

24. Frost, B. W. and Franzen, N. C., *Mar. Ecol. Prog. Ser.*, 1992, **83**, 291–303.
25. Garside, C., *Mar. Chem.*, 1982, **11**, 159–167.
26. Brzezinski, M. A., *Mar. Chem.*, 1987, **20**, 277–288.
27. Harrison, W. G., Harris, L. R. and Irwin, B. D., *Limnol. Oceanogr.*, 1996, **41**, 16–32.
28. Lomas, M. W. and Glibert, P. M., *Mar. Biol.*, 1999, **133**, 541–551.

ACKNOWLEDGEMENTS. This work was carried out as a part of ocean modelling activities at C-MMACS sponsored by the Department of Ocean Development (DOD) under several projects. We thank DOD for financial support. A preliminary version of this work was presented

at the JGOFS Open Science Conference ‘Ocean Biogeochemistry: A New Paradigm’ held in Bergen during 13–17 April 2000. K.S.Y. thanks the conference organizers for financial support. We thank Prof. J. J. McCarthy and co-authors for supplying data on nitrogen uptake kinetics in numerical form. K.S.Y. thanks Dr Peter Burkil and Dr T. R. Anderson and their colleagues in Plymouth Marine Laboratory and Southampton Oceanography Centre, respectively for interesting discussions. We also thank Prof. Lakshmi Kantha, Dr Angelica Peña and Dr P. S. Swathi for useful suggestions.

Received 4 February 2003; revised accepted 18 August 2003

Empirical modelling and forecasting of Indian monsoon rainfall

R. N. Iyengar^{*,†,‡} and S. T. G. Raghu Kanth[‡]

[†]Centre for Atmospheric and Oceanic Sciences, [‡]Department of Civil Engineering, Indian Institute of Science, Bangalore 560 012, India

Indian monsoon rainfall data is modelled as a nonlinear time series. It is demonstrated that the proposed model accounts for about 50% of the inter-annual variability of the rainfall, as observed in eight sets of data representing All India and regional rainfall values. The model is capable of statistically forecasting seasonal rainfall value one year in advance. The model predicts the drought of 2002, with the help of only antecedent data. For the year 2003, the predicted All India rainfall value is 82.65 ± 4.88 cm.

THE importance of the summer monsoon rainfall, also popularly called south-west monsoon (SWM) rainfall, to Indian society and economy is well known. Efforts to forecast the rainfall at different spatial and temporal scales have been in vogue for nearly a century. These are generally based on developing suitable mathematical models which in turn may be broadly classified as empirical or dynamical. The present work is concerned with empirical models only. Any modelling effort will have to be based on an understanding of the variability of past data. Thus, considerable literature is available on analysis of variability of SWM rainfall data. The works of Mooley and Parthasarathy¹, Gregory², Hastenrath and Greischar³, Rupa Kumar *et al.*⁴, Thapliyal⁵, Iyengar and Basak⁶ may be mentioned in this connection. A general discussion on forecasting of monsoon rainfall is available in the paper by Gadgil *et al.*⁷. A review of the literature on empirical modelling and forecasting has been recently presented by

Sahai *et al.*^{8,9} and hence will not be repeated here. A basic feature of rainfall data is its non-gaussianity on several temporal and spatial scales. Weekly, monthly and seasonal data at station levels or at regional scales still exhibit strong non-gaussianity even though they can be treated as sums of large number of random variables. Thus, linear time series models based on past rainfall which capture the behaviour near the mean value fairly well, fail to forecast extreme values, such as floods and droughts which are the ones of main concern to the community at large. This property of rainfall data has been recognized by Kedem and Chiu¹⁰ who argue that at a small time scale rain rate has to be a lognormal random variable. These authors highlight that the lognormal distribution is a natural outcome of the law of proportionate effect,

$$R_{j+1} - R_j = \epsilon_j R_j, \quad (1)$$

where ϵ_j 's are independent identically distributed random variables and are also independent of R_j 's. They demonstrate further that this model fits well the hourly rainfall data obtained from the Global Atlantic Tropical Experiment (GATE). Now, since rainfall at other time scales such as weekly and monthly also exhibit strong non-gaussianity, it is natural to ask how eq. (1) can be generalized to model such data. The present paper addresses this question with respect to Indian monsoon seasonal (June–September) data. It is shown that eq. (1) can be systematically extended to account for year-to-year and long term relationships known to exist in monsoon rainfall data. The new model is shown to account for nearly

*For correspondence. (e-mail: rni@civil.iisc.ernet.in)

50% of the climatic variation of the All India rainfall on a year-to-year basis. The applicability of the model is demonstrated consistently on eight sets of data drawn from different spatial regimes. Thus, the proposed model is claimed to be a robust tool for statistical forecasting of monsoon rainfall. Detailed results are presented to highlight the innovations involved in the present approach. A forecast for the current year monsoon season is also presented.

Database

Eight sets of SWM data as reported by IMD are selected for the present study. All these data are available at the website of IITM-Pune (<http://www.tropmet.res.in/>). Some important reference statistics of the data are shown in Table 1. The data are regional rainfall values obtained by area weighted averaging of subdivisional values. Further details of the subdivisions and data assembly are available at the above website. The mean μ_R is the long term time average of the data. It is also called the *climatic normal*. The standard deviation σ_R reflects the year to year variability of the rainfall with respect to its long term average.

Modelling

When a statistical model is proposed for a time series, one generally incorporates the mean, autocorrelation and other salient features into the model. These statistics in turn would have been computed from the available single sample under the tacit assumption that the process is ergodic. Furthermore, inter-annual correlations are found and tested for their significance, as though for Gaussian data. For the SWM rainfall time series, such steps are questionable, except for treating long term time averages as useful reference quantities. With the above points in view, we avoid invoking concepts connected with Gaussian processes such as anomalies and standard variates, but treat the data series as if it is a generalized lognormal

data. Generalization of eq. (1), which is the simple log-normal model equation, is taken in the form

$$R_{j+1}/R_j = f(R_j) + \epsilon_j. \quad (2)$$

It is easily observed that when $f(R_j)=1$, the above reduces to eq. (1). This equation postulates, given the j th year rainfall R_j , the annual change is proportional to an unknown function of R_j itself. This is verified in Figure 1a, b where the relation between (R_{j+1}/R_j) and R_j is shown for all the eight data sets for the period 1871–1990. It is seen that in all the cases, there is a discernible trend which can be expressed as a cubic polynomial of the form

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

This cubic equation is selected, in comparison with quadratic and linear relations, as the one giving the least variance for the error ϵ found as

$$\sigma_\epsilon^2 = (1/(N-1)) \sum \epsilon_j^2. \quad (4)$$

Eq. (2) is a nonlinear difference equation for the data in the form

$$R_{j+1} = R_j f(R_j) + \epsilon_j R_j. \quad (5)$$

Here ϵ_j is a random time series of unknown structure, that internally drives the system. If ϵ_j is taken to be independent of R_j , it follows that the conditional expectation of R_{j+1} given R_j will be

$$\langle R_{j+1} \rangle = R_j (aR_j^3 + bR_j^2 + cR_j + d). \quad (6)$$

This expression is the natural point predictor for the rainfall in year $(j+1)$ if only the previous year value is known. The error in this predictor is characterized by the conditional standard deviation of R_{j+1} ,

Table 1. SWM rainfall data (1871–2001)

Region	Area (km ²)	μ_R (cm)	σ_R (cm)	Skewness	Kurtosis
All India	2,880,000	85.1180	8.2743	−0.5244	2.9845
Homogeneous	1,596,970	75.4789	11.6785	−0.3701	3.0654
Core monsoon	776,942	85.5215	14.4752	−0.4374	3.0625
WCIND	962,698	92.8718	12.4545	−0.3527	2.9403
CNEIN	573,006	100.1418	11.1060	−0.1855	4.3723
NEIND	267,444	141.8781	12.7023	0.1802	3.0401
NWIND	634,272	49.0790	13.0962	−0.2819	2.8681
Peninsular	442,632	66.2295	9.7911	0.2878	3.0369

WCIND, West Central India; CNEIN, Central North East India; NEIND, North East India; NWIND, North West India.

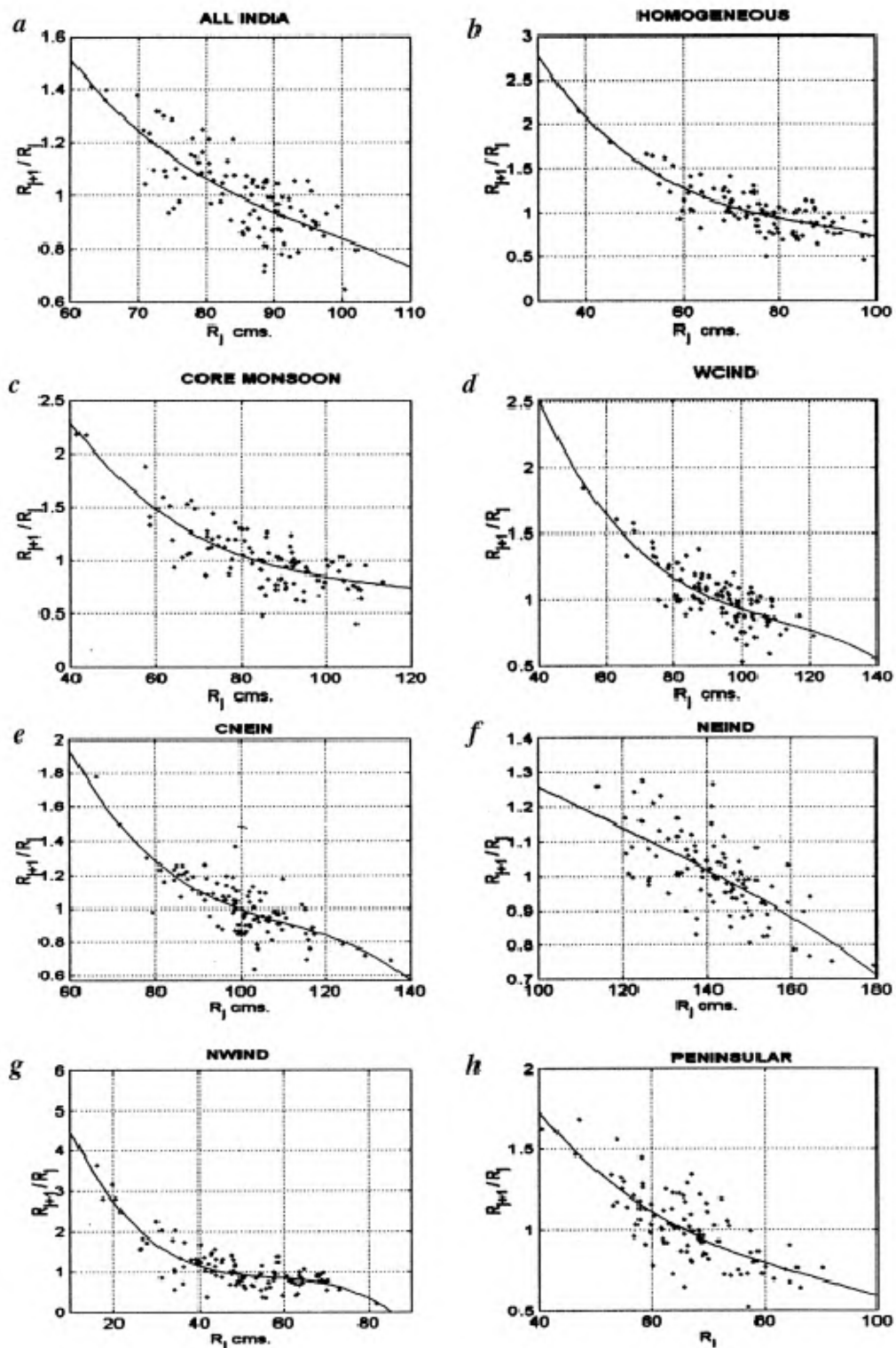


Figure 1. Rate of change of rainfall vs R_j .

$$\sigma_{R_{j+1}} = \sigma_\epsilon R_j. \quad (7)$$

Eq. (6) is a nonlinear map, which on iteration will stabilize at its fixed point. For the SWM data analysed here,

only one real root is possible for the fixed point equation $f(R)=1$. In Table 2, the parameters of the cubic polynomial $f(R)$ and the fixed points of eq. (6) are presented for the eight area rainfall data sets.

Table 2. Parameters of $f(R)$ and R_c

Region	a	b	c	d	α	R_c
All India	-9.55e-6	0.0026	-0.2547	9.3804	0.0996	83.59
Homogeneous	-8.31e-6	0.0021	-0.1924	6.9498	0.1586	73.68
Core monsoon	-4.24e-6	0.0013	-0.1449	6.3353	0.1770	82.53
WCIND	-2.81e-6	0.0010	-0.1180	5.8668	0.1352	92.18
CNEIN	-1.06e-6	0.0004	-0.0591	3.9225	0.1157	98.42
NEIND	-1.61e-7	0.0007	-0.1098	6.9246	0.0859	140.83
NWIND	-3.33e-5	0.0057	-0.3362	7.5651	0.3013	44.92
Peninsular	-2.25e-6	0.0007	-0.0855	4.1087	0.1504	64.91

Table 3. Percentage reduction in climatic variance

Region	Eq. (6)	Eq. (8)	Eq. (10)
All India	2.39	16.943	49.50
Homogeneous	2.46	14.08	57.05
Core monsoon	1.99	8.44	48.28
WCIND	1.27	18.50	65.01
CNEIN	0.46	2.23	50.56
NEIND	1.64	9.81	54.55
NWIND	5.76	7.87	35.79
Peninsular	3.37	16.67	48.78

It is seen that eq. (6) which is a generalization of eq. (1) captures the so-called climatic or long term average of the rainfall data series fairly well. This follows by comparing the fixed points R_c of Table 2 with the mean values μ_R of Table 1. However, the band of variability in any year ($j+1$) depends on the previous rainfall values R_j as given by eq. (7). This can be quite large depending on R_j , even if α were to be small. This indicates that eq. (6) in its current form will not be able to yield good one-year-ahead forecasts, simply because it is attracted towards its long term mean value, namely R_c . While the mean value is simulated, the natural variability is not reflected by the simple function $f(R_j)$ of eq. (3). Nevertheless, it would be interesting to measure the forecast skill of eq. (6) so as to see how this can be improved.

Skill of forecast

Any relation of the form eq. (6) can be used for forecasting the next year value R_{j+1} based only on the present value R_j . The skill in forecast can be measured with respect to the climatic mean and variance. Had there been absolutely no interannual relationship in the rainfall data, we could treat R_j as independent samples of a random variable. In such an extreme case, the mean value μ_R is the only possible prediction in any year and a measure of the prediction error would be the average of $(R_{j+1} - \mu_R)^2$ which is easily seen to be the climatic variance σ_R^2 . Thus, any annual forecast model has to be measured against the reduction of variance it achieves with respect to σ_R^2 . With this in view, for all the eight cases, the error

between the observed rainfall and the value predicted by eq. (6) is found for each year ($j=2$ to $j=120$). The percentage reduction in variance with respect to the climatic normal μ_R is presented in Table 3.

The small reduction in variance, which was expected, goes to indicate that eq. (6), incorporating only consecutive yearly connections, has to be improved to include longer inter-annual relations. This may be done easily by finding correlations between (R_{j+1}/R_j) and lagged yearly values. In Table 4 such correlations between (R_{j+1}/R_j) and $R_{j-1}, R_{j-2}, \dots, R_{j-19}$ are shown. The first row in this table is only a reflection of Figure 1 and eq. (6). It has been previously found by several investigators that SWM rainfall has significant connections at several lags¹¹. These are well reflected in Table 4. As far as statistical significance of the numbers are concerned, we face a difficulty, since the variables considered are non-negative and non-Gaussian. It is known that for two random variables that are jointly exponentially distributed, due to the positivity property, the linear correlation coefficient may be constrained to be much less than unity¹². Thus, Small correlations, which would get rejected for Gaussian random variables, may still be significant for non-Gaussian variables. Keeping in view that for two Gaussian samples with a sample size of 100, the minimum significance level for correlation is around ± 0.2 , here $|\rho| \geq 0.13$ is taken to be significant. In any case the correlations of Table 4 show that the rate of change of rainfall has linear dependence on several past year values. This provides a clue to improve eq. (6) by introducing terms corresponding to $R_{j-1}, R_{j-3}, \dots, R_{j-14}, \dots, R_{j-19}$, as the case may be. Since the length of the available data is limited, lags beyond 19 are not considered here. This analysis indicates that for All India rainfall data, a natural improvement to the climatic mean model of eqs (3) and (6) will be,

$$R_{j+1} = R_j [aR_j^3 + bR_j^2 + cR_j + d + g_1R_{j-1} + g_3R_{j-3} + g_4R_{j-4} + g_{13}R_{j-13} + g_{14}R_{j-14} + g_{17}R_{j-17} + g_{18}R_{j-18} + g_{19}R_{j-19}] + \delta_j R_j. \quad (8)$$

In Table 5, coefficients of eq. (8) found by minimizing the mean square value of the error δ_j are presented for all

Table 4. Correlation coefficients between (R_{j+1}/R_j) and $R_j, R_{j-1}, \dots, R_{j-19}$

Lag-RF	All India	HOMOG	COREM	WCIND	CNEIN	NEIND	NWIND	PENIN
R_j	-0.7629	-0.7618	-0.7534	-0.7615	-0.7341	-0.6846	-0.7339	-0.7671
R_{j-1}	0.1557	0.1913	0.1699	0.2160	0.0394	-0.1886	0.0751	0.0546
R_{j-2}	-0.0144	-0.0688	-0.0674	-0.1092	0.0040	0.1682	-0.0046	0.1268
R_{j-3}	-0.1718	-0.1667	-0.1026	-0.1401	0.0294	0.1113	-0.0991	-0.1039
R_{j-4}	0.1346	0.1036	0.0479	0.1401	-0.0348	-0.0944	-0.0232	0.1299
R_{j-5}	-0.0143	0.0124	-0.0208	0.0197	-0.0345	-0.1936	0.0094	-0.0674
R_{j-6}	-0.0296	-0.0565	-0.0069	-0.0892	0.0091	0.0023	0.0312	0.0806
R_{j-7}	0.0372	0.0716	-0.0597	0.1118	-0.1018	0.2095	0.0279	-0.0536
R_{j-8}	-0.1103	-0.0457	0.0032	-0.0709	0.0343	-0.1009	0.0073	-0.0373
R_{j-9}	0.0634	0.0222	0.0598	0.0089	-0.0594	-0.0670	0.0163	-0.0067
R_{j-10}	0.0025	0.0267	-0.0251	0.0309	0.0709	0.0295	-0.0282	0.0899
R_{j-11}	-0.0403	-0.0396	-0.0320	-0.0704	-0.0830	0.0432	0.0045	-0.1353
R_{j-12}	0.0380	0.0442	0.0368	0.0866	0.0994	0.1162	-0.0048	0.0027
R_{j-13}	0.1612	0.1052	0.0827	0.0592	0.0810	-0.0811	0.0633	0.2705
R_{j-14}	-0.2540	-0.2259	-0.1959	-0.1904	-0.1638	-0.1221	-0.2119	-0.1780
R_{j-15}	0.1055	0.2110	0.1338	0.1657	-0.0113	0.0941	0.1734	-0.1454
R_{j-16}	-0.0034	-0.0799	-0.0896	-0.0789	0.0689	-0.0210	-0.0639	0.1786
R_{j-17}	-0.1316	-0.0853	-0.0423	-0.0909	-0.0874	0.1009	-0.0412	-0.0413
R_{j-18}	0.2161	0.1983	0.1364	0.2341	0.1041	0.0633	0.0952	0.0118
R_{j-19}	-0.1388	-0.1920	-0.1669	-0.2356	0.0702	-0.0710	-0.0711	-0.0438

Table 5. Coefficients of eq. 8

Coeffs.	All India	HOMOG	Core monsoon	WCIND	CNEIN	NEIND	NWIND	Peninsular
a	-6.68e-6	-6.85e-6	-1.75e-6	-3.21e-6	-4.73e-6	-1.71e-6	-3.07e-6	-1.92e-6
b	0.0019	0.0017	0.0007	0.0011	0.0016	-0.0001	0.0053	0.0007
c	-0.1923	-0.1608	-0.0958	-0.1254	-0.1824	0.0136	-0.3110	-0.0828
d	7.5142	6.1107	4.8331	5.9716	8.2724	0.9138	7.0899	3.7494
g_1	0.0015	0.0028	0.0022	0.0031	-	-0.0012	-	-
g_2	-	-	-	-	-	0.0005	-	-
g_3	-0.0012	-0.0022	-	-0.0022	-	-	-	-
g_4	4.64e-5	-	-	-0.0003	-	-	-	-
g_5	-	-	-	-	-	-0.001	-	-
g_7	-	-	-	-	-	0.005	-	-
g_{11}	-	-	-	-	-	-	-	-0.0009
g_{13}	0.0029	-	-	-	-	-	-	0.0048
g_{14}	-0.0012	-0.0020	-0.0013	-0.0015	-0.0006	-	-0.0027	0.0019
g_{15}	-	0.0008	0.0008	0.0001	-	-	0.0024	0.0019
g_{16}	-	-	-	-	-	-	-	0.0009
g_{17}	-0.0018	-	-	-	-	-	-	-
g_{18}	0.0012	0.0012	0.0011	0.0017	-	-	-	-
g_{19}	-0.0005	-0.0001	-0.0006	-0.0008	-	-	-	-
σ	0.0914	0.1514	0.1712	0.1235	0.1064	0.0796	0.2979	0.1327

the eight sets of data. The standard deviation of δ is given in the last row of the table. It is noted that the error standard deviation σ_δ is less than the previous error value σ_ϵ in all the cases. This is an indication of the importance of adding the lag-year terms into the model equation. All the data with the exception of N-E India data series, exhibit long year connections of the order of 14–15 years. Eq. (8) is again a nonlinear map of higher dimension. Initial conditions for twenty past years are needed to solve this equation formally. By iterating the equation sufficient number of

times, one can find the steady state (stable fixed points) of the model. This will exhibit considerable variation about the long-term mean controlled by the cubic polynomial form of $f(R)$. The skill of this new model in forecasting is verified by finding the reduction in variance as explained previously. It is noted here for carrying out this exercise, we have to start from the year 1891 since twenty years of initial conditions are used. Table 3 presents the reduction of variance in a year-to-year forecast achieved by eq. (8). This is consistently higher than that of eq. (6).

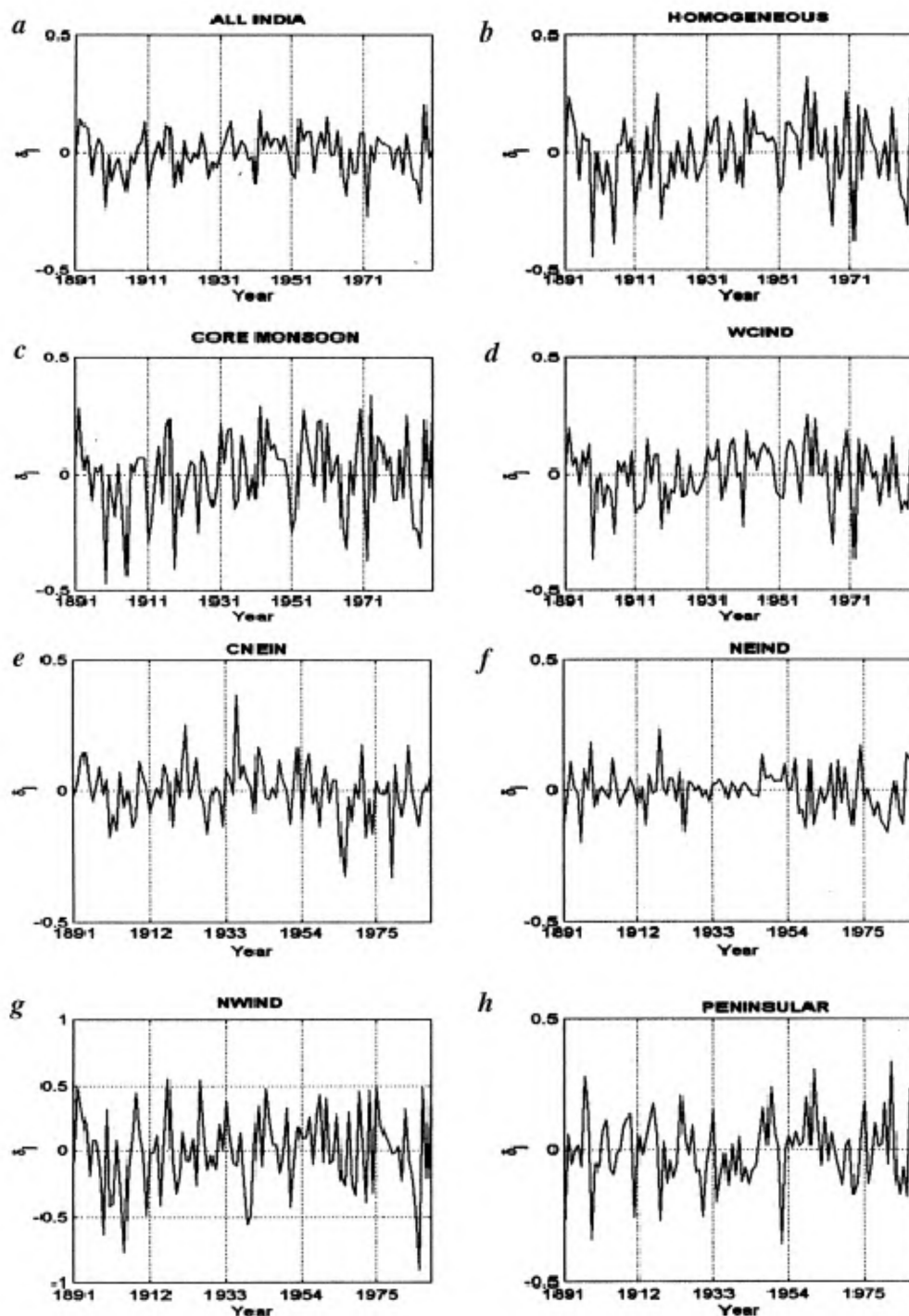


Figure 2. Time series of errors $\hat{\delta}$.

Error series

The efficiency of simulation of SWM rainfall by eq. (8) has been found to be better than eq. (6). But the reduction in variance achieved is still not high enough to warrant its use in forecasting. However, the last two steps provide the clue that it is the error term, which has to be further

studied to find any possible improvement. It is observed that the error term $\hat{\delta}$ appears as a factor of R_j , which is a known value. Thus, the only way the model can be improved is by reducing the standard deviation of this random error time series. This may be possible if $\hat{\delta}$ is not a purely random noise. The error time series $\hat{\delta}$ is shown in Figure 2 *a–h* for all the eight sets of data. This is a non-

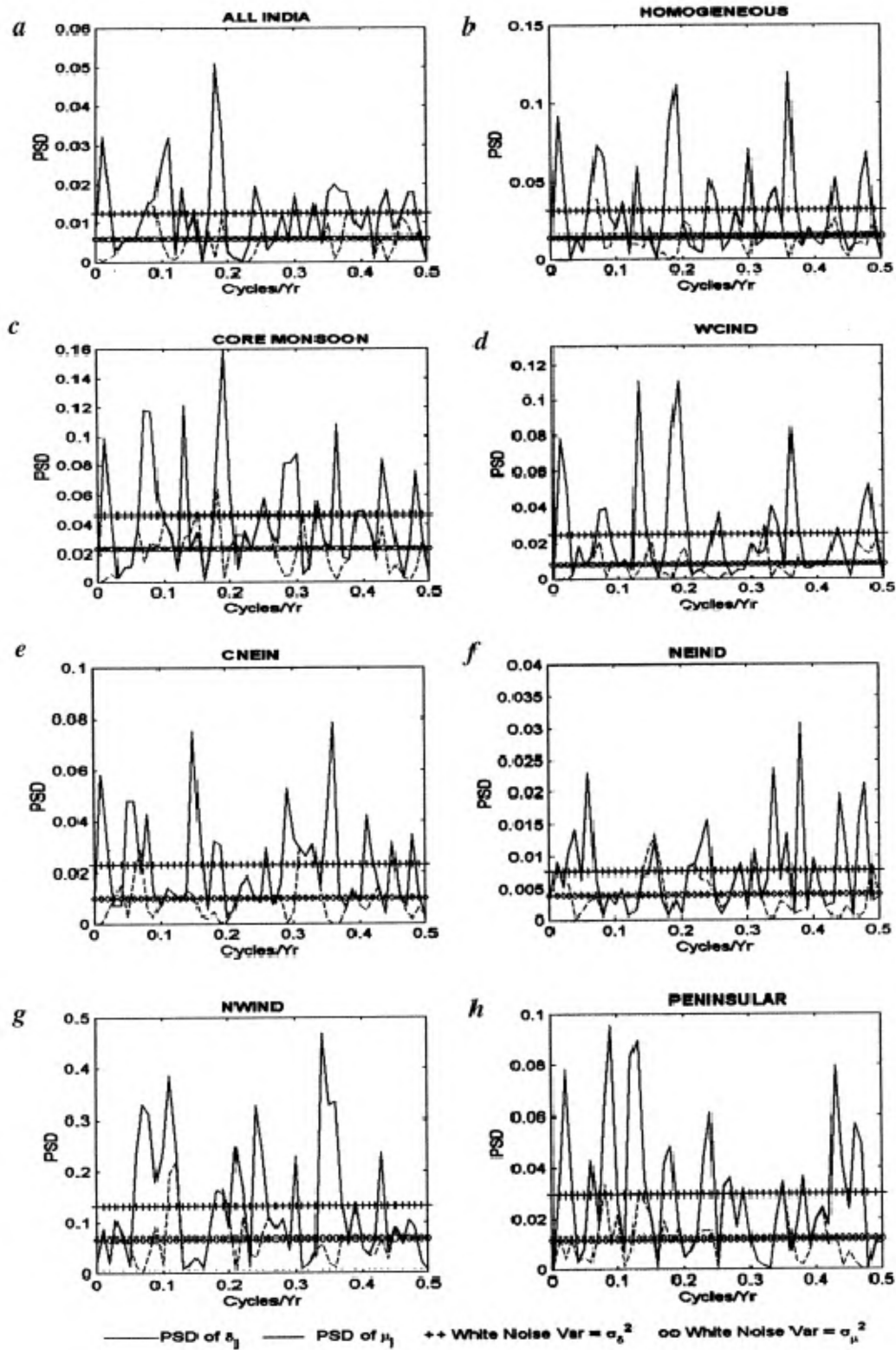


Figure 3. PSD of error time series.

dimensional quantity in $(-1, 1)$ and appears as a multiplicative term in finding the variance of the year-to-year forecast. The dominant feature of δ_j is its apparent periodicity of about 2–4 years. A better picture of the pattern

of δ_j emerges from its power spectral density (PSD) function. For all the eight regional rainfall data, the PSD of δ_j is shown in Figure 3a–h. These spectra exhibit dominant peaks at a few discrete frequencies. The amplitudes of

these peaks are conspicuously above the PSD of a strictly random series (white noise) with the same variance σ_{δ}^2 , which is a horizontal line in the above figures. All except NWIND, exhibit a long term periodic component of the order of several decades. Presence of dominant periodicities in SWM rainfall is reported in the literature^{13–15}. However when the basic time series R_j itself is handled directly, the variance associated with these periodic terms in relation to the overall variance turns out to be too small to be of practical utility. On the other hand, the present study indicates that the long-term periodicities enter the system through a multiplicative type of internal forcing term δ_j . Further improvement of eq. (8) depends on decomposition of δ_j into regular and random noise components. With this in view δ_j is expressed as

$$\delta_j = \sum C_i \sin(2\pi\lambda_i + \phi_i) + \mu_j. \quad (9)$$

In all the eight cases of data, nine predominant frequencies λ_i (cycles/year) are identified sequentially from the PSD function of δ_j . The amplitude of the first term C_1 is initially approximated from the PSD figure, corresponding to the largest peak. This value is used as a trial value to fit in a periodic function in the time period (1891–1990) by minimizing the mean square error. This predominant periodic term is removed from the error series to find the PSD of the remainder. The highest peak of

this new PSD is identified and removed as previously. This way nine most dominant periodic terms along with their (C_i, λ_i, ϕ_i) values, are determined. In Table 6, the values of (C_i, λ_i, ϕ_i) are arranged in decreasing order of C_i values. The resulting final model for the SWM rainfall data series will be

$$R_{j+1} = R_j[f(R_j, R_{j-1}, \dots, R_{j-19}) + \sum C_i \sin(2\pi\lambda_i + \phi_i)] + \mu_j R_j = \langle R_{j+1} \rangle + \mu_j R_j. \quad (10)$$

The PSD of μ_j is shown in Figure 3 *a–h* in dotted lines. This remainder error has been verified to be a white noise. The correlation between μ_j and R_j is found to be negligibly small and hence further improvement of the model at the annual scale appears unlikely. Even though the individual periodic terms in eq. (9) are small in absolute terms, their overall contribution to the variability of the parent series R_j is significant. This can be seen by generating a one-step forecast $\langle R_{j+1} \rangle$ from the expected value of eq. (10). In Figure 4, an example of this exercise is presented for the data of the homogeneous region. The skill of forecast is measured by the reduction in variance as explained previously. These results are presented in Table 3 and compared with the previous results of eqs (6) and (8). The reduction in variance obtained from eq. (10) is large and significant, raising the hope this model will be useful in forecasting monsoon rainfall values.

Table 6. Parameters of the harmonic terms (120-year sample)

<i>I</i>	Parameter	All India	HOMOG	COREM	WCIND	CNEIN	NEIND	NWIND	PENIN
1	C_1	0.0429	0.0663	0.0812	0.0721	0.051	0.0294	0.1562	0.0632
	$1/\lambda_1$	5.6211	5.1414	12.5313	5.1733	2.7902	22.2222	2.9180	8.0775
	ϕ_1	-1.0683	-0.4872	-0.5956	-0.038	-1.5012	-1.8986	-1.4612	-0.1833
2	C_2	0.0404	0.0632	0.0775	0.0548	0.049	0.0278	0.1393	0.0616
	$1/\lambda_2$	2.745	13.089	5.1125	7.4571	6.6181	2.0947	9.2336	10.6838
	ϕ_2	2.5492	-0.2692	-0.7379	-1.4402	1.6101	-0.113	2.1455	-1.1421
3	C_3	0.0368	0.0548	0.0735	0.0529	0.0469	0.0259	0.1393	0.0548
	$1/\lambda_3$	9.2851	2.7732	3.3523	76.9231	21.0084	18.4162	12.9534	42.5532
	ϕ_3	2.9980	2.9887	1.5769	-3.3542	-0.0524	5.108	-5.4044	-1.7240
4	C_4	0.0333	0.0529	0.0566	0.049	0.0442	0.0258	0.1241	0.0510
	$1/\lambda_4$	70.4225	82.6446	2.2925	2.0916	78.125	4.1017	13.6240	2.3015
	ϕ_4	-3.4742	3.1757	-2.7188	2.9081	-2.025	2.3461	5.7362	-2.7543
5	C_5	0.0311	0.0510	0.0566	0.049	0.0422	0.0252	0.12	0.0490
	$1/\lambda_5$	2.7122	2.0934	3.5398	2.7732	3.4483	2.2671	15.9744	2.1552
	ϕ_5	-1.5340	3.1442	-1.2086	2.9740	-0.4865	1.1129	-0.6948	0.0488
6	C_6	0.0299	0.042	0.0548	0.0434	0.0418	0.0252	0.1122	0.0469
	$1/\lambda_6$	2.0942	5.6465	94.3396	13.1752	2.4462	2.6302	4.1459	4.2808
	ϕ_6	3.1302	-0.6518	-2.5930	-0.2581	2.6204	3.2347	-1.5214	2.6901
7	C_7	0.0274	0.0418	0.0529	0.04	0.0369	0.0246	0.1000	0.0446
	$1/\lambda_7$	2.2920	2.2936	2.0912	3.0139	5.3763	2.9612	2.7632	5.7045
	ϕ_7	3.8172	-2.6042	3.0093	-2.3157	-2.6771	-0.9764	3.1310	-0.2201
8	C_8	0.0271	0.0413	0.0529	0.0385	0.0344	0.0233	0.097	0.0353
	$1/\lambda_8$	4.1356	3.3478	8.0645	5.6243	12.8866	6.3171	4.7985	2.6724
	ϕ_8	-2.4022	1.5645	-1.4585	-0.6984	-1.2034	-0.5683	3.7459	2.7906
9	C_9	0.022	0.0371	0.0510	0.0369	0.0315	0.019	0.0762	0.0340
	$1/\lambda_9$	7.4794	7.4460	2.7778	4.0683	2.0991	2.7663	2.3041	16.4474
	ϕ_9	-1.7933	-1.4769	3.4024	2.5227	3.3425	0.1949	-2.3631	-0.9193
σ_{δ}		0.0705	0.1058	0.1274	0.0818	0.0762	0.0565	0.225	0.1035

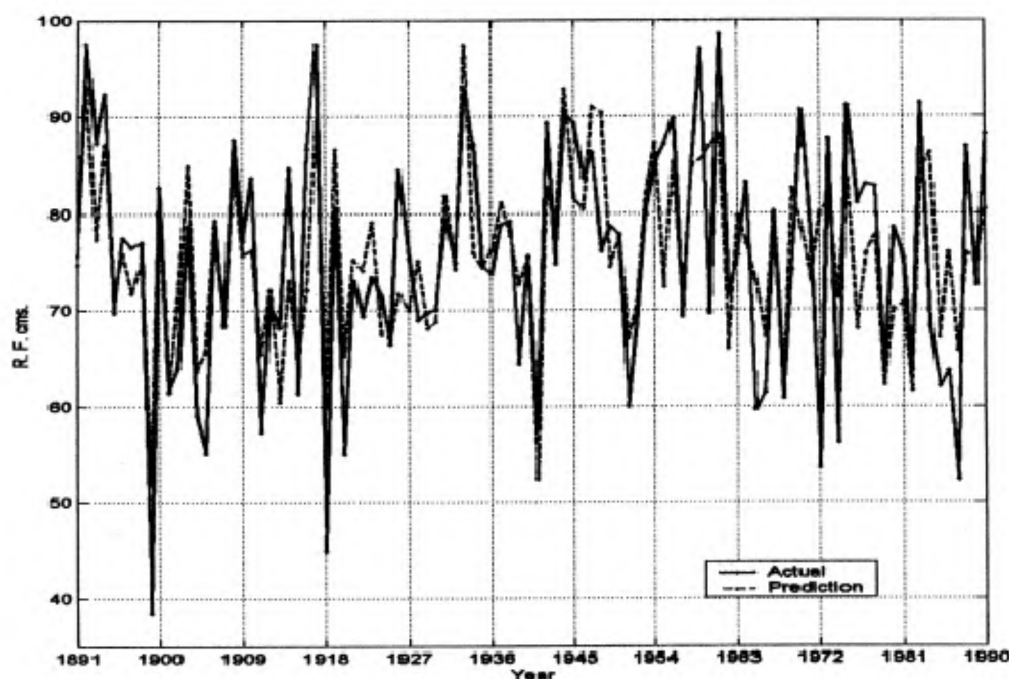


Figure 4. Year-to-year prediction (homogeneous region).

Test forecasting

The model proposed here for SWM rainfall has a deterministic part and a random part. The deterministic part $\langle R_{j+1} \rangle$ is a point predictor for the rainfall in the year $(j+1)$, conditioned on all past values. The statistical variability band about this predictor is $\pm \sigma_R R_j$. Thus, as a forecast we can say that the rainfall in year $(j+1)$ will be in the interval $[\langle R_{j+1} \rangle - \sigma_R R_j, \langle R_{j+1} \rangle + \sigma_R R_j]$ with nearly 67% probability. It is expected that in any year $\sigma_R R_j$ will be less than σ_R . Previously it was mentioned that rainfall data is available for the period 1871–2001. However, only 120 years data have been used for developing the model equation. Thus, the remaining data can be used for an independent verification of the forecasts. It is observed that the predictive part of the model consists of terms with constant coefficients and a time dependent part with harmonic terms. The coefficients of $f(R)$ have been found from a long sample and hence can be expected to remain stable for a few years. But the same cannot be said about the (C_i, λ_i, Φ) , as they are found using the sample length N as a reference quantity. Extrapolation of the harmonic terms by one year will be valid, but beyond that errors would accumulate affecting the forecasts. With this in the background, annual forecasts are presented and compared with observed values in Tables 7a–c. For the year 1991, the previous model has been directly used. For subsequent years, every year the harmonic part is corrected with an extra data added, before making a forecast for the next year. It is found that C_i and λ_i remain constant but

the phase angles show some variation, which improves the accuracy of the forecasts. Totally 88 predictions have been made here. In all the cases the expected value of the prediction is reported along with the one-sigma bands. The standard deviation of the prediction in any year is found to be less than the climatic deviation about the long term mean value, except for a few cases in the NEIND, NWIND data. To verify the statistical significance of the predictions, it is hypothesized that the prediction obtained could be purely due to chance. This null hypothesis (H_0) is tested through the standard χ^2 test. The observed numbers of correct predictions are 60, whereas due to chance the expected value would have been 44. For these frequencies, the observed χ^2 value will be 11.63, which is much higher than the tabulated value of 10.8 at the significance level of 0.001. Hence, the null hypothesis gets rejected and the conclusion that the model has significant forecasting ability stands verified.

Drought of 2002 and forecast for 2003

The model developed captures only the interannual variability of SWM rainfall data. Variabilities due to inter-seasonal, monthly, weekly and smaller scale fluctuations are not modelled here. Apart from this, due to the non-linear oscillatory nature of the phenomenon, bifurcations and attendant sensitivity to parameter errors cannot be avoided. Thus, it is expected that the model has to be updated every year with current data before the next year forecast may be produced. Thus, to foreshadow the

Table 7a. Test forecasting with independent data

Year	All India		HOMOG		Core monsoon	
	Actual	Prediction	Actual	Prediction	Actual	Prediction
1991	78.47	82.91 ± 6.41	64.50	57.18 ± 9.35	74.60	73.21 ± 13.12
1992	78.47	90.5 ± 5.78	73.36	69.25 ± 7.21	81.82	89.65 ± 9.51
1993	89.66	73.07 ± 5.44	79.95	56.01 ± 8.064	90.15	67.73 ± 10.41
1994	93.82	91.64 ± 7.99	91.74	84.63 ± 9.82	103.82	105.78 ± 11.41
1995	79.03	75.38 ± 6.77	64.74	73.06 ± 10.14	71.60	90.77 ± 13.43
1996	85.24	81.79 ± 5.55	73.61	63.58 ± 7.17	83.49	74.51 ± 9.23
1997	87.06	90.04 ± 5.93	71.94	70.47 ± 8.24	81.64	75.91 ± 10.76
1998	85.12	80.77 ± 6.71	74.67	73.59 ± 8.03	86.13	84.51 ± 10.5
1999	82.05	81.89 ± 5.96	67.90	76.62 ± 8.32	78.05	78.05 ± 11.03
2000	77.33	67.72 ± 6.91	63.30	77.79 ± 9.61	67.82	87.83 ± 10.01
2001	80.46	75.83 ± 5.67	67.80	61.65 ± 7.39	69.90	61.73 ± 9.17

Table 7b. Test forecasting with independent data

Year	WCIND		CNEIN		NEIND	
	Actual	Prediction	Actual	Prediction	Actual	Prediction
1991	83.56	85.53 ± 8.71	90.67	91.19 ± 8.07	138.40	143.01 ± 7.82
1992	83.39	96.28 ± 6.92	83.84	87.21 ± 9.38	113.72	150.06 ± 10.64
1993	100.04	73.83 ± 7.38	98.60	91.74 ± 6.42	174.01	133.40 ± 6.50
1994	106.44	104.04 ± 12.04	111.35	105.59 ± 7.62	121.74	118.58 ± 11.08
1995	73.94	87.28 ± 10.15	93.32	103.27 ± 9.13	165.36	132.5 ± 7.75
1996	83.30	93.62 ± 6.22	96.83	99.43 ± 9.84	130.23	142.06 ± 14.37
1997	81.85	84.83 ± 7.80	109.80	104.05 ± 7.68	154.43	142.82 ± 8.85
1998	90.17	94.73 ± 10.04	96.43	92.21 ± 11.31	132.96	132.41 ± 13.44
1999	90.70	97.60 ± 8.36	110.91	89.15 ± 7.75	149.42	132.16 ± 8.72
2000	81.56	112.76 ± 7.72	96.96	98.24 ± 11.58	133.22	130.29 ± 12.95
2001	83.35	78.78 ± 8.16	107.30	102.87 ± 7.87	131.27	128.78 ± 8.77

Table 7c. Test forecasting with independent data

Year	NWIND		Peninsular	
	Actual	Prediction	Actual	Prediction
1991	35.58	80.79 ± 13.66	76.75	71.79 ± 5.81
1992	58.17	59.19 ± 10.28	68.74	70.92 ± 7.98
1993	49.50	49.89 ± 13.17	62.03	60.26 ± 7.12
1994	69.37	57.21 ± 14.99	61.85	64.3 ± 7.86
1995	50.80	45.41 ± 15.78	59.93	59.37 ± 6.03
1996	58.89	63.62 ± 14.57	85.07	74.84 ± 6.18
1997	56.93	50.98 ± 13.83	71.33	80.35 ± 9.49
1998	51.14	67.06 ± 16.64	79.15	66.31 ± 9.67
1999	33.30	44.32 ± 11.87	57.93	68.14 ± 8.98
2000	35.61	51.24 ± 9.42	73.59	71.2 ± 7.98
2001	44.21	33.92 ± 8.38	66.08	58.85 ± 8.31

drought of 2002, it will be more appropriate to use corrected model parameters with a longer sample. Here, this has been done by considering 130 years data of 1891–2000. It is found that $f(R)$, which caters to the stationary part of the model does not change significantly. However, the oscillatory part with (C_i, λ_i, ϕ) gets altered. As examples, the revised model parameters for All India and homogeneous region rainfall are presented below.

All India region

$$\begin{aligned}
 \langle R_{j+1} \rangle = & R_j [-6.19e - 6R_j^3 + 0.0018R_j^2 \\
 & - 0.1810R_j + 7.042 + 0.0012R_{j-1} - 0.0013R_{j-3} \\
 & + 0.006R_{j-4} + 0.0027R_{j-13} - 0.0013R_{j-14} \\
 & + 0.0011R_{j-18} - 0.0003R_{j-19}] \\
 & + R_j [\Sigma C_i \sin(2\pi\lambda_i + \phi)]
 \end{aligned} \quad (11)$$

Homogeneous region

$$\begin{aligned}
\langle R_{j+1} \rangle = & R_j[-6.88e-6R_j^3 + 0.0018R_j^2 - 0.1681R_j \\
& + 6.1506 + 0.002R_{j-1} - 0.0018R_{j-14} \\
& + 0.0009R_{j-15} + 0.0015R_{j-18} - 0.0006R_{j-19}] \\
& + R_j[\Sigma C_i \sin(2\pi\lambda_i + \phi)].
\end{aligned} \quad (12)$$

Ten terms are identified in the time-dependent harmonic part of the series. These are tabulated in Table 8. With the help of the updated model, forecasts are made for three years 2001, 2002 and 2003. It is seen that the present model could have predicted the drought of 2002 unambiguously.

Discussion

The novelty of the present model is that seasonal rainfall is modelled as a generalized log-normal random variable. This is achieved by generalizing the law of proportions, which leads to the basic log-normal distribution. It is found that the long-term climatic mean value is essentially controlled by one step annual connections. This by itself is insufficient to simulate the observed variability of the data series. Hence, lagged year connections and periodicities at select frequencies are introduced to shape the model. The result is a nonlinear difference equation driven internally by a near-periodic term. This model accounts for nearly 50% of the interannual variability in all the eight sets of SWM data studied here. The nonlinear part of the model explaining the climatic average shows that for the estimated parameters there is only one steady state value. However, from Figure 1 e, g it is seen that the rate of change of annual rainfall can have inflection points. This would imply that CNEIN and NWIND may be sensitive to changes in the parameter values. In such a case, the mean part represented by eq. (2) may show two stable fixed points, between which the system would oscillate. The annual forecast in Table 7 is presented as an

expected value along with its standard deviation. The actual value is claimed to be in the predicted interval with a high probability of about 67%. The predicted variation is generally less than the climatic variation. It depends directly on the current year rainfall R_j , with a lesser value of R_j leading to a sharper forecast in the next year. For the year 2001, two different forecasts are presented. In Table 7 the forecast is based on 120 years of data. In Table 9 the forecast is made after updating the model with data up to year 2000. It is seen that the forecast is more favourable in the latter case, even though the two sets of results are not much different. The most stringent test for the model comes from year 2002. Gadgil *et al.*⁷ in their discussion of the drought of 2002 conclude that the deficit rainfall was part of the natural variability of the monsoon, but none of the usual empirical models could foreshadow the drought successfully. However, the results of Table 9 show that the present nonlinear difference model would have captured the low SWM rainfall of 2002, one year in advance. It may be observed that the drought is better reflected in the homogeneous and the core monsoon regions, which contribute significantly to the All India value. The SWM rainfall in the current year 2003 is naturally of immense interest. Since the present model works based on known past values, forecast has been possible only for the All India rainfall series. For the current year 2003 the model predicts an expected value of 82.65 cm, with a deviation of ± 4.88 cm. The meaning of this forecast is better seen in Figure 5, wherein the climatic and the forecast probability distribution functions of 2003 are compared. As the slope of the forecast PDF increases, the forecast becomes sharper and more significant. For the other seven regions also, the forecast for year 2003 has been given. However, these are not based on observed data of year 2002. In the absence of the true values for 2002 the expected values have been used to arrive at the 2003 values. Hence, these seven forecasts are qualitative in nature. Here, it would be informative to compare the present model with two recent empirical models of Sahai *et al.*⁹, proposed for All India rainfall. The first is an ANN model which uses only past rainfall data. Table 2 of this reference shows some statistical details of the observed and predicted standard deviations. Based on this information it is seen that this model explains about 60% of the climatic variance of seasonal AIRF time series. No results of the ANN model are available for the other regional rainfall data series. The second model of Sahai *et al.*¹⁶ is a linear statistical regression model dependent on global SST data. This model explains 72% of the climatic variance, which is higher than the variance explained by the present nonlinear model. However, it has to be noted that the present model uses only past rainfall values, whereas the model of Sahai *et al.*¹⁶ requires widespread sea surface temperature data.

Table 8. Parameters of harmonic terms (130-year sample)

All India			Homogeneous		
C_i	λ	ϕ	C_i	λ	ϕ
0.047	0.177	-0.878	0.072	0.010	3.759
0.042	0.357	-2.005	0.069	0.078	-0.238
0.040	0.112	1.585	0.068	0.062	-0.327
0.038	0.014	2.974	0.063	0.134	-1.424
0.034	0.298	2.151	0.055	0.333	3.650
0.033	0.477	3.168	0.055	0.359	3.751
0.031	0.134	-2.055	0.053	0.196	-0.989
0.027	0.251	2.429	0.049	0.478	3.099
0.028	0.080	-0.935	0.042	0.247	3.277
0.021	0.407	2.260	0.041	0.299	1.602

Table 9. Forecast for the year 2003

Region	Actual 2001	Prediction 2001	Prediction 2002	Forecast 2003
All India	80.46	82.01 ± 5.49	72.18 ± 5.69	$82.65 \pm 4.88^*$
HOMOG	67.8	56.33 ± 6.69	48.36 ± 7.19	$69.05 \pm 5.13^{**}$
Core monsoon	69.9	72.56 ± 7.82	55.78 ± 8.04	$83.16 \pm 6.42^{**}$
WCIND	83.35	76.47 ± 7.87	71.41 ± 8.02	$87.31 \pm 6.87^{**}$
CNEIN	107.3	102.98 ± 7.42	120.16 ± 8.18	$99.56 \pm 9.16^{**}$
NEIND	131.27	139.93 ± 9.33	156.05 ± 9.20	$162.46 \pm 10.93^{**}$
NWIND	44.21	36.13 ± 8.06	31.09 ± 9.99	$39.32 \pm 7.03^{**}$
Peninsular	66.08	67.82 ± 7.18	73.92 ± 6.52	$56.92 \pm 7.29^{**}$

*Forecast for 2003 is based on the actual 2002 value of 68.96 cm.

**Data for 2002 is not available. Forecast for 2003 is based on expected value of 2002.

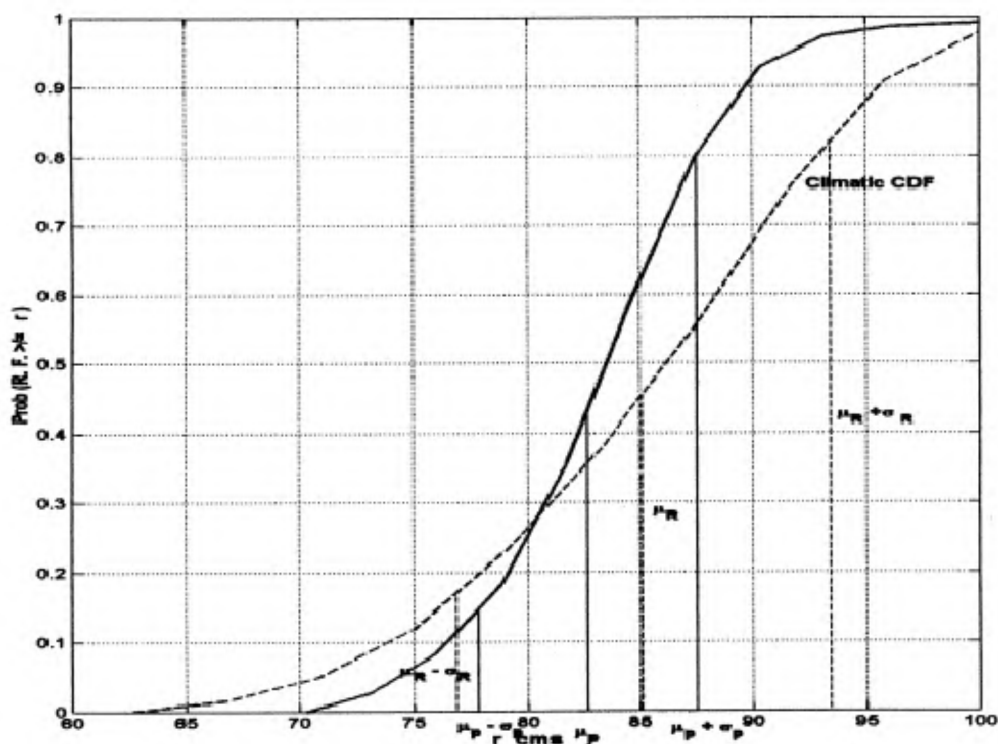


Figure 5. Statistical forecast for year 2003. $\mu_F = \langle R_{2003} \rangle = 82.65$, $\sigma_F = 4.88$.

Summary and conclusion

A new nonlinear dynamic model for SWM rainfall has been proposed in this paper. This is not a mechanistic dynamic model, which incorporates interplay of several atmospheric variables, but it attempts to unravel the structure of the data series so as to analytically extend it by one year. An algorithm for doing this is developed in three stages. The first stage takes care of the climatic mean behaviour. In the second and third steps, important previous year connections and several periodic modulating terms are built into the model. It is demonstrated that in seven out of the eight cases of regional rainfall data considered here, the new model explains successfully

50% of the interannual variability. The year-to-year forecasting ability of the model is verified to be statistically significant on an independent sample of 11 years from all the eight regions. For the year 2003, the All India rainfall is forecast to be (82.65 ± 4.88) cm, with a probability of 67%. The capability of the model to explain rainfall in smaller regions such as coastal Karnataka and at city level is under investigation. At present, only the interannual variability terms have been incorporated into the model. It would be interesting to see how the present random error part can be decomposed into interseasonal, monthly and fortnightly variabilities^{17,18}. It is hoped such efforts will further improve the forecast ability of the present model.

1. Mooley, D. A. and Parthasarathy, B., Fluctuations in all-India summer monsoon rainfall during 1871–1978. *Climatic Change*, 1984, **6**, 287–301.
2. Gregory, S., Macro-regional definition and characteristics of Indian summer monsoon rainfall, 1871–1985. *Int. J. Climatol.*, 1989, **9**, 465–483.
3. Hastenrath, S. and Greischar, L., Changing predictability of Indian monsoon rainfall anomalies? *Proc. Indian Acad. Sci. (Earth Planet. Sci.)*, 1993, **102**, 35–47.
4. Rupa Kumar, K. *et al.*, Spatial and subseasonal patterns of the long-term trends of Indian summer monsoon rainfall. *Int. J. Climatol.*, 1992, **12**, 257–268.
5. Thapliyal, V., Large-scale prediction of summer monsoon rainfall over India: Evolution and development of new models. *Mausam*, 1990, **41**, 339–346.
6. Iyengar, R. N. and Basak, P., Regionalization of Indian monsoon rainfall and long-term variability signals. *Int. J. Climatol.*, 1994, **14**, 1095–1114.
7. Gadgil, S. *et al.*, On forecasting the Indian summer monsoon: The intriguing season of 2002. *Curr. Sci.*, 2002, **83**, 394–403.
8. Sahai, A. K., Grimm, A. M., Satyan, V. and Pant, G. B., Prospects of prediction of Indian summer monsoon rainfall using global SST anomalies. Contribution from Indian Institute of Tropical Meteorology, Report No. RR–093, 2002, Pune, India.
9. Sahai, A. K., Soman, M. K. and Satyan, V., All India summer monsoon rainfall prediction using an artificial neural network. *Climate Dynamics*, 2000, **16**, 291–302.
10. Kedem, B. and Chiu, L. S., On the lognormality of rain rate. *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 901–905.
11. Pathasarathy, B., Rupa Kumar, K. and Munot, A. A., Homogeneous Indian monsoon rainfall: Variability and prediction. *Proc. Indian Acad. Sci. (Earth Planet. Sci.)*, 1993, **102**, 121–155.
12. Johnson, N. L. and Kotz, S., *Distributions in Statistics: Continuous Multivariate Distributions*, New York, Wiley, 1972.
13. Campbell, W. H., Blechman, J. B. and Bryson, R. A., Long-period tidal forcing of Indian monsoon rainfall: An hypothesis. *J. Climate Appl. Meteorol.*, 1983, **22**, 287–296.
14. Bhalme, H. N. and Jadhav, S. K., The double (Hale) sunspot cycle and floods and droughts in India. *Weather*, 1984, **39**, 112.
15. Narasimha, R. and Kailas, S. V., Quasi-cycles in monsoon rainfall by wavelet analysis. *Curr. Sci.*, 2000, **78**, 592–595.
16. Sahai, A. K., Grimm, A. M., Satyan, V. and Pant, G. B., Long-lead prediction of Indian summer monsoon rainfall from global SST evolution. *Climate Dyn.*, 2003, **20**, 855–863.
17. Hartmann, D. L. and Michelsen, M. L., Intraseasonal periodicities in Indian rainfall. *J. Atmos. Sci.*, 1989, **46**, 2838–2862.
18. Ajaya Mohan, R. S. and Goswami, B. N., A common spatial mode for intra-seasonal and inter-annual variation and predictability of the Indian summer monsoon. *Curr. Sci.*, 2000, **79**, 1106–1111.

ACKNOWLEDGEMENT. Thanks are due to Profs. J. Srinivasan, Sulochana Gadgil and B. N. Goswami for useful discussions.

Received 5 June 2003; revised accepted 28 August 2003

Retraction

We write in relation to a paper recently published in your journal (Venugopal, J., Boyle, N. and Kelly, J. P., *Current Science*, 2003, **84**, 1227–1231) and wish to bring the following points to your attention. Neither of us were involved in the preparation of this work for publication. The only experimental work cited in this paper that relates to the Department of Pharmacology occurred under the auspices of a Millennium Grant from NUI, Galway which investigated whether PDE γ was present in mouse brain, and was conducted by Venugopal and Boyle during June–August 2001. This work only represents a single paragraph of this paper and does not include any data. In our opinion, it is clear that the work conducted in the Department of Pharmacology, NUI, Galway during June–August 2001 only contributed in a minor way to this paper, and thus the inclusion of our names as the only co-authors is inappropriate, as is the use of the Department

of Pharmacology as the sole address at which the work was conducted. The same applies for the acknowledgement relating to the funding by the Millennium Grant from NUI, Galway. We wish to take this opportunity to publicly dissociate ourselves from this publication.

JOHN P. KELLY
NOREEN BOYLE

Due to issues relating to the permission for the use of materials that belong to the laboratory of N. J. Pyne, Department of Physiology and Pharmacology, University of Strathclyde, I hereby withdraw the manuscript. I hereby state that all the results are genuine and reproducible.

JOSHI VENUGOPAL
Corresponding author