第 31 卷第 1 期
Vol. 31 No. 1
Jan. 2025
氣候與環境研究
Climatic and Environmental Research

罗凌云, 周一凡, 占明锦, 等. 2025. 融合地理感知注意力的城市内涝预测模型研究 [J]. 气候与环境研究, 31(1): 1−10. LUO Lingyun, ZHOU Yifan, ZHAN Mingjin, et al. 2025. Urban Flooding Prediction Model Integrated with a Geospatial-Aware Attention Module [J]. Climatic and Environmental Research (in Chinese), 31 (1): 1−10. doi:10.3878/j.issn.1006-9585.2025.25054

融合地理感知注意力的城市内涝预测模型研究

罗凌云[1] 周一凡[1] 占明锦[1] 占龙飞[2] 汪宗伟[1]

1 江西信息应用职业技术学院，南昌 330000
2 江西省气候中心，南昌 330000

摘 要 提出了一种基于 Transformer 模型的新型城市雨涝预测模型。利用 Transformer 模型在时序预测任务上的良好性能，并创新性地融入了地理感知注意力模块（Geospatial-aware Attention Module），以更好地捕捉城市内涝的空间异质性和地理特征，增强了模型对城市地形等空间特征的学习能力，结合南昌市 2023 年 6～8 月的降水数据及城市内涝监测站的水位数据开展实验，结果表明本文提出的方法相较原始 Transformer 模型在城市内涝预测任务中表现更优，且模型仍能保持较好的预测效率，均方根误差较原模型降低 0.0006，决定系数较原模型提高了 0.071。这表明本文提出的地理感知注意力模块能够有效提升时序模型在城市内涝预测任务中的性能。

关键词 极端暴雨天气 城市内涝 雨涝预测 Transformer 模型 地理感知注意力
文章编号 1006-9585(2025)01-0001-10 **中图分类号** P49 **文献标识码** A
doi:10.3878/j.issn.1006-9585.2025.25054

Urban Flooding Prediction Model Integrated with a Geospatial-Aware Attention Module

LUO Lingyun[1], ZHOU Yifan[1], ZHAN Mingjin[1], ZHAN Longfei[2], and WANG Zongwei[1]

1 *Jiangxi Vocational and Technical College of Information Application*, *Nanchang* 330000
2 *Jiangxi Climate Center*, *Nanchang* 330000

Abstract This paper proposes a novel, Transformer-based model for predicting urban-rain-induced flooding. The model leverages the remarkable performance of Transformers in time-series prediction tasks and innovatively integrates a geospatial-aware attention module to better capture the spatial heterogeneity and geographical features of urban flooding, thereby enhancing the model's ability to learn spatial features, such as urban terrain. Experiments were conducted using precipitation data from June to August 2023 in Nanchang City and water-level data from urban flood monitoring stations. The results showed that the proposed method outperforms the original Transformer model for urban flooding prediction tasks while maintaining high prediction efficiency. The root mean square error of the proposed model is lower by 0.0006 compared to that of the original model, and the coefficient of determination is higher by 0.071. This indicates that the geospatial-aware attention module proposed herein can effectively improve the performance of time series models in urban flooding prediction tasks.

Keywords Extreme rainstorms, Urban waterlogging, Urban waterlogging prediction, Transformer model, Geospatial-aware attention

收稿日期 2025-04-09；**网络预出版日期** 2025-09-19
作者简介 罗凌云，男，硕士，副教授，主要从事人工智能的研究。E-mail: 11439791@qq.com
通讯作者 占明锦，男，博士，正高级工程师，主要从事气象领域的研究。E-mail: hellorm@126.com
资助项目 江西省重点研发项目 20223BBG71019，中国气象局青年创新团队 CMA2024QN15
Funded by Research and Development Plan of Jiangxi Provincial Department of Science and Technology (Grant 20223BBG71019), China Meteorological Administration Youth Innovation Team Project (Grant CMA2024QN15)

1 引言

近年来，极端强降雨事件多发，给城市排涝带来严峻考验，多地频发城市内涝灾害，如 2019 年深圳“411”大暴雨、2021 年郑州“720”特大暴雨。其中郑州“720”特大暴雨造成地铁系统内涝，最终导致约 380 人遇难，直接经济损失高达人民币 409 亿元（李莹和赵珊珊, 2022）。若能够对城市雨涝进行准确预测，配合气象部门及城市水利部门开展及时有效的预警措施，对于缓解城市内涝灾害具有重要价值。国内外学者曾提出了许多预测模型来解决极端暴雨导致的城市内涝问题，如 IFMS（Integrated Flood Modeling System）、SWMM（Sorm Water Management Model）、HEC-RAS（Hydrologic Engineering Center’s River Analysis System）等（臧文斌等, 2020）。这其中包括基于物理机制的模型，如 SWMM（Sorm Water Management Model）（宋翠萍等, 2015）、FloodArea 模型（谢五三等, 2018）等，广泛用于城市排水系统设计和内涝风险评估，但其依赖高精度地理数据，计算复杂度较高，实时性较差。也包括基于统计学的模型，如回归模型 ARIMA（Auto-Regressive Integrated Moving Average）（Stellwagen and Tashman, 2013）、FSARIMA（fuzzy seasonal Auto-Regressive Integrated Moving Average）（Tseng and Tzeng, 2002），这些模型能够提取数据的非线性特征，一定程度上能够提升水位预测的精确度，但这些模型对降雨量、温度、诸多物理因素等参数依赖性强，限制了它们的实际应用。随着机器学习的不断发展，开始有研究人员使用支持向量机（SVM）（Hearst et al., 1998）等传统的机器学习方法应用于水位预测任务当中。相较于传统数理统计的预测模型，机器学习方法很大程度上简化了模型的构造难度，并且提升了一定的预测准确性。但传统机器学习方法对数据的深层次关联性缺乏有效挖掘，预测效果仍不尽如人意。

近年来，深度学习在气象领域得到了广泛应用。借助深度神经网络强大的非线性关系映射能力以及在时序预测任务上的优异性能，能够挖掘出观测数据与气候现象之间的潜在联系。如 Xu et al.（2019）利用循环神经网络（RNN）对城市积水点进行预测，相较传统方法，模型的预测精度得到了显著提高。

水位预测任务会受到多种因素的影响，其非线性、时序性、复杂性等特点使得简单神经网络难以实现预测精度的要求（刘雪梅等, 2021）。因此有研究人员开始引入时序网络用于水位预测。比如郭燕和赖锡军（2020）将长短时记忆网络（LSTM）（Graves, 2012）应用于水位预测，从模型构建、数据集选取等多个角度讨论模型湖泊水位预测任务中的优化方式，从而准确预测鄱阳湖多个站点的水位。刘惟飞等（2022）将门控循环单元（GRU）（Chung et al., 2014）融入到湖泊水位预测模型中，通过对不同训练集进行研究，发现 GRU 融合模型在小样本数据集下依旧能够保持较高的预测精度。也有研究人员结合卷积与循环网络进行水位预测，如胡昊等（2024）将 CNN 与 GRU 结合，充分提取水位数据的局部空间特征和深层时间特征，预测精度较高。但长短时记忆网络（LSTM）和门控循环单元（GRU）主要缺点是并行运算能力差，导致模型训练效率低，且在训练过程中无法将权重精准地分配给重要信息，因此预测性能还有提升空间。

Transformer 模型（Vaswani et al., 2017）的提出改进了 LSTM 和 GRU 的缺点，其通过自注意力机制（Self-Attention）能够进行快速并行运算。基于此，有研究人员将其应用于水位预测任务中。如冯鹏宇等（2023）提出利用 Transformer 模型提取地下水位在时间序列上包含的关键信息，有效预测了未来 12 个时刻的地下水位值。

受特殊地形影响，南昌夏季容易出现突发性中小尺度对流天气，易发短时强降水，加之城区地势平坦，极易引发城市内涝。本文以南昌市为对象进行城市内涝风险研究。基于 Transformer 的多头注意力机制，提出了一种融合地理感知注意力的城市雨涝预测模型，创新融入城市地理信息注意力机制（Geospatial-aware attention module），使得模型能够挖掘各雨涝点水位之间的空间联系，突出地理信息对城市雨涝潜在的影响。在此基础上将双向长短时记忆网络 BiLSTM 模型（Schuster and Paliwal, 1997）与 Transformer 编码器结合，基于 Transformer 编码器良好的时间序列建模能力，增强降水数据的特征提取能力，有效提高了模型预测的精度。

2 模型构建

2.1 注意力机制

注意力机制（ATT）是由 Bahdanau et al.（2014）提出，旨在突出翻译任务中与当前词临近的词语产生的影响，从而提升翻译精度。受 Bahdanau et al.（2014）的启发，注意力机制被广泛运用于神经网络的设计当中，并且取得了很多成效。Vaswani et al.（2017）提出了 Tranformer 模型，该模型的核心思想是“自注意力（Self-Attention）机制”，它可以同时处理输入序列中的所有元素，捕捉元素之间的关系，这种机制通过计算输入序列中的每个元素对其他元素的“注意力”分数来工作，使得模型能够专注于输入中最相关的部分。此外，Transformer 模型还采用了多头注意力（Multi-Head Attention）机制，增加了模型捕捉不同子空间表示的能力。Transformer 模型结构的核心“多头注意力机制”如图 1 所示。

查询 $\boldsymbol{Q}$（query）、键 $\boldsymbol{K}$（key）、值 $\boldsymbol{V}$（value）是输入序列的嵌入表示矩阵 $\boldsymbol{X}\in\boldsymbol{R}^{L\times d_{\mathrm{m}}}$（$L$为序列长度，$d_{\mathrm{m}}$为模型维度）通过线性变换生成 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$：$\boldsymbol{Q}=\boldsymbol{X}\boldsymbol{W}^{Q},\boldsymbol{K}=\boldsymbol{X}\boldsymbol{W}^{K},\boldsymbol{V}=\boldsymbol{X}\boldsymbol{W}^{V}$，其中，$\boldsymbol{W}^{Q}$, $\boldsymbol{W}^{\mathrm{K}}\in\boldsymbol{R}^{d_{\mathrm{m}}\times d_K}$, $\boldsymbol{W}^{V}\in\boldsymbol{R}^{d_{\mathrm{m}}\times d_V}$，$d_K$和$d_V$分别为$\boldsymbol{K}$和$\boldsymbol{V}$的维度。然后通过查询和健的点积来计算相似度，具体公式如下所示：

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_K}}\right)\boldsymbol{V},\qquad(1)$$

其中，$\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}$表示每个位置的查询与所有键的相似度矩阵，softmax函数会沿着健的维度归一化，从而得到注意力权重矩阵，$1/\sqrt{d_K}$ 为缩放因子，可以避免点积结果过大导致softmax函数梯度不稳定。

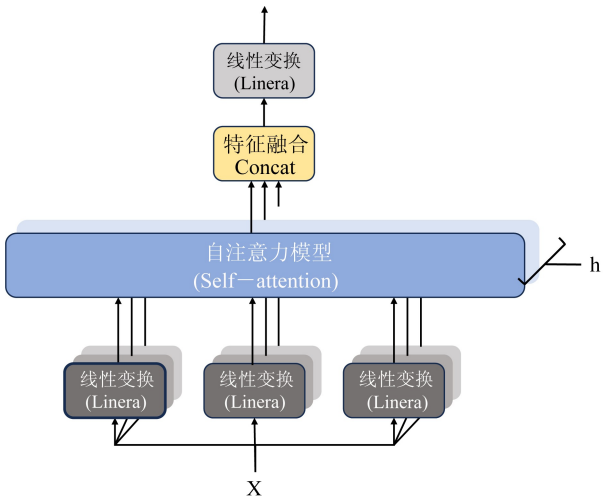


图 1　多头注意力机制。

Fig. 1　Multi-head attention mechanism.

最后输出注意力权重与值V的加权和。

由于$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$均来自同一个输入矩阵$\boldsymbol{X}$，因此该方法被称为“自注意力机制”，通过将$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$拆分到多个子空间并行计算，最后拼接结果，便可实现多头注意力机制。

自注意力机制已被证实在众多时序预测任务中有独特优势，例如机器翻译、命名实体识别等领域（Shaw et al., 2018; 游乐圻等, 2023; 余正涛等, 2025），相比于其他更早的时序预测模型如 LSTM、GRU 等可能由于长距离依赖容易造成的梯度消失的问题，自注意机制能够有效地规避，同时还能解决传统 RNN 必须按序列顺序逐步计算从而无法充分利用 GPU 性能的缺陷，做到并行计算，大幅提升模型训练和推理速度。

2.2　地理感知注意力机制

本文针对城市雨涝会受城市复杂地形影响的特点，提出了融合地理信息的多头感知注意力机制（GeoMultiHeadAttention, GMHA），在 Transformer 模型的多头注意力机制的基础上，添加了地理距离权重，用以挖掘各站点之间潜在的地理空间关系对雨涝预测结果的影响。其结构如下图所示：

图 2 中 $\boldsymbol{X}$ 为降水数据的嵌入表示矩阵，$\boldsymbol{G}$为地理信息嵌入表示矩阵，其具体计算过程如下，首先获取雨涝监测站点的地理坐标（本文使用经纬度坐标），对于其中任意两个坐标点 $P_1(\phi_1,\lambda_1)$和 $P_2(\phi_2,\lambda_2)$，依据下列公式进行距离计算：

$$d(P_1,P_2)=2R\arcsin\left[\sqrt{\sin^2\left(\frac{\delta\phi}{2}\right)+\cos\phi_1\cos\phi_2\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right],\qquad(2)$$

其中，R是地球半径；$\Delta\phi=\phi_1-\phi_2$，表示纬度差；$\Delta\lambda=\lambda_2-\lambda_1$，表示经度差。而对于$n$个站点坐标，距离矩阵$\boldsymbol{D}\in R^{n\times n}$可表示为$D_{ij}=d\left(P_i,P_j\right)$，得到的距离矩阵由于取值从几百米到几十公里不等，范围较大，为了便于模型的训练，距离矩阵还需要经过归一化处理，如下所示：

$$D_{\mathrm{normalized}}=\frac{D-D_{\min}}{D_{\max}-D_{\min}}.\qquad(3)$$

为了确保数据在语义上的直观性以及与实际物理意义的一致性，在得到归一化的距离矩阵后还需要进行相似度转换，如下公式所示：

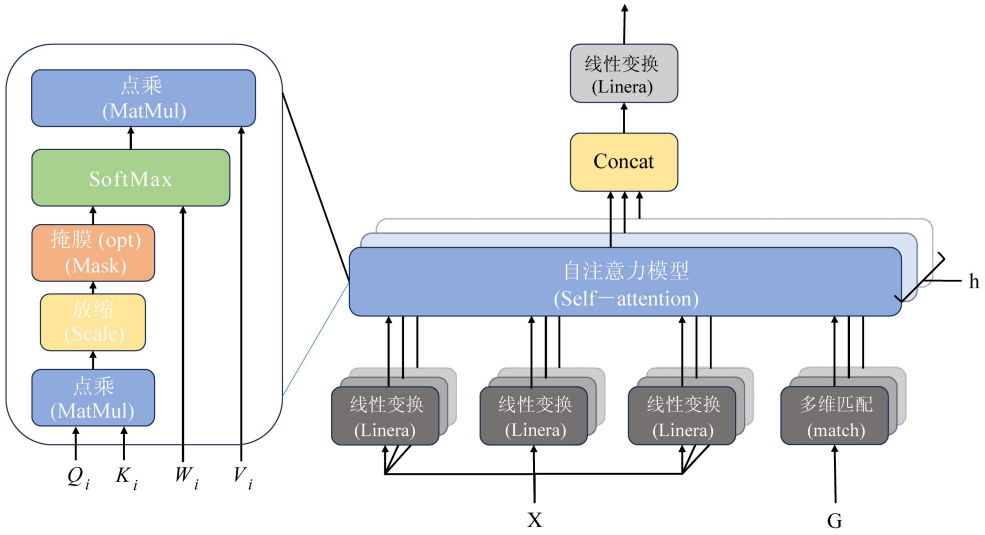


图 2　地理感知多头注意力机制。

Fig. 2　Geoperceptual Multi-Head Attention mechanism.

$$S = 1 - D_{\text{normalized}}. \tag{4}$$

最终的地理感知注意力权重矩阵可表示为

$$\boldsymbol{G} = a\boldsymbol{S}, \tag{5}$$

其中，a为缩放因子，添加缩放因子可以灵活地调整地理距离对模型的影响程度，同时增加数值计算的稳定性。

在得到最终的地理感知注意力权重矩阵$\boldsymbol{G}$后，将其针对输入序列嵌入表示矩阵 $\boldsymbol{X}$ 的线性变换过程进行多维匹配，得到地理权重表示W，并加入到 Transformer 模型的原始自注意力机制中，得到融合地理信息的多头注意力机制，如下公式所示：

$$\text{GeographicAttention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V},\boldsymbol{W}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_K}} + W\right)V. \tag{6}$$

2.3　融合地理感知注意力机制的城市内涝预测模型

融合地理感知注意力机制的城市内涝预测模型多头注意力模型结构如图 3 所示，模型主要由特征提取层、改进的编码器和输出预测层组成。

输入数据首先经过特征提取层以提取数据特征。经过多层非线性变换提取降水数据的深层特征，并在特征中添加时间嵌入表示，用以增强模型对时序模式的感知能力，时间嵌入矩阵定义如下：

$$\boldsymbol{E}_{\text{temporal}} \in \boldsymbol{R}^{1\times L_{\max}\times d_{\mathrm{m}}}, \tag{7}$$

其中，$L_{\max}$表示最大序列长度，本文设置为 1000，d_{m}表示 Transformer 模型内部的特征维度，本文根据所研究的任务复杂性，将其设置为 64，对于输入序列长度L小于或等于$L_{\max}$的，其时间嵌入矩阵可表示为

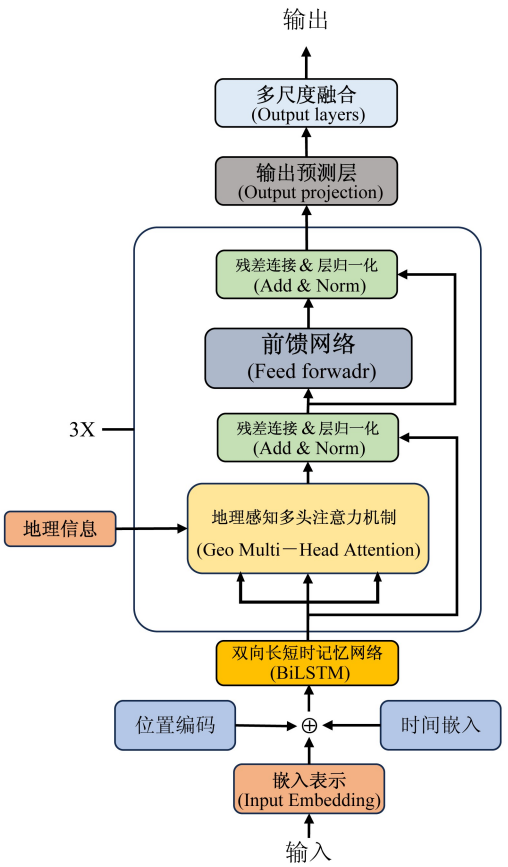


图 3　城市雨涝预测模型结构。

Fig. 3　Structure of urban rainlogging prediction model.

$$E_{\text{temp}} = E_{\text{temporal}}[:,:\mathrm{L},:] \in \boldsymbol{R}^{1\times \mathrm{L}\times \mathrm{d_m}}. \tag{8}$$

对于输入特征嵌入矩阵$\boldsymbol{X} \in \boldsymbol{R}^{L\times d_m}$，如公式所示

添加时间嵌入表示，得到增强后的嵌入矩阵$\boldsymbol{X}'$。

$$\boldsymbol{X}' = \boldsymbol{X} + \boldsymbol{E}_{\text{temp}}. \tag{9}$$

将得到的嵌入表示矩阵$\boldsymbol{X}'$进行位置编码，位置编码能够方便编码器得到不同位置的注意力分数。

Transformer 模型在全局依赖关系推理任务中有一定优势，为和 Transformer 编码器进行特征互补，本文使用双向长短时记忆网络（Bi-directional Long Short-Term Memory, BiLSTM）对位置编码后的嵌入矩阵$\boldsymbol{X}'$进行进一步处理，BiLSTM 擅长捕获局部连续模式，在本文中使用前向 LSTM 捕获过去水位变化趋势，使用后向 LSTM 捕获将来可能的降水影响，并将二者进行信息融合，从而增强预测准确性。

经过 BiLSTM 处理后的特征数据被送入本文基于 Transformer 改进的编码器中。该编码器由 3 层编码层组成。每个编码层由一个地理感知多头注意力模块、两个门控机制、一个前馈网络构成。地理感知在多头注意力模块不仅接收了来自双向长短时记忆网络提供的数据，也集成了地理信息。经过地理感知多头注意力机制处理的数据被送入门控机制进行特征自适应融合，这里利用了残差连接和层归一化，将输入数据残差连接能够防止因网络层数加深导致信息丢失，并且能够缓解梯度消失的问题，加快模型收敛。使用层归一化保证特征分布在合理范围。本文在前馈网络中主要针对数据进行了维度扩展和 Dropout 正则化的处理。以实现高维数据映射增强特征表达，最后再经过一层门控机制得到编码层的输出。

经过 3 层编码层处理后的数据会经过输出预测层进行结果预测，最后经过由 3 层全连接层构成的多尺度融合层得到最终的预测结果。

本文提出的预测模型在模型构建、数据采集与模拟方面有较大优势。传统水文水动力模型在模拟时要考虑多种因素，如不同重现期（李智等, 2019）、不同降水强度（李家科等, 2014），降雨雨型（胡彩虹等, 2022）等，增加了构建模型时的不确定性，在模拟过程中需要选取合适的模型。以 SWMM 模型为例，在使用时需将研究区划分为多个排水区，难以模拟淹没水深等关键致灾水力要素（马勇等, 2021），在进行预测任务前，需要对目标城市的雨水管道、地形、地表铺装信息以及降雨、管道和河道水流量、水位等情况监测资料及进行收集（马洪涛等, 2008），在实际的情景中，城市的雨水管道和地表铺装等信息经过每年市政工程建设会经常发生变化，这会对传统模型的使用造成较大干扰。本文提出的模型基于神经网络搭建，仅需收集城市的降水数据与水位数据，以及固定的水位监测站点经纬度信息，即可进行城市内涝模拟和水位预测，大大减少了模型的搭建难度。

3　实验设置

3.1　数据来源及预处理

本文用到的数据来源于江西省气象局气候中心监测的 2023 年 6～8 月南昌市降水数据、分布在南昌市主城区的 14 个内涝站点监测的水位数据及内涝站点的经纬度。降水数据采集的时间间隔为 5 分钟或 15 分钟。水位数据采集时间为每分钟。降水监测站点与内涝监测站点如图 4 所示。

图 4 中红色圆点表示数据当中的降水监测站点，黄色三角表示内涝水位监测站点，本文使用了全南昌市范围内的所有降水点数据，来预测主要集中在南昌市区的 14 个内涝水位监测点的水位。

为进一步提高模型的训练速度及预测精度，本文对上述实验数据进行了预处理，由于降水监测站在无降水时会降低实时监测的频率，从每 5 分钟记录一次降水量缩小至每 15 分钟记录一次，因此本

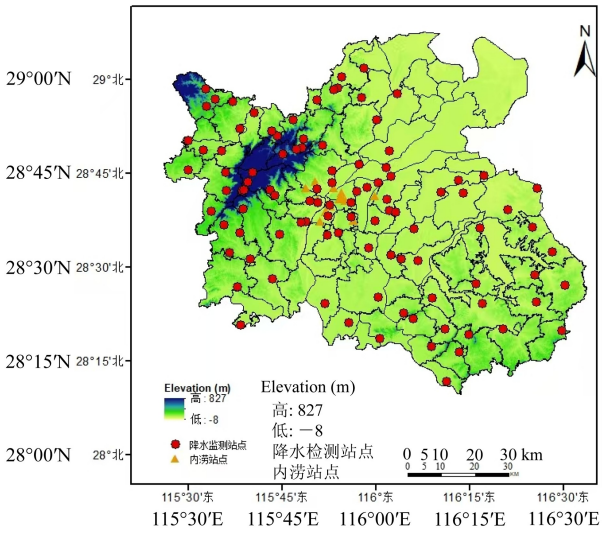


图 4　南昌市主城区降水、内涝监测站点监测分布。

Fig. 4　Distribution of precipitation and waterlogging monitoring stations in main urban area of Nanchang.

文针对上述时间间隔不满每分钟精度的降水数据使用线性插值的方法将降水数据时间精度提升到每分钟。且由于历史降水数据及水位数据中不同数据的量纲差距较大，容易降低模型的训练速度及识别精度，因此本文还对样本数据进行了归一化处理。本文采用的是 Min-Max 归一化方法，公式如下：

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{10}$$

其中，x'为归一化后的数据，x为样本原始数据，$x_{\max}$和$x_{\min}$分别是样本数据中的最大值和最小值。同时为了使预测结果能够反映现实情况，还需要将其进行反归一化处理，公式如下：

$$x = (x_{\max} - x_{\min})\,x' + x_{\min}. \tag{11}$$

经过上述处理后的降水数据与水位数据按照时间点匹配成若干数据样本，使用固定 8∶2 的比例将数据样本划分为训练集和测试集，训练集参与模型的训练和优化，并在每一个批次的训练结束后，使用测试集进行验证，以记录模型优化过程，测试集不参与模型训练。

3.2 训练环境及参数设置

模型的训练及各项处理均在 Windows 11 64 位操作系统上实现。硬件参数如下：CPU 为 Intel(R) Core(TM)i5-12500@2.5Ghz，显卡为NVIDIAGeForce RTX3060，内存为 6 G，并行运算平台为 CUDA11.2。集成开发环境为 Anaconda3.0。采用 Pytorch 深度学习框架构建城市雨涝预测模型。

训练迭代次数 epoch 设置为 50，批处理大小设置为 32，使用 AdamW 算法（Loshchilov and Hutter, 2017）进行模型参数优化。并使用单周期学习率（OneCycleLR）方法（Smith and Topin, 2019）调整学习率。训练过程中使用预测水位与真实水位的加权均方误差作为损失函数，计算公式如下：

$$L = \frac{1}{NTD}\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{d=1}^{D} w_t (y_{n,t,d} - \widehat{y_{n,t,d}})^2, \tag{12}$$

其中，N是批次大小，T是时间步长度，D是特征维度，$y_{n,t,d}$是真实值，$\widehat{y_{n,t,d}}$是预测值，w_t是时间步t的权重，通过线性插值计算，计算公式如下：

$$w_t = 1.0 - 0.5 \cdot \frac{t-1}{T-1}. \tag{13}$$

当$t = 1$时，$w_1 = 1.0$（最近时间步的最大权重），当$t = T$时，$w_T = 0.5$（最远时间步的最小权重），中间时间步的权重会线性递减，加权均方误差损失函数能够让模型在预测时更注重近期的预测准确性，这符合实际应用中对近期预测更重要的需求。

3.3 评价指标

本文选取均方根误差（Root Mean Square Error, RMSE）和决定系数（R^2）作为评价指标，用以衡量模型的预测性能。

RMSE 能够表示模型的预测值与真实值之间的偏差程度，如果越接近于 0，说明模型的预测精度越高。R^2用于衡量模型的预测效果，如果R^2的值越接近于 1，则说明模型的拟合程度越高。对应的计算公式如下：

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2}, \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - p_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{15}$$

其中，y_i表示在 i 时刻积水水位的真实值，p_i表示在 i 时刻积水水位的预测值，$\bar{y}$为真实值的平均值，n表示样本数量。

4 预测结果及分析

本文所使用的数据集中，水位数据来自南昌市 14 个水位监测站的真实监测数据，在模型验证过程中，针对这 14 个站点不同时刻的水位分别都进行了预测，预测时间窗口跨度从 1 分钟到 60 分钟。部分预测结果如图 5 所示。

图 5 中 6 张结果图展示了某一水位站点所有测试数据中在第 10、20、30、40、50、60 分钟时刻的水位预测值与真实值的偏离度。对角虚线为校准线，横坐标为水位的实际值，纵坐标为水位的预测值，蓝色落点为预测点，预测点距离校准线越近，说明预测结果与真实值的偏离越小。图 5 中显示在该水位站点的分时刻预测中，多个时刻的预测精度都在较高水平，如第 10 分钟均方根误差为 0.0023，决定系数为 0.9966，第 30 分钟均方根误差为 0.0023，决定系数为 0.9969，第 50 分钟均方根误差为 0.0023，决定系数为 0.9967，6 个时刻的均方根误差均值为 0.0024，决定系数均值为 0.9965。

由于数据集选取在 6～8 月，正值夏季旱季与短时大到暴雨交替的时段，天气放晴的时段较长，水位站零水位时段占比高，因此 6 个时刻的预测结

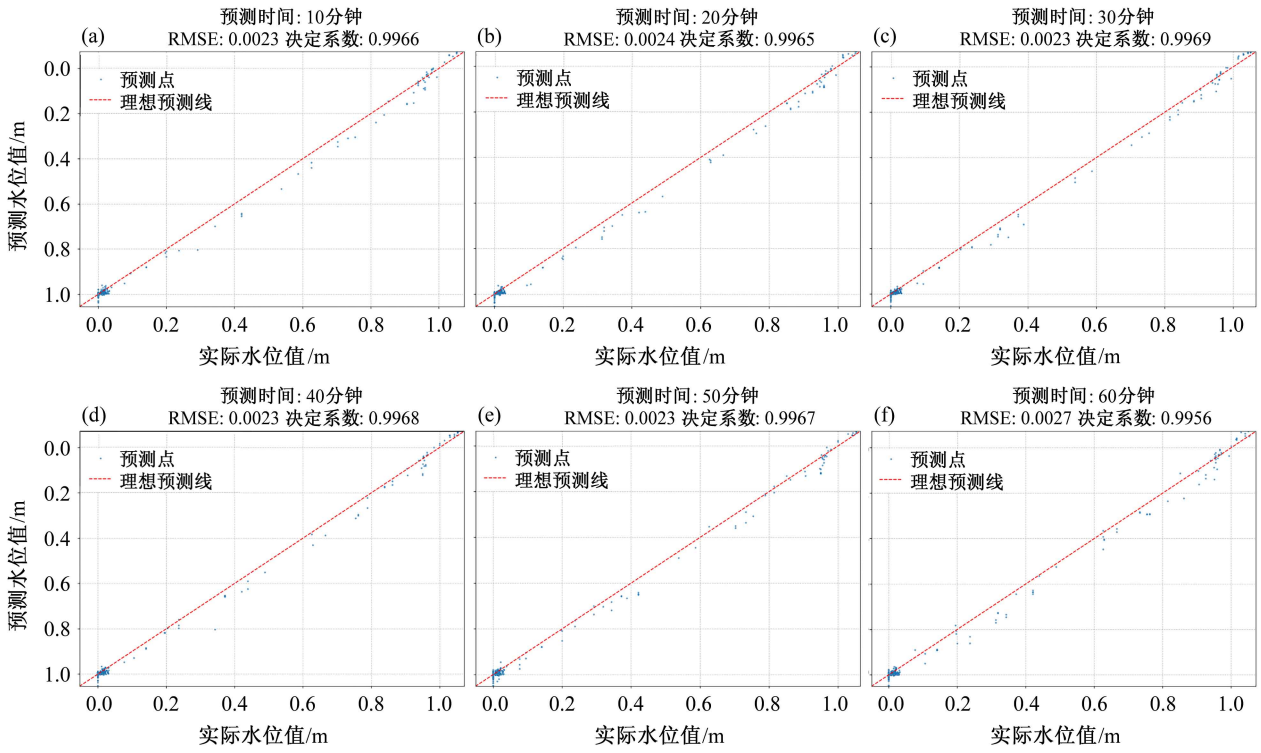


图 5　南昌市单个水位站（玉屏东大街）多时刻水位预测结果对比：（a）第 10 分钟；（b）第 20 分钟；（c）第 30 分钟；（d）第 40 分钟；（e）第 50 分钟；（f）第 60 分钟。

Fig. 5　Comparison chart of water level prediction results at multiple time points for a single water level station (East Yuping Street) in Nanchang: (a) The 10th minute; (b) the 20th minute; (c) the 30th minute; (d) the 40th minute; (e) the 50th minute; (f) the 60th minute.

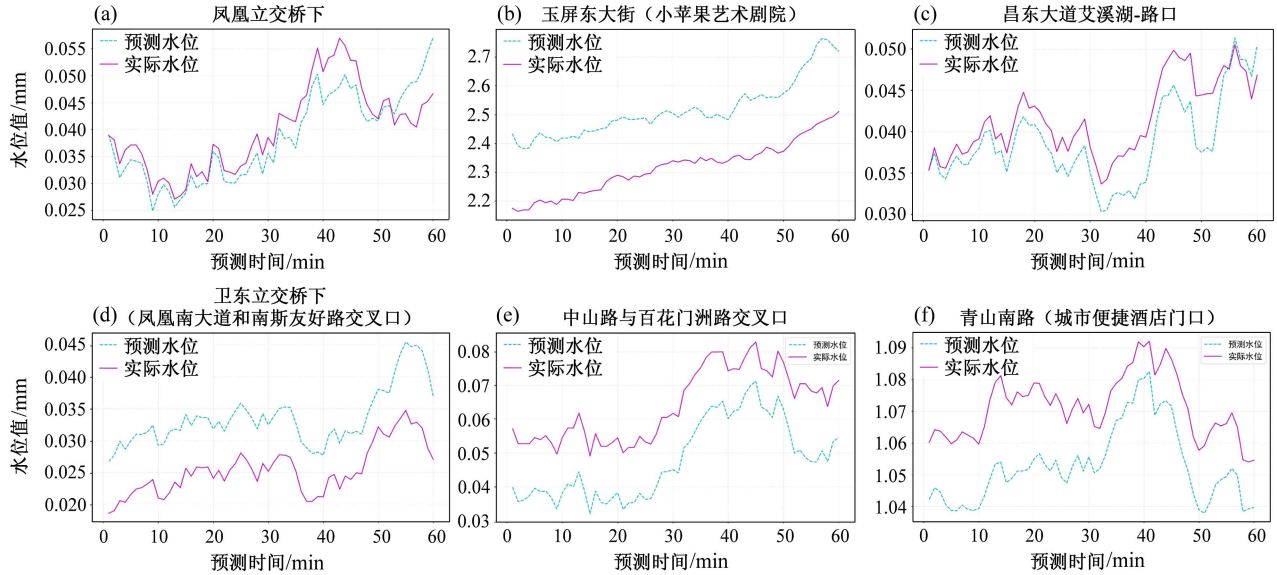


图 6　南昌市多个水位站 60 分钟水位预测走势：（a）凤凰立交桥下；（b）玉屏东大街；（c）昌东大道艾溪湖路口；（d）卫东立交桥下；（e）中山路与百花门洲路交叉口；（f）青山南路。

Fig. 6　The 60-minute water level prediction trends of multiple water level stations in Nanchang: (a) Under the Phoenix Overpass; (b) Yuping East Street; (c) Changdong Avenue, Aixi Lake Intersection; (d) under Weidong Overpass; (e) intersection of Zhongshan Road and Baifamengzhou Road; (f) Qingshan South Road.

果显示在该水位站零水位点附近聚集大量预测点，这符合实际的天气情况。但在有降水的情况下，模型预测点均分布于校准线附近，误差范围较小。

如图 6 所示，本文选取了 14 个站点中的 6 个，

并从每个站点的测试集中随机选取了一段 60 分钟完整预测结果，绘制成了走势图。图 6 横坐标表示预测时间，单位为每分钟，纵坐标表示水位值。紫色实线表示 60 分钟实际水位走势，青色虚线表示 60 分钟预测水位走势。可以看出每个站点的预测值走势均能够较好地拟合实际水位的走势。相较于传统水位水动力模型，如朱一晗（2019）、陈泽伟（2021）使用如 FloodArea 等模型进行城市内涝水位预测，本文方法的预测结果，在有一定偏离度的站点预测走势其偏离精度更优，实际预测精度更高。

模型在训练集和测试集下的损失函数收敛情况如图 7 所示，由图 7 可知，损失函数值随着迭代次数的增加而减小，并且在第 20 个 epoch 基本收敛，收敛的速度较快，损失函数值在下降至 0.0005 后呈缓慢下降趋势，并最终在 0.0001 处趋于稳定。在训练和测试过程中，并没有出现过拟合现象。

为进一步体现本文提出的预测模型在预测准确性上的优势，选取传统机器学习方法随机森林（Breiman, 2001）、时间序列预测模块 GRU 以及原始 Transformer 进行比较。在相同数据集和实验环境下针对上述 3 种模型进行实验，并得到 60 分钟的内涝预测结果。

3 种模型在包含 14 个站点的测试集上所有预测结果的均方根误差与决定系数与本文方法的对比如表 1 所示。

RMSE 为均方根误差，值越接近于 0，说明模型的预测精度越高。R^2为决定系数，值越接近于，说明模型的拟合程度越高。由表 1 可知，本文提出的城市内涝预测模型的决定系数更接近于 1，表明其拟合效果最好，且均方根误差值也最小。相比传统的随机森林，本文方法有极大的优势。在均方根误差上降低了 0.0272，在决定系数上提高了 0.566；相比 GRU 单元，本文方法的均方根误差降低了 0.0008，决定系数提高了 0.272，相比原始模型 Transformer,本文方法的均方根误差减小了 0.0006，决定系数也能够提高 0.071。由此可以得出，本文方法相较其他 3 种方法，预测结果更贴合真实值，能够更好地运用于城市内涝积水点的水位预测。

表 1　各模型预测精度评价

Table 1　Evaluation of prediction accuracy of each model

模型	RMSE	R^2
随机森林（RF）	0.0280	0.409
GRU	0.0016	0.703
Transformer模型	0.0014	0.904
本研究	0.0008	0.975

5　总结

本文提出了一种基于 Transformer 的城市雨涝预测模型，通过在多头注意力中融合地理信息，增强模型对地理信息的感知能力。在输入数据中添加时间嵌入表示，扩展数据的维度，使用双向长短时记忆网络对数据特征进行提取，挖掘降水数据的深层信息。采用江西省气象局气候中心提供的南昌市 2023 年 6～8 月降水数据和 14 个水位站监测到的水位数据对模型进行训练和验证，并与以往的模型进行了对比分析，结果表明本文提出的雨涝预测模型在雨涝预测任务中能够更精准地预测出水位站

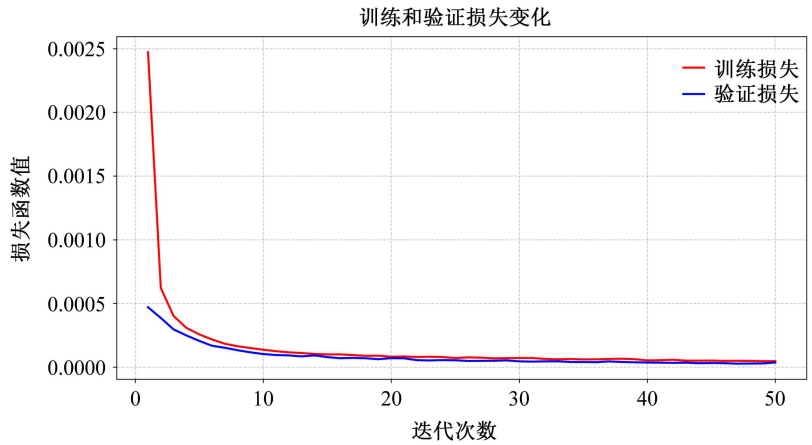


图 7　模型训练与测试损失函数值。

Fig. 7　Model training and test Loss diagram.

点的实时水位。模型在较短时间内实现了收敛，且收敛效果较好，训练损失与测试损失差距较小，没有产生过拟合现象，证明模型有较好的泛化能力。

基于 Transformer 的城市雨涝预测模型依赖城市水文观测数据作为训练样本，在极端降雨天气下监测数据可能存在样本缺失较多、误差较大等情况，对模型训练带来干扰，在未来的工作中，可以拓展训练数据的来源，运用大数据、物联网等技术获取更多实测历史数据来丰富数据来源。并拟将本文方法拓展至气象领域其他有关的时序预测任务中，以更好地拓展模型的适用性和可靠性。

参考文献（References）

Bahdanau D, Cho K, Bengio Y. 2016. Neural machine translation by jointly learning to align and translate [J/OL]. arXiv preprint arXiv: 1409.0473. https://arxiv.org/abs/1409.0473

Breiman L. 2001. Random forests [J]. Machine learning, 45(1): 5−32. doi:10.1023/A:1010933404324

陈泽伟. 2021. 基于不同模拟方法的城市下立交区域内涝模型比对及适用性分析 [J]. 中国市政工程, (5): 42–45, 50, 115–116. Chen Zewei. 2021. Comparison and applicability analysis of waterlogging models in urban lower interchange area based on different simulation methods [J]. China Municipal Engineering (in Chinese), (5): 42–45, 50, 115–116. doi:10.3969/j.issn.1004-4655.2021.05.012

Chung J, Gulcehre C, Cho K H, et al. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling [J/OL]. arXiv preprint arXiv: 1412.3555. https://arxiv.org/abs/1412.3555.

冯鹏宇, 金韬, 沈一选, 等. 2023. 基于 CNN-Transformer 的城区地下水位预测 [J]. 计算机仿真, 40(4): 492−498. Feng Pengyu, Jin Tao, Shen Yixuan, et al. 2023. Prediction of urban groundwater level based on CNN-transformer [J]. Computer Simulation (in Chinese), 40(4): 492−498. doi:10.3969/j.issn.1006-9348.2023.04.094

Graves A. 2012. Long short-term memory [M]//Graves A. Supervised Sequence Labelling with Recurrent Neural Networks. Berlin: Springer, 37–45. doi:10.1007/978-3-642-24797-2_4

郭燕, 赖锡军. 2020. 基于长短时记忆神经网络的鄱阳湖水位预测 [J]. 湖泊科学, 32(3): 865−876. Guo Yan, Lai Xijun. 2020. Water level prediction of Lake Poyang based on long short-term memory neural network [J]. Journal of Lake Sciences (in Chinese), 32(3): 865−876. doi:10.18307/2020.0325

Hearst M A, Dumais S T, Osuna E, et al. 1998. Support vector machines [J]. IEEE Intell. Syst. Their Appl., 13(4): 18−28. doi:10.1109/5254.708428

胡彩虹, 姚依晨, 刘成帅, 等. 2022. 降雨雨型对城市内涝的影响 [J]. 水资源保护, 38(6): 15–21, 87. Hu Caihong, Yao Yichen, Liu Chengshuai, et al. 2022. Effects of rainfall patterns on urban waterlogging [J]. Water Resources Protection (in Chinese), 38(6): 15–21, 87. doi:10.3880/j.issn.1004-6933.2022.06.003

胡昊, 陈军朋, 李擎, 等. 2024. 基于 CNN-GRU-ATT 的城市暴雨积水预测研究 [J]. 华北水利水电大学学报 (自然科学版), 45(4): 27−35. Hu Hao, Chen Junpeng, Li Qing, et al. 2024. Research on urban rainstorm water accumulation prediction based on CNN-GRU-ATT [J].Journal of North China University of Water Resources and Electric Power (Natural Science Edition) (in Chinese), 45(4): 27−35. doi:10.19760/j.ncwu.zk.2024039

李家科, 李亚, 沈冰, 等. 2014. 基于 SWMM 模型的城市雨水花园调控措施的效果模拟 [J]. 水力发电学报, 33(4): 60−67. Li Jiake, Li Ya, Shen Bing, et al. 2014. Simulation of rain garden effects in urbanized area based on SWMM [J]. Journal of Hydroelectric Engineering (in Chinese), 33(4): 60−67. doi:CNKI:SUN:SFXB.0.2014-04-010

李莹, 赵珊珊. 2022. 2001～2020 年中国洪涝灾害损失与致灾危险性研究 [J]. 气候变化研究进展, 18(2): 154−165. Li Ying, Zhao Shanshan. 2022. Floods losses and hazards in China from 2001 to 2020 [J]. Climate Change Research (in Chinese), 18(2): 154−165. doi:10.12006/j.issn.1673-1719.2021.196

李智, 刘玉菲, 任星芮男, 等. 2019. 基于 SWMM 的城市洪涝风险管理研究 [J]. 水利水电技术, 50(11): 35−42. Li Zhi, Liu Yufei, Ren Xingruinan, et al. 2019. Research on urban flood risk management based on SWMM [J]. Water Resources and Hydropower Engineering (in Chinese), 50(11): 35−42. doi:10.13928/j.cnki.wrahe.2019.11.005

刘惟飞, 陈兵, 余周. 2022. 基于 GRU-BP 组合模型的湖泊水位预测方法探索 [J]. 中国农村水利水电, (11): 58−65. Liu Weifei, Chen Bing, Yu Zhou. 2022. Exploration of lake level forecast method based on GRU-BP model combination [J]. China Rural Water Hydropower (in Chinese), (11): 58−65. doi:10.12396/znsd.220194

刘雪梅, 宋文辉, 钱峰, 等. 2021. 基于 VMD-CQPSO-GRU 模型的气象干旱预测方法 [J]. 华北水利水电大学学报 (自然科学版), 42(4): 31−40. Liu Xuemei, Song Wenhui, Qian Feng, et al. 2021. Prediction method of meteorological drought based on VMD-CQPSO-GRU method [J].Journal of North China University of Water Resources and Electric Power (Natural Science Edition) (in Chinese), 42(4): 31−40. doi:10.19760/j.ncwu.zk.2021047

Loshchilov I, Hutter F. (2017-11-14). Decoupled weight decay regularization [J/OL]. arXiv preprint arXiv: 1711.05101. https://arxiv.org/abs/1711.05101.

马洪涛, 张晓昕, 王强. 2008. 基于模型的城市道路积水应急排水措施研究 [J]. 城市道桥与防洪, (9): 42−45. Ma Hongtao, Zhang Xiaoxin, Wang Qiang. 2008. Study of urban road waterlogging emergent drainage measures based on model [J]. Urban Road Bridges Flood Control (in Chinese), (9): 42−45. doi:10.3969/j.issn.1009-7716.2008.09.014

马勇, 顾正华, 王庭辉, 等. 2021. 基于 SWMM 的山丘城区防洪排涝能力分析 [J]. 河海大学学报 (自然科学版), 49(6): 499–505. Ma Yong, Gu Zhenghua, Wang Tinghui, et al. 2021. Analysis of flood control and drainage capacity of hilly city based on SWMM [J]. Journal of Hohai University (Natural Sciences) (in Chinese), 49(6): 499–505. doi:10.3876/j.issn.1000-1980.2021.06.003

Schuster M, Paliwal K K. 1997. Bidirectional recurrent neural networks [J]. IEEE Trans. Signal Process., 45(11): 2673−2681. doi:10.1109/78.650093

Shaw P, Uszkoreit J, Vaswani A. 2018. Self-attention with relative position representations [C]//Proceedings of NAACL-HLT 2018. New Orleans, Louisiana: Association for Computational Linguistics.

Smith L N, Topin N. 2019. Super-convergence: Very fast training of neural networks using large learning rates [C]//Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications. Baltimore: SPIE, 11006: 369–386. doi:10.1117/12.2520589

宋翠萍, 王海潮, 唐德善. 2015. 暴雨洪水管理模型 SWMM 研究进展及发展趋势 [J]. 中国给水排水, 31(16): 16−20. Song Cuiping, Wang Haichao, Tang Deshan. 2015. Research progress and development trend of storm water management model [J]. China Water Wastewater (in Chinese), 31(16): 16−20. doi:10.19853/j.zgjsps.1000-4602.2015.16.005

Stellwagen E, Tashman L. 2013. ARIMA: The models of box and Jenkins [J]. Foresight: The International Journal of Applied Forecasting, (30): 28−33. doi:10.5040/9781666990355.ch-4

Tseng F M, Tzeng G H. 2002. A fuzzy seasonal ARIMA model for forecasting [J]. Fuzzy Sets Syst., 126(3): 367−376. doi:10.1016/S0165-0114(01)00047-1

Vaswani A, Shazeer N, Parmar N, et al. 2017. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc. , 6000-6010. doi:10.5555/3295222.3295349

谢五三, 吴蓉, 丁小俊. 2018. 基于 FloodArea 模型的城市内涝灾害风险评估与预警 [J]. 长江流域资源与环境, 27(12): 2848−2855. Xie Wusan, Wu Rong, Ding Xiaojun. 2018. Risk assessment and early warning of urban waterlogging based on FloodArea model [J]. Resources and Environment in the Yangtze River Basin (in Chinese), 27(12): 2848−2855. doi:10.11870/cjlyzyyhj201812021

Xu G Y, Cheng Y, Liu F, et al. 2019. A water level prediction model based on ARIMA-RNN [C]//IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService). Newark, CA, USA: IEEE, 221–226. doi:10.1109/BigDataService.2019.00038

游乐圻, 裴忠民, 罗章凯. 2023. 融合自注意力的 ALBERT 中文命名实体识别方法 [J]. 计算机工程与设计, 44(2): 605−611. You Leqi, Pei Zhongmin, Luo Zhangkai. 2023. ALBERT Chinese named entity recognition method based on self-attention [J].Computer Engineering and Design (in Chinese), 44(2): 605−611. doi:10.16208/j.issn1000-7024.2023.02.039

余正涛, 孙资钦, 张勇丙, 等. [2025-03-02]. 基于随机自注意力和动量对比学习的自监督序列推荐方法 [J/OL]. 计算机工程, 1–16. Yu Zhengtao, Sun Ziqin, Zhang Yongbing, et al. [2025-03-02]. Self-supervised sequence recommendation method based on random self-attention and momentum contrastive learning [J/OL]. Computer Engineering (in Chinese), 1–16. https://doi.org/10.19678/j.issn.1000-3428.0070193.

臧文斌, 赵雪, 李敏, 等. 2020. 城市洪涝模拟技术研究进展及发展趋势 [J]. 中国防汛抗旱, 30(11): 1−13. Zang Wenbin, Zhao Xue, Li Min, et al. 2020. Research progress and development trend of urban flood simulation technology [J]. China Flood and Drought Management (in Chinese), 30(11): 1−13. doi:10.16867/j.issn.1673-9264.2020328

朱一晗. 2019. 基于 FloodArea 模型的宁波市鄞州区城市内涝模拟及应用研究 [D]. 南京信息工程大学硕士学位论文, 13-28. Zhu Yihan. 2019. Simulation and application of urban waterlogging in Yinzhou district of Ningbo based on floodArea model [D]. M. S. thesis (in Chinese), Nanjing University of Information Science and Technology, 13-28.