UDC 551.509.331:551.577.35:551.553.21"322"(59)

INDEPENDENCE OF MONTHLY AND BIMONTHLY RAINFALL OVER SOUTHEAST ASIA DURING THE SUMMER MONSOON SEASON [1]

D. A. MOOLEY

Institute of Tropical Meteorology, Poona, India

ABSTRACT

Independence of monthly and bimonthly rainfall has been investigated for 39 well-distributed and long-record stations in the field of the Asian summer monsoon. The study reveals: (1) monthly rainfall is pairwise independent but is not tripletwise or quadrupletwise independent and (2) rainfall during the first half of the summer monsoon season is independent of rainfall during the second half. The implications of these results are discussed.

1. INTRODUCTION

If two or more random variables are to be combined, it is essential to know beforehand whether these variables are collectively independent. When rainfall distribution parameters for each of the 2 mo are known and rainfall in 1 mo is known to be independent of rainfall in the other month, then the parameters of the combined 2-mo rainfall distribution can be obtained through inversion of the characteristic function of the combined variable, provided this inversion is feasible. Likewise, from rainfall probabilities for each of the 2 mo, joint rainfall probabilities can be easily obtained if the two variables are independent. Variance assumes the additive property if the variables are pairwise independent. In view of these considerations, it is necessary to investigate the independence of rainfall. In the present paper, independence of monthly rainfall as well as independence of rainfall during the first half and the second half of the summer monsoon season have been investigated. Hereafter, the word "monsoon" will be used in place of summer monsoon season, June to September inclusive.

2. CRITERIA FOR INDEPENDENCE

Kolmogorov (1956) has stated that the necessary and sufficient condition for mutual or collective independence of n events, $A_1, A_2, \ldots, A_n$ is

$$P(A_{i_1}A_{i_2}\ldots A_{i_m})=P(A_{i_1})P(A_{i_2})\ldots P(A_{i_m})=\prod_{i_m=1}^{m} P(A_{i_m}) \tag{1}$$

for any positive integer $i_m \leq n$.

This condition includes $\binom{n}{2}, \binom{n}{3}, \ldots, \binom{n}{n}$ equations for all pairs, triplets, . . ., n-tuplets, respectively, and thus covers 2^n-n-1 equations that are mutually independent. When all these equations are satisfied, mutual or collective or n-tupletwise independence is established. If equations for all pairs hold good, pairwise independence is established. If, in addition, equations for all triplets hold good, tripletwise independence is established, and so on. Given n-tupletwise independence, $(n-1)$-tupletwise independence follows; but $(n+1)$-tupletwise independence does not follow.

The notion of independence of events has been extended to variables. As shown by Gnedenko (1964), the necessary and sufficient conditions for collective independence of n variables are

$$F(x_1, x_2, \ldots, x_m)=F_1(x_1)F_2(x_2)\ldots F_m(x_m)=\prod_{k=1}^{m} F_k(x_k) \tag{2}$$

and

$$P(x_1\ x_2 \ldots x_m)=P_1(x_1)P_2(x_2)\ldots P_m(x_m)=\prod_{k=1}^{m} P_k(x_k) \tag{3}$$

for every positive integer $m \leq n$, where $F_k(x_k)$ and $P_k(x_k)$ are, respectively, distribution function and probability density function for the variable ξ_k. Equation (3) is subject to the condition that the probability density functions exist.

For two normally distributed variables, the necessary and sufficient condition for independence is that their product-moment correlation coefficient is zero. Distribution of rainfall being nonnormal, this simple condition cannot be used for testing independence of rainfall.

Mooley and Appa Rao (1970) have investigated the independence of 5-, 10-, 15-, 20-, and 25-day rainfall at a number of well-distributed stations over the Indian subcontinent during the monsoon. They find that rainfall over the subcontinent over the time scales of 15, 20, and 25 days is pairwise independent. It may be mentioned that pairwise independence of monthly rainfall does not follow from pairwise independence of 15-day rainfall, although it does follow as a corollary that pairwise linear correlation between monthly rainfall is zero as can be shown below.

Let ζ_1, ζ_2, ζ_3, and ζ_4 be rainfall variables for successive 15-day periods. Pairwise independence implies

$$E(\delta\zeta_1\ \delta\zeta_3)=0;\ E(\delta\zeta_2\ \delta\zeta_4)=0;\ E(\delta\zeta_2\ \delta\zeta_3)=0;\ E(\delta\zeta_1\ \delta\zeta_4)=0.$$

When adding these,

$$E[\{\delta(\zeta_1+\zeta_2)\}\{\delta(\zeta_3+\zeta_4)\}]=0,$$

[1] A part of the results of this study has been communicated to the Symposium on Tropical Meteorology in Hawaii, June 1970.

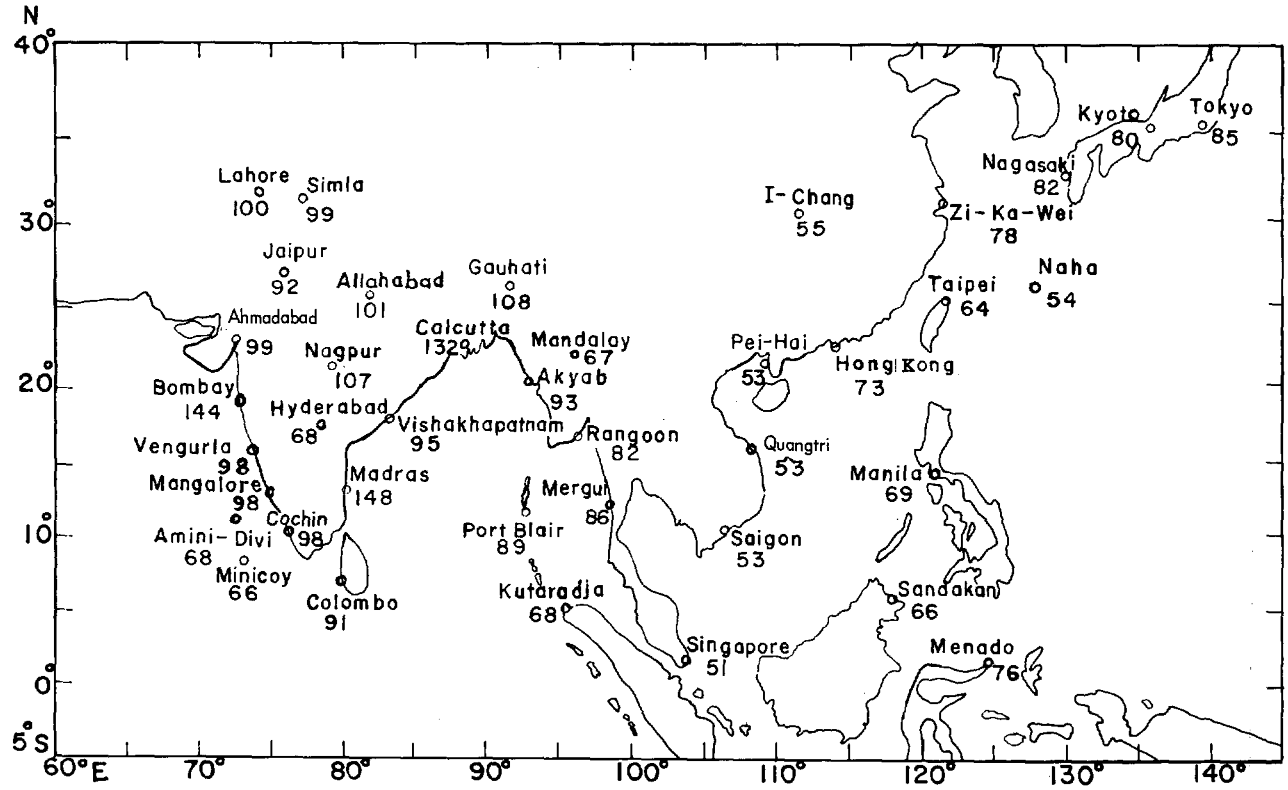

FIGURE 1.—Network of rain gage stations; N, the number of years of rainfall data, is given below the station name.

E being the expected value operator and δ being the differential operator. Hence the product-moment correlation coefficient between $(\zeta_1+\zeta_2)$ and $(\zeta_3+\zeta_4)$ is zero.

As pairwise independence of 15-day rainfall does not imply pairwise independence of monthly rainfall, it is necessary to investigate the pairwise independence of monthly rainfall.

3. RAINFALL DATA

The network of rain gage stations is shown in figure 1. All available data up to 1960 were collected for each of the monsoon months, June to September inclusive, for these stations from the *World Weather Records* of the Environmental Data Service, NOAA. For Singapore alone, data up to 1967 were used to get a rainfall record exceeding 50 yr. Data for Singapore for 1951 to 1967 and for Sandakan, Kutaradja, Menado, and Manila for 1951 to 1960 were obtained from the concerned meteorological services. The number of years of data used for each station has been indicated in figure 1.

4. PROCEDURE FOR INVESTIGATION OF INDEPENDENCE

Let ξ_1, ξ_2, ξ_3, and ξ_4 denote rainfall variables for June, July, August, and September, respectively, at a station. Monsoon rainfall ξ at a place then may be considered as a four-dimensional variable. Let the range on each dimension be divided into k intervals by $k-1$ partitions. For convenience, $k-1$ partitions are taken in such a way that the range on each dimension is divided into k equal-probability intervals. In this way, four dimensional sample space is divided into k^4 cells of a four-dimensional contingency table. Considering any two and any three of the four variables and following the procedure of partitioning as already mentioned, we obtain six two-dimensional and four three-dimensional contingency tables. Thus, there are in all 11 contingency tables; and each of these is to be tested for independence by the chi-square test. Theoretical cell frequencies were computed on the basis of null hypothesis of independence by using eq (3), and the corresponding empirical frequencies were counted from rainfall data. If independence is obtained in all these 11 contingency tables, quadrupletwise independence of monthly rainfall is established. If only the two-dimensional and three-dimensional contingency tables are independent, tripletwise independence is established. However, if independence is found in two-dimensional contingency tables only, pairwise independence is established.

If N is the total number of rainfall observations, then the expected number of observations in a cell in a two-dimensional table is N/k^2. This number in the case of three-dimensional and four-dimensional tables is N/k^3 and N/k^4, respectively. Taking N to be 100, we find that k can be taken as 3 or 4 in the case of two-dimensional tables and 2 only in the case of three-dimensional and four-dimensional contingency tables, so that the theoretical cell frequency for most stations would not fall below 5, the generally accepted lower limit for application of the chi-square test. In the present study, however, this limit was lowered to 4. Craddock (1966) conducted a Monte Carlo study to determine at what stage the approximation of the distribution of the chi-square statistic for a 3×3 contingency table to the chi-square distribution for appropriate degrees of freedom loses validity. The results of this study show that, with the expected frequency of 3 in each of the nine cells, the values of the chi-square statistic still conform to the chi-square distribution and that, if the total number of observations in the contingency table is reduced below 25 or so, the distribution of the chi-square statistic does begin to show systematic departures from the chi-square distribution. On this basis, it is felt that, in the present study, lowering of the limit from 5 to 4 will not affect the results.

In a similar manner, independence of variables η_1 $(=\xi_1+\xi_2)$ and η_2 $(=\xi_3+\xi_4)$, which is rainfall in the first half and second half of the monsoon, has been investigated. In the case of two-dimensional contingency tables, partitioning has been done by terciles and quartiles and, in that of three- and four-dimensional contingency tables, by the median.

Regarding the degrees of freedom of the chi-square value computed for a contingency table, it may be mentioned that, once we obtain four of the expected nine-cell frequencies of a 3 × 3 table, we can deduce the remaining five expected cell frequencies since the marginal totals are known. Thus the number of degrees of freedom for a 3 × 3 table is 4. The general formula for number of degrees of freedom for an $r \times c$ table is $(r-1)$ $(c-1)$ where r is the number of rows and c, the number of columns. For 2 × 2 × 2 and 2 × 2 × 2 × 2 tables, there is only one degree of freedom since, on obtaining one of the expected cell frequencies, the remaining expected cell frequencies can be deduced.

5. RESULTS

MONTHLY RAINFALL

Table 1 gives an analysis of chi-square values obtained in testing the independence of all two-dimensional contingency tables when partitioning was done by terciles. This gives the frequencies for different ranges of $P(\chi^2 \geq \chi_1^2)$ where χ_1^2 is the actual value obtained for the contingency table for each pair of months for each of the stations. Table 2 gives the details of the 14 cases significant at the 5-percent level from table 1. Tables 3 and 4 give an analysis of chi-square values when partitioning is by quartiles.

When considering tables 2 and 4 together, it is seen that chi-square values for only the station Menado for the monthly pairs (June, August; July, September; and August, September) are consistently significant at the 5-percent level. Thus, the null hypothesis of independence is not contradicted. Even when considering tables 1 and 3 separately, the number of values significant at the 5-percent level is generally not different from that which would be expected by chance. In the case of the monthly

TABLE 1.—*Independence of monthly rainfall; partition by terciles; degrees of freedom=4.*

Monthly pair	Frequency for different ranges of $P(\chi^2 \geq \chi_1^2)$							Total
	<0.01	≥0.01 but <0.05	≥0.05 but <0.10	≥0.10 but <0.25	≥0.25 but <0.50	≥0.50 but <0.75	≥0.75	
June, July	1	2	2	7	7	7	13	39
June, Aug.	0	5	4	5	10	9	6	39
June, Sept.	0	1	1	4	12	11	10	39
July, Aug.	1	0	3	7	8	12	8	39
July, Sept.	1	0	3	9	8	8	10	39
Aug., Sept.	1	2	2	5	14	9	6	39
Total	4	10	15	37	59	56	53	234

TABLE 2.—*Details of significant chi-square values; degrees of freedom=4.*

Station	χ^2 values for monthly pair					
	June, July	June, Aug.	June, Sept.	July, Aug.	July, Sept.	Aug., Sept.
Ahmadabad		12.36*				
Akyab		13.16*				14.13†
Colombo				13.74†		10.97*
Hong Kong		11.70*				
Kyoto	15.63†					
Menado		12.58*			14.00†	11.16*
Naha	12.67*					
Simla	10.73*		12.36*			
Vishakhapatnam		10.63*				

*Significance at 5-percent level
†Significance at 1-percent level

TABLE 3.—*Pairwise independence of monthly rainfall; partition by quartiles; degrees of freedom=9.*

Monthly pair	Frequency for different ranges of $P(\chi^2 \geq \chi_1^2)$							Total
	<0.01	≥0.01 but <0.05	≥0.05 but <0.10	≥0.10 but <0.25	≥0.25 but <0.50	≥0.50 but <0.75	≥0.75	
June, July	0	1	1	6	9	10	6	33
June, Aug.	0	3	1	9	7	10	3	33
June, Sept.	0	1	2	7	9	5	9	33
July, Aug.	0	2	2	8	7	10	4	33
July, Sept.	0	3	2	2	6	10	10	33
Aug., Sept.	0	1	3	7	11	7	4	33
Total	0	11	11	39	49	52	36	198

TABLE 4.—*Details of significant chi-square values; degrees of freedom=9.*

Station	χ^2 values for monthly pair					
	June, July	June, Aug.	June, Sept.	July, Aug.	July, Sept.	Aug., Sept.
Allahabad		20.19*				
Mangalore		15.59*			21.18*	
Menado		17.05*			19.16*	18.74*
Nagasaki				19.07*		
Port Blair					18.69*	
Taipei	20.50*					
Vishakhapatnam			17.34*	20.03*		

pair June-August, the percentage of significant chi-square values is slightly higher than that which would be expected by chance; but even in this case, the situation is such that it does not appear to be possible to reject straightway the hypothesis of independence in view of the fact that, for this monthly pair, no value is significant at the 1-percent level. It can thus be seen that monthly rainfall over southeast Asia during the monsoon is pairwise independent. The finding by Mooley and Crutcher (1968) that there is no significant linear correlation between pairs of monthly rainfall at Indian stations during the monsoon is consistent with the result obtained in the present study. Daily rainfall amounts are not independent. Persistence is an important feature of the daily observations. The correlation coefficient with lag one is quite significant, and it falls off with increasing lag. As the time scale of rainfall is increased, persistence falls off; and rainfall tends toward independence.

Table 5 clearly indicates that the null hypothesis of tripletwise independence is contradicted. In about 50 percent of the cases, the chi-square is significant at the 5-percent level; and in about 20 percent of the cases, it is significant at the 1-percent level. Thus, monthly rainfall is not tripletwise independent. From table 5, however, one is unable to determine those stations for which null hypothesis of tripletwise independence holds. Details of such nine cases (of which one is marginal) are given in table 6. This table suggests tripletwise independence of rainfall over and near the southern half of the West Coast of India.

Tripletwise nonindependence of monthly rain implies quadrupletwise nonindependence. However, it would be of interest to know whether monthly rain at the nine stations mentioned in table 6 is quadrupletwise independent. Instead of testing for quadrupletwise independence of monthly rain at these stations only, the testing procedure was applied to all the stations. The results are presented in table 7. Portion (A) of this table covers all stations for which $N \geq 64$, while portion (B) covers the remaining stations. In the latter case, the theoretical cell frequency lies between 3 and 4. Portion (B) of table 7 has been given since rainfall at three of these stations shows tripletwise independence, and one would like to know whether rainfall at these stations suggests quadrupletwise independence. Table 7 shows very clearly that the null hypothesis of independence is rejected. Monthly rainfall at none of the 39 stations is quadrupletwise independent.

TABLE 5.—*Tripletwise independence of monthly rainfall; partition by median; degrees of freedom=1.*

Triplet	Frequency for different ranges of $P(\chi^2 \geq \chi_1^2)$							Total
	<0.01	≥0.01 but <0.05	≥0.05 but <0.10	≥0.10 but <0.25	≥0.25 but <0.50	≥0.50 but <0.75	≥0.75	
June, July, Aug.	10	10	4	12	3	0	0	39
June, July, Sept.	5	13	9	7	4	1	0	39
June, Aug., Sept.	6	16	11	5	1	0	0	39
July, Aug., Sept.	8	8	9	12	2	0	0	39
Total	29	47	33	36	10	1	0	156

TABLE 6.—*Details of cases in which the chi-square is nonsignificant for all triplets*

Station	χ^2 for triplets			
	June, July, Aug.	June, July. Sept.	June, Aug., Sept.	July, Aug., Sept.
Cochin	1.92	2.57	3.06	1.75
Madras	1.62	1.62	1.62	3.57
Mangalore	2.90	2.24	1.73	3.71
Minicoy	2.61	2.36	1.64	2.36
Nagasaki	2.68	1.32	2.29	1.71
Pei-Hai	1.49	3.30	3.00	3.00
Saigon	1.19	2.40	3.60	2.40
Singapore	0.61	2.80	1.23	1.86
Vengurla‡	1.59	3.22	2.73	3.88*

‡Marginal case

TABLE 7.—*Quadrupletwise independence of monthly rainfall; partition by median; degrees of freedom=1.*

Station	χ^2	Station	χ^2	Station	χ^2
		A			
Ahmadabad	13.0†	Jaipur	20.0†	Nagasaki	5.8*
Akyab	9.0†	Kutaradja	7.8†	Nagpur	12.0†
Allahabad	9.0†	Kyoto	11.2†	Port Blair	6.8†
Amini-Divi	13.9†	Lahore	6.6*	Rangoon	11.3†
Bombay	22.0†	Madras	7.2†	Sandakan	10.6†
Calcutta	15.9†	Mandalay	11.1†	Simla	10.0†
Colombo	15.9†	Mangalore	7.1†	Taipei	16.5†
Cochin	5.5*	Manila	14.0†	Tokyo	11.5†
Gauhati	17.6†	Menado	35.2†	Vengulra	9.4†
Hong Kong	7.8†	Mergui	13.7†	Vishakhapatnam	16.4†
Hyderabad	10.1†	Minicoy	6.2*	Zi-Ka-Wei	9.0†
		B			
I-Chang	17.9†	Pei-Hai	10.0†	Saigon	9.0†
Singapore	4.4*	Naha	11.2†	Quangtri	12.9†

RAINFALL DURING THE FIRST HALF AND SECOND HALF OF THE MONSOON

As mentioned toward the end of section 2, it is necessary to investigate the independence of rainfall during the first half and second half of the monsoon season unless the two rainfall distributions become normal. As monthly rainfall distribution at most places in the field of the summer monsoon departs substantially from normal, bimonthly rainfall is not expected to be normal.

Tables 8 and 10 give analyses of the tests in situations when partition is by terciles and quartiles, respectively; and tables 9 and 11 are the corresponding tables giving details of significant chi-square values. Tables 8 and 10 clearly show that the percentages of cases when the chi-square is significant at the 5-percent level are not different from that which would be expected by chance. In addition, tables 9 and 11 show that, in respect to two stations

TABLE 8.—*Pairwise independence of rainfall during the first half and the second half of the monsoon; partition by terciles; degrees of freedom=4.*

Pair	Frequency for different ranges of $P(\chi^2 \geq \chi_1^2)$							Total
	<0.1	≥0.01 but <0.05	≥0.05 but <0.10	≥0.10 but <0.25	≥0.25 but <0.50	≥0.50 but <0.75	≥0.75	
First half and second half of monsoon	1	2	2	6	13	8	7	39

TABLE 9.—*Details of significant chi-square values*

Station	χ^2
Amini-Divi	12.20*
Menado	16.37†
Vishakhapatnam	9.68*

TABLE 10.—*Pairwise independence of rainfall during the first half and the second half of the monsoon; partition by quartiles; degrees of freedom=9.*

Pair	Frequency for different ranges of $P(\chi^2 \geq \chi_1^2)$							Total
	<0.01	≥0.01 but <0.05	≥0.05 but <0.10	≥0.10 but <0.25	≥0.25 but <0.50	≥0.50 but <0.75	≥0.75	
First half and second half of monsoon	1	2	5	5	5	7	8	33

TABLE 11.—*Details of significant chi-square values*

Station	χ^2
Amini-Divi	19.06*
Madras	18.27*
Menado	29.75†

only (viz, Amini-Divi and Menado), values are consistently significant. Thus, the null hypothesis of independence is not contradicted; and rainfall in the first half of the monsoon season can be taken to be independent of rainfall in the second half of the monsoon season.

6. CONCLUSIONS

1. The monthly rainfall over southeast Asia during the monsoon is pairwise independent, but the same is not tripletwise independent. Hence, if distribution functions and rainfall probabilities are available for individual months, it is possible to utilize these to obtain the distribution function (via the characteristic function) for any 2-mo rainfall and joint rainfall probabilities for any 2 mo. However, this monthly information on the distribution function and rainfall probabilities cannot be utilized to obtain similar information for more than a 2-mo period. Variance of rainfall for 2 or more months can be obtained by adding the variance of rainfall for individual months.

2. Rainfall in the first half of the summer monsoon is independent of rainfall during the second half. There is no justification for the belief that excess or deficit of rainfall in the first half will be respectively wiped off or made up during the second half of the summer monsoon season.

ACKNOWLEDGMENTS

The author wishes to express his sincere thanks to the Directors of the Meteorological Services of Indonesia, Malaysia, and Philippines for the supply of rainfall data and to the Director General of Observatories, New Delhi, for permission to publish this paper.

REFERENCES

Craddock, J. M., "Testing the Significance of 3 times 3 Contingency Table," *The Statistician*, Vol. 16, No. 1, London, England, **1966**, pp. 87–94.

Gnedenko, Boris V., *Theory of Probability*, Chelsea Publishing Co., New York, N.Y., **1964**, **471** pp.

Kolmogorov, Andrei N., *Foundations of Theory of Probability*, 2d English Edition, Chelsea Publishing Co., New York, N.Y., **1956**, 84 pp.

Mooley, D. A., and Appa Rao, G., "Independence of Rainfall Over Indian Sub-Continent Over Time Scales of Less Than a Month During Monsoon Season," *India Meteorological Department Scientific Report* No. 129, New Delhi, **1970**, pp. 1–11.

Mooley, D. A., and Crutcher, H. L., "An Application of Gamma Distribution Function to Indian Rainfall," *ESSA Technical Report* EDS 5, Environmental Data Service, Silver Spring, Md., Aug. **1968**, **47** pp.

[*Received July 9, 1970; revised October 8, 1970*]