

拉曼光谱结合偏最小二乘法定量分析水中聚乙烯和聚苯乙烯两种微塑料

丁建明^{1,2} 王鑫¹ 张容玲¹ 周利园¹ 张天龙¹ 汤宏胜^{*1} 李华^{*1,3}

¹(西北大学化学与材料科学学院, 合成与天然功能分子化学教育部重点实验室, 西安 710127)

²(兵器工业卫生研究所, 西安 710065) ³(西安石油大学化学化工学院, 西安 710065)

摘 要 微塑料(MPs)作为一种水环境的新兴污染物,具有极性结构、粒径小(通常小于 5 mm)、比表面积大、稳定性好和生物降解难等特点,能够对水生生物的正常生理活动产生不利影响,并在包括人类在内的生物体内积聚,因此亟需建立水环境中 MPs 的快速且准确的定量分析方法。本研究将拉曼光谱结合偏最小二乘(Partial least squares, PLS)法用于实际水样中聚乙烯(Polyethylene, PE)和聚苯乙烯(Polystyrene, PS)的快速准确的定量分析。首先,制备了 33 个含有不同浓度 MPs 的模拟水样,并采集其拉曼光谱;其次,考察了 6 种光谱预处理方法(归一化、多元散射校正、标准正态变换、一阶导数、二阶导数和小波变换)对 PLS 校正模型预测性能的影响,并分别采用协同区间偏最小二乘算法(Synergy interval partial least squares, SiPLS)、变量重要性投影(Variable importance in projection, VIP)和互信息(Mutual information, MI) 3 种变量选择方法对 PLS 校正模型的输入变量进行选择与优化,采用留一法对 PLS 校正模型的预测能力进行考察与验证。在最优的光谱预处理、变量选择、输入变量和潜变量等条件下,建立了基于蒸馏水的小波变换-偏最小二乘(WT-PLS)校正模型,并对实际水样中 PE 和 PS 的含量进行了预测(PE 和 PS 的预测相关系数(R_p^2)分别为 0.9540 和 0.8472,预测误差(Error_p)分别为 0.0690 和 0.1126);建立了基于混合水样的 MI-PLS 校正模型,对实际水样具有最佳的预测性能(PE 和 PS 的 R_p^2 分别为 0.9776 和 0.9755, Error_p 分别为 0.0360 和 0.0392)。本方法为实际水样中 MPs 以及其它有机污染物的定量分析提供了新思路与新方法。

关键词 微塑料; 偏最小二乘法; 拉曼光谱; 化学计量学

塑料作为一种重要的有机合成高分子材料,具有轻质、多功能性、成本效益高及耐用等特性,在人类社会以及国民经济的发展过程中起着非常重要的推动作用。近年来,塑料的产量急剧攀升,塑料污染与全球气候变化、生物多样性锐减和臭氧破坏等已并列成为全球重大环境问题。为了有效应对这一危机,我国于 2021 年 9 月印发了《“十四五”塑料污染治理行动方案》,进一步完善了塑料污染全链条治理体系。然而,大量的塑料残留物被释放到环境中,经过相互作用、机械磨损^[1]、紫外线(UV)分解^[2]、光氧化和生物降解等方法分解成微塑料(MPs)。MPs 是指粒径 ≤ 5 mm 的塑料颗粒和碎片,作为一种在环境和有机体中积聚的新兴污染物,广泛分布于水体、陆地和大气环境中,MPs 污染已发展成为全球性的环境问题。MPs 可以大量吸附水环境中持久性有机污染物、重金属、抗生素、细菌、真菌以及病毒等,进而形成复合污染物^[3],其中,聚乙烯(Polyethylene, PE)和聚苯乙烯(Polystyrene, PS)是两种典型的 MPs,对多环芳烃的吸附能力较强^[4]。MPs 能够通过多种途径进入生物体内,并在人体组织中积累而引发物理和生物效应^[5],进而对心血管系统^[6]、消化系统^[5]和呼吸系统^[7]等产生严重危害。Prata 等^[8]利用尼罗红(NR)染色方法对不同品牌瓶装水中 MPs 的存在情况进行了评估,并借助 FTIR 技术证实瓶装水样品中存在 PS、PE 和聚酰胺(Polyamide, PA),其中 PE 较常见。更值得关注的是,MPs 在物理、化学和生物等作用下可以逐渐破碎形成粒径更小的纳米级塑料颗粒,更容易进入生物细胞内部并积累在细胞溶酶体中,进而产生比塑料更大的危害^[9]。MPs 来源与其化学组分密切相关,由于来源及形成条件差异,MPs 中所

2024-06-14 收稿; 2024-09-28 接受

国家自然科学基金项目(Nos. 22173071, 22073074)资助。

* E-mail: tanghongsheng@nwu.edu.cn; huali@nwu.edu.cn

含元素种类及其形态存在显著差异,进而呈现不同的毒理学和生物学特性。因此,水环境中 MPs 的定量分析是实现 MPs 溯源、风险评估与降解处理等的重要前提。

目前,水环境中 MPs 的表征主要以定性分析为主,借助目视法、显微镜计数法和光谱法等能够对 MPs 进行直接观察和计数,并统计其尺寸、形态、颜色和颗粒数等指标。目视法具有检测过程简单、成本低和化学危害小等优点,能够通过肉眼或光学显微镜对 MPs 进行初步分类和计数。然而,MPs 的尺寸不断减小以及环境样品中大量有机或无机等杂质的干扰,导致目视法的准确性和检测效率降低^[10],因此,目视法仅适合作为 MPs 分析的辅助方法。红外光谱法作为一种反映样品中化学键和官能团信息的分子振动光谱技术,具有分析速度快、对样本无破坏和操作简单等优势,已广泛应用于 MPs 的定性检测和成分分析。在国家标准(GB/T 40146—2021)^[11]中,红外光谱法被用于化妆品中塑料珠的定性分析,可识别粒径大于 20 μm 的 MPs。辽宁省地方标准(DB21/T 2751—2017)^[12]利用傅立叶变换显微红外光谱法对海水中 0.5~5.0 mm 的 MPs 进行定性分析,通过显微镜下计数的方式进行定量分析。

拉曼光谱(Raman spectroscopy)为水环境中 MPs 的定量分析提供了强有力的技术手段。与红外光谱相比,拉曼光谱具有更低的空间分辨率(0.1~1.0 μm)^[13-14],同时对非极性官能团更敏感,能够更准确地识别具有非极性对称键的化合物,如芳香键、C—H 和 C=C 键^[15]。拉曼光谱对水的干扰不敏感,能够确定聚合物以及 MPs 的类型、颗粒数量、尺寸、形状以及浓度分布^[16],在海洋和淡水^[17]等不同水环境中的 MPs 分析方面具广阔的发展前景。由于 MPs 化学组成多样且复杂,拉曼光谱信号中不仅包含目标组分的特征信息,通常还存在谱峰重叠、基线漂移以及复杂基质等的干扰信号,基于标准曲线法的单变量定量分析方法无法获得令人满意的分析结果。偏最小二乘法(Partial least squares, PLS)作为一种经典的多元校正方法,具有分析精度高、模型参数少和运算速度快等优势,已广泛应用于生物医学^[18]、环境监测^[19]和科技考古^[20]等领域。Zhao 等^[21]利用透射拉曼光谱结合 PLS 建立了双氯芬酸钠片中活性药物成分(Active pharmaceutical ingredient, API)的定量分析方法,其相对误差小于 4%,相对标准偏差为 0.49%。由于拉曼光谱中包含大量无效信息、噪声和杂散光等,其呈现出高维性和复杂性,变量选择是提升 PLS 校正模型预测能力的一种重要的技术手段。目前,常用的变量选择方法包括协同区间偏最小二乘算法(Synergy interval partial least squares, SiPLS)、无信息变量消除法(Uninformative variable elimination, UVE)^[22]、变量重要性投影(Variable importance in projection, VIP)、遗传算法(Genetic algorithm, GA)^[23]以及互信息(Mutual information, MI)等。Ma 等^[24]基于不同混合光谱预处理和变量选择相结合的策略建立了适用于含油污泥中 4 种多环芳烃(Phe、Flt、Flu 和 Nap)定量分析的 PLS 校正模型,获得了较好的预测结果。周围等^[25]建立了 UVE 结合 PLS 模型的柴油纯度检测方法,其预测均方根误差(RMSEP)可达到 0.0239,预测相关系数(R_p^2)可达到 0.9977。由此可见,光谱预处理和变量选择是提升 PLS 校正模型预测能力的一种有效手段。

基于拉曼光谱结合 PLS 在复杂体系定量分析方面具有良好的分析效果,然而尚未见将该方法用于水中 MPs 定量分析的研究报道。本研究分别以蒸馏水、雨水、河水及其混合水样为研究对象,建立了基于拉曼光谱结合 PLS 的实际水样中 PE 和 PS 的快速且准确的定量分析方法。首先,制备了 33 个含有不同浓度 MPs 的模拟水样,并采集其拉曼光谱;其次,考察不同预处理方法(归一化(Nor)、多元散射校正(MSC)、标准正态变换(SNV)、一阶导数(D1st)、二阶导数(D2nd)和小波变换(WT))对 PLS 校正模型预测性能的影响,分别利用 VIP、SiPLS 和 MI 这 3 种变量选择方法对 PLS 校正模型输入变量进行选择与优化,并采用留一法考察 PLS 校正模型的预测能力。在最优的光谱预处理、变量选择、输入变量和潜变量等条件下,建立了基于拉曼光谱结合 PLS 的实际水样中 PE 和 PS 的定量分析方法。

1 实验部分

1.1 仪器与试剂

QEpro6500 型激光拉曼光谱系统(美国 OceanOptics 公司);数控超声波清洗器(昆山市超声仪器有限公司)。

蒸馏水和无水乙醇(广东光华科技股份有限公司);PE 和 PS(东莞市特塑朗胶原料有限公司)。实

际水样包括陕西省西安市西北大学长安校区和太白校区的雨水、陕西省西安市西安理工大学曲江校区的雨水以及陕西省内沔河、灞河和护城河的河水。

1.2 实验方法

分别采集陕西省西安市西北大学长安校区和太白校区以及西安理工大学曲江校区的雨水样品。在采集雨水时,利用烧杯收集雨水样品,严格按照四周无遮挡,最大程度降低外界环境因素对雨水样本的潜在干扰,从而确保雨水样本的纯净度和代表性。河水样品的采样地址为陕西省内沔河、灞河以及护城河,利用水样采集器采集河面以下大约 1 m 深处河水样品,并除去肉眼可见的明显生物杂质。上述雨水和河水样品分别采用定性滤纸和 0.45 μm 滤膜过滤,待用。

为了建立实际水样中 PE 和 PS 的定量分析方法,共制备了 33 个含有不同浓度 PE 和 PS 的模拟水样(如表 1 所示)。具体制备过程如下:称取适量的 PE 和 PS(10 μm)于比色管中,加入 7.00 mL 无水乙醇和 3.00 mL 蒸馏水,超声 20 min,保证 MPs 颗粒均匀分散。其中,1#~24#样品是蒸馏水、河水和雨水的单一基质水样品,25#~33#样品是由蒸馏水、河水和雨水中任意 2 种或全部 3 种单一基质水样混合而成。为了获得高质量的水样品拉曼光谱,采用醋酸-硝酸(CN-CA)混合纤维素酯水性滤膜对水样品进行过滤处理,待用。

表 1 水样中聚乙烯(PE)和聚苯乙烯(PS)的含量信息
Table 1 Reference values of concentrations of polyethylene (PE) and polystyrene (PS) in water samples

序号 No.	基质 Matrix	PE/ (mg/mL)	PS/ (mg/mL)	序号 No.	基质 Matrix	PE/ (mg/mL)	PS/ (mg/mL)
1#*	H	0.15	0.23	18#	Z	0.52	0.64
2#*	Y	0.18	0.25	19#*	H	0.58	0.67
3#	Z	0.20	0.28	20#*	Y	0.63	0.72
4#*	H	0.22	0.30	21#	Z	0.69	0.77
5#*	Y	0.24	0.31	22#*	H	0.73	0.83
6#	Z	0.27	0.36	23#*	Y	0.76	0.88
7#*	H	0.29	0.42	24#	Z	0.84	0.90
8#*	Y	0.34	0.43	25#	Y+Z	0.14	0.15
9#	Z	0.37	0.47	26#	H+Y+Z	0.20	0.24
10#*	H	0.38	0.49	27#	H+Z	0.24	0.33
11#*	Y	0.39	0.50	28#	Y+Z	0.31	0.39
12#	Z	0.42	0.51	29#	H+Y+Z	0.34	0.46
13#*	H	0.44	0.52	30#	H+Z	0.40	0.48
14#*	Y	0.45	0.54	31#	Y+Z	0.44	0.54
15#	Z	0.46	0.56	32#	H+Y+Z	0.47	0.57
16#*	H	0.49	0.57	33#	H+Z	0.56	0.69
17#*	Y	0.50	0.58				

注(Note): *, 测试集(Prediction set); Z, 蒸馏水(Distilled water); Y, 雨水(Rainwater); H, 河水(River water)。

1.3 拉曼光谱采集

利用 QEpro6500 型激光拉曼光谱系统(美国 OceanOptics 公司)采集实际水样的 Raman 光谱。该系统由半导体激光器(785 nm)、光纤(RPBY)、光谱仪(QEpro6500)及光谱处理软件(OceanView)等组成,Raman 光谱采集范围为 0~2000 cm⁻¹,光谱仪分辨率为 4 cm⁻¹,激光器功率为 0~1500 mW 范围内可调。在采集拉曼光谱时,将富集有 MPs 的 CN-CA 混合纤维素酯水性滤膜置于样品台上,为了防止杂散光对拉曼光谱的影响,拉曼光谱采集在暗箱中进行。激光输出功率设置为 0.900 W,探测距离设置为 10.8 mm,积分时间为 20 s。为了减少实验误差,每个样本随机选择 6 个不同位置采集拉曼光谱,环境温度为室温,以空气为背景检测并将其扣除。

采用两种方式划分校正集和测试集: (1) 以 8 个蒸馏水样本为校正集, 8 个河水和 8 个雨水样本为测试集, 考察基于蒸馏水体系下的 PLS 校正模型对实际水样的预测能力; (2) 在 8 个蒸馏水的基础上添加了 9 个混合水样作为校正集, 8 个河水样本和 8 个雨水样本为测试集, 考察混合水样体系下 PLS 校正模型对实际水样的预测能力。

1.4 偏最小二乘法校正模型的建立及评价

PLS 是一种经典多元回归分析方法, 通常以全波段光谱或特征谱峰的所有数据点作为自变量, 待分析组分浓度作为因变量, 建立光谱强度和组分浓度之间的相关关系, 能够在一定程度上克服传统单变量模型的缺陷^[24]。拉曼光谱结合 PLS 定量分析实际水样中 PE 和 PS 的方法如图 1 所示, 具体流程如下: (1) 采集 33 个水样的拉曼光谱, 并采用主成分分析(Principal component analysis, PCA)剔除异常光谱; (2) 按照 1.2 节描述的两种方法划分校正集和测试集; (3) 以全波段拉曼光谱作为输入变量, 建立初始化的 PLS 校正模型; (4) 为了建立预测性能优异的 PLS 校正模型, 考察了光谱预处理、变量选择、模型参数等对 PLS 校正模型预测性能的影响。对于基于蒸馏水的 PLS 校正模型探究了 Nor、MSC、SNV、D1st、D2nd 和 WT 6 种光谱预处理方法对水样 Raman 光谱的预处理效果; 对于基于混合水样的 PLS 校正模型, 重点探究了 MI 的 PLS 校正模型输入变量选择与优化方法。采用留一交叉验证方法考察了所建立 PLS 校正模型的预测性能。PLS 校正模型的评价指标通常包括决定系数(R^2)、均方根误差(RMSE)、平均相对误差(MRE)和平均绝对误差(MAE)。 R^2 越大, 说明 PLS 校正模型的预测性能越好; RMSE、MRE 和 MAE 的值越小, 对应校正模型的预测性能越好。然而, 对于最佳 PLS 校正对应的评价指标不一致的情况, 引入了一个综合指标 Error, 用于评估 PLS 校正模型的整体预测性能, 通常认为, Error 越小, 回归模型的预测性能越好。Error 的计算公式如下:

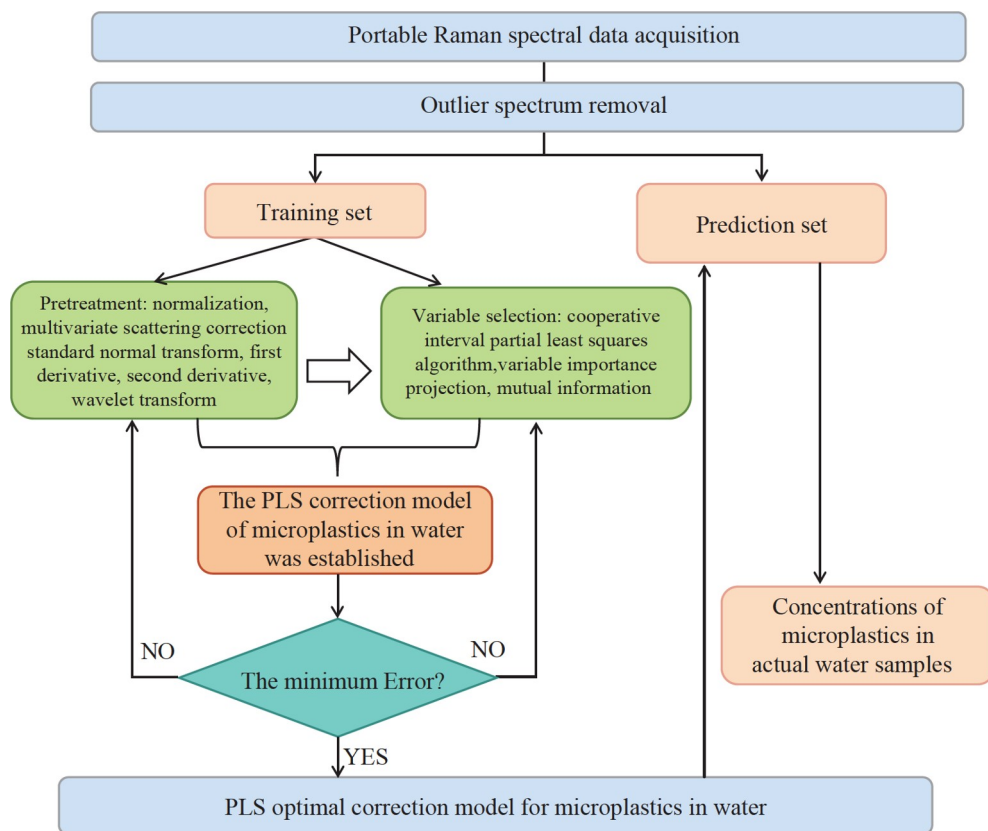


图1 基于拉曼光谱结合偏最小二乘法(PLS)的水中 MPs 定量分析方法示意图

Fig.1 Schematic diagram of quantitative analysis method of microplastics in water based on Raman spectroscopy combined with partial least squares (PLS)

$$\text{Error} = \frac{(1 - R^2) + \text{RMSE} + \text{MRE} + \text{MAE}}{4}$$

其中, RMSE 和 MAE 的大小取决于样品的浓度, 而 R^2 和 MRE 通常小于 1。通常, Error 越小, R^2 越大, RMSE、MRE 和 MAE 越小, 结果越好。

2 结果与讨论

2.1 拉曼光谱解析

含 PS 和 PE 的实际水样的拉曼光谱如图 2A 所示, PS 和 PE 的特征峰光谱如图 2B 所示。PS 的主要特征拉曼峰位于 620、792、998、1029、1179 和 1600 cm^{-1} 处。其中, 620 和 792 cm^{-1} 处为 C—C—C 键振动^[26], 998 cm^{-1} 处为 C—C 苯环呼吸振动, 1029 cm^{-1} 处为 C—H 键的摇摆振动^[27], 1179 cm^{-1} 处为苯环变形振动, 1600 cm^{-1} 处为 C—C 伸缩振动^[26]。PE 的主要拉曼特征峰位于 1061、1128、1293、1416、1437 和 1460 cm^{-1} 处。其中, 1061 和 1128 cm^{-1} 处为 C—C 键伸缩振动, 1293 cm^{-1} 处为 CH_2 弯曲振动, 1416 和 1437 cm^{-1} 处为 CH_2 弯曲振动, 1460 cm^{-1} 处为两个 CH_2 的面内摇摆振动。此外, 由不同基质水样的拉曼光谱(图 2A)可见, 由于样品之间浓度差异较小, 拉曼光谱图重叠严重。因此, 基于标准曲线法的单变量定量分析方法无法实现水中 PE 和 PS 的精准定量分析。

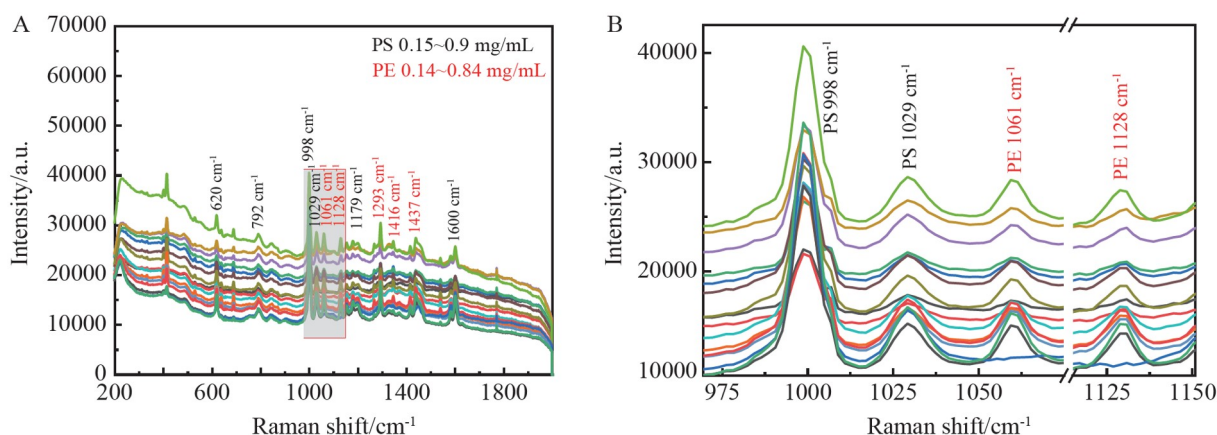


图2 (A) 含 PS 和 PE 的实际水样的拉曼光谱图; (B) PS 和 PE 的特征峰光谱图

Fig.2 (A) Raman spectra of actual water samples containing PS and PE; (B) PS and PE characteristic peak spectra

2.2 基于蒸馏水的小波变换-偏最小二乘(WT-PLS)校正模型

以 8 个蒸馏水样本为校正集、8 个河水样本和 8 个雨水样本为测试集, 探究基于蒸馏水体系下的 PLS 校正模型对实际水样的预测能力。重点考察了 Nor、MSC、SNV、D1st、D2nd 和 WT 这 6 种方法对水样拉曼光谱的预处理效果, 并对比了基于 6 种预处理方法的 PLS 校正模型的预测性能(图 3 和表 2)。以 R^2 和 Error 作为评价指标, 通过留一交叉验证法对导数法和 WT 等方法参数进行优化, 以获得更准确的定量分析结果。PLS 模型参数-潜变量参数范围均为 1~20, D1st 和 D2nd 的平滑点范围为 1~25(奇数), WT 的小波基函数 db2~db14, coif1~coif5, 分解层数 1~15。以 Error 最小为原则对 PLS 校正模型进行优化, 选择最佳的光谱预处理方法。

由图 3 可知, 与基于原始光谱的 PLS 校正模型相比, 基于 6 种不同光谱预处理的 PLS 校正模型预测性能均得到了提升, 其中, 基于 WT 的 PLS 校正模型的预测能力更优异。对于 PE 组分(图 3A), 当小波基函数为 db10、分解层数为 5、PLS 模型潜变量数为 14 时, PLS 校正模型表现出优异的预测性能($R_p^2 = 0.9540$, 预测平均相对误差(MRE_p) = 0.1310, 预测平均绝对误差(MAE_p) = 0.0404, $\text{RMSE}_p = 0.0587$, 预测误差(Error_p) = 0.0690); 对于 PS 组分(图 3B), 当小波基函数为 coif5、分解层数为 1、PLS 模型潜变量数为 14 时, PLS 校正模型表现出优异的预测性能($R_p^2 = 0.8472$, $\text{MRE}_p = 0.1262$, $\text{MAE}_p = 0.0672$, $\text{RMSE}_p = 0.1041$,

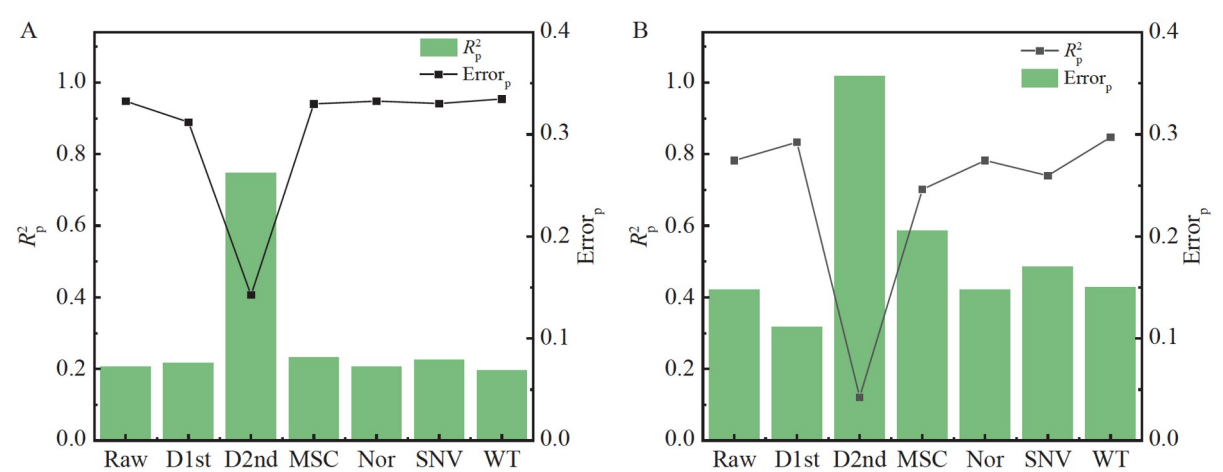


图3 基于不同光谱预处理方法的 PLS 校正模型对 PE 测试集(A)和 PS 测试集(B)的预测结果

Fig.3 Prediction results of (A) PE test set and (B) PS test set by PLS correction model based on different spectral preprocessing methods

R_p^2 : 预测相关系数(Predictive correlation coefficient); $Error_p$: 预测误差 (Prediction error); Raw, 原始数据(Raw data); D1st, 一阶导数(The first derivative); D2nd, 二阶导数(The second derivative); MSC, 多元散射校正(Multiplicative scatter correction); Nor, 归一化(Normalization); SNV, 标准正态变换(Standard normal variate), WT, 小波变换(Wavelet transform)

表 2 基于不同光谱预处理方法的 PLS 校正模型预测性能

Table 2 Predictive performance of PLS calibration model based on different spectral preprocessing methods

预处理方法 Preprocessing methods	PE		PS	
	R_p^2	$Error_p$	R_p^2	$Error_p$
Raw	0.9479	0.0727	0.7823	0.1481
Nor	0.9480	0.0727	0.7823	0.1481
MSC	0.9403	0.0817	0.7014	0.2056
SNV	0.9414	0.0794	0.7400	0.1703
D1st	0.8893	0.0759	0.8332	0.1111
D2nd	0.4066	0.2622	0.1215	0.3573
WT	0.9540	0.0690	0.8472	0.1126

$Error_p=0.1126$)。WT 在复杂水样品定量分析方面表现出色,能够有效克服多组分之间的相互影响。基于蒸馏水体系构建的 WT-PLS 校正模型可实现对实际水样中 PE 和 PS 的精确定量分析。

2.3 基于混合水样的互信息-偏最小二乘(MI-PLS)校正模型

为了建立稳定性好的 PLS 校正模型,分别以 8 个蒸馏水样本和 9 个混合水样本作为校正集、8 个河水样本和 8 个雨水样本为测试集,考察混合水样体系下 PLS 校正模型对实际水样的预测能力。变量选择是提升多元校正模型预测能力的重要策略之一,能够剔除与当前学习任务无关或相关性较低的冗余特征,利用特征重要性评价函数和搜索策略,获得近似最优的特征子集。针对复杂水环境体系 MPs 拉曼光谱的复杂性和高维度性等特征,采用 MI 理论方法进行特征选择^[28],将筛选后剩余有用特征进行 PLS 建模。MI 是衡量相关性的参数,可以反映特征变量之间的任意关系,有助于定量分析。

在利用 MI 进行变量选择时,先以变量数为 50 的间隔优化大致范围,再分别以变量数为 10 和 1 的间隔进行精细筛选,并考察变量数目对 PLS 校正模型预测性能的影响(图 4),以 $Error_p$ 最小为准则优化 MI 的变量数。随着变量数增加,决定系数 R_p^2 呈现先增大后降低的趋势, $Error_p$ 呈现先减小后增大的趋势。当变量数为 1010 时, MI-PLS 校正模型对 PE 表现出了最优的预测性能($R_p^2 = 0.9776$, $Error_p = 0.0360$)(图 4A); 当变量数为 351 时, MI-PLS 校正模型对 PS 表现出了最优的预测性能($R_p^2 = 0.9755$, $Error_p = 0.0392$)(图 4B)。为了进一步考察基于 MI-PLS 校正模型的预测能力,比较了基于 SiPLS 和 VIP 的 PLS 校正模型的预测性能(表 3)。与原始的 PLS 校正模型相比,基于不同变量选择的 PLS 校正模型均表现出优

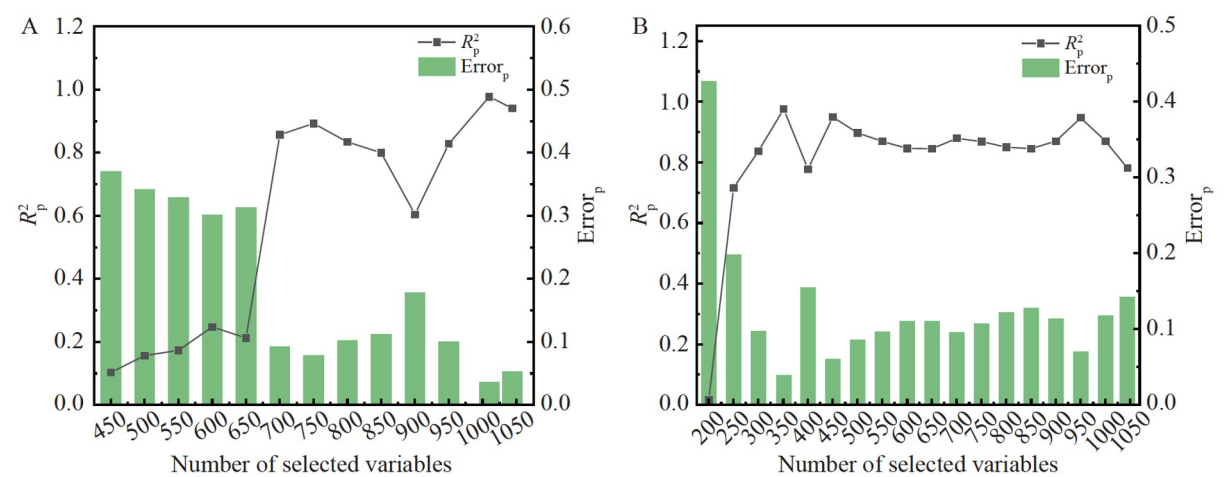


图4 互信息(MI)-PLS 校正模型对实际样品中(A)PE 和(B)PS 的预测结果
Fig.4 The prediction results of mutual information (MI)-PLS calibration model for (A) PE and (B) PS in real samples

表 3 基于不同变量选择方法的 PLS 校正模型预测性能
Table 3 Prediction results of PLS calibration model based on different variable selection methods

方法 Method	PE			PS		
	变量数 Number of variable	R_p^2	Error _p	变量数 Number of variable	R_p^2	Error _p
PLS	1044	0.9408	0.0529	1044	0.7810	0.1424
MI-PLS	1010	0.9776	0.0360	351	0.9755	0.0392
SiPLS-PLS	448	0.9678	0.0436	597	0.8639	0.1019
VIP-PLS	432	0.9661	0.0431	677	0.8810	0.0910

注(Notes): SiPLS, 协同区间偏最小二乘算法(Synergy interval partial least squares); VIP, 变量重要性投影(Variable importance in projection)。

异的预测性能,预测能力均得到提升;相比 SiPLS 和 VIP 两个变量方法,基于 MI 的 PLS 校正模型具有更优异的预测性能。

2.4 不同校正模型预测性能的考察

进一步考察了基于蒸馏水和混合水样的两种 PLS 校正模型对相同测试集的预测能力,采用 R^2 、MRE、MAE、RMSE 和 Error 这 5 个评价指标分别比较了两种 PLS 校正模型的预测结果(表 4 和图 5)。对于基于蒸馏水的 PLS 校正模型,相较于其它预处理方法,基于 WT 的 PLS 校正模型表现出优异的预测性能,然而对于实际水样中 PE 和 PS 的定量分析并未获得令人满意的分析结果。因此,针对复杂实际水样中 PE 和 PS 的定量分析,本研究以混合水样为校正集,通过更加复杂基质样品的 PLS 校正模型训练与优化,进一步利用 MI 方法进行 PLS 校正模型输入变量的选择与优化,筛选出更相关的特征变量。相较于其它方法,基于 MI-PLS 校正模型对实际水样中 PE 和 PS 的预测性能更优,对于 PE 的定量分析, R_p^2 达到

表 4 基于不同基质的 PLS 校正模型预测性能对比
Table 4 Comparison of prediction performance of PLS calibration model based on different matrices

基质 Matrix	校正模型 Calibration model	组分 Constituent	R_p^2	RMSE _p	MAE _p	MRE _p	Error _p
蒸馏水 Distilled water	WT-PLS	PE	0.9540	0.0587	0.0404	0.1310	0.0690
		PS	0.8472	0.1041	0.0672	0.1262	0.1126
混合水样 Mixed water sample	MI-PLS	PE	0.9776	0.0314	0.0231	0.0671	0.0360
		PS	0.9755	0.0365	0.0304	0.0654	0.0392

注(Notes): RMSE_p, 预测均方根误差(Root mean square error of prediction); MAE_p, 预测平均绝对误差(Mean absolute error of prediction); MRE_p, 预测平均相对误差(Average relative error of prediction)。

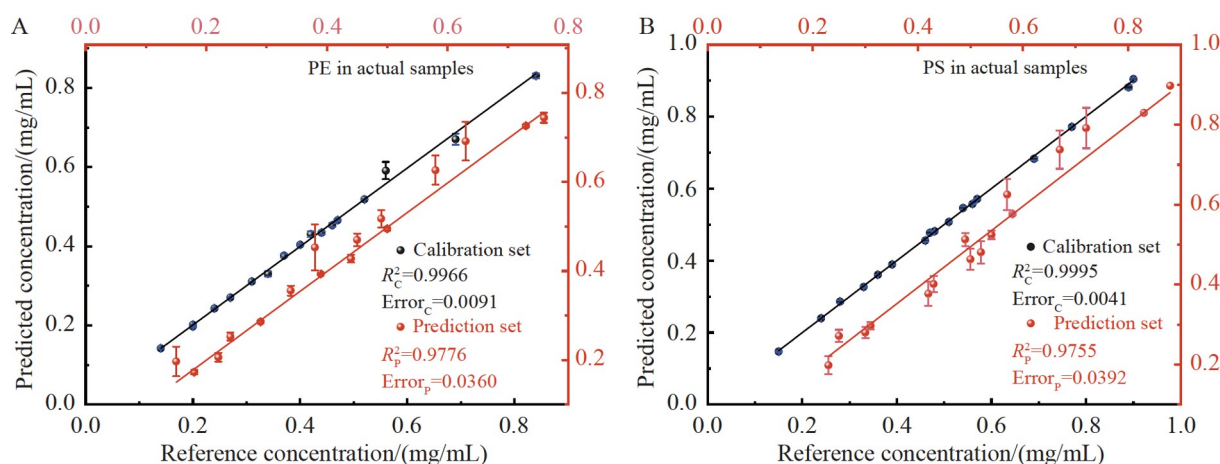


图5 MI-PLS 校正模型对实际样本中(A)PE 和(B)PS 的预测性能

Fig.5 Predictive performance of MI-PLS calibration model for (A) PE and (B) PS in actual samples

0.9776, Error_p 降低至 0.0360 (图 5A); 对于 PS 的定量分析, R_p^2 达到 0.9755, Error_p 降低至 0.0392 (图 5B)。综上,无论是校正集还是测试集,MI-PLS 校正模型对于 PE 和 PS 的定量分析均表现出较好的预测结果,预测值与参考值之间呈较好的相关性。当样品所在水环境更加复杂时,变量选择方法的引入十分必要,不仅可去除冗余变量,而且可以提升建模效率,提高定量分析的准确度。

3 结论

以蒸馏水、雨水和河水等水环境体系为研究对象,建立了拉曼光谱结合 PLS 用于实际水样中 PE 和 PS 的定量分析方法。采集并制备了 33 个含有不同浓度 MPs 的水样品,分别建立了基于蒸馏水和混合水样两种基质的 PLS 校正模型。在 R^2 、RMSE、MRE 和 MAE 4 个单一评价指标的基础上,引入了综合评价指标 Error。对于基于蒸馏水的 PLS 模型,WT-PLS 校正模型表现出优异的预测性能(PE 和 PS 的 R_p^2 分别为 0.9540 和 0.8472, Error_p 分别为 0.0690 和 0.1126)。然而,针对复杂实际水样中 PE 和 PS 的定量分析,MI-PLS 校正模型对实际水样中 PE 和 PS 具有更加优异的预测性能(PE 和 PS 的 R_p^2 分别为 0.9776 和 0.9755, Error_p 分别为 0.0360 和 0.0392)。因此,拉曼光谱结合 PLS 能够实现水环境中 MPs 的快速且准确的定量分析。

References

- [1] EKVALL M T, LUNDQVIST M, KELPSIENE E, SILEIKIS E, GUNNARSSON S B, CEDERVALL T. *Nanoscale Adv.*, 2019, 1(3): 1055-1061.
- [2] LAMBERT S, WAGNER M. *Chemosphere*, 2016, 161: 510-517.
- [3] KINIGOPOULOU V, PASHALIDIS I, KALDERIS D, ANASTOPOULOS I. *J. Mol. Liq.*, 2022, 350: 118580.
- [4] ZHU J H, LIU S Q, SHEN Y, WANG J, WANG H Q, ZHAN X H. *Environ. Pollut.*, 2022, 292: 118472.
- [5] BLACKBURN K, GREEN D. *AMBIO*, 2022, 51(3): 518-530.
- [6] LESLIE H A, VAN VELZEN M J M, BRANDSMA S H, VETHAAK A D, GARCIA-VALLEJO J J, LAMOREE M H. *Environ. Int.*, 2022, 163: 107199.
- [7] VASSE G F, MELGERT B N. *Eur. Respir. Rev.*, 2024, 33(172): 230226.
- [8] PRATA J C, DA COSTA J P, FERNANDES A J S, DA COSTA F M, DUARTE A C, SANTOST R. *Sci. Total Environ.*, 2021, 783: 146979.
- [9] GASPAR T R, CHI R J, PARROW M W, RINGWOOD A H. *Front. Mar. Sci.*, 2018, 5: 345.
- [10] RAN Tai-Shan, LIAO Hong-Kai, LONG Jian, LI Juan, YANG Guo-Mei, ZHAO Yu-Xin. *Plast. Sci. Technol.*, 2022, 50(7): 101-104.
冉泰山, 廖洪凯, 龙健, 李娟, 杨国梅, 赵雨鑫. *塑料科技*, 2022, 50(7): 101-104.
- [11] GB/T 40146-2021. Determination of Plastic Microbead in Cosmetics. National Standards of the People's Republic of China. 化妆品中塑料微珠的测定. 中华人民共和国国家标准. GB/T 40146-2021.

- [12] DB21T 2751–2017. Determination of Microplastics in Seawater-Micro-Fourier Transform Infrared Spectroscopy. Liaoning Provincial Standard.
海水中微塑料的测定 傅立叶变换显微红外光谱法. 辽宁省地方标准. DB21T 2751–2017.
- [13] ARAUJO C F, NOLASCO M M, RIBEIRO A M P, RIBEIRO-CLARO P J A. *Water Res.* , 2018, 142: 426-440.
- [14] KAPPLER A, FISCHER D, OBERBECKMANN S, SCHERNEWSKI G, LABRENZ M, EICHHORN K J, VOIT B. *Anal. Bioanal. Chem.* , 2016, 408(29): 8377-8391.
- [15] XUE Li-Dong, ZHANG Lin-Lin, YU Hai-Bin, ZHU Hong-Xia, YUAN Mao. *Environ. Monit. China*, 2022, 38(5): 9-17.
薛荔栋, 张霖琳, 于海斌, 朱红霞, 袁懋. 中国环境监测, 2022, 38(5): 9-17.
- [16] ENDERS K, LENZ R, STEDMON C A, NIELSEN T G. *Mar. Pollut. Bull.* , 2015, 100(1): 70-81.
- [17] CABERNARD L, ROSCHER L, LORENZ C, GERDTS G, PRIMPKE S. *Environ. Sci. Technol.* , 2018, 52(22): 13279-13288.
- [18] SHARMA A, KAUR I, DHEER D, DHEER D, NAGPAL M, KUMAR P, VENKATESH D N, PURI V, SINGH I. *Carbohydr. Polym.* , 2023, 308: 120448.
- [19] CHENG H. *Environ. Dev. Sustain.* , 2023, 25(7): 6249-6268.
- [20] LEMBERGE P, DE RAEDT I, JANSSENS K H, WEI F Z, VAN ESPEN P J. *J. Chemom.* , 2000, 14(5-6): 751-763.
- [21] ZHAO X J, WANG N, LIU Y T, JIANG Y F, SUN S N, WANG X H, YAO J, SHAN G Z. *J. Chemom.* , 2023, 37(2): e3462.
- [22] CENTNER V, MASSART D L, DENOORD O E, DE JONG S, VANDEGINSTE B M, STERNA C. *Anal. Chem.* , 1996, 68(21): 3851-3858.
- [23] LEARDI R. *J. Chemom.* , 2000, 14(5-6): 643-655.
- [24] MA C F, ZHAI L L, DING J M, LIU Y L, HU S F, ZHANG T L, TANG H S, LI H. *Spectrochim. Acta, Part A*, 2024, 310: 123953.
- [25] ZHOU Wei, LI An-Ji, LYU Wen-Min, YU Tie-Cheng, ZHAO Li-Juan, ZHAO Mei-Qi, YIN Ran. *Control Instrum. Chem. Ind.* , 2024, 51(4): 592-598.
周围, 李安吉, 吕文敏, 俞铁钺, 赵丽娟, 赵美琪, 尹冉. 化工自动化及仪表, 2024, 51(4): 592-598.
- [26] LEVERMORE J M, SMITH T E L, KELLY F J, WRIGHT S L. *Anal. Chem.* , 2020, 92(13): 8732-8740.
- [27] BERTOLDO M D, REYER A, MUSSO M. *Spectrochim. Acta, Part A*, 2018, 190: 433-441.
- [28] VERGARA J R, ESTÉVEZ P A. *Neural Comput. Appl.* , 2014, 24(1): 175-186.

Raman Spectroscopy Combined with Partial Least Squares for Quantitative Analysis of Two Kinds of Microplastics in Water Samples

DING Jian-Ming^{1,2}, WANG Xin¹, ZHANG Rong-Ling¹, ZHOU Li-Yuan¹, ZHANG Tian-Long¹,
TANG Hong-Sheng^{*1}, LI Hua^{*1,3}

¹(Key Laboratory of Synthetic and Natural Functional Molecular Chemistry, Ministry of Education, College of Chemistry & Materials Science, Northwest University, Xi'an 710127, China)

²(Institute for Hygiene of Ordnance Industry, Xi'an 710065, China)

³(College of Chemistry and Chemical Engineering, Xi'an Shiyou University, Xi'an 710065, China)

Abstract Microplastics (MPs) are emerging contaminants in aquatic environments characterized by their polar structure, small particle size (Typically less than 5 mm), large surface area, good stability, and resistance to biodegradation. They pose adverse effects on the normal physiological activities of aquatic organisms and can accumulate in biota, including humans. Therefore, there is an urgent need for rapid and accurate quantitative analysis of MPs in water environments. In this study, Raman spectroscopy combined with partial least squares (PLS) was employed for rapid and accurate quantitative analysis of polyethylene (PE) and polystyrene (PS) MPs in real water samples. Initially, 33 simulated water samples containing different concentrations of MPs were prepared, and their Raman spectra were collected. Six spectral preprocessing methods (Normalization, multiplicative scatter correction, standard normal variate transformation, first derivative, second derivative, and wavelet transform) were investigated for their impact on the predictive performance of PLS calibration models. Subsequently, three variable

selection methods including synergy interval partial least squares (SiPLS), variable importance in projection (VIP) and mutual information (MI) were employed to optimize the input variables of the PLS calibration model. The predictive capability of the PLS calibration model was evaluated and validated using leave-one-out cross-validation. Under the optimal conditions of spectral preprocessing, variable selection, input variables and latent variables, the wavelet transform-partial least squares (WT-PLS) calibration model based on distilled water was established, and the contents of PE and PS in real water samples were predicted with prediction correlation coefficients (R_p^2) of 0.9540 and 0.8472 for PE and PS, respectively, and prediction errors (Error_p) of 0.0690 and 0.1126, respectively. Furthermore, a mixed sample MI-PLS calibration model was developed, demonstrating the best predictive performance in real water samples (With R_p^2 values of 0.9776 and 0.9755 for PE and PS, respectively, and Error_p values of 0.0360 and 0.0392, respectively). This method provided a novel approach and new methodology for quantitative analysis of MPs and other organic pollutants in real water samples.

Keywords Microplastics; Partial least squares; Raman spectroscopy; Chemometrics

(Received 2024-06-14; accepted 2024-09-28)

Supported by the National Natural Science Foundation of China (Nos. 22173071, 22073074).