Transactions on Environment, Energy and Earth Sciences
ISSN: 2960-1924, eISSN: 2960-2300 | Volume 2
ERET 2024

Trends of water and sand fluxes in the Yellow River analyzed based on LSTM model prediction

Peixi Yan[1, *, #], Haotian Luo[2, #], Kaiwen Zheng[3, #]

[1]School of Intelligent Manufacturing, Chongqing Jianzhu College, Chongqing, China, 400072

[2]School of Construction Management, Chongqing Jianzhu College, Chongqing, China, 400072

[3]School of Smart City Design, Chongqing Jianzhu College, Chongqing, China, 400072

*Corresponding author: 18282922349@163.com

#These authors contributed equally.

Abstract. The Yellow River, the river with the highest sand content in China, the change in its water-sand flux has a significant impact on the ecological environment and resource management. In this paper, we mainly use the LSTM neural network model to predict the change rule of water and sand flux in the next two years according to the data from 2016 to 2021, and finally draw time series graphs according to the prediction results, we find that the water and sand flux has seasonality and sudden change, according to this characteristic, we formulate the optimal monitoring program for the dynamic change of water and sand flux in the next two years, to reduce the waste of monitoring resources. Mastering the change rule of water and sand flux in the Yellow River can provide a scientific basis for allocating water resources and constructing an ecological environment in the Yellow River.

Keywords: Water-Sand Flux, LSTM Neural Network Model, Time Series Trend Charts.

1. Introduction

The Yellow River is a vital water resource in northern China [1]. Due to the loose soil quality of the Loess Plateau, it is easy to erode, especially under the scouring of heavy rainfall. A large amount of sediment is brought into the Yellow River, which increases the sediment volume significantly and elevates the riverbed, which is potentially risky for the well-being of the people downstream and even causes substantial economic losses [2-4]. In order to manage and utilize the water resources of the Yellow River scientifically and reasonably, and at the same time to reduce the losses caused by related disasters, it is essential to study the trends and laws of the Yellow River water volume and sand discharge [5].

We reviewed the related literature and found that in order to deeply understand the dynamic change law of water and sand fluxes, some researchers used multiple linear regression model to analyze the influencing factors of water and sand fluxes in the Yellow River, and based on the analysis, they predicted the changing trend in the next two years [6-7]. Some researchers used the Mann-Kendall test to estimate the long-term trend of runoff and sediment time series based on decades of hydrological measured data to analyze the changing trend [8].

We found that although the correlation between water flow and sand discharge is obvious and has a specific linear relationship, the prediction using a multiple linear regression model is easily affected by outliers and outliers, which reduces the robustness of the model and the prediction effect, so it is necessary to preprocess the data. At the same time, it is easy to produce the underfitting problem. Due to the powerful function of the LSTM neural network model, it is easy to deal with outliers and outliers, and it is highly robust. For a vast data set, the LSTM model can divide the data set into training, validation, and test sets, effectively avoiding the underfitting problem and making the prediction results more accurate.

2. Fundamentals of LSTM neural network modeling

2.1. The Structure of the LSTM neural network

The LSTM model is a special kind of RNN model mainly used to solve the gradient dispersion problem of the RNN model when dealing with long-term dependence problems [9].LSTM model controls the loss of information and circulation by introducing a gate mechanism to be able to maintain long-term memory in the sequence, more completely applicable to learning time series, especially the data with poor data fitting, which can be trained continuously to achieve a better fitting data curves and produce accurate results, it predicts nonlinear problems with more accurate predictions.

The fundamentals of LSTM modeling [10]:

A gated output method is used, i.e., input gate, forgetting gate, output gate, where the output value h_{t-1} of the previous moment and the input value x_t of the current moment, from these two parameters, first enter the forgetting gate and get the information f_t which is decided to be discarded, and then enter the input gate and get the decision i_t which is decided to be updated, and finally, the combination of the output values of the two gates get the long-time and short-time information, respectively, and finally for storage. The network parameters of LSTM are given in the following table 1.

Table. 1. The network parameters are listed below

parameters	hidden meaning	parameters	hidden meaning	parameters	hidden meaning
x_t	Input at the current moment	f_t	Oblivion gate control signal	C_{t-1}	Previous Momentary Memory Unit
$\overline{C}_t$	Information Candidate Status	i_t	Inputs control signals	C_t	Current moment memory cell
h_{t-1}	Previous external state	o_t	Output gate control signal	h_t	Current moment external state

The LSTM model cell structure is shown below in Figure 1:

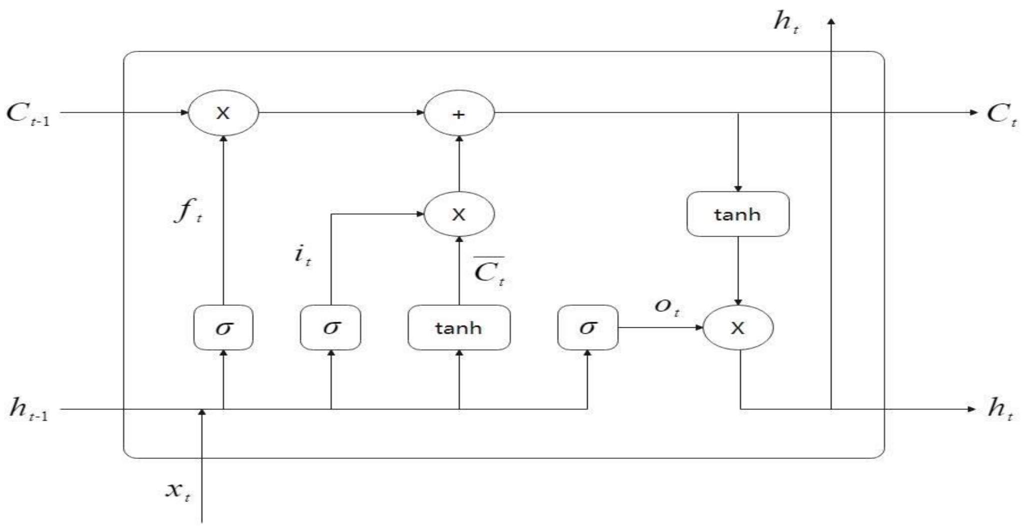


Figure 1. LSTM cell structure diagram

Forgetting gate f: This is to decide how much data should be forgotten in the previous memory unit state before inputting it and storing it to the current memory unit, and randomly forgetting through the sigmoid function tends to 0 to indicate that it should be forgotten. The more it tends to 1 to indicate that it should be retained.

$$f_t = \sigma(\omega_f \cdot [h_{r-1}, x_t] + b_f) \tag{1}$$

Input gate i: Controls the input of total water flow and sand discharge data from 2016 to 2021 at the current moment and saves the data to be predicted to the current memory cell state.

$$\begin{cases} i_t = \sigma(\omega_t \cdot [h_{t-1}, x_t] + b_i) \\ \overline{C_t} = \tanh(\omega_t \cdot [h_{t-1}, x_t] + b_c) \end{cases} \tag{2}$$

The information from the previous moment is retained by adding the retained information from the current input, thus constituting the cell state at the current moment c_t.

$$C_t = f_t \times C_{i-1} + i_t \times \overline{C_t} \tag{3}$$

Output gate o: transforms the input data to the memory cell at the current moment and outputs the data.

$$\begin{cases} o_t = \sigma(\omega_t \cdot [h_{t-1}, x_t] + b_0) \\ h_t = o_t \times \tanh(C_t) \end{cases} \tag{4}$$

3. Results

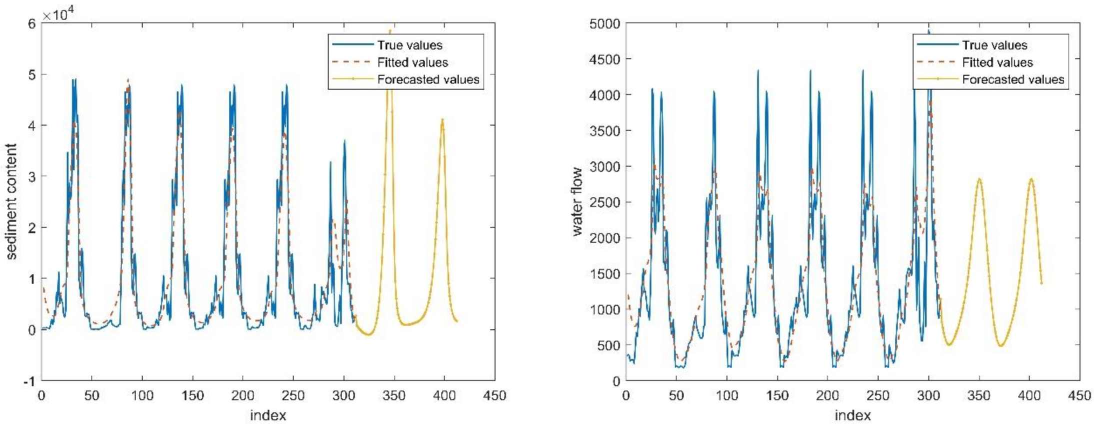
3.1. Prediction and analysis of water and sand fluxes

We constantly adjusted the parameters of the LSTM model to make the best prediction. Therefore, we set the specific parameter values as follows in Table 2:

Table. 2. LSTM neural network parameter settings

parameters	numerical value
Number of LSTM layers	1
attenuation factor	0.2
Number of hidden units	100
Number of iterations	100
Initial learning rate	0.001
Number of rounds between learning rate updates	50
discard layer probability	0.2
random number seed	1234

By setting the parameters of the neural network and using MATLAB toolbox to predict the next two years based on the existing data, the data of water flow and sand discharge in 2022 and 2023 can be obtained. The prediction effect of water flow and sand discharge in the next two years can be obtained as shown in Fig. 2 below:



(a) Water flux (b) Sand Discharge

Figure 2. Projected results for the next two years

Among them, the goodness of fit $R^2=0.9576$ for the water flow in the next two years and $R^2=0.904$ for the sand discharge in the next two years, so the goodness of fit is high, which indicates that the use of the LSTM algorithm for training the model is highly accurate and ideal. Its prediction range is between 300-450.

By predicting the data of water flow in the years 2022 and 2023, the trend graphs can be made using MATLAB respectively:

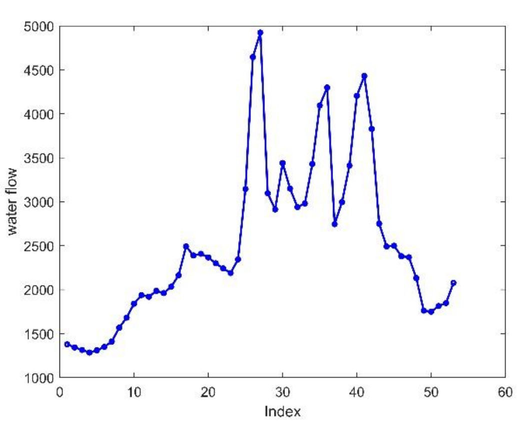


Figure 3. Trends in water flow in 2022

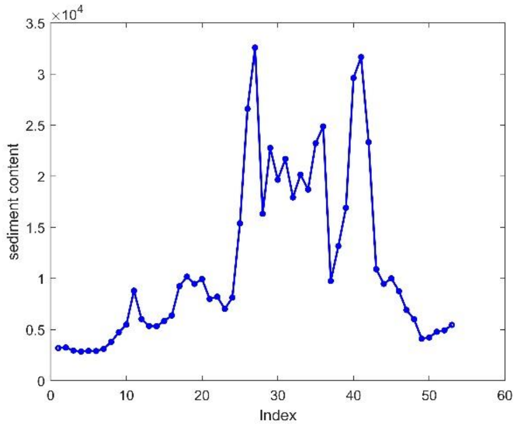
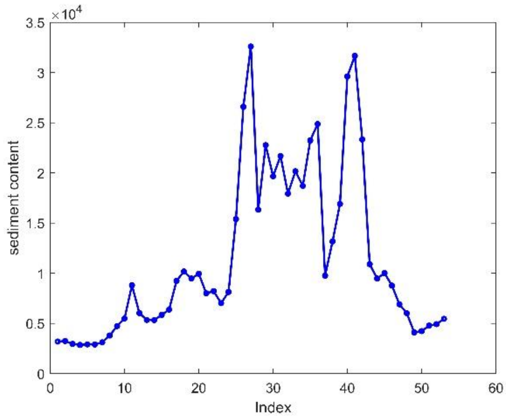


Figure 4. Trends in water flow in 2023

According to the trend graphs in Figures 3 and 4, it is not difficult to find that the two-year trend of water flow is mainly fluctuating and rising during the period of 1-22 weeks per year, and the trend of water flow is not significant; the trend of change is enormous during the period of 23-45 weeks per year; and the trend of fluctuating and declining and then rising is mainly present during 46-53 weeks per year, and the overall degree of change has a distinctly seasonal and abrupt change.

By predicting the data of water flow in the years 2022 and 2023, the trend graphs can be made using MATLAB respectively:



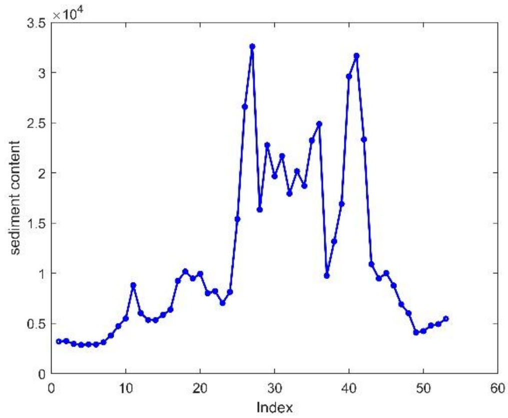


Figure 5. Trends in sand discharge in 2022 **Figure 6.** Trends in sand discharge in 2023

According to the trend graphs in Fig. 5 and Fig. 6, it is not difficult to find that the two-year trend of sand discharge mainly fluctuating during the period of 1-25 weeks per year, and the trend of sand discharge is not significant; during the period of 26-45 weeks per year, the trend of change is enormous; during 46-53 weeks per year, it is mainly fluctuating and decreasing and then increasing, and the overall trend of change is in line with that of the water flow rate. It has apparent seasonality and sudden change.

In summary, we found that the change rule of water and sand flux is as follows: water and sand flux fluctuate in a year, and the change is the largest in summer and fall, while the change is minor in spring and winter, basically holding a stable state. There is a sharp increase and decrease in July and August, with pronounced seasonality and sudden change.

3.2. Development of a water and sand flux monitoring program

A future monitoring program was developed for this hydrological station, considering the need to keep abreast of the dynamics of water and sand fluxes while reducing detection costs. Since the changes in water flux are essentially the same for both years. Since sand discharge is correlated with water flux, it is only necessary to develop a one-year observation program based on a trend graph of water flux.

Table. 3. Schedule of monitoring programs

weeks	timing	ordinal number	stop raining
1-31	8：00	At least once	Immediate monitoring
32-45	8：00	once a day	Immediate monitoring
46-53	8：00	Proper monitoring	Immediate monitoring

According to the monitoring program in Table 3, monitoring should be carried out on sunny days. During weeks 1 to 31, monitoring was carried out at least once a week and immediately after rainfall because the water flow trend was relatively smooth. During weeks 32 through 45, monitoring was conducted daily on sunny days and immediately after the rain stopped because it coincided with the rainfall season in the Yellow River Basin. During weeks 46 to 53, because of the low water flow, weekly monitoring was appropriate and was carried out immediately after the rains.

4. Conclusions and outlooks

In this paper, a water-sand monitoring schedule was developed based on data on water-sand flux over the last six years. The water-sand flux of the Yellow River was predicted by building an LSTM model, and time-series trend diagrams further analyzed the trends and patterns of the Yellow River water-sand flux. We found that the water-sand flux fluctuates in a year, and the change is the largest in summer and fall. In contrast, the change is minor in spring and winter, basically holding a stable state. A sharp increase and decrease occurs in July and August, with pronounced seasonality and sudden change. The LSTM model, comparatively, fully applies to learning the time series. Due to the lack of water and sand flux data in some years, the time series continuity of the data is low. The LSTM network model is more accurate in predicting the data with low continuity, which facilitates the prediction of water and sand fluxes in the lower reaches of the Yellow River and makes it possible to rationally optimize the allocation of water resources in the Yellow River Basin, thus saving economic costs. The model is also applicable to many aspects of economic production. For example, the enterprise is forced to stop production due to a sudden disaster, contributing to the production data's lower temporal continuity. Therefore, the LSTM model can be used to achieve a better fitting curve to improve the accuracy of the prediction, which can bring more significant economic benefits.

LSTM models usually require a large amount of training data and several iterations for convergence, thus leading to long training times. GPU acceleration can be used to increase the speed of training, and model pruning can also be performed to reduce the number of parameters, thus reducing the training time.

References

[1] Lin Z ,Chao Z ,Qing W , et al.Sediment Coarsening in Tidal Flats and Stable Coastline of the Abandoned Southern Yellow River Sub-Delta in Response to Fluvial Sediment Flux Decrease During the Past Decades#13;[J].Frontiers in Marine Science,2021,8

[2] Gao P ,Wang Y ,Li P , et al.Land degradation changes in the Yellow River Delta and its response to streamflow-sediment fluxes since 1976[J].Land Degradation Development, 2018, 29 (9):3212-3220.

[3] Ma X ,Zhang G ,Liu S , et al.Distributions and fluxes of nitrous oxide in lower reaches of Yellow River and its estuary: Impact of water-sediment regulation[J].Estuarine, Coastal and Shelf Science,2016,16822-28.

[4] Coastal Research; Researchers at Ludong University Have Reported New Data on Coastal Research (Long-term Trend and Change Point Analysis On Runoff and Sediment Flux Into the Sea From the Yellow River During the Period of 1950-2018)[J].Energy Ecology,2020,

[5] Coastal Research; Researchers at Ludong University Have Reported New Data on Coastal Research (Long-term Trend and Change Point Analysis On Runoff and Sediment Flux Into the Sea From the Yellow River During the Period of 1950-2018)[J].Energy Ecology,2020,

[6] Dang J ,Cui Z ,Liu J , et al.Analysis of the Change Rule and Future Trend of Yellow River Water and Sand Flux Based on Linear Programming Model[J].Academic Journal of Environment Earth Science,2024,6(2):

[7] Zhang K ,Li Y ,Jiang M , et al.Data-Driven Analysis of Water and Sediment Flux at the Xiaolangdi Station of the Yellow River[J].Academic Journal of Environment Earth Science,2024,6(2):

[8] Sui Y ,Shi H ,You Z , et al.Long-term Trend and Change Point Analysis on Runoff and Sediment Flux into the Sea from the Yellow River during the Period of 1950-2018[J].Journal of Coastal Research,2020,99(sp1):203-207.

[9] Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network[J]. Physica D: Nonlinear Phenomena, 2020, 404: 132306.

[10] Song X, Liu Y, Xue L, et al. Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model[J]. Journal of Petroleum Science and Engineering, 2020, 186: 106682.