

# Hydrologic analysis of the Yellow River based on linear regression and M-K mutation test

Mingyi Chen \*

School of Economics&Management, Shanghai Maritime University, Shanghai, China, 201306

\* Corresponding author: 202211010079@stu.shmtu.edu.cn

**Abstract.** Hydrological analysis of the Yellow River is of great significance to alleviating soil erosion of the Yellow River. In order to better analyze the relationship between water level, water flow and sediment content of the Yellow River downstream of the Xiaolangdi reservoir, based on linear regression and Mann-Kendall mutation test, combined with the advantages of Zhang *et al.* in handling with linear regression and M-K mutation test, a linear regression fitting model and M-K mutation test model for hydrological analysis are constructed. Considering from both micro and macro perspectives, this article obtained linear relationships between the water level, water flow and sediment content of the Yellow River, and them correspond mutations time. The accuracy of these analysis has important reference value for sediment regulation, flood control and alleviation of soil erosion, then disaster prevention.

**Keywords:** Linear Regression; M-K Test; The Yellow River; Hydrologic Analysis.

## 1. Introduction

The Hydrological analysis in the middle and lower reaches of the Yellow River mainly benefit flood control, ice prevention, silt reduction, irrigation, water supply and power generation, and has exerted good social and economic benefits [1]. While previous literature have used multiple regression analysis and time series chart in hydrology analysis. Although this method is applicable, the accuracy needs to be improved [2] [3]. Linear regression is a good method to analyze the relationship between water level, water flow and sediment content of the Yellow River. Meanwhile, the environmental deterioration in the upper reaches aggravates the soil and water loss of the Yellow River, using M-K mutation test to find the mutation time point of soil and water flow can better explore the start time of water and soil loss. Combining linear regression and M-K test can not only effectively analyze the changes of the Yellow River hydrology and formulate relevant countermeasures to alleviate soil erosion, but also reduce the error and increase the accuracy [4] [5] [6] [7]. Based on linear regression and M-K mutation test, this paper studies the internal connection and overall trend of the Yellow River hydrology, establishes a model and analyzes the functional relationship between sediment content, water level and water flow in the lower reaches of Xiaolangdi Reservoir, as well as their overall changes and mutation time points from 2016 to 2021. (Data sources :<https://www.mcm.edu.cn/>).

## 2. Linear regression fitted model's building and visualization

### 2.1. Spearman's correlation analysis

Analysis of the relationship between the water level, water flow, sediment content of the Yellow River requires correlation analysis firstly to determine whether there is a linear relationship between them. Referenced Zhou *et al.*<sup>[8]</sup> in the study of the Muyubao Mount in Sanxia landslide deformation and reservoir water level correlation quantitative analysis, Spearman correlation analysis is used to study the correlation of the Yellow River water level, water flow, sediment content, as is shown in formula (1).

$$\rho = \frac{\frac{1}{n} \sum_{i=1}^n (R(x_i) - \overline{R(x)}) \cdot (R(y_i) - \overline{R(y)})}{\sqrt{\left( \frac{1}{n} \sum_{i=1}^n (R(x_i) - \overline{R(x)})^2 \right) \cdot \left( \frac{1}{n} \sum_{i=1}^n (R(y_i) - \overline{R(y)})^2 \right)}} \quad (1)$$

Furthermore, in order to calculate, formula (1) can be simplified to formula (2).

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2)$$

Among the formula (2),  $n$  is referred to sample data,  $d_i^2$  is referred to the difference between the rank value of the  $i$ -th data pair. For Spearman's correlation coefficient  $\rho$ , the closer to 1, the stronger the correlation. If  $|\rho| \geq 0.8$ , it indicates a high correlation between variables, if  $0.4 \leq |\rho| < 0.8$ , it indicates a moderate correlation, if  $0.2 \leq |\rho| < 0.4$ , it indicates a weak correlation, and if  $|\rho| < 0.2$ , it indicates a very weak correlation.

According to the data of a hydrological monitoring station downstream of Xiaolangdi Reservoir on the Yellow River, using Spearman's correlation coefficient test to analyze the sediment content, water flow and water level then drawing the conclusions is the final step of correlation analysis, which is shown in the Table 1.

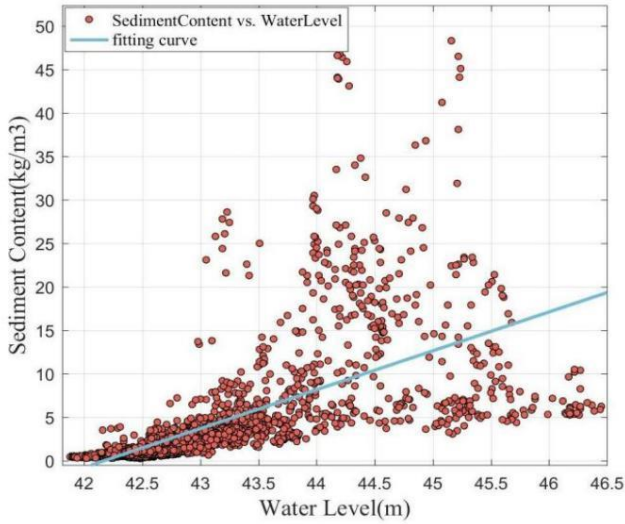
**Table 1.** Spearman correlation coefficient

Table of Spearman correlation coefficient (sample=2159)				
		Sediment content	water level	water flow
Sediment content	correlation coefficient	1.000	0.892	0.926
water level	correlation coefficient	0.892	1.000	0.960
water flow	correlation coefficient	0.926	0.960	1.000

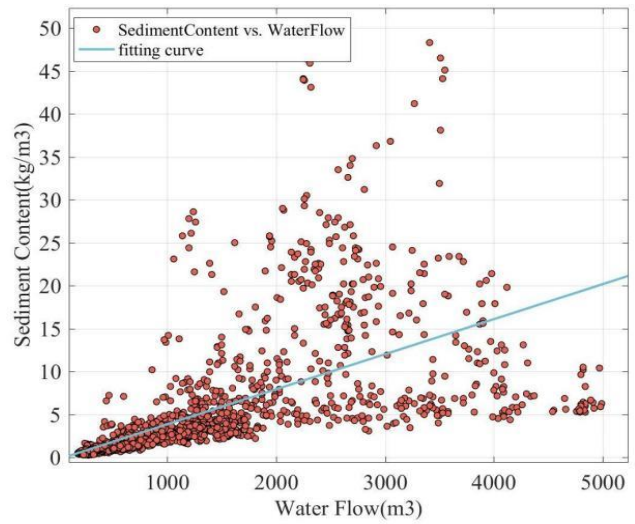
The Spearman correlation coefficient between sediment content and water level is 0.892, the Spearman correlation coefficient between sediment content and water flow is 0.926, and the Spearman correlation coefficient between the water level and the water flow is 0.960, all of which belongs to the range of  $|\rho| \geq 0.8$ , indicating that the sediment content, water level and water flow are highly linear correlation.

## 2.2. The construction of Linear regression fitted model

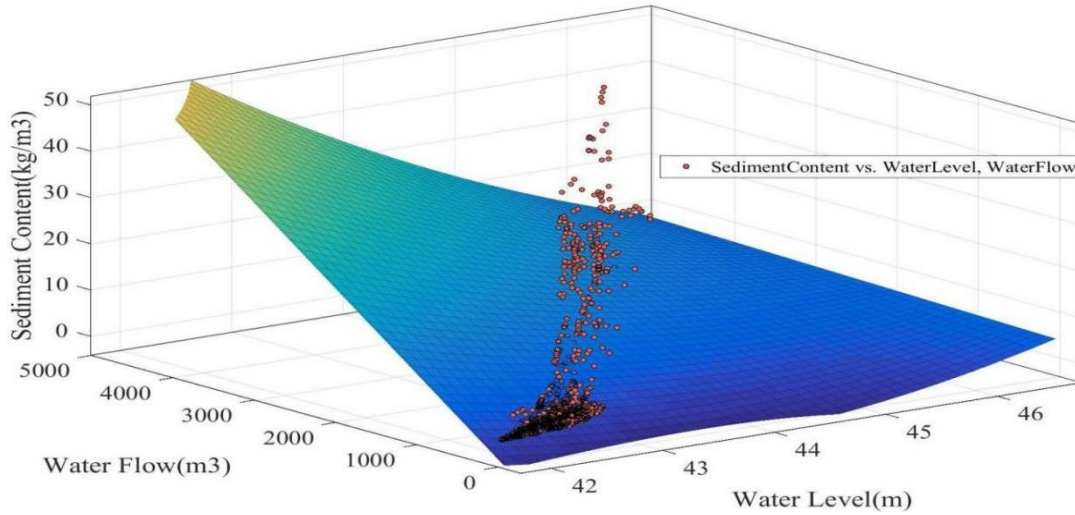
It is examined that the sediment content showed significant relationship with water flow and water level in a linear manner. Since they are a linear relationship, then let the time be  $t$ , which is referred to the monitoring time, let the data of water level, water flow be  $Z(t), Q(t)$ , which is referred to the independent variable  $x$ , and sediment content be  $C(t)$ , which is referred to the dependent variable  $y$ , using least square method to reveal their connection and then make visual processing, which is shown in the Fig 1-3.



**Figure 1.** Scatter plot and linear fit of sediment content and water level



**Figure 2.** Scatter plot and linear fit of sediment content and water flow



**Figure 3.** Scatter plot and linear fit of sediment content, water level and water flow

Wei [9] suggested the multivariate interaction or high-order term fitting can better reduce the residual difference, and the fitting accuracy is relatively high for the regression fitting of multivariate single dependent variable. Considering the interaction between water level and water flow on sand content, Fig 3 was added, and the linear fitting equations for sand content, water level, water flow, the combination of level and flow are obtained, which are shown in the formula (3)-(5).

$$C(t) = 4.469Z(t) - 188.4 \quad (3)$$

$$C(t) = 0.004Q(t) - 0.2285 \quad (4)$$

$$C(t) = 1789 - 82.63Z(t) + 0.1249Q(t) + 0.9512Z^2(t) - 0.002Z(t)Q(t) \quad (5)$$

While obtaining the functional relationship between the water level, water flow and sand content of the Yellow River (formula (3)-(5)), for the micro aspect to say, the prediction and change of sand content can be realized in advance by the change of water level and water flow, which has a forward-looking effect on flood control and alleviation of soil erosion in the lower Yellow River plain.

### 3. M-K mutation test and its visualization

#### 3.1. Time series chart

After analyzing the hydrology of the Yellow River at the micro level above, this paper then analyze the hydrology of the Yellow River from the macro level and make the time-series chart to observe the changes of annual hydrology of the Yellow River based on the data from 2016 to 2021. Before making the chart, processing the dates as counting. For example, let the date of January 1, 2016 be the number 1, and the data corresponding to the number 1 is the hydrological data measured on January 1, 2016. As there is monitoring data once a day, the number of the next day is the number plus 1, and December 31, 2021 is the number 2192, which is shown in the Fig 4-6.

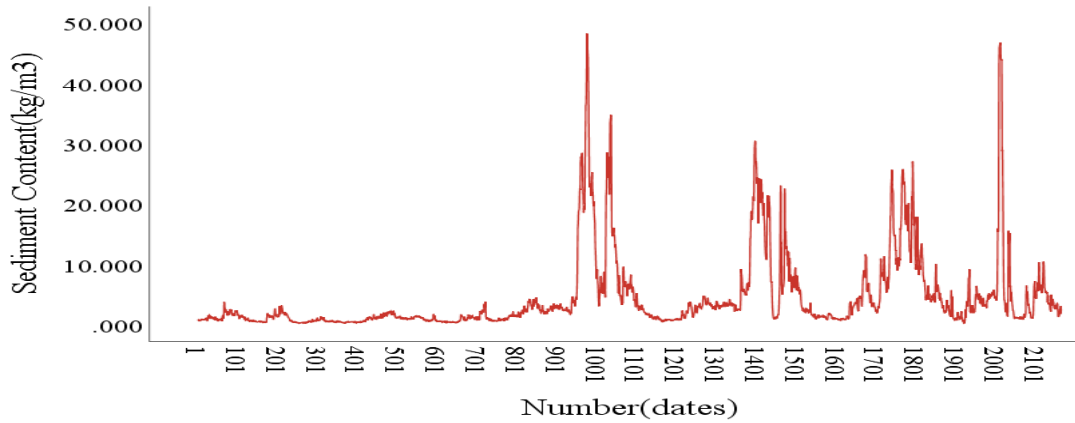


Figure 4. Time series chart of sediment content

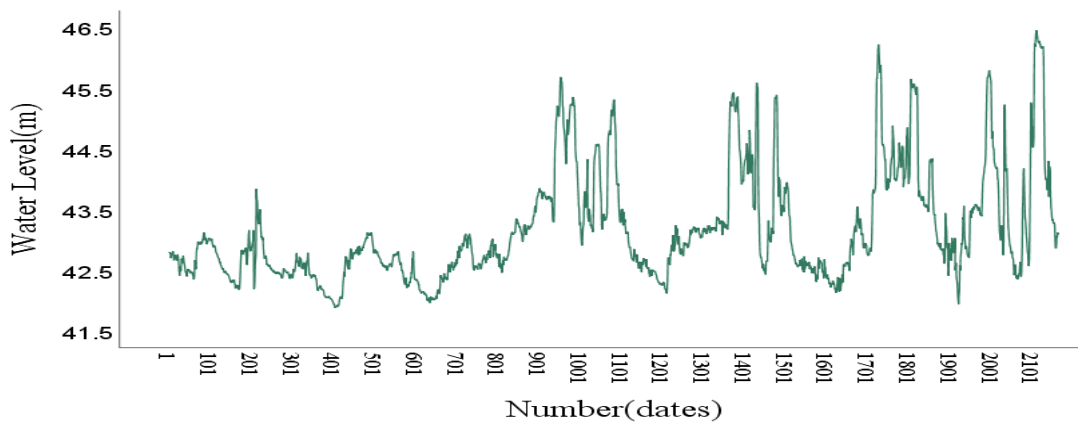


Figure 5. Time series chart of water level

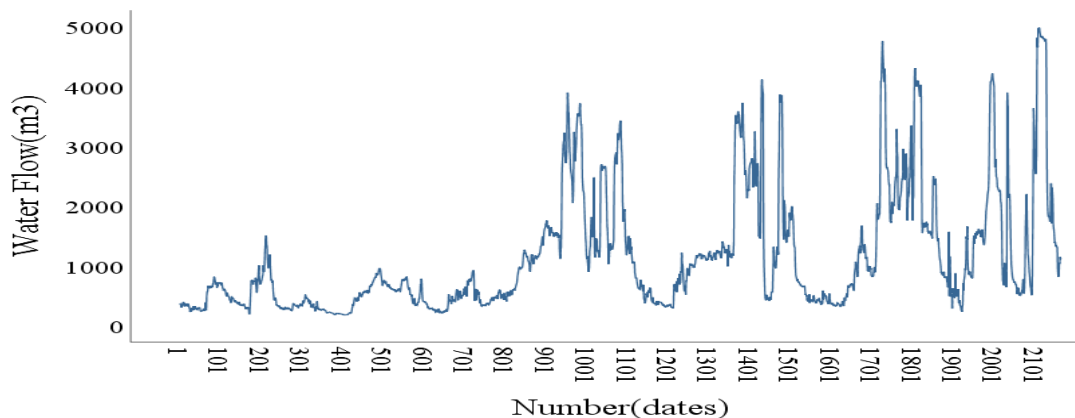


Figure 6. Time series chart of water flow

From these three time series charts, the overall changes in sediment content, water flow, and water level in the lower reaches of the Yellow River over a period of 6 years can be obtained. Moreover, it can be roughly seen from the charts that sediment content, water flow, and water level exhibit an annual periodicity. However, it can be observed that the three have undergone significant changes in numerical values at different times and in a relatively short period of time compared to before. To explore the specific timing of this sudden and significant change, using Mann-Kendall mutation test may be one of the better methods [10].

### 3.2. M-K mutation test model's building and visualization

Man-Kendall test is a nonparametric statistical test method developed by scholars Mann and Kendall. For time series  $X$  ( $X$  contains  $k$  samples), constructing a rank sequence shown in the formula (6)-(8) [11].

$$S_k = \sum_{i=1}^k r_i (k = 1, 2, 3, \dots, n) \quad (6)$$

$$r_i = \sum_{j=1}^i \text{sgn}(x_i - x_j) \quad (7)$$

$$\text{sgn}(x_i - x_j) = \begin{cases} +1, & x_i > x_j \\ 0, & \text{otherwise} \end{cases} (j = 1, 2, \dots, i) \quad (8)$$

The rank sequence  $S_k$  is the accumulation of the number of values when the value at time  $i$  is greater than time  $j$ . Defining statistics  $UF_k$  under the assumption that the time series is random, which is shown in the formula (9).

$$UF_k = \frac{[S_k - E(S_k)]}{\sqrt{\text{Var}(S_k)}} \quad (9)$$

In the formula (9),  $UF_1 = 0$ ,  $E(S_k)$ ,  $\text{Var}(S_k)$  are the mean and variance of  $S_k$ , respectively. When  $x_1, x_2, \dots, x_n$  are independent of each other, they have the same continuous distribution. And the formula(10) can be inferred.

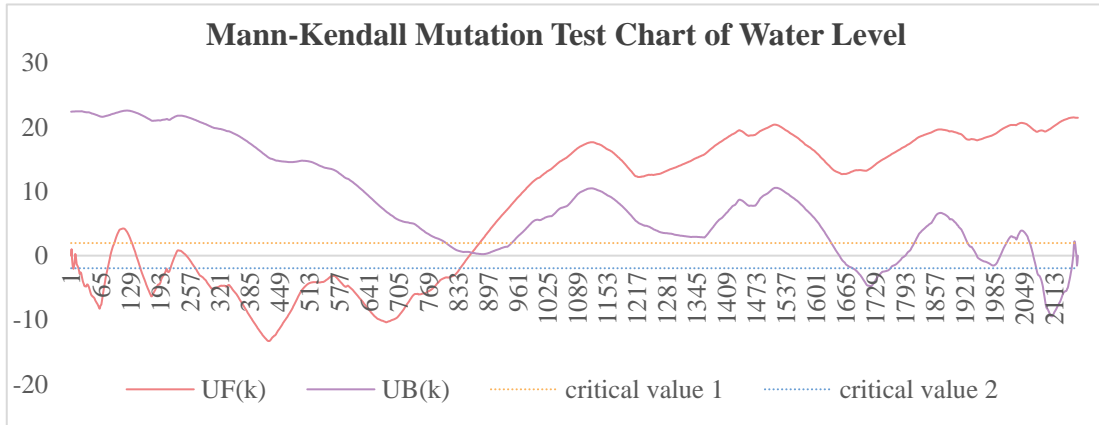
$$E(S_k) = \frac{k(k-1)}{4}, \text{Var}(S_k) = \frac{k(k-1)(2k+5)}{72} (1 \leq k \leq n) \quad (10)$$

$UF_k$  is a statistical sequence calculated in the order of time series  $X$  ( $x_1, x_2, \dots, x_n$ ). Given the significance level  $\alpha$ , query the normal distribution table. If  $UF_k > UF_\alpha$ , it indicates that there is a significant trend change in the time series. Repeating the above operation in reverse order of the time series ( $x_n, x_{n-1} \dots x_1$ ) to determine  $UF'_k$ . Then arranging in reverse order again and take the opposite number  $UB_k, UB_n = 0$ .

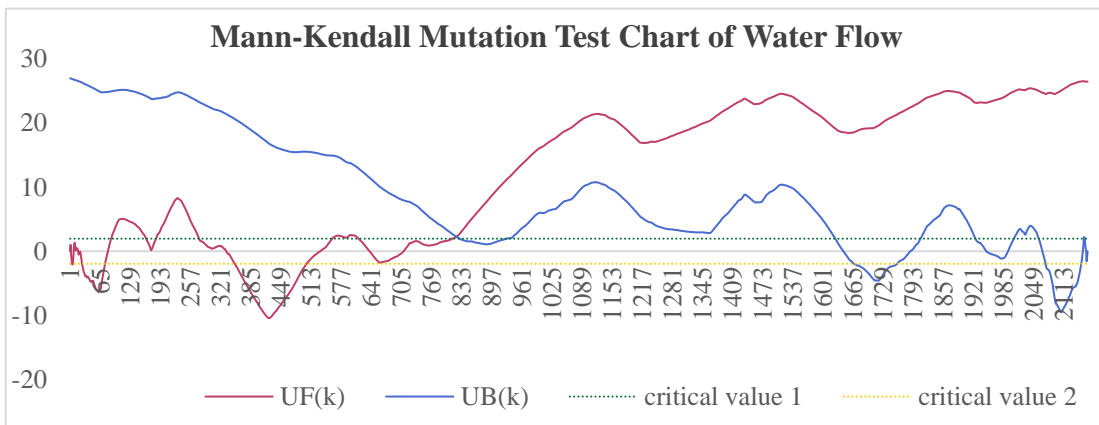
Moreover, plotting the sequence curves of  $UF_k$  and  $UB_k$  statistics, as well as the upper and lower critical values, on the same graph. If the value of  $UF_k$  is greater than 0, it indicates an upward trend in the sequence, and if it is less than 0, it indicates a downward trend. When it exceeds the critical

line, the surface shows a significant upward or downward trend, and the range beyond the critical line is determined as the time zone where a mutation occurs. If the  $UF_k$  and  $UB_k$  curves intersect and the intersection point is between the critical lines, then the time corresponding to the intersection point is the time when the mutation begins.

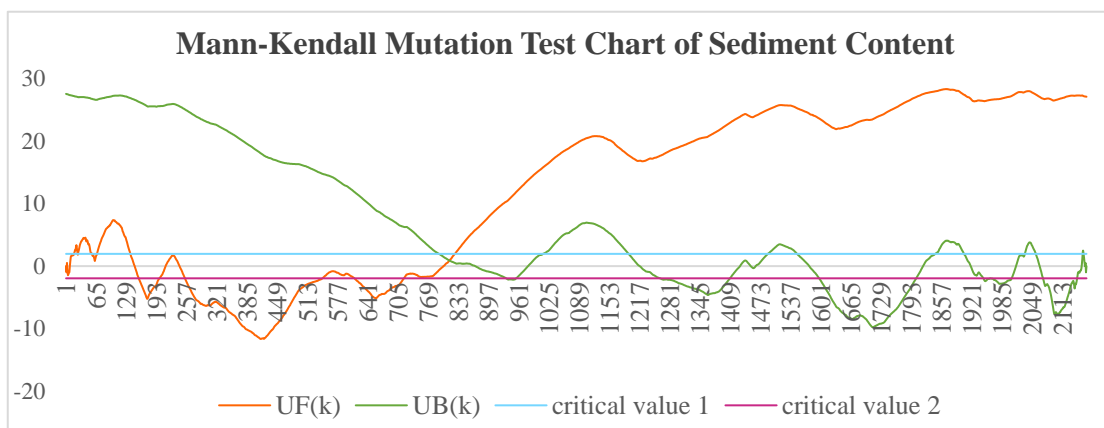
After obtaining all data sets in the M-K test, making the M-K test graph shown in the Fig 7-9. Meanwhile, choosing significance level  $\alpha = 0.05$ , which is widely used in the academic circle, and the critical value is  $UF_{0.05} = \pm 1.96$ .



**Figure 7.** Mann-Kendall Mutation Test Chart of Water Level



**Figure 8.** Mann-Kendall Mutation Test Chart of Water Flow



**Figure 9.** Mann-Kendall Mutation Test Chart of Sediment Content

It can be observed that the  $UF_k$  curves and  $UB_k$  curves of sediment content, water level, and water flow intersect within the critical values and between counting points 800 to 900, indicating mutations

during this period. By searching for various data in the M-K test table, it was found that the time of the mutation of sediment content is at counting point 814, on March 7, 2018, the time of the mutation of water flow is at counting point 823, on March 16, 2018, and the time of the mutation of water level is at counting point 863, on April 26, 2018.

Furthermore, the sequence of mutation occurring is, sediment content, water flow, and water level. The time interval between mutations of sediment content and water flow is one week, and the time interval between mutations of water flow and water level is one month. Therefore, it can be considered that when the sediment content increases sharply, the water flow will increase sharply in about a week, and the water level will rise in about a month. This is of great significance for flood control and sand prevention, water conservancy projects, and coordinating the relationship between humans and the environment. For example, before the flood season of the Yellow River approaches, when hydrological monitoring points detect a sharp increase in sediment content, considering that the water flow will rise a week later, and the water level will also rise a month later, relevant government personnel should check water conservancy projects and the facilities of hydropower stations for normal operation within a week, then take flood prevention measures and alarm nearby residents to prevent floods and the inconvenience caused by rising water levels within a month.

#### 4. Conclusion

The trend of combination of micro and macro analysis to analyze dynamic change processes is developing rapidly in academic circle. This paper combines linear regression and M-K mutation test to analyze hydrology of Yellow River. The linear regression establishes a functional relationship between sediment content, water level, and water flow from a micro perspective, and ensures that all the data from each monitoring points can be linked to other hydrological variables to infer other variables's possible values and make targeted strategies to prevent disasters. The M-K mutation test reveals the changes in the middle and lower reaches of the Yellow River in the past 6 years (2016-2021) from a macro perspective, and the mutations of the sediment content, water flow, and water level occurred on March 7, 2018, March 16, 2018, and April 26, 2018, respectively, which implies that in the first half of 2018, the hydrology of the Yellow River entered a new historical cycle, with three indicators showing significant increases compared to before. For the analysis of current hydrology, the combination of micro layer and macro layer analysis can not only cope with short-term rainstorm and other natural phenomena, but also provide long-term solutions for river management. These conclusions have important reference significance for the overall macro regulation of Yellow River hydrology, the overall trend of Yellow River prevention and control, and future countermeasures. More broadly considering, this kind of analysis can be promoted and applied to other rivers similar to the Yellow River, and also provide some ideas for local researchers and government agencies to solve river problems.

#### References

- [1] Li Xiandong, Liu Shixiang, Liu Zhejie. Analysis of sediment discharge process in Xiaolangdi Reservoir from 2021 to 2023 [J]. People's Yellow River, 2024, 46 (S1): 11 - 12+16.
- [2] Qi Xiaodong. Research on Improved Algorithm for Time Varying Correlation Estimation and Its Application in Hydrological Data Analysis [D]. Dalian Jiaotong University, 2021.
- [3] Jia Guoqing, Cheng Minglong, Yi Huiyue, et al. Research on SSA Algorithm Based on M-K Test and Least Squares Method [J]. Journal of Jiamusi University (Natural Science Edition), 2022, 40 (05): 29 - 33+61.
- [4] Ding Ran. Analysis and Improvement of Error Causes in Small Flow Hydrological Forecasting of Manshui River in Huoshan County [J]. Huai River Management, 2024, (05): 9 - 11.
- [5] Sarina, Sun Yingna, Chen Shi. Analysis of trend changes in root river runoff based on Mann Kendall test method [J]. Gansu Water Resources and Hydropower Technology, 2024, 60 (02): 15 - 19.
- [6] Zhang Yaoxin, Yao Chuanhui, Wang Gang. Analysis of precipitation in Tai'an based on M-K method and SPI index [J]. Haihe Water Resources, 2022, (04): 97 - 100.

- [7] Liu Changming, Tian Wei, Liu Xiaomang, et al. Analysis and Understanding of Changes in Yellow River Runoff over the Past Century [J]. People's Yellow River, 2019, 41 (10): 11 - 15.
- [8] Zhou Yongjian, Feng Wenkai, Yi Xiaoyu, et al. Quantitative analysis of the correlation between deformation of Muyubao landslide and reservoir water level [J]. Studying Technology and Engineering, 2021, 21 (27): 11544 - 11550.
- [9] Wei Juan. Prediction and Analysis of Spring Water Flow Trends of Major Rivers in Bazhou in 2021 [J]. Heilongjiang Water Conservancy Science and Technology, 2024, 52 (02): 84 - 86+113.
- [10] Zhang Yaixin, Tan Xiucui. Analysis of the trend and abrupt change of water level and storage capacity in Dongping Lake in the past 13 years [J]. Haihe Water Resources, 2022, (01): 77 - 80.
- [11] Wang Jiao, Jiang Rengui, Jie Jiancang, et al. Study on the Characteristics of Runoff Changes in the Jinghe River Basin [J]. Journal of Water Resources and Water Engineering, 2018, 29 (06): 88 - 93.