The Annual Average Stream Flow for Hydro-electric Power

By

Tatsuro OKUBO*, Hiroshi NISHIHARA† and Takashi SOMA*

(Received July 30, 1960)

In a large combined hydro-steam power system as in Japan, it is necessary to know what flowing condition of rivers may be expected to hold on an average for a long term, in order to plan for the economical development of hydro-power.

However, the flowing condition not only varies in its total annual stream flow each year, but even seasonally within a given year, which makes it difficult for us to deal directly with the duration curves of stream flow or the daily stream flow curves which have been hitherto employed in developing hydro-power, so that we have to consider as a parameter the annual average stream flow for each year on which these depend.

We need not take the average of all the data obtained in the past in order to estimate the annual average stream flow expected to hold for a long term to come, but must resort to the reasonable method of so-called time series analysis with its theoretical foundation, which takes into consideration the periodicity of the fluctuation contained in the stream flow and other factors.

In this paper we propose to estimate the annual average stream flow expected to hold on an average for a considerably long term by means of an analysis, as easy and practical as possible, of a stream flow of comparatively small sample size.

I. Introduction

In a large combined hydro-steam power system as in Japan, it is necessary to know what flowing condition of rivers may be expected to hold on an average for a long term to come, if we are to plan for the economical development of hydro-power combined with load and steam-power. But the flowing condition of rivers in general, under manifold natural influences, is not only subject to seasonal change, but suffers considerable fluctuation from year to year on account of the peculiar configuration and climate of Japan, so that no satisfactory method for estimating the flowing condition has yet been proposed.

One of the reasons for this lies in the fact that, in Japan at least, no long-term data have been kept and such data are necessary for very accurate stochastic studies. In addition to gathering accurate data year by year, we must strive to

* Department of Electrical Engineering
† Department of Nuclear Engineering

develop a new method for making the best use of them.

In view of these considerations, this paper proposes to estimate the annual average stream flow expected to hold on an average for a long term by means of an analysis, as easy and practical as possible, of a stream flow of comparatively small sample size.

In order to plan for power generation with a high degree of accuracy, efforts towards improvement in the accuracy of the measurement of the stream flow are needed, and we believe that the results obtained by the method given in this paper will increase in reliability as more accurate data on stream flow become available.

II. The Annual Average Stream Flow

The estimation of the annual or monthly generated energy of a hydro-power plant and plans for economical power generation, which take into consideration the auxiliary steam-power corresponding to the hydro-power, are both dependent upon the flowing condition of rivers.

It is possible, in the case of an existing power plant or system, to make an extrapolatory estimation of the annual available power for future years by means of data from the past[1]; but in the case of a plant under planning, we have to estimate the annual available power to be provided by it when connected to a system, upon the basis of the rule of operation imposed in accordance with the flowing condition at the site under planning and the capacity of the plant[2].

In the case of a plant of the run-off-river type, we can estimate the annual available power by assuming beforehand the maximum available stream flow with regard to the flowing condition of the site and regarding any daily stream flow which may exceed that value as equal to it.

In this paper let us assume that there is no excess hydro-power, considering the recent increasing importance of steam-power development and the gradual tendency to develop pumped-storage plants for excess steam-power as well as hydro-power.

In the case of a plant of the pondage or storage type, we have only to estimate an equivalent average stream flow taking into consideration its pondage or storage capacity and the load demand of the power system.

Efficient utilization of stream flow leads to an increase in the maximum available stream flow, so that in a large combined hydro-steam power system, for a given load demand, it inevitably involves the saving of fuel by suspending the steam-power plants during the wet season so as to give time enough for repairs. But the maximum available stream flow must be determined from economical considerations because it also leads to an increase in the annual costs of hydro-power plants, long-distance power transmission lines and substations, together with the increasing possibility of causing an excess hydro-power, in a system in which hydro-power predominates.

However, the flowing condition of rivers, not only varies in its total annual stream flow each year, but even seasonally within a year, as is evident, e.g. from Figs. 1 (a) and (b), showing the annual duration curves of stream flow in every year for the Rivers A and B.

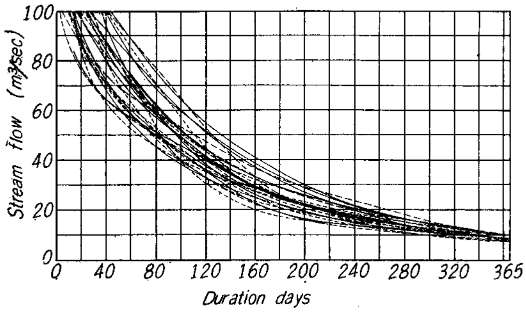

Fig. 1 (a). The annual duration curves of stream flow (the River A).

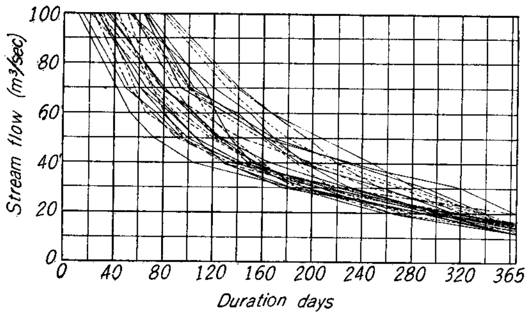

Fig. 1 (b). The annual duration curves of stream flow (the River B).

Hitherto, the duration curve of stream flow has generally been used in developing hydro-power plant, but in estimating the load share for hydro-power in a combined hydro-steam power system, the annual average stream flow alone can to some degree serve the purpose, enabling us to obtain easily more accurate results than by using directly the duration curve of stream flow. In cases where the daily stream flow curve is required, the annual average stream flow also serves as a basis for it.

Two methods are possible for determining the annual average stream flow corresponding to the annual available power, i.e.

(1) We take as the parameter the annual average stream flow where the 3-, 6-, 9-month-flow or the drought-flow—defined in this paper as the average value of the three smallest daily stream flows for each year—is regarded as the maximum available stream flow. This method stresses the annual duration days of stream flow.

(2) We take as the parameter the annual average stream flow where, with the drought-flow of a particular year, e.g. the typical year, as the basis, the stream flow corresponding to its appropriate multiple—the so-called "drought multiple factor" represented by J—is regarded as the maximum available stream flow. This stresses the maximum available stream flow.

Figs. 2 (a) and (b) illustrate the fluctuation curves, represented in order of the calendar years, of the annual average stream flow of the Rivers A and B computed by means of the above two methods, which give much the same results. The results by the **latter method**, however, generally give somewhat smaller degrees of annual fluctuation than the **former**. But it is deemed reasonable to use the **former method** in determining the maximum available stream flow at the site of the projected plant, because the value of J in the **latter method** is also determined by provisionally forecasting the duration days corresponding to a

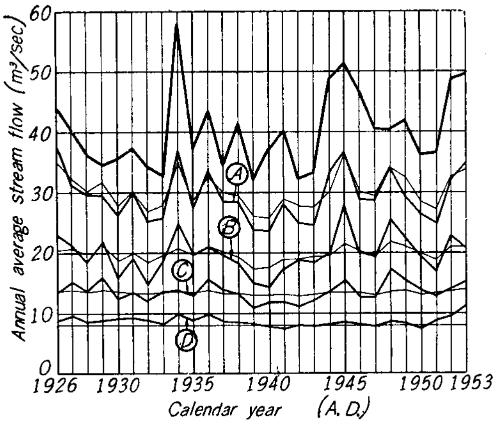

Fig. 2 (a). The fluctuation curves of the annual average stream flow (the River A).

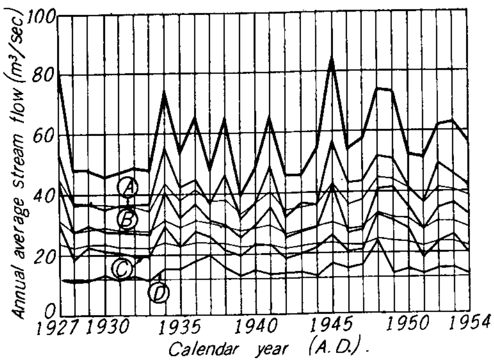

Fig. 2 (b). The fluctuation curves of the annual average stream flow (the River B).

Note: —— The total annual average stream flow
Ⓐ··· where the 3-month-flow is taken as the maximum available
Ⓑ··· " 6-month-flow "
Ⓒ··· " 9-month-flow "
Ⓓ··· " drought-flow "
—— each year
—— typical year

maximum available stream flow. Of course it hardly needs saying that in the case of an existing plant with its definite maximum available stream flow, the annual average stream flow determined by **method (2)** should serve as the basis.

In this paper we propose to inspect the annual average stream flows of the Rivers A and B on the assumption that the total annual average stream flow and the annual average stream flows regarding the 3-, 6-, 9-month-flow or drought-flow for each year as the maximum available stream flow, and regarding 6.90, 3.08, 1.67 for the River A and 4.71, 2.89, 2.06 for the River B as the values of J respectively.

III. The Inspection of the Periodicity

It is generally thought that fluctuations in the stream flow include both the seasonal fluctuations within a year and the year-to-year fluctuations, and that these again consist of a hidden periodic component and a random component.

The annual average stream flows, for example, of the Rivers A and B show, as seen from Figs. 2 (a) and (b), complicated fluctuations from year to year, but it often happens that the stream flow of our rivers has a fairly prominent periodic component of about 10 years, so that in order to estimate the annual average stream flow, we have to take into consideration the periodicity of the fluctuation of the stream flow instead of merely adopting the arithmetical mean of all the data obtained in the past, as is generally done in Japan.

In other words, in dealing statistically with the data of the stream flow—e.g. the annual average stream flow—we have to take into consideration the periodic component (hidden periodic component) and the trend contained in the fluctuation of the stream flow by analyzing the data as a sample of a time series rather as a random sample.

But it is necessary in their statistical treatment, that the past data serving as a basis of computation should be of a longer term than this period, and an even longer one in those cases which include, as in the instances cited in this paper, evidently abnormal years. Where a trend exists, data for a longer term are required. It is difficult, therefore, to decide how much data should be prepared in order to analyze the fluctuation of the stream flow.

Since it often happens that in Japan there is not much available data for a long term to enable us to detect the period, it is desirable to give an all-round judgment after inspecting each instance from various points of view. We shall give an account of practical methods for analyzing the fluctuation of the stream flow in order.

(1) The Correlation Diagram

We obtain the correlation diagram if we plot the value of the annual average stream flow per year and that of k years later respectively in abscissa and ordinate for each year. Though it only helps to give a very rough information for the detection of the period, it is useful as an expedient method of discovering the abnormal years.

Figs. 3 (a) and (b) show the correlation diagrams of the Rivers A and B for the case of $k=11$, in which there seems no remarkable correlation, but the correlation of the 3-year moving average stream flow (cf. the next paragraph), owing to its smoothing effect for random components, are to some degree noticeable except for abnormal years. It is the same in the case of $k=10$. On the other hand, any correlation is hardly noticeable for other values of k.

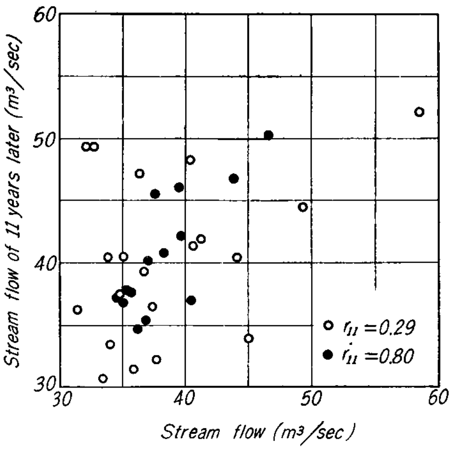

Fig. 3 (a). The correlation diagram between the average stream flow for each year and that of 11 years later (the River A).

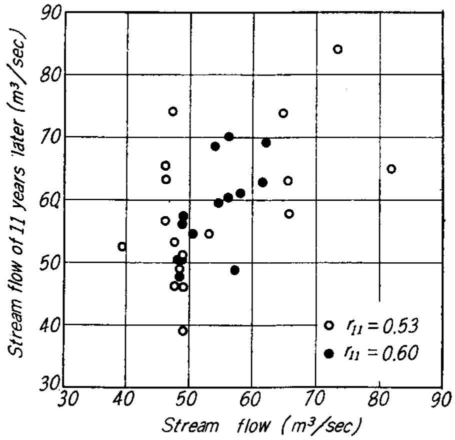

Fig. 3 (b). The correlation diagram between the average stream flow for each year and that of 11 years later (the River B).

Note: ○··· The annual average stream flow
●··· The 3-year moving average stream flow for $K=2$

(2) The Moving Average

The series of moving average serves for smoothing the fluctuation of a sample of time series by removing random components, and sometimes for extracting or removing a particular periodic component contained in the sample.

Representing a sample of time series by $x(1), x(2), \cdots, x(i), \cdots, x(n)$—abbreviated "$\{x(i)\}$" in the following pages—the series of (weighted) $(2s+1)$-year moving average is generally, as is well known, represented by the following equation:

$$y(i) = \sum_{j=-s}^{s} a_j x(i+j), \quad i = s+1, s+2, \cdots, n-s. \tag{1}$$

If, as a particular instance, all weighting coefficients $a_j = 1/(2s+1)$, $(j=0, \pm 1, \pm 2, \cdots, \pm s)$, $y(i)$ is the series of arithmetical moving average. In order, for example, to multiply the amplitudes of respective periodic components—sinusoidal waves—with the period $k_0, k_1, \cdots, k_s$, by respectively $C_0, C_1, \cdots, C_s$ times, we have recourse to the $(2s+1)$-year moving average method with symmetrical coefficients, as

$$y(i) = a_0 x(i) + \sum_{j=1}^{s} a_j \{x(i-j) + x(i+j)\}^*. \tag{2}$$

* The weighting coefficients $a_j (j=0, 1, \cdots, s)$ can be obtained by solving the following equation:

$$\begin{pmatrix} 1 & 2\cos(2\pi/k_0) \cdots 2\cos(2s\pi/k_0) \\ 1 & 2\cos(2\pi/k_1) \cdots 2\cos(2s\pi/k_1) \\ \cdots & \cdots\cdots\cdots \\ 1 & 2\cos(2\pi/k_s) \cdots 2\cos(2s\pi/k_s) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_s \end{pmatrix} = \begin{pmatrix} C_0 \\ C_1 \\ \vdots \\ C_s \end{pmatrix}$$

The simplest moving average, therefore, for eliminating the periodic component with period K—to be referred to as the "K-year periodic component"—is given by

$$y_K(i) = a_1 x(i-1) + a_0 x(i) + a_1 x(i+1)\,,$$
$$a_1 = \frac{1}{2\{1-\cos(2\pi/K)\}}\,, \quad a_0 = -\frac{\cos(2\pi/K)}{1-\cos(2\pi/K)}\,. \tag{3}$$

The weighting coefficients a_1 and a_0 for various values of K are given in Table 1.

Table 1. The weighting coefficients a_1, a_0 for various values of K.

Coefficients \ K	2	3	4	5	6
a_1	1/4	1/3	1/2	0.724	1
a_0	1/2	1/3	0	−0.448	−1

The moving average given by Eq. (3) not only eliminates the K-year periodic component, but affects other periodic components as well, so that the amplitude of the k-year periodic component is generally multiplied as a result of the moving average represented by Eq. (3), by the factor

$$C_K(k) = \frac{\cos(2\pi/k) - \cos(2\pi/K)}{1-\cos(2\pi/K)}\,. \tag{4}$$

The values of the amplitude multiplying factor $C_K(k)$ obtained by Eq. (4) are given in Table 2 and Fig. 4 (a). Further, Fig. 4 (b) shows the amplitude multiplying factor of the k-year periodic component by the successive n operations of the 3-year moving average for $K=2$. These figures illustrate that the periodic

Table 2. The amplitude multiplying factor $C_K(k)$.

k \ $C_K(k)$	$C_2(k)$	$C_3(k)$	$C_4(k)$	$C_5(k)$	$C_6(k)$
2	0	−0.333	−1.000	−1.894	−3.000
3	0.250	0	−0.500	−1.171	−2.000
4	0.500	0.333	0	−0.447	−1.000
5	0.655	0.539	0.309	0	−0.382
6	0.750	0.667	0.500	0.276	0
8	0.854	0.805	0.707	0.576	0.414
10	0.905	0.873	0.809	0.724	0.618
15	0.957	0.942	0.914	0.875	0.827
20	0.976	0.967	0.951	0.929	0.902

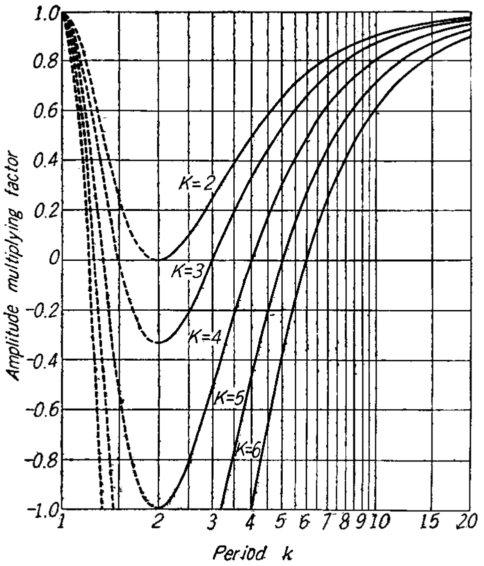

Fig. 4 (a). The amplitude multiplying factor $C_K(k)$.

Note: In the case of the 3-year moving average for various values of K

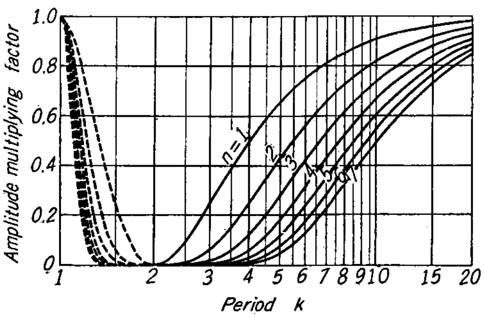

Fig. 4 (b). The amplitude multiplying factor $\{C_2(k)\}^n$.

Note: In the case of the successive n operations of the 3-year moving average for $K=2$

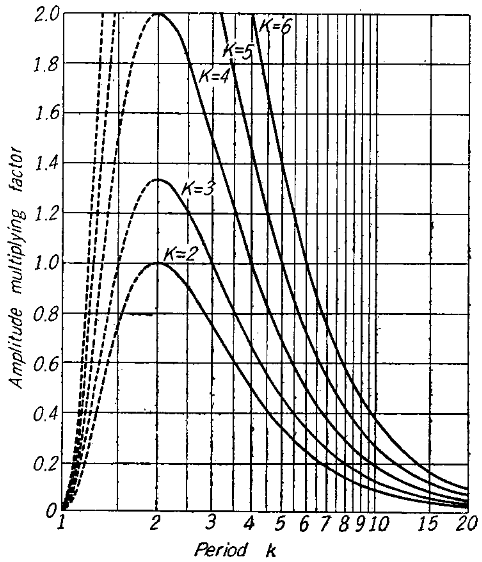

Fig. 4 (c). The amplitude multiplying factor $C_K'(k)$.

Note: In the case of the 3-year moving average of Eq. (6)

components with comparatively short periods can be effectively eliminated while retaining longer periodic components almost in their entirety, by adopting an appropriate moving average.

It is particularly effective to adopt the moving average obtained by putting $K=2$ in Eq. (3), i.e.

$$y_2(i)=\frac{1}{4}\{x(i-1)+2x(i)+x(i+1)\}, \tag{5}$$

in order to smooth the fluctuation curve by removing the random components or short periodic components contained in the fluctuation of the annual average stream flow. Much the same result can be obtained by using the 5-year arithmetic moving average. Figs. 5 (a) and (b) show the total annual average stream flows of the Rivers A and B together with the 3-year moving average for $K=2$ and the 5-year arithmetic moving average of the total annual average stream flows. These moving average stream flows are conspicuously smoothed away as compared with the annual average stream flows, and the existence of a component with a period of about 11 years is brought to light. Since the sample size diminishes by 2 for the moving average of Eq. (3) or (5), and by 4 for the 5-year arithmetic moving average, the former is preferable to the latter unless we are to

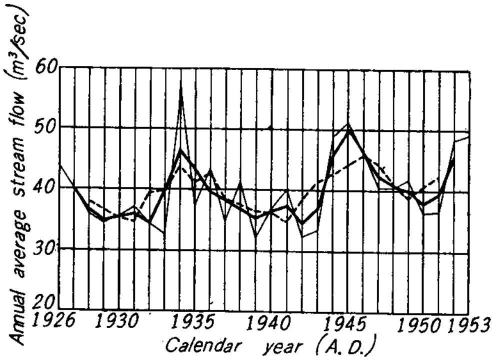

Fig. 5 (a). The total annual average stream flow and its moving average (the River A).

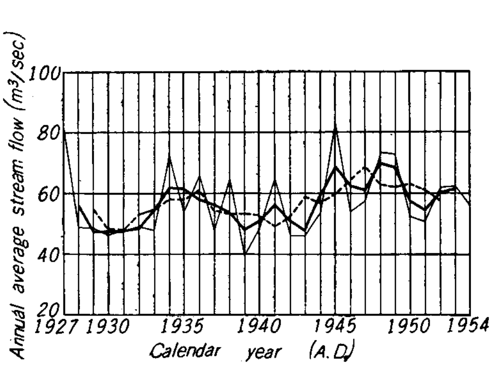

Fig. 5 (b). The total annual average stream flow and its moving average (the River B).

Note: —— The total annual average stream flow
▬ The 3-year moving average stream flow for $K=2$
- - - The 5-year arithmetic moving average stream flow

diminish the sample size. It is especially to be avoided in the instances of our rivers to diminish their sample sizes by operating the moving average.

It is also possible to eliminate, while retaining the amplitude of the K-year periodic component, the d.c. component and the longer periodic components, by

$$y_K'(i) = a_1'x(i-1)+a_0'x(i)+a_1'x(i+1),$$
$$a_1' = -\frac{1}{2\{1-\cos(2\pi/K)\}}, \quad a_0' = \frac{1}{1-\cos(2\pi/K)}. \qquad (6)$$

The values of a_1' and a_0' for various values of K are given in Table 3, where the sum of the weighting coefficients is zero.

Table 3. The weighting coefficients a_1', a_0' for various values of K.

Coefficients \ K	2	3	4	5	6
a_1'	$-1/4$	$-1/3$	$-1/2$	-0.724	-1
a_0'	1/2	2/3	1	1.428	2

The amplitude multiplying factor $C_K'(k)$ is represented by

$$C_K'(k) = \frac{1-\cos(2\pi/k)}{1-\cos(2\pi/K)} . \tag{7}$$

The values of the amplitude multiplying factor for various values of k and K are given in Table 4 and Fig. 4 (c). As is evident from Fig. 4 (c) the trend

Table 4. The amplitude multiplying factor $C_K'(k)$.

k \ $C_K'(k)$	$C_2'(k)$	$C_3'(k)$	$C_4'(k)$	$C_5'(k)$	$C_6'(k)$
2	1.000	1.333	2.000	2.894	4.000
3	0.750	1.000	1.500	2.171	3.000
4	0.500	0.667	1.000	1.447	2.000
5	0.345	0.461	0.691	1.000	1.382
6	0.250	0.333	0.500	0.724	1.000
8	0.146	0.195	0.293	0.424	0.586
10	0.095	0.127	0.191	0.276	0.382
15	0.043	0.058	0.086	0.125	0.173
20	0.024	0.033	0.049	0.071	0.098

and the long periodic components can be effectively eliminated by means of the moving average

$$y_2'(i) = \frac{1}{4}\{-x(i-1)+2x(i)-x(i+1)\} . \tag{8}$$

Moreover, as we obtain nothing else than the moving average of Eq. (5) by subtracting the series of moving average of Eq. (8) from the original series, the series of moving average of Eq. (5) may be substituted for the periodic components. (Cf. **IV.** (**2**) for this instance.)

In order to eliminate both the L- and M-year periodic components contained in the original series, we have only to proceed with the 3-year moving average for $K=L$ and $K=M$ successively, when the periodic component $\xi_k(i)$ is multiplied by $C_L(k)\cdot C_M(k)$. Many periodic components are likewise eliminated. Figs. 6 (a)

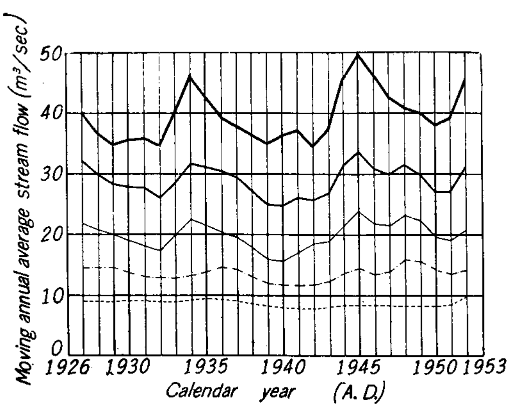

Fig. 6 (a). The 3-year moving average for $K=2$ of the annual average stream flows (the River A).

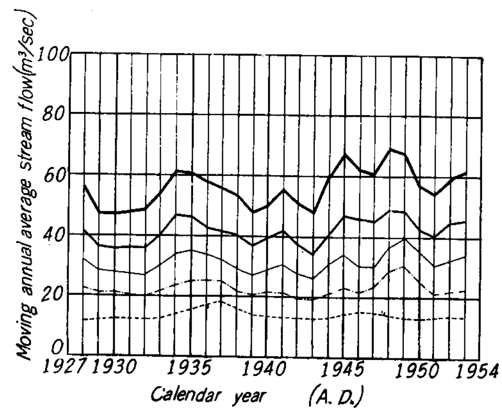

Fig. 6 (b). The 3-year moving average for $K=2$ of the annual average stream flows (the River B).

Note: —— The total annual average stream flow
—— where the 3-month-flow for each year is taken as the maximum available
—— " 6-month-flow " "
- · - " 9-month-flow " "
- - - " drought-flow " "

and (b) show the 3-year moving average for $K=2$ of the annual average stream flows of the Rivers A and B, when as the maximum available stream flow the total annual average stream flow, 3-, 6-, 9-month-flows and the drought-flow for each year are in turn adopted. In any of these cases the periodicity of about 10~13 years is still more evident.

(3) The Correlogram

The correlation coefficient r_k between the annual average stream flow for any year and that of k years later is named "the serial auto-correlation coefficient" of the annual average stream flow. The correlogram means the graph of r_k. The serial auto-correlation coefficient r_k for the sample of time series $x(1)$, $x(2)$, $\cdots, x(n)$ is represented, as is well known, by

$$\begin{aligned} r_k &= \frac{1}{n-k}\sum_{i=1}^{n-k}\{x(i)-\bar{x}_1\}\{x(i+k)-\bar{x}_2\}/S_1S_2, \\ \bar{x}_1 &= \frac{1}{n-k}\sum_{i=1}^{n-k}x(i), \quad \bar{x}_2 = \frac{1}{n-k}\sum_{i=k+1}^{n}x(i), \\ S_1^2 &= \frac{1}{n-k}\sum_{i=1}^{n-k}\{x(i)-\bar{x}_1\}^2, \quad S_2^2 = \frac{1}{n-k}\sum_{i=k+1}^{n}\{x(i)-\bar{x}_2\}^2. \end{aligned} \tag{9}$$

Figs. 7 (a), (b) and Figs. 8 (a), (b) are each the correlogram of the annual average stream flows of the Rivers A and B (shown in Figs. 4 (a), (b)) and that of the 3-year moving average stream flows for $K=2$ (shown in Figs. 6 (a), (b)).

It is easy to detect the hidden periodic components behind the random components, because in a correlogram the periodic components contained in the original time series are manifested. That is, where periodic components exist, $|r_k|$ does not damp to zero when k tends to infinity and the correlogram fluctuates with the same period as the original series.

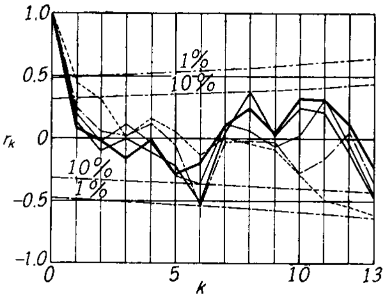

Fig. 7 (a). The correlogram of the annual average stream flows (the River A).

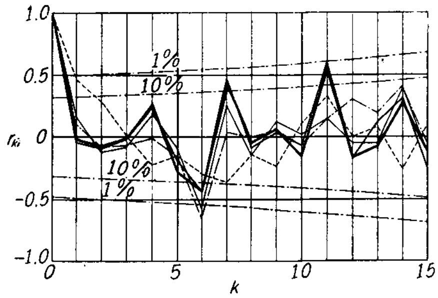

Fig. 7 (b). The correlogram of the annual average stream flows (the River B).

Note:				
━━	The total annual average stream flow			
——	where the 3-month-flow for each year is taken as the maximum available			
——	"	6-month-flow	"	"
-・-	"	9-month-flow	"	"
---	"	drought-flow	"	"

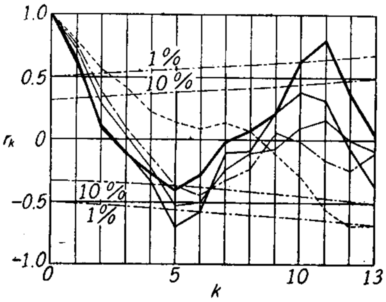

Fig. 8 (a). The correlogram of the 3-year moving average stream flows for $K=2$ (the River A).

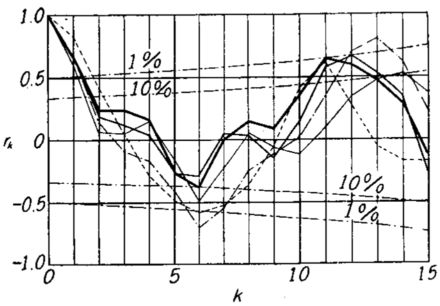

Fig. 8 (b). The correlogram of the 3-year moving average stream flows for $K=2$ (the River B).

Note:				
━━	The total annual average stream flow			
——	where the 3-month-flow for each year is taken as the maximum available			
——	"	6-month-flow	"	"
-・-	"	9-month-flow	"	"
---	"	drought-flow	"	"

In Figs. 7 and 8, the chain lines indicate the 1 or 10% level of significance of the correlation coefficient. The α% level of significance of the correlation coefficient is defined, as is well known, by

$$\alpha(\%) = Pr\{|r|>r_0\},\qquad(10)$$

where r means the sample correlation coefficient out of the non-correlated normal population. In these Figures are illustrated the significance levels for the ordinary correlation coefficient, for no perfect table of the significance level of the serial correlation coefficient has yet been prepared[3]. It is supposed that no great difference exists between them[3]. It is usual in the statistical hypothesis test to employ the 5% level of significance, but in this paper we shall recognize a specially significant correlation only if it exceeds the 1% level, and not at all if it does not reach the 10% level.

According to Figs. 7 (a) and (b), the correlograms for the original series do not reach the 10% level of significance, but according to Figs. 8 (a) and (b) some of the correlograms of the 3-year moving average stream flows for $K=2$ exceed at $k=5$ the 10% level (negative side) of significance and at $k=10\sim11$ the 1% level, so that we have been able to detect the hidden periodicity of 10~11 years contained in the original series by eliminating from it the random or short periodic components.

The fact that the annual average stream flow includes the 10-~11-year periodic components may be connected with the influence of the number of sunspots upon meteorological phenomena.

Further, in Figs. 9 (a) and (b) are shown, for the inspection of the actual condition of the periodic components, the correlograms of the 3-year moving average stream flows of the 5 respective parts which may be given by dividing the duration curve of stream flow for each year, as shown in these Figures, at the levels of each year's 3-, 6-, 9-month-flows and the drought-flow of the Rivers A and B.

As is evident from this, the fluctuation period of the River A generally gets

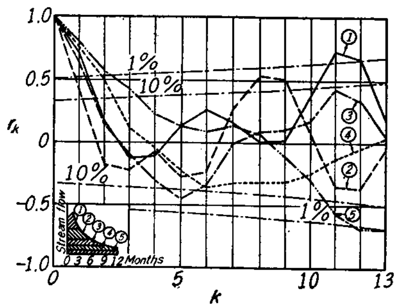

Fig. 9 (a). The correlogram of the 3-year moving average stream flows for $K=2$ of 5 respective parts (the River A).

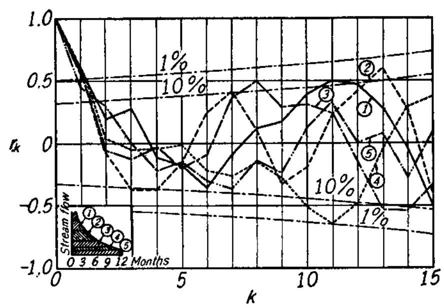

Fig. 9 (b). The correlogram of the 3-year moving average stream flows for $K=2$ of 5 respective parts (the River B).

longer as the duration term increases, while the River B, though a somewhat similar tendency can be discerned, presents a more complex fluctuation.

Likewise in Figs. 10 (a) and (b) are shown the correlograms of the monthly average stream flows of the Rivers A and B for every month, from which it is evident that, while in the River A the above-mentioned periodicity of 10~11 years can be discerned from May to December, and another longer periodicity of 16~20 years is present from January to April, but a periodicity somewhat longer than 10~11 years is also clearly present in the River B during the dry winter months.

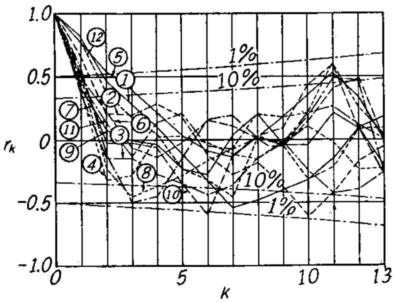

Fig. 10 (a). The correlogram of monthly average stream flows (the River A).

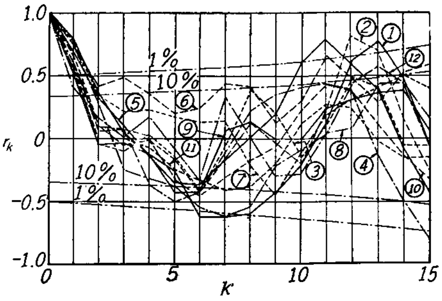

Fig. 10 (b). The correlogram of monthly average stream flows (the River B).

Note: The figure in each circle shows the month

(4) The Periodogram

The hidden periodicity in a time series can also be detected by the periodogram. This is essentially the graph of the spectrum distribution, while the correlogram is that of the serial correlation coefficient.

However, a practical or convenient periodogram is not a graph of the spectrum distribution itself but rather usually one which shows a maximum in the neighbourhood of the hidden period. We shall show a few convenient methods for it.

Let us make a series $y(1), y(2), \cdots, y(k)$ arranging the time series $x(1), x(2), \cdots, x(n)$ as in the following table. (It is advisable to make the mean of the series zero by subtracting the mean beforehand from $\{x(i)\}$, $\{y(i)\}$ etc.)

	$x(1)$	$x(2)$	$\cdots$	$x(k)$
	$x(k+1)$	$x(k+2)$	$\cdots$	$x(2k)$
	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	$x\{(m-1)k+1\}$	$x\{(m-1)k+2\}$	$\cdots$	$x(n)$
sum:	$y(1)$	$y(2)$	$\cdots$	$y(k)$
mean:	$\bar{x}(1)$	$\bar{x}(2)$	$\cdots$	$\bar{x}(k)$

By computing

$$A(k) = \frac{2}{n} \sum_{i=1}^{k} y(i) \cos (2\pi/k)i ,$$
$$B(k) = \frac{2}{n} \sum_{i=1}^{k} y(i) \sin (2\pi/k)i , \tag{11}$$
$$R^2(k) = A^2(k) + B^2(k) ,$$

concerning the series of sum $\{y(i)\}$, we get the maximum (not the greatest) values of the graph of $R(k)$ at the points where the values of k coincide with the sought-for periods. This graph of $R(k)$ is called "Schuster's periodogram".

The ratio of the variation of the series $\{y(i)\}$, $\delta_y^2(k)$, to the variation of the original series, σ^2, is sometimes called "Whittaker-Robinson's periodogram". It also has the maxima near the points where the values of k coincide with the sought-for periods.

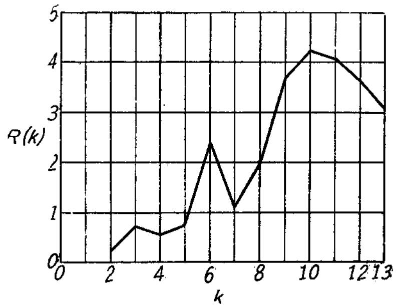

Fig. 11 (a). The periodogram (the River A).

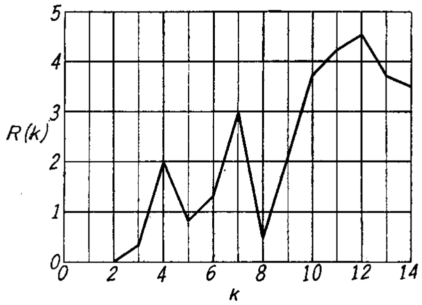

Fig. 11 (b). The periodogram (the River B).

Now, as there is not enough sample size available for inspecting the length of the period in these instances, it is not proper to apply these methods directly. One of the alternatives would be to use, say, the series of mean $\{\bar{x}(i)\}$ instead of the series of sum $\{y(i)\}$.

Figs. 11 (a) and (b) are Schuster's periodograms representing the 3-year moving average for $K=2$ of the total annual average stream flows of the Rivers A and B.

The periodogram is more laborious to construct than the correlogram. In the instances cited in this paper, our purpose has been fulfilled by the correlogram, so we will omit the actual instances of calculation.

IV. The Detection of the Periodic Components

After we have detected the periods of the periodic components contained in the fluctuation curve by the above methods, we can proceed to detect their amplitudes and phases by the two methods that follow,

(1) The Harmonic Analysis

After two periods L and M have been detected, we have generally only to solve, as is well known, the equation

$$\begin{pmatrix} \Sigma\cos^2(\lambda i) & \Sigma\cos(\lambda i)\sin(\lambda i) & \Sigma\cos(\lambda i)\cos(\mu i) & \Sigma\cos(\lambda i)\sin(\mu i) \\ \Sigma\cos(\lambda i)\sin(\lambda i) & \Sigma\sin^2(\lambda i) & \Sigma\sin(\lambda i)\cos(\mu i) & \Sigma\sin(\lambda i)\sin(\mu i) \\ \Sigma\cos(\lambda i)\cos(\mu i) & \Sigma\sin(\lambda i)\cos(\mu i) & \Sigma\cos^2(\mu i) & \Sigma\cos(\mu i)\sin(\mu i) \\ \Sigma\cos(\lambda i)\sin(\mu i) & \Sigma\sin(\lambda i)\sin(\mu i) & \Sigma\cos(\mu i)\sin(\mu i) & \Sigma\sin^2(\mu i) \end{pmatrix} \begin{pmatrix} A_L \\ B_L \\ A_M \\ B_M \end{pmatrix} = \begin{pmatrix} \Sigma x(i)\cos(\lambda i) \\ \Sigma x(i)\sin(\lambda i) \\ \Sigma x(i)\cos(\mu i) \\ \Sigma x(i)\sin(\mu i) \end{pmatrix}, \tag{12}$$

in order to find the periodic components,

$$\begin{aligned} \xi(i) &= \xi_L(i) + \xi_M(i) \\ &= A_L\cos(2\pi i/L) + B_L\sin(2\pi i/L) + A_M\cos(2\pi i/M) + B_M\sin(2\pi i/M), \end{aligned} \tag{13}$$

by the method of least squares, where, in Eq. (12) $\sum$ means $\sum_{i=1}^{n}$, λ and μ being equal to $2\pi/L$ and $2\pi/M$ respectively.

As a special case, Eq. (12) may be somewhat simplified if $\xi_L(i)$ is the fundamental periodic component and $\xi_M(i)$ its higher harmonic. For if the sample size n is a multiple of L, all elements on the principal diagonal of the matrix in Eq. (12) become $n/2$, and all others zero. Thus, we obtain

$$\begin{aligned} A_L &= \frac{2}{n}\sum_{i=1}^{n} x(i)\cos(2\pi i/L), \quad B_L = \frac{2}{n}\sum_{i=1}^{n} x(i)\sin(2\pi i/L), \\ A_M &= \frac{2}{n}\sum_{i=1}^{n} x(i)\cos(2\pi i/M), \quad B_M = \frac{2}{n}\sum_{i=1}^{n} x(i)\sin(2\pi i/M), \end{aligned} \tag{14}$$

and so forth.

This is the same result that we have obtained by the above periodogram.

Therefore, if the sample size n is not a multiple of the fundamental period L, we have only to reduce the size till it becomes a multiple of L, by operating the moving average given in **III. (2)**, in order to facilitate the harmonic analysis by making use of Eq. (14).

The data for twenty-eight years are available for the Rivers A and B, of which the longest fundamental period is 10~11 years. We may, therefore, take it as 11 years (though 10 years gives the same result).

Since the estimation of the so-called trend is difficult due to the smallness of the sample size in this instance, we assume in this paper the period of the trend to be 22 years, which is twice as long as the fundamental period (11 years).

Now $\xi_{22}(i)$ and $\xi_{11}(i)$ can be determined by solving Eq. (12) for $L=22$, $M=11$.

Further, we can decrease the sample size to 22 by operating the 3-year moving average for $K=2$ three time successively, and find out the fundamental periodic component and its arbitrary higher harmonic components by Eq. (14).

Here we must take care to modify these higher harmonic components, for they have been multiplied by $\{C_2(k)\}^3$ times as indicated in Table 2 and Fig. 4 (b) by taking the moving average.

If we prepare, for the River A, the series of 3-year moving average for $K=2$, $\{y_2(i)\}$, calculate $\xi_{22}(i)$ and $\xi_{11}(i)$ by applying to it Eq. (12), then likewise find the 2nd higher harmonic component $\xi_{5.5}(i)$ for the residual series $[y_2(i)-\{\xi_{22}(i)+\xi_{11}(i)\}]$ as shown in Table 5. The modification of these harmonic components by Table 2 or Fig. 4, may bring about a greater accuracy, but these, as they stand, can practically be regarded as the harmonic components of the original series.

Also in Table 5 are given, for reference, the harmonic components as modified by Table 2 and Fig. 4, sought by applying Eq. (14) to the series $y_{2,2,2}(i)$ obtained by repeating three times the 3-year moving average for $K=2$ for the River A. Where, $i=0$ for the year 1926.

Table 5. Harmonic constants (the River A).

Constants / Method of calculation	A_{22}	B_{22}	A_{11}	B_{11}	$A_{5.5}$	$B_{5.5}$
By $y_2(i)$ and Eq. (12)	1.9	−0.4	1.0	−4.0	−2.3	−0.5
By $y_{2,2,2}(i)$ and Eq. (14)	1.6	−1.0	1.7	−5.2	−3.6	0.8

Further analysis as described **III. (1)～(4)** should be repeated for the residual series if, after subtracting from the original series the periodic components thus obtained, the residual series seems to contain still some periodic component. In this paper the residual series, from which all periodic components have been subtracted, are represented by $z(i)$.

In the following investigation, the harmonic components obtained from $y_2(i)$ and Eq. (12) in Table 5 will be adopted.

In this case, as is well known, the original series $x(i)$ can be considered separable as

$$x(i) = m(i) + \xi(i) + z(i)\,, \tag{15}$$

here $m(i)$ is the mean of $x(i)$ which contains a trend, $\{\xi(i)=\xi_{11}(i)+\xi_{5.5}(i)\}$ is the periodic component, and $z(i)$ the component with random fluctuations. ($\xi_{22}(i)$ is regarded as a trend.)

The data $x(i)$ of the annual average stream flows of the Rivers A and B, together with the periodic components obtained by applying the above method, inclusive of $m(i)$, are illustrated in Figs. 12 (a) and (b).

The significance of the harmonic components thus obtained will be tested[3], but is omitted here.

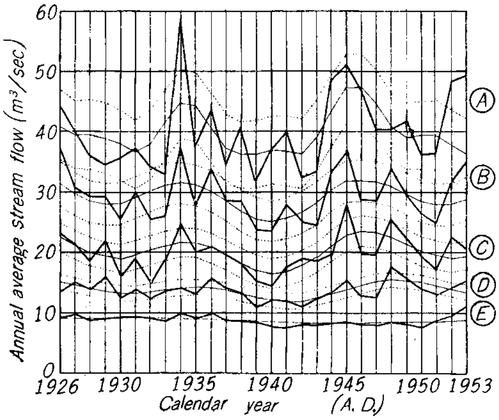

Fig. 12 (a). The fluctuation curves of the annual average stream flows and respective periodic components obtained by harmonic analysis (the River A).

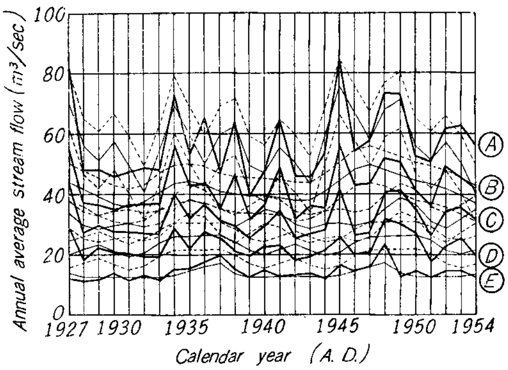

Fig. 12 (b). The fluctuation curves of the annual average stream flows and respective periodic components obtained by harmonic analysis (the River B).

Note: —— Data
—— Periodic components
---- About 68% variation intervals
Ⓐ··· The total annual average stream flow
Ⓑ··· where the 3-month-flow for each year is taken as the maximum available
Ⓒ··· " 6-month-flow " "
Ⓓ··· " 9-month-flow " "
Ⓔ··· " drought-flow " "

(2) The Method regarding the Series of Moving Average of Eq. (5) as the Periodic Component

This is a convenient method which regards the series of moving average of Eq. (5) as the periodic component, by which much the same result may be obtained.

The results we obtain by applying Eq. (5) to the above examples, shown in Figs. 13 (a) and (b), almost agree with that shown in Fig. 12, but no simple conclusion is to be deduced therefrom, since the property of the moving average method is rather complex in spite of its easy operation.

When the above two methods are compared we see that **method (1)**, though theoretically exact, is essentially difficult to apply if the sample size is small, while **method (2)** is not only adapted for the case of small sample size, but

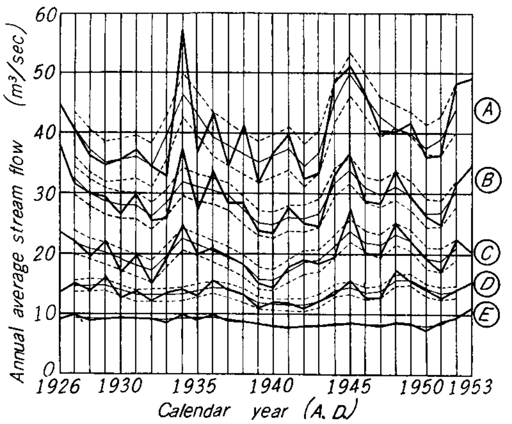

Fig. 13 (a). The fluctuation curves of the annual average stream flows and respective periodic components obtained by the moving average of Eq. (5) (the River A).

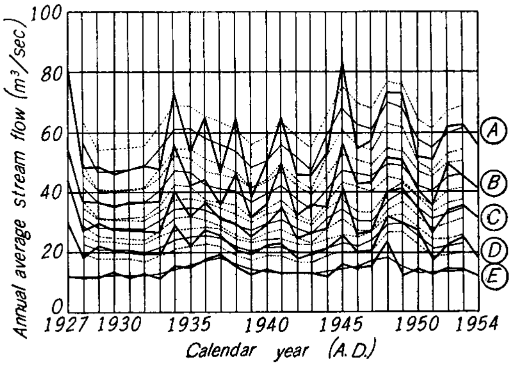

Fig. 13 (b). The fluctuation curves of the annual average stream flows and respective periodic components obtained by the moving average of Eq. (5) (the River B).

Note: —— Data
—— Periodic components
=== About 68% variation intervals
Ⓐ··· The total annual average stream flow
Ⓑ··· where the 3-month-flow for each year is taken as the maximum available
Ⓒ··· " 6-month-flow " "
Ⓓ··· " 9-month-flow " "
Ⓔ··· " drought-flow " "

also is far simpler in respect of calculation, so that it is considered to have more practical utility as a convenient method.

Further, as the skewness of the distribution is supposed to be contained in the periodic component, the asymmetry of distribution can be decreased by subtracting the periodicity, as may easily be inferred from Fig. 14 illustrating the non-exceeding probability of $z(i)$.

V. The Estimation of the Annual Average Stream Flow to hold on an Average for a Long Term

After the existence of the fundamental period, N years, has become manifest by the analysis of the annual average stream flow, we have to take into consideration the periodic components in order to estimate the annual average stream flow expected to hold on an average for a long term to come.

By so doing we are enabled not only to make sufficiently small the fluctuation in the stream flow, whatever starting point we may choose for the years of planning, but also to use the mean of the annual average stream flow for the past N or multiple of N years as the projected value (probable value) of the annual average stream flow for the coming N or multiple of N years.

There are two methods for estimating the expectation of annual average stream flow.

(1) The Method dealing with the Residual Series $\{m(i)+z(i)\}$ obtained by Subtracting from the Annual Average Stream Flow $x(i)$ the Periodic Component $\xi(i)$ whose Mean is Zero

As stated in **IV**, the annual average stream flow $x(i)$ can be divided into three components $m(i)$, $\xi(i)$ and $z(i)$ according to Eq. (15), so that by detecting the periodic component $\xi(i)$ and the d.c. component (containing a trend) $m(i)$, and subtracting these from $x(i)$, we have the random component $z(i)$ left. If these components $z(i)$ of the Rivers A and B are arranged in order of magnitude and their non-exceeding probabilities are plotted according to Thomas' method,

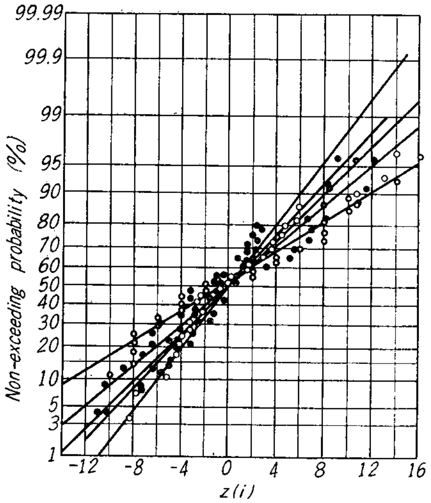

Fig. 14 (a). The probability distribution of the random components $z(i)$ of the annual average stream flows (the River A).

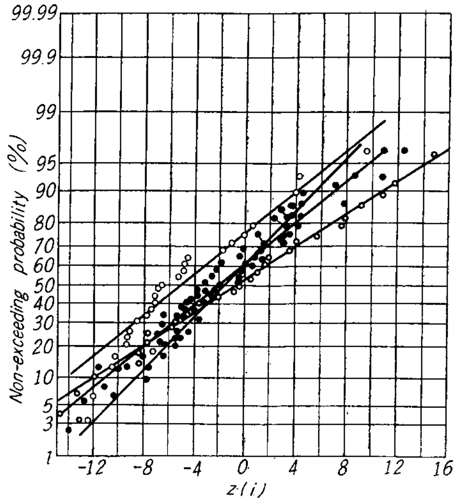

Fig. 14 (b). The probability distribution of the random components $z(i)$ of the annual average stream flows (the River B).

Note: ○··· The total annual average stream flow
●··· where the 3-month-flow for each year is taken as the maximum available
◎··· " 6-month-flow " "
◉··· " 9-month-flow " "
○··· " drought-flow " "

we find that they are distributed nearly as straight lines on normal probability paper, as shown in Figs. 14 (a) and (b). If we remember the meaning of the central limit theorem, we may be safe in concluding that generally the magnitude of $z(i)$ is distributed according to the normal distribution $N(0, \sigma^2)$. It is also considered that the annual average stream flow $x(i)$ is distributed according to the normal distribution $N(0, \sigma^2)$ round the functional component $\{\mu(i)+\xi(i)\}$, because no significant correlation is discerned upon the test between the periodic component $\xi(i)$ and the random (non-functional) component $z(i)$.

If it is possible to look upon $z(i)$ as an independent sample obtained by random sampling from the normal population $N(0, \sigma^2)$, the population mean of $z(i)$ can be estimated. In this case, the sample mean $\bar{z}$ may be adopted as point estimate. Further, we need necessarily not operate the N-year average on $z(i)$.

As there are, however, contained in the stream flow random fluctuation components impossible to predict, it is sometimes more advisable to appoint a certain interval representing the extent of randomness, than to appoint a value as the mean. In such a case, we can determine the $\alpha\%$ confidence interval of the population mean of $z(i)$ by finding t_0 for the $\alpha\%$ level of significance such that, as is well known,

$$P\{t(n-1)\} \equiv Pr\{|t|<t_0\} = \alpha(\%)\,,$$

where n represents the sample size.

Hence the estimates of the population mean $\bar{\mu}_n$ of the annual average stream flow $x(i)$, are given by

$$\text{Point estimate:} \qquad \bar{\mu}_n = \bar{m}+\bar{z}\,,$$

$$\alpha\%\ \text{interval estimate:} \quad \bar{\mu}_n = \left[\bar{m}+\bar{z}+\frac{s}{\sqrt{n-1}}t_0,\ \bar{m}+\bar{z}-\frac{s}{\sqrt{n-1}}t_0\right], \tag{16}$$

where $s^2 = \sum_{i=1}^{n}\{z(i)-\bar{z}\}^2/n$,

$\bar{m}$ is a constant considering the trend (d.c. component).

Table 6 shows the result obtained by applying the method of Eq. (16) to various values of the annual average stream flows of the Rivers A and B.

Table 6. The estimate of the annual average stream flow (figures enclosed in parentheses show the about 68% confidence interval).

Kind of annual average stream flows \ Name of Rivers	A (m^3/sec)	B (m^3/sec)
Total annual average stream flow	40.2 (39.0~41.4)	57.3 (55.8~58.8)
where the 3-month-flow for each year is taken as the maximum available	29.4 (28.7~30.1)	42.2 (40.8~43.6)
" 6-month-flow " " "	20.1 (19.6~20.6)	32.0 (31.0~33.0)
" 9-month-flow " " "	13.7 (13.4~14.0)	23.2 (22.5~23.9)

Here, the figures not enclosed in parentheses show the result of point estimate, those enclosed in parentheses the about 68% confidence interval. In Figs. 15 (a) and (b) are given the mean values and the about 68% confidence intervals of the annual average stream flows. For convenience' sake the $2N$-year periodic function is assumed to be the trend.

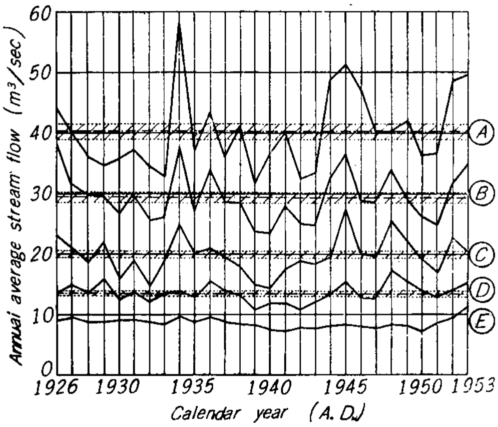

Fig. 15 (a). The expectation of the annual average stream flows (the River A).

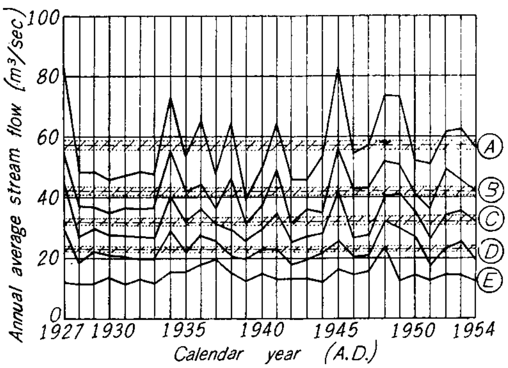

Fig. 15 (b). The expectation of the annual average stream flows (the River B).

Note: —— Data
- • - Point estimates
//// About 68% confidence intervals
Ⓐ··· The total annual average stream flow
Ⓑ··· where the 3-month-flow for each year is taken as the maximum available
Ⓒ··· " 6-month-flow " "
Ⓓ··· " 9-month-flow " "
Ⓔ··· " drought-flow " "

(2) The Method dealing with the N-year Moving Average Stream Flow $y_N(i)$ of the Annual Average Stream Flow $x(i)$

As stated above, the annual average stream flow $x(i)$ is found to be a sample out of the population distributed according to $N(0, \sigma^2)$ round the periodically fluctuating process (functional process) $\{\mu(i)+\xi(i)\}$, so any mean value of stream flows for N years "$x_N(i)$"—abbreviated "N-year average stream flow" in the following—is generally distributed according to $N(0, \sigma^2/N)$ round the N-year moving average of $\{\mu(i)+\xi(i)\}$.

Operating the N-year moving average, however, on $\{\mu(i)+\xi(i)\}$, the N-year periodic component and its higher harmonic components $\xi(i)$ contained in the above-mentioned functional process will be eliminated, leaving only the N-year moving average value $\bar{\mu}_N(i)$ of the population mean $\bar{\mu}(i)$ containing a trend. Therefore $x_N(i)$ may be regarded as a sample out of the population distributed according to $N(0, \sigma^2/N)$ round $\bar{\mu}_N(i)$. If there are other periodic components mixed with the N-year periodic component, they cannot always be removed completely by the N-year moving average, but as a matter of fact they are damped till they are negligible.

If we prepare $y_N(i)$ as a substitute for $x_N(i)$, the functional component to be contained in $y_N(i)$ also becomes $\bar{\mu}_N(i)$, but the N-year moving average of random (non-functional) components cannot be thought of as an independent sample, for some of the neighbouring data overlap each other. It is not therefore permissible, as it stands, to apply the above-mentioned **method (1)** for estimating the population parameter, though it holds good in practice if the sample size n is far larger than N.

In this case, however, the fluctuation in the random components become in general small enough to be approximately regarded as a constant

$$\bar{\mu}_N(i) = \bar{\mu}_N, \quad \bar{m}_N(i) = \bar{m}_N, \tag{17}$$

if we detect and subtract beforehand the trend. Hence it follows that any fluctuation in $y_N(i)$ is entirely caused by its random components.

The mean value, therefore, of the N-year average stream flow—N being usually odd—may be obtained by

$$\bar{x}_N = \frac{1}{n-N+1} \sum_{i=1}^{n-N+1} y_N\left(i + \frac{N-1}{2}\right). \tag{18}$$

Where a trend exists, we can obtain the mean value by subtracting beforehand the trend from $y_N(i)$, calculating Eq. (18) and finally taking the trend into consideration.

The dispersion of the N-year average stream flow can be obtained in the same way as in **method (1)**, but will lack validity unless the sample size n is large enough compared with N. We may notice that in this case we need to take N- (sometimes $2N$- or $3N$- $\cdots$) year average.

By comparing the above two methods we find that **method (1)** is theoretically exact and is better than **method (2)** in such cases where the sample size is small, while **method (2)** can have no less accuracy than **method (1)** if the sample size is large enough so that $n \gg N$.

Moreover, it is seen in our instances that the results obtained by both methods are practically the same, while the latter is far simpler in respect of calculation, so that we can conclude that **method (2)** may be safely adopted as a convenient method for estimating the stream flow for this degree of sample size.

VI. Conclusion

In this paper we have tried to introduce the recent developments of stochastics into the field of the estimation of stream flow for electric power generation, by means of fundamental formulae such as the serial correlation coefficient, the moving average, the periodogram, the detection and separation of periodic component[3),4)]; then by the extensive use of these methods we have estimated the expectation of the annual average stream flow for electric power generation.

In planning power generation we need not take the average of all the available data in the past in order to estimate the stream flow which is expected to hold on an average for a long term to come, but to make statistic considerations by means of a theoretically reasonable analysis.

Since it is difficult, owing to the extreme randomness, to deal directly with the duration curve of stream flow or daily stream flow curve, we have in this paper considered as a parameter the average stream flow for electric power generation for each year, which is suitable to the application of the simple and practical time series analysis. If the duration curve of stream flow or the daily stream flow curve is required on occasion, the annual average stream flow also serves as a basis for it.

After minimizing the influence of the random components by the aid of the moving average, we proceed to examine whether or not there is a fundamental fluctuation period contained in the stream flow by means of the correlogram or periodogram. The correlation diagram is used for a handy detecting method of an abnormal year.

After the fundamental fluctuation period has been detected, one method to obtain an estimation of the annual average stream flow to hold for a long term is as follows; we find the periodic component by means of harmonic analysis or the moving average, subtract from the annual average stream flow the periodic and the d.c. components, thus obtaining a residual component, so as to estimate the population mean, assuming the above-mentioned residual component to be subject to the normal distribution, and finally, we add to the estimate the d.c. component again. Another method is as follows; we operate the moving average for the fundamental period on the annual average stream flows, and then, considering the trend, estimate the mean of the moving average stream flows, looking upon it as the expectation of the annual average stream flow.

References

1) T. Okubo and H. Nishihara: THIS MEMOIRS, **19**, 374 (1957).
2) T. Okubo and H. Nishihara: THIS MEMOIRS, **17**, 139 (1955).
3) M. Ogawara: The Theory of Time Series Analysis and its Application, Applied Statistics, Kokuseido, 1949 (in Japanese).
4) H. Wold: A Study in the Analysis of Stationary Time Series, 2nd ed., Stockholm, 1953.