

基于时间序列预测的水沙检测数据优化模型

摘要

随着气候变化和人民生活的影 响,为了优化黄河流域水资源分配,本文基于黄河水沙流量、水位、含沙量数据的调查,确定自变量时间和因变量流量、水位、含沙量等指标,针对这些数据建立了以预测和优化为目标的 ARIMA、Mann-kendall 突变检验、季节性时间序列、pettitt 等模型,并使用 MATLAB, Python, Excel 对模型进行求解分析,解决了年总水流量,年总排沙量,水沙通量的变化规律,未来两年的监测方案,以及小浪底六、七月份的调水调沙实际效果和未来十年水文站的河底高程变化.

针对问题一: 首先对数据进行预处理,提取出每天的平均水位,平均水流量,平均含沙量,并用 Excel 将时间与含沙量,水位,水流量之间的关系可视化,通过 spsspro 分析含沙量,水位,水流量,时间之间的相关性,并根据每天的平均含沙量估算出年总水流量和年总排沙量.

针对问题二: 对于突变性,采用 Mann-kendall 趋势检验的方法,通过对 UF、UB 统计量的分析,发现突变点在六、七月之间.基于 Mann-kendall 趋势检验,我们又进行了 pettitt 突变点检验法,进行了再次验证.对于季节性和周期性,我们采用季节性时间序列模型,进行了数据分析,通过观察水沙通量的自相关函数图像和偏自相关函数图像,说明了时间序列存在着以 $S=12$ 为周期的季节性变动.并且具有短时间内的周期性变化规律.

针对问题三: 在问题二的基础上,我们使用 Mann-kendall 趋势检验的方法,预测了未来两年水沙通量的变化规律,以前 58 个时间序列点作为训练集,14 个时间序列点作为测试集,得到了精度为 76.3%的预测效果,并依此预测了该水文站未来两年的水沙通量变化趋势,同时结合该水文站水沙通量的突变性、季节性和周期性等特性,给出了该水文站未来两年的最优采样检测方案.

针对问题四: 首先要对每一年六七月份里的水沙通量和河底高层数据整体分析,根据 HEC-RAS 模型进行数据的分析,在依据过去几年的数据对未来十年的调水调沙情况进行分析处理.

关键词: Mann-Kendall 趋势突变; ARIMA 模型; 时间序列; 泥沙运动模型

一、 问题重述

1.1 问题背景

在黄河水沙通量的变化规律中，含沙量通常随着不同因素的变化而变化，如季节，水流量等，黄河水沙通量的变化通常具有不稳定性，从而导致排沙量的不确定性。

1.2 问题提出

与问题一：要求分析含沙量与水位、水流量、时间的关系，估算近 6 年的年总水流量与年总排沙量。

问题二：分析近 6 年该水文站水沙通量的突变性、季节性和周期性等特性，研究水沙通量的变化规律。

问题三：根据该水文站水沙通量的变化规律，预测该水文站未来两年水沙通量的变化趋势，为该水文站制订未来两年最优的采样监测方案，具体到采样监测次数、时间等，不仅要能及时掌握水沙通量的动态变化情况，而且能最大程度减少监测成本。

问题四：根据该水文站的水沙通量与河底高程变化情况，分析每年 6 至 7 月小浪底水库进行“调水调沙”的实际效果。若不进行“调水调沙”，预测 10 年以后该水文站的河底高程将会如何？

二、 问题分析

问题一：首先，需要考虑黄河水流量、时间、水位与含沙量之间的关系，计算出在每一年黄河水沙的变化趋势，将水位，水流量，含沙量随着时间的变化趋势进行可视化处理，并分析他们之间的相关性，由于 2016-2021 年每天的水位、水流量、含沙量的变化都不尽相同，根据分析出每一年的水位，水流量，含沙量变化趋势，通过每天平均水流量和排沙量来估算年总水流量和年总排沙量。

问题二：首先，对数据进行处理，收集每月该水文站水沙通量的数据，对时间序列点，建立检测模型，采用 Mann-kendall 趋势检验法进行突变点检测，并利用 pettitt 突变点检验法进行再次验证。采用季节性时间序列模型对季节性和周期性进行分析。

问题三：首先，根据对问题一问题二历史数据的分析了解水沙通量的变化规律，继续使用时间序列模型，预测未来两年的水沙通量趋势，制定未来两年的采样监测方案，确定每年每季采样监测频率，预测水沙通量的趋势，选择最佳的采样时间，采样地点，

与采样方法,帮助水文站预测未来两年的水沙通量的变化趋势又能及时掌握实时动态变化情况,减少成本资源.

问题四: 首先对每年的 6-7 月份小浪底水库水沙通量和河底高程的变化情况进行分析,分析出“调水调沙”的实际效果,使用泥沙数学模型,模拟水沙通量,流速,水位,河底高程,用初始条件模拟开始的水位,流速等参数.同样使用泥沙数学模型来模拟不进行“调水调沙”的操作估算出水流和含沙量的变化趋势.比较进行“调水调沙”和不进行“调水调沙”的情况下河底高程的变化趋势,预测未来 10 年内河底高程变化情况.

三、 模型假设与符号说明

2.1 模型基本假设

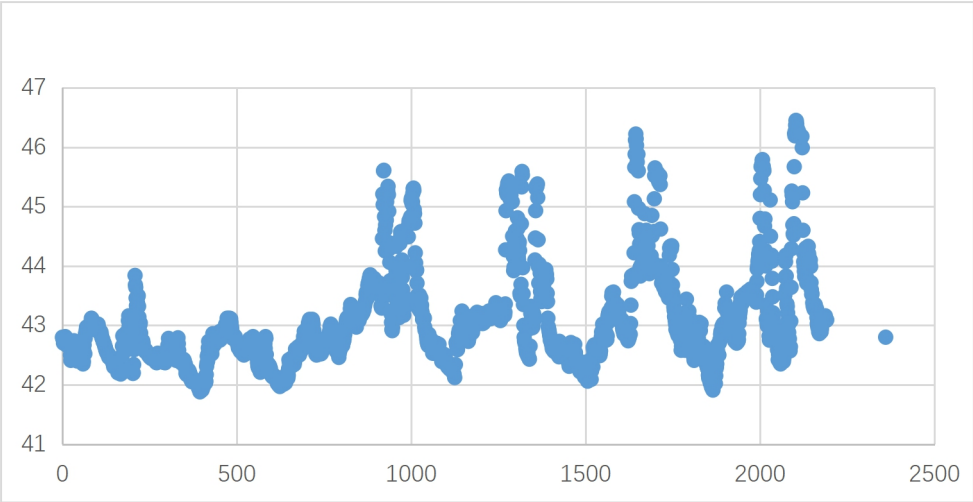
- (1) 每一年的自然灾害对于黄河通量的影响都一成不变;
- (2) 在研究时间范围内黄河通量不会因为人为因素而导致大幅度变化;
- (3) 黄河的水质对于每一年黄河的流量、水位,含沙量影响相同;
- (4) 每年监测的河底高程和起点距离的路程不变.

符号说明

符号	含义
a_t	水沙通量
S_t	季节因子
s	季节周期长度
S_{t+k-s}	最后一年的季节因子
G_t	t 月水流量观测值的趋势分量
Z_t	t 月含沙量观测值的周期分量
Y_t	t 月水位观测值

2.2 数据预处理

1. 首先对数据进行收集,保证整个数据的样本量足够多,保证数据的误差足够小.
2. 识别并处理数据中的异常值和缺失值.
3. 根据不同的模型对数据进行不同的处理,并进行筛选和分析.
4. 最后通过数据分析的结果大致推出走向趋势.



平均每年水位图

四.模型分析与数据建立

Spearman 相关性分析

	时间	水位	水流量	含沙量
时间	0.000(NaN)	0.000(NaN)	0.000(NaN)	0.000(NaN)
水位	0.000(NaN)	1(0.000***)	0.959(0.000***)	0.696(0.000***)
水流量	0.000(NaN)	0.959(0.000***)	1(0.000***)	0.677(0.000***)
含沙量	0.000(NaN)	0.696(0.000***)	0.677(0.000***)	1(0.000***)

注：***、**、*分别代表 1%、5%、10%的显著性水平

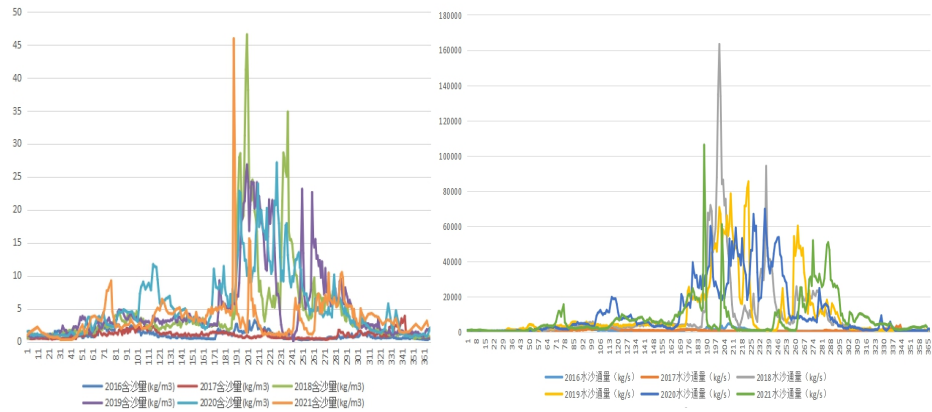


图 2.1

图 2.2

根据图 2.1、2.2 可看出含沙量与水沙通量近六年的走向趋势。

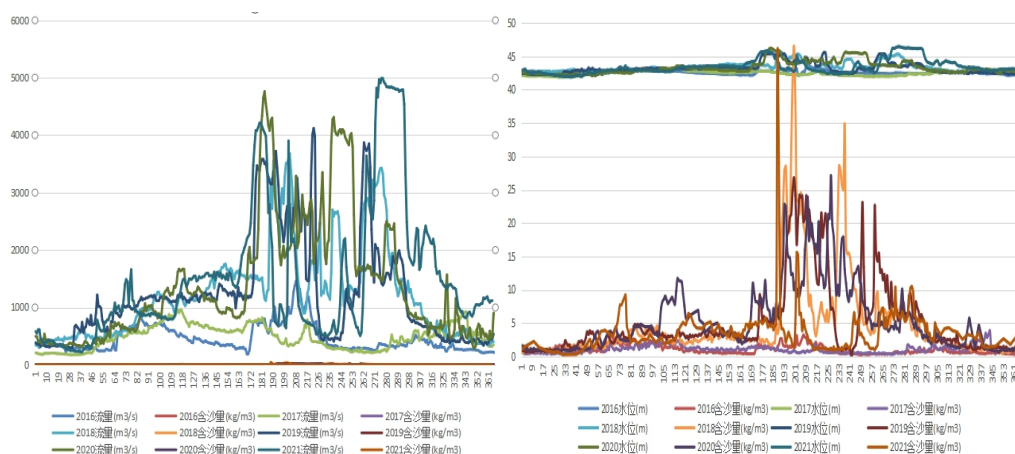


图 2.3

图 2.4

据图 2.3、图 2.4 可以看出近六年流量与含沙量的对比趋势图，以及水位与含沙量的对比趋势图。

年总水流量(m^3) = Σ (每个时间点的水流量(m^3/s))

年总排沙量(kg) = Σ (每个时间点的含沙量(kg/m^3) \times 水流量(m^3/s))

由上述图表(1)、(2)、(3)、(4)与模型可得：

水流量分别为：

2021: $5386611(\text{m}^3/\text{s})$

2020: $4951508(\text{m}^3/\text{s})$

2019: $4601865(\text{m}^3/\text{s})$

2018: $4406970(\text{m}^3/\text{s})$

2017: $1079887(\text{m}^3/\text{s})$

2016: $1089444(\text{m}^3/\text{s})$

排沙量分别为：

2016: $1.410359163000 \times 10^6(\text{t})$

2017: $1.340780092000 \times 10^6(\text{t})$

2018: $4.1052435549000 \times 10^7(\text{t})$

2019: $4.4062298938000 \times 10^7(\text{t})$

2020: $4.5299067361000 \times 10^7(\text{t})$

2021: $2.9971524220000 \times 10^7(\text{t})$

3.1 问题二模型建立与分析：

1. Mann-Kendall 检验

在时间序列趋势分析中，Mann-Kendall 检验不需要样本遵从一定的分布，也不受少数异常值的干扰，适用于水文，气象等非正态分布的数据，计算简便。

2. 季节性时间序列模型

季节性时间序列模型是一种用于分析和预测具有季节性模式的传统时间序列模型，可帮助我们理解时间序列数据中季节性波动的规律。

乘法模型：对于具有季节性周期变化规律的水沙通量模型，我们采用乘法模型描述：

$$Y_t = G_t \times Z_t \times E_t$$

时间 t 与年份 i 、月份 j 满足关系式：

$$t = (i - 1) \times 12 + j$$

式中 $i = 1, 2, \dots, N$ ； N 为观测数据年份个数； $j = 1, 2, \dots, 12$ 。

3.2 问题二模型的求解

MATLAB 实现 Mann-Kendall 趋势突变检验

Mann-Kendall 趋势检验在 Mann-Kendall 检验中，如下图

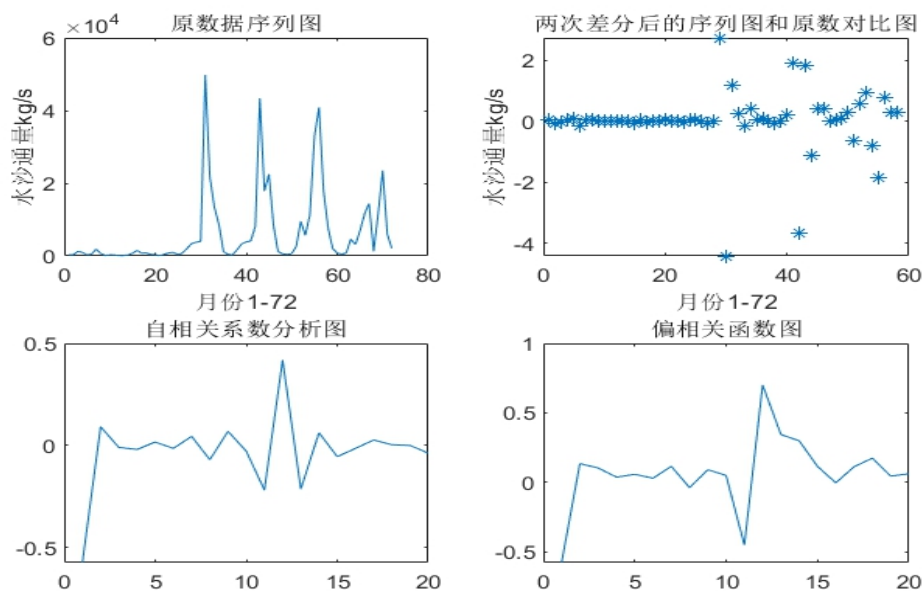


图 3.1

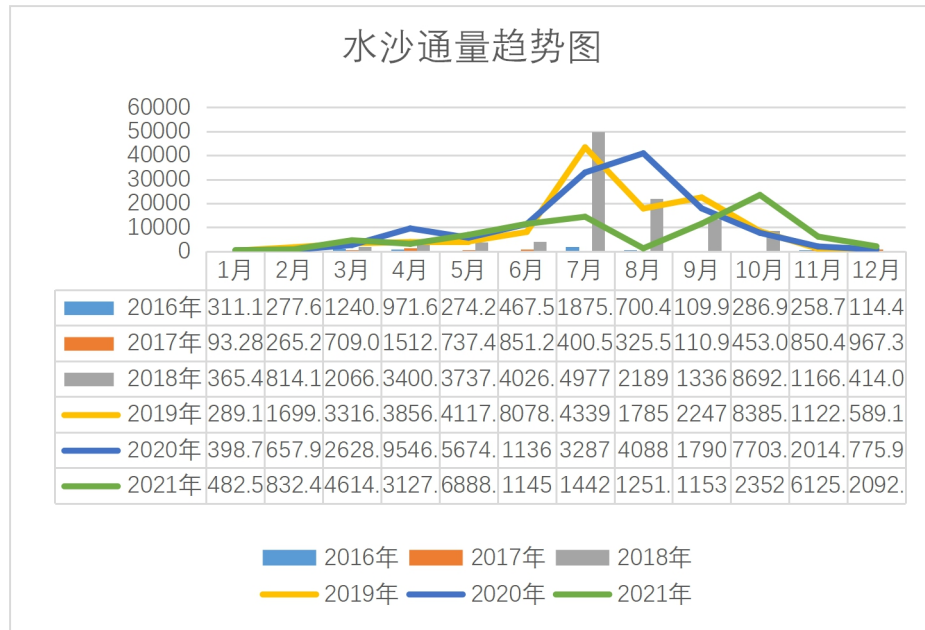


图 3.2

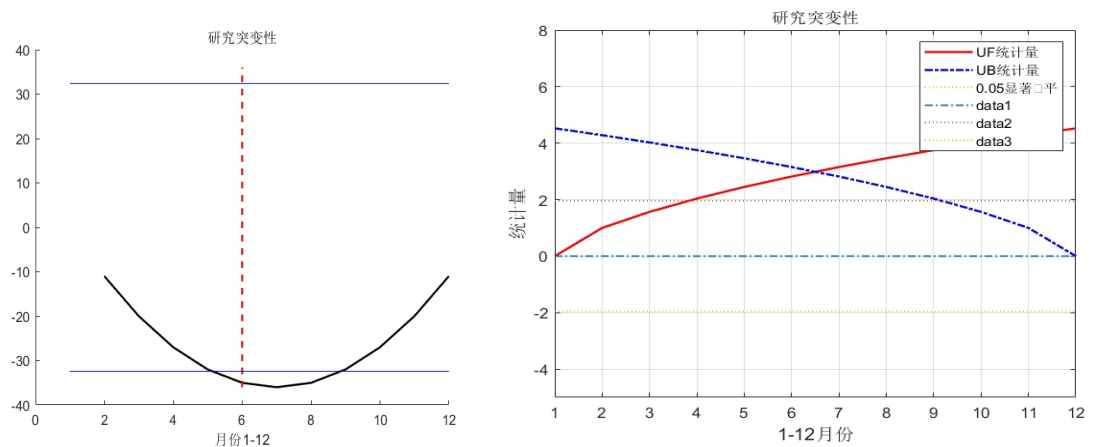
根据图 3.3 可看出 2016 年-2021 年水沙通量的变化趋势.

在进行趋势检验的算法中, 由于数据的容量较大, 则先给数据限定范围, 从而进行部分求解, 最后得出时间序列是下降趋势.

衡量趋势大小的指标 γ

$$\gamma = \text{Median}\left(\frac{x_i - x_j}{i - j}\right), \forall j < i, \quad 1 < j < i < n$$

Mann-Kendall 突变检验:



$$s_k = \sum_{i=1}^k \sum_{j=1}^{i-1} a_{ij}, k = 2, 3, \dots, n$$

按时间序列 t 的顺序，构造一个秩序列：

$$a_{ij} = \begin{cases} 1, x_i > x_j, 1 \leq i \leq j \\ 0, x_i \leq x_j, 1 \leq i \leq j \end{cases}$$

假设时间序列为随机变量，则可以用以下方程计算：

$$UF_k = \frac{S_k - E(S_k)}{\sqrt{Var(S_k)}}, k = 1, 2, \dots, n$$

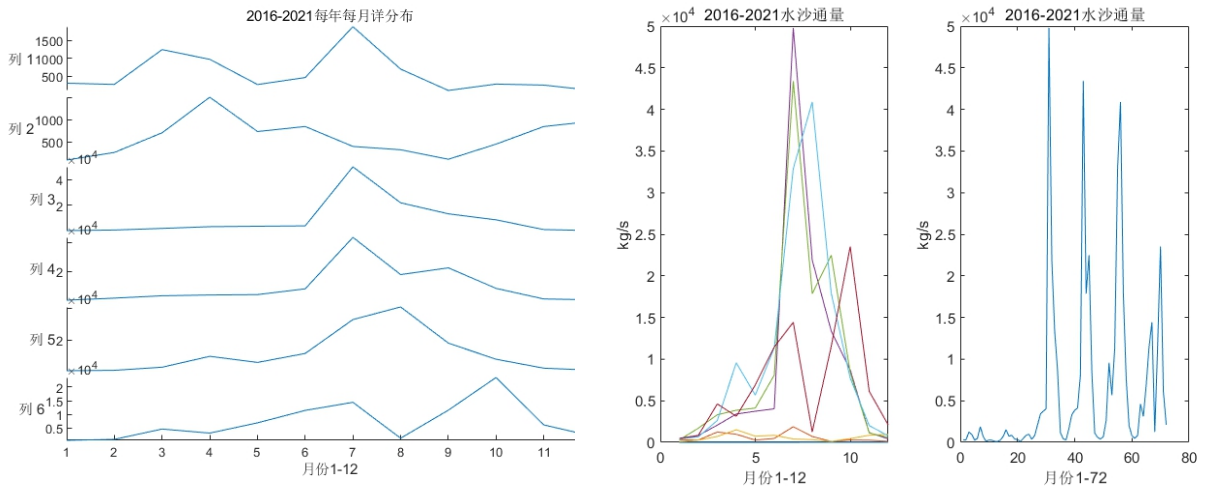
$$E(S_k) = \frac{k(k-1)}{4}$$

$$Var(S_k) = \frac{k(k-1)(2k+5)}{72}$$

若 $|UF_k| > U_\alpha$ ，则表示序列有明显的趋势变化。

根据时间序列 X 呈现出来的规律，重复上述过程，求出 UB_k 。

3.3 问题二模型求解与分析



Holt-Winters 加法模型适用于具有线性趋势和加法季节变化的序列.平滑序列的计算公式为：

$$\hat{y}_{t+k} = a_t + b_t k + S_t + k, t = s+1, s+2, \dots, T$$

其中, a_t 表示截距, b_t 表示斜率, $a_t + b_t k$ 表示趋势, S_t 表示季节因子, s 表示节周期长度, 三个系数 a_t, b_t, S_t 的递归式定义如下：

$$\begin{aligned}a_t &= a(y_t - S_{t-s}) + (1-a)(a_{t-1} + b_{t-1}), \\b_t &= \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}, \\S_t &= \gamma(y_t - a_t) + (1-\gamma)S_{t+k-s},\end{aligned}$$

其中, $k>0$, α , β , λ 在 0-1 之间.如果 $t=T$,预测模型为:

$$\hat{y}_{t+k} = a_t + b_t k + S_{t+k-s},$$

其中, S_{t+k-s} 用样本数据最后一年的季节因子.

Holt - Winters 乘法模型适用于具有线性趋势和乘法季节变化的序列.平滑序列的计算公式为

$$\hat{y}_{t+k} = (a_t + b_t k)S_{t+k}, t = s+1, s+2, \dots, T,$$

三个系数 a_t, b_t, S_t 的定义如下:

$$\begin{aligned}a_t &= \alpha \frac{y_t}{S_{t-s}} + (1-\alpha)(a_{t-1} + b_{t-1}), \\b_t &= \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}, \\S_t &= \gamma \frac{y_t}{a_t} + (1-\gamma)S_{t-s},\end{aligned}$$

其中, $k>0,a,B,y$ 在 0~1 之间.假设 $t=T$ 则算出的预测模型为:

$$\hat{y}_{t+k} = (a_T + b_T k)S_{t+k-s},$$

其中, S_{t+k-s} 用样本数据最后一年的季节因子.

Holt-Winters 乘法模型适用于具有线性趋势和乘法季节变化的序列.平滑序列的计算公式为:

$$\hat{Y}_{t+k} = (a_t + b_t k)S_{t+k}, t = s+1, s+2, \dots, T,$$

a_t, b_t, s_t 的定义如下:

$$\begin{aligned}a_t &= a \frac{y_t}{S_{t-x}} + (1-a)(a_{t-1} + b_{t-1}), \\b_t &= \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}, \\S_t &= \gamma(y_t - a_t) + (1-\gamma)S_{t-s},\end{aligned}$$

其中, $k > 0, \alpha, \beta, \gamma$ 在 0~1 之间。如果 $t = T$,

$$\hat{y}_{t+k} = (a_\gamma + b_\gamma k)s_{\gamma+k-s},$$

Pettitt 突变点检验法:

基于 Mann-Whitney 的统计函数 $U_{t,T}$ ，认为两个样本分别表示水沙通量和排沙量，且他们均来自于同一序列分布，对于这些相对应的式子， $U_{t,T}$ 和 $V_{t,T}$ 由下列公式计算：

$$U_{t,T} = U_{t-1,T} + V_{t,T}, \quad t=2,3,\dots, T \quad (1)$$

其中，

$$U_{t,T} = V_{t,T} \quad (2)$$

$$V_{t,T} = \sum_{j=1}^T \text{sgn}(x_i - x_j) \quad (3)$$

$$\text{sgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

通过 MATLAB 对参数进行修改编程导入附件数据，得出以下图形.

因此， $|U_{t,T}|$ 最大时，对应的 x_t 为可能的突变点（change-point），当突变点 $U_t > 0$ 时，该序列具有向下突变趋势，反之则有向上突变趋势.

4.1 问题三模型建立

ARIMA 模型:

ADF 检验表							
变量	差分阶数	t	P	AIC	临界值		
					1%	5%	10%
日	0	-1.702	0.430	282.188	-3.544	-2.911	-2.593
	1	-7.681	0.000***	279.883	-3.546	-2.912	-2.594
	2	-7.211	0.000***	290.624	-3.553	-2.915	-2.595

注：***、**、*分别代表 1%、5%、10%的显著性水平

4.2 问题三求解

ARIMA (p, d, g)模型具有如下结构:

采样监测方案制定：

根据数据分析得出的预测结果，制定未来两年的采样检测方案，确保及时掌握水沙通量的动态变化，同时减少检测成本.

- 1.采样频率：根据水沙通量趋势图，确定采样的频率，6~9 月份可能要更加频繁的采样，根据研究的突兀性考虑季节变化，夏季需要更加频繁的监测
- 2.采样时间：每天的八点采集监测
- 3.采样方法：采用时间序列方法，对数据进行高精度的筛选，并对数据进行处理
采取简单随机取样，保证数据的普遍性.

$$\begin{cases} (1 - \Phi_1 B - \dots - \Phi_p B^p) \nabla^d x_t = (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \\ E(\varepsilon_t) = 0, \text{var}(\varepsilon_t) = \sigma_\varepsilon^2 \\ \text{cov}(\varepsilon_t, \varepsilon_s) = 0 (s \neq t), \text{cov}(X_s, \varepsilon_t) = 0 (\forall s < t) \end{cases} \quad (2-1)$$

数据清洗

缺失值处理与异常值处理如下

$$\lim_{x \rightarrow \infty} \sqrt{b^2 - 4ac}$$

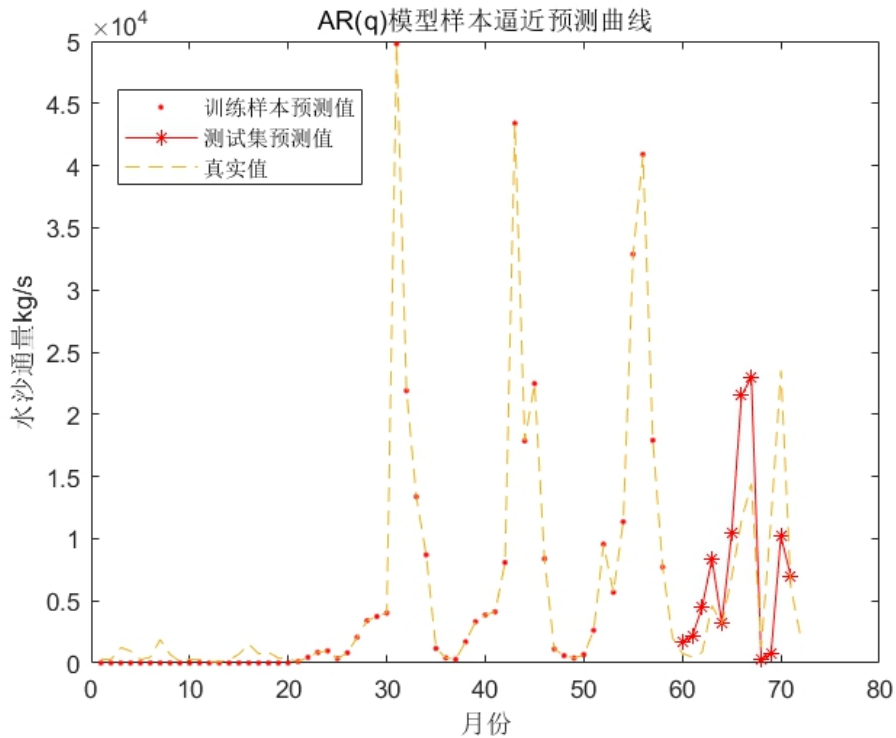
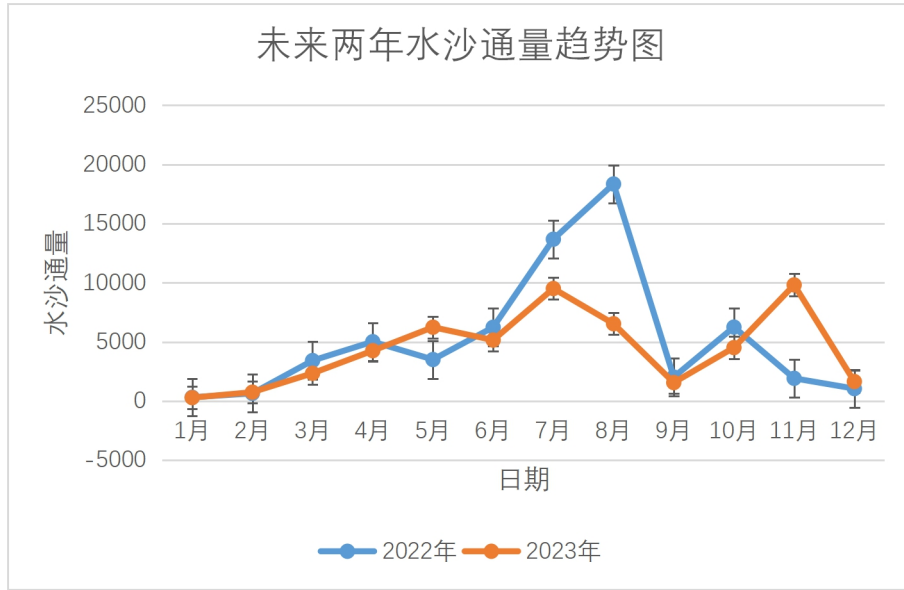


图 3.2



由上图可知：水沙通量在 6-9 月份上升趋势明显，所以未来两年的最优监测方案为在 6-9 月份进行多次监测来保证数据的准确性，其他月份正常监测。

5.1 问题四模型分析

HEA-RAS 模型：HEA-RAS 旨在模拟河流和水体系统的水流、河流行为以及与洪水管理和河流工程相关的各种水文和水力问题。

5.2 问题四模型建立

描述河道非恒定水流运动的一维圣维南方程组为：

$$B \frac{\partial Z}{\partial t} + \frac{\partial Q}{\partial x} = q \quad (1a)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left(\frac{\alpha Q^2}{A} \right) + gA \frac{\partial Z}{\partial x} + gA \frac{|Q|Q}{K^2} = qV_x \quad (1b)$$

式中: q 为旁侧入流; Q 、 A 、 B 和 Z 分别为河道断面流量、过水面积、河宽和水位; K 为流量模数; α 为动量校正系数; qV_x 为旁侧入流流速在干流水流方向上的分量，一般可以近似为零。

利用四点线性隐格式对圣维南方程组离散并线性化，取权重系数为 1.0，得到差分方程组为：

$$-Q_i^{k+1} + Q_{i+1}^{k+1} + C_i Z_i^{k+1} + C_i Z_{i+1}^{k+1} = D_i, \quad (i=1,2,\dots,n-1) \quad (2a)$$

$$E_i Q_i^{k+1} + G_i Q_{i+1}^{k+1} - F_i Z_i^{k+1} + F_i Z_{i+1}^{k+1} = D_i,$$

$$(i=1,2,\dots,n-1) \quad (2b)$$

式中： C_i 、 D_i 、 E_i 、 G_i 、 F_i 和 Φ_i 为系数

每个河段都可拆分出两个差分方程，首、末节点水位都为自由变量，由此可以得出首、末断面的水位和流量，得到首、末断面的流量与首、末节点水位的关系方程为：

$$Q_1 = \alpha_1 + \beta_1 Z_{(N1)} + \zeta_1 Z_{(N2)} \quad (3a)$$

$$Q_n = \theta_n + \eta_n Z_{(N2)} + \gamma_n Z_{(N1)} \quad (3b)$$

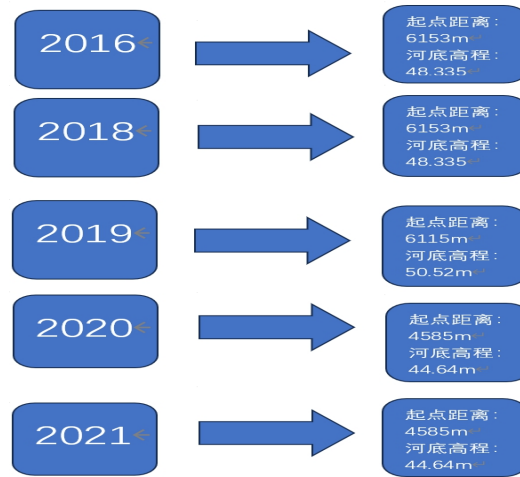
由此可得，同一断面的流量可表达成本断面水位与首、末节点水位的关系：

$$Q_i = \alpha_i + \beta_i Z_i + \zeta_i Z_{(N2)} \quad (4a)$$

$$Q_i = \theta_i + \eta_i Z_i + \gamma_i Z_{(N1)} \quad (4b)$$

即可联立求解相应的断面的水位和流量。

5.3 问题四求解



非恒定流模拟基于连续方程和动量方程，其中连续方程为：

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u_i)}{\partial x_i} = 0 \quad (1)$$

动量方程为：

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = f_i - \frac{1}{\rho} * \frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_i \partial x_i}$$

二维水动力模型计算：

通过该模型模拟水流与水位的变化，以及他们的方向不可控性，基于三维雷诺 Stokes 平均方程沿水深积分的连续和动量方程，可得到：

$$\frac{\partial h}{\partial t} + \frac{\partial hu}{\partial x} + \frac{\partial hv}{\partial y} = hS(2)$$

$$\frac{\partial \bar{hu}}{\partial t} + \frac{\partial \bar{hu}^2}{\partial x} + \frac{\partial \bar{hvu}}{\partial y} = hf\bar{v} - gh\frac{\partial \eta}{\partial x} - \frac{h}{\rho_0}\frac{\partial \rho_\alpha}{\partial x} - \frac{gh^2}{2\rho_0}\frac{\partial \rho}{\partial x} + \frac{T_{sx}}{\rho_0} - \frac{T_{bx}}{\rho_0} - \frac{1}{\rho_0}\left(\frac{\partial S_{xx}}{\partial x} + \frac{\partial S_{xy}}{\partial y}\right) + \frac{\partial}{\partial x}(hT_{xx}) + \frac{\partial}{\partial x}(hT_{xy}) + hu_s S$$

(3)

$$\frac{\partial \bar{hv}}{\partial t} + \frac{\partial \bar{hv}^2}{\partial y} + \frac{\partial \bar{huv}}{\partial x} = hf\bar{u} - gh\frac{\partial \eta}{\partial y} - \frac{h}{\rho_0}\frac{\partial \rho_\alpha}{\partial y} - \frac{gh^2}{2\rho_0}\frac{\partial \rho}{\partial y} + \frac{T_{sy}}{\rho_0} - \frac{T_{by}}{\rho_0} - \frac{1}{\rho_0}\left(\frac{\partial S_{yx}}{\partial x} + \frac{\partial S_{yy}}{\partial y}\right) + \frac{\partial}{\partial x}(hT_{xy}) + \frac{\partial}{\partial y}(hT) + \frac{\partial}{\partial y}(hT_{yy}) + hv_s S$$

(4)

用以上模型进行分析后，可以预测未来十年河底高程趋势会逐渐升高。

五.模型分析检验

5.1 灵敏度分析

	水位	水流量	含沙量
数据之间的标准差	12.45628853	30.19392415	0.361734234
数据之间误差的均值	-0.475222446	0.0693371626	-0.07000922
数据之间绝对误差的标准差	9.544020455	25.77405967	0.24908457
数据之间绝对误差的均值	8.017954471	15.72612245	0.272217647

5.2 误差分析

1. 准备数据：首先，我们收集了大量的水沙通量数据包括含沙量、时间、水位、水流量。
2. 模型训练：我们使用了时间序列模型并使用机器学习进行数据。
3. 模型性能评估：在测试数据上评估模型的性能，发现模型的准确率为百分之 80%。
4. 修复与改进：调整模型的阈值，增加更多关键词，探索更为复杂的模型架构，深度学习模型。

六.模型评价与推广

6.1 模型的优点

(1) 时间序列模型充分结合实际, 简化已知条件, 考虑了诸多重要因素得到合理的模型, 通过对一段时间内的数据进行整合排序, 从而对未来的情况的预测与预估得到的数据贴合实际, 具有较高的应用价值, 可以推广金融学与经济学, 气象学和气候学, 医学等方面;

(2) Mann-kendall 模型运用双向检验和独立变量思想, 抓住影响自变量变化趋势问题的重要因素, 将复杂的水沙监测问题转化为简单的正态统计变量问题, 合理设置参数, 模型的输出结果符合题目要求, 能解决实际问题;

(3) 本文使用的 pettitt 算法具有简单性, 非参数性, 原理直观, 检测多个突变点等多个优点, 对于求解时间序列模型非常适用;

6.2 模型的不足

(1) 本文提出的模型对于现有条件使用效果较好, 由于时间问题没有对其他情况进行检验.对于其他情形(如:随机性太高的预测问题, 数据频率不匹配的问题), 从而可能无法达到较好的效果;

(2) 本文中 Mann-kendall 模型仅能预测短期内的数据变化, 无法对长期的数据进行合理预测.

6.3 模型的推广

推广: 在非定态数据方面, 可以将定态参数替换成非定态参数, 从而解决非定态数据无法确定变量的问题;

改进: 结合参考, 进一步考虑数据对于有限范围的影响, 从而得到更合理的模型; 对于数据频率不匹配的情况下, 要将不同时间尺度的数据合并在一起.

参考文献

- [1] 王船海, 李光炽. 实用河网水流计算[M]. 南京, 中国:河海大学出版社, 2006.
- [2] 张品一,罗春燕,梁锶.基于 GA-BP 神经网络模型的黄金价格仿真预测[J].统计与决策,2018,(17).
- [3] 雨茜,王俊丽,等.循环神经网络研究综述[J].计算机应用,2018,38(S2).

[4] 黄文杰,曹鸿兴,顾岚,等. 时间序列的 ARIMA 季节模型在长期预报中的应用[J]. 科学通报, 1980,25 (22):1030-1032.

[5] 王燕. 时间序列分析:基于 R[M]. 北京:中国人民大学出版社, 2015.

[6] 唐凯,卢埴洵,陈新煜.潮州市汛期降水变化特征分析[J].农业与技术,2020(10):138-140.

附录

附录说明:

Matlab 代码	说明
1.Untitled.m	参数
2.untitled.mlx	6 年水沙通量折线图
3.untitled1.mlx	ARIMA 季节性时间序列预测模型(1)
4.untitled2.mlx	Pettitt 突变检验
5.untitled8.mlx	Mann-Kendall 检验
6.untitled10.mlx	ARIMA 季节性时间序列预测模型(2)
7.untitled11.mlx	每年变化性趋势

1. Untitled.m 文件内容

```
A=[1,311.106419354839,93.2827741935484,365.432677419355,289.196580645161,398.740129032258,482.544693548387;2,277.654551724138,265.283931034483,814.143103448276,1699.71172413793,657.916580459770,832.406251724138;3,1240.20435483871,709.050419354839,2066.22387096774,3316.61838709677,2628.84981182796,4614.08010752688;4,971.667933333333,1512.32500000000,3400.91400000000,3856.30333333333,9546.20211111111,3127.24275000000;5,274.250741935484,737.471032258064,3737.63548387097,4117.70645161290,5674.01489247312,6888.22096774194;6,467.547033333333,851.208000000000,4026.25000000000,8078.25866666667,11364.05400000000,11452.0866666667;7,1875.16961290323,400.508064516129,49777.3780645161,43394.8870967742,32874.0516129032,14423.2974193548;8,700.401677419355,325.573483870968,21894.3912903226,17857.5180967742,40887.3516129032,1251.52977419355;9,109.910100000000,110.956433333333,13364.9366666667,22472.3183333333,17903.2566666667,11530.5796666667;10,286.928935483871,453.087064516129,8692.53870967742,8385.78483870968,7703.79129032258,23521.4430107527;11,258.766200000000,850.401000000000,1166.67213333333,1122.56873333333,2014.67155555556,6125.53777777778;12,114.478161290323,967.334866666667,414.068566666667,589.135033333333,775.980384408603,2092.32706896552]
```

B=A'

```
C=[311.1064194 277.6545517 1240.204355 971.6679333 274.2507419 467.5470333 1875.169613 700.4016774 109.9101 286.9289355 258.7662 114.4781613 93.28277419 265.283931 709.0504194 1512.325 737.4710323 851.208 400.5080645 325.5734839 110.9564333 453.0870645 850.401 967.3348667 365.4326774 814.1431034 2066.223871 3400.914 3737.635484 4026.25 49777.37806 21894.39129 13364.93667 8692.53871 1166.672133 414.0685667 289.1965806 1699.711724 3316.618387 3856.303333 4117.706452 8078.258667 43394.8871 17857.5181 22472.31833 8385.784839 1122.568733 589.1350333 398.740129 657.9165805 2628.849812 9546.202111 5674.014892 11364.054
```



```

32874.05161 40887.35161 17903.25667 7703.79129 2014.671556 775.9803844
482.5446935 832.4062517 4614.080108 3127.24275 6888.220968 11452.08667
14423.29742 1251.529774 11530.57967 23521.44301 6125.537778 2092.327069]
S2=[311.106419354839,93.2827741935484,365.432677419355,289.196580645161,398.7401
29032258,482.544693548387;277.654551724138,265.283931034483,814.143103448276,169
9.71172413793,657.916580459770,832.406251724138;1240.20435483871,709.05041935483
9,2066.22387096774,3316.61838709677,2628.84981182796,4614.08010752688;971.667933
333333,1512.3250000000,3400.9140000000,3856.3033333333,9546.2021111111,3127.2
4275000000;274.250741935484,737.471032258064,3737.63548387097,4117.70645161290,5
674.01489247312,6888.22096774194;467.547033333333,851.208000000000,4026.25000000
000,8078.25866666667,11364.0540000000,11452.0866666667;1875.16961290323,400.5080
64516129,49777.3780645161,43394.8870967742,32874.0516129032,14423.2974193548;700
.401677419355,325.573483870968,21894.3912903226,17857.5180967742,40887.351612903
2,1251.52977419355;109.910100000000,110.956433333333,13364.9366666667,22472.3183
333333,17903.2566666667,11530.5796666667;286.928935483871,453.087064516129,8692.
53870967742,8385.78483870968,7703.79129032258,23521.4430107527;258.766200000000,
850.401000000000,1166.67213333333,1122.56873333333,2014.67155555556,6125.5377777
7778;114.478161290323,967.334866666667,414.068566666667,589.135033333333,775.980
384408603,2092.32706896552]

```

2.untitled.mlx

```

run('D:\Y\Untitled.m')%运行参数所在文件
subplot(1,2,1),plot(A)
xlabel('月份 1-12');
ylabel('kg/s');
title('2016-2021 水沙通量');
subplot(1,2,2),plot(C)
xlabel('月份 1-72');
ylabel('kg/s');
title('2016-2021 水沙通量');

```

3.untitled1.mlx

```

run('D:\Y\Untitled.m');
t=(1:72);
x=C;
X=[ones(72,1) t']; %回归的资料矩阵
y=log(x); %线性化
[B,BINT,R,RINT,STATS] = regress(y,X); %回归
y2= exp(B(1)+B(2).*t) %预测值
%plot(t,x,t,y2,'+'); %回归效果图
B=[-290.4864,0.1510];
STATS=1.0e+003*[0.0010,2.1838,0] ;
r=x-y2; %残差数列
plot(t,r,'O'); %残差散点图
xlabel('时间序列')
ylabel('水沙通量 kg/s');
title('残差散点图 ');
%二次差分后序列

```

```

r1=diff(r);          %残差的一阶差分
r11=[0 r1];         %补数列差分后的项为 0
plot(t,r11,'o');    %一阶差分散点图
xlabel('时间序列')
ylabel('水沙通量 kg/s');
title('一阶差分散点图');
r2=diff(r1);        %二阶差分
r21=[0 0 r2];       %补数列差分后的项为 0
plot(t,r21,'o');    %二阶差分散点图
xlabel('时间序列')
ylabel('水沙通量 kg/s');
title('二阶差分散点图');
%序列 {wt} 的样本自相关函数pk 程序
w=r2-mean(r2);      %零均值化
gamao=var(w);        %求方差
for j=1:27
    gama(j)=w(j+1:end)*w(1:end-j)/27;
end
rho=gama/gamao      %样本自相关系数
bar(rho)            %条状图
title('自相关系数条状图');
f(1,1)=rho(1);
for k=2:27
    s1=rho(k);s2=1; %计算的初始值
    for j=1:k-1
        s1=s1-rho(k-j)*f(k-1,j);
        s2=s2-rho(j)*f(k-1,j);
    end
    f(k,k)=s1/s2;          %对角上的样本偏相关系数
    for j=1:k-1
        f(k,j)=f(k-1,j)-f(k,k)*f(k-1,k-j); %不在对角上的样本偏相关系数
    end
end
pcorr=diag(f)' %提取偏相关函数
bar(pcorr)      %条形图
title('偏自相关函数条状图');

4. untitled2.mlx
run('D:\Y\Untitled.m')
data=A
x=(1:12)
n=length(data);
s=0;
for i=2:n;
    for j=1:n;
        y(i-1,j)=sign(data(i-1,1)-data(j,1));
        g(i-1)=sum(y(i-1,:));
    end
end

```

```

end
end
    Ut=cumsum(g);
    F=Ut'
    Kmax=max(abs(F));
    chg_point=find(abs(F)==Kmax);%找出最大值对应的序列号
    coef=2*exp(-6*Kmax^2/(n^3+n^2));%显著水平计算
    figure(1)
    hold on
    plot(x(1,2:12),F,'K-', 'linewidth',1.5)%画出整个曲线的变化趋势
    xlabel('月份 1-12');
    title('研究突变性');
    %plot(x(2,2:n),F,'b--', 'linewidth',1.5)%画出整个曲线的变化趋势
    plot([x(chg_point),x(chg_point)],[-Kmax,Kmax], 'r--', 'linewidth',1.5)%画出该竖线即为
突变点
    Klim=Kmax-Kmax*0.1
    plot([x(1),x(n)],[-Klim,-Klim], 'b-')
    plot([x(1),x(n)],[Klim,Klim], 'b-')

5.untitled8.mlx
    run('D:\Y\Untitled.m')
    Data=A
    n = length(Data);
    %% 趋势分析线性:Mann-Kendall 检验
    Sgn=zeros(n-1,n-1); %初始化分配内存
    for i=1:n-1
        for j=i+1:n
            if((Data(j)-Data(i))>0)
                Sgn(i,j)=1;
            else
                if((Data(j)-Data(i))==0)
                    Sgn(i,j)=0;
                else
                    if((Data(j)-Data(i))<0)
                        Sgn(i,j)=-1;
                    end
                end
            end
        end
    end
    Smk=sum(sum(Sgn));
    VarS=n*(n-1)*(2*n+5)/18;
    if n>10
        if Smk>0
            Zs=(Smk-1)/sqrt(VarS);
        else
            if Smk==0
                Zs=0;
            else

```

```

if Smk<0
Zs=(Smk+1)/sqrt(VarS);
end
end
end
end
%%% beta 斜率 描述单调趋势
t=1;
for i=2:n
for i=2:n
for j=1:(i-1)
temp(t)=( Data(i)-Data(j) )/( i-j );
t=t+1;
end
end
beta=median( temp );
%%% 突变检验
Sk=zeros(n,1); % 定义累计量序列 Sk
UFk=zeros(n,1); % 定义统计量 UFk
s = 0;
% 正序列计算 start-----
for i=2:n
for j=1:i
if Data(i)>Data(j)
s=s+1;
else
s=s+0;
end
end
Sk(i)=s;
E=i*(i-1)/4; % Sk(i)的均值
Var=i*(i-1)*(2*i+5)/72; % Sk(i)的方差
UFk(i)=(Sk(i)-E)/sqrt(Var);
end
% 正序列计算 end-----
% 逆序列计算 start-----
Sk2=zeros(n); % 定义逆序累计量序列 Sk2
UBk=zeros(n,1);
s=0;
Data2=flipud(Data); % 按时间序列逆转样本 y
for i=2:n
for j=1:i
if Data2(i)>Data2(j)
s=s+1;
else
s=s+0;
end
end
end
end

```

```

Sk2(i)=s;
E=i*(i-1)/4;
Var=i*(i-1)*(2*i+5)/72;
UBk(i,1)=0-(Sk2(i)-E)/sqrt(Var);
end
% 逆序列计算 end-----
UBk2=flipud(UBk);
UFk=UFk';
UBk2=UBk2';
figure(3)%画图
plot(1:n,UFk,'r-','linewidth',1.5);
hold on
plot(1:n,UBk2,'b-','linewidth',1.5);
plot(1:n,1.96*ones(n,1),'.','linewidth',1);
axis([1,n,-5,8]);
legend('UF 统计量','UB 统计量','0.05 显著水平');
xlabel('1-12 月份','FontName','TimesNewRoman','FontSize',12);
ylabel('统计量','FontName','TimesNewRoman','FontSize',12);
title(['研究突变性']);
grid on
hold on
plot(1:n,0*ones(n,1),'-','linewidth',1);
plot(1:n,1.96*ones(n,1),'.','linewidth',1);
plot(1:n,-1.96*ones(n,1),'.','linewidth',1);
End

```

6.untitled10.mlx

clear; clc

run('D:\Y\Untitled.m')

%小白专用, "***** 《需要自己输入》 *****" 仅在有这种注释的地方改成自己的数据即可, 一共有 4 个地方

DD=C;%这里输入自己的单序列数据, 要求行向量***** 《需要自己输入》

N=length(DD);

n=58;%自己选取训练集个数***** 《需要自己输入》 *****

F =DD(1:72);

%-----由于时间序列有不平稳趋势, 进行两次差分运算, 消除趋势性
-----%

Yt=[0,diff(DD,1)];

L=diff(DD,2);%全体, 比原 3 始数据少 2 个, 因为做了差分

Y=L(1:n); %输入

a=length(L)-length(Y);%单步预测步数

aa=a;%多步预测步数

%画图

figure;

plot(DD);

xlabel('月份 1-72');

```

ylabel('水沙通量 kg/s');
title('原数据序列图');
hold on;
figure;
plot(Y,'*');
xlabel('月份 1-72');
ylabel('水沙通量 kg/s');
title('两次差分后的序列图和原数对比图');
%%
%-----对 数 据 标 准 化 处 理
-----%
%处理的算法 : (data - 期望)/方差
Ux=sum(Y)/n ; % 求序列均值
yt=Y-Ux;
b=0;
for i=1:n
    b=yt(i)^2/n+b;
end
v=sqrt(b); % 求序列方差
% Y=yt/v; % 标准化处理公式
Y=zscore(Y);
f=F(1:n);
t=1:n;
%画图
figure;
plot(t,f,t,Y,'r')
title('原始数据和标准化处理后对比图');
xlabel('月份'),ylabel('水沙通量 kg/s');
legend('原始数据 F ','标准化后数据 Y ','Location','best');
%%
%-----对 数 据 标 准 化 处 理
-----%
%-----检验预处理后的数据是否符合 AR 建模要求，计算自相关和偏
相关系数-----%
%-----计算自相关系数-----%
%%
R0=0;
for i=1:n

```