

# 基于深度学习的地下水水质预测方法研究

## ——以大湾区(广州)地下水多层监测基地氨氮为例

戴蔓<sup>1</sup>,胡丹心<sup>1</sup>,李名升<sup>2</sup>,葛楹<sup>1</sup>,张隼<sup>1</sup>,黄霞<sup>1</sup>

1. 广东省广州生态环境监测中心站,广东 广州 510060

2. 中国环境监测总站,国家环境保护环境监测质量控制重点实验室,北京 100012

**摘要:**地下水水质预测可以准确反映水质的未来变化趋势,是地下水污染防治中的重要环节。为提高地下水水质自动监测的预测精度,以大湾区(广州)地下水多层监测基地2022年3层地下水的氨氮监测结果为例,进行数据预处理后,构建将集成经验模态分解(EEMD)与深度学习技术(RNN、LSTM及GRU)相结合的复合深度学习模型,开展传统机器学习模型与不同深度学习模型在水质预测效果方面的对比分析,并探讨深度学习模型多步长预测的预测效果。研究结果表明:①与单一深度学习模型相比,结合EEMD的复合深度学习模型解决了预测滞后性问题,具有更高的预测精度和拟合度。②复合深度学习模型的拟合度高于4种传统机器学习模型。4种传统机器学习模型中,仅MLR与RF的预测拟合度与单一深度学习模型接近。③多步长预测结果表明,复合深度学习模型可以准确地预测地下水水质在3d内的变化趋势。综上,复合深度学习模型展现出更好的预测性能和泛化能力,可为地下水水质预测提供支撑。

**关键词:**地下水水质预测;集成经验模态分解;深度学习

**中图分类号:**X832 **文献标志码:**A **文章编号:**1002-6002(2025)01-0191-10

**DOI:**10.19316/j.issn.1002-6002.2025.01.20

### Research on Groundwater Quality Prediction Method Based on Deep Learning: The Case of Multi-layer Groundwater Monitoring Base, Guangzhou, Greater Bay Area

DAI Wan<sup>1</sup>, HU Danxin<sup>1</sup>, LI Mingsheng<sup>2</sup>, GE Ying<sup>1</sup>, ZHANG Qu<sup>1</sup>, HUANG Xia<sup>1</sup>

1. Guangzhou Sub-branch of Guangdong Ecological and Environmental Monitoring Center, Guangzhou 510060, China

2. State Environmental Protection Key Laboratory of Quality Control in Environmental Monitoring, China National Environmental Monitoring Centre, Beijing 100012, China

**Abstract:** Groundwater quality prediction can reflect the future trend of groundwater quality changes, and is an important link in groundwater pollution prevention and control. To improve the accuracy of water quality prediction in automatic groundwater monitoring, the ammonia nitrogen levels of three layers of groundwater monitored at the multi-layer monitoring base in the Greater Bay Area (Guangzhou) from March to December 2022 were selected, a composite deep learning model combining integrated empirical mode decomposition (EEMD) and deep learning techniques (RNN, LSTM, and GRU) was constructed after data preprocessing. The effect of traditional machine learning model and different deep learning models in water quality prediction were compared and analyzed, and the prediction effect of multi-step prediction of deep learning model was discussed. The results showed that: ① Compared with a single deep learning model, the composite deep learning model solves the problem of prediction lag and has higher prediction accuracy and fitness. ② Fitness of the composite deep learning model is higher than that of the four traditional machine learning models. In the four traditional machine learning models, only the fitness of MLR and RF predictions is close to that of the single deep learning prediction. ③ The multi-step prediction results indicated that the composite deep learning model can accurately predict the change trend of groundwater quality within 3 days. The composite deep learning model exhibits better predictive performance and generalization ability, which can provide support for groundwater quality prediction.

**Keywords:** groundwater quality prediction; ensemble empirical mode decomposition; deep learning

收稿日期:2024-02-06;修订日期:2024-04-23

基金项目:国家重点研发计划项目(2022YFC3700905)

第一作者简介:戴蔓(1995-),男,湖南邵阳人,硕士,工程师。

通讯作者:李名升

水资源是人们日常生活中至关重要的资源,地下水是人类饮用水的重要组成部分。近年来,工业化和城市化的高速发展导致地下水污染问题愈加凸显,对人体健康造成极大潜在威胁<sup>[1-2]</sup>。《2022中国生态环境状况公报》显示,1 890个国家地下水环境质量考核点位中,V类水质点位占比为22.4%,表明当前我国地下水污染形势非常严峻<sup>[3]</sup>。因此,精准预测地下水水质未来趋势,并对可能发生的污染进行预报预警,有助于及早发现地下水环境中存在的问题,对加强地下水污染防治工作有着十分重要的作用<sup>[4]</sup>。

随着机器学习技术的迅猛发展<sup>[5-7]</sup>,多元线性回归<sup>[5]</sup>(Multiple Linear Regression, MLR)、支持向量机<sup>[8]</sup>(Support Vector Machine, SVM)、随机森林<sup>[9]</sup>(Random Forest, RF)、决策树<sup>[5]</sup>(Decision Tree, DT)等机器学习模型为水质预测提供了全新的思路。除上述传统机器学习模型外,近年来,深度学习模型作为机器学习研究的重要组成部分,在水资源研究中受到越来越多的关注<sup>[10]</sup>。深度学习,也称为深度神经网络(Deep Neural Network, DNN),在计算机、智能机器人等诸多研究领域已经得到广泛应用。深度神经网络的特点是通过多层人工神经元来模拟人类大脑的工作方式,使其可以学习到复杂的特征和规律<sup>[11-12]</sup>。CHEN等<sup>[5]</sup>构建了基于2012—2018年中国主要江河湖泊数据的10种机器学习模型,并进行了比较,发现DT、RF和深度级联森林(DCF)等集成模型的水质等级预测效果明显更好。FARUK<sup>[13]</sup>运用传统时间序列方法和反向传播(Back Propagation, BP)人工神经网络构建了混合水质预测模型,发现经模拟训练后的模型的预测效果良好。肖焱等<sup>[9]</sup>利用BP神经网络、RF和SVM等机器学习方法构建了地下水水质预测模型,发现BP神经网络具有良好的精准度。但是,BP神经网络在样本数据时间跨度较大的情况下,可能出现误差偏大的情况<sup>[14]</sup>。由此,循环神经网络(Recurrent Neural Network, RNN)得以提出。RNN是一种非常强大的用于处理和预测时间序列数据的神经网络模型<sup>[15]</sup>。当数据存在一定的时间依赖性时,RNN具有很好的应用价值<sup>[16]</sup>。然而,该方法在应用过程中无法避免梯度消失及梯度爆炸问题的出现。因此,HOCHREITER等<sup>[17]</sup>和CHUNG等<sup>[18]</sup>分别提出了长短期记忆(Long-Short Term Memory, LSTM)神经网络和门控循环(Gate

Recurrent Unit, GRU)神经网络。两者可以被看作是RNN的变种,既继承了深度学习的优势,又具有较强的时间序列数据分析能力<sup>[16]</sup>。有研究直接利用水质时间序列数据构建LSTM神经网络,实现了对水质指标的预报<sup>[19-20]</sup>。VALADKHAN等<sup>[21]</sup>提出了一种基于LSTM、RNN的地下水水质指标预测方法,通过对实际值与预测结果进行比较,发现在大多数情况下,其水质指标预测结果是合理的。但用单特征时间序列数据开展预测会存在滞后性问题,预测值会滞后于真实值。这是由于时间序列数据存在突变,直接使用深度学习模型并不能准确探测到这种突变,而是在下一时刻的预测结果中才会体现出来。集成经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)的提出从根本上解决了这一问题。其本质是将原本的时间序列数据分解为若干条子序列,这些子序列包含了原时间序列数据中的不同时间段的局部特征信息,使得时间序列中的细节得到放大,从而解决了滞后的问题<sup>[22-24]</sup>。目前,EEMD在地下水自动监测数据分析中尚未得到应用。

本文以大湾区(广州)地下水多层监测基地的氨氮在线监测结果为例,先对时间序列数据进行预处理,然后利用EEMD将水质时间序列数据分解为若干个分量,分别输入到RNN、LSTM与GRU水质预测模型中,构建新的复合水质预测模型。同时,构建多种传统机器学习模型(MLR、SVM、DT、RF)与单一深度学习预测模型,探讨传统机器学习模型与不同深度学习模型的预测准确性和可靠性。此外,对深度学习多步长预测结果进行研究,进一步验证不同深度学习模型的预测性能和泛化能力。

## 1 研究区与研究数据

### 1.1 研究区概况

大湾区(广州)地下水多层监测基地位于广州市南部的海陆交互相冲积平原,为地下水自动监测站点。其附近河网密布,水系发育较好。区域气候为亚热带季风气候,降雨集中在夏季,年均降雨量为1 640 mm,全年平均气温为22℃。该基地共监测3个地下含水层。第1层地下水为第四系松散岩类孔隙水,含水层位于6.00~8.50 m深度,岩性为灰褐色、灰黑色淤泥质粉细砂。第2层地下水属于第四系松散岩类孔隙水,含水层位

于 17.30~19.00 m 深度,岩性为灰白色、灰色粉细砂。松散岩类孔隙水的主要补给来源为大气降水入渗补给,以及部分受潮汐影响河流的侧向补给,还有部分农田水塘水入渗补给。第 3 层地下水属于基岩裂隙水,含水层位于 38.16~76.00 m 深度,岩性为砂砾岩和泥质粉砂岩互层。基岩裂隙水的主要补给来源为大气降水,上覆松散岩类在裂隙强烈发育段的侧向径流也是其补给来源之一。

## 1.2 数据来源

水质数据采集时间为 2022 年 1—12 月,监测频次为每天一次,监测指标为氨氮。

表 1 ADF 检验结果  
Table 1 ADF test results

水质指标	层位	ADF 检验值	P 值	1% 临界值	5% 临界值	10% 临界值
氨氮	第 1 层	-3.977 8	0.001 5	-3.450 0	-2.870 2	-2.571 4
	第 2 层	-3.404 6	0.010 8	-3.452 0	-2.871 1	-2.571 9
	第 3 层	-4.362 0	0.000 3	-3.454 0	-2.872 0	-2.572 3

通过平稳性甄别可知,氨氮数据可以用于构建长时间序列水质模型。氨氮指标的描述性统计分析结果如表 2 所示:第 1 层和第 2 层地下水的

## 1.3 数据预处理

### 1.3.1 数据平稳性分析

利用深度学习技术对地下水水质进行预测时,首先要对时间序列数据进行平稳性检验。采用单位根检验方法(ADF 检验)判断时间序列数据是否平稳。在实际应用时,会先计算出统计检验中的  $P$  值。如果  $P$  值小于等于阈值(0.05),表明数据平稳;若高于阈值,则表明数据不平稳。对上述 3 层地下水的氨氮数据进行检验,结果如表 1 所示。检验结果显示, $P$  值均小于 0.05,表明相关数据均为平稳的时间序列数据。

氨氮含量平均值分别为 9.76、5.21 mg/L,水质类别为 V 类;第 3 层地下水的氨氮含量平均值为 0.81 mg/L,水质类别为 IV 类。

表 2 不同层位描述性统计分析  
Table 2 Descriptive statistical analysis for different layers

监测指标	层位	均值/(mg/L)	最小值/(mg/L)	最大值/(mg/L)	标准差/(mg/L)	偏度	峰度
氨氮	第 1 层	9.76	5.04	14.31	2.05	-0.22	-0.69
	第 2 层	5.21	0.61	8.87	1.43	-0.24	0.06
	第 3 层	0.81	0.53	1.02	0.09	-0.29	0.08

### 1.3.2 空缺数据处理

针对空缺数据,使用线性插值进行填补。线性插值法可用于对时间序列空缺数据进行补充,其公式如下<sup>[25]</sup>:

$$x_k = x_f + (x_b - x_f)(k - f)/(b - k) \quad (1)$$

式中: $x_k$  为需填补的值, $x_f$  为前一时刻的值, $x_b$  为后一时刻的值。

### 1.3.3 数据归一化

为降低模型陷入局部最优的可能性,在进行模型训练前,将所有数据进行归一化处理,公式如下<sup>[26]</sup>:

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

式中: $x$ 、 $y$  分别为归一化前、后的值, $x_{\min}$ 、 $x_{\max}$  分别为归一化前数据的最小值、最大值。

### 1.3.4 训练数据与滑动时间窗口

将数据以 9:1 的比例划分为训练数据和测试数据,前 90% 的水质监测数据用于模型训练,后 10% 的水质监测数据用于模型验证。

同时,以  $m$  个时序长度创造若干个滑动时间窗口,通过滑动窗口法实现对下一时刻水质指标的预测。循环获取时间序列数据的方法如图 1 所示。首先,假设第 1 个滑动时间窗口为  $m$ ,用  $m$  个数据去预测第  $m+1$  个结果,并将预测值与实际值进行比对。然后,向后滑动 1 格,直到所有数据均被取出<sup>[27-28]</sup>。

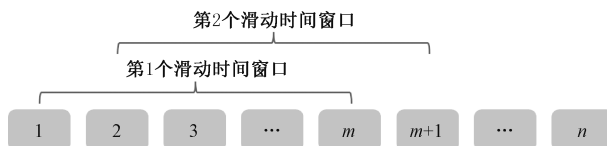


图 1 滑动时间窗口原理

Fig. 1 The principle of sliding time window

## 2 研究方法

### 2.1 EEMD

EEMD 是对经验模态分解 (Empirical Mode Decomposition, EMD) 的改进。EMD 算法常用于分解非线性不平稳数据,其本质是对时间序列数据进行平稳化处理。该算法将原始数据分解为不同频域下的若干个本征模函数 (Intrinsic Mode Function, IMF) 分量,每个 IMF 分量都表征了原始数据的不同时段局部特征。分解结果如下:

$$x(t) = \sum_{i=1}^n \text{IMF}_i(t) + r(t) \quad (3)$$

式中:  $t$  为时间,  $x(t)$  为原始数据,  $r(t)$  为余量。

EMD 算法具有正交性和收敛性,在信号处理领域应用广泛。然而,在 EMD 算法中,单个 IMF 往往会囊括特征时间尺度差异极为显著的成分,进而导致频率分布呈现出较大的离散性。为了改善 EMD 的分解效果,研究人员在 EMD 算法的基础上提出了一种噪声辅助分析方法,即 EEMD 算法<sup>[24]</sup>。EEMD 算法通过多次向原始信号中加入随机白噪声序列,并计算对应分量的平均值,得到相应的分量。其步骤如下:①假设总体分解次数为  $M$ ;②在原时间序列数据中加一个符合标准正态分布的白噪声序列,从而产生一个新的序列;③对新序列进行 EMD 处理;④重复步骤 2、3 各  $M$  次,得到不同的 IMF;⑤将每次得到的 IMF 进行集合平均运算,得到最终的 EEMD 结果。

### 2.2 RNN

RNN 通常用于处理序列数据。将 RNN 按时间序列展开,其结构如图 2 所示。RNN 基本由三部分组成:输入层、隐藏层和输出层。具体的传播过程如公式(4)和公式(5)所示:

$$h_t = \sigma(U \times x_t + V \times h_{t-1}) \quad (4)$$

$$o_t = \sigma(W \times h_t) \quad (5)$$

式中:  $h_t$  为  $t$  时刻的隐藏层状态,  $\sigma$  为 sigmoid 激活函数,  $x_t, o_t$  分别为  $t$  时刻网络的输入值、输出值,  $U, V, W$  分别为此神经元输入的权重、上一个神经元保存的权重、上一个神经元输出的样本权重。由公式(4)和公式(5)可知,  $h_t$  保存了  $t-1$  时刻的信息状态。因此, RNN 隐藏层的神经元具有“记忆能力”,可实现时序信息传递。

### 2.3 LSTM

与 RNN 相比, LSTM 的创新之处在于其在隐

藏层中增加了记忆单元,运用遗忘门、输入门和输出门等门结构进行数据储存和更新。一个标准的 LSTM 记忆单元如图 3 所示。该记忆单元替换了 RNN 隐藏层中的神经元,可实现对时序信息的长期保留。

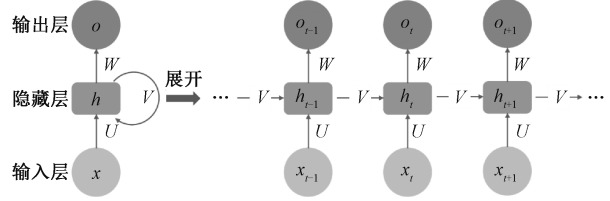


图 2 RNN 结构

Fig. 2 RNN structure

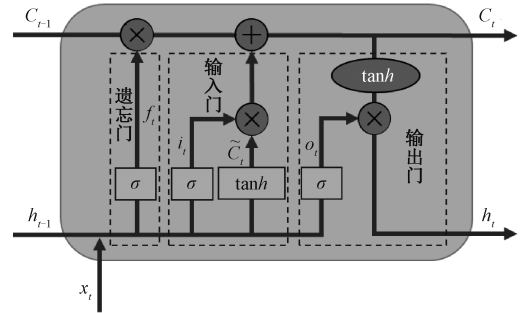


图 3 LSTM 单元结构

Fig. 3 LSTM unit structure

遗忘门  $f_t$  决定了什么样的信息会被从单元状态中丢弃。

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (6)$$

式中:  $W_f$  和  $b_f$  分别为遗忘门的权重和偏置。

输入门  $i_t$  决定了什么样的新信息能被存放在单元状态中。将遗忘门和输入门结合,即可以实现单元状态更新。

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (7)$$

$$\tilde{C}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c) \quad (8)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (9)$$

式中:  $W_i$  和  $b_i$  分别为输入门的权重和偏置,  $\tilde{C}_t$  为新信息,  $C_t$  为  $t$  时刻的单元状态,  $W_c$  和  $b_c$  分别为待更新单元的权重和偏置。

输出门控制着单元状态值 ( $h_t$ ) 的输出。最终输出的  $h_t$  将被输入到下一时刻的 LSTM 单元中,其值由当前的单元状态  $C_t$  和中间输出  $o_t$  共同决定。

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (10)$$

$$h_t = o_t \times \tanh(C_t) \quad (11)$$

式中: $W_o$ 和 $b_o$ 分别为输出门的权重和偏置。

## 2.4 GRU

与 LSTM 相比,GRU 的体系架构更为简化。一个标准的 GRU 记忆单元如图 4 所示。在数据量较小的情况下,GRU 结构能更快地收敛,减少过拟合风险,从而展现出比 LSTM 更好的性能。其变化是用更新门 $z_t$ 和重置门 $r_t$ 替代了 LSTM 单元中的门结构。两个门结构分别控制着信息的通过比例,并将单元状态与输出合二为一。

$$z_t = \sigma(W_z \times [h_{t-1}, x_t] + b_z) \quad (12)$$

$$r_t = \sigma(W_r \times [h_{t-1}, x_t] + b_r) \quad (13)$$

$$g_t = \tanh(W_g \times [r_t \times h_{t-1}, x_t] + b_g) \quad (14)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times g_t \quad (15)$$

式中: $W_r$ 和 $b_r$ 分别为重置门的权重和偏置, $W_z$ 和 $b_z$ 分别为更新门的权重和偏置, $g_t$ 为 $t$ 时刻的候选隐藏信息, $W_g$ 和 $b_g$ 分别为候选隐藏信息的权重和偏置。

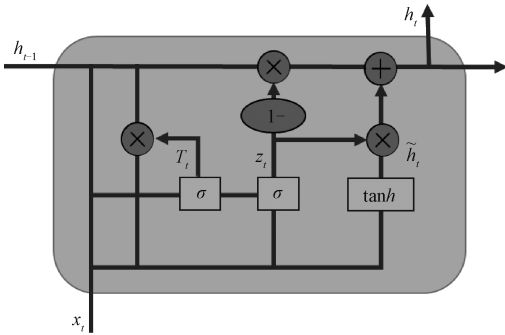


图 4 GRU 单元结构

Fig. 4 GRU unit structure

## 2.5 模型预测效果评价

地下水水质预测模型本质上也是一种回归模型,因此,选用平均绝对误差(MAE)、均方根误差(RMSE)和决定系数( $R^2$ )3个指标来评估模型的预测效果。MAE和RMSE数值越小,意味着模型预测效果越好。 $R^2$ 越接近于1,表示模型拟合能力越强。当 $R^2$ 介于0.75~1之间时,可认为模型拟合效果很好;介于0.65~0.75之间时,可认为模型拟合程度良好;介于0.5~0.65之间时,可认为模型拟合效果令人满意;介于0.4~0.5之间时,可认为模型拟合程度可以接受;低于0.4时,则认为模型不能很好地拟合<sup>[29-32]</sup>。

$$MAE = \frac{1}{N} \sum_{i=1}^N |y'_i - y_i| \quad (16)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y'_i - y_i)^2} \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - y'_i)^2}{\sum_{i=1}^N (\bar{y}_i - y'_i)^2} \quad (18)$$

式中: $y_i$ 为真实值, $y'_i$ 为预测值, $\bar{y}_i$ 为真实值的平均值, $N$ 为被测样本总数。

## 3 模型构建和验证

### 3.1 模型构建环境

研究使用的程序语言为python 3.11,开发环境管理工具是Eclipse,操作系统为Windows 10(64位),CPU配置为Intel(R)Core(TM)i7-8750H,GPU配置为NVIDIA Geforce GTX 1060。RNN、LSTM及GRU地下水水质预测模型基于深度学习库keras 2.12.0(Tensorflow后端)来实现训练。

### 3.2 模型结构与参数

#### 3.2.1 模型构建

模型结构除了包含固定的输入层和输出层,还包含了一层由50个神经元组成的隐藏层。

完成数据预处理后,先对水质数据进行EEMD处理。设置的高斯白噪声标准差为0.05,EEMD次数为100。以第1层地下水的氨氮数据为例,经过EEMD处理后的结果如图5所示。

第1层地下水氨氮指标的时间序列数据通过EEMD处理后,得到8个IMF分量。在此基础上,对分解获得的8个IMF分量分别构建相应的深度学习预测模型,并将所有分量的水质预测值相加求和,获得最终的水质指标预测结果。具体流程如图6所示。

#### 3.2.2 模型参数设置

经过综合考虑和多次试验,将迭代次数设置为100,批量大小设置为16,滑动时间窗口设置为15,并在训练集上使用Adam优化器。

### 3.3 单一与复合深度学习模型预测

分别将3层地下水的测试数据集用未使用EEMD的RNN、LSTM、GRU单一深度学习模型与使用EEMD的EEMD-RNN、EEMD-LSTM、EEMD-GRU复合深度学习模型进行运算,图7、图8、图9分别展示了3层地下水的模型运算结果。

对比结果表明,RNN、LSTM及GRU单一深度学习模型能预测水质时序数据的走势,但存在一定的滞后性,滞后时间为1d;EEMD-RNN、EEMD-LSTM及EEMD-GRU复合深度学习模型预测结果的拟合

程度较高。可以直观地看出,通过 EEMD 处理,模型的预测结果曲线更贴合实际曲线。这说明复合深度

学习模型不仅解决了滞后性问题,在水质时序数据发生突变的情况下,也能给出误差更小的预测数据。

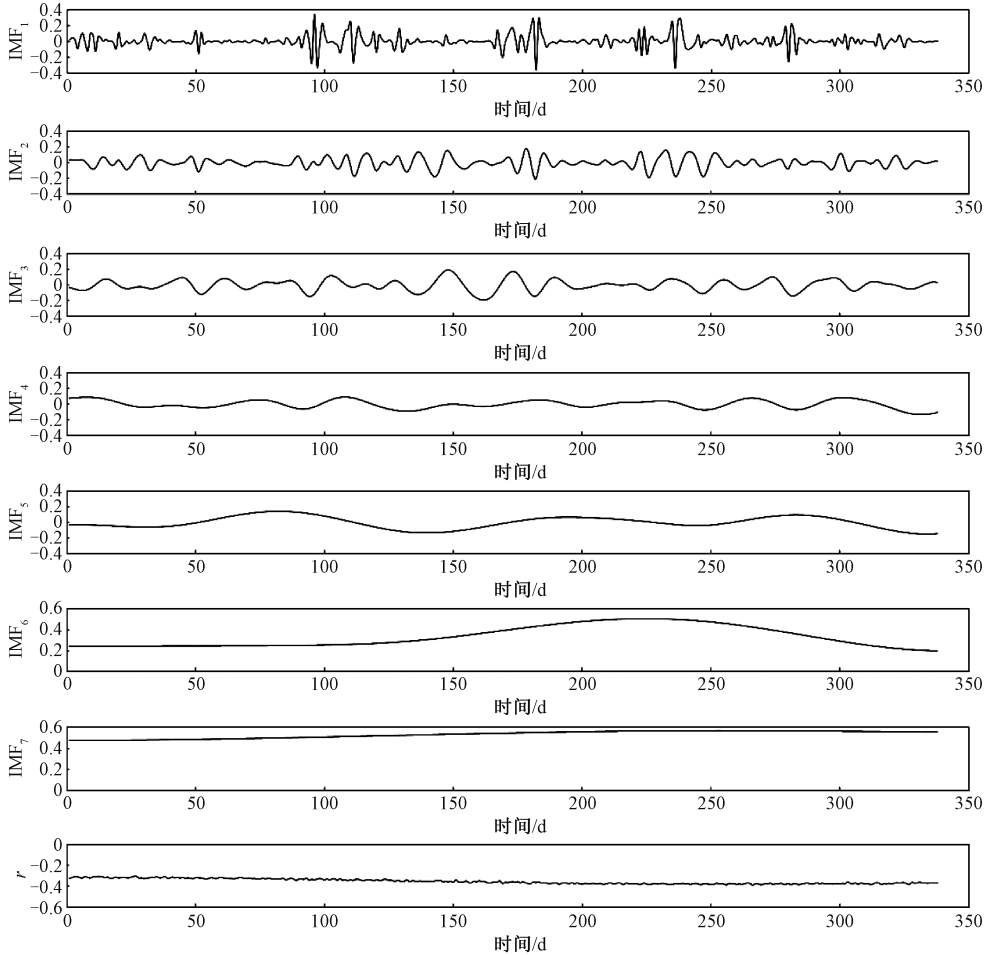


图5 第1层地下水 EEMD 结果

Fig. 5 The first layer of groundwater EEMD results

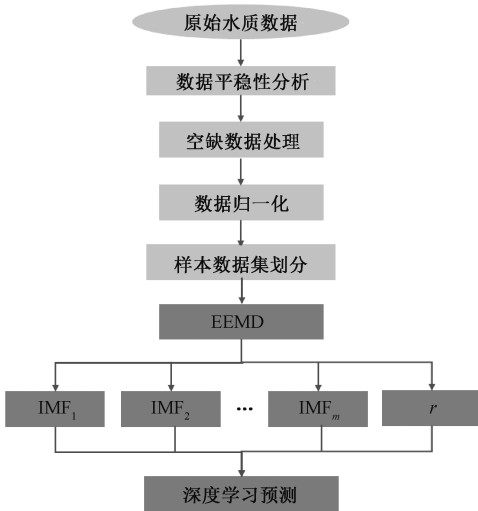


图6 EEMD 复合深度学习模型流程

Fig. 6 EEMD composite deep learning model flow

各层位的水质预测效果评价结果(表3)显示,结合 EEMD 后的深度学习模型的预测误差均有不同程度的降低。EEMD-RNN、EEMD-LSTM、EEMD-GRU 模型第1层地下水预测结果的 RMSE 比对应的单一模型分别降低了 63.1%、68.1%、57.7%,MAE 分别降低了 65.7%、72.3%、55.2%,拟合度分别达到了 0.939、0.957、0.955 ( $R^2$  大于 0.9);第2层地下水预测结果的 RMSE 比对应的单一模型分别降低了 60.7%、65.8%、60.4%,MAE 分别降低了 63.9%、66.0%、59.5%,拟合度分别为 0.683、0.919、0.932 ( $R^2$  介于 0.6~0.9 之间),其中 EEMD-LSTM 及 EEMD-GRU 的拟合度相对更高;第3层地下水预测结果的 RMSE 比对应的单一模型分别降低了 59.2%、56.3%、48.8%,MAE 分别降低了 55.6%、56.9%、45.5%,拟合度分别为 0.821、0.842、0.789 ( $R^2$  介于 0.6~0.9 之间)。

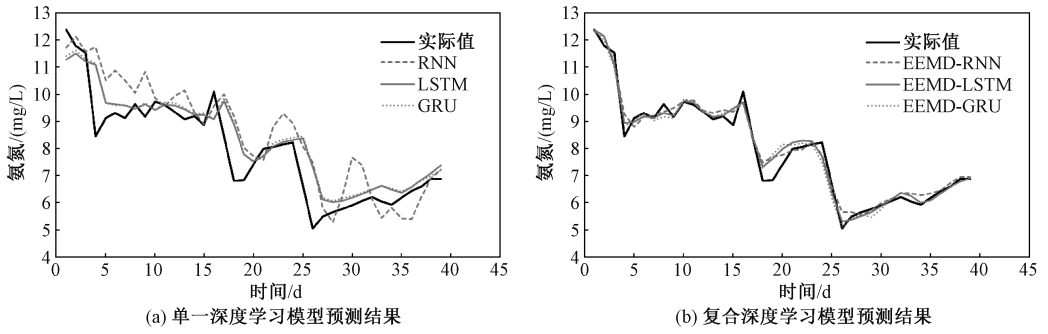


图7 第1层地下水预测结果

Fig. 7 The first layer of groundwater prediction results

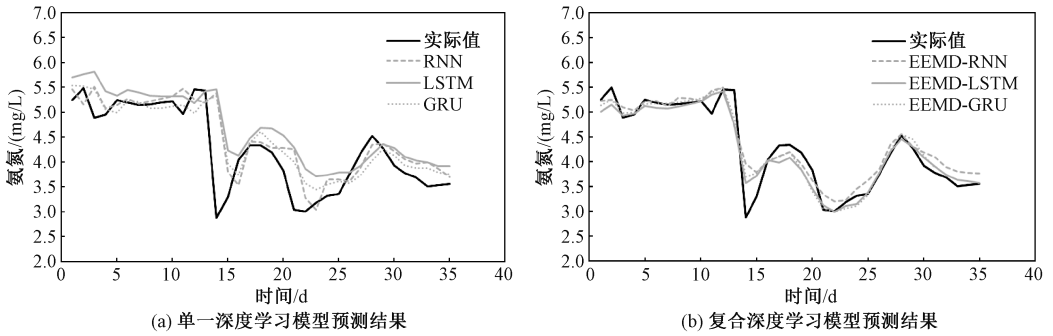


图8 第2层地下水预测结果

Fig. 8 The second layer of groundwater prediction results

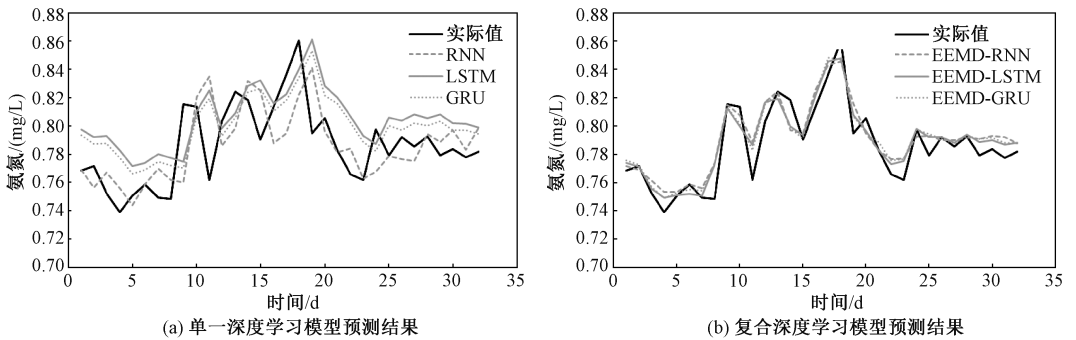


图9 第3层地下水预测结果

Fig. 9 The third layer of groundwater prediction results

表3 不同深度学习模型的预测效果对比

Table 3 Comparison of prediction effects of different deep learning models

层位	评价指标	RNN	LSTM	GRU	EEMD-RNN	EEMD-LSTM	EEMD-GRU
第1层	RMSE/(mg/L)	0.999	0.967	0.748	0.368	0.309	0.316
	MAE/(mg/L)	0.732	0.770	0.463	0.251	0.213	0.207
	$R^2$	0.552	0.580	0.749	0.939	0.957	0.955
第2层	RMSE/(mg/L)	0.700	0.662	0.525	0.362	0.227	0.208
	MAE/(mg/L)	0.515	0.456	0.326	0.279	0.155	0.132
	$R^2$	0.230	0.312	0.568	0.794	0.919	0.932
第3层	RMSE/(mg/L)	0.028	0.024	0.024	0.011	0.011	0.012
	MAE/(mg/L)	0.021	0.019	0.018	0.009	0.008	0.010
	$R^2$	-0.076	0.175	0.196	0.821	0.842	0.789

### 3.4 传统机器学习模型预测

为对比传统机器学习模型与深度学习模型预

测结果的差异,构建了MLR、SVM、RF、DT等4种传统机器学习模型。传统机器学习模型水质预测

效果评价结果如表4所示。第1层MLR与RF模型预测结果的拟合度处于0.6~0.9之间,与单一深度学习模型的预测准确度接近。其余传统机器

学习模型的拟合度均低于0.5,拟合效果欠佳。就预测准确度而言,传统机器学习模型相比复合深度学习模型存在显著差距。

表4 不同传统机器学习模型的预测效果对比

Table 4 Comparison of prediction effects of different traditional machine learning models

层位	预测效果	MLR	SVM	RF	DT
第1层	RMSE/(mg/L)	0.847	1.086	0.846	1.339
	MAE/(mg/L)	0.616	0.927	0.654	0.981
	$R^2$	0.678	0.471	0.678	0.195
第2层	RMSE/(mg/L)	0.591	0.625	0.588	0.900
	MAE/(mg/L)	0.374	0.383	0.396	0.688
	$R^2$	0.451	0.387	0.457	-0.270
第3层	RMSE/(mg/L)	0.023	0.030	0.025	0.046
	MAE/(mg/L)	0.017	0.026	0.019	0.037
	$R^2$	0.272	-0.237	0.134	-1.995

### 3.5 深度学习模型多步长预测

为进一步验证不同深度学习模型的预测性能和泛化能力,继续以第1层地下水的氨氮数据为

例,用过去15d的数据进行了多步长(未来2d和3d)水质预测实验。表5展示了不同步长下的各深度学习模型的预测效果。

表5 第1层地下水多步长预测效果对比

Table 5 Comparison of multi-step prediction effects of groundwater in the first layer

预测时长/d	预测效果	RNN	LSTM	GRU	EEMD-RNN	EEMD-LSTM	EEMD-GRU
2	RMSE/(mg/L)	1.038	1.182	1.041	0.365	0.458	0.407
	MAE/(mg/L)	0.810	0.913	0.703	0.269	0.352	0.302
	$R^2$	0.544	0.408	0.540	0.944	0.911	0.930
3	RMSE/(mg/L)	1.170	1.479	1.315	0.468	0.523	0.467
	MAE/(mg/L)	0.887	1.161	0.968	0.341	0.391	0.333
	$R^2$	0.405	0.050	0.248	0.905	0.881	0.905

从整体来看,随着预测步长的增加,模型的预测误差会逐渐变大,预测效果逐渐变差。相比单一深度学习模型,复合深度学习模型的预测误差相对较小。步长为2(对应未来2d)时,EEMD-RNN、EEMD-LSTM、EEMD-GRU复合深度学习模型预测结果的拟合度分别为0.944、0.911、0.930;步长为3(对应未来3d)时,预测结果的拟合度分别为0.905、0.881、0.905。上述结果均在可接受范围内,可以满足水质预测的精度要求。因此,复合深度学习模型可用于对未来3d的水质进行精准预测。

## 4 结论

1)相比未结合EEMD的单一深度学习预测模型,结合EEMD的复合深度学习模型的水质预测数据的拟合程度更好,预测效果更佳,解决了单一深度学习模型存在的滞后性问题。EEMD-RNN、EEMD-LSTM、EEMD-GRU模型的第1层地

下水预测结果的拟合度分别达到了0.939、0.957、0.955,第2层地下水预测结果的拟合度分别为0.794、0.919、0.932,第3层地下水预测结果的拟合度分别为0.821、0.842、0.789。

2)除第1层MLR与RF模型预测结果的拟合度较好,与单一深度学习模型的预测准确度接近外,其余传统机器学习模型的拟合度均较差。从预测准确度来看,4种传统机器学习模型均与复合深度学习模型存在显著差距。

3)在多步长预测中,EEMD-RNN、EEMD-LSTM、EEMD-GRU三种复合深度学习模型未来2d预测结果的拟合度分别为0.944、0.911、0.930,未来3d预测结果的拟合度分别为0.905、0.881、0.905,预测结果可以满足水质预测的精度要求。

综上所述,复合深度学习模型可以提供及时、精确的地下水污染预测预警信息,为地下水环境管理提供技术支撑,对加强地下水环境保护工作具有重要价值。

## 参考文献 (References):

- [ 1 ] LI P Y, KARUNANIDHI D, SUBRAMANI T, et al. Sources and Consequences of Groundwater Contamination [ J ]. Archives of Environmental Contamination and Toxicology, 2021, 80(1) : 1-10.
- [ 2 ] 田志仁, 李名升, 夏新, 等. 我国地下水环境监测现状和工作建议 [ J ]. 环境监控与预警, 2020, 12(6) : 1-6.  
TIAN Zhiren, LI Mingsheng, XIA Xin, et al. Current Situation and Suggestions of Groundwater Monitoring Work in China [ J ]. Environmental Monitoring and Forewarning, 2020, 12(6) : 1-6.
- [ 3 ] 生态环境部. 2022 中国生态环境状况公报 [ R ]. 北京: 生态环境部, 2023.
- [ 4 ] 任静, 李娟, 席北斗, 等. 我国地下水污染防治现状与对策研究 [ J ]. 中国工程科学, 2022, 24(5) : 161-168.  
REN Jing, LI Juan, XI Beidou, et al. Groundwater Pollution Prevention and Control in China: Current Status and Countermeasures [ J ]. Strategic Study of CAE, 2022, 24(5) : 161-168.
- [ 5 ] CHEN K Y, CHEN H X, ZHOU C L, et al. Comparative Analysis of Surface Water Quality Prediction Performance and Identification of Key Water Parameters Using Different Machine Learning Models Based on Big Data [ J ]. Water Research, 2020, 171 : 115454.
- [ 6 ] ZHU M Y, WANG J W, YANG X, et al. A Review of the Application of Machine Learning in Water Quality Evaluation [ J ]. Eco-Environment & Health, 2022 (2) : 107-116.
- [ 7 ] CHEN Z, LIU L M, WANG Y S, et al. Review of Water Quality Prediction Methods [ C ] // WENG C H. Proceedings of the 8th International Conference on Water Resource and Environment. Singapore: Springer, 2023 : 237-265.
- [ 8 ] JI X L, SHANG X, DAHLGREN R A, et al. Prediction of Dissolved Oxygen Concentration in Hypoxic River Systems Using Support Vector Machine: A Case Study of Wen-Rui Tang River, China [ J ]. Environmental Science and Pollution Research, 2017, 24 : 16 062-16 076.
- [ 9 ] 肖懿, 郭亚会, 李明蔚, 等. 基于机器学习的地下水水质预测研究 [ J ]. 北京师范大学学报 (自然科学版), 2022, 58(2) : 261-268.  
XIAO Yi, GUO Yahui, LI Mingwei, et al. Machine Learning to Predict Groundwater Quality [ J ]. Journal of Beijing Normal University ( Natural Science ), 2022, 58(2) : 261-268.
- [ 10 ] DELGADO A, VRICLIZAR D, MEDINA E. Artificial Intelligence Model Based on Grey Systems to Assess Water Quality from Santa River Watershed [ C ] // 2017 Electronic Congress ( E-CON UNI ). IEEE, 2017 : 1-4.
- [ 11 ] MA C, ZHANG H H, WANG X F. Machine Learning for Big Data Analytics in Plants [ J ]. Trends in Plant Science, 2014, 19(12) : 798-808.
- [ 12 ] LIU Y, ZHAO T L, JU W W, et al. Materials Discovery and Design Using Machine Learning [ J ]. Journal of Materiomics, 2017, 3(3) : 159-177.
- [ 13 ] FARUK D Ö. A Hybrid Neural Network and ARIMA Model for Water Quality Time Series Prediction [ J ]. Engineering Applications of Artificial Intelligence, 2010, 23(4) : 586-594.
- [ 14 ] CHEN Y Y, SONG L H, LIU Y Q, et al. A Review of the Artificial Neural Network Models for Water Quality Prediction [ J ]. Applied Sciences, 2020, 10 (17) : 5776.
- [ 15 ] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning Representations by Back-Propagating Errors [ J ]. Nature, 1986, 323(6 088) : 533-536.
- [ 16 ] WANG Y, ZOU R M, LIU F, et al. A Review of Wind Speed and Wind Power Forecasting with Deep Neural Networks [ J ]. Applied Energy, 2021, 304 (1) : 117766.
- [ 17 ] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory [ J ]. Neural Computation, 1997, 9(8) : 1 735-1 780.
- [ 18 ] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling [ J/OL ]. (2014-12-11) [ 2024-02-06 ]. <https://doi.org/10.48550/arXiv.1412.3555>.
- [ 19 ] WANG Y Y, ZHOU J, CHEN K J, et al. Water Quality Prediction Method Based on LSTM Neural Network [ C ] // 2017 12th International Conference on Intelligent Systems and Knowledge Engineering ( ISKE ). IEEE, 2017 : 1-5.
- [ 20 ] LIU P, WANG J, SANGAIAH A K, et al. Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment [ J ]. Sustainability, 2019, 11(7) : 1-14.
- [ 21 ] VALADKHAN D, MOGHADDASI R, MOHAMMADINEJAD A. Groundwater Quality Prediction Based on LSTM RNN: An Iranian Experience [ J ]. International Journal of Environmental Science and Technology, 2022, 19

- (11);11 397-11 408.
- [22] 张鼎渊. 基于 EEMD-LSTM 水质预测方法的研究与应用[D]. 合肥:中国科学技术大学,2021.
- [23] WU Y X, WU Q B, ZHU J Q. Improved EEMD-Based Crude Oil Price Forecasting Using LSTM Networks [J]. *Physica A: Statistical Mechanics and Its Applications*, 2019, 516:114-124.
- [24] CHEN Z S, YUAN C F, WU H F, et al. An Improved Method Based on EEMD-LSTM to Predict Missing Measured Data of Structural Sensors [J]. *Applied Sciences*, 2022, 12(18):9027.
- [25] 周泉, 胡轩铭, 王东昆, 等. 基于注意力机制优化的 LSTM 河流溶解氧预测模型研究[J]. *环境科学研究*, 2023, 36(6):1 135-1 146.
- ZHOU Quan, HU Xuanming, WANG Dongkun, et al. Prediction of Dissolved Oxygen in Rivers Based on LSTM Model with Improved Attention Mechanism [J]. *Research of Environmental Sciences*, 2023, 36(6):1 135-1 146.
- [26] ZHOU J, WANG Y Y, XIAO F, et al. Water Quality Prediction Method Based on IGRA and LSTM [J]. *Water*, 2018, 10(9):1148.
- [27] 王新民, 张超超. 基于深度学习的旧金山湾水质预测[J]. *吉林大学学报(地球科学版)*, 2021, 51(1):222-230.
- WANG Xinmin, ZHANG Chaochao. Water Quality Prediction of San Francisco Bay Based on Deep Learning [J]. *Journal of Jilin University (Earth Science Edition)*, 2021, 51(1):222-230.
- [28] PANG Z H, NIU F X, ZHENG O. Solar Radiation Prediction Using Recurrent Neural Network and Artificial Neural Network: A Case Study with Comparisons [J]. *Renewable Energy*, 2020, 156:279-289.
- [29] MORIASI D N, ARNOLD J G, VAN LIEW M W, et al. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations [J]. *Transactions of the ASABE*, 2007, 50(3):885-900.
- [30] AHMED A N, OTHMAN F B, AFAN H A, et al. Machine Learning Methods for Better Water Quality Prediction [J]. *Journal of Hydrology*, 2019, 578:124084.
- [31] 顾阔, 焦瑞莉, 薄宇, 等. 基于复合 LSTM 模型的 PM<sub>2.5</sub> 浓度预测[J]. *中国环境监测*, 2023, 39(1):170-180.
- GU Kuo, JIAO Ruili, BO Yu, et al. PM<sub>2.5</sub> Concentration Prediction Based on the Composite LSTM Model [J]. *Environmental Monitoring in China*, 2023, 39(1):170-180.
- [32] BARZEGAR R, AALAMI M T, ADAMOWSKI J. Short-Term Water Quality Variable Prediction Using a Hybrid CNN-LSTM Deep Learning Model [J]. *Stochastic Environmental Research and Risk Assessment*, 2020, 34(2):415-433.