

## An Analog Method for Real-Time Forecasting of Summer Monsoon Subseasonal Variability

PRINCE K. XAVIER

*Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, Paris, France*

B. N. GOSWAMI

*Indian Institute of Tropical Meteorology, Pune, India*

(Manuscript received 10 April 2006, in final form 23 January 2007)

### ABSTRACT

A physically based empirical real-time forecasting strategy to predict the subseasonal variations of the Indian summer monsoon up to four–five pentads (20–25 days) in advance has been developed. The method is based on the event-to-event similarity in the properties of monsoon intraseasonal oscillations (ISOs). This two-tier analog method is applied to NOAA outgoing longwave radiation (OLR) pentad averaged data that have sufficiently long records of observation and are available in nearly real time. High-frequency modes in the data are eliminated by reconstructing the data using the first 10 empirical orthogonal functions (EOFs), which together explain about 75% of the total variance. In the first level of the method, the spatial analogs of initial condition pattern are identified from the modeling data. The principal components (PCs) of these spatial analogs, whose evolution history of the latest five pentads matches that of the initial condition pattern, are considered the temporal PC analogs. Predictions are generated for each PC as the average evolution of PC analogs for the given lead time. Predicted OLR values are constructed using the EOFs and predicted PCs. OLR data for 1979–99 are used as the modeling data and independent hindcasts are generated for the period 2000–05. The skill of anomaly predictions is rather high over the central and northern Indian region for lead times of four–five pentads. The phases and amplitude of intraseasonal convective spells are predicted well, especially the long midseason break of 2002 that resulted in large-scale drought conditions. Skillful predictions can be made up to five pentads when started from an active initial state, whereas the limit of useful predictions is about two–three pentads when started from break initial conditions. An important feature of this method is that unlike some other empirical methods to forecast monsoon ISOs, it uses minimal time filtering to avoid any possible endpoint effects and hence may be readily used for real-time applications. Moreover, as the modeling data grow with time as a result of the increased number of observations, the number of analogs would also increase and eventually the quality of forecasts would improve.

### 1. Introduction

The present-day capabilities of predicting the seasonal mean Indian summer monsoon are limited by a number of factors. The model biases in simulating the mean summer monsoon conditions (Gadgil and Sajani 1998; Kang et al. 2002; Sperber and Palmer 1996; Wang et al. 2004), improper simulation of the boundary-forced interannual variability (Anderson et al. 1999; Fennessy and Shukla 1999; Kumar and Hoerling 1995; Lau 1985; Shukla 1998; Shukla and Wallace 1983), exclusion of the coupled ocean–atmosphere processes

(Wang et al. 2005, 2003), and internal interannual variability generated within the monsoon system (Cherchi and Navarra 2003; Krishnamurthy and Shukla 2000; Sperber and Palmer 1996; Sperber et al. 2000) are at the heart of the problem. The predictability of the interannual variations of the seasonal mean Indian summer monsoon is largely limited by the significant internal variability that arises primarily because of the monsoon intraseasonal oscillations (Goswami and Xavier 2005). Moreover, the utility of forecasts of all (averaged) India summer monsoon rainfall to the user community (i.e., the agriculture and hydrological sector) remains uncertain. While the anomaly of seasonal mean rainfall has the same sign over most of the continent in extreme floods of drought years, it is rather inhomogeneous in “normal” monsoon years. Therefore, it would be diffi-

---

*Corresponding author address:* Prof. B. N. Goswami, Indian Institute of Tropical Meteorology, Pune 411 008, India.  
E-mail: goswami@tropmet.res.in

cult even with a skillful forecast of the all India seasonal rainfall to provide regional communities with information that could be used effectively in agriculture and water resources management (Webster and Hoyos 2004). Alternatively, skillful and timely forecasts of monsoon intraseasonal variability 3–4 weeks in advance may be more useful for regional agriculture and hydrological planning.

The quasiperiodic properties of monsoon intraseasonal oscillations (ISOs) suggest that useful skill of precipitation forecasts could be achieved for lead times of about 3 weeks. The potential predictability limits of ISOs, estimated using AGCM experiments (Liess et al. 2005; Reichler and Roads 2005; Waliser et al. 2003a,b) are about 3–4 weeks and are in agreement with such estimates from observations (Goswami and Xavier 2003). It is also discovered that transitions from monsoon breaks to active conditions are much more chaotic than those from active to breaks, a fundamental property of monsoon ISOs (Goswami and Xavier 2003; Waliser et al. 2003a,b).

Some earlier studies (Cadet and Daniel 1988; Chen et al. 1992; Krishnamurti and Ardunay 1980; Ramasastri et al. 1986; Singh and Kriplani 1990) that attempted to explore the potential for the extended-range prediction of monsoon ISO have been rather inconclusive. Recently, skillful forecasts of the Madden–Julian oscillation (MJO) (Madden and Julian 1994), up to 3 weeks in advance, were made using empirical techniques (Jones et al. 2004; Lo and Hendon 2000; Mo 2001; Waliser et al. 1999; Wheeler and Weickmann 2001). Goswami and Xavier (2003) adopted a similar methodology as that of Lo and Hendon (2000) and demonstrated skillful predictions of rainfall over India at 15–20-day lead times. Webster and Hoyos (2004) developed a model for predicting intraseasonal rainfall variations over India and Brahmaputra–Ganges River discharge to Bangladesh 20–25 days in advance. An important problem to be tackled for real-time prediction of ISOs is the implementation of filtering techniques without the loss of information at the end point of the time series (Huang et al. 1998; Kijewski and Kareem 2002). Some of the aforementioned studies adopt methodologies to overcome the endpoint effects so as to facilitate real-time applications (Lo and Hendon 2000; Mo 2001; Webster and Hoyos 2004; Wheeler and Weickmann 2001).

The results of Goswami and Xavier (2003) and Waliser et al. (2003a,b) indicate that there is certain event-to-event regularity in the evolution of monsoon ISOs up to a certain time frame referred to as the predictability limit. It is known that the ISOs have large-scale patterns associated with them and are slow in their evo-

lution. Furthermore, they are convectively coupled oscillations that evolve coherently with the underlying sea surface temperature (Goswami and Ajayamohan 2001; Goswami et al. 2003; Sengupta et al. 2001). How can we incorporate these properties into a useful predictive tool? The analog method of forecasting employs the philosophy that weather behaves in such a way that the present initial conditions, if found to be similar to a past situation, will evolve in a similar fashion. Therefore, once two similar patterns are found in the past records, the assumption made is that their future development will also be similar. This means that if a “good” analog could be found for the current atmosphere, a forecast could be obtained by using the sequence of previously observed atmospheric states as a reference.

The use of analogs is not a new concept in meteorological forecasting. In the past, a variety of analog schemes have been formulated, employing various predictors and analog selection criteria. The technique has been employed in many different applications: general circulation forecasting (Gutzler and Shukla 1984; Radinovic 1975; Van den Dool 1989); long-range weather (Livezey and Barnston 1988; Schuurmans 1973; Toth 1989), temperature (Bergen and Harnack 1982), and precipitation (Christensen et al. 1981) forecasting; 1–6-day temperature forecasting (Kruizinga and Murphy 1983); long-range prediction of sea ice anomalies (Chapman and Walsh 1991); short-term visibility forecasts in the United States (Chisholm 1976; Tahnk 1975) and Canada (Esterle 1992); short-term mesoscale transport forecasts (Carter and Keislar 2000); and El Niño–Southern Oscillation index forecasts (Drosowsky 1994).

The observed regularities in the evolutions and the similarities in the large-scale spatial patterns of monsoon ISOs have motivated us to attempt the analog method for extended-range monsoon forecasting. Even though one cannot expect exactly identical analogs, as the weather hardly repeats, it should be possible to find closely matching analogs of the large-scale envelope of monsoon intraseasonal variability. The success of the forecasts would depend on the number of such analogs one can isolate from the data. Hence, the constraints for choosing a variable for forecasting must be a reasonably long history observations as well as availability on real time. One such variable that bears close association with the rainfall is the OLR. The following section describes the analog method, section 3 evaluates the forecasts for the June–September season, section 4 examines the dependency of forecasts on the initial conditions, and section 5 analyzes the regional forecasts over India. Results are summarized in section 6.

## 2. Two-tier analog method

We assume that the predictable component of the subseasonal variations is the large-scale envelop of intraseasonal oscillations that contains high-frequency weather fluctuations embedded on it. It is indeed a difficult task to predict the day-to-day weather variations 15–25 days in advance. To highlight the low-frequency intraseasonal variations and to smooth the high-frequency synoptic weather variations, the National Oceanic and Atmospheric Administration (NOAA) interpolated daily outgoing longwave radiation (OLR) data are converted into 5-day averages (pentad means). The model we propose is to predict the intraseasonal variations of pentad averaged data. The total data length is divided into two segments, namely a 21-yr modeling period (1 January 1979–31 December 1999) and a nearly 6-yr hindcast period (1 January 2000–29 August 2005). Pentad OLR data until the beginning of the hindcast period (say  $t = t_0$ ; here, 1 January 2000) are subjected to EOF decomposition (Bjornsson and Venegas 1997; Venegas 2001) into a number of spatial and temporal modes. EOF decomposition is performed over the domain  $15^{\circ}\text{S}$ – $30^{\circ}\text{N}$ ,  $50^{\circ}$ – $110^{\circ}\text{E}$ , as this area represents the maximum subseasonal variability during the summer monsoon season. The annual mean climatology is subtracted from the OLR data prior to the EOF decomposition. The first three EOFs and their corresponding principal components (PCs) are shown in Fig. 1. EOF1 essentially represents the spatial pattern of the annual cycle of OLR variability (Fig. 2) with a unipolar structure over the monsoon domain with maximum loadings over the Bay of Bengal. EOF2 is similar to the classical pattern of monsoon intraseasonal variations but possesses significant seasonality in its time evolution (Fig. 2). The subsequent EOFs show the different modes of subseasonal variability. The separation between the modes in terms of the percentage variance explained are not clear after about 10 modes and cumulatively they contribute about 75% of the total variance (Fig. 3). Higher modes may be considered noise. A second step to filter out the noise from the data is reconstructing the OLR data with the first 10 EOFs and PCs as

$$\text{OLR}_r(x, y, t) = \sum_{n=1}^{10} \text{EOF}_n(x, y) \times \text{PC}_n(t), \quad (1)$$

where  $\text{OLR}_r(x, y, t)$  is the reconstructed OLR, and  $\text{EOF}_n(x, y)$  and  $\text{PC}_n(t)$  are the  $n$ th EOF and PC, respectively. Ten EOFs are chosen as a compromise between maximizing the amount of variance for the reconstructed OLR data and minimizing the noise in the form of higher modes. The seasonal cycle of OLR is

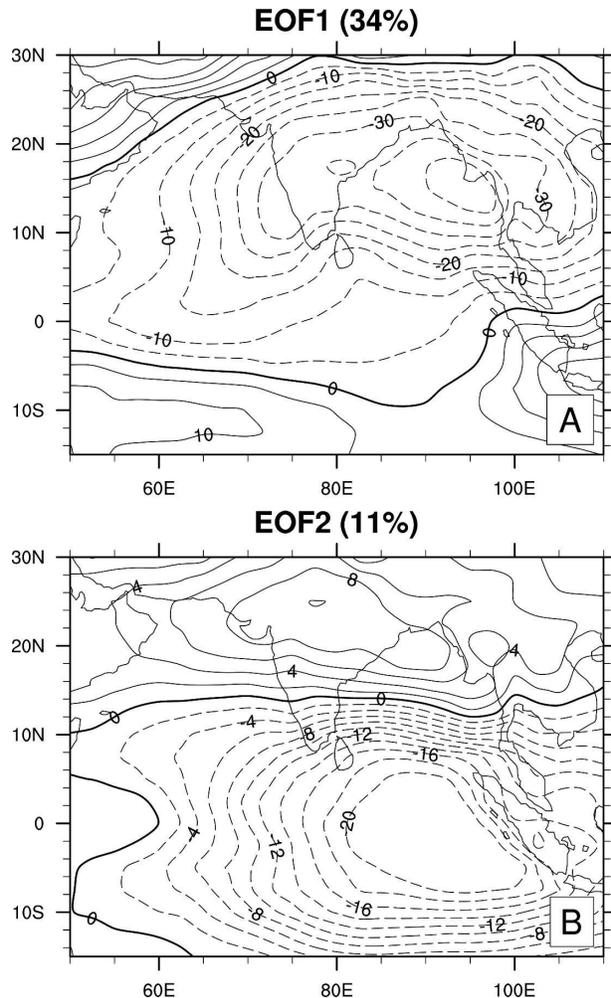


FIG. 1. First two EOFs of pentad average OLR. Percentage of variance explained by each EOF is also noted.

retained in the hindcast experiments. However, the presence of the winter season in the modeling data will not affect predictions of summer values because of the intrinsic property of the analog method that automatically identifies suitable analogs from the corresponding season. This feature is highly advantageous for operational forecasting purposes, as it requires minimum data processing efforts.

The basic algorithm of our method is as follows:

- 1) Consider the spatial pattern of  $t_0$  and find the spatial correlation (in the domain  $15^{\circ}\text{S}$ – $30^{\circ}\text{N}$ ,  $50^{\circ}$ – $110^{\circ}\text{E}$ ) with the spatial patterns at each time step in the modeling period.
- 2) Find the spatial root-mean-square error (RMSE) between the spatial pattern of  $t_0$  and the spatial patterns at each time step in the modeling period.
- 3) Check whether the spatial correlations are above 0.7 and spatial RMSE is less than  $20 \text{ W m}^{-2}$ . These

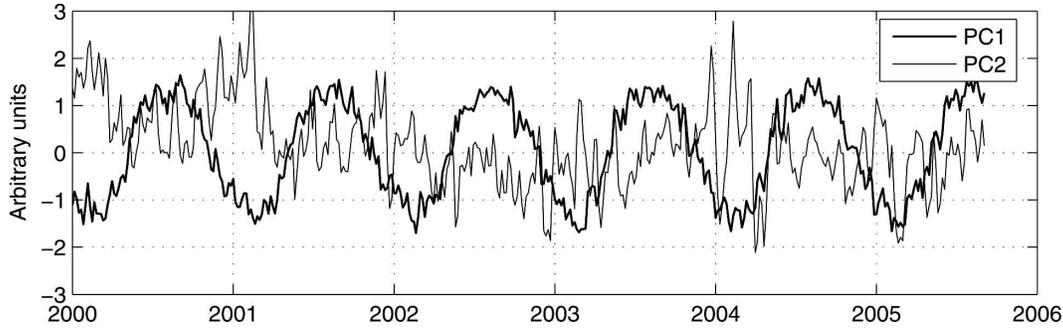


FIG. 2. First two PCs of pentad average OLR.

values are arbitrarily chosen so as to have enough analogs. Those patterns satisfying these criteria are considered the spatial analogs of  $t_0$ . Let  $p_i$ ,  $i = 1, 2, \dots, N$ , where  $N$  is the number of spatial analogs found. Typical values of  $N$  are around 55.

- 4) Consider the evolution of PC1 from  $t_0 - 5$  to  $t_0$  and find the temporal correlation and RMSE between the PC1 from  $p_i - 5$  to  $p_i$ ,  $i = 1, 2, \dots, N$ . If the correlations are greater than 0.5 (arbitrary, yet gives enough analogs) and RMSE is less than the unit standard deviation of PC1, then those are the temporal analogs of PC1 from  $t_0 - 5$  to  $t_0$ . Let them be  $q_j$ ,  $j = 1, 2, \dots, M$ , where  $M$  is the number of temporal analogs (typically on the order of 20) of PC1 and  $M \leq N$ .
- 5) Forecasts of PC1 at lead-time  $\tau$  pentads are generated as

$$\text{PC1}(t_0 + \tau) = \frac{1}{M} \sum_{j=1}^M \text{PC1}(q_j + \tau). \quad (2)$$

- 6) Repeat steps 4–5 for PC2, PC3,  $\dots$ , PC10. Then we have the predicted values of each PC as  $\text{PC}_k(t_0 + \tau)$ ,  $k = 1, 2, \dots, K$ , where  $K$  is the number of EOFs used; here,  $K = 10$ .
- 7) Predicted OLR values for lead-time  $\tau$  are generated as

$$\text{OLR}(x, y, t_0 + \tau) = \sum_{k=1}^K \text{EOF}_k(x, y) \times \text{PC}_k(t_0 + \tau). \quad (3)$$

No forecast is possible if  $N = 0$  or  $M = 0$ . Such time steps are considered unpredictable by this method. However, with the correlations and RMSE criteria used here, no such unpredictable time steps are found during the hindcast period. A comparison of four-pentad-lead predictions and the corresponding observations over central India is shown in Fig. 4. The high degree of accuracy in predicting the strong seasonality of the OLR fluctuations over continental India in the time

scales of the predictions is evident. Since our interest is in predicting the intraseasonal variations embedded on the annual cycle and in order to eliminate any artifacts due to the apparent skill in predicting the annual cycle, the intraseasonal anomalies are extracted from the total OLR predictions and the corresponding observations by removing the observed climatological annual cycle. The predictions are scaled by a factor determined by the ratio of variance explained by the 10 EOFs (EOFs 1–10) to the total OLR variance. Hereafter, all the results presented are based on the intraseasonal OLR anomalies computed as described above.

### 3. Hindcast validation

The performance of the model in predicting the intraseasonal variability in the June–September (JJAS)

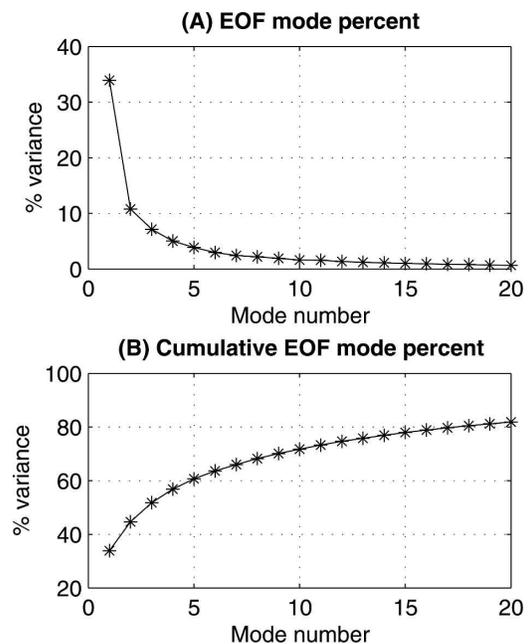


FIG. 3. Percentage of variance explained by (a) the first 20 modes and (b) their cumulative variance.

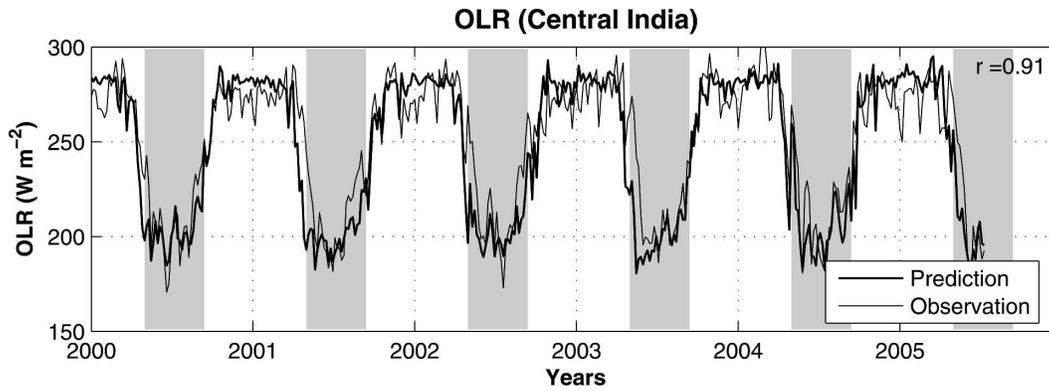


FIG. 4. Four-pentad-lead predictions and observations over central India during the hindcast period. Gray shades indicate the summer monsoon season. Correlation coefficient between the two is also shown.

season is evaluated here. The temporal correlation coefficients (at every grid point) between predictions and observations of intraseasonal OLR anomalies at lead times of two–five pentads are shown in Fig. 5. It may be noted that the correlations over the continental Indian region, especially central India, are high and significant. Correlations remain above 0.6 over a large region of central India even at a five-pentad lead. However, pre-

dictions for the southern Indian states and the oceanic regions do not possess significant skills at four- and five-pentad leads. This is further supported by Fig. 6, where the spatial and average temporal correlations and RMSE for a large region north of 10°N are shown. The correlations remain highly skillful even up to lead times of five pentads. This is a major advantage of our prediction scheme compared to some other prediction

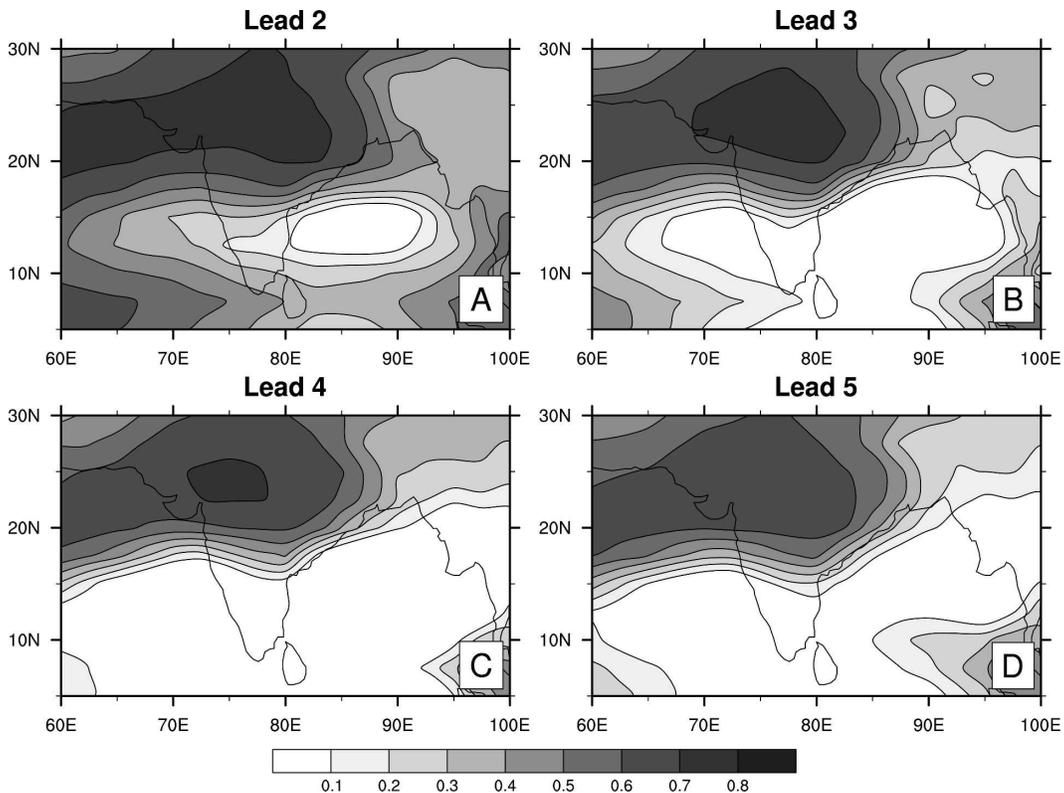


FIG. 5. Temporal correlations between predictions and observations of OLR anomalies for different lead times during the JJAS season of the hindcast period (2000–05).

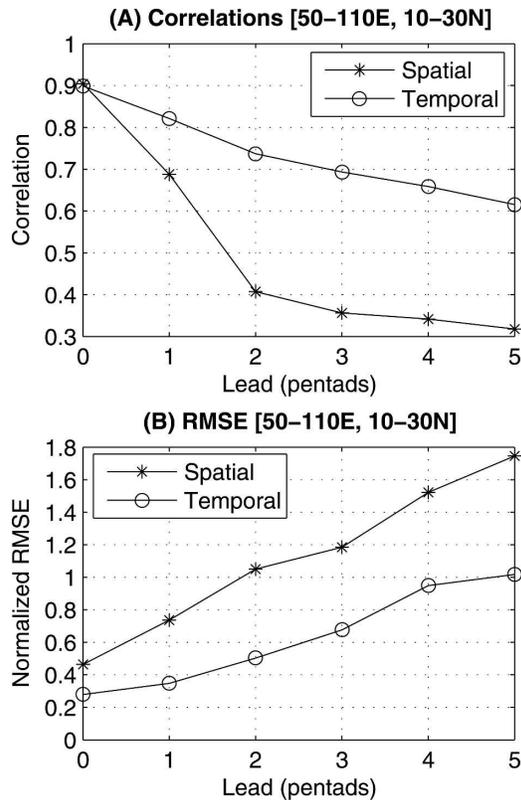


FIG. 6. (a) Spatial and temporal correlations between predictions and observations for JJAS period at different lead times over the region  $10^{\circ}$ – $30^{\circ}$ N,  $50^{\circ}$ – $110^{\circ}$ E. (b) Spatial and temporal RMSE for the same region.

methods in use, since the intraseasonal anomalies are predicted with a high degree of accuracy without employing filtering techniques to isolate the low-frequency temporal evolutions.

The temporal evolution of OLR anomalies predicted four pentads ahead along with the observations aver-

aged over central India ( $20^{\circ}$ – $25^{\circ}$ N,  $75^{\circ}$ – $95^{\circ}$ E) during the JJAS seasons of the six hindcast years is shown in Fig. 7. Overall, the four-pentad-lead predictions foresee the phases of intraseasonal OLR variations quite well in all six hindcast years. Most peaks and troughs align fairly accurately with observations. However, the amplitude of predictions is sometimes overestimated. The skill of predicting the extrema is superior to the predictions of the transition phases of the ISOs. The correlation between the two is 0.66 (significant at the 99% level) and the RMSE is  $7.6 \text{ W m}^{-2}$ . The long midseason breaks of 2002 and 2004, which caused severe droughts over India, were rather accurately forecasted four pentads ahead.

#### 4. Dependency of forecasts on the state of the initial condition

The finding that the predictability of intraseasonal variability in the observations depends on the state of the initial condition from where the forecast is made (Goswami and Xavier 2003) is consolidated by the estimates of Fu et al. (2007) and Waliser et al. (2003a,b) using models. From both observations and model simulations, it is found that the potential predictability limit is rather extended when started from an active initial state in comparison with the limited period of potential for prediction when started from a break initial condition. It is therefore worthwhile examining whether these intrinsic limits of potential predictability affect the skills of predictions. To identify active and break conditions, an index of monsoon intraseasonal variability is defined as the time series of OLR anomalies averaged over the central Indian region and normalized with its own standard deviation. Active (break) phases are identified as the days when the normalized values of OLR anomalies are less than  $-1.5$  (greater than 1.5).

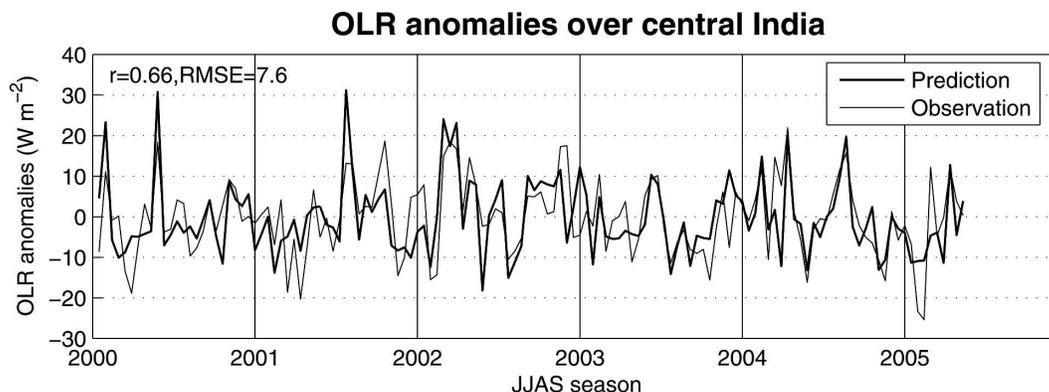


FIG. 7. Prediction of OLR anomalies ( $\text{W m}^{-2}$ ) at four-pentad leads in comparison with observed values over the region  $20^{\circ}$ – $25^{\circ}$ N,  $75^{\circ}$ – $95^{\circ}$ E for the JJAS season of the hindcast period.

Predictions are made from each of these active and break initial conditions until five-pentad leads and their corresponding observed values are noted. Nineteen active and 16 break events are found in the hindcast period. The temporal correlations of predictions from active and break initial conditions with their observed counterparts averaged over the central Indian region are plotted as a function of forecast lead time (Fig. 8a). Correlation values above 0.48 are significant at the 95% level. An important feature to be noted is that the correlations from active initial conditions are better than those from break initial conditions, except for lead times of two pentads, where there is a slight increase in the correlation of predictions from break initial conditions. The quality of forecasts from break initial conditions deteriorates rapidly at three-, four-, and five-pentad leads. However, the skill of forecasts from active initial states remains steady and highly significant even at a five-pentad lead. The average spatial correlations of predictions and the observations over a large region north of 10°N starting from active and break initial conditions shown in Fig. 8b complement this. Spatial correlation is computed for 80 grid boxes and values above 0.23 are significant at the 95% level.

The reduction in spatial correlations with lead time is rather steady for predictions from break initial conditions compared to those from active initial conditions. Spatial correlations from break initial conditions are found to be superior over the active initial condition counterparts at lead times of 2 and 3 and then weaken rapidly. Whereas the correlations from active initial conditions have a dip at two- and three-pentad leads and then peak up at four- and five-pentad leads. In short, the skill of predictions starting from a break initial state have the best skills at two–three-pentad leads. On the other hand, forecasts generated from active initial conditions have high skills at four–five-pentad leads, and most importantly, their skills at four–five-pentad leads are far more superior than those from break initial conditions at two–three-pentad leads. An active (break) phase normally evolves into a break (an active) phase after a period of 15–20 days. Therefore, we may conclude that the breaks are predictable up to five-pentad lead times and the skills of predicting active conditions are restricted to about two–three-pentad leads. Thus, the evidence presented from the model hindcasts establishes that the variable limits of predictability of active and break phases of Indian summer monsoon are intrinsic properties of the system.

Why are the predictability limits higher for the break phase than for the active phase? The different behaviors of error growth in the forecasts from active and

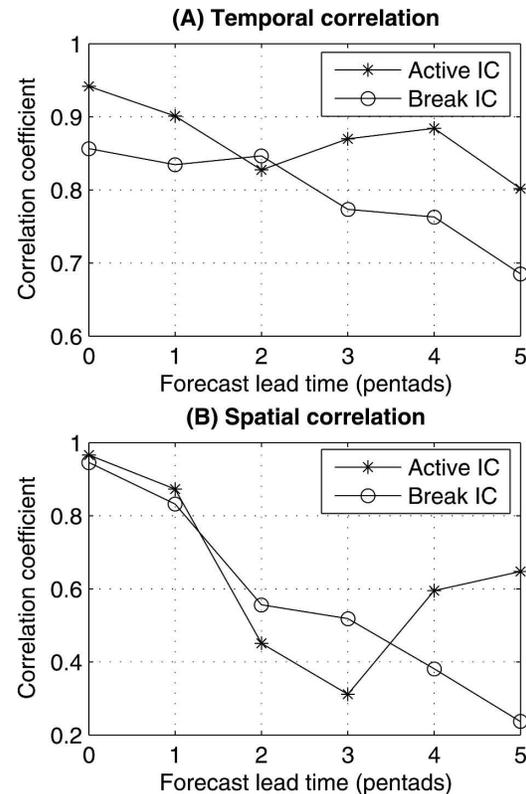


FIG. 8. Temporal and spatial correlations between predictions and observations from active and break initial conditions at different lead times.

break initial conditions may be related to the different physical processes controlling these transitions (Goswami and Xavier 2003; Waliser et al. 2003a,b). The transition from break to active is governed by fast-growing convective instability. On the other hand, the transition from active to break is governed by the evolution of the large-scale Hadley circulation (Goswami and Xavier 2003). If a small perturbation is introduced in the break phase, it grows slowly at first but rapidly later when convective instability becomes active. Whereas if a small perturbation is introduced in the active phase, it grows fast initially but eventually becomes steady (Fu et al. 2007). Different error growth regimes during these two transition periods result in the break phase being more predictable than the active phase.

## 5. Regional forecasts

The need for forecasting intraseasonal monsoon variability on the subseasonal scale and the benefits it can deliver to the agricultural sector of the country has

been discussed in the introduction. The monsoon rainfall has strong regionality in its character and the country has been divided into a number of rainfall zones. The large inhomogeneities in the subseasonal rainfall are one of the factors that questions the utility of seasonal mean forecasts over the country and motivated us to develop a method for regional-scale forecasts four-five pentads in advance.

As shown in Fig. 9a, we may divide the India subcontinent into six regions. Region 1 corresponds to parts of Rajasthan and Gujarat, and region 2 covers large parts of Uttar Pradesh, Madhya Pradesh, Delhi, and the neighboring northern states. Bihar, Bengal, Bangladesh, Assam, and other northeastern states fall into region 3. Region 4 covers Maharashtra and parts of Madhya Pradesh and Gujarat. Region 5 denotes mainly Orissa, and parts of Andhra Pradesh and Karnataka are included in region 6. The anomaly correlations of the four-pentad-lead predictions with observations averaged for each region during the JJAS season of the six hindcast years are given in Fig. 9b. The predictions for northern regions, especially 1 and 2, match closely with the observations as evident from the strong correlations. Regions 3, 4, and 5 also have significant correlations and show the utility of four-pentad forecasts over these regions. However, predictions over the southern region (region 6) do not show any useful prediction skill.

An example of regional four-pentad forecasts for the six JJAS seasons averaged over each of the five regions (1–5) in comparison with observations is shown in Fig. 10. Region 6 is not shown since there is hardly any skill in predicting the OLR anomalies in this region. The summer intraseasonal OLR variations are predicted by the model with remarkable accuracy except over region 3 during most of the years. It may be interesting to note that the predictions for 2002, 2003, and 2004 are clearly superior, of which 2002 and 2004 had produced large-scale droughts over India. The long break conditions in 2002 and 2004 are captured in the four-pentad forecasts. Accurate predictions for such extreme monsoon years would give us confidence in applying the method for operational forecasting so as to foresee the large-scale active/break spells that can offset the seasonal mean to extremes. Another interesting observation is that there is a dependence of skill of the forecasts on the amplitude of ISOs in different regions. Regions 1 and 2 have the largest ISO amplitude and the model predicts the amplitudes and phases of intraseasonal variability fairly accurately. Predictions for regions 4 and 5 are also quite close to the observations, but with lower skill. On the other hand, regions 3 and 6 do not

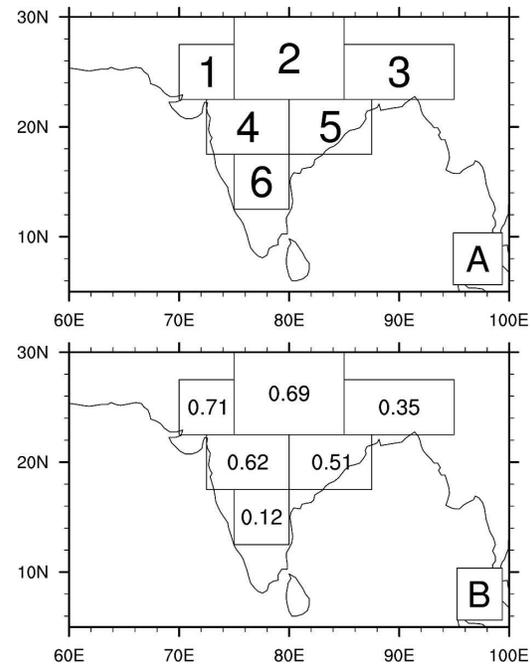


FIG. 9. (a) Regions chosen to evaluate the predictions with observations. (b) The correlation coefficient between four-pentad lead predictions and observations averaged over these regions for the JJAS period.

show any well-defined intraseasonal behavior and the model has marginal skill in predicting it.

## 6. Summary

Based on the premise that the monsoon intraseasonal oscillations exhibit regularity in their evolutions and similarity in the spatial patterns, a new physically based analog method for forecasting the intraseasonal variability four-five pentads in advance has been developed. Motivated by the previous uses of analog methods for weather forecasting, a two-tier analog model is used here that selects analogs of the spatial patterns and their temporal evolutions from a sufficiently long data record. For the best results with the analog method to be applied in real time, the data under consideration should have sufficient past records of observation to be able to find as many analogs as possible, should be as complete as possible for real-time availability, and most importantly, should bear a close relationship with rainfall. We choose NOAA-interpolated OLR, as it meets these criteria. To eliminate the possible contamination of the forecasts due to the high-frequency synoptic disturbances, the daily data are converted to pentad averages. Our method is targeted at operational real-time forecasting purposes and, hence, is devoid of any kind

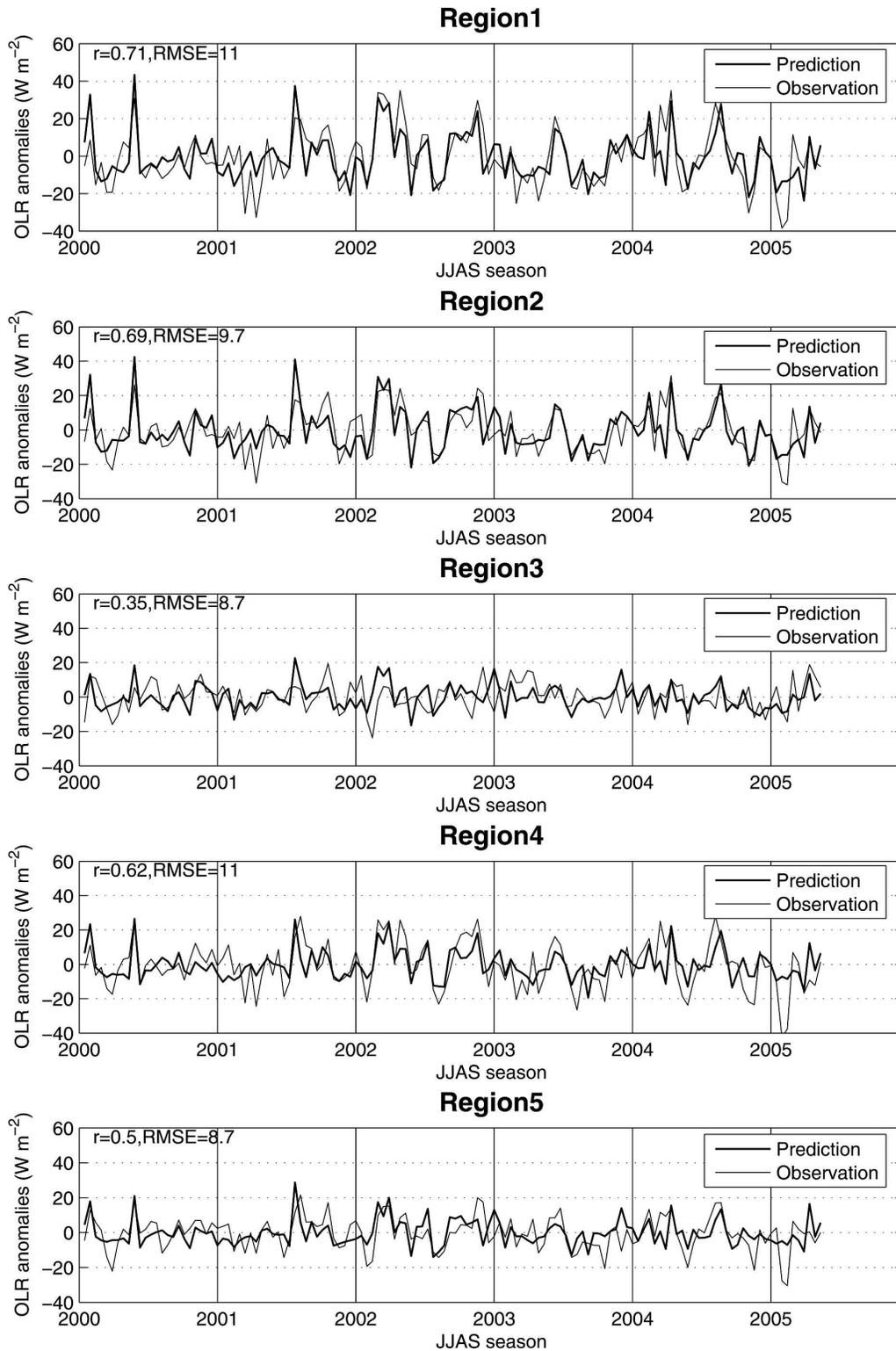


FIG. 10. Predictions at four-pentad lead in comparison with observed values of OLR anomalies ( $W m^{-2}$ ) over regions 1–5 as indicated in Fig. 9 for the JJAS season of the hindcast period. Correlations and RMSE ( $W m^{-2}$ ) between the predictions and observations are also given in the respective panels.

of temporal filtering that introduces endpoint errors. High-frequency modes in the data are eliminated by reconstructing the OLR data with the first 10 EOFs and the corresponding PCs that explain about 75% of the total OLR variance.

Spatial analogs of the initial condition pattern are chosen from the modeling data. Temporal analogs of each PC from five pentads behind the initial condition are found from the set of PCs of spatial analogs. Predictions of each PC for a particular lead time are generated separately as the average values of evolutions of the PC analogs for that lead time. Predicted OLR values are constructed by multiplying each of the predicted PCs with the corresponding EOF. The forecasted OLR anomalies are created by removing the observed climatological annual cycle from the total OLR forecasts. These forecasts show substantial skill in forecasting the intraseasonal OLR anomalies four–five pentads in advance. The phases and amplitude of the intraseasonal variability are predicted with high skill.

There are differences in the skills of forecasts depending on the initial condition from where the forecasts are made. Supporting the findings of Goswami and Xavier (2003) and Waliser et al. (2003a,b), we find that the forecasts starting from an active monsoon initial condition remain skillful even up to five pentads, while those starting from a breaklike initial state show useful skills only up to two–three pentads. Regional forecasts made for six regions of the country indicate high skills for the central and western Indian regions. A case in point is the fairly accurate predictions of the long midseason monsoon break of 2002 that was associated with the unprecedented drought over the country.

Major advantages of this scheme over the existing prediction methods are that it avoids the high-frequency variability in a clever way so as to avoid time filtering of any kind to extract the intraseasonal signal, and due to the intrinsic property of the analogs, it automatically finds out closely matching patterns from the corresponding season. This feature considerably minimizes data processing requirements. The skills demonstrated with this fairly simple method are high and have immense potential for practical purposes. The success of our method is that, due to the long modeling period, we were able to find a large number of closely matching analogs. As the time progresses, the volume of data is growing and so is the number of analogs. Hence, it is expected that the quality of forecasts would improve with time. However, the method has rather limited success in predicting the variability over the southern and northeastern parts of India and the Bay of Bengal where the seasonal mean rainfall is maximum. In-depth

diagnoses on the properties of the analogs over these regions might shed light on the causes of limited predictability over these regions. Even though statistical models provide better forecasts than the dynamical forecasts (Lo and Hendon 2000), certain inabilities are common to all the empirical based methods (e.g., inaccurate representation of the physical processes and inability to accommodate variations in the predictor–predictor relationships). It is expected that the numerical models will improve with time with the wealth of understanding of the phenomena and would be able to simulate and produce skillful extended-range forecasts of the monsoon intraseasonal variability. Until then, a judicious and practical way is to employ such physically based empirical methods for predictions, as the necessity for 2–3-week forecasts over the country is overwhelming.

*Acknowledgments.* This work is partially supported by the Department of Science and Technology, Government of India, New Delhi. Dr. Duane Waliser and two anonymous reviewers are acknowledged for their valuable comments on an earlier version of the paper. NOAA-interpolated OLR data are obtained online ([http://www.cdc.noaa.gov/cdc/data.interp\\_OLR.html](http://www.cdc.noaa.gov/cdc/data.interp_OLR.html)).

#### REFERENCES

- Anderson, J., H. van den Dool, A. Barnston, W. Chen, W. Stern, and J. Ploshay, 1999: Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and prediction. *Bull. Amer. Meteor. Soc.*, **80**, 1349–1361.
- Bergen, R. E., and R. P. Harnack, 1982: Long-range temperature prediction using a simple analog approach. *Mon. Wea. Rev.*, **110**, 1083–1099.
- Bjornsson, H., and S. Venegas, 1997: A manual for EOF and SVD analyses of climate data. McGill University CCGCR Report 97-1, 52 pp.
- Cadet, D. L., and P. Daniel, 1988: Long-range forecast of the break and active summer monsoons. *Tellus*, **40A**, 133–150.
- Carter, R. G., and R. E. Keislar, 2000: Emergency response transport forecasting using historical wind field pattern matching. *J. Appl. Meteor.*, **39**, 446–462.
- Chapman, W. L., and J. E. Walsh, 1991: Long-range prediction of regional sea ice anomalies in the Arctic. *Wea. Forecasting*, **6**, 271–288.
- Chen, T.-C., M.-C. Yen, K.-D. Wu, and T. Ng, 1992: Long-range prediction of the low-frequency mode in the low-level Indian monsoon circulation with a simple statistical method. *Tellus*, **44A**, 324–330.
- Cherchi, A., and A. Navarra, 2003: Reproducibility and predictability of Asian summer monsoon in the ECHAM4-GCM. *Climate Dyn.*, **20**, 365–379.
- Chisholm, D. A., 1976: Objective prediction of mesoscale variations on sensor equivalent visibility during advective situations. *Environ. Res. Pap.*, **569**, 31–42.
- Christensen, R. A., R. F. Eilbert, O. H. Lindgren, and L. L. Rans,

- 1981: Successful hydrologic forecasting for California using an information theoretic model. *J. Appl. Meteor.*, **20**, 706–713.
- Drosowsky, W., 1994: Analog (nonlinear) forecasts of the Southern Oscillation index time series. *Wea. Forecasting*, **9**, 78–84.
- Esterle, G. R., 1992: Adaptive, self-learning statistical interpretation system for the central Asian region. *Ann. Geophys.*, **10**, 924–929.
- Fennessy, M. J., and J. Shukla, 1999: Impact of initial soil wetness on seasonal atmospheric prediction. *J. Climate*, **12**, 3167–3180.
- Fu, X., B. Wang, D. E. Waliser, and L. Tao, 2007: Impact of atmosphere–ocean coupling on the predictability of monsoon intraseasonal oscillations. *J. Atmos. Sci.*, **64**, 157–174.
- Gadgil, S., and S. Sajani, 1998: Monsoon precipitation in the AMIP runs. *Climate Dyn.*, **14**, 659–689.
- Goswami, B. N., and R. S. Ajayamohan, 2001: Intraseasonal oscillations and interannual variability of the Indian summer monsoon. *J. Climate*, **14**, 1180–1198.
- , and P. K. Xavier, 2003: Potential predictability and extended range prediction of Indian summer monsoon breaks. *Geophys. Res. Lett.*, **30**, 1966, doi:10.1029/2003GL017810.
- , and —, 2005: Dynamics of “internal” interannual variability of the Indian summer monsoon in a GCM. *J. Geophys. Res.*, **110**, D24104, doi:10.1029/2005JD006042.
- , R. S. Ajayamohan, P. K. Xavier, and D. Sengupta, 2003: Clustering of synoptic activity by Indian summer monsoon intraseasonal oscillations. *Geophys. Res. Lett.*, **30**, 1431, doi:10.1029/2002GL016734.
- Gutzler, D. S., and J. Shukla, 1984: Analogs in the wintertime 500 mb height field. *J. Atmos. Sci.*, **41**, 177–189.
- Huang, N. E., and Coauthors, 1998: The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. Roy. Soc. London A*, **454**, 903–995.
- Jones, C., L. M. V. Carvalho, R. W. Higgins, D. E. Waliser, and J.-K. E. Schemm, 2004: A statistical forecast model of tropical intraseasonal convective anomalies. *J. Climate*, **17**, 2078–2095.
- Kang, I.-S., and Coauthors, 2002: Intercomparison of atmospheric GCM simulated anomalies associated with the 1997/98 El Niño. *J. Climate*, **15**, 2791–2805.
- Kijewski, T., and A. Kareem, 2002: On the presence of end effects and their melioration in wavelet-based analysis. *J. Sound Vib.*, **256**, 980–988.
- Krishnamurthy, V., and J. Shukla, 2000: Intraseasonal and interannual variability of rainfall over India. *J. Climate*, **13**, 4366–4377.
- Krishnamurti, T. N., and P. Ardunay, 1980: The 10- to 20-day westward propagating mode and “breaks in the monsoons.” *Tellus*, **32**, 15–26.
- Kruizinga, S., and A. H. Murphy, 1983: Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands. *Mon. Wea. Rev.*, **111**, 2244–2254.
- Kumar, A., and M. P. Hoerling, 1995: Prospects and limitations of atmospheric GCM climate predictions. *Bull. Amer. Meteor. Soc.*, **76**, 335–345.
- Lau, N.-C., 1985: Modeling the seasonal dependence of the atmospheric response to observed El Niños in 1962–76. *Mon. Wea. Rev.*, **113**, 1970–1996.
- Liess, S., D. E. Waliser, and S. Schubert, 2005: Predictability studies of the intraseasonal oscillation with the ECHAM5 GCM. *J. Atmos. Sci.*, **62**, 3320–3336.
- Livezey, R. E., and A. G. Barnston, 1988: Operational multifield analog/antianalog prediction system for United States seasonal temperatures. Part I: System design and winter experiments. *J. Geophys. Res.*, **93**, 10 953–10 974.
- Lo, F., and H. H. Hendon, 2000: Empirical extended-range prediction of the Madden–Julian oscillation. *Mon. Wea. Rev.*, **128**, 2528–2543.
- Madden, R. A., and P. R. Julian, 1994: Observations of the 40–50-day tropical oscillation—A review. *Mon. Wea. Rev.*, **122**, 813–837.
- Mo, C. K., 2001: Adaptive filtering and prediction of intraseasonal oscillations. *Mon. Wea. Rev.*, **129**, 802–817.
- Radinovic, D., 1975: An analogue method for weather forecasting using the 500/1000 mb relative topography. *Mon. Wea. Rev.*, **103**, 639–649.
- Ramasastri, A. A., U. S. De, D. Vaidya, and G. Sundari, 1986: Forty day mode and medium range forecasting. *Mausam*, **37**, 305–312.
- Reichler, T., and J. O. Roads, 2005: Long-range predictability in the Tropics. Part II: 30–60-day variability. *J. Climate*, **18**, 634–650.
- Schuermans, C. J. E., 1973: Four-year experiment in long-range weather forecasting, using circulation analogues. *Meteor. Rundsch.*, **26**, 2–4.
- Sengupta, D., B. N. Goswami, and R. Senan, 2001: Coherent intraseasonal oscillations of ocean and atmosphere during the Asian summer monsoon. *Geophys. Res. Lett.*, **28**, 4127–4130.
- Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science*, **282**, 728–731.
- , and J. M. Wallace, 1983: Numerical simulation of the atmospheric response to equatorial Pacific sea surface temperature anomalies. *J. Atmos. Sci.*, **40**, 1613–1630.
- Singh, S. V., and R. H. Kriplani, 1990: Low frequency intraseasonal oscillations in Indian rainfall and outgoing long wave radiation. *Mausam*, **41**, 217–222.
- Sperber, K. R., and T. N. Palmer, 1996: Interannual tropical rainfall variability in general circulation model simulations associated with the Atmospheric Model Intercomparison Project. *J. Climate*, **9**, 2727–2750.
- , J. M. Slingo, and H. Annamalai, 2000: Predictability and the relationship between subseasonal and interannual variability during the Asian summer monsoons. *Quart. J. Roy. Meteor. Soc.*, **126**, 2545–2574.
- Tahnk, W. R., 1975: Objective prediction of fine-scale variations in radiation fog intensity. *Air Force Sur. Geophys.*, **311**, 1–77.
- Toth, Z., 1989: Long-range weather forecasting using an analog approach. *J. Climate*, **2**, 594–607.
- van den Dool, H. M., 1989: A new look at weather forecasting through analogues. *Mon. Wea. Rev.*, **117**, 2230–2247.
- Venegas, S. A., 2001: Statistical methods for signal detection in climate. Danish Center for Earth System Science, University of Copenhagen, DCESS Report 2, 52 pp.
- Waliser, D. E., C. Jones, C. Schemm, and N. E. Graham, 1999: A statistical extended-range tropical forecast model based on the slow evolution of the Madden–Julian oscillation. *J. Climate*, **12**, 1918–1939.
- , K. M. Lau, W. Stern, and C. Jones, 2003a: Potential predictability of the Madden–Julian oscillation. *Bull. Amer. Meteor. Soc.*, **84**, 33–50.
- , W. Stern, S. Schubert, and K. M. Lau, 2003b: Dynamic predictability of intraseasonal variability associated with the Asian summer monsoon. *Quart. J. Roy. Meteor. Soc.*, **129**, 2897–2925.

- Wang, B., R. Wu, and T. Li, 2003: Atmosphere–warm ocean interaction and its impacts on Asian–Australian monsoon variation. *J. Climate*, **16**, 1195–1211.
- , I.-S. Kang, and J.-Y. Lee, 2004: Ensemble simulations of Asian–Australian monsoon variability by 11 AGCMs. *J. Climate*, **17**, 803–818.
- , Q. Ding, X. Fu, I.-S. Kang, K. Jin, J. Shukla, and F. Doblas-Reyes, 2005: Fundamental challenge in simulation and prediction of summer monsoon rainfall. *Geophys. Res. Lett.*, **32**, L15711, doi:10.1029/2005GL022734.
- Webster, P. J., and C. Hoyos, 2004: Prediction of monsoon rainfall and river discharge on 15–30-day time scales. *Bull. Amer. Meteor. Soc.*, **85**, 1745–1767.
- Wheeler, M., and K. M. Weickmann, 2001: Real-time monitoring and prediction of modes of coherent synoptic to intraseasonal tropical variability. *Mon. Wea. Rev.*, **129**, 2677–2694.