TECHNICAL MEMORANDUM

322

Forecasting system performance in summer 1999. Part 2 Impact of changes to the ECMWF system

A.J. Simmons, E. Andersson, M. Fisher, C. Jakob, G.A. Kelly, F. Lalaurette, A.P. McNally, A. Untch and P. Viterbo

Research Department

September 2000

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

Forecasting system performance in summer 1999
Part 2: Impact of changes to the ECMWF system

by **Adrian Simmons, Erik Andersson, Mike Fisher, Christian Jakob, Graeme Kelly, François Lalaurette, Tony McNally, Agathe Untch and Pedro Viterbo**

Abstract

Objective verification of data assimilation and forecast experiments shows that cycle 21r4 of the ECMWF forecasting system, which was introduced into operations on 12 October 1999, gives much better forecasts for late spring and summer 1999 than were provided by the versions of the forecasting system that were operational at the time. This was a period when the operational forecasts from cycle 21r1 (up to 12 July) and cycle 21r2 (from 13 July onwards) were poor compared either with the forecasts in previous years from the ECMWF system or with the forecasts in 1999 from other centres. Two of the factors involved in the change from cycle 21r2 to 21r4, correction of the processing of dew-point observations and revision of the background-error statistics used in the data assimilation, are shown to contribute significantly to the improvement in performance for August 1999, and are related to earlier experimental results obtained for the summer of 1998. Further improvement in performance can be seen in results from cycle 22r3, which was introduced into operations on 27 June 2000.

1. Introduction

The performance of the operational ECMWF forecasting system was unusually poor in the late spring and summer months of 1999. This was identified both by the Centre's routine objective verification and synoptic analysis, and in reports received from Member-State forecasters. Objective scores for the Northern Hemisphere were low in early May, early July and the second half of August. Scores for Europe were exceptionally poor in the August period, and poor also in mid May. Problems were also seen in the forecasts from other operational centres, but not always to the same extent as in the ECMWF forecasts. Performance over Europe was also relatively poor in the same period of 1998.

This paper is one of a number that document and investigate aspects of forecast performance in summer 1999. The principal aim of this contribution is to record the extent to which recent major changes to the forecasting system give improved performance for the period, especially for August. For simplicity the presentation concentrates on a pair of standard forecast scores. There is little discussion of the other benefits of these changes to the forecasting system or of the impacts found in tests carried out for other seasons.

2. Variations in the quality of operational forecasts at the five-day range

Fig. 1 presents time series of running 14-day mean anomaly correlations of the day-5 operational 500hPa height forecasts made during 1999 by ECMWF, the Met. Office and Deutscher Wetterdienst (DWD). They are shown for Europe and for the extratropical Northern Hemisphere. For Europe, very poor scores in the second half of August are evident for ECMWF and to a lesser extent DWD. ECMWF scores drop below those of both other centres in mid-May also. The ECMWF forecasts are generally the most accurate in the first and last four months of the year, with a particularly large advantage in autumn.

Corresponding root-mean-square errors are shown in Fig. 2. These generally confirm the picture provided by anomaly correlations, but do not show ECMWF in as poor a light relative to the other centres at the times of low forecast quality. In particular, the root-mean-square errors of the forecasts for Europe from all centres are very large for several days at the end of August.

Fig. 3 shows anomaly correlations of the day-5 (high-resolution deterministic) ECMWF forecasts as in Fig. 1, together with the scores of the lower resolution control forecasts from the ECMWF ensemble prediction system (EPS). The scores from the two sets of forecasts generally follow each other closely, particularly around the time of exceptionally poor performance over Europe during the second half of August. The higher resolution model gives generally better scores over the first and last four months of the year, as it did in comparison with the forecasts from the Met. Office and DWD. Its scores are, however, slightly worse on average than those of the lower resolution EPS control forecast in the May to August period. It is tempting to ascribe this to the weaker synoptic activity of the lower resolution model, giving less rapid error growth at the times of major forecast failures, but the lower resolution forecasts also score better in late July over Europe, when the forecasts from both versions of the ECMWF model were relatively accurate, though still poorer than those of the Met. Office. Discussion of the performance of the EPS itself in 1999 is given in the companion paper ECMWF Technical Memorandum 324.

Fig. 4 places the results for late spring and summer 1999 in the context of the longer-term performance of the forecasting systems. It shows running four-month averages of day-5 anomaly correlations since 1991 for ECMWF and Met. Office forecasts, and since 1993 for DWD forecasts. The poor performance of the ECMWF system in spells from May to August 1999 is reflected in a sharp dip in the ECMWF curves in mid-1999, when the four-month mean score dropped to a level not seen since 1994, both for Europe and for the Northern Hemisphere. The ECMWF forecasts were also quite poor for Europe in summer 1998, but in contrast to 1999, scores for the Northern Hemisphere were relatively good in summer 1998, in fact at their highest summer level ever. The ECMWF, Met. Office and DWD systems were each upgraded in October 1999, and the scores of all three systems since then have been at a high overall level, the highest ever in the case of ECMWF and the Met. Office, both for Europe and for the Northern Hemisphere.

The unusual nature of 1999 is highlighted by individual monthly means of the day-5 anomaly correlations (not shown). One has to go back to August 1986 to find a month with a day-5 score for Europe as low as recorded for August 1999, yet the corresponding score for September 1999 is better than achieved in any previous September, and better than achieved in any other previous non-winter month. The version of the ECMWF forecasting system used operationally in September 1999 was the same as used in August 1999.

3. Forecasting-system changes

Results will be presented from a number of different versions (or cycles) of the ECMWF forecasting system. The baseline system is cycle 21r1, which was operational from 5 May to 12 July 1999, used with 50-level resolution for data assimilation and the deterministic forecast, and 31-level vertical resolution for the EPS. The changes introduced with subsequent cycles are specified below. Cycle 22r2 is omitted as it was created for convenience as an interim cycle between 22r1 and 22r3, and was not intended for operations. The list for cycle 22r3 refers to all changes made on top of cycle 22r1. Cycle 21r2 became operational on 13 July 1999, 21r4 on 12 October 1999, 22r1 on 11 April 2000 and 22r3 on 27 June 2000. Cycle 21r3 was run in pre-operational trials from 25 July to 3 September 1999, but was superseded by cycle 21r4, and never became operational.

Cycle 21r2:
- New physics/dynamics coupling
- Revised numerics for gravity-wave drag
- New soil-moisture and soil-temperature analysis
- Use of Meteosat high density winds with quality indicator
- Use of US wind-profiler data
- Use of targeted dropsonde data
- Radiosonde temperature bias correction
- Relaxation of quality control of TOVS/ATOVS radiances
- Use of actual ship anemometer height rather than a standard height
- Blacklisting of ship humidities
- Improvements in 1D–Var SSM/I retrieval
- Better specification of humidity background error

Cycle 21r3:
- Introduction of 60-level vertical resolution for the analysis and deterministic forecast, and 40-level vertical resolution for the EPS
- Set of modifications to the cloud and convection schemes
- New global orographies and associated subgrid fields
- Background error statistics derived using new method
- Use of 10m marine wind-speed retrievals from SSM/I radiances
- Revised bias correction of MSU and AMSU-A radiance data

Cycle 21r4:
- Corrected use of sonde and SYNOP humidity observations

Cycle 22r1:
- Better suppression of humidity increments in the stratosphere
- Revised SSM/I quality control, bias correction, thinning and use of second satellite
- Use of coastal ship and buoy winds in the extratropical Southern Hemisphere
- Relaxed quality control of dropsondes
- Set of minor changes to wave model and analysis
- Limit on stratospheric tendency from gravity-wave drag parametrization
- Bugfix for calculation of clear-sky precipitation fraction
- Bugfix for (diagnostic) stratocumulus scheme used in calculation of low-resolution trajectory

Cycle 22r3:
- New parametrization schemes for the land surface, lying snow and sea-ice
- New (RRTM) long-wave radiation scheme
- Improved ozone model
- Improved treatment of precipitation processes in first timestep
- Use of more TOVS/ATOVS data (HIRS-12, AMSU-14; less constraint on AMSU-8; more off-nadir data)
- Use of actual buoy heights
- Revised snow analysis
- Revised observation and background error variances
- Use of digital filter for J_c

Cycles 22r1 and 22r3 also differ from earlier cycles in that they are based on cycle 22, a cycle produced in common with Météo-France. Merging code changes from ECMWF and Météo-France to produce a common cycle generally introduces small numerical differences which means that results from the immediate preceding cycle (21r4 in this case) are not exactly reproducible. Also, in making cycle 22 a minor change to the dynamical configuration used operationally at ECMWF was introduced inadvertently. This was corrected in cycle 22r3.

Many individual changes can be seen to have gone into these new cycles of the forecasting system, and it is clearly impractical to investigate the impact on the forecasts for summer 1999 of each change in isolation, let alone that of different combinations of changes. Each individual change was the subject of preparatory testing to the extent thought necessary given the nature of the change. For those changes anticipated to have a significant affect on overall forecast performance, the testing typically included several weeks or more of data assimilation and forecasts. This mainly had to be carried out for quite recent periods (for the reason discussed in section 7), although use could be made of the prototype ERA-40 assimilation system to test some model and surface-analysis changes over one or more years from the 1980s.

4. Overall impact of cycles 21r2 and 21r4

Tests showing the improvement of cycle 21r2 over the operational cycle 21r1 are available from 7 May until 12 July 1999, after which cycle 21r2 was introduced into operations. Tests of cycle 21r4 are also available for this period. Fig. 5 shows anomaly correlations of 500hPa height for these three versions of the forecasting system. The correlations are plotted as functions of forecast range out to day 7 for Europe and the extratropical Northern Hemisphere, averaged over the whole period. A progressive improvement in going from 21r1 to 21r2 and then to 21r4 can be seen, with more difference between 21r4 and 21r2 than between 21r2 and 21r1.

Table 1 shows the results of standard tests of statistical significance applied to the differences at day 5, for the comparison of 21r2 and 21r1, the comparison of 21r4 and 21r1, and for other comparisons for which mean scores are presented later in this paper. An entry is made for any difference which is found to be significant at the 10, 5, 2, 1, 0.5, 0.2 or 0.1% level. The improvement of 21r4 over 21r1 is identified as being of strong significance, especially over the hemisphere; the smaller improvement of 21r2 over 21r1 is less significant.

Comparison	Period	Europe		Northern Hemisphere	
		Sign test	t test	Sign test	t test
21r2 better than *21r1*	7 May - 12 Jul 1999				5%
21r4 better than *21r1*	7 May - 12 Jul 1999		2%	0.1%	0.1%
21r4 better than *21r1*	13 - 26 May 1999		10%		
21r4 better than *oper*	6 May - 11 Oct 1999		1%		0.1%
21r4 better than *21r2*	7 May - 11 Oct 1999		2%		0.2%
21r4 better than *21r2*	1 - 31 Aug 1999	0.2%	0.1%	0.2%	0.1%
21r4 better than *21r3*	1 - 31 Aug 1999		10%	2%	0.5%
21r3 better than *21r2*	1 - 31 Aug 1999	0.2%	0.2%		5%
21r2+str. fn. better than *21r2*	15 - 31 Aug 1999	1%	2%	1%	0.2%
Suppressed q an better than *oper*	1 - 21 Aug 1998	10%	10%		5%
L50 stat1 better than *L31 oper*	1 - 31 Aug 1998	0.2%	0.1%		2%
L50 stat1 better than *L31oper*	14 May - 16 Sep 1998		0.5%		0.1%
L50 stat2 better than *L31 oper*	1 - 31 Aug 1998				5%
22r3 better than *21r2*	1 - 31 Aug 1999	1%	0.1%	0.2%	0.1%
22r3 better than *21r4*	1 - 31 Aug 1999				10%
22r1 worse than *21r4*	1 - 31 Aug 1999		10%		10%
22r3 better than *22r1*	1 - 31 Aug 1999	5%	10%	1%	0.1%
22r3 better than *oper*	1 Mar - 26 June 2000				0.5%

Table 1: Levels of statistical significance of differences in day-five 500hPa height anomaly correlations

Fig. 6 compares mean anomaly correlations from 21r1, 21r2 and 21r4 for the period from 13 to 26 May. This is the period for which the 14-day mean European scores at day 5 shown in Fig. 1 are at a minimum. Also included in the figure are the corresponding means from the Met. Office and DWD forecasts. Cycle 21r2 gives some improvement over 21r1, making the ECMWF scores for Europe similar to those of DWD. The improvement of 21r4 over 21r1 is substantial, and sufficient to raise the European scores at days 5 and 6 to the levels reached by the Met. Office forecasts. The 21r4 forecasts are nevertheless still poorer than normal over this two-week period.

Cycle 21r4 was tested over the 159-day period from 6 May to 11 October 1999, after which it became operational. Fig. 7 shows mean anomaly correlations for the whole period, together with those from ECMWF operations (a mixture of cycle 21r1 and cycle 21r2), and from the operational forecasts of the Met. Office and DWD. Over this period of almost six months, the operational ECMWF forecasts were better than those of the Met. Office over the Northern Hemisphere, but were slightly worse beyond day 3 over Europe. The forecasts from cycle 21r4 can be seen to be superior to all others in the set.

Fig. 8 shows 500hPa height anomaly correlations computed for Europe comparing results from cycles 21r4 and 21r2 averaged separately for the months of June, July, August and September. They show a substantial variation in the impact of the change from cycle 21r2 to 21r4. Cycle 21r4 improves substantially over 21r2 in August, when operational performance for Europe was unusually poor, but gives worse results than 21r2 in September, when operational performance was unusually good. The new cycle performs better in June, and depending on the forecast range has either neutral or slightly negative impact in July. It is rare for ECMWF (and other operational centres) to test forecasting-system changes over so long a period as here, and this example indicates that very extensive and computationally demanding experimentation may be needed to quantify reliably the extent to which a major change to a forecasting system changes forecast accuracy over Europe. From Table 1 it can be inferred that an impact of high statistical significance in one month does not necessarily imply a similarly significant impact in other months. This is not necessarily a failing of the tests themselves, as a change to a forecasting system may be beneficial in the synoptic regime characteristic of a particular month, but not in a different regime characteristic of another month. An impact judged to be of high statistical significance in an experiment carried out for a particular period can be expected to be reproduced in an experiment over a second period in which a similar synoptic situation (and data coverage) prevails, but may not be found over a period in which the synoptic situation is markedly different.

Hemispheric scores provide a more reliable indication of whether a change increases overall forecast accuracy, although results for a particular month may still not be indicative of those for a larger sample. Fig. 9 shows the impact of the change from 21r2 to 21r4 on monthly-mean anomaly correlations for the Northern Hemisphere. Impact is positive in July as well as June and August. It is largest in August, and essentially neutral in September.

5. Impact of cycle 21r4 for August 1999

Fig. 10 shows the mean anomaly correlations for August from 21r4 and 21r2 forecasts together with the corresponding correlations from the operational forecasts of the Met. Office and DWD. Over Europe, the mean scores from the operational ECMWF (21r2) forecasts are poorer than those of both the Met. Office and DWD beyond day 4, and the corresponding scores for the Northern Hemisphere show little advantage to ECMWF operations beyond day 5. Cycle 21r4 of the ECMWF forecasting system gives mean scores that are better than those from the other centres for Europe up to day 5, though still somewhat worse by day 6. The advantage of 21r4 over the hemispheric domain is clear-cut.

Additional results are available to identify the contributions of some particular components of cycle 21r4 to the improvement in performance for August 1999. As mentioned earlier, cycle 21r3 was originally intended to be the next operational cycle after 21r2, and was run in parallel to operations in August 1999 ahead of the testing of cycle 21r4. The results from the three cycles are compared in Fig. 11. Cycle 21r3 can be seen to improve on 21r2, and the change to cycle 21r4 brings clear further improvement. The change from 21r2 to 21r3 gives the larger impact over Europe, and for the Northern Hemisphere up to day 5. At day 5 the statistical tests indicate greater significance of the change from 21r2 to 21r3 over Europe, but greater significance of the change from 21r3 to 21r4 over the Northern Hemisphere.

Cycle 21r4 differs from 21r3 in only one respect. It corrects a long-standing error in observation processing in the ECMWF system whereby reports of dew-point temperatures from radiosondes and surface measurements were erroneously processed for temperatures below $0^{o}C$ as if they were observations of either a frost point or a parametrized mixed-phase saturation point. The error was such as to produce analyses with a dry bias, and contributed to a tendency for the humidity field to "spin up" to moister values as the forecast proceeded. It was corrected following a query from the Met. Office, but had been known of for some time in other Member States, and it is not entirely clear why the chain of communication did not lead to it being corrected earlier at ECMWF. It seems likely that this was because the error was not expected to have a significant effect on forecast performance. In fact, the improvement seen in going from cycle 21r3 to 21r4 in Fig. 11 indicates that the error had a quite significant effect on forecast quality for August 1999.

There is additional evidence which suggests that this erroneous use of humidity observations may have had a more general detrimental impact on summer forecast quality. The performance of the ECMWF forecasting system was relatively poor over Europe (though not the Northern Hemisphere) in the summer of 1998 as well as 1999, as has been illustrated in Fig. 4. One of the investigations carried out at the time involved an experiment in which the humidity analysis was carried out as normal (to minimize impact on the analysis of other variables) but the resulting analysed humidity field was replaced by the first guess field, thus allowing the specific humidity to evolve freely through data assimilation cycles as is generally the case for model cloud fields. The experiment was run for the first three weeks of August 1998, and the impact on the mean 500hPa anomaly correlations for Europe and the Northern Hemisphere is shown in Fig. 12. Suppression of the humidity analysis resulted in a clear improvement in medium-range forecast scores for both regions. The erroneous interpretation of dew-point measurements may not have been the factor principally responsible, but this is clearly a possibility in the light of the results for August 1999.

One of the important changes introduced in cycle 21r3 was the use of a new method of computing background-error statistics based on an ensemble of analyses using perturbed observations and the EPS's representation of stochastic physics in the background forecasts. This was known from earlier tests to give better forecasts, and following the success of cycles 21r3 and 21r4 in improving the forecasts for August 1999, an assimilation for 15-31 August was carried out using cycle 21r2 plus the new background-error statistics of cycle 21r3. The resulting impact on standard scores is shown in Fig. 13. Use of the new statistics is beneficial for both Europe and the Northern Hemisphere. Its accounts for a substantial part of the improvement brought by 21r3 over the hemisphere, and rather less of the improvement over Europe.

Fig. 13 also shows the results of running the 21r3 (or equivalently the 21r4) version of the forecast model starting from the operational analyses produced running cycle 21r2 in the data assimilation. It shows that the model changes in cycle 21r3 also directly account for some of the improvement in forecast scores, though not as much as derives from the structure-function change. The model changes may also have brought an indirect improvement in the forecasts from the 21r3 system through their use in data assimilation to produce the 21r3 analyses.

Experimental tests of the 50-level vertical resolution (in operational use in summer 1999) were carried out over the summer of 1998, and provide further evidence of the sensitivity of summer forecasts to the specification of the background-error statistics. Fig. 14 presents 500hPa anomaly correlations for Europe and the Northern Hemisphere from an initial trial version of the 50-level system (labelled *L50 stat1*) and from the operational 31-level system, averaged over August 1998. The forecasts from the initial 50-level system score substantially better than those from the 31-level system, especially for Europe. These August results are drawn from an extended experiment run from May to September 1998, and substantial variability was found from month to month in the impact over Europe, much as shown in Fig. 8 for the comparison of cycles 21r4 and 21r2. Again, this is reflected in the statistical significance of the improvement due to the 50-level system being higher for August than for the longer period.

The impact of the initial 50-level system shown in Fig. 14 was much larger than expected from an increase in stratospheric resolution alone. Suspicion fell on the background-error statistics that had to be computed anew for the 50-level data assimilation. The so-called "NCEP" method used at the time to derive these was based on sets of differences between two- and one-day forecasts. The first set of 50-level forecasts used for these calculations was made from initial conditions formed by merging 31-level ECMWF analyses with Met. Office UARS stratospheric analyses. Implied upper-level errors had to be reduced to counter effects of incompatibilities between the ECMWF model and the Met. Office analyses. These background-error statistics were used for two periods of data assimilation and forecasts, and a revised set of statistics was then computed from these 50-level forecasts.

Forecasts for August and the first 23 days of September 1998 were repeated starting from analyses produced using the revised background-error statistics. Hemispheric verification scores were very slightly better than those from the original 50-level assimilation when averaged over the whole period, but there were some marked regional differences. In particular, almost all of the improvement due to the 50-level system disappeared over Europe with the revised background statistics, whereas the forecasts for North America were improved. Scores for Europe and the Northern Hemisphere averaged over August 1998 are included in Fig. 14 (labelled *L50 stat2*). In contrast, the best scores for the period from 1 to 23 September came from the assimilation with the revised background statistics (not shown). This sensitivity to background statistics was not pursued further at the time, as it was decided to concentrate effort on refining the 60-level system, and the new method of deriving the necessary statistics based on perturbed data assimilations was subsequently developed. The operational 50-level assimilation in the summer of 1999 used the revised background statistics computed using the NCEP method.

6. Impact of cycles 22r1 and 22r3 for August 1999

Fig. 15 shows anomaly correlations of 500hPa height averaged for August 1999 from cycles 21r2, 21r4, 22r1 and 22r3. There is a decline in scores in going from cycle 21r4 to 22r1, but this is reversed by the change from 22r1 to 22r3, the latter cycle giving mean scores very close to those from 21r4 for Europe. For the Northern Hemisphere, cycle 22r3 is distinctly better than cycle 21r4 in the mean, although the statistical significance of its improvement is not especially high.

Further information relating to the performance of cycle 22r1 for Europe is given in Fig. 16. In September 1999 it was discovered that the operational data extraction during August had by mistake supplied the data assimilation system with duplicate, but not bit-identical, copies of the MSU data from the NOAA-14 satellite. The consequence of this was, in effect, to give more weight to the MSU data than was intended. The duplicate data were used for all the experiments based on variants of cycle 21 reported above, but were suppressed for the experiments using cycles 22r1 and 22r3. This does not appear, however, to account for 22r1 giving poorer results than 21r4, as rerunning much of August using cycle 21r2 but without the duplicate data gave slightly better mean scores for Europe, as can be seen by comparing the curves labelled *21r2* and *21r2n* in the upper panel of Fig. 16. Note, however, that the small difference at day 5 does not show up as significant at the 10% level. Also included in the upper panel of Fig. 16 is a rerun based on cycle 21r4 but including the improvement to the stratospheric humidity analysis incorporated in cycle 22r1. This component of 22r1 improves the mean European scores for this period, an improvement that at day 5 registers as significant at the 10% level according to the sign test, though corresponding Northern Hemisphere scores (not shown) worsen, a result significant at the 5% level according to the t test for day 5.

A test of cycle 22r1 run over winter and early Spring 2000 prior to operational implementation did not reproduce the negative result found for August 1999. It gave 500hPa scores for Europe almost identical on average to those from cycle 21r4, as illustrated in the lower panel of Fig. 16. Close inspection reveals in fact a slight improvement with 22r1 at day 4, significant at the 10% level according to the t test, and a slight degradation at day 6 which does not register as significant at the 10% level.

Examination of other verification scores for August 1999 indicate a general superiority of cycle 22r3. Anomaly correlations and root-mean-square errors tell essentially the same story, so here only some more results obtained for anomaly correlations are presented. Fig. 17 shows the scores of 1000hPa height forecasts for Europe and the Northern Hemisphere from cycles 21r2, 21r4, 22r1 and 22r3. It provides a picture largely similar to that presented in Fig. 15 for 500hPa, but shows an advantage of 22r3 over 21r4 for Europe.

Fig. 18 shows 500hPa anomaly correlations for North America and the North Atlantic, this time from cycles 22r3, 21r4 and 21r2 of the ECMWF system and from the Met. Office operational forecasts. ECMWF's operational forecasts for North America were also very poor in August 1999, and for this region cycle 21r4 does not improve on 21r2, both cycles giving worse mean results than the Met. Office operational forecasts at days 4 and 5. The forecasts from cycle 22r3 are distinctly better than the others shown. Over the North Atlantic, 21r4 improves substantially over 21r2 and performs at about the same average level as the then-operational Met. Office system. Cycle 22r3 gives a further improvement.

Fig. 19 compares a set of scores for two regions where the operational (21r2) ECMWF forecasts scored much better than those of the Met. Office in August 1999. The regions are the North Pacific and the Southern Hemisphere. Cycle 21r4 improves on 21r2 for both regions. Cycle 22r3 gives much better results than 21r4 for the Southern Hemisphere. It is better than 21r4 for the North Pacific up to day 4 and poorer thereafter, though still better than 21r2.

7. Concluding remarks

The changes made to the operational ECMWF forecasting system with the introductions of cycles 21r2 and 21r4 have been shown by extensive numerical experimentation to improve substantially the performance of the forecasting system in late spring and summer 1999, a time when the operational performance of cycle 21r1 (up to 12 July) and cycle 21r2 (from 13 July onwards) was poor compared either with the performance of ECMWF in previous years or with the performance of other forecasting centres in 1999. Two of the factors involved in the change from 21r2 to 21r4, correction of the processing of dew-point observations and revision of the background-error statistics used in the data assimilation, have been shown to contribute significantly to the improvement in performance for August 1999, and have been related to earlier experimental results obtained for the summer of 1998. In addition, some direct improvement due to the model changes in 21r3 has been illustrated. Operational performance in August 1999 may also have been degraded very slightly by inadvertent supply of duplicate MSU data to the assimilation.

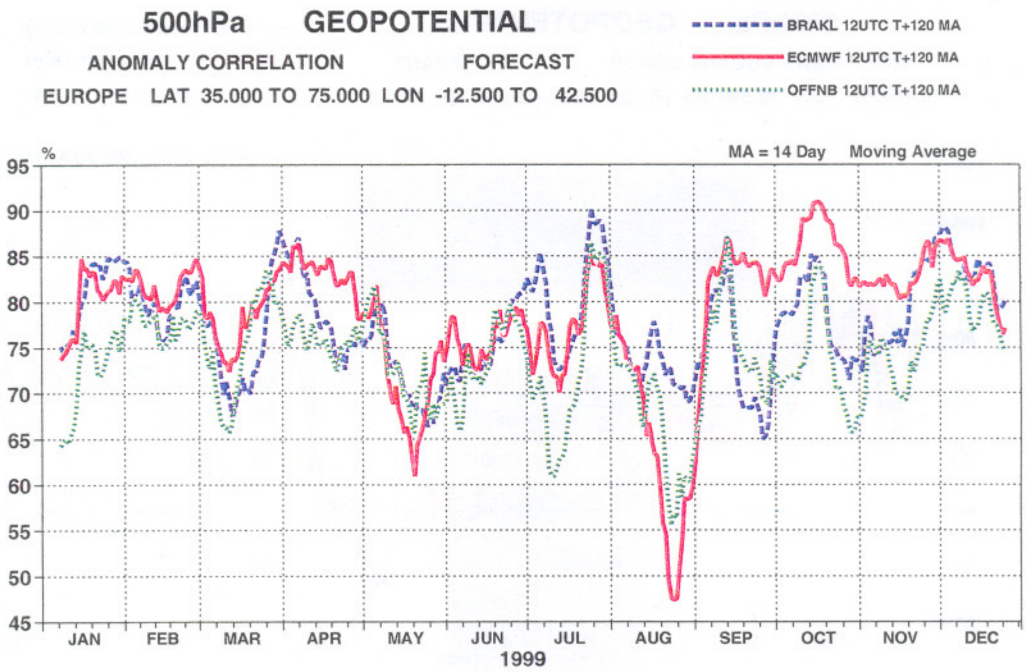
Comprehensive diagnostic examination of quite how the European forecasts benefit from the major changes made in moving from cycle 21r2 to cycle 21r4 is beyond the scope of this particular study. Synoptic analysis has shown that some large improvements in medium-range forecasts for Europe during August 1999 can be traced back to small analysis differences over the Canadian Arctic. Fig. 20 shows that the 21r4 and 21r2 analyses of 500hPa height differ predominantly over parts of the Arctic and Siberia, and that over the Arctic the change from 21r2 to 21r4 tends to bring the ECMWF analyses closer to those of the Met. Office. The monthly-mean 500hPa height field included in Fig. 20 indicates a mean flow which runs southward from the Canadian Arctic and turns eastward to cross Northern Europe. This flow pattern is consistent with the medium-range forecasts for Northern Europe being sensitive to analysis quality over the Canadian Arctic. Further diagnosis and discussion is given in the companion paper ECMWF Technical Memorandum 322..

Cycle 22r1 gives poorer mean scores than cycle 21r4 in the troposphere over Europe and the Northern Hemisphere for August 1999, although this is not seen for a longer winter period. Also, although not illustrated here, cycle 22r1 gives better forecasts at low levels in the Tropics and Southern Hemisphere and performs generally better in the stratosphere. The reduction in some mean forecast scores for August 1999 in moving from 21r4 to 22r1 is compensated by improvements in moving on to cycle 22r3, which according to many measures provides the best performance achieved yet for the month. The subsequent pre-operational trial of cycle 22r3 has confirmed the promising results found for August 1999. This is illustrated in Fig. 21, which compares scores from cycle 22r3 with those from ECMWF operations (a mix of cycles 21r4 and 22r1) and from the operational Met. Office and DWD systems, averaged for the period from 1 March to 26 June 2000.

In addition to results for 1999, rather poor performance of the ECMWF operational forecasting system over Europe has been illustrated for the summer of 1998, and the sensitivity of forecasts for August 1998 to the analysis of humidity and specification of background-error statistics has been discussed. In view of this it would clearly be of interest to run cycles 21r4, 22r1 and 22r3 of the forecasting system for August 1998 to complement the results presented here for August 1999. This is not straightforward, however. Recent cycles of the operational forecasting system have been developed to use the raw radiances from the MSU instrument on the NOAA-14 satellite and from the AMSU-A instrument on NOAA-15. The latter data have been received regularly only since late August 1998. Operations in August 1998 used processed radiances from the MSU, HIRS and SSU instruments on NOAA-11 and NOAA-14. The raw radiance assimilation system has recently been adapted to utilize the HIRS and SSU radiances for ERA-40, making it now possible to carry out a sensible test of the latest cycle of the forecasting system for August 1998.

Acknowledgements

Many people at ECMWF contributed to the various changes whose net effects have been documented in this paper. The listed authors are those principally involved in the running or verification of one or more of the test assimilations and forecasts for which results have been presented.



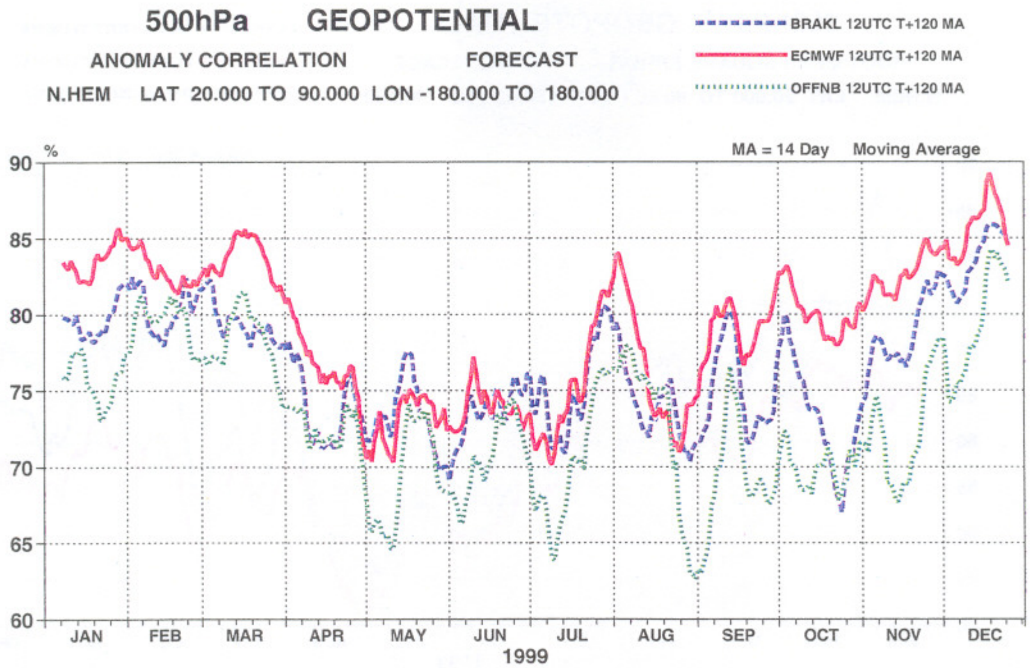
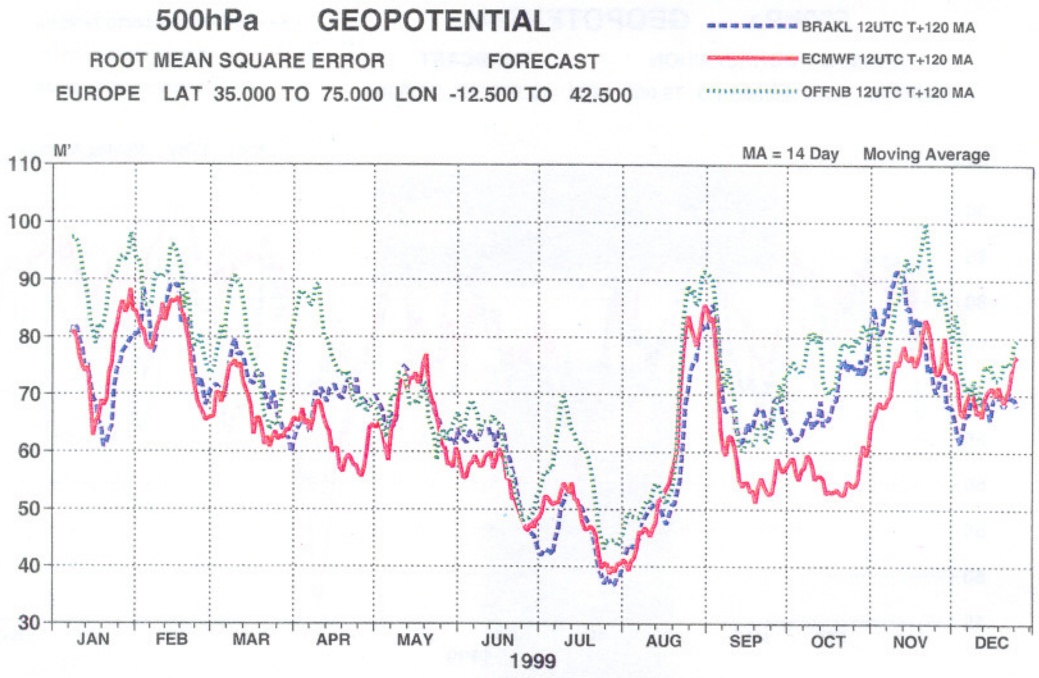


Fig. 1. Time series of anomaly correlations of day-5 operational 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower), from ECMWF (red solid), the Met. Office (blue dashed) and DWD (green dotted). Daily scores for 1999 are smoothed by applying a 14-day running mean.



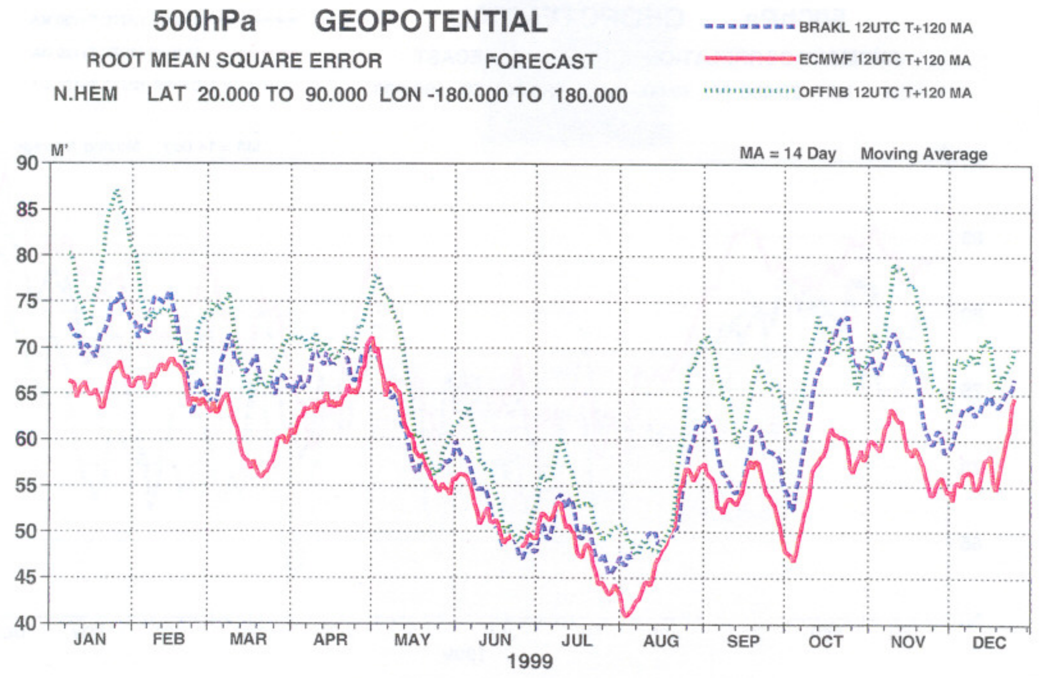
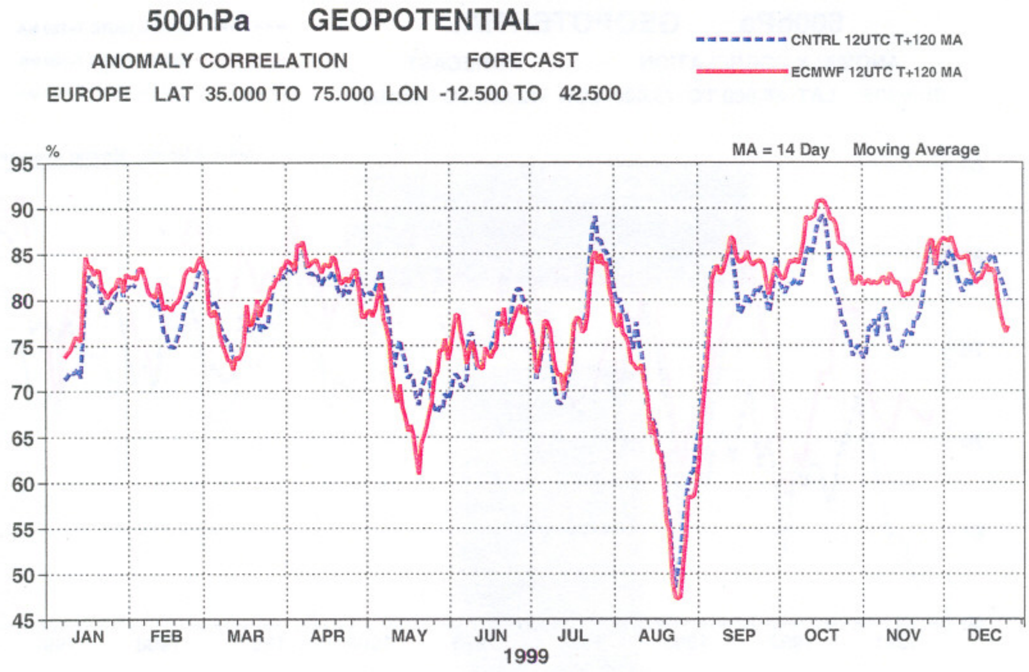


Fig. 2. Time series of root-mean-square errors of day-5 operational 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower), from ECMWF (red solid), the Met. Office (blue dashed) and DWD (green dotted). Daily scores for 1999 are smoothed by applying a 14-day running mean.



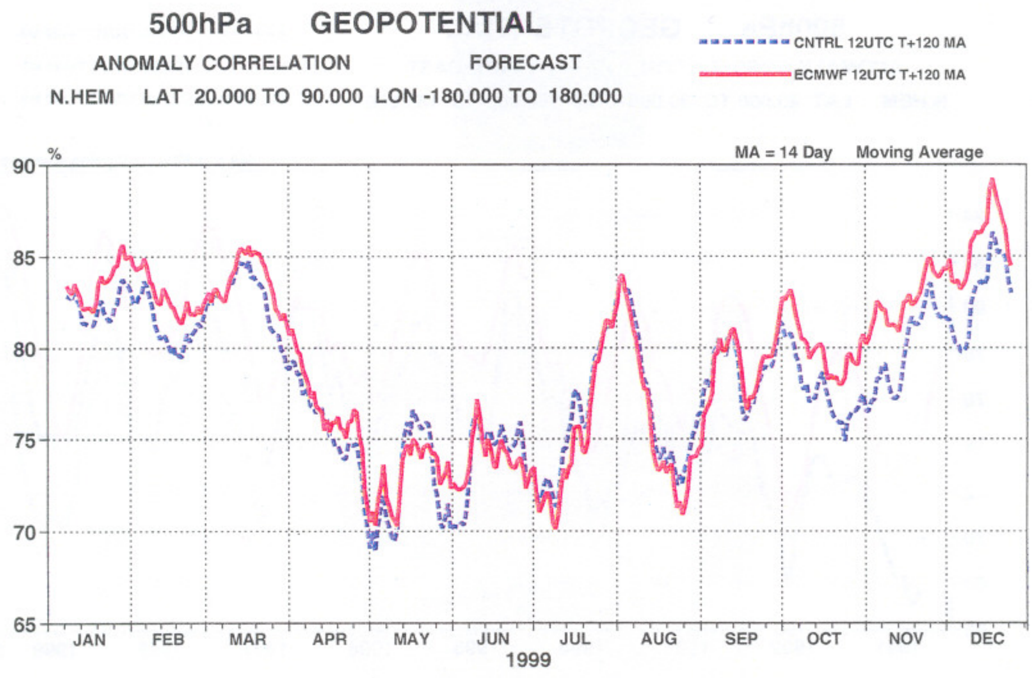
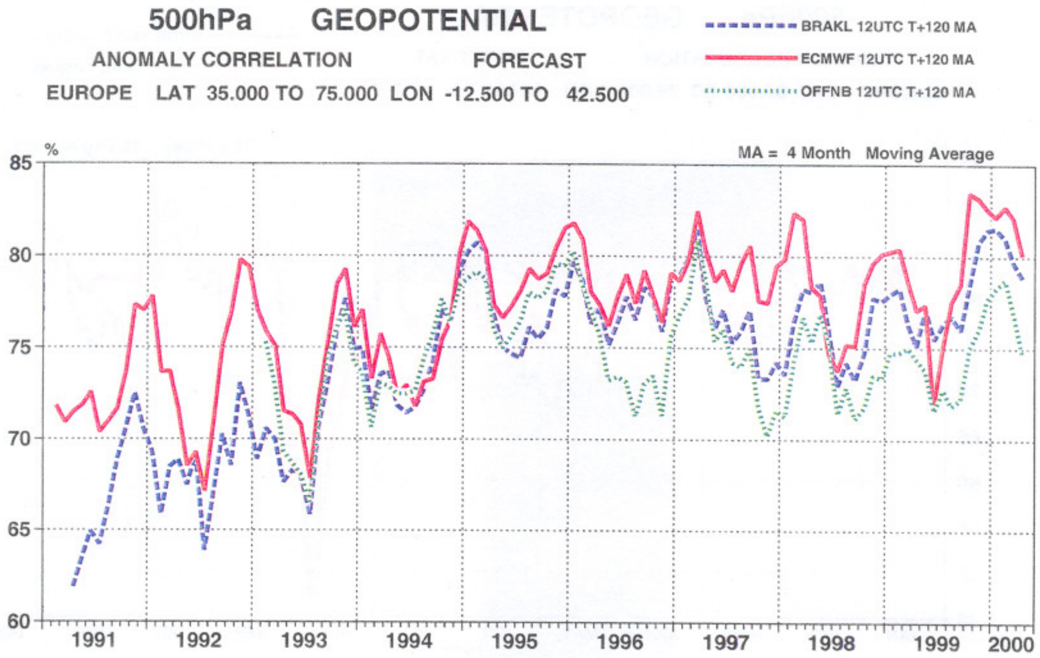


Fig. 3. Time series of anomaly correlations of day-5 operational 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower), from the high-resolution deterministic ECMWF forecasts (red solid) and the lower resolution ECMWF EPS control forecasts (blue dashed). Daily scores for 1999 are smoothed by applying a 14-day running mean.



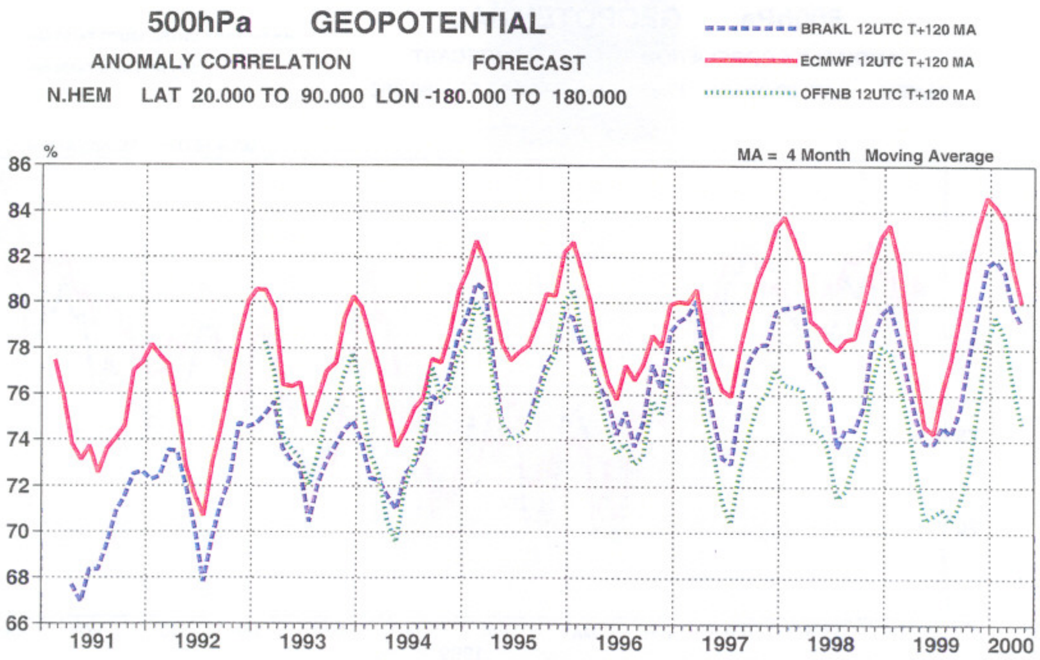
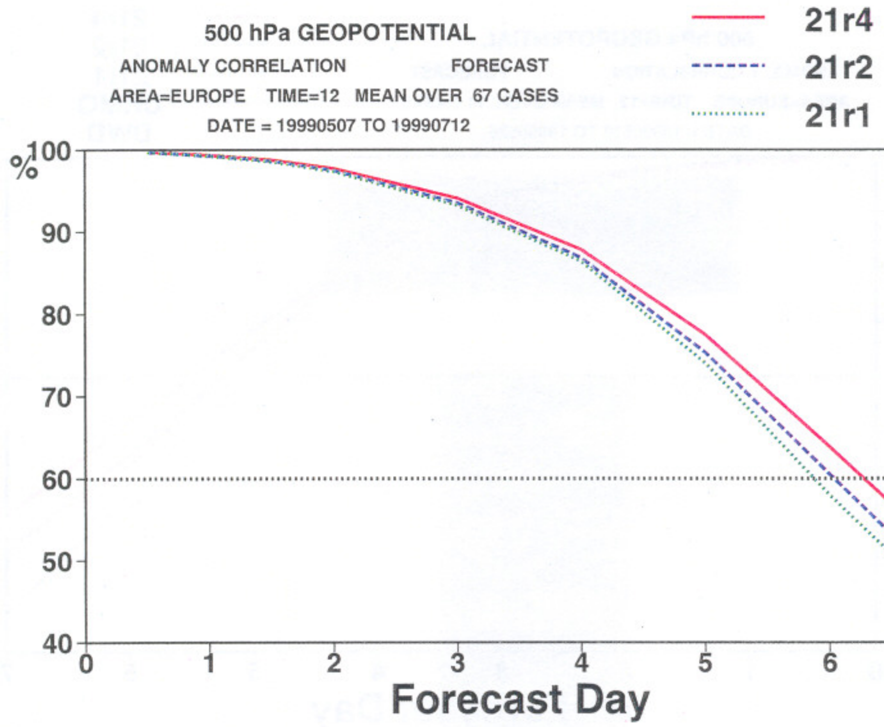


Fig. 4. Time series of anomaly correlations of day-5 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower), from ECMWF (red solid), the Met. Office (blue dashed) and DWD (green dotted). Monthly mean scores available from 1991 onwards are smoothed by applying a 4-month running mean.



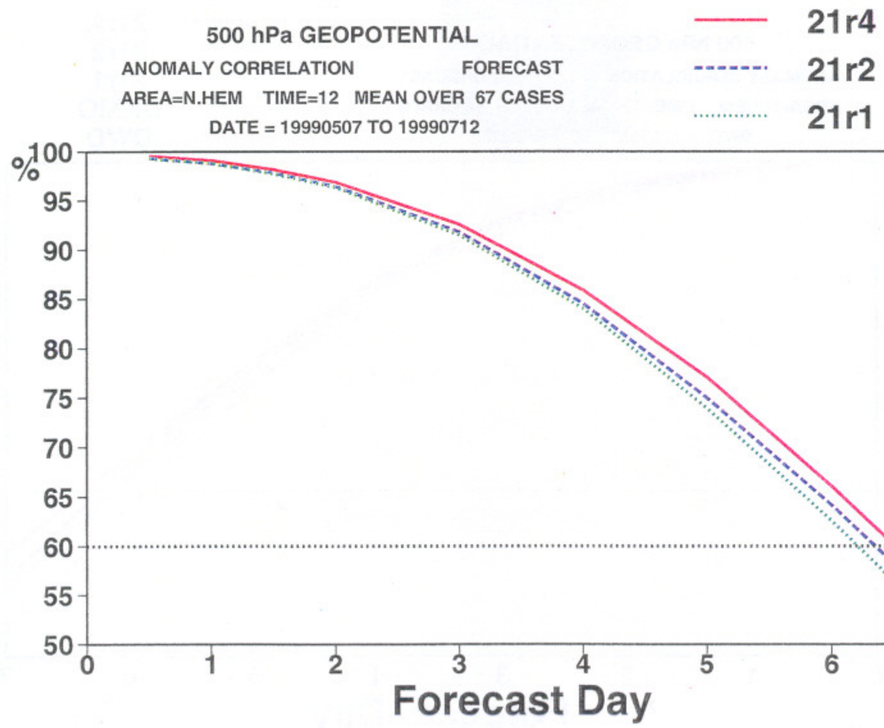
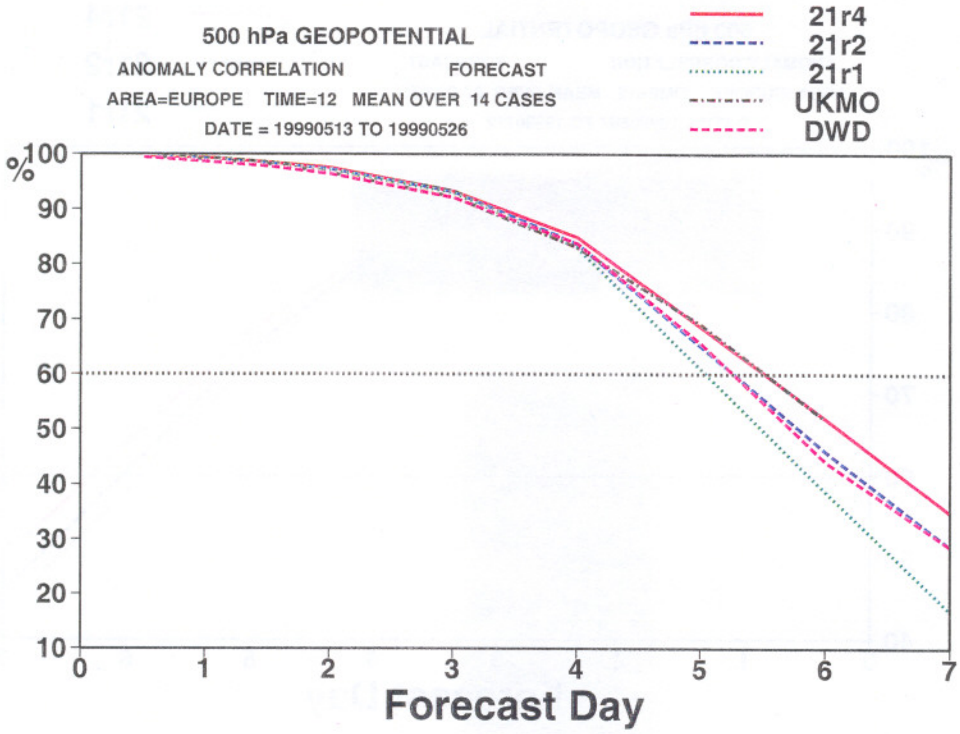


Fig. 5. Anomaly correlations of 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from cycles 21r4 (red solid) and 21r2 (blue dashed) of the ECMWF forecasting system and from ECMWF operations (cycle 21r1, green dotted), averaged over the period from 7 May to 12 July 1999.



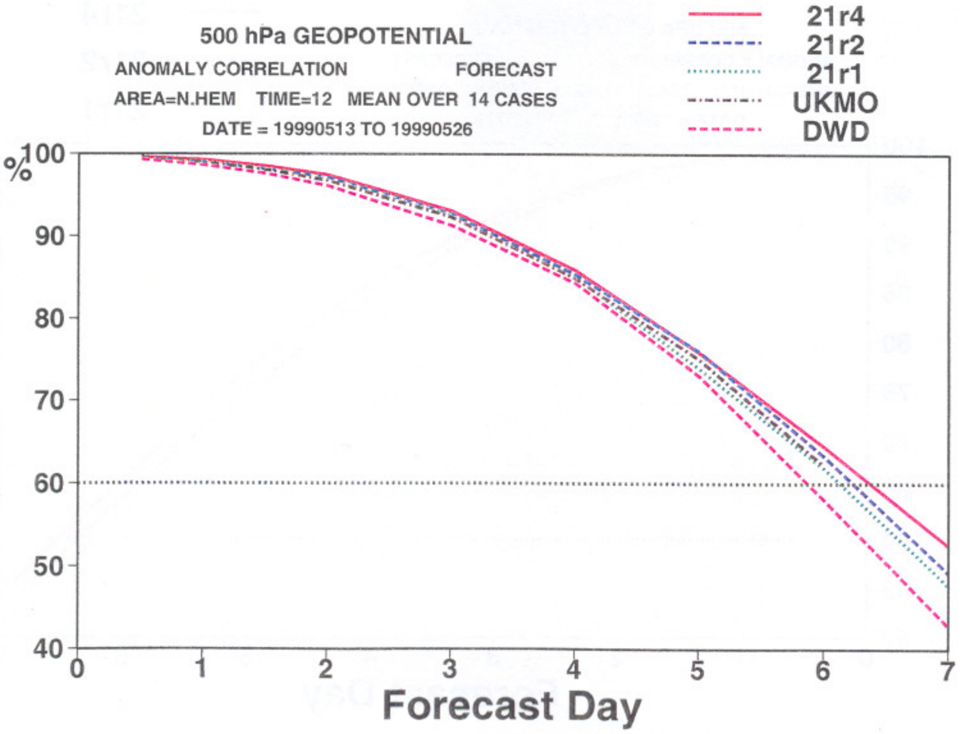
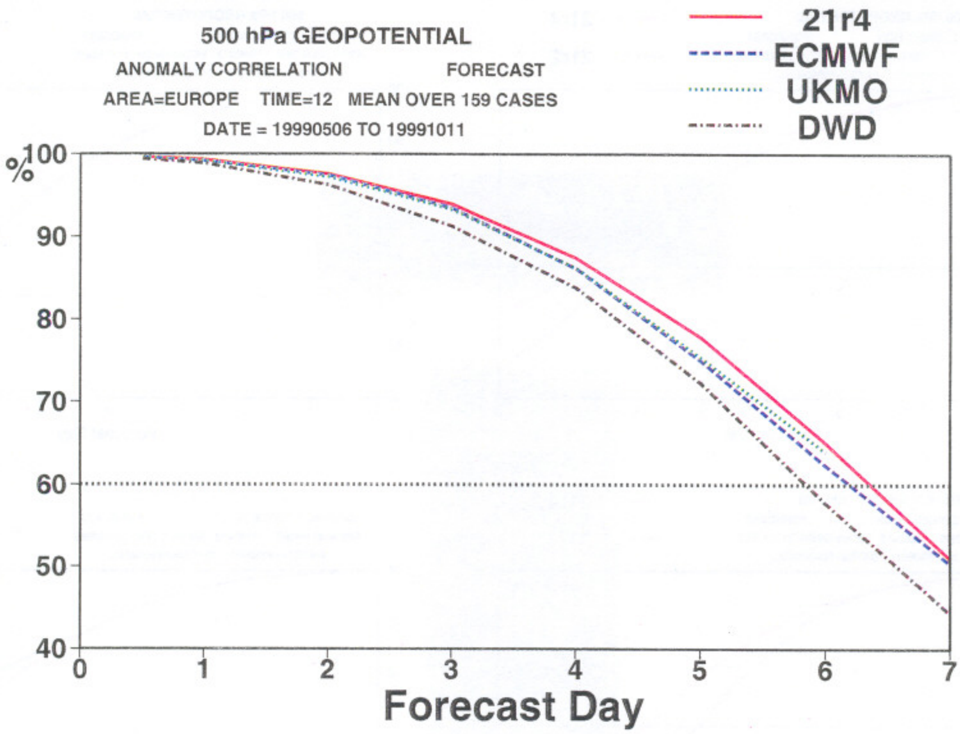


Fig. 6. Anomaly correlations of 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from cycles 21r4 (red solid) and 21r2 (blue dashed) of the ECMWF forecasting system, from ECMWF operations (cycle 21r1, green dotted), and from the operational forecasts of the Met. Office (brown chained) and DWD (magenta dashed), averaged over the period from 13 to 26 May 1999.



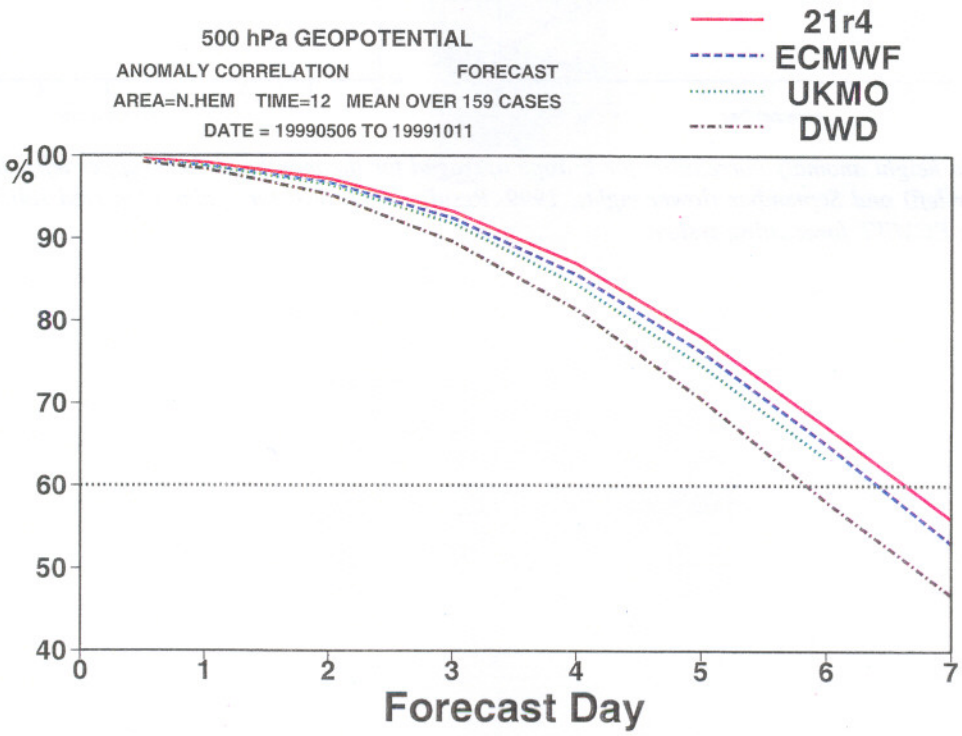


Fig. 7. Anomaly correlations of 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from cycle 21r4 of the ECMWF forecasting system (red solid), ECMWF operations (blue dashed), the Met. Office (green dotted) and DWD (brown chained), averaged over the period from 6 May to 11 October 1999.

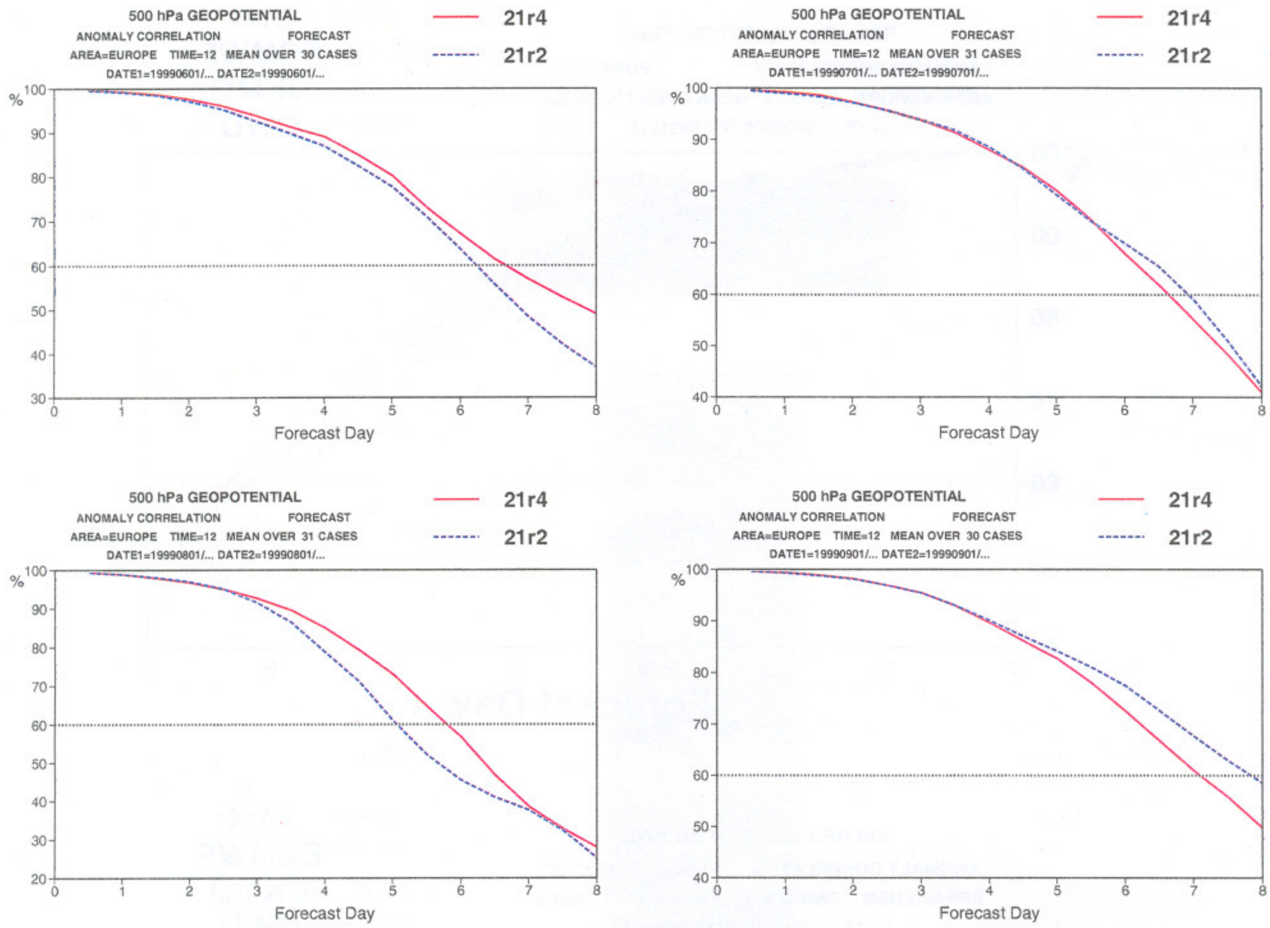


Fig. 8. 500hPa height anomaly correlation for Europe averaged for the months of June (upper left), July (upper right), August (lower left) and September (lower right), 1999. Results are shown for cycles 21r4 (red solid) and 21r2 (blue dashed) of the ECMWF forecasting system.

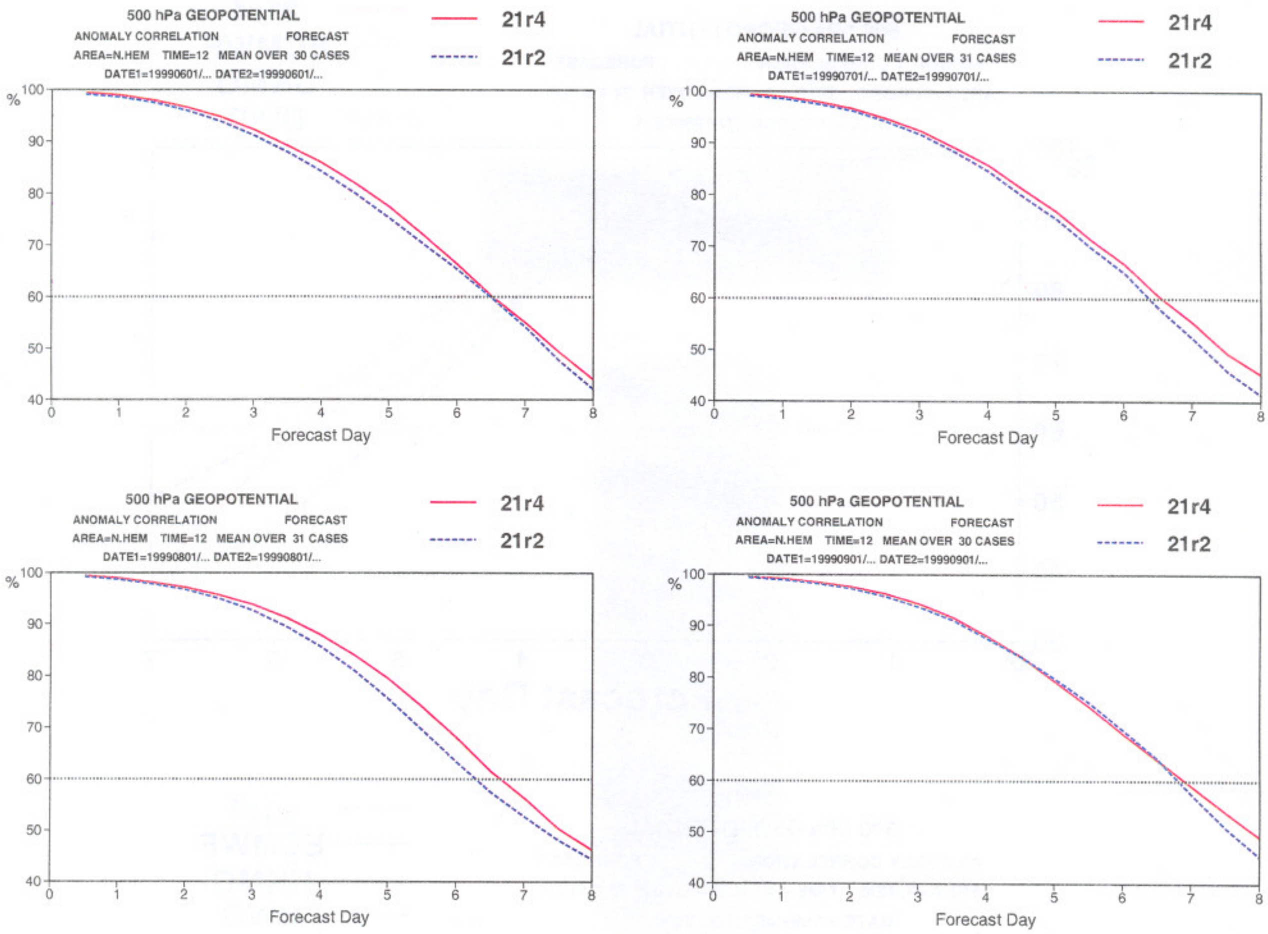
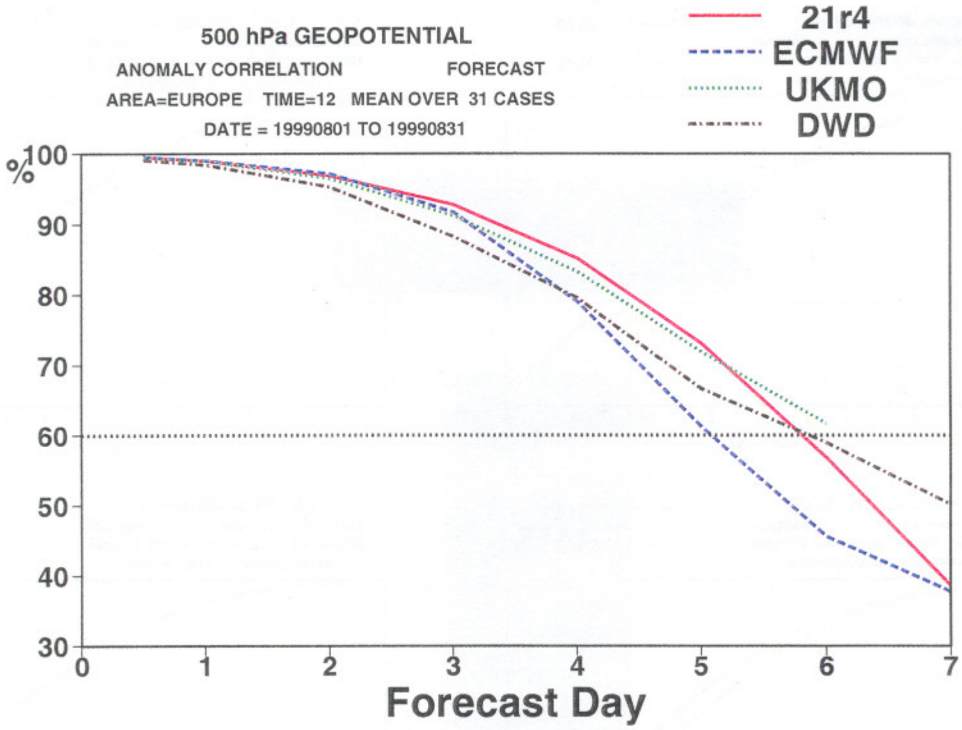


Fig. 9. 500hPa height anomaly correlation for the Northern Hemisphere averaged for the months of June (upper left), July (upper right), August (lower left) and September (lower right), 1999. Results are shown for cycles 21r4 (red solid) and 21r2 (blue dashed) of the ECMWF forecasting system.



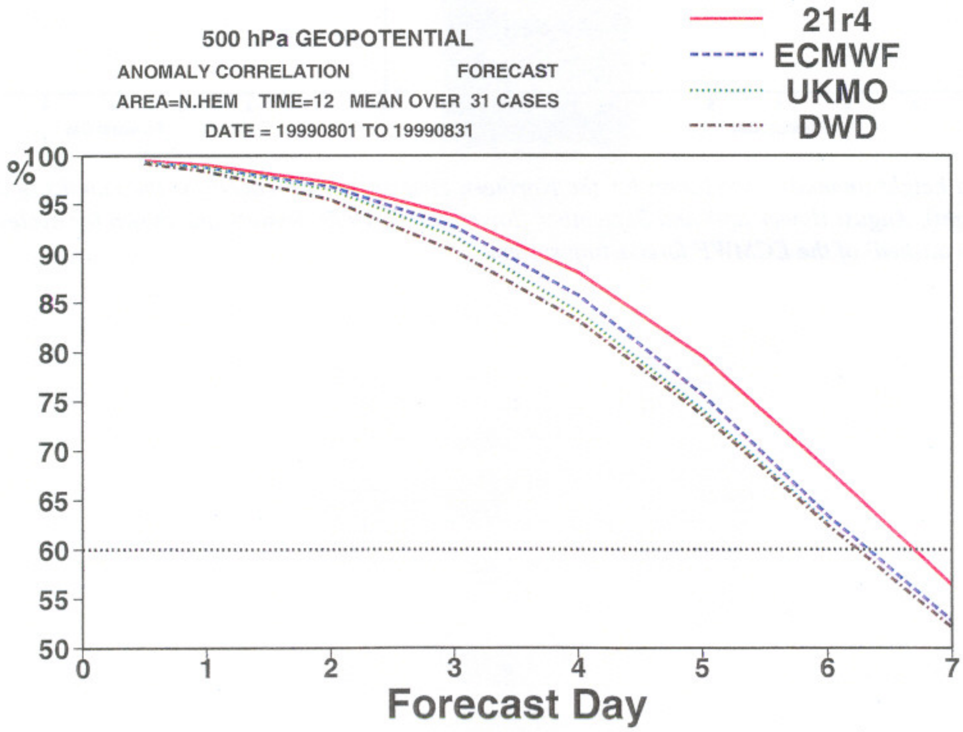
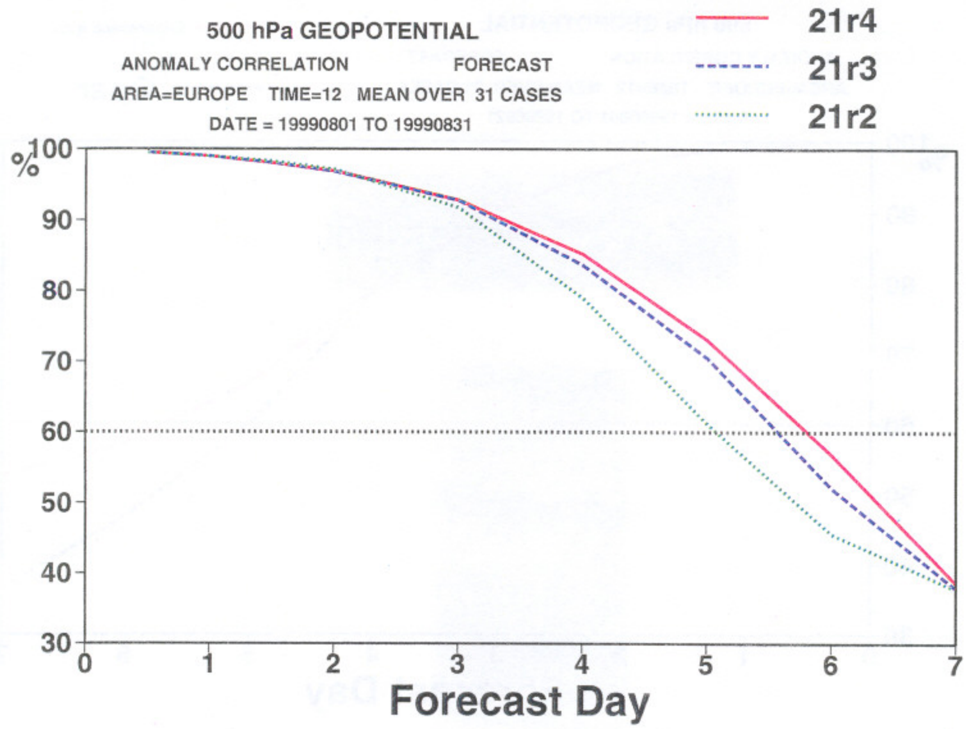


Fig. 10. Anomaly correlations of 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from cycle 21r4 of the ECMWF forecasting system (red solid), ECMWF operations (21r2, blue dashed), the Met. Office (green dotted) and DWD (brown chained), averaged for August 1999.



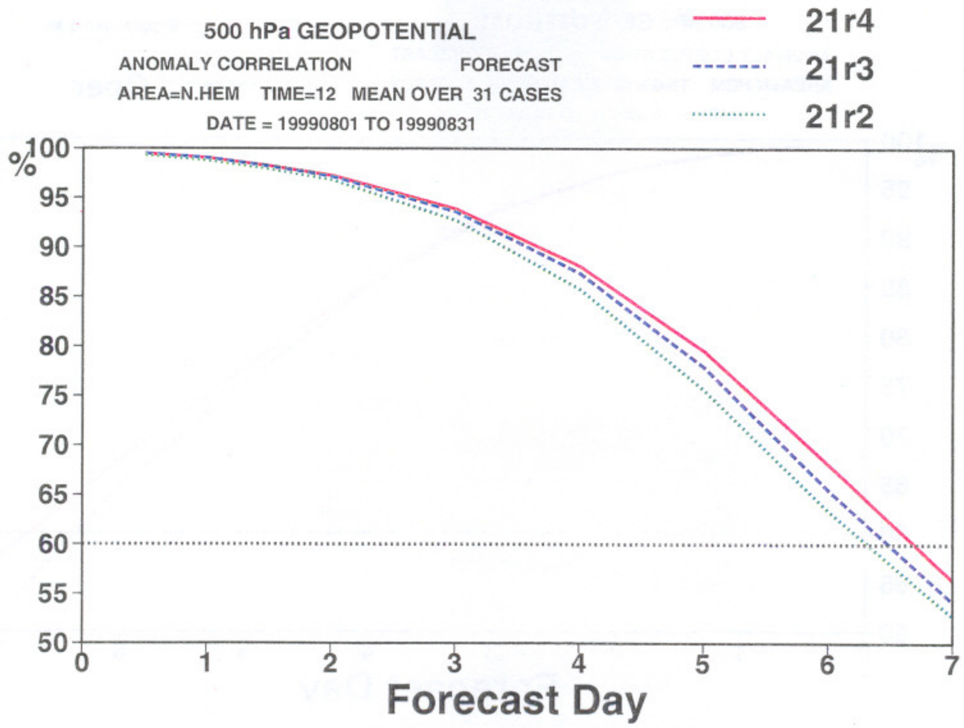
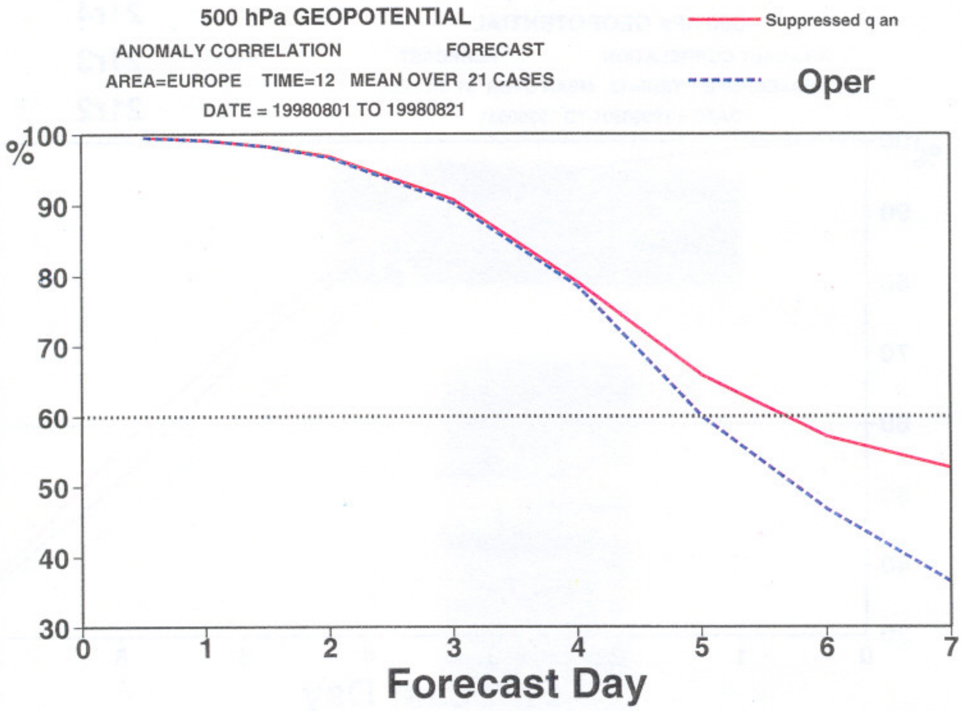


Fig. 11. Anomaly correlations of 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from cycles 21r4 (red solid), 21r3 (blue dashed) and 21r2 (operations, green dotted), averaged for August 1999.



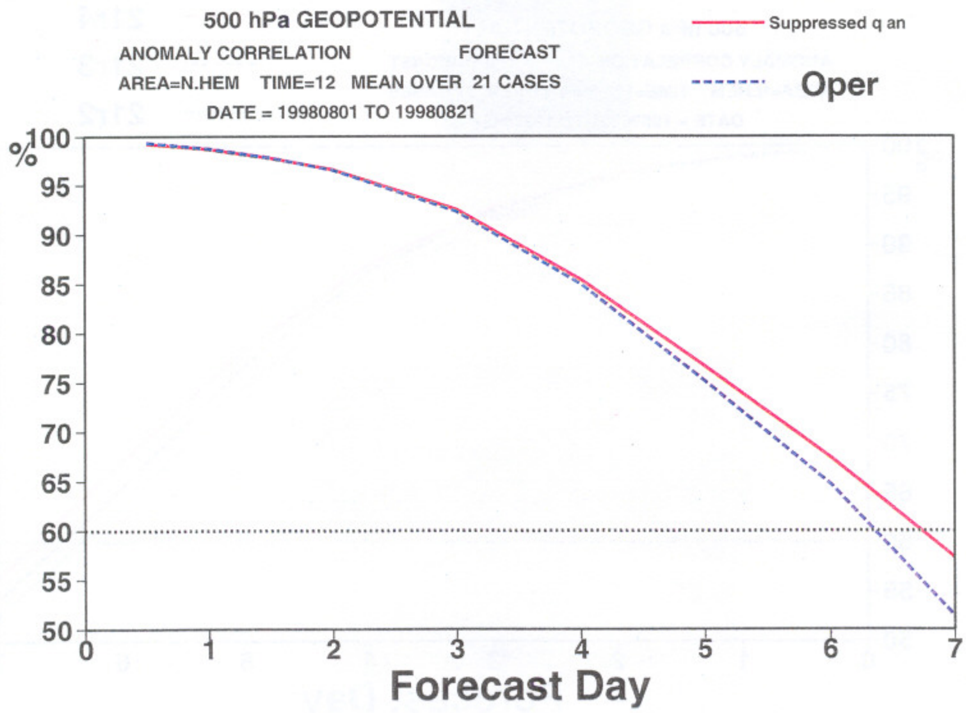
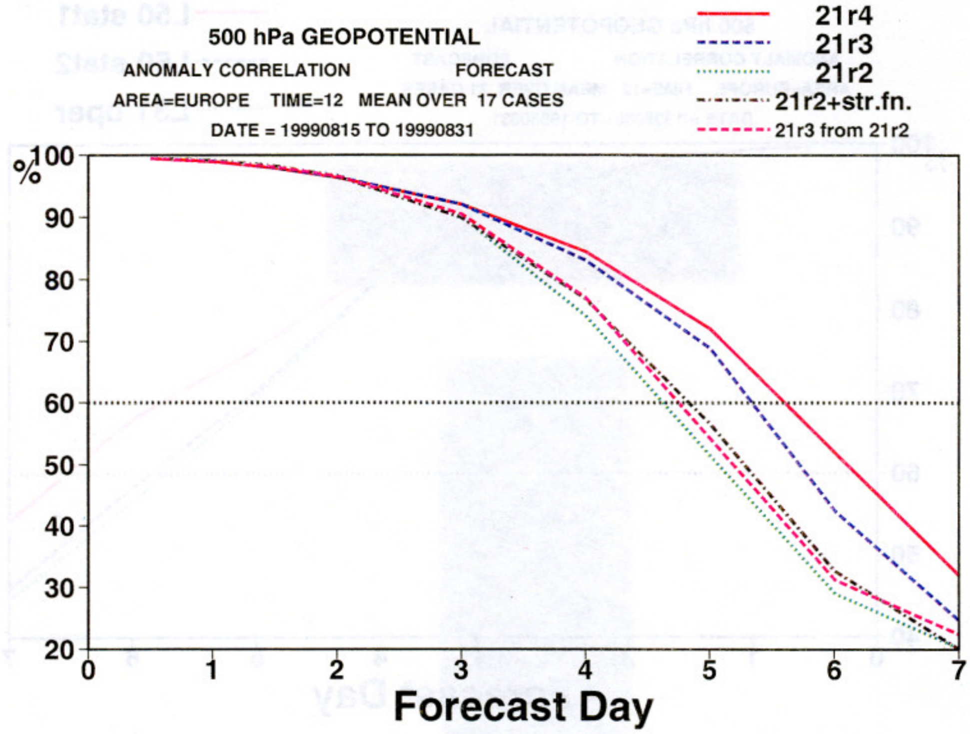


Fig. 12. Anomaly correlations of 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from an assimilation with suppression of the humidity analysis (red solid) and the corresponding operational control (blue dashed), averaged for 1- 21 August 1998.



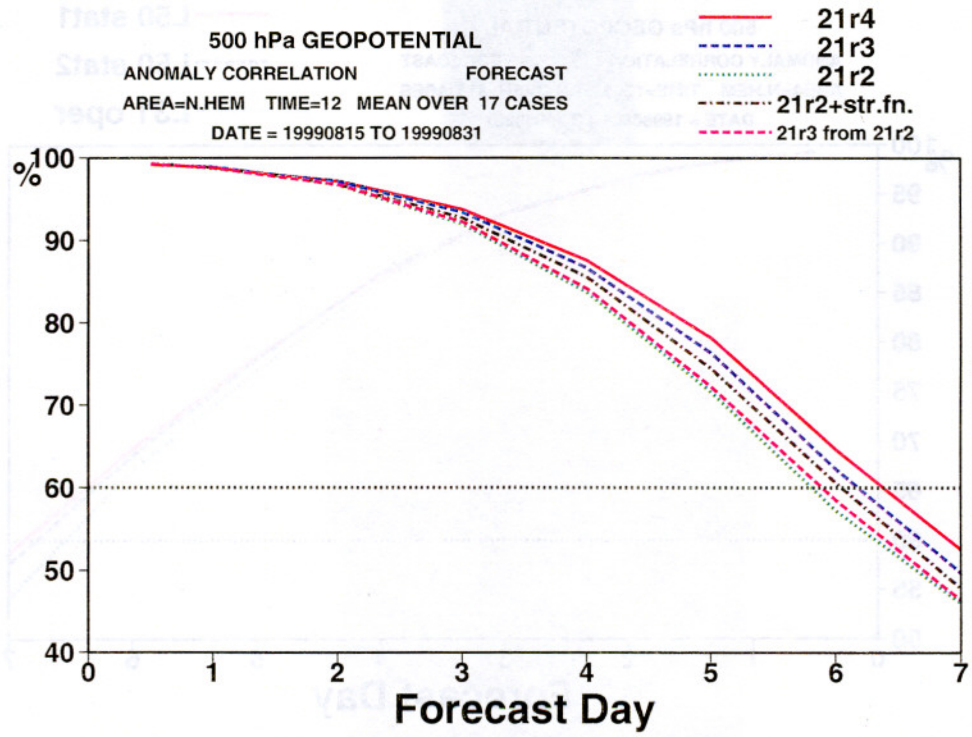
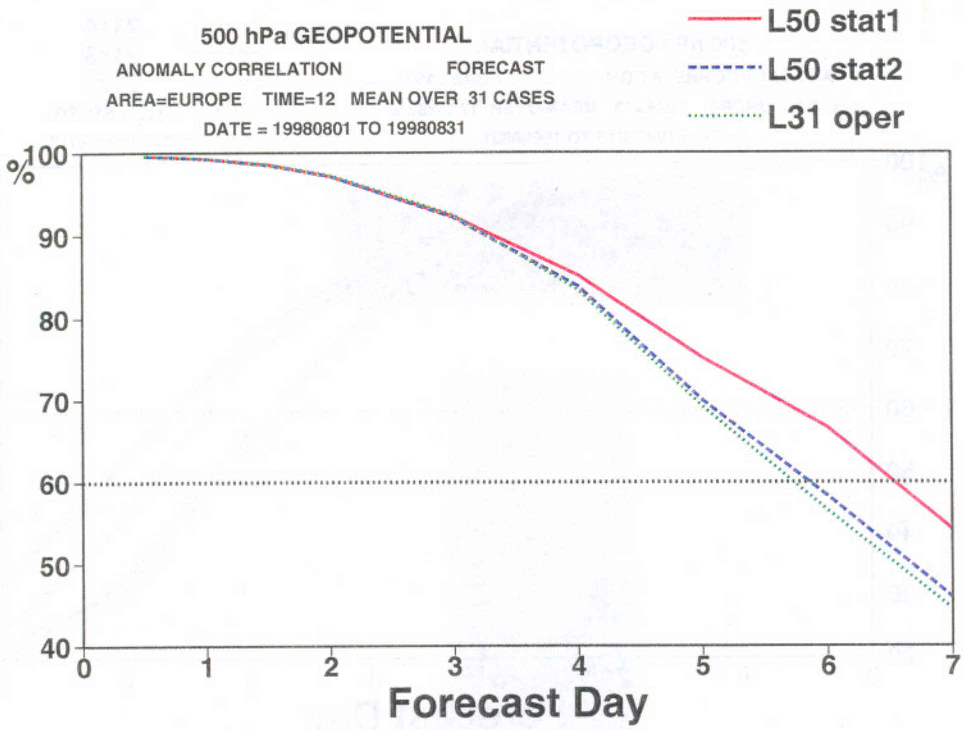


Fig. 13. Anomaly correlations of 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from cycles 21r4 (red solid), 21r3 (blue dashed) and 21r2 (operations, green dotted), and from 21r2 but with new structure functions (brown dashed), and from 21r3 forecasts run from 21r2 analyses (magenta dashed), averaged for 15-31 August 1999.



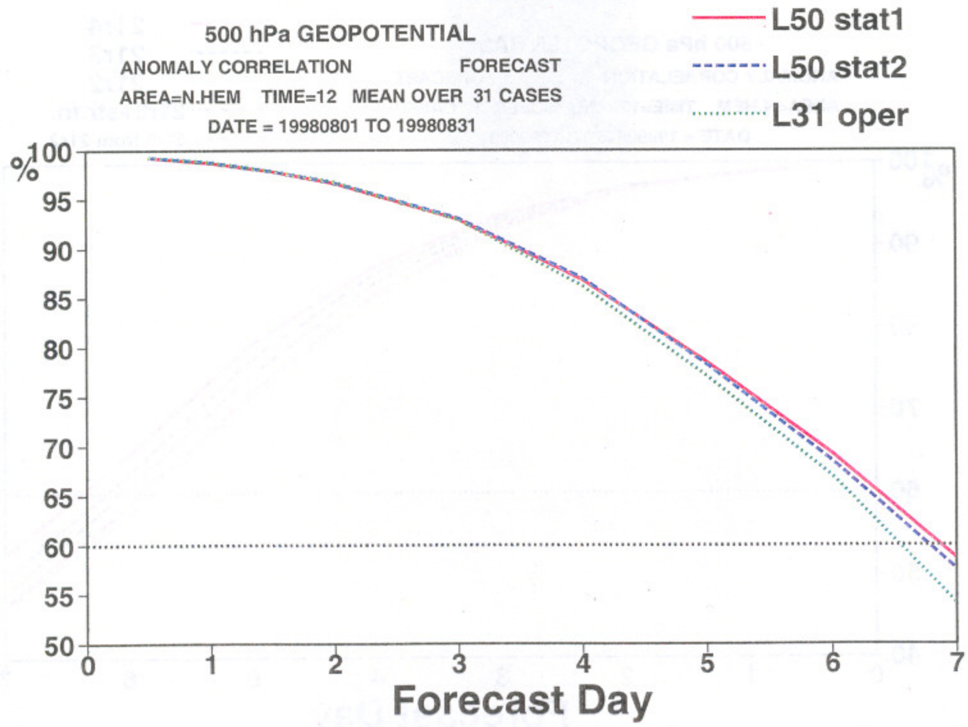
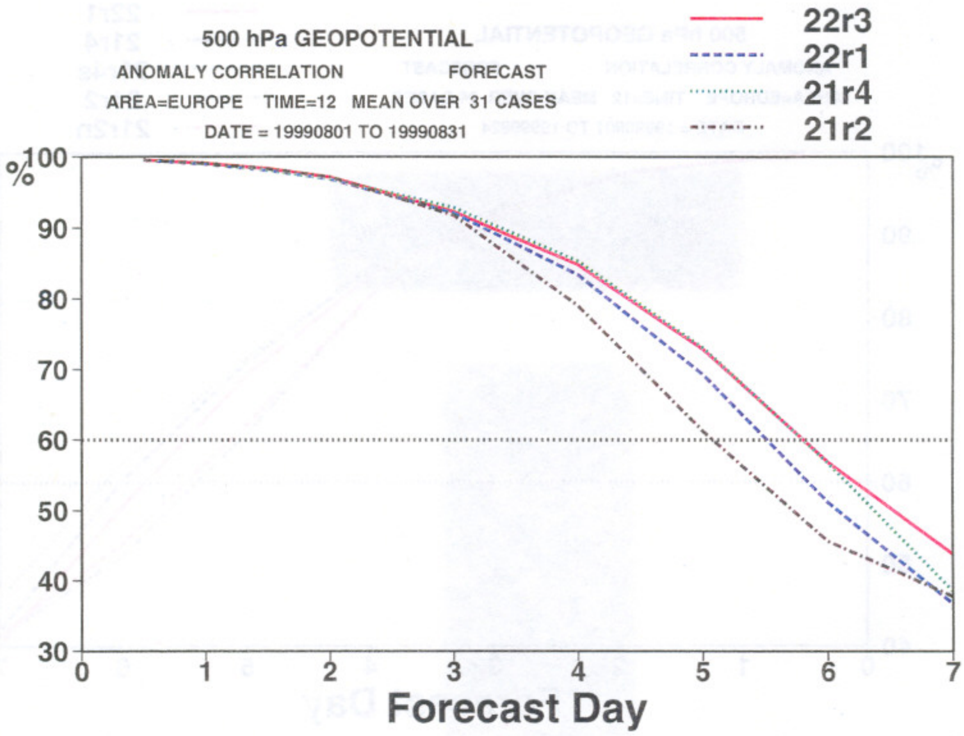


Fig. 14. Anomaly correlations of 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from test assimilations using 50-level vertical resolution with initial background-error statistics (red solid) and revised background-error statistics (blue dashed), and from the corresponding 31-level operational control (green dotted), averaged for August1998.



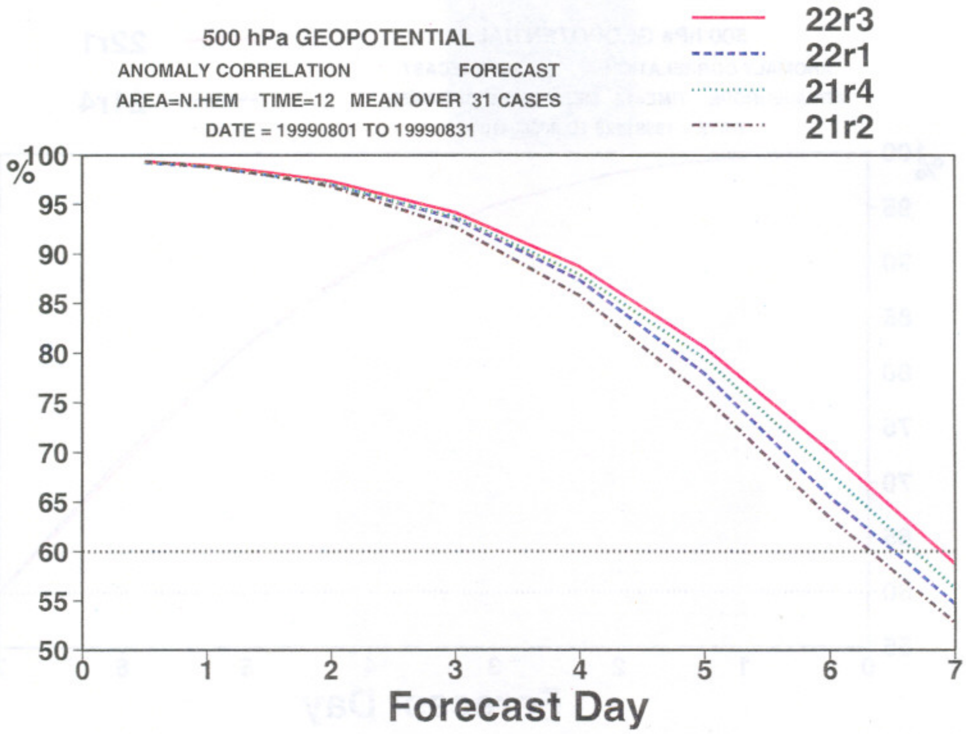


Fig. 15. Anomaly correlations of 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from cycles 22r3 (red solid), 22r1 (blue dashed), 21r4 (green dotted) and 21r2 (operations, brown chained), averaged for August 1999.

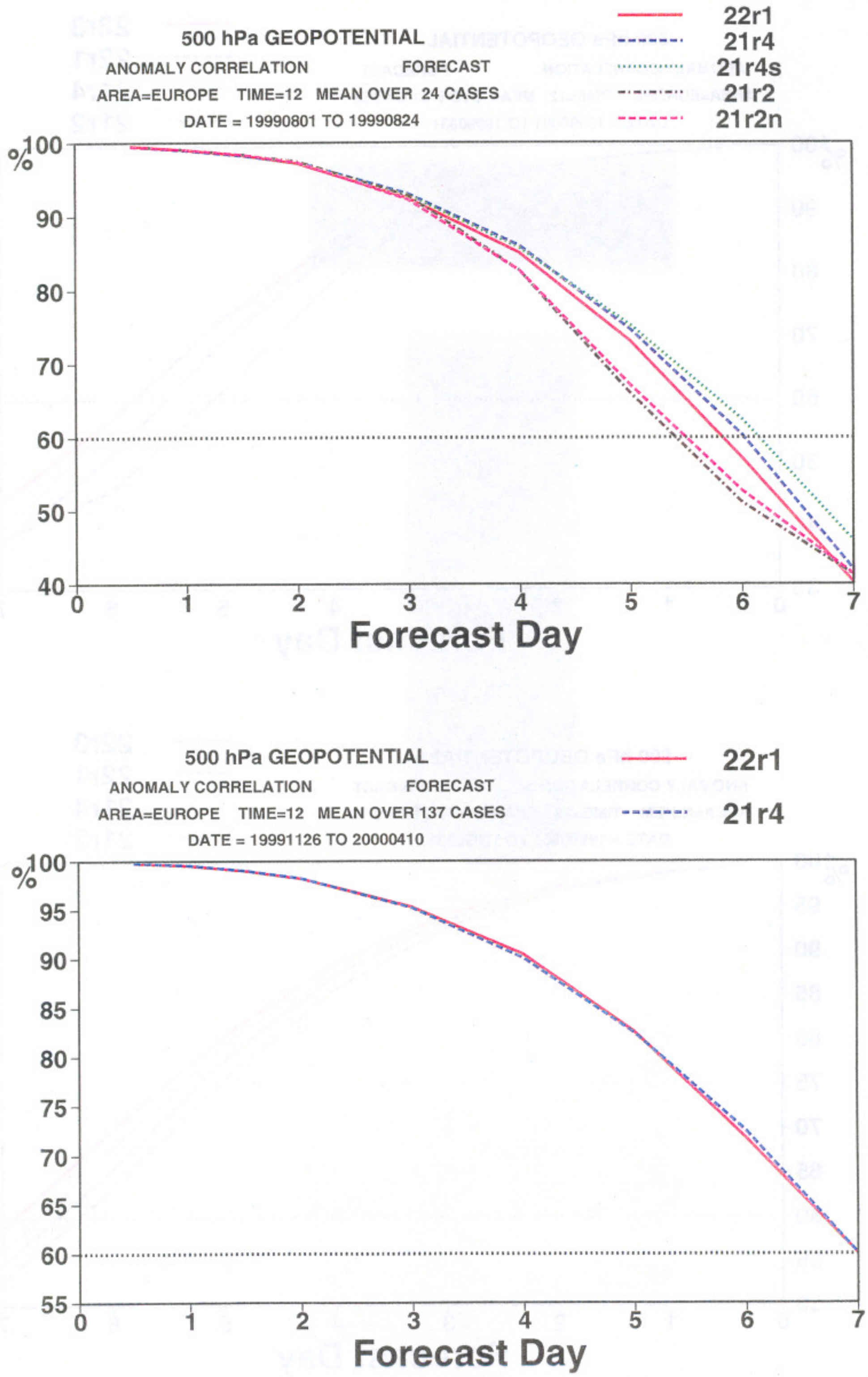
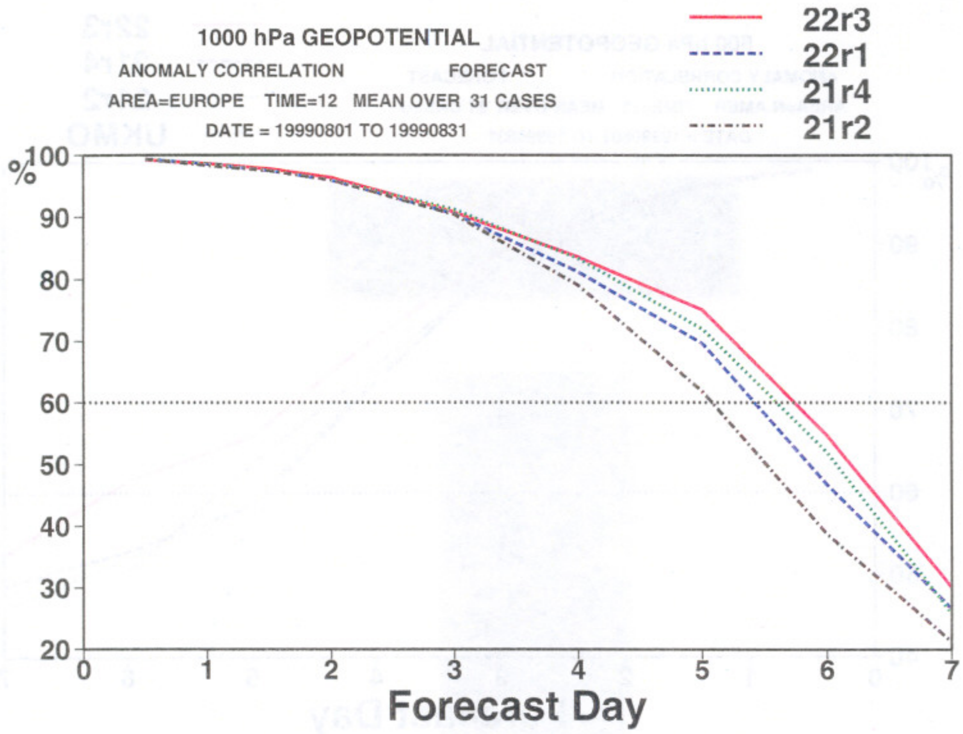


Fig. 16. Anomaly correlations of 500hPa height forecasts for Europe. The upper panel shows results averaged for the period 1-24 August 1999 from cycles 22r1 (red solid), 21r4 (blue dashed) and 21r2 (brown chained), for 21r2 with removal of duplicate MSU observations (21r2n, magenta dashed), as in the 22r1 assimilation, and for 21r4 plus better suppression of stratospheric humidity increments (21r4s, green dotted), also as in the 22r1 assimilation. The lower panel shows results for 22r1 (red solid) and 21r4 (blue dashed) averaged from 26 November 1999 to 10 April 2000.



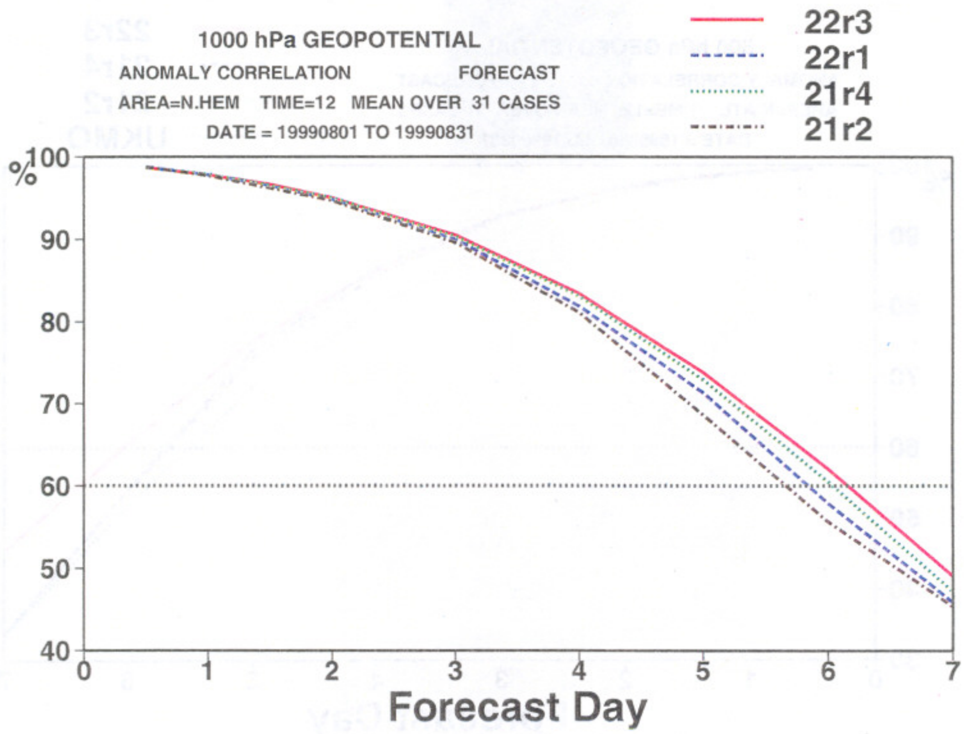
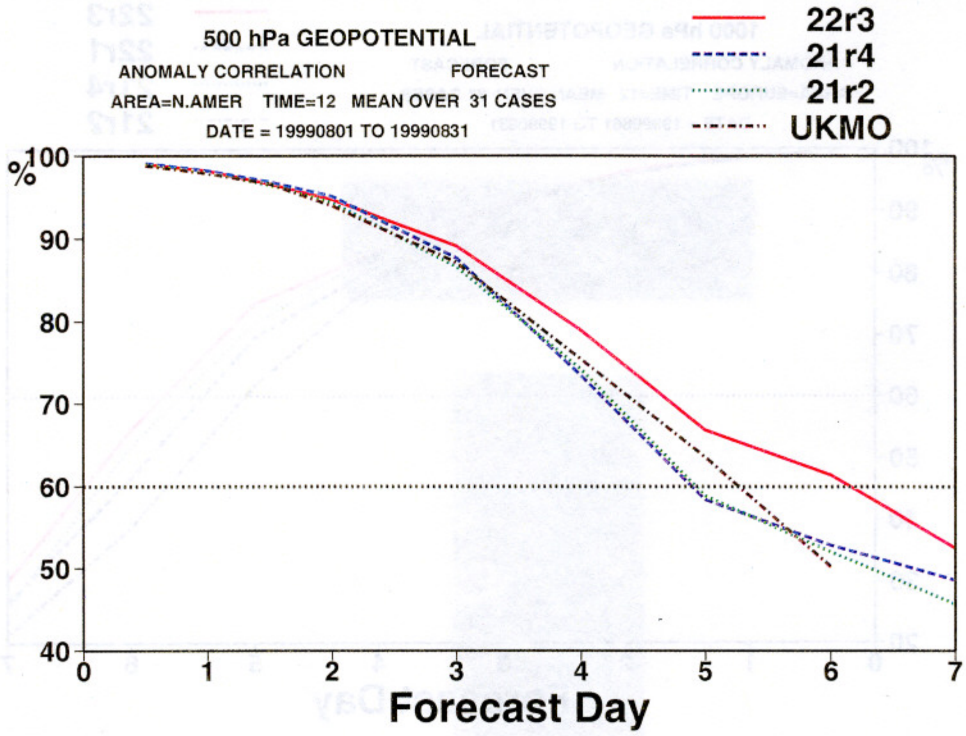


Fig. 17. Anomaly correlations of 1000hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from cycles 22r3 (red solid), 22r1 (blue dashed), 21r4 (green dotted) and 21r2 (operations, brown chained), averaged for August 1999.



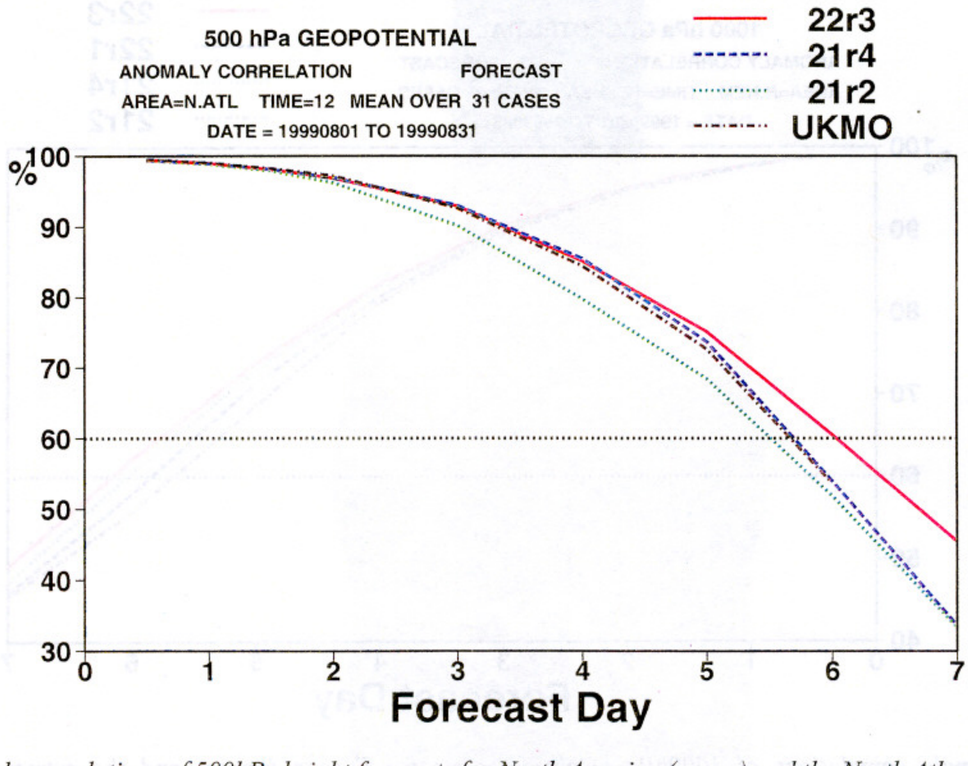


Fig. 18. Anomaly correlations of 500hPa height forecasts for North America (upper) and the North Atlantic (lower) from cycles 22r3 (red solid) and 21r4 (blue dashed) of the ECMWF forecasting system, from ECMWF operations (21r2, green dotted) and from the operational forecasts of the Met. Office (brown chained), averaged for August 1999.

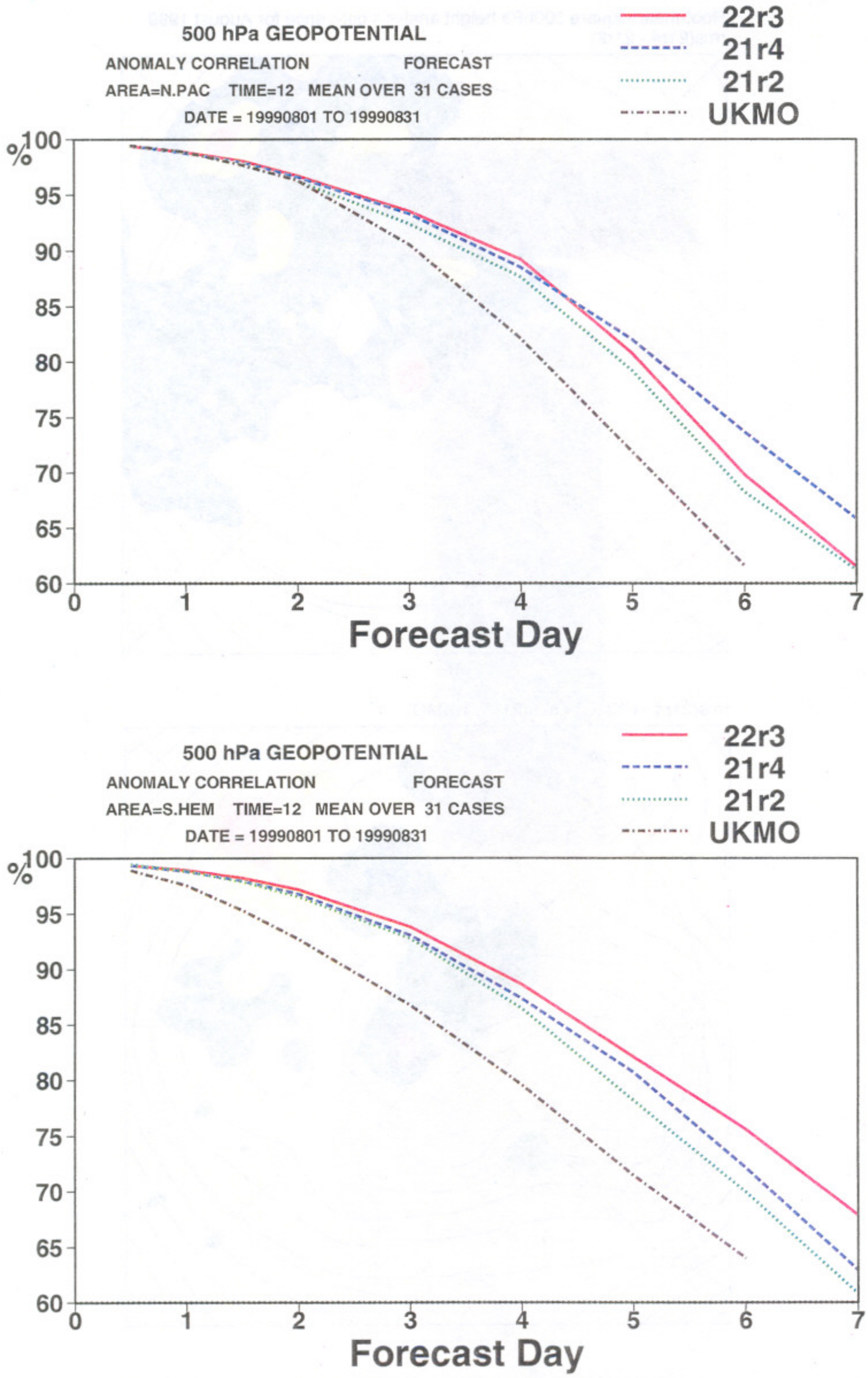


Fig. 19. Anomaly correlations of 500hPa height forecasts for the North Pacific (upper) and the Southern Hemisphere (lower) from cycles 22r3 (red solid) and 21r4 (blue dashed) of the ECMWF forecasting system, from ECMWF operations (21r2, green dotted) and from the operational forecasts of the Met. Office (brown chained), averaged for August 1999.

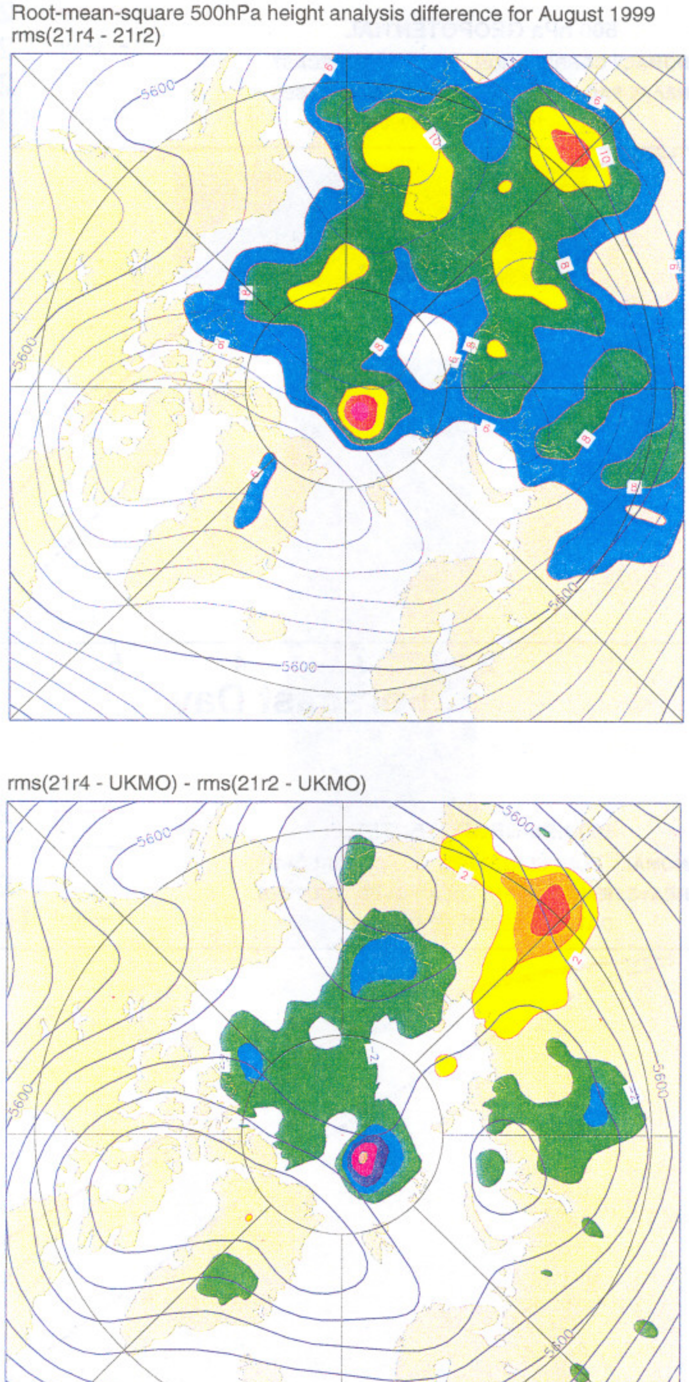
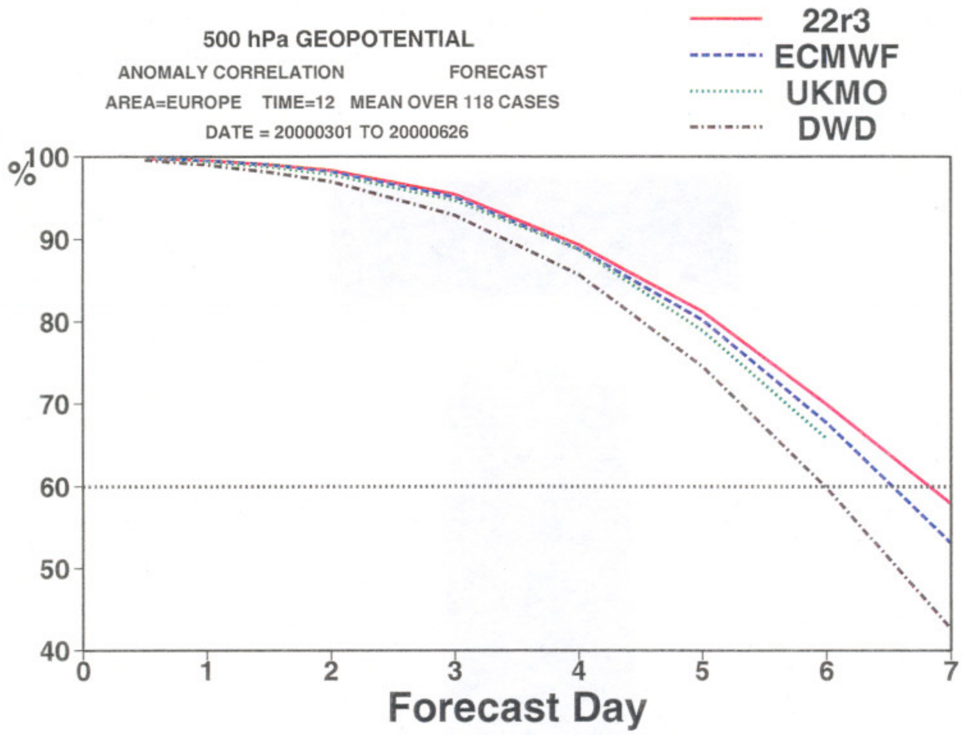


Fig. 20. The upper map shows the root-mean-square (rms) difference between the 21r4 and 21r2 12UTC analyses of 500hPa height for August 1999 (red contours with shading) together with the mean 21r4 analysis for the month (blue contours). The lower map shows the difference between the rms (21r4 - Met. Office) analysis difference and the rms (21r2- Met. Office) analysis difference. Negative values (dashed contours and green/blue shading) indicate where the 21r4 analyses are on average closer to the Met. Office analyses than are the 21r2 analyses, and positive values (solid contours and yellow/red shading) indicate where the 21r2 analyses are closer to the Met. Office analyses. Units: m.



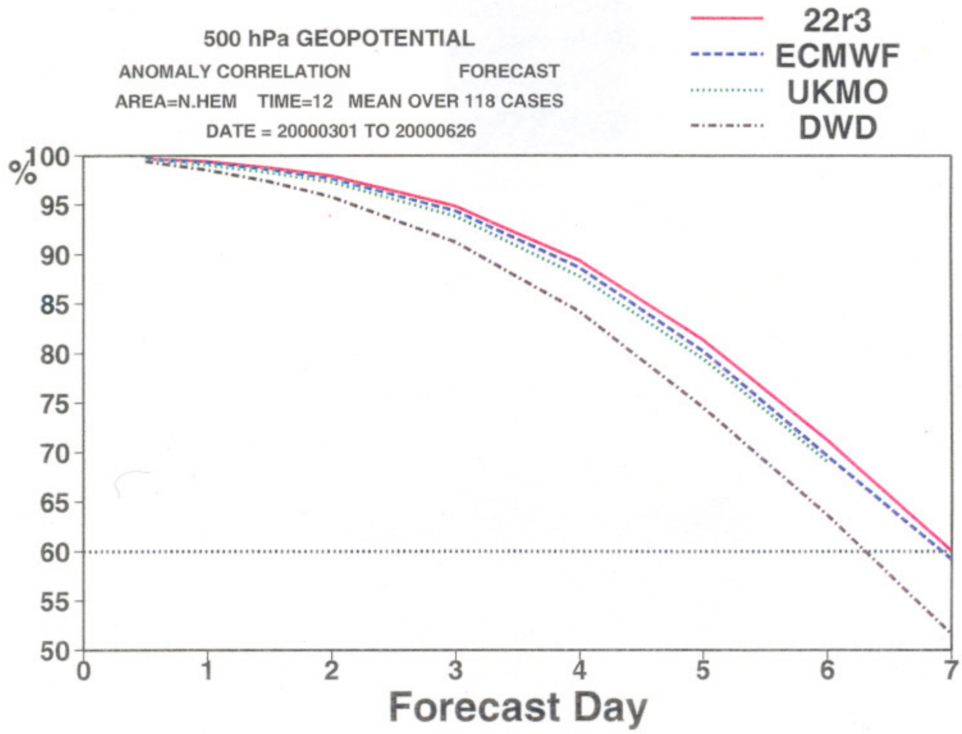


Fig. 21. Anomaly correlations of 500hPa height forecasts for Europe (upper) and the extratropical Northern Hemisphere (lower) from cycle 22r3 of the ECMWF forecasting system (red solid), from ECMWF operations (a mix of 21r4 and 22r1; blue dashed) and from the operational forecasts of the Met. Office (green dotted) and DWD (brown chained), averaged from 1 March to 26 June 2000.