# Improving Global Model Precipitation Forecasts over India Using Downscaling and the FSU Superensemble. Part I: 1–5-Day Forecasts

T. N. Krishnamurti, A. K. Mishra, and A. Chakraborty*

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

M. Rajeevan

*National Climate Centre, India Meteorological Department, Pune, India*

## ABSTRACT

The availability of daily observed rainfall estimates at a resolution of 0.5° × 0.5° latitude–longitude from a collection of over 2100 rain gauge sites over India provided the possibility for carrying out 5-day precipitation forecasts using a downscaling and a multimodel superensemble methodology. This paper addresses the forecast performances and regional distribution of predicted monsoon rains from the downscaling and from the addition of a multimodel superensemble. The extent of rainfall prediction improvements that arise above those of a current suite of operational models are discussed. The design of two algorithms one for downscaling and the other for the construction of multimodel superensembles are both based on the principle of least squares minimization of errors. That combination is shown to provide a robust forecast product through day 5 of the forecast for regional rains over the Indian monsoon region. The equitable threat scores from the downscaled superensemble over India well exceed those noted from the conventional superensemble and member models at current operational large-scale resolution.

## 1. Introduction

This paper combines two features for numerical weather prediction for precipitation forecasts. One of these is the multimodel superensemble based on our previous studies (e.g., Krishnamurti et al. 2000a; Mishra and Krishnamurti 2007). The other feature is a statistical downscaling that relies to the availability of reliable high-resolution rainfall observations. The downscaling of information from larger-scale models toward higher resolution is generally carried out using statistical or dynamical methods (Huth 2002; Druyan et al. 2002; Kanamitsu and Kanamaru 2007; Kanamaru and Kanamitsu 2007). Downscaling has been done mostly for climate modeling where the introduction of higher resolution by one-way nested models utilizes the information from the larger-scale models including features at the higher resolution such as orography, land surface characterization, and modified physical parameterization. In seasonal climate forecasts with large-scale global models (e.g., Chakraborty and Krishnamurti 2009, hereafter Part II) downscaling (statistical or dynamical) is involved using anomalies of different variables. The anomalies (or perturbations) are generally made with respect to the global model products as a reference mean state. Two recent studies by Kanamitsu and Kanamaru (2007) and Kanamaru and Kanamitsu (2007) follow this procedure for the prediction of higher-resolution anomalies. These studies have addressed the California rainfall on seasonal time scales. The statistical downscaling has also been examined for regional weather and climate forecasts (Druyan et al. 2002; Pandey et al. 2000; Huth 2002; von Storch et al. 2000). Recently a spectral nudging technique (von Storch et al. 2000) has been used for dynamical downscaling. In these studies one draws a statistical relationship among the large-scale dynamical forecast products and local (or mesoscale) high-resolution time series based on available information.

---

* Current affiliation: Centre for Atmospheric and Oceanic Sciences, Indian Institute of Science, Bangalore, India.

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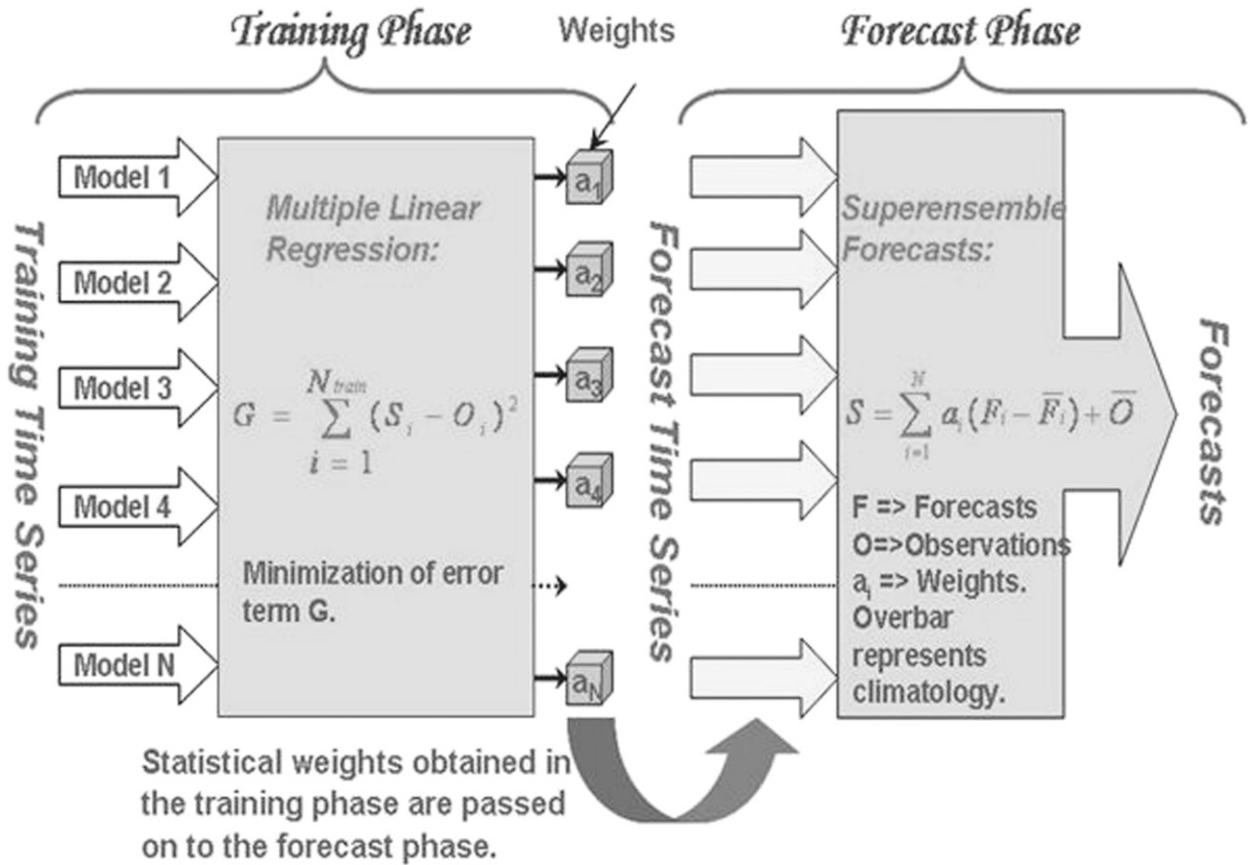E-mail: tnk@io.met.fsu.edu

FIG. 1. Locations of rain gauge stations over India.

The improvement on mesoscale NWP perhaps should be approached using high-resolution mesoscale models directly. Ideally we need an objective assessment on the performance of several mesoscale models run in real time by a large diverse group of modelers. That type of culture exists for the assessment of large-scale operational models with regularly performed skill score assessments to determine the value of a given medium-range forecast. However, these same types of routine model intercomparisons do not exist for mesoscale models. There are several factors that limit an assessment on the accuracy of mesoscale NWP models. The lack of a near uniform distribution of mesoscale observations makes it difficult to perform domain averaged skill scores.

While increasing the horizontal resolution of model forecasts, caution must be exercised when comparing the objective scores of models with different resolutions. Even when increased resolution produces better-resolved mesoscale structures, the increase in the possibility of space and temporal errors often leads to



FIG. 2. Difference in the number of rain gauge stations on 1° × 1° grid box between two versions of IMD data. (V.2 − V.1) of stations: V.2 (2140) and V.1 (1803).

FIG. 3. Schematic representation of the FSU multimodel superensemble methodology. In the training phase, the model forecasts are regressed against the observation to obtain differential weights. These weights are then passed on to the forecast phase to create superensemble forecasts.

larger root-mean-squared errors comparable to lower-resolution operational forecasts. Small phase errors in high-resolution models lead to these well-known double-penalty problems (Anthes 1983; Mass et al. 2002). An increase in resolution often serves to penalize the model's objective score because of increased opportunity to 1) miss the forecast and 2) forecast a false alarm. Several techniques are used to try and circumvent the double penalty problem, such as "smoothening" and "object-oriented verification." Different verification schemes have been developed to evaluate the relative improvement of forecast predictions even when they might suffer from the double-penalty problem if traditional verification tools were used. One of the proposed alternatives for addressing possible double-penalty issue is to assess the conventional skill scores (e.g., equitable threat scores and biases) using adjacent grid points. We did perform this procedure and noted an increase of skill, which again proves the existing double-penalty problem. However, we found that such a use of adjacent grid point was not necessary since the combination of

downscaling and the superensemble did provide superior skills compared to all member models.

The phase errors in our approach are reduced from a contribution of a statistical downscaling and a multimodel superensemble algorithm, both of which are designed on the principle of least squares minimization of errors. An objective of the study is to show that current large-scale precipitation forecasts can be much improved from the use of downscaling of the rainfall forecasts and from the construction of the multimodel superensemble.

## 2. High-resolution gridded daily observed rain gauge datasets over India

This is a special precipitation dataset that was prepared by the National Climate Data Center, India Meteorological Department (IMD), Pune, India. There were two versions of the datasets, one covering 1803 rain gauge (version 1) sites, and the other covering 2140 sites (version 2). They provide daily rainfall total from 1950 to the present. Figure 1 illustrates the rain gauge stations (version 1) across
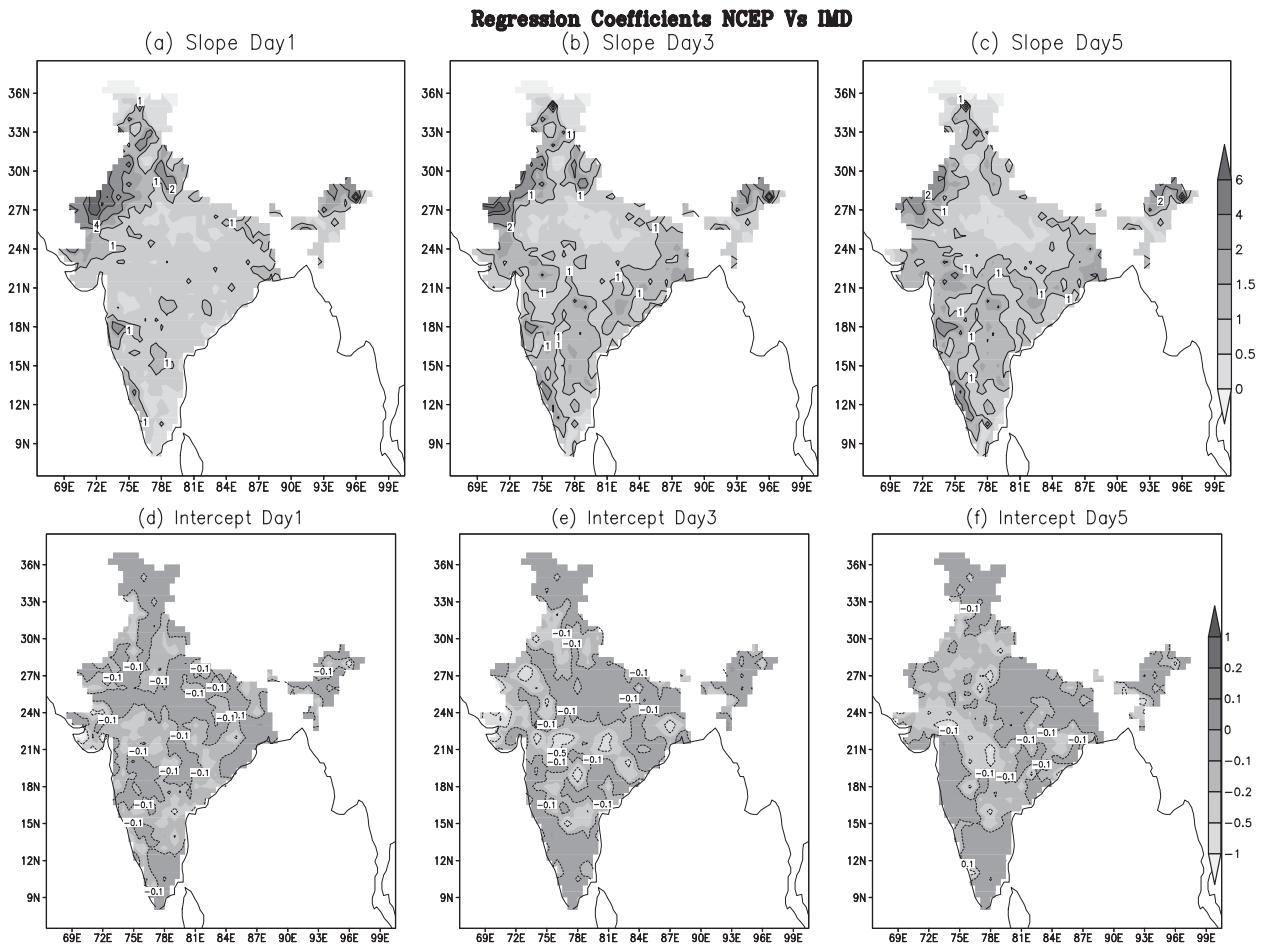
Regression Coefficients NCEP Vs IMD



FIG. 4. Regression coefficients (a)–(c) *a* and (d)–(f) *b* of the downscaling equation [Eq. (1)] for days 1, 3, and 5 of forecasts from the NCEP global forecast model.

India. These datasets were interpolated (Shepard 1968) on a 0.5° × 0.5° latitude–longitude grid (Rajeevan et al. 2006) for research purposes. A separate dataset was also prepared on a 1° × 1° latitude/longitude grid that we have used for seasonal climate forecasts, presented in Part II. Figure 2 shows the station difference between version 2 (V.2) and version 1 (V.1) (i. e., V.2 − V.1) in each grid box (1° × 1°). Rain gauge networks in V.1 have very few rain gauges over the Indo-Gangetic plains while the southern peninsula is densely covered (Fig. 1). The version 2 dataset at 0.5° × 0.5° resolution is used for this study. Differences (in numbers) shown in Fig. 2 are for each grid box (1° × 1°), where positive numbers in each 1 × 1 grid box represent the addition of a rain gauge in that box. It is interesting to note that over the Indo-Gangetic plains, version 2 data has the advantage of more rain gauges. There are some small negative numbers too mainly on the southern part of the country, where the network of the rain gauge was already very dense. This dataset is considered to be one of the most comprehensive datasets for this region.

The precipitation estimated by the rain gauge is influenced by the local effects of orography, surface emmisivity, surface albedo, local lakes and water bodies, and local vegetation. The downscaling of the model rains toward the rain gauge rains implicitly attempt to correct these effects. The total number of grid points over the Indian main land is 1251. Data available in a 0.5° × 0.5° grid box are on these 1251 grid points. For a 5-day forecast we need to carry out computations 5 times at each of these grid points. For this reason we carry out as many as 6255 downscaling algorithm. The frequency of this dataset is once per 24 h. It might have been desirable to have these datasets at more frequent intervals, which was an issue that came up in this study; this is a limitation of this dataset. We shall next discuss the multimodel superensemble methodology.

## 3. Multimodel conventional superensemble

The notion of the multimodel superensemble for weather and seasonal forecasts was first proposed by

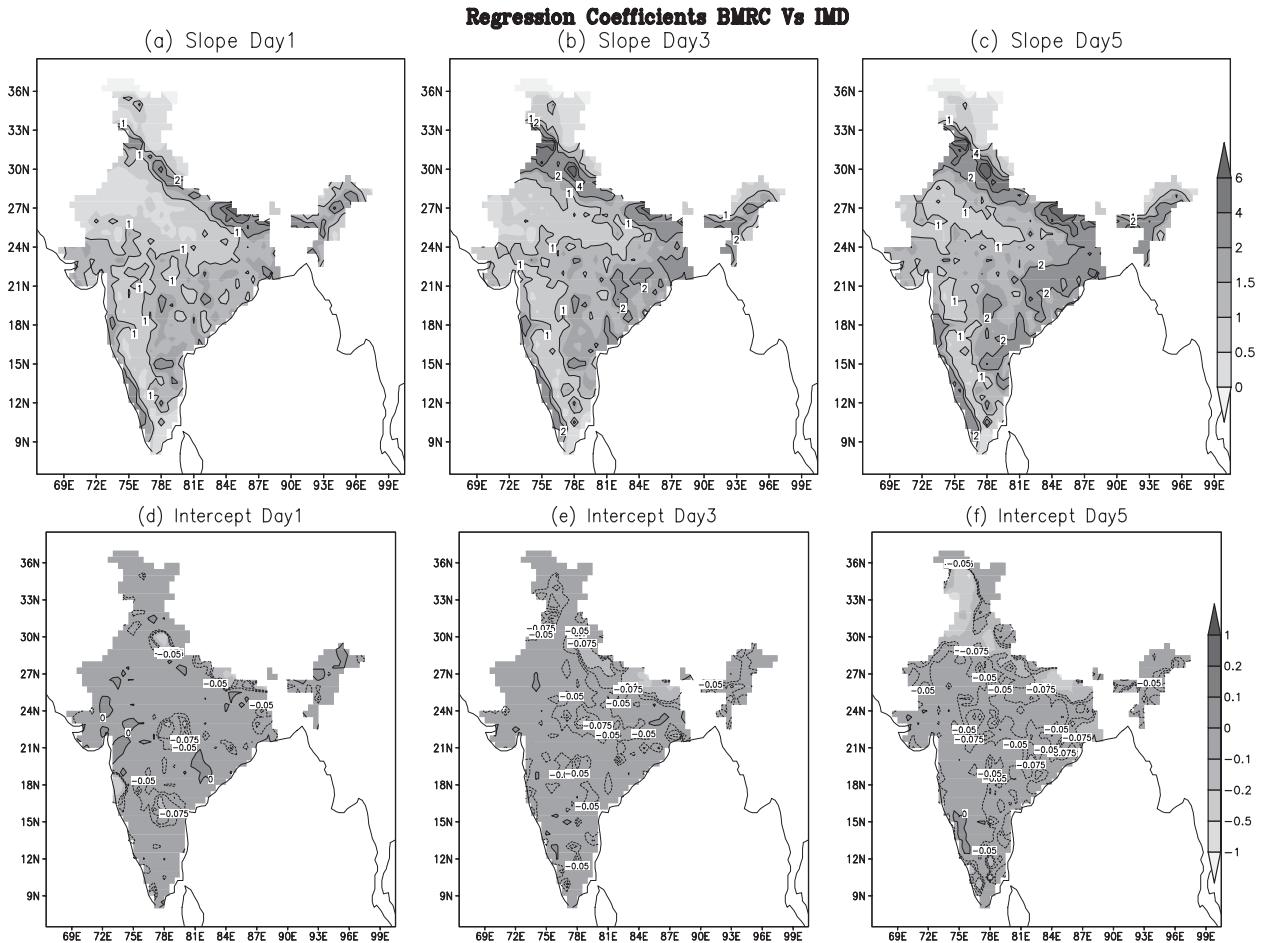**Regression Coefficients BMRC Vs IMD**



FIG. 5. As in Fig. 4, but for the BMRC global forecast model.

Krishnamurti et al. (1999). This method is based on producing a weighted average of model forecasts to construct a superensemble forecast. This procedure carries two phases: training and prediction. During the training phase past forecasts from a number of member models and the corresponding observed (analyzed) fields are used. The training entails determining statistical weights for each grid location in the horizontal, at all vertical levels, for all variables, for each day of forecasts and for each of the member models. For global NWP the procedure brings in as many as $10^7$ statistical weights. These weights arise from a statistical least squares minimization using multiple regressions, where the member model forecasts are regressed against the observed (analyzed) measures. The outcome of this regression is the weights assigned to the individual models in the ensemble, which are then passed on to the forecast phase to construct the superensemble forecasts.

The temporal anomalies of a variable, rather than the full fields, are used in the multiple regression technique. Hence, in formulating the superensemble forecast, the weights are multiplied to the corresponding model anomalies. The constructed forecast is

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

where $\overline{O}$ is the observed climatology; $a_i$ is the weight for the $i$th member in the ensemble; and $F_i$ and $\overline{F_i}$ are the forecasts and forecast climatological values for the training period, respectively, for the $i$th model's forecast. The summation is taken over the $N$ member models of the ensemble. The weight $a_i$ are obtained by minimizing the error term $G$, written as

$$G = \sum_{i=1}^{N_{\text{train}}} (S_i' - O_i')^2, \qquad (2)$$

where $N_{\text{train}}$ is the number of time samples in the training phase, and $S_t'$ and $O_t'$ are the superensemble and observed field anomalies, respectively, at training time $t$.

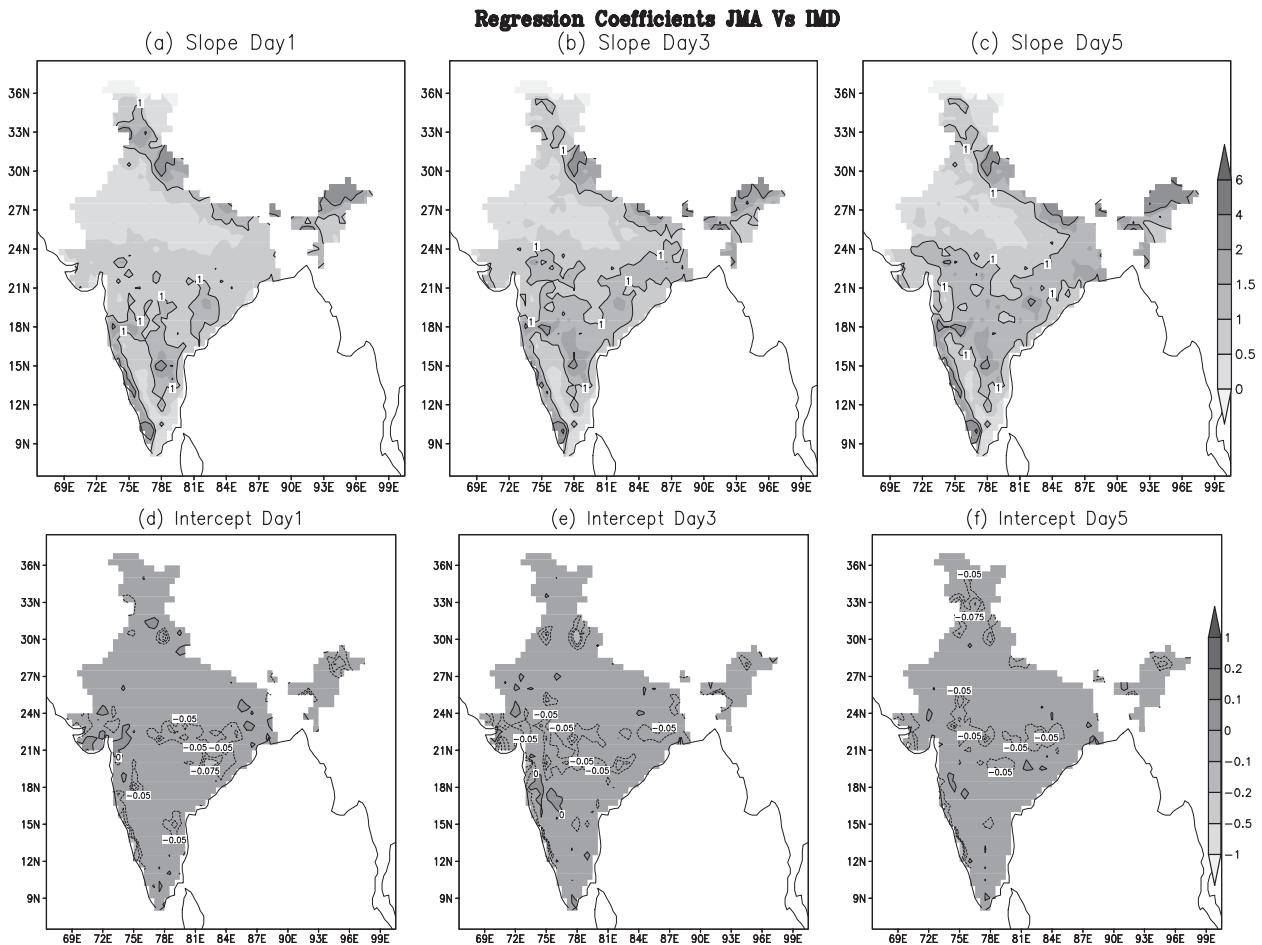## Regression Coefficients JMA Vs IMD



FIG. 6. As in Fig. 4, but for the JMA global forecast model.

This exercise is performed for every grid point and vertical level in the dataset during every forecast phase. In other words, one weight is given to every model at every grid point in the three-dimensional space for each forecast.

Figure 3 provides a schematic outline of the superensemble strategy for the construction of the multimodel superensemble. The method has been applied most recently to improve large-scale NWP forecasts of the monsoon (Mishra and Krishnamurti 2007), hurricane track and intensity forecasts (Krishnamurti et al. 1999, 2000a; Williford et al. 2003; Vijaya Kumar et al. 2003), and seasonal climate forecasts (Krishnamurti et al. 2000b, 2006a; Chakraborty and Krishnamurti 2006). In these studies a common result has been that the multimodel superensemble consistently provides superior forecasts, in terms of skills scores, compared to the participating member models. This, however, does not always assure the usefulness of such forecasts to meet the needs of the user community.

## 4. Member models of NWP

The member models of the present study include the National Centers for Environmental Prediction (NCEP; United States), the European Centre for Medium-Range Weather Forecasts (ECMWF; Europe), the Bureau of Meteorology Research Centre (BMRC; Australia) and the Japan Meteorological Agency (JMA; Japan). Our past experience has shown that four–eight member models can provide enough information for the construction of a multimodel superensemble (Krishnamurti et al. 2000a; Mishra and Krishnamurti 2007). If a single model is subjected to multiple forecasts using perturbed initial states then one may need as many as 25–50 realizations for providing improved forecasts from an ensemble mean (Palmer et al. 1993; Toth and Kalnay 1993). According to Leith (1974) most of the improvement in the ensemble mean is achieved with 8–10 members, whereas ~30 members would be required to estimate second-order statistics. We have noted that as
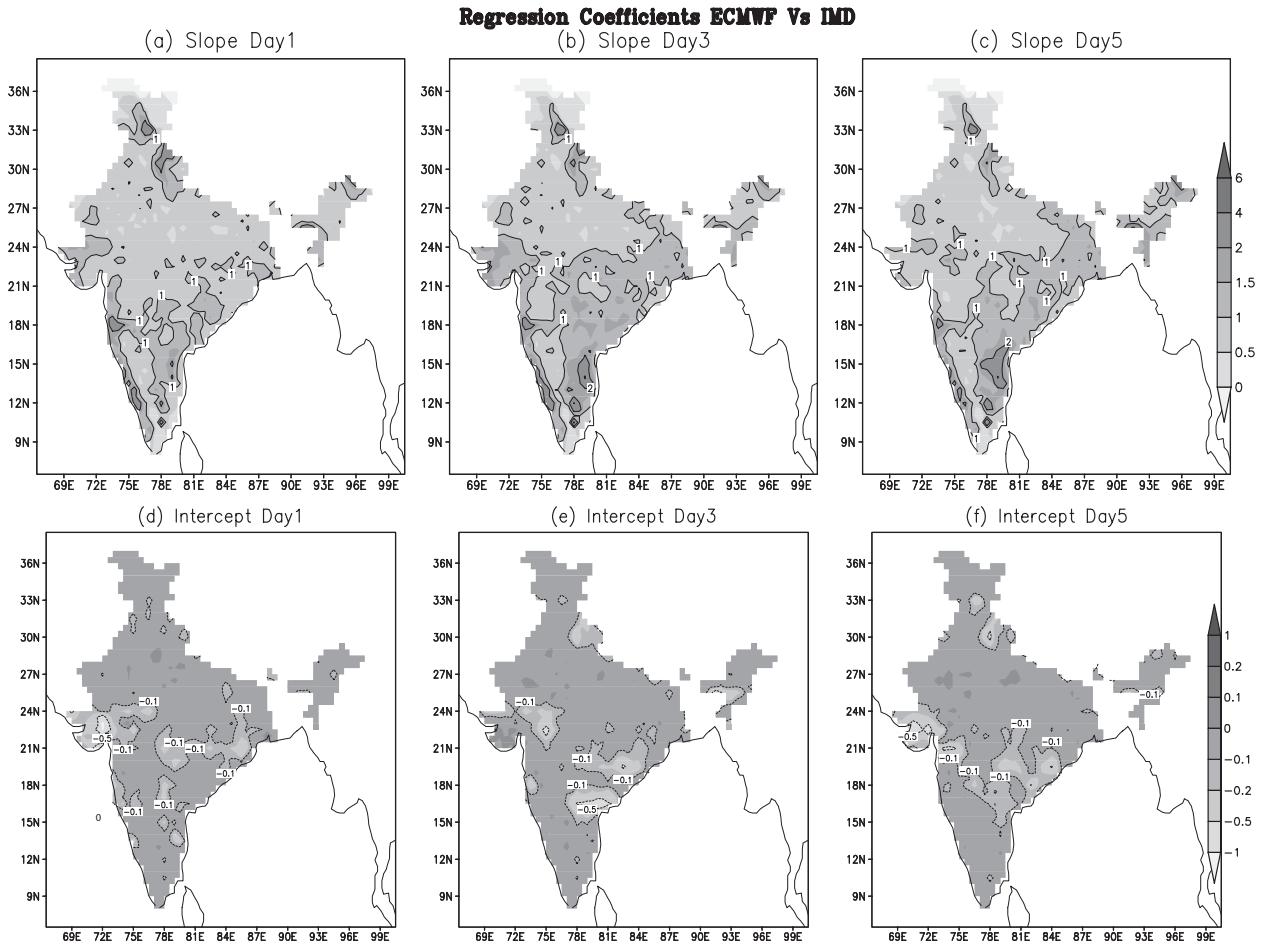
## Regression Coefficients ECMWF Vs IMD



FIG. 7. As in Fig. 4, but for the ECMWF global forecast model.

few as four or five better-performing member models, in terms of skill, can be used to construct a superensemble that carries higher skills than the member models (Krishnamurti et al. 2006b).

## 5. Statistical downscaling

Given the forecasts of precipitation from a number of forecast models with horizontal resolutions of the order of 100 km, our downscaling for model precipitation follows three steps.

1) A simple bilinear interpolation of the model daily rain to the grid points of the IMD rain ($0.5° \times 0.5°$) is performed. This is done for each day of forecast for each model. Where "daily rain" refers to 24-h precipitation accumulation between 1200 and 1200 UTC the next day.

2) A time series of the interpolated rain is made for each model at every grid point and for each day of forecast separately (i.e., the string of day-1 forecasts).

The same procedure is followed to generate strings for the day-2, -3, -4, and -5 forecasts. For each forecast lead time we have a string of high resolution, rain gauge–based rainfall observations. This provides an observational string.

3) The downscaling strategy involves a linear regression of the time series of the data at each grid point:

$$Y_i = aX_i + b, \qquad (3)$$

where $X_i$ are the rainfall forecasts (separately handled for each day and that had been subjected to bilinear interpolation), and $Y_i$ are the observed counterparts.

The training period for our study is from 1 June 2007 to 31 August 2007. Based on 3-month training statistics, superensemble forecasts for the month of September are prepared. We utilize nearly 92 forecasts during the training phase of the superensemble to generate separate coefficients for the different member models. This is similar to our recent experience (Krishnamurti et al.
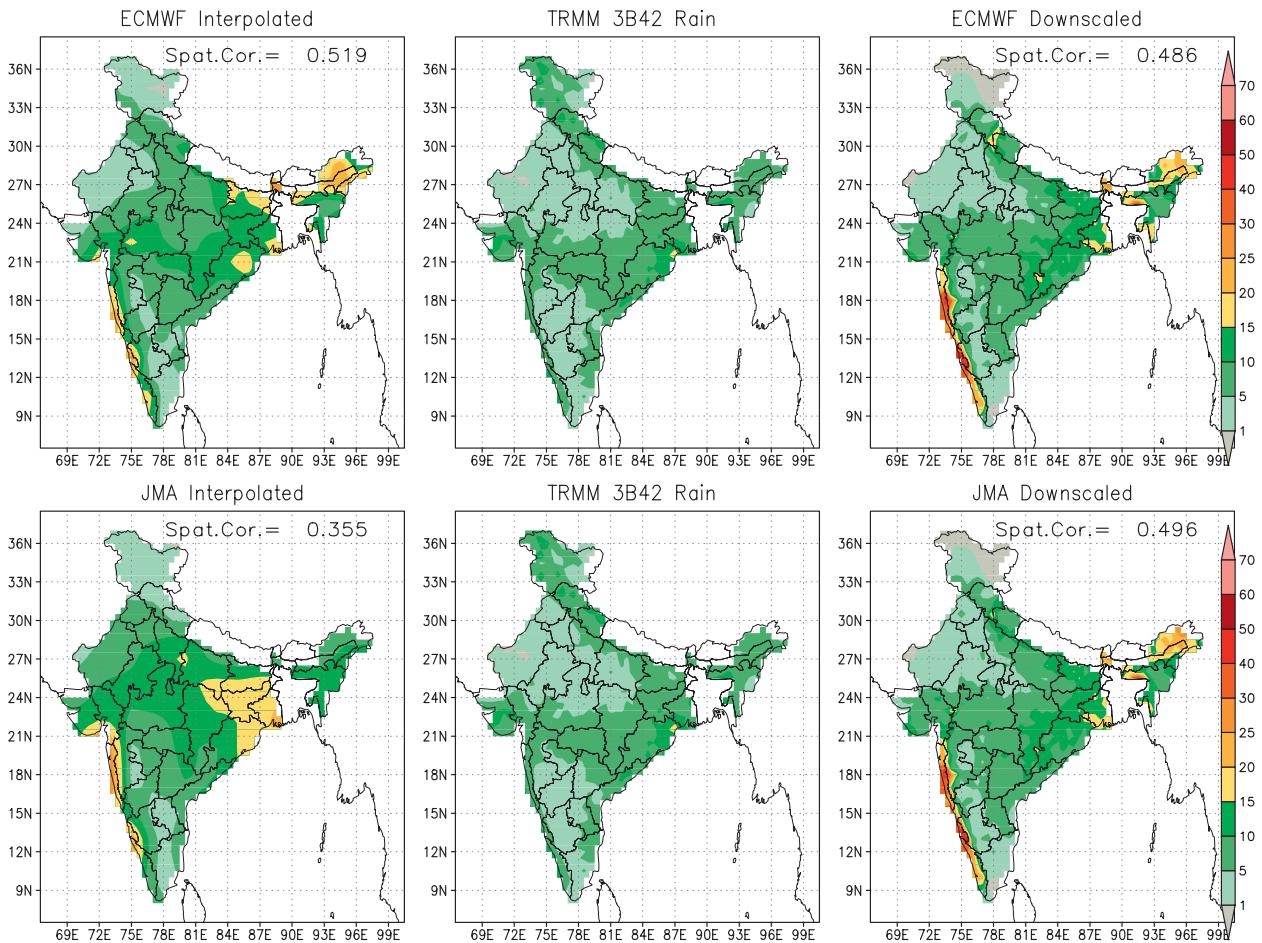
FIG. 8. Climatology (June–August 2007) of day-1 forecasts of precipitation (mm day$^{-1}$) from (top) ECMWF and (bottom) JMA (left) after interpolation and (right) after downscaling. The correlations between forecasts and observations are indicated at the top-right corner of the respective panels.

2006b). These are calculated separately for each grid point of the IMD domain. The coefficients $a$ and $b$ thus vary from one grid location to the next and are also model dependent. The distribution of $a$ and $b$ provides useful information on the precipitation forecast bias of the member model. This exercise provides a bias-corrected rainfall product for each of the member models for each day of forecast. If one downscales the model-predicted precipitation alone, then the downscaled product would generally be an amplified version of the model's precipitation. However, the superensemble that is next constructed using downscaled model forecasts takes it much further.

## 6. Geographical distribution of regression coefficients

The geographical distribution of the regression coefficient conveys information on the bias of the rainfall

forecasts of the member model. For the downscaling regression equation ($Y_i = aX_i + b$), $a$ denotes the ratio of the observed to the modeled rains for different intensities of rain, and $b$ denotes the intercept that conveys underestimates (or overestimates) for the overall model forecast rain depending on its positive or negative values, in other words the slope coefficient $a$ is a measure of the multiplicative bias if the systematic bias $b$ is removed. Downscaling projects the weaker (bilinearly interpolated) rain of large-scale models on to the grids of the higher-resolution rainfall observations and corrects for the slope and intercept biases. These fields are illustrated in Figs. 4–7 for the NCEP, BMRC, JMA, and ECMWF models for forecast lead times of days 1, 3, and 5. Because the nature of the error growth is different for different models and for different days (length of forecasts), the fields of $a$ and $b$ are somewhat different for each member model geographically. The coefficients $a$ and $b$ represent the model biases, $b$ is the constant bias,
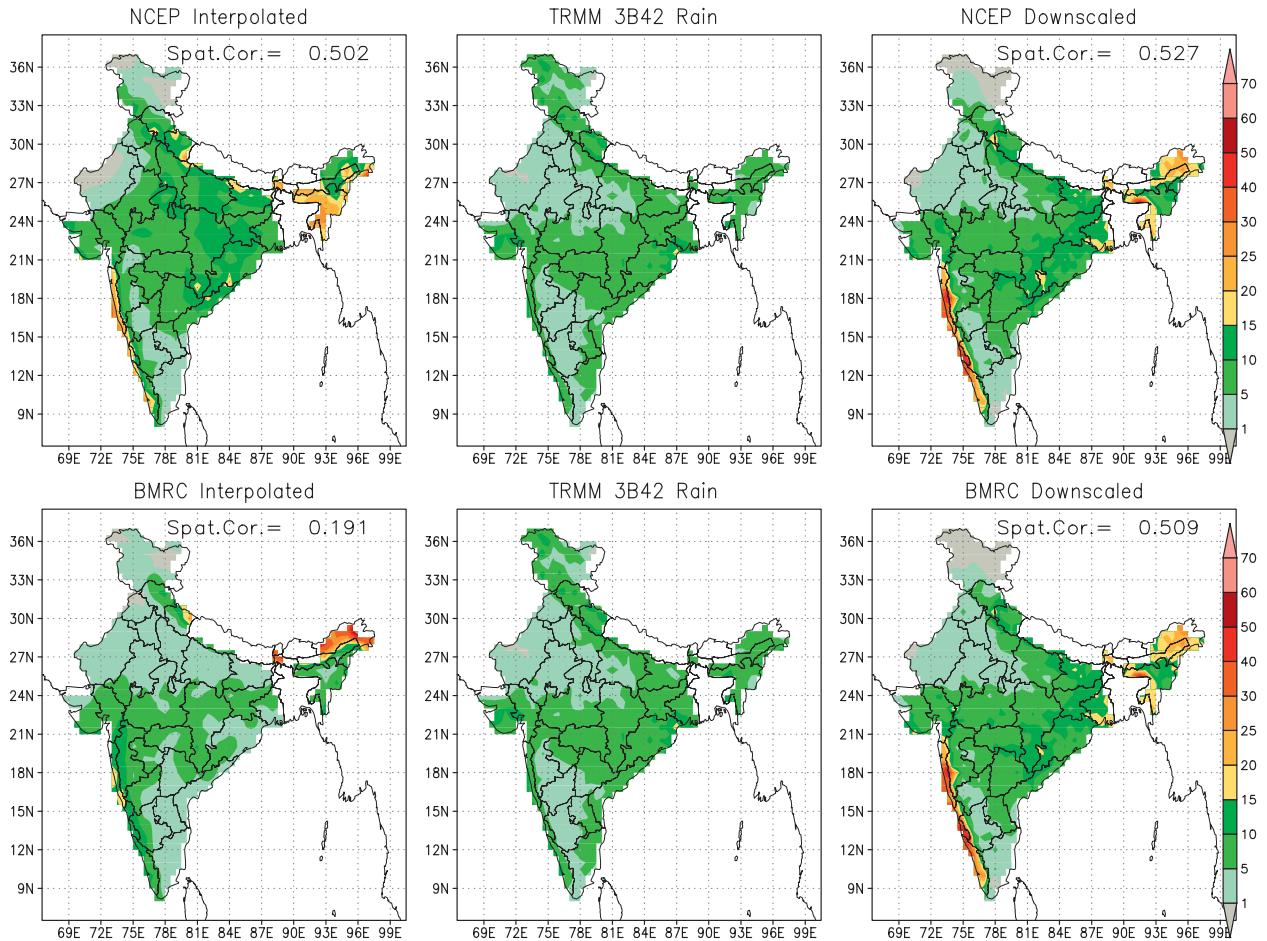
FIG. 9. As in Fig. 8, but for day-5 forecasts.

and $a$ is a bias based on precipitation intensity/amount. Large values of $a$ indicate that the distribution of precipitation intensity in the model is too flat (i.e., there is not enough heavy precipitation or too much light precipitation, or both). High positive values of $a(a > 1)$ denote the model underestimating the rain, while low positive values $(0 < a < 1)$ represent the model overpredicting the rain. Negative values of $a$ denote that the model overpredicts for lower values of observed precipitation and underpredicts for higher values of observed precipitations. If the value of $b$ tends to zero [Eq. (3)], observed precipitation becomes a function of model forecast. A higher value of $b$ denotes the large systematic error in the model. For the NCEP model slope values (i.e., coefficient $a$) lie between the values 0 and 1 mostly, these are the area where the model is overpredicting. There are, however some regions where the slope is greater than 2 (e.g., the northwest, west coast, and northeastern regions). Regions where the model underestimates the rain increase in area by day-5 forecasts. The intercept of the NCEP model lies largely

between 0 and $-0.4$. In these regions the model rain has a slight systematic error. Both the slope and the intercept are adjusted toward a minimum bias from the downscaling. Over some regions these two coefficients try to compensate for each other. The BMRC model forecasts regression statistics (Fig. 5) carry a reasonable slope between 0 and 2 over many regions of India. However, this forecast model has a tendency to overestimate the rainfall totals. Noteworthy are regions along the southern slopes of the Himalayas, where the rainfall is overestimated by factor of 7–10. The regression coefficient $a$ in the BMRC forecasts are, however, generally small ($<5$) except near the southern slopes of the Himalayas where these coefficients increased with duration (or length) of forecasts. Figure 6 shows that the behavior of the JMA model is somewhat different from the NCEP/Environmental Modeling Center (EMC) model. Large areas of India north of 24°N show small slopes, of the order of 0–0.5. These imply large rains for the model forecasts for days 1–5 compared to the observed estimates of rain. In the coastal region of
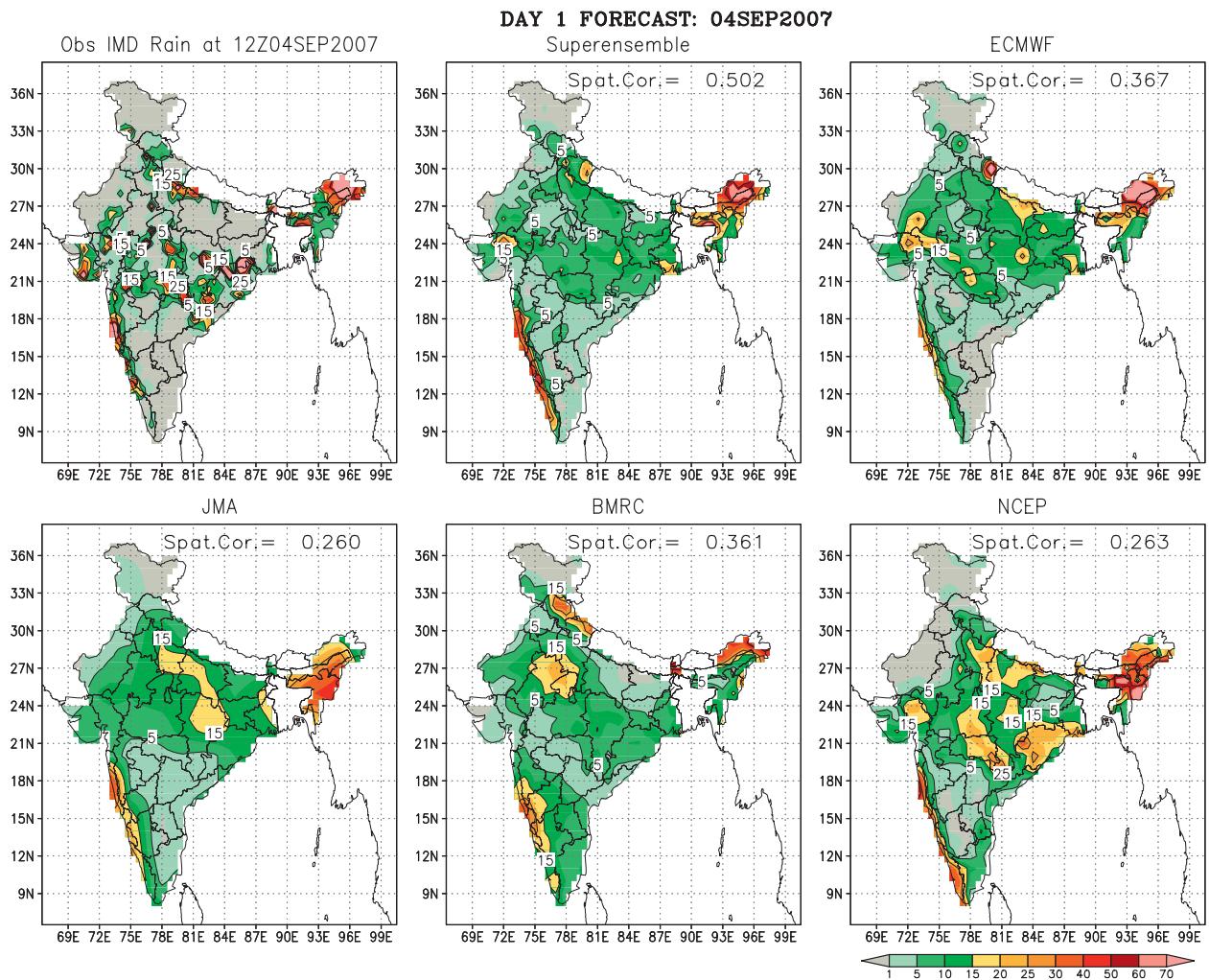
FIG. 10. Observed and forecasted precipitation (mm day$^{-1}$) from superensemble, ECMWF, JMA, BMRC, and NCEP models for day-1 forecasts valid on 4 Sep 2007. The correlations between forecasts and observations are indicated at the top-right corner of the respective panels.

southwest India large slopes ($>1$) are noted where the JMA model underestimates the rain considerably. The intercept values lie between $-0.25$ and 1 and are small compared to the slope coefficients. Over most regions the intercept error is of the order of 0.05. The ECMWF model (Fig. 7) largely carries slopes of the order of 0.5 to 1.5 (i.e., close to 1.0 over most of India and this implies a very small bias). Large slope values ($\approx 2$–3) are seen for the ECMWF model over the southwest coast of India where the model underestimates the rain. Forecast errors of the ECMWF model are also somewhat large over southeast India for day-3–5 forecasts. Here the model underestimates the rain by a factor of 3–3.5. The intercept coefficients of the ECMWF model are generally very small except over central India (along 18°–23°N) where the model systematically overestimates the rain, and these coefficients seem to grow with time in day-3

and -5 of forecasts. This is a region on the western slopes of the Eastern Ghats of India where the sea breeze of the Bay of Bengal and the southwest monsoon show a convergence contributing to the seasonal rains, and the ECMWF model overestimates these.

## 7. Seasonal averages for strings of day 1 and day 5 of forecasts

On comparing the observed (IMD) data with downscaled forecast we found that the climatology of the downscaled rain forecasts for days 1–5 for each of the member models was very close to the observed estimates (figure not shown). We have used the Tropical Rainfall Measuring Mission (TRMM) 3B42 rain as an independent source of observed rain to compare the model interpolated and downscaled forecasts. These
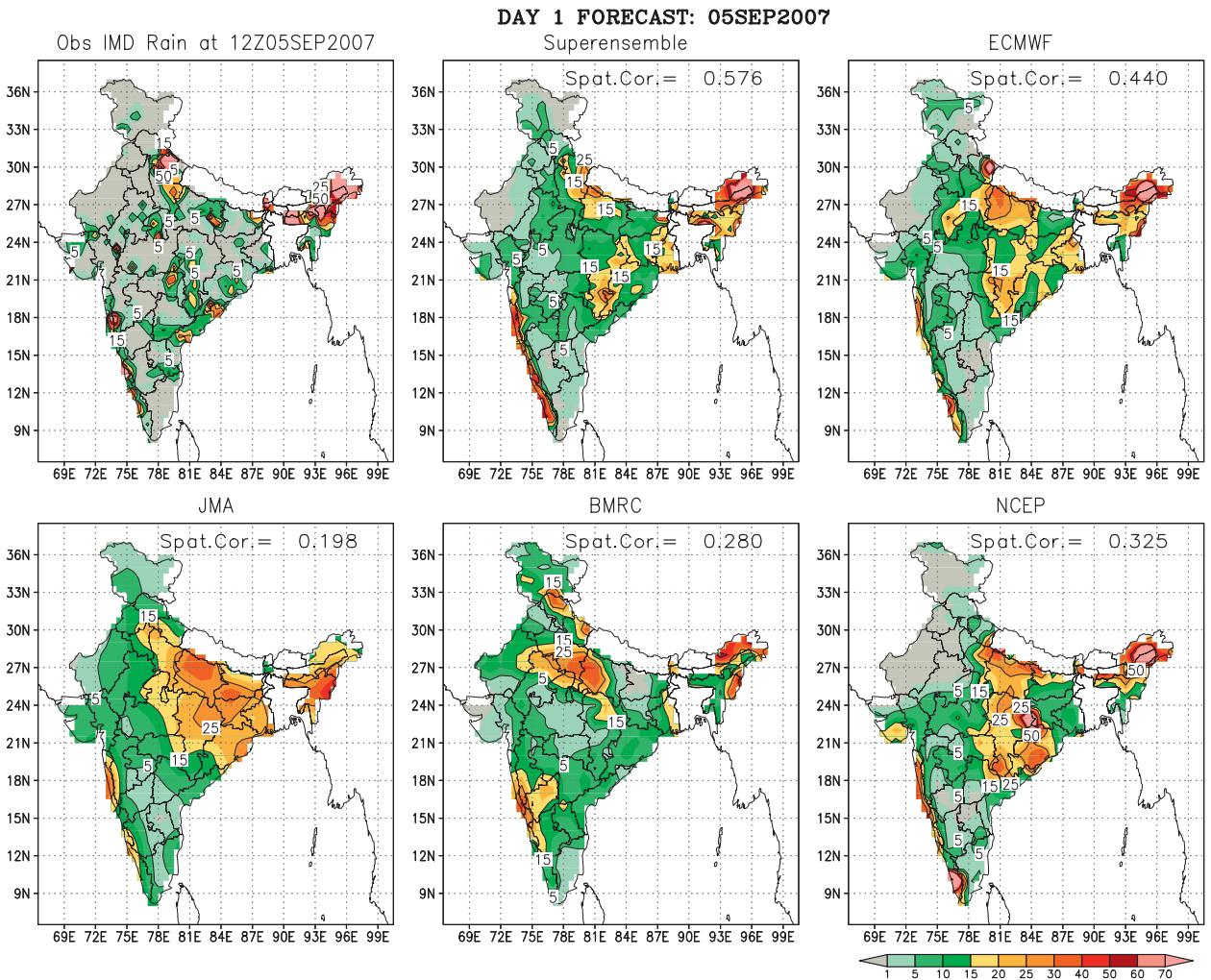
FIG. 11. As in Fig. 10, but for day-1 forecasts valid on 5 Sep 2007.

carry spatial correlations of the order of 0.50 for all of the models. Those scores (Fig. 8 for day-1 forecasts), for spatial correlation for the large-scale model forecasts prior to downscaling were 0.35 for the JMA and 0.51 for the ECMWF model. Those for NCEP and BMRC were 0.54 and 0.24, respectively (not shown). The spatial distribution of seasonal rainfall climatology for June, July, and August 2007 shows improvement from the downscaling. That algorithm recovers features of the observed seasonal rainfall climatology. Some of the large errors in the climatology of the large-scale model rains, seen in the left two panels of Fig. 8, were removed by the downscaling. Those scores for day-5 forecasts are shown in Fig. 9, where again we note an improvement in spatial correlation for rainfall climatology from 0.50 to 0.52 for the NCEP model and 0.19 to 0.50 for the BMRC model. These improvements were also noted for the day-5 forecasts of the ECMWF and JMA models (not

shown here). Overall the proposed downscaling concept is a very robust procedure for the model, which works even for day-5 forecasts.

## 8. Predicted precipitation fields

The geographical distribution of forecasts compared to the observed precipitation fields for individual days is the most important aspect of this study. We illustrate several examples for days 1 and 5 of forecasts in Figs. 10–15. In each of these illustrations we include the observed rainfall, those based on the downscaled superensemble and those from the four member models: ECMWF, JMA, BMRC, and NCEP. The observed and the downscaled predicted rainfall are all presented at the resolution of $0.5° \times 0.5°$ latitude–longitude. Figure 10 illustrates the results for day-1 forecasts, ending at 1200 UTC 4 September 2007. We also include values of spatial
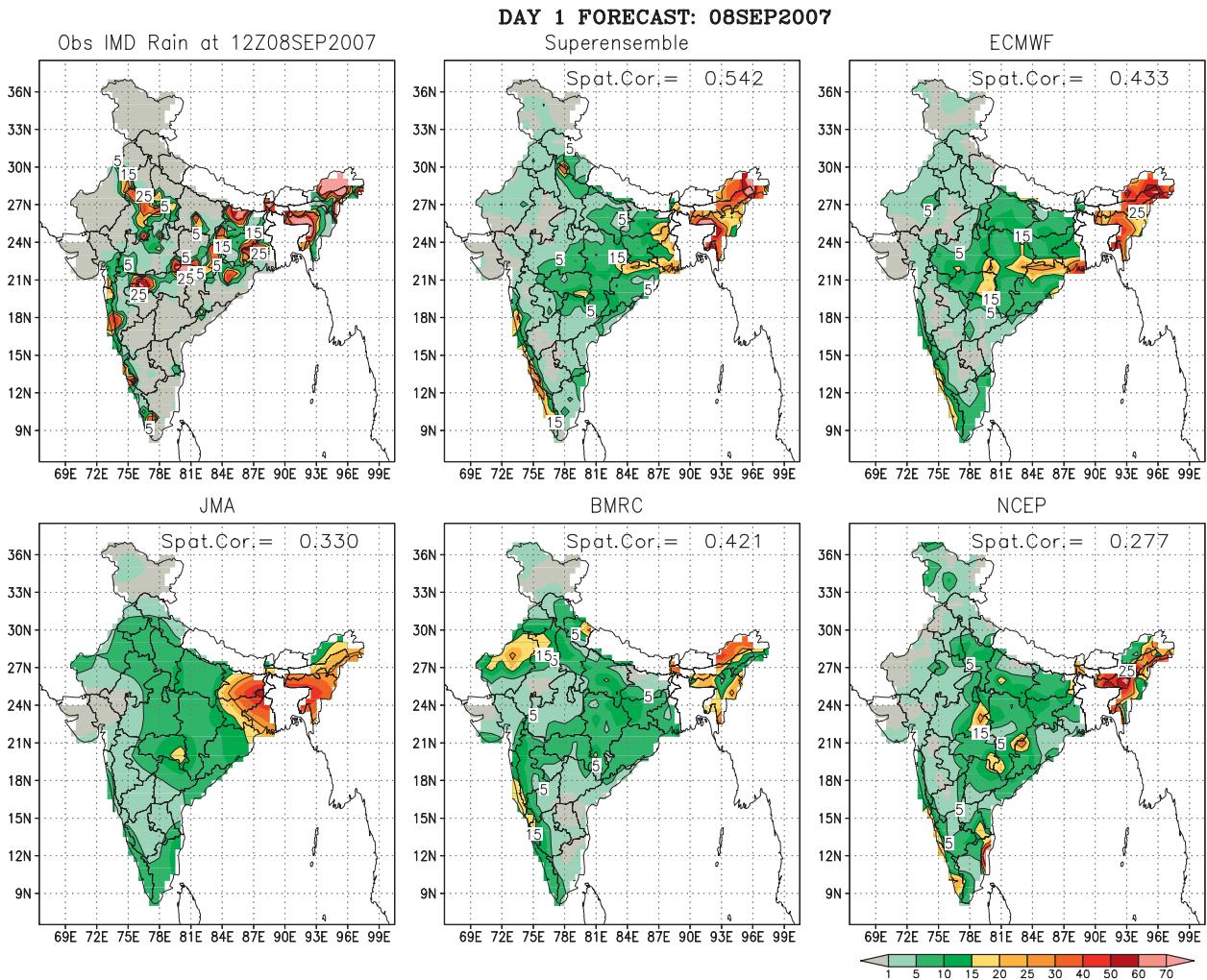
FIG. 12. As in Fig. 10, but for day-1 forecasts valid on 8 Sep 2007.

correlation at the top of each panel. The observed rainfall clearly shows many smaller-scale features of heavy rains that are somewhat smoothed by the member models, as well as by the superensemble. This suggests the possible need for a still higher resolution for the downscaling.

The linear regression within the downscaling evidently contributes to some of this smoother representation of the forecasts. It is however of considerable interest to note that the member models carry spatial correlations 0.36, 0.26, 0.36, and 0.26 whereas the downscaled superensemble is able to enhance that correlation to 0.50 for that forecast (Fig. 10). The JMA, BMRC, and NCEP models carry wide spread rain in excess of 15 mm day$^{-1}$ over central India, and these contributed to lower skill scores for these models. The heaviest observed rain was located along 21°N latitude. Some of the models carried the heaviest rain too far

north (i.e., 27°N as is seen for the BMRC model). Similar latitudinal bias and spread were seen for the JMA model. The second example presented here (Fig. 11) is for 5 September 2007, where the downscaled superensemble provides a spatial correlation of 0.58 compared to those of the member models whose values are 0.44, 0.19, 0.28, and 0.32. Overall we noted a consistency in the downscaled superensemble skills that were always much higher for the spatial correlation as compared to the skills of the member models. Figure 12 illustrates a third example, where again we note similar improvements for precipitation forecasts from the downscaled superensemble. Here member models carry the correlation coefficients 0.43, 0.33, 0.42, and 0.27, while superensemble stands out highest among all with spatial correlations of 0.54. The entire month of forecasts for each day was examined in a similar manner and we noted very similar improvement in skills for the downscaled
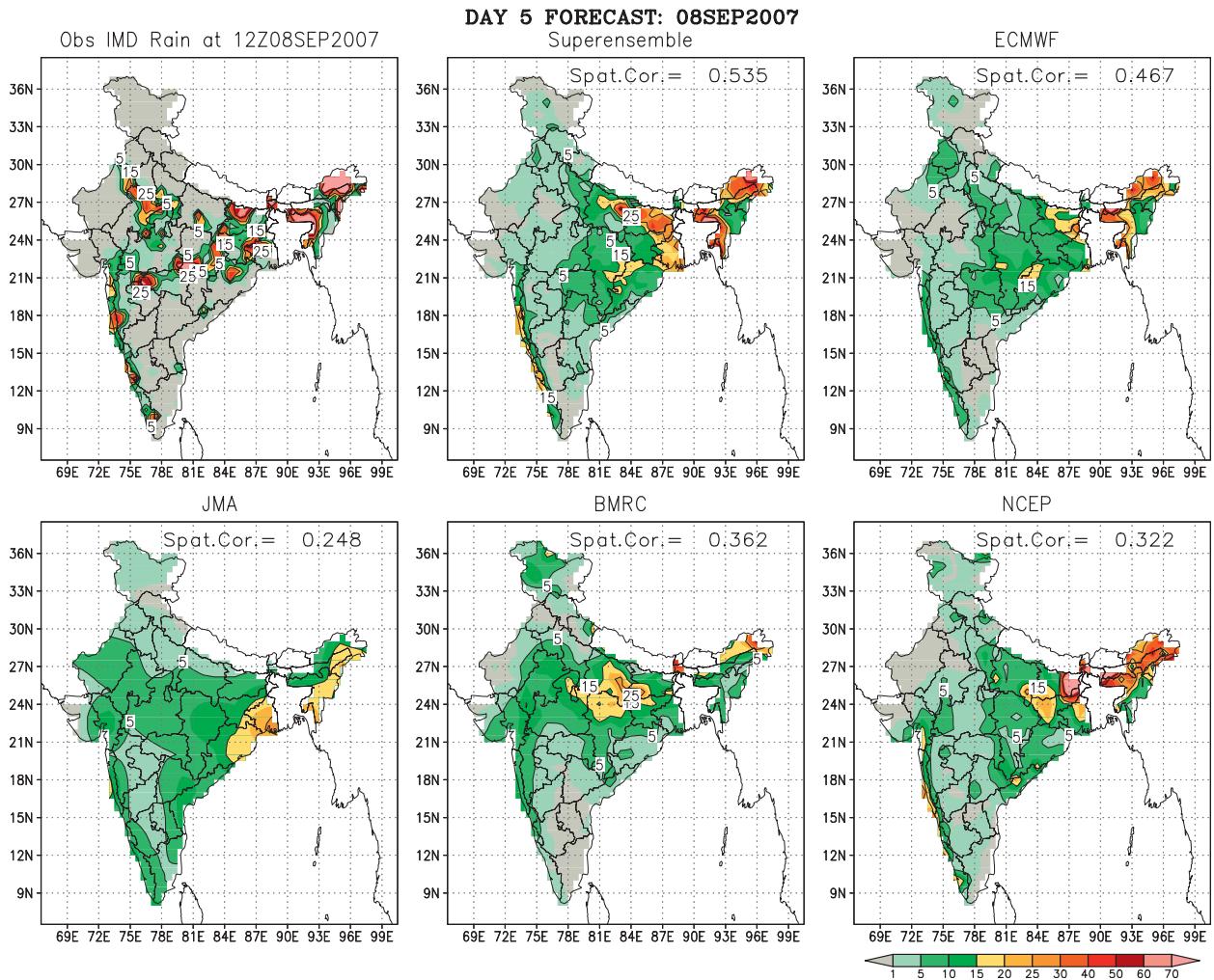
**DAY 5 FORECAST: 08SEP2007**



FIG. 13. As in Fig. 10, but for day-5 forecasts valid on 8 Sep 2007.

superensemble. We also noted that the results of comparison of skills for day-5 forecasts were quite similar to those for day 1 of the forecasts.

Figures 13–15 include examples of heavy rains over eastern India, south-central India, and central India. For day-5 forecasts in each of these cases, the improvements in the spatial correlation (shown as an inset) of the multimodel downscaled superensemble is much higher as in the previous cases. The presented detail on rainfall distribution for day-5 forecasts clearly suggests that the proposed procedure can have useful operational values. On 23 September 2007, the day-5 forecast (Fig. 15) of ECMWF is found to be slightly better than the superensemble; however, the day-5 forecast of ECMWF depicts higher rain amounts (>25 mm) in the Jammu and Kashmir regions of northern India, whereas in the observation only a trace rain amount (<5 mm) was noticed, the same was true for the superensemble.

The drop in skill for the rainfall forecast (at 0.5° × 0.5° latitude–longitude resolution) from day 1 to 5 of the forecasts was very small for the downscaled superensemble. In these examples the day-1 skill for the pattern correlation for the superensemble forecast of all-Indian rainfall was of the order of 0.5. The corresponding number for day-5 forecasts was 0.4. However, the values for the member models were as low as 0.01 for the day-5 forecast.

## 9. Bilinear interpolation versus downscaled precipitation

In Fig. 16 we show the improvements for day-1 forecasts from the downscaling as compared to the bilaterally interpolated rains of the coarse resolution of the individual forecast models. The spatial correlation covers the IMD domain of Fig. 2 (right panel) and represents the all-India rainfall. These illustrations clearly show a
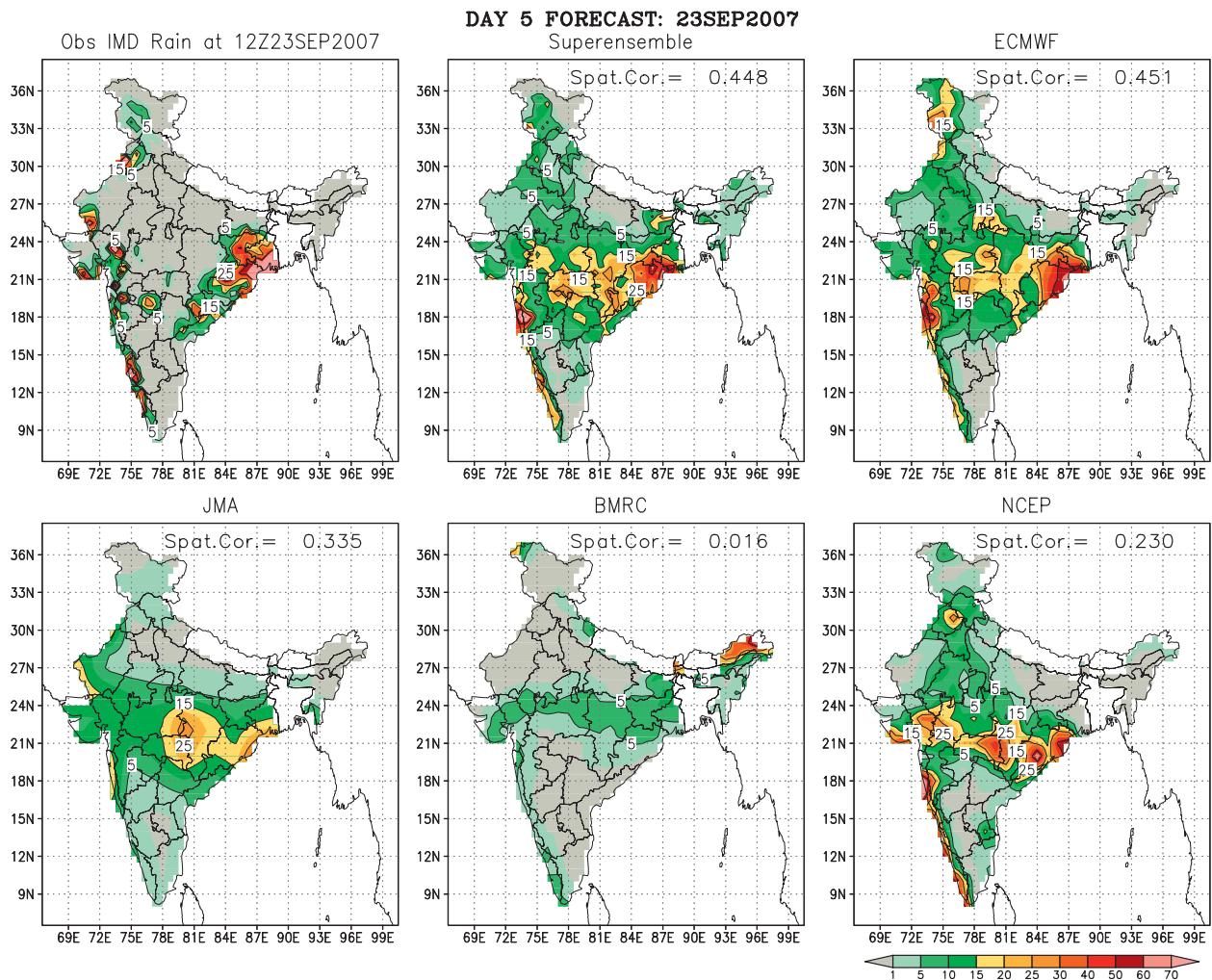
FIG. 14. As in Fig. 10, but for day-5 forecasts valid on 23 Sep 2007.

marked improvement in the spatial correlation in a rather consistent manner for all-India rain. These improvements in correlation were between 10% and 20%. Figure 17 shows the improvements for day-5 forecasts. The spatial correlation of the downscaled all-India rain is clearly 10%–20% higher than the bilaterally interpolated forecast rains. The validations are assessed with respect to the high-resolution observed rain.

## 10. Equitable threat scores

In this section we compare the performance of model skills for rainfall forecasts for the large-scale and the downscaled models using the equitable threat score (ETS; see the appendix). Figure 18 shows the equitable threat scores for the month of September 2007 for day-1–5 forecasts. The equitable threat scores and bias scores (see the appendix), prior to and subsequent to the

downscaling, show a very marked improvement for all of the member models. Here the results of forecasts for days 1–5 for the IMD domain (Fig. 2) are shown. The threat scores (ordinates) improve at all ranges of rainfall rates (abscissa) for all of the member models. The bias scores approach 1.0 for all of the models. A bias score of 1.0 is a perfect score. For low thresholds the bias errors of member models are higher than 1.0, and these were corrected toward 1.0 by the downscaling. The converse is the case for higher thresholds (or heavier rains only) where the bias of member models was less than 1 and was raised toward 1.0 by the downscaling. The largest improvements were noted from this procedure for the BMRC model that carried the lowest threat score prior to downscaling. In general when there is "no rain" or "very small trace rain" both in the model as well as the observations, the skill of the models are higher. Highest ETS skill is achieved for rainfall threshold 5 mm day$^{-1}$. It is
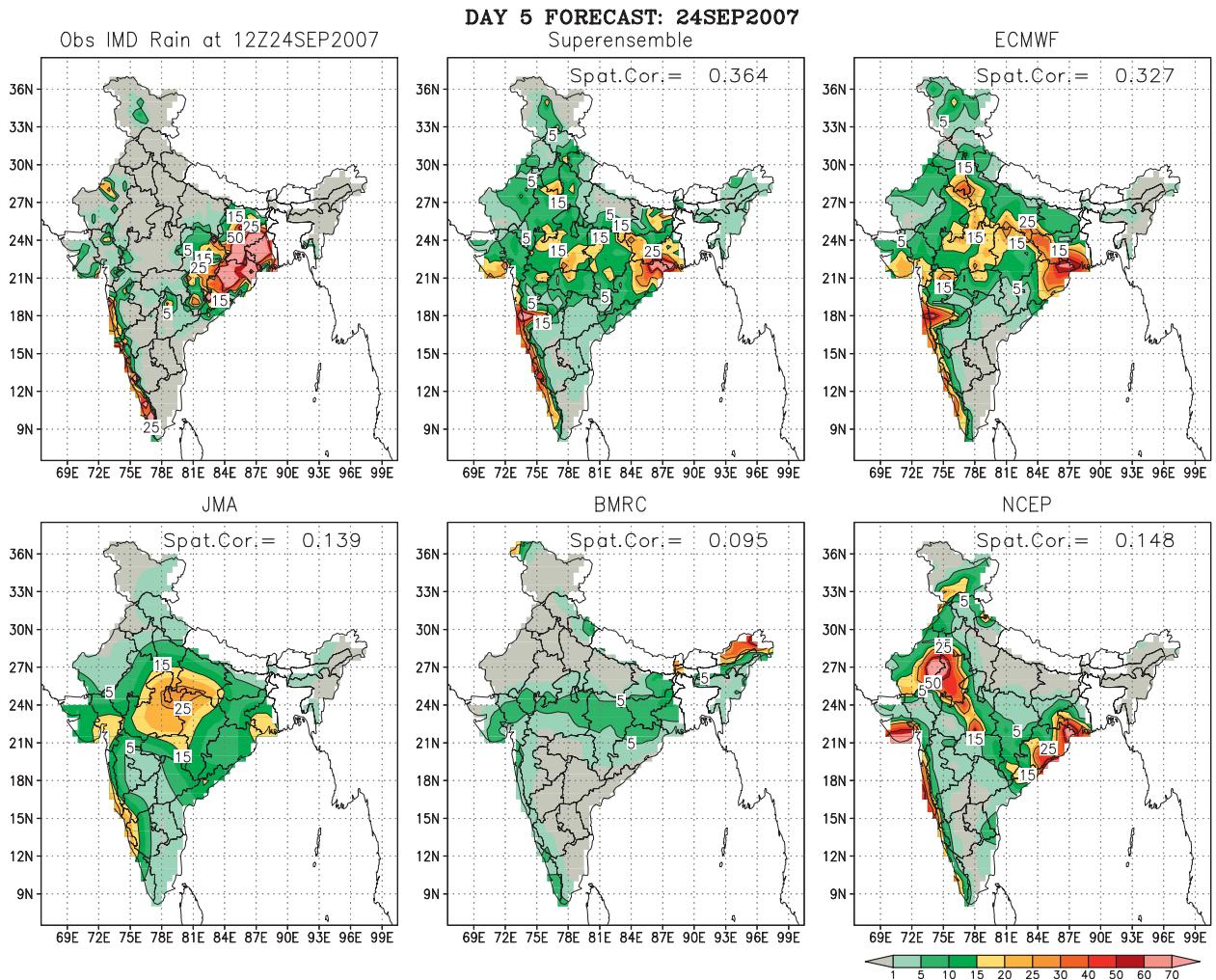
**DAY 5 FORECAST: 24SEP2007**



FIG. 15. As in Fig. 10, but for day-5 forecasts valid on 24 Sep 2007.

important to mention here that ETS and bias are calculated for a threshold value, which means that all rainfall exceeding this threshold are considered for the skill computations. All the models (Fig. 18) show improvement in the ETS for all thresholds after downscaling but there is an interesting point in bias scores; except for the BMRC model, downscaling is making models wet, especially for a smaller threshold. Day-1 forecast of ECMWF model shows a crossover of bias curves (between downscaled and interpolated forecast) at about 20 mm day$^{-1}$ threshold, however for day-2–5 forecasts this crossover point shifts to smaller threshold and is found to be almost constant at around 10 mm day$^{-1}$ threshold. ETS and bias skills of ECMWF's day-1 forecast (Fig. 18a) clearly depicts that downscaling is improving the ETS, at the same time it is also bringing the model (which is wet) closer to observations, however for high rain (flood events) the bias curve is coming up close

1.0, which means it will improve the skill of the model to forecast the heavy rain events. In other words, downscaling results in fewer grid points with light rain and slightly more grid points with very heavy rain, day-2–5 skills for ECMWF shows a different picture than day 1, here ETS is improving like day 1 but bias is having crossover point at lesser thresholds, it clearly shows that for the day-2–5 forecast ECMWF models improved drastically after downscaling for heavy rain but for total rain and or lower thresholds, improvement in bias is very little. The effect of the downscaling is focused on the day-1 forecast, where it shows the highest skills for total rain (and/or lower thresholds) while the greatest impact is on day-2–5 downscaled rain where skills of heavy rain forecast are better than the large-scale product. The NCEP model (Fig. 18b) does not show crossover point in the bias curve of the day-1 forecast, which means that downscaling decreases the rainfall for all the thresholds,
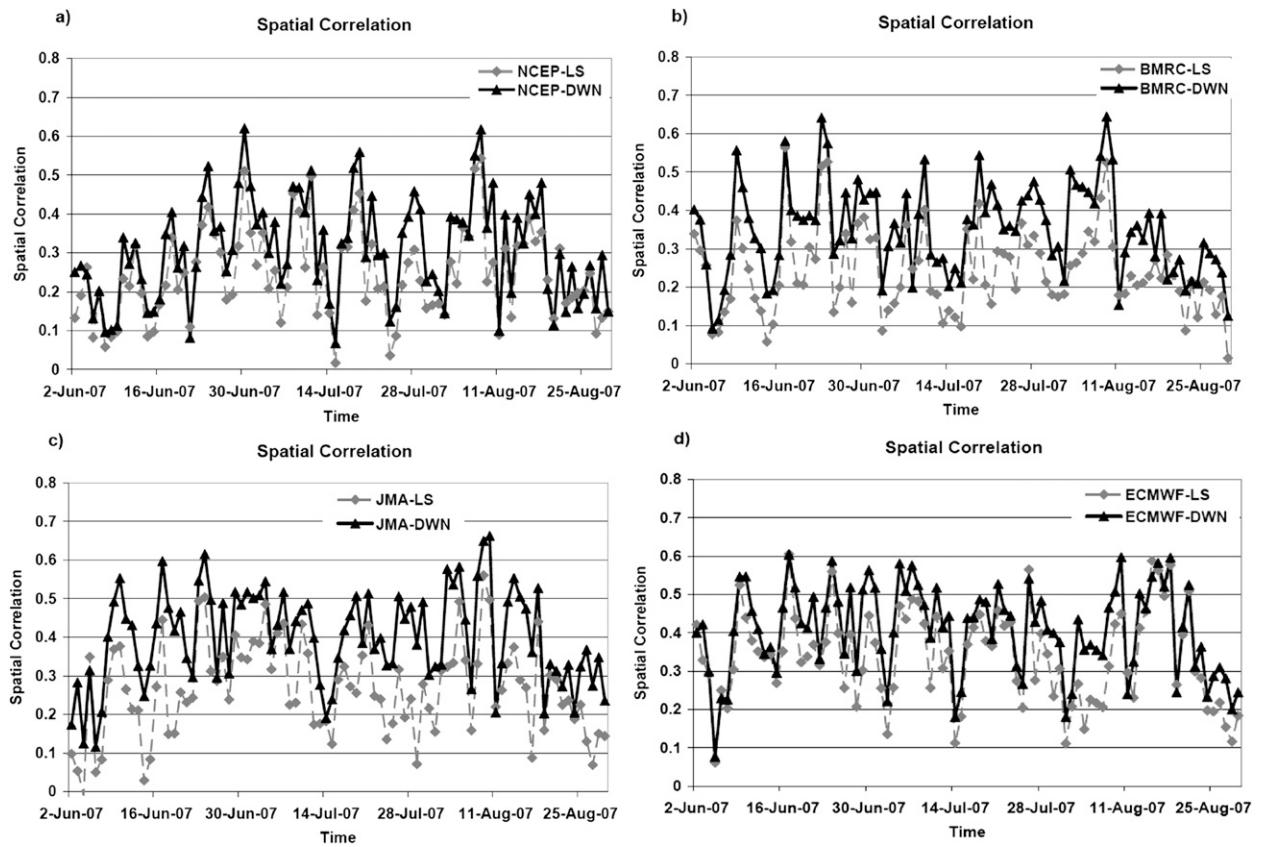
FIG. 16. Time series of correlations between forecasts and observations for day-1 forecasts over the Indian region from (a) NCEP, (b) BMRC, (c) JMA, and (d) ECMWF models. Legends show LS for large-scale model (dark black line) and DWN for downscaled forecast (light dashed gray line).

which made it closer to observation for thresholds less than 25 mm day$^{-1}$. The day-2–5 forecast did show slight improvement, in other words, the downscaled day-1 forecast of the NCEP model is best for total and medium rain assessment. JMA (Fig. 18c) is similar to ECMWF; it follows the same pattern of crossover, but the difference between the bias curves of the downscaled and interpolated forecasts is not much. However it is worth noticing that the day-4 and -5 forecasts of JMA are improved after downscaling. Figure 18d shows the drastic improvements in BMRC model. ETS like other models improved by as much as 100%, however bias is not always in the improving side. For heavy rain and higher thresholds there is certainly a big jump in bias but total rain does not improve. Overall, downscaling improves the equitable threat scores and bias scores of the member models.

## 11. Downscaled superensemble threat scores and regional rainfall forecasts

The ETS (Figs. 19a,b) and bias (Figs. 19c,d) for the day-1 forecast of each day of the entire month of

September 2007 over India are shown in Fig. 19. These are results for rainfall thresholds in excess of 0.1 mm day$^{-1}$ for multimodel forecasts. A large improvement in rainfall predictions over the member models for days 1–5 (not shown) is noted. Figure 19b depicts the same skills as Fig. 19a except for the downscaled multimodels. Figure 19b also shows the interpolated large-scale superensemble (Sup-old) for comparison purpose, which clearly shows a marked improvement in the skills after downscaling. On comparing Figs. 19a,b one can easily see the improvements in the member models skills and in the superensemble after the downscaling. The time series of the bias score is plotted in Fig. 19c (before downscaling), and Fig. 19d (after downscaling). On comparing the left panels with right panels one can easily figure out that the member models as well as superensemble forecasts showed improved skill after downscaling. The multimodel downscaled superensemble (SUP) showed higher skills compared to member model forecasts (before and after downscaling) of BMRC, JMA, ECMWF and NCEP. Also shown in the figure for comparison purpose are the ensemble mean (ENSM) and the
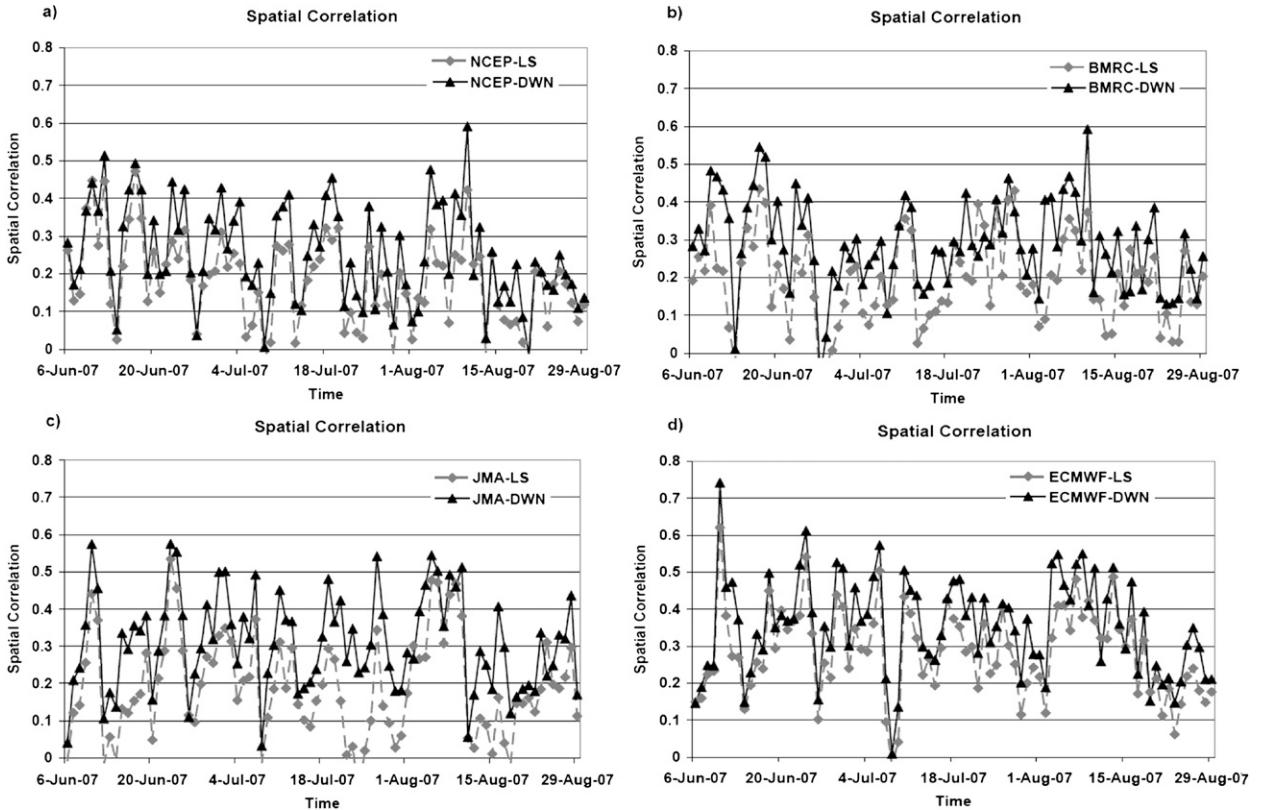
FIG. 17. As in Fig. 16, but for day-5 forecasts.

bias corrected ensemble mean (BcorENSM; the average of all four member models after correcting their biases). We have looked at the performance of the skills of the mesoscale models such as the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5) and the Weather Research and Forecasting (WRF) model in this regard. Further work using a suite of mesoscale models is clearly warranted. The results on the downscaled superensemble based on large-scale operational models shown here appear to be quite promising.

After preparing 30 daily forecasts for each day of September 2007 we came out with an overall summary for the downscaled precipitation superensemble. The equitable threat score and the bias for days 1, 2, 4, and 5 of forecasts are presented in Fig. 20. This summary compares the results for the downscaled superensemble, BMRC, JMA, ECMWF, NCEP, ENSM, and the bias-corrected ensemble mean (BCORR). Also shown in Fig. 20 are the significance levels computed following Chakraborty and Krishnamurti (2006), for the superensemble forecast. To calculate the significance level, the ETS (or bias score) of all the four member models were considered as an ensemble. Standard deviation of these four members was used for the ensemble standard

deviation; the ETS of the ensemble mean were then compared with the ETS of the superensemble to compute the significance level. This method of calculation of the significance level ensures that the ensemble spread is considered (with the use of standard deviation among the members), but a comparison is done between the ensemble mean and the superensemble. This is a conservative estimate because in general ETS (or bias) of the ensemble mean is better than the mean ETS (or bias) of the member models. The significance levels indicated in Fig. 20 were calculated considering the absolute difference between the corresponding values of ensemble mean and the superensemble. Since a higher value of ETS signifies a better forecast, it can be noticed that the superensemble was better than the ensemble mean in almost all threshold ranges. For the bias score, a value closer to 1 is considered to be a better forecast. Figure 20 also suggests that the bias score from the superensemble were generally better than that from the ensemble mean forecasts for all threshold values. It is worth noting that the significance levels of the ETS forecasts are always greater than 95% (>95%), but on days 4 and 5 for the 3 mm day$^{-1}$ threshold it was >90% and for the 20 mm day$^{-1}$ threshold it was >90% on day 4. The bias, however, was not as good as ETS. On day 1 for
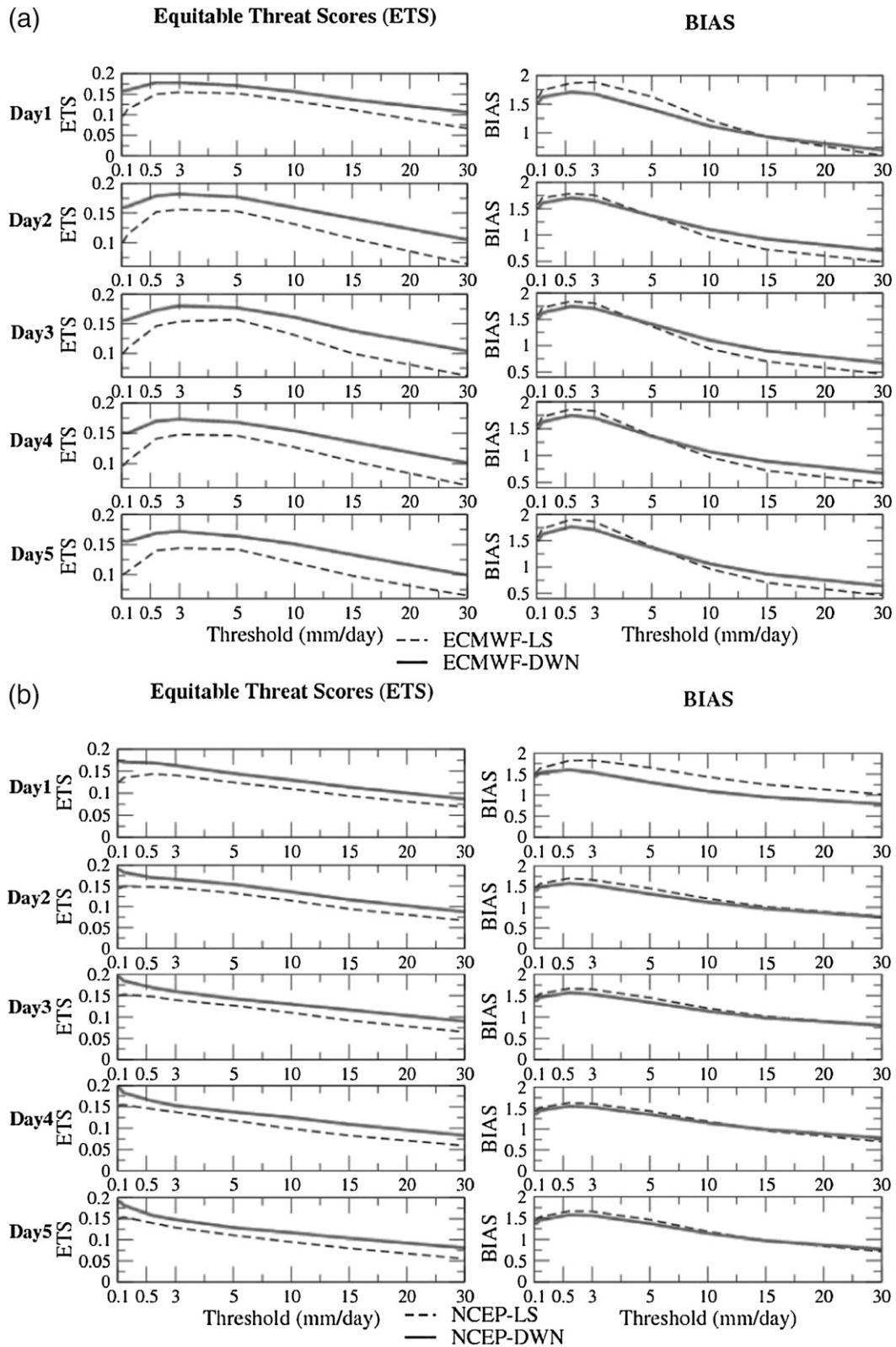
FIG. 18. (left) ETS and (right) bias of day-1–5 precipitation forecasts over the Indian region at various thresholds (mm day$^{-1}$) from (a) ECMWF, (b) NCEP, (c) JMA, and (d) BMRC models.
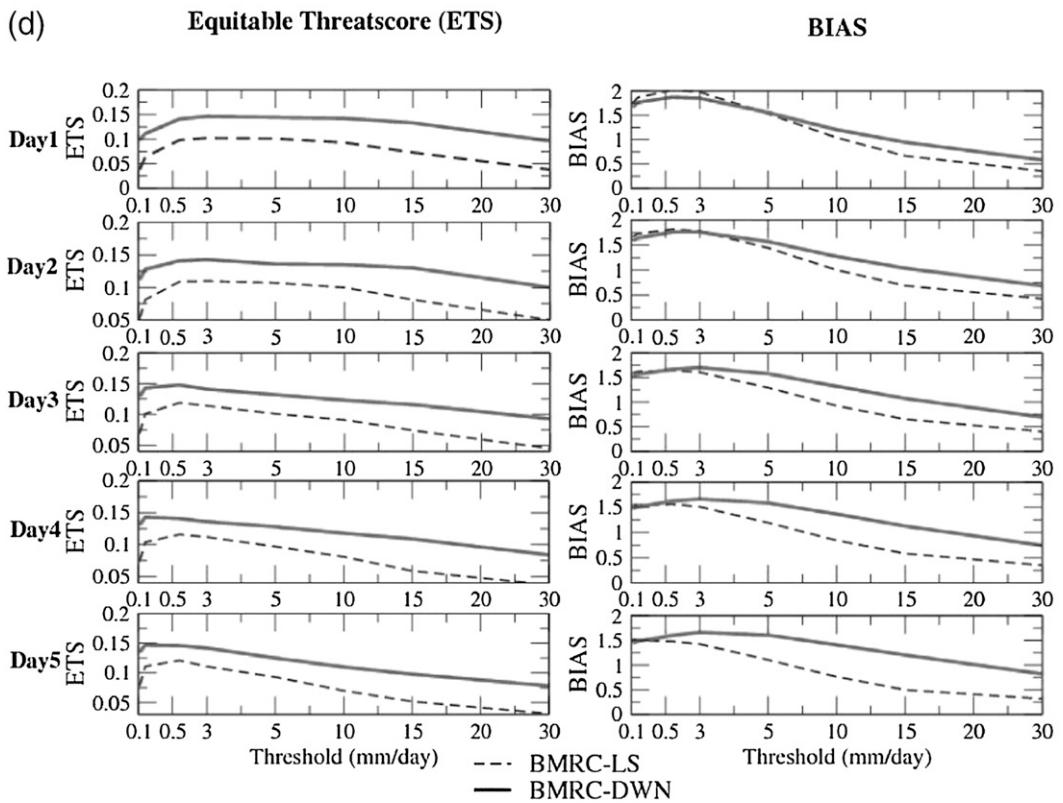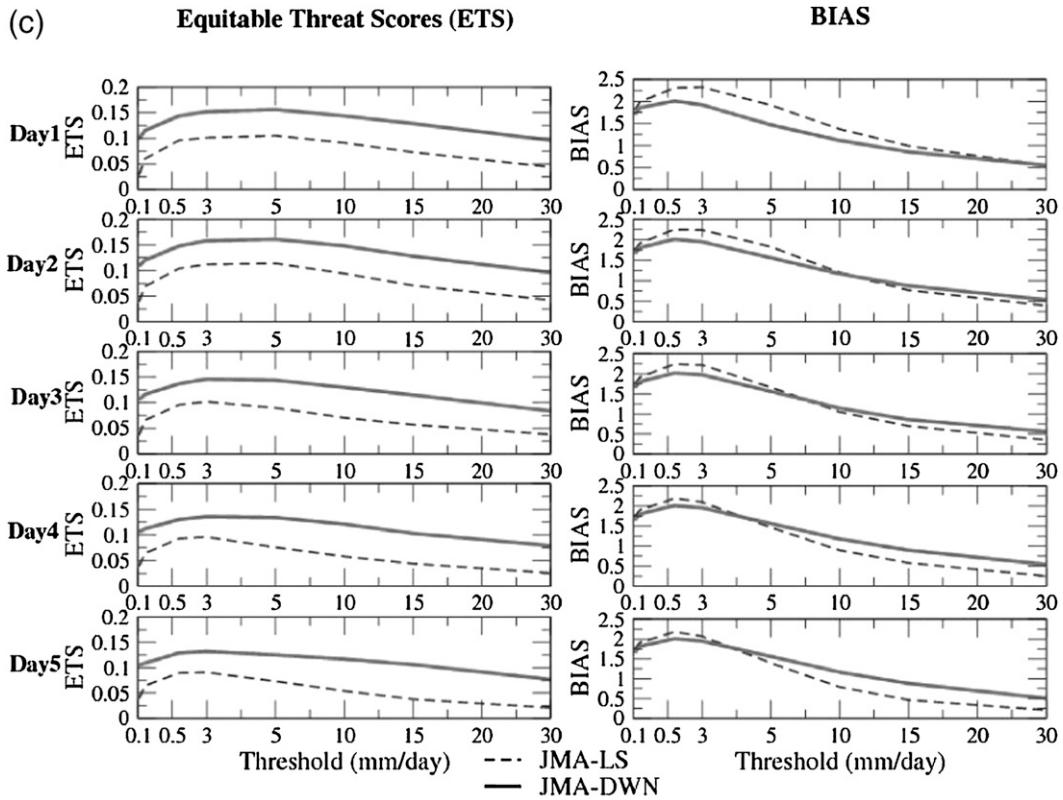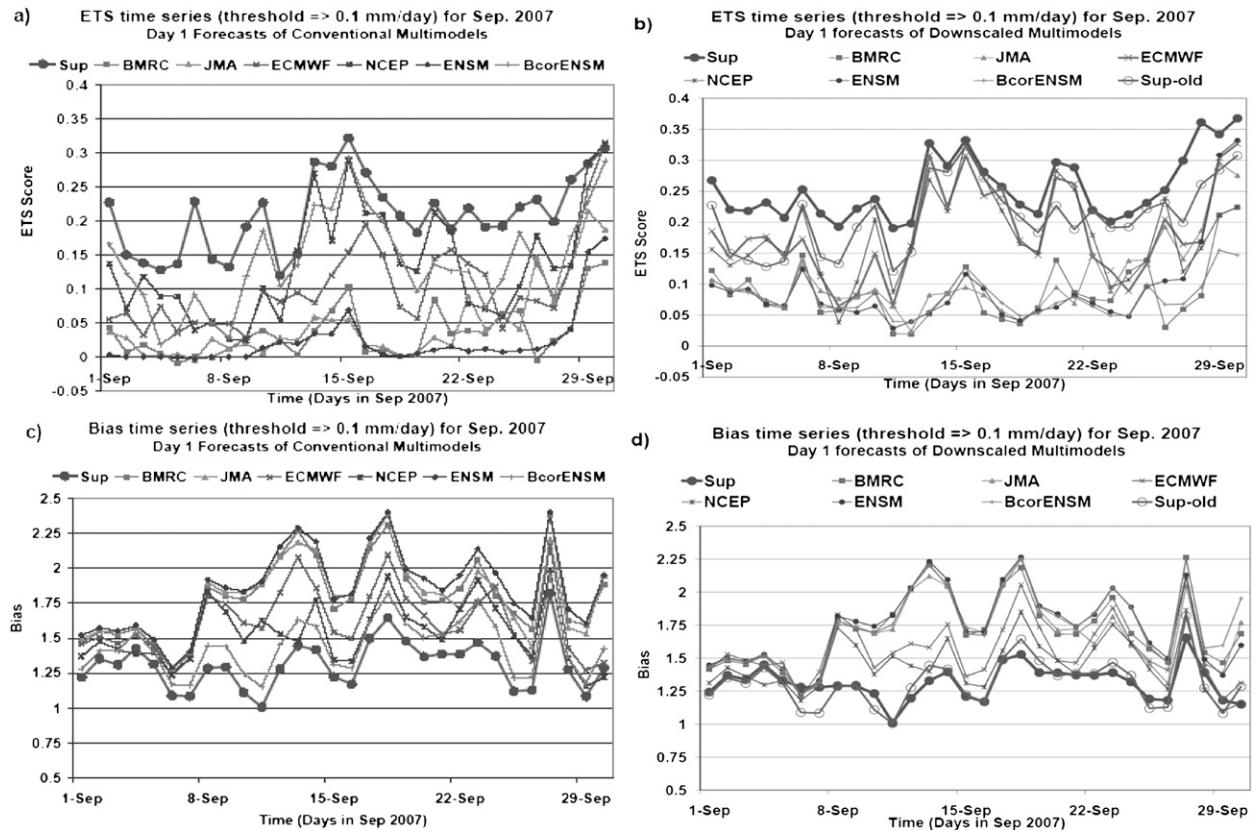
FIG. 18. (*Continued*)

FIG. 19. Time series of daily ETS score of day-1 precipitation forecasts for the 0.1 mm day$^{-1}$ threshold from the superensemble, ensemble mean, and different global models during September 2007 for (a) conventional models forecasts and superensemble and (b) downscaled model forecasts and the superensemble. (c) As in (a), but for bias. (d) As in (b), but for bias.

10 and 15 mm day$^{-1}$ thresholds the significance level was decreased to <80%, on day 2 the bias improved and the significance level of all the thresholds were >95% except for the 15 mm day$^{-1}$ threshold. On day 4 the significance level was again >95% except for the 10 mm day$^{-1}$ threshold. By day 5, the forecast does not remain very good, here the significance level for 3 out of 8 thresholds was decreased to as low as <80%; however, the 0.5, 5.0, 10.0, 20, and 30 mm day$^{-1}$ thresholds had a significance level of >95%. These forecasts show that the superensemble forecasts (for all days of forecasts 1–5) carries the highest equitable threat score and a bias score closest to 1.0. The member model forecasts carry lower skills for the ETS and the bias compared to the ensemble mean and the bias-corrected ensemble mean. Even after a bias correction for each member model an ensemble mean of them does not attain the skills of the multimodel superensemble. The bias-corrected ensemble mean assigns an equal weight of $1/N$ for each model (regardless of their individual skills). Here $N$ is the number of member models. The superensemble is more selective in this regard. It assigns fractional, positive, or negative weights that vary geographically and take into account the varied performing models.

## 12. Concluding remarks and future work

Large-scale medium-range forecasts of monsoon rainfall are generally underestimated because of the larger resolution, ≈100 km, of the operational models. A direct geographically dependent relationship of those member model rain forecasts (bilinearly interpolated) to an analyzed rain field (rain gauge based) at a resolution of $0.5° \times 0.5°$ latitude–longitude is estimated using a simple statistical downscaling algorithm. The motion of weather systems is somewhat reasonably captured by these large-scale operational models. The downscaling algorithm presented in this paper is able to translate the weaker large-scale rains to heavier rains at higher resolution as the monsoon systems traverse in these forecasts. We used two algorithms for rainfall forecasts: one is a downscaling algorithm and the second is a multimodel superensemble. Both of these algorithms call for a least squares minimization of errors. We have noted
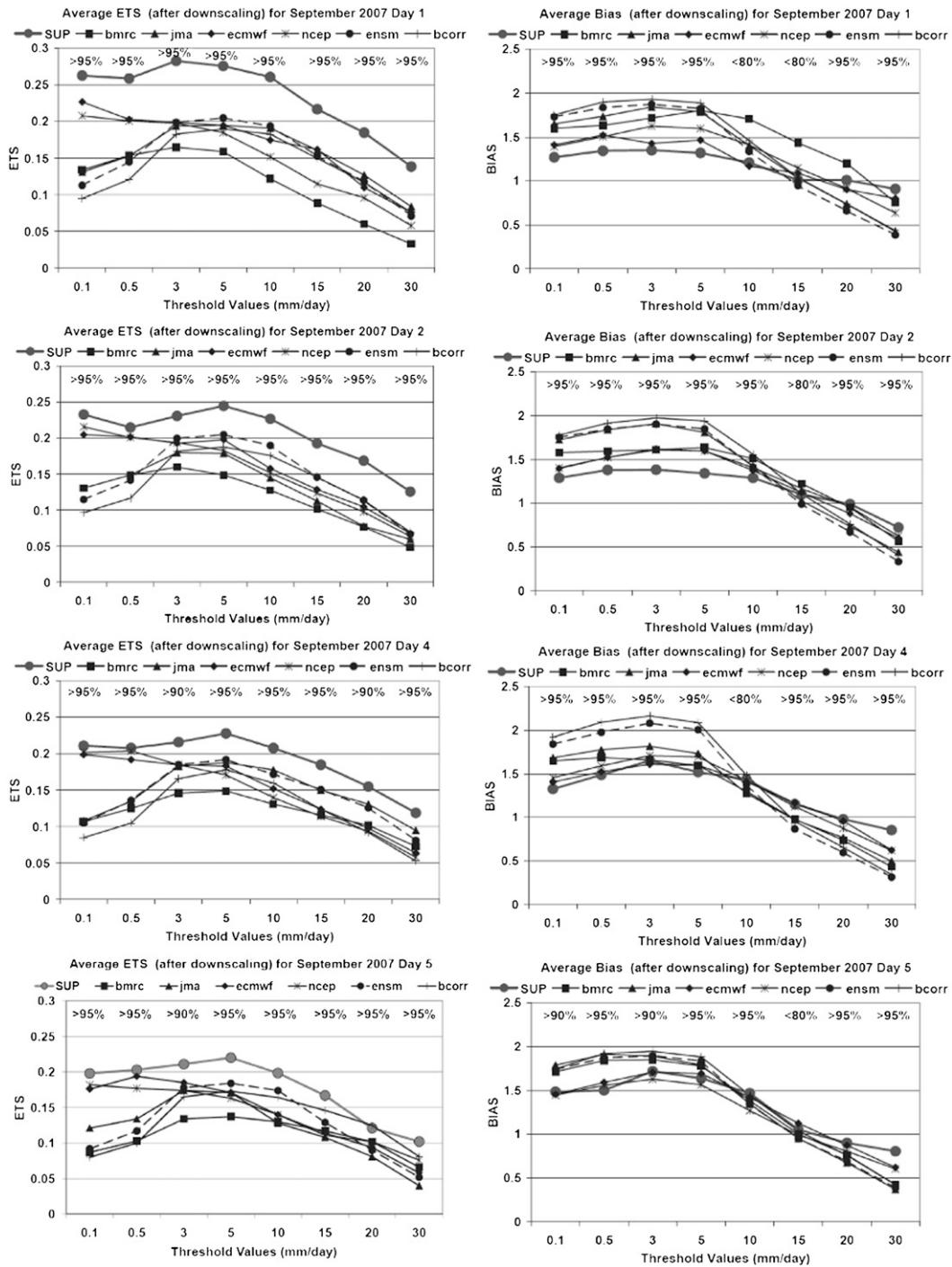
FIG. 20. Monthly mean ETS and bias scores for day-1, -2, -4, and -5 precipitation forecasts from member models and the superensemble after downscaling during September 2007. The numbers on top are the significance levels.

(not published) that if any algorithm takes on a large number of computational steps then the errors propagate through each of the many operations and the final reduction of errors is smaller. By limiting it to two least squares minimization of errors, we found this system to be quite robust. The resulting bias errors are indeed reduced for the total and the heavy rain (even exceeding rates such as 15 mm day$^{-1}$). We noted that for both the total rains and for the heavier rains the downscaled multimodel superensemble increases the skill ($\approx$0.1 for
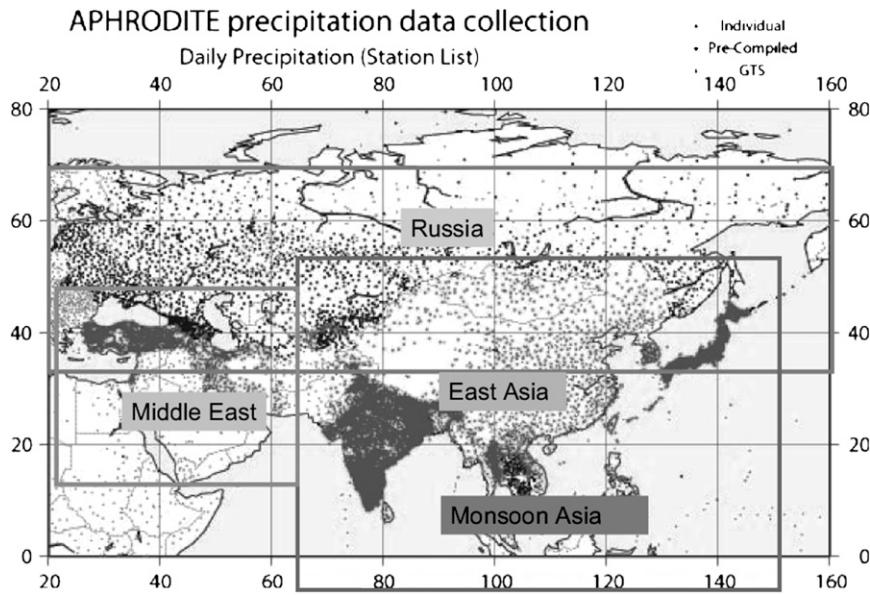
FIG. 21. Rain gauge network over East Asia (Xie et al. 2007).

member models, to ≈0.3 for the multimodel downscaled superensemble) for days 1–5 of the forecasts.

The downscaled multimodel superensemble carried a training phase covering 92 days (June–August 2007) following our earlier studies (Krishnamurti et al. 2000a; Mishra and Krishnamurti 2007). The forecast phase contained 30 days of forecasts that covered the entire month of September 2007. These 5-day forecasts were geographically examined on a daily basis over the entire all-India domain. We noted that even at day 5 the downscaled superensemble was able to differentiate between rainfall events over different parts of India with high skill scores. In these examples the equitable threat scores for total rainfall improved between 20% and 30% compared to forecasts from the downscaled single models. This work would not have been possible without the availability of high-resolution rain gauge network of the India Meteorological Department shown in Fig. 2 that included over 2100 rain gauge sites. Future work on extending this effort to a much larger domain is possible now (Fig. 21). This new rainfall data collection was prepared by Xie et al. (2007). It is also possible to carry out the proposed algorithms on a site-specific basis (rain gauge sites instead of grid points). This would increase the practical utility of the downscaled multimodel superensemble.

The downscaling algorithm provides useful information on the systematic errors of precipitation forecast for the member models. The spatial correlation for the all-Indian rainfall forecast for days 1–5 are about 20%–30% higher for each of the downscaled member models compared to those of prior to downscaling. The equitable threat score for precipitation over India shows 10%–20% improvement for each of the member models subsequent to downscaling. A further major improvement by 100%–200% became possible from the construction of the downscaled multimodel superensemble.

## APPENDIX

### Equitable Threat Score and Bias Score

The equitable threat score is defined as

$$\text{ETS} = \frac{H - \text{CH}}{F + O - H - \text{CH}}, \quad (A1)$$

and the bias score is defined as

$$\text{bias} = \frac{F}{O}, \quad (A2)$$

where $F$ is the number of forecast points above a threshold, $O$ is the number of observed points above a threshold, $H$ is the number of hits above threshold, NUM is the total number of grid points to be verified and CH is the expected number of hits in a random forecast of $F$ points for $O$ observed points, which is equal to

$$\text{CH} = \frac{F \times O}{\text{NUM}}. \tag{A3}$$

A value of 1.0 for ETS indicates a perfect forecast. The minimum value for ETS can be $-1/3$. For the bias score, a value of 1.0 corresponds to a perfect forecast.

## REFERENCES

Anthes, R. A., 1983: Regional models of the atmosphere in middle latitudes. *Mon. Wea. Rev.,* **111,** 1306–1330.

Chakraborty, A., and T. N. Krishnamurti, 2006: Improved seasonal climate forecasts of the South Asian summer monsoon using a suite of 13 coupled ocean–atmosphere models. *Mon. Wea. Rev.,* **134,** 1697–1721.

——, and ——, 2009: Improving global model precipitation forecast over India using downscaling and the FSU superensemble. Part II: Seasonal climate. *Mon. Wea. Rev.,* **137,** 2736–2757.

Druyan, L. M., M. Fulakeza, and L. Patric, 2002: Dynamic downscaling of seasonal climate predictions over Brazil. *J. Climate,* **15,** 84–117.

Huth, R., 2002: Statistical downscaling of daily temperature in central Europe. *J. Climate,* **15,** 1731–1742.

Kanamaru, H., and M. Kanamitsu, 2007: Fifty-seven-year California Reanalysis Downscaling at 10 km (CaRD10). Part II: Comparison with North American Regional Reanalysis. *J. Climate,* **20,** 5572–5592.

Kanamitsu, M., and H. Kanamaru, 2007: Fifty-seven-year California Reanalysis Downscaling at 10 km (CaRD10). Part I: System detail and validation with observations. *J. Climate,* **20,** 5553–5571.

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

——, ——, Z. Zhang, T. LaRow, D. Bachiochi, C. E. Williford, S. Gadgil, and S. Surendran, 2000a: Multimodel superensemble forecasts for weather and seasonal climate. *J. Climate,* **13,** 4196–4216.

——, ——, D. W. Shin, and C. E. Williford, 2000b: Improving tropical precipitation forecasts from a multianalysis superensemble. *J. Climate,* **13,** 4217–4227.

——, A. K. Mitra, W.-T. Yun, and T. S. V. V. Kumar, 2006a: Seasonal climate forecasts of the Asian monsoon using multiple coupled models. *Tellus,* **58A,** 487–507.

——, and Coauthors, 2006b: Weather and seasonal climate forecasts using the super ensemble approach. *Predictability of Weather and Climate,* T. Palmer, Ed., Cambridge Publications, 532–560.

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

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.,* **83,** 407–430.

Mishra, A. K., and T. N. Krishnamurti, 2007: Current status of multimodel superensemble and operational NWP forecast of the Indian summer monsoon. *J. Earth Syst. Sci.,* **116,** 369–384.

Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1993: Ensemble prediction. *Proc. Seminar on Validation of Models over Europe,* Vol. 1, Reading, United Kingdom, ECMWF, 21–66.

Pandey, G. R., D. R. Cayan, M. D. Dettinger, and K. P. Georgakakos, 2000: A hybrid orographic plus statistical model for downscaling daily precipitation in northern California. *J. Hydrometeor.,* **1,** 491–506.

Rajeevan, M., J. Bhate, J. Kale, and B. Lal, 2006: High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells. *Curr. Sci.,* **91,** 296–306.

Shepard, D., 1968: A two-dimensional interpolation function for irregularly spaced data. *Proc. 23rd Natl. Conf. of the Association of Computing Machinery,* Princeton, NJ, ACM, 517–524.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* **74,** 2317–2330.

Vijaya Kumar, T. S. V., T. N. Krishnamurti, M. Fiorino, and M. Nagata, 2003: Multimodel superensemble forecasting of tropical cyclones in the Pacific. *Mon. Wea. Rev.,* **131,** 574–583.

von Storch, H., H. Langenberg, and F. Feser, 2000: A spectral nudging technique for dynamical downscaling purposes. *Mon. Wea. Rev.,* **128,** 3664–3673.

Williford, E. C., T. N. Krishnamurti, R. C. Torres, S. Cocke, Z. Christidis, and T. S. Vijaya Kumar, 2003: Real-time multimodel superensemble forecasts of Atlantic tropical systems of 1999. *Mon. Wea. Rev.,* **131,** 1878–1894.

Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A gauge-based analysis of daily precipitation over East Asia. *J. Hydrometeor.,* **8,** 607–626.