基于棉田原位高光谱数据的土壤 pH 值反演与制图研究

蔡海辉，彭 杰，柳维扬，罗德芳，王玉珍，白建铎，白子金
（塔里木大学 植物科学学院，新疆 阿拉尔 843300）

摘 要：[目的] 研究快速、准确大面积监测农田土壤 pH 值，为大面积土壤改良和实现农田精细化管理提供科学支持。[方法] 以南疆阿拉尔市十二团棉田为研究区，采用网格采样法采集 231 个样点的原位高光谱数据，并同步采集其中 116 个样点的土壤样品；分析了原位高光谱反射率数据经不同预处理模式后的光谱数据与土壤 pH 值的相关性；采用偏最小二乘回归、支持向量机回归和随机森林 3 种建模方法分别建立了土壤 pH 值的高光谱反演模型，根据模型评价指标优选出最优模型对未采集土壤样点的 pH 值进行反演制图。[结果] 反射率经微分处理后可有效改善其与土壤 pH 值的相关性；反射率二阶微分的随机森林模型为所有模型中的最优模型，其验证集的 R^2 为 0.87，RMSE 为 0.04，RPD 为 2.53；最优模型反演的 pH 值数据插值所得数字图与实测 pH 值插值图的空间分布特征高度吻合，能客观反映土壤碱化的空间分布状况。[结论] 随机森林模型为原位反演南疆棉田土壤 pH 值的最优模型，克里金插值能够客观可视化表达研究区土壤 pH 值的分布状况。

关键词：土壤 pH 值；原位高光谱；随机森林；数字制图；棉田；新疆阿拉尔市
文献标识码：B **文章编号**：1000-288X(2021)04-0189-07 **中图分类号**：S153.4

文献参数：蔡海辉，彭杰，柳维扬，等.基于棉田原位高光谱数据的土壤 pH 值反演与制图研究[J].水土保持通报，2021，41(4)：189-195. DOI：10.13961/j.cnki.stbctb.2021.04.027；Cai Haihui，Peng Jie，Liu Weiyang，et al. Inversion and mapping of soil pH valve based on in-situ hyperspectral data in cotton field [J]. Bulletin of Soil and Water Conservation，2021，41(4)：189-195.

Inversion and Mapping of Soil pH Valve Based on In-situ Hyperspectral Data in Cotton field

Cai Haihui，Peng Jie，Liu Weiyang，Luo Defang，Wang Yuzhen，Bai Jianduo，Bai Zijin
(*College of Plant Science，Tarim University，Alaer，Xinjiang* 843300，*China*)

Abstract：[Objective] Rapid and accurate monitoring of farmland soil pH value were explored for large-scale soil improvement and achieving fine management of farmland. [Methods] The cotton fileds of the 12 th regiment at Alar City in the South of Xinjiang Uygur Autonomous Region were selected as the study area，in-situ hyperspectral data of 231 sample points were collected by grid sampling method，and soil samples at 116 sampling points were collected simultaneously. The correlation between in-situ hyperspectral reflectance data after different pretreatment modes and soil pH value was analyzed. Partial least squares regression，support vector machine regression and random forest were used to establish the hyperspectral inversion model of soil pH，respectively. According to the model evaluation indexes，the optimal model was selected and used for inversion and mapping of the pH value of the uncollected soil sample points. [Results] The reflectance after the differentia treatment could effectively improve its correlation with soil pH value. The random forest model with second-order derivative of reflectance was the optimal model among all models with R^2 of 0.87，RMSE of 0.04，and RPD of 2.53. The digital map interpolated by the pH value of optimal model inversion

收稿日期：2020-12-01 **修回日期**：2021-05-05
资助项目：国家重点研发计划项目“土壤综合观测与智能服务平台研发与应用”(2018YFE0107000)；国家自然科学基金项目“盐分对南疆土壤有机质高光谱特征与定量反演的影响及方法”(41361048)；兵团南疆重点产业创新发展支撑计划项目“北斗导航南疆枣园精细施肥关键技术与装备研发”(2020DB003)
第一作者：蔡海辉(1996—)，男(汉族)，河南省开封市人，硕士研究生，研究方向为土壤近地传感技术与制图。Email：tdzkychh@163.com。
通讯作者：彭杰(1977—)，男(汉族)，湖南省邵阳市人，博士，教授，主要从事农业资源高光谱遥感监测与评价方面的研究 Email：pjzky@163.com。

was highly consistent with the spatial distribution characteristics of the actual measurement pH value, which could objectively reflect the spatial distribution of soil alkalinization. [Conclusion] The random forest model is the optimal model for in-situ inversion of soil pH value in cotton fields in South Xinjiang Uygur Autonomous Region, and Kriging interpolation could objectively visualize the soil pH value distribution in the study area.

Keywords: soil pH value; in-situ hyperspectral; random forest; digital mapping; cotton field; Alar City, Xingjiang Wei Autonomous Region

由于受农业灌溉、蒸降比、地下水矿化度等因素影响[1],新疆土壤碱化状况不容乐观。2015 年中国地质调查局资料表明,近 40 a 来,中国有近 2.67×10^7 hm^2 耕地土壤碱化状况加剧,约占耕地面积的 23%,土壤 pH 值上升了 0.6 左右。pH 值是土壤酸碱化的重要评价指标,土壤碱性越强所对应的 pH 值越大,土壤越容易表现出板结、透气性差等现象,从而降低植物根系对养分的吸收效率[2-4],影响植物的生长发育。为更好防治农田土壤碱化,提高农田作物产量、提升作物品质,需要对农田中土壤 pH 值的分布状况进行持续监测。但传统野外采样并进行室内化学分析的方法不能实时、快速获取土壤 pH 值在农田中的分布信息,随着高光谱技术的不断发展,可通过建立土壤 pH 值与土壤反射率反演模型实现土壤 pH 值信息的快速准确获取[5],为农田土壤改良、精准农业发展提供技术支撑[6]。

目前国内外学者在使用高光谱数据预测土壤属性方面有大量研究成果,如李诗朦等[7]使用室内光谱数据预测呼伦贝尔草原的电导率和 pH 值,使用支持向量机回归建模 R^2 大于 0.90,RPD 大于 3.00,具有较好的预测能力;徐驰等[8]研究发现使用室内高光谱数据反演内蒙古河套灌区的地表土(0—5 cm)的 pH 值与含盐量是可行的,模型预测效果良好,R^2 在 0.95 以上;Miles 等[9]通过使用室内光谱数据反演土壤中的速效磷与 pH 值后,认为未来光谱分析可能会取代传统化学分析对土壤属性信息进行获取。虽然室内光谱可以减少土壤水分、土壤质地和一些环境因素对光谱测量与土壤属性预测精度的影响[7],但由于其对野外采集的土壤样品进行预处理后才能测定土壤光谱反射率,需要耗费一定时间因而不能对土壤属性进行快速实时监测,因此一些学者对使用野外实测光谱数据预测土壤属性也表现出了浓厚的兴趣,如翟茂彤[10]、乔娟峰等[11]利用野外光谱数据成功反演了鄱阳湖和阜康地区的有机质含量;贾科利等[12]使用偏最小二乘回归建立土壤 pH 值预测模型的拟合度 R^2 达 0.93,王凯龙等[13]建立的土壤 pH 值预测模型 R^2 为 0.90,RPD 为 2.65,他们使用野外实测光谱数据都较好预测了当地土壤的 pH 值。这些研究都证实了采用野外实测光谱预测土壤属性是可行的。

数字土壤图通常利用栅格的方式来详尽表达土壤属性的空间变化,是一种新兴的、有效的表达土壤属性空间信息的方法[14-15],如廖琪等[16]、Brian 等等[17]使用克里格插值方法分别获取了其研究区的土壤有机质、pH 值空间分布图,但利用田间原位高光谱数据反演土壤 pH 值并制图的研究还较为少见。基于以上分析,本文以南疆阿拉尔市十二团棉田土壤为研究目标,通过野外原位间隔采集光谱数据、土壤样品与室内化学分析,结合偏最小二乘回归、支持向量机回归、随机森林 3 种建模方法,筛选最优模型反演研究区土壤 pH 值并使用普通克里格插值制作土壤 pH 值分布图,为南疆土壤碱化研究与治理提供理论依据。

1 材料与方法

1.1 研究区概况

研究区位于新疆阿拉尔市十二团,其中心地理坐标为东经 81°19′05″,北纬 40°29′20″,地处南疆中部,紧邻塔克拉玛干沙漠边缘,位于塔里木河上游。年均降水量约为 47 mm,年均蒸发量约为 1 988 mm,年平均气温为 10.8 ℃,年均日照时长约为 2 700 h,属典型的大陆性干旱气候。研究区内土壤以砂土和砂壤土为主,依靠膜下滴灌提供作物生育期内所需水分,每年冬春季节都要进行大水漫灌。研究区地势西高东低,灌溉入水口在地势较高处。在阿拉尔市十二团选取能代表当地管理水平且具有一定程度碱化的棉田作为研究区。

1.2 土壤样品采集与测定

土壤样品采集于 2018 年 11 月 3—4 日,采样点远离路边、地边,采用网格布点法均匀分布于研究区内,样点间隔为 20 m,去除表面覆盖的植物残留物、石块等影响光谱反射率的杂物后,采集较为平整、没有明显水分聚集的土壤表面光谱反射率信息。每行采集 21 个样点原位高光谱数据并间隔采集 0—20 cm 深度的棉田土壤,在 8 hm^2 面积内采集 11 行共计 231 个样点原位高光谱数据及其经纬度信息,并同步采集其中 116 个样点的土壤样品,建立原位高光谱数据和土壤 pH 值的反演模型,并利用最优反演模型预测未采集土壤样品的 115 个样点的 pH 值,结合实际测定的土壤 pH 值进行插值制作研究区土壤 pH 值分布图。

将采集所得土样带回实验室风干研磨后过2 mm筛,使用土水比为 1∶2.5 土壤浸提液测定土壤 pH 值[18]。

1.3　原位光谱数据采集与预处理

使用美国 SR-3500 型地物光谱仪采集光谱数据,其波长范围为 350～2 500 nm,在波长区间分别为 350～1 000 nm,1 000～1 900 nm 和 1 900～2 500 nm 时,光谱分辨率分别为 3.5,10 和 7 nm,数据重采样间隔为 1 nm。使用具有内置光源的光纤手柄采集土壤原位光谱数据,将土壤表面的植物残体、残膜、石块等杂物清除干净后,仪器探头紧贴地面测量,每个样点重复测量 10 次,取其算术平均值作为该样点原位光谱数据。每个样点在测量前都要清理手柄的镜面,每 10 个样点进行 1 次白板校准。为了便于描述,将此方法采集的光谱数据简称为"原位数据"。根据国内学者的研究成果去除噪声较大的 350～399 nm,2 401～2 500 nm 波段[19-20],采用 Savitzyk-Golay 平滑滤波去除光谱噪声后[21],使用 Excel 和 The Unscrambler X 10.5.1 软件对原始反射率(R)进行倒数($1/R$)、对数($\lg R$)、倒数对数 $\lg(1/R)$、面积归一化(AN, area normalization)、峰值归一化(MAN, maximum normalization)、多元散射校正(MSC, multiplicative scatter correction)、一阶微分(FDR, first derivative)、二阶微分(SDR, second derivative)等 9 种数据预处理。

1.4　建模方法及评价指标

将土壤 pH 值进行升序排列,取每 3 个相邻样本的中间样本的集合为验证集,其余三分之二为建模集,即 78 个样本用于建模,38 个样本用于验证。建模方法为偏最小二乘回归(PLSR, partial least squares regression)、支持向量机回归(SVMR, support vector regression)、随机森林(RF, random forest) 3 种,其中 PLSR 与 SVMR 建模在 The Unscrambler X 10.5.1 软件中实现,RF 建模在 Rstudio 中实现。

PLSR 集中了主成分分析、典型相关分析、线性回归分析的优点,解决了数据间多重共线性、样本数少于变量数的问题,因而成为了普遍使用的一种线性模型[22]。

SVMR 方法是一种通过一个非线性映射,把样本空间映射到一个高维的特征空间中,使得在样本空间中的非线性可分问题转化为在特征空间中的线性可分问题的机器学习算法[22]。SVMR 方法的关键在于核函数,结合相关文献[23-24],本研究选取 RBF 函数为其核函数,根据模型交叉验证最优效果及模型稳定性确定惩罚系数 C 和 Gamma 的值。

RF 是使用随机方式建立一个具有许多且没有关联的决策树的森林,从 N 个样本中随机选取 n 个样本用于构建回归树,当有样本输入时都要经过每棵分类决策树决策分类,投票最多的一类作为最终分类结果,在进行回归预测时,预测值为所有回归树输出结果的平均值。根据模型稳定性及较高预测效果,并经多次建模验证,确定决策树个数及其余参数设置。

模型稳定性及精度评价指标为决定系数(R^2)、均方根误差(RMSE)、相对分析误差(RPD)。针对 RPD 而言,当 RPD＜1.5 时模型无法对样品进行预测,当 1.5≤RPD＜2.0 时表明模型只能粗略估测样品中高含量和低含量部分,当 2.0≤RPD＜2.5 时表明模型具有较好的预测能力,当 2.5≤RPD 时模型具有很好的预测能力[21]。选择 R^2 和 RPD 大、RMSE 小的模型作为最优模型,进行进一步研究。

2　结果与分析

2.1　土壤 pH 值描述性统计

表 1 为建模集与验证集土壤 pH 值数据统计特征,由表 1 可知,116 个供试样本 pH 值的最大值为 8.09,最小值为 7.48,平均值为 7.77,标准差 0.12,变异系数为 1.54%,根据新疆土壤分析[25]中 pH 值分级标准,6.50≤pH＜7.50 为中性;7.50≤pH＜8.50 为碱性;pH 值≥8.50 为强碱性,变异系数是衡量数据间变异程度的统计量,其可以用于表示样品的离散程度,当 C_v＜10%时为弱变异性;当 10%≤C_v≤100%时为中等变异性;当 100%＜C_v 时为强变异性[26]。研究区土壤 pH 值介于 7.50～8.50 之间,变异系数为 1.54%,为弱变异性的碱性土壤。

表 1　建模集与验证集土壤 pH 值数据统计

数据集	样本数	最小值	最大值	平均值	标准差	变异系数/%
建模集	78	7.48	8.09	7.77	0.12	1.57
验证集	38	7.50	7.99	7.77	0.11	1.44
样 本	116	7.48	8.09	7.77	0.12	1.54

2.2　不同 pH 值的土壤光谱特征分析

图 1 为将 116 个土壤样品的 pH 值升序排列均分为 3 类后(分别为 7.48≤pH≤7.71,7.72≤pH≤7.82,7.83≤pH≤8.09)计算各类野外原位高光谱反射率数据平均值,所得到的土壤样品反射率曲线。从图 1 可以看出在 400～988 nm 波段内不同 pH 值的土样光谱反射率曲线差异较小;在 989～2 400 nm 波段范围内,相同波长条件下土壤 pH 值越大所对应的光谱反射率也越大,说明在 pH 值为 7.48～8.09 范围内土壤的光谱反射率与土壤 pH 值呈正相关,即在一

定 pH 值范围内，在同一波长下随着土壤 pH 值的增大，反射率也随之增大；在 1 450，1 940，2 200 nm 附近存在 3 个明显的特征吸收谷，参照相关学者研究结果，1 450 和 940 nm 是水分吸收谷[27-28]，2 200 nm 附近是用于判断黏土矿物存在的特征谷[29]，在这些吸收谷附近的土壤反射率明显降低，研究区土壤反射率最大值为 27.70%。土壤质地、土壤含水量、环境杂散光是导致土壤光谱反射率数据降低的主要影响因素[27]，这些因素也将影响光谱数据的建模和验证精度。

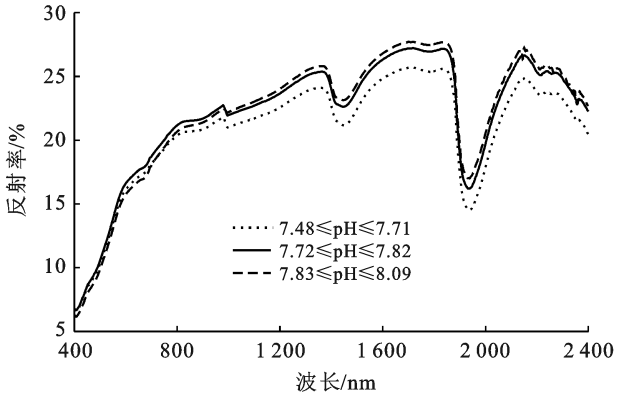


图 1 不同 pH 值范围土样反射率均值曲线

2.3 土壤 pH 与土壤光谱数据相关分析

经建模结果验证，选择建模及验证效果较好的 R，lg$R$，$1/R$，FDR，SDR 5 种数据形式进行后续分析。图 2 为在波长 400～2 400 nm 间，将土壤 pH 值与原位高光谱数据及其 4 种数据变换形式进行相关分析所得的相关系数曲线。由图 2 可知土壤 pH 值与 R，lg$R$ 在 400～745 nm 呈负相关关系，与 $1/R$ 成正相关关系，而在 746～2 400 nm 波段则相反，R，lg$R$ 与土壤 pH 值的相关性系数曲线几乎一致。R，lg$R$，$1/R$ 的最大相关系数均位于 1 965 nm 处，分别为 0.25，0.24 与 −0.23，FDR，SDR 的最大相关系数分别位于 1 942 nm，636 nm 处，分别为 0.45，−0.43。土壤 pH 值与 R，$1/R$，lg$R$，FDR，SDR 的相关系数达显著性的波段总数分别为 421，395，420，819，670 个，FDR，SDR 变换可以大幅度提高土壤 pH 值与光谱数据相关性，但由于其与土壤 pH 值的相关性缺乏规律，需要对光谱进行较为详细的分类，充分挖掘光谱信息才能取得好的建模及预测效果。原位高光谱数据经倒数、对数转换后对于改善光谱与土壤 pH 值相关性作用不大，对反射率进行一阶微分、二阶微分转换可以明显提高光谱与土壤 pH 值的相关性和相关系数达显著性的波段总数，其相关系数最大值为 0.45 比原始反射率提高了 0.20 左右，但相对于一阶微分变换，二阶微分的相关系数分布更为均匀，使用二阶微分建模可能会取得更好的建模效果。

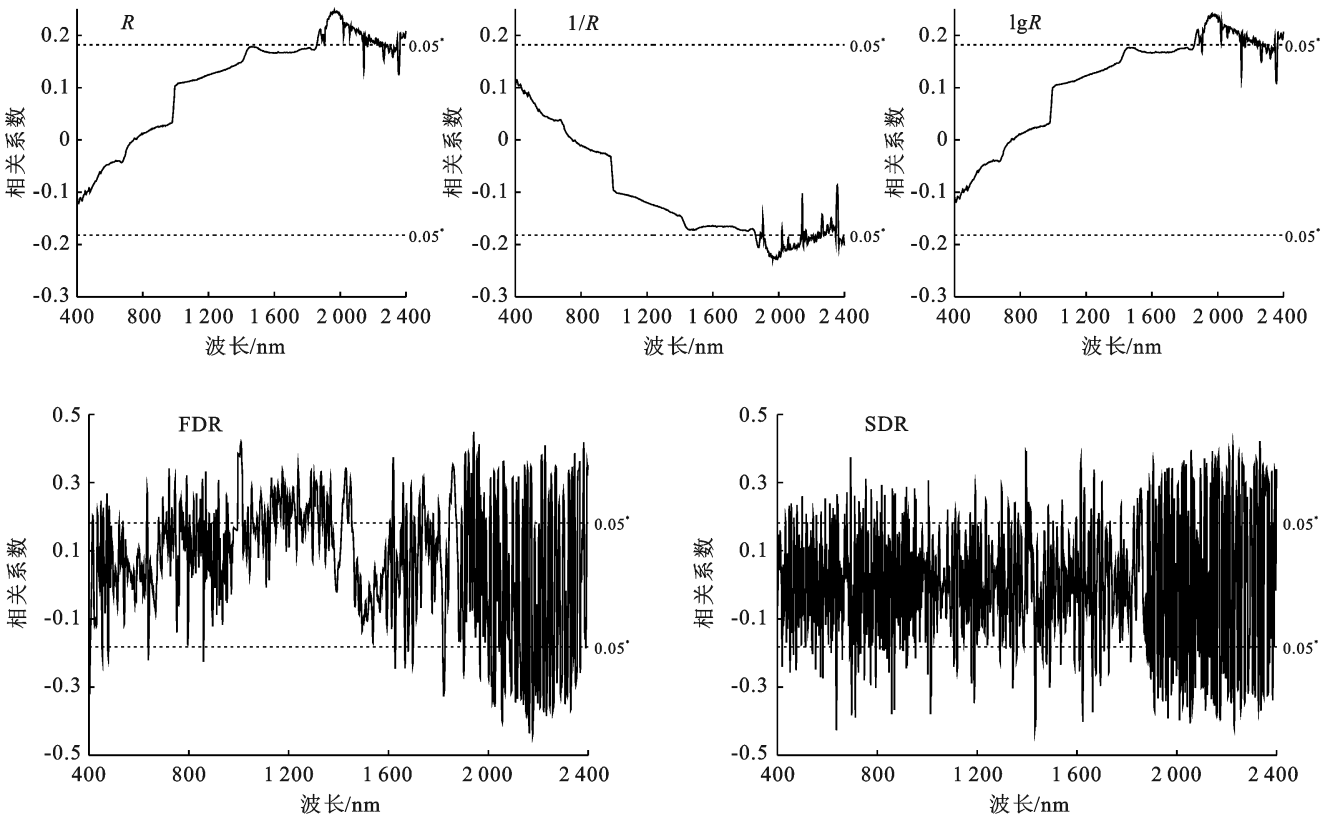


图 2 土壤 pH 值和光谱反射率不同转换形式的相关系数

2.4　土壤 pH 值反演模型的建立与验证

以土壤 pH 值为因变量,所对应原位高光谱反射率数据为自变量,采用 3 种建模方法结合 5 种数据形式建立反演模型。建模集与验证集 pH 值数据详见表 1。表 2 为建立的 3 种模型的精度对比统计。由表 2 可知,微分数据在支持向量机回归建模,随机森林建模中均取得了不错的效果,但在偏最小二乘回归建模时的建模和预测效果较差。相较于微分数据,3 种建模方法在使用反射率倒数、对数建模时的建模与验证效果都不太理想。偏最小二乘回归建模中反射率倒数的建模效果最好,建模集的 R^2 为 0.25,RMSE 为 0.11,验证集 R^2 为 0.43,RMSE 为 0.09,RPD 为 1.26,支持向量机回归建模中一阶微分的建模效果最好,建模集 R^2 为 0.75,RMSE 为 0.08,验证集 R^2 为 0.36,RMSE 为 0.09,RPD 为 1.22,但这两者验证集的 R^2 和 RPD 值较小、RPD 均小于 1.4,无法对样本进行预测。随机森林中最优反演模型的建模集 R^2 为 0.86,RMSE 为 0.04,验证集 R^2 为 0.87,RMSE 为 0.04,RPD 为 2.53 (2.5≤RPD<3.0)模型具有极好的预测能力。相较于 PLSR 模型、SVMR 模型,RF 模型的精度,预测能力均最好,因此选择 RF 模型为最优模型,对未采样的样点进行反演与制图研究。

2.5　土壤 pH 值空间分布数字制图

基于对前文的分析,选取随机森林模型作为反演模型,对只采集原位光谱信息而未采集土壤样品的 115个样点进行反演并制作研究区土壤 pH 值分布图。为进一步验证田间原位光谱的建模精度,将未采集土壤样品样点的原位高光谱数据进行 Savitzyk-Golay 平滑和 SDR 数据处理后使用随机森林模型进行反演土壤 pH 值。克里格插值是地统计学的主要内容之一,同时也是估计未采样位置属性值的最优无偏估计方法,是土壤属性制图中广泛使用的一种插值方法。图 3 为使用随机森林模型模拟的土壤 pH 值数据和实测土壤 pH 值数据分别进行普通克里格插值得到的研究区土壤 pH 值插值图。

表 2　不同土壤 pH 值模型精度对比

建模方法	数据集	建模		验证		
		R^2	RMSE	R^2	RMSE	RPD
偏最小二乘回归(PLSR)	R	0.29	0.11	0.28	0.10	1.12
	$1/R$	0.25	0.11	0.43	0.09	1.26
	lg$R$	0.31	0.10	0.41	0.09	1.21
	FDR	0.25	0.11	0.03	0.11	1.00
	SDR	0.14	0.12	0.09	0.12	0.92
支持向量机回归(SVMR)	R	0.65	0.10	0.22	0.10	1.10
	$1/R$	0.63	0.10	0.28	0.12	0.94
	lg$R$	0.60	0.10	0.26	0.10	1.09
	FDR	0.75	0.08	0.36	0.09	1.22
	SDR	0.73	0.07	0.28	0.10	1.13
随机森林(RF)	R	0.84	0.06	0.80	0.07	2.11
	$1/R$	0.82	0.05	0.81	0.05	2.23
	lg$R$	0.84	0.06	0.80	0.07	2.30
	FDR	0.86	0.04	0.86	0.05	2.45
	SDR	0.86	0.04	0.87	0.04	2.53

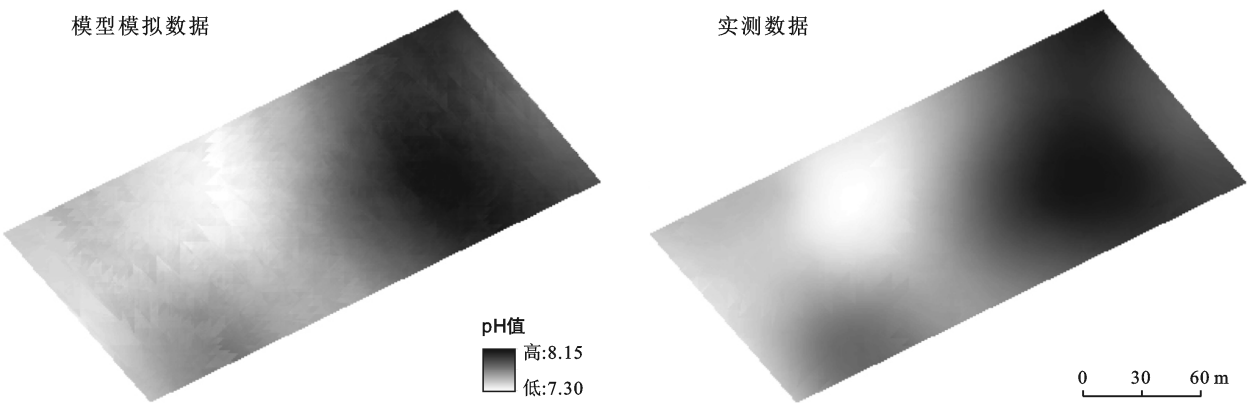


图 3　研究区土壤 pH 值模型模拟插值与实测数据对比

由图 3 可知,模型模拟的土壤 pH 值数据插值图与实测土壤 pH 值数据插值图中土壤 pH 值的分布特征高度吻合,都表现为土壤 pH 值在东西方向上总体呈现逐步减小趋势,南北方向上呈现逐步增加趋势。研究区内土壤 pH 值分布状态为高值部分呈片状集中在研究区东部,低值部分成片集中在研究区的中西部,低值地带和高值地带间有较为明显的缓冲带,出现这种规律可能是因为研究区地势西高东低,在进行冬灌、春灌洗去土壤盐分时,使得土壤中交换性 Na^+ 随水流原因被土壤胶体吸附导致东北区域土壤碱性增强,而表现出土壤 pH 值增加所造成的。研究区选取的是最接近当地生产管理水平的田块,属于轻微碱化,说明该地区土壤可能存在不同程度的碱化现象,因此研究该地区的土壤 pH 值的空间分布信

息，对于精准改良当地的土壤酸碱度，提升作物产量有重要意义。对使用实测 pH 值与模型模拟 pH 值得到的插值图进行预测误差分析，结果见表 3。预测误差、均方根误差、平均标准预测误差越小说明预测效果越好，均方根预测误差越接近于 1 越好[28]。由表 3 可知，实测 pH 值预测误差与模拟 pH 值预测误差相近，都达到了较好的插值效果，说明使用野外原位光谱预测土壤 pH 值是可行的。

表 3 普通克里格插值预测土壤 pH 值误差分析

指 标	预测误差	均方根误差	平均标准预测误差	均方根预测误差
实测 pH 值数据	0.000 1	0.194 4	0.093 5	0.800 1
模型模拟 pH 值数据	−0.000 1	0.173 6	0.073 8	0.831 1

3 讨 论

土壤 pH 值是衡量土壤酸碱化程度最重要的指标之一，使用田间原位高光谱数据对土壤 pH 值进行反演并制作土壤 pH 值分布图，可为土壤养分利用、实现精准农业的实时快速检测提供一定的科技支撑[15,30]。文中对不同数据转换形式的光谱数据与土壤 pH 值的相关系数进行比较，结果表明反射率数据经微分变换后与土壤 pH 值的相关性明显提高，能为建立预测模型提供更多信息，这也与李阳等[19]、李诗朦等[7]、魏雨露等[31]研究结果一致。微分数据在 SVMR，RF 建模中都取得了不错的结果也验证了这一点，但微分数据会放大噪声的干扰，同时也易引进无关因素进而影响建模精度[32]，如张芳等[5]、彭杰等[33]使用微分数据建模的效果不如实测光谱。本文使用 RF 模型预测土壤 pH 值取得了较好的效果，同郭鹏等[34]、张振华等[35]研究结果一致。不少研究证实 RF 在噪声较大、数据量少时仍能建立准确且可靠的模型，其具有处理定量和定性数据的能力，在数字土壤制图中能够发挥出巨大的潜力[36-37]。正是由于随机森林具有较强的抗拟合能力及其对高维数据极强的处理能力[38]，本文在使用其与原位高光谱数据的微分建立的反演模型时取得了不错的结果。PLSR 在处理线性回归时有着独特的优势同时还能解决数据间的多重共线性问题，但对于非线性数据的拟合能力较低，SVMR 适用于小样本的非线性数据建模，但当数据集的噪声过大甚至是成为支持向量时使用 SVMR 预测可能会取得较差的结果[39]，因原位高光谱数据量大且含有土壤水分信息与其他干扰因素，所以本文使用 PLSR，SVMR 建模效果不如 RF 模型。与室内光谱反演土壤属性数据相比，田间原位光谱可以节省室内测定时间，得到更接近实际的土壤属性数据，实现土壤属性的快速、实时检测。原位光谱数据易受环境因素、土壤类型等因素影响，且土壤 pH 值在近红外波段属于间接预测，本文建立的预测模型精度不如李诗朦等[7]、魏雨露等[31]的高。建立的模型受地域、土壤类型等影响较大，在不同的地区应用需建立相应的原位光谱库，进而为精准农田的实现提供一定的技术支持。

4 结 论

本文采用随机森林算法建立新疆阿拉尔市十二团研究区土壤 pH 值的原位高光谱反演模型并利用普通克里格插值方法制作研究区土壤 pH 值插值图，着重讨论了 4 不同数据预处理方式结合 3 种建模方法对于提高建模及验证精度的帮助，确定了使用二阶微分数据预处理后的随机森林反演模型为最优模型。

(1) 原位高光谱反射率数据与土壤 pH 值在 746～2 400 nm 波段呈正相关关系、400～745 nm 波段呈负相关关系，在 1 856～2 400 nm 相关关系达显著性水平，相关系数在 1 965 nm 处达最大值为 0.25。

(2) 相较于偏最小二乘(PLSR)、支持向量机(SVMR)建模，使用 5 种数据形式进行随机森林建模都能取得好的建模及验证效果(RPD 均大于 2.00)。对于不同数据变换来说反射率数据经一阶、二阶微分变换后使用支持向量机回归(SVMR)、随机森林(RF)建模都可以取得较好的建模效果，但支持向量机的模型验证效果较差(R^2 为 0.36，RPD 为 1.22，RPD 小于 1.40 无法预测土壤 pH 值)，随机森林模型验证效果较好 R^2 达 0.87，RPD 达 2.53，能极好的反演未采样点的土壤 pH 值。

(3) 分别利用模型模拟数据和实测数据进行普通克里格插值制作研究区土壤 pH 值插值图，模型模拟数据插值图与实测数据插值图中土壤 pH 值的分布特征相一致，说明田间原位高光谱测量手段可以实现土壤 pH 值空间分布信息的实时快速获取，同时也能为农田碱化防治提供一定依据。

[参 考 文 献]

[1] 李保强，陈绍荣，邵建华，等.新疆土壤盐渍化的综合治理与改良[J].新疆农业科技，2013(6)：15-21.

[2] 林毅，梁颂捷，朱其清.三明烟区土壤 pH 值与土壤有效养分的相关性[J].烟草科技，2003(6)：35-37.

[3] Filippi P, Cattle S R, Bishop T F A, et al. Digital soil monitoring of top-and sub-soil pH with bivariate linear mixed models [J]. Geoderma, 2018, 322: 149-162.

[4] 张维，李启权，王昌全，等.川中丘陵县域土壤 pH 空间变

异及影响因素分析:以四川仁寿县为例[J].长江流域资源与环境,2015,24(7):1192-1199.

[5] 张芳,熊黑钢,栾福明,等.土壤碱化的实测光谱响应特征[J].红外与毫米波学报,2011,30(1):55-60.

[6] 文虎,盛建东,颜安,等.绿洲农田盐碱斑土壤光谱特征分析与建模[J].新疆农业大学学报,2016,39(2):143-148.

[7] 李诗朦,包妮沙,刘善军,等.土壤电导率和 pH 值光谱特征及反演模型:以呼伦贝尔草原干旱半干旱土壤为例[J].测绘科学,2018,43(8):14-22,44.

[8] 徐驰,曾文治,伍靖伟,等.内蒙古河套灌区土壤含盐量和 pH 高光谱反演研究[J].灌溉排水学报,2013,32(3):39-43.

[9] Miles Grafton, Therese Kaul, Alan P, et al. Technical note: Regression analysis of proximal hyperspectral data to predict soil pH and olsen P [J]. Agriculture, 2019,9(3):1-18.

[10] 翟茂彤.基于原位 Vis-NIR 高光谱的鄱阳湖湿地土壤有机碳预测研究[D].江西 南昌:江西财经大学,2020.

[11] 乔娟峰,熊黑钢,王小平,等.新疆阜康荒地土壤有机质高光谱特征及其反演模型研究[J].干旱地区农业研究,2018,36(5):207-214.

[12] 贾科利,张俊华,秦君琴.典型龟裂碱土光谱特征分析及碱化程度预测[J].干旱地区农业研究,2013,31(4):187-192,199.

[13] 王凯龙,熊黑钢,张芳.基于高光谱数据预测土壤碱化程度最佳模型及其影响因素的研究[J].土壤,2014,46(3):544-549.

[14] 周宏冀.江西省耕地土壤 pH 空间变异与管理分区研究[D].江西 南昌:江西财经大学,2019.

[15] 张淑杰,朱阿兴,刘京,等.基于样点的数字土壤属性制图方法及样点设计综述[J].土壤,2012,44(6):917-923.

[16] 廖琪,胡月明,胡小飞,等.广东省典型赤红壤区耕层土壤养分的空间变异[J].水土保持通报,2015,35(6):322-328.

[17] Odhiambo B O, Kenduiywo B K, Were K. Spatial prediction and mapping of soil pH across a tropical afromontane landscape [J]. Applied Geography, 2020, 114: 102-129.

[18] 鲍士旦.土壤农化分析[M].3 版.北京:中国农业出版社,2013.

[19] 李阳,刘新路,彭杰,等.基于可见光近红外光谱的南疆荒漠土壤有机质反演研究[J].土壤通报,2018,49(4):767-772.

[20] 代希君,彭杰,张艳丽,等.基于光谱分类的土壤盐分含量预测[J].土壤学报,2016,53(4):909-918.

[21] 彭杰,刘焕军,史舟,等.盐渍化土壤光谱特征的区域异质性及盐分反演[J].农业工程学报,2014,30(17):167-174.

[22] 史舟.土壤地面高光谱遥感原理与方法[M].北京:科学出版社,2014.

[23] 马利芳,熊黑钢,张芳.基于野外 Vis-NIR 光谱的土壤盐分主要离子预测[J].土壤,2020,52(1):188-194.

[24] 尚天浩,贾萍萍,孙媛,等.宁夏银北地区盐碱化土壤水分光谱特征及模型拟合精度分析[J].水土保持通报,2020,40(4):183-189.

[25] 李培清.新疆土壤分析[M].新疆 乌鲁木齐:新疆人民出版社,1983.

[26] 雷志栋,杨诗秀,谢森传.土壤水动力学[M].北京:清华大学出版社,1988.

[27] 纪文君.基于野外 Vis-NIR 高光谱的土壤属性预测及田间水分影响去除研究[D].浙江 杭州:浙江大学,2014.

[28] 郭燕,纪文君,吴宏海,等.基于野外 Vis-NIR 光谱的土壤有机质预测与制图[J].光谱学与光谱分析,2013,33(4):1135-1140.

[29] Rossel R A V, Cattle S R, Ortega A, et al. In situ measurements of soil colour, mineral composition and clay content by Vis-NIR spectroscopy [J]. Geoderma, 2009, 150(3/4): 253-266.

[30] 杨忠芳,陈岳龙,钱鑫,等.土壤 pH 对镉存在形态影响的模拟试验研究[J].地学前缘,2005,12(1):252-260.

[31] 魏雨露,刘金宝,李劲彬.基于 PLS 的陕西关中地区土壤 pH 高光谱预测[J].西部大开发(土地开发工程研究),2018,3(4):51-57.

[32] 刘炜,常庆瑞,郭曼,等.小波变换在土壤有机质含量可见/近红外光谱分析中的应用[J].干旱地区农业研究,2010,28(5):241-246.

[33] 彭杰,向红英,王家强,等.基于野外实测高光谱数据的干旱区耕作土壤含水量反演研究[J].干旱地区农业研究,2013,31(2):241-246.

[34] 郭鹏,李华,陈红艳,等.基于光谱指数优选的土壤盐分定量光谱估测[J].水土保持通报,2018,38(3):193-199,205.

[35] 张振华,丁建丽,王敬哲,等.集成土壤—环境关系与机器学习的干旱区土壤属性数字制图[J].中国农业科学,2020,53(3):563-573.

[36] Dharumarajan S, Hegde R, Singh S K. Spatial prediction of major soil properties using Random Forest techniques: A case study in semi-arid tropics of South India [J]. Geoderma Regional, 2017,10:154-162.

[37] Vågen T G, Winowiecki L A, Tondoh J E, et al. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance [J]. Geoderma, 2016, 263: 216-225.

[38] 马晓东.基于加权决策树的随机森林模型优化[D].湖北武汉:华中师范大学,2017.

[39] 刘亚秋,陈红艳,王瑞燕,等.基于可见/近红外光谱的黄河口区土壤盐分及其主要离子的定量分析[J].中国农业科学,2016,49(10):1925-1935.