Reprinted from MONTHLY WEATHER REVIEW, Vol. 120, No. 8, August 1992
American Meteorological Society

Medium-Range Forecast Skill Variation and Blocking Transition: A Case Study

M. KIMOTO

Numerical Prediction Division, Japan Meteorological Agency, Tokyo, Japan

H. MUKOUGAWA

Meteorological College, Japan Meteorological Agency, Tokyo, Japan

S. YODEN

Geophysical Institute, Kyoto University, Kyoto, Japan

(Manuscript received 7 March 1991, in final form 29 October 1991)

ABSTRACT

Temporal variability in skill of operational medium-range forecasts in the winter of 1988/89 is examined. The primary concern is to examine the performance of the Japan Meteorological Agency (JMA) global spectral model, which began daily 8-day forecasts in 1988. The model's forecast skill exhibits considerable low-frequency (~a week or longer) temporal variability. During the period under study, root-mean-square error (rmse) of 500-mb height showed a pronounced temporal maximum at the end of January 1989, when the atmosphere over the North Pacific underwent a remarkable transition from zonal to blocked circulation.

The consecutive forecasts during this period showed large dispersion among one another. A linear measure of forecast spread, or phase-space divergence of atmospheric trajectories, is also evaluated after Lorenz using a nondivergent barotropic model. It shows a temporal maximum concurrent with rmse. Therefore, it is likely that the zonal-to-blocking transition was associated with higher-than-average instability of the atmosphere. This impression is strengthened by the fact that the models at two other centers, the European Centre for Medium-Range Weather Forecasts (ECMWF) and the U.S. National Meteorological Center (NMC), showed similar features.

The forecast dispersion prior to the blocking is studied in detail. A set of forecasts initialized every 6 h is used to complement the operational dataset. The principal spatial pattern of the spread showed sensitivity over the blocking high. In addition, it is noted that all three models examined show considerable "reluctance" to enter the blocked regime unless they are initialized 5 days or less prior to the blocking. Once the block is established in the model, however, it tends to persist.

1. Introduction

By now it is well-known that numerical weather prediction (NWP) models show large temporal variation in skill in medium-to-extended range (e.g., Branstator 1986; Curtis et al. 1988; Tracton et al. 1989). Consequently, it is increasingly becoming an important subject to predict the reliability of individual or ensemble forecasts in these forecast ranges (Kalnay and Dalcher 1987; Tennekes et al. 1987). Although there already exist several mainly statistical studies on the feasibility of the skill prediction (Kalnay and Dalcher 1987; Dalcher et al. 1988; Palmer and Tibaldi 1988; Chen 1989), more studies are necessary in order to understand the nature of the problem, especially from theoretical and synoptic points of view.

Corresponding author address: Dr. Masahide Kimoto, Meteorological Research Institute, 1-1 Nagamine, Tsukuba, Ibaraki 305, Japan.

Under a theoretical, perfect-model assumption, the forecast error growing from a small uncertainty in the initial condition is due to intrinsic instability of atmospheric flows. This instability is the divergence of phase-space trajectories passing the neighborhood of a "true" initial state. As long as a small error is assumed, temporal evolutions of a cloud of points around the reference trajectory can be followed by a tangential linear equation of the system. If the initial error distribution is assumed to be hyperspherical, the subsequent evolution deforms it to a hyperellipsoid, with the longest axis pointing to the most likely error pattern (without polarity information since the computation is linear; Lorenz 1965). Even in this idealized context, the temporal variation in error or predictability as a function of circulation seems extremely difficult to study since no general rule is available for the relation between characteristics of the initial state and the evolving hyperellipsoid (Mukougawa et al. 1991). Furthermore, the theoretical computation of tangential

linear growth involves a matrix of size $N \times N$, where N is the degrees of freedom of the system. Therefore, even the linear estimates may be feasible only with models of moderate size.

In the context of operational medium-range forecast (MRF), the small-error assumption may no longer hold. Moreover, the initial distribution of error is unknown. The Monte Carlo approach (Leith 1974; Hoffman and Kalnay 1983) seems practical but is still computationally expensive if we are to study temporal variation in predictability using state-of-the-art numerical weather prediction (NWP) models. The fact that these models are far from perfect adds considerable difficulty to the studies of operational forecast skills. It is likely that theoretical studies (Farrell 1990; Mukougawa et al. 1991), statistical analyses of the largest existing dataset (e.g., ECMWF archive; Palmer and Tibaldi 1988; Tibaldi and Molteni 1990), mechanistic model studies (e.g., Legras and Ghil 1985; Roads 1987, 1988), and intensive case studies (Tibaldi and Ji 1983; Tracton 1990) need to complement each other to ultimately increase our understanding of this subject.

So far, several attempts have been reported investigating possible predictors for the skill variation. These include (i) forecast agreement (or spread; Kalnay and Dalcher 1987; Dalcher et al. 1988; Palmer and Tibaldi 1988), (ii) persistence of forecast flow (Palmer and Tibaldi 1988; Chen 1989), and (iii) large-scale circulation patterns (Palmer 1988; Palmer and Tibaldi 1988). The motivation for the use of (i) is the hope that it serves as a proxy for the phase-space divergence. In the cases of (ii) and (iii), some relations with the stability of the flow appear to be expected. Depending on the period and geographical regions under study, some of these predictors showed modest correlation with observed NWP model skills in terms of a root-mean-square error (rmse) or an anomaly correlation coefficient. However, in the above studies, the reasons for success or failure have not been fully addressed although Palmer (1988) and Palmer and Tibaldi (1988) speculated that the positive correlation between the Pacific–North American (PNA) teleconnection index (Wallace and Gutzler 1981) and the model skill over the North Pacific may be related to the dependence of barotropic instability on the signed amplitude of the index.

The purpose of this study is to describe skill variability in the first winter of operational MRF run every day up to eight forecast days at the Japan Meteorological Agency (JMA). Forecasters realized immediately after the start of the everyday operation that the day-to-day performance of the model fluctuates considerably so that a conventional rule, that of the latest forecast being the best, is frequently violated in the medium range. This motivated us to look into the nature of skill variability. Since only one winter is studied, we focus more on inquiries into the possible nature of the skill variation rather than trying to establish useful measures of skill prediction. In order to avoid being misled by features specific to the JMA model, we complement the analysis by also examining forecasts produced at the European Centre for Medium-Range Weather Forecasts (ECMWF) and at the U.S. National Meteorological Center (NMC) in the same period. Since Palmer (1988) and Frederiksen (1989) have suggested that a barotropic model may be useful in investigating flow instability on some occasions, we use it to aid the description of the full model results. In this study, we refer to the divergence of trajectories as *instability* (cf., Lorenz 1965; also section 4b below).

Section 2 describes the data used in this study. Section 3 documents the low-frequency skill variation in the winter of 1988/89. A comparison of the forecasts produced at the different centers immediately reveals that a considerable portion of the low-frequency variability is independent of the forecast–analysis system used. In particular, a prominent temporal maximum at the end of January 1989 is shown to have been associated with a transition in the North Pacific from zonal to blocked circulation regimes. Section 4 shows that forecast spread and a linear measure of atmospheric instability attained temporal maxima concurrent with that in rmse. In section 5, we study the nature of the forecast dispersion prior to the onset of blocking in more detail. A series of forecasts by the JMA T63 model initialized every 6 h and by a T42 version of the same model initialized every 12 h during the period are used to complement the operational dataset. Section 6 discusses the results, followed by a summary and concluding remarks in section 7.

2. Data

A 16-level global spectral model with triangular 63 truncation (T63) was introduced in March 1988 at JMA and is described by Sugi et al. (1990). Operational 8-day forecasts were produced every 1200 UTC starting 1 March 1988. The forecast data were stored every 24 h. A T42 wave archive of 500-mb geopotential heights and 300-mb winds were used in this study, and 90 verification dates starting 1 December 1988 are investigated. Data slightly before and after this "winter" period are also utilized when temporal filtering is applied. The forecast skill and some circulation indices are evaluated using the 500-mb data. The 300-mb data are used for the barotropic model computation to be presented in section 4. Additional runs were made by the JMA model during the onset of blocking and will be described in section 5.

The ECMWF data consist of daily 10-day forecast fields of 500-mb geopotential height verified during the same 90 days. The forecasts were initialized at 1200 UTC every day, as in the JMA model. NMC forecasts started at 0000 UTC and extended to 10 days. Here 500-mb heights for the same period will be utilized.

During this period, ECMWF and NMC used T106, 19-level and T80, 18-level models, respectively.

The rmse measure of skill is used throughout this study and is evaluated using the 500-mb field. Another common measure, anomaly correlation coefficient, corresponds to the cosine of an angle in phase space between forecast and observed states with respect to the climatological mean. The rms measure is simply a distance in phase space. Therefore, rmse is generally preferred for the ease of interpretation in the context of predictability theory. Generally both measures show similar temporal variability.

3. Skill variation in the 1988/89 winter

Figure 1 shows average growth of rmse computed over the Northern Hemisphere (NH) north of 20°N as a function of forecast day. Ninety JMA forecasts starting from the one initialized at 1200 UTC 1 December were used. The shaded area indicates day-to-day variability of rmse in terms of the average, plus or minus one standard deviation. It is seen that not only the rmse itself but also the variability increases with respect to forecast time. A practical consequence is that beyond 6 days or so the latest forecast may no longer be the best.

Also shown in Fig. 1 by a horizontal, dashed line is the climatological standard deviation of the observed 500-mb heights. It is evaluated using the 90-day winter sample and has a value of 114.6 m. A practical limit of predictability can be defined as the time when the model rmse surpasses that of the climatological forecast. It is seen in the figure that this limit is reached in about 7 days.

Figure 2 illustrates temporal variability of the rmse

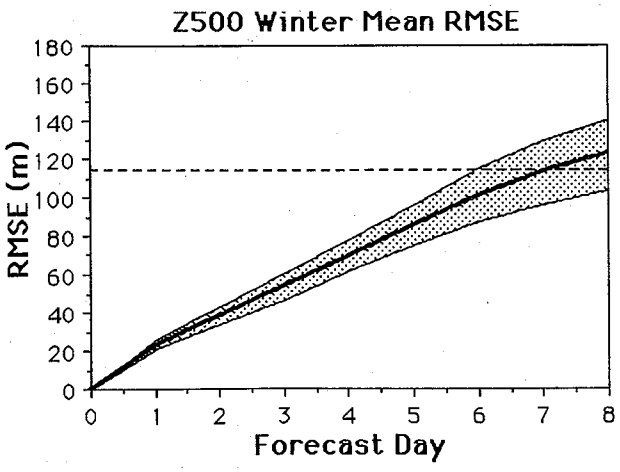

FIG. 1. The rmse of 500-mb geopotential height averaged over the NH north of 20°N. The thick line shows the temporal average over 90 initial dates starting from 1 December 1988 as a function of forecast day. The shaded area indicates the mean plus or minus one standard deviation. For the JMA model. The horizontal dashed line represents the climatological standard deviation.

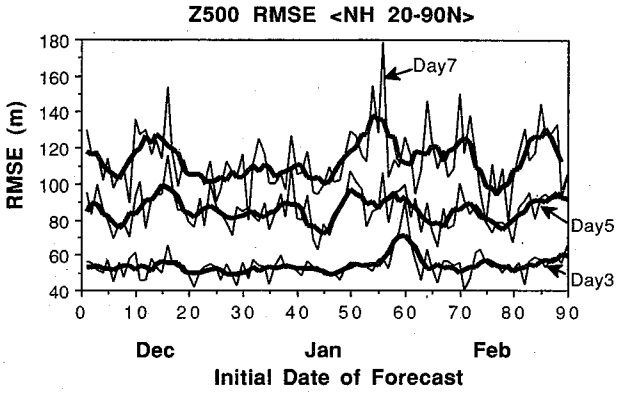

FIG. 2. Day 3, 5, and 7 rmse's of 500-mb geopotential height for the JMA model. The abscissa denotes initial date of forecast, day 1 being 1 December 1988. Thin lines are daily values of rmse. Thick ones are their 5-day running averages. The rmse is taken over the NH north of 20°N as in Fig. 1.

on forecast days 3, 5, and 7 for the JMA model. The day number in the abscissa refers to the initialization date, 1 being 1200 UTC 1 December 1988. Thin lines are daily values and the thick lines show corresponding 5-day running averages. It is seen that in the early part of the MRF, for example, day 3 curve in Fig. 2, the temporal variation in rmse appears more or less random showing little low-frequency variability. However, as the forecast proceeds into day 5 or more, existence of low-frequency variation in the rmse skill becomes more evident.

Figure 3 compares 5-day running averaged rmse's over the NH for day 7 forecasts from the three centers. In this and subsequent figures, the day numbers in the abscissa refer to the verification date, again with 1200 UTC 1 December 1988 being day 1. Curves for the NMC forecasts are shifted by one-half day since they start at 0000 UTC. The figure shows that some simi-

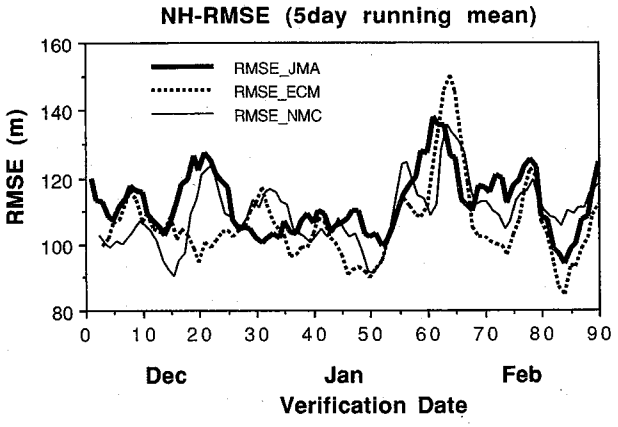

FIG. 3. Day 7 rmse of 500-mb height averaged over the NH north of 20°N for the JMA (thick solid line), ECMWF (dashed line), and NMC (thin solid line) models. The abscissa refers to the verification date, day 1 being 1 December 1988. A 5-day running average is applied for the rmse values.

larity exists among the three curves. Correlated features among all the different models may imply that they are related to the observed circulation characteristics. While all the NWP models could exhibit common deficiencies for particular circulation patterns, it is also likely that they reflected temporal variations in atmosphere's predictability. One of the notable features that appears worthy of discussion is a temporal maximum seen in all the rmse curves around verification dates between 55 and 65.

By examining observed atmospheric evolution during this winter, it was immediately noticed that the rmse maximum was related to a striking development of a blocking in the northeastern Pacific near the Gulf of Alaska, which occurred in the last pentad of January 1989. This is demonstrated by a zonal index defined by the 500-mb geopotential height difference between 40° and 60°N (cf., Lejenäs and Økland 1983) averaged over the longitudes between 160°E and 110°W, as shown in Fig. 4. An abrupt transition from zonal to blocked circulation regimes in the above region, as exemplified by the striking drop of the index around day 60 in Fig. 4, corresponds well to the maximum in rmse over the NH as seen in Fig. 3. This correspondence is also visible even when a zonal average is adopted over all longitudes in the definition of the zonal index (not shown). Thus, although considerations on regionality in atmospheric circulation and in model skills are generally important (Dalcher et al. 1988; Kalnay and Dalcher 1987), they may not be seriously demanded for this outstanding event.

The association between the maximum in the rmse and the blocking transition is further confirmed by Figs. 5 and 6. Figure 5 shows a measure of transience of the observed atmospheric evolution over the NH. The transiency index shown is defined by the rms difference between two 500-mb geopotential height fields

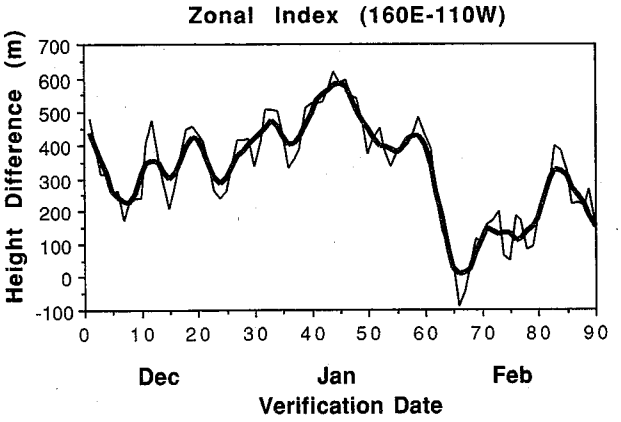

FIG. 4. Observed zonal index, that is, the difference of the 500-mb geopotential heights at 40° and 60°N, averaged over the longitudes between 160°E and 110°W. The abscissa is the same as in Fig. 3. The daily values are shown by the thin solid line. The thick line is the 5-day running averages.

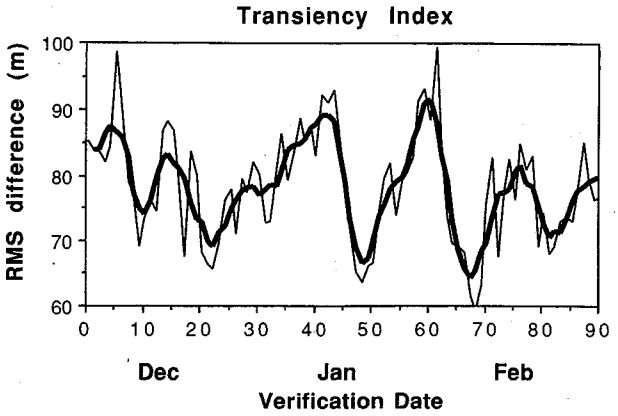

FIG. 5. Transiency index of observed 500-mb fields, defined by the rms difference between consecutive daily analyses. The rms difference is computed over the NH north of 20°N as before. Thin and thick solid lines show daily and 5-day running averaged values. The abscissa is the same as in Fig. 3.

observed 24 h apart. The thin solid curve represents the raw daily values, and the thick solid one shows 5-day running averages. The local maxima in the curve indicate that the observed circulation underwent rapid changes, while minima are characterized by relatively slow evolutions, that is, by persistence. A distinct maximum characterizing large transiency is seen slightly before day 60 in Fig. 5, intervened by two well-defined minima around day 49 and 68. Figure 6 shows 500-mb geopotential height maps on four characteristic days, that is, day 49 (1200 UTC 18 January 1989; hereafter abbreviated as 1812, with the first two digits denoting the calendar date and the last two referring to the maptime) near the bottom of the transiency curve in Fig. 5, day 56 (2512) and 61 (3012) during the "transition" and day 65 (0312) near the bottom of the second transiency minimum. On day 49 (Fig. 6a) the circulation over the North Pacific was zonal with no significant meander. This zonal flow started meandering several days later as seen in Fig. 6b (day 56). A meridional elongation of the disturbance seen in Fig. 6c (day 61) lead to the formation of the Gulf of Alaska block in the beginning of February (Fig. 6d, day 65).

The association between the transiency and NWP model skill observed here during the onset of blocking is similar to the results reported by Palmer and Tibaldi (1988) and Chen (1989). However, this association is not evident for the whole winter. Between days 50 and 90, the transiency curve tends to lead rmse by 2 days. But before day 50, they appear to have a weak anticorrelation without lag.

Operational blocking forecasts by the ECMWF model have recently been investigated by Tibaldi and Molteni (1990) for seven winters. They showed that the model was not, on average, satisfactory in simulating onsets of blocking in the medium range. On the

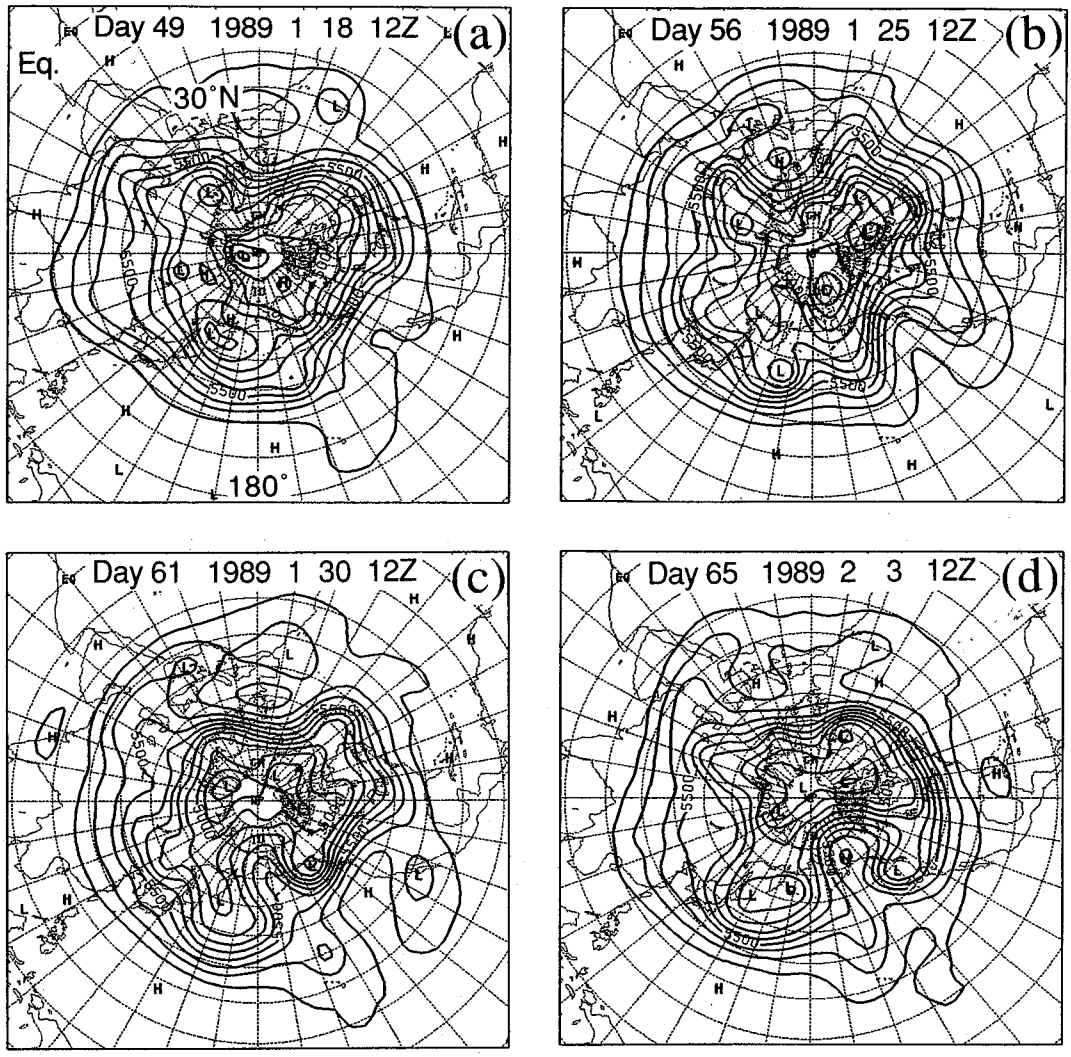

FIG. 6. The 500-mb geopotential heights observed at 1200 UTC of the following dates; (a) 18 January (day 49), (b) 25 January (day 56), (c) 30 January (day 61), and (d) 3 February (day 65) 1989. Contour interval is 100 m. JMA analyses are used.

other hand, Frederiksen (1989) indicated, by performing a three-dimensional, linear instability analysis of observed synoptic flows, that an intrinsic atmospheric instability may have been associated with the onset of a Gulf of Alaska blocking occurring in November 1979. Furthermore, his auxiliary computations suggested that barotropic processes may have been an important factor for this instability. The next two sections will try to show how these two factors, the model deficiency and the atmospheric instability, contributed to the rmse maximum in our case.

4. Forecast spread and atmospheric instability

a. Forecast spread

Figure 7 shows 5-day–averaged time series of the rms difference between 7- and 8-day forecasts verified on the same day for the JMA, ECMWF, and NMC models. An index similar to this was used as a measure of forecast spread by Dalcher et al. (1988), Palmer and Tibaldi (1988), and others. It is seen that a distinct temporal maximum occurs in the JMA curve (thick solid line) between days 57 and 67, which is concurrent with the rmse maximum seen in Fig. 3 above. The ECMWF and NMC curves also show similar bumps relative to adjacent minima, although they do not reach as high a level as the JMA curve. The JMA and ECMWF models show another bump around day 17, a few days prior to rmse maxima of the JMA and NMC models. It is noted that the magnitude of the spread index of the NMC model is consistently smaller than those of the other two centers. These qualitative features did not change when the spread was computed using some other forecast days (not shown).

The correlation between rmse and forecast spread is

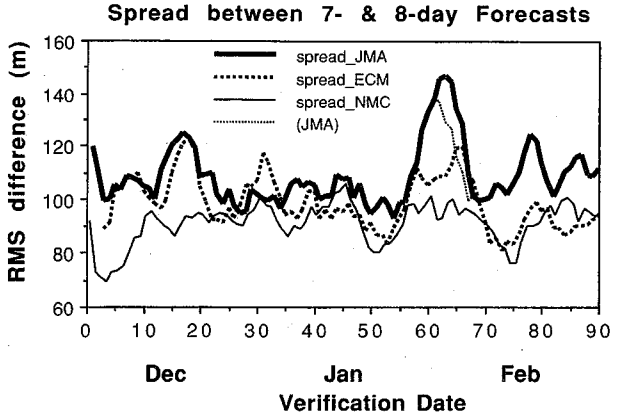

FIG. 7. The rms difference averaged over the NH north of 20°N between the day 7 and day 8 forecast 500-mb heights verified on the same verification dates. Thick solid, thick dashed, and thin solid lines are for the JMA, ECMWF, and NMC models, respectively. A 5-day running average is applied. The abscissa is the same as in Fig. 3. The thin dotted line is explained and discussed in section 5.

not so high for the winter as a whole. Neither is the covariability among the three centers. However, both are definitely better around day 60 than any other period in this winter. The relative maximum in forecast spread common to the three models suggests that the atmospheric instability played an important role during the onset of blocking.

b. A linear measure of forecast spread

Lorenz (1965) presented a theoretical framework with which one can obtain a linear measure of forecast skill within the context of a perfect model. It has also been discussed recently by Farrell (1990) and by Mukougawa et al. (1991) and is briefly outlined below.

An evolution equation of a (discrete) nonlinear dynamical system is written symbolically as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \tag{1}$$

where the dot denotes a time derivative and $\mathbf{f}$ is a nonlinear function of N-dimensional state vector $\mathbf{x}$. The evolution of a small "error", that is, the distance $\epsilon(t)$ from the reference trajectory $\mathbf{x}(t)$, is governed by tangential linear equation of (2):

$$\dot{\epsilon} = \frac{D\mathbf{f}}{D\mathbf{x}}[\mathbf{x}(t)]\epsilon, \tag{2}$$

where $D\mathbf{f}/D\mathbf{x}$ is the Jacobian matrix and is a function of $\mathbf{x}(t)$. Lorenz (1965) showed that those points that lay on a unit hypersphere $|\epsilon| = 1$ at time $t = t_0$ evolve to lie at time $t = t_0 + \tau$ on a hyperellipsoid centered at $\mathbf{x}(t_0 + \tau)$. The principal axes of this ellipsoid are given by the eigenfunctions of the error covariance matrix $\mathbf{X}$ defined as

$$\mathbf{X} = \mathbf{A}\mathbf{A}^{\mathrm{T}}, \tag{3}$$

where the $N \times N$ matrix $\mathbf{A}$ maps $\epsilon(t_0)$ to $\epsilon(t_0 + \tau)$ and is dependent on the initial time t_0 as well as the forecast time τ. The superscript T denotes transpose. Writing eigenvalues of $\mathbf{X}$ as Γ_i ($i = 1, \cdots, N$), the rmse growth rate, or *Lorenz index*, $\alpha(t; \tau)$, is defined by

$$\alpha(t; \tau) \equiv \left(\frac{1}{N}\sum_{i=1}^{N}\Gamma_i\right)^{1/2} = \left[\frac{1}{N}\mathrm{trace}(\mathbf{X})\right]^{1/2}. \tag{4}$$

Obviously, α is also a function of initial state $\mathbf{x}(t_0)$ and the forecast time τ. Although the relation between α and $\mathbf{x}(t_0)$ is not trivial, Mukougawa et al. (1991) discusses a simple relation between the transiency $|\dot{\mathbf{x}}|$ and α when $N = 1$.

While the Lorenz index is theoretically appealing as a measure of error growth, it is difficult to evaluate it with a full NWP model because of the large matrices involved. Use of a full model may not only be impractical but also inappropriate since rapid small-scale instabilities that saturate in a matter of days may contaminate the assessment of larger-scale signatures (cf., Mukougawa et al. 1991). Therefore, in this practical study, the Lorenz index is evaluated using a simple, hemispheric nondivergent barotropic model with T21 truncation ($N = 231$) and with 300-mb streamfunction fields, observed or forecast by the NWP model, as reference trajectories. The barotropic model has no topography, and an Ekman damping of 10-day relaxation time and a very weak ∇^4 hyperviscosity (2 days at the truncation wavenumber) are included.

The Lorenz index (LI) was computed using both observed and JMA forecast 300-mb streamfunction ψ fields as the reference. The reference ψ field was linearly interpolated in time using the 24-h archive. Simply keeping it constant for one day did not change the results. The time integration of (2) is performed using a forward scheme with a time step of 15 min. It is noted that LI is dependent on the choice of inner product (3). Here the areal average of the product over the NH is chosen. Figure 8 shows the results for $\tau = 7$ days. The solid line is for the observed ψ, and the dashed one for the forecast. No running mean is applied. The day number refers again to the verification date. The LI with the observed reference state shows some resemblance to the rmse or the spread curves shown in Figs. 3 and 8, except for a large bump slightly before day 50. Another bump after day 60 appears to correspond to the rmse and spread maxima. While this correspondence gives another support to the involvement of atmospheric instability during the onset of blocking, such an association cannot be claimed for the rest of the period.

The LI with forecast reference states (dashed line) shows a good correlation with the one based on observed reference, but it is much noisier when reflecting the sensitivity of the LI computation to the reference states. Note that in the case of forecast reference states, the daily values are evaluated along separate trajectories, while with the observed reference states they are computed along a single trajectory throughout. Un-

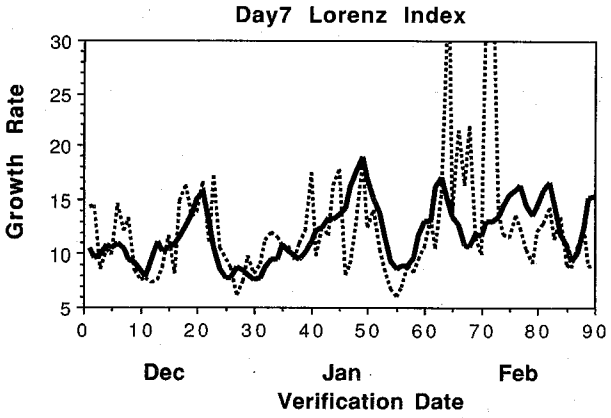

FIG. 8. Lorenz index integrated up to forecast day ($= \tau$) 7 using observed (solid) and JMA forecast (dashed) reference states. The abscissa refers to the verification dates as in Fig. 3. The ordinate is the rms growth rate [cf., Eq. (5)]. No temporal smoothing is applied to the curves.

fortunately, the dashed line shows excessive error growth during the blocked circulation and misses the maximum around day 60. This may have resulted from poor representation of the blocked circulation in the NWP model. It appears that the predictive utility of the index is limited, at least with the simple computational setting adopted here. Neither the inclusion of divergent components in the reference state nor the use of other norms, that is, kinetic energy and enstrophy, changed the qualitative nature of both of the curves shown in Fig. 8.

5. Model behavior during the onset of blocking

This section focuses on the NWP models' behavior during the onset of blocking. In order to complement the operational dataset, 10-day forecasts initialized every 6 h between 2412 and 2806, inclusive, were generated by the JMA model. A T42 version of the model was also integrated starting every 12 h between 2412 and 2712. As will be seen below, a JMA operational forecast initialized at 2612 showed an exceptional performance in simulating the onset of blocking. Two perturbed runs were produced using this integration, called the 2612 run. The first perturbed run was initialized with an arithmetic average between 2612 initial condition and a 6-h forecast initialized at 2606. The second one used an average between the 6-h forecast of 2612 run and 2618 analysis as the initial state. In effect, the analysis increments of the 2612 and 2618 initial conditions have been halved. These perturbed runs will be denoted by 2612′ and 2618′, respectively.

A diagram shown in Fig. 9 conveniently summarizes the features that are discussed. All the available forecasts on day 65, that is, 41 JMA and ECMWF runs verified at 0312 and 10 NMC runs verified at 0300, are plotted according to their zonal-index (ZI) and rmse values computed over the Pacific sector between 160°E and 110°W. Referring back to Figs. 4, 5, and 6, this verification date corresponds to the establishment of the blocking. Those with forecast days $\tau < 5$, $5 \leq \tau \leq 7$, and $7 < \tau \leq 10$ are denoted by open and closed circles and open squares, respectively. The verifying analysis at 0312 is denoted by an open circle on the ordinate. Observed ZI value at 0300 is only 18 m greater than that of 0312 and is omitted.

One of the distinct features seen in Fig. 9 is that the points tend to cluster in two branches, the upper and the lower. The dashed and dotted lines are attached to them for reference. The upper, more-zonal branch accompanied by the dashed line consists of runs that extended more than 5 days, while the lower, blocked branch with the dotted line is formed by those run for 5 days or less, with an exception to be discussed later.

In the forecast time-range of 6–10 days, most of the forecast ZIs did not drop to the level of the verifying analysis; that is, the MRFs tended to remain zonal. Three closed circles lying in the middle of two branches and another one in the lower branch are 5- and 6-day forecasts. A notable exception to this is an 8-day operational JMA forecast initialized at 2612 denoted in the figure by (a). This forecast has the lowest value of ZI and a fairly small value of rmse. Figure 10a shows 500-mb height field of this forecast verified at 0312. The verifying analysis is shown in Fig. 6d. Figure 10b is the same but for the JMA 2712 forecast, which was initialized 24 h later. The clearest contrast can be seen

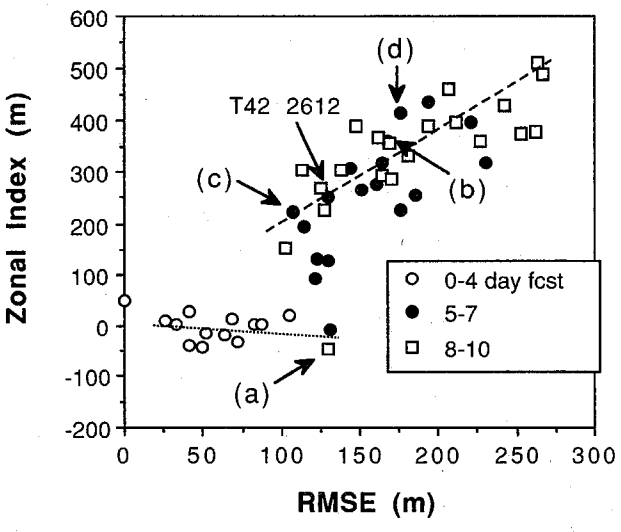

FIG. 9. A scatter diagram of forecasts verified at 0312 (day 65) in terms of rmse (abscissa) and zonal index (ordinate), both of which are computed over the Pacific sector between 160°E and 110°W. The verification time 0300 is used in the case of NMC forecasts. Forecasts of 4 days duration or less are denoted by open circles, while those between 5 and 7 days have closed circles. Forecasts of 8 days or more are shown by open squares. The zonal index of the verification map is indicated on the ordinate. Letters (a), (b), (c), and (d) refer to those forecasts shown in Fig. 10. The JMA T42 run initialized at 2612 is also indicated by an arrow. The dashed and dotted lines mark the upper and lower branch discussed in the text.

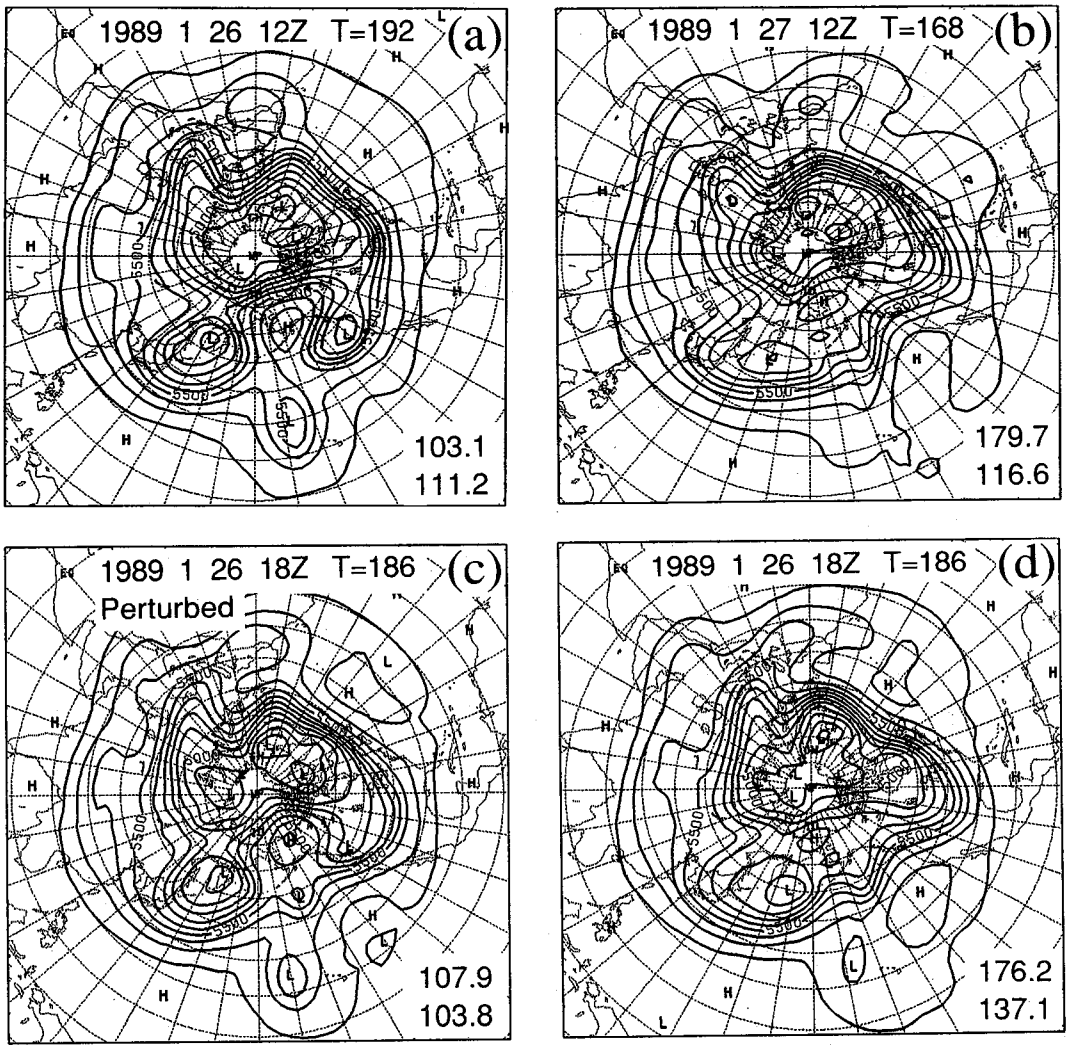

FIG. 10. Forecast 500-mb geopotential height fields for 0312 (day 65). Initialization times are (a) 2612, (b) 2712, (c) 2618′, and (d) 2618. The forecast shown in (c) was started from an average between 2618 analysis and a 6-h forecast initialized at 2612. Numerical values in the bottom right corners denote rmse scores averaged over the North Pacific (160°E–110°W) in the top and over the entire NH in the bottom, respectively.

between them in the blocking region. The former managed to simulate the split flow, but the latter failed.

Shown in Figs. 10c and 10d are 2618′ and 2618 forecasts, respectively. The 2618 (Fig. 10d) remained zonal and looks quite similar to the 2712 in Fig. 10b, while the perturbed run 2618′ (Fig. 10c) is more "blocked." The forecasts shown in Fig. 10 are denoted in the scatter diagram, Fig. 9, by the corresponding letters. It is noted that the 2618 forecast (Fig. 10d) having relatively large ZI and rmse values was brought down along the upper branch to have smaller, therefore better, values in the perturbed run 2618′ (Fig. 10c). This pair serves to demonstrate the large sensitivity to the initial conditions during this period. It was also seen that a lower-resolution T42 run initialized at 2612 (indicated in Fig. 9 by an arrow) was much better than some of the T63 runs initialized later, both in terms of rmse's and synoptic impressions.

While the operational 2612 run was exceptional, it did not affect the qualitative aspect of the spread curve of the JMA model shown by the thick solid line in Fig. 7. The thin dotted line of Fig. 7 is the spread computed using the much poorer 2618 forecast in the upper branch in place of the 2612. This replacement did not make the bump disappear. Since no operational JMA forecasts initialized prior to 2612 entered the lower branch, it is concluded that the temporal maximum in the forecast spread discussed in section 4a refers to the scatter of the points in the upper branch of Fig. 9. Within this branch, points are not aligned chronologically along the dashed line, as can be seen by the mixture of closed circles and open squares. Therefore, although the forecasts in the range 6–10 days were systematically deficient in attaining the observed low ZI value, they showed higher than average sensitivity on initial conditions.

In order to confirm that the principal spatial patterns among the forecasts were related to the blocking in a more objective manner, empirical orthogonal functions (EOFs) of an ensemble of 39 500-mb height maps were computed using 5–10-day forecasts that appeared in Fig. 9. The spatial pattern of the first EOF is shown in Fig. 11. This is associated with 21.1% of the total NH variance around an ensemble average. Clearly, one of the maximum sensitivities is seen over the Gulf of Alaska near the center of the blocking high. (The polarity is arbitrary for an EOF.) The second EOF mode has a much smaller variance of 15.4%. The pattern of the leading EOF is virtually unaltered when three 5-day forecasts and the JMA 2612 run are excluded from the computation.

It is discussed in the Introduction that an ensemble of points, spherically distributed in phase space, evolves initially into an ellipsoid. The ensemble average would coincide with the observed state if the model were perfect. If an NWP model ensemble correctly captures the principal axis of the ellipsoid and if the eccentricity of the ellipsoid is sufficiently large, the projections of individual members of the ensemble on this axis should be correlated with their rmse values. The sign of correlation depends on which side of the axis they reside. When the projection is approximated by ZI, the two branches seen in Fig. 9 may be interpreted as a manifestation of such a feature. The difference in the slope between the branches may suggest asymmetry in probability density distribution around the verifying observation. Forecasts longer than 4 days may have represented the spread on one side of the density distribution.

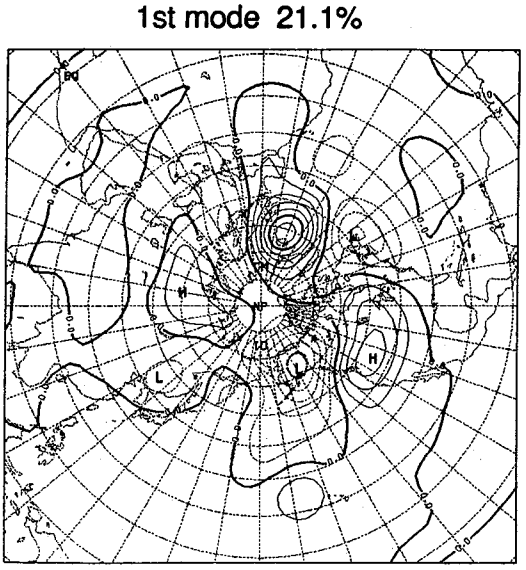

FIG. 11. The leading eigenfunction of an error covariance matrix estimated by taking an ensemble average of 39 forecasts generated by the JMA, ECMWF, and NMC models verified at 0312 (0300 in the case of NMC). Contour interval and polarity are arbitrary.

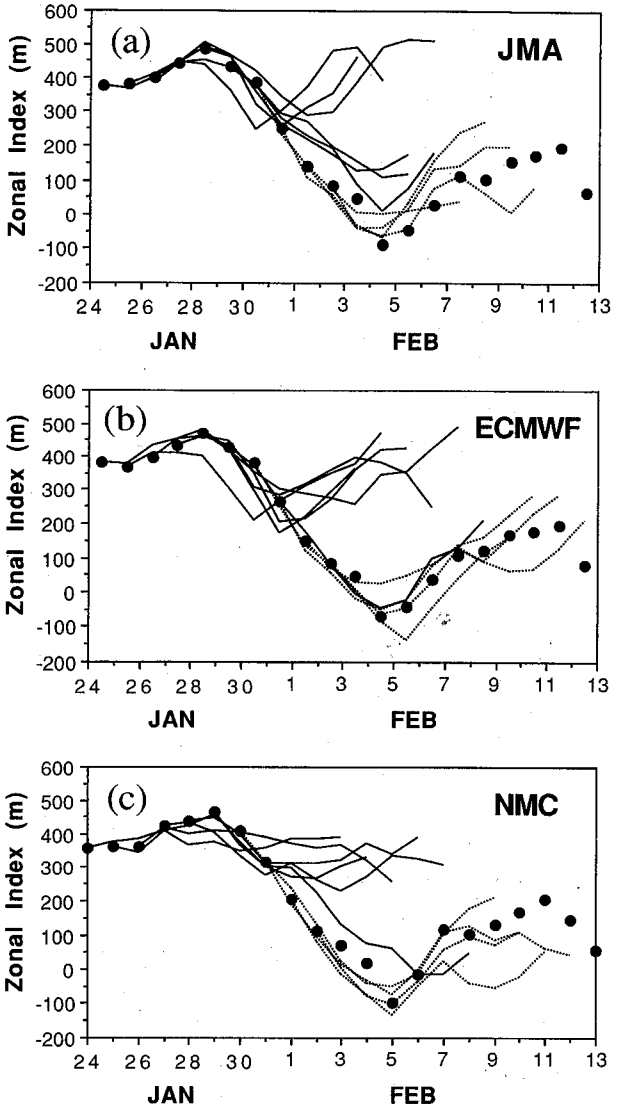

FIG. 12. Zonal indices of observed (closed circles) and forecast (solid and dotted lines) 500-mb heights as defined in Fig. 4. The (a) JMA, (b) ECMWF, and (c) NMC models and analyses are used. Only those forecast that are operational and initialized before 0300 are plotted. The solid and dotted lines denote forecasts initialized before and after 2900, respectively. The abscissa refers to the calender dates from 24 January to 13 February 1989. Tick marks are placed at 0000 UTC.

The forecasts in the lower branch in Fig. 9 have definitely smaller rmse and ZI values than those in the upper branch. However, it is discouraging, from the point of view of forecasting, that they only consist of forecasts extended for less than 6 days except for the JMA 2612 run. This tendency of the NWP models, that is, the "reluctance" to enter the blocking regime, is also noted by Tibaldi and Molteni (1990) and may be related to a common deficiency of the models. We note, however, that the simulated blockings tended to persist once they were established. Figure 12 shows time

evolutions of ZI in the observation and operational forecasts in the three centers separately. The observed values are denoted by closed circles. Only operational forecasts initialized prior to 0312, that is, those that appeared in Fig. 9, are plotted. In Fig. 12, forecasts that had open circles in Fig. 9 are shown by dotted lines, and others are drawn by solid lines. In all panels, the discrete nature of zonal and blocked circulations is suggested; that is, the forecast blockings tended to persist once they were established, while those forecasts that failed consistently remained zonal. Forecasts that had ZI values around 200 m or less at 0312 continued to stay so until 0712 or longer.

Finally, in order to investigate critical synoptic features that lead to the zonal-or-block dispersion, all the forecasts that were integrated at least up to 0312 were examined and grouped into two according to their ZI values. In the first "successful" group of runs, the ZI values remained around or below 200 m during a verification time range between 0312 and 0712. Nine runs satisfied this criterion, and another 30 formed the second group. These two groups' 500-mb height patterns during the onset of blocking are examined in Fig. 13. The two panels show the averages of the "successful" and "unsuccessful" runs, respectively, at a verification time 3012. Note that this verification date lies slightly before the "bifurcation" point in Fig. 12. The largest difference between the two maps is seen along 40°N between 170° and 150°W where weaker westerly and significant deformation prevailed in the "successful" composite (Fig. 13a). In forecasts that had this feature, subsequent synoptic disturbances entering this region were deformed in the north–south direction and transported warmer air mass northward ahead of the cyclones, in such a way as to strengthen the cutoff high (figures not shown). In contrast, the poorer forecasts lacked the deformation region as seen in Fig. 13b and were characterized by a much stronger jet between 170° and 150°W. Disturbances were ducted through this westerly toward the east without being subject to significant north–south stretching. Therefore, the cutoff high being formed in Fig. 13 got detached from the westerly, could not maintain its strength, and weakened considerably by 0312. The positive-feedback mechanism of the synoptic disturbances (e.g., Shutts 1983) seems to have been turned off in the forecasts that ended with zonal circulations.

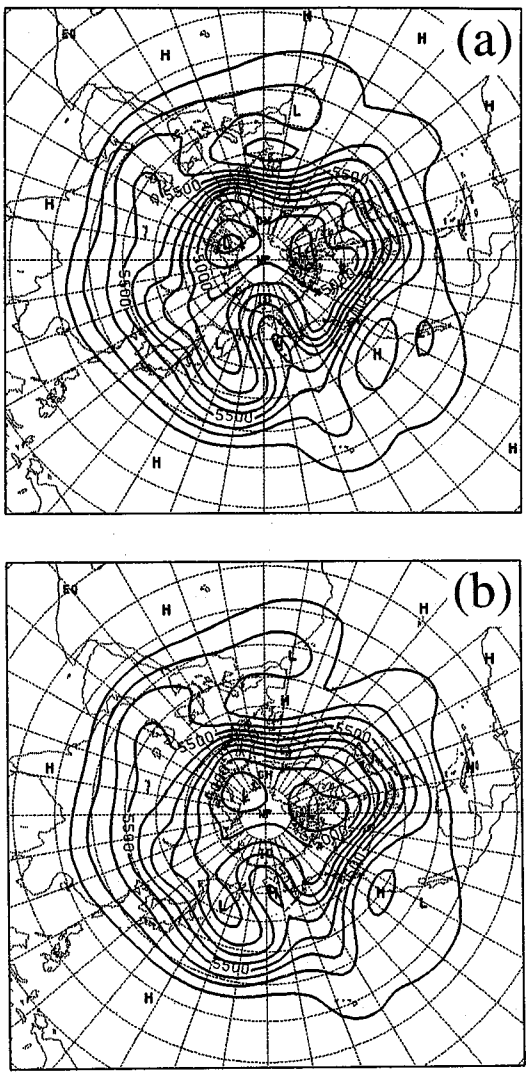

FIG. 13. Averages of 500-mb forecast maps verified at 3012 (day 61; 3000 in the case of NMC forecasts) over (a) 9 "successful" and (b) 30 "poor" ones.

6. Discussion

In a probabilistic sense, the study of forecast skill involves a comparison of density distribution in phase space between model and real atmospheres. Therefore, one has to face an inherent difficulty since reliable estimates are hard to obtain even for the former, not to mention the latter. Nevertheless, the foregoing analysis of the MRFs during the onset of a blocking points to potentially important aspects for better understanding of the extended-range predictability.

First of all, there are indications that the transition from zonal to blocked circulation regimes observed at the end of January 1989 was characterized by an above-average rate of divergence in nearby atmospheric trajectories. Suppose that the largest rmse observed here during the transition was in most part due to this intrinsic instability of the atmosphere, then one is tempted to conclude that extended-range forecasting beyond an occurrence of blocking may be extremely difficult and may not be practically feasible.

On the other hand, the existence of two distinct branches in the scatter diagram shown in Fig. 9 is suggestive of a highly nonlinear nature with regard to the transition. Furthermore, the rather robust residence of the NWP models in either one of the regimes seen in Fig. 12 implies the existence of self-sustaining mechanisms in each of the regimes and processes sen-

sitive to initial conditions at the boundary between the two. If the atmospheric evolution can be viewed in this way, the above pessimism on extended-range predictability should be reserved until such questions are answered as to what their physical and dynamical entities are, how they are turned on, and how much indeterminacy comes into play. Obviously, it is crucial for us to know whether the models are capable of representing such processes.

For the purpose of delineating finite-amplitude behavior of ensemble averages in the extended-range a little more physically, their time evolution is written as follows:

$$\frac{\partial\langle q\rangle}{\partial t} + \langle \mathbf{v}\rangle\cdot\nabla\langle q\rangle + \langle \mathbf{v}'\cdot\nabla q'\rangle = 0, \qquad (5)$$

where the bracket $\langle\ \rangle$ denotes the ensemble average, the prime denotes the deviation from the ensemble average, and q is the potential vorticity. From (5), it can be seen that the evolution of ensemble averaged state acquires the feedback from the primed fields and deviates from the reference trajectory of the tangential linear computation. For now, call the ensemble average a "predictable" component and the deviation an "unpredictable" component. Many of the recent studies (e.g., Shutts 1983; Illari 1984; Mullen 1987) show that the synoptic, "unpredictable" component of the flow feeds back positively to the blocking-type, "predictable" flows. This process is associated with enhanced downscale cascade of enstrophy (Shutts 1983) that leads to the growth of eddy, or "error," enstrophy. When an NWP model underscores this effect, it may produce a biased estimate of density distribution in an ensemble forecast of a blocking, even if its dynamics for the "predictable" component (i.e., $\langle \mathbf{v}\rangle\cdot\nabla\langle q\rangle$ in the above) is perfectly correct. Thus, a model that lacks "vigor" in the unpredictable component and/or misrepresents its interactions with the "predictable" counterpart may be seriously misleading. Several investigators have reported that the onsets of blocking are very sensitive to the synoptic-scale components of the flow (Bengtsson 1981; Tibaldi and Ji 1983; Tracton 1990). Such sensitivity is also noted in section 5 with regard to Fig. 13. More detailed studies are required to clarify the question of the scale-interaction process and predictability.

7. Summary and concluding remarks

In this paper, we examined low-frequency variation in the medium-range forecast skill of operational numerical weather predictions (NWPs). Forecasts from three operational centers were analyzed with emphasis on the Japan Meteorological Agency (JMA) model. Within the case study period of 1988/89 winter, the root-mean-square error (rmse) of operational forecasts showed a pronounced peak during the onset of a Gulf of Alaska blocking at the end of January 1989. This time period was also characterized by a higher than average transience of the atmospheric circulation. The rmse and forecast dispersion computed for the three NWP models showed qualitatively similar temporal variability during this period, suggesting that the large error was related to the characteristics of the atmospheric flows. Forecast dispersion and a barotropic measure of divergence of nearby trajectories in phase space showed temporal maxima concurrent with the rmse peak. Therefore, it is likely that the instability of instantaneous flows intrinsic to the atmosphere contributed to the large rmse. However, these associations did not hold for other time periods.

The forecast dispersion prior to the blocking is studied in detail. High sensitivity on initial conditions during this period is confirmed by making 6-h and some other extra runs by the JMA model. The principal spatial pattern of the spread among different forecasts showed high sensitivity over the blocking region. In addition, it is noted that all three models examined show considerable "reluctance" to enter the blocked regime unless they are initialized 5 days or less prior to the blocking. Once the block is established in the model, however, it tends to persist. While the systematic underestimation of the blocked circulation in the medium range (5–10 days) may be due to NWP models' deficiency, these sensitivities on initial conditions may not be ascribed solely to the models.

The generality of the present results has to be reserved since we have only dealt with one case. The observations made here that are worth further quantification include (i) the covariability of skills among different forecast centers, (ii) the generality of association between relatively poor skill and circulation transitions, and (iii) the relative importance of model deficiency and intrinsic instability of the atmosphere. Item (i) can be investigated with larger datasets. Item (ii) may first be investigated with theoretical and mechanistic models with the perfect-model assumption. Linear computations of the type presented in section 4b may be valuable in such studies. It is also of interest to look at larger operational datasets from such a viewpoint (cf., Tibaldi and Molteni 1990). Item (iii) is the most difficult of all and is unlikely to be resolved thoroughly. However, inquiries into the nature of large rmse through more detailed case studies or statistical analyses should give useful insights into the model development activities and provide a better scope for the extended-range predictability.

Acknowledgments. ECMWF and NMC data were kindly provided by Drs. R. Mureau and S. Saha. Discussions with Drs. M. Ghil, E. Kalnay, K. Miyakoda, K.-C. Mo, and J. Steppeler were helpful. M.K. wishes to thank Drs. D. Rodenhuis, H. van den Dool, and M. Kanamitsu for their hospitality during a two-week stay at NMC. Part of this work was performed while M.K. was staying at the Department of Atmospheric Sciences

of University of California, Los Angeles, as a research associate. He was supported partly by National Science Foundation Grants ATM86-15424 and ATM90-13217 and by NOAA Grant DOC A90AA-D-MC076.

REFERENCES

Bengtsson, L., 1981: Numerical prediction of atmospheric blocking. *Tellus,* **33,** 19–42.

Branstator, G., 1986: The variability in skill of 72-hour global-scale NMC forecasts. *Mon. Wea. Rev.,* **114,** 2628–2639.

Chen, W., 1989: A new method in forecasting forecast skill. *Mon. Wea. Rev.,* **117,** 427–435.

Curtis, J., J. S. Boyle, and C. H. Wash, 1988: Variability in skill in 120 h FNOC 500 mb height forecasts. *Mon. Wea. Rev.,* **116,** 1381–1388.

Dalcher, A., E. Kalnay, and R. Hoffman, 1988: Medium-range lagged average forecasts. *Mon. Wea. Rev.,* **116,** 402–416.

Farrell, B. F., 1990: Small error dynamics and the predictability of atmospheric flows. *J. Atmos. Sci.,* **47,** 2409–2416.

Frederiksen, J. S., 1989: The role of instability during the onset of blocking and cyclogenesis in Northern Hemisphere synoptic flows. *J. Atmos. Sci.,* **46,** 1076–1092.

Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting: An alternative to Monte Carlo forecasting. *Tellus,* **35A,** 100–118.

Illari, L., 1984: A diagnostic study of the potential vorticity in a warm blocking anticyclone. *J. Atmos. Sci.,* **41,** 3518–3526.

Kalnay, E., and A. Dalcher, 1987: Forecasting forecast skill. *Mon. Wea. Rev.,* **115,** 349–356.

Legras, B., and M. Ghil, 1985: Persistent anomalies, blocking, and variations in atmospheric predictability. *J. Atmos. Sci.,* **42,** 433–471.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.,* **102,** 409–418.

Lejenäs, H., and H. Økland, 1983: Characteristics of Northern Hemisphere blocking as determined from a long time series of observational data. *Tellus,* **35A,** 350–362.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus,* **17,** 321–333.

Mukougawa, H., M. Kimoto, and S. Yoden, 1991: A relationship between local error growth and quasi-stationary states: Case study in the Lorenz system. *J. Atmos. Sci.,* **48,** 1231–1237.

Mullen, S. L., 1987: Transient eddy forcing of blocking flows. *J. Atmos. Sci.,* **44,** 3–22.

Palmer, T. N., 1988: Medium and extended range predictability, and stability of the PNA mode. *Quart. J. Roy. Meteor. Soc.,* **114,** 691–713.

——, and S. Tibaldi, 1988: On the prediction of forecast skill. *Mon. Wea. Rev.,* **116,** 2453–2480.

Roads, J. O., 1987: Predictability in the extended range. *J. Atmos. Sci.,* **44,** 3495–3527.

——, 1988: Lagged average predictions in a predictability experiment. *J. Atmos. Sci.,* **45,** 147–162.

Shutts, G. J., 1983: The propagation of eddies in diffluent jet streams: Eddy vorticity forcing of "blocking" flow fields. *Quart. J. Roy. Meteor. Soc.,* **109,** 737–761.

Sugi, M., K. Kuma, K. Tada, K. Tamiya, N. Hasegawa, T. Iwasaki, S. Yamada, and T. Kitade, 1990: Description and performance of the JMA operational global spectral model (JMA-GSM88). *Geophys. Mag.,* **43,** 105–130.

Tennekes, H., A. P. M. Baede, and J. D. Opsteegh, 1987: Forecasting forecast skill. *Proc. ECMWF Workshop on Predictability in the Medium and Extended Range,* Reading, U.K. European Centre for Medium-Range Weather Forecasts, 277–302.

Tibaldi, S., and L. R. Ji, 1983: On the effect of model resolution on numerical simulation of blocking. *Tellus,* **35A,** 28–38.

——, and F. Molteni, 1990: On the operational predictability of blocking. *Tellus,* **42A,** 343–365.

Tracton, M. S., 1990: Predictability and its relationship to scale interaction processes in blocking. *Mon. Wea. Rev.,* **118,** 1666–1695.

——, K.-C. Mo, W. Chen, E. Kalnay, R. Kistler, and G. White, 1989: Dynamical extended range forecasting (DERF) at the National Meteorological Center. *Mon. Wea. Rev.,* **117,** 1604–1635.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.,* **109,** 784–812.