

北京典型耕作土壤养分的近红外光谱分析

李 颖, 张小超^{*}, 苑严伟, 张俊宁

(土壤植物机器系统技术国家重点实验室, 中国农业机械化科学研究院, 北京 100083)

摘 要: 为研究土壤养分含量分布信息, 以从北京郊区一块试验田采集的 72 个土壤样品为试验材料, 应用傅里叶变换近红外光谱技术分析了土样的全氮、全钾、有机质养分含量和 pH 值。采用偏最小二乘法 (PLS) 对光谱数据与土壤养分实测值进行回归分析, 建立预测模型, 以模型决定系数 (R^2)、校正标准差 ($RMSECV$)、预测标准差 ($RMSEP$) 和相对分析误差 (RPD) 作为模型精度的评价指标。结果表明, 利用该模型与光谱数据对土壤全氮、全钾、有机质养分含量和 pH 值进行预测, 结果与实测数据具有较好的一致性, 最高决定系数 R^2 达到 0.9544。偏最小二乘回归方法建立的养分预测模型能准确地对北京地区褐土土质全氮、有机质、全钾和 pH 值 4 种养分进行预测。

关键词: 土壤, 养分, 傅里叶变换, 近红外光谱, 偏最小二乘法

doi: 10.3969/j.issn.1002-6819.2012.02.031

中图分类号: TH83; S15

文献标志码: A

文章编号: 1002-6819(2012)-02-0176-04

李 颖, 张小超, 苑严伟, 等. 北京典型耕作土壤养分的近红外光谱分析[J]. 农业工程学报, 2012, 28(2): 176—179.
Li Jie, Zhang Xiaochao, Yuan Yanwei, et al. Analysis of soil nutrient content based on near infrared reflectance spectroscopy in Beijing region[J]. Transactions of the CSAE, 2012, 28(2): 176—179. (in Chinese with English abstract)

0 引 言

精细农业变量施肥取决于对农田的土壤养分分布信息的了解, 快速获取土壤信息是实施精细农业的基础^[1-2]。近红外光谱 (NIR) 分析方法具有快速、高效、无损和适合在线分析等优点, 近年来, 采用近红外光谱技术获取土壤养分信息, 成为国内外学者研究的重点^[3-5]。Hummel 等分析土壤水分和有机质与近红外光谱之间的关系, 利用逐步多元回归的方法建立了预测模型^[6]。Krishnan 等分析土壤有机质和近红外光谱之间的关系, 利用 4 个波长 (1 130, 1 350, 1 398, 2 210 nm) 进行多元线性回归分析建立模型^[7]。Confalonieri 等用近红外反射光谱法测定了不同土壤中的总有机碳、总氮、可交换钾和速效磷, 认为近红外光谱法可以准确测定土壤中总氮和有机碳含量, 代替常规的化学分析方法, 而可交换钾和速效磷预测效果很不理想, 但或许可以用该法进行样品分组。孙建英等应用近红外光谱分析土壤水分、全氮和有机质, 应用一元和多元线性回归建立模型^[8]。以上研究都是利用一元或多元线性回归的方法建立模型, 由于近红外吸收光谱是由重叠的合频和倍频组成的, 其单个成分吸收峰的分辨率很低^[2], 因此, 需要找到一种更为有效的回归建模方法, 使得模型计算更为简便, 预测稳定性更好。同

时, 京郊是北京市粮食、蔬菜和水果的重要供应基地, 研究京郊土壤养分的预测模型是十分必要的。研究表明, 与主成分回归相比, 偏最小二乘回归方法 (partial least squares regression, PLS) 可采用较少的因子达到最小的均方误差, 它能较好地解决一般多元线性回归难以解决的如计算复杂、多重相关等问题^[9]。本研究探讨应用偏最小二乘法 (PLS) 建立土壤养分预测模型, 并对模型进行验证。

1 材料和方法

1.1 试验材料

北京郊区县土壤覆盖率为 82%, 面积约为 137.8 万 hm^2 , 其中, 褐土面积为 89.1 万 hm^2 , 占北京地区土壤面积的近 65%。为了指导小麦播种时的变量施肥, 须首先获取田间养分的分布情况, 试验中的 72 个土样均采集于北京市昌平区中国农业机械化科学研究院的一块试验田, 土壤样品是典型的褐土土质, 具有很强的代表性。土样的采集范围为北纬 $40^\circ 09' 49.2''$ ~北纬 $40^\circ 10' 34.3''$, 东经 $116^\circ 16' 30.8''$ ~东经 $116^\circ 16' 53.5''$; 采样深度为去除表层土壤后 0~20 cm; 采样质量为 1.4~2.2 kg 之间。将采集得到的土样在室内摊开, 让其自然风干。

为了减少人为操作误差, 本次采集的土样为同属一个批次, 统一风干后的土样经过粉碎机粉碎后过 1 mm 筛并统一编号。用四分法取 A、B 2 份各 72 个土样, A 份进行常规化学测量值分析, B 份进行近红外分析, 用自封袋对土样进行密封保存。

1.2 试验仪器

试验采用德国布鲁克 (BRUKER) 公司的 MPA 傅里叶变换近红外光谱仪, 光谱的测量范围是 $12\ 500\sim 3\ 600\ \text{cm}^{-1}$, 以 $16\ \text{cm}^{-1}$ 的分辨率对样品扫描 64 次, 使用样品旋转台以增加样品扫描范围, 石英样品杯和 PbS 检测器;

收稿日期: 2011-03-08 修订日期: 2011-11-09

基金项目: 973 计划课题 (2010CB735707), 北京市科技计划资助项目 (D101105046310002, D101105046310003)

作者简介: 李 颖 (1984—), 男, 云南红河人。研究方向: 土壤光谱特性分析。北京 土壤植物机器系统技术国家重点实验室, 中国农业机械化科学研究院, 100083。Email: yyw215@163.com

*通信作者: 张小超 (1957—), 研究员, 博士生导师。研究方向: 精细农业。北京 土壤植物机器系统技术国家重点实验室, 中国农业机械化科学研究院, 100083。Email: zxc@caams.org.cn, 农业工程学会会员 (041200093S)

每个样品重复扫描 2 次, 求出平均光谱曲线。

1.3 偏最小二乘回归方法参数

最小二乘回归方法建模方法中, 衡量模型优劣的基本参数包括: 模型决定系数 (R^2), 它的大小决定了预测值与实测值之间相关的密切程度; 校正标准差 ($RMSECV$), 它反映了交叉检验时预测值与实测值之间的偏离程度; 预测标准差 ($RMSEP$), 它反映所建立的模型做预测时预测值与实测值之间的偏离程度; 相对分析误差 (RPD), 其值越高, 说明预测集验证的化学值分布越广, 回归模型的预测能力就越好。一个好的模型, 应该具较高的 R^2 和 RPD 值, 较低且接近的 $RMSECV$ 值和 $RMSEP$ 值^[10]。参数的计算公式如下

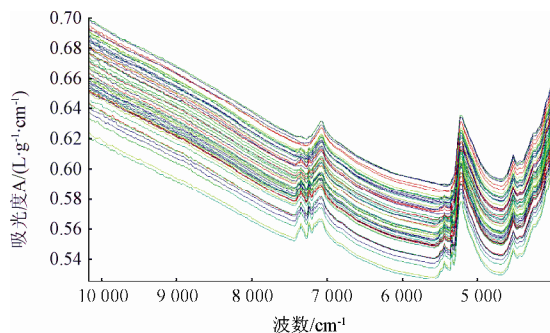
$$R^2 = \left(1 - \frac{\sum_{i=1}^n (y_i - y_{Ci})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right) \times 100\% \quad (1)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - y_{Ci})^2}{n_c}} \quad (2)$$

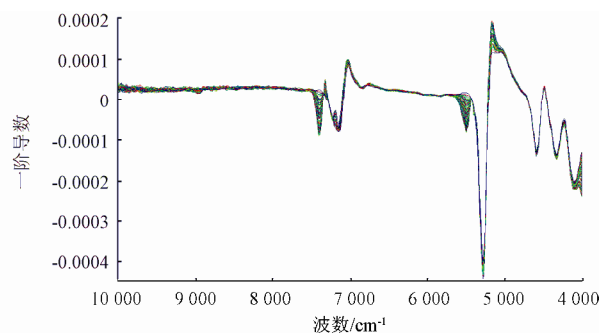
$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - y_{Pi})^2}{n_p}} \quad (3)$$

$$RPD = \frac{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}{\sqrt{\sum_{i=1}^n [(y_i - y_{Pi}) - \frac{\sum_{i=1}^n (y_i - y_{Pi})^2}{n_c}]^2}} \quad (4)$$

式中, y_i 为化学方法测量值, g/kg ; y_{Ci} 为校正集预测值, g/kg ; \bar{y} 为校正集均值, g/kg ; y_{Pi} 为预测集预测值, g/kg ; n_c 为校正集样品数; n_p 为预测集样品数。



a. 近红外吸收光谱



b. 一阶导数光谱

图 1 土壤的近红外吸收光谱和一阶导数光谱

Fig.1 NIR absorbance spectra and first derivative spectra of soil samples

2.3 模型的建立、优化和验证

2.3.1 内部交叉检验

本试验采用定标软件 OPUS, 通过 PLS 方法建立土壤的全氮、有机质、全钾和 pH 值的预测模型。从 72 个样品中随机抽取 54 个作为校正组, 采用内部交叉检验法验证模型。内部交叉检验对建模和检验使用相同的样品系列, 该方法每次剔除建模样本组中的一个或多个样本,

2 结果与分析

2.1 试验测量数据

建立模型的质量好坏很大程度取决于样品真实值的检测精度和样品检测参数的覆盖范围。检测精度越高, 所建模型误差就越小。同时, 样品检测参数范围越广, 样品分布越均匀, 所建模型的适应性就越好。本次研究所采用的土壤样品养分值测量数据均由北京农产品质量检测与农田环境监测技术研究中心提供, 表 1 给出了土样各养分测量值的分布情况。

表 1 土样校正集和预测集养分测量值分布
Table 1 Measured values of calibration group and prediction group

参数	校正集 (54)		预测集 (18)	
	测量值	均值	测量值	均值
全氮/($g \cdot kg^{-1}$)	0.87~1.15	1.02	0.88~1.13	1.06
有机质/($g \cdot kg^{-1}$)	16.02~21.94	18.13	16.86~21.29	17.48
全钾/($g \cdot kg^{-1}$)	11.69~13.50	12.54	11.74~13.46	12.43
pH 值	7.65~8.06	7.83	7.63~7.91	7.77

2.2 光谱数据

在近红外 $12\ 500 \sim 3\ 600\ cm^{-1}$ 的整个谱区内, 低于 $4\ 000\ cm^{-1}$ 的区域贡献噪声, 不存在有用信息; 同样, 高于 $10\ 000\ cm^{-1}$ 的区域无谱图信息。本试验对光谱进行了预处理, 只选择 $10\ 000 \sim 4\ 000\ cm^{-1}$ 区域进行分析, 如图 1b 所示。

同时, 由于一阶导数光谱能够消除一些系统误差和常数误差, 有效地提高建模精度^[11], 本研究对样品的每个吸收光谱都计算了一阶导数光谱, 一阶导数光谱如图 1b 所示。从图中可以看出, 一阶导数光谱很好的去除了常数误差, 并突出显示了出峰位置的差异。

用剩余样本来建模并预测被剔除样本的成分。比较被剔除样本预测值与真实值之间的差异, 由此判断模型的预测准确性。依次循环, 直到全部样品都被用于检验。内部交叉检验可以使用较少的样品量获得一个较好的预测模型, 很好地解决了土壤样品数量少的问题。本试验采用每次剔除建模样本组中的一个样品进行交叉检验建模。

2.3.2 频谱范围的选择和模型优化

PLS 回归法是一个全光谱的方法,然而,有些情况下光谱噪声或样品中出现附加组分,会造成 PLS 算法解析这些特性,使模型变坏,故有必要限定用于 PLS 回归的光谱范围^[12-15]。为了消除测量偏移和减少系统误差,需要对光谱数据进行预处理,依据 *RMSECV* 尽量小和光谱范围尽量宽的原则,选择相应的处理方法。同时,限制最大因子数为 10,以避免过拟合而降低模型质量。详见表 2。

表 2 参数模型的光谱范围选择、因子数选择和光谱预处理方法

参数	光谱范围选择/cm ⁻¹	因子数	光谱预处理方法
全氮模型	7 506~6 094.3, 5 454~4 242.9	4	一阶导数+矢量归一化 (SNV)
机质模型	7 506~6 094.3, 5 454~4 597.7	5	一阶导数
全钾模型	7 506~6 094.3, 5 454~4 242.9	3	一阶导数+矢量归一化 (SNV)
pH 值模型	5 454~4 242.9	5	一阶导数+矢量归一化 (SNV)

2.3.3 模型的建立和验证

利用校正集的 54 个样品光谱建立土壤养分含量的预测模型,同时使用所建立的预测模型对预测集的 18 个土壤样品养分含量进行预测。4 个模型相应的测量值与预测值的关系如图 2 所示。

各个模型的模型参数如表 3 所示,从表中可以看出,土壤的全氮模型 R^2 为 95.44%, *RMSECV* 为 0.014, *RPD* 为 4.68; 有机质模型 R^2 为 89.63%, *RMSECV* 为 0.37, *RPD* 为 3.11; 全钾模型 R^2 为 85.62%, *RMSECV* 为 0.173, *RPD* 为 2.64; pH 值模型 R^2 为 87.33%, *RMSECV* 为 0.031, *RPD* 为 2.81。4 个模型都获得了较好的参数指标。

利用建立的定量模型预测预测集的 18 个样品光谱。应用参数 *RMSEP* 来评价 4 个模型的预测能力,从表 3 可以看出,全氮模型的 *RMSEP* 为 0.0168,有机质模型的 *RMSEP* 为 0.473,全钾模型的 *RMSEP* 为 0.204, pH 值模型的 *RMSEP* 为 0.053,与它们相应的 *RMSECV* 值非常接近,说明 4 个模型都较好地预测了预测集 18 个样品的各个养分值。

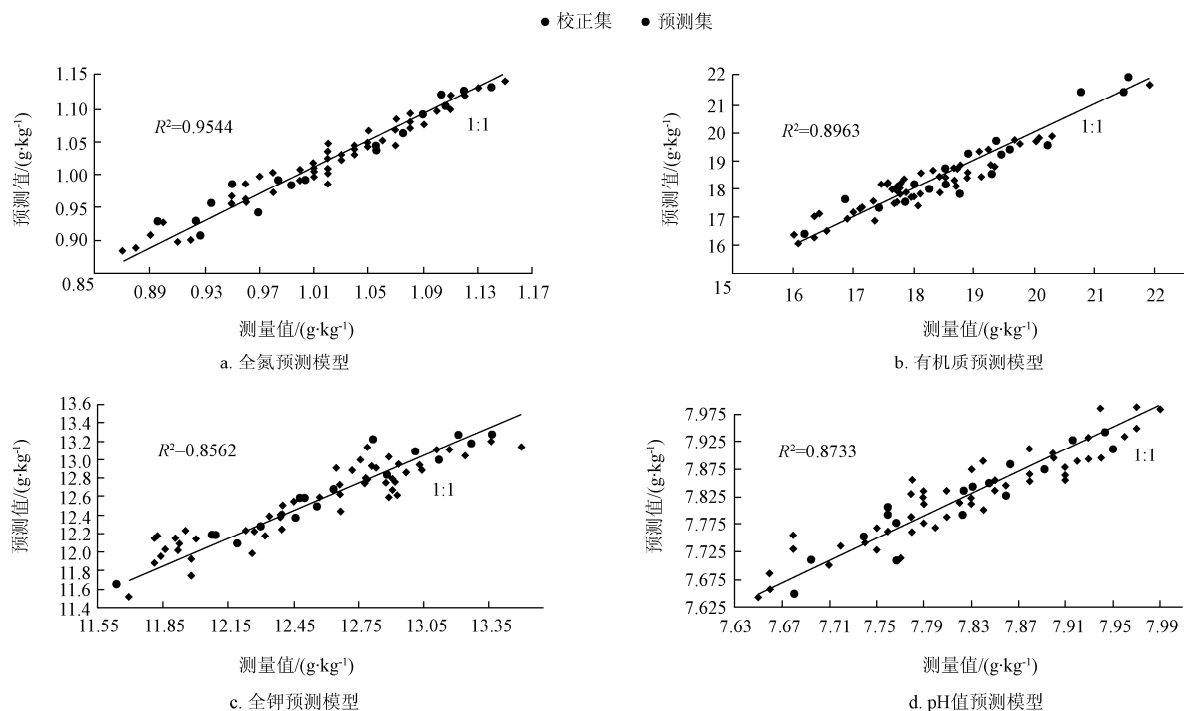


图 2 土壤养分近红外光谱预测模型

Fig.2 NIR spectroscopy prediction models of soil nutrient content

表 3 土壤养分预测模型指标

参数	R^2	<i>RMSECV</i>	<i>RMSEP</i>	<i>RPD</i>
全氮模型	0.9544	0.014	0.0168	4.68
机质模型	0.8963	0.370	0.4730	3.11
全钾模型	0.8562	0.173	0.2040	2.64
pH 值模型	0.8733	0.031	0.0530	2.81

注: R^2 为决定系数; *RMSECV* 为校正标准差; *RMSEP* 为预测标准差; *RPD* 为相对分析误差。

3 结论与讨论

以傅里叶变换近红外光谱技术为基础,利用偏最小

二乘回归方法建立了土壤全氮、有机质、全钾和 pH 值 4 种养分的预测模型。结果表明,利用该模型与光谱数据对土壤全氮、全钾、有机质养分含量和 pH 值进行预测,结果与实测数据具有较好的一致性,最高决定系数 R^2 达到 95.44%。偏最小二乘回归方法建立的养分预测模型能准确地对北京地区褐土土质全氮、有机质、全钾和 pH 值 4 种养分进行预测。

本研究探索了应用 MPA 傅里叶变换近红外光谱仪快速测定土壤中全氮、有机质、全钾和 pH 值的一种新方法,它对占北京地区大部分的褐土土质养分预测效果显著。

对于不同类型、不同季节和不同区域土壤养分的预测及与农作物生长相关的速效氮、速效磷和速效钾的快速测定方法有待在此基础上作进一步的研究。

[参 考 文 献]

- [1] 汪懋华. “精细农业”发展与工程技术创新[J]. 农业工程学报: 1999, 15(1): 1—8.
Wang Maohua. Development of precision agriculture and innovation of engineering technologies[J]. Transactions of the CSAE, 1999, 15(1): 1—8. (in Chinese with English abstract)
- [2] Jerry W J, Lois W. Practical Guide To Interpretive Near-infrared Spectroscopy[M]. Boca Raton: American CRC Press, 2007.
- [3] Ben-Dor E, Banin A. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties[J]. Soil Science of American Journal, 1995, 59(2): 364—372.
- [4] Slaughter D C, Pelletier M G, Upadhyaya S K. Sensing soil moisture using NIR spectroscopy[J]. Applied Engineering of Agriculture, 2001, 17(2): 241—247.
- [5] 陆婉珍. 现代近红外光谱分析技术 (第2版) [M]. 北京: 中国石化出版社, 2006.
- [6] Hummel J W, Sudduth K A, HolliIger S E. Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor[J]. Computers and Electronics in Agriculture, 2001, 32(2): 149—165.
- [7] 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.
- [8] 孙建英. 基于近红外光谱的北方潮土土壤参数实时分析[J]. 光谱学与光谱分析, 2006, 26(3): 426—429.
Sun Jianying. Real-time analysis of soil moisture, soil organic matter, and soil total nitrogen with NIR spectra[J]. Spectroscopy and Spectral Analysis, 2006, 26(3): 426—429. (in Chinese with English abstract)
- [9] 王惠文. 偏最小二乘回归方法及其应用[M]. 北京: 国防工业出版社, 1999.
- [10] 严衍禄. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005, 1.
- [11] 严衍禄, 赵龙莲, 李军会, 等. 现代近红外光谱分析的信息处理技术[J]. 光谱学与光谱分析, 2000, 20(6): 777—780.
Yan Yanlu, Zhao Longlian, Li Junhui, et al. Information technology of modern NIR spectral analysis[J]. Spectroscopy and Spectral Analysis, 2000, 20(6): 777—780. (in Chinese with English abstract)
- [12] He Yong, Song Haiyan, Pereira Annia Garcia. A new approach to predict N, P, K and OM content in a loamy mixed soil by using near infrared reflectance spectroscopy[J]. ICIC, 2005, 3644: 859—867.
- [13] 何绪生. 近红外反射光谱分析在土壤学的应用及前景[J]. 中国农业科技导报, 2004, 6(4): 71—76.
He Xusheng. Advance and prospect of near infrared reflectance spectroscopy applied in soil science[J]. Journal of Agricultural Science and Technology, 2004, 6(4): 71—76. (in Chinese with English abstract)
- [14] Ehsani M R, Upanhyaya S K, Slaughter D. A NIR technique for rapid determination of soil mineral nitrogen[J]. Precision Agriculture, 1999(2): 217—234.
- [15] 傅霞萍, 应义斌, 刘燕德, 等. 水果坚实度的近红外光谱检测分析试验研究[J]. 光谱学与光谱分析, 2006, 26(6): 1038—1041.
Fu Xiaping, Ying Yibin, Liu Yande, et al. Detection of pear firmness using near infrared diffuse reflectance spectroscopy[J]. Spectroscopy and Spectral Analysis 2006, 26(6): 1038—1041. (in Chinese with English abstract)

Analysis of soil nutrient content based on near infrared reflectance spectroscopy in Beijing region

Li Jie, Zhang Xiaochao^{*}, Yuan Yanwei, Zhang Junning

(National Key Laboratory of Soil-Plant-Machine System, Chinese Academy of Agricultural Mechanization Sciences, Beijing 100083, China)

Abstract: To study the distribution of soil nutrients, Fourier transform infrared spectroscopy techniques were used to predict total nitrogen, organic matter, total potassium and pH values of soil. With 72 soil samples collected from the experimental field in the suburbs of Beijing, the models were constructed using partial least-squares (PLS) regression based on the spectral data and measured soil nutrient. The model accuracy was evaluated using determination coefficient (R^2), adjusted standard deviation ($RMSECV$), standard deviation of prediction ($RMSEP$), and residual prediction deviation (RPD). The results showed that, good consistency can be found between the prediction models and the spectral data of total nitrogen, total potassium, organic matter and pH value, the results and the measured data has, the highest coefficient of determination is $R^2=0.9544$. Nutrient prediction model established by PLS could predict total nitrogen, organic matter, and total potassium and pH values accurately.

Key words: soils, nutrients, Fourier transforms, near infrared spectroscopy, partial least squares