

基于分段线性回归的黄河水沙通量变化规律

摘要

黄河是中国重要的河流之一，其水资源对我国经济和社会发展有着重要的意义。研究黄河水沙通量变化趋势，对其调水调沙与防洪防灾，协调人地关系具有一定帮助。

针对问题一，根据附件 1 中近 6 年的监测数据，本文旨在分析含沙量与时间、水位、水流量三者的关系，并计算每年的总输沙量和总水流量。通过研究含沙量与时间，水位，水流量的关系后，其次计算得到每一天的总水量，再对每天的总水量求和即可得到 2016 年至 2021 年的年水流量分别为 14373534260、15373480890、38934144900、38766246670、43383266500、47311947540 m^3 。根据已有的含沙量数据与水流量求出排沙量，得到一年的排沙量，因此 2016-2021 年的年总排沙量为 18239.57225、19104.33858、294456.701、305299.6226、352592.105、226363.7499 kt 。

针对问题二，本文需要分析水沙通量与突变性、季节性与周期性的关系以及变化规律。绘制全年的图形得到径流量与输沙量有着密切的相关性，对再对每天的水沙通量数据按照不同季节进行求平均，则夏季的水沙通量较大，说明夏季为调水调沙的时间与实际相结合夏季为调水调沙工程进行的时间；再次，求出水沙通量的周期性，每年的水沙通量有一定的周期性，周期大约为 21 天；最后，水沙通量会受突变点、季节因素和周期性的影响。

针对问题三，采用分段线性回归方法用 2018 年和 2019 年预测 2022 年数据、2019 年和 2020 年预测 2023 年数据。然后，给出年平均水沙通量结果，详表见正文**错误!未找到引用源。**。发现 2023 年径流量比 2022 年径流量高 9415352.8。整体来看，调水调沙的效果明显。

针对问题四，首先将去除 6 月和 7 月数据后历史数据作为预测数据。接下来，将预测结果与历史数据进行比较，以评估调水调沙的效果。为此，提取了附件 2 和附件 3 中的数据，并利用平均积分方法计算出河底高程。选择调水调沙工程后的数据和下一年调水调沙工程前的自然增长河底高程进行比较。计算结果显示，日平均增长率为 0.0012，十年后河底高程总增长为 5.5 米。这表明调水调沙对小浪底水库的河底高程产生了显著影响。

一、问题背景与问题重述

1.1 问题背景

黄河，作为中华民族的母亲河，不仅是农业灌溉、工业生产、居民生活用水的重要来源，也是中国北方生态系统的关键组成部分。近年来，气候变化和人类活动的加剧使得黄河水沙通量呈现出复杂多变的特征，这对黄河流域的生态平衡、水资源管理和防洪减灾带来了新的挑战。

1.2 问题重述

针对黄河某水文站近六年的水位、水流量与含沙量监测数据，本研究旨在：
问题 1 探索水文站黄河水的含沙量与时间、水位、水流量之间的关联性，并估算该水文站近六年的年总水流量和年总排沙量。

问题 2 分析该水文站水沙通量的突变性、季节性和周期性特性，揭示水沙通量变化的基本规律。

问题 3 基于水沙通量的变化规律，预测该水文站未来两年的水沙通量趋势，并设计最优的采样监测方案，确保监测的效率与成本效益。

问题 4 结合水沙通量和河底高程的数据，评估每年 6-7 月小浪底水库“调水调沙”操作的实际成效，并预测如果不进行调水调沙，十年后河底高程的变化情况。

通过上述研究，旨在为黄河流域的水资源管理、生态环境保护和灾害预防提供科学依据和决策支持。

1.3 已知条件

1. 2016-2021 年黄河某水文站的水位、水流量和含沙量监测数据。
2. 相同时间段内的黄河断面测量数据。
3. 气候和土地利用情况在研究期间相对稳定。
4. 小浪底水库每年 6-7 月进行“调水调沙”。
5. 监测成本与采样次数成正比。

二、模型假设

1. 假设黄河在该水文站的水位、水流量和含沙量的关系是线性的，即含沙量与水位、水流量之间存在一定的函数关系。
2. 假设黄河水沙通量的变化主要受自然因素影响，如降雨、蒸发、地形等，不考虑人为因素（如水库调度、河道整治等）的影响。
3. 假设近 6 年的气候条件、流域内地表覆盖和土地利用情况相对稳定，不会对水沙通量产生显著影响。
4. 假设黄河断面的形状和尺寸在监测期间保持不变，河床材质和河床侵蚀速率相对稳定。
5. 假设小浪底水库的“调水调沙”措施对黄河水沙通量的影响具有短期效应，且每年 6-7 月的调度效果相似。
6. 假设未来两年内，气候条件、流域内地表覆盖和土地利用情况、河道形态等不会发生显著变化，可以依据现有数据预测水沙通量变化趋势。
7. 假设采样监测成本与监测次数成正比，监测成本不受监测时间、地点等其他因素影响。
8. 假设黄河中水的密度为 $2.5 \times 10^3 \text{ kg} / \text{m}^3$
9. 假设在一段时间内水位，流量，含沙量和之前一段时间内都是一致的

三、符号说明

符号	说明	单位
Q	水流量	m/s
S	输沙量	Kg
P	排沙量	Kg
C_i	第 <i>i</i> 天是否监测 ($i = 0,1$)	天
A_i	第 <i>i</i> 天的总输沙量	Kg
\bar{x}	标准差	-

四、问题模型一建立与求解

根据附件 1 中黄河某水文站 6 年实际检测数据，研究该水文站黄河水的含沙量与时间、水位、水流量的关系，并估算近 6 年该水文站的年总水流量和年总排沙量。

4.1 数据预处理

首先对数据进行汇总处理

1. 将附件 1 中缺失的年份日期补充齐全。
2. 附件 1 中存在 6 个数据集，将日期补充齐全后，对其进行合并。
3. 删除重复的数据。

4.1.1 数据处理图表

先将各年的数据进行处理，并对处理好之后的数据进行汇总等操作, 可以发现数据的含沙量在每天的 8:00 时候监测的数据最多，因此本文利用 8:00 的实际监测数据研究黄河河水含沙量与时间，水流量，水位的关系，得到数据如以下图表所示。

÷	年	÷	月	÷	日	÷	时间	÷	水位(m)	÷	流量(m ³ /s)	÷	含沙量(kg/m ³)	÷
0	2016.0		1.0		1.0		0:00		42.79		357		0.825	
1	2016.0		1.0		1.0		4:00		42.80		363		0.825	
2	2016.0		1.0		1.0		8:00		42.80		363		0.796	
3	2016.0		1.0		1.0		12:00		42.81		368		0.796	
4	2016.0		1.0		1.0		14:00		42.84		384		0.796	
...	
2900	2021.0		12.0		31.0		4:00		43.07		1070		3.120	
2901	2021.0		12.0		31.0		8:00		43.09		1110		2.390	
2902	2021.0		12.0		31.0		12:00		43.10		1120		2.390	
2903	2021.0		12.0		31.0		16:00		43.09		1110		2.390	
2904	2021.0		12.0		31.0		20:00		43.07		1070		2.390	

图表 1 数据监测结果

通过整理附件 1，可以发现在 2017 年中数据在 8:00 中存在有部分缺失的情况，其次在 2021 和 2022 年中，存在日期大部分缺失，在 2017 年 1 月 12 日到 1 月 17 日这 6 天并没有监测数据，但是可以发现在 14:00 中监测的数据中存在水位和流量相同，2021 年和 2022 年中天数相差较多，本文采取已知天数进行分析及解答。

4.1.2 模型建立

线性回归是利用数量统计中回归分析，来确定两种或两种以上变量之间相互依存的定量关系的一种统计分析方法。其表达式为 $y=aX+b$, b 为误差服从均值为零的正态分布，回归分析中，只包括一个自变量和一个因变量，且二者关系可用一条直线近似表示，称为线性回归。

4.2 研究水位和含沙量的关系

1. 数据准备

对之前预处理之后 6 年的所有数据，筛选剔除含沙量缺失的行数，从中选取含沙量，水位所在列，作为本次研究数据。

2. 组距分组

在监测数据中，需要将水位的变化范围分成若干区间，每个水位归属与相应的区间，故而进行组距分组，组距分组时，为了更好的获得组距，采用经验公式。

(1) 确定极差: 极差是数据中的最大值与最小值之差。

$$R = \text{最大值} - \text{最小值}$$

(2) 选择组距: 组距是每个组的宽度，通常根据数据的分布和极差来选择。组距的选择可以主观确定，也可以使用以下公式作为参考：

$$d = \frac{R}{k}$$

其中 d 是组距， R 是极差， k 是组数。

(3) 确定组数: 组数 k 的确定可以参考以下经验公式，也称斯特奇斯规则：

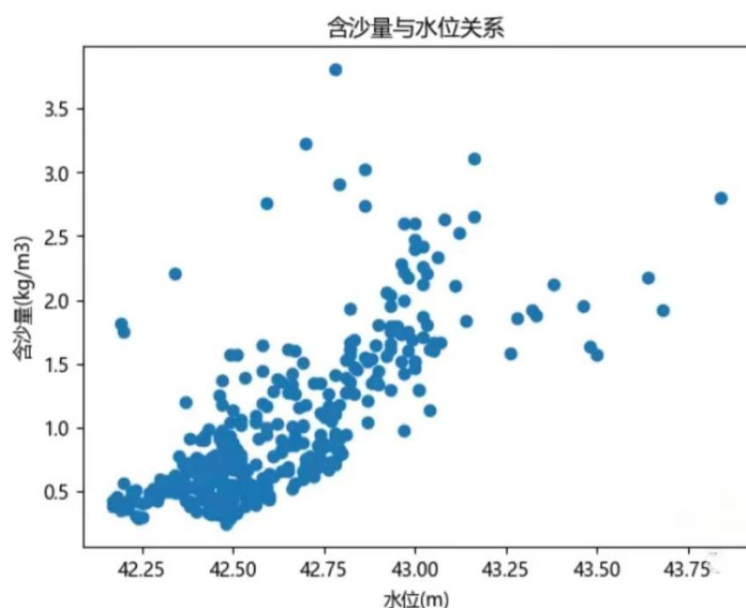
$$k = 1 + 3.322 \log_{10}(n)$$

其中 n 是数据的总个数。这个公式适用于数据量较大且数据分布比较均匀的情况。

(4) 确定组界和组中值

一旦确定了组距和组数，就可以确定每个组的上下限（组界），以及每个组的代表值（组中值）。组界：第一组的下限通常是小于或等于最小观测值的最大整数，上限是下限加上组距减去一个很小的数（通常是 0.1 或 0.01），以避免重叠。后续组的上下限依次类推。

$$\text{组中值} = \frac{\text{上限} + \text{下限}}{2}$$



图表 2 含沙量与水位离散图

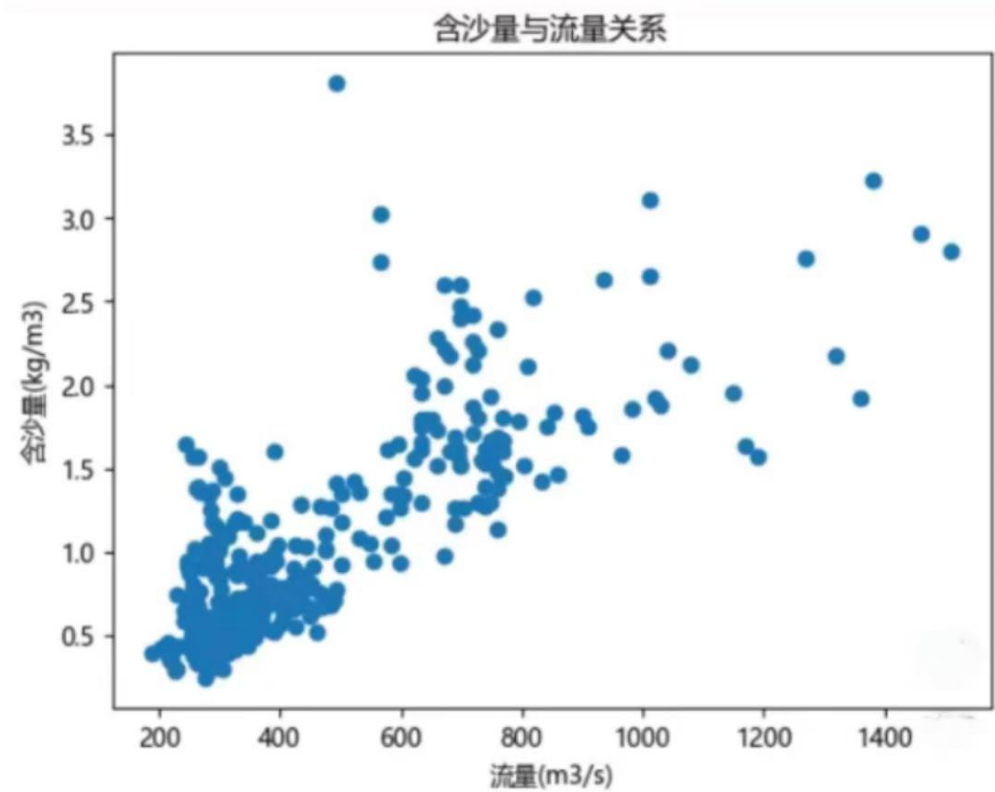
4.3 研究水流量和含沙量的关系

在前期筛选剔除含沙量缺失的行后保留的数据基础上，选取流量，含沙量所在列，作为本次研究对象。而问题一中需要研究“含沙量”与“水流量”的关系，从而补充附件 1 中含沙量对应的水流量作为研究数据，以下为水流量模型：

$$Q_{\text{水}} = Q(1 - \frac{Q_{\text{沙}}}{\rho_{\text{沙}}}), (\rho_{\text{沙}} = 2.58 \times 10^3 \text{kg/m}^3)$$

流量数据的特点与水位类似，属于较多的连续性变量，因此同样需要把水流量的值的变化范围分成若干区间，故而进行分组，通过经验公式 $R = \text{最大值} - \text{最小值}$ 进而获得各组上限和下限。

其变化趋势展示图如下：



图表 3 含沙量与水流量离散图

图中显示了不同流量的数据点与其对应的含沙量之间的关系。从图中可以看出，随着流量的增加，含沙量也呈现出一定的变化趋势。在较低的流量范围内（例如 200 至 400 立方米/秒），含沙量相对较高；而在较高的流量范围内（例如 1200 至 1400 立方米/秒），含沙量则逐渐降低。这可能表明在较低的流量下，河流中的泥沙更容易被水流携带，导致含沙量较高；而在较高的流量下，由于水流的冲刷作用，泥沙可能会沉积下来，从而导致含沙量降低。

4. 4 估算年总水流量和年总排沙量

对采用三次样条插值周的排沙量数据与时间存在关系，则依据插值函数对时间进行积分

$$\bar{y} = \int_1^{34536000} f(x) dt$$

采用上述积分公式，分别对年总水流量和总排沙量进行积分。

(1) 年总水流量计算

年水总流量表示为一年中通过该区域总水量。首先，附件 1 中的监测数据是以小时为单位监测，而流量的单位为 m^3/s ，其次，将每一秒的水流量进行求和即可得到该年的总水流量；最后，利用相同方式即可求出近 6 年中每一年的总水流量。

表 甲近 6 年每年的总流量

年份	年总水流量(m^3)
2016	14373534260
2017	15373480890
2018	38934144900
2019	38766246670
2020	43383266500
2021	47311947540

水流量从 2016 年到 2017 年有所增加，然后在 2018 年显著上升，随后在 2019 年略微下降，之后逐年持续增长直到 2021 年。最显著的增长发生在 2018 年，相较于 2017 年，水流量几乎翻了一倍。这可能是因为当年的降水异常丰富，或者有其他因素影响了水文循环，如水库的调度、气候变化等。从 2018 年开始，尽管增长率不如 2018 年到 2019 年之间的变化那么显著，但水流量仍然呈现出连续增长的趋势。

(2) 年排沙量计算

排沙量为一定时间内排出沙子数量，而排出的沙子数量与水流量和含沙量存在一定的关系，则排沙量的计算公式为：

$$P = L \times H \tag{1}$$

其中， P 表示为排沙量， L 为水流量， H 为含沙量。

年份	排沙量 (kt)
2016	18239.57
2017	19104.34
2018	294456.70
2019	305299.62
2020	352592.11
2021	226363.75

五、问题二模型建立与解析

径流量表示为单位时间内通过断面的水量；输沙量则表示在单位时间内通过断面的沙子数量，而在单位时间内的通过某一面积的径流量或者输沙量可以称之为水沙通量。

针对问题二，先计算出 2016 年-2021 年每年每天的径流量数据，利用径流量的数据可以求出每天的输沙量，利用求解的水沙通量数据对下文进行特性的分析。

径流量的计算公式为

$$w = \int_0^T Q dt$$

其中， Q 表示为水流量， T 表示为单位时间。输沙量的计算公式为

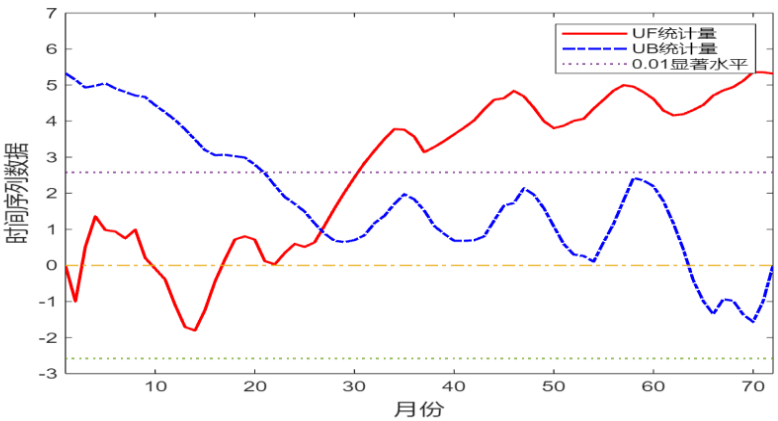
$$S = \int_0^T P dt$$

S 表示为某一单位的输沙量， P 为排沙量，结合两个计算公式即可得到水沙通量的结果。

5.1 近六年该水文站水沙通量突变性

突变性指的是在总体的样本数据中，可能存在某些数据发生急剧的变化，而发生变化的点可以称之为突变点。基于求解出每天的水沙通量，采用 Mann-Kendall 突变法对水沙通量的突变性进行检测，再找到的突变点进行分析，从而得到突变的区间范围。

为判定水沙通量是否存在显著上升或下降趋势，使用 M-K 检验法来判定。水通量的趋势检验结果如下图所示：

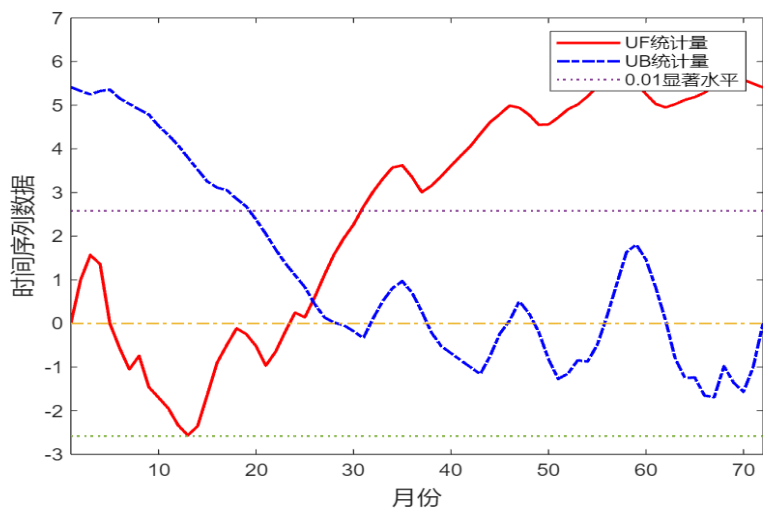


图表 4 水通量的趋势检验

从此图中可得出以下结论：

UF 值从第 23 月份开始稳定大于 0，从第 30 个月份开始超过临界线（显著水平为 0.01）。UF 和 UB 两条曲线在 27 时出现交点，并且靠近下临界线。在 27 时，水流量被分成两类，说明此时为突变点

以下为沙通量的趋势检验结果如下图所示：

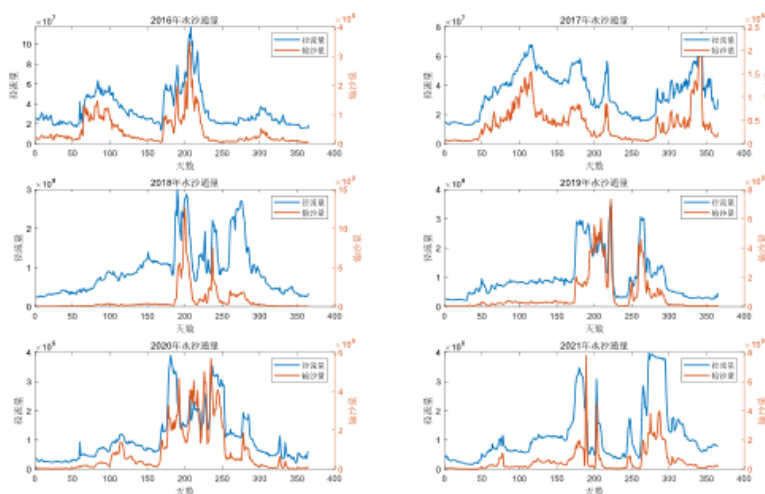


图表 5 沙通量趋势检验 I

可知沙通量趋势检验和水通量检验结果类似。而此时在 30 时沙流量被分成两类，此时为突变点

5.2 近 6 年该水文站水沙通量的季节性

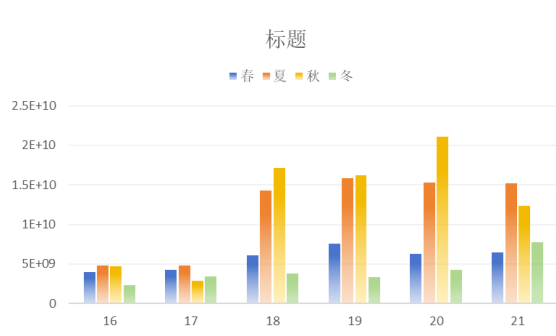
为了研究水沙通量的变化趋势和规律，进一步得到的径流量和输沙量数据与季节性存在的关系。



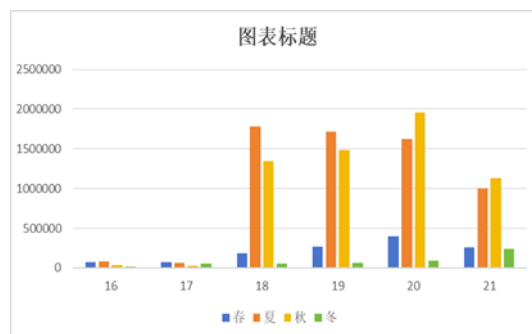
图表 6 各年水沙通量的差异

从每年的水沙通量图中，可以看出径流量与输沙量之间存在的密切的相关性，通过 2018-2021 年的趋势出现的天数大致相同，通过趋势分析在径流量上升的区间中可能为调水调沙的时候，而后半段上升可能是因为汛期，相比之下 16 年和 17 年的趋势差距较大，可能这两年中存在较多的影响因素。

为了进一步探求水沙通量与季节存在的关系，分别对径流量和输沙量按照四季对应的月份进行统计，并求出每一个季节的平均径流量和输沙量，通过比较不同季节给出水沙通量的一个变化。



图表 7 春夏秋冬平均输沙量

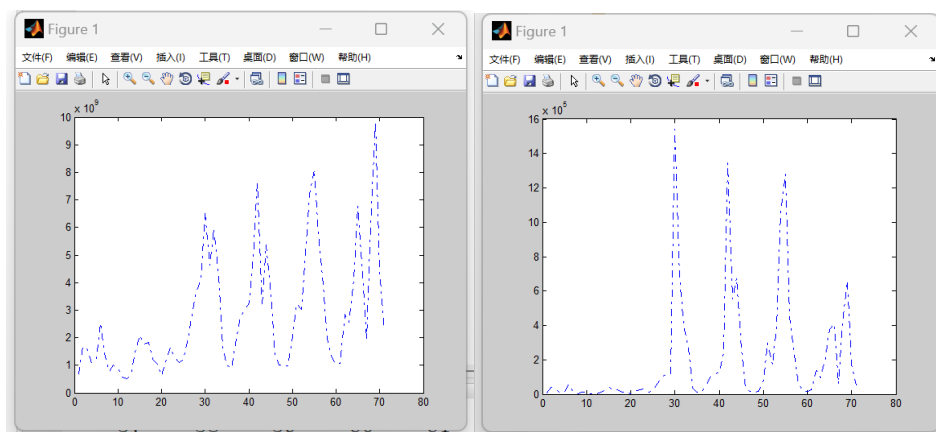


图表 8 春夏秋冬平均径流量

由上图可知，夏季的输沙量较大，通过查阅文献资料说明该水文站正是在夏季进行调水调沙，而秋季的平均输沙量也较大，那是因为汛期导致了输沙量较大。从径流量的趋势也基本符合输沙量的关系，说明水沙通量和季节存在着高度的相关，水文站需要根据季节的变化趋势实现调水调沙以及抗洪等工作。

5.3 近 6 年该水文站水沙通量的周期性

首先画出水沙通量的时序图，如下图所示：



从图中可得出以下结论：

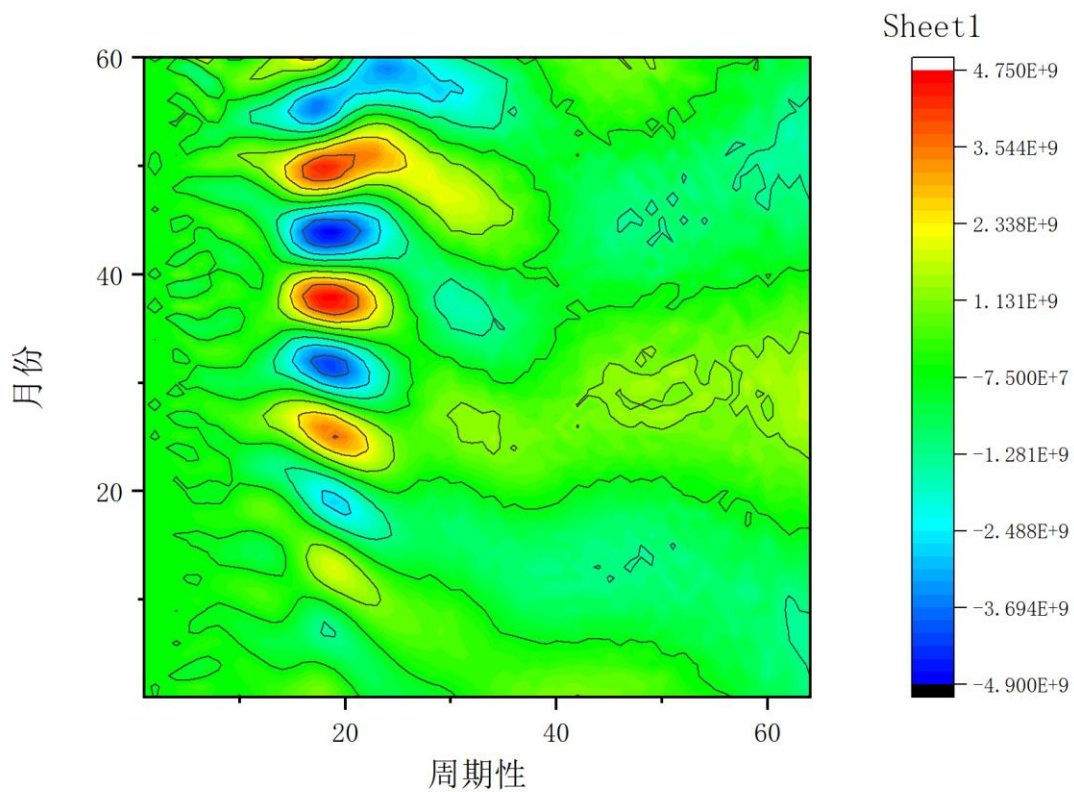
- (1) 水通量具有周期性，周期为 365 天，但在 2018 年有突变现象，呈上升趋势。
- (2) 沙通量具有周期性，周期为 365 天，但在 2018 年有突变现象，呈下降趋势。

5.4 水沙通量时间尺度变化特征

为了详细准确找出突变点和周期性，需要用到小波变换分析。

通过水沙量数据，使用 MATLAB 软件作小波变换，再使用 Origin 软件绘制出图形。

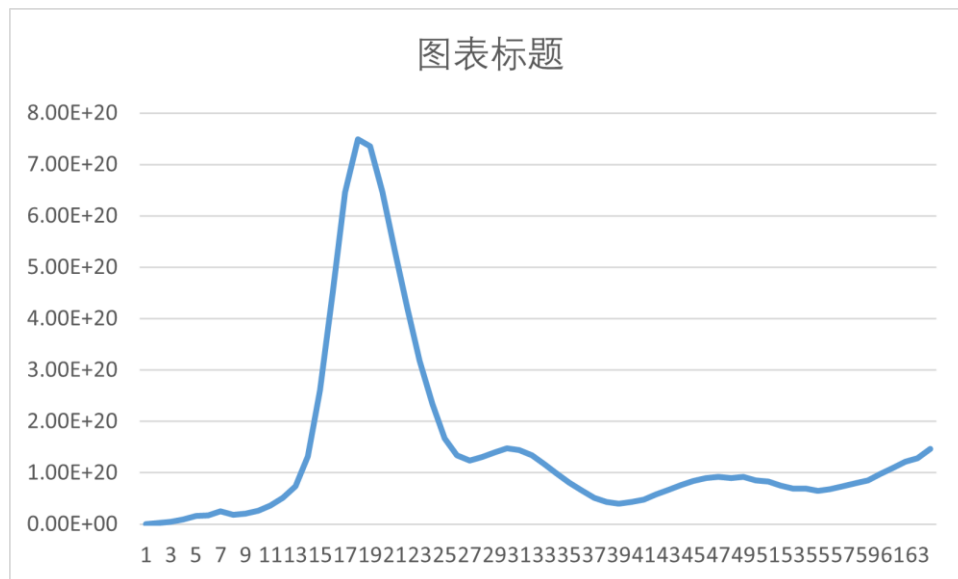
如下图为水通量等值线图：



图表 9 水通量等值线图

水通量在 45~60 月振荡明显，主要呈现三次正负循环交替。

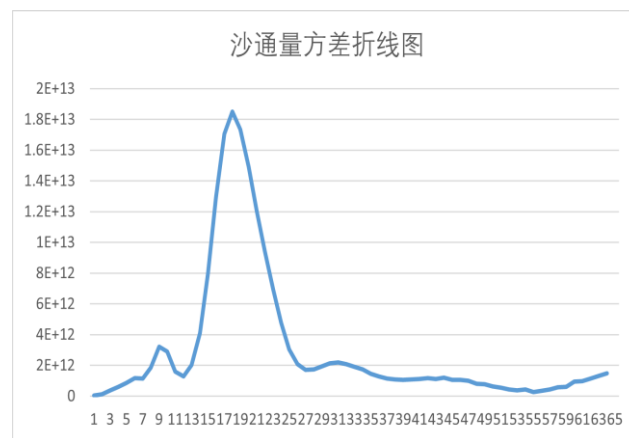
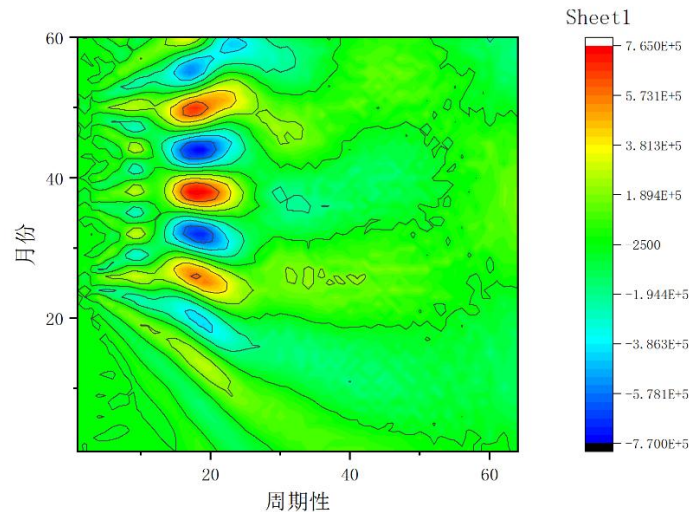
如下为水通量方差分析图：



图表 10 水通量小波方差分析图

水通量以第 20 月为周期振荡，且幅度巨大，其他时间，较为平缓。

画出沙通量等值线图和方差分析图，如下图所示：



从上图中可得出以下结论：

- (1) 沙通量振荡集中在第 15-25 月，并且振荡非常明显，其他与沙通量类似
- (2) 从第 25 月开始，振荡基本消失

六、问题三的模型建立与解析

基于问题二中的分析以及结合气象相关的资料，在 2021 中降水量突发的较为频繁，以至于 2021 年的水沙通量结果与其他的几年的相差较大，所以本文采用 2020 年的数据进行方法的选取。

分段线性回归法

分段线性回归基本思想是将数据点分为若干个区间，各区间的数据点需满足线性关系，对各区间进行回归分析，以此得到对应回归方程。结合问题二中水沙通量季节性的变化规律，分为春、夏、秋、冬四个季节，对每个季节进行回归分析。再进一步基于突变点，选择 7 个突变点将数据划分为八段分别对每段数据进行线性回归，在通过回归的预测出 2020 年一年的结果。

6.1 基于分段线性回归方法预测两年水沙通量的变化趋势

根据 2018 年和 2020 年径流量和输沙量数据预测未来两年的数据，为此采用分段线性回归方法。选择最接近年份下的数据，即用 2018 年和 2019 年预测 2022 年数据，2019 年和 2020 年预测 2023 年数据。

2022 年径流量	2022 年输沙量	2023 年径流量	2023 年输沙量
227957439.6	4864.976338	145342089.9	3247.354632
263109528.1	3928.710306	142957689.7	3021.527318
237720656	3480.15879	192104214.2	2981.405738
221701403.7	3479.811864	159410092.6	2949.641967
219060741.7	3397.10175	174618382	2917.009128
237298570.7	3385.582322	153344582.1	2893.327818
248674413.9	3330.540695	165051790.6	2853.249221
249696704	3226.613867	173615885.3	2843.147504
207575905.6	3161.515632	161125115.8	2758.629554
216813032.1	3047.999407	124048450.9	2730.808022
227355537	2975.402954	173904357.7	2704.841272

图表 11 水沙通量前 10 数据

利用分段回归的方式，可以对两年的水沙通量进行分析和比较。在比较两年的输沙量时，发现 22 年的输沙量和径流量前 10 的结果明显高于 23 年的输沙量。其中两者最大输沙量之间相差 1617.6217，初步说明在 2022 年中前 10 天的水沙通量的结果更好。

年份	年平均径流量	年平均输沙量（kg）
2022	103983338	851.2609546
2023	113398690.8	851.0619995

图表 12 年平均水沙通量

对于年平均的结果，两年的平均输沙量基本相等，但 23 年的径流量比 22 年的径流量高 9415352.8。综合一整年的数据结果，可以发现 2023 年的水沙通量效果会更加优于 22 年。这也从侧面说明调水调沙的效果更加明显。

6. 2 最小监测次数的建立与求解

七、问题四模型的建立与求解

针对问题四，本次采用常用的时间序列预测方法进行预测各年 6 月和 7 月的数据，将历史数据和预测数据对比作为“调水调沙”的实际效果，对附件 3 数据计算河底高程，将河底高程，日期和起点距离提取，计算不同起点距离平均河底高程，并得出结论

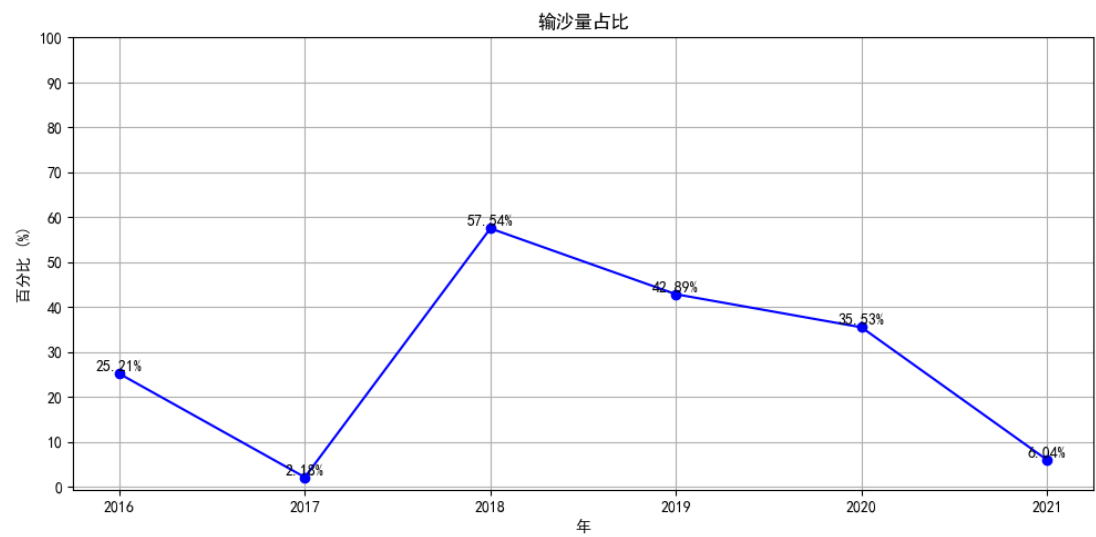
7.1 调水调沙的实际效果

根据题意分析各年 6 月和 7 月小浪底水库水沙通量“调水调沙”的实际效果，则要将历史水沙通量数据去除 6 月和 7 月数据，以剩余的历史数据作为需要预测的数据。其次用每一年各天预测后的数据相加得到每一年的输沙量的总和，再与除去各年 6 月和 7 月的调水调沙总和作差，可得到实际的调水调沙效果。如下：

年份 字段	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年
历史输沙量	15654.918	15858.440	264612.234	276203.898	310674.927	200795.842
预测输沙量	11707.535	15513.456	112362.737	157727.265	200278.846	188667.762
实际输沙量	3947.383	344.984	152249.497	118476.633	110396.081	12128.080

图表 13 调水调沙效果

通过图表可以得出 2018 年的调水调沙效果是最好的，接着进一步求出输沙量占比，如下图所示：



图表 14 实际输沙量占比

从占比角度来看，2018 年的输沙量占比最高，说明该年输沙量相当大，其次是 2019 年的输沙量，最小的是 2017 年，说明调控效果不太明显，但也存在一定影响。

7.2 调水调沙对河底高程的影响

首先，附件 1 中不涉及到河底高程，所以需要结合附件 2 和 3 的数据，对于

附件 3 中的并不存在河底高程，可以水位和水深两个数值计算得出河底高程，提取出附件 3 中的河底高程、起点距离以及日期三者的数据，将起点距离相同的与附件 2 进行积分则可以给出每个平均的河底高程，通过提取后的数据计算出 10 的河底高程变化。

$$\bar{y} = \frac{1}{a-b} \int_a^b f(x) dx$$

其中，该公式求解的为平均河底高程，[a,b]作为底，高为 \bar{y} 同地的矩形面积。

探究调水调沙工程内下降的高程，首先，根据相关的文献资料可以求出调水调沙的工程开始时间，其次利用高程数据，找到对应范围内的高程变化；最后，给出下降的高程大小。

区间	日期	河底高程	日均增长	年均增长
2021-2022	2022/4/18	41.20600345	0.001483184	0.541362272
2020-2021	2020/7/29	41.41214497	0.001259294	0.532642377
2019-2020	2019/10/15	41.2643389	0.001112031	0.40589127

图表 15 高程增长值

在调水调沙工程时间内，只有 2022 与 2021 年的数据存在两天连续的数据，则可以计算出两年的相邻天数每天平均提升河底高程 0.001095 和 0.00111m，调水调沙工程为期一般为 19 天，降低水位线大于 2m。

基于处理好的数据量，选择秋汛之后的数据进行河底高程的变化探究，选择提取后秋汛结束的数据以及河底高程进行分析。

年份	河底高程增长/m
2024	0.531398877
2025	0.532642377
2026	0.541362272
2027	0.542885672
2028	0.544228272
2029	0.558728272
2030	0.561049272
2031	0.562402872
2032	0.564044872
2033	0.566198872
合计	5.504941629

图表 16 未来 10 年河底高程

针对上述给出每年的增长情况，则可以得到不调水调沙后十年的河底高程增长至 5.504 左右，可以说明在调水调沙的加持下可以很好的控制河底的高程。从河底高程的趋势，如果不进行调水调沙的处理则河底的高程会随着时间的增长而增长，其中，第 5 年的河底高程上升趋势更加凸显，说明调水调沙能够很好的控制小浪底水文站的河底高程。

八、模型评价

本文分段线性回归模型的推广可以应用于经济学、环境科学等领域，例如分析经济指标的变化趋势，探索经济发展的阶段性特征、研究气候变化、水资源管

理等方面的数据，为预测和解释变化的趋势。适用于数据存在明显的变化点，可使用该模型来分析和预测数据的变化趋势。

参考文献

[1] 王一鸣.黄河水沙变化及其对黄河三角洲变化的影响分析[D].中国科学院大学(中国科学院教育部水土保持与生态环境研究中心),2018.

[2] 刘静,靳莉君,王鹏.2021 年黄河流域极端秋雨与大气环流和海温的关系[J].水资源与水工程学报,2022,33(05):36-44.