

融合并行时空注意力机制的 LSTM 模型预测黄河潼关含沙量

李珂^{1,2,3}, 李涛^{1,2,3}, 王敏^{1,2,3}, 崔培⁴, 郭焱^{1,2,3}, 祝宾皓^{1,2,3}

(1.黄河水利委员会黄河水利科学研究院, 郑州 450000;

2.河南省智慧水利工程技术研究中心, 郑州 450003;

3. 黄河实验室, 郑州 450003;

4. 水利部小浪底水利枢纽管理中心, 郑州 450099)

摘要: 含沙量过程的准确预测对于水库防汛调度及泥沙管理具有关键作用。以黄河中下游潼关站为研究对象, 构建了融合并行时空注意力机制的长短期记忆网络模型 (Parallel Spatio-Temporal Attention LSTM, PSTA-LSTM)。模型在结构上集成并行时空注意力模块, 同时捕捉含沙量时间序列中的长期时间依赖性和不同流域站点间的空间关联性; 在隐层激活阶段引入改进的自适应分段线性函数 (Smooth-ReLU, SReLU), 提升模型在含沙量快速变化条件下及峰值的特征表达能力。基于 2000–2024 年实测水沙数据进行实验验证, 结果显示, 相较于传统串行 LSTM 结构, 引入并行时空注意力模块可使整体均方根误差 (Root Mean Square Error, RMSE) 下降约 25.6%, 沙峰预测精度 PRE 提升约 12.7%; 引入 SReLU 激活函数显著提升了沙峰的预测精度, 相较于传统的 ReLU 和 softplus, SReLU 能够更有效地处理沙峰值, 纳什效率系数 (Nash-Sutcliffe efficiency coefficient, NSE) 提升超 9%。将样本根据水沙类型划分为丰水丰沙、中水中沙、枯水低沙、丰水低沙和枯水丰沙五类, 并在不同水沙条件下以及汛期与非汛期进行含沙量预测, 验证了按水沙类型分类训练可进一步提高预测精度, RMSE 下降约 15.7%, NSE 达到 80% 以上; 汛期与非汛期对比显示, PSTA-LSTM 在含沙量峰值明显且短时变化幅度较大的阶段具有更强的响应能力。

关键词: 含沙量预测; LSTM 模型; 并行注意力机制; SReLU 激活函数; 黄河中下游

On the sediment concentration prediction at the Tongguan Section of the Yellow River based on parallel spatiotemporal attention LSTM

Li Ke^{1,2,3}, Li Tao^{1,2,3}, Wang Min^{1,2,3}, Cui Pei⁴, Guo Ye^{1,2,3}, Zhu Binhao^{1,2,3}

1: Yellow River Institute of Hydraulic Research, Zhengzhou 450003, China;

2 Engineering Technology Center of Henan Province Smart Water Conservancy, Yellow River Institute of Hydraulic Research, YRCC, Zhengzhou 450003, China;

3. Yellow River Laboratory, Zhengzhou 45000, China;

4. Xiaolangdi Project Construction & Management Center, Ministry of Water Resources, Zhengzhou 450099, China)

Abstract: Accurate prediction of sediment concentration dynamics is critical for effective reservoir flood regulation, sediment management, and ecological conservation in sediment-rich river basins such as the Yellow River. Reliable sediment concentration forecasting is essential to mitigate adverse impacts of reservoir sedimentation, optimize flood control operations, safeguard

*2025-11-10 收稿; 2025-12-31 收修改稿。

国家重点研发计划项目 (2023YFC3209203、2023YFC3209305)、河南省自然科学基金 (252300423358) 黄河水利科学研究院基本业务项目 (HKY-JBYW-2024-12、HKY-YF-2024-05)、数字孪生小浪底项目 (XLDYK22013) 联合资助。

**通信作者: E-mail: lemonlls@163.com

infrastructure safety, and support water quality management. This study investigates sediment concentration dynamics at Tongguan Hydrological Station in the middle and lower Yellow River, a key monitoring site at the confluence of the mainstream with the Weihe and Beiluohe Rivers, immediately upstream of the Sanmenxia Reservoir, thus strongly influencing downstream sedimentation and flood management. To address the dynamic and complex flow-sediment regimes of the Yellow River, this study develops a Parallel Spatio-Temporal Attention Long Short-Term Memory (PSTA-LSTM) model specifically for sediment concentration forecasting. The PSTA-LSTM integrates a parallel spatio-temporal attention mechanism enabling simultaneous capture of multiscale temporal dependencies and spatial correlations across monitoring sites, markedly enhancing simulation of sediment transport processes. An adaptive segmented rectified linear unit (SReLU) is embedded in hidden layers to strengthen nonlinear feature learning and improve representation of rapid sediment concentration fluctuations, especially during high-magnitude sediment peak events, effectively accommodating extreme sediment load variability driven by heterogeneous flow, tributary inputs, and local erosion-deposition dynamics. Model validation used observed hydrological and sediment datasets from 2000 to 2024. Results show that compared with conventional serial LSTM, the parallel spatio-temporal attention mechanism reduces overall Root Mean Square Error (RMSE) by approximately 25.6% and increases Peak Sediment Prediction Accuracy (PRE) by approximately 12.7%. Incorporation of SReLU significantly boosts peak prediction, with Nash-Sutcliffe Efficiency (NSE) improved by over 9%, verifying effectiveness in resolving extreme sediment values. Tongguan Station serves as a critical sediment control section for the Sanmenxia Reservoir and plays a pivotal role in reservoir regulation and downstream transport. Upstream water and sediment mainly derive from the Yellow River above Longmen Station, the Weihe at Huaxian Station, and the Beiluohe at Zhuangtuo Station. Mainstream floods show rapid rise-fall dynamics and high sediment concentrations; Weihe floods have longer duration and flattened peaks; Beiluohe floods feature sharp narrow peaks, high sediment concentrations, and fast-propagating sediment waves. Confluence and superposition of multi-source floods at Tongguan modulate transport intensity and define temporal variability of sediment concentration. Experiments used 2000–2024 data classified into five flow-sediment regimes: high-flow/high-sediment, medium-flow/medium-sediment, low-flow/low-sediment, high-flow/low-sediment, and low-flow/high-sediment. The model was trained and validated under each regime and compared between flood and non-flood seasons to assess adaptability. Results confirm the attention mechanism improves performance with RMSE reduced by 25.6% and PRE increased by 12.7%. SReLU increases NSE by 6–11% relative to traditional ReLU and softplus, showing superior capability in capturing extreme peaks. Regime-specific training further elevates accuracy, reducing RMSE by approximately 15.7% and achieving NSE over 80%. Inter-seasonal comparisons reveal stronger responsiveness during flood seasons, especially under pronounced peaks and rapid fluctuations. These findings validate the model's adaptability to highly dynamic and complex sediment environments and highlight its potential for real-time sediment concentration forecasting in large high-sediment rivers such as the Yellow River.

Keywords: Sediment concentration prediction; LSTM model; parallel attention mechanism; SReLU activation function; middle and lower reaches of the Yellow River

在黄河中游，潼关站是最具代表性的控制性水文泥沙观测断面，被视为进入下游河段的“总控制口门”。潼关含沙量直接反映了渭河、北洛河及黄土丘陵区等主要沙源区的输沙强度，是判断中游来沙状况及其对下游河道淤积演变和行洪能力影响的重要依据^[1-2]。该断面的含沙量过程在汛期尤其敏感，常受暴雨产沙、支流集中来沙以及河道局地冲刷等多因素共同作用而出现短时快速上升，其突变幅度和到达时间对下游调度安全具有直接影响^[3-4]。因此，提升潼关站含沙量的预测能力，对于掌握中游来沙态势、研判下游河道淤积风险以及支撑黄河水沙调度体系的精细化管理具有重要意义。

目前国内外针对含沙量预测已有大量研究，主要方法包括基于物理过程的预测方法、传统经验回归法和机器学习方法。基于物理过程的预测方法（如一维和二维水沙动力学模型）通过耦合水流连续方程、动量方程及泥沙输移方程来模拟含沙量的演变过程^[5-7]，能够从机理上反映输沙过程。但模型在实际应用中通常需要依赖高精度的边界条件、河道地形及糙率等物理参数。对于汛期或洪水过程，上游来水来沙条件、河道糙率与局部冲淤状况存在不确定性，这些因素会降低物理模型在快速变化情境下的响应能力。传统经验回归法不刻画泥沙演变的物理机制^[8-9]，通常利用统计方法建立含沙量与主要影响因子之间的经验关系，在多年平均特征分析中具有一定适用性，如黄河流域的多来多排类经验公式。但由于此类方法依赖预设的函数形式，在含沙量变化较大情况下预测精度会有所下降。

机器学习方法的引入为含沙量预测任务带来了新的视角^[10-11]。由于机器学习模型基于历史数据学习系统的复杂关系，无需依赖先验的工程物理模型，因此有助于克服因工程模型不完善或不准确所带来的问题^[12-14]。郭秀吉等^[15]采用多种机器学习算法对三门峡出库含沙量进行预测，减少对先验物理模型的依

赖,有效提升对常规水文条件下含沙量变化趋势的捕捉能力。陈雪等^[16]选用梯度提升回归、支持向量回归、极端梯度提升和随机森林四种机器学习模型,对延川和子长水文站未来3小时泥沙进行预测,挖掘水文与泥沙数据中的映射关系和多因子耦合特征,提高短时预测精度。尽管机器学习方法及其改进算法在含沙量预测领域取得了一定成果,尤其在单一维度方面。但对于跨断面空间关联、汛期含沙量急剧上升的洪水事件或短历时排沙事件等场景下的耦合效应捕捉不足,限制了其在复杂流域水沙预测中的全面应用^[17]。研究方法逐步从传统经验和浅层机器学习向深度学习等能够更好地刻画时空非线性关系的方向演进。

深度学习方法的引入能够有效利用多维数据。Zaremba等^[18]提出的循环神经网络(Recurrent Neural Network, RNN)在时间复杂度和准确性方面表现出显著优势。然而,随着数据拟合难度增加和网络层数增多,优化过程中往往会出现梯度消失和梯度爆炸问题。长短期记忆网络(Long Short-Term Memory, LSTM)^[19]的出现有效缓解了这些问题,在时间序列分析任务中展现出更强大的能力。在含沙量预测研究中,LSTM能够较好地捕捉重点断面水沙序列的长期依赖特征。林天宙等^[20]使用LSTM模型对寸滩站汛期日含沙量进行预测,缓解了RNN的梯度消失和梯度爆炸问题。魏苗^[21]基于LSTM算法构建含沙量短临预报模型,实现入库含沙量短临预报,验证了该模型在含沙量预测中相比传统机器学习方法具有更高的准确性和更强的动态响应能力。尽管LSTM有助于捕捉长期依赖关系,但它无法在不同时间步对不同的变量给予关注。在含沙量预测研究中,很多工作仅考虑序列数据本身的时间关联^[22-23],或利用单一水文站的观测数据推演未来^[24-25],主要强调时序特征,而较少涉及上游来水来沙对重点断面水沙过程的影响。这种忽视时空耦合的建模方式,使得模型在含沙量急剧上升的洪水事件下预测精度受限。

近年来广泛应用于自然语言处理和计算机视觉领域的注意力机制被引入水文建模中,它能够通过学习自动分配权重,突出关键时间段与特征变量,从而显著提升模型在时空信息捕捉与非线性建模方面的表现。已有研究表明,引入注意力机制的水文预测模型在水文指标建模中能够有效提升预测精度与关键时段敏感性。例如,Dai等^[26]构建LSTM-注意力机制模型预测多个测站水位数据,其RMSE与NSE优于无注意力模型;Zhao等^[27]利用注意力模型提升了在强降雨或洪峰等条件下的流量预测精度;Hu等^[28]引入空间与时间注意力,较好地考虑了上游测站间的空间关联。与流量或水位相比,含沙量过程往往呈现更强的突变性和阶段性特征,尤其在暴雨径流集中的时段,含沙量常在较短时间内出现急剧上升,峰值高、历时短,对模型的敏感性提出更高要求。在串行式注意力结构中,时空信息需逐层向后传递,容易在快速变化或峰值突出的序列中出现权重偏移或信息衰减,导致模型对关键时段响应不足和无法准确捕捉关键站点变化对目标变量的影响。基于此,将注意力结构设计为并行的时间与空间分支,使模型能够分别从时间序列变化和上游多站点空间关联中提取信息,并在后续过程实现有效融合,对于提高潼关含沙量预测精度具有现实意义。

在黄河等多沙河流中,含沙量变化受上游来水过程、河道冲淤演变或支流来沙的共同影响,其时间序列往往呈现明显的非线性特征。激活函数作为深度神经网络的关键组件,能够为模型引入非线性转换,使其在特征提取与模式识别过程中更好地拟合复杂的水沙响应关系。ReLU^[29]因其计算效率高且具备稀疏激活特性,被广泛应用于各类深度神经网络,但其对负值的硬截断可能导致关键信息丢失。softplus等平滑激活函数通过缓解ReLU在负区间的硬截断问题,改善了梯度传递的连续性^[30],但其对负值的压缩可能削弱对含沙量峰值的响应能力。因此,需要设计一种能够融合平滑特性与峰值响应能力的激活函数。

为解决上述问题,提出了一种面向黄河潼关断面含沙量预测的深度学习方法。该方法以LSTM模型为核心,结合同步时空注意力机制,同时改进激活函数,提高对含沙量的预测能力。本研究的主要创新点包括:(1)提出并行时空注意力结构,分别通过时间注意力机制捕捉历史时序数据中的重要时间依赖关系,使用空间注意力机制对贡献较大的测站赋予更高权重,有效保留时空信息的互补性,提升模型在含沙量波动较大条件下及峰值的预测能力;(2)设计平滑非负激活函数Smooth-ReLU(SReLU),融合ReLU与softplus优势,在保证梯度连续性与训练稳定性的同时,提高模型对含沙量峰值敏感性与拟合精度;(3)以黄河中下游重点断面——潼关站为研究对象,结合历史水文及含沙量数据开展预测验证,系统评估模型的适用性。

1 数据来源与数据预处理

研究对象为潼关水文站含沙量过程,研究区域为黄河中游潼关汇流区。潼关站位于黄河干流与渭河、北洛河等主要支流的汇合区域,是黄河中游进入三门峡水库的关键水沙控制断面,对水库调度及下游泥沙输移具有重要影响。潼关以上来水来沙主要来源于黄河干流龙门站以上流域,以及渭华县站和北洛河状头站控制流域,其中干流洪水过程陡涨陡落、含沙量高,渭河洪水历时较长、峰型偏钝,北洛河洪水则具有峰型尖瘦、含沙量高、沙峰传播快等特征。不同来源洪水在潼关汇流区的汇合与叠加,不仅影响泥沙输移强度,也决定了潼关断面含沙量过程的时序变化特征。鉴于汾河来水量相对较小,对潼关断面水沙过程影响有限,未将其作为模型输入变量。基于上述特征,选取龙门、华县和状头站作为潼关断面上游的控制性水文站点,用于表征主要来沙条件并开展含沙量预测研究。

采用潼关水文站2000—2024年的实测含沙量及上游主要控制站点(龙门、华县、状头)的含沙量、流量数据进行模拟预测研究,数据均来源于黄河水利委员会水文局逐日平均流量表、逐日平均含沙量表及洪水水文要素摘录表,经数据插补至分辨率为逐两小时的数据。为消除量纲差异带来的影响,加快模型训练收敛速度,归一化公式如下:

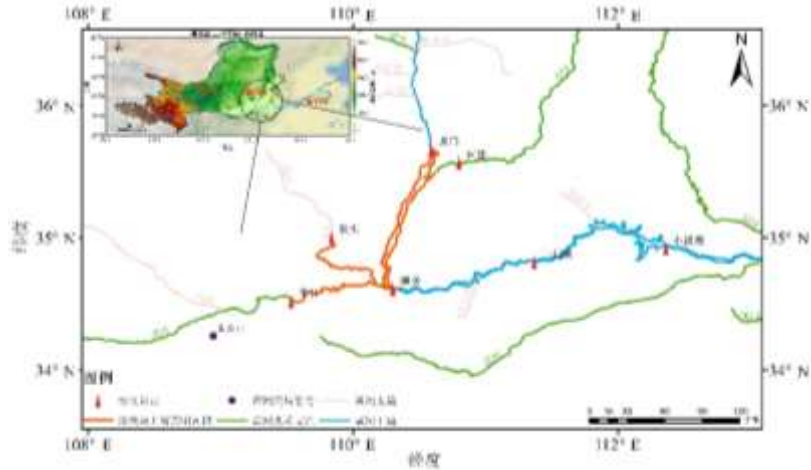


图1 研究区域

Fig.1 Study area

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

其中， x_{norm} 为数据的归一化值； x_{min} 为数据最小值； x_{max} 为数据最大值。

考虑到不同年来水与来沙组合特征对含沙量预测的影响，对研究时段内的水沙特征进行类型划分。

图 2(a)为简化后的水沙类型划分坐标^[31]，图中 W 与 W_s 分别为某年份的总来水量(亿 m^3)与总来沙量(亿 t)， \bar{W} 与 \bar{W}_s 分别为该系列年份总来水量均值(亿 m^3)与总来沙量均值(亿 t)，系列年份指计算平均来水量及来沙量所运用到的所有年份。根据潼关断面 2000—2024（缺少 2014 年数据，因此不做分析）年的来水来沙情况，各年份水沙类型如图 2(b)所示。

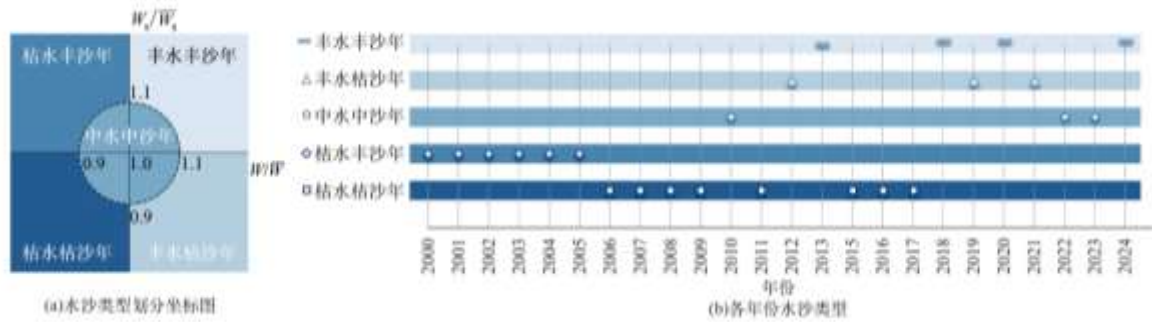


图 2 水沙数据分类

Fig.2 Data classification of water and sediment

在完成水沙类型划分后，依据各年份来水来沙组合特征，对研究时段内的数据进行了训练集和测试集划分。方式 1 中，选取某一年数据作为测试集，其余年份数据作为训练集与验证集。对于水沙分类模型，则在相同类型年份中选取一年作为测试集，其余年份作为训练集与验证集，如在丰水丰沙类型中选取 2024 年为测试集，2013、2018 和 2020 年数据作为训练与验证数据。

表 1 数据集划分情况

Tab.1 Dataset classification and splitting

划分方式	水沙类型	训练集及验证集	测试集	是否交叉验证
方式1	—	2000-2013、2015-2023	2024	是
方式2	丰水丰沙类	2013, 2018, 2020	2024	是
	丰水枯沙类	2012, 2019	2021	是
	中水中沙类	2010, 2022	2023	是
	枯水丰沙类	2000-2004	2005	是
	枯水枯沙类	2006-2009, 2011,2015,2016	2017	是

2 研究方法

2.1 长短时记忆神经网络

长短时记忆神经网络(Long Short-Term Memory, LSTM)是循环神经网络(Recurrent Neural Network, RNN)的一种特殊类型,将隐藏层的RNN神经元替代为LSTM神经元,遗忘无效信息并更新记忆有效信息。记忆单元结构如图3所示。

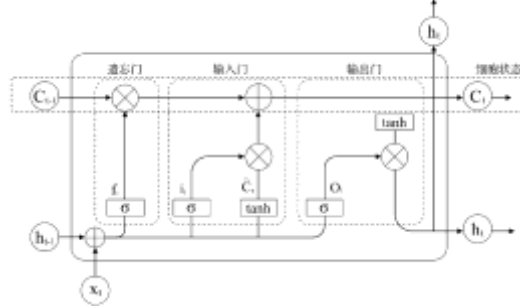


图3 LSTM记忆单元结构

Fig.3 Structure of the LSTM memory cell

LSTM通过构造的三个“门”结构达到更新神经元信息功能。遗忘门用于控制当前时刻是否保留前一时刻的记忆单元状态。如式(2)所示:

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

其中, x_t 是输入向量; h_{t-1} 是上一时刻的隐含状态; f_t 遗忘门输出; W_f 是权重矩阵; b_f 是偏置; σ sigmoid 激活函数。

输入门,对上一状态进行筛选性遗忘的同时,引入当前时刻的重要信息,完成记忆的动态更新。

$$C_t = f_t \square C_{t-1} + i_t \square \tilde{C}_t \quad (3)$$

$$i_t = \sigma(W_i [h_{t-1}, x_t] + b_i) \quad (4)$$

$$\tilde{C}_t = \tanh(W_c [h_{t-1}, x_t] + b_c) \quad (5)$$

其中, C_{t-1} 、 C_t 代表上一时刻与当前时刻的细胞状态; i_t 输入门输出; \tilde{C}_t 候选记忆向量; W_i 和 W_c 是权重矩阵; b_i 和 b_c 是偏置; h_{t-1} 代表上一时刻的隐含状态。

前两个阶段将LSTM的长期记忆与当前记忆相结合,更新单元状态 C_t 。最后是输出门,通过函数 \tanh 处理细胞状态,并将结果值与激活函数输出结果相乘,得到确定会输出的信息。

$$h_t = o_t \square \tanh(C_t) \quad (6)$$

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (7)$$

其中, h_t 当前时刻的隐含状态; C_t 当前时刻的细胞状态; o_t 输出门输出; W_o 权重矩阵; b_o 偏置; h_{t-1} 上一时刻的隐含状态。

2.2 并行时空注意力机制

a. 时间注意力模块

时间注意力机制(Temporal Attention, TA)用于突出序列中对预测目标贡献较大的关键时间步。设长度为 T 的输入序列的LSTM的隐藏状态为:

$$H = [h_1, h_2, \dots, h_T], h_t \in \mathbb{R}^d \quad (5)$$

其中 h_t 表示第 T 个时间步的隐藏状态向量。

计算每个时间步的相关性得分:

$$e_t = \text{ReLU}(W_h h_t + b_h) \quad (6)$$

其中, W_h 是时间注意力的权重矩阵; b_h 为偏置项; e_t 为第 t 个时间步的相关性得分。

通过 Soft max 对 e_t 归一化,生成各特征的注意力权重 β ,并用于加权隐藏状态生成时间上下文向量,如公式(7)和(8)所示。

$$\beta_t = \frac{\exp(e_t)}{\sum_{k=1}^T \exp(e_k)} \quad (7)$$

$$h_{att} = \sum_{t=1}^T \beta_t h_t \quad (8)$$

b. 空间注意力模块

空间注意力机制 (Spatial Attention, SA) 旨在识别不同输入测站对目标站点含沙量的影响程度。设每个时间步的多测站输入向量为:

$$x_t = [f_1^t, f_2^t, \dots, f_m^t]^T \in \mathbb{R}^m \quad (9)$$

其中, f_i^t 表示第 t 个时间步第 i 个测站的特征值, m 为测站数。

计算各测站相关性得分:

$$s_t = W_x x_t + b_x \quad (10)$$

其中, W_x 是时间注意力的权重矩阵; b_x 为偏置项; s_t 为第 t 个时间步的相关性得分。

通过 **Soft max** 对 s_t 归一化, 生成空间注意力权重 α_t , 并用于加权输入向量生成空间上下文向量, 如公式 (11) 和 (12) 所示。

$$\alpha_t = \text{Soft max}(s_t), \quad \alpha_t \in \mathbb{R}^m, \quad \sum_{i=1}^m \alpha_i^t = 1 \quad (11)$$

$$x_t^{\text{SA}} = \alpha_t \odot x_t \quad (12)$$

其中, x_t^{SA} 为空间增强后的输入, 突出关键测站信息; \odot 表示逐元素乘积。

将序列 $\{x_1^{\text{SA}}, x_2^{\text{SA}}, \dots, x_T^{\text{SA}}\}$ 输入至 LSTM, 得到隐藏状态序列 $\{h_1^{(s)}, h_2^{(s)}, \dots, h_T^{(s)}\}, h_t^{(s)} \in \mathbb{R}^d$, d 为 LSTM 隐含层维度。对隐藏状态 $\{h_1^{(s)}, h_2^{(s)}, \dots, h_T^{(s)}\}, h_t^{(s)} \in \mathbb{R}^d$ 采用池化操作获得空间路径的全局表示:

$$h^{(s)} = \text{Pool}(\{h_1^{(s)}, h_2^{(s)}, \dots, h_T^{(s)}\}) \quad (13)$$

其中, Pool 为池化操作, 用于在综合各时间步的空间特征。 $h^{(s)}$ 为空间注意力结果。

(3) 并行融合

采用向量拼接操作进行联合表示:

$$c = \text{Concat}(h^{(s)}, h_{att}^{(s)}), c \in \mathbb{R}^{2d} \quad (14)$$

其中, $h^{(s)}$ 为空间路径的全局表示, $h_{att}^{(s)}$ 为时间路径的上下文表示。

考虑到两类特征存在尺度差异, 拼接后引入一层非线性映射:

$$z = \tanh(W_f c + b_f), z \in \mathbb{R}^{d_f} \quad (15)$$

其中, $W_f \in \mathbb{R}^{d_f \times 2d}$ 为权重, $b_f \in \mathbb{R}^{d_f}$ 为偏置项。 $\tanh(\bullet)$ 用于增强特征交互与非线性表达能力, d_f 为融合后的特征维度。

2.3 TA-LSTM、SA-LSTM 与 SSTA-LSTM 模型

2.3.1 TA-LSTM 模型

时间注意力 LSTM (TA-LSTM) 通过在 LSTM 输出层引入可学习的时间注意力权重, 对所有历史时间步的隐藏状态进行相关性计算和加权。在模型预测当前含沙量时, 每个历史时间步的信息都会参与注意力分配, 但其最终权重由该时间步对当前预测的重要性决定。这样模型能够自动突出具有代表性的含沙量时序特征, 即那些对当前含沙量变化具有关键指示作用的历史过程片段, 增强对水沙过程非线性变化的捕捉。

2.3.2 SA-LSTM 模型

空间注意力 LSTM (SA-LSTM) 在 LSTM 主干网络前引入空间注意力层, 用于潼关上游多个控制站的输入特征的差异化加权。输入的空间特征包括各上游测站的流量和含沙量等关键水沙指标。模型通过计算每个站点特征与潼关含沙量预测任务之间的相关性, 自动学习空间注意力权重向量。权重值越大, 表示该站点的水沙过程对潼关含沙量变化的贡献越高。通过这种加权组合, 模型能够综合反映上游各站来水、来沙的相互作用及其对潼关断面水沙演变的影响, 更准确地刻画多源来沙输入对潼关含沙量的空间耦合效应。

2.3.3 SSTA-LSTM 模型

串行时空注意力结构（Sequential Spatio-Temporal Attention, SSTA）将时间注意力与空间注意力依次嵌入LSTM网络。首先引入空间注意力机制，对同一时刻来自不同上游控制站点的水沙特征进行加权，以突出对潼关断面含沙量贡献更为显著的站点；随后，利用 LSTM 对加权后的水沙序列进行时序建模；在此基础上，引入时间注意力机制，对不同历史时间步的隐藏状态进行重要性分配，进一步突出对当前含沙量预测最具指示意义的关键历史过程片段。

2.4 PSTA-LSTM 模型

PSTA-LSTM（Parallel Spatio-Temporal Attention, PSTA）模型是一种融合空间注意力机制、时间注意力机制与LSTM的混合神经网络结构。与SSTA-LSTM模型不同，PSTA-LSTM模型在空间注意力和时间注意力机制上采用并行结构，使模型能够同时识别输入数据中的关键空间特征和时间依赖关系，动态衡量各测站与历史时刻对目标断面的贡献。

2.4.1 模型建立

以空间注意力机制作为空间特征提取器，以时间注意力机制作为时序特征提取器，LSTM作为特征承载与序列建模单元，用于捕获输入序列的动态依赖。并行注意力机制与LSTM组合的基本思路为：空间与时间注意力支路并行运行，在空间支路中，利用空间注意力机制对输入特征在空间维度上加权计算，形成空间上下文表示；在时间支路中，将输入序列送入LSTM提取隐藏态，再基于隐藏态计算时间注意力权重，并加权聚合形成时间上下文表示。两支路输出的特征在融合层中拼接，并经非线性映射实现特征交互与尺度统一。最终，通过全连接层将融合特征映射为未来若干时刻的含沙量预测结果。

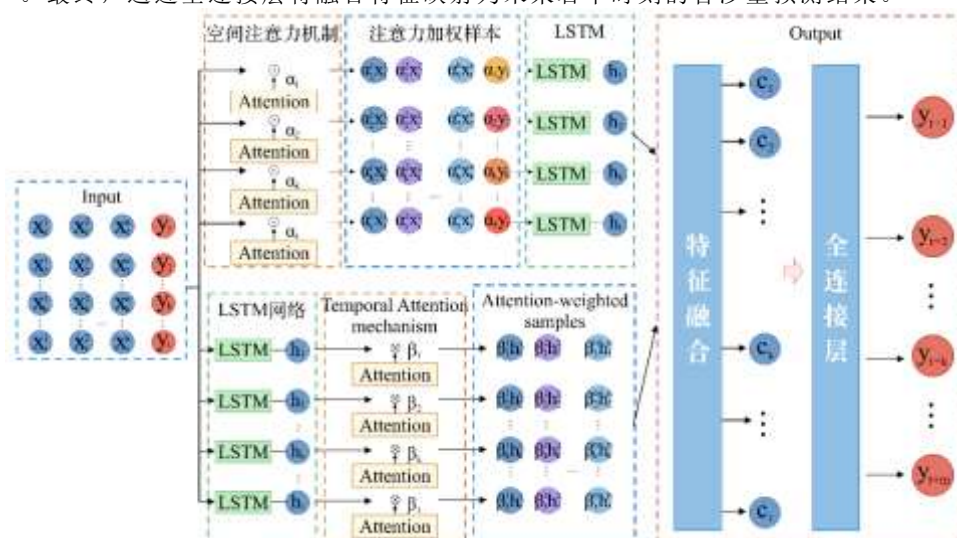


图4 PSTA-LSTM的含沙量预测模型

Fig.4 Sediment concentration prediction model based on PSTA-LSTM

2.4.2 SReLU 激活函数

ReLU函数在低值区的“硬截断”特性导致梯度不连续，影响模型训练的稳定性；softplus 函数在低值区保证梯度连续，但在高值区增长过慢。因此设计一种融合ReLU与softplus优势的平滑非负激活函数Smooth-ReLU（SReLU）。利用softplus处理ReLU的低值区间，保证梯度平滑与非负性。高值区采用线性增强段，类似ReLU对正值的快速响应，捕捉含沙量峰值变化转折点处函数值及一阶导数连续，实现了softplus与ReLU的融合，兼顾训练稳定性与高值响应能力。

$$SReLU = \begin{cases} k(x - \tau) + \ln(1 + e^\tau), & x > \tau \\ \ln(1 + e^x), & x \leq \tau \end{cases} \quad (16)$$

其中， τ 为含沙量操作阈值， kg/m^3 。依据潼关站含沙量数据的统计分布，同时兼顾小浪底水库关闭发电洞的工程操作要求确定其候选范围，再通过交叉验证选取最优值，最终确定为 $42\text{kg}/\text{m}^3$ 。k 为高值区线性段斜率，用于增强对峰值含沙量的敏感性，通过交叉验证确定为1.2。即：

$$SReLU = \begin{cases} 1.2(x - 42) + \ln(1 + e^{42}), & x > 42 \\ \ln(1 + e^x), & x \leq 42 \end{cases} \quad (17)$$

2.5 输入特征构建

潼关站含沙量变化主要受其控制流域内黄河干流、渭河及北洛河来水来沙共同影响。为表征上游来水来沙边界条件，选取三个具有代表性的控制水文站作为模型输入源。龙门站，位于黄河干流进入潼关前的关键节点，反映主河道水沙输入；华县站是渭河下游控制站，表征渭河流域对潼关的水沙贡献；状头站为北洛河出口控制站，代表北洛河水沙汇入过程。

模型输入仅包含各测站的原始观测变量：流量与含沙量。在每个时间步 t ，输入向量定义为：

$$x_t = [Q_{\text{龙门}}(t), S_{\text{龙门}}(t), Q_{\text{华县}}(t), S_{\text{华县}}(t), Q_{\text{状头}}(t), S_{\text{状头}}(t)]^T \in \mathbb{R}^m$$

其中， m 为输入步长，所有输入变量均使用公式（1），将数值映射至[0,1]区间，消除不同量纲差异并提升模型训练稳定性。

基于皮尔逊相关系数，确定模型输入序列长度。假设模型接收输入长度为 m 的时间序列作为输入，即：

$$X = [x_{t-m}, x_{t-m+1}, \dots, x_t] \in \mathbb{R}^{m \times m}$$

2.6 评价指标

采用平均绝对误差(Mean Absolute Error, MAE)、均方根误差(RMSE)、纳什效率系数(NSE)、峰值相对误差(Peak Relative Error, PRE)及决定系数(Coefficient of Determination, R^2)评估模型的预测表现。

MAE用来评估模型预测结果与实际数据的平均偏差，反映模型对总体误差的鲁棒性。

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (18)$$

RMSE量化模型的预测结果与实际数据的方差，反映预测值和真实值的总体误差。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (19)$$

NSE衡量模型预测值相对于观测均值的拟合优度，反映模型对观测数据变异性的解释能力。

$$NSE = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2} \quad (20)$$

PRE衡量模型对时间序列极值的预测准确性，直观反映模型对最大值的拟合精度。

$$PRE = \frac{|\hat{y}_i - P_{obs}|}{P_{obs}} \quad (21)$$

R^2 是衡量模型拟合优度，取值为[0,1]，越接近1表示模型拟合越好。

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (22)$$

上述式中： y_i 为含沙量实测值， kg/m^3 ； \hat{y}_i 为含沙量预测值， kg/m^3 ； \bar{y} 为实测含沙量的平均值， kg/m^3 ； $\hat{\bar{y}}$ 为预测含沙量的平均值， kg/m^3 ； P_{obs} 为实测峰值， kg/m^3 。

3 结果分析

3.1 输入因子相关性特征分析

图 5 给出了潼关含沙量与各上游测站输入因子的皮尔逊相关系数热力图。结果显示，潼关含沙量与龙门站、华县站的含沙量呈显著正相关，相关系数较高，说明黄河干流及渭河来沙对潼关含沙过程具有主导作用。这主要是由于黄河干流和渭河作为主要支流，其流域面积大、汇沙量高，使其含沙变化能够较快且显著地反映在潼关断面。相比，状头站的含沙量与潼关含沙量相关性为中等水平，表明北洛河的来沙对潼关具有阶段性补给作用，其影响可能在局地降雨或枯水期相对增强。

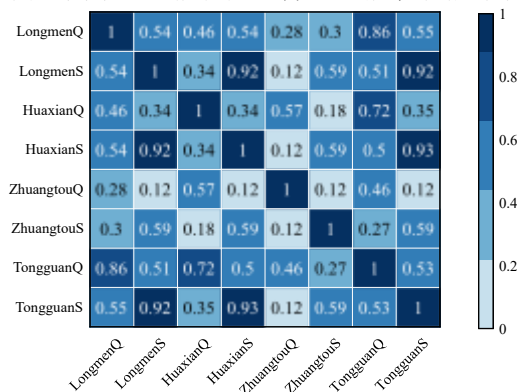


图 5 预报因子相关系数热力图

Fig.5 Correlation heatmap of forecasting factors

3.2 模型参数性能分析

3.2.1 参数确定与多模型统一比较

在实验中，以方式一的数据划分方式进行验证。采用网格搜索法分别测试了不同输入序列长度和隐层层数对模型预测性能的影响，模型输入序列长度最终选定为6步，对应前12小时的观测信息；LSTM隐层层数确定为2层。在此基础上，进一步考察了不同隐含层单元数与训练轮数组合的性能。模型学习率固定为 1×10^{-3} ，优化器采用Adam算法。表2为模型训练结束后，测试集RMSE平均值随LSTM层隐含层单元数和网络迭代轮数组合变化的趋势，并绘制其三维曲面图。结果如图6所示。

从表2与图6可以看出，LSTM的预测精度随着隐含层单元数和训练轮数的增加呈现“先降后升”的趋势。在参数较小，如8个单元、50轮时，模型误差偏高，说明网络容量不足；当参数提升至中等规模时，性能显著改善，RMSE较初始组合降低约55%，其中在64个隐含单元、100轮训练时达到最优，表现出良好的特征表达与泛化能力。但当参数进一步增大时，误差回升，说明复杂度过高。三维曲面图呈现出两端高、中间低的“山谷”形态，说明当模型复杂度过低或过高时，都会导致拟合能力不足或过拟合。故最终确定隐含层单元数为64、训练轮数为100。

表2 不同隐含层单元数和网络迭代轮数组合下 RMSE 均值统计表

Tab.2 Table of RMSE for different combinations of hidden layer units and network training epochs

迭代次数	隐含层单元数					
	8	16	32	64	128	256
50	91.73	82.41	68.56	54.28	59.47	66.23
100	74.12	63.38	48.67	41.29	44.78	52.65
150	75.04	64.85	50.01	42.11	46.37	54.93
200	76.82	66.94	51.34	43.26	48.16	57.42
250	78.69	69.37	53.02	44.73	50.24	60.15

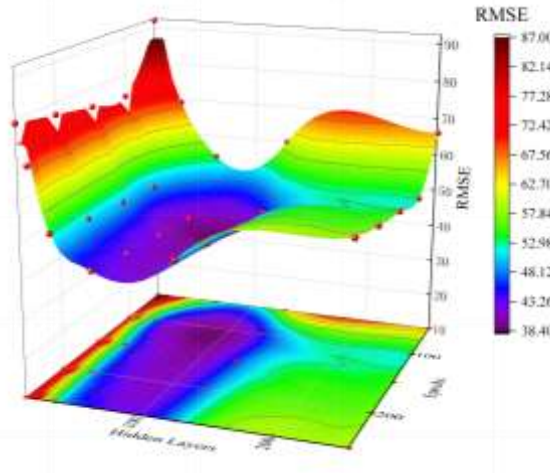


图6 RMSE 随隐含层单元数和网络迭代轮数变化示意图

Fig.6 RMSE variation with hidden layer units and training epochs

当空间注意力维度较低（16或32）时，无论时间维度如何增加，RMSE 始终维持在 41.29-91.73 的区间，整体水平偏高。这表明空间特征在捕捉上游来沙分布及支流汇入差异方面起关键作用，空间信息不足会削弱模型对河道输沙路径与上游来沙贡献的捕捉能力。当空间维度提高至64或128后，模型性能显著改善，RMSE下降至 42.11-51.34，说明高维空间特征有助于增强跨测站信息的耦合和对主要来沙测站贡献的区分能力，使模型能够充分反映上游来沙在河道内的叠加效应及其对下游含沙量变化的影响作用。时间注意力维度的增加对预测精度改善有限，这是因为潼关站含沙量对上游变化的响应主要集中在最近若干时间步，超过该范围后历史信息的贡献显著下降。因此，扩大时间维度对精度提升作用有限。

表3 空间注意力维度和时间注意力维度组合下 RMSE 均值统计表

Tab.3 Table of RMSE for combinations of spatial and temporal attention dimensions

空间注意力维度	时间注意力维度					
	8	16	32	64	128	256
8	98.0	92.5	85.3	82.0	81.5	83.0
16	91.7	82.4	68.6	65.2	64.8	66.0
32	74.1	63.4	58.5	57.3	56.9	58.2
64	75.0	64.9	46.8	48.0	47.5	49.0
128	76.8	66.9	49.0	49.5	48.8	50.5
256	78.7	69.4	51.5	52.0	51.8	53.2

选取4个模型进行对比实验。4个模型的基础结构参数均采用前述实验中确定的最优组合：输入序列长度为6步、LSTM隐层层数为2、隐含单元数为64、迭代次数为100，学习率为 1×10^{-3} ，优化器为Adam。具体比较模型如下：a. LSTM模型：不含注意力机制，用于对比基准。b. 单注意力机制模型：TA-LSTM模型的时间注意力维度设置为32；SA-LSTM模型的空间注意力维度设置为64。c. 双注意力机制模型：

SSTA-LSTM空间注意力维度设置为 64，时间注意力维度为 32；PSTA-LSTM空间注意力维度设置为 128，时间注意力维度为 64。

3.2.2 PSTA-LSTM 模型的分类型参数

在3.2.1节中确定的参数范围基础上，对不同水沙类型的样本分别开展独立的网格搜索与交叉验证。表4列出了不同水沙类型下的最优参数配置。从表4可见，不同水沙条件对模型在空间特征提取和时间序列依赖上的要求不同。

丰水丰沙阶段：上游来沙量大且峰值高，多支流贡献显著，河道输沙空间分布复杂。模型采用3层、128单元的LSTM网络，空间注意力维度128，时间注意力维度64，输入序列长度18。高维空间注意力可识别多个上游测站和支流的贡献差异；时间注意力结合18步历史观测信息，反应上游来沙对下游断面含沙量的影响，提高模型对高峰值和快速变化过程的预测能力。

丰水枯沙阶段：上游来沙量较低，河道含沙量短期变化幅度小，河道输沙空间分布相对简单。模型采用2层、96单元的LSTM网络，空间注意力维度96，时间注意力维度128，输入序列长度12。中等空间注意力可捕捉主要测站和支流贡献；时间注意力结合12步历史观测信息，使模型保持整体趋势，减少短期微小波动对预测曲线的影响。

中水中沙阶段：模型采用2层、64单元的LSTM网络，空间注意力维度96，时间注意力维度64，输入序列长度10。空间注意力整合主要测站贡献权重，时间注意力结合历史观测信息，可反映中水中沙条件下含沙量连续变化趋势，同时兼顾空间特征融合和时间序列依赖。

枯水丰沙阶段：上游来沙量高但河道流量较低，短期变化对下游影响明显，河道输沙空间分布复杂。模型采用3层、128单元的LSTM网络，空间注意力维度128，时间注意力维度64，输入序列长度6。高维空间注意力能够区分各测站和支流贡献，表达空间差异；时间注意力结合近期历史观测信息，使模型对短时含沙量快速变化敏感。

枯水枯沙阶段：上游来沙量低，河道流量有限，含沙量变化平稳且幅度小，空间分布简单。模型采用1层、32单元的LSTM网络，空间和时间注意力均为64，输入序列长度6。低维空间注意力可捕捉主要测站核心贡献；时间注意力结合历史观测信息，反映低含沙量整体趋势。

表4 最优超参数设定情况

Tab.4 Optimal hyperparameters setting

水沙类型	隐含层层数	隐含层单元数	输入序列长度	迭代次数	空间维度	时间维度	学习率
丰水丰沙类	3	128	18	200	128	64	1e-3
丰水枯沙类	2	96	12	150	96	128	1e-3
中水中沙类	2	64	10	120	96	64	1e-3
枯水丰沙类	3	128	6	150	128	64	1e-3
枯水枯沙类	1	32	6	100	64	64	1e-3

3.3 模型预测结果及对比

3.3.1 模型结构对比分析

在方式1中，将测试样本输入训练好的模型，得到预测结果并与实测值对比（图7）。总体来看，各模型均能再现潼关含沙量的主要变化趋势，但峰值阶段的幅度存在差异。基础LSTM模型能够捕捉总体波动，但在峰值阶段，预测的峰值时刻较实测峰值出现的时刻有所延迟，且峰值幅度偏低，曲线过于平滑。结果表明，缺乏注意力机制时，模型难以识别上游来沙的剧烈变化过程，峰值响应不足，典型表现为“峰值钝化”。引入空间注意力的SA-LSTM模型在含沙量高峰阶段的总体趋势拟合较好，表明空间特征提取增强了对上游各支流来沙对潼关含沙量影响差异的识别，但由于未引入时间注意力机制，模型无法对历史时间步的重要性进行加权，因而对峰值阶段的响应仍有限。

TA-LSTM模型引入时间注意力后，对峰值阶段的含沙量拟合更接近实测值，说明模型能够利用历史观测信息改善高峰预测。然而，对于两个峰值之间的含沙量波动变化，模型预测仍存在偏差，这表明仅依赖时间注意力难以充分捕捉不同上游测站对下游含沙量的空间贡献差异。SSTA-LSTM模型相比仅使用时间注意力或空间注意力的模型，在峰值拟合和连续波动段的表现均有所改善。但由于空间注意力仅对当前时间步的测站信息进行加权，而时间注意力在串行结构中依次处理这些加权后的历史状态，因此模型在含沙量快速变化时的响应仍存在一定滞后。PSTA-LSTM模型采用并行时空注意力结构，在峰值幅度和峰值时刻的预测上均优于其他模型。该结构能够同时利用不同上游测站的水沙信息进行加权，增强空间特征提取能力；同时通过时间注意力机制突出关键历史时间步的信息，使模型能够更准确地模拟潼关断面含沙量的峰值及变化趋势。

表5列出了五种模型的性能指标。随着结构从基础LSTM演化至PSTA-LSTM，RMSE呈下降趋势，NSE持续上升，决定系数 R^2 达到0.891，体现了并行时空注意力结构在信息融合和泥沙输移规律表达上的优势。

3.3.2 不同激活函数对比

以方式一的数据划分方式对比ReLU、SReLU和softplus激活函数在PSTA-LSTM模型中的表现。从图8和图9中可以看出，各激活函数的预测精度存在明显差异。

在含沙量变化相对平稳的时段，ReLU激活函数能够较好拟合整体变化趋势；但在含沙量快速上升的阶段，其预测误差明显增大，对峰值存在低估，预测曲线相较实测结果表现出一定的响应滞后，难以及时反映含沙量的快速变化过程。softplus激活函数由于输出连续平滑，在数值变化较为平缓的阶段具有较好的稳定性，但在含沙量剧烈上升或峰值阶段，其响应速度相对较慢，预测峰值偏低。相比之下，

SReLU 激活函数通过分段非线性调节,提高了模型对含沙量快速变化区间的表达能力,在峰值幅度和预测峰值时刻上,模型的预测结果与实测值更为接近。

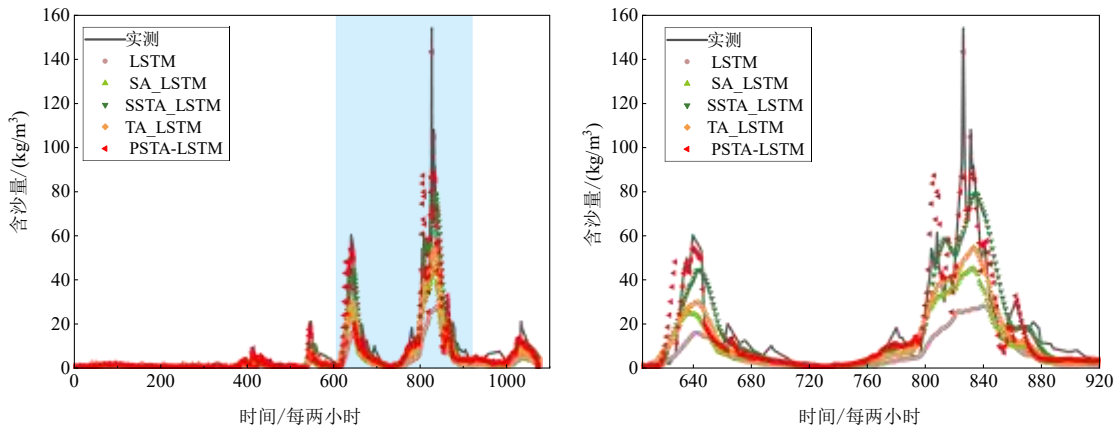


图 7 不同模型结构预测对比

Fig.7 Comparison of predictions among different model structures

表 5 不同模型结构评价指标

Tab.5 Evaluation metrics for different model architectures

模型	训练期					测试期				
	MAE	RMSE	NSE	PRE	R ²	MAE	RMSE	NSE	PRE	R ²
LSTM	13.556	19.894	0.725	0.376	0.794	15.583	24.642	0.689	0.403	0.771
SA-LSTM	12.051	19.045	0.750	0.344	0.823	14.035	23.046	0.738	0.367	0.796
TA-LSTM	11.652	18.799	0.794	0.312	0.855	13.838	22.761	0.773	0.329	0.838
SSTA-LSTM	11.349	18.424	0.839	0.261	0.896	12.803	20.153	0.811	0.283	0.867
PSTA-LSTM	9.173	13.216	0.892	0.228	0.914	10.385	14.975	0.865	0.247	0.891

从图9可见, SReLU在各项评价指标上均优于ReLU和softplus,表明SReLU在处理含沙量峰值及快速变化阶段时,具有更好的稳定性和适应性,提升了模型的预测能力。

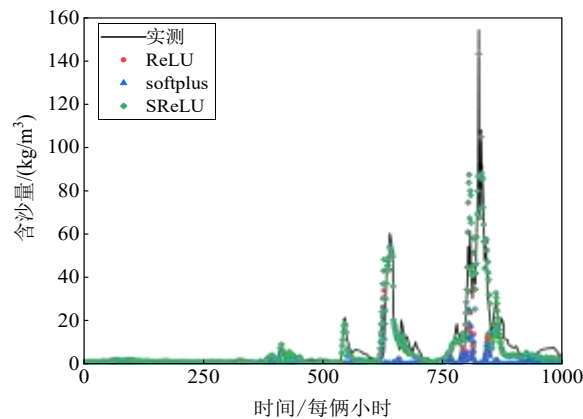


图 8 不同激活函数预测结果

Fig.8 Prediction results with different activation functions

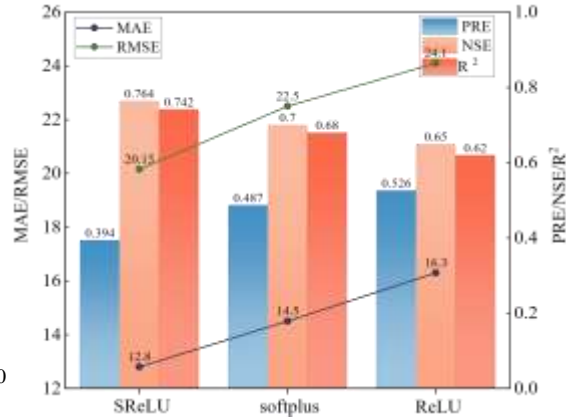


图 9 不同激活函数评价指标

Fig.9 Evaluation metrics for different activation functions

3.3.3 不同数据分类方式对比

通过两种数据集划分方式,比较了PSTA-LSTM模型在不同训练策略下的预测性能,二者均以2024年(丰水丰沙类)水沙过程作为测试集。方式1采用连续年份的训练集与验证集,方式2则根据水沙类型选取相应年份作为训练集与验证集。

图10展示了两种划分方式的预测结果对比。总体来看,模型在两种划分下较好地反应含沙量变化趋势及峰值时刻,但在峰值拟合精度方面存在明显差异。方式2的预测曲线峰值及峰现时刻上与实测结果更为一致,表明基于水沙类型的数据划分方式有助于提升模型的预测能力。

以丰水丰沙类型为例,潼关断面在汛期常出现含沙量短历时急剧上升的典型过程。表现为图10横轴为800-1000时刻之间,对应的是2024年8月11日洪峰,潼关站含沙量在00:00至08:00的8小时内由65.15 kg/m³迅速攀升至156 kg/m³,增幅达139%,呈现出显著的快速增沙特征。同期,上游龙门、华县及状头三站含沙量同步抬升,表明此次潼关断面的含沙量骤增主要源于上游高含沙水流的集中输移,而非本地河床冲刷所致。在此典型事件中,采用方式1对此次沙峰的预测相对误差为9.6%,最大预测峰值为141

kg/m³，低于实测峰值；而采用方式2的PSTA-LSTM模型，沙峰预测误差降低至6.4%，预测峰值提升至146 kg/m³，更接近实测值。

图10中左图汇总了两种数据集划分方式下模型的综合评价指标。蓝色为方式1的指标，红色为方式2的指标。由指标数值可知，方式2在所有主要指标上均优于方式1，峰值误差下降约15.7%，NSE提升约2.3%，R²提升约2.9%，这表明基于水沙类型划分的方式能有效提升模型对含沙量变化及峰值的预测能力。方式2的性能提升主要由于“分情景建模”的策略。不同水沙类型对应的含沙量变化特征不同，统一建模方式导致模型在不同水沙条件下的拟合需求可能相互干扰。而分类型训练能针对不同水沙条件分别学习相应的时空变化规律，从而提高预测精度。

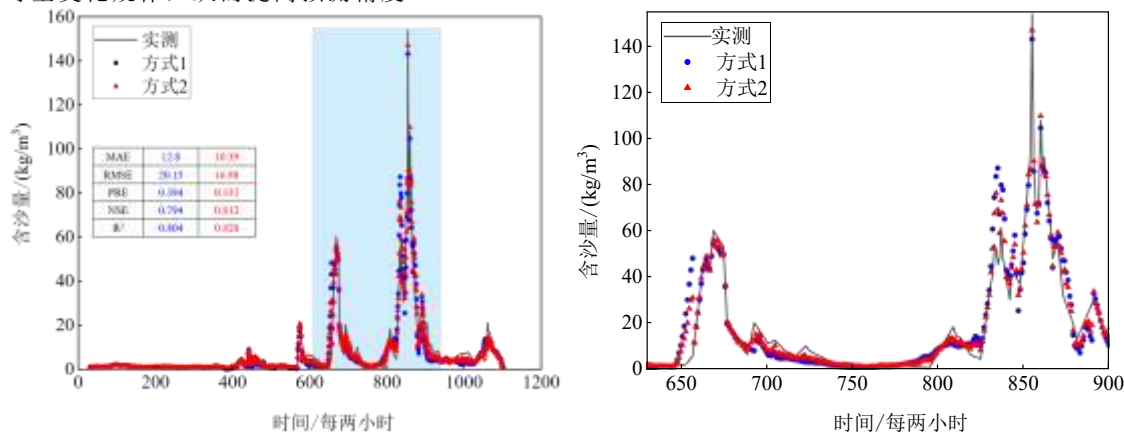


图 10 不同划分方式下的计算结果比较

Fig.10 Comparisons of calculation results of different classification methods

3.3.4 不同水沙类型预测对比分析

对PSTA-LSTM模型在丰水丰沙、丰水枯沙、枯水丰沙、中水中沙及枯水枯沙五类典型水沙情势下的预测结果进行了对比分析(图11和表6)，其中丰水丰沙结果见图10。总体来看，PSTA-LSTM模型在不同水沙类型条件内会存在预测值与实测值差异较大的情况，但预测值与实测值总体吻合，能够反映出含沙量随时间的变化过程，但不同类型之间的预测效果仍存在差别，表明模型对不同水沙条件下含沙量变化具有一定的区分能力。在枯水丰沙阶段(图11a)中，含沙量峰值明显且呈多级波动，模型能够较准确地再现峰值出现的时间和幅度，峰值时序与实测曲线基本一致。在枯水枯沙阶段(图11b)，含沙量总体偏低且变化幅度有限，模型能够再现主要峰值及其出现时段，但对低幅度波动的拟合略为平滑，预测曲线在低值段与实测数据基本吻合。在丰水枯沙阶段(图11c)中，含沙量主要峰值被模型有效捕捉，预测序列与实测曲线在峰值位置和上升趋势上保持一致，但在峰值段存在轻微低估。在中水中沙阶段(图11d)，模型能够再现峰值时刻及幅度变化，整体趋势与实测曲线高度一致。因此，PSTA-LSTM在丰水丰沙和枯水丰沙阶段，即含沙量峰值明显且短期变化剧烈的条件下，能够较为准确模拟峰值位置并反映含沙量的变化趋势；而在丰水枯沙和枯水枯沙阶段，含沙量较低且波动幅度有限时，模型的响应相对较弱。

表6 预测结果评价

Tab.6 Evaluation of prediction results

水沙类型	RMSE	MAE	PRE	NSE	R ²
丰水丰沙(2024年)	20.15	12.80	0.332	0.812	0.828
丰水枯沙(2021年)	18.56	11.94	0.314	0.836	0.836
中水中沙(2023年)	18.01	11.23	0.294	0.842	0.838
枯水丰沙(2005年)	16.42	10.75	0.301	0.859	0.854
枯水枯沙(2017年)	15.13	9.87	0.286	0.871	0.866

由表6可知，模型在不同水沙条件下的评价指标存在明显差异。在枯水枯沙阶段，模型的RMSE、MAE较低，NSE和R²超过0.85，这可能由于在枯水枯沙阶段，含沙量较低且变化不大，导致模型的NSE和R²偏大。正如图11所示，尽快整体预测趋势满足，但模型的含沙量峰值预测结果存在低估。随着水沙类型的变化，从丰水枯沙到中水中沙至丰水丰沙，RMSE、MAE逐渐增大，NSE和R²有所下降，模型的预测精度随水沙条件的复杂性的增加而下降，但丰水丰沙条件下的NSE和R²在80%以上，这表明，PSTA-LSTM对含沙量变化较大的水沙类型具有较高的敏感性，而在水沙波动较小的情况下，模型的预测误差相对增大。

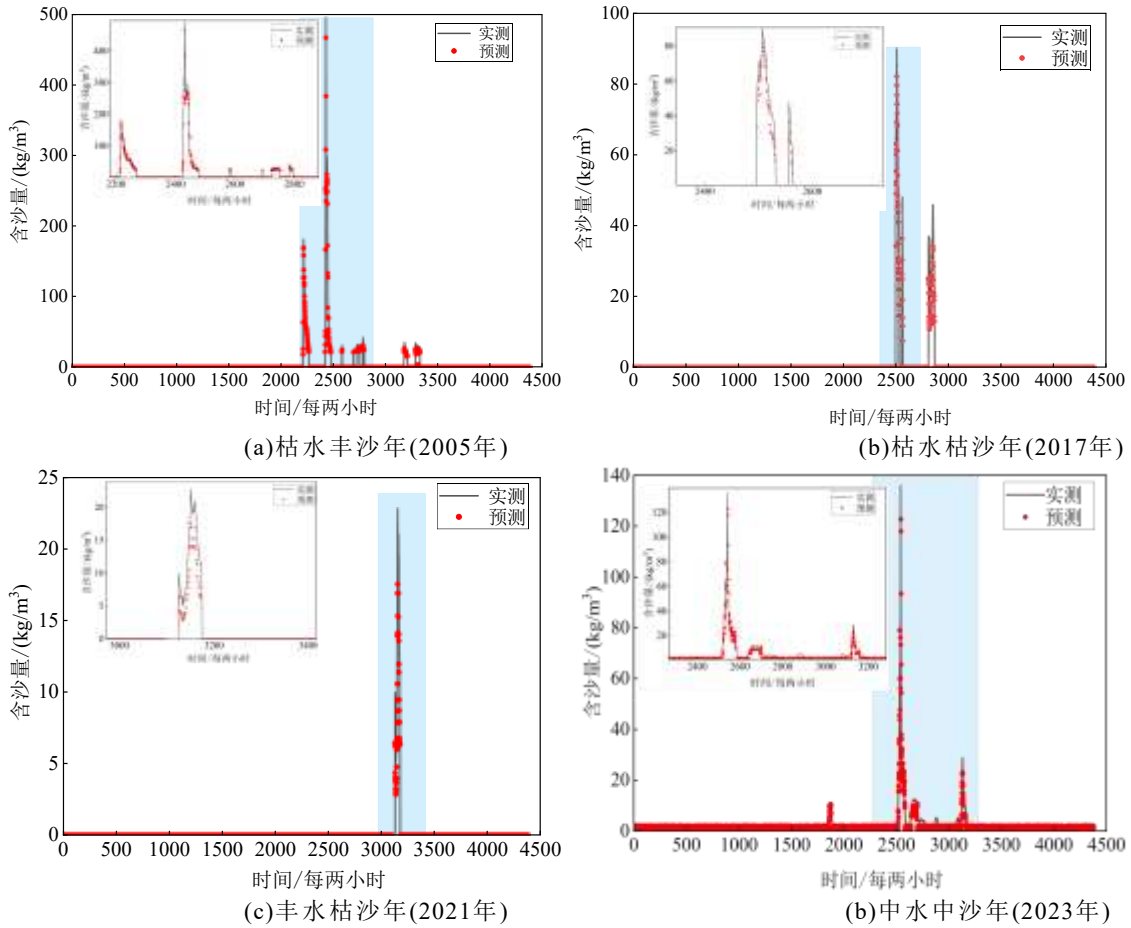


图 11 不同水沙类型年份的预测结果

Fig.11 Prediction results of different water and sediment types in different years

3.3.5 不同水沙类型汛期与非汛期预测性能对比

黄河流域的汛期一般为 7-10 月，分别计算预测结果在汛期与非汛期时段内的各类评价指标(图12)。PSTA-LSTM在汛期与非汛期的表现存在显著差异，非汛期指标明显优于汛期，这是因为汛期含沙量变化过程较为剧烈，模型预测值产生的误差也随之变大，而非汛期含沙量本身较小且变化不大，会导致指标表现较优。整体上，模型在汛期面临复杂水沙过程的条件下，仍保持良好的预测能力，NSE达80%以上，表明模型能有效捕捉含沙量过程的整体趋势变化，PRE约为30%，表明模型在预测含沙量峰值时存在一定的误差，但能够较为准确地反映峰值波动。

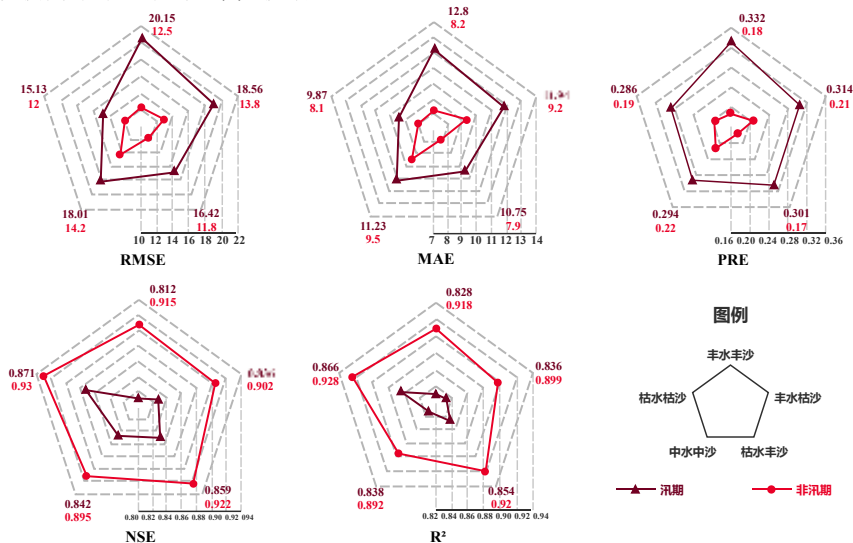


图 12 不同水沙类型汛期与非汛期预测性能对比

Fig.12 Comparison of flood season and non-flood season forecasts

不同水沙类型的指标对比显示,在丰水丰沙阶段,含沙量变化大,波动较多,模型的峰值预测能力最强,NSE达0.856, R^2 为0.872,体现出PSTA-LSTM模型对含沙量峰值的敏锐响应。

4 结论

基于黄河潼关站实测含沙量数据,构建了融合并行时空注意力机制的长短期记忆网络模型(PSTA-LSTM)。通过对该模型的训练和测试,得到以下主要结论:

(1)并行时空注意力机制使模型能够综合历史序列和上游来沙信息,使模型在含沙量快速变化及峰值阶段的预测精度明显提高。与传统串行结构相比,PSTA-LSTM的RMSE降低了25.6%,NSE提升了6.7%, R^2 提升至0.891,表明模型在含沙量变化剧烈的情况下拟合效果显著改善。

(2)基于水沙类型划分的数据集能够更精确地反映模型在不同水沙条件下的表现。特别在丰水丰沙阶段,模型对潼关站含沙量整体趋势的预测和峰值拟合表现最佳, RMSE为16.98, NSE为0.812, R^2 为0.828,表明模型对含沙量变化剧烈阶段较为敏锐。

(3)SReLU激活函数在各项评价指标上均优于ReLU和softplus,特别是在含沙量峰值。SReLU的RMSE降低了约10.4-16.3%,NSE提升至0.764, R^2 最大提升了约19.6%,显示出较好的稳定性和适应性。

5参考文献

- [1] 侯素珍,胡恬,王平,等. 近年来三门峡库区冲淤及潼关高程变化分析. 泥沙研究, 2024, 49 (06): 30-36. DOI:10.16239/j.cnki.0468-155x.2024.06.005.(HOU Suzhen, HU Tian, WANG Ping, et al. Analysis of erosion and sedimentation in the Sanmenxia Reservoir and Tongguan elevation change in recent years. Journal of Sediment Research, 2024, 49 (06): 30-36.)
- [2] 胡春宏,张晓明,于坤霞,等. 黄河流域水沙变化趋势多模型预测及其集合评估. 水利学报, 2023, 54 (07): 763-774. DOI:10.13243/j.cnki.slxb.20220717.(HU Chunhong, ZHANG Xiaoming, YU Kunxia, et al. Multi model prediction and its ensemble evaluation of runoff and sediment change trend in the Yellow River Basin. Journal of Hydraulic Engineering, 2023, 54 (07): 763-774.)
- [3] 胡春宏,张晓明,赵阳. 黄河泥沙百年演变特征与近期波动变化成因解析. 水科学进展, 2020, 31 (05): 725-733. DOI:10.14042/j.cnki.32.1309.2020.05.009.(HU Chunhong, ZHANG Xiaoming, ZHAO Yang. Cause analysis of the centennial trend and recent fluctuation of the Yellow River sediment load. Advances in Water Science, 2020, 31 (05): 725-733.)
- [4] 何娟,郑珊,吴保生. 三门峡水库冲淤重心变化及其对潼关高程的影响. 湖泊科学, 2023, 35 (01): 338-348. DOI: 10.18307/2023.0126 (HE Juan, ZHENG Shan, WU Baosheng. Changes of erosion/deposition centers in the Sanmenxia Reservoir and its influence on the riverbed elevation at Tongguan. Journal of Lake Sciences, 2023, 35 (01): 338-348.)
- [5] 夏军强,张贤梓依,王增辉,等. 黄河三门峡水库水沙电耦合模型研究及应用. 水科学进展, 2023, 34 (02): 238-249. DOI:10.14042/j.cnki.32.1309.2023.02.008.(XIA Junqiang, ZHANG Xianziyi, WANG Zenghui, et al. Study and application of a coupled modelling of flow-sediment transport and hydropower generation in the Sanmenxia Reservoir. Advances in Water Science, 2023, 34 (02): 238-249.)
- [6] 周美蓉,夏军强,邓珊珊,等. 考虑河道整治工程影响的一维水沙数学模型及其应用. 湖泊科学, 2021, 33 (02): 571-580. DOI: 10.18307/2021.0222(ZHOU Meirong, XIA Junqiang, DENG Shanshan, et al. One-dimensional morphodynamic model of channel deformation with the consideration of river regulation works and its application. Journal of Lake Sciences, 2021, 33 (02): 571-580.)
- [7] 余明辉,祝求,杨国录,等. 考虑溯源冲刷的一维水沙数学模型拓展及应用. 泥沙研究, 2021, 46 (02): 1-6. DOI:10.16239/j.cnki.0468-155x.2021.02.001.(YU Minghui, ZHU Qiu, YANG Guolu, et al. Improvement and application of onedimensional numerical model of flow and sediment with retrogressive erosion. Journal of Sediment Research, 2021, 46 (02): 1-6.)
- [8] AFAN H A, EL-SHAFIE A, MOHTAR W H M W, et al. Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. Journal of Hydrology, 2016, 541: 902-913. DOI: 10.1016/j.jhydrol.2016.07.048
- [9] KHOSRAVI K, GOLKARIAN A, MELESSE A M, et al. Suspended sediment load modeling using advanced hybrid rotation forest based elastic network approach. Journal of Hydrology, 2022, 610: 127963. DOI: 10.1016/j.jhydrol.2022.127963
- [10] ASSEM, H., GHARIBA, S., MAKRAI, G., et al. Urban water flow and water level prediction based on deep learning[C]// In Proceedings of the Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, 18-22 September 2017; Proceedings, Part III 10. pp. 317-329. DOI: 10.1007/978-3-319-71273-4_26

- [11] CISTY, M., SOLDANOVA V. Flow prediction versus flow simulation using machine learning algorithms. In Machine Learning and Data Mining in Pattern Recognition[C]//In Proceedings of the 14th International Conference, MLDM 2018, New York, NY, USA, 15-19 July 2018; Proceedings, Part II 14; pp. 369-382. DOI: 10.1007/978-3-319-96133-0_28
- [12] YUAN Xiaohui; CHEN Chen, LEI Xiaohui, et al. Monthly runoff forecasting based on LSTM-ALO model. *Stochastic Environmental Research and Risk Assessment*. 2018, 32, 2199-2212. DOI: 10.1007/s00477-018-1560-y
- [13] AHMED A N, OTHMAN F B, AFAN H A, et al. Machine learning methods for better water quality prediction. *Journal of Hydrology*, 2019, 578: 124084. DOI: 10.1016/j.jhydrol.2019.124084
- [14] LI Xia, XU Wei, REN Minglei, et al. Hybrid CNN-LSTM models for river flow prediction. *Water Supply*, 2022, 22(5), 4902-4919. DOI: 10.2166/ws.2022.170
- [15] 郭秀吉,孙龙飞,王平等. 基于机器学习的三门峡水库多因子排沙预测模型比较研究. *人民黄河*, 2024, 46 (12): 37-43. DOI: 10.3969/j.issn.1000-1379.2024.12.006(GUO Xiuji, SUN Longfei, WANG Ping, et al. Comparative research on multifactor sediment discharge prediction models of sanmenxia reservoir based on machine learning. *Yellow River*, 2024, 46 (12): 37-43.)
- [16] 陈雪,胡宏昌,刘慧等. 基于机器学习算法的清涧河流域径流及含沙量预测研究. *中国水利水电科学研究院学报(中英文)*, 2024, 22 (05): 520-529+538. DOI:10.13244/j.cnki.jiwhr.20240056.(CHEN Xue, HU Hongchang, LIU Hui, et al. Prediction of runoff and sediment concentration in Qingjian River basin based on machine learning algorithm. *Journal of China Institute of Water Resources and Hydropower Research*, 2024, 22 (05): 520-529+538.)
- [17] NEJC B, KLAUDIJA L, YUN B, et al. Using machine learning to predict suspended sediment transport under climate change. *Water Resources Management*, 2025, 39: 3311-3326. DOI: 10.1007/s11269-025-04108-7
- [18] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization. *Neural and Evolutionary Computing*, 2014, arXiv:1409.2329. DOI: 10.48550/arXiv.1409.2329
- [19] HOCHREITER S, SCHMIDHUBER J. Long short-term memory. *Neural Computation*, 1997, 9(8), 1735-1780. DOI: 10.1162/neco.1997.9.8.1735
- [20] 林天宙, 彭杨, 罗诗琦, 等. 基于 RF-LSTM 模型的长江上游汛期日含沙量预测. *中国农村水利水电*, 2024, (10): 32-39. DOI: 10.12396/znsd.240312 (LIN Tianzhou, PENG Yang, LUO Shiqi, et al. Prediction of daily sediment concentrations in the upper reaches of yangtze river during flood season based on rf-lstm model. *China Rural Water and Hydropower*, 2024, (10): 32-39.)
- [21] 魏苗,胡新源,周聂,等. 基于机器学习算法的含沙量短临预报模型研究. *中国农村水利水电*, 2024, (09): 60-67. DOI: 10.12396/znsd.240082 (WEI Miao, HU Xinyuan, ZHOU Nie, et al. Short-Term forecasting of suspended sediment concentration based on machine learning. *China Rural Water and Hydropower*, 2024, (09): 60-67)
- [22] 鲍振鑫,张建云,王国庆,等. 基于水文模型与机器学习集合模拟的水沙变异归因定量识别——以黄河中游窟野河流域为例. *水科学进展*, 2021, 32(4): 485-496. DOI: 10.14042/j.cnki.32.1309.2021.04.001 (BAO Zhenxin, ZHANG Jianyun, WANG Guoqing, et al. Quantitative assessment of the attribution of runoff and sediment changes based on hydrologic model and machine learning: a case study of the Kuye River in the Middle Yellow River Basin. *Advances in Water Science*, 2021, 32(4): 485-496.)
- [23] CHOUBIN B, DARABI H, RAHMATI O, et al. River suspended sediment modelling using the CART model: a comparative study of machine learning techniques. *Science of the Total Environment*, 2018, 615: 272-281. DOI: 10.1016/j.scitotenv.2017.09.293
- [24] KAVEH K, KAVEH H, BUI M D, et al. Long shortterm memory for predicting daily suspended sediment concentration. *Engineering with Computers*, 2021, 37(3): 2013-2027. DOI: 10.1007/s00366-019-00921-y
- [25] 全胜蓝,江衍铭. 基于神经网络模型的输沙过程推估. *中国农村水利水电*, 2020(9):112-116. DOI: 10.3969/j.issn.1007-2284.2020.09.022 (QUAN Shenglan, JIANG Yanming. Estimation of suspended sediment loads by artificial neural network methods. *China Rural Water and Hydropower*, 2020 (9): 112-116.)
- [26] DAI Z, ZHANG M, NEDJAH N, et al. A Hydrological Data Prediction Model Based on LSTM with Attention Mechanism. *Water*, 2023,15: 670. DOI: 10.3390/w15040670

- [27] ZHAO Xinfeng, DONG Shengwen, RAO Hui, et al. Water flow forecasting model based on bidirectional long- and short-term memory and attention mechanism. *Water*, 2025, 17: 2118. DOI: 10.3390/w17142118
- [28] HU Wenwen, YU Yongchuan, YAN Jianzhao, et al. Water Flow Prediction Based on Improved Spatiotemporal Attention Mechanism of Long Short-Term Memory Network. *Water*, 2024, 16:1600. DOI: 10.3390/w16111600
- [29] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]// ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning, 2010: 807-814.
- [30] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]//14th International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, USA, 2010:315-323. DOI: <https://doi.org/10.1.1.208.6449>
- [31] 陈立,明宗富.河流动力学.武汉:武汉大学出版社,2001