# 基于 BP 神经网络的不同时间尺度泵站前池水位预测模型

薛萍[1],张召[2],雷晓辉[2],卢龙彬[1],颜培儒[3],李月强[4]

(1. 济南大学水利与环境学院,济南 250022;2. 中国水利水电科学研究院水资源所,北京 100038;
3. 天津大学建筑工程学院,天津 300072;4. 河海大学水利水电学院,南京 210098)

**摘要:**针对建筑物调控下的水位预测难题,建立基于 BP 神经网络的泵站前池水位预测模型,在不同时间尺度下分析时间序列、影响因子、预报因子对水位预测精度的影响。将构建模型应用于胶东调水工程东宋泵站,研究结果表明:在数据总量一定的情况下,训练期和预测期之比为 7:3 时,预测结果较好;数据量越大,保持一定预报精度所需要的正相关的影响因子数量越多;短时间内,预测时间间隔与数据本身时间间隔相同时,预测效果更好。该构建模型能够满足明渠调水工程泵站前池的水位动态预测需求,实现泵站前池水位的 2 h 准确预测和 4 h 一般准确预测,同时可在其他类似明渠调水工程中推广应用。

**关键词:**泵站前池;水位预测;BP 神经网络;时间序列;比例

**中图分类号:**TV675　　　**文献标志码:**A　　　开放科学(资源服务)标志码(OSID):

明渠调水工程在进行长距离输水调度时,一般通过在渠道中设置泵站、节制闸、倒虹吸等水工建筑物解除地形条件对输水限制的影响,同时在建筑物前设置水位计、流量计等监测设备获取水情信息监控通水安全。相比于实时水位监测,高精度的水位预测更能在水量调度过程中为调度人员提供科学指导,尤其是泵站前池水位预测,对泵站调控、水量调度、渠道安全均具有重要意义。受气候、温度、人类活动等多种因素影响,监测设备采集到的水情序列往往呈现出非线性和不确定性的特点,常规方法很难对其进行规律分析和趋势预测。学者[1-3]曾通过建立水力学模型模拟渠道水流的变化过程,但建模要求完整且准确的地形资料、工程参数和实测数据,糙率率定过程也较为反复和繁琐[4],因此存在较大

的局限性。随着人工智能技术和机器学习方法的不断进步,采用数据驱动的方法进行预测可避免水力学建模的多方面要求和诸多限制,直接探索数据间的内在规律[5]。

到目前为止,大部分学者[6-9]通过构建神经网络模型进行水位预测,如采用优化后的 RBF 神经网络、LSTM 神经网络模型、小波神经网络等应用于地下水位预测,预测精度高且预测效果处于较优水平;虽可建立向量机 RVM 预测模型[10]、Mike 模型[11]、相似模型[12]、统计模型[13]、贝叶斯模型[14]等进行水位预测,但使用时受限于一定的条件,故应用于调水工程中水位预测时不太广泛;因神经网络已广泛应用于水位预测,发展逐渐趋于成熟,也有众多学者[15-26]通过将神经网络模型、算法组合或者改进

算法的方式进行水位预测,如吴美玲[27]等将 KNN、GA、BP 相结合,对秦淮河的洪水位进行预测,相比于未组合的神经网络模型预测精度提高但略为复杂,但未组合的神经网络模型较为简单实用,如高学平等[28]利用 BP 神经网络对泵站站前水位进行预测,发现 BP 神经网络在解决非线性问题上有很大优势,在智能预测方面存在巨大潜力。同时,常用的评价指标有 $E_{RMS}$(均方根误差)、$R^2$(决定系数)[29]等。

综上可知,构建神经网络进行水位预测是一种切实可行的研究方法。人工神经网络等智能算法在水文预测应用中具有一定的适用性条件,如:ANN 有强大非线性能力,但结构简单不能保存前时信息而无法学习时间序列数据;RNN 能保持先前时刻的水位预测,可有效处理序列数据,但梯度传递中存在缺陷;LSTM 具有长短期记忆功能,在一定程度上解决梯度消失和梯度爆炸,但长序列依旧存在问题且不能并行;受信息单向流动特点的限制,经典 BP 神经网络考虑有限数量的历史信息,仅适用于短时预测,但结构稳定,具有多功能性和简便性的特征,可灵活处理非线性问题并达到较高的预测精度,具有极强的非线性映射能力;而水文预测中的水情序列因受人为因素影响较大,呈现出较大的非线性特点,故 BP 神经网络适用于水文预测。BP 神经网络自 1986 年被 Rumelhart 等[30]提出后,已被广泛应用于水文预测领域的研究。本文通过建立 BP 神经网络,利用历史数据预测泵站前池未来时刻的水位,分析时间序列比例及影响因子对水位预测的影响,预测结果既可为泵站前池水位预测提供一种预测方式,也给泵站前池水位变化趋势提供参考数据。

# 1　研究方法

选取泵站前池水位为研究对象,利用相关性分析确定影响因子,并将其作为输入进行 BP 神经网络模型构建,预测结果用各指标参数情况评判优劣。

## 1.1　影响因子识别

受各种水力因素(断面面积、水力比降、糙率等)影响,渠道内断面流量和水位之间存在对应关系。泵站前池水位作为监测断面之一,与相邻断面的水位、泵站的流量、上游流量、流量差等均可能存在水力联系。将这些相关的水位、流量等作为变量,对各变量与预测因子进行相关性分析,识别出具有一定关联度的影响因子。

采取的影响因子识别方法有皮尔逊(Pearson)相关系数法、肯德尔(Kendall)相关性系数法、斯皮尔曼(Spearman)等级相关系数法及灰关联分析。皮尔逊相关系数法用于度量 2 个变量之间的相关程度,2 个变量之间的皮尔逊相关系数定义为 2 个变量之间的协方差和标准差的商;肯德尔相关性系数法是表示多列等级变量相关程度的一种方法,若 $n$ 个同类的统计对象按特定属性排序,其他属性通常是乱序的,同序对和异序对之差与总对数$[n(n-1)/2]$的比值定义为肯德尔系数;斯皮尔曼等级相关系数法是根据等级资料研究 2 个变量间相关关系的方法,依据 2 列成对等级的各对等级数之差来进行计算,利用单调方程评价 2 个统计变量的相关性。上述 3 种方法的相关性指标或相关系数为 $-1\sim1$:绝对值越接近 1,相关性越高;绝对值等于 0 时,不具备相关性。灰关联分析是一种分析系统中各因子关联程度的量化方法,根据不同变量序列间发展趋势的相似或相异程度,衡量因素间关联程度。灰色关联度小于 0.6 时,不具有相关性;灰色关联度越趋近 1,相关性程度越高。

## 1.2　BP 神经网络

BP 神经网络是一个利用误差反向传播算法进行训练的多层前馈神经网络,一般包括输入层、隐含层、输出层 3 部分。输入层具有信息接入即信号接收功能,信号接收完成后将信息传递到隐含层,输入层神经元的个数为输入影响因子的数量 $n$;隐含层负责信息处理、信息变换,隐含层神经元的个数为 $m$,小于 $N-1$($N$ 是训练样本数),在 MATLAB 中经测试取值;经隐含层后信息传递到输出层,输出层将结果对外输出,1 个 3 层的典型网络结构见图 1。
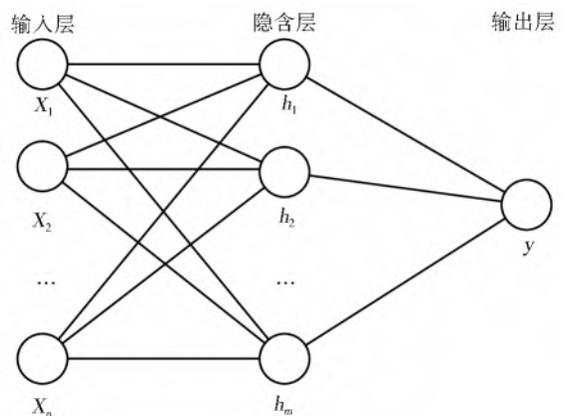


图 1　BP 神经网络模型结构

神经网络结构参数设置有:最大训练次数=100,训练要求精度=$1\times10^{-8}$,学习率=0.01。参

数设置完成后,网络利用误差的反向传播自动调整权重和阈值,驱使 BP 神经网络中表达函数能够得到最优解,最后输出预测结果及评判结果的各项指标值。

## 1.3 预测结果评判标准

以 $R^2$(决定系数)、$E_{RMS}$(均方根误差)、$E_{MA}$(平均绝对误差)为评判标准对预测结果的优劣进行评判,$R^2$ 越趋近 1,$E_{RMS}$ 和 $E_{MA}$ 越趋近 0,说明预测精度越高。

## 2 研究区概况

胶东调水工程是山东省水利建设的重要组成部分,包括引黄调水工程和引黄济青工程 2 条输水线路。引黄济青工程于 1986 年 4 月 15 日开工兴建,1989 年 11 月 25 日正式通水;引黄调水工程于 2003 年 12 月 19 日开工,2013 年 7 月全线贯通,2013 年 12 月主体工程建成通水。其中,引黄调水工程包括明渠段和管道段两部分,明渠段以宋庄分水闸为起点,以黄水河泵站为终点,途经灰埠、东宋、辛庄 3 座泵站及若干倒虹吸、渡槽等输水建筑物,全长约 160 km。本文所选研究区为引黄调水工程明渠段,具体研究区域为东宋泵站前后,其上游控制节点为灰埠泵站,下游控制节点为埠上节制闸,该渠段及沿线建筑物分布情况见图 2。
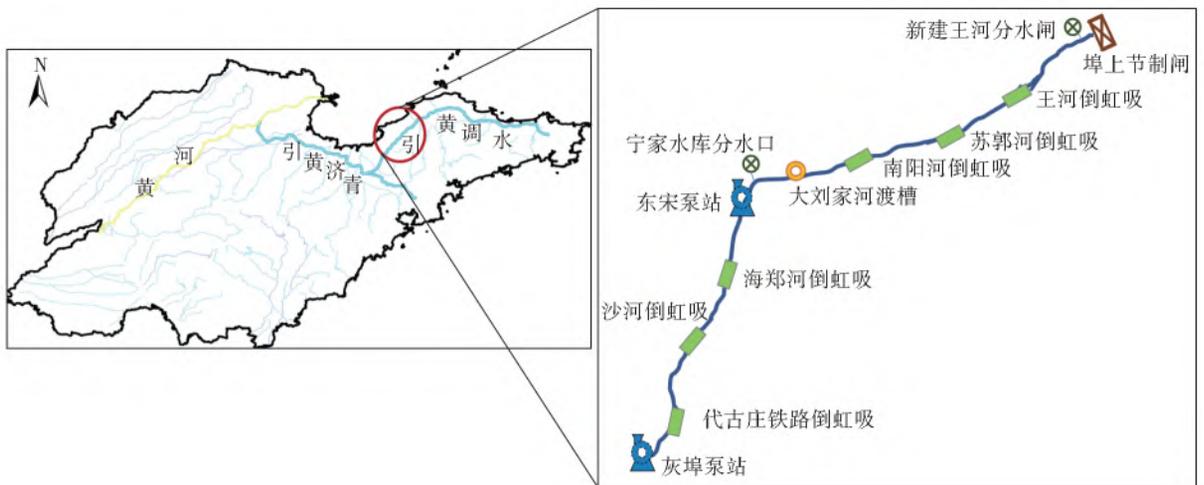


图 2 研究渠段及沿线建筑物分布

## 3 结果与讨论

### 3.1 影响因子识别结果

研究东宋泵站未来时刻的前池水位时,考虑到水位流量间关系及人为因素影响,除选取相邻断面水位作为影响因子外,还选取东宋泵站流量、灰埠泵站流量、灰埠-东宋 2 级泵站流量差为影响因子进行预测,且影响因子均为当前时刻的影响因子。表 1 为不同方法下的各因子与前池水位间的相关性分析结果。

表 1 影响因子相关性分析结果
Tab.1 Correlation analysis results of impact factors

| 相关性指标 | 海郑河倒虹吸上游水位 | 海郑河倒虹吸下游水位 | 东宋泵站站前水位 | 东宋泵站站后水位 | 泵站流量 | 2 级泵站流量差 | 上游泵站流量 |
|---|---|---|---|---|---|---|---|
| 皮尔逊相关系数 | 0.846 | 0.876 | 0.911 | −0.045 | 0.224 | 0.911 | −0.034 |
| 肯德尔相关性系数 | 0.648 | 0.642 | 0.747 | −0.067 | 0.233 | 0.747 | −0.067 |
| 斯皮尔曼等级相关系数 | 0.829 | 0.825 | 0.901 | −0.067 | 0.322 | 0.901 | −0.067 |
| 灰色关联度 | 0.792 | 0.920 | 0.934 | 0.764 | 0.807 | 0.893 | 0.731 |

由表 1 可知,影响因子相关性排序从高到低依次为东宋泵站站前水位、2 级泵站流量差、海郑河倒虹下游水位、海郑河倒虹上游水位、泵站流量、东宋泵站流量及上游泵站流量。前 4 项影响因子的各系数均为 0.8～0.9,识别为相关性较高的影响因子,建模时优先考虑;后 3 项影响因子的指标中仅灰色关联度表明其相关性程度较高,故识别为相关性较低的影响因子,建模时可考虑在内,但不重点考虑。

### 3.2 水位预测结果分析

利用 BP 神经网络模型进行泵站前池水位预测,预测结果从时间序列、影响因子 2 个方面进行分析。

### 3.2.1 时间序列

将不同时间尺度的数据按照一定的比例进行训练和验证,对比训练时长和预测精度。结果表明,训练期和预见期的最优比例为 7∶3,减小该比例会使预测精度降低,增大该比例预测精度与之相差无几,且数据需求量大幅提升。

采用 3 600 个数据预测未来 2 h 的水位变化,7∶3 比例下的 $R^2$、$E_{RMS}$、$E_{MA}$ 分别维持在 0.95、0.04、0.03 左右。增大该比例时各指标预测效果略有提高,但相差不大,高于5∶1时其预测精度基本不提高。具体对比见图 3 和图 4。
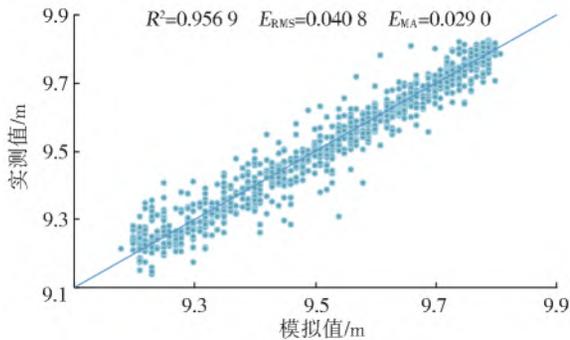
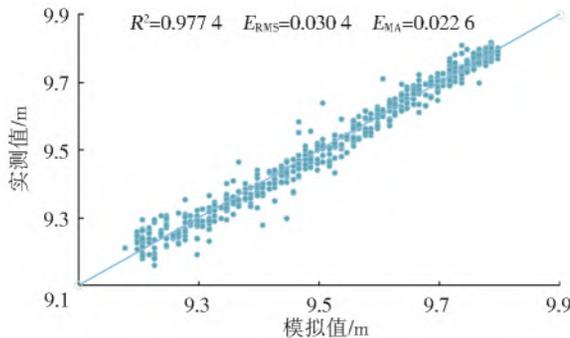

图 3 未来 2 h 水位预测结果(7∶3)



图 4 未来 2 h 水位预测结果(5∶1)

以 7∶3 的比例分别对 3 组 3 个月的数据进行训练和验证,$R^2$ 维持在 0.93~0.98,$E_{RMS}$ 维持在 0.02~0.05、$E_{MA}$ 维持在 0.02~0.04,预测结果见图 5。
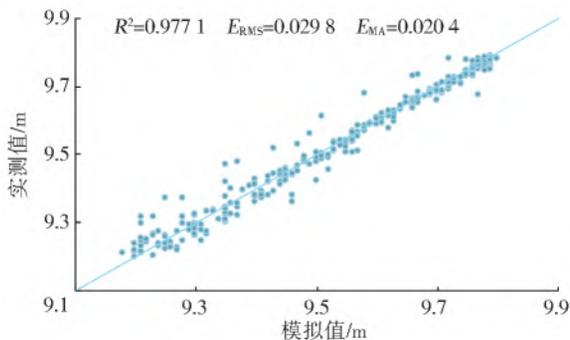


图 5 未来 2 h 水位预测结果(7∶3)

以 7∶3 的比例对 1 个月的数据进行验证,验证结果表明该比例对 1 个月的数据量依旧适用,具体见图 6。
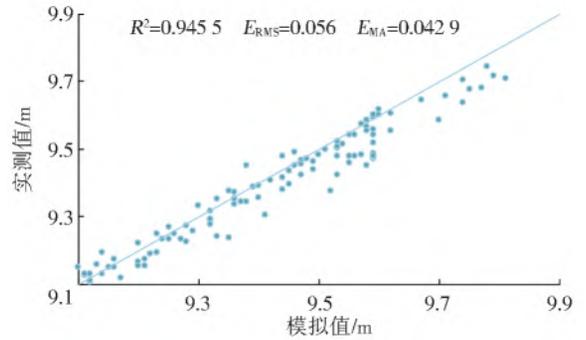


图 6 未来 2 h 水位预测结果(7∶3)

由上述可知,最优比例适用于不同时间尺度的数据,且最优比例的确定既可节省神经网络学习的时间,又能提高预测精度,在模型中具有较大影响力。

### 3.2.2 影响因子

影响因子数量。当影响因子与预测因子之间都具有较高相关性时,影响因子数量越多,预测结果越精确。但影响因子的数量会增加训练期的数据需求量,为减少数据需求量且保证预测精度,利用不同数量的影响因子进行训练和验证,验证结果表明:短期(1~3 个月)内至少选择 3~5 个影响因子进行训练,3 个月至 1 a 的数据量则至少需要 5~7 个相关性最大的影响因子。

影响因子种类。研究表明,选取相关性最高的影响因子构建模型,预测精度更高。由影响因子相关性分析结果可知,相关性最高的 3 个影响因子为泵站当前时刻的水位、上游相邻节点的水位、流量差。采用 3 个影响因子对 1 个月的数据进行训练和预测,上述 3 个影响因子的预测效果最佳,具体见图 7。



图 7 3 个因子水位预测结果(7∶3)

影响因子的时间间隔。数据间隔均为 2 h 时,对东宋泵站未来时刻的水位进行预测:未来 2 h 的水位预测结果较稳定,$R^2$ 结果均在 0.9 以上,$E_{RMS}$ 和 $E_{MA}$ 也较小;未来 4 h 的水位预测结果一般,$R^2$ 为 0.8~0.9,$E_{RMS}$ 和 $E_{MA}$ 比 2 h 预测略大;未来 6 h

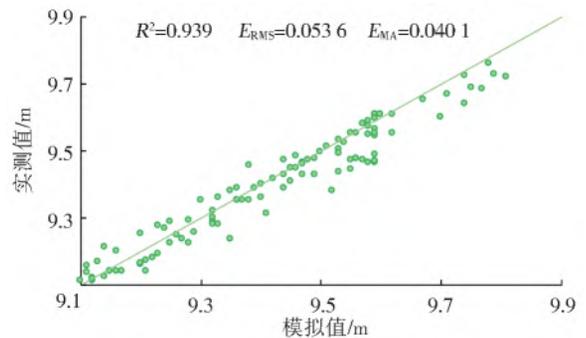的水位预测结果较差,$R^2$ 不稳定且变化区间较大,结果较好时也仅为 0.7 左右,$E_{RMS}$ 和 $E_{MA}$ 则预测结果偏大,分别在 0.11 和 0.09 左右。即训练期内数据不发生改变时,预测时间越长,预测精度越低。对 3 个月的数据进行筛选,使 2 h 间隔转为 4 h 间隔,并预测东宋泵站未来 4 h 的水位,预测结果见图 8。
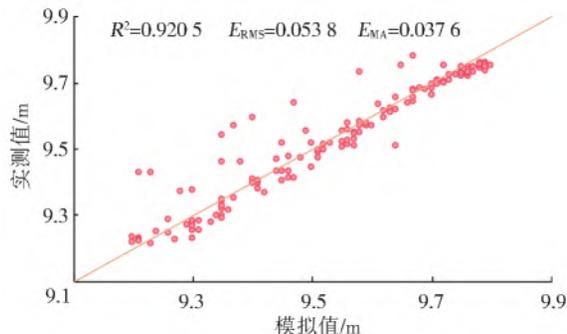


**图 8  未来 4 h 水位变化结果(7:3)**
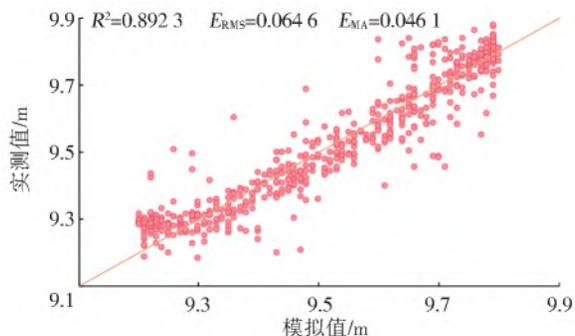
对 10 个月的数据进行 4 h 间隔的筛选并预测未来 4 h 水位,预测结果见图 9。



**图 9  未来 4 h 水位变化结果(7:3)**

研究结果表明,与采用 2 h 间隔的数据直接预测相比,采用 4 h 间隔的数据预测泵站未来 4 h 的水位,其预测精度更高,$R^2$ 变化基本维持在 0.82~0.93,$E_{RMS}$ 和 $E_{MA}$ 分别维持在 0.05~0.06、0.04~0.05。

对 1 a 的数据进行筛选,使 2 h 间隔转为 6 h 间隔,并预测东宋泵站未来 6 h 的水位。预测结果表明,筛选后进行预测比用 2 h 的数据直接预测其预测效果更差。经分析,上述现象是由 6 h 的时间间隔太长不能完全反映各因子变化规律导致,所以筛选后进行预测的结果比直接采用 2 h 间隔的数据进行预测结果更差。

# 4  结  论

时间序列比例对水位预测结果的影响:训练期和预测期的最佳比例为 7:3,提高比例其预测精度无明显变化,降低比例则预测效果变差。

影响因子对预测结果的影响:数据量与影响因子数量呈对应关系,3 个月的数据量需 3~5 个影响因子进行训练,3 个月至 1 a 的数据量则需 5~7 个影响因子确保相同预测效果。

数据的时间间隔对预测结果的影响:一般情况下,数据间隔不变,预测精度随预测时间的增加而逐渐降低;但当数据能够反映各因子变化规律时,数据间隔和预测时间相同,预测效果更佳。

**参考文献(References):**

[1] 张成,倪春飞,刘林. 输水系统闸前常水位控制下的区间调度研究[J]. 应用基础与工程科学学报,2015,23(S1):110-121. DOI:10.16058/j.issn.1005-0930.2015.s1.012.

[2] 周龙才,程千,李娜. 灌区长渠道梯级泵站运行优化及仿真[J]. 灌溉排水学报,2014,33(Z1):390-393. DOI:10.13522/j.cnki.ggps.2014.04/05.084.

[3] 黄会勇,刘子慧,范杰,等. 南水北调中线工程总干渠水力学仿真模型研究[J]. 水利水电技术,2013,44(12):111-115,122. DOI:10.13928/j.cnki.wrahe.2013.12.008.

[4] 郑和震,马雄威,雷晓辉,等. 南水北调中线工程渠道糙率计算方法研究[J]. 南水北调与水利科技,2018,16(1):158-163. DOI:10.13476/j.cnki.nsbdqk.20180024.

[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] 王晓伟,赵斌,戈天剑,等. 基于 RBF 神经网络地下水位预测[J]. 农业科技与信息,2018(4):60-61,71. DOI:10.15979/j.cnki.cn62-1057/s.2018.04.028.

[7] 闫佰忠,孙剑,王昕洲,等. 基于多变量 LSTM 神经网络的地下水水位预测[J]. 吉林大学学报(地球科学版),2020,50(1):208-216. DOI:10.13278/j.cnki.jjuese.20190055.

[8] 徐强,束龙仓,杨桂莲,等. 基于遗传算法优化的小波神经网络在地下水位预测中的应用[J]. 水文,2010,30(1):27-30. DOI:10.3969/j.issn.1000-0852.2010.01.006.

[9] 齐慧. 基于 KNN-LSTM 模型的常武地区地下水位时空预测方法研究[D]. 南京:南京师范大学,2020.

[10] 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.

[11] TU Z J,GAO X G,XU J,et al. A novel method for regional short-term forecasting of water level[J]. Water,2021,13(6),820. DOI:10.3390/W13060820.

［12］　徐源蔚,李祚泳,汪嘉杨. 基于集对分析的相似模型在地下水位预测中的应用[J]. 水文,2015,35(6):6-10. DOI:10. 3969/j. issn. 1000-0852. 2015. 06. 002.

［13］　MITSUHIRO N,SHOICHI S,YUKI K,et al. Evaluation of real-time water level prediction technology using statistical models for reducing urban flood risk: Special issue on actions toward futuristic urban flood risk research and management[J]. Journal of Disaster Research,2021,16(3):387-394. DOI:10. 20965/JDR. 2021. P0387.

［14］　LIU Z Y,CHENG L Y,LIN K R,et al. A hybrid Bayesian vine model for water level prediction[J]. Environmental Modelling & Software,2021,142,105075. DOI:10. 1016/J. ENVSOFT. 2021. 105075.

［15］　刘墨阳,李巧玲,李致家,等. 基于小波分析的 NARX 神经网络在水位预测中的应用[J]. 南水北调与水利科技,2019,17(5):56-63. DOI:10. 13476/j. cnki. nsbdqk. 2019. 0110.

［16］　许国艳,周星熠,司存友,等. 基于 GRU 和 LightGBM 特征选择的水位时间序列预测模型[J]. 计算机应用与软件,2020,37(2):25-31,53. DOI:10. 3969/j. issn. 1000-386x. 2020. 02. 005.

［17］　龚莎,彭宏玉. 基于 PCA-ESN 模型的潘家口水库水位预测研究[J]. 唐山学院学报,2020,33(3):37-41,67. DOI:10. 16160/j. cnki. tsxyxb. 2020. 03. 008.

［18］　刘青松,严华,卢文龙. 基于 AR-RNN 的多变量水位预测模型研究[J]. 人民长江,2020,51(10):94-99. DOI:10. 16232/j. cnki. 1001-4179. 2020. 10. 016.

［19］　王亦斌,孙涛,梁雪春,等. 基于 EMD-LSTM 模型的河流水量水位预测[J]. 水利水电科技进展,2020,40(6):40-47. DOI:10. 3880/j. issn. 1006-7647. 2020. 06. 008.

［20］　REN T,LIU X F,NIU J W,et al. Real-time water level prediction of cascaded channels based on multilayer perception and recurrent neural network[J]. Journal of Hydrology,2020,585,124783. DOI:10. 1016/j. jhydrol. 2020. 124783.

［21］　GAN M,PAN S Q,CHEN Y P,et al. Application of the machine learning LightGBM model to the prediction of the water levels of the lower Columbia River[J]. Journal of Marine Science and Engineering,2021,9(5),496. DOI:10. 3390/JMSE9050496.

［22］　LUO Y,DONG Z C,LIU Y H,et al. Research on stage-divided water level prediction technology of rivers-connected lake based on machine learning:A case study of Hongze Lake,China[J]. Stochastic Environmental Research and Risk Assessment,2021,35(10):2049-2065. DOI:10. 1007/S00477-021-01974-6.

［23］　BAEK S S,PYO J C,CHUN J A. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach[J]. Water,2020,12(12),3399. DOI:10. 3390/W12123399.

［24］　CHEN Y P,GAN M,PAN S Q,et al. Application of auto-regressive (AR) analysis to improve short-term prediction of water levels in the Yangtze estuary[J]. Journal of Hydrology,2020,590,125386. DOI:10. 1016/J. JHYDROL. 2020. 125386.

［25］　EHTERAM M,FERDOWSI A,FARAMARZPOUR M,et al. Hybridization of artificial intelligence models with nature inspired optimization algorithms for lake water level prediction and uncertainty analysis[J]. Alexandria Engineering Journal,2021,60(2):2193-2208. DOI:10. 1016/J. AEJ. 2020. 12. 034.

［26］　吴美玲,杨侃,张铖铖. 基于 KG-BP 神经网络在秦淮河洪水水位预测中的应用[J]. 水电能源科学,2019,37(2):74-77,81. DOI:CNKI:SUN:SDNY. 0. 2019-02-019.

［27］　XIONG B,LI R P,REN D,et al. Prediction of flooding in the downstream of the Three Gorges Reservoir based on a back propagation neural network optimized using the AdaBoost algorithm[J]. Natural Hazards,2021,107(2):1559-1575. DOI:10. 1007/S11069-021-04646-4.

［28］　高学平,闫晨丹,张岩,等. 基于 BP 神经网络的调水工程调蓄水位预测模型[J]. 南水北调与水利科技,2018,16(1):8-13. DOI:10. 13476/j. cnki. nsbdqk. 20180002.

［29］　郭燕,赖锡军. 基于长短时记忆神经网络的鄱阳湖水位预测[J]. 湖泊科学,2020,32(3):865-876. DOI:10. 18307/2020. 0325.

［30］　RUMELHART D E,HINTON G E,WILLIAMS R J,et al. Learning representations by back-propagating errors[J]. Nature,1986(323):533-536. DOI:10. 1038/323533a0.

第 20 卷　第 2 期
2022 年 4 月

南水北调与水利科技（中英文）
South-to-North Water Transfers and Water Science & Technology

Vol.20　No.2
Apr. 2022

・译文・

# Prediction model for forebay water level of pumping stations with different time scales based on BP neural networks

XUE Ping[1], ZHANG Zhao[2], LEI Xiaohui[2], LU Longbin[1], YAN Peiru[3], LI Yueqiang[4]

(1. School of Water Conservancy and Environment, University of Jinan, Jinan 250022, China; 2. Institute of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China; 3. School of Civil Engineering, Tianjin University, Tianjin 300072, China; 4. College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China)

**Abstract:** Considering the difficulty in water level prediction under building control, a water level prediction model for the forebay of a pumping station was built on the basis of back-propagation (BP) neural networks, and the influence of time series and impact factors on the accuracy of water level prediction was analyzed under different time scales. The constructed model was applied to the Dongsong Pumping Station of the Jiaodong Water Transfer Project. The research results revealed that: when the total amount of data was fixed, and the ratio of the training period to the prediction period was 7 : 3, the prediction result was good; a larger amount of data was accompanied by a greater number of positively correlated impact factors required for certain prediction accuracy; in a short period of time, when the prediction time interval was the same as the time interval of the data itself, the prediction effect was better. The constructed model can meet the demand for dynamic prediction of the water level in the forebay of the open channel water transfer project and can achieve the 2 h accurate prediction of the forebay water level of the pumping station and the 4 h general accurate prediction. Additionally, it can be popularized and applied in other similar open channel water transfer projects.

**Key words:** forebay of pump station; water level prediction; BP neural network; time series; proportion

For the long-distance water dispatching of an open channel water transfer project, hydraulic structures such as pumping stations, control gates, and inverted siphons are generally set up in the channel to relieve the influence of terrain conditions on water transfer restrictions. Meanwhile, monitoring equipment such as water level meters and flow meters are installed in front of buildings to obtain water information and monitor water safety. Compared with real-time water level monitoring, high-precision water level prediction can provide more scientific guidance for dispatchers in the process of water dispatching, especially the water level prediction in the forebay of pumping stations, which is of great significance to the regulation of pumping stations, water dispatching, and channel safety. Affected by various factors such as climate, temperature, and human activities, the

hydrological sequence collected by monitoring equipment often presents the characteristics of nonlinearity and uncertainty, and it is difficult to analyze the laws and predict the trend by conventional methods. Scholars[1-3] have built hydraulic models to simulate the changing process of channel water flow, but the modeling requires complete and accurate topographic data, engineering parameters, and measured data; moreover, the calibration process of the roughness rate is also repetitive and cumbersome[4], and thus there are huge limitations. With the continuous progress of artificial intelligence technology and machine learning methods, the data-driven methods used for the prediction can avoid many requirements and limitations of hydraulic modeling and directly explore the inherent laws between data[5].

Up to now, most scholars[6-9] have built neural network models for water level prediction, such as the optimized RBF neural network, LSTM neural network model, and wavelet neural network applied in groundwater level prediction, with high prediction accuracy and an excellent prediction effect. Although the relevance vector machine (RVM) prediction model[10], Mike model[11], similarity model[12], statistical model[13], and Bayesian model[14] can be constructed for water level prediction, their applications are limited to a certain extant, and hence they are not widely used in water level prediction for water transfer projects. As the neural network has been commonly used in water level prediction, and its development has gradually matured, many scholars[15-26] have made water level predictions by combining neural network models and algorithms or improving algorithms. For instance, Wu et al. [27] combined KNN, GA, and BP to predict the flood level of the Qinhuai River, and compared with the neural network model without combination, the combined method has higher prediction accuracy but is slightly more complicated. In other words, the uncombined neural network models are simple and practical. For example, Gao et al. [28] used the BP neural network to predict the water level in front of the pumping station and found that the BP neural network has great advan-

tages in solving nonlinear problems and has significant potential in intelligent prediction. In addition, the commonly used evaluation indicators include the root mean square error ($E_{RMS}$) and determination coefficient ($R^2$)[29].

In summary, it is a feasible research method to construct a neural network for water level prediction. Moreover, intelligent algorithms such as artificial neural networks have certain applicability conditions in hydrological prediction applications. For example, ANN has a strong nonlinear ability, but due to its simple structure, previous information can not be saved, and time series data can not be learned. RNN can retain the water level prediction at the previous moment and can effectively process sequence data, but there are defects in gradient transfer. LSTM has long and short-term memory functions and can solve gradient disappearance and gradient explosion to a certain extent, but there are still problems in long sequences, and it can not be parallelized. Restricted by the one-way flow of information, the classical BP neural network considers a limited amount of historical information and is only suitable for short-term prediction, but it has a stable structure and features versatility and simplicity, which can flexibly deal with nonlinear problems, achieve high prediction accuracy, and has strong nonlinear mapping ability. As the hydrological sequence in hydrological forecasting is greatly affected by human factors and presents a prominent nonlinear characteristic, and the BP neural network is suitable for hydrological forecasting. Since BP neural network was proposed by Rumelhart et al. [30] in 1986, it has been widely used in research on hydrological prediction. In this paper, a BP neural network was established. We used historical data to predict the water level in the forebay of the pumping station and analyzed the influence of the time series proportion and impact factors on the water level prediction. The research results can provide a new method for water level prediction and reference data for the changing trend of the water level in the forebay of the pumping station.

# 1 Research method

The water level in the forebay of the pumping station is selected as the research object. The impact factors are determined by correlation analysis and are used as the input to construct the BP neural network model，and then the prediction results are judged by the parameters of each indicator.

## 1. 1 Impact factor identification

Under the influence of various hydraulic factors（section area，hydraulic gradient，roughness，etc.），there is a corresponding relationship between the section flow and the water level in the channel. As one of the monitoring sections，the water level in the forebay of the pumping station may have a hydraulic connection with the water level of the adjacent section，the flow of the pumping station，the upstream flow，and the flow difference. Taking these relevant water levels and flow as variables，we conduct a correlation analysis of each variable and the predictor，and the impact factors with a certain degree of correlation are identified.

The impact factor identification methods adopted include Pearson's correlation coefficient，Kendall's correlation coefficient，Spearman's rank correlation coefficient，and grey relational analysis （GRA）. Pearson's correlation coefficient is used to measure the degree of correlation between two variables，and Pearson's correlation coefficient between two variables is defined as the quotient of the covariance and standard deviation between the two variables. Kendall's correlation coefficient is a method to represent the degree of correlation of multi-column rank variables. If $n$ similar statistical objects are sorted by a specific attribute，other attributes are usually out of order，and the ratio of the difference between same-order pairs and out-of-order pairs to the total number of pairs$[n(n-1)/2]$is defined as Kendall's coefficient. Spearman's rank correlation coefficient is a method to study the correlation between two variables according to the rank data；in other words，it is calculated according to the rank difference between each pair of two-column paired ranks，and the monotone equation is used to evaluate the correlation of the two statisti-

cal variables. The range of the correlation indicator or correlation coefficient of the above three methods is from $-1$ to 1；When the absolute value of the correlation coefficient is closer to 1，the correlation is higher；when it is equal to zero，there is no correlation. GRA is a quantitative method for analyzing the correlation degree of each factor in the system，which measures the degree of correlation between factors according to the degree of similarity or dissimilarity in development trends among different variable sequences. When GRA is less than 0. 6，it is considered that there is no correlation，and when it is closer to 1，the correlation degree is higher.

## 1. 2 BP neural networks

A BP neural network is a multilayer feedforward neural network trained by an error backpropagation algorithm，generally including the input layer，hidden layer，and output layer. The input layer has the function of information access，i. e.，signal reception. When the signal reception is completed，the information is transmitted to the hidden layer，and the number of neurons in the input layer is the number $n$ of input impact factors. The hidden layer is responsible for information processing and information transformation，and the number of neurons in the hidden layer is $m$，which is less than $N-1$ （$N$ is the number of training samples），whose value is tested in MATLAB. Then，the information is transmitted from the hidden layer to the output layer，and the output layer outputs the results. The typical structure of a three-layer network is shown in Fig. 1.
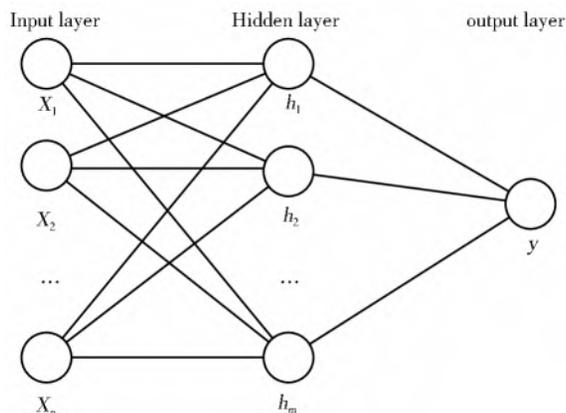


Fig. 1 BP neural network model structure

The neural network structure parameters are set as follows: maximum training times＝100; required accuracy of training＝$1×10^{-8}$; learning rate＝0.01. Upon the parameter setting, the network automatically adjusts the weights and thresholds by the back-propagation of errors, which drives the expression function in the BP neural network to obtain the optimal solution, and finally, it outputs the prediction results and the indicator values of the evaluation results.

## 1.3   Evaluation criteria of prediction results

$R^2$, $E_{RMS}$, and the mean absolute error ($E_{MA}$) are used as the evaluation criteria to judge the strengths and weaknesses of the prediction results. When $R^2$ is closer to 1, and $E_{RMS}$ and $E_{MA}$ are closer to zero, the prediction accuracy is higher.

## 2   Overview of study area

The Jiaodong Water Transfer Project is an important part of the water conservancy construction in Shandong Province, including two water transmission lines: the Yellow River Transfer Project and the Water Transfer Project from the Yellow River to Qingdao. The latter started on April 15, 1986, and it was officially put into operation on November 25, 1989; the Yellow River Water Transfer Project started on December 19, 2003, and the whole line was completed in July 2013, with the main project put into operation in December. The Yellow River Transfer Project includes two parts: the open channel section and the pipeline section. The open channel section starts from the Songzhuang Transfer Gate and terminates at the Huangshuihe Pumping Station, passing through three pumping stations in Huibu, Dongsong, and Xinzhuang, several inverted siphons, aqueducts, and other water transfer structures, with a total length of about 160 km. The study area selected in this paper is the open channel section of the Yellow River Water Transfer Project. Specifically, the study area is around the Dongsong Pumping Station, with the upstream control node as the Huibu Pumping Station and the downstream control node as the control gate on the port. The building distribution of this section and buildings along the line are shown in Fig. 2.
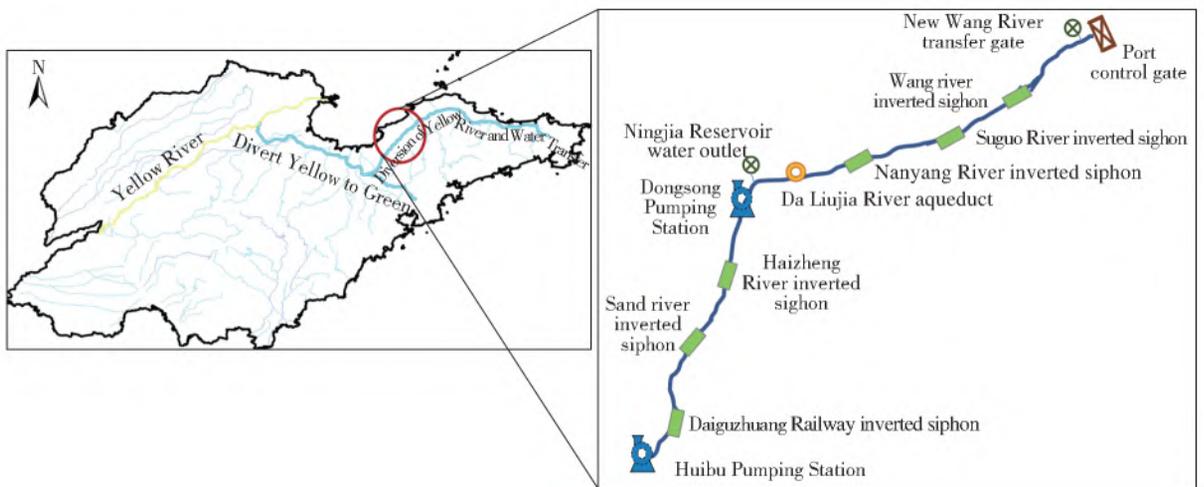


Fig. 2   Canal section and building along the distribution

## 3   Results and discussion

### 3.1   Identification results of impact factors

The relationship between the water level and flow rate and the influence of human factors were considered when studying the water level in the forebay of the Dongsong Pumping Station in the future. In addition to the water level of the adjacent section, the flow of the Dongsong Pumping Station, the flow of the Huibu Pumping Station, and the flow difference between the two pumping stations were also selected as the impact factors for prediction. The impact factors are all the impact factors at the current time. Tab. 1 shows the correlation analysis results between each factor and the water level of the forebay under different methods.

Tab. 1　Correlation analysis of impact factors

| Correlation indicators | Upstream water level of the Haizheng River inverted siphon | Downstream water level of the Haizheng River inverted siphon | Water level in front of Dongsong Pumping Station | Water level behind the Dongsong Pumping Station | Flow of pumping station | Flow difference between the two pumping stations | Upstream flow of pumping station |
|---|---|---|---|---|---|---|---|
| Pearson's correlation coefficient | 0.846 | 0.876 | 0.911 | −0.045 | 0.224 | 0.911 | −0.034 |
| Kendall's correlation coefficient | 0.648 | 0.642 | 0.747 | −0.067 | 0.233 | 0.747 | −0.067 |
| Spearman's rank correlation coefficient | 0.829 | 0.825 | 0.901 | −0.067 | 0.322 | 0.901 | −0.067 |
| GRA | 0.792 | 0.920 | 0.934 | 0.764 | 0.807 | 0.893 | 0.731 |

It can be seen from Tab. 1 that the order of the correlation of impact factors from high to low is the water level in front of the Dongsong Pumping Station, the flow difference of the two pumping stations, downstream water level of the Haizheng River inverted siphon, upstream water level of the Haizheng River inverted siphon, the flow of the pumping station, flow of the Dongsong Pumping Station, and upstream flow of the pumping station. The coefficients of the first four impact factors are all between 0.8 and 0.9, which are identified as impact factors with a high correlation and are given priority when modeling. Considering the indicators of the last three impact factors, only GRA indicates that the degree of correlation is high, and thus they are identified as impact factors with a low correlation, which can be considered in modeling but are not importantly considered.

## 3.2　Analysis of water level prediction results

The BP neural network model was used to predict the water level in the forebay of pumping stations, and the prediction results were analyzed from the aspects of time series and impact factors.

### 3.2.1　Time series

The data of different time scales were trained and verified according to a certain proportion, and the training duration and prediction accuracy were compared. The results indicate that the optimal ratio of the training period to the prediction period is 7∶3. Reducing the ratio will lessen the prediction accuracy, while increasing the ratio almost does not change the prediction accuracy, and the required data volume is significantly raised.

We used 3 600 data to predict the water level change in the next two hours, and $R^2$, $E_{\mathrm{RMS}}$, and $E_{\mathrm{MA}}$ at the ratio of 7∶3 were maintained at about 0.95, 0.04, and 0.03, respectively. When the ratio was increased, the prediction effect of each indicator was slightly improved, but the difference was not large; when the ratio was higher than 5∶1, the prediction accuracy basically would not see a rise. The specific comparison is shown in Fig. 3 and Fig. 4.
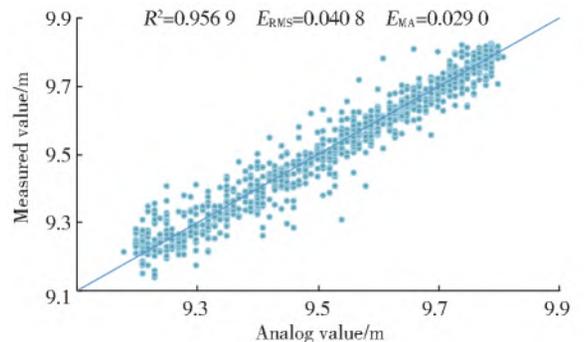


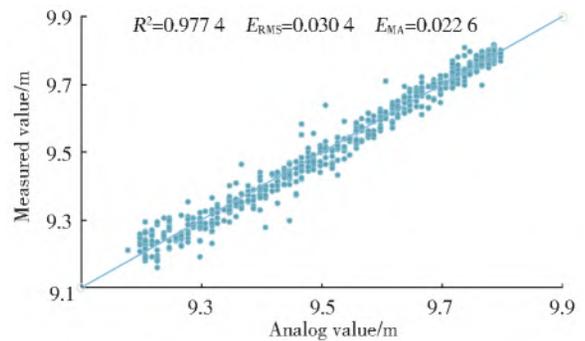Fig. 3　The result of water level forecast in the next 2 h (7∶3)



Fig. 4　The result of water level forecast in the next 2 h (5∶1)

Three groups of three-month data were trained and validated at a ratio of 7∶3. $R^2$ was maintained at 0.93-0.98, $E_{\mathrm{RMS}}$ at 0.02-0.05, and $E_{\mathrm{MA}}$ at 0.02-0.04. The prediction results are shown in Fig. 5.
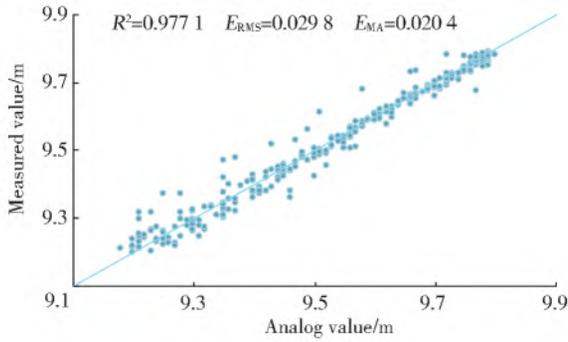
Fig. 5   The result of water forecast change in the next 2 h (7 : 3)

The data of one month was verified at a ratio of 7 : 3, and the verification results indicated that the ratio could still be applied to the amount of data of one month, as shown in Fig. 6.
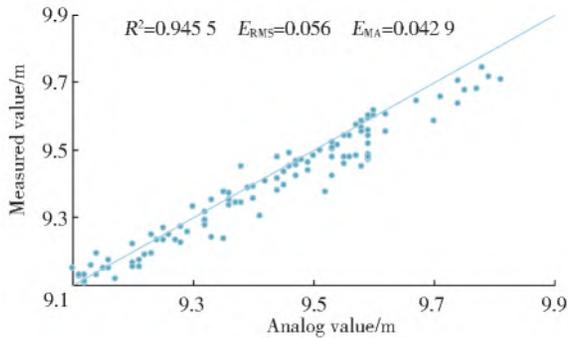


Fig. 6   The result of water level forecast in the next 2 h (7 : 3)

It can be seen from the above that the optimal ratio is suitable for data of different time scales, and the determination of the optimal ratio can not only save the learning time of the neural network but also improve the prediction accuracy, which has a great influence on the model.

### 3. 2. 2   Impact factors

The number of impact factors. When there is a high correlation between impact factors and predictors, a higher number of impact factors leads to more accurate prediction results. However, the increase in the number of impact factors can elevate the data demand during the training period. Therefore, to reduce the data demand and ensure prediction accuracy, we employed different numbers of impact factors for training and verification. The verification results revealed that at least three to five impact factors should be selected for training in the short term (one to three months), and at least five to seven impact factors with the greatest correlation were required for the data volume of

three months to a year.

Types of impact factors. Studies have shown that higher prediction accuracy can be achieved when the most relevant impact factors are selected for modeling. According to the correlation analysis results of the impact factors, the three impact factors with the highest correlation are the water level of the pumping station at the current moment, the water level of the upstream adjacent nodes, and the flow difference. Three impact factors were applied to train and predict data of one month, and the above three impact factors registered the best prediction effect, as shown in Fig. 7.
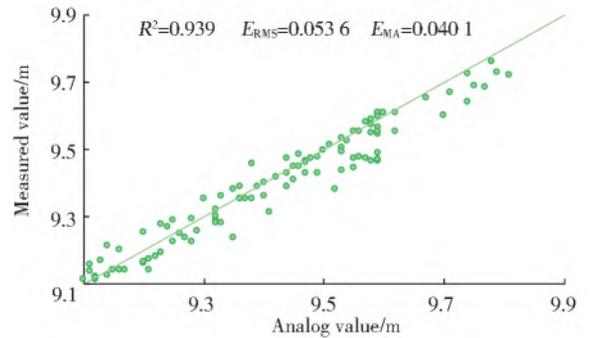


Fig. 7   3-factor water level prediction result map (7 : 3)

The time interval of the impact factors. When the data interval was 2 h, the water level of the Dongsong Pumping Station in the future was predicted: The water level prediction results in the next two hours were relatively stable, with $R^2$ greater than 0. 9 and small $E_{RMS}$ and $E_{MA}$; the prediction results of water levels in the next four hours were general, with $R^2$ of 0. 8-0. 9 and $E_{RMS}$ and $E_{MA}$ slightly larger than those predicted in two hours; the prediction results of the water level in the next six hours were poor; $R^2$ was unstable and had a large variation range, and it was only about 0. 7 when the results were good, while $E_{RMS}$ and $E_{MA}$ were overly great. In other words, when the data does not change during the training period, a longer prediction time is accompanied by lower prediction accuracy. The three-month data were screened to change the interval from 2 h to 4 h, and the water level of the Dongsong Pumping Station in the next 4 h was predicted. The prediction results are shown in Fig. 8.
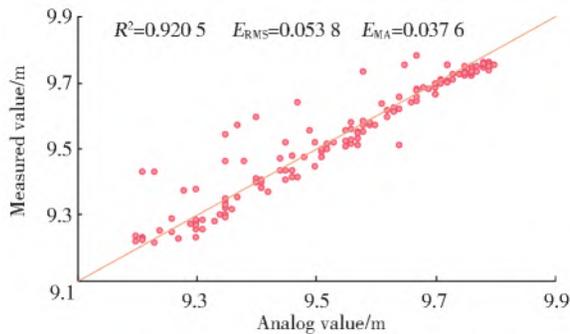
Fig. 8　The result of water level change in the next 4 h (7∶3)

The ten-month data were screened at an interval of 4 h，and the water level in the next 4 h was predicted. The prediction results are shown in Fig. 9.
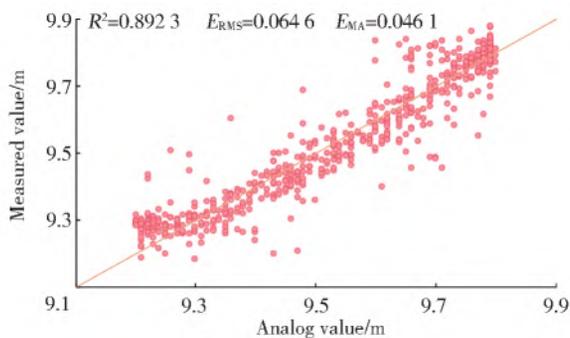


Fig. 9　The result of water level change in the next 4 h (7∶3)

The research results demonstrate that compared with the direct prediction using the data at an interval of 2 h，the prediction using the data at an interval of 4 h registers higher accuracy in predicting the water level of the pumping station in the next 4 h，with $R^2$，$E_{RMS}$，and $E_{MA}$ in the range of 0. 82-0. 93，0. 05-0. 06，and 0. 04-0. 05，respectively.

The one-year data were screened to convert the interval from 2 h to 6 h，and the water level of the Dongsong Pumping Station in the next 6 h was predicted. The prediction results show that the prediction effect after screening is worse than that of the direct prediction using data at an interval of 2 h. Upon analysis，the above phenomenon is caused by the overly long interval of 6 h，which can not fully reflect the changing laws of each factor. Therefore，the prediction result after screening is worse than that using the data at an interval of 2 h directly.

## 4　Conclusion

The influence of the time series ratio on the water level prediction results：The optimal ratio of the training period to the prediction period is 7∶3，and the increase in the ratio cannot significantly change the prediction accuracy，while the decrease in the ratio can lead to a worse prediction effect.

The effect of impact factors on the prediction results：The amount of data corresponds to the number of impact factors. The data volume of three months requires three to five impact factors for training，and the data volume of three months to a year requires five to seven impact factors to ensure the same prediction effect.

The influence of the data interval on the prediction results：In general，when the data interval remains unchanged，the prediction accuracy gradually decreases with the increase in the prediction time，but when the data can reflect the changing laws of each factor，the data interval and the prediction time are the same，and the prediction effect is better.

**References：**

[1]　ZHANG C，NI C F，LIU L. Research on interval scheduling under constant water level control in front of water delivery system gate[J]. Journal of Applied Basic and Engineering Sciences，2015，23(S1)：110-121. (in Chinese)) DOI：10. 16058/j. issn. 1005-0930. 2015. s1. 012.

[2]　ZHOU L C，CHENG Q，LI N. Operation optimization and simulation of cascade pumping stations in long channels in irrigation areas[J]. Journal of Irrigation and Drainage，2014，33(Z1)：390-393. (in Chinese)) DOI：10. 13522/j. cnki. ggps. 2014. 04/05. 084.

[3]　HUANG H Y，LIU Z H，FAN J，et al. Research on the hydraulic simulation model of the main canal of the Middle Route of the South-to-North Water Transfer Project[J]. Water Resources and Hydropower Technology，2013，44(12)：111-115，122. (in Chinese)) DOI：10. 13928/j. cnki. wrahe. 2013. 12. 008.

[4]　ZHENG H Z，MA X W，LEI X H，et al. Study on the calculation method of channel roughness of the Middle Route of the South-to-North Water Transfer Project [J]. South-to-North Water Transfers and Water Science ＆ Technology，2018，16(1)：158-163. (in Chinese)) DOI：10. 13476/j. cnki. nsbdqk. 20180024.

[5]　TANG M，LEI X H，LONG Y，et al. Water level prediction of the Middle Route of the South-to-North Water Transfer Project based on long short-term memory

Network (LSTM)[J]. China Rural Water and Hydropower,2020(10):189-193. (in Chinese)) DOI:10. 3969/j. issn. 1007-2284. 2020. 10. 034.

[6]　WANG X W,ZHAO B,GE T J,et al. Groundwater level prediction based on RBF neural network[J]. Agricultural Science and Technology and Information, 2018(4):60-61,71. (in Chinese)) DOI:10. 15979/j. cnki. cn62-1057/s. 2018. 04. 028.

[7]　YAN B Z,SUN J,WANG X Z,et al. Groundwater level prediction based on multivariate LSTM neural network[J]. Journal of Jilin University (Earth Science Edition),2020,50(1):208-216. (in Chinese)) DOI:10. 13278/j. cnki. jjuese. 20190055.

[8]　XU Q,SHU L C,YANG G L,et al. Application of wavelet neural network optimized by genetic algorithm in groundwater level prediction[J]. Hydrology,2010, 30(1):27-30. (in Chinese)) DOI:10. 3969/j. issn. 1000-0852. 2010. 01. 006.

[9]　QI H. Research on spatio-temporal prediction method of groundwater level in Changwu area based on KNN-LSTM model[D]. Nanjing:Nanjing Normal University,2020. (in Chinese))

[10]　ZHANG Y,LIAO Y F,WANG P P,et al. Prediction model for water level adjustment and storage of water transfer project based on correlation vector machine [J]. South-to-North Water Transfers and Water Science & Technology,2021,19(4):814-821. (in Chinese)) DOI:10. 13476/j. cnki. nsbdqk. 2021. 0085.

[11]　TU Z J,GAO X G,XU J,et al. A novel method for regional short-term forecasting of water level[J]. Water,2021,13(6),820. DOI:10. 3390/W13060820.

[12]　XU Y W,LI Z Y,WANG J Y. Application of similarity model based on set pair analysis in groundwater level prediction[J]. Hydrology,2015,35(6):6-10. (in Chinese)) DOI:10. 3969/j. issn. 1000-0852. 2015. 06. 002.

[13]　MITSUHIRO N,SHOICHI S,YUKI K,et al. Evaluation of real-time water level prediction technology using statistical models for reducing urban flood risk: Special issue on actions toward futuristic urban flood risk research and management[J]. Journal of Disaster Research,2021,16(3):387-394. DOI:10. 20965/JDR. 2021. P0387.

[14]　LIU Z Y,CHENG L Y,LIN K R,et al. A hybrid bayesian vine model for water level prediction[J]. Environmental Modelling & Software, 2021, 142, 105075. DOI:10. 1016/J. ENVSOFT. 2021. 105075.

[15]　LIU M Y,LI Q L,LI Z J,et al. Application of NARX neural network based on wavelet analysis in water level prediction[J]. South-to-North Water Transfers and Water Science & Technology,2019,17(5):56-63. (in Chinese)) DOI:10. 13476/j. cnki. nsbdqk. 2019. 0110.

[16]　XU G Y,ZHOU X Y,SI C Y,et al. Water level time series prediction model based on feature selection of GRU and LightGBM[J]. Computer Applications and Software,2020,37(2):25-31,53. (in Chinese)) DOI: 10. 3969/j. issn. 1000-386x. 2020. 02. 005.

[17]　GONG S,PENG H Y. Panjiakou reservoir water level prediction research based on PCA-ESN model[J]. Journal of Tangshan University,2020,33(3):37-41, 67. (in Chinese)) DOI:10. 16160/j. cnki. tsxyxb. 2020. 03. 008.

[18]　LIU Q S,YAN H,LU W L. Research on multivariate water level prediction model based on AR-RNN[J]. Yangtze River,2020,51(10):94-99. (in Chinese)) DOI: 10. 16232/j. cnki. 1001-4179. 2020. 10. 016.

[19]　WANG Y B,SUN T,LIANG X C,et al. Prediction of river water volume and level based on EMD-LSTM model[J]. Progress in Water Resources and Hydropower Science and Technology,2020,40(6):40-47. (in Chinese)) DOI:10. 3880/j. issn. 1006-7647. 2020. 06. 008.

[20]　REN T,LIU X F,NIU J W,et al. Real-time water level prediction of cascaded channels based on multilayer perception and recurrent neural network[J]. Journal of Hydrology, 2020, 585, 124783. DOI: 10. 1016/j. jhydrol. 2020. 124783.

[21]　GAN M,PAN S Q,CHEN Y P,et al. Application of the machine learning LightGBM model to the prediction of the water levels of the lower Columbia River [J]. Journal of Marine Science and Engineering,2021, 9(5),496. DOI:10. 3390/JMSE9050496.

[22]　LUO Y,DONG Z C,LIU Y H,et al. Research on stage-divided water level prediction technology of rivers-connected lake based on machine learning:A case study of Hongze Lake,China[J]. Stochastic Environmental Research and Risk Assessment,2021,35(10): 2049-2065. DOI:10. 1007/S00477-021-01974-6.

[23]　BAEK S S,PYO J C,CHUN J A. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach[J]. Water, 2020, 12 (12), 3399. DOI:10. 3390/W12123399.

[24]　CHEN Y P,GAN M,PAN S Q,et al. Application of auto-regressive (AR) analysis to improve short-term prediction of water levels in the Yangtze estuary[J]. Journal of Hydrology, 2020, 590, 125386. DOI: 10. 1016/J. JHYDROL. 2020. 125386.

[25]　EHTERAM M,FERDOWSI A,FARAMARZPOUR

M，et al. Hybridization of artificial intelligence models with nature inspired optimization algorithms for lake water level prediction and uncertainty analysis[J]. Alexandria Engineering Journal，2021，60（2）：2193-2208. DOI：10. 1016/J. AEJ. 2020. 12. 034.

[26] WU M L，YANG K，ZHANG C C. Application of Qinhuai River flood level prediction based on KG-BP neural network[J]. Hydropower Energy Science，2019，37（2）：74-77，81. （in Chinese）) DOI：CNKI：SUN：SDNY. 0. 2019-02-019.

[27] XIONG B，LI R P，REN D，et al. Prediction of flooding in the downstream of the Three Gorges Reservoir based on a back propagation neural network optimized using the AdaBoost algorithm[J]. Natural Hazards，2021，107（2）：1559-1575. DOI：10. 1007/S11069-021-04646-4.

[28] GAO X P，YAN C D，ZHANG Y，et al. Prediction model of storage water level of water transfer project based on BP neural network[J]. South-to-North Water Transfers and Water Science & Technology，2018，16（1）：8-13. （in Chinese）) DOI：10. 13476/j. cnki. nsbdqk. 20180002.

[29] GUO Y，LAI X J. Prediction of Poyang Lake water level based on long-and short-term memory neural network[J]. Lake Science，2020，32（3）：865-876. （in Chinese）) DOI：10. 18307/2020. 0325.

[30] RUMELHART D E，HINTON G E，WILLIAMS R J，et al. Learning representations by back-propagating errors[J]. Nature，1986（323）：533-536. DOI：10. 1038/323533a0.

推进地下水超采综合治理
复苏河湖生态环境