ISSN 2616-5872 Vol.6, Issue 2: 59-65, DOI: 10.25236/AJEE.2024.060208

Data-Driven Analysis of Water and Sediment Flux at the Xiaolangdi Station of the Yellow River

Kun Zhang*, **Yuanmeng Li, Mingxuan Jiang, Yi Jin**

Luoyang Polytechnic, Luoyang, 471000, China
**Corresponding author: zhangkun9595@163.com*

Abstract: *The study of the change rule of the water and sand flux of the Yellow River has an important impact and significance on the environmental management, people's life, and climate change along the Yellow River basin, and also has an important theoretical guiding significance for optimizing the water resources allocation, water and sand transfer, flood prevention and disaster mitigation in the Yellow River basin. This paper analyzes and researches the water and sand fluxes as well as the change trends of a hydrological station downstream of Xiaolangdi Reservoir on the Yellow River. The main purpose is to use the actual monitoring data of water level, water flux, and sand content of a hydrological station downstream of the Yellow River Xiaolangdi Reservoir for the past 6 years, pre-process the data using MATLAB, process the missing values or abnormal values, and set up mathematical models to calculate the long-term trend of the sand content of the Yellow River, the total water flux and the total sand discharge of the hydrological station of the Yellow River in the past 6 years, as well as to predict the dynamic change of the water-sand flux and the change trend of the sand content of the Yellow River based on the calculated and analyzed data. The dynamic change of water and sand fluxes and the situation of sudden change, seasonal and cyclic patterns of sand content can set up a reasonable monitoring program for the management of the Yellow River. The research in this paper can provide reliable predictions in a short time, which can be generalized to river management and water resource allocation, and has important reference value for future river management.*

Keywords: *Yellow River water and Sediment flux, water and sand monitoring, change trend, mathematical model, time series prediction model*

1. Introduction

The water and sand content of the Yellow River is affected by natural and anthropogenic factors. With global climate change and human activities, the water-sand content of the Yellow River changes, which has a significant impact on the environment, climate and people's life in the Yellow River basin. Therefore, the monitoring of the water sand content of the Yellow River is of great significance to understand the water resources of the Yellow River and the prevention of water and sand disasters. The monitoring of Yellow River water and sand flux is an important means to understand the water resources of the Yellow River and the water and sand disaster, and the monitoring of Yellow River water and sand flux can be carried out by means of hydrological stations and remote sensing technology. The hydrological station collects water samples from the Yellow River, measures its sand content and water level, and provides real-time data to provide a basis for flood prevention and disaster reduction and water resource management.

Many scholars at home and abroad have done a lot of research on the hydrological situation of the Yellow River. Chen Yongche and Cui Yanhong [1] from Tsinghua University studied the evolution of vegetation and hydrological elements in the Inner Mongolia section of the Yellow River Basin under the changing environment, and put forward suggestions for the ecological protection and restoration and optimal allocation of water resources in the Inner Mongolia section of the Yellow River Basin. Fan Yongyong and Cui Wenjuan [2] studied the impact of the 2020 Yellow River water dispatching on the hydrological characteristics of the estuary and the sedimentation effect. Based on a large number of hydrological measurements and topographic data in 2020, they analyzed the changing pattern of temperature and salinity distribution in the spatial and temporal distribution of the sea area of the Yellow River Estuary and the evolution of the siltation characteristics of the underwater topography and explored the effect of the Yellow River inflow of seawater sands on the salinity distribution of the sea area of the Yellow River Estuary and the underwater topography. Wu Li and Song Baihui [3] took Xiaolangdi

ISSN 2616-5872 Vol.6, Issue 2: 59-65, DOI: 10.25236/AJEE.2024.060208

Reservoir as the research object and studied the effect of water and sand regulation on dissolved organic matter in the water body of the reservoir area. An Xindai [4] proposed the role of reservoir silt reduction based on the analysis of Xiaolangdi Reservoir and downstream river sediment washout, calculated the reservoir flood control and silt reduction benefits, and the results showed the reasonableness of the construction of the Yellow River water and sand regulation backbone project.

The data used in this paper come from the official website of the China University Students Mathematical Modeling Contest(www.mcm.edu.cn), based on the research of some scholars, mainly analyzed the trend of water and sand fluxes and the optimal monitoring scheme of Xiaolangdi Reservoir of the Yellow River, calculated the total water fluxes and total sand discharges of Xiaolangdi Hydrological Station of the Yellow River in the recent 6 years, and predicted the dynamic changes of water and sand fluxes based on the calculated and analyzed data, set a reasonable monitoring scheme and predicted the changes of riverbed elevation in the next 10 years, and set a reasonable monitoring scheme. A reasonable monitoring program is set up and the change of the riverbed elevation of the Yellow River in the next 10 years is predicted.

2. Yellow River water and sand data visualization and analysis

2.1 The structure of the BP neural network

The hydrological station provides real-time data by collecting water samples from the Yellow River, measuring its sand content and water level, and providing a basis for flood prevention disaster mitigation, and water resource management. A hydrological station located in the Yellow River downstream of Xiaolangdi Reservoir recorded the measurement data of water level, water flow, sand content, and Yellow River section for the last 6 years, First, the relationship between date, time, water level, flow and sand content contained in the data needs to be analyzed, and the problems included are the relationship between sand content and time, the relationship between sand content and water level, and the relationship between sand content and water flow, and then the annual average total water flow and annual average total sand discharge of this hydrological station were estimated for the last 6 years. Secondly, to estimate the annual average total water flow and annual average total sand discharge of the hydrological station in the past 6 years, the main problem is to calculate the total sand discharge and total water flow, which can be theoretically obtained through the calculation of water flow and sand content. Therefore the focus is still on analyzing the situation between annual total water flow and total sand content. This paper needs to consider the effect of different data collection methods and time scales on the results. A common approach is to use a multiple linear regression model in which water level and water flow are used as independent variables and sand content is the dependent variable. Such models can be used to estimate the extent to which water level and water flow affect sand content. Finally, to better understand the relationship between water level, water flow, and sand content, visualization charts such as scatter plots and regression lines can be drawn. These graphs can visualize the distribution and trend of the data and help to interpret the results of the analysis. Several aspects such as data collection, model selection, time series analysis, and visualization of results need to be considered to fully understand the relationship between hydrological data. This can help to predict sand content more accurately and provide important insights into water resource management and ecosystem protection.

However, to account for possible non-linear relationships, more advanced machine learning algorithms, such as random forests or deep learning models, can be used to construct more accurate predictive models. These algorithms can capture complex data patterns and interactions, which improves the predictive power of the model. In addition, the dynamics of these data can be studied through time series analysis methods. Time series analysis can help us identify seasonal and cyclical patterns, which are critical to understanding the changing patterns of hydrologic data. These patterns can be revealed through the analysis of autocorrelation functions and partial autocorrelation functions.

In the research process of this paper, to construct the mathematical model and solve the problem, the sudden changes caused by special typhoons, geological disasters, and other force majeure factors are not taken into account, and it is recognized that the basic policy of "water transfer and sand regulation" of Xiaolangdi Reservoir remains unchanged during the research period. Pre-processing of the source data of this paper, in which the sand content data may contain some noise, first of all, needs to be smoothed, with some statistical methods to remove random fluctuations in the data on the sand content of the impact of the data, the amount of data uniform and accurate. y_t The data volume is unified and accurate to the precision in days. The linear regression analysis of the processed data can produce the data visualization of water level, flow rate, and sand content with time as the independent variable as shown in Fig. 1, and

ISSN 2616-5872 Vol.6, Issue 2: 59-65, DOI: 10.25236/AJEE.2024.060208

the further analysis of sand content and its trend over time as shown in Fig. 2.

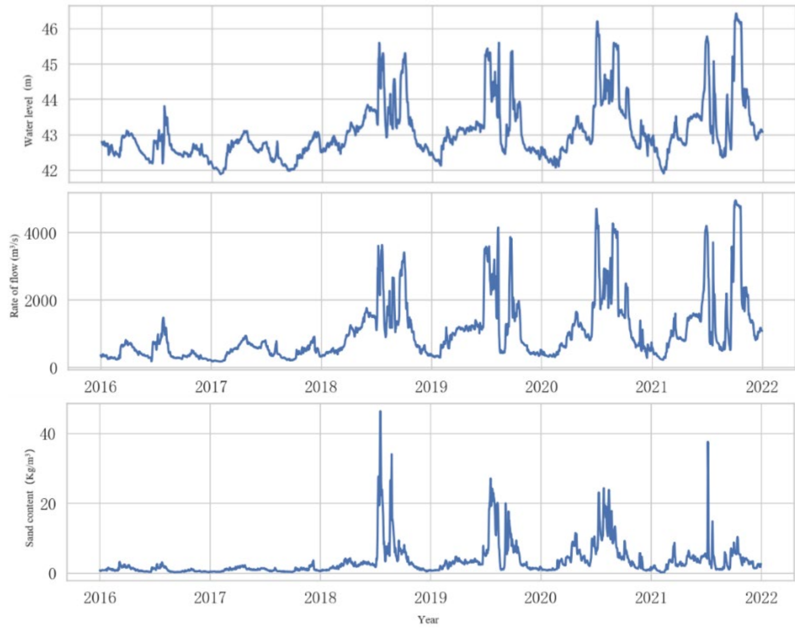

Figure 1: Trends in water level, flow, and sand content over time

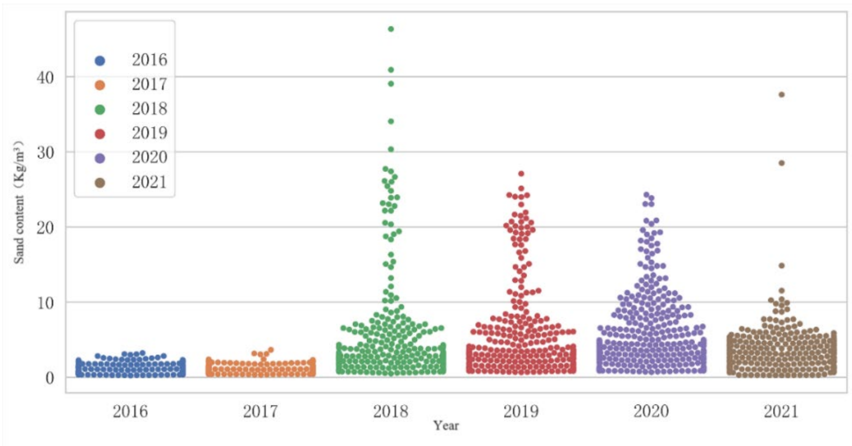

Figure 2: Trends in sand content over time years

The change in the sand content characteristics of the Yellow River can be found through the analysis of the visualization map, and the sand content is generally high from 2018 to 2021.

3. Time series investigation of water and sand fluxes in the Yellow River

3.1 Linear regression modeling of sediment content versus water level and water flow rate

In the linear regression model, the one-way linear regression model for the dependent variable y and the independent variable x is shown in Formula 1.

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1}$$

Among them. β_0 is the regression constant, and β_1 is the regression coefficient of $\varepsilon \sim N(0, \delta^2)$. The regression equation is that.

$$y = \hat{\beta}_0 + \hat{\beta}_1 x \tag{2}$$

In analyzing and calculating the relationship between sand content and water level, assuming that y denotes the sand content (Kg/m^3) and x_1 denotes the water level (m), the double logarithmic regression model for y and x_1 is.

ISSN 2616-5872 Vol.6, Issue 2: 59-65, DOI: 10.25236/AJEE.2024.060208

$$lny = \beta_0 + \beta_1 lnx_1 + \varepsilon \tag{3}$$

Using MATLAB software, given a significance level of 0.05, parameter estimation was carried out using the least squares method, and the results are shown in Table 1.

Table 1: Parameter estimates and their test results

ratio	estimated value	The lower limit of 95% confidence interval	The upper limit of 95% confidence interval
β_0	-135.95	-140.98	-130.92
β_1	36.31	34.97	37.64
$R^2 = 0.5979, P= 0.0000,\ S_y = 0.31$			

As can be seen from Table 1, the 95% confidence intervals of the 2 regression coefficients do not contain 0, so the regression coefficient test meets the requirements. The goodness of fit$R^2 = 0.5979$, indicating that the fitting accuracy meets the calculation requirements, and the companion probability of F-test P=0.0000<0.05, indicating that the sand content and water level $linx_1$,The linear relationship between sand content and water level was significantly established. So the regression equation can be determined as.

$$lny = -135.95 + 36.31 lnx_1,\ \ x_1 \in (0,\ \ +\infty) \tag{4}$$

Based on equation (4), the following conclusions can be drawn.

(1) Sand content and water level are positively correlated.

(2) The sand content increased by 36.31% for every 1% increase in water level, and the sand content was very sensitive to the change in water level.

Using the same method to analyze and calculate the relationship between sand content and water flow rate, assuming that y denotes the sand content (Kg/m^3) and x_2 denotes the water flow rate (m^3 /s), the regression equation can be determined as follows.

$$lny = -6.35 + 1.06 lnx_2,\ \ x_2 \in (0,\ \ +\infty) \tag{5}$$

Based on equation (5), the following conclusions can be drawn.

(1) Sand content and water flow are positively correlated.

(2) For every 1% increase in water flow, the sand content increases by 1.06%, and the sand content is not sensitive to the cross-fertilization of water level.

3.2 Calculation of total annual water flow and total sand content

Let x_2 denote the water flow (m^3 /s) at the time (day) t of a certain year, and z denote the total water flow (m^3) of the year, then the total water flow of a certain year is.

$$z = \frac{365 \times 24 \times 3600}{365 \times 10^8} \sum_{t=1}^{365} x_t = 0.000864 \sum_{t=1}^{365} x_t \tag{6}$$

Let $_{it}$ denote the sand content (Kg/m^3) at time t (day), and M denotes the total sand content (tons) in each year, then the total sand content in that year is.

$$M = \frac{\sum_{t=1}^{365} y_t z_t}{10^{11}} \tag{7}$$

The results of the calculation of total annual water flow and total annual sand discharge are shown in Table 2.

ISSN 2616-5872 Vol.6, Issue 2: 59-65, DOI: 10.25236/AJEE.2024.060208

Table 2: Total annual water flow and total annual sand discharge

vintages	2016	2017	2018	2019	2020	2021	average value
Total flow (billion m^3)	142.13	151.36	383.04	381.20	424.24	466.61	324.76
Total sand content (billion tons)	0.16	0.18	2.92	3.02	3.41	2.26	1.99

3.3 Changing patterns of water and sand fluxes

According to the change rule of water and sand flux, we can predict the change of water and sand flux in the future, and according to the prediction result, we can grasp the dynamic change of water and sand flux in the Yellow River and design a reasonable monitoring program.

For analyzing the change rule of water and sand flux in the hydrological station in the past 6 years, the following algorithm is used:

(1) ADF test

ADF Test (Augmented Dickey-Fuller Test) Algorithm: the ADF test is used to detect whether the time series data is smooth or not. It checks for the presence of unit roots in the data. The algorithm can test the smoothness of the time series and can determine if differencing or other preprocessing steps are required for further time series analysis. If the time series is not smooth, steps can be taken to make it smooth. The ADF test determines smoothness by checking for the presence of a unit root in the time series. Specifically, the ADF test assumes an original hypothesis (null hypothesis) that the time series has a unit root, indicating that the data is not smooth. If the result of the test rejects the null hypothesis, it implies that the time series is smooth. Otherwise, if the null hypothesis is not rejected, differencing or other preprocessing steps are required to smooth the time series. The application of the ADF test helps to determine the steps required for further time series analysis and provides important information about the smoothness of the data, as shown in Table 3 and Table 4.

Table 3: Statistical results of the ADF test

descriptive statistics							
Series	N statisticians	Realm statisticians	Minimum value	Maximum value	average value		standard deviation
					statisticians	standard error	
water level	13381	4.36	41.88	46.24	43.2037	0.0080	0.9268
Rate of flow	13381	4791	179	4970	1188.8892	8.9778	1038.5196
Sand content	13381	48.29	0.01	48.3	5.0557	0.0465	5.3750
flux	13381	164700.23	2.77	164703	11114.3096	165.0503	19092.4049

Table 4: Component analysis table

Series	ADF statistic	p-value	1%threshold	5% threshold	10%threshold	Whether smooth (p <= 0.05)
water level	-4.24665	0.00055	-3.43084	-2.86176	-2.56689	True
Rate of flow	-4.22715	0.00059	-3.43084	-2.86176	-2.56689	True
Sand content	-4.32463	0.00040	-3.43084	-2.86176	-2.56689	True
flux	-5.54186	0.00000	-3.43084	-2.86176	-2.56689	True

These are the results of the ADF (Augmented Dickey-Fuller) test for determining whether time series data has a unit root, i.e. whether it is smooth. The following is a detailed analysis of these results:

1) ADF test results for water level.

The ADF statistic is -4.2467. p-value is 0.0005, which is less than the significance level of 0.05, indicating that the original hypothesis is rejected, suggesting that the water level data is smooth. The critical values are 1%, 5%, and 10% which are less than the ADF statistic and also support that the water level data is smooth.

2) ADF test results for water flow.

ISSN 2616-5872 Vol.6, Issue 2: 59-65, DOI: 10.25236/AJEE.2024.060208

The ADF statistic is -4.2272. p-value is 0.0006, which is less than the significance level of 0.05, indicating that the original hypothesis is rejected, suggesting that the flow of data is smooth. The critical values are 1%, 5%, and 10% which are less than the ADF statistic and also support that the flow data is smooth.

3) ADF test results for sand content.

The ADF statistic is -4.3246. p-value is 0.0004 which is less than the significance level of 0.05 indicating that the original hypothesis is rejected indicating that the sand content data is smooth. The critical values of 1%, 5%, and 10% are less than the ADF statistic, which also supports that the sand content data is smooth.

4) ADF test results for water-sand fluxes.

The ADF statistic is -5.5419. p-value is very small, close to zero (1.69e-06), and much less than the significance level of 0.05, indicating that the original hypothesis is rejected, suggesting that the water-sand flux data is smooth. The critical values of 1%, 5%, and 10%, respectively, are all less than the ADF statistic and also support that the water-sand flux data are smooth.

(2) Time series algorithms

Time series plot algorithm: a time series plot of water and sand fluxes using the matplotlib library is shown in Figure 3. The algorithm can time series plot can visualize the fluctuations and trends in the data and help in the initial characterization of the data. This helps in identifying seasonal and trend components. So that the characteristics of the data can be better understood. The time series plot can help to get a preliminary understanding of the fluctuations, trends, and seasonal patterns of the data.

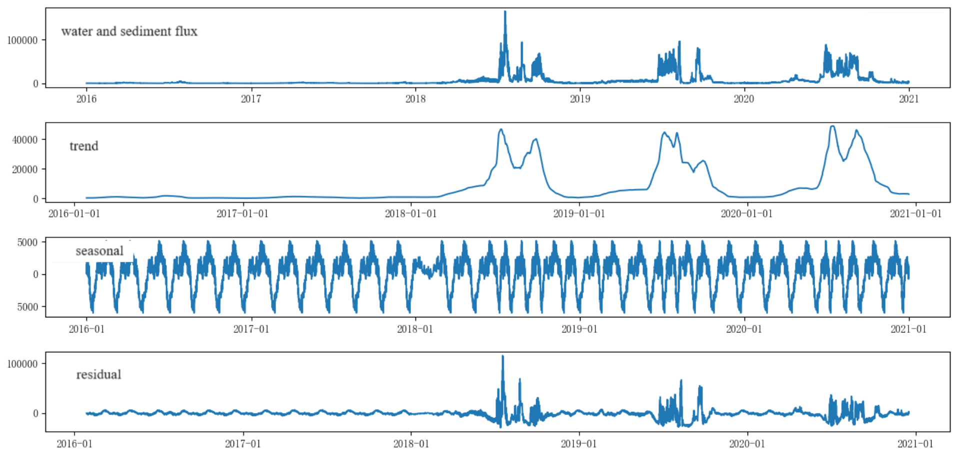

Figure 3: Time series plot of water-sand fluxes

The analysis of the water-sand flux curve shown in the figure shows that the water-sand flux curve shows seasonal variations during the year. During the flood season, the water-sand flux value is higher, which means that the amount of sediment carried in the river is higher. While in the dry season, the water-sand flux value is lower. Analyzing the residual curves shows that the residual curves also show similar seasonal variations. During the flood season, the water level and flow rate varied more due to the influence of flowing sand and the topography of the river bottom. In contrast, these variations are relatively small during the dry season.

4. Conclusions

This paper takes a hydrological station of the Yellow River downstream of Xiaolangdi Reservoir as the research object, calculates the water-sand flux of the Yellow River by establishing a mathematical model, as well as analyzes and researches the change rule of the water-sand flux.

The seasonal variation of water and sand fluxes in the Yellow River can be seen from the study of this paper, and the water and sand fluxes show obvious differences in different seasons of the year. In general, the water-sand flux is higher in the flood season and lower in the dry season. This seasonal variation is related to factors such as rainfall, temperature, and evaporation. During the flood season, the flow and velocity of the river increase, so the water-sand flux increases accordingly. The water-sand flux

ISSN 2616-5872 Vol.6, Issue 2: 59-65, DOI: 10.25236/AJEE.2024.060208

of the Yellow River also shows cyclic changes, which may be caused by cyclic changes in the river's water level, flow rate, and changes in sand content. In general, the changing pattern of water-sand flux is that the flood season usually occurs between May and July. During this period, the water level and flow velocity of the river increase, and the water-sand flux increases accordingly. During the dry season, which usually occurs between November and April, the water level and flow velocity of the river decrease, and the water-sand flux decreases accordingly.

The research in this paper can be able to provide reliable predictions in a short period, which can be generalized to river management and water allocation, and will be an important reference value for river management in the future.

References

[1] Chen Yongche, Cui Yanhong, Zhang Caijin, et al. Spatial and temporal co-evolution of vegetation and hydrological elements in the Inner Mongolia section of the Yellow River Basin under changing environment [J]. Water Resources Conservation, 2024,(01)1-15.

[2] Fan Yongyong, Wu Wenjuan, Bi Naisheng et al. Impacts of Yellow River water dispatching on estuarine hydrological characteristics and depositional effects in 2020[J]. Frontiers of Marine Geology, 2023, 39(10):43-51.

[3] Wu Li, Song Baihui, Yang Meng et al. Effects of water-sand regulation on dissolved organic matter in Xiaolangdi Reservoir of the Yellow River[J]. Journal of Environmental Science, 2023,43(03):216-225.

[4] An Xindai. Tracking analysis of flood control and siltation reduction of Xiaolangdi Reservoir on the Yellow River[J]. People's Yellow River, 2023,45(08):37-42.

[5] Zhou Yan, Wei Xiaowen. Exploring the Yellow River water and sand regulation model to better meet the Yellow River governance and protection requirements[N]. China Water Resources News, 2022-12-08(005).