



环境工程

Environmental Engineering

ISSN 1000-8942, CN 11-2097/X

《环境工程》网络首发论文

题目：基于注意力机制的时间卷积网络河流总磷预测
作者：黎园园，刘海隆
收稿日期：2022-07-26
网络首发日期：2023-01-18
引用格式：黎园园，刘海隆. 基于注意力机制的时间卷积网络河流总磷预测[J/OL]. 环境工程. <https://kns.cnki.net/kcms/detail//11.2097.X.20230117.1550.005.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于注意力机制的时间卷积网络河流总磷预测

黎园园 刘海隆*

(电子科技大学资源与环境学院, 成都 611731)

摘要：深度学习在水质污染监测中的应用研究成为当前热点。为了解决水体污染物浓度在上下游之间存在不同时间步关联的问题,以四川彭山区岷江水系的总磷预测为研究目标,提出了基于时间模式注意力(TPA)机制的时间卷积网络(TCN)预测模型。首先,利用 Pearson 相关系数对上下游站点处总磷浓度进行时空关系分析;然后,利用 TCN 网络的扩张因果卷积来提取时序数据之间的依赖关系,利用 TPA 机制从 TCN 网络中学习空间不同站点总磷浓度时序之间的复杂关系,从而获取时序数据中不同时间步的站点权重;最后,将该模型应用于河流总磷浓度预测。研究表明:多站点输入的 TPA-TCN 模型的评价指标 RMSE、MAE 和 MAPE 比单站点输入的 TPA-TCN 模型分别降低了 36.29%、28.18%和 25.26%,相比于融入传统注意力机制的 TCN 模型,融入 TPA 机制的 TCN 模型这三种评价指标分别降低了 10.24%、10.78%和 9.94%,说明多站点输入的 TPA-TCN 模型对河流总磷浓度预测具有一定的优势。TPA-TCN 模型可以有效应用于水质监测之中,该研究对水质污染物预测具有重要参考意义。

关键词：水质预测; 时间卷积网络; 时间模式注意力机制; Pearson 相关系数

Prediction of Total Phosphorus in Rivers Based on Attention Mechanism of Temporal Convolutional Networks

LI Yuan-yuan, LIU Hai-long*

(School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: The application of deep learning in water pollution monitoring has become a current hottopic. Taking the total phosphorus prediction of Min River water system in Pengshan District, Sichuan Province as the research goal, a temporal convolutional network (TCN) prediction model based on temporal pattern attention (TPA) mechanism was proposed to solve the problem that the concentration of water pollutants was related to different time steps between upstream and downstream. Firstly, using Pearson correlation coefficient to analyze the spatiotemporal relationship of total phosphorus concentration at upstream and downstream sites. Then, the dilated convolution and causal convolution of TCN were used to extract the dependencies of the time series data. And the TPA mechanism was used to learn the complex spatial relationship between the time series of total phosphorus concentration at different stations from the TCN, which can obtain the station weights at different time steps in the time series data. Finally, this model was applied to predict the total phosphorus concentration of rivers. The research results show that the multi-site input TPA-TCN model reduced RMSE by 36.29%, MAE by 28.18% and MAPE by 25.26% than that of the single-site input TPA-TCN model. Compared with the TCN model integrated with the traditional

收稿日期: 2022-07-26

基金项目: 第三次新疆综合科学考察项目(2021xjkk1404); 四川省国际合作项目(2020YFH0067)

第一作者: 黎园园(1996-), 女, 硕士研究生, 主要研究方向为生态环境智能监测与信息处理。yuan_yuan_li20@163.com

* 通信作者: 刘海隆(1974-), 男, 教授, 主要研究方向为生态环境监测。liuhl@uestc.edu.cn

attention mechanism, the three evaluation metrics of the TCN model integrated with the TPA mechanism reduced by 10.24%, 10.78% and 9.94%, indicating that the multi-site input TPA-TCN model has certain advantages in predicting the total phosphorus concentration in rivers. The TPA-TCN model can be effectively applied to water quality monitoring, which has important reference significance for the prediction of water pollutants.

Key words: water quality prediction; temporal convolutional network; temporal pattern attention mechanism; Pearson correlation coefficient

0 引言

水质预测评估是环境监管和治理的重要内容。基于数学物理模型的水质预测模型,需要利用大量的基础数据资料来建立,许多参数无法精准测量和率定,导致模型的应用存在一定的局限性^[1]。近年来,非机理性方法在水质预测中取得了良好的效果,如灰色系统模型^[2]、自回归综合移动平均模型 (ARIMA)^[3, 4]、支持向量机 (SVM)^[5]、人工神经网络 (ANN)^[6]和深度学习^[7]等,其中深度学习发展极为迅速。

水质污染具有显著的时空间关联特征。由于传统神经网络很难捕捉到长跨度的关键信息,现有水质的预测主要是基于循环神经网络 (Recurrent Neural Network, RNN) 和卷积神经网络 (Convolutional Neural Network, CNN) 的方法来进行^[8, 9]。张贻婷从单个站点水质数据之间的相关性和上游站点的影响两方面,分别建立 LSTM 河流水质预测模型,结果表明,对输入变量进行特征选择及加入上游监测站的双测站的 LSTM 模型预测效果更好^[10]。Barzegar 等人提出了一种结合 CNN 和 LSTM 的混合模型,该模型将水质数据序列作为卷积层结构的输入,进行特征提取,然后通过扁平化层作为 LSTM 的输入,获得目标的预测值,通过实验对比,这个混合模型集成了 CNN 和 RNN 两种经典模型的优点^[11]。Yang 等人提出了一种基于注意力机制的 CNN-LSTM 混合模型来进行水质预测,混合的 CNN-LSTM 模型具有很好解决非线性时间序列预测问题的能力,注意力机制可以捕获较长的时间依赖性,实验结果表明,该模型优于其他模型,并且可以在不同的时间滞后条件下进行稳定预测^[12]。

时间卷积网络 (Temporal Convolutional Network, TCN) 是近年来提出的具有处理时间序列能力的网络,结合了 CNN 和 RNN 的优点,即 CNN 的并行处理和 RNN 的时域建模能力^[13]。TCN 的结构对历

史信息的感受域更加灵活,可以获取更长时间序列的信息^[14]。Zhang 等人利用 TCN 网络的并行特点,提出了一个多任务时间卷积网络来预测多个水质变量^[15]。Fu 等人利用基于 TCN 的水质长期预测模型来预测溶解氧、水温和 pH,实验结果表明相比基于循环神经网络、双向简单循环单元、门控循环单元以及 LSTM 网络模型,基于 TCN 的长期预测方法具有更高的精度和较低的时间复杂度^[16]。Li 等人结合 LSTM 和 TCN 网络的特点,提出一种基于 LSTM-TCN 的组合预测模型,对溶解氧进行预测,结果表明随着溶解氧序列时间窗的增加,该组合模型可以更好地捕捉到历史数据的特征,具有较高的预测性能^[17]。这些预测方法都没有考虑上下游水质之间的相关性,没有充分利用上游的水质数据。

由于河流上游不同水质监测站点在不同的时间步上对下游预测目标站点的水质影响不同,传统的注意力机制不能提取不同站点之间在不同时间上的非线性的复杂关系。Shih 等人提出了时间模式注意力 (temporal pattern attention, TPA) 机制,TPA 能够为不同时间序列选择不同时间步,使得模型为不同变量在不同时间步上进行差异化权重分配,实验结果表明基于 TPA 的模型取得了更好的预测效果^[18]。

因此,本文在探究河流上下游不同监测站点处的河流总磷浓度相关性的基础上,构建基于时间卷积网络的河流总磷浓度预测模型,同时引入时间模式注意力机制,捕捉不同站点总磷浓度在不同时间步的相关性,从而高效地学习上游其他站点和目标站点的历史总磷浓度数据,以提高河流总磷浓度的预测精度和效率。

1 数据与方法

1.1 数据来源和预处理

本文以四川省眉山市彭山区为研究区域,该区域的河流属于长江上游的重要支流——岷江的支流,收集 15 个水质监测站点的总磷浓度历史数据,这

15 个监测站点分布于彭山区的涪洲河、毛河及西干渠上,采集时间均为 2021 年 7 月至 2022 年 1 月,每隔 4h 采集一次。水质监测站点在彭山区的分布如图 1 所示,圆点表示监测站点。该区域的河流流向从北往南和从西往东。在实际的数据采集中,常常会出现监测点失效导致数据缺失的问题,对于数据的缺失值,采用邻近均值法实现缺失值填充。

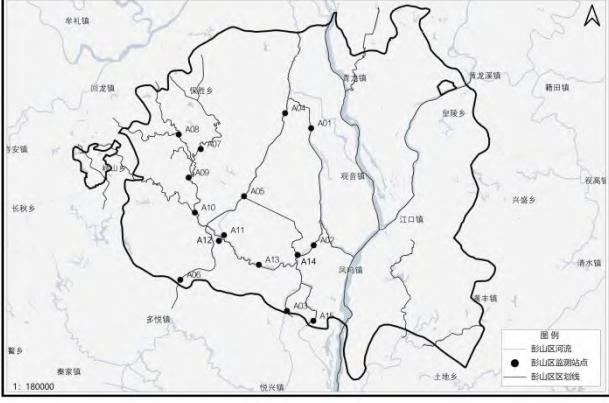


图 1 彭山区水质监测站点分布图

Figure 1 Distribution diagram of water quality monitoring stations in Pengshan District

1.2 Pearson 相关系数

Pearson 相关系数法常用于表明变量之间的线性相关强弱^[16]。利用 Pearson 相关系数法可以分析不同站点之间总磷浓度的相关性,及趋势变化的关系。本文利用 Pearson 相关系数法作为一个筛选上游站点的方法,可以减少模型的输入特征个数,提高模型精确度,提升模型效率,也有助于增加模型的可解释性^[10, 19, 20]。

Pearson 相关系数是一种用于衡量两个变量之间相关程度的指标,该值处于-1 到 1 之间^[21]。两个变量之间的 Pearson 相关系数 r 是由两个变量 X 、 Y 之间的协方差和标准差的商确定。

如果 $r > 0$, 则变量 X 、 Y 正相关,即 Y 会随 X 的增加而增加,反之, $r < 0$, 则 X 、 Y 负相关。若 $|r| \geq 0.8$, 则 X 、 Y 极强相关;若 $0.6 \leq |r| < 0.8$, 则 X 、 Y 强相关;若 $0.4 \leq |r| < 0.6$, 则 X 、 Y 中度相关;若 $0.2 \leq |r| < 0.4$, 则 X 、 Y 相关程度弱;若 $0 \leq |r| < 0.2$, 则 X 、 Y 基本不相关^[20]。

1.3 时间卷积网络结构

时间卷积网络(TCN)是将卷积神经网络(CNN)思想引入时序数据处理的一种方法,主要由因果卷积、扩张卷积和残差块三个核心模块构成^[22]。

假设输入时间序列为 $X^{T+1} = (x_0, x_1, \dots, x_T)$, 神

经元的输出为 $Y^{T+1} = (y_0, y_1, \dots, y_T)$, 卷积核公式为 $F = (f_1, f_2, \dots, f_K)$, 其中 K 为卷积核的大小,则 x_T 处的因果卷积表示为式 (1):

$$F(x_T) = \sum_{k=1}^K f_k \cdot x_{T-K+k} \quad (1)$$

由于使用因果卷积,对于任意时刻输出的 y_t 只能与 x_0, x_1, \dots, x_t 有关,与 $x_{t+1}, x_{t+2}, \dots, x_T$ 无关,即确保该时刻之后的信息不会用来预测,同时也保证了之前的信息不会被遗忘^[23]。当追溯的历史时间越久远,卷积层层数就越多,以致网络架构就越深。因此,由于网络层数的限制,序列往往不能过长,时间卷积网络就引入了扩张卷积来加以改善,使网络可以获得更长的时间序列依赖关系^[24]。

扩张卷积主要是对卷积的输入进行间隔采样,等同于跳过部分输入来改变卷积核的大小,以扩展 TCN 的感受野,获得更长的序列依赖关系,其中采样率由扩张因子 d 决定。扩张因果卷积表示为式 (2):

$$F_d(x_T) = \sum_{k=1}^K f_k \cdot x_{T-(K-k)d} \quad (2)$$

TCN 使用残差模块代替了卷积层。设定 X 为残差模块的输入值,跨层时潜在的恒等映射函数为 F ,其结果会被添加至输入值 X ,则残差模块的输出值 o 可表示为式 (3):

$$o = \text{Activation}(X + F(X)) \quad (3)$$

一般来说,一个残差模块由 2 层扩张因果卷积、WeightNorm 单元、ReLU 单元以及 Dropout 单元组成。TCN 的残差连接网络由多个这样的残差块堆叠而成。TCN 模型感受野的增加依赖于网络深度,而残差结构可以很好地抑制网络层增加导致的梯度消失或者爆炸的后果,且同时扩张因果卷积的引入使得 TCN 在时序建模任务具有优势^[25]。

1.4 时间模式注意力机制 (TPA)

TPA 可以采用一维卷积网络从 TCN 隐藏状态行向量获取时间序列数据包含的时间模式信息,使模型能够在不同的时间步上学习多个水质数据之间的相互关系。图 2 展示了 TPA 机制的结构^[18]。

原始水质时间序列数据经过 TCN 网络后,得到隐藏状态向量 $h_{t-w}, h_{t-w+1}, \dots, h_t$, 组成隐藏状态矩阵,其中 h_t 表示时间步 t 时的状态, w 是注意力所关注的时间序列长度。时间步 t 之前的隐藏状态向量组成隐藏状态矩阵 $H = \{h_{t-w}, h_{t-w+1}, \dots, h_t\}$, 隐藏状态矩阵的行向量表示单个变量在时间步上的状态,列向量是同一个时间步下 TCN 内部神经元参数构成的。

根据图 2, TPA 机制作用原理的描述如下。

首先,使用一维 CNN 对隐藏状态矩阵进行卷积

计算。将 k 个一维卷积核分别沿着隐藏状态矩阵 H 的行向量进行卷积，得到时间模式矩阵 H^C ，表示如式 (4)。

$$H_{i,j}^C = \sum_{l=1}^w H_{i,(t-w-1+l)} \times C_{j,(T-w+1+l)} \quad (4)$$

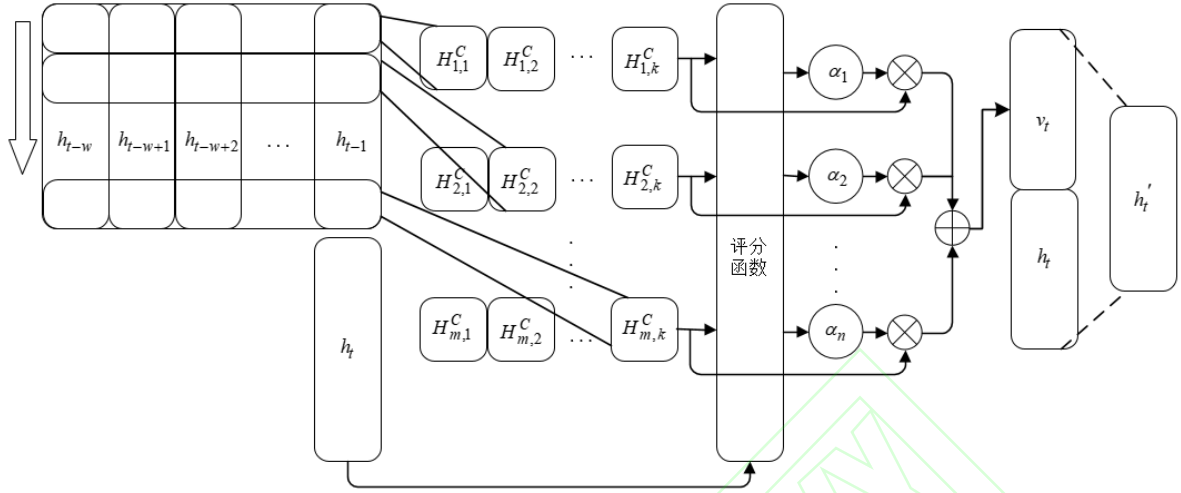


图 2 TPA 机制结构图

Figure 2 Structure of TPA

其中, C_j 表示第 j 个长度为 T 的滤波器, T 为注意力需要关注的最大长度, 通常 T 取 w ; $H_{i,j}^C$ 表示第 j 个滤波器对隐藏状态矩阵的第 i 行进行卷积计算得到的时间模式矩阵。这个时间模式矩阵体现了不同变量的时间序列之间的复杂关系。

其次, 利用评分函数对时间模式进行评分, 计算注意力权重。评分函数 f 为:

$$f(H_i^C, h_t) = (H_i^C)^T W_\alpha h_t \quad (5)$$

其中, H_i^C 表示时间模式矩阵 H^C 的第 i 行行向量, W_α 表示权重矩阵, h_t 表示时间步 t 时的隐藏状态。时间模式矩阵第 i 行的注意力权重 α_i 的计算公式如式 (6), 使用的激活函数为 sigmoid 函数。

$$\alpha_i = \text{sigmoid}(f(H_i^C, h_t)) \quad (6)$$

然后, 计算上下文向量。利用注意力权重 α_i 对 H^C 的每一行进行加权求和, 得到上下文向量 v_t , 如式 (7) 所示。

$$v_t = \sum_{i=1}^n \alpha_i H_i^C \quad (7)$$

最后, 将上下文变量 v_t 与时间步 t 时的隐藏状态 h_t 线性映射后相加得到最终的预测值, 如式 (8) 所示。

$$y_{t-1+\tau} = W_{h'}(W_h h_t + W_v v_t) \quad (8)$$

其中, $y_{t-1+\tau}$ 表示预测值, τ 表示预测的时间尺度, $W_{h'}$, W_h 以及 W_v 表示对应变量的权重矩阵。

在根据不同站点来进行河流总磷浓度预测的时候, TPA 机制利用一维 CNN 对 TCN 网络的隐藏状态的行向量进行提取, 就能够同时提取不同站点总

磷浓度与时间序列之间的复杂关系, 选择对目标站点总磷浓度更相关的站点。上下文向量是包含时间信息的时间模式矩阵行向量的加权和, 因此模型能够选择不同时间步的相关信息^[26]。

1.5 基于 TPA 的 TCN 河流总磷预测模型的构建

基于时间模式注意力机制的 TCN 河流总磷预测算法流程如图 3 所示, 其算法步骤如下:

- 1) 从彭山区水质监测站点获取每个站点的总磷浓度时间序列数据。
- 2) 对各个站点的数据做预处理, 包括缺失数据填充和数据标准化。对于缺失的数据采用邻近均值法实现缺失值填充。为了降低不同监测站点总磷数值波动范围的影响, 同时减少计算的复杂度, 加快神经网络的收敛速度, 本文采用极差标准化方法^[12, 27, 28], 计算公式如式 (9), 其中 x'_i 表示第 i 个水质站点总磷浓度数据标准化后的值, x_i 为第 i 个水质站点总磷浓度数据的实际值, x_{imax} 为第 i 个水质站点总磷浓度数据的最大值, x_{imin} 为 i 个水质站点总磷浓度数据的最小值。

$$x'_i = \frac{x_i - x_{imin}}{x_{imax} - x_{imin}} \quad (9)$$

- 3) 利用 Pearson 相关系数计算河流各个水质站点总磷浓度之间的相关性, 选择上游相关系数大于 0.7 的站点作为目标站点预测的辅助站点。

- 4) 常用的数据集划分比例有 7: 3 和 8: 2, 本文的数据集样本较小, 采用 8: 2 的比例划分数据集,

可以使用更多的数据来训练模型，得到模型的最优参数^[12, 29]。故本文将经过上述处理后的数据按照 8:2 的比例划分为训练集和测试集。

5) 构建 TPA-TCN 的河流总磷浓度预测模型，该模型包括 TCN 层、TPA 层以及 Dense 输出层。将训练集数据送入 TCN 网络中，得到隐藏状态矩阵 H 和时间步 t 时的隐藏状态输入到 TPA 层，通过卷积计算得到时间模式矩阵，然后通过评分函数得到注意力权重，利用时间模式矩阵和注意力权重加权求和得到上下文向量，将上下文向量和隐藏状态行向量分别进行线性映射相加后的结果送入 Dense 输出层，得到最终的预测值。

6) 用训练集训练 TPA-TCN 的河流总磷浓度预测模型和结构参数，得到训练好的预测模型。

7) 在训练好的河流总磷浓度预测模型中输入测试集，得到预测结果。

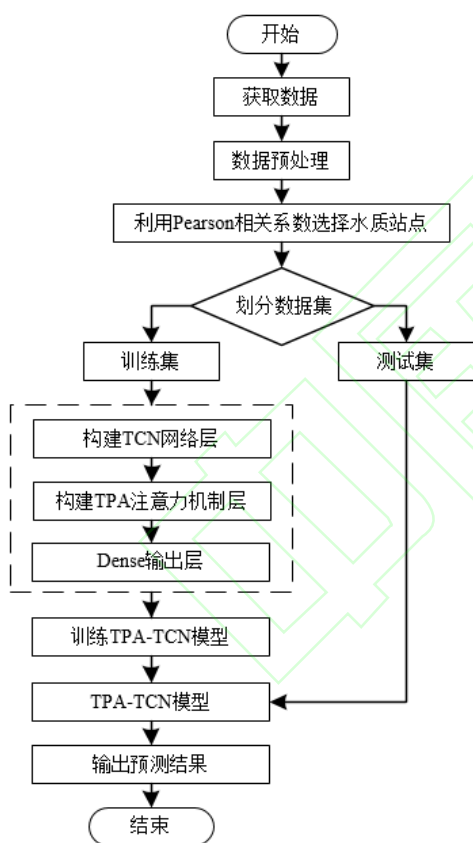


图3 基于 TPA 的 TCN 河流总磷预测算法流程
Figure 3 Algorithm flow of river total phosphorus prediction based on TPA of TCN

1.6 模型评价指标

本文采用均方根误差 (RMSE, root mean squared error)、平均绝对误差 (MAE, mean absolute error)

和平均绝对百分比误差 (MAPE, mean absolute percentage error) 作为河流总磷浓度预测结果准确度的评价指标。RMSE、MAE 和 MAPE 的计算公式如式 (10)、式 (11) 和式 (12)。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (10)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (11)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (12)$$

其中, n 表示样本数量, y_i 表示实测值, \hat{y}_i 表示预测值。

2 结果与分析

2.1 彭山区河流水质特征

本研究区域彭山区的这 15 个水质监测站点中, 站点 1、2、3 和 14 位于湄洲河, 站点 4、5 和 6 位于西干渠, 站点 7、8、9、10、11、12、13 和 15 位于毛河。这些站点所处的三条河流都有交汇的地方, 水的流动性以及河流周围环境的相似性, 也会导致不同空间位置的污染情况具有一定的相似性。

2021 年 7 月至 2022 年 1 月彭山区河流 15 个监测站点总磷监测数据的描述性统计结果如表 1。

表 1 总磷数据的描述性统计情况

Table 1 Descriptive statistics of total phosphorus

监测 点位	总磷/(mg/L)			
	均值	标准差	超过Ⅲ类水 比例 (%)	超过Ⅴ类水 比例 (%)
A01	0.191	0.0341	33.8	0
A02	0.188	0.0433	29.07	0
A03	0.210	0.0472	51.78	0
A04	0.196	0.0354	38.84	0
A05	0.199	0.0399	43.49	0
A06	0.192	0.0424	36.20	0
A07	0.341	0.2557	65.89	26.82
A08	1.154	0.4556	100	97.67
A09	0.435	0.1633	89.69	55.89
A10	0.184	0.1248	35.19	9.61
A11	0.213	0.0468	51.55	0
A12	0.166	0.0535	20.70	0
A13	0.205	0.0526	44.57	1.01
A14	0.238	0.0726	65.43	4.34
A15	0.208	0.0563	47.83	0

从表 1 可以看出, 湄洲河的总磷浓度平均值范围在 0.191~0.238 mg/L, 在《我国地表水环境质量标准》(GB3838-2002)III~IV类水标准。以地表III类水质为标准, 站点 1 和 2 超标比例较小, 而站点 3 和 14 超标比例较大。以地表V类水质为标准, 只有站点 14 有很少量的时间未达标, 湄洲河其余站点均达标。西干渠的总磷浓度平均值范围在 0.192~0.196 mg/L, 在《我国地表水环境质量标准》(GB3838-2002)III类水标准。以地表III类水质为标准, 西干渠的总磷超标比例均不超过 50%。以地表V类水质为标准, 西干渠的总磷均达标, 该河流的水质状况较好。毛河的总磷浓度平均值范围在 0.166~1.154 mg/L, 在《我国地表水环境质量标准》(GB3838-2002)III~劣V类水标准。以地表III类水质为标准, 毛河上游的站点 7, 8, 9 的超标比例较严重, 站点 8 甚至全部超标, 而毛河中下游的超标比例较少。以地表V类水质为标准, 上游的站点 8 超标比例仍然严重。从标准差来看, 站点 7, 8, 9, 10 的总磷标准差较大, 数据波动较大。

2.2 河流水质监测站点总磷数据的时空关系分析

为了探究彭山区河流上下游监测站点处总磷浓度的关系, 本文对不同空间位置的监测站点总磷浓度数据进行 Pearson 相关系数检验, 各个监测站点之间总磷浓度数据的 Pearson 相关系数如图 4 所示。由图 4 可知, 彭山区这些站点处水域的总磷状况存在一定相关性, 上下游站点之间的相关性更强。

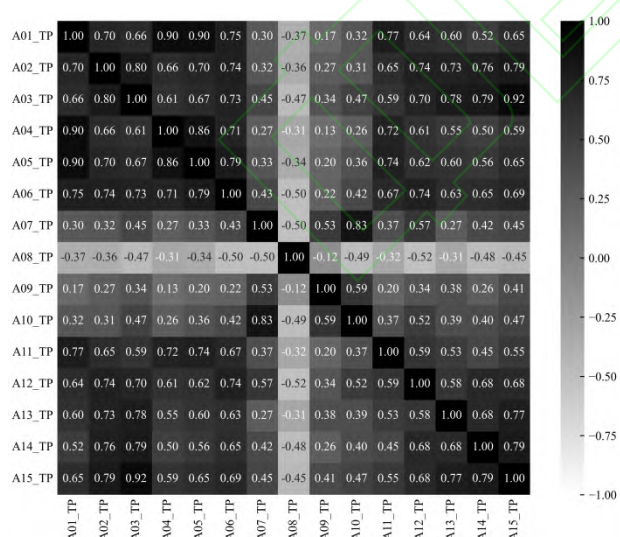


图 4 监测站点间总磷浓度数据的 Pearson 相关系数

Figure 4 Pearson correlation coefficient of total phosphorus concentration data between monitoring stations

位于湄洲河的监测站点, 总磷浓度的相关性比较强, 相关性系数在 0.52~0.8 之间。位于西干渠的三个站点, 总磷浓度的相关性也比较强, 相关性系数在 0.71~0.86 之间。由于站点 1 和站点 4 是由其上游河流的两个分支, 这两个站点的总磷浓度相关性很强, 相关系数达到了 0.9, 也从侧面说明上游到这两个站点的支流上的其他污染源可能比较少, 没有影响站点 1 和站点 4 的总磷浓度。位于毛河的监测站点中, 站点 7、8、9 和 10 位于上游, 站点 11、12 和 13 位于中游, 站点 15 位于下游且有其他河流的支流汇集而成。站点 8 的总磷浓度与其他站点的总磷浓度相关性很低且为负相关, 由实地考察得知, 站点 8 处河岸有一家养猪场, 该养猪场导致此处的水质很差, 总磷的浓度也很高。毛河上游与中下游的总磷浓度相关性不高, 说明在没有外部污染源的影响下, 由于河流自净作用, 中下游的总磷浓度有所下降。

本文选择受外部影响因素较小的流域处水质作为研究对象, 受外部污染源影响较大的监测站点的总磷不作为本文预测模型的研究对象。本文根据 Pearson 相关性分析, 以及上下游的位置关系, 筛选出相关系数较大的几个上游站点作为预测模型的特征数据。本文以站点 3 为研究对象进行总磷浓度预测, 选择站点 2、12、13、14、15 的总磷浓度时间序列数据以及站点 3 的数据作为模型数据集。

2.3 TCN-TPA 模型超参数设置

TCN-TPA 河流总磷浓度预测模型的结构主要由时间模式注意力机制层和时间卷积网络层组成。模型代码以 Python 语言编写, 采用 Tensorflow 框架和 keras 构建深度学习网络模型。模型中的本模型设计的超参数有 TCN 残差模块数量、卷积核数量、卷积核大小、扩张因子、迭代次数、批次大小和 dropout 率。将残差模块数量设置为 nb_stacks=[1,2,3], 卷积核数量设置为 nb_filters=[32,64,128], 卷积核大小设置为 kernel_size=[1,2,3,4,5,6], 扩张因子设置为 dilations=[3,4,5,6,7]。

本文调用 SKlearn 库中的 GridSearchCV 方法对超参数进行网格搜索, 得到如下的最优参数: TCN 层由一个残差模块堆叠, 卷积核数量为 64, 卷积核大小为 3, 扩张因子设置为 7, 模型训练的迭代次数为 80, 批次大小为 14, dropout 为 0.3。采用均方误差 MSE 作为模型的损失函数, 初始的学习率设置为 0.001, 采用自适应的方式调整学习率, 当模型性能不再提升时减小学习率, 缩放率设置为 0.7, 模型采

用优化器 Adam 来优化参数。

对于输入的时间步长 T ，本文分别设置为 $T=[6,8,10,12]$ ，利用模型进行预测，得到如下表 2 的预测结果。

表 2 不同输入时间步长的预测结果

Table 2 Prediction results of different input timesteps

输入时间步长 T	RMSE	MAE	MAPE
6	0.01125	0.00765	3.806%
8	0.01062	0.00738	3.617%
10	0.00955	0.00637	3.136%
12	0.01123	0.00756	3.664%

由表 2 可知，当输入的时间步长 $T=10$ 时，基于 TCN-TPA 河流总磷浓度预测模型的预测精度最高。因此，本文选用输入时间步长为 10，即利用前 40 小时的数据来预测总磷浓度的未来值。

2.4 基于 TCN-TPA 模型的总磷预测结果与分析

为了说明本文所提出的多站点 TPA-TCN 预测模型的优势，设置如下对比模型进行实验。

1) 为了验证 TPA 机制在河流总磷浓度预测上的提升，利用基于传统注意力机制的 AM-TCN 模型作为对比模型。

2) 不采用任何注意力机制的时间卷积网络 (TCN) 模型。

3) 长短期记忆网络 (LSTM) 模型，具有记忆的功能，可以缓解循环神经网络模型中梯度消失问题，是深度学习中时间序列预测最常用的模型^[30]。

由于目标站点总磷浓度的未来值受周围上游水环境的影响，为 TPA-TCN 模型和上述对比模型设置单站点输入的预测和多站点输入的预测，分别利用单站点总磷浓度和利用多个相关站点总磷浓度进行预测，TPA-TCN 模型与其他时间序列预测模型的河流总磷浓度的预测精度如表 3 所示，这几种模型预测结果和真实值的对比如图 5 所示。

由图 5 可知，采用 TCN-TPA 模型进行单站点预

测和多站点预测，总磷浓度预测值的变化趋势与真实值的变化趋势一致，但是多站点预测较单站点预测的拟合程度更好，预测值也更接近于真实值。由表 3 的河流总磷浓度预测评价指标结果可知，这几种单站点输入模型比多站点输入模型的 RMSE、MAE 和 MAPE 值均要高一些，其中多站点输入的 TPA-TCN 模型预测得到的这三个评价指标相比单站点输入的 TPA-TCN 模型分别降低了 36.29%、28.18%和 25.26%。多站点预测效果比单站点预测效果好，说明加入了周围上游水域相关性较强的总磷浓度时序数据，可以更好地进行预测。

对于多站点输入，由表 3 可知 TPA-TCN 模型的预测精度要高于其他三种模型。TCN 模型相较于 LSTM 模型，RMSE 指标、MAE 指标和 MAPE 指标均要低一些。在 TCN 网络中加入注意力机制后，AM-TCN 模型和 TPA-TCN 模型相较于 TCN 模型预测得到的误差更小一些。相对于 TCN 模型，多站点输入的 TPA-TCN 模型的 RMSE 指标降低了 13.65%，MAE 指标降低了 13.92%，MAPE 指标降低了 12.30%。相比于 AM-TCN 模型，多站点输入的 TPA-TCN 模型的这三种评价指标分别降低了 10.24%、10.78%和 9.94%。实验结果说明在时间序列预测模型中，加入注意力机制可以捕捉到时间序列中更重要的历史信息，可以提高模型的预测效果。相比于加入传统的注意力机制 (AM)，在模型中加入时间模式注意力机制 (TPA) 得到了更低的预测误差，提高了总磷浓度预测的准确率。由图 5(b)可以看出，AM-TCN 模型预测的曲线在有些时刻相对真实值偏离较大，而 TPA-TCN 模型的预测结果曲线更加贴近实际监测得到的值。

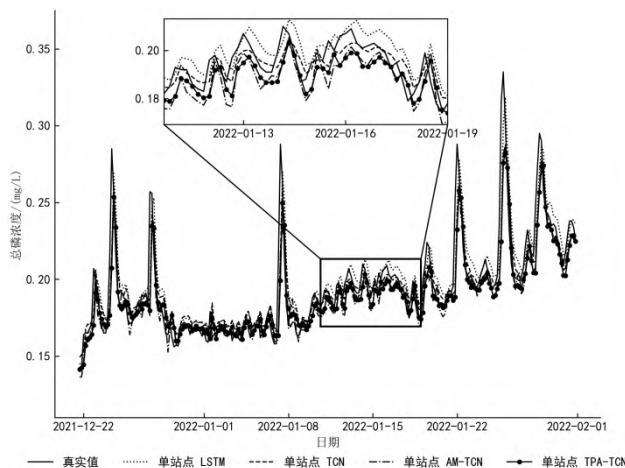
整体来看，考虑周围水域总磷时空相关性的多站点输入模型的预测误差均小于单站点输入模型的预测误差；TPA-TCN 模型的预测误差<AM-TCN 模型的预测误差<TCN 模型的预测误差<LSTM 模型的预测误差。结果证明，本文提出的多站点输入 TPA-TCN 河流总磷预测模型有着较优的预测精度。

表 3 河流总磷浓度预测精度

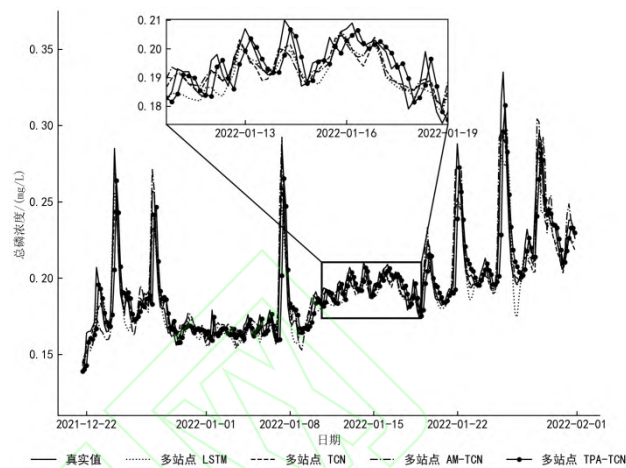
Table 3 Prediction accuracy of total phosphorus concentration in rivers

数据来源	预测模型	RMSE	MAE	MAPE
单站点	LSTM	0.01589	0.00980	4.726%
	TCN	0.01520	0.00913	4.417%
	AM-TCN	0.01561	0.00920	4.379%
	TPA-TCN	0.01499	0.00887	4.196%

多站点	LSTM	0.01128	0.00767	3.725%
	TCN	0.01106	0.00740	3.576%
	AM-TCN	0.01064	0.00714	3.482%
	TPA-TCN	0.00955	0.00637	3.136%



a—单站点输入预测结果



b—多站点输入预测结果

图 5 河流总磷浓度预测结果

Figure 5 Prediction results of total phosphorus concentration in rivers

3 讨论

由于周围上游水域的总磷浓度对目标站点处的总磷浓度值有一定影响，多站点输入模型比单站点输入模型的预测效果更好。输入周围上游相关性较强的总磷浓度时序数据，增加了输入维度，模型能够提取到更多特征，预测效果也就更好。许佳辉等人利用图神经网络构建水质监测站点在空间上的复杂关系，在预测水质的时候可以将周围监测站点的数据输入模型中，相比单序列的深度学习预测模型同时考虑了时间和空间两方面的依赖关系，也获得了更好地预测效果^[31]。这与本研究分析河流水质监测站点时空关系，使用输入周围相关性较强的监测站点总磷浓度时序数据的多站点输入模型来进行总磷预测的结论一致。

深度学习中有许多时间序列预测的算法模型，这些模型越来越多地被应用到水质预测中。TCN 网络是近些年提出的一个可以进行时间序列预测的模型，TCN 由残差模块构成，且残差模块中使用了扩张因果卷积，可以获取更大的感受野，利用更多的历史信息来预测未来时刻的值。本研究提出的多站点输入的 TPA-TCN 网络模型对河流总磷预测的效果比 LSTM 模型和 TCN 模型的预测效果更好。

在 TCN 网络中融入注意力机制，获得了更高的

预测精度，这是因为注意力机制能够捕捉到时间序列中不同时间步各个输入站点总磷浓度值对目标站点未来时刻的总磷浓度值的影响，利用权重的分配，计算得到当前时间步更重要的信息，从而提高模型预测的效果^[32]。本研究提出的模型是加入了时间模式注意力机制，相比于加入传统的注意力机制，本文的 TPA-TCN 模型的预测结果曲线更加贴近真实情况。这是由于传统的时间注意力机制对周围相关性较强的站点分配了相同的时序注意力。如果这些站点对目标站点的总磷浓度的影响在时间步上是一致的，那么预测的效果就会好一些，但是，这些站点对目标站点总磷浓度的影响不同，则就会产生较大的误差。而 TPA 机制通过一维卷积提取了隐藏状态矩阵行向量的特征，捕捉到了其他站点与目标站点总磷浓度之间的复杂关系，注意力机制函数生成上下文向量来选择不同时间上更重要的信息，从而充分利用输入序列所包含的信息^[33]。因此在分配注意力权重的时候，可以在不同的时间步上为不同的站点使用不同的注意力权重，从而达到了较优的预测精度，预测曲线也更加贴近于真实值。

4 结论

本文提出了基于时间模式注意力的时间卷积网络河流总磷浓度预测模型，利用 Pearson 相关系数进

行水质时空关系分析,选择与目标站点处总磷浓度相关性较强的周围上游站点总磷浓度时序数据作为输入,使用时间卷积网络对历史信息进行特征提取,使用时间模式注意力机制为不同历史时间步的不同站点的时序数据分配不同的权重,实现对四川省彭山区岷江支流监测站点处总磷浓度的预测。具体结论如下:

1) 彭山区河流上下游总磷浓度存在一定相关性,上下游站点之间的总磷浓度相关性更强。

2) 相比单站点输入模型,利用周围水域上游的数据来构建多站点输入的深度学习模型更能反映河流总磷浓度的变化。

3) TCN 网络对彭山区河流总磷浓度的预测性能优于 LSTM 网络,TPA 机制相比于传统注意力机制,其在处理多站点输入的预测任务上有着更好的性能。因此,本文提出的多站点 TPA-TCN 模型有着更高的预测精度。

本文的方法主要考虑了周围水域的水质时序数据,没有考虑影响河流水质的其他因素,如站点周边污染排放信息和气象数据等,因此将影响河流水质的其他因素引入模型是下一步研究的重点。

参考文献

- [1] HORN AL, RUEDA F, HORMANN G, et al. Implementing river water quality modelling issues in mesoscale watershed models for water policy demands—an overview on current concepts, deficits, and future tasks[J]. *Physics and Chemistry of the Earth*, 2004, **29**(11-12): 725-737.
- [2] 刘东君, 邹志红. 最优加权组合预测法在水质预测中的应用研究[J]. *环境科学学报*, 2012, **32**(12): 3128-3132.
- [3] 李娜, 李勇, 冯家成, 等. 太湖水体 Chl-a 预测模型 ARIMA 的构建及应用优化[J]. *环境科学*, 2021, **42**(5): 2223-2231.
- [4] 曾一川, 曾会国, 袁伟皓, 等. 长江口入海通道水质综合分析与模型预测[J]. *环境工程*, 2022, **40**(5): 95-102+108.
- [5] GUO T, HE W, JIANG Z L, et al. An improved LSSVM model for intelligent prediction of the daily water level[J]. *Energies*, 2019, **12**(1): 112.
- [6] 刘世存, 杨薇, 田凯, 等. 基于多层全连接神经网络的白洋淀水质预测[J]. *农业环境科学学报*, 2020, **39**(6): 1283-1292.
- [7] 刘攀, 郑雅莲, 谢康, 等. 水文水资源领域深度学习研究进展综述[J]. *人民长江*, 2021, **52**(10): 76-83.
- [8] HUAN J, CHEN B, XU X G, et al. River dissolved oxygen prediction based on Random Forest and LSTM[J]. *Applied Engineering in Agriculture*, 2021, **37**(5): 901-910.
- [9] 董泉汐. 基于深度学习的水环境时间序列预测方法研究[D]. 北京: 北京工业大学, 2020.
- [10] 张贻婷, 李天宏. 基于长短时记忆神经网络的河流水质预测研究[J]. *环境科学与技术*, 2021, **44**(8): 163-169.
- [11] 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.
- [12] YANG Y R, XIONG Q Y, WU C, et al. A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism[J]. *Environmental Science and Pollution Research*, 2021, **28**(39): 55129-55139.
- [13] BAI S, KOLTER JZ, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [C]. arXiv:1803.01271, 2018.
- [14] 李荆, 刘钰, 邹磊. 基于时空建模的动态图卷积神经网络[J]. *北京大学学报(自然科学版)*, 2021, **57**(4): 605-613.
- [15] ZHANG Y F, THORBURN P J, FITCH P. Multi-task temporal convolutional network for predicting water quality sensor data[C]// *International Conference on Neural Information Processing*, Springer, 2019: 122-130.
- [16] FU Y X, HU Z H, ZHAO Y C, et al. A long-term water quality prediction method based on the temporal convolutional network in smart mariculture[J]. *Water*, 2021, **13**(20): 2907.
- [17] LI W S, WEI Y G, AN D, et al. LSTM-TCN: dissolved oxygen prediction in aquaculture, based on combined model of long short-term memory network and temporal convolutional network[J]. *Environmental Science and Pollution Research*, 2022, **29**(26): 39545-39556.
- [18] SHIH S Y, SUN F K, LEE H Y. Temporal pattern attention for multivariate time series forecasting[J]. *Machine Learning*, 2019, **108**(8-9): 1421-1441.
- [19] 崔鸿雁, 徐帅, 张利锋, 等. 机器学习中的特征选择方法研究及展望[J]. *北京邮电大学学报*, 2018, **41**(1): 1-12.
- [20] 张阳, 洗慧婷, 赵志杰. 基于空间相关性和神经网络模型的实时河流水质预测模型[J]. *北京大学学报(自然科学版)*, 2022, **58**(2): 337-344.
- [21] SCHÖBER P, BOER C, SCHWARTE L A. Correlation coefficients: appropriate use and interpretation[J]. *Anesthesia and Analgesia*, 2018, **126**(5): 1763-1768.
- [22] PANTISKAS L, VERSTOEP K, BAL H. Interpretable multivariate time series forecasting with temporal attention convolutional neural networks[C]// *IEEE Symposium Series on Computational Intelligence*, 2020: 1687-1694.
- [23] ZHAI N J, ZHOU X F. Temperature prediction of heating furnace based on deep transfer learning[J]. *Sensors*, 2020, **20**(17): 4676-4702.

-
- [24] WAN R Z, MEI S P, WANG J, et al. Multivariate temporal convolutional network: a deep neural networks approach for multivariate time series forecasting[J]. Electronics, 2019, **8**(8): 876-893.
- [25] ZHOU X H, WANG J P, CAO X K, et al. Simulation of future dissolved oxygen distribution in pond culture based on sliding window-temporal convolutional network and trend surface analysis[J]. Aquacultural Engineering, 2021, **95**: 102200.
- [26] 王渝红, 史云翔, 周旭, 等. 基于时间模式注意力机制的 BiLSTM 多风电机组超短期功率预测[J]. 高电压技术, 2022, **48**(5): 1884-1892.
- [27] SONG C G, YAO L H, HUA C Y, et al. A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long short-term memory[J]. Journal of Hydrology, 2021, **603**(A): 126879.
- [28] 李光, 吴祈宗. 基于结论一致的综合评价数据标准化研究[J]. 数学的实践与认识, 2011, **41**(3): 72-77.
- [29] LIU Y Q, ZHANG Q, SONG L H, et al. Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction[J]. Computers and Electronics in Agriculture, 2019, **165**: 104964.
- [30] ZHOU J, WANG Y Y, XIAO F, et al. Water quality prediction method based on IGRA and LSTM[J]. Water, 2018, **10**(9): 1148.
- [31] 许佳辉, 王敬昌, 陈岭, 等. 基于图神经网络的地表水水质预测模型[J]. 浙江大学学报(工学版), 2021, **55**(4): 601-607.
- [32] 陈海涵, 吴国栋, 李景霞, 等. 基于注意力机制的深度学习推荐研究进展[J]. 计算机工程与科学, 2021, **43**(2): 370-380.
- [33] 王竹荣, 薛伟, 牛亚邦, 等. 基于注意力机制的泊位占有率预测模型研究[J]. 通信学报, 2020, **41**(12): 182-192.