DOI:10.11918/202511103

寒区流域复合环境压力驱动微生物群落季节性变化分析

孙 笛[1,2],叶雪松[1,2],张 枫[1,2],董双石[1,2],赵振豪[1,2],姜晶晶[1,2]

(1. 地下水资源与环境教育部重点实验室(吉林大学),长春 130021;
2. 吉林省水资源与水环境重点实验室(吉林大学),长春 130021)

摘 要:为支撑寒区流域水生态精细化管控,明确冬夏季节水质与抗生素对微生物群落的差异化驱动特征及关键环境因子浓度效应拐点,以某典型寒区流域为研究对象,整合132条有效样本(夏季68条、冬季64条),构建包含8种水质指标及11种抗生素与微生物群落的关联数据集。采用随机森林(Random Forest,RF)模型分析环境因子与群落多样性的关系,结合SHAP(SHapley Additive exPlanations)可解释性分析识别关键驱动因子及其浓度效应拐点与多变量交互作用,并通过典范对应分析(Canonical Correspondence Analysis,CCA)解析群落功能潜力的季节差异。结果表明:该寒区流域冬夏环境与群落特征呈现显著季节差异,冬季溶解氧(DO,均值10.32 mg/L)显著高于夏季(6.95 mg/L),夏季诺氟沙星(NOR,均值156.76 ng/L)等抗生素质量浓度波动远高于冬季(1.23 ng/L),群落多样性和功能潜力季节差异显著;TN在冬夏均为驱动Shannon指数变化的关键因子,PD_whole_tree指数夏季以大环内酯类、四环素类抗生素主导,冬季以DO与喹诺酮类抗生素主导,微生物群落多样性驱动因子呈现季节特异性,SHAP分析进一步识别出DO(夏季7.41 mg/L、冬季10.26 mg/L)、TN(冬季1.01 mg/L)等关键因子可能存在的浓度效应拐点,初步揭示其对微生物群落多样性产生促进或抑制效应的浓度区间;夏季多种抗生素与DO、冬季TN与土霉素(OTC)之间存在协同效应,复杂环境中多因子复合压力对微生物群落有叠加影响;冬季群落信息存储与处理功能受DO、温度主导,夏季细胞过程与代谢功能受多种抗生素调控,形成“冬季水质主导-夏季抗生素调控”的季节分异模式。本研究在方法上融合机器学习与可解释性分析,揭示了寒区流域微生物群落对复合环境压力的季节性响应规律,提出的浓度效应拐点为流域分季节生态风险管控提供了科学参考。

关键词:寒区流域;微生物群落;季节变化;随机森林;可解释性分析
中图分类号:X522 **文献标志码**:A **文章编号**:0367-6234(2026)06-0068-12

Analysis of seasonal variations in microbial communities driven by combined environmental pressures in cold-region river basins

SUN Di[1,2], YE Xuesong[1,2], ZHANG Feng[1,2], DONG Shuangshi[1,2], ZHAO Zhenhao[1,2], JIANG Jingjing[1,2]

(1. Key Laboratory of Groundwater Resources and Environment (Jilin University), Ministry of Education, Changchun 130021, China;
2. Jilin Provincial Key Laboratory of Water Resources and Environment (Jilin University), Changchun 130021, China)

Abstract: To support the refined management of water ecology in cold-region river basins and clarify the differential driving characteristics of water quality and antibiotics on microbial communities in winter and summer, as well as the concentration effect inflection points of key environmental factors, this paper focused on a typical cold-region river basin, integrated 132 valid samples (68 in summer and 64 in winter), and constructed an association dataset containing eight water quality indicators, 11 antibiotics, and microbial communities. This paper applied the Random Forest (RF) model to analyze the relationship between environmental factors and community diversity, combined SHapley Additive exPlanations (SHAP) interpretable analysis to identify key driving factors, their concentration effect inflection points, and multivariate interactions, and analyzed the seasonal differences in community function potential through Canonical Correspondence Analysis (CCA). The results indicate that the environmental and community characteristics of this cold-region river basin exhibit significant seasonal differences; the dissolved oxygen (DO, mean value of 10.32 mg/L) in winter is significantly higher than that in summer (6.95 mg/L); the mass concentration fluctuation of antibiotics such as norfloxacin (NOR, mean value of 156.76 ng/L) in summer is much higher than that in winter (1.23 ng/L), and the community diversity and functional potential

收稿日期:2025-11-30;**录用日期**:2026-02-03;**网络首发日期**:2026-03-19
网络首发地址:https://link.cnki.net/urlid/23.1235.t.20260318.1139.004
基金项目:国家自然科学基金(5240100851,52570079,W2521110);吉林省科技厅青年人才培养项目(20260602015RC);吉林省科技发展计划(20260102033JC)
作者简介:孙 笛(1998—),女,硕士研究生;董双石(1980—),男,教授,博士生导师
通信作者:赵振豪,zhaozhenhao0621@163.com;姜晶晶,Jingjingjiang@jlu.edu.cn

show significant seasonal differences; TN is the key factor driving the changes in Shannon index in both winter and summer, while PD_whole_tree index is dominated by macrolide and tetracycline antibiotics in summer and by DO and quinolone antibiotics in winter, indicating that the driving factors of microbial community diversity exhibit seasonal specificity; SHAP analysis further identifies the potential concentration effect inflection points of key factors such as DO (7.41 mg/L in summer, 10.26 mg/L in winter), and TN (1.01 mg/L in winter), preliminarily revealing their concentration ranges that exert promoting or inhibitory effects on microbial community diversity; there are synergistic effects between multiple antibiotics and DO in summer and between TN and oxytetracycline (OTC) in winter, and the multi-factor combined pressures in complex environments have a superimposed impact on microbial communities; the community information storage and processing functions in winter are dominated by DO and temperature, while the cellular processes and metabolic functions in summer are regulated by multiple antibiotics, forming a seasonal differentiation pattern of "water quality dominance in winter and antibiotic regulation in summer". Methodologically, this paper integrated machine learning and interpretable analysis and revealed the seasonal response patterns of microbial communities to combined environmental pressures in cold-region river basins. The proposed concentration effect inflection points provide a scientific reference for the season-specific ecological risk management of river basins.

Keywords: cold-region river basin; microbial community; seasonal variation; Random Forest; interpretable analysis

微生物群落是维系水生态系统物质循环与能量流动的关键载体,其结构组成与功能潜力可直接反映河流生态健康状态[1-2],因此,解析环境因子对微生物群落的影响机制,对于流域水生态风险评估与精细化管控具有重要意义。寒区流域因独特的水文气候特征呈现显著的季节差异。冬季漫长冰封期内,水体维持低温(2~4 ℃)、高溶解氧(10~12 mg/L)状态,微生物代谢活性降低,群落结构趋于稳定;夏季高温和农业面源污染输入导致有机污染物及抗生素质量浓度显著上升,微生物活性增强,群落结构变化剧烈[3]。这种季节差异使微生物群落面临的环境压力呈现明显的时间异质性,亟需针对性研究支撑分季节精细化管控。已有研究表明,冰封条件下微生物群落通过增强跨域网络关联性提升对低温胁迫的适应性[4],而春季融冰期冰体中累积的污染物释放可能加剧水体生态风险[5]。因此,解析冬夏季节差异下微生物群落的动态变化及环境因子的驱动机制,对寒区流域生态风险评价研究具有重要的理论与实践价值。

当前寒区流域的研究虽已关注季节动态特征,但多单独分析水质指标或单一抗生素对群落的影响[6],缺乏对多种抗生素与常规水质指标共同构成的复合压力的系统考量。同时,现有流域生态风险评价体系尚未将“复合压力-微生物群落响应”这一关联纳入评估范畴[7]。即便部分研究涉及寒区环境因子与微生物的关联[8],也多采用线性分析方法,难以有效捕捉环境因子与微生物群落间的非线性关系,且缺乏对关键因子浓度效应拐点的定量识别,无法满足精细化管控对多因子综合分析的需求。

近年来,机器学习方法在环境微生物学领域展现出独特优势。随机森林(RF)模型具有适配小样本数据、精准捕捉非线性关联的能力,可有效处理多因子复杂关系[9-10];SHAP 可解释性分析则能定量解析各因子对预测结果的贡献,并识别关键因子的阈值效应与多因子交互作用[11],为环境因子-微生物群落关联研究提供了新的方法支撑。然而,现有研究多侧重于模型预测精度,对驱动机制的季节分异研究尚不充分,尤其缺乏针对寒区流域这一特殊地理单元的系统的、多指标的综合分析。

本研究以中国东北部某典型寒区流域为研究对象,整合冬夏两季水质理化指标与多种抗生素质量浓度数据,结合微生物群落多样性与功能潜力数据,采用 RF 模型构建环境因子与群落结构多样性的关联,结合 SHAP 分析识别关键驱动因子及其浓度效应拐点、多环境变量交互作用对群落结构的影响,并通过典范对应分析(Canonical Correspondence Analysis,CCA)解析冬夏微生物群落功能潜力的驱动差异。研究揭示了该寒区流域冬夏环境因子与微生物群落的季节差异特征,识别影响群落多样性的关键环境因子及其阈值效应,阐明冬夏群落功能潜力的差异化驱动机制,以期为寒区流域分季节生态风险管控提供科学依据,并为理解类似复杂环境系统中多因子对微生物群落的协同影响提供方法参考。

1　模型与方法

1.1　研究区域概况

选取中国东北部某典型寒区流域为研究对象,地理范围介于 121°47′~134°39′E、42°39′~51°13′N 之间,流域面积约 56 万 km^2。该流域属大陆性季风气候,四季分明,冬夏气候差异显著。多年平均气温 2.7 ℃,1 月均温低于 -20 ℃,7 月温度 20~24 ℃;多年平均降水量 531 mm,6—9 月汛期降水占全年

70% ~80%。流域内广泛分布季节性冻土,冻融期一般为11月—次年4月,部分可延长至5月,土壤冻结厚度1~1.5 m。流域径流补给以降水为主,春季积雪融水与冻土消融水为重要补充,年内径流量分配呈明显季节性,3—4月积雪与冻土消融形成春汛,6—9月集中降水形成夏汛,且夏汛期径流量占全年总量的60%以上。这种独特的水文过程使其成为研究寒区流域季节性特征的理想区域。

该流域总人口超7 000万,城镇化率约50%。流域内农业活动频繁,畜牧业发达,农业面源污染和生活污水是抗生素等新污染物的主要来源。冬季水体冰封期间污染物降解速率降低,而春季融冰期污染物集中释放[12],使得该流域水环境管控面临独特挑战。

1.2　数据收集与处理

微生物数据来源于NCBI数据库(https://www.ncbi.nlm.nih.gov/),通过关键词"寒区流域""河流微生物""16S rRNA gene sequencing"筛选并下载16S rRNA基因测序原始数据,共获取447条样本数据,覆盖研究区冬夏两季典型采样点位。

微生物数据处理采用QIIME2 pipeline(2023.9)。测序质量评估阶段,使用qiime demux插件对原始测序数据进行接头序列去除(基于Illumina接头数据库),采用qiime quality-filter q-score插件进行质量剪切(质量阈值 $q=20$),剔除长度<250 bp的短reads,通过qiime vsearch merge-pairs合并高质量双端reads。嵌合体去除与扩增子序列变体(Amplicon Sequence Variant,ASV)聚类阶段,采用DADA2算法构建误差模型,实现ASV聚类,同时去除嵌合体序列。物种注释与功能预测阶段,使用qiime feature-classifier classify-sklearn插件基于Silva138数据库对ASV序列进行属水平物种注释,通过PICRUSt2工具预测微生物群落功能(基于COG功能注释体系)。最终剔除测序深度<10 000 reads/样本的低质量数据,保留有效样本132条,其中,夏季(7—9月)68条,冬季(11月—次年2月)64条。

水质数据来源于国家水质自动综合监管平台(https://szzdjc.cnemc.cn:8070/GJZ/Business/Publish/Main.html),选取与微生物数据采样点时空匹配的断面监测数据,涵盖两类核心指标。理化指标为水温(Temperature)、pH、溶解氧(DO);营养盐指标为高锰酸盐指数(COD_{Mn})、氨氮(NH_3-N)、总磷(TP)、总氮(TN)、总有机碳(TOC)。

抗生素数据通过文献检索获取,检索平台包括Web of Science核心合集与中国知网(CNKI),检索主题词包括"river""antibiotic""occurrence""cold region"等,通过阅读标题和摘要进行初步筛选,得到涉及研究区域水体抗生素检测的研究51篇。进一步筛选遵循严格的数据完整性标准:研究区域需与微生物样本采集河流一致,且文献需明确提供抗生素种类、浓度数值及采样时间信息,确保数据与微生物、水质数据的时空关联性。最终获取6大类共11种典型抗生素数据(表1),包括磺胺类(磺胺嘧啶SDZ)、大环内酯类(红霉素ERY、泰乐菌素TYL、替米考星TIL)、四环素类(土霉素OTC、四环素TCH)、喹诺酮类(诺氟沙星NOR、恩诺沙星ENR、环丙沙星CIP)、β-内酰胺类(阿莫西林AMX)、氯霉素类(氯霉素CAP)。值得注意的是,本研究整合的抗生素数据来源于独立研究,基于该数据集识别的抗生素质量浓度效应拐点,主要反映统计关联规律,其生态学有效性与实际调控意义,可通过后续实地同步采样检测,结合机制解析,进一步深化认知。

表1　原始数据集信息

Tab.1　Information of raw dataset

数据类别	变量	单位	夏季			冬季		
			最小值	最大值	均值	最小值	最大值	均值
环境数据	水温	℃	11.40	22.30	16.10	1.90	5.50	4.20
	pH		4.90	8.20	6.80	8.00	8.30	8.10
	DO	mg/L	5.16	9.67	6.95	9.61	11.04	10.32
	COD_{Mn}	mg/L	4.72	6.68	5.85	6.17	7.50	6.79
	NH_3-N	mg/L	0.09	27.25	7.06	0.04	0.89	0.13
	TP	mg/L	0.02	0.09	0.04	0.01	0.08	0.05
	TN	mg/L	0.85	1.52	1.13	0.62	1.28	0.92
	TOC	mg/L	12.63	12.96	12.80	12.48	13.12	12.81

表 1(续)

数据类别	变量	单位	夏季			冬季		
			最小值	最大值	均值	最小值	最大值	均值
抗生素数据	SDZ	ng/L	0.53	9.82	4.36	0.54	69.13	7.33
	ERY	ng/L	0.61	40.18	3.97	—	1.45	0.46
	TYL	ng/L	—	1.37	0.12	—	—	—
	TIL	ng/L	—	32.85	2.51	—	0.07	0.02
	OTC	ng/L	—	34.07	14.41	3.48	58.58	21.01
	TCH	ng/L	—	19.12	2.99	2.42	17.65	8.79
	NOR	ng/L	0.71	1 553.20	156.76	0.09	3.92	1.23
	ENR	ng/L	—	15.30	1.24	0.31	0.98	0.56
	CIP	ng/L	—	0.50	0.12	—	0.55	0.19
	AMX	ng/L	2.20	128.59	52.69	1.55	30.95	14.48
	CAP	ng/L	2.23	24.24	9.90	1.78	17.10	6.68
微生物数据	Shannon		4.10	8.58	5.60	3.08	6.68	5.33
	PD_whole_tree		1.88	44.14	18.95	1.58	38.81	15.41

为确保数据时空一致性，本研究对微生物、水质和抗生素 3 类数据进行严格匹配。时间匹配：夏季定义为 7—9 月，冬季定义为 11 月—次年 2 月，3 类数据采样时间偏差严格控制在≤15 d，确保同一样本的微生物、水质、抗生素数据反映相同环境背景；空间匹配：以采样点所在断面为匹配单元，确保同一样本的 3 类数据来自相同或邻近(< 5 km)的监测点位；数据整合：最终形成 132 条有效样本的关联数据集(夏季 68 条、冬季 64 条)，用于后续模型构建与分析。

1.3　模型构建

本研究聚焦寒区流域冬夏季节差异，构建区分冬季、夏季两个季节以及 Shannon 指数、PD_whole_tree 指数分别为目标变量的 4 个独立模型，以解析环境因子对微生物群落多样性的驱动机制。4 个模型遵循统一的分析框架，完整流程如图 1 所示，包括数据预处理、特征工程、模型构建与可解释性分析，各模型在相同流程下独立执行所有分析步骤。

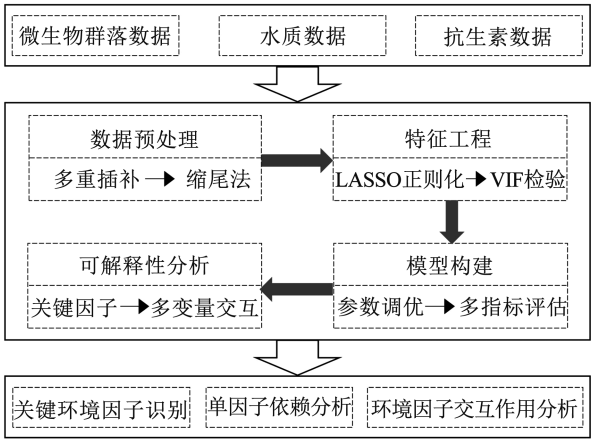


图 1　本研究采用的机器学习流程示意

Fig. 1　Schematic diagram of machine learning process used in this paper

1.3.1　数据预处理

针对各环境因子的缺失值，采用基于链式方程的多重插补(Multiple Imputation by Chained Equations，MICE)方法填充，通过捕捉特征间相关性提升插补精度，设置估计器为贝叶斯岭回归(Bayesian Ridge Regression)，最大迭代次数 10 次、近邻特征数 10 个，插补后验证结果显示无缺失值残留。异常值处理采用 1% 分位数缩尾法，基于各变量 1% 和 99% 分位数确定截断边界，将极端值约束在合理范围，保留数据整体分布特征。数据集划分采用简单随机抽样，按 7∶3 划分训练集与测试集。特征标准化采用 Z-score 方法，仅基于各模型的训练集数据拟合参数，随后应用于对应训练集与测试集，消除量级差异对模型训练的干扰。

1.3.2　特征工程

特征选择采用 LASSO 正则化方法，设置正则化参数 $\lambda = 0.01$，通过压缩系数剔除冗余特征，保留系数绝对值最大的 6 ~ 8 个核心特征，平衡模型复杂度与解释力。共线性检验采用方差膨胀因子(VIF)，4 个模型各自针对核心特征变量计算 VIF 值，设定阈值为 5，所有核心变量的 VIF 值为 1.17 ~ 3.25，无严重共线性问题，符合生态学研究中变量独立性要求。

1.3.3　模型构建与参数调优

采用 RF 模型构建环境因子与微生物群落多样性的关联模型。RF 模型通过集成多棵决策树的预测结果进行最终预测，具有适配小样本数据、捕捉非线性关系、抗过拟合能力强的优势。模型构建基于 Python scikit-learn 库(1.3.0 版本)实现。参数调优采用网格搜索(Grid Search)结合 5 折交叉验证，以交叉验证的平均 R^2 最大化为优化目标，统一设置参数搜索范围：决策树数量(n_estimators) ∈ {30，40，

50,100,130,150}、最大深度(max_depth) ∈ {4,5,6,7}、最小叶节点样本数(min_samples_leaf) ∈ {2,3,5}、最小分裂样本数(min_samples_split) ∈ {2,3,5},确定各模型最优参数(表 2)。

表 2 随机森林模型最优参数

Tab. 2 Optimal parameters of Random Forest model

季节	指标	n_estimators	max_depth	min_samples_split	min_samples_leaf	random_state
冬	Shannon	40	6	3	3	42
	PD_whole_tree	50	5	2	2	42
夏	Shannon	130	6	2	2	42
	PD_whole_tree	50	6	2	3	42

1.3.4 模型评估

采用 4 项常用指标综合评估 RF 模型的拟合精度与泛化能力。决定系数(R^2)衡量模型对响应变量变异的解释程度,取值范围为 0 ~ 1,值越大表明模型解释力越强;均方根误差(RMSE,E_{RMS})可直观体现平均预测误差的量级;平均绝对误差(MAE,E_{MA})通过计算偏差绝对值的均值,减少极端值对误差评估的过度影响;平均绝对百分比误差(MAPE,E_{MAP})以百分比形式反映相对误差,便于跨数据集比较模型性能。本研究同时报告训练集和测试集的评估指标,以全面评价模型的拟合能力与泛化性能。另外,为量化小样本条件下模型性能的不确定性,4 个模型均执行 500 次重复随机划分评估,每次按 7∶3划分训练集与测试集,采用对应模型的最优参数训练后,计算测试集 R^2 值,系统反映模型泛化能力的波动范围。

$$E_{\mathrm{RMS}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$E_{\mathrm{MA}} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

$$E_{\mathrm{MAP}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

式中:y_i为第 i 个样本的响应变量真实值,$\hat{y}_i$为对应预测值,$\bar{y}$ 为响应变量真实值的均值,n 为样本数。

1.3.5 可解释性分析

SHAP 分析基于博弈论中的 Shapley 值原理,可定量解析各环境因子对微生物群落多样性预测结果的贡献,并识别关键因子的阈值效应与多因子交互作用。本研究采用 Python 的 SHAP 库(0.42.1 版本),使用 Tree Explainer 适配随机森林模型以提升计算效率与解释精度。

1)特征重要性分析。通过计算各因子平均 SHAP 绝对值(|SHAP|_mean)衡量其对多样性指数的整体贡献,值越大表明该因子对预测结果的影响越显著。

2)单变量依赖分析。通过绘制 SHAP 依赖图(dependence plot)解析单一因子浓度变化对预测结果的边际效应,结合广义可加模型(GAM)拟合曲线,识别关键因子的生态阈值,定义为 SHAP 值由正转负或由负转正的临界点,反映因子对群落结构从“促进”到“抑制”的转变浓度。

3)多变量交互分析。通过计算 SHAP 交互值(interaction value)量化双因子间的协同或拮抗效应(SHAP 交互值的计算基于 Shapley 值的扩展,通过比较特征在所有可能排序组合中的联合贡献与各自独立贡献的差异来量化交互效应),以交互值绝对值衡量交互强度(值越大表明双因子共同作用越显著),明确不同季节水质指标与抗生素间的组合效应及强度排序特征。

1.3.6 典范对应分析

典范对应分析(CCA)用于解析微生物群落功能潜力与环境因子的关联。以 COG 功能类别(信息存储与处理 ISP、细胞过程与信号传导 CPS、代谢功能 MET、未知功能 PC)为响应变量,以水质指标与抗生素为解释变量,通过 Canoco5.0 软件实现。以前两轴累计解释变异率评估模型拟合效果,明确冬夏季节环境因子对群落功能的差异化驱动机制。

2 结果与讨论

2.1 微生物群落结构与功能的季节性差异

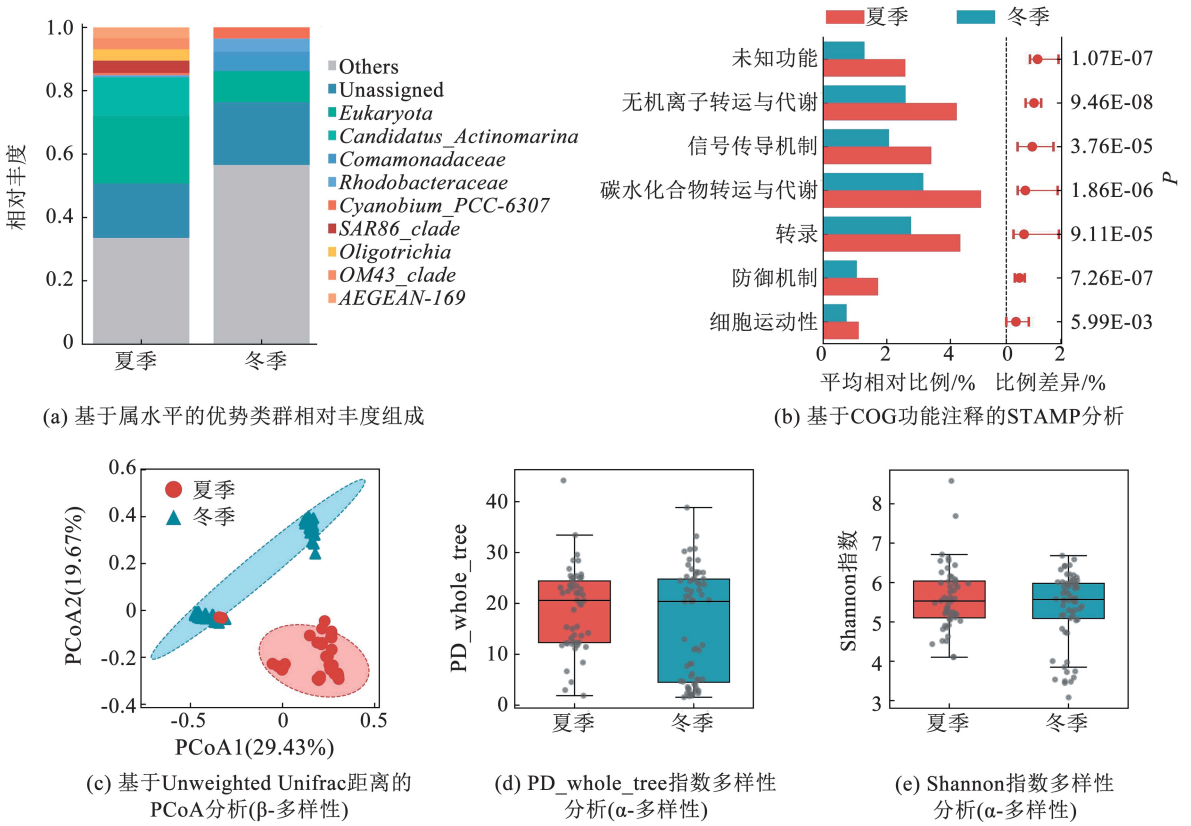
冬夏微生物群落的组成呈现显著的季节性分化(图 2(a))。夏季优势类群以 *Candidatus Actinomarina* 和 *SAR86 clade* 等浮游生境关联类群为主,这些类群通常在温暖、富营养的水体中具有竞争优势,其丰度与夏季较高的水温和有机质输入密切相关[13]。冬

季则以 *Comamonadaceae* 和 *Rhodobacteraceae* 等好氧、兼性厌氧菌占优，反映了低温高溶解氧环境下微生物群落的适应性调整。值得注意的是，冬夏样本中均存在大量未分类（Unassigned）类群（相对丰度 >30%），体现该寒区流域微生物群落具有较高的分类学未知性，可能含有尚未被鉴定的寒区特有类群。这种群落组成的季节差异与已有研究结论一致，Zhang 等[4]对内蒙古湖泊的研究发现，冰封期与非冰封期细菌群落结构存在显著差异，水体 pH 是影响群落组成的关键因子。Baker 等[14]对阿拉斯加北极潟湖的研究也表明，微生物群落的系统发育组成随季节显著变化，以适应极端的环境波动。

基于 PICRUSt2 预测的 COG 功能谱，通过 STAMP 软件进行组间差异检验显示，群落功能潜力同样呈现显著的季节分化（图 2（b））。夏季微生物群落在多数功能类别中相对丰度更高，尤其在碳水化合物转运与代谢（$P < 10^{-5}$）、无机离子转运与代谢（$P < 10^{-7}$）等功能上更为活跃，反映了夏季高温环境下微生物代谢活性的整体提升。类似地，Li 等[15]在东北黑土带河流的研究表明，季节变化通过影响有机质可利用性调控微生物呼吸活性。

基于 Unweighted Unifrac 距离的 PCoA 分析结果表明（图 2（c）），冬夏样本在排序空间中聚类明显分离，前两轴累计解释 49.1% 的群落变异，证实季节是驱动群落 β-多样性变化的关键因素。这种显著的季节聚类模式表明，尽管同一季节内不同采样点之间存在空间异质性，但季节差异对群落结构的影响仍占主导地位。这提示在寒区流域生态管控中，应充分考虑季节差异对微生物群落的影响，制定分季节的差异化管控策略。

α-多样性分析显示（图 2（d）、（e）），冬季 PD_whole_tree 指数（均值 15.41）显著低于夏季（均值 18.95，$P = 0.044$），但差异程度有限。这一结果表明群落的物种丰富度和均匀度在两季节间相对稳定。该现象可能与冬季水体环境相对稳定有关，低温条件下微生物代谢活性降低，竞争强度减弱，使群落能够维持相对稳定的多样性水平[16]。相比之下，夏季虽然环境因子波动剧烈，但高温和丰富的营养供给也为多种微生物类群提供了生态位，抵消了环境胁迫对多样性的负面影响。整体而言，季节对流域微生物群落 α-多样性的影响弱于对群落组成和 β-多样性的影响。



(a) 基于属水平的优势类群相对丰度组成

(b) 基于COG功能注释的STAMP分析

(c) 基于Unweighted Unifrac距离的PCoA分析(β-多样性)

(d) PD_whole_tree指数多样性分析(α-多样性)

(e) Shannon指数多样性分析(α-多样性)

图 2　某寒区流域微生物群落结构和功能的季节性差异

Fig. 2　Seasonal difference in microbial community structure and function in a cold-region river basin

2.2 环境因子对群落多样性的季节性影响

整合水质、抗生素与微生物群落数据构建的 RF 预测模型，经多项指标评估表现出稳健的拟合与泛化能力（表 3）。总体来看，训练集 R^2 均高于 0.75，测试集 R^2 为 0.51 ~ 0.84，MAE 和 MAPE 均处于较低水平，表明所构建的 RF 模型能有效捕捉环境因子与微生物群落结构的非线性关联。PD_whole_tree 指数模型的整体拟合与泛化能力相对更优，测试集 R^2 较 Shannon 指数模型高 47% ~ 65%。从季节角度看，Shannon 指数模型的冬夏测试集 R^2 接近（约为 0.51），但夏季 MAPE（6.61%）较冬季（11.69%）降低约 43%，这种季节差异与流域水文气候特征密切相关。夏季受高温与农业面源污染影响，环境因子梯度更显著：COD_{Mn} 波动范围为 4.72 ~ 6.68 mg/L、NOR 质量浓度范围为 0.71 ~ 1 553.20 ng/L（表 1），相比之下，冬季冰封期水体处于低温（1.9 ~ 5.5 ℃）、高 DO（9.61 ~ 11.04 mg/L）的相对稳定状态，环境因子波动幅度窄，导致因子-群落关联性减弱。这一模型性能特征与机器学习在复杂环境系统中的应用规律一致，即因子梯度越显著，模型对非线性关系的捕捉能力越强[17]。

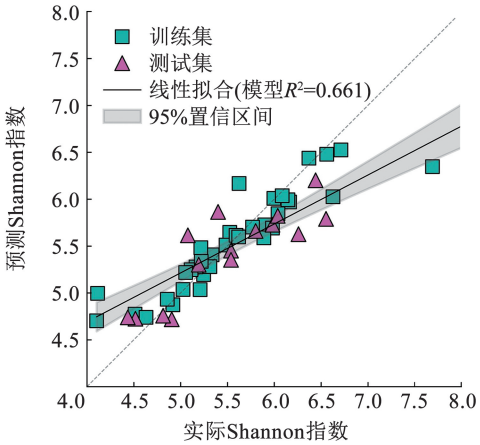
表 3 训练集与测试集上的 Shannon 指数与 PD_whole_tree 指数预测性能

Tab. 3 Prediction performance of Shannon index and PD_whole_tree index on training and test sets

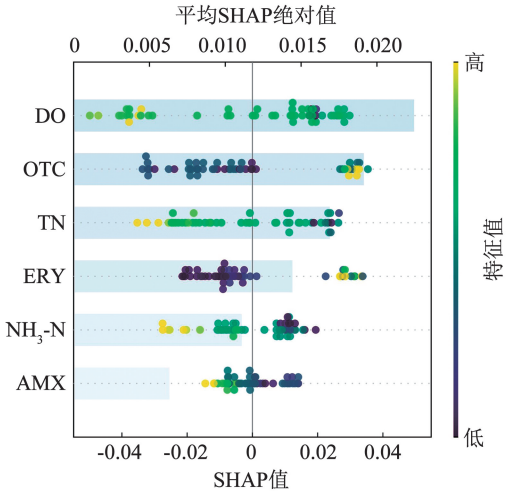
季节	预测指标	训练集				测试集					
		R^2	MAE	RMSE	MAPE	R^2	R^2:95% CI	R^2:mean	MAE	RMSE	MAPE
冬	Shannon	0.745 7	0.338 4	0.438 0	0.070 2	0.512 5	0.412 2 ~ 0.738 6	0.420 9	0.541 0	0.686 4	0.116 9
	PD_whole_tree	0.877 3	2.726 0	3.624 1	0.212 2	0.836 7	0.382 2 ~ 0.848 5	0.661 0	3.482 6	4.198 2	0.237 2
夏	Shannon	0.779 0	0.210 4	0.342 9	0.038 2	0.509 8	0.383 7 ~ 0.606 7	0.410 9	0.421 2	0.685 0	0.066 1
	PD_whole_tree	0.884 6	1.887 3	2.479 5	0.161 7	0.753 2	0.422 3 ~ 0.856 3	0.701 2	3.113 7	4.202 8	0.269 8

2.2.1 Shannon 指数环境驱动因子的季节特征

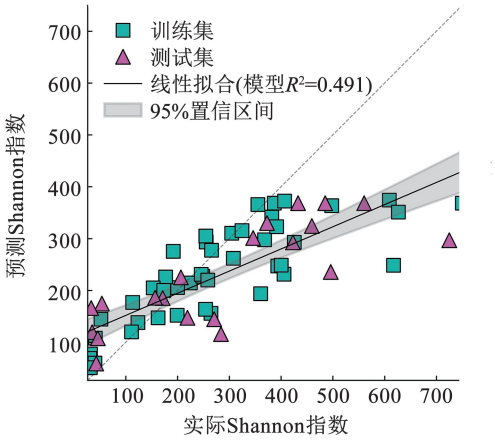
SHAP 特征重要性分析揭示了冬夏两季 Shannon 指数驱动因子的共性与差异（图 3（b）、（d））。夏季 DO（SHAP 均值 0.022）、OTC（0.019）、TN（0.017）为核心驱动因子，Wang 等[11]在城市受纳河流中通过 SHAP 分析发现 DO 是群落 Chao1 指数和 Shannon 指数的关键预测因子，本研究进一步揭示了抗生素压力的突出作用。冬季 TN（0.157）、TIL（0.148）和 NOR（0.128）为主导因子，表明 TN 作为基础水质因子，通过氮源供给影响微生物生长与代谢[18]，是贯穿两季调控微生物群落物种均匀度的关键因素，但其相对重要性存在季节差异。冬季 TN 的影响尤为突出，反映低温环境下微生物对氮源波动的敏感性增强，这可能与冬季微生物代谢活性降低、对环境因子变化的耐受范围收窄有关[19]。



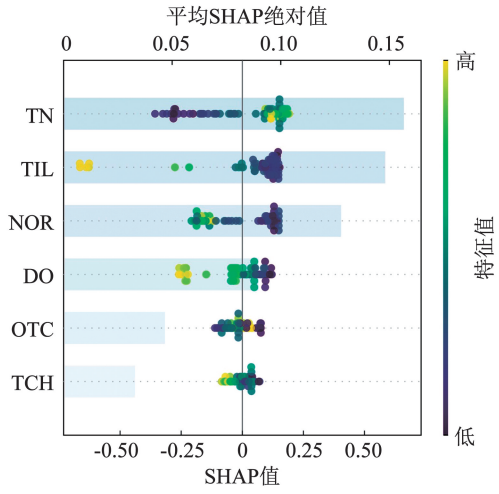
(a) 夏季模型的预测性能



(b) 夏季变量的SHAP特征重要性



(c) 冬季模型的预测性能



(d) 冬季变量的SHAP特征重要性

注：(a)、(c)中 R^2 为全集数据评估。

图 3　Shannon 指数的模型预测性能与 SHAP 特征重要性评估

Fig. 3　Assessment of model prediction performance and SHAP feature importance for Shannon index

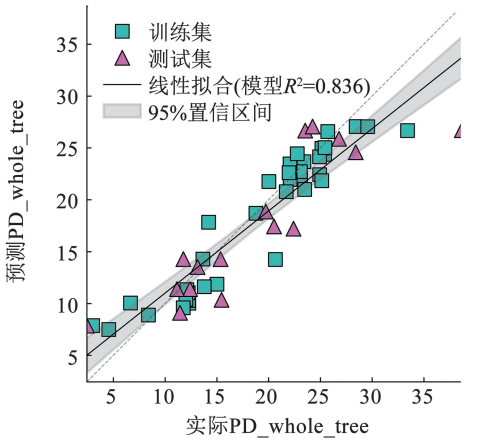
2.2.2　PD_whole_tree 指数环境驱动因子的季节模式

与 Shannon 指数不同，PD_whole_tree 指数的驱动因子呈现更显著的季节特异性（图 4（b）、（d））。夏季群落系统发育多样性调控以 TIL（SHAP 均值 0.136）、OTC（0.102）和 TYL（0.078），即大环内酯类和四环素类抗生素为主导；冬季则以 DO（0.175）、NOR（0.155）和 pH（0.128），即水质与喹诺酮类抗生素协同调控。值得注意的是，冬季 DO 重要性居首，提示低温环境下氧化还原状态对群落系统发育结构的重要影响，冬季 NOR 虽然平均浓度较低，但喹诺酮类抗生素在低温条件下降解速率显著降低（可降低 60% 以上）[20]，同时，寒区河流冰封期存在冰体封存与融冰期污染物集中释放的过程[5]，使 NOR 能够持续作用于水体微生物，pH 则通过影响细胞膜电位和酶活性调控微生物生理功能[21]。PD_whole_tree 指数与 Shannon 指数驱动因子的分离，说明物种均匀度和系统发育多样性反映了群落结构的不同维度，提示需在多指标框架下评估环境效应。

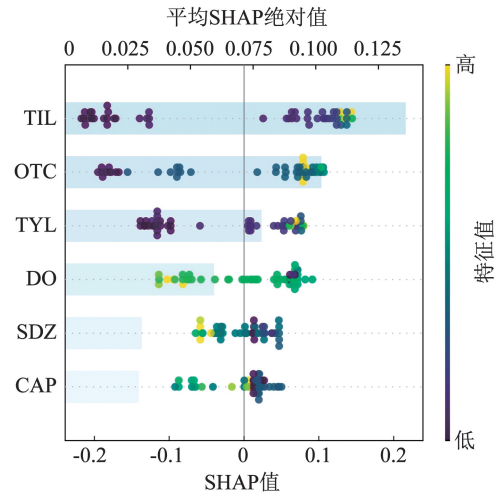
2.3　复合环境压力下的群落响应特征

2.3.1　Shannon 指数关键因子的阈值效应

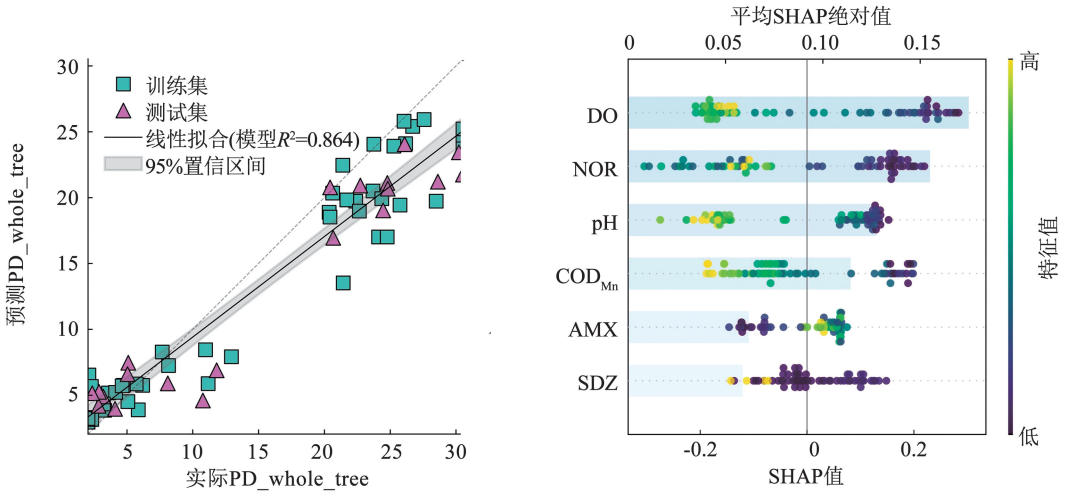
SHAP 单变量依赖分析初步揭示了关键环境因子对群落 Shannon 指数可能存在的浓度效应拐点（图 5）。夏季 DO 存在阈值 7.41 mg/L（图 5（a）），当 DO 低于该值时 SHAP 值为正，对 Shannon 指数产生正面影响，促进物种均匀度，超过则转为抑制（R^2 =94.6%）。这一阈值效应表明夏季适宜 DO 质量浓度（<7.41 mg/L）可维持好氧与兼性厌氧微生物的生态位平衡，而过高 DO 可能改变氧化还原环境，抑制部分厌氧或微好氧类群，简化群落结构。冬季 TN 质量浓度低于 1.01 mg/L 时 SHAP 值为负，氮源不足限制微生物生长（图 5（d）），超过则转为促进。Huang 等[22]针对鄱阳湖河湖交错带的研究也证实适宜氮质量浓度促进硝化细菌相对富集，氮源可用性是调控微生物群落的关键因子。



(a) 夏季模型的预测性能



(b) 夏季变量的SHAP特征重要性

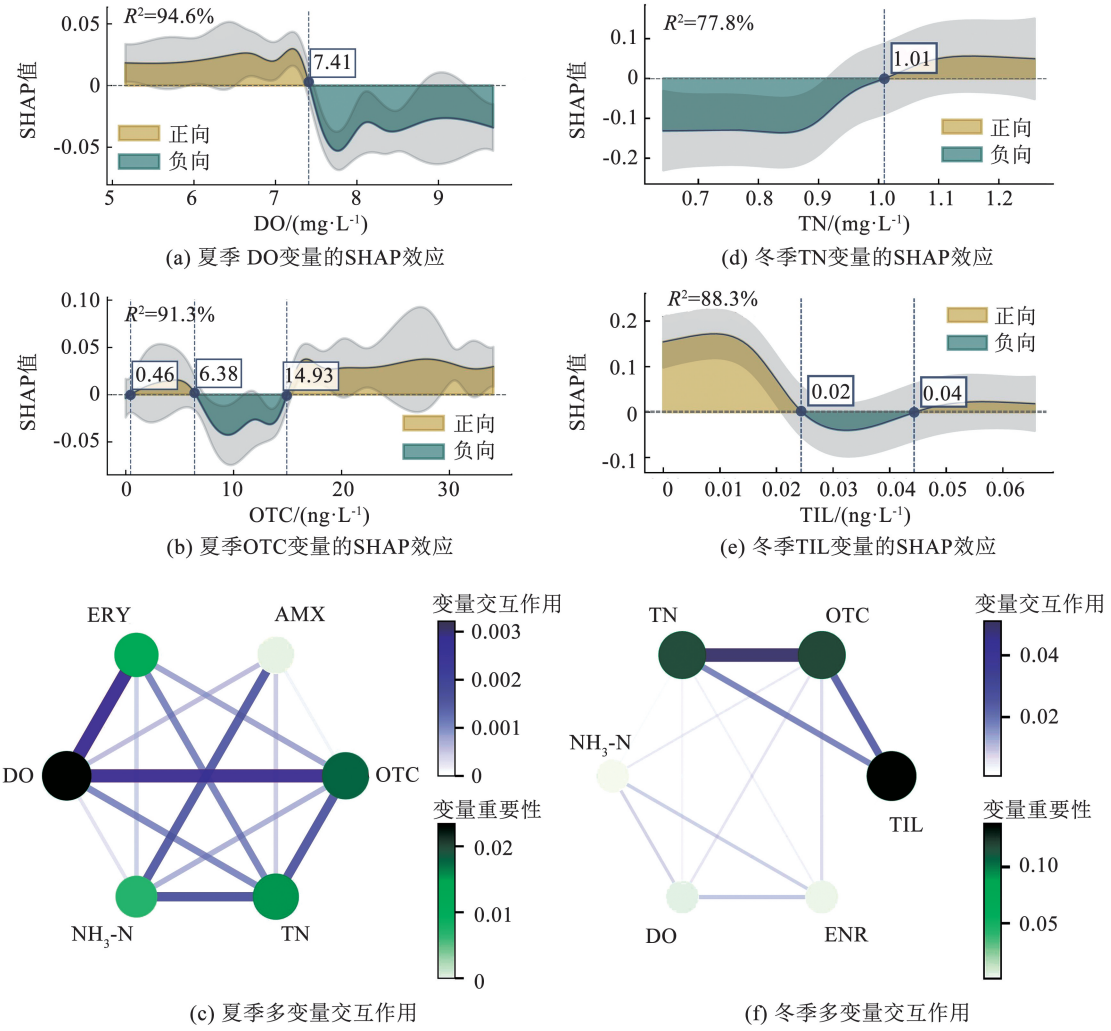


(c) 冬季模型的预测性能　　(d) 冬季变量的SHAP特征重要性

注：(a)、(c)中 R^2 为全集数据评估。

图 4　PD_whole_tree 指数的模型预测性能与 SHAP 特征重要性评估

Fig. 4　Assessment of model prediction performance and SHAP feature importance for PD_whole_tree index



(a) 夏季 DO变量的SHAP效应　　(d) 冬季TN变量的SHAP效应

(b) 夏季OTC变量的SHAP效应　　(e) 冬季TIL变量的SHAP效应

(c) 夏季多变量交互作用　　(f) 冬季多变量交互作用

图 5　Shannon 指数预测模型的 SHAP 单变量依赖性及多变量交互作用分析

Fig. 5　Analysis of SHAP univariate dependency and multivariate interaction of Shannon index prediction model

抗生素变化与 SHAP 响应的波动性表明其阈值不明确，且冬夏的群落多样性与抗生素变化间存在复杂关联，从抗生素单特征依赖图可知，存在多个 SHAP 值 0 点，无法界定统一浓度阈值以区分其对

Shannon 指数的正负影响，表明抗生素对微生物群落多样性的影响机制较 DO、TN 等环境因子更为复杂。

2.3.2　PD_whole_tree 指数关键因子的阈值效应

SHAP 分析同样识别出影响 PD_whole_tree 指数的关键因子浓度拐点（图 6）。夏季 TIL（阈值 1.71 ng/L，图 6（a））和 OTC（阈值 14.14 ng/L，图 6（b））呈现近似的波动模式，低于时抑制系统发育多样性，超过则促进（OTC 影响程度相对较弱），反映抗生素质量浓度与群落进化结构的非线性关联。冬季 DO（阈值 10.26 mg/L，图 6（d））和 NOR（阈值 0.8 ng/L，图 6（e））则呈现近似且与夏季相反的调控模式，二者处于低值时对群落系统发育多样性有促进作用。需注意的是，冬季 NOR 均值（1.23 ng/L）已超阈值，提示需重点关注喹诺酮类抗生素的冰封期富集风险。

不同因子在不同季节表现出特异性影响模式，综合来看抗生素因子在夏季、理化因子在冬季分别占据主导地位。这种季节特异性的驱动模式揭示了寒区流域微生物群落对环境压力的差异化响应，后续研究需通过更系统的观测数据进一步验证这些发现的普适性。

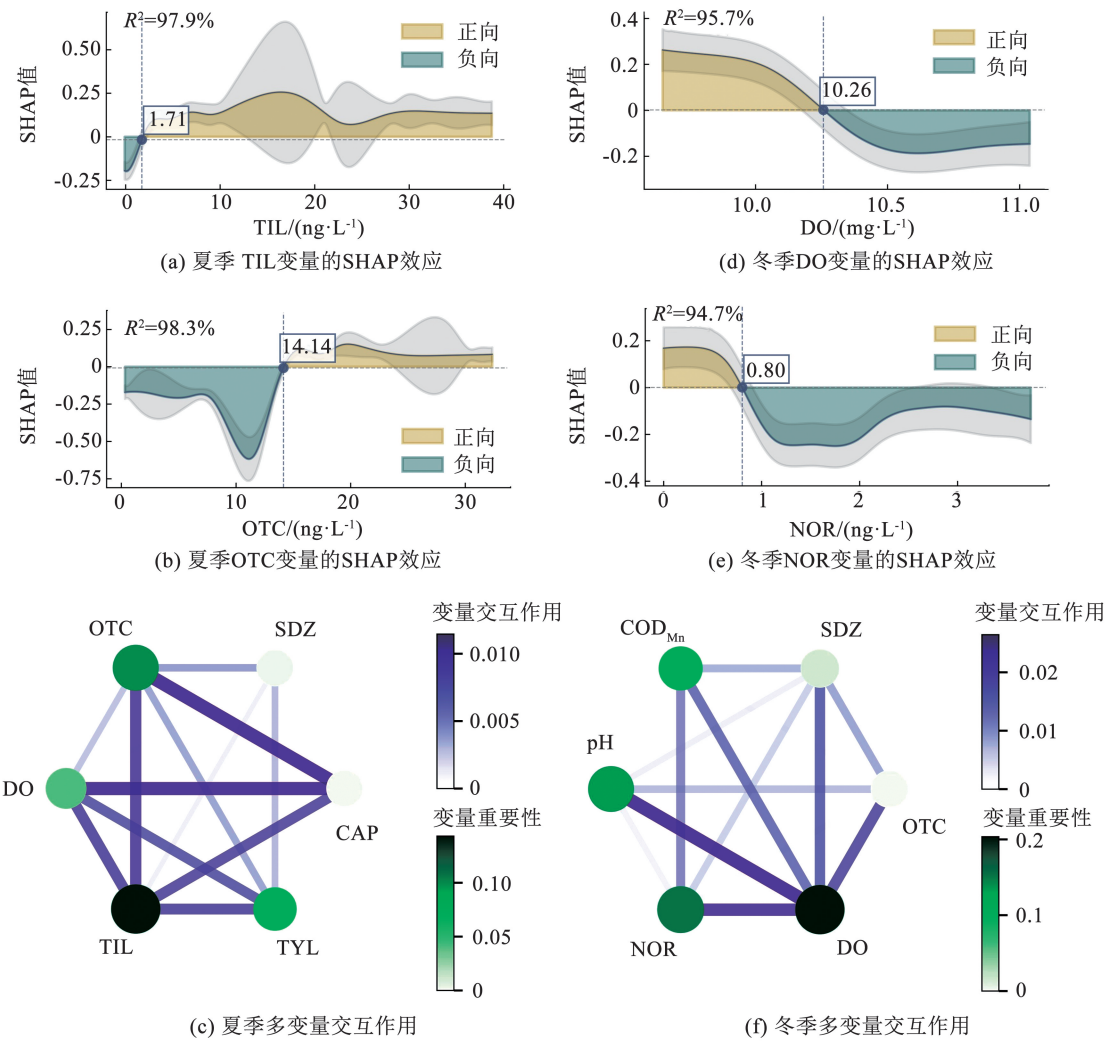


图 6　PD_whole_tree 指数预测模型的 SHAP 单变量依赖性及多变量交互作用分析

Fig. 6　Analysis of SHAP univariate dependency and multivariate interaction of PD_whole_tree index prediction model

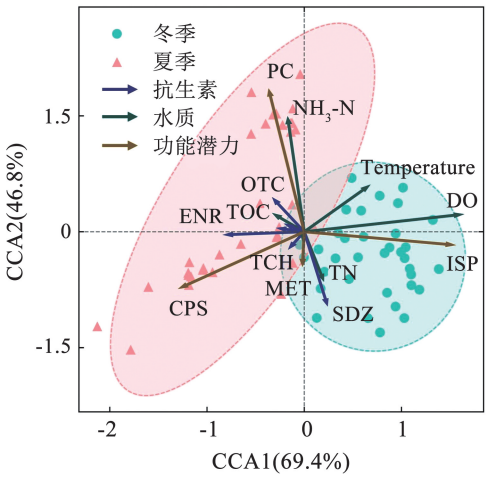
2.3.3　多变量交互作用分析

SHAP 交互分析揭示了环境因子间的协同与拮抗效应（图 5（c）、（f），图 6（c）、（f））。夏季 Shannon 指数最强交互为 ERY 与 DO（交互强度 0.001 6），当 ERY >7.6 ng/L 且 DO >7.4 mg/L 时，两因子协同抑制 Shannon 指数，叠加效应比单一因子提升 4.20%。这表明夏季抗生素污染与高 DO 环境的复合胁迫对微生物群落的负面影响大于单一压力。冬季 Shannon 指数交互以 TN 与 OTC 为主导（交互强度 0.025 5），较单一因子提升 10.80%（图 5（f））。夏季 PD_whole_tree 指数交互效应按 CAP-OTC（0.005 7）、CAP-DO（0.005 6）、OTC-TIL（0.004 8）

递减(图 6(c)),反映不同种类抗生素间的协同作用压力。冬季 PD_whole_tree 交互以 DO-pH(0.013 2)、DO-NOR(0.012 3)为核心(图 6(f)),且各交互呈协同效应,提示冬季需关注多因子复合污染的生态风险,寒区流域管理需关注多压力协同调控。

2.4 群落功能潜力的环境驱动

CCA 分析结果揭示了冬夏两季微生物群落功能潜力与环境因子的差异化关联(图 7),前两轴累计解释 116.2% 的变异。冬夏样本在排序空间中呈现明显分离,证实季节是微生物群落功能潜力差异的核心驱动因素。



注:CPS—cellular processes and signaling;ISP—information storage and processing;MET—metabolism;PC—poorly characterized。

图 7 冬夏微生物群落功能潜力的 CCA 分析

Fig. 7 CCA of microbial community function potential in winter and summer

夏季群落的细胞过程与信号传导(CPS)、代谢功能(MET)和氨氮及四环素类、喹诺酮类抗生素高度相关。夏季微生物群落代谢活性和功能多样性的提升,与高温环境下微生物活跃的物质代谢和能量转化相一致。夏季农业活动导致的营养盐和抗生素复合输入,可能诱导微生物群落的代谢重组,以适应复杂的环境压力[23]。冬季样本的群落功能潜力则与信息存储和处理(ISP)显著相关,且受温度和 DO 影响,这表明冬季微生物群落采取以遗传信息维护为核心的调控策略,通过增强 DNA 修复和基因表达调控等机制来维持细胞功能完整性,以应对低温和环境胁迫[15]。这种功能策略的季节分化体现了微生物群落对环境条件的适应性调整。

综合以上分析,该寒区流域微生物群落功能呈现“冬季水质主导-夏季抗生素调控”的季节分异模式。从生态适应视角来看,冬季微生物群落通过维持遗传信息处理功能来应对温度、DO 等水体理化指标的限制,体现遗传保守型策略;夏季微生物则在抗生素、营养盐的多重环境压力下表现出更活跃的代谢功能,通过提升代谢活性和功能多样性来适应变化的环境条件,体现代谢活跃型策略。为寒区流域分季节管控策略的制定提供了群落功能层面的科学依据。

3 结 论

1)环境变量与群落特征呈显著季节差异。冬季水体维持低温(2 ~4 ℃)、高 DO(均值 10.32 mg/L)状态,夏季受高温与农业面源污染影响,NOR 等抗生素质量浓度波动远高于冬季。微生物群落 β-多样性和功能潜力则呈显著季节分化,季节是驱动群落特征变化的关键因素。

2)群落多样性驱动因子存在季节差异。SHAP 分析表明,TN 在冬夏两季均对 Shannon 指数有重要影响,但其相对重要性存在季节差异,冬季 TN 的影响尤为突出。PD_whole_tree 指数驱动因子季节特异性更强,夏季以大环内酯类和四环素类抗生素主导,冬季以 DO 和喹诺酮类抗生素为主。模型在不同季节的表现差异,可能反映了环境因子生态梯度季节性变化的复杂性。SHAP 分析初步揭示关键环境因子对微生物多样性可能存在的浓度效应拐点(DO 夏季 7.41 mg/L、冬季 10.26 mg/L;TN 冬季 1.01 mg/L),提示维持因子浓度在一定范围内可能有利于提升微生物群落多样性。

3)多变量交互作用呈现协同效应。SHAP 交互分析表明,夏季 ERY 与 DO、冬季 TN 与 OTC 等因子组合均表现出协同效应,其交互强度最高可达单一因子效应的 10% 以上,揭示了寒区流域中水质指标与抗生素的复合压力对微生物群落具有叠加影响。

4)群落功能呈“冬季水质-夏季抗生素”驱动模式。CCA 分析表明,冬季群落信息存储与处理功能(ISP)受 DO、温度主导,体现遗传保守型适应策略;夏季细胞过程与代谢功能(CPS、MET)受多种抗生素调控,体现代谢活跃型适应策略。

本研究识别了寒区流域微生物群落的季节驱动模式与关键因子,夏季应重点关注多种抗生素排放与其复合压力,冬季应重点关注 NOR 等喹诺酮类抗生素的冰-水界面富集释放风险,同时监测 pH 和 DO 水平,为流域分季节管控提供了参考。未来研究可重点关注春季融冰期这一关键过渡阶段的环境与群落动态,结合多组学技术深入解析微生物代谢通路对环境压力的响应机制,更有效地支撑寒区流域精细化管理。此外,鉴于流域内地理水文与土地利用的空间异质性,未来可通过覆盖流域内不同地理单元的实地采样,获取更充分的样本数据,深入揭

示环境因子与微生物群落关联的空间分异规律，以提升寒区流域水生态风险预警的准确性与管理策略的适用性。

参考文献

[1] AYLAGAS E, BORJA Á, TANGHERLINI M, et al. A bacterial community-based index to assess the ecological status of estuarine and coastal environments[J]. Marine Pollution Bulletin, 2017, 114: 679. DOI:10.1016/j.marpolbul.2016.10.050

[2] CHEN Fangzhou, KOH X P, TANG M L Y, et al. Microbiological assessment of ecological status in the Pearl River Estuary, China[J]. Ecological Indicators, 2021, 130: 108084. DOI: 10.1016/j.ecolind.2021.108084

[3] JU Hanyu, LIU Ling, LIU Xuemei, et al. A comprehensive study of the source, occurrence, and spatio-seasonal dynamics of 12 target antibiotics and their potential risks in a cold semi-arid catchment[J]. Water Research, 2023, 229: 119433. DOI: 10.1016/j.watres.2022.119433

[4] ZHANG Mingyu, BAI Long, YAO Zhi, et al. Seasonal lake ice cover drives the restructuring of bacteria-archaea and bacteria-fungi interdomain ecological networks across diverse habitats[J]. Environmental Research, 2025, 269: 120907. DOI: 10.1016/j.envres.2025.120907

[5] LI Tianzhi, OUYANG Wei, LIN Chunye, et al. Occurrence, distribution, and potential ecological risks of antibiotics in a seasonal freeze-thaw basin[J]. Journal of Hazardous Materials, 2023, 459: 132301. DOI:10.1016/j.jhazmat.2023.132301

[6] KAMANMALEK S, RICE-BOAYUE J. Development of a national antibiotic multimetric index for identifying watersheds vulnerable to antibiotic pollution[J]. Environmental Pollution, 2023, 339: 122670. DOI:10.1016/j.envpol.2023.122670

[7] KARIMIAN H, ZOU Wenmin, CHEN Youliang, et al. Landscape ecological risk assessment and driving factor analysis in Dongjiang river watershed[J]. Chemosphere, 2022, 307: 135835. DOI:10.1016/j.chemosphere.2022.135835

[8] CHANG Chao, HU En, SHI Yifei, et al. Linking microbial community coalescence to ecological diversity, community assembly and species coexistence in a typical subhumid river catchment in northern China[J]. Science of the Total Environment, 2024, 938: 173367. DOI:10.1016/j.scitotenv.2024.173367

[9] KANG Jian, ZHANG Zhenyan, CHEN Yiling, et al. Machine learning predicts the impact of antibiotic properties on the composition and functioning of bacterial community in aquatic habitats[J]. Science of the Total Environment, 2022, 828: 154412. DOI:10.1016/j.scitotenv.2022.154412

[10] NOI P T, KAPPAS M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery[J]. Sensors, 2018, 18(1): 18. DOI:10.3390/s18010018

[11] WANG Longfei, YIN Haojie, LI Yi, et al. Prediction of microbial activity and abundance using interpretable machine learning models in the hyporheic zone of effluent-dominated receiving rivers[J]. Journal of Environmental Management, 2024, 357: 120627. DOI: 10.1016/j.jenvman.2024.120627

[12] ZHANG Liwen, DU Siying, ZHANG Xun, et al. Occurrence, distribution, and ecological risk of pharmaceuticals in a seasonally ice-sealed river: from ice formation to melting[J]. Journal of Hazardous Materials, 2020, 389: 122083. DOI: 10.1016/j.jhazmat.2020.122083

[13] YU Minda, LIU Sijia, LI Guowen, et al. Municipal wastewater effluent influences dissolved organic matter quality and microbial community composition in an urbanized stream[J]. Science of the Total Environment, 2020, 705: 135952. DOI: 10.1016/j.scitotenv.2019.135952

[14] BAKER K D, KELLOGG C T E, MCCLELLAND J W, et al. The genomic capabilities of microbial communities track seasonal variation in environmental conditions of Arctic Lagoons[J]. Frontiers in Microbiology, 2021, 12: 601901. DOI: 10.3389/fmicb.2021.601901

[15] LI Chunlan, PI Kunfu, VAN CAPPELLEN P, et al. Mollisol erosion-driven efflux of energetic organic carbon and microflora increases greenhouse gas emissions from cold-region rivers[J]. Environmental Science & Technology, 2024, 58: 10298. DOI:10.1021/acs.est.4c02082

[16] SABORIMANESH N, XIN Qin, RIDENOUR C, et al. Response of microbial communities in North Saskatchewan River to diluted bitumen and conventional crude under freeze-thaw-refreeze cycle[J]. Environmental Pollution, 2023, 323: 121256. DOI: 10.1016/j.envpol.2023.121256

[17] ZHANG Wengang, WU Chongzhi, ZHONG Haiyi, et al. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization[J]. Geoscience Frontiers, 2021, 12(1): 469. DOI:10.1016/j.gsf.2020.03.007

[18] QUICK A M, REEDER W J, FARRELL T B, et al. Nitrous oxide from streams and rivers: a review of primary biogeochemical pathways and environmental variables[J]. Earth-Science Reviews, 2019, 191: 224. DOI:10.1016/j.earscirev.2019.02.021

[19] ZHAO Ke, WANG Qian, QIAN Shifeng, et al. Spatial and temporal distribution characteristics of antibiotics and heavy metals in the Yitong River basin and ecological risk assessment[J]. Scientific Reports, 2023, 13: 4202. DOI:10.1038/s41598-023-31471-5

[20] CHEN Meixia, WEN Fang, WANG Hui, et al. Effect of various storage conditions on the stability of quinolones in raw milk[J]. Food Additives & Contaminants: Part A, 2016, 33(7): 1147. DOI:10.1080/19440049.2016.1184051

[21] CHEN Di, FENG Qiyan, LI Wenbo, et al. Effects of acid drainage from abandoned coal mines on the microbial community of Shandi River sediment, Shanxi Province[J]. International Journal of Coal Science & Technology, 2021, 8: 756. DOI:10.1007/s40789-021-00433-5

[22] HUANG Xiaolan, HU Baoqing, WANG Peng, et al. Microbial diversity in lake-river ecotone of Poyang Lake, China[J]. Environmental Earth Sciences, 2016, 75: 965. DOI: 10.1007/s12665-016-5473-0

[23] SHI Kun, ZHANG Jiafeng, WU Chenbin, et al. Vertical dynamics of DOM-specialized bacteria and fungi drive stability in stratified reservoirs: mechanisms revealed by machine learning[J]. Water Research, 2025, 287: 124334. DOI: 10.1016/j.watres.2025.124334

（编辑　刘　彤）