

一种基于 EWMA-PCA 的水质光谱数据标准化方法研究

周思寒^{1,2}, 胡新宇^{1,2}, 汤斌^{1,2,3*}, 赵明富^{1,2},
李奉笑^{1,2}, 汪仁杰^{1,2}, 肖棋森^{1,2}, 肖渝^{1,2}

1. 重庆理工大学现代光电检测技术与仪器重点实验室, 重庆 400054
2. 重庆理工大学光纤传感与光电检测重庆市重点实验室, 重庆 400054
3. 重庆大学光电技术及系统教育部重点实验室, 重庆 400044

摘要 模型传递对于解决由于样本与各仪器响应函数不同所导致的量测信号不一致具有重要意义, 而解决模型传递的有效方法是仪器或数据标准化。针对现有的光谱标准化方法中, 鲜有对紫外-可见吸收光谱的应用研究, 且紫外-可见光谱法水质多参数检测光谱探测单元存在分辨率、精度、响应范围不统一, 难以进行不同仪器间测试数据的比对及多参数数据拟合的问题, 提出采用 EWMA-PCA 归一化算法, 实现紫外-可见水质光谱在不同仪器上的模型传递。EWMA(exponentially weighted moving-average)是一种指数加权平均移动算法, 用以寻找以较高概率产生观察紫外-可见水质光谱数据的系统发生树, 最大概率复原理论紫外-可见水质光谱数据, 使紫外-可见光谱特征不丢失、不偏移, 减小由于数据处理对紫外-可见水质光谱数据的影响。采用不同浓度的邻苯二甲酸氢钾溶液, 对日本滨松 C10082CAH 光谱仪、美国海洋光学 Maya2000Pro 光谱仪以及厦门奥谱天成 ATP2000 光谱仪进行对比测试实验。对比组 1 选取源机滨松 C10082CAH 光谱仪和目标机海洋 Maya2000Pro 光谱仪, 对比组 2 选取源机滨松 C10082CAH 光谱仪和目标机奥谱天成 ATP2000 光谱仪, 对比组 3 选取源机海洋 Maya2000Pro 光谱仪和目标机奥谱天成 ATP2000 光谱仪。三组比对实验结果表明, 该算法能很好地应用于不同的比对光谱仪中, 在采用 EWMA-PCA 归一化算法对水质吸收光谱数据标准化后, 相关系数达到 99.576 5%, 方差达到 0.082 3%, 且波峰偏移量可降低至 0.000 5%, 基于 EWMA-PCA 归一化光谱标准化算法具适应性广、所需传递样本少、传递精度高等优点, 研究结果对光谱法水质检测仪器的广泛应用具有重要理论指导意义和工程应用参考价值。

关键词 紫外-可见水质光谱; 指数加权平均移动-主成分分析; 标准化; 归一化; 模型传递

中图分类号: X52 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)11-3443-08

引言

传统的化学方法水质检测精度高, 但检测周期较长、需专业人员操作、所使用化学试剂容易产生二次污染^[1-2]。近年来, 紫外-可见光谱法广泛应用于水质检测中, 其具有快速、无二次污染、指纹检测、可污染溯源等优点, 已成为水质检测领域的研究热点^[2]。在紫外-可见光谱法水质检测系统中, 为了提高整个检测系统的精度及稳定性, 对数据的光谱探测单元提出更高要求。不同仪器生产厂家制作工艺不同, 同一厂家产品批次不同, 以及检测设备老化程度不同等

影响因素, 对同一样本检测结果会出现差异。在一台仪器(称源机, Master)建立的模型, 将无法适用于另一台仪器(称目标机, Slave)。引起该模型“失效”的原因主要是样本与各仪器响应函数不同所引起的量测信号不一致。因此, 实际监测环境需针对不同监测点、不同光谱仪所测得的紫外-可见水质光谱数据分别建模, 会导致效率低下, 模型维护成本增加, 而模型传递能够有效解决上述问题。

国内外学者对拉曼光谱、高光谱、近红外光谱中的模型传递问题有着广泛深入的研究^[3]。如结合标准正态变换, 利用分段直接标准化实现拉曼光谱模型传递; 有研究^[4]实现被测物的拉曼光谱的预测模型在同类物种之间传递。陈奕云

收稿日期: 2019-10-16, 修订日期: 2020-02-05

基金项目: 国家自然科学基金项目(61805029), 重庆市社会事业与民生保障科技创新专项(cstc2017shmsA1497), 重庆市基础与前沿研究计划基金项目(cstc2016jcyjA0308)资助

作者简介: 周思寒, 1994年生, 重庆理工大学现代光电检测技术与仪器重点实验室硕士研究生 e-mail: 631694566@qq.com

* 通讯联系人 e-mail: tangbin@cqut.edu.cn

等^[5]运用直接标准化对湿土的高光谱进行校正;翁海勇等^[6]利用直接标准化算法进行图像校正,以此研究柑橘溃疡病高光谱模型普适性差、预测精度低的问题。张晓羽等^[7]提出基于稳定竞争自适应重加权采样的无标模型传递方法,提高谷物脂肪含量与蛋白质含量的近红外光谱预测精度;徐惠荣等^[8]采用直接标准化算法,基于平均光谱差值进行模型转换,降低了皇冠梨糖度的预测均方差根误差;有报道^[9]采用 FIR 算法实现土壤的近红外模型在不同测试条件下的传递;Qin 等^[10]研究了烟叶总糖的均相粉末模型转移和混合模型方法;有报道提出 SSCW (screening stable and consistent wavelength) 波长优化方法,筛选出对仪器而言光谱信号一致性好且稳定的波长,以实现近红外光谱模型的无标样传递;贾一飞等^[11]提出斜率截距校正结合正交信号校正,实现近红外定量模型传递。上述方法能够较好地实现拉曼光谱、高光谱、近红外光谱的模型传递,并提高了对应的分析精度。

解决模型传递的有效方法是将仪器或数据标准化。针对现有的光谱标准化方法中,鲜有对紫外-可见吸收光谱的应用研究,且在近红外光谱标准化中,未考虑实验过程中产生的噪声对特征峰偏移的影响,以及不同型号的仪器对同一样本数据测得的量程制定范围的不同。本工作对水质检测仪器的标准化工作开展研究,采用基于 EWMA-PCA 的归一化算法处理水质光谱数据,对提高模型传递的利用率,提高水质检测精度,以及水质检测仪器的标准化工作具有十分重要的意义。

1 实验部分

1.1 样品及光谱数据采集

实验采用三种典型的应用于水质光谱检测的光谱仪进行样品的检测,分别为:滨松 C10082CAH 光谱仪,海洋 Maya 2000Pro 光谱仪以及奥谱天成 ATP2000 光谱仪,其中滨松 C10082CAH 光谱仪,测量波段 200~800 nm,共测量 2 048 个点;海洋 Maya2000Pro 光谱仪,测量波段 200~1 100 nm,共测量 2 068 个点;奥谱天成 ATP2000 光谱仪,测量波段 200~1 100 nm,共测量 2 048 个点水质检测系统如图 1 所示。为了验证算法的有效性,实验除光谱仪不同外,其余实验环境及试验设备均相同:采用相同的光源(型号为 Ocean optics DH-2000-BAL),设计相同的光路(光学支架型号为 Ocean optics ACH-CUV-VA R),光程为 9.5 cm,采用海洋抗紫外光纤(型号为 Ocean optics Inc S/N OOS-004354-05)。实验所用 8 组样本溶液,均为用化学法配备的不同浓度的邻苯二甲酸氢钾溶液,并以其作为标准浓度的化学耗氧量(COD)溶液以供实验测量使用,其浓度分别为 20, 40, 60, 120, 160, 200, 400 和 800 $\text{mg} \cdot \text{L}^{-1}$ 。

每个样本经光谱仪采集软件多次测量后取平均。图 2(a)为滨松 C10082CAH 光谱仪所测光谱图,在 20~800 $\text{mg} \cdot \text{L}^{-1}$ 范围内,第一个特征峰吸光度范围在 2.4~3.25 之间,波段范围在 230~255 nm 之间,第二个特征峰吸光度范围在 0.4~3.1 之间,波段在 280 nm 附近;图 2(b)为海洋 Maya2000Pro 光谱仪所测光谱图,在 20~800 $\text{mg} \cdot \text{L}^{-1}$ 范围内,

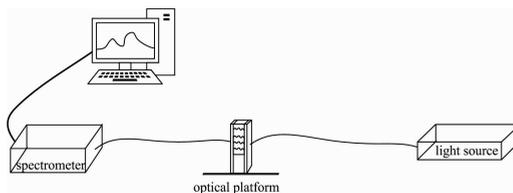


图 1 紫外-可见光谱法水质检测系统框图

Fig. 1 UV-Visible spectrum water quality detection system block diagram

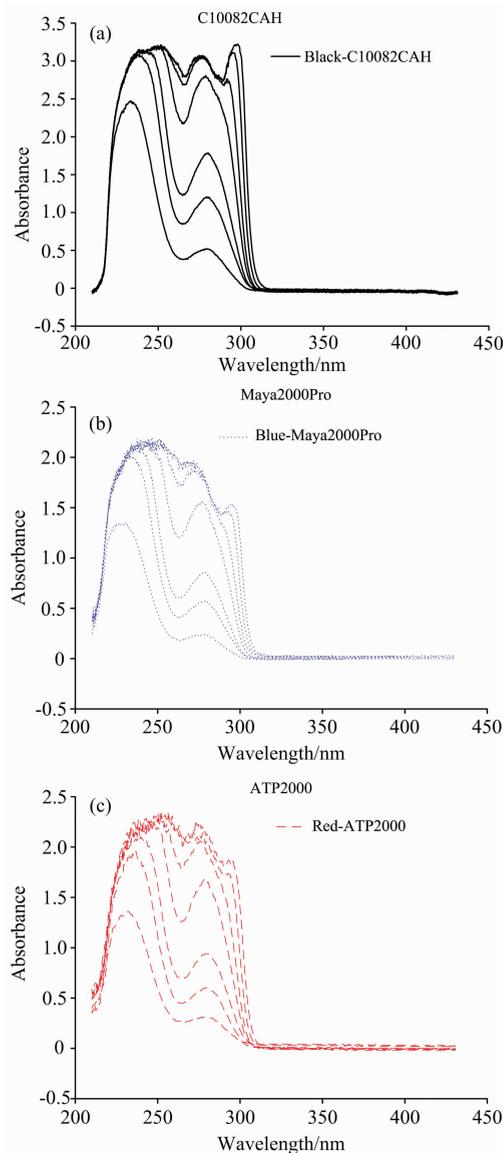


图 2 三个光谱仪下测得的 8 个标准样本溶液的光谱图
(a): 滨松 C10082CAH 光谱仪; (b): 海洋 Maya200Pro 光谱仪;
(c): 奥谱天成 ATP2000 光谱仪

Fig. 2 Spectra of 8 standard sample solutions measured under three spectrometers

(a) Hamamatsu C10082CAH spectrometer;
(b): Ocean Maya 2000Pro spectrometer;
(c): Optosky ATP2000 spectrometer

第一个特征峰吸光度范围在 1.3~2.2 之间, 波段范围在 235~260 nm, 第二个特征峰吸光度范围在 0.2~2 之间, 波段范围在 270~300 nm 之间; 图 2(c) 为奥谱天成 ATP2000 光谱仪所测光谱图, 在 20~800 mg·L⁻¹ 范围内, 第一个特征峰吸光度范围在 1.3~2.4 之间, 波段范围在 230~255 nm 之间, 第二个特征峰吸光度范围在 0.3~2.3 之间, 波段范围在 275~300 nm 之间。

1.2 EWMA-PCA 标准化算法原理

首先利用 EWMA 对样本水质光谱数据滤波, 并对数据进行降维提取主成分; 再利用直接标准化算法对源机与目标机两组数据进行矩阵变换, 最后进行 Z-score 标准化与归一化数据处理。

EWMA(exponentially weighted moving-average) 是一种指数加权平均移动算法。该算法可平滑短期波动, 保留波形长期发展趋势, 引入权重因子对数据进行最大似然估计, 寻找以较高概率产生观察紫外-可见水质光谱数据的系统发生树, 最大概率情况复原理论紫外-可见水质光谱数据。

$$\text{EWMA}(N) = \lambda Y(N) + (1 - \lambda)\text{EWMA}(N - 1) \quad (1)$$

$$i = 1, 2, \dots, n$$

式(1)中, EWMA(N) 为波段点 N 处的光谱数据估计值; Y(N) 为波段点 N 处的光谱数据的测量值; λ 为权重因子, 0 < λ < 1, 且随着波长点 N 的数目增加而指数递减; n 为所测量的波长点 N 的个数总和。

主成分分析(principal components analysis, PCA) 是一种常见的用于特征提取的多元分析方法, 其对水质光谱数据 A_{n×m} 进行降维, 从而输出含 k 个特征维度的光谱矩阵 A'_{n×m}。

考虑到实验所采取的光谱仪采样点非同步设置, 采用三次多项式插值以获得相同波长点的吸光度值。波长范围选用三台光谱仪测量范围的交集。设三台光谱仪的测量波段范围分别为 L₁=[b₁, b₂], L₂=[m₁, m₂], L₃=[a₁, a₂]。比较范围值, 若 b₁ < m₁ < a₁, b₂ > m₂ > a₂, 则波长测量范围取 [b₁, a₂], 即在 [b₁, a₂] 波长范围内进行插值处理, 以取得相同波长点的吸光度值。直接标准化原理: 将源机插值后的矩阵定义为 A_m, 目标机插值后得到的矩阵定义为 A_t, 通过转换矩阵 F 建立起 A_m 与 A_t 两组矩阵的联系。

$$\mathbf{A}_m = \mathbf{A}_t \mathbf{F} \quad (2)$$

式(2)中 A_t⁺ 为 A_t 的广义逆矩阵, 转移矩阵 F 由此可得: F = A_t⁺ × A_m。目标机矩阵 A_t 通过转移矩阵 F, 转化为 A_m, 即源机矩阵。由此可得, 当用目标仪器去测量一个未知样本溶液, 所得的光谱数据矩阵 A_{unknown}, 亦可通过转移矩阵 F, 转化为源仪器对该未知样本溶液的光谱数据矩阵 A_{std}, 理论上分析, 源仪器对该未知样本溶液所测得的光谱矩阵 A'_{std} 应与 A_{std} 相同, 见式(3)

$$\mathbf{A}'_{\text{std}} = \mathbf{A}_{\text{std}} = \mathbf{A}_{\text{unknown}} \mathbf{F} \quad (3)$$

Z-score 标准化就是将直接标准化后的光谱数据中每个特征平均值变为 0, 标准值变为 1, 使超出取值范围的吸光度值回归于正常值即 A* [见式(4)], 其中 μ 为所有吸光度数据的均值, σ 为所有吸光度数据的标准差, σ =

$$\sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - \mu)^2}$$

$$A^* = \frac{A - \mu}{\sigma} \quad (4)$$

$$A^* = \frac{A^* - A^*_{\min}}{A^*_{\max} - A^*_{\min}} \quad (5)$$

Z-score 标准化之后, 进行数据归一化处理, 使吸光度值映射到 [0-1] 之间。式(5)中 A_{max}^{*} 为吸光度的最大值, A_{min}^{*} 为吸光度的最小值。该做法有利于水质光谱数据的初始化进行, 可简化计算, 缩小量值, 提高水质光谱数据的表现能力。该法在后续模型传递过程中, 可以加快梯度下降的求解速度, 提升模型的收敛速度。

2 结果与讨论

实验中 3 个光谱仪测得的八组待测邻苯二甲酸氢钾溶液的水质光谱数据, 图 3(a) 对应 20 mg·L⁻¹ 的样本溶液, 图 3(b) 对应 40 mg·L⁻¹ 的样本溶液, 图 3(c) 对应 60 mg·L⁻¹ 的样本溶液, 图 3(d) 对应 120 mg·L⁻¹ 的样本溶液, 图 3(e) 对应 160 mg·L⁻¹ 的样本溶液, 图 3(f) 对应 200 mg·L⁻¹ 的样本溶液, 图 3(g) 对应化学耗氧量 COD 400 mg·L⁻¹ 的样本溶液, 图 3(h) 对应 800 mg·L⁻¹ 的样本溶液。

从整体波形趋势来看, 对于 8 组样本溶液, 三种光谱仪测得的原始光谱数据在 210~300 nm 范围内吸光度比较高, 且在此范围内波形变化趋势明显: 对于浓度 160 mg·L⁻¹ 及其以下的样本溶液中, 可见两个明显的特征峰, 波段在 250 与 280 nm 左右的范围内, 对于浓度在 400 mg·L⁻¹ 以上的样本溶液, 清晰可见, 在 300 nm 处新增第三特征峰(对于浓度在 200 mg·L⁻¹ 的溶液, Maya2000Pro 光谱仪与 ATP2000 光谱仪并无明显吸收峰, 分析原因可能是由于三台光谱仪的仪器状态如检测精度不同等造成的); 在 315~430 nm 范围内吸光度值变化幅度较小。对同一样本溶液而言, 三种光谱仪仍有不同: 对于第一特征峰, C10082CAH 光谱仪测得的吸光度值可达到 2.4~3.25; 但 Maya2000Pro 光谱仪与 ATP2000 光谱仪所测得的第一特征峰的吸光度值只在 1.3~2.2 与 1.3~2.4, 且对于同一样本溶液吸光度值差异更明显, 如图 3(a) COD 20 mg·L⁻¹ 的样本溶液, C10082CAH 光谱仪测得的吸光度可达到 2.4, 但 Maya2000Pro 光谱仪与 ATP2000 光谱仪测得的吸光度值仅可达到 1.3; 对于第二特征峰, C10082CAH 测得到的吸光度值可达到 0.4~3.1, Maya2000Pro 光谱仪与 ATP2000 光谱仪测得的吸光度值分别在 0.2~2 与 0.3~2.3, 且对应波段亦有不同。同时差异性对于同一样本溶液吸光度值更明显, 如图 3(c) COD 60 mg·L⁻¹, C10082CAH 光谱仪测得的吸光度值远大于 1.5, 但 Maya2000Pro 光谱仪与 ATP2000 光谱仪测得的吸光度值远小于 1, 且 Maya2000Pro 光谱仪与 ATP2000 光谱仪测得的吸光度值也有差距。正是由于对于同一样本溶液吸光度值差异较大, 且出现波形趋势不同的情况[如图 3(f) COD 200 mg·L⁻¹], 对于单一样本数据的准确度而言, 需要对水质光谱数据进行标准化及归一化, 从而实现对仪器标准化, 才能为后续稳定数据处理提供依据, 实现模型传递。

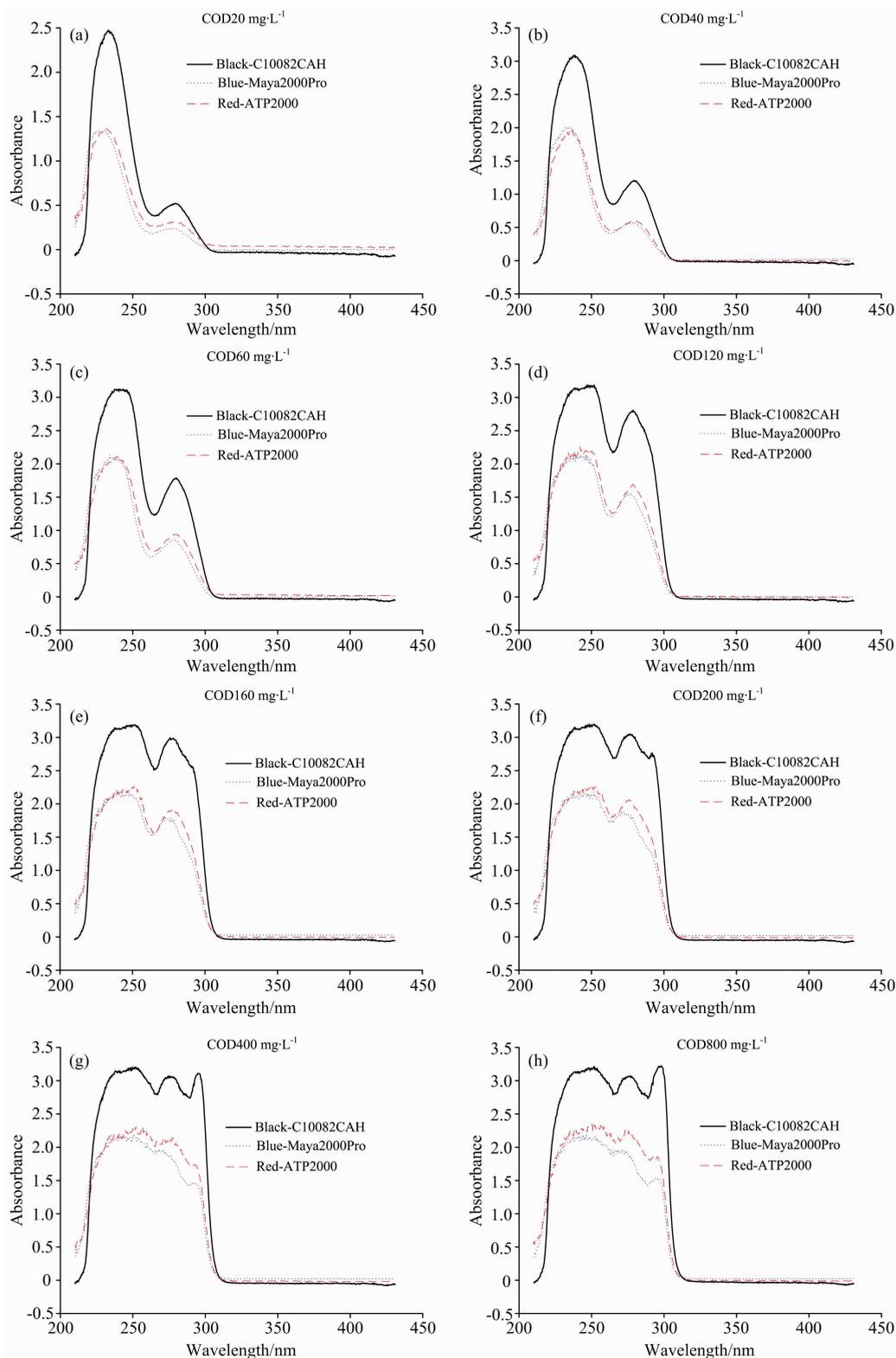


图 3 八组邻苯二甲酸氢钾溶液的水质光谱数据

(a): COD $20 \text{ mg} \cdot \text{L}^{-1}$; (b): COD $40 \text{ mg} \cdot \text{L}^{-1}$; (c): COD $60 \text{ mg} \cdot \text{L}^{-1}$; (d): COD $120 \text{ mg} \cdot \text{L}^{-1}$;
 (e): COD $160 \text{ mg} \cdot \text{L}^{-1}$; (f): COD $200 \text{ mg} \cdot \text{L}^{-1}$; (g): COD $400 \text{ mg} \cdot \text{L}^{-1}$; (h): COD $800 \text{ mg} \cdot \text{L}^{-1}$

Fig. 3 Water quality spectral data of 8 groups of potassium hydrogen phthalate solution

(a): COD $20 \text{ mg} \cdot \text{L}^{-1}$; (b): COD $40 \text{ mg} \cdot \text{L}^{-1}$; (c): COD $60 \text{ mg} \cdot \text{L}^{-1}$; (d): COD $120 \text{ mg} \cdot \text{L}^{-1}$;
 (e): COD $160 \text{ mg} \cdot \text{L}^{-1}$; (f): COD $200 \text{ mg} \cdot \text{L}^{-1}$; (g): COD $400 \text{ mg} \cdot \text{L}^{-1}$; (h): COD $801 \text{ mg} \cdot \text{L}^{-1}$

标准化的评价指标采用传统意义上的相关系数与方差进行比较。同时,由于第一特征峰特征明显,具有能够代表此段待标准化的水质吸收光谱数据的特性,因此提出第一特征峰偏移量的评价指标(以下均称波峰偏移量, Peak offset, P_o),用以评价该方法的优劣性。对于源机和目标机所测得的同一水质样本溶液,由于实验过程中易产生噪声,对于实验所测得的水质光谱数据的第一特征峰的峰值对应波段的判断往往会存在一个区间值,即由于噪声的干扰,极大波峰不止一个,而在去噪的过程中,有时候亦会去除真实的波峰值,而由其邻近或较远的波段的吸光度的值代替,因此,在数据处理过程中,保证特征峰不偏移或偏移较小往往显得很重要。

波峰偏移量 P_o 定义如式(6)

$$P_o = \left| 1 - \frac{\Delta_m}{\Delta_i} \right| \quad (6)$$

式(6)中, Δ_m 为源机数据标准化后的波峰比值 $\Delta_m = \frac{\lambda'_m}{\lambda_m}$, λ_m 为源机测得的原始数据的第一特征波峰对应的波段值, λ'_m 为源机测得的数据标准化处理后第一特征峰对应的波段值; Δ_i 为目标机数据标准化后的波峰比值, $\Delta_i = \frac{\lambda'_i}{\lambda_i}$, λ_i 为目标机测得的原始数据的第一特征波峰对应的波段值, λ'_i 为目标机测得的数据标准化处理后第一特征峰对应的波段值。 P_o 为该算法处理下的波峰偏移量,若 $P_o = 0$,从理论分析上,该算法在标准化的同时未改变光谱数据波形;若远大于 0,则该算法已使波形产生改变,即标准化效果较差。

实验结果分别如图 4 所示,实验结果统计如表 1,表 2 和表 3 所示,其中图 4(a) 对应组 1: 此组实验中, C10082CAH 光谱仪为源机, Maya2000Pro 光谱仪为目标机;图 4(b) 对应组 2: 此组实验中, C10082CAH 光谱仪为源机, ATP2000 光谱仪为目标机;图 4(c) 对应组 3: 此组实验中, Maya2000Pro 光谱仪为源机, ATP2000 光谱仪为目标机。

为了更直观展示实验统计结果,将表 1, 2, 3 绘制成图 5。从表 1, 2, 3 可以看出, 3 组仪器对应的八组样本溶液的紫外-可见吸收光谱数据标准化指标评价中,除表 1 中 $20 \text{ mg} \cdot \text{L}^{-1}$ 相关系数保持不变外,表 2, 表 3 浓度分布从 $20 \sim 800 \text{ mg} \cdot \text{L}^{-1}$, 标准化数据的相关系数较原始数据的相关系数均增加,从图 5(a) 中可明显看出, 标准化后的值高于标准化之前的值,即两曲线的相关系数提高,即该算法标准化使得曲线的相似度提高;从方差指标来看,浓度分布从 $20 \sim 800 \text{ mg} \cdot \text{L}^{-1}$, 8 组样本溶液的水质数据, 标准化后较标准化前方差大幅降低,从图 5(b) 中可明显看出, 表明该算法标准化使得两曲线间方差减小,则两曲线对应波段点数据波动减小,即标准化效果越好;从波峰偏移量来分析, 8 组样本溶液在三组对比实验中,波峰偏移量均约为 1,从图 5(c) 中可明显看出, 对比组实验 1 中,与 1 之间的误差在 $0.0028\% \sim 0.0736\%$,对比组 2,与 1 之间的误差在 $0.0005\% \sim 0.054\%$,对比组 3,与 1 之间的误差在 $0.004\% \sim 0.045\%$,误差较小,说明该标准化算法在进行曲线标准化时,未牺牲

曲线特征。经计算,采用基于 EWMA-PCA 的归一化算法标准化后,相关系数可达到 99.5765% ,方差可达到 0.0823% ,波峰偏移量可降低至 0.0005% ,以上三个数据均表明该算法在 C10082CAH 光谱仪与 Maya2000Pro 光谱仪、C10082CAH 光谱仪与 ATP2000 光谱仪、Maya2000Pro 光谱仪与 ATP2000 光谱仪三组仪器及数据标准化过程中呈良性发展,该标准化算法有效,且表现较好。

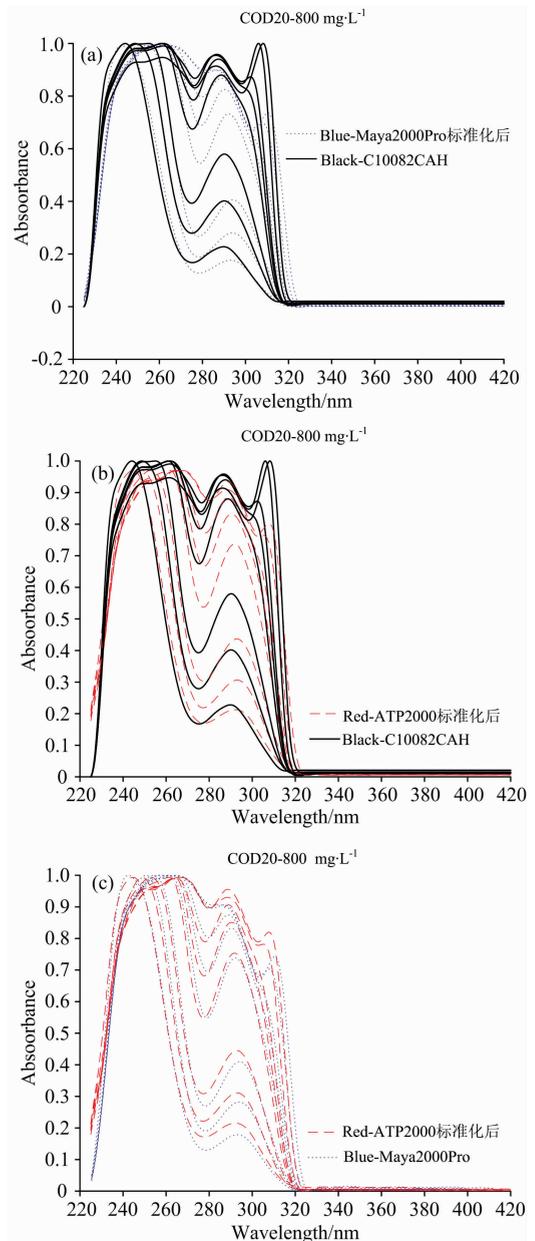


图 4 三组对比实验测得各浓度水质吸收光谱标准化数据
(a): 对应组 1; (b): 对应组 2; (c): 对应组

Fig. 4 Three sets of comparison experiments to determine the normalized data of water absorption spectra of each concentration

(a): Comparison 1; (b): Comparison 2; (c): Comparison 3

表 1 C10082CAH 光谱仪与 Maya2000Pro 光谱仪测得水质吸收光谱数据标准化评价

Table 1 Standardized evaluation of water absorption spectrum data measured by C10082CAH and Maya2000Pro spectrometer

Concentration /(mg · L ⁻¹)	Correlation Coefficient		Variance		Peak offset/%
	Original	Standardized	Original	Standardized	
20	0.994 229	0.994 229	0.098 594	0.001 016	0.010 8
40	0.990 752	0.990 766	0.131 140	0.001 903	0.035 4
60	0.986 116	0.986 128	0.161 782	0.003 592	0.002 8
120	0.989 171	0.989 173	0.239 228	0.004 028	0.097 7
160	0.991 622	0.991 623	0.252 841	0.003 540	0.006 6
200	0.991 255	0.991 256	0.256 969	0.003 659	0.073 6
400	0.985 524	0.985 524	0.286 155	0.005 789	0.009 5
800	0.982 170	0.982 170	0.305 703	0.006 357	0.070 9

表 2 C10082CAH 光谱仪与 ATP2000 光谱仪测得水质吸收光谱数据标准化评价

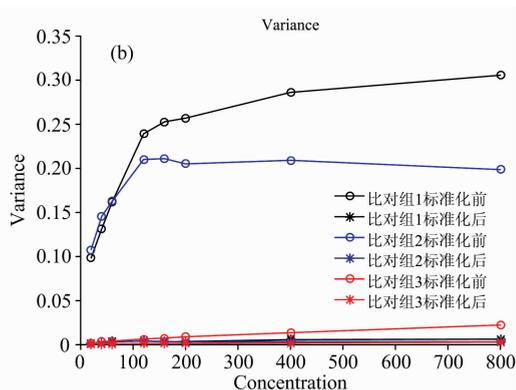
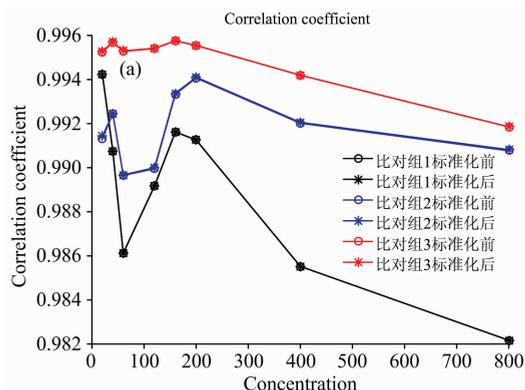
Table 2 Standardized evaluation of water absorption spectrum data measured by C10082CAH spectrometer and ATP2000 spectrometer

Concentration /(mg · L ⁻¹)	Correlation Coefficient		Variance		Peak offset/%
	Original	Standardized	Original	Standardized	
20	0.991 321	0.991 437	0.107 919	0.001 462	0.010 6
40	0.992 446	0.992 460	0.145 516	0.001 786	0.031 7
60	0.989 650	0.989 674	0.162 702	0.003 006	0.011 4
120	0.989 949	0.990 001	0.210 102	0.004 373	0.054 0
160	0.993 332	0.993 381	0.210 997	0.003 653	0.029 0
200	0.994 065	0.994 107	0.205 069	0.003 262	0.029 0
400	0.992 017	0.992 058	0.209 195	0.003 576	0.032 3
800	0.990 767	0.990 820	0.198 635	0.003 291	0.000 5

表 3 Maya2000Pro 光谱仪与 ATP2000 光谱仪测得水质吸收光谱数据标准化评价

Table 3 Standardized evaluation of water absorption spectrum data measured by Maya2000Pro and ATP2000 spectrometer

Concentration /(mg · L ⁻¹)	Correlation Coefficient		Variance		Peak offset/%
	Original	Standardized	Original	Standardized	
20	0.995 236	0.995 267	0.001 574	0.000 841	0.021 5
40	0.995 680	0.995 707	0.003 461	0.000 823	0.003 7
60	0.995 278	0.995 304	0.004 107	0.001 005	0.014 2
120	0.995 401	0.995 418	0.006 540	0.001 301	0.043 7
160	0.995 752	0.995 765	0.007 556	0.001 338	0.022 4
200	0.995 540	0.995 550	0.009 044	0.001 494	0.044 6
400	0.994 192	0.994 200	0.013 650	0.002 087	0.022 8
800	0.991 849	0.991 857	0.022 398	0.003 049	0.003 7



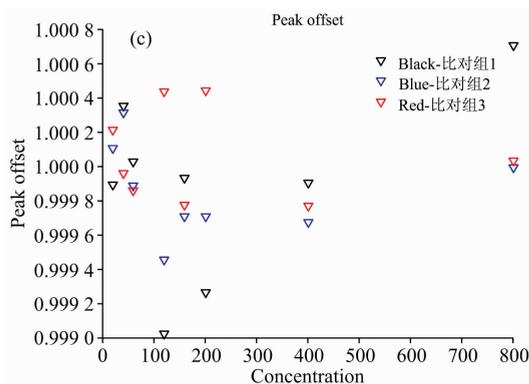


图 5 三组对比实验测得各浓度水质吸收光谱标准化评价

(a): 相关系数; (b): 方差; (c): 峰的偏移量

Fig. 5 Standardized evaluation of water absorption spectra of each concentration measured by three sets of comparison experiments

(a): Correlation coefficient; (b): Variance; (c): Peak offset

References

- [1] Zhao H, Hou D, Huang P, et al. *Water, Air & Soil Pollution*, 2014, 225(11): 2183.
- [2] TANG Bin, ZHAO Jing-xiao, WEI Biao, et al(汤斌, 赵敬晓, 魏彪, 等). *China Environmental Science(中国环境科学)*, 2015, 35(2): 478.
- [3] SHI Yun-ying, LI Jing-yan, CHU Xiao-li(史云颖, 李敬岩, 褚小立). *Chinese Journal of Analytical Chemistry(分析化学)*, 2019, 47(4): 479.
- [4] Dong X, Dong J, Li Y, et al. *Computers and Electronics in Agriculture*, 2019, 156: 669.
- [5] CHEN Yi-yun, QI Tian-ci, HUANG Ying-jing, et al(陈奕云, 齐天赐, 黄颖菁, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2017, 33(6): 107.
- [6] Haiyong W, Jingwen L, Haiyan C, et al. *Sensors and Actuators B: Chemical*, 2018, 275: 50.
- [7] Pasquini, Celio. *Analytica Chimica Acta*, 2018, 1026: 8.
- [8] XU Hui-rong, LI Qing-jing(徐惠荣, 李青菁). *Transactions of the Chinese Society for Agricultural Machinery(农业机械学报)*, 2017, (9): 312.
- [9] Qin Y, Gong H. *Infrared Physics & Technology*, 2016, 77: 239.
- [10] NI Li-jun, HAN Ming-yue, ZHANG Li-guo, et al(倪力军, 韩明月, 张立国, 等). *Chinese Journal of Analytical Chemistry(分析化学)*, 2018, 46(10): 163.
- [11] JIA Yi-fei, ZHANG Ying-ying, XU Bing, et al(贾一飞, 张盈盈, 徐冰, 等). *China Journal of Chinese Materia Medica(中国中药杂志)*, 2017, 42(12): 2298.

3 结 论

通过八组溶液, 3 组光谱仪的对比实验, 探讨了基于 EWMA-PCA 的归一化算法在水质光谱数据标准化中的应用研究, 并提出相关系数、方差、波峰偏移量等指标, 用以评价标准化结果。研究的主要结论是: 基于 EWMA-PCA 的归一化算法在标准化过程中, 控制波峰波段特征不变的同时很大程度上提高波形相似度, 降低数据波动, 减小两波形间方差; 且将该算法应用于 3 台不同的光谱仪中, 效果显著, 证明该算法的有效性及其适应性。

Application Research of Normalization Algorithm Based on EWMA-PCA in Standardization of Water Quality Spectral Data

ZHOU Si-han^{1,2}, HU Xin-yu^{1,2}, TANG Bin^{1,2,3*}, ZHAO Ming-fu^{1,2}, LI Feng-xiao^{1,2}, WANG Ren-jie^{1,2}, XIAO Qi-sen^{1,2}, XIAO Yu^{1,2}

1. Key Laboratory of Modern Optoelectronic Detection Technology and Instrument, Chongqing University of Technology, Chongqing 400054, China
2. Chongqing Key Laboratory of Optical Fiber Sensing and Photoelectric Detection, Chongqing University of Technology, Chongqing 400054, China
3. Key Laboratory of Optoelectronic Technology and Systems of Ministry of Education, Chongqing University, Chongqing 400044, China

Abstract Model transfer is important for solving the inconsistency of measurement signals due to differences in sample and instrument response functions, and an effective method for solving model transfer is the instrument or data standardization. For the existing spectral standardization methods, there are few applied research on UV-visible absorption spectrum, and the UV-visible spectroscopy water quality multi-parameter detection spectrum detection unit has inconsistent resolution, accuracy, and response range, which is difficult to perform between different instruments. The problem of the comparison of test data and the fitting of multi-parameter data, this paper proposes to use the EWMA-PCA normalization algorithm to achieve the model transfer of UV-visible water quality spectra on different instruments' (Exponentially Weighted Moving-Average) is an exponentially weighted average moving algorithm for finding phylogenetic trees with a high probability of generating UV-Visible water spectral data and recovering theoretical UV-visible water quality spectral data with maximum probability. The UV-Vis spectral characteristics are not lost or offset, reducing the impact of data processing on the UV-visible water spectral data. In this experiment, different concentrations of potassium hydrogen phthalate solution were used to compare and test the Japanese Hamamatsu C10082CAH spectrometer, the US Ocean Optics Maya2000Pro spectrometer and the Xiamen Optosky ATP2000 spectrometer. In the comparison group 1, the source machine Hamamatsu C10082CAH spectrometer and the target machine ocean Maya2000Pro spectrometer were selected, the comparison group 2 selected the source machine Hamamatsu C10082CAH spectrometer and the target machine Optosky ATP2000 spectrometer, and the comparison group 3 selected the source machine marine Maya2000Pro spectrometer and the target machine Optosky ATP2000 spectrometer. The experimental results of three sets of experiments show that the algorithm can be applied to different ratiometric spectrometers. After the EWMA-PCA normalization algorithm is used to standardize the water absorption spectrum data, the correlation coefficient reaches 99.576 5%, and the variance reaches 0.082 3%, and the peak offset can be reduced to 0.000 5%. Based on the EWMA-PCA normalized spectral normalization algorithm, it has wide adaptability, less need to transfer samples, and high transmission precision. The research results are widely used in spectroscopy water quality testing instruments. It has important theoretical guiding significance and reference value for engineering application.

Keywords Ultraviolet water spectroscopy; EWMA-PCA; Standardization; Normalization; Model transfer

(Received Oct. 16, 2019; accepted Feb. 5, 2020)

* Corresponding author