An Analysis of Forecast Error in the NMC Hemispheric Primitive Equation Model

JOHN M. WALLACE AND JAMES K. WOESSNER

Department of Atmospheric Sciences, University of Washington, Seattle 98195

(Manuscript received 19 May 1981, in final form 30 September 1981)

ABSTRACT

Errors in 72 h forecasts of 500 and 1000 mb height and 1000–500 mb thickness from the NMC hemispheric primitive equation model during six recent winter seasons are examined in order to determine the geographical distribution of the systematic and non-systematic components of the forecast error. The systematic errors are in the sense as to make the equator-to-pole temperature gradient, the time-averaged jet streams, and the standing waves weaker in the forecasts than in the observed climatology. The non-systematic forecast errors are closely related to the observed high-frequency temporal variability, with largest values in the storm tracks. The systematic error accounts for up to about one-third of the mean-square forecast error in middle latitudes.

1. Introduction

In June 1966, a six-layer, hemispheric, grid-point primitive equation (PE) model went into operational use at the U.S. National Meteorological Center to provide guidance in the preparation of weather forecasts. The specifics of an early version of the PE model are described in Shuman and Hovermale (1968) and subsequent changes in the model formulation are documented in a series of Technical Procedures Bulletins issued by the Weather Analysis and Prediction Division of NMC. A seventh layer was added in 1977 and the grid size was halved in 1978. The model remained operational until August 1980, when it was replaced by a nine-layer spectral model. The extensive archive of forecast products derived from the PE model constitutes a valuable resource for developing procedures for model diagnostics.

In discussing the performance of numerical weather prediction models the term systematic error has been used in two rather different contexts. Authors such as Leary (1971), Colucci and Bosart (1979) and Silberberg and Bosart (1981) have used the term to describe the tendency of a forecast model to show a bias in the treatment of specific kinds of synoptic entities (e.g., the failure to predict sufficient deepening of cyclones, or the tendency to move systems too slow or too fast), whereas other authors such as Fawcett (1969) and Hollingsworth et al. (1980) have used it to describe differences between the mean fields generated by a forecast model (run out to a given forecast interval from many sets of initial conditions) and their observational counterparts. The systematic errors, as defined by the two different kinds of investigations may, in some cases, be related: for example, the tendency for a forecast model to predict excessive deepening of cyclones in a given geographical region could contribute to a negative bias in the mean sea level pressure field in the 72 forecasts for that region. On the other hand, as Hollingsworth et al. point out, errors in the mean “climate” generated by a forecast model might also be a result of improper treatment of the linear forcing of the stationary waves. Speaking in more general terms, it seems reasonable to expect that any errors in the parameterization of physical processes in a forecast model should produce systematic errors of both types. In the present study we will be concerned with systematic errors in the NMC PE model, as used in the second context.

Fawcett (1969) documented the performance of early versions of the PE model, based on an analysis of mean forecast errors for a selection of months during the interval 1966–68. He concluded that the model's forecasts exhibited systematic regional biases: positive in the regions of the climatological mean troughs and negative in the mean ridges, so that the amplitude of the climatological mean standing waves was systematically underestimated. He also noted that the model tended to damp out the thickness gradient between low and high latitudes, thus reducing the strength of the zonally averaged thermal wind and tropospheric jet stream.

---

1 Contribution No. 596, Department of Atmospheric Sciences, University of Washington.

2 Present affiliation: Staff Meteorologist, Air Force Engineering and Services Center, Tyndall Air Force Base, Panama City, FL 32403.

0027-0644/81/122444-06$05.50
© 1982 American Meteorological Society
2. Analysis procedures

We will consider the geographical distribution of what we will refer to as the systematic component of the forecast error, as defined by

\[ \bar{x}_E = \bar{x}_E - \bar{\bar{x}}_E, \]

where \( x \) is the meteorological parameter under consideration (e.g., 500 mb height). The subscript \( F \) refers to the forecast value of \( x \), \( O \) to the observed value at verification time and \( E \) to the forecast error. The overbar denotes an average over a large number of forecasts made with the same model but with different initial conditions. We also will consider the non-systematic component of the forecast error, as defined by

\[ x'_E = x_E - \bar{x}_E. \]

It is convenient to display this non-systematic component in terms of its root mean square (rms) value

\[ (x'_E)^2)_{1/2} = [(x_E - \bar{x}_E)^2]_{1/2}. \]

Note that the mean-square error \( \bar{x}_E^2 \), a conventional measure of forecast performance, is equal to the sum of the squares of the systematic and non-systematic components, as defined above, i.e.,

\[ \bar{x}_E^2 = \bar{x}_E^2 + \bar{x}'_E^2. \]

We will use the above formalism to document the errors in the NMC hemispheric PE model during the later part of its history—the winter seasons beginning with 1973–74 and ending with 1979–80. Our analysis included 24, 48 and 72 h forecasts. In general, we observed an approximately linear growth of the systematic errors with time, with the geographical patterns being very much the same for all three forecast intervals, apart from differences in amplitude. Therefore, we will show in this paper only the results for the 72 h forecasts. All calculations were made only for the winter months (December, January, February). All data used in this study were obtained from the NCAR data library. About 550 individual 72 h forecasts were available, most of these verifying at 0000 GMT.

The reproducibility of the results described in the following sections was tested by breaking the data into subsets for individual winters and for 0000 and 1200 GMT verification times, and making separate computations for the various subsets. We were particularly concerned about possible differences in the pattern of systematic error between earlier and later parts of the data record, due to the evolution of the forecast model during this period or changes in the objective analysis procedures. This preliminary analysis indicated that the geographical patterns of the systematic component of the error are highly reproducible in individual winters, and for the 0000 and 1200 GMT verification times separately, at least insofar as the large-scale features are concerned. Nor did we find any evidence of major shifts in the pattern between the earlier and later parts of the 6-year record. For the sake of brevity we will show in subsequent sections only the results based on the entire six-winter data set.

3. Systematic errors

The mean 72 h forecast error map for the 500 mb height field is shown in Fig. 1a. The systematic component of the error exhibits a well-defined pattern which is characterized by a tendency for negative bias (heights forecast lower than observed) in low latitudes and positive bias (height forecast higher than observed) in high latitudes. There are also strong longitudinal gradients, with positive bias in the standing wave troughs over the east coasts of the continents and negative bias in the standing wave ridge over western Europe. The pattern of the systematic component of the error is such that the forecast 500 mb flow tends to weaken with time nearly everywhere in the hemisphere. On the basis of calculations of the geostrophic wind at representative locations we estimate that 500 mb wind speeds near the jet streams are on the order of 5–10% too low in the 72 h forecasts.

The 1000 mb height pattern, shown in Fig. 1b, is dominated by positive biases of the PE model 72 h forecasts over the Icelandic and Aleutian lows and negative biases over the continents. The largest negative biases over North America are found to the east of the Rockies, in agreement with Fawcett's findings. As at 500 mb, the general sense of the forecast errors is such that the climatological mean flow pattern becomes increasingly damped as the forecast interval lengthens. Underestimates of geostrophic wind speeds, in terms of percent, are larger than those at 500 mb.

The 1000/500 mb thickness pattern, shown in Fig. 1c, is dominated by negative biases over the oceans and at low latitudes, and larger positive biases over the high-latitude continents. Again, the pattern resembles the inverse of the climatological mean pattern, but the land–sea contrasts are more prominent, particularly over the Atlantic sector. In interpreting the magnitude of these errors, it is helpful to note that each 10 m contour interval is equivalent to an increment of 0.5 K in the bias in the mean temperature of the 1000/500 mb layer; hence, for example, over the northern continents the model produces a fictitious warming of more than 3 K over a 3-day period.

4. Non-systematic errors

The temporal variance of the 72 h forecast 500 mb height about its own climatological mean was
Fig. 1. Distribution of systematic error in 72 h forecasts produced by the NMC PE model, based on ~550 forecasts made during the winter months (December–February) of the winters 1973–74 through 1979–80. (a) 500 mb height, (b) 1000 mb height, and (c) 1000/500 mb thickness. Contour interval 10 m.

compared with the corresponding temporal variance of the observed 500 mb height at each grid point (not shown). Poleward of 30° latitude the ratio of the variances is close to unity at most grid points. The forecasts exhibit slightly (up to 25%) less variance than climatology over North America and up to 50% more variance than climatology over the Eurasian sector. Equatorward of 30°N the forecasts exhibit more temporal variability than the NMC analysis. The interpretation of this result is not clear, since the NMC analyses may well be deficient near the periphery of the grid.

The 72 h forecast fields produced by the NMC model tend to be overly smooth; that is, deficient in disturbances with relatively small space scales. The above findings are not necessarily in conflict with this
general observation because, even though small-scale disturbances are of considerable importance in day-to-day weather forecasting, they account for only a very small fraction of the temporal variance of the 500 mb height field.

Fig. 2 shows the non-systematic component of the 72 h forecast errors in 500 mb height, shown in rms format. The pattern strongly resembles the distribution of the climatological mean temporal variance of 500 mb height (e.g., see Fig. 3a of Blackmon (1976)). Hence the non-systematic forecast errors tend to be largest in the regions of large transient variability, which are centered over the northern oceans. [As shown in Blackmon et al. (1976) and in Lau (1979), these regions of large variability are associated with the midlatitude storm tracks poleward and downstream of the climatological mean jet streams.] In these active regions of the hemisphere the 72 h forecast errors amount to slightly more than half the rms climatological variability. Very similar results were obtained for the forecasts of 1000 mb height (not shown).

Fig. 3 shows the distribution of the non-systematic component of the error for 500 mb height, normalized by dividing the values at each grid point in Fig. 2 by the corresponding climatological mean standard deviation, computed from the same ~550 analyses against which the forecasts were verified. Regions

where the rms non-systematic forecast error exceeds the climatological mean variance are shaded. These may be viewed as regions in which the forecast model has no skill relative to climatology when run out to 72 h. Poleward of about 30°N (40°N over Asia), the normalized error for 500 mb height is rather uniform and mostly below 0.8; the smallest normalized errors tend to coincide with the storm tracks. The normalized 1000 mb height errors (not shown) exhibit a rather similar pattern, but they tend to be somewhat larger, particularly over the southwestern United States, Korea and Japan, where they are greater than 1. For both 500 and 1000 mb field heights, the normalized errors become progressively larger as one approaches the periphery of the grid.

5. Discussion

The relative magnitudes of the systematic and non-systematic components of the forecast error in the 500 mb height field can be estimated by comparing values in Figs. 1a and 2 in various geographical regions. For example, over Newfoundland and just east of Japan, where the systematic component of the error reaches maximum values of ~70 m, the non-systematic component is on the order of 100–120 m; hence, the systematic component accounts for up to
about one-third of the mean-squared forecast error in localized regions of middle latitudes. Over the Arctic and over portions of the subtropics, the systematic and non-systematic components are of comparable magnitude.

The pattern of the systematic component of the error is remarkably similar to that obtained by Fawcett (1969) for an early version of the model. It is characterized by a progressive weakening of the meridional temperature gradient and the land–sea thermal contrast with increasing forecast interval, and a weakening of the stationary wave pattern, both in the mid-troposphere and at the earth’s surface.

Hollingsworth et al. (1980) recently reported on the systematic component of the error in the European Center for Medium Range Forecasting (ECMWF) model for wintertime conditions. In agreement with the above results for the NMC model, the systematic errors exhibited an approximately linear growth with time, and their geographical distribution was such as to contribute to a progressive weakening of the stationary wave pattern for forecast intervals up to 10 days. However, the model did not exhibit a tendency for warming over the high-latitude continents and a concomitant weakening of the zonally averaged jetstream, as the NMC model did; in fact, zonal kinetic energy and zonal available potential energy tended to increase throughout the 10-day forecast interval. It should perhaps be stressed that the results of Hollingsworth et al. are somewhat preliminary, since they are based on an ensemble of only seven forecasts. Nevertheless, the similarity of the models, with respect to their inability to maintain the stationary wave pattern, is interesting, and perhaps suggestive of common deficiencies in the model formulations. On the other hand, the NMC model appears to behave differently from the ECMWF model with respect to the maintenance of the zonally averaged jet stream and the associated meridional temperature gradient.

Hollingsworth et al. do not show the distribution of the systematic component of the error in the temperature field per se, but comparison of their distributions for 1000 and 500 mb height gives the impression that, to a first approximation, the systematic errors in the ECMWF model have an equivalent barotropic structure such that the systematic errors in the 1000/500 mb thickness field have a distribution somewhat similar to those in 500 mb heights and are perhaps half as large. It is evident from an inspection of Fig. 1 that the systematic component of the error in the NMC model displays a more complicated and more baroclinic vertical structure with errors in thickness being about as large as those in 500 mb height. The strong bias toward warming over the high-latitude continents in the NMC model is suggestive of deficiencies in the treatment of infrared radiative transfer.

6. Concluding remarks

The statistical approach used in this study and in the related studies proposed below complements the "case study" approach which has been much more commonly used for the diagnosis of forecast error (see, e.g., Baumhefner and Downey, 1978). Although they involve much larger data volumes than those needed for case studies, statistical analyses of forecast errors are rather simple and inexpensive to carry out, provided that the forecasts are archived in a convenient format. Such studies could provide a rational basis for developing statistical regression schemes for correcting model output (Leith, 1978), and they might also help to motivate improvements in the representation of physical or dynamical processes in numerical prediction models. It is important that such statistical analyses be based on time series of sufficient length to establish statistical significance, and that the results be presented in the form of maps and/or cross sections (as opposed to area or volume integrals) so that the relationships to climatological mean fields can be diagnosed.

The breakdown of the error into systematic and non-systematic components appears to be a useful formalism for diagnostic purposes.

It has come to the authors’ attention that an extensive comparison of 72 h forecasts from eight different numerical prediction models, including the NMC PE model, has been carried out by the Finnish Meteorological Institute under the auspices of the World Meteorological Organization Committee on Atmospheric Sciences (WMO/CAS) and results are being prepared for publication as a WMO/CAS report, under the authorship of L. Bengtsson and A. Lange. The format for the presentation of results is similar to that adopted here, with separate maps for systematic and non-systematic components of the errors. Results are presented for four different seasons and two different years. The results for the NMC model are in close agreement with those presented here.

Acknowledgments. We wish to thank Richard J. Reed, staff members at the European Centre for Medium Range Weather Forecasts, Reading, England, and the reviewers for a number of helpful comments. The paper was completed while the senior author was on leave to the European Centre. Support for the computing was provided by the University of Washington and by the Climate Dynamics Program Office of the National Science Foundation under Grant 78-07369.
REFERENCES