PROC. OF JSCE,
No. 230, OCT. 1974

ESTIMATION OF MEAN AREAL PRECIPITATION BY PRINCIPAL AXIS METHOD

By Kiyoshi HOSHI*

1. INTRODUCTION

Determination of the total amount of water which falls on a watershed for a given time period is basic to many hydrologic studies such as rainfall-runoff relationships in which the areal precipitation plays an important role as a watershed system input. However, there is actually very little information of sufficient detail and accuracy on how much water falls onto a watershed, primarily because the true shape of precipitation distribution is never known. An approximation must therefore be made of the mean areal precipitation from point precipitation values. The density of rain gages varies greatly from region to region, and the data so obtained represent only a scattered sample of precipitation distribution over an area. Thus, the degree of reliability of estimates for mean areal precipitation by transformation of point precipitation values depends to a large extent on whether the data at sample points represent sufficiently variations in precipitation distribution over a watershed. Determining how many rain gages are needed for an accurate estimate of mean areal precipitation is the central problem of network design. It is clearly evident that a network of rain gages should be planned so as to give an accurate picture of the areal distribution of precipitation, but whether the network of rain gages is representative or not can not be revealed by the study of estimation of mean areal precipitation alone. In order to have a full understanding of areal variability of precipitation, data must be accumulated for longer periods. And correlation analysis is a useful technique to examine a regional consistency in precipitation patterns for longer periods of time. Space variations of precipitation have been examined in terms of various time units for inferences as to a network design of rain gages.[1),2),3)] Some sophisticated techniques have been also advocated for estimating mean areal precipitation accurately.[4),5),6)] The major deficiency of the traditional techniques, however, is due to the fact that approaches for estimating mean areal precipitation and space variations of precipitation have been developed independently. For this reason, such analyses do not permit the evaluation of the effect of areal variabilities on the reliability of areal precipitation estimates. In order to have a better idea as to how far areal variability of precipitation introduces uncertainty in relation to the estimation of mean areal precipitation, an alternative mathematical scheme should be devised, which must contain a framework that is sufficiently flexible to respond to all needs with respect to the evaluation of the reliability of estimates and a network design of rain gages.

To clarify some problems involved in estimating mean areal precipitation, the mathematical features of various models are examined. The isohyetal method is the most accurate one for computing mean areal precipitation for individual precipitation events. However, this method is extremely time consuming; it is not adapted to objective computational routines; and a great deal of personal judgement is left to the individual who draws and interpolates isohyets, especially when the number of rain gages is small and the shape of the isohyetal pattern is not definitely known. Furthermore, the chief disadvantage is that this method can not be employed to examine the characteristics of space variations of precipitation. For practical convenience, the Thiessen method has been made use of almost exclusively for estimating mean areal precipitation. To date, emphasis has been placed only on its simplicity and the uniqueness of the Thiessen polygon network, and no consideration has been given to the examination of areal variability of precipitation before computing mean depths of precipitation

* Research Associate, Department of Civil Engineering, Hokkaido University.

over a watershed. The degree of reliability of estimates can not be measured in the Thiessen method, because the effective area assumed to be represented by each rain gage is always constant, independent of the choice of time unit and the feature of space variations of precipitation. The Thiessen method assumes that the precipitation at any site can be applied halfway to the next rain gage in any direction. In general, it is quite well understood that if the areal distribution of precipitation is uniform and the rain gages are evenly distributed within a watershed, this method will yield fairly accurate results, but some techniques should be introduced to test the Thiessen hypothesis, because the precipitation is unevenly distributed both in time and in space. The Thiessen method does not offer information essential to the assessment of existing networks of rain gages, unless some provision is introduced to account for space variations of precipitation.

As shown in the above discussion, no approach has been developed to an extent sufficient to handle completely the actual problems involved in the studies on the estimation of mean areal precipitation and space variations of precipitation. Correlation analysis indicates that within a given watershed, the cross correlations among precipitation data at different rain gages are unlikely to be zero, and the degree of interdependences among them has a varied seasonal pattern. Assuming that the cross correlations among rain gages are available information to account for areal variability of precipitation, incorporation of the cross correlations in the estimation of mean areal precipitation makes it possible to establish quantitative relationships between the areal variability of precipitation and the reliability of estimates. As a result, the adequacy or otherwise of a network of rain gages is to be assessed by examination of the reliability obtained. The objective of this paper is to provide a standard method for estimating mean areal precipitation and to elucidate some of the factors which contribute to prediction of the network design of rain gages. The mean areal precipitation can be approximated through the formulation of a linear transformation of point precipitation values. Under the assumption that the effective area represented by the rain gage varies with areal variability of precipitation according to the season of the year, its area expressed as a percentage of the whole area is determined by the principal axis method. The use of the proposed method for estimating mean areal precipitation allows areal variabilities of precipitation to be followed readily in various networks. And moreover, there are considerable theoretical advantages to be gained in the degree of reliability of estimates. The measure of reliability by the model in describing mean depths of precipitation over a watershed is given by the percentage contribution of the obtained maximum variance to the total variance. An alpha coefficient is introduced to examine representativeness of the network of rain gages in the estimation of mean areal precipitation. This parameter is easily calculated from a two-way analysis of variance. An alpha coefficient can be used to determine how many rain gages are required to estimate mean areal precipitation with the desired reliability. Therefore, the derivation of a high alpha coefficient is of prime importance in determining the adequacy of rain gage networks. The proposed model can be easily extended to situations where mean areal precipitation has to be computed for the whole watershed comprising a greater number of subwatershed systems.

2. PROBLEMS INVOLVED IN THE ESTIMATION OF MEAN AREAL PRECIPITATION

A mathematical model should be devised to give a closer approximation to the total volume of precipitation falling on a watershed, because the true shape of precipitation distribution is never known. If the areal precipitation can be represented by a linear transformation of precipitation values at a number of sample points in a watershed, approximating the unknown areal precipitation corresponds to the problem of how to evaluate the effective area which each rain gage is assumed to represent. If the area expressed as a percentage of the total area of a watershed is determined in an objective manner, the mean depth of precipitation over a watershed is the sum of point precipitation amounts, each multiplied by its assigned percentage of area. A linear model is expressed as follows:

$$R_j=\sum_{i=1}^{n}\beta_i x_{ij} \quad (j=1, 2, 3, \cdots, N) \quad \cdots\cdots(1)$$

where R_j, mean areal precipitation in the jth period; β_i, effective area expressed as a percentage of the total area at the ith gage; x_{ij}, observed precipitation depth by the ith gage in the jth period; n, number of rain gages; N, number of observations. The weighting factors at the rain gages must fulfill the following conditions;

$$\sum_{i=1}^{n}\beta_i=1 \quad \cdots\cdots(2)$$

and

$$\beta_i > 0 \quad (i=1, 2, 3, \cdots, n) \cdots\cdots\cdots\cdots\cdots\cdots (3)$$

Some criterion is needed to evaluate the gage weightings to satisfy the required conditions of Eqs. (2) and (3). The Thiessen method assumes that the precipitation at any site is best represented by the gage nearest to it, and accordingly the effective area assumed to be controlled by each rain gage is determined only from the configuration of rain gages. Correlation-distance relationships[7),8)] indicate that the Thiessen hypothesis is not necessarily valid, and an alternative approach is required to provide a more accurate weighting method. It is well known that the variability of precipitation decreases with increase in the time unit being considered, and there is a marked seasonal variation in precipitation types. In the Thiessen method, however, the size of the effective area at the rain gage is independent of the choice of time unit and the feature of space variations of precipitation. Although a fundamental requirement of mean areal precipitation is a knowledge of the reliability of estimates, quantitative assessments of the results by the Thiessen method have been very limited. The inability to evaluate the degree of reliability of estimates is attributed to the fact that factors to account for areal variability of precipitation are not introduced to determine the area assumed to be represented by the rain gage in the Thiessen method. The Thiessen method may be a reasonable assumption in some areas where a uniform distribution of precipitation prevails, but this may not always be correct. Something more than a mere qualitative assessment as to the reliability of estimates is therefore desirable.

This paper suggests improvements in the present techniques for estimating mean areal precipitation by combining correlation analysis. As a distinguished feature of precipitation data, precipitation values at the rain gages in a watershed are more or less correlated with each other, and the degree of associations among the rain gages varies widely from season to season. The use of correlation analysis for estimating mean areal precipitation will make it more flexible than the Thiessen method to evaluate the effect of areal variability of precipitation on the reliability of estimates. As a result, this measure of reliability can be used to examine unrepresentativeness of the network of rain gages. An underlying assumption in this paper is that the effective area assumed to be controlled by the rain gage depends on the feature of space variations of precipitation. A method of determining weighting factors in Eq. (1) is developed through the use of a variance-covariance matrix of precipitation, which is available information to account for space variations of precipitation. In areas where there is a marked seasonal variation in precipitation patterns, better results can be obtained by using seasonal values rather than the Thiessen method. Equation (1) is of the same form as the Thiessen model, so that the proposed model can be used to test the validity of the Thiessen hypothesis. If the difference of estimates between the proposed and Thiessen methods is reasonably small, a measure of reliability given by the proposed method can be used as that by the Thiessen method.

An approach to the estimation of mean areal precipitation and its reliability through the principal axis method demonstrates that the desired solution with regard to the weighting factors, which must meet the required conditions of Eqs. (2) and (3), is given by an eigenvector corresponding to the largest eigenvalue of a variance-covariance matrix of precipitation among the rain gages.

3. MODEL EQUATION

This paper examines the possibility of a technique for computing mean areal precipitation as an alternative to the commonly used technique. The technique is analogous to the principal axis method of multivariate analysis. Obviously, some form of modification is needed to give estimates for mean areal precipitation. Assume there are n rain gages within a watershed, each of which has an N-year record of precipitation depths. Let x_{ij} denote the jth observation on the ith gage. Geometrically viewed, N points can be plotted in the n-dimensional hyperspace by taking $x_1, x_2, \cdots, x_n$ as co-ordinate axes. The principal axis method involves the rotation of co-ordinate axes to a new frame of reference so that this new rotated axis might be preferable for the purpose of interpreting the basic dimension of the domain measured by N observations. The objective solution in terms of finding the new reference axis is obtained by projecting all points perpendicularly on the principal axis along which the sum of squares of distances becomes minimum. Fig. 1 clarifies the meaning of the principal axis method.

An equation of the principal axis passing through an arbitrarily fixed point A: $(b_1, b_2, \cdots, b_n)$ and with the direction cosines of $(w_1, w_2, \cdots,$

w_n) is expressed as follows:

$$f:\ \frac{x_1-b_1}{w_1}=\frac{x_2-b_2}{w_2}=\cdots=\frac{x_n-b_n}{w_n}\ \cdots(4)$$

where b_i and w_i are unknown parameters.

A plane perpendicular to the line expressed by Eq. (4), passing through a given point C: $(x_{1j}, x_{2j}, \cdots, x_{nj})$ is given by

$$\sum_{i=1}^{n} w_i(x_i-x_{ij})=0 \cdots\cdots(5)$$

or

$$p_j=\sum_{i=1}^{n} w_i x_{ij}=\sum_{i=1}^{n} w_i x_i \quad (j=1, 2, 3, \cdots, N) \cdots\cdots(6)$$

It is of interest to note that Eq. (6) is of similar form as shown in Eq. (1). If a set of positive direction cosines for the rain gage is available, these coefficients can be transformed to yield the gage weightings with their sum equal to unity as follows:

$$\beta_i=w_i\Big/\sum_{i=1}^{n} w_i \cdots\cdots(7)$$

The mean depth of precipitation over a watershed is the sum of the individual gage amounts, each multiplied by its weighting factor of Eq. (7).

For convenience, p_j expressed by Eq. (6) is referred to as areal precipitation, and the major effort herein is devoted to the problem of how to evaluate the parameter, w_i in a linear model. The direction cosine, w_i is determined in such a way that when N points are projected perpendicularly onto the particular line of Eq. (4), the sum of squares of distances is minimum. As an index measuring how much the observed record x_{ij} deviates from the principal axis in the jth period, the distance, D_j as shown in Fig. 1 is calculated. This distance is calculated by elementary geometry as follows:

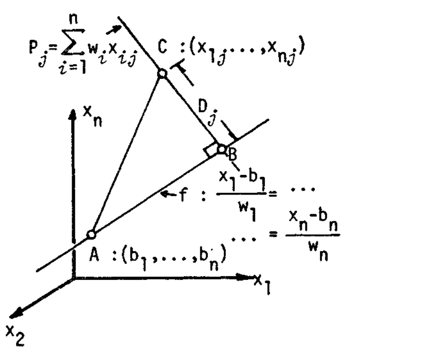

Fig. 1 Geometrical Interpretation of Model Equation.

$$D_j^2=\overline{BC}^2=\overline{AC}^2-\overline{AB}^2$$

$$=\sum_{i=1}^{n}(x_{ij}-b_i)^2-\left[\sum_{i=1}^{n} w_i(x_{ij}-b_i)\right]^2 \cdots(8)$$

The sum of squares of perpendiculars over N periods, which is denoted by D^2, is given by

$$D^2=\sum_{j=1}^{N} D_j^2=\sum_{j=1}^{N}\sum_{i=1}^{n}(x_{ij}-b_i)^2-\sum_{j=1}^{N}\left[\sum_{i=1}^{n} w_i(x_{ij}-b_i)\right]^2 \longrightarrow \text{Mini}\cdots(9)$$

Since the parameter w_i of Eq. (4) gives the direction cosine for the n variates, the required condition with respect to the plane expressed by Eq. (6) is given by

$$\sum_{i=1}^{n} w_i^2=1 \cdots\cdots(10)$$

And in order to determine the physically realizable gage weightings for estimating mean areal precipitation, the following condition must be fulfilled:

$$w_i>0 \quad (i=1, 2, 3, \cdots, n) \cdots\cdots(11)$$

The mean and variance of areal precipitation expressed by Eq. (6) over N periods are given by

$$m_p=\sum_{i=1}^{n} w_i \bar{x}_i \cdots\cdots(12)$$

$$\sigma_p^2=\sum_{j=1}^{N}(p_j-m_p)^2/N$$

$$=\sum_{j=1}^{N}\left[\sum_{i=1}^{n} w_i(x_{ij}-\bar{x}_i)\right]^2\Big/N \cdots\cdots(13)$$

$$=\sum_{i=1}^{n}\sum_{k=1}^{n} w_i w_k \sigma_{ik} \cdots\cdots(14)$$

where

$$\bar{x}_i=\sum_{j=1}^{N} x_{ij}\Big/N \cdots\cdots(15)$$

$$\sigma_{ii}=\sigma_i^2, \quad \sigma_{ik}=\sigma_{ki} \quad (i, k=1, 2, 3, \cdots, n) \cdots\cdots(16)$$

m_p and σ_p^2, mean and variance of areal precipitation over N periods, respectively; $\bar{x}_i$ and σ_i^2, mean and variance of precipitation depths at the ith rain gage, respectively; σ_{ik}, covariance of precipitation depths at gages i and k.

It is now required to determine the parameters of b_i and w_i. The minimum of the quantity, D^2 with respect to b_i and w_i is determined by solving the following simultaneous equations:

$$\partial D^2/\partial b_i=0 \quad \text{and} \quad \partial D^2/\partial w_i=0 \cdots\cdots(17), (18)$$

From Eq. (9), Eq. (17) can be expanded in the following way

$$\sum_{j=1}^{N}(x_{ij}-b_i)=\sum_{j=1}^{N} w_i \sum_{i=1}^{n} w_i(x_{ij}-b_i) \cdots\cdots(19)$$

and by using Eqs. (12) and (15), the resulting equation is

$$N(\bar{x}_i-b_i)=w_iN\left(m_p-\sum_{i=1}^{n}w_ib_i\right) \quad\cdots\cdots\cdots(20)$$

Without loss of generality, both sides of Eq. (20) are equal to zero in the case of b_i being equal to $\bar{x}_i$ with help of Eq. (12). In other words, the desired parameter of b_i is equal to the mean of precipitation depths at the ith rain gage. When the parameter of b_i is replaced by the mean value of $\bar{x}_i$ in Eq. (9), the quantity of D^2 can be expressed as follows:

$$D^2/N=\sum_{i=1}^{n}\sigma_i^2-\sigma_p^2 \longrightarrow \text{Mini}\cdots\cdots\cdots\cdots\cdots(21)$$

$$=\sum_{i=1}^{n}(1-w_i^2)\sigma_i^2-\sum_{i=1}^{n}\sum_{k=1}^{n}w_iw_k\sigma_{ik} \quad (i\neq k) \quad\cdots\cdots\cdots\cdots(22)$$

In the right-hand side of Eq. (22) the first term represents the effect of point variations of precipitation at the individual gages, while the second term accounts for areal variabilities of precipitation among the rain gages. Therefore, the smaller the differences between point and areal variations of precipitation, the smaller is the sum of squares of distances when all points are projected perpendicularly onto the principal axis. And the resulting plane expressed by Eq. (6) will produce a high degree of reliability of estimates for areal precipitation. Since the quantity of the first term in the right-hand side of Eq. (21) is constant with data available at n gages in N periods, there is no need to proceed to the scheme of Eq. (18). Instead, the parameter of w_i has to be chosen in such a way as to make the quantity, σ_p^2 a maximum.

For the numerical elaboration of a maximization scheme, Eqs. (6), (10), and (14) are written in matrix notation:

$$\boldsymbol{P}=\boldsymbol{W}\boldsymbol{X} \quad\cdots\cdots\cdots\cdots(23)$$

$$\boldsymbol{W}\boldsymbol{W}^T=1 \quad\cdots\cdots\cdots\cdots(24)$$

$$\sigma_p^2=\boldsymbol{W}\boldsymbol{Q}\boldsymbol{W}^T \quad\cdots\cdots\cdots\cdots(25)$$

where

$$\boldsymbol{Q}=[\boldsymbol{X}-\bar{\boldsymbol{X}}\boldsymbol{1}][\boldsymbol{X}-\bar{\boldsymbol{X}}\boldsymbol{1}]^T/N \quad\cdots\cdots\cdots\cdots(26)$$

$\boldsymbol{P}$, row vector with N elements; $\boldsymbol{W}$, row vector with n elements; $\boldsymbol{X}$, matrix of observed data with n rows and N columns; $\bar{\boldsymbol{X}}$, column vector of mean values at the gages with n elements; $\boldsymbol{1}$, row vector of unities with N elements; $\boldsymbol{Q}$, variance-covariance matrix of observed data between the gages with n rows and n columns; T, transpose of matrix as a superscript.

Following the maximum variance criterion as described, main attention is directed to the system in which the quantity expressed by Eq. (25) must be maximized under the condition of Eq. (24). Subject to a restriction that is introduced by use of a Lagrange multiplier, the resulting equation is written as

$$[\boldsymbol{Q}-\lambda\boldsymbol{I}]\boldsymbol{W}^T=\boldsymbol{0} \quad\cdots\cdots\cdots\cdots(27)$$

where $\boldsymbol{I}$, identity matrix; $\boldsymbol{0}$, null vector; λ, Lagrange multiplier.

Premultiplying Eq. (27) by $\boldsymbol{W}$ and using Eqs. (24) and (25) give

$$\lambda=\boldsymbol{W}\boldsymbol{Q}\boldsymbol{W}^T=\sigma_p^2 \quad\cdots\cdots\cdots\cdots(28)$$

Equation (27) indicates that λ and $\boldsymbol{W}$ are equal to an eigenvalue and corresponding eigenvector of a variance-covariance matrix, respectively. Equation (28) means that an eigenvalue, λ, is precisely the variance of areal precipitation as expressed by Eq. (25), which is to be maximized. A necessary and sufficient condition for the nontrivial solution of Eq. (27) is given by

$$\det(\boldsymbol{Q}-\lambda\boldsymbol{I})=0 \quad\cdots\cdots\cdots\cdots(29)$$

For a known matrix of $\boldsymbol{Q}$, a characteristic equation of Eq. (29) gives, in general, n roots in λ. According to the maximum variance criterion, the desired root in the characteristic equation is the largest one, which gives the maximum variance of the particular linear combination of precipitation depths at the gages. And the eigenvector corresponding to the largest eigenvalue provides the solution for the parameter of w_i.

However, there is some question of whether a set of positive values with regard to w_i can be obtained. The following theorem concerning the fundamental properties of a square matrix is useful to answer the above question:

Theorem; "If a real symmetric matrix is nonnegative definite, an eigenvalue is positive and the elements of the eigenvector corresponding to the largest eigenvalue are all positive."

As for a variance-covariance matrix of precipitation values among the rain gages, all elements of this matrix are positive in most cases. Therefore, it follows from the above discussion that a linear combination of point precipitation depths for estimating areal precipitation is determined in such a way that the desired maximum variance is equal to the largest root of the characteristic equation and the gage weightings are associated with the eigenvector corresponding to the largest eigenvalue of a variance-covariance matrix.

Correlation coefficients between estimates of areal precipitation and precipitation depths at different gages are expressed by use of Eqs. (27)

and (28) as follows:

$$A=[\sigma_p^{-1}(P-m_p1)][E^{-1}(X-\bar{X}1)]^T/N \cdots(30)$$

$$=\sqrt{\lambda_{\max}}\,WE^{-1} \cdots\cdots(31)$$

where A, row vector of correlation coefficients between areal precipitation depths and precipitation depths at the gages with n elements; E, diagonal matrix of standard deviations of precipitation at the gages with n rows and n columns; $\lambda_{\max}$, largest eigenvalue of Eq. (29); E^{-1}, inverse matrix of E.

Correlation coefficients can be used to measure the relative contributions of point precipitation to the prediction of areal precipitation as a check on the representativeness of various networks of rain gages.

It is required to measure the degree of reliability of estimates for areal precipitation, which is related with the magnitude of the maximum variance obtained. From a property with respect to n eigenvalues of Eq. (29), the following expressions are derived:

$$\sum_{i=1}^{n}\lambda_i=\sum_{i=1}^{n}\sigma_i^2 \cdots\cdots(32)$$

$$\sigma_r^2=\sum_{i=1}^{n}\sigma_i^2-\lambda_{\max}=\sum_{i=1}^{n}\sigma_i^2-\sigma_p^2 \cdots\cdots(33)$$

$$P_{\max}=100\,\lambda_{\max}\Big/\sum_{i=1}^{n}\sigma_i^2 \cdots\cdots(34)$$

where λ_i, eigenvalues of Eq. (29); σ_r^2, sum of remaining variances except the maximum variance; $P_{\max}$, percentage contribution of the maximum variance obtained.

Equation (32) indicates that the sum of n eigenvalues is equal to the total variance of precipitation depths at n gages. Thus, the remaining variations unexplained by the maximum variance are given by the quantity, σ_r^2 as shown in Eq. (33). Equation (34) states that the percentage contribution of the maximum variance to the total variance can be used as a measure of the reliability of estimates. In other words, the sum of remaining variances is regarded as uncertainty about the feature of areal variability of precipitation. From Eq. (33), Eq. (21) is reduced to

$$D^2/N=\sigma_r^2 \cdots\cdots(35)$$

Equation (35) can be used as the standard error of estimates for areal precipitation. In order to clarify the foregoing ideas, consider the special case of $\sigma_i^2=\sigma_{ij}=\sigma^2$ in a variance-covariance matrix $(i, j=1, 2, \cdots, n)$. As a solution for this case, Eq. (29) gives $\lambda_{\max}=n\sigma^2$ and $\lambda_i=0$ $(i=2, 3, \cdots, n)$. From Eq. (27), the elements of an eigenvector, w_i are all equal to $1/\sqrt{n}$. As a result, the maximum variance is equal to the total variance. Presumably such a situation would not arise in hydrologic practice, but this example suggests that the less areal variability of precipitation, the larger is the percentage contribution of the maximum variance and accordingly a degree of the reliability of estimates is larger.

The degree of reliability of estimates for areal precipitation by transformation of precipitation depths at the rain gages depends largely on whether the selected gages represent sufficiently variations in precipitation distribution over a watershed. The derivation of a useful index is, therefore, of prime importance in determining the adequacy of the network of rain gages. If the weighting factors of w_i to satisfy the required conditions are derived, the weighted elements of point precipitation depth are represented as follows:

$$Y=\begin{pmatrix} w_1x_{11} & w_1x_{12} & w_1x_{13} & \cdots & w_1x_{1N} \\ w_2x_{21} & w_2x_{22} & w_2x_{23} & \cdots & w_2x_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ w_nx_{n1} & w_nx_{n2} & w_nx_{n3} & \cdots & w_nx_{nN} \end{pmatrix} \cdots\cdots(36)$$

The areal precipitation depth for the particular period is equal to the sum of elements in each column of the above matrix. It is shown in the analysis of variance that the linear model by Eq. (6) corresponds to a partitioning of the total sum of squares of deviations from the mean involved in the weighted elements of Y into three parts, two of which are ascribed to differences among the gage means and to differences among the period means, respectively, while the third measures the variations of the residuals. This partitioning is represented in a two-way analysis of variance as illustrated in Table 1.

Table 1 Analysis of Variance for Data from n Gages and N Periods.

	Sum of Squares	Degrees of Freedom
Total	$S_t=\sum_{i=1}^{n}w_i^2\sum_{j=1}^{N}x_{ij}^2-\frac{N}{n}m_p^2$	$Nn-1$
Gages	$S_n=N\sum_{i=1}^{n}(w_i\bar{x}_i)^2-\frac{N}{n}m_p^2$	$n-1$
Periods	$S_N=\frac{1}{n}\sum_{j=1}^{N}\left(\sum_{i=1}^{n}w_ix_{ij}\right)^2-\frac{N}{n}m_p^2$	$N-1$
Residual	$S_r=S_t-S_n-S_N$	$(N-1)(n-1)$

From this table, an index for measuring homogeneity of the selected gages in estimating mean depths of precipitation over a watershed is given[9] by

$$\alpha=1-\left[\frac{S_r}{(N-1)(n-1)}\Big/\frac{S_N}{N-1}\right]$$
$$=\frac{n}{n-1}\left(1-\frac{S_t-S_n}{nS_N}\right)\cdots\cdots\cdots\cdots(37)$$

where α is an alpha coefficient.

An alpha coefficient can be used to examine the adequacy of a network of rain gages. If requirements for mean depths of precipitation over a watershed can be expressed in terms of the allowable level of an alpha coefficient, network density and configuration of rain gages can be efficiently designed to meet these requirements by this parameter.

A technique for estimating precipitation depth over a single watershed is developed in the above discussion. This theory can be extended to practical situations where mean depths of precipitation have to be estimated for a larger watershed consisting of a greater number of subwatersheds. If precipitation depths for each subwatershed are independently estimated, determination of the parameters in a restricted linear model leads naturally to the equation for the solution of the maximum variance and corresponding eigenvector of a variance-covariance matrix of areal precipitation depths among the subwatersheds. Interest is now confined to different subwatersheds k and m for which several quantities are presented. Estimates of precipitation for each subwatershed are given by

$$\boldsymbol{P}_i=\boldsymbol{W}_i\boldsymbol{X}_i\cdots\cdots\cdots\cdots(38)$$

where $\boldsymbol{P}_i$, row vector of estimates of precipitation for subwatershed i with N elements; $\boldsymbol{W}_i$, row vector of direction cosines at the n_i gages with n_i elements; $\boldsymbol{X}_i$, matrix of observed data at the gages within a subwatershed with n_i rows and N columns; n_i, number of gages in subwatershed i; N, number of observations.

The covariance of areal precipitation depths between subwatersheds k and m is given by

$$\sigma_{km}=[\boldsymbol{P}_k-m_k\boldsymbol{1}][\boldsymbol{P}_m-m_m\boldsymbol{1}]^T/N$$
$$=\boldsymbol{W}_k\boldsymbol{X}_k\boldsymbol{X}_m^T\boldsymbol{W}_m^T/N-\boldsymbol{W}_k\bar{\boldsymbol{X}}_k\bar{\boldsymbol{X}}_m^T\boldsymbol{W}_m^T\cdots(39)$$
$$=\boldsymbol{W}_k\boldsymbol{Q}_{km}\boldsymbol{W}_m^T\cdots\cdots\cdots\cdots(40)$$

where

$$\boldsymbol{Q}_{km}=[\boldsymbol{X}_k-\bar{\boldsymbol{X}}_k\boldsymbol{1}][\boldsymbol{X}_m-\bar{\boldsymbol{X}}_m\boldsymbol{1}]^T/N\cdots\cdots(41)$$

σ_{km}, covariance of areal precipitation depths between subwatersheds k and m; m_k, mean value of areal precipitation depths is subwatershed k; $\bar{\boldsymbol{X}}_k$, column vector of mean values in precipitation at the gages in subwatershed k with n_k elements; $\boldsymbol{Q}_{km}$, covariance matrix of observations among gages in subwatersheds k and m.

As shown in Eq. (39), the relevant information from observed data among the gages can be used to estimate precipitation depths for the combined subwatersheds. Therefore, the removal and addition of the rain gages for estimates of areal precipitation provide only minor changes in a variance-covariance matrix of precipitation depths at the gages. When the required sub-matrices for the various networks of gages are arbitrarily chosen from the entire matrix, areal precipitation can be estimated and its reliability can be evaluated for the chosen network. A correlation coefficient of areal precipitation depths between the subwatersheds is given by

$$a_{km}=\sigma_{km}/\sigma_k\sigma_m\cdots\cdots\cdots\cdots(42)$$

where a_{km}, correlation coefficient of estimates of areal precipitation between subwatersheds k and m; σ_k, standard deviation of areal precipitation depths in subwatershed k.

4. CASE STUDY

In order to test the validity of the model equation developed in the preceding section, the Ishikari River Watershed in Hokkaido was chosen for investigation. This watershed covers an area of 12 700 km². Fig. 2 shows the location of the selected watershed and twenty-nine rain gages with the station numbers assigned. The station

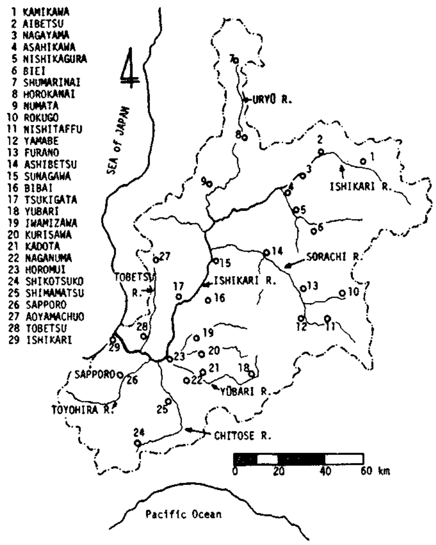

Fig. 2 Location Map of the ISHIKARI RIVER WATERSHED and the Selected Raingage Stations.

numbers in this figure correspond to those in the subsequent results shown in several figures and tables. Data were collected from the rain gages of the Meteorological Observatory at Sapporo in Hokkaido. The areal variability of precipitation varies itself according to the time interval taken, the variability on an annual basis being much less than on a monthly or daily basis. Since the study is mainly concerned with the provision of natural water supplies in the water-resource development planning, the time unit taken is the month. The data of monthly total precipitation for the period of 21 years from 1951 to 1972 were used for the present study, as the network was completely in operation. A water year is the period from November to October, because precipitation occurs as a form of snow in winter months in Hokkaido.

Taking into account the stochastic nature of the precipitation event, the estimation of mean areal precipitation will be affected by two factors: (1) the number of rain gages. (2) the length of available data. In estimating mean areal precipitation the question arises as to the required density of rain gage networks or representativeness of the selected gages. It has been always acknowledged that it is necessary to establish a greater network density in mountainous areas for the same accuracy of estimation of mean areal precipitation than in flat ones. Mountains not only affect the quantity and distribution of precipitation, but also the areal variability. Quantitative assessments of the effects of topography and meteorology on the areal variability have been very limited,[10] because these effects can not be revealed by a study of precipitation records alone. The establishment of too dense a network will be avoided in view of the economics of operating a network. A less dense network of rain gages gives only rough information about the actual precipitation falling on a watershed. Thus, a decision as to the optimum density of a network depends on scientific and practical considerations. Another difficulty encountered in setting up hydrologic investigations is the problem of whether the data for a long-term period are available. When relating the gage weightings to the estimation of mean areal precipitation, the parameter stability must be taken into consideration. A time series analysis indicates that the predominant periodicity of a year is clearly discerned in a correlogram of monthly precipitation values. This fact suggests that if the seasonal variations of precipitation do not widely vary from year to year, it is possible to get stable values of the gage weighting even from a small number of records. In the present study an examination is made concerning two cases which meet the above-mentioned requirements.

As a check on the representativeness of various networks of rain gages in calculating monthly total precipitation over the Ishikari River Watershed, the total watershed area was divided into four subwatersheds according to the natural watershed areas. Table 2 shows the station numbers in the entire network and a less dense network within these four subwatersheds. Over the total watershed area the entire network consists of 29 gages, while 18 gages are selected in the subnetwork. The subnetwork of 18 gages as shown in Table 2 remained unchanged to compute the mean depth of precipitation over the Ishikari River Watershed in every month.

Table 2 Four Divisions of the ISHIKARI RIVER WATERSHED and the Selected Raingage Networks.

Subwatershed No.	Raingage Number in the Entire Network	Raingage Number in the Subnetwork
I	1, 2, 3, 4, 5, 6, 7, 8, 9	2, 4, 6, 8, 9
II	10, 11, 12, 13, 14, 15, 16, 17	11, 13, 14, 15, 16
III	18, 19, 20, 21, 22, 23, 24, 25	19, 20, 22, 23, 25
IV	26, 27, 28, 29	27, 28, 29

The total sample size of 21 years gave positive values for the direction cosine in the model equation in both cases of the entire networks and subnetworks for these four subwatersheds in every month, so that the equation parameters could be transformed to yield the gage weightings by use of Eq. (7). The sum of the gage weighting multiplied by the precipitation depth at each gage gives the mean depth of monthly precipitation over each subwatershed. The mean depth of monthly precipitation over the entire Ishikari River Watershed was calculated by a linear combination of mean depths of precipitation over four subwatersheds, subject to some restrictions to be satisfied.

For subwatershed I a few examples of monthly variations in the statistical quantities are presented. Figs. 3 and 4 show the percentage contribution of the maximum variance and an alpha coefficient in the entire network of 9 gages and two subnetworks of 5 gages, respectively. In these figures, subnetwork A consists of stations

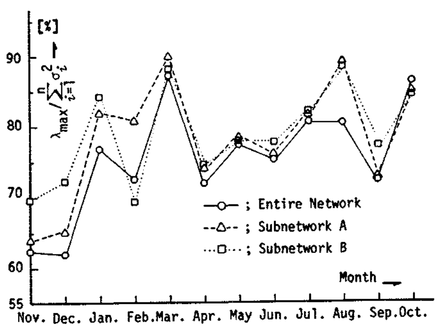

Fig. 3 Changes of the Maximum Eigenvalue due to Different Density of Raingage Networks in the Watershed I.

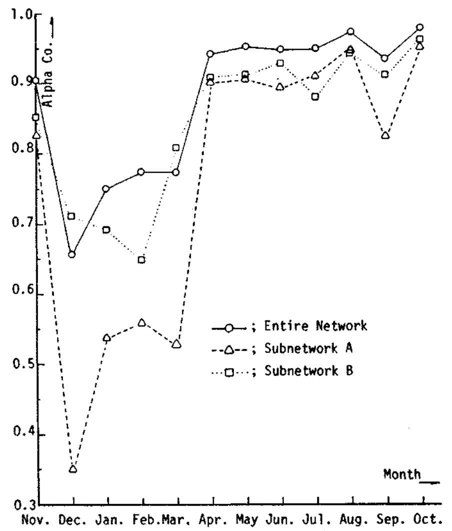

Fig. 4 Changes of Alpha Coefficient due to Different Density of Raingage Networks in the Watershed I.

(1, 4, 6, 7, 9), while subnetwork B consists of the stations indicated in Table 2. The maximum variance accounts for more than 70% of the total variance for the entire network except in November and December. As stated before, the magnitude of the maximum variance depends on the elements of a variance-covariance matrix. The less areal variability of precipitation, the larger is the percentage contribution of the maximum variance. A visual inspection of correlation coefficients matrices among the rain gages shows that off-diagonal elements have small values in winter months, compared with larger values in summer months. This suggests that there are considerable local variations within a watershed due to the nature of topography during the winter. In contrast, meteorologic factors predominate over topographic features during the summer, producing uniform precipitation over an area. The small value of the percentage contribution in November and December is attributed to the large areal variability of precipitation. A greater number of rain gages are required to give estimates of mean areal precipitation with a high reliability in winter months. Alpha coefficients are small during the winter, while they have larger values during the summer. Fig. 4 also shows that the magnitudes of an alpha coefficient are largely affected by the selected rain gages. Careful consideration has to be given to the selection of gages for estimating areal precipitation with a smaller number of the gages in winter months. Fig. 5 gives correlation coefficients between estimates of mean areal precipitation and point precipitation depths for the entire network of 9 gages in subwatershed I. If the isohyets cover a fairly large area, homogeneous situations would occur in most cases and accordingly pairs of gages would have about the same magnitude in the correlation coefficients. Fig. 5 makes it clear that there are large differences in the monthly variations of correlation coefficients. Correlation coefficients are distributed in a wide range in December, resulting in the percentage contribution and an alpha coefficient having small values in Fig. 3 and Fig. 4, respectively. The marked decrease of an alpha coefficient for subnetwork A in December is mainly due to the large difference of correlation coefficients between gages 1 and 7. Correlation coefficients are large and become rather uniform

Table 3 Correlation Coefficient of Mean Areal Precipitation Depths between Four Subwatersheds.

	Entire Network of 29 Raingages		Subnetwork of 18 Raingages		
		November		November	
I	1.00	1.00 .876 .580 .715	1.00	1.00 .859 .606 .729	I
II	.762 1.00	1.00 .695 .792	.771 1.00	1.00 .705 .769	II
III	.512 .770 1.00	1.00 .863	.560 .617 1.00	1.00 .837	III
IV	.506 .595 .648 1.00	1.00	.524 .513 .665 1.00	1.00	IV
	February		February		
		May		May	
I	1.00	1.00 .865 .365 .583	1.00	1.00 .858 .529 .572	I
II	.876 1.00	1.00 .579 .759	.875 1.00	1.00 .608 .613	II
III	.762 .928 1.00	1.00 .847	.744 .881 1.00	1.00 .917	III
IV	.800 .948 .930 1.00	1.00	.812 .942 .906 1.00	1.00	IV
	August		August		

Table 4 Changes of Subwatershed Weighting due to Different Density of Raingage Networks in the Entire ISHIKARI RIVER WATERSHED.

Month	February		May		August	
Subwatershed No.	Entire Network	Subnetwork	Entire Network	Subnetwork	Entire Network	Subnetwork
I	0.136	0.120	0.184	0.216	0.266	0.268
II	0.121	0.105	0.209	0.225	0.262	0.252
III	0.189	0.173	0.353	0.271	0.249	0.250
IV	0.555	0.602	0.254	0.288	0.223	0.231

in summer months. Correlation coefficients at stations 1 and 7 are small in September, causing an alpha coefficient to become small for subnetwork A in Fig. 4. Correlation coefficients between estimates of mean areal precipitation and point precipitation depths can be used to check the representativeness of rain gage networks. Judging from Figs. 3 and 4, subnetwork B is considered more representative of the network of the gages, compared with subnetwork A in subwatershed I.

Table 3 shows a few examples of correlation coefficients of estimates for mean areal precipitation among four subwatersheds in the entire network and a subnetwork as indicated in Table 2. There are no significant differences in the correlation coefficients between the entire network and the subnetwork. The mean depths of precipitation over four subwatersheds were combined to yield the mean depth of precipitation over the entire Ishikari River Watershed. Table 4 gives the weighting factors for each subwatershed to calculate the mean depth of precipitation over the total area in the entire network and the subnetwork. Due to large areal variabilities of precipitation caused by topographic features within each subwatershed, there are slight differences of the weighting factor between the entire network and the subnetwork for four subwatersheds in winter months. On the other hand, during the summer, no significant differences of the weighting factor would be recognized between two networks due to the predominance of meteorologic factors in the production of rainfall. Fig. 6 shows the mean and the standard deviation of estimates of monthly precipitation over the entire Ishikari River Watershed in the entire network of 29 gages and the subnetwork of 18 gages. Since point precipitation depths are smoothed

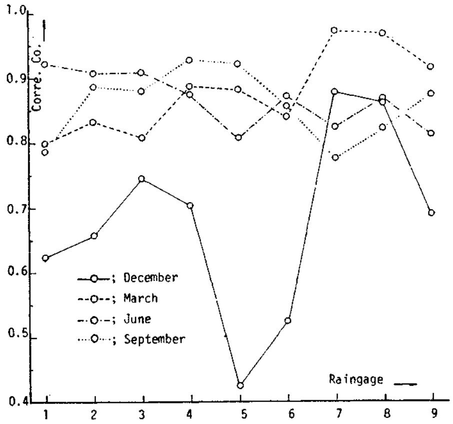

Fig. 5 Correlation Coefficient between Mean Watershed Precipitation and Point Precipitation Depths in the Watershed I.

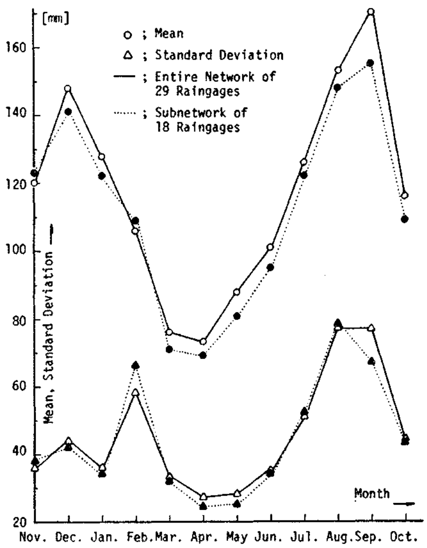

Fig. 6 Changes of Mean and Standard Deviation due to Different Density of Raingage Networks in the Entire ISHIKARI RIVER WATERSHED.

out with the watershed area being larger, these statistical quantities are not largely affected by the selected rain gages. Fig. 7 gives the absolute difference of estimates between the entire network and subnetwork over the total area for the period of 21 years. The maximum difference of about 65 mm is discerned in September. The larger differences of the statistical quantities in Fig. 6 and mean depths in Fig. 7 in September may be due to the fact that there were considerable differences in several quantities between the entire network of 8 gages and the subnetwork of 5 gages in subwatershed III. Table 5 shows the difference of the statistical quantities between two networks for subwatershed III in September. As a result, these large differences are reflected in Figs. 6 and 7. It is noted in Table 5, however, that the less dense network gives larger values than the entire network in terms of the percentage contribution and an alpha coefficient and accordingly the subnetwork will give more accurate results in estimating mean depths of monthly precipitation over subwatershed III in September.

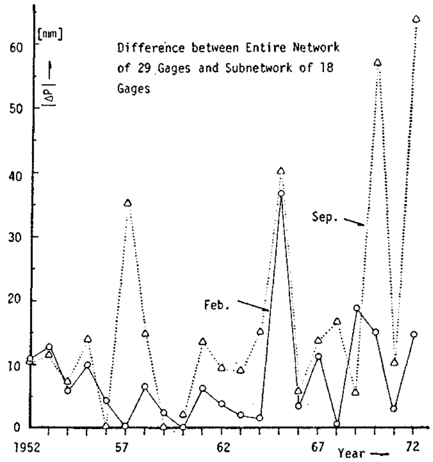

Fig. 7 Absolute Difference of Estimates due to Different Density of Raingage Networks in the Entire ISHIKARI RIVER WATERSHED.

Table 5 Changes of Statistical Quantities due to Different Density of Raingage Networks in the Watershed III.

(September)

Parameter \ Number of Gages	8 Gages	5 Gages
$\lambda_{max} / \sum_{i=1}^{n} \sigma_i^2$ (%)	78.32	93.54
Alpha Coefficient	0.777	0.959
Mean of Areal Pre.	191.5	154.8
Standard Deviation of Areal Pre.	101.6	84.0

From the above results on the representativeness of the network for estimation of mean areal precipitation, the entire network does not necessarily provide accurate results even within a small watershed area, because the areal variability of precipitation is largely affected by the meteorologic and topographic characteristics. If the requirements for estimating mean areal precipitation can be expressed in terms of percentage contributions and alpha coefficients, it is desirable to arrange the network density so as to make these quantities larger. If a large number of rain gages are available, they can be stratified into homogeneous groups in which the variation among the gages within a group is much less than that among the groups by examining the magnitudes of the parameters presented in this study.

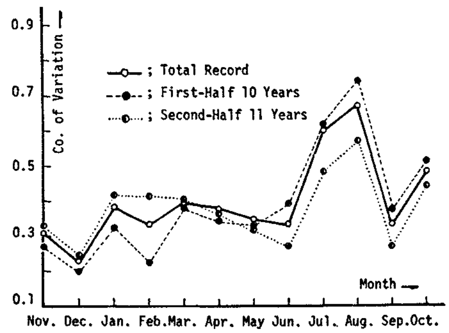

Fig. 8 (a) Changes of Coefficient of Variation due to Varying Length of Records at ASAHIKAWA.

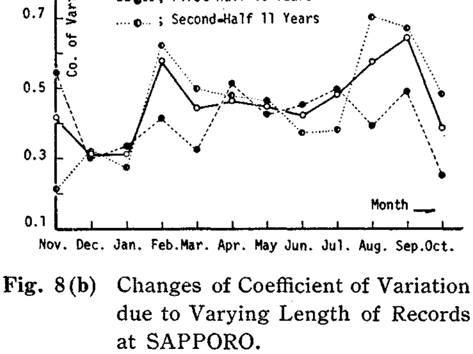

Fig. 8 (b) Changes of Coefficient of Variation due to Varying Length of Records at SAPPORO.

A further check of a sample size on the reliability of estimates for monthly total precipitation over a watershed was made through split-record tests. The total sample size of 21 years was divided into two segments, that is, the first-half records of 10 years and the second-half records of 11 years. A typical split-record comparison of the coefficient of variation in monthly precipitation at two gages is illustrated in Fig. 8. Neither the means nor variances of these sets of records differ significantly at the 5% levels, and the records could be homogeneous without seriously affecting the analysis. This also corresponds to the fact that a correlogram of monthly precipitation values produces the predominant periodicity of a year. Although the seasonal variation of precipitation at each gage did not widely vary from year to year, the large areal variability was revealed when the lengths of data were varied. This led to practical difficulties for the evaluation of model parameters, because a set of positive direction cosines could not be obtained during the winter in the entire network for each subwatershed as shown in Table 2. Such cases are: (1) for subwatershed I, the first half of December. (2) for subwatershed II, the first half of December and January. (3) for subwatershed III, the second half of December and two halves of January. (4) for subwatershed IV, the second half of November and the first half of January. As a result, approximating mean areal precipitation is of limited success and applicability in winter months. A few examples of the results for subwatershed II are shown in Figs. 9~12. Figs. 9 and 10 give monthly variations of the percentage contribution and an alpha coefficient with the lengths of data varied, respectively. In these figures, the quantities are omitted in December and January, because a set of positive gage weightings was unavailable for the first-half records of 10 years in these months. Both the percentage contribution and an alpha coefficient have large values in summer months for three cases. This suggests that the areal variability of rainfall would not be considerably affected by the lengths of available data during the summer.

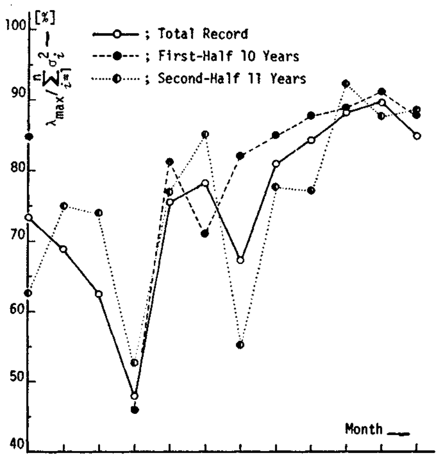

Fig. 9 Changes of the Maximum Eigenvalue due to Varying Length of Records in the Watershed II.

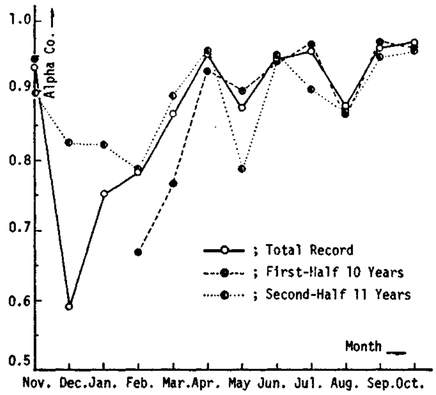

Fig. 10 Changes of Alpha Coefficient due to Varying Length of Records in the Watershed II.

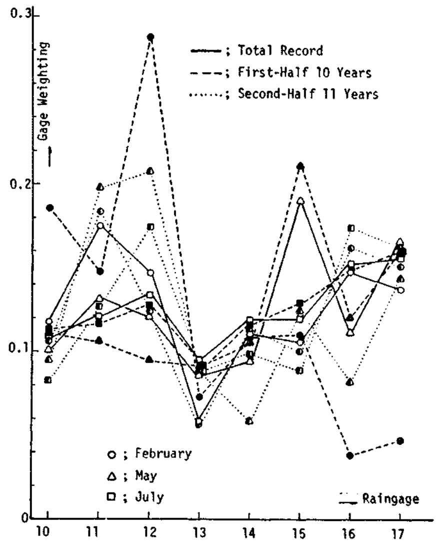

Fig. 11 Changes of Gage Weighting due to Varying Length of Records in the Watershed II.

Therefore, it is possible to estimate mean areal rainfall with a high precision, even when a large number of records are unavailable. Fig. 11 gives the changes of the gage weightings. In February, there are considerable differences in gage weightings between the total and the first-half records in stations (12, 16, 17), which may result in a decrease of an alpha coefficient in Fig. 10. In May, the larger differences in stations (11, 12, 15) between the total and the second-ealf records may cause the quantities to have small values in Figs. 9 and 10. In July, no significant differences are revealed between the three cases of a split-record test. Fig. 12 shows absolute differences of estimates between the total record and two halves of records. The larger differences of estimates are discerned between the total and the first-half records in February, and the second-half records in May. These results correspond to the differences of the statistical quantities in Figs. 9 and 10.

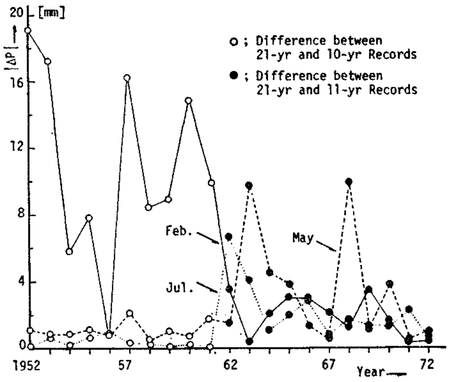

Fig. 12 Absolute Difference of Estimates due to Varying Length of Records in the Watershed II.

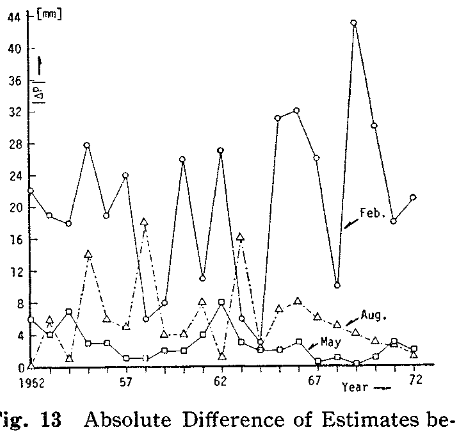

Fig. 13 Absolute Difference of Estimates between Proposed Method and Thiessen Method in the Watershed I.

As a final check of the validity of the model equation, the absolute difference of estimates between the proposed and the Thiessen methods is presented in Fig. 13. The difference of estimates between two models is larger in winter months than in summer months corresponding to the magnitude of the percentage contribution and an alpha coefficient as shown in Figs. 3 and 4. The maximum variance can be used as a measure of reliability of estimates given by the Thiessen method, as the difference of estimates between two methods is not much larger during the summer.

The numerical analysis of this study was carried on the FACOM 230-60 system at the Computing center of Hokkaido University.

5. CONCLUSIONS

(1) The mathematical features of traditional techniques for estimating mean areal precipitation do not contain a framework that is sufficiently flexible to respond to all needs with regard to the evaluation of the reliability of estimates and a current network of rain gages. The present study assumes that the mean areal precipitation can be represented by a linear combination of precipitation depths at the rain gages, and the size of the effective area controlled by the rain gage will depend on the choice of time unit and the feature of space variations of precipitation. In order to elaborate the estimating technique so as to yield better estimates of mean areal precipitation, the cross correlations among the rain gages are used to determine weighting factors at the rain gages.

(2) An approach to the estimation of mean areal precipitation and its reliability through the principal axis method demonstrates that only an eigenvector corresponding to the largest eigenvalue has to be selected in a variance-covariance matrix of precipitation depths at the rain gages. Modification of this solution gives the effective area expressed as a percentage of the whole area of a watershed. The advantage of the present study lies in the fact that if the rain gages are added or removed from service, only some minor changes are required in the data preparation.

(3) The proposed technique not only allows a measure of reliability to be evaluated but also permits the design of networks to give better estimates within specified limits. Two parameters are required to express the reliability of estimates: percentage contribution of the maximum

variance; a measure of the degree of reliability, and alpha coefficient; a measure of representativeness of various networks.

(4) A magnitude of the largest eigenvalue depends on the elements of a variance-covariance matrix. The less the areal variability of precipitation, the larger is the percentage contribution of the maximum variance to the total variance. Differences of estimates between the proposed and Thiessen models correspond to the magnitude of the maximum variance, that is, differences are small during the summer season, while they are greater in the winter months. If the difference of estimates between two methods is regarded small, the maximum variance can be used as a measure of reliability of estimates given by the Thiessen method.

ACKNOWLEDGEMENT

The author wishes to express his sincere gratitude to Professor Isao Yamaoka, Department of Civil Engineering, Hokkaido University, for providing constant encouragement, help, and guidance during the study.

REFERENCES

1) Hendrick, R. L. and Comer, G. H.: Space Variations of Precipitation and Implications for Raingage Network Design; Journal of Hydrology, 10, 1970.
2) Ishihara, T. and Ikebuchi, S.: A Statistical Study on Network of Precipitation Stations; Disaster Prevention Research Institute Annuals, Kyoto University, No. 13 B, 1970 (in Japanese).
3) Hoshi, K. and Yagi, S.: Correlation Analysis of Monthly Precipitation by the Group Principal-Axis Method; Proc. 26th Annual Meeting of JSCE, II, 1971 (in Japanese).
4) Chidley, T. R. E. and Keys, K. M.: A Rapid Method of Computing Areal Rainfall; Journal of Hydrology, 12, 1970.
5) McGuinness, J. L.: Accuracy of Estimating Watershed Mean Rainfall; Journal of Geophysical Research, Vol. 68, No. 16, 1963.
6) Clarke, R. T. and Edwards, K. A.: The Application of the Analysis of Variance to Mean Areal Rainfall Estimation; Journal of Hydrology, 15, 1972.
7) Stol, Ph. Th.: The Relative Efficiency of the Density of Rain-Gage Networks; Journal of Hydrology, 15, 1972.
8) Hutchinson, P.: A Contribution to the Problem of Spacing Raingauges in Rugged Terrain; Journal of Hydrology, 12, 1970.
9) Shiba, S.: Correlation Analysis in the Behavioral Sciences; Third Edition, the University of Tokyo Press, 1970 (in Japanese).
10) Hoshi, K.: A Study on the Analysis and Simulation of Precipitation by the Multivariate Statistical Model; Proc. JSCE, No. 213, 1973.

(*Received Jan. 23, 1974*)