

改进 SVR 的内陆水体 COD 高光谱遥感反演

盛 辉¹, 池海旭¹, 许明明^{1*}, 刘善伟¹, 万剑华¹, 王锦锦²

1. 中国石油大学(华东), 海洋与空间信息学院, 山东 青岛 266580
2. 珠海欧比特宇航科技股份有限公司, 广东 珠海 519080

摘 要 高光谱数据可以捕获内陆水体中不同浓度的化学需氧量(COD)引起的光谱变化, 因此研究光谱反射率与 COD 浓度之间的关系对于 COD 的遥感估算至关重要。支持向量回归模型(SVR)具有适合小样本、泛化能力好的特点, 基于 SVR 模型能够更加准确获得 COD 浓度和光谱数据之间的关系, 但仍然存在参数选取困难和易陷入局部极值的问题。为了解决这个问题, 将模拟退火-粒子群算法(SA-PSO)引入到支持向量回归机的参数优化过程中, 提出了一种改进 SVR(SA-PSO-SVR)的内陆水体 COD 高光谱遥感反演方法。以淮河流域为研究区域, 通过野外测量获得了 COD 浓度和水表面光谱反射率。首先根据光谱反射率对 COD 的响应来确定敏感因子, 把 SA-PSO 算法引入 SVR 的参数优化过程中建立了 COD 浓度与敏感因子之间的反演模型。最后利用珠海一号高光谱数据验证模型的准确性, 进而获得了 COD 浓度的分布情况。通过光谱分析, 可知该区域实测的水面光谱具有典型的二类水体特征, 光谱曲线形状呈现明显的双峰特征, 当浓度增加时, 反射峰具有向短波长方向移动而反射谷向长波长方向移动的趋势。通过计算 Pearson 相关系数分析 COD 浓度和光谱之间的相关性, 结果表明最佳的反演因子为 518 nm/940.4 nm, 623.6 nm/636.8 nm, 729.2 nm/890.9 nm 和 752.3 nm/857.9 nm 的四个波段比值组合; 经过 SA-PSO-SVR 方法建立的 COD 估计模型的平均相对误差(MRE)和均方根误差(RMSE)分别为 1.62% 和 2.99 mg · L⁻¹ (R² = 0.86), 反演结果优于其他模型(SVR、BP 神经网络和线性回归模型)。将实测水面光谱建立的最优模型应用于高光谱卫星影像上, RMSE 和 MRE 分别为 4.47 mg · L⁻¹ 和 11.87%。获得的淮河—峡山水库区域的 COD 反演结果显示: COD 的整体浓度介于 17~42 mg · L⁻¹ 之间, 韩信坝、峡山水库的东北部、渠河注入淮河的交汇处等区域的 COD 浓度高于其他水域。证实了 SA-PSO-SVR 是一种有效的 COD 高光谱反演方法, 可供淮河流域水资源管理提供参考。

关键词 化学需氧量; 支持向量回归; 模拟退火-粒子群; 高光谱数据

中图分类号: P237 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)11-3565-07

引 言

河流及水库作为内陆水资源的重要组成部分, 对涵养水源、保护生态具有重要的作用。但近几十年来, 自然因素和人为因素的双重压力使水环境在不同程度上受到了污染, 水体富营养化的趋势愈加严峻, 对水质进行动态监测和精细化的管理迫在眉睫。

传统的水质监测工作主要是获得监测区域内某些点的水质浓度信息, 这些有限点的信息并不能反映出整个区域的情况, 而遥感水质监测具有实时高效、监测范围广和适合于长

期动态监测等特点^[1]。化学需氧量(chemical oxygen demand, COD)可以用来表征水中有机物的含量, 是遥感水质监测的重要指标。COD 遥感监测主要是研究水体反射的光谱特征和 COD 浓度之间的关系, 从而建立反演算法。水质参数反演的方法主要有两类: 半经验方法和半分析方法^[2]。一些学者利用半经验的分析方法研究 COD 和遥感参数之间的关系, 通过构建的线性及生物光学等模型反演得到了内陆湖泊 COD 含量的分布情况, 并进一步分析得到了 COD 的来源及影响^[3]。但半经验模型的拟合精度范围大多在 0.6~0.8 之间, 因此模型的准确性还有待提高。国内外的部分学者研究评估机器学习进行水质遥感的可行性^[4-5], 表明在充足样本

收稿日期: 2020-10-22, 修订日期: 2021-02-16

基金项目: 国家自然科学基金项目(U1906217), “珠海一号”遥感星座建设、运营与应用项目(2017BT01G115)资助

作者简介: 盛 辉, 1972 年生, 中国石油大学(华东)副教授 e-mail: sheng@upc.edu.cn

* 通讯作者 e-mail: xumingming@upc.edu.cn

的情况下, 神经网络能够对水质参数进行有效的反演分析。由于某些机器学习算法参数选择的不确定性以及 COD 光谱特征的微弱性, 通常需要多次重复实验来获得最优参数从而提高模型的准确性。另外, 研究人员在野外实验中需要花费大量的时间和精力才能够收集到足够的水质参数数据。因此, 构建结构简单、适合小样本的水质参数反演模型尤为重要。研究表明, 经过优化的支持向量回归机 (support vector regression, SVR) 具有结构简单、全局最优的特点, 已经广泛应用到叶绿素、悬浮物、溶解氧和水体富营养化的水质研究中^[6-7], 但是针对 COD 的反演研究较少。因此, 本工作采用改进的 SVR 方法进行 COD 反演。

用于 COD 反演的数据源主要集中于 Landsat, Modis 和 Sentinel-2 等多光谱数据^[8]。高光谱数据可以捕获由不同水质浓度引起的光谱变化, 并在内陆水质监测中显示出巨大的潜力^[9-10]。但由于水体光谱影响机理的复杂性和其他水质参数的影响^[11], 狭小区域的 COD 反演和变化监测需要兼有高空间分辨率和高光谱分辨率的数据提供支持。实测的水表面光谱反射率具有受大气影响较小的特征^[12], 将其与高光谱图像结合可以更准确地估算 COD 浓度^[13]。林建远等通过航空高光谱数据和实测的水表面反射率反演得到了杭嘉湖平原河道的 COD 含量, 但是航空高光谱数据的获取难度大、成本高。珠海一号高光谱卫星 (Orbita hyper spectral, OHS) 数据具有高空间分辨率、高时间分辨率和易于获取的特点。因此, 采用现场实测光谱反射率数据结合珠海一号高光谱卫星数据用于 COD 浓度的反演。

针对 COD 光谱特征的微弱性以及 SVR 参数选取困难、易陷入局部极值的问题, 基于高光谱数据, 提出了一种模拟退火-粒子群算法 (simulated annealing-particle swarm algorithm, SA-PSO) 优化的 SVR 方法进行 COD 浓度反演。首先, 基于实测的光谱数据建立淮河—峡山水库区域的 COD 高光谱反演模型, 减少大气校正对反演精度的影响。然后应用于 OHS 高光谱数据对该区域的 COD 含量进行遥感估算, 以期对淮河流域的水质监测与综合管理和 OHS 卫星数据在内陆水质的反演应用提供基础数据与科学参考。

1 研究区域与数据

淮河及峡山水库位于山东半岛, 是胶东地区的战略水源地之一。淮河是流经峡山水库的主要河流之一, 发源于沂水县和莒县, 最终注入渤海, 河长 200 多公里, 流域面积近万平方公里; 峡山水库的注入河流有淮河、渠河和浞河等河流, 入库量为 $8 \times 10^8 \text{ m}^3$ 左右。峡山水库周围及淮河上下游区域分布着众多村庄和工厂, 该区域为周边地区的居民饮水, 农业灌溉和工业发展做出了重要贡献。同时, 该地区的百姓生活与经济发展也对水质造成了一定的影响, 采用遥感技术对该地区的水质进行监测显得尤为必要。

1.1 实测数据

2019 年 10 月 26 日—10 月 28 日, 分别在淮河流域的古县大桥、峡山水库、辉村和金口大桥附近水域进行水体取样和光谱采集, 采集光谱数据 12 组, 28 日有效测得 COD 浓度

数据 22 组。将采水器采集到的表面水样装在琥珀瓶中, 在瓶体贴标签记录好采样地点、时间和当时的天气情况并拍照保存, 最终在实验室化验得到各采样点 COD 的浓度值。

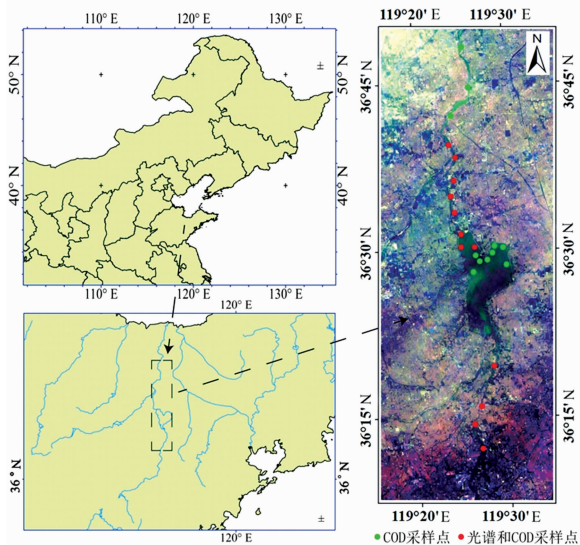


图 1 淮河及峡山水库现场实验点位图

Fig. 1 The maps of Weihe and Xiashan reservoirs

光谱测量采用 YW-TRIOS-AWRMMS 水面移动测量系统 (光谱范围 320~943.7 nm, 共 190 个波段), 该仪器内置 GPS 定位系统, 携带方便可用于各种情况的光谱测量。测量过程中, 保持伸缩杆方向和太阳入射平面 135° 的夹角可使 3 根传感器分别有效测量太阳辐照度、水面辐照度和天空辐照度三种参数, 并可在显示屏上查看光谱曲线。在每个采样点测量 10 组光谱以最小化不确定性, 通过光谱数据的处理程序及 Mobley(1999) 的方法对光谱数据进行处理, 即可得到离水辐亮度和遥感反射率等光谱数据信息参数。

1.2 遥感数据

OHS 卫星于 2018 年 4 月 26 日成功发射升空, 弥补了我国在高光谱数据上的不足, 开启了商业航天遥感的新时代。OHS 卫星空间分辨率为 10 m, 4 颗高光谱卫星的重访周期为 2.5 d, 单次成像范围为 $150 \text{ km} \times 320 \text{ km}$, 光谱分辨率为 3~8 nm, 光谱范围为 466~940 nm, 共 32 个波段。由于其具有幅宽大、高空间分辨率、高光谱分辨率和高时间分辨率的特点, 非常有利于反演光学特性复杂多变的内陆水体、小型水库和河流的水质参数等相关研究。OHS 数据是当前能够进行小区域水质参数反演的最佳数据源。

采用珠海欧比特宇航科技股份有限公司提供的 2 景淮河流域 2019 年 10 月 28 日的 OHS 高光谱数据。首先对影像数据进行辐射定标、大气校正、影像拼接与陆地掩膜等处理, 最终得到仅有淮河流域区域在内的影像产品。大气校正是为了消除大气分子和气溶胶的影响, 是准确获得水质遥感信息的前提。FLAASH 模型用于遥感影像的大气纠正具有很好的应用效果; 故使用 FLAASH 模型进行大气校正, 然后根据采样点的坐标位置, 在影像上提取各点对应的反射率信息。

2 改进 SVR 的内陆水体 COD 高光谱遥感反演模型

建立 COD 的反演模型, 首先要对光谱数据进行归一化和相关性分析等预处理, 然后基于对 COD 敏感的波段结合 SA-PSO 算法建立 SVR 反演模型。

2.1 光谱数据与 COD 的相关性分析

为了削弱野外环境给光谱测量造成的影响, 同时也使光谱数据和 OHS 影像光谱范围相对应, 采用归一化的方法对 466~940 nm 的光谱数据进行预处理。归一化的公式如式 (1)

$$R_N(\lambda_i) = \frac{R(\lambda_i)}{\frac{1}{n} \sum_{i=466}^{940} R(\lambda_i)} \quad (1)$$

式(1)中: $R_N(\lambda_i)$ 为水体遥感反射率的归一化结果, $R(\lambda_i)$ 为原始的遥感反射率, λ_i 为 i 处波长, n 为 466~940 nm 处的波段数。

Pearson 相关性分析是从统计学的角度研究两个或多个随机变量间关联度强弱的方法。相关系数的大小可以描述变量间的密切程度, 两个变量的相关系数表达式如式(2)

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

式(2)中: y_i 是各个采样点水质参数的浓度, x_i 是各个采样点的地表反射率。

2.2 基于 SA-PSO 算法优化的 SVR 模型

SVR 是基于机器学习的一种算法, 通过核函数将非线性数据映射到高维空间构造决策函数进行线性回归, 在解决小样本、非线性和高维模式识别问题方面具有其独特的优势, 经常用于小样本水质参数的反演^[14]。SVR 模型的数学关系为式(3)

$$f(x_i) = \omega^T \varphi(x_i) + B \quad (3)$$

式(3)中, x_i 为样本数据, ω 为待辨识的权重, $\varphi(x_i)$ 为非线性映射, $f(x_i)$ 为特征空间中的线性函数, B 为常数项。

由于 SVR 模型存在参数选取难的问题, 如何确定最优参数直接影响到 SVR 模型的学习和泛化能力。根据鸟群觅食行为提出的粒子群算法 PSO 常用于对 SVR 模型参数进行优化选取, 将只有位置和速度两个属性的粒子模拟为鸟, 每个粒子在空间中的极值 P_{best} 作为潜在的最优解 G_{best} ; 所有粒子根据式(4)不断调整各自的位置和速度, 直至获得最优解。同时, 在粒子群更新过程中引入模拟退火算法 SA 可提高 PSO 算法的全局寻优能力, 即利用其在寻优过程中的突跳能力对粒子群算法进行改进。SA 算法本质是模拟高温物体退火过程中寻找全局最优解的过程。初始温度 T 的确定采用适应度和接受概率的方法, 由式(5)决定

$$\begin{aligned} V_{i+1} &= \omega \times V_i + c_1 \times \text{rand} \times (P_{best} - X_i) + \\ &\quad c_2 \times \text{rand} \times (G_{best} - X_i) \\ X_{i+1} &= X_i + V_{i+1} \end{aligned} \quad (4)$$

$$T = (f_{max} - f_{min}) / \ln p = -|\Delta f| / \ln p \quad (5)$$

式中: f_{max} 和 f_{min} 和 Δf 为初始粒子群最大、最小适应度值及其差值, p 为初始接受概率, $p=0.8$ 。

建立基于 SA-PSO 算法优化的 SVR 模型, 关键是利用 SA-PSO 算法优化支持向量回归模型中判别函数的惩罚因子 C 和核函数中核的宽度 g 。建立 SA-PSO 优化 SVR 模型的主要步骤如图 2 所示。

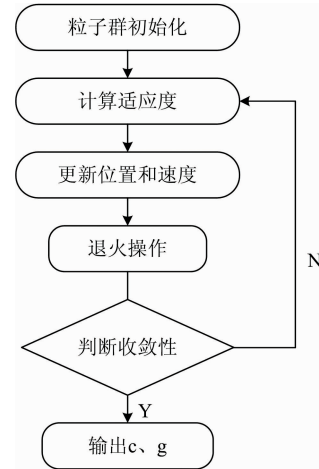


图 2 SA-PSO 优化 SVR 流程

Fig. 2 Optimization SVR model using SA-PSO

(1) 粒子群的初始化。对粒子速度, 位置, 惩罚因子 C 和 RBF 核函数宽度等参数进行初始化。设置初始退火温度 T , 并将最大迭代次数设置为 200。

(2) 计算群体中每个粒子的适应度。如果新的适应性值好于原始值, 则接受新位置。否则, 将保留旧位置。

(3) 根据式(5)更新粒子的位置和速度。

(4) 收敛性的判断。如果满足终止条件, 则算法停止。否则, 执行退火操作, 转到(2)。

(5) 输出最优参数 C 和 g 。

3 实验与结果讨论

进行 COD 的高光谱反演, 首先要对光谱进行归一化和相关性分析来确定敏感因子, 然后根据敏感因子和 COD 浓度建立反演模型。最后通过使用与采样点对应的 OHS 反射率数据来分析模型的准确性, 并将 SA-PSO-SVR 应用于淮河流域的 OHS 数据进行 COD 的空间分析。

3.1 光谱数据与 COD 的相关性分析

实测的原始光谱曲线如图 3(a) 所示, 根据等式(1)进行归一化的结果如图 3(b) 所示。与原始光谱曲线相比, 归一化光谱曲线的形状发生了变化, 反射谷和反射峰更加明显。从 560, 680 和 710 nm 的反射率值可以发现, 当浓度增加时, 反射峰具有向短波长方向移动而反射谷向长波长方向移动的趋势。

Pearson 相关性分析的输入反射率形式包括单波段和波段比值等形式。研究表明, 波段比值组合可部分消除水表面

光滑度和微波等其他环境因素的影响^[15],能在一定程度上有效提高水质参数反演的精度。对各个采样点的 COD 浓度分别与每组光谱数据的波段比值组合做 Pearson 相关性分析,得到 COD 浓度与波段比值组合的相关性分析结果如图 4 所示,最佳反演因子是 518 nm/940.4 nm, 663.6 nm/636.8 nm, 729.2 nm/890.9 nm 和 752.3 nm/857.9 nm 四种波段比值组合。

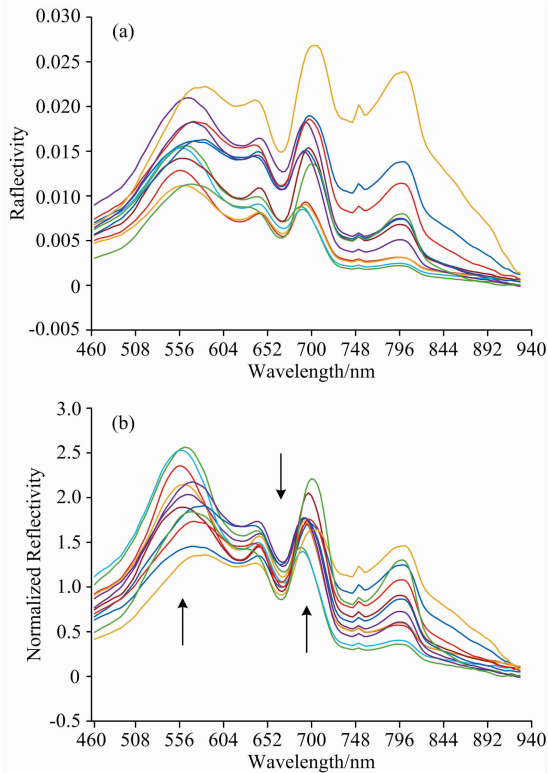


图 3 原始光谱 (a) 和归一化后的光谱 (b)

Fig. 3 Measured reflectance spectra (a); and normalized spectra (b)

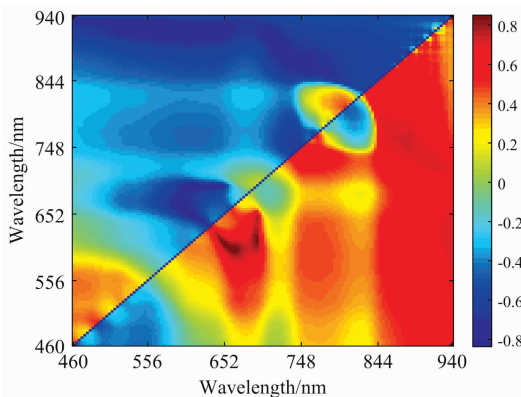


图 4 466~940 nm 的实测光谱比值与 COD 浓度的 Pearson 相关性分析结果

Fig. 4 Pearson's correlation coefficient (r) between band ratios $R_{rs}(\lambda_2)/R_{rs}(\lambda_1)$ and COD in the spectral range of 466~940 nm

3.2 COD 反演

根据选取的最佳反演因子,用 12 个采样点的光谱数据来训练 SA-PSO-SVR 模型。根据相关研究及反复实验,设置种群规模为 50,进化次数为 200,选取参数 c_1 和 c_2 均为 0.5,SVR 模型的 C 和 g 的取值范围为 $[0.1, 1000]$ 和 $[0.1, 10]$ 。对训练数据进行归一化处理,经过 SA-PSO 优化获得的惩罚因子 C 与核参数 g 分别为 151.09 和 0.36。

利用 28 日测得的 22 个采样点数据作为检查点,在 OHS 影像上提取与采样点对应的反射率,导入 SA-PSO-SVR 模型得到各个检查点的预测值。图 5 为 SA-PSO-SVR 精度分析结果。红色点为利用光谱数据建立模型得到的实测值和预测值的分布情况,虽然某些采样点的预测值与实测值具有一定差异,但依然可以看出 SA-PSO-SVR 模型的反演结果与实测值呈现较好的相关一致性,其模型决定系数为 0.86。绿色点为检查点数据的预测值和实测值的分布情况,平均相对误差(MRE)和均方误差(RMSE)分别为 9.04% 和 $3.64 \text{ mg} \cdot \text{L}^{-1}$,说明该模型可以实现对 COD 浓度的有效反演。

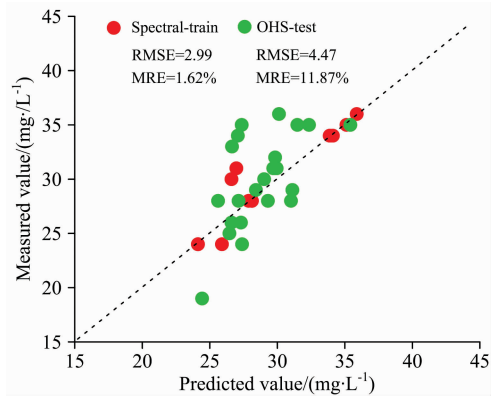
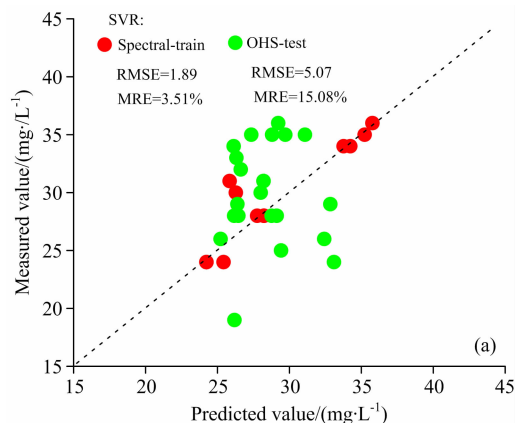


图 5 SA-PSO-SVR 精度分析

Fig. 5 SA-PSO-SVR accuracy evaluation

3.3 SA-PSO-SVR 精度分析

从图 5 可以看出, OHS 数据部分反演结果与实测值差距略大,推测原因为数据质量参差不齐。由于 TRIOS 光谱测量仪和 OHS 卫星是两种截然不同的传感器,传感器的信噪比等参数并不一致,因此两种仪器获得的数据质量并不一致。研究表明, COD 对水体的光学特性响应很微弱,为了减



(a)

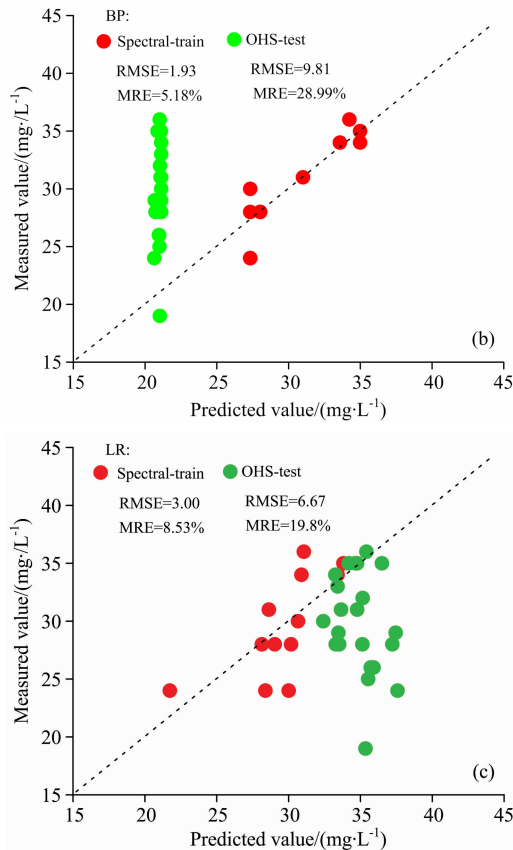


图 6 SVR, BP 神经网络, LR 模型预测评估
Fig. 6 Evaluations of predictions using SVR, BP neural network and LR method

少河道两边其他非水体像元的影响, 采样点的像元位置尽可能的选在距离监测点近的河道中央, 但由于河道中央流速较快, 可能导致河中央同采样点的 COD 浓度存在一定差异, 进而造成部分反演结果与实测值的差异较大。

另外, 基于光谱数据分别建立 SVR 和 BP 神经网络和线性回归(LR)模型, 将 SVR 模型的惩罚因子 C 与核参数 g 设置为 109.85 和 0.000 1; 根据相关研究及反复实验, 设置 BP 神经网络模型的隐含层节点为 12, 最大训练次数为 2 000, 期望误差为 0.000 1, 学习速率为 0.01。SVR 和 BP 神经网络和 LR 模型的决定系数分别为 0.712, 0.598 和 0.494。从图 6(a) 可以看出, SVR 模型的整体反演效果较好, 但是 OHS 数据的部分反演结果低于实际测量值。从图 6(b) 可以看出 BP 神经网络模型的 RMSE 和 MRE 分别为 $1.93 \text{ mg} \cdot \text{L}^{-1}$ 和 5.18%, 与其他两种模型的结果相差不大; 但是从 OHS 数据的反演情况看, BP 神经网络模型的结果陷入了局部最优, 由此也证明 BP 神经网络模型确实存在参数选取困难的问题。另外, 从图 6(c) 可以看出, 光谱数据的 RMSE 和 MRE 分别为 $3.00 \text{ mg} \cdot \text{L}^{-1}$ 和 8.53%, 与其他模型结果相近, 但 OHS 影像的反演结果要普遍高于实测值。综合来看, SA-PSO-SVR 模型的反演效果优于其他模型。

3.4 COD 浓度空间分布分析

将 SA-PSO-SVR 模型应用于 OHS 数据, 得到的 COD 浓度空戒分布如图 7 所示。可以看出, COD 的浓度呈现部分区域浓度高的特点, 在韩信坝、峡山水库的东北部区域、渠河与潍河的交叉口及辉村与金口大桥之间的浓度明显偏高。

结合潍河流域地理情况并查阅相关资料发现, 潍河流域附近分布着众多工厂, 大部分企业将工业污水排入河中, 这些河流(潍河、渠河等)携带污染物最终汇入峡山水库, 是导致峡山水库东部沿岸区域的 COD 浓度明显高于其他区域的原因之一。另外, 库区及潍河等其他河流两岸分布着众多村庄, 百姓的农业生产活动依赖于该区域的水资源, 农业生产带来的化肥、农药等污染物也是造成该局部区域 COD 浓度过高的原因之一。同时, 研究表明农作物腐烂死亡降解产生的有机物也会造成局部区域的 COD 浓度增高。另外, 在韩信坝和山阳村附近区域有多处拦河大坝, 部分区域河流径流量减少, 遥感影像获取的 COD 反射信息不充分, 造成该区域的 COD 反演结果偏高。

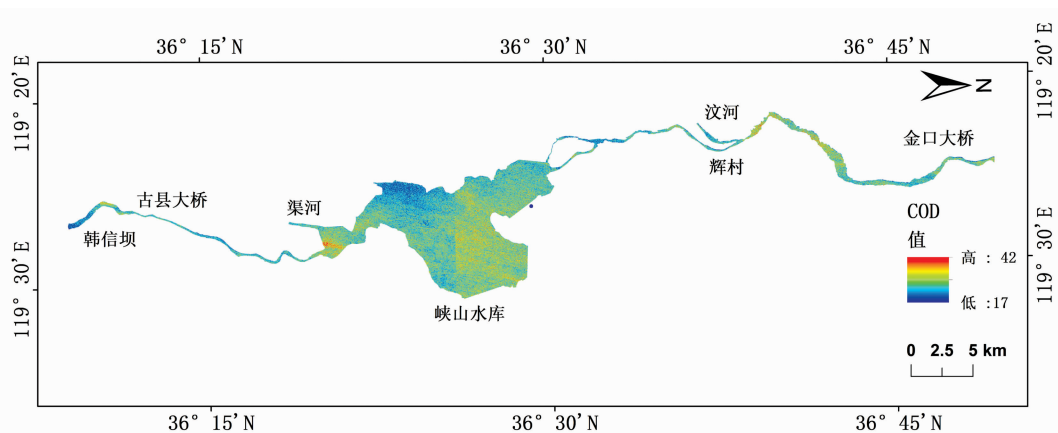


图 7 COD 浓度分布情况

Fig. 7 Spatial distribution of COD concentration obtained from OHS data

4 结 论

针对 SVR 模型参数选取困难和易陷入局部极值的情况, 本文基于水面以上的光谱数据建立了 SA-PSO-SVR 模型, 通过珠海一号卫星数据反演得到了淮河—峡山水库流域 COD 的分布情况, 得到以下结论:

(1) 通过对淮河—峡山水库区域的实测光谱进行分析, 结果表明该区域实测的水面光谱具有典型内陆湖泊水域复杂浑浊水体的光谱特征, 560 和 710 nm 附近的光谱曲线形状呈现明显的双峰特征, 反射率幅值较大。当浓度增加时, 反射峰具有向短波长方向移动而反射谷向长波长方向移动的趋势。

(2) 引入 SA-PSO 算法对 SVR 模型的参数进行优化选取, 解决了局部最优的问题, 其反演效果明显也优于其他模型。将基于实测光谱建立的 SA-PSO-SVR 模型应用在珠海

一号卫星数据上可以进行 COD 遥感估算, 说明了基于实测光谱和高光谱卫星影像的水质参数反演方法具有良好的应用前景和推广价值。同时, 反演得到的 COD 浓度分布情况可以为珠海一号卫星数据的内陆水质参数反演和淮河流域的综合管理提供科学依据。

珠海一号数据的验证结果说明了 COD 反演的有效性, 但由于星载传感器和地面光谱仪信噪比等仪器参数的不同, 并受到其他复杂的大气和光学水体组分的影响, 该方法还需要有足够的论据对此进行进一步的详细分析。此外, 虽然基于实测光谱建立的 SA-PSO-SVR 模型在淮河流域表现出了相对较好的性能, 但在其他复杂的水环境的应用性能还有待研究。因此, 接下来的工作将多次收集淮河流域的光谱、COD 浓度等其他水体成分的信息, 并根据长期观测资料进一步探讨 COD 的时空变化; 同时, 在包含多个内陆水域的综合数据集的基础上, 对 SA-PSO-SVR 方法进行更全面的评估。

References

- [1] Gholizadeh M H, Melesse A M, Reddi L. *Sensors*, 2016, 16(8): 43.
- [2] Bernardo N, Carmo A, Park E. *Remote Sensing*, 2019, 11(9): 22.
- [3] Shen M, Duan H T, Cao Z G. *Remote Sensing of Environment*, 2020, 247: 17.
- [4] Agrawal R K, Bawane Narendra G. *Applied Soft Computing Journal*, 2015, 28: 217.
- [5] Hafeez S, Wong M S, Ho H C, et al. *Remote Sensing*, 2019, 11(6): 23.
- [6] Garcia-Nieto P J, Garcia-Gonzalo E, Sanchez L F, et al. *Journal of Computational and Applied Mathematics*, 2020, 366: 19.
- [7] Li N, Zhu X F, Pan Y Z, et al. *Journal of Remote Sensing*, 2018, 22(4): 559.
- [8] Kahru M, Kudela R M, Anderson C R, et al. *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(11): 2282.
- [9] Bioucas-Dias J M, Plaza A, Camps-Valls G, et al. *IEEE Geoscience and Remote Sensing Letters*, 2013, 1(2): 6.
- [10] Babankumar B, Rangoli S, Ritula T. *Spatial Information Research*, 2018, 26(2): 203.
- [11] CHEN Yao, HUANG Chang-ping, ZHANG Li-fu, et al(陈 瑶, 黄长平, 张立福, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2020, 40(3): 824.
- [12] Pyo J C, Duan H T, Ligaray M, et al. *Remote Sensing*, 2020, 12(7): 23.
- [13] Gursoy O, Atun R. *Polish Journal of Environmental Studies*, 2019, 28(4): 2139.
- [14] Ansari M, Akhoondzadeh M. *Advances in Space Research*, 2020, 65(5): 1490.
- [15] El-Rawy M, Fathi H, Abdalla F. *Environmental Geochemistry and Health*, 2019, 42(7): 2101.

Inland Water Chemical Oxygen Demand Estimation Based on Improved SVR for Hyperspectral Data

SHENG Hui¹, CHI Hai-xu¹, XU Ming-ming^{1*}, LIU Shan-wei¹, WAN Jian-hua¹, WANG Jin-jin²

1. College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China

2. Zhuhai Orbita Aerospace Science & Technology Co., Ltd., Zhuhai 519080, China

Abstract Hyperspectral data can capture the spectral changes caused by different concentrations of chemical oxygen demand (COD) in inland water bodies, and it is important to study the relationship between spectrum and COD concentrations for COD estimation. Support Vector Regression (SVR) model has the advantages of being suitable for small samples and good generalization ability, but it is difficult to select a parameter and prone to fall into the local extremum. In order to solve this problem, this study introduced Simulated Annealing-Particle Swarm Optimization (SA-PSO) into the parameter optimization process of SVR and proposed an improved SVR (SA-PSO-SVR) method to estimate the inland waters COD. This paper takes the Weihe River Basin as the research area, obtained the COD concentrations and spectral curves through field measurement. The sensitive band was determined by analyzing the response of spectral reflectance to COD at first in this paper, and the Simulated Annealing-Particle Swarm Optimization (SA-PSO) was introduced into the parameter optimization process of Support Vector Regression (SVR) to established an inversion model between the cod concentration and the sensitivity factor. The Orbita Hyper Spectral (OHS) hyperspectral data was used to verify the accuracy, and the distribution of COD concentration was obtained at last. Through spectral analysis, it can be seen that the measured above-surface spectra in this area demonstrated typical spectral signatures of second-class water, and the shape of the spectrum curve shows obvious double-peak characteristics. When the concentration increases, the reflection peak tends to move to the short wavelength direction and the reflection valley to the long wavelength direction. The Pearson's correlation coefficient was used to analyze COD concentration and the spectral, the result showed that the best inversion factors are four band combinations of 518 nm/940.4 nm, 623.6 nm/636.8 nm, 729.2 nm/890.9 nm and 752.3 nm/857.9 nm. The model established by the SA-PSO-SVR method is accurate compared with models established by SVR, Back Propagation neural network, and linear regression method. The Mean-Relative-Error (MRE) and Root-Mean-Square-Error (RMSE) of the COD estimation model established by the SA-PSO-SVR method are 1.62% and 2.99 mg · L⁻¹ ($R^2 = 0.86$), respectively. The optimal model established by the measured water surface spectra was applied to the hyperspectral satellite image. The RMSE and MRE are 4.47 mg · L⁻¹ and 11.87% respectively. The obtained COD inversion results of the Weihe-Xiashan reservoir area show: the overall concentration of COD is between 17 and 42 mg · L⁻¹, and COD concentration in the Hanxinba, the northeast region of XiaShan Reservoir, the confluence of the Qu River into the Wei River are higher than other waters. The experimental results show that SA-PSO-SVR is a feasible approach for the COD inversion of hyperspectral data, providing a reference for water resources management in the Weihe River Basin.

Keywords Chemical oxygen demand; Support vector regression; Simulated annealing-particle swarm optimization; Hyperspectral data

(Received Oct. 22, 2020; accepted Feb. 16, 2021)

* Corresponding author