

Jennifer C. Roman
University of Utah, Salt Lake City, Utah

Gonzalo Miguez-Macho
Rutgers University, New Brunswick, New Jersey

Lee A. Byerle and Jan Paegle *
University of Utah, Salt Lake City, Utah

1. INTRODUCTION

A number of atmospheric forecast models have been developed and now display considerable skill in weather prediction. The underlying philosophy of these developments is improved models and more accurate initial conditions should provide better forecasts. The importance of specific model improvements relative to specific observational enhancements may, nonetheless, still be inadequately understood. White et al. (1999) addressed some of these questions, and suggested the error spread among models of very different configuration and resolution is generally less than the magnitude of the error in any single relatively advanced model. Increased resolution in regional models does lead to some improvement in skill, particularly for heavier categories of precipitation that are not simulated at cruder resolution. This benefit, however, appears to be relatively small for other forecast variables, and is evident only for short-term forecasts in point validations against observations, even in the most highly resolved and dynamically sophisticated approaches.

Various hypotheses have been proposed to explain this result. Some of these note current observing systems contain inadequate resolution of local regions of pronounced dynamic instability. Other explanations of the marginally superior performance of high-resolution models point to the evident difficulties in validation of forecast features against coarsely spaced observations. All but one of the models studied by White et al. (1999) were limited area models and these are known to be strongly influenced by lateral boundary conditions supplied at their perimeter (Warner et al. 1997; Paegle et al. 1997). The Medium Range Forecast (MRF) model was the only global model evaluated by White et al. (1999), and that model also provided, or strongly influenced, the lateral boundary conditions of the tested limited area models.

The initial stages of the present research compared

two global models (the Utah Global Model [UGM] and the MRF), and followed a hypothesis suggested by Miguez-Macho and Paegle (2000, MMP henceforth). This perspective, which is based upon down-scale uncertainty growth is supported by MMP and by Miguez-Macho and Paegle (2001), and is rooted in early barotropic model predictability studies of Thompson (1957) and Lorenz (1969). Results obtained by MMP contrast with other recent literature emphasizing the importance of relatively smaller-scale instabilities (e.g. Palmer et al. 1998, Hartmann et al. 1995) of the initial state, and other local error sources (e.g. Rabier et al. 1996). These perspectives and contrasting theories are summarized by MMP, who presented preliminary work suggesting that the dominant source of short-term forecast errors may be the uncertainty of relatively large scales of the initial state.

The present research extends MMP's studies to sort out the relative roles of modeling technique and initial state uncertainty and continues to probe the limitation to deterministic weather prediction due to inadequate observation of relatively large scales of the atmosphere. Other global model studies attempting to prioritize the relative contributions of initial errors, boundary errors, and model errors to total forecast error include work by Reynolds et al. (1994) and Hacker et al. (2003). Simmons and Hollingsworth (2002) show substantial improvement in forecast accuracy in global operational models over the past decade, and present an extensive list of forecast system changes which may have contributed to forecast improvement over the past two decades.

The first goal of the present research is to compare UGM and MRF forecasts in more detail. Our initial studies of this problem focused on a 17 case forecast sample during winter 1993, and compared research model forecasts executed at horizontal wavenumber 42 triangular truncation with MRF forecasts at horizontal wavenumber 62 triangular truncation. Those results were consistent with the possibility that uncertainty of the initial state represented the greatest limitation to forecast accuracy, and implied that details of model formulation were not equally important. That suggestion is supported in Fig. 1 which displays globally averaged bias-corrected anomaly correlations of the 500-mb height forecasts made by the

* *Corresponding author address:* Jan Paegle, University of Utah, Department of Meteorology, 135 South 1460 East, Room 819, Salt Lake City, UT 84112-0110; e-mail: jpaegle@met.utah.edu

UGM and MRF with analyses, and also correlations of UGM forecasts with MRF forecasts. Bias corrections are

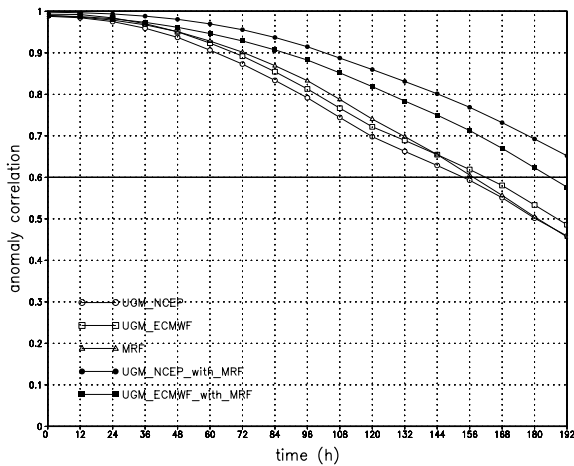


Fig. 1. Time evolution of global, case-averaged anomaly correlations of the 500-mb geopotential height, for the UGM initialized with NCEP (open circles) and ECMWF (open squares) reanalyses and the MRF (triangles), each correlated with analyses, and for the UGM initialized with NCEP (closed circles) and ECMWF (closed squares) reanalyses, each correlated with the MRF.

applied by extracting the case average error from the individual forecasts:

$$F^n - \frac{1}{N} \sum_{n=1}^N (F^n - O^n) \quad (1)$$

where F^n is the forecasted variable and O^n is the analyzed variable for run number n , and N is the total number of runs in the set of cases (for Figure 1, N is 17). The model bias may also be calculated using a “training period” applied to subsequent independent cases. For example, the bias can be computed from the average error for the 30 days prior to a given date, and then applied to that date. This “a priori” method will be employed in Section 3.

The bottom three curves of Fig. 1, based on calculations performed by MMP, display bias-corrected anomaly correlations obtained with the two research model forecasts initialized separately with NCEP and ECMWF reanalyses, and of the MRF model initialized with NCEP reanalyses. As expected, the MRF model is more accurate than the research model globally, but only by a relatively small amount. The bias-corrected UGM lags the bias-corrected MRF model predictions by only about 12 hours, and both show similar skill after 7 days when the anomaly correlations drop below 0.6 for each model. When calculating model errors used for statistics, fore-

casts from each model were not truncated to the same resolution. Although this may have unfairly penalized the higher resolution model (MRF) which has more fine-scale structure, the MRF still outperforms the UGM in nearly all comparisons, and without the bias correction, the MRF forecasts retain predictability skill 24 hours longer than UGM forecasts. The top two curves of Fig. 1 display bias-corrected anomaly correlations of two UGM forecasts (initialized with NCEP and ECMWF reanalyses) validated against the MRF model forecasts. It is noteworthy that the anomaly correlation of the UGM forecasts with the MRF model predictions are higher than the anomaly correlation of either model with the analysis. One interpretation of these results is the research model can be used to predict the behavior of the MRF model with more accuracy than either model can be used to predict the state of the atmosphere. At 8 days, the research model anticipates the MRF model evolution about as well as either model anticipates the analyzed evolution at only 6 days. Another interpretation is the error fields of the MRF and UGM exhibit strong correlation through 8 days, implying a similar error source for both models.

The UGM retains lower horizontal resolution and fewer levels than the MRF model, and has undergone much less extensive development and calibration. The fact that a relatively unsophisticated and more crudely resolved model possesses such high skill in anticipating the behavior of a much more sophisticated model implies that something other than model complexity provides the fundamental limitation in forecast accuracy for this set of 17 individual cases of winter 1993. A possible inference is that both models suffer from the common deficiencies of imperfect initial state specification, and this may represent the primary forecast limitation for these cases. This inference carries important implications for deployment of observing systems and for forecast model development, and merits close scrutiny.

Operational model sophistication, resolution, and initial state specification have undergone substantial improvement in the past decade (Simmons and Hollingsworth 2002). The second goal of the current research is to evaluate this improvement using an up-to-date version of operational model products from NCEP and to compare these results to the research model. The comparison allows a crude estimate of value added by the improved observing system and data assimilation system since 1993, in comparison to model enhancements. Section 2 describes the datasets and models used in the prior 1993 comparisons, and more recent datasets and models in real-time comparisons. Section 3 presents updated diagrams similar to Fig. 1 and quantifies comparisons of recent research and operational products. These results

suggest extension of forecast skill by approximately 6 hours in the low-resolution UGM initialized with winter 2003 analyses, and a more substantial extension of approximately 36 hours in a more developed, high-resolution version of the MRF model, due both to initial state and model improvements.

There is, therefore, evidence for the important positive role of both improved initial states as well as improved model resolution and formulation. Error diagnostics still suggest strong similarity of error structure, consistent with the possibility that initial state error is yet a strongly limiting element. Section 5 continues past studies by MMP of the rate of growth of initial state uncertainty. MMP's experiments were limited by several factors. First, they were integrated for only 5 days, and it was not possible to establish the time scale on which the uncertainty saturates. Second, they did not demonstrate the relative sensitivity of the forecasts to imposed initial uncertainty in comparison to the sensitivity to model formulation. The integrations were also performed at a relatively coarse wavenumber 42 resolution, which could have limited error growth.

Therefore, a third goal of the present study is to perform a series of further predictability studies with the UGM and compare the results with those from the MRF, with experiments to address each of these issues. In particular, all wavenumber 42 UGM experiments performed by MMP are now extended from five days to two weeks duration; two sets of experiments impose doubled horizontal resolution, and the sensitivity of forecast evolution to the model used is now compared with the sensitivity to initial state uncertainty.

The study is organized as follows. Section 2 presents brief overviews of the UGM and datasets. Section 3 describes anomaly correlations of the UGM predictions in comparison to predictions by the MRF model for a winter period in 2002-2003. Section 4 quantifies the sensitivity of global rms estimates of height and wind forecast changes produced by switching between available models and switching between available initial states, and compares both to analyzed and forecasted changes and to the errors of the changes. Section 5 studies sensitivity to initial state uncertainty and section 6 presents conclusions.

2. DATASETS AND MODELS

a. Datasets

The National Centers for Environmental Prediction (NCEP, Kalnay et al. 1996) and the European Centre for Medium-Range Weather Forecasts (ECMWF, Gibson et al. 1997) have performed gridded retrospective analyses, based upon all available observations by a frozen

state-of-the-art global data assimilation system. Present estimates of initial state uncertainty are obtained from the difference of these two equally credible analyses, and are assumed to be reasonable for this study. However, it is likely that this method of characterizing initial state uncertainty underestimates actual values at all scales. In fact, NCEP reanalyses have been truncated at wavenumber 36, and, to the degree that ECMWF and NCEP use the same observations, the difference in their analyses will underestimate the total error. This limitation to the study is further discussed in Section 5.

The 17 cases previously used by MMP were selected for model initialization. These start on 1 January 1993, and continue at 5-day intervals through March 1993. The particular dates are chosen because 8-day predictions by the version of the MRF used in the reanalyses are available within the NCEP reanalysis archive. This allows comparison of the UGM research model with the MRF model, which is well-documented within NCEP technical reports.

In addition to the earlier 1993 ensemble, a set of 52 cases was selected from winter 2002-2003 from operational NCEP analyses that were available in near-real time. These analyses are archived at 2.5 degree resolution on 26 vertical levels and extend from 4 January 2003 to 26 February 2003 (two dates were excluded). The set of 52 cases was selected based upon continuous analysis data availability to compute bias statistics, allowing an a-priori bias extraction based upon a training period of 30 days duration prior to each UGM and MRF prediction. All errors and anomaly correlations shown subsequently represent bias-corrected values for both the UGM and the MRF unless otherwise stated.

b. Models

The UGM is based upon Galerkin approximations applied separately in each spatial dimension. Latitude and vertical structure are depicted by finite elements and longitude variability by Fourier series. The method retains the high accuracy and conservative properties of alternative Galerkin approximations, such as those used in the MRF that are based upon spherical harmonic expansions.

The dynamical core of the model uses a hydrostatic set of primitive equations in which vorticity, divergence, and thermal fields are predicted on pressure-based sigma-coordinates. This approach is similar to that used in global models in operational centers, with the exception of numerical methods outlined above. Model physical parameterizations of convective and stable precipitation are similar to those used by the NCAR Community Climate Model 1 (Bath et al. 1987) introduced in 1987, and parameterizations of radiative and surface processes are also relatively less advanced and

follow methods that have been used by other models more than two decades ago. Vertical mixing coefficients are calculated from a low-order turbulent energy equation, and radiation processes include cloud radiation interactions, as described for another model by Nicolini et al. (1993). Moist convective processes use simple convective adjustment, and cloud fraction and condensation criteria are based upon local relative humidity without explicit treatment of cloud microphysics. Sea-surface temperature is maintained at the initial value, and land surface evaporation is extracted from the daily analysis files. Each of these processes is treated in a more modern fashion within the MRF. Present applications retain 20 vertical levels and wavenumber 42 resolution. One set of experiments was repeated at wavenumber 84 resolution.

The UGM was originally designed by Paegle (1989) to address predictability questions. It has been used to study impact of wind data voids on objective analyses (Paegle and Horel 1991); for predictability work (Vukicevic and Paegle 1989; Paegle et al. 1997; Miguez-Macho and Paegle 2000, 2001); for idealized global simulations of tropical-extratropical interactions (Buchmann et al. 1995); to study orographically forced regional circulations (Nogues-Paegle et al. 1998; Byerle and Paegle 2003); and for initial data impact investigations of the 1993 “storm of the century” (Miguez-Macho and Paegle 1999a, 1999b).

The MRF model used in the 1993 intercomparisons corresponds to the version used operationally at NCEP until about 1995, and is truncated (triangular) at horizontal wavenumber 62 on 27 vertical levels. The acronym for the MRF changed to GFS (Global Forecasting System) recently, but products from the real-time NCEP forecasts used for the winter 2002-2003 cases will continue to be referred to as MRF predictions in this study. These forecasts were executed at wavenumber T254 resolution for the first 3.5 days and subsequently at wavenumber T170 resolution with triangular truncation, on 64 vertical levels (Lord, personal communication). The MRF is a spectral model, based upon a set of primitive equations similar to those used in the UGM.

3. ANOMALY CORRELATIONS

Figure 2 compares anomaly correlations for the recent 52 case real-time forecasts of winter 2002-2003 with those for the 1993 experiments. The anomaly correlations of forecasts calibrated against observations (in this case, analyses) are computed as:

$$\frac{(\overline{S - S_{cl}})(\overline{S_{ob} - S_{cl}})}{[(\overline{S - S_{cl}})^2(\overline{S_{ob} - S_{cl}})^2]^{1/2}} \quad (2)$$

where S_{cl} represents the climatological value of S , and S_{ob} is the analyzed value, in this case from Global Data Assimilation System analyses. Here $\overline{(\quad)}$ indicates a global or hemispheric average, and the 50-year (1951-2000) monthly average derived from NCEP reanalyses is defined as climatology for this purpose. S represents the forecasted variable from the UGM and MRF. An a priori bias correction is applied to forecast variables using biases calculated from the 30 daily forecasts available immediately prior to the day of interest. The anomaly correlations of UGM forecasts calibrated against MRF forecasts is given by a similar formula, with S replaced by the UGM prediction and S_{ob} by the MRF prediction. Anomaly correlations of the UGM forecasts with the MRF model predictions exceed anomaly correlations of UGM forecasts with analyses when errors of the UGM possess similar structure to errors of the MRF.

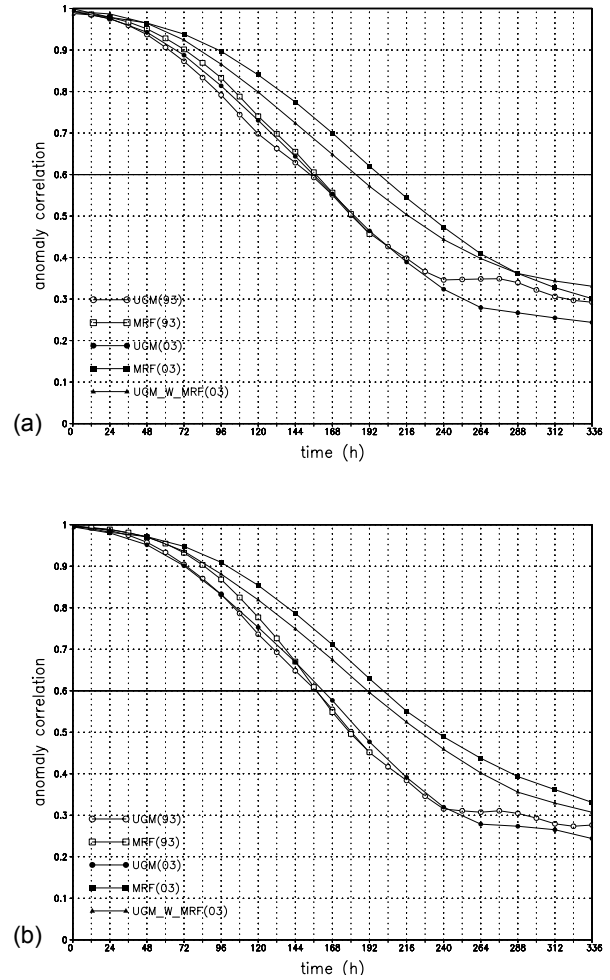


Fig. 2. (a) Time evolution of global, case-averaged anomaly correlations of the 500-mb geopotential height, for the 1993 cases of the UGM (open circles) and the

MRF (open squares), and for the winter 2002-2003 cases of the UGM (closed circles) and MRF (closed squares), each correlated with analyses; and for the winter 2002-2003 cases of the UGM correlated with the MRF (closed triangles). (b) Same as in (a) but averaged over the latitude band from 20° to 90° N.

The curves with solid squares in Fig. 2 represent anomaly correlations of MRF predictions with analyses for the 52 selected cases from winter 2002-2003. Substantial improvement is seen with respect to similar results for the 1993 ensemble (curves with open squares). Bias-corrected MRF anomaly correlations for winter 2002-2003 drop below 0.6 after approximately 192 hours, compared to approximately 156 hours for the 1993 cases, corresponding to a 36 hour gain in predictive skill at this level by the MRF.

The UGM also produces more skillful forecasts in 2003 than it does in 1993. In the winter hemisphere (Fig. 2b), UGM forecasts produce anomaly correlations of 0.6 to approximately 162 hours (curve with solid circles), and represent 6 hours longer skill than for the 1993 ensemble (open circles).

The MRF has clearly benefitted more from advances made at NCEP during the past decade than has the UGM. In the case of the MRF, these advances include substantially higher horizontal and vertical resolution in addition to improved initial state analyses. The only improvement in the UGM is the availability of more modern analyses of the initial state, since UGM resolution and physical formulation are frozen between the 1993 and more recent applications displayed in Fig. 2.

A superficial examination of Fig. 2 suggests that observation and/or analysis improvements have produced only a 6 hour improvement in forecasts over the past decade, since the frozen UGM exhibits only 6 hours forecast improvement in that period. It is unclear, however, how much of the additional 30 hours of forecast improvement in the MRF is due to initial state advances relative to model advances over the past decade. A substantial portion of that model's improvement may be due to improved specification of the initial state in addition to model advances. It is also possible that winter 2003 was more predictable than winter 1993 and this may improve recent skill scores in all models.

The curves with triangles in Fig. 2 represent anomaly correlations of the UGM forecasts validated with forecasts made by the MRF for the winter 2002-2003 cases. This curve is only slightly below the anomaly correlation of the MRF forecasts with respect to analyses (curve with solid squares) in the winter hemisphere (Fig. 2b). Consequently, the low-resolution UGM forecasts have almost as much skill in predicting the MRF forecasts as

the latter have in anticipating the state of the atmosphere. This implies correlation of error patterns of forecasts made by the UGM and by the operational model. The similarity of error patterns of two very different models suggests common limitations, including limitations associated with use of the same initial state.

Fig. 3 shows an example of the bias-corrected error fields for the 500-mb meridional wind for day 14 of forecasts made by the UGM (Fig. 3a) and MRF (Fig. 3b).

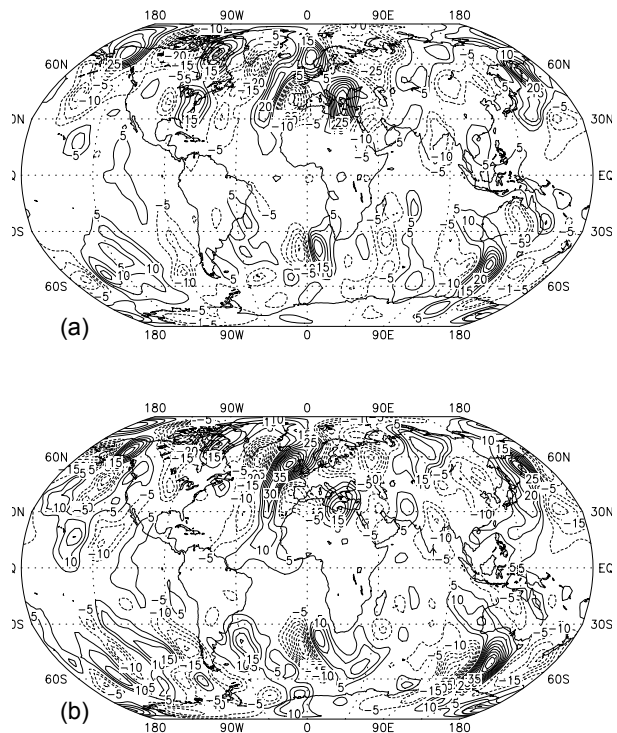


Fig. 3. Bias-corrected error-fields for the 500-mb meridional wind at forecast hour 336 for the (a) UGM and (b) MRF, both initialized at 00Z on 26 February 2003. Contour interval is 5 m s⁻¹.

Accurate prediction of the meridional flow is important for accurate depiction of wave amplitude and phase, and this variable is more representative of synoptic scale features than geopotential height, which carries most amplitude on global scales. Error field structural similarities in the 500-mb meridional flow are particularly evident in northern Canada, over the Eurasian subcontinent, and the eastern coast of Asia in the northern hemisphere. In the southern hemisphere, two strong error field couplets exist off the southwestern coasts of Africa and Australia in both forecasts. Correlation coefficients (discussed below) for the bias-corrected error fields shown in Fig. 3 are 0.61 for global and northern hemisphere averages,

and 0.62 for the southern hemisphere average. Similar agreement in error patterns of the two model forecasts is evident on many other days.

These error fields are computed after removing from each model the a priori 30-day average biases, shown in Figs. 4a and 4b. The bias fields for 500-mb meridional

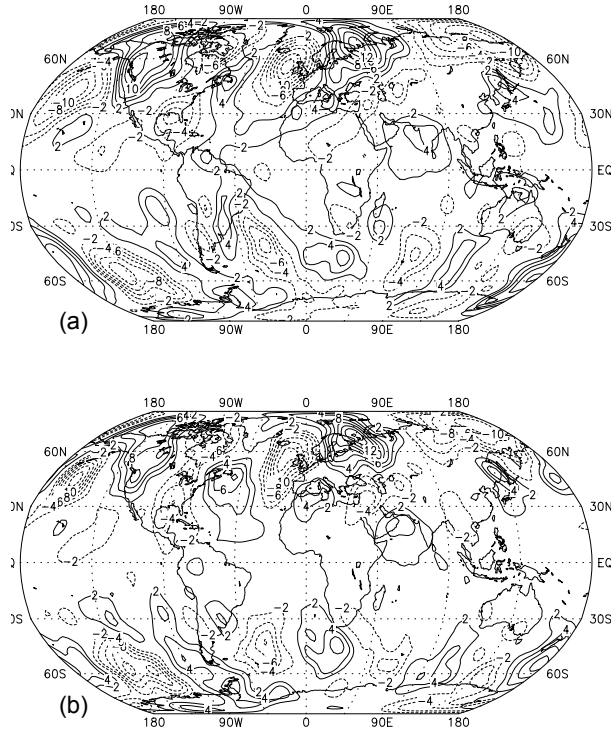


Fig. 4. 30-day average bias fields for the 500-mb meridional wind at forecast hour 336 for the (a) UGM and (b) MRF, for the 30 days prior to 26 February 2003. Contour interval is 2 m s^{-1} .

wind also display remarkable similarity between the UGM (Fig. 4a) and MRF (Fig. 4b) at day 14, despite the difference in model complexity and resolution, implying some common deficiency. Correlation coefficients for the 30-day average bias fields shown in Fig. 4 are 0.74, 0.72, and 0.81 for the global average, northern hemisphere and southern hemisphere, respectively clearly exceeding the asymptotic correlation limit of 0.5. Furthermore, the overall magnitudes of the 30-day average meridional flow biases from the models (Figs 4a and 4b) are comparable to the magnitude of the 30-day average 500-mb meridional wind analysis (Fig. 5a). The model biases, therefore, are not negligible.

Comparison of departures from the zonal mean of 30-day averages of 500-mb forecasted heights from the UGM (Fig. 6a) and MRF (Fig. 6b) with the departures

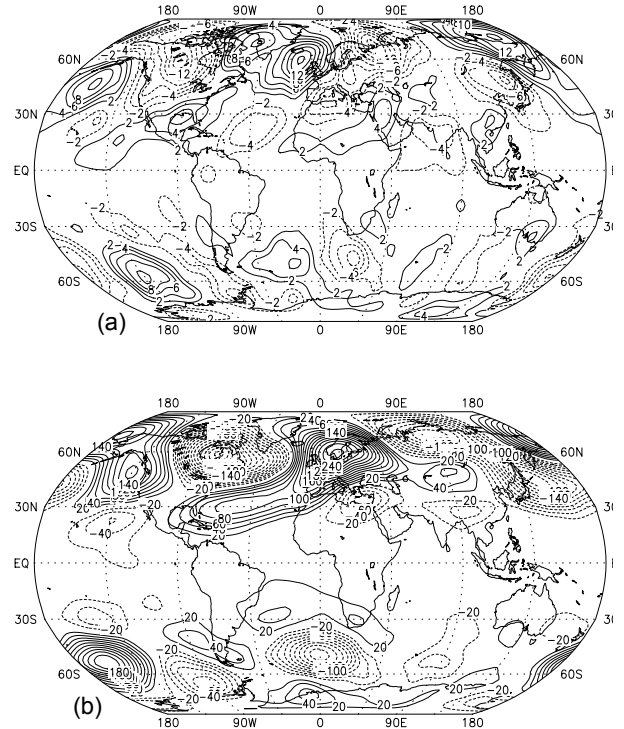


Fig. 5. The 500-mb (a) meridional wind flow and (b) departure from the zonal mean height analysis, averaged over 30 days prior to 26 February 2003. Contour interval is 2 m s^{-1} in (a) and 20 m in (b).

from the zonal mean for 30-day average 500-mb heights from the reanalysis (Fig. 5b) also shows a stronger similarity between model forecasts than between either model and the reanalysis. Both the UGM (Fig. 6a) and MRF (Fig. 6b) tend to forecast a ridge over the northern Rockies, while the reanalysis (Fig. 5b), indicates actual ridge placement off the western coast of the U.S. and Canada. Additionally, neither model forecasts the blocking pattern over Europe, and both models underforecast the amplitude of most of the wave structure in the 500-mb pattern.

To further quantify the similarities, Fig. 7 displays correlation coefficients of the error and bias of the 500-mb meridional flow forecast by the UGM correlated with the error and bias of the same variable forecast by the MRF:

$$\frac{\overline{(S_{UGM})(S_{MRF})}}{[(\overline{(S_{UGM})^2})(\overline{(S_{MRF})^2})]^{1/2}} \quad (3)$$

where S_{UGM} represents either the error or bias of the 500-mb flow forecasted by the UGM, and S_{MRF} is either the error or bias of the 500-mb flow forecasted by the MRF. Again, $\overline{(\quad)}$ indicates a global or hemispheric average. This measure of correlation asymptotes to a non-

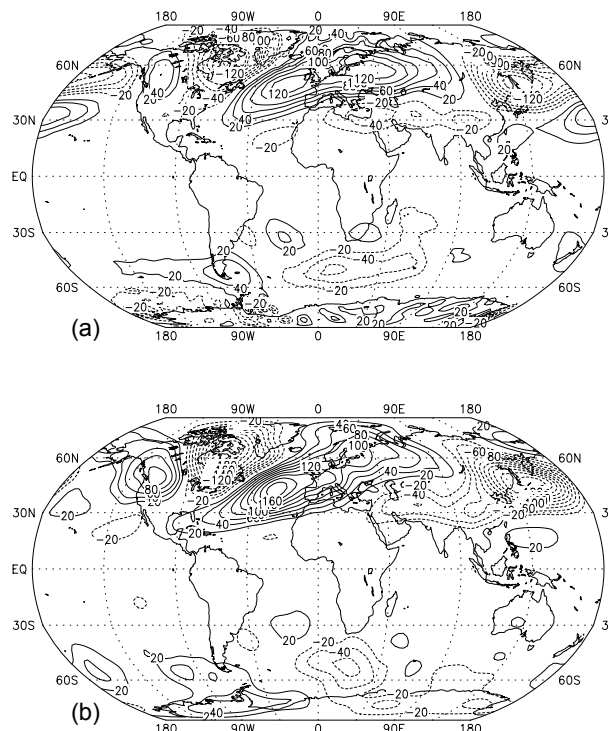


Fig. 6. Departures from zonal mean 500-mb height field at forecast hour 336 for the (a) UGM and (b) MRF, averaged over 30 model runs prior to 26 February 2003. Contour interval is 20 m.

zero value at long times, when predictability is lost due to initial state uncertainty and the chaotic nature of the atmosphere. The time evolution of the case-averaged error field correlation coefficients for a 30-case sample in winter 2003 is shown in Fig 7a. The fact that the correlations of errors continue to rise throughout the 14 day forecasts despite initial data uncertainties and chaos may be explained by the growing similarities in model bias between the UGM and MRF (Fig. 7b). As yet, however, it is unclear to what the similarity in bias structure can be attributed.

The similarity of the UGM and MRF forecast and bias errors suggests that models characterized by very different resolution and physical parameterizations still may possess a common, fundamental deficiency. Both the UGM and MRF are based upon primitive equation dynamical cores and both incorporate hydrostatic and incompressibility assumptions. In order to ascertain whether the dynamical simplifications inherent to primitive equation approaches may produce similarity in bias errors, a 15-case subset of the 2003 winter experiments was also integrated with a new experimental model being developed within our projects. The experimental

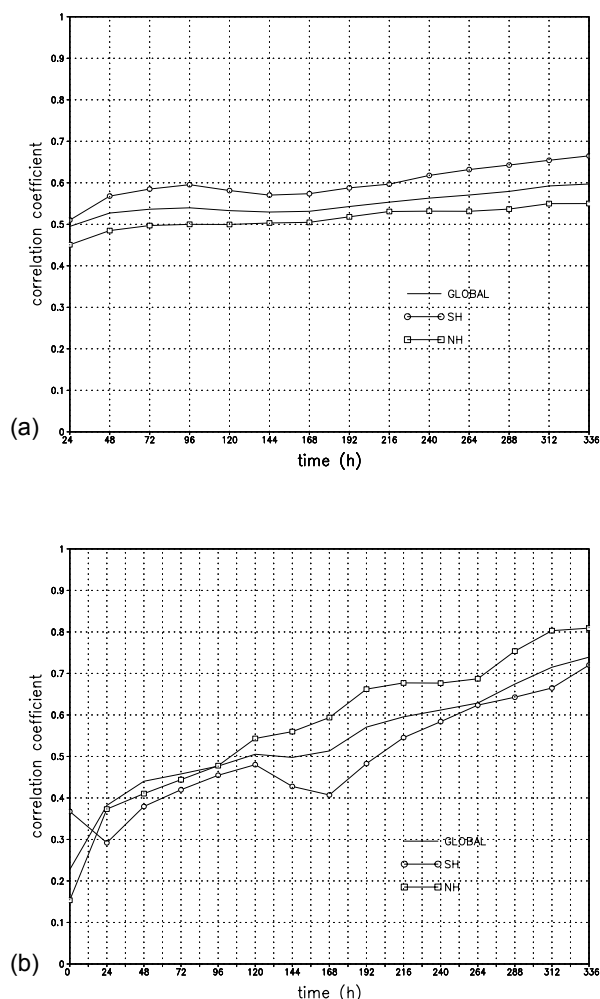


Fig. 7. (a) Time evolution of case-averaged correlation coefficients (not bias corrected) of the 500-mb meridional wind error for 30 days preceding 26 February 2003, averaged globally (solid), over the latitude band 90° to 20° S (open circles), and over the latitude band 20° to 90° N (open squares). (b) Time evolution of the correlation coefficients of the 500-mb meridional wind bias, computed over the 30 days preceding 26 February 2003, averaged globally (solid), over the latitude band 90° to 20° S (open circles), and over the latitude band 20° to 90° N (open squares).

model is a non-hydrostatic and compressible Euler model that contains a more general dynamical core than does the primitive equation UGM, while including the same physical parameterizations and vertical differencing. The Utah euler model incorporates the b-grid distribution of variables, uses fourth-order finite element methods in all advection terms, and retains 30 vertical

levels in a height-based terrain-following coordinate.

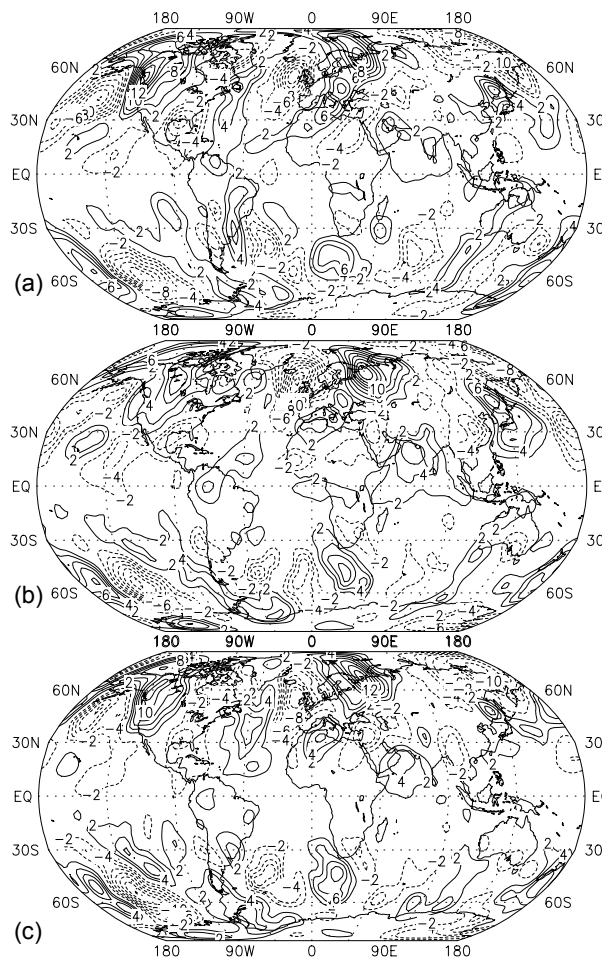


Fig. 8. 15-day average bias fields for the 500-mb meridional wind at forecast hour 336 for the (a) UGM, (b) Euler model and (c) MRF, for alternate days starting on 27 January 2003. Contour interval is 2 m s^{-1} .

Figure 8 displays the meridional wind bias of the UGM, MRF and Euler models at forecast hour 336 for a 15-case subset initialized from analyses taken every other day starting on 27 January 2003. All model forecasts exhibit similar magnitude and phase in the bias fields. This implies the sources of these errors may not be simply due to dynamical simplifications inherent to primitive equation formulations and may also be expected in nonhydrostatic and compressible treatments. We are continuing experimentation with the Euler model for more comprehensive conclusions.

Thus, given the similarity of errors in models of highly differing complexity and design, it is tempting to speculate a fundamental limitation of forecast accuracy is uncertainty of the initial state, a problem common to all

global models initialized with the same or similar data, or inherent errors in model physics or parameterizations common to the global models in question. The next section describes other measures of the forecast sensitivity to model and initial state changes.

4. GLOBAL ROOT-MEAN SQUARE FORECAST AND ANALYZED EVOLUTION

Although UGM and MRF model error fields display substantial correlation, the forecasts also exhibit differences and these are most easily summarized in terms of their global rms values:

$$((X - Y)^2)^{1/2} \quad (4)$$

where X and Y are some time-specific variable (as described below), and $(\overline{\quad})$ indicates a global average. The present section displays global rms changes of the analyzed height (X is analyzed height at time t , Y is analyzed height at initial time) as well as the global rms magnitude of the forecast change (X is forecast value at time t , Y is analyzed value at initial time), of the forecast error (X is forecast value at time t , Y is analyzed value at time t), and of sensitivity to model (X is UGM forecast at time t , Y is MRF forecast at time t) and initial condition changes (X is NCEP-initialized UGM forecast at time t , Y is ECMWF-initialized UGM forecast at time t).

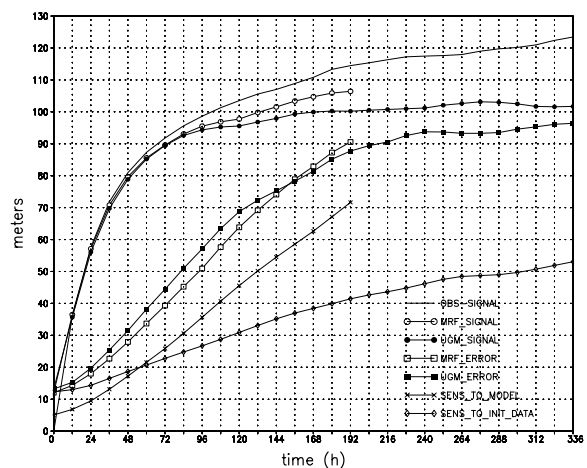


Fig. 9. Time evolution of global, case-averaged root-mean square 500-mb heights, for the 1993 cases. The top curve (solid) depicts the differences of the analyzed height at the time indicated on the abscissa minus the value at 0 hours. The second curve (open circles) depicts the differences of the MRF forecast heights at the indicated time minus the analysis at 0 hours. The third curve (closed circles) depicts the differences of the UGM (initialized with NCEP) height at the indicated time minus the analysis at 0 hours. The fourth curve (open

squares) depicts the rms error of the height field forecasted by the MRF. The fifth curve (closed squares) depicts the rms error of the height field forecasted by the UGM (initialized with NCEP). The sixth curve (x's) depicts the difference of the UGM (initialized with NCEP) forecast height minus the MRF forecast height, and measures the sensitivity of the forecasts to model used. The seventh and bottom curve (diamonds) depicts the difference of the UGM forecast height initialized with NCEP minus that of the UGM initialized with ECMWF, and measures the sensitivity of the forecasts to the initial dataset used.

The top (solid) curves of Fig. 9 depict the global rms evolution of the 1993 analyzed differences of the 500-mb height fields at the time indicated on the abscissa minus the initial value. This curve is an approximate measure of the observed change in the "signal" associated with each field (i.e., the error of a persistence forecast) and represents the average for the 17 cases. The curve on each diagram immediately below this (open circles), and stopping at 192 hours, displays the rms evolution of the height field from initial to forecast time for the 8 day bias-corrected MRF forecasts. The fact that these curves are slightly lower than the topmost curves suggests that the MRF model maintains slightly less variability than analyzed in the atmosphere.

This deficiency is even more pronounced in the low-resolution UGM, for which the bias-corrected forecast signal is displayed in the third curve from the top (closed circles) in Fig 9. The forecasted signal of the UGM is approximately 10% -20% weaker than the analyzed signal; the rms change from the analyzed initial state continues to grow past day 14, while this measure of evolution grows very little past day 7 in the UGM (even though the UGM predicted fields continue to change after day 7).

The fourth curve from the top (open squares), stopping at 192 hours, depicts the globally computed rms error of the 500-mb geopotential height for the MRF. The fifth curve from the top (closed squares) depicts the globally computed rms error of the 500-mb geopotential height for the NCEP initialized UGM. These curves asymptote toward the forecasted signal after day 8.

The bottom two curves in Fig 9 display the sensitivity of the forecast to the model used (second from the bottom, x's) and to the initial dataset used (diamonds). The curve depicting the sensitivity to model used is obtained by differencing the forecast produced by the MRF model from that produced by the low-resolution UGM and computing the globally averaged rms of the result for the 17 cases. Through the 8 day period for which this difference field can be calculated, this measure of sensitivity to model is less than the forecast error of either the UGM or the MRF predictions. This supports the conjecture that

forecasts by different models sometimes resemble each other more closely than they resemble reality.

The bottom curve of Fig 9 shows sensitivity to a data switch from NCEP to ECMWF reanalyses for initial data. The fact that this curve lies well below the two curves directly above it (forecast error and model sensitivity) may imply that initial data uncertainty is not a very important problem in the 1993 cases. This inference does not have strong support, since certain other features of the forecast evolution imply sensitivity to the initial data, as discussed in the previous section. It is also likely that the present estimate of initial uncertainty (from the difference of ECMWF and NCEP analyses of essentially the same observations) underestimates the actual initial state errors. Additionally, the rate of increase of this curve is related to model resolution, and may be a function of model complexity as well. These aspects are discussed in more detail in Section 5.

Similar results were obtained for the 52 selected cases of winter 2002-2003 (not shown here). In general, most of the discussion for the 1993 cases also applies here, although the model error curves have smaller values in 2003 than in 1993 for both the UGM and the MRF, while the sensitivity to the model has increased slightly. For winter 2002-2003, the sensitivity of the forecast to the selected model is about as great as the error of the MRF model, and only slightly smaller than the error of the UGM through the first week of prediction. This result and the substantial positive correlation between UGM and MRF predicted error fields discussed in the last section suggest that the two models produce errors that have similar pattern but differ in magnitude. This "distinct tendency for errors in each model to be highly correlated with one another," though remarkable, is not new and was noted in forecast intercomparisons over two decades ago (Baumhefner and Downey 1978).

5. INITIAL STATE ERROR EVOLUTION

MMP (2000) demonstrated that the influence of initial state uncertainty is far from saturation after five days of simulation by the Utah model. They studied forecast sensitivity to initial state uncertainty due to different complementary wave groups in the total initial uncertainty. The uncertainty was estimated from the difference of two equally credible analyses of the atmosphere, provided by NCEP (T36 truncation) and ECMWF reanalyses of the same state. These analyses, which use the same observations, are very similar over land areas of good observational coverage (Fig 10 for 500-mb height and wind analysis differences), but they differ more substantially in regions of poor observations, such as the southern oceans.

In qualitative terms, the spatial structures of analysis differences displayed in Fig. 10, represent the expected geographical distribution of initial state uncertainty. In particular, the analysis differences are small over well-observed continents and larger over poorly observed regions of the globe. For this study, it is assumed this difference field provides a reasonable estimate of the actual observation uncertainty. The influence of this level of initial state uncertainty is determined by repeating separate forecasts made by the low-resolution UGM with NCEP and with ECMWF initializations, and studying the evolving difference fields of the predictions. Following MMP (2000), sequences of experimental cases are performed in which separate spectral wave groups are modified in the initial data, and results are normalized by dividing by the variance of the predicted difference produced by initial state modifications of the entire spectrum. Variances, rather than rms values of the difference fields are computed since spectral contributions to the

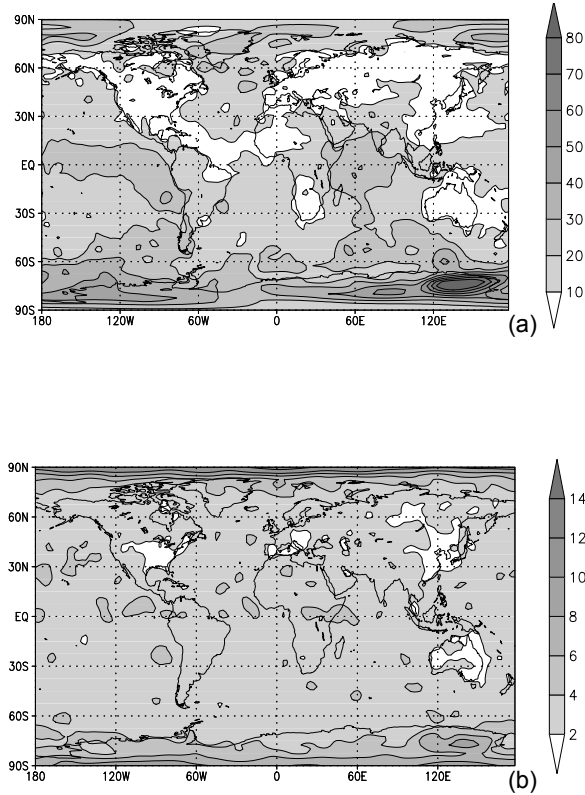


Fig. 10. Initial state 500-mb (a) geopotential height and (b) wind rms differences between NCEP and ECMWF reanalyses for the 17 selected cases of 1993. The units are m in (a) and m s^{-1} in (b).

total variance are additive. Present results emphasize response in the meridional wind since this effectively

represents the positioning and magnitude of synoptic scale waves. This quantity also better represents variability in the tropics than does the height field.

The results are displayed in Fig. 11, which depicts normalized forecast responses to uncertainties retained in different portions of the global spectrum. If the forecasts were extended over a sufficiently long period, error growth would saturate, and all the curves of Fig. 11a and Fig. 11b would asymptote to one, and the sums of the two curves would then asymptote to 2.

MMP (2000) produced the results of Fig. 11a and 11b out to five days, and noted that sums of normalized uncertainty growth for complementary wave groups approximately equal 1 rather than 2. They concluded that the first five days of prediction by the Utah model were characterized by slow, linear error growth, which is far from saturation through five days. Within this regime of unsaturated error growth, the effect of changing long-wave components of the uncertainty spectrum produces

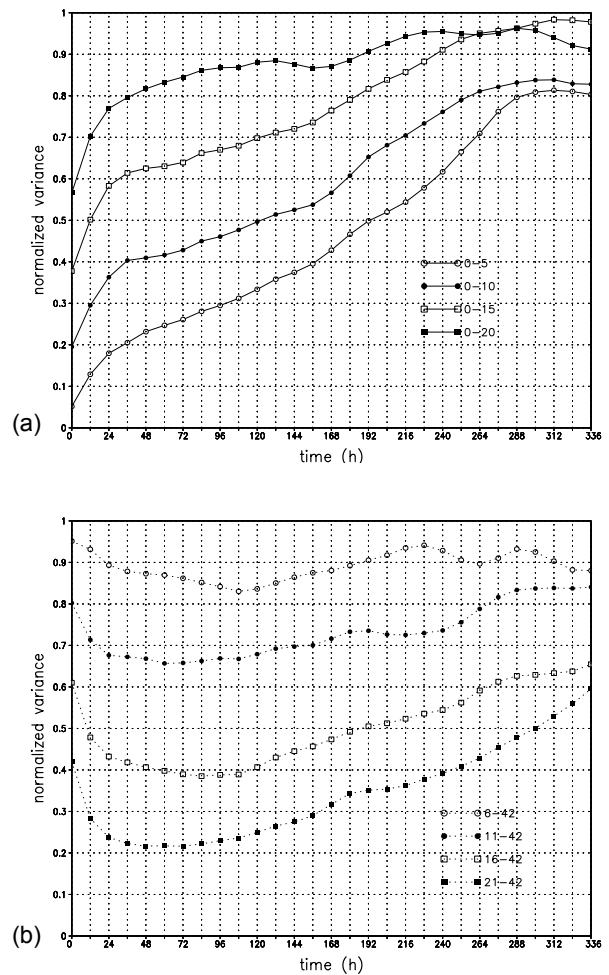


Fig. 11. Time evolution of global variances at sigma = 0.53, of forecast meridional flow differences produced by

initial data changes from NCEP to ECMWF reanalyses. Curves in (a) show the impact of changing initial data only for global wavenumbers 0-5 (open circles), 0-10 (closed circles), 0-15 (open squares), and 0-20 (closed squares). Curves in (b) display the impact of changing initial data only for the complementary wavenumbers 6-42 (open circles), 11-42 (closed circles), 16-42 (open squares), and 21-42 (closed squares). Results in both (a) and (b) have been normalized by dividing by the variance of the forecast meridional flow differences produced by initial state change of the entire spectrum.

an error contribution that grows more rapidly in a relative sense than the effect of changing shorter waves of the initial state. This can be deduced from the initial upward trends of curves in Fig. 11a and the initial downward trends in curves in Fig. 11b.

The present extension to 14 day predictions provides important modifications of these conclusions. In particular, pronounced growth of the sums of the normalized uncertainty contributions in Fig. 11 occurs after 120 hours, when the sums are on the order of 1.1 or less. By 336 hours, the sums range from 1.5 to nearly 1.7 suggesting that initial state errors have effectively spread across the full spectrum resolved in the present integrations. Even for this extended prediction the longer wave uncertainty explains at least as much, or more of the total error growth than does uncertainty in the shorter waves. Waves 0-15, for example, account for less than 40% of the initial uncertainty in the present experiments in Fig. 11a, but produce about 50% more sensitivity at 14 days than do waves 16-42 (Fig. 11b).

The sums of relative uncertainty in different spectral groups increases toward approximately 1.5-1.7. These curves should asymptote to 2 in the limit when errors associated with the chaotic nature of the atmosphere would finally saturate. Simmons and Hollingsworth (2002) suggest that the ECMWF model is close to error saturation after approximately 21 days. They base this conclusion upon the rate at which model solutions initialized on consecutive days diverge. The present experiments are integrated to only 14 days and use a different method to study error saturation, but an extrapolation of the sums of curves from Fig. 11 from the second week shown toward 21 days appears to be consistent with Simmons and Hollingsworth's (2002) conclusion.

The most important limitations to present conclusions are that they are made using a forecast model truncated at relatively low resolution, and that the differences of NCEP and ECMWF analyses, which use the same observations, may not adequately reflect the actual uncertainty of the initial state. Higher resolution models generally allow larger uncertainty growth because model diffusion coefficients need to increase with coarser reso-

lution to control spectral blocking and to limit accumulation of energy at the shortest resolvable scales due to nonlinear energy cascade.

It is likely that the present method of initial state uncertainty specification underestimates the actual uncertainty in both large and small scales. In particular, neither analysis contains much amplitude in higher wavenumber, smaller scale components (wavelengths on the order of 1000 km) and their differences may systematically underestimate the impact of errors at these scales.

White et al (1999) estimated errors of the MRF initial state over the intermountain west of the U.S. in a region of good radiosonde density. Table 2 of that study suggests initial rms error values for the wind exceed 4 m s^{-1} in the mid troposphere in a region of reliable observations. These uncertainties are estimated from the fit of initial analyses interpolated to radiosonde observation sites using a horizontal grid corresponding to approximately global wavenumber 100. Presumably, the uncertainties are even larger in regions of sparse observations. The initial data uncertainties (Fig. 10) used in present experiments are smaller than 4 m s^{-1} over most of the globe, and impose smaller initial state uncertainties than found by White et al. (1999) in a data-rich region.

In order to address these issues, three other sets of experiments were conducted. In one, the initial state uncertainty estimated from the ECMWF - NCEP forecast differences was retained only in wave groups 16-42, as in the earlier experiment sequence, but the amplitude of this difference field was arbitrarily doubled. A similar approach was taken by doubling the initial state uncertainty only in waves 0-15. Both experimental sequences were then repeated at low (wavenumber 42) resolution. A third experiment set was created by doubling the horizontal resolution in both longitude and latitude, and repeating the 17 forecasts for both ECMWF and NCEP initializations.

The results are summarized in Fig. 12, which is in the same format as Fig. 9, and displays the relative contributions of various model/data configurations to the forecast discussed in Section 4. Each curve of Fig. 9 is repeated here, and three new curves are added: the sensitivity to doubled initial uncertainty in waves 0-15 (open triangles), the sensitivity to doubled initial uncertainty in waves 16-42 (closed triangles), and the sensitivity to initial state for doubled horizontal resolution (plus signs).

The most important change comes from doubling our prior estimate of uncertainty in waves 0-15. Doubling uncertainty in waves 16-42 is also important, but does not produce as much influence at longer forecast times as found in modifications of the longer wave group. Dou-

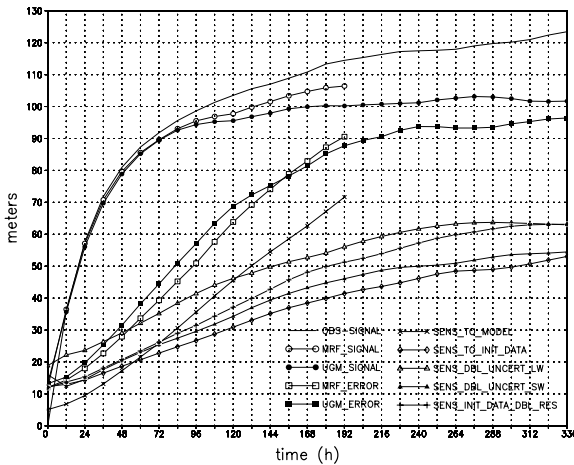


Fig. 12. Time evolution of global, case-averaged root-mean square 500-mb heights, for the 1993 cases. The top curve (solid) depicts the differences of the analyzed height at the time indicated on the abscissa minus the value at 0 hours. The second curve (open circles) depicts the differences of the MRF forecast height at the indicated time minus the analysis at 0 hours. The third curve (closed circles) depicts the differences of the UGM (initialized with NCEP) height at the indicated time minus the analysis at 0 hours. The fourth curve (open squares) depicts the rms error of the height field forecasted by the MRF. The fifth curve (closed squares) depicts the rms error of the height field forecasted by the UGM (initialized with NCEP). The sixth curve (x's) depicts the difference of the UGM (initialized with NCEP) forecast height minus the MRF forecast height, and measures the sensitivity of the forecasts to model used. The seventh curve (open triangles) depicts the NCEP-initialized UGM forecast difference due to doubled initial uncertainty in wavenumbers 0-15. The eighth curve (plus signs) depicts the difference of the high-resolution UGM forecast initialized with NCEP minus that of the high-resolution UGM initialized with ECMWF, and measures the sensitivity of the high-resolution forecasts to the initial dataset used. The ninth curve (closed triangles) depicts the NCEP-initialized UGM forecast difference due to doubled initial uncertainty in wavenumbers 16-42. The tenth and bottom curve (diamonds) depicts the difference of the low-resolution UGM forecast initialized with NCEP minus that of the low-resolution UGM initialized with ECMWF, and measures the sensitivity of the low-resolution forecasts to the initial dataset used.

bling model resolution also enhances error growth rates.

None of these experiments produces as much 8 day forecast sensitivity to the initial state as the model selection (x's). As noted previously, the relative magnitude and structure of errors of the MRF and Utah model appear to be similar through 8 days, but

their forecast differences are larger than sensitivity to presently estimated measures of initial state uncertainty. This conforms with Reynold's et al.'s (1994) conclusion for the tropics, but conflicts with their findings for the extra-tropics. Sensitivity to initial state uncertainty may also be a function of model complexity and to be fair in the relative weighting of initial condition differences versus model differences, the initial condition sensitivity experiments should be done with a more complex, higher resolution model.

6. SUMMARY AND CONCLUSIONS

This study has investigated predictability using three different global forecast models, and three different estimates of the initial state selected from reanalyses for 1993 and operational data for winter 2002-2003. Our approach has been to compare the different model forecasts and errors, and the sensitivity of forecasts to model and to initial state. One of the models (the experimental UGM) was applied with all three analyses (ECMWF and NCEP reanalyses for 1993 and NCEP operational analyses for 2003). The other models are different versions of formerly and currently operational MRF models initialized with NCEP analyses.

The comparison of results of experimental models with operational models allows more predictability inferences than would the use of a single operational model. The strategy has been to maintain substantially lower resolution and simplified physical parameterization in the UGM than is available in the MRF. This allows assessment of the roles of resolution and of model sophistication relative to improvements of the initial state specification. The relative contributions of these separate improvements cannot be determined from a single forecast system in which model initialization as well as sophistication are changing simultaneously. The first guess is such an important component of the data assimilation cycle that model advances will lead to initial state advances. Consequently, the value of individual contributions are not entirely clear even in the present approach, but the strategy does help sharpen the questions and diagnostic methods.

Earlier low-resolution experiments by MMP were expanded from 5 days to 14 days. Additionally, some experiments were performed with doubled resolution, and modified estimates of initial state uncertainty. The relative magnitude of forecast uncertainty associated with model changes was compared to the role of initial state uncertainty estimates. The main results are summarized below.

The results of the 1993 reanalysis cases suggest that model bias extraction produces more benefit to 14 days

than other tested modifications of model heating. Bias corrected anomaly correlations produced by the UGM lag MRF bias corrected anomaly correlations by about 12 hours to day 6, and become very similar to the MRF bias corrected anomaly correlations by day 7. Bias corrected anomaly correlations fall below 0.6 after 6.5 days for the UGM initialized with NCEP reanalyses and fall below 0.6 after 7 days for the UGM initialized with ECMWF reanalyses. These results suggest that bias-corrected forecasts of a relatively simple research model such as the UGM have almost as much value in predicting large-scale features as those obtained from a more highly developed model such as the MRF that was used operationally until approximately 1995.

The results for the 2003 real-time sample suggest that the operational MRF has gained approximately 36 hours of forecast value since the earlier experiments, and exhibits bias-corrected anomaly correlations of approximately 0.6 after 8 days for the presently studied sample. Meanwhile the UGM executed with the more recent data has increased the period of useful predictability by only approximately 6 hours relative to the 1993 data. It is possible that the extra 30 hours gain made by the MRF is due to the model advances at NCEP rather than improvements of the initial state. Although this possibility cannot be discounted, the error patterns of both the low-resolution UGM and high-resolution MRF show strong positive correlation through the prediction, suggesting that much of the remaining error may be explained by some common deficiency of the models.

The MRF contains certain inherent advantages in the balance of initial fields, since the same model is incorporated to provide the initial guess fields for the analyses. The UGM linearly interpolates the reanalyses values to UGM grid points, and the short-term evolution includes a spin-up time to adjust to small local imbalances. In view of these simplifications it is somewhat surprising the length of skillful forecasts does not differ by more than roughly 12h for the 1993 sample. This result is consistent with the possibility that those predictions were not strongly limited by differences in model sophistication.

To better understand sensitivity to model heating, all UGM experiments presently shown were repeated by substituting climatological values obtained from reanalysis archives for model-generated values of radiative heating of the atmosphere. All statistics matching those displayed in this paper, as well as individual forecasts were remarkably similar to the UGM forecasts in which internally consistent physical parameterizations were computed rather than pre-specified, and are not shown.

Our experiments indirectly suggest that a major benefit enjoyed by the current version of the MRF is the extra

resolution of that model relative to the other models tested in the present experiments. As previously discussed, higher resolution also produces more sensitivity to initial state, possibly allowing more complete accounting for the many initial state improvements that have been incorporated since 1993.

These speculations should be checked by repeating experiments displayed here using a much higher resolution model for experimental forecasts and systematically modifying its input and model parameterizations to quantify the role of model resolution and sophistication on predictive skill. Comparison of forecast and error fields from a third, independent model with those from the UGM and MRF would also allow further assessment of the role of improvements in initial state specification. Finally, repeating the spectrally-binned experiments with a much higher resolution model is necessary to verify results regarding initial state error evolution in different wave groups.

Acknowledgements. This research was supported by NSF Grants ATM0109241 and ATM0106776 to the University of Utah. The 1993 reanalysis data were obtained from the National Centers for Atmospheric Research with assistance from Dr. J. N. Paegle. The 2003 analyses and MRF forecasts were obtained from the National Centers for Environmental Prediction with assistance from Dr. W. James Steenburgh. Invaluable computer assistance was provided by Bryan White.

- Bath, L. M., M. A. Dias, D. L. Williamson, G. S. Williamson, and R. J. Wolski, 1987: User's Guide to NCAR CCM1, Technical Report NCAR/TN-286+IA, National Center for Atmospheric Research, 173 pp.
- Baumhefner, D., and P. Downey, 1978: Forecast inter comparisons from three numerical weather prediction models. *Mon. Wea. Rev.*, **106**, 1245-1279.
- Buchmann, J., L. E. Buja, J. Nogues-Paegle, and J. Paegle, 1995: The dynamical basis of regional vertical motion fields surrounding localized tropical heating. *J. Climate*, **8**, 1217-1234.
- Byerle, L. A., and J. Paegle, 2003: Description of the seasonal cycle of low-level flows flanking the Andes and their interannual variability. *Meteorologica*, **27**, 71-88.
- Gibson, J. K., P. Kallberg, S. Uppala, A. Nomura, A. Hernandez, and E. Serrano, 1997: ERA description. ECMWF Re-Analysis Project Report Series, Vol. 1, 58 pp.
- Hacker, J. P., E. S. Krayenhoff, and R. B. Stull, 2003: Ensemble experiments on numerical weather prediction error and uncertainty for a North Pacific forecast failure. *Wea. Forecasting*, **18**, 12-31.
- Hartmann, D. L., R. Buizza, and T. N. Palmer, 1995: Sin

- gular vectors: The effect of spatial scale on linear growth of disturbances. *J. Atmos. Sci.*, **52**, 3885-3894.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **53**, 252-264.
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289-307.
- Miguez-Macho, G., and J. Paegle, 1999a: Optimal observation distribution for numerical weather prediction. Preprints, *Third Symp. on Integrated Observing Systems*, Dallas, TX, Amer. Meteor. Soc., 18-23.
- , and -----, 1999b: Optimal observation distribution for numerical weather prediction. Preprints, *13th Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 23-26.
- , and -----, 2000: Sensitivity of a global forecast model to initializations with reanalysis datasets. *Mon. Wea. Rev.*, **128**, 3879-3889.
- , and -----, 2001: Sensitivity of North American numerical weather prediction to initial state uncertainty in selected upstream subdomains. *Mon. Wea. Rev.*, **129**, 2005-2022.
- Nicolini, M., K. M. Waldron, and J. Paegle, 1993: Diurnal variations of low-level jets, vertical motion and precipitation: A model case study. *Mon. Wea. Rev.*, **121**, 2588-2610.
- Nogues-Paegle, J., K. C. Mo, and J. Paegle, 1998: Predictability of the NCEP-NCAR reanalysis model during austral summer. *Mon. Wea. Rev.*, **126**, 3135-3152.
- Paegle, J., 1989: A variable resolution global model based upon Fourier and finite element representation. *Mon. Wea. Rev.*, **117**, 583-606.
- , and J. Horel, 1991: The influence of observational uncertainty upon wind-based analyses. Preprints, *Ninth Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 779-782.
- , Q. Yang, and M. Wang, 1997: Predictability in limited area and global models. *Meteor. Atmos. Phys.*, **63**, 53-69.
- Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633-653.
- Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **122**, 121-150.
- Reynolds, C. A., P. J. Webster, and E. Kalnay, 1994: Random error growth in NMC's global forecasts. *Mon. Wea. Rev.*, **122**, 1281-1305.
- Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **128**, 647-677.
- Thompson, P. D., 1957: Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, **9**, 275-295.
- Vukicevic, T., and J. Paegle, 1989: The influence of one-way interacting lateral boundary conditions on predictability of flow in bounded numerical models. *Mon. Wea. Rev.*, **117**, 340-350.
- Warner, T. T., R. A. Peterson, and R. E. Treadon, 1997: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599-2617.
- White, B. G., J. Paegle, W. J. Steenburgh, J. D. Horel, R. T. Swanson, L. K. Cook, D. J. Onton, and J. G. Miles, 1999: Short-term forecast validation of six models. *Mon. Wea. Forecasting*, **14**, 84-108.