

基于 LSSVM 的土壤重金属定量分析

林晓梅¹, 黄玉涛¹, 林京君^{2*}, 陶思宇¹, 车长金¹

1. 长春工业大学电气与电子工程学院, 吉林 长春 130012

2. 长春工业大学机电工程学院, 吉林 长春 130012

摘要 为了提高土壤定量分析的精度, 分别把偏最小二乘法(PLS)和最小二乘支持向量机(LSSVM)与激光诱导等离子体技术相结合对土壤中的 Cu 元素进行分析。对比分析了 Cu I 324.75 nm 和 Cu I 327.40 nm 两条特征谱线, 最终选择 Cu I 324.75 nm 作为分析谱线。首先对实验参数进行优化。通过对比激光能量、采集延时与信噪比之间的关系, 确定最佳能量为 90 mJ, 最佳采集延时为 1 000 ns。然后在最佳实验条件下采集五个不同浓度样品的特征光谱, 并用内标法、PLS 和 LSSVM 建立定标模型。对比三种模型的拟合系数、均方根误差和平均相对误差, 发现由于土壤基体效应和自吸收效应的影响, 内标法的定标模型性能较差, 拟合程度未达到实验要求, 而均方根误差和平均相对误差的数值过大, 无法满足实验对于精确度和稳定性的要求。用 PLS 对定标模型进行校准, 相对于内标法而言, 定标模型的精确度和稳定性均有明显的提高, R^2 由 0.870 1 提高到 0.985 1, 训练集和预测集的均方根误差均下降到了 0.1 Wt%量级, 平均相对误差虽有所下降, 但仍然无法达到实验要求, 说明 PLS 虽然可以在一定程度上提高定标模型的精确度, 但在提高稳定性方面仍有欠缺, 并不能很好的降低土壤的基体效应与自吸收效应。与内标法和 PLS 的定标模型相比, LSSVM 定标模型的精确度和稳定性最好, R^2 提高到了 0.997 6, 模型中的数据点基本分布在拟合曲线上, 具有良好的线性相关性。相比于内标法, LSSVM 定标模型训练集的均方根误差由 3.448 8 Wt% 下降到 0.018 7 Wt%, 预测集的均方根误差由 1.280 7 Wt% 下降到 0.149 1 Wt%, 体现稳定性的平均相对误差降低了 6.24 倍。与 PLS 定标模型相比, LSSVM 定标模型的各个参数均有大幅降低, 特别是平均相对误差由 7.455 6% 下降到 2.137 0%, 可以满足稳定性要求。说明在提高定标模型精确度与稳定性方面, LSSVM 算法更具有优势, 能够更好地降低土壤基体效应和自吸收效应带来的影响。

关键词 激光诱导等离子体技术; 内标法; 偏最小二乘法; 最小二乘支持向量机; 土壤

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)05-1523-05

引言

土壤是维持人类生存和生活的重要组成部分。随着工业化程度的不断提高, 工业生产中产生的重金属废弃物也越来越多, 并在土壤中不断累积, 对土壤造成很严重的污染。同时这些重金属还会通过动植物进入食物链, 最终进入人体, 对身体健康造成很大的影响^[1]。因此, 提高土壤重金属定量分析的精确度和稳定性意义重大。

在检测土壤成分含量的过程中, 常用的方法有原子吸收分光光度法、X 射线荧光光谱法、电感耦合等离子体质谱法和中子活化法等。激光诱导等离子体技术(LIBS)是一种成分

含量分析光谱技术^[2]。相对于以上几种检测方法, LIBS 技术具有无需制样、检测时间短、可多元素同时检测等优势^[3], 被逐渐应用于土壤定性、定量分析领域。Ramli Muliadi 等^[4]用 LIBS 技术检测土壤中的 Cs 元素, 检测限为 0.3 ppm。Edilene C Ferreira 等^[5]用人工神经网络对 Cu 元素的定标曲线进行校准, 使 Cu 的检测限降到 2.3 ppm。Li 等^[6]通过把电火花放电和 LIBS 技术相结合, 进行土壤中 Cr 等元素的检测, 把光谱的信噪比提高了 2~3 倍。谷艳红等^[7]把 CF-LIBS 与 LIBS 技术相结合, 提高了土壤中 Cr 元素含量分布的准确度。这些研究侧重于检测技术的结合和样品检出限的提高, 但对样品精确度和稳定性的研究较少。由于土壤的基体效应较大^[8], 会对定量分析产生很大影响, 降低 LIBS 的分析精

收稿日期: 2019-04-17, 修订日期: 2019-08-20

基金项目: 国家重大科学仪器开发专项(2014YQ120351), 吉林省科技厅(20180414017GH, 20180101283JC)资助

作者简介: 林晓梅, 女, 1965 年生, 长春工业大学电气与电子工程学院教授 e-mail: linxiaomei@ccut.edu.cn

* 通讯联系人 e-mail: 1124270941@qq.com

度。本工作主要研究了偏最小二乘法(PLS)和最小二乘支持向量机(LSSVM)算法在提高土壤精确度和稳定性检测方面的效果。

首先研究了能量和延时对谱线强度和信噪比的影响,同时为了降低土壤基体效应对谱线的影响,提高 LIBS 的重复性和检测精度,在最佳实验条件下,分析对比了 PLS 和 LSSVM 算法对定标模型的影响。通过拟合系数,均方根误差和平均相对误差等参数对结果进行评价,得到最佳的定标模型。

1 实验部分

1.1 装置

实验装置如图 1 所示。用输出波长为 1 064 nm, 脉冲宽度为 8 ns, 脉冲频率为 1~10 Hz 的 Nd:YAG 激光器(Nimma-400 型, 中国镭宝光电)作为光源。激光经过焦距为 150 mm 的聚焦透镜, 将能量聚焦在样品表面, 使样品激发产生等离子体, 等离子体的光谱信号经光纤探头耦合至具有 CCD 的四通道光谱仪(Avaspec-ULS2048CL-4-EVO 型, 荷兰 Avantes)进行分光等处理。实验中用三维移动平台(YMC400, 中国江云光电)来控制样品的移动, 激光器与光谱仪由数字延时发生器(BNC575, 美国 Berkeley Nucleonics Corp)同步控制, 触发信号为激光器的调 Q 信号。光谱仪的探测延时为 0.9 μ s。为了降低因能量波动对实验带来的影响, 取不同位置光谱的平均值作为实验最终谱线。

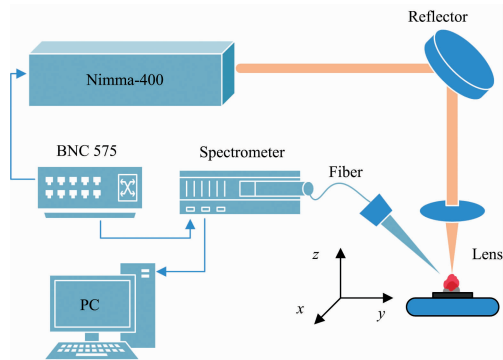


图 1 实验系统原理图

Fig. 1 Schematic diagram of the experimental system

1.2 样品制备

实验中所用的样品由标准土壤样品 GBW07403 和结晶状的 $\text{Cu}(\text{NO}_3)_2$ 混合而成。先将混合好的样品反复研磨, 再对样品进行烘干处理。取 5 g 样品放入模具中, 用压片机施加 30 MPa 的压强, 持续 30 min, 压制成直径 30 mm, 厚度 3 mm 的圆饼形薄片。压制好的样品如图 2 所示, 样品浓度如表 1 所示。

1.3 算法

偏最小二乘法(PLS)是一种多元线性回归方法, 通过提取元素浓度和谱线强度的主成分, 实现多因变量对多自变量的回归建模。最小二乘支持向量机(LSSVM)是基于支持向量机(SVM)发展而来的机器学习方法, 把 SVM 中的不等式

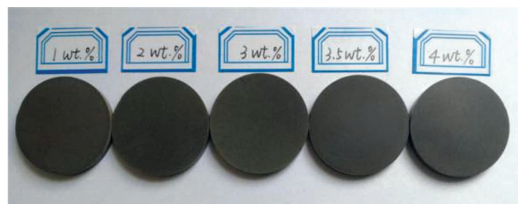


图 2 实验样品图

Fig. 2 Pictures of experimental samples

表 1 样品中 Cu 元素的浓度

Table 1 Certified concentration (Wt%) of Cu in the samples

Sample number	Concentration of Cu/(mg · g ⁻¹)
1 #	10
2 #	20
3 #	30
4 #	35
5 #	40

约束替换为等式约束, 训练过程的复杂度大大降低, 并且能够较好地解决高维度和非线性的问题^[9]。在 LSSVM 算法中, 把径向基函数作为核函数, 用留一验证法进行参数寻优。该模型的优化问题可以表示为

$$\min J(\omega, e) = \frac{1}{2} \omega^T \omega + \frac{\lambda}{2} \sum_{i=1}^N e_i^2 \quad (1)$$

式(1)中, ω 为权向量, e_i 为误差向量, λ 为正正则化参数, 用来平衡最优超平面与最小偏差量, 该式的约束条件为

$$y_i [\omega^T \varphi(x_i) + b] = 1 - e_i, \quad i = 1, \dots, N \quad (2)$$

式(2)中, b 为偏差向量, $\varphi(x_i)$ 为核空间映射函数, 提取原始空间中的数据特征, 把原始样本数据映射到高维空间中进行处理。

最终可得 LSSVM 回归模型为

$$y(x) = \sum_{i=1}^N a_i k(x, x_i) + b \quad (3)$$

式(3)中, a_i 为 Lagrange 乘子, $k(x, x_i)$ 为满足 mercer 条件下的核函数。

本实验对每种浓度样品取 9 组数据, 其中 6 组作为训练集, 3 组作为预测集, 建立定标模型。并对数据进行 Lorenz 拟合, 用预测集的偏离程度验证两种算法在降低基体效应和自吸收效应方面的效果。

2 结果与讨论

2.1 特征谱线选取

实验检测了样品在 190~557 nm 范围内的光谱数据^[10]。由于土壤中 Cu 元素的特征谱线受附近谱线的影响较大, 因此对比各特征谱线受影响程度, 选择 Cu I 324.75 nm 和 Cu I 327.40 nm 两条较好的特征谱线进行分析。对 Cu I 324.75 nm 谱线产生影响的是 Ti II 324.86 nm 谱线, 由于选取的 GBW07403 标准土样含 Ti 量很低, 且衰减很快, 因此把 Ti II 324.86 nm 对 Cu I 324.75 nm 的影响忽略不计。综合光谱强度和美国原子光谱数据库(NIST)的部分参数,

选择 Cu I 324.75 nm 谱线作为 Cu 的分析线。两条 Cu 线的参数对比如表 2 所示。由于 Cu 在土壤中属于微量元素，而 Fe 是土壤中的基体元素，含量相对稳定，因此选择 Fe I 393.26 nm 作为内标元素。

表 2 Cu I 324.75 nm 与 Cu I 327.40 nm 的参数对比
Table 2 Comparison of the parameters of Cu I 324.75 nm and Cu I 327.40 nm

Spectral line/nm	Intensity /counts	Transition probability/s ⁻¹	Energylevels /eV
Cu I 324.75	6 602	1.395×10 ⁸	3.817
Cu I 327.40	4 491	1.376×10 ⁸	3.786

2.2 参数优化

2.2.1 激光能量对特征谱线的影响

激光诱导产生等离子体的过程中，激光能量的大小会对特征谱线和信噪比(SNR)产生影响^[11]。Cu I 324.75 nm 的谱线强度和信噪比随激光能量的变化如图 3 所示。由图可以看出，能量在 30~90 mJ 之间，谱线强度和信噪比随能量的增加而增加。分析认为在此过程中，等离子体密度不断增加，并逐渐达到临界值。激光能量在到达 90 mJ 后，谱线强度和信噪比开始趋于平缓。这是由于过高的能量使等离子体产生屏蔽效应，对后沿激光产生一定程度的散射，激光能量不能完全到达样品表面。因此，选择 90 mJ 作为本实验使用的最佳能量。

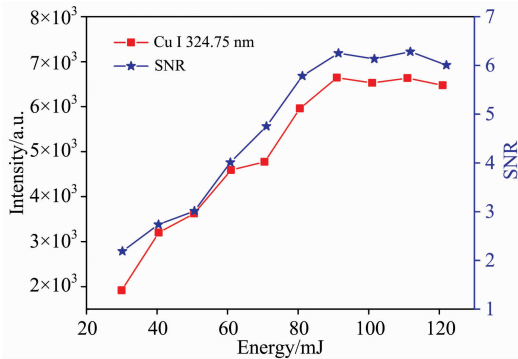


图 3 激光能量对谱线强度和 SNR 的影响
Fig. 3 The influence of laser energy on the spectral intensity and SNR

2.2.2 采集延时对特征谱线的影响

在等离子体光谱信号的采集过程中，采集延时对光谱强度和信噪比会产生很大的影响^[12]。具体变化如图 4 所示，采集延时在 0 ns 时，光谱的信噪比较低。推测等离子体产生初期，由于韧致辐射和复合辐射的作用，信号光谱会有很高的背景噪声。随着采集延时的不断增加，韧致辐射和复合辐射的逐渐衰减，并且衰减速度比原子发射光谱的衰减速度快，因此，采集延时在 0~1 000 ns 时，信噪比逐渐增加并达到最高值。采集延时在 1 000 ns 之后，由于等离子体不断冷却扩散，使得谱线强度和信噪比不断降低。因此，选择 1 000 ns 作为本次使用的最佳采集延时。后续实验均在最佳能量和

最佳延时条件下进行。

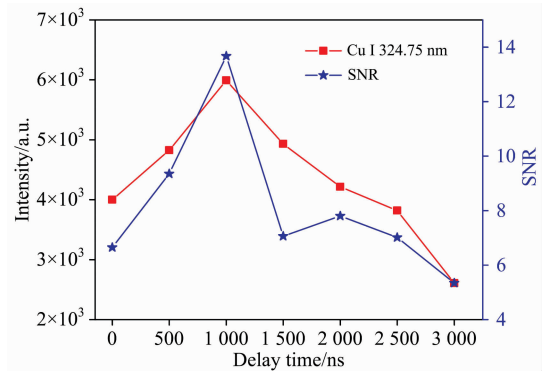


图 4 采集延时对谱线强度和 SNR 的影响
Fig. 4 The influence of delay time on the spectral intensity and SNR

2.3 定量分析

2.3.1 基于内标法的定标模型

内标法是最常用的定量分析方法之一，尤其是土壤的基体效应比较明显，内标法可以在一定程度上抑制基体效应，降低实验环境变动对谱线的影响^[13]。在用内标法做定标模型时，把每个浓度的样品各采集 9 组数据，以其中 6 组数据做训练集，3 组数据做预测集。把训练集和预测集的全部数据拟合在一起，通过拟合系数 R^2 来观察模型的拟合效果。拟合结果如图 5 所示。由图可知，训练集的均方根误差(RMSEC)为 3.448 8 Wt%，预测集的均方根误差(RMSEP)为 1.280 7 Wt%，表明该模型的精确度较低。平均相对误差(ARE)为 13.340 0%，说明该模型的稳定性不高，而且从拟合系数可以看出，拟合效果还有待提高，特别是浓度在 2 Wt% 和 4 Wt% 时，预测集浓度出现大幅度的偏离。

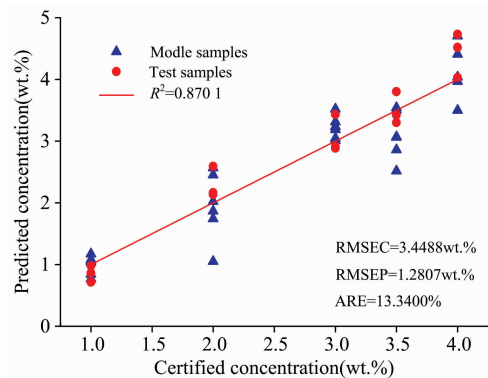


图 5 基于内标法的定标模型

Fig. 5 Calibration model based on internal standard method

2.3.2 基于 PLS 的定标模型

偏最小二乘法是一种常用的数学优化方法，可以在一定程度上降低土壤基体效应带来的误差，对定标模型进行校准。用实验采集的所有数据建立实际浓度与预测浓度之间的关系曲线。拟合结果如图 6 所示。由图可以看出，定标曲线的 R^2 有了很大幅度的提高，提高到了 0.985 1，RMSEC 和 RMSEP 均有了很大程度的降低，降低到了 0.1 Wt% 量级，

大大提高了模型的精确度。同时该模型的 ARE 降低到了 7.455 6%，说明该模型的稳定性有了一定的提升，但仍然无法满足实验要求，需要进一步优化。

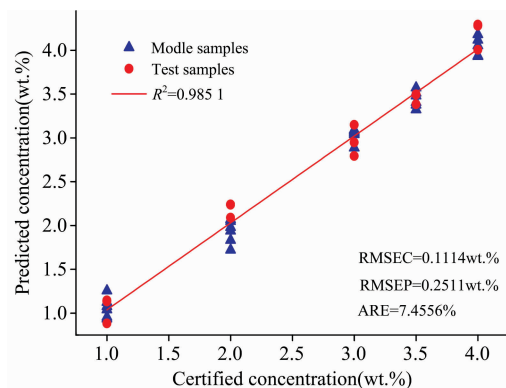


图 6 基于 PLS 的定标模型

Fig. 6 Calibration model based on PLS

2.3.3 基于 LSSVM 的定标模型

PLS 算法的引入虽然在一定程度上提高了模型的精确度和稳定性，但提高程度仍然不够，因此引入了 LSSVM 算法来对定标模型进行校准。校准结果如图 7 所示。从图中可以看出，定标曲线的 R^2 有了明显的提高，提高到了 0.997 6，

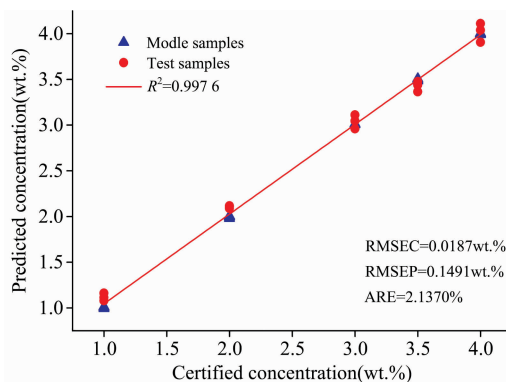


图 7 基于 LSSVM 的定标模型

Fig. 7 Calibration model based on LSSVM

拟合效果较好。同时，RMSEC 和 RMSEP 分别下降到了 0.018 7 Wt% 和 0.149 1 Wt%，ARE 下降到了 2.137 0%。相比于内标法和 PLS，LSSVM 的 RMSE 和 ASE 均得到了很大程度的降低，模型的稳定性和精确度都得到大幅度的提高，基本消除了因基体效应和自吸收效应给实验结果带来的影响。具体对比结果如表 3 所示。

表 3 内标法、PLS 和 LSSVM 数据对比

Table 3 Internal standard method, PLS and LSSVM data comparison

方法	R^2	RMSEC /Wt%	RMSEP /Wt%	ARE /%
内标法	0.870 1	3.448 8	1.280 7	13.340 0
偏最小二乘法	0.985 1	0.111 4	0.251 1	7.455 6
最小二乘支持向量机	0.997 6	0.018 7	0.149 1	2.137 0

3 结 论

实验首先对激光能量和采集延时进行优化，选取最佳能量为 90 mJ，最佳延时为 1 000 ns，在此最佳实验条件下分别利用内标法、PLS 和 LSSVM 对土壤中的 Cu 元素建立定标模型。对比可以看出，内标法定标模型的稳定性和精确度都很不理想，拟合系数 R^2 只有 0.870 1，RMSEC 和 RMSEP 分别为 3.448 8 Wt% 和 1.280 7 Wt%。说明实验结果波动性较大，而 ARE 高达 13.340 0%，说明模型稳定性较差，很难实现定标曲线的精确校准。用 PLS 建立的定标模型， R^2 提高到了 0.985 1，RMSEC 和 RMSEP 分别降低为 0.111 4 Wt% 和 0.251 1 Wt%，和内标法相比精确度有了很大的提高。ARE 相比于内标法虽然也有了很大降低，但是模型的稳定性依然较差，无法满足实验要求。而用 LSSVM 建立的定标模型，与内标法和 PLS 相比，模型的精确度和稳定性都有很好的提高。 R^2 提高到 0.997 6，RMSEC 和 RMSEP 分别下降到 0.018 7 Wt% 和 0.149 1 Wt%。同时，ARE 与内标法的相比提高了 6.24 倍，说明 LSSVM 的稳定性较好，能够满足实验要求，适合用以提高定标模型的精确度和稳定性。

References

- [1] DU Chuang, GAO Xun, SHAO Yan, et al(杜 闯, 高 勋, 邵 妍, 等). Acta Physica Sinica(物理学报), 2013, 62(4): 221.
- [2] Guo L B, Zhu Z H, Li J M, et al. Optics Express, 2018, 26(3): 2634.
- [3] Zhang G Y, Song H, Liu Y, et al. Optik, 2018, 165(5): 87.
- [4] Ramli M, Khumaeni A, Kurniawan K H, et al. Spectrochimica Acta Part B: Atomic Spectroscopy, 2017, 132(6): 8.
- [5] Ferreira E C, Milori D, Ferreira E J, et al. Spectrochimica Acta Part B: Atomic Spectroscopy, 2008, 63(10): 1216.
- [6] Li K X, Zhou W D, Shen Q M, et al. Journal of Analytical Atomic Spectrometry, 2010, 25(9): 1475.
- [7] GU Yan-hong, ZHAO Nan-jing, MA Ming-jun, et al(谷艳红, 赵南京, 马明俊, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(3): 982.
- [8] Shakeel H, Haq S U, Abbas Q, et al. Spectrochimica Acta Part B: Atomic Spectroscopy, 2018, 146(8): 101.
- [9] Lin J J, Lin X M, Guo L B, et al. Journal of Analytical Atomic Spectrometry, 2018, 33(7): 1545.
- [10] Jiang H, Dong L, J Benchi, et al. Journal of Applied Spectroscopy, 2017, 84(1): 103.
- [11] Guo Y M, Deng L M, Yang X Y, et al. Journal of Analytical Atomic Spectrometry, 2017, 32(12): 2401.

[12] Robledo M A, Sobral H, Garcia V A. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2018, 144(6): 7.

[13] Lepore K H, Fassett C I, Breves E A, et al. *Applied Spectroscopy*, 2017, 71(4): 1.

Quantitative Analysis of Soil Heavy Metals Based on LSSVM

LIN Xiao-mei¹, HUANG Yu-tao¹, LIN Jing-jun^{2*}, TAO Si-yu¹, CHE Chang-jin¹

1. College of Electronics and Electrical Engineering, Changchun University of Technology, Changchun 130012, China

2. College of Mechanical and Electrical Engineering, Changchun University of Technology, Changchun 130012, China

Abstract In order to improve the accuracy of soil quantitative analysis, the partial least squares (PLS) and least squares support vector machine (LSSVM) were combined with laser-induced plasma technology to analyze the Cu elements in the soil. Two characteristic lines of Cu I 324.75 nm and Cu I 327.40 nm were compared and analyzed, and Cu I 324.75 nm was selected as the analytical line. Firstly, the experimental parameters were optimized. The relationship between laser energy, acquisition delay time and signal-to-noise ratio were compared. The optimal energy and optimal acquisition delay time were 90 mJ and 1 000 ns, respectively. Then, the characteristic spectra of five different concentrations of samples were collected under the optimal experimental conditions. The calibration model was established by standard internal method, PLS and LSSVM. By comparing the fitting coefficient, root mean square error and average relative error of the three models, the calibration model of the standard internal method had poor performance due to the influence of soil matrix effect and self-absorption effect. And the fitting degree did not meet the experimental requirements. The values of the root mean square error and the average relative error were too large to meet the accuracy and stability requirements of the experiment. The calibration model was calibrated with PLS. Compared with the standard internal method, the accuracy and stability of the calibration model were significantly improved. R^2 was increased from 0.870 1 to 0.985 1. The root mean square error of the training set and the root mean square error of the prediction set were reduced to the order of 0.1 Wt%. But the decrease of the average relative error with PLS model can't meet the experimental requirement. It indicated that PLS could improve the accuracy of the calibration model rather than the stability of the calibration model. The matrix effect and self-absorption effect of soil cannot be reduced. Compared with the former calibration models, the LSSVM calibration model has better accuracy and stability. The R^2 increased to 0.997 6. The data points in the model were basically distributed on the fitted curve with good linear correlation. Compared with the standard internal method, the root mean square error of the LSSVM training set decreased from 3.448 8 Wt% to 0.018 7 Wt%. The root mean square error of the prediction set decreased from 1.280 7 Wt% to 0.149 1 Wt%. The average relative error reduced by 6.24 times. Compared with the PLS calibration model, the parameters of the LSSVM calibration model were greatly reduced. The average relative error reduced from 7.455 6% to 2.137%, which can meet the stability requirements. It shows that the LSSVM algorithm has advantages for improving the accuracy and stability of the calibration model. It can reduce the matrix effects and the self-absorption effects of soil.

Keywords Laser-induced plasma technology; Internal standard method; Partial least squares; Least squares support vector machine; Soil

(Received Apr. 17, 2019; accepted Aug. 20, 2019)

* Corresponding author