# Bayesian dynamic modeling for monthly Indian summer monsoon rainfall using El Niño−Southern Oscillation (ENSO) and Equatorial Indian Ocean Oscillation (EQUINOO)

Rajib Maity[1] and D. Nagesh Kumar[1]

[1] There is an established evidence of climatic teleconnection between El Niño−Southern Oscillation (ENSO) and Indian summer monsoon rainfall (ISMR) during June through September. Against the long-recognized negative correlation between ISMR and ENSO, unusual experiences of some recent years motivate the search for some other causal climatic variable, influencing the rainfall over the Indian subcontinent. Influence of recently identified Equatorial Indian Ocean Oscillation (EQUINOO, atmospheric part of Indian Ocean Dipole mode) is being investigated in this regard. However, the dynamic nature of cause-effect relationship burdens a robust and consistent prediction. In this study, (1) a Bayesian dynamic linear model (BDLM) is proposed to capture the dynamic relationship between large-scale circulation indices and monthly variation of ISMR and (2) EQUINOO is used along with ENSO information to establish their concurrent effect on monthly variation of ISMR. This large-scale circulation information is used in the form of corresponding indices as exogenous input to BDLM, to predict the monthly ISMR. It is shown that the Indian monthly rainfall can be modeled in a better way using these two climatic variables concurrently (correlation coefficient between observed and predicted rainfall is 0.82), especially in those years when negative correlation between ENSO and ISMR is not well reflected (i.e., 1997, 2002, etc.). Apart from the efficacy of capturing the dynamic relationship by BDLM, this study further establishes that monthly variation of ISMR is influenced by the concurrent effects of ENSO and EQUINOO.

## 1. Introduction

[2] Indian summer monsoon rainfall (ISMR) is crucial for socio-economic status of India. For instance, deficit of 19% in the Indian summer monsoon in 2002 caused a significant decrease in agricultural production and economic status of the country. Although reliable prediction of monthly rainfall is more challenging as compared to seasonal forecast, it is essential for planning and devising agricultural strategies, decision making and better management of water resources. Among the various approaches of rainfall prediction, incorporation of climate information as an external input influencing the rainfall is gaining more and more interest in recent years.

### 1.1. El Niño−Southern Oscillation and Indian Summer Monsoon

[3] El Niño−Southern Oscillation (ENSO) is one of the main sources of interannual variability in weather and climate around the world [*Kiladis and Diaz*, 1989]. Attempts were made to forecast hydrologic variables, like rainfall, streamflow, etc., using ENSO information all over the world [*Ropelewski and Halpert*, 1987; *Kahya and Dracup*, 1993; *Dracup and Kahya*, 1994; *Eltahir*, 1996; *Jain and Lall*, 2001]. Association of Southern Oscillation and ISMR was recognized long ago [*Walker*, 1923, 1924; *Normand*, 1953]. Significant correlation between ENSO and ISMR was a major advancement to explain interannual variation of the ISMR [*Pant and Parthasarathy*, 1981; *Rasmusson and Carpenter*, 1983]. General impact of an El Niño event on the Indian monsoon is shown to be associated with lower-than-normal rainfall and opposite in case of a La Niña event [*Rasmusson and Carpenter*, 1983; *Khandekar and Neralla*, 1984; *Mooley and Paolino*, 1989]. Regarding the teleconnection mechanism between ISMR and ENSO events, *Rasmusson and Carpenter* [1983] concluded that "episodes of above normal SST's (Sea Surface Temperature) over the eastern and central equatorial Pacific are associated with a low SOI (Southern Oscillation Index), i.e., negative pressure anomalies in the southeast Pacific, and positive anomalies over the Indian Ocean region, weaker than normal southwest monsoon over the Arabian Sea, and below normal rainfall over India." However,

[1]Department of Civil Engineering, Indian Institute of Science, Bangalore, India.

contrary to the long-recognized negative correlation between the ISMR and El Niño events, India received slightly above normal rainfall in 1997 [*Li et al.*, 2001] when El Niño was observed in the Pacific Ocean. On the other hand, in 2002 the failure of Indian summer monsoon was completely unanticipated, although it was a weak El Niño year. No model could predict such a large deficit of ISMR in 2002 [*Gadgil et al.*, 2003]. According to *Kane* [1998], the relationship is not unique for all the El Niño events. According to him, "…. in some years, some other factors might be playing important roles……" *Krishna Kumar et al.* [1999] have shown that the historical relationship has been broken after 1970. These unanticipated experiences suggest that the response of monsoon to El Niño is not yet assessed adequately [*Gadgil et al.*, 2003; *Gadgil*, 2003] or more importantly that there are some other causative climate forcing events which are also influencing the Indian rainfall concurrently.

## 1.2. Equatorial Indian Ocean Oscillation (EQUINOO)

[4] Equatorial Indian Ocean Oscillation (EQUINOO) is the atmospheric part of Indian Ocean Dipole (IOD) mode [*Gadgil et al.*, 2004]. IOD mode is basically an anomalous state of air-sea interaction over eastern and western part of tropical Indian Ocean, which alters the atmospheric circulation as well as the weather pattern over Indian Ocean and its surroundings [*Saji et al.*, 1999; *Webster et al.*, 1999; *Ashok et al.*, 2001]. During the summer monsoon season (June–September), the convection over the eastern part of the equatorial Indian Ocean (EEIO, $90°E–110°E$, $10°S$–equator) is negatively correlated to that over the western part of the equatorial Indian Ocean (WEIO, $50°E–70°E$, $10°S–10°N$). The anomalies in the sea level pressure and the zonal component of the surface wind along the equator are consistent with the convection anomalies. When the convection is enhanced (suppressed) over the WEIO, the anomalous surface pressure gradient, high to low, is toward the west (east) so that the anomalous surface wind along the equator becomes easterly (westerly). The oscillation between these two states is called the Equatorial Indian Ocean Oscillation (EQUINOO) and equatorial zonal wind index (EQWIN) is considered as an index of EQUINOO. EQWIN is defined as the negative of the anomaly of the zonal component of surface wind in the equatorial Indian Ocean region ($60°E–90°E$, $2.5°S–2.5°N$) normalized by its standard deviation [*Gadgil et al.*, 2003, 2004].

[5] Recent meteorological observations indicate a strong link between ISMR and EQUINOO. This may be due to the association of large-scale monsoon rainfall over the Indian region with the northward propagation of convective system generated over the Indian Ocean region [*Gadgil et al.*, 2003, 2004]. *Gadgil et al.* [2004] suggested that an educated guess could be made about the Indian summer monsoon rainfall by knowing the prior EQUINOO status. For example, in 1983 and 1994, ENSO was small but EQUINOO was positive and India received excess rainfall. On the other hand, in 1979 and 1985, ENSO signal was favorable for monsoon but EQUINOO was unfavorable and India received drought. In 2002, both the ENSO and EQUINOO were unfavorable and a severe drought occurred. *Gadgil et al.* [2004] show that all the extremes in the Indian summer

monsoon rainfall (greater than ±1 standard deviation) from 1958 to 2003 are statistically associated with favorable (unfavorable) phases of ENSO or EQUINOO or both. Hence EQUINOO has significant relation with Indian monsoon along with ENSO.

## 1.3. Bayesian Dynamic Linear Models (BDLMs)

[6] The relationship between ISMR and large-scale atmospheric circulation is dynamic and nonstationary owing to the effect of climate change. A simple non-Bayesian multiple linear regression model is investigated to predict the ISMR using the information of ENSO and EQUINOO. However, owing to its inherent stationarity assumption and static nature, it is unable to capture the dynamic relationship and shows poor prediction performance (please refer to section 5). On the other hand, dynamic nature of the relationship between climate information and corresponding responses of hydrological events motivates use of a dynamic model with Bayesian updating of the parameters. Research, development and usability of this type of model can be found elsewhere [*Pole et al.*, 1994; *Besag et al.*, 1995]. Dynamic nature of such models makes them potential for various application fields [*Bernier*, 1994; *Berger and Insua*, 1998; *Krishnaswamy et al.*, 2000; *Berliner et al.*, 2000; *Krishnaswamy et al.*, 2001]. Stationarity assumption of data set can be relaxed for these models which is a very useful property. Such models also allow incorporation of exogenous inputs. Finally, incorporation of all prior information, with allowance for on-line external intervention, makes it possible to use the principle of management by exception, which is fundamental in the Bayesian forecasting philosophy [*West and Harrison*, 1997]. In the present study, a Bayesian dynamic linear model (BDLM) is proposed and used to capture the monthly variation of Indian summer monsoon rainfall using ENSO and EQUINOO index as external input. A complete description of the model is presented in section 3.

[7] The objective of this paper is to capture the dynamic relationship between monthly all India summer monsoon rainfall and large-scale circulation indices of ENSO and EQUINOO using BDLM. Useful transformation of raw climate index series is also explained for effective extraction of climatic signal for dynamic linear model.

[8] Organization of this paper is as follows. Data used for this study and their sources are mentioned in sections 2. In section 3, Bayesian dynamic linear model is described. Methodology is described in section 4. In section 5, results and discussions are presented. Finally, conclusions are presented in section 6.

## 2. Data

[9] To represent the ENSO, sea surface temperature anomaly from Niño 3.4 region ($5°S–5°N$, $170°W–120°W$) (Niño 3.4 SSTA) is used in this study (Figure 1). Monthly Niño 3.4 SSTA data are obtained from the web site of National Weather Service, Climate Prediction Centre of NOAA (http://www.cpc.noaa.gov/data/indices/) for the period January 1958 to December 2003. Negative of Niño 3.4 SSTA data is used to represent the ENSO index to make it positively correlated with Indian rainfall during monsoon months.
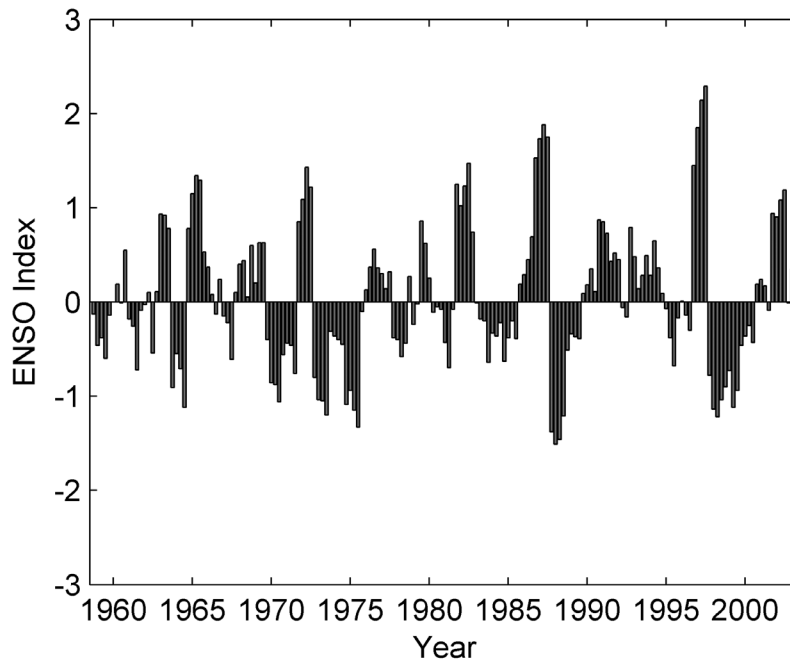
**Figure 1.**  Plot of ENSO index during monsoon months for the period 1959–2003.

[10]  On the basis of *Gadgil et al.* [2004], equatorial zonal wind index (EQWIN) is used to represent EQUINOO index (Figure 2). To compute this index, monthly surface wind data were obtained from National Center for Environmental Prediction [*Kalnay et al.*, 1996] (http://www.cdc.noaa.gov/ Data sets) for the period January 1958 to December 2003.

[11]  All India monthly rainfall data are obtained from the web site of Indian Institute of Tropical Meteorology, Pune, India (http://www.tropmet.res.in/data.html) for the period January 1901 to December 2003. These data are also available on the web site of International Research Institute for Climate Prediction (http://iridl.ldeo.columbia.edu/).
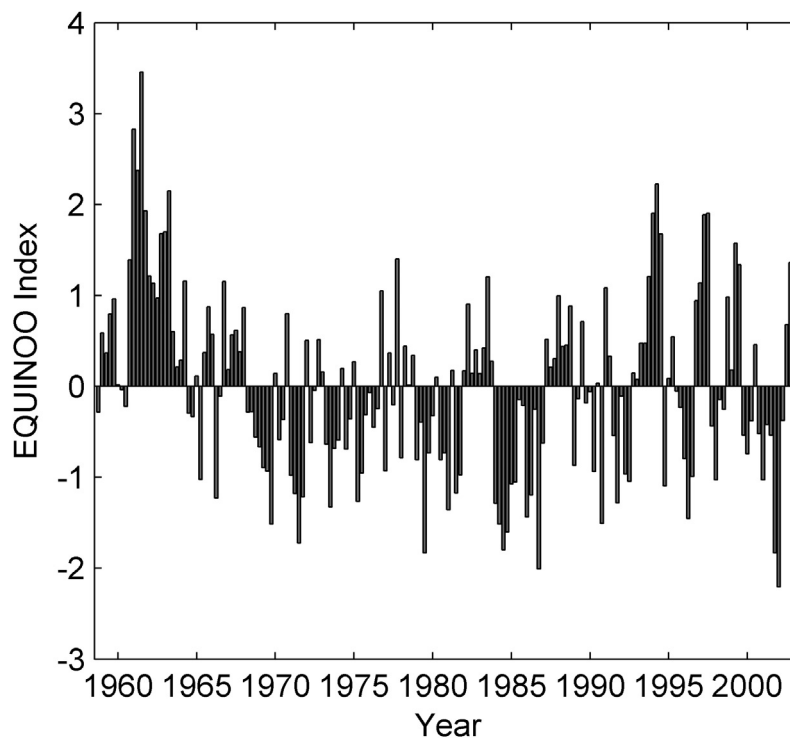


**Figure 2.**  Plot of EQUINOO index during monsoon months for the period 1959–2003.

Month-wise long-term (1901–1958) mean rainfall values are shown in Table 1. These are considered as long-term climatological mean values and are used to obtain the corresponding monthly anomalies.

## 3. Bayesian Dynamic Linear Models

[12] Mathematical framework of Bayesian dynamic linear models (BDLMs) was developed by *West and Harrison* [1997]. In this paper, detailed mathematical proof and related issues are not presented. Rather the model has been discussed for two different causative variables, which will be used in the present study. Continuously changing relationship between the target time series of Indian monthly rainfall and two regressor time series (ENSO index and EQUINOO index) is captured by this model. The target time series is assumed to have a deterministic part (monthly climatological mean) and a stochastic part. The stochastic part of the time series is being captured by using the information from two regressor time series and added to the deterministic part. The complete model is described below. Observation equation is

$$Y_{i,j} = F_{i,j}^T \Theta_{i,j} + \nu_{i,j}, \qquad (1)$$

where $Y_{i,j}$ is the observed value of the target time series for the $j$th month ($j = 1, \ldots, 12$) of $i$th year; $F_{i,j}^T$ is the transpose of regression vector for $j$th month of $i$th year; $\Theta_{i,j}$ is the regression parameter vector for $j$th month of $i$th year; $\nu_{i,j}$ is the normally distributed observational error for $j$th month of $i$th year with mean 0 and unknown variance V, i.e., $\nu_{i,j} \sim$ N[0, $V$].

[13] Regression vector for $j$th month of $i$th year is

$$F_{i,j} = \begin{bmatrix} 1 \\ f * EN_{i,j-\kappa} \\ (1-f) * EQ_{i,j-\lambda} \end{bmatrix}, \qquad (2)$$

where $EN_{i,j-\kappa}$ is the transformed ENSO index for $\kappa$ months prior to $j$th month of $i$th year; $EQ_{i,j-\lambda}$ is the transformed EQUINOO index for $\lambda$ months prior to $j$th month of $i$th year; $\kappa$ and $\lambda$ are the lead times (in months) for ENSO index and EQUINOO index respectively; $f$ is the relative weightage factor for ENSO index; and (1 − f) is the relative weightage factor for EQUINOO index.

[14] Regression parameter vector for $j$th month of $i$th year is

$$\Theta_{i,j} = \begin{bmatrix} \overline{Y}_j \\ \theta_{i,j}^{en} \\ \theta_{i,j}^{eq} \end{bmatrix}, \qquad (3)$$

where $\overline{Y}_j$ is the long-term climatological mean value for $j$th month; $\theta_{i,j}^{en}$ is the regression parameter for ENSO index for $j$th month of $i$th year; $\theta_{i,j}^{eq}$ is the regression parameter for EQUINOO index for $j$th month of $i$th year.

[15] Relative weightage factors for ENSO index and EQUINOO index are incorporated to the model to investigate their relative influences on monthly variation of all India rainfall. Instead of using raw circulation indices

**Table 1.** Long-Term (1901–1958) Mean Monthly Rainfall

| Month | Mean Monthly Rainfall, cm |
|---|---|
| Jan | 1.26 |
| Feb | 1.44 |
| March | 1.52 |
| April | 2.72 |
| May | 5.24 |
| June | 16.0 |
| July | 27.67 |
| Aug | 24.35 |
| Sep | 17.09 |
| Oct | 7.84 |
| Nov | 3.28 |
| Dec | 1.08 |

(ENSO and EQUINOO), transformed indices will be used, as described later.

[16] As the first element of $\Theta_{i,j}$, i.e., $\overline{Y}_j$, is the known long-term climatological mean value for $j$th month, step by step updating of it is done by substituting the known long-term mean value for the corresponding month. Other elements, i.e., $\theta_{i,j}^{en}$ and $\theta_{i,j}^{eq}$, are updated sequentially by the system of equations at each time step. For generality in description, suffixes "en" and "eq" are omitted for those equations which are applicable for both $\theta_{i,j}^{en}$ and $\theta_{i,j}^{eq}$.

[17] System of equations is as follows:

$$\theta_{i,j} = \theta_{i,j-1} + \omega_{i,j} \quad \text{for} \quad \forall\, i \quad \text{and} \quad j = 2 \ldots 12, \qquad (4a)$$

$$\theta_{i,j} = \theta_{i-1,12} + \omega_{i,j} \quad \text{for} \quad \forall\, i \quad \text{and} \quad j = 1 \qquad (4b)$$

where $\omega_{i,j}$ is the Student-T distributed system evolution error with degree of freedom $n$, for $j$th month of $i$th year with parameter 0 and $W_{i,j}$, i.e., $\omega_{i,j} \sim T_n \lfloor 0, W_{i,j} \rfloor$. Degree of freedom n, for $j$th month of $i$th year is

$$n = (i - 1) * 12 + j - 1. \qquad (5)$$

[18] Model parameters are updated at each time step, to obtain one-step-ahead forecast and posterior distribution for the next time step. Procedure to update the model parameters is described below. Suffixes of the parameters indicate the time step. In general, the first suffix denotes year and the second denotes month. In the following expressions, $T_n$, $N$ and $G$ stand for Student-T distribution with $n$ degrees of freedom, Normal distribution and Gamma distribution, respectively.

[19] 1. To initiate the model, initial information $D_{1,0}$ is to be provided by the forecaster with which initial distributional form of regression parameters will be decided, i.e., $(\theta_{1,0}/D_{1,0}) \sim T_0 \lfloor m_{1,0}, C_{1,0} \rfloor$. Again, $(\phi/D_{1,0}) \sim G \lfloor n_{1,0}/2, d_{1,0}/2 \rfloor$, where $\phi$ is the precision parameter and defined as $\phi = V^{-1}$. The parameters $m_{1,0}$, $C_{1,0}$, $n_{1,0}$ and $d_{1,0}$ are the initial beliefs of the forecaster. Generally, initial information $D_{1,0}$ consists of past experience and information available to the forecaster. In general, $D_{i,j}$ consists of all available information and observation up to the end of $j$th month of $i$th year. Thus initial information is improved at each time step as new observations are made.

[20] 2. To discuss the parameter updating methodology, without the loss of generality, let us assume, at some
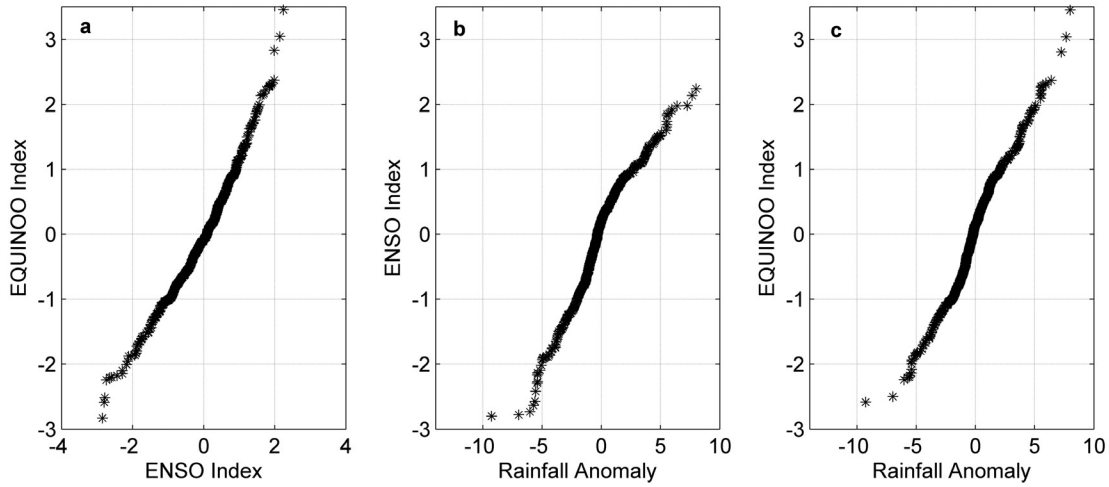
**Figure 3.** Q-Q plots between different combinations among ENSO index, EQUINOO index, and rainfall anomaly.

particular time step $(i, j - 1)$, the posterior distribution for regression parameters $\theta_{i,j-1}$ and distributional form of precision parameter $\phi$ is known.

$$\left(\theta_{i,j-1}/D_{i,j-1}\right) \sim T_n\left[m_{i,j-1}, C_{i,j-1}\right]$$

$$\text{and } \left(\phi/D_{i,j-1}\right) \sim G\left[n_{i,j-1}/2, d_{i,j-1}/2\right],$$

with $n$ as defined in equation (5) for the time step $(i, j - 1)$.

[21] 3. Prior distribution for $\theta_{i,j}$ is obtained as follows:

$$\left(\theta_{i,j}/D_{i,j-1}\right) \sim T_n\left[m_{i,j-1}, R_{i,j}\right],$$

with $n$ as defined in equation (5) for the time step $(i, j - 1)$ and

$$R_{i,j} = C_{i,j-1} + W_{i,j}, \tag{6}$$

where $W_{i,j}$ is the system evolution variance for $j$th month of $i$th year. It is practically difficult to assign the sequence of evolution variance $\{W_{i,j}\}$ without relating it to some previous known variance. Therefore a discount factor $\delta$ $(0 < \delta < 1)$ for both $\theta_{i,j}^{en}$ and $\theta_{i,j}^{eq}$ is introduced such that

$$R_{i,j} = C_{i,j-1}/\delta. \tag{7}$$

This reflects the real fact that $R_{i,j} > C_{i,j-1}$. Using equation (7) in equation (6) it can be shown that

$$W_{i,j} = C_{i,j-1}\left(\delta^{-1} - 1\right). \tag{8}$$

[22] 4. One-step ahead forecast distribution is obtained as follows:

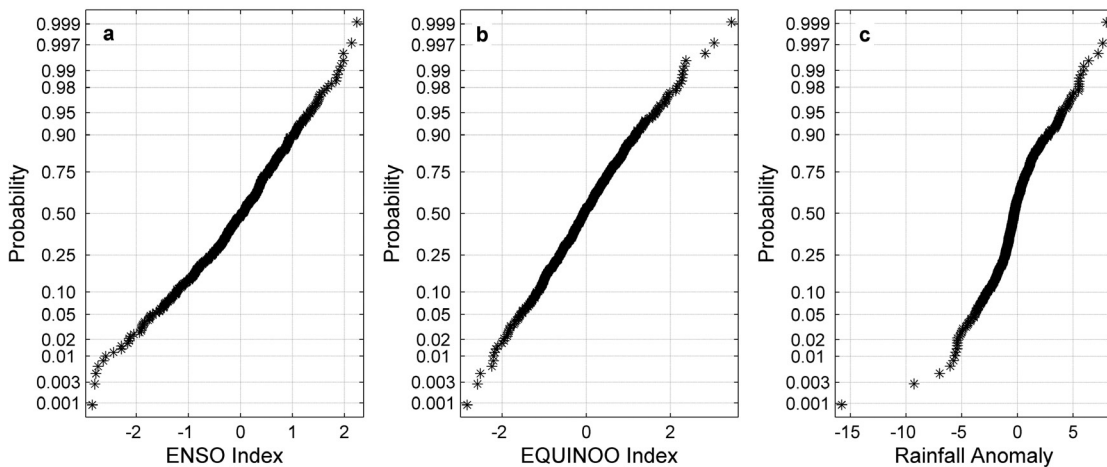$$\left(Y_{i,j}/D_{i,j-1}\right) \sim T_n\left[F_{i,j}, Q_{i,j}\right],$$



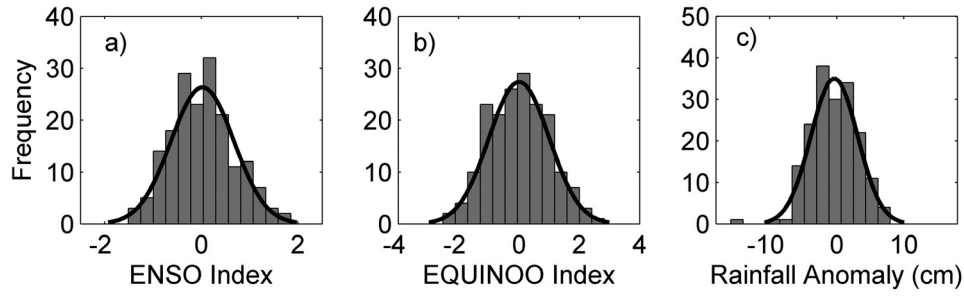**Figure 4.** Normal probability plots for ENSO index, EQUINOO index, and rainfall anomaly.

**Figure 5.** Histogram distributions with superimposed fitted normal density for ENSO index, EQUINOO index, and rainfall anomaly.

with $n$ as defined in equation (5) for the time step $(i, j)$ and

$$F_{i,j} = \overline{Y}_j + EN_{i,j-\kappa} * f * m^{en}_{i,j-1} + EQ_{i,j-\lambda} * (1 - f) * m^{eq}_{i,j-1} \quad (9)$$

$$Q_{i,j} = (EN_{i,j-\kappa} * f)^2 * R^{en}_{i,j} + \{EQ_{i,j-\lambda}*(1-f)\}^2 * R^{eq}_{i,j} + S_{i,j-1}, \quad (10)$$

where

$$S_{i,j-1} = \frac{d_{i,j-1}}{n_{i,j-1}}. \quad (11)$$

[23] 5. Posterior distribution for $\theta_{i,j}$ is obtained as follows:

$$(\theta_{i,j}/D_{i,j}) \sim T_n[m_{i,j}, C_{i,j}],$$

with $n$ as defined in equation (5) for the time step $(i, j)$ and

$$m_{i,j} = m_{i,j-1} + A_{i,j}e_{i,j}, \quad (12)$$

$$C_{i,j} = R_{i,j}S_{i,j}/Q_{i,j}, \quad (13)$$

$$S_{i,j} = d_{i,j}/n_{i,j}, \quad (14)$$

$$e_{i,j} = Y_{i,j} - F_{i,j}, \quad (15)$$

$$A^{en}_{i,j} = EN_{i,j-\kappa} * f * R_{i,j}/Q_{i,j}, \quad (16)$$

$$A^{eq}_{i,j} = EQ_{i,j-\lambda} * (1 - f) * R_{i,j}/Q_{i,j}. \quad (17)$$

**Table 2.** Scale Factors and Shift Factors for ENSO Index and EQUINOO Index

| | Period of Considered Monthly Rainfall Anomaly | | | |
|---|---|---|---|---|
| | All Months in the Year | | Only Monsoon Months (June Through September) | |
| Indices | Scale Factor | Shift Factor | Scale Factor | Shift Factor |
| ENSO index | 2.62 | −0.025 | 3.90 | −0.184 |
| EQUINOO index | 2.46 | −0.104 | 3.66 | −0.301 |

Distributional form of scale parameter $\phi$ at time step $(i, j)$ is obtained as follows:

$$(\phi/D_{i,j}) \sim G[n_{i,j}/2, d_{i,j}/2],$$

where

$$n_{i,j} = n_{i,j-1} + 1 \quad (18)$$

$$d_{i,j} = d_{i,j-1} + S_{i,j-1}e^2_{i,j}/Q_{i,j}. \quad (19)$$

[24] It is also important to note that the discount factor $\delta$ plays an important role, indicating the loss of information between successive observations. For instance, large value of $\delta$ indicates smaller rate of decay of past information. In fact, $\delta = 1$ yields a static model. On the other hand, small value of $\delta$ implies faster rate of decay of past information [*West and Harrison*, 1997]. Optimal choice of $\delta$ can be made on the basis of the model performances.

[25] Another important point is that the model could predict the uncertain future rainfall values as a distributional

**Table 3.** Performance Statistics for Different Combinations of Lead Times for ENSO and EQUINOO[a]

| EQUINOO Lead Time, Months | ENSO Lead Time, Months | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | 0.49 | **0.50** | 0.46 | 0.39 |
| | 0.87 | **0.87** | 0.87 | 0.85 |
| | −279.47 | **−277.08** | −280.01 | −286.07 |
| | 8.68 | **8.61** | 8.97 | 9.76 |
| 2 | 0.41 | 0.40 | 0.37 | 0.27 |
| | 0.86 | 0.85 | 0.85 | 0.83 |
| | −283.87 | −282.33 | −285.40 | −292.41 |
| | 9.60 | 9.67 | 9.93 | 11.08 |
| 3 | 0.40 | 0.39 | 0.35 | 0.27 |
| | 0.85 | 0.85 | 0.85 | 0.83 |
| | −282.27 | −281.47 | −284.40 | −290.60 |
| | 9.70 | 9.77 | 10.11 | 11.00 |
| 4 | 0.45 | 0.44 | 0.40 | 0.31 |
| | 0.86 | 0.86 | 0.86 | 0.84 |
| | −277.92 | −277.77 | −280.95 | −286.62 |
| | 9.14 | 9.26 | 9.60 | 10.49 |

[a]From top to bottom in each cell, the statistics are as follows: correlation coefficient (CC) between observed and predicted rainfall anomaly; CC between observed and predicted rainfall; log likelihood; and mean square error (MSE). Higher values of first three statistics and lower value of last statistics indicate better model performance. Performance statistics in the cell corresponding to best combinations of lead times are shown in boldface.

**Table 4.** Best Values of the Subjective Parameters

| Data Set | C Values | Relative Weightages | δ Values | Lead Time, months |
|---|---|---|---|---|
| Rainfall anomaly | 15 | ... | ... | ... |
| ENSO index | 20 | 0.61 | 0.89 | 2 |
| EQUINOO index | 20 | 0.39 | 0.95 | 1 |

form as against the point forecast by most of the other models like multiple linear regression, Box-Jenkins model, etc. Prediction as a distributional form is useful for making probabilistic inferences which are of considerable practical use. Computation of confidence interval (CI) for the forecasted values is possible, which is an advantageous property for many fields of application.

## 4.  Methodology

[26]  Raw climate indices are transformed to extract the climate signal more effectively. It is done by transforming both the regressor series in such a way that the transformed regressor series will have identical distribution to that of the target series. Before transforming the raw data set, it should be verified that all the regressor series as well as target series have similar distribution.

[27]  To illustrate the data transformation, let two series, $X$ and $Y$, have similar distribution but not necessarily identical parameters. Also, let $Y$ be the target series and $X$ be the regressor series. Series $X$ will be transformed in such a way that the transformed $X$ series will have identical distribution as that of $Y$. To do this $X$ is transformed using following equations:

$$X'(i) = X(i) * \frac{\sigma_Y}{\sigma_X} \qquad (20)$$

$$X''(i) = X'(i) + (\mu_Y - \mu_{X'}) + C, \qquad (21)$$

where $\sigma_X$, $\sigma_Y$ are the standard deviations of series $X$ and $Y$ respectively; $\mu_Y$ is the mean of series $Y$; $\mu_{X'}$ is the mean of the transformed series $X'$; $C$ is any constant to shift the entire time series by some desirable amount. In the above expressions, $\frac{\sigma_y}{\sigma_x}$ is known as scale factor and $(\mu_Y - \mu_{X'}) + C$ is called shift factor. After transforming the series $X$ using equations (20) and (21) (putting any value of $C$), the obtained series $X''$ will have identical parameters as those of series $Y$ (both mean and standard deviation).

[28]  It can be shown that $X''$ and $Y$ will be closer to each other as compared to $X$ and $Y$, irrespective of the nature of relationship between them. If there is some relationship between the target and the regressor time series, then an obtained closeness between the target series and the transformed regressor series will be better. If $X''$ is used, instead of $X$, as the causal variable of $Y$, it will be more effective and it is observed that the performance of the model has enhanced by using the transformed regressor time series.

[29]  The technique described above is used in the present study to transform the time series of ENSO index and EQUINOO index, to have same mean and standard deviation as that of rainfall anomaly series. However, time series of ISMR, ENSO index and EQUINOO index are statistically tested to check whether they have similar distribution or not. From the Q-Q plots between different combinations among ENSO index, EQUINOO index and rainfall anomaly (Figure 3), it can be inferred that all these data sets follow a similar distribution. Again normal probability plots (Figure 4) and histogram distributions with superimposed fitted normal density (Figure 5) indicate that all of them follow normal distribution, if few outliers are neglected.

[30]  Both ENSO and EQUINOO index are transformed using equations (20) and (21), so that finally all three series (rainfall anomaly, ENSO index and EQUINOO index) become identically distributed. From this analysis a set of scale factors and shift factors for ENSO index and
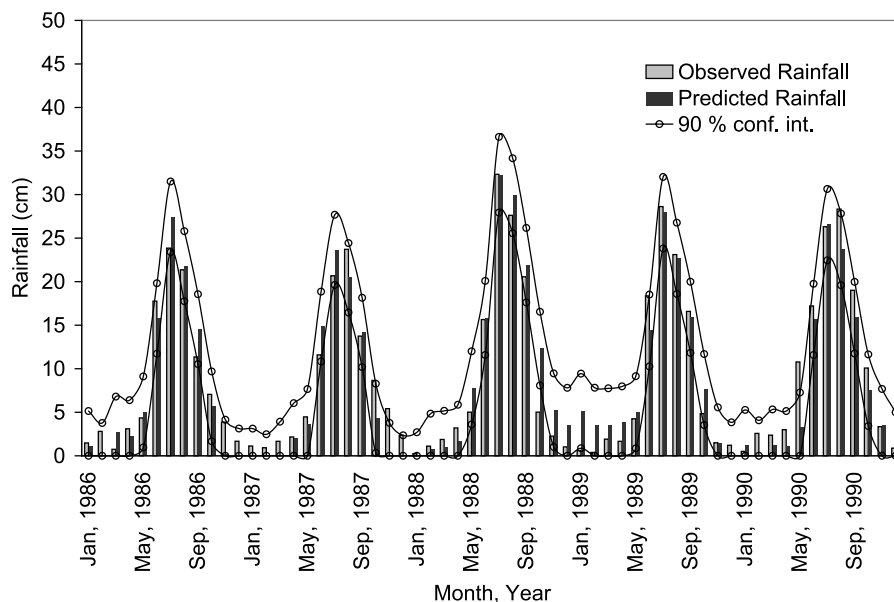


**Figure 6a.**  Comparison between observed and predicted monthly rainfall (January 1986 to December 1990).
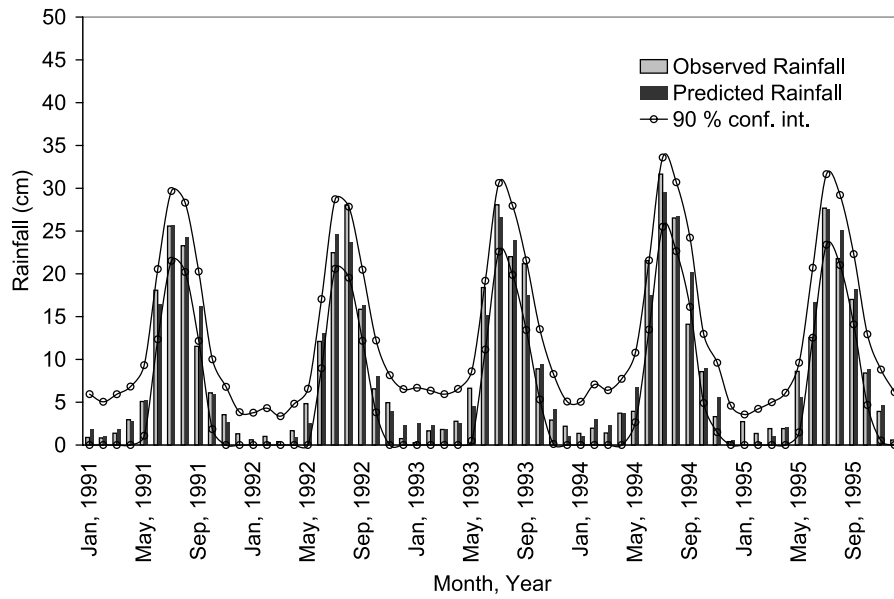
**Figure 6b.** Comparison between observed and predicted monthly rainfall (January 1991 to December 1995).

EQUINOO index is obtained (Table 2). However, India receives more than 80% of annual rainfall during the monsoon months of June, July, August and September (JJAS). Hence a similar analysis is performed considering monthly rainfall anomaly series for these four monsoon months. From this analysis another set of scale factors and shift factors for ENSO index and EQUINOO index is obtained. Both the sets of scale factors and shift factors are shown in Table 2. It is noticed from Table 2 that there is significant difference between two sets of scale factors. As it is necessary to give more prominence to the monsoon months, the ENSO index and the EQUINOO index are transformed as per the second case. The variable C for the above analysis was assumed to be

0. However, its value will be determined by the model performance during calibration of the model. Other subjective variables will also be determined in a similar way.

## 5. Results and Discussions

[31] The model is calibrated on the basis of prediction performance of monthly rainfall for monsoon months (June through September) for the period 1959–1985. Subjective parameters, like discount factors $\delta$, relative weightage factors ($f$ and $1 - f$) and values of C are determined on the basis of the model performance during this period. Model performances are judged on the basis of mean square
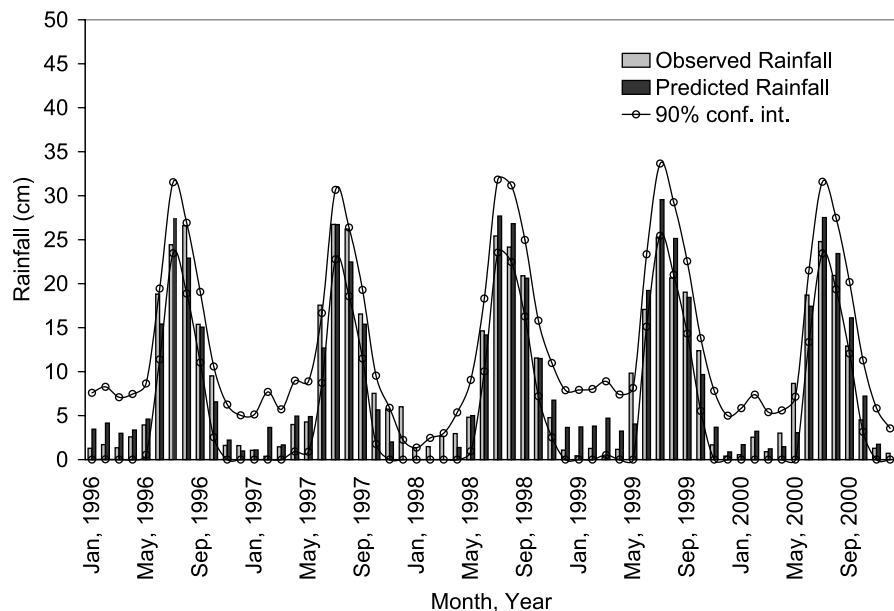


**Figure 6c.** Comparison between observed and predicted monthly rainfall (January 1996 to December 2000).
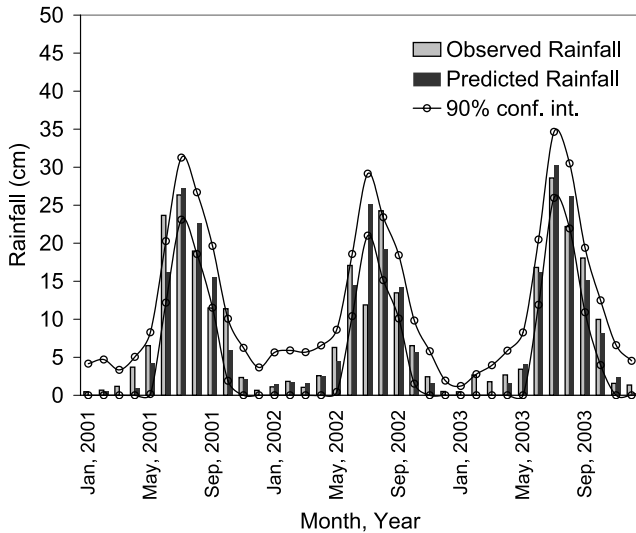
**Figure 6d.** Comparison between observed and predicted monthly rainfall (January 2001 to December 2003).

error (MSE), log likelihood and correlation coefficient between predicted and observed monthly rainfall during monsoon months. Best values of lead times ($\kappa$ and $\lambda$) are selected after investigating all the possible combinations of lead times varying from 1 to 4 months for both ENSO and EQUINOO during the model calibration period. Table 3 shows the results for different combinations of lead times. Each cell shows four different statistics to measure the model performance. From top to bottom these are (1) Correlation coefficient (CC) between observed and predicted rainfall anomaly, (2) CC between observed and predicted rainfall (3) Log likelihood and (4) Mean Square Error (MSE). Higher values of first three statistics and lower value of last statistic indicate better model performance. It can be noticed from Table 3, that lead times of 2 months and 1 month for ENSO and EQUINOO, respectively, produce the best performance. It can be also noticed that there is a minor improvement in performance statistics for EQUINOO lead time of 4 months, compared to 3 months. However, existing literature and practical intuition says that EQUI-NOO is more immediate factor compared to ENSO for ISMR. Thus lead time for EQUINOO should be less than or equal to the lead time for ENSO (i.e., 2 months). Considering all these factors a lead time of 2 months for ENSO and 1 month for EQUINOO is selected.

[32] The best values of all the subjective parameters are shown in Table 4. Using these parameters, model performance is investigated for the period 1986–2003. Correlation coefficient between predicted and observed monthly rainfall is 0.82 during monsoon months of this period. Apart from the statistical correlation coefficient, comparison between predicted and observed monthly rainfall are presented by bar plots (Figures 6a–6d).

[33] As mentioned in section 1.3, a simple non-Bayesian multiple linear regression model, which has the mathematical form

$$R_{i,j} = \alpha.EN_{i,j-\kappa} + \beta.EQ_{i,j-\lambda} + \nu_{i,j}, \qquad (22)$$

is investigated to compare its performance with BDLM. In equation (22), $R_{i,j}$ is the observed value of the target time series (ISMR) for the $j$th month ($j = 1, \ldots., 12$) of $i$th year; $EN_{i,j-\kappa}$ is the ENSO index for $\kappa$ months prior to $j$th month of $i$th year; $EQ_{i,j-\lambda}$ is the EQUINOO index for $\lambda$ months prior to $j$th month of $i$th year and $\nu_{i,j}$ is the noise term for $j$th month of $i$th year. As it is observed that 2 months and 1 month lead time for ENSO and EQUINOO index, respectively, are most effective, values of $\kappa$ and $\lambda$ are used as 2 and 1, respectively, in this approach also. First, the coefficients, $\alpha$ and $\beta$ are estimated by least squares method based on the period 1959–1985 and the estimated model is used for prediction for the period 1986–2003 to test the model performance, as done in BDLM, too. By doing this, estimated values of $\alpha$ and $\beta$ are calculated as −6.173 and 3.056. Thus the estimated model has the mathematical form as

$$\hat{R}_{i,j} = -6.173.EN_{i,j-\kappa} + 3.056.EQ_{i,j-\lambda}, \qquad (23)$$

where $\hat{R}_{i,j}$ is the predicted value of the target time series (ISMR). Now, using this model, prediction is made for the period 1986 to 2003. Correlation coefficient (CC) between the observed and this predicted rainfall is obtained as 0.27, whereas in case of BDLM, CC between observed and predicted rainfall during this period was obtained as 0.82. Thus it is clear that BDLM is really superior to a simple non-Bayesian multiple linear regression model. As it was also indicated in section 1.3, the extra skill of Bayesian model comes from its dynamic nature [*West and Harrison*, 1997] and capability of handling nonstationarity [*Bernier*, 1994].

[34] As the model can predict the uncertain future rainfall values as a distributional form, computation of confidence interval (CI) for the forecasted values is possible, which helps in making a decision with required statistical confidence level. Such predictions are much more advantageous than a point prediction. In Figures 6a–6d, 90% confidence intervals are also shown. It is noticed that the observed monthly rainfall is well captured by this confidence interval for almost all the years (except July 2002). This indicates the capability of the model to capture dynamic relationship and successfully predict the monthly rainfall, using both the large-scale circulation information of ENSO and EQUINOO.

[35] From Figures 6a–6d, concurrent effect of ENSO and EQUINOO can be observed. For instance, in 1987, India was expected to receive a lower-than-normal rainfall due to the El-Niño event (Figure 1). It was further intensified owing to the presence of negative EQUINOO index in the same year (Figure 2) and the model has correctly predicted low values of monthly rainfall for that year (Figure 6a). In the very next year, higher-than-normal rainfall has been successfully predicted (Figure 6a) owing to the joint effect of La Niña (Figure 1) and occurrence of positive EQUINOO index (Figure 2). In 1997, normal rainfall values are predicted (Figure 6c), although a drought was expected owing to the occurrence of El-Niño (Figure 1) and this correct prediction is made owing to the presence of positive EQUINOO index (Figure 2). In 2002, joint occurrence of El Niño (Figure 1) and very high negative EQUINOO index (Figure 2) produced a severe drought over India. In that
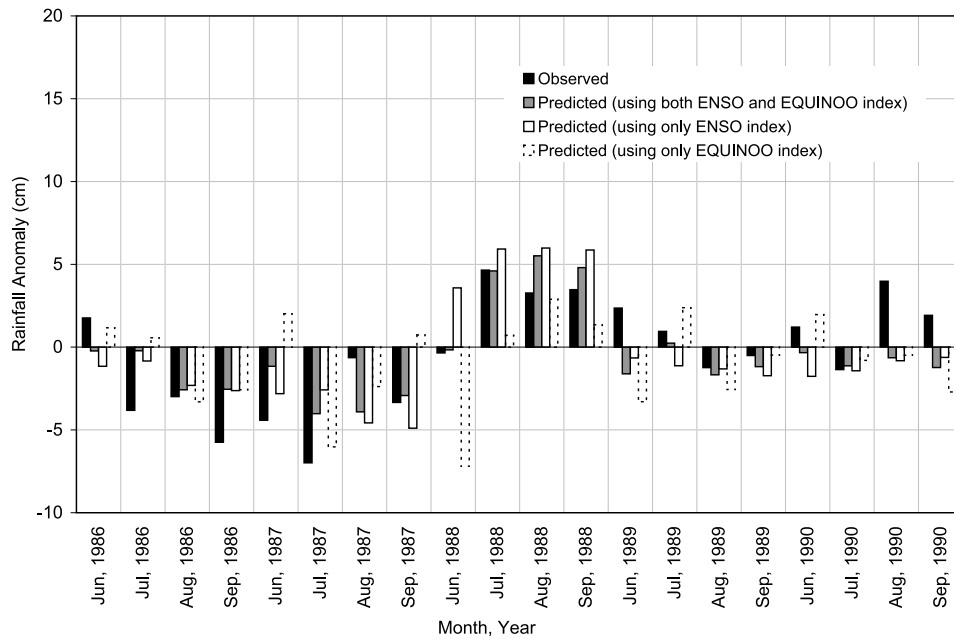
**Figure 7a.** Comparison between observed and predicted monthly rainfall anomaly for monsoon months during 1986–1990 using only ENSO index, only EQUINOO index, and both ENSO and EQUINOO indices.

year, the EQUINOO was lowest in the available record (1958–2003) for monsoon months. As a consequence, lower-than-normal values for all the monsoon months of 2002 were predicted. For the month of July 2002, rainfall was the lowest recorded. This model also predicted very low value for this month but could not predict as low as it was observed (Figure 6d). This may be owing to some other

local physical phenomena which occur at smaller scale, compared to the large-scale atmospheric circulation like ENSO and EQUINOO, for example, sudden change in total water vapor in air column [*Gadgil et al.*, 2002] or sudden change in cloud system over Bay of Bengal [*Srinivasan and Nanjundiah*, 2002]. The reasons behind unanticipated deficit in 2002 are still being investigated, which is not within
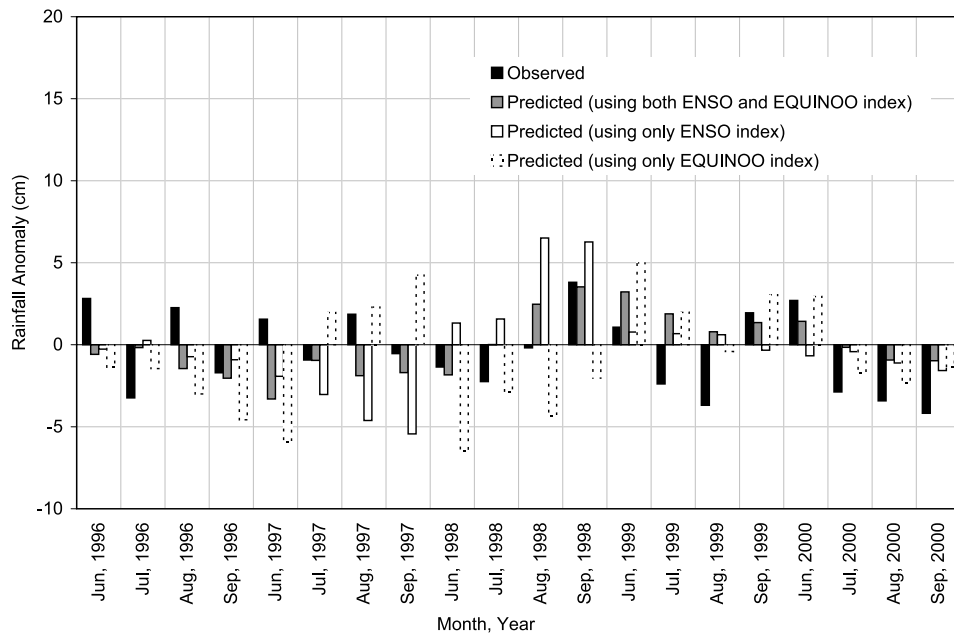


**Figure 7b.** Comparison between observed and predicted monthly rainfall anomaly for monsoon months during 1996–2000 using only ENSO index, only EQUINOO index, and both ENSO and EQUINOO indices.
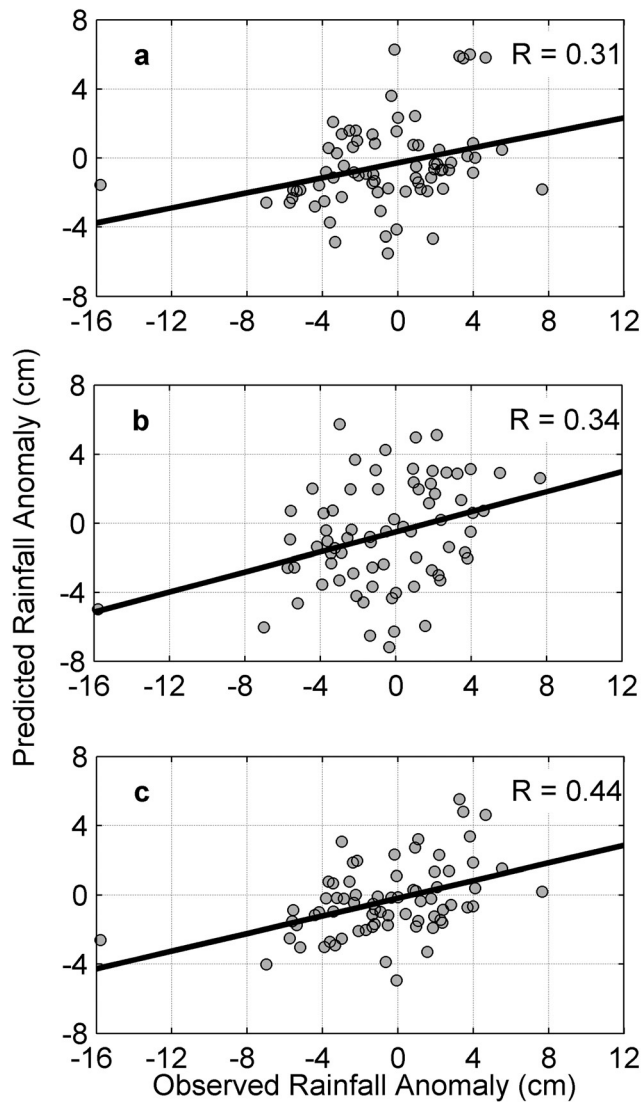
**Figure 8.** Scatterplot between observed and predicted monthly rainfall anomaly for monsoon months during 1986−2003 using (a) only ENSO index, (b) only EQUINOO index, and (c) both ENSO and EQUINOO indices.

the scope of this study. However, model has successfully predicted lower-than-normal rainfall for remaining three monsoon months in 2002. In general, a concurrent effect of ENSO and EQUINOO on the Indian monthly rainfall can be observed from this study.

[36] To further investigate concurrent effect of ENSO and EQUINOO, predictions were made using these indices one at a time. A comparison between observed and predicted rainfall anomaly using the ENSO index, the EQUINOO index and both ENSO and EQUINOO indices is studied for the monsoon months of all the years (1986−2003). However, for the purpose of discussion, such plots are presented in Figures 7a and 7b for the periods 1986−1990 and 1996−2000, respectively. It is observed that prediction performances using only ENSO index and only EQUINOO index are poorer than those obtained by considering both of them concurrently. It is visually noticed that observed and pre-

dicted rainfall anomalies for a few months are out of phase. These out-of-phase behaviors are seen generally in the month of June, which is the starting month of monsoon period. Month of June receives considerably high rainfall as compared to the previous months (i.e., March, April and May). The out-of-phase behavior may be attributed to the sudden jump of the rainfall series in the month of June. However, in general, predictions are observed to be better when a combination of ENSO and EQUINOO indices is used. As it is difficult to appreciate the results by visual inspection, scatterplots for monsoon months for (1) the ENSO index, (2) the EQUINOO index and (3) both the ENSO and EQUINOO indices between observed and predicted monthly rainfall anomaly for the period 1986−2003 are shown in Figure 8. It is observed that correlation coefficient between observed and predicted rainfall anomaly is 0.44 in case of using both ENSO and EQUINOO indices as against 0.31 and 0.34 in case of using only ENSO index and only EQUINOO index, respectively. Thus it can be concluded that the monthly variability of all India rainfall can be explained in a better way by using both the large-scale circulation information from Indian Ocean (EQUINOO index) and that from Pacific Ocean (ENSO index). Another point is that, on the basis of the correlation coefficient, it can be said that addition of EQUINOO is increasing the lead time for ENSO as compared to ENSO only case (1 month for ENSO only case and 2 months for both ENSO along with EQUINOO) but MSE and Log likelihood indicate a lead time of 2 months for ENSO while considering ENSO alone. However, when ENSO is considered along with EQUINOO, better model performance is achieved.

[37] It may be noted here that ENSO and EQUINOO indices are poorly correlated to each other (correlation coefficient = 0.0013) for monsoon months with 2 months lag for ENSO index and 1 month lag for EQUINOO index. These lags are found to be the best leading times, compared to rainfall anomalies, for ENSO and EQUINOO indices as discussed earlier (Tables 3 and 4).

[38] It is necessary to mention here that although the model performance is analyzed on the basis of monthly prediction during four monsoon months only, predictions evolved continuously for all 12 months to maintain the time continuity for updating the parameters at each month. Visual inspection and the statistics obtained above indicate a close association between the predicted and observed monthly rainfall, proving the capability of the model to capture the dynamic relationship between monthly variability of all India rainfall and both the large-scale circulation information of ENSO and EQUINOO.

## 6. Conclusions

[39] In this study, Indian summer monsoon rainfall (ISMR), which corresponds to June, July, August, September, is predicted by incorporating two different large-scale climate circulation indices, ENSO and EQUINOO. Both the circulation indices are used as exogenous input to Bayesian dynamic linear model (BDLM). This model is shown to capture the dynamic relationship between these circulation indices and rainfall phenomenon, which is important to climate change studies. Moreover, predictions of uncertain

future values are made in a distributional form by Bayesian dynamic linear models. Hence forecast intervals can be obtained at any desired confidence level as shown in this study (90% confidence intervals). This is an advantage for many application fields like hydrometeorology, water resources management etc, where decisions need to be taken with some statistical confidence level.

[40] It is observed that the monthly rainfall can be well predicted by using ENSO and EQUINOO indices concurrently. Unusual recent experiences, against the long-recognized negative correlation between ENSO and Indian summer monsoon rainfall, are satisfactorily explained by this approach. This study indicates that both ENSO and EQUINOO have significant influence on monthly Indian summer monsoon rainfall, the relative weightages being 0.61 and 0.39, respectively. However, it is not claimed that these two are the only large-scale circulation indices effecting Indian rainfall, but it can be concluded that instead of using only ENSO information, if ENSO and EQUINOO information are concurrently used, a better prediction can be made.

[41] Apart from the concurrent influence of ENSO and EQUINOO index on monthly rainfall variation, efficacy of the BDLM to capture the time varying dynamic relationship between monthly rainfall anomaly and circulation indices is shown in this study. Such models can be used in any similar application field where it is necessary to deal with time varying dynamic relationship between regressor and target time series.

## References

Ashok, K., Z. Guan, and T. Yamagata (2001), Impact of Indian Ocean dipole on the relationship between the Indian monsoon rainfall and ENSO, *Geophys. Res. Lett.*, 28, 4499–4502.

Berger, J., and D. R. Insua (1998), Recent developments in Bayesian inference with applications in hydrology, in *Statistical and Bayesian Methods in Hydrological Sciences*, edited by E. Parent et al., pp. 43–62, UNESCO Press, Paris.

Berliner, L. M., C. K. Wikle, and N. Cressie (2000), Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling, *J. Clim.*, 13, 3953–3968.

Bernier, J. (1994), Statistical detection of changes in geophysical series, in *Engineering Risk in Natural Resources Management With Special References to Hydrosystems Under Changes of Physical or Climatic Environment, NATO ASI Ser. E, Applied Sciences*, vol. 275, edited by L. Duckstein and E Parent, pp. 343–357, Elsevier, New York.

Besag, J., P. Green, D. Higdon, and K. Mengersen (1995), Bayesian computation and stochastic systems, *Stat. Sci.*, 10, 1–58.

Dracup, J. A., and E. Kahya (1994), The relationship between U.S. streamflow and La Niña events, *Water Resour. Res.*, 30(7), 2133–2141.

Eltahir, E. A. B. (1996), El Niño and the natural variability in the flow of the Nile River, *Water Resour. Res.*, 32(1), 131–137.

Gadgil, S. (2003), The Indian Monsoon and its variability, Annu, *Rev. Earth Planet. Sci.*, 31, 429–467, doi:10.1146/annurev.earth.31.100901.141251.

Gadgil, S., J. Srinivasan, R. S. Nanjundiah, K. Krishna Kumar, A. A. Munot, and K. Rupa Kumar (2002), On forecasting the Indian summer monsoon: The intriguing season of 2002, *Curr. Sci.*, 83(4), 394–403.

Gadgil, S., P. N. Vinayachandran, and P. A. Francis (2003), Droughts of the Indian summer monsoon: Role of clouds over the Indian ocean, *Curr. Sci.*, 85(12), 1713–1719.

Gadgil, S., P. N. Vinayachandran, P. A. Francis, and S. Gadgil (2004), Extremes of the Indian summer monsoon rainfall, ENSO, and equatorial Indian Ocean oscillation, *Geophys. Res. Lett.*, 31, L12213, doi:10.1029/2004GL019733.

Jain, S., and U. Lall (2001), Floods in a changing climate: Does the past represent the future?, *Water Resour. Res.*, 37(12), 3193–3205.

Kahya, E., and J. A. Dracup (1993), U.S. streamflow patterns in relation to the El Niño/Southern Oscillation, *Water Resour. Res.*, 29(8), 2491–2503.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, 2, 437–471.

Kane, R. P. (1998), Extremes of the ENSO phenomenon and Indian summer monsoon rainfall, *Int. J. Climatol.*, 18, 775–791.

Khandekar, M. L., and V. R. Neralla (1984), On the relationship between the sea surface temperatures in the equatorial Pacific and the Indian monsoon rainfall, *Geophys. Res. Lett.*, 11, 1137–1140.

Kiladis, G. N., and H. F. Diaz (1989), Global climatic anomalies associated with extremes in the Southern Oscillation, *J. Clim.*, 2, 1069–1090.

Krishna Kumar, K., B. Rajagopalan, and M. A. Cane (1999), On the weakening relationship between the Indian Monsoon and ENSO, *Science*, 284, 5423, doi:10.1126/science.284.5423.2156.

Krishnaswamy, J., M. Lavine, D. D. Richter, and K. Korfmacher (2000), Dynamic modeling of long-term sedimentation in the Yadkin River basin, *Adv. Water Resour.*, 23, 881–892.

Krishnaswamy, J., P. N. Halpin, D. D. Richter, and K. Korfmacher (2001), Dynamics of sediment discharge in relation to land-use and hydro-climatology in a humid tropical watershed in Costa Rica, *J. Hydrol.*, 353, 91–109.

Li, T., Y. S. Zhang, C. P. Chang, and B. Wang (2001), On the relationship between Indian Ocean sea surface temperature and Asian summer monsoon, *Geophys. Res. Lett.*, 28, 2843–2846.

Mooley, D. A., and D. A. Paolino (1989), The response of the Indian monsoon associated with the change in sea surface temperature over the eastern south equatorial Pacific, *Mausam*, 40, 369–380.

Normand, C. (1953), Monsoon seasonal forecasting, *Q. J. R. Meteorol. Soc.*, 79, 463–473.

Pant, G. B., and B. Parthasarathy (1981), Some aspect of an association between the Southern Oscillation and Indian summer monsoon, *Arch. Meteorol. Geophys. Bioklimatol., Ser. B*, 89, 179–195.

Pole, A., M. West, and J. Harrison (1994), *Applied Bayesian Forecasting and Time Series Analysis*, CRC Press, Boca Raton, Fla.

Rasmusson, E. M., and T. H. Carpenter (1983), The relationship between eastern equatorial Pacific sea surface temperature and rainfall over India and Sri Lanka, *Mon. Weather Rev.*, 111, 517–528.

Ropelewski, C. F., and M. S. Halpert (1987), Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation, *Mon. Weather Rev.*, 115, 1606–1626.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata (1999), A dipole mode in the tropical Indian Ocean, *Nature*, 401, 360–363.

Srinivasan, J., and R. S. Nanjundiah (2002), The evolution of Indian summer monsoon in 1997 and 1983, *Meteorol. Atmos. Phys.*, 79, 243–257.

Walker, G. T. (1923), Correlation in seasonal variations of weather: III. A preliminary study of world weather, *Mem. India Meteorol. Dep.*, 24, 75–131.

Walker, G. T. (1924), Correlation in seasonal variations of weather: IV. A further study of world weather, *Mem. India Meteorol. Dep.*, 24, 275–332.

Webster, P. J., A. M. Moore, J. P. Loschnigg, and R. R. Leben (1999), Coupled oceanic-atmospheric dynamics in the Indian Ocean during 1997–98, *Nature*, 401, 356–360.

West, M., and P. J. Harrison (1997), *Bayesian Forecasting and Dynamic Models*, 2nd ed., Springer, New York.

————————————

D. Nagesh Kumar and R. Maity, Department of Civil Engineering, Indian Institute of Science, Bangalore 560 012, India. (nagesh@civil.iisc.ernet.in)