

黄语芯, 赵治东, 刘峰. 基于逐类型随机森林的土种与成土环境关系研究[J]. 土壤学报, 2027, 64 (1)

HUANG Yuxin, ZHAO Zhidong, LIU Feng. Study on the Relationship Between Soil Species and Soil-Forming Environments Based on a Type-by-Type Random Forest Modeling Strategy[J]. Acta Pedologica Sinica, 2027, 64 (1)

基于逐类型随机森林的土种与成土环境关系研究*

黄语芯^{1,2}, 赵治东³, 刘峰^{1,2†}

(1.土壤与农业可持续发展全国重点实验室(中国科学院南京土壤研究所), 南京 211135; 2. 中国科学院大学, 北京 100049; 3. 中国农业大学土地科学与技术学院, 北京 100193)

摘要:土壤类型数字制图通常采用多类型一次性建模策略, 但是这种策略难以充分考虑每个土壤类型与其成土环境的复杂关系。对此, 本研究提出一种逐类型的随机森林建模策略, 并以江苏省仪征市为研究区, 利用 21 个环境协变量和从历史土壤图土种图斑提取 860 个典型虚拟样点, 逐个土种构建随机森林分类模型, 输出各土种预测概率的空间分布, 取概率值最大的土种类型作为该像元的预测结果。将该方法与常规多类型一次性随机森林建模策略进行了对比。结果表明, 逐类型多次建模策略的总体精度与 Kappa 系数分别为 0.62 和 0.58, 常规建模策略为 0.61 和 0.57, 二者总体精度在该研究区差距微弱, 但是前者不确定性明显减小, 逐类型多次建模策略的最大概率空间均值为 0.54, 高于常规建模策略的最大概率空间均值 0.08, 而且逐类型多次建模策略能够输出每个像元概率次大的土种可能性及混淆指数, 有助于识别北部岗地等概率次大土种可能性偏高的易混淆区域。逐类型多次建模策略对马肝土、黄刚土、黄杂土、油泥土等的预测精度高于常规建模策略, 但未在所有土种上取得一致提升。另外, 逐类型多次建模策略可输出单个土种的环境变量重要性, 有助于土种与环境因素关系分析, 结果显示高程、地形湿度指数和距长江距离可指示大部分土种空间变异。因此, 与常规一次性建模策略相比, 逐类型多次随机森林模型可显著降低土壤类型预测的不确定性, 提高预测结果可靠性, 可作为理解每个土种与成土环境之间关系的有用工具。

关键词: 数字土壤制图; 土壤类型; 逐类型多次建模策略; 低地形起伏区; 预测不确定性

中图分类号: S159

文献标志码: A

Study on the Relationship Between Soil Species and Soil-Forming Environments Based on a Type-by-Type Random Forest Modeling Strategy

HUANG Yuxin^{1,2}, ZHAO Zhidong³, LIU Feng^{1,2†}

(1. State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 211135, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. College of Land Science and Technology, China Agricultural University, Beijing 100193, China)

Abstract: 【Objective】 Digital soil mapping of soil types commonly adopts a multi-type one-time modeling

* 国家自然科学基金项目(42571077)资助 Supported by the National Natural Science Foundation of China (No.42571077)

† 通讯作者 Corresponding author, E-mail: fliu@issas.ac.cn

作者简介: 黄语芯(2000—), 女, 福建南平人, 硕士研究生, 主要从事数字土壤制图研究。E-mail: huangyuxin23@mails.ucas.ac.cn

收稿日期: 2026-03-13; 收到修改稿日期: 2026-10-21; 网络首发日期(www.cnki.net):

<http://pedologica.issas.ac.cn>

strategy, but this strategy struggles to fully consider the complex relationship between each soil type and its soil-forming environment. To address this gap, this study proposes a type-by-type random forest modeling strategy.

【Method】 Taking Yizheng City, Jiangsu Province, as the study area, 21 environmental covariates and 860 typical virtual sample points extracted from soil species polygons in historical soil maps were used to construct random forest classification models for each soil species individually. The spatial distribution of predicted probabilities for each soil species was output, and the soil species with the highest probability was taken as the predicted result for each pixel. This method was compared with the conventional multi-type one-time random forest modeling strategy. 【Result】 Results showed that the overall accuracy and Kappa coefficient of the type-by-type multiple modeling strategy were 0.62 and 0.58, respectively, while those of the conventional modeling strategy were 0.61 and 0.57. The difference in overall accuracy between the two strategies was very small in the study area. However, the uncertainty of the former decreased significantly. The spatial mean of the maximum probability under the type-by-type multiple modeling strategy was 0.54, which was 0.08 higher than the spatial mean of the maximum probability under the conventional modeling strategy. Moreover, the type-by-type multiple modeling strategy could output, for each pixel, the probability of the second-most-probable soil species and the confusion index, which helps identify confusion-prone areas such as the northern uplands, where the probability of the second-most-probable soil species is relatively high. The type-by-type multiple modeling strategy had higher prediction accuracy for Horse-liver Soil, Yellow Rigid Soil, Yellow Variegated Soil, Oily Silty Soil, etc., than the conventional modeling strategy, but it did not achieve consistent improvement for all soil species. In addition, the type-by-type multiple modeling strategy could output the importance of environmental variables for individual soil species, facilitating analysis of the relationships between soil species and environmental factors. The results showed that elevation, topographic wetness index, and distance to the Yangtze River could indicate the spatial variation of most soil species. 【Conclusion】 Therefore, compared with the conventional one-time modeling strategy, the type-by-type multiple random forest model can significantly reduce the uncertainty of soil type prediction, improve the reliability of prediction results, and serve as a useful tool for understanding the relationship between each soil species and its soil-forming environment.

Key words: Digital soil mapping; Soil types; Type-by-type multiple modeling strategy; Low-relief areas; Prediction uncertainty

数字土壤制图 (Digital Soil Mapping, DSM) 是以土壤-景观模型为理论基础, 利用与土壤发生相关的环境协变量, 构建土壤与环境之间的定量关系模型, 预测生成栅格格式的土壤类型图^[1]。相较于依赖专家经验、手工勾绘的传统土壤制图, DSM 极大地提高了制图效率, 降低了人力与时间成本, 能更好地表达空间异质性。在平原、岗地等低地形起伏区土壤类型制图难度相对较大。这类区域环境梯度小, 地形、植被等常用环境变量与土壤类型空间变异的协同程度较低, 难以有效捕捉土壤类型的差异^[2]。同时, 土壤类型与成土环境变量之间的关系通常比较复杂, 模型预测能力不足。

为应对上述挑战, 学者们主要从引入新环境变量和优化预测模型两个方向进行探讨。在环境变量方面, 利用遥感技术引入地表动态反馈方法^[3]、时间序列遥感影像纹理特征^[4]等新型协同变量, 以增强对土壤空间分异信息的捕捉能力。在预测模型方面, 机器学习, 特别是集成学习模型, 已成为提升预测精度的核心手段。集成模型通过组合多个基模型以综合其优势, 显著提升性能。其中, 随机森林作为 Bagging 算法的代表, 能够有效处理非线性关系, 可减小过拟合, 在土壤类型制图中取得了良好效果^[5-7]。Boosting 和 Stacking 等集成算法也展现出较大潜力^[8]。土种是土壤分类系统中基层分类的基本单元, 同一土种处于相同或相似的景观部位, 其剖面性态特征、理化性质也相同或近似^[9]。针对土种的土壤类型预测, 大多采用基于随机森林的策略, 通过构建土壤样点与环境因素之间的关系, 一次性对所有土种进

行预测,得到土种分布图^[6-7]。李坤使用随机森林模型预测土种分布,总体精度为 0.5^[10]。当土壤类型过多时,其与成土环境变量之间的关系更为复杂多样,一次性建模难以充分考虑所有类型与成土环境变量之间的关系,预测效果下降,不确定性增加^[5]。

基于上述问题,本研究提出一种逐类型多次随机森林建模策略,以江苏仪征市为例,基于常规多边形土壤图提取虚拟样点,结合环境数据采用不同建模策略构建土壤景观模型,对仪征市土种进行预测制图。对比逐类型多次随机森林建模策略与常规多类型一次性随机森林建模策略的制图准确性及不确定性,分析制图策略适用性,提升低地形起伏区县级土壤类型制图准确性。

1 材料与方法

1.1 研究区概况

本文以江苏省扬州市仪征市(32°14′—32°36′N, 119°02′—119°22′E)县域范围(图 1)为研究区。仪征市南临长江,总面积 857 km²,属北亚热带季风气候,年均温 15.7℃,年降水量 1 058.9 mm,气候温润,夏季炎热。地势西北高、东南低,地貌分低丘陵、岗地、平原三大类型,丘陵地区岗、塍、冲相间,地貌复杂^[11]。地面高程相差较小,为低地形起伏区:平原 3.5~5 m,岗地 40~45 m,仅少数丘陵 100 m 以上。成土母质在低丘陵和岗地区域为基岩风化残积坡积物质、下蜀黄土及黄土淤积物,在平原圩区为长江冲积物。仪征市共有 4 个土类、15 个土种,其中水稻土面积最大,广泛分布于全境;潮土主要位于长江冲积平原;黄褐土和粗骨土位于低丘陵高岗地。

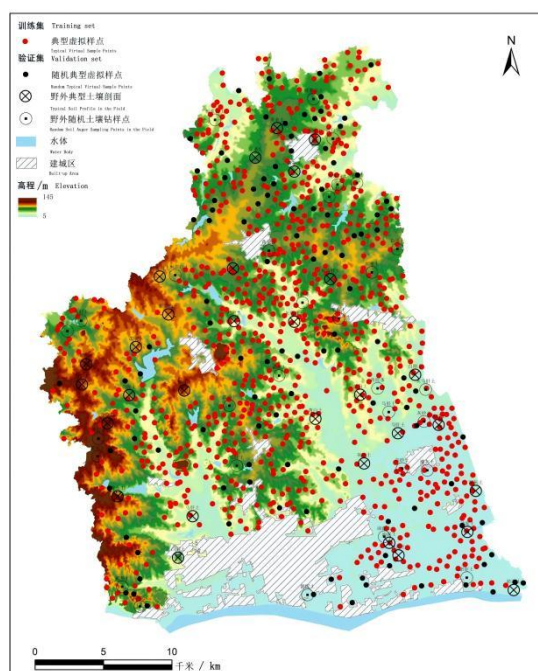


图 1 研究区地形高程及训练样点和验证样点分布(以彩图发表)

Fig. 1 Study area and training and validation samples

1.2 数据来源

仪征市土壤共划分 4 个土类、6 个亚类、9 个土属及 15 个土种。本研究以第二次全国土壤普查仪征市土壤图(图 2)为基础提取虚拟样点^[12],按土壤图斑面积分级(表 1)划定虚拟样点数量,同时保证两套模型策略训练集内的各土种虚拟样点数量一致。验证集共 172 个

用于独立验证（图1）：（1）30个野外典型土壤剖面，来自近年仪征市土壤调查；（2）30个随机采集的野外土钻样点；（3）112个典型虚拟样点，这些虚拟样点以500 m为最小间距随机抽取于典型景观区域，不参与建模。

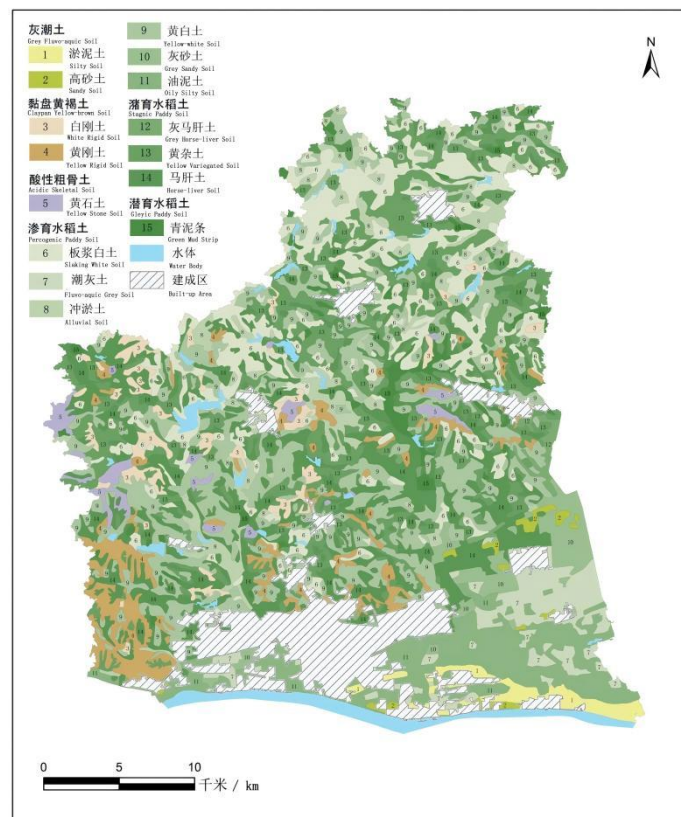


图2 研究区传统土壤图（以彩图发表）

Fig. 2 Historical soil map of the study area

表 1 不同土种类型的虚拟样点数量

Table 1 Number of virtual sampling points for each soil series

土类 Soil Great Group	亚类 Soil Subgroup	土属 Soil family	土种 Soil series	虚拟样点数量 Number of virtual sample points
潮土 Fluvo-aquic Soil	灰潮土 Grey Fluvo-aquic Soil	石灰性灰潮砂土 Calcareous Grey Fluvo-aquic Sandy Soil	高砂土 Sandy Soil	10
		石灰性灰潮黏土 Calcareous Grey Fluvo-aquic Clayey Soil	淤泥土 Silty Soil	10
粗骨土 Skeletal Soil	酸性粗骨土 Acidic Skeletal Soil	硅质酸性粗骨土 Siliceous Acidic Skeletal Soil	黄石土 Yellow Stone Soil	10
黄褐土 Yellow-brown Soil	黏盘黄褐土 Claypan Yellow- brown Soil	黄土质黏盘黄褐土 Loessial Claypan Yellow- brown Soil	白刚土 White Rigid Soil	40
			黄刚土 Yellow Rigid Soil	40
水稻土 Paddy Soil	潜育水稻土 Gleyic Paddy Soil	青马肝泥田 Green Horse-liver Mud Paddy Soil	青泥条 Green Mud Strip	10
	渗育水稻土 Percogenic Paddy Soil	渗潮泥田 Percogenic Fluvo-aquic Paddy Soil	潮灰土 Fluvo- aquic Grey Soil	40
			冲淤土 Alluvial Soil	40
			灰砂土 Grey Sandy Soil	40
			油泥土 Oily Silty Soil	40
	潴育水稻土 Stagnic Paddy Soil	渗马肝泥田 Percogenic Horse-liver Paddy Soil	板浆白土 Slaking White Soil	90
			黄白土 Yellow- white Soil	160
		潮泥田 Fluvo-aquic Paddy Soil	黄杂土 Yellow Variegated Soil	160
			灰马肝土 Grey Horse-liver Soil	10
		马肝泥田 Horse-liver Mud Paddy Soil	马肝土 Horse-liver Soil	160
	总计 Total			

根据土壤景观关系理论,地形、母质、植被、土地利用、地表环境、关键土壤属性是仪征市土种空间变异的主要环境因素。本研究使用了 21 个环境协变量表征仪征市土壤类型的发生环境,分别描述如下。

地形因素使用高程、地形湿度指数、坡度、剖面曲率及坡位进行表征,基于 12.5 mALOS DEM 数据通过 SAGA GIS 7.9 软件空间分析计算得到。母质类型来源于仪征市二普土壤调查历史资料,主要母质类型有砂砾岩、玄武岩等残坡积物、第四纪下蜀黄土、近代长江冲积物、黄土冲积淤积物等。植被因素使用归一化植被指数(NDVI)、增强植被指数(Enhanced Vegetation Index, EVI)标准差及增强植被指数增长速率,均基于 Sentinel-2A 影像,经去云、筛选等预处理后,通过 Google Earth Engine(GEE)平台提取 2021—2023 年生长季峰值 NDVI

值, 基于 12 个月增强植被指数时间序列计算增强植被指数标准差和增长速率。土地利用数据, 来源于市自然资源局提供的第三次全国国土调查数据, 按水田、水浇地、旱地、园地、林地、草地及其他用地分为 7 类。地表环境因素使用地表温度标准差、短波红外波段、距长江距离、距内部水系距离、距山体距离、斜坐标, 其中地表温度标准差和短波红外波段是使用 GEE 平台对 2018—2022 年近 5 年 Landsat 8 卫星时间序列影像进行月均值合成后得到。距长江距离通过 ArcGIS 10.7 基于长江中心线生成欧氏距离栅格。距内部水系距离基于哨兵 2 号遥感影像提取归一化差分水体指数 (NDWI) 生成水体图层, 提取仪征市内部宽度大于 10 m 的动态水体信息, 利用 ArcGIS 10.7 空间分析计算得到欧氏距离栅格。距山体距离根据《江苏省仪征县土壤志》^[13]所描绘的山丘几何中心点, 利用 ArcGIS 10.7 生成欧氏距离栅格。距内部水系距离、距长江距离及距山体距离, 单位均为千米。斜坐标是仪征市地形栅格图层各像元投影坐标系下横纵坐标平方和的平方根。土壤属性变量包括表层土壤 pH、有机质、阳离子交换量、砂粒、黏粒含量。数据来源于仪征市近年土壤调查数据空间插值生成。

由此, 本文使用的环境协变量有: 高程、坡度、剖面曲率、坡位、地形湿度指数; 母质类型; 归一化植被指数、增强植被指数标准差、增强植被指数增长速率; 土地利用; 地表温度标准差、短波红外波段; 距长江距离、距内部水系距离、距山体距离、斜坐标; 土壤有机质、pH、阳离子交换量、砂粒、黏粒含量。为确保数据的精确性和一致性, 对所有环境协变量进行了双线性插值重采样, 统一至 30 m 分辨率栅格, 空间投影坐标系为高斯-克吕格投影 6 度分带。

1.3 研究方法

本研究技术路线是 (图 3): 从传统矢量土壤图斑拾取典型虚拟样点, 从虚拟样点中随机抽取 112 个, 与野外剖面样点和野外土钻样点共同作为验证样点集, 其他虚拟样点作为训练样点集; 然后, 基于训练样点与环境协变量, 分别采用常规多类型一次性随机森林建模策略和逐类型多次随机森林建模策略进行仪征市土种类型空间预测。逐类型多次随机森林建模策略将传统一次性多分类问题拆解为多个二分类问题, 推测每个土种类型在各像元位置出现的概率, 选择概率值最高的土种类型作为该像元预测结果, 输出不确定性。最后, 基于模型导出的环境协变量重要性排序, 对各土种与成土环境因素之间的关系进行分析。

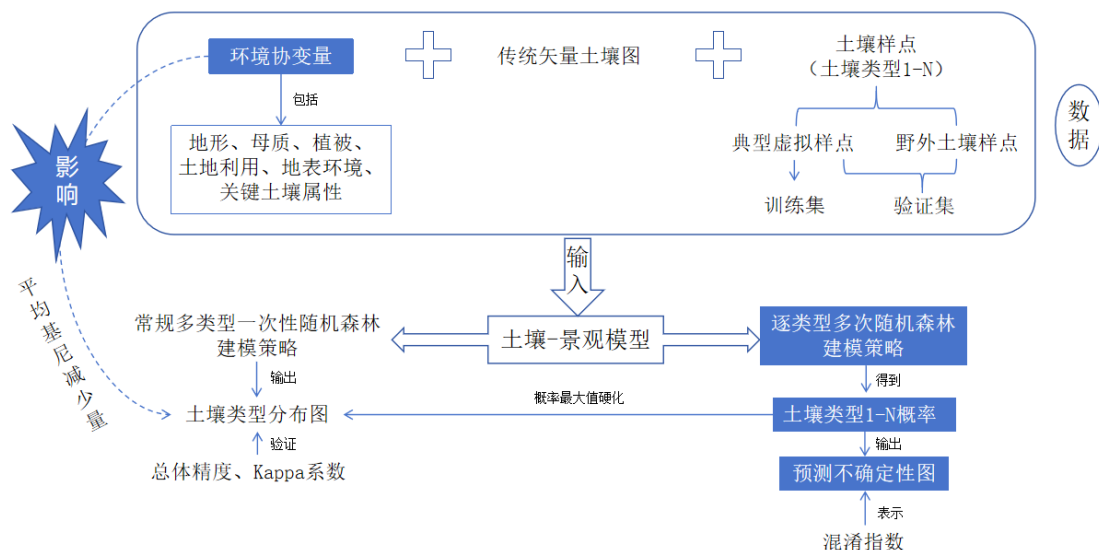


图 3 技术路线图 (以彩图发表)

Fig. 3 Technology roadmap

随机森林模型在分类问题中通常表现较好^[7]。它建立多个决策树, 统计每棵树的预测结果, 通过投票法确定最终结果。其优点在于数值处理时具有抗噪声、抗过拟合等特性, 且能

细致刻画土壤类型分布。常规多类型一次性随机森林建模策略是最为广泛使用的策略，它一次性建立所有土壤类型与环境协变量之间的一个集总关系模型，生成制图区域土壤类型空间分布图。

逐类型多次随机森林建模策略是对每个土壤类型都建立一个随机森林模型，推测该土壤类型在整个制图区域的概率空间分布，综合每个土壤类型的概率分布图层，生成制图区域土壤类型分布图。具体做法是将多土壤类型的多元分类问题简化为二元分类问题：针对某目标土壤类型进行预测时，将其余土壤类型设置为 0，目标土壤类型设置为 1，通过预测模型输出该土壤类型的空间分布概率图层。以此类推，输出各土壤类型的分布概率图层。在每个栅格像元上，将各土壤类型分布概率进行对比，选择概率值最大者作为该像元预测土壤类型，即硬化处理。

在 R 4.1.1 中使用 randomForest 包构建随机森林模型。随机森林模型两个关键参数需要设定，分别是每个节点中随机选择特征变量的数量（mtry）和决策树的数量（ntree）^[14]。终端节点最小样本数（nodesize）用于控制树的生长深度。根据 Breiman^[14]、Liaw 和 Wiener^[15] 的研究，在回归问题中 mtry 的默认推荐设置为环境协变量总数的三分之一。ntree 的取值通常设为 500，已有研究表明当树模型数量大于等于 500 时模型袋外误差趋于稳定，继续增加树的数量对精度提升有限。nodesize 在分类问题中的默认值为 1。在本研究中，将 mtry 数值设置为环境协变量数量的三分之一，ntree 数值设置为 500，nodesize 采用默认值 1。两个模型策略在模型参数设置上保持一致。

基于验证样点对土壤类型制图精度进行评价。计算不同制图策略土壤类型预测的混淆矩阵，采用总体精度、Kappa 系数、用户精度和生产者精度指标，采用混淆指数（Confusion Index, CI）量化预测结果的不确定性^[16]。

2 结果

2.1 不同策略制图结果与精度评价

不同制图策略输出的土壤图的独立验证效果（表 2）仅采用总体精度和 Kappa 系数评价，制图效果差异不明显，但两者最大概率空间均值差异达 0.46。

表 2 不同制图策略的准确性量化评估

Table 2 Quantitative evaluation of accuracy for different mapping strategies			
制图策略 Mapping Strategy	总体精度 OA	KAPPA	最大概率空间均值 Maximum Probability Spatial Mean
常规多类型一次性随机森林建模策略 Conventional multi-class one-time random forest modeling strategy	0.61	0.57	0.08
逐类型多次随机森林建模策略 Type-by-type multiple random forest modeling strategy	0.62	0.58	0.54

不同土种在两种策略中的建模精度（表 3）显示，两种模型在不同土种上的生产者精度和用户精度差异较大，均呈现对部分土种识别能力较强、对另一些土种识别能力较弱的现象。逐类型多次建模策略相较于常规建模策略并未在所有土种上取得一致提升，部分土种预测性能改善，部分则下降，未表现出稳定的优越性，但对面积最大的马肝土，逐类型多次建模策略分类效果最优。板浆白土、潮灰土、灰砂土和马肝土的生产者精度在逐类型多次建模策略中高于常规建模策略；白刚土、黄刚土、黄石土、黄杂土、青泥条、油泥土和淤泥土的生产

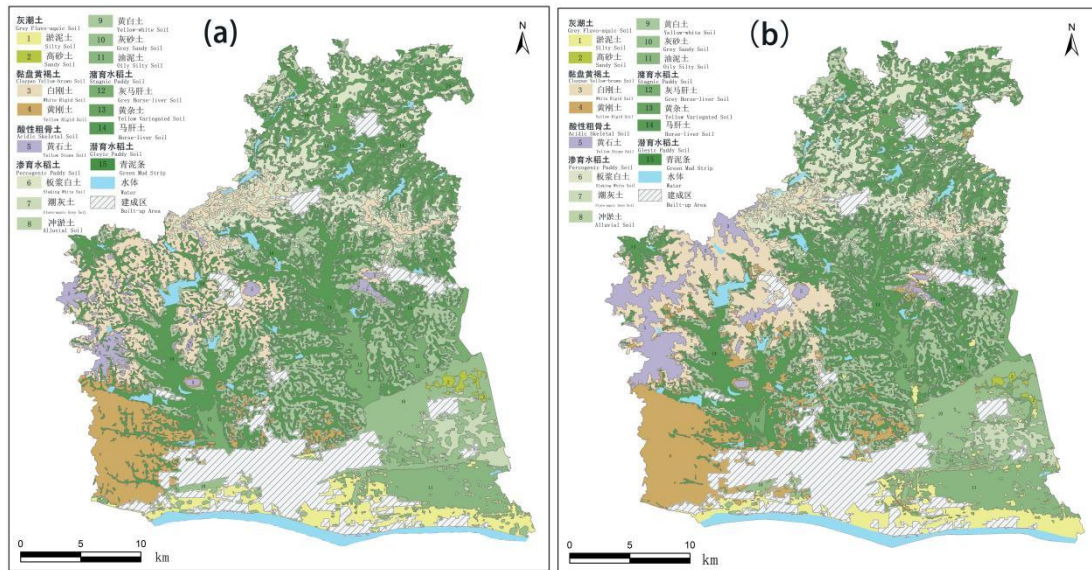
者精度则低于常规建模策略；冲淤土、高砂土、黄白土和灰马肝土在两种模型中的生产者精度相等。马肝土、黄刚土、黄杂土及油泥土的用户精度在逐类型多次建模策略中高于常规建模策略；黄白土、白刚土、板浆白土及青泥条的用户精度则低于常规建模策略；淤泥土、灰砂土、黄石土、灰马肝土、潮灰土、冲淤土及高砂土在两种模型中的用户精度相等。

表 3 各土种类型在不同策略中的预测建模精度

Table 3 Prediction modeling accuracy of different soil series under various strategies

土种 Soil series	常规多类型一次性随机森林建模策略 Conventional multi-class one-time random forest modeling strategy		逐类型多次随机森林建模策略 Type-by-type multiple random forest modeling strategy	
	生产者精度		生产者精度	
	用户精度		用户精度	
	PA	UA	PA	UA
白刚土 White Rigid Soil	0.22	0.29	0.20	0.14
板浆白土 Slaking White Soil	0.89	0.44	0.94	0.41
潮灰土 Fluvo-aquic Grey Soil	0.50	0.33	1.00	0.33
冲淤土 Alluvial Soil	1.00	0.33	1.00	0.33
高砂土 Sandy Soil	1.00	0.50	1.00	0.50
黄白土 Yellow-white Soil	0.50	0.61	0.50	0.56
黄刚土 Yellow Rigid Soil	0.80	0.75	0.76	0.81
黄石土 Yellow Stone Soil	1.00	0.86	0.75	0.86
黄杂土 Yellow Variegated Soil	0.45	0.75	0.44	0.79
灰马肝土 Grey Horse-liver Soil	0.25	0.33	0.25	0.33
灰砂土 Grey Sandy Soil	0.44	1.00	0.50	1.00
马肝土 Horse-liver Soil	0.53	0.59	0.55	0.65
青泥条 Green Mud Strip	0.50	0.25	0.00	0.00
油泥土 Oily Silty Soil	0.91	0.91	0.85	1.00
淤泥土 Silty Soil	1.00	0.86	0.92	0.86

两种制图策略的土壤类型预测结果（图 4）显示，逐类型多次建模策略（图 4（b））在部分土种的预测上具更大优势。在灰马肝土和马肝土区分上，冲田区域灰马肝土和马肝土的主要差异在于 60 cm 土层以下是否存在腐泥层，该差异在表层土壤属性数据中很难区分，加之二者成土环境差异较小，模型预测时常将二者混淆，逐类型多次建模策略在灰马肝土预测精度不变的情况下，细化了灰马肝土分布又提升了马肝土预测精度。在潮灰土和灰砂土的区分上，两者均位于平原与岗地的过渡区域，成土母质同为长江冲积物，主要区别在于土壤质地，逐类型多次建模策略相比常规建模更有效地利用了土壤属性变量，更好地识别了潮灰土和灰砂土的真实分布，提升了两者的生产者精度。



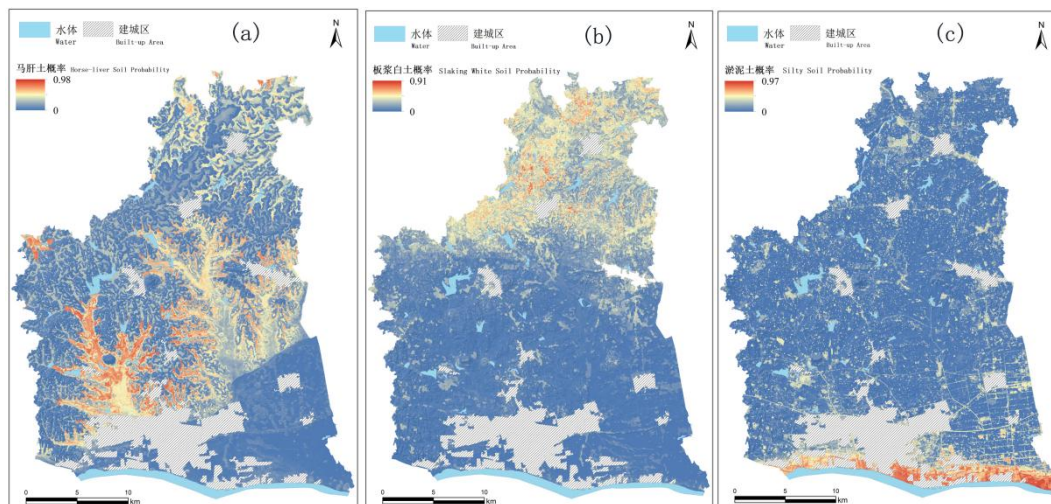
注: (a)常规多类型一次性随机森林建模策略更新后的土壤图, (b)逐类型多次随机森林建模策略更新后的土壤图。Note:(a) Soil map updated using the conventional one-time multi-class random forest modeling strategy, (b) Soil map updated using the type-by-type multi-run random forest modeling strategy.

图4 土壤类型图对比 (以彩图发表)

Fig. 4 Comparison of soil type maps

2.2 基于逐类型多次随机森林建模策略的土种结果

由逐类型多次建模策略更新后的土壤图(图4b)可见,水稻土是仪征市面积最大的土类,其亚类自东南向西北由渗育经潜育过渡到潴育。潜育水稻土青泥条主要分布于东部河谷地带的低洼地区和低丘陵地带,也零星分布在其他地区。潴育水稻土主要分布于仪征市低岗冲田区域,包括灰马肝土、马肝土、黄杂土,灰马肝土土种分布于东部低平岗地的冲田部位,马肝土(图5a)分布于丘陵岗地底部的冲田部位,黄杂土通常分布在岗地中下部位。渗育型水稻土包括板浆白土、黄白土、冲淤土、潮灰土、灰砂土、油泥土。板浆白土(图5b)主要分布在东北部高平岗地区。黄白土主要分布在低丘陵山区。冲淤土在中北部平缓岗地区域及南部河谷平原呈零星分布。潮灰土主要分布在东南部平原与岗地过渡地带。灰砂土主要分布于平原与岗地过渡带、长江冲积平原圩田区域,呈东西带状分布。油泥土主要分布在长江冲积平原沿岸,呈东西带状分布。旱作土壤包括黄褐土、潮土、粗骨土土类。黄褐土主要分布于仪征西部低丘陵高岗地与中部低山坡地,包括黄刚土、白刚土。黄刚土和白刚土均集中分布在西部高岗地,前者主要位于西部低丘陵地带上坡和中坡,后者主要位于中坡和零星下坡水改旱区域。潮土主要分布于长江冲积平原,包括高砂土、淤泥土。高砂土主要位于平原-岗地过渡带及长江冲积平原圩区上部,呈西南-东北向条带状展布,沿江与内陆河道两岸亦有零星斑块。淤泥土(图5c)多分布于滨江圩田,与长江平行呈带状分布。粗骨土为仪征市面积最小的土类,黄石土主要分布在仪征的山丘顶部,其次是西部低丘陵与高岗地的过渡地带及中部突起的低山坡地。



注：(a) 马肝土、(b) 板浆白土、(c) 淤泥土。Note: (a) Horse-liver Soil, (b) Slaking White Soil, and (c) Silty Soil.

图 5 马肝土、板浆白土、淤泥土的空间分布概率图（以彩图发表）

Fig. 5 Probability maps of spatial distribution for Horse-liver Soil, Slaking White Soil and Silty Soil

2.3 环境变量重要性排序

仪征市土种类型环境变量重要性排序（图 6）显示，高程、地形湿度指数、距长江距离等变量在黄杂土、灰马肝土、马肝土等多个土壤类型中呈现深红色，说明这些变量对这类土壤的形成、特征识别或演化具有关键影响。岩性类型、土地利用类型等变量在大多数土种中重要性均较低，说明其并非主要成土因素。

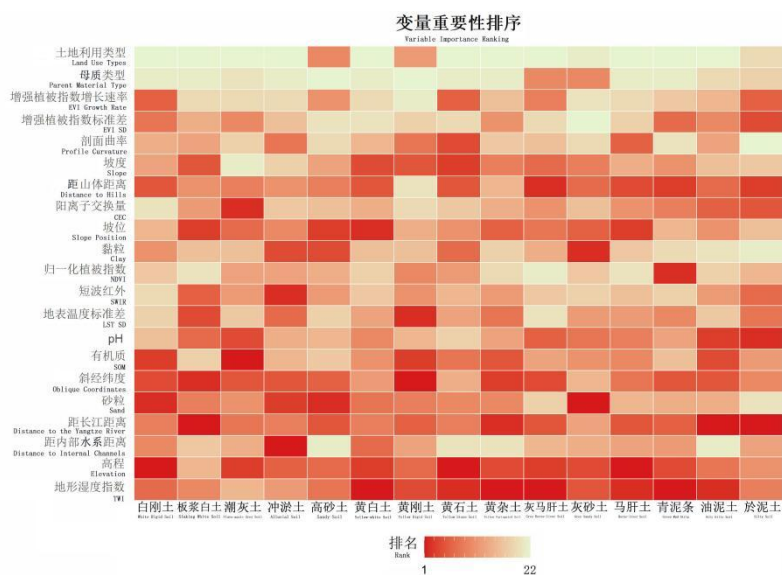


图 6 不同环境变量对预测土种的重要性（以彩图发表）

Fig. 6 Variable importance for soil type prediction

3 讨论

3.1 不同建模策略预测性能比较

随机森林模型虽可获得每个样本在每棵决策树上的预测结果及其对应的投票结果,但由于随机森林由多个决策树组成,每棵决策树基于随机特征的子集进行训练,决策路径各异。这种随机性使得难以追踪模型的具体决策路径,虽可得到每个样本的预测结果和对应的投票概率,但难以清楚了解模型的具体决策路径^[17-18]。

常规多类型一次性随机森林建模策略针对分类问题采用投票机制,即每个决策树对样本进行分类,最终输出得票数最多的类别作为整个森林的预测结果,可解释性差,难以清晰解释每个样本的决策分类过程。逐类型多次随机森林建模策略可获取各类型在整个地理空间出现的概率值,通过多个预测概率图层的栅格堆栈,得知每个像元上各土壤类型可能出现的概率,较好地评估分类结果的不确定性。逐类型多次建模策略的预测结果在南部展现出较好的预测性能,表明该策略对识别平原岗地过渡区域具有较好的预测性能。然而,北部区域概率次大的土种可能性上升,概率最大的土种可能性下降,表明模型在这些低地形起伏的岗地地貌中易产生混淆。尽管常规建模策略和逐类型多次建模策略预测的土壤类型图在县域尺度上呈现出相似分布格局,但是后者预测结果的不确定性相对较低。这可能说明逐类型多次建模策略能够更为充分地考虑土壤类型与成土环境条件之间的复杂关系,提高了土壤预测的可靠性。

3.2 土种空间分异与成土环境的关系分析

环境变量对预测土种空间分布具有重要作用。常规建模策略输出的是总体的环境变量重要性,而逐类型多次建模策略可输出单个土种的环境变量重要性,有助于分析单个土种与其成土环境因素之间的关系。本文环境协变量的贡献度具有非平稳性,同一变量对不同土种的贡献度差异较大。这意味着一个环境变量的重要性并非在整个研究区或对所有土壤类型均恒定不变,而是随着地理位置和土种类型的变化而变化。高程、地形湿度指数及距长江距离是最重要的环境协变量。其中,高程直接影响土壤的垂直分布;地形湿度指数能够反映地形对水分分布的控制作用,而土种形成与水分条件密切相关;距长江距离同样能够间接指示土壤水文条件和土壤质地差异,尤其是距离长江最远和最近的土种预测效果较明显。通过地形、水文及土壤质地等环境变量可有效预测仪征市大部分土种。

潜育水稻土的青泥条土种(UA=0),地形低洼,排水不畅,作物长势差,地形湿度指数值较大,归一化植被指数较低,这两个变量最为重要。该土种分布一般较为分散,斑块较小,预测土种精度极低。潜育水稻土包括灰马肝土、马肝土、黄杂土。其中,灰马肝土(UA=0.33)的重要环境变量是地形湿度指数值和距主要山体距离。在高地形湿度指数值区域汇水条件优越、地下水位埋深浅,利于灰马肝土发育。距主要山体较近区域地形坡降大,可指示下冲部位的土壤类型。灰马肝土实际分布面积较小,虚拟样点训练样本数量较少,一定程度上造成了该土种预测精度偏低。对马肝土(UA=0.65)而言,最重要的是高程,这与马肝土的地理分布位置相关。其次为地形湿度指数,高值区汇水充沛,促进马肝土潜育层发育。随后是坡位,马肝土多发育于下坡位置。距主要山体距离、距长江距离同样体现地形水文的控制作用。马肝土与灰马肝土在剖面特征上的区别在于60 cm土层以下腐泥层的有无,该特征是土壤历史形成过程的产物,难以被现有环境协变量指示,加之二者成土环境相近,易出现混淆。黄杂土(UA=0.79)分布坡位略高于马肝土,黄杂土潜育层发育棕黄色胶膜及丰富的铁锰新生体,地形湿度指数通过调控水分分配,主导了铁富集与黏粒胶膜发育等。

渗育型水稻土随母质分为多个类型,下蜀黄土母质发育的板浆白土、黄白土,黄土淤积物发育的冲淤土,长江冲积物发育的潮灰土、灰砂土与油泥土。板浆白土(UA=0.41)由下蜀黄土风化壳受侵蚀、下部网纹层出露地表后经成土作用形成。距长江距离指示的是侧渗漂洗地形环境,斜坐标、坡位、坡度可表征倾斜地形下板浆白土的空间变异,也可一定程度上指示下蜀黄土母质分布范围。板浆白土土壤呈砂粒质微酸性,pH与土壤质地分效果较好。黄白土(UA=0.56)的白土层源于黏粒与铁锰氧化物随水分向下淋溶淀积。地形湿度指数、

坡位及坡度等局部地形因子共同调控淋溶强度,高地形湿度指数区域汇水条件好,坡顶排水通畅,淋溶作用强烈,白土层易于发育。黄白土多发育于高程 20~30 m 垆田,因而高程、距主要山体距离等地形变量重要性次之。黄白土与板浆白土地形部位相近,但预测概率高值区差异显著,根源在于表层土壤属性差异。黄白土有机质含量高于板浆白土,土壤属性变量可有效区分二者。冲淤土(UA=0.33)最重要的环境协变量是反映地表湿度的短波红外波段,临近内部水系区域沉积物堆积显著,冲淤特征明显;同时其砂质壤土质地也是其重要特征,土壤质地变量区分效果较好。潮灰土(UA=0.33)分布于平原与岗地过渡带,土壤有机质和地形高程有较好的区分效果。灰砂土(UA=1)由石灰性灰潮砂土土属的高砂土土种演化而来,与潮灰土处于同一部位,二者在土壤质地上有差异,灰砂土通体砂质,土壤质地变量重要性最高。油泥土(UA=1)土体上黏下砂,其分布沿着长江呈条带状分布,因而土壤质地、距长江距离和地形湿度指数具有较好区分效果。

黄褐土包括黄刚土与白刚土,发育于下蜀黄土母质,土体黏粒含量较高,黄刚土有机质积累水平高于白刚土。对于黄刚土(UA=0.81),斜坐标、地表温度标准差和土壤有机质变量的重要性较高,斜坐标表征了黄刚土分布区集中在西南部,黏粒和有机质含量相对较高,地表温度标准差较小。白刚土(UA=0.14)分布较为零散,地形部位比黄刚土低,多见于坡下水田改造形成的园地与旱地。

潮土中的高砂土与淤泥土均发育于长江冲积物母质。高砂土(UA=0.5)属砂质土,质地可有效表征其空间分布,坡位可指示砂质物质易于堆积的强沉积动力位置,距长江距离是控制颗粒分选的重要因子。淤泥土(UA=0.86)发育与长江密切相关,距长江距离控制沉积动力与物质来源,距主要山体距离约束沉积边界,二者共同反映宏观沉积环境。淤泥土呈微碱性,保肥性能良好,使 pH 与 CEC 等土壤理化属性具备较高的变量重要性。增强植被指数、短波红外波段、地表温度标准差表征了水热和植被状况,也对淤泥土分布具备较好的指示能力。

粗骨土(UA=0.86)为仪征市面积最小的土类,其下的黄石土发育于基岩风化残积坡积物。高程、地形湿度指数、坡度、剖面曲率、距主要山体距离主导了黄石土的空间分布,均具有较高变量重要性。黄石土所处地形部位侵蚀较强,在地形湿度指数低值、剖面曲率为正的凸形坡位置,径流冲刷与侵蚀作用强烈,土壤粗骨化、砾石占比升高,该土种的出现概率更高。

4 结 论

(1) 逐类型多次建模策略能够更为充分地考虑每个土壤类型与其成土环境之间的复杂关系,在土种预测制图中具有显著优势。第一,尽管在本研究区其总体精度与 Kappa 系数与常规建模策略相当,但其最大概率空间均值 0.54 明显高于常规建模策略的 0.08,预测不确定性显著降低,预测可靠性得到提升。第二,常规建模策略只能输出总体环境变量重要性,而逐类型多次建模策略可输出每个单个土种的环境变量重要性排序,有助于理解和认识单个土种与其成土环境的关系。(2) 逐类型多次建模策略对土种类型易混淆的区域及易混淆的土种的区分效果更好。对于马肝土与灰马肝土等成土环境相似、表层属性难以区分的土种,通过逐类型建模提升了土种预测精度,马肝土精度从 0.59 提升至 0.65,并能够识别北部岗地等易混淆区域的概率次大的土种。在后续的研究中,可将概率次大的土种出现可能性较高的区域作为补充采样的重点区域,提升对易混淆土种的推测效果。(3) 高程、地形湿度指数和距长江距离对仪征市大部分土种空间预测具有重要的指示作用。对于环境变量指示能力弱的土种,仍需进一步研究构建新的环境协变量。

参考文献 (References)

- [1] Zhang G L, Shi Z, Wang Q B, et al. Development of soil geography in the new era and its future[J]. *Acta Pedologica Sinica*, 2023, 60(5): 1264-1276. [张甘霖, 史舟, 王秋兵, 等. 新时代土壤地理学的发展现状与趋势[J]. *土壤学报*, 2023, 60(5): 1264-1276.]
- [2] Wang J Y, Liu F, Song X D, et al. Mapping soil properties using the land surface temperature in an arid plain[J]. *Chinese Journal of Soil Science*, 2018, 49(6): 1270-1278. [王俊雅, 刘峰, 宋效东, 等. 基于地表温度的干旱平缓区土壤属性制图[J]. *土壤通报*, 2018, 49(6): 1270-1278.]
- [3] Liu F. Study on soil spatial inference in gentle terrain areas based on Land Surface Dynamic Feedback[D]. Beijing: University of Chinese Academy of Sciences, 2010. [刘峰. 基于地表动态反馈的平缓地区土壤空间推测研究[D]. 北京: 中国科学院大学, 2010.]
- [4] Duan M Q, Guo Z Q, Zhang X G, et al. Influences of different environmental covariates on county-scale soil type identification using remote sensing images[J]. *Ecological Indicators*, 2022, 139: 108951.
- [5] Yao D J, Yang J, Zhan X J. Feature selection algorithm based on random forest[J]. *Journal of Jilin University(Engineering and Technology Edition)*, 2014, 44(1): 137-141. [姚登举, 杨静, 詹晓娟. 基于随机森林的特征选择算法[J]. *吉林大学学报(工学版)*, 2014, 44(1): 137-141.]
- [6] Yang Y F, Han H W, Chen R, et al. Inference methods of digital soil mapping[J]. *Chinese Journal of Soil Science*, 2020, 51(5): 1016-1023. [杨雨菲, 韩浩武, 陈荣, 等. 数字土壤制图的推理方法对比研究[J]. *土壤通报*, 2020, 51(5): 1016-1023.]
- [7] Zhang X T, Huang W, Fu P H, et al. Research on digital soil mapping based on feature selection algorithm[J]. *Acta Pedologica Sinica*, 2024, 61(3): 635-647. [张晓婷, 黄魏, 傅佩红, 等. 基于特征筛选算法的数字土壤制图研究[J]. *土壤学报*, 2024, 61(3): 635-647.]
- [8] Li C P, Han M, Meng F X, et al. Survey of online ensemble classification algorithms for complex data streams[J]. *Application Research of Computers*, 2024, 41(3): 641-651. [李春鹏, 韩萌, 孟凡兴, 等. 复杂数据流在线集成分类算法综述[J]. *计算机应用研究*, 2024, 41(3): 641-651.]
- [9] Zhu A X, Lü G N, Zhou C H, et al. Geographic similarity: Third law of geography[J]. *Journal of Geo-information Science*, 2020, 22(4): 673-679. [朱阿兴, 闫国年, 周成虎, 等. 地理相似性: 地理学的第三定律[J]. *地球信息科学学报*, 2020, 22(4): 673-679.]
- [10] Li K, Huang W, Fu P H, et al. Research on soil type inference based on combinatorial cartography method[J]. *Acta Pedologica Sinica*, 2025, 62(2): 348-361. [李坤, 黄魏, 傅佩红, 等. 基于结合型制图方法的土壤类型推理研究[J]. *土壤学报*, 2025, 62(2): 348-361.]
- [11] Xu D Y, Jiang X S, Pan J J, et al. Research on the method of soil environment evaluation under different landform on a county-wide scale—A case study of Yizheng City, Jiangsu Province[J]. *Journal of Mountain Science*, 2007, 25(1): 45-53. [许端阳, 姜小三, 潘剑君, 等. 地貌差异条件下的县级尺度土壤环境综合评价方法——以江苏省仪征市为例[J]. *山地学报*, 2007, 25(1): 45-53.]
- [12] Zhao Z D. Study on updating method of historical soil map at county level in low topographic relief area[D]. Huainan, Anhui: Anhui University of Science & Technology, 2024. [赵治东. 低地形起伏区县级历史土壤图更新方法研究[D]. 安徽淮南: 安徽理工大学, 2024.]
- [13] Soil Survey Office of Yizheng County. Soil Survey of Yizheng County, Jiangsu Province[M]. Yizheng, Jiangsu: Soil Survey Office of Yizheng County, 1986. [仪征县土壤志普查办公室. 江苏省仪征县土壤志[M]. 江苏仪征: 仪征县土壤志普查办公室, 1986.]
- [14] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [15] Liaw A, Wiener M. Classification and regression by random forest[J]. *R News*, 2002, 2(3): 18-22.
- [16] Burrough P A, van Gaans P F M, Hootsmans R. Continuous classification in soil survey: Spatial correlation, confusion and boundaries[J]. *Geoderma*, 1997, 77(2): 115-135.

[17] Shao W, Yue W H, Zhang Y, et al. The application of machine learning techniques in geotechnical engineering: A review and comparison[J]. Mathematics, 2023, 11(18): 3976.

[18] Bergen S, Huso M M, Duerr A E, et al. A review of supervised learning methods for classifying animal behavioural states from environmental features[J]. Methods in Ecology and Evolution, 2023, 14(1): 189-202.

(责任编辑: 檀满枝)