Precursory Strong Signal of Abnormal Flood in Xi River and Characteristic Diagnosis of Extreme Value of Flood Water Level

Duan Peng1*, Yu Zhang2, Chunlin Xu3, Huadi Zhou1 and Qianmin Liang1

1 Zhaoqing Meteorological Bureau, Zhaoqing, Guangdong, 526060, China
2 Guangdong climate center, Guangzhou, 510080, China
3 Sihui Meteorological Bureau, Sihui, Guangdong, 526200, China
*Email: pend_073@163.com

Abstract. The article uses the signal-to-noise ratio (SNR) method to identify the precursor strong signal of the atmospheric circulation in the early stage of the abnormal flood in the Xi River. And by using the equation derived from the Cross theory, the extreme frequency, duration and interval time of the flood water level of each stage of the Xi River are calculated. This study will help improve the accuracy of mid-and long-term prediction of abnormal floods in the Xi River Basin and provide a basis for government departments to make better water resources planning and flood prevention.

1. Introduction

The West River has a long history and is a major tributary of the Pearl River. It runs from west to east, originating in Zhanyi County in the east of Yunnan Province, flowing through Guizhou, Guangxi, and into the South China Sea. The total length is of the Xi River is about 2,075km, and the drainage area is 353,100 square kilometers. The geographical location and natural environment of the area where the Xi River Basin is located make the upper reaches of the Xi River a flood-prone area, prone to large floods and even catastrophic floods. The abnormal flood disaster not only poses great threats to the production and life of the entire Xi River Basin, but also has great impacts on the economic development and people’s livelihood in the Pearl River Delta downstream. Many scholars have conducted relevant researches on the relationship between river basin floods and atmospheric circulation and the temporal and spatial distribution characteristics of flood-causing rainstorms [1-6], and the strong signals of the atmospheric circulation in the early stage of abnormal flooding and the extreme flood water level characteristics have been conducted. However, there are very few studies on the strong signal of the atmospheric circulation in the early period of abnormal flooding and the characteristics of extreme flood water level. The research results of this work aim to provide a scientific and objective basis for the long-term planning and construction of government departments and has application values.

Crossover theory is a powerful tool for studying time series, and it’s suitable for describing the changing pattern of extreme values. Since Rice [7] made the pioneering work on intersection theory, the intersection theory of time series has been developed and continuously improved in the stochastic process theory, and it has been applied to study the regularity of sequences in hydrology [8-10].

As climate series often contain vibration signals of various frequencies, their contributions to the extreme value characteristics are different from each other. The crossover theory can study various
characteristic quantities such as the frequency of extreme values, the waiting time, the duration of the extreme value processes, and the law of change. The literature [11] discussed preliminarily the transition law of weather and climate state under the normal crossover theory. Ding Yuguo [11] et al. used this method to estimate the normal stationary sequence Southern Oscillation Index to explore the number, duration, and interval of strong El Niño, and at the same time analyzed the characteristics of summer floods in Nanjing.

This paper used the signal-to-noise ratio method to identify the strong signals of atmospheric circulation in the early stage of abnormal flooding. The highest water level data of the Wuzhou Station of the past 120 years was selected as the investigation sequence, used the equation derived from the crossover theory to calculate the number of occurrences, duration, and interval of extreme values of all levels of big floods of the Xi River. Results were compared with the measured time sequence for error analysis to assess its reliability.

2. Material and Method

2.1. Geographic range and basic data processing
Select average precipitation data of June-July, 1954-2019, of 10 stations in Xi River Basin (Mengzi, Chengjiang, Xingren, Rongjiang, Liuzhou, Guilin, Baise, Nanning, Wuzhou, Gaoyao) and the annual highest water level data of 1933-2019 in the downstream Gaoyao.

The 2.5°×2.5° 500 hPa monthly average altitude field and monthly average sea temperature field reanalyzed by NCEP. Calculations show that during the period of 1954 to 2019, the water level at the Gaoyao station of the lower reach of the Xi River was significantly positively correlated with the June to July precipitation in the Xi River Basin, with a correlation coefficient of 0.5626, which passed the 0.01 reliability test. It is reasonable to use the water level sequence at Gaoyao Station having a long-period of data record to analyze the drought and flood in the Xi River Basin.

Select the annual maximum water level data of Wuzhou Station 1900-2019 as the time series to study the characteristics of the extreme values of the flood water level of the Xi River.

2.2. Method of identifying strong signals [12]
Using the commonly used signal-to-noise ratio to detect signal strength, discuss the precursor signals that may appear at the height of 500 hPa in the northern hemisphere in the early stage of abnormal flooding in the Xi River Basin. The signal-to-noise ratio at a certain grid point height is defined as:

$$R_i = \frac{(X_{ij} - \bar{X}_i)}{S_i}$$  \hspace{1cm} (1)

Among them, $X_{ij}$ $(i=1,2...M, j=1,2...N)$ is the standardized height sequence of the grid point in a certain year, $M$ is the number of grid points, $N$ is the length of the data, $\bar{X}_i$ is the grid point sequence of multi-year average, $S_i$ is the standard deviation of the grid point sequence.

The meaning of signal-to-noise ratio is to use the difference between the height value at a certain grid point and the multi-year average at the same grid point to reflect the climate signal, and the standard deviation of the multi-year series of the grid point is used to reflect the climate noise. Their ratio is called the signal-to-noise ratio, referred to as signal for short. Its value reflects the day-to-day climate anomaly of the altitude field [12] (here is month to month). The signal value of each grid point constitutes the climate anomaly signal field, referred to as the signal field [12]. Assuming that $X_{ij}$ obeys a normal distribution, when $|R_i| > 1.645$, it reveals that $R_i$ passes the significance level test of $\alpha=0.10$; when $|R_i| > 1.96$, it shows that $R_i$ passes the significance level test of $\alpha=0.05$. When analyzing the signal field, we approximate the range of $|R_i| > 1.50$ as the significant signal area, and the range of $|R_i| > 2.00$ as the strong signal area. According to literature [12], in general, the positive
saliency area corresponds to the ridge area of the 500 hPa height field. Contrarily, the negative saliency signal area corresponds to the groove area of the 500 hPa height field.

2.3. Diagnosis and calculation method of extreme value feature for normal stationary process

Suppose $X(t)$ is a zero-mean normal-stationary continuous parameter process. According to the differentiability of the stationary normal mean square of the second-order moment process, Rice has derived the famous crossover expectation equation:

$$ N(t) = \frac{\sigma_x}{\pi \sigma^2} \exp \left( -\frac{u^2}{2\sigma^2} \right) $$

(2)

In the equation, $N(u)$ is the average number of crossings of $X(t)$ with the horizontal line $X = u$ per unit time, $\sigma_x$ is the standard deviation of $X(t)$, $\sigma_x$ is the standard deviation of the first derivative of $\sigma_x$. According to the second-order moment process theory, there should be $\sigma^2_x = -R''(0)$ and $\sigma^2_x = R(0)$, where $R(t)$ is the autocovariance function of $X(t)$. $R''(0)$ is its second derivative when $\tau = 0$. Apparently, there should be $\rho^*(0) = R''(0) / R(0)$, where $\rho^*(0)$ is the second derivative of the autocorrelation function $\rho(\tau)$ of $X(t)$ at $\tau = 0$. Therefore, equation (1) can be converted to

$$ N(u) = \frac{1}{\pi} \sqrt{-\rho^*(0)} \exp \left( -\frac{u^2}{2\sigma^2_x} \right). $$

(3)

Now take $u$ as different critical values to derive the equation for the average number of crossings corresponding to a certain critical value. For example, for $u = \sigma_x$, then equation (2) can be converted to

$$ N(\sigma_x) = \frac{1}{\pi} \sqrt{-\rho^*(0)} \exp \left( -\frac{1}{2} \right) $$

(4)

In the same way, when $u = 2\sigma_x$ and $u = 3\sigma_x$, there is

$$ N(2\sigma_x) = \frac{1}{\pi} \sqrt{-\rho^*(0)} \exp (-2) $$

(5)

$$ N(3\sigma_x) = \frac{1}{\pi} \sqrt{-\rho^*(0)} \exp (-2) $$

(6)

In terms of weather and climate sequence, the anomaly value exceeding $\sigma_x$, $2\sigma_x$, and $3\sigma_x$ from its average means the occurrence of small probability events. Considering that the extreme values in weather and climate sequence are generally not continuous, for convenience, this article stipulates that when the points that exceed a certain horizontal axis are in a series, which was used as a maximum value process, and it is recorded as "a maximum value appears". Therefore, in the case of a large value, every two crossing points can be assumed to be a "one maximum value" process. The average maximum value frequency per unit time is recorded as $\mu(u)$, which is

$$ \mu(u) = \frac{N(u)}{2} = \frac{1}{2\pi} \sqrt{\rho^*(0)} \exp \left( -\frac{u^2}{2\sigma^2_x} \right) $$

(7)

This article mainly discusses the characteristics of extremely large values. As for extremely small values, similar methods can be used to derive them; no further discuss will be provided here. To further describe the "time span" of the extreme value, the duration of a certain maximum value process exceeding the value is defined as, then its "average maximum value duration" can be recorded as, according to the literature \[9\], the assumption is Down, there are
Here, symbol $L_u$ represents the duration of the extreme value process (exceeding the critical value), and the + sign is the maximum value process.

Similarly, let $B_u^+$ denote the value of two consecutive rises of $X(t)$ that exceeds the time interval of $u$, then define the "average time interval of maximum value" as $E(B_u^+)$, which can be proved as

$$E(B_u^+) = \frac{1}{\mu(u)}$$

Apparently, if $X(t)$ is set as a strictly stationary normal sequence, then when is at the critical extreme value, $E(L_u^+)$ and $E(B_u^+)$ can be regarded as the average durations of exceeding the maximum value and the time interval between two adjacent maximum values.

The normality test is performed on the sequence, and the calculated absolute values of the skewness coefficient and kurtosis coefficient are all relatively small (0.08 and 0.228, respectively), and both accept the assumptions, so the Wuzhou water level sequence can be considered as following a normal distribution.

The stationarity of the series is tested by the reverse order method. The results also show that the stationarity hypothesis is accepted under the significant level of reliability $\alpha=0.05$.

The annual maximum water level sequence in Wuzhou can be regarded as a normal stationary sequence.

### 3. Strong Early Signal of Summer Abnormal Drought and Flood in Xi River Basin

If the precursor signals of abnormal floods or droughts can be identified from the changes of previous circulation and sea temperature field, it can provide a basis for the prediction of droughts and floods.

Due to the inter-decadal differences in atmospheric circulation, we analyzed the abnormal flood year 1994 for the past 25 years and the severe drought year of 2003 which is second only to the 1963 severe drought. Use equation (1) to calculate and plot the 500 hPa height signal field in May of the two years, as shown in Figure 1. From this, we can find the strong signal of the 500 hPa height field during the first month of the abnormal drought and flood in the summer of the Xi River Basin. Figure 1a showed that in the early stage of the abnormal flooding in the Xi River Basin, the signal field at the height of 500 hPa, from high latitude to low latitude. It showed a "negative, positive, and negative" distribution from west to east. There is a negative signal area from the polar regions to the European Black Sea. From the southern side of the Ural Mountains to the Balks Lake extending to the southeast of China is a positive signal area; the northeast of China and area extending to the Sea of Japan are negative signal areas; the east side of the Sea of Japan is a weak positive signal area. In the signal field, East Asia presents a "positive west and negative east" distribution, which is "high west and low east" distribution in the 500hPa altitude field synoptic analysis. The strongest significant signal is located at the Qinghai-Tibet Plateau in China. From the east to the southeast side of the ridge, the R value reached 1.5, indicating that the high pressure ridge in the area on the altitude field of the month has a very abnormal performance than usual, and the height has an abnormal increase. The negative value area from northeast China to the Sea of Japan is 1.5, indicating that the height of the regional trough area is abnormally reduced. At the same time, the negative signal area, from Moscow to the Black Sea, did not exceed the significance level of 0.10.
Figure 1. The signal field of 500hPa height in the month before the abnormal flood (a) and drought (b) in the Xi River Basin.

Figure 1b showed that in the first month before the abnormal drought occurred in the Xi River Basin, the signal field at the height of 500 hPa was contrary to the anomaly of flood. From high latitudes to low latitudes, from west to east, there was a "positive, negative, and positive" distribution. The area near the Black Sea in southern Europe is a positive signal area, from the Central Siberian Plateau through the Mongolian Plateau to the Qinghai-Tibet Plateau is a negative signal area, where the northeast China to Japan region is a positive signal area, and its positive signal area extends to the north Pacific region. In East Asia, the distribution is “negative in the west and positive in the east”, the distribution of the 500hPa height field is “low in the west and high in the east”, which is just the opposite of the abnormal flood year. Negative strong signals exceeding the significance level of 0.05 are located in area from the south side of the Qinghai-Tibet Plateau to the Bay of Bengal, and the R value reaches -2.0, indicating that the troughs of this area on the monthly altitude field have very abnormal performances compared with normal years, and the altitude exhibits abnormal decrease. The R value of the positive signal center from the northeastern part of China to the Sea of Japan is +2.0, indicating that the high-pressure ridge in the altitude field in this area has been abnormally enhanced. The R value of the positive strong signal near the Black Sea in southern Europe (40-50°N, 30-40°E) reaches +2.5, indicating that the high-pressure ridge region in the height field of this region is abnormally enhanced. The area near Ural Mountains and the North Pacific (30-50°N latitude, 165-180°E longitude) is a significant positive signal area, exceeding the significance level of 0.10, indicating that the area is a large-scale ridge area.

In the early period of abnormal floods and droughts in the Xi River Basin, the atmospheric circulation in East Asia was opposite. Flood years were "low in the east and high in the west", and drought years were "high in the east and low in the west". In other words, when the East Asia trough was active and strengthened in May, the high-pressure system on the east and southeast sides of the Qinghai-Tibet Plateau continued to strengthen, which was a precursor to flooding in the Xi River basin from June to July. On the contrary, when the East Asian trough weakened in May, and there are high-pressure ridges developing, and the south and east sides of the Qinghai-Tibet Plateau is controlled by a low-pressure system, which is likely to cause drought in the Xi River Basin in the summer. The abnormal atmospheric circulation in May led to abnormal droughts and floods in the Xi River basin from June to July, and the mechanism of its impact needs further study.

Only the precursor signals that occurred in May of the year of abnormal drought and flood in the Xi River Basin are analyzed. In fact, the abnormal performances of the circulation and sea temperature field in the previous few months is related to the abnormal drought and flood in the Xi River Basin. This part of the work needs to be further explored in the future.

4. Characteristics of extreme value of graded flood level

According to equations (7), (8), (9), the maximum value exceeding the critical value \( u \) per unit time, its average frequency \( \mu(u) \), average duration \( E(L_u^-) \), and average time interval \( E(B^+_u) \) can be calculated. According to the actual situation, the critical value \( u \) may be set as \( \sigma_x, 1.5\sigma_x, 2\sigma_x \), and
3\sigma_x$, and $\rho^*(0)$ can be calculated using the following approximate equation, namely

$$\rho^*(0) = 2\rho(1) - 2$$

(10)

Where $\rho(1)$ is the first-order autocorrelation coefficient, which can be estimated from samples. The above calculation results are shown in Table 1, where $T_\mu(u)$ represents the mathematical expectation of the number of occurrences of the maximum value exceeding $u$ in the period T, where T is set as 120 years. It can be seen from Table 1 that in the period of 120 years, the average number of maximum occurrences of the annual maximum water level exceeding $\sigma_x$, $1.5\sigma_x$, $2\sigma_x$, and $3\sigma_x$, are 20.330, 10.221, 4.548, and 0.375, respectively. Since the sequence is assumed to be normally stationary, the number of occurrences of the minimum values less than $-\sigma_x$, $-1.5\sigma_x$, $-2\sigma_x$, and $-3\sigma_x$ in the 120 years assumed the above values. Figure 1 showed the actual measurement sequence. The numbers of occurrence of extreme values exceeding $\sigma_x$, $1.5\sigma_x$, $2\sigma_x$, and $3\sigma_x$ in the sequence are 17, 6, 3, and 0, respectively, and the actual error is +2.9%, +3.7%, and 1.0%, all errors are less than 5%. Therefore, the theoretically derived calculation results are reliable. According to the calculation results, some other characteristics of the flood level of the Xi River need further calculated.

Table 1. Results of characteristic calculation of the extreme value of the Xi River water level.

| Calculation | $\mu(u)$ | $T_\mu(u)$ | $E(L'_u)$ | $E(B'_u)$ |
|-------------|----------|------------|-----------|-----------|
| $\sigma_x$  | 0.1905   | 20.330     | 1.034     | 5.25      |
| $1.5\sigma_x$ | 0.0955   | 10.221     | 0.00      | 10.47     |
| $2\sigma_x$ | 0.0425   | 4.548      | 0.00      | 23.53     |
| $3\sigma_x$ | 0.0035   | 0.375      | 0.00      | 286.68    |

According to the regulations of the World Meteorological Organization (WMO) and some countries, events with sequence anomalies reaching twice the standard deviation (2\sigma) are called anomalies. Events with anomalies exceeding 1.3\sigma are called serious events. According to the actual situation, 25 meters is defined as the warning water level, so we take 25 as the critical value when it reaches 25 (warning water level) = $1.5|\sigma|$. Calculate the standard deviation of Wuzhou annual flood water level anomalies of individual years at $|\sigma|$ equals to 2.98, 1.5$|\sigma|$ equals to 25.00 (warning water level), and 2$|\sigma|$ equals to 5.97. We define the flood classification of the Xi River as shown in Table 2, where H is the highest annual water level value.

Table 2. Definition of Xi River flood classification.

| Classification | Water Level (m) | Definition     |
|----------------|-----------------|----------------|
| $|\sigma|<H<1.5|\sigma|$ | 23.38$\leq$H$<25.00$ | Major Flood    |
| 1.5$|\sigma|<H<2|\sigma|$ | 25.00$\leq$H$<26.37$ | Super Flood    |
| $H$ $\geq$2$|\sigma|$ | H$\geq$26.37 | Extreme Flood  |

The calculation results show that (Table 1 and Table 2), there are roughly 1.9 major floods every 10 years. Among them, there are 0.9 super floods and 1.0 major floods. The duration of a major flood can last 1.034 years, and a major flood can occur once every five to six years. While the interval between super floods occurs every 10 to 11 years, and extreme floods occur every 23 to 24 years.

5. Conclusions

The month before the abnormal flood and drought in Xi River Basin, the distribution of strong signals shown by the atmospheric circulation was basically the opposite. At the same time, the abnormal atmospheric circulation in East Asia, especially on the east and south side of the Qinghai-Tibet Plateau, the abnormal height changes from northeast of China to the Sea of Japan, is a strong precursor to the
occurrence of abnormal droughts and floods in the Xi River Basin. They have opposite distribution characteristics: flood years are low in the east and high in the west, and drought years are high in the east and low in the west.

Based on statistical theory, this paper defines the graded flood of the Xi River and calculates the characteristics of extreme flood value at all levels. The calculations show that roughly 1.9 major floods occur every 10 years. Among them, there are 0.9 super floods and 1.0 major floods. The duration of a major flood can last for 1.034 years, and super flood occurs once every five to six years. The time interval between two super floods is 10 to 11 years, and extreme flood occurs every 23 to 24 years.

Since the forecast of extreme weather and climate events is the most difficult and crucial, theoretical derivation and example calculations show that using the cross theory method to diagnose the characteristics of extreme values of the Xi River flood level has good applicability and reliability. The error is less than 5%, hence it has certain application value for the long-term forecast of flood level of the Xi River.

Under the assumption of normality, there is a certain corresponding relationship between the number of occurrences, duration, interval and other statistical characteristics of extreme values derived in this paper and the various periodic vibration frequencies implicit in the sequence, which need to be further explored and evaluated.

Acknowledgment
This work is financially supported by Guangdong Natural Science Foundation - Major Basic Research and Fostering Project (2015A030308014).

Reference
[1] Yang Hui and Li Chongyin 2003 Adv Atmos Sci 20(4) 540-553
[2] Zhong Lihua, Li Yong, Li Zhongyi, et al. 2018 Torrential Rain and Disasters 37(1) 24-31
[3] Zou X K and Ren F M 2015 Advances in Atmospheric Sciences 32(5) 704-714
[4] Chen Hong and Zhao Si xiong 2000 Chinese Journal of Atmospheric Sciences (in Chinese) 24 (2) 238 -252
[5] Xu Ming, Zhao Yuchun, Wang Xiaofang, et al. 2016 Torrential Rain and Disasters 35(2) 109-118
[6] Huang Haihong, Lin Kaiping, Gao Anyu, et al.2012 Technology and method of weather forecast in Guangxi (Bei Jing: Meteorological Press) p 32
[7] Rice, s.o.1945 Mathematical analysis of random noise (Bell.Sys.Tech.J.) p 24–156
[8] Desmond, A.F.and B.t.Guy 1991 Water Resources Research 27 2791–2797
[9] Yi Li and Barbara J. Lence. 2008 Journal of Hydrologic Engineering 13(9) 776–780
[10] Yi Li and Barbara J. Lence. 2007 Water Resources Research 43 7
[11] Ding Yuguo, Jin Lianji and LiuJingmiao 2002 Chinese Journal of Atmospheric Sciences, 26(3) 343–351
[12] Huang Jiayou, Yang Yang and Zhou Guoliang 2002 Chinese Journal of Atmospheric Sciences, 26(2) 222