文章编号：1672 - 3317（2020）09 - 0126 - 06

基于 GA-PLS 算法的河网水体化学需氧量高光谱反演

蔡建楠[1,2]，刘海龙[3]，姜 波[3]，何甜辉[1]，陈文杰[2]，冯志伟[2]，黎倬琳[2]，邢前国[3*]
（1.中山市环境监测站，广东 中山 528403; 2.中山市生态环境局，广东 中山 528403;
3.中国科学院 烟台海岸带研究所/中国科学院海岸带环境过程与生态修复重点实验室，山东 烟台 264003）

摘 要：【目的】建立河网水体化学需氧量（COD）高光谱反演模型，验证遗传-偏最小二乘（GA-PLS）算法对建模效果的改善作用。【方法】采集广东省中山市 146 个点位的水体高光谱数据和 COD 质量浓度实测数据，通过 GA-PLS 算法对高光谱反射率数据进行特征波段筛选后建立 COD 质量浓度反演模型，并比较输入变量为不同特征波段组合时模型反演效果差异。【结果】基于 GA-PLS 算法的 COD 质量浓度高光谱模型反演效果优于全谱段 PLS 模型，验证集 *RMSEP* 最小为 4.887 mg/L，较全谱段 PLS 模型降低 11.4%；以筛选得到的 74 个波段（占全波段数的 2.9%）作为输入变量时，模型仍可保持良好的稳定性和反演精度；GA-PLS 算法筛选得出的部分特征波段与水体中藻类、悬浮颗粒物的吸收特征波段一致，筛选结果具有合理性和指示意义。【结论】通过 GA-PLS 算法可对高光谱数据进行特征波段筛选，实现数据降维优化，进一步简化模型；在样本 COD 质量浓度主要分布范围内，GA-PLS 算法模型有良好的反演精度和水质类别分类准确性。该方法在河流 COD 快速监测中具有良好的应用前景。

关 键 词：高光谱；遗传算法；偏最小二乘法；化学需氧量；河网水体
中图分类号：X832 **文献标志码：**A doi：10.13522/j.cnki.ggps.2020063

蔡建楠，刘海龙，姜波，等. 基于 GA-PLS 算法的河网水体化学需氧量高光谱反演[J]. **灌溉排水学报**, 2020, 39(9): 126-131.

CAI Jiannan, LIU Hailong, JIANG Bo, et al. Using Hyperspectral Imagery and GA-PLS Algorithm to Estimate Chemical Oxygen Demand Concentration of Water in River Network[J]. Journal of Irrigation and Drainage, 2020, 39(9): 126-131.

0 引 言

【研究意义】水质监测是河流水污染防治工作的重要组成部分。传统的水质监测方法存在成本高昂、采样及分析过程需时长、时空代表性差等缺点，难以满足当前形势下水环境管理的需要[1-2]。近年来，高光谱水质遥感以其检测速度快、无二次污染、精确度高、可与航空技术结合、可实现大面积覆盖等优点，在水质参数定量化监测中具有较大的应用潜力[3-4]。高光谱水质监测主要基于水体实测光谱丰富的谱域信息，通过建立光谱特征和水质指标的关系模型，实现对水质指标的反演[3]；但高光谱数据也存在数据冗余、易受环境噪声影响、不同波段间数据存在共线性等问题，影响水质反演精度及稳定性[5-6]。采取合适的数学建模方法进行数据优化和波段筛选，实现冗余变量的剔除和模型的优化，是水质反演模型建立的关键环节[7-8]。

【研究进展】高光谱环境遥感数据的波段筛选方法主要有遗传算法（GA）、无信息变量消除法（UVE）、竞争性自适应重加权算法（CARS）、连续投影算法（SPA）等[9]。其中 GA 是一种模拟自然进化搜索最优解方法，GA 与偏最小二乘法（Partial Least Squares, PLS）结合而成的 GA-PLS 算法，综合了 GA 全局优化搜索能力和 PLS 可有效解决变量间多重共线性问题的能力，成为光谱反演模型建立的研究热点，被广泛应用于化工、食品、农业等领域[8,10-12]。

【切入点】在水质遥感监测领域，GA-PLS 主要应用于湖泊水库中透明度、叶绿素 a、悬浮物和浊度等光学活性参数的反演研究[13-14]，而针对河网水体水质综合指标的高光谱反演研究报道较少。本研究以广东省中山市有代表性的河流、水渠为监测对象，以化学需氧量（COD）为水质综合评价指标，通过现场同步监测获得水体高光谱数据和 COD 质量浓度，建立河网水体化学需氧量（COD）质量浓度高光谱反演模型。

收稿日期：2020-02-13
基金项目：中国科学院重点仪器项目（YJKYYQ20170048）；国家自然科学基金项目（41676171）；2020 年广东省科技创新战略专项（PDJH2020b1091）
作者简介：蔡建楠（1982-），男。高级工程师，主要从事环境监测遥感应用研究。E-mail: jncai@foxmail.com
通信作者：邢前国（1975-），男。研究员，主要从事资源环境遥感应用研究。E-mail: qgxing@yic.ac.cn

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

【拟解决的关键问题】通过本研究验证 GA-PLS 算法对高光谱水质反演模型预测效果的改善作用，从而为河网水体水质高光谱监测提供参考。

1 材料与方法

1.1 研究区概况

中山市地处粤港澳大湾区几何中心，珠江口西岸，北连广州，毗邻港澳；属珠三角感潮河网区，辖区内河流密布，除主要外江河道外，有各种规模的内河涌、排灌渠等超过 1 000 条。中山市经济发达，产业集群密集，近年来随着城市快速发展和人口不断增加，河网水体多数受到一定程度的污染，水质复杂多变。

1.2 光谱测量及水质监测

光谱测量和水质监测于 2019 年 7—9 月实施，监测对象尽可能覆盖中山市不同规模的河流和渠道。剔除个别离群值后，共获得 146 个点位的现场实测光谱数据和水质样品。使用的光谱测量设备为美国 Ocean Optics 公司 USB4000 光谱仪，光谱分辨率<0.2 nm。测量时选择晴天无云或少云天气状况下，距 12:00 约 3 h 的时段内进行，以最大限度地降低太阳耀光影响。在每个点位离河流水面 0.5 m 处，测量河流水体上行辐射亮度（L_u）、天空辐射亮度（L_s）和太阳下行辐照度（E_d）。测量前使用标准白板进行仪器定标，后进行 3 次测量后取平均值作为光谱测定结果。计算测点光谱反射率 R_{rs} 计算式为[15]：

$$R_{rs}=\frac{L_u-\rho L_s}{E_d}。\qquad（1）$$

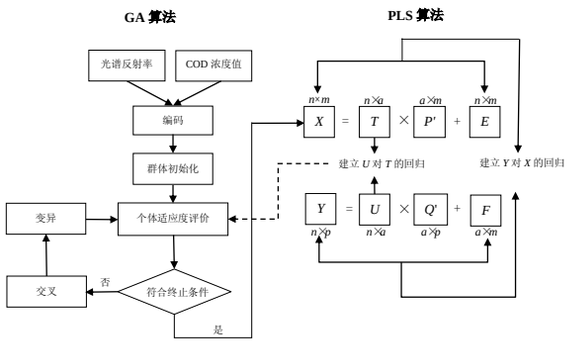
去除高光谱数据中噪声较大、信噪比低的两端波段，最终每个测点保留 400~900 nm 光谱范围共 2 534 个波长的反射率数据。

监测点的水样采集与光谱测量同步进行，采集河流表层 0~20 cm 处混合样，加入固定剂保存并送至实验室分析，监测项目为反映河流水质的综合性指标 COD，分析方法依据《水质化学需氧量的测定重铬酸钾法》（HJ 828—2017）。

1.3 基于 GA-PLS 的波段优选

1.3.1 GA-PLS 算法原理及实现

GA 是一种非导数的随机优化方法，借鉴生物界自然选择和进化机制，利用选择、交换和突变等算子的操作，随着不断的遗传迭代，使目标函数值较优的变量被保留，较差的变量被淘汰，最终达到最优结果[8, 11]。PLS 是一种基于高维投影思想的非参数回归方法，综合了多元线性回归分析、典型相关分析和主成分分析的基本功能，特别是当各自变量集合内部存在较高程度相关性时，通过提取主成分能对系统中的数据信息进行分解和筛选，提取对因变量解释性最强的综合变量，辨识系统中的信息和噪声，有助于建立最优模型[16-17]。将 GA 和 PLS 结合可融合 2 种算法的优势，建立更加稳定、简便、预测能力更强的模型。GA-PLS 算法的基本思想是将 PLS 算法对因变量预测效果的评价指标作为 GA 的适应度函数，用 GA 进行波段筛选，再用偏最小二乘法方法对筛选后的自变量建立反演模型。GA-PLS 算法实现流程见图 1。



注　图中 X 为自变量数据矩阵；Y 为因变量数据矩阵；T、U 分别为 X、Y 中提取的主成分矩阵；P'、Q'为荷载矩阵；E、F 为残差矩阵；n、m、a、p 为矩阵行列数。

Note　In the figure, X is the independent variable data matrix, Y is the dependent variable data matrix, T and U are the principal component matrix extracted from X and Y, $P'$and Q' are the load matrix, and E and F are the residual matrix. n, m, a, p are the number of rows and columns in the matrices.

图 1　GA-PLS 算法实现流程图

Fig.1　GA-PLS algorithm implementation flowchart

1.3.2 GA-PLS 算法参数设定及结果处理

GA 算法采用 PLS 交叉校验均方根误差（*RMSECV*）作为适应度函数，控制参数设定为：群体规模为 30，最大迭代次数为 100，交叉概率为 50%，变异概率为 1%，以遗传迭代次数为收敛终止条件。PLS 算法采用舍一交叉验证方法（PRESS）确定主成分的选取个数。由于 GA 算法单次计算结果具有随机性，为确保得到稳定的筛选结果，对数据执行 10 次 GA-PLS 重复运算并累积记录每次运算得到的入选波段和入选频率。

1.4 模型的建立与评价

将 146 个样本按照 COD 质量浓度升序排列，按照 1∶4 的间隔依序选择验证数据和建模数据，可将数据划分为建模集（116 个样本）和验证集（30 个样本）。其中建模集用于 GA-PLS 运算以筛选特征波段并建立 PLS 回归模型；验证集用于验证筛选得到的特征波段及回归模型对水体 COD 质量浓度的反演效果。同时将全谱段范围的反射率数据建立 PLS 回归模型，以比较 GA-PLS 算法和常规 PLS 算法的效果差异。模型建立和验证效果分别采用模型决定系数 R^2、建模集方根误差 *RMSEC* 和验证集方根误差 *RMSEP* 进行评价。各评价指标计算式为：

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$$R^2=\frac{\sum_{i=1}^{n}(y_i'-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}, \quad (2)$$

$$RMSEP\ (RMSEC)=\sqrt{\frac{\sum_{i=1}^{n}(y_i-y_i')^2}{n}}, \quad (3)$$

式中：y_i为建模集（验证集）第 i 个样本 COD 质量浓度实测值；y_i'为建模集（验证集）第 i 个样本 COD 质量浓度反演值；$\bar{y}$为建模集 COD 质量浓度实测值平均值；n 为样本数。

2 结果与分析

2.1 光谱反射率与 COD 质量浓度的关系

根据水质监测结果，146 个水质样品测得的 COD 质量浓度范围为 5~32 mg/L，平均值为 13.7 mg/L，标准差为 6.3 mg/L，偏度和峰度分别为 0.799 和-0.004。按照《地表水环境质量标准》（GB 3838—2002）Ⅱ～Ⅲ类水的限值作为区间划分标准，将各测点的高光谱反射率数据分别按照不同的 COD 质量浓度分类，以分析不同 COD 质量浓度水平下水体高光谱特征差异，（图 2）。

在 400~700 nm 波段范围内水体光谱反射率对低 COD 质量浓度水体有较好的区分度，Ⅱ类水质水体光谱反射率在此波段范围内均显著高于其他水质类别的水体；而在波长大于 720 nm 的波段范围内，则出现Ⅳ类水质水体的光谱分辨率高于Ⅲ类水质水体的现象。表明水体光谱反射率与 COD 质量浓度存在一定的相关性，但在不同波段范围内相关关系特征规律不同；总体上表明基于光谱分辨率的水体 COD 质量浓度反演具有较好的可行性。

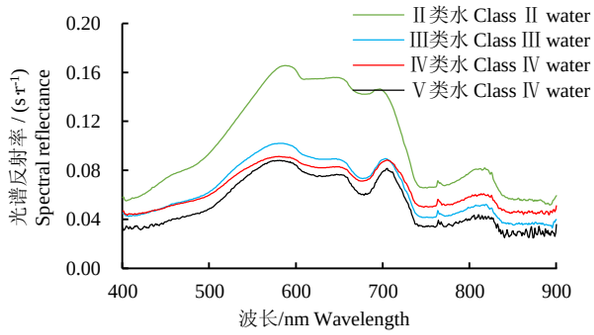


图 2 不同 COD 质量浓度水平对应的水体高光谱反射率

Fig.2 Hyperspectral reflectance of water body with different COD concentration levels

2.2 波段筛选过程及结果

样本集数据经过 10 次 GA-PLS 重复运算，得到的入选波段及累积入选频率见图 3。图 3 显示在 610~660、680~730、840~870 nm 等波段范围内均有连续的波段入选；入选频率最高为 650、720、763、818、842、873 nm 附近波长的波段，上述波段入选频率均超过 20%；入选频率较低的波段范围为 400~430、530~560、780~800 nm，表明该上述波段范围内的水体光谱反射率可能与 COD 质量浓度相关性较低。

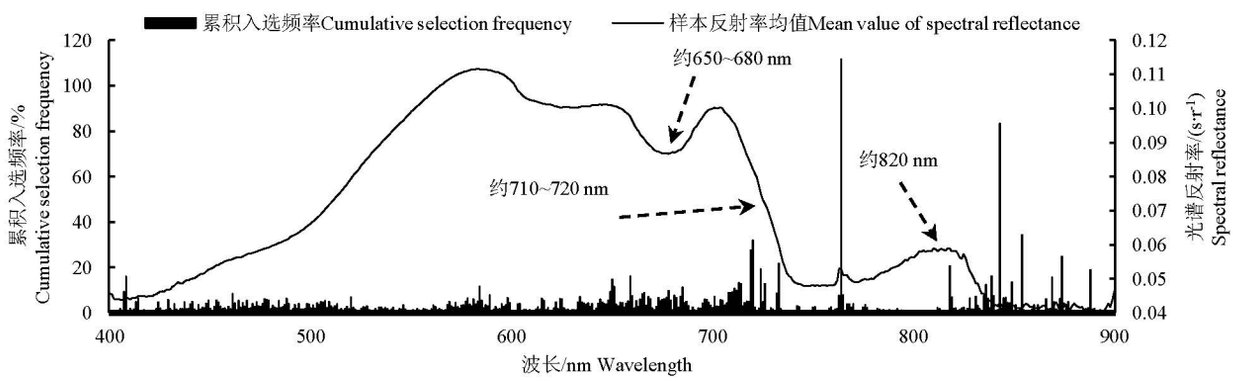


图3 GA-PLS运算得出的入选波段及频率

Fig.3 Selected bands and frequencies from GA-PLS calculation

2.3 水质反演模型预测效果及评价

为探索不同入选波段数量对模型建立及验证效果的差异，分别选取累积入选频率大于 15%（共 74 个波段）、大于 10%（共 140 个波段）、大于 5%（共 471 个波段）和大于 2%（共 1 235 个波段）的波段作为输入变量进行 PLS 算法建模，并和基于全谱段 PLS 模型的预测效果相比较，评价结果见表 1。由表 1 可知，与全谱段 PLS 模型相比，经过 GA-PLS 算法筛选后得到的不同波段组合作为 PLS 模型的输入变量，模型稳定性基本保持不变或有所提高。在各种波段组合中，以 471 个波段作为输入变量可得到最佳的验证效果，验证集 *RMSEP* 为 4.887 mg/L，较全谱段 PLS 模型降低 11.4%；而以 74 个波段（占全波段数的 2.9%）作为输入变量时，模型仍然保持良好的稳定性和验证精度，模型建立和验证效果评价指标均优于全谱段 PLS 模型，实现了对输入参数的有效缩减。

水质类别判定是水环境管理和评价的重要工作，故本研究按照《地表水环境质量标准》（GB3838—2002）

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

所划分的水质类别，将全部 146 个样的模型反演值予以分类，并与实测值对应水质类别进行比较，结果（图 4）显示，模型对全部水质样品的水体类别分类准确率为 72.0%，其中对Ⅱ、Ⅲ类水质水体分类准确性可达 83.9%，但对Ⅳ、Ⅴ类水质水体的分类准确性较差，为 21.4%。从全体样本 COD 实测值与模型反演值对比结果看（图 5），该模型对 COD 质量浓度小于 20 mg/L 的水体样本有较好的反演精度，但对于较高 COD 质量浓度样本的反演结果偏低。由于本研究中样本 COD 质量浓度主要处于Ⅱ～Ⅲ类水质范围内，说明在样本 COD 质量浓度主要分布范围内，基于 GA-PLS 的高光谱水质反演模型具有较高的验证精度。

表 1 模型建立和预测效果评价

Table 1 Evaluation of model establishment and prediction effect

模型类型及变量数 Model type and number of variables	*RMSEC*/(mg·L^{-1})	R^2	*RMSEP*/(mg·L^{-1})
全谱段PLS（2 354个波段） Full-spectrum PLS with 2 354 bands	4.330	0.497	5.514
GA-PLS（1 235个波段） GA-PLS with 1 235 bands	4.603	0.432	5.399
GA-PLS（471个波段） GA-PLS with 471 bands	4.117	0.546	4.887
GA-PLS（140个波段） GA-PLS with 140 bands	4.016	0.568	5.046
GA-PLS（74个波段） GA-PLS with 74 bands	4.086	0.553	5.167

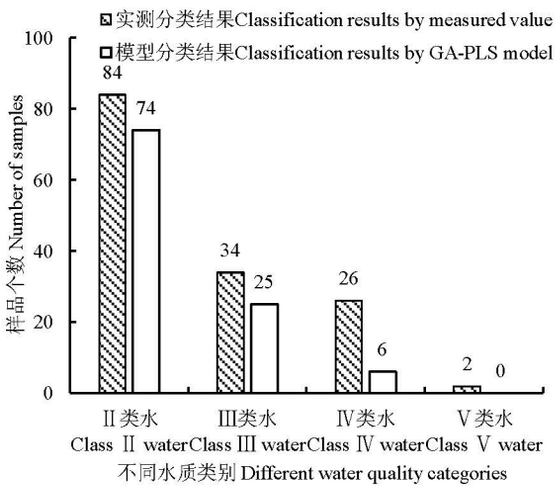


图 4 GA-PLS 模型水质分类结果与实测分类结果比较

Fig.4 Comparison of water quality clasification results by GA-PLS model with measured value

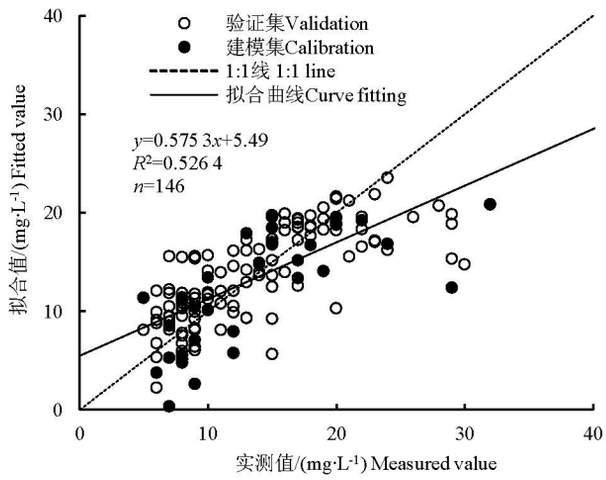


图 5 GA-PLS 模型反演值与实测值比较

Fig.5 Comparison of GA-PLS model retrieval and measured values

3 讨 论

本研究采用 GA-PLS 算法对高光谱数据进行波段筛选，实现了高光谱数据的降维，当采用数量仅为全谱段3.4%的优选波段作为自变量进行PLS建模时，模型建立及验证效果仍然优于全谱段 PLS 建模，说明 GA-PLS 算法可有效去除与水体 COD 质量浓度相关性小的波段数据，从而提高水体 COD 质量浓度高光谱反演模型的稳定性和预测能力。

对 GA-PLS 算法筛选得到的部分特征波段（图 3 中箭头标示处）做进一步分析可见，筛选得到的部分特征波段与水中藻类、悬浮颗粒物的光谱吸收、反射特征波段一致：如 650~680 nm 波长范围内的反射低谷由藻类吸收所致[18]；710~720 nm 波长范围内反射率下降与水分子吸收过程密切相关，反映水体洁净程度；约 820 nm 波长处的反射峰则是水体悬浮颗粒物散射程度加强而形成[19]。上述结果表明受测水体中悬浮颗粒物、藻类对 COD 质量浓度影响较大，是水体 COD 的主要来源；也反映 GA-PLS 算法选取的特征波段具有较好的合理性和指示意义。此种特性使得 GA-PLS 算法在水质高光谱监测中具有广阔的应用前景，可为同类型水体光谱监测的波段选择或多光谱传感器的波段设计提供重要参考，也有助于深入开展水体 COD 光谱吸收特征及机理研究。

由于 GA-PLS 算法结果具有随机性，多次运算后可能得到不同的波段筛选结果，选择何种方法确定最终输入变量对模型效果有一定的影响。本研究以各波段累积入选频率为标准筛选特征波段，可综合多次运算的结果，方法简便易行。但最优变量组合的确定仍然有必要结合目标污染物的光谱吸收特征、不同波段

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

间反射率的相关性和预测精度的要求等，开展进一步研究。

4 结 论

1）基于 GA-PLS 算法的河网水体 COD 质量浓度高光谱反演模型效果优于全谱段 PLS 模型。采用不同数量的筛选波段作为输入变量，验证集 *RMSEP* 最小为 4.887 mg/L，较全谱段 PLS 模型降低 11.4%；以筛选后的 74 个波段（占全波段数的 2.9%）作为输入变量时，模型仍可保持良好的稳定性和反演精度，表明 GA-PLS 算法可以实现对模型输入参数的有效缩减。

2）模型对全部水质样品的水体类别分类准确率为 72.0%，对Ⅱ、Ⅲ类水质水体分类准确性达 83.9%，表明在样本 COD 质量浓度主要分布范围内，GA-PLS 算法模型有良好水质类别分类准确性。

3）通过 GA-PLS 算法可对高光谱数据进行特征波段筛选，实现数据的降维优化，并进一步简化模型；GA-PLS 算法筛选得出的部分特征波段具有较好的合理性，有助于揭示 COD 的主要来源构成，可为水质高光谱监测传感器设计和水质反演机理研究提供参考，在河流 COD 快速监测中具有广阔的应用前景。

参考文献：

[1] 黄李童, 陈江, 朱渭宁, 等. 基于 Landsat-8 的城市湖泊水体总悬浮物吸收系数的遥感反演：以杭州西湖为例[J]. 环境科学学报, 2018, 38(10): 4 073-4 082.
HUANG Litong, CHEN Jiang, ZHU Weining, et al. Remote sensing inversion of total suspended matter absorption coefficient in coastal urban lakes using Landsat-8: Case study of west lake in Hangzhou[J]. Acta Scientiae Circumstantiae, 2018, 38(10): 4 073-4 082.

[2] 阿如娜, 青松, 包玉海. 基于 Landsat-8 OLI 数据的乌梁素海总溶解性固体质量浓度遥感反演[J]. 灌溉排水学报, 2018, 37(4): 99-105.
A Runa, QING Song, BAO Yuhai. Estimating the concentration of Dissolved solid in the wuliangsuhai lake based on the landsat-8 OLI images[J]. Journal of Irrigation and Drainage, 2018, 37(4): 99-105.

[3] 彭建, 徐飞雄, 邓凯, 等. 琅琊山景区不同指标浓度下水质光谱差异分析[J]. 光谱学与光谱分析, 2018, 38(5): 1 499-1 507.
PENG Jian, XU Feixiong, DENG Kai, et al. Spectral differences of water quality at different index concentrations: In langya mountain scenic area[J]. Spectroscopy and Spectral Analysis, 2018, 38(5): 1 499-1 507.

[4] FICHOT C G, DOWNING B D, BERGAMASCHI B A, et al. High-Resolution Remote Sensing of Water Quality in the San Francisco Bay-Delta Estuary[J]. Environmental Science & Technology, 2015, 50(2): 573.

[5] JAY S, GUILLAUME M, MINGHELLI A, et al. Hyperspectral remote sensing of shallow waters: Considering environmental noise and bottom intra-class variability for modeling and inversion of water reflectance[J]. Remote Sensing of Environment, 2017, 200: 352-367.

[6] 刘志刚, 徐勤超. 基于高光谱技术的基质含水率快速测定方法[J]. 灌溉排水学报, 2017, 36(10): 82-86.
LIU Zhigang, XU Qinchao. Measuring moisture content of rice husk using hyperspectral technology[J]. Journal of Irrigation and Drainage, 2017, 36(10): 82-86.

[7] 邹小波, 赵杰文. 用遗传算法快速提取近红外光谱特征区域和特征波长[J]. 光学学报, 2007, 27(7): 1 316-1 321.
ZOU Xiaobo, ZHAO Jiewen. Methods of characteristic wavelength region and wavelength selection based on genetic algorithm[J]. Acta Optica Sinica, 2007, 27(7): 1 316-1 321.

[8] 褚小立, 袁洪福, 王艳斌, 等. 遗传算法用于偏最小二乘方法建模中的变量筛选[J]. 分析化学, 2001, 29(4): 437-442.
CHU Xiaoli, YUAN Hongfu, WANG Yanbin, et al. Variable selection for partial least squares modeling by genetic algorithms[J]. Chinese Journal of Analytieal Chemistry, 2001, 29(4): 437-442.

[9] 于雷, 洪永胜, 周勇, 等. 高光谱估算土壤有机质含量的波长变量筛选方法[J]. 农业工程学报, 2016, 32(13): 95-102.
YU Lei, HONG Yongsheng, ZHOU Yong, et al. Wavelength variable selection methods for estimation of soil organic matter content using hyperspectral technique[J]. Transactions of the CSAE, 2016, 32(13): 95-102.

[10] 刘燕德, 张光伟, 蔡丽君. 基于高光谱的 GA 和 SPA 算法对赣南脐橙叶绿素定量分析[J]. 光谱学与光谱分析, 2012, 32(12): 3 377-3 380.
LIU Yande, ZHANG Guangwei, CAI Lijun. Analysis of chlorophyll in Gannan navel orange with algorithm of GA and SPA based on hyperspectral[J]. Spectroscopy and Spectral Analysis, 2012, 32(12): 3 377-3 380.

[11] 乔天, 吕成文, 肖文凭, 等. 基于遗传算法的土壤质地高光谱预测模型研究[J]. 土壤通报, 2018, 49(4): 773-778.
QIAO Tian, LYU Chengwen, XIAO Wenping, et al. Hyperspectral prediction modeling of soil texture based on genetic algorithm[J]. Chinese Journal of Soil Science, 2018, 49(4): 773-778.

[12] LEARDI R, GONZALEZ A L. Genetic algorithms applied to feature selection in PLS regression: How and when to use them[J]. Chemometrics and Intelligent Laboratory Systems, 1998, 41(2): 195-207.

[13] SONG K S, LI L, LI S, et al. Hyperspectral retrieval of phycocyanin in potable water sources using genetic algorithm-partial least squares (GA-PLS) modeling[J]. International Journal of Applied Earth Observation and Geoinformation, 2012, 18: 368-385.

[14] SUDDUTH K A, JANG G, LERCH R N, et al. Long-term agroecosystem research in the central Mississippi river basin: hyperspectral remote sensing of reservoir water quality[J]. Journal of Environmental Quality, 2015, 44(1): 71-83.

[15] MOBLEY C D. Estimation of the remote-sensing reflectance from

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

above-surface measurements.[J]. Applied Optics, 1999, 38(36): 7 442-7 455.

[16] 罗批, 郭继昌, 李锵, 等. 基于偏最小二乘回归建模的探讨[J]. 天津大学学报, 2002, 35(6): 783-786.
LUO Pi, GUO Jichang, LI Qiang, et al. Modeling construction based on partial least-squares regression[J]. Journal of Tianjin University, 2002, 35(6): 783-786.

[17] 高惠璇. 两个多重相关变量组的统计分析(3)(偏最小二乘回归与 PLS 过程)[J]. 数理统计与管理, 2002, 21(3): 58-64.
GAO Huixuan. Statistical analyses for multiple correlation variables of two sets(3) (Partial least-sguares Regression and PLS Procedure)[J]. Application of Statistics and Management, 2002, 21(3): 58-64.

[18] 刘志国, 周云轩, 沈芳. 河口水体泥沙浓度的水面光谱统计模式分析[J]. 水利学报, 2007, 38(7): 799-805.
LIU Zhiguo, ZHOU Yunxuan, SHEN Fang. Statistic analysis of suspended sediment concentration in offshore waters based on field measurement of reflectance hyper-spectral[J]. Journal of Hydraulic Engineering, 2007, 38(7): 799-805.

[19] 陈雯扬, 许惠平. 基于高光谱数据的东海近海悬浮泥沙浓度估测研究[J]. 水运工程, 2010 (2): 31-35.
CHEN Wenyang, XU Huiping. Estimation on suspended sedimentation concentration of East Sea based on hyperspectral data[J]. Portt & Waterway Engineering, 2010 (2): 31-35.

Using Hyperspectral Imagery and GA-PLS Algorithm to Estimate Chemical Oxygen Demand Concentration of Water in River Network

CAI Jiannan[1,2], LIU Hailong[3], JIANG Bo[3], HE Tianhui[1],
CHEN Wenjie[2], FENG Zhiwei[2], LI Zhuolin[2], XING Qianguo[3*]

(1.Zhongshan Environmental Monitoring Station, Zhongshan 528403, China; 2. Zhongshan Ecology and Environmental Agency, Zhongshan 528403, China; 3. Key Laboratory of Coastal Environmental Processes and Ecological Remediation, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003, China)

Abstract:【Objective】The hyperspectral remote sensing has proven potential to monitor water quality, but issues such as data redundancy and susceptibility to environmental variation could affect its accuracy and reliability. The genetic algorithm-partial least squares (GA-PLS) algorithm with a function to select sensitive spectral variables could resolve these problems. The GA-PLS algorithm was mainly used in retrieval of the optically active parameters such as transparency, chlorophyll-a, suspended matter and turbidity in surface water bodies. The purpose of this paper is to combine it with hyperspectral retrieval model to estimate chemical oxygen demand (COD) concentration of water in the river network in the Pearl River estuary.【Method】Hyperspectral imageries and COD concentration of 146 samples taken from water bodies in the Pearl River estuary were collected, and the characteristic bands of the hyperspectral reflectance data were screened using the GA-PLS algorithm to retrieve the COD concentration. The differences in retrieval accuracy between different band combinations were compared.【Result】The COD concentration retrieved from the hyperspectral imageries based on the GA-PLS algorithm is more accurate than that calculated using the full-spectrum PLS model. The minimum *RMSEP* of the method was 4.887 mg/L, 11.4% less than that of the full-spectrum PLS model. Using 74 filtered bands, accounting for 2.9% of the full bands, the model was still stable and accurate. Some characteristic bands obtained by the GA-PLS algorithm have physical interpretation, indicating that the screening results were rational.【Conclusion】The GA-PLS algorithm can be used to screen characteristic bands from the hyperspectral imageries to reduce the number of data and simplify the model as a result. It can accurately estimate COD of water in river networks.

Key words: hyperspectral imagery; genetic algorithms; partial least squares; chemical oxygen demand; river network waters

责任编辑：白芳芳

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net