

不同土壤类型的有机质含量的可见-近红外光谱 检测模型传递方法研究

胡国田^{1,2,3}, 尚会威^{1,2,3}, 谭瑞虹¹, 许翔虎¹, 潘伟东¹

1. 西北农林科技大学机械与电子工程学院, 陕西 杨凌 712100
2. 农业农村部农业物联网重点实验室, 陕西 杨凌 712100
3. 陕西省农业信息感知与智能服务重点实验室, 陕西 杨凌 712100

摘要 利用可见-近红外光谱分析技术可以准确快速的获取土壤养分含量,但不同类型土壤间养分含量校正模型的普适性是亟待解决的关键问题。为提高有机质含量光谱校正模型在多类型土壤之间的普适性和农田在线检测有机质含量速度,利用美国 M107B 区 66 个样品建立基于可见-近红外光谱的土壤有机质含量的粒子群-最小二乘支持向量机(PSO-LSSVM)校正模型,预测 M107B 区的 23 个验证集样品的决定系数 $R^2 = 0.859$, 相对分析误差 $RPD = 2.660$; 将 M107B 区 89 个土壤样品作为校正集建模后对 N116B 区 20 个验证集样品的有机质含量预测,预测 $R^2 = 0.562$, 预测 $RPD = 0.952$, 模型的预测 R^2 和预测 RPD 分别降低 34.6% 和 64.2%, 表明 M107B 区土壤有机质含量的可见-近红外光谱校正模型直接用于 N116B 区时,预测精度显著降低; 将 N116B 区部分土壤样品加入到 M107B 区样品集后重新建模,并预测 N116B 区 20 个验证集样品的有机质含量,当加入的 N116B 区土壤样品数量达到 35 以上,预测 $R^2 > 0.80$, 预测 $RPD > 2.0$; 加入到校正集的 N116B 区土壤样品数量从 0 增加到 50, 模型预测 R^2 从 0.562 增加到 0.811, 预测 RPD 从 0.952 增加到 2.274, 精度逐渐提高。结果表明,在 M107B 区校正模型中加入 N116B 区部分土壤样品建模,能够有效提高 M107B 区土壤校正模型对 N116B 区土壤有机质含量的预测精度; 加入的 N116B 区土壤样品数量达到 50 以上,模型预测性能趋于稳定,预测精度达到实用要求,成功将 M107B 区土壤有机质含量校正模型传递给 N116B 区土壤; 优先选择与 M107B 区土壤样品的有机质含量或光谱曲线差异较大的 N116B 区土壤样品参与建模,可有效避免模型传递时模型性能出现突变。提出的方法能够有效提高 M107B 区土壤的有机质校正模型对 N116B 区土壤的预测精度,为基于可见-近红外光谱的农田土壤有机质含量实时检测提供一种新的经济可行的模型传递方法,为提高多类型土壤的有机质含量检测模型的普适性提供一种有效的解决方案。

关键词 可见-近红外光谱; 精细农业; 土壤有机质; 粒子群-最小二乘支持向量机; 模型传递

中图分类号: O657.33 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)10-3148-07

引言

土壤有机质(soil organic matter, SOM)是评价土壤肥力和养分的重要指标,是科学精准施肥需获取的重要信息。准确获取土壤有机质含量是保持土壤肥力、维持良好耕地质量的基础。传统的土壤有机质含量测量方法虽然精度比较高,但是操作复杂、周期长、破坏性大、成本较高,不适合大面

积测量。土壤可见-近红外光谱曲线包含丰富的光谱信息,能够综合反映多种土壤信息。近年来,国内外许多学者针对某一地区或某种土壤开展有机质含量光谱预测研究,有机质含量的光谱预测精度不断提高^[1-3],但在进行大范围多类型土壤的有机质含量预测时,不同类型土壤的光谱差异会影响预测精度。

对多类型土壤建立有机质含量校正模型主要有两种方法,即用大范围多类型的土壤样品建模和模型传递。Brown

收稿日期: 2021-08-07, 修订日期: 2022-03-03

基金项目: 陕西省重点研发计划(一般项目)(2017NY-170), 财政部和农业农村部国家现代农业产业技术体系项目(CARS-23-C07), 陕西省科技创新团队(资源高效设施农业创新团队)项目(2021TD-34)资助

作者简介: 胡国田, 1977 年生, 西北农林科技大学机械与电子工程学院副教授, 美国农业部农业研究所访问学者

e-mail: hugttrh@163.com

等^[4]从全球五大洲采集 4 184 个独立样本对阳离子交换量、土壤有机碳、土壤无机碳、粘土含量、游离铁含量等 7 种土壤属性进行预测,证明了可见-近红外光谱在全球土壤表征方面的基本可行性。Krishnan 等^[5]利用逐步多元线性回归对美国伊利诺斯州 4 种类型土壤的光谱反射率数据分析,发现 564 和 623 nm 是有机质的敏感波段。陈昊宇等^[6]采用连续小波变换对多种土壤类型光谱数据进行预处理,提升光谱数据与有机质含量之间的相关系数,进而提高模型精度。纪文君等^[7]提取了 7 组不同地区不同类型土壤样品共同的有机质响应波段,为建立多类型土壤的有机质含量校正模型提供参考。但若采集全球所有类型土壤光谱并进行土壤有机质含量预测并不可行,且模型精度也可能随着土壤类型数量的增加而降低。针对某一土壤类型的光谱校正模型应用到其他土壤类型的养分含量检测精度显著下降问题,模型传递提供了一种有效的解决方案。Padarian 等^[8]论证了模型传递的有效性,模型传递前后土壤有机碳、阳离子交换量、粘土含量和 pH 值的预测均方根误差都降低了 10% 以上。Panchuk 等^[9]通过直接校正算法(DS)成功将一定波长范围的杏仁糖近红外光谱定量模型应用于另一台近红外仪器。国内外许多学者也尝试使用其他模型传递算法如典型相关分析算法^[10]、斜率截距校正法^[11]等。

目前,大多数模型传递研究以不同仪器之间模型传递为主,主要在改良算法或开发新算法范畴内进行,通常需要大量的样品来使模型传递更加可靠^[12]。这些算法在解决不同类型土壤之间的模型传递时具有局限性,难以适应田间土壤养分含量的快速检测需求。本研究探索在已有土壤类型的有机质含量光谱预测模型基础上,通过适当增加另一土壤类型的样品参与建模,优化已有模型,提高已有模型对新土壤类型的有机质含量检测精度,为基于可见-近红外光谱的农田多类型土壤的有机质含量实时检测提供一种新的经济可行的模型传递方法,提高模型的普适性,为建立大面积小样本的多类型土壤的有机质含量校正模型提供参考。

1 实验部分

1.1 样本采集

主要土地资源区(major land resource areas, MLRAs)是美国农业部定义的用来描述相似气候、土壤和土地用途的区域,是依据土壤、气候、水资源或土地利用的显著地理特征进行划分^[13]。本研究土壤样品采自 N116B 区和 M107B 区,两区的地理位置如图 1 所示。

N116B 区位于斯普林菲尔德平原,属于亚热带常绿阔叶林气候地带,年降水集中,年均降水量为 660~1 040 mm,年均气温 8~13 °C,牧场面积占比近 50%。M107B 区位于爱荷华州和密苏里州中部的黄土丘陵,属于温带草原气候地带,其年均降水量 1 040~1 145 mm,年均气温 13~15 °C,该地区农场较多,存在水土流失、土壤中有机的枯竭和水质差等资源问题。共采集了 259 份土壤样品,其中 M107B 区土壤样品 89 份, N116B 区土壤样品 170 份。所有土壤样品的采样深度均为 0~15 cm。所有样品用烤箱烘干后研磨,过 2

mm 孔筛,再分成两份分别用于实验室理化检测和光谱扫描。土壤有机质的理化检测使用灼烧法,在密苏里大学的土壤与植物检测实验室完成,理化检测得到的源土壤 M107B 区和目标土壤 N116B 区的土壤有机质含量统计特征见表 1。

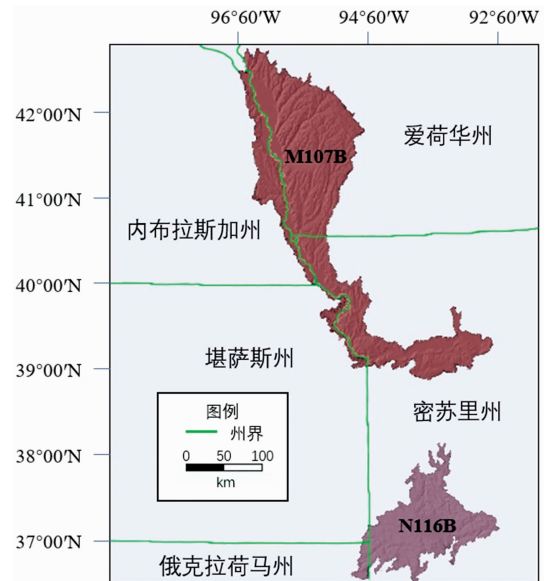


图 1 N116B 区和 M107B 区在美国地图上的位置
Fig. 1 Locations of N116B & M107B on the map of USA

表 1 土壤有机质含量理化分析结果统计
Table 1 Statistics of physical and chemical analysis results of soil organic matter

土壤类型	样品数量	土壤有机质含量/%			
		最大值	最小值	均值	标准偏差
源土壤 M107B 区	89	5.700	0.800	2.723	1.121
N116B 区校正集	150	6.900	1.000	3.357	1.035
N116B 区验证集	20	5.800	1.700	3.380	0.931

从研磨过筛的土壤样品中取约 15 cm³ 土壤装入底部为玻璃的样品杯中,摇晃样品杯使较细颗粒土壤位于杯底,卤素灯的光通过玻璃照射土壤样品表面,反射光通过光纤传输到光谱仪。光谱仪采用美国 ASD 公司的 FieldSpec Pro FR 光谱仪,该光谱仪由 3 个不同波段的独立光谱仪组合而成,波段分别为 350~1 100, 1 000~1 900 和 1 700~2 500 nm,其总波长范围是 350~2 500 nm,光谱仪分辨率为 3 nm(350~1 000 nm), 10 nm(1 000~2 500 nm); 采样间隔为 1.4 nm(350~1 000 nm), 2 nm(1 000~2 500 nm); 采样输出数据间隔为 1 nm。计算机安装光谱采集软件 FieldSpec RS³ 记录土壤样品的光谱数据。采集光谱时,FieldSpec RS³ 的参数设置为:光谱平均数 30 次,白板扫描平均数 50 次,暗电流扫描平均数 50 次。暗电流扫描在每个扫描阶段开始时进行,之后每 30 min 最少扫描 1 次。按 0°, 45° 和 90° 放置样品杯,采集各角度下的土壤样品光谱。若 3 次所得反射光谱相似,取其平均值作为该土壤样品的光谱数据;若 3 次反射光谱中有 1 次获得的光谱反射率与另外 2 次有显著差异,则去除该次光谱扫描数据,并以另外 2 次光谱反射率的均值作为该土壤样

品的光谱数据。每扫描 5 个土壤样品后进行白板校正以消除参考光谱误差。光谱扫描完成后去除所有土壤样品的光谱曲线中光谱数据信噪比低的边缘波段 350~400 和 2 451~2 500 nm。光谱扫描在密苏里大学校内的美国农业部农业研究所完成。

图 2 为源土壤 M107B 区和目标土壤 N116B 区两组土壤样品的平均光谱反射率曲线。从 401~2 450 nm 全波段范围的光谱反射率来看, 两个地区土壤的光谱曲线在 1 400, 1 900 和 2 200 nm 附近都存在明显的水分吸收谷。N116B 区土壤类型主要为淋育土(Alfisol)、极育土(Ultisol)和黑沃土(Mollisol), 大部分土地用地为农场和牧场, 土壤的有机质含量较高; M107B 区土壤类型主要为黑沃土, 存在水土流失、土壤中有有机物耗竭等资源问题, 土壤的有机质含量较低; 且 M107B 区表层土壤的颜色比 N116B 区的更深。故 N116B 区土壤有机质含量高于 M107B 区(如表 1 所示), 但 N116B 区土壤的光谱反射率却比较高。

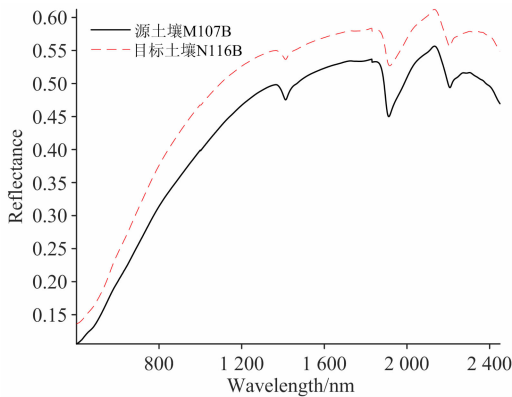


图 2 源土壤 M107B 区和目标土壤 N116B 区的土壤反射率均值曲线

Fig. 2 Mean reflectance spectra of soil samples from source soil M107B and target soil N116B

1.2 数据处理

1.2.1 样本集划分

利用 Kennard-Stone(KS)算法将 M107B 区 89 个土壤样品按照 3:1 的比例划分为校正集和验证集。随机挑选出 N116B 区 20 个土壤样品作为 N116B 区验证集, 本研究中后续对 N116B 区土壤有机质含量进行预测, 均是指对这 20 个验证集土壤样品进行预测; N116B 区剩余 150 个土壤样品随机划分成 30 组, 每组包含 5 个土壤样品, 再根据实验需要加入到 M107B 区样品集中。

1.2.2 特征波长提取

去除光谱曲线中光谱数据信噪比低的边缘波段 350~400 和 2 451~2 500 nm 后, 采用窗口竞争性自适应重加权采样(window competitive adaptive reweighted sampling, WCARS)和连续投影算法(successive projections algorithm, SPA)进行波长选择以提高预测精度, 简化模型^[14]。WCARS 结合“窗口”与 CARS 方法的优势, 能有效增强特征波长变量选择的准确性和稳定性。SPA 通过提取最低限度的冗余信息的波长组合, 消除共线问题。因此, 采用 WCARS 联合 SPA

算法获取较少的特征波长。将 M107B 区 89 个土壤样品的光谱数据通过 WCARS+SPA 挑选后, 确定 544, 728, 1 486, 1 830, 1 895, 2 285 和 2 450 nm 等 7 个特征波长, 并将这 7 个波长作为 SOM 特征波长。

1.3 模型构建

最小二乘支持向量机(least squares support vector machines, LSSVM)建模时, 如何对正则参数 λ 和核参数 σ 进行优化是不可避免的问题。粒子群算法(particle swarm optimization, PSO)是一种通过更新速度和位置来不断更新到参数最优解的群体优化智能算法, 自提出以来被广泛应用^[15]。本研究通过 PSO 优化 LSSVM, 建立土壤有机质含量校正模型。

1.4 方法

以 M107B 区土壤样品为源土壤, 以 N116B 区土壤样品为目标土壤, 从 N116B 区 170 个土壤样品中随机挑选出 20 个作为验证集。将 M107B 区 89 个土壤样品采用 KS 算法按照 3:1 划分校正集和验证集, 用 WCARS+SPA 提取 SOM 特征波长后, 从 3 个方面进行实验研究, 研究流程如图 3 所示。(1)通过 PSO-LSSVM 算法对 M107B 区 66 个校正集样品建立有机质含量校正模型, 并用该区的验证集验证模型精度; (2)用 M107B 区 89 个土壤样品, 建立 SOM 含量的 PSO-LSSVM 校正模型, 用该模型预测 N116B 区的 20 个验证集土壤样品的有机质含量; (3)将 N116B 区剩余的 150 个土壤样品随机等分成 30 组, 在由 M107B 区的 89 个土壤样品组成的样品集中, 每次加入 N116B 区的 1 组 5 个土壤样品组成新的样品集, 以新的样品集为校正集建立校正模型, 再用校正模型对 N116B 区 20 个验证集样品进行 SOM 含量预测, 直至全部加入 N116B 区的 30 组土壤样品。比较 M107B 区校正模型对 M107B 区及 N116B 区 SOM 含量预测精度, 分析加入建模的 N116B 区土壤样品数量与模型性能之间的关系。

1.5 模型评价指标

以预测决定系数(coefficient of determination, R^2)、预测均方根误差(root mean square error of prediction, RMSEP)和预测相对分析误差(ratio of standard deviation to RMSEP, RPD)对模型评价。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

$$RMSEP = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$RPD = \frac{SD}{RMSEP}$$

式中, n 为验证集样本数; y_i 为验证集第 i 个样品用化学方法测定的 SOM 含量实际值; \hat{y}_i 为验证集第 i 个样品的 SOM 含量预测值; \bar{y}_i 为验证集样品 SOM 含量均值; SD 为验证集样品 SOM 含量标准差。预测决定系数 R^2 用于衡量预测值与实际值之间相关程度, 越接近 1 则模型相关性越高; RMSEP 用于衡量验证集样品 SOM 含量的预测值和实际值之间的误差大小, 越接近 0 则模型预测能力越强; RPD 用于比较不同

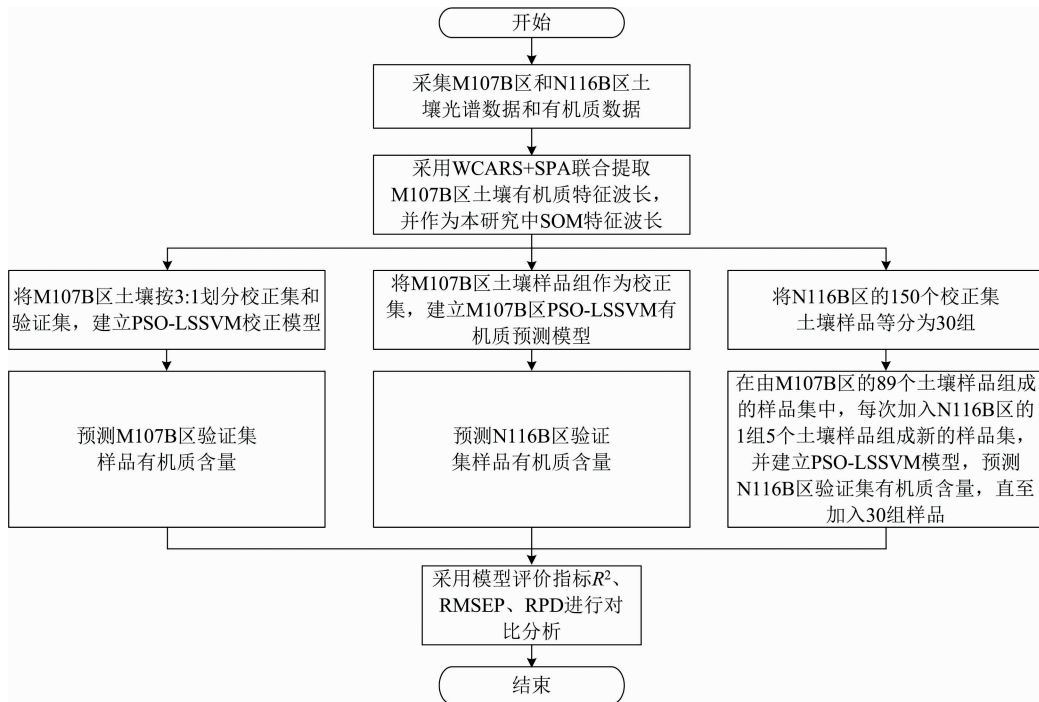


图 3 实验流程图

Fig. 3 Experimental flow chart

组分或不同样品的模型预测效果, $RPD > 2.0$ 时, 模型能成功预测土壤样品的有机质含量。

2 结果与讨论

2.1 未加入目标土壤样品的模型传递

实验(1)建立 M107B 区土壤有机质 PSO-LSSVM 校正模型并预测验证集土壤有机质含量, 结果如表 2 所示, 预测 $R^2 = 0.859$, 预测 $RPD = 2.660$, 表明模型能有效预测 M107B 区内部的土壤有机质含量。实验(2)以 M107B 区 89 个土壤样品作为校正集建立土壤有机质含量校正模型, 对 N116B 区 20 个验证集土壤样品进行预测, 结果如表 2 所示, 其模型预测 $R^2 = 0.562$, 预测 $RPD = 0.952$, 表明 M107B 区的校正模型不能有效预测 N116B 区的土壤有机质含量。实验(1)和实验(2)建模过程中均只使用 M107B 区土壤样品, 直接将 M107B 区土壤有机质含量校正模型用于 N116B 区土壤样品时模型预测精度显著降低, 预测决定系数 R^2 降低了 34.6%, 预测相对分析误差 RPD 降低了 64.2%, 模型传递失败。

表 2 实验(1)和实验(2)的模型性能

Table 2 Model performances of experiments (1) and (2)

实验	校正集 样品数量	验证集 样品数量	PSO-LSSVM		
			R^2	RMSEP/%	RPD
实验(1)	66	23	0.859	0.334	2.660
实验(2)	89	20	0.562	0.977	0.952

2.2 加入目标土壤样品后的模型传递

实验(3)用 M107B 区土壤样品建立有机质含量校正模型时加入部分 N116B 区土壤校正集样品, 建立校正模型后预测 N116B 区的验证集土壤有机质含量。校正模型预测 N116B 区的验证集土壤有机质含量的 R^2 , RMSEP 和 RPD 值随加入的 N116B 区土壤校正集样品数量的变化结果如表 3 和图 4 所示, 表 3 中仅列出 RPD 值变化超过 1% 的实验数据。随着加入的 N116B 区土壤校正集样品数量增加, 校正模型的预测 R^2 和 RPD 逐渐提高, RMSEP 逐渐降低。当加入样品数量达到 35 时, 校正模型的预测 $R^2 > 0.80$, $RPD > 2.0$; 当加入样品数量达到 50 以上, 模型精度变化趋于平稳。但是, 在加入 101—105 号 N116B 区土壤校正集样品后, 校正模型预测 R^2 发生了突变, 这些样本可能对 N116B 区土壤的预测精度有较大影响, 因此需对 101—105 样本进一步研究。

2.3 模型传递的稳定性

加入 N116B 区 101—105 号土壤校正集样品后, 模型精度出现突变(如图 4 所示)。为了探明加入单个样品对模型精度的影响, 本节研究将 101—110 号目标土壤样品提取出来, 每次只新增 1 个样品到校正集中参与建模, 即: 第一次对 M107B 区 89 个土壤样品和 N116B 区 1—101 号样品组成新的校正集建立校正模型, 第二次对 M107B 区 89 个土壤样品和 N116B 区 1—102 号样品组成新的校正集建立校正模型, 以此类推, 直到 101—110 号样品全部参与建模。再对 N116B 区的 20 个验证集样品进行土壤有机含量预测, 预测 R^2 如图 5 所示。结果表明: 当加入 105 号目标土壤校正集样品时模

型 R^2 由 0.810 提高到 0.867。将 105 号样品去除后重新建模, 模型决定系数 R^2 如图 5, 模型 R^2 不再提高, 因此确定 105 号样品引起模型 R^2 发生突变。

表 3 实验(3)模型性能

Table 3 Model performances of experiment (3)

实验	校正集 样品数量	验证集 样品数量	PSO-LSSVM			
			R^2	RMSEP/%	RPD	
实验(3)	89+5	20	0.562	0.897	1.038	
	89+10		0.563	0.690	1.349	
	89+15		0.657	0.656	1.419	
	89+20		0.766	0.503	1.850	
	89+25		0.794	0.466	1.999	
	89+30		0.793	0.438	2.125	
	89+35		0.805	0.443	2.101	
	89+40		0.814	0.413	2.255	
	89+45		0.805	0.430	2.166	
	89+50		0.811	0.409	2.274	
	...					
	89+100		0.809	0.399	2.329	
	89+105		0.867	0.334	2.783	
	89+110		0.865	0.334	2.783	
	89+115		0.840	0.363	2.561	
...						

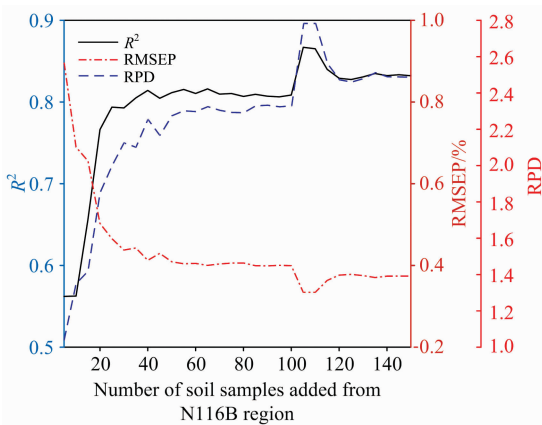


图 4 加入不同数量的 N116B 区土壤样品后的模型性能
Fig. 4 Model performances when adding different numbers of soil samples from N116B region

为找出平稳区影响模型 R^2 值发生 0.05 以上突变的样品的特征, 本研究按照实验(3)方法进行多次重复实验, 共提取出 20 个与 105 号样品类似的能使模型 R^2 值提高 0.05 以上的 N116B 区校正集土壤样品, 其有机质含量及其特征波长的光谱反射率如表 4 所示。由于源土壤 M107B 区样品的 SOM 平均浓度低于目标土壤 N116B 区样品的 SOM 平均浓度(如表 1 所示), 且光谱反射率均值低于目标土壤 N116B 区(如图 2 所示), 因此, 造成校正模型预测 R^2 值突然提高 0.05 的原因可能是 N116B 区这 20 个土壤样品较 M107B 区土壤样品的 SOM 浓度较高或光谱反射率较高。该 20 个样品

对模型传递精度影响较大, 因此后续研究中, 在加入 N116B 区土壤样品建模时将这 20 个特殊样品的优先级提高, 即先加入这 20 个特殊样品, 再把 N116B 区其余 130 个土壤校正集样品随机等分成 26 组, 每次加入 1 组 5 个土壤样品到 M107B 区土壤样品集中作为新的样品集, 以新的样品集为校正集建立校正模型, 再用校正模型对 N116B 区 20 个验证集土壤样品进行有机质含量预测, 直至加入 N116B 区所有校正集土壤样品。校正模型预测 N116B 区的验证集土壤有机质含量的 R^2 , RMSEP 和 RPD 值随加入的 N116B 区土壤校正集样品数量的变化结果如图 6 所示, 当加入的 N116B 区土壤校正集样品数量由 5 增加到 45 时, 校正模型预测 N116B 区验证集土壤有机质含量的 R^2 从 0.690 逐渐增加到 0.800; 当加入的样品数量达到 45 以后, 校正模型预测 R^2 在 ± 0.015 范围内波动, 逐渐趋于稳定, 未出现模型精度突变情况。因此, 在将 M107B 区土壤有机质校正模型传递给 N116B 区时, 应优先加入比 M107B 区土壤样品的 SOM 浓度高或光谱反射率高的 N116B 区土壤样品参与建模, 以提高模型预测精度和避免模型传递时模型精度出现突变。

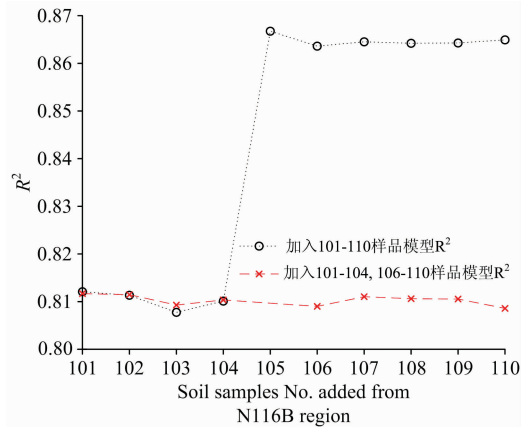


图 5 加入 105 号样品和不加入时的模型预测 R^2 对比
Fig. 5 Comparison of R^2 with or without No. 105 sample

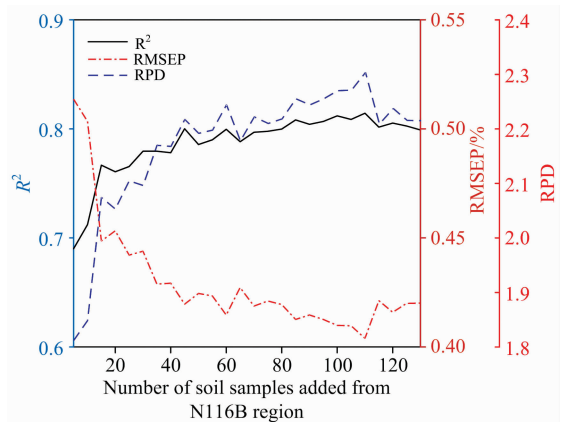


图 6 加入 N116B 区除特殊样品外的 130 个土壤样品的模型性能
Fig. 6 Model performances when adding 130 soil samples which are not special samples in N116B region

表 4 提高模型 R^2 值 0.05 以上的 N116B 区土壤校正集样品数据
Table 4 The data of calibration samples in N116B which improve R^2 above 0.05

土壤有机质含量/%	反射率						
	544 nm	728 nm	1486 nm	1830 nm	1 895 nm	2 285 nm	2 450 nm
1.0	0.303	0.457	0.608	0.629	0.602	0.633	0.614
1.4	0.306	0.461	0.600	0.642	0.621	0.646	0.628
2.7	0.265	0.412	0.577	0.599	0.569	0.592	0.565
2.8	0.295	0.435	0.598	0.618	0.599	0.624	0.612
2.8	0.286	0.435	0.617	0.640	0.602	0.626	0.594
3.0	0.296	0.437	0.607	0.632	0.597	0.620	0.592
3.2	0.265	0.414	0.598	0.623	0.598	0.623	0.605
4.8	0.197	0.328	0.586	0.617	0.580	0.600	0.570
4.8	0.212	0.361	0.558	0.584	0.562	0.581	0.563
4.9	0.193	0.310	0.580	0.614	0.583	0.608	0.520
4.9	0.262	0.398	0.587	0.613	0.590	0.606	0.591
4.9	0.132	0.209	0.492	0.536	0.506	0.538	0.507
4.9	0.215	0.363	0.590	0.617	0.586	0.603	0.579
5.0	0.171	0.296	0.551	0.580	0.539	0.557	0.519
5.5	0.228	0.356	0.562	0.591	0.565	0.585	0.566
5.7	0.138	0.225	0.509	0.549	0.514	0.538	0.500
5.9	0.202	0.338	0.551	0.582	0.559	0.581	0.566
6.3	0.117	0.193	0.450	0.492	0.468	0.494	0.469
6.4	0.210	0.340	0.533	0.563	0.540	0.560	0.544
6.9	0.221	0.342	0.552	0.583	0.550	0.564	0.537

3 结 论

基于可见-近红外光谱建立 M107B 区的土壤有机质含量校正模型, 探索将 M107B 区土壤有机质校正模型应用到 N116B 区土壤有机质含量预测的模型传递方法。研究结果表明, M107B 区土壤有机质的 PSO-LSSVM 校正模型 $R^2 = 0.859$, $RMSEP = 0.334\%$, $RPD = 2.660$, 可以有效预测 M107B 区土壤有机质含量。用 M107B 区土壤有机质校正模型预测 N116B 区的土壤验证集样品的有机质含量和预测 M107B 区土壤验证集样品的有机质含量相比, 预测 R^2 和预测 RPD 分别降低了 34.6% 和 64.2%, 且预测精度很低, 不能有效预测 N116B 区土壤有机质含量。将部分 N116B 区土壤样品加入 M107B 区土壤样品集后重新建模, 模型对

N116B 区土壤的有机质含量预测精度随加入的 N116B 区土壤样品数量增加逐渐提高。在加入样品数量达到 50 后模型精度变化趋于稳定, 且模型预测 $R^2 > 0.80$, $RPD > 2.0$, 能有效预测土壤有机质含量, 成功实现 M107B 区模型预测 N116B 区土壤有机质含量的模型传递。为提高模型传递时的预测精度, 应优先加入比 M107B 区土壤样品的 SOM 浓度高或光谱反射率高的 N116B 区土壤样品参与建模。本文研究 M107B 区土壤有机质含量校正模型预测 N116B 区土壤有机质含量的建模方法, 提出了一种新的基于可见-近红外光谱的模型传递方法, 为建立大面积小样本的多类型土壤有机质含量校正模型提供一种有效的解决方案, 为提高大范围多类型土壤的有机质含量农田在线光谱检测的准确性、减少模型计算量、降低检测成本提供一种实用方法。

References

- [1] Ji W J, Rossel R A V, Shi Z. European Journal of Soil Science, 2015, 66(3): 555.
- [2] Wang J Z, Tiyyip T, Ding J L, et al. Journal of Spectroscopy, 2017, (1/2): 1.
- [3] Tan Y, Jiang Q G, Yu L F, et al. IEEE Access, 2021, 9: 5895.
- [4] Brown D J, Shepherd K D, Walsh M G, et al. Geoderma, 2005, 132(3): 273.
- [5] Krishnan P, Alexander J D, Butler B J, et al. Soil Science Society of America Journal, 1980, 44(6): 1282.
- [6] CHEN Hao-yu, YANG Guang, HAN Xue-ying, et al(陈昊宇, 杨光, 韩雪莹, 等). Journal of Agricultural Science and Technology(中国农业科技导报), 2021, 23(5): 132.
- [7] JI Wen-jun, SHI Zhou, ZHOU Qing, et al(纪文君, 史舟, 周清, 等). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2012, 31(3): 277.
- [8] Padarian J, Minasny B, McBratney A B. Geoderma, 2019, 340: 279.
- [9] Panchuk V, Kirsanov D, Oleneva E, et al. Talanta, 2017, 170: 457.

- [10] Zheng K Y, Zhang X, Iqbal J, et al. *Journal of Chemometrics*, 2014, 28(10): 773.
- [11] Dong X G, Dong J, Li Y L, et al. *Computers and Electronics in Agriculture*, 2019, 156: 669.
- [12] Vitaly P, Dmitry K, Ekaterina O, et al. *Talanta*, 2017, 170: 457.
- [13] USDA Natural Resources Conservation Service. *Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin. Agricultural Handbook 296*, Washington, D. C. 2006, 332-334, 375-377.
- [14] LIU Yan-de, WANG Jun-zheng, JIANG Xiao-gang, et al(刘燕德, 王军政, 姜小刚, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2021, 41(7): 2064.
- [15] ZHANG Wen, WU Zhi-bin, XU Jiu-ping(张 雯, 吴志彬, 徐玖平). *Control and Decision(控制与决策)*, 2022, 37(7): 1837.

Research on Model Transfer Method of Organic Matter Content Estimation of Different Soils Using VNIR Spectroscopy

HU Guo-tian^{1, 2, 3}, SHANG Hui-wei^{1, 2, 3}, TAN Rui-hong¹, XU Xiang-hu¹, PAN Wei-dong¹

1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling 712100, China

2. Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling 712100, China

3. Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling 712100, China

Abstract Soil properties can be estimated accurately and quickly using visible and near-infrared (VNIR) diffuse reflectance spectroscopy. However, a key problem is the lack of universal nutrient content calibration models for different soils. To improve the universality of the soil organic matter (SOM) content calibration model for different types of soils and the speed of online detection of the SOM in farmland, sixty-six samples of soil from M107B in the United States were used to establish the SOM content. Calibration model using the particle swarm optimization-based least squares support vector machines (PSO-LSSVM) method using VNIR spectroscopy. Then this calibration model predicted 23 samples of the validation set from M107B. The results gave the coefficient of determination (R^2) and the ratio of standard deviation to root mean square error of prediction (RPD) of 0.859 and 2.660, respectively. Subsequently, we predicted the SOM content of the validation set, including 20 samples from N116B, by the PSO-LSSVM calibration model of all 89 soil samples from M107B. The results showed decreases in the R^2 -value (0.562) and RPD (0.952). These decreases in R^2 and RPD values by 34.6% and 64.2%, respectively, indicated that the prediction accuracy was significantly decreased when the PSO-LSSVM calibration model of SOM content in M107B was directly used to predict SOM content in N116B. The PSO-LSSVM calibration model established by the calibration set, a combination of some soil samples from N116B and all 89 samples from M107B was also used to predict SOM content of the previous validation set from N116B and gave the R^2 values that were more than 0.80 and RPD values that were more than 2.0 when the number of soil samples from N116B was added over 35. In addition, R^2 increased from 0.562 to 0.811. RPD increased from 0.952 to 2.274 when the number of soil samples from N116B added to the calibration set increased from 0 to 50. The results showed that calibration model accuracy could be effectively improved by adding some soil samples from N116B to M107B calibration set when predicting SOM content in N116B. The prediction performance of models was stable, whereas the prediction accuracy met practical requirements when the number of soil samples from N116B added to the calibration set was more than 50. In addition, the calibration model of SOM in M107B was successfully transferred to the soil in N116B, and the samples in N116B with large differences in organic matter content or spectral curve from samples in M107B are preferred to adding to the calibration set because this method can effectively avoid the mutation of model transfer performance. In conclusion, the results provided a method to improve the SOM prediction accuracy of N116B soil using the SOM calibration model of M107B soil. Furthermore, the results provided a new, economical and feasible model transfer method for real-time estimating of SOM content in farmland based on VNIR. The results also provided an effective solution to improve the universality of the SOM content calibration model for different soil types.

Keywords Visible and near-infrared spectroscopy; Precision agriculture; Soil organic matter; Particle swarm optimization-based least squares support vector machines; Model transfer

(Received Aug. 7, 2021; accepted Mar. 3, 2022)