

城市雨水管理

DOI: 10.19853/j.zgjsps.1000-4602.2023.23.018

# 生物滞留系统去除典型重金属的影响因素识别

刘霖皓<sup>1</sup>, 程麒铭<sup>1</sup>, 陈 垚<sup>1,2</sup>, 袁绍春<sup>1,2</sup>, 吴 琼<sup>1</sup>

(1. 重庆交通大学 河海学院, 重庆 400074; 2. 重庆交通大学 环境水利工程重庆市工程  
实验室, 重庆 400074)

**摘 要:** 生物滞留系统具有雨水径流削减和污染控制双重功能,但其对重金属的去除效果易受设计参数和环境因素的影响而不稳定。基于文献数据驱动,利用CART算法构建了生物滞留系统设计参数和环境变量的二叉树机器学习模型,并对生物滞留系统去除Cu、Zn和Pb等典型重金属的影响因素进行识别。结果表明,影响Cu和Pb去除的最敏感因素为入流浓度,而影响Zn去除的最敏感因素为介质土深度。二叉树模型对3种重金属影响因素的识别准确率( $p_0$ )分别为0.86、0.80和0.74,分类性能均取得了中等以上的一致性,Cohen's Kappa系数( $K_a$ )分别可达到0.72、0.60和0.48。研究证实,单变量相关性分析法难以识别出生物滞留系统去除典型重金属的敏感因素,而基于文献数据驱动的机器学习方法不仅可有效挖掘出生物滞留系统中敏感因素的影响程度,还能识别出相应的阈值,可为后续优化设计和运维管理提供一定参考。

**关键词:** 生物滞留; CART算法; 机器学习; 二叉树模型; 重金属; 设计参数; 环境变量

**中图分类号:** TU992 **文献标识码:** A **文章编号:** 1000-4602(2023)23-0124-09

## Identification of Influencing Factors for Removal of Typical Heavy Metals in Bioretention System

LIU Lin-hao<sup>1</sup>, CHENG Qi-ming<sup>1</sup>, CHEN Yao<sup>1,2</sup>, YUAN Shao-chun<sup>1,2</sup>, WU Qiong<sup>1</sup>

(1. School of River and Ocean Engineering, Chongqing Jiaotong University, Chongqing 400074, China; 2. Engineering Laboratory of Environmental Hydraulic Engineering of Chongqing Municipal Development and Reform Commission, Chongqing Jiaotong University, Chongqing 400074, China)

**Abstract:** Bioretention system has dual functions of stormwater runoff reduction and pollution control. However, its removal performance of heavy metals is susceptible to be affected by design parameters and environmental variables. Based on the literature data, a binary tree machine learning model was constructed by CART algorithm for determining the design parameters and environmental variables of biological retention system, and the influencing factors of biological retention system for removal of Cu, Zn, Pb and other typical heavy metals were identified. The most sensitive factor affecting the removal of Cu and Pb was the inflow concentration, while the most sensitive factor affecting the

**基金项目:** 国家自然科学基金资助项目(51709024); 重庆市建设科技计划项目(城科学2020第5-7); 重庆市青少年创新人才培养雏鹰计划项目(CY230708); 重庆市自然科学基金资助项目(cstc2020jcyj-msxmX1000)

**通信作者:** 陈垚 E-mail: chenyaoc@cqjtu.edu.cn

removal of Zn was the depth of soil medium. The accuracy ( $p_0$ ) of the binary tree model for identification of the three heavy metal influencing factors was 0.86, 0.80 and 0.74, respectively, the classification consistency was above medium level, and the Cohen's Kappa coefficient ( $K_a$ ) was 0.72, 0.60 and 0.48, respectively. Univariate correlation analysis was difficult to identify the sensitive factors of bioretention system for the removal of typical heavy metals. In contrast, the machine learning method based on literature data could not only effectively mine the influence degree of sensitive factors in bioretention systems, but also identify the corresponding threshold, which could provide some reference for the subsequent optimization design and operation and maintenance management.

**Key words:** bioretention; CART algorithm; machine learning; binary tree model; heavy metal; design parameter; environment variable

生物滞留系统是一种源头式雨洪控制措施,已成为我国海绵城市建设中常用的低影响开发措施之一<sup>[1]</sup>。研究表明,生物滞留系统可去除雨水径流中的氮磷营养物、病原微生物和重金属等污染物<sup>[2-3]</sup>,其中对重金属的去除易受生物代谢和填料表面官能团等电子转化过程的影响而表现出不稳定,甚至发生淋溶现象<sup>[3-4]</sup>。虽然在填料介质中添加相关改良剂能提高生物滞留系统对重金属的去除能力<sup>[5]</sup>,但因受介质土类型、植物种类、水力渗透性能、雨前干旱时间(ADWP)等因素的影响<sup>[4,6]</sup>,最终造成一些强化措施或优化设计并未达到预期目标<sup>[7-8]</sup>。显然,传统的模拟实验手段难以涵盖所有设计参数和环境变量,导致研究结果难以识别出影响重金属去除效果的敏感因素,在一定程度上降低了设计优化措施的有效性,甚至可能会导致设计优化方向出现偏离。

目前,研究学者对生物滞留系统运行性能与影响因素的耦合关系模型进行了相关研究,提出了RECARGA模型<sup>[9]</sup>和基于数据驱动的建模方法<sup>[10]</sup>。但前者属于基于过程的物理模型,需要大量实验数据进行参数率定<sup>[11]</sup>,识别结果准确性受样本量的显著影响;而后者虽有效识别了影响各污染物去除的敏感因素,但其主要以TN和TP作为目标污染物进行特征性分析,且未考虑服务面积比率和积水深度等变量,并简化了植物多样性等无序分类变量。机器学习因其具有强大的数据挖掘能力,已广泛用于多变量特征识别领域<sup>[12]</sup>。其中,决策树模型属于典型的机器学习方法<sup>[13]</sup>,该方法可通过多叉树识别出敏感特征,实现特征分类与最优特征值识别<sup>[14]</sup>。但传统多叉树存在树枝规模大、生成效率低等问题,

而基于CART算法构建的二叉树模型可将多叉树剪枝成二叉树,大幅提升决策树的生成效率<sup>[14]</sup>,并且可直接处理连续性变量,使预测精度更高<sup>[13]</sup>。鉴于此,笔者基于文献数据驱动,利用CART算法构建了生物滞留系统设计参数和环境变量的二叉树机器学习模型,并通过模型学习训练与测试对生物滞留系统去除Cu、Zn和Pb等典型重金属影响因素的最优特征进行识别,确定去除各重金属的敏感影响因素及其阈值,以期对生物滞留设施的优化设计和运维管理提供参考。

## 1 研究方法

### 1.1 数据来源及预处理

利用Web of Science(核心合集数据库)对2007年—2022年发表的生物滞留相关文献进行检索,采用布尔逻辑运算符“AND”“OR”“\*”对相关关键词进行组合搜索,搜索词为:theme=((bioretention\*) OR (biofiltration\*) OR (raingarden\*) OR (rain garden\*) OR (biofilter\*)) AND (((heavy metal\*) OR (toxic metal\*)) AND (((stormwater\*) OR (storm water\*) OR (rainwater\*) OR (rain water\*))),文献类型为研究文章(article),检索年份为2007年—2022年,删除重复文献后,一共检索到110篇文献。由于本研究主要针对Cu、Zn和Pb这3种典型重金属,且从设计参数和环境变量中识别最优特征,因此文献需提供相关设计参数和环境变量的数值。经筛选剔除不满足主题或某些变量未能给出的文献后,最终得到13篇文献<sup>[8,15-26]</sup>,其中实验室研究文献10篇、野外现场研究文献3篇。利用文献提供的原始数据或利用Origin软件数据点拾取功能,分别获得3种重金属的入流和出流浓度数据。若文献仅提供入流(出流)

浓度和去除率,则通过去除率计算公式求出另一个浓度数据。同时,根据Fang等人<sup>[27]</sup>采用的生物滞留设施设计参数和环境变量划分依据,对重金属去除效果的影响因素进行分类统计,结果如表1所示。

表1 生物滞留系统去除重金属的影响因素汇总

Tab.1 Summary of factors affecting the removal of heavy metals in bioretention system

影响因素(变量)	符号	描述
设计参数	服务面积比率	Ratio 生物滞留设施表面积/集水区面积
	植物类型	PT 单子叶植物(1), 双子叶植物(2), 混合植物(3), 无植物(0)*
	介质土类型	M 腐殖土(有机质含量>5%)(0), 表层土(1), 砂介质(2)*
	介质土深度	MD 介质土设置总深度
	淹没区深度	SZD 淹没区设置深度
环境变量	积水深度	PD 设施表面积积水深度
	雨前干旱时间	ADWP 两场降雨或进水的时间间隔
	入流量	IF 径流量/设施表面积
	入流浓度	Cin 重金属进水浓度
	平均入渗速率	AIR 介质土水力渗透率

注: “\*”表示括号内的数字为无序变量分类值。

样本集统计信息见表2。

表2 样本集统计信息

Tab.2 Statistics of the sample sets

影响因素(变量)		Cu	Zn	Pb
设计参数	Ratio	0.025(0.01~0.05) <sup>#</sup>	0.03(0.01~0.05)	0.02(0.01~0.05)
	PT	0, 1, 2, 3 <sup>&amp;</sup>	0, 1, 2, 3	0, 1, 2, 3
	<i>M</i>	0, 1, 2	0, 1, 2	0, 1, 2
	MD/m	0.7(0.3~1.1)	0.7(0.3~1.1)	0.6(0.3~0.8)
	SZD/m	0(0~0.6)	0(0~0.6)	0(0~0.6)
环境变量	PD/m	0.2(0~0.5)	0.15(0~0.5)	0.15(0~0.5)
	ADWP/d	3(1~27)	3.5(1~27)	3(1~27)
	IF/mm	226(100~636)	226(100~636)	226(100~452)
	Cin/(μg·L <sup>-1</sup> )	95.1(10~1116)	587.3(100~3461)	138.9(0.88~327.1)
	AIR/(mm·h <sup>-1</sup> )	190(12.5~1224)	190(12.5~1224)	302(12.5~1224)
去除效果	出流浓度/(μg·L <sup>-1</sup> )	8.46(0.53~85.68)	10.46(0.83~363)	1.14(0.01~25.73)
	去除率/%	92.89(74.18~99)	97.48(88.9~99.4)	98.67(90~99.1)
	样本量	145	145	112

注：“#”表示中位数(最小值~最大值);“&”表示无序变量(PT、*M*)的分类数据。

由于数据来源于不同文献,对同一种重金属的去除效果可能存在较大的差异性。为避免异常数据对模型的训练结果产生影响,本研究利用IQR方

法对文献中获得的数据进行预处理,剔除明显偏离其他观测值的异常数据,最终获得用于模型训练和测试的数据集。经数据预处理后,生物滞留系统对Cu、Zn和Pb的去除率中位数分别为92.89%、97.48%、98.67%。

## 1.2 单变量相关性分析法

数据预处理后,通过单变量分析法识别生物滞留系统对3种重金属的去除率与10个变量(5个设计参数、5个环境因素)之间的相关性与显著程度,以探究单变量分析法对影响因素的识别准确性。考虑到本研究连续分布的数据集并不满足正态分布特性,更适合采用斯皮尔曼相关系数法<sup>[28]</sup>,因此,除M和PT这两个无序分类变量采用单因素方差(ANOVA,即F检验<sup>[29]</sup>)进行相关性分析外,其余8个变量采用斯皮尔曼相关系数法进行相关性分析。

## 1.3 基于CART算法的二叉树模型构建方法

根据Python软件中的SKLEARN库,利用CART算法构建二叉树模型,以识别对生物滞留系统去除重金属起主导作用的影响因子。具体构建过程如下<sup>[13]</sup>:

① 将所有样本集作为根节点(一级子节点),并以Y作为目标变量(分类变量),以 $X_i(X_1, X_2, \dots, X_p)$ 作为自变量(解释变量)。

② 通过分割函数选择其中一个解释变量 $X_i$ 及分割阈值 $S_i$ ,并将 $S_i$ 的P维空间分割为两部分空间,其中空间 $P_1$ 包含的点均满足 $X_i \leq S_i$ ,而空间 $P_2$ 中的点则满足 $X_i > S_i$ ,便可得根节点下的2个子节点(二级子节点)。

③ 通过再次选择解释变量和分割阈值,将第②步得到的每个子节点以相似的方式再划分为两部分空间(三级子节点)。

④ 重复上述步骤,直至将整个X空间尽可能地划分为互补重叠的小矩形。

为不断减少子节点的不纯度,采用Gini指数对节点不纯度进行分割。该指数表示在样本集中随机选中一个样本被分错的概率,值越小表明样本被分错的概率越小,即样本集不纯度越低,反之则表明样本集不纯度越高。Gini指数(节点不纯度)计算方法如下:

$$\text{Gini}(P) = \sum_{k=1}^K P_k(1 - P_k) = 1 - \sum_{k=1}^K P_k^2 \quad (1)$$

式中: $k$ 为类, $k=1, 2, \dots, K$ ,本研究为二分类问



题,即类别个数为2; $P_k$ 表示选中的样本属于 $k$ 类的概率,该样本被分错的概率是 $1-P_k$ 。

由于本研究的目标变量并不是重金属去除率的具体数值,而是以样本集的去除率中位数为基准,将去除率分为“高”和“低”两个类别,并进行赋值,其中,“高”去除率为类别1、“低”去除率为类别0。为将10个解释变量生成二叉树,将样本集分为训练集和测试集,并随机选择20%的样本作为测试集。由于构建的模型为典型的二分类模型,选择混淆矩阵 $M$ 来判断模型的总体准确率,并以Cohen's Kappa系数<sup>[30]</sup>作为衡量分类精度的指标,计算公式如下:

$$K_a = \frac{p_0 - p_e}{1 - p_e} \quad (2)$$

式中: $K_a$ 为Cohen's Kappa系数,用于表征分类一致性,其中, $K_a < 0$ 表示缺少一致性, $K_a$ 为0.01~0.20表示一致性较差, $K_a$ 为0.21~0.40表示一致性一般, $K_a$ 为0.41~0.60表示一致性中等, $K_a$ 为0.61~

$$p_e = \frac{(TN + FP) \times (TN + FN) + (FN + TP) \times (FP + TP)}{N \times N}$$

0.80表示一致性较强, $K_a$ 为0.81~1.00表示一致性强; $p_e$ 为机遇一致率; $p_0$ 为观察一致率(模型识别准确率)。 $p_e$ 和 $p_0$ 可通过混淆矩阵 $M$ 计算获得,如下所示:

$$M = \begin{bmatrix} TN & FP & TN + FP \\ FN & TP & FN + TP \\ TN + FN & FP + TP & N \end{bmatrix} \quad (3)$$

式中:TN为类别0中真阴性(预测结果为0,且实际结果为0)的样本数;TP为类别1中真阳性(预测结果为1,且实际结果为1)的样本数;FN为类别0中假阴性(预测结果为0,但实际结果为1)的样本数;FP为类别1中假阳性(预测结果为1,实际结果为0)的样本数;TN+FN为预测结果为0的样本数;FP+TP为预测结果为1的样本数;TN+FP为实际结果为0的样本数;FN+TP为实际结果为1的样本数; $N$ 为总样本数,且 $N=TN+TP+FN+FP$ 。根据该矩阵,可求得观察一致率 $p_0$ 和机遇一致率 $p_e$ ,分别如下:

$$p_0 = \frac{TP + TN}{N} \quad (4)$$

$$(5)$$

## 2 结果与讨论

### 2.1 单变量结果分析

斯皮尔曼相关系数分析结果表明,ADWP、Cin和AIR这3个变量与Cu、Zn和Pb的去除率均存在显著相关性,且去除率与ADWP呈负相关,而与其他2个变量呈正相关,如图1所示(“\*\*\*”表示 $P < 0.01$ ;

“\*\*”表示 $0.01 < P < 0.05$ ;“\*”表示 $0.05 < P < 0.1$ )。这是因为ADWP越长,生物滞留系统中介质土的含水率越低,在一定程度上降低了重金属在介质土中的滞留时间,减小了对重金属的截留吸附作用,从而呈现出重金属去除率与ADWP存在显著负相关的关系。

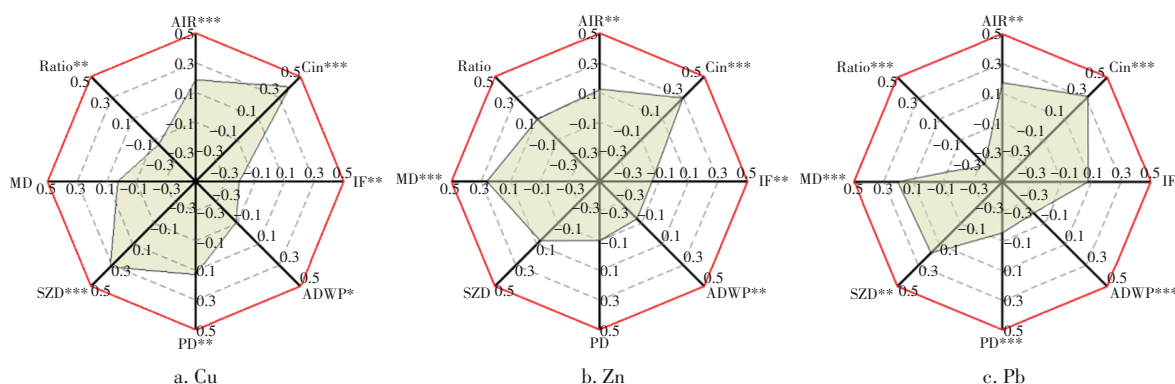


图1 重金属去除率与单变量的相关性及其显著性

Fig.1 Correlation and significance of heavy metal removal rate and univariate

生物滞留系统的Ratio变量与Cu和Pb去除率呈显著负相关( $P < 0.05$ ),而与Zn去除率呈正相关但不显著。MD变量与Zn和Pb去除率呈极显著正相关( $P < 0.01$ ),而与Cu去除率无显著相关性。PD

变量与Cu去除率呈显著正相关( $P < 0.05$ ),与Pb去除率呈极显著负相关( $P < 0.01$ ),而与Zn去除率呈负相关但不显著。IF变量与Cu和Zn去除率均呈显著负相关( $P < 0.05$ ),而与Pb去除率呈正相关但不

显著。SZD变量与Cu和Pb去除率均呈显著正相关( $P < 0.05$ ),而与Zn去除率的相关性不显著。单变量分析结果表明,同一变量对不同重金属去除效果的影响甚至出现截然不同的结果。

由图1可知,与Cu去除率存在极显著相关的变量数为3个,与Zn去除率存在极显著相关的变量数为2个,而与Pb去除率存在极显著相关的变量数高达5个,表明斯皮尔曼相关系数分析并不能准确识别出影响生物滞留系统去除重金属的敏感因素,甚至出现与研究结果相悖的情况。例如,介质土的AIR越小,滞留时间越长,介质土对重金属的截留吸附量就越大,即生物滞留系统对重金属的去除效果应与介质土AIR呈负相关,但本研究中单变量分析得出的结果却是两者呈显著正相关。同时,Ratio变量取值过大或过小均会显著影响生物滞留系统的运行性能,通过单变量分析因其无法确定阈值,使得获得的相关性并不一定可靠。此外,采用斯皮尔曼相关系数的单变量分析法将各变量相互独立,并未进行变量间的耦联分析,导致获得的结果存在一定的局限性,难以识别出影响生物滞留系统去除重金属的重要因子。

由于文献中采用的植物几乎是单子叶植物或单子叶与双子叶的混合植物,极少单独采用双子叶植物(仅1篇文献),从侧面反映了单子叶植物或混合植物在重金属去除能力上较双子叶植物更强。因此,在进行方差分析时本研究暂未考虑双子叶植物样本。无序变量PT和M方差分析结果表明(见表3),生物滞留系统对Cu的去除能力受植物类型(变量PT)变量显著影响( $P < 0.01$ ),而对Zn和Pb的去除无显著影响,这可能是由于生物滞留系统对Cu的去除主要通过植物富集作用,而在Zn和Pb的去除上植物仅起到一定的滞留作用而不是吸收富集作用<sup>[17,19]</sup>。介质土类型(变量M)并不会显著影响生物滞留系统对重金属的去除性能,这与Fang等的研究结果相似<sup>[27]</sup>。这可能是由于本研究对介质土的分类过于概化(见表1),并不能区分介质土是否含强吸附材料(如椰子壳、蛭石、珍珠岩等)等差异,而这些材料对重金属具有很强的吸附能力,可提升生物滞留系统对重金属的去除率<sup>[18,23]</sup>。此外,不同文献研究所处的实验地点不同,其土壤、砂石等理化性质也会存在较大差异。

表3 无序变量方差分析结果(P值)

Tab.3 ANOVA results for unordered variables (P value)

无序变量	Cu	Zn	Pb
PT	0.000 2***	0.88	0.19
M	0.15	0.63	0.66

## 2.2 二叉树结果分析

利用样本集数据,对构建的二叉树模型进行训练后,可获得模型对3种重金属影响因素的总体识别准确率( $p_0$ 值)和分类一致性情况( $K_a$ 值),见表4;并得到基于CART模型的生物滞留系统去除典型重金属的影响因素二叉树模型,如图2所示。

表4 模型验证结果

Tab.4 Model validation results

项 目		Cu	Zn	Pb
样本数	TN	13	9	10
	FP	2	5	4
	FN	2	1	2
	TP	12	14	7
$p_0$		0.86	0.80	0.74
$p_e$		0.50	0.51	0.50
$K_s$		0.72	0.60	0.48

影响Cu去除的最敏感变量为Cin,其余两个变量依次为SZD和ADWP,如图2所示。虽然二叉树模型识别的敏感变量数与单变量分析识别的极显著相关变量数均为3个,但两者识别的变量并不一致[见图1(a)]。同时,二叉树对敏感变量的识别准确率高, $p_0$ 为0.86,分类一致性较强( $K_a$ 为0.72)。以上结果表明,Cin决定了生物滞留系统对Cu的去除能力,该结果与单变量分析结果相一致[见图1(a)],并与Blecken等<sup>[21]</sup>研究认为生物滞留系统出水重金属平均浓度与入流浓度存在正相关这一结果相一致,但与其他相关文献研究结果相悖。如Feng等<sup>[8]</sup>和王建龙等<sup>[31]</sup>研究认为,由于生物滞留设施对重金属的去除主要发生在介质土表层位置,在设计参数固定不变的条件下,重金属出流浓度并不受其入流浓度的影响。这也进一步佐证了生物滞留系统对重金属的去除受多变量的复杂调控影响。二叉树模型结果显示,当Cin作为分割变量且阈值高于196  $\mu\text{g/L}$ 时,在节点3处Gini不纯度为0.145,被分割为类别1(高处理率)的样本比例可达92.2%,即当Cu的入流浓度高于196  $\mu\text{g/L}$ 时,生物滞留系统可取得较高的处理效能,Cu去除率高于中

位数92.89%。作为影响Cu去除的第二敏感变量,模型建议SZD的设置阈值为0.4 m,此时在节点5分割为类别1的样本比例为87.0%,Gini不纯度为0.227。SZD变量对Zn和Pb去除的影响并不敏感,这与此前文献报道的结果相一致,即设置淹没区仅能显著提升系统对Cu的去除能力,而不会改善系统对Zn或Pb的去除性能,并建议淹没区设置深度至少应为0.45 m<sup>[15,21]</sup>。ADWP也是影响系统去除Cu的敏感因素,模型建议当生物滞留设施的SZD设置值低于0.4 m时,场次间隔降雨时间应控制在24 d

内,才能实现较高的Cu去除率(类别1的样本比例为88.9%),与单变量分析结果一致。这与Blecken等<sup>[20]</sup>研究认为生物滞留设施经历3~4周的干旱期后其运行性能会显著恶化的结论相一致。因此在实际运行过程中,可通过人为浇灌来保持土壤一定的含水率,确保生物滞留设施不会因干旱期过长而影响根际微生物的活动和植物根系生长,并避免后续湿润过程中发生优先流现象而改变污染物的运移路径<sup>[32]</sup>,造成重金属在介质土内滞留时间缩短,从而导致除污性能发生恶化。

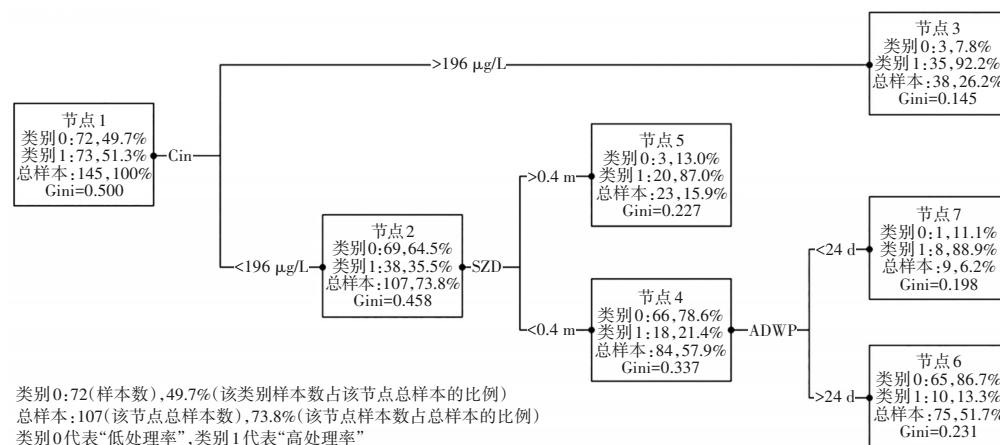


图2 重金属Cu的二叉树模型结果

Fig.2 Binary tree model results for heavy metal Cu

影响Zn去除的最敏感变量为MD,其余两个变量依次为IF和ADWP。而单变量分析识别的极显著变量仅有2个,分别为Cin和MD变量[见图1(b)]。同时,二叉树对重要变量的识别准确率高, $p_0$ 为0.80,但分类一致性中等( $K_a$ 为0.60)。当MD作为分割变量且阈值大于0.95 m时,在节点3处Gini不纯度为0,被分割的样本全部属于类别0(低处理率),即介质土设置深度超过0.95 m时,生物滞留系统对Zn的去除能力较差,去除率低于中位数97.48%。而在节点2处,类别1的样本比例为70.2%,因此模型建议MD不宜大于0.95 m。这是因为Zn等重金属的去除主要发生在介质土表层,如距顶部0.3 m区域内<sup>[23]</sup>。若介质土设置过深,反而会造成Zn、Pb等重金属在介质土中发生浸溶和迁移现象而导致出流浓度升高;同时研究还证实,当MD从0.7 m降至0.3 m时,生物滞留系统中Zn出流浓度也会发生剧增<sup>[8]</sup>。这可能是由于介质土过浅导致重金属滞留时间过短而影响介质土对Zn的吸附过程,因此MD也不宜过小,建议控制在0.70~0.95 m。

当IF作为分割变量时,模型得出当IF>189 mm时,生物滞留系统才能实现较高的Zn去除率,类别1的样本比例为83.7%。这是因为雨水径流中的重金属主要来自下垫面的冲刷过程,其浓度与径流量存在正相关性<sup>[33]</sup>。对于同一个生物滞留设施而言,径流量越大其IF值也越大,使得进入生物滞留系统的Zn浓度也越高。单变量分析表明,Zn去除率与Cin变量呈极显著正相关[见图1(b)],从而呈现出较高的IF值反而具有“高”去除率。同时,IF变量不仅取决于服务汇水区域内的径流量,还受生物滞留设施尺寸规模的影响,即该变量既属于环境变量,又属于设计参数,使IF变量对Zn的去除呈现出多重影响,且影响程度还受气候条件的显著影响。如规模尺寸过大(或过小)的系统在自然条件下难免会遭受长期干旱(或寒冷气候)等恶劣气候条件的冲击,进而对系统的水力性能和除污性能产生不同程度的影响,甚至是长期的影响<sup>[16,19-20,34]</sup>。当ADWP作为分割变量且阈值低于17.5 d时,节点6处类别1的样本比例高达87.8%,Gini不纯度为0.214,再次



佐证Cu、Zn等重金属去除率与ADWP存在显著负相关,即重金属去除率随ADWP的减少而增强。此外,李鹏的研究也证实ADWP是影响生物滞留系统去除Zn的重要因素<sup>[35]</sup>。

影响Pb去除的最敏感变量为Cin,其余两个变量依次为AIR和Ratio。相较于单变量分析识别的5个极显著相关变量[见图1(c)],AIR变量并未被单变量分析法识别出。同时,二叉树对重要变量的识别准确率较高, $p_0$ 为0.74,分类一致性中等( $K_a$ 为0.48)。根据二叉树模型结果,当以Cin作为分割变量时,Pb的入流浓度应大于46.8  $\mu\text{g/L}$ ,此时节点3的样本均为类别0,Gini不纯度为0。在节点2处,以AIR变量(阈值为200 mm/h)进行分割时,节点4中类别1的样本数占比为77.6%,即模型建议生物滞留系统的AIR应控制在200 mm/h内。分析认为,AIR变量与M变量(即介质土类型)密切相关,其受介质土的有机质含量、含沙量等理化性质的影响,较低的入渗率反而可以为介质土与金属离子提供更多的接触时间,进而强化生物滞留系统对重金属的去除,并影响重金属在介质土内部的竖向迁移过程<sup>[36]</sup>。当以Ratio作为分割变量时,模型识别的分隔阈值为0.022,并应将生物滞留系统的Ratio变量值控制在0.022以上,可获得100%的类别1样本(节点7的Gini不纯度为0),这与Bratieres等<sup>[37]</sup>研究认为生物滞留系统的规模尺寸不应低于其服务汇水面积的2%这一结论相一致。考虑到生物滞留设施建设成本问题,在实际建设过程中建议将Ratio值控制在0.022~0.040范围内。

综上,相较于单变量分析法,基于CART算法的二叉树模型不仅能充分考虑变量间的耦联关系,通过机器学习过程有效挖掘出敏感变量,并确定敏感变量的影响程度,还能清晰展示出影响生物滞留系统去除典型重金属相关敏感变量的控制阈值,这有助于后续的优化设计和运维管理。

### 3 结论

① 单变量斯皮尔曼相关系数及显著性检验表明,生物滞留系统对Cu的去除率与3个变量存在极显著相关,对Zn的去除率与2个变量存在极显著相关,而对Pb的去除率则与5个变量存在极显著相关,且均与入流浓度(Cin)变量存在极显著相关。该方法并不能有效识别出最敏感变量,且尚未考虑各

变量间的耦联作用。

② 基于CART算法的二叉树模型显示,影响Cu去除的敏感变量依次为Cin、淹没区深度(SZD)和雨前干旱时间(ADWP),变量的控制阈值分别为Cin>196  $\mu\text{g/L}$ 、SZD>0.4 m、ADWP<24 d;影响Zn去除的敏感变量依次为介质土深度(MD)、入流量(IF)和ADWP,变量的控制阈值分别为MD<0.95 m(建议值为0.70~0.95 m)、IF>189 mm、ADWP<17.5 d;影响Pb去除的敏感变量依次为Cin、平均入渗率(AIR)和服务面积比率(Ratio),变量的控制阈值分别为Cin>46.8  $\mu\text{g/L}$ 、AIR<200 mm/h、Ratio>0.022(建议值为0.022~0.040)。同时,模型对Cu、Zn、Pb这3种典型重金属去除率影响因素的识别准确率( $p_0$ )分别为0.86、0.80和0.74,分类性能均取得了中等以上的一致性,Cohen's Kappa系数( $K_a$ )分别可达到0.72、0.60和0.48。

③ 研究结果证实,相较于单变量的斯皮尔曼相关系数法,基于CART算法的二叉树模型不仅考虑了变量间的联系,还能够有效识别出敏感变量的影响程度,并通过变量特征的求解给出相应变量的最佳阈值,可通过这些阈值为生物滞留设施的优化设计和运维管理提供一些建议和科学依据。但本研究构建的模型在机器学习过程中因样本集数据量有限,在一定程度上影响了分类结果的准确性,导致部分节点的Gini不纯度值较大。

### 参考文献:

- [1] YIN D K, CHEN Y, JIA H F, *et al.* Sponge city practice in China: a review of construction, assessment, operational and maintenance [J]. *Journal of Cleaner Production*, 2021, 280: 124963.
- [2] 王亚军,耿冲冲,许妍,等. 不同强化手段对生物滞留池脱氮除磷性能的影响 [J]. *中国给水排水*, 2020, 36(19): 77-82.  
WANG Yajun, GENG Chongchong, XU Yan, *et al.* Effect of different enhanced methods on efficiency of denitrification and phosphorus removal in bioretention cell [J]. *China Water & Wastewater*, 2020, 36(19): 77-82 (in Chinese).
- [3] LEFEVRE G H, PAUS K H, NATARAJAN P, *et al.* Review of dissolved pollutants in urban storm water and their removal and fate in bioretention cells [J]. *Journal of Environmental Engineering*, 2015, 141 (1):

- 04014050.
- [4] 黄恒粤, 陈垚, 袁绍春, 等. 基于文献分析的生物滞留研究现状与展望 [J]. 环境污染与防治, 2022, 44 (5): 675–681.
- HUANG Hengyue, CHEN Yao, YUAN Shaochun, *et al.* Advances and prospect in the studies on bioretention based on related literatures analysis [J]. Environmental Pollution & Control, 2022, 44 (5): 675–681 (in Chinese).
- [5] LI Y J, WEN M, LI J K, *et al.* Reduction and accumulative characteristics of dissolved heavy metals in modified bioretention media [J]. Water, 2018, 10 (10): 1488.
- [6] 李迪, 陈垚, 吕波. 生物滞留系统对溶解性污染物的去除特性及优化途径 [J]. 环境工程, 2020, 38(10): 120–127.
- LI Di, CHEN Yao, LÜ Bo. Characteristics and optimization approach for removal of dissolved pollutants in bioretention systems [J]. Environmental Engineering, 2020, 38(10): 120–127 (in Chinese).
- [7] READ J, WEVILL T, FLETCHER T, *et al.* Variation among plant species in pollutant removal from stormwater in biofiltration systems [J]. Water Research, 2008, 42(4/5): 893–902.
- [8] FENG W, HATT B E, MCCARTHY D T, *et al.* Biofilters for stormwater harvesting: understanding the treatment performance of key metals that pose a risk for water use [J]. Environmental Science & Technology, 2012, 46(9): 5100–5108.
- [9] 涂安国, 李英, 莫明浩, 等. 基于RECARGA模型生物滞留池设计参数的水文效应 [J]. 水土保持学报, 2020, 34(1): 149–153.
- TU Anguo, LI Ying, MO Minghao, *et al.* Hydrological effects of design parameters optimization of bioretention facility based on RECARGA model [J]. Journal of Soil and Water Conservation, 2020, 34(1): 149–153 (in Chinese).
- [10] ZHANG K, LIU Y, DELETIC A, *et al.* The impact of stormwater biofilter design and operational variables on nutrient removal—a statistical modelling approach [J]. Water Research, 2021, 188: 116486.
- [11] KHAN U T, VALEO C, CHU A, *et al.* A data driven approach to bioretention cell performance: prediction and design [J]. Water, 2013, 5(1): 13–28.
- [12] 芮栋妮, 马燕燕, 叶林. 机器学习方法在污水处理系统中的应用 [J]. 环境工程, 2022, 40(6): 145–153.
- RUI Dongni, MA Yanyan, YE Lin. Application of machine learning methods in wastewater treatment systems [J]. Environmental Engineering, 2022, 40 (6): 145–153 (in Chinese).
- [13] 陈平, 徐星. 基于CART算法的带钢抗拉强度影响因素研究 [J]. 控制工程, 2015, 22(2): 276–281.
- CHEN Ping, XU Xing. Research of tensile strength of strip steel based on CART [J]. Control Engineering of China, 2015, 22(2): 276–281 (in Chinese).
- [14] 常辉, 胡修林, 张蕴玉. 基于CART算法的卫星星座原始构型选择策略 [J]. 华中科技大学学报(自然科学版), 2011, 39(6): 1–5.
- CHANG Hui, HU Xiulin, ZHANG Yunyu. Strategy of selecting original configuration for satellite constellation using CART algorithm [J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2011, 39(6): 1–5 (in Chinese).
- [15] WANG M, ZHANG M, LI Y, *et al.* Effect of a submerged zone and carbon source on nutrient and metal removal for stormwater by bioretention cells [J]. Water, 2018, 10: 1629–1642.
- [16] VALTANEN M, SILLANPAA N, SETALA H. A large-scale lysimeter study of stormwater biofiltration under cold climatic conditions [J]. Ecological Engineering, 2017, 100: 89–98.
- [17] MUTHANNA TM, VIKLANDER M, GJESDAHL N, *et al.* Heavy metal removal in cold climate bioretention [J]. Water, Air, & Soil Pollution, 2007, 183: 391–402.
- [18] LIM H S, LIM W, HU J Y, *et al.* Comparison of filter media materials for heavy metal removal from urban stormwater runoff using biofiltration systems [J]. Journal of Environmental Management, 2015, 147: 24–33.
- [19] BLECKEN G, MARSALEK J, VIKLANDER M. Laboratory study of stormwater biofiltration in low temperatures: total and dissolved metal removals and fates [J]. Water, Air, & Soil Pollution, 2011, 219: 303–317.
- [20] BLECKEN G T, ZINGER Y, DELETIC A, *et al.* Influence of intermittent wetting and drying conditions on heavy metal removal by stormwater biofilters [J]. Water Research, 2009, 43(18): 4590–4598.
- [21] BLECKEN G T, ZINGER Y, DELETIC A, *et al.* Impact of a submerged zone and a carbon source on heavy metal removal in stormwater biofilters [J]. Ecological Engineering, 2009, 35(5): 769–778.



- [22] GOOD J F, O' SULLIVAN A D, WICKE D, *et al.* Contaminant removal and hydraulic conductivity of laboratory rain garden systems for stormwater treatment [J]. *Water Science and Technology*, 2012, 65(12): 2154-2161.
- [23] HATT B E, DELETIC A, FLETCHER T D. Stormwater reuse: designing biofiltration systems for reliable treatment [J]. *Water Science and Technology*, 2007, 55(4): 201-209.
- [24] HATT B E, FLETCHER T D, DELETIC A. Hydrologic and pollutant removal performance of stormwater biofiltration systems at the field scale [J]. *Journal of Hydrology*, 2008, 365(3/4): 310-321.
- [25] HERMAWAN A A, TALEI A, LEONG J Y C, *et al.* Performance assessment of a laboratory scale prototype biofiltration system in tropical region [J]. *Sustainability*, 2019, 11: 1947.
- [26] WANG J L, ZHAO J L, YANG L Q, *et al.* Removal of heavy metals from urban stormwater runoff using bioretention media mix [J]. *Water*, 2017, 9: 854-876.
- [27] FANG H, JAMALI B, DELETIC A, *et al.* Machine learning approaches for predicting the performance of stormwater biofilters in heavy metal removal and risk mitigation [J]. *Water Research*, 2021, 200: 117273.
- [28] XIAO C, YE J, ESTEVES R M, *et al.* Using Spearman's correlation coefficients for exploratory data analysis on big dataset [J]. *Concurrency and Computation: Practice and Experience*, 2016, 28(14): 3866-3878.
- [29] 戴金辉. 单因素方差分析中异方差的检验与修正 [J]. *统计与决策*, 2017(8): 23-26.  
DAI Jinhui. Test and correction of heteroscedasticity in the single-factor variance analysis [J]. *Statistics & Decision*, 2017(8): 23-26 (in Chinese).
- [30] COHEN J. A coefficient of agreement for nominal scales [J]. *Educational and Psychological Measurement*, 1960, 20(1): 37-46.
- [31] 王建龙, 杨丽琼, 黄涛. 复合生物滞留介质对雨水径流中重金属净化效果 [J]. *环境工程学报*, 2015, 9(7): 3086-3092.  
WANG Jianlong, YANG Liqiong, HUANG Tao. Effect of composite bioretention media on heavy metals removal from urban stormwater runoff [J]. *Chinese Journal of Environmental Engineering*, 2015, 9(7): 3086-3092 (in Chinese).
- [32] 陈焱, 程启洪, 甘春娟, 等. 雨水生物滞留系统氮素过程研究现状与展望 [J]. *环境工程*, 2018, 36(7): 11-16.  
CHEN Yao, CHENG Qihong, GAN Chunjuan, *et al.* Research and prospect on processes of nitrogen removal in bioretention systems [J]. *Environmental Engineering*, 2018, 36(7): 11-16 (in Chinese).
- [33] 骆辉, 章泽宇, 胡小波, 等. 道路雨水径流中重金属特征研究进展 [J]. *应用化工*, 2019, 48(10): 2456-2461.  
LUO Hui, ZHANG Zeyu, HU Xiaobo, *et al.* Study on the characteristics of heavy metals in road stormwater runoff [J]. *Applied Chemical Industry*, 2019, 48(10): 2456-2461 (in Chinese).
- [34] GEHENIAU N, FUAMBA M, MAHAUT V, *et al.* Monitoring of a rain garden in cold climate: case study of a parking lot near Montréal [J]. *Journal of Irrigation and Drainage Engineering*, 2015, 141(6): 04014073.
- [35] 李鹏. 生物滞留系统对重金属的净化效果及出水回用风险评估 [D]. 西安: 西安理工大学, 2017.  
LI Peng. Research on the Purification Efficiency of Bioretention System for Heavy Metal Pollutions and Risk Assessment of Effluent Reuse [D]. Xi'an: Xi'an University of Technology, 2017 (in Chinese).
- [36] SAUVE S, HENDERSHOT W, ALLEN H E, *et al.* Solid-solution partitioning of metals in contaminated soils: dependence on pH, total metal burden, and organic matter [J]. *Environmental Science & Technology*, 2000, 34(7): 1125-1131.
- [37] BRATIERES K, FLETCHER T D, DELETIC A, *et al.* Nutrient and sediment removal by stormwater biofilters: a large-scale design optimisation study [J]. *Water Research*, 2008, 42(14): 3930-3940.

作者简介: 刘霖皓(1998- ), 男, 重庆人, 硕士研究生,  
主要从事水污染防控技术研究。

E-mail: 916928407@qq.com

收稿日期: 2022-08-21

修回日期: 2022-10-24

(编辑: 刘贵春)