



A SST based large multi-model ensemble forecasting system for Indian summer monsoon rainfall

A. K. Sahai,¹ R. Chattopadhyay,¹ and B. N. Goswami¹

Received 28 July 2008; revised 19 August 2008; accepted 26 August 2008; published 3 October 2008.

[1] An ensemble mean and probabilistic approach is essential for reliable forecast of the All India Summer Monsoon Rainfall (AIR) due to the seminal role played by internal fast processes in interannual variability (IAV) of the monsoon. In this paper, we transform a previously used empirical model to construct a large ensemble of models to deliver useful probabilistic forecast of AIR. The empirical model picks up predictors only from global sea surface temperature (SST). Methodology of construction implicitly incorporates uncertainty arising from internal variability as well as from the decadal variability of the predictor-predictand relationship. The forecast system demonstrates the capability of predicting monsoon droughts with high degree of confidence. Results during independent verification period (1999–2008) suggest a roadmap for generating empirical probabilistic forecast of monsoon IAV for practical delivery to the user community. **Citation:** Sahai, A. K., R. Chattopadhyay, and B. N. Goswami (2008), A SST based large multi-model ensemble forecasting system for Indian summer monsoon rainfall, *Geophys. Res. Lett.*, 35, L19705, doi:10.1029/2008GL035461.

1. Introduction

[2] Although the complexity involved in the interannual variability (IAV) of all India summer monsoon rainfall (AIR) in the form of large scale droughts and floods makes its accurate prediction a challenging task [Goswami *et al.*, 2006, Xavier and Goswami, 2007], there is a great demand for long range forecast of AIR from policy makers as it adversely affects country's agricultural production, economy and GDP [Gadgil and Gadgil, 2006]. For this reason, efforts on prediction of IAV of AIR started more than a century ago [Blanford, 1884] and received immense attention in past three decades through use of empirical as well as dynamical models. Prediction of seasonal mean AIR is useful during extreme monsoon years (droughts and floods) when the rainfall anomaly is homogeneous over the country. However, it is not very useful for any regional hydro-meteorological applications during 'normal' monsoon years, when the rainfall anomaly is quite inhomogeneous within the country [Xavier and Goswami, 2007]. Any long range prediction system for AIR, therefore, should have useful skill in predicting the extremes. Most empirical models [Gowariker *et al.*, 1989; Iyengar and Raghukanth, 2004; Goswami and Srividya, 1996; Rajeevan *et al.*, 2004; Gadgil *et al.*, 2005] tend to be biased towards predicting the

'average' climate and fail to predict extremes with useful skill. Currently, most dynamical models also have poor skills in predicting the AIR [Kang *et al.*, 2004; Kang and Shukla, 2005; Krishna Kumar *et al.*, 2005]. Multi-model super-ensemble forecasting [Krishnamurti *et al.*, 1999; Chakraborty and Krishnamurti, 2006] shows promise of improving the dynamical forecasts beyond the skill of individual models. However, further significant improvement of skill would be required before these models could be used for operational long range predictions.

[3] The basis for long range predictability of IAV of monsoon comes from slowly varying large scale external boundary forcing [Charney and Shukla, 1981; Shukla, 1998; Goswami and Xavier, 2005] arising from ocean-atmosphere interactions, however it is limited by 'internally' generated IAV arising from convective feedback and scale interactions involving fast processes [Goswami *et al.*, 2006]. Many studies [Goswami, 1998; Kang *et al.*, 2004; Cherchi and Navarra, 2006; Goswami and Xavier, 2005] have brought out that the challenge in predicting the Indian summer monsoon arises from the fact that 'internal' IAV contributes to a large fraction of IAV of the Indian summer monsoon. The seminal role played by 'internal' variability in the predictability of AIR demands a probabilistic approach for prediction of AIR. As a result of limited skill of current dynamical models in predicting AIR, empirical models are still needed to provide useful guidance [Rajeevan *et al.*, 2007]. However, all empirical models are generally used as deterministic ones for predicting the AIR [e.g., Gowariker *et al.*, 1989; DelSole and Shukla, 2002; Sahai *et al.*, 2003; Goswami and Srividya, 1996; Iyengar and Raghukanth, 2004; Kishtawal *et al.*, 2003]. The predictor-predictand relationship exploited in most of these models essentially try to capture the predictable teleconnection patterns but contains no method to incorporate the uncertainty arising due to sensitivity to initial conditions. Another uncertainty in these models arises from the non-stationarity and inter-decadal variability of the predictor-predictand relationship [Sahai *et al.*, 2000; Rajeevan, 2001; Krishna Kumar *et al.*, 1999]. In the context of this background, it is logical to shift from the current deterministic approach in operational forecasting and formulate an ensemble prediction strategy using empirical models that takes into account such uncertainties and allow generation of a probabilistic forecast of the predictand [e.g., Rajeevan *et al.*, 2004]. For such a system of forecast to be useful it must have high confidence in predicting the monsoon extremes such as droughts and floods. In this study, a strategy to construct such a large multi-model ensemble for predicting AIR based entirely on past sea surface temperature (SST) is presented. Unlike earlier empirical models, this probabilistic forecast system

¹Indian Institute of Tropical Meteorology, Pune, India.

is also capable of predicting the monsoon extremes (droughts and flood) with higher degree of confidence.

2. The Strategy for Constructing the Multi-Model Ensemble, Data and Methodology

[4] Instead of using different parameters to represent different teleconnections with the Indian monsoon, *Sahai et al.* [2003] constructed an empirical model for long lead prediction of AIR where all predictors are SST at a number of different geographical locations with lags varying up to 4 years. The model showed good skill in the independent verification period as well as during real-time forecast of AIR (generated in the month of March) for the last six years [*Sahai et al.*, 2007] (auxiliary material¹, Table S4). This strategy is based on the hypothesis that footprint of all conceivable teleconnections with monsoon could be found in some geographical locations of global SST with appropriate lag as all teleconnections owe their origin to some slow coupled ocean-atmosphere oscillation. The multiple regression (MR) technique used in this model, however, could neither include uncertainty in the prediction arising from non-stationarity of the predictor-predictand relationship nor from ‘internal’ variability. Continuing with the philosophy of using SST alone to represent all possible teleconnections, here we present a new methodology to create a large ensemble of empirical models. Data used and details of the method are described below.

[5] Monthly SST data taken from ERSSTv2 [Smith and Reynolds, 2004] were averaged over boxes of 10° latitude \times 20° longitude whose centers are 5° latitude \times 10° longitude apart [*Sahai et al.*, 2002, 2003] and monthly mean rainfall data constructed from well distributed 1476 rain gauge stations were taken from IMD [*Guhathakurta and Rajeevan*, 2007]. AIR was then constructed from JJAS seasonal rainfall as percentage departure of long term mean (LTM). The spatial extent of SST data were considered for the region 30° S to 50° N and temporal extent from March 1918 to February 2008. Following steps are taken to construct the models.

[6] a) The 78 year period from 1921 to 1998 is taken in development set and 9 year period from 1999 to 2007 in forecast set. The development set is further divided into two sets, a training set and a test set. The training set consists of randomly selected 58 years out of 78 and the test set remaining 20 years. This random partitioning is done 10,000 times.

[7] b) Correlation coefficients, CC, were calculated between SST (seasonal average of three months DJF, MAM, JJA and SON, persisted for two season DJF+MAM, MAM+JJA, JJA+SON, SON+DJF and for seasonal tendency MAM-DJF, JJA-MAM, SON-JJA, DJF-SON) of each 10° \times 20° grid of ERSST data and AIR. CC was calculated from 1 season lag to 3 years lag. This is expected to capture the biennial and ENSO variability. Regions and seasons were identified, if CC is significant at 5% level for the training and the test set and of the same sign.

[8] c) Best predictors were selected based on leave-one-out cross validated step wise regression selection method in

the training set [*DelSole and Shukla*, 2002]. The whole procedure was repeated for all randomly selected 10000 training sets, resulting in those many models. This is expected to overcome the non-stationarity problem of predictor-predictand relationship.

[9] d) Hindcasts were obtained from all 10,000 models for the development period (1921–1998) using leave-one-out cross validation MR as suggested in a WMO manual [*World Meteorological Organization (WMO)*, 2006] on the Standardised Verification System for Long-Range Forecasts (SVSLRF) and independent forecasts for the forecast period (1999–2008) is obtained using MR equation developed on the whole development period. It is a general tendency of the empirical models that the forecasts are largely concentrated near normal resulting in smaller variance than the observations. Therefore all forecasted values were subjected to bias removal and variance correction with the following formula:

$$F_{\text{inf}} = (F - \bar{F}) \times (O_{sd}/F_{sd}) + \bar{O}$$

where F and O are forecasted and observed values, \bar{F} and \bar{O} are their means over the development period and F_{sd} and O_{sd} are standard deviations respectively. F_{inf} is the variance inflated bias corrected forecast.

[10] e) So far we have not incorporated decadal and inter-decadal variability in the model. This is introduced by selecting 1,000 best models out of 10,000 models based on their performance in the immediate past 30 years [*Sahai et al.*, 2002; *DelSole and Shukla*, 2002]. For example, to obtain the forecast of year 1999, RMSE were calculated for the forecasts from 1969 to 1998 and 1,000 models with lowest RMSE were selected. Similarly, 1,000 forecasts were obtained and their averages were calculated for other years.

[11] By selecting the SST-AIR relationship in an ensemble of subset of randomly selected years, an attempt is made to model the spread in the predictions due to internal variability through the spread in predictions of the ensemble of models. A detailed example of how a model is constructed may be found in supporting online material.

3. Results and Discussion

[12] In order to get a feeling of why such large number of dynamic predictors are required, we consider two forecasts for the years 2002 and 2004 obtained from selecting the area averaged SST (as predictors) for two different training sets as shown in Figures 1a and 1b. The two sets of predictors, however, are so chosen that both of them gave good forecast of AIR during the model development period (1921–1998) as can be seen from Figure 1c. Both these forecasts (Set 1 and Set 2) have identical correlation (0.62) and RMSE (7.2%) with respect to observation for the same period, but one of them gave correct estimate in the independent validation years (2002 and 2004) while the other did not. As discussed in the introduction, such a discrepancy is, however, unavoidable because of the secular variability in the predictability and uncertainty in the models. In this case, such a decadal variability can be seen in a 11 year running correlation of the two sets of predictors with respect to observation (Figure 1d). At times both the predictors are able to predict correctly but for certain period

¹Auxiliary materials are available in the HTML. doi:10.1029/2008GL035461.

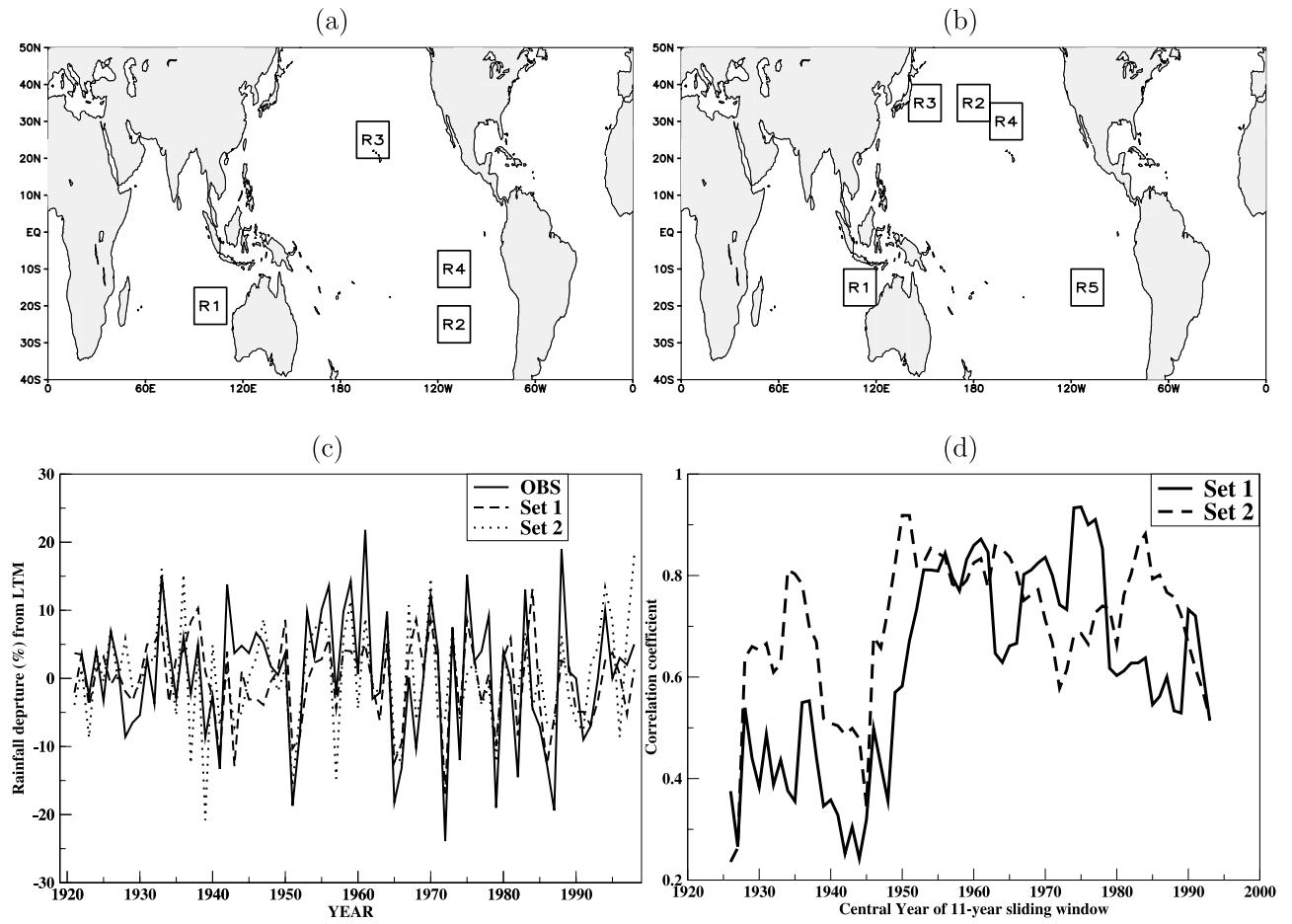


Figure 1. Example of selection of different regions shown in (a) Set 1 and in (b) Set 2 for different years in the training period. The area averaged SST chosen as predictors from these regions (R1, R2,..etc.) include different lag seasons (see auxiliary material, Table S3). Multiple regression equations were calculated in the training period for the predictors in both the sets. Predicted rainfall values (c) for the period 1921–1998 gave almost equal correlation coefficient (0.63) and root mean square error (7.2%) with the observed rainfall in both the cases. (d) Interdecadal variation in predictability depicted through 11-year running CC also shows similar behavior for both set of predictors. Despite the similar performance in development period, the predicted values in 2002 and 2004 were (−17 and −6 %) for Set 1 and (+11 and +9%) for Set 2 against observation (−19 and −13%).

(including the years 2002 and 2004) both have different level of efficiency in capturing the actual variability. Thus, a probabilistic forecast from an ensemble of many independent forecasts seems to be a logical choice. In addition to this a more confident deterministic forecast can also be obtained from these as an ensemble mean.

[13] As described in the last section, 1,000 hindcasts for the period 1951–1998 and those many forecasts for the period 1999–2008 were generated for the AIR using the dynamic SST predictor-predictand relationship. Thus we have 57 years (1951–2007) of independent 1,000 forecasts for verification purpose. In the manual [WMO, 2006] on the Standardized Verification System for Long-Range Forecasts (SVSLRF) of WMO it has been suggested to plot the Relative Operating Characteristics (ROC) curve and the area under curve (AUC) as the skill score for verification of multi-category probabilistic forecasts. The ROC AUC is characterized as 1.0 as a perfect forecast and 0.5 as a climatology forecast (no information). For AIR, 5 categories

(shown in Table 1) have been defined by India Meteorology Department (IMD). The ROC curves for all these five categories are plotted in Figure 2a. The ROC curves are based on the ROC contingency tables obtained by binning the forecast probabilities into 10 equal bins of 0–10%, 10–20%, ..., 90–100%. The curves depict the high success rate of our multi-model ensemble approach. The skill score (ROC AUC) and 5% significance value obtained from bootstrap procedures are also shown in Figure 2a. The skill score is significantly high for all the categories, except the AN category. It is worth noting that the skill of the model is exceptionally good for predicting droughts.

Table 1. Limits of the Five Mutually Exclusive Categories for the Rainfall Departure as a Percentage of LTM

Category	Drought (DR)	Below Normal (BN)	Near Normal (NN)	Above Normal (AN)	Flood (FL)
Limits	≤ -10	> -10 and < -4	< -4 and ≥ 4	> 4 and < 10	≥ 10

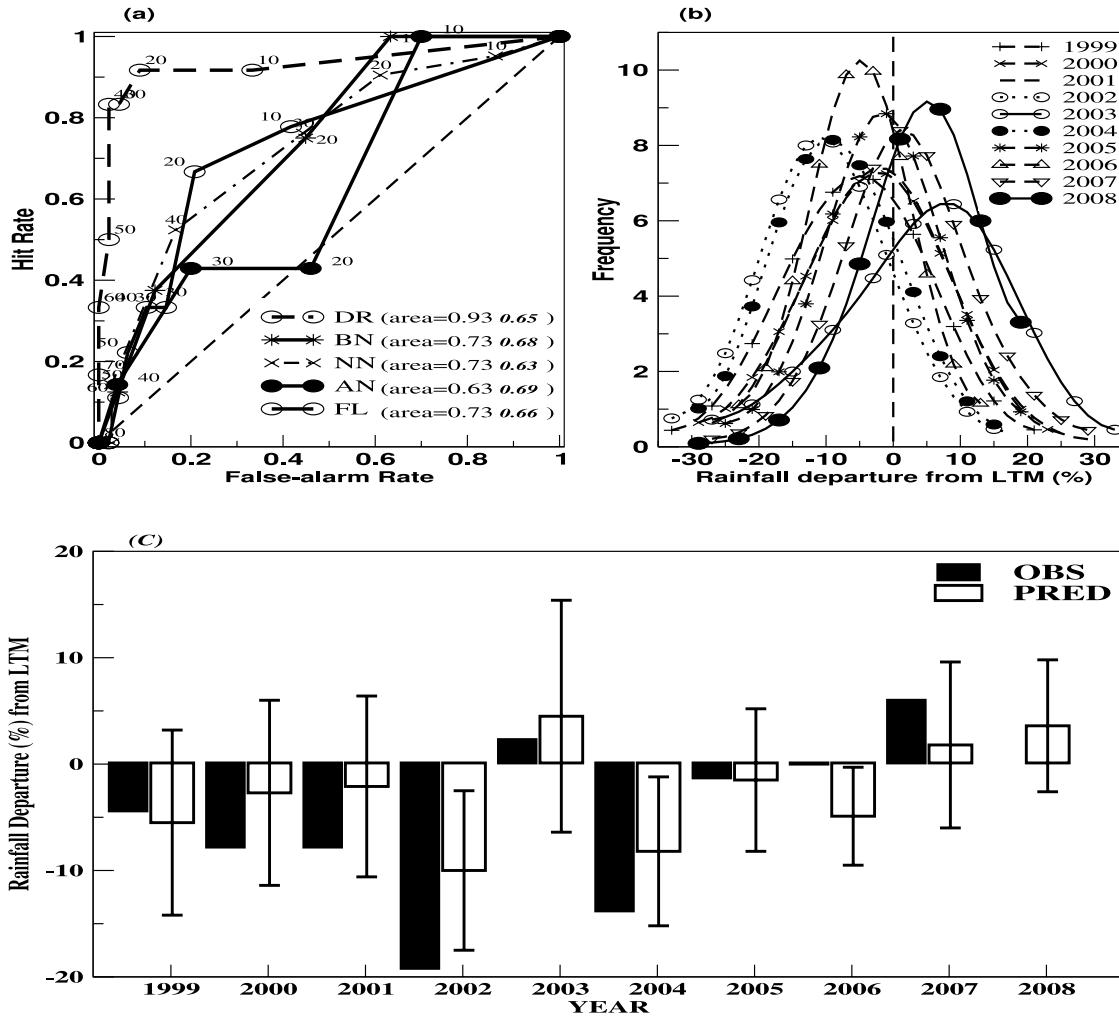


Figure 2. (a) The ROC curve constructed on the basis of forecast categories described in Table 1 for 1000 hindcast. Area under the curve (AUC) for each category is also denoted in braces along with their 95% significance level (in bold italics) for comparison. (b) Probability distribution functions of forecasts for each forecast year from 1999 to 2008. (c) Ensemble mean of 1000 forecasts and their spread represented as one standard deviation limit (thin bars) are shown along with the observed rainfall for the forecast period.

[14] To get a clearer idea of the probabilistic forecast skill of the current model, we have plotted probability distribution function (PDF) in Figure 2b and also shown the percentage frequencies for different categories in Table 2 for the independent forecast period (1999–2008). We see that PDFs of drought years shown in dotted line in Figure 2 (2002 and 2004) are clearly separated with good monsoon (continuous line) and normal (dashed line) years. The same information can be extracted from Table 2. Thus the rainfall departure in drought years is well predicted with higher degree of confidence. Such forecast would be valuable for policy decision, water resource and food security management. It is also worth noting that the maximum error is through one category shift from actual observation. This depicts the dynamic role of SST in the global Tropics and sub-Tropics in regulating AIR. For the year 2008, the maximum forecasts are in the excess category, thus a normal to excess monsoon is expected in this year.

[15] The multi-model ensemble mean forecasts for the forecast period (1999–2008) are shown in Figure 2c and are compared with observation. The spread amongst the ensemble members is indicated by thin bars representing one

Table 2. Percentage Probability of Occurrence of Five Categories^a

Year	Observed Category	Forecasted Category				
		DR	BN	NN	AN	FL
1999	BN	32	23	30	13	2
2000	BN	20	24	33	15	8
2001	BN	20	19	38	15	8
2002	DR	53	26	17	3	1
2003	NN	13	8	22	23	34
2004	DR	47	24	25	4	1
2005	NN	9	27	47	12	5
2006	NN	12	48	37	3	0
2007	AN	5	14	49	18	14
2008	—	3	9	38	45	5

^aBold is the most likelihood forecasted category.

standard deviation of the spread. It may be seen that the spread is comparatively less in some years (viz. 2002 and 2004) while it is more in some years (viz. 2003). This is clearer from Table 2 for the drought years 2002 and 2004 where almost all 1,000 forecasts tend to predict the seasonal droughts one season in advance. This ensemble method, thus, increases the confidence of the monsoon forecast. The observation shows that 2002 and 2004 were drought years with rainfall deficit more than 10% of LTM; 2003, 2005 and 2006 were normal years with rainfall between $\pm 4\%$ of LTM; 2007 being normal to excess year with rainfall departure between 4% to 10% surplus of LTM and the years 1999–2001 were normal to deficient years with rainfall between -4.0 to -10.0% of LTM. Thus, the last nine years include a wide spectrum of AIR variability. The model is able to foresee the variability in the observations with sufficient confidence. The skill scores determined through CC (0.9) and the RMSE (5%) are high. The probabilistic forecast indicate that for the year 2008, there is a 45% probability for the monsoon to be in AN category and 38% probability for the NN category (Table 2), while the ensemble mean obtained from the same model indicate a departure of $+3.6\%$ of LTM which shows that this year will experience a normal to excess AIR.

4. Conclusions

[16] As a result of seminal role played by internal IAV in the predictability of AIR, the need for a probabilistic approach has been recognized. While, probabilistic forecast can be made using dynamical models rather easily by making a large ensemble of forecasts with different initial conditions, no concentrated effort has so far been made to generate large ensemble of empirical forecasts to produce probabilistic forecasts of AIR. We present here a methodology to achieve this. We accomplish it by constructing a large ensemble of empirical models (in place of different initial conditions of dynamical models). With the premise that all teleconnections arise from some coupled ocean-atmosphere oscillations, the SST based predictors for AIR from different geographical locations and with various lags is considered to be footprints of these teleconnections with AIR. As the predictors for each model are based on SST-AIR relationship on a randomly selected subset of the time history, it is believed that the spread in the predictions by the model ensemble represents the uncertainty arising partly from internal variability and partly from interdecadal variability in the predictor-predictand relationship.

[17] The model allows better lead time as forecasts are made in March. The ensemble mean forecast during the recent nine year independent verification period not only shows high skill in predicting AIR but also provides level of confidence of the ensemble mean forecast. Another strength of the ensemble prediction system is its ability to predict droughts with high degree of confidence. This has been demonstrated by good predictions of droughts of 2002 and 2004 in the independent forecast validation period as well. Another advantage of the system is that the forecast (for any category) is made with a quantitative confidence level (probability) which could be used for decision making. The probability distribution function of forecast for 2008

indicates a high degree of confidence in the ensemble mean forecast of 3.6% above LTM.

References

Blanford, H. H. (1884), On the connection of Himalayan snowfall and seasons of drought in India, *Proc. R. Soc. London*, *37*, 3–22.

Chakraborty, A., and T. N. Krishnamurti (2006), Improved seasonal climate forecasts of south Asian summer monsoon using a suite of 13 coupled ocean-atmosphere models, *Mon. Weather Rev.*, *134*, 1697–1721.

Charney, J. G., and J. Shukla (1981), Predictability of monsoons, in *Monsoon Dynamics*, edited by J. Lighthill and R. P. Pearce, pp. 99–110, Cambridge Univ. Press, Cambridge, U. K.

Cherchi, A., and A. Navarra (2006), Sensitivity of the Asian summer monsoon to the horizontal resolution: Differences between AMIP-type and coupled model experiments, *Clim. Dyn.*, *28*, 273–290.

DeSole, T., and J. Shukla (2002), Linear prediction of Indian monsoon rainfall, *J. Clim.*, *15*, 3645–3658.

Gadgil, S., and S. Gadgil (2006), The Indian monsoon, GDP and agriculture, *Econ. Polit. Weekly*, *41*, 4887–4895.

Gadgil, S., M. Rajeevan, and R. Nanjundiah (2005), Monsoon prediction: Why yet another failure?, *Curr. Sci.*, *88*, 1389–1400.

Goswami, B. N. (1998), Interannual variation of Indian summer monsoon in a GCM: External conditions versus internal feedbacks, *J. Clim.*, *11*, 501–522.

Goswami, P., and P. Srividya (1996), A neural network design for long range prediction of rainfall pattern, *Curr. Sci.*, *70*, 447–457.

Goswami, B. N., and P. K. Xavier (2005), Dynamics of ‘internal’ interannual variability of the Indian summer monsoon in a GCM, *J. Geophys. Res.*, *110*, D24104, doi:10.1029/2005JD006042.

Goswami, B. N., G. Wu, and T. Yasunari (2006), Annual cycle, intraseasonal oscillations and roadblock to seasonal predictability of the Asian summer monsoon, *J. Clim.*, *19*, 5078–5099.

Gowariker, V., V. Thapliyal, R. P. Sarker, G. S. Mandal, and D. R. Sikka (1989), Parametric and power regression models: New approach to long range forecasting of monsoon rainfall in India, *Mausam*, *40*, 115–122.

Guhathakurta, P., and M. Rajeevan (2007), Trends in the rainfall pattern over India, *Int. J. Climatol.*, *28*, 1453–1469, doi:10.1002/joc.1640.

Iyengar, R. N., and S. T. G. Raghukanth (2004), Intrinsic mode functions and a strategy for forecasting Indian monsoon rainfall, *Meteorol. Atmos. Phys.*, *90*, 17–36.

Kang, I.-S. and J. Shukla (2005), Dynamical seasonal prediction and predictability of monsoon, in *The Asian Monsoon*, edited by B. Wang, pp. 585–612, Praxis, Chichester, U. K.

Kang, I. S., J. Y. Lee, and C. K. Park (2004), Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction, *J. Clim.*, *17*, 834–844.

Kishtawal, C. M., S. Basu, F. Patadia, and P. K. Thapliyal (2003), Forecasting summer rainfall over India using genetic algorithm, *Geophys. Res. Lett.*, *30*(23), 2203, doi:10.1029/2003GL018504.

Krishna Kumar, K., B. Rajgopalan, and M. K. Cane (1999), On the weakening relationship between the Indian monsoon and ENSO, *Science*, *284*, 2156–2159.

Krishna Kumar, K., M. Hoerling, and B. Rajgopalan (2005), Advancing dynamical prediction of Indian monsoon rainfall, *Geophys. Res. Lett.*, *32*, L08704, doi:10.1029/2004GL021979.

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 multimodel superensemble, *Science*, *285*, 1548–1550.

Rajeevan, M. (2001), Prediction of Indian summer monsoon: Status, problems and prospects, *Curr. Sci.*, *81*, 1451–1457.

Rajeevan, M., D. S. Pai, S. K. Dikshit, and R. R. Kelkar (2004), IMD’s new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003, *Curr. Sci.*, *86*, 422–431.

Rajeevan, M., D. S. Pai, R. A. Kumar, and B. Lal (2007), New statistical models for long-range forecasting of southwest monsoon rainfall over India, *Clim. Dyn.*, *28*, 813–828.

Sahai, A. K., M. K. Soman, and V. Satyan (2000), All India summer monsoon rainfall prediction using an artificial neural network, *Clim. Dyn.*, *16*, 291–302.

Sahai, A. K., A. M. Grimm, V. Satyan, and G. B. Pant (2002), Prospects of prediction of Indian summer monsoon rainfall using global SST anomalies, *Res. Rep. RR-093*, Indian Inst. of Trop. Meteorol., Pune, India. (Available at <http://www.tropmet.res.in/~lip/Publication/RR-pdf/RR-93.pdf>).

Sahai, A. K., A. M. Grimm, V. Satyan, and G. B. Pant (2003), Long-lead prediction of Indian summer monsoon rainfall from global SST evolution, *Clim. Dyn.*, *20*, 855–863.

Sahai, A. K., S. K. Mandke, M. A. Shinde, R. Chattopadhyay, S. Joseph, and B. N. Goswami (2007), Experimental seasonal forecast of Indian

summer monsoon 2007: Statistical and dynamical models, *Res. Rep. RR-120*, Indian Inst. of Trop. Meteorol., Pune, India. (Available at <http://www.tropmet.res.in/~lip/Publication/RR-pdf/RR-120.pdf>).

Shukla, J. (1998), Predictability in the midst of chaos: A scientific basis for climate forecasting, *Science*, 282, 728–731.

Smith, T. M., and R. W. Reynolds (2004), Improved extended reconstruction of SST (1854–1997), *J. Clim.*, 17, 2466–2477.

Xavier, P. K., and B. N. Goswami (2007), Analog method for realtime forecasting of summer monsoon subseasonal variability, *Mon. Weather Rev.*, 135, 4149–4160.

World Meteorological Organization (WMO) (2006), Standardised verification system (SVS) for long-range forecasts (LRF). New attachment II-8 to the manual on the GDPFS, *Rep. 485*, vol. 1, 83 pp., Geneva, Switzerland.

R. Chattopadhyay, B. N. Goswami, and A. K. Sahai, Indian Institute of Tropical Meteorology, Pune 411008, India. (sahai@tropmet.res.in)