

## 高光谱成像的褐土土壤速效钾含量预测

王文俊<sup>1</sup>, 李志伟<sup>1</sup>, 王 璨<sup>1</sup>, 郑德聪<sup>1\*</sup>, 杜慧玲<sup>2\*</sup>

1. 山西农业大学工学院, 山西 太谷 030801

2. 山西农业大学文理学院, 山西 太谷 030801

**摘要** 精细农业变量施肥取决于对农田的土壤养分分布的了解, 快速获取土壤信息是实施精细农业的基础。速效钾是土壤肥力的重要参数, 是植物生长发育所必需的营养元素。对土壤速效钾含量进行测量, 是了解土壤肥力的重要途径, 是实现精细农业的必要条件。以山西典型褐土土壤为研究对象, 采集农田耕层褐土土壤样品共 169 份, 样品经风干处理, 手动捏碎较大的土粒并去除杂质后, 未经研磨过筛处理而直接用于土壤近红外高光谱的测量。根据实验室速效钾含量测定结果, 将所有土壤样品分为两类: 其中速效钾含量低于  $100 \text{ mg} \cdot \text{kg}^{-1}$  的样品共 144 个, 随机选取 108 个作为低含量建模集( $L_c$ ), 剩余 36 个作为低含量验证集( $L_p$ ); 速效钾含量高于  $100 \text{ mg} \cdot \text{kg}^{-1}$  的样品共 25 个, 随机选取 19 个作为高含量建模集( $H_c$ ), 剩余 6 个作为高含量验证集( $H_p$ )。其中  $L_c$  和  $H_c$  统称为所有含量建模集( $T_c$ ),  $L_p$  和  $H_p$  统称为所有含量验证集( $T_p$ )。获取所有土壤样本  $950 \sim 1650 \text{ nm}$  范围内的近红外高光谱图像。分别采用平均光谱曲线(R)、平均光谱曲线的一阶导数(FD)、平均光谱曲线与一阶导数共同建模( $R \& FD$ )、平均光谱曲线与一阶导数的乘积( $R * FD$ )、平均光谱曲线与一阶导数的商( $R / FD$ )等五种光谱数据预处理方法, 结合偏最小二乘法(PLS), 分别对建模集  $T_c$ 、 $L_c$  及  $H_c$  建模, 然后分别对验证集  $T_p$ 、 $L_p$  及  $H_p$  进行验证。结果表明: 土壤的平均光谱反射率随速效钾含量的增大呈现先增加后减小的趋势。当速效钾含量低于  $100 \text{ mg} \cdot \text{kg}^{-1}$  时, 所有波段的光谱反射率随速效钾含量的增加而增加; 当速效钾含量在  $100 \sim 200 \text{ mg} \cdot \text{kg}^{-1}$  之间时, 所有波段的光谱反射率均达到最大值。当速效钾含量超过  $200 \text{ mg} \cdot \text{kg}^{-1}$  时,  $950 \sim 1400 \text{ nm}$  的光谱反射率急剧减小, 但曲线的整体斜率显著增加; 且速效钾含量越高, 曲线整体斜率越大。当速效钾含量高于  $100 \text{ mg} \cdot \text{kg}^{-1}$  时, 平均光谱曲线的一阶导数显著增大, 且随速效钾含量的增加而增加。该研究建立的 PLS 模型, 可以对整体(所有速效钾含量)和高含量( $\geq 100 \text{ mg} \cdot \text{kg}^{-1}$ )速效钾进行有效预测, 但无法对低含量( $\leq 100 \text{ mg} \cdot \text{kg}^{-1}$ )速效钾进行预测。建模效果最好的光谱预处理方法为  $R * FD$ , 其次为 FD, R, 而  $R \& FD$ ,  $R / FD$  预测效果相对较差。最优建模方式为:  $R * FD$  结合  $T_c$  建模, 其 PLS 主因子个数为 2 个,  $\text{RMSE}_c = 29.293$ ,  $\text{RPD}_c = 4.669$ ,  $R_c^2 = 0.956$ ; 对  $T_p$  的验证效果为  $\text{RMSE}_p = 29.438$ ,  $\text{RPD}_p = 4.740$ ,  $R_p^2 = 0.958$ ; 对  $H_p$  的验证效果为  $\text{RMSE}_p = 23.033$ ,  $\text{RPD}_p = 3.199$ ,  $R_p^2 = 0.915$ 。该模型能够根据土壤速效钾的含量对土壤进行分类: 当预测值小于  $100 \text{ mg} \cdot \text{kg}^{-1}$  时, 表明土壤速效钾含量低于  $100 \text{ mg} \cdot \text{kg}^{-1}$ , 具体含量不确定; 当预测值大于  $100 \text{ mg} \cdot \text{kg}^{-1}$  时, 预测值则能够很好反映土壤速效钾的真实含量。由于选用的土壤样本未经研磨和过筛处理, 因而能够大大缩短样本制备时间, 提高预测效率。该研究结果可为近红外高光谱成像应用于褐土土壤除速效钾含量以外其他营养成分的快速预测提供参考。

**关键词** 近红外; 高光谱成像; 速效钾; 偏最小二乘

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)05-1579-07

收稿日期: 2018-03-30, 修订日期: 2018-08-12

基金项目: 国家重点研发计划项目(2017YFD0701501), 山西省自然科学基金面上项目(201701D121099), 中国博士后基金项目资助

作者简介: 王文俊, 1986 年生, 山西农业大学工学院博士后 e-mail: wwjbuoy03@163.com

\* 通讯联系人 e-mail: zhengdecong@126.com; duhuiling66@163.com

## 引言

精细农业变量施肥取决于对农田的土壤养分分布的了解,快速获取土壤信息是实施精细农业的基础<sup>[1]</sup>。速效钾是土壤肥力的重要参数,是植物生长发育所必需的营养元素。对土壤速效钾含量进行测量,是了解土壤肥力的重要途径,是实现精细农业的必要条件。传统土壤养分的测量方法为化学分析法,该方法费时、费力,且成本较高。近红外光谱分析具有快速、高效、无损和适合在线分析等优点,采用近红外光谱获取土壤养分信息已成为国内外学者的研究热点<sup>[2-11]</sup>。

刘燕德<sup>[7]</sup>等采用近红外漫反射技术对赣南脐橙果园的土壤进行研究,分别采用偏最小二乘(PLS)、主成分回归(PCR)和最小二乘支持向量机(LS-SVM)等 3 种方法,建立赣南脐橙果园土壤全钾快速定量检测模型。结果表明:PLS、PCR 和 LS-SVM 建立的模型对全钾的预测效果均较为理想。刘雪梅等<sup>[8]</sup>采用可见-近红外光谱预测土壤中速效氮和速效钾的含量,结果表明:LS-SVM 对速效磷和速效钾的预测效果要好于 PLS。对速效钾的最佳预测效果为:决定系数  $R^2$  为 0.72,预测均方根误差 RMSE<sub>p</sub> 为 15.0。对原始光谱平滑后采用蒙特卡罗无信息变量消除方法(MC-UVE)结合 PLS 建立的速效钾预测模型要好于使用全谱和 PLS 建立的模型。王昶等<sup>[9]</sup>以长江中下游粮食主产区水稻土为研究对象,在 350~2 500 nm 范围,利用 PLS 建立了近红外漫反射光谱与多个土壤指标之间的定量分析模型。研究发现:PLS 对速效钾的预测效果较好,决定系数  $R^2$  为 0.82,相对分析误差 RPD<sub>p</sub> 为 2.53。胡国田等<sup>[10]</sup>利用可见-近红外光谱对土壤磷和钾含量监测存在精度不高的问题进行了研究,结果表明:采用直接正交信号校正(DOSC)能够较好消除由于土壤质地、类型等引起的噪声信息,提高模型预测精度。

目前,土壤营养元素的近红外高光谱反演主要集中在有机质<sup>[2-3]</sup>、氮<sup>[4-6]</sup>、水分等方面,而针对土壤速效钾含量的定

量预测仍然较少;同时,几乎所有实验室条件下的研究都要对土壤样本进行磨碎和过筛处理,而直接采集和利用原始土壤高光谱数据进行建模的极少;此外,针对土壤速效钾含量分段建立预测模型的较少。

本工作以山西典型褐土土壤为研究对象,所有土壤样品经风干后,未经研磨过筛处理而直接装入样品杯进行近红外高光谱图像的采集,并提取平均光谱曲线(R)。采用 5 种光谱数据前处理方法,结合偏最小二乘法(PLS)分别建立针对不同速效钾含量的预测模型。对比分析多种光谱数据前处理方法及对不同速效钾含量预测集的预测准确度,旨在构建一种快速检测土壤速效钾含量的定量预测模型,为快速获取土壤速效钾含量提供依据。

## 1 实验部分

### 1.1 样本准备及分类

实验用土壤样品采集自山西农业大学试验田内,采样面积约 67 000 m<sup>2</sup>,土壤类型均为褐土土壤。采样方法为规则栅格采样,采样栅格面积为 400 m<sup>2</sup>,栅格长度和宽度均为 20 m;采样深度为 0~0.2 m,即农田表层土壤。采样点个数为 169 个,共采集褐土土壤样本 169 份。所有样品经自然风干,手动捏碎较大的土粒并去除杂质后,采用四分法取样,一式两份:一份研磨过筛,用于实验室速效钾含量的测定,另一份直接用于土壤近红外高光谱图像的采集。

根据实验室速效钾含量测定结果,将所有土壤样品分为两类:其中速效钾含量低于 100 mg·kg<sup>-1</sup>的样品共 144 个,随机选取 108 个作为低含量建模集( $L_c$ ),剩余 36 个作为低含量验证集( $L_p$ );速效钾含量高于 100 mg·kg<sup>-1</sup>的样品共 25 个,随机选取 19 个作为高含量建模集( $H_c$ ),剩余 6 个作为高含量验证集( $H_p$ )。其中  $L_c$  和  $H_c$  统称为所有建模集( $T_c$ ), $L_p$  和  $H_p$  统称为所有验证集( $T_p$ )。土壤速效钾含量的特征统计见表 1。

表 1 土壤速效钾含量的特征统计

Table 1 Characteristic statistics of available potassium content in soil

速效钾含量 /(mg·kg <sup>-1</sup> )	样本分类	样本数	最小值 min /(mg·kg <sup>-1</sup> )	最大值 max /(mg·kg <sup>-1</sup> )	平均值 $E$ /(mg·kg <sup>-1</sup> )	标准差 $\sigma$ /(mg·kg <sup>-1</sup> )
所有	所有	169	23.81	678.57	95.17	139.48
	建模集 $T_c$	127	23.81	678.57	95.21	139.60
	验证集 $T_p$	42	27.78	654.76	95.05	140.82
≤100 低含量	所有	144	23.81	87.30	47.29	10.58
	建模集 $L_c$	108	23.81	87.30	47.10	10.14
	验证集 $L_p$	36	27.78	83.33	47.84	11.96
>100 高含量	所有	25	115.08	678.57	370.95	206.25
	建模集 $H_c$	19	115.08	678.57	368.63	207.39
	验证集 $H_p$	6	119.05	654.76	378.31	221.98

### 1.2 高光谱图像采集与校正

使用美国 Headwall Photonics 公司的 Starter Kit 室内移

动扫描平台进行近红外高光谱图像的采集。该系统由室内移动扫描平台、微型近红外高光谱成像仪(光圈 1.4, 焦距 25

mm)、光源、控制器和电脑等组成。高光谱图像数据的采集参数为:物距 320 mm,平台移动速度  $15.55 \text{ mm} \cdot \text{s}^{-1}$ ,曝光时间 0.9 ms,光谱范围 900~1 700 nm,入射狭缝宽度 30  $\mu\text{m}$ ,光谱分辨率 4.715 nm,共 170 个波段。由于在靠近光谱仪量程的两端( $<900 \text{ nm}$ ,  $>1 700 \text{ nm}$ ),光谱反射率会出现振荡,误差较大,因此选取用于建模的光谱范围为 950~1 650 nm,共 148 个波段。

土壤样本装入直径约 3 cm,深约 1 cm 的样品杯,抹平并压实,然后放在移动扫描平台上进行高光谱图像采集。每个土壤样本均采集 1 幅高光谱图像。

在实验前,保持所有采集参数不变,首先扣上镜头盖采集暗背景图像 B;然后打开镜头盖并扫描 99% 反射率的标准白板得到白背景图像 W;在实验时,采集系统会根据计算公式  $X=(I-B)/(W-B)$ ,自动将采集到的绝对图像 I 校正为相对图像 X。实验时每采集 10 个土壤样本,重复一次光谱校正。

### 1.3 光谱数据的提取

高光谱图像同时包含了土壤样本的光谱信息和图像信息。图像上的每一个像素点,都有一条漫反射光谱曲线与之相对应。采用 SpectralView 软件,提取高光谱图像上土壤样本范围内所有像素点的漫反射光谱曲线,并取算术平均值,得到平均光谱曲线(R)。

### 1.4 速效钾含量的实验室测定

根据《土壤速效钾和缓效钾含量的测定》(NY/T 889—2004)中性乙酸铵溶液浸提、火焰光度计法测定土壤中的速效钾含量。

### 1.5 光谱数据前处理方法

采用五种光谱数据前处理方法:

- (1)平均光谱曲线(R),由高光谱图像上土壤样品范围内多个像素点的光谱曲线取算术平均值得到;
- (2)平均光谱曲线的一阶导数(FD);
- (3)平均光谱曲线与一阶导数共同建模(R&FD),即 R 和 FD 同时作为输入参数参与建模;
- (4)平均光谱曲线与一阶导数的乘积( $R * FD$ );
- (5)平均光谱曲线与一阶导数的商( $R/FD$ )。

### 1.6 建模方法

偏最小二乘(partial least square, PLS)方法是结合多元线性回归和主成分分析的化学计量学方法,其本质是消除光谱共线性后再做回归。其中 PLS 的主因子个数是其最重要的参数,它代表了样品光谱数据加权组合的组数。当主因子个数过少时,会出现欠拟合现象,模型不能充分反映样品主要信息,预测精度降低;而当主因子个数过多时,会出现过拟合现象,会引入噪声干扰的主要成分的数据,同样会导致模型的预测能力降低。

采用 PLS 方法,分别使用建模集  $T_c$ ,  $L_c$  及  $H_c$  进行建模,然后分别对验证集  $T_p$ ,  $L_p$  及  $H_p$  进行验证。

### 1.7 模型评价标准

采用建模集决定系数( $R_c^2$ )、验证集决定系数( $R_p^2$ )、建模集均方根误差( $RMSE_c$ )、验证集均方根误差( $RMSE_p$ )、建模集相对分析误差( $RPD_c$ )、验证集相对分析误差( $RPD_p$ )等指

数来对比分析评价模型的优劣。

决定系数( $R^2$ )的值越大越好,但不能超过 1,越接近 1 表明模型的预测能力越强。均方根误差(RMSE)越小,说明模型越好。相对分析误差(RPD)越大,说明模型越好:当  $RPD \geq 2.0$  时,表明模型较好,可用于土壤速效钾的定量预测;当 RPD 在 1.0~2.0 之间时,表明模型预测效果较差;当  $RPD \leq 1.0$  时,表明模型预测效果极差,不能用于土壤速效钾的定量预测<sup>[11]</sup>。

## 2 结果与讨论

### 2.1 不同速效钾含量的光谱响应特征

将所有的土壤样本按实验室实测速效钾含量进行分类,分别统计速效钾含量在 0~50, 50~100, 100~200, 200~500 及  $500 \text{ mg} \cdot \text{kg}^{-1}$  以上的光谱反射率,计算平均光谱曲线。同时对平均光谱曲线求导,获取每类光谱的一阶导数曲线。结果如图 1 和图 2 所示。

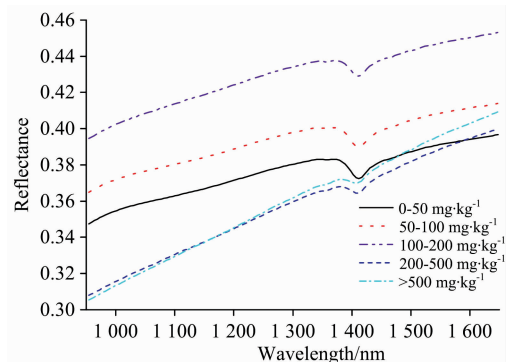


图 1 不同速效钾含量的平均光谱曲线

Fig. 1 Average spectral curves of different available potassium contents

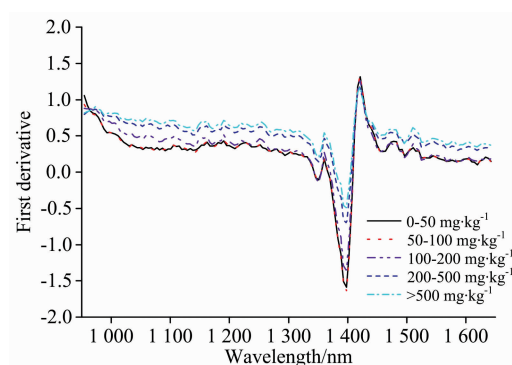


图 2 不同速效钾含量的平均光谱曲线的一阶导数

Fig. 2 First derivative of the average spectral curve of different available potassium contents

由图 1 可知,随波长的增加,土壤的光谱反射率逐渐增大;在 1 400 nm 附近有明显的 OH 键和  $\text{H}_2\text{O}$  的水吸收特征,这可能与残余在土壤中的少量水分有关。随速效钾含量的增加,平均光谱曲线呈先增加后减小的趋势。当速效钾含量低

于  $100 \text{ mg} \cdot \text{kg}^{-1}$  时, 所有波段的光谱反射率随速效钾含量的增加而增加; 当速效钾含量在  $100 \sim 200 \text{ mg} \cdot \text{kg}^{-1}$  之间时, 所有波段的光谱反射率均达到最大值。当速效钾含量超过  $200 \text{ mg} \cdot \text{kg}^{-1}$  时,  $950 \sim 1400 \text{ nm}$  的光谱反射率急剧减小, 但曲线的整体斜率显著增加; 且速效钾含量越高, 曲线整体斜率越大。

由图 2 可知, 速效钾含量为  $0 \sim 50$  和  $50 \sim 100 \text{ mg} \cdot \text{kg}^{-1}$  的平均光谱曲线的一阶导数基本完全重合; 当速效钾含量在  $100 \sim 200 \text{ mg} \cdot \text{kg}^{-1}$  之间时,  $950 \sim 1350 \text{ nm}$  之间的一阶导数显著增加; 当速效钾含量高于  $200 \text{ mg} \cdot \text{kg}^{-1}$  时, 所有波长范围内一阶导数均有显著提高; 速效钾含量高于  $500 \text{ mg} \cdot \text{kg}^{-1}$  的一阶导数显著高于  $200 \sim 500 \text{ mg} \cdot \text{kg}^{-1}$  的一阶导数。这与图 1 平均光谱曲线所得结论相同。

平均光谱曲线随速效钾含量先增大后减小的特点, 可能

导致两个速效钾含量差别极大的平均光谱曲线出现部分重合, 可能影响平均光谱曲线对速效钾含量的预测效果。低含量速效钾的一阶导数曲线基本重合, 因此采用一阶导数对速效钾含量低于  $100 \text{ mg} \cdot \text{kg}^{-1}$  的预测效果可能较差; 而高含量速效钾的一阶导数随速效钾含量的变化有显著区别, 因此一阶导数对速效钾含量高于  $100 \text{ mg} \cdot \text{kg}^{-1}$  的预测效果可能较好。将平均光谱曲线和一阶导数组合使用, 可能建立更好的预测模型, 获得更好的预测效果。

## 2.2 不同 PLS 模型及对所有速效钾含量验证集 ( $T_p$ ) 的验证效果

分别采用 R, FD, R&FD, R/FD, R\*FD 等五种光谱数据预处理方法,  $T_c$ ,  $L_c$  和  $H_c$  等三类建模集, 结合 PLS 方法进行建模, 并对验证集  $T_p$  进行验证。不同 PLS 模型及对验证集  $T_p$  的验证结果见表 2。

表 2 不同 PLS 模型及对所有钾含量验证集 ( $T_p$ ) 的验证结果

Table 2 Different PLS models and validation results of all potassium content validation sets ( $T_p$ )

光谱预处理方式	建模					验证		
	建模集	主因子	RMSE <sub>c</sub>	RPD <sub>c</sub>	R <sub>c</sub> <sup>2</sup>	RMSE <sub>p</sub>	RPD <sub>p</sub>	R <sub>p</sub> <sup>2</sup>
R	$T_c$	3	29.785	4.588	0.954	32.151	4.254	0.945
	$L_c$	18	1.856	5.350	0.966	161.100	0.141	0.427
	$H_c$	4	32.472	6.161	0.974	502.924	0.735	0.563
FD	$T_c$	2	30.250	4.514	0.953	30.087	4.527	0.952
	$L_c$	13	2.231	4.417	0.951	135.175	0.135	0.380
	$H_c$	3	13.442	15.045	0.995	108.422	1.548	0.893
R&FD	$T_c$	5	22.810	6.050	0.973	37.041	3.760	0.935
	$L_c$	14	2.042	4.847	0.959	136.477	0.128	0.326
	$H_c$	4	8.534	23.729	0.998	110.932	1.532	0.906
R/FD	$T_c$	3	30.261	4.512	0.953	54.285	2.277	0.842
	$L_c$	11	1.948	5.089	0.962	210.919	0.424	0.303
	$H_c$	5	12.256	16.507	0.996	509.905	0.991	0.147
R*FD	$T_c$	2	29.293	4.669	0.956	29.438	4.740	0.958
	$L_c$	15	2.107	4.690	0.956	132.605	0.158	0.435
	$H_c$	3	16.515	12.232	0.993	77.549	2.045	0.928

由表 2 可知, 当采用  $T_c$  进行建模时, 5 种光谱预处理方法所建立的 PLS 模型的最小 RPD<sub>p</sub> 为 2.277, 最小 R<sub>p</sub><sup>2</sup> 为 0.842, 即所有模型对  $T_p$  的验证效果均较好。光谱预处理方法从好到差依次是: R\*FD > FD > R > R&FD > R/FD。其中, 采用 R\*FD 光谱预处理方法所建立的模型, PLS 主因子个数只有 2 个, 为所有模型中最少; RMSE<sub>c</sub> = 29.293, RMSE<sub>p</sub> = 29.438, 两者取值较小且最为接近; RPD<sub>c</sub> = 4.669, R<sub>c</sub><sup>2</sup> = 0.956, RPD<sub>p</sub> = 4.740, R<sub>p</sub><sup>2</sup> = 0.958 均为所有模型中最高, 使用建模集  $T_c$  建立的 PLS 模型对  $T_p$  的验证效果最好。

当采用建模集  $L_c$  建模时, 所有模型对  $T_p$  的验证效果均极差。当采用建模集  $H_c$  建模时, 只有使用 R\*FD 光谱预处理方法建立的 PLS 模型的 RPD<sub>c</sub> = 2.045 > 2.0, 但 RMSE<sub>p</sub> 相对较大, 表明该模型稳定性稍差。

## 2.3 不同 PLS 模型及对低含量验证集 ( $L_p$ ) 的验证效果

分别采用 R, FD, R&FD, R/FD, R\*FD 等五种光谱数

据预处理方法,  $T_c$ ,  $L_c$  和  $H_c$  等三类建模集, 结合 PLS 方法进行建模, 并对验证集  $L_p$  进行验证。不同 PLS 模型及对验证集  $L_p$  的验证结果见表 3。

由表 3 可知, 所有模型对  $L_p$  的验证效果为: 最大 RPD<sub>p</sub> 仅有 1.052, 最高 R<sub>p</sub><sup>2</sup> 只有 0.288。因此, 无论采用何种光谱预处理方法, 使用何种建模集进行 PLS 建模, 都无法对速效钾含量低于  $100 \text{ mg} \cdot \text{kg}^{-1}$  的褐土土壤的速效钾含量进行有效预测。其主要原因可能与低含量速效钾对光谱反射率的影响不明显, 土壤内其他成分对光谱的影响掩盖了速效钾对光谱的影响有关。

此外, 采用 R 与  $T_c$ , FD 与  $T_c$ , FD 与  $H_c$ , R&FD 与  $T_c$ , R\*FD 与  $T_c$ , R\*FD 与  $H_c$  组合建立的 PLS 模型, 对  $L_p$  的预测值均小于  $100 \text{ mg} \cdot \text{kg}^{-1}$ 。

## 2.4 不同 PLS 模型及对高含量验证集 ( $H_p$ ) 的验证效果

分别采用 R, FD, R&FD, R/FD, R\*FD 等五种光谱数

据预处理方法,  $T_c$ ,  $L_c$  和  $H_c$  等三类建模集, 结合 PLS 方法进行建模, 并对验证集  $H_p$  进行验证。不同 PLS 模型及对验证集  $H_p$  的验证结果见表 4。

表 3 不同 PLS 模型及对钾含量低于  $100 \text{ mg} \cdot \text{kg}^{-1}$  验证集 ( $L_p$ ) 的验证结果

Table 3 Different PLS models and validation results of potassium content less than  $100 \text{ mg} \cdot \text{kg}^{-1}$  validation set ( $L_p$ )

光谱预处理方式	建模					验证		
	建模集	主因子	RMSE <sub>c</sub>	RPD <sub>c</sub>	R <sub>c</sub> <sup>2</sup>	RMSE <sub>p</sub>	RPD <sub>p</sub>	R <sub>p</sub> <sup>2</sup>
R	$T_c$	3	29.785	4.588	0.954	22.498	0.954	0.031
	$L_c$	18	1.856	5.350	0.966	15.078	0.944	0.131
	$H_c$	4	32.472	6.161	0.974	540.902	0.440	0.029
FD	$T_c$	2	30.250	4.514	0.953	18.876	0.818	0.005
	$L_c$	13	2.231	4.417	0.951	17.037	0.749	0.044
	$H_c$	3	13.442	15.045	0.995	104.636	0.437	0.073
R&FD	$T_c$	5	22.810	6.050	0.973	28.169	1.042	0.288
	$L_c$	14	2.042	4.847	0.959	17.069	0.748	0.043
	$H_c$	4	8.534	23.729	0.998	116.680	0.386	0.209
R/FD	$T_c$	3	30.261	4.512	0.953	36.072	1.052	0.129
	$L_c$	11	1.948	5.089	0.962	18.825	0.840	0.133
	$H_c$	5	12.256	16.507	0.996	548.073	0.910	0.004
R * FD	$T_c$	2	29.293	4.669	0.956	19.802	0.789	0.001
	$L_c$	15	2.107	4.690	0.956	17.879	0.755	0.033
	$H_c$	3	16.515	12.232	0.993	79.610	0.433	0.108

表 4 不同 PLS 模型及对钾含量高于  $100 \text{ mg} \cdot \text{kg}^{-1}$  验证集 ( $H_p$ ) 的验证结果

Table 4 Different PLS models and validation results of potassium content more than  $100 \text{ mg} \cdot \text{kg}^{-1}$  validation set ( $H_p$ )

光谱预处理方式	建模					验证		
	建模集	主因子	RMSE <sub>c</sub>	RPD <sub>c</sub>	R <sub>c</sub> <sup>2</sup>	RMSE <sub>p</sub>	RPD <sub>p</sub>	R <sub>p</sub> <sup>2</sup>
R	$T_c$	3	29.785	4.588	0.954	64.799	2.905	0.895
	$L_c$	18	1.856	5.350	0.966	424.628	0.100	0.844
	$H_c$	4	32.472	6.161	0.974	122.783	2.101	0.786
FD	$T_c$	2	30.250	4.514	0.953	64.796	2.886	0.897
	$L_c$	13	2.231	4.417	0.951	355.196	0.029	0.001
	$H_c$	3	13.442	15.045	0.995	68.098	2.798	0.885
R&FD	$T_c$	5	22.810	6.050	0.973	26.304	2.626	0.879
	$L_c$	14	2.042	4.847	0.959	358.656	0.027	0.043
	$H_c$	4	8.534	23.729	0.998	66.757	2.795	0.890
R/FD	$T_c$	3	30.261	4.512	0.953	42.796	1.417	0.698
	$L_c$	11	1.948	5.089	0.962	556.131	0.405	0.411
	$H_c$	5	12.256	16.507	0.996	133.088	2.449	0.616
R * FD	$T_c$	2	29.293	4.669	0.956	23.033	3.199	0.915
	$L_c$	15	2.107	4.690	0.956	348.096	0.033	0.006
	$H_c$	3	16.515	12.232	0.993	63.801	3.019	0.899

由表 4 可知, 采用 R, FD, R&FD, R \* FD 等 4 种光谱预处理方法, 使用建模集  $T_c$  建模, 均可获得较好的预测效果, 光谱预处理方法从好到差依次是: R \* FD > R > FD > R&FD > R/FD; 使用建模集  $H_c$  建模, 也可获得较好的预测效果, 光谱预处理方法从好到差依次是: R \* FD > FD > R&FD > R > R/FD。其中最佳模型为: 采用 R \* FD 光谱预处理方法, 使用建模集  $T_c$  建立的 PLS 模型, PLS 主成分因子为 2 个, 为所有模型中最少; RMSE<sub>c</sub> = 29.293, RMSE<sub>p</sub> =

23.033, 两者取值较小且最为接近; RPD<sub>c</sub> = 4.669, R<sub>c</sub><sup>2</sup> = 0.956, RPD<sub>p</sub> = 3.199, R<sub>p</sub><sup>2</sup> = 0.915 均为所有模型中最高。

此外, 采用 R 与  $T_c$ , FD 与  $T_c$ , FD 与  $H_c$ , R&FD 与  $T_c$ , R&FD 与  $H_c$ , R \* FD 与  $T_c$ , R \* FD 与  $H_c$ , R/FD 与  $T_c$  组合建立的 PLS 模型, 对  $H_p$  的预测值均大于  $100 \text{ mg} \cdot \text{kg}^{-1}$ 。

结合之前的试验结果, 采用 R 与  $T_c$ , FD 与  $T_c$ , FD 与  $H_c$ , R&FD 与  $T_c$ , R \* FD 与  $T_c$ , R \* FD 与  $H_c$  组合建立的

PLS 模型, 对验证集  $L_p$  的预测值均小于  $100 \text{ mg} \cdot \text{kg}^{-1}$ , 对验证集  $H_p$  的预测值均大于  $100 \text{ mg} \cdot \text{kg}^{-1}$ , 即能够对速效钾含量低于和高于  $100 \text{ mg} \cdot \text{kg}^{-1}$  的土壤进行分类。

### 3 结 论

以山西典型褐土土壤为研究对象, 所有土壤样品经风干后, 未经研磨过筛处理, 直接装入样品杯进行测量。获取所有土壤样本的近红外高光谱图像, 并提取平均光谱曲线(R)。分别采用 R, FD, R&FD, R\*FD 及 R/FD 等 5 种光谱数据前处理方法,  $T_c$ ,  $L_c$  和  $H_c$  等三种建模集建立 PLS 预测模型, 并对  $T_p$ ,  $L_p$  及  $H_p$  等三种验证集进行验证。主要结论如下:

(1) 土壤的平均光谱反射率随速效钾含量的增大呈现先增加后减小的趋势。当速效钾含量在  $100 \sim 200 \text{ mg} \cdot \text{kg}^{-1}$  时, 光谱反射率达到最大。当速效钾含量高于  $100 \text{ mg} \cdot \text{kg}^{-1}$  时, 平均光谱曲线的一阶导数显著增大, 且随速效钾含量的增加而增加。

(2) 本工作所建立的 PLS 模型, 均无法对速效钾含量低于  $100 \text{ mg} \cdot \text{kg}^{-1}$  的褐土土壤的速效钾含量进行有效预测。主要原因可能与低含量速效钾对光谱反射率的影响不明显,

土壤内其他成分对光谱的影响掩盖了速效钾对光谱的影响有关。未来仍需对低含量速效钾的预测方法和预测模型做进一步研究。

(3) 建模效果最好的光谱预处理方法为 R\*FD, 其次为 FD, R, 而 R&FD, R/FD 效果相对较差。其中最优建模方式为: R\*FD 结合  $T_c$  建模, 其 PLS 主因子个数 2 个,  $\text{RMSE}_c = 29.293$ ,  $\text{RPD}_c = 4.669$ ,  $R_c^2 = 0.956$ ; 对  $T_p$  的验证效果为  $\text{RMSE}_p = 29.438$ ,  $\text{RPD}_p = 4.740$ ,  $R_p^2 = 0.958$ ; 对  $H_p$  的验证效果为  $\text{RMSE}_p = 23.033$ ,  $\text{RPD}_p = 3.199$ ,  $R_p^2 = 0.915$ 。该预处理方法对土壤其他营养元素、其他土壤类型的预测效果有待进一步研究。

(4) 采用 R\*FD 光谱预处理方法建立的 PLS 模型, 能够根据土壤速效钾的含量对土壤进行分类: 当预测值小于  $100 \text{ mg} \cdot \text{kg}^{-1}$  时, 表明土壤速效钾含量低于  $100 \text{ mg} \cdot \text{kg}^{-1}$ , 具体含量不确定; 当预测值大于  $100 \text{ mg} \cdot \text{kg}^{-1}$  时, 预测值则能够很好反映土壤速效钾的真实含量。

选用的土壤样本未经研磨和过筛处理, 能够大大缩短样本制备时间, 提高预测效率。研究结果可为近红外高光谱成像技术应用于褐土土壤除速效钾含量外的营养成份的快速预测提供参考。

### References

- [1] CEN Yi-lang, SONG Tao, HE Yong, et al(岑益郎, 宋 韬, 何 勇). Journal of Zhejiang University · Agric & Life Sci(浙江大学学报·农业与生命科学版), 2011, 37(3): 300.
- [2] Xu Dongyun, Ma Wanzhu, Chen Songchao, et al. Computers and Electronics in Agriculture, 2018, 144: 1.
- [3] Sorenson P T, Quideau S A, Rivard B. Geoderma, 2018, 315: 170.
- [4] Said Nawar, Abdul M Mouazen. Catena, 2017, 151: 118.
- [5] CHEN Chen, LI Zhi-wei, DONG Da-ming(陈 晨, 李志伟, 董大明). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(11): 3557.
- [6] LI Shuo, WANG Shan-qin, ZHANG Mei-qin(李 硕, 汪善勤, 张美琴). Acta Optica Sinica(光学学报), 2012, 32(8): 297.
- [7] LIU Yan-de, XIONG Song-sheng, WU Zhi-jing, et al(刘燕德, 熊松盛, 吴至境, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2013, 29(18): 156.
- [8] LIU Xue-mei, LIU Jian-she(刘雪梅, 柳建设). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2013, 44(3): 88.
- [9] WANG Chang, HUANG Chi-chao, YU Guang-hui, et al(王 昶, 黄驰超, 余光辉, 等). Acta Pedologica Sinica(土壤学报), 2013, 50(5): 881.
- [10] HU Guo-tian, HE Dong-jian, Kenneth A Sudduth(胡国田, 何东健, Kenneth A Sudduth). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2015, 46(7): 139.
- [11] ZHANG Juan-juan, TIAN Yong-chao, YAO Xia, et al(张娟娟, 田永超, 姚 霞, 等). Acta Pedologica Sinica(土壤学报), 2012, 49(1): 50.

## Prediction of Available Potassium Content in Cinnamon Soil Using Hyperspectral Imaging Technology

WANG Wen-jun<sup>1</sup>, LI Zhi-wei<sup>1</sup>, WANG Can<sup>1</sup>, ZHENG De-cong<sup>1\*</sup>, DU Hui-ling<sup>2\*</sup>

1. College of Engineering, Shanxi Agricultural University, Taigu 030801, China

2. College of Arts and Sciences, Shanxi Agricultural University, Taigu 030801, China

**Abstract** Variable rate fertilization in precision agriculture depends on the understanding of the distribution of soil nutrients in the farmland. The rapid acquisition of soil information is the basis for the application of precision agriculture. Available potassium is an important parameter of soil fertility, and it is a necessary nutrient element for plant growth. The measurement of the content of available potassium in soil is an important way to understand the soil fertility, and it is a precondition of the realization of precision agriculture. In this paper, a total of 169 farmland plough cinnamon soil samples were collected in Shanxi province. All samples were air dried, with the larger soil particles crumbled and impurities removed manually and directly used for measuring soil near infrared hyperspectral without grinding and sieving. According to the measuring results of the available potassium content in the laboratory, all soil samples were divided into two parts. There were 144 samples of available potassium content less than  $100 \text{ mg} \cdot \text{kg}^{-1}$ , and 108 samples were randomly selected as the low content modeling sets ( $L_c$ ), and the remaining 36 samples as the low content validation sets ( $L_p$ ). There were 25 samples of available potassium content more than  $100 \text{ mg} \cdot \text{kg}^{-1}$ , and 19 samples were randomly selected as the high content modeling sets ( $H_c$ ), and the remaining 6 samples as the high content validation sets ( $H_p$ ).  $L_c$  and  $H_c$  were collectively known as all modeling sets ( $T_c$ ), and  $L_p$  and  $H_p$  as all validation sets ( $T_p$ ). Near infrared hyperspectral imaging technology was used to obtain near infrared hyperspectral images in the range of  $950 \sim 1650 \text{ nm}$  of all soil samples. There were five different spectral data preprocessing methods used in this paper: the average spectral curve (R), the first derivative of the average spectral curve (FD), the average spectral curve and the first derivative co-modeled (R&FD), the product of the average spectral curve and the first derivative ( $R * \text{FD}$ ) and the quotient of average spectral curve and first derivative ( $R/\text{FD}$ ). Combined with partial least squares (PLS) method, the models were built using the modeling set  $T_c$ ,  $L_c$  and  $H_c$  respectively. The validation sets  $T_p$ ,  $L_p$  and  $H_p$  were verified respectively. The results showed that: along with the increase of available potassium content, the average spectral reflectance of soil increased first and then decreased. When the content of available potassium was less than  $100 \text{ mg} \cdot \text{kg}^{-1}$ , the spectral reflectance of all bands increased with the increase of available potassium content. When the content of available potassium was between  $100 \sim 200 \text{ mg} \cdot \text{kg}^{-1}$ , the spectral reflectance of all bands reached the maximum. When the available potassium content was more than  $200 \text{ mg} \cdot \text{kg}^{-1}$ , the spectral reflectance of  $950 \sim 1400 \text{ nm}$  decreased sharply, but the overall slope of the curve increased significantly. The higher the available potassium content was, the larger the overall slope of the curve was. When the content of available potassium was higher than  $100 \text{ mg} \cdot \text{kg}^{-1}$ , the first derivative of the average spectral curve increased significantly, and increased with the increase of available potassium content. The PLS models proposed in this paper could predict the whole (all available potassium content) and high content ( $\geq 100 \text{ mg} \cdot \text{kg}^{-1}$ ) of available potassium effectively; but could not predict the low content ( $\leq 100 \text{ mg} \cdot \text{kg}^{-1}$ ) of available potassium. The best spectral preprocessing method was:  $R * \text{FD}$ , followed by FD and R. The predict results of R&FD and  $R/\text{FD}$  were relatively poor. The optimal modeling method was  $R * \text{FD}$  combined with  $T_c$ . The number of PLS principal factors was 2,  $\text{RMSE}_c = 29.293$ ,  $\text{RPD}_c = 4.669$ ,  $R_c^2 = 0.956$ ;  $\text{RMSE}_p = 29.438$ ,  $\text{RPD}_p = 4.740$ ,  $R_p^2 = 0.958$  for the validation sets of  $T_p$ ;  $\text{RMSE}_p = 23.033$ ,  $\text{RPD}_p = 3.199$ ,  $R_p^2 = 0.915$  for the validation sets of  $H_p$ . This model could classify soil according to the content of available potassium. When the predicted value was less than  $100 \text{ mg} \cdot \text{kg}^{-1}$ , it indicated that the content of available potassium in soil was less than  $100 \text{ mg} \cdot \text{kg}^{-1}$ , and the specific content was uncertain; while when the predicted value was higher than  $100 \text{ mg} \cdot \text{kg}^{-1}$ , the predicted value could reflect the real content of soil available potassium well. Because the soil samples selected in this paper were used without ground or sifted, the time of sample preparation could be greatly shortened and the prediction efficiency could be improved greatly. The results in this study can provide a reference for the rapid prediction of nutrients including available potassium content in cinnamon soil using near infrared hyperspectral imaging technology.

**Keywords** Near infrared; Hyperspectral imaging; Available potassium; Partial least square

\* Corresponding authors

(Received Mar. 30, 2018; accepted Aug. 12, 2018)