

# 黄河水沙监测数据分析

## 摘 要

黄河是中华民族的母亲河，孕育了丰富的物产资源，带动沿岸的发展，但是泥沙问题一直是黄河的症结所在，研究黄河水沙通量的变化规律具有重大意义，对此我们进行以下研究。

本文针对因素关系问题，采用相关性分析、归一化处理以及最小二乘法拟合，得出各因素之间关系；针对变化规律总结问题，采用小波相干分析、M-K 检验和最小二乘法拟合，基于分段拟合思想，建立线性函数、季节性 ARIMA 与 BP 神经网络模型，求解分析得出变化规律；针对趋势预测问题，基于分段拟合函数建立预测模型，预测得出未来趋势，基于满足及时掌握水沙通量动态变化以及最大程度减少检测成本的要求，给出优化模型得到最优采样检测方案。

**针对问题一** 为分析得出含沙量与时间，水位及水流量之间关系，首先对各因素之间进行相关性分析，含沙量与水位、含沙量与水流量以及水流量与水位相关强。分别建立含沙量与三个变量之间的线性拟合函数，采用最小二乘法进行求解；针对估算 6 年水流量及水排沙量，按照日水流量及含沙量的数据，得到每年总水流量及排沙量，再求和得出 6 年总水流量及排沙量分别为 19812805.6979 和 12125.08593000 万吨。

**针对问题二** 为研究水沙通量的变化规律，采用小波相干法分析得到周期性规律情况；采用 M-K 检验研究突变性，得出突变点；基于数据变化趋势，利用分段拟合研究季节性，建立线性、季节性 ARIMA 模型与 BP 神经网络模型，最后求解并进行精度分析，对比得出最优预测模型为季节性 ARIMA 模型，模型的适配度高，拟合精度高。

**针对问题三** 已知 6 年水沙通量变化，预测未来两年变化，结合问题 2 建立的最优拟合模型，给出后两年水沙通量预测值。结合附件 1 及附件 3 内容，设定目标值，将干旱时期及洪涝时期分段研究推出既可观察动态变化，又减少监测成本的最优方案，以表格形式在正文给出。

**针对问题四** 为研究每年进行调水调沙的实际效果，基于附件 2 求河底高程平均值以及峰值，使用差值法分析调水调沙实施前后河底高程的变化趋势。基于前两问结果，分析水沙变化趋势，量化给出调水调沙的实际效果；并估算 10 年后河底高程增加值为 1.47m。

**关键词：**M-K 检验，小波相干分析、BP 神经网络、季节性 ARIMA 模型

## 一、 问题重述

### 1.1 问题背景

黄河是中华民族的母亲河，孕育了一代又一代的人们，但是泥沙问题一直是黄河的症结所在，研究黄河水沙通量的变化规律，对治理环境，研究气候变化，改善人们生活，及协调资源的分配，防止自然灾害具有重大意义，对此我们进行以下研究

### 1.2 问题重述

**问题一** 研究一家特定水文站的黄河水文数据（附件一）内容，通过分析黄河水含沙量与时间，水位，水流量三者之间的关系并估算 2016~2021 年水文站的年中水流量及总排沙量

**问题二** 分别从突变性，季节性和周期性三个因素，研究 2016 年到 2021 年水沙通量的变化并总结规律

**问题三** 根据问题二所得数据，预测分析水文站 2022 年到 2023 年两年水沙通量的变化趋势，并根据附件 3 历年检测方案为该水文站 2022~2023 制定既掌握动态变化，又可减少监测资源的最优采样方案

**问题四** 根据问题三所给数据及附件 2 和附件 3 所给的信息分析，判断是否需进行调水调沙，若需进行，那实际效果呈现如何，若不进行 10 年之后河底高程又会发生如何变化

## 二、 问题分析

### 2.1 问题一的分析

1.对附件一的数据进行预处理，保留每日 8:00 测量数据，且对水位，水流量及含沙量数据进行归一化处理

2.对保留数据进行相关性分析，判断含沙量分别与时间，水流量及水位的相关性强弱

3.根据对应的强弱性特点，分别使水位，水流量及时间与含沙量一一对应采用最小二乘法拟合分析,得出与含沙量的各自关系

4.通过保留数据中水流量及含沙量数据求和计算得出每年总水流量和年总排沙量，再求和得出六年总水流量及排沙量

### 2.2 问题二的分析

1.针对突变性特性，采用 M-K 统计方式分析得出年突变点及天突变点区间，得出规律

2.针对周期性特性，采用六年每时刻水沙通量值进行小波相干分析，得出年间规律

3.针对季节性特性，将出现变化区间分别采用线性方式，BP 神经网络进行分段处理，得出变化规律

4.总结上述规律，得出水沙通量变化规律

2.3 问题三的分析

- 1.在问题 2 水沙通量变化规律基础上,为准确预测水文站未来两年水闸通量,建立 BP 神经网络模型,线性拟合模型及非线性拟合模型;
- 2.各自模型通过每年各月的水沙通量变化数据,推算出预测年各月的水沙通量;
- 3.将预测年各月水沙通量综合统计得到预测年水沙通量变化图;
- 4.将三类模型所得趋势预测进行比较,更符合现实性及科学性,则代表预测模型合适,以该预测模型为预测年作水沙通量变化趋势;
- 5.根据上述变化趋势及水沙通量的变化规律,结合附件 1, 3 内容得出既掌握水沙通量动态变化又可降低成本的最优采样检测方案区间段历年监测手段得出最优采样监测方案。

2.4 问题四的分析

- 1.根据国家水文局相关报告,得出调水调查工程,一般开始于 6 月 19 日至 7 月 19 日,为分析每年六七月份进行调水调沙的实际效果,以附件二数据为基础,使用数学软件工具箱对附件 2 内容数据预处理。
- 2.预处理数据之后,使用差值计算方式得出各年间水沙通量的变化
- 3.通过数据变化可分析判断,得出调水调沙的实际效果。
- 4.为明确不进行调水调沙会造成什么影响,计算得出不进行调水调沙 10 年是水文站河底高程深度

三、基本假设

- 1. 假设忽略黄河截面。
- 2. 题目各附件所给的数据真实可靠。
- 3. 假设年与年之间总含沙量变化趋势较小,则认为调水调沙力度不够,认为没有进行调水调沙。

四、符号说明

符号定义	符号说明
$Q$	年总水流量
$t$	时间
$q$	日水流量
$P$	年总排沙量
$a$	日排沙量

## 五、模型建立与求解

### 5.1 问题一的模型建立与求解

数据的预处理：利用插值法对每日每个时刻的含沙量空白数据进行补充。且对水位，水流量及含沙量数据进行 Min-Max 归一化，对原始数据进行线性变换，使结果值映射到[0, 1]之间。转换函数如下：其中  $\max$  为样本数据的最大值， $\min$  为样本数据的最小值，公式如下：

$$\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

#### 5.1.1 含沙量与时间，水流量及水位相关性分析

附件 1 给出 2016-2021 共计 16743 条数据，每条数据含流量，水位，时间及含沙量。在 EXCEL 中预处理数据，筛选得出含沙量未空白的数据，用 Python 对数据中六年含沙量、时间、水位及水流量进行相关性分析得出下图 1。

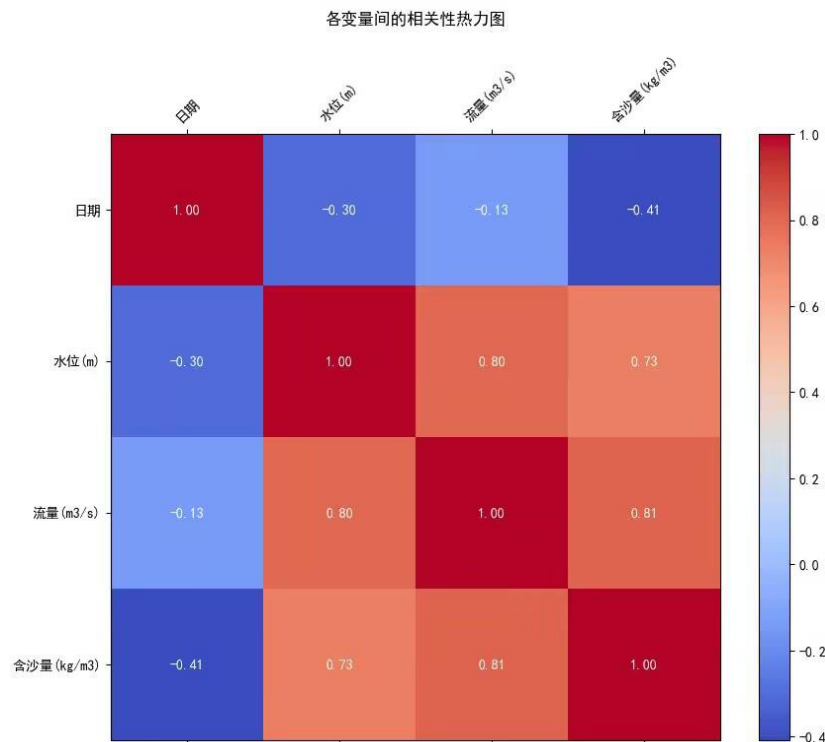


图 1 含沙量，时间，水位及水流量四者相关性热力图

从图 1 中可得出 [0.95,1]区间代表强相关性，[0.5,0.7]区间代表较强相关性，[0.3,0.5]区间代表较弱相关性，[-0.4,0.3]区间代表弱相关性。图中含沙量与水流量相关性系数为 0.79 为强相关性，含沙量与水位相关性系数为 0.73 为强相关性，由此可得出含沙量时间相关系数为-0.4 为弱相关性，水位与水流量相关性系数为 0.81 为强相关性。

根据相关性强弱，为直观得出含沙量与水流量，含沙量与水位，含沙量与时间，水位与水流关系，使用最小二乘法一一进行拟合。

### 5.1.2 含沙量与时间的关系分析

通过图 2 可分析得出，2016-2021 年含沙量与时间整体变换趋势幅度趋于一致，得出每年含沙量与时间关系变化幅度相似，可调取六年其中一年数据进行时间与含沙量的关系分析。

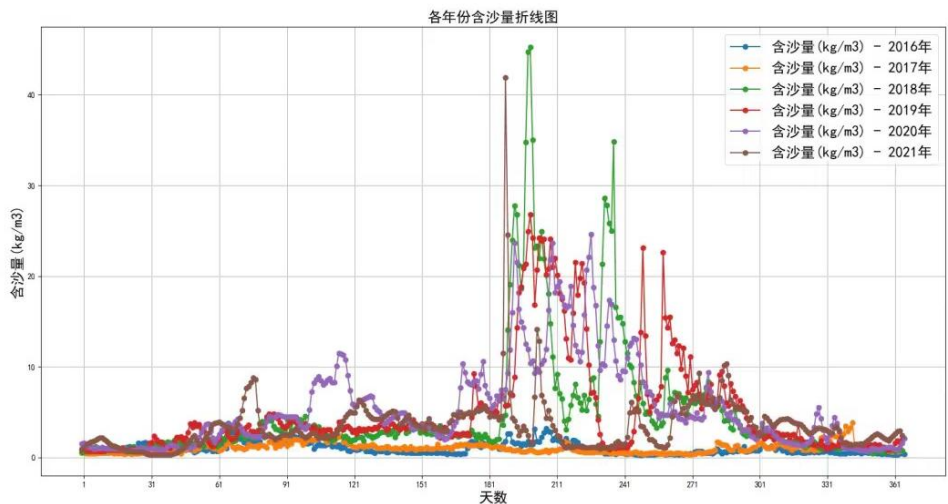


图 2 2016-2021 各年含沙量与时间变化趋势图

为更直观发现时间与含沙量的关系，在于 2016-2021 曲线变化图中对比，2016 变化较稳定，调选出 2016 年含沙量与时间的关系图（图 3）进行以下研究：

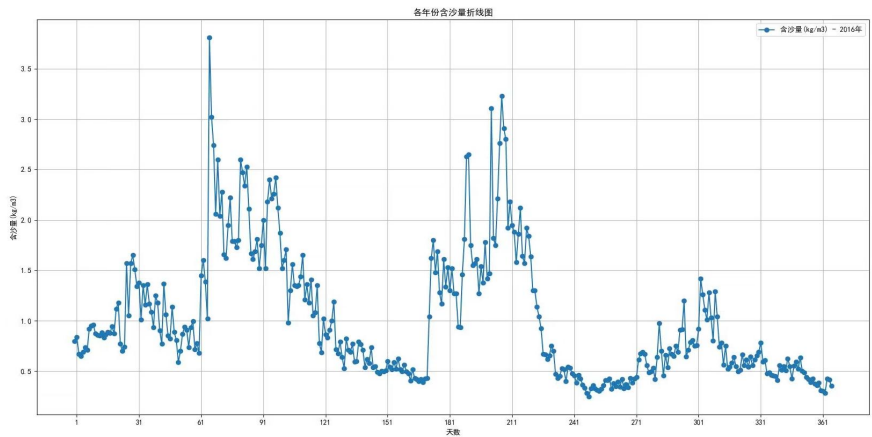


图 3 2016 含沙量与时间变化趋势图

从图 3 中，可得出在四月份左右与七月份左右含沙量会出现剧烈变化，根据相关文献说明，四月份黄河处于春汛时期，而七月份黄河处于夏汛时间，因此每逢春汛，夏汛时间段含沙量会增多，但含沙量与时间并无强相关性联系

### 5.1.3 含沙量与水流量的关系分析

通过上述分析，可得出含沙量与水流量具有强相关性，即存在线性联系，为

更加直观得出含沙量与水流量的关系，使用 Python 对数据中六年含沙量及水流量采用最小二乘法进行一元拟合分析，得出以下数据图例（图 4）

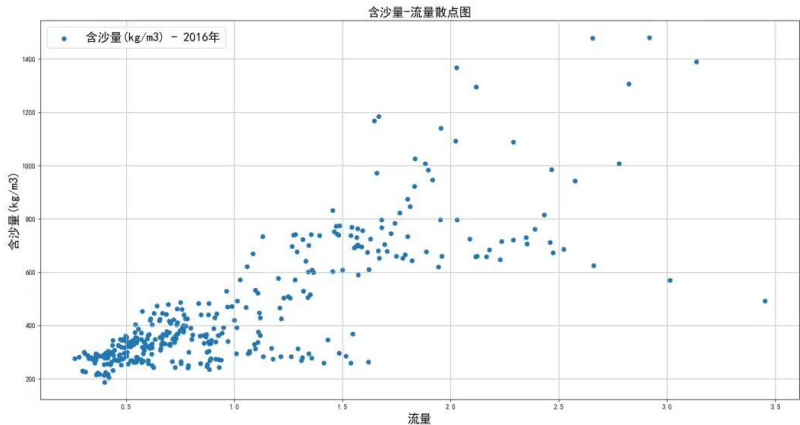


图 4 含沙量-流量散点图

并推导得出水位与水流量之间的一元拟合公式：

$$Y = 138.446 + 312.612x \tag{2}$$

#### 5.1.4 含沙量与水位的关系分析

通过上述分析，可得出含沙量与水位具有强相关性，即存在线性联系，为更加直观得出含沙量与水位的关系，采用 Python 对数据中六年含沙量及水位采用最小二乘法进行一元拟合分析，得出以下数据图例（图 5）

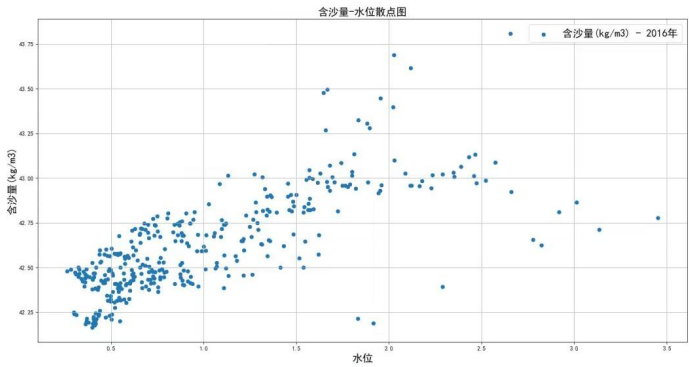


图 5 含沙量-水位散点图

并推导得出含沙量与水位之间的一元拟合公式：

$$Y = 42.295 + 0.328x$$

#### 5.1.5 水位与水流量的关系分析

由于水位与水流量有强相关性，可预估水位及水流量可能存在线性联系，为明确水位与水流量相关关系，利用 Python 工具箱对预处理后的数据（附件 1）采用最小二乘法进行拟合，如下图 6。

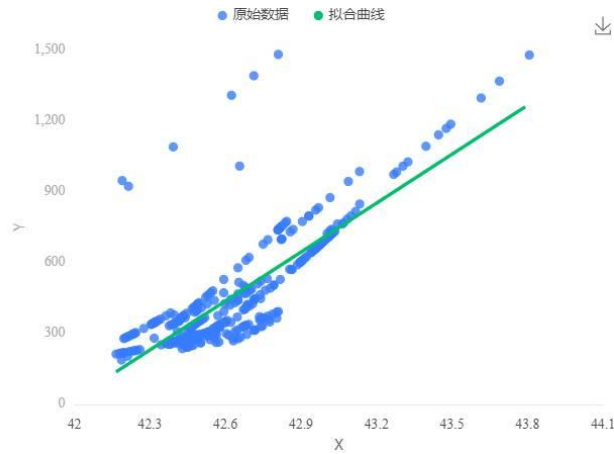


图 6 水流量与水位拟合曲线

并推导得出水位与水流量之间的一元拟合公式：

$$Y = 36.95 + 0.082x \quad (3)$$

### 5.1.6 含沙量，水位，水流量的关系分析

通过上述分析，已得出含沙量及时间两者不存在具体关系，在分析含沙量与水位、水流量、时间的关系时，可简化分析为含沙量与水位，水流量之间关系。

为直观看出三者之间关系，采用 Python 对于附件一所相关数据进行多元拟合（图 7）

假设含沙量与水位及水流量之间的多元拟合公式：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + e, \quad (4)$$

其中， $y$  是因变量，也就是含沙量(kg/m<sup>3</sup>)；

$\beta_0$  是截距项表示当所有自变量都为 0 时因变量的预测值；

$\beta_1, \beta_2, \dots, \beta_n$  表示各个自变量系数，表示自变量对因变量的影响；

$x_1, x_2, \dots, x_n$  是归一化后的自变量,及水位(m)和水流量(m<sup>3</sup>/s)；

$e$  是误差项，表示模型无法解释的随机误差。

OLS Regression Results						
Dep. Variable:	含沙量(kg/m3)		R-squared:	0.678		
Model:	OLS		Adj. R-squared:	0.676		
Method:	Least Squares		F-statistic:	381.4		
Date:	Sat, 09 Sep 2023		Prob (F-statistic):	6.02e-90		
Time:	10:27:35		Log-Likelihood:	-133.73		
No. Observations:	366		AIC:	273.5		
Df Residuals:	363		BIC:	285.2		
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.3354	0.036	9.221	0.000	0.264	0.407
x1	2.0718	0.166	12.505	0.000	1.746	2.398
x2	0.8848	0.182	4.863	0.000	0.527	1.243
Omnibus:	145.575	Durbin-Watson:	0.295			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	809.831			
Skew:	1.589	Prob(JB):	1.40e-176			
Kurtosis:	9.558	Cond. No.	13.6			

图 7 含沙量与水位及水流量流量的多元拟合方程

得出含沙量与水位及水流量流量的多元拟合方程：

$$Y = 0.3354 + 2.718x_1 + 0.8848x_2 \quad (5)$$

### 5.1.7 年总水流量的估算

水流量的单位是立方米每秒，简而言之，日流量 $Q$ 可以通过小时水流量 $q$ 乘以一天的秒数（3600 秒）和一天的小时数（24 小时）来计算，从而得到：

$$Q = 3600 \times 24 \times q \quad (6)$$

年水流量则是将所有的日水流量累加，从而得到一年的总流量

表 1 年总水流量

年份	年总水流量/万立方米	六年总水流量/万立方米
2016	1434585.6000	19812805.6979
2017	1536865.9200	
2018	3898134.7200	
2019	3878098.5600	
2020	4336513.0276	
2021	4728607.8703	

### 5.1.8 年总排沙量的估算

含沙量表示单位体积水中所含沙的质量，日排沙量( $P$ )可以通过日含沙量( $a$ )与日流量( $q$ )相乘来计算： $P = q \times a$  表示每日排放的含沙量，以千克为单位。

年排沙量则是将所有的日排沙量累加，从而得到一年总排沙量



表 2 年总排沙量

年份	年总排沙量/万吨	六年总排沙量/万吨
2016	182.44868285	12125.08593000
2017	190.72690358	
2018	2940.99961304	
2019	3050.83775635	
2020	3495.04092842	
2021	2265.03204997	

## 5.2 问题二的模型建立与求解

### 5.2.1 针对周期性的分析

为更加清晰看出周期性规律,采用小波相干分析方法对六年水沙通量进行分析判断如下图 8

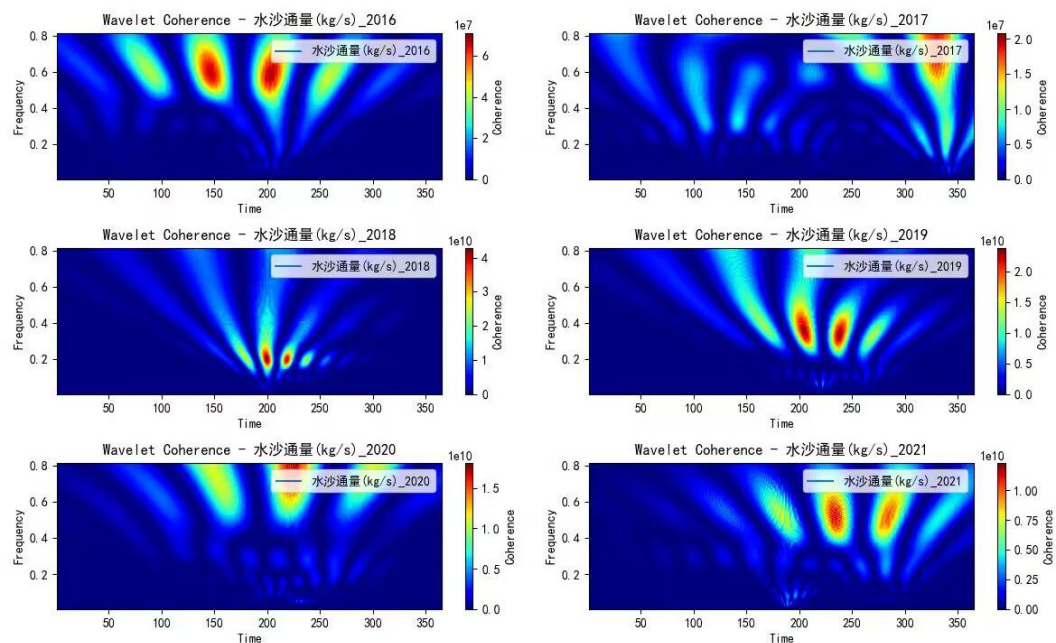


图 8 六年小波相干谱图

由六年水沙通量的小波相干谱可以发现,时间和水沙通量共同的高能量周期 [150,300], 并且在该区间的周期尺度通过了 95%显著性检验,且二者呈正相位关系;时间和水沙通量在整个时频空间低能量区内未存在显著的共振周期,且低能量区的显著相关性远低于高能量区,其中的时间尺度在整个时间段内均通过 95%的显著性检验(两端时间受边界影响),且二者以正相位为主,说明水沙通量和时间具有极强的一致性。

综上,水沙通量在在 [150,300]具有周期性。

### 5.2.2 针对突变性的分析

为分析黄河水沙通量的趋势,对 2016-2021 年小浪底水沙时间序列做趋势分

析和滑动平均处理，近六年水沙通量呈现波动的趋势见（图9）。在2016-2017年水沙通量表现为平稳趋势，在2018-2021水沙通量表现为曲折变化，于2017-2020呈现上升趋势但在2020-2021又突然呈现剧烈下降趋势。

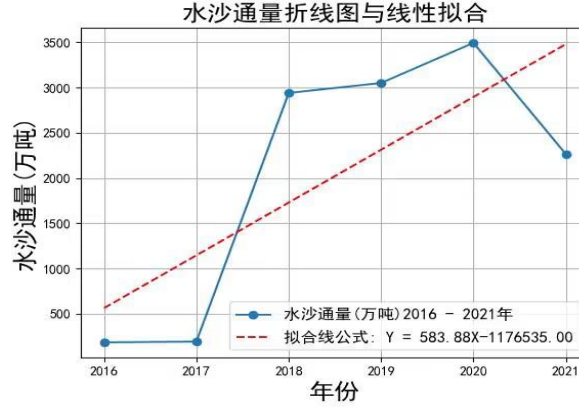


图9 水沙通量变化趋势

在分析2016-2021年黄河水沙通量变化的基础上，通过M-K非参数检验法对水沙通量的突变年份进行确定。

对于具有六个样本量的时间序列 $x_1, x_2, x_3, x_4, x_5, x_6$ ，构建一个秩序列：

$$S_6 = \sum_{i=1}^6 R_i \quad (7)$$

式中： $R_i$ 表示 $x_i$ 大于 $x_j$  ( $1 \leq j \leq i$ ) 的累积数。在时间序列随即独立的假设下，定义统计量：

$$UF_k = \frac{d_k - E(d_k)}{\sqrt{\text{var}(d_k)}} \quad (k = 1, 2, \dots, 6) \quad (8)$$

式中：当 $k=1$ 时， $UF_1=0, d(sk), \text{Var}(dk)$ 是累计数 $Sk$ 的均值和方差，在 $x_1, x_2, x_3, x_4, x_5, x_6$ 相互独立，且有相同连续分布时，它们可由下式算出：

$$\begin{cases} E(d_k) = k(k-1)/4 \\ \text{var}(d_k) = k(k-1)(2k+5) \end{cases} \quad (9)$$

按时间逆序列 $x_6, x_5, x_4, x_3, x_2, x_1$ ，再重复上述过程，同时使 $UB_k = -UF_k, k = n, n-1, \dots, 1$ ，且当 $K=1$ 时， $UB_1 = 0$

分析在近六年水沙通量 $UB_k, UF_k$ 曲线图中， $UB_k, UF_k$ 两条曲线出现交点在 $\{2, 5\}$ 两点置信度水平内（见图八），即代表交点处出现突变，因此，对应的2017

年及 2020 年可能为水沙时间序列的突变年份。

为了进一步阐明近六年水沙通量的突变情况，结合水沙通量序列的累积距平曲线（图九），发现 2016-2021 年水沙时间序列的累积距平曲线大致呈“V”型，同时其拐点出现在 2017 年与 2020 年，说明在 2017 年，入海水沙量呈现增长的趋势，在 2020 年之后开始出现下降，整体上经历了由“丰”到“枯”的变化，因此，在结合多种方法的基础上，认为水沙通量在 2017 年及 2021 年发生变异，尽管两年水沙通量的时间序列具有较强的相关性，但两者的变异年份仍有一定时差，这与黄河流域“水沙异源”特性以及人类活动干扰等因素有关。

综上：2017 年及 2020 年为水沙通量的突变年份

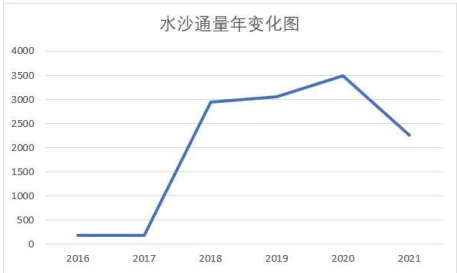


图 10 2016-2021 水沙通量的年变化图

根据 2016-2021 水沙通量的年变化图（图 10），可以得出 2016-2017 两年含沙量及水流量变化平缓，但在 2018 年后出现剧烈变化。

为更加客观分析水沙通量突变性特性，以 2018 年为界限，将 2016-2017 年定为干旱年，2018-2021 定为洪涝年分别进行讨论。

为更加直观分析水沙通量突变特点，将 2016-2021 年内水沙通量变化图逐一作出进行分析判断。

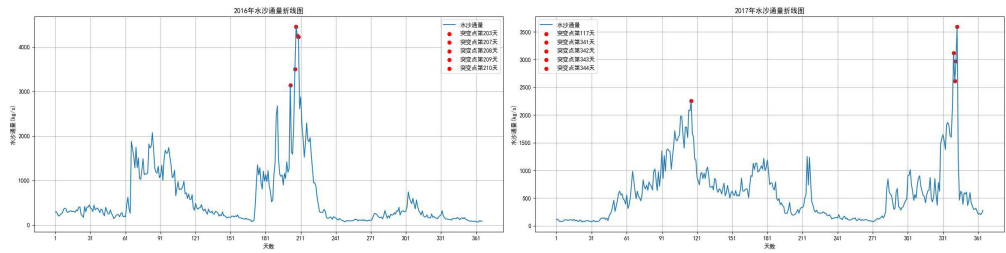


图 11 2016-2017 年水沙通量变化图

通过观察 2016 年的水沙通量变化图，可发现突变点出现在第[200,215]区间内；通过观察 2017 年的水沙通量变化图，可发现突变点出现在[115,120]和[320,330]区间内。

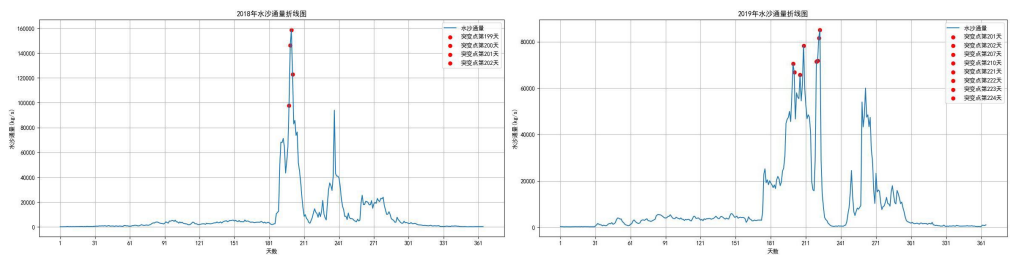


图 12 2018-2019 水沙通量变化图

观察 2018 年水沙通量变化趋势图，可发现突变点出现在[200,210]区间内；观察 2019 年水沙通量变化趋势图，可发现突变点出现在[190,220]区间内。

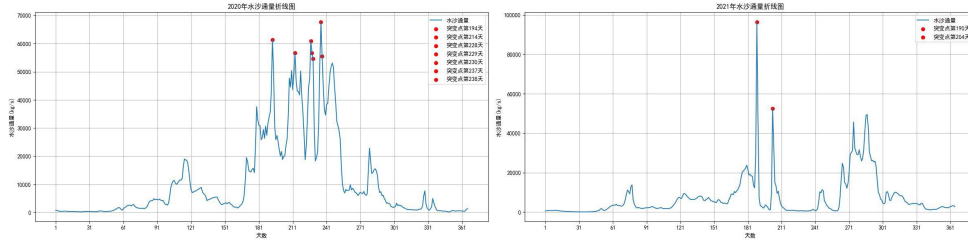


图 13 2020-2021 水沙通量变化图

观察 2020 年水沙通量变化图趋势，可发现突变点出现在 [190, 240]区间内；观察 2021 年是啥通量变化图，可发现突变点出现在 [190,211]区间内。

### 5.2.3 针对季节性的分析

BP 神经网络模型和 ARIMA 模型都可直观体现季节性的变化趋势。采取两者方法进行分析，比较最后季节性分析显著的结果，确定季节性的变化规律。

#### 5.2.3.1 BP 神经网络分析

根据上述 2016-2021 水沙通量的变化研究，得出 180 天之后出现峰值，为此，将一年分段为两部分分别为[0,180]及[180,365]，对于[0,180]采用线性分析，对于[180,365]采用神经网络 BP 分析。

BP 神经网络具备出色的学习能力，能够学习和存储大量的输入-输出模式映射关系，无需预先了解描述这种映射关系的数学方程。BP 网络的学习规则采用最速下降法，利用反向传播技术，不断调整网络的权重和阈值，以使网络的均方误差最小化。以下是 BP 神经网络的算法流程图：

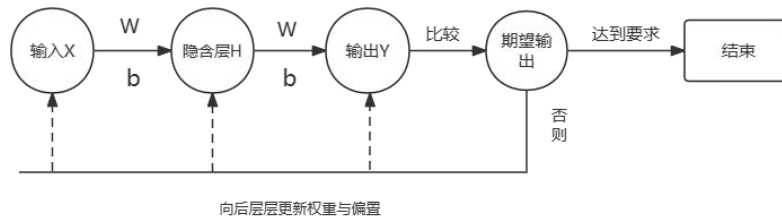


图 14 算法流程图

#### 5.2.3.2 季节性 ARIMA 模型分析

ARIMA 模型是 AR 模型和 MA 模型的直接结合，ARIMA 模型的公式可以表示为：

$$Y_t = c + r_1 Y_{t-1} + r_2 Y_{t-2} + r_1 Y_{t-1} + ..... + r_n Y_{t-n} + \varepsilon + b_1 Y_{t-1} + b_2 Y_{t-2} + r_1 Y_{t-1} + ..... + r_m Y_{t-m} + b_t \quad (12)$$

其中： $Y_t$  是我们正在考虑的时间序列数据。 $r_1$  到  $r_2$  是 AR 模型的参数，这些参数用来描述当前值与过去  $n$  个时间点值之间的关系。 $b_1$  到  $b_2$  是 MA 模型的参数，这些参数用来描述当前值与过去  $m$  个时间点的误差之间的关系。 $b_t$  是在  $t$  时间点的误差项  $c$  是一个常数项。

基于季节性 ARIMA 模型的流程如图所示：

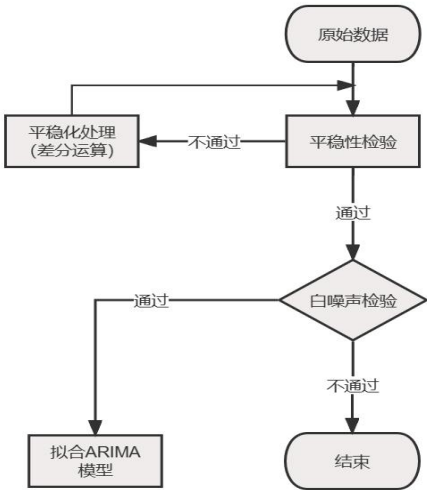


图 15 算法流程图

5.3 问题三 的模型建立与求解

为了科学分析预测未来两年水沙通量的变化趋势，采用 BP 神经网络，线性拟合方式，非线性拟合方式三者进行比对，观察哪类方式预测效果好，并得出预测趋势图。

5.3.1 线性拟合分析预测

通过各年份 1 月相关数据进行拟合分析，可推断出未来 2022 年至 2023 年 1 月份的整体变化趋势。同理，根据 1~12 每月份拟合分析逐月推断出预测年各月变化趋势如下图 15，16

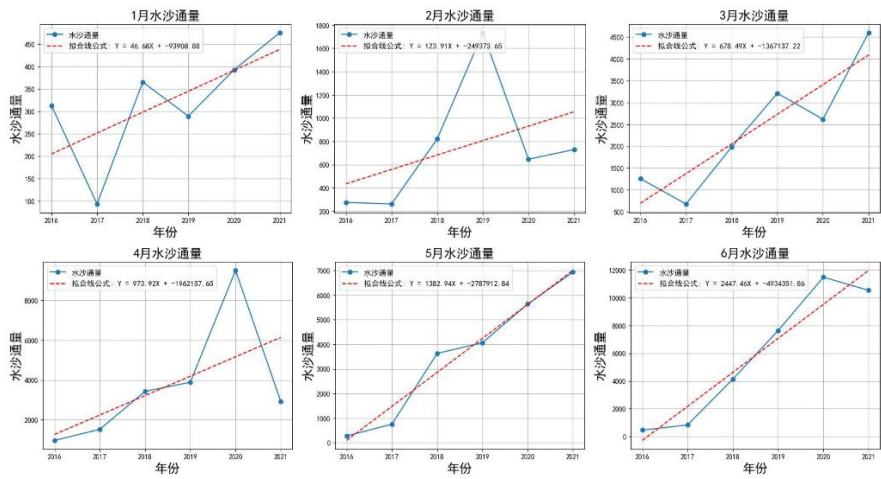


图 16 各年 2016-2021 年 1-6 月的拟合分析



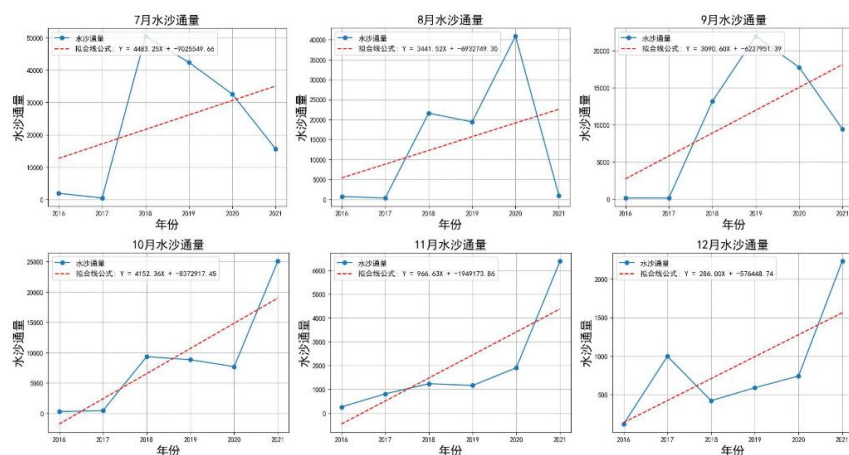


图 17 各年 2016-2021 年 7-12 月的拟合分析  
最后预测得出 2022 年至 2023 年水流通量的年变化趋势如下

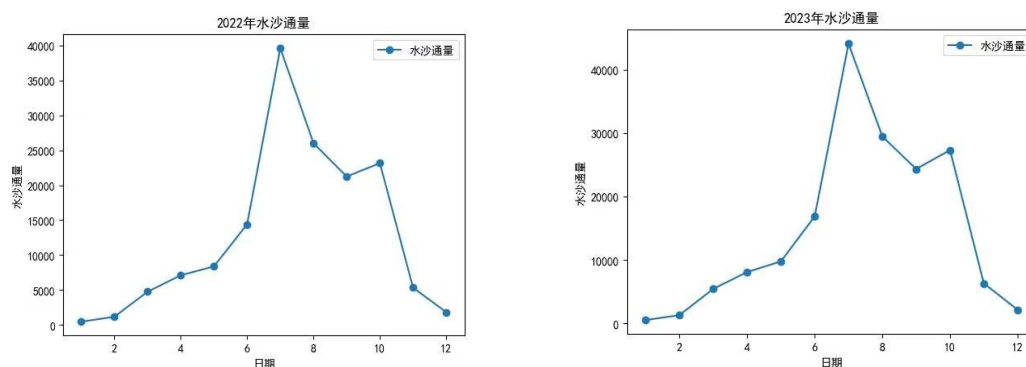


图 18 2022 与 2023 年水沙通量变化预测图

### 5.3.3 ARIMA 模型分析预测

与问题二模型使用方法同理，可得预测值并得出规律曲线为：

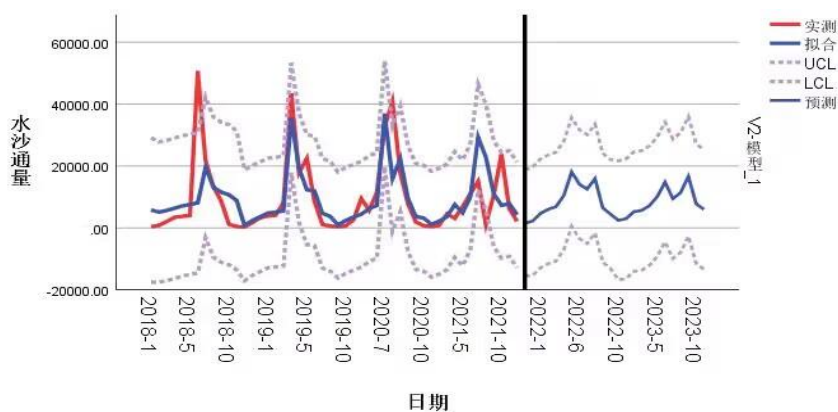


图 19 基于 ARIMA 模型的后两年预测图

综上：通过三类模型之间的比对，得出 BP 神经网络缺少大量数据的支撑，无法成功预测后两年的变化趋势；线性拟合模型最后所得预测后两年变化趋势图重合度太高，不宜判断；ARIMA 模型所得预测趋势符合本题现实规律，得出后两年预测趋势

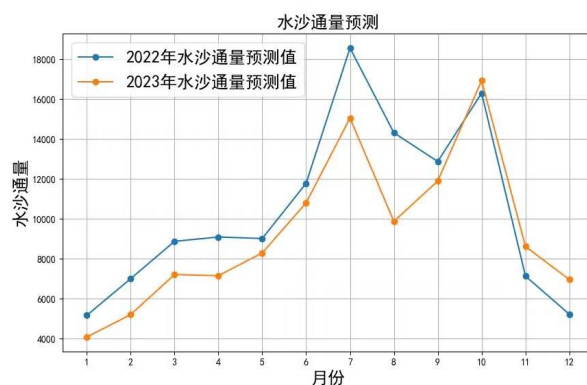


图 20 2022 及 2023 年水沙通量预测图

通过预测图我们可以看出，在七月和十月的时候水沙通量到达了两个顶峰，所以这两个月着重需要监测每天每隔一个小时监测一次，根据附录 3 的数据我们在 2022 年选取 7.7 和 10.15，在 2023 年选取 7.7 和 10.18，从每天早上的 8:00 到 22:00，三月黄河流域出现春汛也需要监测，但是由于预测水沙通量不多所以每隔 3 个小时监测一次从 12:00 到 18:00，在 2022 年选取 3.12，在 2023 年选取 3.22，水沙通量在 2022 年七月到达顶峰后下降到九月才继续升高，所以九月是一个转折点，我们选取 9.7 作为监测日期从 10:00 到 20:00 一天监测 5 次，在 2023 年水沙通量从七月同样骤然下降到八月，八月作为转折点，我们选取 8.5 作为监测日期从 10:00 到 20.00 一天监测五次。

表 4 2022 年监测方案

日期	时刻	时刻	监测次数
2022. 3. 12	12:00	18:00	2
2022. 6. 15	8:00	18:00	5
2022. 7. 23	8:00	22:00	14
2022. 9. 7	10:00	20:00	5
2022. 10. 15	8:00	22:00	14

表 5 2023 年监测方案

日期	时刻	时刻	监测次数
2023. 3. 22	12:00	18:00	2
2022. 7. 13	8:00	18:00	5
2023. 7. 7	8:00	22:00	14
2023. 8. 5	10:00	20:00	5
2023. 10. 18	8:00	22:00	14

这样既能在水沙通量到达顶峰时着重观测，在发生转折时也能及时发现，在发生汛期时也能合理监测达到最大程度地减少监测成本

## 5.4 问题四的模型建立与求解

调水调沙是水库水域泥沙流量调节的两种功能，来形成接近泥沙输运理想值的人造洪峰过程即用最小的水流量来运输最大的泥沙量，而水沙通量代表水流对泥沙的运输能力强度

因此通过观察河底高程的变化趋势,可分析调水调沙对于泥沙量及水流量的影响

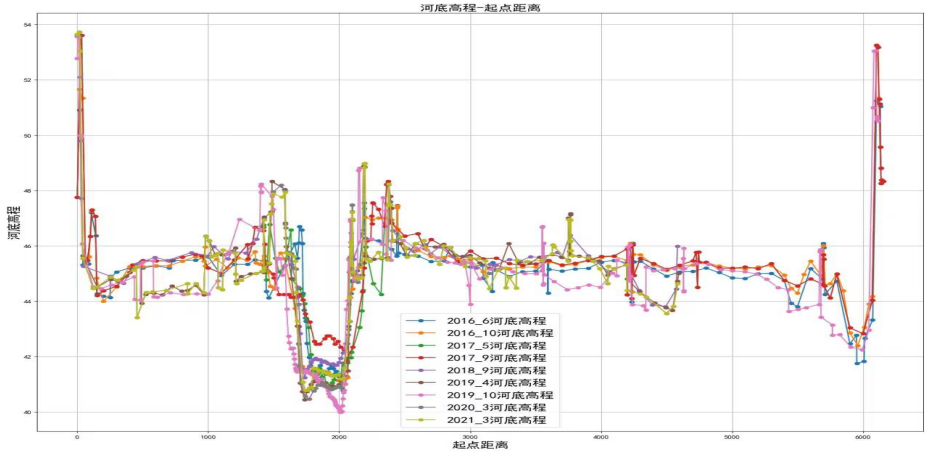


图 21 各年河底高程趋势图

结合文献,日常进行调水调沙一般在六月至七月之间,可推断 6 月份至 7 月河底高程会出现变化。

根据相关文献,可得 6~7 月份正值黄河下旬汛期,因此造成降水量的增多,在结合问题一中泥沙的曲线变化图,可得泥沙含量在 6~7 月份出现较大波动,结合问题二中水沙通量的变化,降水趋势变化及泥沙含量变化,可以推断得出由于汛期的到来,降水量的增多,使得水流冲击能力变强,带走大量泥沙

而 7 月至 12 月之后水沙通量曲线趋于稳定,排沙量曲线也较 6~7 月份出现明显降低,可得 7 月份后调水调沙工程临近收尾。

最后 12 月~次年 6 月的水沙通量变化较 7~12 月曲线图波动明显,且可观察得出 12 月~次年 6 月的含沙量较 7~12 月含沙量出现增长,再结合 12 月至次年 6 月黄河流域的气象观察,可得知 12 月至 6 月,黄河流域存在凌汛及春汛两次汛期,由此可推断得出 7~12 月调水调沙工程完成后,12 月至 6 月又因汛期带来大量泥沙,导致河床的抬高。

表 6 各年日河底高程平均值

日期	河底高程平均值/m
2016/6/08	44.7914234
2016/10/20	45.4634034
2017/5/11	45.2419559
2017/9/05	45.6101908
2018/9/13	45.1810000
2019/4/13	45.0398211
2019/10/15	44.1786111
2020/3/19	45.1481295
2021/3/14	45.0826619

为了更好地观察河底高程的变化情况,且结合附件 2 所给出的相关数据,首先为保证更加准确计算邻近两年数据的差值,对每年河底高程数据进行平均化处理,得出平均值如表 7。

表 7 各年日河底高程平均值

日期	河底高程平均值/m
----	-----------



2016/6/08	44.7914234
2016/10/20	<b>45.4634034</b>
2017/5/11	45.2419559
2017/9/05	<b>45.6101908</b>
2018/9/13	45.1810000
2019/4/13	<b>45.0398211</b>
2019/10/15	44.1786111
2020/3/19	<b>45.1481295</b>
2021/3/14	45.0826619

再者通过平均值来作为代表分析邻近两年数据的差值，通过差值之间的联系，结合上述表达及国家水文局相关数据可知黄河调水调沙一般启动在在 6 月 19 日至 7 月 19 日时间区间内。

表 8 各年日河底高程平均值

日期	日期	差值
2016/6/08	2016/10/20	0.67198000
2016/10/20	2017/5/11	<b>-0.22144750</b>
2017/5/11	2017/9/05	0.36823496
2018/9/13	2019/4/13	<b>-0.42919080</b>
2019/4/13	2019/10/15	-0.86121000
2019/10/15	2020/3/19	<b>0.96951839</b>
2020/3/19	2021/3/14	-0.06546760

通过各年河底高程的差值变换趋势分析，得出 2018 年数据无法衔接上下年关系，因此判断 2018 年数据为异常值，进行剔除分析。

剔除后结合表中数据，根据上述调水调沙整体时间段，将 2019 年 4 月 13 日与 2020 年 3 月 19 日河底高程对比，得出未进行调水调沙河底高程变化趋势一致，将 2019 年 4 月 13 日与 2019 年 10 月 15 日河底高程对比，得出调水调沙后出现明显变化，因此调水调沙效果显著。

综上 6 月-7 月，河底高程明显降低，但 7 月后随着水流的冲刷沉积，河底高程又会逐渐上升。

根据国家水文数据，得出黄河流经黄土高原，每年进入黄河下游的泥沙多达 16 亿吨，其中有 4 亿吨将沉积河底，致使下游河段平均每年以 10 厘米的速率抬升。多年研究表明，黄河下游洪水在大余 2500 立方米每秒时，加上其他相关因素能够对下水域起到重大作用，从而减少黄河淤积，甚至达到充裕平衡，由此可从根本上遏制河床抬高，因此调水调沙有必要。

通过图 22 各年年总排沙量的变化趋势增长，得出 2016 年及 2017 年较后 4 年变化幅度平缓，可推算出 2016 年至 2017 年相较于后 4 年调水调沙工程较弱。

在此基础上，假设 2016 年与 2017 年调水调沙工程并未进行，以 2017 与 2016 年之间差值作为河底淤积一年深度预估初始值，从而估算 10 年未进行调水调沙河地高程的变化增加 1.47m。

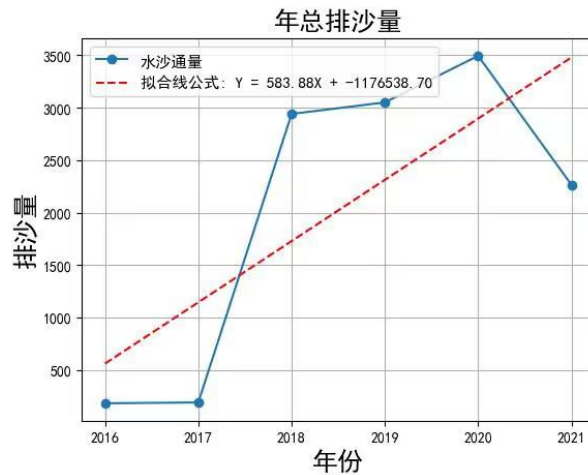


图 22 各年年总排沙量的变化趋势增长图

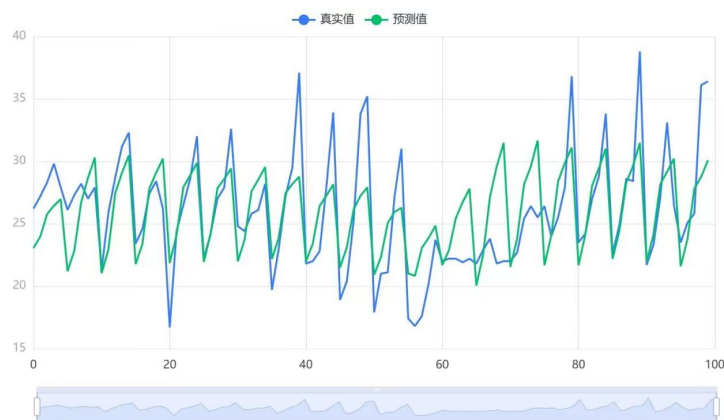


图 23 水底高程拟合图

由上图 23 可以得出起点距离、测点水流速、测点水深还有水深四个数据的拟合效果很好，由此可以得出水流量越快，水位就越高，沉淀的泥沙越多。

黄河泥沙顺着水流沉积在河床，调水调沙的目的是为了打破黄河水沙的时空分布规律，降低水文站泥沙淤积，所以如果不进行“调水调沙”，水底高程将会提高。

## 六、 模型评价与推广

### 6.1 模型评价

对于神经网络处理含有季节性规律的数据：神经网络对季节性数据要进行适当的预处理。这包括进行季节性差分以减少季节性影响，并进行数据标准化或归一化，以确保所有特征具有相似的尺度选择适当的神经网络架构对于模型的性能至关重要。需要决定网络的层数、每个隐藏层的神经元数量以及激活函数的选择。不同的架构可能会对季节性数据产生不同的影响。

对于季节性数据，考虑将季节性信息引入神经网络模型。一种方法是使用虚拟变量来表示季节性，或者将季节性信息作为附加的输入特征对于 ARIMA 模型处理含有季节性规律的数据：要确定时间序列数据的季节性周期，这是数据中季节性模式的重要特征，通常以季度、月份或周为单位。

对数据进行季节性差分,以减少季节性影响。季节性差分意味着将数据减去一定季节性周期前的值。这可以通过使用 `diff` 函数来实现,通常使用周期性差分的次数表示为  $d$ 。还要使用 ACF 和 PACF 图来帮助确定 ARIMA 模型的  $p$  和  $q$  参数。这些图可以帮助识别自回归和移动平均部分的阶数。观察 ACF 和 PACF 图中的显著滞后值,这些值可能表示 AR 和 MA 部分的阶数。在处理季节性数据时,通常需要使用季节性 ARIMA 模型。季节性 ARIMA 模型包括季节性差分、季节性自回归部分 ( $P$ )、季节性移动平均部分 ( $Q$ ) 和非季节性自回归部分 ( $p$ )、非季节性移动平均部分 ( $q$ ) 等参数。根据分析确定这些参数的值。ARIMA 模型的性能在很大程度上取决于选择的阶数,包括自回归阶数 ( $p$ )、差分阶数 ( $d$ ) 和移动平均阶数 ( $q$ )。对于季节性数据,ARIMA 模型的性能对这些阶数的选择非常敏感,因此 ARIMA 模型对比于神经网络更适合处理。

### 6.1.1 模型的优势

ARIMA 模型基于对时间序列数据的自回归 (AR) 和移动平均 (MA) 成分进行建模。这些成分有助于捕捉数据中的趋势、季节性以及随机波动。还能够生成预测区间,这些区间可用于识别模型的不确定性范围,而不仅限于提供单一点预测值。

最小二乘法提供了对模型参数的最优估计,这些估计是使得观测数据与模型的拟合误差平方和最小化的估计值,这些估计值有良好的统计效果。

Mann-Kendall 检验是一种非参数检验方法,不需要对数据分布进行假设且可以有效地检测时间序列中的弱趋势,即使趋势的形状不明显

### 6.1.2 模型的劣势

ARIMA 模型的局限性,对非线性关系的拟合较差,对于某些复杂的非线性时间序列模式可能无法很好地拟合和捕捉趋势,此外,ARIMA 模型在处理时间序列中的异常值或离群点时可能表现不佳,因为它对数据的拟合受到过去观测值的线性组合的约束。

最小二乘法的限制,它对数据的一些假设,如线性性和正态分布误差的假设,可能在某些情况下不成立。因此,在使用最小二乘法之前,需要仔细考虑数据的性质和适用性。

Mann-Kendall 检验的限制,它不能提供关于趋势的具体信息,只能告诉我们是否存在趋势,所以在应用时需要综合考虑数据的特点和分析的目的。

### 6.1.3 模型推广

本文建立的模型与实际情况高度吻合,具有显著的指导价值。此外,我们还可以对模型进行更深入的分析,揭示其在解决问题时的内在机制。该模型在预测方面表现出色,精度较高,且在不同情境下的泛化性能也出众。

## 参考文献

- [1]王俊杰,拾兵,巴彦斌.近 70 年黄河入海水沙通量演变特征[J].水土保持研究,2020,27(03):57-62+69.DOI:10.13869/j.cnki.rswc.2020.03.009.
- [2]李海华,邢静,李喜柱等.基于 BP 神经网络的黄河小浪底济源断面水质评价

模型研究[J].节水灌溉,2014(06):57-59.

[3] 李聚方,宋新龙.黄河下游高程基准面拟合探讨[J].水利水电技术,2010,41(09):105-107.DOI:10.13928/j.cnki.wrahe.2010.09.004.

[4] 卢金友,朱勇辉.水利枢纽下游河床冲刷与再造过程研究进展[J].长江科学院院报,2019,36(12):1-9.

[5] Kong D, Miao C, Borthwick A G L, et al. Evolution of the Yellow River Delta and its relationship with runoff and sediment load from 1983 to 2011[J]. Journal of Hydrology, 2015, 520: 157-167.

[6] Bi N, Sun Z, Wang H, et al. Response of channel scouring and deposition to the regulation of large reservoirs: A case study of the lower reaches of the Yellow River (Huanghe)[J]. Journal of Hydrology, 2019, 568: 972-984.

[7] Yin S, Gao G, Ran L, et al. Spatiotemporal variations of sediment discharge and in-reach sediment budget in the Yellow River from the headwater to the delta[J]. Water Resources Research, 2021, 57(11): e2021WR030130.

## 附 录

```
## 插值法填充
```

```
import pandas as pd
```

```
df = pd.read_excel('附件 1.xlsx', sheet_name="2020")
```

```
df['时间'] = df['时间'].str.strip()
```

```
df['时间'] = pd.to_datetime(df['时间'], format='%H:%M', errors='coerce')
```

```
df['含沙量(kg/m3)'].interpolate(method='linear', inplace=True)
```

```
print(df)
```

```
df.to_excel('data.xlsx', index=False)
```

```
## 绘制折线图
```

```
import matplotlib
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
plt.rcParams['font.sans-serif'] = ['SimHei'] # 使用中文字体
```

```
matplotlib.rcParams['axes.unicode_minus'] = False # 显示负号
```

```
M = pd.read_excel("2022-2023 预测.xlsx")
```

```
plt.plot(np.arange(1,13), M[2022], label='2022 年水沙通量预测值', marker='o')
```

```
plt.plot(np.arange(1,13), M[2023], label='2023 年水沙通量预测值', marker='o')
```

```
plt.legend(fontsize=18)
```

```
plt.title('水沙通量预测', fontsize=18)
```

```
plt.xlabel('月份', fontsize=18)
```

```
plt.ylabel('水沙通量', fontsize=18)
```

```
plt.xticks(np.arange(1, 13, 1))
```

```

plt.grid(True)

plt.show()

##小波相干分析图代码
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import pywt
from scipy import signal

np.random.seed(0)
t = np.linspace(2016, 2022)

data = pd.read_excel("2016-2021.xlsx")
df = {'Signal1': data["水沙通量(kg/s)_2016"], 'Signal2': data["水沙通量(kg/s)_2017"], 'Signal3': data["水沙通量(kg/s)_2018"], 'Signal4': data["水沙通量(kg/s)_2019"], 'Signal5': data["水沙通量(kg/s)_2020"], 'Signal6': data["水沙通量(kg/s)_2021"]}

wavelet = 'db4'
level = 4
freqs = np.linspace(1, 128, 128)
cwt1 = signal.cwt(df['Signal1'], signal.ricker, freqs)
cwt2 = signal.cwt(df['Signal2'], signal.ricker, freqs)
cwt3 = signal.cwt(df['Signal3'], signal.ricker, freqs)
cwt4 = signal.cwt(df['Signal4'], signal.ricker, freqs)
cwt5 = signal.cwt(df['Signal5'], signal.ricker, freqs)
cwt6 = signal.cwt(df['Signal6'], signal.ricker, freqs)
cross_power = np.abs(cwt1 * cwt2 * cwt3 * cwt4 * cwt5 * cwt6.conj())

plt.subplot(2, 1, 2)
plt.imshow(cross_power, extent=[t.min(), t.max(), freqs.min(), freqs.max()],
           aspect='auto', cmap='jet', origin='lower')
plt.title('Wavelet Coherence')
plt.xlabel('Time')
plt.ylabel('Frequency')
plt.colorbar(label='Coherence')

plt.tight_layout()
plt.show()

```

```

##总排沙量折线图与拟合线
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.linear_model import LinearRegression

plt.rcParams['font.sans-serif'] = ['SimHei']
M = pd.read_excel("")

plt.plot(np.arange(2016, 2022, 1), M["水沙通量(万吨)"], label='排沙量',
marker='o')

X = np.arange(2016, 2022, 1).reshape(-1, 1)
y = M["水沙通量(万吨)"].values.reshape(-1, 1)
reg = LinearRegression().fit(X, y)
y_pred = reg.predict(X)
slope = reg.coef_[0][0]
intercept = reg.intercept_[0]
fit_equation = f'拟合线公式: Y = {slope:.2f}X + {intercept:.2f}'
plt.plot(np.arange(2016, 2022, 1), y_pred, color='red', linestyle='--',
label=fit_equation)
plt.legend(loc='upper left', fontsize=11)
plt.title('年总排沙量', fontsize=18)
plt.xlabel('年份', fontsize=18)
plt.ylabel('排沙量', fontsize=18)
plt.grid(True)
plt.show()

print(fit_equation)

```

### 热力图

```

import scipy.stats as stats
df = pd.read_excel("")

```

```

plt.rcParams['font.sans-serif'] = ['SimHei']

```

```

corr = df.corr()

```

```

fig, ax = plt.subplots(figsize=(10, 8))

```

```

cax = ax.matshow(corr, cmap='coolwarm')

fig.colorbar(cax)

ax.set_xticks(np.arange(len(corr.columns)))
ax.set_yticks(np.arange(len(corr.columns)))
ax.set_xticklabels(corr.columns, rotation=45)
ax.set_yticklabels(corr.columns)

for i in range(len(corr.columns)):
    for j in range(len(corr.columns)):
        ax.text(j, i, f"{corr.iloc[i, j]:.2f}", ha="center", va="center", color="w",
fontsize=10)

ax.set_title('各变量间的相关性热力图', pad=20)
plt.show()

```

## 河底高程-起点距离图

```

import pandas as pd
import matplotlib
import matplotlib.pyplot as plt

df = pd.read_excel()
column_names = df.columns.tolist()

year_distance = []
year_hight = []

for i in range(0, 18, 2):
    distance = df.iloc[1::, i].tolist()
    distance = [x for x in distance if not pd.isna(x)]
    hight = df.iloc[1::, i + 1].tolist()
    hight = [x for x in hight if not pd.isna(x)]
    year_distance.append(distance)
    year_hight.append(hight)

plt.rcParams['font.sans-serif'] = ['SimHei']
matplotlib.rcParams['axes.unicode_minus'] = False

```



```

plt.figure(figsize=(20, 16))

year_list =
["2016_6","2016_10","2017_5","2017_9","2018_9","2019_4","2019_10","2020_3","
2021_3"]
for i in range(9):
    plt.plot(year_distance[i],year_hight[i], label=f'{year_list[i]} 河底高程 ',
marker='o')

print(len(year_distance))
plt.legend(fontsize=18)
plt.title('河底高程-起点距离', fontsize=18)
plt.xlabel('起点距离', fontsize=18)
plt.ylabel('河底高程', fontsize=18)
plt.grid(True)

## 训练数据归一化
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

excel_file = pd.ExcelFile('train_data.xlsx')

columns_to_normalize = ['水位(m)', '流量(m3/s)', '含沙量(kg/m3)', '水沙通量
(kg/s)']

normalized_train_data = pd.DataFrame()
normalized_test_data = pd.DataFrame()

for sheet_name in excel_file.sheet_names:
    data = pd.read_excel(excel_file, sheet_name=sheet_name)
    print(data)

    scaler = MinMaxScaler()
    data[columns_to_normalize] =
scaler.fit_transform(data[columns_to_normalize])

    if '2021' in sheet_name:
        normalized_test_data = data
    else:
        normalized_train_data = pd.concat([normalized_train_data, data])

print("归一化后的训练数据:")

```

```
print(normalized_train_data)
print("\n 测试数据:")
print(normalized_test_data)
```