PROC. OF JSCE,
NO. 213, MAY 1973

A STUDY ON THE ANALYSIS AND SIMULATION OF PRECIPITATION BY THE MULTIVARIATE STATISTICAL MODEL

By Kiyoshi HOSHI*

1. INTRODUCTION

The explosive population growth along with an accelerating shift from agricultural land use to urban and industrial land use has encouraged rapid technological developments in water resources. Among possible technological advances are the possibilities of conserving or augmenting natural water supplies. These supplies are, however, poorly distributed, and some regions are already short of water, and other areas will be in short supply in a relatively short time. One obvious alternative is to limit the development of such areas to that which can be accommodated within the available water supplies where techniques should be included for increasing the utility of available supplies by conservation or regulation which permits greater use. For many reasons this alternative may be undesirable, and a second alternative is to import water from some nearby regions of surplus via extensive aqueduct systems. From this point of view, there is a need for a wider range of regional evaluation of the quantities of available water in place of local control, and it is necessary to provide more accurate predictions of regional interrelationships of precipitation which is the major important source to the surface runoff phase in order to meet rapidly increasing future demands for water.

In recent years the literature has abounded with papers devoted to synthesis of various time series[1]. Many of these papers have concentrated upon generating sequences with autocorrelation functions, and subsequently proposed Markov chain models of various orders on a monthly or an annual basis. A decrease in the time unit of a hydrologic time series usually leads to increased problems in fitting of data generating models where extensions of Markov chain models have met with limited success mainly due to the high variance of a time series.

The analysis of the autocorrelation function indicates that after deseasonalising, a precipitation time series contains no statistically significant persistence at a single location[2],[3]. Therefore, there is a limitation on the adaptability of a Markov model when a stationary time series of precipitation is considered to be a pure random process from the correlogram. And moreover, the defect in hydrologic data such as missing data, in particular, during the winter season is a principal obstacle to the effectiveness of the recursion relation of a Markov model in the simulation process.

With increasing complexity of water resources systems the problems arising in their projection and control involve substantially the large-scale water supply systems on a region-by-region basis where an adequate description of regional differences of precipitation pattern over a large area is required, and an adequate synthesis requires precipitation data to be simulated simultaneously at several stations, their series being related to each other in an appropriate manner[4].

For the purposes of the study which motivates this work, the ability to generate a synthetic sequence at a single station is not of great value. This approach to region problems and solutions seeks to develop a guide to future developments for the optimum use, or combination of water resources to meet foreseeable long-term needs in terms of chosen objectives. The study described in this paper concerns itself with the formation of spatial relationships of precipitation over a wider area from the viewpoint of mathematical statistics, and demonstrates that the multivariate statistical analysis and simulation[5],[6],[7] will enable investigators to produce satisfactorily ac-

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

curate results with a minimum of time and effort expended.

2. CHARACTERISTICS OF MULTIVARIATE ANALYSIS

Current water resources development requires the best engineering procedures for the evaluation, prediction, and control of water resources. Lack of adequate data, in part, hampers fulfillment of these objectives. As a solution to this problem, operational hydrology plays an important role in water resources systems engineering because it helps to achieve efficient and economic designs of water resources systems through the use of analytical or simulation models. Based on the statistics of historical records, synthetic hydrologic sequences are simulated which preserve the significant statistical properties of historical records. These sequences will be used as inputs to subsequent assessment of the response of water resources systems design and management.

In the simplest stochastic model, a station with the longest observed record is selected as a key station, and the concurrent hydrologic sequences at all other satellite stations are developed only from the key station after least-squares coefficients are calculated for overlapping records between a key station and a satellite station. It is impossible to utilize relevant information among the satellite stations in this linear-regression model for generating synthetic sequences at their stations. The author proposed a technique of simultaneous generation of monthly precipitation at several satellite stations from a key station where use was made of interrelationships among all the satellite stations with the consequent reduction in the random components[8)].

To establish relations among a set of variables, multiple regression techniques have been made use of almost exclusively. Although the variables seldom satisfy all the underlying assumptions, the techniques have proved useful where the relations are used for predictive purposes. Prediction is the primary purpose of a regression relation. However, the inability to interpret a regression relation in terms of cause-and-effect is often attributed to the lack of independence among the variables. Moreover, a multiple regression approach is not probably preferable in the analysis of regional distribution of precipitation in which it is impossible to distinguish the dependent variable from the independent variables.

Analysis on the modern water resources problems should encompass relationships between regions in the long-run time periods in the practical hydrologic engineering. When regional interrelationships of precipitation are ignored, it is not sufficient to make a fuller assessment of water resources projects and a more rational use of water to meet future demands for the surface runoff.

The optimum use of an existing water supply depends on an accurate quantitative assessment of the possible techniques. In developing models, a lack of understanding of the basic process occurring simultaneously within a precipitation system, or a lack of physical data from which relationships can be established, constitutes a limitation on the application of physical hydrology to the assessment of regional precipitation pattern over a large area. One of the difficulties in describing precipitation pattern over a wider area which results from complicated interactions of meteorological and topographical characteristics is how to evaluate regional differences of precipitation at several stations. This difficulty is conveniently overcome by a suitable choice of a multivariate statistical model by which the structure of dependence among the variates can be approached. A distinguished feature of a multivariate technique adopted in this paper is an orthogonality property which is a useful strategy for evaluating regional differences of precipitation pattern at several stations, and for classifying the stations into some subgroups in an objective manner which have the distinct properties of precipitation characteristics.

3. FUNDAMENTAL CONCEPTS

The main object of the proposed multivariate technique is to describe adequately regional differences of precipitation pattern at the adopted stations, thereby applying the parameters estimated from it to a simultaneous simulation model of precipitation at the stations. To do this, a mathematical model of a linear transformation is constructed as follows:

$$G=WZ \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots (1)$$

$$[g_1\ g_2\ g_3 \cdots\cdots g_n]^T=[w_1\ w_2\ w_3 \cdots\cdots w_n]^T Z \quad (2)$$

where

$$G=[g_1\ g_2\ g_3 \cdots\cdots g_n]^T \quad \cdots\cdots (3)$$

$$W=[w_1\ w_2\ w_3 \cdots\cdots w_n]^T \quad \cdots (4)$$

$$g_i=w_i Z \cdots\cdots\cdots\cdots\cdots\cdots\cdots (5)$$

g_i, ith standardized composite variate (ith component) and a $(1 \times N)$ matrix; w_i, ith standard weight vector and a $(1 \times n)$ matrix; Z, standardized data array among the selected stations and an $(n \times N)$ matrix; G, an $(n \times N)$ matrix; W, an $(n \times n)$ matrix; n, number of stations; N, number of observations.

Before the detailed statistical analysis of the model equation, the fundamental properties of the generalized form of Eq. (5) are presented. Since the ith composite variate of g_i should be standardized with unit variance by definition, postmultiplying Eq. (5) by g_i^T/N gives

$$w_i R w_i^T = 1 \quad \cdots\cdots (6)$$

$$R = ZZ^T/N \quad \cdots\cdots (7)$$

where R denotes a correlation coefficients matrix among the stations concerned, and Z^T and w_i^T are the transpose of Z and w_i, respectively.

To obtain a correlation coefficient between the composite variate and each station, postmultiplying Eq. (5) by Z^T/N gives

$$a_i = w_i R \quad \cdots\cdots (8)$$

$$a_i = g_i Z^T/N \quad \cdots\cdots (9)$$

where a_i is defined as a structure vector consisting of correlation coefficients between the composite variate and each station $(-1 \leq a_i \leq 1)$ and it is a $(1 \times n)$ matrix. The structure vector is a useful parameter to account for the internal structure of correlations among the stations and to represent the relative contribution of each station to a liner composite.

The sum of the squared elements of a structure vector indicates the total contribution of the stations to the composite variate. This sum is given by

$$V_i = a_i a_i^T = w_i R^2 w_i^T \quad \cdots\cdots (10)$$

where $R = R^T$ and $R^2 = RR$.

To extract regional differences of precipitation pattern from the available records at the stations, an orthogonal condition that two composite variates are uncorrelated is introduced. This expression is given by

$$g_i g_j^T/N = w_i R w_j^T = 0 \qquad (i \neq j) \quad \cdots\cdots (11)$$

The adoption of an orthogonal condition, from a hydrologic point of view, indicates that the stations under consideration could be classified into some subgroups with the distinct properties of precipitation characteristics, so that a weight vector is high in one component, while it should be low in the other and vice versa.

The procedure to estimate a weight vector in the first component is to choose w_1 so as to make V_1 a maximum under the condition of Eq. (6). This criterion is written as follows:

$$2T_1 = w_1 R^2 w_1^T - \lambda_1 (w_1 R w_1^T - 1) \quad \cdots\cdots (12)$$

where λ_1 is a Lagrange multiplier.

Then, set the partial derivative of this new function T_1 with respect to each element of w_1 equal to zero, namely,

$$R^2 w_1^T - \lambda_1 R w_1^T = 0 \quad \cdots\cdots (13)$$

Substituting the transpose of Eq. (8) into Eq. (13) gives

$$(R - \lambda_1 I) a_1^T = 0 \quad \cdots\cdots (14)$$

where I and 0 denote an identity matrix and a null vector, respectively.

Premultiplying Eq. (13) by w_1 and using Eqs. (6) and (10) give

$$V_1 = \lambda_1 = a_1 a_1^T \quad \cdots\cdots (15)$$

From Eqs. (8) and (14) the following expression is derived:

$$w_1 = a_1/\lambda_1 \quad \cdots\cdots (16)$$

The maximization of Eq. (10) under the condition of Eq. (6) leads to the system of Eq. (14) for the solution of an eigenvalue (λ_1) and corresponding eigenvector (a_1^T) of a correlation matrix R. An important feature of Eq. (15) is that an eigenvalue of λ_1 is precisely V_1, the quantity which is to be maximized. In other words, V_1 is equal to one of the roots of the characteristic equation, namely, the largest root λ_1. Eq. (16) gives the solution of a standard weight vector in the first component.

The procedures to estimate weight vectors in the other components are subject to the conditions, which are the restriction of Eq. (6) and the orthogonal condition of Eq. (11). The quantity of Eq. (10) should be maximized under the conditions of Eqs. (6) and (11). The functions necessary for obtaining the weight vector w_j $(j > i)$ in Eq. (5) corresponding to Eqs. (12) and (13) are expressed as follow:

$$2T_j = w_j R^2 w_j^T - \lambda_j (w_j R w_j^T - 1) - 2\theta w_i R w_j^T \qquad (j > i) \quad (17)$$

$$R^2 w_j^T - \lambda_j R w_j^T - \theta R w_i^T = 0 \quad (j > i) \quad (18)$$

where λ_j and θ are Lagrange multipliers.

Premultiplying Eq. (18) by w_j and using Eqs. (6), (10), and (11) yield

$$V_j = \lambda_j = a_j a_j^T \quad \cdots\cdots (19)$$

Premultiplying Eq. (18) by w_i $(i \neq j)$ and using Eqs. (6) and (11) yield

$$\theta = w_j R^2 w_i^T \quad \cdots\cdots (20)$$

Suppose $i = 1$ and $j = 2$, then substituting Eq. (13) into Eq. (20) and using Eq. (11) yield

$$\theta = 0 \quad \cdots\cdots (21)$$

In general, Eq. (21) can hold for the chosen pair of i and j. Therefore, Eq. (18) reduces to

$$R^2 w_j^T - \lambda_j R w_j^T = 0 \quad \cdots\cdots(22)$$

Substituting Eq. (8) into Eq. (22) yields

$$(R - \lambda_j I) a_j^T = 0 \quad \cdots\cdots(23)$$

Eq. (23) is explicitly identical with Eq. (14) in a mathematical notation.

From Eqs. (19) and (23) it is clear that λ_j and a_j^T are the jth largest eigenvalue, and its associated eigenvector of the correlation matrix, respectively. The generalized expression of weight vectors corresponding to Eq. (16) is written as follows:

$$w_j = a_j / \lambda_j \quad \cdots\cdots(24)$$

The advantage of technique proposed herein is that all weight vectors in Eq. (5) are obtained directly from the successive eigenvalues and their associated eigenvectors of the original correlation matrix. Most scientific subroutine libraries on the computers may facilitate the evaluation for the solution of λ_j and a_j in Eq. (23). However, the elements of each eigenvector in Eq. (23) must be normalized to satisfy the relation of Eq. (19), so that they are divided by the square root of the sum of their squares and then multiplied by $\sqrt{\lambda_j}$.

Some distinguishing properties[9),10)] of eigenvalues and eigenvectors in the system of a real symmetric matrix (the correlation matrix) are: (a) The eigenvalues are all real, positive, and distinct. (b) If a_i and a_j are eigenvectors corresponding to the distinct eigenvalues λ_i and λ_j, then $a_i a_j^T = 0$. (c) For the correlation matrix with unities in the diagonal, the λ_i means the variance of each component, and the sum of the variances of all n components is equal to the trace of R, the trace being the sum of the diagonal elements. Since the trace of R is the total variance (equal to n) to be accounted for, the cumulative sum of eigenvalues divided by the trace is the proportion of variance accounted for by the resulting components.

From the property of (b), and Eqs. (19) and (24), expressions with respect to the structure vectors and weight vectors are written as follow:

$$\begin{cases} a_i a_i^T = \lambda_i \\ a_i a_j^T = 0 \qquad (i \neq j) \end{cases} \quad \cdots\cdots(25)$$
$$(i, j = 1, 2, 3, \cdots\cdots, n)$$

and

$$\begin{cases} w_i w_i^T = 1/\lambda_i \\ w_i w_j^T = 0 \qquad (i \neq j) \end{cases} \quad \cdots\cdots(26)$$
$$(i, j = 1, 2, 3, \cdots\cdots, n)$$

According to the property of (c), equations with respect to the eigenvalues are derived as follow:

$$\sum_{i=1}^{n} \lambda_i = n \quad \cdots\cdots(27)$$

and

$$\sum_{i=1}^{m} \lambda_i / n \times 100\% = P_m \qquad (m < n) \quad \cdots\cdots(28)$$

where P_m represents the percentage of variation accounted for by the mth component.

From Eq. (26) row vectors of W in Eq. (1) are linearly independent, so that the determinant of W is different from zero and W has an inverse. Premultiplying Eq. (1) by W^{-1} gives

$$Z = W^{-1} G \quad \cdots\cdots(29)$$

where W^{-1} denotes the inverse matrix of W.

Eq. (29) produces a number of properly correlated series of precipitation of the length desired for the specified time points at n stations when the elements of an $(n \times n)$ matrix of W are estimated from the available records at the stations with help of Eqs. (23) and (24). Numerical procedures are very simple and consist only of generating random numbers of G because a set of n composite variates are mutually orthogonal (uncorrelated) by Eq. (11).

4. PRACTICAL APPLICATION

The application of the theory developed in the preceding section to the problem of the analysis and synthesis of ten-day total precipitation (ten-day is referred to as "Jun" in Japanese) on the available records was made.

Ten gaging stations scattered over Hokkaido, Japan, were selected to illustrate various features in the regional distribution of precipitation, where simultaneous records of ten-day total precipitation were available for the period of 27 years from 1943 to 1970.

Fig. 1 demonstrates the location of stations un-

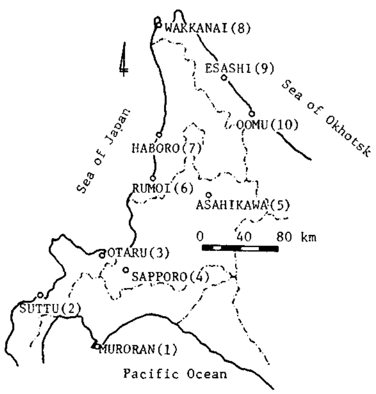

Fig. 1 Location Map of Selected Stations.

der consideration with station numbers assigned. The station numbers in this figure correspond to numbers in the subsequent results shown in several figures.

Before proceeding into the detailed statistical analysis of the model equation, results by the traditional techniques of frequency distributions and a time series analysis are presented.

Fig. 2 gives a typical illustration of the annual variation in the means and the coefficients of variation of ten-day total precipitation at three stations which are considered as representatives of different meteorological regions of Hokkaido. The complex pattern of precipitation in Hokkaido reflects several interacting influences. The marked variations in the annual precipitation cycle are:

(a) In the Western areas, heavier precipitation occurs during the winter, where the importance of the cold continental high-pressure and mountains as factors in the production of precipitation is evident. The coefficients of variation are far more stable during the winter. And melting snow provides the major portion of the annual runoff in the side of the Sea of Japan.

(b) A summer maximum of precipitation is observed in the sides of the Pacific Ocean and the Sea of Okhotsk, where the low-pressure center recedes northward during the summer. The coefficients of variation show large seasonal fluctuations in the Pacific Ocean side.

However, the different characteristics of statistical quantities in the individual regions could not be employed for the simultaneous syntheses of precipitation over a fairly large area.

Some examples of correlograms are presented

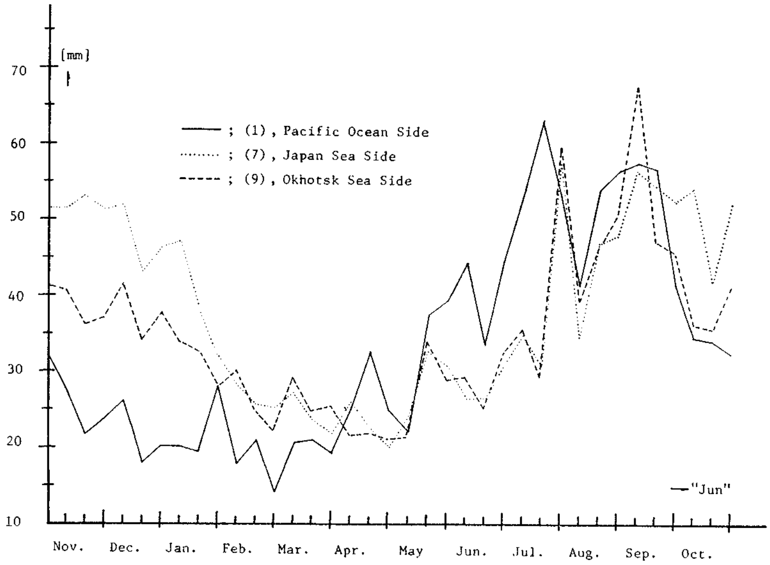

Fig. 2 (a) Mean of ten-day precipitation.

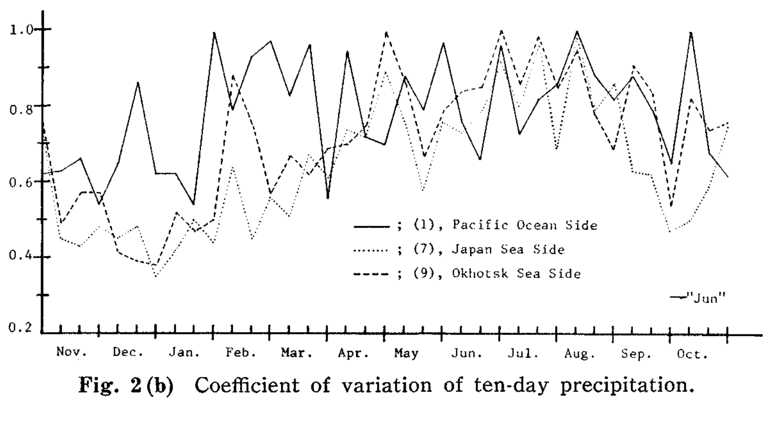

Fig. 2 (b) Coefficient of variation of ten-day precipitation.

not only to demonstrate the properties of theoretical stochastic and deterministic processes, but also to examine the use of a Markov chain model. The solid line in Fig. 3 represents the serial correlation coefficients of an observed time series for ten-day total precipitation with the sample size of 972. From these correlograms the structure of an original time series indicates the following facts:

(a) The correlograms tend to reach their maximum values at lags that are multiples of 36 "Juns" (12 months).

(b) The maximum amplitudes of correlograms lie within 10 to 20% of the total variance so that seasonal variations in the observed time series are assumed to account for 10

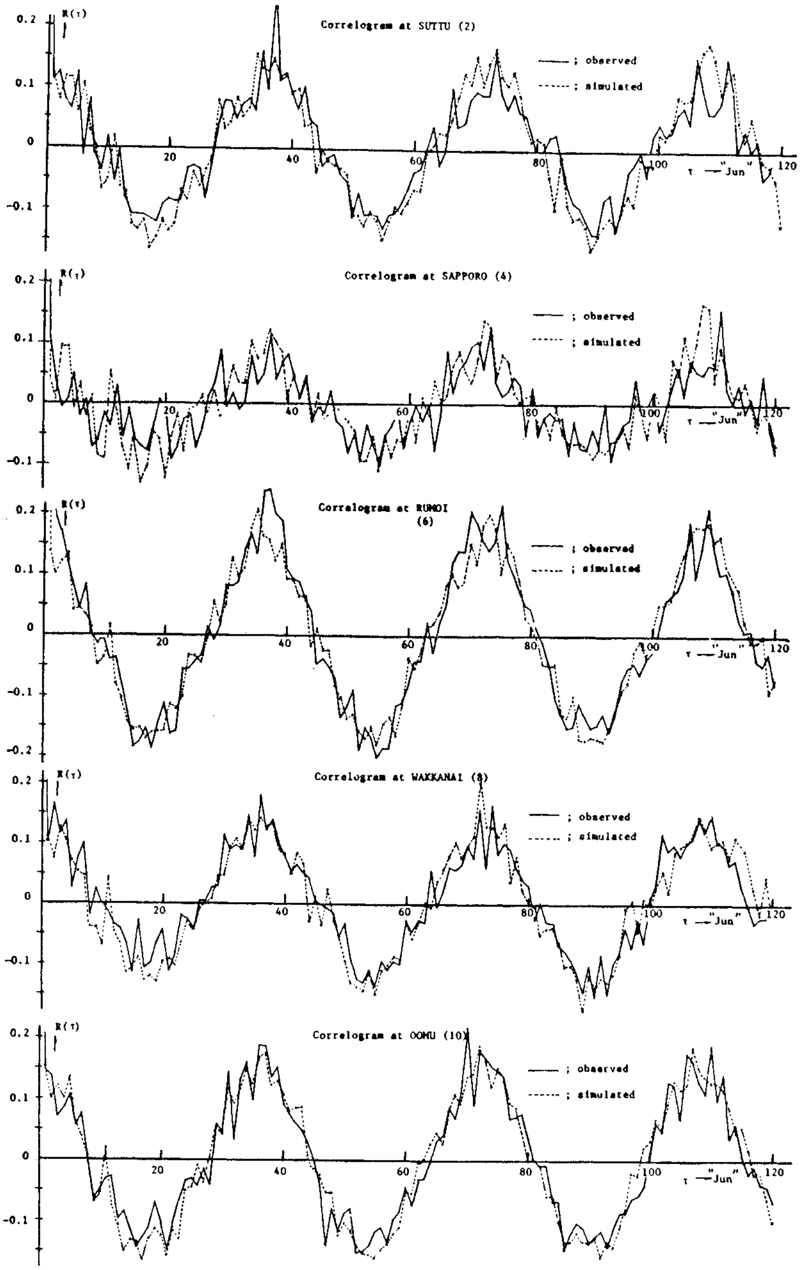

Fig. 3 Correlogram of an Original Time Series.

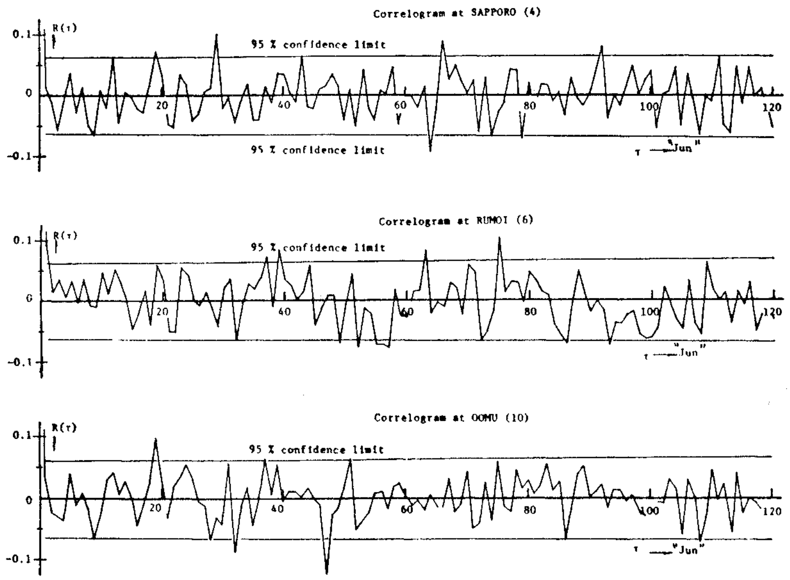

Fig. 4 Correlogram of a Standardized Time Series.

to 20%, while stochastic random components account for 80 to 90%.

The fact of (b) indicates that a generation model of ten-day total precipitation by the time series theory could fail to make the statistics of simulated data compatible with those of observed records because of high variances in a precipitation time series.

Fig. 4 gives the correlogram of a standardized time series where the seasonal variation at a station is removed in such a manner that the mean is subtracted from an original series for each "Jun", and the difference is divided by the standard deviation for each "Jun". The serial correlation coefficients were tested for significance on a normal random time series of N values[11] (=972). The confidence limits (CL) for a computed value of serial correlation coefficients ($R(\tau)$) are given by

$$CL(R(\tau))=\frac{-1\pm1.960\sqrt{N-\tau-2}}{N-\tau-1}\quad\cdots\cdots\cdots(30)$$

where τ denotes lag time, and the value of 1.960 is the standardized normal variate corresponding to the probability level of 95%. The confidence limits of Eq. (30) are also shown in Fig. 4. Since most of the serial correlation coefficients of a standardized time series for ten-day precipitation lie within the confidence limits, $R(\tau)$ is considered to be insignificantly different from zero at the probability level of 95%. Therefore, the reliability of a Markov chain model for generating sequences of ten-day total precipitation is questioned because a time series with seasonal variations removed is regarded as a pure random process.

Some results by the multivariate technique are presented. Fig. 5 gives the percentage contributions of the first four components to the total variance for ten-day precipitation where each eigenvalue is divided by the total variance of 10. The four components account for more than 75% in every month. The contribution of the first component to the total variance is the largest in every month, indicating an important general factor of precipitation pattern among the stations. Seasonal differences in terms of variance of each component are noticed:

(a) During the winter season, the first component contributes 35 to 40% to the total variance, and the contributions of the second and third components are equal to about 20%.

(b) During the summer, the first component accounts for about 70%, which shows the higher contribution in contrast with one during the winter. The second and third components account for about 20% and 10%, respectively.

(c) Relative large contributions of the second and third components during the winter suggest that regional differences of precipitation distribution among a given set of stations would be recognized. This problem is solved from the subsequent result in terms of a structure vector.

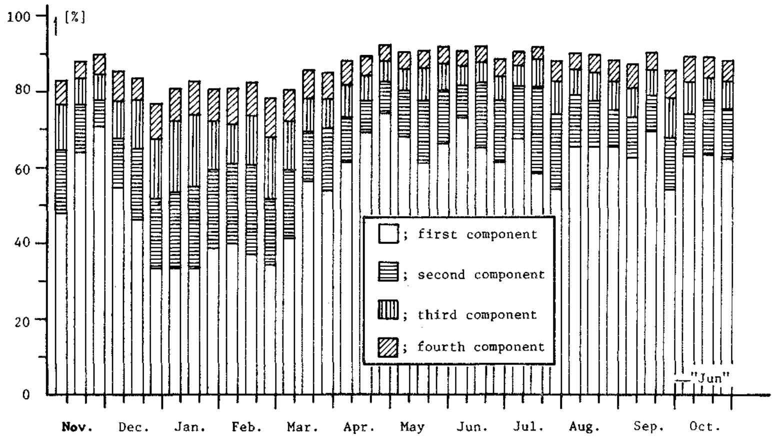

Fig. 5 Contribution of each component to total variance.

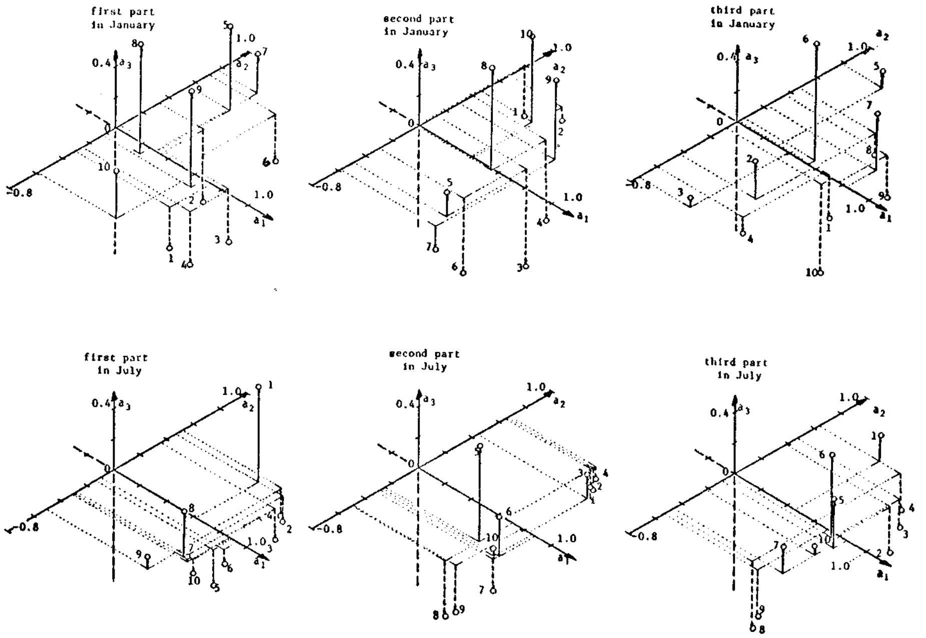

Fig. 6 Distribution of Three Structure Vectors.

Fig. 6 illustrates the distribution of structure vectors of the first three components in the three dimensional systems. The distribution of structure vectors which are mutually orthogonal facilitates the evaluation of regional differences of precipitation over a wide area, and the clustering of the stations. Three structure vectors, for example, are given by $(a_1, a_2, a_3)=(0.736, -0.305, -0.249)$ for MURORAN (station 1), $(a_1, a_2, a_3)=(0.337, 0.326, -0.453)$ for SUTTU (station 2), and $(a_1, a_2, a_3)=(0.820, 0.025, -0.331)$ for OTARU (station 3) in the first "Jun" of January and so on.

Equation (25) is an important property in which the contribution (equivalent to the eigenvalue λ_i) is related to the structure vector of each component, so that it is possible to evaluate the effect of a structure vector to the variance in one component in which some elements of a structure vector have large values and the others have small values. Some characteristics with respect to the distribution of three structure vectors are:

(a) A structure vector of the first component has positive elements in every "Jun" because (almost) all correlation coefficients are positive in the original correlation matrix.

During the summer, a structure vector has large coefficients of about 0.8 for all ten stations, causing the λ_1 to have the large contribution to the total variance. No regional differences of rainfall are recognized and a large-scale pressure area is uniformly continuous during the summer season. On the other hand, during the winter, a structure vector has a wide range of coefficients between 0.3 and 0.8, resulting in the λ_1 having the smaller contribution. This result suggests that snowfall may be influenced by the local variations inherent to the in dividual regions during the winter.

(b) A structure vector of the second component has positive coefficients for the stations in the Sourthern part of Hokkaido (stations; (1), (2), (3), and (4)), and negative coefficients for the stations in the Northern part (stations: (5), (6), (7), (8), (9), and (10)) during the summer. This division may result from the different meteorogical factor. As shown in Fig. 5, the contribution of λ_2 is large in the second "Jun" of July. This effect is interpreted from the fact that the stations (1), (2), and (4) have large positive coefficients of 0.7, while the stations (8) and (9) have large negative coefficients of -0.7 in Fig. 6. On the other hand, during the winter, a structure vector of the second component provides a means for the divison of the stations into two groups in the sides of the Sea of Japan, and the Sea of Okhotsk. The Mountains running in the north-south direction as a factor in the production of precipitation is of great importance in the second component.

(c) In the third component, one subgroup of stations (5) and (6), and another subgroup of stations (8) and (9) contribute to make the λ_3 large during the summer. During the winter, a structure vector has larger coefficients for the stations (8), (9), and (10) in the side of the Sea of Okhotsk.

(d) The distribution of these three structure vectors plays an important role in clustering the stations. A set of ten stations are made up of some distinct subgroups. For example, they consist of 4 subgroups of (1, 4), (2, 3, 6), (5, 7), and (8, 9, 10) in the first "Jun" of January, while 4 subgroups of (1), (2, 3, 4), (5, 6, 10), and (7, 8, 9) in the first "Jun" of July and so on.

The distribution of three weight vectors is similar to that of structure vectors, since elements of a weight vector are proportional to those of a structure vector as expressed by Eq. (24).

Since it is possible to express the contribution of each component to the total variance by an eigenvalue, and also to assess independent effects of weight vectors by an orthogonality property in the multivariate analysis, the simulation technique by Eq. (29) is valuable for simultaneous syntheses of ten-day total precipitation at 10 stations. One of the difficulties encountered in the practical study was that ten composite variates which are mutually orthogonal were generally skewed, even though the available records for each "Jun" were normalized by a square root transformation. To overcome this situation in the simulation process, normal random numbers on a gamma distribution were generated. Such a system is represented[6] by

$$\varepsilon_{ij}=\frac{2}{\gamma_i}\left(1+\frac{\gamma_i t_{ij}}{6}-\frac{\gamma_i^2}{36}\right)^3-\frac{2}{\gamma_i} \qquad \cdots\cdots\cdots(31)$$

$$(i=1,2,3,\cdots\cdots,10)$$

$$(j=1,2,3,\cdots\cdots,50)$$

where γ_i denotes the coefficient of skewness in the ith composite variate of Eq. (5), t_{ij} is a normally distributed random number with zero mean and unit variance, and ε_{ij} is a random number on a gamma distribution with zero mean, unit variance, and skewness γ_i.

Ten-day total precipitations at 10 stations were simultaneously generated for the period of 50 years by Eq. (29), where elements of a (10×10) matrix of W were estimated from the available records by use of Eqs. (23) and (24), and random numbers of G were generated by Eq. (31). As a check on the validity of the multi-site simulation model by Eq. (29), the means and standard deviations of the observed records compared with those of the simulated data for 50 years of operations are illustrated in Fig. 7 for some stations. The serial correlation coefficients of the observed (sample size=972) and the simulated (sample size =1 800) for ten-day total precipitation are shown in Fig. 3. Results by Eq. (29) reveal fairly good agreements with the essential statistics of the observed records.

Even though the use of simulation techniques can not entirely remove the chance elements in water resources projects, the good agreement between statistical properties of the observed and generated data indicates that the simulation model of correlated simultaneous ten-day total precipitation over any specified length for any number of stations proposed herein is useful for decision making in the simulation process.

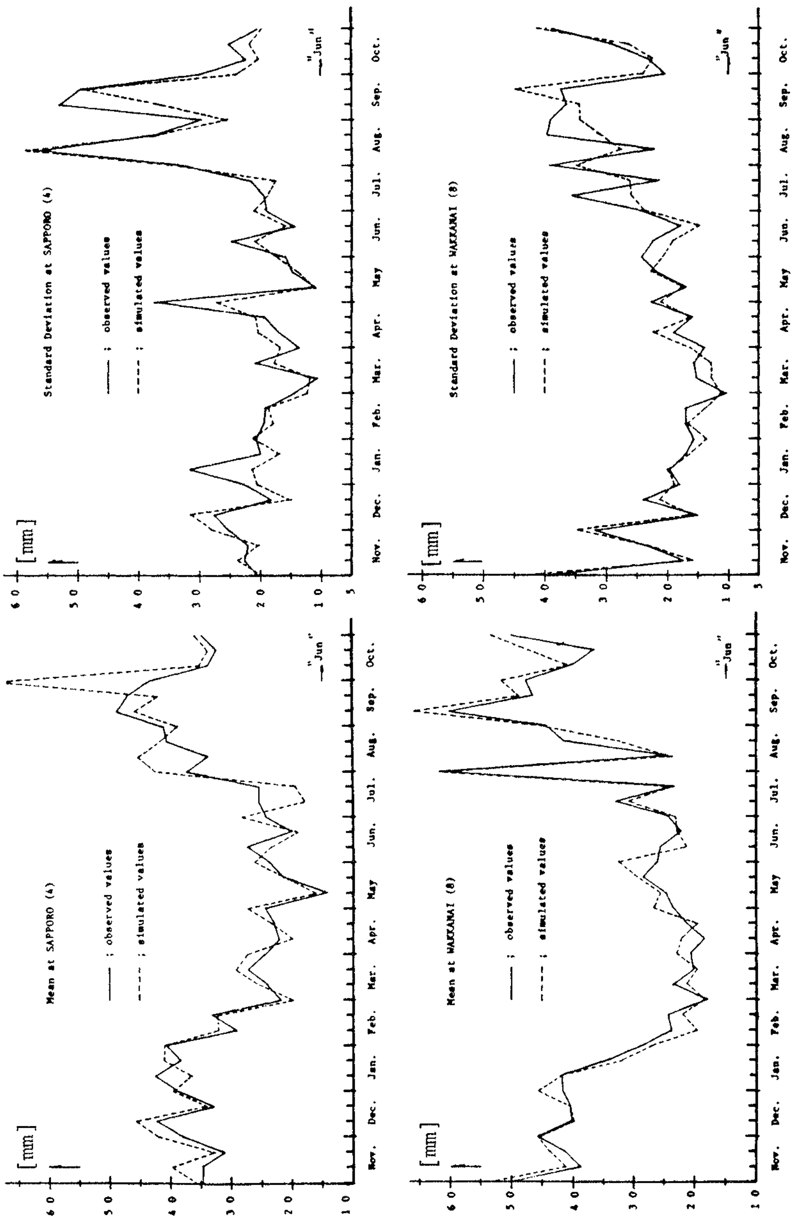

Fig. 7 Mean and Standard Deviation.

All computations involved in this study were performed on the FACOM 230-60 system in the computing center of Hakkaido University.

5. CONCLUSIONS

On the basis of some results from this study, the following facts are satisfactorily presented:

(a) The multivariate statistical analysis by matrix representation is a very useful tool for extracting information from a given set of stations.

(b) The system proposed herein ultimately leads to the solution of a linear-equation system, with the problem of a matrix inversion, and the solution of eigenvalues and corresponding eigenvectors of a correlation matrix. Therefore, the theory is simple and the computation procedures are suitable in practical situations when investigators can utilize the subroutine libraries on the high-speed computer.

(c) The contribution of each component to the total variance is expressed by an eigenvalue which is a useful index for evaluating seasonal variations of precipitation distribution over a fairly wide area.

(d) The structure vectors which are mutually orthogonal facilitate the evaluation of regional differences of precipitation pattern among the stations. And the distribution of the structure vectors reflects several interacting influences of meteorology and topography as factors in the production of precipitation, and it offers the solution to the problem of clustering the stations even when there exist high interrelationships among them.

(e) A decrease in the time unit of a hydrologic time series usually leads to increased problems in fitting of data generating models mainly due to the high variance of a time series. The inverse transformation of a linear system proposed herein provides a means for simultaneous syntheses of any specified length for any number of stations, where the total variance is expressible and independent effects of weight vectors can be assessed. The model equation improves some shortcomings of the traditional simulation techniques when the time unit of a precipitation time series decreases.

ACKNOWLEDGEMENT

The author is greatly indebted to Professor Isao Yamaoka, Department of Civil Engineering, Hokkaido University, for his encouragement and advice. And the author also wishes to express his grateful appreciation to the Meteorological Observatory at Sapporo which has very kindly supplied valuable data to me.

REFERENCES

1) D. G. Decoursey: The Stochastic Approach to Watershed Modeling; Nordic Hydrology, II, 1971.
2) S. Ramaseshan: Stochastic Modeling of Storm Precipitation; Nordic Hydrology, II, 1971.
3) T. Ishihara & S. Ikebuchi: A Study on the Stochastic Structures of Daily Precipitation and Their Simulation; Proc. JSCE, No. 197, 1972 (in Japanese).
4) A. Murota, T. Eto & M. Kadoyu: Space Structure Ananlyses and Simulation of Monthly Hydrologic Data; Proc. JSCE, No. 203, 1972 (in Japanese).
5) M. B. Fiering: Multivariate Technique for Synthetic Hydrology; Proc. ASCE, HY5, 1964.
6) N. C. Matalas: Mathematical Assessment of Synthetic Hydrology; Water Resources Research, Vol. 3, No. 4, 1967.
7) S. Shiba: Correlation Analysis in the Behavioral Sciences; Third Edition, the University of Tokyo Press, 1970 (in Japanese)
8) K. Hoshi & I. Yamaoka: A Simulation Technique of Monthly Runoff by Use of Precipitation Time Series at Multistations; Proc. 2nd International Hydrology Symposium, Ft. Collins, Colo., U.S.A., 1972.
9) C. R. Wylie: Advanced Engineering Mathematics; Third Edition, Mcgraw-Hill Book Company, 1966.
10) H. H. Harman: Modern Factor Analysis; Second Edition, Revised, the University of Chicago Press, 1968.
11) L. A. Roesner & V. M. Yevdjevich: Mathematical Models for Time Series of Monthly Precipitation and Monthly Runoff; Colorado State University Paper, No. 15, Fort Collins, Colorado, 1966.

(*Received July 4, 1972*)