

文章编号 1004-924X(2007)12-1946-06

土壤有机质近红外光谱分析组合波长的优选

丁海泉^{1,2}, 卢启鹏¹, 朴仁官¹, 陈星旦¹

- (1. 中国科学院 长春光学精密机械与物理研究所 应用光学国家重点实验室, 吉林 长春 130033;
2. 中国科学院 研究生院, 北京 100039)

摘要:以分析 85 份吉林省北部黑质土壤样品有机质含量为例,研究了利用多元散射校正和相关光谱法对定标波长的优选,并根据经验和参考波长理论对所建立的模型进行逐步优化。最终得到定量模型的相关系数为 0.90,定标准差(RMSEC)为 0.36%,交叉检验标准差(RMSECV)为 0.40%,该结果与采用偏最小二乘算法(PLS)对土壤连续谱进行定标的结果相当。结果表明,利用该方法建立的模型简单、稳定,可以为吉林省北部黑土质土壤养分状况普查和便携仪器的研制提供参考。

关键词:土壤养分;近红外光谱技术;精准农业;多元散射校正;相关光谱

中图分类号:TH744.1 **文献标识码:**A

Optimum choice of combination wavelengths in near infrared analysis for soil organic matter

DING Hai-quan^{1,2}, LU Qi-peng¹, PIAO Ren-guan¹, CHEN Xing-dan¹

- (1. *State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;*
2. *Graduate School of the Chinese Academy of Sciences, Beijing 100039, China*)

Abstract: By taking the 85 black soil samples in north of Jilin province as the examples of analyzing organic matter, a optimum choosing method of combination wavelengths for calibration was presented by the ways of Multiplicative Scatter Correction (MSC) and correlation spectrum, and the model was optimized progressively based on experience and reference wavelength theory. Finally, the model shows the correlation coefficient of 0.90, Root Mean Square Error of Calibration(RMSEC) of 0.36%, and Root Mean Square Error of Calibration Value(RMSECV) of 0.40%, which approaches to the calibration of continuous spectrum by Partial Least Square(PLS). Results indicate that the model is simple and stable, and can offer a reference for general black soil condition investigation in north of Jilin province.

Key words: soil nutrition; near infrared spectroscopy; precision agriculture; multiplicative scatter correction; correlation spectrum

收稿日期:2007-08-14; **修订日期:**2007-09-17.

基金项目:中国科学院知识创新工程领域前沿资助项目(No. O62Y32Q060);应用光学国家重点实验室主任基金资助项目(No. O6133FQ066);吉林省科技发展计划资助项目(No. 20040324-2)

1 引 言

现代农业技术领域的一个重要发展方向是“数字农业”或叫“精准农业”。精准农业是近年来国际上农业科学研究的热点领域,是现代信息技术和传统农业紧密结合的产物。近红外光谱(Near Infrared—NIR)技术作为一种无损、快速的分析手段,已经在国民经济发展的各个领域得到应用^[1-4],在土壤学领域的应用也逐渐得到人们的重视,是实现精准农业的有效手段之一,在预测土壤水分、有机质和总氮等方面取得了良好的效果^[5-10],显示出其在土壤肥力及植被营养方面的应用前景十分广阔。偏最小二乘算法(Partial Least Square,PLS)是目前比较常用的近红外定标分析方法之一,其优点是可以对光谱数据进行有效压缩,利用它对连续谱进行分析能够得到令人满意的结果。由于普查性质的大范围土壤定位管理所采集的数据量过于庞大,利用 PLS 算法对其连续谱进行定标分析无疑是一项“浩大的工程”。因此有必要采用适当的化学计量学算法对连续谱进行筛选,即根据待测组分选用尽可能少的波长组合来对该成分进行定量分析。本文通过实验,利用多元散射校正(Multiplicative Scatter Correction,MSC)法对土壤有机质成分全谱数据进行处理,有效地消除了由样品状态、测量条件引起的散射影响,并利用相关光谱法优选出若干可用于有机质成分分析的最优组合波长,得到了比较满意的结果。

2 理论方法分析

2.1 多元散射校正

由于样品颗粒不均匀常造成同一批样品不同次装样得到的近红外光谱有很大差异,在许多情况下,由散射引起的光谱变化不可忽略。当待测组分含量较低时,由其引起的光谱变化几乎完全被散射效应所覆盖,严重影响分析质量。所以,Martens^[11]等人首先提出了 MSC,用于消除由于散射引起的基线漂移和光谱的不重复。其具体算法如下:

(1)计算样品原始光谱矩阵的平均光谱,并近似地认为平均光谱即为“理想”光谱:

$$\bar{\mathbf{A}} = \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i, \quad (1)$$

(2)对每个样品光谱与“理想”光谱做一元线性回归:

$$\mathbf{A}_i = m_i \bar{\mathbf{A}} + \mathbf{b}_i, \quad (2)$$

(3)求出 MSC 校正后光谱矩阵:

$$\mathbf{A}_{i, \text{MSC}} = \frac{\mathbf{A}_i - \mathbf{b}_i}{m_i}. \quad (3)$$

其中, \mathbf{A} 为原始光谱矩阵, \mathbf{A}_i 为第 i 个样品光谱, m_i 和 \mathbf{b}_i 分别是第 i 个样品的相对偏移系数和平移量。

MSC 假设波长及样品浓度变化对散射无影响,因此对组分性质变化宽的样品处理效果较差^[12]。但本文中土壤有机质不属于此类情况,可以进行 MSC 校正。

2.2 相关光谱法

采用全波段光谱数据建立多变量定标模型可以获得较多的样品成分信息,但同时也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,影响模型的预测精度。进行近红外光谱分析应当遵循“少而精”的原则,即提高分析信号中的有效信息率,用尽可能少的波长信息来准确描述待测成分与其化学浓度之间的关系。

相关光谱法是 NIR 分析中常用的一种光谱解析手段。通过对定标样品集每一个波长点处的吸光度值与相应的化学成分值做一元线性回归,并求出近红外分析值与真实值之间的相关系数,得到相关系数与对应波长点的相关光谱曲线图。根据相关光谱图可以对定标谱区和波长点进行优选。

3 实验样品及仪器

收集吉林省北部地区具有代表性的黑质土壤样品 85 份,有机质成分含量在 0.8%~4.7%,平均值为 2.49%。风干后由旋风磨粉碎,并过 1 mm 孔筛。样品分成 2 份,1 份用于实验室化学分析,1 份用于近红外定标分析。

实验仪器采用丹麦 Foss 公司的 XDS Rapid Content™ 光栅型近红外光谱分析仪,光谱采集范围 400~2 500 nm,光谱宽度为 10 nm,探测器为 Si 和 PbS,环境温度为室温(26±1)℃。样品装入培养皿后由专用装置固定,测量时每隔 120°旋

转培养皿,每个位置扫描 32 次漫反射光谱并自动求其平均光谱,相当于每个样品测量 96 次后求平均,从而减少仪器本身和装样不均匀引起的误差。保存数据用于土壤有机质定量分析。化学计量分析软件分别采用 WinISI III v1.5e 和基于 Matlab 自行编写的定标分析程序。

4 结果与讨论

4.1 近红外光谱测量结果

85 份吉林省北部地区黑质土壤粉末样品近红外光谱如图 1 所示。

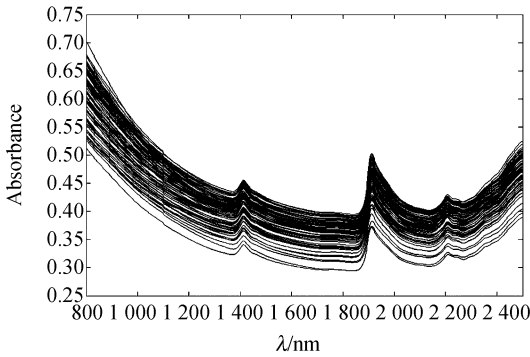
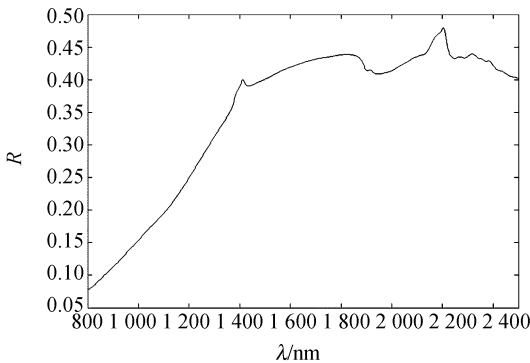


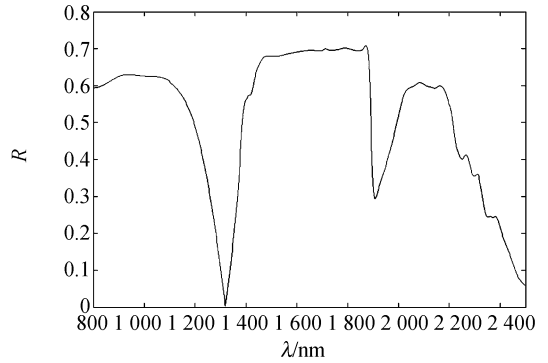
图 1 85 份土壤样品近红外吸收光谱
Fig. 1 Absorbance spectra of 85 soil samples

4.2 应用 MSC 校正和相关光谱法对土壤近红外光谱处理结果

土壤有机质含量化学值分别与原始吸光度和散射校正后吸光度的相关曲线如图 2(a、b)所示。校正前,散射效应的影响致使大量有用信息被掩盖,使得样品吸光度与化学值之间的相关性受到干扰,造成相关系数不高。散射校正后,部分减弱



(a)原始吸光度相关曲线
(a)Original correlation spectra



(b)散射校正后吸光度的相关曲线
(b)MSC correlation spectra

图 2 土壤有机质含量化学值与原始吸光度和散射校正后吸光度的相关曲线

Fig. 2 Correction spectra of soil organic mater

的信息被较好地解析出来,出现了 3 个连续的高相关区,从长波到短波分别对应近红外波段的合频、1 级倍频和 2 级倍频区。相关系数和分辨率与校正前相比都有了不同程度的提高,为谱区优选^[13-14]提供了良好的参考信息。

4.3 采用 PLS 算法对连续谱分析结果

对定标谱区的选取原则上应选用相关系数较高处的波长点,这些波长点处的光谱所包含的样品成分信息相对比较丰富,用于定标容易得到较好的模型。由于土壤的测量为漫反射方式,实验中只关心近红外长波波段的信息。因此根据图 2 (b)中土壤有机质相关光谱曲线所示,选取波长范围为 1 660~1 850 nm,2 000~2 320 nm 的光谱数据对土壤有机质成分进行定标分析,定标软件采用 Foss 公司提供的 WinISI III,光谱预处理方式为 MSC,定标方法采用 PLS 算法。选取 55 份样品进行定标,另外 30 份样品为预测集。当主因子数为 8 时得到了最好的回归结果,RMSEC 为 0.34%,相关系数 R 为 0.91, RMSEP 为 0.30%。

利用 PLS 算法对有机质成分进行定量分析虽然得到了比较好的效果,但当样品数量较多时,计算将变得十分复杂,耗时较长。因此,需要利用相关光谱法进一步对光谱数据进行压缩。

4.4 波长逐级优选及定标分析结果

如何选择尽可能少的波长组合准确反映待测组分信息是近红外光谱分析的难点。波长选用少了会出现欠拟合现象,所建模型与预测效果都不

好;波长选用过多容易产生过拟合,建立模型的定标结果可能比较好,但不利于进行稳定的预测分析。实验研究表明:参与建立定标模型的波长个数在 3~5 个之间为好。

采用相关光谱法对定标波长点进行选取不能像选取定标谱区那样简单的通过设定阈值来实现。这是因为相邻波长点处的吸光度值之间具有一定的相关性,在同一高相关区域选取的波长点不宜过多,应该尽量选择相关曲线峰值处所对应的波长点,并根据实际情况进一步筛选。

根据图 2(b)中曲线选出若干相对较高的相关峰,对应的波长分别为 1 710、1 790、1 870、2 085、2 165、2 270、2 310 nm,如图 3 所示。采用多元线性回归算法建立这 7 个波长点处吸光度矩阵与有机质化学值之间的定标模型,并利用交叉检验均方标准差 (Root Mean Square Error of

Calibration Value, RMSECV)来衡量模型的预测能力。结果如表 1、图 4。

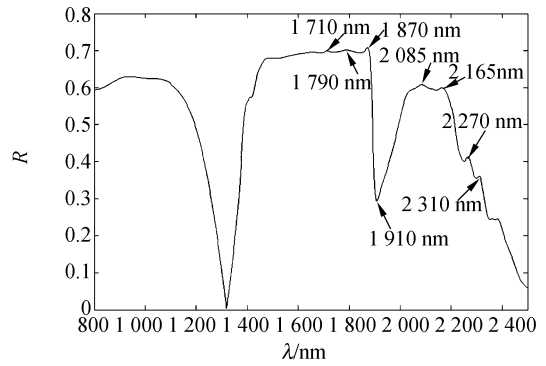


图 3 土壤有机质散射校正后的相关光谱及挑选出的定标波长点

Fig. 3 MSC correlation spectra of soil organic matter and chosen calibration wavelength points

表 1 选取 7 个波长点定标的模型参数

Tab. 1 Calibration model parameters of 7 wavelength points

| | | | | | | | |
|---------|--|------|--------|------|------|--------|-------|
| 波长点/nm: | 1710 | 1790 | 1870 | 2085 | 2165 | 2270 | 2310 |
| 回归系数: | 994 | 1276 | -282.1 | -0.6 | -414 | -226.3 | 493.4 |
| 回归方程: | $Y = 58.7 - 994 * X_{1710} + 1276 * X_{1790} - 282.1 * X_{1870} - 0.6 * X_{2085} - 414 * X_{2165} - 226.3 * X_{2270} + 493.4 * X_{2310}$ | | | | | | |

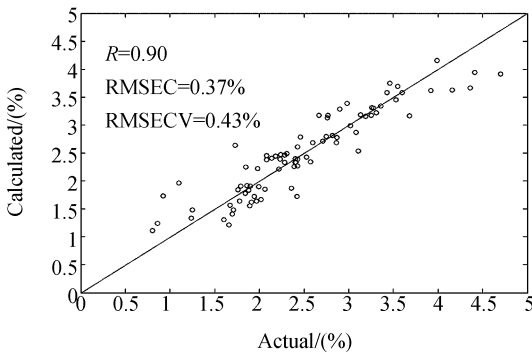


图 4 选取 7 个波长点定标的散点图

Fig. 4 Scatter plots of calibration with 7 wavelength points

从结果上看:相关系数比较令人满意,达到了 0.90,但 1 710 nm 和 1 790 nm 这两个波长点的回归系数明显偏高。一般认为某波长点处的回归

系数超过 500 时,该点处吸光度的变化会对方程的结果产生较大影响,这样的定标模型不够稳定,不是一个合理的模型,需要通过减少参与定标的波长个数来降低回归系数。因此依次去除 2 085、2 270、1 870 nm 这 3 个对方程影响相对较小的波长点,只采用 1 710、1 790、2 165、2 310 nm 建立定标模型。结果如表 2、图 5。

表 2 选取 4 个波长点定标的模型参数

Tab. 2 Calibration model parameters of 4 wavelength points

| | | | | |
|---------|--|-------|--------|-------|
| 波长点/nm: | 1 710 | 1 790 | 2 165 | 2 310 |
| 回归系数: | 238.3 | 340.1 | -392.3 | 254.5 |
| 回归方程: | $Y = 16.1 - 238.3 * X_{1710} + 340.1 * X_{1790} - 392.3 * X_{2165} + 254.5 * X_{2310}$ | | | |

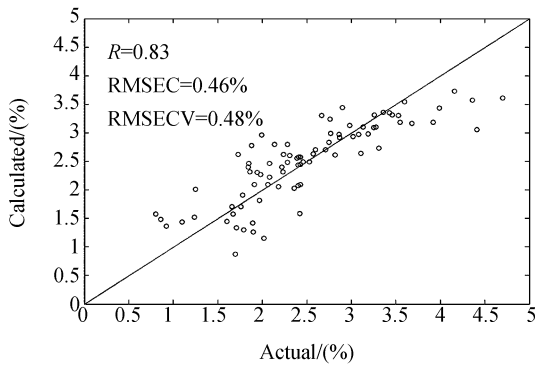


图 5 选取 4 个波长点定标的散点图

Fig. 5 Scatter plots of calibration with 4 wavelength points

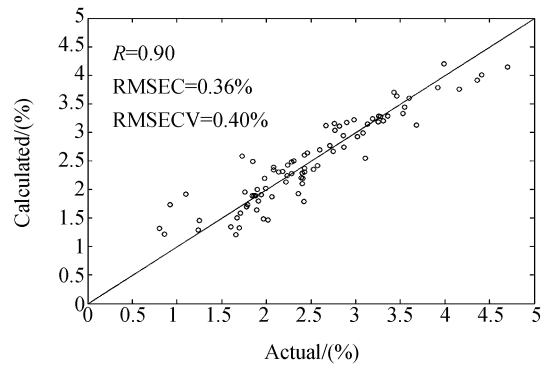


图 6 选取 5 个波长点定标的散点图

Fig. 6 Scatter plots of calibration with 5 wavelength points

调整参与定标的波长点后,回归系数大幅度下降,提高了模型的稳定性,然而相关系数却降到了 0.83,模型的预测能力随之降低,显然这是违反初衷的。为此,本文引入弱相关波长为参考波长,参与回归运算。弱相关波长是指相关系数曲线上较低的点(如图 3 中的 1 910 nm),它对待测组分化学含量的变化不敏感,但由于与高相关点受到的散射及其它物理状态因素影响基本相同,这些波长点在参与回归运算中可以起到参考波长的作用,消除其他波长点所受散射等物理因素的影响,从而提高模型的分析精度。引入参考波长后的回归结果如表 3、图 6:

表 3 选取 5 个波长点定标的模型参数

Tab. 3 Calibration model parameters of 5 wavelength points

| | | | | | |
|---------|--|-------|-------|--------|-------|
| 波长点/nm: | 1 710 | 1 790 | 1 910 | 2 165 | 2 310 |
| 回归系数: | -275.8 | 316.4 | -54.0 | -351.9 | 295.3 |
| 回归方程: | $Y = 31.5 - 275.8 * X_{1710} + 316.4 * X_{1790} - 54.0 * X_{1910} - 351.9 * X_{2165} + 295.3 * X_{2310}$ | | | | |

采用 MSC 校正和相关光谱法首先确定出若干待选波长点进行多元线性回归,通过减少参与回归的波长点个数成功地降低了模型的回归系数,并引入了参考波长点。参考波长较好地抑制了化学成分之外其它因素对各波长点的影响,使

模型相关系数上升到 0.90,回归结果明显改善。所选择的波长中,1 710 nm 和 2 310 nm 与有机质在合频和一倍频区的特征吸收峰接近,从而验证了该方法的合理性。最终得到的模型的预测能力和稳定性与之前相比均有大幅提升,接近本文采用 PLS 算法对连续谱进行分析的结果。

5 结 论

由于土壤的类型(化学形态、物理形态如颗粒度)及所处环境(温度、湿度、pH 值等)比较复杂,未经预处理的原始光谱容易受到散射影响,导致待测成分与其化学值之间的部分相关信息被掩盖。而利用多元散射校正可以很好地剔除散射引起的光谱基线漂移现象,提高光谱分辨率和信噪比。采用相关光谱法选出测量土壤有机质含量的波长组合,并根据经验和参考波长理论对所建立的模型进行逐步优化,最终获得了比较满意的结果,模型相关系数达到 0.90, RMSEC 为 0.36%, RMSECV 为 0.40%,接近 PLS 算法的分析结果。采用该方法需要的数据量较少,建立的模型简单、稳定,在对土壤定位管理所采集的海量光谱数据进行定量分析中具有一定的优势,该方法可以为吉林省北部地区黑土质土壤养分状况普查和便携仪器的研制提供参考。

致谢:感谢暨南大学近红外光谱组的有益探讨及大力支持。

参考文献:

- [1] 陈洁梅,潘涛,陈星旦. 二阶导数光谱预处理在用 FTIR/ATR 方法定量测定葡萄糖-6-磷酸和果糖-6-磷酸中的应用[J]. 光学精密工程, 2006, 14(1): 1-7.

- CHEN J M, PAN T, CHEN X D. Application of second derivative spectrum prepares in quantification measuring glucose-6-phosphate and fructose-6-phosphate using a FTIR/ATR method[J]. *Opt. Precision Eng.*, 2006, 14(1): 1-7. (in Chinese)
- [2] 陈华才, 吕进, 陈星旦, 等. 基于径向基函数网络的茶多酚总儿茶素近红外光谱检测模型的研究[J]. *光学精密工程*, 2006, 14(1): 58-62.
- CHEN H C, LU J, CHEN X D, *et al.*. Near infrared spectroscopic model for determinating total catechins in tea polyphenol powder based on radical basis function network[J]. *Opt. Precision Eng.*, 2006, 14(1): 58-62. (in Chinese)
- [3] 黄英华, 乐焕龙, 傅宜磊, 等. 注射用奥美拉唑钠水分含量的近红外漫反射光谱法测定[J]. *中国医药工业杂志*, 2005, 36(6): 354-356.
- HUANG Y H, LE H L, FU Y L, *et al.*. Determination of water content in omeprazole sodium for injection by near infrared diffuse reflectance spectroscopy[J]. *Chinese Journal of Pharmaceuticals*, 2005, 36(6): 354-356. (in Chinese)
- [4] 郑咏梅, 张军, 陈星旦, 等. 短波近红外光谱的整粒小麦蛋白质 PLS 方法的定量分析[J]. *光谱学与光谱分析*, 2004, 24(9): 1047-1049.
- ZHENG Y M, ZHANG J, CHEN X D, *et al.*. The PLS analysis with short wave near-infrared of whole wheat for protein[J]. *Spectroscopy and Spectral Analysis*, 2004, 24(9): 1047-1049. (in Chinese)
- [5] KRISHNAN P, ALEXANDER J D, BUTLER B J, *et al.*. Reflectance technique for predicting soil organic matter [J]. *Soil Sci. Soc. Am. J.*, 1980, 44: 1282-1285.
- [6] DALAI R C, HENRY R J. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry[J]. *Soil Sci. Soc. Am. J.*, 1986, 50: 120-123.
- [7] MALLEY D F, YESMIN L, RAY D W, *et al.*. Application of near infrared spectroscopy in analysis of soil mineral nutrients[J]. *Communications in Soil Science and Plant Analysis*, 1999, 30: 999-1012.
- [8] CHENG W C, LAIRD D A, MAURICE J. Mausbach, and Charles R. Hurburgh, Jr. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties[J]. *Soil Sci. Soc. Am. J.*, 2001, 65: 480-490.
- [9] MORON A, COZZOLINO D. Application of near infrared reflectance spectroscopy for the analysis of organic C, total N and pH in soils of Uruguay[J]. *Journal of Near Infrared Spectroscopy*, 2002, 10(3): 215-221.
- [10] BEN-DOR E, GOLDSHLEGER N, BENYAMINI Y, *et al.*. The spectral reflectance properties of soil structural crusts in the 1.2- to 2.5-um spectral region[J]. *Soil Sci. Soc. Am. J.*, 2003, 67: 289-299.
- [11] GEI ADI P, MACDOUGALL D, MARTENS H. Linearization and scatter-correction for near infrared reflectance spectra of meat[J]. *Appl. Spectrosc.*, 1985, 39(3): 491-500.
- [12] 陆婉珍. 现代近红外光谱分析技术(第二版)[M]. 北京: 中国石化出版社, 2007: 36.
- LU W ZH. *Modern Near Infrared Spectroscopy Analytical Technology (Second Edition)* [M]. Beijing: China Petrochemical Press, 2007: 36. (in Chinese)
- [13] 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 Q, ZHANG G L, CHEN X D. Effects of multiplicative scatter correction on a calibration model of near infrared spectral analysis[J]. *Opt. Precision Eng.*, 2005, 13(1): 53-58. (in Chinese)
- [14] 芦永军, 张军, 朴仁官, 等. 组合生成算法与多元线性回归相结合用于近红外光谱波长的优选[J]. *分析化学*, 2004, 32(8): 1116-1119.
- LU Y J, ZHANG J, PIAO R G, *et al.*. Choose optimal wavelengths for calibration by combining the combination-making algorithm with multivariate linear regression[J]. *Chinese Journal of Analytical Chemistry*, 2004, 32(8): 1116-1119. (in Chinese)

作者简介: 丁海泉(1982—), 男, 黑龙江人, 中国科学院长春光学精密机械与物理研究所博士研究生, 主要从事土壤养分的快速近红外光谱分析方法技术研究。E-mail: haiquan_ding@163.com

通讯作者: 卢启鹏(1964—), 男, 中国科学院长春光学精密机械与物理研究所研究员, 博士生导师, 长期从事光谱分析及光谱仪器等相关技术方面的研究。E-mail: luqp@ciomp.ac.cn