

薛萍, 廖丽莎, 廖卫红, 等. 南水北调中线工程水位的水动力-神经网络耦合预测模型[J]. 南水北调与水利科技(中英文), 2023, 21(6): 1116-1125. XUE P, LIAO L S, LIAO W H, et al. Hydrodynamic-neural network coupling prediction model of water level for the Middle Route of South-to-North Water Transfers Project[J]. South-to-North Water Transfers and Water Science & Technology, 2023, 21(6): 1116-1125. (in Chinese)

# 南水北调中线工程水位的水动力-神经网络 耦合预测模型

薛萍<sup>1</sup>, 廖丽莎<sup>2</sup>, 廖卫红<sup>2</sup>, 位文涛<sup>3</sup>, 景象<sup>4</sup>

(1. 河海大学人工智能与自动化学院, 南京 210098; 2. 中国水利水电科学研究院水资源研究所, 北京 100038;  
3. 武汉虹信技术服务有限责任公司, 武汉 430043; 4. 天津大学建筑工程学院, 天津 300072)

**摘要:**南水北调中线工程通常以闸前常水位调度运行, 而水位在闸门调控影响下多数处于非平稳状态, 探索其变化规律对于监测数据和研究方法均有一定限制和要求。监测数据方面, 针对大量的高频监测数据选取均值滤波、滑动平均值滤波、递推中位值平均滤波法、滑动小波变换进行数据预处理, 提高数据质量、增强数据预测的可行性。研究方法层面, 以 BP 神经网络模型和长短期记忆(long short-term memory, LSTM)网络模型为主体框架, 以水动力模型的模拟数据为辅助支撑, 对比单神经网络在不同工况下的预测效果, 输出水动力-神经网络组合预测结果。结果表明: 数据预处理是数据分析和预测的必要环节, 高频数据滤波处理再预测可以提高数据预测的精度; 均值滤波、递推中位数均值滤波方式对数据预处理的效果最好, 指标合理时滤后决定系数( $R^2$ )精度均超过 0.95, 且均方根误差(RMSE)和平均绝对误差(MAE)不超过 0.02, 准确性高; 基于滤后数据进行模型构建, 通过对比数据驱动模型和数据-机理双重驱动模型的计算结果,  $R^2$  维持在 0.98 附近, RMSE、MAE 维持在 0.01 左右, 耦合模型具有更好的稳定性和准确性。

**关键词:**南水北调中线工程; 数据滤波; 神经网络; 水位预测

**中图分类号:** TV672 **文献标志码:** A **DOI:** 10.13476/j.cnki.nsbdqk.2023.0109

南水北调中线工程利用地形优势从丹江口水库北上 1 432 km 调水至北京团城湖<sup>[1]</sup>, 而长距离水量输送需要调度人员的丰富经验, 更需要良好的监测条件。南水北调中线工程依托由沿线近 300 座闸站的监测设备组成的监测站网, 在近 9 年的平稳运行下积累了海量的监测信息。其中, 水情、工情的历史监测序列运用到模型中, 对于调度方案的制定、预报预警和水力调控等均具有重要意义。限于闸前常水位的调度运行条件, 水位数据对中线工程影响较为直接, 且受闸门调控影响, 其变化趋势呈现出不确定性和非线性。目前已存在较多的水位预测方式, 其中人工智能算法在大量数据可用且计算成本降低的基础上, 以其性能上的独特优势, 得到

了推广和普及。对神经网络方法而言, 采用长短期记忆(long short-term memory, LSTM)网络模型及其耦合模型进行短期水位预测的应用较为广泛, 包括但不限于调水工程水位、水库水位<sup>[2]</sup>、湖泊水位<sup>[3]</sup>、地下水水位<sup>[4]</sup>等; 唐鸣等<sup>[5]</sup>采用 LSTM 网络对南水北调中线工程进行水位预测, 发现该网络可以高精度地完成预测情况; 王亦斌等<sup>[6]</sup>利用中值滤波进行数据预处理, 并将 EMD 分解和 LSTM 结合叠加预测, 实现水位、流量等的连续 12 h 的准确预测; 徐杨等<sup>[7]</sup>利用 LSTM 实现三峡水库上游水位的短期精准预测; 吕海峰等<sup>[8]</sup>构建了 splice-LSTM 的多因素水位预测模型研究西江干线, 提高了对线性成分的敏感性和水位峰值预测的准确性; 陈珺等<sup>[9]</sup>、汪松松等<sup>[10]</sup>

收稿日期: 2023-09-15 修回日期: 2023-10-30 网络出版时间: 2023-11-29

网络出版地址: <https://link.cnki.net/urlid/13.1430.TV.20231128.0916.004>

基金项目: 国家自然科学基金青年科学基金项目(52209046)

作者简介: 薛萍(1998—), 女, 山东潍坊人, 博士研究生, 主要从事智慧流域及水资源智能调控研究。E-mail: 2857487127@qq.com

通信作者: 廖卫红(1986—), 女, 湖北荆州人, 正高级工程师, 博士, 主要从事水文模型预报等研究。E-mail: liaowh@jwhr.com

利用卷积神经网络 (convolutional neural network, CNN)、循环神经网络 (recurrent neural network, RNN) 水位预测模型也可进行水位预测,但相比较而言, LSTM 既能解决 RNN 常面临的梯度消失和爆炸问题,又能灵活处理输入输出的维度,比 CNN 更具通用性。可构建 LSTM 模型进行水位预测,以期获得较好的预测效果。BP 神经网络易理解和实现,非线性映射能力强,利用 BP 神经网络及其耦合算法对泵站<sup>[11]</sup>、水电站<sup>[12]</sup>的水位进行预测研究,短期水位预测精度较高。

按建模方式进行分类,模型可分为数据驱动模型和机理模型,神经网络模型属于数据驱动模型。数据驱动模型依赖于大量的数据进行预测研究,但受限于过拟合和数据质量问题等约束条件;机理模型依赖于专业领域的理论知识并通过数学方程形式对结果进行展示,但受限于参数估计等诸多限制。将数据驱动模型和机理模型结合可以综合发挥两种模型的优势,适应复杂环境下的多重情景,提高预测精度。数据与机理双重驱动模式的优越性在特性分析<sup>[13-14]</sup>、水文预报<sup>[15]</sup>、系统建模<sup>[16]</sup>、预测控制<sup>[17-20]</sup> 等多方面均得到了体现,参照该种组合的优势条件,在研究 LSTM 神经网络、BP 神经网络这两种数据驱动模型在中线工程水位预测规律时,将机理模型-水动力模型与之相结合,以期获得更精准的水位预测效果,给调水工程的调度运行和水位预测规律研究提供依据。

## 1 数据预处理

神经网络模型属于数据驱动模型,该类模型依托于大数量且高质量的数据输入,故数据预处理是必要环节。针对时间序列种类的多样性,数据预处理的方法也各有不同,黄颖等<sup>[21]</sup>通过孤立森林和随机森林回归算法实现了对数据的预处理过程,位文涛等<sup>[22]</sup>通过水量动态平衡原理及区间流量最长序列法对流量数据进行数据预处理,提高流量数据质量;杨晓玲等<sup>[23]</sup>通过反  $k$  近邻树监测出异常值存在并处理,提高时间序列的质量;Long 等<sup>[24]</sup>通过色彩空间的转换和图像特征的监测进行数据清洗,风机数据中的异常数据可检测并清洗;吴月超<sup>[25]</sup>利用 IMDP 内清洗框架对异常数据进行监测和清洗;数据滤波可以去除噪声、平滑信号、提高数据质量,韩国锋<sup>[26]</sup>对时间序列数据滤波处理后预测效果较佳。测试滤波方法如下。

### 1.1 滤波方法

#### 1.1.1 均值滤波

均值滤波属于线性滤波算法,将当前像素点周围  $m \times m$  个像素点的均值取代原有像素点值,然后将该项操作遍历图像区间的每一个像素点。

#### 1.1.2 滑动平均值滤波

滑动平均值滤波,建立一个数据缓冲区,按照顺序存放  $N$  个数据,采取一个新数据时,最早收集的旧数据丢弃,求包括新数据在内的  $N$  个数据平均值。

#### 1.1.3 递推中位值平均滤波法

递推中位值平均滤波法是中位值滤波和递推滤波的结合,该方法把连续  $N$  个采样值看成一个队列,其长度固定不变,按照先进先出的顺序,去掉队列内的最大值和最小值后取  $(N-2)$  个数的平均值。

#### 1.1.4 滑动小波变换

小波变换通过变换突出特征、局部化分析时空频率、伸缩平移运算对函数逐步进行多尺度细化,可得高频时间细分、低频处频率细分处理结果,适应信号分析要求。滑动小波变换主要结合滑动窗口和小波变换的思想进行信号处理,即在小波变换的基础上添加滑动窗口使其更具应用性。

### 1.2 滤波结果

针对闸前水位、闸后水位以及流量数据的历史时间序列,选择不同的滤波方式进行测试,展示可视化图像,分析滤波效果。因数据频次较高,滤波前后对比不明显,故图内添加滤波前后差值的绝对值指标进行展示。

#### 1.2.1 秒级数据

采用秒级数据对水情数据进行滤波计算,以放水河 4 h 左右的闸前水位为例,4 种滤波之后的效果见图 1。分析图 1 可知:

整体而言,滤波具备平滑效果。均值滤波、滑动平均值滤波结果极为相似,差值大多聚集在 0.02 附近,即平滑效果偏小;超过 0.4 的数据量较多,即峰值处平滑效果更大。

递推中位值平均滤波在平滑效果处理上与前两种较为相近,但超过 0.4 的数据量较少,即峰值处平滑效果小于前两种滤波。

小波变换的差值接近 0.1 的较为密集,且偏大的差值极少,即滤波时对所有数据均进行了平滑处理,峰值有一定的平滑效果,但小于均值滤波。

4 h 内秒级水位波动范围 0.2 m 左右,接近平缓。相较于预测效果带来的差距,秒级数据的处理带来的不便更为突出,故预测不建议采用秒级数据处理。

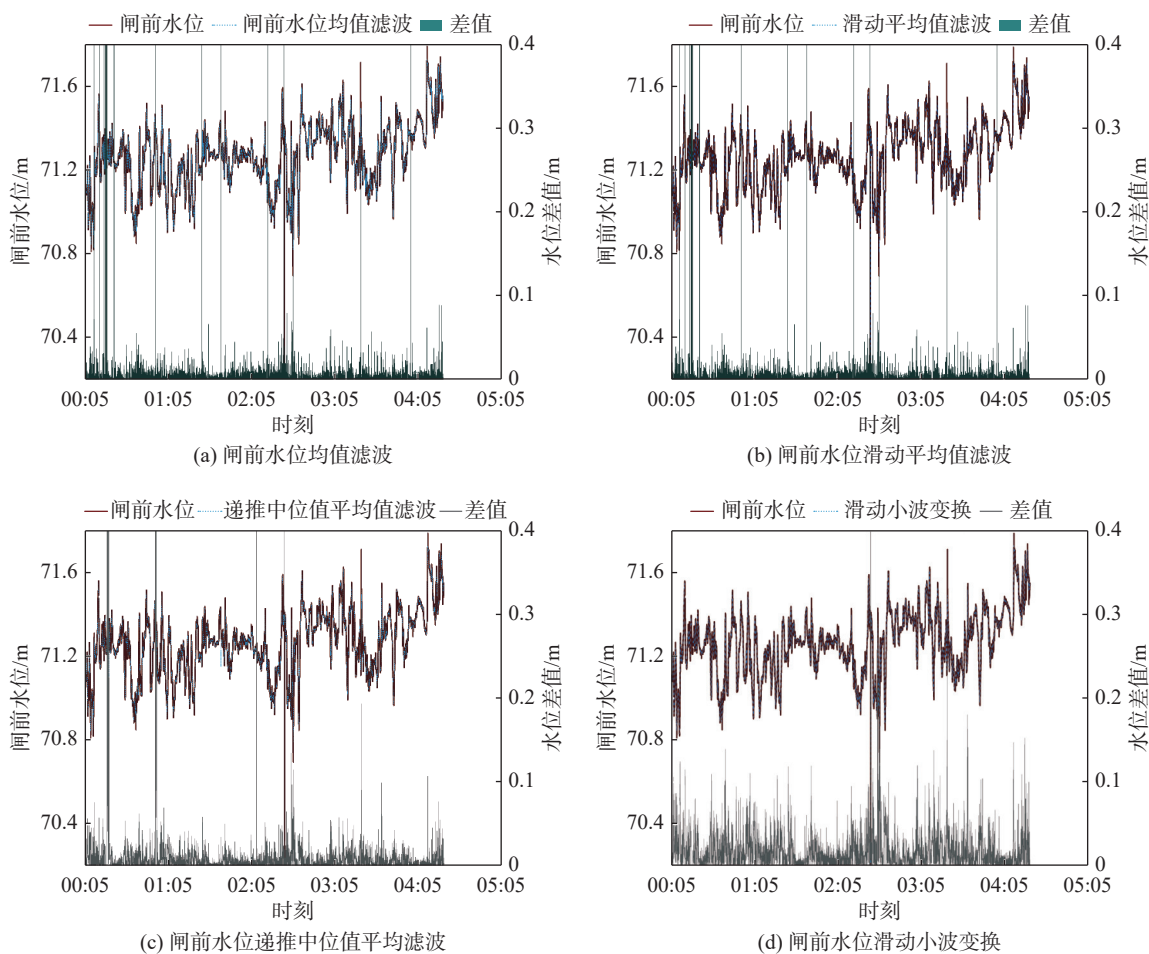


图 1 秒级数据滤波结果

Fig. 1 Second-level data filtering results

### 1.2.2 分钟级数据

采用分钟级数据对闸前水位、闸后水位、流量数据进行滤波计算,以放水河 2020 年 10 月数据为例,不同滤波之间的效果对比见图 2。

图 2 中绿色柱状图为差值,按列对比不同类型数据滤波效果,由差值起伏可知,滤波后流量差值振幅大于水位差值,闸后水位差值大于闸前水位差值;按行对比滤波种类的效果,由差值高度可知,均值滤波对振幅的平滑效果最大,滑动平均值滤波、滑动小波变换次之,递推中位值平均滤波平滑效果最小。

闸门是常见的水力调控建筑物之一,滤波后数据可清晰展示闸门前后的水位、流量变化规律,减少噪声影响,填补数据中的缺失值。考虑数据总量、变化程度及实际应用,训练数据最低频次选择 15 min。数据滤波对水位预测的影响最大程度体现在水位预测精度上,故本节仅通过可视化图像展示滤波前后效果。

## 2 研究方法

南水北调中线工程主要采用闸控方式进行调水,上下游闸门之间关系紧密。选取闸门本身的闸前、闸后水位、开度、流量数据及上游闸门的水位数据

作为模型输入因子,模型输出因子为闸门未来 2 h 的上游水位,滤波对水位预测的影响体现在预测精度中。

### 2.1 BP 神经网络模型

BP 神经网络一般为 3 层的前馈神经网络,通常采用误差反向传播算法进行训练。有效信息  $(x_1, x_2, \dots, x_n)$  通过输入层进行接收,并进一步传递到隐含层  $(h_1, h_2, \dots, h_m)$  进行信息处理,处理完成的数据通过输出层进行输出,其实质为映射功能的实现。BP 神经网络模型在泛化性、容错性以及非线性映射方面均具有较强特性,故在非线性序列的预测方面应用广泛,是一种典型的神经网络模型,1 个 3 层的典型网络结构见图 3。

### 2.2 LSTM 神经网络模型

LSTM 神经网络是一种经典的循环神经网络,网络模型通过记忆单元模块实现较长时间的记忆。因 LSTM 解决了 RNN 在训练中易存在的梯度消失和梯度爆炸问题,故被广泛应用于非线性、长时序的预测问题。整体网络结构包括 3 个门(遗忘门、输入门、输出门)和 1 个记忆单元,具体见图 4。

单元状态。图 4 方框上方的水平线称为单元状态,控制信息将上一细胞状态  $C_{t-1}$  传递给下一时刻  $C_t$ 。

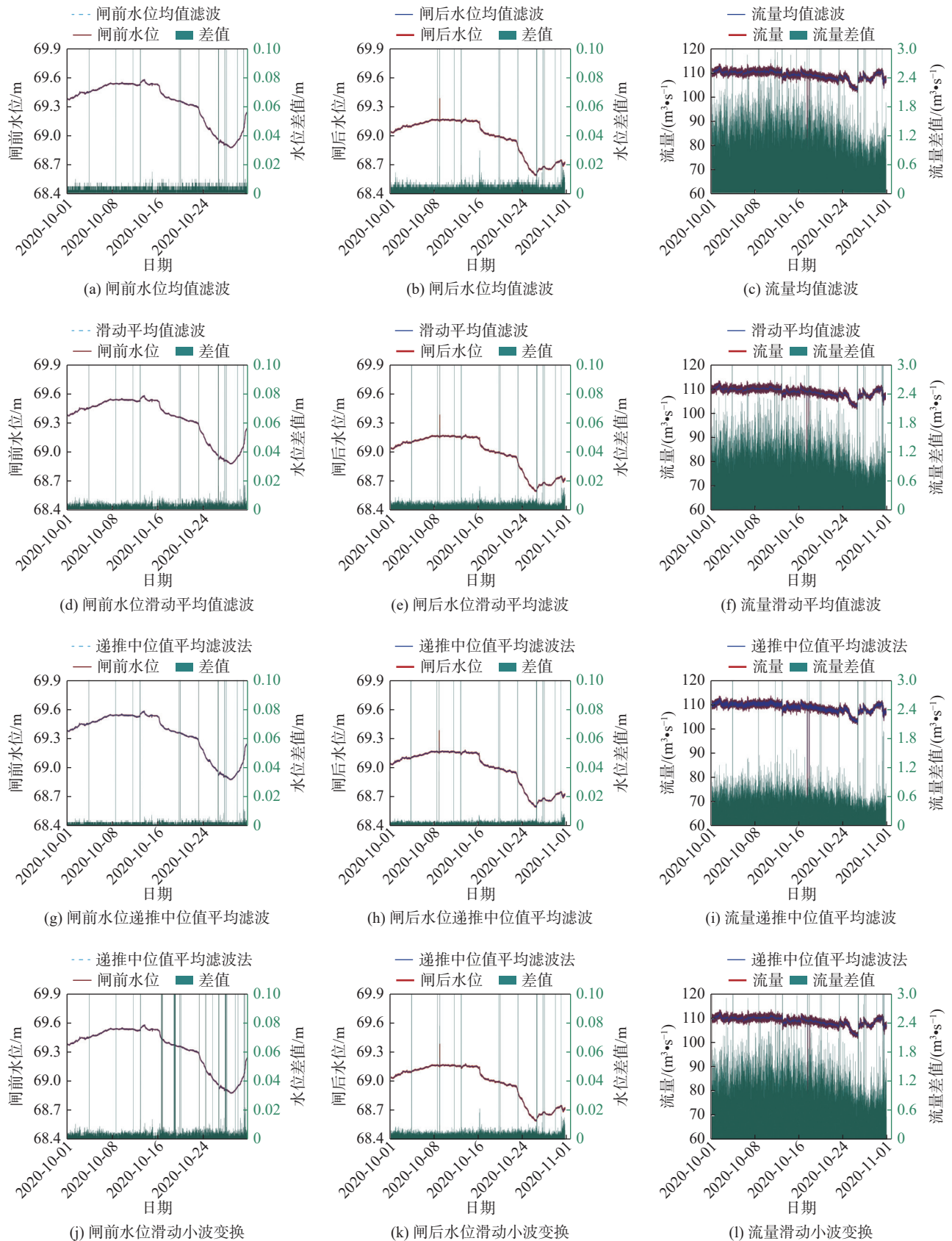


图2 分钟级数据滤波结果

Fig. 2 Minute-level data filtering results

遗忘门。输入本序列数据  $x_t$  上一系列的隐藏状态  $h_{t-1}$  和, 利用 Sigmoid 激活函数得到遗忘门的输出  $f_t$ ,  $f_t$  可决定  $C_{t-1}$  的遗忘程度:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

式中:  $W_f$  表示从遗忘门到输入门的权值矩阵;  $b_f$  表示遗忘门的偏置向量。

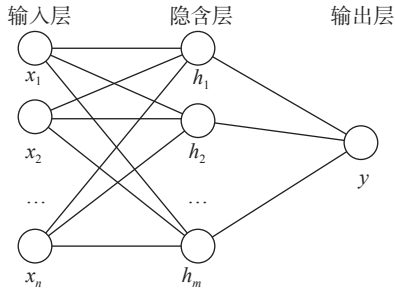


图 3 BP 神经网络模型结构

Fig. 3 BP neural network model structure

输入门。  $h_{t-1}$  和  $x_t$  经过 Sigmoid 激活函数和 tanh 层分别得到  $i_t$  和  $\tilde{C}_t$ ,  $i_t$  对更新到细胞状态的信息起决定作用,  $\tilde{C}_t$  为 tanh 层创建的新的候选值向量, 联合遗忘门处理过的上一时刻的细胞状态, 对细胞状态进行更新。

$$C_t = f_i C_{t-1} + i_t \tilde{C}_t \quad (3)$$

输出门。  $h_{t-1}$  和  $x_t$  经过 Sigmoid 层决定细胞状态的输出部分, 后细胞状态经过 tanh 层处理并与 sigmoid 门的输出相乘, 得到输出  $h_t$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t \tanh C_t \quad (5)$$

式中:  $W_o$  表示从遗忘门到输出门的权值矩阵;  $b_o$  表示输出门的偏置向量。

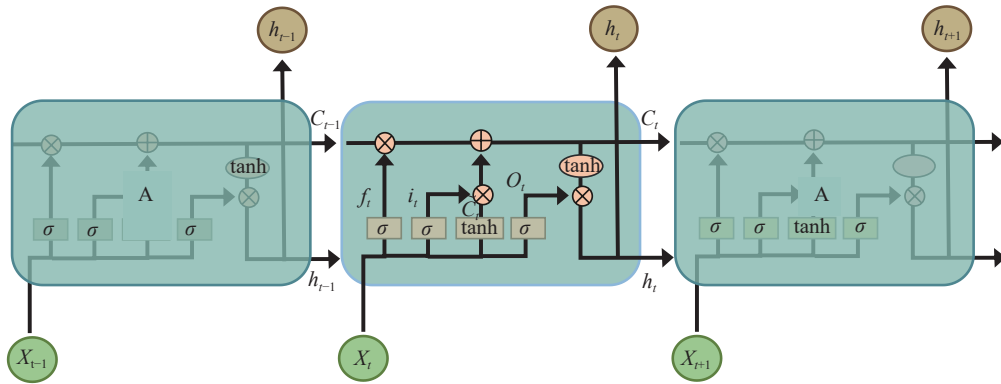


图 4 LSTM 神经网络模型结构

Fig. 4 LSTM neural network model structure

### 2.3 水动力-神经网络

南水北调中线工程渠道输水具有长距离的典型特征, 相较于宽度和深度而言, 其水流流态更具一维水动力特性。遵循质量守恒定律和动量守恒定律, 构建一维水动力模型, 采用的基本控制方程为一维非恒定流的圣维南方程组, 利用 Preissmann 四点带权隐式差分进行求解。

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

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

式中:  $B$  为明渠过水断面表面宽度, m;  $Z$  为水位, m;  $t$  为时间, s;  $Q$  为流量,  $m^3/s$ ;  $X$  为沿水流方向的渠道纵向距离, m;  $q$  为渠道旁侧入流,  $m^3/s$ ;  $\alpha$  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。

水动力-神经网络方法的本质为数据模型-机理模型双重驱动方法, 优势互补, 可产生良好的模型

效果。利用一维水动力模型生成部分训练数据, 结合历史调度数据, 构成神经网络训练数据集, 进行历史训练, 并检验预测效果。

## 3 结果与讨论

### 3.1 神经网络模型预测结果分析

#### 3.1.1 BP 神经网络模型预测结果

不同滤波参数对滤波结果有一定影响, 数据频次越高影响越小。输入数据为 2 h 频次, 选取开始时间节点为每天的调控时间节点, 考虑实际运行时按天调控, 以此为依据设定滤波  $n$  值, 即  $n \times \text{频次} \leq 24 \text{ h}$ 。输入数据包括预测闸门和上游闸门共 4 个, 闸门总数分别选取 2、3、4 闸门进行分析, 考虑 4 种滤波方式, 对比效果见图 5。

基于 2 h 频次的监测数据, BP 神经网络模型的预测结果已经处于较优水平, 仅图 5(j)、5(l) 的  $R^2$  远小于 0.95(表 1), 故此处不以更高频次的输入数据进行实验。由图示结果可得, 采用 BP 进行网络构建时, 采用 2、3 个闸门最佳, 建议滤波方式为均值滤波、递推中位值平均滤波。

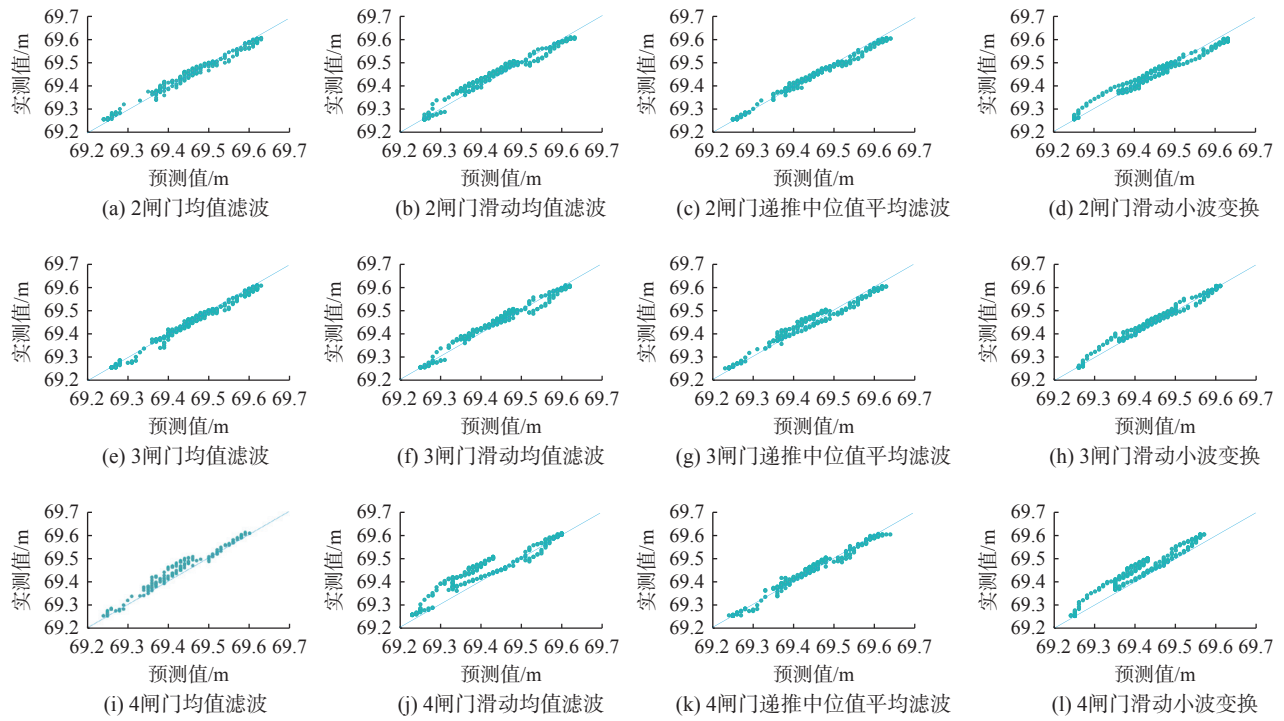


图5 120 min-BP神经网络预测结果

Fig. 5 Schematic diagram of the prediction results of the 120 min-BP neural network

表1 单网络模型误差分析结果

Tab. 1 Error analysis results of a single network model

模型	时间/min	闸门总数	指标	均值滤波	滑动均值滤波	递推中位值 平均滤波法	滑动小波变换
BP	120	2	$R^2$	0.970 9	0.960 7	0.973 8	0.950 5
			MAE/m	0.014 1	0.017 0	0.012 3	0.017 5
			RMSE/m	0.017 1	0.019 8	0.016 2	0.020 8
		3	$R^2$	0.974 4	0.967 1	0.954 8	0.975 5
			MAE/m	0.012 6	0.015 1	0.019 1	0.011 3
			RMSE/m	0.016 0	0.018 1	0.021 3	0.014 7
	4	$R^2$	0.925 0	0.767 5	0.975 4	0.779 0	
		MAE/m	0.022 4	0.037 9	0.012 8	0.037 9	
		RMSE/m	0.027 4	0.048 2	0.015 7	0.044 0	
	15	2	$R^2$	0.602 7	0.937 9	0.988 2	0.954 7
			MAE/m	0.024 9	0.009 2	0.003 6	0.007 5
			RMSE/m	0.026 7	0.010 6	0.004 6	0.009 0
3			$R^2$	0.808 5	0.549 7	0.738 3	0.578 7
			MAE/m	0.016 2	0.027 5	0.021 3	0.026 7
			RMSE/m	0.018 6	0.028 5	0.021 7	0.027 5
LSTM	4	$R^2$	0.925 9	0.962 1	0.789 9	0.745 9	
		MAE/m	0.009 5	0.007 1	0.017 1	0.020 6	
		RMSE/m	0.011 5	0.008 3	0.019 4	0.021 4	
		2	$R^2$	0.228 5	0.282 2	0.434 8	0.180 9
			MAE/m	0.040 6	0.033 3	0.030 5	0.042 2
			RMSE/m	0.041 4	0.0335	0.031 2	0.042 6
30	3	$R^2$	0.961 4	0.878 8	0.779 7	0.770 7	
		MAE/m	0.007 2	0.012 9	0.017 5	0.021 4	
		RMSE/m	0.009 3	0.013 8	0.019 5	0.022 6	

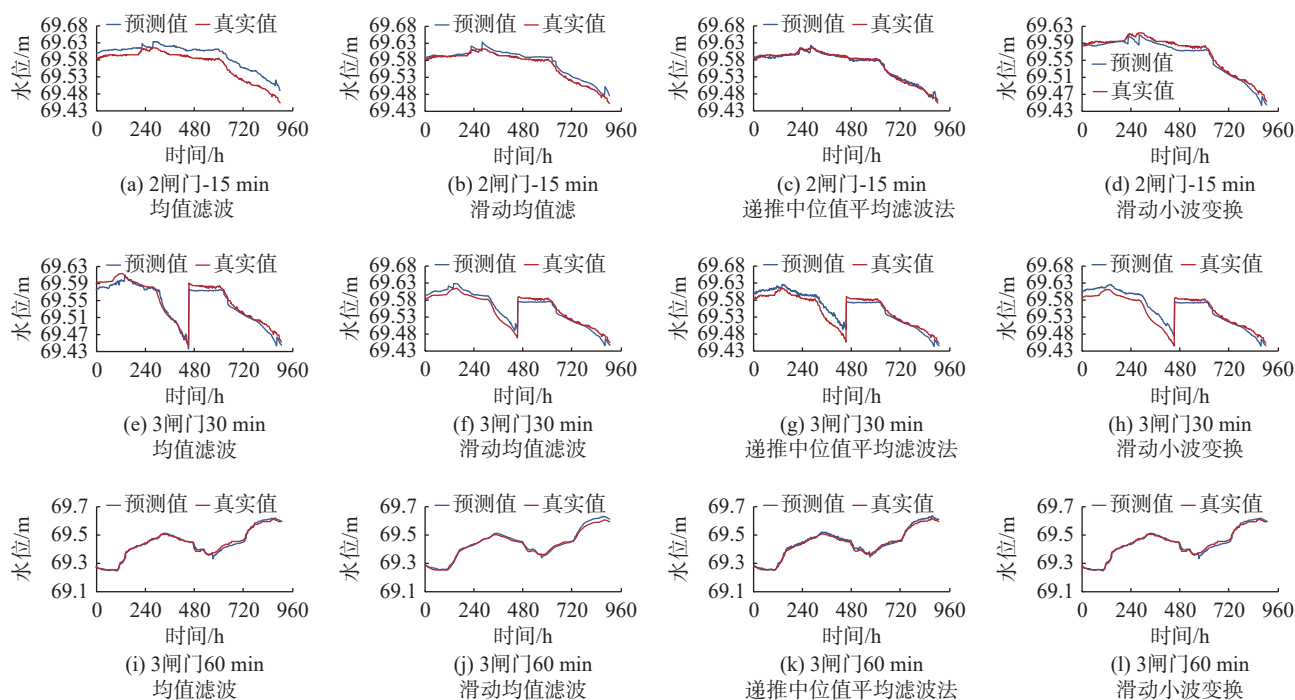
表 1(续)  
Tab. 1 (Continued)

模型	时间/min	闸门总数	指标	均值滤波	滑动均值滤波	递推中位值平均滤波法	滑动小波变换
LSTM	30	4	$R^2$	0.267 4	0.905 7	0.850 8	0.507 4
			MAE	0.037 9	0.008 7	0.010 9	0.031 3
			RMSE	0.040 3	0.012 2	0.016 0	0.033 1
		2	$R^2$	0.957 1	0.961 6	0.968 0	0.954 1
			MAE/m	0.015 6	0.014 7	0.013 3	0.015 7
			RMSE/m	0.020 6	0.019 7	0.017 9	0.020 7
	60	3	$R^2$	0.985 1	0.979 3	0.985 5	0.964 0
			MAE/m	0.010 4	0.011 8	0.010 3	0.015 4
			RMSE/m	0.012 1	0.014 5	0.012 0	0.018 4
		4	$R^2$	0.976 6	0.951 2	0.974 3	0.959 6
			MAE/m	0.012 4	0.017 5	0.011 9	0.015 5
			RMSE/m	0.015 2	0.022 2	0.016 0	0.019 5
	120	2	$R^2$	0.968 9	0.966 0	0.963 6	0.956 6
			MAE/m	0.013 7	0.015 2	0.015 2	0.015 7
			RMSE/m	0.017 6	0.018 5	0.019 1	0.019 5
		3	$R^2$	0.918 2	0.965 1	0.985 6	0.974 7
MAE/m			0.023 7	0.015 9	0.009 5	0.011 6	
RMSE/m			0.028 6	0.018 7	0.012 0	0.014 9	
	4	$R^2$	0.864 6	0.976 5	0.948 9	0.894 4	
		MAE/m	0.033 1	0.011 2	0.018 4	0.027 4	
		RMSE/m	0.036 8	0.015 3	0.022 6	0.030 4	

### 3.1.2 LSTM 神经网络模型预测结果

输入数据时间尺度包括 15、30、60、120 min, 其余工况划分同上。各时间尺度仅展示一行最优

结果, 效果见图 6。误差分析结果见表 1, 指标包括  $R^2$ 、RMSE、MAE。



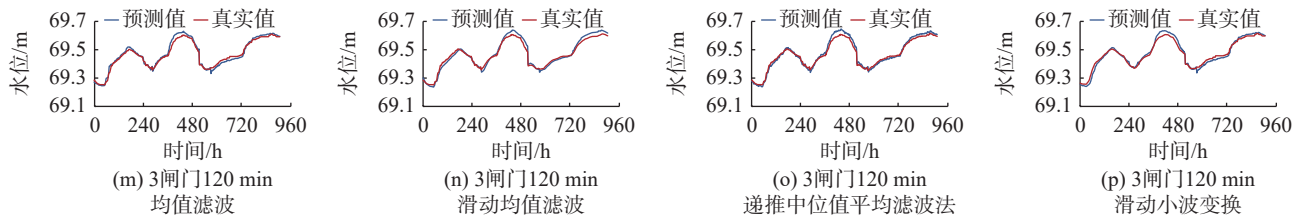


图6 LSTM神经网络预测结果

Fig. 6 Schematic diagram of the prediction results of the LSTM neural network

对比输入因子时间序列,整体效果而言,以60 min的预测结果最佳;对比闸门个数,分钟级频次下2、3个闸门情况均可,小时级数据下3个闸门数据为佳;对比滤波方式,以递推中位值平均滤波算法为佳,滑动小波变换效果最差;整组预测结果中以图6(k)、6(o)为最佳预测结果,即最优预测的输入选项为小时级输入、递推中位值平均滤波法和3个闸门输入。其中,LSTM神经网络在30 min频次、2个闸门时训练和预测时效果极差,实际过程中应尽量避免。

神经网络模型在调参后,选定时间内可较好地反映数据变化过程。但预测数据仅体现训练数据所能展现出来的规律,调参后的神经网络模型不适合任意场景下的数据,若此模型在工程实际中应用,需包含模型在不同范围内的调参情景。

### 3.2 水动力-神经网络预测结果

由上述可知,BP和LSTM模型预测效果均十分稳定,选取水动力模型和BP神经网络组合进行预测。构建包含预测闸门段的水动力模型,采取水动力模型模拟输出的30 min数据为穿插,1 h间隔为实测,选用3个闸门数据以及递推中位值滤波方式,误差分析结果见表2,预测效果见图7。

表2 水动力-BP神经网络模型误差分析结果

Tab. 2 Error analysis results of hydrodynamic-BP neural network model

min	闸门总数	$R^2$	RMSE/m	MAE/m
60	3	0.986 8	0.009	0.011

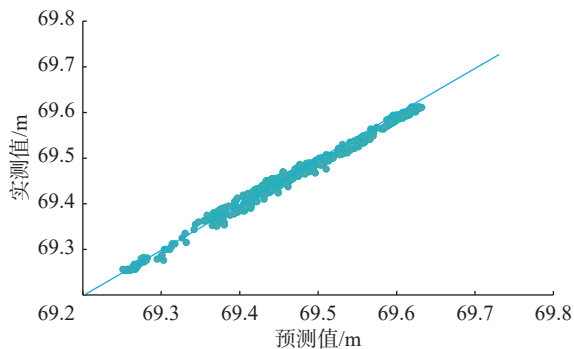


图7 水动力-BP神经网络预测结果

Fig. 7 Schematic diagram of the prediction results of hydrodynamic-BP neural network

与单神经网络相比,水动力-神经网络不仅预测效果较为稳定不容易出现偏差较大的情况,而且预测效果较好基本为单神经网络效果最佳情况。水动力-神经网络在神经网络模型的基础上提高了预测精度和稳定性,但其模型输入数据整合过程工作较为复杂,因此在工程中应用中建议采用单神经网络。

## 4 结论

数据滤波是一种减少噪声的合理处置方式,建议采用均值滤波、递推中位值均值滤波方式进行水情数据预处理。

基于模型调参和数据预处理,单神经网络模型预测在指标合理时滤后 $R^2$ 精度基本均超过0.95,且RMSE和MAE不超过0.02,准确性高;水动力-BP神经网络比BP神经网络稳定, $R^2$ 维持在0.98附近,RMSE和MAE在0.01左右。

增加训练的闸门总数其数据量增加,应选择合适的数量进行训练,频次较高时2、3个闸门的数据即可精准预测未来2 h尺度水位结果,添加距离较远的闸门数据甚至会影响数据预测精度。

### 参考文献:

- [1] 李红艳,付景保,褚钰,等.基于人水和谐的南水北调中线运行效果评价:以河南典型受水区为例[J].南水北调与水利科技(中英文),2022,20(1):93-101. DOI: 10.13476/j.cnki.nsbdkq.2022.0011.
- [2] 汤文强.基于LSTM模型的响洪甸水库旱警水位预测与效果检验[J].云南水力发电,2023(7):339-342. DOI: 10.3969/j.issn.1006-3951.2023.07.082.
- [3] 马飞,涂振宇,朱松挺,等.基于改进注意力机制的LSTM水位预测模型研究[J].江西水利科技,2023(3):162-166,175. DOI: 10.3969/j.issn.1004-4701.2023.03-02.
- [4] 袁志洪,陈雨.基于LSTM-TCN的地下水水位数据修复及应用[J].现代计算机,2023(8):20-26. DOI: 10.3969/j.issn.1007-1423.2023.08.003.



- [5] 唐鸣, 雷晓辉, 龙岩, 等. 基于长短时记忆网络(LSTM)的南水北调中线水位预测[J]. *中国农村水利水电*, 2020(10): 189-193. DOI: 10.3969/j.issn.1007-2284.2020.10.034.
- [6] 王亦斌, 孙涛, 梁雪春, 等. 基于EMD-LSTM模型的河流量水位预测[J]. *水利水电科技进展*, 2020, 40(6): 40-47. DOI: 10.3880/j.issn.1006-7647.2020.06.008.
- [7] 徐杨, 刘亚新, 汪涛, 等. 基于LSTM的三峡水库短期上游水位预测方法研究[J]. *水利水电快报*, 2022(10): 13-18. DOI: 10.15974/j.cnki.slsdkb.2022.10.003.
- [8] 吕海峰, 冀肖榆, 丁勇. 基于splice-LSTM的多因素西江水位预测模型研究[J]. *人民长江*, 2023, 54(7): 81-88. DOI: 10.16232/j.cnki.1001-4179.2023.07.012.
- [9] 陈琨, 黄燕华, 洪朋, 等. 基于机器学习模型的河道水位预测[J]. *水利水电科技进展*, 2023(3): 9-14. DOI: 10.3880/j.issn.1006-7647.2023.03.002.
- [10] 汪松松, 徐蓬博, 胡书源, 等. 基于深度学习的小流域山洪水位预测模型研究[J]. *电脑知识与技术*, 2022(20): 89-91. DOI: 10.14004/j.cnki.ckt.2022.1420.
- [11] 薛萍, 张召, 雷晓辉, 等. 基于BP神经网络的不同时间尺度泵站前池水位预测模型[J]. *南水北调与水利科技(中英文)*, 2022, 20(2): 393-407. DOI: 10.13476/j.cnki.nsbdqk.2022.0040.
- [12] 张钰彬, 练继建, 王孝群, 等. 基于PSO-水量平衡-BP耦合模型的短期水位预测[J]. *人民长江*, 2023, 54(3): 90-95. DOI: 10.16232/j.cnki.1001-4179.2023.03.014.
- [13] 郑翔, 王慧芳, 姜宽, 等. 机理与数据融合驱动的含IIDG配电网短路电流计算方法[J]. *电力自动化设备*, 2021, 41(1): 41-48. DOI: 10.16081/j.epae.202010018.
- [14] 王锁瑞. 机理-数据混合驱动的机床进给系统建模与特性分析[D]. 济南: 山东建筑大学, 2023. DOI: 10.27273/d.cnki.gsajc.2023.000197
- [15] 周茜. 机理和数据驱动的水文预报模型应用研究: 以汉江流域为例[D]. 邯郸: 河北工程大学, 2022. DOI: 10.27104/d.cnki.ghbjy.2022.000535
- [16] 王一茗, 马振华, 杨萌祺, 等. 机理和数据混合驱动的排水系统控制模型构建方法[J]. *环境工程*, 2022(6): 204-211, 225. DOI: 10.13205/j.hjgc.202206026.
- [17] 思迪. 机理与数据驱动的缸盖铣削表面质量预测与控制[D]. 沈阳: 沈阳工业大学, 2022. DOI: 10.27322/d.cnki.gsgyu.2022.000526
- [18] 龙梓楠. 基于机理数据双驱动模型的丹江口水库入库日径流预测研究[D]. 武汉: 华中科技大学, 2022. DOI: 10.27157/d.cnki.ghzku.2022.003252
- [19] 王彪, 吕洋, 陈中, 等. 考虑信息时移的分布式光伏机理-数据混合驱动短期功率预测[J]. *电力系统自动化*, 2022(11): 67-74. DOI: 10.7500/AEPS20211020004.
- [20] GHAITH M, SIAM A, LI Z, et al. Hybrid hydrological data-driven approach for daily streamflow forecasting[J]. *Journal of Hydrological Engineering*, 2020, 25(2): 0419063. DOI: 10.1061/(ASCE)HE.1943-5584.0001866.
- [21] 黄颖, 许剑, 周子祺, 等. 高效长序列水位预测模型的研究与实现[J]. *电子科技大学学报*, 2023, 52(4): 595-601. DOI: 10.12178/1001-0548.2022133.
- [22] 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. DOI: 10.13476/j.cnki.nsbdqk.2022.0114.
- [23] 杨晓玲, 冯山, 袁钟. 基于相对距离的反k近邻树离群点检测[J]. *电子学报*, 2020, 48(5): 937-945. DOI: 10.3969/j.issn.0372-2112.2020.05.014.
- [24] LONG H, XU S H, GU W. An abnormal wind turbine data cleaning algorithm based on color space conversion and image feature detection[J]. *Applied Energy*, 2022, 311(4): 118594. DOI: 10.1016/j.apenergy.2022.118594.
- [25] 吴月超. 面向智能水电站的监测数据集成及运行分析研究与实践[D]. 武汉: 华中科技大学, 2018.
- [26] 韩国锋. 基坑变形在数据滤波处理条件下的组合预测研究[J]. *岩土工程技术*, 2023, 37(4): 392-396. DOI: 10.3969/j.issn.1007-2993.2023.04.003.

## Hydrodynamic-neural network coupling prediction model of water level for the Middle Route of South-to-North Water Transfers Project

XUE Ping<sup>1</sup>, LIAO Lisha<sup>2</sup>, LIAO Weihong<sup>2</sup>, WEI Wentao<sup>3</sup>, JING Xiang<sup>4</sup>

(1. School of Artificial Intelligence and Automation, Hohai University, Nanjing 210098, China; 2. Department of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China; 3. Wuhan Hongxin Technical Service Co. Ltd., Wuhan 430043, China; 4. School of Civil Engineering, Tianjin University, Tianjin 300072, China)

**Abstract:** The Middle Route of South-to-North Water Transfers Project has made a significant contribution to

mitigating the water scarcity challenges prevalent in the central and northern regions of China, and in the process of the project scheduling and operation, it usually operates in accordance with the normal water level before the gate. Under the influence of gate regulation, the water level before and after the gate is in a non-stationary process most of the time, so exploring its regular changes has certain limitations and requirements on the monitoring data and research methods. To enhance the precision of water level forecasting within the ambit of the South-to-North Water Transfers Project, the monitoring data and research methodology are improved respectively, with a view to obtaining better prediction results.

For a large amount of high-frequency monitoring data, mean filter, sliding mean filter, recursive median mean filter, and sliding wavelet transform are selected for data preprocessing to improve the data quality and enhance the feasibility of data prediction. The data prediction framework leverages two primary neural network models, namely the BP neural network model and LSTM neural network model, and the hydrodynamic model simulation data as the auxiliary support, and selects the pre-gate, and post-gate water level, openness, and flow rate data of the upstream gate itself as the model input factor for prediction. The predictive output factor pertains to the upstream water level of the gate within the subsequent 2 h. The assessment of predictive performance is predicated upon key indicators, namely the coefficient of determination, root-mean-square error, and average absolute error. Indicators compare the single neural network prediction results and network-hydrodynamic combination prediction results and analyze the accuracy and stability of the prediction results.

The accuracy of data prediction can be improved after data filtering preprocessing of high-frequency data, and the selection of appropriate water-level data filtering methods can significantly improve the prediction effect. After filtering the data can more clearly show the water level before and after the gate, flow change rule, and the minimum frequency of training data can be selected for 15 minutes of data for filtering and data processing. Constructing BP and LSTM neural networks based on the monitoring data, a comparative analysis is conducted encompassing the number of gate inputs, temporal scales, and data filtering methodologies. The investigation reveals the following insights: The prediction results under the hourly data can already reach the optimal state; The number of gates can be 2 or 3 gates to reach the optimal state, which is related to the frequency of the data; Comparing the prediction results after a variety of filtered data, the recursive median-mean filtering algorithm is the best, and the mean filtering is the worst. The sliding wavelet transform has the worst effect, so it is suggested that the filtering methods are mean filtering and recursive median-mean filtering. The combined hydrodynamic-neural network prediction results are better than the single network prediction results.

The computational outcomes prove that high-frequency data preprocessing is a necessary part of data analysis, and the suggested filtering methods mean filter and recursive median-mean filter can be applied in the water condition data processing of the water transfer project. These two filtering methods yield commendable outcomes in data processing. The neural network model necessitates tailoring to the specific parameters corresponding to distinct objects and varying temporal cycles, and after adjusting the parameters, it can better reflect the data change process in a period of time, and the prediction effect is better. Different neural network models have different prediction characteristics, but the prediction accuracy is higher under the condition of sufficient high correlation data. Moreover, the data mechanism dual-drive model can play the advantages of the hydrodynamic model and neural network model at the same time, and the prediction accuracy is higher.

**Key words:** Middle Route of South-to-North Water Transfers Project; data filtering; neural network; water level prediction