

基于特征区域下凸点提取的藻类荧光光谱波长选择方法

张永彬¹, 朱丹丹¹, 陈颖^{1*}, 刘喆¹, 段玮靓¹, 李少华²

1. 燕山大学电气工程学院河北省测试计量技术及仪器重点实验室, 河北 秦皇岛 066004
2. 河北先河环保科技股份有限公司, 河北 石家庄 050000

摘要 藻华现象的频繁发生严重影响了海洋环境和人类的生产活动, 因此对水体浮游植物的监测十分重要。三维荧光光谱被广泛应用于水体浮游植物中藻类的群落组成分析和浓度定量分析, 然而三维荧光光谱数据中的信息冗余给藻类定性定量分析带来了一定的影响。针对光谱信息冗余问题, 提出了特征区域积分与凸点提取相结合的三维荧光光谱波长选择方法。以抑食金球藻、细长聚球藻、小球藻为研究对象, 采用 Savitzky-Golay 卷积平滑法对三维荧光光谱进行预处理, 解决了因外界因素造成的光谱噪声问题, 采用马氏距离法剔除三维荧光光谱数据集中的异常光谱样本, 运用浓度残差法剔除三维荧光光谱数据集中的异常浓度值样本, 然后通过偏最小二乘回归模型的内部交叉验证均方根误差衡量不同特征区域下凸点的可靠性进行波长变量的选择。为验证波长筛选方法的有效性, 对三种藻类建立偏最小二乘回归模型, 以内部交叉验证决定系数(R^2)、内部交叉验证均方根误差(RMSECV)作为模型评价指标。与全光谱数据建立的回归模型进行了比较, 抑食金球藻、小球藻、细长聚球藻的波长变量由全谱的 1071 个分别减少到 77 个、75 个、67 个, R^2 分别提高了 0.016 4, 0.002 和 0.032 4, RMSECV 分别降低了 1.8×10^5 , 2.0×10^5 , 2.6×10^5 。与 UVE 方法相比, 抑食金球藻、小球藻、细长聚球藻的波长变量分别减少了 599 个、357 个、317 个, R^2 分别提高了 0.014 5, 0.000 4 和 0.012 3, RMSECV 分别降低了 1.6×10^5 , 7.0×10^4 和 1.6×10^5 。经过该方法进行波长变量选择后, 减少了冗余信息, 提高了模型预测能力。

关键词 浮游植物; 三维荧光光谱; 特征区域; 凸点; 波长选择

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)10-3031-08

引言

近年来我国工农业发展迅速, 海水富营养化加剧, 水体中浮游藻类大量繁殖, 频繁引发藻华现象, 褐潮成因藻抑食金球藻、常见藻华中的小球藻和细长聚球藻已经成为海洋生态环境的主要污染, 影响人类健康安全和环境发展, 所以藻类种类和数量的测量对生态环境检测具有重要意义^[1-2]。浮游植物中含有的色素具有荧光效应, 因此荧光光谱分析是一种有效的测量方法, 其中三维荧光光谱法在藻类测量中应用广泛, 如何选择光谱波长变量从而提高三维荧光光谱的信噪比是目前一个重要研究方向。

三维荧光光谱是在已设置的激发波长和发射波长范围内

扫描所有波长点而得到的数据矩阵, 其数据较多, 但并不是光谱区域下的所有波长变量都含有有用信息, 只有少数光谱区域含有物质特征。若冗余区域波长变量参与建模会影响预测精度, 增加模型的计算, 光谱预处理虽然可以提高信噪比, 但并没有去除不相关的波长变量, 不能最大程度提高模型的精度, 没有减小模型复杂度。

传统的波长选择方法是峰值法, 对固定波长点下的荧光峰进行分析, 但该方法所选择的波长变量过少, 原光谱的数据信息没有得到充分利用, 无法有效反映整体区域荧光强度变化情况^[3]。目前的波长选择方法有无信息变量消除法(uninformative variables elimination, UVE)^[4]、连续投影算法(successive projections algorithm, SPA)^[5-6]、遗传算法(genetic algorithm, GA)^[7-8]等。SPA 通过投影方式选取线性

收稿日期: 2021-08-09, 修订日期: 2021-10-22

基金项目: 国家重点研发计划项目(2016YFC1400601), 河北省重点研发计划项目(19273901D, 20373301D), 河北省自然科学基金项目(F2020203066), 中国博士后基金项目(2018M630279), 河北省博士后择优资助项目(D2018003028), 河北省高等学校科学技术研究项目(ZD2018243)资助

作者简介: 张永彬, 1992 年生, 燕山大学电气工程学院硕士研究生 e-mail: 2824272678@qq.com

* 通讯作者 e-mail: chenying@ysu.edu.cn

关系最小的波长组合,减少光谱信息中的波长变量,但 SPA 所选择的波长变量的个数不能超过所用样品数量,其受样品数量的限制较大。UVE 将随机噪声矩阵加入光谱矩阵并建立偏最小二乘 (partial least-square, PLS) 回归模型,通过回归系数的稳定性进行波长变量选择。GA 是对初始群体通过自然遗传机制进行选择、交换、突变等算子操作产生新群体的过程,过程中适应度值较优的变量被保留。由于 UVE 的人工随机噪声矩阵和 GA 的随机初始种群的随机性,导致了每次波长变量选择的结果不同,且经过 UVE 波长选择后剩余的波长变量依然较多^[9]。

荧光区域积分结合凸点提取的波长选择方法是根据整体区域荧光强度变化选择出能够表征藻类色素的荧光特性的波长变量,同时避免选择过程的随机性。本文以抑食金球藻、小球藻、细长聚球藻为研究对象,在对 3 种藻的原始光谱进行预处理后,结合荧光区域积分和二元凸函数判定方法,进行特征区域波长变量筛选,并将筛选后的光谱数据作为输入建立 PLS 回归模型,验证其有效性。

1 基于凸点提取的波长选择原理

特征区域下的凸点提取是通过在主要荧光区域中提取曲面峰值点及峰值周围的点来达到去除荧光光谱冗余信息的目的。其整体过程分为两步,首先通过荧光区域积分^[10]找到荧光特征区域,通过阈值改变荧光特征区域的大小;然后在特征区域内投票统计所有波长点进行凸点提取后的得分,设定阈值衡量得分可靠性。调节特征区域和凸点投票统计后得分的阈值,大于阈值的波长变量被认为利于模型建立,予以保留。其实现过程包括:

Step1: 识别光谱数据中的凸点。三维荧光光谱可以看作是一个关于发射波长(x 轴)、激发波长(y 轴)、荧光强度(z 轴)的二元函数 $z = f(x, y)$,采用 Savitzky-Golay 多项式拟合微分法求偏导,计算 x 的偏导数时固定 y 轴,计算 y 的偏导数时固定 x 轴。根据二元凸函数^[11]判定定理,三维光谱的任一点(x, y)为凸点需要满足以下条件

$$\frac{\partial^2 f(x, y)}{\partial x^2} > 0, \frac{\partial^2 f(x, y)}{\partial y^2} > 0 \quad (1)$$

$$\frac{\partial^2 f(x, y)}{\partial x^2} \frac{\partial^2 f(x, y)}{\partial y^2} - \left[\frac{\partial^2 f(x, y)}{\partial x \partial y} \right]^2 \geq 0 \quad (2)$$

对所有样本三维荧光光谱的每个波长点的凸点进行统计,将

得到的统计值与样本总数的比值作为个波长点凸点得分。

Step2: 将整个荧光区域划分为若干个单位区域,由于三维荧光光谱数据是离散数据,其单位区域的体积分

$$\Phi_i = \sum_{EX} \sum_{EM} I(\lambda_{EX}, \lambda_{EM}) \Delta \lambda_{EX} \Delta \lambda_{EM} \quad (3)$$

式(3)中, $I(\lambda_{EX}, \lambda_{EM})$ 为在激发波长 λ_{EX} 、发射波长 λ_{EM} 处的荧光强度, $\Delta \lambda_{EX}$ 为激发波长间隔(取 5 nm), $\Delta \lambda_{EM}$ 为发射波长间隔(取 5 nm)。

设最大单位区域体积分的 $\frac{1}{\eta_i}$ ($\eta_i = 10, 9, 8, 7, 6, 5, 4, 3$) 作为阈值,大于该阈值的所有单位区域体积分的并集作为特征区域。

Step3: 求光谱数据的凸点得分与特征区域的交集,得到特征区域下的光谱数据凸点得分结果。

Step4: 设定特征区域下的光谱数据凸点得分阈值 η ($0.3 < \eta < 1$), 小于设定阈值,则认为该凸点是不利于建模的波长变量或者是由于噪声引起的波长变量。

Step5: 将选择的波长变量建立偏最小二乘回归模型,采用留一交叉验证法计算内部交叉验证均方根误差 (RMSECV)。

Step6: 遍历所有 η_i 值,重复 Step2—Step5 过程,当 RMSECV 值最小时所选择的波长变量为最终的波长变量选择结果

2 实验部分

所用的抑食金球藻和小球藻由秦皇岛海洋环境监测中心提供,细长聚球藻购自中国科学院淡水藻种库。采用 FS920 三维荧光光谱仪对三藻种进行光谱测量,测量条件:激发波长范围为 400~650 nm,发射波长范围为 630~730 nm,激发步长和发射步长均为 5 nm,狭缝宽度为 5 nm,设置发射波长滞后激发波长 10 nm,用 1 cm 石英比色池每天取样一次进行光谱测量。

实验中所选藻种的培养周期均为 14 d,为采集到更多丰度下的光谱信息,每种藻培养 3 份,每份保持不同起始丰度。经过多周期测量,最终获得藻种光谱数据共 306 个。不同色素有不同的荧光特性,藻类中色素的种类决定荧光峰的位置,色素的含量决定荧光峰的强弱,图 1 给出了三种藻类的高线谱。图 1(a)中 440/685, 450/685 和 470/685 nm 三处

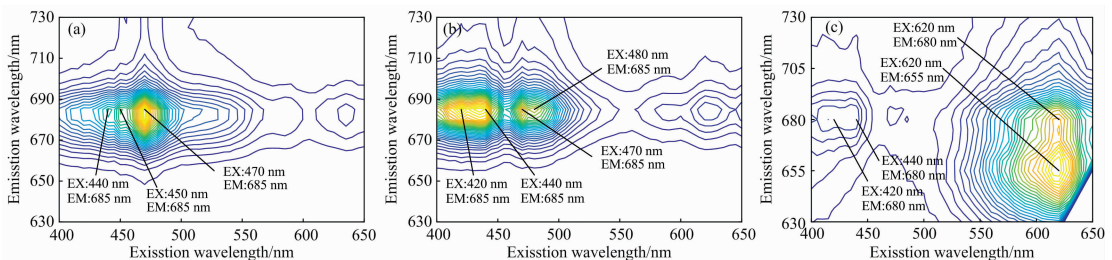


图 1 三藻种等高线光谱图

(a): 抑食金球藻; (b): 小球藻; (c): 细长聚球藻

Fig. 1 Contour spectra of 3 phytoplankton

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

的荧光峰分别由抑食金球藻的叶绿素 a (Chla)、叶绿素 c (Chlc) 和叶绿素 b (Chlb) 造成；图 1(b) 中 440/685, 470/685, 480/685 nm 三处的荧光峰分别由小球藻的 Chla、Chlb 和类胡萝卜素造成，420/685 nm 处的荧光峰为 Chla 的次吸收带引起的次荧光峰；图 1(c) 中 420/680 和 440/680 nm 分别为细长聚球藻中 Chla 的次荧光峰和主荧光峰，620/655 和 620/680 nm 处的荧光峰均由藻蓝蛋白(PC)引起，但 PC 对光的吸收和释放特性主要表现在 620/655 nm 处的强荧光峰。Chla 存在于三种藻中，但在细长聚球藻中 Chla 的荧光发射波长为 680 nm，可能是由于细长聚球藻中的 Chla 与蛋白质结合，改变了荧光的响应特性，从而引起发射光波长的偏移。

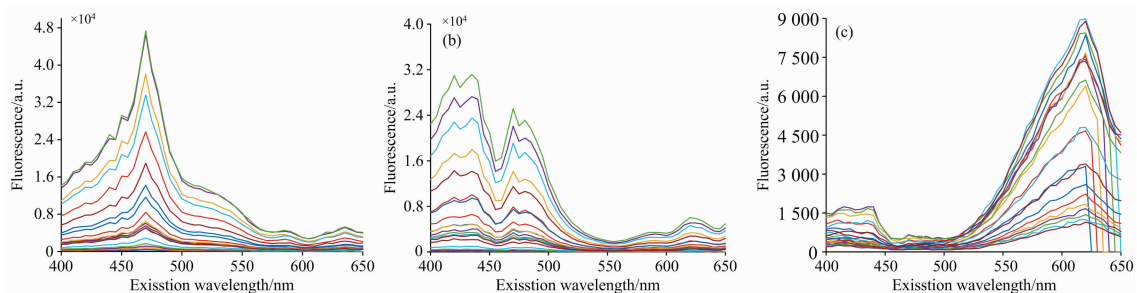


图 2 三藻种平滑去噪前的激发光谱图

(a): 抑食金球藻；(b): 小球藻；(c): 细长聚球藻

Fig. 2 Excitation spectra of 3 phytoplankton before smooth denoising

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

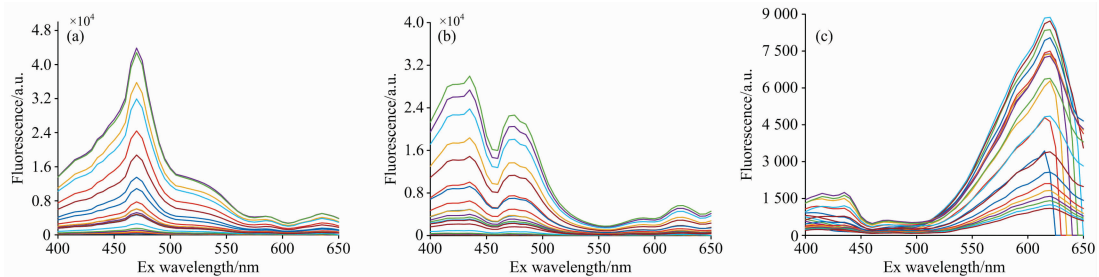


图 3 三藻种平滑去噪后的激发光谱图

(a): 抑食金球藻；(b): 小球藻；(c): 细长聚球藻

Fig. 3 Excitation spectra of 3 phytoplankton after smooth denoising

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

3.2 异常样本数据剔除

在光谱采集过程中，由于测量人员、测量环境、测量仪器等因素的影响，导致所测量的样本中存在一些光谱异常样本和化学值异常样本，这些异常数据会影响模型的精度，需要在建模之前剔除^[12]。

光谱异常样本通过马氏距离法剔除。计算出各个样本到平均光谱的马氏距离，通过平均值和标准差来设定阈值，调节阈值权重系数来改变阈值范围，剔除不利于建模的样本。光谱异常样本剔除结果如图 4 所示，本文中抑食金球藻、小球藻、细长聚球藻的阈值权重系数分别取 0.5, 0 和 0.5，剔除样本数分别为 17 个、19 个、13 个。

剩余的样本中可能存在化学值异常样本，本实验中化学

3 结果与讨论

3.1 光谱数据去噪

由于一些因素的影响，通过光谱仪获得的荧光光谱既有有用信号，同时也有噪声信号。本文通过 Savitzky-Golay 卷积平滑法去除噪声信号，平滑窗口宽度为 5，多项式次数为 2。图 2 和图 3 为三种藻类样本激发光谱平滑前后对比图，可看出各藻种光谱的大体形状没有发生变化，消除了光谱上的毛刺，将误差信息重新分配到整个光谱中，使光谱更加平滑，降低了噪声干扰。

值指藻种浓度，可用浓度残差法剔除。通过留一交叉验证法依次获得数据集各样本的浓度残差组成浓度残差向量，设置浓度残差的 F 统计性检验的显著性水平阈值，浓度残差大于该阈值的样本为化学值异常样本。浓度残差剔除异常样本的结果如图 5，抑食金球藻、小球藻、细长聚球藻的显著性水平均为 0.5，剔除的浓度异常样本分别为 32 个、34 个、34 个，最终可用于建模的样本数分别为 53 个、55 个、49 个。

3.3 基于荧光区域积分和凸点提取的波长选择方法

3.3.1 全波长凸点提取

对各藻类样本数据集下的全谱波长点依次进行凸点统计，如果满足凸点条件，则将该波长点保留并将凸点数加 1，以每个波长点统计的凸点数与样本总数的比值作为该点的凸

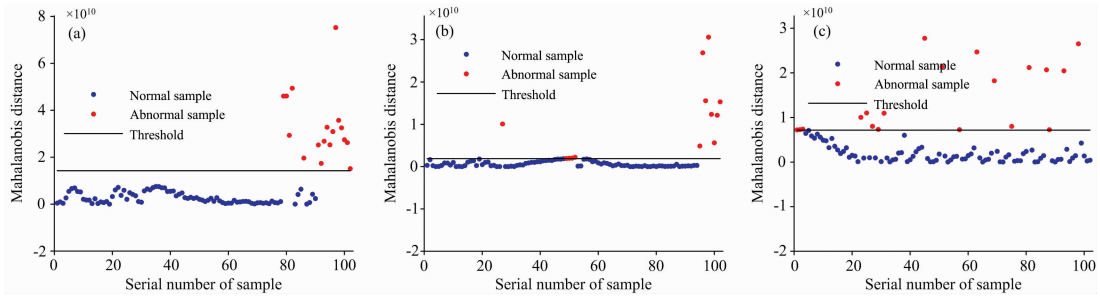


图 4 三藻种马氏距离分布图

(a): 抑食金球藻; (b): 小球藻; (c): 细长聚球藻

Fig. 4 Mahalanobis distance distribution of 3 phytoplankton

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

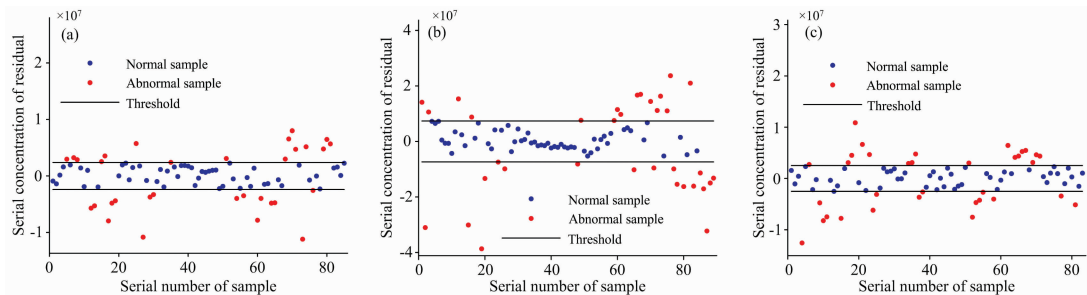


图 5 三藻种浓度残差分布图

(a): 抑食金球藻; (b): 小球藻; (c): 细长聚球藻

Fig. 5 Concentration residual distribution of 3 phytoplankton

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

点统计得分,图 6 为各藻类凸点得分统计图,抑食金球藻、小球藻、细长聚球藻的全谱波长点均为 1 071 个,被判定为凸点的波长点个数分别为 369 个、422 个和 501 个。

3.3.2 初始特征区域的确定

计算各藻类全谱范围下所有单位荧光区域的体积积分,初始特征区域阈值取最大单位区域体积积分的 1/10,大于该阈值的单位体积积分的并集作为初始特征区域,图 7 为各藻

类提取到的初始候选特征区域结果图,抑食金球藻、小球藻、细长聚球藻在初始特征区域下的波长点个数分别为 340 个、409 个、525 个,由图 6 和图 7 可知虽然主要的色素峰值点既被判定为凸点,又包含在初始特征区域内,但在初始候选特征区域外,有些波长点既不是色素的特征峰又无其他明显的光谱特征信息,却多次被判定为凸点。

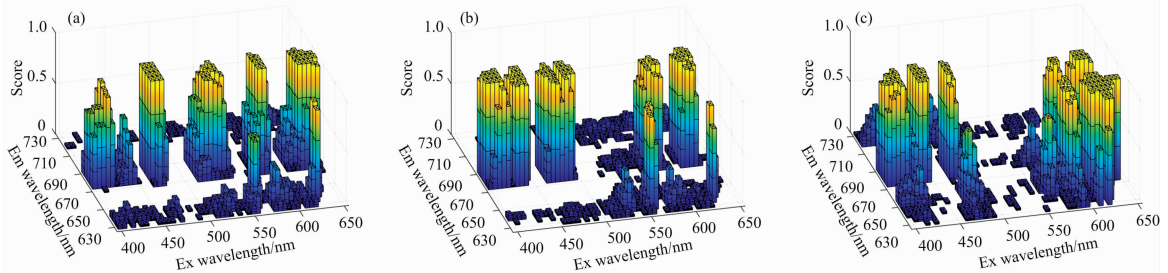


图 6 三藻种凸点统计图

(a): 抑食金球藻; (b): 小球藻; (c): 细长聚球藻

Fig. 6 Convex points statistics of 3 phytoplankton

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

3.3.3 特征区域和凸点提取阈值的确定

图 8 为初始特征区域下的凸点得分,抑食金球藻、小球藻、细长聚球藻在该区域下被判定为凸点的波长个数分别为 131 个、142 个、259 个,已有效去除不含特征信息的凸点,

但此时由噪声引起的光谱凸点波长变量包含在内,同时由于特征区域阈值选取问题,初始特征区域并不精确,区域内可能存在与藻种信息相关性不大的波长点,依旧有光谱冗余。图 9 为阈值 n_0 和 n_1 取不同值时 RMSECV 的变化情况, RM-

SECV 值越小，模型的预测能力越好，合适的 n_q 和 n_l 值可以提高模型的预测能力。图 9(a)中 n_q 取 1/8, 1/7, 1/6 和 1/5 时 RMSECV 值相等，图 9(b)中 n_q 取 1/4 和 1/3 时 RMSECV 值相等，是由于在这几个特征区域阈值下，光谱的积分区域发生了变化，但波长变量的数量并没有改变，这时模型的预测能力由凸点的阈值 n_l 决定。当抑食金球藻、小球

藻、细长聚球藻的候选特征区域阈值分别为 1/10, 1/6 和 1/8, 凸点得分阈值分别为 0.4, 0.5 和 0.9 时 RMSECV 最佳。图 10 为最佳候选特征区域阈值和最佳凸点得分阈值下各藻种筛选波长的最终结果，抑食金球藻、小球藻、细长聚球藻剩余波长点数分别为 77, 75 和 67。

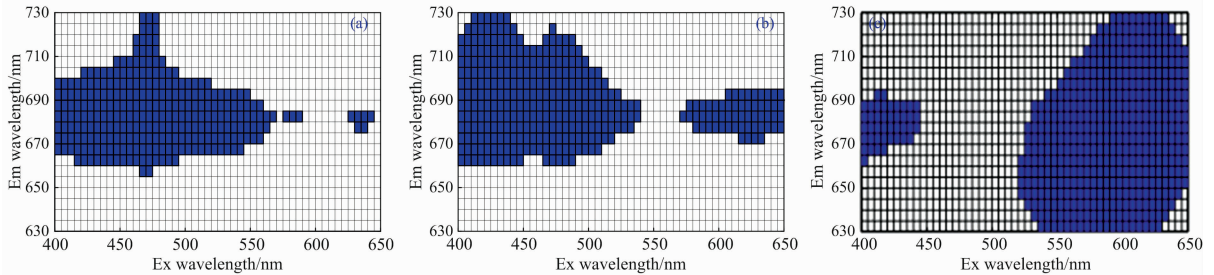


图 7 初始候选特征区域

(a): 抑食金球藻; (b): 小球藻; (c): 细长聚球藻

Fig. 7 Initial candidate feature region

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

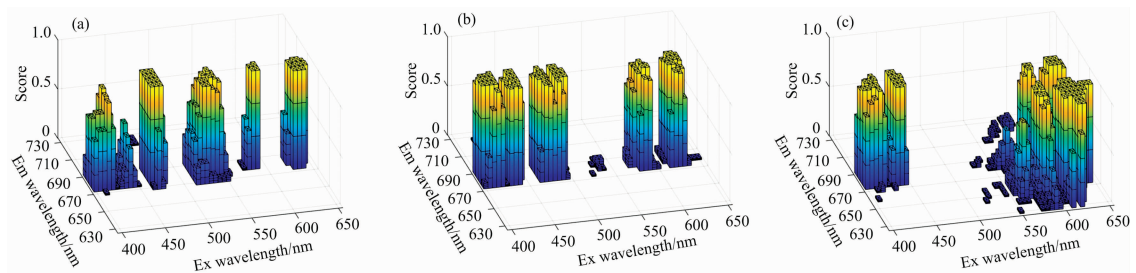


图 8 初始候选特征区域下的凸点统计

(a): 抑食金球藻; (b): 小球藻; (c): 细长聚球藻

Fig. 8 Statistics of convex points under the initial candidate feature region

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

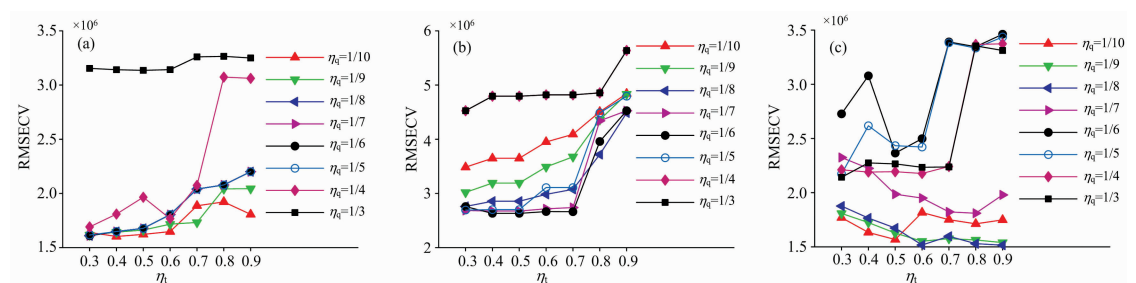


图 9 η_q 和 η_l 取不同值时 RMSECV 的变化曲线

(a): 抑食金球藻; (b): 小球藻; (c): 细长聚球藻

Fig. 9 Change curve of RMSECV with different η_q and η_l

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

3.4 建模方法及模型评价

分别用全波长数据、特征区域结合凸点提取后波长数据、UVE 提取后波长数据建立 PLS 回归模型，采用留一交叉验证法进行内部交叉验证，以内部交叉验证决定系数 (R^2)、内部交叉验证均方根误差 (RMSECV) 作为评价指标评估波长选择的效果，其中 R^2 反映模型验证的稳定性，

RMSECV 反映模型的预测能力， R^2 越接近 1, RMSECV 越接近 0, 模型的预测性能及稳定性越好。图 11 为经过 UVE 提取的波长变量结果，表 1 为各 PLS 回归模型的 R^2 、RMSECV 值。抑食金球藻经该方法波长筛选后的 R^2 相对全谱提高了 0.016 4, RMSECV 降低了 1.8×10^5 , 波长点由全谱的 1 071 个减少到 77 个，经 UVE 波长筛选后的 R^2 相对全

谱提高了 0.001 9, RMSECV 降低了 3.0×10^4 , 波长点由全谱的 1 071 个减少到 676 个; 小球藻经该方法波长筛选后的 R^2 相对全谱提高了 0.002, RMSECV 降低了 2.0×10^5 , 波长点由全谱的 1 071 个减少到 75 个, 经 UVE 波长筛选后的 R^2 相对全谱提高了 0.001 6, RMSECV 降低了 1.3×10^5 , 波长点由全谱的 1 071 个减少到 432 个; 细长聚球藻经该方法波长筛选后的 R^2 相对全谱提高了 0.032 4, RMSECV 降低了 2.6×10^5 , 波长点由全谱的 1 071 个减少到 67 个, 经 UVE 波长筛选后的 R^2 相对全谱提高了 0.010 1, RMSECV

降低了 1.0×10^5 , 波长点由全谱的 1 071 个减少到 384 个。由以上分析可知, 三藻种在经过特征区域结合凸点提取后 R^2 均得到提高, 其中小球藻的 R^2 提高效果并不显著, 但在没有降低模型预测能力的基础上波长变量减小到原来的 7%, 该波长提取方法是有效的。相比 UVE, 特征区域结合凸点提取所得到的模型精度更高, 所选择的波长变量更少, 且均分布在含有荧光特征的区域, 保留了所含色素的特征峰。

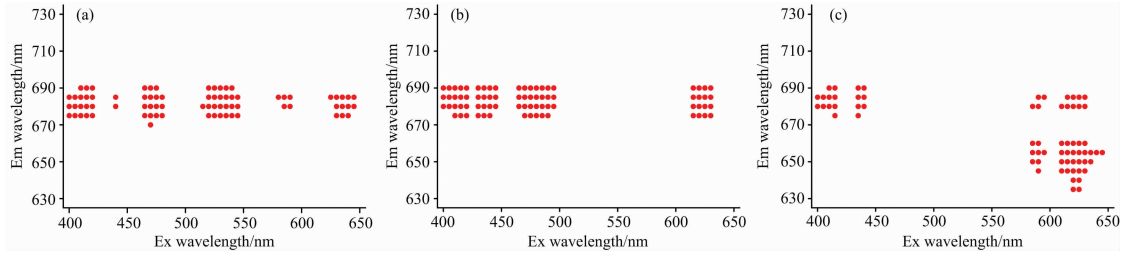


图 10 特征区域结合凸点提取波长选择结果

(a): 抑食金球藻; (b): 小球藻; (c): 细长聚球藻

Fig. 10 Wavelength selection results of feature region combined with convex point extraction

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

表 1 三种方法的 R^2 和 RMSECV

Table 1 R^2 and RMSECV of three methods

Phytoplankton	Method	Wavelength point	R^2	RMSECV/ (10^6 cells \cdot mL $^{-1}$)
<i>Aureococcus anophagefferens</i>	Total-PLS	1 071	0.8143	1.55
	Feature region+convex point-PLS	77	0.830 7	1.37
	UVE-PLS	676	0.816 2	1.53
<i>Chlorella vulgaris</i>	Total-PLS	1 071	0.983 3	2.80
	Feature region+convex point-PLS	75	0.985 3	2.60
	UVE-PLS	432	0.984 9	2.67
<i>Synechococcus elongatus</i>	Total-PLS	1 071	0.875 2	1.76
	Feature region+convex point-PLS	67	0.907 6	1.50
	UVE-PLS	384	0.895 3	1.66

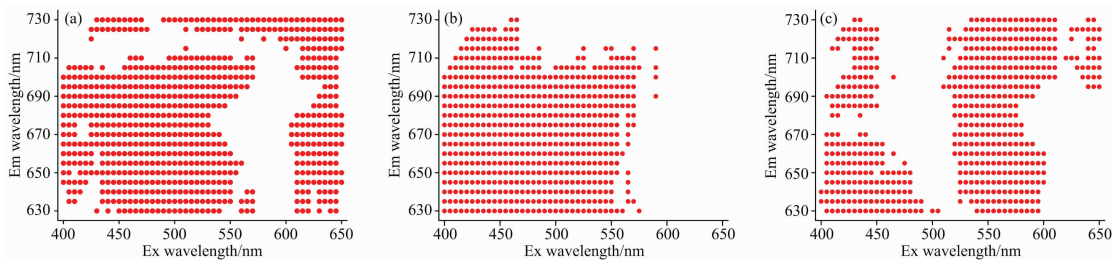


图 11 UVE 波长选择结果

(a): 抑食金球藻; (b): 小球藻; (c): 细长聚球藻

Fig. 11 Wavelength selection results of UVE

(a): *Aureococcus anophagefferens*; (b): *Chlorella vulgaris*; (c): *Synechococcus elongatus*

4 结 论

利用实验得到的三种藻类不同浓度下的三维荧光光谱数

据, 提出了特征区域结合凸点提取的波长变量选择方法。根据藻类光谱找到包含荧光峰的初始特征区域, 利用二元凸函数判定定理统计初始特征区域内的凸点, 设定特征区域和凸点统计结果的不同阈值, 通过计算 PLS 回归模型的 RM-

SECV 值衡量特征区域下波长变量凸点提取的可靠性从而选择波长去留。结果表明, 用该方法得到的波长变量建立的 PLS 模型优于用全谱波长变量建立的 PLS 模型, 且选择的

波长变量在色素荧光峰附近, 能够很好地表征藻类色素的荧光特性, 为三维荧光光谱分析提供了依据。

References

- [1] Nunes S, Latasa M, Delgado M, et al. Deep-Sea Research Part I, 2019, 151(5): 103059.
- [2] Nankabirwa A, DeCrop W, Vander Meeren T, et al. Ecological Indicators, 2019, 107: 105563.
- [3] PU Shu-juan, HE Pei-xiang, ZHANG Yue, et al(蒲淑娟, 贺培翔, 张悦, 等). Petrochemical Industry Application(石油化工应用), 2021, 40(5): 102.
- [4] CHEN Yuan-yuan, WANG Zhi-bin, WANG Zhao-ba(陈媛媛, 王志斌, 王召巴). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(1): 299.
- [5] Zhang Chu, Liu Fei, Kong Wenwen, et al. Sensors, 2015, 15(7): 16576.
- [6] XIONG Zhi-xin, MA Pu-fan, LIANG Long, et al(熊智新, 马璞璠, 梁龙, 等). Transactions of China Pulp and Paper(中国造纸学报), 2019, 34(4): 46.
- [7] HUA Chen-zhi, ZHAO Ling, SONG Jian-jun(花晨芝, 赵凌, 宋建军). Journal of Sichuan Normal University(四川师范大学学报), 2019, 42(6): 825.
- [8] MAI Shu-kui, YANG Yang, ZHAO Xiao-bo, et al(买书魁, 杨洋, 赵小波, 等). Food Science and Technology(食品科技), 2019, 44(2): 301.
- [9] Chang Haitao, Zhu Lianqing, Lou Xiaoping, et al, Sensors, 2016, 16(6): 827.
- [10] CHEN Shi-yu, LI Yan, LI Ai-min(陈诗雨, 李燕, 李爱民). Environmental Science & Technology(环境科学与技术), 2015, 38(5): 64.
- [11] ZHANG Xue-mao(张学茂). Journal of Langfang Teachers College(廊坊师范学院学报), 2011, 11(4): 18.
- [12] SHI Lu-zhen, ZHANG Jing-chuan, WANG Yan-qun, et al(石鲁珍, 张景川, 王彦群, 等). Journal of Chinese Agricultural Mechanization(中国农机化学报), 2016, 37(6): 99.

Wavelength Selection Method of Algal Fluorescence Spectrum Based on Convex Point Extraction From Feature Region

ZHANG Yong-bin¹, ZHU Dan-dan¹, CHEN Ying^{1*}, LIU Zhe¹, DUAN Wei-liang¹, LI Shao-hua²

1. Hebei Province Key Laboratory of Test/Measurement Technology and Instrument, School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China
2. Hebei Sailhero Environmental Protection Hi-tech Co., Ltd., Shijiazhuang 050000, China

Abstract The frequent occurrence of algal bloom seriously affects the Marine environment and human production activities, so it is very important to monitor the phytoplankton in water. 3D fluorescence spectroscopy has been widely used in the analysis of algae community composition and the quantitative analysis of algae concentration in water phytoplankton. However, the information redundancy in 3D fluorescence spectrum data has significantly impacted the qualitative and quantitative analysis of algae. In order to solve the problem of spectral information redundancy, a new wavelength selection method of 3D fluorescence spectrum based on the combination of feature region and convex point extraction is proposed. Taking *Aureococcus anophagefferens*, *Chlorella Vulgaris*, and *Synechococcus elongatus* as the research object, the Savitzky-Golay convolution smoothing method was used to preprocess the 3D fluorescence spectrum to solve the problem of spectral noise caused by external factors. The Mahalanobis distance method was used to eliminate the abnormal spectral samples in the 3D fluorescence spectrum data set. The residual concentration method was used to eliminate the abnormal concentration value samples in the 3D fluorescence spectrum data set. Then the reliability of the convex points under different characteristic regions was measured by the root mean square error of cross-validation (RMSECV) of the PLS regression model, and the wavelength variable was selected. In order to verify the effectiveness of the wavelength selection method, the PLS regression model was established for the three algae species, and the determination coefficient (R^2) and root mean square error of cross-validation (RMSECV) were used as the evaluation indexes of the model. Compared with the regression model established with the full spectrum data, the wavelength variables of *Aureococcus anophagefferens*, *Chlorella Vulgaris*, and *Synechococcus elongatus* respectively decreased

from 1 071 to 77, 75 and 67, and R^2 respectively increased by 0.016 4, 0.002 and 0.032 4. RMSECV was respectively reduced by 1.8×10^5 , 2.0×10^5 and 2.6×10^5 . Compared with the UVE method, the wavelength variables of *Aureococcus anophagefferens*, *Chlorella Vulgaris*, and *Synechococcus elongatus* were respectively reduced by 599, 357 and 317, and R^2 was respectively increased by 0.014 5, 0.000 4 and 0.012 3, RMSECV was respectively decreased by 1.6×10^5 , 7.0×10^4 and 1.6×10^5 . After the selection of wavelength variables by the method of feature region combined with convex point extraction, the redundant information is reduced, and the model's prediction ability is improved.

Keywords Phytoplankton; 3D fluorescence spectroscopy; Feature region; Convex point extraction; Wavelength selection

(Received Aug. 9, 2021; accepted Oct. 22, 2021)

* Corresponding author

(上接 3023)

论文摘要提交具体步骤

1. 请您在光谱网上(<http://www.sinospectroscopy.org.cn>)用真实姓名注册,注册系统已经设置认证功能,请用手机号码或邮箱注册。
2. 点击光谱网会议会展栏目,选择《第 22 届全国分子光谱学学术会议暨 2022 年光谱年会》。
3. 点击会议基本情况下的会议快捷通道中的“会议投稿”。
4. 输入用户名和密码登陆,在页面下选择稿件提交。
5. 按照提示提交稿件。

报告形式

为充分提高会议学术交流的效率,会议将采用“口头报告”和“墙报展示”两种方式进行学术交流。无论是口头报告还是墙报展示,均属大会同等学术交流。为尊重个人意见和便于组委会的安排,请大家在会议注册时,提交“口头报告”或“墙报”的题目。为了鼓励博士、硕士研究生积极参与学术交流活动,本次会议将继续设立“优秀青年论文奖”和“优秀墙报奖”,表彰那些研究水平高、能突出研究内容要点、条理清晰的“口头报告”和“墙报”,大会将给获奖作者颁发优秀论文证书和奖金。同时会议还将邀请国内外知名专家学者就光谱有关学术领域的前沿热点问题作大会报告和主题报告。

主要报告形式有:

1. 大会邀请报告:主要邀请国内外知名专家学者报告光谱分析的前沿技术在各个领域的最新研究进展。
2. 主题邀请报告:本次会议将选择光谱技术的热点应用领域,开设多个专题论坛,邀请在该领域的知名专家作论坛主题报告。
3. 口头报告:由参会代表申请、组委会审核方式确定报告人选。
4. 青年论坛报告:为博士、硕士研究生开设交流平台,并评选“优秀青年论文奖”。
5. 墙报展示:作为本次会议的主要交流和展示形式。会议统一安排墙报讲解时间,希望作者按时到位讲解。

重要时间

开通会议注册系统:2022 年 5 月 10 日

论文截稿日期:2022 年 6 月 30 日

第二轮会议通知:2022 年 5 月

第三轮会议通知:2022 年 10 月

会议召开期:2022 年 11 月 11 日—14 日

会议组织机构、注册费及缴纳方式、宾馆住宿介绍及住房预定等信息将在 2022 年 5 月初在会议主页上发布,请您经常浏览光谱网上会议主页,了解会议筹备情况和会议具体安排。网址:<http://www.sinospectroscopy.org.cn>

(下转 3067 页)