

基于注意力机制的循环卷积方法 在长江感潮河段潮位预测中的应用

张典豪¹, 冯思敏¹, 吴艳鸣²

(1. 南通大学, 226000, 南通; 2. 江苏省水文水资源勘测局南通分局, 226000, 南通)

摘要:针对传统潮位预测模型在水文特征深度提取和长时序依赖关系建模中的不足, 研究构建了融合深度学习技术的预测方法, 以提升长江下游复杂水文条件下的潮位预报精度。基于卷积神经网络 (CNN)、双向长短期记忆网络 (BiLSTM) 与 transformer 的协同机制, 提出 CNN-BiLSTM-transformer 耦合模型, 模型采用 CNN 提取水位序列多尺度空间特征, 通过 BiLSTM 捕捉双向时序关联, 并引入 transformer 注意力机制解析长程依赖关系。基于天生港潮位站 2012—2024 年逐时观测数据开展实证研究, 结果表明, 在 24 h、48 h、60 h、80 h 预见期下, 模型的均方根误差 (MSE) 较 BiLSTM 基准模型分别降低了 88.38%、89.85%、91.04%、93.90%。随着预测周期的延长, 模型预测精度进一步提升, 验证了模型在长周期预测中的有效性。这主要得益于时空耦合注意力机制对空间和时间依赖性的协同建模, 以及 transformer 注意力机制对长程依赖关系的有效捕捉。本模型有效解决了水文预测中的时空特征耦合难题, 其预测成果可为长江口防洪调度提供高精度的决策支持。

关键词:潮位预测; 卷积神经网络 (CNN); 循环神经网络; 双向长短期记忆神经网络 (BiLSTM); 自注意力模型 (transformer); 长江下游; 潮位

Application of a CNN-BiLSTM-transformer model based on attention mechanism in tidal level prediction for the tidal reach of the Yangtze River//ZHANG Dianhao, FENG Simin, WU Yanming

Abstract: To address the limitations of traditional tidal level prediction models in deep extraction of hydrological features and modeling of long-term temporal dependencies, a forecasting method integrating deep learning techniques was developed to improve the accuracy of tidal level prediction under complex hydrological conditions in the lower Yangtze River. A coupled CNN-BiLSTM-transformer model is proposed based on the collaborative mechanism of Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and transformer. CNN is used to extract multi-scale spatial features from water level sequences, BiLSTM captures bidirectional temporal correlations, and the transformer attention mechanism is introduced to model long-range dependencies. An empirical study based on hourly observational data from 2012 to 2024 at the Tianshenggang Tidal Station demonstrates that the model's mean squared error (MSE) decreases by 88.38%, 89.85%, 91.04%, and 93.90%, respectively, for forecast horizons of 24 h, 48 h, 60 h, and 80 h, compared with the BiLSTM benchmark model. As the forecast horizon extends, the model's prediction accuracy continues to improve, validating its effectiveness in long-term forecasting. These improvements are mainly attributed to the collaborative modeling of spatial and temporal dependencies by the spatiotemporal attention mechanism and the effective capture of long-range dependencies by the transformer. This model effectively addresses the challenge of spatiotemporal feature coupling in hydrological forecasting and provides high-accuracy decision-making support for flood control regulation in the Yangtze River Estuary.

Keywords: tidal level prediction; convolutional neural network (CNN); recurrent neural network; bidirectional

收稿日期: 2025-05-29 修改日期: 2025-07-21

作者简介: 张典豪, 博士, 研究方向为智能计算及水文分析。E-mail: dzhang07@qub.ac.uk

long short-term memory neural network (BiLSTM); self-attention model (transformer); lower Yangtze River; tidal level

中图分类号: TV882.2+P731.34+TP183 文献标识码: A 文章编号: 1000-1123(2025)14-0047-09

DOI: 10.3969/j.issn.1000-1123.2025.14.009

长江下游受径流、潮汐、风暴潮等多重因素影响, 极端水文事件频发, 对沿线防汛安全构成严重威胁。准确实时的潮位预报对于长江下游沿线防汛安全、港口工程建设、航运安全等起着至关重要的作用。基于物理机制的传统数值模型虽能揭示潮波传播规律, 但在实时预报中面临模型计算效率、参数率定精度、初始和驱动条件等多方面制约, 预报精度受限。

近年来, 时间序列分析、数据同化、人工智能方法等数据驱动技术逐渐被引入潮位预报领域, 以提升预报精度。在深度学习领域, 潮位特征提取与时间依赖建模是潮位预测的两大核心问题。早期研究通过卷积神经网络(CNN)挖掘潮位序列的局部时空特征, 如Pan等人构建多尺度卷积结构捕捉潮汐波动特征, Baek等人设计时空融合架构实现多源数据交互建模。为进一步提升这些方法对长程时序依赖的表征能力, 研究者引入循环神经网络(RNN)及transformer进行潮位预测, 如Xu等人采用transformer架构提升长序列处理能力, Li等人结合迁移学习增强模型对水文要素耦合关系的解析。transformer内的位置编码较好地把握了时间点之间的空间关系, 在位置编码的发展历程中, Zhang等人在输入特征中添加了绝对位置编码, 以捕获序列中元素的顺序信息; 后来Wu等人在计算机视觉领域应用相对位置编码, 以解决绝对位置编码固定的局限性。

然而, 现有混合模型多采用级联式特征融合策略, 没有可以有效协调时空特征的耦合机制, 在长江下游短期高精度潮位预报中仍面临较大挑战。本文针对现有潮位预报方法在时空特征协同建模方面的不足, 提出CNN-BiLSTM-transformer耦合架构:

①通过CNN-BiLSTM并行特征提取层实现多尺度空间特征与双向时序依赖的协同建模。该模块采用一维卷积网络提取局部空间特征, 同时利用双向LSTM捕获时序依赖, 二者并行处理, 实现特征互补。

②设计位置编码模块, 将监测站点空间与潮汐周期信息嵌入transformer注意力机制。通过正余弦编码显式注入空间位置和周期性特征, 增强模型对潮汐规律的识别能力。

③基于天生港潮位站实测数据的验证表明, 本模型在不同预见期的预测误差较传统方法均有降低, 特别是在潮位转折阶段仍保持稳定精度, 验证了模型的有效性。

对比来看, 在周期性变化的时间序列(如潮汐数据)中, 传统的卷积方法虽能提取局部特征, 但难以建模长程依赖; 而单纯的循环神经网络(如LSTM)在长期依赖建模中易出现梯度消失问题。相比之下, 注意力机制结合循环卷积的方法能够有效融合局部特征提取与全局依赖建模的优势。

通过引入位置编码(如正余弦编码), 注意力机制可将周期性特征(如每日潮汐周期)显式注入模型, 增强模型对潮汐周期规律的识别能力。此外, 在潮位转折阶段或极端水文事件中, 注意力机制能够动态调整关注点, 降低对异常数据的敏感性, 从而提升模型在复杂场景下的稳定性与泛化能力。

本研究以天生港站为例进行模拟分析。天生港站位于长江澄通河段, 是长江下游干流的重要水情控制站, 距长江入海口约130 km, 上游约400 km处为长江下游重要控制站大通站。天生港站潮位主要受天文潮与长江径流的共同影响, 每日呈两涨两落的周期性变化, 具有明显的潮汐特征。

一、数据与研究方法

1. 数据资料

潮位资料采用天生港潮位站的逐时实测数据, 数据集的时间范围从2012年1月1日到2024年1月1日。为削弱人类活动、河口工程及海平面上升等因素的影响, 进行了数据归一化、时间特征提取和深度学习建模: 首先, 通过MinMaxScaler将潮位数据归一化至统一区间, 减小不同因素导致的量纲差异; 其次, 提取年、月、日等时间特征作为输入, 将周期性潮汐变化与非周期性干扰分离; 最后, 利用滑动窗口构建时间序列特征, 通过LSTM-transformer模型聚焦潮汐的短期依赖关系, 自适应学习长期趋势。这些步骤共同作用, 减弱了工程扰动和海平面上升等复杂因素对模型的影响, 使模型专注于潮汐规律预测, 提升了模型的鲁棒性和准确性。

2. 模型架构

本文使用一种融合多模态深度学习的潮位预测模型(CNN-BiLSTM-transformer),针对传统方法在长时序依赖建模与多源特征耦合方面的不足,通过协同优化时空特征提取机制提升潮位预测精度。模型采用三级耦合架构,主要结构如图1所示。

(1) 空间特征提取层

基于卷积神经网络(CNN)的多层空洞卷积模块,同步捕获长江监测站点水位数据的空间特征。空洞卷积公式见式(1)。

$$y_{i,j} = \sum_{m=1}^k \sum_{n=1}^k \omega_{m,n} \cdot x_{i+r \cdot m, j+r \cdot n} + b \quad (1)$$

式中, $y_{i,j}$ 表示输出特征图在位置 (i, j) 的值; $\omega_{m,n}$ 是卷积核在位置 (m, n) 的权重参数; $x_{i+r \cdot m, j+r \cdot n}$ 是输入特征图在位置 $(i+r \cdot m, j+r \cdot n)$ 的值; k 表示卷积核的大小; r 为空洞率,用于扩大感受野; b 为偏置项。

(2) 时序依赖解析层

通过双向长短期记忆网络(BiLSTM)提取潮汐涨落过程中的双向时序关联特征,解析潮位变化的非线性响应规律。前向LSTM单元计算过程见式(2)~(6)。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (5)$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

式中, x_t 是 t 时刻输入特征; f_t , i_t , O_t 分别为遗忘门、输入门、输出门的输出; W_f , W_i , W_o 分别为遗忘门、输入门、输出门的权重矩阵; b_f , b_i , b_o 分别为遗忘门、输入门、输出门的偏置项; W_c 为候选状态权重矩阵, b_c 为对应偏置项; $[h_{t-1}, x_t]$ 表示隐藏状态和当前输入的拼接; $\tilde{C}_t$ 为候选细胞状态,包含当前输入的新信息; C_t 是更新后的细胞状态,用来存储潮汐周期记忆; σ 为 sigmoid 激活函数,输出范围 $[0,1]$ 。

(3) 长程周期建模层

引入 transformer 编码器,利用注意力机制动态捕捉不同预见期内潮位的空间关系,有效解决传统循环神经网络(RNN)因梯度消失导致的长周期特征衰减问题。最终通过特征融合注意力机制整合时空特征,经全连接层输出高精度潮位预测值。其中自注意力机制公式见式(7)。

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

式中, Q 、 K 、 V 为由输入序列线性变换得到的查询、键、值矩阵; d_k 为键向量维度(用于缩放)。

3. 主要模块

本研究提出的潮汐预测模型通过深度融合 CNN、BiLSTM 和 transformer 三大核心架构,构建了一个层次化的多尺度时空特征解析框架,实现了对潮位变化的高精度预测。各模块基于不同的数学原理和物理机制,共同构建了一个层次化的潮位预测系统。

(1) 基于 CNN 的特征提取模块

CNN 模块的核心特点是通过局部感受野和权重共享机制,高效提取输入数据的空间或时间局部特征。该架构通常由卷积层、激活函数、池化层及全连接层构成,能够逐层捕获从简单到复杂的层次化特征模式。CNN 的关键组件包括卷积核、激活函数和池化操作,其计算过程可通过式(8)~(11)描述。

$$Z_{i,j,k} = \sum_m \sum_n (W_k \cdot X_{i+m,j+n}) + b_k \quad (8)$$

$$A_{i,j,k} = ReLU(Z_{i,j,k}) = \max(0, Z_{i,j,k}) \quad (9)$$

$$P_{i,j,k} = \max_{m,n \in Z^2} (A_{i+m,j+n,k}) \quad (10)$$

$$y = \sigma(W_{fc} \cdot P_{flat} + b_{fc}) \quad (11)$$

式中, $Z_{i,j,k}$ 为第 k 个卷积核在位置 (i, j) 的线性输出; m 、 n 为池化窗口内的局部索引; W_k 为第 k 个卷积核的权重; $X_{i+m,j+n}$ 为输入特征; b_k 为偏置; $A_{i,j,k}$ 为 $ReLU$ 激活后的特征值; $P_{i,j,k}$ 为最大池化后的输出; Z^2 表示池化窗口的整数坐标范围; P_{flat} 为展平后的池化输出; W_{fc} 和 b_{fc} 为全连

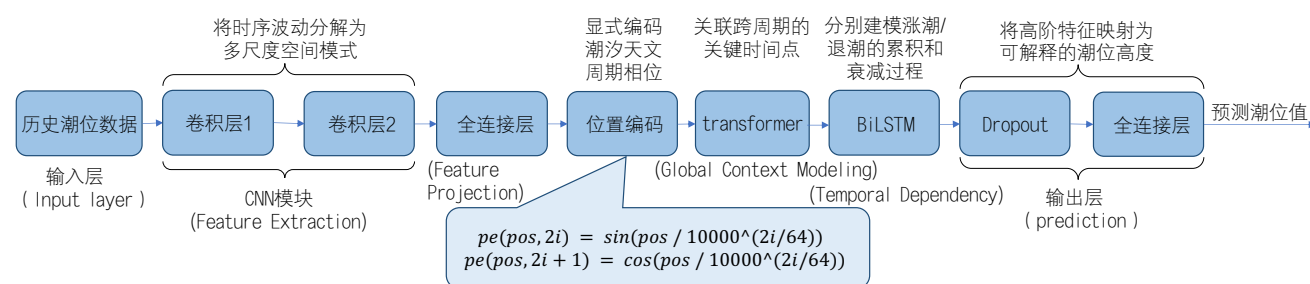


图1 模型整体架构

接层的权重和偏置; σ 为 sigmoid 激活函数; y 为最终预测结果。

在潮汐水位预测任务中, CNN 可有效捕捉历史序列中的局部波动模式和周期性趋势, 通过空间相关性建模提升预测精度。此外, 其并行计算特性使模型在处理高维数据时更具效率优势。

(2) 基于 BiLSTM 的时序预测模块

BiLSTM 架构的特点是集成了两个 LSTM 层, 一个处理正向序列, 另一个处理反向序列, 允许模型同时捕获正向和反向时间依赖关系。BiLSTM 的关键组件包括存储单元、输入门、输出门和遗忘门, 它们使模型能够保持过去输入的某种存储形式。

LSTM 中的密集层或全连接层旨在学习从输入数据到输出数据的映射, 从而促进通过线性或非线性运算进行特征转换。密集层的公式见式 (12)。

$$y = \sigma(Wx + b) \quad (12)$$

式中, y 和 x 分别是输出向量和输入向量; W 和 b 是可训练的权重矩阵和偏置; σ 是激活函数。

该模型以预测的潮汐水位为输出变量, 以历史潮汐序列为输入变量, 能够捕捉潮汐水位的空间关系和趋势, 从而提高预测的准确性。

(3) 基于注意力机制的全局上下文编码

Transformer 模型捕获长期依赖关系的能力对于解决短期水位预报中的高时间分辨率需求和广泛的依赖关系挑战至关重要。该 transformer 由几个关键组件组成, 包括层归一化、注意力机制、卷积和 dropout。多头注意力机制的公式见式 (13)。

$$\begin{aligned} & \text{MultiHead}(Q, K, V) \\ &= \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O \end{aligned} \quad (13)$$

其中每个 head_i 的计算方式见式 (14)。

$$\text{head}_i = \text{Attention}(Q_i W_i^Q, K_i W_i^K, V_i W_i^V) \quad (14)$$

而 Attention 函数定义见式 (15)。

$$\text{Attention}(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \quad (15)$$

式中, $Q=xW_Q$ 、 $K=xW_K$ 和 $V=xW_V$ 是输入 x 的线性变换, 分别生成查询向量、键向量和值向量; W_Q 、 W_K 和 W_V 是可学习的权重矩阵; d_k 表示查询和关键向量的维数, 是一个缩放因子, 可以防止内积增长过大; Q_i 、 K_i 和 V_i 是第 i 个头的变换输入矩阵, 通过将输入 x 与相应的权重矩阵 W_i^Q 、 W_i^K 和 W_i^V 相乘而获得。然后, 每个头的输出由输出权重矩阵 W^O 连接和转换, 以产生多头注意力机制的最终输出。

在 transformer 模型中, 位置编码是一种基本机

制, 它使模型能够理解序列中元素之间的位置关系。Transformer 论文中提出的一种位置编码方法使用不同频率的正弦和余弦函数来计算位置编码, 称为“正弦位置编码”。具体来说, 对于位置编码中的位置 pos 和维度 i , 计算公式见式 (16) ~ (17)。

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (16)$$

$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (17)$$

式中, pos 是序列中位置的索引; i 是位置编码维度的索引; d_{model} 是模型的维度。

整个模型的预测流程体现了一个从局部到全局、从空间到时序的特征解析过程。原始潮位数据首先经过 CNN 模块提取局部空间特征, transformer 模块在此基础上建立全局的时间依赖关系, BiLSTM 模块最后完成时序动态的精细建模。这种层次化的处理方式使得模型能够同时捕捉潮位变化中的短时波动和长期趋势, 实现了对复杂潮汐动态的准确预测。实验结果表明, 该模型在各种水文条件下都能保持较高的预测精度, 特别是在潮位转折阶段和气象扰动时期表现出优越的性能。

4. 实验设置

整个数据集以 9:1 的比例分为训练集和测试集。本研究将上述模型与 BiLSTM 模型和 BiLSTM-transformer 模型进行了比较, 采用相关研究中广泛使用的评价指标, 并将均方根误差 (MSE)、平均绝对误差 (MAE) 作为模型预精度判定的评价标准。实验参数配置见表 1。

在天生港站实测数据验证中, 该框架展现出卓越性能, 特别是在大潮期和小潮期的过渡阶段, 模型能够准确捕捉潮位变化的细微特征, 有效解决了复杂水文条件下潮位预测的多尺度耦合问题。

二、结果分析

1. 实验结果

经过对水文预测模型性能的深入分析, 本研究通过可视化对比和定量评估揭示了不同模型在长江下游水文预测中的表现特征。

图 2 是天生港潮位站的实测潮位曲线。由于数据周期较长, 故实验中随机选取一段用来与预测数据进行对比, 使得对比效果更加清晰。

图 3、图 4、图 5 分别展示了三种不同模型的预测

表1 参数配置表

模块	参数名称	参数值/配置	说明
transformer 编码器	num_layers	1	单层编码器
	num_heads	4	4头注意力机制
	d_model	64	位置编码维度
学习率策略	initial_learning_rate	0.001	初始学习率
	final_learning_rate	$0.01 \times \text{learning_rate}$	终止学习率
	decay_schedule	$(\text{total_epochs} - \text{epoch}) / \text{total_epochs}$	线性衰减策略
	total_epochs	100	总训练轮数
CNN 模块	cnn_channels(首层)	32	首层卷积通道数
	kernel_size	3×1	卷积核尺寸
	padding	1	对称填充
	activation	ReLU	激活函数
	cnn_channels(第二层)	64	第二层卷积通道数
	adaptive_pooling_size	16	自适应最大池化输出尺寸
BiLSTM 模块	hidden_size1	64	单层双向 LSTM 隐藏单元数
	bidirectional	True	双向结构
	batch_first	True	批处理优先模式
	output_dim	128 (64×2)	双向输出拼接后的维度
位置编码	max_len	105 000	最大序列长度
	div_term	$\exp(-\log(10000) / d_model)$	对数间隔频率项
正则化	dropout_rate	0.01	Dropout 随机失活率
输出层	fc_input_dim	128	全连接层输入维度
	fc_output_dim	1	输出维度 (预测值)

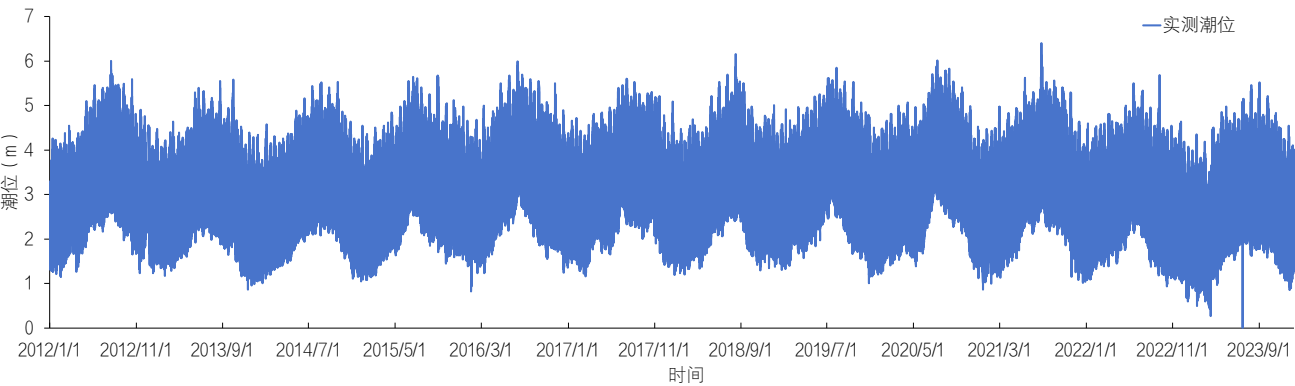


图2 天生港潮位站的实测潮位曲线

水位值与实际观测值的对比情况。可以直观发现，相较于其他两种模型，本文提出的CNN-BiLSTM-transformer模型在预测精度上实现了显著提升。具体表现为，该模型的预测曲线与实际观测曲线在形态上高度一致，两者之间的偏离度极小，尤其是在关键时间节点和水位波动剧烈的区域，这种重合度更为明显，这表明模型对水位变化的捕捉能力更强，预测结果更

为可靠。

图6的箱线图则从量化角度进一步验证了上述观察。该图清晰地对比了不同模型预测误差的分布情况。箱线图显示，本研究提出的CNN-BiLSTM-transformer模型的误差箱体更窄，且中位数更接近0，同时上下须的范围也相对较小，甚至可能没有异常值。这量化地表明，该模型预测结果不仅平均误差更小，

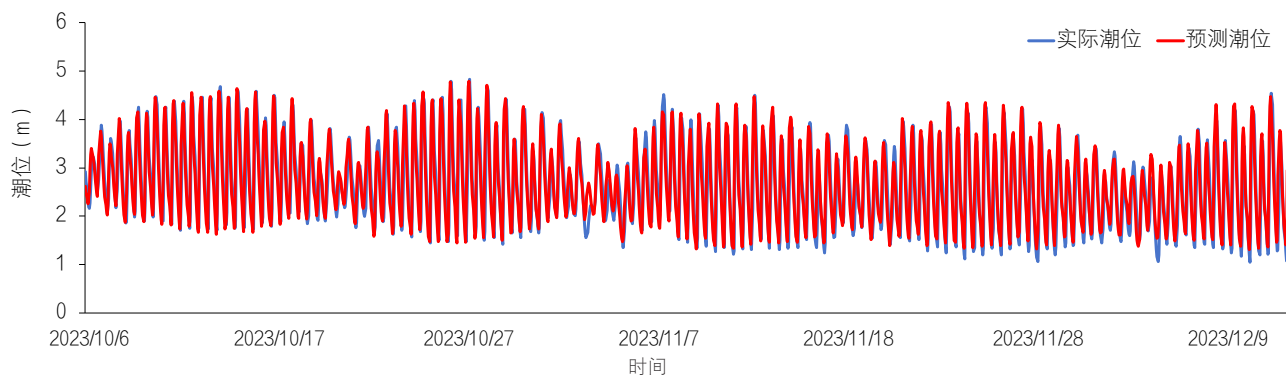


图3 BiLSTM模型预测值与实测值对比

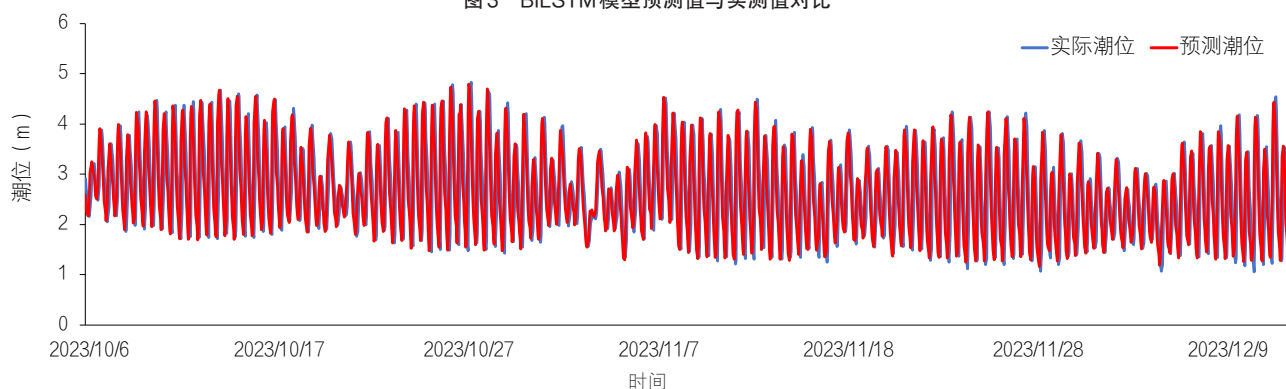


图4 BiLSTM-transformer模型预测值与实测值对比

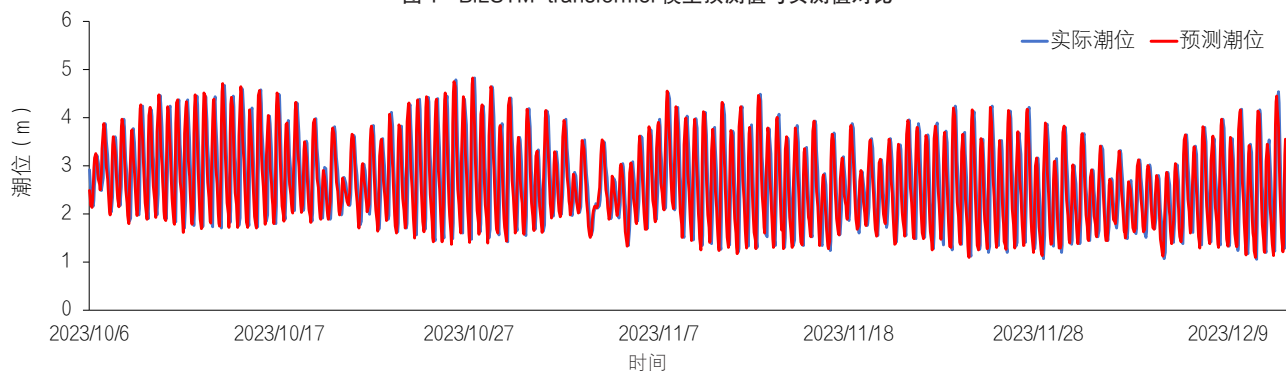


图5 CNN-BiLSTM-transformer模型预测值与实测值对比

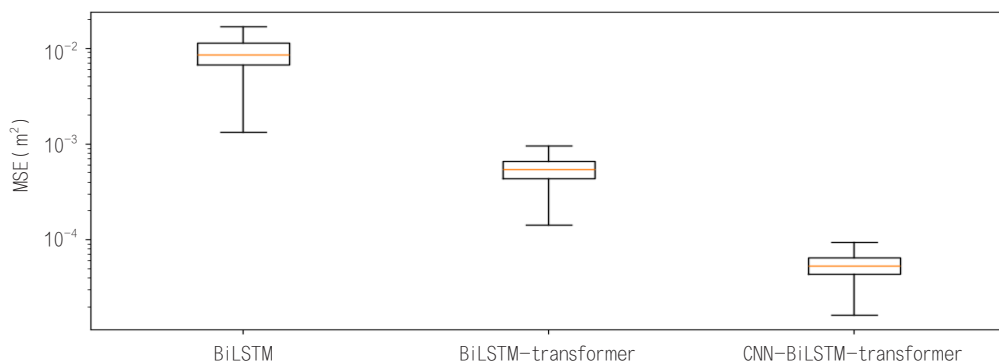


图6 三种模型的预测误差对比

而且稳定性更高,误差分布更集中,波动性更低。特别是在处理典型特征值(如高低潮的预测)时,这种高精度和稳定性表现得尤为突出,验证了融合时空特

征的CNN-BiLSTM-transformer模型在潮汐预测任务中的优越性能。

相较之下,传统双向长短期记忆网络模型(BiLSTM)预测值较实测数据有一定的误差,与本研究提出的模型相比呈现数量级增长;而结合位置编码的BiLSTM-transformer混合模型虽在精度上比BiLSTM有所提升,但仍存在系统性偏差,其均方根误差(MSE)较本研究模型高82.70%。

本研究中模型架构的改进对长时序预测性能提升尤为显著。如图7所示,在200 h延伸预测任务中,CNN-BiLSTM-transformer模型的预测误差增长速率(-0.017 m/h)显著低于对比模型(BiLSTM为0.023 m/h,BiLSTM-transformer为0.022 m/h)。这一现象揭示了传统循环神经网络在长期记忆保持方面的固有缺陷,以及现有transformer架构在水文时序特征提取中的局限性。具体而言,BiLSTM模型由于梯度消失问题导致超过60 h的预测可靠性显著下降,而BiLSTM-transformer模型虽通过注意力机制增强了全局特征捕捉能力,但未能有效处理水文序列中突变事件的局部关联特性。

表2的定量分析进一步验证了模型的鲁棒性。在24 h到80 h的预测过程中,BiLSTM模型MSE值逐渐增加,从 $1.53 \times 10^{-3} \text{ m}^2$ 上升到 $2.67 \times 10^{-3} \text{ m}^2$ 。这表明随着预测时间的增长,BiLSTM模型的预测准确性有所下降,这是由于仅有BiLSTM难以捕捉长时间序列中的复杂模式。

BiLSTM-transformer模型 的MSE值虽然低于BiLSTM模型,但也呈现出随预测时长增加先降低后上升的趋势,从 $2.15 \times 10^{-4} \text{ m}^2$ 增加到 $3.13 \times 10^{-4} \text{ m}^2$ 。这说明尽管引入了transformer结构,BiLSTM-transformer模型在面对更长的时间跨度时尚有不足。

CNN-BiLSTM-transformer模型的MSE值随着预测时长的增加反而减小,从 $4.53 \times 10^{-5} \text{ m}^2$ 降低到 $2.21 \times 10^{-5} \text{ m}^2$ 。造成这种现象的原因是多方面的。

BiLSTM模型的结构仅包含双向长短期记忆网络,虽然能够捕捉时序数据中的双向依赖关系,但对于长期依赖的处理能力却显得捉襟见肘。随着预测周期的延长,水文过程的复杂性逐渐增加,BiLSTM模型难以有效应对这种复杂性,

其预测误差自然呈现出上升的趋势。

相比之下,BiLSTM-transformer模型在BiLSTM的基础上加入了transformer模块,引入了注意力机制。然而,这种改进可能并未充分整合多尺度特征提取,导致在长期预测中对局部关联特征的捕捉不够精准。因此,尽管BiLSTM-transformer模型在短期预测中表现尚可,但随着预测周期的延长,其预测误差同样呈现出上升的趋势。

而CNN-BiLSTM-transformer模型则巧妙地结合了CNN、BiLSTM和transformer的优势。CNN模块能够从潮位序列中提取多尺度的空间特征,这些特征对于理解潮位变化的局部模式至关重要。BiLSTM模块则负责捕捉时序动态,进一步丰富了特征表示。而transformer模块则建立全局的时间依赖关系,解决了传统RNN因梯度消失导致的长周期特征衰减问题。此外,位置编码模块将监测站点的空间与潮汐周期信息嵌入transformer注意力机制,进一步增强了模型对长期依赖关系的建模能力。

然而,在预见期达到大约240 h之后,CNN-BiLSTM-transformer模型预测误差开始呈现波动上升趋势,如图8所示。这一现象可能由以下几个因素导致:首先,随着预测周期的延长,天文潮与气象潮的非线性相互作用加剧,模型对超长期复杂耦合效应的建模能

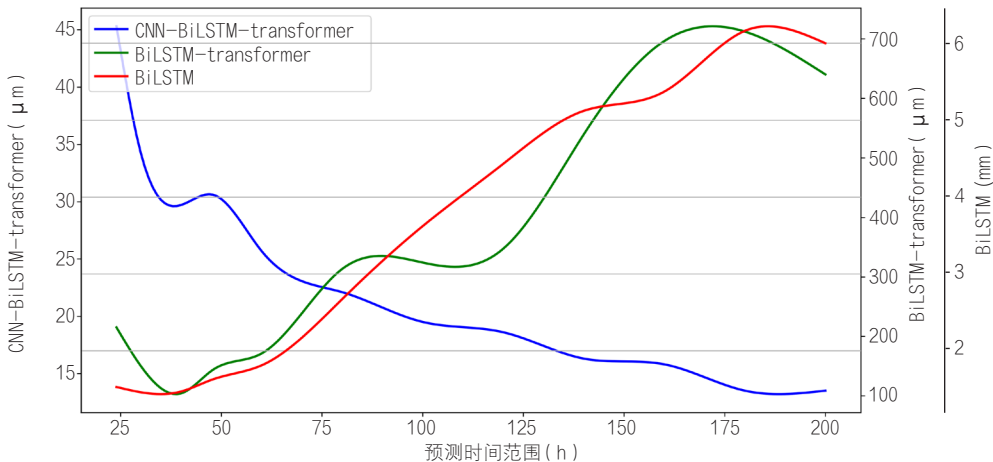


图7 三种模型的长时序预测误差对比

表2 不同模型在不同预测时长下的误差对比

模型	MSE($\times 10^{-4} \text{ m}^2$)			
	预测时长24 h	预测时长48 h	预测时长60 h	预测时长80 h
BiLSTM	15.3	16.3	18.1	26.7
BiLSTM-transformer	2.15	1.44	1.71	3.13
CNN-BiLSTM-transformer	0.453	0.306	0.257	0.221

力尚有不足;其次,位置编码的周期性假设在超长期预测中可能出现偏差,导致相位误差累积;最后,训练数据中极端水文事件(如风暴潮)的样本不足,影响了模型在超长期预测中的泛化能力。

综上,CNN-BiLSTM-transformer模型在结构设计、特征提取、长期依赖处理和位置编码等方面均展现出显著优势。这些优势使得该模型在预测周期延长时能够更好地应对复杂的水文过程,捕捉和理解潮位变化的规律,从而实现预测误差的下降。特别是在潮位转折阶段和气象扰动时期,该模型表现出优越的性能,进一步验证了其鲁棒性和准确性。

此外,研究结果表明,模型各组件之间存在显著的协同互补效应。具体而言,当仅采用BiLSTM模块进行建模时,虽然该模块能够有效提取潮位序列的双向时序关联特征,但由于其固有的循环神经网络结构限制,在长程依赖关系建模方面存在明显不足,这一缺陷直接导致模型在短期预测(24 h至48 h时间尺度)中的精度下降,MSE增加了6.54%。相比之下,引入transformer模块后,其特有的自注意力机制能够全局捕捉水位变化过程中的长程时空依赖关系,特别是对水位渐变过程的精细化建模能力显著提升。这种模块间的功能互补特性形成了良好的协同效应,使得完整模型在应对复杂水文条件(如极端天气事件、河口地形变化等)时展现出卓越的综合预测性能,其预测准

确性显著优于当前主流的技术方案。

2. 消融实验

为系统评估模型架构中关键组件的贡献度,本研究设计了层次化消融实验方案。如表3所示,基于天生港潮位站的实测数据,分别测试了不同模块组合对水位预测性能的影响,验证各组间MSE差异的统计学显著性。

实验结果表明,模型组件的协同作用对预测精度具有决定性影响。例如,完整架构(CNN-BiLSTM-transformer)在60 h预见期的MSE为 $2.57 \times 10^{-5} \text{ m}^2$,显著优于单一模块架构(BiLSTM为 $1.81 \times 10^{-3} \text{ m}^2$,BiLSTM-transformer为 $1.71 \times 10^{-4} \text{ m}^2$)。特别值得注意的是,当移除CNN特征提取模块时,模型的空间特征捕捉能力显著下降,导致MSE急剧上升至 $1.71 \times 10^{-4} \text{ m}^2$ (增幅82.70%),这印证了CNN在提取空间性特征中的核心作用。进一步分析显示,CNN模块的缺失会削弱模型的响应灵敏度,使预测误差显著增加。

多模块融合机制的有效性在研究中得到充分体现。对比实验表明,单独使用BiLSTM-transformer架构时的预测误差数量级在 10^{-4} 。而引入CNN模块后,模型通过空间卷积核实现了数据的特征融合,预测误差数量级降低到 10^{-5} ,预测精度提升了45.26%。这得益于CNN的局部感受野特性可有效捕捉流域内的空间关联模式,而transformer注意力机制则强化了上下游水位传播时序依赖关系的建模。

三、结论与未来研究方向

本研究针对长江下游河段水位预测精度提升需求,提出一种基于transformer架构的混合递归预测模型。该模型通过多模块协同机制实现水文特征的有效提取:采用卷积神经网络(CNN)捕捉潮位序列的局部时空关联特征,利用双向长短期记忆网络(BiLSTM)建模水位序列的长程时间依赖性,并引入transformer的注意力机制解析全局依赖关系。值得注意的是,模型在transformer层中创新性嵌入位置编码策略,显著提升了潮位相位特征的识别精度。基于天生港潮位监测站的实测数据验证表明,本模型较传统的BiLSTM基准模型在不同预见期内的预测精度显著提升,在24 h、48 h、60 h、80 h预见期预测中,均方根误差(MSE)分别降低了88.38%、89.85%、91.04%、93.90%。

潮位预测精度的提升不仅可优化洪水预警系统的响应时效,还能为航运交通提供决策支持。当前模型已整合局部时空特征,未来若进一步融合高分辨率气

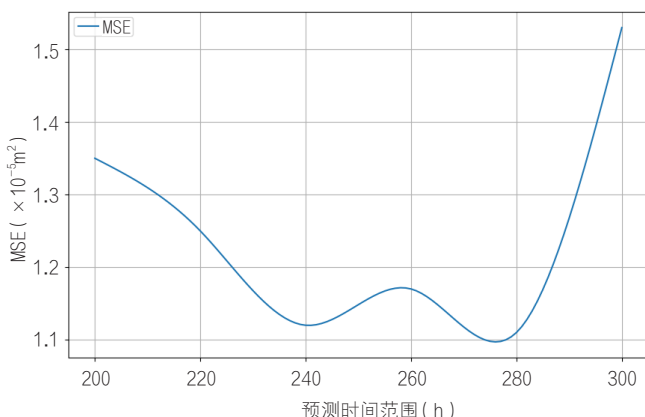


图8 CNN-BiLSTM-transformer模型的预测误差曲线

表3 不同模块配置的消融实验结果

LSTM	Transformer	CNN	MSE(m^2)	MAE(m)
✓			1.81×10^{-3}	1.43×10^{-3}
✓	✓		1.71×10^{-4}	1.37×10^{-4}
✓	✓	✓	2.57×10^{-5}	1.98×10^{-5}

象预报数据、地质构造特征及数字高程模型 (DEM) 等多元环境因子, 可有望突破复杂水文耦合效应的量化瓶颈。后续研究将进一步构建多源异构数据融合框架, 量化降雨、径流对潮汐波动的非线性影响; 开发自适应特征选择机制, 动态识别不同水文情景下的主导影响因子。这些技术路径的实施将有效提升模型在极端水文事件中的泛化能力, 为长江大保护战略实施提供更精准的技术支撑。

参考文献:

- [1] 杨晓坤, 唐刚, 李大刚. 基于 LSTM 神经网络的潮位预报及分析[J]. 港口科技, 2022(5): 39-44+46.
- [2] 俞肇元, 袁林旺, 谢志仁, 等. 基于 SSA 和 AR 模型的海面变化预测试验[J]. 海洋湖沼通报, 2007(4): 14-20.
- [3] 欧素英, 杨清书. 人工神经网络模型在航道、港口潮水位预报中的应用[J]. 水利水运工程学报, 2008(2): 67-70.
- [4] PAN M, ZHOU H, CAO J, et al. Water level prediction model based on GRU and CNN[J]. IEEE Access, 2020, 8: 60090-60100.
- [5] BAEK S S, PYO J, CHUN J A. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach[J]. Water, 2020, 12(12): 3399.
- [6] NIE Q, WAN D, WANG R. CNN-BiLSTM water level prediction method with attention mechanism[C]// Journal of Physics: Conference Series. IOP Publishing, 2021, 2078(1): 012032.
- [7] ZHANG X, WANG T, WANG W, et al. A multi-site tide level prediction model based on graph convolutional recurrent networks[J]. Ocean Engineering, 2023, 269: 113579.
- [8] XU G, CHENG Y, LIU F, et al. A water level prediction model based on ARIMA-RNN[C]// 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (Big Data Service). IEEE, 2019: 221-226.
- [9] REN T, LIU X, NIU J, 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.
- [10] XU J, FAN H, LUO M, et al. Transformer based water level prediction in Poyang Lake, China[J]. Water, 2023, 15(3): 576.
- [11] LI Y, LIU S, ZHANG Y, et al. A Ladder Water Level Prediction Model for the Yangtze River Based on Transfer Learning and Transformer[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024.
- [12] SUN W, CHANG L C, CHANG F J. Deep dive into predictive excellence: Transformer's impact on groundwater level prediction[J]. Journal of Hydrology, 2024, 636: 131250.
- [13] ZHANG Q, GE M, ZHU H, et al. An empirical study on the impact of positional encoding in transformer-based monaural speech enhancement[C]// ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024: 1001-1005.
- [14] WU K, PENG H, CHEN M, et al. Rethinking and improving relative position encoding for vision transformer[C]// Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10033-10041.
- [15] CHEN Y, ZHANG Z, ZHONG W, et al. A Multi-step Water Level Prediction Model Based On CNN-LSTM-ATTENTION Combined With Wavelet Transform[C]// 2024 4th International Conference on Neural Networks, Information and Communication Engineering (NNICE). IEEE, 2024: 992-996.
- [16] GOLOVNEVA O, WANG T, WESTON J, et al. Contextual Position Encoding: Learning to Count What's Important[DB/OL]. (2024-05-29)[2025-05-26]. <https://arxiv.org/abs/2405.18719>.
- [17] 王振宇, 赵倩. 基于三维地球模型的中国大陆重力潮汐因子研究[J]. 地球与行星物理论评(中英文), 2025, 56(1): 83-93.
- [18] 战鹏祥, 黄飞虎, 廖思睿, 等. 基于物理机理引导的数据驱动潮流计算方法[J]. 电网技术, 2024, 48(12): 5034-5045.
- [19] 李明昌, 梁书秀, 孙昭晨, 等. 基于数据驱动模型的潮位和潮流预测方法研究[J]. 北京理工大学学报, 2010, 30(7): 864-868.

责任编辑 王 慧