

杨培杰,韩保栋,张玉燕,等. Sentinel-2 影像结合空间关联随机森林模型反演裸土期耕地土壤全氮含量[J]. 江苏农业科学,2023,51(11):185-191.
doi:10.15889/j.issn.1002-1302.2023.11.027

Sentinel-2 影像结合空间关联随机森林模型反演 裸土期耕地土壤全氮含量

杨培杰^{1,2}, 韩保栋^{1,2}, 张玉燕^{1,2}, 赵 菲^{1,2}, 王庆兵^{1,2}, 翟晓燕^{3,4}

(1. 山东省国土空间生态修复中心, 山东济南 250000; 2. 自然资源部黄河三角洲土地利用安全野外科学观测研究站, 山东滨州 251900;
3. 枣庄学院旅游与资源环境学院, 山东枣庄 277160; 4. 黎刹大学, 菲律宾曼达卢永市 1552)

摘要:为快速准确地获取区域内土壤全氮的含量信息和空间分布特征。选取山东省济南市章丘区刁镇为研究区, 系统采集 64 个土壤样品并获取同期 Sentinel-2 (哨兵 2 号) 影像数据, 进一步利用皮尔逊相关分析法选择土壤全氮的敏感光谱参量作为自变量, 测试得到的土壤全氮含量为因变量, 分别建立基于随机森林和空间关联随机森林算法的反演模型, 完成区域尺度的土壤全氮含量的遥感反演和数字制图。结果表明: 哨兵 2 号影像的多光谱波段与土壤全氮含量相关性较低, 通过波段间比值变换能够显著增强土壤全氮含量对光谱信号的响应能力, 光谱指数 b_6/b_{11} 、 b_{8a}/b_{12} 、 b_8/b_9 、 b_{8a}/b_9 和 $(b_9 - b_{11})/(b_9 + b_{11})$ 对土壤的全氮含量信息最为敏感; 空间关联随机森林模型的反演精度指标 R^2 和 $RMSE$ 分别为 0.90 和 0.11, 相对比随机森林模型精度分别提升 11.11% 和 26.67%, 使反演模型的结构和计算效率均得到了优化; 土壤全氮含量在田块尺度上的空间变异较大, 与土地利用状况关系密切, 村居建筑物周边土壤全氮含量处于低水平 (<0.80 g/kg), 远离建筑物的大片耕地区域土壤全氮含量则较高 (>1.20 g/kg)。哨兵 2 号影像与空间关联随机森林算法相结合的遥感反演技术可为区域土壤环境信息的监测与制图分析提供有效的方法支持。

关键词: 土壤; 全氮; Sentinel-2 影像; 空间关联; 随机森林模型; 裸土期; 耕地

中图分类号: X87; S127 **文献标志码:** A **文章编号:** 1002-1302(2023)11-0185-07

土壤全氮含量是衡量土壤肥力的重要指标, 也是维持植物正常生命活动的基础营养元素, 实现快速准确的土壤全氮含量和空间分布监测是建设高标准农田和保证粮食安全的重要途径。土壤光谱反射特性与土壤物质构成密切相关, 土壤全氮含量越高, 光谱反射率便也随之降低^[1-2]。因此运用光谱反射率与土壤全氮间的内在联系, 可以进行土壤全氮含量的快速估算, 实现区域尺度的土壤全氮含量遥感反演和空间分布监测。

传统土壤全氮含量信息的调查依赖于大规模土壤样点布设和实验室分析, 不仅耗时耗力效率低, 且在大尺度上的应用和监测受到限制^[3]。遥感

技术具备覆盖面积大、时效性强和高时空分辨率等特点, 已经广泛应用于地表土壤参数的反演制图^[4]。Krishnan 等在 1980 年首次采用土壤的室内高光谱及多元回归技术, 完成了对土壤关键组分的含量和性质反演估算, 并且分析得到相应组分在不同波谱区间上的特定响应敏感性^[5]。张平等也尝试构建反射光谱模型来反演土壤全氮含量, 确定了土壤全氮在一定含量范围内与反射光谱值存在线性关系^[6]。卢艳丽等基于实测土壤光谱反射数据对土壤全氮含量进行建模反演, 取得较好的估算效果并且证实了光谱技术用于土壤全氮定量研究的能力^[7-9]。马驰进一步厘清了卫星数据的波段反射率值与地面样本全氮含量之间的统计回归联系, 完成了土壤全氮的反演制图, 为土壤全氮在区域尺度上的空间分布研究提供了理论支持^[10-11]。

在对土壤全氮的遥感定量分析中, 因受到复杂的地表反射率辐射传输过程影响, 卫星光谱波段的反射率值与地表土壤理化参数之间未必是确定的线性关系, 所以基于非线性回归算法的机器学习遥感反演模型在精度和稳定性具有明显的优势。例如, 雷浩川等利用影像的波段光谱反射率为输入自

收稿日期: 2022-08-11

基金项目: 山东省教育厅公派出国留学基金 (编号: 2019027); 环境演变与自然灾害教育部重点实验室开放课题项目 (编号: 2022-KF-14)。

作者简介: 杨培杰 (1988—), 男, 山东费县人, 高级工程师, 主要研究方向为生态修复和地质环境监测。E-mail: yangpj1232022@163.com。

通信作者: 翟晓燕, 女, 博士, 讲师, 主要研究方向为生态旅游规划。E-mail: zxy18263230964@126.com。

变量,地面土壤实测全氮含量为响应变量,构建基于人工神经网络、随机森林和支持向量机等机器学习算法的遥感反演模型,得到区域尺度上精度相对较高的土壤全氮含量信息分布图^[12-14]。通常来说,机器学习模型可以通过加入隐藏层结构、决策树和自由参数等方式,确定光谱特征与土壤属性信息之间的映射关系,故而能够提供相对较高的预测精度。但是,机器学习模型在对真实的地表环境过程进行模拟时,易受到水汽、成土母质、数据采集过程等人类活动的作用下的产生的特异性地理标签数据的影响,模型常会出现过拟合和局部极值问题,限制了反演模型精度和性能的提升。针对此问题,Liu 等通过随机抽取、光谱特征、土壤类型和含量欧式距离划分等方式来优化建模样本集的选择,减弱或避免特异性样本数据的干扰,在保证验证集精度的同时最大程度提升训练集准确性^[15-16]。然而上述对特异样本数据的识别均是基于土壤属性的含量特征进行的,地理样本数据除具有描述性的含量特征值以外,还具有显著的空间位置特征信息^[17-18]。地理学第一定律指出,地物间的空间距离越近,所表现出的空间相关性越显著,而特异性数据则表现出了显著的非空间关联性^[19]。因此在随机森林算法的数据输入侧引入空间关联函数,即在决策树生成之前先对土壤样本数据间的空间关联度进行评估,其后根据样本关联度权值来选取最佳的训练样本数据集,以此减少数据计算量和提高模型的拟合精度,使其能够准确完成区域尺度的土壤全氮含量信息反演与分布制图。

选取山东省刁镇为研究区,系统采集 64 处土壤样点并获取同期 Sentinel-2 (哨兵 2 号) 多光谱影像数据,通过波段间差值、比值和归一化值变换来探寻土壤全氮的光谱响应信号,进一步运用相关分析法优选光谱参量作为输入自变量,测得土壤全氮含量为因变量,分别构建基于随机森林算法和空间关联随机森林算法的遥感反演模型,最后推广至哨兵 2 号卫星实现区域尺度的土壤全氮含量的反演制图,以期土壤资源环境管理提供理论依据和技术支持。

1 材料与方法

1.1 研究区及土壤样品采集

研究区位于山东省济南市章丘区刁镇,地理位置 $36^{\circ}51'58.2''\text{N} \sim 36^{\circ}59'1.0''\text{N}$, $117^{\circ}25'16.3''\text{E} \sim$

$117^{\circ}35'1.5''\text{E}$, 占地面积约为 102.3 km^2 (图 1)。研究区气候类型为典型的暖温带大陆性季风气候,四季分明,雨热同季,多年平均气温和降水量分别 12.8°C 和 600.8 mm ^[20]。研究区地形以平原为主,土壤类型主要为褐土,肥力适中,保肥持水能力强。土地利用除村居建筑物外,其他均为耕地,主要种植作物为玉米、小麦和蔬菜。

在综合考虑土地利用、道路可达性和代表性原则,在 ArcGIS 10.2 矢量底图上完成 64 处样点布设。采样过程中,土壤样品的采集依据多点混合采样法,在 5 m 半径内选取 4 处表层土壤样点 (0 ~ 20 cm),简单移除土块杂草等杂质后均匀混合至约 1 kg,装至聚乙烯密封袋后送往实验室待测,另留存 1 袋作为备份对照。同时,使用手持 GPS 记录样点的真实坐标 (图 1),并用相机记录采样周边环境。土壤样品的采集在 2019 年 3 月上旬完成。在实验室内,土壤样品经自然风干、研磨过筛等处理后,依据凯氏定氮法^[2],完成样品土壤全氮含量的实验室测定。

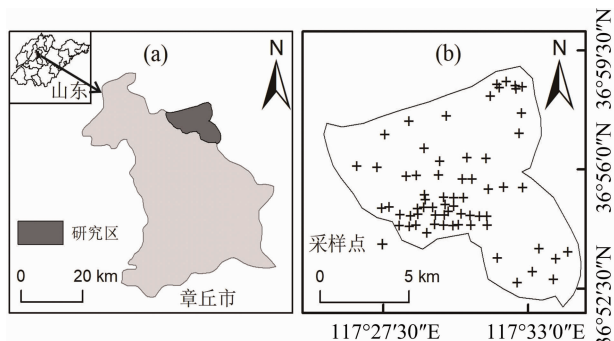


图1 研究区及采样点概况

1.2 影像获取与预处理

Sentinel-2 A/B (哨兵 2 号) 是欧洲航天局主导的哥白尼计划中发射的第 2 组卫星,其搭载的多光谱成像仪 (MSI) 是美国陆地卫星 Landsat 和中国高分卫星计划的拓展和延续,可更有效地对全球进行环境监测和灾害预警。哨兵 2 号的 MSI 传感器具有 4 个 10 m 空间分辨率波段 (b_2 、 b_3 、 b_4 、 b_8)、6 个 20 m 空间分辨率波段 (b_5 、 b_6 、 b_7 、 b_{8a} 、 b_{11} 、 b_{12}) 和 3 个 60 m 空间分辨率波段 (b_1 、 b_9 、 b_{10}),额外增加的红边波段对地表土壤植被变化信息的响应更加敏感^[21]。

哨兵 2 号影像数据在美国地质调查局网站 (<https://glovis.usgs.gov/>) 免费对公众提供 L1C 级产品的下载,研究选取在 2019 年 3 月 23 日的少云晴朗天气状态下一景哨兵 2 号 MSI 影像数据。此时土地覆盖以裸土为主,便于进行土壤全氮的遥感

估算;部分存在植被的区域,依据裸土的判别依据[归一化植被指数(NDVI) < 0.2]^[22],去除影像中非裸土区域。下载到影像的天顶反射率须要进行辐射定标和大气校正等处理才可以转换到地面真实反射率 L2A 级产品。处理过程均在哨兵数据专用处理软件(sentinel application platform, SNAP) 和大气校正 Sen2Cor 插件中完成。处理过程中,卷云波段(band 10)主要参与大气特征的检测,并不能反映土壤全氮的光谱特性,因此在接下来的处理计算中将其排除;此外,为获取更多可用的波段信息,裁剪后的波段均进行了 10 m 分辨率的重采样,同时进行地理校正操作,确保地面样点与影像像元间的偏移误差不超过 0.5 像元^[23]。

1.3 光谱指数构建及优选

全氮的光谱吸收特征在土壤各类组分(例如有机质、铁和盐分)的合频和倍频作用下,在可见光和近红外光谱范围通常是表现出非特异性的,产生的光谱特征具有覆盖范围广和信号微弱的特点^[24-25]。因此,本研究在影像原有的 12 个波段的基础上,分别构建光谱差值指数(DI)、光谱比值指数(RI)和归一化光谱指数(ND)用于探寻土壤全氮含量相关的光谱指示信息,计算公式如下:

$$DI = b_i - b_j; \quad (1)$$

$$RI = b_i / b_j; \quad (2)$$

$$ND = (b_i - b_j) / (b_i + b_j)。 \quad (3)$$

式中: b_i 和 b_j 分别表示哨兵数据的第 i 波段和第 j 波段的光谱反射率。光谱指数的计算是基于 ENVI 5.3 软件的 IDL 平台编程完成。

由公式(1)~公式(3)计算得到的 197 个光谱指数(DI 65 个;RI 67 个;ND 65 个)和 12 个光谱波段均能一定程度地反映土壤全氮的含量状态,但仍有必要进一步选择能够显著提升估算模型精度且具有与全氮显著关联的最佳光谱参量。皮尔逊相关系数分析法(Pearson's correlation coefficient, PCC)广泛用于定量化测度变量间的相关强度^[2,15],计算公式如下:

$$r = \frac{\sum (X - \bar{X}) \times (Y - \bar{Y})}{\sqrt{(\sum (X - \bar{X})^2) \times (\sum (Y - \bar{Y})^2)}}。 \quad (4)$$

式中: X 与 Y 表示变量集合; \bar{X} 与 \bar{Y} 分别为 X 与 Y 变量集合的均值。

1.4 空间关联随机森林模型

空间关联随机森林模型的构建是以随机森林算法^[26]为基础,通过引入空间权重函数^[27-28]对输

入的样本数据进行空间关联度评估[公式(5)、公式(6)],并赋以权重,依据关联度的强弱形成最优样本数据组合,以此为样本输入完成模型的训练和计算的一种遥感反演技术。具体实现流程见图 2。

$$F_i(\phi) = \beta_i(\mu_i, \nu_i); \quad (5)$$

$$c = \sqrt{\sum_{i=1}^n [L - F_0(\phi) - \sum_{j=1}^n F_i(\phi) X_j - \delta]^2}。 \quad (6)$$

式中: $F_i(\phi)$ 表示位置 i 处的地理样本的空间权重高斯函数; (μ_i, ν_i) 为位置 i 处的地理样本的经纬度坐标值; β_i 表示地理样本在位置 i 处存在的空间函数; $F_0(\phi)$ 为局部区域内的地理样本之间的空间权重函数; n 表示局部区域内的地理样本数目; L 和 δ 分别代表位置 i 处的土壤样本的实测土壤全氮含量和随机误差项; c 表示样本数据间的空间关联度,数值越大,样本数据间空间关联度越弱,表现出的空间特异性也就越强。

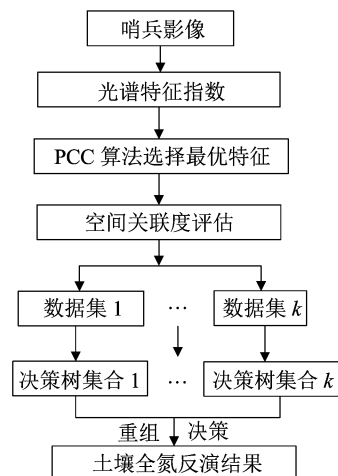


图2 基于空间关联随机森林算法的土壤全氮遥感反演流程

决策树的数量(n tree)、分裂点变量的数量(m try)和地理信息协同变量(X_j)是空间关联随机森林模型构建过程中的 3 个关键参数。构建决策树类模型的核心思想是首先生成若干棵决策树,在每棵决策树上的各节点处,从 M 个特征中进行数据剪枝选取 m 个特征集用作预测变量,在特征的分枝过程中,根据袋外样本误差选取出最优的 m 个特征集作为预测变量,最后汇总投票得出结果^[29]。此外,为增强空间权重函数对样本数据的属性和空间信息的评估能力, X_j 设置为典型且易获取的 4 种环境要素辅助信息,分别为高程、归一化植被指数、土壤湿度指数和地表温度。高程数据免费获取自地理空间数据云网站(<http://www.gscloud.cn/>),归一化植被指数、土壤湿度指数和地表温度数据的获取与采样

时间保持一致,具体计算方法参见文献[30–31]。空间关联随机森林模型在 Python 3.8 平台中经过多次测试拟合,选取误差最小的模型,本研究中最优模型的 ntree 和 mtry 参数分别设定为 1 000 和 3。

1.5 模型评价

选用留一交叉验证法用于模型的精度评估,即每一个样本点都会参与模型建立和验证。确定系数(R^2)和均方根误差($RMSE$)是模型性能的量化评价的重要指标, R^2 表示模型中自变量所能解释因变量信息的比例,数值越大,所能解释模型信息量越多,模型效果越好。 $RMSE$ 表示模型所得的预测值与实测值之间差值, $RMSE$ 与实测含量值具有相同单位,故直接可与实测值估算值进行对比评价模型的好坏。总的来说, R^2 越大、 $RMSE$ 越小,模型精度和稳定度越高^[32]。

2 结果与讨论

2.1 土壤全氮描述性统计分析

如图 3 所示,研究区土壤全氮含量处在 0.37 ~ 1.86 g/kg 之间,均值与中值基本保持一致,表明区域内土壤全氮的含量状态稳定在 1.20 g/kg 左右。变异系数反映了土壤样点数据间的离散状态^[33]。土壤全氮含量的变异系数为 0.30,处于中度变异水平^[34],表明农业生产等外部活动对土壤全氮的分布状态产生了一定影响。全氮含量的偏度为 -0.53,呈现出负偏状态,与外部扰动相关,进一步印证了变异系数的判断。此外,模型的训练样本集表现出的一定变异性,也将有助于加速反演模型的收敛,提高模型的稳定性^[35]。

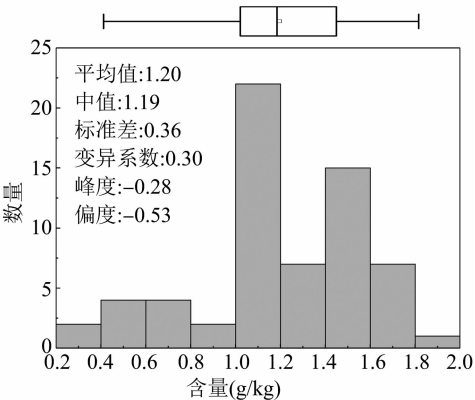


图3 土壤全氮含量描述性统计特征

2.2 最佳光谱响应特征

哨兵 2 号的 12 个多光谱波段与土壤全氮含量的相关系数均小于 0.1 (未通过 0.01 水平上的显著

性检验),无法有效地反映土壤全氮的光谱特征信息。与土壤全氮含量的相关系数排名的前 10 的光谱指数见表 1,其中 b_6/b_{11} 和 b_{8a}/b_{12} 的相关系数分别达到 0.76 和 0.75,并且通过 0.01 水平上的显著性检验。此外,获得到的光谱指数虽然能够有效地反映土壤全氮特征信息,但是将其全部输入反演模型用作预测自变量,会造成光谱信息冗余和训练复杂度上升等问题。因而此基础上仍须要筛选出能够显著提高模型运行精度和效率的最优光谱指数,故相关系数排名前 5 的光谱指数 [b_6/b_{11} 、 b_{8a}/b_{12} 、 b_8/b_9 、 b_{8a}/b_9 和 $(b_9 - b_{11})/(b_9 + b_{11})$] 被选作为最优光谱参量,参与遥感反演模型的构建。

表 1 与土壤全氮含量的相关系数排名前 10 的光谱指数

排序	光谱指数	相关系数
1	b_6/b_{11}	0.76 **
2	b_{8a}/b_{12}	0.75 **
3	b_8/b_9	0.15 **
4	b_{8a}/b_9	0.13 *
5	$(b_9 - b_{11})/(b_9 + b_{11})$	0.18 *
6	b_2/b_{11}	0.10
7	$b_8 - b_9$	0.05
8	b_8/b_{11}	0.03
9	$(b_{11} - b_{12})/(b_{11} + b_{12})$	0.02
10	$b_1 - b_{8a}$	0.02

注: **、* 分别表示在 0.01、0.05 水平上显著相关。

2.3 反演模型建立

优选得到的 5 个光谱指数作为输入自变量,土壤全氮实测值为因变量,分别建立以随机森林和空间关联随机森林算法为基础的土壤全氮遥感反演模型。由图 4 可以看出,随机森林模型(图 4-a)估算出的土壤全氮含量较为分散,土壤全氮含量估算值与实测值差距较大,估算精度指标 R^2 和 $RMSE$ 分别为 0.81、0.15;对比来看,空间关联随机森林模型(图 4-b)的精度指标 R^2 、 $RMSE$ 分别为 0.90、0.11,反演得到的土壤全氮估算值与实测值更为接近,反演精度较高,相对随机森林反演模型 R^2 上升了 11.11%, $RMSE$ 下降了 26.67%,能够实现土壤全氮含量的遥感反演。

空间关联随机森林模型在构建中充分考虑了区域地理环境中土壤样本的特异性和空间关联性,有助于确定随机森林的最佳决策树结构和参数。土壤理化组分的含量信息与影像波段的光谱反射率存在着确定的对应关系,但因受到水汽、植被和地面粗糙度等因素的影响使得这样对应关系变得

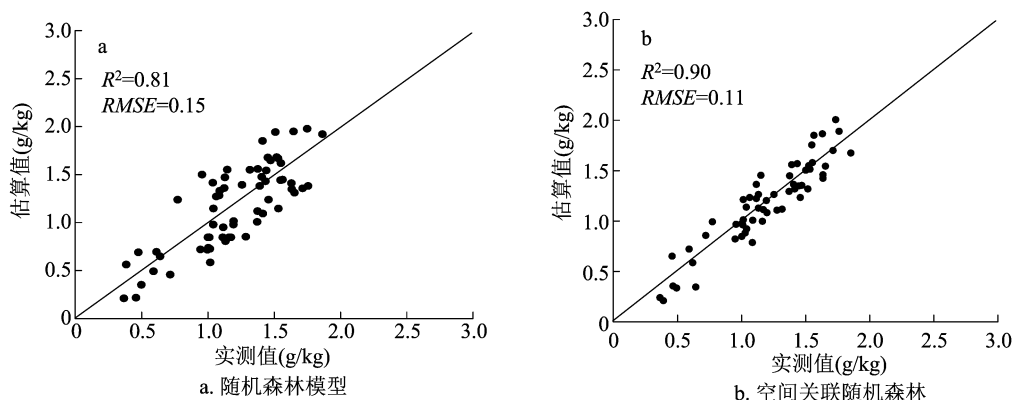
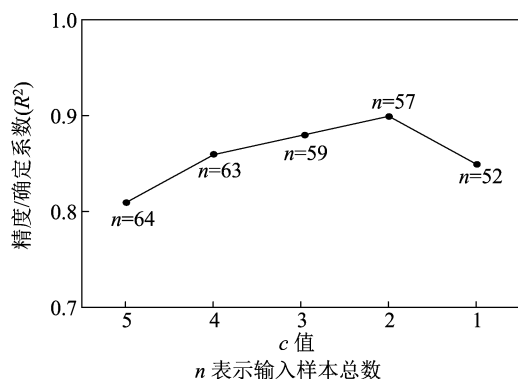


图4 不同模型反演土壤全氮含量估算值与实测值的对比散点

异常复杂^[36]。Lobell 等指出随着土壤中水分含量的上升,光在水中的折射次数将会增加,使从土壤到空气中的反射光减少,呈现出的土壤光谱反射率会近乎平行地下降,改变了原有土壤的光学性质;另外,粗糙度和作物残积物等因素则会改变土壤表面的结构,产生各种微小阴影,对漫反射光谱的角度、方向和强度都会产生影响^[37]。此外,这种土壤-光谱映射关系还具有典型的区域变异特性,即在不同空间维度的各个地理单元内的这种土壤-光谱映射关系也存在一定差异,地理环境数据在二维空间上展现出的这种特殊性增加了土壤组分信息遥感反演模型的建立难度^[15,38]。传统的随机森林遥感反演模型在构建中通常是将地理样本在空间上假设为独立存在的,通过不断改变决策树的数量和节点数目去拟合这种真实的地表环境过程,进而实现光谱信号到土壤属性值的信息转换^[39]。但很明显,在真实的地表环境中,地理样本间会存在一定关联性,并且与位置信息和相距距离密切相关,因此直接将随机森林模型应用至真实地表环境会产生一定偏差。所以本研究在随机森林算法中通过加入空间关联函数模块将地理样本数据的位置信息考虑进模型中去,以此帮助分析土壤全氮样本间关联性与异质度,形成最优的输入样本组合。减少空间关联随机森林模型在校准过程中决策树和预测节点的复杂选择和设置,提升计算效率和精度,使反演模型的稳健性得到上升,准确地完成区域土壤全氮含量的遥感反演。

在不同空间关联度(c 值)下,建立的空间关联随机森林模型的精度变化见图5。可以看出,随着 c 值的减小,输入模型的样本数据之间的空间关联度不断增强,反演模型的精度也有提升,在当 c 值为2时模型达到了最高的性能。随着输入样本数据的

空间关联度的继续增强,输入样本的数量已无法满足模型的拟合需求,反演精度迅速下降。总体来看,空间关联随机森林模型通过引入空间关联函数模型,选取出最优的输入样本组合,使模型结构得到优化,有效的提升在区域尺度上土壤组分信息遥感反演和制图能力。

图5 不同 c 值下空间关联随机森林模型的反演精度变化

2.4 土壤全氮含量空间分布

为进一步验证构建的反演模型的稳定性和应用能力,将局地尺度上建立的最优空间关联随机森林遥感反演模型推广至整个区域,得到研究区的土壤全氮含量的空间分布(图6-a)。结合研究区土地利用概况(图6-b)可以看出,土壤全氮含量在田块尺度上的空间分异明显,与土地利用状况关系密切,斑块状分布较均质的为村居建筑物,周边包围有大量耕地。田块尺度上耕地土壤全氮含量的空间变化明显,距离村居建筑物越近,土壤的全氮含量越低(<0.80 g/kg),而距离建筑物较远的大片耕地,土壤全氮含量值较高(>1.20 g/kg)。这可能与刁镇的土地利用现状相关,山东生活上济南市章丘区是国内知名的大葱生产基地,大面积集中分布的耕地主要以粮食和蔬菜生产为主,农家肥和有机肥

料的持续施用,使得土壤养分含量明显提升。同时建筑物旁边往往保留一定量的土地用于经济开发,此类土地缺少农家肥料养护,故土壤全氮含量低于平均水平。但是,土壤中全氮的形成与变化不仅是自然的过程(母质演化、地形、植被、降水、温度等要素的长期相互作用),而且受到人类活动的影响,即土地利用变化、耕作干扰、有机肥施用以及种植制度和轮作方式,甚至灌溉栽培方式等人类活动的影响。所以,尽管土壤全氮含量分布与土地利用状况的关联关系十分明显,但是根据土壤全氮含量空间分布图仍无法判断出,影响土壤全氮含量的主要因素,需要结合自然地理和农业生产等相关知识继续深入探究。

以优选后的光谱指数作为输入变量,构建得到的基于空间关联随机森林框架的反演模型拓展至区域卫星影像,所得分布图能够反映全氮的空间变异特征。但值得注意的是,反演制图精度的近一步提升受到诸多因素的限制,例如地表水汽、样本采集和混合像元。本研究中获取哨兵影像生成时间为 2019 年 3 月,此时处于冬春之交,降水逐渐丰沛,地表水汽会吸收掉部分光谱,使波段反射率下降,

尽管通过光谱指数构建可以减弱环境因素的干扰,但水汽吸收的影响仍不可忽视^[40]。刘焕军等研究也表明,地表水汽会对土壤数字制图产生干扰,通过构建新型光谱指数和多期影像指数,可以减弱水汽的干扰,在下一步的研究中可以尝试改进^[41]。此外,采集到的土壤样本数目会对反演模型的训练产生影响,尽管研究中土壤样本的采集充分考虑了区域代表性,选取对各种样本集适应能力强的随机森林作为反演建模的算法基础,但是数量更多和分布更密集的土壤样本将会明显提升模型的稳健性^[42]。高光谱卫星影像具有更高的光谱分辨率,对土壤信息有更加灵敏的响应能力,虽然当前国内外高光谱卫星研究正处于起步阶段,数据的获取、处理和应用存在一定难度,但接下来机器学习算法和高光谱影像相结合的遥感反演方式必定会受到越来越多的关注。遥感反演制图中存在的光谱混合像元是无法回避的问题,尽管研究选取了较为均质的耕地区域作为研究区,并且哨兵 2 号多光谱影像的空间分辨率达到 10 m,但是道路及植被仍然可能会对目标物质的光谱特性产生影响,混合像元分解用于反演制图精度的提升在接下来的研究中须要深入探究。

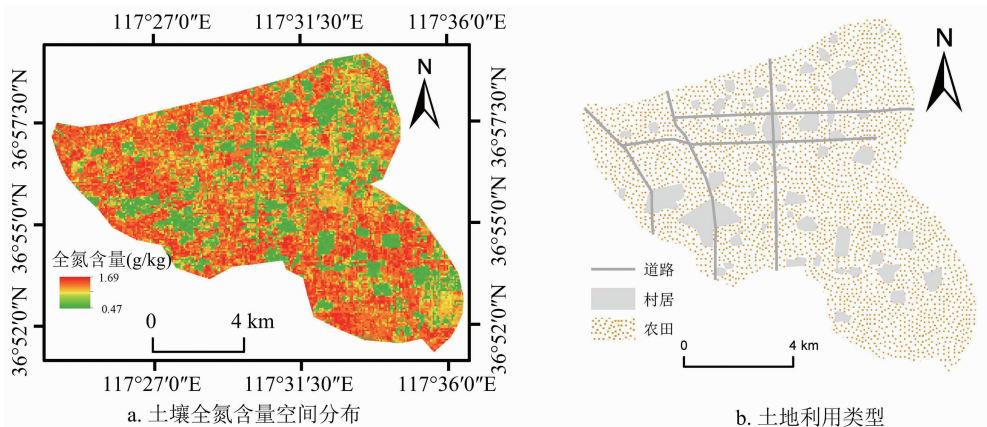


图6 研究区土壤全氮含量的分布图及土地利用类型概况

3 结论

哨兵 2 号影像的单波段与土壤全氮含量相关性较低,波段间进行比值变换能够显著增强光谱信号对土壤全氮信息的响应能力,对土壤全氮含量最敏感的光谱指数分别为 b_6/b_{11} 、 b_{8a}/b_{12} 、 b_8/b_9 、 b_{8a}/b_9 和 $(b_9 - b_{11})/(b_9 + b_{11})$,与土壤全氮含量显著或极显著相关;单波段与土壤全氮含量的相关性在 0.01 水平上均不显著。

考虑地理位置信息的空间关联随机森林模型

的精度指标 R^2 和 $RMSE$ 分别为 0.90 和 0.11,相对比随机森林算法支持下的遥感反演模型 ($R^2 = 0.81$; $RMSE = 0.15$),精度分别提升了 11.11% 和 26.67%,模型的结构得到优化,并且提升了计算效率,可以实现区域尺度的土壤全氮含量的反演制图。

土壤全氮含量在田块尺度上的空间变异明显,变异趋势与土地利用状况关系密切。村居建筑物周边土壤全氮含量处于较低水平 (<0.80 g/kg),远离村居的耕地区域土壤全氮含量较高 (>1.20 g/kg)。

参考文献:

- [1] Zhang Y, Li M Z, Zheng L H, et al. Spectral features extraction for estimation of soil total nitrogen content based on modified ant colony optimization algorithm[J]. *Geoderma*, 2019, 333: 23–34.
- [2] 彭远新, 蒙永辉, 徐夕博, 等. 基于随机森林算法的淮北平原土壤全氮高光谱估算[J]. *安全与环境学报*, 2020, 20(5): 1975–1983.
- [3] Wang X P, Zhang F, Johnson V C. New methods for improving the remote sensing estimation of soil organic matter content (SOMC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR) in northwest China[J]. *Remote Sensing of Environment*, 2018, 218: 104–118.
- [4] Zhang Y, Sui B, Shen H O, et al. Mapping stocks of soil total nitrogen using remote sensing data: a comparison of random forest models with different predictors[J]. *Computers and Electronics in Agriculture*, 2019, 160: 23–30.
- [5] Krishnan P, Alexander J D, Butler B J, et al. Reflectance technique for predicting soil organic matter[J]. *Soil Science Society of America Journal*, 1980, 44(6): 1282–1285.
- [6] 张平, 闫宏涛. 反射光谱法测定土壤和作物中氮[J]. *仪器仪表与分析监测*, 1999(1): 37–39.
- [7] 卢艳丽, 白由路, 王磊, 等. 黑土土壤中全氮含量的高光谱预测分析[J]. *农业工程学报*, 2010, 26(1): 256–261.
- [8] 张强, 张国龙, 张泽, 等. 基于高光谱的土壤全氮含量估测[J]. *山西农业科学*, 2016, 44(7): 972–976.
- [9] 李焱, 王让会, 管延龙, 等. 基于高光谱反射特性的土壤全氮含量预测分析[J]. *遥感技术与应用*, 2017, 32(1): 173–179.
- [10] 马驰. HJ-1A HSI 与 Sentinel-2A 遥感数据土壤全氮含量反演精度的对比研究[J]. *无线电工程*, 2021, 51(12): 1540–1548.
- [11] 贾伟, 高小红, 杨扬, 等. 基于 Landsat 8 OLI 影像的三江源区表层土壤全氮空间格局反演[J]. *干旱区研究*, 2015, 32(5): 890–896.
- [12] 雷浩川, 刘尊方, 于晓晶, 等. 基于 Landsat 5 影像的青海省大通县土壤表层全氮空间格局反演[J]. *青海大学学报*, 2021, 39(6): 79–86.
- [13] Zhang Y, Sui B, Shen H O, et al. Mapping stocks of soil total nitrogen using remote sensing data: a comparison of random forest models with different predictors[J]. *Computers and Electronics in Agriculture*, 2019, 160: 23–30.
- [14] Zhou T, Geng Y J, Chen J, et al. Mapping of soil total nitrogen content in the middle reaches of the Heihe River Basin in China using multi-source remote sensing-derived variables[J]. *Remote Sensing*, 2019, 11(24): 2934.
- [15] Liu X Q, Zhu A X, Yang L, et al. A graded proportion method of training sample selection for updating conventional soil maps[J]. *Geoderma*, 2020, 357: 113939.
- [16] Bao Y L, Meng X T, Ustin S, et al. Vis-SWIR spectral prediction model for soil organic matter with different grouping strategies[J]. *Catena*, 2020, 195: 104703.
- [17] Wadoux A M J C, Brus D J, Heuvelink G B M. Sampling design optimization for soil mapping with random forest[J]. *Geoderma*, 2019, 355: 113913.
- [18] Lagacherie P, Arrouays D, Bourennane H, et al. Analysing the impact of soil spatial sampling on the performances of Digital Soil Mapping models and their evaluation: a numerical experiment on Quantile Random Forest using clay contents obtained from Vis-NIR-SWIR hyperspectral imagery[J]. *Geoderma*, 2020, 375: 114503.
- [19] Tobler W R. A computer movie simulating urban growth in the Detroit region[J]. *Economic Geography*, 1970, 46: 234.
- [20] 韩琳, 徐夕博. 基于 PMF 模型及地统计的土壤重金属健康风险定量评价[J]. *环境科学*, 2020, 41(11): 5114–5124.
- [21] 田颖, 陈卓奇, 惠凤鸣, 等. 欧空局哨兵卫星 Sentinel-2A/B 数据特征及应用前景分析[J]. *北京师范大学学报(自然科学版)*, 2019, 55(1): 57–65.
- [22] Davis E, Wang C, Dow K. Comparing Sentinel-2 MSI and Landsat 8 OLI in soil salinity detection: a case study of agricultural lands in coastal North Carolina[J]. *International Journal of Remote Sensing*, 2019, 40(16): 6134–6153.
- [23] Wang J, Ding J, YU D, et al. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China[J]. *Geoderma*, 2019, 353: 172–187.
- [24] Rossel R A V, Behrens T. Using data mining to model and interpret soil diffuse reflectance spectra[J]. *Geoderma*, 2010, 158(1/2): 46–54.
- [25] Zhang S, Lu X, Zhang Y Z, et al. Estimation of soil organic matter, total nitrogen and total carbon in sustainable coastal wetlands[J]. *Sustainability*, 2019, 11(3): 667.
- [26] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5–32.
- [27] Stewart F A, Brunson C, Charlton M. Geographically weighted regression: the analysis of spatially varying relationships[M]. Chichester: John Wiley & Sons, 2002.
- [28] Yang S H, Liu F, Song X D, et al. Mapping topsoil electrical conductivity by a mixed geographically weighted regression kriging: a case study in the Heihe River Basin, northwest China[J]. *Ecological Indicators*, 2019, 102: 252–264.
- [29] 刘焕军, 张美薇, 杨昊轩, 等. 多光谱遥感结合随机森林算法反演耕作土壤有机质含量[J]. *农业工程学报*, 2020, 36(10): 134–140.
- [30] Chi Y, Shi H H, Zheng W et al. Simulating spatial distribution of coastal soil carbon content using a comprehensive land surface factor system based on remote sensing[J]. *Science of the Total Environment*, 2018, 628/629: 384–399.
- [31] 侯宇初, 张冬有. 基于 Landsat8 遥感影像的地表温度反演方法对比研究[J]. *中国农学通报*, 2019, 35(10): 142–147.
- [32] Gholizadeh A, Saberioon M, Ben-Dor E, et al. Modelling potentially toxic elements in forest soils with vis-NIR spectra and learning algorithms[J]. *Environmental Pollution*, 2020, 267: 115574.