

## 大面积高寒山区土壤养分空间预测与管理分区<sup>①</sup>

杜龙全<sup>1,2,3</sup>, 刘 峰<sup>3,4</sup>, 史 舟<sup>5</sup>, 赵 霞<sup>1,2\*</sup>, 李德成<sup>3</sup>, 张甘霖<sup>3,4,6</sup>, 董晋鹏<sup>1,2</sup>, 陈东升<sup>1,2</sup>

(1 青海师范大学, 青海土壤数字服务中心, 青海省自然地理与环境过程重点实验室, 西宁 810008; 2 青海省人民政府-北京师范大学高原科学与可持续发展研究院-高原土壤信息科学研究团队, 西宁 810008; 3 土壤与农业可持续发展国家重点实验室(中国科学院南京土壤研究所), 南京 210008; 4 中国科学院大学, 北京 100049; 5 浙江大学环境与资源学院, 杭州 310058; 6 中国科学院南京地理与湖泊研究所流域地理学重点实验室, 南京 210008)

**摘要:** 本文基于青海省土系调查的 205 个土壤剖面数据, 利用随机森林模型, 分别建立了表层(0~20 cm)土壤全氮(TN)、全钾(TK)、全磷(TP)与环境因素变量(地形、气候、植被、遥感数据)之间的定量关系, 对青海省土壤养分含量的空间分布进行了预测; 在此基础上, 结合全国土壤养分的分级标准, 利用投影寻踪法, 生成了土壤养分的管理分区。留一交叉验证结果显示, TN、TK、TP 空间预测的  $R^2$  分别是 0.89、0.85、0.82, 可解释土壤养分空间变异的 80% 以上, 表明随机森林模型和环境因素变量结合可以在稀疏样本条件下有效地预测大面积高寒山区土壤养分空间变异。青海省土壤养分呈现东高西低的分布模式, 土壤综合养分高等级出现在南部的玉树、果洛、黄南和东部的湟水谷地地区; 低等级主要分布在柴达木盆地、可可西里、海南州中南部; 全省土壤综合养分分级均在中上等级以上, 占全省面积的 81%。分析发现, 植被是影响青海省表层土壤 TN、TP、TK 空间分布的主要环境因素, 其中年均降水量、地表温度是影响青海省表层土壤 TN 空间模式的重要因素; 地表植被、海拔和地表温度等环境因子对表层土壤 TP 的空间变异占主导作用; 年均降水量、昼夜温差是影响表层土壤 TK 的空间模式的重要因素。

**关键词:** 土壤养分; 随机森林; 投影寻踪法; 管理分区; 数字土壤制图

中图分类号: S159 文献标志码: A

## Spatial Prediction and Management Zoning of Soil Nutrients in Large-scale Alpine Mountainous Areas

DU Longquan<sup>1,2,3</sup>, LIU Feng<sup>3,4</sup>, SHI Zhou<sup>5</sup>, ZHAO Xia<sup>1,2\*</sup>, LI Decheng<sup>3</sup>, ZHANG Ganlin<sup>3,4,6</sup>, DONG Jinpeng<sup>1,2</sup>, CHEN Dongsheng<sup>1,2</sup>

(1 Qinghai Normal University, Qinghai Soil Digital Service Center, Physical Geography and Environmental Process Key Laboratory of Qinhai Province, Xining 810008, China; 2 Qinghai Normal University Academy of Plateau Science and Sustainability-Plateau Soil Information Science Research Team, People's Government of Qinghai Province & Beijing Normal University, Xining 810008, China; 3 State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China; 4 University of Chinese Academy of Sciences, Beijing 100049, China; 5 College of Environmental & Resource Sciences, Zhejiang University, Hangzhou 310058, China; 6 Key Laboratory of Watershed Geography Sciences, Nanjing Institute of Geography & Limnology, Chinese Academy of Sciences, Nanjing 210008, China)

**Abstract:** Based on the data of 205 sample points of soil series survey in Qinghai Province in recent years, the quantitative relationships between the contents of total nitrogen (TN), total potassium (TK) and total phosphorus (TP) of topsoils (0~20 cm) and environmental factor variables (terrain, climate, vegetation and remote sensing data) were established respectively by using random forest model, and the spatial distribution of soil nutrient contents in Qinghai Province was predicted, then the management zoning of soil nutrients was generated by using the projection pursuit method and the national soil nutrient classification standard. The cross validation results show that  $R^2$  of spatial prediction of TN, TK and TP are 0.89, 0.85 and 0.82, respectively. The model can explain more than 80% of the spatial variation of soil nutrients, indicating that the combination of random forest model and environmental factor variables can effectively predict the spatial variation of soil nutrients in large-area

<sup>①</sup>基金项目: 国家自然科学基金重点项目(41930754)、国家自然科学基金地区项目(42067001)、青海省科技厅科技合作专项(2018-HZ-804)资助。

\* 通讯作者(zhaoxia@qhnu.edu.cn)

作者简介: 杜龙全(1994—), 男, 山西吕梁人, 硕士研究生, 主要从事资源环境与生态评估研究。E-mail: 15135132501@163.com

alpine mountainous areas under the condition of sparse samples. The distribution pattern of soil nutrients in Qinghai Province is high in the east and low in the west. The high levels of soil comprehensive nutrients appear in Yushu, Guoluo, Huangnan in the south and Huangshui Valley in the east; The lower grades are mainly distributed in Qaidam Basin, Hoh Xil and the south central part of Hainan prefecture; The soil nutrient classification of the whole province is above the middle and upper grades, accounting for 81% of the total area of the whole province. It is found that vegetation is the main environmental factor affecting the spatial distribution of soil nutrients in topsoil in Qinghai Province, among which annual precipitation and surface temperature are important factors affecting the spatial model of TN in topsoil in Qinghai Province. The spatial variation of TP in topsoil was dominated by environmental factors such as surface cover, altitude and surface temperature. Annual precipitation and temperature difference between day and night are important factors affecting the spatial model of TK in topsoil.

**Key words:** Soil nutrients; Random forest; Projection pursuit method; Management partition; Digital soil mapping

土壤养分是土壤提供的植物生长所必需的营养元素<sup>[1]</sup>, 是土壤生产力的基础, 了解土壤养分的空间变异, 对于研究土壤退化、生物多样性缺失、气候变化和制定土地管理计划具有重要意义<sup>[2-3]</sup>。我国西部省区, 尤其是高寒山区, 土壤景观关系复杂, 土壤空间变异程度高<sup>[4]</sup>, 但缺乏准确的土壤空间信息, 制约着农牧业的科学生产与管理。

数字土壤制图是将土壤-景观模型通过计算机辅助生成的以栅格数据为表达形式的制图方法, 可反映较为准确的土壤空间信息。制图模型主要包括地统计、专家知识和机器学习方法, 研究内容集中于土壤的多种属性。在区域从小到大、景观从简单到复杂、制图结果从二维到三维(多维)的制图研究中, 环境变量选取的是相对较少且易获取的协变量集, 主要集中于地形因子与气候因子以及用遥感影像来表征的生物因子。涉及大面积区域时, 成土母质这一因子的获取较难, 往往被忽略; 同时, 表征土壤和环境之间关系的一些成土因子定量化很难表示(如时间因子)<sup>[5-12]</sup>。因此, 要实现大尺度、高精度的制图, 开发挖掘丰富的环境协变量数据库是一个重要的突破口。在现有的机器学习(集成)模型甚至是人工智能模型中, 只能反映小区域、局部景观环境与土壤的关系, 模型对于与研究区环境条件相似的地区可能适用性较好, 对于复杂的大面积区域适用性较低<sup>[12-14]</sup>。要精确地模拟土壤性质和过程, 有必要开发具有处理不同区域土壤与环境之间非线性关系的针对性的模型。目前国内数字土壤制图工作主要集中在我国中部和东部地区, 地形梯度相对较小, 可达性较好, 土壤调查样本通常数量较多且分布相对均衡。例如, 杨煜岑等<sup>[5]</sup>在陕西省蓝田县用多元线性回归模型预测了陕西省蓝田县土壤养分的空间分布; Dong 等<sup>[6]</sup>将机器学习算法与基于地块的 DSM 框架相结合, 对宁夏冲积平原农业区表层土壤养分(土壤有机质、全

氮等)特征进行了预测; Liu 等<sup>[7]</sup>用地表动态反馈方法预测了江苏省中部平原地区土壤有机质含量等属性的空间分布。然而, 在我国西部大面积高寒山区, 路网不发达, 可达性较差, 土壤调查样本较为稀疏, 已有工作主要集中在景观或流域尺度以及单一土壤养分属性(如有机碳或全氮), 如代子俊等<sup>[8]</sup>针对湟水流域分析了近 30 a 的土壤全氮的时空变异影响因素; 张欢和高小红<sup>[9]</sup>基于湟水流域土壤有机质属性对比分析了地统计学进行预测制图的精度; 周伟等<sup>[10]</sup>在三江源地区将原始光谱反射率数据及其不同数据变换形式下的光谱分别与土壤有机质含量进行了相关分析; 庞龙辉等<sup>[11]</sup>对青海省土壤全氮、有机碳、粉粒和 pH 进行了预测制图。如何对大面积高寒山区土壤养分空间分布准确地预测, 并在此基础上面向农牧业进行合理的生产管理分区, 是一个值得研究的问题。

本文以地处青藏高原东北部的青海全省为研究区, 将机器学习方法与成土环境协同变量相结合, 探讨土壤养分(全氮(TN)、全钾(TK)、全磷(TP))含量的空间预测制图及生产管理分区, 以期为青海省土壤养分空间管理及农牧业可持续发展提供必要参考。

## 1 材料与方法

### 1.1 研究区概况

青海省地处我国西北部内陆腹地, 地理位置为  $31^{\circ}39' \sim 39^{\circ}19'N$ 、 $89^{\circ}35' \sim 103^{\circ}04'E$ , 总面积 72.23 万  $km^2$ , 大部分区域为高寒山区, 地势总趋势为南高北低<sup>[15]</sup>, 海拔介于 1 686 ~ 6 477 m, 大部分地区海拔在 3 000 m 以上。地貌上复杂多样兼有高原、山地、草原、湖泊; 气候上, 虽然总体上属于高原大陆性气候, 但属暖温带向亚热带过渡型, 不同海拔高度和不同地区气候有明显差异, 年均温度  $-5.1 \sim 9.0^{\circ}C$ , 年均降水量  $15 \sim 750 mm$ , 降水量由东向西逐渐递减。按大地构造全省大体可分为三类不同地形区: 南部青

南高原、西部柴达木盆地和北部与东部平行岭谷区<sup>[4]</sup>。区域内成土母质类型较多, 成因各异。受地形、气候、土壤环境的影响, 草地面积占全省的3/5, 畜牧业得

以发展, 全省的森林、草原、湿地等在应对气候变化上有显著的作用, 是我国重要的生态屏障。研究区采样点空间分布见图1。

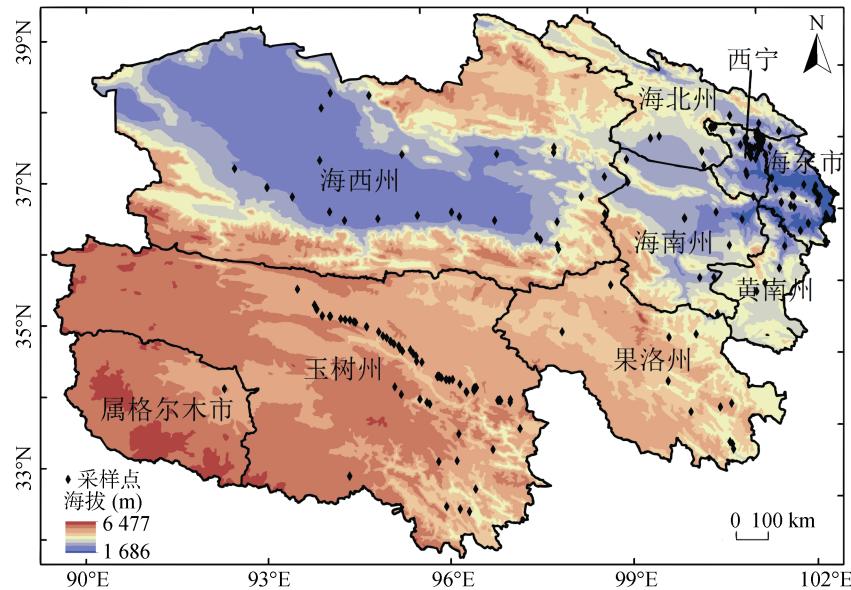


图1 研究区采样点空间分布  
Fig. 1 Spatial distribution of sampling points in study area

## 1.2 数据来源与处理

土壤剖面数据来源于科技部基础专项“我国土系调查与《中国土系志·青海卷》编制”项目中的205个样点, 主要包括采样点位置和景观信息以及土壤理化性质(TN, TP, TK, SOC, pH, 黏粒、粉粒、砂粒含量等)。土壤TN采用重铬酸钾、硫酸消化-蒸馏法(凯氏蒸馏法)测定; TK使用碱熔-火焰光度法测定; TP使用碱熔-钼锑抗比色法测定<sup>[16]</sup>。对剖面各发生层测定的数据进行深度加权, 生成表层(0~20 cm)的TN、TK、TP数据。

环境变量的选取是以土壤发生学理论为依据, 主要依照土壤形成的五大要素(母质、地形、生物、气候和时间)进行选取代表性环境变量。由于目前尚无统一的定量指标来描述时间信息, 因此本研究暂不考虑时间类因子。用土壤类型(soil types)近似表征母质的差异; 地形因子在土壤发生与演变过程中影响着地表物质能量的循环转换; 遥感数据具有实时、易获取的特点, 其特征波段及不同的组合可以用来表征生物景观信息的综合环境(地表景观信息)。气候因子(降水量、气温)是直接影响成土过程中的水热条件。本研究通过表层土壤属性及其与环境变量之间的相关性确定了比较关键的21个环境变量。其中, 地形因子: DEM数据来自于资源环境科学与数据中心(<http://www.resdc.cn/>), 基于DEM数据利用

SAGA软件(<https://sourceforge.net/>)计算提取海拔(Elev)、坡度(Slp)、坡向(Asp)、剖面曲率(Curprf)、地形湿度指数(Twi)、地形开放度(Openness)、多尺度山谷指数(Mrvbf); 气候数据: 年均温(MAT)、年均降水量(MAP)、气温季节性(Ts)、昼夜温差(Mdr)、潜在太阳辐射(Radition)数据来自于世界气象数据库网<sup>[17]</sup>(<http://worldclim.org/version2>); 土壤类型(Soiltype)和土地覆盖类型(Lucc)均来源于资源环境科学与数据中心(<http://www.resdc.cn/>); 植被指数均值(Evm), 植被指数标准差(Evs), 白天地表温度(LSTday), 夜间地表温度(LSTnight), 7、8月地表温度(LST78)和地表反射率(band4/7)的处理是在Google Earth Engine(GEE)云平台通过JavaScript编程完成, 选取时间为2015年7—8月的MODIS影像数据集, 处理包括影像去云、掩膜裁剪、投影变换。所有栅格环境变量均重采样(最近邻法)为1 km。

## 1.3 土壤养分空间预测方法

随机森林<sup>[18]</sup>是通过建立多个决策树并将其融合成一个更加准确和稳定的模型, 能够根据样本之间的相似性来衡量特征交互, 广泛应用于分类和回归。随机森林建模在R.4.1中Random Forest包进行<sup>[19]</sup>, 该模型主要有3个参数需要定义, 即树的数量(ntree)、叶节点包含最少样本的个数(nodesize)、随机选取变量的个数(mtry)。使用caret包对nodesize和mtry两

个参数进行优化, 优化值分别是 5 和 3, ntree 采用默认值 500, 足以产生稳定的预测。%IncMSE 作为判断预测变量重要性的指标, 该值越大表示该变量的重要性越大, 即对土壤养分含量的影响度就越大。

本文中制图预测精度采用留一交叉验证法 LOOCV(leave one out cross validation)。其原理是: 在个数为  $N$  数据集中随机抽取 1 个样本去验证  $N-1$  个数据所训练的模型, 重复  $N$  次取平均值<sup>[20]</sup>。评价指标包括: 决定系数  $R^2$ 、一致性指数(CC)、均方根误差(RMSE)、预测偏差值(ME)。其中,  $R^2$  表示模型对自变量解释的比例, 取值范围  $0 \sim 1$ ; CC 表示预测误差的方差值占实测值的比, 取值范围  $0 \sim 1$ , 1 表示预测结果最佳, 0 表示结果不太符合; RMSE 是预测值与实测值之间的均方根误差; ME 是预测值与实测值之间误差值的平均, 反映预测结果情况是否存在高估或低估。公式如下:

$$R^2 = \frac{\sum_{i=1}^N \{p(x_i) - \bar{p}\}^2}{\sum_{i=1}^N \{v(x_i) - \bar{p}\}^2} \quad (1)$$

$$CC = \frac{2R\sigma_p\sigma_v}{\sigma_p^2 + \sigma_v^2 + (\bar{v} - \bar{P})^2} \quad (2)$$

$$ME = \frac{1}{N} \sum_{i=1}^N \{p(x_i) - v(x_i)\} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \{p(x_i) - v(x_i)\}^2} \quad (4)$$

式中:  $N$  为样本的数量,  $p(x_i)$  为样本的实测值,  $v(x_i)$  为样本的预测值,  $\bar{P}$  为样本实测值的平均值,  $\bar{v}$  为预测值的平均值,  $\sigma_p^2$  为样本实测值的方差,  $\sigma_v^2$  为样本预测值的方差。

#### 1.4 土壤养分分区评价方法

投影寻踪是将高维数据向低维空间投影, 通过分析低维空间的投影特性来研究高维数据的特征, 找到反映数据结构特征的最优投影, 是处理多指标问题的统计方法<sup>[21]</sup>。其具体思路是将影响问题的多因素指标通过投影寻踪分析得到反映其综合指标特性的最优投影特征值, 然后建立投影特征值与因变量的一一对应关系, 通过分析一维的投影值来对样本做出更加合理的分级和评价<sup>[21]</sup>。

首先, 假定  $\{x_{ij} / i=1, \dots, n; j=1, \dots, m\}$  是样本集,  $x_{ij}$  是第  $i$  个样本的第  $j$  个指标值, 使用每个指标数据的最大值对该指标进行标准化处理, 以计算  $\frac{x_{ij}}{x_{j\max}}$  得

到标准化后的样本数据  $x_{ij}^*$ 。

然后, 使用数量为  $x$  的多维单位向量  $a(a_1, a_2, \dots, a_m)$  表示某一投影方向, 用公式(5)可计算样本  $i$  在该方向上的投影值, 用公式(6)计算投影值的标准差  $S(a)$ , 用公式(7)计算投影值的局部密度  $D(a)$ , 用投影值的标准差  $S(a)$  和局部密度  $D(a)$  的乘积表示为投影指标函数  $Q(a)$ , 即公式(8), 目的是当把  $m$  维数据综合成某一方向上的一维投影值时, 投影值散布特征为局部投影点尽可能密集, 而整体上各个点团之间又尽可能分散<sup>[21]</sup>。

$$z_i = \sum_{j=1}^m a_j x_{ij}^*, i = 1, 2, 3, \dots, n \quad (5)$$

$$S(a) = \sqrt{\frac{\sum_{i=1}^n (Z_i - \bar{Z}_i)^2}{n-1}} \quad (6)$$

$$D(a) = \sum_{i=1}^n \sum_{k=1}^n (R - r_{ik}) f(R - r_{ik}) \quad (7)$$

式中:  $\bar{Z}_i$  为投影方向上投影值的平均值,  $r_{ik} = |z_i - z_k|$ ,  $R$  为局部密度的窗口半径, 可取  $r_{\max} + \frac{m}{2} \leq R \leq 2m$ , 函数  $f$  为单位跃阶函数,  $R - r_{ik} \geq 0$  时, 其值为 1,  $R - r_{ik} < 0$  时, 其值为 0。

$$Q(a) = S(a)D(a) \quad (8)$$

考虑到  $Q(a)$  值与投影方向  $a$  有关, 当  $Q(a)$  取最大值时的  $a$  方向是最能反映数据结构特征的方向, 因而利用遗传算法, 通过求解投影指标函数优化问题识别最佳投影方向, 公式如下

$$\begin{cases} \max Q(a) \\ s.t. \|a\| = 1 \end{cases} \quad (9)$$

最后, 基于最佳投影方向采用公式(5)计算全国土壤养分含量分级标准值的投影值, 建立基于投影值的等级评价和分区方法。

依据全国第二次土壤普查的分级标准(表 1)<sup>[22]</sup>, 将土壤 TN、TP、TK 的预测结果进行叠加分析、归一化处理, 基于 DPS9.5 软件(<http://www.dpsw.cn/>)计

表 1 全国土壤养分含量分级  
Table 1 Grades of soil nutrient contents in China

TN (g/kg)	TP (g/kg)	TK (g/kg)	级别	丰缺
<0.50	<0.50	<3.00	1	极低
0.50 ~ 0.75	0.50 ~ 0.75	3.00 ~ 5.00	2	低
0.75 ~ 1.00	0.75 ~ 1.00	5.00 ~ 10.00	3	中
1.00 ~ 1.50	1.00 ~ 1.50	10.00 ~ 15.00	4	中上
1.50 ~ 2.00	1.50 ~ 2.00	15.00 ~ 20.00	5	高
>2.00	>2.00	>20.00	6	极高

算最优投影方向以及最优投影方向的投影值, 较为客观地对土壤养分分区进行划分。

## 2 结果与分析

### 2.1 表层土壤养分的统计特征

表2为表层样本的TN、TP、TK的描述性统计情况。Kolmogorov-Smirnov检验结果显示, TN、TP经对数转换后均符合正态分布, TK含量符合正态分布。样点表层土壤TN含量范围0.04~13.50 g/kg, 平均值2.56 g/kg; TP含量范围0.49~3.39 g/kg, 平均值1.52 g/kg; TK范围4.67~26.95 g/kg, 平均值17.23 g/kg。

表2 表层土壤养分含量描述性统计  
Table 2 Descriptive statistics of topsoil nutrient contents

土壤养分	最大值	最小值	平均值	标准差	方差	偏度	峰度	变异系数
TN(g/kg)	13.50	0.04	2.56	2.41	5.81	1.83	4.04	0.94
lg(TN)	1.13	-1.39	0.22	0.45	0.19	-0.53	0.40	2.07
TP(g/kg)	3.39	0.49	1.52	0.48	0.23	0.73	0.99	0.32
lg(TP)	0.53	-0.31	0.16	0.14	0.02	-0.44	0.86	0.88
TK(g/kg)	26.95	4.67	17.23	2.84	8.09	-0.02	2.76	0.65

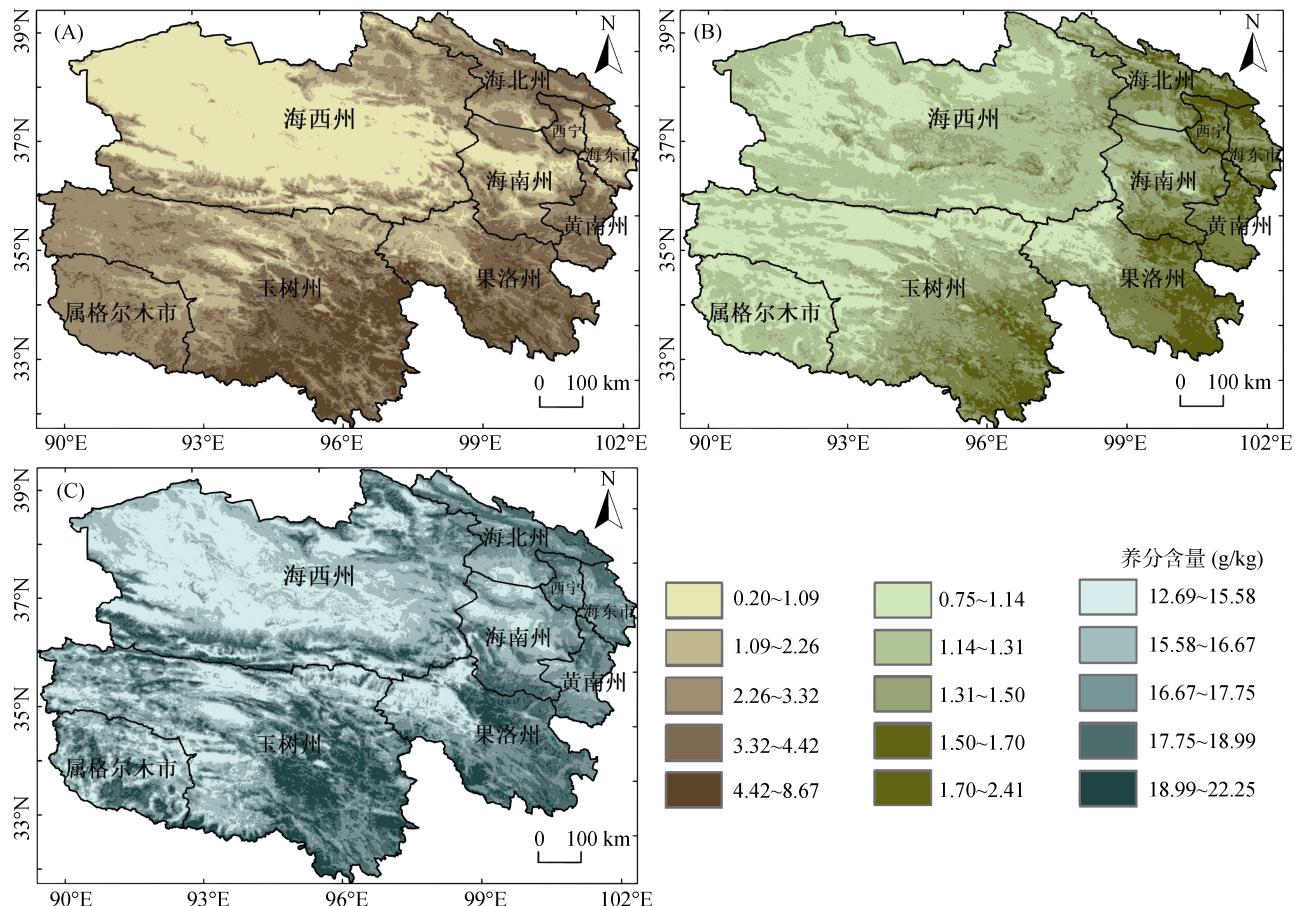


图2 土壤养分空间预测图(A. TN; B. TP; C. TK)

Fig. 2 Spatial prediction maps of soil nutrients

从变异系数来看, 表层土壤养分的变异系数均处于0.1~1, 表明土壤养分均呈中等程度变异, 其排序为TN>TP>TK, 其中TN的变异程度最大, 达到0.94。

### 2.2 土壤养分预测制图与结果

利用随机森林模型预测青海省表层土壤养分空间分布(图2)表明, 土壤TN预测值的范围为0.20~8.67 g/kg, 均值2.48 g/kg, 在空间上整体呈现东南向西北递减的趋势, 其中青海省东南、中部和东北部土壤TN含量较高, 西北部及湟水谷地的含量较低。土壤TP的预测结果范围0.75~2.41 g/kg, 均值1.33 g/kg, 空间分布上的趋势是东高西低, 土壤TP

含量高的区域是在玉树地区的东南部、东北部的祁连山地区和全省东部，而西南部可可西里地区、西北部的茫崖地区和中部昆仑山区含量较低。土壤 TK 预测的结果范围  $12.70 \sim 22.25 \text{ g/kg}$ ，均值  $16.94 \text{ g/kg}$ ，整体呈现分布趋势为东南高西北低，空间分布上较为分散，土壤 TK 含量高值主要分布在青南高原和祁连山山区，在唐古拉山中段地区也呈现高值，低值主要分布柴达木地区、昆仑山脉南部、黄南州的西北部和青海省西南部的格尔木市地区。

交叉验证结果(表 3)表明，土壤 TN、TP、TK 模型拟合度  $R^2$  比较高，分别为  $0.89$ 、 $0.85$ 、 $0.82$ ，即可以解释土壤养分含量的 80% 以上；从一致性系数(CC)来看，三者均在 0.87 以上；预测偏差(ME)三者接近 0 值，说明预测的值有被低估。由此，机器学习方法与成土环境协同变量相结合对青海省表层土壤养分的预测，可以很好地解释土壤养分的空间变异。

表 3 土壤养分预测的交叉验证结果

Table 3 Cross validation results of soil nutrient prediction

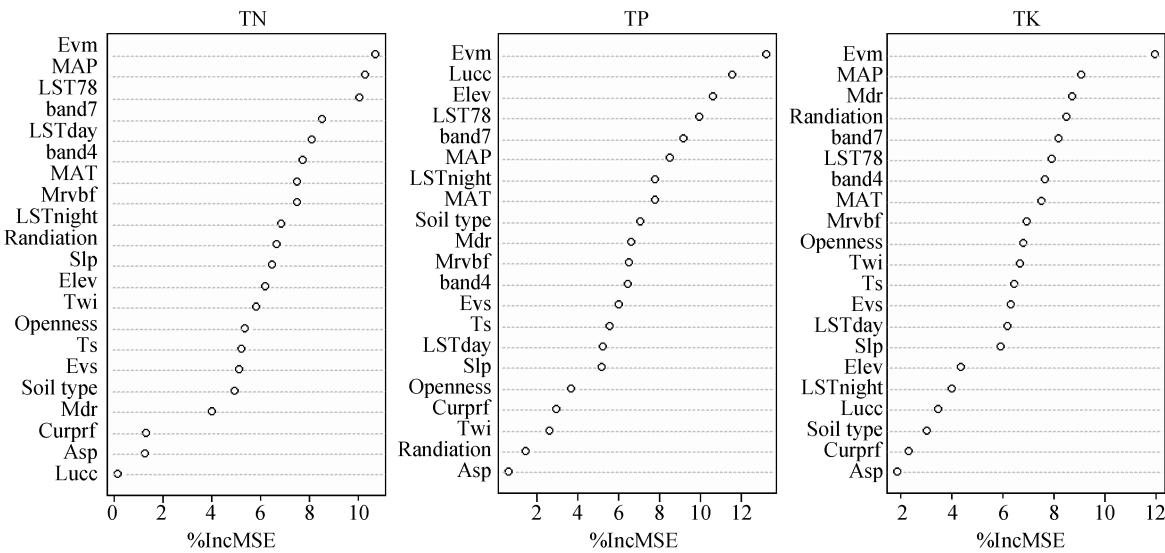
土壤养分	$R^2$	CC	ME	RMSE
Log_TN	0.89	0.93	0.046	0.77
Log_TP	0.85	0.90	0.003	0.18
TK	0.82	0.87	0.035	0.28

### 2.3 土壤养分空间变异的影响因素

图 3 显示了土壤养分空间预测中环境协同变量的相对重要性排序。对于青海省表层土壤 TN 的空

分布格局，植被指数均值(Evm)为主导因子，年均降水量(MAP)次之。通常在空间范围较大研究区域内，气候因素的地带性可以直观反映作物的生长状况，间接影响着土壤养分元素的迁移转化过程。在青海省东北部的祁连山地、湟水谷地及果洛玉树高原区域的降水量较高，土壤水分条件好，植被覆盖度高，且随海拔增加，覆盖率增加，地表的枯枝落叶增加，腐殖质层较厚，土壤微生物活性较强，土壤养分来源更加丰富，积累的 TN 含量普遍较高。在青海省西部地区，气候干旱炎热，降水量稀少且地表温度高，水热条件的不均衡导致植被覆盖度小，影响着土壤中微生物的活动，TN 含量低。其次，西北区域分布着广泛大面积的盐渍土、风沙土、砂质土等，且其土壤保水保肥能力较差，土壤养分分解速度快，植物受高温干旱和盐分制约生存，氮素的来源和累积弱，土壤 TN 较低。

从 TP 来看，植被指数均值(Evm)、地表覆盖(Lucc)和海拔(Elev)对其含量重要性排名前三。气候、地形地貌等环境因子对土壤中磷元素的循环产生重要影响，玉树地区的东南部、东北部的祁连山地区为青海省的植被覆盖度高的区域，土壤中有机质高，可减少磷的吸附位点，增加磷的移动性，其次土壤磷形态和含量受土壤微生物活性的直接影响，土壤中的微生物，能够通过其生理活动促进土壤中含磷化合物中的磷释放，从而改变土壤中磷元素形态和含量。中北部柴达木盆地，存在绿洲所以出现高值；西南部可可西里地区、西北部的茫崖地区和中部昆仑山区，气候干



(Soil type: 土壤类型；Lucc: 地表覆盖类型；Evm: 植被指数均值；Evs: 植被指数标准差；Elev: 海拔；Slp: 坡度；Asp: 坡向；Curprf: 剖面曲率；Radiation: 潜在太阳辐射；Twi: 地形湿度指数；Openness: 地形开放度；Mrvbf: 多尺度山谷指数；MAP: 年均降水量；MAT: 年均温；Mdr: 昼夜温差；Ts: 气温季节性；LSTday: 白天地表温度；LSTnight: 夜间地表温度；LST78: 7、8月地表温度；band4、band7: 地表反射第 4 和第 7 波段)

图 3 土壤养分空间预测中环境协同变量的相对重要性

Fig. 3 Relative importance of environmental covariates in spatial prediction of soil nutrients

旱, 同时受海拔、风力的作用增强, 土壤中砂粒含量高, 温度流失较快, 土壤磷素的转换率低。青南高原河谷地区降水量相对充足, 此外海拔(Elev)对降水(MAP)分布模式产生影响, 进而土壤 pH 空间分布也受之影响, 同时气温受海拔的影响, 在此区域的蒸发量远低于降水量, 使得土壤淋溶过程加速, 促进土壤酸性升高, 在中、酸性的土壤中, 磷素形态转换速率快, 土壤 TP 含量较高; 西部柴达木盆地气候干燥, 蒸发量大, 容易形成盐沼, 导致碳酸盐集聚, 土壤 TP 含量偏低; 在东北部的河湟谷地区域土壤 TP 含量高, 局部变异也大, 该区域是青海省主要农业区, 长时间的人类农耕活动作用, 导致土壤碱性增强, 土壤 TP 含量高。

对于 TK 而言, 植被指数均值(Evm)、年均降水量(MAP)、昼夜温差(Mdr)的影响作用很大。青海省南部的玉树、东南部的果洛地区和东北部的祁连山山区植被覆盖度高, 植被覆盖较好, 土壤受淋溶作用相对较弱, 加之林地土壤中的充足草木枯萎被微生物分解, 有利于土壤速效钾的涵养与保持。中部的昆仑山区和唐古拉山中段地区的昼夜温差(Mdr)较大, 风雪气候常见, 存在严重的土壤侵蚀, 由于土壤全钾主要是由矿物钾组成, 土壤钾素会随土壤侵蚀流失, 土壤中 TK 含量较低; 在西南部的格尔木地区, 海拔较高, 受风力与水流的风化作用较强, 风化物中又富含较多的云母、长石等含钾结晶体, 富钾风化残积物混入土壤后使得含量变高。潜在太阳辐射(Radiation)、年均降水量(MAP)等协变量因子对土壤的湿润和淹水时间持续的长短起到关键的作用, 进而影响钾素固定和

释放过程, 土壤失水变干, 会产生释钾现象, 青海省地形起伏度大, 同时受气候和人类活动影响, 由此全省的 TK 含量局部上有所变异。

## 2.4 土壤养分分区结果

从表 4 可知, 土壤养分分级标准值对应的投影值分别为 0.141、0.234、0.411、0.682 和 0.814, 据此对土壤养分投影值划分等级, 其结果为: 1 级( $<0.141$ )、2 级( $0.141 \sim 0.234$ )、3 级( $0.234 \sim 0.411$ )、4 级( $0.411 \sim 0.682$ )、5 级( $0.682 \sim 0.814$ )、6 级( $>0.814$ )。

表 4 土壤养分的投影特征值及土壤养分分级  
Table 4 Projection eigenvalues and grades of soil nutrients

TN (g/kg)	TP (g/kg)	TK (g/kg)	投影特征值
2.0	2.0	20	0.814
1.5	1.5	15	0.682
1.0	1.0	10	0.411
0.75	0.75	5	0.234
0.5	0.5	3	0.141

图 4 为青海省土壤养分分区图, 表 5 是每个等级和分区所占面积的比重。土壤养分等级划分为极低、低、中、中上、高、极高 6 个等级, 管理分区分为 6 个分区(Q1 ~ Q6)。全省土壤养分等级划分的大致空间格局为: 整体上呈现东南-西北均匀递减趋势, 极高的主要分布在玉树的南部地区、果洛地区和东北部的祁连山, 其分布区域植被覆盖度高; 昆仑山段也存在极高的分布, 面积比例占全省的 4.18%, 这表明该地区的土壤人为活动较少, 主要是自然发育的土壤, 主要受气候因子和地形因子影响作用较大; 养分等级

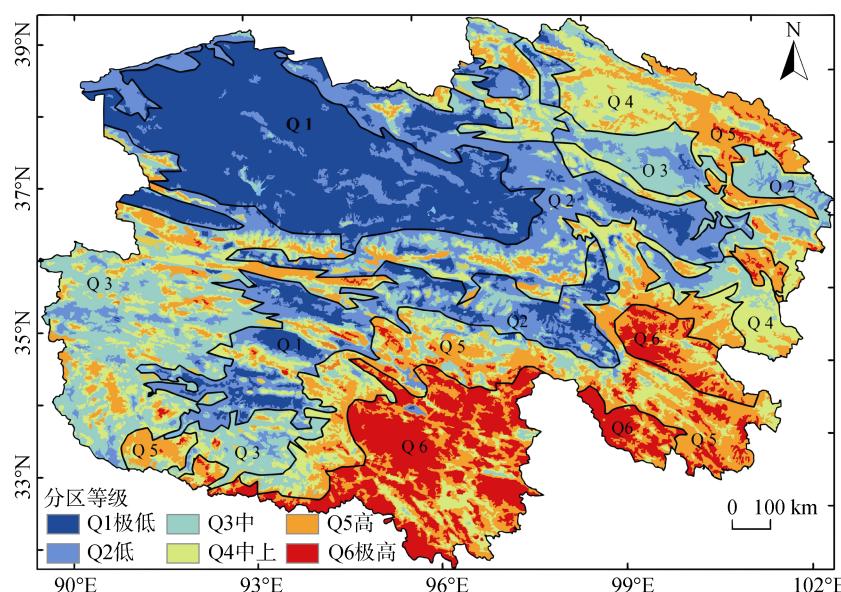


图 4 土壤养分分区图  
Fig. 4 Spatial distribution of soil nutrient zoning

表 5 土壤养分分区及面积比例  
Table 5 Classification and area ratio of soil nutrient zoning

等级	分区	面积比例(%)	面积累计比例(%)
极低	Q1	3.04	3.04
低	Q2	15.73	18.77
中	Q3	24.21	42.97
中上	Q4	39.94	82.91
高	Q5	12.91	95.82
极高	Q6	4.18	100

高的区域面积比例占全省的面积的 13%，其空间分布不均，主要分布在海北州、格木市飞地、玉树州的西南部和黄南州的西北部等地区，这些地区有大面积的林区、草场和耕地，同时受人类活动作用影响比较大；养分等级中至中上的区域主要分布西北柴达木盆地周边、东部湟水谷地、青海湖周围以及玉树州的西南区域，占面积比例 65%，这与该区域内的地形坡度变化及人工施肥量有关；土壤养分低等级的区域主要分布在柴达木盆地、西南的可可西里区域、扎陵湖和鄂陵湖周边和海南州的中北部，占面积比 16%。土壤养分等级从高到低的面积比例为：4：13：40：24：16：3，养分等级低和极低的总面积占全省的 19%，表明青海省土壤养分属于中等级及偏上，主要分布在全省的中南部和东北部。因此，需要因地制宜，科学地划分土壤利用区域，调整放牧和草场施肥结构，加强对土壤养分含量高的区域的管理与调整，恢复土壤养分含量低的区域，增加生态保护措施。

### 3 讨论

通常在研究区面积相对较小、景观相对单一、样点较多且分布相对均衡的条件下，只利用样点而不用环境变量进行简单插值预测，预测效果比较好。如张铁婵等<sup>[23]</sup>和石小华等<sup>[24]</sup>采用空间插值法分别对榆林市榆阳区和周至县北部猕猴桃适生区的土壤养分元素进行空间预测，在样点布设大的地区，插值精度很好，反之插值精度不是很理想。插值结果受样点布设密度的影响较大，而预测精度受不同的插值方法的影响较小。对于大面积高寒地区，样点稀少且分布不均，显然单纯基于样点的插值不适合，需要发挥环境因素辅助变量的作用。本文研究发现，通过把丰富的成土环境因素变量与复杂的机器学习方法结合，能够在大面积高寒山区实现较为准确的土壤养分预测制图，捕捉土壤空间变异。植被能有效地指示土壤养分在大面积区域的空间变异。植被指数均值(Evm)对青海省表层土壤养分的空间变异具有主导作用。Pouladi 等<sup>[25]</sup>

在丹麦日德兰半岛，利用 NDVI、SAVI、DVI、RVI 和地形相关变量预测表层 SOC 含量，发现这些遥感变量解释了 89% 的 SOC 空间变化。Wang 等<sup>[26]</sup>在辽宁省表层土壤含量制图中发现结合仅利用遥感数据的随机森林模型预测效果明显好于仅使用地形和气候变量的随机森林模型。在后续工作中，在景观复杂且草地大面积发育的青海省，遥感数据作为主要环境变量的土壤制图值得探究。地表温度是地表对于太阳辐射的响应或反馈<sup>[27]</sup>，可以作为环境变量有效地对土壤环境的空间差异进行揭示，对于本研究的高寒山区，地形地貌复杂多变，降水少，植被以牧草为主，这种反馈将主要取决于太阳辐射、地形、降水、植被和土壤的综合作用。这也说明该区域地表温度(LST)和年均降水量(MAP)是影响表层土壤养分空间变异的主要环境因子。地形因子的贡献率在青海省表层土壤养分的空间变异中相对较低，主要的原因可能是在景观和小流域尺度上，母质、气候和植被相对一致，地形通过径流、土壤侵蚀间接影响土壤属性的空间分布<sup>[28]</sup>，地形是影响小面积区域土壤属性的主要因素，如孙孝林等<sup>[29]</sup>和杨琳等<sup>[30]</sup>在小区域土壤养分预测中只考虑地形属性变量就可以进行有效的土壤制图。在大面积高寒山区，土壤属性受到多种环境因素的协同影响，如母质、气候、地形和植被等。各成土因素变量对土壤养分都有贡献，有较强的成土因素综合作用。土壤属性与环境变量之间存在非常复杂的关系，这与小面积区域土壤制图不同。在后续的工作中，需要发掘更为有效的成土环境因素协同变量，来提高高寒山区数字土壤制图的准确性。

另一方面，预测模型选取应考虑具体的土壤样点数据和土壤景观情况。大面积高寒山区地貌景观复杂多样，土壤空间变异程度高，样点比较稀少且空间不均衡，导致这类地区数字土壤制图结果难度往往比较大，简单景观是指某个成土因素主导土壤空间分布，土壤-环境关系近于线性。复杂景观是指多个成土环境因素，土壤-环境关系是非线性的、非平稳的<sup>[31]</sup>。对于前者，线性回归可以较好地解释协变量(如地形属性)与土壤属性之间的关系。对于后者，可以使用机器学习算法来模拟土壤变化。青海省作为大面积高寒地区，土壤-环境关系复杂，因此预测模型的选取至关重要，张欢和高小红<sup>[9]</sup>比较了地统计中协同克里格、地理加权回归克里格以及结合土地利用类型的均值、中值修正协同克里格等方法对湟水流域土壤有机质属性的预测精度，忽略了复杂的土壤环境关系，在空间分布上能体现其变化规律，但预测值结果存在预高估和低估现象甚至负值。地统计学模型在土壤样本

均匀分布且满足空间自相关的条件下,模型的预测精度较好,这在地形复杂、土壤调查样本较为稀疏的青海省并不适用。

土壤的利用和管理需要对土壤养分进行生产管理分区。通常土壤属性指标和分区方法的选取不同,会忽略土壤综合养分的空间差异、土壤质量因子的变异性以及土壤环境的特殊性,进而管理分区结果也因此产生不同,大部分学者采用了主成分分析、地统计、神经网络和模糊聚类模型等方法结合土壤属性的单因子来进行分区管理,黄婷等<sup>[32]</sup>通过主成分分析来计算土壤综合质量指数进行了线性回归分析,忽略了土壤-景观关系的复杂多变性;朱昌达等<sup>[33]</sup>运用地统计方法分析土壤属性空间变异的同时采用模糊聚类来进行分区,未能考虑土壤属性的综合作用,计算分区指标权重和综合值。由于土壤综合养分具有多维性和综合性,而综合定量研究在较大程度上可以避免研究者主观因素的影响,已然成为土壤综合养分分区与管理的一个趋势<sup>[34-36]</sup>。对于大面积高寒山区,土壤养分与环境变量之间的关系较为复杂,需要建立土壤养分与评价值之间较为准确的关系,重视土壤养分的综合指标的选取信息,来提高管理分区结果的准确性,本研究基于数字土壤制图得到的土壤养分分布图,采用投影寻踪法对土壤养分进行了管理分区。尽管投影寻踪法减少了人为因素干扰,提高了养分管理分区的客观和准确性,但是不同投影方向优选算法可能会影响评价结果,有待进一步探讨。

## 4 结论

1)在土壤调查样本较为稀疏的大面积高寒山区,通过把丰富的成土环境因素变量与复杂的机器学习方法相结合,模型解释度达83%以上,能够实现较为准确的土壤养分预测制图,捕捉土壤空间变异。

2)青海省表层土壤养分在整体上呈现东南-西北均匀递减趋势。植被能有效地指示青海省表层土壤养分全氮(TN)、全磷(TP)、全钾(TK)空间分布模式,其中年均降水量、地表温度是影响青海省表层土壤TN空间模式的重要因素;地表植被、海拔和地表温度等环境因子对表层土壤TP的空间变异占主导作用;年均降水量、昼夜温差对表层土壤TK的空间变异占主导作用。

3)土壤养分预测结果高等级出现在南部的玉树、果洛、黄南和东部的湟水谷地地区,低等级主要分布在柴达木盆地、可可西里、海南州中南部,全省养分等级呈中上等级以上,占全省面积的81%。由此建立

的土壤养分管理分区,可为全省土壤养分的监测与农牧业区划的管理提供参考依据。

## 参考文献:

- [1] 贾三满,杜丽娜,冯辉,等.北京大清河流域生态涵养区生态地质环境质量综合评价[J].北京地质,2020,15(2): 125-136.
- [2] 李莹莹,赵正勇,杨旗.数字土壤制图在土壤养分方面的研究综述[J].江西农业学报,2021,33(7): 61-67.
- [3] Dai F Q, Zhou Q G, Lv Z Q, et al. Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau[J]. Ecological Indicators, 2014, 45: 184-194.
- [4] 李德成,赵霞,张甘霖.中国土系志·青海卷[M].北京:科学出版社,2018.
- [5] 杨煜岑,杨联安,王晶,等.基于多元线性回归模型的土壤养分空间预测——以陕西省蓝田县农耕区为例[J].土壤通报,2017,48(5): 1102-1113.
- [6] Dong W, Wu T J, Luo J C, et al. Land parcel-based digital soil mapping of soil nutrient properties in an alluvial-diluvia plain agricultural area in China[J]. Geoderma, 2019, 340: 234-248.
- [7] Liu F, Rossiter D G, Song X D, et al. An approach for broad-scale predictive soil properties mapping in low-relief areas based on responses to solar radiation[J]. Soil Science Society of America Journal, 2020, 84(1): 144-162.
- [8] 代子俊,赵霞,李德成,等.近30年湟水流域土壤全氮时空变异及影响因素[J].土壤学报,2018,55(2): 338-350.
- [9] 张欢,高小红.复杂地形区土壤有机质空间变异性分析及制图[J].水土保持研究,2020,27(5): 93-100.
- [10] 周伟,谢利娟,杨晗,等.基于高光谱的三江源区土壤有机质含量反演[J].土壤通报,2021,52(3): 564-574.
- [11] 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.
- [12] 陈荣,韩浩武,傅佩红,等.基于多时相遥感影像和随机森林算法的土壤制图[J].土壤,2021,53(5): 1087-1094.
- [13] Nussbaum M, Spiess K, Baltensweiler A, et al. Evaluation of digital soil mapping approaches with large sets of environmental covariates[J]. Soil, 2018, 4(1): 1-22.
- [14] 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.
- [15] 青海省统计局.青海统计年鉴·2002(总第18期)[M].北京:中国统计出版社,2002.
- [16] 张甘霖,龚子同.土壤调查实验室分析方法[M].北京:科学出版社,2012.
- [17] Fick S E, Hijmans R J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas[J]. International Journal of Climatology, 2017, 37(12): 4302-4315.
- [18] Leo Breiman. Random Forests[J]. Machine Learning, 2001, 45(1): 5-32.

- [19] R Development Core Team R A Language and Environment for Statistical Computing[M]. Vienna: R Foundation for Statistical Computing, 2014.
- [20] Picard R R, Cook R D. Cross-validation of regression models[J]. *Journal of the American Statistical Association*, 1984, 79(387): 575–583.
- [21] 谢贤健, 韩光中. 基于普通克里格和投影寻踪模型的城市土壤重金属污染评价[J]. *生态环境学报*, 2017, 26(9): 1584–1590.
- [22] 全国土壤普查办公室. 中国土壤[M]. 北京: 中国农业出版社, 1998.
- [23] 张铁婵, 常庆瑞, 刘京. 土壤养分元素空间分布不同插值方法研究——以榆林市榆阳区为例[J]. *干旱地区农业研究*, 2010, 28(2): 177–182.
- [24] 石小华, 杨联安, 张蕾. 土壤速效钾养分含量空间插值方法比较研究[J]. *水土保持学报*, 2006, 20(2): 68–72.
- [25] Pouladi N, Møller A B, Tabatabai S, et al. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging[J]. *Geoderma*, 2019, 342: 85–92.
- [26] Wang S, Jin X X, Adhikari K, et al. Mapping total soil nitrogen from a site in northeastern China[J]. *CATENA*, 2018, 166: 134–146.
- [27] 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.
- [28] 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, 202, 201.
- [29] 孙孝林, 赵玉国, 赵量, 等. 应用土壤-景观定量模型预测土壤属性空间分布及制图[J]. *土壤*, 2008, 40(5): 837–842.
- [30] 杨琳, 朱阿兴, 秦承志, 等. 运用模糊隶属度进行土壤属性制图的研究——以黑龙江鹤山农场研究区为例[J]. *土壤学报*, 2009, 46(1): 9–15.
- [31] Brungard C W, Boettinger J L, Duniway M C, et al. Machine learning for predicting soil classes in three semi-arid landscapes[J]. *Geoderma*, 2015, 239/240: 68–83.
- [32] 黄婷, 岳西杰, 葛玺祖, 等. 基于主成分分析的黄土沟壑区土壤肥力质量评价——以长武县耕地土壤为例[J]. *干旱地区农业研究*, 2010, 28(3): 141–147, 187.
- [33] 朱昌达, 高明秀, 王文倩, 等. 基于 GIS 的滨海盐渍化农田土壤空间变异及其分区管理[J]. *生态学报*, 2020, 40(19): 6982–6990.
- [34] 张彬, 杨联安, 杨粉莉, 等. 基于投影寻踪的土壤养分综合评价及影响因素研究[J]. *土壤*, 2020, 52(6): 1239–1247.
- [35] 梁斌, 齐实. 北京山区土壤养分空间变化特征研究[J]. *土壤*, 2018, 50(4): 769–777.
- [36] Li H, Leng W J, Zhou Y B, et al. Evaluation models for soil nutrient based on support vector machine and artificial neural networks[J]. *The Scientific World Journal*, 2014, 2014: 478569.