Improved Skill for the Anomaly Correlation of Geopotential Heights at 500 hPa

T. N. KRISHNAMURTI, K. RAJENDRAN, AND T. S. V. VIJAYA KUMAR

Department of Meteorology, The Florida State University, Tallahassee, Florida

STEPHEN LORD, AND ZOLTAN TOTH

Environmental Modeling Center, National Centers for Environmental Prediction, Camp Springs, Maryland

XIAOLEI ZOU, STEVEN COCKE, AND JON E. AHLQUIST

Department of Meteorology, The Florida State University, Tallahassee, Florida

I. MICHAEL NAVON

CSIT and Department of Mathematics, The Florida State University, Tallahassee, Florida

(Manuscript received 7 October 2002, in final form 27 November 2002)

ABSTRACT

This paper addresses the anomaly correlation of the 500-hPa geopotential heights from a suite of global multimodels and from a model-weighted ensemble mean called the superensemble. This procedure follows a number of current studies on weather and seasonal climate forecasting that are being pursued. This study includes a slightly different procedure from that used in other current experimental forecasts for other variables. Here a superensemble for the ∇^2 of the geopotential based on the daily forecasts of the geopotential fields at the 500-hPa level is constructed. The geopotential of the superensemble is recovered from the solution of the Poisson equation. This procedure appears to improve the skill for those scales where the variance of the geopotential is large and contributes to a marked improvement in the skill of the anomaly correlation. Especially large improvements over the Southern Hemisphere are noted. Consistent day-6 forecast skill above 0.80 is achieved on a day to day basis. The superensemble skills are higher than those of the best model and the ensemble mean. For days 1–6 the percent improvement in anomaly correlations of the superensemble over the best model are 0.3, 0.8, 2.25, 4.75, 8.6, and 14.6, respectively, for the Northern Hemisphere. The corresponding numbers for the Southern Hemisphere are 1.12, 1.66, 2.69, 4.48, 7.11, and 12.17. Major improvement of anomaly correlation skills is realized by the superensemble at days 5 and 6 of forecasts. The collective regional strengths of the member models, which is reflected in the proposed superensemble, provide a useful consensus product that may be useful for future operational guidance.

1. Introduction

In several operational numerical weather prediction centers, the anomaly correlation of 500-hPa forecasts has always been used as a measure of the models' overall performance. Professor Fred Sanders from MIT (see Table 1 for a list of acronyms and their definitions) has been a frequent lecturer at FSU in recent years. One of his favorite comments was that the 500-hPa anomaly correlation (a measure of skill of 500-hPa geopotential height forecasts) was not weather. He always wondered why so much was said on that skill parameter while assessing the relative performance of models when in fact what mattered was the rain and the severe weather. The counter argument he usually received was that if the troughs and ridges were not correctly placed, the likelihood of a good weather forecast were slim. For what it may have been worth, an anomaly correlation index has been used uniformly over several decades by the weather services of the world to assess the performance of their models.

The motivation for this paper emerged from our recent examination of anomaly correlations of 500-hPa heights in the context of a multimodel superensemble following several of our recent studies (Krishnamurti et al. 1999, 2000a,b, 2001). The multimodel superensemble is best explained by the schematics in Fig. 1. The word ''superensemble'' was used by the first author of this paper in a series of publications, only to stress the fact that this ensemble does carry the highest skill compared to participating member models of the ensemble

Corresponding author address: T. N. Krishnamurti, Dept. of Meteorology, The Florida State University, Tallahassee, FL 32306-4520. E-mail: tnk@io.met.fsu.edu

TABLE 1. List of acronyms.

Acronym	Definition
AMIP	Atmospheric Model Intercomparison Project
BMRC	Bureau of Meteorology Research Center
DMSP	Defense Meteorological Satellite Program
ECMWF	European Centre for Medium-Range Weather Forecasts
EMC	Environmental Modeling Center
EOF	Empirical orthogonal function
EPS	Ensemble Prediction System
FSU	The Florida State University
hPa	Hectopascals
JMA	Japan Meteorological Agency
MIT	Massachusetts Institute of Technology
NCEP	National Centers for Environmental Prediction
NOGAPS	Navy Operational Global Atmospheric Prediction System
NRL	Naval Research Laboratory
NWP	Numerical weather prediction
Rmse	Root-mean-square error
RPN	Recherché en Prévision Numérique
SSM/I	Special Sensor Microwave Imager
SVD	Single value decomposition
TRMM	Tropical Rainfall Measuring Mission
TSDIS	TRMM Science Data and Information System
UKMET	U.K. Met office
UTC	Universal time coordinated

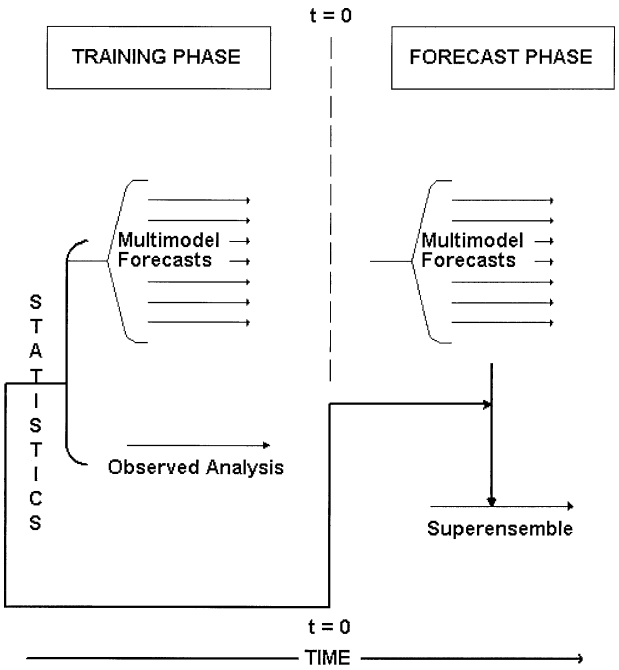

FIG. 1. A schematic diagram that illustrates the division of the time line; prior to day 0 is the training phase that includes 120 experiments from the multimodels whose daily results are regressed against an observed ''analysis'' benchmark. To the right of the zero line is the future forecasts phase that make use of the statistics from the training phase.

and also carries skills above those of the bias-removed ensemble mean representations. Here the multimodel forecasts are divided into two categories: (i) past medium-range forecasts and (ii) current real-time medium-range forecasts. The past covers roughly 100 recent past forecasts made by the multimodels. Given a benchmark analysis for this entire period, it is possible to obtain a model-weighted bias of forecasts of the multimodels at each geographical location. This is done following Krishnamurti et al. (2000b), by using multiple regressions of the model forecasts (applied on individual ensemble members) against a benchmark analysis. In this study, ECMWF analysis was selected as the benchmark. This superensemble, based on selective weights assigned to the member models, can be viewed in a probabilistic sense as providing information from a number of models (Stefanova and Krishnamurti 2002).

An anomaly correlation of 0.6 is generally regarded as an indication of a useful forecast. This threshold value comes from experience gained watching the forecast charts. A forecast with a skill greater than 0.6 generally implies that troughs and ridges at 500 hPa are beginning to be properly placed in that forecast.

Reviewing the anomaly correlations of U.S. operational forecasts, Kalnay et al. (1991) reported the summaries for the decade of the 1980s. During this decade, the anomaly correlations for 5-day forecasts increased from values such as 0.6 to approximately 0.75 over the Northern hemisphere and from approximately 0.4 to 0.625 over the Southern Hemisphere, as seen in Fig. 2a. Here, only the first 12 zonal wavenumbers were included in their analysis. If we take the first 12 wavenumbers (from forecasts and the analysis), the current skill during November 2000 for the best model and the proposed superensemble are approximately 0.85 and 0.90. This implies that major improvements for the 5-day forecasts are continually being realized. This progress of the forecast quality was attributed to factors such as improved computing power, improved models, data coverage, and assimilation methodologies.

Figure 2b shows the 500-hPa anomaly correlation skill over the European region for October–December 1981 carried out with the ECMWF model (Nieminen 1983). What is most encouraging here is the steady maintenance of very high skill in day-3 forecasts where the anomaly correlations are around 0.9 for almost the entire 3-month period. This speaks for the high quality of the modeling and data assimilation of the ECMWF system. Variability in skill from one day to the next increases with skills at day 5 for forecasts ranging from 0.4 to 0.9 and for day 7 for forecasts ranging from -0.4 to 0.7. This is nearly the state of the art of these forecasts at 500 hPa from the current best models.

Kalnay et al. (1998) also reported on the anomaly correlation at the 500-hPa level, covering a more extensive period of nearly 43 yr of forecasts. Here, the forecasts were based on the NCEP–NCAR reanalysis datasets, described by Kalnay et al. (1996), and the 1998 version of the NCEP forecast model runs at T62 reso-

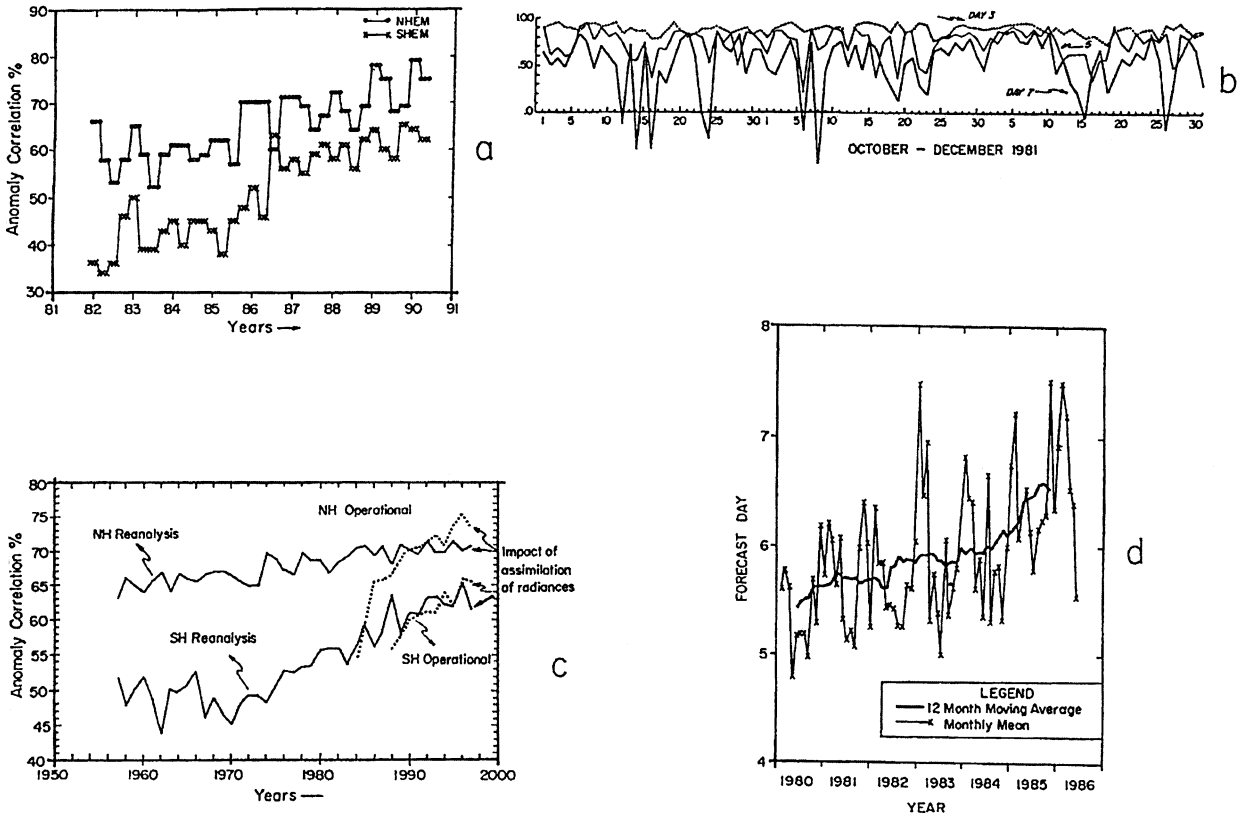

FIG. 2. Historical progress on anomaly correlations of the 500-hPa level. (a) The 5-day forecast 500-hPa height anomaly correlation (seasonally averaged), including zonal wavenumbers 0–12 for both the Northern and Southern Hemispheres (Kalnay et al. 1991). (b) Daily variation of the anomaly correlation of error of the 500-hPa height for forecast days 1, 3, 5, and 7 from 1 Oct to 31 Dec 1981 in Europe (Nieminer 1983). (c) Comparison of operational and reanalysis 5-day forecast anomaly correlations for the Northern and Southern Hemispheres (Kalnay et al. 1998). (d) ECMWF operational forecast skill since 1980 as represented by the forecast day on which the 500-hPa geopotential height anomaly over the Northern Hemisphere has dropped to 0.6. Heavy line is the 12-month average of the monthly mean values (Brown 1987).

lution for day-5 forecasts. Results for the Southern and Northern Hemispheres (20°–80° latitude) are presented in Fig. 2c, which shows a slow increase in skill over the Northern Hemisphere with skills reaching around 0.7 and around 0.6 for the Southern Hemisphere. The recent operational scores (also averaged over yearly periods) are shown by the dashed lines. The latter are based on higher-resolution (operational) models and reveal slightly higher skills.

TABLE 2. Outline of multimodels used in this study.

Model	Vertical levels	Horizontal resolution
ECMWF	31	T213
UKMET	30	0.8333° lon × 0.5555° lat
BMRC	29	T239
JMA	40	T213
FSU	14	T126
NCEP	42	T170
NRL	24	T159
RPN	28	0.9° lon × 0.9° lat

Brown (1987) has provided a review of the rapid progress in the improvements of global numerical weather prediction during the 1980s. The anomaly correlation at 500 hPa was one of the measures historically monitored by the weather services during that decade. Figure 2d from Brown's paper shows rapid increase in 500-hPa skill of the ECMWF forecasts. Here the length of forecast period (along the ordinate) for which the forecasts with anomaly correlation greater than 0.6 are reached during different years (along the abscissa) are shown. As the analysis and forecasting system has been improved, the useful length of the prediction has increased from 5.5 days in 1980 to 6.5 days in 1985, which is reflected in the 12-month average of the forecast length where the anomaly correlation of the 500-hPa geopotential height over the Northern Hemisphere drops to a value of 0.6. The monthly mean values indicated that the longer forecast skills are realized in the colder months.

Anomaly correlations of 500-hPa geopotential heights from ensemble forecasts at various operational

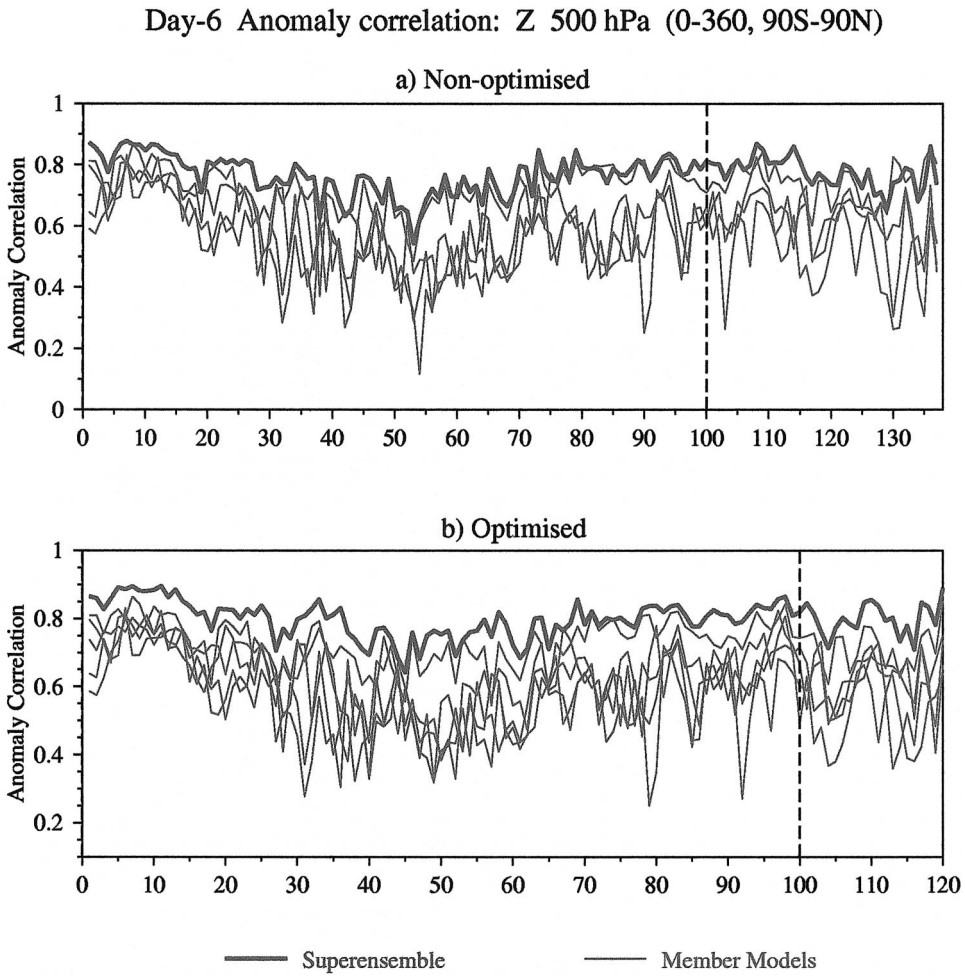

FIG. 3. Comparison of two graphs of anomaly correlation skill: (a) nonoptimized training and (b) optimized training. The ordinate denotes anomaly correlation and the abscissa denotes forecast days. The dashed vertical line separates the training and the forecast phases of the superensemble.

centers also have shown major improvements in skill in recent years. The ECMWF forecasts now show that 3-day skills can be close to 0.9. This implies that on day 3 these forecasts are placing the troughs, ridges, and contours right on top of the observed fields. This is an extraordinary accomplishment when we compare results obtained in the 1970s and 1980s from single models. Recent skills scores for the NCEP EPS and several other models also show similar major improvements in recent years. It is interesting to note that although there are major differences in the horizontal resolution of these models, somewhat comparable anomaly correlations are being achieved by these groups simply from overall model improvements.

2. Superensemble methodology

A main tool of this study is a multimodel superensemble that was recently developed at The Florida State University (Krishnamurti et al. 1999, 2000a,b, 2001). The superensemble is developed by using a number of forecasts from a variety of weather and climate models. Along with a benchmark observed (analysis) field, past forecasts are used to derive statistics on the past behavior of these models. These statistics, combined with multimodel forecasts, enable us to construct a superensemble forecast.

Given a set of past multimodel forecasts, we used a multiple regression technique (for the multimodels), in which the model forecasts were regressed against an observed (analysis) field. We then used least squares minimization of the difference between the anomalies of the model and the analysis fields in order to determine the weights. We carried out this minimization at all vertical levels, at all geographic locations (the grid points of the multimodels), and for all model variables. In all, some six million statistical coefficients describe the past behavior.

The motivation for this approach came from the construction of a multimodel superensemble from a low-

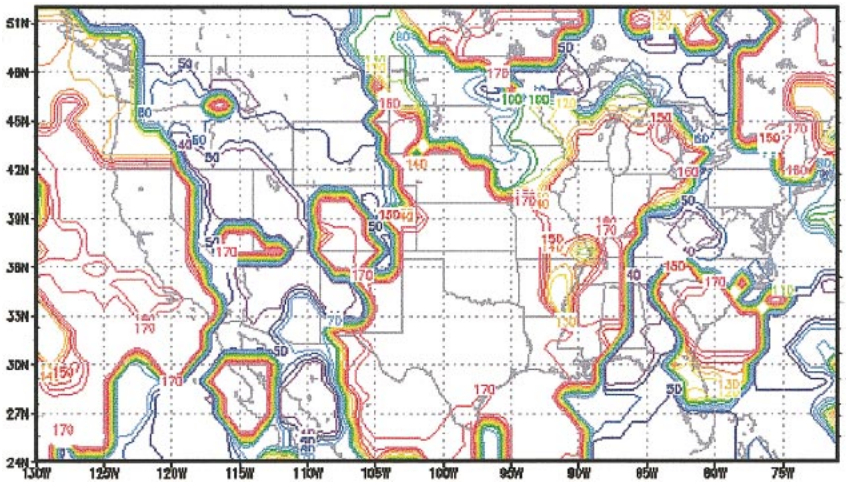

FIG. 4. Distribution of the optimal number of days for the training phase that provides the highest anomaly correlations at the 500-hPa surface over the North American Region for a day-6 forecast.

order spectral model (Lorenz 1963). In this low-order model, it was shown that it is possible to introduce various (proxy) versions of cumulus parameterization (or model physics) by simply altering forcing terms (Krishnamurti et al. 1999). Time integration of this multimodel system showed that the multiple regression coefficients of these multimodels (regressed against a nature run) carry a marked time invariance. This time invariance was a key element for the success of the proposed method.

We have used many models at diverse horizontal and vertical resolutions. Model output was interpolated to a common grid of the lowest-resolution multimodel. These global models include several different parameterizations of physical processes; effects of ocean, snow, and ice cover, as well as treatment of orography. The observed (or the analysis) fields are used only during the control period to determine the weights, and the verification of forecasts in the forecast phase of the superensemble. The training period for global weather comprises about 120 forecast experiments for each of the multimodels.

Approximately seven or eight multimodel forecasts were found to be minimally needed to produce very effective superensemble forecasts. The effectiveness of weather and seasonal climate forecasts has been assessed from measures of standard skill scores such as correlation against observed fields, root-mean-square errors, anomaly correlations, and the so-called Brier skill scores for climate forecasts (assessing skills above those of climatology). The horizontal resolution of the superensemble is around 125 km (a common denominator of the multimodel resolutions).

In some sense the construction of the proposed superensemble is a postprocessing of multimodel forecasts. This is still a viable forecast product that is being prepared experimentally on real time at The Florida State University using 11-member models. Thus it is a useful product for people to see and is currently available online on real-time basis (http://lexxy.met.fsu.edu/rtnwp).

3. Summary of past results

The following is a summary of computations based on our past publications (Krishnamurti et al. 1999, 2000a,b, 2001).

It is consistently noted in our past studies that the superensemble forecasts generally have higher skill compared to all participating multimodels and the ensemble mean. The ensemble mean assigns a weight of $1/N$ to all the member models (N is the number of models) everywhere (and for all variables), including several poorer models. As a result, assigning the same weight of $1/N$ to some poorer models was noted to degrade the skill of the ensemble mean. It is possible to remove the bias of models individually (at all locations and for all variables) and to perform an ensemble mean of the bias-removed models. This too has somewhat lower skill compared to the superensemble, which carries selective weights distributed in space, multimodels, and variables. A poorer model does not reach the levels of the best models after its bias removal.

Training is a major component of this forecast initiative. We have compared training with the best quality past ''observed'' datasets versus training deliberately with poorer datasets. This has shown that forecasts are improved when higher quality training datasets are deployed for the evaluation of the multimodel bias statistics. It was felt that the skill during ''forecast phase'' could be degraded if the training was executed with either poorer analyses or poorer forecasts. That was noted in our recent work on precipitation forecasts where we had shown that the use of poorer rainfall estimates

during the training period affected the superensemble forecasts during the forecast phase (Krishnamurti et al. 2001).

In medium-range real-time global weather forecasts, the largest skill improvement is seen for precipitation forecasts both regionally and globally. The overall skill of the superensemble is 40%–120% higher than the precipitation forecast skills of the best global models. The rms error and the equitable threat scores were the skill parameters used in that study. The training datasets for precipitation came from the daily TSDIS operational files of TRMM microwave radiometer-based rainfall estimates. These were augmented from the use of the U.S. Air Force polar-orbiting DMSP satellites that provided SSM/I data from a number of current satellites (*F11, F13, F14,* and *F15*) in order to extend the global coverage. An application of these precipitation forecasts included the forecast guidance for some recent flood episodes.

In real-time global weather forecasts the superensemble exhibited major improvements in skill for the divergent part of the wind and the temperature distributions. Tropical latitudes show major improvements with the superensemble for daily weather forecasts. For most variables, we have used the operational ECMWF analysis at 0.5° latitude–longitude for the training phase in these previous studies.

Real-time hurricane track and intensity forecasts are another major component of superensemble modeling. This approach of carrying out a training phase followed by real-time forecasts has shown improved forecasts for the tracks and intensity (up to 5 days) for the Atlantic hurricanes. Improvements in track forecasts were 25%–35% better than those of the participating member models. The intensity forecasts for hurricanes have been only marginally better than the best models. In some recent real-time tests during 1999, marked skill in the forecasts of difficult storms such as Floyd and Lennie was noted, where the performance of the superensemble was considerably better than that of the member models.

The area of seasonal climate simulations has only been addressed recently in the context of atmospheric climate models where the sea surface temperatures and sea ice were prescribed, such as the AMIP datasets. In this context, given a training period of some 8 yr and a training database from the ECMWF the results exhibited improved skill compared to the member models and the ensemble mean, which were based on seasonal and multiseasonal forecasts of monthly mean precipitation, temperatures, winds, and sea level pressure distributions. Further extension of this work is currently being pursued in the area of improved multimodel seasonal forecasts using coupled climate models.

4. Computational methodology

A main tool for this study is a multimodel superensemble that was recently developed at The Florida State University (Krishnamurti et al. 1999, 2000a,b, 2001). The proposed superensemble is defined by

$$S = \overline{O} + \sum_{i-i}^{N} a_i(F_i - \overline{F}_i), \tag{1}$$

where S = superensemble prediction, $\overline{O}$ = time mean of ''observed'' state, a_i is the weight for model i with i being the model index, N is the number of models, $\overline{F}_i$ is the time mean of prediction by model i, and F_i is the prediction by model i. Here the coefficient a_i is determined from the use of the least square minimization procedure. The weights a_i are computed at each grid point by minimizing the following function:

$$G = \sum_{t=0}^{t-\text{train}} (S_t - O_t)^2, \tag{2}$$

where O = observed state, t = time, and t − train = length of training period (100 days in the current case for NWP).

The variance of the final geopotential field was obtained using two different methods: (i) the use of height field Z to construct the superensemble and (ii) the use of the $\nabla^2 Z$ field to construct the superensemble. Here instead of constructing a superensemble of the geopotential height Z at the 500-hPa level, we have first constructed the superensemble of the $\nabla^2 Z$ field. The reason behind this was to extract some extra skill from the geopotential gradients and its Laplacian. The geopotential is thereafter recovered from a solution of the Poisson equation using the spectral transform:

$$(\nabla^2 Z)_n^m = -n(n+1)Z_n^m Y_n^m, \tag{3}$$

where m is the zonal wavenumber and n is the index of the associate Legendre function (which denotes the degree of its polynomial). As one proceeds to smaller and smaller scales, n becomes larger and we note then $\nabla^2 Z$ is directly proportional to n^2, whereas Z is inversely proportional to n^2; thus, this property would be reflected in the two-dimensional spectral distribution of $\nabla^2 Z$ and Z.

This method appears to improve the superensemble solution for the geopotential compared to a direct construction of the superensemble of Z. We are furthermore able to assess the skills where this method appears to contribute to the improvement of forecasts.

We have also carried out a comparison of the classical bias versus that of the proposed superensemble for the forecasts. A simple way of finding the bias for the NWP of a specific model is to take a daily string of NWP forecasts, obtain a monthly average of these, and compare those with the analysis (or observed) mean for that month. This procedure has been used by most weather services to assess whether the model has a cold, warm, moist, or dry bias, etc. (e.g., Heckley 1985; Sumi and Kanamitsu 1984; Kanamitsu 1985). This is the classical bias of a forecast. The proposed superensemble does not do quite the same thing. The multiple-regression-based

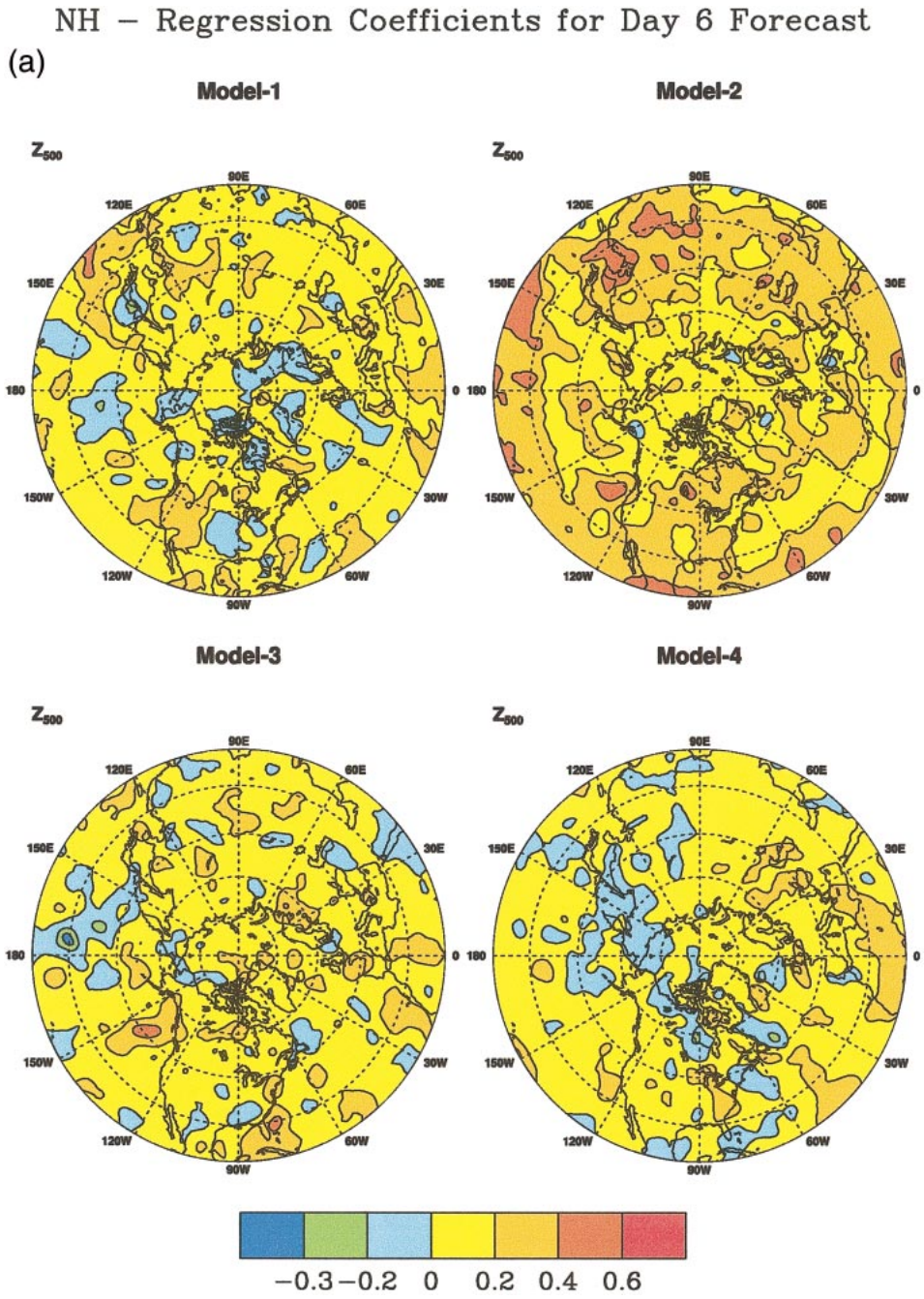

FIG. 5. Geographical distribution of statistical weights for different member models: (a) Northern Hemisphere and (b) Southern Hemisphere. Color scale of the fractional weights is shown at the bottom.

least squares minimization of errors is different from the simple bias correction in the following manner.

The simple classical bias is given by

$$\text{CB (classical bias)} = \{\overline{Z}_{\text{Fn}}(\lambda, \phi) - \overline{Z}_{\text{on}}(\lambda, \phi)\}/N. \quad (4)$$

Here the nth-day forecast bias for a total number of N days is considered; $\overline{Z}_{\text{Fn}}$ is the average forecasted geopotential height value at 500 hPa for a period of N days while $\overline{Z}_{\text{On}}$ is the averaged of the observed (analyzed) geopotential height for that period.

The superensemble-based bias, following Eq. (1), is given by

$$\text{SB (superensemble bias)} = \overline{Z}_{\text{sn}} - \overline{Z}_{\text{on}} = \sum_{t=1}^{t=N} \sum_{1}^{P\times Q} a_i(Z_{\text{Fi}} - \overline{Z}_{\text{FMi}}). \quad (5)$$

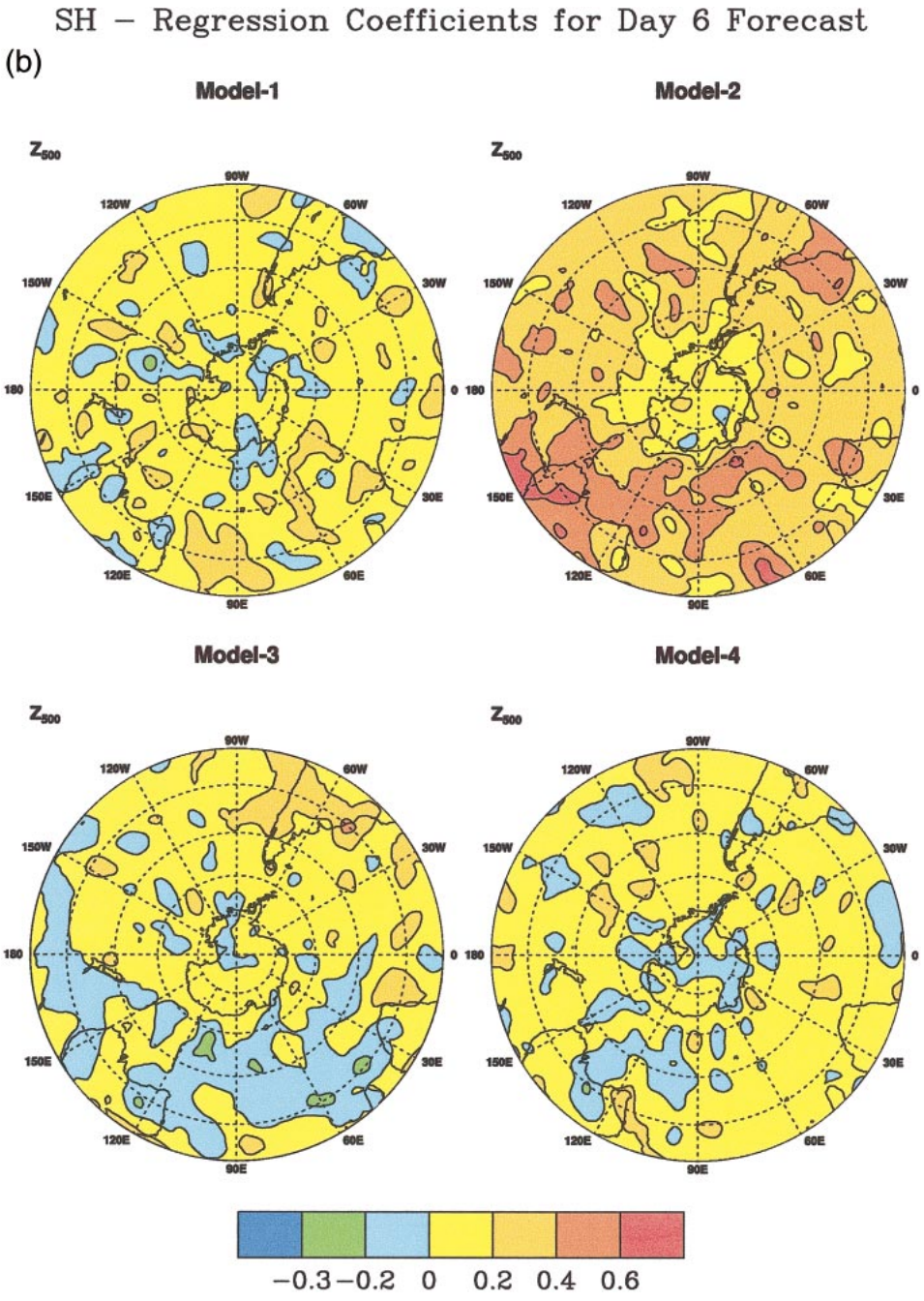

FIG. 5. (*Continued*)

Here $\overline{Z}_{\mathrm{Sn}}$ is the average geopotential height obtained from superensemble forecasts for the period of N days, $a_i s$ are the coefficients (weights) for individual member models, Z_{Fi} is the forecast from model i, and $\overline{Z}_{\mathrm{FMi}}$ is the simple forecast mean of geopotential height from the ith member model. The summation is done both on spatial ($P \times Q$ grid points, latitude–longitude) and temporal (days) scales for all member models. If the a_i were all set equal to $1/N$, then the classical bias and the superensemble-based bias are very close to each other. The weights take into account the local relative error characteristics of each model in the formulation of the superensemble.

The anomaly correlation coefficient is computed for individual model forecasts and the ensemble based on the method suggested by Brankovic et al. (1990). Anomaly correlation for forecast variables is defined as the correlation between the predicted and analyzed anom-

alies of the variables. Here anomalies are deviations from the mean climatological values. The following expression is used for computing the anomaly correlation of geopotential height at 500 hPa:

$$\mathrm{ACC} = \frac{\sum \{[(Z_F - Z_C) - \overline{(Z_F - Z_C)}][(Z_V - Z_C) - \overline{(Z_V - Z_C)}]\}}{\sqrt{\sum [(Z_F - Z_C) - \overline{(Z_F - Z_C)}]^2 \sum [(Z_V - Z_C) - \overline{(Z_V - Z_C)}]^2}}. \quad (6)$$

Here suffix F denotes forecast, suffix C denotes climatology, and suffix V stands for verifying analysis. The overbar is the area mean and Z is the geopotential height at 500 hPa.

5. Datasets

The datasets used in this study are identical to those used in a recent study on precipitation forecasts (Krishnamurti et al. 2001). The daily global analysis and forecasts from the following prediction centers were used: NCEP (Washington, D.C.), RPN (Canada), NOGAPS (NRL, Monterey, California), BMRC (Australia), UKMET (Reading, United Kingdom), and JMA (Japan). In addition to these, we used six daily forecasts from the FSU global spectral model that utilizes different initial analyses. The differences in the initial analysis are obtained through the physical

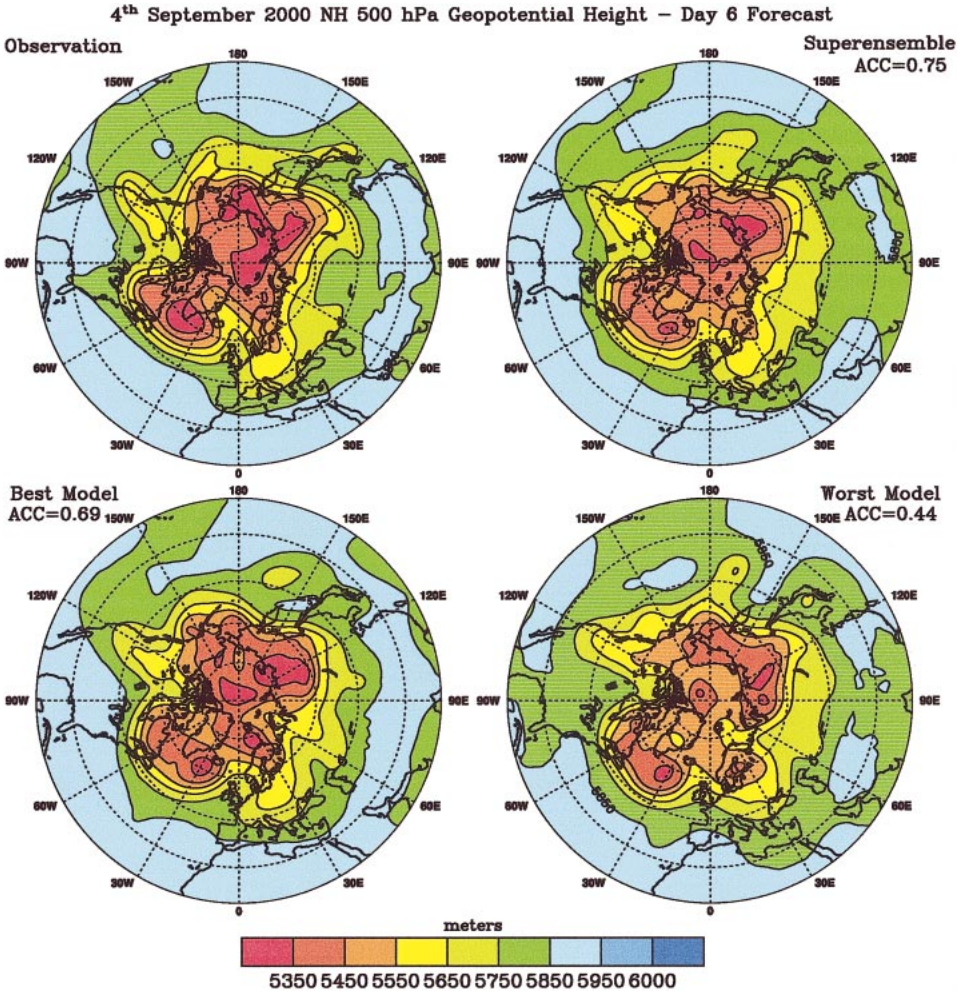

FIG. 6. A typical example of a forecast on day 6 over the Northern Hemisphere valid on 4 Sep 2000: (top left) analysis, (bottom left) the best model, (top right) superensemble, and (bottom right) the model with the lowest skill. Contour interval is 100 m.

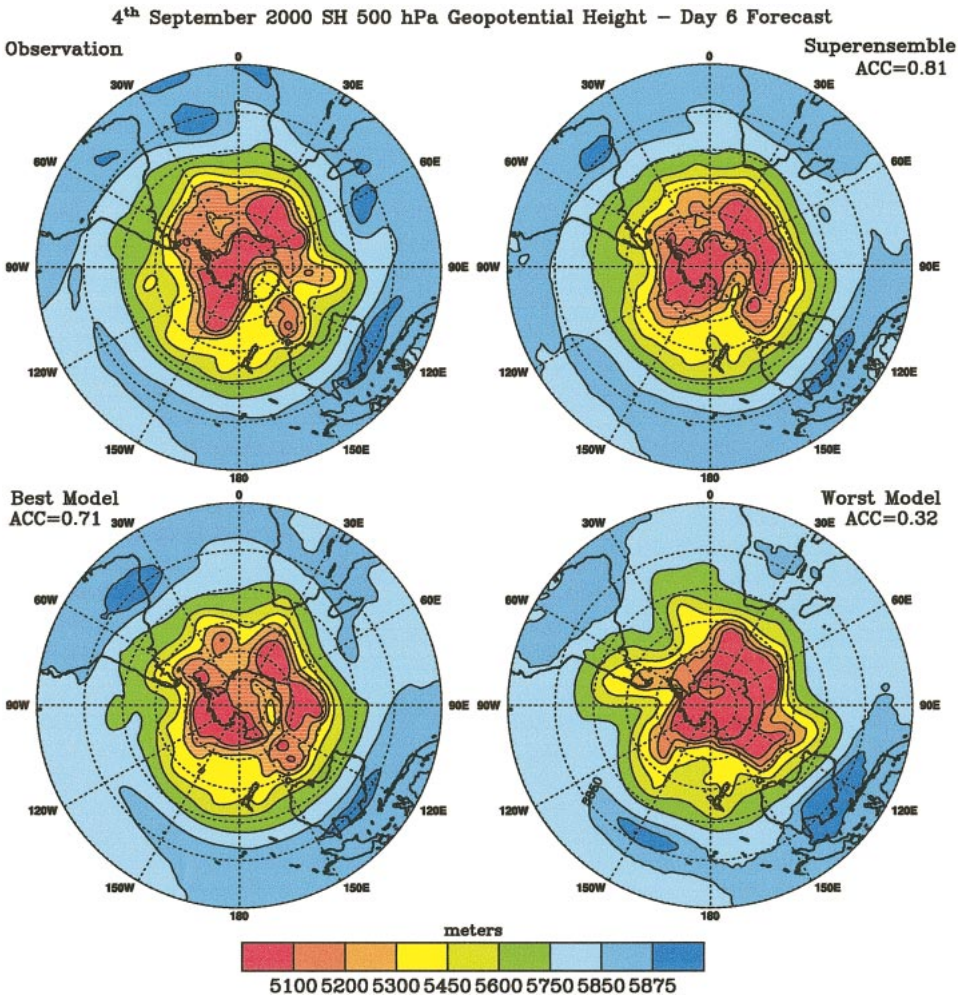

FIG. 7. Same as in Fig. 6 but for the Southern Hemisphere.

initialization of observed rain rates (Krishnamurti et al. 1991) using different rain-rate algorithms (Krishnamurti et al. 2001). All of these forecasts commenced at 1200 UTC and in most cases 6-day forecast datasets were available for extended periods. In addition to these, we had access to the initial assimilated datasets from the European Centre for Medium-Range Weather Forecasts for extended periods. It should be noted that in this study we have used ECMWF datasets in the following context. (i) For the training phase and forecast verification we have used the analyzed fields from ECMWF. (ii) We have not used ECMWF forecast datasets in the training phase. (iii) We had access to the daily anomaly correlation of geopotential height at 500 hPa of ECMWF forecasts from a Web site provided by NCEP. These are included in our tabulations of anomaly correlation skills but were not a part of the superensemble.

Formal computation of the anomaly correlation for most of these models was possible on a daily basis. Table 2 provides an outline of all the models used in this study. These models deploy different horizontal and vertical resolutions, topography, physical parameterizations, and modes of dynamic computations. The superensemble was constructed at an interpolated horizontal resolution of T126 (triangular truncation at 126 waves around the globe) at the 500-hPa surface for this study.

The Environmental Modeling Center of the U.S. National Weather Service maintains daily records of the 500-hPa anomaly correlation. Some of these datasets for the diverse models were extracted from their Web site and used in this study. In the computation of anomaly correlation, it should be noted that the anomaly correlations do depend on the choice of the source of climatology and also the verifying analysis. In this study, both ECMWF and the NCEP analyses were used separately for the training and forecast validation and in the evaluation of anomaly correlation skills. The results from the use of these different analyses for training were quite similar; hence, those based on the use of ECMWF only are shown here.

Since we carry a daily inventory of rms errors of

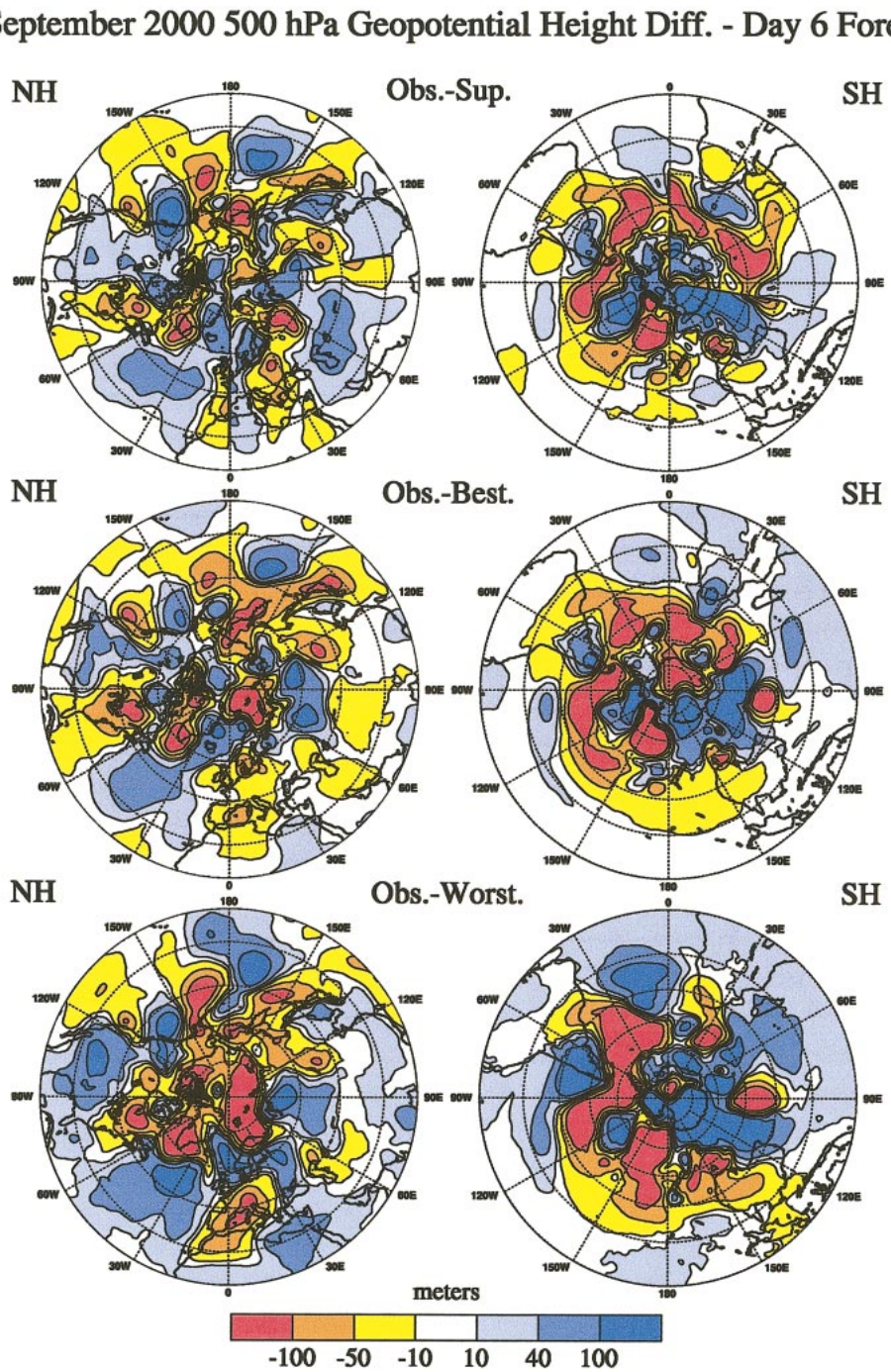

FIG. 8. Differences between forecast and analysis from Figs. 6 and 7: (top) superensemble forecasts, (middle) forecast from the best model, and (bottom) forecast from the model with the lowest skill. Units, m; color scale in m shown at the bottom.

forecasts during the training phase, it was possible to identify days where the forecasts skills were low. We arbitrarily removed those dates where the rms errors of the training superensemble were greater than 120 m for the 500-hPa geopotential heights. Using 100 days of ''higher skill'' training days, we noted that the skill of the superensemble was much improved. Figure 3 shows the results of computations of anomaly correlations with and without this optimization of the training phase, respectively. It is clear from this procedure that some further improvement of the skill of the superensemble is achievable.

6. Results of computations

In this section we present some of our results on the multimodel forecasts covering several months of forecasts. This is an ongoing effort, which is being carried out in real time with the datasets from the models described in section 5. Unless stated otherwise, most computations and results presented here are based on a training period covering 20 March to 15 July 2000 and the forecasts covered the period from 16 July to 17 September 2000. This period includes some missing days where data were not available.

a. Optimizing the number of training days

We had noted that the forecast skill degrades somewhat having either too few or too many training days for the construction of the superensemble. Thus it was felt that a certain optimal number of days could be determined in order to obtain the best statistics for the training phase, which in turn could provide the highest forecast skill. In order to determine this, we worked separately at each grid point and assessed the optimal number of days that provided the highest skill for the superensemble. Figure 4 shows the distribution of the optimal number of days for the best anomaly correlations of geopotential height at the 500-hPa level for day 6 of the forecasts over the North American region. The forecasts here were carried out for the entire month of August 2000; the optimal training days preceded that. It is interesting to note that a large-scale distribution of the optimal number of days is present in this analysis. This number around North America shows large-scale variation from oceans to mountains to the Great Plains. This behavior reflects the possible effects of bias errors of the member models over these different geographical regions. We have not addressed the issues of seasonality with respect to the optimal number of training days; this may need to be addressed for practical applications.

b. The distribution of weights

The essence of the proposed superensemble lies with the distribution of multiple regression weights for the member models. The training phase provides these weights. These weights exhibit a distribution of positive and negative fractional values. Figures 5a and 5b illustrate some of these distributions based on the training period covering the months April–July 2000. We have arbitrarily selected four of the member models, whose weights for the Northern and Southern Hemisphere are displayed in Figs. 5a and 5b, respectively. These weights display positive and negative fractional distributions. The scales of the centers of maxima and minima of these weights appear to be smaller over the Northern Hemisphere compared to those of the Southern Hemisphere. Some interesting features seen here are, for instance, that the weights for model 1 over North America show positive fractional weights over the western United States and negative fractional weights over the eastern United States. Over the South Pacific and Australia, the weights for model 2 are predominantly positive whereas they are negative for model 3. Collectively, positive or negative signs are contributed by the different models.

If one were to contrast these models with the ensemble mean, then all these distributions would bear the same constant value $1/N$ (where N denotes the number of ensemble models). This is the major difference between a superensemble and a member mean. The past history of performance of the member models makes the superensemble superior in skill, which is reflected by the geographical distribution of the statistical weights. This exercise of determining the weights is based on training of the $\nabla^2 Z$ fields and not the geopotential Z.

c. On the number of models

We noted that differences in the design on the models arise from the choice of physics, resolution, air–sea interaction, and the definition of orography. Thus the question of the optimal minimum number of models is important in the construction of the superensemble. If more than two of the models are included, the errors of the ensemble mean start to increase (see Krishnamurti et al. 2000b). This growth in error arises from assigning an equal weight of 1.0 to all models including the models with relatively lower performance levels. However, the error of the superensemble decreases as these additional models are included, since all models seem to have something to contribute over different regions. The error growth rate starts to decrease as these models are included, and beyond the inclusion of six top models, the error reduction almost stops. The reduction of error from six models is substantial and much higher in comparison to the ensemble mean. The reason for this behavior of the superensemble arises from its selective use of fractional and negative weights for the member models of the superensemble. Thus, we feel that improved 500-hPa anomaly correlations require minimally six models. These results are quite similar to what had already been noted from the datasets of 1998 (Krishnamurti et al. 2000b).

d. Predicted maps on day 6 of forecasts

Figures 6–8 illustrate a typical 6-day forecast. In Fig. 6 we show a forecast over the Northern Hemisphere for day 6 of a forecast from the superensemble (panel b), the best model (panel c), and a model with the lowest anomaly correlation skill (panel d). These are to be compared to the analysis valid on that date (i.e., 1200 UTC 4 September 2000). The anomaly correlations for these three respective forecast categories were 0.75, 0.69, and 0.44. We can see a slight improvement in the forecast features over the best model and a considerable im-

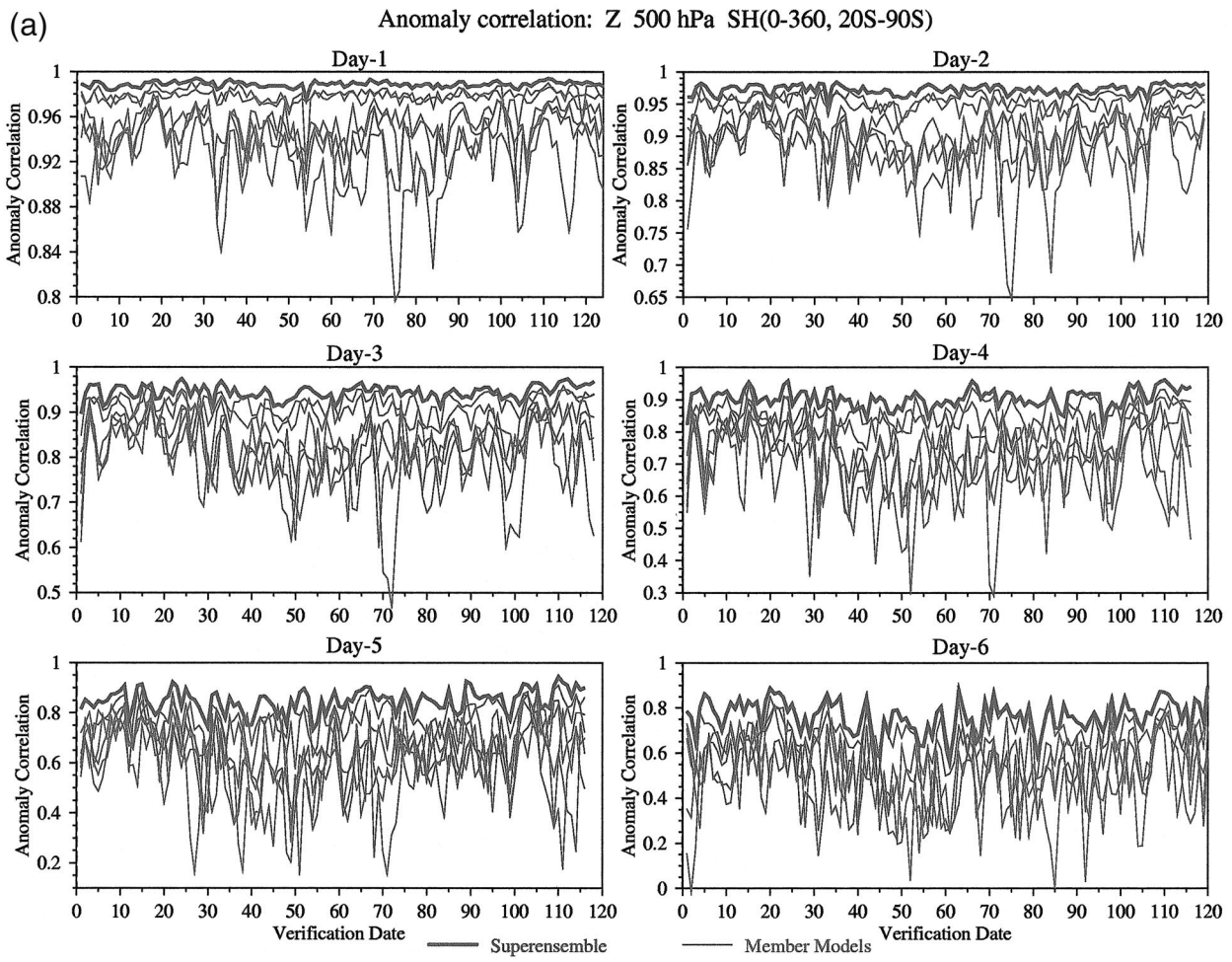

FIG. 9. Anomaly correlation as a function of forecast days: thin lines show results for member models and thick line to the top shows results for the superensemble. The different panels show forecasts for days 1–6.

provement with respect to the model with the lowest skill. A similar example of the day 6 forecast for the Southern Hemisphere is shown in Figs. 7a–d. The skill of the superensemble was generally quite high over the Southern Hemisphere. The respective anomaly correlations for the superensemble, the best, and the lowest skill models for this day-6 forecast were 0.81, 0.71, and 0.32.

Given that roughly 10%–15% improvement in anomaly correlation skill is possible from the superensemble over the best model, does it convey any useful synoptic information? Inspecting numerous 500-hPa forecasts from the best model and from the superensemble, we find that there is generally some more information content in the superensemble forecasts. The superensemble places the trough along 120°W more accurately, as compared to the best model, which moves it somewhat farther east in 6 days. The ridge over North America west of Lake Michigan shows a short-wave trough (riding the ridge) that is captured by the superensemble; the best model fails to capture that very short wave feature. In general, the model with the lowest skill has much lower heights over the entire United States. The trough along 120°E in the analysis is placed too far east, near 140°E, by the best model and near 125°E by the superensemble. In the highest latitudes, north of 50°N, the difference between the forecasts of the superensemble and the best model does not appear to be large.

The above features are more clearly seen from the differences between the forecasts and the analysis fields. Those fields for the Northern and Southern Hemispheres are shown in Figs. 8a–e. The preponderance of dark blue and dark brown coloring shows the large forecast errors for the model with the lowest skill. This coloring diminishes somewhat as we proceed to the best model and the fewest errors are seen for the superensemble. This was an example that was selected randomly. There are several other instances where the forecast skills on day 6 were strikingly larger for the superensemble compared to the best model where these differences are even sharper than what are shown in Fig. 8.

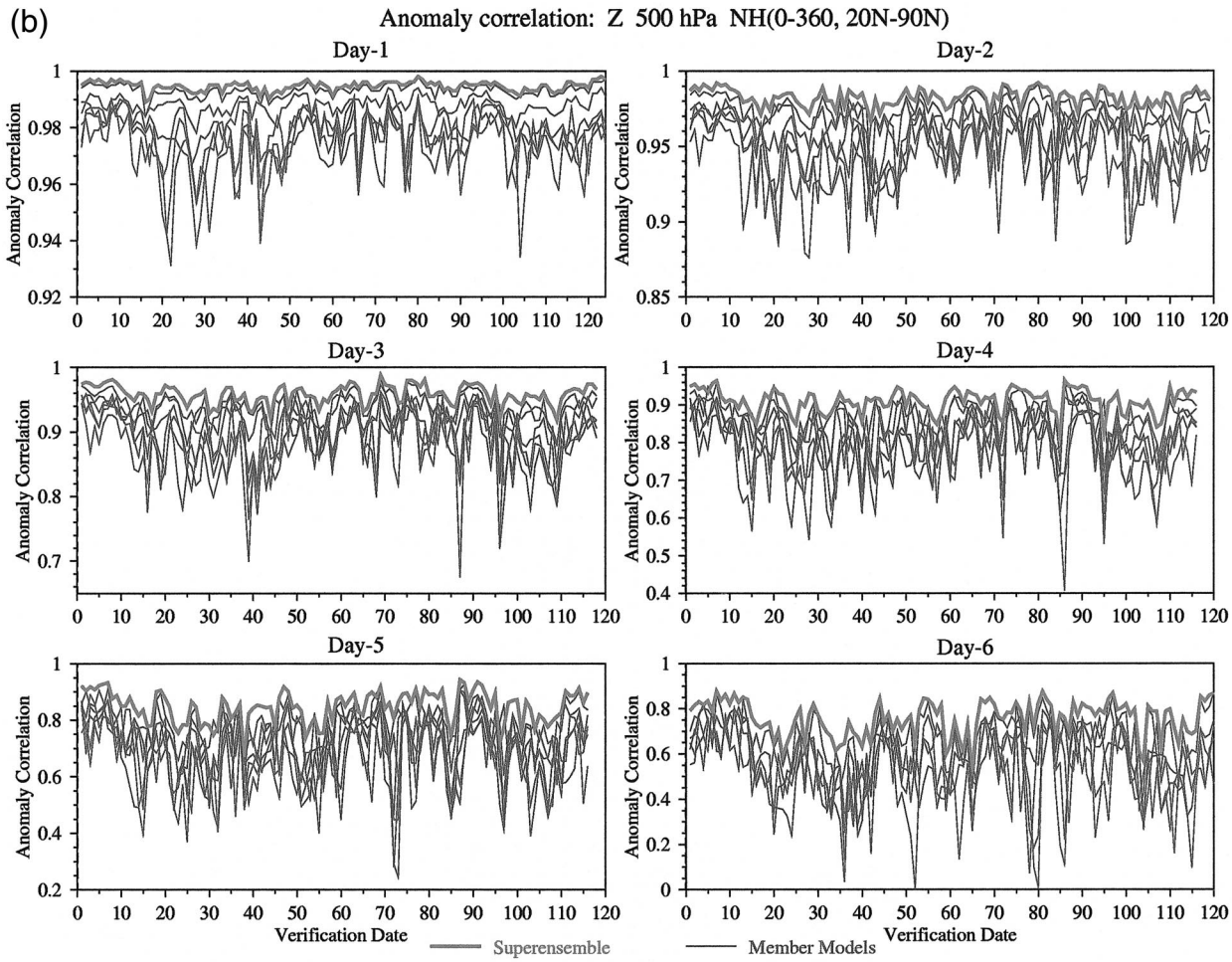

FIG. 9. (*Continued*)

e. Anomaly correlations at 500 hPa

Figures 9a–c illustrate the anomaly correlation of the geopotential height at 500 hPa for the member models (thin lines) and for the superensemble (thick line) for a recent 20-day period (this includes 100 days of training and 20 days of forecast). These results include those models that directly participated in providing daily datasets, apart from those models whose anomaly correlations were available on the NCEP Web site. Although the member models exhibit considerable variation in their anomaly correlation, varying from 0.1 to 0.8 for these forecasts, we note more consistent results for the superensemble. The $\nabla^2 Z$-based superensemble does show higher skill on almost all occasions for days 1–6 of forecasts. Figure 9a shows the results over the Southern Hemisphere, while Figs. 9b and 9c show the results for the Northern Hemisphere and the whole globe, respectively. Through day 3 of forecasts, the anomaly correlation skill for superensemble is consistently higher than 0.9. This shows the major improvement in the state of numerical weather prediction from some of the member models and from the superensemble. Overall the results for the superensemble are quite impressive. A noteworthy feature of these forecasts is the consistent higher skills over the Southern Hemisphere that is well above those of the last two decades shown earlier. Although we present a sample of results here, we have noted a major uniformity of these results in our continued computation of this algorithm on a real-time basis.

Table 3 provides a summary of these results for the Southern Hemisphere, Northern Hemisphere, and the global belt. Here the entries for the anomaly correlation skills covering a forecast period from 20 August to 17 September 2000 are presented. Results for the member models, the ensemble mean, and the superensemble are included here. Results for days 1–6 of forecasts are provided in these tables.

The wintertime skills of the anomaly correlation are generally higher than those for the summer season. The overall member model skills over the Southern Hemisphere are quite high. Over the Northern Hemisphere during this period, the best model's skill at days 1 and

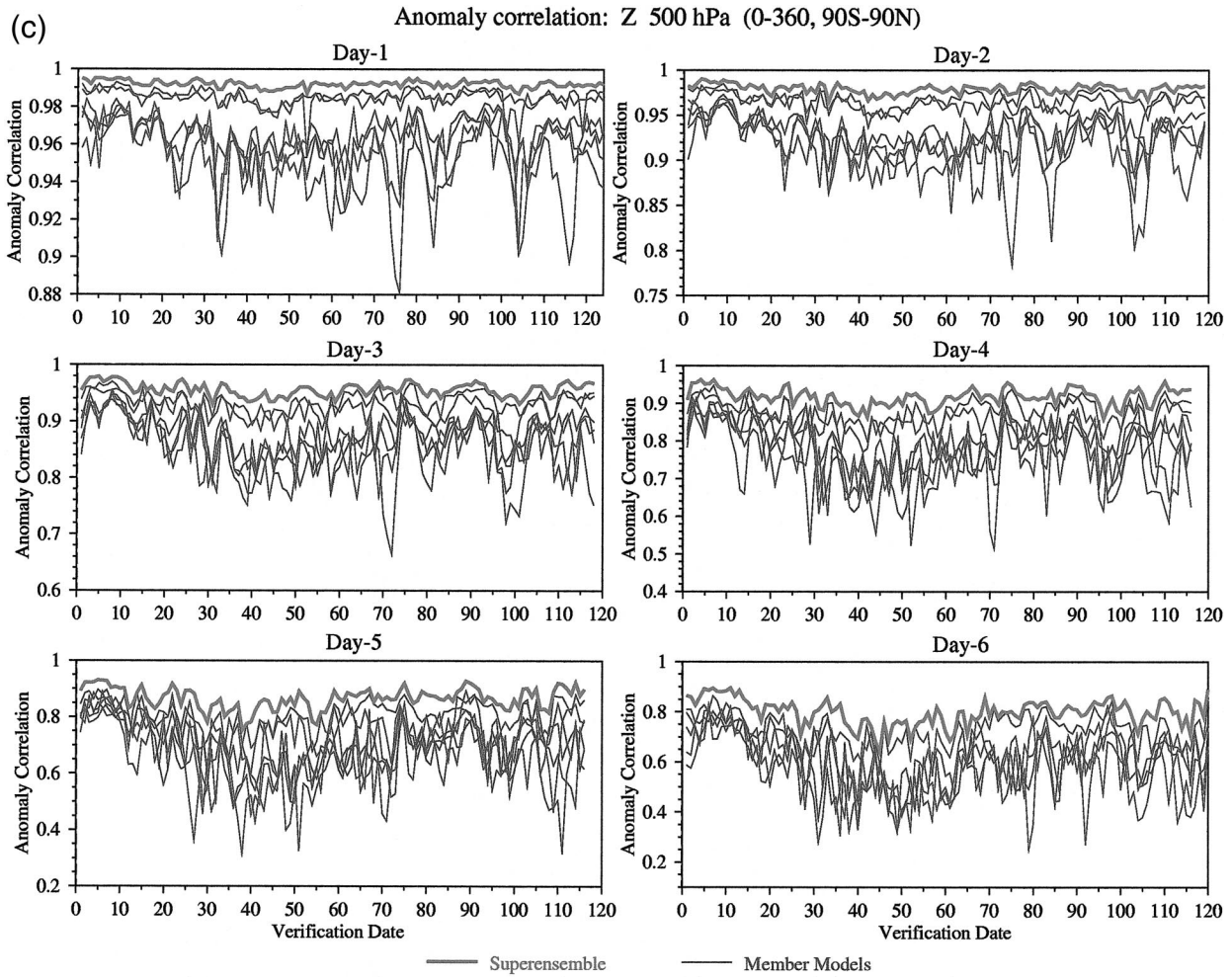

FIG. 9. (*Continued*)

6 were 0.992 and 0.653, respectively. The corresponding numbers for the superensemble were 0.995 and 0.748 for days 1 and 6. This shows roughly a 14% improvement on day 6 of forecasts. The corresponding figures for the Southern Hemisphere are an improvement for days 1 and 6 from 0.979 and 0.715 (for the best model) to 0.990 to 0.802 (for the superensemble, i.e., roughly a 13% improvement on day 6). Also shown in this table are the entrees for the ensemble mean, which lie roughly halfway in between the best model and the superensemble. Thus it appears that a substantial improvement in skill is possible from the use of the proposed superensemble. The global results presented in Table 3 basically confirm these same findings. The fall of skill from the best to the poorest model on day 6 of forecasts for the Northern Hemisphere can be seen ranging from 0.65 to 0.45 and for the Southern Hemisphere from 0.71 to 0.53. We have noted a consistently high skill for the superensemble around 0.75–0.8 for day 6 in our experimental runs.

Table 3d describes the global results for the case where two of the member models exhibiting the highest anomaly correlation skills are entirely excluded. The entries in this table are to be compared with the entries in Table 3c where these high skill member models are included. It appears as though the addition of the best models contributes to roughly 1%–2% improvement for the superensemble, while the overall improvement of the superensemble over the best (available) model is around 10%. This improvement in the superensemble is a result of the selective weighting of the available models during the training phase.

f. Reduction of systematic errors

In Fig. 10, we show the systematic errors (predicted minus observed mean) for day 6 of forecasts for the entire month of August 2000. Here the results for a member model with the least systematic error are compared with those of the superensemble for the Northern and Southern Hemisphere. In this illustration, the distribution of colors from blue to red displays the negative

TABLE 3. (a) The 500-hPa geopotential height anomaly correlation values from the superensemble (bold), ensemble mean (italic), and the multimodels averaged for the period 20 Aug–17 Sep 2000 for the Northern Hemisphere for forecasts from day 1 to day 6. (b) Same as in (a) except for the Southern Hemisphere. (c) Same as (a) except for total globe. (d) Same as in (c) except the two best models are excluded from the superensemble suite.

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
Superensemble	**0.995**	**0.981**	**0.956**	**0.905**	**0.843**	**0.748**
Ensemble mean	*0.989*	*0.970*	*0.936*	*0.880*	*0.807*	*0.700*
Model 1	0.992	0.973	0.935	0.864	0.776	0.653
Model 2	0.987	0.965	0.929	0.857	0.755	0.538
Model 3	0.975	0.951	0.892	0.812	0.721	0.526
Model 4	0.974	0.937	0.887	0.776	0.645	0.499
Model 5	0.969	0.937	0.873	0.762	0.636	0.453
Model 6	0.978	0.931	0.866	0.740	0.622	

(b)

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
Superensemble	**0.990**	**0.978**	**0.956**	**0.933**	**0.889**	**0.802**
Ensemble mean	*0.979*	*0.956*	*0.931*	*0.888*	*0.822*	*0.756*
Model 1	0.979	0.962	0.931	0893	0.830	0.715
Model 2	0.978	0.950	0.929	0.882	0.799	0.636
Model 3	0.957	0.925	0.888	0.826	0.723	0.596
Model 4	0.956	0.914	0.856	0.770	0.693	0.558
Model 5	0.945	0.903	0.837	0.733	0.615	0.535
Model 6	0.921	0.852	0.813	0.724	0.613	

(c)

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
Superensemble	**0.992**	**0.979**	**0.958**	**0.928**	**0.881**	**0.799**
Ensemble mean	*0.983*	*0.962*	*0.935*	*0.891*	*0.827*	*0.756*
Model 1	0.984	0.967	0.936	0.889	0.824	0.713
Model 2	0.981	0.957	0.932	0.880	0.796	0.623
Model 3	0.963	0.930	0.885	0.815	0.706	0.579
Model 4	0.962	0.925	0.871	0.786	0.697	0.578
Model 5	0.956	0.918	0.858	0.767	0.665	0.549
Model 6	0.941	0.889	0.846	0.739	0.632	

(d)

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
Superensemble	**0.978**	**0.957**	**0.927**	**0.891**	**0.834**	**0.740**
Ensemble mean	*0.970*	*0.948*	*0.915*	*0.866*	*0.798*	*0.684*
Model 1	0.963	0.930	0.885	0.815	0.706	0.579
Model 2	0.962	0.925	0.871	0.786	0.697	0.578
Model 3	0.956	0.918	0.858	0.767	0.665	0.549
Model 4	0.941	0.889	0.846	0.739	0.632	

to positive spread of systematic errors. A preponderance of dark blue color over the Southern Hemisphere reflects large positive systematic errors in excess of 40 m for the best model. It can be clearly seen that the systematic error of the model with the lowest skill is quite large (as seen by the coloring). Even the best model has rather large systematic errors. The superensemble is not free of these errors; however, the error is small compared to that of the member models. Over the Northern Hemisphere the preponderance of white, yellow, and light brown colors for the superensemble clearly reflect a large reduction of the systematic error. These same features have been monitored on a real-time basis for the last 2 yr. It is also clear from these computations that the superensemble does not simply remove the classical bias (i.e., forecast mean minus observed mean equal to zero). The proposed superensemble, although not systematic error free, is able to reduce the forecast error with obvious practical advantages but that kind of an artificial bias removal is not possible in a truly predictive sense. The bias correction is an after-the-fact correction and cannot be implemented for forecasts of any practical utility.

g. Percent improvement in rms errors from superensemble forecasts

Figure 11 shows the percent improvements from the superensemble forecasts compared to those of the ensemble mean over the best model. These are for 2- and 5-day forecasts of the global geopotential heights at 500 hPa. These improvements are related to the respective rms errors. It is clear that the improvements for the superensemble are quite large at day 2 of the forecasts, approaching about 40%, whereas the corresponding improvement for the ensemble mean is around 28%. At day 5, the improvements appear significant and larger; here the corresponding numbers are 52% and 46% for the superensemble and the best model, respectively. The mean rms errors at 48 and 120 h of forecasts for several of the member models and for the ensemble mean and superensemble are shown in Fig. 11b. These are results over the tropical belt 30°S–30°N. The period covered for the results shown in Figs. 11a and 11b includes 92 days starting from 1 June to 31 August 2000 (averaged). It is clear that the errors of the superensemble are smaller than the other representations shown here.

h. Improvements in the planetary and synoptic scales

When we examine the zonal harmonics of the geopotential at the 500-hPa level (Fig. 12), we note that the superensemble (based on the use of $\nabla^2 Z$) carries a larger portion of the variance compared to the individual models (assessed in terms of their respective anomaly correlation) for the first seven zonal wavenumbers. These are the scales where the zonal harmonics of the 500-hPa geopotential carries the largest proportion of the total variance. The predicted geopotential field for the worst model is quite flat by day 6 of forecasts compared to its variance field, shown in Fig. 12, and it appears that the long waves are somewhat misrepresented. The best model exhibits some improvement for the zonal percent variances while the superensemble, holding the highest anomaly correlation skill, exhibits a robust structure for the planetary and synoptic scale waves.

These same features can be viewed using two-dimensional variances in the triangular truncation space. In Figure 13 we show these variances as a function of

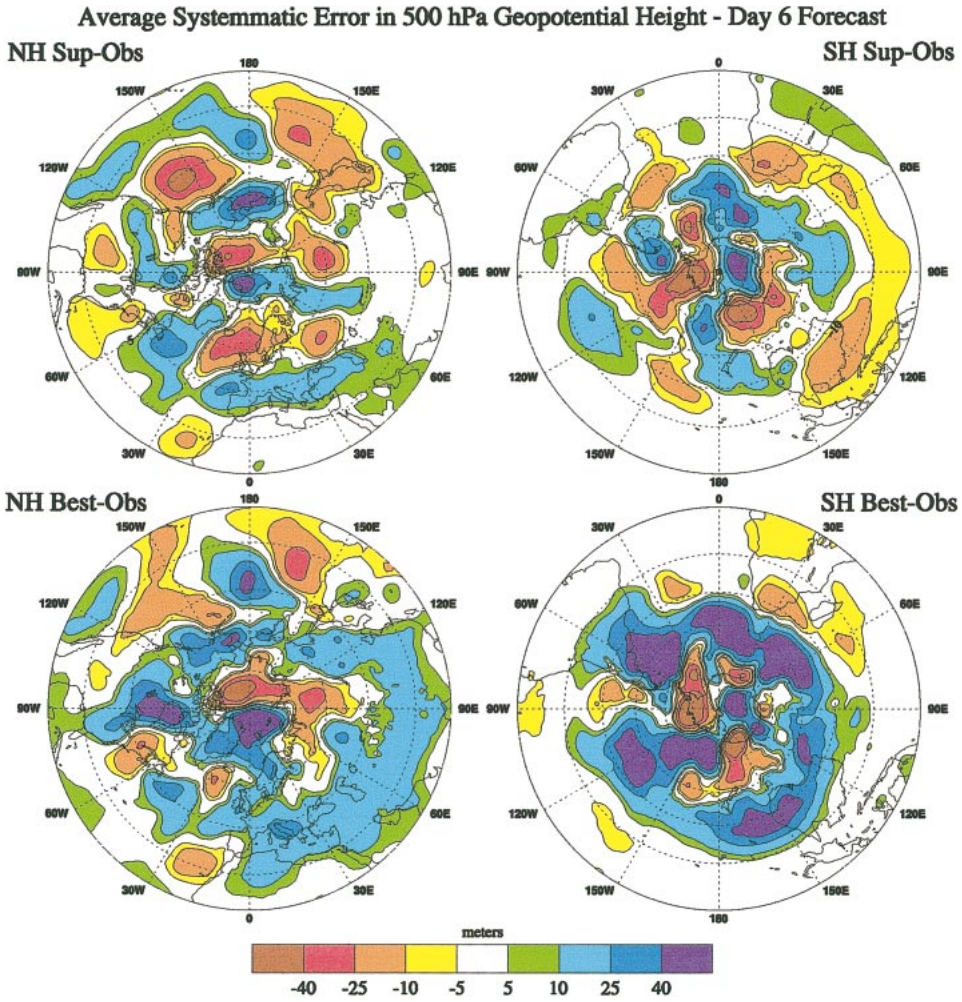

FIG. 10. Systematic errors: forecast minus analysis over 120 days over the 500-hPa level (top left) Northern Hemisphere superensemble, (bottom left) Northern Hemisphere best model, (top right) Southern Hemisphere superensemble, and (bottom left) Southern Hemisphere best model. Units, m; color scale in m shown at the bottom.

the east–west and north–south wavenumber *m* and *n*. Here again we note a very robust structure for the variances centered around zonal wavenumbers 2–5 and meridional wavenumbers 4–12. The variances for the best and the worst models are much lower in comparison. When we first started on this exercise we felt that the construction of the superensemble would enhance the structure of the smaller scales (wavenumbers greater than 10) since the $\nabla^2 Z$ field exhibits many smaller scales compared to the Z field. This exercise revealed that large geopotential gradients on the planetary and synoptic scales were much improved by the construction of the superensemble of the $\nabla^2 Z$ field.

7. Future outlook

The results on an anomaly correlation at the 500-hPa level are a part of an ongoing real-time numerical weather prediction exercise that is being pursued at The Florida State University. The total problem includes 11 different models where all the variables at 12 vertical levels are being subjected to the construction of the superensemble. For these weather models, some 10^7 statistical weights, which are being used, carry the past behavior. The reason for this large volume of statistics has to do with the model's performance. Some models appear to handle local water bodies better; others have greater skill over orographic regions while others seem to describe the oceanic convection and rainfall better. The systematic errors, in detail, vary from one region to another for these diverse models. Results for all of the variables, including daily precipitation (Krishnamurti et al. 2001), show a rather similar behavior; that is, the superensemble generally exhibits a much higher skill compared to the ensemble mean and the member models.

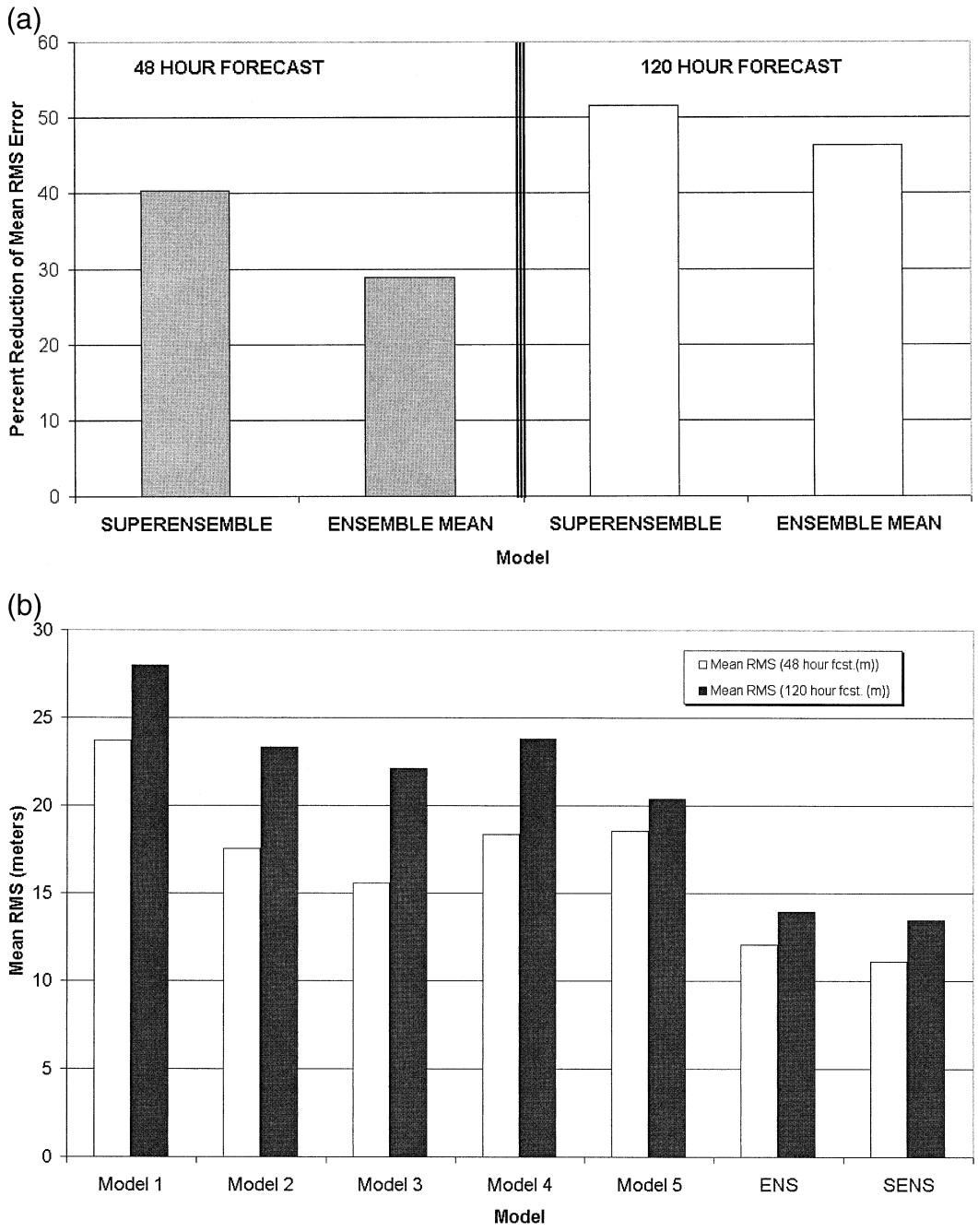

FIG. 11. (a) Percent reduction of mean rms errors at the 500-hPa surface over the best model by the superensemble and by the ensemble mean for Jul 2000 over the global tropical belt 30°S–30°N. (left) Results of 48-h forecast and (right) results at 120 h. (b) Mean rms errors of the respective member models, the ensemble mean, and the superensemble at hours 48 and 120 of forecasts over the tropical belt 30°S–30°N.

The emphasis on the 500-hPa geopotential is based on historical and practical reasons. Noting that it is now possible to construct a superensemble of the geopotential heights at 500 hPa with an accuracy of 0.95–0.80 between days 1–6 of forecasts implies that troughs and ridges are very nearly accurately placed close to their correct locations for the medium-range forecasts. Achieving these skills in a consistent manner appears to be an accomplishment of current NWP systems combined with the superensemble postprocessing method. This may be one of the reasons why operational forecasters should consider the implementation of this simple procedure of construction of superensemble forecasts. The datasets provided by the superensemble con-

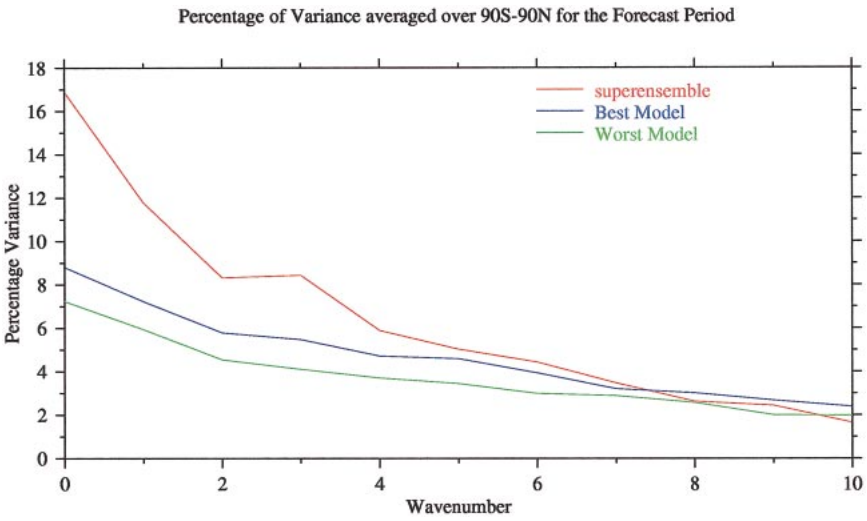

FIG. 12. Meridionally averaged percent variance of the geopotential heights as a function of zonal wavenumber.

tain the weighted averages that are carried out independently at each grid point of the domain of calculation for each day of forecast separately. The issue of dynamical consistency of this dataset has been raised following our first study (Krishnamurthi et al. 2000b), where we examined the quasi static and quasigeostrophic fields (over middle latitudes) of this data and noted that the balance is quite acceptable. It should be noted in this context that the construction of the superensemble is based on Eq. (1) where the regression coefficients are based on anomalies with respect to a time mean and not from a direct use of the full geopotential fields. It should also be noted that the individual models have indeed improved considerably over the last three decades. Further improvement, no doubt, will occur as the models improve their data assimilation, resolution, dynamical representations, and physical parameterizations. From what we have seen in our recent work it would seem that further improvements of the superensemble forecast would also follow.

An area of future work would be to explore the usefulness of singular value decomposition (SVD) method to address the removal of cases of possible ill-conditioned matrices in the current superensemble. The current multiple regression procedure in the training phase involves the solution of matrices at each location and uses the Gauss–Jordan method. We have essentially bypassed the difficulty at a few hundred grid points using the ensemble mean wherever ill conditioning was encountered. The SVD method, as well as several other methods that invoke EOFs, *Z* transforms, the Kalman filter, and cyclostationary EOFs can also be used to remove the degeneracy. An example of anomaly correlation improvement from the SVD method is shown in Table 4. These are the results for the same period as those presented earlier in Table 3. We notice that an improvement of over 5% from the use of SVD over the Southern and Northern Hemispheres and the global belt. On day 6 of forecasts, an increase of anomaly correlation from 0.799 to 0.845 can be seen over the global belt; while it increased from 0.748 to 0.801 and 0.802 and 0.849 over the Northern and Southern hemispheres, respectively, using the SVD method. We propose to incorporate these refinements in our future studies and hopefully they would also be useful for operational forecasts.

TABLE 4. The 500-hPa geopotential anomaly correlation averaged for the period from 20 Aug to 17 Sep 2000 for the global belt, the Northern Hemisphere, and the Southern Hemisphere for forecasts from day 1 through day 6. Here the results from two different methods, the conventional superensemble (SENS) and the SVD-based superensemble (SE_SVD), are presented.

	Global		NH		SH	
Day	SENS	SE_SVD	SENS	SE_SVD	SENS	SE_SVD
1	0.992	0.994	0.995	0.996	0.990	0.993
2	0.979	0.983	0.981	0.983	0.978	0.982
3	0.958	0.960	0.956	0.959	0.956	0.961
4	0.928	0.952	0.905	0.948	0.933	0.953
5	0.881	0.911	0.843	0.869	0.889	0.921
6	0.799	0.845	0.748	0.801	0.802	0.849

It is worth noting that the anomaly correlation skills over the Southern Hemisphere are beginning to exceed those of the Northern Hemisphere. At first we thought that this might have been peculiar for the period we had investigated, but it appears that as the member models are improving, the Southern Hemisphere forecast skills have indeed been going up in recent years.

Currently the multimodel datasets have only been available to us through day 6 of forecasts. It should be possible to obtain these datasets through day 10 of the

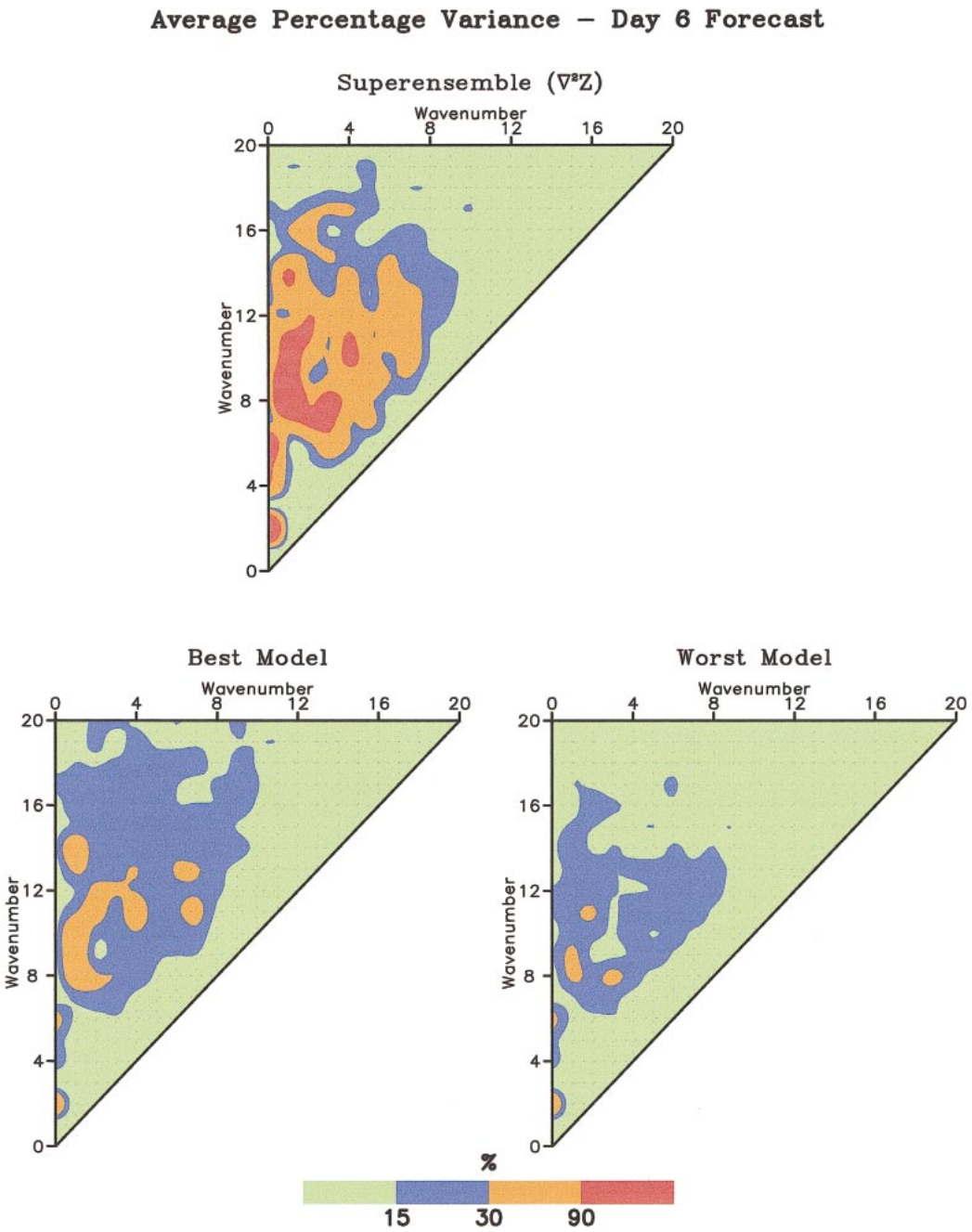

FIG. 13. Variances of the geopotential Z in the two-dimensional triangular truncation space. The ordinate denotes a meridional wavenumber n and the abscissa denotes the zonal wavenumber m. (top) Results for the superensemble. (bottom left) Variances for the best model and (bottom right) those for the model with the lowest skill.

forecasts. It may also be possible to receive the datasets as ensemble forecasts for the modeling groups. Given such datasets we expect further improvements in the forecasts of the 500-hPa anomaly correlations. Having such a forecast on the operational suite of products can provide useful guidance for the weather especially over the Tropics and midlatitudes.

Acknowledgments. The research reported here was funded by NASA Grants NAG5-9662 and NAG8-1537, NSF Grant ATM-0108741, and FSURF COE. The authors wish to express their thanks to Dr. Ricardo Correa Torres for computing the optimal training periods for this study. We wish to acknowledge the operational global modeling community for the datasets

they have provided. Special thanks go to Tony Hollingsworth for the ECMWF datasets used for model verification.

REFERENCES

Brankovic, C., T. N. Palmer, F. Molteni, S. Tibaldi, and U. Cubasch, 1990: Extended-range predictions with ECMWF models—Time-lagged ensemble forecasting. *Quart. J. Roy. Meteor. Soc.,* **116,** 867–912.

Brown, J. A., 1987: Operational numerical weather prediction. *Rev. Geophys.,* **25,** 312–322.

Heckley, W. A., 1985: Systematic errors of the ECMWF operational forecasting model in tropical regions. *Quart. J. Roy. Meteor. Soc.,* **111,** 709–738.

Kalnay, E., R. Petersen, M. Kanamitsu, and W. E. Baker, 1991: United States operational numerical weather prediction. *Rev. Geophys.,* **29,** 104–114.

——, and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

——, S. J. Lord, and R. D. McPherson, 1998: Maturity of operational numerical weather prediction: Medium range. *Bull. Amer. Meteor. Soc.,* **79,** 2753–2892.

Kanamitsu, M., 1985: A study of the predictability of the ECMWF Operational Forecast Model in the tropics. *J. Meteor. Soc. Japan,* **63,** 779–804.

Krishnamurti, T. N., J. Xue, H. S. Bedi, and D. Oosterhof, 1991: Physical initialization for numerical weather prediction over the Tropics. *Tellus,* **43A–B,** 51–81.

——, C. M. Kishtawal, T. LaRow, D. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved skills for weather and seasonal climate forecasts from multimodel superensemble. *Science,* **285,** 1548–1550.

——, ——, ——, ——, ——, ——, ——, and ——, 2000a: Improving tropical precipitation forecasts from a multianalysis superensemble. *J. Climate,* **13,** 4217–4227.

——, ——, D. W. Shin, and C. E. Williford, 2000b: Multimodel superensemble forecasts for weather and seasonal climate. *J. Climate,* **13,** 4196–4216.

——, and Coauthors, 2001: Real time multianalysis/multimodel superensemble forecasts of precipitation using TRMM and SSM/I products. *Mon. Wea. Rev.,* **129,** 2861–2883.

Lorenz, E. N., 1963: Deterministic non-periodic flow. *J. Atmos. Sci.,* **20,** 130–141.

Nieminen, R., 1983: Operational verification of ECMWF forecast fields and results for 1980–1981. ECMWF Tech. Rep. 36, 40 pp. [Available from European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, United Kingdom.]

Stefanova, L., and T. N. Krishnamurti, 2002: Interpretation of seasonal climate forecast using Brier skill score, FSU superensemble, and the AMIP-1 dataset. *J. Climate,* **15,** 537–544.

Sumi, A., and M. Kanamitsu, 1984: A study of systematic-errors in a numerical weather prediction model. 1. General aspects of the systematic errors and their relation with the transient eddies. *J. Meteor. Soc. Japan,* **62,** 234–251.