can be placed on some of the results presented earlier in this paper. In this section a set of results is presented comparing ECMWF analyses and forecasts with corresponding Met Office products provided daily to ECMWF on a 2.5° latitude/longitude grid.

Table 3 presents r.m.s. differences between the ECMWF and Met Office analyses for the extratropical northern and southern hemispheres. Results are shown for the periods from 12 December to 11 March and from 1 June to 31 August examined in the preceding section, and cover the past four years. Differences between the analyses have been reduced very substantially over these years, particularly for the southern

TABLE 3. R.M.S. DIFFERENCES (m) BETWEEN ECMWF AND MET OFFICE 500 hPa HEIGHT ANALYSES OVER THE EXTRATROPICS FOR THE PERIODS 12 DECEMBER TO 11 MARCH AND 1 JUNE TO 31 AUGUST FOR FOUR YEARS

	12 December–11 March				1 June–31 August			
Hemisphere	1997/98	1998/99	1999/00	2000/01	1998	1999	2000	2001
Northern	14.1	15.3	12.8	9.7	10.6	10.3	8.9	8.2
Southern	20.7	21.4	16.1	11.5	29.7	27.4	17.0	14.2

hemisphere. The fact that the analyses from the two centres have become much closer to each other does not necessarily imply that both sets have become much more accurate, but given the substantial improvements in forecast accuracy achieved by both centres it may be inferred that both centres' analyses have indeed become significantly closer to the truth. Similar conclusions are drawn from comparisons of ECMWF and NCEP analyses*.

Maps of the mean-square differences between the ECMWF and Met Office analyses from 12 December 2000 to 11 March 2001 are shown in Figs. 13(a) and (b). The contribution from systematic differences has been excluded. The mean-square differences are below 50 m^2 over substantial areas of the northern continental land masses, but are larger in a band stretching from the Arctic southwards over central Asia, where radiosonde coverage is lower than over other northern land areas, and also over western and northern North America, where the background fields of the data assimilation suffer from the lower accuracy of upstream analyses over the Pacific and Arctic oceans. Differences are larger over the mid- and high-latitude oceans, particularly in the southern hemisphere, and are generally largest over the Arctic and Antarctic. Local minima can also be seen, for example at the South Pole and around parts of the coastline of Antarctica, reflecting the availability of radiosonde observations which both analyses fit closely. In the northern hemisphere maxima exceed 150 m^2 only over two small regions of the Pacific, over Siberia and around the North Pole, where the difference almost reaches 400 m^2. A maximum of almost 400 m^2 also occurs in the African sector of the Southern Ocean. The largest difference exceeds 700 m^2 over Antarctica.

Figures 13(c) and (d) show corresponding mean-square differences between the 1-day forecasts of the two centres. The two sets of forecasts diverge rapidly over the first day as differences amplify and spread downstream. The maxima in mean-square differences almost quadruple over the Pacific, triple over the Atlantic, and more than double in the southern westerlies. There is also a substantial increase of differences over land areas, with mean-square differences more than tripling over Europe and almost tripling over North America. Differences also generally grow at high latitudes, though more slowly. The maxima over the Arctic and Antarctic change little, growing slightly in magnitude in the case of the Arctic, but decaying a little over Antarctica.

The analysis differences and corresponding short-range forecast errors may be used to provide an indication of the overall accuracy of each set of analyses. The r.m.s. difference between two sets of analyses, d, satisfies

$$d^2 = a_{1E}^2 + a_{2E}^2 - 2a_{1E}a_{2E}c_{12},$$

where a_{1E} is the r.m.s. error of the first set of analyses, a_{2E} is the r.m.s. error of the second set of analyses and c_{12} is the pattern correlation between the errors of the two

* Numerical results for the NCEP analyses are not presented as these analyses are provided routinely to ECMWF only on a coarser 5° grid.

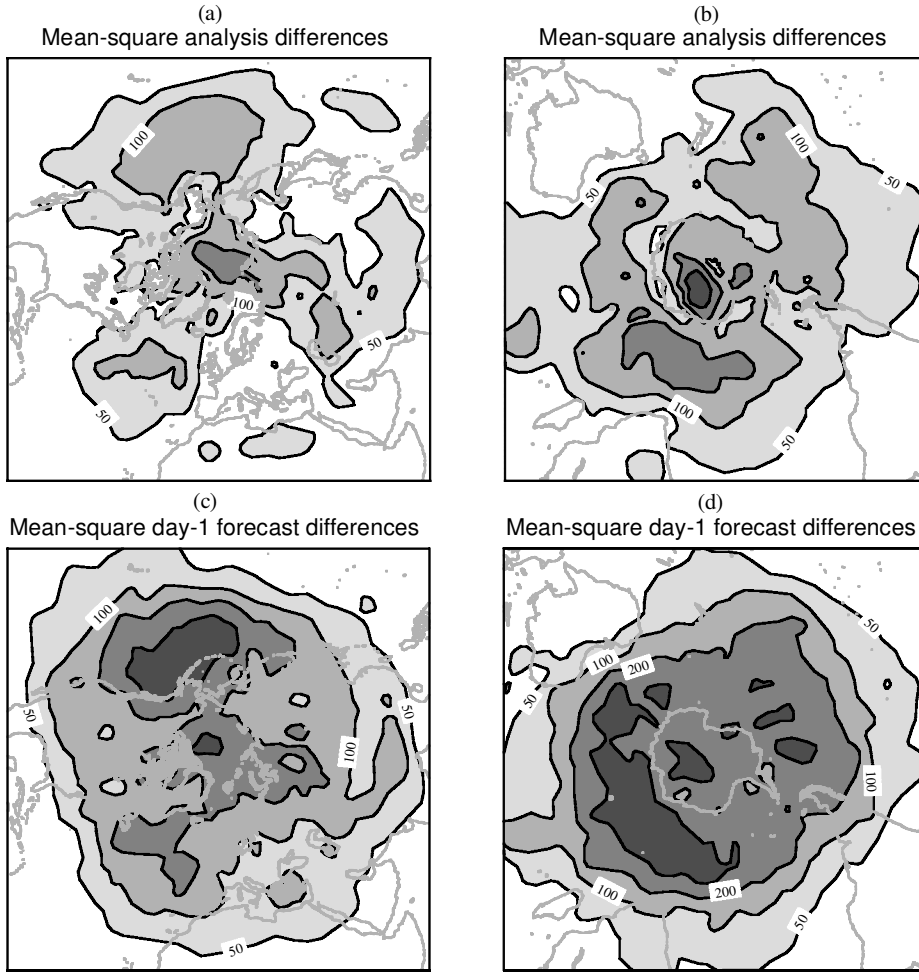

Figure 13. (a) Mean-square differences between ECMWF and Met Office 12 UTC 500 hPa height analyses over the northern hemisphere for the period 12 December 2000 to 11 March 2001. Contour values are 50, 100, 200 and 400 m^2. Seasonal mean differences have been excluded and smoothing has been applied to the 2.5° data to reduce the amplitude of small-scale features. (b) As (a) but for the southern hemisphere; (c) and (d) as (a) and (b), respectively, but for 1-day forecasts of 500 hPa height.

sets of analyses. Knowing d, estimates of a_{1E} and a_{2E} can be made by assuming that both their ratio and c_{12} are the same as the ratio and correlation calculated from short-range forecast errors.

For the period 12 December 2000 to 11 March 2001, the r.m.s. 500 hPa height errors of 1-day ECMWF and Met Office forecasts (verified against their respective analyses) are 10 and 14 m respectively for the northern hemisphere and 10 and 15 m for the southern hemisphere. The correlations between the 1-day forecast errors are 0.40 for the northern hemisphere and 0.36 for the southern hemisphere*. These values indicate r.m.s. analysis errors of the order of 7 m for ECMWF and 10 m for the Met

* Correlations increase monotonically from 0.3 at a 12-hour range to 0.55 at a 6-day range for the northern hemisphere, and similarly from 0.28 to 0.51 for the southern hemisphere. The asymptotic limit for this correlation is 0.5 for large sample size and long forecast range, in the absence of systematic error in the means and variances of model and analysis fields.

Office for the (winter) northern hemisphere and 8 m for ECMWF and 12 m for the Met Office for the (summer) southern hemisphere. The estimates are reduced by 1 m if the lower correlation of 0.2 is used. For both hemispheres, 0.2 is the value derived for the analysis time by linear extrapolation of the correlations computed for 12- and 24-hour forecast errors. Corresponding calculations for 1 June to 31 August 2001 give analysis error estimates of 5 to 6 m for the northern hemisphere and 7 to 8 m for the southern hemisphere in the case of ECMWF.

Alternative estimates of analysis error can be made by extrapolating either the forecast errors or the forecast differences back to the initial forecast time using values at days one, two and perhaps three. This gives similar or somewhat lower estimates of analysis error, depending on quite how the calculation is carried out.

Having derived an estimate of analysis error, an estimate can also be made of the extent to which error in the verifying analysis affects the measured short-range forecast error. The true r.m.s. forecast error, f_{T}, satisfies

$$f_{\mathrm{T}}^2 - 2 f_{\mathrm{T}} a_E c_{fa} + a_E^2 - f^2 = 0,$$

where f is the r.m.s. forecast error verified against analyses, a_E is the r.m.s. error of the verifying analyses and c_{fa} is the correlation between the true forecast errors and the errors of the verifying analyses. At the start of the forecast, f_{T} is equal to a_E, the correlation c_{fa} is 1.0 and the measured error is zero. The correlation decreases with increasing forecast range as the verifying analysis is a forecast state that has been corrected by assimilation of the observations available over the forecast period. An estimate of c_{fa} is provided by correlating differences between ECMWF and Met Office forecasts with the corresponding differences in verifying analyses. This gives correlations of 0.37 for the northern hemisphere and 0.47 for the southern hemisphere at 1-day range for the period 12 December 2000 to 11 March 2001. Corresponding correlations are 0.46 and 0.50 for 1 June to 31 August 2001. Substituting values for the measured 1-day forecast errors and the estimated analysis errors shows that there is an approximate cancellation between the terms $2 f_{\mathrm{T}} a_E c_{fa}$ and a_{E}^2. The measured 1-day r.m.s. errors of recent 500 hPa height forecasts are thus estimated to be within a metre or so of the true errors. As the forecast error approximately doubles and the correlation approximately halves from day one to day two, a similar cancellation occurs at day two.

Conversely, verification against radiosondes gives 1-day errors that are significantly inflated by errors in the radiosonde measurements, estimated by Lönnberg and Hollingsworth (1986) to be of the order of 10 m for the 500 hPa height. Thus verification of 1-day ECMWF forecasts for the northern hemisphere gives r.m.s. values of around 10 m against analyses but 14 m against radiosondes for winter 2001. This is a larger discrepancy than seen in the early medium range (Figs. 1 and 3), as errors in verifying observations or analyses contribute decreasingly to measured r.m.s. forecast error as the latter grows.

7. Discussion

A substantial recent improvement in the accuracy of forecasts for the extratropical northern and southern hemispheres has been illustrated. The improvement is particularly evident in the southern hemisphere and is, to a degree, common to several global forecasting centres. A factor in this has been the extent of cooperation among institutions worldwide, which has been particularly strong in the development of variational assimilation of radiance data from satellites. Among shared developments within Europe are those between ECMWF and Météo-France in forecasting-system software to

enable *inter alia* direct (3D-Var and 4D-Var) assimilation of radiances (Andersson *et al.* 1994), and between the Met Office, ECMWF and Météo-France in radiative transfer modelling for radiance assimilation (Saunders *et al.* 1999). Other collaborative developments include those between ECMWF and NCEP in data assimilation (Derber and Bouttier 1999) and the assimilation of raw radiance data in particular (McNally *et al.* 2000), and between ECMWF and the Australian Bureau of Meteorology in radiance bias correction (Harris and Kelly 2001). Specific recent changes that have reduced 1-day ECMWF forecast errors have been identified and their impacts quantified using results from pre-operational trials.

Differences between the centres' analyses have also been reduced substantially, especially for the southern hemisphere. The pre-operational trial results from ECMWF (see appendix) and observing-system experiments (English *et al.* 2000; Bouttier and Kelly 2001; McNally, personal communication) provide evidence that the reduction of analysis differences is due largely to improvements in data assimilation methods and to the use of new and improved types of satellite data. There have also been other observing-system improvements, such as increasing numbers of reports from commercial aircraft. These developments have helped in particular to reduce uncertainties over the oceans, where analysis differences are generally larger than over land masses that have a good coverage of *in situ* observations.

The implied substantial increase in analysis accuracy places new demands on the accuracy and resolution both of observing and of data assimilation systems if the potential for future analysis and forecast improvements is to be realized. The launch of high-resolution satellite-borne sounding instruments, such as the Advanced Infra-Red Sounder (AIRS) and the Infrared Atmospheric Sounding Interferometer (IASI), and the establishment of programmes for targeted *in situ* observations, offer the promise of improving analyses of the shallow structures over oceans that can still cause rapid growth of error when misrepresented (e.g. Prunet *et al.* 1998; Montani *et al.* 1999). Complementary data assimilation developments include work aimed at improving the use of key observations by implementing improved, flow-dependent estimation of the variances and covariances of background errors, especially in regions of strong sensitivity to analysis error (e.g. Fisher 1998). This targeted reduction in the growth of short-range forecast errors is expected to lead to a significant downstream reduction in medium-range forecast errors. Reduction of slower-growing initial errors may, nevertheless, also be of importance for improving the accuracy of medium-range forecasts. For example, errors initially tend to be large in Arctic regions and, although growing relatively slowly at first, may propagate into the midlatitude westerlies and subsequently amplify rapidly, as happened over the Atlantic in the summer of 1999.

The results presented here indicate that identifiable improvements in the ECMWF forecasting system that have significantly reduced analysis and short-range forecast errors, have indeed also led to substantially lower medium-range forecast errors. Diagnosed error-growth rates have, nevertheless, continued to increase. Intrinsic error-doubling times computed from the divergence of northern hemisphere forecasts started 1-day apart show a small overall reduction over the past 10 years from day two onwards, and a much larger reduction at day one. Doubling times for the southern hemisphere have become generally shorter and quite similar to those for the northern hemisphere. One-day forecast errors have been reduced very substantially for the southern hemisphere, however, and the reduction has been large enough to give lower errors across the whole of the 10-day forecast range despite the increase in error-growth rate.

The approach to saturation of forecast error beyond the 10-day range has been examined for sets of 21-day forecasts. Care is needed in the interpretation of the results,

TABLE 4. R.M.S. HEIGHT ERRORS (m) AT 500 hPa OF 1-DAY FORECASTS FOR 1989 FROM ERA-40, ERA-15 AND OPERATIONS AT THE TIME, AND FROM CURRENT OPERATIONS FOR THE YEAR TO 31 AUGUST 2001

Hemisphere	ERA-40 1989	ERA-15 1989	Operations 1989	Operations 2000/01
Northern	12.8	14.4	14.7	9.1
Southern	16.5	22.0	22.0	10.7

but the model 500 hPa height fields remain quite realistic at the 3-week range in several respects. The most obvious discrepancy in mean climate is in the Pacific/North-American sector, and variance is too high in the southern hemisphere. Further study of such forecasts will be made for subsequent seasons as results become available, and planned monthly forecasts using a T159 version of the atmospheric model coupled to an ocean model will provide further material for the study of predictability beyond the 10-day range.

Recent changes to forecasting systems and observational usage have been substantial and frequent, and at several points in this paper questions have arisen as to the extent to which interannual differences in forecasting-system performance are due to these changes or to interannual variations in the atmospheric circulation. The improvements measured for the 1-day operational ECMWF forecasts are of the order expected from the pre-operational testing of changes, but the situation is less clear as regards the medium-range forecasts. For example, an unusually low level of variance of the northern hemisphere circulation for winter 2000/01 has been noted, which may well have been reflected in lower medium-range forecast errors than would otherwise have occurred.

These issues are being explored further in the context of the ERA-40 re-analysis project (Simmons and Gibson 2000). The project entails analysis of observations taken over the period from mid-1957 to the present, using a fixed data assimilation system. It benefits from many but not all of the recent forecasting-system changes listed in the appendix. For reasons of cost and timely production it uses 3D-Var rather than 4D-Var data assimilation and an assimilating model with lower (T159) horizontal resolution. In ERA-40 production, T159 forecasts are being run daily to a 36-hour range from the analyses for 00 and 12 UTC. Averaged over 1989, the first year of full production, the r.m.s. errors of the 1-day 500 hPa height forecasts from 12 UTC are significantly lower than both those of the forecasts run as part of the ERA-15 project (Gibson *et al.* 1997) and those of the operational forecasts made in 1989, though higher than current operations, as illustrated in Table 4.

ERA-15 used a version of the ECMWF forecasting system which differed from that operational in mid-1995 only in its lower (T106) horizontal resolution. The short-range forecast improvements between ERA-15 and ERA-40 thus reflect many of the changes made to the operational forecasting system over the past 5 years or so. ERA-40 also benefits from a better collection of conventional *in situ* data. The lower errors of current ECMWF operations reflect not only the impact of 4D-Var and higher horizontal resolution but also the impact of a number of enhancements to the observing system made since 1989, although the period has also seen a decline of the radiosonde network over the former Soviet Union. There may also be some impact of circulation differences between the years.

An ERA-40 follow-up activity is planned in which the 12 UTC forecasts will be extended to 10 days ahead. The aim is to explore interannual variations in medium-range predictability that are not obscured by forecasting-system changes. Observing-system experiments are also planned, to help separate variations in predictability that

arise from variations in atmospheric or surface conditions from those that arise from the major changes to the observing system introduced over the period of the re-analysis.

ACKNOWLEDGEMENTS

Very many people at ECMWF and elsewhere have contributed to the forecasting improvements discussed here. David Richardson is thanked for his maintenance of the special datasets that formed the basis for much of the study. We thank Martin Ehrendorfer and Martin Leutbecher whose questions prompted our re-examination of the error-growth model. Comments from Erik Andersson, Horst Böttger, Eugenia Kalnay, Graeme Kelly, Anders Persson and an anonymous referee are gratefully acknowledged. The ERA-40 project is partially funded by the European Union through contract EVK2-CT-1999-00027.

APPENDIX

Recent changes to the ECMWF forecasting system

New versions or cycles of the operational ECMWF forecasting system typically comprise a set of changes each of which is first tested in isolation and then perhaps tested together with one or more of the other changes. The extent of this testing depends on the nature and anticipated effect of each change. The changes are then merged to form the new cycle, which is subjected to more comprehensive testing. This includes a trial in which the new cycle is run in parallel with the established operational cycle for several weeks or months, usually followed immediately by operational implementation. The results from these extensive pre-operational trials provide the primary evidence available as to the sources of the improvements seen in operational performance, although for cycles in which several major changes are combined the evidence as to the impact of individual changes is less clear.

Caution has to be exercised in drawing conclusions regarding the impact on medium-range forecasts of changes tested in the shorter pre-operational trials, as the impact of a new cycle that is tested particularly extensively can be quite different from one month to the next, particularly when evaluated for smaller regions (Simmons 2000), and the impact may also depend on the season in which the trial was carried out. More consistent and reliable measures of improvement are found from study of short-range impacts.

One-day r.m.s. errors of mean-sea-level pressure and 500 hPa height for the extra-tropical northern and southern hemispheres have been examined for all pre-operational trials carried out from that of 4D-Var in autumn 1997 onwards. All new cycles that changed the 500 hPa height error by more than 0.25 m on average for at least one of the hemispheres over the trial period are listed in Table A.1. The table also lists the start date of the trial period, the operational implementation date and the reductions (positive numbers) or increase (negative number) in r.m.s. 1-day forecast error for both hemispheres, whilst notes below indicate the principal changes made for each cycle. Each forecast is verified against the analysis produced using the same cycle as the forecast, and each trial was run up to the date of operational implementation. The final date of the trial for the purpose of the 1-day forecast verifications is accordingly 2 days prior to operational implementation.

These results show a larger overall reduction in error for the southern than for the northern hemisphere. The reductions are highly consistent in magnitude with the reductions in 1-day operational forecast errors that have occurred since 1997, as shown in Fig. 2. One of the changes listed gave an increase in 500 hPa height error for one

TABLE A.1. REDUCTIONS IN R.M.S. 1-DAY FORECAST ERRORS MEASURED IN PRE-OPERATIONAL TRIALS OF RECENT CHANGES TO THE ECMWF FORECASTING SYSTEM

	Cycle	18r1	18r6	19r2	21r1	21r4	22r1	22r3	23r1	23r3
	Principal changes (see below)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Start date of trial period	9 Oct 1997	16 Apr 1998	1 Jan 1999	10 Mar 1999	7 May 1999	5 Feb 2000	1 Mar 2000	1 Jul 2000	1 Jul 2000
	Date implemented	25 Nov 1997	29 Jun 1998	9 Mar 1999	5 May 1999	12 Oct 1999	8 Apr 2000	27 Jun 2000	12 Sep 2000	21 Nov 2000
NH	Z500	0.5	0.4	0.4	0.2	1.5	−0.1	0.6	0.4	0.8
	MSLP	0.06	0.02	0.01	0.00	0.13	0.00	0.07	0.04	0.07
SH	Z500	1.8	0.0	0.4	0.8	2.3	0.5	1.9	0.7	0.7
	MSLP	0.27	0.02	0.01	0.09	0.39	0.02	0.19	0.08	0.07

SH and NH are the southern and northern hemispheres, respectively. Z500 is the 500 hPa height forecast error (m). MSLP is the mean-sea-level pressure forecast error (hPa).
1. 4D-Var replaces 3D-Var (with 6-hourly cycling).
2. Introduction of coupling with ocean-wave model. Improved utilization of radiosonde data. Use of hourly surface data. Assimilation of humidity from SSM/I.
3. Increased vertical resolution in the stratosphere.
4. Assimilation of raw MSU and AMSU-A radiances.
5. Increased vertical resolution in PBL and improved representation of clouds, convection and orography. New background-error statistics for 4D-Var. Corrected processing of humidity observations. Assimilation of marine winds from SSM/I.
6. Improved quality control of SSM/I data and assimilation of the data from a second satellite. Corrected stratospheric humidity analysis
7. New background- and observation-error variances. Assimilation of additional ATOVS radiance data. Improved representation of the land surface, sea-ice and radiation
8. Twelve-hourly cycling of 4D-Var.
9. T511 horizontal resolution for atmospheric model and T159 analysis increments. Doubled angular resolution for ocean-wave model.

hemisphere, but the increase is small in magnitude and is flagged as a worsening of forecast accuracy only at the 10% confidence level according to a t-test applied to the set of individual forecast differences, assuming them to be temporally uncorrelated. All other changes in 500 hPa height error shown in Table A.1 are recognized as improvements at a confidence level of better than 0.1%, with the exception of the small northern hemisphere improvement found with cycle 21r1, which is significant at the 0.5% level.

REFERENCES

Andersson, E., Pailleux, J., Thépaut, J.-N., Eyre, J. R., McNally, A. P., Kelly, G. A. and Courtier, P. 1994 Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, **120,** 627–653

Bengtsson, L. and Simmons, A. J. 1983 'Medium-range weather prediction—operational experience at ECMWF'. Pp. 337–363 in *Large-scale dynamical processes in the atmosphere.* Eds. B. J. Hoskins and R. P. Pearce. Academic Press, London, UK

Bouttier, F. and Kelly, G. A. 2001 Observing system experiments in the ECMWF 4D-Var data assimilation system. *Q. J. R. Meteorol. Soc.*, **127,** 1469–1488

Buizza, R. and Hollingsworth, A. 2002 Storm prediction over Europe using the ECMWF Ensemble Prediction System. *Met. Appl.*, in press

Buizza, R. and Palmer, T. N. 1995 The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52,** 1434–1456

Corti, S., Molteni, F. and Palmer, T. N. 1999 Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, **398,** 799–802

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

Derber, J. C. and Bouttier, F. 1999 A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A,** 195–222

English, S. J., Renshaw, R. J. Dibben, P. C., Smith, A. J., Rayer, P. C., Poulsen, C., Saunders, F. W. and Eyre, J. R. 2000 A comparison of the impact of TOVS and ATOVS satellite sounding data on the accuracy of numerical weather forecasts. *Q. J. R. Meteorol. Soc.*, **126,** 2911–2931

Ferranti, L., Klinker, E., Hollingsworth, A. and Hoskins, B. J. 2002 Diagnosis of systematic forecast errors dependent on flow pattern. *Q. J. R. Meteorol. Soc.*, in press

Fisher, M. 1998 'Development of a simplified Kalman filter'. ECMWF Tech. Memo. No. 260. Available from ECMWF, Shinfield Park, Reading, UK

Gérard, E. and Saunders, R. W. 1999 Four-dimensional variational assimilation of Special Sensor Microwave/Imager total column water vapour in the ECMWF model. *Q. J. R. Meteorol. Soc.*, **125,** 3077–3101

Gibson, J. K., Kållberg, P., Uppala, S., Nomura, A., Hernandez, A. and Serrano, E. 1997 'ERA Description'. ECMWF Re-Analysis Final Report Series 1. Available from ECMWF, Shinfield Park, Reading, UK

Gregory, D., Morcrette, J.-J., Jakob, C., Beljaars, A. C. M. and Stockdale, T. 2000 Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecasting System. *Q. J. R. Meteorol. Soc.*, **126,** 1685–1710

Harris, B. A. and Kelly, G. A. 2001 A satellite radiance bias correction scheme for data assimilation. *Q. J. R. Meteorol. Soc.*, **127,** 1453–1468

Hollingsworth, A., Arpe, K., Tiedtke, M., Capaldo, M. and Savijärvi, H. 1980 The performance of a medium-range forecast model in winter—impact of physical parameterizations. *Mon. Weather Rev.*, **108,** 1736–1773

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

Jakob, C. and Klein, S. A. 2000 A parametrization of the effects of cloud and precipitation overlap for use in general-circulation models. *Q. J. R. Meteorol. Soc.*, **126,** 2525–2544

Janssen, P. A. E. M. 1999 'Wave modelling and altimeter wave height data'. ECMWF Tech. Memo. No. 269. Available from ECMWF, Shinfield Park, Reading, UK

Janssen, P. A. E. M., Doyle, J. D., Bidlot, J., Hansen, B., Isaksen, L. and Viterbo, P. 2002 Impact and feedback of ocean waves on the atmosphere. *Adv. Fluid. Mech.*, in press

Järvinen, H., Andersson, E. and Bouttier, F. 1999 Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus*, **51A,** 469–488

Kelly, G. A. 1997 Influence of observations on the operational ECMWF system. *WMO Bulletin*, **46,** 336–342

Klinker, E. and Ferranti, L. 2001 'Forecasting system performance in summer 1999. Part 1: Diagnostics related to the forecast performance during spring and summer 1999'. ECMWF Tech. Memo. 321. Available from ECMWF, Shinfield Park, Reading, UK

Lange, A. and Hellsten, E. 1984 'Results of the WMO/CAS NWP Data Study and Intercomparison Project for forecasts for the northern hemisphere'. WMO, Geneva, Switzerland

Leith, C. E. 1978 Objective methods for weather prediction. *Ann. Rev. Fluid Mech.*, **10,** 107–128

Lönnberg, P. and Hollingsworth, A. 1986 The statistical structure of short range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus*, **38A,** 137–161

Lorenc, A. C., Ballard, S. P., Bell, R. S., Ingleby, N. B., Andrews, P. L. F., Barker, D. M., Bray, J. R., Clayton, A. M., Dalby, T., Li, D., Payne, T. J. and Saunders, F. W. 2000 The Met. Office global three-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **126,** 2991–3012

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

McNally, A. P., Andersson, E., Kelly, G. A. and Saunders, R. W. 1999 'The use of raw TOVS/ATOVS radiances in the ECMWF 4D-Var assimilation system'. Pp. 2–7 in ECMWF Newsletter No. 83. Available from ECMWF, Shinfield Park, Reading, UK

McNally, A. P., Derber, J. C., Wu, W. and Katz, B. B. 2000 The use of TOVS level-1b radiances in the NCEP SSI analysis system. *Q. J. R. Meteorol. Soc.*, **126,** 689–724

Mahfouf, J.-F. and Rabier, F. 2000 The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics. *Q. J. R. Meteorol. Soc.*, **126,** 1171–1190

Miller, M., Hortal, M. and Jakob, C. 1995 'A major operational forecast model change'. Pp. 2–8 in ECMWF Newsletter No. 70. Available from ECMWF, Shinfield Park, Reading, UK

Montani, A., Thorpe, A. J., Buizza, R. and Undén, P. 1999 Forecast skill of the ECMWF model using targeted observations during FASTEX. *Q. J. R. Meteorol. Soc.*, **125,** 3219–3240

Morcrette, J.-J., Mlawer, E. J., Iacono, M. J. and Clough, S. A. 2001 'Impact of RRTM in the ECMWF forecast system'. In ECMWF Newsletter No. 91. Available from ECMWF, Shinfield Park, Reading, UK

Parrish, D. F. and Derber, J. C. 1992 The National Meteorological Center's Spectral Statistical-Interpolation analysis system. *Mon. Weather Rev.*, **120,** 1747–1763

Persson, A. 2000 'Synoptic–dynamic diagnosis of medium-range weather forecast systems'. Pp. 123–137 in Proceedings of 1999 ECMWF seminar on diagnosis of models and data assimilation systems. Available from ECMWF, Shinfield Park, Reading, UK

Prunet, P., Thépaut, J.-N. and Cassé, V. 1998 The information content of clear-sky IASI radiances and their potential for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **124,** 211–241

Rabier, F., Klinker, E., Courtier, P. and Hollingsworth, A. 1996 Sensitivity of forecast errors to initial conditions. *Q. J. R. Meteorol. Soc.*, **122,** 121–150

Reynolds, C. A., Webster, P. J. and Kalnay, E. 1994 Random error growth in NMC's global forecasts. *Mon. Weather Rev.*, **122,** 1281–1305

Ritchie, H., Temperton, C., Simmons, A. J., Hortal, M., Davies, T., Dent, D. and Hamrud, M. 1995 Implementation of the semi-Lagrangian method in a high resolution version of the ECMWF forecast model. *Mon. Weather Rev.*, **123,** 489–514

Rohn, M., Kelly, G. A. and Saunders, R. W. 2001 Impact of new cloud motion wind products from Meteosat on NWP analyses and forecasts. *Mon. Weather Rev.*, **129,** 2392–2403

Saunders, R. W., Matricardi, M. and Brunei, P. 1999 An improved fast radiative transfer model for assimilation of satellite radiance observations. *Q. J. R. Meteorol. Soc.*, **125,** 1407–1425

Savijärvi, H. 1995 Error growth in a large numerical forecast system. *Mon. Weather Rev.*, **123,** 212–221

Simmons, A. J. 1986 Numerical prediction: Some results from operational forecasting at ECMWF. *Adv. Geophys.*, **29,** 305–338

2000 'Objective verification of deterministic forecasts'. Pp. 385–404 in Proceedings of 1999 ECMWF seminar on diagnosis of models and data assimilation systems. Available from ECMWF, Shinfield Park, Reading, UK

Simmons, A. J. and Gibson, J. K (Eds.) 2000 'The ERA-40 Project Plan'. In ERA-40 Project Report Series No. 1. Available from ECMWF, Shinfield Park, Reading, UK

Simmons, A. J., Mureau, R. and Petroliagis, T. 1995 Error growth and predictability estimates for the ECMWF forecasting system. *Q. J. R. Meteorol. Soc.*, **121,** 1739–1771

Simmons, A. J., Andersson, E., Fisher, M., Jakob, C., Kelly, G. A., Lalaurette, F., McNally, A. P., Untch, A. and Viterbo, P. 2001 'Forecasting system performance in summer 1999. Part 2—Impact of changes to the ECMWF forecasting system'. ECMWF Tech. Memo. No. 322. Available from ECMWF, Shinfield Park, Reading, UK

Teixeira, J. 1999 'The impact of increased boundary layer resolution on the ECMWF forecast system'. ECMWF Tech. Memo. No. 268. Available from ECMWF, Shinfield Park, Reading, UK

Tomassini, M., LeMeur, D. and Saunders, R. W. 1998 Near-surface satellite wind observations of hurricanes and their impact on ECMWF model analyses and forecasts. *Mon. Weather Rev.*, **126,** 1274–1286

Tomassini, M., Kelly, G. A. and Saunders, R. W. 1999 Use and impact of satellite atmospheric motion winds on ECMWF analyses and forecasts. *Mon. Weather Rev.*, **127,** 971–986

Untch, A. and Simmons, A. J. 1999 'Increased stratospheric resolution in the ECMWF forecasting system'. Pp. 2–8 in ECMWF Newsletter No. 82. Available from ECMWF, Shinfield Park, Reading, UK

van den Hurk, B. J. J. M., Viterbo, P., Beljaars, A. C. M. and Betts, A. K. 2000 'Offline validation of the ERA40 surface scheme'. ECMWF Tech. Memo. No. 295. Available from ECMWF, Shinfield Park, Reading, UK

Wallace, J. M. and Gutzler, D. S. 1981 Teleconnections in the geopotential height field during the northern hemisphere winter. *Mon. Weather Rev.*, **109,** 785–812