

三种水质动态预测模型在米山水库的应用与结果对比

黄林显¹, 张明芳², 钱永^{3,4*}, 邢学睿⁵, 邢立亭¹, 韩忠⁶

(1. 济南大学 水利与环境学院, 济南 250022; 2. 威海市水文中心, 山东 威海 264209;
3. 中国地质科学院 水文地质环境地质研究所, 石家庄 050061; 4. 河北省/地调局地下水污染
机理与修复重点实验室, 石家庄 050061; 5. 山东正元地质资源勘查有限责任公司,
济南 250101; 6. 山东省第六地质矿产勘查院, 山东 威海 264209)

摘要: 【目的】分析不同水质预测模型的预测精度, 探寻最优的水库水质预测方法。【方法】分别构建了季节性差分自回归滑动平均模型(SARIMA)、霍尔特-温特(Holt-Winters)模型和长短时记忆(LSTM)神经网络模型, 利用米山水库2012—2018年的月平均电导率观测数据对模型进行训练, 利用2019年月电导率实测数据对模型进行验证, 考察3种预测模型的准确性和稳定性。【结果】SARIMA模型和Holt-Winters模型仅能考察水质数据的时序演化趋势, 预测精度较低; 相比之下, LSTM神经网络模型能同时考察水质数据的时序演化趋势及不同时刻之间的前后依赖关系, 具有较强的非线性映射能力, 预测精度最高。【结论】LSTM神经网络预测模型仅在电导率值突变处误差相对较大, 但整体预测效果较为理想, 因此在水质预测中更加具有推广价值。

关键词: 时间序列模型; LSTM模型; 电导率; 水质预测; 米山水库

中图分类号: P641.2

文献标志码: A

doi: 10.13522/j.cnki.ggps.2022653

OSID:



黄林显, 张明芳, 钱永, 等. 三种水质动态预测模型在米山水库的应用与结果对比[J]. 灌溉排水学报, 2023, 42(11): 140-144.

HUANG Linxian, ZHANG Mingfang, QIAN Yong, et al. Comparison of Three Models for Predicting Water Quality in Mishan Reservoir[J]. Journal of Irrigation and Drainage, 2023, 42(11): 140-144.

0 引言

【研究意义】地表水质预测能够有效揭示水环境的演化趋势, 及时评估水体污染状况并制定相应治理措施, 对水生态环境保护和水资源管理具有重要意义^[1]。水质受到区外调水、降水、气温和水动力条件等诸多因素的影响, 其动态变化不仅具有稳定的周期性和季节性特征, 同时也存在较强的随机性。因此, 水质预测一直是水环境研究领域的难点之一^[2]。【研究进展】宫殿林等^[3]研究表明, 多元线性回归模型能够获得最佳的水质预测效果; 王泽斌等^[4]应用GM模型对阿什河入松花江口断面中的高锰酸盐指数和铵态氮浓度变化趋势进行了预测, 取得了较好的效果; 刘建霞等^[5]采用数值模拟法对青岛大沽河地下水水质进行了预测, 提出了相应的地下水污染防治措施; Najah等^[6]分别采用线性回归、多层感知器(MP)神经网络和径向基函数(RBF)神

经网络对水质进行了预测, 指出RBF神经网络具有较优的预测性能; 王梓璇等^[7]研究指出, 通过小波分解和遗传算法进行数据前处理能够有效提高模型的预测精度。上述研究虽然从不同角度对水质预测方法进行了探讨, 但存在一定局限性。线性回归和GM模型仅采用简单的线性或指数函数预测水质变化趋势, 无法体现水质动态的随机性; 数值模型法在获取模型相关参数过程中需要耗费大量的人力、物力和时间; RBF神经网络和BP神经网络等方法只考虑模型输入变量的影响, 无法反映数据的时序依赖性, 因此仅适用于短时间水质的动态预测^[8]。【切入点】时间序列模型(SARIMA模型和Holt-Winters模型)和LSTM神经网络模型由于具有建模过程简单、预测精度高等优势而被广泛应用于各个领域^[9-11]。此类模型主要通过对时间序列数据进行统计分析以找出其变化规律并揭示其未来演变趋势, 在水文领域已被广泛应用于降水^[12]、径流^[13]和地下水位^[14]的预测研究。然而, 此类模型尚未应用于具有强烈非平稳性和随机性特征水体的水质预测, 不同模型预测精度的比较研究甚少。【拟解决的关键问题】鉴于此, 以山东省威海市米山水库为研究区, 分别利用SARIMA模型、Holt-Winters模型和LSTM神经网络模型对区外调水影响下的米山水库的水质动态演化

收稿日期: 2022-11-25 修回日期: 2023-05-05 网络出版日期: 2023-11-10
基金项目: 国家自然科学基金项目(42272288); 山东省高校院所创新团队项目(2021GXRC070); 河北省/地调局地下水污染机理与修复重点实验室开发基金项目(SK202103KF01)
作者简介: 黄林显(1982-), 男, 副教授, 主要从事水资源及其环境问题研究。E-mail: stu_huanglx@ujn.edu.cn
通信作者: 钱永(1980-), 男, 高级工程师, 主要从事水文地质研究。E-mail: desertqy@163.com
©《灌溉排水学报》编辑部, 开放获取 CC BY-NC-ND 协议

规律进行预测，对比分析3种模型的建模过程与预测精度，评估不同模型在水质预测中的优劣，为水质监测管理及水生态环境保护提供技术支持。

1 材料与方法

1.1 研究区概况

米山水库位于山东省威海市文登区米山镇，是用于防洪、城镇供水、发电的大型水库，控制流域面积为440 km²，总库容为2.8亿m³，死库容为507万m³（图1）。米山水库作为胶东调水工程末端的调蓄水库，承担着威海市区和文登区的供水任务，其水质状况对于维持城市供水安全和区域水生态环境稳定具有重要意义。米山水库在实施区外调水后，其水质状况发生了显著变化，调水前（2015年）和调水后（2018年）的水库部分水质指标检测结果如表1所示。

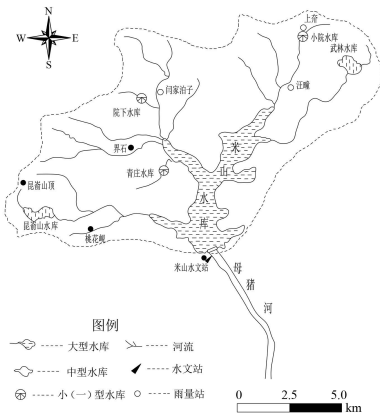


图1 米山水库及其控制流域分布

Fig.1 Watershed map of Mishan Reservoir

表1 米山水库2015年和2018年平均水质指标对比

Table 1 Comparison of annual average water quality indicators of Mishan Reservoir in 2015 and 2018

指标	2015年 (调水前)	2018年 (调水后)	变化率/%	变化趋势
电导率/($\mu\text{S cm}^{-1}$)	335	858	156	上升
钠/(mg L^{-1})	17.6	115	553	上升
硫酸盐/(mg L^{-1})	39.4	189	380	上升
氯化物/(mg L^{-1})	36.5	126	245	上升
总硬度/(mg L^{-1})	113	247	119	上升
重碳酸盐/(mg L^{-1})	81.6	164	101	上升
钙/(mg L^{-1})	26.2	45.5	74	上升
钾/(mg L^{-1})	5.15	8.86	72	上升

1.2 研究指标

电导率用于表征水体传导电流能力的强弱，通过电导率值可以推测水中带电荷物质的总浓度，可作为反映水体污染状况的关键指标^[15-16]。因此，本研究主要关注米山水库 2012—2019 年水体月电导率的动态变化。2013 年电导率数据缺失，采用样条插值进行插补。

由图 2 可知，米山水库电导率在 2016 年之前呈小幅度波动趋势，波动范围主要介于 200~400 $\mu\text{S/cm}$ 。2016 年开始实施区外调水后，电导率迅速上升，并在 2017 年达到最大，随后其动态变化不仅具有周期性和季节性特征，同时还存在较强的随机性，波动范围主要介于 500~1 600 $\mu\text{S/cm}$ 。

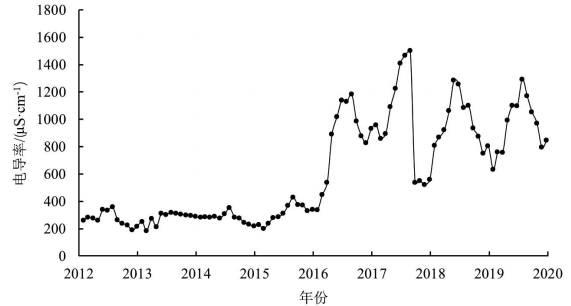


图2 2012—2019 年米山水库月平均电导率的变化趋势

Fig.2 Trend of monthly average conductivity of Mishan Reservoir from 2012 to 2019

1.3 研究方法

1.3.1 SARIMA模型

SARIMA模型是一种典型的预测非稳定时间序列的方法^[17-18]，通过拟合逼近观测值来建立数学关系，从而实现对时间序列的预测^[19-20]。SARIMA模型将水质观测值作为时间序列的一组随机数据，根据时间序列的特征，将其简化为自回归模型（AR）、移动平均模型（MA）、自回归移动平均模型（ARMA），反映水质的动态变化特征。

1.3.2 Holt-Winters模型

Holt-Winters模型同样是一种广泛应用的时间序列分析模型，主要利用指数平滑法使模型参数不断拟合非平稳序列，适用于带有趋势性和季节性时间序列的预测^[21-22]。该模型将水质观测值的时间序列分解为趋势项、季节项和随机项，并分别使用指数平滑法进行计算，最终实现对水质演变趋势的预测。

1.3.3 LSTM神经网络模型

LSTM神经网络模型能够解决长时间序列的传递问题^[23]。通过将水质训练样本输入到LSTM神经网络模型，不断调节优化权重矩阵和偏差，使水质模拟值与观测值不断接近，利用优化好的模型可实现对水质的动态预测。

1.4 模型评价指标

分别采用均方根误差（RMSE）和纳什效率系数（NSE）来评价3种水质预测模型的预测精度。其中，RMSE越小，预测精度越高；NSE越接近于1，预测值与观测值之间的一致性越强。

1.5 模型参数确定

采用米山水库 2012—2018 年的月平均电导率数

据作为训练样本对 3 种水质预测模型进行训练, 利用训练好的模型预测 2019 年的月电导率 (预测样本) 并与实际观测值进行对比, 以此验证 3 种预测模型的准确性。

通过对 SARIMA 模型进行反复调参、验证, 最终确定差分次数为 1, 滑动平均项数为 1, 季节性周期差分次数为 1, 季节性周期滑动平均项数为 1, 季节性周期长度为 12 时, 模型的拟合效果最优。Holt-Winters 模型经过反复调参、验证后, 最终确定采用加法模型时其拟合精度最高。LSTM 神经网络模型采用 4 层模型结构, 每层设置 80 个神经单元, 神经元失活率设置为 20%, 学习率设置为 0.001, 并利用各时刻前 4 个时刻的电导率值作为模型的输入数据。

2 结果与分析

2.1 训练样本结果分析

SARIMA 模型、Holt-Winters 模型、LSTM 神经网络模型的 $RMSE$ 分别为 127、121、105 $\mu\text{S}/\text{cm}$, NSE 分别为 0.88、0.87、0.91。LSTM 神经网络模型的 $RMSE$ 最小, 且 NSE 最接近于 1, 说明其与电导率观测值的拟合效果最佳。

3 种模型的拟合误差见图 3。SARIMA 模型与 Holt-Winters 模型的箱型图高度较为接近, 说明其拟合误差范围较为一致 (介于 3%~17%), 拟合误差的平均值分别为 11.87% 和 12.35%。与 SARIMA 模型相比, Holt-Winters 模型的拟合误差异常值更多 (最大异常值达到 58.6%), 说明 SARIMA 模型的拟合效果更加稳定。LSTM 神经网络模型的箱型图高度最低, 说明拟合误差范围更小 (介于 4%~15%), 其拟合误差的平均值在 3 种模型中也最小 (11.05%)。总体来说, LSTM 神经网络模型的拟合效果最优, 能够准确反映电导率的动态变化规律。

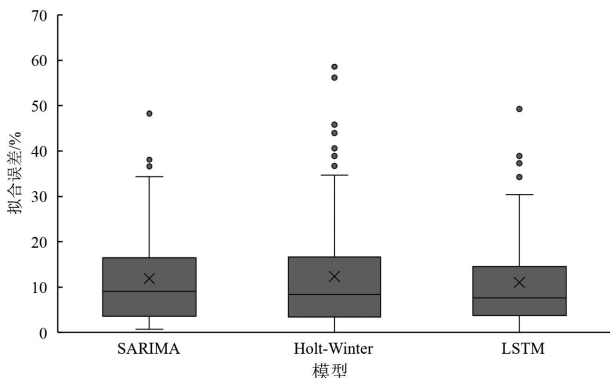


图3 模型训练拟合误差分布

Fig.3 Training fitting error box plots

2.2 预测样本结果分析

SARIMA 模型、Holt-Winters 模型、LSTM 神经网络模型的 $RMSE$ 分别 163、138、112 $\mu\text{S}/\text{cm}$, NSE

分别为 0.78、0.81、0.85, 可见 LSTM 神经网络模型的预测精度最高。此外, 预测样本的 $RMSE$ 普遍高于训练样本的 $RMSE$, 而 NSE 则更加接近于 1, 这主要是因为模型训练是一个参数反复调整优化的过程, 并会选择对训练样本拟合误差最优的参数作为模型输入参数, 因此预测精度较高; 在进行数据预测时, 最优参数会偏离预测样本, 导致预测精度略微降低。

由图 4 可知, SARIMA 模型的拟合误差范围较大 (介于 7%~23%), 同时其拟合误差平均值在 3 种模型中最大 (16.04%), 说明其预测精度较差。Holt-Winters 模型的拟合误差范围介于 6%~15%, 拟合误差平均值 (12.47%) 均介于其他 2 种模型之间, 说明其预测效果居中。LSTM 神经网络模型的拟合误差范围更小 (介于 6%~13%), 拟合误差平均值在 3 种模型中也最小 (9.95%), 说明 LSTM 神经网络模型的预测精度最高。

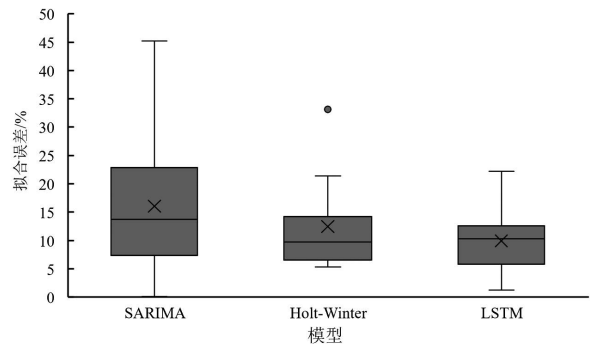


图4 模型预测误差分布

Fig.4 Prediction error box plots

3 种模型的电导率预测结果如图 5 所示。3 种模型的电导率预测曲线与观测曲线拟合效果较好, 均能够体现出电导率的季节性变化。其中, LSTM 神经网络模型的拟合度最高 (仅在水位突变处误差较大)。

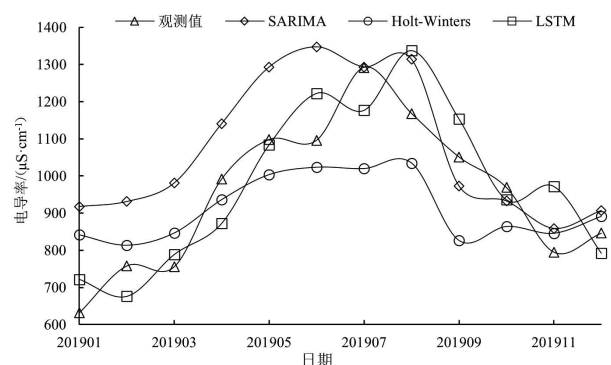


图5 不同模型电导率预测曲线

Fig.5 Conductivity prediction curves of three models

3 讨论

水质受人类活动、降水、气温和水动力条件等多种因素影响, 水质动态不仅体现出稳定的周期性

和季节性特征，同时还表现出复杂的随机性。为验证不同模型的水质预测性能，分别构建了 SARIMA、Holt-Winters、LSTM 神经网络水质预测模型，利用米山水库 2012—2018 年月平均电导率数据对模型进行训练，对 2019 年的月电导率数据进行预测，以此考察 3 种模型在水质预测中的准确性和稳定性。

SARIMA 模型和 Holt-Winters 模型的本质是通过拟合逼近时间序列观测值来模拟其规律性，而米山水库自 2016 年开始实施区外调水后，其受人为因素的影响不断加剧，导致电导率动态变化的规律性减弱，随机性增强，因此会一定程度上影响此 2 种模型的预测精度。

水体电导率的变化是一个连续过程，某一时刻电导率值往往受其前一时刻电导率值的高度影响，LSTM 神经网络模型将前 4 个时刻的电导率值作为输入数据，能同时考虑电导率的时序演化规律和不同时刻之间的前后依赖关系，因此预测精度最高；但由于区外调水后电导率动态的随机性增强，造成 LSTM 预测模型在电导率值突变处的模拟误差较大。

在未来研究中，建议采用集合经验模态分解或小波分析方法对电导率时序数据进行降噪处理，通过降低原始数据序列的非线性和非平稳波动特征，解决对变化细节和突变点预测精度不高的缺陷。此外，还可将影响电导率动态的主要因素（如降水、气温、水动力条件和人类活动）作为 LSTM 神经网络模型的输入数据，进一步提高模型的预测效果。长时间序列的水质监测数据往往存在大量的观测值缺失，利用 LSTM 神经网络模型能实现水质缺失值的准确插补，可有效提高数据挖掘方法的准确性，为水质评价提供科学依据。

4 结论

1) LSTM 神经网络模型的拟合误差范围更小，拟合误差的平均值最小，异常值较少，拟合效果最优。

2) LSTM 神经网络模型能同时考虑电导率时序演化规律及不同时刻之间的前后依赖关系，因此预测精度更高。

（作者声明本文无实际或潜在利益冲突）

参考文献：

[1] 张贻婷, 李天宏. 基于长短时记忆神经网络的河流水质预测研究[J]. 环境科学与技术, 2021, 44(8): 163-169.
ZHANG Yiting, LI Tianhong. River water quality prediction based on long short-term memory neural network[J]. Environmental Science & Technology, 2021, 44(8): 163-169.

[2] 程琳琳, 杨开宇, 杜鹃, 等. 河北省河流水质时空变化特征及受人类活动的影响分析[J]. 水资源与水工程学报, 2015, 26(1): 1-7.
CHENG Linlin, YANG Kaiyu, DU Juan, et al. Feature of spatial and temporal variation of water quality and influence of human activities on it in Hebei Province[J]. Journal of Water Resources and Water Engineering, 2015, 26(1): 1-7.

[3] 宫殿林, 洪曦, 曾冠军, 等. 亚热带典型农业流域河流水质多元线性回归预测[J]. 生态与农村环境学报, 2017, 33(6): 509-518.
GONG Dianlin, HONG Xi, ZENG Guanjun, et al. Prediction of water quality in rivers in agricultural regions typical of subtropics in China using multivariate linear regression model[J]. Journal of Ecology and Rural Environment, 2017, 33(6): 509-518.

[4] 王泽斌, 马云, 叶珍, 等. 应用 GM(1, 1) 模型预测阿什河水质变化趋势[J]. 环境科学与管理, 2011, 36(4): 24-27, 39.
WANG Zebin, MA Yun, YE Zhen, et al. Water quality prediction of Ashi river with GM(1,1) model[J]. Environmental Science and Management, 2011, 36(4): 24-27, 39.

[5] 刘建霞, 袁西龙. 青岛大沽河源地地下水水质的数值模型预测[J]. 海洋地质动态, 2006, 22(2): 9-14, 38.
LIU Jianxia, YUAN Xilong. Application of numerical modeling to prediction of water quality in Dagu River groundwater source area, Qingdao[J]. Marine Geology Letters, 2006, 22(2): 9-14, 38.

[6] NAJAH A, EL-SHAFFIE A, KARIM O A, et al. Application of artificial neural networks for water quality prediction[J]. Neural Computing and Applications, 2013, 22(1): 187-201.

[7] 王梓璇, 王圃, 王颖, 等. 基于小波分解及遗传 BPNN 耦合模型的地表水 As 浓度预测研究[J]. 环境科学学报, 2021, 41(7): 2 942-2 950.
WANG Zixuan, WANG Pu, WANG Ying, et al. Wavelet decomposition and genetic BPNN hybrid model based modelling approach for As concentration prediction in surface water[J]. Acta Scientiae Circumstantiae, 2021, 41(7): 2 942-2 950.

[8] WUNSCH A, LIESCH T, BRODA S. Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX)[J]. Hydrology and Earth System Sciences, 2021, 25(3): 1 671-1 687.

[9] HAMILTON J D. Time series analysis[M]. Princeton: Princeton University Press, 1994.

[10] LOPEZ BERNAL J, CUMMINS S, GASPARRINI A. The use of controls in interrupted time series studies of public health interventions[J]. International Journal of Epidemiology, 2018, 47(6): 2 082-2 093.

[11] 王宇, 卢文喜, 卞建民, 等. 三种地下水位动态预测模型在吉林西部的应用与对比[J]. 吉林大学学报(地球科学版), 2015, 45(3): 886-891.
WANG Yu, LU Wenxi, BIAN Jianmin, et al. Application and comparison of three dynamic models for groundwater in Western Jilin[J]. Journal of Jilin University (Earth Science Edition), 2015, 45(3): 886-891.

[12] 周生辉, 刘廷玺, 段利民, 等. 基于时间序列的海流兔河流域降水量模拟预测分析[J]. 水土保持研究, 2021, 28(5): 88-94.
ZHOU Shenghui, LIU Tingxi, DUAN Limin, et al. Simulation and prediction of precipitation in Hailiutu River Basin based on time series[J]. Research of Soil and Water Conservation, 2021, 28(5): 88-94.

[13] 雷昌宁. 基于 SARIMA 和 SVR 混合模型的黑河径流量预测分析[D]. 兰州: 兰州大学, 2018.
LEI Changning. The runoff forecasting of Heihe River based on SARIMA and SVR hybrid model[D]. Lanzhou: Lanzhou University, 2018.

[14] 张朝逢, 陈皓锐, 岳中奇. 基于长短时记忆网络(LSTM)的地下水埋深模拟预测: 以关中平原为例的实例分析[J]. 中国农村水利水电, 2020(9): 127-131, 137.

- ZHANG Zhaofeng, CHEN Haorui, YUE Zhongqi. Modelling groundwater table depth in Guanzhong, Shaanxi Province by long short-term memory (LSTM) based model term memory (LSTM) based model[J]. *China Rural Water and Hydropower*, 2020(9): 127-131, 137.
- [15] KNEY A D, BRANDES D. A graphical screening method for assessing stream water quality using specific conductivity and alkalinity data[J]. *Journal of Environmental Management*, 2007, 82(4): 519-528.
- [16] 姜伟, 周川, 纪道斌, 等. 三峡库区澎溪河与磨刀溪电导率等水质特征与水华的关系比较[J]. *环境科学*, 2017, 38(6): 2 326-2 335.
- JIANG Wei, ZHOU Chuan, JI Daobin, et al. Comparison of relationship between conduction and algal bloom in Pengxi River and Modao River in Three Gorges Reservoir[J]. *Environmental Science*, 2017, 38(6): 2 326-2 335.
- [17] 黄玥, 黄志霖, 肖文发, 等. 三峡水库水位调度对出库水质影响分析与水质预测[J]. *水资源与水工程学报*, 2020, 31(4): 78-85.
- HUANG Yue, HUANG Zhilin, XIAO Wenfa, et al. Analysis and prediction of effects of Three Gorges Reservoir water level scheduling on the outflow water quality[J]. *Journal of Water Resources and Water Engineering*, 2020, 31(4): 78-85.
- [18] VALIPOUR M. Long-term runoff study using SARIMA and ARIMA models in the United States[J]. *Meteorological Applications*, 2015, 22(3): 592-598.
- [19] DABRAL P P, MURRY M Z. Modelling and forecasting of rainfall time series using SARIMA[J]. *Environmental Processes*, 2017, 4(2): 399-419.
- [20] 徐瑞. 基于 SARIMA-LSTM 的北仑河口水质预测方法研究及应用[D]. 重庆: 重庆大学, 2019.
- XU Rui. Study and application of water quality prediction for Beilun estuary based on SARIMA-LSTM[D]. Chongqing: Chongqing University, 2019.
- [21] 袁嘉晨, 陈元芳, 顾圣华, 等. 基于 Holt-Winters 方法的年径流序列周期识别[J]. *水电能源科学*, 2016, 34(3): 28-31.
- YUAN Jiachen, CHEN Yuanfang, GU Shenghua, et al. Cycle identification of annual runoff time series based on Holt-Winters method[J]. *Water Resources and Power*, 2016, 34(3): 28-31.
- [22] 沙川. Holt-Winters 时间序列模型参数估计和预测[D]. 南京: 南京大学, 2011.
- SHA Chuan. Estimation and prediction for Holt-Winters time series models[D]. Nanjing: Nanjing University, 2011.
- [23] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: Continual prediction with LSTM[J]. *Neural Computation*, 2000, 12(10): 2 451-2 471.

Comparison of Three Models for Predicting Water Quality in Mishan Reservoir

HUANG Linxian¹, ZHANG Mingfang², QIAN Yong^{3,4*}, XING Xuerui⁵, XING Liting¹, HAN Zhong⁶

(1. School of Water Conservancy and Environment, University of Jinan, Jinan 250022, China; 2. Weihai Hydrological Center, Weihai 264209, China; 3. Institute of Hydrogeology and Environmental Geology, Chinese Academy of Geological Sciences, Shijiazhuang 050061, China; 4. Key Laboratory of Groundwater Remediation of Hebei Province and China Geological Survey, Shijiazhuang 050061, China; 5. Shandong Zhengyuan Geological Resources Exploration Co. Ltd., Jinan 250101, China; 6. No.6 Institution of Geology and Mineral Resources Exploration of Shandong Province, Weihai 264209, China)

Abstract: 【Objective】 Accurate prediction of water quality is of great importance for protecting water ecological environment and improving water resource management, but difficult due to the combined influence of various factors including human activities, rainfall, temperature and hydrodynamic conditions which are complex and uncertain. In this paper we compare three models, the SARIMA model, Holt-Winters model and LSTM neural network model, for predicting water quality of reservoir. 【Method】 Electric conductivity of water was used as a proxy for water quality. Data measured monthly from 2012—2018 from Mishan Reservoir was used for model training, and the data measured in 2019 was used to test the models. 【Result】 The SARIMA and Holt-Winters models are comparatively simple, but because they only consider time series of water quality data, their accuracy is low. In contrast, the LSTM neural network model considers factors that affect water quality and implicitly represents the nonlinearity of the factors in their impact on water quality, it is more accurate than other two models. 【Conclusion】 In general, the LSTM neural network model is reliably, giving rise to large error only when there were sudden changes in the conductivity. Overall, it is reliable and accurate for predicting change in water quality induced by variations in the environment.

Key words: time series model; LSTM model; conductivity; water quality prediction; Mishan Reservoir

责任编辑: 韩洋