Current status of multimodel superensemble and operational NWP forecast of the Indian summer monsoon

Akhilesh Kumar Mishra* and T N Krishnamurti

Department of Meteorology, Florida State University, Tallahassee, Florida 32306, USA.
**e-mail: akhil@io.met.fsu.edu*

In the last thirty years great strides have been made by large-scale operational numerical weather prediction models towards improving skills for the medium range time-scale of 7 days. This paper illustrates the use of these current forecasts towards the construction of a consensus multimodel forecast product called the superensemble. This procedure utilizes 120 of the recent-past forecasts from these models to arrive at the training phase statistics. These statistics are described by roughly 10^7 weights. Use of these weights provides the possibility for real-time medium range forecasts with the superensemble. We show the recent status of this procedure towards real-time forecasts for the Asian summer monsoon. The member models of our suite include ECMWF, NCEP/EMC, JMA, NOGAPS (US Navy), BMRC, RPN (Canada) and an FSU global spectral forecast model. We show in this paper the skill scores for day 1 through day 6 of forecasts from standard variables such as winds, temperature, 500 hPa geopotential height, sea level pressure and precipitation. In all cases we noted that the superensemble carries a higher skill compared to each of the member models and their ensemble mean. The skill matrices we use include the RMS errors, the anomaly correlations and equitable threat scores. For many of these forecasts the improvements of skill for the superensemble over the best model was found to be quite substantial. This real-time product is being provided to many interested research groups. The FSU multimodel superensemble, in real-time, stands out for providing the least errors among all of the operational large scale models.

1. Introduction

Some of the basic characterizations of weather and climate systems over the region of the Asian summer monsoon may be obtained from Rao (1976) and Krishnamurti (1979) among several authors. Our study focuses on daily weather predictability over the diverse geographical regions of Asian monsoon. Most operational weather centers routinely provide 5 to 6-day NWP forecast over the global domain. It is thus possible to examine the current state of the art skills of real-time operational global models over this belt. As many as seven global models were available for the prediction of weather on real-time. Our study entails the construction of ensemble and superensemble forecasts following Krishnamurti *et al* (2000a). Skills of member models and those of the ensemble and superensemble are constructed at the end of days 1, 2, 3, 4, 5, and 6 of the forecasts. Traditionally, monsoon forecasts have encountered many difficulties that stem from numerous issues such as lack of adequate upper air observation over the neighboring oceans, mesoscale nature of convection, proper resolution, radiative interactions, planetary boundary layer physics, mesoscale air–sea fluxes, and representation of orography. Uncertainties in any of these areas lead to large systematic errors; those are reduced to some extent by the construction of the multimodel superensemble. These syntheses

Keywords. Monsoon forecast; ensemble modeling for monsoon; inspiring monsoon forecast beyond operational skills.

J. Earth Syst. Sci. **116**, No. 5, October 2007, pp. 369–384

of multimodel forecasts provide us with a somewhat better product for the current state of the art. In this paper, we present some of these current improvements that were made possible.

We shall next present the relative performance of several of these models with respect to each other. This will describe the current state of the art. Figure 1 shows the skills computed by EMC's global climate and weather modeling branch for January 2006. Models considered in this comparison study are global weather forecasting center models such as NCEP (GFS); ECMWF (ECM); UKMet (UKM); Meteorological Service of Canada (MSC) and NOGAPS (NGP). In figure 1(a) and (b) the anomaly correlations for the 500 hPa geopotential height forecasts (day 1 through 6) are illustrated for the northern and southern hemisphere respectively. The forecast from the European center (ECMWF) carries the highest skills for all of the 6 days. It should be noted that these forecast validations were carried out using the analysis by NCEP/EMC and hence this is to be viewed as a current state of operational forecast skills. These skills for January 2006 for the northern hemisphere were slightly larger than those for the southern hemisphere. Higher skills for the anomaly correlation > 0.9 were carried by the GFS (the US model), UKMet office model and the ECMWF through day 4 of forecasts over the northern hemisphere. These are indeed very high skills that have been achieved in recent years. Southern hemispheric skills are greater than 0.9 through day 3 of the forecasts by these models. Figure 1 (c and d) illustrates the skills of forecast over the tropics for zonal and meridional winds at the 850 hPa and 200 hPa levels for these same models through day 3 of forecasts. Anomaly correlation skills at these two vertical levels are comparable in magnitudes. The skills of the ECMWF again standout compared to all other models. Figure 1 (e through h) describes RMS error for the geopotential heights at the 500 hPa level for the northern hemisphere and southern hemisphere as well as zonal and meridional winds at the 850 hPa and 200 hPa levels. In all of these illustrations, the least RMS error is systematically provided by the ECMWF. These errors grow with the lengths of the forecast periods, between day 1 and day 3 of forecasts. Figure 1 was extracted from the web files of NCEP/EMC in order to illustrate a current state of the art. The purpose is to show that it is possible to provide a multimodel consensus forecast on real-time that provides additional skill beyond what is seen here.

2. Superensemble methodology

Predictions from an ensemble of slightly different initial conditions and/or various versions of models using a single base model are often carried out by the weather services. An ensemble mean is defined as the average of all models involved in the ensemble suite. Another type of ensemble mean is the bias-removed ensemble mean, where the bias of each model is removed prior to the execution of an ensemble average. In both of these cases all models are given an equal weight of $1/N$, where N denotes the total number of models. The superensemble approach is a recent contribution to the general area of weather and climate forecasting developed at FSU; this has been discussed in a series of publications, Krishnamurti *et al* (1999, 2000a, 2000b and 2001). This technique entails the partition of a time line into two parts. One part is a 'training' phase, where forecasts by a set of member models are compared to the observed or the analysis fields with the objective of developing statistics (i.e., weights a_i) on the least squares fit of the forecasts to the observations. The second part is the forecast phase where estimates for a_i from the training phase are used to create the superensemble. The performance of the individual models is obtained in the training phase using multiple linear regressions against observed (analysis) fields. 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 model anomalies of the variables are regressed against the observed anomalies when formulating the superensemble forecasts, and the weights are multiplied to the corresponding model anomalies. The constructed forecast is:

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

where $\overline{O}$ is the observed climatology over the training period; a_i is the weight for the i^{th} member in the ensemble; and F_i and $\overline{F_i}$ are the i^{th} model's forecasts and the forecast mean (over the training period) respectively. N is the number of member models. The weights a_i's are obtained by minimizing the error term G, where G is expressed as:

$$G = \sum_{t=1}^{N_{\text{train}}} (S_t' - O_t')^2. \tag{2}$$

Here N_{train} is the number of time samples in the training phase, and S_t' and O_t' are the respective superensemble and observed field anomalies at training time 't'.

This exercise is performed at all model grid points. A fit performed for all model variables at all model grid points at all vertical levels typically yields close to 10^7 regression weights. These spread

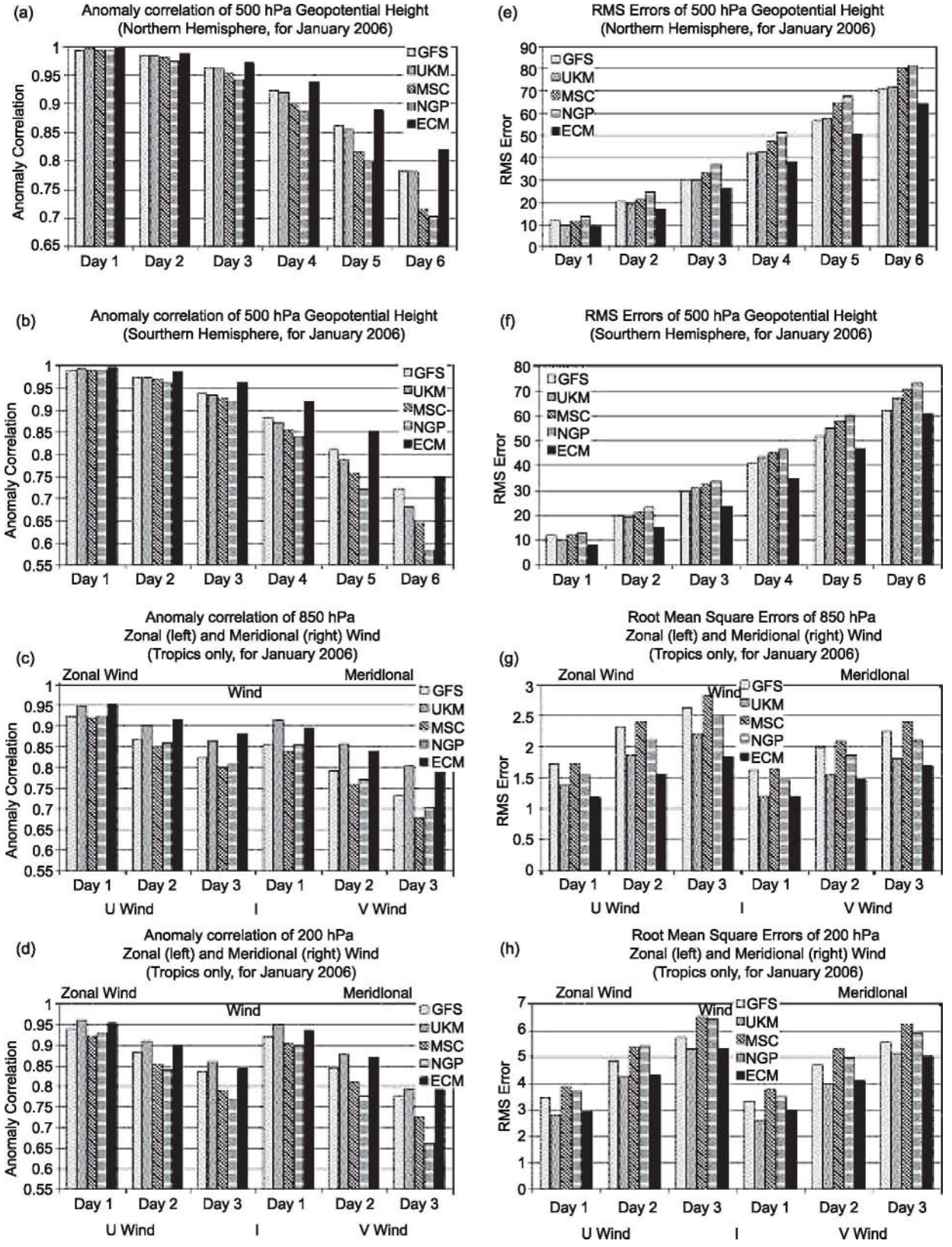

Figure 1. Current skills of the operational forecasting models for January 2006 (obtained from EMC/NCEP).

of weights are fractional, positive or negative. This large number arises from the number of transform grid points, number of vertical levels, number of basic variables and the number of models. Over many such locations we have noted diverse performance characteristics of the member models that

arises from differences in horizontal and vertical discretization, treatment of physics, handling of inhomogeneity of land surface, orography, water bodies, surface physics and boundary conditions. All such peculiarities tend to leave their signature in the error distributions and hence on these weights. These may be thought of as a collective bias correction procedure. The second part of the time line is composed of model predictions. The superensemble approach combines each of these forecasts according to the weights determined during the training phase using the formulation. The prediction 'S' is referred to as the 'superensemble' forecast. This forecast should be contrasted with the more standard anomaly forecasts known as the bias-removed ensemble mean or ensemble mean forecast

$$E = \overline{O} + \frac{1}{N}\sum_{i=1}^{N}(F_i - \overline{F}_i)$$

or

$$\hat{E} = \overline{O} + \frac{1}{N}\sum_{i=1}^{N}(F_i - \overline{O}). \qquad (3)$$

The skill of the multimodel superensemble method significantly depends on the error covariance matrix, since the weights of each model are computed from a designed covariance matrix. The classical method for the construction of the superensemble utilizes a least square minimization principle within a multiple regression of model output against observed 'analysis' estimates. This entails a matrix inversion that is solved by the Gauss Jordan elimination technique. That matrix can be ill-conditioned and singular depending on the interrelationships of the member models of the superensemble. We have recently designed a singular value decomposition (SVD) method (Wilks 1995) for the multimodel superensemble that overcomes this problem and removes the ill conditioning of the covariance matrix entirely (Yun *et al* 2002). Tests of this method have shown great skills in weather and seasonal climate forecasts compared to the Gauss Jordan elimination method.

3. Models participating in NWP superensemble suite

Seven global models used in the construction of real-time superensemble include the European Center for Medium-Range Weather Forecasts model (ECMWF), the Global Forecast System of National Centers for Environmental Prediction (NCEP/GFS), Global spectral model of Japan Meteorological Agency (JMA), Global Environmental Multiscale Model from Canadian Meteorological Center (GEM), the Fleet Numerical Meteorology and Oceanography Center model (NOGAPS), the model of Australian Bureau of Meteorology (GASP) and an FSU Global Spectral Model. In India, the National Center for Medium Range Weather Forecasting (NCMRWF) has been carrying out routine numerical weather predictions on real-time over the last 12 years. They issue medium range global forecasts, i.e., 1 to 6 days into the future. They have steadily improved the state of numerical prediction and its applications for the use of the agricultural community in India. The current FSU multimodel superensemble does not include this model within its suite of models. Thus a direct comparison of this model with the superensemble was not possible.

Table 1 illustrates the layout of the various participating models of our NWP Superensemble suite. This table shows the horizontal and vertical resolutions, various physical parameterization schemes and the land surface schemes deployed by each of these participating models. Overall, we note a diversity of operational models that carry a range of resolution and physics. Table 1 describes the models which were used to prepare the forecast used in this study, some of the operational centers are presently using an advanced version of the model.

3.1 *Data assimilation and satellite datasets*

A variety of data assimilation schemes are used by different operational NWP centers around the world. The initial datasets used by these models are generally the same for the surface-based observations. However, the satellite-based data content of the different models do vary. Furthermore, there are major differences in the way the satellite data are being assimilated by the different centers. Direct use of radiances is part of the satellite data assimilation. NCEP and FNMOC/NOGAPS utilize the satellite-based datasets in their 3-DVar data assimilation system. JMA had been using 3-DVar until February 2005 in their global analysis system, although 4-DVar was operational in their mesoscale analysis system. From March 2005 onwards JMA has implemented 4-DVar in the global operational forecast and analysis system. ECMWF carried out very comprehensive 4-DVar data assimilation. Canadian model has 4-DVar operational; recently on 31 October 2006 they made major changes in their operational Global Environmental Multiscale (GEM) model that includes increasing significantly horizontal and vertical resolutions, improving its physical parameterization, improving the condensation and precipitation package, introducing the more sophisticated ISBA (interactions soil-biosphere-atmosphere) surface scheme and improving the

Table 1. *Descriptions of NWP models participating in multimodel FSU superensemble.*

Model	Resolution		Cumulus param.	PBL and surface Physics	Radiation	Land surface processes
	Horizontal	Vertical				
BMRC	T239	29 levels upto 10 hPa	Kuo 1974	Stability dependent constant flux layer (Monin–Obukhov similarity theory)	Radiative transfer (Schwarzkopf and Fels 1991) (Edwards and Slingo 1996)	1. Gravity wave drag parameterization 2. Interactive soil moisture (bucket type) 3. Land Albedo climatology
FSU	T170	14 level 50–1000 hPa	Modified Kuo (Krishnamurti *et al* 1983)	Sfc flux via similarity theory (Businger *et al* 1971) Vertical diffusion of flux (Louis 1979)	Harshvardhan and Corsetti (1984) Lacis and Hancen (1974)	BATS
JMA	T319	40 levels up to 0.4 hPa	Prognostic Arakawa Schubert Cum. Param. (1974)	Mellor Yamada level 2, Monin–Obukhov similarity theory	Short wave (every hour) infrared (3 hourly),	Simple Biosphere Model (SiB) Gravity wave drag
NCEP	T255	42 sigma levels up to 0.27 hPa	Pan and Wu (1994), based on Arakawa Schubert as simplified by Grell (1993)	PBL (Troen and Mahrt 1986), roughness length by (Charnock 1955) and Monin–Obukhov similarity theory	Long wave computation every 3 hours, short wave every hour Schwarzkopf and Fels (1991), Harshvardhan *et al* (1989)	Gravity wave drag SST; 5 day running mean analysis Snow cover from NESDIS
NRL	T239	24 sigma levels up to 1 hPa, 5 sigma levels below 850 hPa	Relaxed Arakawa Schubert (RAS) (Moorthi and Suarez 1992)	After ECMWF's vertical mixing parameterization (Louis *et al* 1982) Sfc flux param. (Louis 1979)	Harshvardhan *et al* (1989)	Gravity wave drag (Palmer *et al* 1986) SST: US Navy Ice cover percentage: NAVICE center weekly analysis
RPN	T199	28 η levels up to 10 hPa	Kuo type deep convection scheme	Sfc layer based on Monin–Obukhov similarity theory	Solar and IR radiation interactive with water vapor, CO_2, O_3 and clouds	Gravity wave drag, prediction of sfc temperature over land (force-restore method)

Table 2. *ECMWF data assimilation system.*

	Satellite or instruments		
Conventional observations	Platform	Data	List of sensors
• Synop surface pressure	NOAA15-16-17 & DMSP13-14-15	AMSU/HIRS + SSM/I raw radiances	• 3 × AMSUA (NOAA 15/16/17) - radiances
• Ship surface pressure and wind			• 3 SSMI (F-13/14/15) - radiances
• Buoy surface pressure and wind	NASA- QuikScat	Sea winds	• 2 × HIRS (NOAA-16/17) - radiances
• PAOB surface pressure	Geostationary platform	AMVs + WV radiances	• Radiances from 2 × GEOS (Met-7 GOES-10)
• Radiosonde temperature wind and humidity	Polar platform	AMVs from MODIS-TERRA, AQUA products	• Winds from 3 × GEOS (Met-7/5 GOES-10) and MODIS/TERRA
• Pilot wind			• Sea winds on QuikScat
• Aircraft temperature and wind	GOME and SBUV-NOAA16	Ozone profiles/columns	• ERS-2 Altimeter/SAR
• American and European profiler winds			• SBUV (NOAA-16)
			• AIRS AMSUA

GOME – Global Ozone Monitoring Experiment, SBUV – Solar Backscatter Ultraviolet, AMSUA – Advanced Microwave Sounding Unit-A, AIRS – Atmospheric InfraRed Sounder, HIRS – High Resolution Infrared Radiation Sounder, and ERS – European Remote Sensing satellite.

model physics. In addition to the model changes, a number of improvements have also been introduced into the analysis used by the model, e.g., use of a new set of background error statistics. BMRC is using the 1-DVar in their operational Global Analysis and Prediction (GASP) system. BMRC is planning to adopt the 3-DVar, where a set of radiances is analyzed simultaneously in a three-dimensional sense. Table 2 describes the data assimilation system of ECMWF. Table 2 explains conventional observations (left column) and satellite data considered for the assimilation. Table 2 also explains list of sensors deployed in different platforms to facilitate the existing observing system.

3.2 *FSU model (physical initialization)*

The Florida State University Global Spectral Model (FSUGSM) also produces a 6-day forecast for inclusion into the superensemble. In this model, the initial fields have been physically initialized using the Ferraro and Marks rain rate algorithm (1995), also called the NOAA–NESDIS (National Oceanic and Atmospheric Administration)–(National Environmental Satellite Data and Information Service) SSM/I (Special Sensor Microwave Instrument) algorithm. Physical initialization is a powerful tool which primarily assimilates satellite-derived observed rainfall distributions along with calculated surface fluxes of moisture to produce a physically consistent and more realistic spin-up of the initial state. The key to the spin-up procedure is a Newtonian relaxation, where selected variables are 'nudged' toward prescribed values during a 24-hour pre-integration period. Upon integrating the FSUGSM, the rainfall patterns and surface moisture fluxes as well as the wind and mass fields appear more robust. For example, experiments have shown that within the ITCZ, monsoonal flows, and typhoons, the moisture flux is stronger and more organized. Thus, the precipitation areas are better defined. In the physical initialization technique, day minus one (one day prior to the initial condition) and day zero (initial time) ECMWF (European Centre for Medium-Range Weather Forecasts) operational analyses are collected. Next, they are augmented with microwave radiance datasets from five satellites, including the NASA TRMM (Tropical Rainfall Measuring Mission) satellite. From these radiances, rain rates are derived. Then, via a number of reverse algorithms, including reverse similarity theory and reverse cumulus parameterization, the observed interpolated rainfall rates at each time step and at every location are used to derive a physically consistent initial state at hour zero. This initialization procedure includes the following five computational areas:

Step 1: The surface fluxes of moisture and heat are obtained from the vertical integration of the apparent moisture sink Q_2, and the apparent heat source Q_1, following the

analysis of Yanai *et al* (1973). The proposed procedure using this moisture balance relation improves these estimates as one applies it repeatedly since the divergence, heating, and the moisture fields adjust to the imposed rain rates with time. The vertical integrals of Q_1 and Q_2 are carried out from the top of the atmosphere, $\sigma = 0$ to the earth's surface $\sigma = 1$, where σ is the vertical coordinate.

Step 2: A reverse similarity theory is invoked during each time step of the physical initialization. The rain rates are assumed to be known and are used to provide the specific humidity and the potential temperature at the top of the constant flux layer. An assimilation of this data at the top of the constant flux layer promotes a consistency between the calculated flux and the imposed rain rate.

Step 3: A reverse cumulus parameterization algorithm restructures the moisture variable following the vertical coordinate which is consistent with the imposed rain rate. This restructuring of the moisture variable is achieved between the cloud base (the lifting condensation level) and the cloud top (the level at which the local moist adiabatic and the cloud-environment sounding intersect).

Step 4: In the upper troposphere ($\sigma < 0.5$) the moisture variable is restructured to minimize the local difference between the model and the satellite inferred fields of outgoing long wave radiation. A bisection method determines a parameter which defines the structure of the moisture variable in the upper troposphere.

Step 5: These components of physical initialization are executed within data assimilation phases of the model initialization. During this assimilation of the surface fluxes, the outgoing longwave radiation, the divergence field, the diabatic heating and the surface pressure evolves in a consistent manner. The details of the above procedure are presented in the two papers Krishnamurti *et al* (1991, 1993).

Lacking a unified column model that integrates all of the physical process (i.e., the surface flux parameterization, the cumulus parameterization and the radiative transfer scheme), the current FSU model has separate computational algorithms for these components. The sequence of computations within the physical initialization modifies the moisture variable separately to satisfy each requirement, i.e., fluxes, rain rate and the earth radiation budget. The sequence of reverse algorithms is: reverse similarity followed by OLR matching and then the reverse cumulus parameterization. The time interpolated rain rates are an input to the reverse similarity and the reverse cumulus parameterization algorithms. The satellite based OLR enters the matching algorithm. A one time step, forward integration follows the execution of the reverse cumulus parameterization algorithm. Since the reverse cumulus and the normal cumulus are carefully designed to be almost completely reversible, the "observed" rain rate is reproduced by the model at the end of the time step. Thus, an accumulation over 24-hour of simulated assimilated rainfall appears realistic.

Currently missed by the physical initialization are instances where the model calls for no precipitation (i.e., no net moisture convergence or absence of conditional instability) whereas the satellite or raingauge observations imply rain. We impose saturation as an upper limit for the modification of the specific humidity. This constrains the extent to which rainfall can be recovered in these instances. Further work is ongoing in both of these above areas to overcome these difficulties towards further improvement of the now casting skill.

3.3 *Optimizing the number of training days*

Because of the nature of atmospheric variability over different regions, the minimum number of training days required for stabilizing the statistical weights varies from region to region. Krishnamurti *et al* (2003) illustrated the nature of this geographical dependence. They studied the distribution of the minimum number of days that were required for the training phase of NWP multimodels. This distribution varies from roughly 90 days to 120 days; over the land area more days of training were required compared to those over the ocean. This led to our accepting a number of 120 days for the length of training the phase for NWP. We noted that roughly 120 days of training provided stability for the statistical weights. This number does vary somewhat from one vertical level to the next and from one atmospheric variable to the next. Overall a use of 120 days appeared to cover the needs for stable statistics. Another important aspect of these coefficients is that it is better to compute these from data of similar previous seasons. For instance a use of dry season pre-monsoon data over 120 days (such as March, April, May and June) may not be best suited for the rainy and wet months of July and August. In such instances, several previous years of July August datasets can be used. This of course also requires that the models of our forecast suite have not altered much over these several years used for the training phase.We have

used the operational ECMWF analysis at 0.5° latitude/longitude as the benchmark fields for most variables for the training phase. The observed measures of precipitation are derived from the so-called morphing technique of NOAA CPC (CMORPH) that is described in some detail in Joyce *et al* (2004) and within the references stated therein. In the light of work by Krishnamurti *et al* (2000b, 2001), CMORPH dataset is used for training and validation of a new precipitation superensemble product in real-time.

4. Results and discussions

The anomaly correlation and the RMS errors for the sea level pressure, 500 hPa level geopotential heights, and the zonal wind at the 850 hPa level for the Asian monsoon domain covering the monsoon months June, July, August and September (2004) are illustrated in figure 2 (a through f). The left panels carry the results for the anomaly correlation, and the right panels carry the RMS error. Figure 3 (a through f) illustrates the anomaly correlation and RMS error for the different components of the wind fields (i.e., the meridional wind at 850 hPa level, the zonal wind at the 200 hPa level, and the meridional wind at the 200 hPa level). We note again similar very high skills for the anomaly correlations, > 0.9 for day 4 of forecast for all these variables. In all cases, the RMS errors are the least for the multimodel superensemble compared to those of the member models and the ensemble. Forecasts for all these elements show very high skills for the multimodel superensemble that is shown by the dark bars. These carry the highest anomaly correlation and the lowest RMS errors for each of the forecast days. Here the skills for seven of the best models are compared with those of the superensemble (far right) and for the ensemble means (shown next to the superensemble). The results of the superensemble appear clearly better than those of the ensemble mean. The most striking results are the large improvements in the anomaly correlation from the multimodel superensemble, values as high as 0.90 to 0.94 on day 4 of the forecast are worth noting. The implications of such scores are discussed below.

Figure 4(a–b) (over plots) shows a superposition of the observed and predicted sea level pressure and 500 hPa geopotential heights for a day 4 of forecasts when the anomaly correlation was of the order of 0.92. The isobars of the forecasts, line for line, seem to lie on top of the observed isobars. We have chosen anonymous dates to show the pattern matching. The entire season of forecasts for day 4 and day 5 carry high anomaly correlations (see figure 4c–d). Figure 4(c–d) depicts time series of the anomaly correlation of day 4 and day 5 MSLP forecast using the multimodel superensemble along with member models. This shows very high skills of the daily values. The member models do not carry that degree of consistency in their forecasts through days 4 and 5. These skills show a spread between 0.35 and 0.85 on different days whereas that of superensemble stays close to 0.9 or higher in most days.

Monsoon lows and depressions are large scale disturbances with scale > 1000 km. These are well resolved at the resolution of the multimodel superensemble (whose horizontal resolution is of the order of 80 km). The consistent skill of the superensemble provides a useful guidance for the mesoscale modelers. Figure 5 was checked during day 4 where one such disturbance propagated across the Indian subcontinent (dotted lines are analysis while solid lines are day 4 forecast of multimodel superensemble) that occurred during 11 to 13 June 2004. Checking the anomaly correlation for these dates from figure 4(c–d), we noted that in most of these instances the day 4 and day 5 of forecast skills of the anomaly correlations was around 0.9 for the multimodel superensemble. This shows that it is possible to predict the isobaric distribution during disturbance passages with a very high degree of accuracy. Figure 5 was deliberately selected to show the superposition of isobars (forecast and observed) on day 4 of the forecast where a depression was present over northern India. This shows the usefulness of the superensemble product. A high skill forecast for up to day 5 may not meet the needs of agriculture where at least two weeks forecast is generally demanded. However, it is satisfying to see that we can provide a very useful guide to the mesoscale modelers through day 5. This overall consistent improvement of day-to-day skill for the superensemble is the special feature of these real-time products.

In the context of the usefulness of the forecasts of monsoon, precipitation is the most important variable. Precipitation forecasts are generally evaluated using a probabilistic skill, which is called the Equitable Threat Scores (see appendix 1). Skill score is evaluated for rainfall thresholds greater than a preset value. A bias score is also generally evaluated for each of the thresholds of rainfall amounts. The normalized bias score (see appendix 1) is designed such that a value of 1.0 defines a good forecast. Overestimates/underestimates of rainfall amounts carry bias score > 1 or < 1. In figure 6(a–d), we present these results of equitable threat scores and the bias scores for the season (June to September 2004). We show here the equitable threat scores in the range of thresholds

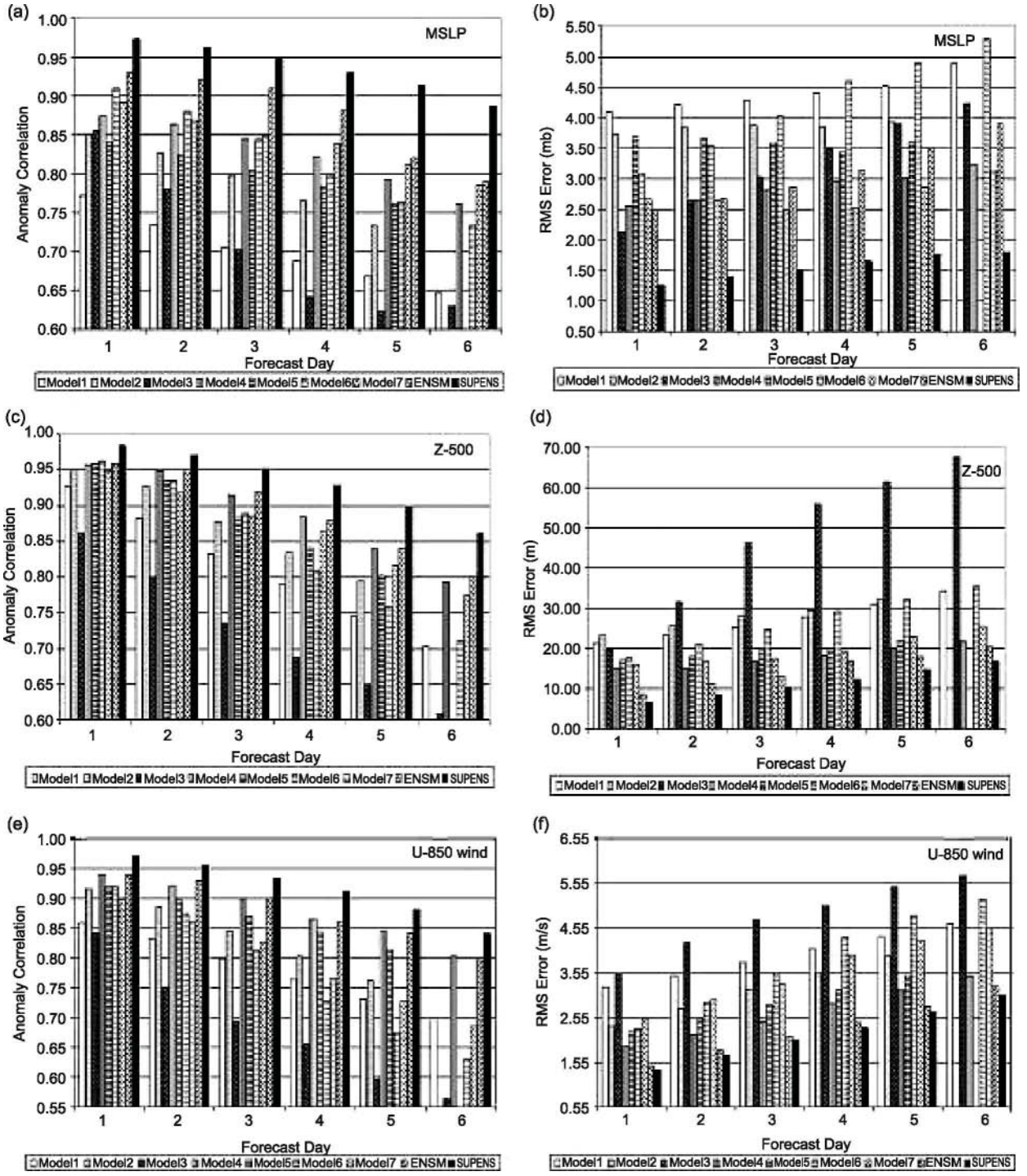

Figure 2. Forecast skills of the multimodel FSU superensemble compared with member models.

between 0.2 and 2.0 mm/day. The highest ETS are clearly carried by the multimodel superensemble compared to the individual models for day 1 through 5 of forecasts. The precipitation bias score of the multimodel superensemble is generally close to 1.0 for day 1 through 5 of forecasts; one of the member models does have a very high bias score, the performance of the bias score of the multimodel superensemble is close to that of the best model. A threshold of 0.2 mm/day implies the skill for the total rain whereas the thresholds 2 mm/day ignores the trace rain that many member models have an abundance of. The day 1 ETS for most member models ranges from 0.2 to 0.35. With the multimodel superensemble, it is possible to retain a threat score greater than or equal to 0.3 through day 4 of forecasts. How good is an ETS of 0.3? That is the current now casting

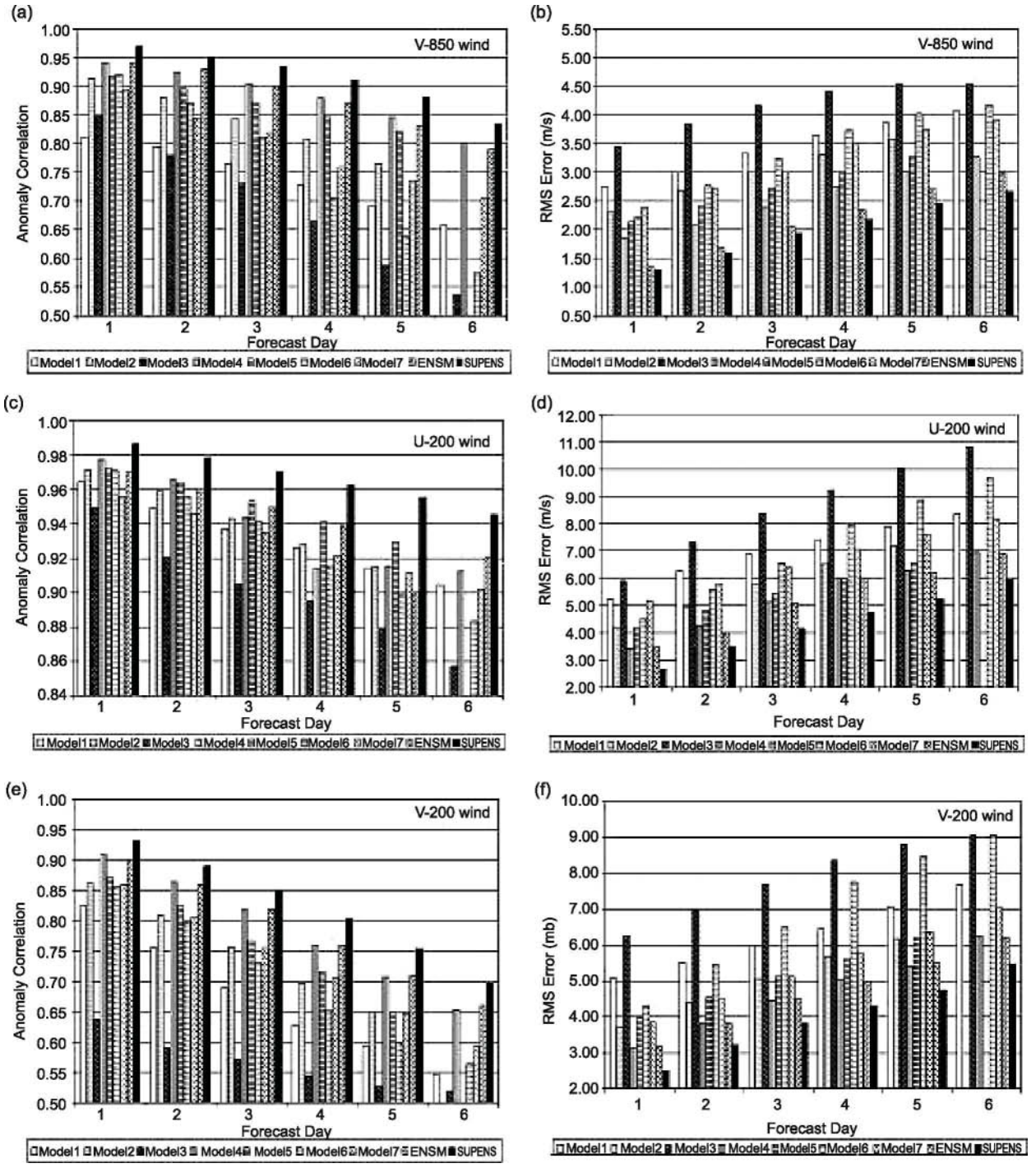

Figure 3. Forecast skills of the multimodel FSU superensemble compared with member models.

skill for most operational models. This score does signify some degree of usefulness of forecast over a large-scale model whose horizontal resolution is of the order of 100 km. This skill cannot be directly compared to that of mesoscale models that are generally evaluated against radar-based rainfall estimates.

The ETS and bias scores for a case study of heavy rains from the passage of a monsoon depression are shown in figure 6(e–h). Here the initial day 1 skills of the ETS for the multimodel superensemble for rainfall thresholds > 2.0 and > 5.0 mm/day are both close to 0.4. These skills are clearly higher than those for the member models. For this case study, there were noted minima of skill for day 3 of forecast. That sharp drop of skill between day 2 and day 3 of the member models was the period of very heavy rains where their bias scores (for days 2 and 3) showed values somewhat higher than 1.2. The bias score of the superensemble remained close

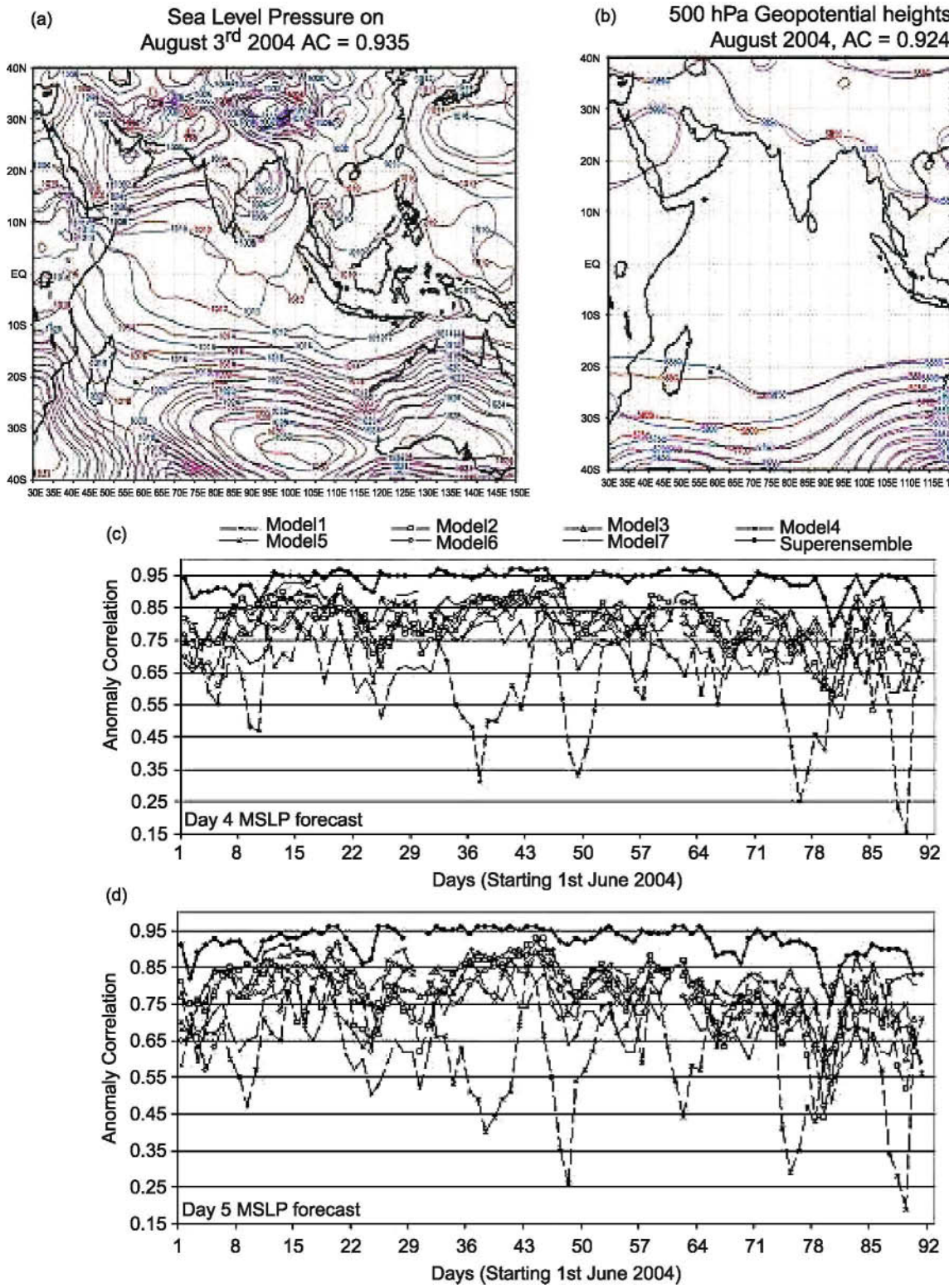

Figure 4. **(a)** Over plot of day 4 superensemble SLP forecast (red line) with analysis (blue line) on 3 August 2004, **(b)** Over plot of day 4 superensemble 500 hPa geopotential heights forecast with analysis on 25 August 2004, **(c)** Day 4 MSLP anomaly correlation for 2004 monsoon season and **(d)** same as (c) except for day 5.

to 1.0 (for days 4 and 5 of forecasts). Overall, for this case study, the bias score of the multimodel superensemble were clearly better than those of the member models.

5. Conclusion

In medium range real-time global weather forecasts, the largest skill improvements are seen for

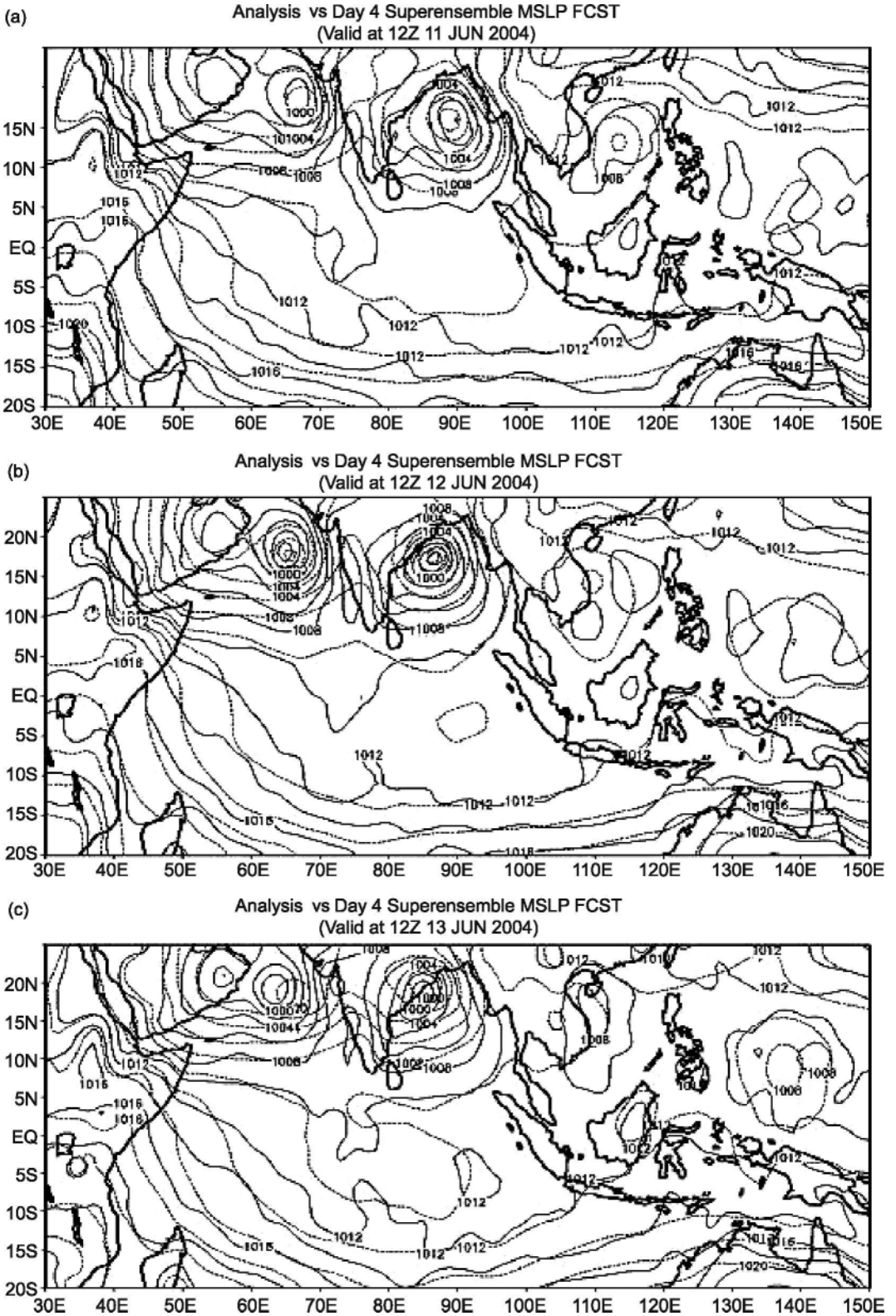

Figure 5. Analysis (dotted line) plotted against day 4 MSLP superensemble forecast (continuous line) during the passage of monsoon depression **(a)** 11 June, **(b)** 12 June and **(c)** 13 June 2004.

precipitation forecasts both regionally and globally. The overall skill of the superensemble is 40–120% higher than the precipitation forecast skills of the best global models for precipitation forecasts. The superensemble shows major improvements in skill for the divergent part of the wind and the temperature distributions. Tropical latitudes show major improvements in daily weather

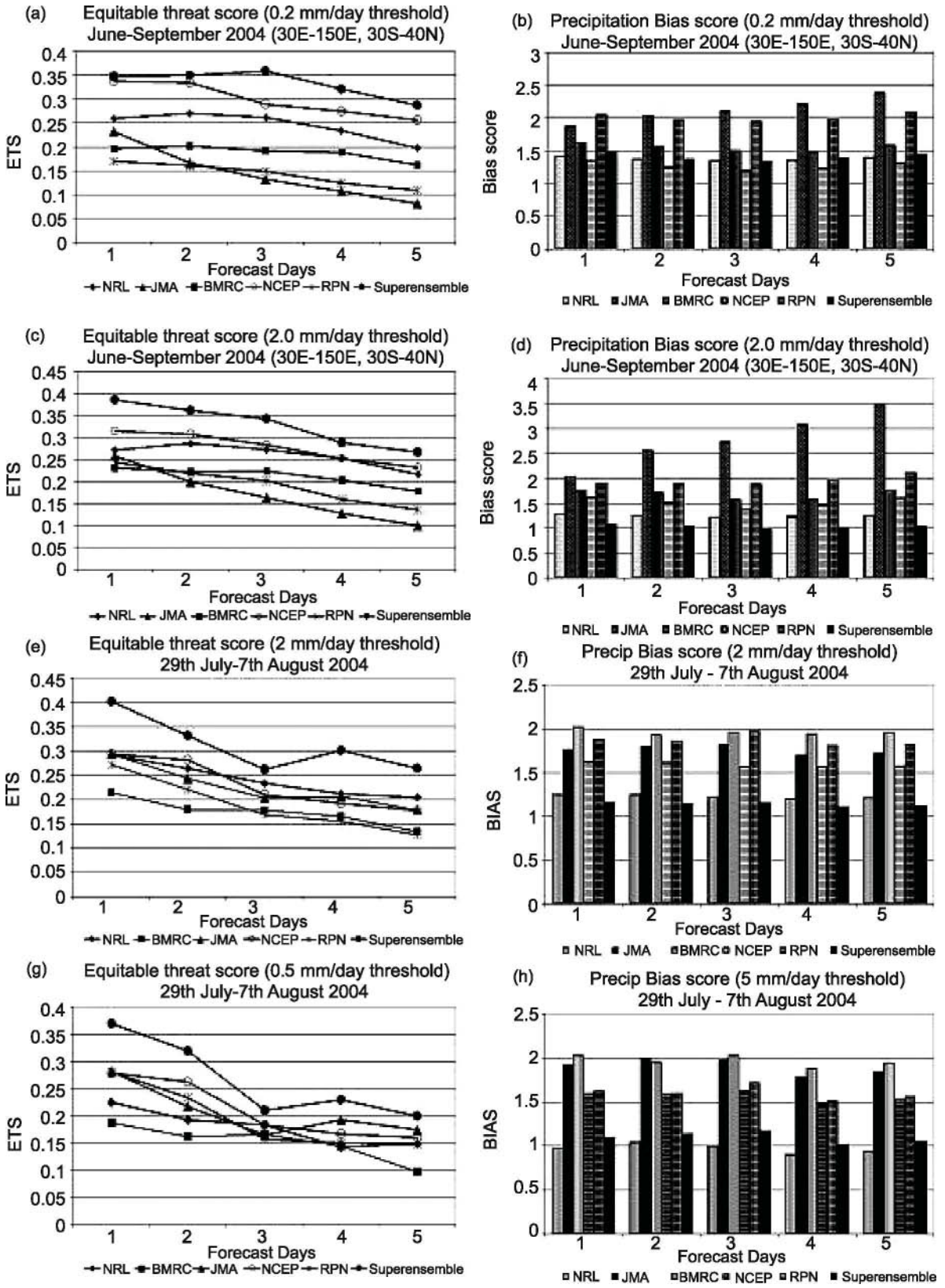

Figure 6. Forecast skills of multimodel superensemble precipitation compared against the member models.

forecasts. It has been possible to maintain a real-time website (http://lexxy.met.fsu.edu/rtnwp/) for global numerical weather prediction from the construction of a multimodel superensemble. These are the current best known global operational NWP models. It is possible to obtain the highest

skills on a daily basis out to 6 days both globally and regionally from the superensemble. The fields we have looked at include the winds, temperature, mean sea level pressure, geopotential heights at 500 hPa, humidity, and precipitation. The skill matrices include RMS error, anomaly correlation, and probabilistic measure such as equitable threat scores and bias score for precipitation. We have noted some extraordinary improvements in skill over all models. The overlaying of the geopotential height isopleths over the globe for day 5 of forecasts and corresponding analysis (observation) display a line-for-line match. These carry anomaly correlation that exceeds 0.9 on day 5. These improvements in global NWP are well beyond those that were possible a few years ago. We have also seen major improvements in the forecasts over the southern hemisphere where the skills of forecasts are close to those of northern hemisphere.

The monsoon forecasts by member models especially for monsoon rainfall were one of the difficult areas for the global NWP. Here the errors of the member models were quite large. The use of the multimodel superensemble reduced these errors considerably, and it became possible to obtain much higher equitable threat scores for the day 1 through day 5 forecasts over a monsoon domain. On day 5 of forecast, it became possible to consistently maintain anomaly correlations for sea level pressure over the monsoon region at values close to 0.92. The member models best values were considerably lower. This illustrates the robust strength of the multimodel superensemble. Similar results were noted for the wind components at 850 hPa and 200 hPa levels.

The reduction of collective bias error by the superensemble (derived from the statistics of the training phase) makes it possible to reduce errors in the real time forecast phase. We have also examined the error of the ensemble mean for some variables. We noted that those were generally lower compared to those of the member models. However, when compared to the errors of the multimodel superensemble the errors of the ensemble mean were higher consistently. The ensemble mean assigns weights of $1/N$ (N being the number of member models) uniformly over the globe for all models. The superensemble, in this regard, is very selective regionally, and assigns fractional, positive and even negative weights at different grid locations for the member models. In all we make use of as many as 10^7 weights for 3-dimensional grid location ($384 \times 142 \times 10$) for roughly 7 variables and 7 member models. The error correction plays a major role in providing the best global NWP product on real time. We also illustrate a case study of a monsoon depression where the improvements of forecast details from the superensemble over those of the member models are highlighted.

Acknowledgements

We gratefully acknowledge the ECMWF for providing analysis. We wish to acknowledge the member modeling groups whose forecast datasets were used in this paper. The research reported here was supported by NSF grant numbers ATM-0419618 and NAG5-13563. We are also thankful to our reviewers for their useful suggestions.

Appendix 1

Equitable Threat Score

The equitable threat score (Schaefer 1990) is defined as

$$ETS = (H - CH)/(F + O - H - CH)$$

where F = the number of grid boxes that forecast more than the threshold, O = the number of grid boxes that observe more than the threshold, H = the number of grid boxes that correctly forecast more than the threshold, CH = the expected number of correct forecasts due to chance = $F \times O/T$, where T = the total number of grid boxes inside the verification domain.

The equitable threat score seems to be a good estimate for overall forecast skill. The higher the value, the better the forecast model skill for that particular threshold. The equitable threat score can vary from a small negative number to 1.0, where 1.0 represents a perfect forecast. This is basically the ratio of the correct forecast area to the total area of the forecast and observed precipitation. The model gets penalized for forecasting rain in the wrong place as well as not forecasting rain in the right place. Thus, the model with the highest score is generally the model with the best forecast skill.

Bias score

The bias score is a very simple equation, defined as simply as F/O. This score does not comment at all on the skill of a model forecast in terms of the placement of precipitation, but does give an indication if a model is consistently over- or under-forecasting areas of precipitation. The best model is generally the one that remains near the 1.0 line, which means that the model does not generally over-forecast precipitation or under-forecast precipitation. If the model verifies over 1.0, it is over-predicting precipitation, and if below 1.0, it is under-predicting precipitation.

List of Acronyms

3D-VAR	Three Dimensional Variational Data Assimilation
4D-VAR	Four Dimensional Variational Data Assimilation
AGCM	Atmospheric General Circulation Model
AIRS	Advanced InfraRed Sounder
AMSU	Advanced Microwave Sounding Unit
BMRC	Bureau of Meteorology Research Centre, Australia
CMC	Canadian Meteorological Centre
CMORPH	CPC MORPHing technique
DMSP	Defense Meteorological Satellite Program
ECMWF	European Centre for Medium-Range Weather Forecasts
EMC	Environmental Modeling Center
ERS	European Remote Sensing
FNMOC	Fleet Numerical Meteorology and Oceanography Center
FSU	Florida State University
GASP	Global AnalysiS and Prediction
GEM	Global Environmental Multiscale Model
GFS	Global Forecast System
GOES	Geostationary Operational Environmental Satellites
GOMOS	Global Ozone Monitoring by Occultation of Stars
GSM	Global Spectral Model
HIRS	High Resolution Infrared Radiation Sounder
ITCZ	Inter Tropical Convergence Zone
JMA	Japan Meteorological Agency
MODIS	Moderate Resolution Imaging Spectroradiometer
MSU	Microwave Sounding Unit
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Center for Environmental Prediction
NCMRWF	National Centre for Medium Range Weather Forecasting
NESDIS	National Environmental Satellite, Data, and Information Service
NOAA	National Oceanic & Atmospheric Administration
NOGAPS	Navy Operational Global Atmospheric Prediction System
NRL	Naval Research Laboratory
QuikScat	Quick Scatterometer
RMS Error	Root Mean Square Error
RPN	Recherche Prévision Numérique
SBUV	Solar Backscatter Ultraviolet radiometer
SLP	Sea Level Pressure
SSM/I	Special Sensor Microwave Instrument
SST	Sea Surface Temperature
SVD	Singular Value Decomposition
SYNOP	Traditional Synoptic Weather Observations
TRMM	Tropical Rainfall Measuring Mission
UKMet	United Kingdom Met. Office

References

Arakawa A and Schubert W H 1974 Interaction of a cumulus cloud ensemble with the large-scale environment. Part I; *J. Atmos. Sci.* **31** 674–701.

Businger J A, Wyngard J C, Izumi Y and Bradley E F 1971 Flux profile relationship in the atmospheric surface layer; *J. Atmos. Sci.* **28** 181–189.

Charnock H 1955 Wind stress over a water surface; *Quart. J. Roy. Meteor. Soc.* **81** 639–640.

Edwards J M and Slingo A 1996 Studies with a flexible new radiation code I: choosing a configuration for a large-scale model; *Quart. J. Roy. Meteor. Soc.* **122** 689–719.

Ferraro R R and Marks G F 1995 The development of SSM/I rain-rate retrieval algorithms using ground-based radar measurements; *J. Atmos. Oceanic Technol.* **12** 755–770.

Grell G A 1993 Prognostic evaluation of assumptions used by cumulus parameterizations; *Mon. Wea. Rev.* **121** 764–787.

Harshvardan and Corsetti T G 1984 Longwave parameterization of the UCLA/GLAS GCM; *NASA Technical Memorandum* 86072, Goddard Space Flight Center, Greenbelt, MD, 52pp.

Harshvardhan, David A Randall and Thomas G Corsetti 1989 Earth Radiation Budget and Cloudiness Simulations with a General Circulation Model; *J. Atmos. Sci.* **46** 1922–1942.

Joyce R J, Janowiak J E, Arkin P A and Xie P 2004 CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution; *J. Hydromet.* **5** 487–503.

Krishnamurti T N 1979 Tropical meteorology. *WMO Publ. No. 364, Compendium of Meteorology*, Part 4, 428pp.

Krishnamurti T N, Low-Nam S and Pasch R 1983 Cumulus parameterization and rainfall rates II; *Mon. Wea. Rev.* **111** 816–828.

Krishnamurti T N, Xue J, Bedi H S, Ingles K and Oosterhof D 1991 Physical initialization for numerical weather prediction over the tropics; *Tellus* **43** 53–81.

Krishnamurti T N, Bedi H S and Ingles K 1993 Physical initialization using the SSM/I rain rates; *Tellus* **45A** 247–269.

Krishnamurti T N, Kishtawal C M, LaRow T E, Bachiochi D R, Zhang Z, Williford C E, Gadgil S, Surendran S 1999 Improved weather and daily and medium Range climate forecasts from multi-model superensemble; *Science* **285** 1548–1550.

Krishnamurti T N, Kishtawal C M, Shin D W and Williford C E 2000a Multimodel ensemble forecasts for weather and seasonal climate; *J Climate* **13** 4196–4216.

Krishnamurti T N, Kishtawal C M, Shin D W and Williford C E 2000b Improving tropical precipitation forecasts from a Multianalysis Superensemble; *J Climate* **13** 4217–4227.

Krishnamurti T N, Surendran S, Shin D W, Correa-Torres R J, Vijay Kumar T S V, Williford E, Kummerow C, Adler R F, Simpson J, Kakar R, Olson W S and Turk F J 2001 Real-time multianalysis-multimodel superensemble forecasts of precipitation using TRMM and SSM/I products; *Mon. Wea. Rev.* **129** 2861–2883.

Krishnamurti T N, Rajendran K, Vijaykumar T S V, Stephen Lord, Zoltan Toth, Xiaolei Zou, Steven Cocke, Jon E Ahlquist and Michael Navon I 2003 Improved Skill for the Anomaly Correlation of Geopotential Heights at 500 hPa, *Mon. Wea. Rev.* **131** 1082–1102.

Kuo H L 1974 Further studies of the parameterization of the influence of cumulus convection on large-scale flow; *J. Atmos. Sci.* **31** 1232–1240.

Lacis A A and Hansen J E 1974 A parameterization for the absorption of solar radiation in the Earth's atmosphere; *J. Atmos. Sci.* **31** 118–133.

Louis J F 1979 A parametric model of vertical eddy fluxes in the atmosphere; *Bound.-Layer Meteor.* **17** 187–202.

Louis J F, Tiedtke M and Geleyn J F 1982 A short history of the operational PBL parameterization at ECMWF; *ECMWF Workshop on Planetary Boundary Parameterizations*, 59–79.

Moorthi S and Suarez M J 1992 Relaxed Arakawa–Schubert: a parameterization of moist convection for general circulation models; *Mon. Wea. Rev.* **120** 978–1002.

Palmer T N, Shutts G J and Swinbank R 1986 Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parameterization; *Quart. J. Roy. Meteor. Soc.* **112** 1001–1039.

Pan Hua-lu and Wan-Shu Wu 1994 Implementing a mass flux convective parameterization package for the NCEP Medium-range forecast model. Preprints of the *10th AMS Conf. on Num. Weather Pred.*, July 18–22, 1994, Portland OR (Pan H-L and Wu W-S 1995 NMC Office Note, No. 409, 40pp.)

Rao Y P 1976 South West Monsoon *IMD Publications.*

Schaefer J T 1990 The critical success index as an indicator of warning skill; *Wea. Forecasting* **5** 570–575.

Schwarzkopf M D and Fels S B 1991 The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes; *J. Geophys. Res.* **96** 9075–9096.

Troen I and Mahrt L 1986 A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation; *Bound.-Layer Meteorol.* **37** 129–148.

Wilks D S 1995 *Statistical Methods in the Atmospheric Sciences*, Academic Press, 467pp.

Yanai M, Esbensen S and Chu J 1973 Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets; *J. Atmos. Sci.* **30** 611–627.

Yun W T, Stefanova L, Mitra A K, Vijay Kumar T S V, Dewar W and Krishnamurti T N 2005 Multi-Model Synthetic Superensemble Algorithm for Seasonal Climate Prediction using DEMETER Forecasts; *Tellus* **57** 280–289.

MS received 20 March 2006; revised 10 March 2007; accepted 10 March 2007