

Multimodel Ensemble Forecasts for Weather and Seasonal Climate

T. N. KRISHNAMURTI, C. M. KISHTAWAL, ZHAN ZHANG, TIMOTHY LAROW,
DAVID BACHIOCHI, AND ERIC WILLIFORD

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

SULOCHANA GADGIL AND SAJANI SURENDRAN

Center for Atmospheric and Oceanic Sciences, Indian Institute of Science, Bangalore, India

(Manuscript received 8 November 1999, in final form 19 July 2000)

ABSTRACT

In this paper the performance of a multimodel ensemble forecast analysis that shows superior forecast skills is illustrated and compared to all individual models used. The model comparisons include global weather, hurricane track and intensity forecasts, and seasonal climate simulations. The performance improvements are completely attributed to the collective information of all models used in the statistical algorithm.

The proposed concept is first illustrated for a low-order spectral model from which the multimodels and a "nature run" were constructed. Two hundred time units are divided into a training period (70 time units) and a forecast period (130 time units). The multimodel forecasts and the observed fields (the nature run) during the training period are subjected to a simple linear multiple regression to derive the statistical weights for the member models. The multimodel forecasts, generated for the next 130 forecast units, outperform all the individual models. This procedure was deployed for the multimodel forecasts of global weather, multiseasonal climate simulations, and hurricane track and intensity forecasts. For each type an improvement of the multimodel analysis is demonstrated and compared to the performance of the individual models. Seasonal and multiseasonal simulations demonstrate a major success of this approach for the atmospheric general circulation models where the sea surface temperatures and the sea ice are prescribed. In many instances, a major improvement in skill over the best models is noted.

1. Introduction

The notion of multimodel forecasting was evident in the studies of Lorenz (1963), where he examined the initial state uncertainties in a simple nonlinear system. Much progress has been made in multimodel forecasting for the conventional weather prediction problems, using the singular-vector-based perturbations, Molteni et al. (1996) and the use of breeding modes, Toth and Kalnay (1997). Several other formulations have appeared in the current literature, including the simpler Monte Carlo methods, Mullen and Baumhefner (1994). In seasonal climate forecasts, the multimodel forecasts are normally constructed by using initial perturbations from adjacent start dates, LaRow and Krishnamurti (1998).

Our primary interest is not in the evaluation of the ensemble mean and its probability characteristics, but in the evaluation of a statistical combination of various

models in a postprocessing procedure (super forecast/analysis). The skill of this technique appears to far exceed a conventional average of model fields. A number of model forecasts, solutions based on initial state perturbations, and the mean all cluster in one region of the system phase space. The errors of these models are sufficiently large generally such that the "observed" field often lies in another part of this phase space. A multimodel analysis tends to lie closer to the observed state compared to all of these other solutions. This single algorithm is used for all of the proposed problems, that is, global weather, hurricane track and intensity forecasts, and seasonal climate simulations.

This notion first emerged from the construction of a low-order spectral model that has been discussed extensively in the literature; see Lorenz (1963) and Palmer (1999). Section 2 of this paper discusses the low-order multimodel forecast. Here we show the time invariance of a statistic that makes it possible to project solutions, during a forecast phase, and exhibit fewer errors than most conventional models. This idea of resilience has been tested further, in this paper, using a variety of datasets on weather forecasts and seasonal to multiseasonal climate simulations. The seasonal simulations are

Corresponding author address: Prof. T. N. Krishnamurti, Department of Meteorology, The Florida State University, Tallahassee, FL 32306.

E-mail: tnk@io.met.fsu.edu

TABLE 1. Lorenz model parameters.

	Nature run	Multimodel range of values
R	24.74	22.24–27.74
B	24.74	8.35–18.35
θ	45°	42.5–47.5
σ	50.0	37.5–62.5

based on AGCMs (a list of acronyms is provided in appendix B). These are multimodel, 10-yr-long integrations based on AMIP datasets, Gates et al. (1999). The global numerical weather prediction applications of this algorithm (appendix A) are based on 1998 datasets from a large number of operational weather centers: ECMWF, UKMO, NCEP, RPN, JMA, NRL (NOGAPS), BMRC, and those developed from the FSU global spectral model forecasts. Also included in this paper are applications of the same algorithm for hurricane tracks and intensity forecasts during the 1998 season.

2. Motivation

The motivation for this paper came from the examination of the solution of a low-order spectral model, following Lorenz (1963), where we constructed models by assigning different values to the parameters of this problem. These parameters provide different values for the implied heat sources and sinks of this system. Assignment of different values for these parameters enabled us to construct an ensemble of models. One of these member models was arbitrarily defined as a nature run (i.e., a proxy for the real atmosphere).

The low-order system of Lorenz is described by the following equations:

$$\dot{X} = -\sigma X + \sigma Y + f \cos \theta \quad (1)$$

$$\dot{Y} = -XZ + rX - Y + f \sin \theta \quad (2)$$

$$\dot{Z} = -XY - bZ, \quad (3)$$

where X , Y , and Z are the spectral amplitudes that are a function of time only, the dot denotes a time derivative. For our purposes we can regard σ and r as denoting source terms, as a dissipation/diffusion coefficient, r as a heating term and b as the inverse of a scale height. In the original study of Lorenz, σ is the Prandtl number (a ratio of eddy diffusion coefficient to the thermal diffusion coefficient), r was the ratio of the Rayleigh number to a critical Rayleigh number, and b denoted the inverse of a length scale, that is, the size of thermal convective cells ($f \cos \theta$ and $f \sin \theta$ are forcing terms). The value of f was fixed at 2.5. Random perturbations were introduced for the values of r , b , θ , and σ within the range given in Table 1. The initial state was defined by $X = 0$, $Y = 10$, and $Z = 0$.

This is a simple numerical system that can be integrated forward in time using the standard leapfrog, time-differencing scheme with an occasional forward differencing (to dampen the computational mode). The model

ensemble can be thought of as set of models with different versions of physical parameterization (and/or diffusion) for each model. The different model versions were generated using the values shown in Table 1, which also includes the initial state definition that was altered slightly by random perturbations. The different initial states reflect different analyses from different forecasts systems.

At this point we define the superensemble that is central to this paper. A long-term integration through a time T (of the multimodels) is arbitrarily divided into two time periods, T1 and T2. Here T1 is regarded as a training period and T2 is regarded as the test (or forecast) period. During the training period, the multimodel variables were regressed against the observed (i.e., the nature run) variables X , Y , and Z . This is simply a least square minimization of the differences of the time series of the multimodels and of the nature run. This procedure provides weights from the regression for the individual multimodels.

There is an essential time invariance of the weights during the training and the test period; see Fig. 1. The abscissa in Fig. 1 denotes time. The training period was arbitrarily assigned as 70 time units (nondimensional time following Lorenz), and the test period is between time units 71 and 200. The time histories of weights shown in Fig. 1 were calculated using a cross-validation technique. To accomplish this, the weights for any given date are calculated from all other dates. This time invariance is an important feature, which if confirmed for weather and seasonal climate forecasts, could provide the possibility for a major improvement in the skill of our current prediction capability.

Figure 2 represents the principal variables X , Y , and Z for the Lorenz model runs. The thin, dashed lines represent the individual model runs, while the solid, heavy line shows the solution for the nature run. The heavy, dashed line starting at time 70 shows the solution for the superensemble. It is clear that the best solution, that is, the closest to that of the nature run, is seen for the superensemble, whereas the individual models exhibit large errors since their solutions carry large phase errors. The solution of the superensemble begins to exhibit a slow growth of error after around 190 time units. That type of behavior has also been noted in the application of this procedure for weather and seasonal climate forecasts. However, this growth of error begins to occur so late that one can make a major improvement in the skill of forecasts over periods of interest. That is a goal of this paper. The key element in this proposed technique is the time invariance (resilience) of the statistics displayed in Fig. 1, and that appears to be present if systems are not highly chaotic. The systems studied by Lorenz (1963) and Palmer (1999) were considerably more chaotic (for the choice of their parameters) compared to the examples we have presented here. We feel that our choice of parameters is more relevant for the atmospheric behavior as will be shown from the display

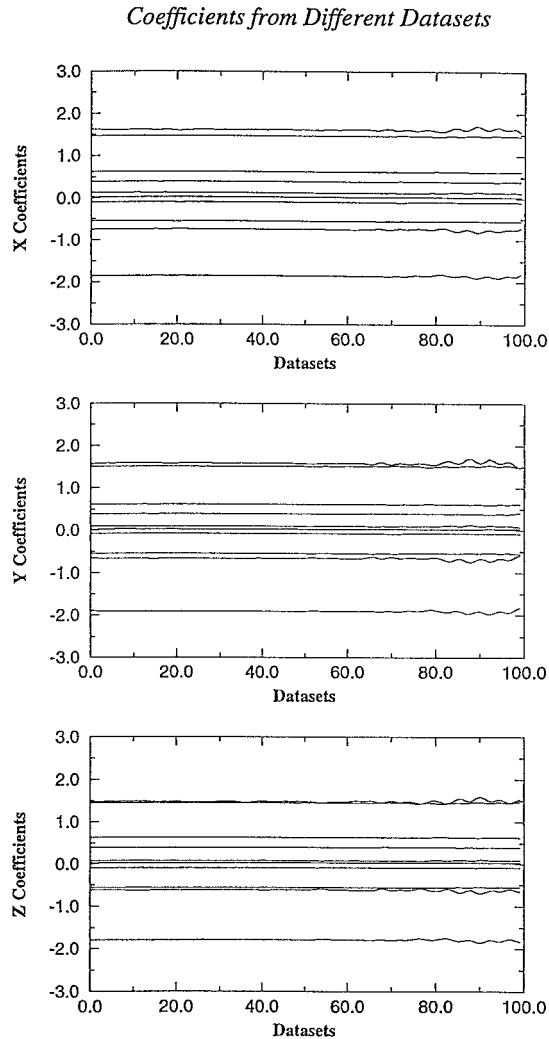


FIG. 1. Time history of the statistical weights for the Lorenz experiment based on 10 ensemble members. The computations shown here were made using the cross-validation approach.

of the time invariance of the statistical weights for the weather and seasonal climate forecasts.

The procedure above is defined by training and test phases using a simple multiple regression of the model anomaly forecast fields with respect to the observed fields. The general procedure is provided in appendix A and may be thought of as a forecast postanalysis procedure. During the test phase, the individual model forecasts together with the aforementioned statistics provide a superensemble forecast.

3. Datasets

The following datasets were used for the three components of the modeling studies reported here.

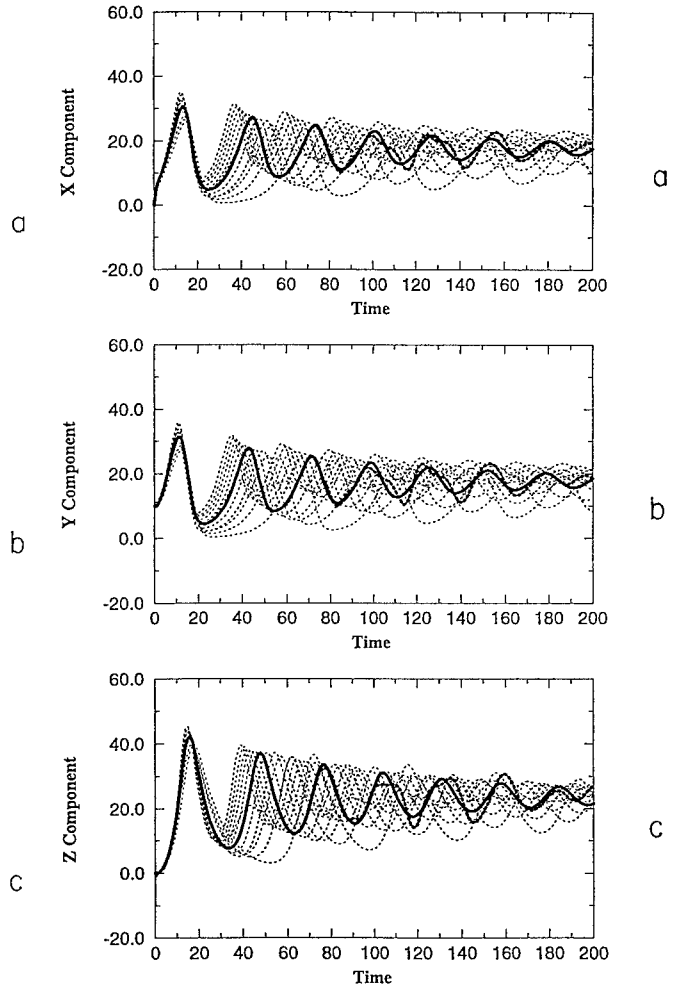


FIG. 2. The multimodel Lorenz solutions (dashed lines), the solution of the nature run (heavy dark line), and of the superensemble solution (for time > 100 units) shown by heavy dashed lines.

a. Seasonal climate

The AMIP datasets were provided to the research community by the Lawrence Livermore Laboratory, Livermore, California. This includes all basic model variables (such as winds, temperature, sea level pressure, geopotential heights, moisture, and precipitation) at intervals of one month (i.e., monthly averages) covering a 10-yr period, January 1979–December 1988. The multimodels of AMIP used in our study are listed in Table 2. The observed analysis fields were based on the ECMWF reanalysis and these were used in our studies. All of these observed analysis fields and the multimodel forecast fields were interpolated to a common resolution of 2.5° latitude by 2.5° longitude and 10 vertical levels for the monthly mean time intervals. In addition to these datasets, the monthly mean global rainfall totals were also an important data component. Gadgil and Sajani (1998) prepared the precipitation datasets.

TABLE 2. AMIP1 models for seasonal climate simulations.

Model	Vertical levels	Horizontal resolution	Radiation scheme				Land surface			
			Cumulus scheme	Shortwave	Longwave	PBL scheme	Soil model	Vegetation model	Orography	
UKMO (HAD-AM1)	19	2.5×3.75	Gregory and Rowntree (1990)	Ingram (1993)	Slingo and Wilder-spin (1986)	Stability dependent	Heat diffusion, single moisture reservoir	Warrilow et al. (1986); Shuttleworth (1988)	U.S. Navy 10-min terrain grid box averaged and 1-2-1 filtered poleward of 60°	
CSIRO (CSIRO9 Mark1)	9	T21	Arakawa (1972) Relaxed moist adjustment	Lacis and Hansen (1974); Sasamori et al. (1972)	Schwarzkopf and Fels (1991)	Stability dependent	Deardorff (1977)	None	Truncated to model resolution from Gates and Nelson (1975)	
ECMWF (Cycle36)	19	T42	Tiedtke (1989)	Morcrette (1991)	Morcrette (1991)	Stability dependent	Diffusion of heat/moisture, prescribed deep values	Blondin and Berger (1987)	Smoothed envelope orography	
GFDL (CDG1)	14	R30	Manabe et al. (1965)	Lacis and Hansen (1974)	Rodgers and Walshaw (1966); Stone and Manabe (1968)	Stability dependent	Bucket, no storage of heat	None	U.S. Navy 10-min terrain smoothed using Fourier filter transformed to R30	
NMC (MRF)	18	T40	Kuo (1965)	Lacis and Hansen (1974); Sasamori et al. (1972)	Schwarzkopf and Fels (1991)	Stability dependent	Heat diffusion and bucket	Pan (1990)	U.S. Navy 10-min terrain area averaged to T126 then spectrally truncated to T40	
BMRC (BMRC3.7)	17	R31	Modified Tiedtke (1989)	Lacis and Hansen (1974)	Schwarzkopf and Fels (1991)	Stability dependent	Prescribed deep temperature and bucket	None	1×1 U.S. Navy terrain is 3-point smoothed using Cressman truncated down to R31	
LMD (LMD6b)	11	3.6×5.6	Kuo (1965)	Fouquart and Bonnel (1980)	Morcrette (1991); Clough et al. (1980)	Stability dependent	Polcher (1994); Ducoudre et al. (1993)	Docoudre et al. (1993)	U.S. Navy 10-min terrain area averaged over grid box	
MPI (ECHAM4)	19	T42	Modified Tiedtke (1989)	Fouquart and Bonnel (1980)	Morcrette (1991)	Stability dependent	Warrilow et al. (1986); Dumenil and Todini (1992)	Blondin and Berger (1987)	U.S. Navy 10-min terrain Gaussian filtered with radius 50 km, truncated to T42	

TABLE 3. Resolution of NWP models (1998).

Models	ECMWF	UKMO	JMA	NCEP	NOGAPS	BMRC	RPN
Horizontal resolution	T213	~60 km	T106	T170	T159	T79	~35 km
Vertical resolution	31	30	21	42	18	19	28

For precipitation, a merged dataset supplied by NCEP (Schemm et al. 1992; Kalnay et al. 1996; henceforth referred to as NCEP merged data) comprising station data over land and precipitation over the oceans estimated from the Microwave Sounding Unit (Spencer 1993) has been used.

b. Global NWP

Two separate datasets were available for the superensemble forecasts over the globe.

- 1) Daily winds at 850 hPa (analysis and forecasts through day 3) for the three months June–July–August (JJA) 1998 (at 1200 UTC). These were available from the multimodeling centers: NCEP/MRF, ECMWF, JMA, RPN, BMRC, NOGAPS/NRL and the UKMO (Table 3).
- 2) Daily global 500-hPa geopotential heights for the months January–February–March (JFM) 1999 (at 1200 UTC), including the analysis and daily forecasts through day 5. These datasets were used to generate the superensemble forecasts and then to verify these forecasts using anomaly correlations. They were made available to us by the following centers: ECMWF, JMA, NCEP, RPN (Canada), BMRC, NOGAPS, and includes also the results from the global model of FSU.

c. Hurricanes

The datasets for the hurricane component of this paper comprise 12-hourly multimodel forecasts of storm positions and intensity for hours 0–144 as they became available. The multimodels include those listed in Table 4. In addition to those, we have also received the official “observations” called the BEST track and intensity, the Official (subjective) forecasts of the NHC and the superensemble forecasts of this study. Finally we also include hurricane intensity forecast datasets obtained from two in-house models of the NHC called SHIPS and SHIFOR.

4. Seasonal/multiseasonal precipitation simulations

We shall next present the superensemble forecast procedure based on the simulation datasets of AMIP. Phillips (1994, 1996) described the AMIP models. AMIP consisted of 31 different global models; the eight models selected are shown in Table 2. This choice was somewhat arbitrary. This table shows the resolution of the models and their salient physical parameterizations. These are all atmospheric general circulation models of the late-1980s vintage. This was one of the most complete datasets presently available for carrying out our proposed study on the superensemble. The 10-yr integration from these respective models had a start date around 1 January 1979 but was not all the same. The initial analyses (common to all models) were provided by ECMWF. Other common datasets shared by all these models included the distribution of monthly mean sea surface temperatures and sea ice. A common grid was used for the proposed superensemble studies. Data at all vertical levels and for all basic variables were interpolated to a common 2.5° latitude–longitude grid that was close to the horizontal resolution of most of these models. In the vertical, these datasets were next interpolated to common standard (so-called mandatory) vertical levels (i.e., 1000, 850, 700, 500, 300, and 200 hPa).

We have used two different time options to construct the superensemble from the 10-yr-long AMIP runs. Option 1 uses the last 8 yr of the dataset as a control (or the training period) and the first 2 yr as the test period. Here we make use of the monthly mean simulations, along with the monthly mean analysis fields (provided by ECMWF) to generate the “anomaly multiregression coefficients” (defined in appendix A). These are regression coefficients that may be separately generated for all grid points for all vertical levels and all basic variables of the multimodels. It should be noted that these weights vary in the three space dimensions. Option 2 simply uses the first 8 yr as control and the last 2 yr as the test period. We generated the respective weights for each model using the multiple regression technique during the training periods. We assume here that the long-term behavior, that is, a relationship of the mul-

TABLE 4. Resolution of hurricane forecast models.

Models	FSU1 (global)	FSU2 (global w/PI)	FSU3 (regional spectral)	NOGAPS	GFDL	UKMO
Horizontal resolution	T126	T126	~50 km	T159	1/6 deg	~60 km
Vertical resolution	14	14	14	18	19	30

timodels to the analysis fields, is defined in terms of the resilience of these weights. We next make use of this relationship to produce forecasts for an independent forecast period. For comparison purpose, results shown hereafter (excluding Figs. 5 and 6) use the following anomaly expressions (following the notation of appendix A) for the superensemble forecast (S'), the individual model forecasts (F'_i) and the ensemble mean (M'):

$$S' = S - \bar{O}, \quad F'_i = F_i - \bar{F}_i, \quad \text{and} \\ M' = \frac{1}{N} \sum_{i=1}^N (F_i - \bar{F}_i). \quad (4)$$

An overbar indicates a time mean. It is also important to understand that when full field plots are shown, the model climatology is added back.

Figure 3a illustrates the monthly mean tropical precipitation skill for option 1. The individual model precipitation rms errors lie between 3 and 5.5 day^{-1} . All AMIP rms errors displayed are calculated using anomalies from the respective model's monthly mean "climatology" that is based on the training period. The error of the ensemble mean, around 2.5 mm day^{-1} , is superior to those of the individual models. The error of the superensemble is approximately 2 mm day^{-1} and is superior to all other measures shown here. Options 1 and 2 have nearly identical results for the errors of the superensemble during each of their respective training and test periods. That is because 6 yr of data were common during their respective training periods.

Several questions naturally arise at this stage: How good is an error of 2 mm day^{-1} ? Is there a real skill for the climate simulations? An rms errors of monthly precipitation totals of the order of 1–2 mm day^{-1} does imply some useful skill, since the range of values of monthly totals lie between 0 and roughly 20 mm day^{-1} . The issue of its usefulness is further addressed in the geographical distribution of these simulations separately. We found that all simulations for different variables did not exhibit any large distinctions between options 1 and 2. Hence, we shall limit our future discussions to one or the other option.

Figure 3c illustrates the skill for the meridional wind over the global Tropics (30°S – 30°N) at 850 hPa for option 1. These are rms errors from continuous 10-yr simulations of the models. The individual model errors lie in the range of 1.5–2.5 m s^{-1} , whereas the errors of the superensemble (thick black line) lie around 1 m s^{-1} . Also shown in the illustrations are the errors of the ensemble mean. Those errors are of the order of 1.3–1.4 m s^{-1} ; again we note that the errors of the superensemble are distinctively lower than those of the ensemble mean. Also shown in this diagram are the errors of climatology based on observations (17-yr average from 1980 to 1996 based on ECMWF analysis). Similar results are seen in Fig. 3b for option 2. Overall, the proposed superensemble procedure has the least errors compared to all models, the ensemble mean and the

observed climatology. Very similar results are found for the meridional wind at 850 hPa over a monsoon domain (Fig. 3d) bounded by 30°S – 35°N , 50° – 120°E . The results show a clear separation of errors of the superensemble compared to those of the individual models. The monsoon region is of special interest because of the rainfall variability. The rms error of the rainfall over the monsoon region (Fig. 3e) lies roughly between 3 and 6.5 mm day^{-1} for the individual models and around 2 mm day^{-1} for the superensemble. The errors for the mean of the models are between 3 and 4 mm day^{-1} . This holds for both the training and test periods.

Next, we analyzed results for the monthly mean over various selected domains. Precipitation simulation rms errors of the monthly are briefly discussed here. The domains of analysis (Figs. 3f,g,h,i,j) include the Northern Hemisphere, Southern Hemisphere, the globe, Europe (between 30° – 55°N and 0° – 50°E) and North America (between 25° – 55°N and 70° – 125°W). Observed monthly rainfall estimates were obtained from the GPCP estimates [Arkin and Janowiak (1991)]. The precipitation simulation skill over the globe, Northern Hemisphere, North American domain, and the Tropics from the superensemble technique stand out and are superior to those of the individual model errors. The monthly errors during the entire 2 yr of the test period (January 1979–December 1980) are quite small, that is, of the order of 1–2 mm day^{-1} for the superensemble, whereas the errors of the ensemble average were around 2.5 mm day^{-1} . The range of individual model rms errors is between 2 and 5 mm day^{-1} over the various domains. Furthermore, we note that over several periods, the improvement in skill of the superensemble over the best model exceeds the difference in skill between the best and the worst individual model. Among all of the fields we analyzed, we noted that the best results were obtained for the superensemble statistics of the meridional wind errors over the Tropics, where that error of the superensemble was around 1 m s^{-1} throughout the 10-yr period (training as well as the test period). This is how far the multiple regression pushes the superensemble solutions toward the observed or the analysis state during the training phase and that accuracy is nearly retained during the test period. The tropical results are expected to be better due to the strong boundary forcing from the large oceans, whereas in the extratropics, there are typically large external errors. An ensemble mean does not have access to spatially weighted models. The procedure selects the best weighting at each individual point in space and for each separate variable, creating a collective, localized reduction of errors.

A comparison of the rms errors of the superensemble and the ensemble mean for the global domain are shown in Figs. 3k and 3m for precipitation and 850-hPa zonal wind, respectively. Here the results for option 2 are shown, but again, it is emphasized that the option does not change the representation of the results. In both cases, the rms errors of the superensemble are less than

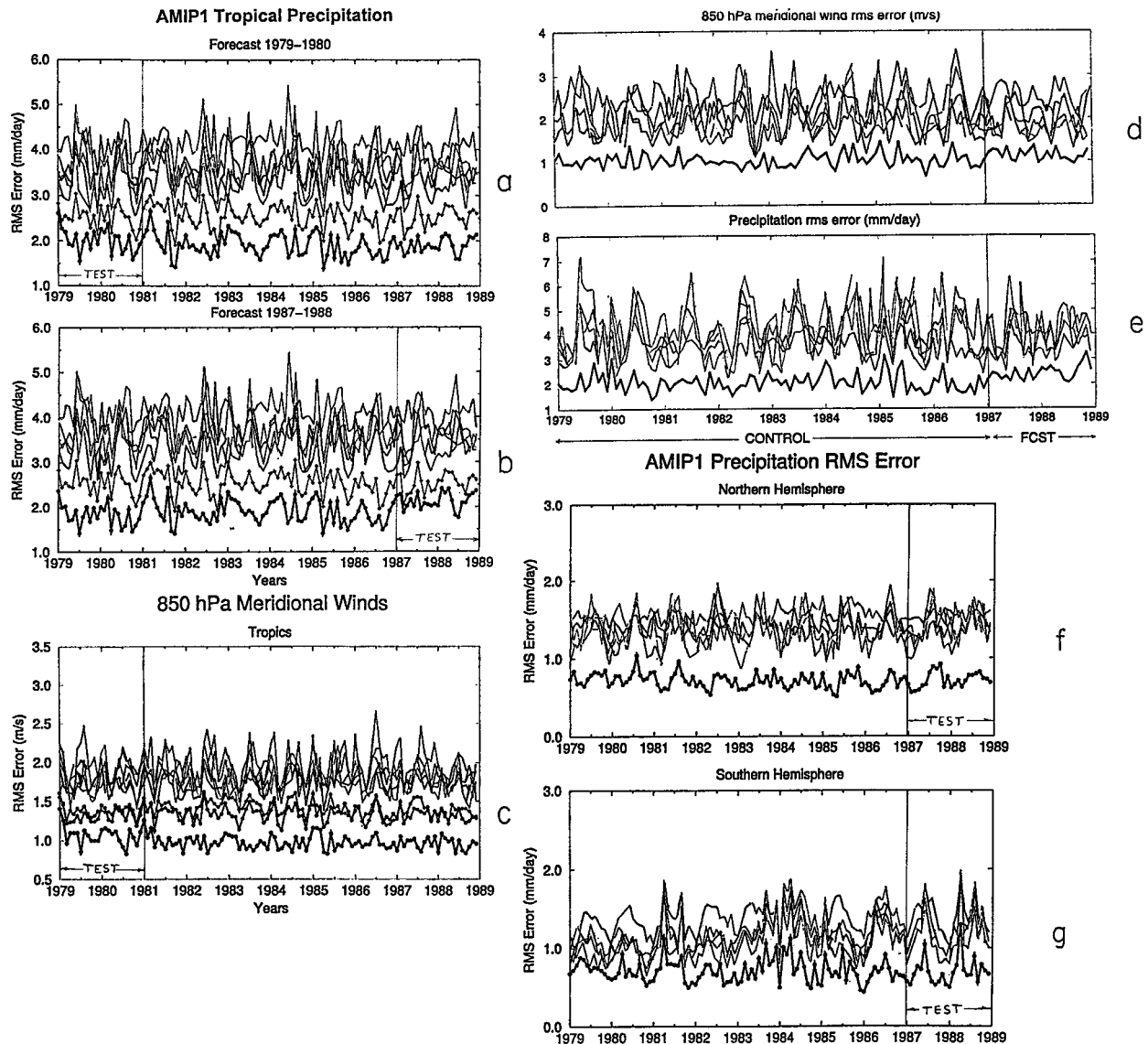


FIG. 3. (a), (b) The rms errors of precipitation for the multimodels and for the superensemble (lower curve). Results over the tropical belt (30°S – 30°N) are displayed. Lower dark line shows the superensemble, the next line above it is for the ensemble mean, and all other lines denote the results of various models. Units; mm day^{-1} . (a) Denotes option 1 and (b) denotes option 2. (c) The rms errors of the meridional wind (m s^{-1}) at 850 hPa over the entire Tropics (30°S – 30°N). Thin black lines: The rms error of the various models. Heavy black line: The rms error of the superensemble. Two lines above superensemble line: Climatology and ensemble mean. (d) The rms error of the meridional wind over a monsoon domain (m s^{-1}). Lower heavy line is for the superensemble, all other curves are for the individual models. (e) Monthly monsoon precipitation (mm day^{-1}) where the first 8 yr are used for control and the last 2 yr are used for forecast (Domain: 30°S – 35°N , 50° – 120°E). (f), (g), (h), (i), (j) The rms errors of precipitation (mm day^{-1}) for the multimodels for option 2. Results are shown for Northern Hemisphere, Southern Hemisphere, globe, Europe, and North America. (k), (l), (m), (n), (o), (p) A comparison of rms errors for the precipitation (mm day^{-1}), 850-hPa zonal winds (m s^{-1}), and 850-hPa meridional winds (m s^{-1}). Top curve in each box shows error of the ensemble mean, whereas the lower curve shows the error of the superensemble.

those of the ensemble mean. The reduction in error of the superensemble over those of the ensemble mean is of the order of 40%–100%. These results are very similar to the results obtained for the meridional wind (Figs. 3o and 3p), as well as in the Indian monsoon domain for all variables (Figs. 3l,n,p).

This study raises several questions. Are we seeing a useful skill here, in terms of seasonal and multiseasonal

climate simulations? Why is this skill so high? Given these very low rms errors for the superensemble, a mapping of the predicted fields from the superensemble leaves no doubt that they carry much more skill compared to the individual models. The collective, useful information content of multimodels is extracted in the construction of the proposed postprocessing superensemble technique comparing these to the analysis fields

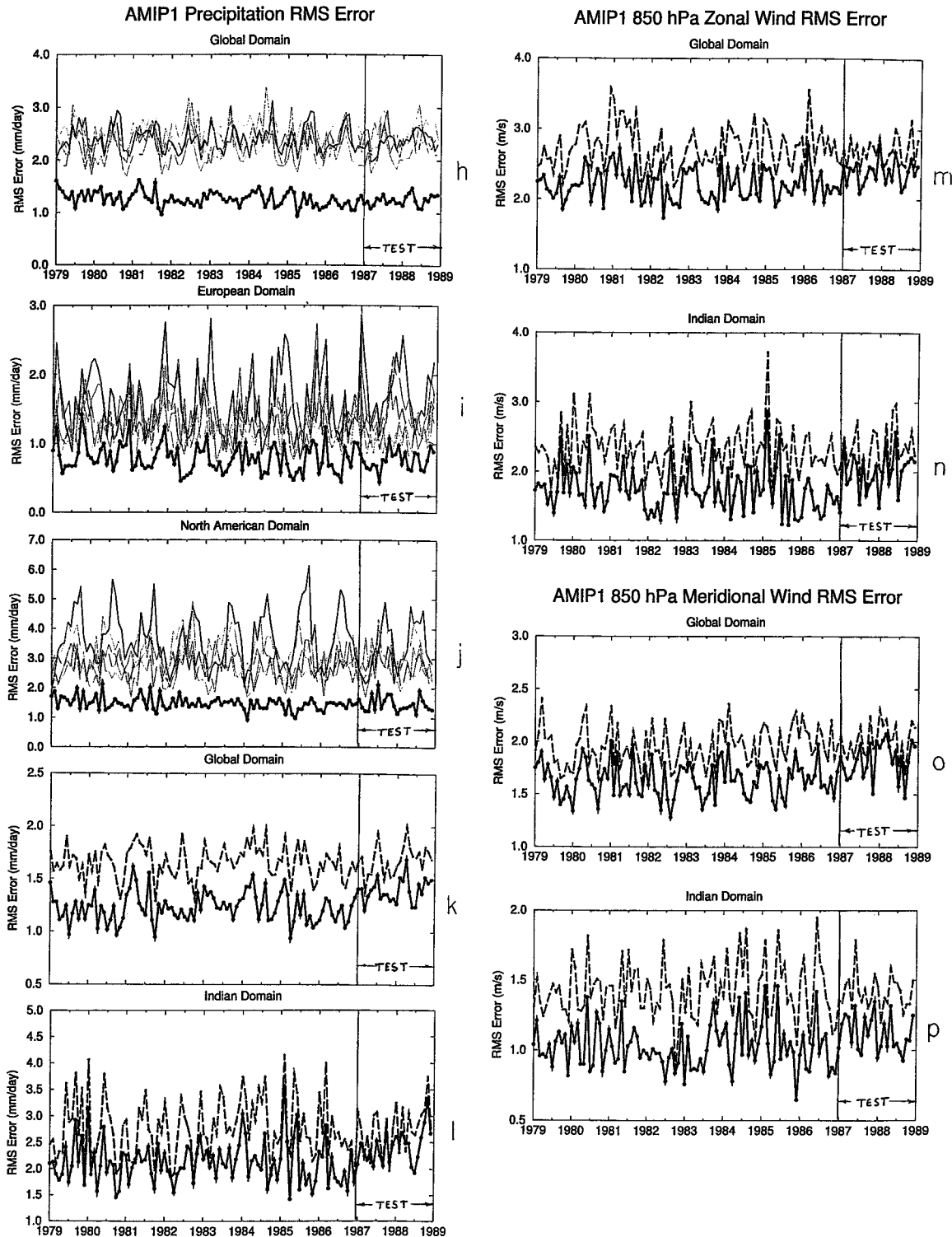


FIG. 3. (Continued)

during a training period. The removal of biases for each geographical region, at each vertical level and for each variable, appears to make this a very powerful scheme. The AMIP results from the individual models did not

seem very impressive. This procedure has been able to extract some very useful information from the results of the individual models from local regions where their skill was higher. As stated earlier, the success of this

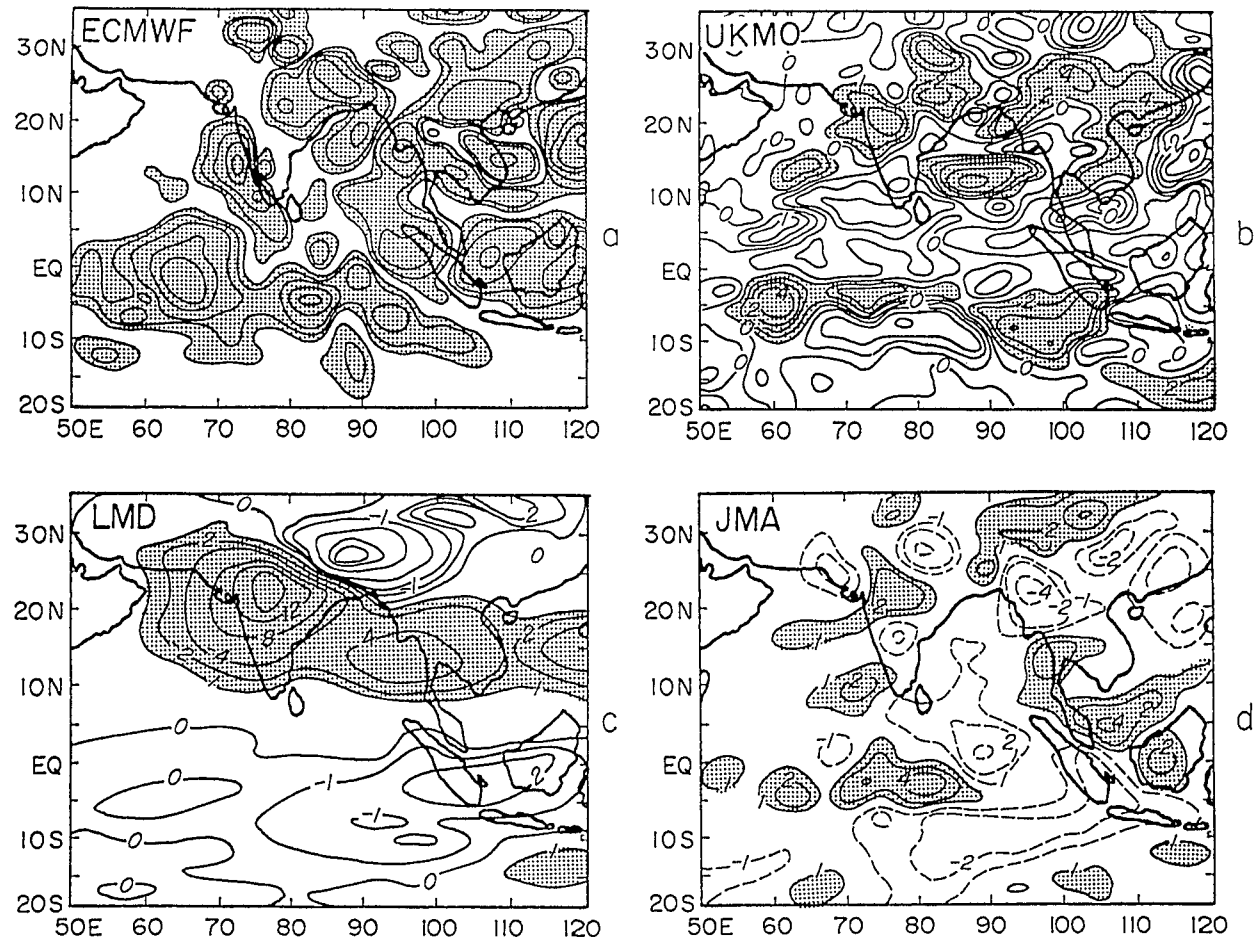


FIG. 4. (a), (b), (c), (d), (e), (f), (g) Difference in rainfall between JJA 1988 and JJA 1987, for multimodels, the superensemble, and the observed fields (mm day^{-1}).

scheme lies in the resilience of a relationship between the individual models and the analysis field. Further work is needed to understand the significance of this relationship and further improve the performance.

Monsoon rainfall differences during 1987 and 1988

A model intercomparison of monsoon simulations was organized by Dr. Tim Palmer of the ECMWF under a WMO/WGNE initiative, World Climate Research Program (1992). Eight modeling groups participated in these experiments. A start date of 1 June for each of the years (1987 and 1988) was used by all the modeling efforts. Seasonal simulations for these two years were carried out where the initial data for 1, 2 and 3 June were provided by ECMWF from their reanalysis. Additional datasets for the sea surface temperature and sea ice were also provided. The various models used different resolutions. One of the most difficult areas in seasonal climate forecasting is that of monsoon precipitation. Climate models have an inherent drift and model climatology (i.e., monthly and seasonal means) usually

exhibits large biases. In order to compare model performance against observed (i.e., analysis) fields such a bias needs to be removed. In the WMO/WGNE initiative, seasonal rainfall differences (JJA) for the years 1987 and 1988 exist for the various modeling groups. Such differences are expected to remove the bias to some extent. We have somewhat arbitrarily selected the following models, ECMWF, UKMO, LMD, JMA, and BMRC, which show these differences. Figures 4a–g show these model simulations on the differences 1988 minus 1987 seasonal rainfall covering the months JJA. These are based on a one-season-long simulation of the precipitation. Figure 4g shows the corresponding observed rainfall differences based on the GPCP. Those were largely the OLR-based precipitation algorithms. As was noted by Palmer and Anderson (1994), the current state of seasonal monsoon precipitation forecast for the different models is quite poor.

The multimodel superensemble test phase results shown in Fig. 4f were carried out from a start date of 1 January 1987, whereas the model simulations shown in Figs. 4a–e started around 1 June 1979 with observed

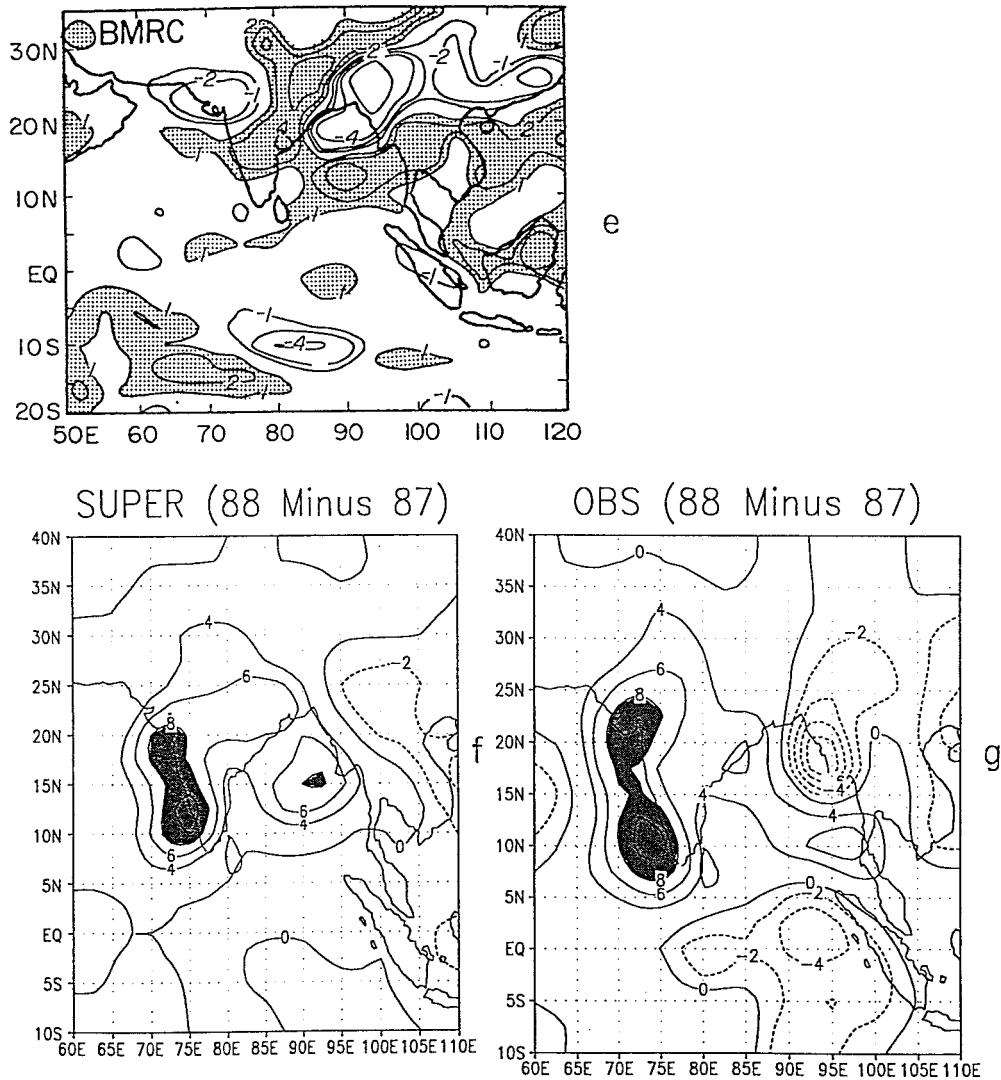


FIG. 4. (Continued)

boundary conditions. Figure 4g shows the observed differences in the seasonal precipitation for 1988 and the 1987 season over the domain of the Asian monsoon. We find that superensemble approach predicts the interannual variability of monsoon rainfall (1988 minus 1987) better than all of the multimodels used in comparison here. That was also apparent in the statistics of rms errors of precipitation that we presented earlier. Figure 5 shows the time sequence of the rms error for the meridional wind over the same domain. The arrow at the bottom right indicates the test period.

5. Global numerical weather prediction

Two datasets covering the periods JJA 1998 and JFM 1999 were used for the postprocessing superensemble forecasts from various global weather prediction models. Here we show the results for the global winds at

850 hPa during JJA 1998. Table 3 lists the models whose forecasts were available for this period. In total, we had seven models, each making 92 forecasts with a start time at 1200 UTC for each day. The 61 days during June and July were treated as the training period and the 31 3-day “forecasts” were prepared during the entire month of August 1998.

Given the differences in the physics and complexity of weather models, it is not surprising that one notes varying skills in the performance of these models. Figure 6 illustrates the 850-mb rms forecast errors on day 3, for various regions. These include the global belt, Tropics (30°S–30°N); the Asian monsoon region (between 30°S–40°N and 30°–150°E); USA (25°–55°N, 125°–70°W); Europe (30°–55°N, 0°–40°E); and the entire Northern and Southern Hemispheres. Here we include day-3 histograms of skills for the seven models as well as those for the multimodel ensemble (orange).

RMS Error of 850 hPa V-Wind Difference (50°E-120°E;30°S-35°N Average)

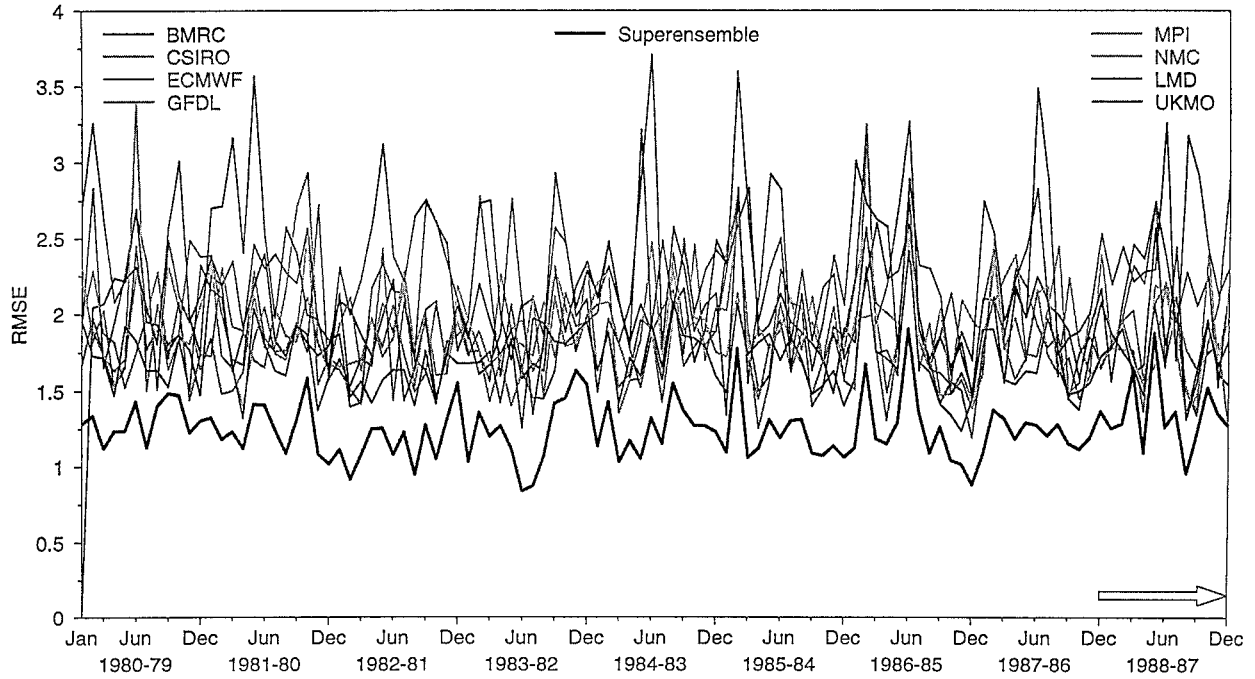


FIG. 5. The rms errors of the 1-yr difference in meridional wind for the multimodels and the superensemble forecasts (of these differences; mm day^{-1}).

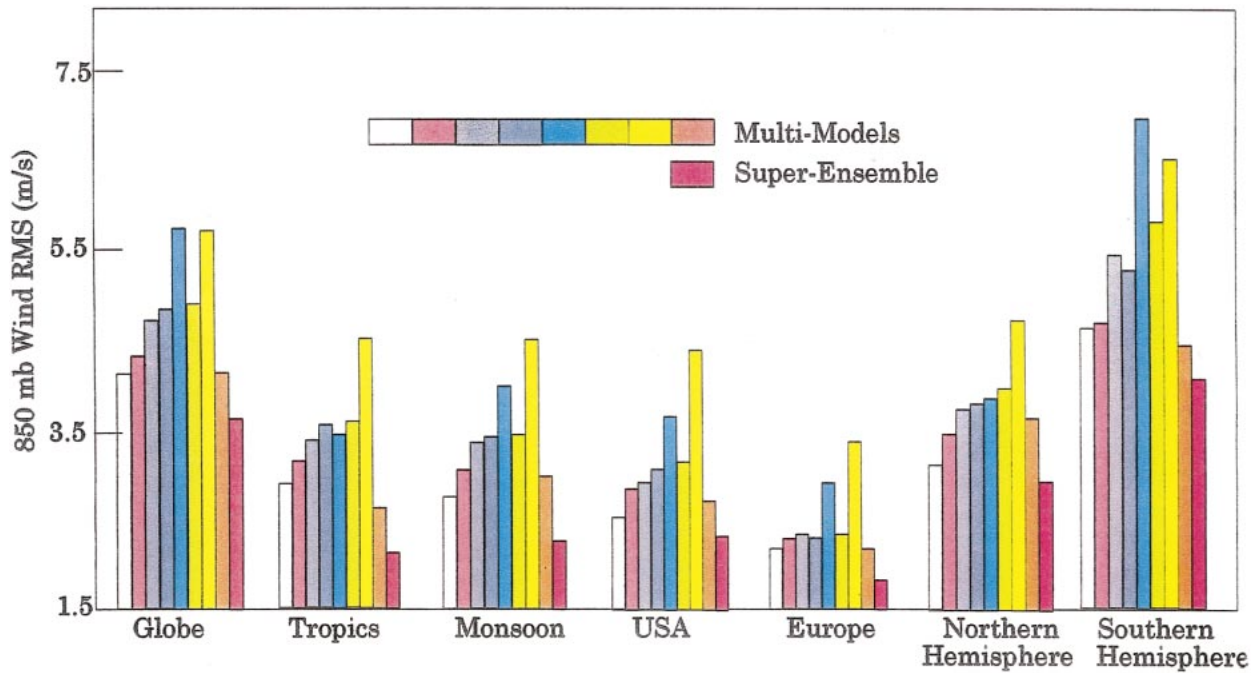


FIG. 6. The rms error of the 850-hPa winds on day 3 of the forecasts during Aug 1998. The results for the multimodels follow from left to right, and the results for the ensemble mean and the superensemble are shown in the far right, respectively, (m s^{-1}).

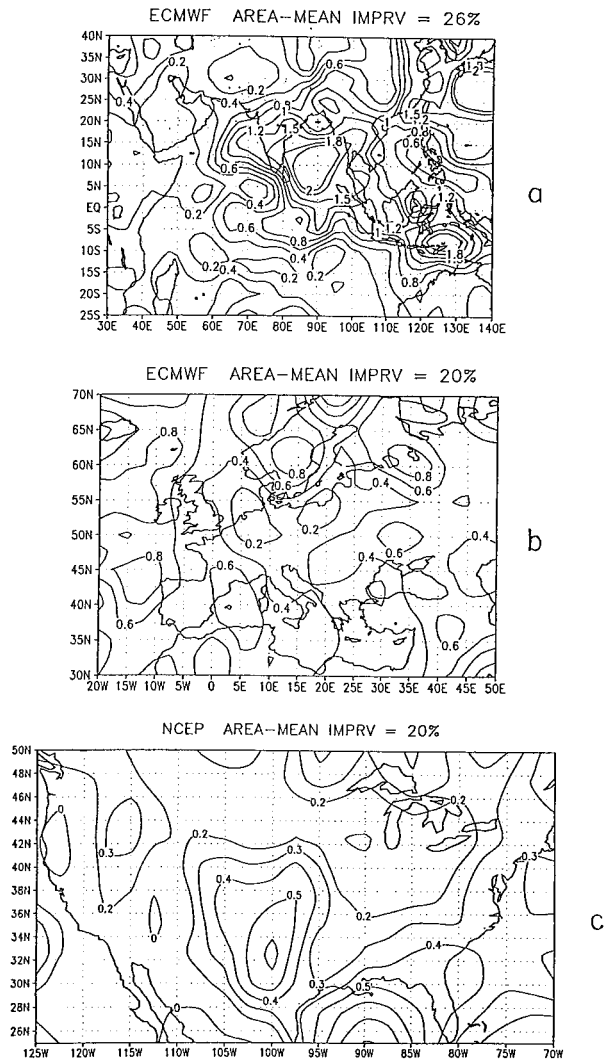


FIG. 7. Improvement on 3-day wind forecast (rms in m s^{-1}), averaged over an entire month, Aug 1998, with respect to (a) ECMWF forecasts over India, (b) ECMWF forecasts over Europe, and (c) NCEP forecasts over North America. Respective verification analysis was used for calculating rms errors.

Verification skills are calculated against ECMWF analysis. For the individual models, the rms wind error increases from roughly 2 to 6 m s^{-1} during the 3-day forecast. The multimodel superensemble has the least error compared to each of the models. Furthermore, we noted that the 3-day skill of the superensemble (rms error 1.6–4 m s^{-1}) is comparable to the day-1 errors for several of the models. We also show the rms errors of the ensemble mean (orange) in these histograms. The superensemble has a lower error compared to that of the ensemble mean in all cases. Over many regions, the ensemble mean has a higher error compared to some of the best models. This is one of the most promising results of the proposed approach.

Figure 7b shows an example of a typical 3-day fore-

cast percent improvement over Europe for the 850-hPa winds. Although the ECMWF and superensemble forecasts are visually similar, there was a 26% area average improvement of the rms wind error for the superensemble over the ECMWF prediction. This is also typical of most forecasts as seen in Figs. 7a and 7c over the monsoon domain and the United States, respectively (Note: NCEP forecast used in 11c). Small discrepancies over most domains account for roughly 20% differences in skill. The superensemble always shows a closer agreement to the analysis field compared to the multimodels by invoking the models' bias corrections.

The rms error reduction of the zonal wind at 850 hPa over the entire Tropics for August 1998 is shown in Fig. 8a. The different lines show the percent improvement over the performance of respective models for forecast days 1–3. The analysis field of the respective models is used to calculate the rms errors. Overall, we note a 10%–30% improvement in the 850-hPa zonal wind rms errors from the superensemble. The ensemble means percentages lie in the range of -20% to $+20\%$. In all cases, the superensemble exhibits a much larger reduction of error with respect to the multimodel errors and the ensemble mean errors. It is also interesting to note that the percent improvement increases with the number of days of forecast.

To assess how many models are minimally needed to improve the skill of the multimodel superensemble, we examined sequentially the issue using one to seven models. Results for global wind rms errors at 850 mb are shown in Fig. 8b. We sequentially added models with lower and lower skill as we proceeded from one model to seven models. The dashed line shows the error for the ensemble mean and the solid line indicates that of the superensemble. The superensemble skill is higher than that of the ensemble mean for any selection of the number of ensemble members. The skill of the superensemble between 4 and 7 is small, that is, around 3.6 m s^{-1} . The ensembles mean error increases as we add more ensemble members beyond 3. That increase came from the gradual addition of models with low skill. For two and three models, the skill differences are the smallest. That rapid increase of error beyond three models is not seen for the superensemble since it automatically assigns low weights to the model's low skills. It is also worth noting that half the skill improvement comes from a single model for this procedure. That is roughly a 5% improvement for the 850-mb winds for 3-day forecasts.

The ensemble of one is essentially the procedure for the removal of the bias of a single model. Here the regression utilizes the past history of the model for each geographical location, each vertical level, and for each variable separately. We carried out this exercise for each of the seven models separately. We then performed an ensemble mean of these individual bias removed fields. This was carried out separately for days 1, 2, and 3 of forecasts. The rms errors of the tropical (30°S – 30°N) 850-mb winds for this ensemble mean were compared

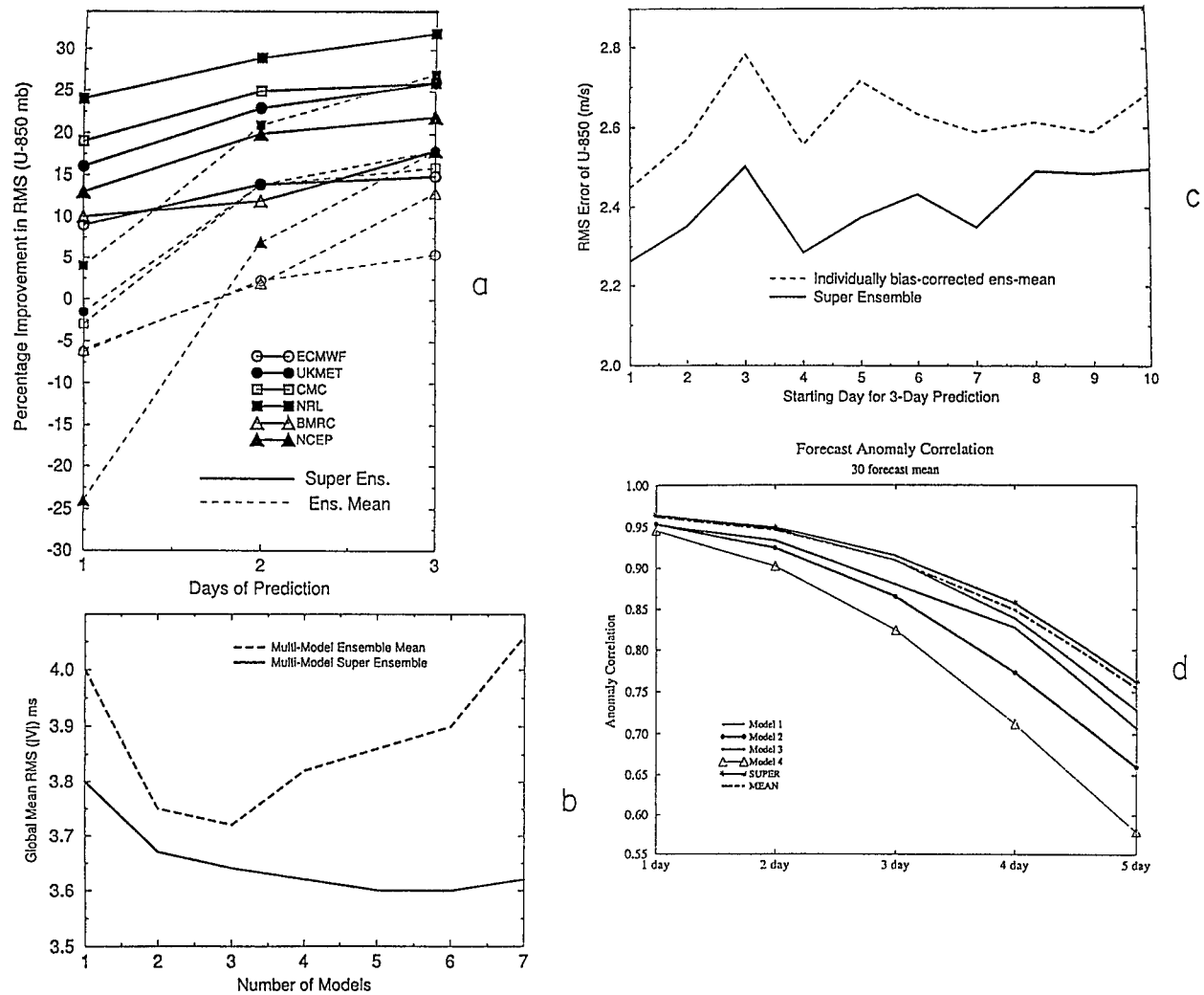


FIG. 8. (a) The percentage improvement from the superensemble (solid lines) and of the ensemble mean (dashed lines) with respect to different models. Results for different member multimodels use the respective analysis for verification. Abscissa denotes days of forecast. (b) Global mean forecast rms error of the total wind at 850 hPa during Aug 1998 for 3-day forecasts (m s^{-1}). The abscissa denotes the number of models. The dashed line denotes the error of the ensemble mean, whereas the solid line denotes the error of the superensemble. (c) The rms errors at 850-mb wind during Aug 1998 for 3-day forecasts for a sequence of days (m s^{-1}). The dashed line denotes the results for the ensemble mean where the bias of the individual models was first removed. The solid line denotes the results of the superensemble where the collective bias of all models was removed. (d) Anomaly correlations at 500 hPa during JFM 1999. The top curve is that for the superensemble, the dashed line denotes the error of the superensemble, and the other curves belong to several multimodels.

to that of the statistical multimodel superensemble, Fig. 8c. We clearly see that removal of collective bias is superior to doing that separately.

When we remove the bias of each model separately, we find many interesting results. The skill of forecast of each model is increased. The poorer models, however, do not compare as well as the best model after the removal of the individual biases. If we perform an ensemble mean of all the models, after the bias is removed individually, the mean still includes the poorer models. That is because an equal weight is assigned to each model for the ensemble averaging (after the bias removal). That has an effect of degrading the results. This

is apparent in Fig. 8b, where as more models are included, the results eventually become inferior to using a single (best) model. The proposed multimodel forecast technique is superior to the “individual bias removed ensemble mean” because the superensemble does not assign an equal weight of 1.0 to all the models. Thus the poorer models are not excluded in this averaging; however, in those geographical locations and vertical levels for certain variables where they may have a higher skill, this inclusion is helpful. This global bias removal (geographically, vertically, and for different variables) is a dominant aspect of the superensemble. For proper application of such a technique, a priori determination

of which models perform well and at what locations could greatly enhance the procession time.

NWP, 500-mb anomaly correlations

Global 500-mb geopotential height for days 0–6 of forecasts for the following models were available: NCEP, ECMWF, NOGAPS, FSU, and UKMO. These daily forecasts covered the period between 1 January and 31 March 1999. These were the datasets that were readily available to us on GTS. Anomaly correlations were defined by using an independent 10-yr ECMWF daily analysis database to define the climatological 500-mb geopotential heights for the months of JFM for the extraction of the anomalies. Figure 8d illustrates the anomaly correlation of the 500-mb geopotential heights averaged over JFM 1999. The respective multimodel forecast results, the anomaly correlation of the ensemble mean and that of the statistical superensemble are shown here. Among the individual models the anomaly correlation of the European Center stands out in terms of skill. The verifications were carried out using the ECMWF analysis that is a bit unfair to the other models, but used as illustration. The anomaly correlation of the ensemble averaged geopotential heights performs better than all the individual models. The results of the superensemble show slightly better results. The smaller systematic bias of height fields and the smaller improvement for this variable reemphasizes the importance of bias removal on this technique. The same procedure appears to work equally well for other variables and hence can, in principle, be used at all vertical levels for all of the model variables including precipitation forecasts.

6. Hurricane track and intensity forecasts from the superensemble

Over the Atlantic, Caribbean, and the Gulf of Mexico, there were 14 named tropical systems during the 1998 hurricane season. Of these, four became tropical storms and the remaining reached hurricane strength. The 1998 hurricane season was active between 27 July and 1 December. Our interest here is on the superensemble forecasts skill for the tracks and intensity of these storms. Ideally, it would have been desirable to develop the statistics of a control (a training) period from one or more years of past history. It is imperative that no major model changes occur within the multimodels during the course of these “control” and “forecast” periods. A number of models generally provide hurricane forecast information; these include NCEP/NOAA, GFDL, UKMO, NOGAPS/NRL, official forecasts of the NHC and the FSU suite of models. Table 4 provides a short summary of these models. In addition to these forecasts, there exists some in-house prediction models at the NHC that routinely provide hurricane intensity forecasts (SHIPS, SHIFOR) De Maria and Kaplan (1999). It was

not possible to acquire uniform datasets, for a string of years, without major model changes. Four of the aforementioned models, NCEP/MRF, FSU, GFDL, and NOGAPS, underwent major resolution changes after the 1997 season. Thus, it was not possible to derive the statistics based on the multimodel performance with the 1997 datasets and use those to forecast during the 1998 hurricane season. It was still possible to carry out forecasts for each and every storm of the 1998 season using a cross-validation approach, which excludes all days of a particular hurricane while computing the weights for its forecast. All other hurricanes of the season are included in the evaluation of the weights for the storm being forecasted. This entails deriving the multimodel statistics from all storms of 1998, sequentially, excluding the specific storm that is being forecasted. This is a robust approach for assessing the validity of the proposed approach for the forecasts of the storms of 1998. Each of the 14 named storms in 1998 lasted over several days. Thus, it was possible to develop a multimodel forecast database for 68 sets of forecasts for this entire season. The methodology for the calculation of regression coefficients is identical to that described in appendix A. The databases are the model forecasts, the observed, and the official forecast estimate of the track (position) and intensity (maximum wind estimate) every 12 h, starting from an initial time and ending at hours 72–144, as dictated by the termination of such forecasts. The weights for the multimodels vary for each forecast period (i.e., hour 12, 24, 36, 48, and 72). We found that this gave better results than providing a single set of multimodel weights for all three days.

First we shall look at the overall statistics of the track and intensity forecasts for the entire year of 1998. Here we show the day 1, day 2, and day 3 skills for these forecasts. The results for the training period are shown in Fig. 9a, and those for the superensemble forecasts (cross validation) are presented in Fig. 9b. During 1998, the best-track forecasts came from the NHC official component. These are in fact subjective forecasts made by the forecasters of the NHC in Miami. They are essentially based on consensus from among the suite of model forecasts available to them. In addition, they also make use of their past experience on subjective hurricane forecasting while arriving at these official forecasts. The histograms on the left panels show, respectively, the track forecasts. The superensemble track forecasts are superior to those of all other models and official forecasts for each of the three days. The superensemble, in the training phase, has position errors of the order of 0.85° , 1.5° , and 1.9° for days 1, 2, and 3 of forecasts, respectively. The corresponding position errors for the superensemble forecasts (Fig. 9b) are 1.25° , 1.9° , and 2.6° latitude for days 1, 2, and 3, respectively. Similar results hold for the intensity forecasts as well. The rms errors of intensity for the training and forecasts from the superensemble are better than those of all other models including the ensemble mean.

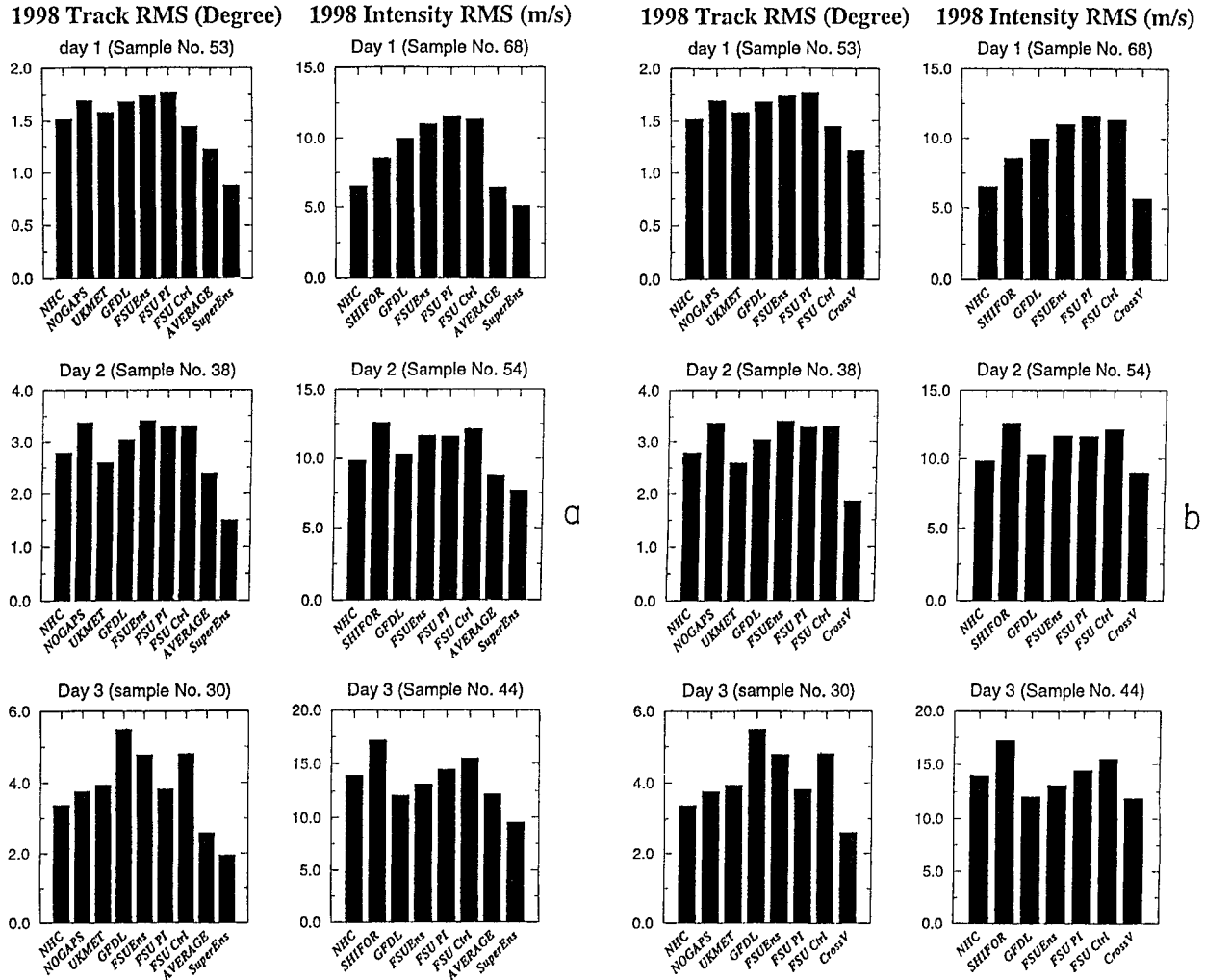


FIG. 9. (a) The rms error for the position (left panel) and the intensity (right panel) for 3-day forecasts of all hurricanes of 1998. Day 1, day 2, and day 3 forecast errors for the multimodels, the ensemble mean, and the superensemble are shown here. This illustration shows the results for the control period. (b) Same as (a) but for the forecast period using the cross-validation technique.

In Figs. 10a–f we show several examples of predicted tracks for Atlantic hurricanes such as Alex, Bonnie, Danielle, Georges, and Mitch of 1998. Here the predicted tracks from several models are displayed. These also include three models from FSU as well, which are

- 1) a control forecast with the global spectral model at the resolution T126 with no physical initialization,
- 2) a forecast at the resolution T126 that includes physical initialization (Krishnamurti et al. 1991),
- 3) an ensemble forecast with a high-resolution regional spectral model following Zhang and Krishnamurti (1999).

In these illustrations, the tracks use the following symbols:

- BEST denotes the observed best track;
- FSUC denotes the FSU control forecast that does not include rain-rate initialization;

- FSUP denotes the results for FSU global spectral model that includes physical initialization;
- FSUE denotes the ensemble averaged track for an FSU regional spectral model, Cocks (1998), which is nested within the FSU global spectral model;
- OFCL denotes the official forecast from the NHC;
- GFDL denotes the forecast made from the geophysical fluid dynamic or Multiple-Mesh Model of Princeton (NCEP/NOAA);
- NGPS denotes the U.S. Navy’s NOGAPS model;
- UKMT denotes the UKMO’s global model; and
- SENS denotes the track based on the superensemble forecasts.

In Fig. 10 we note that the tracks of the superensemble are superior to the forecasts made by the multimodels. A higher accuracy of track forecasts results from the improved timing for the locations of the superensemble with respect to the “best track.” In some instances the

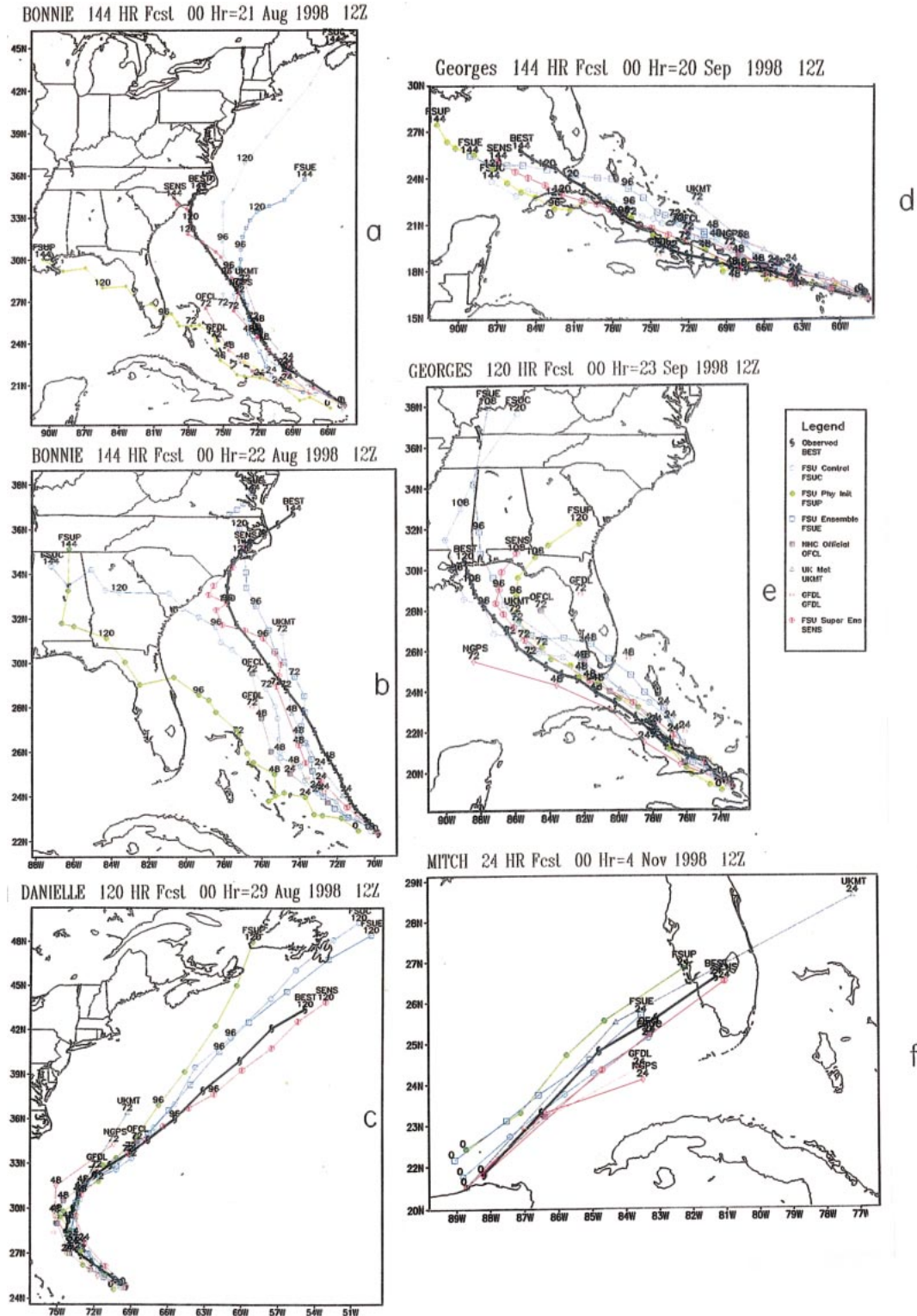


FIG. 10. (a), (b), (c), (d), (e), (f) The predicted hurricane tracks for a number of multimodels. Also shown are the predicted tracks from the Official NHC's estimate and the superensemble. The observed best track is shown in black.

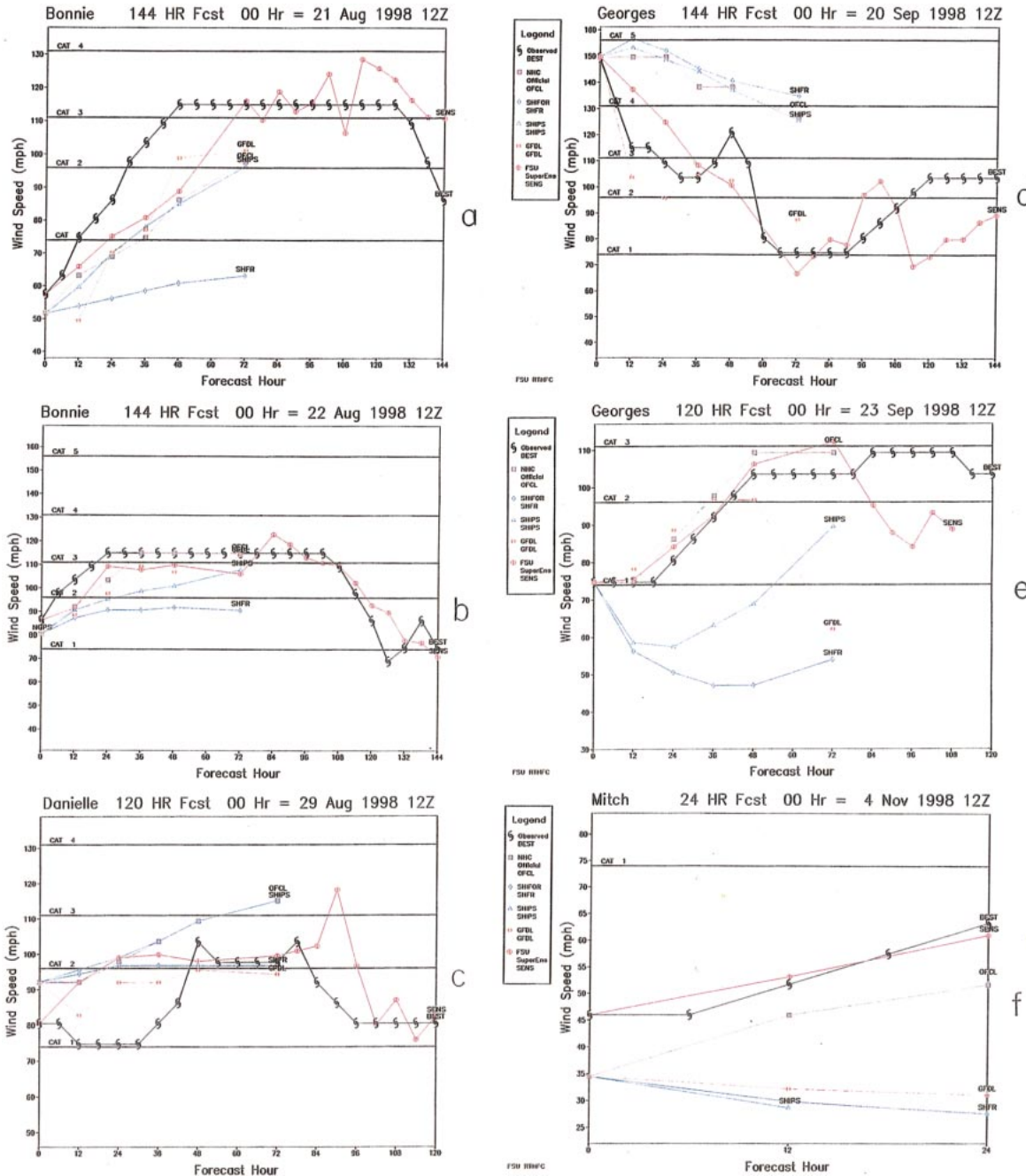


FIG. 11. (a), (b), (c), (d), (e), (f) Same as Fig. 10, but for the intensity of various storms, the official best estimate, and the superensemble forecasts.

heading of some of the multimodel storms appears more accurate, but the timing of their locations compared to the best track is less accurate compared to the superensemble.

Several specific examples of intensity forecasts from the superensemble are shown in Figs. 11a–f. A number of these multimodels provided 3-day forecasts; these included FSU (three models) through day 6 of forecasts, NGPS, and GFDL (not all are illustrated in Fig. 11).

We have illustrated two other intensity-forecast estimates provided by the NHC, the SHIPS and SHIFOR statistical models; see De Maria and Kaplan (1999). SHIFOR is a statistical hurricane forecast model. It is based on climatology and persistence and applies only to storms over the ocean. SHIPS is another simple climatological scheme of the Hurricane Research Division that makes use of parameters such as maximum possible intensity, current intensity, vertical shear of the tropo-

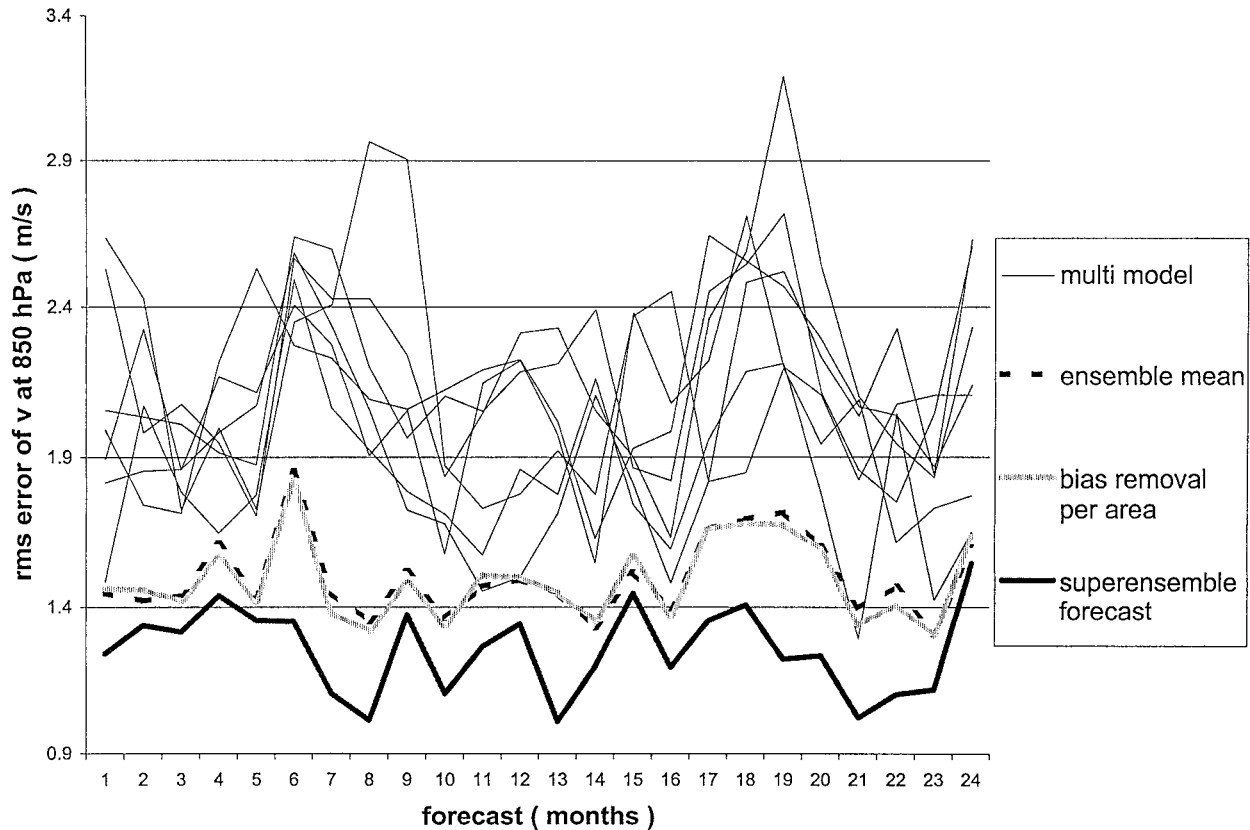


FIG. 12. The rms error of the meridional wind at 850 hPa (m s^{-1}) for the AMIP data comparing the superensemble (thick dark line) to the individual models (thin lines), the ensemble mean (dashed line), and the ensemble with individual model bias removed.

spheric horizontal wind, persistence of intensity change over the last 12 h, eddy flux convergence of momentum at 200 mb, zonal wind, and temperature within 1000 km of the storm center. The forecasts for days 4, 5, and 6 came from the FSU models (global without physical initialization, global with physical initialization, and an ensemble average intensity). Overall, a skill in the intensity forecasts is clearly evident from the superensemble. The FSU models contribute to the skill of the superensemble during days 4, 5, and 6, providing very useful information for these extended superensemble forecasts of intensity. The intensity forecasts from SHIPS and SHIFOR are reasonable for the first three days of forecasts; however, it is apparent from the overall statistics, presented earlier (Fig. 9), that the superensemble outperforms all other models in the intensity forecast. It is generally recognized that numerical models have rather poor skill in intensity forecasts. This was reflected in the multimodel forecasts of the storms of 1998. In most instances the intensity forecasts of the superensemble are within a “category” (as indicated by the horizontal lines on the plots) of the observed best estimates of intensity. Overall these intensity forecasts are quite impressive through day 6 of forecasts, given that the current state of the prediction is accepted as

being quite poor. SHIPS and SHIFOR often underestimate the storm intensity by several categories.

7. Concluding remarks and future outlook

One of the major issues we have addressed here relates to the removal of the collective errors of all models. That approach exhibits different characteristics from an ensemble average of the results. The straight average approach assigns an equal weight of 1.0 to all the models and may include several poor models. The mean of these poor models degrades the overall results, as more of such models are included. Our approach of the superensemble assigns weights to each model, based on its performance, geographically, vertically, and for each variable separately. This would not assign a high weight to the poorer models over selected regions based on their past performance.

A postprocessing algorithm based on multiple regression of multimodel solutions toward observed fields during a training period shows promise for various applications during a subsequent test period. The resulting superensemble reduces the forecast errors below those of the multimodels. Application of the algorithm to the simulation of seasonal climate, global weather, and hur-

ricane track and intensity shows the most promise. Root-mean-square errors of the seasonal superensemble simulations for winds are reduced to less than one-half the errors of the best models. Precipitation forecast errors of the order of a few millimeters per day are around 300% better than the individual models. The results herein clearly show that the superensemble has higher skills. This performance is attributed to the superior statistical combination of model anomaly fields compared to the standard technique.

One reason for the algorithm's superior performance is its use of rms error for defining the statistical forecast weights. The same metric is used to define rms error skill scores. The individual model anomalies are not as good as the removal of the statistical reconstruction of all the models using the regression technique. The rms error of the superensemble is shown to be considerably less than that a simple reconstruction of models. Initially, these results seemed strange since we are dealing with a linear regression. Upon closer examination, we noted that the computation of the root-mean-square differences in obtaining the statistical weights implies that this is not a simple linear problem.

The seasonal climate simulations using the AMIP dataset show that the superensemble rms errors for the monthly mean fields (such as winds, temperature, and precipitation) are quite small compared to all other models. Due to regional interannual variability in temperature and precipitation, we examined the AMIP and WMO/WGNE seasonal monsoon datasets in the context of the present modeling. We prepared a dataset from 1-yr differences for the models and the observed fields during the control (or training) phase. We created a simulation of differences using the statistics resulting from the multiple regression technique. The rms errors of the difference fields of the superensemble were far less than those of the individual models. Monsoon rainfall differences between 1988 (heavy rainfall year) and 1987 (a weak rainfall year) were very promising, again revealing that the superensemble had less errors compared to the multimodels. The AMIP includes atmospheric general circulation models that utilize prescribed SSTs and sea ice. Therefore, similar work is needed to explore the skill of the proposed methodology for coupled atmospheric-ocean models.

Our analysis on the number of models needed to improve the skill of the superensemble shows that roughly six models produce the lowest rms errors for the superensemble global NWP. The number of models needed is likely to change with changes in the models making up the multimodel set. In the global NWP application, the superensemble stands out, providing roughly a 20% improvement over the best models. While examining local improvements arising from the superensemble, we noted that it extracts the best information from a number of models. An example of that shown in Fig. 7b illustrates that an erroneous forecast over a given region from the best model is corrected from information ex-

tracted from another model that carries a higher weighting over that region. Therefore, improvement by collective inclusion of good features makes the superensemble stand out.

Forecasting the Atlantic hurricanes of 1998 required using that season's dataset for both the control and the forecast phase. This was based on the cross-validation technique where no data for a storm being forecasted was used for the evaluation of the statistical weights. These forecasts showed the best results compared to all other multimodels. There was a major improvement in the position and intensity of each storm compared to most multimodels and in the overall seasonal statistics; the superensemble outperformed all other models.

Further exploration of methods other than the proposed simple linear multiple regression deserves to be explored. We have started to examine the use of nonlinear regression and neural network training methods for possible further improvement in skill. We have also tested seasonally varying weights for climate applications, which prove to be very promising, but the training size is greatly reduced. Further work is needed on detailed analysis of the probabilistic properties and application for the technique and the weights obtained, particularly for climate. This may also lead to a more efficient and superior technique.

Finally, a detailed separate study on the probabilistic estimates is in preparation using the so-called Brier skill scores among others. In that work we will show that the superensemble does have measurably much higher skills for seasonal climate forecasts compared to all individual models, climatology, and the ensemble mean. These results will be published in a separate paper.

Acknowledgments. This work could not have been completed without the data support of the weather services of the world. We are especially thankful to the ECMWF which has provided, over many years, the real-time data, for the execution of FSU models. This work was supported by the following grants: NASA Grant NAG8-1199, NASA Grant NAG5-4729, NOAA Grant NA86GP0031, NOAA Grant NA77WA0571, NSF Grant ATM-9612894, and NSF Grant ATM-9710336.

APPENDIX A

Creation of a Multimodel Superensemble Prediction at a Given Grid Point

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

where S = superensemble prediction, \bar{O} = time mean of observed state, a_i = weight for model i , i = model index, N = number of models, \bar{F}_i = time mean of prediction by model i , and F_i = prediction by model i . The weights a_i are computed at each grid point by minimizing the following function:

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

where O = observed state, t = time, and t -train = length of training period (96 months in present case for seasonal prediction).

APPENDIX B

List of Acronyms

Acronym	Meaning
AGCMs	Atmospheric general circulation models
AMIP	Atmospheric Model Intercomparison Project
BEST	National Hurricane Centers Best Track Estimate
BMRC	Bureau of Meteorology Research Centre
CCM2	Community Climate Model 2 of NCAR
CCM3	Community Climate Model 3 of NCAR
DERF	Deterministic Extended Range Forecast
ECMWF	European Centre for Medium-Range Weather Forecasts
FSU	The Florida State University
FSUC	FSU control experiment
FSUE	FSU ensemble experiments
FSUP	FSU physical initialization experiment
GFDL	Geophysical Fluid Dynamics Laboratory
GPCP	Global Precipitation Climatology Project
GTS	Global Telecommunication System
hPa	Hectopascals
ITCZ	Intertropical convergence zone
JMA	Japan Meteorological Agency
LMD	Laboratoire de Meteorologie Dynamique
mb	Millibar
mm	Millimeter
MRF	Medium range forecasts
$m\ s^{-1}$	Meters per second
NCAR	National Center for Atmospheric Research
NGPS	U.S. Navy's NOGAPS model
NHC	National Hurricane Center
NOAA	National Oceanic and Atmospheric Administration
NOGAPS	Navy Operational Global Atmospheric Prediction System
NRL	Naval Research Laboratory
NWP	Numerical Weather Prediction
OFCL	Official
OLR	Outgoing longwave radiation
rms error	Root-mean-square error
RPN	Recherche en Prévision Numérique
SENS	Superensemble
SHIFOR	Statistical Hurricane Intensity Forecast
SHIPS	Statistical Hurricane Intensity Prediction System
UKMO	United Kingdom Meteorological Office
UTC	Universal standard time

WGNE	Working Group on Numerical Experimentation
WMO	World Meteorological Organization

REFERENCES

- Arakawa, A., 1972: Design of the UCLA general circulation model. Tech. Rep. 7, Department of Meteorology, University of California, Los Angeles, 116 pp. [Available from Department of Atmospheric Sciences, University of California, Los Angeles, Los Angeles, CA 90024.]
- Arkin, P., and J. Janowiak, 1991: Analyses of the global distribution of precipitation. *Dyn. Atmos. Oceans*, **16**, 5–16.
- Blondin, C., and H. Bottger, 1987: The surface and subsurface parameterization scheme in the ECMWF forecasting system: Revision and operational assessment of weather elements. ECMWF Tech. Memo. 135, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, 112 pp.
- Clough, S. A., F. X. Kneizys, R. Davies, R. Gernache, and R. Tipping, 1980: Theoretical line shape for H_2O vapor: Application to continuum. *Atmospheric Water Vapor*, T. D. Wilkerson and L. H. Ruhnke, Eds., Academic Press, 695 pp.
- Cocke, S., 1998: Case study of Erin using the FSU Nested Regional Spectral Model. *Mon. Wea. Rev.*, **126**, 1337–1346.
- Deardorff, J. W., 1977: A parameterization of ground-surface moisture content for use in atmospheric prediction models. *J. Appl. Meteor.*, **16**, 1182–1185.
- De Maria, M., and J. Kaplan, 1999: An updated Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **14**, 326–337.
- Ducoudre, N., K. Laval, and A. Perrier, 1993: SECHIBA, a new set of parameterizations of the hydrologic exchanges at the land-atmosphere interface within the LMD atmospheric general circulation model. *J. Climate*, **6**, 248–273.
- Dumenil, L., and E. Todini, 1992: A rainfall-runoff scheme for use in the Hamburg climate model. *Advances in Theoretical Hydrology: A Tribute to James Dooge*, J. P. O'Kane, Ed., European Geophysical Society Series on Hydrological Sciences, Vol. 1, Elsevier Press, 129–157.
- Fouquart, Y., and B. Bonnel, 1980: Computation of solar heating of the Earth's atmosphere: A new parameterization. *Beitr. Phys. Atmos.*, **53**, 35–62.
- Gadgil, S., and S. Sajani, 1998: Monsoon precipitation in AMIP runs (Results from an AMIP diagnostic subproject). World Climate Research Programme—100, WMO/TD-837, 87 pp.
- Gates, W. L., and A. B. Nelson, 1975: A new (revised) tabulation of the Scripps topography on a one-degree global grid. Part 1: Terrain heights. Tech. Rep. R-1276-1-ARPA, The Rand Corporation, Santa Monica, CA, 132 pp. [Available from The Rand Corporation, Santa Monica, CA 90407.]
- , and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29–55.
- Gregory, D., and P. R. Rowntree, 1990: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Wea. Rev.*, **118**, 1483–1506.
- Ingram, W. J., 1993: Radiation, version 1. Unified Model Documentation Paper 23, The Met. Office, 254 pp. [Available from The Met. Office, Bracknell, Berkshire RG12 2SZ, United Kingdom.]
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Krishnamurti, T. N., J. S. Xue, H. S. Bedi, K. Ingles, and D. Oosterhof, 1991: Physical initialization for numerical weather prediction over the tropics. *Tellus*, **43AB**, 53–81.
- Kuo, H. L., 1965: On formation and intensification of tropical cyclones through latent heat release by cumulus convection. *J. Atmos. Sci.*, **22**, 40–63.
- Lacis, A. A., and J. E. Hansen, 1974: A parameterization for the

- absorption of solar radiation in the Earth's atmosphere. *J. Atmos. Sci.*, **31**, 118–133.
- LaRow, T. E., and T. N. Krishnamurti, 1998: Initial conditions and ENSO prediction using a coupled ocean–atmosphere model. *Tellus*, **50A**, 76–94.
- Lorenz, E. N., 1963: Deterministic non-periodic flow. *J. Atmos. Sci.*, **20**, 130–141.
- Manabe, S., J. Smagorinsky, and R. F. Strickler, 1965: Simulated climatology of a general circulation model with a hydrologic cycle. *Mon. Wea. Rev.*, **93**, 769–798.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliaigis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Morcrette, J.-J., 1991: Radiation and cloud radiative properties in the ECMWF operational weather forecast model. *J. Geophys. Res.*, **96**, 9121–9132.
- Mullen, S. L., and D. P. Baumhefner, 1994: Monte Carlo simulations of explosive cyclogenesis. *Mon. Wea. Rev.*, **122**, 1548–1567.
- Palmer, T. N., 1999: A nonlinear dynamical perspective on climate prediction. *J. Climate*, **12**, 575–591.
- , and D. L. T. Anderson, 1994: The prospects for seasonal forecasting—A review paper. *Quart. J. Roy. Meteor. Soc.*, **120**, 755–793.
- Pan, H.-L., 1990: A simple parameterization scheme of evapotranspiration over land for the NMC Medium-Range Forecast Model. *Mon. Wea. Rev.*, **118**, 2500–2512.
- Phillips, T. J., 1994: A summary documentation of the AMIP models. PCMDI Rep. 18, Lawrence Livermore National Laboratory, 343 pp. [Available from PCMDI, Lawrence Livermore National Laboratory, Livermore, CA 94550.]
- , 1996: Documentation of the AMIP models on the World Wide Web. *Bull. Amer. Meteor. Soc.*, **77**, 1191–1196.
- Polcher, J., 1994: Etude de la sensibilit  du climat tropical a la deforestation. Ph.D. dissertation, Universit  Pierre et Marie Curie, Paris, France, 215 pp. [Available from Universit  Pierre et Marie Curie, Paris 6-4, place Jussieu, 75005 Paris, France.]
- Rodgers, C. D., and C. D. Walshaw, 1966: The computation of infrared cooling rate in planetary atmospheres. *Quart. J. Roy. Meteor. Soc.*, **92**, 67–92.
- Sasamori, T., J. London, and D. V. Hoyt, 1972: Radiation budget of the Southern Hemisphere. *Southern Hemisphere Meteorology, Meteor. Monog.*, No. 35, Amer. Meteor. Soc., 9–22.
- Schemm, J. K., S. Schubert, J. Terry, and S. Bloom, 1992: Estimates of monthly mean soil moisture for 1979–1988. NASA Tech. Memo. 104571, GSFC, Greenbelt, MD, 260 pp.
- Schwarzko pt, M. D., and S. B. Fels, 1991: The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes. *J. Geophys. Res.*, **96**, 9075–9096.
- Shuttleworth, W. J., 1988: Macrohydrology: The new challenge for process hydrology. *J. Hydrol.*, **100**, 31–56.
- Slingo, A., and R. C. Wilderspin, 1986: Development of a revised longwave radiation scheme for an atmospheric general circulation model. *Quart. J. Roy. Meteor. Soc.*, **112**, 371–386.
- Spencer, R. W., 1993: Global oceanic precipitation from MSU during 1979–91 and comparison to other climatologies. *J. Climate*, **6**, 1301–1326.
- Stone, H. M., and S. Manabe, 1968: Comparison among various numerical models designed for computing infrared cooling. *Mon. Wea. Rev.*, **96**, 735–741.
- Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800.
- Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- Warrilow, D. A., A. B. Sangster, and A. Slingo, 1986: Modelling of land surface processes and their influence on European climate. Dynamical Climatology Branch, The Met. Office, DCTN 38, 80 pp. [Available from The Met. Office, Bracknell, Berkshire RG12 2SZ, United Kingdom.]
- World Climate Research Program, 1992: Simulation of interannual and intraseasonal monsoon variability. World Climate Research Program (WCRP) Report to the World Meteorological Organization, WMO Rep. WMO/ID-470, 218 pp.
- Zhang, Z., and T. N. Krishnamurti, 1999: A perturbation method for hurricane ensemble predictions. *Mon. Wea. Rev.*, **127**, 447–469.