

高怡然, 卢龙彬, 桑国庆, 等. 基于输入因子影响分析的 LSTM 梯级泵站前水位实时预测模型[J]. 南水北调与水利科技(中英文), 2025, 23(1): 157-165. GAO Y R, LU L B, SANG G Q, et al. LSTM model for the real-time prediction of pre-station water level based on input factor analysis[J]. South-to-North Water Transfers and Water Science & Technology, 2025, 23(1): 157-165. (in Chinese)

基于输入因子影响分析的 LSTM 梯级泵站 站前水位实时预测模型

高怡然, 卢龙彬, 桑国庆, 王维林, 刘鲁霞

(济南大学水利与环境学院, 济南 250022)

摘要: 相比于传统的渠道水动力学方程求解站前水位, 机器学习模型因其响应快、稳态误差小等特点, 被广泛应用于渠道水位预测。然而, 输入数据庞杂、时序预测困难等原因直接限制和影响了预测精度, 如何准确评价输入数据对预测结果的影响, 是水位预测模型应用亟待解决的难题之一。以南水北调东线两湖段典型工程为研究对象, 提出一种测试输入数据对预测结果影响分析的框架, 构建基于长短时记忆网络(long short-term memory, LSTM)模型的泵站站前水位实时预测模型, 通过对各输入因子的敏感性分析, 实现预测模型数据输入集的优化, 并基于该数据组合对不同预见期下的邓楼泵站站前水位进行预测分析。结果表明: 以长沟站前水位、长沟泵站流量、邓楼站前水位指标作为输入参数的组合最优, 平均绝对误差(mean absolute error, E_{MA})、均方误差(mean squared error, E_{MS})、均方根误差(root mean square error, E_{RMS})和平均绝对百分比误差(mean absolute percentage error, E_{MAP})⁴项指标均小于 0.05, 相关系数(r)和决定系数(R^2)均大于 0.95, 剔除突变数据后最大水位误差不超过 0.02 m。

关键词: LSTM 神经网络; 站前水位; 水位预测; 敏感性分析; 时间序列

中图分类号: TV675 **文献标志码:** A **DOI:** 10.13476/j.cnki.nsbdkq.2025.0016

邓楼泵站作为南四湖—东平湖段的关键枢纽之一, 其站前水位的精确预测是保障整个输水工程安全运行、实现南水北调东线工程通水目标的关键前提^[1-2], 也是调度人员进行泵站调控和水量调度科学决策的重要依据。传统预测方法通常以一、二维水动力学模型为基础, 其优点在于模型参数具有一定实际物理意义, 但模型精确度的保证需要大量合理分布的数据和适当的分析方法^[3]。以神经网络为代表的机器学习模型通过采集海量数据形成决策模型, 有助于精确预测站前水位, 提高泵站运行效率, 实现智能调度, 对水网建设具有重要意义^[4]。

许多学者尝试通过构建神经网络模型进行水位预测, 并取得一定的成果, 其中: 陈奂良等^[5]基于 GA-BP(genetic algorithm-back propagation)神经网络

调蓄水位预测模型精确预测岩溶泉动态水位变化; 刘晓伟等^[6]提出基于 GRA-NARX(grey relation analysis-nonlinear auto-regressive model with exogenous inputs)神经网络的综合预测模型, 在密云水库水位预测中表现出高精度和广泛适用性; 薛萍等^[7]通过机理与数据混合驱动的水位预测方法成功提高闸前水位预测的精度; 杨婷婷等^[8]提出基于 Embedding-GRU(embedding-gated recurrent unit)的预测模型, 在水库水位预测中展现出较高的预测精度和泛化能力。以上研究探讨了不同神经网络水位预测方法, 但如何实现实时水位预测仍是亟须攻克的难题。

长短时记忆网络(long short-term memory, LSTM)模型作为一种特殊的循环神经网络结构, 既有神经网络的非线性特性, 又有记忆性和长期依赖

收稿日期: 2024-07-03 修回日期: 2024-11-14 网络出版时间: 2024-11-18

网络出版地址: <https://link.cnki.net/urlid/13.1430.tv.20241118.1158.003>

基金项目: 国家自然科学基金青年项目(42301046); 山东省自然科学基金项目(ZR2020ME249); 山东省自然科学基金青年项目(ZR2020QE282)

作者简介: 高怡然(2001—), 女, 山东济南人, 主要从事水文学及水力控制研究。E-mail: 1427402591@qq.com

通信作者: 卢龙彬(1989—), 男, 山东日照人, 讲师, 博士, 主要从事梯级泵站调水工程水力仿真及联合优化调度研究。E-mail: llb198804@126.com

性,相较于其他类似算法,在处理较长序列数据、解决梯度消失问题方面具有独特优势。引入实时神经网络预测模型可以很好地解决传统预测方法对外部条件影响的忽略问题。张睿芝等^[9]利用 LSTM 神经网络对洞庭湖不同湖区的水位进行预测,证明 LSTM 模型的预测精度较高且预测效果较优;Hrnjica 等^[10]采用 LSTM 模型预测了位于克罗地亚的 Vrana 湖的逐月水位,证实了 LSTM 模型结果验证的连续性;胡昊等^[11]设计了基于 DRSN-LSTM 神经网络的站点周边水位预测系统,分析站点下游不同时间尺度的水位变化,验证了 LSTM 的时序性。由上述研究可知,LSTM 在预测精度、时序性上均表现良好,但如何解决水位预测输入因子繁冗问题仍未有明确答案。

本文从探寻 LSTM 神经网络模型的实时预测能力和寻求输入因子最优组合出发,建立基于 LSTM 的实时泵站前水位预测模型,并使用模型对不同预见期下邓楼泵站前水位进行模拟预测^[12],以期为梯级泵站前水位实时预测提供新的方法和思路。

1 研究方法

本研究提出一种测试输入数据对预测结果影响分析框架,筛选最优输入数据组合^[13],在此基础上构建基于 LSTM 模型的站前水位实时预测模型,对不同预见期的邓楼泵站前水位进行实时预测,继而通过 one-at-a-time(OAT)法对各输入因子进行敏感性分析,优化预测前的数据输入集合,同时采用不同评价指标对模型结果的优劣进行评判^[14]。

1.1 LSTM 神经网络

LSTM 模型本质上是一种特殊设计的循环神经网络(recurrent neural network, RNN),旨在解决传统 RNN 在处理长序列数据时遇到的梯度消失和短期记忆问题,引入门控机制,使得 LSTM 能够更有效地捕获和保留长距离的时序依赖关系^[15-16]。

在 LSTM 中,有 3 个关键的门控单元:输入门、遗忘门和输出门。这些门控单元以及记忆单元共同构成了 LSTM 的核心结构,协同工作来控制信息的流动^[17]。LSTM 神经网络模型结构见图 1。

设隐藏单元数为 h ,在 t 时刻输入, $X_t \in R^{n \times d}$ (其中 n 为样本数, d 为输入长度),前一时刻的隐藏状态 $H_{t-1} \in R^{n \times h}$ (可视为短期状态),则在 t 时刻输入门 $I_t \in R^{n \times h}$,遗忘门 $F_t \in R^{n \times h}$ 和输出门 $O_t \in R^{n \times h}$ 表达式为

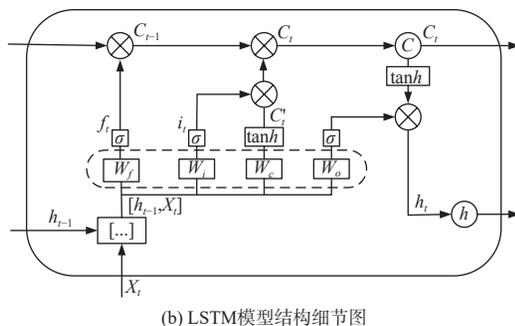
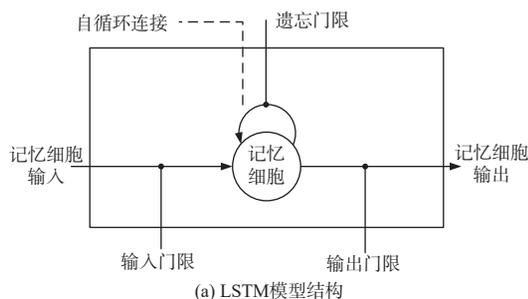


图 1 LSTM 神经网络模型结构
Fig. 1 LSTM neural network model structure

$$I_t = \sigma(X_t W_{XI} + H_{t-1} W_{HI} + b_I) \quad (1)$$

$$F_t = \sigma(X_t W_{XF} + H_{t-1} W_{HF} + b_F) \quad (2)$$

$$O_t = \sigma(X_t W_{XO} + H_{t-1} W_{HO} + b_O) \quad (3)$$

式中: I_t 、 F_t 、 O_t 都由 Sigmoid 激活函数控制,它们的输出值都在 0~1。记忆细胞 G_t (可视为长期状态) 可由式(5)计算:

$$G_t = \tanh(X_t W_{XG} + H_{t-1} W_{HG} + b_G) \quad (4)$$

$$c_t = F_t \otimes c_{t-1} + I_t \otimes G_t \quad (5)$$

式中: W_{XI} 、 W_{XF} 、 W_{XO} 、 $W_{XG} \in R^{d \times h}$ 为 4 个全连接层关于它们的输入向量的权重矩阵; W_{HI} 、 W_{HF} 、 W_{HO} 、 $W_{HG} \in R^{n \times h}$ 为 4 个全连接层关于它们的短期状态 H_{t-1} 的权重矩阵; b_I 、 b_F 、 b_O 、 $b_G \in R^{1 \times h}$ 为 4 个全连接层的线性偏倚; $\sigma(\cdot)$ 和 $\tanh(\cdot)$ 为 Sigmoid 函数和双曲正切函数; \otimes 表示向量间的逐元素相乘。隐藏状态和输出的表达式为

$$Y_t = H_t = O_t \otimes \tanh(c_t) \quad (6)$$

1.2 评价指标的选取

以平均绝对误差(mean absolute error, E_{MA})、均方误差(mean squared error, E_{MS})、均方根误差(root mean square error, E_{RMS})、平均绝对百分比误差(mean absolute percentage error, E_{MAP})、相关系数(r)和决定系数(R^2)为指标进行模型性能评估^[18]。相关指标计算公式如下:

$$E_{MA} = \frac{1}{m} \sum_{i=1}^m |f_i - y_i| \quad (7)$$

$$E_{MS} = \frac{1}{m} \sum_{i=1}^m (f_i - y_i)^2 \quad (8)$$

$$E_{RMS} = \sqrt{\frac{1}{m} \sum_{i=1}^m (f_i - y_i)^2} \quad (9)$$

$$E_{MAP} = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - f_i}{y_i} \right| \quad (10)$$

$$r = \frac{\sum_{i=1}^m (y_i - \bar{y})(f_i - \bar{f}_i)}{\sqrt{\sum_{i=1}^m (y_i - \bar{y})^2 \sum_{i=1}^m (f_i - \bar{f}_i)^2}} \quad (11)$$

$$R^2 = \frac{\sum_{i=1}^m (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (12)$$

式中: f_i 为预测值; y_i 为实测值; m 为样本数; \bar{y} 为测试开始之前的平均值。其中: E_{MA} 、 E_{MS} 、 E_{RMS} 和 E_{MAP} 取值范围在 0~1, 值越接近于 0, 说明模型拟合越好, 模型预测准确率越高; r 与 R^2 的取值范围在 0~1, 越接近于 1, 说明模型的实测值与模拟值线性相关度越好。

2 水位预测实例

以南水北调东线工程核心调蓄节点邓楼泵站为例, 对其站前水位进行预测研究^[19-20]。围绕邓楼泵站建立站前水位实时预测模型, 实现对该泵站水位变化的精准把握, 从而为工程调度提供科学依据, 保障输水过程的平稳进行^[21]。研究区域及渠段剖面见图 2。

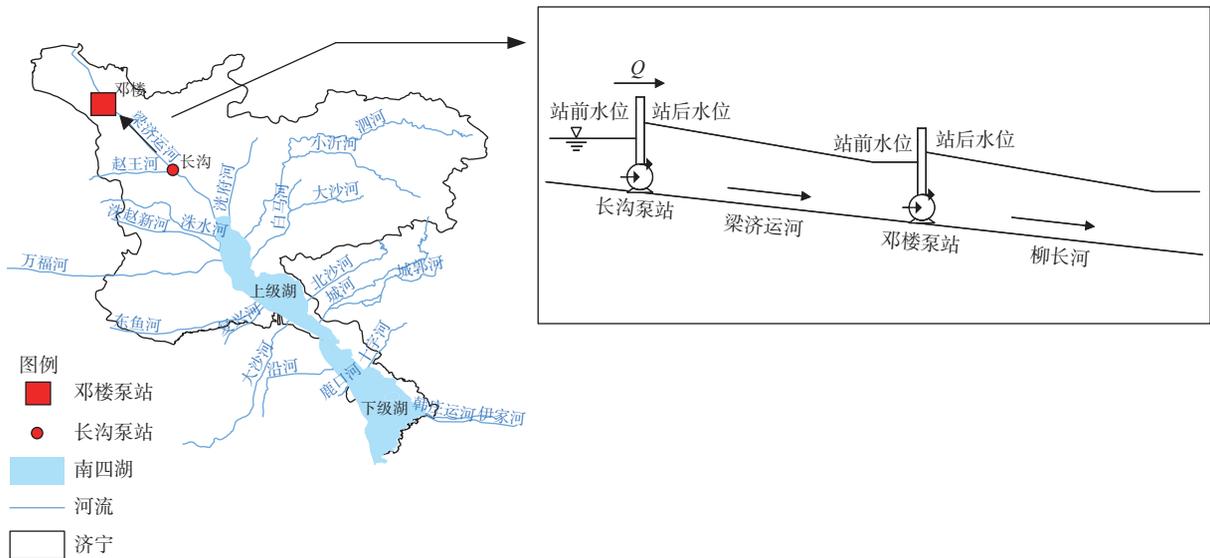


图 2 研究区域及渠段剖面

Fig. 2 Study area and drainage section profiles

2.1 数据来源与处理

选取山东省南水北调 2013—2023 年的实测泵站水位数据, 并采用归一法^[22]进行数据预处理, 然后通过分析不同影响因子组合对站前水位实时预测模型精度的影响, 寻求不同预见期下的站前水位变化规律^[23]。处理方式如下:

$$N_{ew}(x) = \left[\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right] \cdot (D - C) + C \quad (13)$$

式中: $N_{ew}(x)$ 为归一化后的值, x 为归一化前的值; D 为归一化合适范围的最大值; C 为归一化合适范围的最小值。

将处理后 2 381 组数据中的 70% 作为训练集, 30% 作为测试集^[24], 在进行传播时间计算验证后, 对

邓楼泵站未来 2、4 和 6 h 水位的预测精度进行探讨和分析, 寻求不同预见期对邓楼泵站站前水位预测精度的影响。

2.2 模型构建

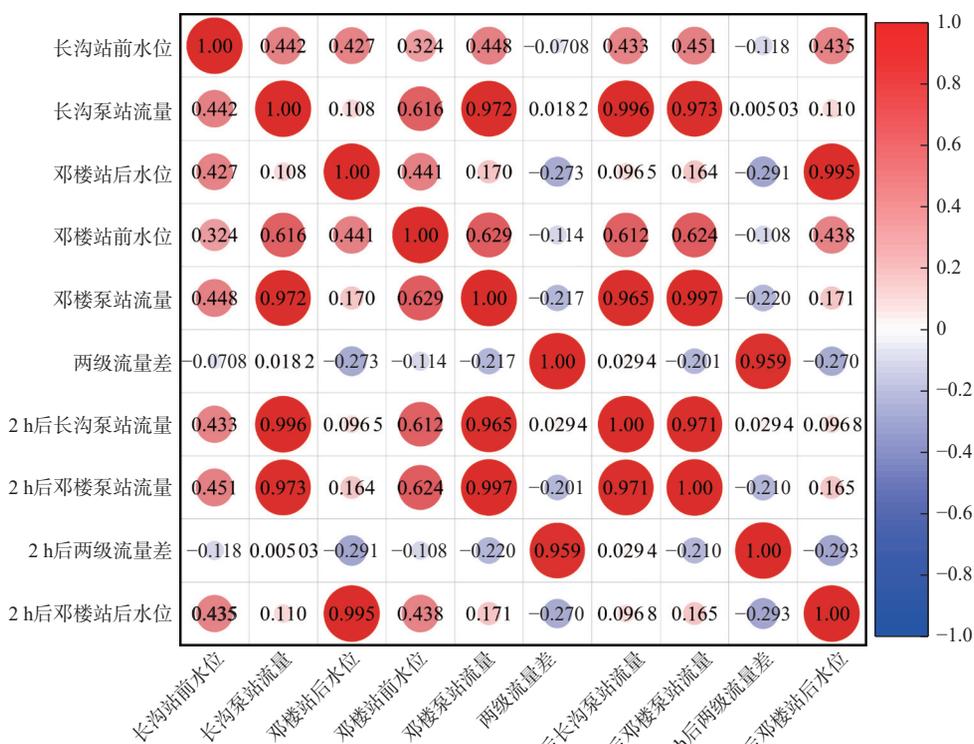
搭建的 LSTM 神经网络实时预测模型参数设置如下: 隐藏层大小为 64, 批处理值大小为 500, 列长度为 1, 损失函数为 E_{MS} , 步长根据预见期 (2、4 和 6 h) 大小设置, 进行 500 次迭代, 并对原始数据进行归一化处理。将站前水位、泵站流量等实测数据作为模型输入, 通过调节优化权重矩阵和偏差项使其达到期望预测精度, 即完成模型的训练和优化^[25-26]; 基于优化后的模型, 输入预测集数据, 最终实现对站前水位的预测。

3 结果分析

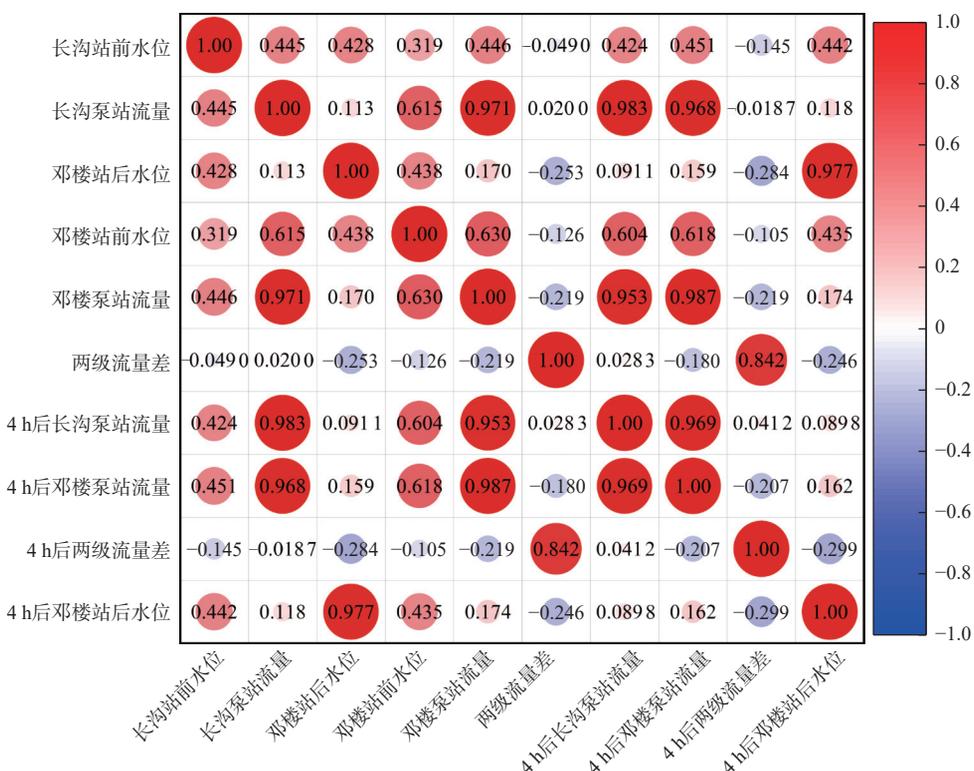
3.1 输入变量影响分析

在泵站-泵站的 LSTM 神经网络实时水位预测模型中有 10 个参数能够用于独立性分析, 其中每个

参数都在一定程度上参与了水力要素分析的过程。但基于本文研究目标, 使用相关系数图对输入参数进行相关性分析, 见图 3, 根据系数图中不同方块颜色对应的相关系数大小, 可以判断出变量之间相关性的大小。



(a) 2 h 相关系数图



(b) 4 h 相关系数图

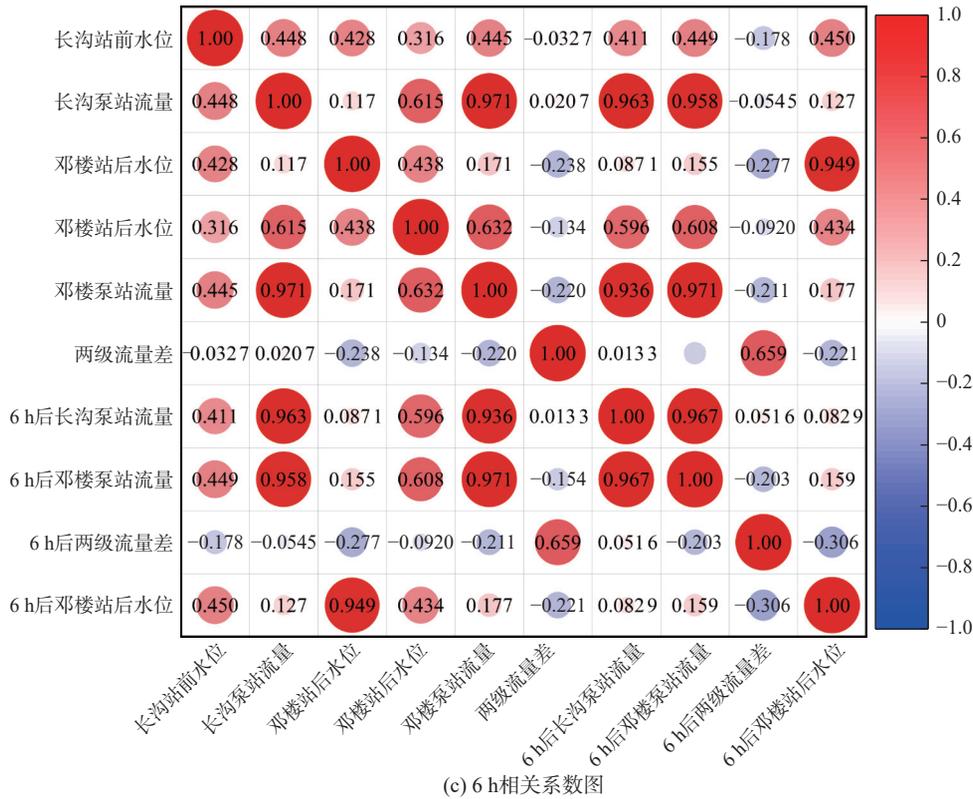


图3 各预见期相关系数

Fig. 3 Plot of correlation coefficients for each forecast period

2个变量之间相关系数的计算公式为

$$\rho_{XY} = \frac{C_{ov(X,Y)}}{\sigma_X \sigma_Y} = \frac{E_{XY} - E_X * E_Y}{\sigma_X \sigma_Y} \quad (14)$$

式中： ρ 表示相关系数； C_{ov} 表示协方差； σ 表示标准差； E 表示数学期望/均值。

初步选择变量间线性相关程度高的长沟站前水位、长沟泵站流量、邓楼站前水位、邓楼泵站流量4个参数作为模型输入；通过输入变量的增加或减少模拟不同输入变量对模型精度的影响，重点考虑

敏感性系数较大的变量，以此降低模型的复杂度，减少数据分析处理的工作量，以期在一定程度上提高模型的精度。研究人员可利用各因子敏感性系数的排序结果解决相应的调度决策问题。由于篇幅有限，仅以预见期2h为例进行输入变量影响分析，4和6h分析过程与2h相同。

在输入变量为 t 时刻，输出变量为 t +预见期(2h)工况下，对实测数据和预测数据进行拟合，并计算其评价指标。输入组合和计算结果见表1。

表1 不同输入组合下预测模型评价指标

Tab. 1 Evaluation metrics of prediction models under different combinations of inputs

序号	输入组合	E_{MA}	E_{MS}	E_{RMS}	E_{MAP}	r	R^2
1	长沟站前水位、泵站流量、邓楼站前水位、泵站流量	0.017 9	0.001 0	0.034 3	0.000 5	0.983 2	0.967 2
2	长沟站前水位、邓楼站前水位、邓楼泵站流量	0.014 2	0.000 9	0.024 1	0.000 4	0.982 9	0.965 7
3	长沟站前水位、长沟泵站流量、邓楼泵站流量	0.016 3	0.001 0	0.032 5	0.000 5	0.982 3	0.964 3
4	长沟站前水位、长沟泵站流量、邓楼站前水位	0.017 3	0.001 1	0.032 7	0.000 5	0.982 7	0.966 2

由表1模型评价指标计算结果可知，构建的LSTM神经网络实时预测模型精度符合模型要求，可以实现对邓楼泵站站前水位的实时精确预测。

3.2 输入变量敏感性分析

根据3.1计算结果可知：影响因子数量越多，

预测结果精确度越高；当输入因子分别减少长沟泵站流量、邓楼站前水位、邓楼泵站流量时，模型评价指标下降程度不同。根据不同下降程度，得到表2。

综上所述，选取长沟站前水位、长沟泵站流量、邓楼站前水位为输入因子。

表 2 不同输入因子相关性

Tab. 2 Correlation of different input factors

输入因子	等级	灵敏度分级
长沟站前水位	1	极为敏感
长沟泵站流量	2	敏感
邓楼站前水位	3	敏感
邓楼泵站流量	4	一般敏感

3.3 不同预见期下站前水位预测结果分析

3.3.1 结果分析

根据 3.2 所得结果,选择表 1 中输入组合 4 进

行系统分析,以寻求不同预见期下邓楼泵站站前水位实时预测结果。不同预见期下站前水位预测分析结果见图 4。经模型计算,预见期为 2、4 和 6 h 的预测精度分析结果见表 3。由此可见,无论预见期如何变化,模型预测值与实测值间的相对误差保持稳定,不仅体现了模型在控制相对误差上的一致性,而且证明了模型性能对于预见期的非敏感性,即无论预测时段的长短,模型在相对误差上均表现稳定。

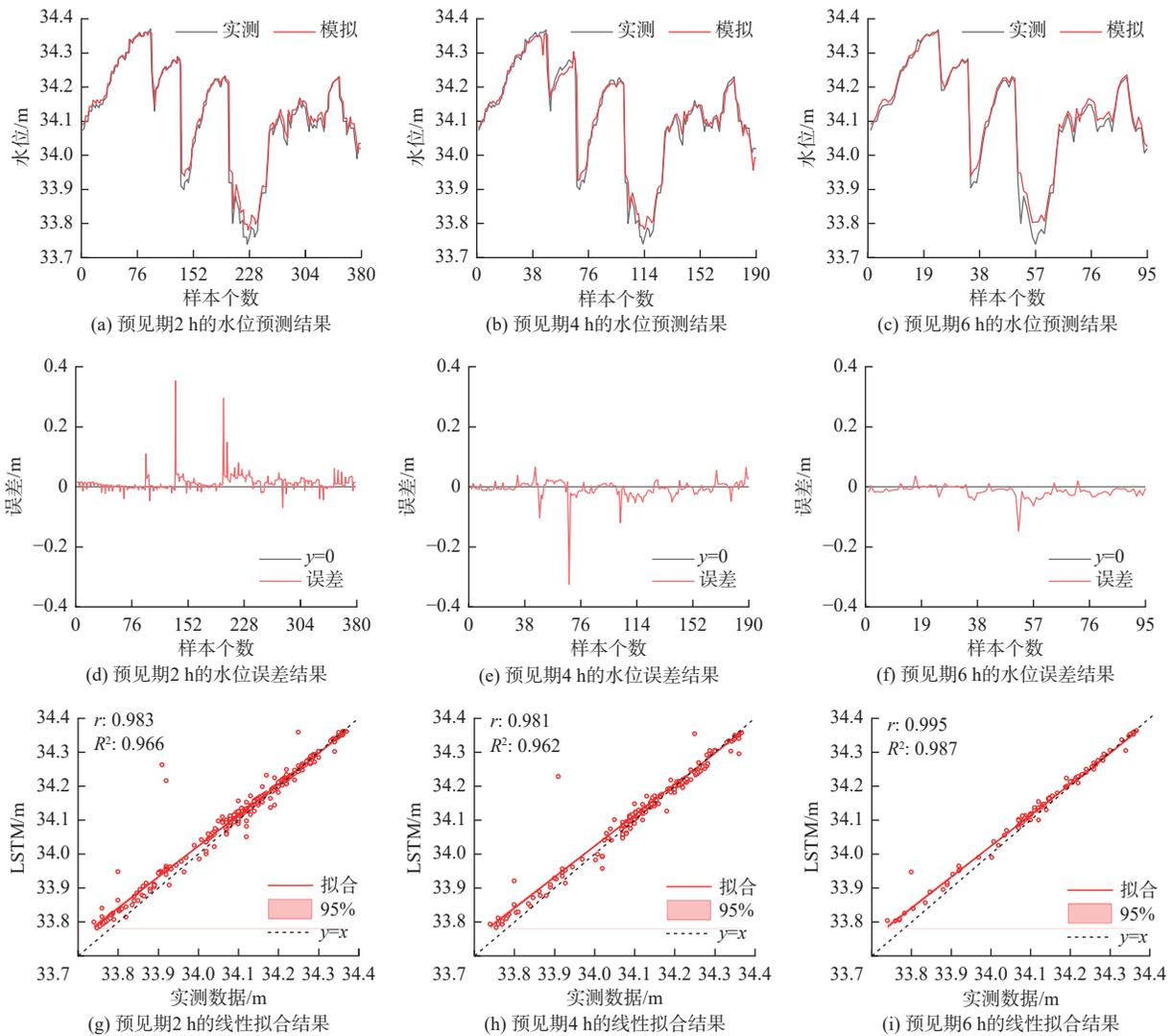


图 4 各预见期站前水位预测结果分析

Fig. 4 Analysis of the results of the prediction of pre-station water levels for each forecast period

表 3 预见期 2、4 和 6 h 预测精度分析

Tab. 3 Analysis of prediction accuracy in the prediction period of 2, 4, and 6 h

预见期/h	最大误差/m	截距	斜率	E_{MA}	E_{MS}	E_{RMS}	E_{MAP}
2	0.353 6	3.009 5	0.912 2	0.017 3	0.001 1	0.032 7	0.000 5
4	0.325 5	3.076 4	0.910 0	0.017 7	0.001 0	0.032 3	0.000 5
6	0.147 2	3.121 6	0.908 9	0.017 6	0.000 6	0.025 7	0.000 5

综上所述, LSTM 神经网络实时预测模型在跨越多个预见期的预测任务中展现出了良好的预测精度、高度的稳定性以及准确的长期趋势捕捉能力,充分证明了其在处理复杂时间序列预测问题时的可靠性,能够实现对邓楼泵站站前水位的精准预测。

3.3.2 误差分析

由图 4 水位误差结果可知,水位预测误差点较少,但误差值较大,因此进一步讨论造成平均误差较大的原因。图 5 为水位误差箱线分布情况,预见期为 2 h 时,误差值中位数为 0.009 5 m,方差为 0.000 9 m,平均值为 0.013 2 m;预见期为 4 h 时,误差值中位数为 -0.005 4 m,方差为 0.000 1 m,平均值为 -0.006 7 m;预见期为 6 h 时,误差值中位数为 -0.013 3 m,方差为 0.000 4 m,平均值为 -0.015 3 m。由此可见,不同预见期下的误差值中位数较小,其绝对值最大不超过 0.02 m。对原始数据进行分析,图 4 水位误差结果图的最大误差较大,其原因在于人工调控导致数据异常。除去异常点之外,整体误差较小,预见期为 4 h 时 LSTM 神经网络实时预测模型精度最高。

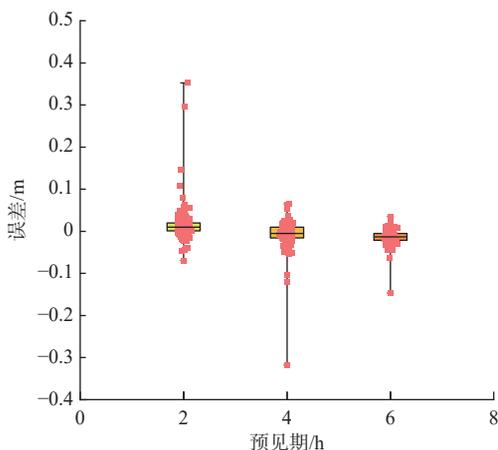


图 5 水位误差箱线图

Fig. 5 box line diagram of water level error

综上所述,模型预测结果与实测值具有较高的拟合度,其误差在合理范围之内, LSTM 神经网络预测模型预测效果稳定,可以用于邓楼泵站站前水位的实时预测。

4 结论

本文基于时间序列分析理论构建了多变量 LSTM 神经网络站前水位实时预测模型,并将模型应用于南水北调东线山东段邓楼泵站,得出结论如下:

研究构建的 LSTM 神经网络实时模型能够精

确预测邓楼泵站站前水位,且与实测值拟合较好,预测效果显著,可为梯级泵站站前水位实时预测提供新的方法与思路;

由于 LSTM 神经网络记忆单元的长期依赖性和参数共享性,选择 OAT 法优化输入参数。当选择恰当的输入参数组合时,模型的 E_{MA} 、 E_{MS} 、 E_{RMS} 和 E_{MAP} 这 4 项指标均小于 0.05, r 和 R^2 均大于 0.95;

由于人工调控会导致部分时刻水位突变,通过探寻误差箱线分布情况,寻找并剔除异常数据,所得结果进一步验证了所构建的 LSTM 神经网络站前水位实时预测模型的预测精度。

参考文献:

- [1] 杜梦盈,张召,李谷涵,等.大型水利工程梯级泵站短期优化调度方案[J].排灌机械工程学报,2024,42(2):194-200. DOI: 10.3969/j.issn.1674-8530.23.0049.
- [2] 岳俊杰,桑国庆,张林,等.南水北调工程对南四湖下级湖水体交换特征影响研究[J].水利水电技术(中英文),2024,55(6):72-83. DOI: 10.13928/j.cnki.wrahe.2024.06.006.
- [3] YAN P, ZHANG Z, HOU Q, et al. A novel IBAS-ELM model for prediction of water levels in front of pumping stations[J]. Journal of Hydrology, 2023 (616): 128810. DOI: 10.1016/j.jhydrol.2022.128810.
- [4] 王富建,尚露露,马海华.梯级泵站输水系统调蓄水池容量变化特征[J].水电能源科学,2024,42(5):166-169. DOI: 10.20040/j.cnki.1000-7709.2024.20231194.
- [5] 陈奂良,李常锁,高帅,等.基于遗传算法优化 BP 神经网络的岩溶泉水位预测研究[J].水资源与水工程学报,2022,33(4):50-57. DOI: 10.11705/j.issn.1672-643X.2022.04.07.
- [6] 刘晓伟,哈明虎,雷晓辉,等.基于 GRA-NARX 神经网络的泵站站前水位预测模型[J].南水北调与水利科技(中英文),2022,20(4):773-781. DOI: 10.13476/j.cnki.nsbdkq.2022.0079.
- [7] 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. DOI: 10.13476/j.cnki.nsbdkq.2023.0109.
- [8] 杨婷婷,虞佳颖,肖姚,等.基于 Embedding-GRU 的水库水位预测模型[J].南水北调与水利科技(中英文),2023,21(5):940-950. DOI: 10.13476/j.cnki.nsbdkq.2023.0100.
- [9] 张睿芝,戴凌全,戴会超,等.基于改进 LSTM 模型的三峡水库蓄水对洞庭湖水位影响的空间异质性分析[J].水利水电技术(中英文),2022,53(2):98-108.

- DOI: [10.13928/j.cnki.wrahe.2022.02.010](https://doi.org/10.13928/j.cnki.wrahe.2022.02.010).
- [10] HRNJICA B, BONACCI O. Lake level prediction using feed forward and recurrent neural networks[J]. *Water Resources Management*, 2019, 33(7): 2471-2484. DOI: [10.1007/s11269-019-02255-2](https://doi.org/10.1007/s11269-019-02255-2).
- [11] 胡昊, 马鑫, 徐杨, 等. 基于权重修正和 DRSN-LSTM 模型的向家坝下游水位多时间尺度预测[J]. *水利水电技术(中英文)*, 2022, 53(7): 46-57. DOI: [10.13928/j.cnki.wrahe.2022.07.005](https://doi.org/10.13928/j.cnki.wrahe.2022.07.005).
- [12] 张伟, 邓彬彬, 仇建春, 等. 基于 ESMD-FE-AJSO-LSTM 算法的水闸深基坑变形预测[J]. *南水北调与水利科技(中英文)*, 2024, 22(2): 378-387. DOI: [10.13476/j.cnki.nsbdqk.2024.0039](https://doi.org/10.13476/j.cnki.nsbdqk.2024.0039).
- [13] LIU Y, ZHAO Q, HU C, et al. Prediction of storm surge water level based on machine learning methods[J]. *Atmosphere*, 2023, 14(10): 1568-1588. DOI: [10.3390/atmos14101568](https://doi.org/10.3390/atmos14101568).
- [14] 袁旦, 谭尧耕, 朱艳霞, 等. 改进 LSTM 在金沙江流域中长期径流预报研究[J]. *水利水电技术(中英文)*, 2024, 55(1): 28-38. DOI: [10.13928/j.cnki.wrahe.2024.S1.004](https://doi.org/10.13928/j.cnki.wrahe.2024.S1.004).
- [15] 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](https://doi.org/10.3969/j.issn.1007-2284.2020.10.034).
- [16] 武少振, 任智慧, 赵雪花, 等. 基于多模型比选耦合的降水预测[J]. *南水北调与水利科技(中英文)*, 2024, 22(1): 99-109. DOI: [10.13476/j.cnki.nsbdqk.2024.0012](https://doi.org/10.13476/j.cnki.nsbdqk.2024.0012).
- [17] MA X, HU H, REN Y. A hybrid deep learning model based on feature capture of water level influencing factors and prediction error correction for water level prediction of cascade hydropower stations under multiple time scales[J]. *Journal of Hydrology*, 2023, 617: 129044. DOI: [10.1016/j.jhydrol.2022.129044](https://doi.org/10.1016/j.jhydrol.2022.129044).
- [18] 张研, 廖逸夫, 王鹏鹏, 等. 基于相关向量机的调水工程调蓄水位预测模型[J]. *南水北调与水利科技(中英文)*, 2021, 19(4): 814-821. DOI: [10.13476/j.cnki.nsbdqk.2021.0085](https://doi.org/10.13476/j.cnki.nsbdqk.2021.0085).
- [19] 杨靖仁, 王超, 雷晓辉, 等. 南水北调东线江苏段典型泵站运行效率模拟模型[J]. *南水北调与水利科技(中英文)*, 2024, 22(2): 388-398. DOI: [10.13476/j.cnki.nsbdqk.2024.0040](https://doi.org/10.13476/j.cnki.nsbdqk.2024.0040).
- [20] 王林威, 苏柳, 谭培影, 等. 东平湖调蓄南水北调东线来水的可行性[J]. *南水北调与水利科技(中英文)*, 2024, 22(3): 445-454. DOI: [10.13476/j.cnki.nsbdqk.2024.0046](https://doi.org/10.13476/j.cnki.nsbdqk.2024.0046).
- [21] 王维林. 两湖段梯级泵站优化运行研究[D]. 济南: 济南大学, 2020. DOI: [10.27166/d.cnki.gsdcc.2020.000179](https://doi.org/10.27166/d.cnki.gsdcc.2020.000179).
- [22] WANG W, SANG G, ZHAO Q, et al. Water level prediction of pumping station pre-station based on machine learning methods[J]. *Water Supply*, 2023, 23(10): 4092-4111. DOI: [10.2166/ws.2023.242](https://doi.org/10.2166/ws.2023.242).
- [23] LIU M, ZHANG M, ZHANG P, et al. Water level prediction model based on blockchain and LSTM[J]. *Journal of Intelligent & Fuzzy Systems*, 2024, 46(1): 2371-2380. DOI: [10.3233/JIFS-231411](https://doi.org/10.3233/JIFS-231411).
- [24] 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](https://doi.org/10.13476/j.cnki.nsbdqk.2022.0040).
- [25] 闫孟婷, 黄炜斌, 张天遥, 等. 基于 JMI-CNN-LSTM 耦合模型的梯级水电站间流量动态滞时关系[J]. *水利水电技术(中英文)*, 2023, 54(3): 154-164. DOI: [10.13928/j.cnki.wrahe.2023.03.014](https://doi.org/10.13928/j.cnki.wrahe.2023.03.014).
- [26] 石朋, 丁松, 司伟, 等. 新安江模型敏感参数动态变化规律研究[J]. *河海大学学报(自然科学版)*, 2024, 52(3): 1-6. DOI: [10.3876/j.issn.1000-1980.2024.03.001](https://doi.org/10.3876/j.issn.1000-1980.2024.03.001).

LSTM model for the real-time prediction of pre-station water level based on input factor analysis

GAO Yiran, LU Longbin, SANG Guoqing, WANG Weilin, LIU Luxia

(School of Water Conservancy and Environment, University of Jinan, Jinan 250022, China)

Abstract: The step pumping station system is a very complex system that consists of the pumping station, barrage and canal section facilities, for which accurate water level prediction is recommended for the safe operation of water transfer projects. Machine learning models are widely used for water level prediction because of their fast response and small overshoot as compared with the traditional hydrodynamic equations. However, the complicated input data and the difficulty of time series prediction directly limit and affect the prediction accuracy, and raises the question as how to accurately evaluate the influence of input data on the prediction results to be solve in water level prediction model. The establishment of machine learning model helps to accurately predict the water level in front of the pump,

realize intelligent dispatching, improve the operation efficiency of the pumping station, and is of great significance to the construction of water network.

Long Short-Term Memory (LSTM), as a special recurrent neural network structure, has both the nonlinear properties of neural networks, and memory and long-term dependence, which facilitates its processing of time series data. The Changgou pumping station to Denglou pumping station in the Nansi-Dongping Lake section of the South-to-North Water Transfer East Route is taken as the research object. A framework for analyzing the influence of test input data on prediction results is proposed to construct a real-time prediction model of water level in front of pumping station based on LSTM model, and the optimization of the data input set of the prediction model is realized through the sensitivity analysis of each input factor, while the prediction analysis of the water level in front of Denglou Pumping Station under different foresight periods is carried out based on the above data combination.

Four input factors, namely water level in front of Changgou station, flow rate of Changgou pumping station, water level in front of Denglou station and flow rate of Denglou pumping station, were initially determined as model inputs by the correlation coefficient method. The water level prediction model in front of Denglou pump station with a foresight period of 2 h is taken as an example, and the combination of input factors is further analyzed, with a Mean Absolute Error of 0.017 3, Mean Squared Error of 0.001 1, Root Mean Square Error of 0.032 7, and Mean Absolute Percentage Error of 0.000 5. The inputs for the pre-pumping water level prediction model were ultimately established as the water level in front of Denglou station, the flow rate at Changgou pumping station, and the water level in front of Changgou station. Following this, the accuracy of the prediction model for foresight periods of 2 hours, 4 hours, and 6 hours were analyzed, along with an error assessment based on these results. The results show that the prediction model with a foresight period of 6 h has higher accuracy. For the data anomalies existing in the model, most of them are caused by the manual regulation process, after removing the anomalies, the overall error of the model was reduced, and the prediction model with a foresight period of 4 h had the highest accuracy.

The LSTM neural network real-time model constructed can accurately predict the water level in front of Denglou pumping station, and it fits well with the measured value, with high prediction accuracy, with the absolute value of the relative water depth error of less than 0.02 m except in the case of the abnormal value. The input parameters are optimized through sensitivity analysis, and when the input parameters are optimal combinations, the four indexes of the model were less than 0.05, and the correlation coefficient and the coefficient of determination were greater than 0.95. The LSTM model can reduce the complexity of the combination of input factors and ensure the accuracy of prediction. In summary, the model prediction results have a high degree of fit with the measured values, and their errors are within a reasonable range, and the LSTM prediction model had a stable prediction effect, which can be used for real-time prediction of the water level in front of the station of the terrace pumping station system.

Key words: LSTM neural network; station water level; water level prediction; sensitivity analysis; time series