

# 基于近红外光谱和正交信号-偏最小二乘法对土壤的分类

宋海燕<sup>1</sup>, 秦刚<sup>2</sup>, 韩小平<sup>1</sup>, 刘海芹<sup>1</sup>

(1. 山西农业大学工学院, 太谷 030801; 2. 山西农业大学林学院, 太谷 030801)

**摘要:** 不同质地的土壤, 由于蓄水能力和土壤颗粒大小的不同使得其光谱特性不同, 这为采用近红外光谱技术对土壤质地进行判别分析提供了依据。该研究利用正交信号校正(OSC)方法可以获得与浓度有关的谱图信息这一优势, 将其与偏最小二乘法(PLS)结合, 采用近红外光谱技术对不同质地的土壤: 砂土、壤土、黏土进行判别分析。结果表明: 建模样本的相关系数可达 0.965, 采用该模型对其余 45 个样本分别进行了预测, 三种土壤预测样本的判别正确率分别为: 93.3%, 86.6%和 86.6%。说明 OSC 方法可以提取谱图中的微弱的质地信息, 实现土壤质地的快速鉴别分析。

**关键词:** 土壤, 质地, 分类, OSC-PLS, 近红外光谱

doi: 10.3969/j.issn.1002-6819.2012.07.028

中图分类号: S152.3

文献标志码: A

文章编号: 1002-6819(2012)-07-0168-04

宋海燕, 秦刚, 韩小平, 等. 基于近红外光谱和正交信号-偏最小二乘法对土壤的分类[J]. 农业工程学报, 2012, 28(7): 168-171.

Song Haiyan, Qin Gang, Han Xiaoping, et al. Soil classification based on near infrared reflectance spectroscopy and orthogonal signal correction-partial least square[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(7): 168-171. (in Chinese with English abstract)

## 0 引言

近红外光谱(near infrared reflectance spectroscopy, 缩写 NIR)分析技术是一种高效、快速的现代分析技术。由于近红外谱区(780~2 526 nm)几乎包括了有机物中所有含氢基团的信息, 蕴涵着分子的结构、组成状态等信息。因此, 采用近红外光谱分析技术对土壤进行定性定量检测已成为当前光谱检测的热点<sup>[1-8]</sup>。

土壤质地是指土壤中不同大小直径的矿物颗粒的组合状况, 一般分为砂土、壤土和黏土 3 类, 是土壤物理性质之一, 它是影响土壤光谱特征的主要因素之一。这是因为, 一方面是由于它影响土壤蓄水能力, 较大的颗粒之间能容纳更多的空气和水, 另一方面是土壤颗粒大小对土壤反射率有显著影响; 颗粒越小, 彼此的结合越紧密, 土壤表面也就越平滑, 反射率就越大<sup>[9]</sup>。为此, 对不同类型土壤的光谱特性及对预测成分的影响也开展了一系列研究工作<sup>[10-14]</sup>。但以上研究大多都集中在土壤类型及土壤类型对土壤成分预测含量的影响上, 对采用近红外光谱进行快速判别土壤质地的研究上还很少。

正交信号校正(orthogonal signal correction, 缩写 OSC)方法是 1998 年由 S.Wold 提出的<sup>[15]</sup>。这类预处理方法的基本原理是: 在建立定量校正模型前, 通过正交的数学方法将光谱阵用浓度阵正交, 滤除光谱与浓度阵不相关的正交主成分, 然后在此基础上再进行定量分析。近年来国内外

运用 OSC 作为谱图预处理方法已开展了一些研究工作<sup>[16-18]</sup>。本文采用近红外光谱技术对收集的 3 种不同质地的土壤进行了分析, 并将 OSC 和偏最小二乘(partial least square, 缩写 PLS)结合对其进行了判别分类研究。

## 1 材料与方法

### 1.1 土壤样本的获取

本研究主要从山西省境内的关帝山、太谷五马河、山西农业大学试验田、右玉等地采集了 156 个样本。将样本一分为二, 一部分送于资源环境学院, 采用比重计法进行质地测定, 按照中国质地分类标准<sup>[19]</sup>将其分为 3 类。砂土: 1~0.05 mm 的砂粒占有矿物质土粒 60%~70%, 样本数为 51; 黏土: 小于 0.001 mm 的黏粒质量分数为 35%~40%, 样本数为 54; 壤土: 1~0.05 mm 的砂粒占有矿物质土粒 15%~20%, 小于 0.001 mm 的黏粒质量分数为 20%~30%, 样本数为 51。另一部分在室内摊开, 自然风干, 经 2.5 mm 筛预处理, 用于光谱分析。

### 1.2 测量仪器及方法

试验使用 ASD 公司的 FieldSpec3 光谱仪, 波长范围为 350~2 500 nm, 扫描获得样本的反射率值。试验在室温为 20℃左右的环境下进行, 测量时将土壤样本置于直径为 12.5 cm 的培养皿中, 并放入一自制的暗室内进行。光谱采集软件为 Rs3, 数据导出为 ASC II, 数据分析软件为 The UnscramblerX 10.1 (CAMO AS Oslo, Norway), Matlab V7.0 (The Math Works, Natick, USA)。

### 1.3 判别分析方法的选取

考虑到土壤光谱的特征曲线在反映质地信息方面要比化学信息和物理信息少<sup>[13]</sup>, 若采用常见谱图预处理方法对获得的光谱图进行处理, 很有可能会将反映质地信

收稿日期: 2011-06-08 修订日期: 2012-02-03

基金项目: 基金项目: 国家科技支撑计划(2009BADB5B03); 山西省科技攻关(20100311066-5), 科技创新(2010011)的资助

作者简介: 宋海燕(1977-), 女, 山西介休人, 副教授, 博士, 主要从事田间信息检测技术研究。太谷 山西农业大学工学院, 030801。

Email: yybbao@163.com

息的谱图特征误认为噪声滤除，为此引入 OSC 对谱图进行预处理。

PLS 在光谱技术建模分析中是一种常用的多元分析方法<sup>[20]</sup>，该方法是将因子分析和回归分析相结合的方法。本文采用 PLS 方法建立光谱数据与土壤质地的判别模型。模型性能的优劣以对预测集样本的准确判别率为标准，正确判别率越高，说明模型性能越好。

## 2 试验结果与分析

### 2.1 不同类型土壤近红外吸收光谱特性

3 类土壤（砂土、壤土和黏土）的近红外光谱曲线如图 1 所示，图中横坐标为波长范围 400~2 500 nm，纵坐标为吸光度（ $\log 1/R$ ，R 为反射率）。从图中可以看出，3 类土壤的可见近红外谱图趋势非常相似，很难识别由土壤质地不同带来的谱图差异。

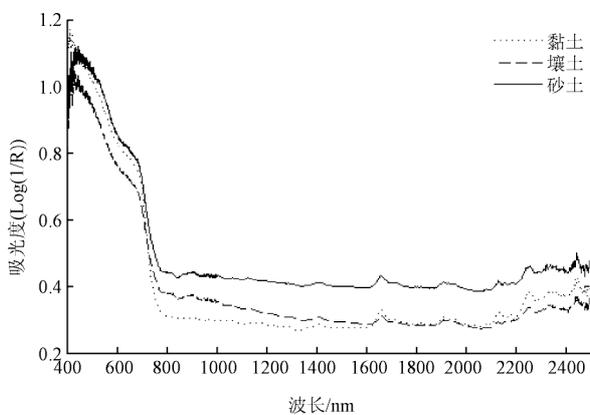


图 1 不同质地土壤样本光谱图  
Fig.1 Spectra of different textural soil samples

### 2.2 OSC 谱图预处理

OSC 有多种算法，其中 DO 算法是直接光谱阵与浓度阵正交来滤除无关信号的方法，该方法算法简单，运行速度快<sup>[21]</sup>，因此本文选用了该方法。

DO 运算步骤如下：

1) 将原始校正集光谱阵  $X(n \times m)$  和浓度阵  $Y(n \times 1)$  进行标准化处理；

2) 计算  $M = XY(Y'Y)^{-1}$ ；

3) 计算  $Z = X - YM'$ ；

4) 对 Z 进行主成分分析，取前  $f$  个需正交处理的得分矩阵  $T_f$  和载荷矩阵  $P_f$ ；

5) 计算新的  $T_f^{new}$ ， $T_f^{new} = XP_f$ ；

6)  $X_{OD} = X - T_f^{new} P_f'$ ；

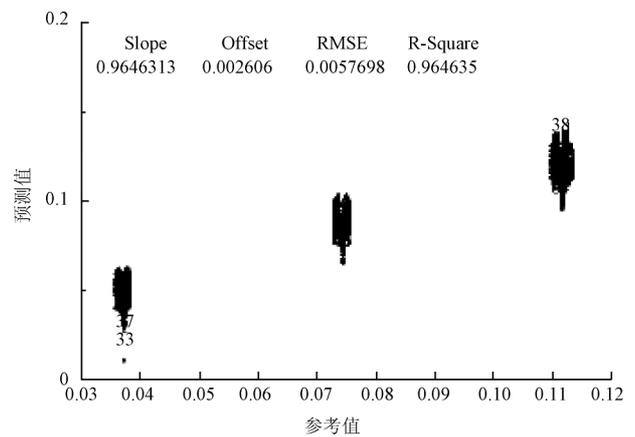
7) 对于预测向量  $x_{new}$ ，由载荷  $P_f$  求出校正后的光谱：  
 $T = x_{new} P_f, x'_{OD} = x_{new} - T P_f'$

为了能从定量角度判别分类效果，本文将黏土、壤土、砂土 3 类土壤分别赋予以下 3 个虚拟变量：1，2 和 3 作为其浓度信息。根据 DO 运算步骤（1），浓度阵标准化处理后新形成的虚拟变量分别为：0.0373，0.0745，和 0.112，即黏土的虚拟参考值变为 0.0373、壤土的虚拟参考值变为 0.0745、砂土的虚拟参考值变为 0.112。

OSC 预处理在 MATLAB 环境中实现。在使用 OSC 对光谱进行预处理时，光谱阵正交处理所用主因子数的选取非常重要，本文根据累计贡献率来确定。当主因子数为 4 时，累计贡献率达到 98%。

### 2.3 PLS 判别模型的建立

将上述 OSC 处理后的光谱数据，随机选取砂土、黏土和壤土各 15 个作为预测样本，其余作为建模样本导入数据分析软件 UnscramblerX 10.1 进行 PLS 分析。其建模样本参考值和预测值相关图如图 2。



注：slope, offset, RMSE, R-Square 分别表示建模样本的斜率、偏移量、均方根误差和相关系数，每个参考值所对应预测值的样本号见表 1。

图 2 建模样本参考值和预测值相关图  
Fig.2 Predicted vs. reference plot for calibration samples

### 2.4 预测结果及分析

采用上述 PLS 判别模型对其余 45 个样本（砂土、黏土和壤土各 15 个）进行预测。设定预测结果偏差在  $\pm 0.01$  以内，将预测值介于 0.03~0.05 之间都认为黏土，介于 0.06~0.08 之间的都认为壤土，0.10~0.12 之间的都认为砂土，其预测结果见表 1。

表 1 不同质地土壤预测分析结果

Table 1 Prediction results of different texture soil samples

黏土			壤土			砂土		
样本	参考值	预测值	样本	参考值	预测值	样本	参考值	预测值
1	0.0373	0.0376	16	0.0747	0.0736	31	0.1120	0.1111
2	0.0373	0.0424	17	0.0747	0.0808	32	0.1120	0.1066
3	0.0373	0.0462	18	0.0747	0.0744	33	0.1120	0.1102
4	0.0373	0.0477	19	0.0747	0.0871	34	0.1120	0.1115
5	0.0373	0.0575	20	0.0747	0.0678	35	0.1120	0.1049
6	0.0373	0.0377	21	0.0747	0.0758	36	0.1120	0.1101
7	0.0373	0.0466	22	0.0747	0.0722	37	0.1120	0.1144
8	0.0373	0.0369	23	0.0747	0.0730	38	0.1120	0.1033
9	0.0373	0.0430	24	0.0747	0.0757	39	0.1120	0.1144
10	0.0373	0.0438	25	0.0747	0.0921	40	0.1120	0.1084
11	0.0373	0.0407	26	0.0747	0.0808	41	0.1120	0.1083
12	0.0373	0.0398	27	0.0747	0.0771	42	0.1120	0.1083
13	0.0373	0.0316	28	0.0747	0.0776	43	0.1120	0.0998
14	0.0373	0.0340	29	0.0747	0.0755	44	0.1120	0.0836
15	0.0373	0.0305	30	0.0747	0.0743	45	0.1120	0.0855
Crr	93.3%		86.6%			86.6%		

注：Crr(Correct recognition ratio)表示正确识别率

从表中可以看出：黏土中的样本 5，其预测值为 0.0575，而参考值为 0.0373，超出偏差范围，很容易被误判为壤土；壤土中的样本 19 和 25，同样超出偏差范围，很容易被误判为砂土；砂土样本 44 和 45 也超出偏差范围，很容易被误判为壤土。根据以下计算方法：

$$\frac{\text{正确识别样本数}}{\text{该类样本总数}} \times 100\%$$
，可得各类土壤的正确识别率分别为：93.3%，86.6%和 86.6%。

### 3 结 论

1) 土壤反射光谱特性是土壤的基本特征之一，它与土壤的化学信息，物理信息都有密切的关系。不同质地的土壤，由于蓄水能力和土壤颗粒大小的不同使得其光谱特性不同。因此采用近红外光谱特性曲线来区分土壤质地从理论上是可行的。

2) 与化学信息相比，质地信息在谱图曲线上的反应程度很低，因此，必须采用一些特殊的光谱处理技术才能获得其信息，并对其进行分析。

3) 本研究利用 OSC 方法可以获得与浓度有关的谱图信息这一优势，将其与 PLS 结合，采用近红外光谱技术对不同质地的土壤进行了判别分析，其判别准确率分别为 93.3%，86.6%和 86.6%。

#### [参 考 文 献]

- [1] 朱登胜, 吴迪, 宋海燕, 等. 应用近红外光谱法测定土壤的有机质和 pH 值[J]. 农业工程学报, 2008, 24(6): 196—198.  
Zhu Dengsheng, Wu Di, Song Haiyan, et al. Determination of organic matter contents and pH values of soil using near infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2008, 24(6): 196—199. (in Chinese with English abstract)
- [2] 郑立华, 李民赞, 安晓飞, 等. 基于近红外光谱和支持向量机的土壤参数预测[J]. 农业工程学报, 2010, 26(增刊 2): 81—87.  
Zheng Lihua, Li Minzan, An Xiaofei, et al. Forecasting soil parameters based on NIR and SVM[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2010, 26(Supp.2): 81—87. (in Chinese with English abstract)
- [3] 刘炜, 常庆瑞, 郭曼, 等. 土壤导数光谱小波去噪与有机质吸收特征提取[J]. 光谱学与光谱分析, 2011, 31(1): 100—104.  
Liu Wei, Chang Qingrui, Guo Man, et al. Extraction of first derivative spectrum features of soil organic matter via wavelet de-noising[J]. Spectroscopy and Spectral Analysis, 2011, 31(1): 100—104. (in Chinese with English abstract)
- [4] 沈掌泉, 王珂, Huang Xuewen. 用近红外光谱预测土壤碳含量的研究[J]. 红外与毫米波学报, 2010, 29(1): 32—37.  
Sheng Zhangquan, Wang Ke, Huang Xuewen. Estimation the content of soil carbon by using near-infrared spectra[J]. Journal of Infrared and Millimeter Waves, 2010, 29(1): 32—37. (in Chinese with English abstract)
- [5] 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.  
Lu Yanli, Bai Youlu, Wang Lei, et al. Determination for total nitrogen content in black soil using hyperspectral data[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2010, 26(1): 256—261. (in Chinese with English abstract)
- [6] Johanna Wetterlind, Bo Stenberg, Mats Söderström. Increased sample point density in farm soil mapping by local calibration of visible and near infrared prediction models[J]. Geoderma, 2010, 156(3/4): 152—160.
- [7] Volkan Bilgili A., H.M. van Es, F. Akbas, et al. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey[J]. Journal of Arid Environments, 2010, 74(2): 229—238.
- [8] Barthes B G, Didier B, Edmond H, et al. Determining the distributions of soil carbon and nitrogen in particle size fractions using near-infrared reflectance spectrum of bulk soil samples[J]. Soil Biology and Biochemistry, 2008, 40(6): 1533—1537.
- [9] 吴昫昭, 田庆久, 季峻峰, 等. 土壤光学遥感的理论、方法及应用[J]. 遥感信息, 2003(1): 40—47.  
Wu Junzhao, Tian Qingjiu, Ji Junfeng, et al. Soil remote sensing research theory method and application[J]. Remote Sensing Information, 2003(1): 40—47. (in Chinese with English abstract)
- [10] 张焕军, 张柏, 张渊智, 等. 基于反射光谱特性的土壤分类研究[J]. 光谱学与光谱分析, 2008, 28(3): 624—628.  
Zhang Huanjun, Zhang Bo, Zhang Yuanzhi, et al. soil taxonomy on the basis of reflectance spectral characteristics[J]. Spectroscopy and Spectral Analysis, 2008, 28(3): 624—628. (in Chinese with English abstract)
- [11] 张娟娟, 田永超, 朱艳, 等. 不同类型土壤的光谱特征及其有机质含量预测[J]. 中国农业科学, 2009, 42(9): 3154—3163.  
Zhang Juanjuan, Tian Yongchao, Zhu Yan, et al. Spectral characteristics and estimation of organic matter contents of different soil types[J]. Scientia Agricultura Sinica, 2009, 42(9): 3154—3163. (in Chinese with English abstract)
- [12] 张雪莲, 李晓娜, 武菊英, 等. 不同类型土壤总氮的近红外光谱技术测定研究[J]. 光谱学与光谱分析, 2010, 30(4): 906—910.  
Zhang Xuelian, Li Xiaona, Wu Juying, et al. study on the determination of total nitrogen(TN) in different types of soil by near infrared spectroscopy (NIS)[J]. Spectroscopy and Spectral Analysis, 2010, 30(4): 906—910. (in Chinese with English abstract)
- [13] 曾庆猛, 孙宇瑞, 严红兵. 土壤质地分类的近红外光谱分析方法研究[J]. 光谱学与光谱分析, 2009, 29(7): 1759—1763.  
Zeng Qingmeng, Sun Yurui, Yan Hongbing. NIR spectral analysis for soil textural classification[J]. Spectroscopy and Spectral Analysis, 2009, 29(7): 1759—1763. (in Chinese with English abstract)
- [14] 王遵义, 金春华, 刘飞, 等. 基于光谱技术的土壤快速分

- 类方法研究[J]. 浙江大学学报: 农业与生命版, 2010, 36(3): 282—286.
- Wang Zunyi, Jin Chuhua, Liu Fei. et al. Rapid discrimination of soil variety based on spectroscopic techniques[J]. Journal of Zhejiang University: Agric. and Life Sci., 2010, 36(3): 282—286. (in Chinese with English abstract)
- [15] Wold S, Antti H, Lindgren F, et al. Orthogonal signal correction of near-infrared spectra[J]. Chemom Intell Lab Syts, 1998, 44(1/2): 175—185.
- [16] 宋海燕, 何勇. 基于 OSC 和 PLS 的土壤有机质近红外光谱测定[J]. 农业机械学报, 2007, 38(12): 113—115.
- Song Haiyan, He Yong. Near infrared determination of organic matter content in soils based on OSC and PLS[J]. Transactions of the Chinese Society of Agricultural Machinery, 2007, 38(12): 113—115. (in Chinese with English abstract)
- [17] 张娴, 袁洪福, 郭峥, 等. 正交信号校正应用于多元线性回归建模的研究[J]. 光谱学与光谱分析, 2011, 31(12): 3228—3231.
- Zhang Xian, Yuan Hongfu, Guo Zheng, et al. Study on building MLR model using orthogonal signal correction[J]. Spectroscopy and Spectral Analysis, 2011, 31(12): 3228—3231. (in Chinese with English abstract)
- [18] 黄亚伟, 王加华, Jacqueline J. Shan, 等. 近红外光谱测定人参与西洋参的主要皂甙总量[J]. 分析化学, 2011, 39(3): 377—381.
- Huang Yawei, Wang Jiahua, Jacqueline J. Shan, et al. Determination of total main ginsenosides contents in american ginseng and chinese ginseng using near infrared spectroscopy[J]. Chinese Journal of Analytical Chemistry, 2011, 39(3): 377—381. (in Chinese with English abstract)
- [19] 黄昌勇. 土壤学[M]. 北京: 中国农业出版社, 2000: 77.
- [20] 刘飞, 孙光明, 何勇. 基于波长选择方法 Modeling Power 的黑木耳产地判别研究[J]. 光谱学与光谱分析, 2010, 30(1): 62—65.
- Liu Fei, Sun Guangming, He Yong. Geographical origin discrimination of Auricularia Auricula using variable selection method of modeling power[J]. Spectroscopy and Spectral Analysis, 2010, 30(1): 62—65. (in Chinese with English abstract)
- [21] 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. 化学进展, 2004, 16(4): 528—542.
- Chu Xiaoli, Yuan Hongfu, Lu Wanzhen. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique[J]. Progress in chemistry, 2004, 16(4): 528—542. (in Chinese with English abstract)

## Soil classification based on near infrared reflectance spectroscopy and orthogonal signal correction-partial least square

Song Haiyan<sup>1</sup>, Qin Gang<sup>2</sup>, Han Xiaoping<sup>1</sup>, Liu Haiqin<sup>1</sup>

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

2. College of Forestry, Shanxi Agricultural University, Taigu 030801, China)

**Abstract:** Soils with different soil textures have different spectra because of their different moisture-holding capacity and particle size. This feature provides a scientific basis for soil textural classification by Near-infrared Spectroscopy (NIR) technology. In the paper, the spectra of three different soil textural samples (sand, loam and clay soils) were analyzed. Since texture information in spectra is less than chemical information, the orthogonal signal correction (OSC) was applied in this research because it can keep down concentration information by orthogonal processing, and the partial least square (PLS) classification model was obtained. The results showed that the correlation coefficient of validation model was 0.946, and the correct recognition ratio for the three kinds of predicted samples were 93%, 86.6% and 86.6% respectively. The research indicates that OSC can extract texture information from weak spectra, so as to realize soil texture classification.

**Key words:** soils, texture, classification, OSC-PLS, near infrared spectroscopy