Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 6, No. 1, 2025

Optimisation of Water and Sand Flux Evolution and Monitoring in the Lower Reaches of the Yellow River Xiaolangdi Based on Multivariate Statistics and Time Series Analysis

Junpeng Yuan[1, *, #], Hang Zeng[2, #], Yiting Qiu[1, #], Dang Gan[1, #], Yun Huang[1, #], Chentong Wen[1, #], Jiongyi He[1, #], Jingmin Lan[1, #], Tingyan Wang[1, #], Manni Li[1, #]

[1]School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China

[2]School of Economics and Statistic, Guangzhou University, Guangzhou 510006, China

[*] **Corresponding author**: Junpeng Yuan (Email: 13257516286@163.com)

[#]**These authors contributed equally.**

Abstract: In this paper, the water and sand monitoring data of a hydrological station in the lower reaches of Xiaolangdi of the Yellow River for the past 6 years are studied in depth based on multivariate statistics and time series analysis. By constructing multiple linear regression model, Pearson correlation analysis, sliding mean, fast Fourier transform (FFT) and SARIMA model, the relationship between sand content and water level and flow was systematically analysed, the sudden change, seasonal and cyclic characteristics of water-sand flux were revealed, and the water-sand flux in the next two years was predicted, and the optimised sampling and monitoring scheme was proposed. The results provide strong theoretical support and practical guidance for the scientific management and monitoring of water-sand fluxes in the lower reaches of the Yellow River, and help to improve the ecological management and resource allocation capacity of the Yellow River Basin. For the relationship between sand content, water level and flow at a hydrological station on the Yellow River, this paper establishes a multiple linear regression model and quantifies the relationship between the variables through Pearson correlation analysis. The results show that the flow rate is weakly and positively correlated with the sand content (Pearson correlation coefficient is about 0.25-0.3), and the water level has less influence on the sand content. The total annual water flow and total annual sand discharge at this hydrological station were further estimated from 2016 to 2021 using the time series aggregation method, with the total annual water flow fluctuating between 1.98 and 2.72 billion cubic metres (mcm), and the total annual sand discharge varying between 0.98 million tonnes (mt) and 1.95 million tonnes (mt). These results provide basic data support for subsequent water resource management and water-sand flux studies.

Keywords: Gravity Model, In-depth Learning, The Belt and Road, Route Planning.

1. Introduction

The Yellow River is regarded as the mother river of the Chinese nation, and the dynamic changes of its water and sand fluxes not only directly affect the environmental management and climate change along the Yellow River Basin, but also are closely related to people's life, water resource allocation in the Yellow River Basin, the coordination of human-land relationship, and the major projects, such as water and sand transfer and flood control and disaster mitigation. The quantitative study of the changing law of water and sand fluxes in the main hydrological stations of the Yellow River has very important theoretical and practical values for the scientific formulation of basin management measures and the enhancement of ecological security and governance capacity. Based on the water level, water flux, sand content, cross section and monitoring point data of a hydrological station of the Yellow River from 2016 to 2021 provided by the title, this paper establishes mathematical models and data analysis methods around the following three core issues: Based on the provided actual monitoring data of water level, water flow rate and sand content at a hydrological station on the Yellow River in the past 6 years, the interrelationships between sand content and time, water level and water flow rate are explored, and the corresponding mathematical model is constructed to quantitatively describe such relationships[1]. On this basis, the model is further used to estimate the total annual water flow and total annual sand discharge of the hydrological station during the six-year period, which will provide basic data support for subsequent water resources management and water-sand flux studies[2].

2. Research on the relationship between water level and sediment concentration

The purpose of this problem is to systematically portray the statistical relationship between "water level", "water flow" and "sand content" in a hydrological station of the Yellow River in the past 6 years, and to estimate the total water flow and total sand discharge in the past years. The core idea is to estimate the total water flow and total sand discharge in the past years through statistical regression[3]. The core idea is to quantify the effects of water level and flow on sand content through statistical regression and correlation analyses, and to estimate the annual water and sand fluxes and reveal the trends by using the time series aggregation method.

The following mathematical models were developed to fully characterise the links between the variables:

a) Multiple linear regression model

Multiple linear regression relies on the least squares method, which is a classical statistical method for portraying and predicting the strength of the influence of a set of independent variables (e.g., water level and flow) on a dependent variable (e.g., sand content), and is used to quantify the independent influence of flow and water level on sand content, which is conducive to the understanding of the dynamical principal controlling factors of the sand transport in the river.

Mathematical expression:

$$C=\beta_0+\beta_1 Q+\beta_2 H+\varepsilon \quad (1)$$

Where C denotes sand content; Q denotes flow; H denotes water level; β_0 denotes intercept term; β_1 and β_2 denote regression coefficients; and ε denotes random error term.

b) Pearson correlation analysis

Theoretical basis: Pearson's correlation coefficient is used to measure the strength of linear correlation between two (or more) continuous variables, and is used to quantitatively analyse the synchronicity and coupling characteristics between variables such as water level and flow rate, flow rate and sand content.

Mathematical expression:

$$r_{XY}=\frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad (2)$$

r_{XY} denotes the Pearson correlation coefficient between variables (X) and (Y); Cov(X, Y) denotes covariance; σ_X and σ_Y denote standard deviation.

c) Modelling of total annual water flow and total annual sand discharge estimation

Integrate daily or high-frequency flow with the product of sand content to obtain the total amount for the desired time period. The total annual/seasonal water and sand transport is obtained intuitively, which facilitates the analysis of basin resource management and management countermeasures.

Mathematical expression (discretisation):

$$Q_{year}=\sum_{i=1}^{n} Q_i \Delta t_i \quad (3)$$

$$S_{year}=\sum_{i=1}^{n} Q_i \times C_i \times \Delta t_i \quad (4)$$

Q_{year}: total annual water flow; S_{year}: total annual sand discharge; Q_i: flow in the (i) time period; C_i : sand content in the (i) time period; Δt_i : length of the (i) time period.

2.1. Solving the model

(1) Algorithm selection

The multiple linear regression model uses the least squares method for parameter estimation, because this algorithm can efficiently know the primary and secondary relationships between variables. Pearson's correlation coefficient was chosen for correlation analysis, which is widely used in linear correlation analysis and is easy to interpret. The annual totals are estimated using the direct accumulation method, which can reflect the overall changes in each year.

Detailed steps

a) Data standardisation and field correction

Ensure that the units of variables are uniform (e.g., water level, flow rate and sand content are all standardised), and correct the outliers or extreme values of the original data.

b) Correlation analysis

Calculate the Pearson correlation coefficients between water level[4-6], flow rate and sand content, generate the correlation matrix and visualise it. - Fit multiple linear regression, obtain the regression coefficients of the respective variables on sand content, and judge the significance.

c) Time series statistics

Aggregate flows and sand discharges by year and month, generate trend lines and yearly/monthly averages - to facilitate identification of seasonal periodicity and extreme event effects.

d) Visualisation output

Generate heat maps of variable relationships, scatter plots[5], trend lines with key statistical Table 1.

Table 1 Summary of Estimated Water Flow and Sediment Discharge from 2016 to 2021

Year	Total water flow (billion m³)	Total sand discharge (million tonnes)
2016	22.3	118
2017	23.9	125
2018	27.2	195
2019	20.4	110
2020	19.8	98
2021	21.1	104

3. Conclusions

Through the above model analysis, the linear correlation between "water level - flow rate - sand content" at a hydrological station on the Yellow River in the past 6 years is significant, in which the flow rate and water level are highly synchronous (Pearson's coefficient>0.95), and the sand content is weakly and positively correlated with the water level/flow rate (the coefficient is about 0.25-0.3).

Linear regression results showed that there was a slight tendency for sand content to increase when flow increased, which was particularly affected by external sources (e.g., water and sand transfers or sudden rainfall).

The overall fluctuation of total annual flow and sand discharge was not significant, but the unusually high sand discharge in 2018 may be closely related to extreme weather events or artificial interventions by upstream sand transfer.

This analysis provides basic data support and trend basis for subsequent river management, water resource allocation and flood control decision-making; at the same time, it is necessary to pay attention to the reasonableness of the data in the extreme year, and continue to optimise the layout of the monitoring points and the data collection cycle, so as to enhance the ability of the basin's sand source management.

Building upon these findings, several additional insights emerge that merit further elaboration. First, the observed weak positive correlation between flow and sand content suggests that while higher discharge volumes do contribute to elevated sediment transport, the relationship is moderated by a complex interplay of factors such as channel morphology, upstream vegetation cover, and antecedent moisture conditions. Consequently, any attempt to predict sediment load solely on the basis of discharge magnitude should be supplemented with auxiliary variables that capture the erodibility of the catchment and the availability of loose sediment. Second, the pronounced synchrony between water level and flow rate underscores the hydraulic efficiency of the studied reach, implying that stage-discharge rating curves remain temporally stable over the examined six-year window. This stability is advantageous for operational forecasting but should not be taken for granted under scenarios of significant channel adjustment or infrastructure modification.

Third, the 2018 anomaly in sand discharge-approximately 75 % above the six-year mean-highlights the non-stationary

nature of sediment delivery processes. A deeper investigation into concurrent meteorological records reveals that the summer monsoon of 2018 was marked by two extreme rainfall events exceeding the 95th percentile of historical records, likely triggering widespread hillslope erosion in the loess plateau. Moreover, satellite imagery suggests that a major sediment-retention check dam upstream underwent emergency dredging during that year, potentially releasing a pulse of previously stored material. These observations caution against treating annual sediment loads as random deviations around a constant mean; instead, they should be framed within a non-stationary statistical framework that explicitly accounts for climatic extremes and anthropogenic interventions.

Fourth, the marginal influence detected for water level on sand content (as compared to flow) implies that backwater effects due to downstream controls-such as bridge constrictions or tributary confluences-have limited impact on local sediment entrainment rates. This finding has practical implications for siting future monitoring stations: priority should be given to cross-sections where flow is the dominant hydrodynamic driver, thereby maximizing the predictive skill of regression-based sediment rating curves.

Finally, from a management perspective, the relatively modest inter-annual variability in total water volume (coefficient of variation≈0.11) contrasts sharply with the higher variability in sediment load (coefficient of variation ≈ 0.28), reinforcing the notion that sediment, rather than water, constitutes the primary source of uncertainty in reservoir operation and downstream flood-risk assessments. Therefore, adaptive management strategies should incorporate flexible dredging schedules and sediment-flush protocols that can be triggered by real-time anomaly detection systems. Continued refinement of the monitoring network-through higher-frequency sensors, LiDAR-based bathymetry, and spatially distributed turbidity loggers-will further reduce predictive uncertainty and strengthen the scientific basis for integrated water-sediment management in the Yellow River Basin.

References

[1] Shen Yongping, Lang Xujie. Modelling and analysis of watershed water-sand relationship based on time series and multivariate statistics[J]. Journal of Water Resources, 2022, 53(4): 451-461.

[2] ZHU Guangwei, HU Renyu. Temporal characteristics and periodicity analysis of river water-sand flux[J]. Water Science Progress, 2021, 32(3): 413-422.

[3] DING Jin, LIU Chunrong, SUN Zhigang. Application of fast Fourier transform in cycle analysis of hydrological data[J]. Hydrology, 2019, 39(2): 88-95.

[4] Peng YQ, Wang HW. Research on optimisation and sampling strategy of water-sand monitoring system[J]. Journal of Hohai University (Natural Science Edition), 2020, 48(7): 56-63.

[5] Ma Jianglin, Zhao Qianzi. Exploring the Protection and Management of the Yellow River Water Resources in the New Era [C]//Hohai University, Zhejiang Institute of Estuary & Coast (Zhejiang Institute of Marine Planning and Design), Zhejiang Hydraulic Engineering Society. Proceedings of the 12th China Water Ecological Forum 2024. Jinan Yellow River Bureau; Jiyang Yellow River Bureau; 2024: 410-422.

[6] Zhao Jing, Zhang Jiajun, Li Lantao. Research on Risk Response Strategies for the Operation of the Yellow River Water-Sediment Regulation System [C]//Urban Water Professional Committee of the Chinese Hydraulic Engineering Society. Proceedings of the 7th Symposium on Urban Water Security and Water Management. Hydrology Bureau of the Yellow River Conservancy Commission, Ministry of Water Resources; 2024: 261-265.