

基于红外光谱聚类分析的纳滤膜污染动态发展行为研究

李梦晨¹, 肖康², 黄霞^{1*}

1. 清华大学环境学院环境模拟与污染控制国家重点联合实验室, 北京 100084
2. 中国科学院大学资源与环境学院, 北京 100049

摘要 污水再生利用是解决水资源短缺问题的有效对策。纳滤技术由于能够生产高质量的再生水, 成为污水深度处理、再生利用的有效方法之一。然而, 在纳滤过程中存在复杂的、动态的膜污染现象, 会导致产水通量、产水质量下降等问题。研究膜污染动态发展的行为, 对于膜污染的分阶段针对性控制具有重要意义。有机物是污染层动态发展过程的重要指示性成分, 红外光谱是表征污染层发展过程中表面有机物官能团变化情况的重要手段。但由于红外光谱中峰的数量多, 系列样品之间峰强度的差别较小(尤其是当膜污染过程中的采样间隔较小时), 利用直观观察不易甄别不同样品间的谱图差异及其变化趋势, 在此水平上难以对膜污染阶段进行准确识别、对各阶段特征进行有说服力的分类概括。为探索膜污染的动态发展过程, 本研究将傅里叶变换红外光谱与统计学聚类分析相结合, 对膜污染过程中不同时间点的膜样本进行红外光谱分析, 再对红外光谱数据进行一系列预处理和系统聚类分析, 从而客观解读膜污染动态发展过程中系列样品红外光谱分阶段变化规律。考虑到类别间距离度量方法、红外吸收峰强度标准化、峰之间自相关性、峰与样本之间交互作用等因素的影响, 研究采用对应分析对红外数据进行预处理, 提取各样本在主要维度上的得分, 随后基于标准化欧式距离对各样本进行聚类。在为期一个月的城市污水深度处理纳滤试验过程中, 由于污染物在膜表面累积, 纳滤膜发生了较为严重的污染。通过对13个不同时间点的膜样本进行红外光谱聚类发现, 膜污染可清晰划分为如下阶段: 空白膜、阶段I(3 h~8 d)、阶段II(10~15 d)和阶段III(20~30 d)。采用红外聚类, 得到膜表面X射线光电子能谱(XPS)和三磷酸腺苷(ATP)含量分析等方法的交互验证。结果表明, 随着膜污染的发展, 膜表面有机物成分与共存微生物量发生协同变化, 各阶段大致特征为: 阶段I各类有机污染物初步覆盖, 微生物开始富集; 阶段II多糖类污染物比例上升, 微生物的富集趋于稳定; 阶段III整体污染趋于成熟, 有机污染物氢键特征更加明显。该研究通过对红外数据进行聚类分析, 能够灵敏地探测各红外图谱之间的差别, 有助于对红外光谱规律的深度挖掘, 为膜污染阶段的识别和划分提供了一种客观、自动、量化的辅助性方法, 并且有助于归纳出不同阶段的污染层特征, 可作为膜污染时序特征的侦查手段。此外, 除了膜污染的研究, 在材料、吸附等领域, 只要有一系列变化的红外光谱, 均可尝试采用红外光谱聚类分析方法, 获取基于红外特征的定类信息或分阶段规律。

关键词 红外光谱; 聚类分析; 纳滤膜污染; 动态发展

中图分类号: X131.2 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)02-0421-07

引言

污水再生利用是解决水资源短缺问题的有效对策。纳滤技术由于能够产生高质量的出水, 广泛应用于水处理领域^[1-2], 是污水再生利用的有效方法之一。但在应用过程中, 纳滤膜污染会导致膜通量和产水质量下降等问题, 限制了纳

滤工艺的广泛应用^[3-4]。膜污染是一个复杂的、动态的过程, 污水中的无机物、有机物和微生物均可能参与膜表面污染层的形成。污染层的动态发展涉及不同阶段, 膜污染的主要作用机理由初始阶段的污染物-膜之间相互作用, 逐渐演变为后续阶段的污染物-污染物之间相互作用^[5]。膜污染动态发展过程的研究对于解析膜污染的机理非常重要, 而目前系统、全面地解析纳滤膜污染动态发展过程的研究仍然较少。

收稿日期: 2018-02-10, 修订日期: 2018-06-24

基金项目: 国家自然科学基金项目(51678336)资助

作者简介: 李梦晨, 女, 1993年生, 清华大学环境学院环境模拟与污染控制国家重点联合实验室硕士研究生 e-mail: lmc1993@126.com

* 通讯联系人 e-mail: xhuang@tsinghua.edu.cn

Khan 等^[6]的研究发现,在膜污染的不同阶段,污染物组成比例变化明显:在初始阶段(数小时),膜表面的优势污染物为腐殖酸;在中间阶段(数天),优势污染物为多糖;在生物膜形成阶段(数周),优势污染物则为蛋白质。但在其研究中膜污染阶段的时间划分主要是通过直观判断,缺乏科学充分的划分依据。对膜污染阶段进行客观、自动、量化的划分与识别,对于纳滤膜污染动态过程的系统研究非常重要。

有机物是膜污染的重要成分,不仅来源于污水混合液,随膜污染过程在膜表面不断富集,而且可通过污染层中微生物的二次代谢发生成分和性质上的演变,这使得有机物成为污染层动态发展过程的重要指示性成分^[7-8]。红外光谱由于能够表征污染层表面有机物的官能团情况,并且具有灵敏度高、分辨率高、扫描速度快、对样品无破坏性等优点,是普遍使用的有机污染物分析手段^[6, 9-10]。在膜污染发展过程中,膜表面污染物组成的变化使得红外光谱发生变化,会导致新峰的出现,或一些峰被覆盖,从而能够进行污染层有机物成分动态变化的研究。但由于红外光谱中峰的数量多,系列样品之间峰强度的差异较小(尤其是当膜污染过程中的采样间隔较小时),利用直观观察不易甄别不同样品间的谱图差异及其变化趋势,在此水平上难以对膜污染阶段进行准确识别、对各阶段特征进行有说服力的分类概括。

目前已有研究将红外光谱和统计学聚类分析结合,利用聚类分析较为灵敏地探测出不同样品之间红外图谱的差别。该方法已在药品、食品等领域的分类和源谱的解析等方面有所应用^[11-12]。例如,徐永群等^[11]以赤芍的红外指纹图谱为依据,对来自 18 个产地的赤芍进行聚类分析,分类结果具有明确的物理意义,并且与地理位置有较明显的对应关系,可作为对赤芍药材质量评价的辅助方法。然而目前几乎没有研究将红外光谱和聚类分析相结合的方法应用于膜污染相关的研究领域。

本研究以傅里叶变换红外光谱的聚类分析为基础方法,以主成分分析、对应分析等多元统计学方法为辅助性优化手段,对纳滤膜污染层的系列红外光谱进行解析,从而挖掘膜污染的动态发展规律。数据处理过程中考虑到数据标准化、距离度量方法、峰之间的自相关性、样本之间的交互作用等可能对结果产生的影响,不断改进聚类分析方法,最终实现了利用该方法进行纳滤膜污染阶段的划分,从红外光谱的角度为膜污染发展研究提供了额外信息,从而为分阶段针对性的膜污染精细控制策略提供支持。除膜污染研究外,这种方法对于其他领域的系列红外光谱样品数据的解析也能具备一定的参考意义。

1 实验部分

1.1 膜污染试验

采用聚酰胺复合纳滤膜(DF30,北京碧水源)进行膜污染试验,膜孔径 0.61 nm,透水系数为 $(1.30 \sim 2.74) \times 10^{-6} \text{ m} \cdot \text{s}^{-1} \cdot \text{bar}^{-1}$ 。采用恒压错流过滤的模式在室温条件下运行纳滤装置,原水为某城市污水处理厂的微滤膜出水,总有机碳(total organic carbon, TOC)为 $3 \sim 4 \text{ mg} \cdot \text{L}^{-1}$,电导率

为 $500 \sim 700 \mu\text{S} \cdot \text{cm}^{-1}$ 。纳滤过程中跨膜压差维持为 0.5 MPa,膜表面错流速率为 $12 \text{ cm} \cdot \text{s}^{-1}$,初始膜通量为 $6 \text{ L} \cdot (\text{m}^2 \cdot \text{h})^{-1}$ 。通过 Darcy 定律计算膜过滤阻力

$$R = \frac{\Delta P}{\mu J} \quad (1)$$

式(1)中 R 为膜过滤阻力 (m^{-1}), ΔP 为跨膜压差 (Pa), μ 为水的粘度 ($\text{Pa} \cdot \text{s}$), J 为膜通量 ($\text{m} \cdot \text{s}^{-1}$)。

在 30 d 的运行过程中,先后采取 12 个不同时间点的污染膜样品,采样时间点分别为 3, 6 和 12 h, 及 1, 2, 4, 6, 8, 10, 15, 20 和 30 d。污染纳滤膜样品经过干燥后,进行红外光谱测定及其他理化、生化性质表征。

1.2 仪器及参数

膜样品的红外光谱测量采用美国 Thermo Nicolet 公司的 Nicolet 5700 衰减全反射傅里叶变换红外光谱(attenuated total reflectance-Fourier transform infrared spectroscopy, ATR-FTIR)仪器,分辨率为 4 cm^{-1} ,扫描范围 $400 \sim 4\,000 \text{ cm}^{-1}$ 。X 射线光电子能谱分析(X-ray photoelectron spectroscopy, XPS)采用美国 Thermo Fisher Escalab 250Xi 型电子能谱仪,用于测定污染层中的元素组成和键合情况^[4]。三磷酸腺苷(adenosine triphosphate, ATP)的荧光检测采用 BacTiter-Glo™ Microbial Cell Viability Assay (Promega)试剂盒,用于表征污染层中的生物学特性^[13]。

1.3 数据分析

红外吸收(峰强度)的数据呈现在以膜污染时间和红外波数张开的二维表格中,运用 SPSS 16.0 软件对数据进行聚类分析、主成分分析以及对应分析。

2 结果与讨论

2.1 膜污染概况

纳滤膜运行过程中膜过滤通量和阻力的变化情况如图 1 所示。膜通量和阻力是膜污染情况的外在直观体现。经过 30 d 的运行,膜通量下降了 33%,可见由于污染物在膜表面积,纳滤膜发生了较为严重的污染。在这 30 d 中,膜污染经历了由轻到重的过程,但是由于膜通量的下降速率及阻力的

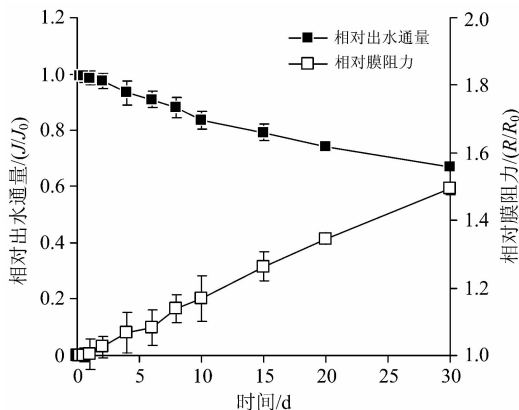


图 1 纳滤膜过滤通量和阻力随时间的变化

Fig. 1 The variation of permeate flux and resistance of nano-filtration (NF) membranes with operating time

上升速率没有明显的变化,仅根据图 1 曲线难以直接辨别膜污染的不同阶段,更是无从概括膜污染的分阶段特征。为此,红外光谱有望为污染层化学性质的变化提供重要参考信息。

2.2 污染膜的红外光谱

不同污染时间的纳滤膜样品的傅里叶变换红外光谱如图 2 所示。膜样品红外光谱的吸收峰主要集中在 $560 \sim 1760 \text{ cm}^{-1}$ 之间,可以看出污染膜样品与空白膜有明显区别, 1050 和 1740 cm^{-1} 附近的吸收峰差异尤其明显,很可能是空白膜上聚酰胺的特征峰^[14]。但是在不同的污染膜样品之间,红外图谱的差别则较为细微,难以通过肉眼观察准确区分。随着污染时间的延长, 1240 和 1490 cm^{-1} 附近的峰值存在一定的衰减趋势(与时间显著相关, Spearman 相关系数分别为 -0.896 和 -0.912),有可能对应腐殖酸的吸收峰^[15],但仍然很难据此区分膜污染的特征阶段。红外光谱信息的进一步挖掘,可通过聚类分析等多元统计分析方法进行尝试。

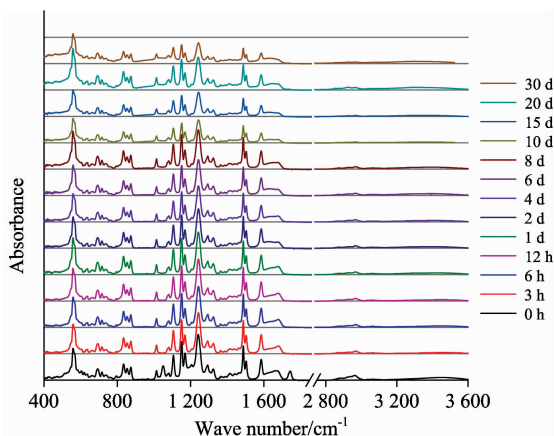


图 2 不同污染时间纳滤膜样品的红外光谱

Fig. 2 ATR-FTIR spectra of NF membranes fouled after different fouling time

2.3 红外光谱聚类分析结果

在存在多个样本和多个变量的情况下,聚类分析能够根据样本特点,将样本概括为若干类别,以样本与样本之间的距离(或相似性)作为聚类的依据,距离最近的优先聚为一类。本研究以不同时间的污染膜作为样本,以不同波数下的红外吸收作为变量,对红外数据进行聚类分析。

2.3.1 样本间距离的度量

欧式距离是度量样本之间距离的常用指标,第 i 个样本和第 j 个样本之间的欧式距离表达为

$$D_{ij(\text{Eu})} = \left[\sum_{k=1}^p (x_{ik} - x_{jk})^2 \right]^{1/2} \quad (2)$$

式(2)中 x_{ik} 是指第 i 个样本在第 k 个波数下的红外吸收,每个样本的红外图谱由 p 个波数来描述。式(2)表示的即为 p 维空间中两个点的欧式距离。本研究中,原始的红外数据共有 14 936 个波数值,即 p 为 14 936。利用红外的原始数据,首先计算在 14936 维空间中,不同时间的污染膜样品与空白样品(未污染膜)之间的欧式距离,初步查看膜污染红外特征随时间变化的大致规律,如图 3 所示。从图中可见,不同样

品之间表现出一定的变化趋势。除空白膜外, $3 \text{ h} \sim 8 \text{ d}$ 的膜样品较为接近, $10 \sim 30 \text{ d}$ 的样品较为接近,而 $8 \sim 10 \text{ d}$ 之间似乎存在膜污染阶段的大致分界线。从整体上看,欧式距离的阶段差异性,说明膜表面有机污染物的性质发生了变化。

由污染膜样品的原始红外图谱(图 2)可见,不同波数的特征峰吸光度差别较大,因此在计算欧式距离时各峰所占权重不同,影响计算结果。采用标准化欧式距离对各峰权重校正如式(3)

$$D_{ij(\text{Eu}^*)} = \left[\sum_{k=1}^p \left(\frac{x_{ik} - x_{jk}}{s_k} \right)^2 \right]^{1/2} \quad (3)$$

其中 s_k 是指第 k 个波数下红外吸收在各样本之间的标准差。标准化欧式距离在欧式距离的基础上消除了量纲,考虑了 p 维空间中各维数据的分布离散度不同对欧式距离计算的影响,其计算结果如图 3 所示。与欧式距离的曲线相比,仍然是 $3 \text{ h} \sim 8 \text{ d}$ 的样品较为接近, $10 \sim 30 \text{ d}$ 的样品较为接近,但其中 20 d 的样品与 10 和 15 d 的差距缩小,而与 30 d 样品的差距增大。这表明各峰权重对 20 和 30 d 样品的欧式距离计算结果有较大影响,利用标准化欧式距离能对此方面的影响进行修正。

2.3.2 主成分分析和对应分析对距离的优化

在红外图谱中一种有机物可能同时存在多个官能团的特征峰,例如 1650 和 1550 cm^{-1} 处分别为酰胺 I 和酰胺 II 的特征峰^[6],这两种官能团均存在于蛋白质中,因此当蛋白质的含量发生变化时,酰胺 I 和酰胺 II 的吸光度会出现同增或同减的情况,即红外图谱中部分峰与峰之间可能存在相关性,导致计算欧式距离时各维度之间不独立。为排除峰之间相关性对样本之间距离计算的干扰,有两种改进方法:一是将标准化欧式距离改进为马氏距离,用峰之间的协方差或相关系数对标准化欧式距离进行校正;二是进行压缩降维,采用主成分分析提取各峰之间的独立主成分,再根据主成分数值计算样本间的标准化欧式距离。在计算马氏距离的过程中,为保证相关系数矩阵满秩,变量数量应小于等于样本数量,由于纳滤膜样本共有 13 个(空白膜+12 个时间点的膜样品),所以能够纳入分析的峰最多为 13 个,远小于实际的特征峰数量,在计算过程中信息丢失较为严重。因此不考虑采用马氏距离的方法,而采用主成分欧式距离的方法。

主成分分析是一种有效的数据降维方法,能够将原始的多个变量线性组合成一组有效而独立的新变量(即一组主成分),仅用部分主成分就可高效率地保留原有变量的主要信息,而且由于主成分之间严格正交(独立),避免了自变量间相关性对欧式距离计算的干扰。从红外光谱提取总共 23 个特征吸收峰的吸光度值,计算峰之间相关系数矩阵的特征值,以特征值大于 1 作为主成分的遴选标准。最终提取出两个主成分,其方差累积贡献率达到 94.5%,损失的信息仅 5.5%。因此,利用主成分分析将 23 维的信息压缩为 2 维,这两个主成分就能概括原始数据的大部分信息内容,不存在严重的信息丢失。利用得到的主成分结果计算欧式距离如式(4)

$$D_{ij(\text{PC})} = \left[\sum_{k=1}^q (y_{ik} - y_{jk})^2 \right]^{1/2} \quad (4)$$

其中 q 为主成分数量, y_{ik} 为第 i 个样本在第 k 主成分上的标准化得分(由 SPSS 软件给出)。基于主成分的欧式距离计算结果如图 3 所示。从图中可见, 3 h~8 d 的膜样品之间的差异更加明显, 说明在此期间内, 膜上的有机污染物性质也存在一定变化。与普通的标准化欧式距离相比, 主成分欧式距离更灵敏地甄别出了该变化。

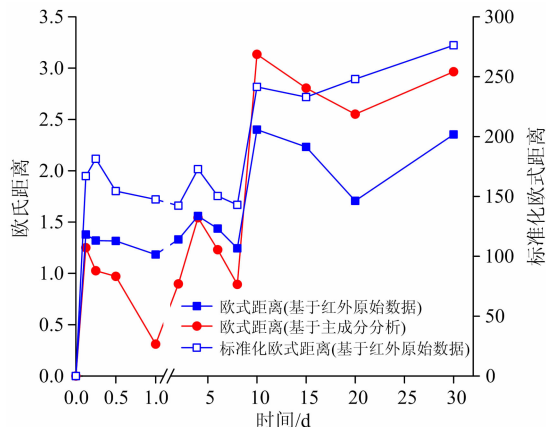


图 3 污染膜与空白膜之间的红外光谱欧式距离、标准化欧式距离和主成分分析欧式距离随时间的变化

Fig. 3 The Euclidean distance, normalized Euclidean distance and Euclidean distance based on principal component analysis between fouled membranes and virgin membrane based on ATR-FTIR spectra

除了峰之间的相关性, 由于本研究中的纳滤膜样本分别代表了不同污染时间的结果, 整体上反映了膜污染的进程, 因此不同样本间可能也存在相关性。对应分析在主成分分析的基础上, 能够同时反映变量和变量、样本和样本、以及变量和样本之间的关系。变量和样本之间的交互作用, 对标准化欧式距离计算过程中的标准差以及主成分提取过程中的相关系数矩阵均有潜在影响。利用从红外光谱提取的 23 个特征峰进行对应分析, 得到的结果如图 4(a) 所示。由此提取出的两个主要维度则校正了变量和样本之间的交互作用。两个维度共有 96% 的方差贡献率, 其中维度 1 贡献了 84%, 维度 2 贡献了 12%。图 4(b) 是根据各样本点的维度 1 得分和维度 2 得分绘制的各维得分随时间变化曲线。从空白膜至 30 d 膜样品, 维度 1 得分随时间逐渐降低, 这说明维度 1 得分可能与膜污染总体趋势有关, 维度 1 得分单调下降表明膜污染不断加重。维度 1 得分与膜通量之间的高度显著相关性(Pearson 相关系数 0.969, 显著性水平 0.000)也印证了这一点。维度 2 得分随时间的变化则大致体现为三个台阶: 3 h~8 d, 10~15 d, 20~30 d。这说明维度 2 的得分可能与膜污染阶段划分有关。利用维度 1 和维度 2 的得分, 可进而计算欧式距离如式(5)

$$D_{ij(\text{CA})} = \left[\sum_{k=1}^r (z_{ik} - z_{jk})^2 \right]^{1/2} \quad (5)$$

其中 r 为对应分析维度, z_{ik} 为第 i 个样本在第 k 维度上的标准化得分(由 SPSS 软件给出)。

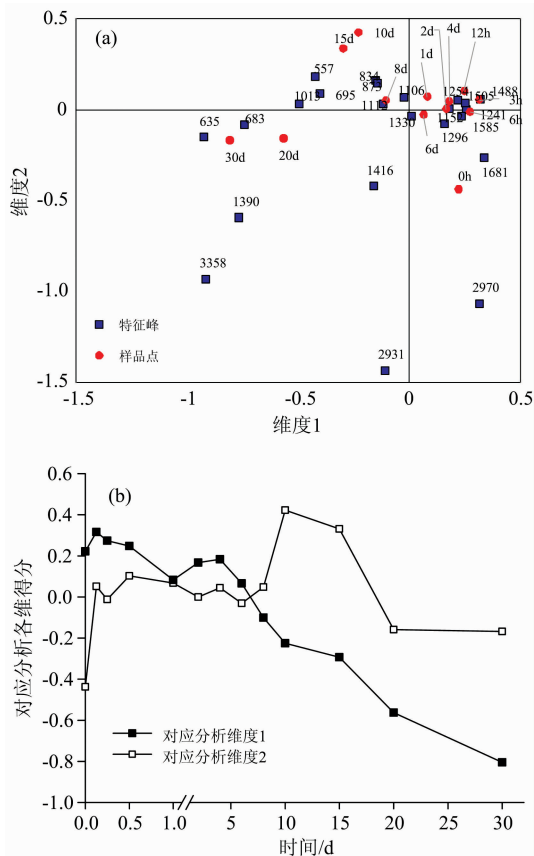


图 4 纳滤膜样本红外光谱对应分析结果

(a) 特征峰及样本点在各维度上的分布;

(b) 样本点各维得分随时间的变化

Fig. 4 The correspondence analysis of NF membranes based on ATR-FTIR spectra

(a): Distribution of characteristic peaks and sample points along different dimensions; (b): Variation of dimension scores of sample points with operating time

2.3.3 聚类分析结果及验证

根据对应分析的各维得分, 基于类间平均欧式距离对 13 个纳滤膜样品进行聚类分析, 结果如图 5 所示。系统聚类法的树状图清晰地展示了聚类过程。若取阈值为 2~3, 膜样本可分为四类: 空白膜, 阶段 I (3 h~8 d), 阶段 II (10~15 d), 阶段 III (20~30 d)。在此基础上复核红外图谱(图 2)可见, 阶段 I 可认为是初期污染, 主要体现为污染物对空白膜信号的显著覆盖(如 1 050 和 1 740 cm^{-1} 处的吸收峰); 阶段 II 与阶段 I 的差别主要为 1 240 和 1 490 cm^{-1} 等处的峰值大幅降低, 可能是因为蛋白、腐殖质等不饱和污染物的相对份额降低; 阶段 III 作为后期污染, 与阶段 II 相比, 2 800 cm^{-1} 以上的吸收有所增强, 可能与富含氢键的污染物进一步累积有关。

聚类结果的物理意义, 可进一步通过膜污染过程的其他表征手段来交叉验证。膜污染红外光谱特征的分阶段差异, 不仅归因于不同阶段有机污染物成分和性质的不同, 还可能与共存微生物(例如可通过二次代谢降解有机物并分泌胞外

多聚物)在膜表面的不断累积和发展有关^[16]。鉴于此,测定了不同阶段的膜样品表面的材料学(XPS)和生物学(ATP)特性,结果如图 6 所示。其中,沿空白膜→阶段 I→阶段 II→阶段 III 的方向, O/C 元素比例逐步升高(Jonckheere-Terpstra 检验的 J-T 统计量 2.40, 显著性水平 <0.05), C—(O, N, OH)单键和 C—(C, H)单键的相对占比分别呈现显著的单调递增和单调递减的趋势(J-T 统计量分别为 3.38 和 -3.10, 显著性水平 <0.01), 而各阶段之间 N/C 元素比例和羰基 [O=C—(R, N, O)] 相对占比的差异却不显著。这说明在蛋白、腐殖质、多糖这几类常见的膜污染物中, 含有饱和 C—O 键的多糖类物质更有可能呈现阶梯式增长的特征。ATP 也呈现显著的整体增长趋势(J-T 统计量 2.68, 显著性水平 <0.01), 阶段 III 和阶段 II 之间的差异较小, 在阶段 II 膜表面微生物的富集已趋稳定, 在阶段 III 微生物则可能具有稳定的生长代谢过程。综合以上分析, 可大致推测各阶段的膜污染特征: 阶段 I 为各类有机污染物的初步覆盖, 微

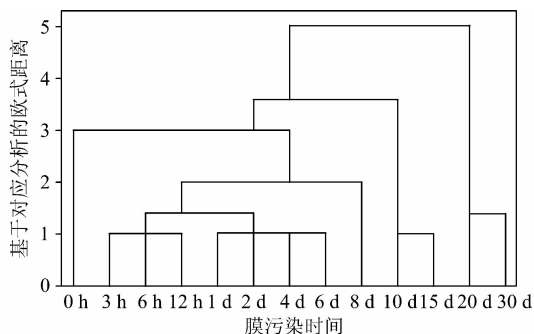


图 5 基于对应分析的纳滤膜污染样本聚类结果

Fig. 5 The hierarchical clustering results of fouled NF membranes based on correspondence analysis

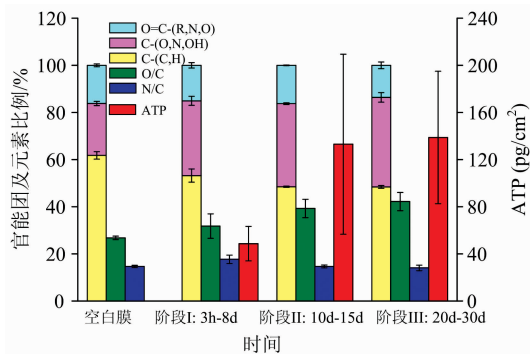


图 6 不同阶段的膜样品表面的 X 射线光电子能谱(XPS)元素比例(O/C 比和 N/C 比)、化学键相对占比(O=C—(R, N, O), C—(O, N, OH)和 C—(C, H))以及三磷酸腺苷(ATP)含量

Fig. 6 The atomic concentration ratios (O/C and N/C ratios), carbon-containing functional groups percentages (O=C—(R, N, O), C—(O, N, OH) and C—(C, H)) obtained by X-ray photoelectron spectroscopy (XPS) and adenosine triphosphate (ATP) content of NF membranes at different fouling stages

生物开始富集; 阶段 II 为多糖类污染物的比例上升, 微生物的富集趋于稳定; 阶段 III 为整体污染趋于成熟, 污染物组成比例稳定, 微生物自身的生长代谢成为膜污染的主要原因, 有机污染物的氢键特征更加明显。

红外光谱所包含的吸收峰信息复杂, 当样本间红外光谱差别较为细微时, 难以通过直观观察比较样本之间的整体差异性, 难以摒除结论的主观性。对红外数据进行聚类分析, 则能够灵敏地探测各红外图谱之间的差别, 结论更加客观而有说服力, 有助于红外光谱规律的深度挖掘。但在进行聚类分析时, 有诸多因素会对聚类结果产生影响, 例如是否对数据进行标准化, 选取何种距离度量方法(欧式距离、马氏距离等), 是否校正峰与峰之间相关性、峰与样本之间交互作用的干扰等。因此在进行聚类分析时, 应注意这些因素, 保障数据处理方法的严谨性。

研究中, 利用欧式距离能够直观判断出样本间红外光谱特征的整体变化规律, 但考虑到光谱中各吸收峰的权重, 进一步采用了标准化欧式距离进行比较。此外, 考虑到峰之间的自相关, 并且避免信息丢失, 采用了主成分分析进行修正。除了峰之间的相关性, 由于本研究中的纳滤膜样品代表了不同污染时间的结果, 整体上反映了膜污染的过程, 因此样本间也可能存在相关性, 从而对欧式距离的标准化和主成分的提取产生潜在影响。对应分析可进一步校正这方面的影响。结果表明, 经过不断改进最终的数据处理方法能够实现红外光谱的有效聚类, 系统地甄别不同膜样品之间的差异, 并清晰地膜污染划分为不同阶段。红外光谱聚类分析结果的合理性得到 XPS 和 ATP 等其他理化和生化分析方法的验证, 几种方法所得的膜污染发展规律一致。

红外光谱聚类分析对纳滤膜污染动态过程的解析具有指示意义。红外光谱聚类分析能够区分膜污染的不同阶段, 并有助于归纳出不同阶段的污染层特征, 可作为膜污染时序特征的侦查手段。在红外光谱的基础上进行聚类分析, 无需投入额外仪器成本, 即可深入挖掘红外光谱规律。若再结合其他的理化和生化表征手段, 则可进一步确证不同阶段污染层的物质成分和性质变化, 解析膜污染的发展过程和机理, 识别不同阶段的关键污染物成分, 从而为膜污染的针对性控制和程式化清洗提供支持。

此外, 除了膜污染的研究, 在材料、吸附等领域, 只要有一系列变化的红外光谱, 均可尝试采用外光谱聚类分析方法, 获取基于红外特征的定义信息或分阶段规律。

3 结论

红外光谱聚类分析可用于挖掘系列样本的红外光谱变化规律。在方法运用的过程中, 需考虑数据标准化、距离度量方法(欧式距离、马氏距离等)、峰之间自相关性、峰与样本之间交互作用等因素的影响。建议采用对应分析对红外数据进行预处理, 提取各样本在主要维度上的得分, 随后基于标准化欧式距离对各样本进行聚类。

将该方法应用于纳滤膜污染动态发展过程的解析, 可有效地对不同时间点的污染膜样本进行分类, 从而较为客观地

识别膜污染的不同阶段。经其他理化和生化表征手段的交互验证表明, 红外光谱聚类分析结果的可解释性强, 具有显著的物理意义。该方法为膜污染研究提供了额外信息, 有望为

分阶段针对性的膜污染精细控制提供支持。此外, 对于其他领域的系列红外光谱样本数据解析, 本方法也可供参考。

References

- [1] Xu K, Ren H Q, Ding L L, et al. *Environmental Science and Pollution Research*, 2013, 20(2): 771.
- [2] Zhao Y Y, Kong F X, Wang Z, et al. *Frontiers of Environmental Science & Engineering*, 2017, 11(6): 20.
- [3] Li C, Yang Y, Ding S Y, et al. *Journal of Membrane Science*, 2016, 514: 264.
- [4] Sari M A, Chellam S. *Journal of Membrane Science*, 2013, 437(12): 249.
- [5] Tang C Y, Kwon Y N, Leckie J O. *Journal of Membrane Science*, 2014, 326(2): 526.
- [6] Khan M T, Manes C L D O, Aubry C, et al. *Environmental Science & Technology*, 2013, 47(19): 10884.
- [7] Ivnitsky H, Katz I, Minz D, et al. *Water Research*, 2007, 41(17): 3924.
- [8] Herzberg M. *Environmental Science & Technology*, 2009, 43(12): 4393.
- [9] GUO Yu, WANG Xiao-mao(郭 驭, 王小毛). *Water & Wastewater Engineering(给水排水)*, 2017, (9): 120.
- [10] HUANG Hong-ying, YIN Qi-he(黄红英, 尹齐和). *Journal of the Graduates, Sun Yat-sen University • Natural Sciences, Medicine(中山大学研究生学刊 • 自然科学与医学版)*, 2011, (1): 20.
- [11] XU Yong-qun, HUANG Hao, ZHOU Qun, et al(徐永群, 黄 昊, 周 群, 等). *Chinese Journal of Analytical Chemistry(分析化学)*, 2003, 31(1): 5.
- [12] WANG Yan, WANG Bin, XU Yin-feng, et al(王 燕, 王 斌, 徐银峰, 等). *Journal of Chinese Institute of Food Science and Technology(中国食品学报)*, 2013, 13(1): 178.
- [13] Li G Q, Yu T, Wu Q Y, et al. *Water Research*, 2017, 123: 345.
- [14] Carvalho S I, Otero M, Duarte A C, et al. *Chemosphere*, 2008, 73(11): 1845.
- [15] ZHU Hong-tao, WEN Xiang-hua, HUANG Xia(朱洪涛, 文湘华, 黄 霞). *Environmental Science(环境科学)*, 2009, 30(1): 302.
- [16] Heffernan R, Habimana O, Semião A J C, et al. *Water Research*, 2014, 67: 118.

Tracking the Dynamic Evolution of NF Membrane Fouling Through Clustering Analysis Based on ATR-FTIR Spectra

LI Meng-chen¹, XIAO Kang², HUANG Xia^{1*}

1. State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China
2. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract Sewage reclamation is one of the effective countermeasures for solving the water shortage problem. Nanofiltration (NF) process is an effective method for reclaiming secondary effluent since it can provide high-qualified water. However, during the nanofiltration process, complicated and dynamic membrane fouling occurred, which can cause the decrease in flux and effluent quality. Tracking the dynamic evolution of NF membrane fouling in water treatment is important for controlling the membrane fouling depending on different fouling stages. Organic matters are important indicative contaminant for the dynamic evolution of fouling layer. Attenuated total reflectance-Fourier transform infrared spectroscopy (ATR-FTIR) is one of the most significant methods to characterize the change of functional groups in fouling layer. However, the peaks of ATR-FTIR are complex and the variations between different samples are tiny, especially when the fouled membrane samples are similar in time. It is difficult to directly discriminate the variation and trend of different ATR-FTIR spectra and cannot be the convictive evidence for fouling stages recognition. To investigate the dynamic evolution of membrane fouling, this study categorized the NF membranes samples obtained at different fouling time through combining the ATR-FTIR spectra and clustering analysis. Considering the influence of distance measurement method between categories, normalization of ATR-FTIR peak absorbance, correlation between peaks, and interaction between peaks and samples, this study utilized the correspondence analysis as the pretreatment of the ATR-FTIR spectra to obtain the scores of different membrane samples along main dimensions and then clustered the samples based on normalized Euclidian distance. During the 1-month NF experiment using sewage secondary effluent, because of the deposition of foulants, membrane fouling occurred and 13 fouled membranes were obtained at different time. Based on the hierarchi-

cal clustering of ATR-FTIR spectra, the fouling process can be clearly divided into the stages of: virgin membrane, stage I (3 h ~8 d), stage II (10~15 d), and stage III (20~30 d). The results of the clustering analysis was further interpreted by X-ray photoelectron spectroscopy (XPS) and adenosine triphosphate (ATP) content on the membrane surface. It was shown that with the evolution of membrane fouling, the organic composition and coexisting microorganism amount on the membrane surface changed concordantly. The characteristics of different stages may be interpreted as: in stage I, the membrane was initially covered by organic foulants, and microorganism began to gather; in stage II, the proportion of polysaccharide-like substance increased and the gathering of microorganism became stable; in stage III, the membrane fouling became mature and the hydrogen bond characteristics of organic foulants became more evident. In this study, the variation of different ATR-FTIR spectra was detected sensitively through clustering analysis. The study provides an objective, automatic and measurable auxiliary method for recognizing and characterizing membrane fouling stages. Besides, it is meaningful for investigating the ATR-FTIR spectra of a series of samples in not only membrane fouling research but also other fields such as materials science and adsorption research.

Keywords ATR-FTIR spectra; Clustering analysis; Nanofiltration membrane fouling; Dynamic evolution

(Received Feb. 10, 2018; accepted Jun. 24, 2018)

* Corresponding author