494

Proc. Indian Acad. Sci. (Earth Planet. Sci.), Vol. 103, No. 4, December 1994, pp. 439–448.

Singular spectral analysis of homogeneous Indian monsoon (HIM) rainfall

G K RANGARAJAN

Indian Institute of Geomagnetism, Colaba, Bombay 400 005, India

MS received 6 December 1993; revised 1 July 1994

Abstract. Homogeneous Indian Monsoon region rainfall for the epoch 1871–1990 has been analysed using Singular Spectral Analysis. It is shown that the HIM time series is simple in structure with only the annual oscillation and its first two harmonics accounting for almost the entire variability. Longer period oscillations related to lunar tidal forcing, solar activity and quasibiennial variation are conspicuously absent. It is also shown that the singular spectral decomposition is closely similar to complex demodulation and thus provides variations in the signals which evolve only slowly with time. As the rainfall series is marked by several jerky changes, predictability of HIM rainfall through the principal components derived from SSA appears impossible.

Keywords. Singular spectral analysis; monsoon; rainfall; complex demodulation.

1. Introduction

This investigation is an outcome of two interesting communications in the special issue on 'Climate and global warming' of the Proceedings of the Indian Academy of Sciences (Subramaniam Moten 1993 and Parthasarathy *et al* 1993). To study the multiple time scales in rainfall variability in Malaysia, Subramaniam Moten utilized the technique of Singular Spectral Analysis (SSA) and determined the different modes of oscillation in the rainfall series in peninsular Malaysia. Parthasarathy *et al* (1993) prepared a spatially coherent monsoon rainfall series using fourteen sub-divisions covering north-western and central parts of India (about 55% of the total area of the country). In view of their similar rainfall characteristics and associations with regional/global parameters, the area-weighted mean values constitute the monthly and seasonal Homogeneous Indian Monsoon (HIM) rainfall for a long period covering the epoch 1871–1990. In their statistical analysis, Parthasarathy *et al* (1993) found the HIM series to be free from persistence, that the recent three decades is marked by high variability and that it has a significant quasibiennial oscillation (QBO). They have also listed the monthly, seasonal and annual rainfall of the HIM region for 1871–1990 "for the benefit of further research".

It is now our effort to use the advantages of SSA in the study of HIM rainfall in India to isolate dominant principal components and see if the rainfall can be predicted with reasonable certainty using these signals.

2. Singular spectral analysis: Formulation and advantages

Vautard and Ghill (1989) have highlighted how SSA provides quantitative and qualitative information about the deterministic and stochastic parts in a time series.

Their development was a modification of the originally proposed application of SSA to the problems of dynamical systems theory by Broomhead and King (1986), which showed that the statistical dimension S of the system is dependent on the length of the state vectors (M). In fact, the term 'Statistical Dimension' was introduced by Vautard and Ghill (1989).

We begin with a finite time series $y(t)$ of length N

$$y(t) = y(Kt_s) \qquad k = 1, 2, \ldots N \tag{1}$$

where t_s is the sampling interval.

This series after normalisation using the mean (Y) and standard deviation (σ_Y) yields the series

$$X(t) = \frac{y(t) - Y}{\sigma_y} \qquad t = 1, 2, \ldots N \tag{2}$$

The sampled time series is then embedded in an M-dimensional space by taking as state vectors the consecutive sequences

$$Z = \begin{bmatrix} X_1 & X_2 & \cdots & \cdots & X_M \\ X_2 & X_3 & \cdots & \cdots & X_{M+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ X_{N-M+1} & \cdots & \cdots & \cdots & X_N \end{bmatrix} \tag{3}$$

The matrix so derived is known as Trajectory Matrix.

For different choices of M, we can have different trajectory matrices. However M should be larger than the autocorrelation time (the lag at which the first zero occurs). According to Broomhead and King (1986), the reconstruction of the attractor is then guaranteed under certain hypothesis on M (the viewing window), T_s (the sampling interval) and the smoothness of the time series.

From the trajectory matrix defined in (3), we can generate the correlation matrix C taking the product of each column with all the others successively. These are, in effect, the lagged auto correlation coefficients. Alternately the singular values can be obtained directly by singular value decomposition of the data matrix (see Reyment and Joreskog 1993 for e.g.). However, when we use the normalized values as shown in equation (2), the covariance matrix generated from the data and the correlation matrix will yield the same results.

Broomhead and King (1986) have given useful recurrence relation for determining the elements of the covariance matrix as outlined below:

C_{kl} an element of the covariance matrix is given by

$$C_{kl} = \sum_{i=1}^{N} v_{i+k-1} v_{i+l-1}. \tag{4}$$

Then

$$C_{k+1\,l+1} = C_{kl} + \frac{1}{N}(v_{N+k} v_{N+l} - v_k v_l) \tag{5}$$

Thus only the elements of the first row in the matrix need to be directly calculated. The amount of data required for calculating the lagged autocorrelation should ensure convergence of the autocorrelation function of a stationary process. This implies an upper bound on M sufficiently less than N.

Once the covariance matrix is obtained, the next step is the simple one of finding its M eigen values in descending order and the corresponding M eigen vectors. Ideally, the number of non-zero eigen values would correspond to the number of independent variables. In the presence of noise in the data, the others will be close to zero (but not actually zero) and will define a noise level (Sharma *et al* 1993).

The square root of the eigen values are called 'Singular Values' and their set is called 'Singular Spectrum'. The successive singular values can be arranged in monotonically decreasing order and the noise level then appears in the Singular Spectrum as a flat 'floor' at its tail. These are associated with non-significant principal components (Vautard and Ghill 1989).

The projections of the time series along the directions of the eigen vectors then yield the time variations of individual components. In other words, the eigen vectors serve as data adaptive filters whose transfer functions delineate sharp spectral peaks within narrow individual bands, thus providing a significantly more flexible tool than conventional spectral analysis (Vautard and Ghill 1989). If the eigen vectors appear as even and odd pair in phase quadrature, it is indicative of pure oscillations in the time series with the periods less than the length of the viewing window (Subramaniam Moten 1993). Also, the distance of the pairs of singular values above the noise floor is indicative of the fact that the corresponding modes are deterministic and not stochastic (Schlesinger and Ramankutty 1994).

3. Data analysis

The monthly values of HIM rainfall from 1871 to 1990 (in mm) are the inputs to the Singular Spectral Analysis. The time series, after normalisation, are used to generate trajectory matrices of varying lengths with $M = 25$, 49, 61, 73. The corresponding singular spectrums all showed a noise floor beyond $M = 10$. For subsequent analysis we use $M = 49$.

4. Results and discussion

The singular spectrum for order length upto 30 is shown in figure 1. All the four spectra have the same shape and noise floor. The confidence intervals for the eigen values are derived from the error estimate given by

$$\delta\lambda_k = 2\lambda_k(2/N)^{1/2}$$

(Vautard and Ghill 1989). The 95% confidence level indicated in the figure corresponds to the largest eigen value. With this criterion, only first six eigen values turn out to be significant. Examination of the principal components associated with the singular values for order 8 and above indicates that the time series are contaminated largely by noise and therefore we can estimate the upper limit of the statistical dimension for the HIM data as 8, but it is likely to be closer to 6, as seen from figure 1.

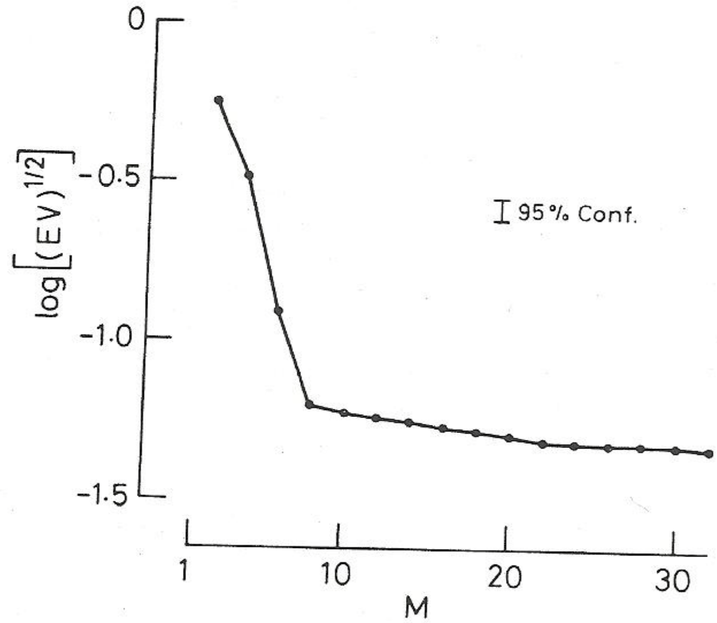

Figure 1. The singular spectrum for HIM rainfall series for M upto 30. The ordinate scale is log of the square root of the eigen values in descending order. As the singular values for sinusoidal signals appear as even and odd pairs of nearly same values, only alternate values are shown.

The first six modes together account for 90% of the total variability, the annual components for 65%, the semiannual components for 22% and the 4-monthly (triannual) component for 3%. The remaining 10% is accounted for by several subsequent components none of which is significant individually.

The eigen vectors (even and odd parts) for each of these three major oscillations are shown in figure 2. Apart from annual variation in HIM and its first two harmonics, there are no other periodicities in the entire data length. This is in sharp contrast to the reported longer period oscillation in Indian rainfall and peninsular Malaysian rainfall with a periodicity of 18·5 years apparently due to long-term lunar tidal potential caused by sun-earth-moon system (Subramaniam Moten 1993). This conspicuous absence could be due to the procedure of deriving HIM series which is weighted dominantly in favour of south west monsoon rainfall, with July-August the wettest and the seven months November through May being the driest.

In view of the simple structure with only 3 major components for the rainfall it becomes interesting to study the time variations in the individual sequences. For this purpose the eigen vectors (only even vectors need to be considered, here) are used as data-adaptive filters to derive the filtered time series i.e. the three principal components. From the filtered series for each year, the range – the difference between maximum and minimum normalized rainfall – is computed and their time evolutions are shown in figure 3. These curves provide a good visual picture of the variability in HIM rainfall in terms of its individual components. In contrast to the annual component, the semi-annual and triannual terms show less variability. The epochs 1892–1919 and 1952–1982 show very large swings in the range of annual precipitation and for part of the duration, between 1930 and 1964, there is a discernible cyclicity. These features tempt us to use autoregressive technique of analysis to predict the

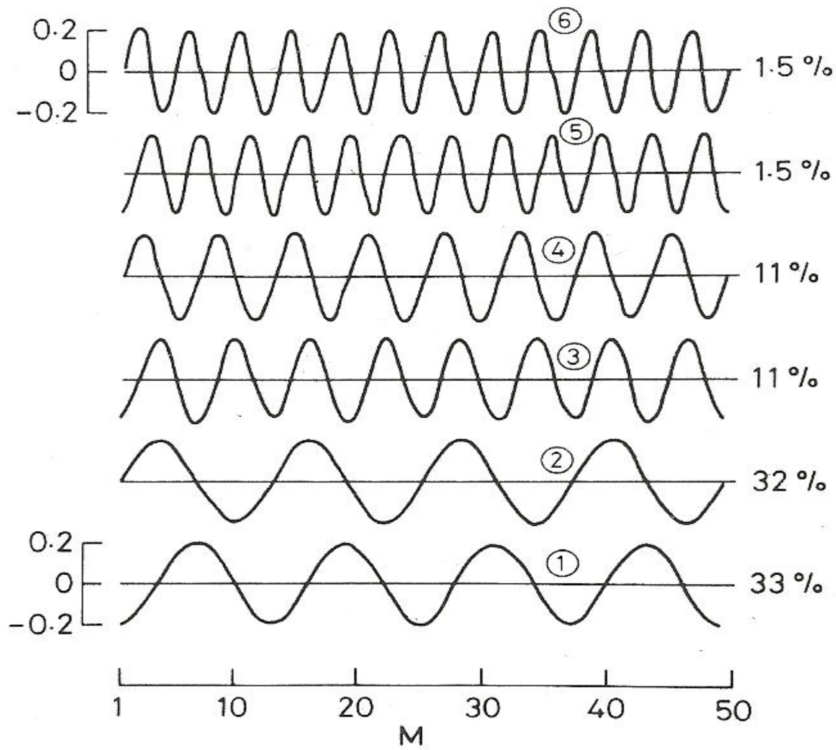

Figure 2. The first 6 eigen vectors corresponding to the six largest eigen values. The vectors are plotted for $M = 49$. Note the exact phase quadrature between individual pairs 1–2, 3–4 and 5–6 corresponding to 12-, 6- and 4-month oscillations.

individual principal components and therefore make an educated guess about the total annual rainfall in the HIM region.

Since the dominant periodicities in the HIM series have been identified with the annual cycle and its harmonics, we could use the complex demodulation analysis to get the time variations in their amplitudes and phases. Details of the method of complex demodulation are provided in Banks (1975). Agarwal *et al* (1980) showed the relationship between narrow bandpass filtering and the demodulate estimates of amplitude and phase.

Using a version of Fast Fourier Transform valid when the number of data points can be given by $N = 2^{k1} \times 3^{k2}$ where $k1$ and $k2$ are integers, we are able to derive the demodulates for narrow bands centred precisely on 12-, 6- and 4-month periodicities (Yfantis and Borgman 1981).

The demodulate amplitudes for the 3 principal components are also shown in figure 3. The variability in the amplitudes is quite similar to that in the range of the normalized HIM rainfall. The phase angles, on the other hand show excellent stability with average values $252° \pm 2°$, $58° \pm 6°$, and $240° \pm 16°$. These correspond to maxima for annual component near July 31, that for semi-annual near January 30 and July 30, and that for the four-month component near March 15, July 15 and November 15 respectively.

We then attempt to model the demodulates using Maximum Entropy Method. The periodicities corresponding to the spectral peaks their amplitudes and phases as also the percentages accounted for by these sinusoidal oscillations in the total variance

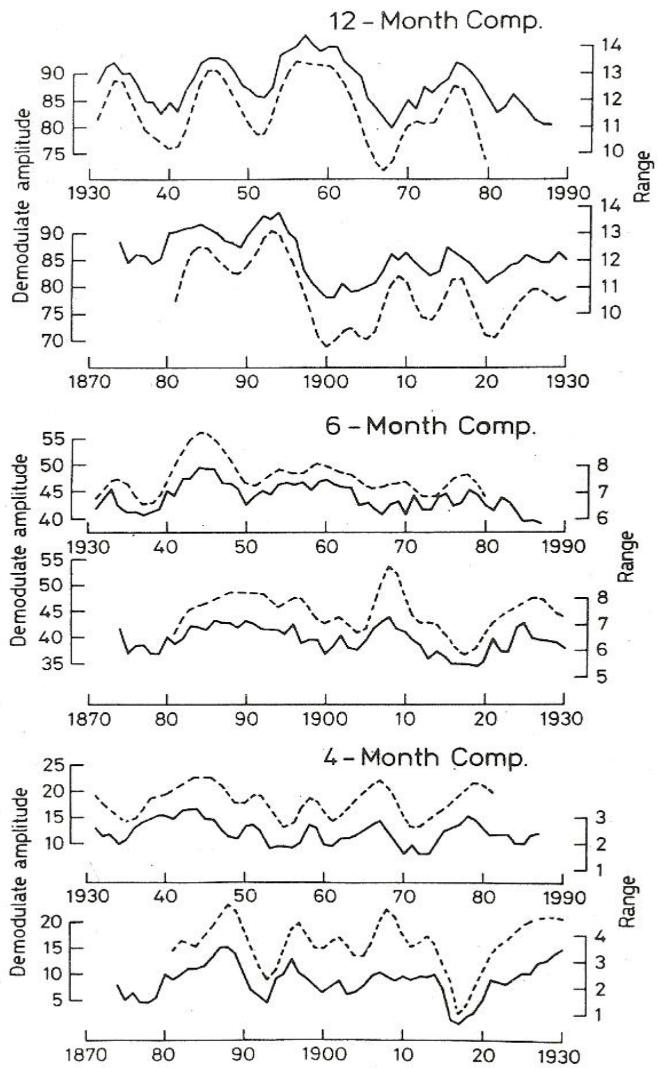

Figure 3. Time variation in the range of annual variation (maximum − minimum in the normalized HIM rainfall) for every 12 months computed from the three principal components (continuous line with scale to the right). Also shown are the demodulate amplitudes centred on 12-, 6- and 4-month periodicities (broken line with scale to the left).

of the demodulate amplitude series are shown in table 1. The validity of the model is for the epoch 1881–1980 as the first 10% and last 10% of the data are not realistically reproduced due to the cosine bell taper adopted prior to FFT to avoid undesirable leakages (Bloomfield 1976).

As the time variations of the amplitude of the three periodicities (12-, 6- and 4 months) could be closely approximated by the cyclic components in table 1, we can compute the total annual variation for each year by deriving the expected

Table 1. Periodicities in the demodulate amplitude of the Annual Component and its first two harmonics. Percentage of total variance accounted by individual cyclic component is also given.

Demodulate component	Period (years)	Amplitude	Phase (degree)	Percentage accounted
Annual	81·82	4·98	133	33
	22·22	4·14	300	23
	15·00	2·21	8	6
	13·04	3·32	82	15
	10·34	2·00	350	5
	7·38	1·41	112	3
Semiannual	54·55	2·34	11	21
	18·18	1·39	219	7
	17·14	2·61	242	26
	11·76	1·73	295	11
	6·32	0·90	334	3
Triannual	18·95	3·21	309	35
	13·95	1·60	11	9
	11·69	1·05	273	4
	8·57	1·29	115	6
	6·36	1·36	11	6

amplitudes for the individual year and using the constant phases as shown below:

$$A_{12}\sin(30t + 252^\circ) + A_6\sin(60t + 58^\circ) + A_4\sin(90t + 240^\circ) + A_G, t = 0, 1, \ldots 11 \quad (6)$$

where A_G is the gross average annual rainfall taken as 72·3 mm (Parthasarathy *et al* 1993).

In view of the known scaling down of the demodulate amplitudes in relation to the expected values even for sinusoidal cycles (Agarwal *et al.* 1980) we utilised normalization factors for A_{12}, A_6 and A_4 by taking ratios of the average amplitudes obtained from harmonic analysis for individual years and the average of the demodulate amplitudes. The values adopted are 107/81, 63/46 and 25/17 respectively.

Despite the fact that the HIM series is dominated only by 3 components, the computed annual rainfall could not match the observed rainfall even for the closed loop period of analysis 1881–1980 so that predictability beyond this range becomes meaningless. This mismatch was initially quite unexpected. A closer examination of the temporal evolution of the individual harmonic components derived taking 12 monthly values at a time shows very drastic changes in amplitude from one year to the other. For example the annual rainfall for successive years like 1877 and 1878 (760 and 1045 mm) 1894 and 1895 (1070 and 792 mm), 1898–99 and 1990 (855, 428, 890 mm), 1917–1918 (1225 and 541 mm) etc. are so widely different that the fundamental assumption in complex demodulation that the signal should have *slowly varying* amplitude and phase (Bloomfield 1976) is immediately violated.

In a way, this non-predictability of the HIM rainfall despite its simple structure brings to the fore an aspect of Singular Spectral Analysis which has not been brought out hitherto. We conclude that the principal components derived from SSA can highlight only the slowly varying signal similar to the complex demodulation and hence care has to be taken while interpreting the amplitude variation of any signal. The failure of predictability of the rapidly changing annual HIM rainfall can thus be explained.

5. Spectrum of annual rainfall

Subramaniam Moten (1993) has shown that in peninsular Malaysia, longer period oscillations with periods 18·5 years, 7–10 years, 3–5 years related to the ENSO mode, and QBO could be clearly detected through SSA. Parthasarathy *et al* (1993) however list only a weak QBO signal in the HIM rainfall series. The monthly values of HIM analysed here also does not indicate any longer period variations in terms of significant principal components.

Singular spectrum is known to be suitable when the time series is short and noisy (Vautard and Ghill 1989). Similarly the data-adaptive Maximum Entropy Method of spectral analysis can also be utilised for detecting periodicities in short series, such as even a truncated sinusoid (Ulrych and Bishop 1975).

For singular spectrum of the different annual rainfall series (columns 14 to 18 of table 3 of Parthasarathy *et al* 1993) we use a viewing window of 25 for the total length 120. The singular spectrum (not shown here) has no discernible noise floor with the singular values decreasing almost monotonically. The variance accounted

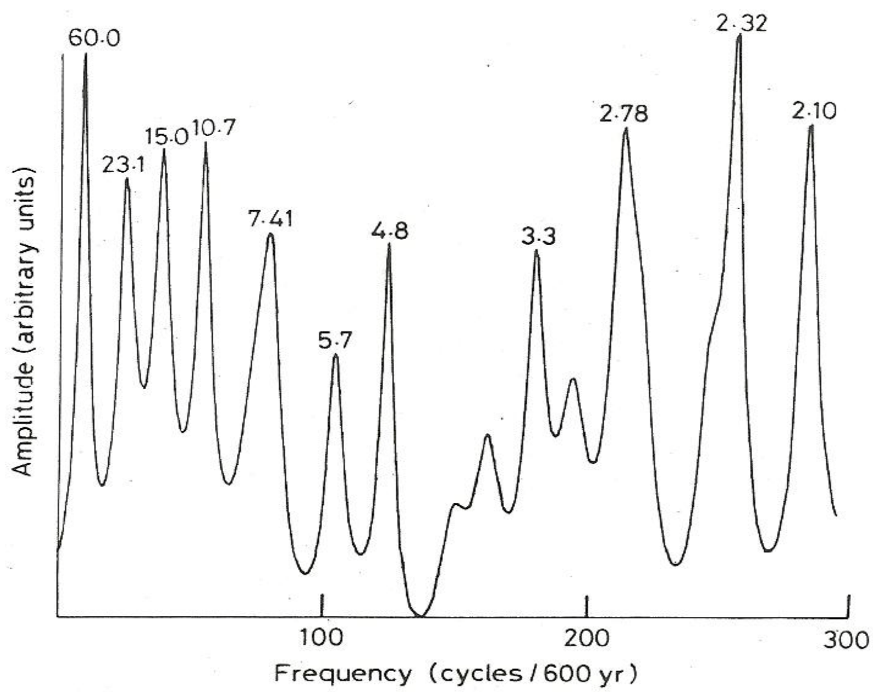

Figure 4. A typical high resolution spectrum of annual rainfall series from Maximum Entropy Method. The numbers shown on individual spectral peaks correspond to the dominant periodicities.

by individual vectors varies between 8·8% and 1·5% almost equally distributed among all the vectors. In view of this, no upper limit on the statistical dimension of the annual rainfall series could be estimated successfully.

The MEM spectrum for the total annual precipitation with length of the prediction error filter of 50, for $N = 120$ is shown in figure 4. Apparent periodicities (in years) are indicated. The other annual mean series exhibit different apparent periodicities. The spectrum is a clear indication that the annual rainfall series of HIM region is almost a random series with no dominant periodicities corroborating the inference based on the singular spectrum. Due to this random behaviour, any prediction into the future based on the limited data of annual rainfall will be futile.

6. Conclusion

The Homogeneous Indian Monsoon rainfall series derived from a judicious choice of meteorological subdivisions is simple in structure which could be accounted for in terms of an annual cycle and its first two harmonics only. This is largely due to the fact that the HIM series is dominated by monsoon rainfall unlike the rainfall series for peninsular Malaysia, for instance. The annual rainfall series also appears to be free of any dominant periodicities and random in nature. Even the quasibiennial oscillation is not clearly seen though in the MEM spectrum the largest signals are in the period range 2·78 to 2·10 years.

The ranges of the annual variation in the 3 major principal components show some interannual variability and a highly consistent phase but attempts to model the HIM rainfall from the amplitudes derived from complex demodulation and the constant phase does not succeed in accurate prediction of the HIM for any epoch. This helps us to highlight one aspect of SSA, not mentioned in literature earlier. The principal components from SSA can delineate only slow and smooth changes in the signal but cannot account for sharp and rapid fluctuations.

Acknowledgements

The contribution of Parthasarathy *et al* (1993) in listing the monthly values of HIM rainfall which paved way for this paper is gratefully acknowledged. The comments of referees which helped in improvements of the text is also acknowledged with thanks.

References

Agarwal A K, Singh B P and Nityananda N 1980 An application of complex demodulation technique to geomagnetic data and conductivity studies; *Proc. Indian Acad. Sci.* (*Earth Planet. Sci.*) **89** 67–77

Banks R J 1975 Complex demodulation of geomagnetic data and the estimation of transfer functions; *Geophys. J.R. Astron. Soc.*, **43** 87–101

Bloomfield P 1976 *Fourier analysis of time series: An introduction*, (New York: John Wiley & Sons)

Broomhead D S and King G P 1986 Extracting qualitative dynamics from experimental data; *Physica* **D20** 217–236

Parthasarathy B, Rupakumar K and Munnot A A 1993 Homogeneous Indian monsoon rainfall; *Proc. Indian Acad. Sci.* (*Earth Planet. Sci.*) **102** 121–155

Reyment R and Joreskog K G 1993 *Applied factor analysis in the natural sciences* (U.K.: Cambridge Press)
Schlesinger M E and Ramankutty N 1994 An oscillation in the global climate system of period 65–70 years; *Nature* (*London*), **367** 723–726
Sharma A S, Vassiliadis D and Papadapoulos K 1993 Reconstruction of low dimensional magnetospheric dynamics by singular spectral analysis 1993; *Geophys. Res. Lett.* **20** 335–338
Subramaniam Moten 1993 Multiple time scales in rainfall variability; *Proc. Indian Acad. Sci.* (*Earth Planet. Sci.*) **102** 249–263
Ulrych T J and Bishop T N 1975 Maximum entropy spectral analysis and auto regressive decomposition; *Rev. Geophys.* **13** 183–200
Yfantis E A and Borgman L E 1981 Fast fourier transform 2-3-5; *Computers & Geosci.* **7** 99–108
Vautard R and Ghill M 1989 Singular spectrum analysis in non linear dynamics with application to palaeoclimatic time series; *Physica* **D35** 395–424