Raingauge Network Design for a Basin

Pankaj Mani
Scientist "D",
Centre for Flood Management Studies,
National Institute of Hydrology,
Patna, Bihar (India)
mailofpmani@yahoo.com

1. Introduction

A hydrological network is an organized system for collection of information of specific kinds such as precipitation, run off, water quality, sedimentation and other climate parameters. The accuracy in the decision making in the water project design depends on how much information are available for the region concerned. Having enough relevant and accurate hydrologic information reduces the chances of under-design or overdesign and thus minimizes the economic losses, which leads to the overall increase in the benefit/ cost ratio. The failure of many capital intensive projects throughout the world can be attributed in part to an inadequate record length, the sparseness of the network, and/ or inaccuracy of the hydrologic information. It has not, so far, been possible to define the optimum level of hydrologic information required for planning, design and development of a specific project in a region, due to difficulties in developing a benefit cost function of hydrologic information. It is, therefore, difficult to attain an optimum balance between, on one hand the economic risk arising from inadequate information and on the other hand, the cost of a hydrologic network capable of transmitting the required information. Unless techniques to evaluate such a balance are developed, the network design methods sited in the literature cannot be universally applied.

There are several ways to define the objectives of the hydrological network design, but the fundamental theme, in most cases, is the selection of an optimum number of stations and their optimum locations. Other considerations that can arise in the network design are; achieving an adequate record length prior to utilizing the data, developing a mechanism to transfer information from gauged to ungauged locations when the need arises and estimating the probable magnitude of error or regional hydrologic uncertainty arising from the network density, distribution and record length. Another difficulty in developing a network design methodology is related to the complexity in dealing with the multi variate interaction of hydrologic events in the domains of space and time. The stochastic nature of the hydrologic variables complicates the problems further.

The objective of providing a network of raingauge is to adequately sample the rainfall and explain its variability within the area of concern. The rainfall variability depends on topography, wind, direction of storm movement and type of storm. The location and spacing of gauge depends not only on the above factors but also upon the use of the data for that region. For example in the tea garden areas of the North East region, the density of precipitation gauge is found to be much higher, even in the valleys, in comparison to the hills where jhoom cultivation are practiced, since the data are used for irrigation planning of the tea gardens. Network design covers following three main aspects (WMO, 1976):

a. Number of data acquisition points required.

b. Location of data acquisition points and

c. Duration of data acquisition from a network

Measurement stations are divided into three main categories by WMO, namely:

1. Primary stations: These are long term reliable stations expected to give good and reliable records.
2. Secondary or Auxiliary stations: These are placed to define the variability over an area. The data observed at these stations are correlated with the primary stations, and if and when consistent correlations are obtained secondary stations can be discontinued or removed.
3. Special stations: These are established for specific studies and do not form a part of minimum network or standard network.

The World Meteorological Organisation (1976) has recommended the minimum network densities for general hydro-meteorological practices.

1. For plain regions of temperate Mediterranean and tropical zones one station for 600-900 sq. km.

2. For mountainous region of temperate Mediterranean and tropical zones one station for 100-250 sq. km.

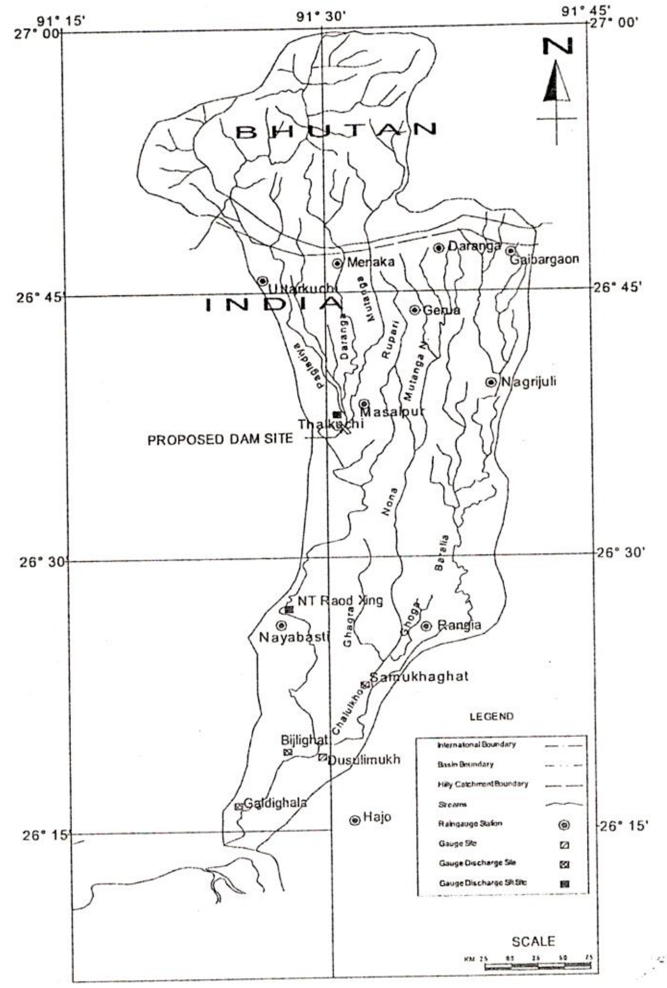
3. For arid and polar region one station for 1,500-10,000 sq. km.

Table I.2.6. Recommended minimum densities of stations (area in km^2 per station)

Physiographic unit	*Precipitation*		*Evaporation*	*Streamflow*	*Sediments*	*Water quality*
	Non-recording	*Recording*				
Coastal	900	9000	50000	2750	18300	55000
Mountains	250	2500	50000	1000	6700	20000
Interior plains	575	5750	5000	1875	12500	37500
Hilly/undulating	575	5750	50000	1875	12500	47500
Small islands	25	250	50000	300	2000	6000
Urban areas	–	10–20	–	–	–	–
Polar/arid	10000	100000	100000	20000	200000	200000

Source: Guide to Hydrological Practices, Volume-I: "Hydrology – From Measurement to Hydrological Information", WMO No. 168

In this lecture, the different methods of raingauge network design have been illustrated through a case study for a basin in the north-east part of India. In the study area, the rains are of long duration and occur mostly between March and October. During March to April the rainfall is sporadic, but it is steady and heavy or very heavy during May to October. Annual rainfall in the basin is around 2300 mm. There is wide spatial variation in this catchment. Various methods have been tried which takes into account the location and type of basin and its climate, precipitation characteristics from the existing raingauge stations etc. Apart from BIS recommendations, C_v method, Key station method, spatial correlation method and entropy concept have been tried for determining the adequate number of raingauge stations and their suitable location.



2. Study Area

Pagladiya river is one of the major tributary of Brahmaputra on its north bank. The river originates on southern slopes in hills of Bhutan at an altitude of 3000 m above mean sea level (msl). After traversing through the Bhutan territory, it enters the Nalbari district of Assam near Chowki. It meets Brahmaputra near Lowpara village. The basin is lying between 26°14' - 27°0' N latitude and 91°18' - 91°42' E longitude. The catchment area of the basin is 1507 km^2 of which 1084 km^2 is in India and remaining 423 km^2 is in Bhutan. The hilly part of the catchment area comprises of 465 km^2 (423 km^2 in Bhutan and 42 km^2 in India). The river length is 196.8 km out of which it flows for a length of 19 km in the hilly tracts of Bhutan and the rest 177.8 km in plains of Nalbari district of Assam. The major tributaries of the river are Mutunga (L- 30 km, A- 130 km^2), Dimla (L-25.5 km, A- 48.75 km^2), Nona/Mutunga (L- 63 km, A- 268 km^2) and Chowlkhowa (L- 75 km, A- 538 km^2) on its left bank.

Rainfall in the catchment occurs mostly during June to October. There are also some pre-monsoon and post-monsoon showers. The average annual rainfall in the catchment is in the order of 2300 mm. Daily rainfall data for ten stations have been used in the study. only. The details of data used in the study is given in Table - 1. For Nagarjuli, Darrang, Giobergaon and Rangiya, rainfall data are available for a period of about 22 years, from 1975 to 1996 (rainfall for few days are missing) while for Maneka, Uttarkuchi, Hajo and Masalpur, data are available for about 15 years. For Gerua and Nayabasti, data are available from 1991 to 1994 only. The concurrent daily rainfall is available for a limited period only (from 1991 to 1994) for 9 stations (except Masalpur). Most of the hills are in Bhutan and no raingauge station is available in this region. It has been assumed that the data collected at Uttarkuchi, Menaka, Darang and Goibergaon, all along the foothills, exhibit the rainfall characteristics of hilly portion of the basin. Average annual rainfall in the catchment is 2300 mm and its seasonal distribution is pre-monsoon 26.66%, monsoon 64.86%, post monsoon 5.5% and winter 2.97%.

Table – 1

Data availability in the study area

Station	Year
Darrang	1975-96
Menaka TE	1976, 1983-96
Uttarkuchi	1980-96
Gaibargaon	1975-96 (except 1990)
Nagarjuli TE	1975-96
Uttarkuchi	1980-96
Masalpur	1975-88 (except 1979)
Nayabasti (Nalbari)	1991-94
Rangiya	1976-95 (except 1990)
Hajo	1980-95(except 1990)

3. Methodology

The following methodologies have been used in the present study:

(i) IS: 4986-1968 guidelines

The Bureau of Indian Standard (BIS) suggests that one raingauge up to 500 sq. km might be sufficient in non-orographic regions. In regions of moderate elevation (up to 1000 m above msl), the network density might be one raingauge for 260 - 390 sq. km. In predominantly hilly areas and areas of heavy rainfall, the density recommended is one for 130 sq. km.

(ii) C_v method

The problem of ascertaining the optimum number of raingauges in various basins is of statistical nature and depends on spatial variation of rainfall. Thus, the coefficient of spatial variation of rainfall from the existing stations is utilised for determining the optimum number of raingauges. If there are already some raingauges in the catchment, the optimal number of stations that should exist to have an assigned percentage of error in the estimation of mean rainfall is obtained by statistical analysis as:

$$N = \left(\frac{C_v}{P}\right)^2$$

where,

N = optimal number of stations,

p = allowable degree of error in the estimate of mean rainfall and

C_v = coefficient of variation of rainfall values at the existing m stations.

If there are m stations in a catchment and P_1, P_2, ...P_m is the recorded rainfall at a known time at 1, 2, ...m station, then the coefficient of variation C_v is calculated as:

$$C_v = \frac{100 \times \sigma_{m-1}}{\overline{P}}$$

where

$$\sigma_{m-1} = \frac{\sqrt{\sum_{i=1}^{m} P_i^2 - m \times \overline{P}^2}}{(m-1)}$$

P_i is monthly average precipitation at i th station and P is the average rainfall of 'm' number of stations, given by:

$$\overline{P} = \frac{\sum_{i=1}^{n} P_i}{m}$$

It is usual practice to take p =10%. σ_{m-1} is used for calculation of C_v when number of stations, m, in the network is less than 30 otherwise σ_m can also be used.

(iii) Key station network method

One of the most rational methods for determination of key station is as suggested by Hall (1972). In this method, at first, the correlation coefficient between the average of storm rainfall and the individual station rainfall are found. The stations are then arranged in the order of their decreasing correlation coefficients and the station exhibiting highest correlation coefficient is called the first key station and its data is removed for determination of next key station. The procedure is repeated by considering the average rainfall of the remaining stations. The station showing the highest correlation coefficient after removing the data of first key station is called the second key station. Similarly third and successive key stations are determined after removing the data of already selected key stations. Now the sum of the squares of deviations of the estimated values of average rainfall from the actual rainfall in respect of 1st, 2nd, 3rd key station etc. is determined and a graph is plotted between the sum of the square of deviation and corresponding number of stations in combinations. It will be seen that a stage comes when the improvement in the sum of squares of deviation is very little with the addition of more stations. The corresponding number of stations at that stage is taken to be representative and key stations for the network in the catchment/ basin.

(iv) Spatial correlation method

Under the assumption that spatial variability of rainfall can be quantified through a spatial correlation function, a network of raingauges can be designed to meet a specified error criterion (Kagan, 1966 and WMO, 1972). However in applying such an approach, care must be taken to ensure that condition necessary for the existence of spatial correlation function, such as hydrological homogeneity and isotropy are fulfilled; flat areas with a relatively homogeneous surface are more appropriate for the application of the technique. A general theoretical spatial correlation method for the planning of meteorological networks has been given by Gandin (1970). Some details of the specific approach and its application have been given by Kagan (WMO, 1972). The basis of this method is the correlation function ρ(d) which is a function of the distance between the stations, and the form of which depends on the characteristics of the area under consideration and on the type of precipitation. The function ρ(d) can frequently be described by the following exponential form:

$$\rho(d) = \rho(0)e^{-d/d_0}$$

where ρ(0) is the correlation corresponding to zero distance and d_0 is the correlation radius or distance at which the correlation is ρ(0)/e. Theoretically, ρ(0) should equal to unity but is rarely found so in the practice due to random errors in precipitation measurement and micro climatic irregularities over an area. The variance of those random errors has been given by Kagan (1966) as:

$$\sigma_1^2 = \left[1 - \sigma(0)\right]\sigma_h^2$$

where σ_h^2 is the variance of the precipitation time series at a fixed point. The quantities ρ(0) and d_0 provide the basis for assessing the accuracy provided by a network. In this context, two accuracy criterion may be of interest.

Criterion 1: The accuracy with which the average rainfall over a given area may be obtained is to be evaluated. For an area 's' with the centre station, and assuming ρ(d) exists and described as above, the variance of the error in the average precipitation over 's' is given by Kagan (1966) as:

$$v = [1 - \rho(0)]\sigma_h^2 + 0.23\sigma_h^2 \frac{\sqrt{s}}{d_0}$$

where, the first term is attributed to random error and second term with spatial variation in the precipitation field.

For an area 'S' with 'N' stations evenly distributed so that S = N × s, the variance of the error in the average rainfall in the area 'S' is given by

$$v_n = \frac{\sigma_h^2}{n}\left[1 - \rho(0) + 0.23\frac{\sqrt{s}}{d_0} \times \sqrt{n}\right]$$

The relative root mean square error is then defined as:

$$z_1 = \frac{\sqrt{v_n}}{\bar{h}} = \frac{C_v}{n}\sqrt{1 - \rho(0) + 0.23\frac{\sqrt{s}}{d_0} \times \sqrt{n}}$$

where $C_v = \sigma_h/h$ and h is the average precipitation over the area S. From the above equation the value of 'n' to meet a specified error criterion Z_1 can be obtained if the values of ρ(0) and d_0 are known or vice versa. The uniform spacing of station over the area S is such that S = N ×s can be achieved on the basis of a square grid for which the spacing between the station is: L= √S/N. However a triangular grid is usually more convenient if the area S has a complex configuration and its spacing is given L=1.07 √S/N

Criterion 2: The accuracy of spatial interpolation is to be evaluated. Kagan (WMO, 1972) has given the relative errors associated with linear interpolation between two points and interpolation at the center of square and triangle, where the maximum error of interpolation occur. For a triangular grid with a spacing one, the relative error is given by as:

$$z_2 = C_v\sqrt{\frac{[1 - \rho(0)]}{3} + \frac{0.52 \times \rho(0) \times \sqrt{s}}{\sqrt{n} \times d_0}}$$

Application: The derivation of z_1 and z_2 in a particular case requires the estimation of ρ(0) from which ρ(d) and d_0 can in turn be derived. The function ρ(d) can be evaluated by calculating the correlation ρ(i,j) as a function of distance between stations. The value of ρ(i,j) is calculated as:

$$\rho(i,j)=\frac{\sum(h_i-\bar{h}_i)(h_j-\bar{h}_j)}{\sqrt{\sum(h_i-\bar{h}_i)^2}\sqrt{\sum(h_j-\bar{h}_j)^2}}$$

where the summations are taken from 1 to m and m is the number of pairs of observations. For determination of ρ(0) and d_0 distance is plotted against correlation and an exponential curve is drawn through the points. The value of ρ(0) is found out by extrapolating ρ(d) to zero distance, and d_0 is calculated as the distance corresponding to a correlation of ρ(0)/e. Alternatively, ln[ρ(d)] may be plotted against d which should result in a linear plot with slope ($-1/d_0$) and intercept is ln[ρ(0)]. The objective of fitting of a straight line to the plotted points by least square may result in values of ρ(0) greater than unity which would be nonessential. Consequently, a subjective approach such as fitting by eye is apparently only the alternative.

(v) Entropy Method

One of the important objectives of network design is to estimate the probable magnitude of error or regional hydrological uncertainty arising from network density, distribution and record length. In entropy method, the hydrological information and regional uncertainty associated with a set of precipitation station are estimated using Shannon's entropy concept. Since time series of rainfall data can be represented by gamma distribution due to the presence of skewness, the entropy term is derived using single and bivariate gamma distribution.

Let **X** represents precipitation variables measured at a station with events $\mathbf{x_1}$, $\mathbf{x_2}$, ... $\mathbf{x_n}$ and with the probability of occurrence of $\mathbf{j^{th}}$ event denoted by $\mathbf{p(x_j)}$. Shannon and Weaver (1949) defined the average uncertainty, also known as entropy with **n** event as

$$H(X)=-\sum_{j=1}^{n}p(x_j)\log_2 p(x_j)$$

where, **H(X)** is the uncertainty or entropy of variable X.

Similarly, the joint entropy in a region with m station and precipitation variables (X_1, X_2, ...X_m) can be extended to

$$H(x_1,x_2,......x_m)=-\sum_{j=1}^{n}p(x_j^1,x_j^2,....x_j^m)\log_2 p(x_1,x_2,......x_m)$$

where, X_1, X_2, ... X_m represents precipitation variables measured at m station and $p(x_j^1, x_j^2, ...x_j^m)$ is the joint probability of occurrence of j^{th} event at m^{th} station.

Computation of Entropy

A random variable X is said to be gamma distributed if its probability density function is given by,

for **x, s, l > 0** $$f(x,\sigma,\lambda) = \frac{1}{\sigma\Gamma(\lambda)}\left(\frac{x}{\sigma}\right)^{\sigma-1} e^{-\frac{x}{\sigma}}$$

where **s** is scale parameter, **l** is shape parameter and **G(l)** is the gamma function and equals to

$$\int_0^{\lambda} x^{\lambda-1}e^{-x}dx$$

The parameters **s** and **l** can be calculated by following method:

Expected value of **X** = $\mu_1(X) = \sigma\lambda$

The variance of **X** = $\mu_2(X) = \sigma^2\lambda$

Similarly, the third and forth central moments of X are

$$\mu_3(X) = 2\sigma^3\lambda$$

$$\mu_4(X) = 3\sigma^4\lambda(\lambda+2)$$

The entropy of gamma probability density function is calculated as

$$H(X) = -\int_0^{\infty} f_G(x,\sigma,\lambda)\ln f_G(x,\sigma,\lambda)dx$$

which is simplified to (Husain, 1987)

$$H(X) = -(\lambda-1)\psi(\lambda)+\Gamma(\lambda+1)/\Gamma(\lambda)+\ln(\sigma\Gamma(\lambda))$$

where $\psi(\lambda)$ is the digamma function:

$$\psi(\lambda) = \partial/\partial\lambda(\ln\Gamma(\lambda))$$

The bivariate gamma distribution, as proposed by Moran (1969) can be transformed to normalised variates z and w as follows:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-0.5t^2}dt = \int_0^{x} f(t,\sigma_x,\lambda_x)dt$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-0.5t^2}dt=\int_{0}^{y} f(t,\sigma_y,\lambda_y)dt$$

In the above equation, X and Y are variables with univariate gamma distribution and with their parameters as (s_x, l_x) and (s_y, l_y) respectively. Z and w are normalized variates of X and Y respectively, with a mean of zero and standard deviation of unity. If r_{zw} is the correlation coefficient between z and w, then the information transmitted by variable Y about X, T(X,Y) or by variable X about Y, T(Y,X) is given by Shannon and Weaver (1949):

$$T(X,Y)=T(Y,X)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_n(z,w)\ln\frac{f_n(z,w)}{f_n(z)f_n(w)}dzdw$$

where $f_n(z,w)$ is the bivariate normal probability density function of normalized variates z and w and is transformed from the bivariate gamma probability density function $f_g(x,y;s_x,l_x,s_y,l_y)$. $f_n(z)$ is the normal probability density function of normalized variate z and is transformed from the gamma probability density function $f_g(x;s_x,l_x)$. Similarly, $f_n(w)$ is the normal probability density function of the normalized variate w and is transformed from the gamma probability density function $f_g(y;\ s_y,l_y)$. The information transmission relationship is simplified to:

$$T(X,Y)=T(Y,X)=-\frac{1}{2}\ln(1-\rho_{zw}^2)$$

Now, considering a triangular element (Figure - 1) formed by joining three precipitation station (i,j,k) measuring precipitation X_i, X_j, X_k respectively. The entropy $H(X_i)$ and $H(X_j)$ for the line i-j can be written as:

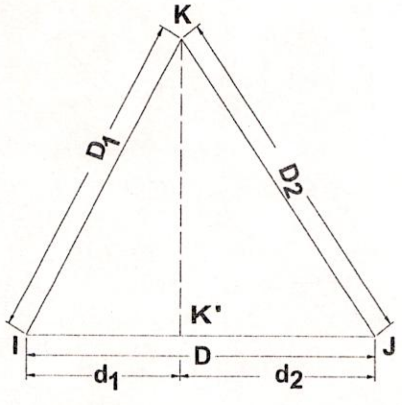


Figure – 1: Triangular element

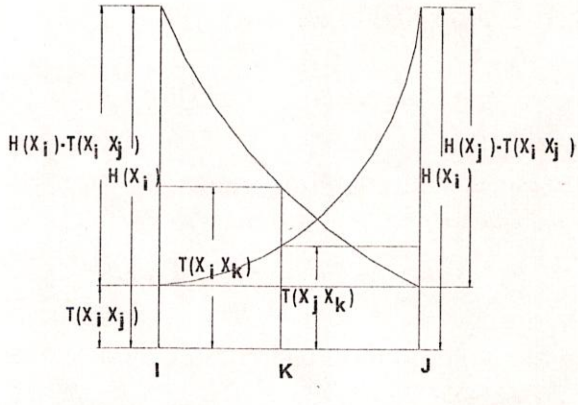


Figure – 2: Information transformation relationship

$$H(X_i) = T(X_i, X_j) + H(X_i, X_j) - H(X_j)$$

$$H(X_j) = T(X_i, X_j) + H(X_i, X_j) - H(X_i)$$

It means, the entropy $H(X_i)$ at a station, which is also the transmitted information about the variable X_i at station I can be decomposed into: (a) common information $T(X_i, X_j)$ between station pair i and j which will remain constant throughout the segment i - j and (b) the net information $H(X_i, X_j) - H(X_j)$ at i, considering station pair (i, j), which depends on the distance between the station pair and is maximum at station location. Similarly for station pair (i, k) and (j, k), the following relationships can be written:

$$H(X_i) = T(X_i, X_k) + H(X_i, X_k) - H(X_k)$$

$$H(X_k) = T(X_i, X_k) + H(X_i, X_k) - H(X_i)$$

$$H(X_j) = T(X_j, X_k) + H(X_j, X_k) - H(X_k)$$

$$H(X_k) = T(X_j, X_k) + H(X_j, X_k) - H(X_j)$$

The common information between station pairs for gamma distributed variables can be computed using the relationship derived by transforming the gamma variates to normalized variates. The net information term in the above equations is a function of distance from the stations. The variation of the net information term with distance has been found to be exponentially related as follows:

$$Y_i = ce^{-(bd_1/(D-d_1))}$$

where, Y_i is the net information ordinate at a distance d_i from the i station, c is the net information ordinate at station location and b is the coefficient of exponential decay curve, which is determined as:

Case (a) the coefficient b (originating from i to j)

$$b = [(D-d_1)/d_1][H(X_i) - T(X_i, X_j)] / [T(X_i, X_k) - T(X_i, X_j)]$$

Case (b) the coefficient b (originating from j to i)

$$b' = [d_1/(D-d_1)][H(X_j) - T(X_i, X_j)] / [T(X_j, X_k) - T(X_i, X_j)]$$

Hence, the information at any point on line i - j, say at a distance d_l from i can be interpolated as:

$$H(X_l) = T(X_i, X_j) + [H(X_i) - T(X_i, X_j)]e^{-b[d_1/(D-d_1)]} + [H(X_j) - T(X_i, X_j)]e^{-b[(D-d_1)/d_1]}$$

4. Results

Estimation of the number and location of the raingauge stations have been analysed by IS guidelines, C_v method, key stations method, correlation method and using entropy concept. The study area, Pagladiya basin having nine raingauge station inside and one station outside the catchment area of about 1507 sq. km. The major hilly portion of the catchment (about 423 sq. km) is lying on the other side of international border in Bhutan and there is no raingauge station in this part. In the hills inside Indian territory (about an extent of 42 sq. km), no station is existing and all along the foot hills, there are four stations which, obviously may not represent the rainfall characteristics of the hills. In the plain portion of the basin (area about 1042 sq. km) altogether there are ten stations and it can be assumed that data of ten stations should be sufficient to represent the rainfall in the basin. The rainfall in the catchment shows a wide variation in space and time. The location of these stations are not uniform over the catchment, therefore even if there is relatively large number of raingauge in some part of the basin, it does not give the adequate representation of the basin rainfall. Therefore, proper network design is required to sample the precipitation in the basin. The results and observations for each methods are discussed in details as follows:

IS guidelines

The catchment lies in the region of heavy rainfall zone. It is recommended as per IS4987-1968 that there should be one raingauge for every 130 sq. km of hilly area and one for every 260 sq. km of plain area. As per this guidelines, the minimum number of raingauge in the hilly area (423 sq. km) should be equal to four and in plain (1042 sq. km) another four. Therefore, the method prescribed total number of eight stations in the basin. Out of existing ten precipitation stations all are inside the catchment except Hajo. These stations are not well distributed. The north hilly part of the catchment is totally unrepresented. Though the major portion of the hills are in Bhutan, there in no station even in the 42 sq. km hilly terrain inside India. Therefore this method suggest the relocation of raingauge stations in the basin.

C_v method

The coefficient of variation evaluated for the average annual rainfall for the existing raingauge network is 0.43 and with N = 10, the probable error in the observation of the rainfall is 13.3%. To keep it within 10% of probable error 19 numbers of station is required. When the same exercise was done on the rainfall values of monsoon season only C_v has been evaluated as 0.39 and probable error as 12.4% and required number of station is 16 to keep it within 10% of probable error.

Key station network method

It has been tried to select the period so that concurrent rainfall data at maximum number of stations are available. **Table-2** gives the details of combination of stations for determination of key stations and also their correlation. Determination of key stations helps in the reduction of network. It helps in deciding which stations are to be retained in the network to attain a certain accuracy level. The decrease in root mean squares error (RMSE) with the increase in the number of stations has been shown in Table - 3. Figure - 5 shows the graphical representation of Table - 3 from which it can be seen that the RMSE decreses at faster rates till the seventh station and after the introduction of the eighth station in the combination the rate of decrease in RMSE is insignificant. It means 7 keys stations namely; Daranga, Goibargaon, Rangiya, Nagrijuli, Menaka, Hajo, Nayabasti may be considered as adequate and economical for estimation of average areal rainfall in the basin.

Thus the method suggests the exclusion of Uttarkuchi, Gerua stations from the network. This input is of vital importance when the redistribution of raingauge stations are determined. However, it is to be mentioned here that this methodology does not give any idea about the station (Masalpur) for which data is not available for the concurrent period.

Table – 2

Correlation coefficient with different combination of stations

S. No.	No. of stations in combination	Combination of stations	Key station	Correlation of key station
1	9	Daranga, Goibargaon, Rangiya, Nagrijuli, Menaka, Hajo, Nayabasti, Uttarkuchi, Gerua	Daranga	0.762385
2	8	Goibargaon, Rangiya, Nagrijuli, Menaka, Hajo, Nayabasti, Uttarkuchi, Gerua	Goibargaon	0.763693
3	7	Rangiya, Nagrijuli, Menaka, Hajo, Nayabasti, Uttarkuchi, Gerua	Rangiya	0.727482
4	6	Nagrijuli, Menaka, Hajo, Nayabasti, Uttarkuchi, Gerua	Nagrijuli	0.739477
5	5	Menaka, Hajo, Nayabasti, Uttarkuchi, Gerua	Menaka	0.776523
6	4	Hajo, Nayabasti, Uttarkuchi, Gerua	Hajo	0.82318
7	3	Nayabasti, Uttarkuchi, Gerua	Nayabasti	0.872339
8	2	Uttarkuchi, Gerua	Uttarkuchi	0.825515

Table – 3

Change in the sum of square with increase in number of stations

No of stations	RMSE
1	21.45172
2	11.64701
3	6.956837
4	6.892627
5	5.19651
6	3.812188
7	2.047332
8	1.148477

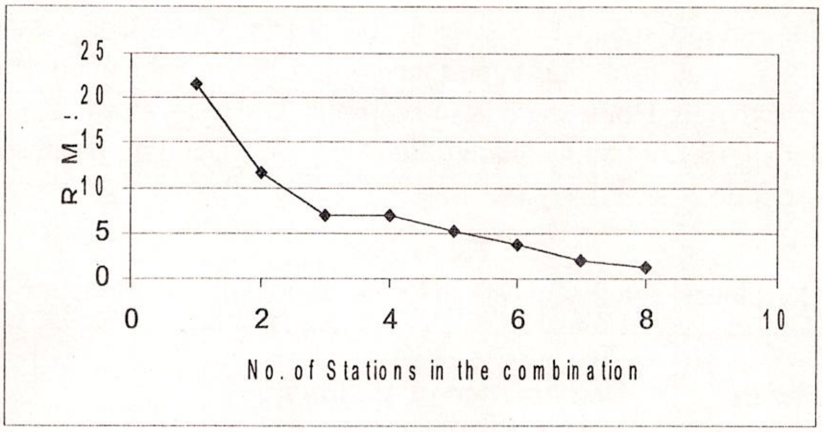


Figure – 3: Decrease in RMSE with increase in no. of stations

Spatial correlation method

For spatial correlation method daily rainfall data has been used. As there are ten stations in the basin total 45 number of combination is possible out of which there is no concurrent data available for Gerua & Masalpur and Nayabasti & Masalpur, therefore, inter-station distance and their correlation for 43 combinations has been calculated and tabulated in Table - 4. Figure - 6 shows the line of best fit of the correlation and the distance (on semi log scale). The slope of this line is equal to -0.030996314 which is equal $1/d_0$ and the Y intercept is 0.009585797 which is $\log[\rho(0)]$. It means the value of d_0 is 32.2619 km and $\rho(0)$ is 0.990460. Then the relative error (root mean square error, RMSE) has been derived and tabulated in Table - 5 which shows the decrease in relative error with increase in number of raingauge stations. This is also shown in Figure - 7. This figure shows that the gain in the accuracy of rainfall estimation is very insignificant after introduction of ninth station. It means eight stations can be recommended for the basin based on this study.

Table – 4
Inter-station distance and their correlation

Station Name	Station Name	Distance	Correlation
Gerua	Daranga	6.635933	0.281523
Goibargaon	Daranga	6.720109	0.487851
Menaka	Uttarkuchi	7.234885	0.448008
Menaka	Gerua	8.698046	0.280573
Menaka	Daranga	9.633291	0.480566
Nagrijuli	Gerua	10.21089	0.17029
Goibargaon	Gerua	10.71486	0.346644
Masalpur	Nagarijuli	12.02968	0.254124
Nagrijuli	Goibargaon	13.37846	0.355068
Nayabasti	Rangiya	13.61854	0.368347
Nagrijuli	Daranga	14.43904	0.296527
Menaka	Masalpur	14.46311	0.645556
Uttarkuchi	Gerua	14.61809	0.24556
Masalpur	Uttarkuchi	15.72948	0.476863
Menaka	Goibargaon	16.2681	0.427164
Uttarkuchi	Daranga	16.85225	0.382571
Masalpur	Daranga	17.21371	0.451507
Menaka	Nagarijuli	18.79007	0.25819
Masalpur	Goibargaon	20.59654	0.468624
Hajo	Rangiya	20.63936	0.396006
Hajo	Nayabasti	20.87076	0.281303
Masalpur	Rangiya	23.24876	0.343275
Uttarkuchi	Goibargaon	23.42803	0.415388
Uttarkuchi	Nagarijuli	23.76972	0.24444
Nagrijuli	Rangiya	25.3396	0.281992
Nagrijuli	Nayabasti	31.36899	0.214145
Gerua	Rangiya	31.99795	0.299989
Nayabasti	Gerua	34.21938	0.907909
Uttarkuchi	Nayabasti	34.91605	0.257129
Menaka	Nayabasti	36.98174	0.282765
Menaka	Rangiya	37.68367	0.422015
Uttarkuchi	Rangiya	38.2408	0.369154
Daranga	Rangiya	38.24285	0.399103
Goibargaon	Rangiya	38.67579	0.39844
Nayabasti	Daranga	40.85275	0.298737
Masalpur	Hajo	42.0461	0.248119
Goibargaon	Nayabasti	43.41789	0.335313
Hajo	Nagarijuli	45.93341	0.193902
Hajo	Gerua	51.81685	0.265757
Hajo	Uttarkuchi	55.24055	0.177415
Menaka	Hajo	56.30718	0.430246
Hajo	Daranga	58.28824	0.36881
Hajo	Goibargaon	59.21265	0.335097

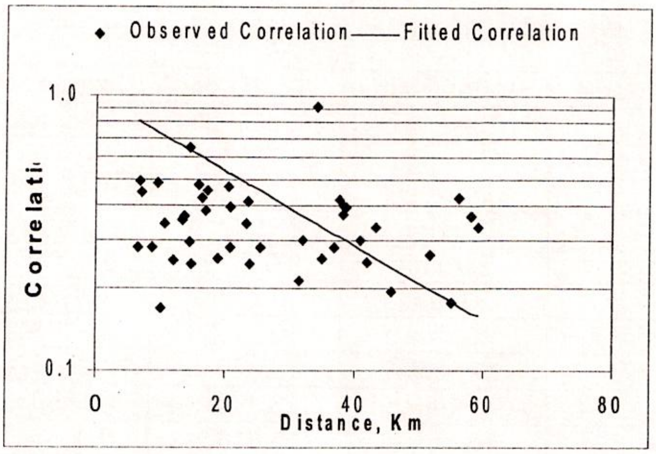


Figure – 4: Variation of correlation with inter-station distance

Table – 5

Variation of relative root mean square with number of stations

Number of station	RMSE
1	0.776178
2	0.388089
3	0.258726
4	0.194045
5	0.155236
6	0.129363
7	0.110883
8	0.097022
9	0.086242
10	0.077618
11	0.070562
12	0.064682
13	0.059706
14	0.055441
15	0.051745

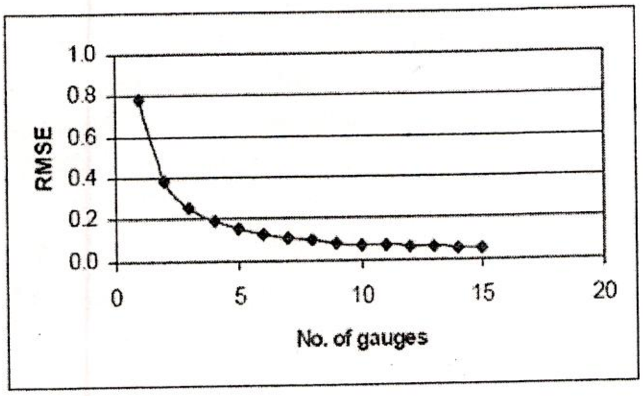


Figure – 5: Variation of relative root mean square error (RMSE) with number of stations

Entropy method

The total annual rainfall at each station for all the available years were computed and used as basic data for this method. Since this method uses some numerical approximations for calculation of entropy, a computer programs was developed specifically for it. The detailed steps followed are presented below.

Using the basic time series data, the mean and variance for each station were estimated. Substituting these, the scale and shape parameters (σ and λ) at each point were estimated. The entropy of rainfall variable measured at each station, H(X) were computed. This equation contains two functions viz. the digamma function ($\psi(X)$) and gamma function ($\Gamma(X)$). The values of these functions cannot be estimated analytically for the entire range of X. Hence the numerical approximations for these functions were used to estimate their values. The estimated values of H(X) are presented in Table - 6.

Table – 6

The estimated values of entropy, H(X) at each station location

S. No.	Name of Station	H(x)
1	Rangiya	6.5572
2	Daranga	7.6584
3	Gerua	7.1210
4	Nayabasti	7.2387
5	Goibargaon	6.6690
6	Nagrijuli	8.1847
7	Uttarkuchi	7.4961
8	Hajo	7.1237
9	Masalpur	6.6924
10	Menaka	6.8139

Using the mean and variance of the time series data, the rainfall at each station were transformed into normal variates and the correlation coefficient (ρ_{zw}) for each pair of station were computed from the normal variates. However, the correlation of some of the combinations, as explained in the previous section, could not be obtained because of paucity of concurrent data. Substituting ρ_{zw} values, the information transmission, $T(Xi;X_j)$ for all the combination of stations were computed. The $T(Xi;X_j)$ matrix are presented in Table - 7.

The entire study area was divided into seven triangles. The vertices of the triangles were decided in such a way that (a) correlation between all three stations at the vertices exist, and (b) none of triangles has any obtuse angle. The selection of triangles is shown in Figure - 6.

The information transformation at 10 points (equally spaced) along each side of all the triangles are then computed. Using these information, information transmission contours are plotted as shown in Figure - 7. The location of redundant or additional stations is determined based on these contour maps.

Table – 7

Information transmission matrix

Stn No. Stn No.	1	2	3	4	5	6	7	8	9	10
1	6.5573	0.0192	0.2154	0.0932	0.3585	0.0054	0.1767	0.0407	0.0643	0.0178
2		7.6584	0.2815	0.4118	0.2297	0.0240	0.1229	0.0501	0.0474	0.0938
3			7.1210	1.5565	0.0143	0.0819	-	0.1874	-	0.3763
4				7.2387	0.0008	0.1821	-	0.0848	-	0.6166
5					6.6690	0.0040	0.0021	0.0446	0.0004	0.2025
6						8.1847	0.0122	0.0722	0.0064	0.0374
7							7.4961	0.3331	0.0370	0.0329
8								7.1237	0.0039	0.0276
9									6.6924	1.2792
10										6.8139

- indicates non existence of correlation/ information transmission from the available data

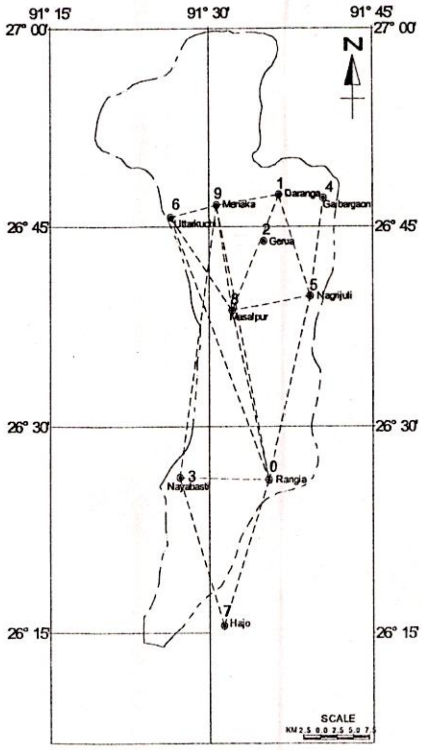


Figure – 6: Feasible triangles

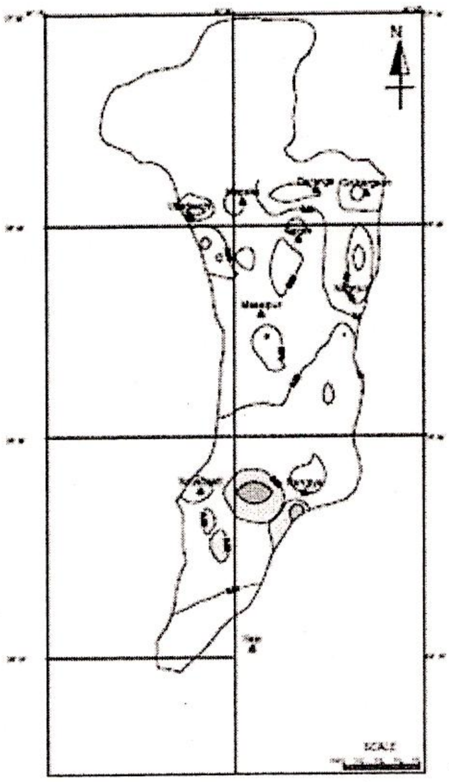


Figure – 7: Information contours