

基于随机森林方法研究鄱阳湖典型洲滩植被群落分布与表层土壤因子耦合关系^①

郑利林^{1,2}, 徐金英^{1,2}, 王晓龙^{1*}, 刘宝贵^{1,2}

(1 中国科学院流域地理学重点实验室, 中国科学院南京地理与湖泊研究所, 南京 210008; 2 中国科学院大学, 北京 100049)

摘要: 准确识别湿地植被与土壤的相互作用是湿地恢复和保护的重要前提。水文情势是影响鄱阳湖地区植被分布的关键因子, 而植被分布格局则会对湿地土壤养分累积与赋存形态产生影响。本文利用优化的随机森林(Random Forests)算法, 基于多环境变量预测了鄱阳湖典型低滩植物(虉草, *Phalaris arundinacea* Linn)和高滩植物(南荻, *Triarrhena lutarioriparia* L. Liu)的分布, 进而分析这两种植被表层土壤养分累积差异。结果表明: 随机森林模型对虉草和南荻预测的精度分别达到了 89.6% 和 89.3%。模型给出了土壤因子的重要性排序。按土壤因子与虉草分布的密切相关程度, 重要性依次为全钾、氨氮、有机质、含水率、全氮、有效磷、全磷、pH 和硝态氮; 按土壤因子与南荻分布的密切相关程度, 重要性依次为全钾、pH、有机质、全氮、全磷、硝态氮、氨氮、含水率和有效磷。从植被分布与土壤因子的偏依赖图中可得出, 南荻分布区较虉草分布区土壤酸性更强; 貉草分布与土壤全氮、氨氮含量呈负相关关系, 南荻分布则与土壤全氮含量呈正相关关系, 而与氨氮关系不显著; 貉草分布与土壤全磷含量正相关, 而南荻则与全磷负相关; 貉草和南荻与土壤有效磷相互作用关系较弱; 此外虉草分布区钾含量低, 二者负相关, 而南荻分布区钾含量高, 二者正相关。随机森林方法适用于模拟复杂的非线性关系, 给出了单个土壤因子与植被之间的偏依赖关系, 易于给出生态学意义上的解释, 在研究湿地植被与环境因素的相互作用关系中有极大的推广价值。

关键词: 貉草; 南荻; 土壤因子; 随机森林

中图分类号: K903 文献标志码: A

Study on Relationships Between Vegetation Community Distribution and Topsoil Factors Based on Random Forests in Shoaly Wetlands of Poyang Lake

ZHENG Lilin^{1,2}, XU Jinying^{1,2}, WANG Xiaolong^{1*}, LIU Baogui^{1,2}

(1 Key Laboratory of Watershed Geographic Sciences, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China; 2 University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Accurate identification of the interaction between vegetation and soil is an important premise of wetland restoration and protection. Hydrologic situation is the key factor affecting the distribution of wetland vegetation, while vegetation distribution pattern can affect the accumulation and occurrence form of soil nutrients. Based on soil environmental factors, optimized Random Forests was used to predict the distribution of *Phalaris arundinacea* Linn and *Triarrhena lutarioriparia* L. Liu, which grow in low elevation and high elevation of shoaly wetlands of Poyang Lake respectively. And then we analyzed soil nutrient accumulation under the two vegetation. Results showed that the predication accuracy by Random Forest was 89.6% for *Phalaris arundinacea* Linn and 89.3% for *Triarrhena lutarioriparia* L. Liu. According to the model, The importance of soil factors which closely related to *Phalaris arundinacea* Linn was in order of total potassium > ammonia nitrogen > organic matter > soil water content > total nitrogen > available phosphorus > total phosphorus > pH > nitrate nitrogen, and that for *Triarrhena lutarioriparia* L. Liu was in order of total potassium > pH > organic matter > total nitrogen > total phosphorus > nitrate nitrogen > ammonia nitrogen > soil water content > available phosphorus. From the partial dependent plot, pH value under *Phalaris*

^①基金项目: 中国科学院科技服务网络计划(STS)重点项目(KFJ-STS-ZDTP-011)、国家科技基础性工作专项(2013FY111800)和国家自然科学基金项目(41171024)资助。

* 通讯作者(wangxl@niglas.ac.cn)

作者简介: 郑利林(1994—), 女, 安徽黄山人, 硕士研究生, 主要从事湿地生态研究。E-mail: zhenglilin17@mails.ucas.ac.cn

arundinacea Linn was significantly higher than that under *Triarrhena lutarioriparia* L. Liu. Distribution of *Phalaris arundinacea* Linn was negatively correlated to total nitrogen and ammonia nitrogen; distribution of *Triarrhena lutarioriparia* L. Liu was positively correlated to total nitrogen while weak relationship between *Triarrhena lutarioriparia* L. Liu and ammonia nitrogen was found. Total phosphorus was positively correlated to *Phalaris arundinacea* Linn while negatively correlated to *Triarrhena lutarioriparia* L. Liu. Weak relationships were found between available phosphorus and the two kinds of vegetation. Total potassium was negatively correlated with *Phalaris arundinacea* Linn while positively correlated with *Triarrhena lutarioriparia* L. Liu. Soil water content was positively correlated to *Phalaris arundinacea* Linn while negatively correlated to *Triarrhena lutarioriparia* L. Liu. Random Forests is suitable for simulating complex nonlinear relation, and can show the partial dependence relationship between individual soil factors and vegetation, so can explain the results in the ecological sense. Random Forests is of great value in the study of the interaction between wetland vegetation and environmental factors.

Key words: *Phalaris arundinacea* Linn; *Triarrhena lutarioriparia* L. Liu; Soil factors; Random Forests

湿地土壤、水文、植被与气候等因素在不同的时空尺度上相互作用,共同影响湿地生态系统过程与功能^[1]。湿地水文过程主导着湿地生态系统的格局和过程^[2],是制约湿地植被生长和繁殖的关键因素^[3-4]。湿地土壤是长期以来湿地生态系统各因素相互作用的产物^[5-6]。土壤因子与湿地植被的相关作用关系一直是湿地研究的热点问题。已有的研究表明土壤养分与植被类型、植被发育年限、群落演替阶段等相关。张天雨等^[7]研究表明长江口典型滨海湿地盐沼植被类型显著影响土壤各层次总有机碳和全氮储量的分布。张文敏等^[8]的研究表明杭州湾湿地互花米草的入侵增强了滩涂湿地的固碳能力。张骞等^[9]的研究表明随长江口典型潮滩湿地崇明东滩植被发育年限增加,土壤有机碳含量逐年增加,植被发育 14 a 的土壤有机碳含量约为发育 1 a 的 1.3 倍。罗琰等^[10]研究了辉河湿地不同演替阶段的草甸植物群落中同种元素含量的差异,认为过渡草甸对土壤碳、氮、磷的吸收和滞留能力最强。其中植被类型与土壤养分的关系是研究的热点,已有的研究多基于不同的土壤类型^[11]、土地利用类型^[12-13]、高程梯度^[14]等选择代表性的植被类型研究植被与土壤因子的关系。现有的研究较少涉及湿地不同高程代表性植被与土壤的作用关系。

判别植被在特定生境下是否存在经典方法是逻辑斯蒂回归。逻辑斯蒂回归不能很好地处理大量特征变量;对自变量多重共线性敏感,不能很好地处理缺失值^[15]。2001 年, Breiman^[16]和 Cutler 等^[17]借鉴 Ho^[18]提出的随机决策森林方法,把分类树组合成随机森林,生成很多分类树,再汇总分类树的结果。随机森林可以处理复杂的较多变量间的非线性关系,对缺失值、变量多元共线性、异常值不敏感。机器学习方法不能给出预测变量与响应变量的简单表达式,导致生态学解释上的困难性。但不同于一般的“黑箱”

模型,随机森林可以给出变量的重要性排序以及变量间的偏依赖关系,从而能识别重要的生态因子并进行解释。Prasad 等^[19]认为最好把随机森林称为“灰箱”模型。近年来,随机森林算法在生态学方面的应用逐渐增长^[20],然而在处理物种分布数据时,随机森林通常在少数类物种的预测中表现不佳^[21]。可以基于抽样的方法修改训练数据集增加少数类的比重以解决这一问题。2002 年 Chawla 等^[22]提出的 SMOTE (Synthetic Minority Over-sampling Technique) 算法是较为成熟的小样本重采样技术,SMOTE 算法的基本思想是对少数类样本进行分析并根据少数类样本人工合成新样本添加到数据集中,增加少数样本在抽样过程中的比重。SOMTE 和随机森林相结合在植被分布预测中鲜少有相关研究。但已有的研究表明,二者的结合可以有效地处理和分析非平衡数据^[21,23]。

鄱阳湖位于江西省北部,是我国最大的淡水湖泊。鄱阳湖湿地是中国首批公布的国家重点湿地保护地之一,1992 年被列入《国际重要湿地名录》^[24]。鄱阳湖承接上游赣、抚、信、饶、修五河之水,由湖口北注长江。由于受到五河流域水系和长江水位的共同影响,鄱阳湖水位年内差异极大,周期性的水位高程变化使鄱阳湖洲滩形成了特定的生境梯度^[25]。鄱阳湖年内水位的剧烈波动使鄱阳湖洲滩植被经历着出露-淹没-出露的过程^[26]。不同高程洲滩出露时间的差异,使得鄱阳湖湿地洲滩植被呈现明显的带状、弧状或环带状分布^[27]。不同的植物群落带对土壤养分的累积差异进一步加剧了植被生境异质性和梯度性。南荻(*Triarrhena lutarioriparia* L. Liu)和虉草(*Phalaris arundinacea* Linn)是鄱阳湖洲滩湿地的主要建群种。其中南荻是多年生竹状草本,具十分发达的根状茎,分布高程相对较高,多与芦苇混生,为典型高滩植物;虉草则分布高程相对较低,多靠近湖心分

布, 为典型低滩植被。已有的研究主要基于排序或相关性分析研究鄱阳湖湿地植被生长繁殖及其与土壤因子的相关关系^[1,25,28-31], 较少涉及鄱阳湖高滩植被和低滩植被与土壤因子的耦合关系研究。比较以南荻为代表的高滩植被和以虉草为代表的低滩植被与土壤因子的相互作用对探究鄱阳湖湿地生态系统格局具有重要意义。

本文将土壤环境因子作为预测变量, 利用优化的随机森林分类模型对鄱阳湖湿地以南荻为代表的高滩植被和以虉草为代表的低滩植被在样方中是否存在进行判别。基于此找到这两种植被分布下累积的土壤养分差异, 分析这两种植被与各土壤因子相互作用的程度和方向。以期为鄱阳湖生态系统的管理、保护以及可持续发展提供一定的数据支持。

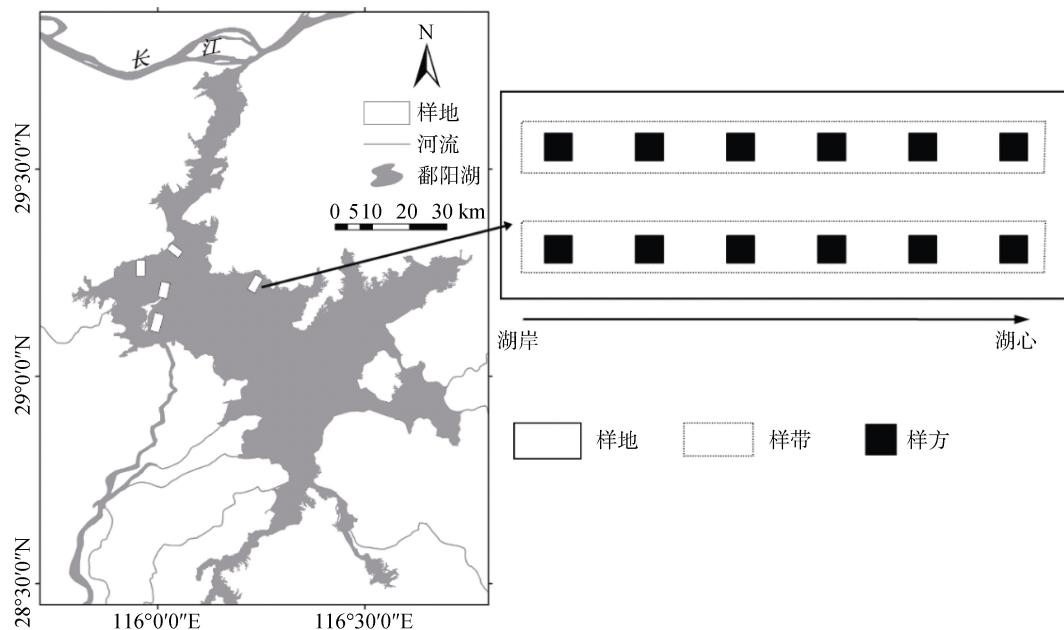


图 1 鄱阳湖典型湿地植被调查样方分布示意图

Fig.1 Distribution vegetation survey quadrats in typical wetlands of Poyang Lake

1.2 调查与采样

于 2017 年春季, 选取了研究区域 5 个样地, 按照垂直于湖岸线的地形梯度进行样带设置。在每个样地中确定 2 条 3 m 宽的样带, 样带之间间距为 200 m。

1 研究区域与研究方法

1.1 鄱阳湖典型洲滩概况

本研究选取的洲滩位于鄱阳湖主湖北部(图), 属于亚热带季风气候区, 年平均气温在 17 ℃左右, 年降水量平均为 1 500 mm。降水年内分配极不均匀, 4—6 月降水约占全年的 48%, 6 月最大占全年 17%, 12 月最小只占 3%。丰水期(4—9 月)洲滩几乎全部淹没, 10 月至次年 3 月为枯水期, 进入枯水期后, 不同高程的洲滩相继出露。鄱阳湖植被分布主要受到水文情势的影响, 沿泥滩到阶地, 主要的植物群落分别为狗牙根群落、芦苇-南荻群落、薹草群落、刚毛荸荠群落、虉草-蓼子草-水田碎米荠群落、苔菜-竹叶眼子菜-轮叶黑藻-苦草群落^[31]。

根据湖岸到湖心水文梯度的变化, 在每条样带上等距离布设 6 个 1 m × 1 m 的样方(图 1), 共布设 60 个样方。记录各样方的主要建群种和伴生种。其中虉草群落和南荻群落的建群种和伴生种如下表(表 1)所示。

表 1 鄱阳湖典型洲滩虉草群落和南荻群落的建群种和伴生种

Table 1 Dominant and companion species of *Phalaris arundinacea* Linn and *Triarrhena lutarioriparia* L. Liu communities in typical beaches in Poyang Lake area

植物群落	建群种	伴生种
虉草群落 <i>Ass. Phalaris arundinacea</i> Linn	虉草 (<i>Phalaris arundinacea</i> Linn)	灰化薹(<i>Carex cinerascens</i> Kükenth.)、 蓠蒿、水田碎米荠(<i>C. lyrate</i> Bunge)
南荻群落 <i>Ass. Triarrhena lutarioriparia</i> L. Liu	南荻 (<i>Triarrhena lutarioriparia</i> L. Liu)	灰化薹(<i>Carex cinerascens</i> Kükenth.)、 蓠蒿、下江委陵菜(<i>P. imprichtii</i> J. Krause)、 水蓼(<i>P. hydropiper</i> L)、狗牙根(<i>Cynodon</i>)

1.3 土壤样品测定方法

对植被样方进行调查的同时,对样方内土壤进行采样。取土样时先去除土壤表层凋落物,用土钻(内径5 cm)对表层(0~15 cm)土壤进行取样,将同个样方随机选取的3~5个土样混合(约500 g),装入无菌塑料袋中并封口,共采集土样60份,置于有冰袋的保鲜盒中运回实验室。测定土壤养分含量,包括土壤pH、全氮、氨氮、硝态氮、全磷、有效磷、全钾、有机质、含水率。pH采用无CO₂水浸提,酸度计测定,土:水为1:2.5;土壤全氮含量测定采用凯氏定氮法测定;土壤氨氮、硝态氮含量采用KCl溶液浸提,流动注射分析仪测定;土壤全磷含量采用硫酸-高氯酸消煮,钼锑抗比色法测定;土壤有效磷采用氟化铵-盐酸浸提法测定;土壤全钾含量采用碳酸钠熔融法测定;有机质采用稀释热法测定;含水率采用105℃烘干法测定^[32]。

1.4 数据处理和分析方法

本文将植被南荻和藨草在某样方中是否存在作为因变量(二分变量),把土壤养分状况(全氮、氨氮、硝态氮、全磷、有效磷、全钾、有机质)、pH、含水率、洲滩类型和植被样方的经纬度作为自变量,建立了随机森林判分类模型。在自变量和样本量确定的情况下,可以通过改变分类节点m和迭代次数k减小袋外误差(OOB)。袋外误差是分类误差的无偏估计^[16]。Beriman^[16]和Cutler等^[17]认为m=√自变量数量时分类效果达到最优。随机森林基于分类的结果,给出自变量重要性排序^[16]。重要性的衡量方法为:将一个变量的取值变为随机数,随机森林预测准确性降低的程度(mean decrease accuracy)^[16]。

本研究中存在藨草的样方占总样方比例大致为43%,而存在南荻的样方占总样方的比例仅为18%左右,是典型的非平衡数据集。随机森林在处理非平衡数据时,由于抽样的随机性使其对少数类样本往往不能很好地区分^[21]。为了提高分类精度,更精确地分析土壤因子与植被分布的关系,本研究利用SMOTE算法对原始样本进行重新采样,使新的数据重新达到平衡。重采样是对少数类样本进行分析并根据少数类样本人工合成新样本添加到数据集中,合成的策略是对每个少数类样本a,从它的最近邻中随机选一个样本b,然后基于a、b之间的距离选择合适的点作为新合成的少数类样本^[22]。然后基于新的数据集建立随机森林预测模型。

利用Excel 2016软件进行数据整理;利用SPSS 22软件方差分析(ANOVA)模块进行差异显著性检

验;对非平衡数据的预处理利用R语言中的caret软件包和DMwR软件包;基于土壤因子的植被分布预测利用R语言中的random Forest软件包。

2 结果与分析

2.1 薜草和南荻分布区土壤养分特征

藨草和南荻分布区土壤pH均值分别为5.53和4.32(表2),其中藨草分布区pH显著高于南荻分布区($P<0.05$)。藨草分布区土壤全氮、氨氮与硝态氮含量分别为3 080.17、139.46和14.83 mg/kg;南荻分布区土壤全氮、氨氮与硝态氮含量分别为3 155.35、121.61和22.48 mg/kg。全氮和氨氮在这两种植被分布区上的差别不显著,而南荻分布区土壤硝态氮略高于藨草分布区($P<0.1$)。藨草分布区土壤全磷和有效磷分别为534.01、4.60 mg/kg,南荻分布区为449.71和4.81 mg/kg。两种植被分布区全磷和有效磷含量没有显著差别。藨草分布区土壤中全钾含量为26.47 mg/kg,显著低于南荻分布区的35.41 mg/kg($P<0.05$)。藨草分布区和南荻分布区有机质含量没有显著差别,但藨草分布区土壤含水率显著高于南荻分布区($P<0.05$)。

表2 薜草分布区和南荻分布区土壤养分含量特征

Table 2 Soil nutrient contents under *Phalaris arundinacea* Linn and *Triarrhena lutarioriparia* L. Liu

土壤因子	藨草		南荻	
	平均值	标准差	平均值	标准差
pH	5.53	0.69	4.32	0.27
全氮(mg/kg)	3 080.17	1 079.40	3 155.35	631.86
氨氮(mg/kg)	139.46	42.84	121.61	31.34
硝态氮(mg/kg)	14.83	5.36	22.48	19.57
全磷(mg/kg)	534.01	266.01	449.71	163.70
有效磷(mg/kg)	4.60	1.60	4.81	5.91
全钾(mg/kg)	26.47	8.02	35.41	7.27
有机质(mg/kg)	8.84	2.73	9.52	2.02
含水率(%)	40.3	10.2	26.7	7.4

2.2 随机森林预测结果

本研究中自变量的数量为12个,设置随机森林模型树的数量为1 000,每棵树分类节点数m=3。基于12个自变量和60个样本建立的随机森林模型对藨草和南荻是否存在判别精度分别为89.58%和89.29%,为准确分析植被分布概率与土壤因子相互作用的关系做了准备。随机森林模型给出了每个土壤因子对植被出现概率预测的重要性程度(图2)。重要性体现为将一个变量的取值变为随机数后随机森林预测准确性降低的程度。重要性程度高的土壤因子能显著区分出该植被存在或不存在,体现了植被与土壤高

的相互作用程度和密切相关程度。与虉草分布密切相关的土壤因子重要性依次为全钾、氨氮、有机质、含水率、全氮、有效磷、全磷、pH 和硝态氮；与南荻分布密切相关的土壤因子的重要性依次为全钾、pH、有机质、全氮、全磷、硝态氮、氨氮、含水率和有效磷。

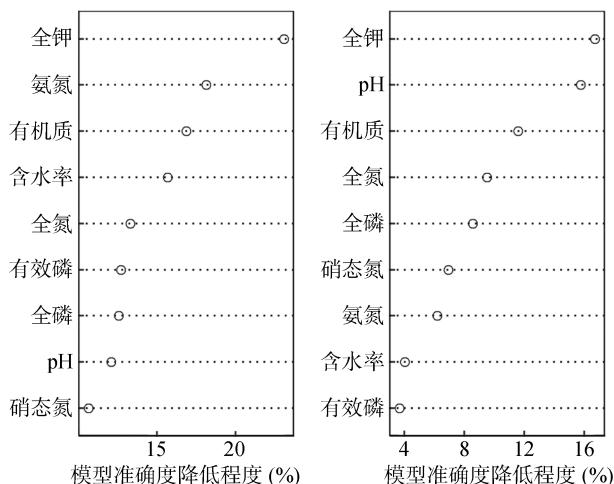


图 2 土壤因子与虉草(左)和南荻(右)分布预测中的重要性排序

Fig.2 Importance ranking of soil factors for distribution prediction of *Phalaris arundinacea* Linn and *Triarrhena lutarioriparia* L. Liu

2.3 植被分布与土壤养分的耦合作用

鄱阳湖属于洪泛型湿地，淹水历时、淹没深度和淹水频率等关键水情因子的时空变化对湿地植物的生长和繁殖有重要的影响^[33-35]。鄱阳湖低滩和高滩受到显著不同的水文情势的作用，植被的类型与分布有很大差异^[36]。不同高程植物和土壤通过长期的物质和能量的交换，加剧高低滩生境的空间异质性。因此，随机森林模型可以基于土壤环境因子能对低滩植被虉草和高滩植被南荻进行精准判别。基于随机森林分类的结果，模型自动计算了每个土壤因子作为预测变量对虉草和南荻判别能力的大小，以及植被与土壤因子的作用关系。图 3 中给出了植被分布与土壤因子的偏依赖图(partial dependent plot)及其解释。偏依赖图描述了某一预测变量在排除其他预测变量的情况下，对被预测变量出现概率在统计结果上的影响，在本研究中实质上体现了低滩植被和高滩植被对土壤养分的累积差异。偏依赖图纵坐标表示植被出现的可能性大小，横坐标表示调查的 60 个样方中土壤因子分布。

如图 3 所示，各土壤因子与虉草和南荻的相互作用不同。虉草分布区土壤 pH 呈弱酸性或中性；而南荻分布区土壤酸性较强。虉草在样方中出现的概率与土壤全氮含量呈负相关，虉草分布区土壤全氮含量低；南荻在样方中出现的概率与土壤全氮含量有较弱

的正相关，其分布区土壤全氮含量适中。虉草和南荻分布区氨氮含量都较低。虉草和南荻在样方中出现的概率与土壤中硝态氮含量呈正相关关系，分布区硝态氮含量都较高。虉草在样方中出现的概率与土壤全磷含量正相关，其分布区全磷含量较高；南荻出现的概率则与全磷含量负相关，其分布区全磷含量较低。虉草分布和南荻分布区与土壤有效磷关系不显著。虉草出现的概率与土壤全钾含量负相关，其分布区钾含量较低；南荻出现的概率则与土壤全钾含量正相关，其分布区钾的含量高。虉草分布区和南荻分布区有机质含量都比较低。虉草出现的概率与土壤含水率正相关，因此其分布区含水率高；南荻出现的概率则与土壤含水率负相关，因而其分布区含水率较低。

3 讨论

鄱阳湖水位周期性波动形成的水陆交替过渡带具有复杂多样的生境以及沿水位高程变化的湿地物质和能量结构^[25,31,37]。从湿地中心到湿地边缘，淹水时间变短，淹水频率降低，植被由低滩植被逐渐演变为高滩植被。不同植被类型通过对土壤养分和水分的吸收、根系分泌物的释放、残体降解等过程对土壤养分的组成产生重要作用。本文基于随机森林模型找到了低滩植被虉草和高滩植被南荻分布下土壤养分的累积差异。虉草分布区土壤偏中性或软酸性，这与虉草自身生理特性、虉草与土壤的相互作用有关。徐高峰等^[38]认为虉草在 pH 为 6.0 时萌发率最高。张永亮和骆秀梅^[39]的研究也表明，虉草耐 pH 范围在 4.9 ~ 8.2；南荻分布区土壤酸性较强，与 2010 年刘芳鹏等^[40]在鄱阳湖南矶山湿地的调查(4.83 ~ 5.72)相比，本研究显示南荻分布区土壤酸性程度更强(pH < 4.50)。虉草分布区土壤全氮和氨氮较低，硝态氮含量稍高。但虉草与硝态氮的关系不显著。相应的，硝态氮在预测变量重要性排序中最不重要。贾有余等^[41]则认为在控制实验中，施氮肥会提高虉草的产量，这可能与鄱阳湖虉草长期适应低滩区域低氮环境^[42]，从而改变了自身的生理生态特性有关；南荻分布区土壤全氮和硝态氮较高，与氨氮作用不显著。张全军等^[28]认为鄱阳湖代表性群落植株重量与土壤氮素含量存在显著的负相关关系，但未能揭示出每种植被分别与土壤氮素的相关关系。虉草分布区土壤全磷和有效磷含量高，而南荻则与土壤磷的作用不显著。郝建朝等^[43]的研究也表明不同类型的湿地群落中只有虉草与有效磷分布和全磷分布有显著的正相关，且虉草具有极强的脱磷能力。虉草和南荻与土壤中钾的作用

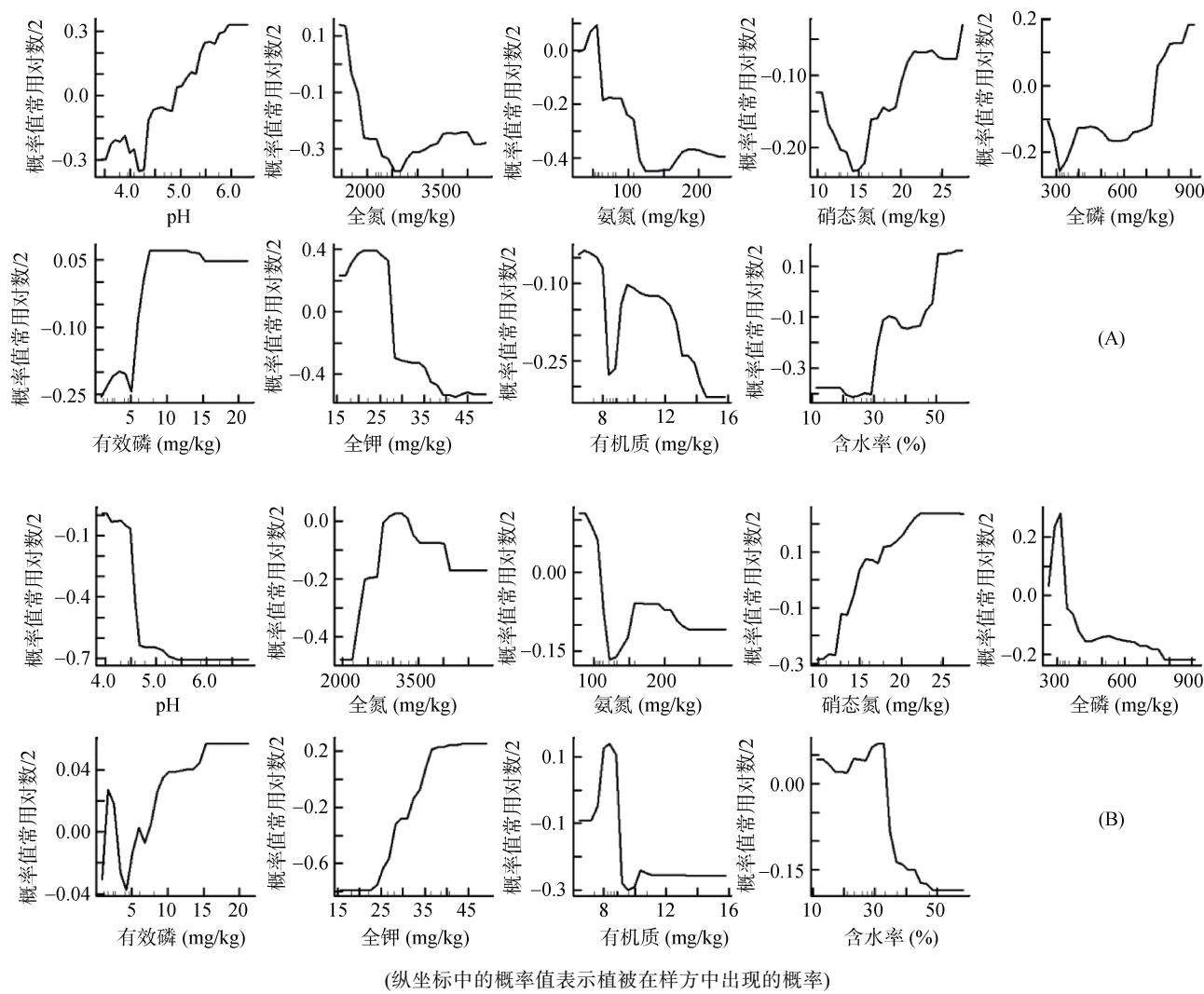


图 3 土壤因子与植被分布的偏依赖图(A: 藦草; B: 南荻)

Fig. 3 Partial dependent plots for soil factors and vegetation distribution (A: *Phalaris arundinacea* Linn; B: *Triarrhena lutarioriparia* L. Liu)

极为显著。藺草分布与钾的含量呈负相关关系, 藦草分布区土壤含量低; 南荻分布与钾的含量有显著的正相关关系, 南荻分布区土壤钾的含量高。董磊等^[1]、张全军等^[28]也认为鄱阳湖南荻分布区土壤钾的含量高于藺草分布区, 但没有揭示出具体的相关关系。藺草和南荻与有机质的作用显著, 两者分布区有机质含量低。张全军等^[28]的研究也表明有机质含量与鄱阳湖湿地群落植株重量有非常显著的负相关关系, 但没有阐明不同植被与有机质的相关关系。藺草分布区含水率高; 南荻分布区含水率低。这与这两种植被分布的区域密切相关, 藦草靠近湖心分布, 南荻分布的区域海拔稍高。董磊等^[1]认为含水率是影响鄱阳湖植被带状分异的最重要因素。

生态学问题通常涉及多维变量间的复杂关系, 传统的统计学方法例如逻辑斯蒂回归难以揭示复杂过程中透露的格局和关系^[44]。随机森林能很好地处理

非线性关系、高维数据以及缺失值, 并能对分类结果进行可视化, 给出合理的生态学解释, 在处理生态学问题中有极大的应用价值。Peters 等^[45]使用多重逻辑斯蒂回归模型和随机森林模型找到了桑松低洼湿地地区沿环境梯度下的植被分布的主要限制性要素, 研究表明随机森林的综合表现更好。Long 等^[46]基于随机森林模型找到了美国西部大盐湖芦苇分布的主要限制性环境因子。张雷等^[47]以云南松分布模拟为例, 突出了随机森林算法与传统的统计方法以及其他机器学习方法相比, 在解决生态学问题上的优势。本研究表明随机森林方法适用于植被分布的预测, 清晰地展示了植被分布与土壤因子的依赖关系, 进而结合研究区域的特点进行合理的生态学解释。

4 结论

1) 随机森林基于土壤环境因子对藺草和南荻在

样方中存在与否进行了准确的判别(预测精度分别为 89.6% 和 89.3%), 较好辨识了以𬟁草为代表的低滩植被和以南荻为代表的高滩植被对土壤养分的累积差异。

2) 𬟁草与土壤中的全钾、氨氮、有机质密切相关; 南荻与土壤全钾、pH、有机质密切相关(将自变量按重要性排序后, 选取前 3 个土壤因子)。𬟁草分布与钾、氨氮和有机质都呈负相关关系, 且分布区这几种土壤因子含量都比较低; 南荻分布与全钾正相关, 而与有机质和 pH 负相关。南荻分布区全钾含量较高, 土壤酸性强且有机质含量较低。

3) 随机森林算法在处理生态学高维数据中具有显著的优越性, 能克服传统统计学方法在处理复杂数据方面的不足, 对大量相互作用的数据进行挖掘, 并得到简单易于解释的可视化结果, 在解决生态学问题中具有显著的优越性。

参考文献:

- [1] 董磊, 徐力刚, 许加星, 等. 鄱阳湖典型洲滩湿地土壤环境因子对植被分布影响研究[J]. 土壤学报, 2014, 51(3): 618–626.
- [2] 徐金英, 陈海梅, 王晓龙. 水深对湿地植物生长和繁殖影响研究进展[J]. 湿地科学, 2016, 14(5): 725–732.
- [3] 李文, 王鑫, 潘艺雯, 等. 不同水淹深度对鄱阳湖洲滩湿地植物生长及营养繁殖的影响[J]. 生态学报, 2018, 38(9): 3016–3021.
- [4] Voesenek L A C J, Rijnders J H G M, Peeters A J M, et al. Plant hormones regulate fast shoot elongation under water: From genes to communities[J]. Ecology, 2004, 85(1): 16–27.
- [5] 曹煦彬, 林娣, 蔡璐, 等. 鄱阳湖南矶山湿地不同植被类型对土壤碳组分、羧化酶及 cbcl 基因的影响[J]. 土壤学报, 2017, 54(5): 1269–1279.
- [6] 罗琰, 苏德荣, 吕世海, 等. 辉河湿地河岸带土壤养分与酶活性特征及相关性研究[J]. 土壤, 2017, 49(1): 203–207.
- [7] 张天雨, 葛振鸣, 张利权, 等. 崇明东滩湿地植被类型和沉积特征对土壤碳、氮分布的影响[J]. 环境科学学报, 2015, 35(3): 836–843.
- [8] 张文敏, 吴明, 王蒙, 等. 杭州湾湿地不同植被类型下土壤有机碳及其组分分布特征[J]. 土壤学报, 2014, 51(6): 1351–1360.
- [9] 张骞, 姜俊彦, 李秀珍, 等. 潮滩湿地植被发育年限对土壤有机碳含量及积累速率的影响[J]. 生态学杂志, 2017, 36(5): 1173–1179.
- [10] 罗琰, 苏德荣, 纪宝明, 等. 辉河湿地不同草甸植被群落特征及其与土壤因子的关系[J]. 草业学报, 2018, 27(3): 33–43.
- [11] 李志, 李新国, 毛东雷, 等. 博斯腾湖西岸湖滨带不同植被类型土壤剖面盐分特征分析[J]. 西北农业学报, 2018, 27(2): 260–268.
- [12] 曾全超, 李鑫, 董扬红, 等. 黄土高原延河流域不同植被类型下土壤生态化学计量学特征[J]. 自然资源学报, 2016, 31(11): 1881–1891.
- [13] 孙倩, 阿丽亚·拜都热拉, 依力亚斯江·努尔麦麦提. 新疆艾比湖湿地 LUCC 对景观格局演变的影响[J]. 土壤, 2018, 50(3): 583–592.
- [14] 马文瑛, 赵传燕, 彭守璋, 等. 黑河天涝池五种植被类型土壤呼吸速率动态特征及其影响因子[J]. 生态学报, 2015, 35(17): 5654–5665.
- [15] 李欣海. 随机森林模型在分类与回归分析中的应用[J]. 应用昆虫学报, 2013, 50(4): 1190–1197.
- [16] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5–32.
- [17] Cutler A, Cutler D R, Stevens J R. Random forests[M]. New York: Springer US, 2012: 1–12.
- [18] Ho T K. The random subspace method for constructing decision forests[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(8): 832–844.
- [19] Prasad A M, Iverson L R, Liaw A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction[J]. Ecosystems, 2006, 9(2): 181–199.
- [20] Cutler D R, Jr Edwards T C E, Beard K H, et al. Random forests for classification in ecology[J]. Ecology, 2007, 88(11): 2783–2792.
- [21] Chen C, Breiman L. Using random forest to learn imbalanced data[J]. Univ. California, Berkeley, 2004: 1–12.
- [22] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: Synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16(1): 321–357.
- [23] 马骊. 随机森林算法的优化改进研究[D]. 广州: 暨南大学, 2016: 3–15.
- [24] 王晓鸿. 鄱阳湖湿地生态系统评估[M]. 北京: 科学出版社, 2004: 3–55.
- [25] 张全军, 于秀波, 胡斌华. 鄱阳湖南矶湿地植物群落分布特征研究[J]. 资源科学, 2013, 35(1): 42–49.
- [26] 冯文娟, 徐力刚, 王晓龙, 等. 鄱阳湖洲滩湿地地下水位对灰化薹草种群的影响[J]. 生态学报, 2016, 36(16): 5109–5115.
- [27] 楼倩. 鄱阳湖湿地植被群落 β -多样性格局[D]. 南昌: 南昌大学, 2016: 5–32.
- [28] 张全军, 于秀波, 钱建鑫, 等. 鄱阳湖南矶湿地优势植物群落及土壤有机质和营养元素分布特征[J]. 生态学报, 2012, 32(12): 3656–3669.
- [29] 葛刚, 赵安娜, 钟义勇, 等. 鄱阳湖洲滩优势植物种群的分布格局[J]. 湿地科学, 2011, 9(1): 19–25.
- [30] 谭志强, 张奇, 李云良, 等. 鄱阳湖湿地典型植物群落沿高程分布特征[J]. 湿地科学, 2016, 14(4): 506–515.
- [31] 胡振鹏, 葛刚, 刘成林, 等. 鄱阳湖湿地植物生态系统结构及湖水位对其影响研究[J]. 长江流域资源与环境, 2010, 19(6): 597–605.
- [32] 刘光崧. 土壤理化分析与剖面描述[M]. 北京: 中国标准出版社, 1996.
- [33] 谭志强, 许秀丽, 李云良, 等. 长江中游大型通江湖泊湿地景观格局演变特征[J]. 长江流域资源与环境, 2017, 26(10): 1619–1629.

[34] Chapin D M, Paige D K. Response of delta vegetation to water level changes in a regulated mountain lake, Washington State, USA[J]. *Wetlands*, 2013, 33(3): 431–444.

[35] Tan Z Q, Zhang Q, Li M F, et al. A study of the relationship between wetland vegetation communities and water regimes using a combined remote sensing and hydraulic modeling approach[J]. *Hydrology Research*, 2016, 47(1): 278–292.

[36] 叶春, 刘元波, 赵晓松, 等. 基于 MODIS 的鄱阳湖湿地植被变化及其对水位的响应研究[J]. 长江流域资源与环境, 2013, 22(6): 705–712.

[37] 葛刚, 徐燕花, 赵磊, 等. 鄱阳湖典型湿地土壤有机质及氮素空间分布特征[J]. 长江流域资源与环境, 2010, 19(6): 619–622.

[38] 徐高峰, 申时才, 张付斗, 等. 外来入侵植物小子薹草研究进展与展望[J]. 中国生态农业学报, 2015, 23(9): 1083–1092.

[39] 张永亮, 骆秀梅. 蒿草的研究进展[J]. 草地学报, 2008, 16(6): 659–666.

[40] 刘芳鹏, 许光勤, 刘倩纯, 等. 南矶山湿地典型植物群落土壤细菌群落结构变化[J]. 湿地科学, 2015, 13(4): 444–450.

[41] 贾有余, 骆秀梅, 张永亮. 氮肥与行距对不同生育期薹草产量与品质的调控效应[J]. 草业科学, 2016, 33(11): 2312–2318.

[42] 雷学明, 段洪浪, 刘文飞, 等. 鄱阳湖湿地碟形湖泊沿高程梯度土壤养分及化学计量研究[J]. 土壤, 2017, 49(1): 40–48.

[43] 郝建朝, 吴沿友, 连宾, 等. 不同植物下湿地土壤磷状况与植物脱磷效应[J]. 地球与环境, 2006, 34(1): 44–48.

[44] De'Ath G, Fabricius K E. Classification and regression trees: A powerful yet simple technique for ecological data analysis[J]. *Ecology*, 2000, 81(11): 3178–3192.

[45] Peters J, Verhoest N E C, Samson R, et al. Wetland vegetation distribution modelling for the identification of constraining environmental variables[J]. *Landscape Ecology*, 2008, 23(9): 1049–1065.

[46] Long A L, Kettenring K M, Hawkins C P, et al. Distribution and drivers of a widespread, invasive wetland grass, *Phragmites australis*, in wetlands of the Great Salt Lake, Utah, USA[J]. *Wetlands*, 2016, 37(1): 1–13.

[47] 张雷, 王琳琳, 张旭东, 等. 随机森林算法基本思想及其在生态学中的应用——以云南松分布模拟为例[J]. 生态学报, 2014, 34(3): 650–659.