Monsoonal Rainfall of Sub-Himalayan West Bengal: Empirical Modelling and Forecasting

Pijush Basak
Dept. of Mathematics, Narula Institute of Technology
E-mail: pijushbasak@rediffmail.com

ABSTRACT

The South West Monsoonal rainfall data of the subdivisions of Sub-Himalayan West Bengal is modelled as a non-linear time series. As observed through set of rainfall data of the regions, more than 60% of the inter-annual variability of rainfall is accounted for by the proposed non-linear model. The proposed model is capable of forecasting South West Monsoon rainfall one year in advance in the respective regions. The model indicates the high/low rainfall with the help of antecedent data. For the year 2001, the predicted South West Monsoon rainfall in Sub-Himalayan West Bengal is 168.21±21.36 (cm).

Keywords*: South West Monsoon rainfall, Lognormal distribution, Skill of forecast, modelling, Climatic normal.*

1. Introduction

The occurrence of South West Monsoon (SWM) rainfall or the summer monsoon rainfall over the meteorological subdivisions of Sub-Himalayan West Bengal (SHWB: No.5) is an important phenomenon; also its impact on the agriculture, economy and society is well known. Efforts are made to quantify the variability and forecast of monsoonal phenomenon at various temporal and spatial scales till long time. A modelling exercise effort is ought to be focussed on understanding of variability of historical data. A number of literatures are available on analysis of variability of SWM rainfall data. One may refer to the architectural works of Mooley and Parthasarathy; 1984; Gregory, 1989; Thapliyal, 1990; Iyenger and Basak, 1994; Iyenger and Raghukant, 2003 and Basak, 2014. However, a discussion in general in case of All India rainfall is well documented in the works of Gadgil et al., 2002. A review of literature extensively and exclusively on empirical modelling and forecasting has been presented by Sahai et al. 2000 and Hastenrath and Greischar, 1993.

The basic characteristics of SWM rainfall data is non-Gaussianness on several temporal and spatial scales. In fact, non-Gaussianness in weekly, monthly and seasonal scales persists even though those data can be treated as sum of large number of random variables. As explained in the earlier works (Iyenger and Basask, 1994; Basak, 2014) for the non-Gaussian property, the statistical significance test, namely, auto-correlation, power spectral density show poor signal that may be helpful in establishing linear time series model. The linear time scale models based on past rainfall reflect the behaviour near the mean value reasonably well; but fail to forecast at the extreme values such as flood or draught relevant to the community at large. The alternative approach such as Principal Component Analysis (PCA) with versatile variation in space and time has been attempted by few researchers (Iyenger and Basak, 1994) for All India and particular regions. The Empirical Mode Function (IMF) approach has been utilized for All India level and few subdivisions (Iyenger and Raghukant, 2003).for rainfall based on past rainfall reasonably well. However, little attention is given to the present approach to the sub-division of SHWB (No. 6) and few other sub-divisions. The present paper stresses on prediction of SWM rainfall on the basis of empirical modelling developed.

Kedem and Chiu (1987), however, proposes that in a small time scale, rain rate has to be Lognormal random variable and indicate that lognormal distribution is a natural outcome of the law of the proportionate effect applied to rain rate, namely,

$$R_{j+1} - R_j = \varepsilon_j R_j \qquad (1)$$

where ε_j's are independent identically distributed random variable and are independent of Rj's.

The present work establishes the logic for the SWM rainfall (June-September) of the subdivision SHWB. It is explained that the equation (1) can be systematically extended to account for year-to-year and long term relationship known to persist in monsoonal rainfall. The applicability of the model is demonstrated consistently on SWM rainfall of the subdivision. The proposed model is understood to behave like a 'useful tool' for statistical forecasting of monsoonal rainfall. The detailed approach is presented in the consequent sections. A forecast of the recent year is also presented.

2. Data

The SWM rainfall data of the subdivisions such as SHWB for the year 1871-2001 is extracted from IMD data base available at IITM-IMD site (http://www.tropmet.res.in/) with the details of data assembly are available in the concerned website.

3. Modelling

The initial statistical properties of the sub divisional SWM rainfall are presented in Table 1. The mean μ_R is the long term time average of the data and may be recognised as 'climatic normal'. The standard deviation δ_R reflects year-to-year variability of the SWM rainfall with respect to long term average.

The autocorrelation and other salient features helpful for modelling are extracted. These statistics would have been computed from the available single sample under the assumption of ergodic process. The inter-annual correlations are studied and tested as if those are Gaussian data. For SWM rainfall series of the subdivisions with moderate or strong non-Gaussianness such steps may not be valid except for long-term average as useful reference quantities (Iyenger and Basak, 1994). With these in view, we do not refer to the concepts connected with Gaussian processes such as standard variate; but treat data series as if it is generalised lognormal data.

The equation (1) is now generalised and converted to simple Lognormal model equation, namely,

$$R_{j+1} / R_j = f(R_j) + \varepsilon_j \qquad (2)$$

where $f(R_j)$ = an appropriate function of R_j.

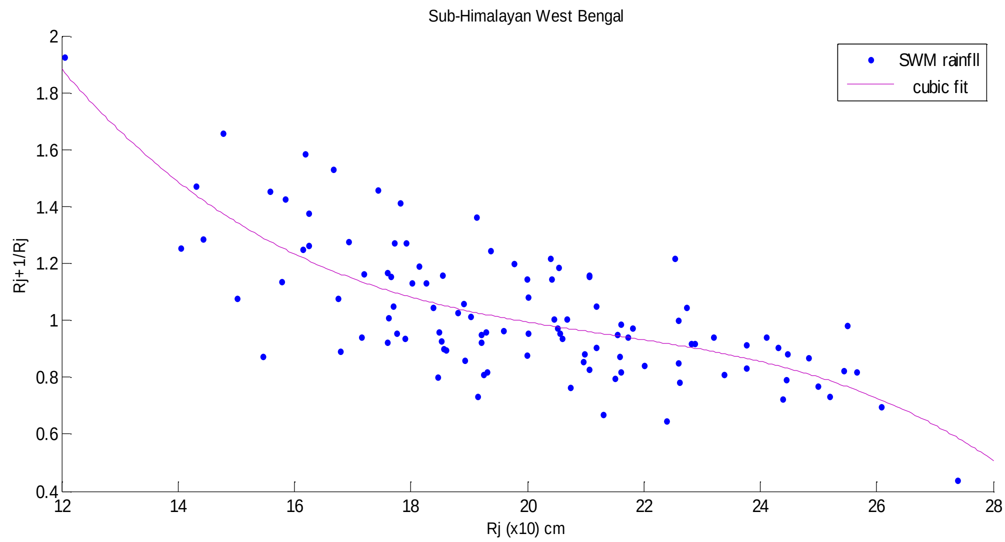

Figure 1: (R_{j+1}/R_j) vs. R_{j+1} plot of Sub-Himalayan West Bengal SWM rainfall

It may be observed that equation (2) is converted to equation (1) when $f(R_j) = 1$. It indicates that given j^{th} year rainfall R_j, the annual change is proportional to an unknown function of R_j itself.

In Fig. 1, the relation between the (R_{j+1}/R_j) and R_j, for sub-divisional SWM rainfalls of SHWB is presented for the period 1871-1990; the figures indicate that there is a clear cut discernable trend that can be expressed as a cubic polynomial of the form

$$f(R) = aR^3 + bR^2 + cR + d \qquad (3)$$

The cubic equation is selected in comparison with quadratic and linear equations due to least variance for the error ε obtained as

$$\sigma_\varepsilon^2 = (1/(N-1)).\Sigma\varepsilon_j^2 \qquad (4)$$

where ε_j is a random time series of unknown structure that internally directs (traces) the system.

The equation (2) is now a non-linear difference equation for the data in the form

$$R_{j+1} = R_j.f(R_j) + \varepsilon_j.R_j = R_j(aR_j^3 + bR_j^2 + cR_j + d) + \varepsilon_j.R_j \qquad (5)$$

If ε_j is taken to be independent of R_j, it follows that the conditional expectation of R_{j+1} given R_j would be

$$<R_{j+1}> = R_j..(aR_j^3+bR_j^2+cR_j+d) \qquad (6)$$

The above expression defines the natural point predictor for the rainfall in year (j+1) if only the previous year value is known.

The error in the predictor is perhaps the conditional standard deviation of R_{j+1}, namely,

$$\sigma_{Rj+1} = \sigma_\varepsilon.R_j \qquad (7)$$

For SWM rainfall of the subdivisions analysed in the paper, only one real root is possible for the fixed point equation f(R) = 1 and ultimately equation (6), on iteration would stabilise at that fixed point. The parameters of the proposed cubic polynomial f(R) and the fixed point (R_c) of equation (6) are presented for the SWM data series of the subdivisions in Table 2.

Table 1. SWM rainfall data (1871-2001)

Region	Area (Square Km)	μ_R	σ_R	Skewness	Kurtosis
SHWB	21625	199.1141	20.0709	0.1255	3.1025

SHWB: Sub Himalayan West Bengal

It is understood that equation (6) or in general equation (1) extracts the climatic or long-term average of the data series reasonably well.

Thereafter, the fixed point R_c of Table 2 is compared with the mean value μ_R of Table 1.

Table 2. Parameters of f(R) and R_c

Region	a	b	c	d	σ_ε	R_c
SHWB	-0.0010	0.0655	-1.4225	11.4302	0.1858	199.1141

SHWB: Sub Himalayan West Bengal

However, the variability band in any year (j+1) depends on the previous rainfall values R_j given by equation (7). This may be considered large enough although σ_ε is too small. This, however, implies that equation (6) in its current form would not yield good one-year ahead forecast due to plausible attraction towards long-term mean value (fixed point), namely R_c. It indicates that the natural variability is not reflected by equation (3) containing $f(R_j)$. Further, the form of $f(R_j)$ has to be rectified and it would be interesting to measure the forecast skill of equation (6) so as to see an improvement in forecast.

4. Improvement in Forecast Skill

A relation in the form of equation (6) is utilized for forecasting the SWM rainfall in the next year i.e. R_{j+1}, whenever the rainfall in the previous year R_j is known. The skill in forecast can be measured with reference to climatic mean μ_R and variance σ_R^2. If there had been no inter-annual relationship in the SWM rainfall, then R_j may be treated as independent samples of random variable. In such a case, the only possible prediction is a year is its mean value μ_R and the prediction error may be measured as $(R_{j+1} - \mu_R)^2$ which is obviously the climatic variance σ_R^2. Consequently, any improvement in forecast has to be acquired by reduction of variance with respect to σ_R^2. With this in view, the error between the observed rainfall and the value predicted by equation (6) is compared for each year with j=2 to j=120 (1872-1990). The percentage reduction in variance with respect to climatic normal μ_R are computed and presented in Table 3.

Table 3. Percentage reduction in climatic variance

Region	Equation (6)	Equation (8)
SHWB	3.1163	3.2049

SHWB: Sub Himalayan West Bengal

Table 4. Correlation Coefficient (R_{j+1}/ R_j) and SWM rainfall at different lags R_{j-1}, R_{j-2},...,R_{j-19} (* Significant)

Lag-RF	SHWB
R_j	0.7096*
R_{j-1}	-0.0336
R_{j-2}	0.0823
R_{j-3}	0.0978
R_{j-4}	-0.1303
R_{j-5}	-0.0379
R_{j-6}	0.1559*
R_{j-7}	-0.1002
R_{j-8}	-0.0984
R_{j-9}	0.0244
R_{j-10}	0.0922
R_{j-11}	-0.0326
R_{j-12}	-0.0680
R_{j-13}	0.0211
R_{j-14}	0.0668
R_{j-15}	0.0171
R_{j-16}	-0.0564
R_{j-17}	-0.0684
R_{j-18}	0.0652
R_{j-19}	0.0361

Table 5. Coefficients of equation (8)

Coefficients	Magnitude
a	0.0005
b	-0.0256
c	0.3606
d	1.3039e-3
g_1	-
g_2	-
g_3	-
g_4	-
g_5	-
g_6	2.5144e-3
g_7	-
g_8	-
g_9	-
g_{10}	-
g_{11}	-
g_{12}	-
g_{13}	-
g_{14}	-
g_{15}	-
g_{16}	-
g_{17}	-
g_{18}	-
g_{19}	-
g_{20}	2.6201e-3
σ_δ	0.1514

In fact, a small reduction in variance as observed in Table 3 indicates that equation (6) incorporating only consecutive reduction yearly connection, namely R_{j+1} and R_j has to be improved by incorporating the longer inter-annual relations. This may be executed by finding correlation between (R_{j+1}/R_j) and lagged yearly values. Previously, it is reported by investigators (Parthasarthy et al., 1994; Iyenger and Basak' 1994; Basak' 2014) that SWM rainfall posses significant correlations at several lags at a number of regions. For the present paper, the inter-relations are investigated for the present sub-divisional SWM rainfalls. As the correlations are small numbers and series concerned are non-negative and non-Gaussian, the statistical significance level for the Gaussian series may not be valid and would be less than that of Gaussian series (Johnson and Kotz, 1972). Also, due to positivity property, the random variables (different series) may be jointly exponentially distributed and thus the linear correlations for significance of those variables would be considerably less than linear correlation of Gaussian series (Johnson and Kotz, 1972). The small correlations which are rejected for Gaussian series at 5% level (±0.20 for a data series of length around 120) may still be valid for non-Gaussian series relevant to our study. Thus, for non-Gaussian series as per our investigation, for a sample size of 120, the significant value at 5% level is accepted as ±0.13.

The above discussion evolves a clue for the improvement of the model (6) by incorporating up to twentieth lagged terms, namely, R_{j-1} , R_{j-2} , R_{j-3} ,..., , R_{j-18}, R_{j-19}, R_{j-20}. As the length of data is limited to 120 as discussed, the correlations up to lag 20 is considered in our study.

In Table 4, significant lagged correlations between (R_{j+1}/R_j) and rainfall is significant for 6th and 20th lags for SHWB (Table 4) suggesting a model of the form

$$R_{j+1} = R_j(aR_j^3 + bR_j^2 + cR_j + d + g_6R_{j-6} + g_{20}R_{j-20}) + \delta_j.R_j \quad (8)$$

It may be noted that all the data series under consideration posses significant lag-correlation in the range of 14-20 years. Upon minimization of the mean square value of the error δ_j for the respective subdivisions, the coefficients of the equations (8) are computed and are presented in Table 5.

The last row in Table 5, the standard deviation σ_δ is presented for the SWM rainfall of SHWB. It may be interestingly observed that the standard deviation σ_δ is less than the previous error value σ_ε's for all the corresponding case (Tables 2 and 5). Naturally, the significance of adding lagged terms in the model (6) is established. The SWM series of subdivision of SHWB model consists of two lagged correlated terms.

Equation (8) provides a non-linear map of higher dimension. Initial conditions of 20 past values are required for solving the equation (8) and then iterating sufficient number of times, one can find the steady state (stable fixed points) of the model. This process possesses considerable variation controlled by the cubic polynomial f(R) in equation (8). The skill of this new model in forecasting is verified by finding the reduction in variances as discussed earlier. The reduction in variance in a year-to-year forecast achieved by equation (8) along with equation (6) for SWM rainfalls of SHWB is presented in Table 3. It is observed that in all the cases, the percent reductions in variances for equations (8) are always higher than equation (6).

5. Test Forecasting

The models (8) proposed by SWM rainfall of the subdivisions has a clear cut deterministic part and a random part. The deterministic part $<R_{j+1}>$ is a point predictor for the rainfall in the year (j+1) conditioned on all the previous values. The statistical variability band about this predictor leads to $\pm\, \sigma_\delta.R_j$. Consequently, as a forecast we understand that the rainfall in year (j+1) will be with 67% probability in the interval $[<R_{j+1}> -\sigma_\delta.R_j,\ <R_{j+1}> +\sigma_\delta.R_j]$. Obviously, it is expected that $\sigma_\delta.R_j$ is less than σ_R. It is pointed out that the rainfall data is available over the period 1871-2001. In the prediction exercise, for the first prediction year, 1991, the previous model has been directly used and also for the subsequent 10 years (1992-2001), the actual rainfall of the previous year has been utilized for a possible prediction. It is observed that most of the predicted values are well within the range of prediction. In Table 6, the predictions of the SWM rainfall are presented.

Totally 11 predictions have been made. The expected value of the prediction is reported against the 1-sigma band (67% probability). The observed numbers of correct frequencies are 8, Regarding statistical significance, the Null Hypothesis (H_0) that correct prediction would be purely due to chance, is tested with Chi-square test. If the correct prediction is purely due to chance, then expected correct prediction would be 10.5. However, the number of correct prediction is 8 which are about 73% of the total predictions. For the concerned frequencies, the Chi-square is 7.36 compared to tabulated Chi-square of 3.84 at 5% level of significance. Consequently, H_0 gets rejected and the modelling capability of the model is verified. It is also observed surprisingly that the standard deviation of the prediction in any year is less than the climatic deviation about the long term mean value.

6. Discussion

The model developed in the paper carries importance in the sense that SWM rainfalls of the subdivision SHWB is modelled as generalised Lognormal random variables. It is of course, the generalization of the basic law of proportion leading to basic Lognormal distribution. It is shown that long-term climatic mean is essentially controlled by one-step annual connection. This model accounts for nearly 60% of the inter-annual variability of the subdivision. The inter-relation between the present rainfall value and past values are utilized to shape the model for the prediction purpose.

The impact of the previous rainfall values on the future values of rainfall has been introduced in the model. Imperatively, it is stressed that the present and future rainfall values are the cumulative effect of the present and past values; although the basic law of proportion is the main consideration the model.

The SWM rainfall forecasts of the subdivision SHWB are presented in Table 6 as an expected value along with its standard deviation. The actual value is expected to be within 1-σ prediction interval with high probabilities of 67%. As explained in the Table, all the actual values of SWM rainfall for the prediction exercise period (1991-2001) are within the 1-σ interval. The prediction variation is less than the climatic variation as observed in the present case.

Consequently, it would be relevant to mention the works of Sahai et al., 2000, Iyenger and Raghukant, 2003 proposed empirical model for All India and sub-divisional SWM rainfall. The works of Sahai et al. (2000) is an ANN model for the All India SWM rainfall data on the basis of past rainfall data and the model explain about 60% of climatic variance of All India SWM series. No results are found for the other regional series. In another work, regression model depending on SST data explaining about 72% of variance is introduced by Sahai et al. (2003), Iyenger and Raghukant (2003) in their work suggested a regression model for All India and different zonal SWM rainfall series such as West Central India. However, in our study, the meteorological subdivisions consisting of SHWB with the effecting the back correlation is undertaken. It is worth to mention that model of Sahai et al. (2003)' dependant on SST data; whereas, the works of Iyenger and Raghukant (2003) is dependent past rainfall data.

Table 6. Test forecasting with independent data

	SHWB	
Year	Actual (mm)	Prediction (mm)
1991	124.51	154.42±11.90
1992	132.26	154.94±19.13
1993	154.68	152.98±18.77
1994	197.20	154.18±19.04
1995	122.18	143.05±17.67
1996	157.05	151.19±18.67
1997	148.82	153.38±18.00
1998	130.16	153.38±18.00
1999	165.56	154.79±19.12
2000	156.24	153.50±18.96
2001	161.25	168.21±21.36

SHWB= Sub Himalayan West Bengal

7. Summary and Conclusions

Keeping in view of basic law of proportion, a new nonlinear dynamic model for SWM rainfall of the subdivisions of North Eastern region namely SHWB has been proposed in this paper. Along with the mechanistic dynamic model, the interplay of several atmospheric variables are carefully considered; then attempts are made to open up the structure of the data series with possible expansion. The corresponding algorithm for doing this is developed. The first stage takes care of the climatic mean behaviour; it is followed by interplay of previous year connections in the model. It is seen that the new model successfully explains an overall nearly 60% of the inter-annual variability. The year-to-year forecasting ability of the model verified to be effective on an independent sample of 11 years. For the year 2001, the forecast for the SHWB is 168.21 ± 21.36 (cm) with a probability of 67%. It would be interesting to see how the present random error part can be decomposed into the intra seasonal variability (Ajoy Mohan , Hartmann and Michelson) in the impending regions. It is hoped that such effort would further improve the forecast ability of the present model.

References

Ajay Mohan R. S. and Goswami B. N., 2000, 'A common spatial mode for intra-seasonal and inter-seasonal variation and predictability of the Indian summer monsoon', *Curr. Sc*., 79(8), 1106-1111.

Basak P., 2014, 'Variability of south west monsoon rainfall in West Bengal: An application of principal component analysis', *Mausam,* 65(4), 559-568.

Gadgil S., Srinivasan J. and Nanjundiah R, S,, 2002, 'On forecasting the Indian summer monsoon: the intriguing season of 2000', *Curr. Sc*., 83, 394-403.

Gregory S,. 1989. .Macro-regional definition and characteristics of Indian summer monsoon rainfall 1871-1975', *Int. J. Climatol*., 9, 465-483.

Hartmann D. L. and Michelson M. L., 1989, 'Intraseasonal periodicities in Indian rainfall', *J. Atmos. Sc.,* 46**,** 2838-2862.

Hastenrath S. and Greischar L., 1993, 'Changing predictability of Indian monsoon rainfall anomalies?', *Proc. Ind. Acad. Sci. (Earth Planet Sci.),* 102**,** 35-47.

Iyenger R. N. and Basak P., 1994, 'Regionalization of Indian monsoon rainfall and long-term variability signals', *Int. J. Climatol*., 14, 1095-1114.

Iyenger R. N. and Raghukant S. T. G., 2003, 'Empirical modelling and forecasting of Indian monsoon rainfall', *Curr. Sc*., 8, 1189-1201.

Johnson N. L. and Kotz S., 1972, *'Distribution in Statistics: Continuous Multivariate Distribution'*, New York, Wiley.

Evaluation and development of new model', *Mausam,* 41, 339-346.

Kedem B. and Chiu L. S., 1987, 'On the log normality of rain rate', *Proc. Natl. Acad. Sci*., USA, 84, 901-905.

Mooley D. A. and Parthasarathy B., 1984, 'Fluctuations in all-India summer monsoon rainfall during 1871-1978'; *Climatic Change,* 6, 287-301.

Parthasarathy B., Munot A. A. and Kothawale D. R., 1994, 'All-India Monthly and Seasonal Rainfall Series: 1871-1993', Theo. Appl. Climatol., 49(4), 217–224.

Sahai A. K., Soman M. K. and Satyen V., 2000, 'All India summer monsoon rainfall prediction using an artificial neural network', *Clim. Dynam.,* 16, 291-302.

Sahai A. K., Grimm A. K., Satyen V. and Pant G. B., 2003, 'Long-lead prediction of Indian summer monsoon rainfall from SST evolution'; *Climate Dynamics,* 20, 855-863.

Thapliyal V., 1990, 'Large scale prediction of summer monsoon rainfall over India: