Mausam, (1987), **38**, 4, 453-458

551.577.11 : 551.501.777

Statistical analysis of weekly rainfall

R. NARAYANA IYENGAR

Centre for Atmospheric Sciences, I. I. Sci., Bangalore

(*Received 18 June 1986*)

सार — इस शोधपत्र में कर्नाटक में आठ केन्द्रों पर साप्ताहिक वर्षा के एक संख्यात्मक विश्लेषण को प्रस्तुत किया गया है। सामान्य (प्रायिक) माध्य और मानक विचलन के अतिरिक्त घातवर्णक्रम घनत्व फलनों की संगणना की गई हैं। यह पाया गया है कि मंगलूर में वर्षा की लगभग चालीस दिनों की प्रमुख अवधि हैं। इसके अलावा यह भी पाया गया है कि कुछ केन्द्र सप्ताह से सप्ताह पर सार्थक निर्भरता दर्शाते हैं।

ABSTRACT. This paper presents a statistical analysis of weekly rainfall at eight stations in Karnataka. Apart from the usual mean and standard deviations, the power spectral density functions are computed. It is found that the rainfall at Mangalore shows a dominant period of about 40 days. Further it is found that some stations show significant week to week dependence.

1. Introduction

Study of rainfall data on a weekly time scale is of interest to meteorologists, engineers and agriculturists. Routine statistical analysis easily leads to long term weekly averages (normals), standard deviations and probability distribution. While these informations are of interest, they do not give an insight into the temporal variation of the rainfall process. A study of the time-wise variations would start with auto-correlation and power spectral analysis of data as time series. These are also the initial steps in building stochastic models. However, generally the power spectrum is found under the assumption of stationarity which may not be strictly valid. One of the questions often asked about weekly rainfall is its predictability. If the auto-correlation function is significantly different from zero, one could expect the rainfall in a given week to be statistically predictable in terms of observed prior values. Even when this hope is well founded, nonstationarity and nongaussianness of the rainfall sample brings in many hurdles to a suitable prediction. Also, it is plausible that the relations between weeks are nonlinear and hence, are not well identified by the linear auto-correlation function. With these points in the background, in the present paper an exploratory study of weekly rainfall at eight stations in south India is undertaken. Previously Mooley and Appa Rao (1970, 1972) have studied the statistical distribution of rainfall. Also they have investigated independence of rainfall over time scales of 5, 10, 15, 20 and 25 days. Their major findings are (a) the rainfall is nearly gamma distributed; (b) over time scales of 5 and 10 days rainfall is pairwise dependent for consecutive periods and independent for non-consecutive periods and (c) over still longer time scales of 15, 20 and 25 days rainfall is pairwise independent.

2. The data

The data consists of weekly rainfall series at eight stations in the State of Karnataka. In Table 1 the location of the stations, duration of data, the mean annual rainfall and the mean rainfall during the southwest monsoon (June-September) season are presented. It is observed that the SW monsoon season accounts for the major rainfall of the year except in areas to the south of Bangalore. However, in areas adjoining the west and south of Bangalore, southwest monsoon rainfall is less than 50% of annual and to the south of Mysore the percentage is less than 40. Among the stations selected here, Mangalore is on the west coast, while the others are in the interior arid and semi-arid regions of Karnataka. The sub-humid stations close to the Western Ghats are not represented in the above selection.

3. Time series analysis

The weekly data is collected on the basis of 52 standard weeks for a year. In Table 2 for sake of reference, the weeks and their dates are shown. A sample time series of the data R_{ij} (i=1, 2, .., 52 wk; j=1, 2.., N yr) is shown in Fig. 1 for Mysore starting from the first week of 1901. The presence of the annual cycle and the seasonality can be easily observed in the data. The well known procedure for removing the annual cycle is to standardize the data with respect to the weekly mean and standard deviations. These quantities, namely the weekly mean m_i and weekly standard deviation s_i have been found for all the eight stations but not presented here. Now, a new data series:

$$y_{ij} = (R_{ij} - m_i)/s_i \quad (1)$$

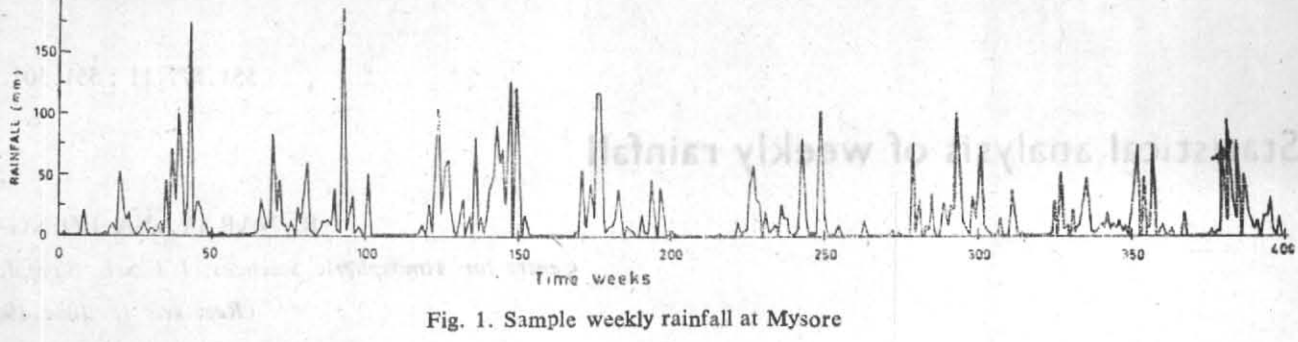
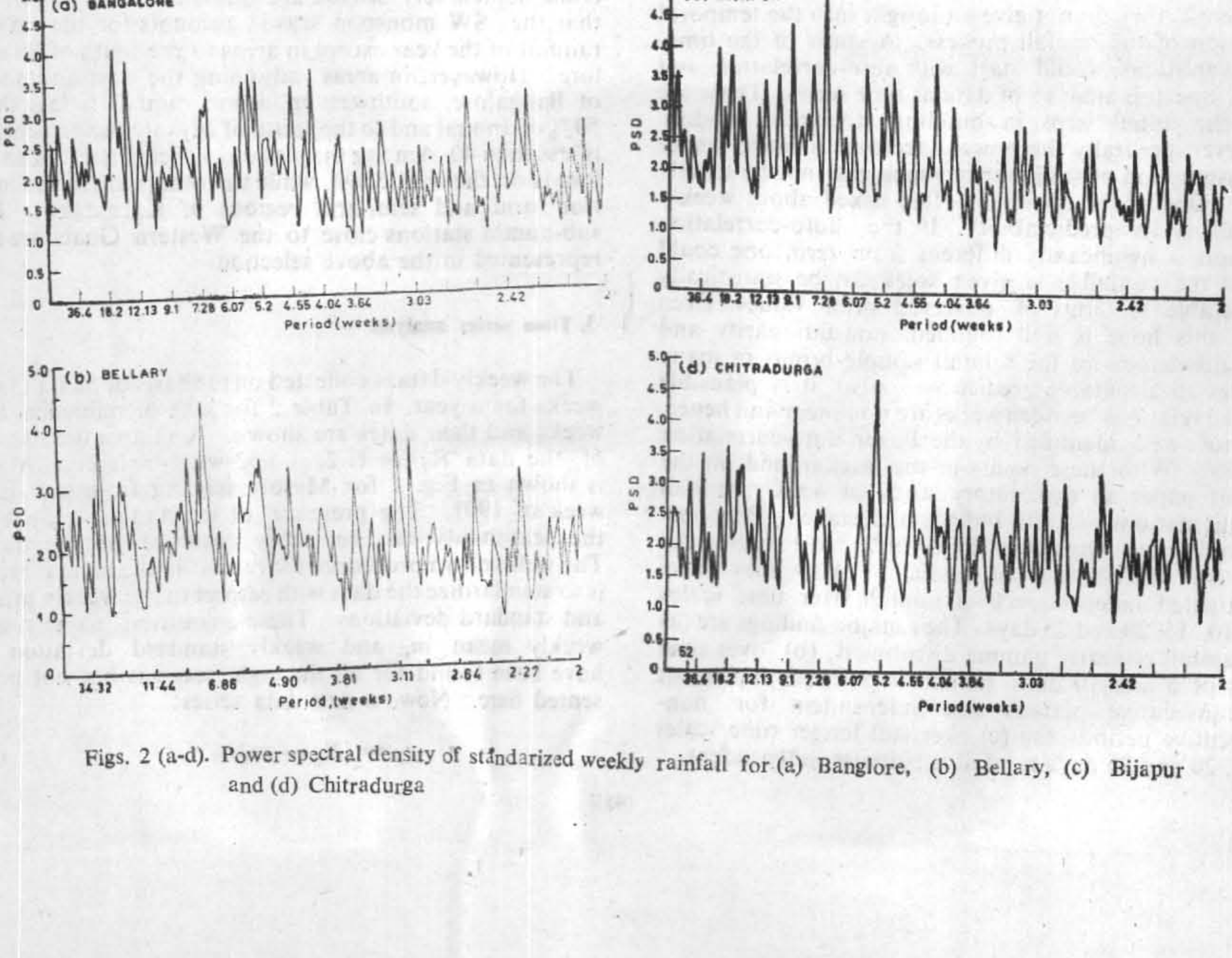


Fig. 1. Sample weekly rainfall at Mysore



Figs. 2 (a-d). Power spectral density of standarized weekly rainfall for (a) Banglore, (b) Bellary, (c) Bijapur and (d) Chitradurga

TABLE 1

Station	Location		Data period	Mean rainfall*	
	Lat. (°N)	Long. (°E)		Annual (mm)	SWM (mm)
Bangalore	12°58′	77°35′	1901-70	909.5	490.8
Bellary	15°09′	76°51′	1901-66	525.6	277.3
Bijapur	16°49′	75°43′	1901-70	586.0	379.2
Chitradurga	14°14′	76°26′	1901-70	637.9	320.1
Gulbarga	17°21′	76°51′	1901-70	763.0	584.7
Mangalore	12°52′	74°51′	1901-65	3494.3	2961.6
Mysore	12°18′	76°42′	1901-68	787.3	323.0
Raichur	16°12′	77°21′	1901-68	691.6	506.3

*Based on the period 1901-1982

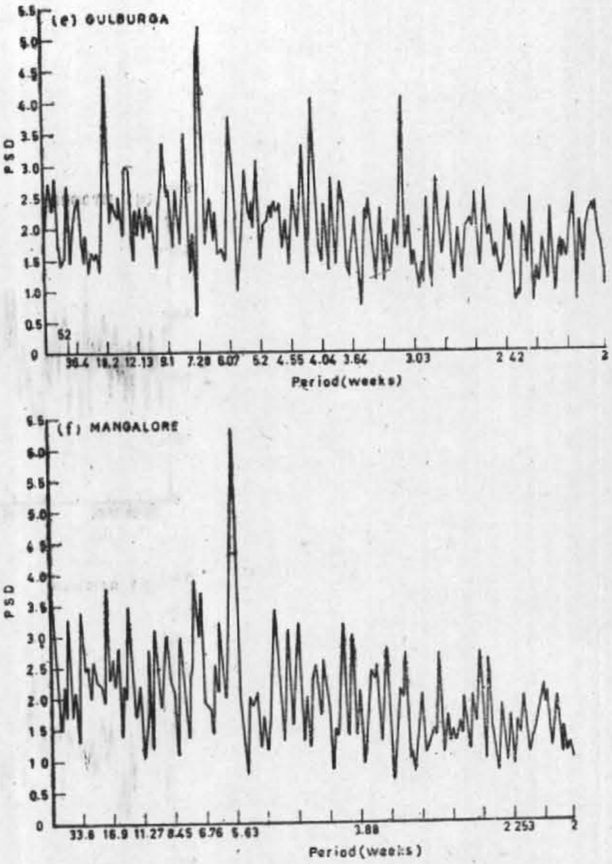
TABLE 2

The standard weeks

Week No.	Dates	Week No.	Dates
	Jan		Jul
1	1-7	27	2-8
2	8-14	28	9-15
3	15-21	29	16-22
4	22-28	30	23-29
5	29-4	31	30-5
	Feb		Aug
6	5-11	32	6-12
7	12-18	33	13-19
8	19-25	34	20-26
9	26-4*	35	27-2
	Mar		Sep
10	5-11	36	3-9
11	12-18	37	10-16
12	19-25	38	17-23
13	26-1	39	24-30
	Apr		Oct
14	2-8	40	1-7
15	9-15	41	8-14
16	16-22	42	15-21
17	23-29	43	22-28
18	30-6	44	29-4
	May		Nov
19	7-13	45	5-11
20	14-20	46	12-18
21	21-27	47	19-25
22	28-3	48	26-2
	Jun		Dec
23	4-10	49	3-9
24	11-17	50	10-16
25	18-24	51	17-23
26	25-1	52	24-31†

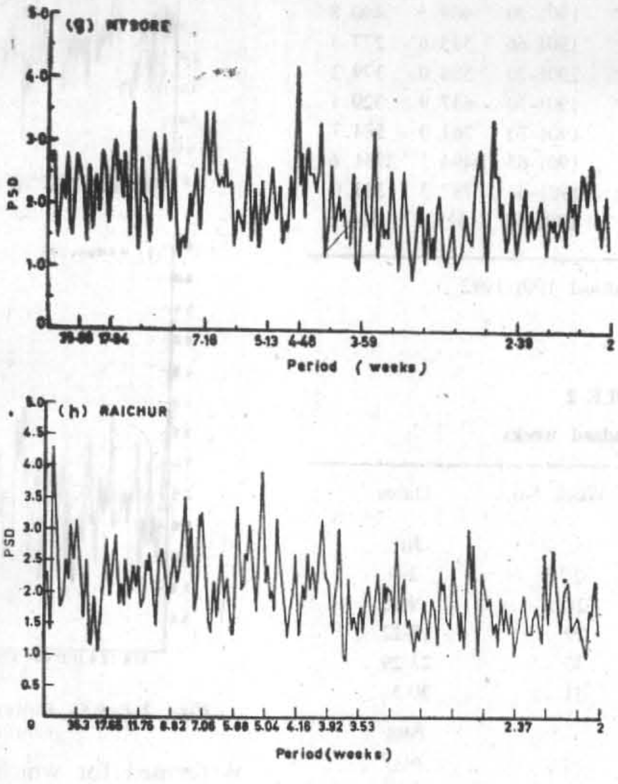
*In leap years the week No. 9 will be 26 February to 4 March, *i.e.*, 8 days instead of 7.

†Last week will have 8 days, *i.e.*, 24 to 31 December.



Figs. 2 (e & f). Power spectral density of standarized weekly rainfall for (e) Gulbarga and (f) Mangalore

is formed for which the auto-correlation and power spectral density (PSD) functions are computed. The power spectral density function of all the eight standardized data are shown in Figs. 2(a-h). It is interesting to observe that Mangalore which is a typical southwest monsoon station exhibits a dominant period of about 40 days. The spectra of other interior stations, namely Bijapur, Gulbarga and Chitradurga show a tendency to peak in the 30-50 day period interval. The presence of dominant periods may be helpful in fitting in trigonometric models which have been found to be useful forecasting aids (Dyer 1977). The interpretation of the PSD results except for the above qualitative observations is difficult. This is mainly due to the nonstationarity persisting even after the standardization and the strong non-gaussian nature of the rainfall distribution. Thus, no great significance can be attached to any peak other than inferring that much energy is carried at these frequencies. It may be noted here that inclusion of the pre and post SWM seasons in the data series has introduced much fluctuation particularly at the high frequency end of the PSD function. Notwithstanding this, the Mangalore data shows a very predominant peak. The existence of this has been also checked by dividing the data into two series of 32 and 33 years length. The PSD functions obtained have been checked for their significance by postulating the null hypothesis that the data series are white noise. This is done using the Kologorov-Smirnov statistics as suggested by Jenkins and Watts (1969). It turns out that at the 95% level of significance all the data series have to be accepted as non-white processes.



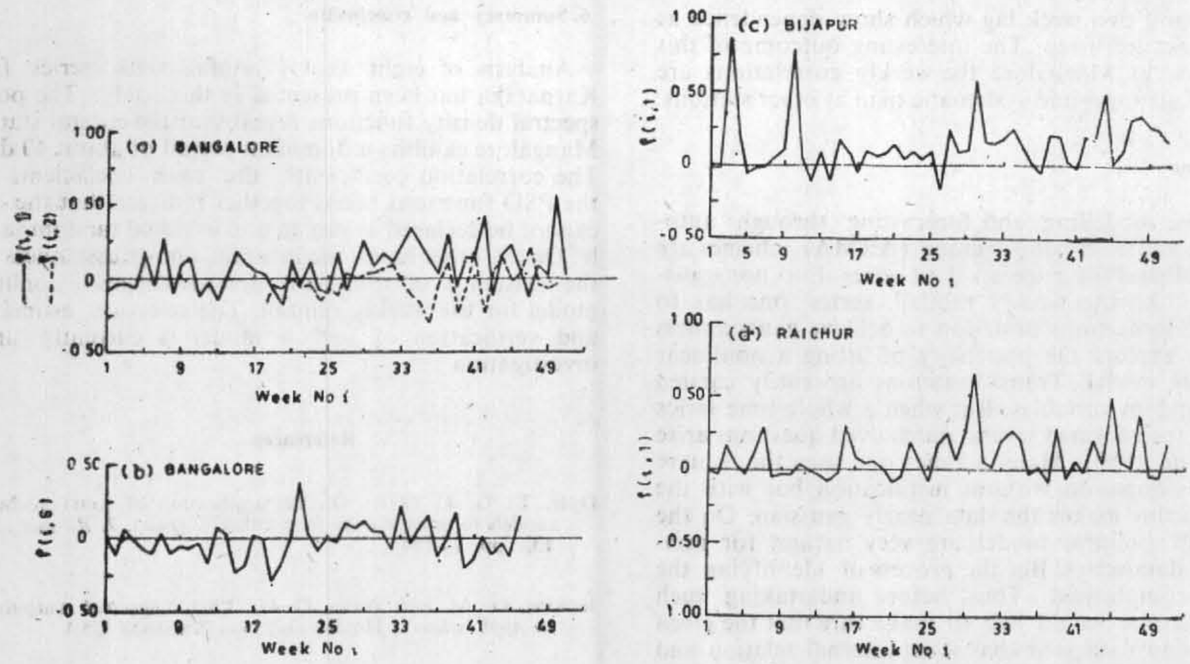
Figs. 2 (g&h). Power spectral density of standarized weekly rainfall for (g) Mysore and (h) Raichur

4. Structure of correlation

It is common practice to compute correlation coefficients for pairwise data and auto-correlation functions for time series data. These quantities refer to linear relations and hence are generally meaningful for gaussian (normal) random processes. The weekly process is not gaussian distributed and hence the correlation coefficients obtained from the usual product moment formula are difficult to interpret. However, these linear correlation coefficients could help in qualitatively establishing associations among weeks. With this in view the correlation coefficient defined as :

$$\rho\,(i,\,k) = N^{-1} \sum_{j=1}^{N} (R_{ij} - m_i)\,(R_{kj} - m_k)\,(s_i\, s_k)^{-1} \qquad (i,\,k = 1,\,2,\ldots,52) \qquad (2)$$

has been computed. A few typical ρ' s for fixed lag k are shown in Figs. 3(a-d) as functions of the week number i. A stationary process would have shown a more or less constant value of ρ with respect to i, once k is fixed. The present results on the other hand, show strong non-stationary tendencies in the data. From a quantitative angle, perhaps, the correlation at one week lag (k=1) is of interest. If a significance test, as on a normal population, is applied on $\rho(i,\,1)$ a value of about $|\rho| \geqslant 0.25$ has to be accepted as significant . It is observed that all the stations possess weeks which have significant $\rho(i,\,1)$ values. Another interesting result is that this value when significant is always positive. This is an indication of persistence at the weekly time scale. The correlation coefficients at longer lags such as $\rho\,(i,\,2)$, $\rho\,(i,\,6)$ etc also, some time show significant values, if tested under the assumption of normality. However, it must be noted that the assumption of normality is not valid for weekly rainfall. This further brings up the necessity of using other approaches for understanding associations between weeks. One possible method is to postulate that the two consecutive weekly rainfall are statistically independent. This hypothesis can be tested by the Spearman's rank test. Here this has been carried out for all the eight stations. For every pair of weeks the null hypothesis that they are independent is tested by the chi-square test at 95% level. In Table 3 the pairs of weeks



Figs. 3 (a-d). Correlation in weekly rainfall for (a) Bangalore, (b) Bangalore, (c) Bijapur and (d) Raichur

TABLE 3

Related weeks and the Spearman's rank coefficient r_s

Station	Correlated consecutive or alternate pairs of weeks and correlation coefficient								
Bangalore	(33, 34) .39	(42, 43) .36	(44, 45) .36	(45, 46) .36					
Bellary	(33, 34) .35	(44,45) .35	(45, 46) .43						
Bijapur	(19, 20) .37	(19, 21) .34	(27, 28) .36	(30, 31) .36	(32, 33) .32	(34, 35) .33	(42, 43) .41	(43, 44) .33	(46, 48) .32
Chitradurga	(34, 35) .32	(35, 36) .34	(36,38) −.38	(42, 43) .32	(45, 46) .50				
Gulbarga	(19,20) .41	(30, 31) .31	(31, 32) .32	(39, 40) .40	(42, 43) .35	(46, 47) .30			
Mangalore	(20, 21) .48	(21, 22) .31	(23, 25) −.32	(29, 30) .31	(33, 34) .39	(36, 37) .36	(37, 38) .32	(39, 40) .34	(45, 46) .35
Mysore	(18, 20) .29	(45, 46) .35							
Raichur	(18, 20) .31	(27, 28) .32	(30, 31) .34	(31, 32) .33	(39,40) .37				

with one and two week lag which show dependence as per this test are listed. The interesting outcome of this test is that at Mangalore the weekly correlations are somewhat stronger and systematic than at other stations.

5. Discussion

Analysis, modelling and forecasting through autoregressive and or moving average (ARMA) schemes are well established for gaussian data series. For non-gaussian data like the weekly rainfall series one has to either use some transformation to achieve gaussianness or has to explore the possibility of fitting a nonlinear time series model. Transformations are easily carried out on random variables. But when a whole time series has to be transformed several unresolved questions arise (Salas *et al.* 1980). Many a times one uses the square root transformation without justification but with the hope that this makes the data nearly gaussian. On the other hand nonlinear models are very natural for non-gaussian data series. But the process of identifying the model is complicated. Thus, before undertaking such an exercise one would like to make sure that the given non-gaussian data series has some internal relation and is not just an uncorrelated data set. The present study is motivated by this desire, to explore and verify whether the weekly station rainfall in Karnataka has some internal association or not. This has been carried out through the estimation of three statistics namely, the product moment correlation coefficient, the PSD function and the rank coefficient. The interpretation of all the three results can only be qualitative. However, it is observed that all the three statistics seem to indicate the existence of some stochastic internal relationship in the data series.

6. Summary and conclusion

Analysis of eight weekly rainfall time series from Karnataka has been presented in this study. The power spectral density functions reveal that the coastal station, Mangalore exhibits a dominant period of about 40 days. The correlation coefficients, the rank coefficients and the PSD functions taken together indicate that the data cannot be declared as just an uncorrelated random sample. On the other hand, the inherent non-gaussianness and the existence of dominant periods suggest a nonlinear model for the weekly rainfall. The selection, estimation and verification of such a model is currently under investigation.

References

Dyer, T. G. J., 1977, "On the application of some stochastic models to precipitation forecasting", *Quart. J. R. met. Soc.*, **13**, pp. 77-184.

Jenkins, G. M. and Watts, D. G., 1969, "*Spectral Analysis and its Applications*", Holden-Day, San Francisco, USA.

Mooley, D. A. and Appa Rao, G., 1970, "Statistical distribution of pentad rainfall over India during monsoon season", *Indian J. Met. Geophys.*, **21**, 2, pp. 219-230.

Mooley, D. A. and Appa Rao, G., 1972 "Independence of rainfall over Indian subcontinent over time scales of less than a month during monsoon season", *Indian J. Met. Geophys.*, **23**, 1, pp. 51-56.

Salas, J. D., Delleur, J. W., Yevejevich, V. and Lane, W. L., 1980, "*Applied Modelling of Hydrologic Time Series*", Water Resources Publications, Colorado, USA.