Mausam, (1982), 33, 1, 65-72

511.54

A study of the errors in the areal mean rainfall arising out of missing observations

D. A. MOOLEY and A. A. MUNOT

Indian Institute of Tropical Meteorology, Pune

(*Received 23 March 1981*)

सार-इस लेख में प्रेक्षणों के छूट जाने से क्षेत्रीय वर्षा के माध्य की त्रुटियों का प्रायोगिक अध्ययन किया गया है। इसके लिए विदर्भ के अमरावती जिले को चुना गया है जहां वर्षा मापियों का अच्छा खासा संजाल है और अवधि माह जून से सितम्बर तथा मानसून ऋतु रखी गई है। प्रत्येक वर्ष में जब कोई एक, दो, तीन, चार या पांच स्टेशन छूटने की ऐसी 5 परिस्थितियों में किसी क्षेत्र की वर्षा को ज्ञात करने में होने वाली त्रुटियों का अलग अलग परीक्षण किया गया है। प्रतिशत आपेक्षिक त्रुटि का आयत चित्र दिया गया है और प्रतिशत आपेक्षिक त्रुटि के अंत मानों को निकाला गया है एवं त्रुटि बंटन के दशमक की सीमाओं को ज्ञात किया गया है। प्रत्येक परिस्थिति में एक, दो, तीन, चार या पांच स्टेशनों के यादृच्छिक रूप से छूट जाने की हालत में प्रतिशत वर्ग माध्य मूल त्रुटि की विशिष्ट मानसून माह, जुलाई और मानसून ऋतु के लिए स्थान और समय के संदर्भ में गणना की गई है। इसी प्रकार जुलाई में एक से पांच तक प्रेक्षणों के छूट जाने पर 2.8 से 8.8 प्रतिशत तक और ऋतु में 2.2 से 7.0 प्रतिशत तक परिवर्तन पाया गया। अन्तत: जब छूटे हुए स्टेशन (शून्य से पांच) यादृच्छिक हों और स्थान एवं समय के संदर्भ में स्टेशनों का छूट जाना यादृच्छिक होने की परिस्थितियों में प्रतिशत वर्ग माध्य मूल त्रुटि की गणना की गई है विशिष्ट मानसून माह जुलाई में यह त्रुटि 6.2 प्रतिशत तथा मानसून ऋतु में 5.6 प्रतिशत पाई गई है।

ABSTRACT. This is an experimental study of the mean errors in the areal rainfall in situations of missing observations. The area selected is Amravati district in Vidarbha having a very good network of raingauges and the periods considered are the months June to September and the monsoon season. Errors in the areal rainfall have been examined separately in each of the five situations when any one, two, three, four or five stations are missing in each year, and histograms of the percentage relative error have been given, and extreme values of the percentage relative error have been brought out. Limits on the deciles of the error distribution have been obtained. Percentage Root Mean Square Error (RMSE) has also been computed for the typical monsoon month, July, and for the monsoon season for each of the situations of one, two, three, four and five stations missing randomly in space and time; the same for July is found to vary from 2.8 to 8.8 per cent as the number of missing observations varies from one to five, and for the season 2.2 to 7.0 per cent. Finally the percentage RMSE has been computed in a situation when the number of missing stations (zero to five) is random and the missing of stations is random in space and time; this error for the typical monsoon month, July, is 6.2 per cent and for the monsoon season, 5.6 per cent.

1. Introduction

For a planning of the water resources and for a proper appraisal of the occurrence and spatial distribution of rainfall a good network of raingauges and a complete record of all the gauges for a reasonably long period are essential. The problem of minimum network of raingauges necessary for a proper appraisal of the occurrence of rainfall during the pre-monsoon, monsoon and post-monsoon seasons was examined by Rai Sircar and Hariharan (1954). They considered daily rainfall occurrence over an area of 80 km radius (about 20,000 sq km) around Calcutta, Delhi, Bangalore and Tiruchirapalli during the period 1941-45. The number of raingauge station varied from 27 to 50. They found that the occurrence

Fig. 1. Amravati district and the raingauge stations

or non-occurrence of rainfall on any day could not be adequately described for the area on the basis of the central station alone and that a certain minimum network was necessary for adequately describing the occurrence or non-occurrence of rainfall. The study was conducted for the purpose of examining the existing criteria for the assessment of rainfall forecasts for the small areas included in the *Farmers' Weather Bulletins*. Hariharan (1956 a) extended this study to areas of the same size around Ahmedabad, Ahmedanagar, Bijapur and Nizamabad where the number of raingauges varied from 11 to 25; however, the study was confined to the monsoon season only. In another study, Hariharan (1956 b) examined the spatial distribution of daily rainfall over the then meteorological sub-division of south Hyderabad (presently, part of north interior Karnataka and part of Telangana) during the monsoon seasons of the years 1941-43 on the basis of data from 6 observatories and from 23 raingauges. The categories of the spatial distributions considered were widespread, fairly widespread, local, scattered and no rain. In addition, he considered six intensities of the mean daily rainfall for the sub-division on the basis of the data from the observatories as well as the data from the raingauges. Through contingency tables he brought out the agreement between the two sets of networks in respect of the spatial distribution and the rainfall intensity. The agreement was found to be 57 and 50 per cent respectively.

Bhalla and Narayanan (1978) have examined the question of minimum network for representative areal rainfall over a district for weekly and fortnightly periods during the monsoon season, using multiple correlation approach. They used rainfall data for the period 1940-70 for the available stations in three selected districts. The number of available stations in the districts varied from 7 to 11. They found that it is possible to select a suitable network accounting for 95 per cent of the variance in the average rainfall over the district.

In spite of the best efforts to maintain the complete rainfall records of all the observatories in an area, we have to face the problem of a part of the records missing in space and time. This situation results in errors in the areal mean rainfall. While planning for water resources of an area it is essential that we should have an estimate of the error introduced by missing record. The main purpose of this study is to obtain an estimate of the mean errors in the areal rainfall of a district in different situations of missing observations. The study is of an experimental and exploratory nature.

2. Area of study and rainfall data used

The area considered is Amravati district in the meteorological sub-division of Vidarbha. Fig. 1 shows the location of Amravati district in Vidarbha sub-division and the 13 raingauge stations. The area of Amravati district is 12,200 sq km. The monthly rainfall data for these stations are available for the period 1901-70 except for a few gaps which were filled up by the "normal ratio method" as suggested by Paulhus and Kohler (1952) and McDonald (1957). Data in punched card form were obtained from the office of the Deputy Director General of Meteorology (Climatology & Geophysics), Pune.

3. Variability in space of rainfall over the district

In any particular month of an year rainfall varies even over a small area of the size of a district. It is necessary to know about this variation when we are considering errors arising out of missing observations. To make the rainfall at a place comparable with that at other places, it has been expressed as percentage departure from long-period mean, *i.e.*, the normal. This has been done for all the stations over the area for the months June to September and for the monsoon season and for each of the years during the period 1901-70. The results, however, are being presented only for the typical monsoon month, July, and for the season.

In each of the years, the highest and the lowest percentage departures from normal over the area are noted for July and the season along with the numbers of the stations recording these. Fig. 2(a) shows the highest and the lowest percentage departures of July rainfall from normal over the area in each of the years during the period 1901-70. Similar departures for the seasonal rainfall are given in Fig. 2(b). Quite large variations are noticed. For example, during 1949, the highest and the lowest percentage departures from normal for July rainfall are 151 and — 21 respectively. The corresponding values for the seasonal rainfall during 1964 are 40 and — 59 respectively.

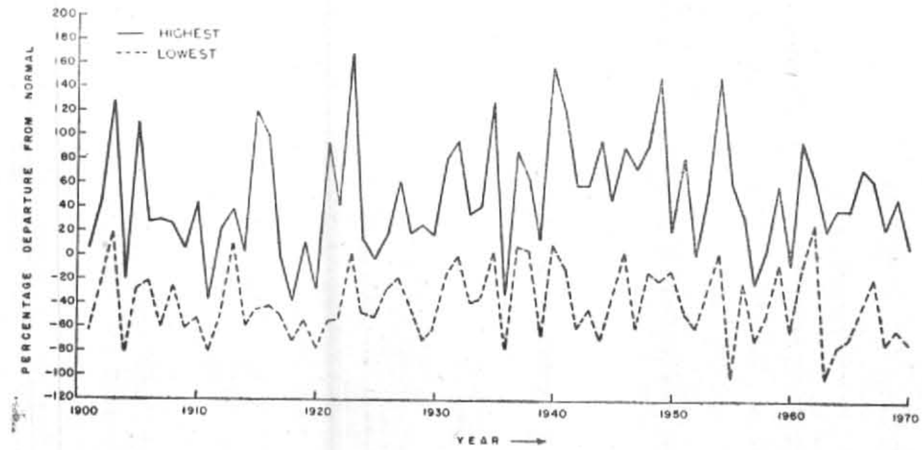

Fig. 2(a). Highest and lowest percentage departure of July rainfall from normal over Amravati district in each of the years during the period 1901-70 (Highest-continuous line, lowest-broken line)

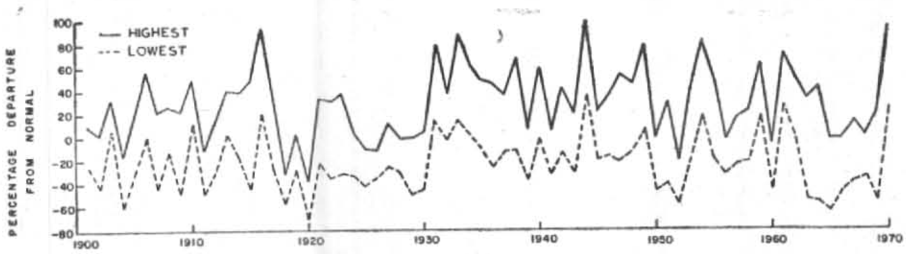

Fig. 2(b). Highest and lowest percentage departure of monsoon seasonal rainfall from normal over Amravati district in each of the years during the period 1901-70 (Highest-continuous line, lowest-broken line)

TABLE 1

Results of application S-E and M-K tests to the series of locations of highest as well as lowest percentage departure of July and seasonal rainfall from normal over Amravati district

Series	Period of rainfall	Test statistic	
		S-E	M-K
Location of highest percentage departure	July	41	−0.13
Do.	Season	39	−0.08
Location of lowest percentage departure	July	43	−0.06
Do.	Season	34	−0.09

The location of the occurrence of the highest as well as the lowest percentage departure of rainfall from normal over the area in different years has been examined for randomness or otherwise. This has been done for July and seasonal rainfall. Swed and Eisenhart's (S-E) test of runs above and below the median and the Mann-Kendall (M-K) rank statistic test for randomness against trend has been applied to the series of station numbers recording the highest percentage depature of rainfall and also to the series of station numbers recording the lowest percentage depature of rainfall. The results of the tests are given in Table 1. As seen from the statistical tables by Owen (1962), for $m=n=$ 35, 5 and 95 per cent confidence limits on the number of runs above and below the median are 28 and 43 respectively. The actual numbers of runs obtained is within these limits and as such no non-randomness is suggested by this test. The value of the M-K test statistic lying in the interval $\pm$ 0.15 is not significant at 5 per cent level. All the four values of the M-K test statistic are not significant at 5 per cent level. Results from both the tests suggest that the location of the highest as well as the lowest percentage departure of July/seasonal rainfall from normal is random over the area of Amravati district. Thus, the occurrence of the highest or the lowest percentage departure of rainfall from normal over the area is not confined to a few specific stations.

4. Examination of the errors in areal rainfall

In this study, the errors in the areal rainfall of Amravati district arising as a result of using networks with 1, 2, 3, 4 and 5 stations missing have been considered. The total number of networks with i missing stations is $\binom{13}{i}$. Hence the numbers of networks with 1, 2, 3, 4 and 5 stations missing are 13, 78, 286, 715 and 1287. For each of these several networks, areal rainfall has been computed for June, July, August, Sep-

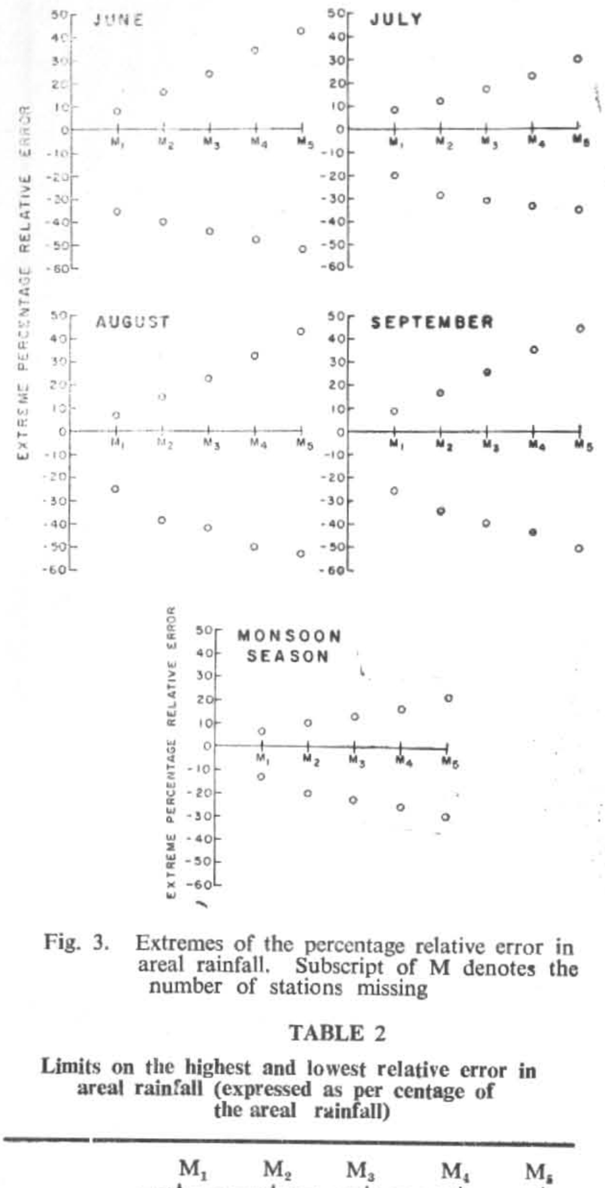

Fig. 3. Extremes of the percentage relative error in areal rainfall. Subscript of M denotes the number of stations missing

TABLE 2

Limits on the highest and lowest relative error in areal rainfall (expressed as per centage of the areal rainfall)

	M_1		M_2		M_3		M_4		M_5	
	U	L	U	L	U	L	U	L	U	L
(a) Highest										
Jun	8	6	16	4	24	6	34	7	42	9
Jul	8	4	12	0	17	0	23	–1	30	–2
Aug	7	0	15	–1	23	–3	33	–4	43	–6
Sep	8	5	17	6	26	6	35	7	44	8
Season	6	–1	10	–2	13	–5	16	–6	22	–6
(b) Lowest										
Jun	–6	–35	–7	–40	–7	–44	–7	–48	–8	–52
Jul	–3	–20	–2	–29	–1	–31	–1	–33	–1	–35
Aug	–2	–25	–1	–38	2	–42	2	–50	3	–53
Sep	–4	–26	–4	–34	–5	–39	–5	–43	–5	–50
Season	–1	–13	0	–20	1	–23	3	–26	4	–29

Note — M_1, M_2, M_3, M_4, M_5 denote 1, 2, 3, 4, 5 stations missing respectively and U and L are upper and lower limits.

tember and the monsoon season (June to September) for each of the years 1901-70. Hereafter, the months June to September will be referred to as the monsoon months and the monsoon season as the season. Areal rainfall obtained by using all the 13 raingauge stations has been taken as the correct areal rainfall and error is obtained in each of the cases of missing station(s) and the same is expressed as a percentage of the correct areal rainfall. Hereafter, this error will be referred to as Percentage Relative Error (PRE).

4.1. *Extremes of percentage relative error*

We have 13 networks with one station missing and 70 years' data. This gives 910 PREs. The highest and the lowest of these have been noted. In the same way, the highest and the lowest PREs have been obtained when 2, 3, 4 and 5 stations are missing. Fig. 3 shows these extreme PREs for the monsoon months and the season when the number of stations missing is 1, 2, 3, 4 and 5. It is seen that the range of extremes increases with the number of missing stations. Amongst the monsoon months, the range of extremes is much smaller for July than that for each of the remaining months, and for each of the situations of 1, 2, 3, 4 and 5 stations missing. The range of extremes for June, August and September generally differs little from each other, for each of the situations of missing stations. For the monsoon season, the range is much lower than that for July.

4.2. *Limits on the highest and the lowest percentage relative error*

Corresponding to each of the 13 sub-networks with one station missing we get a series of PREs and this series contains 70 terms. We note down the highest for each of the 13 series. From these 13 values we obtain the upper and the lower limits on the highest error. In the same manner we obtain the upper and the lower limits on the lowest error. It may be noted that the upper limit on the highest error and the lower limit on the lowest error are the extremes as discussed in the preceding sub-section. Proceeding in the same way, upper and lower limits on the highest and the lowest values have been obtained when 2, 3, 4 and 5 stations are missing. These upper and lower limits are given in Table 2 for each of the monsoon months and for situations of 1, 2, 3, 4 and 5 stations missing. In Table 1, M_1, M_2, M_3,

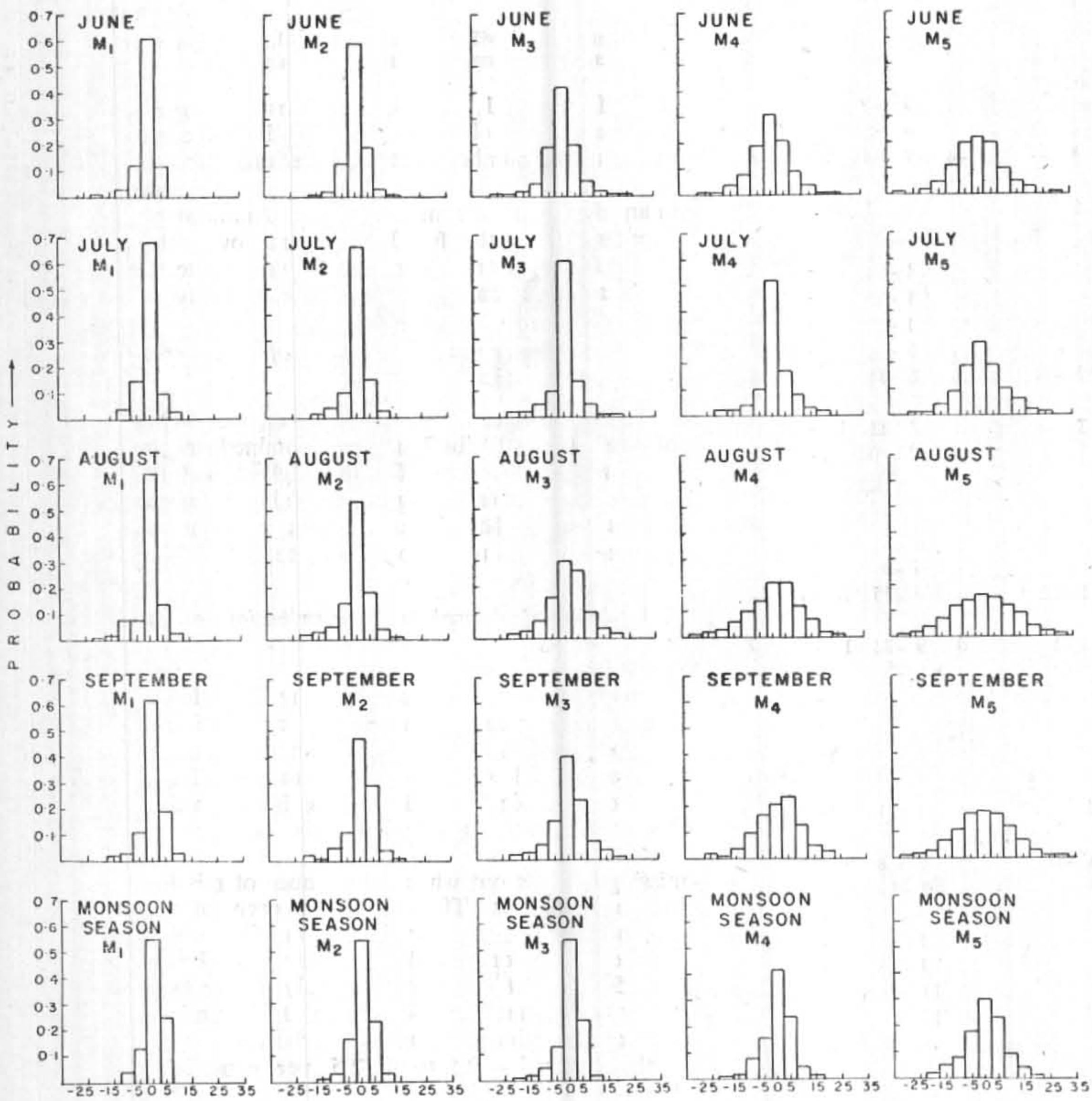

Fig. 4. Probability distribution of the percentage relative error in areal rainfall. Subscript of M denotes the number of stations missing

TABLE 3

Limits of deciles of the distribution of error in the areal rainfall expressed as percentage of areal rainfall

Decile	M_1 L	M_1 U	M_2 L	M_2 U	M_3 L	M_3 U	M_4 L	M_4 U	M_5 L	M_5 U
					June					
1	−7	−1	−12	−1	−15	−1	−19	−2	−21	−1
2	−5	0	−8	0	−10	0	−13	1	−17	1
3	−5	1	−6	2	−8	3	−11	3	−13	4
4	−4	1	−6	3	−7	4	−8	5	−10	6
5	−3	2	−4	3	−5	5	−7	7	−8	7
6	−3	2	−4	4	−4	6	−5	8	−6	10
7	−2	3	−3	5	−4	7	−4	9	−4	12
8	0	4	−1	6	−2	9	−1	11	−1	14
9	2	5	1	8	1	11	1	16	1	19
					July					
1	−12	0	−19	1	−21	2	−24	3	−27	4
2	−10	1	−15	2	−19	3	−21	5	−22	7
3	−9	1	−14	2	−16	4	−18	6	−19	9
4	−8	1	−13	3	−15	5	−16	7	−17	9
5	−7	2	−11	4	−13	6	−14	8	−16	11
6	−7	2	−9	4	−11	6	−13	9	−13	12
7	−6	3	−8	6	−9	8	−11	11	−12	14
8	−5	3	−7	7	−8	10	−10	14	−10	17
9	−3	4	−4	8	−6	12	−6	16	−8	20
					August					
1	−16	0	−26	2	−31	4	−35	6	−37	8
2	−14	1	−23	3	−27	6	−30	8	−33	11
3	−12	2	−22	4	−25	7	−25	10	−28	14
4	−11	3	−20	5	−22	8	−24	11	−25	15
5	−10	3	−16	6	−20	9	−22	13	−23	17
6	−10	3	−15	7	−17	10	−20	14	−22	18
7	−9	4	−14	8	−15	12	−17	15	−29	20
8	−7	4	−11	8	−13	13	−15	17	−15	23
9	−6	6	−8	10	−10	15	−11	21	−13	28
					September					
1	−14	−1	−21	0	−24	0	−27	0	−28	1
2	−12	0	−17	1	−21	2	−22	3	−25	4
3	−10	1	−14	2	−16	4	−18	5	−20	7
4	−7	1	−12	3	−15	5	−15	7	−18	10
5	−7	2	−10	4	−11	6	−13	9	−14	12
6	−5	3	−7	5	−9	8	−11	11	−13	14
7	−5	3	−6	6	−7	9	−9	12	−10	16
8	−3	4	−4	7	−5	11	−6	15	−7	18
9	−2	5	−2	9	−2	12	−3	17	−4	22
					Season					
1	−10	0	−16	2	−18	3	−19	5	−21	7
2	−9	1	−14	2	−16	4	−18	6	−19	8
3	−8	2	−13	3	−15	5	−16	7	−18	9
4	−8	2	−12	3	−14	5	−16	7	−16	10
5	−7	2	−11	4	−13	6	−14	8	−15	10
6	−7	2	−10	4	−12	6	−14	9	−14	12
7	−6	2	−9	5	−11	7	−12	10	−12	13
8	−6	3	−8	5	−10	8	−11	11	−12	14
9	−5	3	− 7	6	−8	9	−9	12	−10	16

Note — M_1, M_2, M_3, M_4, M_5 denote 1, 2, 3, 4, 5 stations missing respectively and L & U are lower and upper limits

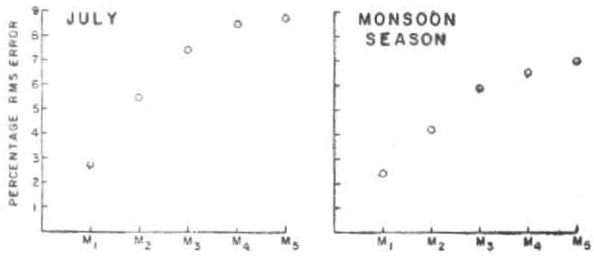

Fig. 5. Percentage RMS error in areal rainfall. Subscript of M denotes the number of stations missing randomly

M_4, M_5 denote 1, 2 3, 4, 5 stations missing respectively. It is observed that the limits become wider as the number of missing stations increases, and that the limits for the lowest error are wider than those for the highest error. Amongst the months, the limits for July are narrower than those for the remaining months and those for the season are narrower than those for July.

4.3. *Limits on the deciles of the error distribution*

The upper and the lower limits on the deciles of the error distribution were obtained in the same manner as those for the highest and the lowest error. These are given in Table 3 for the deciles 1 to 9. These limits become wider as the number of missing stations increases.

4.4. *Empirical probability distribution of the PREs*

The probabilities of errors lying in different ranges have been computed for each of the monsoon months and for the season for the situations of 1, 2, 3, 4 and 5 stations missing. These are shown in Fig. 4 by histograms. It is seen that the error which is predominantly confined to range —7.5 to +7.5 per cent when 1 station is missing spreads out when the number of missing stations increases. This holds for each of the monsoon months and for the season. The probability of the error lying in the interval — 2.5 to + 2.5 per cent is highest for July, being even higher than that for the season. It is noticed that for July the probability for the error lying in the interval —2.5 to + 2.5 per cent falls sharply from the situation of 4 missing stations to the situation of 5 missing stations. The error distributions are generally symmetrical as expected.

4.5. *Root Mean Square Error (RMSE) when stations are missed randomly*

So far, we considered all the cases arising out of each of the five situations of any 1, 2, 3, 4, and 5 missing stations, every year and obtained

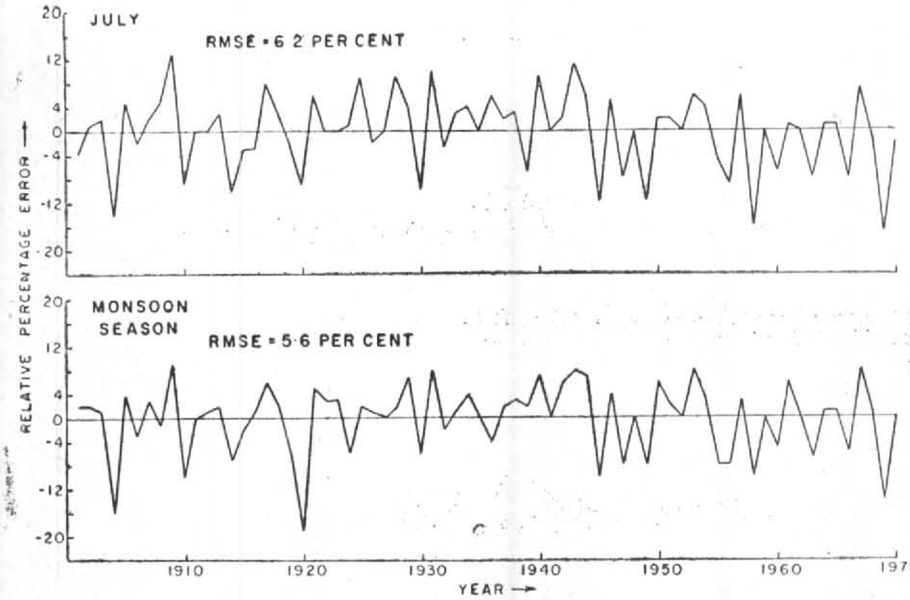

Fig. 6. Relative percentage error in each year when the number of missing stations is random and stations missed are random in space and time

the errors in areal rainfall. We shall now consider separately situations when 1, 2, 3, 4 and 5 stations are missed randomly and obtain the RMSE. To start with, we obtain a series of networks of 70 terms in such a way that one station is missed randomly in space and time, *i.e.,* which of the 13 stations will be missed and when it will be missed are both random. This series of networks is made to correspond with the series of the years 1901 to 1970, and the areal rainfall is calculated for each of the years. From this areal rainfall series PRE is obtained for each of the years for the typical monsoon month, July, and for the season. In the same manner, PRE is obtained for July and for the season in each of the situations of 2, 3, 4 and 5 stations missing randomly in space and time. From these PREs, RMSEs are computed for July and the season for the situations of 1, 2, 3, 4 and 5 stations missing randomly. These are showing Fig. 5. The figure shows that RMSE varies from 2.8 to 8.8 per cent, for July and from 2.2 to 7.0 per cent for the season, as the number of stations missing randomly varies from 1 to 5. RMSE increases with the number of missing stations; however, the rate of increase falls off slowly.

Finally, we consider a series of networks of 70 terms in which the number of stations missing randomly in space and time may be anything from 0 to 5. This series of networks is made to correspond with the years 1901-70, and PRE is obtained for each of the years for July and for the season and the same is shown in Fig. 6 for July and for the season along with the RMSE. For July, in most of the years, the error is between — 8 and + 8 per cent; however, in a few years the error has attained the values + 13 and —17 per cent. For the season, the error lies mostly between — 6 and + 6; however, in a few years the error has touched + 9 and —19 per cent. The RMSEs for July and for the season are 6.2 and 5.6 per cent respectively.

5. Conclusions

(*i*) For July, the extreme percentage errors are (+8, —20), (+12, —29), (+17, —31), (+23, —33), (+30, —35) for 1, 2, 3, 4 and 5 missing stations respectively; for the other monsoon months, the extreme errors are higher in magnitude, but those for the season are smaller

(*ii*) If the number of missing stations is fixed and the stations are missed randomly in space and time, then the RMSE varies from about 3 to 9 per cent for the typical monsoon month, July, and from about 2 to 7 per cent for the season, when the number of missing stations is increased from 1 to 5.

(*iii*) If the number of missing stations varies randomly from 0 to 5 and the stations are missed randomly in space and time, then RMSEs for July and the season are near 6 per cent.

Acknowledgement

The authors convey their thanks to the Director for the facilities to pursue this study, to the Deputy Director General of Meteorology (Climatology & Geophysics) for the supply of the rainfall data and to Mrs. S. P. Lakade for typing of the paper.

References

Bhalla, V. K. and (Mrs.) Narayanan, P., 1978, Proc. of the Symp. on Hydrol. of Rivers with small and Medium Catchments, New Delhi, 17-19 April 1978, Vol. II, Technical papers on Hydrological planning, Design and Management, pp. 19-34.

Hariharan, P. S., 1956 (*a*), *Indian. J. Met. Geophys.*, **7**, pp. 129-136.

Hariharan, P. S., 1956 (*b*), *Indian J. Met. Geophys.*, **7**, pp. 285-288.

Mc Donald, J. E., 1957, *Trans. Am. geophys. Un*, **38**, pp. 657-661.

Owen, D. B., 1962, *Handbook of statistical Tables*, Sandia Corporation, Addison-Wesley Publishing Co., Inc. Pergamon Press, London, p. 374.

Paulhus, J. L. H. and Kohler, M. A., 1952, *Mon. Weath. Rev.*, **80**, pp. 129-133.

Rai Sircar, N. C. and Hariharan, P. S., 1954, *Indian J. Met. Geophys.*, **5**, pp. 219-229.