

Article

Data Stream Approach for Exploration of Droughts and Floods Driving Forces in the Dongting Lake Wetland

Yeqing Zhai ^{1,2} , Jie Liang ^{1,2,*} , Zhenyu An ³, Xin Li ^{1,2}, Ziqian Zhu ^{1,2}, Wanting Wang ^{1,2}, Yuru Yi ^{1,2} and Suhang Yang ^{1,2}

¹ College of Environmental Science and Engineering, Hunan University, Changsha 410082, China

² Key Laboratory of Environmental Biology and Pollution Control, Hunan University, Ministry of Education, Changsha 410082, China

³ Hunan Water Resources and Hydropower Survey, Design, Planning and Research Co., Ltd., Changsha 410007, China

* Correspondence: liangjie@hnu.edu.cn

Abstract: Wetlands are important environmental resources that are vulnerable to droughts and floods. Studying drought-flood events and their driving factors is essential for wetland resource planning and management. However, climate change and human activities present dynamic challenges that traditional approaches are unable to simulate dynamically in a rapidly changing environment. This makes quantitative analysis difficult. Our research focused on the innovative use of the data stream model, namely online bagging of Hoeffding adaptive trees, to quantify drought and flood drivers in response to climate change and human activity. The proposed approach was applied to a river-lake system, the Dongting Lake wetland. The frequency and duration characteristics of drought-flood events were analyzed. In addition, the cyclical changes of droughts and floods were analyzed by wavelet analysis. Then, drought-flood indicators as well as climatic and hydrological factors were entered into a dynamic data stream model for quantitative calculations. The results showed that the water conservancy projects largely reduced flood events while aggravating droughts. The frequency of floods decreased by 4.91% and the frequency of droughts increased by 6.81% following the construction of the Gezhouba Hydro-project and the Three Gorges Dam. Precipitation and Sankou streamflow were two dominant factors in the Dongting Lake drought and flood events, both of which had a feature importance value of approximately 0.3. This research showed how the data stream model can be used in a changing environment and the applicability of the conclusions reached through real-world instances. Moreover, these quantitative outputs can help in the sustainable utilization of Dongting Lake wetland resources.

Keywords: drought-flood; data stream; driving forces; standardized precipitation index; standardized streamflow index; human influence; wetland



Citation: Zhai, Y.; Liang, J.; An, Z.; Li, X.; Zhu, Z.; Wang, W.; Yi, Y.; Yang, S. Data Stream Approach for Exploration of Droughts and Floods Driving Forces in the Dongting Lake Wetland. *Sustainability* **2022**, *14*, 16778. <https://doi.org/10.3390/su142416778>

Academic Editor: Andrzej Walega

Received: 11 October 2022

Accepted: 11 December 2022

Published: 14 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Floods and droughts are extreme natural hydrological disasters, referring to the persistent water shortage or excess phenomenon caused by water imbalance between supply and demand [1]. Over the past century, under the influence of environmental changes (climate change and human activities), the frequency of major floods and extreme droughts has tended to increase, which seriously threatens the ecological environment and restricts the economic development of society [2,3]. Therefore, there is an urgent need to analyze disaster drivers quickly and precisely in a changing environment to reduce losses. Furthermore, the occurrence of drought and flood events is complicated due to the interaction of numerous factors, such as meteorological, hydrological, and human influences (especially hydraulic projects such as reservoirs and dams) [4,5]. As a result, determining the reliable causes of droughts and floods is critical for assessing disaster risk and improving policy management.

Numerous efforts have been made to develop objective indicators for drought-flood quantification, monitoring, and analysis, including the standardized precipitation index (SPI) [6], the standardized precipitation evapotranspiration index (SPEI) [7], the Palmer drought severity index (PDSI) [8], streamflow drought index (SDI) [9], and standardized streamflow index (SSI) [10]. Nowadays, the standardized precipitation index (SPI) endorsed by the World Meteorological Organization (WMO), is one of the most widely used indicators [11]. Compared with other indices, the SPI calculates the probability of precipitation occurrence and can identify abnormal dry season and rainy season more accurately than others [12]. Given the long-time scale of our study and the fact that the drought and flood events identified by the indicators would be verified against historically realistic information, we chose the SPI/SSI for more accurate conclusions. With the standardized process, just a single precipitation input parameter is required to characterize drought across numerous time scales, such as months, seasons, and years, making it simple to compute and assess the wet and dry periods between adjacent seasons in various climatic zones [6,13]. SPI remains widely used in combination with remote sensing [14], hydrological models [15], or machine learning [16] methods to monitor and forecast drought and flood risks all over the world due to its robustness and convenience. Similar to the principle of the SPI, the standardized streamflow index (SSI) has the apparent benefit of fewer data required, convenient calculation, and the ability to calculate under various time scales [7,17]. It can be utilized for spatial and temporal comparison of various river conditions and streamflow characteristics.

So far, there is still no clear answer as to which approach would best quantify the drivers and identify the major influences. Since the impact of human activities on the environment is extensive and difficult to quantify, there are generally two approaches to studying the impact of human activities on drought and flood. The first compares drought and flood characteristics of the same region before and after the completion of the engineering in hydraulics, or of the upstream basin not disturbed by the reservoir and the downstream basin disturbed [17–19]. The second employs hydrological models to forecast and simulate reservoir operations, either in conjunction with several scenarios to quantitatively assess the impact of river engineering [15,20,21]. The hydrological model has been widely used and well-received in recent years. However, the hydrological model simulation process is subject to extra physical mechanisms derived from a priori understanding of the system's function, implying that the model contains parameter and structural errors [22,23]. Furthermore, the most general challenge is that hydrological simulations supporting scenario analysis with uncertainty estimates are typically too time-consuming and computationally intensive [24].

To overcome the computational barriers in real-time forecasting and scenario analysis, data-driven modeling approaches, such as those collectively referred to as machine learning (ML) approaches, have proven to be effective alternatives to hydrological models that provide accurate simulations. Commonly utilized methods include random forest, artificial neural network (ANN), long short-term memory (LSTM), and others [16,23,25]. These algorithms typically quantify the relationship between the object and the factors by uncovering hidden patterns in large amounts of data. The limitation is that the ML approaches are founded on the crucial assumption that the interaction between the objects is stable at the determined stage. In reality, this relationship is not always fixed since the total environment is continually changing [24,26,27]. Consequently, a method is required to record the dynamic evolution in real time.

The data stream models can assist in answering such questions. Data streams are distinguished by continuity, temporality, breadth, and temporal evolution, with data patterns/concepts changing over time [28,29]. Data stream models, which can handle a continuous stream of input data, can be used for clustering, regression, classification, and time series analysis [30]. By detecting concept drift (alterations in the conditional distributions of goal variables in response to the input data) and applying a retraining process, data stream models enable designs to adapt in response to changing data

items [31,32]. To generate updated models, the data stream might discard unnecessary past data and incorporate new instances [33]. Furthermore, in the situation of changing data streams, the algorithm must be able to recognize and classify related events. Since the year 2000, data stream has been a hot research direction in the top database field and data mining conferences like ICDE (IEEE International Conference on Data Engineering). Currently, data stream research findings are employed in a variety of applications, such as sensor networks, telecommunication data, web logs, and stock trading. This approach was first introduced to the environmental field.

The Dongting Lake Wetland (DTL), an important regulating lake and water source, locates in the middle and lower reaches of the Yangtze River. In recent years, the DTL ecosystem has been deeply affected by the construction of dams and other water projects, especially the Three Gorges Dam, the world's largest water project [34]. Besides, the Yangtze River and the DTL have a complex interaction that includes plentiful water exchange rates, the evolution of sediment erosion and sedimentation, and the exchange of materials and energy [35,36]. Therefore, it has an irreplaceable diversion and regulation effect in the prevention of drought-floods and the utilization of water resources. Our understanding of the DTL's ever-complicated relationship with the Yangtze River continues to be incomplete under the influence of human activities in recent years [37]. Studying drought-flood events in the context of the river system and the DTL itself provides an interesting perspective for such research. Given that current hydrological models are insufficient for such a complex situation, the superior data stream method would appropriately apply to the environmental characteristics of the DTL, highlighting its dynamic modeling characteristics in the changing environment.

The main objectives of this research are: (1) to analyze the characteristics of drought-flood events in the DTL under the long time series from 1959 to 2019; (2) to explore the periodic change and evolution mechanism of drought and flood in the DTL under the change in river-lake relationship caused by man-made water conservancy project construction; and (3) to build a data stream model to quantify the factors that contribute to drought and flood occurrences in the DTL.

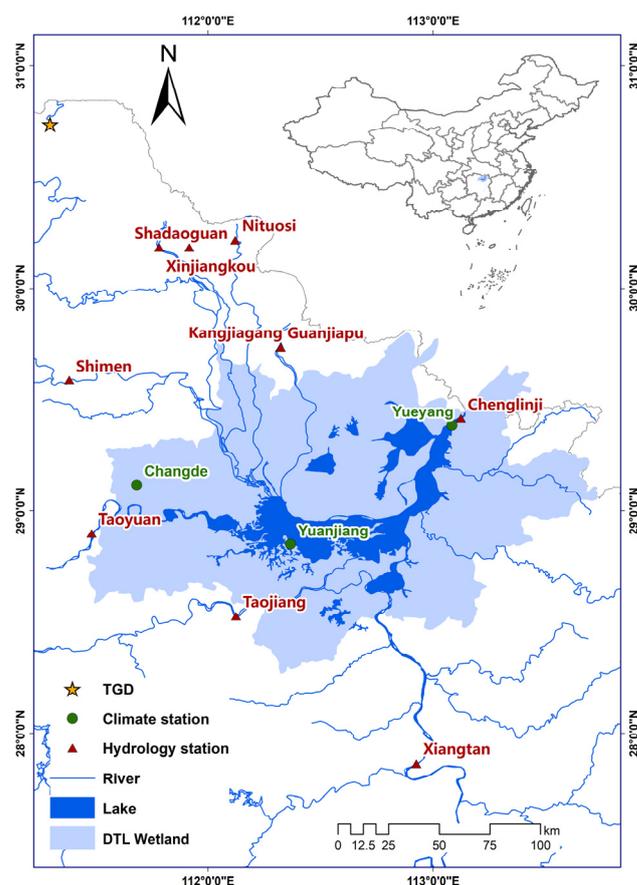
2. Materials and Methods

2.1. Study Area Description

The Dongting Lake Wetland (28°44'–29°35' N, 111°53'–113°05' E, Figure 1) is the second largest freshwater lake in China, located in the north of Hunan Province. The Dongting Lake Wetland has a total area of 2625 km² and a total volume of 16.7 billion m³. The average water altitude of the Dongting Lake is 33.5 m. The climate of the lake region belongs to the typical subtropical monsoon climate, which lies in the intersection zone of the Southeast monsoon and the Northwest monsoon. It receives an average annual precipitation of 1148–1837 mm, and the mean annual temperature ranges from 16 to 18 °C. The lake collects water from the Yangtze River through “Sankou” (three channels: Songzi, Taiping, and Ouchi) and “Sishui” (four rivers: Xiang, Zi, Yuan, and Li) and discharges it into the Yangtze River at Chenglingji station. The lake water is substantially affected by four large-scale water conservation projects: the Tiaoxian Port Blockage (Launching year, 1958), the Lower Jing River Cut-off (Launching year, 1966), the Gezhouba Hydro-Project (Launching year, 1970) and the Three Gorges Dam (Launching year, 1994). The study period (1959–2019) was therefore divided into four segments, each corresponding to the time interval before and after the construction of the projects (Table 1).

Table 1. Four study periods according to typical large water conservancy projects.

| Year | Water Conservancy Projects |
|-----------|----------------------------|
| 1959–1972 | Tiaoxian Port Blocking |
| 1973–1981 | Lower Jingjiang Cutting |
| 1982–2002 | Gezhouba Hydro-project |
| 2003–2019 | Three Gorges Dam |

**Figure 1.** Location of the Dongting Lake region and hydroclimatic stations.

For this research, data sets from ten hydrological sites and three meteorological stations in the Dongting Lake Wetland were chosen, including daily streamflow, daily precipitation, and daily temperature. The hydrological and meteorological data were collected from the Hunan Hydrology and Water Resources Survey Center (<https://www.hnsw.com.cn>, accessed on 1 December 2022) and the China Meteorological Data Center (<http://data.cma.cn>, accessed on 1 December 2022). The hydrological stations included the monitoring stations of “Sankou” and “Sishui” and the Chenglingji station, a typical representative hydrological station of the DTL (Figure 1). The bathymetry of the lake and the longitudinal slope of the water level at the monitoring site vary gently, and the streamflow variation at the Chenglingji station can reflect the changes of the whole DTL. Compared to the “Sishui” streamflow, the “Sankou” streamflow was relatively low and greatly influenced by the project. As a result, the sum of the “Sankou” (three-channel) streamflow was adopted as a study variable for human effect. Then, Yueyang, Changde, and Yuanjiang were chosen as representative meteorological stations for the east, south, and east of the DTL, respectively, and the average data value represented the climate of the entire region.

2.2. Standardized Precipitation Index (SPI)

After inputting precipitation parameters, we obtained SPI on multiple time scales (e.g., month, season, and year) using the standardized process, allowing us to compute and assess the wet and dry periods between adjacent seasons. In this study, the Gamma probability density function was used to fit the monthly precipitation data series [6]. The Gamma distribution was followed to convert to a normal distribution with a mean of zero and a standard deviation of one [38]. The probability density function is computed in the following manner:

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, (x > 0) \quad (1)$$

where α , β denote the shape and scale parameter of the gamma function, respectively. More details on the standardized precipitation index can be found in Supplementary Materials.

2.3. Standardized Streamflow Index (SSI)

Streamflow is a critical element for drought and flood monitoring. As with all standardized indexes, the standardized streamflow index (SSI) was able to assess the spatial and temporal variability of hydrological droughts and floods throughout the study area. The calculation procedure and classification level of SSI depends on long-term observed or simulated streamflow records. According to the previous research on the DTL, the distribution of streamflow values followed the Pearson type III distribution, so that the flood-drought index could be determined by normalizing the streamflow data series [7]. Prior work by Vicente-Serrano [39] included detailed calculations for the SSI series. Table 2 shows the classification of the SPI and SSI levels [6].

Table 2. Classification levels based on SPI/SSI.

| SPI/SSI Range | Classification | Probability |
|---------------|------------------|-------------|
| >2 | Extreme flood | 2.3% |
| 1.5–2 | Severe flood | 4.4% |
| 1~1.5 | Moderate flood | 9.2% |
| –1–1 | Near normal | 68.2% |
| –1.5––1 | Moderate drought | 9.2% |
| –2––1.5 | Severe drought | 4.4% |
| <–2 | Extreme drought | 2.3% |

When evaluating the SPI/SSI, it is essential to mention that the wet and dry situations are related to the historical average, not to the total precipitation/streamflow at a particular location. At the same time, the drought and flood events that were revealed through the SPI/SSI drought indices were compared to historic events. The comparison between SSI and SPI will provide the required time for the precipitation deficit to spread through the hydrological cycle and culminate in a discharge shortfall. Then, the duration of droughts and floods (periods when SPI/SSI last continuously above or below 0), severity (cumulative sum of the absolute values of SPI/SSI), and intensity (average severity of the entire duration) were approached by run theory [40] to illustrate drought-flood characteristics.

2.4. Wavelet Analysis Method

The wavelet analysis converts one-dimensional signals (like time series) into time and frequency domains in two-dimensional space and provides both time and frequency observation simultaneously [41,42]. This method, which was generally used in hydrological series analysis to reveal the characteristics of the time series, can reveal the non-stationary trend and periodicity hidden in the time series and estimate the time position of the time scale changes [43,44]. Here, wavelet analysis was used to check the scale and time of variations in droughts and floods. Specifically, the wavelet power spectrum (WPS) and global wavelet spectrum (GWS) were used to investigate the periodicity of drought and wetness, as well as changes in the power cycle of the monthly SPI and SSI series over time.

The Morlet wavelet was used here for wavelet decomposition. The calculation formula can be found in Supplementary Materials.

2.5. Data Stream Model

When the environmental context changes, the relationship between drought-flood and the influencing factors would shift unexpectedly, correlating to concept drift in the data stream [45]. We used an online bagging of Hoeffding adaptive trees (OBHAT) model to dynamically detect the response of drought and flood events to human activities and climate change. Hoeffding adaptive trees (HAT) is a decision tree algorithm that can rapidly process data streams. The HAT makes decisions based on new incoming instances rather than reusing them, with ADWIN (adaptive sliding window) serving as a change detector and error estimator for each split node in order to reset trees when their associated detectors generate a drift signal [46]. Therefore, we considered $S = \{s1, s2, \dots, sT, \dots\}$ to be an open-ended data sequence entered over time, containing input examples $X = \{x1, x2, \dots, xt, \dots\}$, where x_t denoted eigenvectors at time t . Each instance X contains precipitation, temperature, streamflow data, and other influencing variables. Each instance in X had a corresponding input in y , which was a sequence of class labels defined by SPI or SSI classification levels. In the end, the data stream OBHAT model was constructed by building incremental trees, training and testing the instances with the online bagging method, and discarding poorly performing trees based on the error rate as the concept changed [28]. We selected the software environment platform Massive Online Analysis (MOA) to implement the above model [47].

2.6. Online Feature Importance

The OBHAT model that we used is a tree model and is inherently interpretable. To measure the importance of each feature on the feature set to which it belongs (the influence of the factors), the tree structure model first needed to find the best node and the best branching method, and then calculate the feature “impurity” based on the nodes. Mean decrease in impurity (MDI) was later determined to evaluate the feature importance [48], calculating as follows:

$$H(k) = - \sum_{p \in V(S(k))} \frac{N_p^k}{N^k} \log_2 \frac{N_p^k}{N^k} \quad (2)$$

$$DI(k) = H(k) - N^{k_l} H(k_l) - N^{k_r} H(k_r) \quad (3)$$

$$MDI(f_i) = \frac{1}{T_n} \sum_{t \in T} \sum_{k \in t, S(k)=f_i} DI(k) \quad (4)$$

where k, k_l, k_r denote the split node and number of instances that traverse k node ‘s left and right children, respectively. $V_{(f_i)}$ is a collection of possible values for f_i . N, N_k, N_p^k denotes the whole number of instances, instances arriving at node k and a subset of N_k where feature $S(k)$ equals p . T and T_n denote a forest of trees and the number of trees within it, respectively.

Therefore, we constructed the OBHAT that provided MDI scores when updating the tree model. When it comes to updating or calculating the split node statistics on a continuous basis, we referred to the solution of Gomes et al. [27] for the exact principle. Following prequential cross-validation, the classification performance was evaluated by using the accuracy metric Kappa coefficient [49]. The schematic of the study methodology was shown in Figure 2.

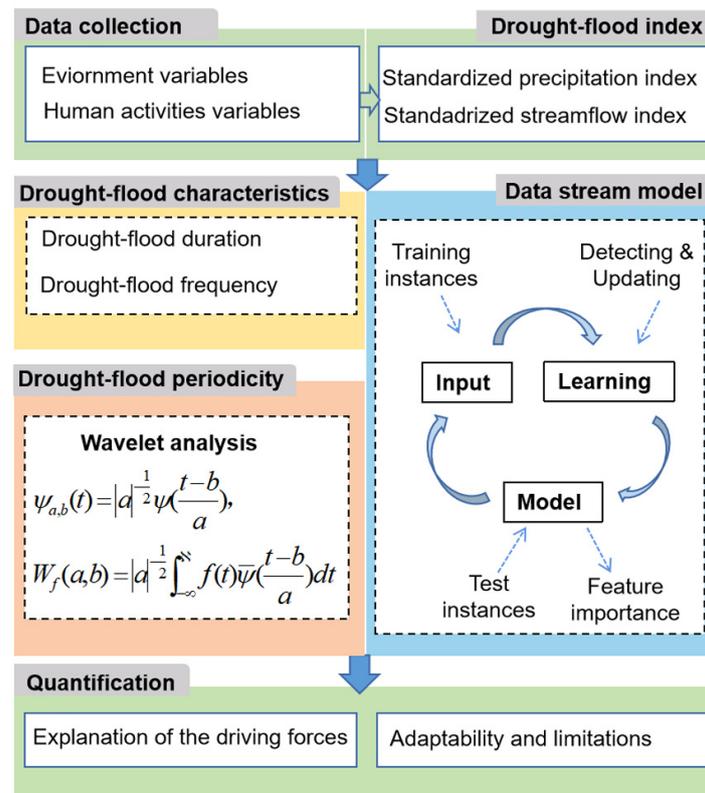


Figure 2. Schematic of the study methodology.

3. Results and Discussions

3.1. Time Series of Standardized Drought Indicators

The results of one-, three-, six-, 12-, and 24-month SPI/SSIs calculated for the DTL from 1959 to 2019 are shown in Figure 3. Drought-flood alternation was frequent and of short duration on shorter time scales, but less frequent and of longer duration on longer time scales (e.g., the SPI/SSI values for one and three months often fluctuated above and below zero, while fluctuations became slower on the 12- and 24-month time scales). Because the precipitation or streamflow variables are less persistent and more stochastic over short time scales, resulting in larger variability, oppositely, SPI/SSI over longer time scales include more prior effects of precipitation or streamflow. In addition, it was observed that the SPI and SSI show inconsistent relationships on different time scales. The difference between them reduced as the accumulation time lengthened, with the average correlation between SPI and SSI increasing from 0.338 (one-month) to 0.574. (12-month). The difference between them decreased with a longer accumulation period, as the average correlation between SPI and SSI changed from 0.338 (one-month) to 0.574 (12-month). The typical drought and flood years and their duration could become more prominent as the time scale was extended, such as the severe flood years of 1970, 1983, 1998, and 2002, and the severe drought years of 1972, 1978, 1986, and 2011.

The ability of SPI/SSI to identify drought/flood events on different time scales varied significantly. In this paper, historical drought and flood disaster information were compared with the events identified by two hydrological extreme indicator thresholds -1.50 and 1.50 for supporting evidence (Tables S1 and S2). It could be observed that one- and three-month SPI and SSI were effective in identifying drought and flood years that were more in line with the actual situation of the DTL. Since the standardized index analysis at shorter time scales was mainly used to monitor past wet and dry conditions, it helped reconstruct the flood history of the region. Besides, longer time scale data can mask the signal of short-term extreme precipitation events, making it difficult to detect major drought/flood events. Considering that most drought events lasted for several months, the one-month SPI/SSI

may not be capable of identifying drought events well. We emphasized the three-month index's utility in identifying drought and flood disasters.

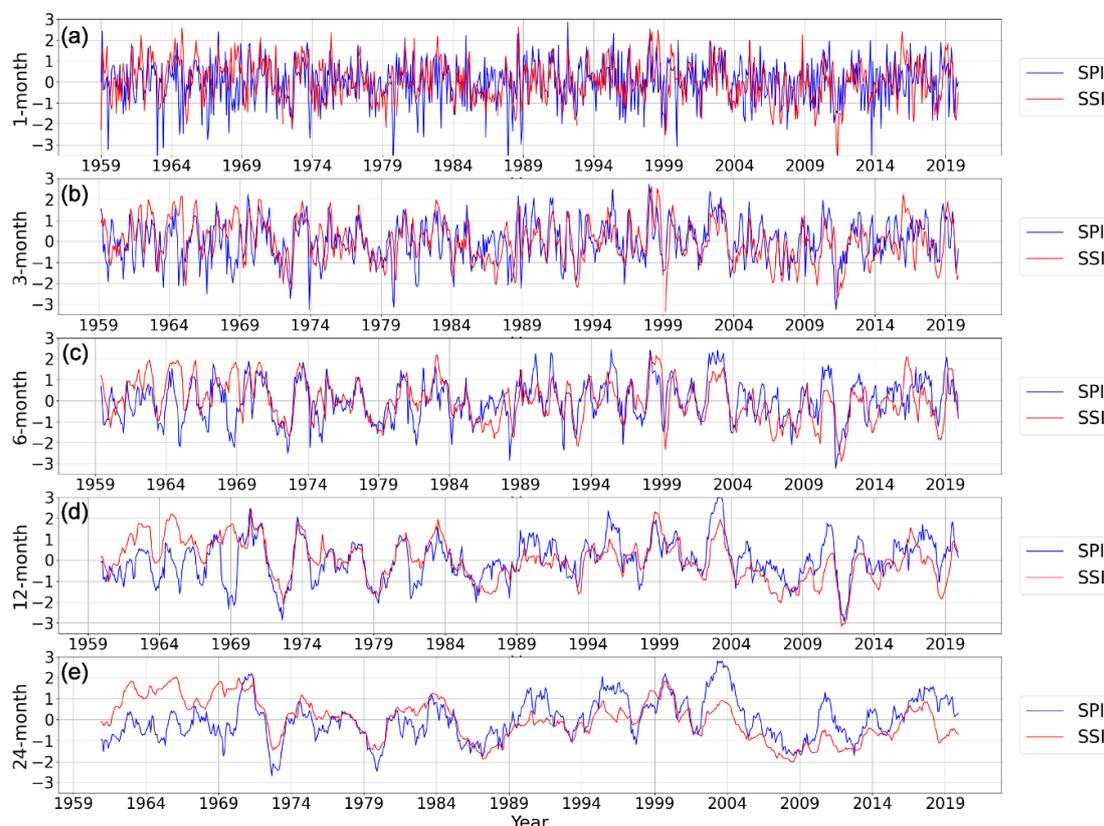


Figure 3. The SPI (solid blue line) and SSI (solid red line) at 1 (a), 3 (b), 6 (c), 12 (d), and 24 (e) month timescales in the DTL from 1959 to 2019.

3.2. Temporal Characteristics of Drought and Flood Events

Based on the identification of drought-floods presented in Table 2, the frequency and duration of the events that continued at DTL from 1959 to 2019 were obtained based on the three-month SPI/SSI. To show the trends more clearly, we also analyzed the statistics of drought and flood frequency and duration in the four study periods (Table 3, Figures S1 and S2). There were great differences in the frequency of drought and flood events from the perspective of meteorology and hydrology (Table 2). According to Table 3, the frequency of droughts in Dongting Lake Wetland was consistently higher than that of floods between 1959 and 1981. In 1982–2019, the opposite was true, with floods occurring more frequently than droughts. As a result, Dongting Lake was prone to drought before 1982 and flooding after 1982, which means the DTL showed the trend of droughts before floods in meteorology, especially the frequency of flood events from 1982 to 2002, which was as high as 22.62%. In terms of hydrology, it showed the opposite pattern of floods followed by droughts, with the highest frequency of floods and droughts in 1959–1973 and 2003–2019, respectively. After the construction of the Gezhouba Hydro-project and the Three Gorges Dam, the flood phenomenon decreased significantly by 4.91%, while the drought increased by 6.81%.

According to the frequency variation of drought and flood levels (Figure S1), the occurrence of meteorologically extreme drought and severe drought events gradually declined, except for 1973–1981, when 5.56% of extreme drought events occurred. There has been a minor increase in moderate floods, and as previously stated, the frequent flood events in 1982–2002 included 3.57% extreme floods. The change in the frequency of extreme events in hydrology was not significant, but the frequency of severe droughts increased

by 9.69%, and severe and moderate floods both reduced by roughly 8%. We also utilized the box-line plot to investigate the distribution of the duration of drought and flood events through four periods (Figure S2) and discovered that the characteristics of the change in duration and frequency might be similar. The average duration of meteorological droughts and floods fluctuated up and down for five months. Corresponding to the frequency characteristics of drought followed by flood, the duration of drought slightly decreased and the duration of flood slightly increased. While the duration of hydrological droughts and floods was generally longer than meteorological droughts and floods, the average duration of drought increased by 1.4 months throughout the period. The longest duration of floods was within 1959–1973, and the duration of floods in DTL greatly reduced in 2003–2019 after three major hydraulic projects. Although the flood control properties of artificially constructed hydraulic projects were entirely realized, the projects somehow extended the drought.

Table 3. Drought and flood events frequency in four time period.

| Index | Frequency | 1959–1972 | 1973–1981 | 1982–2002 | 2003–2019 |
|-------|-----------|-----------|-----------|-----------|-----------|
| SPI-3 | Drought | 19.28% | 14.81% | 12.30% | 16.18% |
| | Flood | 16.27% | 12.96% | 22.62% | 18.14% |
| SSI-3 | Drought | 9.64% | 15.74% | 13.49% | 22.55% |
| | Flood | 28.92% | 16.67% | 14.68% | 11.76% |

3.3. The Periodicity of Drought-Flood Events Reflected by SPI and SSI

The scale and time of variation in droughts and floods were checked by wavelet analysis in the DTL. Specifically, wavelet power spectrum and global wavelet spectrum were used to study the periodicity of drought and wetness, as well as the changes in the power cycle of monthly SPI and SSI series over time (Figure 4). During 1970–1977, the area surrounded by thick contours with relatively strong wavelet power oscillated periodically for about 28–48 months, and the determined significant power indicated that a drought occurred during this time frame. The figure also shows large wavelet power values at 95% confidence levels over 4–8 months cycles for a short duration, reflecting intermittent quasi-periodic oscillations. Similar to the wavelet power of SPI, the wavelet power of SSI also exhibited periodic features (Figure 4b). Figure 4b shows a period of 64–128 months from 1970 to 1995 with a 95% confidence level. Similar to the SPI, the periods 1962–1968 and 2009–2014 showed distinct 12–24-month cycles. The region surrounded by a thick contour with the strongest wavelet power oscillates periodically for about 90 months between 1982 and 1984. The periodic quasiperiodic oscillation also existed in the SSI sequence, which was characterized by 6–12 months period of wavelet power with a larger value but shorter duration. In general, SPI and SSI series had similar cyclic evolution characteristics, in which high-frequency or short-scale models (such as four- and eight-month components) were stronger. This suggested that the DTL was greatly affected by the short-cycle climate, where the analysis of short-term drought and flood cycles was a key procedure for annual water management research, which was consistent with the conclusion of Oguntunde et al. [42].

We also performed seasonal wavelet analysis of the DTL droughts and floods (Figure 4c–j). It was observed that the spatiotemporal patterns of seasonal drought and flood events had no obvious periodicity at large scale, and the main periodicity in spring was about nine and 23 months (Figure 4c,d) at small and medium scale, respectively. Although the wavelet power spectrum of SPI and SSI simultaneously depicted the enhanced power of the frequency band from six to 10 months in 1968–1972, the enhanced power of SSI was stronger in the frequency band from four to 12 months in 1998–2005 than 2–6 months when compared with the power of SPI. The main summer cycle of SPI was 8–16 months, occurring in 1968–1985 and 1998–2001, whereas the main summer cycle of SSI was 4–8 months, occurring in 1962–1973 and 2008–2012 (Figure 4e,f). At the same time, the SPI sequence had intermittent quasi-periodic oscillation with a short duration of 2–8 months, and the

SSI sequence only had a long duration period. The main periodic characteristics of SPI in the autumn were about four and 20 months (Figure 4g) respectively, and the relatively strongest wavelet power periodic oscillation was about 18–24 months in the two years around 1990. The autumn cycle of the SSI sequence with the large difference from spring and summer occurred later, and the small-scale feature of 4–8 months appeared only after 1990. The main periodic characteristics of winter drought and flood events were about six and 12 months (Figure 4i,j), and the short-time quasi-periodic oscillation was obvious.

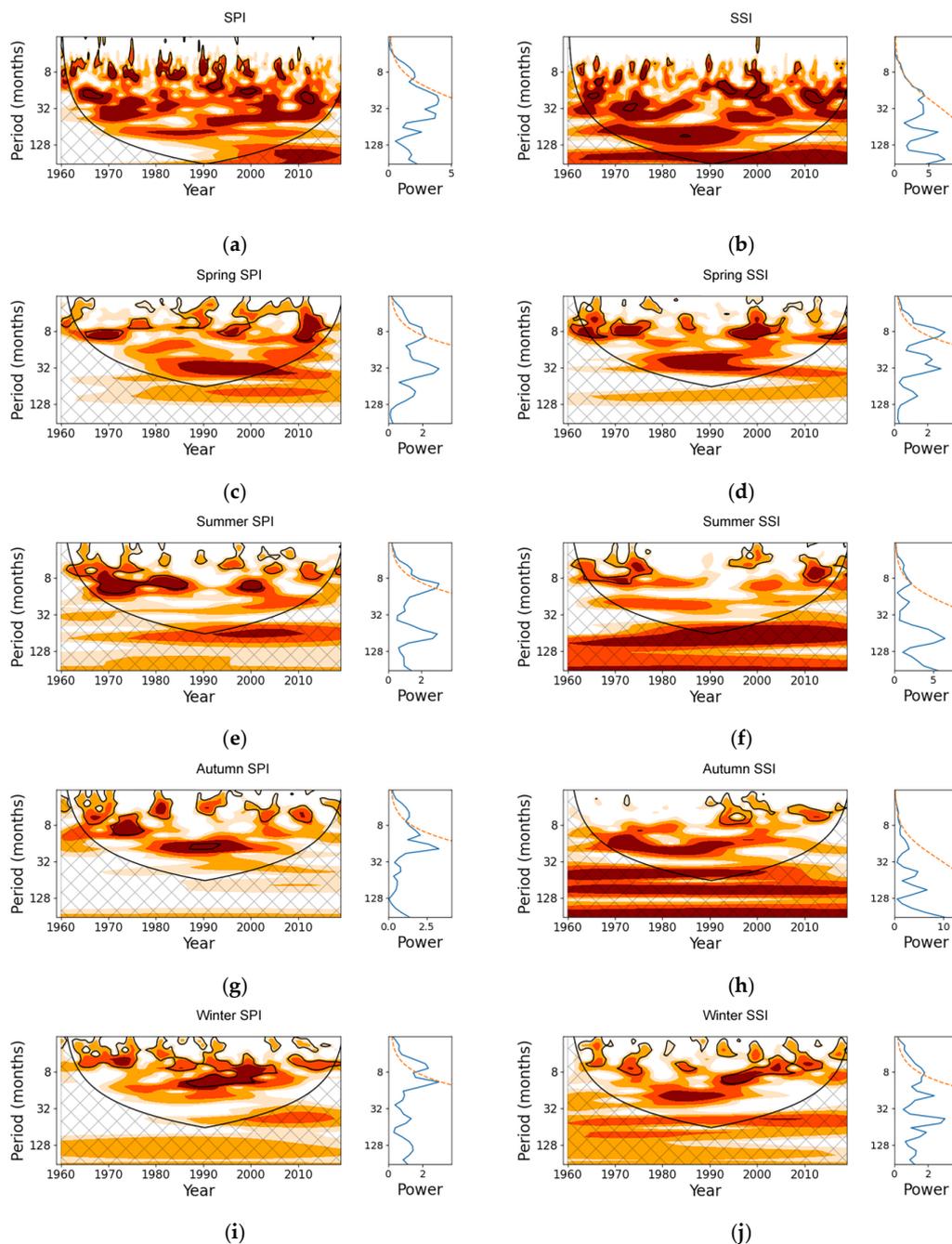


Figure 4. The wavelet power spectrums of SPI (left column) and SSI (right column) at multiyear (a,b) and seasonal (c–j) scales. Each subplot consists of a wavelet power spectrum (left panel) with the corresponding global wavelet spectrum (right panel). The colour codes for power values from light orange (low values) to dark orange (high values). The black circle lines show the maxima of the undulations of the wavelet power spectrum. Black concave contours indicate regions with 5% significance level.

3.4. Quantifying Drought-Flood Driving Factors by Data Stream Model

We counted the feature importance values of each variable for four time periods (Figure 5). For SPI, the Kappa coefficient of the OBHAT was 0.83, and the main factors affecting droughts and floods in the study area were precipitation, Sankou streamflow, and temperature. At the beginning of the study period, the importance of the Sankou streamflow characteristics accounted for the largest proportion of 39%. During 1959–1981, the importance proportion increased slowly, but the influence decreased by about 7.5% in the last 30 years. The proportion of the precipitation factor continued to increase until it was 35.3%, exceeding the Sankou by 3% as the main driver. The importance of temperature continued to decline in the period 1959–2002 and then increased after 2003. For SSI, the Kappa coefficient of the OBHAT was 0.80, and the main factors affecting droughts and floods in the study area were similar to SPI, including precipitation, Sankou streamflow, temperature, and Xiang River streamflow. The specific gravity of the Sankou streamflow fluctuated violently throughout the study period and dropped to 20.9% after two periods of large ups and downs. On the contrary, the trend of precipitation factors continued to increase until it exceeded 11.7% of the Sankou. The importance of the temperature factor decreased slightly, as did the SPI. It was worth noting that the Xiang factor showed an upward trend in accordance with precipitation at 22.6%, exceeding the Sankou and temperature to become the second most influential factor.

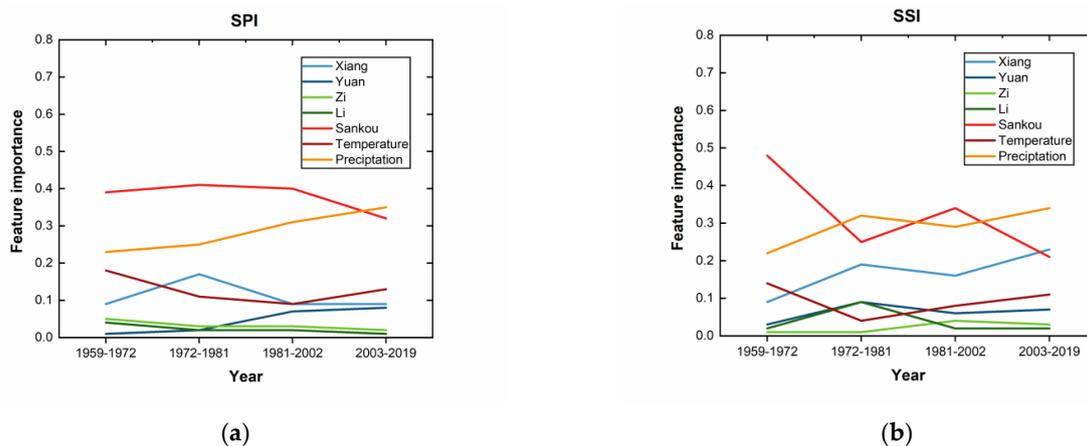


Figure 5. The variation of the feature importance value of each variable in the OBHAT model for SPI (a) and SSI (b) during the four periods of 1959–1972, 1973–1981, 1982–2002 and 2003–2019.

Figures 6 and 7 also showed the degree of drought and flood corresponding to the importance of each variable on the monthly scale to better explain the driving mechanism. We found that under the extreme humid conditions of SPI, the importance of the Sankou was always greater than other factors, and the maximum could reach 60.9% in August 1988. On the contrary, under extreme drought conditions, the precipitation factor exceeded the Sankou, such as in 1966, 1981, and 2013. The situation reflected by SSI was different from that of SPI. Under humid conditions, the characteristic importance values of the Sankou and precipitation were approximately similar, and there was no obvious prominent variable. In the case of extreme drought, 2006 was the key moment for the greatest change in the impact factors. Before that, the Sankou adjusted the regional drought to a large extent, and after that, precipitation became the biggest driving force for drought. In addition, temperature as a meteorological variable played a relatively large role in the drought and flood events identified by SPI or SSI. Temperature was typically more important during meteorologically moderate droughts and floods. Interestingly, the Xiang effect was similar to the temperature in the 1980s and 2010s, with stronger responses to meteorological droughts and floods.

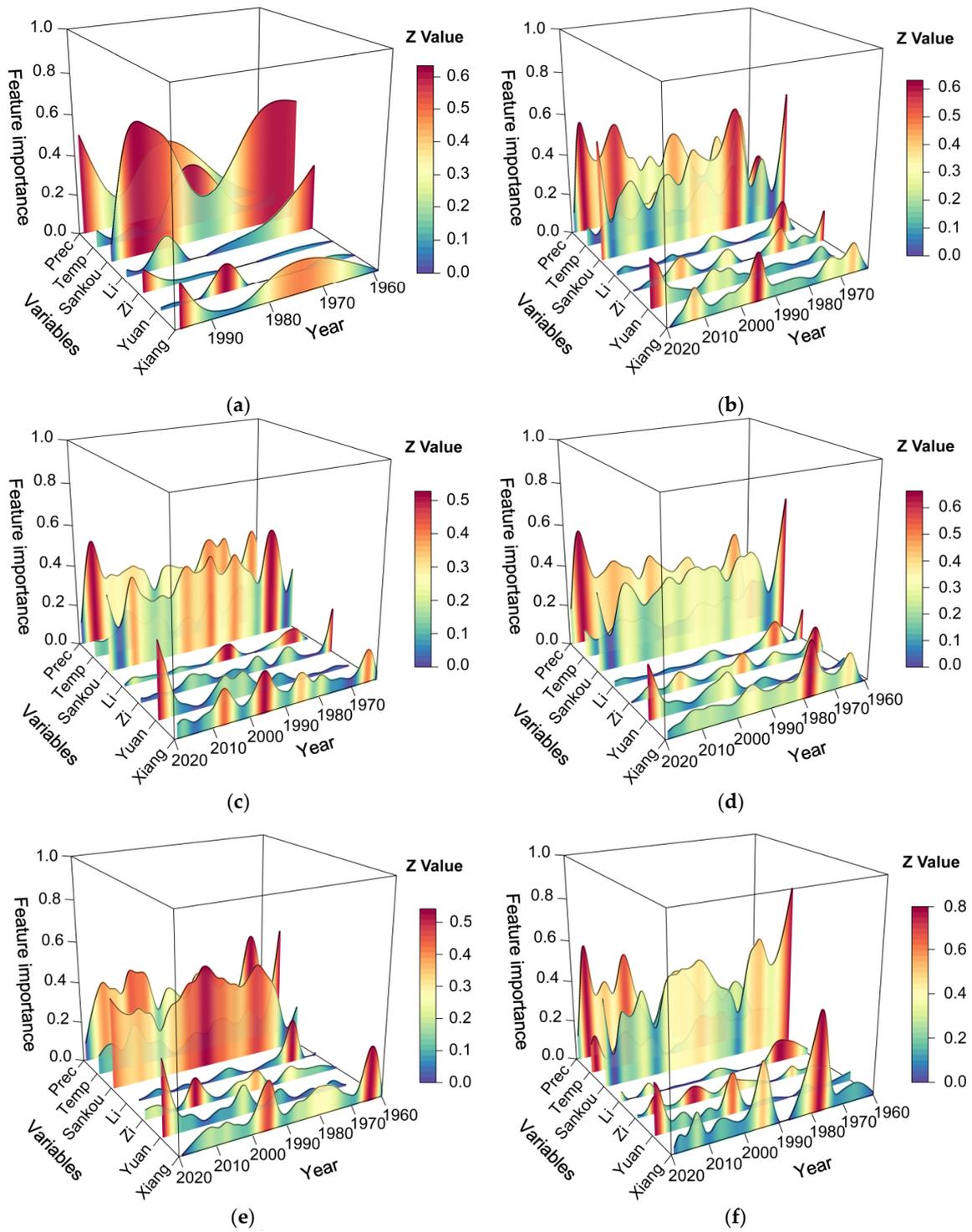


Figure 6. Cont.

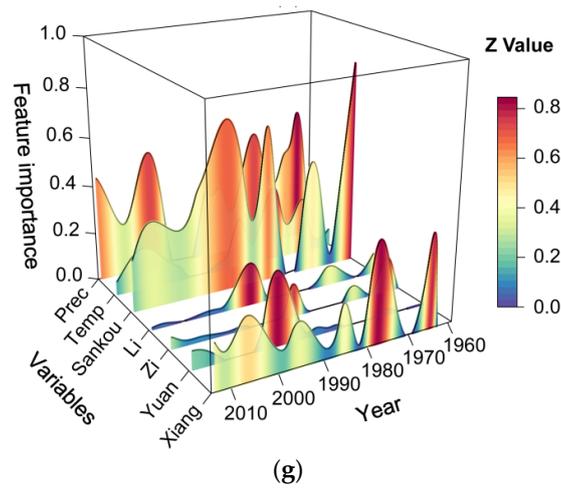


Figure 6. The feature importance values of each variable in the OBHAT based on different SPI level, (a) Extreme flood, (b) Severe flood, (c) Moderate flood, (d) Near normal, (e) Moderate drought, (f) Severe drought, (g) Extreme drought.

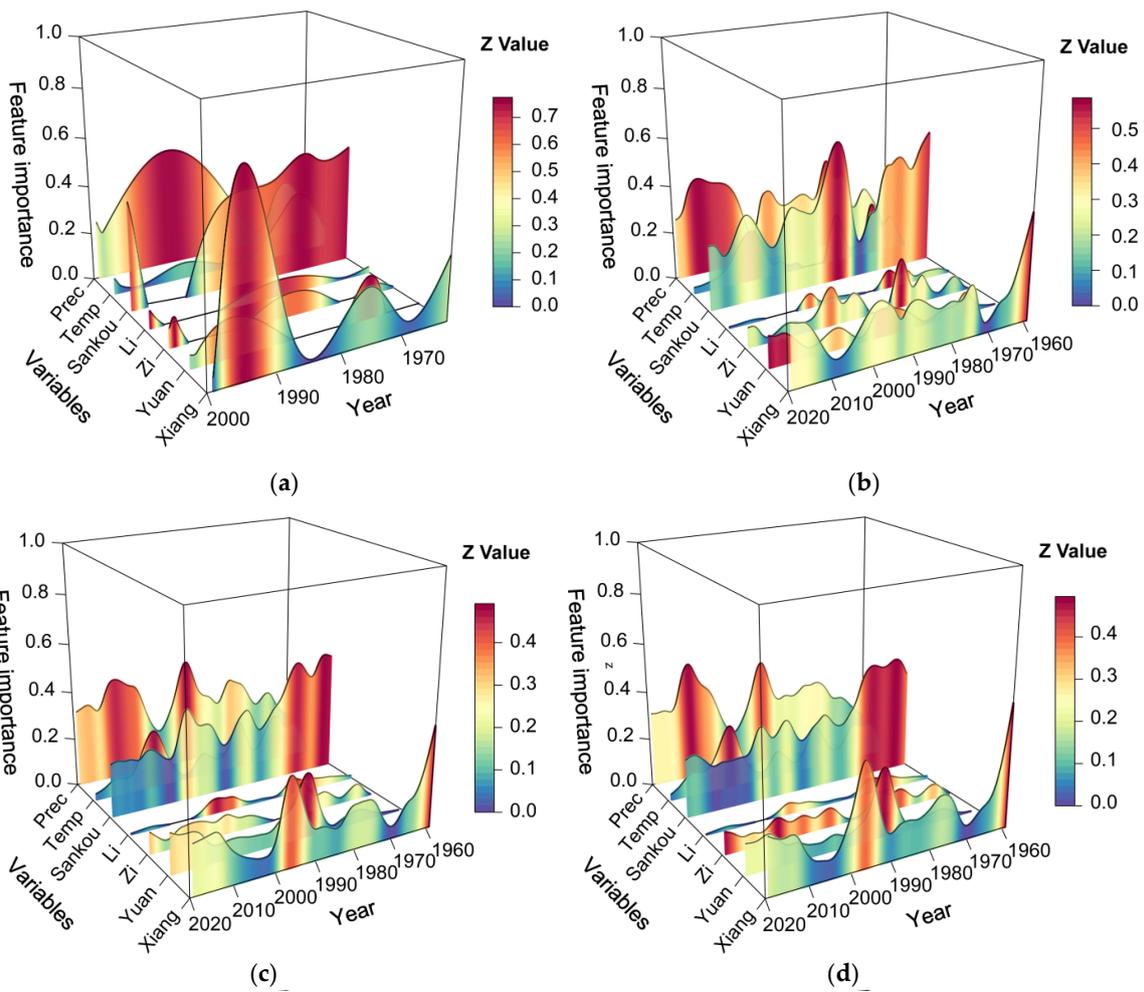


Figure 7. Cont.

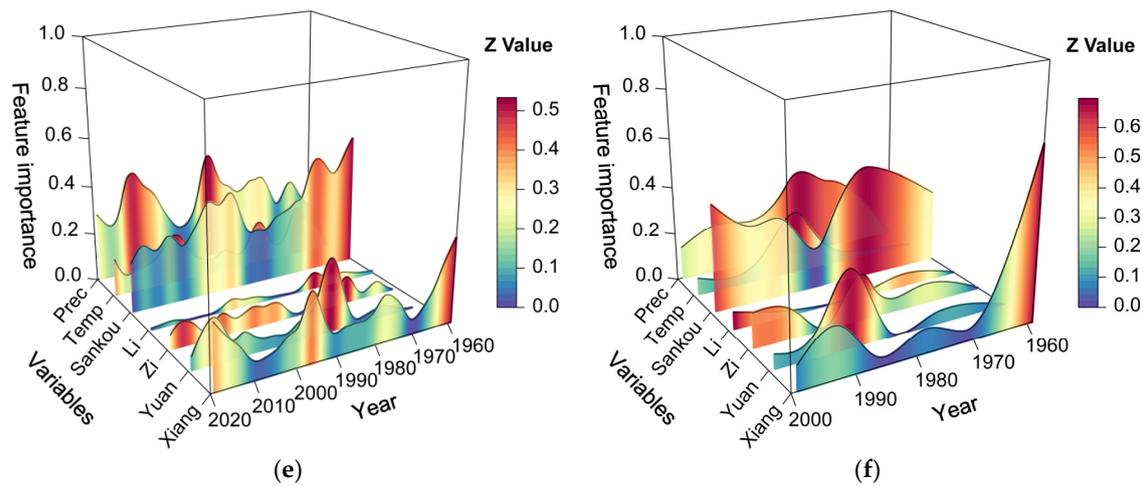


Figure 7. The feature importance values of each variable in the OBHAT based on different SPI level, (a) Extreme flood, (b) Severe flood, (c) Moderate flood, (d) Near normal, (e) Moderate drought, (f) Severe drought.

The case studies help us to identify the main drivers of drought and flood events in a changing and complex river-lake system. According to the results, Sankou streamflow, as an anthropogenically influenced study variable, had a significant impact on droughts and floods in the lake area. Despite the fact that the overall volume of streamflow in four rivers (Xiang, Zi, Yuan, and Li) on a multi-year annual average is twice as high as the Sankou, their feature importance is significantly lower at around a quarter of that of the Sankou. In addition, precipitation was the variable with the highest feature importance from 2003 to 2019, indicating that the influence of uncontrollable climatic factors on droughts and floods in the DTL was gradually increasing. While the Sankou streamflow's importance was decreasing, it proved that water conservancy projects had a role in the regulation of drought-floods. The operation of the TGD first attenuates the peak streamflow in July and August, effectively reducing the flood rate. Moreover, impoundment of the TGD during September and November plays a role in mitigating drought [37]. Hence, we suggest that the rolling forecast frequency of precipitation and the water regime in the Dongting Lake Wetland be strengthened to provide early guarantees for the decision. Besides, enhancing the scientific and optimal operation of the Gezhouba Hydro-project and the Three Gorges Dam will give full play to the regulation and storage roles of the Dongting Lake. We hope these measures will provide an important safety guarantee for Dongting Lake Wetland in flood control and drought mitigation. This study used the OBHAT data stream model to dynamically quantify the responses of drought-floods to environmental and anthropogenic conditions, providing new insights into disaster-driven simulations in changing environments, which would be valuable for future water management control and planning.

3.5. Adaptability and Limitations of the Data Stream Model

Due to climate change and human activity, traditional hydrological models and machine learning methods are no longer applicable to the changing environment. The advantage of data stream model lies in their dynamism and flexibility. By taking into account the non-stationary natural environment and human activities, the data stream model may dynamically simulate the impact of environmental changes on drought-flood events by incrementing and updating models. However, unlike most normal hydrological models, the data stream model lacks an explanation of meteorological and hydrological processes. Due to the nature of data mining [50,51], the proposed model is suitable for watersheds where long observation data sets exist. While data streams can contain additional drivers, only the obtained meteorological and hydrological data were tested in this study. For better understanding this interaction in the complex system, it is necessary to determine

the quantitative attribution of the factors leading to drought and flood. In the future, we recommend incorporating additional influencing factors, such as evaporation, soil moisture, and reservoir capacity, into the model and combining them with spatial scale analysis for a more comprehensive study.

4. Conclusions

It is vital to detect the drought-flood occurrences in a changing wetland system and to discover potential influencing factors by employing the dynamic data stream method. The results showed that the ability of SPI/SSI to identify individual drought-flood events varies significantly across time scales, with drought-flood events being more easily detected at three-month time scales. Short-term extreme precipitation events could be obscured by long-term data that contain precipitation or streamflow accumulation effects. This results in a failure to recognize catastrophic drought and flood occurrences.

Extreme droughts and floods were significantly reduced in the study area between 1959 and 2019, according to the 3-SPI/SSI, and the incidence of floods decreased significantly, but the frequency of mild drought increased. After the construction of the Gezhouba Hydro-Project and the Three Gorges Dam, the frequency and duration of floods were greatly reduced, while the frequency of droughts increased. Wavelet analysis revealed that DTL was susceptible to the short-period climate, and the SPI and SSI series had similar cyclic evolutionary characteristics, with high-frequency or short-scale patterns (e.g., four- or eight-month components) showing stronger performance.

In a changing environment, the data stream model provides a promising method for quantifying the causes of regional droughts and floods. According to the real-time dynamic detection of the feature importance of each variable by the stream model, precipitation and Sankou streamflow were the most important factors influencing the occurrence of drought and flood events in the DTL. We expect that the data stream model will provide novel insights into disaster-driven simulations in changing environments, as well as valuable information for future wetland management planning.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su142416778/s1>, Figure S1: The frequency of different drought and flood events levels revealed by counting 3-month SPI/SSI: (a) 1959–1972, (b) 1973–1981 and (c) 1982–2002, and (d) 2003–2019; Figure S2: The duration of drought and flood events revealed by counting 3-month SPI/SSI: (a) 1959–1972, (b) 1973–1981 and (c) 1982–2002, and (d) 2003–2019; Table S1: Validation of SPI and SSI for major DTL drought events in the last 59 years; Table S2: Validation of SPI and SSI for major DTL flood events in the last 59 years. References [11,12] are mentioned in Supplementary Materials file.

Author Contributions: Conceptualization, J.L.; methodology, Y.Z.; software, Y.Z.; formal analysis, Y.Z.; resources, Z.A.; writing—original draft preparation, Y.Z.; writing—review and editing, J.L., X.L., Z.Z. and W.W.; data curation, Y.Y. and S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (51979101), the Science and Technology Project of Hunan Water Resources Department (XSKJ2021000-06, XSKJ2022068-21), the Natural Science Foundation of Hunan Province (2019JJ20002).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, Y.; You, Q.; Lin, H.; Chen, C. Analysis of dry/wet conditions in the Gan River Basin, China, and their association with large-scale atmospheric circulation. *Glob. Planet. Chang.* **2015**, *133*, 309–317. [\[CrossRef\]](#)
2. Mishra, A.K.; Singh, V.P. Review paper A review of drought concepts. *J. Hydrol.* **2010**, *391*, 202–216. [\[CrossRef\]](#)
3. Huang, S.; Huang, Q.; Chang, J.; Leng, G.; Xing, L. The response of agricultural drought to meteorological drought and the influencing factors: A case study in the Wei River Basin, China. *Agric. Water Manag.* **2015**, *159*, 45–54. [\[CrossRef\]](#)
4. Wanders, N.; Wada, Y. Human and climate impacts on the 21st century hydrological drought. *J. Hydrol.* **2015**, *526*, 208–220. [\[CrossRef\]](#)
5. Lu, Y.; Yan, D.; Qin, T.; Song, Y.; Weng, B.; Yuan, Y.; Dong, G. Assessment of drought evolution characteristics and drought coping ability of water conservancy projects in Huang-Huai-Hai River Basin, China. *Water* **2016**, *8*, 378. [\[CrossRef\]](#)
6. McKee, T.B.; Doesken, N.J.; Kleist, J. The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference on Applied Climatology, Anaheim, CA, USA, 17–22 January 1993; pp. 179–183.
7. Vicente-Serrano, S.M.; López-Moreno, J.I.; Beguería, S.; Lorenzo-Lacruz, J.; Azorin-Molina, C.; Morán-Tejeda, E. Accurate Computation of a Streamflow Drought Index. *J. Hydrol. Eng.* **2012**, *17*, 318–332. [\[CrossRef\]](#)
8. Palmer, W.C. *Meteorological Drought*; Research Paper, No. 45; U.S. Weather Bureau: Washington, DC, USA, 1965; p. 58.
9. Nalbantis, I.; Tsakiris, G. Assessment of hydrological drought revisited. *Water Resour. Manag.* **2009**, *23*, 881–897. [\[CrossRef\]](#)
10. Shukla, S.; Wood, A.W. Use of a standardized runoff index for characterizing hydrologic drought. *Geophys. Res. Lett.* **2008**, *35*, 1–7. [\[CrossRef\]](#)
11. Seiler, R.A.; Hayes, M.; Bressan, L. Using the standardized precipitation index for flood risk monitoring. *Int. J. Climatol.* **2002**, *22*, 1365–1376. [\[CrossRef\]](#)
12. Wang, Y.; Chen, X.; Chen, Y.; Liu, M.; Gao, L. Flood/drought event identification using an effective indicator based on the correlations between multiple time scales of the Standardized Precipitation Index and river discharge. *Theor. Appl. Climatol.* **2017**, *128*, 159–168. [\[CrossRef\]](#)
13. Heim, R.R. A Review of Twentieth-Century Drought Indices Used in the United States. *Bull. Am. Meteorol. Soc.* **2002**, *83*, 1149–1166. [\[CrossRef\]](#)
14. Javed, T.; Li, Y.; Rashid, S.; Li, F.; Hu, Q.; Feng, H.; Chen, X.; Ahmad, S.; Liu, F.; Pulatov, B. Performance and relationship of four different agricultural drought indices for drought monitoring in China’s mainland using remote sensing data. *Sci. Total Environ.* **2021**, *759*, 143530. [\[CrossRef\]](#)
15. Qi, P.; Xu, Y.J.; Wang, G. Quantifying the individual contributions of climate change, dam construction, and land use/land cover change to hydrological drought in a marshy river. *Sustainability* **2020**, *12*, 3777. [\[CrossRef\]](#)
16. Wu, Z.; Zhou, Y.; Wang, H.; Jiang, Z. Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse. *Sci. Total Environ.* **2020**, *716*, 137077. [\[CrossRef\]](#)
17. Wu, J.; Chen, X.; Yao, H.; Gao, L.; Chen, Y.; Liu, M. Non-linear relationship of hydrological drought responding to meteorological drought and impact of a large reservoir. *J. Hydrol.* **2017**, *551*, 495–507. [\[CrossRef\]](#)
18. López-Moreno, J.I.; Vicente-Serrano, S.M.; Beguería, S.; Garcia-Ruiz, J.M.; Portela, M.M.; Almeida, A.B. Dam effects on droughts magnitude and duration in a transboundary basin: The lower river tagus, Spain and Portugal. *Water Resour. Res.* **2009**, *45*, 1–13. [\[CrossRef\]](#)
19. Wen, L.; Rogers, K.; Ling, J.; Saintilan, N. The impacts of river regulation and water diversion on the hydrological drought characteristics in the Lower Murrumbidgee River, Australia. *J. Hydrol.* **2011**, *405*, 382–391. [\[CrossRef\]](#)
20. Mei, X.; Van Gelder, P.H.A.J.M.; Dai, Z.; Tang, Z. Impact of dams on flood occurrence of selected rivers in the United States. *Front. Earth Sci.* **2017**, *11*, 268–282. [\[CrossRef\]](#)
21. Jiao, D.; Wang, D.; Lv, H. Effects of human activities on hydrological drought patterns in the Yangtze River Basin, China. *Nat. Hazards* **2020**, *104*, 1111–1124. [\[CrossRef\]](#)
22. Nearing, G.S.; Tian, Y.; Gupta, H.V.; Clark, M.P.; Harrison, K.W.; Weijs, S.V. A philosophical basis for hydrological uncertainty. *Hydrol. Sci. J.* **2016**, *61*, 1666–1678. [\[CrossRef\]](#)
23. Schmidt, L.; Heße, F.; Attinger, S.; Kumar, R. Challenges in Applying Machine Learning Models for Hydrological Inference: A Case Study for Flooding Events Across Germany. *Water Resour. Res.* **2020**, *56*, e2019WR025924. [\[CrossRef\]](#)
24. Yang, Q.; Zhang, H.; Wang, G.; Luo, S.; Chen, D.; Peng, W.; Shao, J. Dynamic runoff simulation in a changing environment: A data stream approach. *Environ. Model. Softw.* **2019**, *112*, 157–165. [\[CrossRef\]](#)
25. Adikari, K.E.; Shrestha, S.; Ratnayake, D.T.; Budhathoki, A.; Mohanasundaram, S.; Dailey, M.N. Evaluation of artificial intelligence models for flood and drought forecasting in arid and tropical regions. *Environ. Model. Softw.* **2021**, *144*, 105136. [\[CrossRef\]](#)
26. Cassidy, A.P.; Deviney, F.A. Calculating feature importance in data streams with concept drift using Online Random Forest. In Proceedings of the 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 27–30 October 2014; pp. 23–28. [\[CrossRef\]](#)
27. Gomes, H.M.; De Mello, R.F.; Pfahringer, B.; Bifet, A. Feature Scoring using Tree-Based Ensembles for Evolving Data Streams. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 761–769. [\[CrossRef\]](#)
28. Orduña Cabrera, F.; Sánchez-Marrè, M. Environmental data stream mining through a case-based stochastic learning approach. *Environ. Model. Softw.* **2018**, *106*, 22–34. [\[CrossRef\]](#)

29. Shao, J.; Huang, F.; Yang, Q.; Luo, G. Robust Prototype-Based Learning on Data Streams. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 978–991. [[CrossRef](#)]
30. Gama, J. *Knowledge Discovery from Data Streams*; CRC Press: Boca Raton, FL, USA, 2007; pp. 1–235. [[CrossRef](#)]
31. Widmer, G. Learning in the presence of concept drift and hidden contexts. *Mach. Learn.* **1996**, *23*, 69–101. [[CrossRef](#)]
32. Gama, J.; Indr, L.; Bifet, A.; Pechenizkiy, M.; Bouchachia, A. A Survey on Concept Drift Adaptation. *Curr. Proteom.* **2014**, *7*, 258–264. [[CrossRef](#)]
33. Razmjoo, A.; Xanthopoulos, P.; Zheng, Q.P. Online feature importance ranking based on sensitivity analysis. *Expert Syst. Appl.* **2017**, *85*, 397–406. [[CrossRef](#)]
34. Zhang, Q.; Zhou, Y.; Singh, V.P.; Chen, X. The influence of dam and lakes on the Yangtze River streamflow: Long-range correlation and complexity analyses. *Hydrol. Process.* **2012**, *26*, 436–444. [[CrossRef](#)]
35. Yuan, Y.; Zeng, G.; Liang, J.; Huang, L.; Hua, S.; Li, F.; Zhu, Y.; Wu, H.; Liu, J.; He, X.; et al. Variation of water level in Dongting Lake over a 50-year period: Implications for the impacts of anthropogenic and climatic factors. *J. Hydrol.* **2015**, *525*, 450–456. [[CrossRef](#)]
36. Liang, J.; Yi, Y.; Li, X.; Yuan, Y.; Yang, S.; Li, X.; Zhu, Z.; Lei, M.; Meng, Q.; Zhai, Y. Detecting changes in water level caused by climate, land cover and dam construction in interconnected river–lake systems. *Sci. Total Environ.* **2021**, *788*. [[CrossRef](#)]
37. Liang, J.; Meng, Q.; Li, X.; Yuan, Y.; Peng, Y.; Li, X.; Li, S.; Zhu, Z.; Yan, M. The influence of hydrological variables, climatic variables and food availability on Anatidae in interconnected river-lake systems, the middle and lower reaches of the Yangtze River floodplain. *Sci. Total Environ.* **2021**, *768*, 144534. [[CrossRef](#)]
38. Kao, S.C.; Govindaraju, R.S. A copula-based joint deficit index for droughts. *J. Hydrol.* **2010**, *380*, 121–134. [[CrossRef](#)]
39. Vicente-Serrano, S.M. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *J. Clim.* **2009**, *23*, 1696–1718. [[CrossRef](#)]
40. Yevjevich, V. An objective approach to definitions and investigations of continental hydrologic droughts. *J. Hydrol.* **1969**, *7*, 353. [[CrossRef](#)]
41. Kang, S.; Lin, H. Wavelet analysis of hydrological and water quality signals in an agricultural watershed. *J. Hydrol.* **2007**, *338*, 1–14. [[CrossRef](#)]
42. Oguntunde, P.G.; Abiodun, B.J.; Lischeid, G. Impacts of climate change on hydro-meteorological drought over the Volta Basin, West Africa. *Glob. Planet. Chang.* **2017**, *155*, 121–132. [[CrossRef](#)]
43. Torrence, C.; Compo, G.P. A Practical Guide to Wavelet Analysis. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 61–78. [[CrossRef](#)]
44. Grossmann, A.; Morle, J.; Paul, T. Transforms associated to square integrable group representations. I. general results. *Fundam. Pap. Wavelet Theory* **2009**, *2473*, 140–146. [[CrossRef](#)]
45. Folino, G.; Pisani, F.S.; Pontieri, L. A GP-based ensemble classification framework for time-changing streams of intrusion detection data. *Soft Comput.* **2020**, *24*, 17541–17560. [[CrossRef](#)]
46. Bifet, A.; Gavaldà, R. Learning from time-changing data with adaptive windowing. In Proceedings of the 2007 SIAM International Conference on Data Mining (SDM), Minneapolis, MN, USA, 26–28 April 2007; pp. 443–448. [[CrossRef](#)]
47. Bifet, A.; Holmes, G.; Pfahringer, B.; Kranen, P.; Kremer, H.; Jansen, T.; Seidl, T. Moa: Massive online analysis, a framework for stream classification and clustering. In Proceedings of the First Workshop on Applications of Pattern Analysis, Windsor, UK, 1–3 September 2010.
48. Louppe, G.; Wehenkel, L.; Sutura, A.; Geurts, P. Understanding variable importances in Forests of randomized trees. *Adv. Neural Inf. Process. Syst.* **2013**, *1*, 431–439.
49. Bifet, A.; De Francisci Morales, G.; Read, J.; Holmes, G.; Pfahringer, B. Efficient online evaluation of big data stream classifiers. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; pp. 59–68. [[CrossRef](#)]
50. Montanari, A.; Young, G.; Savenije, H.H.G.; Hughes, D.; Wagener, T.; Ren, L.L.; Koutsoyiannis, D.; Cudennec, C.; Toth, E.; Grimaldi, S.; et al. “Panta Rhei-Everything Flows”: Change in hydrology and society-The IAHS Scientific Decade 2013-2022. *Hydrol. Sci. J.* **2013**, *58*, 1256–1275. [[CrossRef](#)]
51. Gibert, K.; Izquierdo, J.; Sánchez-Marrè, M.; Hamilton, S.H.; Rodríguez-Roda, I.; Holmes, G. Which method to use? An assessment of data mining methods in Environmental Data Science. *Environ. Model. Softw.* **2018**, *110*, 3–27. [[CrossRef](#)]