

# 结合光谱降维的 IPSO-SVR 水体 总磷浓度预测模型

王彩玲, 张国浩

(西安石油大学 计算机学院, 陕西 西安 710065)

**摘要:** [目的] 选择最优模型对水体中总磷浓度进行预测, 为准确、实时、高效检测水资源状况提供支持。[方法] 以 2021 年在长江中下游武汉—安徽地区采集的水质样本作为研究对象, 首先, 对采集到的长江光谱数据进行最大最小归一化和均值中心化两种预处理操作以便统一数据的范围和均值点, 并使用核主成分分析(KPCA)技术对预处理后的光谱数据进行降维操作。选取方差解释率为 99.6% 下的 6 个特征向量进行后续预测模型的训练, 接着在原有粒子群算法的基础上引入自适应惯性权重更新公式和遗传—模拟退火变异思想, 提高算法的寻优能力。使用改进的粒子群优化算法对支持向量回归模型中的超参数组合进行寻优, 对支持向量回归模型使用输出的结果进行预测模型的训练, 最后使用测试集数据进行总磷浓度的预测。[结果] 提出了一种结合光谱降维的改进粒子群优化算法(IPSO)结合支持向量回归(SVR)的水体总磷含量预测模型。通过和当前预测性能较好的几种机器学习模型进行精度的比较发现, 该试验模型对长江水体总磷浓度进行预测时决定系数( $R^2$ )为 0.973 920, 均方根差(RMSE)为 0.003 012, 平均绝对误差(MAE)为 0.002 105。[结论] 使用光谱数据结合降维技术、粒子群优化算法和机器学习模型的算法融合模型检测水体总磷浓度可行性强, 精确度高, 且拟合效果良好。

**关键词:** 高光谱; IPSO-SVR 模型; KPCA 降维; 长江水质; 总磷浓度检测

**文献标识码:** B

**文章编号:** 1000-288X(2024)02-0196-09

**中图分类号:** O657.3, P237, 131.2

**文献参数:** 等. 结合光谱降维的 IPSO-SVR 水体总磷浓度预测模型[J]. 水土保持通报, 2024, 44(2): 196-204. DOI: 10.13961/j.cnki.stbctb.2024.02.021; Wang Cailing, Zhang Guohao. Predictive model of total phosphorus concentration in IPSO-SVR waters combined with spectral downscaling [J]. Bulletin of Soil and Water Conservation, 2024, 44(2): 196-204.

## Predictive Model of Total Phosphorus Concentration in IPSO-SVR Waters Combined with Spectral Downscaling

Wang Cailing, Zhang Guohao

(College of Computer Science, Xi'an Shiyou University, Xi'an, Shaanxi 710065, China)

**Abstract:** [Objective] The optimal model for predicting total phosphorus concentration in water bodies was studied in order to provide support for accurate, real-time, and efficient monitoring of water resources. [Methods] Water quality samples were collected in 2021 from the Wuhan-Anhui region in the middle and lower reaches of the Yangtze River. Firstly, the collected spectral data of the Yangtze River was preprocessed by both maximum-minimum normalization and mean centering to unify the range and mean point of the data. Kernel principal component analysis (KPCA) was then used to perform dimensionality reduction on the preprocessed spectral data. Six feature vectors were selected based on a variance explanation rate of 99.6% for training the subsequent prediction model. Next, an improved particle swarm optimization (IPSO) algorithm was proposed by introducing an adaptive inertia weight updating formula and a genetic-simulated annealing mutation concept to enhance the optimization ability of the algorithm. The improved particle swarm optimization algorithm was used to optimize the hyperparameter combinations in the support vector regression

收稿日期: 2023-10-10

修回日期: 2023-11-19

资助项目: 陕西省重点研发计划项目“结合深度学习的高光谱水质监测模型研究”(2023-YBSF-437); 国家自然科学基金(61401439; 41301382; 31160475; 62002286; 62276213)

第一作者: 王彩玲(1984—), 女(汉族), 宁夏回族自治区吴忠市人, 博士, 副教授, 硕士生导师, 主要研究方向为高光谱成像中的信息提取技术、深度学习。Email: azerling@163.com.

(SVR) model. The support vector regression model was trained using the output results to predict the total phosphorus concentration. Finally, the test set data were used to predict total phosphorus concentration. [Results] A prediction model for total phosphorus content in water using an improved particle swarm optimization (IPSO) combined with support vector regression (SVR) and spectral dimensionality reduction was proposed. The experimental model achieved an  $R^2$  of 0.973 920, a root mean square error of 0.003 012, and a mean absolute error of 0.002 105 when predicting total phosphorus concentrations in the Yangtze River water. [Conclusion] The proposed method of using spectral data combined with dimensionality reduction techniques, particle swarm optimization algorithms, and machine learning models was determined to be feasible and effective in total phosphorus concentration measurement. The accuracy and fitting effects of the model were better than the accuracy obtained with several well-performing machine learning models.

**Keywords:** hyperspectral; IPSO-SVR modeling; KPCA downscaling; Yangtze River water quality; total phosphorus concentration measurement

目前,随着经济快速发展和人类活动的加剧,农业活动、城市排放、工业废水和生活污水等因素导致水体中的含氮磷物质不断增加,当总磷浓度过高时,会导致藻类和水生植物过度生长。这些大量的藻类和水生植物会消耗水中的氧气,并在其死亡和分解过程中释放出有机物质,使水体的溶解氧含量降低,进一步加剧水体富营养化问题。此外,过高的总磷浓度还可能引发藻毒素的产生,对水生生物及人类健康构成潜在风险。因此,监测和控制水体中的总磷浓度对于及早发现和应对水体富营养化问题至关重要。通过准确实时地检测总磷浓度,可以及时地根据当前水体状况采取相应的管理和治理措施,从而促进水资源的可持续利用。

作为常见的水质污染物之一,总磷浓度的大小能够直观反映出水质是否健康。目前主要使用分光光度法、化学发光法、电化学发光法和色谱法等来对总磷含量进行检测,但这些方法都存在灵敏度低、不方便携带、有容量限制等问题,不能准确、实时、高效地对样本中的总磷浓度进行检测分析<sup>[1]</sup>。为了更好地实现水资源的保护,选择一种合适的水质总磷浓度检测方法具有良好的理论研究和实践应用价值。

高光谱检测技术可以在很短时间内高效地对样本中光谱数据进行检测,因此很多学者提出结合机器学习支持向量机、随机森林等技术对所测得的高光谱数据进行模型训练的研究。机器学习中支持向量回归模型(support vector regression, SVR)对于回归问题具有十分优秀的表现性能,许多研究将 SVR 等机器学习方法应用在水质中物质含量的检测领域,促进了水质检测的发展。薛同来等人对 SVR 模型进行 GA 优化从而实现水质预测,预测模型的平均误差为 0.009 443,具有十分优秀的性能<sup>[2]</sup>;李建文等<sup>[3]</sup>提出了 EEMD-SVR 模型用于对渔业养殖时的水质进

行预测,满足了实际渔业养殖水质精细化管理的需求;康贝等人<sup>[4]</sup>利用局部线性嵌入结合支持向量回归法提出了 LLE-SVR 模型,提高了预测精度,实现了水质中 COD 的检测;康成龙<sup>[5]</sup>基于 SERS 技术和 GWO-SVR 模型实现了水中污染物的检测;姜吉光等人<sup>[6]</sup>采用连续投影算法对水质样本进行筛选后结合支持向量回归对水质中污染物含量进行解耦预测,实现了水质中物质的同步预测。但在现有的文献报告中,利用水体透射光谱数据结合机器学习方法对水质中总磷浓度预测模型进行训练的报道较少。为了更好地对水资源进行检测和保护,本文提出了一种新的策略。该策略在使用降维技术降低光谱数据维度的基础上,将改进的粒子群优化(improved particle swarm optimization, IPSO)算法和 SVR 模型进行结合,并使用该 IPSO-SVR 融合算法进行水体中总磷浓度的检测。

长江作为中国两大河流之一,水体健康程度直接关系到沿岸居民的健康和经济发展。本试验以长江中下游武汉—安徽地区水体为研究对象,对采集到的水样首先使用光谱仪获取原始透射光谱数据,接着使用核主成分分析技术对光谱数据进行降维,选择合适的主成分个数。最后使用本文所提出的 IPSO-SVR 模型进行水质总磷浓度预测。同时将该模型和当前预测性能较好的几个机器学习模型进行预测精度的比较,选择出最优模型来进行水体中总磷浓度的预测。研究所测得的高光谱数据对水体中总磷浓度进行估计的可行性以及最优方法,为实时检测水资源状况提供了新思路。

## 1 材料与方 法

本节将详细介绍本试验所使用的方法及原理,具体流程框架如图 1 所示。

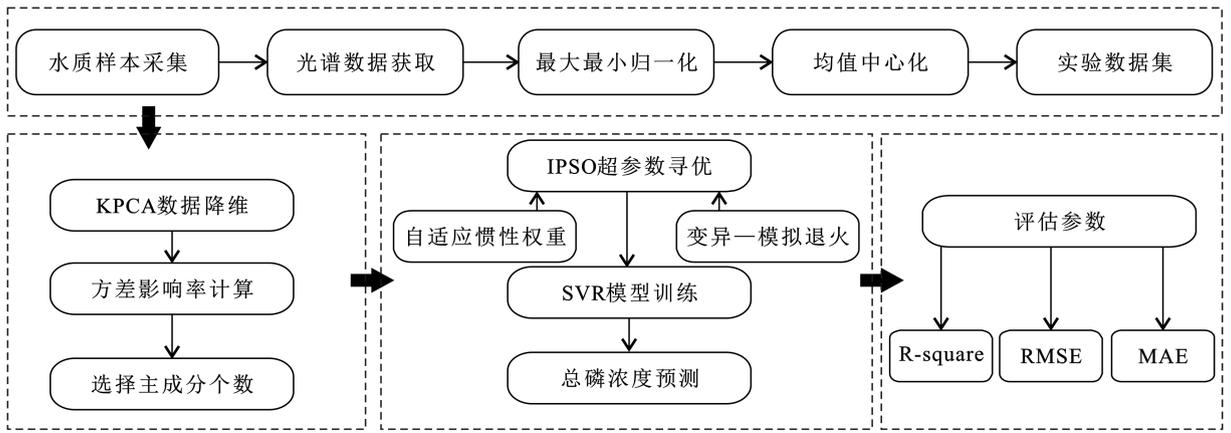


图 1 水体污染物预测模型流程框架图

Fig.1 Framework diagram of water pollutant prediction model in water bodies

## 1.1 光谱数据的获取

1.1.1 采集区域 选择长江流域中湖北—安徽地段作为研究区域,该区域是长江中下游的重要部分,涵盖了长江经济带的核心区域,在中国具有代表性和重要性。本试验选定研究区域中 47 个地点进行水质的分析研究。采样时间为 2021 年 5 月 25 日到 2021 年 6 月 10 日,其中每个地点均采集 5 次,共获得 235 个样本数据,以确保样本的代表性和可靠性。图 2 展示了本试验选择的采集区域的空间分布。通过在区域内选择多个地点进行采集,可以更全面地了解该区域的水质特征,并为管理和保护水资源提供科学依据。

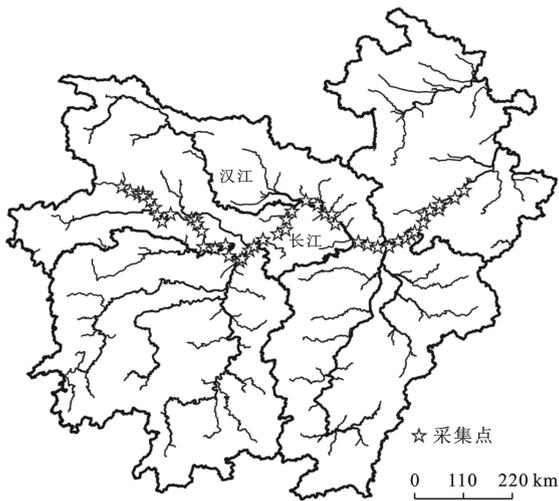


图 2 样本采集区域示意图

Fig.2 Schematic diagram of sample collection area

1.1.2 光谱数据获取 首先使用 OCEAN-HDX-XR 微型光纤光谱仪(美国 Ocean Optics 公司)采集长江样本的光谱数据,该光谱仪具有高分辨率,微型化和宽波长范围等特点,是一种常用的光谱测量设备。接着进行了背景校正操作,即使用光谱仪测量去离子水

(纯水)下的背景光谱,后将样本的光谱值与背景光谱值相除得到样本的光谱透射率大小,该操作消除了背景光谱的影响,从而获取样本光谱的相对强度信息。本试验联合中国科学院西安光学精密机械研究所于涛课题组进行样本采集和光谱数据获取,以确保数据集的有效性和可靠性。

## 1.2 IPSO-SVR 基本原理

1.2.1 SVR 算法 SVR 是一种基于支持向量机(support vector machine, SVM)的回归算法,它通过寻找最优的超平面来拟合输入数据与输出数据之间的关系,主要用于解决回归问题。SVR 是基于核函数的非线性方法,核心思想在于通过核函数将低纬度的数据映射到高维度空间,从而让原始数据在高纬度空间下线性可分,后在高纬度下对数据进行处理<sup>[7-8]</sup>。SVR 可以有效地对异常点问题进行解决,泛化性能好,可以适应各种不同类型的数据,且不容易出现过拟合的现象,对于数据的要求低。SVR 模型主要通过惩罚参数( $C$ )、 $\epsilon$ -不敏感区间(epsilon)以及内核函数等超参数构成,在进行 SVR 模型的训练时,选择合适的内核函数以及超参数的大小非常重要,对于线性可分的数据,多项式核函数可以处理具有多项式特征的数据,径向基函数核是最常用的内核函数之一,可以用来处理非线性可分的数据,Sigmoid 核函数可以用来处理具有 S 形特征的数据<sup>[9-10]</sup>。

1.2.2 IPSO 算法 本研究所提出的 IPSO 算法是在粒子群优化(particle swarm optimization, PSO)算法的基础上引入了新的改进策略,以增强算法的寻优能力,使其更适用于复杂的优化问题。PSO 算法是一种基于群体智能的优化算法,其主要是利用所有粒子群体间的信息共享从而使所有粒子的运动在求解空间中产生从无序到有序的演化过程,进而求得最优解,目前被广泛应用于机器学习模型的优化、神经网络

络的训练、图像和信号处理等领域<sup>[11]</sup>。PSO 算法的基本思想为:首先在搜索空间中根据所设定的粒子个数对各个粒子的速度和位置进行随机初始化,之后每个粒子根据自身的历史最优位置和粒子群的全局最优位置来对自己的速度和位置进行更新,同时更新每个粒子的历史最优位置和粒子群中的全局最优位置,通过不断迭代粒子群逐渐在搜索空间中找到全局最优解<sup>[12]</sup>。在 PSO 算法中,粒子被赋予速度和位置,个体的历史最优位置(Pbest)与全局最优位置(Gbest)用于引导粒子的运动,假设空间的维度为  $D$ ,则第  $i$  个粒子的位置和位置表示如公式(1)–(2)所示:

$$X_i = (x_{i1}, x_{i2}, x_{i3} \cdots x_{iD}), \quad (i=1, 2 \cdots N) \quad (1)$$

$$V_i = (v_{i1}, v_{i2}, v_{i3} \cdots v_{iD}), \quad (i=1, 2 \cdots N) \quad (2)$$

式中: $X_i, V_i$  分别表示第  $i$  个粒子的位置和速度大小; $x_{i1}, x_{i2}, x_{i3} \cdots x_{iD}$  和  $v_{i1}, v_{i2}, v_{i3} \cdots v_{iD}$  分别为第  $i$  个粒子在所有维度下的分量大小。

第  $i$  个粒子的第  $d+1$  次迭代的速度和位置更新公式如公式(3)–(4) 所示:

$$V_i^{d+1} = \omega v_i^d + c_1 \cdot r_1 \cdot (Pbest_i^d - x_i^d) + c_2 \cdot r_2 \cdot (Gbest_i^d - x_i^d) \quad (3)$$

$$x_i^{d+1} = x_i^d + V_i^{d+1} \quad (4)$$

式中: $Pbest_i^d$  是第  $i$  个粒子在第  $d$  次迭代时的历史最优解; $Gbest_i^d$  是第  $i$  个粒子在第  $d+1$  次迭代时的群体最优解; $v_i^d, x_i^d$  分别为第  $i$  个粒子在第  $d$  次迭代时的速度和位置大小; $v_i^{d+1}, x_i^{d+1}$  为第  $i$  个粒子在第  $d+1$  次迭代时的速度和位置大小; $\omega$  为惯性权重; $c_1, c_2$  分别为个体学习因子和社会学习因子; $r_1, r_2$  为  $0 \sim 1$  范围内的随机数。

由于 PSO 算法具有全局搜索能力强,并行性好,对问题的适应性强和参数设置简单等优点,被广泛应用于函数优化、机器学习和图像处理等方面。然而,该算法却存在容易陷入局部最优解,对高维问题的搜索能力有限等不足。为进一步提升算法在复杂问题中的表现性能,本文在原有算法的基础上引入惯性权重参数自适应更新和粒子遗传—模拟退火变异策略,以获得更好的优化结果。

在原始的粒子群算法中,惯性权重通常为固定的常数。较小的惯性权重会导致粒子陷入局部最优解,而较大的惯性权重会导致粒子在搜索空间内震荡,不利于算法的寻优和收敛<sup>[13-14]</sup>。为解决这一问题,本试验提出惯性权重自适应更新策略,让其根据当前的训练效果动态地调整自身大小<sup>[15]</sup>。该更新策略如公式(5)–(6)所示。

$$\omega = \omega - k \cdot (\omega_{\max} - \omega_{\min}) \cdot \text{softsign} \left( \frac{f_i - f_{\text{avg}}}{f_{\text{avg}} - f_{\min}} \right) \quad (5)$$

$$\text{softsign}(x) = x / (1 + |x|) \quad (6)$$

式中: $\omega_{\max}, \omega_{\min}$  表示惯性权重  $\omega$  的最大值和最小值; $f_i$  表示当前粒子的函数适应值; $f_{\min}$  表示粒子群体中最小的函数适应值,相当于此时粒子群中的全局最优粒子; $f_{\text{avg}}$  表示粒子群中的所有粒子的平均适应值; $k$  为  $0 \sim 1$  范围内的随机数。

通过公式(5)可知,该更新策略是根据粒子自身适应度值、群体所有粒子的平均适应度值和最小适应度值间的大小关系来进行判断。如果当前粒子的适应度值大于群体平均适应度值,表示粒子距离全局最优解较远,此时应增大惯性权重,让粒子更好地进行探索。当粒子的适应度值小于平均适应度值,表示此时粒子距离全局最优解较近,此时应减少惯性权重,从而加快粒子的收敛速度。

引入自适应惯性权重更新公式后粒子群体可以根据当前的寻优程度自适应地调整参数的大小,以获得最佳的效果。但由于群体中的粒子在每轮迭代时都会受到自身历史最优解和全局最优解的吸引,因此随着迭代次数的增加,所有粒子会逐渐收敛到较小的局部区域中,该现象导致粒子群体容易陷入局部最优解,为了进一步减少粒子在迭代过程中陷入局部最优解的问题,增加粒子寻优过程中的多样性,本试验在遗传算法<sup>[16]</sup>和模拟退火算法<sup>[17-18]</sup>的思想提出遗传—模拟退火变异策略。

该策略在每轮迭代时都会依次遍历群体中的所有粒子,并根据变异规则来判断粒子是否进行变异。该变异规则为:①随机生成一个  $-1 \sim 1$  内的随机数作为扰动变量,后粒子的每个维度都与该扰动变量相乘得到和粒子位置大小相同的扰动向量,将该扰动向量和原始粒子位置相加即为变异后新粒子的位置。②使用新粒子的适应度值减去旧粒子的适应值得到差值,根据该差值的大小和随机数两个条件来判断粒子是否进行变异。③如果差值小于  $0$ ,表明新粒子的位置更优,此时进行粒子的变异;如果差值大于  $0$ ,表明旧粒子的位置更优,此时随机生成一个  $0 \sim 1$  内的随机数,如果该随机数小于  $0.1$ ,则进行粒子的变异,如果大于  $0.1$ ,则不进行更新,以增加粒子的随机性。扰动向量和新粒子位置的计算公式为:

$$e = r \cdot X(i) \quad (7)$$

$$P_{\text{new}}(i) = P(i) + e \quad (8)$$

式中: $r$  为  $-1 \sim 1$  间的随机数; $e$  为扰动向量; $P(i)$  为粒子的原始位置; $P_{\text{new}}(i)$  为新粒子位置。

1.2.3 IPSO-SVR 融合算法 针对水质中总磷浓度预测时具有的复杂性和非线性等特点,本试验选择使用 SVR 模型对浓度进行预测。然而,SVR 算法的性

能很大程度上依赖于其超参数的选择,但传统的参数寻优方法如网格搜索和随机搜索需要对超参数空间进行遍历,对计算资源和空间需求较大。为了解决这一问题,本试验采用了 IPSO 算法对 SVR 的超参数进行寻优,提高模型的优化效果和计算效率。这种优化算法和回归算法的结合可以更有效地利用各自的优势,优化预测模型的性能。

## 2 结果与讨论

### 2.1 原始光谱数据获取

图 3 展示了采集的长江水体原始光谱数据图。从图 3 可以观察到,不同样本下的光谱透射率虽然存在一定的差异,但相对较小,仅凭肉眼观察很难对不同浓度下的数据进行明显区分。样本光谱数据在波段区间为 178.82~350,480~530,530~750 nm 中呈先下降后上升的趋势,光谱曲线波谷分布于 180~210 nm 范围内,在 570 nm 波长周围出现极大的吸收峰,浓度不同吸收峰的高度也有所不同,主要表现为随着总磷浓度的增加,总磷在各个波段的光谱透射率值逐渐升高。同时,在使用光谱仪进行数据采集时,试验结果可能会受到环境中的噪声和其他因素的干扰,例如周围光线的干扰、水样本的颗粒物等。这些因素可能会对光谱数据采集造成干扰,从而影响数据的准确性和可靠性。

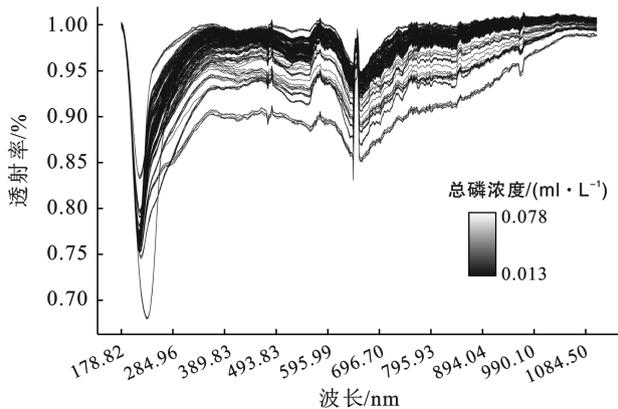


图 3 采集水体样本原始透射光谱

Fig.3 Original transmittance spectrum of water sample collection

### 2.2 光谱数据处理

2.2.1 光谱数据预处理 原始的光谱数据中不同特征之间的数值和量级存在差异。这些差异在后进行数据模型的构建时,会对试验结果造成影响,影响模型的性能。因此本试验首先使用最大最小归一化操作,将所有数据缩放到一个固定的范围中,使不同特征在构建模型时的权重更加平等,从而降低模型

过拟合的风险并提高模型的收敛速度。接着进行均值中心化操作,从而使所有数据的中心位置都位于原点附近,消除不同特征之间由于均值不同导致的平移差距<sup>[19-20]</sup>。最大最小归一化和均值中心化的计算公式为:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

式中: $x'$ 表示单个数据在进行最大最小归一化后的取值; $x$ 为该数据的原始取值; $x_{\min}$ 表示数据所在列的最小值; $x_{\max}$ 表示该数据所在列的最大值。

$$x' = \frac{(x - x_{\text{mean}})}{x_{\text{std}}} \quad (10)$$

式中: $x'$ 表示单个数据在进行均值中心化后的取值; $x$ 为该数据的原始取值; $x_{\text{mean}}$ 表示数据所在列的平均值; $x_{\text{std}}$ 表示该数据所在列的标准差。

原始光谱数据进行两种预处理操作后的光谱数据如图 4 所示。

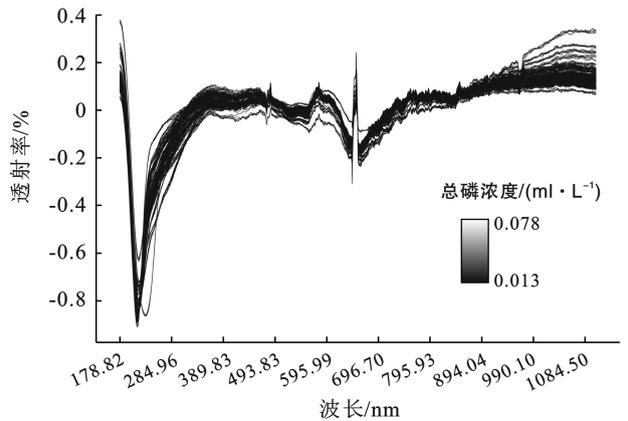


图 4 水体样本数据预处理光谱

Fig.4 Preprocessed spectrum of water sample data

2.2.2 数据降维 本试验的光谱范围为 178.2~1104.2 nm,具有很高的维度。高维度数据可能会增加计算复杂性,增加模型训练的难度,并且不同维度的特征对于模型训练的影响存在差异。因此在构建预测模型时,需要选择适当的降维技术来减少数据的维度并选择主要的特征。降低维度的目标是在保留数据信息的同时,减少冗余和噪声,提取具有较高信息量的特征,并去除对试验造成干扰的特征。核主成分分析(kernel principal component analysis, KPCA)是一种常用的降维方法,该方法在传统主成分分析(principal component analysis, PCA)的基础上引入了核技巧,通过将高维度的数据映射到更高维度的空间中使得原始数据变得线性可分,之后在新空间中进行主成分分析操作,在尽可能保留数据关键信息的基础上提取主成分<sup>[21-22]</sup>。在进行 KPCA 操作后,由于

不同主成分的个数对原始数据的解释率存在差异,因此选择合适的主成分个数十分重要。方差解释率是指每个主成分对于总方差的贡献度,代表了每个主成分的重要性,可以通过观察方差解释率曲线来进行主成分个数的选择<sup>[14]</sup>。在本试验中,使用 KPCA 技术对总磷光谱数据降维后前 15 个主成分的方差解释率和累积方差值曲线如图 5 所示。通过图 5 可以看出,随着主成分个数的增加,对总磷光谱数据的解释比例也逐渐增加。特别是前两个主成分的方差解释率分别为 0.631 0 和 0.227 0,包含了较高的信息量。这意味着前两个主成分能够捕捉原始数据中大部分的信息。相比之下,从第 3 个主成分开始,各主成分所包含的信息量都低于 0.1,这表示这些主成分与原始数据的相关性较弱。为了确保模型训练的稳定性和准确性,本试验选择了能够代表原始数据 99.60% 信息量的前 6 个主成分作为后续试验模型的输入。

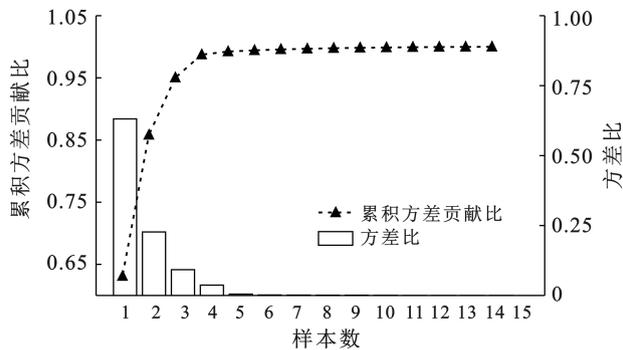


图 5 预处理样本数据方差解释率曲线

Fig.5 Explained variance curve of preprocessed sample data

## 2.3 IPSO-SVR 总磷预测模型的建立

2.3.1 粒子数目和迭代次数的选择 IPSO 算法中迭代次数和粒子数目的设定对于算法的性能和输出结果有很大影响。粒子数目和迭代次数越多,算法搜索的精度就越高,能够更好地在搜索空间中进行探索,但同时算法的时间和空间复杂度越大,可能会造成资源浪费<sup>[23-24]</sup>。因此,为了选择合适的迭代次数和粒子数目,本试验分别使用不同组合进行 SVR 超参数寻优,后使用输出的最优参数在测试集中进行模型预测效果的验证。其中 SVR 模型选择“RBF”作为内核函数,输出的结果为  $C$ ,  $\epsilon$  和  $\gamma$  这 3 个超参数的最优值。同时选择决定系数( $R^2$ ),均方根误差(RMSE)、平均绝对误差(MAE)这 3 个参数对模型预测效果进行评估,其中均方根误差 RMSE 和平均绝对误差(MAE)越小,说明模型的拟合效果越好;决定系数  $R^2$  越接近 1,模型的拟合效果越好。不同迭代次数和粒子数目下 IPSO-SVR 模型的对比结果详见

表 1。通过表 1 可以看出,在迭代次数为 20 时,随着粒子数目的增多,模型的预测效果在不断提升,即在相同的迭代次数下粒子数目越多,模型的预测精度越好。同时,在粒子数目为 50 和 60 时,进行 30 次迭代和 40 次迭代下模型的预测效果相近,表明此时再增加粒子的数目和迭代的次数已经没有特别大的提升,因此在本试验中运行 IPSO 算法时使用 50 个粒子迭代 30 次来进行最优参数组合的寻找。

表 1 不同组合下模型精度检验结果  
Table 1 Model accuracy test results under different combinations

粒子数	迭代次数	$R^2$	RMSE	MAE
40	20	0.931 565	0.004 880	0.003 616
50	20	0.935 879	0.004 724	0.002 888
60	20	0.942 351	0.004 479	0.002 747
40	30	0.960 095	0.003 726	0.002 685
50	30	0.973 920	0.003 012	0.002 105
60	30	0.972 878	0.003 214	0.002 303
40	40	0.960 635	0.003 801	0.002 804
50	40	0.974 608	0.002 924	0.002 439
60	40	0.975 479	0.002 836	0.002 045

2.3.2 预测模型构建 使用 IPSO-SVR 进行总磷水质预测模型的构建时,首先使用 IPSO 算法在给定的搜索空间中通过 50 个粒子迭代 30 次来寻找 3 个超参数的最优组合并输出至 SVR 模型中,之后在 SVR 中进行总磷预测模型的训练。在模型的训练和预测时,其中训练集占总样本的 95,测试集占总样本的 5,在训练集中使用留一交叉验证法<sup>[25]</sup>进行模型的训练和调优,最后使用测试集进行模型预测效果的验证。并将本试验所构建 IPSO-SVR 模型同当前预测性能较好的随机森林(random forest, RF)模型<sup>[17]</sup>、BPNN 模型、IPSO-BPNN 模型进行预测效果的对比,分别计算  $R^2$ , RMSE 和 MSE 这 3 个评估参数的大小。3 个模型的参数对比结果详见表 2。

表 2 模型预测精度对比

Table 2 Comparison of model prediction accuracy

预测模型	$R^2$	RMSE	MAE
IPSO-SVR	0.973 920	0.003 012	0.002 105
IPSO-BPNN	0.935 838	0.004 412	0.003 021
BPNN	0.891 490	0.006 145	0.004 533
RF	0.914 306	0.005 461	0.003 868

图 6 为测试集中 4 种模型的预测结果,通过该图可知,本试验所提出的 IPSO-SVR 模型对于长江水

体的总磷浓度预测具有很好的效果。通过观察表 2 和图 6 可知,使用 IPSO 算法对 SVR 模型中的 3 种超参数进行优化,能够找到合适的超参数值,从而提高了模型的预测性能。同时将该模型同当前预测性能较好的 RF 模型、BPNN 模型和 IPSO-BPNN 模型进行预

测精度的对比,结果发现 IPSO-SVR 模型的  $R^2$  为 0.973 920, RMSE 为 0.003 012, MAE 为 0.002 105,其预测精度以及拟合效果明显优于其他模型,能够准确预测长江水体总磷浓度。因此选择 IPSO-SVR 模型作为水体中总磷含量的检测模型。

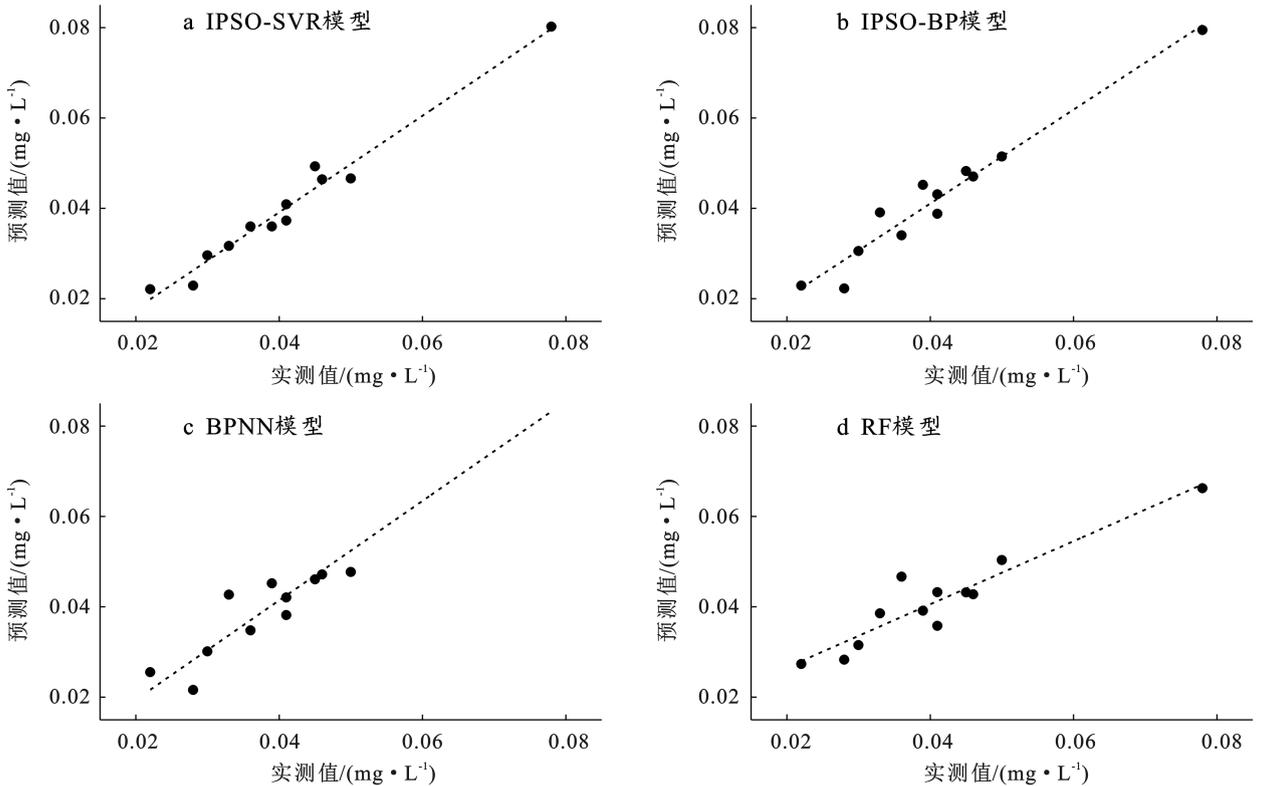


图 6 模型预测结果对比

Fig.6 Comparison of model prediction results

### 3 结论

本文提出了 IPSO-SVR 水质污染物预测模型,以 2021 年长江中下游武汉—安徽地区采集的水质样本为研究对象,分别进行数据预处理、降维、模型的训练和预测操作。经过对比分析,本文所提出模型的  $R^2$  为 0.973 920, RMSE 为 0.003 012, MAE 为 0.002 105,表现出更小的预测误差和更高的准确率,为水质中污染物浓度的预测提供了新的思路和方法。然而,水质的变化会受采集时间,区域和季节等多个因素的影响。因此,未来的研究将重点结合不同代表性河段,综合不同季节等条件进行影响结果因素的对比分析讨论,从而更好地了解水质的时空变化规律,以进一步提高模型的适用性,为水资源的管理和保护提供更全面科学的信息和依据。

### 参考文献 (References)

[1] 钟婷.4-氨基硫代酚调控的 Au@Ag 纳米颗粒的制备及其

对亚硝酸盐的检测应用[D].四川 成都:成都大学,2023.  
Zhong Ting. Preparation of Au@Ag nanoparticles regulated by 4-aminothiophenol and its application to the detection of nitrite [D]. Chengdu, Sichuan: Chengdu University, 2023.

[2] 薛同来,赵冬晖,韩菲.基于 GA 优化的 SVR 水质预测模型研究[J].环境工程,2020,38(3):123-127.

Xue Tonglai, Zhao Donghui, Han Fei. SVR water quality prediction model based on GA optimization [J]. Environmental Engineering, 2020,38(3):123-127.

[3] 李建文,孟连子,刘娜,等.基于 EEMD-SVR 的渔业养殖水质预测模型[J].天津理工大学学报,2018,34(5):14-19,24.

Li Jianwen, Meng Lianzi, Liu Na, et al. The prediction model of fishery water quality based on EEMD-SVR [J]. Journal of Tianjin University of Technology, 2018, 34(5):14-19,24.

[4] 康贝,马洁.基于 LLE-SVR 的水质 COD 紫外光谱检测方法研究[J].传感器世界,2018,24(9):11-15.

- Kang Bei, Ma Jie. Study of UV visible spectrum-based COD detection method for water quality monitoring based on LLE-SVR [J]. *Sensor World*, 2018,24(9):11-15.
- [5] 康成龙.基于 SERS 技术和 GWO-SVR 算法对水中污染物的检测研究[D].河北 秦皇岛:燕山大学,2021.
- Kang Chenglong. Study on determination of pollutants in water based on SERS Technology and GWO-SVR algorithm [D]. Qinhuangdao, Hebei: Yanshan University, 2021.
- [6] 姜吉光,石磊,苏成志,等.基于 SPA-SVR 的紫外光谱水质污染物含量解耦预测方法[J].*激光与光电子学进展*, 2023,60(7):396-403.
- Jiang Jiguang, Shi Lei, Su Chengzhi, et al. Decoupled prediction method for water pollutant concentration based on SPA-SVR using ultraviolet spectroscopy [J]. *Laser & Optoelectronics Progress*, 2023,60(7):396-403.
- [7] 刘豫,王顺钰,赵全顺.基于改进 PSO-SVR 的气温预测模型研究[J].*青海科技*, 2022,29(5):148-153.
- Liu Yu, Wang Shunyu, Zhao Quanshun. Research on temperature prediction model based on improved PSO-SVR [J]. *Qinghai Science and Technology*, 2022, 29(5):148-153.
- [8] 刘合兵,王垒,王一飞,等.基于 PSO-SVR 模型的河南省小麦产量预测方法[J].*江苏农业科学*, 2023,51(8):157-163.
- Liu Hebing, Wang Lei, Wang Yifei, et al. Wheat yield prediction method for Henan Province based on PSO-SVR model [J]. *Jiangsu Agricultural Sciences*, 2023,51(8):157-163.
- [9] 朱成杰,汪正权.基于 SPA-GA-SVR 模型的土壤水分及温度预测[J].*中国农村水利水电*, 2024(1):30-36.
- Zhu Chengjie, Wang Zhengquan. Soil moisture content and temperature prediction based on SPA-GA-SVR model [J]. *China Rural Water and Hydropower*, 2024(1):30-36.
- [10] 任远芳,牛坤,丁静等.基于改进 PSO 算法优化 SVR 的信息安全风险评估研究[J].*贵州大学学报(自然科学版)*, 2024,41(1):103-109.
- Ren Yuanfang, Niu Kun, Ding Jing, et al. Research on risk assessment of information security based on SVR optimized by improved PSO algorithm [J]. *Journal of Guizhou University (Natural Sciences)*, 2024, 41(1): 103-109.
- [11] 杨俊祺,范晓军,赵跃华等.基于 PSO-BP 神经网络的山西省碳排放预测[J].*环境工程技术学报*, 2023,13(6):2016-2024.
- Yang Junqi, Fan Xiaojun, Zhao Yuehua, et al. Prediction of carbon emissions in Shanxi Province based on PSO-BP neural network [J]. *Journal of Environmental Engineering Technology*, 2023,13(6):2016-2024.
- [12] 王寿元,李积元,郎永存,等.基于 PSO 优化 SVM 数控机床主轴系统故障诊断的研究[J].*组合机床与自动化加工技术*, 2023(9):151-155,159.
- Wang Shouyuan, Li Jiuyan, Lang Yongcun, et al. Research on fault diagnosis of CNC machine tool spindle system based on SVM Optimized by PSO [J]. *Modular Machine Tool & Automatic Manufacturing Technique*, 2023,9:151-155,159.
- [13] 石延辉,杨洋,廖毅,等.基于改进粒子群算法优化 SVM 的变压器故障诊断[J].*武汉大学学报(工学版)*, 2023,56(10):1238-1244.
- Shi Yanhui, Yang Yang, Liao Yi, et al. Transformer fault diagnosis based on improved particle swarm optimization SVM [J]. *Engineering Journal of Wuhan University*, 2023,56(10):1238-1244.
- [14] 龚泽威一,饶桐,王钢,等.基于改进粒子群优化 XG-Boost 的变压器故障诊断方法[J].*高压电器*, 2023,59(8):61-69.
- Gong Zeweiyi, Rao Tong, Wang Gang, et al. Fault diagnosis method of transformer based on improved particle swarm optimization XGBoost [J]. *High Voltage Apparatus*, 2023,59(8):61-69.
- [15] 刘洋,张鸿,徐娟,等.改进 PSO 的 SVM 回归模型及在气温预测中的应用[J].*计算机系统应用*, 2023,32(9):203-210.
- Liu Yang, Zhang Hong, Xu Juan, et al. Improved PSO SVM regression model and its application in temperature prediction [J]. *Computer Systems & Applications*, 2023,32(9):203-210.
- [16] 赵海越,曹岩,黄亮,等.基于遗传算法的 ZL205A 铝合金切削参数多目标优化[J].*兵器材料科学与工程*, 2023,46(5):58-65.
- Zhao Haiyue, Cao Yan, Huang Liang, et al. Multi-objective optimisation of cutting parameters of ZL205 A aluminium alloy based on genetic algorithm [J]. *Ordnance Material Science and Engineering*, 2023, 46(5):58-65.
- [17] 李艳波,李若尘,史博,等.基于改进模拟退火遗传算法的高速公路服务区自洽能源系统高效能优化[J].*西安交通大学学报*, 2024,58(1):197-207,216.
- Li Yanbo, Li Ruochen, Shi Bo, et al. Optimization of high energy efficiency for self-consistent energy system in highway service area via simulated annealing algorithm-genetic algorithm [J]. *Journal of Xi'an Jiaotong University*, 2024,58(1):197-207.
- [18] 王殿熙,杨波,成保忠,等.基于精英回溯-模拟退火混合算法的参数耦合变更路径优选[J].*机械工程学报*, 1-18.(2023-09-22)[2024-05-12].<http://kns.cnki.net/kcms/detail/11.2187.TH.20230922.1511.050.html>.

- Wang Dianxi, Yang Bo, Cheng Baozhong, et al. Optimization of parameter coupling change propagation path based on elite backtracking-simulated annealing hybrid algorithm[J]. Journal of Mechanical Engineering, 1-18. (2023-09-22)[2024-05-12].<http://kns.cnki.net/kcms/detail/11.2187.TH.20230922.1511.050.html>.
- [19] 孙嘉豪, 张伟, 施鉴芬, 等. 光谱数据预处理策略选择及应用[J]. 计量学报, 2023, 44(8): 1284-1292.
- Sun Jiahao, Zhang Wei, Shi Jianqin, et al. Selection and application of spectral data preprocessing strategy [J]. Acta Metrologica Sinica, 2023, 44(8): 1284-1292.
- [20] 阚相成, 李耀翔, 王立海, 等. 基于光谱预处理的低温水曲柳原木含水率检测[J]. 中南林业科技大学学报, 2022, 42(11): 154-163.
- Kan Xiangcheng, Li Yaoxiang, Wang Lihai, et al. Moisture content detection of Fraxinus mandshurica logs at low temperatures based on different spectrum pretreatments [J]. Journal of Central South University of Forestry & Technology, 2022, 42(11): 154-163.
- [21] 陈晓红, 王辉, 李喜华. 基于 KPCA-CNN-DBiGRU 模型的短期负荷预测方法[J]. 管理工程学报, 2024(2): 221-231.
- Chen Xiaohong, Wang Hui, Li Xihua. A short-term load forecasting method based on KPCA-CNN-DBiGRU model [J]. Journal of Industrial Engineering and Engineering Management, 2024(2): 221-231.
- [22] 葛盼猛, 陈波, 刘庭赫, 等. 基于 KPCA-ICFSFDP-MOGP 的拱坝多测点变形数据预处理及预测方法[J]. 水利水
- 电科技进展, 2023, 43(4): 92-97.
- Ge Panmeng, Chen Bo, Liu Tinghe, et al. Preprocessing and prediction method for multi-point deformation data of arch dams based on KPCA-ICFSFDP-MOGP [J]. Advances in Science and Technology of Water Resources, 2023, 43(4): 92-97.
- [23] 谢金燕, 刘丽星, 杨欣, 等. 改进粒子群优化算法的果园割草机作业路径规划[J]. 中国农业大学学报, 2023, 28(11): 182-191.
- Xie Jinyan, Liu Lixing, Yang Xin, et al. Orchard lawn mower operation path planning based on improved particle swarm optimization algorithm [J]. Journal of China Agricultural University, 2023, 28(11): 182-191.
- [24] 刘孝保, 严清秀, 易斌, 等. 基于集成学习和改进粒子群优化算法的流程制造工艺参数优化[J]. 中国机械工程, 2023, 34(23): 2842-2853.
- Liu Xiaobao, Yan Qingxiu, Yi Bin, et al. Optimization of process parameters in process manufacturing based on ensemble learning and improved particle swarm optimization algorithm [J]. China Mechanical Engineering, 2023, 34(23): 2842-2853.
- [25] 邵宁宁, 王英. 基于 Adam 优化的改进 PSO-RBF 牵引变压器故障诊断研究[J]. 电气工程学报, 2023, 18(4): 209-216.
- Shao Ningning, Wang Ying. Research on PSO-RBF traction transformer fault diagnosis based on Adam optimization [J]. Journal of Electrical Engineering, 2023, 18(4): 209-216.

(上接第 195 页)

- [18] 杜永强, 王悦, 李晶洁等. 基于主成分分析的经济绿色发展评价研究[J]. 科技和产业, 2018, 18(10): 50-59.
- Du Yongqiang, Wang Yue, Li Jingjie, et al. Research on economic green development evaluation based on principal component analysis [J]. Science Technology and Industry, 2018, 18(10): 50-59.
- [19] 贾科, 杨哲, 魏超, 等. 基于斯皮尔曼等级相关系数的新能源送出线路纵联保护[J]. 电力系统自动化, 2020, 44(15): 103-111.
- Jia Ke, Yang Zhe, Wei Chao, et al. Pilot protection based on spearman rank correlation coefficient for transmission line connected to renewable energy source [J]. Automation of Electric Power Systems, 2020, 44(15): 103-111.
- [20] 李久龙, 李玲, 刘瑞敏, 等. 一元线性回归分析方法在压力传感器数据拟合中的应用[J]. 计测技术, 2022, 42(2): 40-49.
- Li Jiulong, Li Ling, Liu Ruimin, et al. Application of univariate linear regression analysis in data fitting of pressure sensor [J]. Metrology & Measurement Technology, 2022, 42(2): 40-49.
- [21] 荀梦君, 李进岚, 黄爱菊, 等. ARIMA 模型和 Holt-Winters 指数平滑法在贵州省肺结核发病预测中的应用[J]. 中国预防医学杂志, 2023, 24(7): 678-682.
- Xun Mengjun, Li Jinlan, Huang Aiju, et al. Application of ARIMA and Holt-Winters exponential smoothing in the prediction of tuberculosis in Guizhou Province [J]. Chinese Preventive Medicine, 2023, 24(7): 678-682.