APPLICATION OF MULTI-DIMENSIONAL SEQUENCE SIMILARITY METHOD IN METEOROLOGY

Gong Jiuding (貢九丁)

Zhejiang Research Institute of Meteorology, Hangzhou

Received July 29, 1984

ABSTRACT

The present work is devoted to multi-dimensional sequence similarity method with a view to predicting a typhoon analogous in both rainfall and track to the historical event. The result can be used in operational forecasts.

I. INTRODUCTION

Similarity analysis finds its wide use in routine forecasts at weather stations. As a rule, the so-called multi-dimensional similarity method is used when similarity of a number of factors in the past record is searched for and the sequence similarity method is adopted if that of successive values of a certain factor taken from the record is to be found. In studying the multi-dimensional sequence, the similarity method does not seem to be employed frequently except the cluster method. This may be due to the fact that the characterizing quantities available are difficult to use in multi-dimensional sequence similarity.[1] An attempt is therefore made to give a suitable characterizing factor for similarity, the matrix approximability vector, and thereby the difference in similarity of multi-dimensional sequences can be shown in a quantitative way.

If $\mathbf{X}_i$ is defined as the series of the ith factor, then multi-dimensional series can be factually expressed in terms of the matrix

$$\mathbf{X}=\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_m \end{pmatrix}.$$

However, matrix similarity in mathematics is quite different from multi-dimensional sequence similarity in meteorology. Since an equivalence class is formed by a set of analogous matrixes (which can hence include an endless set of matrixes), it is necessary to introduce the concepts of approximability and the approximability vector (AV) in matrix study when the multi-dimensional series similarity problem is dealt with in terms of matrix.

II. WAYS OF GETTING APPROXIMABILITY AND AV

1. *Approximability between Vectors and Its Expression*

If **A** and **B** are let to be two n-dimensional vectors, then Euclidean distance between them is

$$x = |\mathbf{A} - \mathbf{B}|$$

and the cosine of the included angle is

$$y = \frac{(\mathbf{A}, \mathbf{B})}{|\mathbf{A}| \cdot |\mathbf{B}|}.$$

If $\mathbf{A} = \mathbf{B} \neq 0$, then $x = 0$ and $y = 1$. In the case of $\mathbf{A} \neq \mathbf{B}$, $s = s(x, y)$, where s is for the difference between them. The domain of x values is $[0, M]$, M being a sufficiently large positive number, and that of y is $[-1, 1]$. The domains $[0, M]$ and $[-1, 1]$ are divided by L times and two L-term series are obtained

$$\{x\}: x_1, x_2, \cdots, x_L;$$
$$\{y\}: y_1, y_2, \cdots, y_L.$$

In operation $x_1 = M$, $x_L = 0$; $y_1 = -1$, $y_L = 1$, $x_p > x_{p+1}$ and $y_p < y_{p+1}$ are given. Obviously, each term in the series $\{x\}$ and $\{y\}$ corresponds to each number of a limited natural-number subset $\{1, 2, \cdots, L\}$[2].

s, approximability between the vectors $\mathbf{A}$ and $\mathbf{B}$, is defined as one of the natural numbers in $\{1, 2, \cdots, 2L\}$. When $x \leqslant x_p$ and $y < y_p$, $s = 2P - 1$. When $x \leqslant x_p$ and $y \geqslant y_p$, $s = 2P$ with $P \leqslant L$.

It is apparent that when $x \leqslant x_L = 0$ and $y \geqslant y_L = 1$ are true, $\mathbf{A} = \mathbf{B}$ is available, in which case $s = 2L$, reaching its maximum. s has a simplified form.

Definition 1:

$$s = \begin{cases} 2P - 1 & \text{if } (x \leqslant x_p) \wedge (y < y_p) \\ 2P & \text{if } (x \leqslant x_p) \wedge (y \geqslant y_p) \end{cases}$$

This definition is really nothing less than the accumulated of P by performing term-by-term comparison until $x > x_p$ or $y < y_p$ of the number couple (x, y) acquired through calculation with the number couple sequence consisting of $\{x\}$ and $\{y\}$ corresponding terms (x_1, y_1), (x_2, y_2), $\cdots$, (x_L, y_L). The more the terms for comparison, the greater the value of P and hence s, implying the closer the two vectors to each other, i. e. the smaller their difference would be. But comparison is done at most $s = 2L$ times and only on this occasion are the vectors equal, i. e. $\mathbf{A} = \mathbf{B}$.

The known number couple sequence $\{(x_p, y_p)\}$ $(P = 1, 2, \cdots, L)$ is designated as the approach sequence of the number couple (x, y). During the comparison, s, if given an odd number, would satisfy the requirement of Euclidean distance and s, given an even one, of the angle cosine. Then

$s = 1, 3, \cdots, 2L - 1$, (suitable for Euclidean distance)

$s = 2, 4, \cdots, 2L$, (suitable for the angle cosine)

where s represents the mutually synchronous approach of the vectors in norm and direction.

Definition 2: $\mathbf{S} = (s_1, s_2, \cdots, s_m)$ is let to be an AV.

Each component s_i of S denotes approximability of the vector $\mathbf{B}$ to every element $\mathbf{A}_i$ of the vector set $\{\mathbf{A}_i | i = 1, 2, \cdots, m\}$.

If the set is viewed as a matrix, then $\mathbf{S}$ is the AV of $\mathbf{B}$ to the matrix $\mathbf{A}^{m \times n}$.

2. *The AV and Approximability of Two Matrixes (Multi-Dimensional Sequence)*

$\mathbf{A}^{m \times n}$ and $\mathbf{B}^{m \times n}$ are both matrixes of m rows and n columns with $m \geqslant n$. Now all column vectors are transformed into the row ones, which are placed in order below the originally last row vector, with the gaps filled with zero. Then we have

$$(\quad)\, m \Longrightarrow (\quad)\, m+n$$

In this case $\mathbf{A}^{m\times n} \Longrightarrow \mathbf{A}^{(m+n)\times n}$ and $\mathbf{B}^{m\times n} \Longrightarrow \mathbf{B}^{(m+n)\times n}$. Apparently, all row vectors (with a total of $m+n$) in the right side of the arrow represent the matrix in the left.

Calculating approximability of corresponding row vectors of $\mathbf{A}^{(m+n)\times n}$ and $\mathbf{B}^{(m+n)\times n}$ gives the AV of the two matrixes. Then we come to the following.

Definition 3: $\mathbf{S}=(s_1, s_2, \cdots, s_{m+n})$ is defined as the AV of $\mathbf{B}^{m\times n}$ to $\mathbf{A}^{m\times n}$.

Definition 1 shows that the maximal approximability $s_{max}=2L$, and hence the maximal AV should be $\mathbf{S}_{max}=(2L, 2L, \cdots, 2L)$ (whose total is $m+n$). If definition is used again for $\mathbf{S}$ and $\mathbf{S}_{max}$, we have the following definition.

Definition 4: $\bar{s}$ is approximability of $\mathbf{S}$ to $\mathbf{S}_{max}$.

Then S, the AV, is dimensionally decreased to $\bar{s}$, which can characterize the approximability of $\mathbf{B}^{m\times n}$ to $\mathbf{A}^{m\times n}$ as well, and when $\mathbf{B}^{m\times n}=\mathbf{A}^{m\times n}$, $\bar{s}=2L$.

3. *The AV of $B^{m\times n}$ to the Matrix Set $\{\mathbf{A}_j^{m\times n} \mid j=1, 2, \cdots, k\}$*

It is noticed that $\mathbf{B}^{m\times n}$ is a multi-dimensional series of the sample to be predicted and the matrix set is that of the sample from the past data. Let the AV be

$$\tilde{\mathbf{S}}=(\bar{s}_1, \bar{s}_2, \cdots, \bar{s}_k),$$

$\bar{s}_j$, being the jth component of $\tilde{\mathbf{S}}$, is the approximability of $\mathbf{B}^{m\times n}$ to $\mathbf{A}_j^{m\times n}$.

The transformation mentioned in the preceding part includes

$$\{A_j^{m\times n} \mid_{j=1,2,\cdots,k}\} \Longrightarrow \{A_j^{(m+n)\times n} \mid_{j=1,2,\cdots,k}\}$$
$$\mathbf{B}^{m\times n} \Longrightarrow \mathbf{B}^{(m+n)\times n}$$

For multi-dimensional sequences, 1—m rows of the matrix at the right side of the arrow represent time series of m factors while $m+1$ to $m+n$ rows values of these factors at n different times.

If row vectors having the same serial number of all elements in the set $\{\mathbf{A}_j^{(m+n)\times n} \mid j=1, 2, \cdots, k\}$ comprise a matrix, then we have a further matrix set $\{\bar{\mathbf{A}}_i^{k\times n} \mid i=1, 2, \cdots, m+n\}$, where the ith element $\bar{\mathbf{A}}_i^{k\times n}$ is made up of the ith row vectors (whose total is k) of each element in $\{\mathbf{A}_j^{(m+n)\times n} \mid j=1, 2, \cdots, k\}$ which is a matrix composed of time series as row vectors of the same factor of all samples involved. It is apparent that $\bar{\mathbf{A}}_1^{k\times n}$, $\bar{\mathbf{A}}_2^{k\times n}$, $\cdots$, $\bar{\mathbf{A}}_m^{k\times n}$ are marked by time series of the same factor (a sum of m) of different samples (of k) as their row vectors, and $\bar{\mathbf{A}}_{m+1}^{k\times n}$, $\bar{\mathbf{A}}_{m+2}^{k\times n}$, $\cdots\cdots$, $\bar{\mathbf{A}}_{m+n}^{k\times n}$ by simultaneous values of various factors from different samples as their row vectors.

$\mathbf{B}^{(m+n)\times n}$ can be rewritten in a form of vector[5]

$\mathbf{B}^{(m+n)\times n}=(\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_{m+n})$.

similarly, the n-dimensional column vectors $\mathbf{B}_1$, $\mathbf{B}_2$, $\cdots$, $\mathbf{B}_m$ signify time series of different factors (with a total of m) in the sample to be forecasted while $\mathbf{B}_{m+1}$, $\mathbf{B}_{m+2}$, $\cdots$, $\mathbf{B}_{m+n}$ the simultaneous values of various factors of the sample to be predicted.

Based on Definitions 1 and 2 the AV of $\mathbf{B}_i$ to $\bar{\mathbf{A}}_i$ can be acquired:

$$\begin{matrix} s_{11}, & s_{12}, & \cdots, & s_{1k} \\ s_{21}, & s_{22}, & \cdots, & s_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ s_{(m+n)1}, & s_{(m+n)2}, & \cdots, & s_{(m+n)k} \end{matrix}$$

It should be noted here that s_{1j} is the approximability of $\mathbf{B}_1$ to the jth row of $\bar{\mathbf{A}}_1^{k\times n}$ and s_{2j} is that of $\mathbf{B}_2$ to the jth row of $\bar{\mathbf{A}}_2^{k\times n}$ and so on. All the jth rows (a sum of $m+n$) of $\bar{\mathbf{A}}_1^{k\times n}$, $\bar{\mathbf{A}}_2^{k\times n}$, ..., $\bar{\mathbf{A}}_{m+n}^{k\times n}$ make up the matrix $\mathbf{A}_j^{(m+n)\times n}$ perfectly.

According to Definition 3, $\mathbf{S}_j=(s_{1j}, s_{2j}, \cdots, s_{(m+n)j})$ is the AV of $\mathbf{B}^{m\times n}$ to $\mathbf{A}_j^{m\times n}$.

$\bar{s}_j$, the approximability of the dimensionally decreased $\mathbf{S}_j$, is obtained by use of Definition 4. If j is made to go through 1, 2, $\cdots$, k, then the AV, $\tilde{\mathbf{S}}=(\bar{s}_1, \bar{s}_2, \cdots, \bar{s}_k)$, of $\mathbf{B}^{m\times n}$ to the matrix set $\{\mathbf{A}_j^{m\times n}|_j=1,2,\cdots,k\}$ is derived. The component $\bar{s}_j$ is a similarity characterizing quantity for the multi-dimensional sequences of the sample to be predicted and that for the jth sample found in the historical data. The value reflects the similarity.

III. APPLICATION IN OPERATIONAL FORECASTS

The multi-dimensional sequence similarity method aims at providing a base for total typhoon rainfall forecast, which is based on the fact that the time is defined as t_0 when a typhoon comes across the arc line, which is drawn with Wenzhou, Zhejiang Province as the center and the Wenzhou-Okinawa distance as the radius (see Fig. 1) and then one or more historical samples closest to the typhoon are picked out by means of values of m_1, m_2, $\cdots$, m_{12} at t_0, t_{-1}, t_{-2}, t_{-3} and t_{-4}. In General, when the typhoon comes to the position at t_0, precipitation is likely to occur 18—36 hr later, implying that the forecast is *valid for about 24 hr in advance, if rainfall is possible over the Province.*

130 typhoons sorted out from the given area over the period of 1960—1983 are used for practice and as reference samples for operational forecasts.

1. *Factors Suitable for the Purpose*

- m_1 eastward extension from 105°E of the northwest or northern wind in the region covered by thick lines in Fig. 1 on the 500-hPa map.
- m_2 southward extension from 40°N of these winds in the same area.
- m_3 sea-level pressure of the typhoon's center.
- m_4 the direction in which the storm is moving.
- m_5 height of the 500-hPa isobaric surface at Quxian County, Zhejiang Province
- m_6 number of stations with unstable precipitation over the Province.
- m_7 distance from the typhoon's center to the 1000-hPa isobar taken on the center-Wenzhou connecting line on the surface map.
- m_8 6-hr variable in pressure of the center.
- m_9 eastward convergence of moisture flux along the path indicated by points 4, 5, 6 and 7 in Fig. 1.
- m_{10} eastward divergence at the same points.
- m_{11} the SE component of the wind vector at point 2 in the figure.
- m_{12} its SW component.

These factors are taken at intervals of 6 hr (denoted at t_i) and surface maps for 5 different observation times and 500-hPa charts are used for the purpose.

2. *Calculation in Steps*

Sequences of 12 factors and their values at 5 different times are written as 17 12-dimensional row vectors, with I_j denoting the number of rows, namely

$$
\begin{array}{lllllll}
I_1: & m_{11}, & \cdots, & m_{15}, & 0, & \cdots, & 0 \\
I_2: & m_{21}, & \cdots, & m_{25}, & 0, & \cdots, & 0 \\
\vdots & \vdots & & \vdots & & & \vdots \\
I_{12}: & m_{121}, & \cdots, & m_{125}, & 0, & \cdots, & 0 \\
I_{13}: & t_{01}, & \cdots & \cdots & \cdots & \cdots, & t_{012} \\
I_{14}: & t_{-11}, & \cdots & \cdots & \cdots & \cdots, & t_{-112} \\
\vdots & \vdots & & & & & \vdots \\
I_{17}: & t_{-41}, & \cdots & \cdots & \cdots & \cdots, & t_{-412}
\end{array}
$$

where I_1—I_{12} represent the sequences of these 12 factors and I_{13}—I_{17} their values at 5 observational times.

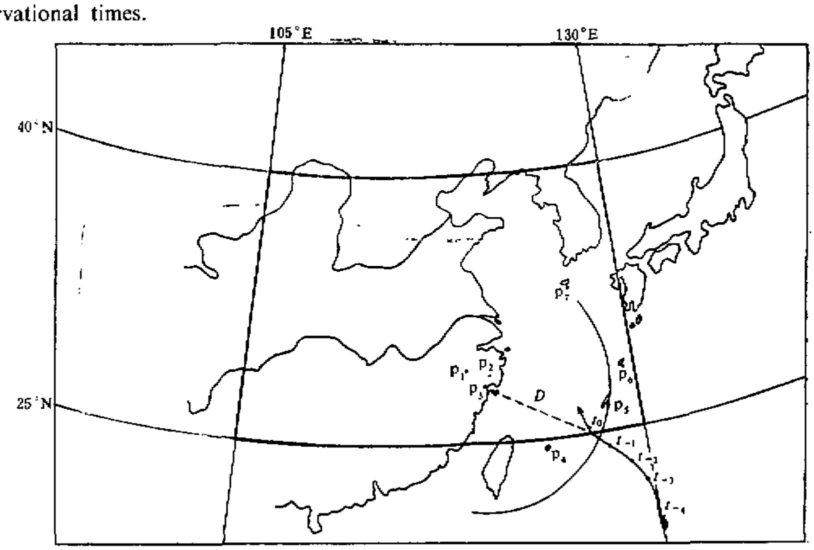

Fig. 1. The illustration for typhoon forecast. The forecast starting line is taken to be the arc having p_3 as a circle centre and the distance p_3—p_5 as a radius. The area covered by the thickened solid lines is that related to the determination of facfors m_1 and m_2 (see text). p_1–Quxian County, p_2–Kuocang Mountain, p_3–Wenzhou, p_4–Shihuan, p_5–Okinawa, p_6–Amami Gunto, p_7–Jeju, D–the connection between p_3 and the centre of a typhoon. t_i–the surface position of typhoon centre at corresponding time.

Calculation is generally made in two steps.

Step 1. To find the AV of the typhoon to be predicted to one or more of 130 historical typhoons in I_i $(i=1, 2, \cdots, 17)$

$$\mathbf{S}_i=(s_{i1}, s_{i2}, \cdots, s_{i130})$$

where s_{ij} $(j=1, 2, \cdots, 130)$ represents approximability of the typhoon in question to the jth historical typhoon in I_i. Then we have 17 130–dimensional AV's, with C_j denoting the number of columns, that is,

$$
\begin{array}{cccc}
c_1 & c_2 & \cdots & c_{130} \\
\vdots & \vdots & \vdots & \vdots \\
s_{11} & s_{12} & \cdots & s_{1130} \\
s_{21} & s_{22} & \cdots & s_{2130} \\
\vdots & \vdots & \vdots & \vdots \\
s_{171} & s_{172} & \cdots & s_{17130}.
\end{array}
$$

Step 2. To find $\tilde{s}_j$, the approximability of C_j ($j=1, 2, \cdots, 130$) to the maximal AV, $S_{max}=(2L, 2L, \cdots, 2L)$ (whose sum is 17). Then we have

$$\tilde{S}=(\tilde{s}_1, \tilde{s}_2, \cdots, \tilde{s}_{130}).$$

If $\tilde{s}_j$ is the maximum or one of the maxima of the 130 components, then the jth historical sample is none other than the one closest to or one of the closest samples to the typhoon to be expected.

3. *Determination of Approach Sequences*

(1) *The sequence length L (the number of terms)*

Obviously, the longer the L is and the more the sifting, the more clear the approximability would be of the sample to be predicted to the historical one. In fact, it is impossible to make L large enough and this is unnecessary since the concept of similarity is relative. If the only maximum approximability of the sample to be forecasted to a historical one is obtained through long-sequence approach, then it is also the maximum for a shorter-sequence approach, only with the defect of simultaneous occurrence of several maxima possible because of a shorter sequence and coarse sifting. So $L=20$ is taken for the present work.

(2) *The first term and its step length*

In view of the discretization of the factors different approach sequences are used for different I_i. The number of I_i determines that of approach seqences. In this study I_i has 17 sequences, each with the same number of terms ($L=20$) differing from each other only in the first term and its step length. Take I_i for example. For all I_i in the historical samples the maximal Euclidean distance $(x_i)_{max}$ and the minimal angle cosine $(y_i)_{min}$ are to be found which comprise the first term of the ith approach sequence$((x_i)_1, (y_i)_1)$

$=((x_i)_{max}, (y_i)_{min})$, whose step length is $\left|-\frac{(x_i)_{max}}{20}\right|$ and $\left|\frac{1-(y_i)_{min}}{20}\right|$.

4. *Application of This Method in the Typhoon Seasons of* 1982—1984

The initial form was worked out early in 1982 and tried out in the typhoon season. It was improved in 1983 and put into routine use for the typhoon period of 1984.

Four typhoons in 1984 are inside the arc line in Fig. 1. They are numbered as 8403, 06, 07 and 09. Their respective AV are respectively given in th Tables 1—4, where the bold-faced figures are the maximal components of corresponding vectors, which mean a historical typhoon closest to the one to be predicted.

Note that two maximal components are shown in the AV of the Typhoon 8403, One is $\tilde{s}_{53}=13$ and the other is $\tilde{s}_{55}=13$. The former is the approximability of Typhoon 8403 to 6813 and the latter of 8403 to 6909. Typhoon 6813 is one shifting northward at the position south of 25°N and east of 125°E, thus exerting no influence on this province. Fig. 2 depicts precipitation along the tracks of Typhoons 8403 and 6909.

The maximum of AV of Typhoon 8406 is $\tilde{s}_{41}=31$, being approximability of 8406 to 6408, one of which landed and the other did not on this province. Nevertheless, both had almost the same influence on this region especially in the total of precipitation, not exceeding 100 mm (see Fig. 3).

Table 1. The AV of the Typhoon 8403

$\tilde{S}_1=$	(5	7	1	7	7	9	5	5	3	1	11	7	7	9	7	5	5	3	6	3	5	3
	7	7	5	5	5	7	3	7	9	5	1	7	7	7	7	5	9	3	5	9	5	5
	11	5	5	9	5	5	3	7	13	5	13	11	9	1	9	3	7	9	1	7	9	1
	7	9	3	5	7	9	9	7	7	7	5	7	5	4	7	5	5	9	7	5	9	7
	7	5	7	7	5	5	7	9	9	5	7	5	5	7	11	7	5	9	7	3	1	9
	7	3	9	9	5	7	5	11	7	1	7	1	5	9	1	7	3	7	6	5)		

Table 2. The AV of the Typhoon 8406

$\tilde{S}_2=$	(17	15	19	17	13	19	13	17	21	15	15	23	25	19	21	19	17	19	19	15	15	15
	21	15	19	15	17	15	23	19	17	19	15	19	15	17	21	19	21	21	15	19	19	15
	15	17	17	21	21	15	17	15	17	17	21	17	15	15	13	17	31	19	15	19	13	15
	17	17	15	17	17	23	17	21	17	17	21	13	15	17	19	21	21	19	17	21	15	17
	19	17	23	19	21	17	17	13	15	19	19	19	17	19	21	19	19	21	19	17	11	15
	19	17	17	15	17	17	15	17	15	17	17	15	15	21	11	21	15	17	17	17)		

Table 3. The AV of the Typhoon 8407

$\tilde{S}_3=$	(19	17	13	19	17	19	17	15	15	13	19	19	15	19	15	15	15	15	19	15	15	15
	21	19	15	17	15	15	17	19	19	17	13	19	17	19	21	17	19	17	17	21	15	15
	19	17	19	19	17	15	17	17	19	19	21	21	19	13	19	15	17	25	13	21	19	13
	13	17	13	19	15	15	19	19	19	21	15	19	19	13	15	17	19	21	19	15	19	15
	21	17	17	19	13	19	17	17	21	17	19	17	17	17	19	19	13	19	17	19	13	19
	15	17	15	19	17	17	17	19	19	15	21	15	15	21	11	19	13	17	19	17)		

Table 4. The AV of the Typhoon 8409

$\tilde{S}_4=$	(15	17	17	21	17	19	15	17	23	11	15	23	23	19	21	19	21	17	21	19	17	17
	21	15	21	19	17	15	21	17	19	19	17	17	15	21	21	17	17	19	17	19	15	15
	17	21	21	25	25	15	23	17	17	19	21	21	17	13	15	19	19	23	15	21	15	17
	17	17	17	19	13	19	21	23	19	21	21	15	19	15	21	19	23	23	19	21	19	17
	19	19	23	21	23	21	15	13	17	21	29	19	19	21	27	21	17	21	19	15	11	19
	21	17	19	19	17	17	15	17	19	19	19	15	17	19	11	19	17	19	17	17)		

In the case of Typhoon 8407 the maximum is $\tilde{s}_{62}=25$, which is approximability of 8407 to 7010. Both landed between Northern and Central Fujian Province causing rainfall of 100—200 mm in the south-east of Zhejiang. But the region enclosed by the 100-mm isopluvial of Typhoon 8407 is as large as that by the 50-mm isoline of 7010. This is associated

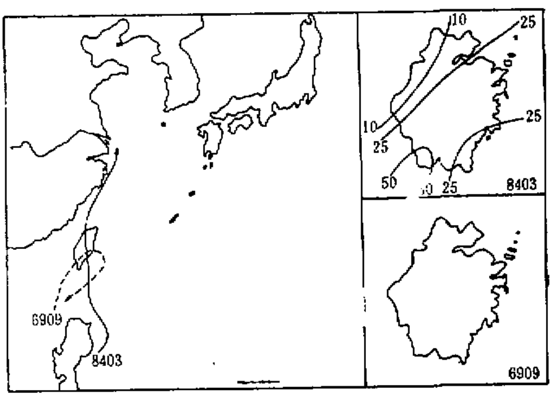

Fig. 2. The comparison between forecasted typhoon 8403 and its similar sample 6909. Shown in the left half of the figure is the comparison of typhoon tracks between the forecasted (solid line) and its similar sample (dashed line). The right half is that of the total precipitations (in mm).

with Typhoon 8407 landing slightly further north and going straight northward (see Fig. 4).

For Typhoon 8409, the maximum is $\tilde{s}_{99}=29$. This is approximability of Typhoon 8409 to 7909. whose courses and rainfall are close to each other. Typhoon 7909 produced more precipitation because of its course being further west with respect to the other (see Fig. 5).

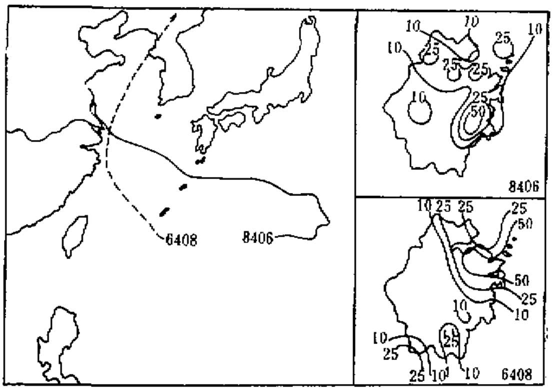

Fig. 3. As in Fig. 2, except for the forecasted Typhoon 8406 and its similar sample 6408.

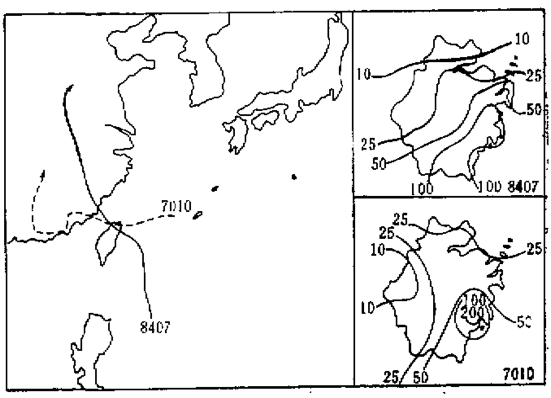

Fig. 4. As in Fig. 2, except for the forecasted Typhoon 8407 and its similar sample 7010.

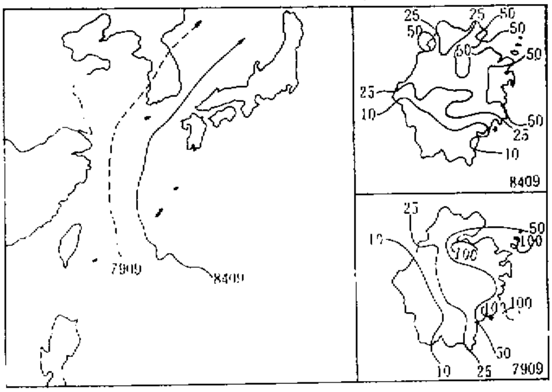

Fig. 5. As in Fig. 2, except for the forecasted Typhoon 8409 and its similar sample 7909.

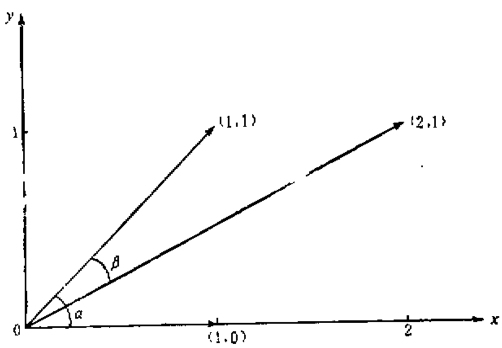

Fig. 6. The schematic diagram for the relationship between two vectors.

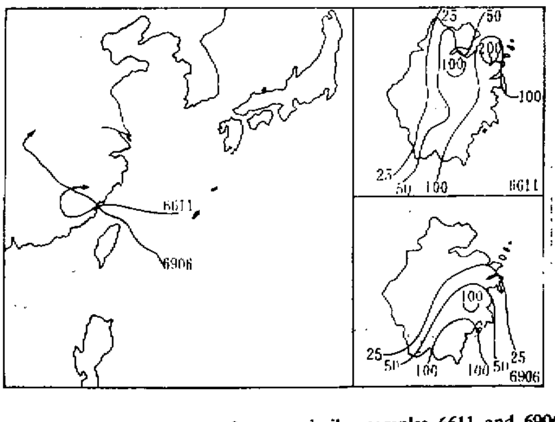

Fig. 7. As in Fig. 2, except for two similar samples 6611 and 6906.

IV. REMARKS ON THIS METHOD

1. *Diminution of Similar Samples Circle Searched for*

The principal advantage of the method lies in that Euclidean distance and the included angle cosine are alternately used for the vectors. Yet cluster analysis (e. g. the K-mean method) employs merely Euclidean distance. Theoretically, that one vector has the same distance to any other two does not necessarily mean that the included angle cosine has the same value for them (see Fig. 6). The vectors (1,1) and (2,1) have the distance of 1 and the vectors (1,1) and (1,0) have the same but clearly their included angle values of cosine are different. However, in cluster analysis with Euclidean distance allowed for only similarity of the vector (1, 1) to (2,1) is exactly the same as that of (1,1) to (1,0) . The method presented here is able to show that the former is higher on account of $\beta<\alpha$. Undoubtedly, this makes the size of sieve mesh smaller, thus leading to the decrease of the number of samples available for choice.

Take, for example, Typhoon 8407. When the K-mean method is adopted with $K=12$. It will be related to 13 past typhoons. In the subsequent work diminution of Euclidean distance is repeatedly done with the 13 tempests and eventually 4 typhoons of these are sorted out which are the closest to Typhoon 8407 in nature. They are 6611, 6906, 6909 and 7010, the third of which, quite different from the others, brought little, if any, precipitation to this province. This presents difficulty in making decision of forecast.

On the other hand, the multi-dimensional sequence similarity method can exclude Typhoon 6909 (and meanwhile Typhoons 6611 and 6906), thereby objectiveness of forecast-making is improved, but limitation of forecast effectiveness is increased. Compare Fig. 4 with 7. In the case of the value of the rainfall center and area covered, 7010 is closest to 8407; as far as the region enclosed by the 100-mm isopluvial is concerned, 6611 is most similar to 8407; and for the track and landing spot, 6906 is in close proximity to 8407. All these show the weakness of similarity analysis because wholly similar samples are hardly available in the historical events of the real atmosphere. Therefore, the problem concerning how to use this method in conjunction with the general cluster analysis method needs to be dealt with for prediction purposes.

2. *Establishment of an Objective Similarity Criterion*

It is apparent from the components of $\bar{s}_2$, $\bar{s}_3$ and $\bar{s}_4$ that their maxima show greater magnitude of $\geqslant 25$ and hence the similar samples obtained are closer to that to be forecasted while $\tilde{S}_1$ has the maximal component of 13 only, causing greater difference between both.

Therefore, if a threshold of approximability, say, 20, is established beforehand, then with an AV derived assessment can be made of the effectiveness of the multi-dimensional sequence similarity method for the forecast that has been done. In case an AV has its maximum below the threshold, it means that this method is proved to be useless and other methods should be suggested. If the maximal component lies above it, the method is of value. The greater the maximum, the closer the historical sample found would be to the one to be predicted.

3. *Adjustment of Factor Weighing by Altering Magnitude of the First Term in an Approach Sequence*

If Euclidean distance x_i of the first term in the ith approach sequence is increased and the included angle cosine y_i decreased (<0), then the step length $\left|\frac{0-x_i}{20}\right|$ and $\left|\frac{1-y_i}{20}\right|$ will grow in their absolute magnitude. This means that the size of sieve mesh for the ith factor would be larger, resulting in smaller difference between approximabilities of the ith factor $(s_{i1}, s_{i2}, \cdots, s_{i130})$. The result would be that during the next calculation of $\bar{s}_j$ the ith factor has less effect, thus achieving the diminution of its weighing.

If x_i is large enough to be $x_i=20\ (x_i)_{max}$ and y_i small $y_i=20\ (y_i)_{min}-19$, the 19th term (i. e. the last but one) of the ith approach sequence will be $((x_i)_{max}, (y_i)_{min})$. In this case the size of sieve mesh is so large that any Euclidean distance and the cosine can be approximated to the end by the ith approach sequence. This means that the ith factor, after the first-step operation, will assume (40, 40, ⋯, 40) (with a sum of 130), and for the next calculation have no effect, i. e. zero weighing. If augment of the weighing were required, the first term would be given smaller Euclidean distance and greater cosine value.

REFERENCES

[1] 曹鸿兴，局地天气预报的数据分析方法，气象出版社，1983.
[2] 江泽坚，吴智泉，实变函数论，人民教育出版社，1978.
[3] G.W. 斯图尔特，矩阵计算引论，上海科学技术出版社，1980.