International Conference on Estuaries and Coasts
November 9-11, 2003, Hangzhou, China

VARIATION IDENTIFICATION OF HIGH TIDE LEVEL IN HANGZHOU BAY

Aiju YOU
Zhejiang Institute of Hydraulic & Estuary, Hangzhou 310020,China
Tel.:+86 0571 86438373, Fax:+86 0571 86960283, E-mail:yajqjs@163.net

Abstract*:* Variation of hydrological regime occurred timely or spatially due to natural hydrologic cycle change and human activities. It is necessary to comprehend and distinguish when or where the variation occurs in hydrologic simulation, forecasting and hydrologic computation in order to adopt feasible measures. This paper adopts differential information theory to identify change points in time series. Key step of variation identification is to built up comparing sequence. Two methods are introduced, one is comparing series formed of moving partition step by step, another is comparing series formed of order clustering optimum bi-partition. The two means are applied in analysis of change point of annual maximum tidal level series at Ganpu station and get the unanimous result.

Key words: Variation, Human activity, Hydrological time series, Differential information theory

1. INTRODUCTION

Variation of hydrological regime occurred timely or spatially due to hydrologic cycle change and human activities. For example, statistical parameters of hydrological sequence like mean value and variance change a lot, which results in the destruction of the consistency of hydrologic sequence. It is necessary to decide when or where the variation occurs in hydrologic simulation, forecasting and hydrologic computation in order to adopt applicable measures.

Comparing sequences used for studying hydrological variation are transformed from original hydrological sequence. How to define comparing sequences is significant when theory of differential information is applied because unsuitable comparing series would lead to incorrect conclusion. It is important to define comparing sequences bases on hydrological background and practical target to be solved.

Hydrological variation identification includes variation identification between elements, variation identification between several sequences and identification variation with time of one hydrological sequence. Two methods for identifying variation with time of hydrological sequence is illustrated as following, taking the annual maximum tide level of GanPu Station as the studying example. The site of Ganpu station is the demarcation of Qiantang River and Hangzhou Bay, located at Haiyan city, Zhejiang Province, see Fig. 1.

2. COMPARING SERIES OF MOVING PARTITION STEP BY STEP

Comparing series of the annual maximum tidal levels of Ganpu station are built up through moving partition step by step of the original sequence, and change point of the original sequence is identified through calculating differential information among variant sequences. Procedures of making comparing series and calculating differential information are as following.

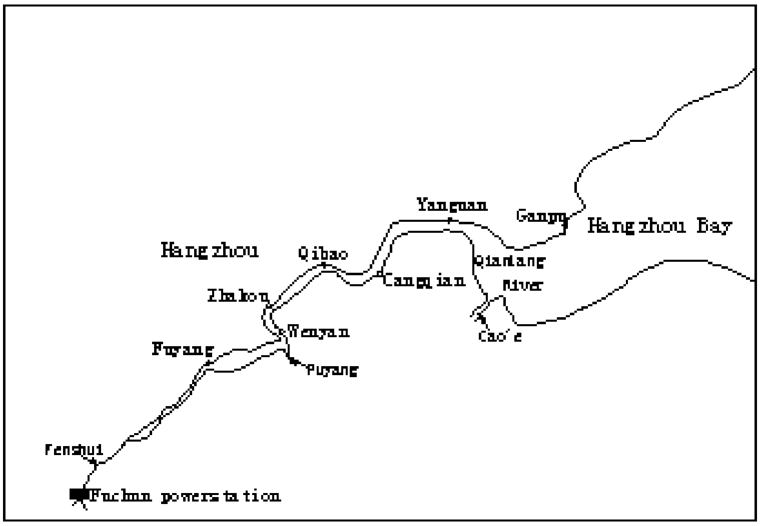

Fig. 1 The Qiantang Estuary

a. Comparing sequences formed step by step moving from point j, the elements ahead of j remain the same as the origin value, but all elements behind j replaced by statistical parameter such as mean value. If the original sequence as:

$$X_0 = [x(1), x(2), ..., x(j-1), x(j), x(j+1), ..., x(n)] \tag{1}$$

The comparing sequence X_j defined as:

$$X_j = [x(1), x(2), \cdots, x(j), \bar{x}(j), \cdots, \bar{x}(j)) ,\ j = 1,2,\ldots\ldots,N \tag{2}$$

Where,

$$\bar{x}(j) = \sum_{k=1}^{j} x(k) / j$$

The comparing sequence set is given by:

$$A_1 = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x(1), \bar{x}(1), \bar{x}(1), \cdots, \bar{x}(1) \\ x(1), x(2), \bar{x}(2), \cdots, \bar{x}(2) \\ \cdots \\ x(1), x(2), x(3), \cdots, x(n) \end{bmatrix} \tag{3}$$

Where n is the length of the original sample.

b. Calculate the differential information entropy $I(X_j)$,differential information measure $I_d(X_j)$ and relative differential information measure $I_a(X_j)$ of each X_j sequence.

$$I(X_j) = -K \sum_{j=1}^{S} y_j \ln y_j \tag{4}$$

Herein,

$$y_j = f(x_j) = (\frac{1}{1+x_j}) / (\sum_{i=1}^{s} \frac{1}{1+x_i}) \tag{5}$$

Maximum differential information entropy $I_{max}(X_j)$ must be calculated before computation differential information measure $I_d(X_j)$ and relative differential information measure $I_a(X_j)$.

$$I_{\max}(X_j) = K \ln(n) \tag{6}$$

And

$$I_d(X_j) = I_{\max} - I(X_j) \tag{7}$$

$$I_a(X_j)=\frac{I_d(X_j)}{I_{d,\max}(X_j)}\times 100\%=\frac{I_{\max}(X_j)-I(X_j)}{I_{\max}(X_j)-I_{\min}(X_j)}\times 100\% \tag{8}$$

The value of I_{min} usually can be obtained through referencing the fixed ratio 0.50−0.95 with the minimum value of calculated sequence $I(X_j)$.

c. Plot the graph $I_a(X_j)$ −j as Fig. 2. From the viewpoint of cluster analysis theory, the points with the same slope can be treated as the same sequence zone, that is to say, the principle of the initial sequence is keep unchanged in this zone.

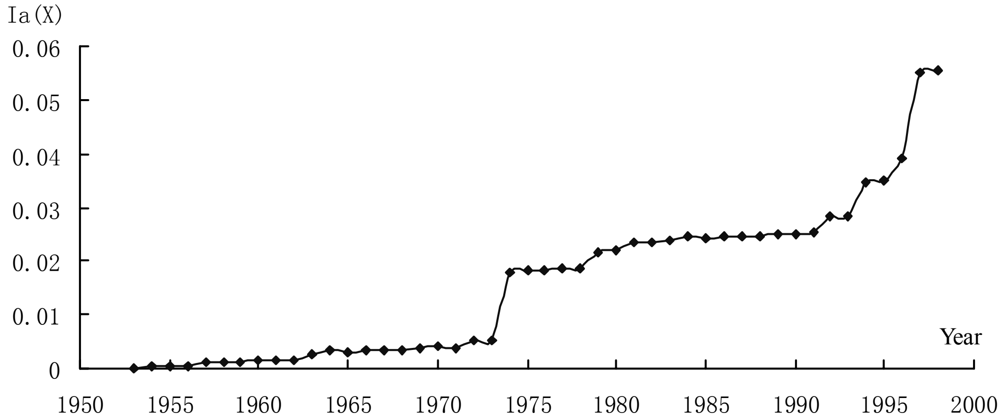

Fig. 2 Relative measure $Ia(X_j)$ of annual maximum tidal level differential informationat Ganpu station

The above method is effective because the difference between X_j and X_{j+1} can be identified from the sequence $I_d(X_j)$ or $I_a(X_j)$. If no systematic variance in sequence X_j, then the elements before j has the same average value as that of the elements after j. Otherwise, if there is systematic variance in sequence X_j, and happens at point j, then the mean value of the first j elements has different average value from the value of total elements. The aggravate of the variability leads to the increase of measure value of $I_d(X_j)$ or $I_a(X_j)$, and its slop coefficient changed obviously while compared with no variance sequence. From Fig. 2, the slope coefficient of $I_a(X_j)$ changes significantly around year 1974 and 1992, so the two years could be judged as the change point.

3. COMPARING SERIES OF ORDER CLUSTERING OPTIMUM BI-PARTITION

Comparing series of order clustering optimum bi-partition is another method to identify time variance or spatial variance of a time series. For a given hydrology time series X=(x(1),x(2),……,x(n)), supposing the change point τ , comparing series of deviation square before and after partition could be defined as:

$$Z_\tau=\left[(x(1)-\overline{x_\tau})^2,(x(2)-\overline{x_\tau})^2,\cdots,(x(\tau)-\overline{x_\tau})^2,(x(\tau+1)-\overline{x}_{n-\tau})^2,\cdots,(x(n)-\overline{x}_{n-\tau})^2\right] \tag{9}$$

Where,

$$\overline{x}_\tau=\frac{1}{\tau}\sum_{i=1}^{\tau}x_i\ ,\ \ \overline{x}_{n-\tau}=\frac{1}{n-\tau}\sum_{i=\tau+1}^{n}x_i \tag{10}$$

To obtain the most possible variant point τ_0 with order classification method means finding the optimum partition point where the square of deviation between the same types is smaller and which between different types is larger. This is a reciprocal expression of sequence difference. For an original series with definite sample capacity n, n groups of hydrological sequence can be got with τ varies from 1 to n.

$$A_2 = \begin{bmatrix} (x_1 - \overline{x}_1)^2, (x_2 - \overline{x}_{n-1})^2, \cdots, (x_\tau - \overline{x}_{n-1})^2, \cdots, (x_n - \overline{x}_{n-1}) \\ (x_1 - \overline{x}_2)^2, (x_1 - \overline{x}_2)^2, \cdots, (x_\tau - \overline{x}_{n-2})^2, \cdots, (x_n - \overline{x}_{n-2})^2 \\ \vdots \\ (x_1 - \overline{x}_n)^2, (x_1 - \overline{x}_n)^2, \cdots, (x_\tau - \overline{x}_n)^2, \cdots, (x_n - \overline{x}_n)^2 \end{bmatrix} \tag{11}$$

Calculate the information entropy relative measure I_e and differential information relative measure I_a for each X_τ. The point τ where I_e is maximum or I_a is minimum is just the variant point makes the original sequence divided into two clustering. This can be recognized and identificationd with differential information measure I_a directly.

The differential information measure $I_a(X_\tau)$ indicates the difference among comparing sequences X_τ which means the global difference of a hydrological sequence after partition transforming, where $I_a(X_n)$ shows the degree of global difference of the original sequence with no partition.

Supposing

$$C(Z_\tau) = I_a(Z_\tau) / I_a(Z_n) \tag{12}$$

$$B(\tau) = C(Z_\tau) / (\overline{C} - 1) \tag{13}$$

Where,

$$\overline{C} = \sum_{\tau=1}^{n} C(Z_\tau) \tag{14}$$

$B(\tau)$ is the range of differential information, denotes the degree of increased or decreased systematic difference of hydrological series due to the "classification" of hydrological series system after partition at the assumed variant point. Plot the points $B(\tau) \sim \tau$ as Fig. 3, where we can identification the difference of sequence system explicitly.

Figure 3 shows that there are variant points in the sequence, the year of variation located at the trough of the graph, and the deeper is the trough, the larger is the degree of variation. Because if point τ is the change point, then the division at point τ will decrease the degree of the difference increment of the original hydrological sequence which result from the existed variant phase. Judge from Fig. 3, there is an obvious variant point in 1973 and another change phase is between 1992 and 1995.

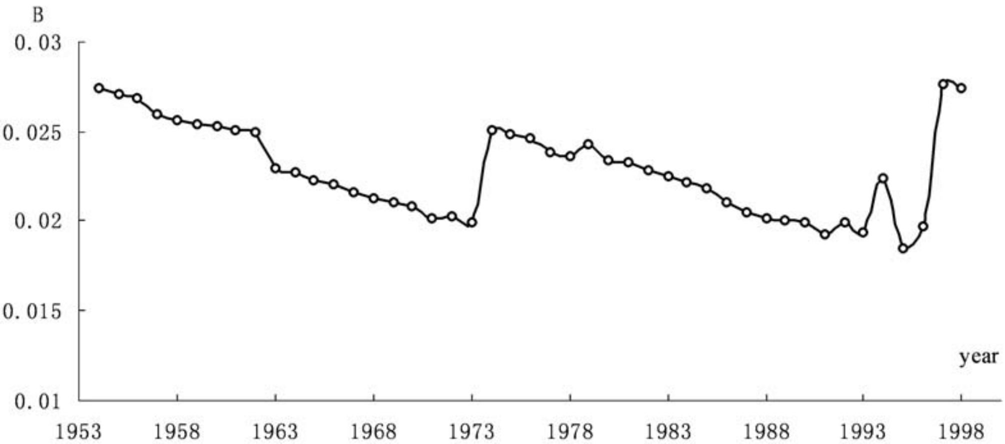

Fig. 3 Annual maximum tide level differential information amplitude B at GanPu station

Whether the variation of tide level series at Ganpu station is caused by runoff or not can be determined by contrast it with the change point of runoff. Fig. 4 shows $B(\tau)-\tau$ of runoff at FuChunJiang Power Station, and the variant point appears at year 1954, but there is no obvious variant point near year 1973 and 1993. This truth illustrates clearly that the variation of GanPu high tide level dose not caused by the variation of runoff.

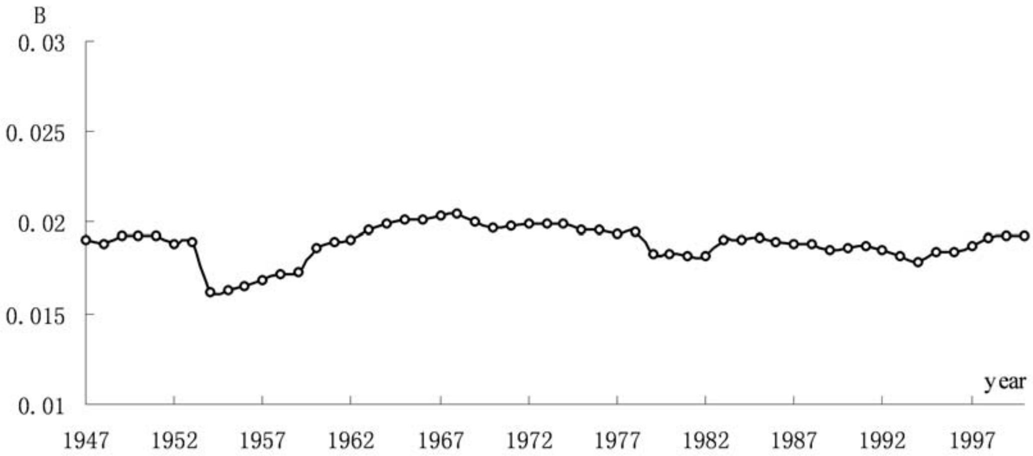

Fig. 4 Discharge differential information amplitude B at FuChunJiang Power Plant

4. CONCLUSION

As mentioned before, because of hydrologic cycle change and human activities, variation of hydrological regime occurred in time or in space frequently. This paper presents practical and simple approaches to identify variant point of hydrological time series with differential information theory. The principal step to identify change point is to define feasible comparing series and two means are introduced in this essay, comparing series of moving partition step by step and comparing series of order clustering optimum bi-partition. The two means are applied in analysis change point of annual maximum tidal level series at Ganpu station and the conclusion is that two changes point in the series, around year 1974 and 1993. At the same time, it is proved that upstream runoff is not the cause of variance of tidal level because the change points of the two series do not occur at the same year. The methods demonstrated in this paper are very simple and effective.

REFERENCES

Dooge J C I. Linear theory of hydrological system. Tech Bull, 1468. Agr Res Serv US Dep of Agr, Washinton, D C, 1973.

Xia Jun, Huang G H, Brad B. Combination of differented prediction approach and interval analysis for the prediction of weather variables under uncertainty. Journal environmental management, 1997(1), 95−106

Xia Jun. A grey clustering approach applied to environmental impact prediction to wild animals, IAHS Publication, No 230 1995.227−236.

Hu Baoqing, Xia Jun. Interval-valued fuzzy number and its operations. Proc. of the international Conference on System control information: methodologies & applications, SCI'94, Huangzhong University of Science & Technology Press, 1994.284−28.

Xia Jun, Hu Baoqing. A grey system model for predicting trend change of urban wastewater load. Journal environment hydrology,1997(5):1−6