GEOPHYSICAL RESEARCH LETTERS, VOL. 33, L18801, doi:10.1029/2006GL027035, 2006

Click Here for Full Article

An empirical rule for extended range prediction of duration of Indian summer monsoon breaks

Suneet Dwivedi,[1] Ashok Kumar Mittal,[1] and B. N. Goswami[2]

Received 30 May 2006; revised 2 August 2006; accepted 9 August 2006; published 16 September 2006.

[1] Prediction of the duration of the Indian summer monsoon breaks is highly desirable. It will help in planning water resource management, sowing and harvesting. Applicability of the recently discovered regime transition rules for the Lorenz model in predicting the duration of monsoon breaks, is explored in this paper. Using several indices of the observed summer monsoon intraseasonal oscillation (ISO), it is shown that the peak anomaly in an active regime can be used as a predictor for the duration of the subsequent break spell. It is also found that the average growth rate around the threshold to an active condition can be used as a predictor of the peak anomaly in the active spell. Average growth around the threshold to an active condition can give useful prediction of the duration of the following break, on an average, about 23 days (38 days) in advance of its commencement (end).

Citation: Dwivedi, S., A. K. Mittal, and B. N. Goswami (2006), An empirical rule for extended range prediction of duration of Indian summer monsoon breaks, *Geophys. Res. Lett.*, *33*, L18801, doi:10.1029/2006GL027035.

1. Introduction

[2] Aperiodic oscillations between active spells of abundant rain and break spells of scanty rain characterize the Indian summer monsoon season [*Rao*, 1976; *Ramamurthy*, 1969]. These oscillations have been the subject of extensive studies [*Webster et al.*, 1998; *Goswami and Ajayamohan*, 2001; *Waliser et al.*, 2003; *Goswami*, 2005]. Frequent or prolonged breaks lead to drought conditions that substantially reduce the agricultural yield [*Gadgil and Rao*, 2000]. Therefore, skillful and timely forecasts of the duration of break spells could be of great value for agriculture planning, disaster and water resource management. Although researches in recent years have exploited the large-scale quasi-periodic character of the monsoon ISOs to develop empirical models for extended range prediction of the Indian summer monsoon ISOs [*Goswami and Xavier*, 2003; *Webster and Hoyos*, 2004], no method is so far available for predicting the duration of breaks. The objective of the present study is to develop an empirical technique for extended range prediction of the duration of monsoon breaks. This effort has been motivated by the recent findings of rules of regime transitions and the duration of regimes in some idealized two regime dynamical systems such as the Lorenz model [*Evans et al.*, 2004; *Yadav et al.*, 2005].

[3] The Lorenz model [*Lorenz*, 1963] exhibits some of the important features of the weather and climate systems such as sensitive dependence on initial conditions, multiple time scales and distinct quasi-stationary regimes. *Evans et al.* [2004] presented two forecasting rules for regime changes in the Lorenz model using bred vectors [*Kalnay*, 2003]. *Yadav et al.* [2005] discovered empirically the following simpler, and more accurate, forecasting rules. Firstly, when the variable $|x(t)|$, of the Lorenz model, crosses a critical value, the current regime will end after it completes the current orbit. Secondly, the duration of the next regime increases monotonically with the maximum value that the variable attains in the previous regime. They also found similar forecasting rules for some other two-regime attractors such as the forced Lorenz model [*Palmer*, 1993], the Rucklidge attractor and the ACT attractor [*Sprott*, 2003]. Thus, these prediction rules seem to represent certain universal properties often exhibited by diverse nonlinear dynamical systems. Although the monsoon ISO is not as clean a two-regime system as the Lorenz model, approximately 85% (80%) of break (active) conditions are immediately followed by active (break) conditions. Hence, the monsoon ISO could be considered as an approximate two-regime system. This suggests the possible applicability of the prediction rules, found in simple two regime systems, to transitions of active-break regimes and for predicting the duration of monsoon breaks. In this study, we examine long time series of several indices of the monsoon intraseasonal variability derived from a number of observed datasets. We show that similar prediction rules do indeed apply to monsoon ISO time series and suggest a method for predicting the duration of monsoon breaks.

[4] To gain some insight into the nonlinear dynamical character of these ISOs, a stochastically forced Lorenz model is proposed as a paradigm model for the monsoon ISOs. Data obtained from this model were subjected to exactly the same treatment as each of the observed indices of monsoon ISOs. We find that different correlations and forecast skill measures for this model are very similar to those for the observed datasets. This shows that a stochastically forced Lorenz model is able to capture some of the salient statistical properties of the Indian monsoon intraseasonal oscillations.

2. Data and Methods

[5] The datasets, which we used for our analysis, and some of their characteristics, are summarized in Table 1. The pentad Climate Prediction Center Merged Analysis of Precipitation (CMAP) [*Xie and Arkin*, 1996] data for 22

---

[1] M. N. Saha Centre of Space Studies and K. Banerjee Centre of Atmospheric and Ocean Studies, Institute of Interdisciplinary Studies, and Department of Physics, University of Allahabad, Allahabad, India.

[2] Indian Institute of Tropical Meteorology, Pune, India.

Copyright 2006 by the American Geophysical Union.
0094-8276/06/2006GL027035$05.00

Table 1. Monsoon ISO Datasets Analyzed

S. No.	Time Series	Area (Used for Box Averaging)	Period	Standard Deviation of Unfiltered Data	Standard Deviation of Filtered Data
1	CMAP daily anomaly from May–October	70–90E, 15–25N	1979–2000	2.35 mm/day	2.25 mm/day
2	OLR daily anomaly from May–October	70–95E, 5–25N	1979–2000	17.62 W/m^2	13.78 W/m^2
3	U850 wind daily anomaly from May–October	80–100E, 10–15N	1979–2001	3.48 m/sec	3.03 m/sec
4	IMD gridded station rainfall daily anomaly from May–October	70–90E, 15–25N	1954–2003	3.99 mm/day	2.84 mm/day

years (1979–2000) is linearly interpolated to daily values. It is found to reasonably well represent the observed ISO in rainfall over India [*Goswami*, 2005]. Other datasets analyzed were: daily interpolated long wave radiation (OLR) data for 22 years (1979–2000) [*Liebmann and Catherine*, 1996], Zonal (U) component of daily NCEP/NCAR reanalyzed wind data at 850hPa level for 23 years (1979–2001) [*Kalnay et al.*, 1996] and daily gridded raingauge data [*Rajeevan et al.*, 2006] analyzed into regular grid boxes over the Indian continent for 50 years (1954–2003).

[6] In order to look at the low frequency intra-seasonal component, a 10–90 day bandpass Lanczos filter is applied to the daily anomalies defined as departures from the annual cycle (sum of annual mean and first three harmonics). An index of ISO is defined by the filtered anomaly averaged over the region described in Table 1. Rainfall based indices (CMAP and gridded daily raingauge) are averaged over the monsoon trough area (70–90E, 15–25N). OLR is similar to the rainfall, but is a smoother field. Hence the OLR index is averaged over a slightly larger area (70–95E, 5–25N). The box (80–100E, 10–15N) for the zonal wind index is taken, keeping in mind that the heating due to precipitation produces a strong wind response, slightly to the south of the heat source. Data, from 1 May–31 October (184 days), from each of the data sets, for the available periods, is taken and normalized by its own standard deviation. A portion (five-year period) of the normalized anomaly index time-series so obtained for the CMAP dataset is shown in Figure 1. Following previous studies [*Goswami and Xavier*, 2003; *Goswami and Ajayamohan*, 2001], we define the active (break) conditions by normalized anomaly index values greater (less) than 1. The peak anomaly in an active (break) spell is denoted by R_{ma} (R_{mb}), duration of the subsequent break (active) spell by T_b (T_a) and the average growth rate around the threshold +1 (−1) of an active (break) spell by γ_a (γ_b). These variables are illustrated in Figure 2, which is a magnification of the inset portion of Figure 1.

[7] In about 15% (20%) cases, a break (active) spell is not immediately followed by an active (break) spell, but by another break (active) spell. In such cases we have added the duration of these break (active) spells to obtain T_b (T_a) and treated them as a single break (active) spell. We have denoted the largest peak anomaly amongst such active (break) spells by R_{ma} (R_{mb}) and the corresponding growth rate around the threshold by γ_a (γ_b). Such a case is shown in Figure 3 for a portion of the CMAP time series.

3. Correlation Analysis and Forecasts

[8] The regime transition prediction rules, for the Lorenz model given by *Yadav et al.* [2005] and *Evans et al.* [2004], suggest the use of R_{ma} and γ_a as the predictors of T_b. We explore this possibility first with the CMAP dataset. Using approximately half the length of the data set (1979–1989), we calculate the correlation between prospective predictors and T_b and construct a simple forecast model (linear regression equation) for predicting it. These results are presented in Tables 2 and 3. The fidelity of this forecast model is verified on the other half of the data set (1990–2000). The results are summarized in Table 3. Other datasets were analyzed and forecasts made in exactly the same way. For the gridded daily rainfall over India, first 30 years are used for model development and the last 20 years are used for verification. These results are also summarized in Tables 2 and 3.

3.1. R_{ma} as a predictor of T_b

[9] We find that R_{ma} is strongly correlated with T_b (r = 0.75). The best-fit regression equation between them is,

$$T_b = 4.12R_{ma} - 0.03 \quad (1)$$

[10] Using this equation, for each of the peaks R_{ma}, the forecast value $T_{b\_forecast}$ of the subsequent break duration is

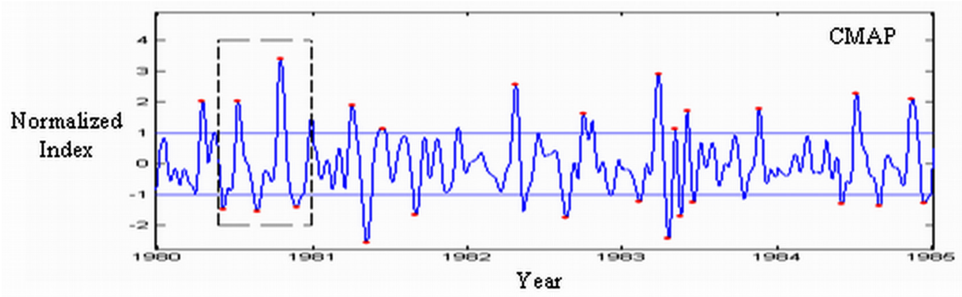

Figure 1. 10–90 day filtered normalized index time series averaged over the monsoon trough region for CMAP daily (May–October) anomaly for five years (1980–1985). Active (break) monsoon condition correspond to the index > +1 (<−1). Red points in the figure indicate maximum anomaly value in active/break regime.

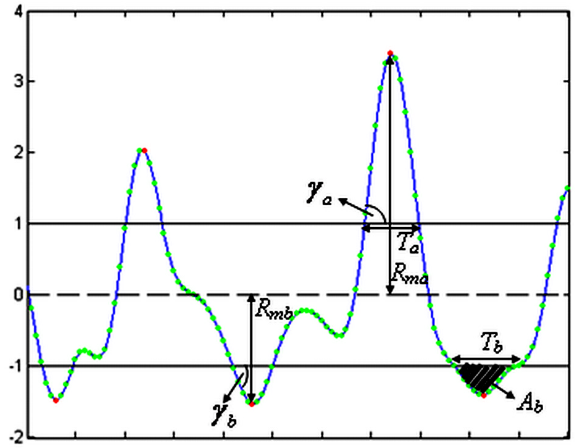

Figure 2. The inset portion of CMAP time series of Figure 1 is magnified to illustrate, γ_a, γ_b, R_{ma}, R_{mb}, T_a and T_b.

computed from the second part of the data set. Figure 4 shows the plot between the actual observed value $T_{b\_observed}$ and $T_{b\_forecast}$. The correlation skill score (ρ_{FP}) between the forecast and the observed T_b is 0.6, which is statistically significant and useful. For the forecast F and the verifying observation P, the root mean square error in the forecast of T_b defined by

$$S_{FP} = \left\{E\left((F-P)^2\right)\right\}^{1/2} \qquad (2)$$

(where E denotes the expectation value) is less than the standard deviation of the observed T_b indicating a good forecast skill.

[11] We have also made a two-class categorical forecast [*von Storch and Zwiers*, 2003] predicting the break spell duration, either as above average or as below average, using R_{ma} as a predictor. Out of the 14(16) observed events above (below) average, 9(14) are correctly predicted by R_{ma}. In Figure 4, the points in the upper right quadrant and the lower left quadrant represent correct categorical forecasts, whereas the points in the other quadrants represent wrong categorical forecasts.

[12] It can also be seen (Table 2) that the correlation between R_{mb} and T_a is weak, compared to that between R_{ma} and T_b. This is in agreement with the observation of *Goswami and Xavier* [2003], that transitions from break to active spells are more chaotic and less predictable than those from active to break spells. It is also found, that the aggregate deficiency A_b of the rainfall from the normal, during a break spell, correlates well with R_{ma}.

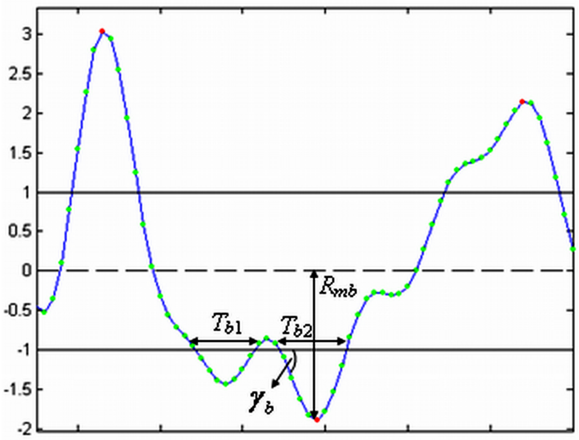

Figure 3. A portion of CMAP time series showing such a case, where a break spell is not immediately followed by an active spell, but by another break spell. The duration of these break spells is added and they are treated as a single break spell. Here, $T_b = T_{b1} + T_{b2}$.

3.2. Observed γ_a and R_{ma} Together as Predictors

[13] One may expect to improve the forecast by jointly using the observed values of γ_a and R_{ma} as predictors. Assuming that T_b is a linear function of the two variables γ_a and R_{ma}, the best-fit equation obtained from the first/training part of the CMAP data is found to be

$$T_b = 6.46\gamma_a + 3.58R_{ma} - 0.66 \qquad (3)$$

[14] Using equation (3) as the prediction equation for T_b of the second/test part and the observed values of γ_a and R_{ma}, the value of S_{FP} is found to be 3.86. Comparing this value with that in Table 3, we find only a slight improvement in the forecast skill by using equation (3) instead of equation (1).

Table 2. Correlation Coefficients Between Different Variables

		Correlation Coefficients (r)				
Variable 1	Variable 2	CMAP	OLR	U850 Wind	IMD Gridded Station Rainfall	Stochastic Forced Lorenz Model
Number of pairs		30	34	53	95	24
R_{ma}	T_b	0.75^a	0.48^a	0.32^b	0.32^a	0.80^a
R_{mb}	T_a	−0.07	0.27	0.25	0.16	0.21
γ_a	R_{ma}	0.58^a	0.32^b	0.46^a	0.45^a	0.69^a
γ_b	R_{mb}	0.47^a	0.49^a	0.47^a	0.42^a	0.48^a
R_{ma}	T_a	0.40^b	0.62^a	0.67^a	0.44^a	0.48^b
R_{mb}	T_b	0.61^a	0.63^a	0.66^a	0.68^a	0.59^a
R_{ma}	A_b	0.63^a	0.45^a	0.36^b	0.36^a	0.56^a

[a] Indicates statistical significance level greater than 99.9%.
[b] Indicates statistical significance level greater than 99.0%.

Table 3. Forecast Equations and Forecast Skill Measures

Time Series	Best Fit Equations Between T_b and R_{ma}	No. of Pairs (T_b, R_{ma})	Mean ± Std. Deviation of the Duration of Breaks	Root Mean Square Error in Forecast (S_{FP})	Correlation Skill Score (ρ_{FP})
CMAP	$T_b = 4.12R_{ma} - 0.03$	30	9.13 ± 4.81	3.93	0.60[a]
OLR	$T_b = 5.92R_{ma} - 1.4$	35	8.40 ± 5.07	4.60	0.45[b]
U850	$T_b = 3.52R_{ma} + 1.75$	44	7.33 ± 4.76	4.26	0.46[b]
IMD gridded station rainfall	$T_b = 1.82R_{ma} + 3.14$	61	7.02 ± 4.60	4.30	0.40[b]
Stochastic forced Lorenz model	$T_b = 6.39R_{ma} - 4.24$	25	10.04 ± 3.85	2.62	0.75[a]

[a] Indicates statistical significance level greater than 99.9%.

[b] Indicates statistical significance level greater than 99.5%.

3.3. γ_a as Predictor

[15] In addition, we also found that γ_a correlates well with R_{ma}. Similarly, R_{mb} is well correlated with γ_b. The best-fit empirical relation between γ_a and R_{ma} obtained from the first part of the CMAP data is

$$R_{ma} = 4.73\ \gamma_a + 0.82 \tag{4}$$

[16] To obtain a greater lead-time, use of γ_a as a predictor of T_b is desirable. For this, we used the observed γ_a to calculate R_{ma} by equation (4), and then used this calculated R_{ma} in equation (3) to predict T_b. The correlation between γ_a and T_b reduces to 0.44, though it is still statistically significant at 99% level. Although γ_a is not a better predictor of T_b than R_{ma}, it provides on an average 23 ± 14 days (38 ± 19 days) of lead-time for predicting the commencement (end) of breaks as compared to about 16 ± 11 days (31 ± 17 days) using R_{ma} alone. A possible prediction scheme is illustrated in Figure 5.

[17] Similar results were obtained for the other datasets. The best-fit line between T_b and R_{ma} determined from the first part for different datasets is given in Table 3. It is noted that a robust and statistically significant correlation is found in each of the indices, between the peak anomaly in an active phase (R_{ma}) and the duration of the following break (T_b). While prediction of the duration of the active spell from the peak anomaly of the last break spell may be a bit difficult, skillful prediction of the duration of break spells does appear a good possibility.

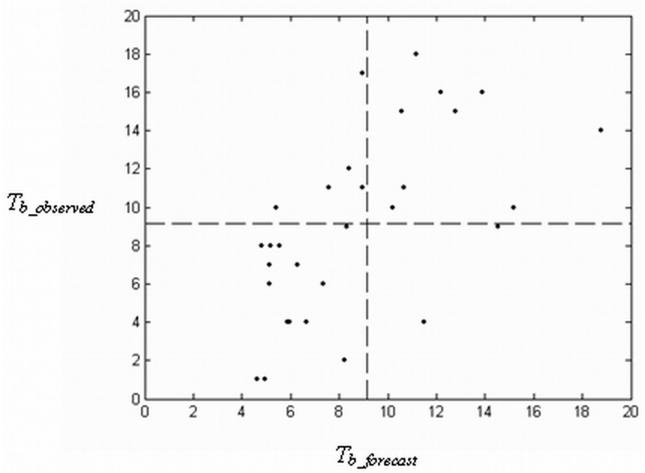

Figure 4. Actual observed value of the duration of breaks $T_{b\_observed}$ is plotted as a function of the forecast value $T_{b\_forecast}$ for the CMAP time series using R_{ma} as predictor. Points in the top right quadrant and bottom left quadrant represent correct categorical forecast. Dashed horizontal and vertical lines represent mean value of the precipitation anomaly.

4. A 'Nonlinear Dynamical' Model for the Monsoon ISO

[18] It is rather interesting, that the transition rules originally derived for the Lorenz model seem applicable to the observed monsoon ISO. Strong seasonality of the monsoon ISO or its dependence on the background mean flow, however, indicates that the monsoon ISO is a forced dynamical system. Due to the intrinsic nonlinearity of the ocean-atmosphere system, the background mean flow (or the external forcing for the ISO) has a stochastic uncertainty. Therefore, we propose that a paradigm model of the monsoon ISO could be a forced Lorenz model of the type used by *Palmer* [1993] with a forcing that contains a stochastic component.

[19] The forced Lorenz model concerned is governed by:

$$\begin{aligned} dx/dt &= -ax + ay + aF \\ dy/dt &= -xz + rx - y - F \\ dz/dt &= xy - bz \end{aligned} \tag{5}$$

[20] Here $F = -1 + \varepsilon(t)$, where ε is an independent random number chosen at each time step, from the range $[-0.1, 0.1]$. A time series $x(t)$ of 16,192 data points (after discarding the initial transients) is generated using the fourth-order Runge-Kutta algorithm. We considered a unit non-dimensional time in the forced Lorenz model as corresponding to about 25 days of atmospheric observations. Thus, in our simulation with a time step of 0.01, four time steps correspond to one day. Therefore, we applied a four time step averaging on the series $x(t)$. The resulting time series corresponds to 4,048 days of the Indian summer monsoon season (184 days) of 22 years. This time series is filtered using the 10–90 day Lanczos filter and normalized by its own standard deviation. The resulting time-series is analyzed in exactly the same way as in section 3 for the observed datasets. The results obtained are summarized in Table 2 (last column) and Table 3 (last row).

[21] It is evident from Tables 2 and 3 that the results obtained from the stochastically forced Lorenz model time series are similar to those for the CMAP dataset. Thus, the stochastically forced Lorenz model seems to be reasonable in representing the basic dynamical character of the monsoon ISO.

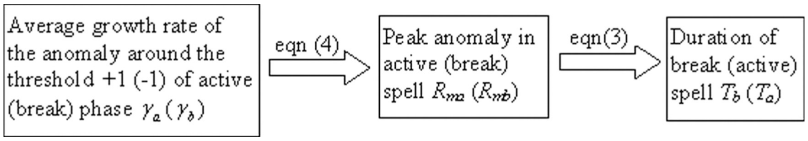

Figure 5. A possible prediction scheme for predicting the commencement (end) of breaks.

[22] Such a model may be useful in deriving insight regarding the underlying physical processes responsible for regime transitions of the monsoon ISOs. Recalling that ISOs result from a convective-radiative-dynamical feedback, the relationship between R_{ma} and T_b is indicative of such an underlying mechanism. Higher the intensity of the active condition, larger is the rainout level (drying) and stabilization of the atmosphere. The radiative and moistening processes would take that much longer to recharge for triggering a new active episode, thereby causing longer breaks.

5. Conclusions

[23] Prediction of the duration of monsoon break spells is very important. However, building a forecast model for the duration of monsoon breaks has not been obvious. The empirical rules to predict regime transitions in the Lorenz model and the forced Lorenz model provided the necessary insight to look for useful predictors for the duration of monsoon break spells. Using several different indices of the Indian summer monsoon ISOs, it is shown that the peak anomaly in an active regime can be used as a predictor for the duration of the following break spell. A simple linear regression based on this relationship shows significant and useful skill in predicting the duration of breaks, although very long breaks are slightly underestimated.

[24] The average growth rate around the threshold of active condition can also be used as a predictor of the maximum anomaly in the active spell, and through it, of the subsequent break spell duration. Although, this does not provide as good forecasts as those obtained from the maximum anomaly, it has the advantage of a greater lead-time. A good strategy would be to use the average growth rate around the threshold as an early predictor followed by the maximum anomaly as a more reliable predictor.

[25] A time series obtained from a stochastically forced Lorenz model was subjected to the same treatment as the observed datasets pertaining to the Indian monsoon intraseasonal oscillations. The correlations between different variables and the forecast skill measures for the stochastically forced Lorenz model are all very similar to those for the observed datasets. Moreover, as in the observed datasets, the breaks are more predictable than the active conditions for the stochastically forced Lorenz model too. Thus, the stochastically forced Lorenz model may be a useful tool to study some of the salient dynamical properties of the Indian summer monsoon intraseasonal oscillations.

[26] The results of the present study are based on time-filtered data (10–90 day). This is necessary to bring out intraseasonal oscillations and their transition properties. Real time forecasts of the duration of monsoon breaks require methodology to find correctly filtered data at the end point. We are currently exploring various techniques to cope with this problem.

[27] **Acknowledgment.** SD and AKM thank ISRO and NCAOR/DOD for providing financial support.

References

Evans, E., N. Bhatti, J. Kinney, L. Pann, M. Pena, S.-C. Yang, E. Kalnay, and J. Hansen (2004), RISE undergraduates find that regime changes in Lorenz's model are predictable, *Bull. Am. Meteorol. Soc.*, *85*(4), 520–524.

Gadgil, S., and P. R. S. Rao (2000), Famine strategies for a variable climate - A challenge, *Curr. Sci.*, *78*, 1203–1215.

Goswami, B. N. (2005), South Asian Monsoon, in *Intraseasonal Variability of the Atmosphere-Ocean Climate System*, edited by W. K. M. Lau and D. E. Waliser, pp. 19–61, Springer, New York.

Goswami, B. N., and R. S. Ajayamohan (2001), Intraseasonal oscillations and interannual variability of the Indian summer monsoon, *J. Clim.*, *14*, 1180–1198.

Goswami, B. N., and P. K. Xavier (2003), Potential predictability and extended range prediction of Indian summer monsoon breaks, *Geophys. Res. Lett.*, *30*(18), 1966, doi:10.1029/2003GL017810.

Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge Univ. Press, New York.

Kalnay, E., et al. (1996), The NCEP/NCAR 40–year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471.

Liebmann, B., and A. S. Catherine (1996), Description of a complete (interpolated) outgoing longwave radiation dataset, *Bull. Am. Meteorol. Soc.*, *77*, 1275–1277.

Lorenz, E. N. (1963), Deterministic non-periodic flow, *J. Atmos. Sci.*, *20*, 130–141.

Palmer, T. N. (1993), Extended-range atmospheric prediction and the Lorenz model, *Bull. Am. Meteorol. Soc.*, *74*, 49–66.

Ramamurthy, K. (1969), Monsoon of India: Some aspects of 'break' in the Indian south west monsoon during July and August, in *Forecasting Manual*, part IV18.3, India Meteorol. Dep., New Delhi.

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

Rao, Y. P. (1976), Southwest Monsoon, in *Meteorological Monograph*, *Synoptic Meteorol.* no. 1/1976, India Meteorol. Dep., New Delhi.

Sprott, J. C. (2003), *Chaos and Time Series Analysis*, Oxford Univ. Press, New York.

von Storch, H., and F. W. Zwiers (2003), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, New York.

Waliser, D., K. Lau, W. Stern, and C. Jones (2003), Potential predictability of the Madden-Julian oscillation, *Bull. Am. Meteorol. Soc.*, *84*, 33–50.

Webster, P. J., and C. Hoyos (2004), Prediction of monsoon rainfall and river discharge on 15–30 day time scales, *Bull. Am. Meteorol. Soc.*, *85*, 1745–1765.

Webster, P. J., V. O. Magana, T. N. Palmer, J. Shukla, R. T. Tomas, M. Yanai, and T. Yasunari (1998), Monsoons: Processes, predictability and the prospects of prediction, *J. Geophys. Res.*, *103*(C7), 14,451–14,510.

Xie, P., and P. A. Arkin (1996), Analyses of global monthly precipitation using gauge observations, satellite estimates and numerical predictions, *J. Clim.*, *9*, 840–858.

Yadav, R. S., S. Dwivedi, and A. K. Mittal (2005), Prediction rules for regime changes and length in a new regime for the Lorenz model, *J. Atmos. Sci.*, *62*, 2316–2321.

S. Dwivedi and A. K. Mittal, M. N. Saha Centre of Space Studies and K. Banerjee Centre of Atmospheric and Ocean Studies, Institute of Interdisciplinary Studies, and Department of Physics, University of Allahabad, Allahabad, UP 211002, India. (dwivedisuneet@rediffmail.com)

B. N. Goswami, Indian Institute of Tropical Meteorology, Pune 411008, India.