

DOI: 10.13476/j.cnki.nsqdqk.2022.0114

位文涛, 靳燕国, 张召, 等. 南水北调中线工程流量监测站点倒挂数据清洗模型及应用[J]. 南水北调与水利科技(中英文), 2022, 20(6): 1158-1167. WEI W T, JIN Y G, ZHANG Z, et al. Application of inverted data cleaning for flow monitoring stations in the middle route of the South-to-North Water Transfer Project[J]. South-to-North Water Transfers and Water Science & Technology, 2022, 20(6): 1158-1167. (in Chinese)

南水北调中线工程流量监测站点 倒挂数据清洗模型及应用

位文涛¹, 靳燕国², 张召³, 雷晓辉^{1,3}, 薛萍⁴, 王艺霖³

(1. 河北工程大学水利水电学院, 河北 邯郸 056038; 2. 中国南水北调集团中线有限公司, 北京 100038;
3. 中国水利水电科学研究院, 北京 100038; 4. 河海大学计算机与信息学院, 南京 211100)

摘要:以南水北调中线工程为研究对象, 针对调度运行中出现监测站点流量倒挂的数据异常现象, 基于水量动态平衡原理及区间流量最长序列法, 构建流量监测站点倒挂数据清洗模型。以中线工程白河至黄金河河段为例, 基于模型清洗后的流量数据作为上边界条件, 构建水动力模型进行验证。结果表明: 清洗模型消除了流量监测数据中存在的倒挂现象, 提高了流量监测数据的质量, 并发现节制闸流量监测设备存在的监测数据偏差及闸门调控的干扰是造成该现象的主要原因; 水动力数值模拟的闸前水位平均绝对误差减少 0.075 7 m, 均方根误差减小 0.089 5 m, 清洗后的流量数据在数据空间的逻辑性、一致性等方面优于实测流量数据。

关键词: 南水北调中线工程; 数据清洗; 流量倒挂; 水量平衡; 输水损失; 水动力模型

中图分类号: TV698 文献标志码: A 开放科学(资源服务)标识码(OSID):



近年来, 随着南水北调中线工程信息化水平的逐步提高, 工程沿线监测站点的分布越来越密集, 水情信息监测及监测数据质量越来越受到人们关注^[1]。水情监测数据是调水工程中开展科学研究和进行调度决策的基础^[2]。然而, 在水情数据监测过程中, 手动数据记录错误、监测设备故障和数据传输错误等主观及客观因素的影响^[3-4]造成数据监测异常频发, 直接影响数据监测的质量, 并间接影响输水过程中调度决策的制定, 增加调水成本。流量监测作为水情信息监测的重要组成部分, 对渠道的水力控制、输水系统水力特性分析等均具有重要意义^[5]。明渠调水过程中连续多个时刻, 下游节制闸过闸流量持续大于上游节制闸过闸流量, 称之为发生流量监测点位数据倒挂, 即“流量倒挂”现象。该种现象将对以流量数据为载体的工程应用与理论

研究产生不利影响: 如渠道蓄量推求, 连续长时间的“流量倒挂”现象, 将导致计算得到的渠池蓄水量低于实际蓄水量, 影响现实工程运行安全; 如一维水动力模拟, 若采用流量数据为上边界条件且数据存在“流量倒挂”现象, 将导致模拟计算得到的沿线水位高于实际, 加剧模拟误差。

水情数据清洗可看作数据清洗领域在水利行业的应用与拓展。在已有的与水利行业相关的数据清洗研究中: 侍建国等^[6]结合区域水文数据具有时间系列长和数据量大的特点, 采用拉伊达准则法剔除异常值, 并取得明显效果; 陈琼妮^[7]针对水情数据监测过程中可能存在的缺失、离群等数据问题, 提出了一种基于 Hadoop 的分布式、清洗策略可选的清洗方案, 实验证明该种方案能显著提高数据清洗算法性能, 对缺失数据的填充精度提高约 10%。近

收稿日期: 2022-07-20 修回日期: 2022-10-29 网络出版时间: 2022-11-22

网络出版地址: <https://kns.cnki.net/kcms/detail/13.1430.TV.20221121.1414.004.html>

基金项目: 中国水利水电科学研究院基本科研业务费专项项目(01882106); 极端工况下中线工程预测预演与应急调控关键技术(ZXJ/ZD/YW-2022-004)

作者简介: 位文涛(1996—), 男, 河北辛集人, 主要从事水文与水资源研究。E-mail: 13271498985@163.com

通信作者: 雷晓辉(1974—), 男, 陕西渭南人, 正高级工程师, 博士, 博士生导师, 主要从事明渠调度及水力控制等研究。E-mail: lxh@jwhr.com

些年,人工智能技术的快速发展使得数据异常诊断和插补方向往人工智能、机器学习、深度学习的领域延伸^[8-14]。金容鑫等^[15]针对水电机组状态提出了监测数据的清洗方法,并通过实测监测参数验证了方法的有效性;闫亚男等^[16]充分考虑时间序列数据特性,利用DBSCAN算法进行异常检测,将随机森林预测与时间序列分析结合进行缺失值填充,完成对水电遥测数据的清洗工作;郑霞忠等^[17]提出了一种考虑关联规则约束的聚类算法识别大坝位移监测异常值,并利用PSO-LSSVM模型重构异常数据,提高了大坝安全监测数据的准确性。水情监测数据按数据排列方式不同可分为时间序列排序和空间序列排序,且不同的数据排列方式可表现出不同的物理意义。上述清洗研究大多以数据的时间序列关系为主要清洗目标,如水电机组状态和水电遥测等数据,未曾对数据的空间序列关系进行考虑。明渠调水工程中流量数据以时间序列关系进行排列可表示节制闸不同时期的工作状态,以空间序列关系排列可反映当前渠段水量关系。因此,为保证流量数据的逻辑性和一致性,需要在时间和空间两个尺度上对流量数据进行清洗。在考虑时间序列关系的前提下,将数据的空间联系纳入分析,即考虑了同一渠道不同空间位置流量数据之间的变化规律及联系,构建流量监测站点倒挂清洗模型,实现对流量监测数据全方位诊断和清洗。

针对南水北调中线工程中出现的流量监测站点数据倒挂现象,考虑流量数据的空间尺度关系,运用水量动态平衡原理和区间流量最长序列法,建立监测站点数据倒挂清洗模型,对倒挂数据进行成因分析,并利用清洗前后流量数据作为上边界条件验证数据清洗对水动力模型模拟结果的影响,以期为中线工程的水力计算及调度运行提供可靠的数据支撑。

1 研究区域及数据选取

1.1 研究区域

为支撑南水北调中线工程的运行调度管理,在沿线布设大量监测站点,形成了包括闸前水位、闸后水位和过闸流量等水情数据及工情数据的监测站网。其中,流量监测点位共计161个,多位于节制闸或分退水口上、下游100~500 m范围内。查阅工程设计参数得知,部分节制闸实测流量数据为节制闸流量与前方退水闸流量之和,需在流量计算时

加以考虑。研究范围内的白河节制闸至黄金河节制闸段属于河南段,长43 km,采用明渠自流输水,包含2个渠池,白河节制闸、东赵河节制闸及黄金河节制闸共3座节制闸,以及半坡分水口、大营分水口、十里庙分水口和清河退水闸共4个分退水口,研究范围所在位置见图1。

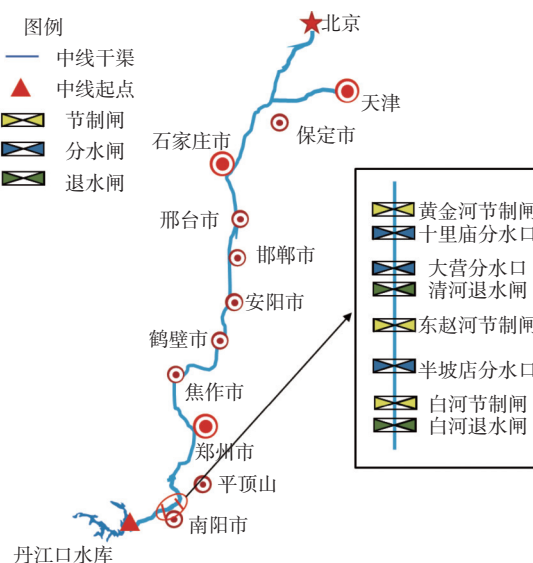


图1 南水北调中线工程输水线路

Fig. 1 Water transmission line of the middle route of the South-to-North Water Transfer Project

1.2 研究数据选择

为便于开展监测点位数据倒挂清洗研究,避免降水入渠、冰期冰冻消融和渠段沿线蒸发^[18]等因素对输水损失率^[19]的直接影响和对倒挂数据清洗的间接影响,以全线无冰覆盖、降雨较少和沿线气候差异较小为标准,选择研究时段。综合分析后确定2018年4月份作为研究时段,并根据4月份调度运行数据,选择流量倒挂次数较多的23日14时数据进行修正,其中23日14时全线流量数据见图2。

2 研究方法

2.1 输水损失率

对于南水北调中线工程,任意两时段输水损失可依据水量动态平衡原理进行推求,公式为

$$Q_L = \frac{(Q_{in} - Q_{out} - Q_r)\Delta t + V_1 - V_2}{\Delta t} \quad (1)$$

式中: Q_L 为时段内渠段输水损失流量, m^3/s ; Δt 为时段长, s ; Q_{in} 、 Q_{out} 分别为渠段入渠流量和出渠流量, m^3/s ; Q_r 为时段内分水流量和退水流量之和, m^3/s ; V_1 、 V_2 分别为时段初始和终止时刻渠段蓄水体积, m^3 ,可根据实测水位采用分段求和法计算求得。用

输水损失流量除以入渠流量,可得无量纲化参数输水损失率 R_L :

$$R_L = \frac{Q_L}{Q_{in}} \quad (2)$$

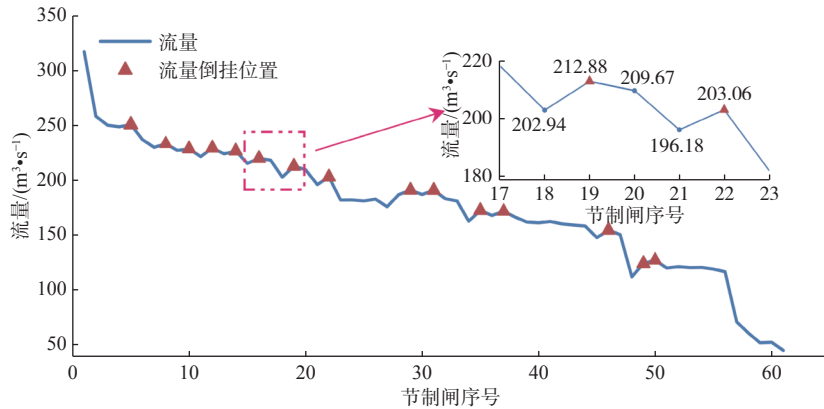


图 2 研究时段节制闸过闸流量

Fig. 2 Control gate over-gate flow during the study period

2.2 数据清洗方法

监测站点倒挂数据清洗需推求出各个节制闸理论流量值,作为划分区间的依据。为保证节制闸理论流量数据计算结果的可靠性,需选取一个可靠监测断面的流量监测数据为基准开展清洗工作。倒挂数据清洗具体步骤如下。

步骤 1: 依据水量动态平衡原理计算各个渠段的输水损失率。

步骤 2: 节制闸理论流量推求。基于水量动态平衡原理由所选取的特定监测断面向下游推求各个节制闸(目标节制闸)理论流量值,并求得无量纲流量偏差系数,计算公式为

$$Q_{up,actual} - Q_s - Q_{up} \cdot R_L \cdot \frac{l}{L} = Q_{down,theory} \quad (3)$$

$$q_l = \frac{Q_{down,actual} - Q_{down,theory}}{Q_{down,theory}} \quad (4)$$

式中: L 为渠段总长度, m; l 为渠段内首个节制闸至目标节制闸距离, m; $Q_{up,actual}$ 为研究时段选取监测断面流量, m^3/s ; $Q_{down,theory}$ 为经水量平衡计算推求得到目标节制闸理论流量, m^3/s ; Q_s 为选取监测断面至目标节制闸区间分退水流量之和, m^3/s ; $Q_{down,actual}$ 为目标节制闸实测流量, m^3/s ; q_l 为无量纲流量偏差系数。

步骤 3: 清洗区间划分。统计全线各节制闸流量偏差系数,在保证下游监测断面流量小于上游监测断面流量的前提下,考虑中线节制闸流量监测存在 $\pm 5\%$ 偏差^[20],故按流量偏差系数绝对值小于 1% 将中线总干渠划分为若干区间。

步骤 4: 区间内待修正流量数据筛选。依次对

各个区间内流量数据进行检查和筛选。考虑到倒挂数据的清洗存在“牵一发,动全身”的特性,即对当前节制闸流量数据的修正,会对与其相邻的其他节制闸数据造成影响。因此,待修正流量数据的筛选采用区间流量最长序列法。区间最长序列法如下。

确定选择标准。考虑沿线部分节制闸流量监测设备的可靠性,故以区段无倒挂数据且流量数据修改最少为数据选择标准。

筛选待清洗流量数据。依次以区间内首个节制闸、第二个节制闸、第三个节制闸……为参考目标,按所选定的标准,向下游排列计数。统计各个节制闸为参考时,待清洗流量个数。若参考节制闸非区间内首个节制闸,则将区间重新划分为两段,重复该步骤对两段进行流量个数统计。

统计结束后,按标准进行选择,确定参考节制闸,得到待清洗数据。

步骤 5: 区间流量清洗。根据查找得到的最长序列,利用水量动态平衡原理对区间内节制闸数据进行清洗,清洗公式为

$$Q_{up} - Q_{fen} - Q_{up} \cdot R_L \cdot \frac{l_s}{L_s} = Q_{down} \quad (5)$$

式中: Q_{up} 为区间首断面节制闸流量, m^3/s ; Q_{fen} 为区间内首断面到待清洗节制闸分水流量和退水流量之和, m^3/s ; Q_{down} 为待清洗节制闸推求流量值, m^3/s ; l_s 为区间首断面至待清洗节制闸距离, m; L_s 为区间长度, m。

2.3 一维水动力模型

水情监测和水动力仿真模拟是常用的获取调水工程监测断面水情状态的手段,旨在利用水动力模

型检验流量监测点位倒挂数据清洗前后对仿真模拟精度的影响作用。在明渠调水工程中,调水长度往往大于输水断面的宽度和深度,因此可将调水系统的水动力仿真近似概化为一维问题。一维水动力模型采用的基本控制方程为一维非恒定流的圣维南方程组^[21]:

$$\begin{cases} B\frac{\partial Z}{\partial t} + \frac{\partial Q}{\partial X} = q \\ \frac{\partial Q}{\partial t} + \frac{\partial(\alpha\frac{Q^2}{A})}{\partial X} + gA\frac{\partial Z}{\partial X} + \frac{gQ|Q|}{C^2AR} = 0 \end{cases} \quad (6)$$

$$C = \frac{1}{n}\sqrt{R} \quad (7)$$

式中: B 为过水断面表面宽度, m; Z 为水位, m; t 为时间, s; Q 为流量, m^3/s ; X 为沿主流向的渠道纵向距离, m; q 为渠道旁侧入流, m^3/s ; α 为动量校正系数; A 为过水面积, m^2 ; g 为重力加速度, m/s^2 ; C 为谢才系数; R 为水力半径, m; n 为输水渠道曼宁糙率系数。

控制方程可采用 Preissman 四点带权隐式差分格式^[22-23]对圣维南方程进行离散,将水位和流量交替布置在网格上,后利用“追赶法”进行求解差分方程^[24]。

3 研究实例

3.1 监测站点流量数据倒挂成因分析

统计中线工程各节制闸发生监测站点流量数据倒挂次数及占比情况,见表1。统计结果显示:全线节制闸4月份监测站点流量数据平均倒挂次数为118次,其中:刁河、澧河、小洪河、孟坟河、鹤壁淇河、安阳河、午河及西黑山共8座节制闸均未发生监测站点流量数据倒挂现象;草墩河、澎河、双泊河、潮河、黄水河及漳河发生监测站点流量数据倒挂次数较多,占比分别为:89.72%、100%、99.72%、100%、95.83%、99.72%;其余节制闸监测站点流量数据倒挂次数发生较少,占比均小于20%。

表1 南水北调中线工程全线节制闸流量倒挂情况统计

Tab. 1 Statistical table of flow inversion of control gates across the whole line of the middle route of the South-to-North Water Transfer Project

节制闸名称	流量倒挂个数/个	占比/%	节制闸名称	流量倒挂个数/个	占比/%
陶岔渠首	0	0	孟坟河倒虹吸出口	0	0
刁河渡槽进口	0	0	香泉河倒虹吸出口	39	10.83
湍河渡槽进口	34	9.44	鹤壁淇河倒虹吸出口	0	0
严陵河渡槽进口	44	12.22	汤河涵洞式渡槽进口	255	70.83
淇河倒虹吸出口	209	58.06	安阳河倒虹吸出口	0	0
十二里河渡槽进口	1	0.28	漳河倒虹吸出口	359	99.72
白河倒虹吸出口	92	25.56	牯牛河南支渡槽进口	2	0.56
东赵河倒虹吸出口	230	63.88	沁河倒虹吸出口	43	11.94
黄金河倒虹吸出口	23	6.39	洺河渡槽进口	250	69.44
草墩河倒虹吸出口	323	89.72	南沙河倒虹吸出口	265	73.61
澧河渡槽进口	0	0	七里河倒虹吸出口	55	15.28
澎河渡槽进口	360	100	白马河倒虹吸出口	235	65.28
沙河渡槽进口	18	5.00	李阳河倒虹吸出口	65	18.06
玉带河倒虹吸出口	243	67.50	午河渡槽进口	0	0
北汝河倒虹吸出口	60	16.67	槐河(一)倒虹吸出口	52	14.44
兰河涵洞进口	201	55.83	洹河倒虹吸出口	1	0.28
颍河倒虹吸出口	53	14.72	古运河暗渠进口	2	0.56
小洪河倒虹吸出口	0	0	薄沱河倒虹吸出口	134	37.22
双泊河渡槽进口	359	99.72	磁河倒虹吸出口	242	67.22
梅河倒虹吸出口	81	22.50	沙河(北)倒虹吸出口	193	53.61
丈八沟倒虹吸出口	27	7.50	漠道沟倒虹吸出口	177	49.17
潮河倒虹吸出口	360	100	唐河倒虹吸出口	36	10.00
金水河倒虹吸出口	0	0	放水河渡槽进口	246	68.33

表 1 (续)
Tab. 1 (Continued)

节制闸名称	流量倒挂个数/个	占比/%	节制闸名称	流量倒挂个数/个	占比/%
须水河倒虹吸出口	218	60.56	蒲阳河倒虹吸出口	7	1.94
索河渡槽进口	135	37.50	岗头隧洞进口	24	6.67
穿黄隧洞出口	177	49.17	西黑山	0	0
济河倒虹吸出口	15	4.17	瀑河倒虹吸出口	182	50.56
闫河倒虹吸出口	156	43.33	北易水倒虹吸出口	7	1.94
寨河倒虹吸出口	281	78.06	坟庄河倒虹吸出口	270	75.00
峪河暗渠进口	41	11.39	北拒马暗渠进口	1	0.28
黄水河支倒虹吸出口	345	95.83			

3.1.1 监测设备影响分析

监测设备在数据监测过程中往往存在系统误差,导致流量监测数值较理论值偏大或偏小。因此,选取监测站点流量倒挂次数占比大于 80% 的节制闸进行监测设备影响分析,计算各个节制闸及上游节制闸流量相对理论流量偏差占比情况并求平均值(平均相对偏差),计算结果见表 2。

表 2 所选节制闸及上游节制闸平均相对偏差统计
Tab. 2 Statistics of the average relative deviation of the selected control gate and the upstream control gate %

节制闸	平均相对偏差
草墩河倒虹吸出口节制闸	-0.60
澎河渡槽进口节制闸	1.12
双泊河渡槽进口节制闸	6.80
潮河倒虹吸出口节制闸	0.15
黄水河支倒虹吸出口节制闸	2.59
漳河倒虹吸出口节制闸	-0.62
黄金河倒虹吸出口节制闸	-3.84
澧河渡槽进口节制闸	-3.78
小洪河倒虹吸出口节制闸	-0.88
丈八沟倒虹吸出口节制闸	-5.15
峪河暗渠进口节制闸	0.42
安阳河倒虹吸出口节制闸	-6.02

黄金河倒虹吸出口节制闸作为草墩河倒虹吸进口节制闸的上游节制闸,其 4 月份监测站点流量倒挂数据个数为 23,占比为 6.39%。黄金河与草墩河 4 月份实测流量数据与理论流量数据对比见图 3。由图 3 可知:黄金河的实际流量比理论值要小,经计算平均相对偏差为-3.84%;草墩河实测流量与理论值较为接近,平均相对偏差为-0.6%。综上所述,可得草墩河倒虹吸进口节制闸发生监测站点流量倒挂现象的主要原因是上游黄金河流量监测设备观测值偏小。

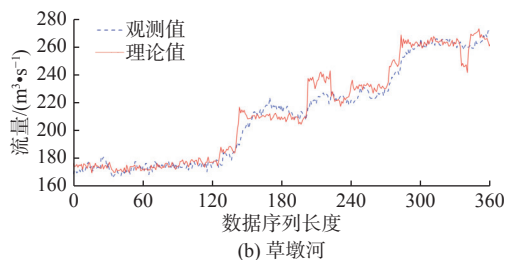
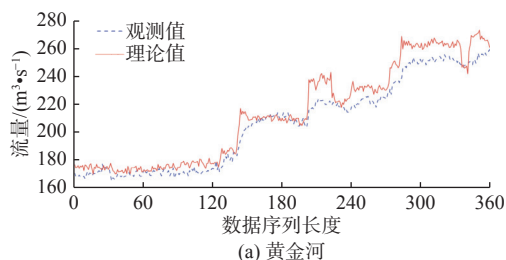


图 3 黄金河与草墩河实测流量与理论流量对比

Fig. 3 Comparison of measured and theoretical flow between Huangjin River and Caodun River

分别对澎河渡槽进口节制闸、双泊河渡槽进口节制闸、潮河倒虹吸出口节制闸、黄水河支倒虹吸出口节制闸及漳河倒虹吸出口节制闸进行分析,得出结论如下:澎河渡槽进口节制闸、漳河倒虹吸出口节制闸和潮河倒虹吸出口节制闸监测站点流量数据倒挂产生的原因,均为上游节制闸监测设备观测值偏低所致;双泊河渡槽进口节制闸和黄水河倒虹吸出口节制闸监测点位流量数据倒挂产生的原因则为自身监测设备观测值偏高所致。

3.1.2 节制闸调控影响分析

选取监测站点流量倒挂次数占比小于 20% 的节制闸进行节制闸调控影响分析,统计节制闸的总调控次数和倒挂事件中发生的调控次数,统计结果见表 3。十二里河、牯牛河南支、洹河、古运河、蒲阳河、北易水及北拒马 7 座节制闸,由于流量倒挂事件中的调控次数较少,故舍弃。其中:沙河渡槽进口节制闸共发生 18 次监测站点流量倒挂事件,包

括闸门调控 13 次,占比为 72.2%;丈八沟倒虹吸出口节制闸共发生 27 次监测站点流量倒挂事件,包括闸门调控 17 次,占比为 62.9%;岗头隧洞进口节制闸共发生 24 次监测站点流量倒挂事件,包括闸门调

控 10 次,占比为 41.67%。因此,可得出沙河渡槽进口节制闸、丈八沟倒虹吸出口节制闸及岗头隧洞进口节制闸监测点位倒挂事件的发生受到节制闸调控干扰的影响。

表 3 4 月份节制闸调控次数统计

Tab. 3 Statistics on the number of control gates in April

节制闸名称	总调控次数	流量倒挂事件中调控次数	节制闸名称	总调控次数	流量倒挂事件中调控次数
湍河渡槽进口	33	3	洧河倒虹吸出口	38	0
十二里河渡槽进口	28	0	古运河暗渠进口	35	0
沙河渡槽进口	18	13	蒲阳河倒虹吸出口	49	1
丈八沟倒虹吸出口	27	17	岗头隧洞进口	24	10
济河倒虹吸出口	41	5	北易水倒虹吸出口	68	0
牯牛河南支渡槽进口	29	1	北拒马暗渠进口	0	0

3.2 清洗结果展示

研究范围内共包含 63 座节制闸,故可将总干渠划分为 62 个渠段。考虑到沿线不同渠段的工程参数设计及布置方式具有一定的相似性^[25],且在中线工程年度水量调度计划制定及水量调度时,以陶盆渠首、刁河渡槽进口节制闸、草墩河倒虹吸进口节制闸、穿黄隧洞出口节制闸、漳河倒虹吸出口节制闸、岗头隧洞进口节制闸和北拒马河暗渠进口节制

闸作为主要控制断面。因此,将中线工程自陶盆渠首至北拒马暗渠进口节制闸划分为 5 段并计算各渠段内输水损失率。通过对南水北调中线工程沿线节制闸和分退水口水情数据进行一致性检验,剔除监测设备断电导致的监测数据跳变,并利用水量动态平衡原理推求得到各个渠段对应输水损失率,计算结果见表 4。

表 4 南水北调中线工程各渠段输水损失率

Tab. 4 Water loss rate of each canal section in the middle route of the South-to-North Water Transfer Project

渠段名称	陶盆至草墩河	草墩河至穿黄	穿黄至漳河	漳河至岗头	岗头至北拒马
输水损失率	0.968	5.836	2.744	3.625	-1.773

陶盆渠首出口流量监测数据是南水北调中线工程水资源跨流域入渠水量监测的重要依据,为南水北调中线工程引水流量计算提供了精确的数据支

撑^[26]。因此,可靠监测断面选择陶盆渠首监测断面。利用上述研究方法对 2018 年 4 月份数据进行修正,23 日 14 时修正效果图见图 4。

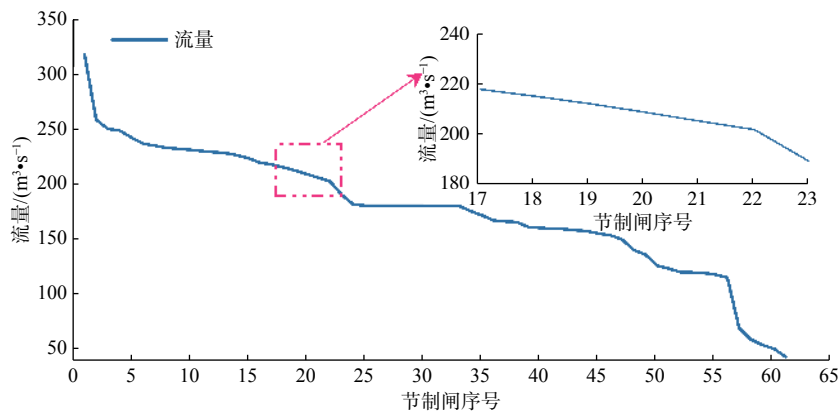


图 4 23 日 14 时数据清洗结果

Fig. 4 Data cleaning result chart at 14 : 00 on the 23rd

3.3 一维水动力模型验证

分别利用清洗前后的流量数据作为上游边界条

件,构建一维水动力模型,检验清洗模型对水动力模拟的影响。本次仿真模拟所选渠段为南水北调

中线工程白河至黄金河段,模型上边界条件采用白河倒虹吸出口节制闸过闸流量,下边界条件采用黄金河倒虹吸出口节制闸闸前水位,模拟东赵河倒虹吸出口节制闸未来1个月的闸前水位的变化过程,并将清洗前后的模拟结果与实测数据进行对比。水动力模拟结果见图5。

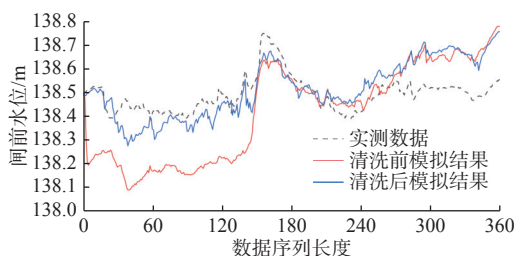


图5 东赵河节制闸水动力模拟结果对比

Fig. 5 Comparison of hydrodynamic simulation results of Dongzhao River control gate

对图5进行分析发现:在模拟结果对比中,存在前155条数据清洗后仿真模拟结果优于清洗前,而后205条数据清洗前后仿真模拟结果相近的现象。对该现象成因进行分析:前155条数据中,清洗前上游白河节制闸流量数据均小于下游东赵河节制闸。通过清洗模型对该时刻全线流量数据进行清洗,清洗后白河节制闸流量数据均发生改变且大于东赵河流量数据,清洗结果保证了流量数据的逻辑性,故最终模拟结果优于清洗前。后205条数据中,清洗前上游白河节制闸流量数据小于东赵河节制闸流量数据的占比为9.7%,清洗后流量数据发生改变,且在水动力模拟过程中,边界数据的改变对仿真模拟结果有直接影响,故清洗前后仿真模拟结果相近。

统计水动力模型模拟结果,基于清洗前流量数据的仿真模拟结果中,闸前水位绝对误差范围为0.000 6~0.373 0 m,平均绝对误差为0.153 7 m,相对误差范围为-0.27%~0.17%,均方根误差为0.184 5 m;基于清洗后数据的仿真模拟结果中,闸前水位绝对误差范围为0.000 5~0.207 0 m,平均绝对误差为0.078 0 m,相对误差范围为-0.14%~0.15%,均方根误差为0.095 0 m。通过对比可看出,基于清洗后的数据进行水动力模拟仿真精度更高,证明了流量倒挂数据清洗模型的合理性。

4 结论

基于水量动态平衡原理及区间流量最长序列法对2018年4月份历史数据进行清洗,结果表明该种

清洗方法能有效解决流量倒挂数据异常问题。

对2018年4月份的倒挂数据进行分析,得到流量监测设备监测值的偏差及闸门调控的干扰是导致监测站点流量倒挂现象的主要原因。

以白河节制闸清洗前后流量数据为上边界条件构建水动力模型,通过对东赵河节制闸闸前水位的仿真模拟结果进行对比,发现基于清洗后的流量数据作为上边界条件能有效提高模拟精度。研究结果证明清洗后的流量数据在数据空间的逻辑性、一致性等方面优于实测流量数据,可在实际工程中推广应用。

参考文献(References):

- [1] 李月强,马芳平,薛萍,等.明渠输水流量不平衡量感知及水位超限预警方法研究[J].水利水电技术(中英文),2022,53(1):73-82.(LI Y Q, MA F P, XUE P, et al. Research on perception of open channel water flow imbalance and early warning method of water level overrun[J]. *Water Resources and Hydropower Engineering*, 2022, 53(1): 73-82. (in Chinese)) DOI: 10.13928/j.cnki.wrahe.2022.01.008.
- [2] 展培尧.南水北调东线工程山东段调度运行水情监测与管理系统的分析与设计[D].济南:山东大学,2014.(ZHAN P Y. The analysis and design of the east route of South-to-North Water Transfer Project in Shandong hydrological monitoring and dispatching management system[D]. Jinan: Shandong University, 2014. (in Chinese))
- [3] 廖书妍.数据清洗研究综述[J].电脑知识与技术,2020,16(20):44-47.(LIAO S Y. A review of data cleansing research[J]. *Computer knowledge and technology*, 2020, 16(20): 44-47. (in Chinese)) DOI: 10.14004/j.cnki.ckt.2020.2361.
- [4] ELSHORBAGY A, SIMONOVIC S P, PANU U S. Estimation of missing streamflow data using principles of chaos theory[J]. *Journal of Hydrology*, 2002, 255(1-4): 123-133. DOI: 10.1016/s0022-1694(01)00513-3.
- [5] 王艺霖,靳燕国,陈晓楠,等.LSTM神经网络和量纲分析法在弧形闸门过流计算中的对比[J].南水北调与水利科技(中英文),2022,20(3):590-599.(WANG Y L, JIN Y G, CHEN X N, et al. Comparison between LSTM neural network and dimensional analysis method in discharge calculation of arc

- gates[J]. *South-to-North Water Transfers and Water Science & Technology*, 2022, 20(3): 590-599. (in Chinese)) DOI: [10.13476/j.cnki.nsbdqk.2022.0060](https://doi.org/10.13476/j.cnki.nsbdqk.2022.0060).
- [6] 侍建国,张亦飞.拉依达准则在处理区域水文数据异常值中的应用[J].*海河水利*,2016(5):49-51.(SHI J G,ZHANG Y F.Application of Pauta criterion in handling regional hydrological data outliers[J].*Haihe Water Resources*,2016(5):49-51.(in Chinese)) DOI: [10.3969/j.issn.1004-7328.2016.05.017](https://doi.org/10.3969/j.issn.1004-7328.2016.05.017).
- [7] 陈琼妮.基于Hadoop的水情数据清洗方案研究[D].武汉:武汉理工大学,2016.(CHEN Q N.S Research of hydrologic data cleansing scheme based on Hadoop[D].Wuhan:Wuhan University of Technology,2016.(in Chinese))
- [8] CHOONG M K, CHARBI M, YAN H. Autoregressive-model-based missing value estimation for DNA microarray time series data[J]. *IEEE Transactions on Information Technology in Biomedicine*, 2009, 13(1): 131-137. DOI: [10.5353/th_b3936451](https://doi.org/10.5353/th_b3936451).
- [9] AMIRI M, JENSEN R. Missing data imputation using fuzzy-rough methods[J]. *Neurocomputing*, 2016, 205: 152-164. DOI: [10.4018/978-1-60566-336-4.ch005](https://doi.org/10.4018/978-1-60566-336-4.ch005).
- [10] 余泽禹.基于随机森林的大数据下数据缺失插补方法[D].荆州:长江大学,2018.(YU Z Y.Random forest missing data algorithms in big data[D].Jingzhou:Yangtze University,2018.(in Chinese))
- [11] 郝雨微.基于深度学习的医疗时序数据补值模型研究[D].长春:吉林大学,2019.(HAO Y W.Research on deep learning based imputation model for clinical time series[D].Changchun:Jilin University,2019.(in Chinese))
- [12] REN T, LIU X F, NIU J W, et al. Real-time water level prediction of cascaded channels based on multi-layer perception and recurrent neural network-ScienceDirect[J]. *Journal of Hydrology*, 2020, 585: 124783. DOI: [10.1016/j.jhydrol.2020.124783](https://doi.org/10.1016/j.jhydrol.2020.124783).
- [13] SHIGEYUKI O, MADA-AKI S, ICHIRO T, et al. A Bayesian missing value estimation method for gene expression profile data[J]. *Bioinformatics*, 2003, 19(16): 2088-2096. DOI: [10.1093/bioinformatics/btg287](https://doi.org/10.1093/bioinformatics/btg287).
- [14] KIM K Y, KIM B J, YI G S. Reuse of imputed data in microarray analysis increases imputation efficiency[J]. *Bmc Bioinformatics*, 2004, 5(1): 160. DOI: [10.1186/1471-2105-5-160](https://doi.org/10.1186/1471-2105-5-160).
- [15] 金容鑫,娄岱松,黄华德,等.水电机组状态监测数据清洗方法[J].*中国农村水利水电*,2022(7):187-192.(JIN R J,LOU D S,HUANG H D,et al.Research on cleaning method of condition monitoring data of hydropower units[J].*China Rural Water and Hydropower*,2022(7):187-192.(in Chinese))
- [16] 闫亚男,韩长霖,陈小松,等.融合时间序列特性的水电遥测数据清洗优化框架[J].*水力发电*,2021,47(11):79-83,95.(YAN Y N,HAN C L,CHEN X S,et al.Improved data cleaning framework of hydropower telemetry dataset based on time series characteristics[J].*Water Power*,2021,47(11):79-83,95.(in Chinese)) DOI: [10.3969/j.issn.0559-9342.2021.11.016](https://doi.org/10.3969/j.issn.0559-9342.2021.11.016).
- [17] 郑霞忠,陈国梁,邹韬.考虑时间序列关联的大坝监测异常数据清洗[J].*水力发电*,2020,46(4):111-114,125.(ZHENG X Z,CHEN G L,ZOU T.Cleaning of abnormal dam monitoring data considering time series association[J].*Water Power*,2020,46(4):111-114,125.(in Chinese)) DOI: [10.3969/j.issn.0559-9342.2020.04.023](https://doi.org/10.3969/j.issn.0559-9342.2020.04.023).
- [18] 王亚慧,黄鹏年,李致家,等.不同蒸发数据在降雨径流模拟中的适用性[J].*南水北调与水利科技(中英文)*,2020,18(3):48-56.(WANG Y H,HUANG P N,LI Z J,et al.Study on applicability of different evaporation data in rainfall runoff simulation[J].*South-to-North Water Transfers and Water Science & Technology*,2020,18(3):48-56.(in Chinese)) DOI: [10.13476/j.cnki.nsbdqk.2020.0049](https://doi.org/10.13476/j.cnki.nsbdqk.2020.0049).
- [19] 田景环,常思源,黄鑫,等.南水北调京石段应急供水输水损失规律分析[J].*南水北调与水利科技*,2015,13(2):395-398.(TIAN J H,CHANG S Y,HUANG X,et al.Analysis of water conveyance loss of Beijing-Shijiazhuang emergency water supply section in South-to-North Water Diversion Project[J].*South-to-North Water Transfers and Water Science & Technology*,2015,13(2):395-398.(in Chinese)) DOI: [10.13476/j.cnki.nsbdqk.2015.02.045](https://doi.org/10.13476/j.cnki.nsbdqk.2015.02.045).
- [20] 李景刚,乔雨,陈晓楠,等.南水北调中线干线节制闸过流公式率定及曲线绘制[J].*人民长江*,2019,50(8):224-227.(LI J G,QIAO Y,CHEN X N,et al.Flow-through equation calibration and curve draw-

- ing for Diaohu check gates in middle route Project of South-to-North Water Diversion[J]. *Yangtze River*, 2019, 50(8): 224-227. (in Chinese) DOI: [10.16232/j.cnki.1001-4179.2019.08.039](https://doi.org/10.16232/j.cnki.1001-4179.2019.08.039).
- [21] 刘孟凯, 王长德, 冯晓波. 长距离控制渠系结冰期的水力响应分析[J]. *农业工程学报*, 2011, 27(2): 20-27. (LIU M K, WANG C D, FENG X B. Analysis on the hydraulic response of long distance canal control system during ice period[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2011, 27(2): 20-27. (in Chinese) DOI: [CNKI:SUN:NYGU.0.2011-02-005](https://doi.org/CNKI:SUN:NYGU.0.2011-02-005)).
- [22] MU J, ZHANG X. Real-time flood forecasting method with 1-D unsteady flow model[J]. *Journal of Hydrodynamics*, 2007, 19(2): 150-154. DOI: [10.1016/S1001-6058\(07\)60041-9](https://doi.org/10.1016/S1001-6058(07)60041-9).
- [23] 顾峰峰, 倪汉根. 四点时空偏心隐格式的改进求解[J]. *大连理工大学学报*, 2007(3): 419-423. (GU F F, NI H G. Two improved calculation methods of Preissmann four-point linear implicit scheme[J]. *Journal of Dalian University of Technology*, 2007(3): 419-423. (in Chinese) DOI: [10.3321/j.issn:1000-8608.2007.03.022](https://doi.org/10.3321/j.issn:1000-8608.2007.03.022)).
- [24] 孙亚楠, 叶磊, 吴剑, 等. 基于一维水动力模型的洪水顶托影响分析[J]. *南水北调与水利科技*, 2018, 16(6): 194-201. (SUN Y N, YE L, WU J, et al. Analysis of backwater effects of flood based on one-dimensional hydrodynamic model[J]. *South-to-North Water Transfers and Water Science & Technology*, 2018, 16(6): 194-201. (in Chinese) DOI: [10.13476/j.cnki.nsbdqk.2018.0171](https://doi.org/10.13476/j.cnki.nsbdqk.2018.0171)).
- [25] 冯志勇, 李立群, 吴永妍, 等. 南水北调中线一期工程总干渠输水损失变化规律[J]. *南水北调与水利科技(中英文)*, 2022, 20(3): 600-609. (FENG Z Y, LI L Q, WU Y Y, et al. The variation of water loss in the main canal of middle route of South-to-North Water Transfer First Stage Project[J]. *South-to-North Water Transfers and Water Science & Technology*, 2022, 20(3): 600-609. (in Chinese) DOI: [10.13476/j.cnki.nsbdqk.2022.0061](https://doi.org/10.13476/j.cnki.nsbdqk.2022.0061)).
- [26] 邓山, 赵昕, 张莉, 等. 南水北调工程陶岔站时差法流量计推流技术研究[J]. *人民长江*, 2022, 53(4): 86-90. (DENG S, ZHAO X, ZHANG L, et al. Research on discharge calculation by ultrasonic transit time method flowmeter at Taocha Hydrological Station, canal head of South-to-North Water Diversion Project[J]. *Yangtze River*, 2022, 53(4): 86-90. (in Chinese) DOI: [10.16232/j.cnki.1001-4179.2022.04.014](https://doi.org/10.16232/j.cnki.1001-4179.2022.04.014)).

Application of inverted data cleaning for flow monitoring stations in the middle route of the South-to-North Water Transfer Project

WEI Wentao¹, JIN Yanguo², ZHANG Zhao³, LEI Xiaohui^{1,3}, XUE Ping⁴, WANG Yilin³

(1. School of water Resources and Electric Power, Hebei University of Engineering, Handan 056038, China; 2. China South to North Water Transfer Group Middle Line Co., LTD, Beijing 100038, China; 3. China Institute of Water Resources and Hydropower Research, Beijing 100038, China;

4. College of Computer and Information, Hohai University, Nanjing 211100, China)

Abstract: With the further development of the informatization level of long-distance open channel water transfer projects, the distribution of monitoring stations along the project has become denser and denser, and the monitoring of water regime information and the quality of monitoring data has attracted more attention. The reliability and consistency of the flow monitoring data along the long-distance open channel water transfer project in time and space directly affect the accuracy of the water condition study and hydrodynamic numerical simulation.

Taking the middle route project of the South-to-North Water Transfer Project as the research object, aiming at the abnormal phenomenon of the inverted flow of the monitoring station in the dispatching operation, the principle of dynamic balance of water volume and the longest sequence method of interval flow is adopted to construct the inverted data cleaning model of the flow monitoring station. The principle of water dynamic balance is used to calculate the water loss rate of the canal pool along the middle line project, and according to the water loss rate, the theoretical flow value of each control gate along the line is determined. By calculating the deviation of the actual

monitoring data of each control gate relative to the theoretical flow data, the relative deviation is selected to be less than 1% for interval division. The interval longest sequence method is used to screen out the data to be modified. The water balance principle is used to modify the modified data. By comparing and analyzing the measured flow data and the theoretical flow data of the control gate, the number of control times of the control gate and the number of control times that occurred in the inverted flow phenomenon are counted, and the causes of the inverted flow phenomenon are summarized. A one-dimensional hydrodynamic model was constructed to verify the impact of data cleaning on the simulation model by taking the flow data before and after cleaning as the upper boundary conditions.

A hydrodynamic model is constructed for verification based on the flow data after model cleaning as the upper boundary condition in the Baihe River to Huanghe River section of the middle line project. The results show that the cleaning model eliminates the inverted phenomenon in the flow monitoring data and improves the quality of the flow monitoring data. The mean absolute error of the water level in the front of the gate is reduced by 0.075 7 m and the root mean square error is reduced by 0.089 5 m in the hydrodynamic numerical simulation.

Through the analysis of the causes of the inverted data, it is found that the monitoring data deviation of the control gate flow monitoring equipment and the interference of the gate control are the main influencing factors of the phenomenon. By comparing the results of hydrodynamic simulation, it is found that the upper boundary condition of flow data after cleaning can effectively improve the simulation accuracy. The results show that the cleaned flow data is superior to the measured flow data in data space logicity and consistency, which can be popularized and applied in practical engineering.

Key words: middle route of South-to-North Water Transfer Project; data cleaning; traffic upside down; water balance; water loss; hydrodynamic model



积极稳妥推进碳达峰碳中和