Mausam, (1992), **43**, 3,269-272

551.577.3. 551.501.777

Behaviour of spatial rainfall correlation for short distances

J. N. KANAUJIA, SURINDER KAUR and D. S. UPADHYAY

India Meteorological Department, New Delhi

(Received 18 June 1991)

सार – दो स्टेशन जो कि लघु दूरी पर स्थित है, पर रिकार्ड की गई वर्षा की दो श्रेणियों के मध्य सहसम्बन्ध प्राय: सार्थक पाया गया है। आँकड़ा अन्तर्वेशन, संजाल अभिकल्पना, अनुपलब्ध आँकड़ों के सम्बन्ध में सूचना का स्थानान्तरण और बिन्दु मान से व्युत्पन्न क्षेत्रीय वर्षा में इस सूचना की महत्वपूर्ण उपयुक्तता है। इस शोध-पत्र में भारत के लगभग 1500 स्टेशनों के लिए 70 वर्षों (1901–1970) के वार्षिक वर्षा आंकड़ों का विश्लेषण किया गया है। 40 कि० मी० की दूरी के अन्दर स्थित स्टेशनों के लिए सहसम्बन्ध गुणांक (r) के बंटन प्राप्त किए गए थे। r के सैद्धान्तिक निदर्श को व्युत्पन्न करने का प्रयत्न किया गया है। इस उद्देश्य के लिए दो बंटनों (1) दो प्राचल β बंटन और, (2) दो प्राचल परिबद्ध बंटन, क्योंकि दोनों ही स्थितियों में चर का परिसर 0 और 1 बीच है।

ABSTRACT. The correlation between two series of rainfall recorded at two stations which are at short distance, is usually found significant. This information has important applicability in the areas of data interpolation, network design, transfer of information in respect of missing data and deriving areal rainfall from point values. In this paper 70-year (1901–1970) annual rainfall data for about 1500 stations in India have been analysed. The distribution of correlation coefficient (r) for the stations located within a distance of 40 km were obtained. Attempt has been made to derive theoretical model of r. For this purpose two distributions, (1) a two parameter β—distribution and (2) a two parameter bounded distribution, have been chosen as in both cases the variable ranges from 0 to 1.

Key words — Spatial rainfall correlation, β and bounded distribution, Network design

1. Introduction

The variation of rainfall in space and in time are well known. However, it is generally expected that a direct correlation may exist between the rainfall at stations close to each other and the correlation may decrease when the distances are large. Attempts have been made by various authors in the past to establish a functional relationship between distance and rainfall correlation. Upadhyay *et al.* (1990) determined the rainfall correlation structure over the different regions of India for distances up to 600 km and established that space rainfall correlation decreases exponentially with distance and projected practical use of such an information in the field of rainfall estimation, network design, rainfall analysis and transfer of rainfall information from one station to another including interpolation. Ramanathan *et al.* (1981) used correlation structure for determining optimum network density. Rodrigues and Mejia (1974) used this correlation concept for determining error factors in estimating areal rainfall from point rainfall. Hershfield (1965) used space correlation for determining proper spacing between gauges.

The present work is a further attempt in the earlier direction to bring out exclusive behaviour of space rainfall correlation (r) for short distances (< 40 km) over plains of India excluding north eastern States. This study of correlation distribution, where high degree of positive relationship is observed, will augment information on transfer of rainfall from one station to another and interpolation of missing rainfall data in spatial series.

2. Data

70-year (1901–1970) annual rainfall data of about 1500 stations distributed over plains of India have been used for this study. The correlation coefficient between all pairs of stations within 40 km of distance were worked out.

3. Behaviour of correlation coefficient at short distances

Suppose there are 'n' stations. From the 'n' annual rainfall series recorded at these stations, ${}^nC_2$ correlations coefficients (r) can be obtained. Also, there will be ${}^nC_2$ values of distances (s) between these pairs of stations. Assuming 'r' and 's' as random variables, a bivariate frequency distribution can be worked out.

A conditional distribution with '*s*' lying between 0 and 40 km has been prepared and shown below :

Class (r)	0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0	Total
Frequency (f)	161	427	1017	1055	288	2948

It has been found out that most correlations for such short distances are positive. It is also a logical deduction if we consider the scale with which a pressure system affects over a region. Under this assumption, a few negative values of 'r' obtained during the process of calculation have been ignored. This keeps the range of 'r' between 0 and 1 instead of -1 and $+1$.

The statistical nature of this distribution is summarised below :

Mean		$= 0.56$
Standard deviation		$= 0.20$
Skewness	($\sqrt{\beta_1}$)	$= -0.34$
Kurtosis	($\beta_2 - 3$)	$= 0.62$

This distribution is unimodal having long tail towards left as apparent from definite and significant negative value of skewness.

The average value of correlation (0.56) also is highly significant and is large enough to be applied in practical cases. The Kurtosis value (0.62) indicates slight elongated peak although it does not appear to be significantly different from normal peak.

4. Statistical modelling of 'r'

With a view to enable the distribution for further algebraic treatment and to enhance its applicability, it is desirable to derive a sampling distribution of correlation coefficients under the conditions described in section (3). Since the sampling distribution is bounded between 0 and 1 the following models have been attempted.

(A) Two parameter β — distribution of Type I.

(B) Two parameter bounded distribution whose probability distribution function (p.d.f.) is given as :

$$f(r) = ab\, r^{a-1}\ (1 - r^{a})^{b-1} \tag{1}$$

where, $a \geqslant 1,\ b \geqslant 1$ and $0 \leqslant r \leqslant 1$

4.1. *Fitting of β — distribution Type I*

We have considered β — distribution with p.d.f.

$$\beta(r) = \frac{1}{\beta(a, b)}\, r^{a-1}\ (1 - r)^{b-1} \tag{2}$$

where, $a > 1$, $b > 1$ and $0 < r \leqslant 1$

The parameters a and b were chosen so as to satisfy the following 3 equations

$$\frac{a}{a + b} = \bar{r}\ (= 0.56) \tag{3}$$

$$\frac{ab}{(a + b)^2\ (a + b + 1)} = \sigma_r^2\ (= 0.0409) \tag{4}$$

and $$\frac{2(b - a)}{(a + b + 2)}\sqrt{\frac{(a + b + 1)}{ab}} = \sqrt{\beta_1}\ (= -0.34) \tag{5}$$

Solving algebraically the following values give best fit to the above equations

$a = 2.9$ and $b = 1.9$

Hence the distribution function

$$B(r) = \frac{1}{\beta(2.9,\ 1.9)} \int_0^r r^{1.9}\ (1 - r)^{0.9}\, dr \tag{6}$$

The following approximations for evaluating incomplete β—function $B(r)$ is given below as suggested by Milton *et al.* (1964) :

$$B(r) = F(y) \tag{7}$$

where $$y = \frac{3\left[w_1\left(1 - \frac{1}{9b}\right) - w_2\left(1 - \frac{1}{9a}\right)\right]}{\sqrt{\frac{w_1^2}{b} + \frac{w_2^2}{a}}}$$

and $w_1 = (br)^{\frac{1}{3}}$, $w_2 = [a\,(1 - r)]^{\frac{1}{3}}$

The procedure is applicable if $(a + b - 1)(1 - r) \geqslant 0.8$, $F(y)$ is the normal probability function given as ;

$$F(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}\, dt \tag{8}$$

The values of $F(y)$ can be obtained from standard statistical tables.

When $(a + b - 1)(1 - r) < .8$ then $B(r) = Q\left(\frac{\chi^2}{\nu}\right)$

where, $\chi^2 = (a + b - 1)(1 - r)(3 - r) - (1 - r)(b - 1)$

and $\nu = 2b$

The values of $Q\ (\chi^2/v)$ can be obtained from the standard tables.

The results are summarised in Tables 1 & 2.

TABLE 1

'r'	w_1	w_2	y	$F(y)$	Cumulative frequency [N.F.(y)]
0	0	1.4260	—4.9100	0	0
0.2	0.7243	1.3238	—1.8900	0.02939	87
0.4	0.9126	1.2027	—0.9216	0.17879	527
0.6	1.0446	1.0507	—0.8280	0.46812	1380
0.8	1.1498	0.8340	—0.8699	0.80780	2381
1.0	1.2385	0	—0.3890	0.99950	2948

TABLE 2

Class (r)	Actual frequencies		Percentage frequencies	
	Obs (O)	Exp. (E)	Obs. (O)	Exp. (E)
0-0.2	161	87	5	3
0.2-0.4	427	440	15	15
0.4-0.6	1017	853	34	29
0.6-0.8	1055	1001	36	34
0.8-1.0	288	567	10	19
Total	2948	2948	100	100

β—distribution generates expected frequencies having similar statistical features to those of observed. These expected frequencies alongwith the observed frequencies are plotted in Fig. 1. However, the χ^2—test of goodness of fit does not show that the difference between the observed and expected frequencies is insignificant. When χ^2—test is applied to percentage frequencies, the difference is insignificant. (χ^2 calculated$=6.6$; $\chi^2_{0\cdot05}=9.5$ at 4 degrees of freedom).

4.2. *Fitting of bounded distribution*

For fitting the bounded distribution suggested in section 4(B), the parameters a and b are estimated using mode, mean and standard deviation of the sample. These parameters describe the central tendency, variability and skewness of population.

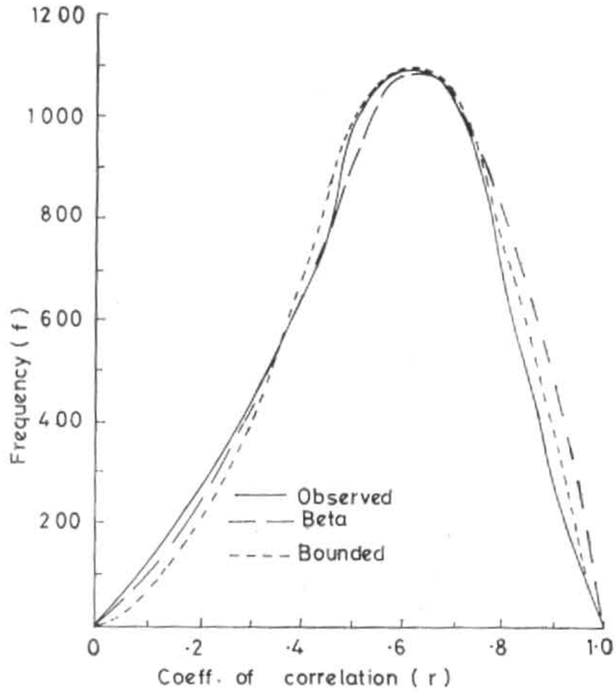

Fig. 1

$$\text{Mode} = \left(\frac{a-1}{ab-1}\right)^{\frac{1}{a}} = 0.63 \tag{9}$$

$$\text{Mean} = b\beta\left(\frac{1}{a}+1,\, b\right) = 0.56 \tag{10}$$

$$\text{Variance} = b\beta\left(\frac{2}{a}+1,\, b\right) - b^2\left[\beta\left(\frac{1}{a}+1,\, b\right)\right]^2 = 0.0409 \tag{11}$$

It may be algebraically deduced that $a=3$ and $b=3$ provide best fit to the above set of equations.

Hence, the probability density function can be expressed as :

$$f(r) = 9r^2\,(1-r^3)^2$$

$$\text{or}\quad f(r) = 9(r^2 - 2r^5 + r^8) \tag{12}$$

$$\text{It can be shown that} \int_0^1 f(r)\,dr = 1 \tag{13}$$

The distribution function

$$F(r) = r^9 - 3r^6 + 3r^3 \quad \text{where, } 0 \leqslant r \leqslant 1 \tag{14}$$

from which expected frequencies can be calculated. The results are presented in Table 3.

TABLE 3

Class (r)	$F(r)$	Cumulative frequencies	Actual frequencies		Percentage frequencies	
			Obs (O)	Exp (E)	Obs (O)	Exp (E)
0.0-0.2	0.0238	70	161	70	5	2
0.2-0.4	0.1799	531	427	461	15	16
0.4-0.6	0.5181	1527	1017	996	34	34
0.6-0.8	0.8838	2605	1055	1078	36	36
0.8-1.0	1.0000	2948	288	343	10	12
Total			2948	2948	100	100

These expected frequencies are plotted in Fig. 1.

If we apply χ^2—test on actual frequencies by pooling the first and last frequencies with those of adjacent classes.

χ^2 calculated$=10.2$ against $\chi^2_{.01}=9.2$ at 2-degree of freedom.

This shows that expected frequencies are marginally different from observed ones at 1% probability level. However, when χ^2-test is applied on percentage frequencies, χ^2 - calculated$=3.2$ as against $\chi^2_{.01}=13.3$ showing an insignificant difference.

From Fig. 1 it can be seen that the bounded distribution gives a better fit to the observed distribution of 'r' than the two parameter β—distribution of Type I.

5. Applications

Once we obtain a probability density function of correlation coefficient 'r' we can generate theoretical frequencies and moments of any order describing completely statistical behaviour of the correlation pattern. In the present case the moment about the origin is given by :

For bounded distribution :

$$\mu'_k = b\,\beta\left(\frac{k}{a}+1,\, b\right) \tag{15}$$

and for β—distribution of Type I

$$\mu'_k = \frac{\beta(k+a,\, b)}{\beta(a,\, b)} \tag{16}$$

These results will facilitate the application of space correlation structure in rainfall series in various fields. Some are provided below :

(a) *The information contents regarding population mean* (μ)

It can be deduced that sampling variance of $\bar{r}$ computed from 'n' set of observations, is given by

$$V(\bar{r}) = \frac{\sigma^2}{n}[1+(n-1)\,\bar{r}\,]$$

The information content regarding μ is given by

$$I_\mu = [1+(n-1)\,\bar{r}\,]^{-1}$$

(b) *Network design*

If we are estimating the areal rainfall using 'n' points of observations in respect of one rainfall event, the variance of estimate is reduced to :

$$V(P_A) = \frac{\sigma^2}{n}(1-\bar{r})$$

where, σ^2 is the population variance for point rainfall process. A network design scheme should aim at determining the network density 'n' for minimum acceptable variance. The cost of maintaining n —stations will also come into considerations.

(c) *Areal to point rainfall ratio*

Many researchers (Upadhyay *et al.* 1990) believe that P_A/P_O varies with k^{th} root of $\bar{r}$. k may be 2 or 3 or even more. Therefore,

$$\frac{P_A}{P_O} \propto (\bar{r})^{1/k}$$

It may be seen that in all the applications indicated above, r plays a critical role. The theoretical model not only facilitates computation of $\bar{r}$ easily, but also provides for the sampling variance of $\bar{r}$ yielding an interval estimate. These interval estimates provide flexible applications in interpolation, estimation of missing observations and evaluation of forecast accuracy.

References

Hershfield, D.M., 1965, "On the spacing of rainguages" Symp. on design of Hydrol. Networks, Int. Assoc. Sci. Hyd. Publ. 67, pp. 72-75.

Milton *et al.* 1964, *Handbook of Math. fns. NBS,* Publ., New York, pp. 945.

Ramanathan, A.S. *et al.*, 1981, "Precipitation network design for hydrological purposes," Part I, IMD, Pre publ. Sci. Rep.

Rodrigues—Iturbe, I. and Mejia, J.M., 1974, "The design of rainfall network in time and space" *Water Resour. Res.*, **10**, pp. 713-728.

Upadhyay, D.S. *et al.*, 1984, "The depth area relationship a theoretical approach," Proc. of Sem. on Hyd., Pune.

Upadhyay, D.S. *et al.*, 1990, "Space correlation structure of rainfall over India", *Mausam*, **41**, 4, pp 523-530.