The response of numerical weather prediction systems to FGGE level IIb data. Part II: Forecast verifications and implications for predictability

By K. ARPE\textsuperscript{1}, A. HOLLINGSWORTH\textsuperscript{1}, M S. TRACTON\textsuperscript{2}, A. C. LORENC\textsuperscript{1}, S. UPPALA\textsuperscript{1} and P. KÅLLBERG\textsuperscript{1*}

\textsuperscript{1} European Centre for Medium Range Weather Forecasts, Reading; \textsuperscript{2} Meteorological Office, Bracknell; \textsuperscript{3} National Meteorological Center, Washington, U.S.A.

(Received 28 February 1984; revised 3 September 1984)

SUMMARY

The purpose of our two-part study is to assess the importance of the differences between three independent analyses of the same FGGE level IIb observational dataset. Part I is concerned with the nature and origin of the differences among the analyses. Part II of the study is concerned with the implications of the analysis differences for forecast skill, and for estimates of predictability.

Our experimental material is a set of forecasts by two models from the three ensembles of analyses. The energetics of the analyses are different, and this is reflected in differences in the energetics of the forecasts based on the analyses.

A study of the objective forecast verifications shows that there are clear differences in the forecast skill of the two models we used. The verifications also show a marked sensitivity to the choice of initial data.

We studied in some detail how forecasts made with the same model from different analyses diverge from one another, with particular attention to three aspects: (i) the growth rate at small amplitude; (ii) the time taken for the differences to reach an asymptotic level; (iii) the amplitude of this level relative to that of persistence. The results for small amplitude growth rates are consistent with earlier results in giving a doubling time of 2-0 days in the height field in the day 1–2 forecasts but with slower growth rates (doubling times of 2-6 days) in the wind field. The baroclinic waves show faster growth, with about 0-25 days shorter doubling times. The forecast differences reach an asymptotic level more rapidly in this than in earlier investigations. This suggests that earlier estimates of the purely dynamical potential for predictability of instantaneous weather patterns may be overoptimistic.

We combined the measures of forecast skill and forecast divergence in an error budget which separated the contributions to forecast error arising from model error and analysis error, subject to some simplifying assumptions. The results show that between day 2 and day 5 the model error grows linearly in time and the analysis error grows exponentially. The model is the main source of forecast error in this time range. The analysis error is the main source of forecast error in the short-range forecasts, and is a substantial contributor after about day 5. These results are very sensitive to the magnitude of the initial analysis errors.

Taken together with the synoptic results of part I, these statistical results emphasize the importance of analysis technique for the success of forecasting, particularly in the medium range. Given identical data of good quality, three advanced analysis systems produced analyses which lead to rather different forecasts.

The results are based on a small sample over a period of 24 days, which represent essentially the same synoptic situation; therefore the generality of the results should be tested on further experiments.

1. INTRODUCTION

This study discusses the forecast results from initial datasets which were produced by the analysis systems of three organizations; the European Centre for Medium Range Weather Forecasts (EC); the National Meteorological Center, Washington (US); and the Meteorological Office (UK), for the period 15–19 February 1979. All three analysis systems used as input the main FGGE level IIb observational dataset (Bengtsson et al. 1982). Forecasts were performed with the EC and the US models. The third model, UK, was not designed or implemented as an operational forecast model, and a complete set of forecasts was not available. The level IIb observational data, the analysis systems, and the forecasting models are described in part I of this study (Hollingsworth et al. 1985) and the description will not be repeated here.

* Present address: Swedish Meteorological and Hydrological Institut, Norrköping, Sweden.
The analysis material considered in part I covered a five-day period (15–19 February 1979) during the first Special Observing Period (SOP I) of the FGGE year. The forecasts were all made from the second half of this five-day period; the initial times for the forecasts were 00 GMT on 17 and 18 February, and 12 GMT on 18 and 19 February. The forecast material for this part of the study comprises 20 forecasts as follows: four forecasts with the EC model, for each of the above times, from each of the EC, US and UK analyses (12 forecasts); together with four forecasts with the US model from each of the EC and US analyses (8 forecasts).

The objective of part I was to quantify differences between the analyses and to assess the reasons for them. Forecasts from the different analyses were used only to amplify and identify the important differences.

The objective of part II is to study the implications of these analysis differences for overall forecast skill and predictability, particularly in the medium range (days 2–10). The verification measures we use are designed for this purpose and are discussed in section 2. In section 3 we investigate some aspects of the energetics of the forecasts. These results show that there are important differences in the energetics associated with the various analyses, which may affect the forecast skill. In section 4 the forecasts are compared objectively with verifying analyses. We find some important differences between the forecasting ability of the two models. The objective verifications also show sensitivity to the analyses used as initial data. In section 5 we discuss the rate at which forecasts with the same model from different analyses diverge from one another. Three aspects of these forecast divergences are particularly important: the rate of growth at small amplitude; the time taken for the forecast divergences to reach their asymptotic level; and the level relative to persistence at which the divergences asymptote. In section 6 we use the forecast verification and forecast divergence data, together with some simplifying assumptions, to partition the mean square forecast error into a part arising from model error, and a part arising from analysis error. Using the results of sections 4, 5 and 6 we discuss the implications of our results for estimates of the predictability of atmospheric flow.

Most of the results presented are for the northern hemisphere between 20° and 82.5°N. There is also some discussion of the southern hemisphere forecasts. Analysis and forecast errors in the tropics warrant a separate study, and are not discussed here.

In the course of the discussion we find it convenient to use expressions such as ‘divergence of forecasts’ for the more precise ‘growth of differences between forecasts’. From the context it should be clear that we do not mean ‘divergence’ in the sense of fluid dynamics.

In considering the results of this study the sampling problem inherent in dealing with a small sample of forecasts during a period of typically the same synoptic situation must be borne in mind. The results need not be general. This sampling problem is probably more serious for this part of the study than for part I where differences in the analysed fields could be related to differences in the analysis systems. Nevertheless it is worthwhile to consider the results of this study, as it is the most extensive investigation of this kind to date.

2. Verification methods

In this section we discuss the properties of the scores which are used for the verification of the medium-range forecasts. The ways in which the choice of verifying analysis can affect the results of the verifications will be demonstrated.
The skill scores are the anomaly correlation coefficient and standard deviation of height and temperature and the r.m.s. vector wind difference. All calculations are carried out in the zonal wavenumber domain up to zonal wavenumber 20 to facilitate a separation into contributions by wavenumber groups, specifically wavenumbers 1–3, 4–9, 10–20 and the zonal mean. The definitions of the scores are given in Miyakoda et al. (1972) and Arpe (1980). For theoretical reasons Baumhefner (1983) and others suggest a scale separation by two-dimensional wavenumber rather than by zonal wavenumber. We did not use this approach because it would have included the tropics in our verifications, and we believe that this area needs a separate investigation. In the standard deviations we subtract the area mean differences, because vertical interpolations from model coordinates to standard pressure levels can cause differences in the mean fields which are unimportant for this comparison. A further discussion of the skill scores can be found in the appendix.

The forecasts have been verified against the EC initialized level IIIb analyses because this was the most complete data set available to us; we had to verify forecasts for ten days beyond the end of the analysis intercomparison period. We found that the choice of the verifying analysis favours the forecasts based on that analysis in the first two days; at later times the influence is small. In order to estimate the extent of such biases in our verifications we have also verified the northern hemisphere height and temperature fields against DWD (German Meteorological Office) level IIIa analyses. The two sets of verifications are shown in Fig. 1 for forecasts of the geopotential height field with the EC model. The hatching indicates in each set the spread of scores when using EC, US or UK analyses as initial data for the forecasts.

Initial values here and in most other figures in this part of the study, when EC model output is concerned, refer to the analysis data after they have gone through the following process. The analysis data were interpolated from FGGE level IIIb format to

![Figure 1. Standard deviation of 500 mb height fields (20°N–82.5°N) in the EC forecast on three analysis ensembles using two different verifying analyses: DWD level IIIa and the EC initialized level IIIb analyses. The hatching indicates the spread of the scores of the forecasts from the three sets of analyses. Units: m.](image-url)
model grid (some climatological data were used, where necessary, in the definition of the model levels above 50 mb). Then the fields were initialized (see part I) and one forecasting step was performed. The data were then interpolated back to standard pressure levels and are used here as initial data. The process has a smoothing effect in the EC system, as shown in the appendix of part I. This process results in a non-zero error at day 0 of Fig. 1 even for the EC analyses. Day-0 forecast results from the US model have gone through a similar interpolation process as outlined above for the EC model except for omission of a forecasting step.

It is interesting to note that all day-0 forecasts are nearer to the FGGE level IIIb analyses than any of them to the DWD analyses. After day 3 it is clear that the two verifications give essentially the same results.

Some systematic differences in the analyses persist for a longer period in the models, notably in the zonal-mean zonal wind. As shown in part I and in Fig. 2(a), the zonal wind speed in the stratosphere equatorward of 40°N is larger in the US than in the EC analyses. At 50 mb 20°N the differences are more than 6 m/s. As indicated in part I these differences are due at least in part to the problem of interpolating the US (and UK) fields to 50 mb, as the sigma levels are at about 25 and 75 mb. Biases from the forecast model in the assimilation system may also be important.

This difference in the initial analyses persists during the forecasts, and may be found even in the day-5 forecast; a verifying analysis clearly favours the forecast started from its 'own' analysis for this field.

The verification of the forecasts will concentrate mainly on 500 and 1000 mb height and wind fields and therefore will not be much affected by these stratospheric features in the analyses. The energetic calculations used in section 3 are performed for the 200–1000 mb layer.

3. ENERGETICS OF THE FORECASTS

A necessary requirement of a good dynamical forecast is that it should give a good simulation of the atmospheric energy cycle. It should preserve the amplitude of each of the main components of the energy budget, and should represent the energy conversions in a satisfactory way. Failure to do so is indicative of weaknesses in the analyses or the forecast models, and can affect the objective verifications, as discussed for r.m.s. errors in the appendix. For this reason we discuss the evolution of the energetics in our five ensembles of forecasts, three ensembles made with the EC model (from the EC, US and UK analyses) and two ensembles made with the US model (from the EC and US analyses). Some differences are immediately apparent among the energetics of the forecast ensembles, and are indicative of important differences among the analyses and forecast models. We concentrate our attention on those aspects of the basic energetic quantities where the differences can be seen most clearly; we have examined the energy cycle in its entirety in each case, but space limitations require brevity.

Forecast data from the EC model forecasts are available every 12 hours up to 10 days, while the forecasts from the US model are available every 24 hours up to day 6. Thus diagrams using the US forecast data will have different time axes than those based on EC forecast data.

(a) Forecasts with the EC model

To examine the ability of the EC model to simulate the variability of the wind field, the eddy kinetic energy of all three forecast ensembles is shown in Fig. 3. It is interesting to note that the initial values from different analysis ensembles are much closer to each
Figure 2. Zonal mean differences of zonal winds in the EC model forecasts on EC and US analyses.
Units: m/s.
(a) Initial analyses differences. (b) and (c) Differences between day 3 forecasts and the verifying EC initialized analyses. (d) and (e) Same as (b) and (c) for day 5.
other than the kinetic energies of the original analyses discussed in part I, due to initial data interpolation and initialization as described above. This process reduces the kinetic energy, most strongly in the US analyses (~5%), and to a smaller extent in the EC analyses (~1%).

Figure 3 shows the systematic tendency of the EC model to underestimate the eddy kinetic energy (Arpe 1983). The only exception to this is in the waves 4–9 beyond day 6. The low values for the long waves are of particular concern. Wallace et al. (1983) found that the decay of the quasi-stationary wave amplitudes could be attributed to insufficient orographic forcing.

The most interesting aspect of the energetics of these three ensembles of forecasts is the behaviour of the ensemble from the UK initial data relative to the others. The initial eddy kinetic energy in the UK-based forecasts is similar to that in the US- and EC-based forecasts. During the forecasts from the UK data the level of the eddy kinetic energy drops below that in the EC- or US-based forecasts. For the medium and short waves this holds only for one or two days. Along with this lesser eddy kinetic energy one finds an excessive increase of zonal available potential energy in the UK-based forecasts (Fig. 4, upper panel). This suggests lower conversions from zonal available potential energy, via eddy available potential energy, to eddy kinetic energy in the initial stages of the UK-based forecasts. The lesser conversions from zonal to eddy available potential energy (Fig. 5) support this interpretation, especially for the short waves, if we assume that the generation of zonal available potential energy is not too different in all ensembles. The lower conversions are consistent with a much smaller thermal advection noted in case study 5(c) of part I. After 4 days the UK-based forecasts overcompensate for these lower conversions. This overreaction is strongest and soonest in the wavenumber group 10–20 and can be noticed in the conversion from zonal to eddy available potential energy and in the amount of eddy kinetic energy. The conversion from eddy available potential to eddy kinetic energy is not shown because it is difficult to verify but it should, in adiabatic and stationary conditions, be identical to the conversion shown here.

We show only northern hemisphere data, because the signals in the energetics are clearer in the winter months. The feature described above can, however, be found also in the southern hemisphere.

One reason for the reduced initial eddy activity in the UK-based forecasts may be that the UK analyses are less balanced than the other analyses, as found in part I. The initialization reduces the kinetic energy of the UK analyses by 2%, but we believe that a significant amount of the imbalance occurs in higher vertical modes than the five gravest modes which are initialized. This part is likely to be dissipated during the forecast, contributing to a decline in kinetic energy. This is supported by our finding in section 5 that differences in the wind fields between forecasts on different analyses decrease during the first forecast day. This decrease is strongest in the UK–EC analyses comparison.

These tendencies in the energy levels and conversion rates in the UK-based forecasts are also visible to a lesser extent in the forecasts from the EC and US analyses. This suggests either a deficiency in the EC model or similar deficiencies in all analyses. Too large zonal kinetic energies in the forecasts from the UK and US analyses up to day 8 may be connected with the same problem.

(b) Forecasts with the US model

The eddy kinetic energy in the forecasts with the US model is shown in Fig. 6. The interpolation and initialization process in the initial data reduces the kinetic energy in
Figure 3. Kinetic energy by wavenumber groups integrated from 20° to 82.5°N and 1000 to 200 mb in the forecasts with the EC model. Solid line: verifying FGGE level IIIb initialized analyses; dotted line: forecasts from EC analyses; dashed line: forecasts from US analyses; dashed-dotted line: forecasts from UK analyses. Units: kJ/m².

Figure 4. Available potential energy by wavenumber groups integrated from 20° to 82.5°N and 1000 to 200 mb in the forecasts with the EC model. Solid line: verifying FGGE level IIIb initialized analyses; dotted line: forecasts from EC analyses; dashed line: forecasts from US analyses; dashed-dotted line: forecasts from UK analyses. Units: kJ/m².
the EC analyses by 5% while the kinetic energy is increased in the US analyses by 2%. The reason for the increase is not clear.

In the 1982 operational configuration of the US model which is used here, the horizontal resolution is reduced from rhomboidal-30 to rhomboidal-24 wave truncation after 48 hours, while the vertical resolution is reduced from 12 levels to 6 levels after 84 hours of integration. This change in the vertical resolution is apparent as a sudden drop in the kinetic energy in Fig. 6, which is especially marked in the zonal kinetic energy. A tendency to underestimate the eddy kinetic energy, which was found for the EC model, is also apparent in the US model forecasts, but to a lesser extent for medium and shorter waves. We also find that the zonal kinetic energy is overestimated in the forecasts with the US model from US analyses between day 1 and 3.
Figure 6. Kinetic energy by wavenumber groups integrated from 20° to 82.5°N and 1000 to 200 mb in the forecasts with the US model. Solid line: verifying FGGE level IIIb initialized analyses; dotted line: forecasts from EC analyses; dashed line: forecasts from US analyses. Units: kJ/m².
Summary of the energetic calculations

The main results of this section are that the UK analyses produce an energy cycle in the EC model which is different from that produced by the other two ensembles of analyses; that the changes of truncation in the configuration of the US model used in this study lead to large discontinuities in the energy cycle of the forecasts with that model; and that all the forecast ensembles with the EC model deviate from the observed energy levels, in particular underestimating the variance in the long waves. This last point will affect our estimates of the asymptotic limit of error growth in section 5(b).

We shall see below that the UK-based forecast ensemble is distinctive in a number of other important respects. Similarly, we shall find that the reductions of resolution in the US model lead to other effects on the forecasts besides those discussed here.

4. OBJECTIVE VERIFICATIONS OF THE FORECASTS

In this section we wish to determine the sensitivity of the objective verifications to differences between the models and analyses we have used for our study. In sections 4(a) to (c) we discuss the verification of the ensemble of sixteen forecasts (using EC and US analyses and models) for which we can explicitly separate the effects of model and analysis. Section 4(a) compares the impact of the models, by comparing mean scores for the ensembles of eight forecasts made with the EC and US models. Section 4(b) compares the impact of the analyses, by examining the mean scores for the ensembles of eight forecasts made from the EC and US analyses. Section 4(c) considers the internal consistency of the results by examining the case dependence of the verification scores. Finally in section 4(d) we extend the discussion of the analysis dependence of the scores to include the results from the UK analyses with the EC model.

We have used either the anomaly correlation score or the standard deviation score as the main vehicle for the discussion in each sub-section. The results of the discussion are independent of which measure we use; for brevity we have restricted the discussion to the results of just one score in each sub-section.

(a) Impact of EC and US model differences

In order to investigate the differences in the skill of the forecast models in the medium range, we averaged for each model the scores of eight forecasts made with that model (four forecasts from the EC analyses and four from the US analyses). Figure 7 shows the anomaly correlation coefficient for height, averaged between 20° and 82.5°N and between 1000 and 200 mb. The EC model has a clear advantage over the US model in all wavenumber groups, particularly after day 2, as perhaps might be expected considering the differences in resolution and in the completeness of the physical parametrizations (Brown 1983). Experiments with a version of the US model which has a rhomboidal-30 horizontal resolution and a twelve-layer vertical resolution throughout the first six days yield better forecasts, but the EC model retains its superiority (Tracton 1983).

Defining the limit of useful predictive skill to be the time until the anomaly correlation coefficients drop below 60%, one finds a span of useful predictability of 4½ days for the US model, and a span of 5½ days for the EC model, for the total field. For the (mainly baroclinic) waves with wavenumbers 4–9 the corresponding figures are 3½ and 4½ days respectively. In both cases the EC model gives slightly more than one day longer useful
Figure 7. Anomaly correlation coefficients averaged between 20° and 82.5°N and between 1000 and 200 mb. Solid line: mean scores of the forecasts with the EC model; dotted line: mean scores of the forecasts with the US model; dashed line: intercomparison between forecasts from the same analyses.

forecasts. This agrees with comparisons by Baumhefner (1983) between operational forecasts by both centres on much larger forecast ensembles.

It has been found in earlier work (Baumhefner and Downey 1978 and Cullen et al. 1981) that forecasts by different models from the same analyses resemble each other much more than they resemble the verifying analyses. To test if that is the case here we have included in Fig. 7 an intercomparison between the forecast ensembles, which measures the rate of divergence of pairs of forecasts with the EC and US models from the same analyses, averaged for four EC analysis and four US analysis pairs. For the most part these curves are very close to the curves which give the scores of the EC model in forecasting the real atmosphere. This means that the forecasts by our two models from the same data are about as close to each other as the EC forecast is to the
verifying analysis. The results of Fig. 7, together with further results (omitted for space limitations) such as the standard deviations of height and temperature and the r.m.s. vector wind differences, show that the result of Baumhefner and Downey is not true for the models used here. Their results have been studied by Daley et al. (1981) who concluded that a large part of the shared errors of the models arose from spurious large-scale external Rossby waves which were excited by deficiencies in the generation of the initial data or because of aliasing effects arising from the use of a hemispheric domain. In the present experiments the sources of excitation of the spurious Rossby waves that were considered by Daley et al. presumably do not occur. Cullen et al. found that a difference in the horizontal resolution of forecast models was an important factor for differences in the forecasts; the differences in resolution of the EC and US models may therefore be of particular importance for these calculations.

(b) Impact of EC and US analysis differences

Figure 8 demonstrates the impact of the analyses on the forecast skill. For each analysis set eight forecasts (four with the EC model and four with the US model) have been averaged. A slight advantage in using the EC analyses can be found for all wavenumber groups. This advantage, especially for the wavenumber group 10–20, might be thought to arise from the choice of the verifying analyses. However, verifications against DWD analyses, instead of EC analyses, yield very similar results throughout.

Also included in Fig. 8 is an intercomparison between forecasts from both ensembles of analyses, showing the rate of divergence of pairs of forecasts with the same model from the EC and US analyses, averaged for four EC model and four US model pairs. It is clear that in Fig. 8 both ensembles of forecasts are nearer to each other than to the verifying analyses. This is confirmed by other measures of skill. These results are discussed further in section 5.

From the results of this and the previous section we can deduce that for the EC and US ensembles of forecasts and analyses, the differences between the models have more impact on the objective verifications than the differences between the analyses.

(c) The variability of the scores in the EC and US ensembles

Having studied the ensemble-mean scores, we now examine the variability of the scores. To display the variability of the results within each ensemble, the standard deviations of the 500 mb and 1000 mb height fields at day 3 and day 5 for each forecast are shown on the scatter diagrams of Fig. 9. Forecasts with both models from the EC and US analyses are compared. Points falling below the diagonal indicate that the forecasts from EC analyses score better than those from US analyses, and vice versa. We also indicate the mean score for each model, averaged over each set of four forecasts. In all the diagrams it is clear that the EC model has better scores than the US model.

We note some difference of the dependence of the forecast results between 500 and 1000 mb. At 500 mb, there is a clear tendency for both the US and EC model to prefer the EC models. At 1000 mb the models show somewhat better results for the US analyses, particularly the EC model at day 5. We have no satisfactory explanation for this result.

(d) EC model forecasts from UK, EC and US analyses

In Figs. 10 and 11 the northern hemisphere forecast scores with the EC model from UK analyses are compared with those from EC and US analyses. Although there is a
Figure 8. Anomaly correlation coefficients averaged between 20° and 82.5°N and between 1000 and 200 mb. Solid line: mean scores of the forecasts from the EC analyses; dotted line: mean scores of the forecasts from the US analyses; dashed line: intercomparison between forecasts from the two analysis ensembles for the same date with the same model.
spread in the scores, the scores for the forecasts from UK analyses are clearly worse than the other scores. Such a clear difference was not apparent in the subjective assessment of selected cases of shorter period forecasts in part I. As mentioned in the introduction, in this part we concentrate more on medium-range forecast scores, which are more sensitive to large-scale differences. It was shown in part I that the UK analyses have a large-scale mean height error over continents. The differences in the wind scores are smaller than the differences in the height scores, as shown in Fig. 12 and further discussed in sections 5 and 6. There it will also be shown that the wind scores are more affected by medium-scale waves. This, together with the results from the southern hemisphere, discussed below, suggests that the biases in the height fields cause the lower scores for the UK-based forecasts. We have also seen in section 3 that there are important differences between the energetics of the forecast ensemble based on the UK analyses and the energetics of the other ensembles. These differences may also contribute to the differences in objective scores. The differences in scores we have discussed here are not as large as those found when different forecast models are used, but they are nevertheless significant.

In the southern hemisphere the EC- and UK-based forecasts had very similar scores, with the UK scores being slightly better; the US-based forecasts had clearly worse scores than the other two. In the mean statistics, the UK and EC analyses tend to be more similar to each other over the southern oceans than either is to the US analyses. A case study (section 5(b) part I) of a short-range forecast in the southern hemisphere indicates that despite this general agreement there are nevertheless large local differences between
Figure 10. Standard deviations of height fields of forecasts with the EC model from EC and UK analyses. Markers as in Fig. 9. Units: m.

Figure 11. Standard deviations of height fields of forecasts with the EC model from US and UK analyses. Markers as in Fig. 9. Units: m.
Figure 12. Standard deviation of heights (top) and r.m.s. of winds (bottom) of forecast verifications and forecast divergences with the EC model.Forecasts are compared with the verifying analyses from DWD for the height and from EC for the wind. Averages between 20° and 62°N for the levels 1000 to 200 mb over four cases are presented. The heavy line shows the error of a persistence forecast.

the UK and EC analyses and forecasts. Possible causes for the relatively worse performance of the forecasts from the US analyses in the southern hemisphere are the treatment of the SATEM data which predominate there and the greater importance of the assimilation model in data-sparse regions.

It seems that at least in some cases the qualities necessary in an analysis for a good medium-range forecast are not readily apparent to subjective study either of the analysis or the short-range forecast.
5. The amplification of analysis differences in the forecasts

In this section we want to investigate how the initial differences, due to the different analysis systems, grow in the forecasts. Three aspects of the growth of the analysis differences are of particular interest for predictability studies: (i), the growth rates of the analysis differences, when their amplitudes are small, considered in section 5(a); (ii), the time for the forecast differences to reach their asymptotic level, discussed in section 5(b); as is (iii), the level relative to persistence at which the forecast differences asymptote.

Growth rates of differences in the initial data have been studied by several authors (e.g. Smagorinsky 1969; Lorenz 1969; Shukla 1981) using twin forecast experiments, where pairs of forecasts were performed from ensembles of initial data which differed by small randomly generated amounts; the differences were intended to represent the inaccuracy of analyses. A related study was made by Lorenz (1982) who calculated the rate at which forecasts from successive analyses diverged from each other. In his study the initial differences were day-1 forecast errors which were taken to represent the inaccuracy of the analyses. If the model is realistic, the divergence between the forecasts based on these initial data can be used to estimate the error growth due to the uncertainty in the analyses.

All these authors emphasize that they did not use realistic analysis errors. Our forecast experiments are amongst the first experiments with realistic differences in their analyses. Therefore we felt it worthwhile to study the growth rates of the analysis differences in our forecasts, although we have only a small sample over a period of 24 days which represent essentially the same synoptic situation. We wish to obtain estimates of the significance of the analysis differences which are as realistic as possible. As shown in section 4, there are significant differences in the skill of the two forecast models. Thus it seems reasonable to focus the discussion in this and the following sections on the results from the EC model, which showed more accurate forecasts, and for which we have a complete set of forecasts. We investigate only the northern hemisphere results because of the higher accuracy of the analyses there.

(a) Growth of small differences

Figure 12 shows the forecast verification scores and the forecast divergences for the height and wind fields. Averages for the area 20° to 82.5°N and the levels 1000 to 200 mb are used but there is hardly any difference to the scores for the 500 mb level. The scores of the persistence forecasts are given for comparison.

During the first five days the forecast errors increase fairly linearly and after day 6 the growth rates begin to decrease, especially in the wind verifications. The growth of divergence between the forecasts is more of an exponential form during the first 5 days. The error of the persistence forecasts grows very rapidly at first and then more slowly. By day 10 the forecast errors are close to the level of the persistence forecast, while the level of the forecast divergences are significantly lower.

To quantify the growth rates of forecast divergence in the early stage of the forecasts we fitted our data to two error growth laws: (i), an exponential (linear on a logarithmic plot) growth law; (ii), the nonlinear law described by Lorenz (1982). Lorenz’s method uses the data from the whole 10-day forecast length; the simple exponential fit uses only the data from the early stages of the forecast because differences start to level off after
day 5. The earliest (day 0–1) forecast differences cannot be used with either fitting method because there is no growth of differences between day 0 and day $\frac{1}{2}$ in the height field, and there is a decrease in the wind field. Such a decrease of the forecast differences in the first day has been observed in earlier experiments (e.g. Smagorinsky 1969). It can arise for several reasons. The initial differences can consist of both amplifying and decaying quasi-geostrophic modes, as well as gravity-wave noise. Dispersion and damping may contribute to a decrease of the differences in the early stages, before the growing modes begin to dominate the difference fields. In section 3(a) we noted that the initialization procedure for the EC model has a damping effect on the level of eddy kinetic energy and that the eddy kinetic energy decreases further during the early forecasts, especially in the UK-based forecasts. This is consistent with our results here that the forecast wind fields converge during the first day.

In Fig. 13 we show the 500mb height r.m.s. forecast differences for the three forecast ensemble intercomparisons, together with the linear and nonlinear growth rate fitting curves. For the simple exponential fit we experimented with several forecast ranges up to 5 days. Increasing the forecast range used for fitting an exponential curve resulted in longer doubling times (smaller growth rates), as may be expected from the distribution of the data in Fig. 13. The growth rates from the exponential fit (Table 1) are for the forecast range day $1\frac{1}{2}$ to 2. On a logarithmic plot the exponential curve is a straight line and fits the data in this range, and even out to day 5, very well.

The curve obtained from the method of Lorenz (1982) fits the data for the whole range. To plot this curve we had to define a constant of integration representing the error at day 0. This was done by a simple averaging procedure, which did not affect the calculations of growth rates and need not be discussed further.

The growth rate of differences in Fig. 13 is represented by the slopes of the fitting curves. During the early forecasts both fitting curves have very similar slopes, which is also exhibited by the doubling times in Table 1. For the nonlinear fit the doubling time at day 1 is given, to avoid extrapolation to zero amplitude. If we assume that the growth rates of differences with the EC model agree with those of a perfect model, we can deduce that with a perfect model a halving of the present total analysis error would lead to a gain of predictability of 2-0 days in the height field for all forecast ranges. The scale of motion on which the improvements occur is, however, important; when halving the analysis errors of the medium, mostly baroclinic, waves only 1-7 days of predictability will be gained. This faster growth rate of the baroclinic waves agrees with our synoptic evaluation of the forecast experiments in part I. The potential for gains in predictability will be discussed further in section 6.

Calculations of the doubling times of wind differences between the forecasts resulted in longer doubling times (lower growth rates) as shown in Table 1. This is probably due to the fact that the initial analysis errors in the wind field are, relatively speaking, nearer to the asymptotic differences than those in the height field. If in Fig. 12 we consider the average ratios of the day-0 to the day-10 forecast differences these ratios are 0-14 for the height and 0-29 for the wind field. From Fig. 13 it is obvious that the growth rates decrease with increasing amplitude. The growth rate of the height field divergence at day $2\frac{1}{2}$ (using the Lorenz-type fit) is the same as the growth rate of the wind field divergence at day 1. Moreover the ratio of the height field divergence at day $2\frac{1}{2}$ to that at day 10 is also 0-29. So it is plausible to find a lower rate for the relatively larger initial differences in the wind field. We found much lower growth rates in the southern hemisphere than in the northern hemisphere. This we interpret as being due to the larger analysis errors in the southern hemisphere.

A study of the data points in Fig. 13 (or the curves for the divergence between
Figure 13. Root mean square divergences, on a logarithmic plot, of 500mb height forecast with the EC model from EC, US and UK analyses. Averages between 20° and 82-5°N. Data have been fitted to two growth laws: thick line, exponential growth law using only data from day 14 to 21; thin line, a nonlinear law described by Lorenz (1982).

### TABLE 1. DOUBLING TIMES FOR THE DIFFERENCES BETWEEN THE FORECASTS BASED ON ANALYSES BY THE EC, US AND UK SYSTEMS

<table>
<thead>
<tr>
<th></th>
<th>Exponential fit</th>
<th>Lorenz-type fit for day 1 differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>day 1-5 to 2-5</td>
<td>(days)</td>
</tr>
<tr>
<td>Height</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2.01</td>
<td>1.98</td>
</tr>
<tr>
<td>Waves 1–3</td>
<td>2.02</td>
<td>2.15</td>
</tr>
<tr>
<td>Waves 4–9</td>
<td>1.76</td>
<td>1.73</td>
</tr>
<tr>
<td>Wind</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2.67</td>
<td>2.59</td>
</tr>
<tr>
<td>Waves 1–3</td>
<td>2.53</td>
<td>2.59</td>
</tr>
<tr>
<td>Waves 4–9</td>
<td>2.36</td>
<td>2.40</td>
</tr>
</tbody>
</table>
forecasts in Fig. 12) reveals the variations of the results between different intercomparisons. For the total field, and for the contribution by the long waves, the EC- and US-based forecasts diverge much faster from each other than either from the UK-based forecasts. The EC-US comparisons have the smallest differences up to day 1 and the largest during the last three days of the forecasts. The EC-UK forecast differences grow more slowly for these fields. In the lower panel of Fig. 13 showing the contribution by the mostly baroclinic waves all forecast comparisons have very similar growth rates. As discussed in section 6 these variations in the growth rates of divergence between different ensembles may be connected with correlations between errors in the analyses of the different systems.

The doubling times of differences between the forecasts discussed above are those for analysis errors which are representative of the current amplitude of analysis differences. To estimate the growth rate of infinitesimal differences one has to extrapolate to zero amplitude. The doubling time derived with the (linear) exponential fitting curve is independent of amplitude, but Lorenz's method gives shorter doubling times for smaller amplitudes; with our data it leads to very much shorter doubling times: 1-6 days for the total field, 1-8 days for the long waves, and 1-4 days for the mostly baroclinic waves. We regard these very short doubling times for infinitesimal differences as being unrealistic due to biases in the extrapolation method. Plots of differences for separate intercomparisons showed almost exponential growth in the early part of the forecasts, followed by a sudden change over to a final asymptotic level. This sudden transition could not be represented by Lorenz's fitting curve and leads to an apparent bias towards an overestimation of growth rates (underestimation of doubling times) at zero amplitude. The transition to saturation occurred at different times for different ensembles of forecast comparisons, and so the averaged results show a more gradual transition to saturation, with less tendency towards a bias at small amplitudes. The effect would probably be even less noticeable with larger ensembles.

If we want to compare our calculations of doubling times with those by Lorenz (1982) we have to use the extrapolated values because he did not quote the growth rates for finite amplitudes. Lorenz estimated that the doubling time for infinitesimal errors is 1-85 days, whereas we found 1-6 days. It has been found that the predictive skill of the operational ECMWF forecasts varies considerably with the synoptic situation (Bengtsson and Simmons 1983), and the skill scores presented in section 4 for our cases are lower than mean operational winter scores. To determine if our results are case dependent with respect to error growth due to analysis differences, we repeated Lorenz's calculations with our data by comparing forecasts on successive analyses from the same system, the forecasts being verified at the same time. Unlike Lorenz we considered only r.m.s. errors north of 20°N and we did not modify the error fields. For our small sample the growth rate for infinitesimal errors calculated in this way turned out to be almost exactly the same as that given by Lorenz for the whole winter season. The doubling time in our sample was 1-86 days instead of 1-85 days. This suggests that our period was not atypical.

If we compare the longer doubling time, 1-85 days for infinitesimal amplitudes, given by Lorenz and confirmed by us, when using forecast errors as initial disturbances, with the value of 1-6 days, when using realistic analysis errors, we find that the realistic analysis differences contain components which grow faster than others. From our recalculation with Lorenz's configuration we also have doubling times at finite amplitudes, which lead to the same conclusion, e.g. the doubling time with the difference amplitude of 19-1 m, which is the amplitude used in Table 1 for the total field, is 2-30 days, when using forecast errors as initial differences, and 2-01 days, when using our analysis differences. We do not have an explanation for this difference in growth rates.
(b) The asymptotic level of forecast differences

In Fig. 12 we have seen that the growth of differences between the forecasts slows down after day 6, indicating that the forecast differences are beginning to level off, although they have not at all reached the level of persistence forecast errors. The time taken to reach the asymptotic level of differences can be used as an ultimate limit for predictive skill as far as instantaneous weather patterns are concerned, where the limit arises from uncertainties in the initial analysis. With a perfect model one would expect this level to be the asymptotic error level of a persistence forecast.

Using the results from fitting Lorenz’s nonlinear error growth model to the data we can estimate the time for the forecast divergence to reach a specified fraction of its implied asymptotic level. Table 2 presents such a calculation for the different wavenumber groups in the height field. For the wavenumbers 4–9 the levelling off after day 7 can be seen clearly in Fig. 13. These results suggest that the forecast divergence is probably close to its asymptotic level over the important parts of the spectrum within a few days after day 10.

<table>
<thead>
<tr>
<th></th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-20</td>
<td>8½</td>
<td>~10½</td>
</tr>
<tr>
<td>1-3</td>
<td>9¼</td>
<td>&gt;10</td>
</tr>
<tr>
<td>4-9</td>
<td>7½</td>
<td>9½</td>
</tr>
</tbody>
</table>

Shukla (1981) calculated results similar in nature to ours, the main difference being that the initial perturbations were randomly generated in the wind field and that he used a different model. He presents forecast divergences for the height field for the zone 40–60°N, rather than 20–82.5°N, and for a slightly different wavenumber grouping (k = 0–4 and k = 5–12). For comparison we show Fig. 14, which can be compared directly with his Fig. 3. The results of Fig. 14 are generally similar to the results of Fig. 12, and show that the forecast divergence in the medium wave group (k = 5–12) reaches an asymptotic level after 6 days. The asymptotic level of the persistence for these waves in our as well as in Shukla’s experiments is about 81 m. The time for the forecast divergence to reach 80% of this level is 5-8 days in this study and 14 days in Shukla’s study. This suggests a much shorter forecast length with predictive skill in our experiments.

In the long wave group (k = 0–4) we find a clear retardation in the growth of divergences after day 8 in our calculations, but without longer integration we cannot know if this really indicates the approach to an asymptotic level. In our calculations the r.m.s. divergence reaches about 105 m at day 10, which is about 65% of the asymptotic persistence error, while in Shukla’s calculations, which show a similar asymptotic level for the persistence forecast, this value is not reached before day 23. Consequently we find also for the long waves that it takes less than half the time to reach the same level of divergence in our experiments compared to Shukla’s.

Part of this discrepancy may be due to the initial perturbations used. Shukla generated his perturbations by introducing random perturbations only in the wind field. From an experiment by Hollingsworth (1980) we know that the initialization damps a large part of a random perturbation in the height field although it was supported by
approximately geostrophic winds. Daley (1981) has shown that only the projection of the perturbations on the Rossby components of the flow will be relevant for predictability studies. For these reasons Shukla's divergence curves reach a specified small error at a much later time than in our study, e.g. for the waves 0–4 the 20 m level is reached after 7 days in his and after 2 days in our experiments. Perhaps more important are the growth rates after the amplitudes have reached 20 m. The time for the differences for the waves 0–4 to grow from 20 to 90 m is 5 1/2 days in our and 11 days in Shukla's study. This difference in the growth rates may be due to differences of the models used. Because of the small and dependent sample in our experiments the synoptic situation could also be important. Similar experiments by Baumberger (personal communication) with the NCAR model give growth rates of forecast differences which are very similar to ours in the amplitude range 20–90 m.
The retardation of the divergence between the EC model forecasts of the long waves after day 8 in Fig. 14 can also be interpreted as a final approach to an asymptotic level which is different from the asymptotic level of persistence. The different asymptotic levels might be due to the systematic error of the EC model in underestimating the variance of the height field, which was demonstrated in section 3 for the eddy kinetic energy. The fact that the forecast errors at day 10 are much closer to the persistence level than the forecast divergences does not alter this view, since the forecast verifications have an extra contribution from the mean errors of the model, discussed by Hollingsworth et al. (1980), which the forecast divergences do not have. This interpretation puts Shukla’s and our results even further apart.

![Figure 15](image)

**Figure 15.** Contribution of wavenumber groups to the total forecast differences in the height and wind field, averaged over the three ensembles and averaged between 20° and 82.5°N and for the levels 1000 to 200 mb.

To round off our discussion of the asymptotic behaviour of the forecast divergences we consider Fig. 15, which compares the contribution of wavenumber groups to the forecast divergence in the height and wind fields r.m.s.-averaged for all ensembles. With this figure one can easily estimate subjectively the time until the forecast differences reach their asymptotic level. For the wavenumber group 10–20 this clearly happens after
day 5 and for the medium waves around day 7 in accordance with our findings above. In the period between day 7 and day 10 the medium waves in the wind field still show some growth, while the medium waves in the height field have effectively reached the asymptotic level in the same time period.

Another important aspect is noteworthy. In the height field divergences the long waves and medium waves make roughly equal contributions to the total forecast divergence until day 7; thereafter the long waves make the larger contribution. Throughout the ten-day period the contribution of the short waves is negligible. In the wind field on the other hand, the medium waves make much the largest contribution to the divergence, with the long and short waves making similar contributions up to day 7. After day 7 the contribution of the short waves reaches an asymptotic level while the long wave contribution continues to increase, but it remains clearly smaller than the contribution of the medium waves.

6. THE CONTRIBUTION OF ANALYSIS ERROR TO FORECAST ERROR

The forecast divergence statistics just discussed give us information about the growth in time of the analysis errors. The forecast verification statistics presented in section 4 include the effect of both analysis errors and inadequacies of the forecast model. In this section, by making some simplifying assumptions, we combine the two sorts of statistics and derive an error budget separating the contributions of the model deficiencies and of the initial analysis errors to the total forecast error. The assumptions underlying the error budget are suggested by the results of part I, but are not rigorously justified. The implications of the budget are therefore qualitative, but are nonetheless of interest.

The contribution of analysis errors to forecast errors has been discussed previously by Robert (1976). His discussion was concerned with short-range forecast errors at a single location in a data-rich area, while our discussion considers the northern hemisphere extra-tropics. It is therefore difficult to compare the two studies.

As in section 5 our calculations are made for the EC model. For our discussion we use the mean square geopotential height and wind differences over the area 20°–82.5°N and 200–1000 mb. In both cases the statistics are averaged over the ensembles of four forecasts. We use these two statistics because their spectral distributions are different, as we have seen already in section 5. The height fields were verified against the DWD analyses to eliminate any bias. For the wind fields we used the EC analyses, as the DWD analyses did not provide winds. This gave unrealistically low forecast errors for the EC-based forecasts for the first day. To remedy this we assumed up to day 1 that the EC-based forecasts had r.m.s. errors which were 90% of the average of the other two ensembles.

(a) Error budget equations

Let us express the error $A$ of a forecast produced by an imperfect model from one of the imperfect analyses of analysis system 'a' in the form

$$A = M + AE + Xa$$

where $M$ represents the error there would be if the model alone were imperfect, and will be called the model error; $AE$ represents the error there would be if the analysis alone were imperfect, and will be called the analysis error; $Xa$ represents the interaction between the other two errors if both are non-zero.
If $B$ represents the error of a forecast by the same model from an analysis produced by a second analysis system ‘b’ then we may write

$$B = M + BE + Xb$$

so that the difference between the forecasts $a$ and $b$ is

$$A - B = AE - BE + Xa - Xb.$$ 

Most of the predictability studies cited above assume that mean forecast differences are independent of the model imperfections; this assumption will be made here also. Defining $\overline{AB^2}$ as the area and ensemble average of the difference between forecasts based on the analyses of systems $a$ and $b$, it follows that

$$\overline{AB^2} = \overline{AE^2} + \overline{BE^2} - 2VAB,$$

where $VAB$ is the covariance of $AE$ and $BE$. This equation has useful asymptotic properties for long forecast ranges.

Miyakoda et al. (1972) show that the asymptotic level of mean square forecast error (when all skill is lost) is $2S^2$ where $S^2$ is the climatological variance, provided the model is unbiased, as we shall assume. Miyakoda et al. further show that the asymptotic level of the tendency correlation is 0.5. Their arguments can be used to show that if $P, Q, R$ are randomly chosen sets of atmospheric states, then the temporal and area averaged covariance of the differences $(P-R)$ and $(Q-R)$ is $S^2$, and the correlation of $(P-R)$ and $(Q-R)$ is 0.5. It follows that the asymptotic level of the covariance of the errors of two independent sets of forecasts is $S^2$. Consequently the above expression for $\overline{AB^2}$ is valid at large times, and the sum of the neglected terms is asymptotic to zero.

For forecasts with the same model from three sets of analyses $a, b, c$, one may write

$$\begin{align*}
\overline{AB^2} &= \overline{AE^2} + \overline{BE^2} - 2VAB \\
\overline{BC^2} &= \overline{BE^2} + \overline{CE^2} - 2VBC \\
\overline{CA^2} &= \overline{CE^2} + \overline{AE^2} - 2VCA
\end{align*}$$

(1)

where $\overline{AB^2}, \overline{BC^2}, \overline{CA^2}$ represent the mean square forecast differences; $\overline{AE^2}, \overline{BE^2}, \overline{CE^2}$, represent the mean square forecast errors with a perfect model, and $VAB, VBC, VCA$ represent the covariances of $AE$ with $BE$, etc.

Consider now the area and ensemble mean square forecast error of the forecasts from the analyses of system $a$:

$$\overline{A^2} = \overline{M^2} + \overline{AE^2} + (2VMa + X^2a + 2VMXa + 2VAXa)$$

where $VMA, VMXa, VAXa$ are the covariances of $M$ with $AE$, $M$ with $Xa$ and $Xa$ with $AE$.

For long forecast times the sum of the bracketed terms must be equal to $-2VMA$, because $\overline{A^2}, \overline{M^2}$ and $\overline{AE^2}$ all tend to $2S^2$ (since the model is assumed to be unbiased). For forecast times up to three to five days it was found in several examples in part I that analysis differences could grow in a similar fashion in different models. For these forecast ranges the dominant terms in the forecast error budget must therefore be $\overline{M^2}$ and $\overline{AE^2}$; $VMA$ and the other terms being small.

These considerations suggest the approximation that

$$\overline{A^2} = \overline{M^2} + \overline{AE^2} - 2VMA.$$ 

The error in this simplification will vanish at large forecast ranges, and will be modest at forecast ranges up to about day 5, when $\overline{M^2}$ and $\overline{AE^2}$ are the dominant terms. It is
difficult to estimate the error of the approximation in the period between these forecast ranges; considerations of continuity suggest it may not be too large.

Stated simply, the assumption is that the direct effects of analysis error and model error are the only significant contributors to forecast error; the covariance term takes account of the fact that whether there is imperfection in the model, or in the analysis, or in both, the asymptotic level of forecast error is still $25^2$.

The ensemble mean square error of forecasts with the same model from three sets of analyses a, b, c, may then be written

$$\begin{align*}
A^2 &= M^2 + AE^2 - 2VMA \\
B^2 &= M^2 + BE^2 - 2VMB \\
C^2 &= M^2 + CE^2 - 2VMC
\end{align*}$$

where $VMA$, $VMB$ and $VMC$ are the covariances of $M$ with $AE$, $BE$ and $CE$ respectively. Each of these equations should have an additional term to account for the error $D^2$ of the verifying analysis, which is assumed independent of the other terms and constant in time. If we suppose that $M^2(0) = 0$, we can use the statistics at time 0 to give information about $D^2$. In our discussion below the value for $D^2$ has been estimated in this way and has been accounted for by subtracting it from all scores. This enables us to proceed as if $D^2$ were zero.

If we sum the three Eqs. (2) we obtain a simple expression for $M^2(t)$:

$$M^2(t) = \{[(A^2(t) + B^2(t) + C^2(t))] + 2[VMA(t) + VMB(t) + VMC(t)] -$$

$$- [AE^2(t) + BE^2(t) + CE^2(t)]\}.$$ 

(3)

Thus the model error calculation depends on the averages of the total forecast error, the analysis error, and of the model-error/analysis-error covariances. The sum of the three Eqs. (1) implies that

$$AE^2(t) + BE^2(t) + CE^2(t)$$

$$= \{AB^2(t) + BC^2(t) + CA^2(t)\}/2 + VAB(t) + VBC(t) + VCA(t).$$

(4)

The calculation of the average analysis mean square error depends on the average of the mean square forecast divergence, and the average of the analysis-error/analysis-error covariances; but it is independent of the model error. Much of our discussion will be concerned with the average analysis mean square error; we therefore need knowledge of, or assumptions about, the average of these covariances.

Equations (1) and (2) give us six equations for ten unknowns. Obviously we cannot make progress without some simplifying assumptions about the covariances in the problem. In choosing such simplifying assumptions it is useful to consider the asymptotic limits of Eqs. (1) and (2). All the correlations in Eqs. (1) and (2) have the asymptotic limit of 0-5, provided there are no biases, or systematic errors, in the model. Although our model is not unbiased we shall in all calculations below assume that the model-error/analysis-error correlations in Eqs. (2) are equal, that they are zero initially, and that they grow linearly in time to reach an asymptotic value of 0-5 in 12-5 days. We do not consider it worthwhile to try to discriminate between these correlations, as the model error calculation depends only on the average of these covariances, and the average analysis error is independent of them.

Having specified these three quantities, we need only specify one further variable in order to solve the system. It is convenient to specify the sum of the analysis error
covariances through a variable \( q(t) \) as follows:

\[
VAB(t) + VBC(t) + VCA(t) = q(t) \left( \overline{AE^2(t)} + \overline{BE^2(t)} + \overline{CE^2(t)} \right)
\]

so that Eq. (4) becomes

\[
\overline{AE^2(t)} + \overline{BE^2(t)} + \overline{CE^2(t)} = 0.5 \left( \frac{4AB^2(t) + 4BC^2(t) + 4CA^2(t)}{1 - q(t)} \right).
\]  

(5)

The quantity \( q(t) \) can be thought of as a variance-weighted average of the analysis error correlations. It is convenient to specify the time variation of \( q \), which in the asymptotic limit tends to 0.5. We consider two simple specifications for \( q(t) \) as follows:

- **case 1:** \( q(t) = 0.2 + 0.3t/12.5 \) where \( t \) is in days;
- **case 2:** \( q(t) = 0.5 \).

These two cases probably encompass the likely range of variation of the average analysis error intercorrelations. Case 1 was chosen because it is the smallest choice for which the individual correlations are always positive, for both height and wind; we do not expect that there will be significant negative correlations between the analysis errors. Case 2 was chosen because with this choice of \( q \), Eq. (5) implies that the average analysis error is exactly the average forecast divergence. If we increase the correlations in the early stages beyond this value, the implied values for the analysis error exceed the measured forecast errors for an increasing range in the earlier part of the forecast.

At this point we recall the results of part I of this study. In the northern hemisphere, with its good data coverage, the striking aspect of the daily analyses was the diversity between the analyses. Large local differences could always be found in every variable. Verifications of the analyses against the observations supported the subjective view that correlation between the analysis errors was likely to be low. These results suggest that case 2 represents an upper bound to the analysis errors. We feel that these two cases provide reasonable limits for the range of analysis error intercorrelations that are consistent with our subjective judgement, with our expectation that the correlations should be positive, and with our expectation that the asymptotic values are reached soon after day 10.

Having specified the correlations, the solution of the system proceeds by eliminating the analysis error terms in Eq. (3), to obtain an equation in a single unknown, \((\overline{M^2})^{1/2}\). The method of elimination differs for the linear and quadratic terms. The quadratic term \( \overline{AE^2} + \overline{BE^2} + \overline{CE^2} \) is eliminated using the forecast convergence data (Eq. 5). Expressions for the analysis errors \((\overline{AE^2})^{1/2}, (\overline{BE^2})^{1/2}, (\overline{CE^2})^{1/2}\) in terms of \((\overline{M^2})^{1/2}\) may be obtained from Eq. (2) by solving quadratics; these expressions are used to simplify the covariance terms of Eq. (3), which are linear in the analysis errors. The resulting nonlinear equation for \((\overline{M^2})^{1/2}\) is solved numerically for the smallest positive solution. It is worth noting that the equations are solved independently at each 12-hour time step, so that the time continuity of the solutions derived from the data gives us confidence in the method.

(b) The solutions of the error budget equations

Figure 16 shows the average forecast error and the contributions to forecast error from the model error and the three analysis errors in case 1, where we have done the calculations for both the wind data and the height data. We can see that the results for the analysis errors reflect the verification results of section 4, with UK showing the largest contribution of analysis error to the total forecast error, out to day 6, and EC the lowest. Of particular interest is the fact that the UK analysis error behaves rather differently in the wind and height results. With the wind data there is much closer qualitative agreement between the results for the three analysis systems. We have already
Figure 16. Time evolution in case 1 of the average total forecast error, model error, and the three analysis errors in the height and wind field. The model error is the forecast error that would be made by the EC model from a perfect analysis, while the analysis errors are the forecast errors that would be made by a perfect model from the UK, US and EC analyses. Calculated from standard deviations of height and r.m.s. of wind averaged between 20° and 82.5°N and 1000 and 200 mb.

noted that in the northern hemisphere there are large-scale biases in the UK analyses of height (cf. part I). These biases may contribute to the differences in the UK results in Fig. 16, since we have seen already that the height data has significant contributions from the larger scales, while the wind data is dominated by shorter scales.

In what follows we confine ourselves to a discussion of the model error and the average analysis error, because these quantities depend on the averages of the assumed correlations, of the total forecast error, and of the forecast divergence.

Figure 17 shows the mean square model and analysis error, where both have been normalized by the averaged total forecast error, \( [(\tilde{A}^2(t) + \tilde{B}^2(t) + \tilde{C}^2(t))/3] \), for the two cases specified above. These quantities do not sum to 1, because the model-error/analysis-error intercorrelations are specified to be non-zero. There is excellent agreement
Figure 17. Time evolution of the relative contribution of model error and analysis error to total mean square forecast error in the wind field in the two cases discussed in the text (case 1: solid lines, $q = 0.2 + 0.3q/12.5$; case 2: dotted lines, $q = 0.5$). The shading represents the range of solutions within these limits. Calculated from r.m.s. differences of wind averaged between 20° and 82.5°N and 1000 and 200 mb.

Figure 18. Time evolution of the average r.m.s. forecast error and the average r.m.s. analysis error in the wind field in the two cases discussed in the text (case 1: dashed lines, $q = 0.2 + 0.3q/12.5$; case 2: dotted lines, $q = 0.5$). Shading represents the range of solutions within these limits. Calculated from r.m.s. differences of wind averaged between 20° and 82.5°N and 1000 and 200 mb.
between the wind and height results regarding the relative importance of analysis error and model error as a function of forecast range, so we show only the results for the wind. The area between the curves for both cases is shaded to indicate the range of the results. In discussing Fig. 17 we may distinguish three forecast ranges. In the short-range forecasts (up to day 1) analysis errors predominate, while the model errors grow rapidly in importance. In the early medium-range forecasts (days 2 to 5) the model error predominates over analysis error; the latter typically amounts to 30% to 50% of the total forecast mean square error, while the former amounts to 60% to 80%. In the later medium-range (after day 5) the analysis error increases in relative magnitude. In the case of an unbiased model, the normalized model error and average analysis error would asymptote to 1. The asymptotic level of the analysis error in our calculation is determined by the asymptotic level of the forecast divergences. As discussed in section 5(b), the forecast divergences asymptote near day 10 to a level that is significantly lower than the level of a persistence forecast.

The predominance of model error in the early medium-range forecasts gives grounds for optimism that model improvements will lead to significant improvements in forecast skill. To estimate the potential for such improvement we consider Fig. 18, which shows the average total r.m.s. forecast error, and the average r.m.s. analysis error for our two cases. Following the arguments of Lorenz (1982), the horizontal distance between the forecast curve and the analysis curves indicates the potential for improvement in forecast skill by elimination of model error. The distance between the two analysis error curves would vanish at day 12·5 since we assume that the average correlations reach their asymptotic values then. In the early medium-range (days 2–5) Fig. 18 indicates that in the case with larger initial analysis errors (case 2) the expected gain in predictive skill in a r.m.s. sense would vary from 1·5 days at day 2 to 2·5 days at day 5. In case 1, which has lower initial analysis errors, the corresponding figures are 2 days at day 2 and 4 days at day 5. It should be borne in mind that a halving of the r.m.s. model error would realize much of these gains between day 2 and day 5, since the model error is so dominant in the total error field.

The two components of the forecast error have quite different growth characteristics in time, as can be seen in Figs. 16 to 18; and these characteristics are not very sensitive to our assumptions about the analysis error correlations. In the early medium range the r.m.s. model error grows linearly in time. As we have seen in the last section, the forecast divergence data have a marked exponential growth in the early medium range. Since the forecast divergence data control the average analysis error in our calculation, the average analysis error also has a marked exponential character in the early medium range, with doubling times of order 2 days in the height field in the early forecasts.

We note that unresolved small-scale motion may contribute to forecast errors on resolved scales after some days but will not contribute to forecast divergences. Such error sources would be attributed to model errors in our calculations.

The result that the model error grows approximately linearly in time is significant. Linear error growth suggests a constant source of error, independent of the error magnitude. Such an interpretation is consistent with the results of Wallace et al. (1983) on topographic forcing in the EC model, and with the results of Wergen (1984), who showed how errors in the tropical convective forcing in the EC model can lead to approximately linear growth of large-scale errors in the tropics. It is unlikely that this list of phenomena giving rise to linear model-error growth is exhaustive; further research will probably identify others. The fact that the r.m.s. model-error growth is linear in time gives grounds for optimism that the major sources of model error can eventually be identified, and their effects reduced.
Better forecast models will obviously give more useful medium-range forecasts but for short-range forecasts and for forecasts beyond day 5 we also require improvements in the analyses. As indicated by Leith (1981), analysis quality is dependent on the quality of the model in the assimilation scheme, so model improvements will probably lead also to analysis improvements.

Our solution of the budget equations requires that we specify the average analysis error intercorrelation, \( q(t) \); the solution then defines the time evolution of the individual analysis error intercorrelations. For both height and wind, the largest correlation always occurs for the errors of the UK and EC analyses. This is because these ensembles of forecasts show the lowest rates of divergence from each other and because the forecasts with the UK analyses have the largest total errors. In the mass field, the smallest correlation is always the US/EC correlation; while in the wind field the UK/US and US/EC correlations are approximately equal. This is because the largest rates of divergence are seen between the US-based forecast ensemble and one of the other two ensembles. For case 1 the correlation of analysis error for the wind fields in the early forecasts is 0·30 between EC and UK, and 0·20 between the other pairs; in case 2 the corresponding values are 0·60 and 0·45. We can only speculate about the source of the high correlation between EC and UK analysis errors; the most important aspects that set the US analyses apart from the other two are, possibly, the lower resolution of the assimilating model and the use of SATEMs as heights rather than thicknesses or temperatures.

The method we have used above to discuss the error budget for certain ranges of the correlations is a simple form of linear programming. We have examined the equations using somewhat more general techniques, imposing reasonable constraints that variances be positive, that correlations be less than 1 etc. The presentation used above is simpler and encompasses the main results; we believe these results to be robust in the sense that they cover the most relevant part of the parameter space for which solutions exist.

It is worth noting that an approximate solution to the budget equations can be found as follows: use Eq. (4) to define an average analysis error \( \bar{A}E \), and approximate the covariance terms in Eq. (3) by

\[
2(\text{VMA} + \text{VMB} + \text{VMC})/3 = 2z(M^2)^{1/2} \bar{A}E
\]

where \( z \) is the specified model/analysis error correlation. The approximation permits a simple calculation for the model error: we use the approximation (6) to simplify Eq. (3) so that all unknowns except \((M^2)^{1/2}\) can be replaced from Eq. (4). This simplification reduces Eq. (3) to a simple quadratic which is easily solved for \((M^2)^{1/2}\). This solution for \((M^2)^{1/2}\) is accurate to better than 5% for the first 6 days.

7. CONCLUSIONS

The purpose of our two-part study has been to assess the significance and implications of the differences between analyses made by the EC, UK and US analysis systems using the same FGGE level IIb observational dataset. An important feature of these inter-comparisons has been that each analysis system was presented with an identical set of observations, and the impact of the analysis differences was studied on each model. Thus we have isolated the impact of differences in analysis technique. In part I we discussed their impact on analyses and the implications for forecasts from the synoptic point of view. In part II we have made a statistical study of the impact of the differences in analysis method on medium-range forecasts. This study highlighted differences in the analyses which are not always apparent in the synoptic studies.
We began with a study of the energetics of the forecast ensembles from the two models and three sets of analyses. The UK analyses have lower baroclinic conversion rates. In the EC model forecasts this causes an initial increase of zonal available potential energy and a decrease of eddy kinetic energy. Later one finds an overreaction of the model with too high energy conversions and too high eddy kinetic energy in the baroclinic waves. In the US model it was shown that the changes of resolution in the course of the forecasts have a marked effect on the energy cycle. The EC model was shown to have a systematic tendency to decrease the kinetic energy of the long waves.

We examined the forecast skill of the EC and US models using different analysis ensembles. It was shown that in terms of importance for the forecast skill, the differences between the EC and US models outweighed the differences in the analyses. The EC model was clearly superior to the US model, which is consistent with findings by Baumhefner (1983), and Brown (1983) and could have been expected because of the differences in resolution and complexity of the physical parametrizations.

Forecasts using the EC and US models diverged as rapidly from each other as the EC forecasts diverged from the analyses, and we saw little evidence of forecast problems in mid-latitudes arising because of the barotropic phenomena discussed by Daley et al. (1981). The rapid divergence of the forecasts from the two models may be due to the differences in resolution as discussed by Cullen et al. (1981).

In terms of sensitivity to the EC and US analyses, the US model showed more sensitivity than the EC model. Both models performed better at 500 mb when running from the EC analyses. At 1000 mb the US model gave similar results from both analyses, while the EC model performed better from the US analyses. The forecast verifications showed that the UK-based forecasts scored poorly, relative to the others, in the northern hemisphere and scored well in the southern hemisphere.

A study of the divergence between forecasts from different analyses with the EC model showed that the doubling time of the analysis differences in the early forecasts was 2-0 days in the height field, and 2-6 days in the wind field. There was a marked scale dependence in the results, with the baroclinic waves showing about 0-25 days shorter doubling times. To test if the small sample over a short period we used was atypical we repeated Lorenz's growth rate calculations using forecast errors as initial perturbations. The growth rate calculations reproduced his results almost exactly.

To calculate the growth rates of forecast divergence we used linear and nonlinear fitting curves, which yielded similar values for the early forecast ranges. The nonlinear method suggested by Lorenz gives a useful description of the variation of growth rate with amplitude. We used this model to estimate the asymptotic level of forecast divergence and found that the differences are close to this level over most of the spectrum by day 10. The forecast divergences reach a specified fraction of the asymptotic level of persistence in less than half the time found by Shukla (1981). Consequently Shukla's estimates of the potential for predictability, based purely on dynamical processes and ignoring all boundary forcing effects, may be too optimistic because his perturbations of the initial analyses were probably unrealistic. Differences in the models and differences in the synoptic situations may be important too for the discrepancies between both studies.

We used the forecast verification data and the forecast divergence data, together with some simplifying assumptions, to partition the forecast error into a part arising from model error and a part arising from analysis error. The r.m.s. contribution due to model errors grew linearly between day 1 and day 6, while the contribution arising from analysis error grew approximately exponentially in the same period. The model error was the dominant source of forecast error between days 2 and 5. The analysis error was
the dominant source of forecast error in the short-range forecasts, and made an important contribution after day 5. Substantial gains in predictability could be made between day 2 and day 5 through the elimination of model error. A halving of the r.m.s. model error would realize a large part of this potential. These results support the conclusions of Lorenz (1982) on the potential for forecast improvements in the early medium range. Given the current level of model error, a halving of the analysis error would improve the short-range forecasts, and would have an impact on forecast skill after day 5. However, an increase in analysis error would have a very adverse effect on forecast skill because of its exponential growth rate. With a perfect model a halving of the analysis error would lead to a further increase of predictability of about 2 days.

It is difficult to estimate the potential for improvement in forecast skill beyond day 7 because of a systematic error in the model which underestimates the atmospheric variance in the later stages of the forecasts. If this error could be corrected, a further study would be worthwhile. The essential question is whether the correction of this error would lead to the forecast divergences asymptoting at a higher level, closer to the persistence error, at day 10, or whether the divergence between forecasts would grow at the rate we have observed, and reach their asymptotic levels much later.

In both parts of this study various types of analysis differences have been identified which affected the subsequent forecasts in different ways at different time scales. The results of the case studies in part I indicated that the analysis differences had most impact on the forecasts if they occurred in baroclinically unstable regions. For the objective medium-range verification scores used in part II the baroclinic eddy activity is also very important, as these scales of motion show the most rapid growth rates, and contribute most to the growth of forecast differences.

Although we had only a small sample available, we can conclude from the forecast scores and the energetics that for improving medium-range forecasts it is not sufficient to improve the forecast models but emphasis should also be given to the observation and analysis of baroclinically active regions. The results of part I suggest that this is most important over the oceans.

APPENDIX

Discussion of skill scores

The anomaly correlation is sensitive to phase differences. In the extreme case where the predicted and observed field consists of waves of one wavelength only, say $A \cos(nx)$ and $B \cos(nx + dx)$ respectively, the anomaly correlation coefficient reduces to

$$R = 100 \cos(dx) \text{ (\%)}$$  \hspace{1cm} (A1)

showing no dependence on the amplitudes $A$ or $B$. To see the dependence of the score on the magnitude of the error, one has to separate the error into the contribution which is in phase with the observed anomaly and the one which is $\pi/2$ out of phase. The contribution which is in phase has no impact on the correlation, as shown by Eq. (A1). The error contribution out of phase, $Dx_s$, can be related to the anomaly correlation, when $Dx_s$ is small, by

$$R = \cos(dx) = \{1 - (Dx_s/B)\}^{1/2} \approx 1 - 0.5(Dx_s/B).$$  \hspace{1cm} (A2)

This result indicates that the anomaly correlation score depends not only on the magnitude of the out-of-phase component, but also on the magnitude of the anomaly one is trying to forecast. The suggestion is that for a given magnitude of the forecast error the anomaly correlation will be large if the anomaly is large.
If there is a linear increase (with time) of the phase error of a single wave, then the graph of the anomaly correlation coefficient is a cosine curve decreasing from 100% to −100%, with largest slope at the 0% level. Miyakoda et al. (1972) show (for an unbiased model) that with many waves of random phase, the asymptotic value will be 0% instead of −100%. The strongest decrease is then around the value of 50%. This last point suggests enhanced sensitivity in the score near the 50–60% value, which is often used to indicate the limit for useful predictive skill (Hollingsworth et al. 1980). Root mean square errors or standard deviations have the advantage of responding linearly to error growth.

Unlike the anomaly correlation coefficient, the r.m.s. score has the disadvantage that it favours forecasts which underestimate atmospheric variability. Its asymptotic value after the loss of predictive skill varies considerably with synoptic situation, so it is not terribly useful for defining a limit of useful predictive skill. For a large sample the asymptotic r.m.s. value of an unbiased model tends to \( \sqrt{2} \) of the r.m.s. variability of the atmosphere about its climatological mean (Miyakoda et al. 1972). This relation has been used by Baumhefner (1983) to define a normalization which may help to provide a practical definition of the limit of useful skill.

ACKNOWLEDGMENTS

We are grateful to the Directors of our respective institutions, Dr Lennart Bengtsson, Dr William Bonner and Sir John Mason, for their support and encouragement in the course of the work. We are grateful to Andrew Gilchrist, Adrian Simmons, David Burridge, Norman Phillips, Ronald McPherson and Michael Tiedtke for valuable discussions. We are grateful for the assistance of many individuals at ECMWF, NMC and the Meteorological Office. Particular thanks are due to Geoffrey DiMego, Clifford Dey, Lauren Morone, John Stackpole and Carl McCala. The research was partly funded at NMC by the NOAA Office of Climate and Atmospheric Research.

REFERENCES


Daley, R.

Daley, R., Tribbia, J. and Williamson, D. L.

Hollingsworth, A.

Hollingsworth, A., Arpe, K., Tiedtke, M., Capaldo, M. and Savijarvi, H.

Hollingsworth, A. et al.

Leith, C. E.

Lorenz, E. N.

Miyakoda, K., Hembree, G. D., Strickler, R. F. and Shulman, I.

Robert, A.

Shukla, J.

Smagorinsky, J.

Taction, M. S.

Wallace, J. M., Tibaldi, S. and Simmons, A. J.

Wergen, W.


1981 Predictability experiments with a baroclinic model. *Atmosphere-Ocean*, 19, 77-89


1980 'An experiment in Monte Carlo forecasting'. ECMWF workshop proceedings on stochastic dynamic forecasting, 17-19 October 1979, 65-86


1981 'Statistical methods for the verification of long and short range forecasts'. In proceedings of the ECMWF seminar on problems and prospects in long and medium range weather forecasting, 14-18 September 1981, 313-333


1982 Atmospheric predictability experiments with a large numerical model. *Ibid.*, 34, 505-513


1983 Reduction of systematic forecast errors in the ECMWF model through the introduction of an envelope orography. *Quart. J. R. Met. Soc.*, 109, 683-717

1984 'Forced motion in the tropics'. ECMWF workshop on current problems in data assimilation, 8-10 November 1982, 183-206