

FPXRF——偏最小二乘法定量分析土壤中的铅含量

黄启厅¹, 周炼清^{1*}, 史舟¹, 李震宇², 顾群³

1. 浙江大学环境与资源学院, 浙江 杭州 310029
2. 浙江省环境监测中心站, 浙江 杭州 310012
3. 上海泽泉科技有限公司, 上海 200052

摘要 在实验室条件下, 利用 NITON XLt920 型便携式 X 射线荧光光谱 (field portable X-ray fluorescence, FPXRF) 仪获取土壤样品的 X 射线荧光光谱数据, 并采用偏最小二乘法 (PLS) 建立土壤 Pb 含量的预测模型。模型所用的光谱范围为与土壤中 Pb 元素密切相关的两个波段: 10.40~10.70 keV 和 12.41~12.80 keV; 最佳主成分数为 6。模型经交互验证, 其预测结果与实测值之间的相关系数为 0.966 6, 预测均方根误差 (RMSEP) 为 0.873 2。另外为了与偏最小二乘法做比较, 还分别利用仪器直接获取的 Pb 含量读数以及 X 射线荧光光谱数据中 Pb 的 $L\alpha$ 和 $L\beta$ 线的强度与 ICP 测定值进行一元线性和多元线性回归, 相关系数分别为 0.680 5 和 0.730 2, 均低于 PLS 模型的预测结果。研究表明, 相比较传统的原子吸收等测试方法, 便携式 XRF 仪在保证一定测试精度基础上, 具有方便、快速、无损和耗费少等优势, 可作为进一步分析前有力的筛选手段。

关键词 X 射线荧光光谱; 偏最小二乘法; 土壤; 重金属污染

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2009)05-1434-05

引言

我国受 Cd, As, Cr, Pb 等重金属污染的耕地面积近 2 000 万 hm^2 , 约占总耕地面积的五分之一; 其中工业“三废”污染耕地 1 000 万 hm^2 , 污水灌溉的农田面积已达 330 万 hm^2 。除了耕地污染之外, 我国的工矿区、城市还存在土壤(或土地)污染问题^[1]。土壤重金属污染已经引起了人们的重视, 并对污染土壤的治理和修复进行了大量的研究。

对污染土壤中各种重金属含量的测定是进行治理和修复的首要环节, 测定土壤重金属元素含量的传统方法是原子吸收分光光度法 (AAS) 和等离子体发射光谱仪 (ICP) 法。这两种方法都需要事先进行大量的样品预处理工作, 如需要加强酸消煮、准确配制标准溶液、准确配制样品溶液等。而且一般的土壤重金属调查采样量比较大, 需要大量的人力、物力和资金投入。目前 X 射线荧光光谱技术 (XRF) 已经相当成熟, 室内 XRF 分析已经进入实用阶段, 在地质、生物、食品、合金和表面薄膜等领域有着广泛的应用^[2-4]。滕

彦国等^[5]用 XRF 方法成功地对攀枝花地区土壤中的重金属含量进行了测定并评价了该区土壤重金属污染的特征。随着 XRF 激发源和探测器等相关技术的发展, 便携式 XRF 仪已经开始投入使用。Clark 等^[6]应用 FPXRF 考察了印度 Karnataka 和 Gujarat 两个城市的土壤和涂料中的 Pb 含量, 该方法可用于类似的环境中 Pb 含量的调查。Sterling 等^[7]利用 FPXRF 分析了粉尘样品中的 Pb 含量, 结果表明 FPXRF 法测定的结果与原子吸收分光光度法的测定结果没有显著差异 ($P=0.272$), 并且其准确度和精确度均随着样品中 Pb 含量的升高而提高。Argyaki^[8]在用 FPXRF 测定污染土壤中 Pb 的研究中发现土样的湿度和表面粗糙度会影响测定结果的准确度, 但经过校正后 FPXRF 法与 ICP-AES 法的测定结果能很好地吻合。前人的很多研究都是围绕着如何减轻或者消除基质效应等展开, 本实验利用便携式 XRF 仪对野外采集的 47 个土壤样品进行 Pb 元素的测定, 并引入偏最小二乘法对 X 射线荧光光谱数据进行分析研究, 期望为进一步的野外现场测定提供有益的方法和经验。

收稿日期: 2007-11-20, 修订日期: 2008-02-22

基金项目: 国家自然科学基金项目 (40571066), “十一五”国家支撑计划项目 (2006BAD10A09), 浙江省科技计划项目 (2004C32066), 浙江省教育厅科研项目 (20070228) 和泽泉科技基金资助

作者简介: 黄启厅, 1983 年生, 浙江大学环境与资源学院硕士研究生 e-mail: huangqiting830112@163.com

* 通讯联系人 e-mail: lianqing@zju.edu.cn

$$y = x(UX)'BQ \quad (4)$$

1 材料与方法

1.1 实验土样

土样来自 2004 年前期所做项目, 采自浙江省优势农产区, 采样及预处理方法见文献[9]。本实验从中随机抽取 47 个土样, 其中 30 个用于构建模型, 另外 17 个用于检验模型的预测能力。进行土壤 X 荧光光谱测定的同时, 采用等离子体发射光谱仪(ICP)测定其 Pb 含量。

1.2 实验仪器

本实验使用美国 NITON 公司 XLt920 型便携式 XRF 仪, 采用微型 X 射线管做激发源, 探测器使用了高性能的 Si-PIN, 热电(Peltier)制冷, 分辨率为 220 eV。可以同时测定 22 号元素钛(Ti)到 94 号元素钚(Pu)范围内的 25 种基本元素, 并可存储 3 000 个监测数据(带 X 射线光谱)。仪器对土壤 Pb 含量的定性检出限为 $12.0 \text{ mg} \cdot \text{kg}^{-1}$, 定量检出限为 $40.0 \text{ mg} \cdot \text{kg}^{-1}$ 。

1.3 光谱测量

取风干土样填满塑料样品盒, 样品盒高约 5 cm, 直径约 7 cm。土面略为整平后直接置于探头下部与之平行接触, 测定时间为 2 min。每个土样重复测定 3 次, 最后取其平均值作为该土样的 X 射线荧光光谱数据。

1.4 偏最小二乘法的基本原理

偏最小二乘法(PLS)是一种多元统计数据分析方法 由于其具有较强的提供信息的能力而成为化学计量学中的多变量校正法, 它集多元线性回归分析、典型相关分析和主成分分析的基本功能于“一体”, 能有效克服一般最小二乘回归分析方法无法解决的问题^[10]。PLS 从自变量矩阵和应变量矩阵中提取偏最小二乘成分, 有效地降维并消除自变量间可能存在的复共线关系, 明显改善了数据结果的可靠性和准确度^[11]。研究表明, 在对烟草、谷物等的近红外光谱分析中, PLS 方法的预测结果都优于一般的线性模型, 且对于连续光谱有明显的优势^[12]。谢丽娟等^[13]利用近红外光谱结合 PLS 法对杨梅汁可溶性固形物含量进行测定, 所建模型的相关系数达到 0.957 85, 预测均方根误差为 0.431, 结果满意。偏最小二乘法的基本原理如下。

将 n 个样品 m 个组分的浓度矩阵 $Y = (y_{ij})_{n \times m}$, 和仪器测定 n 个样品 p 个波长点处的 X 射线荧光强度矩阵 $X = (x_{ij})_{n \times p}$ 分解成特征向量的形式

$$Y = UQ + F \quad (1)$$

$$X = TP + E \quad (2)$$

式中 U , T —分别为 n 行 d 列(d 为抽象组分数)的浓度特征因子矩阵和 X 射线荧光强度特征因子矩阵; Q — $d \times m$ 阶浓度载荷矩阵; P — $d \times p$ 阶荧光强度载荷矩阵; F , E —分别为 $n \times m$, $n \times p$ 阶浓度残差矩阵和荧光强度残差矩阵。

偏最小二乘法(PLS)是根据特征向量的相关性分解 Y 和 X , 建立回归模型

$$U = TB + Ed \quad (3)$$

式中 Ed —随机误差矩阵; B — d 维对角回归系数矩阵。对待测样品, 如果测得其荧光强度向量为 x , 则浓度 y 为

1.5 数据处理与模型构建

本实验采用留一交互验证法(Leave-one-out), 模型的预测能力以模型对预测集的预测值与其真实值的相关系数 r 、预测均方根误差(root mean square error of prediction, RMSEP)来考核。为提高模型的稳定性和可靠性, 建模过程中分别用基线校正(baseline correction, BC)、Savitzky-Golay 平滑、多次散射校正(MSC)、变量标准归一化(SNV)、Nomalize 等预处理方法与直接利用原光谱数据建模进行比较。同时也用一元线性回归和多元线性回归分析与 PLS 回归分析做对比。

本实验对 X 射线荧光光谱数据的各种预处理、PLS 模型的构建均在 Unscrambler 9.7 (CAMO ASA, Trondheim, Norway)上实现。与 PLS 回归分析对比所做的线性回归和多元回归在 Microsoft EXCEL 2003 中实现。

2 结果与讨论

2.1 偏最小二乘法回归分析

2.1.1 最佳波段和主成分数的选择

Pb 元素的 X 射线荧光特征谱线 $K\alpha = 74.96 \text{ keV}$, $K\beta = 84.92 \text{ keV}$, $L\alpha = 10.55 \text{ keV}$, $L\beta = 12.61 \text{ keV}$, 从图 1 的光谱全谱图中可知, $K\alpha$ 和 $K\beta$ 处的强度几乎为零, 可忽略不计。这里只取 $L\alpha$ 和 $L\beta$ 两线。由于 X 射线在传输的过程中使空气电离, 能量会随着距离的增大而减弱, 另外土样中的其他元素也会对 Pb 的 X 射线荧光能量起到增强或者减弱的作用, $L\alpha$ 和 $L\beta$ 两线的强度并不能代表土壤中所有 Pb 的荧光能量强度, 所以需要在 Pb 的 $L\alpha$ 线和 $L\beta$ 线之外确定一个与 Pb 相关的光谱范围。从仪器显示的读数信息可以知道除了 $L\alpha$ 线和 $L\beta$ 线外, 与土壤中 Pb 密切相关的光谱范围有两个, 分别为: $10.40 \sim 10.70 \text{ keV}$ 和 $12.41 \sim 12.80 \text{ keV}$, 仪器每隔 0.025 keV 采一次样, 一共 29 个采样点。在这两个波段范围内 As, Ga, Os, Ir, Ge, Bi, Se, Pt, Au 九种元素的光谱与 Pb 的光谱有重叠, 均对 Pb 的测定造成干扰。

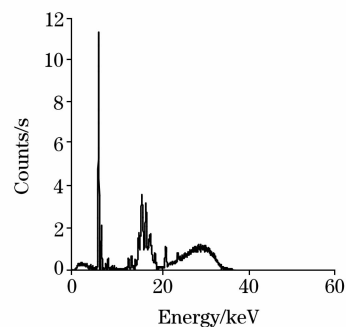


Fig. 1 X-ray fluorescence spectrogram

在构建 PLS 模型的过程中主成分数的选择关系到模型的预测能力, 选择的主成分数过少则不能充分反映样品中的光谱信息, 会造成不充分拟合; 选择的主成分数过多, 则把一些噪声信息也带入模型使得模型的预测误差加大, 造成过

分拟合。图 2 表示了不同成分数对预测精度的影响, 由图 2 可知当主成份数为 6 个时模型的预测误差最小, 故模型的最佳主成份数为 6。

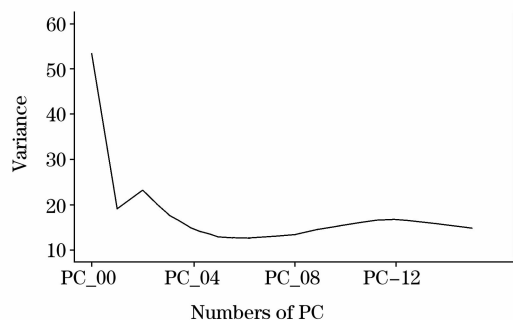


Fig. 2 Influence of different factors on the prediction precision

2.1.2 光谱预处理方法的选择

逯家辉等^[14]在用近红外光谱结合偏最小二乘法无损定量分析异烟肼片的研究中对原始光谱数据进行卷积平滑(Savitzky-Golay)和一阶导数处理, 结果较理想。李君霞等^[15]在水稻蛋白质近红外定量模型的创建及在育种中的应用的研究中, 对光谱数据进行一阶导数和变量标准化处理后所建的模型最佳。光谱数据预处理的方法很多, 为了减少和消除一些非目标因素的影响, 本文以斜率(Slope)、验证相关系数(r)、交互验证均方根误差(RMSEC)、验证标准差(SEC)4个参数作为模型的优化参数, 考察不同的预处理方法对模型预测精度的影响。结果见表 1。

Table 1 Influence of different pre-processing methods on the precision of model

Pre-processing method	Slope	r	RMSEC	SEC
NO*	0.981 2	0.990 6	0.963 7	0.982 1
BC*	0.403 6	0.635 3	6.526 9	6.646 6
Savitzky-Golay*	0.696 9	0.834 5	3.874 9	3.945 0
MSC*	0.511 2	0.715 0	5.908 7	6.077 2
SNV*	0.528 3	0.726 8	6.718 0	6.836 9
Normalize	0.422 0	0.649 6	5.690 9	5.799 3

* NO: Raw X-ray fluorescence spectrum date; * BC: Baseline correction; * Savitzky-Golay: Savitzky-Golay smooth; * MSC: Multiplicative scatter correction; * SNV: Standard normalized variate

从表 1 可知, 直接利用 X 射线荧光光谱数据进行建模比对光谱数据进行各种预处理后再建模效果要好, 交互验证均方根误差(RMSEC)和验证标准差(SEC)最小, 分别为 0.963 7 和 0.982 1; 交互验证相关系数 r 可以达到 0.990 6, 远高于经过各预处理后的相关系数。究其原因, 可能是相对于可见-近红外光谱, X 射线荧光光谱为特征谱, 谱线比较简单并且受到漫反射等的干扰比较少的缘故。

2.1.3 模型的构建和预测

用 30 个土样在 10.40~10.70 keV 和 12.41~12.80 keV 这两个光谱范围内的 X 射线荧光光谱数据以 6 个主成分数

建立 PLS 模型, 并对其余 17 个样品进行预测, 结果如表 2 所示。在对 17 个土样的预测中, 绝对误差的绝对值最大为 1.633 mg·kg⁻¹, 最小为 0.01 mg·kg⁻¹, 平均绝对误差 0.732 mg·kg⁻¹; 相对误差绝对值最大为 5.832%, 最小为 0.029%, 平均相对误差为 2.291%; 并且绝对误差和相对误差从整体上有随土样 Pb 含量的降低而增大的趋势。模型的预测均方根误差(RMSEP)等于 0.873 2, 预测值与真实值的相关系数 r 达到 0.9666, 结果令人满意。

Table 2 Statistical results predicted by PLS model

Predicted value /(mg·kg ⁻¹)	Measured value /(mg·kg ⁻¹)	Absolute error /(mg·kg ⁻¹)	Relative error/%
26.367	28.0	-1.633	-5.832
31.442	30.0	1.442	4.807
31.449	32.0	-0.551	-1.722
31.576	30.0	1.576	5.253
31.746	33.0	-1.254	-3.800
31.914	33.0	-1.086	-3.291
32.042	31.0	1.042	3.361
32.573	33.0	-0.427	-1.294
33.217	33.0	0.217	0.658
33.492	33.0	0.492	1.491
33.990	34.0	-0.010	-0.029
34.166	34.0	0.166	0.488
36.592	36.0	0.592	1.644
36.988	36.0	0.988	2.744
37.670	37.0	0.670	1.811
39.018	39.0	0.018	0.046
41.279	41.0	0.279	0.680
Mean		0.732*	2.291*

* The average of absolute value of absolute or relative errors

2.2 一元线性回归与多元线性回归分析

作为比较, 本实验用一元线性回归和多元线性回归分析法对全部 47 个土样也建立了相应的模型。一元线性回归用仪器直接获取的土壤 Pb 浓度读数与其 ICP 测定值做回归分析, 结果见式(5)。多元线性回归利用土壤样品 X 荧光光谱数据中 Pb 的 L_{α} 线和 L_{β} 线的强度数值与其 ICP 测定值进行回归分析, 结果见(6)式。

$$y = 0.456 2x + 13.34 \quad r = 0.680 5 \quad (5)$$

$$y = 237.149x_1 + 162.095x_2 - 0.470 6, \quad r = 0.730 2 \quad (6)$$

式中 y 为土壤样品 Pb 的 ICP 测定值, x 为 NITON XLt920 型 FPXRF 仪器直接获取的土壤 Pb 浓度读数, x_1 是 L_{α} 线的强度值, x_2 是 L_{β} 线的强度值。

结果表明, 用仪器所获得的浓度读数与 ICP 测定值作线性回归的效果最差, 相关系数仅为 0.680 5; 用 Pb 的 L_{α} 线和 L_{β} 线的强度数值与 ICP 值做线性回归的效果较好, 相关系数可以提高到 0.730 2, 但两者均远低于 PLS 的相关系数 0.966 6。图 3 是三种回归分析方法对预测集中 17 个样品的预测值与 ICP 值的散点图, 由图可知 PLS 的预测值最逼近 ICP 值, 其次是多元线性回归的预测值, 仪器读数偏离 ICP

值最远。说明 PLS 法能减少或消除噪声因素的干扰,比一元线性回归和多元线性回归能更有效地提取和利用光谱中的重要信息。

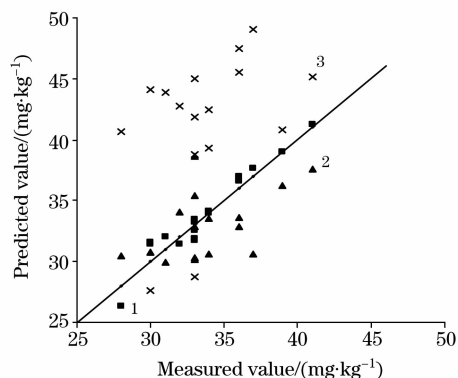


Fig. 3 Scatter plot between measured and predicted Pb value by three models

- 1: Predicted by PLS; 2: Predicted by multivariate regression;
3: Instrument reading

3 结 论

本研究利用 NITON XLt920 型便携式 XRF 仪所获取的 X 射线荧光光谱数据与土壤 Pb 的 ICP 测定数据构建了相应的 Pb 含量 PLS 预测模型,并从 X 射线荧光光谱数据中选择了与土壤中 Pb 密切相关的两个光谱范围,即 10.40~10.70 keV 和 12.41~12.80 keV。直接利用原光谱数据来建模效果要好于事先对光谱数据进行基线校正、多次散射校正、变量标准归一化等各种数据预处理。与传统的一元线性回归和多元线性回归分析相比,PLS 回归分析能明显提高模型预测的准确度。

本实验是在实验室相对稳定的条件下进行的,在野外的现场测定中很多因素都可能影响到测定的准确性,比如土壤的水分含量、土壤颗粒大小、土壤表面的状况等。这些因子对野外现场测定有什么影响、影响程度如何,还需要进一步的研究。另外,如何利用 X 射线荧光光谱建立适当的模型来对土壤中其他的重金属进行测定也是下一步研究的重点。

参 考 文 献

- [1] CHEN Tong-bin(陈同斌). Metal World(金属世界), 1999, (3): 10.
- [2] CHEN Yuan-pan(陈远盘). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 1995, 15(2): 103.
- [3] GUAN Ying, ZHAO Hai-yan, DING Xi-feng, et al(关颖, 赵海燕, 丁喜峰, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(5): 1029.
- [4] LI Zhong-shan(李忠山). Chinese Journal of Analysis Laboratory(分析实验室), 1998, 17(4): 73.
- [5] TENG Yan-guo, NI Shi-jun, ZHANG Cheng-jiang, et al(滕彦国, 倪师军, 张成江, 等). Nuclear Technology(核技术), 2003, 26(6): 440.
- [6] Clark C S, Thuppil V, Clark R, et al. Journal of Occupational and Environmental Hygiene, 2005, 2(1): 38.
- [7] Sterling D A, Lewis R D, Luke D A, et al. Environmental Research, 2000, 83(2): 174.
- [8] Argyraki A, Potts P J, Ramsey M H. The Analyst, 1997, 122(8): 743.
- [9] CHENG Jie-liang, SHI Zhou, ZHU You-wei(程街亮, 史舟, 朱有为). Journal of Soil and Water Conservation(水土保持学报), 2006, 20(1): 103.
- [10] TANG Qi-yi, FENG Ming-guang(唐启义, 冯明光). DPS Data Processing System for Practical Statistics(实用统计分析及其 DPS 数据处理系统). Beijing: Science Press(北京: 科学出版社), 2002. 35.
- [11] QU Hai-bin, LIU Quan, CHENG Yi-yu(瞿海斌, 刘全, 程翼宇). Chinese J. Anal. Chem. (分析化学), 2004, 32(4): 477.
- [12] WANG Hui-wen(王惠文). Partial Least Square Regression Method and Its Application(偏最小二乘回归方法及其应用). Beijing: National Defence Industry Press(北京: 国防工业出版社), 1999. 45.
- [13] XIE Li-juan, LIU Dong-hong, ZHANG Yu-huan, et al(谢丽娟, 刘东红, 张宇环, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(7): 1332.
- [14] LU Jia-hui, LÜ Xin, WANG Yue-xi, et al(逯家辉, 吕昕, 王跃溪, 等). Journal of Jilin University • Science Edition(吉林大学学报 • 理科版), 2006, 44(3): 485.
- [15] LI Jun-xia, ZHANG Hong-liang, YAN Yan-lu, et al(李君霞, 张洪亮, 严衍禄, 等). Agricultural Sciences in China(中国农业科学), 2006, 39(4): 836.

Analysis of Lead in Soil with Partial Least Square Regression (PLS) Method and Field Portable X-Ray Fluorescence(FPXRF) Analyzer

HUANG Qi-ting¹, ZHOU Lian-qing^{1*}, SHI Zhou¹, LI Zhen-yu², GU Qun³

1. College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310029, China

2. Zhejiang Environmental Monitoring Center, Hangzhou 310012, China

3. Zeal Quest Scientific Technology Co., Ltd, Shanghai 200052, China

Abstract In the present study, soil samples were scanned by NITON XLt920 field portable X-ray fluorescence (FPXRF) analyzer, and the relationship between the X-ray fluorescence spectra and the concentration of Pb in soil was studied. For predicating the Pb concentration in soil, a partial least square regression model (PLS) was established with 6 optimal factors and two closely relevant electron volt ranges: 10.40-10.70 keV and 12.41-12.80 keV. After cross-calibration, the correlation coefficient of value predicted by PLS model against that measured by ICP was 0.966 6, and the root mean square error of prediction (RMSEP) was 0.873 2. Meanwhile, the univariate linear regression and multivariate linear regression models were also built with the correlation coefficient of 0.680 5 and 0.730 2, respectively. Obviously, the PLS method was better than the other two methods for predication. Comparing to the conventional approach of atomic absorption spectroscopy(AAS), FPXRF has the advantages of rapidness, non-destruction and relatively low cost with the acceptable accuracy. It would be a powerful tool to decide which sample is needs for further analysis.

Keywords X-ray fluorescence(XRF)spectrum; Partial least square regression; Soil, Contamination by heavy metals

(Received Nov. 20, 2007; accepted Feb. 22, 2008)

* Corresponding author