

长江中游城陵矶河段 2016—2020 年汛期水位非正常波动: 影响因子及滤波修正*

程海云¹, 香天元¹, 唐 聪^{2**}

(1: 长江水利委员会水文局, 武汉 430012)

(2: 长江水利委员会水文局长江中游水文水资源勘测局, 武汉 430012)

摘要: 长江经济带的人口和经济总量均超过全国的 40%, 长江中游城市群受地理位置与水文条件的综合影响更易遭受洪涝灾害的威胁. 莲花塘水位站作为长江中游城陵矶河段蓄滞洪区启用指标控制站, 当三峡水库水位达 155.0 m 后, 如莲花塘站水位达到 34.40 m 并继续上涨, 则城陵矶附近区蓄滞洪区需采取分洪措施. 然而受河流形态和江湖关系(长江与洞庭湖)的双重影响, 莲花塘水位站水位呈现出“潮汐”式不规则周期性波动变化(俗称“假潮”), 给蓄滞洪区启用等工作带来一定决策困扰. 为了研究洞庭湖出口长江城陵矶河段高水期洪水水位波动现象相关因子和水位滤波方法, 本文对莲花塘水位站近 5 年 18 个月(2016 年 6—8 月、2017 年 7—10 月、2018 年 5—8 月、2019 年 6—8 月、2020 年 6—9 月)5 min 水位自记数据进行分析, 首先明确了单日水位波动值 $D > 0.05$ m 为发生“假潮”现象的判别标准; 通过二元 Logistic 回归分析表明, 在监利站高水位($Z_j \geq 34.00$ m)期间, 当监利至莲花塘水位落差(ΔZ_{jl})大于 2.75 m 时, “假潮”发生比例为 100%, 模型预测结果综合百分比校正值为 96.4%, 表明 ΔZ_{jl} 为发生“假潮”的主要影响因子; 然后通过快速傅里叶变换法、局部加权回归算法两种数据滤波方法对水位波动序列进行处理, 研究表明, 两种方法均适用于不同振幅的波形态数据分析, 不受短时间数据丢失影响, 不受水位瞬时值的影响, 具有较大的应用价值; 从误差值的波动情况来看, 局部加权回归算法误差值波动更小, 更适用于对波动水位数据的处理.

关键词: 洞庭湖; 假潮; 二元 Logistic 回归模型; 快速傅里叶变换法; 局部加权回归算法; 滤波

Abnormal fluctuation of water level in the Chenglingji reach of the middle Yangtze River during flood seasons, 2016–2020: Influencing factors and filtering correction*

Cheng Haiyun¹, Xiang Tianyuan¹ & Tang Cong^{2**}

(1: Water Resources Survey, CWRC, Wuhan 430012, P.R. China)

(2: Middle Yangtze River Bureau of Hydrology and Water Resources Survey, CWRC, Wuhan 430012, P.R. China)

Abstract: The Yangtze economic belt accounts for more than 40% of China's total population and economy, and the urban agglomeration in the middle reaches of the Yangtze River is more vulnerable to the threat of flood disasters due to the comprehensive influence of geographical location and hydrological conditions. The water level of the Lianhuatang station is served as the index control station of the flood storage and detention area in the middle reaches of the Yangtze River. When the water level of the Three Gorges Reservoir reaches 155.0 m, if the water level of the Lianhuatang station reaches 34.40 m and continues to rise, flood diversion measures should be taken in the flood storage and detention area near Chenglingji. While under the dual influence of the river form and the relationship between the river and the lake (the Yangtze River and Lake Dongting), the water level of the Lianhuatang station presents tidal irregular periodic fluctuation (commonly known as “false tide”), which brings decision-making difficulties to the flood storage and detention area. To study the related factors of flood level fluctuation phenomenon and water level filtering method in the Chenglingji reach of the Yangtze River at the outlet of Lake Dongting, this paper analyzed the self-recorded water level data of Lianhuatang water station in the last 5 years and 18 months (June–August 2016, July–October 2017, May–August

* 2021-01-24 收稿; 2021-05-17 收修改稿.

湖南省水利科技项目((2013)243-15)资助.

** 通信作者; E-mail: cong_tang@163.com.

2018, June–August 2019, and June–September 2020). Firstly, the single-day water level fluctuation value $D > 0.05$ m is defined as the criterion for the occurrence of “false tide”. And the Binary Logistic regression analysis showed that during the period of high water level ($Z_j \geq 34.00$ m) at Jianli station, when the water level drop from Jianli to Lianhuatang (ΔZ_{ji}) is greater than 2.75 m, the proportion of “false tide” is 100%. The comprehensive correction value of model prediction results was 96.4%, indicating that ΔZ_{ji} was the main factor influencing the false tide. Then, two data filtering methods, fast fourier transform and local regression, are used to process the water level fluctuation sequence, which shows that both of the two methods are applicable to the analysis of wave shape data with different amplitudes, and are not affected by short-term data loss and instantaneous water level value, which may have great application value. From the perspective of the fluctuation of the error value, the locally weighted regression algorithm has a smaller fluctuation of the error value and is more suitable for the processing of the fluctuating water level data.

Keywords: Lake Dongting; false tide; binary Logistic regression analysis; fast fourier transform filter analysis; local regression; filtering method

长江是中华民族的生命河,也是中华民族发展的重要支撑^[1]. 长江经济带覆盖面积占全国的 21%, 人口和经济总量均超过全国的 40%, 是我国经济中心所在、活力所在^[2]. 其中, 作为长江经济带“一轴、两翼、三极、多点”格局中“三极”之一的长江中游城市群, 承东启西、连南接北, 以全国 3.4% 的土地面积和 9.0% 的人口数量创造了 9.6% 的经济总量, 是长江经济带的重要组成部分, 在我国区域发展格局中占有重要地位^[3]. 受地理位置与水文条件的综合影响, 长江中下游承接长江上游、洞庭湖、汉江、鄱阳湖等洪水, 洪水峰高量大、持续时间长^[4,6], 在极端气候条件下, 易遭受洪涝灾害的威胁, 为区域经济的良好发展带来了不安因素. 为确保长江流域广大地区人民生命财产安全和经济社会发展, 国务院分别于 2011 和 2015 年批复了《长江洪水调度方案》和《长江防御洪水方案》, 明确长江中下游干流沿岸设有 42 处蓄滞洪区, 总面积约为 1.2 万 km^2 , 包括耕地 4745.3 km^2 , 人口约 632.5 万人, 有效蓄洪容积为 589.7 亿 m^3 . 其中, 城陵矶河段(城陵矶—东荆河口)蓄滞洪区调度规定, 当三峡水库水位达 155.0 m 后, 如莲花塘站水位达到 34.40 m 并继续上涨, 则需采取城陵矶附近区蓄滞洪区分洪措施. 因此, 作为城陵矶河段蓄滞洪区启用指标控制站, 莲花塘水位站的水位值极其关键. 然而, 受河流形态和江湖关系(长江与洞庭湖)的双重影响, 在特殊水情下, 莲花塘水位站水位呈现出“潮汐”式不规则周期性波动变化导致水位波动, 俗称“假潮”. 已有研究表明, 莲花塘水位站“假潮”波动周期一般在 0.8~1.0 h 之间, 最大波动变化幅度可达 0.29 m^[7]. “假潮”现象给蓄滞洪区启用等工作带来一定决策困扰, 如 2020 年长江干流第 2, 3 次编号洪水期间, 莲花塘水位站自 7 月 21 日起, 在“假潮”作用影响下其水位波动瞬时最大值已达 34.40 m 并有继续上涨的趋势, 根据上述两方案规定, 若采纳该“假潮”影响下的瞬时水位值, 城陵矶附近区蓄滞洪区需采取分洪措施, 然而下一时间段莲花塘水位站水位或将下降并低于指标水位 34.40 m, 表明不应采取分洪措施, 是否启用分蓄洪区对地区国民经济影响重大, 受“假潮”影响的水位值是否符合启用蓄滞洪区的判定条件成为了防汛决策部门考量的难题. 因此, 通过对城陵矶河段水文特性及水位数据的研究分析, 探索“假潮”现象的特点与规律, 研究“假潮”水位滤波处理方法具有十分重要的意义.

“假潮”现象早已引起水文水利行业学者的广泛关注, 现有研究表明, “假潮”大致划分为两类, 原发型“假潮”和传播型“假潮”^[8], 主要发生在长江中下游、黄河中下游河段, 目前以黄河山东段对“假潮”的研究为主. 宋士强等^[8]结合黄河下游孙口水文站实际观测资料, 对黄河下游地区较小流量级下发生的“假潮”现象进行了系统分析, 得出了“假潮”是由于沙波运动在弯道处因控导工程作用产生雍水和回水而引发的一种水位、流速、含沙量在短时间内发生较大幅度变化的水文现象; 刘超等^[7]对长江中游莲花塘水位站“假潮”现象水位平滑处理方法进行了研究, 通过对多种处理方法的综合研究表明, 采用 3 h 加权平滑及综合处理效果较为理想; 谷源泽等^[9]对黄河山东段“假潮”期水文测报方法进行了分析, 认为“假潮”期水文测验用前一日平均水位代表法较为方便实用, 而过程测验法尽管报汛精度较高, 但由于受各种条件限制, 目前还难以实现; 李春莲等^[10]对黄河下游河段水文资料整编中“假潮”现象的数据分析与处理进行了研究, 提出了利用水量平衡原理进行平差计算, 较好地解决了“假潮”引起的系统误差问题. 与黄河中下游山东段“假潮”现象时机、成因不同, 长江中游城陵矶河段“假潮”主要在主汛期发生, 在越接近洪峰时水位波动越大, 刘超等^[7]认为该河段主要受弯道干扰波影响, 是一种出现在河道两岸交相传播的一种能量波, 但尚未对该河段“假潮”

规律进行研究. 本文结合莲花塘水位站地理位置和已有研究成果, 选取监利水文站、城陵矶(七里山)水文站、螺山水文站和莲花塘水位站近 5 年 18 个月的水文数据进行分析, 研究“假潮”发生判别标准与“假潮”现象出现规律, 计算“假潮”发生模型, 并分析“假潮”水位数据滤波处理方法, 以期降低“假潮”现象对莲花塘水位站水位报讯数据的影响, 为蓄滞洪区启用等工作的合理决策提供有力的数据支撑.

1 研究区域概况

长江是我国第一大河, 全长 6300 余千米, 坐西向东流经 11 个省(区、市), 支流涉及 8 个省(区), 流域面积 180 万 km^2 , 长江防洪安全关系流域广大地区人民生命财产安全和经济社会发展. 长江中游河段自宜昌至江西鄱阳湖口, 长 955 km, 集水面积约 68 万 km^2 . 洞庭湖($28^{\circ}44' \sim 29^{\circ}35' \text{N}$, $111^{\circ}53' \sim 113^{\circ}05' \text{E}$)是我国第二大淡水湖, 也是长江中游重要的过水吞吐湖泊^[11], 长江来水自松滋口、藕池口、太平口入湖, 自城陵矶出湖, 誉为“长江之肾”, 兼具河流与湖泊双重属性^[12], 作为长江流域重要的调蓄湖泊, 洞庭湖具有强大的蓄洪能力^[13]. 城陵矶(七里山)水文站是洞庭湖出口汇入长江的重要控制站^[14], 被誉为洞庭湖及长江流域水情“晴雨表”, 多年平均年径流量达 2600 亿 m^3 , 占长江中游干流螺山水文站多年平均年径流量的 41.10%. 江湖交汇口建有莲花塘水位站, 为长江干流主要控制站, 也是长江中游城陵矶河段蓄滞洪区启用的指标控制站, 该站水位在防汛调度与决策上占有十分重要的作用.

2 材料与方法

2.1 数据来源

根据洞庭湖出湖口河流分布特征, 本研究以莲花塘水位站为中心, 选取其长江干流上游监利水文站、下游螺山水文站和洞庭湖出口城陵矶(七里山)水文站作为研究对象, 测站分布见图 1. 根据莲花塘水位站水位等级, 选取 2016 年 6—8 月、2017 年 7—10 月、2018 年 5—8 月、2019 年 6—8 月、2020 年 6—9 月连续 5 年高水期水位作为研究对象, 数据来源于长江水利委员会水文局长江中游水文水资源勘测局水位自记数据及整编数据(水位数据采用冻结基面高程, 水位自记仪满足精度要求).

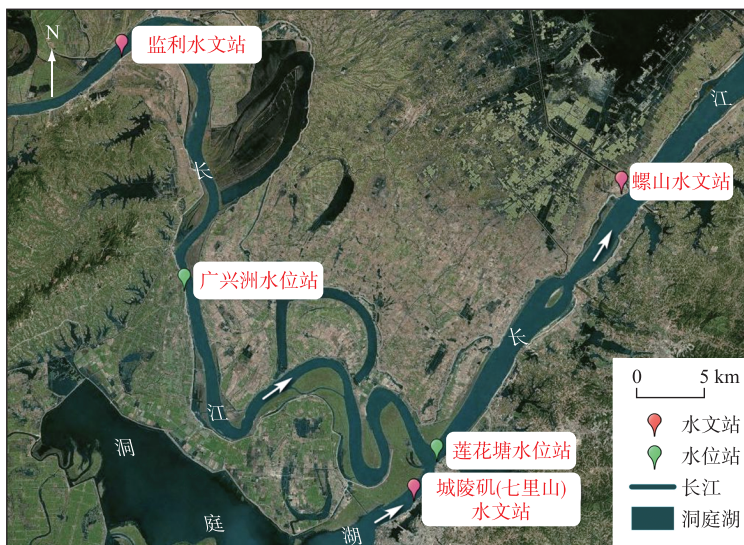


图 1 洞庭湖出口河段测站布置

Fig.1 Hydrometric station layout of Lake Dongting outlet reach

通过对莲花塘水位站 2016—2020 年中 18 个月的 5 min 水位自记仪数据进行分析, 根据水位数据“小涨落多、大涨落少”和渐变特征(图 2), 提取每日 24 h 内连续水位变幅的第 10 个百分点数(记为 $D_{10\%}$)作为当日水位波动值(记为 D). 通过对 2020 年莲花塘水位站 1—5 月 5 min 水位自记数据的研究发现, 当日水位平

均波动均值为 0.03 m, 其中 92.76% 的水位波动值 ≤ 0.05 m, 我们认为该波动主要受莲花塘水位站周边来往船舶及风浪对水位数据的影响. 根据《水位观测标准》(GB/T 50138—2010), 自记水位计的比测结果要求置信水平 95% 的综合不确定度不应超过 3 cm, 校测水位与自记水位系统偏差不应超出 ± 2 cm 范围, 因此, 本文约定单日水位波动值 $D > 0.05$ m 时, 视为发生“假潮”现象.

为研究与“假潮”现象产生的关联因子, 本文选取洞庭湖出口控制站城陵矶(七里山)水文站、长江干流上游监利水文站、下游螺山水文站的日平均流量、水位数据, 以及莲花塘日平均水位数据, 提取流量占比 $Q\%$ 、相邻站水位落差 ΔZ 、和各站流量 Q 、水位 Z 、单日涨落率 J 作为潜在自变量, 计算监利占螺山流量比 ($Q_{jl}\%$)、城陵矶(七里山)占螺山流量比 ($Q_{cl}\%$)、监利至莲花塘水位落差 (ΔZ_{jl})、城陵矶(七里山)至莲花塘水位落差 (ΔZ_{cl})、莲花塘至螺山水位落差 (ΔZ_{ls}), 以及各站流量 (Q_x)、水位 (Z_x)、单日涨落率 (J_x) (x 为站名首字母, 螺山站用 ls 表示). 经 Pearson 相关分析后, 去除非显著相关和相关性弱的参数, 再进行二元 Logistic 回归分析.

2.2 研究方法

2.2.1 二元 Logistic 回归分析法 二元 Logistic 回归分析法是一种用于检验因果关系假设和两分类预测的分析方法, 适用于因变量 Y 为定性数据, 并且只有两种状态. 一般要求样本数量大于自变量 X 数量的 15 倍以上, 因变量 Y 中发生数量占样本总数的 15% 以上. 本文主要采用该方法研究“假潮”发生的影响因子, 并提出判定模型. 其分析模型为:

$$p = \frac{1}{1 + e^{-(\alpha + \beta \cdot x)}} \quad (1)$$

变形后:

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta \cdot x \quad (2)$$

一般表示为:

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \cdots + \beta_n \cdot x_n \quad (3)$$

式中, p 表示 $Y=1$ 的概率; α 表示为回归方程的截距; β_i 为回归系数, 表示某一因素 x_i 改变一个单位时, 事件发生与不发生的概率之比的对数变化值.

2.2.2 快速傅立叶变换法 快速傅立叶变换(FFT)作为离散傅里叶变换(DFT)的快速算法, 一直以来都是频谱分析的重要工具^[15], 1965 年由美国科学家 J.W. 库利和 T.W. 图基提出. 本文依据 FFT 的滤波原理对水位波动数据进行滤波处理, 通过快速傅立叶变换将时间域信号转化为频率域信号, 再将频率域中的某些频率成分的振幅置零, 最后运用快速傅立叶逆变化转化为时间域实现滤波. 其主要计算公式为:

$$y(n) \approx \sum_{i=0}^M x_{n-1}(i) e^{-j\frac{2\pi}{N}ki} \quad (4)$$

式中, $y(n)$ 为滤波输出值; x_n 为时域离散信号值; M 为时域离散信号值 x_n 之前的信号点数; $N = M + 1$; $k = 0, 1, \cdots, M$.

2.2.3 局部加权回归算法 局部加权回归算法(local regression)^[16]由 Cleveland 于 1979 年提出, 是对二维散点图进行平滑的常用方法, 它结合了传统线性回归的简洁性和非线性回归的灵活性. 当要估计某个响应变量值时, 先从其预测变量附近取一个数据子集, 然后对该子集进行线性回归或二次回归, 回归时采用加权最

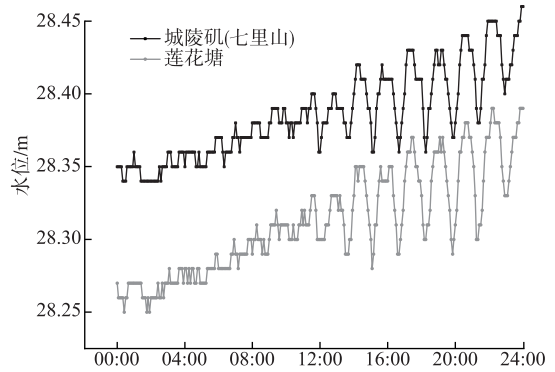


图 2 2020 年 6 月 19 日城陵矶(七里山)水文站和莲花塘水位站 24 h 5 min 水位波动

Fig.2 Water level fluctuation of Chenglingji (Qilishan) hydrology station and Lianhuatang water level station in 24 h and 5 min on June 19, 2020

小二乘法,即越靠近估计点的值其权重越大,最后利用得到的局部回归模型来估计响应变量的值,用这种方法进行逐点运算得到整条拟合曲线.

本文通过 R 语言中 loess 函数完成水位数据滤波计算,其中权重 w_i 采用 tri-cubed weight function 确定,公式如下:

$$w_i = \begin{cases} 1 - \left(\frac{d_i}{d_{max}}\right)^3 & \text{当 } d_i < d_{max} \text{ 时} \\ 0 & \text{当 } d_i \geq d_{max} \text{ 时} \end{cases} \quad (5)$$

式中, d_i 为观测点 x_i 到拟合点 x^* 的距离,即 $d_i = |x_i - x^*|$; d_{max} 为最大距离,即权重函数趋近于零的距离,或称为半窗口宽度; d_{max} 确定时,常与 R 语言中的滑动参数 span 有关.

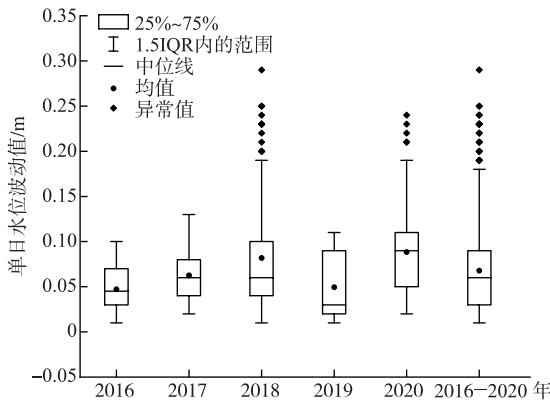


图3 莲花塘站 2016—2020 年单日水位波动值箱线图
Fig.3 Box plot of single-day water level fluctuation of Lianhuatang station from 2016 to 2020

范围内变化,2020 年波动均值最大,2016 年、2019 年波动均值最小,近 5 年单日水位波动值 D 均值为 0.07 m,中位数为 0.06 m,第 25%~75% 个数据集集中在 0.03~0.09 m 之间.

研究对象自变量特征值(表 1)可知监利站、城陵矶(七里山)站、螺山站日均流量分别在 9180~39300、1150~49200、16800~59300 m^3/s 之间;监利占螺山流量比 $Q_{j1}\%$ 在 15.83%~97.75% 之间,变化幅度达 81.92%;各站日均水位差值均在 9.59~10.75 m 之间,水位变化幅度基本一致;监利至莲花塘水位落差最小值为 0.61 m,发生在 2017 年 7 月 5 日,监利站日均水位为 34.66 m,莲花塘站日均水位为 34.05 m,当日莲花塘至螺山水位落差为 0.86 m,接近落差最小值(0.84 m),说明长江干流出流不畅,顶托现象严重. 监利、城陵矶(七里山)、莲花塘三站单日水位涨落率特征值基本一致.

3.2 二元 Logistic 回归分析结果

对研究对象运用 SPSS 20.0 软件进行单因素筛查分析,采用 Pearson 相关系数法对各变量进行相关性分析,显著性检验为双侧检验,选取水位波动值与各变量相关性分析结果见表 2.

由表 2 可知,自变量与水位波动值均呈现显著相关,除 ΔZ_{in} 外,其余自变量均在 $P < 0.01$ 水平上显著相关,其中监利日均流量相关系数最大,为 0.816,其余项相关系数均小于 0.7,整体而言相关性较弱,解释能力不强.造成该结果的可能原因其一是影响水位波动现象的关键因素不在本次分析之列,其二是由于样本数量较多,水位变化幅度大使得数据典型性不强.仅从各站间相关系数分析而言,可以看到监利站的流量、水位、落差数据相关性明显强于城陵矶(七里山)站、莲花塘站和螺山站.对此,结合本文研究目的,选取满足监利站高水位($Z_j \geq 34.00$ m)的 167 个样本数据进行 Pearson 相关性分析.结果表明,除 Q_{is} 、 J_j 外,其余自变量均在 $P < 0.01$ 水平上与水位波动值显著相关,相关系数较整体样本显著提升,其中强相关性的指标有监利日均流量 Q_j 、监利占螺山流量比 $Q_{j1}\%$ 、城陵矶(七里山)占螺山流量比 $Q_{c1}\%$ 、监利至莲花塘水位落差 ΔZ_{j1} .

3 结果与讨论

3.1 水位波动值与水文数据特征

从莲花塘水位站 2016—2020 年高水期 5 min 水位自记数据来看,水位波动瞬时值在 0.01~0.37 m 范围内浮动,最大波动瞬时值发生在 2018 年 7 月 7 日;经数据处理后,单日水位波动值范围在 0.01~0.29 m 之间,最大值发生在 2018 年 7 月 6 日.单日水位波动值 D 样本数 552 个,其中 $D > 0.05$ m 样本数为 282 个,占样本总数的 51.09%,2016—2020 年水位波动数据箱线图见图 3.由图 3 可知,分析月份中,从近 5 年单日水位波动变化幅度来看,年际间水位波动范围差异明显,表现为 $D_{2018} > D_{2020} > D_{2017} > D_{2019} > D_{2016}$,其中 2018 年最大值为 2016 年最大值的 2.9 倍;从单日水位波动均值来看,年际间水位波动较为集中,在 0.05~0.09 m

表 1 自变量特征值

Tab.1 Characteristic values of independent variables

项目	符号	单位	最大值	最小值	均值	差值
监利日均流量	Q_j	m^3/s	39300	9180	20200	30100
城陵矶(七里山)日均流量	Q_c	m^3/s	49200	1150	14400	48000
螺山日均流量	Q_{ls}	m^3/s	59300	16800	34500	42500
监利占螺山流量比	$Q_{jl}\%$	%	97.75%	15.83%	59.95%	81.92%
城陵矶占螺山流量比	$Q_{cl}\%$	%	82.97%	3.70%	40.39%	79.27%
监利日均水位	Z_j	m	37.25	27.66	32.60	9.59
城陵矶(七里山)日均水位	Z_c	m	34.71	23.96	29.71	10.75
莲花塘日均水位	Z_l	m	34.55	23.94	29.64	10.61
螺山日均水位	Z_{ls}	m	33.63	22.90	28.64	10.73
监利至莲花塘水位落差	ΔZ_{jl}	m	4.42	0.61	2.97	3.81
城陵矶至莲花塘水位落差	ΔZ_{cl}	m	0.53	-0.01	0.07	0.54
莲花塘至螺山水位落差	ΔZ_{ls}	m	1.99	0.84	1.00	1.15
监利涨落率	J_j	m/d	0.70	-0.51	—	1.21
城陵矶(七里山)涨落率	J_c	m/d	0.86	-0.39	—	1.25
莲花塘涨落率	J_l	m/d	0.77	-0.37	—	1.14

表 2 Pearson 相关性检验结果

Tab.2 Pearson correlation test results

项目	符号	Pearson 相关性	显著性(双侧)
水位波动值	D	1	
监利日均流量	Q_j	0.816**	<0.001
城陵矶(七里山)日均流量	Q_c	-0.172**	<0.001
螺山日均流量	Q_{ls}	0.358**	<0.001
监利占螺山流量比	$Q_{jl}\%$	0.565**	<0.001
城陵矶占螺山流量比	$Q_{cl}\%$	-0.577**	<0.001
监利日均水位	Z_j	0.459**	<0.001
城陵矶(七里山)日均水位	Z_c	0.294**	<0.001
莲花塘日均水位	Z_l	0.302**	<0.001
螺山日均水位	Z_{ls}	0.292**	<0.001
监利至莲花塘水位落差	ΔZ_{jl}	0.400**	<0.001
城陵矶至莲花塘水位落差	ΔZ_{cl}	-0.177**	<0.001
莲花塘至螺山水位落差	ΔZ_{ls}	0.101*	0.018
监利涨落率	J_j	0.208**	<0.001
城陵矶(七里山)涨落率	J_c	0.254**	<0.001
莲花塘涨落率	J_l	0.259**	<0.001

根据发生“假潮”判别标准,将水位波动值 D 进行二元分类,即 $D > 0.05$ m 表示为 1,反之表示为 0. 在 167 个样本数据中,分类为 1 的样本数 114 个,占比 68.26%. 根据单因素筛查分析结果,选取监利日均流量 Q_j 、监利占螺山流量比 $Q_{jl}\%$ 、监利至莲花塘水位落差 ΔZ_{jl} 和监利日均水位 Z_j 与二元分类变量运用 SPSS 20.0 软件进行二元 Logistic 回归模型分析,分析方法为 Enter. Hosmer 和 Lemeshow 检验结果($P=0.955$)表明模型和真实数据拟合良好(表 3).

结果表明,监利日均流量 Q_j 、监利占螺山流量比 $Q_{jl}\%$ 、监利至莲花塘水位落差 ΔZ_{jl} 和监利日均水位 Z_j 4 个变量中,仅 ΔZ_{jl} 显著($P < 0.05$),表明仅 ΔZ_{jl} 是发生“假潮”的独立影响因素. 通过对监利站高水位($Z_j \geq 34.00$ m)的 167 样本数据进行统计,结果如图 4,研究发现当 $\Delta Z_{jl} > 2.50$ m 时,发生“假潮”样本数为 114 个,

表3 二元 Logistic 回归变量结果
Tab.3 Binary Logistic regression variable results

项目	符号	Sig.	Exp (B)
监利日均流量	Q_j	0.121	1.002
监利占螺山流量比	$Q_{j1}\%$	0.386	0.750
监利日均水位	Z_j	0.334	0.048
监利至莲花塘水位落差	ΔZ_{j1}	0.021	7.537×10^{13}
常量		0.890	0

发生比例为 95.00%, 当 $\Delta Z_{j1} > 2.75$ m 时, “假潮”发生比例为 100%. 反之表明本次统计样本中, 发生“假潮”时均满足 $\Delta Z_{j1} > 2.50$ m.

对是否发生“假潮”进行预测模型分析, 方法采用 Forward: conditional, 分割点取 0.5, Hosmer 和 Lemeshow 检验结果 ($p = 0.672$) 表明模型和真实数据拟合良好. 其中单个变量 ΔZ_{j1} 的模型分析结果, $\alpha = -69.98, \beta_1 = 27.16$, 方程表达式为:

$$P = \frac{1}{1 + e^{69.98 - 27.16\Delta Z_{j1}}} \quad (\text{当 } Z_j \geq 34.00 \text{ 时}) \quad (6)$$

式中, P 为“假潮”发生概率; e 为自然常数; ΔZ_{j1} 为监利至莲花塘水位落差.

根据预测模型结果, 由公式 (6) 可知, ΔZ_{j1} 为影响“假潮”发生概率的主要影响因子. 通过对公式 (6) 进行验证分析 (计算结果 $P > 0.5$ 时, 预测值 = 1, 反之为 0), 结果表明, 当观测结果为 0 时 (即 $D \leq 0.05$ m, 未发生“假潮”), 该模型预测结果为 0 的百分比校正值为 92.5% (49/53), 当观测结果为 1 时 (即 $D > 0.05$ m, 发生“假潮”), 该模型预测结果为 1 的百分比校正值为 98.2% (112/114), 综合百分比校正值为 96.4% (161/167), 表明模型建立效果理想, ΔZ_{j1} 为发生“假潮”的主要影响因子.

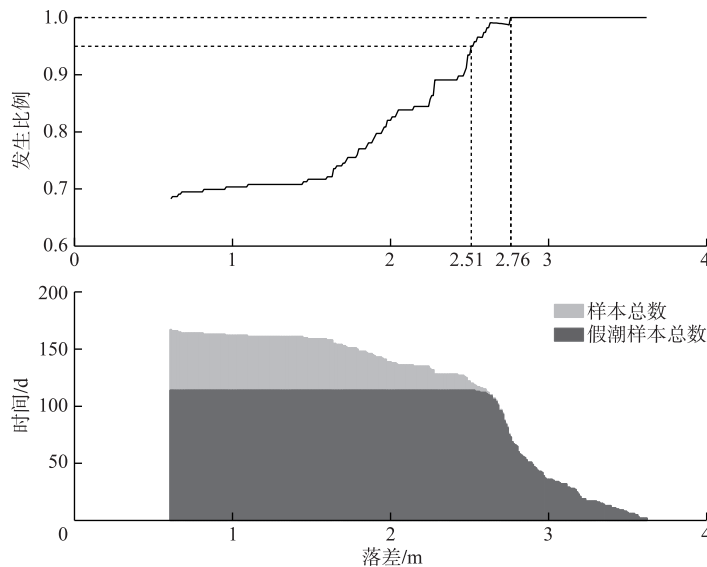


图4 监利站高水位期间“假潮”发生比例与落差 ΔZ_{j1} 的关系

Fig.4 The relationship between the occurrence ratio of “false tide” and ΔZ_{j1} drift at Jianli station during high water level

3.3 FFT、local regression 滤波结果

根据单口水位波动值特征, 本文分别运用 OriginPro 2016 与 R 语言对莲花塘水位站 2020 年 6 月 19 日 (2020 年首日出现“假潮”)、7 月 21 日 (2020 年首日 5 min 水位瞬时数据破保证水位)、21 日 (模拟 4~6 h、11 h、22~22 h 数据丢失)、28 日 (2020 年最高水位出现日) 单日 5 min 自记水位数据采用 FFT、local regression 方法进行滤波处理. 根据滤波后的线型变化趋势, 以及与人工拟合结果的一致性, 研究确定 FFT 分析中 pts 取值 13, local regression 分析中 span 取值 0.75. 两种方法滤波结果如图 5 所示, 滤波后采用值与人工拟合值进行误差计算 (视人工拟合数据为真值), 误差结果箱线图如图 6 所示.

通过对滤波后曲线与人工拟合曲线比较分析, 结果表明, FFT、local regression 滤波结果与人工拟合曲线

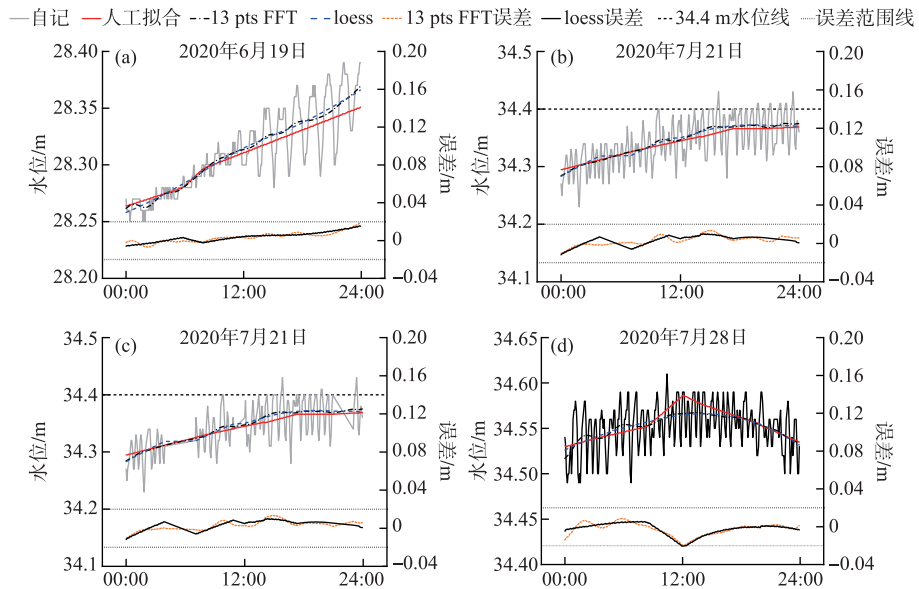


图 5 莲花塘水位站 2020 年特征日变化水位滤波结果

Fig.5 Filtering results of characteristic diurnal variation water level of Lianhuatang water level station in 2020

水位差均绝对值小于 0.02 m, 滤波结果可靠. 从图 5a 可以看到, 当瞬时波动值 ≤ 0.05 或 > 0.5 时, 采用 FFT、local regression 滤波方法对波动数据进行平滑处理的方法均适用, 说明该两种方法适用于不同振幅的波形态数据分析. 从图 5b 可以看到, 当莲花塘水位瞬时数据超过保证水位 34.40 时, 滤波后数据不受瞬时数据影响.

另一方面, 与时间加权平均法需参照最近 2 点的趋势延长来获取合理的水位不同^[7], 从图 5 中 24 h 数据可以看出, FFT、local regression 滤波法计算结果均不受水位瞬时值的影响, 可以准确地计算出相应水位. 从图 5b、c 比较可知, 当设备数据采集端或接收端故障导致数据空缺时(如空缺 1、2、3 h 数据), 拟合计算结果不受数据丢失影响.

从快速傅里叶变换法与局部加权回归算法误差值箱线图来看, 两种方法误差均值均为 -0.003 m, 中位数值均为 -0.001 m, 表明滤波结果整体误差较小; 快速傅里叶变换法上、下四分位数值(25%~75%)集中在 -0.007~0.002 m 之间, 局部加权回归算法上、下四分位数据(25%~75%)集中在 -0.005~0.001 m 之间, 表明局部加权回归算法误差值波动情况小, 数据更集中. 两种滤波方法的误差绝对值均在 0.02 m 之间, 表明两种方法均满足生产要求, 具有一定的实际用途, 但从误差值的波动情况来看, 局部加权回归算法更适用于对波动水位数据的处理.

3.4 水位波动成因分析与对策展望

城陵矶河段发生“假潮”现象是弯曲河道水力学特点所决定的. 从图 1 可以看到, 长江中游监利至莲花塘河段河道极为弯曲, 河流弯曲系数 Ka 为 1.74, 广兴洲至莲花塘河段 Ka 达 1.91. 同时, 在长江干堤的保护

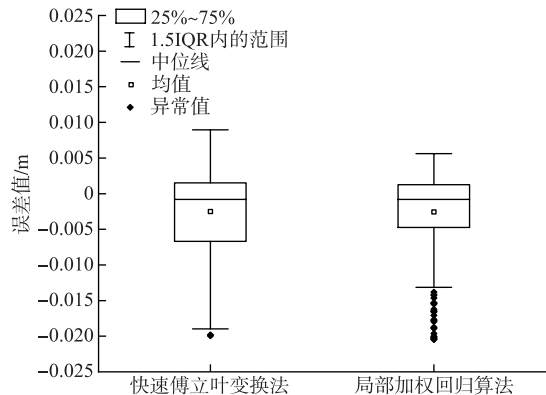


图 6 快速傅里叶变换法与局部加权回归算法误差值箱线图

Fig.6 Box plot of error values of fast Fourier transform and local regression algorithm

下,城陵矶河段水面宽度变化不大,当水流过弯时,流速降低,水面抬高,导致弯前河段水面比降减小,上游来水会短时间在弯道内蓄积,当水位抬升到一定高度后,弯后河段水面比降增大,出口流速增大,流量增加,弯道内槽蓄水量进入下一个弯道,此过程在每一个弯道内循环,调蓄河道来水.最终,莲花塘水位站上游河段来水在多个弯道串联调蓄作用下进入三江口(莲花塘水位站所在位置)弯道,在洞庭湖出口水流叠加影响下,两股水流能量相互抵消,弯道内流速降低、水位抬高作用更加显著,进一步放大了水位波动,形成“假潮”,与黄河中下游孙口至泇口河段形成“假潮”原因相似^[6].城陵矶(七里山)水文站对岸七弓岭河段自1998年修建防洪大堤后,长江来水不能直接从七弓岭抵达三江口,而需“掉头”北上后,再南下抵达莲花塘站,进一步加剧了“假潮”现象的水位波动值.

目前对于有过多“噪声”的数据,根据数据本身的完整性、波动幅度和时间连续性等一般采用滑动平均法(mean average)、指数滑动平均法(exponential mean average)、SG滤波法(Savitzky Golay Filter)、时间加权平均法(time weighted average)和平滑处理(smooth)等方法,刘超等^[7]根据时间加权平均法采用最接近报讯时段的前1.5 h的2 h加权平滑值,并参照最接近2点的趋势延长得到了较为合理的报讯水位,基本解决了数据波动问题,但该方法中根据最接近2点做趋势延长的做法实质是一种“预测”.本研究中,莲花塘水位站的报讯数据直接来源于水位自记设备采集的5 min瞬时数据,从图5可以看到,当水位呈现波动变化时,报讯数据则具有一定的不确定性.本文采用通讯工程中常使用的FFT滤波平滑处理方法和局部加权回归法对不规则波动水位数据进行处理,分析结果表明,滤波后平滑曲线走势与人工拟合曲线基本吻合,水位差值均小于0.02 m,表明平滑处理后的数据可靠,报讯水位与人工拟合“走中线”最大误差可达0.14 m,而基于该两种算法处理后的水位可以将该误差降低至0.02 m以下.同时,该方法适用于不同振幅的波形态数据分析,在汛期和非汛期均有效果,但中低枯水位级下,水位滤波必要性不足.此外,滤波后数据不受短时间数据丢失影响,不受瞬时数据影响,是实测数据的分析结果,本研究中选取的是3段24 h数据独立分析,数据时间跨度长,数据量丰富,计算结果可信度高.若将该算法应用到现有水情报讯系统中,在数据采集端或分中心接收端预置,则可以有效降低水位报讯误差,为防汛主管部门提供科学合理的实时水位数据,避免波动数据给防汛决策带来的干扰.

本次研究中,仅对历年中部分数据进行了系统性分析,难免会因样本量不足而出现分析误差,另一方面,对于两种滤波方法的适用性分析中,分析所选取的3个代表天水位数据基本完整,虽然对1~3 h内数据丢失的情况进行分析,但没有对存在明显异常数据(如一天中出现大量0值)的24 h数据进行分析,这些都需要在后续的工作中进一步论证、分析和完善.若通过对更长时间数据丢失进行计算发现滤波处理结果仍然可靠的话,对于类似于因仪器故障导致的0值的异常值,可以在数据处理之前进行预处理,舍弃异常数据后再进行分析.

4 结论

因受河流形态和江湖关系(长江与洞庭湖)的双重影响,作为长江中游城陵矶河段蓄滞洪区启用指标控制站的莲花塘水位站,其水位呈现出的“假潮”变化给蓄滞洪区启用等工作带来了一定决策困扰.为此,本文通过对莲花塘水位站近5年18个月份的5 min水位自记数据进行分析,确定了发生“假潮”现象的判别标准,再通过二元Logistic回归分析揭示影响“假潮”的主要因子,最后通过对比快速傅里叶变换法与局部加权回归两种算法,提出了应对“假潮”发生后水位序列处理的有效方法.具体主要结论如下:

1) 单口水位波动值 $D > 0.05$ m为发生“假潮”现象的判别标准;近5年单口水位波动值 D 均值为0.07 m,中位数为0.06 m,第25%~75%个数集中在0.03~0.09 m之间.

2) 二元Logistic回归分析结果表明,监利站高水位($Z_j \geq 34.00$ m)期间,当 $\Delta Z_{jt} > 2.50$ m时,“假潮”发生比例为95.00%,当 $\Delta Z_{jt} > 2.75$ m时,“假潮”发生比例为100%.模型预测结果综合百分比校正值为96.4%,表明模型建立效果理想, ΔZ_{jt} 为发生“假潮”的主要影响因素.

3) 快速傅里叶变换法和局部加权回归算法滤波平滑处理结果表明,24 h 5 min数据采用两种滤波方法进行滤波处理结果可靠;选择方法适用于不同振幅的波形态数据分析,不受短时间数据丢失影响,不受水位瞬时值的影响,具有较大的实用价值.从误差值的波动情况来看,局部加权回归算法误差值波动更小,更适

用于对波动水位数据的处理.

5 参考文献

- [1] Zhang WW, Ma Q, Li MZ *et al.* Hatch dates and growth rates of grass carp (*Ctenopharyngodon idellus*) and silver carp (*Hypophthalmichthys molitrix*) juveniles in Three Gorges Reservoir, Lake Dongting and Lake Poyang. *J Lake Sci*, 2020, **32**(3): 804-812. DOI: 10.18307/2020.0320. [张文武, 马琴, 黎明政等. 三峡水库和长江中下游通江湖泊(洞庭湖和鄱阳湖)草鱼、鲢的孵化日期及早期生长特征. 湖泊科学, 2020, **32**(3): 804-812.]
- [2] XI J. A speech at the forum on further promoting the development of Yangtze River Economic Zone(4/26/2018). Party Construction Articles; first half of the month, 2018, (7): 4-9.
- [3] State Council. The State Council's reply to the development plan for urban agglomerations in the middle reaches of the Yangtze River. *Gazette of the State Council of the People's Republic of China*, 2015, (11): 22-23. [国务院. 国务院关于长江中游城市群发展规划的批复. 中华人民共和国国务院公报, 2015, (11): 22-23.]
- [4] Xiao X, Mao BP, Wu SQ. Dynamics of hydrological characteristics in Changjiang River and Lake Dongting confluence reach in the past 100 years. *J Lake Sci*, 2021, **33**(1): 266-276. DOI: 10.18307/2021.0119. [肖潇, 毛北平, 吴时强. 近 100 年来长江与洞庭湖汇流河段水文特征演变. 湖泊科学, 2020, **33**(1): 266-276.]
- [5] Gao Y, Xie YH, Zou DS *et al.* Changes and driving factors of water surface area for inner lakes in Lake Dongting district in the past 40 years. *J Lake Sci*, 2019, **31**(3): 755-765. DOI: 10.18307/2019.0314. [高耶, 谢永宏, 邹冬生等. 近 40 年洞庭湖区内湖水面积变化及其驱动因素. 湖泊科学, 2019, **31**(3): 755-765.]
- [6] Fang CM, Zhong ZQ. The influence of capacity decrease on flood stages in Dongting Lake and in Yangtze River. *Journal of Hydraulic Engineering*, 2001, **32**(11): 70-74. [方春明, 钟正琴. 洞庭湖容积减小对洞庭湖和长江洪水位的影响. 水利学报, 2001, **32**(11): 70-74.]
- [7] Liu C, Peng YH, Peng YM *et al.* Research on smooth treatment of false tide of water level. *Journal of Changjiang Engineering Vocational College*, 2011, **28**(1): 27-28, 34. [刘超, 彭艳华, 彭艳明等. 水位“假潮”现象平滑处理方法研究. 长江工程职业技术学院学报, 2011, **28**(1): 27-28, 34.]
- [8] Song SQ, Li FJ. Analysis on the cause of “false tide” in the lower reaches of the Yellow River?. *Zhihuang Keji Xinx*, 2017, **246**(2): 3-5. [宋士强, 李福军. 黄河下游河段“假潮”成因分析. 治黄科技信息, 2017, **246**(2): 3-5.]
- [9] Gu YZ, Liu YQ, Cui CJ *et al.* Analysis of hydrological prediction methods in the “false tide periods” in Shandong reaches of the Yellow River. *Yellow River*, 2001, **23**(3): 22-23. [谷源泽, 刘以泉, 崔传杰等. 黄河山东段“假潮”期水文测报方法分析. 人民黄河, 2001, **23**(3): 22-23.]
- [10] Li CL, Fu CL, Zhang JM *et al.* Analysis and treatment on “false tide” in hydrologic data reorganization. *Journal of Water Resources and Water Engineering*, 2009, **20**(2): 153-155. [李春莲, 付春兰, 张建民等. 水文资料整编中对“假潮”现象的分析与处理. 水资源与水工程学报, 2009, **20**(2): 153-155.]
- [11] Fu X, Zhao QX, Sun ZH. Effects of 175 m experimental operation of the Three Gorges Reservoir on the storage capacity of Lake Dongting. *J Lake Sci*, 2019, **31**(6): 1713-1725. DOI: 10.18307/2019.0603. [付湘, 赵秋湘, 孙昭华. 三峡水库 175m 试验性蓄水期调度运行对洞庭湖蓄水量变化的影响. 湖泊科学, 2019, **31**(6): 1713-1725.]
- [12] Tang C, Qian B, Li WQ *et al.* Pollution characteristics and risk evaluation of heavy metals in surface sediments of Dongting Lake. *Yangtze River*, 2020, **51**(6): 49-56, 62. [唐聪, 钱宝, 李炜钦等. 洞庭湖区表层沉积物重金属污染特征与风险评价. 人民长江, 2020, **51**(6): 49-56, 62.]
- [13] Qian B, Tang C, Yang Y *et al.* Pollution characteristics and risk assessment of heavy metals in the surface sediments of Dongting Lake water system during normal water period. *European Journal of Remote Sensing*, 2021, **54**(sup2): 211-221. DOI: 10.1080/22797254.2020.1763207.
- [14] Ke WL, Chen CZ, Ji HX *et al.* A loop-like relationship between water surface area of Lake Dongting and water level at Chenglingji, the Yangtze River. *J Lake Sci*, 2017, **29**(3): 753-764. DOI: 10.18307/2017.0325. [柯文莉, 陈成忠, 吉红霞等. 洞庭湖水面积与城陵矶水位之间的绳套关系. 湖泊科学, 2017, **29**(3): 753-764.]
- [15] Yang J, Lv JQ. Study on FFT filtering error. *Communications Technology*, 2010, **43**(6): 235-237, 240. [杨君, 吕镜清. FFT 滤波误差分析. 通信技术, 2010, **43**(6): 235-237, 240.]
- [16] Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 1979, **74**(368): 829-836. DOI: 10.1080/01621459.1979.10481038.