Mausam, (1982), 33, 1, 107-112

551.579

An operational key station network for the Sone catchment

K. PRASAD

Flood Meteorological Office, Patna

(*Received 25 July 1977*)

सार — सोन जलग्रहण के लिए पदश: बहुसमाश्रयण विश्लेषण से एक प्रचालन की स्टेशन संजाल का अभिकल्पन किया गया है। क्षेत्रीय औसत वर्षा के आकलन के लिए समाश्रयण समीकरणों को हल किया गया है।

ABSTRACT. An operational key station network has been designed for the Sone catchment by stepwise multiple regression analysis. Regression equations have been worked out for the estimation of areal average rainfall.

1. Introduction

The Sone is the principal right bank tributary of the Ganga after the Yamuna. Originating in the Maikala range of hills in Vindhyan Plateau (Madhya Pradesh) it is a source of frequent floods in its lower reaches as it debouches in the plains of Bihar. The low lying areas near its confluence with the Ganga are particularly vulnerable to serious floods. A part of the capital town of Patna also happens to lie in the flood plains of the Sone. The river has caused serious floods in this part of the town on many occasions in the past. In the recent past, the river experienced very serious floods in 1971, 1975 and 1976. The floods of 1975 actually inundated a major part of the town and caused substantial damage to life and property. Those in 1971 and 1976 posed a serious threat to the town. Following these major floods, the forecasting of floods in the Sone has assumed considerable importance.

2. A key station network — need and approach

One of the basic inputs in any scientific flood forecasting scheme is the estimate of areal average rainfall for a catchment. The success of a forecast would depend upon, how quickly, after the occurrence of a flood producing rainstorm, can the real-time rainfall reports within the catchment are obtained. This can be ensured by providing reliable communication channels between the rainfall stations and the forecasting centre. However, economic considerations impose constraints in having a large number of such stations. As such, there is a need for designing a small network of stations from the existing number of stations, which can provide a reliable estimate of the true value of areal average rainfall within permissible limits of statistical error. The objective approach to determine such a network is to pick out the stations which have got a high degree of correlation with the areal average rainfall of the basin. Such stations are called representative stations or key stations and the aggregate of these stations is known as a key station network. A combination of raingauge stations is selected in order of their importance and entered into the multiple regression equation with average basin rainfall as the dependent variable and individual station rainfall as independent variables.

Rao and Bhalla (1977) made an attempt to determine the key station network for Yamuna and Damodar basing using this approach. Mccouloch and Booth (1970) have applied regression analysis technique for the estimation of average monthly basin precipitation over the Great Lakes in Canada. The technique of multiple regression analysis has been adopted in the present study for determining the key station network for the Sone catchment.

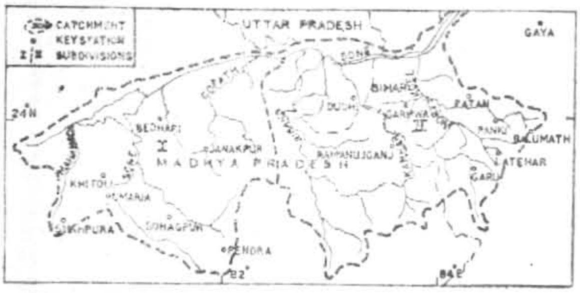

Fig. 1. Sone catchment sub-divisions and key station network

3. The technique of multiple regression analysis (MRA) as applied to the key station network problem

The technique adopted by Rao and Bhalla (1977) consists of ranking the raingauge stations on the basis of *'independent order'* of correlation of individual station rainfall with areal average rainfall of the basin. The first key station is that which has the highest correlation coefficient with the areal rainfall. This station is then removed from the set of independent variables and the areal average rainfall is recalculated with the remaining set. The second key station is one which gives the highest correlation with the recalculated areal average. The process is repeated to select the third, fourth key stations and so on. In the present study the station selected at the first step is the one which accounts for the maximum sum of squares of the areal average rainfall. This is equivalent to the selection of the station on the basis of the highest correlation coefficient with areal average rainfall (sum of squares or variance accounted $=100r^2$ per cent, where, r is the correlation coefficient). The linear effect of this station is then removed from the dependent as well as independent variables. The residuals (dependent and remaining independent variables left over after removal of the linear effect of the variable selected in the first step) are tested for accountability of the sum of squares of areal rainfall by the remaining stations and the one accounting the highest is selected as the next station to enter the regression analysis. The linear effect of the second station is now removed from the residuals and the third station selected in the same manner. This process is carried out till all the stations have been ranked. A few stations of higher ranks, which together account for a desired level of the sum of squares (Variance) of the areal rainfall, are picked up to form the key station network.

The basic difference between the approach followed by Rao and Bhalla (1977) and the one followed in the present study is that while in the former the independent variables are removed physically from the data set in the successive steps, thus changing the basic values of the dependent variable each time, in the later the independent variables are removed mathematically in the linear effect sense and therefore the ranking of stations is done, on the basis of 'partial correlations'.

4. Selection of storm rainfall data used for regression analysis

The Sone catchment contains an area of 71259 sq km and is almost rectangular in shape having a maximum length of about 460 km in the east-west direction and a width of 220 km in north-south direction. The catchment tapers off to a narrow strip downstream. The catchment was divided into two parts, the upper catchment (sub-division-I) comprising the subcatchments of the main river *Sone*, the Mahanadi and the Gopath, and the lower catchment (sub-division-II) comprising the subcatchments of the Rihand, the Kanhar and the north Koel (Fig. 1). The portion forming the narrow strip downstream was not included as the contribution to storm runoff from this small area would be of no significance.

There are thirteen raingauge stations in sub-division-I and eighteen in sub-division-II for which long-term continuous rainfall records were available. Rainfall events, when one-day point rainfall at a raingauge station exceeded 5 cm were picked up from rainfall records for the 20-year period from 1951-70. Dates on which three or more stations recorded rainfall exceeding 5 cm were compiled. The rainfall data of these rain spells were used as the basic data for regression analysis. The areal average rainfall was computed by taking arithmetic mean of all the available stations.

The data sample for the regression analysis was taken from the 10-year period 1951-60. The remaining data were used for testing the regression equations. The development sample consisted of 27 observations for sub-division-I and 40 for sub-division-II; the test samples had 22 and 18 observations respectively.

TABLE 1

Raingauge stations arranged in order of importance with proportion of sum of squares accounted by each and corresponding correlation coefficients as obtained at the first step of STPRG

S. No.	Station	Sum of squares accounted (%)	Corr. coeff.
	Sub-division-I		
1	Umaria	49.3	0.7025
2	Khitoli	40.5	0.6368
3	Niwar	39.8	0.6312
4	Sohagpur	21.0	0.4584
5	Pendra	20.8	0.4565
6	Amari	19.9	0.4459
7	Shahpura	12.1	0.3478
8	Sidhi	7.6	0.2749
9	Beohari	6.9	0.2620
10	Pushparajgarh	2.5	0.1573
11	Murwara	1.8	0.1358
12	Janakpur	0.8	0.0921
13	Jaiwan	—	0.1587
	Sub-division-II		
1	Garu	42.3	0.6507
2	Latehar	38.4	0.6194
3	Mahuadanr	29.8	0.5459
4	Netarhat	29.2	0.5407
5	Balumath	25.8	0.5082
6	Daltonganj	25.3	0.5029
7	Patan	13.9	0.3723
8	Panki	12.2	0.3490
9	Garhwa	10.5	0.3246
10	Kusmi	10.5	0.3241
11	Dudhi	9.6	0.3096
12	Bishrampur	8.9	0.2987
13	Bhandaria	6.3	0.2514
14	Lesliganj	5.8	0.2417
15	Ramanujganj	5.4	0.2321
16	Ambikapur	5.0	0.2229
17	Ranka	4.3	0.2078
18	Nagerutari	0.02	0.0147

5. MRA computational scheme

Denoting the areal average rainfall (dependent variable of the problem under consideration) by Y and individual station rainfall by X_j's ($j=1, 2..........r$), where, r is the number of stations (independent variables), the linear regression model for Y on X_j's may be written as

$Y = a + \sum_{j=1}^{r} b_j x_j + e$ which can also be put in form

$$Y = \overline{Y} + \sum_{j=1}^{r} b_j (X_j - \overline{X}_j) + e \text{ or}$$

$$y = \sum_{j=1}^{r} b_j x_j + e \qquad (1)$$

where, y, x_j's are measured from their respective means. 'e' is the 'error'.

$a = \overline{Y} - \sum_{j=1}^{r} b_j \overline{X}_j$ is constant in the regression-equation

b_j's are the regression coefficients to be estimated by minimising the sum of squared residuals

$$\Sigma e^2 = \Sigma (y - \sum_{j=1}^{r} b_j x_j)^2$$

Partial differentiation of Σe^2 with respect to b_j's and equating to zero gives a set of r normal equations which can be put in the matrix form as

$$\begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1r} \\ a_{21} & a_{22} & \cdots\cdots & a_{2r} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a_{r1} & a_{r2} & \cdots\cdots & a_{rr} \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_r \end{bmatrix} = \begin{bmatrix} a_{1y} \\ a_{2y} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_{ry} \end{bmatrix}$$

$$\text{or, } \underset{\sim}{A} . \underset{\sim}{B} = \underset{\sim}{A}_y \qquad (2)$$

$\underset{\sim}{A}$ is $r \times r$ covariance matrix, $\underset{\sim}{B}$ is $r \times 1$ matrix of regression coefficients and $\underset{\sim}{A}_y$ is $r \times 1$ matrix of cross products of deviations of X_j's and Y from their respective means where

$$\underset{\sim}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1r} \\ a_{21} & a_{22} & \cdots\cdots & a_{2r} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a_{r1} & a_{r2} & \cdots & a_{rr} \end{bmatrix} = \begin{bmatrix} \Sigma x_1^2 & \Sigma x_1 x_2 & \cdots & \Sigma x_1 x_r \\ \Sigma x_2 x_1 & \Sigma x_2^2 & \cdots & \Sigma x_2 x_r \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \Sigma x_r x_1 & \Sigma x_r x_2 & \cdots & \Sigma x^2_r \end{bmatrix}$$

$$\underset{\sim}{B} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_r \end{bmatrix} \quad \underset{\sim}{A}_y = \begin{bmatrix} a_{1y} \\ a_{2y} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_{ry} \end{bmatrix} = \begin{bmatrix} \Sigma x_{1y} \\ \Sigma x_{2y} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \Sigma x_{ry} \end{bmatrix}$$

To determine $\underset{\sim}{B}$ Eqn. (2) is written as

$\underset{\sim}{A}^{-1} \underset{\sim}{A} \underset{\sim}{B} = \underset{\sim}{A}^{-1} \underset{\sim}{A}_y$ where $\underset{\sim}{A}^{-1}$ is the inverse of matrix $\underset{\sim}{A}$ so that $\underset{\sim}{B} = \underset{\sim}{A}^{-1} \underset{\sim}{A}_y$ (3)

The regression coefficients are determined by the process of matrix inversion.

In the stepwise MRA technique, the independent variable entering the regression equation in the first step is selected, first, by computing the amount of sum of squares of dependent variatble accounted by each independent variable as

$$C_j = \frac{a^2_{jy}}{a_{jj}} \qquad (4)$$

(for $j = 1,2.....r$ independent variables; a_{jy} and a_{jj} are defined previously).

and, second, by finding the maximum (over j) of C_j

The linear effect of the variable selected in the first step is now eliminated from all the remaining independent variables as well as the dependent variable and the same process is repeated to select the second, third, and so on, independent variables to enter the regression analysis.

6. Results

The computations were carried out by using a computer subroutine (STPRG) for stepwise multiple regression analysis. The raingauge stations arranged in decreasing order of the proportion (per cent) sum of squares of areal rainfall accounted for by each and the corresponding

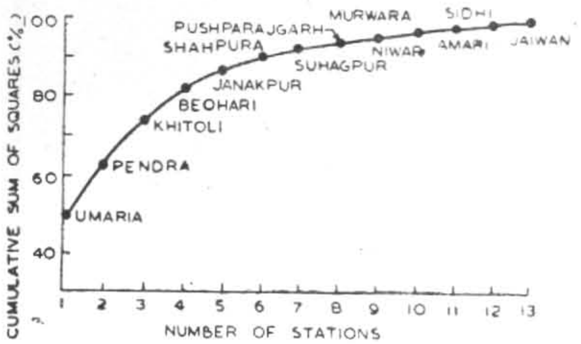

Fig. 2. Cumulative sum of squares (%) accounted for by the regression at successive steps (upper Sone catchment)

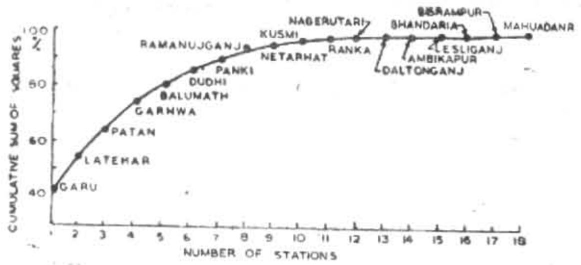

Fig. 3. Cumulative sum of squares (%) accounted for by the regression at successive steps (lower Sone catchment)

TABLE 2 (a)

Sub-division I

Key station network and associated regression constants

Net-Work No.	Regression coefficients							'a'	TSS	SSR	SSR	MCC	SE
	X1	X2	X3	X4	X5	X6	X7	(mm)	(mm²)	(mm²)	(%)		(mm)
1	.0761	.1398	.1438	.0899	.1047	.0646	.0792	11.77		4330.16	91.3	.9556	5.63
2	.0840	.1678	.1461	.0314	.0978	.0759	—	12.64		4237.73	89.4	.9454	6.23
3	.0997	.1294	.1454	.0761	.1055	—	.0971	13.98		4210.39	88.8	.9423	6.39
4	.0542	.1666	.1399	.0950	. —	.0655	.0664	17.63	4741.56	4103.11	86.5	.9302	7.01
5	.1224	.1327	.1321	. —	.1178	.0230	.0514	16.29		3829.05	80.7	.8986	8.38
6	.2036	.0981	. —	.0809	.0970	.0683	.0877	15.54		3714.00	78.3	.8850	8.89
7	.1201	. —	.1124	.0858	.1446	.0644	.1564	12.25		3966.15	83.6	.9146	7.72
8	—	.1559	.1788	.0998	.0885	.0796	.0871	12.85		4236.53	89.3	.9452	6.23

X1 : Umaria, X2 : Pendra, X3 : Khitoli, X4 : Beohari, X5 : Janakpur, X6 :Shahpura, X7 :Sohagpur

TABLE 2 (b)

Sub-division II

Key station net work and associated regression constants

	X1	X2	X3	X4	X5	X6	X7	X8	'a' (mm)	TSS (mm²)	SSR (mm²)	SSR (%)	MCC	SE (mm)
1	.0978	.1064	.0864	.0973	.1090	.1164	.0718	.0959	6.76		4234.52	93.2	.9653	3.84
2	.1015	.0950	.0865	.1036	.1201	.1233	.0678	—	9.82		4068.84	89.5	.9462	4.76
3	.1035	.1325	.1111	.0947	.0933	.1041	—	.0890	8.52		4066.15	89.5	.9459	4.77
4	.0989	.1200	.0885	.1167	.0973	—	.0624	.1050	8.83	4544.45	3957.38	87.1	.9332	5.29
5	.1309	.1122	.0875	.1284	—	.1015	.0567	.1144	8.06		3908.32	86.0	.9274	5.50
6	.0862	.1213	.0766	—	.1308	.1337	.0701	.1033	9.08		4029.19	88.7	.9416	4.95
7	.0996	.1085	—	.0778	.1106	.1202	.1051	.0963	10.17		3878.65	85.3	.9238	5.63
8	.1359	—	.0876	.1143	.1137	.1302	.0919	.0807	7.63		3857.84	84.9	.9214	5.72
9	—	.1593	.0877	.0796	.1445	.1179	.0777	.1025	8.03		3960.72	86.0	.9270	5.51

X1 : Garu, X2 : Latehar, X3 : Patan, X4 : Garwa, X5 : Balumath, X6 : Dudhi X7 : Panki & X8 : Ramanujganj

'a' : Costant in the regression equation. TSS : Total sum of squares of dependent variable. SSR : Sum of squares due to regression. MCC : Multiple correlation coefficient. SE : Standard error.

correlation coefficients as obtained in the first step of STPRG are shown in Table 1 for the two subdivisions. The raingauge station Umaria in sub-division-I and Garu in sub-division-II which account for the maximum sum of squares (highest correlation coefficients) are, therefore, selected as the first key stations to enter the regression analysis. Subsequent steps follow as outlined in the computational scheme above.

The Cumulative sum of squares (expressed as percentage of the total) obtained at successive steps of STPRG plotted against the number of stations entering into the regression are shown in Figs. 2 and 3.

It will be seen that the curves rise steeply at first and become asymptotic after inclusion of a few stations, implying that the addition of a new

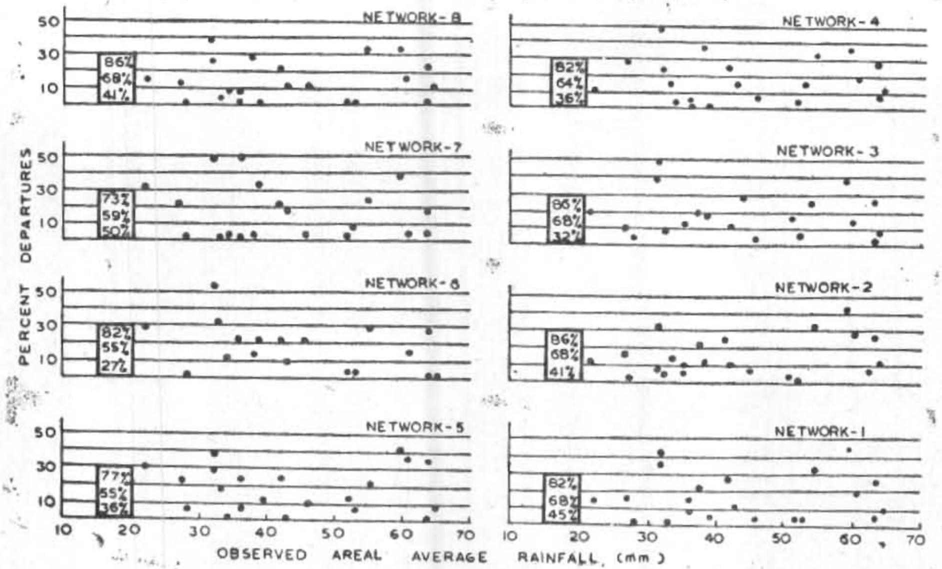

Fig. 4. Test of regression quations on independent data sample — observed areal average rainfall and percentage departures of individual areal estimates from observed values under various key networks

station after a certain stage does not make any significant contribution towards reduction of the unaccounted sum of squares.

The stations which accounted for 90 per cent sum of squares of areal rainfall were finally selected as key stations. The criterion gave a combination of 6 stations for the upper Sone catchment (sub-division-I) and 7 stations for the lower catchment (sub-division-II) as forming the key station network. These stations arranged in order of their importance are:

Upper catchment (sub-division-I) : Umaria, Pendra, Khitoli, Beohari, Janakpur, Shahpura.

Lower catchment (sub-division-II) : Garu, Latehar, Patan, Garhwa, Balumath, Dudhi, Panki.

The locations of these stations are also shown in Fig. 1.

7. The problem of missing station — alternate networks

Provision has to be made for the contingency that the real time-rainfall report from a station out of the key stations selected may be missing on any occasion. One way to meet this contingency would be to designate an alternate station for each station in the key network. This approach will, however, involve keeping as many additional stations as the key stations in the operational network and the very purpose of working out a key network will be defeated.

The approach adopted in the present case was to add one station to the network of key stations originally worked out; the station added being the one immediately succeeding the last station in the key network. Thus in the event of any station missing, the number of stations in the network remains same as in the original key network. Regression equations for each set of stations can be worked out.

With this approach the total number of operational key stations was taken as seven and eight respectively for the upper and lower catchments. The regression coefficients and related constants for the combinations of all stations and with one station missing were computed.

The key station network, the alternate networks with one station missing and the related regression constants are shown in Table 2 (a & b) for sub-divisions I and II. It will be seen from these tables that the proportion sum of squares accounted by the regression, with the complete set of seven and eight stations taken into consideration and from various combinations with any one station missing, differ by only small amounts. This shows that even when the best combination, *viz.*, the net work originally worked out is not available for estimation of areal rainfall, the other combinations used would not cause much loss of accuracy. This approach has, therefore, the obvious advantage that by taking just one additional station in the operational network the problem arising from any station missing from the network can be tackled by using the appropriate combination of remaining stations.

8. Test of regression on independent data

Test of the regression equations on independent data were carried out in respect of all the networks for each of the two sub-divisions. Results are presented diagramatically in Figs. 4 and 5. The figures show percentage departure of the estimated values against observed areal rainfall amounts and also the percentage of cases lying

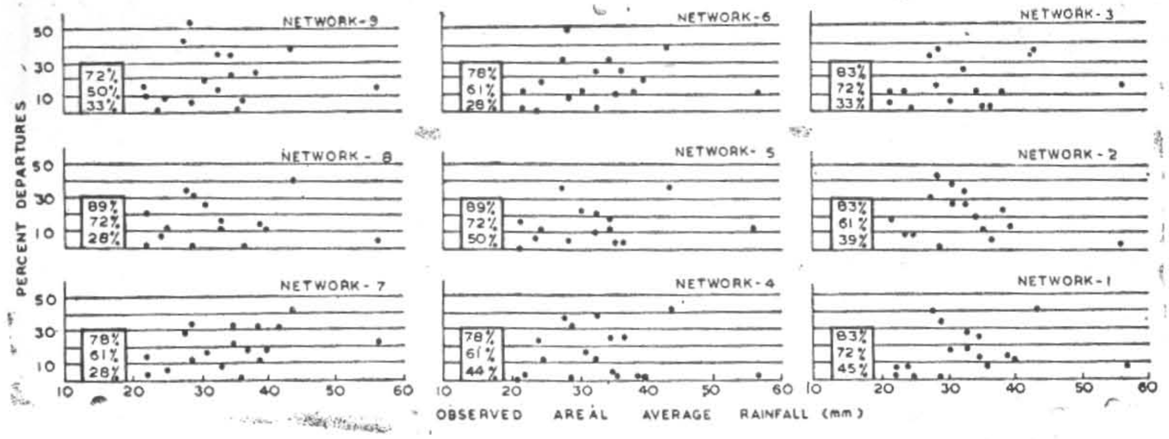

Fig. 5. Test of regression equations on independent data sample — observed areal average rainfall and percentage departures of individual areal estimates from observed values under various key networks (lower Sone catchment — Sub-division-II) (figures in rectangles are the proportion of cases within 10%, 20% and 30% departures)

within 10 per cent, 20 per cent and 30 per cent departures. The agreement between the estimated and observed values is reasonably good.

9. The order of importance of raingauge stations as obtained on the basis of simple correlation coefficients and STPRG — a comparison

It will be seen that the order of stations arranged according to their simple correlation coefficients with the areal rainfall as indicated in Table 1 and according to partial correlations as indicated in Figs 2 and 3 are different from each other. The STPRG (partial correlation) scheme for determining the order of importance of independent variable being superior to the one of simple correlation coefficients can be seen from the fact that the Regression Sum of Squares computed by taking into consideration first seven stations of sub-division-I and eight stations of sub-division-II (Table 1) arranged according to simple correlation coefficients, worked out to be 79 per cent (m.c.c. 0.8887) and 78.7 per cent (m.c.c. 0.8869) respectively as against the corresponding figures of 91.3 per cent (m.c.c. 0.9556) and 93.2 per cent (m.c.c. 0.9653) under STPRG scheme.

10. The applicability of regression equations

It need hardly be mentioned that the regression equations derived above are applicable to only those situations where significant rainfall occurs in the catchment.

It is generally said that the presence of constant in the regression equations renders them unrealistic for estimation of areal rainfall since the calculated value of areal rainfall does not become zero when all the individual point rainfall at stations composing the regression equations are zero. This argument is untenable in view of the fact that the regression equations developed are applicable to only storm rainfall situations since the character of the basic data has been defined that way. In any case, the question of nonapplicability of the regression equations to zero or insignificant rainfall situations is by itself hypothetical, since there is no need for estimation of areal rainfall, as far as flood forecasting is concerned, when the rainfall amounts are of insignificant order of magnitude.

11. Conclusions

The stepwise multiple regression analysis has given six stations for the upper catchment of the Sone and seven for the lower catchment as forming the key station network; the regression sum of squares being about 90 per cent. It has been demonstrated that the problem of any station's rainfall report not being available on any occasion can be tackled by adding only one station to the original key network.

The agreement between the estimated values and the observed values of areal rainfall has been found to be reasonably good.

Acknowledgements

The author is grateful to Dr. A. S. Ramanathan, Director (Hydrometeorology), New Delhi for providing facilities in his Directorate for carrying out the work. The author is also indebted to Shri A. K. Sen Sarma, Meteorologist-in-Charge, Flood Meteorological Office, Patna for constant encouragement and invaluable suggestions. Thanks are also due to Shri H. C. Mehra and R. S. Pareek of the Hydromet Directorate, New Delhi for their active co-operation in execution of the computer programme and Shri S. R. Singh for typing the manuscript.

References

Mcculoch, I. A. W. and Booth, M., 1970, Estimation of Basin precipitation by Regression Equation, *Water Resources Res.*, **6**, 6, pp. 1753-1758.

Rao, D. V. L. N. and Bhalla, V. K., 1977, Design of representative and key raingauge stations network for flood forecasting in Yamuna and Damodar rivers, *Indian J. Met. Hydrol. Geophys.*, **28**, 1, pp. 63-73.