

甄诚,王海燕,雷相东,等.基于随机森林模型的旺业甸实验林场土壤全氮数字制图[J].华中农业大学学报,2024,43(3):249-257.
DOI:10.13300/j.cnki.hnlkxb.2024.03.026

基于随机森林模型的旺业甸实验林场土壤全氮数字制图

甄诚¹,王海燕¹,雷相东²,赵晗¹,董齐琪¹,崔雪¹,仇皓雷¹

1.北京林业大学林学院/林木资源高效生产全国重点实验室/森林培育与保护教育部重点实验室,北京 100083;
2.中国林业科学研究院资源信息所,北京 100091

摘要 为探究林场土壤全氮含量的空间分布特征及对环境因素的响应,以旺业甸实验林场为研究区,采用随机森林模型和Cubist模型建立了不同土层深度(0~10、10~30、30~50 cm)土壤全氮含量与环境协变量(海拔、归一化植被指数、年平均降水量、年平均气温、y坐标和坡向)之间的定量关系模型,对该区土壤全氮含量进行预测并制图,并分析了影响土壤全氮空间变异的控制性因素。研究结果显示:0~10、10~30、30~50 cm土层实测全氮含量的均值分别为3.20、2.02、1.47 g/kg,土壤全氮的平均含量随土层深度的增加而降低;3个土层深度土壤全氮预测随机森林模型的决定系数 R^2 分别为0.59、0.42和0.39,均优于决定系数 R^2 分别为0.56、0.38和0.34的Cubist模型,2种模型预测精度都随土层深度的增加而降低,各环境因素对土壤全氮空间分布的影响作用随土层深度的增加而减小;从随机森林模型土壤全氮预测图来看,不同土层深度土壤全氮含量均呈现西部、北部和中部低,西南、东南和东部高的空间格局,不确定性图显示随机森林模型预测土壤全氮含量分布具有较低的标准差;海拔对土壤全氮含量的影响最大,其他依次为:归一化植被指数>年平均降水量>年平均气温>y坐标>坡向。结果表明,随机森林模型可以作为有效预测该林场不同土层深度土壤全氮含量的方法。

关键词 土壤全氮;随机森林模型;空间分布;土层深度;数字土壤制图

中图分类号 S714 **文献标识码** A **文章编号** 1000-2421(2024)03-0249-09

氮是生物地球化学循环的关键元素之一,影响土壤质量和其他土壤性质^[1-2],同时氮素缺乏对植物生长发育的影响很大,科学评价土壤氮素状况对土壤肥力管理和植物营养尤其重要。土壤氮库是生态系统中氮循环的重要组成部分,深层土壤氮库拥有巨大的氮储量,因此,了解土壤全氮(total nitrogen, TN)在垂直梯度上的分布及其控制因素的可靠估计对于研究土壤氮储量和氮循环极为必要^[3]。

基于土壤调查的传统土壤制图方法既耗时又昂贵,导致土壤地图很难更新。近年来,随着高精度数字森林土壤数据库的建立和科学利用土地的要求日益提高,数字土壤制图(digital soil mapping, DSM)技术作为一种为土壤评估框架提供数据和信息的新方法被用来描述世界各地土壤属性的空间分布,此法方便、经济高效^[4]。DSM基于土壤属性和预测变量之间的数字关系,使用空间分析和数学方法来理解土壤属性的空间格局^[5],也可以结合实验室土壤分析

和遥感光谱指数等环境变量,建立高分辨率、高精度的土壤属性预测模型。

基于McBratney提出的数字土壤制图框架,许多机器学习算法已被成功用于预测土壤属性^[4]。随机森林(random forest, RF)模型具有不易过拟合的优势,相比于其他机器学习方法更加有效稳定。如,姜赛平等^[6]运用3种空间预测模型对海南岛不同土层深度土壤有机质(soil organic matter, SOM)空间分布进行了预测,结果表明0~5、0~20、20~40 cm土层深度RF模型表现优于其他模型(决定系数 R^2 为0.21~0.37),40~60 cm土层深度普通克里格方法最优,但该方法图斑较大,无法详细描述SOM空间变异的细节信息。庞龙辉等^[7]使用RF模型对青海省表层土壤TN、粉粒、SOM和pH的空间分布格局进行预测,发现RF模型预测土壤属性空间格局效果优异, R^2 分别为0.611、0.474、0.532和0.542。Keskin等^[8]使用8种模型预测佛罗里达州土壤碳组分空间格局,结果

收稿日期:2023-11-30

基金项目:国家科技基础资源调查专项(2021FY100801);亚太森林组织项目

甄诚, E-mail: zhencheng@bjfu.edu.cn

通信作者: 王海燕, E-mail: haiyanwang72@aliyun.com

表明RF模型的预测效果最好, R^2 达到0.72,远高于其他模型。Cubist模型具有学习数据中的非线性关系的能力^[9-10]。Shahbazi等^[11]在伊朗东阿塞拜疆省使用RF模型和Cubist模型对4个标准化深度(0~15、15~30、30~60、60~100 cm)氮、磷和硼的空间分布进行建模和预测,结果表明Cubist模型预测效果优于RF模型。可以看出,不同模型在同一研究区域内效果不同,同一模型在不同研究区域内效果也并不相同,具有不确定性,且国内缺乏森林经营单位(如林场)层次的相关研究,需要更多的验证。

本研究以内蒙古赤峰市旺业甸实验林场为对象,基于土壤实测数据,结合多源数据集作为环境协变量,使用RF模型和Cubist模型探索影响土壤TN含量的环境协变量,对该林场不同土层深度土壤TN含量进行空间预测,形成林场TN数字地图,以为林场尺度森林土壤养分管理和可持续利用提供参考。

1 数据与方法

1.1 研究区概况

旺业甸实验林场(41°33'~41°49'N, 118°07'~118°30'E)位于内蒙古自治区赤峰市喀喇沁旗西南部,处于草原向森林的过渡地带。地貌类型为中山山地,海拔高度877~1890 m。属暖温带半干旱地区,为明显的大陆性季风气候。年平均降水量430~560 mm,年平均气温1.8~6.2℃,无霜期117 d。林场经营面积281 km²,有林面积约235 km²,其中人工林约118 km²,天然林约117 km²。土壤类型以典型棕壤为主,还包括暗棕壤、褐土和草甸土等。成土母岩主要有岩浆岩和沉积岩等,少数发育在黄土母质上。

1.2 土壤样品采集与测定

2021年9月,综合考虑林场地形地貌和林分类型等因素,采用典型取样方法,在该林场设置了147个采样点(图1),其中人工林71个,天然林76个,在每个样点设置一块20 m×30 m的样地,在样地内多点采集0~10、10~30、30~50 cm 3个土层深度的土壤,分层混匀后带回实验室。同时记录样点经纬度、海拔、坡度、坡向和林分类型等信息。混合土样经风干、去除杂质、研磨,通过孔径0.25 mm筛,采用凯氏定氮法测定土壤TN含量^[12]。

1.3 环境协变量数据来源及处理

本研究初步选择使用12个环境协变量作为预测土壤TN的因子(表1),包括x坐标、y坐标、地形、气

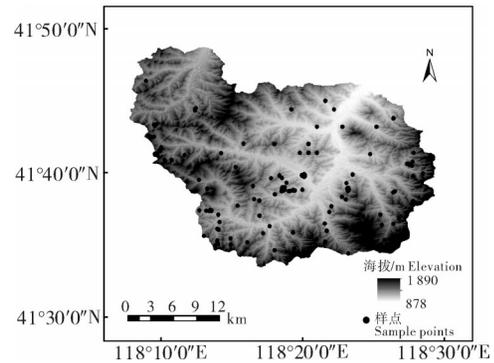


图1 旺业甸实验林场海拔及样点分布图

Fig. 1 Elevation and distribution of sampling sites in Wangyedian Experimental Forest Farm

候和生物变量数据。使用ArcGIS 10.5生成环境变量并将其传输到栅格层。地形属性是数字土壤制图中使用最广泛的环境预测因子^[11]。本研究中使用6个地形变量,包括海拔(elevation)、坡度(slope)、坡向(aspect)、剖面曲率(profile curvature)、地形粗糙指数(topographic roughness index, TRI)和地形湿度指数(topographic wetness index, TWI)。数据来源于地理空间数据云(<https://www.gscloud.cn/search>) GDEM V3数字高程数据。TRI和TWI使用SAGA GIS生成,其他4个变量使用ArcGIS 10.5生成。湿度和温度影响土壤氮的积累速度,被广泛应用于TN预测^[13]。采用研究区1970—2000年30 a的年平均气温(mean annual temperature, MAT)和年平均降水量(mean annual precipitation, MAP)作为气候变量,数据来源于世界气象数据库网站(<https://worldclim.org/>),在ArcGIS 10.5软件中使用最临近法将MAT和MAP数据重采样到30 m分辨率。生物变量使用归一化植被指数(normalized difference vegetation index, NDVI)和红边叶绿素植被指数(red edge chlorophyll index, RECI)进行表征^[14]。数据来源于地理空间数据云(<https://www.gscloud.cn/search>) Landsat 8 OLI图像,图像采集于2020年10月15日。使用ENVI 5.6对图像进行辐射定界和大气校正,采集近红外波段5(B5, 0.85~0.89 μm)和红外波段4(B4, 0.63~0.68 μm)用于计算NDVI和RECI,构建模型。

1.4 建模方法

随机森林(RF)是Breiman提出的一种基于树的集成学习技术^[15],RF模型中的节点纯度(IncNodePurity)可表征各环境因子对目标变量的贡献率,该值越大表示该变量的重要性越大^[16]。本研究

表1 环境协变量的来源和分辨率

Table 1 Source and resolution of environmental covariates

类型 Type	名称 Name	来源 Source	分辨率/m Resolution
地理坐标 Geographical coordinate	<i>x</i> 坐标 <i>x</i> -coordinate	实地测量 Field measurement	
	<i>y</i> 坐标 <i>y</i> -coordinate		
地形变量 Topographic variable	海拔 Elevation	GDEM V3	30
	坡度 Slope		30
	坡向 Aspect		30
	剖面曲率 Profile curvature		30
	地形粗糙指数 TRI		30
	地形湿度指数 TWI		30
气候变量 Climatic variable	年平均气温 MAT	WorldClim	1 000
	年平均降水量 MAP		1 000
生物变量 Biological variable	归一化植被指数 NDVI	Landsat 8	30
	红边叶绿素植被指数 RECI		30

中 RF 模型的 2 个重要参数 *mtry* 和 *ntree* 通过网格搜索确定为 2 和 800, 用于预测林场土壤 TN 含量分布。

Cubist 模型是一种先进的基于树的回归算法^[11]。本研究中 Cubist 模型的 2 个重要参数 *committees* 和 *neighbours* 分别设置为 10 和 9, 用于预测林场 TN 含

表2 不同土层深度土壤全氮含量的描述性统计

Table 2 Descriptive statistics of soil total nitrogen content at varied soil depths

土层深度/cm Soil depth	最小值/(g/kg) Minimum	最大值/(g/kg) Maximum	均值/(g/kg) Mean	标准差/(g/kg) Standard deviation	变异系数/% Coefficient of variation	偏度 Skewness	峰度 Kurtosis
0~10	0.83	9.01	3.20	1.68	52.38	0.90	0.54
10~30	0.46	6.23	2.02	1.05	51.86	1.05	1.36
30~50	0.31	5.89	1.47	0.83	56.36	1.95	6.25

2.2 环境协变量的选择

采用 Pearson 相关系数法进行环境协变量的选择。不同土层深度 TN 含量与环境协变量之间的 Pearson 相关分析结果表明(表 3), 0~10 cm 土层 TN 含量与海拔($r=0.533$)、坡向($r=0.178$)和年平均降水量($r=0.553$)呈正相关, 与 *y* 坐标($r=-0.205$)、年平均气温($r=-0.501$)和归一化植被指数($r=-0.265$)表现出负相关关系。*x* 坐标、坡度、剖面曲率、地形粗糙指数、地形湿度指数和红边叶绿素植被指数与 TN 含量相关性不显著, 给予剔除。

除 *y* 坐标和坡向与 30~50 cm 土层 TN 含量相关性不显著, 10~30 和 30~50 cm 土层 TN 含量与环境协变量的关系与 0~10 cm 土层基本一致。因此, 土

壤 TN 含量最终使用海拔、坡向、年平均降水量、*y* 坐标、年平均气温和归一化植被指数作为环境协变量进行建模预测。

1.5 模型验证与预测不确定性评估

使用交叉验证来评估模型的预测性能, 该方法的优点是在较小的数据集上执行, 可靠且无偏倚^[17]。本研究使用十折交叉验证, 重复 10 次, 用平均绝对误差(MAE)、均方根误差(RMSE)和 R^2 3 个统计指标来评估模型在预测土壤 TN 方面的性能。使用模型迭代 100 次预测的土壤全氮含量的标准差绘制模型预测的不确定性分布图。

2 结果与分析

2.1 土壤全氮的描述性统计分析

研究区 147 个样点 0~10、10~30 和 30~50 cm 3 个土层深度土壤 TN 的描述性统计信息见表 2。土壤 TN 含量的均值和标准差随着土层深度的增加而降低, 均值分别为 3.20、2.02、1.47 g/kg。3 个土层深度 TN 含量的变异系数为 51.86%~56.36%, 均属于中等变异^[18]。土壤 TN 经对数转换后符合正态分布。

不同林分类型土壤 TN 含量的分布情况如图 2 所示。从整体看, 不同土层深度天然林土壤 TN 含量高于人工林土壤 TN 含量。

壤 TN 含量最终使用海拔、坡向、年平均降水量、*y* 坐标、年平均气温和归一化植被指数作为环境协变量进行建模预测。

2.3 不同土层深度土壤全氮含量预测模型结果

利用样地数据训练的 RF 模型和 Cubist 模型的交叉验证结果如表 4 所示, 在不同土层深度下, RF 模型的预测效果均优于 Cubist 模型, 可以解释不同土层 TN 含量的空间变化范围(39%~59%), 且具有较高的 R^2 和较低的 RMSE 和 MAE。因此选用该 RF 模型对林场土壤 TN 的分布格局进行预测并制图。对于 0~10 cm 土层 TN, RF 模型的解释率超过 50%, 而对 30~50 cm 土层 TN 则表现出较低的预测性能($R^2=0.39$), RMSE 和 MAE 也随土层深度的增加而

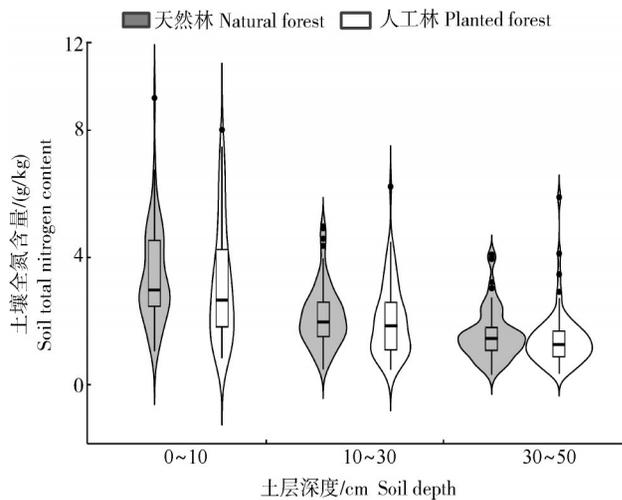


图2 不同林分类型不同土层深度下土壤全氮含量小提琴图
Fig. 2 Violin map of soil total nitrogen content under different soil depths and stand types

表3 不同土层深度下土壤全氮含量与环境协变量的相关性分析

Table 3 Correlation analysis of soil total nitrogen content and environmental covariates at varied soil depths

环境协变量 Environmental covariate	土层深度/cm Soil depth		
	0~10	10~30	30~50
x坐标 x-coordinate	-0.132	-0.145	-0.196*
y坐标 y-coordinate	-0.205*	-0.172*	-0.056
海拔 Elevation	0.533**	0.465**	0.518**
坡度 Slope	0.096	0.035	-0.071
坡向 Aspect	0.178*	0.155*	0.109
剖面曲率 Profile curvature	0.094	0.069	0.051
地形粗糙指数 TRI	0.099	0.064	-0.033
地形湿度指数 TWI	0.059	0.067	0.123
年平均气温 MAT	-0.501**	-0.400**	-0.464**
年平均降水量 MAP	0.553**	0.446**	0.484**
归一化植被指数 NDVI	-0.265**	-0.262**	-0.278**
红边叶绿素植被指数 RECI	-0.135	-0.146	-0.170*

注: *和**分别表示在0.05和0.01水平上显著相关。Note: * and ** indicate a significant correlation at 0.05 and 0.01 levels, respectively.

增大,这可能是因为用于建立与土壤属性关系的环境协变量多采集于地表,环境协变量的解释能力随着土层深度的增加而降低所致。

2.4 林场土壤全氮制图及不确定性分布图

由图3可知,0~10 cm土层TN含量预测值变化范围为1.08~5.89 g/kg,均值2.87 g/kg;10~30 cm土层TN预测值变化范围为0.64~4.37 g/kg,均值1.84 g/kg;30~50 cm土层TN预测值变化范围为0.58~3.74 g/kg,均值1.28 g/kg。土壤全氮的平均含量随土层深度的不断增加而降低。预测的3个土层

表4 不同土层深度土壤全氮含量预测的交叉验证结果

Table 4 Cross-validation results of prediction of soil total nitrogen content at varied soil depths

土层深度/cm Soil depth	RF			Cubist		
	R ²	RMSE	MAE	R ²	RMSE	MAE
0~10	0.59	0.36	0.28	0.56	0.37	0.30
10~30	0.41	0.41	0.33	0.38	0.44	0.35
30~50	0.39	0.43	0.34	0.34	0.46	0.38

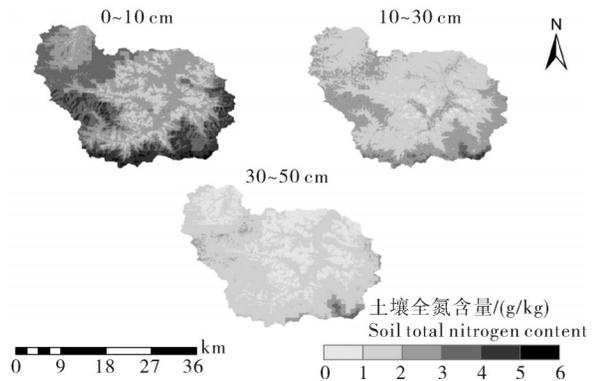


图3 不同土层深度土壤全氮空间预测图

Fig. 3 Spatial prediction of soil total nitrogen at varied soil depths

深度TN含量空间分布的总体变化趋势基本一致,均表现为西部、北部和中部低,西南、东南和东部高的空间格局,且与海拔的变化特征大致相符。

由图4可知,RF模型预测土壤TN的标准差的平均值随土层深度的增加不断提高,依次为0.013、0.018和0.019 g/kg,即使是标准差的最大值(0.032、0.035和0.036 g/kg)仍较低,总体上RF模型的预测不确定性很低。

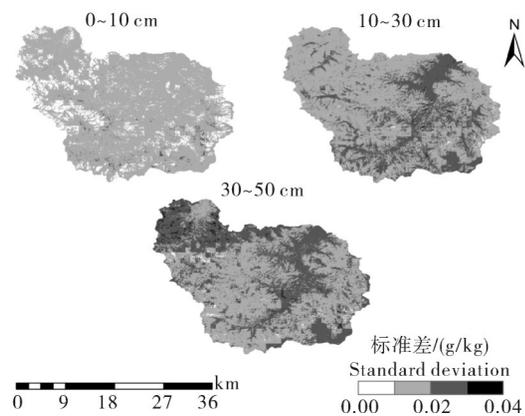


图4 随机森林模型预测土壤全氮分布的不确定性图

Fig. 4 Uncertainty map of soil total nitrogen distribution predicted by random forest model

2.5 土壤全氮对各环境协变量的响应

基于研究区土壤TN的预测数据和提取的环境

协变量数据,使用RF模型绘制了土壤TN对各环境协变量的响应曲线(图5),用于表征土壤TN与各环境协变量之间的关系。由图5可知,土壤TN与各环境协变量呈差异较大的非线性关系,不同土层深度土壤TN与各变量之间的响应曲线在趋势上大体相同。总体来看,土壤TN含量随海拔和年平均降水量的增大而增加,随年平均气温的增大而减小,这与

Pearson相关分析结果基本一致。土壤TN与坡向和y坐标表现为波动变化趋势。土壤TN与归一化植被指数之间的响应曲线较为特殊,随着归一化植被指数的增大,土壤TN先升高达到第一个峰值后迅速下降,再逐步上升到第二个峰值后递减,这可能是由于人工林和天然林2种林分类型的差异造成土壤TN与归一化植被指数之间呈复杂的响应关系。

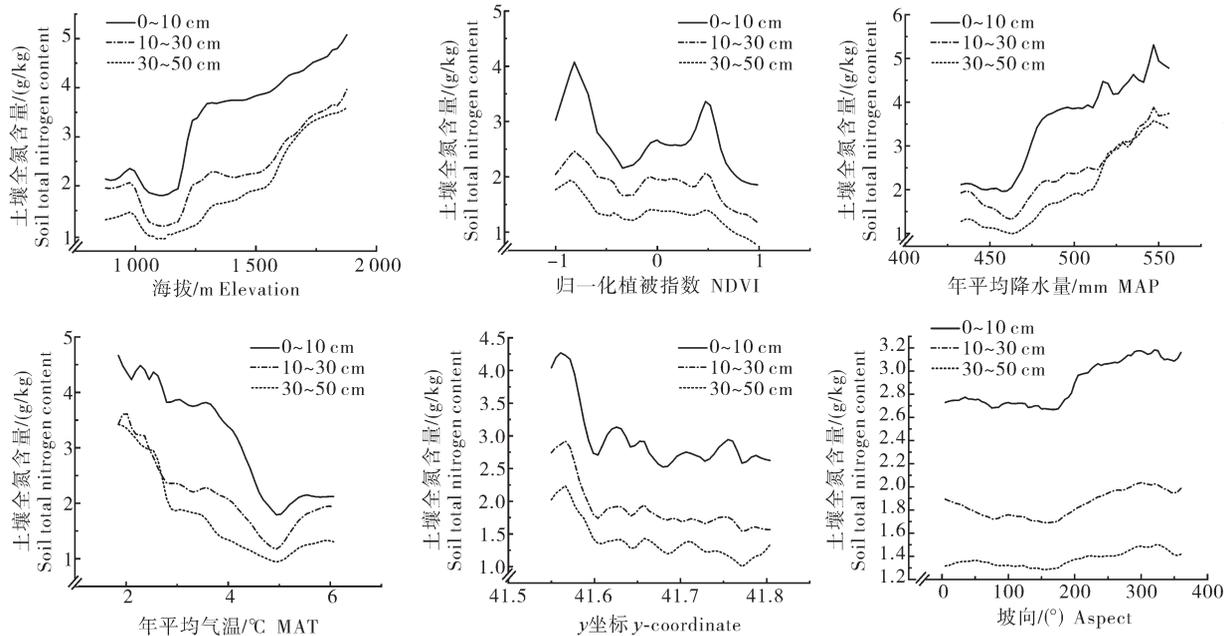


图5 不同土层深度土壤全氮含量对各环境协变量的响应曲线

Fig. 5 Response of soil total nitrogen content to environmental covariates at varied soil depths

2.6 环境协变量对土壤全氮影响的重要性

使用RF算法中环境协变量的节点纯度表征对应环境协变量的相对重要性。由图6可知,3个土层深度TN含量的分布格局基本一致。海拔对林场土壤TN含量影响最大,其他依次为:归一化植被指数>年平均降水量>年平均气温>y坐标>坡向。

旺业甸实验林场海拔877~1890 m,四周向中部逐渐递减,西南、东南和东部土壤在海拔高度的影响下人为干扰较少且水热条件较好^[19],分解和转化有机质的细菌的活性和种类在低温下受到限制^[20],土壤氮素得到累积;中部和北部主要为居民区,处于低海拔地区,同时植被覆盖较差,不利于土壤氮素的累积。

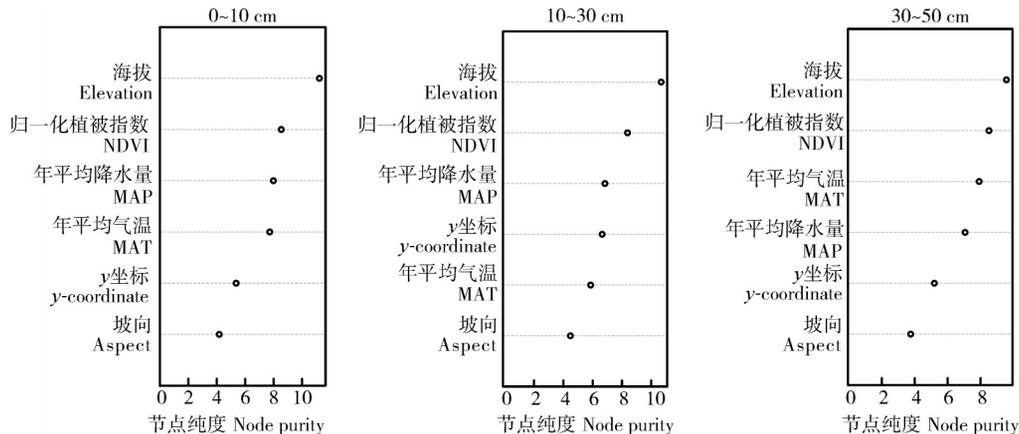


图6 随机森林模型环境协变量节点纯度

Fig. 6 The node purity of environmental covariates using random forest model

3 讨论

本研究在地形复杂的林场尺度下,对旺业甸实验林场147个样点,结合海拔、归一化植被指数、年平均气温、年平均降水量、坡向和 y 坐标等环境变量,利用RF模型和Cubist模型对0~10、10~30和30~50 cm 3个土层深度土壤TN含量的空间分布进行预测,并使用十折交叉验证方法进行验证,同时探究影响土壤全氮空间分布的主要变量。结果显示:①0~10、10~30和30~50 cm土层实测TN含量随着土层深度的增加逐渐减小;②相较于Cubist模型,RF模型能更好地预测3个土层深度土壤TN含量的空间分布,能解释39%~59%的土壤TN空间变异;③用RF模型预测土壤TN含量,不同土层深度TN含量均呈现西部、北部和中部低,西南、东南和东部高的空间格局;④土壤TN受到多种环境变量的综合影响,海拔是影响该林场土壤TN含量空间分布格局的最主要因素,其他依次为:归一化植被指数>年平均降水量>年平均气温> y 坐标>坡向。环境变量对土壤TN的影响随土层深度增加而减小。

在所有环境协变量中,地形变量对研究区内土壤TN含量空间分布的影响最大。这是因为对于小尺度数字土壤制图研究,地形变量不仅可以描述地形属性和地貌部位信息,也可以反映主要由地形地貌变化引起的局部小气候差异对土壤形成的影响,通常是小尺度下土壤属性空间变异的主导因素^[21]。地形是土壤形成的五大自然因素之一,对土壤TN的影响主要体现在影响地表植被分布和水热交换这2个方面^[22-23]。本研究发现,在所有地形变量中,海拔在预测旺业甸实验林场土壤TN空间分布的过程中发挥最重要的作用,旺业甸实验林场属中山山地地形,海拔高度差近1 000 m,地理坐标的改变会使森林组成产生较大的垂直地带性变化,进而影响土壤TN含量,与前人的研究结果一致^[24-26]。气温随海拔下降而上升,直接导致土壤温度升高,促进土壤微生物的活性,Yu等^[27]研究表明森林土壤反硝化作用产生气态氮损失随温度的升高而增加,减少了土壤TN的积累,导致不同海拔条件下TN含量存在差异。降水量随海拔的升高而上升,致使土壤水分含量增加,厌氧条件下微生物分解能力降低,有利于氮素积累。坡向主要制约土壤水热条件,进而对TN含量产生影响^[28]。在北半球,南坡比北坡受到更多的太阳辐射,植被蒸腾作用强烈,因此南坡水分蒸发快,土壤水分少,植被覆盖性差,土壤氮素积累较少,同时南坡的

强烈光照使土壤温度较高,土壤TN的矿化速率加快,不利于氮素积累。响应曲线显示,旺业甸实验林场土壤TN随坡向的增加呈现出先减小后增加的变化趋势,这是因为随着坡向值的增加,坡向由正北方向沿顺时针再次返回到正北方向。

地理坐标对土壤TN也有一定的影响, y 坐标与土壤TN呈负相关,表明研究区南部的土壤TN含量比研究区北部的要高。响应曲线则表现出随 y 坐标的增大,土壤TN先迅速下降,再表现出有波动的缓慢下降趋势,这与该地区复杂的地形有关。研究区南部为高海拔地区,基本不含居民区,而随着纬度的增加,地域逐渐开阔,密集的居民区也增加了人为因素对土壤TN的影响,土壤TN的变化范围也随之扩大。

一般情况下,在小尺度上,大生物气候因素对土壤属性空间变异的影响基本一致,但由于研究区内海拔梯度陡峭,在短距离内形成了温度和湿度等非生物因素梯度导致研究区生态系统更容易受到气候因素的影响^[29]。年平均降水量对土壤TN的空间分布具有较强的影响,本研究区年平均降水量的极差高达105 mm,差异明显,降水丰富将有效增加土壤含水量,进而增加土壤TN的积累。研究区年平均气温的极差为3.9℃,差异较小,对于土壤TN含量的空间分布的影响较弱。

在本研究中,生物变量对土壤TN的影响仅次于地形变量。NDVI数据表征植被生长状况和植被覆盖度等植被活动^[30-31]。Pearson相关分析结果表明,土壤TN与NDVI呈显著的负相关关系,这与前人的研究结果不一致^[32],究其原因可能是研究区为小尺度的林场,有林地面积超过83%,人工林和天然林面积相近,而人工林林分密度大,郁闭度大,因此具有较高的NDVI。Chen等^[33]研究指出树木多样性与土壤N积累具有正相关关系。以自然更新为主的天然林群落结构复杂,具有丰富的枯落物,枯落物直接归还的有机质多,有利于土壤氮素积累,因此,天然林土壤TN含量高于人工林,NDVI高的地区土壤TN低。

RF模型在0~10 cm土层深度精度指标值均优于10~30、30~50 cm土层深度。模型对土壤属性空间分布的预测精度主要受环境的复杂性、土壤属性的空间变异性和模型的预测性能的影响^[34]。人为活动会在土壤表层产生相对均匀的环境,同时本研究最终选定的用于与土壤TN建立关系的环境协变量

大多数是从地表收集的,这些因素有助于提高模型在0~10 cm 土层深度的预测精度。在一般情况下,深层土壤环境复杂,模型的预测性能较低。土壤属性在垂直梯度上的分布及环境因素对其控制性影响仍是土壤学研究的重点,故应加强对深层土壤环境的研究并提高深层土壤属性的预测精度。

本研究在构建土壤 TN 含量的预测模型时,在环境协变量的选择上未包含对土壤 TN 含量影响较强的土壤其他属性(pH、有机质、速效钾等),土壤数据多是通过不同制图方法获取,本身存在误差和不确定性,进而可能影响土壤 TN 含量预测的准确性;在土壤 TN 制图时未考虑人为因素的影响,林场的业务涉及森林经营、种苗培育、木材生产和旅游开发等人类活动,而本研究选取的环境协变量多为自然因素,缺乏能很好表征人为因素的变量,未来若可以将林场经营模式、采伐强度等因素纳入模型中,可进一步完善林场土壤属性预测模型;研究区内天然林土壤 TN 明显高于人工林,若后期能获得该林场林分类型数据,设置哑变量结合其他环境变量数据进行模型建模,可能会在一定程度上提高模型预测效果和精度。

参考文献 References

- [1] WANG K, ZHANG C R, LI W D. Predictive mapping of soil total nitrogen at a regional scale: a comparison between geographically weighted regression and cokriging [J]. *Applied geography*, 2013, 42: 73-85.
- [2] YANG X, REYNOLDS W D, DRURY C F, et al. Organic carbon and nitrogen stocks in a clay loam soil 10 years after a single compost application [J]. *Canadian journal of soil science*, 2014, 94(3): 357-363.
- [3] 李龙, 秦富仓, 姜丽娜, 等. 区县域尺度土壤全氮的空间分布格局分析[J]. *生态学报*, 2020, 40(5): 1572-1579. LI L, QIN F C, JIANG L N, et al. Spatial distribution patterns of soil total nitrogen at a county scale [J]. *Acta ecologica sinica*, 2020, 40(5): 1572-1579 (in Chinese with English abstract).
- [4] MCBRATNEY A, SANTOS M, MINASNY B. On digital soil mapping [J]. *Geoderma*, 2003, 117(1/2): 3-52.
- [5] WADOUX M J C, SAMUEL-ROSA A, POGGIO L, et al. A note on knowledge discovery and machine learning in digital soil mapping [J]. *European journal of soil science*, 2020, 71(2): 133-136.
- [6] 姜赛平, 张怀志, 张认连, 等. 基于三种空间预测模型的海南岛土壤有机质空间分布研究[J]. *土壤学报*, 2018, 55(4): 1007-1017. JIANG S P, ZHANG H Z, ZHANG R L, et al. Research on spatial distribution of soil organic matter in Hainan Island based on three spatial prediction models [J]. *Acta pedologica sinica*, 2018, 55(4): 1007-1017 (in Chinese with English abstract).
- [7] 庞龙辉, 刘峰, 赵霞, 等. 青海省表层土壤属性数字制图[J]. *土壤通报*, 2019, 50(3): 505-513. PANG L H, LIU F, ZHAO X, et al. Digital mapping of topsoil attributes in Qinghai Province [J]. *Chinese journal of soil science*, 2019, 50(3): 505-513 (in Chinese with English abstract).
- [8] KESKIN H, GRUNWALD S, HARRIS W G. Digital mapping of soil carbon fractions with machine learning [J]. *Geoderma*, 2019, 339: 40-58.
- [9] MA Y X, MINASNY B, WU C F. Mapping key soil properties to support agricultural production in Eastern China [J]. *Geoderma regional*, 2017, 10: 144-153.
- [10] RUDIYANTO, MINASNY B, SETIAWAN B I, et al. Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands [J]. *Geoderma*, 2018, 313: 25-40.
- [11] SHAHBAZI F, HUGHES P, MCBRATNEY A B, et al. Evaluating the spatial and vertical distribution of agriculturally important nutrients: nitrogen, phosphorous and boron in North West Iran [J]. *Catena*, 2019, 173: 71-82.
- [12] 鲍士旦. 土壤农化分析 [M]. 3版. 北京: 中国农业出版社, 2000: 42-49. BAO S D. Soil and agricultural chemistry analysis [M]. 3rd ed. Beijing: China Agriculture Press, 2000: 42-49 (in Chinese).
- [13] WANG S, ZHUANG Q L, WANG Q B, et al. Mapping stocks of soil organic carbon and soil total nitrogen in Liaoning Province of China [J]. *Geoderma*, 2017, 305: 250-263.
- [14] BARDGETT R D, STREETER T C, BOL R. Soil microbes compete effectively with plants for organic-nitrogen inputs to temperate grasslands [J]. *Ecology*, 2003, 84(5): 1277-1287.
- [15] BREIMAN L. Random forests [J]. *Machine learning*, 2001, 45(1): 5-32.
- [16] 雷相东. 机器学习算法在森林生长收获预估中的应用 [J]. *北京林业大学学报*, 2019, 41(12): 23-36. LEI X D. Applications of machine learning algorithms in forest growth and yield prediction [J]. *Journal of Beijing Forestry University*, 2019, 41(12): 23-36 (in Chinese with English abstract).
- [17] WONG T T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation [J]. *Pattern recognition*, 2015, 48(9): 2839-2846.
- [18] 刘继龙, 刘璐, 马孝义, 等. 不同尺度不同土层土壤盐分的空间变异性研究 [J]. *应用基础与工程科学学报*, 2018, 26(2): 305-312. LIU J L, LIU L, MA X Y, et al. Spatial variability of soil salt in different soil layers at different scales [J]. *Journal of basic science and engineering*, 2018, 26(2): 305-312 (in Chinese with English abstract).
- [19] ZHAO Z Z, ZHANG X F, DONG S K, et al. Soil organic carbon and total nitrogen stocks in alpine ecosystems of Altun

- Mountain National Nature Reserve in dry China[J/OL]. Environmental monitoring and assessment, 2018, 191(1): 40 [2023-11-30]. <https://doi.org/10.1007/s10661-018-7138-9>.
- [20] TSUI C, CHEN Z, HSIEH C. Relationships between soil properties and slope position in a lowland rain forest of southern Taiwan[J]. Geoderma, 2004, 123(1/2): 131-142.
- [21] 朱阿兴, 杨琳, 樊乃卿, 等. 数字土壤制图研究综述与展望[J]. 地理科学进展, 2018, 37(1): 66-78. ZHU A X, YANG L, FAN N Q, et al. The review and outlook of digital soil mapping [J]. Progress in geography, 2018, 37(1): 66-78 (in Chinese with English abstract).
- [22] YANG R, ZHANG G, LIU F, et al. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem[J]. Ecological indicators, 2016, 60(8): 870-878.
- [23] 程先富, 史学正. 亚热带典型地区土壤全氮和地形、母岩的关系研究: 以江西省兴国县为例[J]. 水土保持学报, 2004, 18(2): 137-139. CHENG X F, SHI X Z. Relationship between total nitrogen and topographic, parent material in typical subtropical region—a case study in Xingguo County, Jiangxi Province [J]. Journal of soil water conservation, 2004, 18(2): 137-139 (in Chinese with English abstract).
- [24] 郭澎涛, 李茂芬, 罗微, 等. 基于多源环境变量和随机森林的橡胶园土壤全氮含量预测[J]. 农业工程学报, 2015, 31(5): 194-200. GUO P T, LI M F, LUO W, et al. Prediction of soil total nitrogen for rubber plantation at regional scale based on environmental variables and random forest approach[J]. Transactions of the CSAE, 2015, 31(5): 194-200 (in Chinese with English abstract).
- [25] 李龙, 姚云峰, 秦富仓, 等. 黄花甸子流域土壤全氮含量空间分布及其影响因素[J]. 应用生态学报, 2015, 26(5): 1306-1312. LI L, YAO Y F, QIN F C, et al. Spatial distribution pattern of soil nitrogen in Huanghuadianzi watershed and related affecting factors[J]. Chinese journal of applied ecology, 2015, 26(5): 1306-1312 (in Chinese with English abstract).
- [26] 周洋, 赵小敏, 郭熙. 基于多源辅助变量和随机森林模型的表层土壤全氮分布预测[J]. 土壤学报, 2022, 59(2): 451-460. ZHOU Y, ZHAO X M, GUO X. Prediction of total nitrogen distribution in surface soil based on multi-source auxiliary variables and random forest approach[J]. Acta pedologica sinica, 2022, 59(2): 451-460 (in Chinese with English abstract).
- [27] YU H M, DUAN Y H, MULDER J, et al. Universal temperature sensitivity of denitrification nitrogen losses in forest soils [J]. Nature climate change, 2023, 13: 726-734.
- [28] 郭永龙, 毕如田, 王瑾, 等. 华北典型山区忻州不同小生境坡耕地土壤肥力特征[J]. 水土保持学报, 2013, 27(5): 205-208. GUO Y L, BI R T, WANG J, et al. Soil fertility development trends in sloping lands under different habitats in Xinzhou of North China[J]. Journal of soil and water conservation, 2013, 27(5): 205-208 (in Chinese with English abstract).
- [29] TASHI S, SINGH B, KEITEL C, et al. Soil carbon and nitrogen stocks in forests along an altitudinal gradient in the eastern Himalayas and a meta-analysis of global data [J]. Global change biology, 2016, 22(6): 2255-2268.
- [30] 高江波, 焦珂伟, 吴绍洪. 1982—2013年中国植被NDVI空间异质性的气候影响分析[J]. 地理学报, 2019, 74(3): 534-543. GAO J B, JIAO K W, WU S H. Revealing the climatic impacts on spatial heterogeneity of NDVI in China during 1982—2013 [J]. Acta geographica sinica, 2019, 74(3): 534-543 (in Chinese with English abstract).
- [31] 韩雅, 朱文博, 李双成. 基于GWR模型的中国NDVI与气候因子的相关分析[J]. 北京大学学报(自然科学版), 2016, 52(6): 1125-1133. HAN Y, ZHU W B, LI S C. Modelling relationship between NDVI and climatic factors in China using geographically weighted regression [J]. Acta scientiarum naturalium universitatis Pekinensis, 2016, 52(6): 1125-1133 (in Chinese with English abstract).
- [32] WEI X, HAO M, SHAO M, et al. Changes in soil properties and the availability of soil micronutrients after 18 years of cropping and fertilization [J]. Soil & tillage research, 2006, 91(1/2): 120-130.
- [33] CHEN X L, TAYLOR A R, REICH P B, et al. Tree diversity increases decadal forest soil carbon and nitrogen accrual [J]. Nature, 2023, 618(7963): 94-101.
- [34] ZHANG Y K, JI W J, SAURETTE D D, et al. Three-dimensional digital soil mapping of multiple soil properties at a field-scale using regression Kriging [J/OL]. Geoderma, 2020, 366: 114253 [2023-11-30]. <https://doi.org/10.1016/j.geoderma.2020.114253>.

Digital mapping of soil total nitrogen in Wangyedian experimental forest farm based on random forest model

ZHEN Cheng¹, WANG Haiyan¹, LEI Xiangdong², ZHAO Han¹, DONG Qiqi¹, CUI Xue¹, QIU Haolei¹

1. *College of Forestry/State Key Laboratory of Efficient Production of Forest Resources/Key Laboratory for Silviculture and Conservation of Ministry of Education, Beijing Forestry University, Beijing 100083, China;*

2. *Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China*

Abstract To explore the spatial distribution characteristics of soil total nitrogen content in forest farm and its response to environmental factors, random forest model and Cubist model were used to establish a quantitative relationship between soil total nitrogen content and environmental covariates including elevation, normalized difference vegetation index, mean annual precipitation, mean annual temperature, y-coordinate and aspect at soil depths of 0-10, 10-30 and 30-50 cm in Wangyedian experimental forest farm. Soil total nitrogen content of the area was predicted and mapped, and the controlling factors affecting the spatial variation of soil total nitrogen were analyzed. The results showed that the average content of soil total nitrogen at soil depth of 0-10, 10-30 and 30-50 cm was 3.20, 2.02 and 1.47 g/kg, respectively. It decreased with the increase of soil depth. The results of cross-validation showed that the R^2 of the random forest model for predicting soil total nitrogen at the three soil depths was 0.59, 0.42, and 0.39, respectively, better than the R^2 of Cubist model with 0.56, 0.38, and 0.34, respectively. The prediction accuracy of both models decreased with the increase of soil depth. The influence of various environmental factors on the spatial distribution of soil total nitrogen decreased with the increase of soil depth. From the prediction map of soil total nitrogen with the random forest model, the content of soil total nitrogen at different soil depth showed a spatial pattern of low in the western, northern, and central regions, and high in the southwestern, southeastern, and eastern regions. The uncertainty map showed that the random forest model had a low standard deviation in predicting the distribution of the content of soil total nitrogen. The elevation had the greatest impact on the content of soil total nitrogen, followed by the normalized difference vegetation index > mean annual precipitation > mean annual temperature > y-coordinate > aspect. It is indicated that the random forest model can serve as an effective method for predicting the content of soil total nitrogen at different soil depth in the forest farm.

Keywords soil total nitrogen; random forest model; spatial distribution; soil depth; digital soil mapping

(责任编辑:葛晓霞)