

# 基于近红外光谱的土壤全氮含量估算模型

张娟娟<sup>1,2</sup>, 田永超<sup>2</sup>, 姚霞<sup>2</sup>, 曹卫星<sup>2</sup>, 马新明<sup>1</sup>, 朱艳<sup>2\*</sup>

(1. 河南农业大学信息与管理科学学院, 郑州 450002; 2. 南京农业大学国家信息农业工程技术中心, 江苏省信息农业高技术研究重点实验室, 南京 210095)

**摘要:** 土壤全氮是诊断土壤肥力水平和指导作物精确施肥所需的重要信息, 建立土壤全氮的近红外光谱估测模型并对建模波段进行优化选择对于土壤养分信息快速获取和精确农业发展具有重要意义。该研究以中国中、东部地区 5 种主要类型土壤为研究对象, 利用近红外光谱仪采集土壤样品的光谱信息, 结合近红外区域分子振动特点选取全谱、合频、一倍频、二倍频和 N-H 基团及其组合的 8 个波段, 采用多元散射校正等多种预处理方法组合进行处理, 结合偏最小二乘法 (PLS) 对每个波谱区域进行定标建模。结果表明, 利用 4 000~5 500  $\text{cm}^{-1}$  波谱区域结合附加散射校正处理过的原始光谱建立的模型精度表现最好, 其内部互验证决定系数达到 0.90, 均方根误差为 0.16。经不同类型土壤的观测资料检验, 模型验证决定系数为 0.91, 均方根误差为 0.15, 相对分析误差 *RPD* 为 3.40, 表明模型具有极好的预测能力。因此, 利用近红外光谱可以实现土壤全氮的快速估测, 且以合频波段 (4 000~5 500  $\text{cm}^{-1}$ ) 为建模区域可以得到更好的预测效果。

**关键词:** 土壤, 全氮, 近红外光谱, 模型, 偏最小二乘法

doi: 10.3969/j.issn.1002-6819.2012.12.030

中图分类号: S153.6; TN219

文献标志码: A

文章编号: 1002-6819(2012)-12-0183-06

张娟娟, 田永超, 朱艳, 等. 基于近红外光谱的土壤全氮含量估算模型[J]. 农业工程学报, 2012, 28(12): 183-188.  
Zhang Juanjuan, Tian Yongchao, Yao Xia, et al. Estimating model of soil total nitrogen content based on near-infrared spectroscopy analysis[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(12): 183-188. (in Chinese with English abstract)

## 0 引言

快速准确地获取土壤养分信息是精确农业的关键之一。土壤氮素丰缺状况直接影响作物的生长状况及产量品质形成。Kjeldahl 或 Bremner 法等传统土壤氮测定方法耗时、费工、有污染, 且时效性弱<sup>[1]</sup>, 难以满足实时监测土壤全氮含量的需要, 开发土壤相关参数测定的精密快速的实时分析仪器显得尤为重要。

近红外反射光谱 (NIRS) 技术具有快速高效监测的特点, 已在农业及其它许多领域得到广泛应用<sup>[2-4]</sup>。通过分析土壤全氮含量与土壤近红外反射光谱间的关系建立监测模型, 具有较高的预测精度<sup>[5-13]</sup>。然而, 以往土壤近红外光谱模型研究多以全波段区域的光谱信息建立多变量定标模型, 虽可以获得较多的样品成分信息, 但同时容易受到冗余无关信息和非线性信息的干扰, 建立的模型不稳定, 很少在实际生产中推广应用。因此筛选在近红外区域与土壤全氮含量相关性

较高的波段并构建其估测模型, 对于快速获取土壤养分信息, 实施作物肥料精确管理尤为重要。部分研究基于近红外光谱和高光谱技术, 阐明了土壤有机质与可见光-近红外光谱之间的关系, 并利用光谱指数等方法建立了有机质的光谱估测模型<sup>[14-16]</sup>, 同时, 基于 ASD FieldSpec Pro 地物高光谱数据<sup>[17]</sup>, 比较了不同建模方法对土壤全氮估测的效果。但利用 NIRS 对土壤全氮进行预测时, 如何结合近红外区域分子振动的特点, 基于不同波谱区域结合不同的预处理方法进行土壤全氮近红外建模的研究还鲜有报道。

为此, 本研究利用傅立叶 NIRS 技术采集不同类型土壤样品的光谱信息, 运用偏最小二乘法结合不同的预处理方法优化选择土壤全氮的建模区域, 建立适合 5 种土壤样本全氮含量的适宜估测模型, 从而为土壤养分信息获取和肥料精确管理提供技术支持。

## 1 材料与方法

### 1.1 土壤光谱采集

运用 Antaris 型傅立叶变换近红外光谱仪<sup>[14]</sup>进行土壤光谱采集。测试方法参见参考文献[14]。

### 1.2 土壤样品与全氮化学分析

试验土壤样品采集于我国中、东部地区郑州、滑县、宿州、南京、杭州、大丰和上海等地, 共包括 5 种主要土壤类型: 水稻土、潮土、砂浆黑土、盐潮土和盐土。土壤样品分别经过风干、磨碎及过筛处理供光谱分析和化学测试使用。土壤样本基本信息和土壤全氮含量的特

收稿日期: 2011-06-22 修订日期: 2012-05-14

基金项目: 国家自然科学基金项目 (30900868); 国家 863 计划项目 (2010AA10A301); 江苏高校优势学科建设工程资助项目 (PAPD); 河南科技攻关项目 (112102110030) 资助

作者简介: 张娟娟 (1979-), 女, 河南博爱人, 博士, 讲师, 研究方向为农业遥感监测。郑州 河南农业大学信息与管理科学学院, 450002。

Email: zhangjuan\_2003@126.com.

\*通信作者: 朱艳 (1976-), 女, 江苏南通人, 博士, 教授, 研究方向为信息农业。南京 南京农业大学/国家信息农业工程技术中心, 210095。

Email: yanzhu@njau.edu.cn.

征统计参见文献<sup>[15-17]</sup>。

### 1.3 光谱预处理

利用 Savitzky-Golay 7 点平滑<sup>[18]</sup>结合一阶导数和二阶导数来移除噪音对光谱曲线的影响,此外,较多研究<sup>[19-20]</sup>表明附加散射校正(multiplicative scatter correction, MSC)可以消除由于样品颗粒不均匀造成的装样误差和颗粒散射对光谱带来的影响,降低土壤样品的不均匀性带来的光谱差异。本研究通过比较选取最优的方法进行建模,进而确定最优的建模波谱区域。

### 1.4 数据分析方法

采用偏最小二乘法进行建模,先用校正样品集(建模样本,共 331 个)建立校正模型,再做交叉检验(采用 Leave-one-out 方法),最后采用独立的数据对模型进行外部验证(测试样本,共 129 个),建模和测试样本数据的选取参见文献<sup>[17]</sup>。具体方法:1)选取近红外区域全谱、合频、一倍频、二倍频和 N-H 基团及其组合的 8 个波段(如图 1),采用偏最小二乘法对每个波段分别建立定标模型;2)为了提高模型精度,对不同预处理下的光谱进行定标建模,进而确定定标效果最好的区域。采用以下指标评定模型效果:内部交互验证决定系数  $Rcv^2$  及均方根误差  $RMSECV$ 、外部验证决定系数  $Rv^2$ 、预测均方根误差  $RMSEV$  及预测相对分析误差  $RPD$ 。各指标的评价方法参见<sup>[21]</sup>。

## 2 结果与分析

### 2.1 土壤全氮含量与近红外光谱的相关分析

图 1 给出不同类型土壤的光谱特征曲线。不同类型土壤整体表现平缓,在近红外区域的光谱曲线在  $7\ 142\ cm^{-1}$  ( $1\ 400\ nm$ )、 $5\ 263\ cm^{-1}$  ( $1\ 900\ nm$ ) 和  $4\ 545\ cm^{-1}$  ( $2\ 200\ nm$ ) 存在有明显的吸收峰,这主要是由于水分的吸收引起的。由于土壤全氮的标志基团 N-H 基团在近红外谱区的特征吸收波长主要有  $9\ 800\sim 9\ 430$ 、 $7\ 000\sim 6\ 370$ 、 $5\ 100\sim 4\ 360\ cm^{-1}$ (参见图 1),因此该区域也作为土壤全氮定标建模区域。土壤全氮含量与近红外光谱吸光度之间的相关系数见图 2。从图上可以看出,吸光度与全氮含量化学值的相关系数绝对值低于 0.32,表明多种类型土壤样本的全氮含量与单波长点的吸光度相关不显著,必须考虑用多个谱区内的光谱数据来建立定标模型。

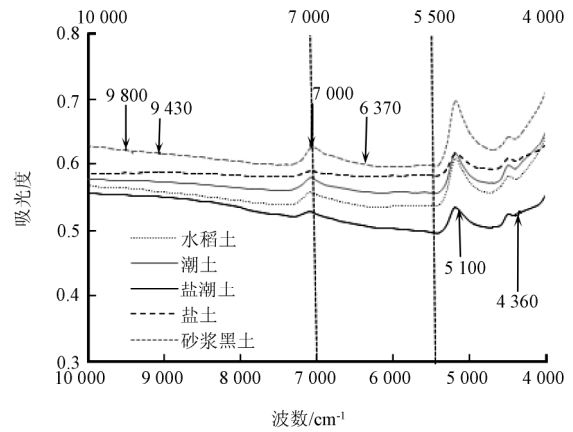


图 1 不同类型土壤的近红外吸收光谱  
Fig.1 Absorption spectra of different soil types

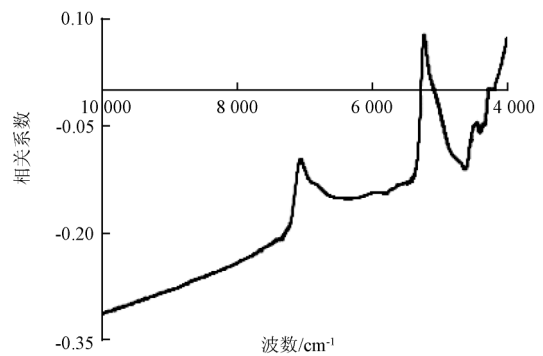


图 2 土壤全氮含量与近红外光谱吸光度之间的相关系数  
Fig.2 Correlation coefficient between total nitrogen content and spectrum for absorbance of soil samples

### 2.2 不同区域 PLS 校正模型

从图 1 土壤样品在仪器测定范围 ( $4\ 000\sim 10\ 000\ cm^{-1}$ ) 内的吸收光谱可知,对不同的区域进行组合共有 8 个区域:全谱( $4\ 000\sim 10\ 000\ cm^{-1}$ )、一倍频( $5\ 500\sim 7\ 000\ cm^{-1}$ )、二倍频( $7\ 000\sim 10\ 000\ cm^{-1}$ )、合频( $4\ 000\sim 5\ 500\ cm^{-1}$ )、一倍频+合频( $5\ 500\sim 7\ 000\ cm^{-1}+4\ 000\sim 5\ 500\ cm^{-1}$ )、二倍频+合频( $7\ 000\sim 10\ 000\ cm^{-1}+4\ 000\sim 5\ 500\ cm^{-1}$ )、一倍频+二倍频( $5\ 500\sim 7\ 000\ cm^{-1}+7\ 000\sim 10\ 000\ cm^{-1}$ ),另外还有“N-H 基团”区域。为提高模型精度,对光谱进行不同的预处理,包括附加散射校正、导数及平滑处理,以确定最优的建模波段。表 2 为不同预处理方法下 8 个建模区域的定标及外部验证结果。

表 2 基于偏最小二乘法建立的不同区域及预处理方法的土壤全氮含量的标定模型及检验结果

Table 2 Calibration and validation results for spectral models of N developed by PLS with different wavebands and pretreatment methods (n=331)

建模区域 波数/cm <sup>-1</sup>	预处理方法	主因子个数	标定模型		外部验证		
			决定系数	均方根误差	决定系数	均方根误差	相对分析误差
4 000~10 000	原始光谱	14	0.9	0.15	0.85	0.18	2.75
	SG+FD	8	0.76	0.27	0.74	0.22	2.27
	SG+SD	4	0.54	0.37	0.56	0.27	1.88
	MSC+LOG	12	0.9	0.16	0.9	0.16	3.11
	MSC+SG+FD	9	0.83	0.23	0.79	0.18	2.73
	MSC+SG+SD	5	0.59	0.36	0.58	0.25	2

续表

建模区域 波数/cm <sup>-1</sup>	预处理方法	主因子个数	标定模型		外部验证		
			决定系数	均方根误差	决定系数	均方根误差	相对分析误差
10 000~7 000	原始光谱	9	0.78	0.26	0.74	0.24	2.08
	SG+FD	4	0.56	0.37	0.55	0.28	1.8
	SG+SD	4	0.34	0.45	0.26	0.37	1.34
	MSC+LOG	9	0.79	0.25	0.78	0.2	2.54
	MSC+SG+FD	4	0.66	0.32	0.65	0.24	2.13
	MSC+SG+SD	4	0.4	0.43	0.22	0.37	1.36
7 000~5 500	原始光谱	7	0.79	0.25	0.65	0.28	1.79
	SG+FD	3	0.44	0.42	0.19	0.39	1.27
	SG+SD	2	0.22	0.5	0.04	0.47	1.08
	MSC+LOG	7	0.77	0.26	0.62	0.24	2.07
	MSC+SG+FD	4	0.59	0.35	0.48	0.29	1.73
	MSC+SG+SD	2	0.26	0.48	0.06	0.4	1.24
5 500~4 000	原始光谱	12	0.91	0.16	0.82	0.19	2.66
	SG+FD	9	0.85	0.21	0.76	0.22	2.28
	SG+SD	6	0.53	0.38	0.59	0.25	1.98
	MSC+LOG	13	0.9	0.16	0.91	0.15	3.4
	MSC+SG+FD	9	0.88	0.19	0.82	0.18	2.78
	MSC+SG+SD	5	0.58	0.36	0.6	0.24	2.07
7 000~4 000	原始光谱	13	0.92	0.16	0.82	0.19	2.69
	SG+FD	8	0.83	0.23	0.76	0.22	2.24
	SG+SD	4	0.51	0.39	0.56	0.27	1.89
	MSC+LOG	12	0.91	0.16	0.84	0.16	3.23
	MSC+SG+FD	9	0.86	0.21	0.79	0.19	2.62
	MSC+SG+SD	5	0.58	0.36	0.6	0.24	2.07
10 000~5 500	原始光谱	10	0.85	0.21	0.75	0.23	2.19
	SG+FD	4	0.57	0.36	0.55	0.28	1.82
	SG+SD	4	0.36	0.45	0.26	0.37	1.35
	MSC+LOG	12	0.85	0.22	0.82	0.18	2.78
	MSC+SG+FD	4	0.67	0.32	0.66	0.23	2.2
	MSC+SG+SD	4	0.42	0.42	0.23	0.36	1.39
(10 000~7 000) +5 500~4 000)	原始光谱	13	0.92	0.16	0.84	0.19	2.66
	SG+FD	8	0.76	0.27	0.74	0.22	2.24
	SG+SD	5	0.55	0.38	0.52	0.3	1.65
	MSC+LOG	12	0.92	0.16	0.86	0.16	3.07
	MSC+SG+FD	10	0.84	0.22	0.79	0.18	2.75
	MSC+SG+SD	5	0.58	0.36	0.58	0.25	2.01
9 800~9 430、 7 000~6 370、 5 100~4 360	原始光谱	11	0.86	0.2	0.75	0.21	2.43
	SG+FD	8	0.74	0.28	0.6	0.29	1.74
	SG+SD	4	0.49	0.4	0.53	0.27	1.85
	MSC+LOG	10	0.88	0.19	0.81	0.17	2.94
	MSC+SG+FD	8	0.84	0.22	0.6	0.26	1.93
	MSC+SG+SD	7	0.46	0.41	0.28	0.38	1.32

注：MSC 为多元散射校正，multiplicative scatter correction；SG 为 Savitzky-Golay 7 点平滑，Savitzky-Golay 7 data point filter with 3rd-order polynomial；FD 为一阶导数，first derivative；SD 为二阶导数，second derivative；LOG 为吸光度，absorbance。

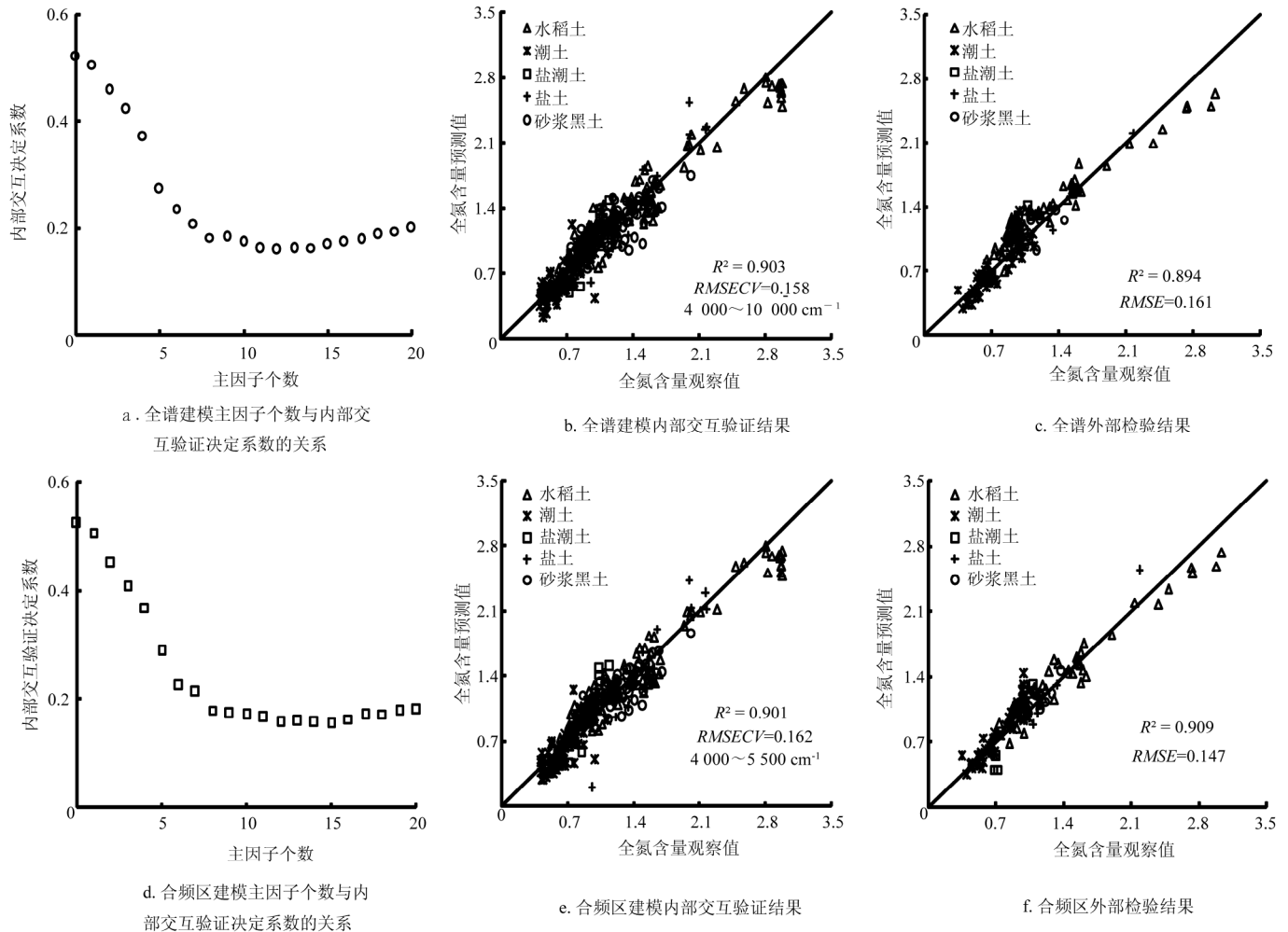
从表 2 可以看出，不同的预处理方法对建模及检验结果有较大影响，相比较原始光谱而言，一阶导数和二阶导数并没有增加模型的预测能力，分析原因可能是经过导数变换后虽然使谱图的肩峰更加突出，但却影响了主因子有效个数的选取，经过一阶导数和二阶导数变换后，最大选取的主因子数为 9。经过附加散射校正后，原始光谱、一阶导数光谱及二阶导数比未经变换前获得了

更好的预测能力，特别是多元散射校正后原始光谱的预测能力更好，而二阶导数建模效果仍然表现较差，如在 4 000~10 000 cm<sup>-1</sup> 范围内经过多元散射校正和二阶导数处理后相对分析误差 RPD 为 2.00，而附加散射校正后原始光谱建模验证 RPD 达到 3.11。

从 8 个不同的建模区域看，预测能力最差的建模区域为一倍频 7 000~5 500 cm<sup>-1</sup>，除经过多元散射校正光谱

$RPD > 2$  外,其余预处理方法下标定模型的外部验证  $RPD$  均小于 2; 其次是二倍频区 ( $7000 \sim 10000 \text{ cm}^{-1}$ ), 经过附加散射校正后的原光谱  $RPD$  为 2.54; 其余区域预测能力相对较好, 建模精度由高到低分别是  $4000 \sim 5500$ 、 $4000 \sim 7000$ 、 $4000 \sim 10000$  和  $7000 \sim 10000 + 4000 \sim 5500$ 、 $10000 \sim 5500 \text{ cm}^{-1}$  以及基团区域 ( $9800 \sim 9430$ 、 $7000 \sim 6370$ 、 $5100 \sim 4360 \text{ cm}^{-1}$ ), 均表现为经过附加散射校正后的光谱预测能力最好, 其次是原光谱。最好的建模区域为  $5500 \sim 4000 \text{ cm}^{-1}$ , 经过附加散射校正的原始光谱内部交互验证决定系数达到 0.90,  $RMSE_{CV}$  为

0.162, 验证决定系数达到 0.91, 验证  $RPD$  为 3.40, 说明模型具有极好的预测能力。此外在  $4000 \sim 7000$ 、 $4000 \sim 10000 \text{ cm}^{-1}$  区域, 经附加散射校正后的原始光谱也获得了较好的预测结果。图 3 为采用全谱 ( $4000 \sim 10000 \text{ cm}^{-1}$ ) 和合频 ( $4000 \sim 5500 \text{ cm}^{-1}$ ) 的标定及外部验证结果, 可以看出, 采用全波段和合频波段建立的模型精度基本相同, 但全波段外部验证时的结果并不稳定, 预测均方根误差为 0.161, 大于采用合频波段的外部验证结果, 可见采用合频波段可以代替全谱波段获得更好的定标效果。



注: 全谱为  $4000 \sim 10000 \text{ cm}^{-1}$ ; 合频区为  $4000 \sim 5500 \text{ cm}^{-1}$ 。

图 3 基于全谱和合频区建立的土壤全氮含量标定模型预测值和化学值的比较

Fig.3 Comparison of measured and estimated values of soil TN contents based on best models obtained for whole spectral region and combination region

### 3 讨论

以往较多研究试图通过选取具体的波段对全氮进行预测, 如 Dalal 等<sup>[7]</sup>用 NIRS 预测澳大利亚风干磨筛过土壤的总氮含量, 筛选 1744、1870 和 2052 nm 波长预测土壤全氮的结果良好; Lee 等<sup>[22]</sup>研究认为土壤氮的吸收波段在 510 nm 左右, 但不同的土壤类型有较大差异。本文通过对不同类型土壤光谱吸光度与全氮含量的相关性分析可以看出: 不同类型土壤全氮吸收的敏感波段有较大差异, 当所有样本放在一起与后, 在每一个波长点上吸

光度与全氮含量化学值的相关系数都很低 (绝对值低于 0.32)。因此, 对多种类型土壤样本直接从谱图中选择一个波长点来确定其含量是不可行的。

不同类型土壤由于质地等因素影响, 即使经过磨样、过筛等处理后仍然存在颗粒大小不一致等, 因此, 在测试光谱的过程中受到散射影响从而导致土壤待测成分与光谱之间的相关信息被掩盖。在本文中, 经附加散射校正后的原光谱能获得了更好地预测能力,  $RPD$  达到 3 以上。表明利用附加散射校正可以消除土壤在近红外光谱测试过程中由于颗粒、环境等不一致而引起的误差, 提

高光谱分辨率和信噪比。此外, 以往较多研究采用全波段光谱建模, 本文根据近红外区域分子振动的特点, 采用偏最小二乘法对不同区域的建模结果作了比较分析。结果显示采用合频波段 ( $4\ 000\sim 5\ 500\ \text{cm}^{-1}$ ) 的定标效果比全谱波段的定标效果要好, 经过附加散射校正的吸光度建模内部交互验证决定系数达到 0.90,  $RMSECV$  为 0.16, 验证决定系数达到 0.91, 验证  $PRD$  为 3.40, 表明模型具有极好的预测能力。因此, 对于土壤全氮的近红外光谱分析, 用  $4\ 000\sim 5\ 500\ \text{cm}^{-1}$  的光谱可以代替全谱获取更好的效果。

#### 4 结 论

利用傅里叶变换型近红外光谱仪对 5 种不同土壤类型 460 个土壤样品采集光谱。根据近红外区域分子振动的特点结合偏最小二乘法分析建立了不同波谱区域的土壤全氮近红外监测模型, 结果表明: 基于合频波段 ( $4\ 000\sim 5\ 500\ \text{cm}^{-1}$ ) 的建立的土壤全氮近红外光谱估测模型预测效果表现最好, 经过附加散射校正的原始光谱建模内部交互验证决定系数达到 0.90, 预测均方根误差为 0.16, 验证决定系数达到 0.91, 验证相对分析误差  $RPD$  为 3.40, 表明模型具有极好的预测能力。

#### [参 考 文 献]

- [1] 李生秀. 关于土壤供氮指标的研究: I. 对几种测定土壤供氮能力方法的评价[J]. 土壤学报, 1990, 27(3): 233—240. Li Shengxiu. Studies on the indices of soil nitrogen supplying capacities[J]. Acta Pedologica Sinica, 1990, 27(3): 233—240. (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(14): 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(14): 81—87. (in Chinese with English abstract)
- [3] 王静, 刘湘南, 黄方, 等. 基于 ANN 技术和高光谱遥感的盐渍土盐分预测[J]. 农业工程学报, 2009, 25(12): 161—165. Wang Jing, Liu Xiangnan, Huang Fang, et al. Salinity forecasting of saline soil based on ANN and hyperspectral remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2009, 25(12): 161—165. (in Chinese with English abstract)
- [4] 石吉勇, 邹小波, 赵杰文, 等. 近红外光谱技术快速无损诊断黄瓜植株氮、镁元素亏缺[J]. 农业工程学报, 2011, 27(8): 283—287. Shi Jiyong, Zhou Xiaobo, Zhao Jiewen, et al. Rapid and non-destructive diagnostics of nitrogen and magnesium deficiencies in cucumber plants by near-infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2011, 27(8): 283—287. (in Chinese with English abstract)
- [5] Chang C, Laird D A. Near-infrared reflectance spectroscopic analysis of soil C and N[J]. Soil Science, 2002, 167(2): 110—116.
- [6] Reeves III J B, McCarty G W, Mimmo T. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils[J]. Environmental Pollution, 2002, 116(1): 277—284.
- [7] Dalal R C, Henry R J. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry[J]. Soil Science Society of America Journal, 1986, 50: 120—123.
- [8] Bernard G, Barthe's. 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] Brunet D, Bernoux M, Barthe's B G. Comparison between predictions of C and N contents in tropical soils using a Vis-NIR spectrometer including a fibre-optic probe versus a NIR spectrometer including a sample transport module[J]. Biosystems engineering, 2008, 100: 448—452.
- [10] Reeves III J B, McCarty G W, Reeves V B. Mid-infrared diffuse reflectance spectroscopy for the quantitative analysis of agricultural soils[J]. Food Chemistry, 2001, 49(2): 766—772.
- [11] 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)
- [12] Brunet D, Barthe's B G, Chotte J L, et al. Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity[J]. Geoderma, 2007, 139(1/2): 106—117.
- [13] Ludwig B, Khanna P K, Bauhus J, et al. Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability[J]. Forest Ecology and Management, 2002, 171(1/2): 121—132.
- [14] 张娟娟, 田永超, 朱艳, 等. 一种预测土壤有机质含量的近红外光谱参数[J]. 应用生态学报, 2009, 20(8): 1896—1904. Zhang Juanjuan, Tian Yongchao, Zhu Yan, et al. A novel spectral index for estimating soil organic matter content using near-infrared spectroscopy[J]. Chinese Journal of Applied Ecology, 2009, 20(8): 1896—1904. (in Chinese with English abstract)
- [15] 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研究[J]. 中国农业科学, 2009, 42(9): 3154—3163. Zhang Juanjuan, Tian Yongchao, Zhu Yan, et al. Spectral characteristics and estimating of organic matter contents of different soil types[J]. Scientia Agricultura Sinica, 2009, 42(9): 3154—3163. (in Chinese with English abstract)
- [16] 田永超, 张娟娟, 姚霞, 等. 基于近红外光声光谱的土壤有机质含量定量建模方法[J]. 农业工程学报, 2012, 28(1): 145—152. Zhang Juanjuan, Tian Yongchao, Yao Xia, et al. Quantitative modeling method of soil organic matter content based on

- near-infrared photoacoustic spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(1): 145—152. (in Chinese with English abstract)
- [17] 张娟娟, 田永超, 姚霞, 等. 基于高光谱的土壤全氮含量估测[J]. 自然资源学报, 2011, 26(5): 881—889.  
Zhang Juanjuan, Tian Yongchao, Yao Xia, et al. Estimating soil total nitrogen content based on hyperspectral analysis technology[J]. Journal of Natural Resources, 2011, 26(5): 881—889. (in Chinese with English abstract)
- [18] Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least-squares procedures[J]. Analytical Chemistry, 1964, 36(8): 1627—1639.
- [19] 赵强, 张工力, 陈星旦. 多元散射校正对近红外光谱分析定标模型的影响[J]. 光学精密工程, 2005, 13(1): 53—58.  
Zhao Qiang, Zhang Gongli, Chen Xingdan. Effects of multiplicative scatter correction on a calibration model of near infrared spectral analysis[J]. Optics and Precision Engineering, 2005, 13(1): 53—58. (in Chinese with English abstract)
- [20] 丁海泉, 卢启鹏, 朴仁官, 等. 土壤有机质近红外光谱分析组合波长的优选[J]. 光学精密工程, 2007, 15(12): 1946—1951.  
Ding Haiquan, Lu Qipeng, Piao Renguan, et al. Optimum choice of combination wavelengths in near infrared analysis for soil organic matter[J]. Optics and Precision Engineering, 2007, 15(12): 1946—1951. (in Chinese with English abstract)
- [21] Chang C W, Laird D A, Mausbach M J, et al. Near-infrared reflectance spectroscopy principal components regression analysis of soil properties[J]. Soil Science Society of America Journal, 2001, 65(2): 480—490.
- [22] Lee W S, Sanchez J F, Mylavarapu R S, et al. Estimating chemical properties of Florida soil using spectral reflectance[J]. Transactions of the ASAE, 2003, 46(5): 1443—1453.

## Estimating model of soil total nitrogen content based on near-infrared spectroscopy analysis

Zhang Juanjuan<sup>1,2</sup>, Tian Yongchao<sup>2</sup>, Yao Xia<sup>2</sup>, Cao Weixing<sup>2</sup>, Ma Xinming<sup>1</sup>, Zhu Yan<sup>2\*</sup>

(1. College of Information and Management, Henan Agricultural University, Zhengzhou 450002, China; 2. Jiangsu Key Laboratory for Information Agriculture, National Engineering and Technology Center for Information Agriculture, Nanjing Agricultural University, Nanjing 210095, China)

**Abstract:** Soil total nitrogen content is regarded as an important index for soil fertility diagnosis and crop growth management direction, the rapid monitoring model and establishing optimum bands on soil nitrogen content is the key to swiftly obtain soil nutrient information and develop precision agriculture. Five types of soil (Paddy soil, Fluvo-aquic soil, Salinized fluvo-aquic soil, Saline soil, Dark soil with lime concretion) collected from the middle and eastern of China were measured in the near-infrared region (4 000-10 000  $\text{cm}^{-1}$ ). According to molecular vibration characteristics of full spectrum in the near infrared region, 8 wave bands (the whole region, combination region, the first overtone region, the second overtone region including the combination of bands referred above, and the protein functional group bands) were designed for establishing calibration models adopting multiplicative scatter correction(MSC) method. The calibration models for nitrogen content in soil samples were established by Partial Least Square (PLS) regression. The results showed that the prediction effect using MSC spectra in the combination region of 4 000-5 500  $\text{cm}^{-1}$  was the best, the coefficient of determination and the root mean square error of cross-validation (RMSECV) for the calibration model were 0.90 and 0.16, respectively. Testing of the monitoring models with independent data of different types of soil samples indicated that determination coefficient and RMSE of validation were 0.91 and 0.15, respectively; *RPD* was 3.40. So, near-infrared diffuse reflection spectroscopy analysis technology can swiftly estimate total nitrogen content, and taking the combination band (4 000-5 500  $\text{cm}^{-1}$ ) as modeling regional can get better calibration effect.

**Key words:** soils, nitrogen, near infrared spectroscopy, models, partial least squares