

## 不同 pH 值土壤中铅含量的太赫兹光谱反演建模研究

李超<sup>1,2</sup>, 李斌<sup>2,3\*</sup>, 张丽琼<sup>2</sup>, 叶大鹏<sup>1\*</sup>, 郑书河<sup>1</sup>

1. 福建农林大学机电工程学院, 福建 福州 350012
2. 北京农业信息技术研究中心, 北京 100097
3. 农业部农业遥感机理与定量遥感重点实验室, 北京 100097

**摘要** 针对土壤中铅含量的定量检测问题, 本研究基于太赫兹光谱技术对不同 pH 下土壤中铅含量的最佳反演预测模型进行了探索性研究。分别制备了 pH 为 8.5, 7.0 和 5.5 的含铅土壤样品, 采集样品的太赫兹光谱数据, 并对光谱数据做了多元散射校正(MSC)、基线校正和 Savitzky-Golay 平滑等预处理。对预处理后的光谱数据, 采用连续投影法(SPA)选取光谱数据的特征频率。基于选取的特征频率分别采用偏最小二乘法(PLS)、支持向量机(SVM)和误差反向传播神经网络(BPNN)建立土壤中铅含量的反演预测模型, 采用校正集相关系数( $R_c$ )、校正集均方根误差(RMSEC)、预测集相关系数( $R_p$ )、预测集均方根误差(RMSEP)和剩余预测偏差(RPD)作为评价参数对模型性能进行评估, 确定铅在不同 pH 土壤中的最佳预测模型。实验结果表明: 在经过 SPA 选择特征频率后的建模效果普遍比全光谱的效果好。其中 pH 8.5 的样品最佳预测模型为 SPA-PLS,  $R_c$ ,  $R_p$ , RMSEC, RMSEP 和 RPD 分别为 0.997 7, 0.994 6, 14.52 mg · kg<sup>-1</sup>, 22.70 mg · kg<sup>-1</sup> 和 9.63; pH 7.0 的样品最佳预测模型为 SPA-SVM,  $R_c$ ,  $R_p$ , RMSEC, RMSEP 和 RPD 分别为 0.996 2, 0.975 7, 20.25 mg · kg<sup>-1</sup>, 33.04 mg · kg<sup>-1</sup> 和 4.56; pH 5.5 的样品最佳预测模型为 SPA-BPNN,  $R_c$ ,  $R_p$ , RMSEC, RMSEP 和 RPD 分别为 0.968 7, 0.974 4, 48.83 mg · kg<sup>-1</sup>, 55.03 mg · kg<sup>-1</sup> 和 4.44。该研究结果为不同 pH 土壤中铅含量的光谱反演预测提供了一种新思路, 亦可为其他重金属在不同 pH 土壤中的含量反演预测模型提供理论方法和技术支持。

**关键词** 土壤; 铅; 太赫兹光谱; 偏最小二乘法; 支持向量机; 神经网络

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)08-2397-06

## 引言

重金属具有生物毒性、不可降解性及在生物体中的累积性, 通过废水、大气沉降持续释放至周围环境的表层土壤中, 不断累积的重金属在土壤中通过有机体转化为有机复合物, 最终对自然生态系统和人类健康造成威胁<sup>[1-2]</sup>。铅作为环境中典型的重金属污染元素之一, 在进入土壤后, 容易通过溶解、沉淀、络合和吸附等方式以 5 种(可交换态、碳酸盐结合态、铁锰氧化物结合态、有机结合态和残渣态)不同的化学结合态富集于土壤中<sup>[3]</sup>。因此土壤中铅含量的检测对于防控土壤重金属污染具有重要意义。

目前利用光谱反演土壤重金属含量的研究主要集中在重

金属种类、重金属分布区域等方面。相关研究人员以河滩平原<sup>[4]</sup>、矿区<sup>[5-6]</sup>、农田<sup>[7]</sup>为例, 利用近红外光谱和高光谱探究了土壤光谱与 Cd, Cu, As, Ni 和 Pb 等重金属含量之间的关系, 通过光谱反演建立了含量预测模型。尽管上述土壤重金属含量光谱反演建模都取得了不错的预测效果, 然而对于利用太赫兹对不同 pH 土壤中重金属含量反演的研究却少有报道。相关研究<sup>[8-9]</sup>表明, 土壤光谱信息与重金属化学结合态的组分具有一定相关性, 而 pH 值是影响土壤重金属化学结合态的重要因素之一。因此, 有必要探索并建立重金属在不同 pH 土壤中的最佳反演模型, 以实现土壤重金属含量的准确预测。

李斌<sup>[10]</sup>等开展了土壤中铅含量的初步研究, 发现土壤样品中铅含量与对应的太赫兹吸收谱之间存在一定的对应关

收稿日期: 2019-07-16, 修订日期: 2019-11-28

基金项目: 北京市自然科学基金面上项目(6182012), 北京市农林科学院创新能力建设专项(KJCX20180421), 国家“十三五”重点研发计划项目(2016YFD0702002)资助

作者简介: 李超, 1993年生, 福建农林大学机电工程学院工程硕士研究生 e-mail: 819752082@qq.com

\* 通讯联系人 e-mail: lib@nercita.org.cn

系,因而得出了利用太赫兹光谱技术对土壤重金属含量进行测定具有可行性的结论。以此为理论依据,本文尝试利用太赫兹光谱技术建立不同 pH 土壤中铅含量的反演预测模型。首先采集土壤样品的太赫兹光谱,并对光谱做一定预处理,然后对预处理后的光谱数据,利用 SPA 选取光谱的特征频率,最后分别采用 PLS, SVM 和 BPNN 建立铅含量的反演预测模型,并通过比较模型效果,确定铅在不同 pH 土壤中的最佳预测模型。

## 1 实验部分

### 1.1 样品制备

制备样品所需的纯净土壤采集于北京市农林科学院院内实验田,采集 0~20 cm 的表层土壤,约 2.5 kg。取部分采集到的土壤样品送至北京中科联宏检测技术公司进行土壤的主要成分检测,根据国家标准《土壤环境质量农用地土壤污染风险管控标准》<sup>[11]</sup>检测土壤样品中的重金属元素是否已经超标。经检测后,采集的土壤样品中铅含量背景值为 29.476 mg·kg<sup>-1</sup>,符合国家标准。

为模拟铅离子自然状态下进入不同 pH 的土壤,利用 NaOH 溶液和 HCl 溶液分别配制 pH 为 5.5, 7.0, 8.5 的土壤,选用 0.1 mol·L<sup>-1</sup>的 Pb(NO<sub>3</sub>)<sub>2</sub> 标准溶液作为铅源加入土壤中。每种 pH 条件下按铅含量为 50~1 000 mg·kg<sup>-1</sup>, 梯度为 50 mg·kg<sup>-1</sup> 配制含铅土壤样品 20 组,每组 2 个,共计 120 个。另配制 5 个不含铅的土壤样品用作比较和分析铅进入土壤前后的太赫兹光谱曲线变化。含铅土壤经过干燥、研磨后称取 200 mg 含铅土壤样品,使用手动液压型压片机 (Specac GS15011, 英国)进行压片,压力为 3.5 T,压制时间 3 min。由于土壤样品本身特性,在压片过程中容易出现松散现象。为解决这一问题,在土壤中加入少量聚乙烯粉末实现样品制备工艺的优化。最终成型的压片,厚度约为 1.0~1.1 mm,直径约为 13 mm。

### 1.2 光谱数据采集与预处理

实验采用德国 Menlo Systems 公司的太赫兹时域光谱系统 TERA K15。为避免空气中的水蒸气对太赫兹波的强烈吸收所产生的干扰,将太赫兹波的产生与探测装置置于透明亚克力密闭箱内,并连续充入干燥氮气,保证测试环境相对湿度小于 5%,环境温度在 23 °C 左右。将压片固定在样品架上,然后置于 THz-TDS 系统中进行扫描以采集 0~80 ps 的时域光谱数据,每一个压片取不同部位 3 次测量的平均值作为该样品的光谱数据。最后将时域光谱数据输入到 PC 端的 Teralyzer 软件程序处理,即可从时域光谱数据中得到相应的吸收光谱数据。

样品的光谱数据在首端和高频段的信噪比较低,因此本文只取 0.075~2 THz 波段的数据用于后续分析。应用 MSC、基线校正和 Savitzky-Golay 平滑对光谱数据进行预处理,减少噪声、样品粒度和光程变化等因素对光谱产生的影响。其中,采用基线倾斜的方法对吸收光谱曲线进行基线校正,采用三次多项式 5 点平滑的方法对吸收光谱曲线进行 Savitzky-Golay 平滑。

### 1.3 模型建立与评价参数

PLS 是一种结合了多元线性回归分析、典型相关分析和主成分分析三种算法优点的化学计量分析法,广泛应用于光谱分析中<sup>[12]</sup>。在本研究的 PLS 建模中,根据样品数量和维度,初始设定 15 个主因子数,采用留一交互验证法对校正集进行交互验证,取交互验证均方根误差 (RMSECV) 最小值对应的主因子个数作为建模的最佳主因子数。

SVM 是一种有监督的学习方法,用于数据分析和模式识别<sup>[13]</sup>。本研究以径向基函数 (RBF) 为核函数,采用粒子群算法计算 *c* 和 *g*, 初始设置种群的粒子为 20, 学习因子 *c1* 和 *c2* 分别为 1.5 和 1.7; 设定惯性权重  $\lambda$  初始值为 0.9, 终止迭代次数为 200。粒子群算法在粒子的适应度值趋于稳定或达到最大迭代次数后,寻优结束,得出最佳的 SVM 模型参数 *c* 和 *g*。

BPNN 是一种基于误差反向传播算法的非线性多层前馈神经网络,它包括输入层、隐含层和输出层<sup>[14]</sup>。本研究初始设置输入层节点为样本的特征维度,而输出层节点、目标误差、最大迭代次数和学习率等参数分别为 10, 100, 200 和 0.01, 隐含层节点数根据经验公式 (1) 选取, 根据网络训练效果确定最优的隐含层节点数。当网络训练达到目标误差或最大迭代次数时,网络训练停止。

$$m = \sqrt{n+l} + \alpha \quad (1)$$

式 (1) 中, *m* 为隐含层节点数, *n* 为输入层节点数, *l* 为输出层节点数,  $\alpha$  为 1~10 之间的常数。

在 Matlab 2016a (MathWorks, USA) 软件中完成对原始光谱的预处理、特征频率的筛选、PLS, SVM 和 BPNN 模型的建立。在建立模型时,每种 pH 条件下的 40 个样本,按照 3:1 的比例划分成校正集和预测集。本文选用校正集相关系数 (*R<sub>c</sub>*), 校正集均方根误差 (RMSEC), 预测集相关系数 (*R<sub>p</sub>*), 预测集均方根误差 (RMSEP) 和剩余预测偏差 (RPD) 作为铅含量预测效果的评价指标。

## 2 结果与讨论

### 2.1 光谱特征及预处理

重金属铅进入土壤前后的吸收光谱曲线如图 1(a) 所示。可以观察到, 铅含量为 100 mg·kg<sup>-1</sup> 的土壤样品与纯净的土壤样品相比, 吸收系数增大且在 1.75 THz 处尤为明显, 这说明铅在进入土壤后能对太赫兹光谱产生一定影响。样品的原始吸收曲线经过 MSC、基线校正和 Savitzky-Golay 平滑等预处理后的光谱曲线如图 1(b) 所示。从图中可以看出, 经过预处理后的光谱曲线, 无效噪声得到抑制, 样品之间的分辨信息得到增强。且不同样品之间整体表现为吸收系数随铅含量的增加而逐渐增加。

### 2.2 全光谱的 PLS 建模分析

全光谱的 PLS 模型最佳主因子个数和预测结果如表 1 所示。pH 8.5 的样品模型预测结果最好, *R<sub>c</sub>*, RMSEC, *R<sub>p</sub>* 和 RMSEP 分别为 0.991 3, 27.51 mg·kg<sup>-1</sup>, 0.983 9 和 33.35 mg·kg<sup>-1</sup>, RPD 为 6.85。pH 7.0 的样品相关系数和 RPD 结果较好, 但均方根误差偏大。而 pH 5.5 的样品模型

预测结果较差,  $R_c$ , RMSEC,  $R_p$  和 RMSEP 分别为 0.688 9, 133.53  $\text{mg} \cdot \text{kg}^{-1}$ , 0.604 3 和 164.88  $\text{mg} \cdot \text{kg}^{-1}$ , RPD 仅为 1.25。结果表明 pH 7.0 和 pH 5.5 的样品模型预测结果需要进一步提高。

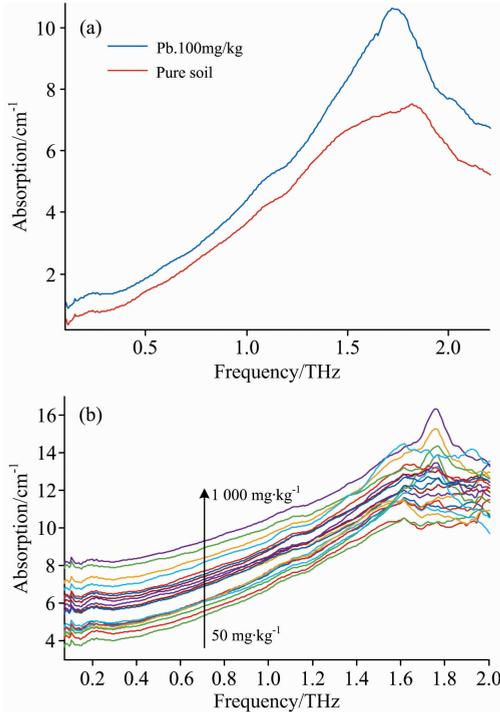


图 1 样品吸收曲线

(a): 铅进入土壤前后的吸收曲线; (b): 预处理后的吸收曲线

Fig. 1 Absorption curves of samples

(a): Absorption curves of lead before and after entering soil;

(b): Absorption curves after pretreatment

表 1 全光谱的 PLS 建模及预测结果

Table 1 The PLS modeling and prediction results of full spectrum

样品类型	主因子数	$R_c$	RMSEC/ ( $\text{mg} \cdot \text{kg}^{-1}$ )	$R_p$	RMSEP/ ( $\text{mg} \cdot \text{kg}^{-1}$ )	RPD
pH 5.5	11	0.688 9	133.53	0.604 3	164.88	1.25
pH 7.0	6	0.945 9	69.07	0.953 1	72.78	3.30
pH 8.5	8	0.991 3	27.51	0.983 9	33.35	6.85

### 2.3 SPA 特征频率选择

实验中发现, 0.075~2.0 THz 的光谱数据量较大且数据之间存在共线性和大量冗余的问题, 容易造成模型不稳定, 预测效果较差。因此需要对原始数据进行降维以提取特征变量, 减少冗余以及共线性数据的影响<sup>[11]</sup>。连续投影算法(SPA)是一种使矢量空间共线性最小化的前向变量选择算法, 在光谱分析中有广泛的应用。本文对样品的光谱数据进行特征频率选择, 选取性能好的特征频率, 将进一步提高模型分析的准确性。

如图 2 所示, 以 SPA 算法对样品光谱数据进行筛选, 得到了特征频率的分布图, 根据对应编号最终筛选出特征频率。pH 8.5 的样品筛选出 11 个特征频率(0.2, 0.419, 0.6,

0.788, 0.888, 1, 1.206, 1.306, 1.406, 1.469 和 1.544 THz)。pH 7.0 和 pH 5.5 的样品, 分别筛选了 10 个特征频率(0.144, 0.581, 0.775, 1.038, 1.094, 1.206, 1.325, 1.425, 1.531, 1.569 THz)和 13 个特征频率(0.081, 0.219, 1.356, 1.431, 1.513, 1.588, 1.656, 1.731, 1.8, 1.844, 1.875, 1.925 和 1.95 THz)。

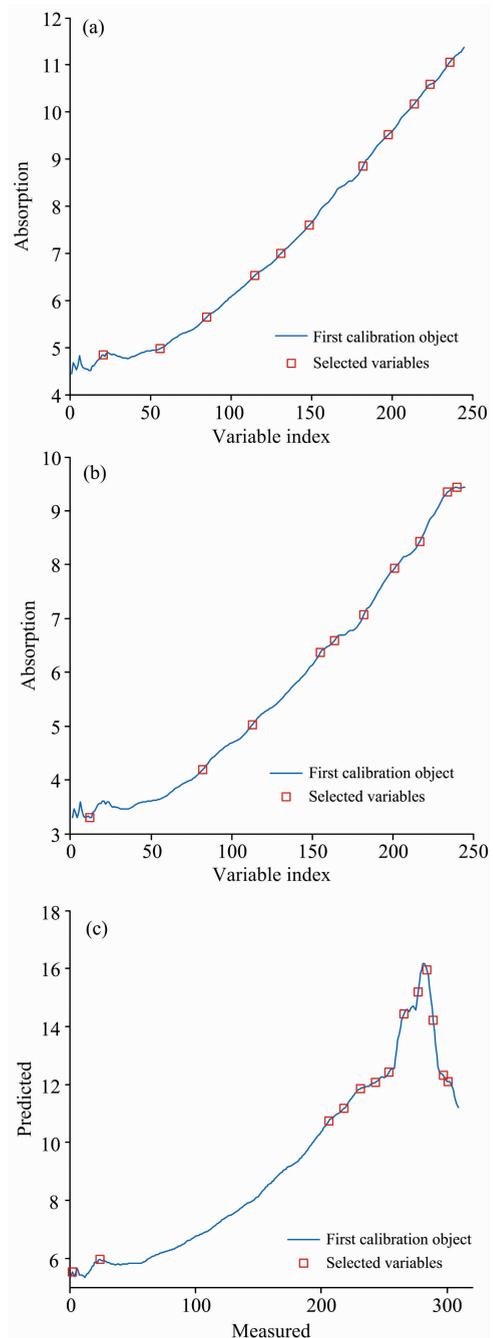


图 2 土壤样本的特征频率分布图

(a): pH 8.5; (b): pH 7.0; (c): pH 5.5

Fig. 2 Distribution of variables selected plot by SPA

(a): pH 8.5; (b): pH 7.0; (c): pH 5.5

### 2.4 基于特征频率的预测模型

基于 SPA 选择的特征频率建立了 SPA-PLS, SPA-SVM

表 2 基于特征频率的建模及预测结果

Table 2 The modeling and prediction results based on sensitive frequencies

样品类型	建模方法	参数选择	$R_c$	RMSEC/( $\text{mg} \cdot \text{kg}^{-1}$ )	$R_p$	RMSEP/( $\text{mg} \cdot \text{kg}^{-1}$ )	RPD
pH 5.5	SPA-PLS	主因子数=10	0.980 0	42.58	0.890 3	96.79	2.19
	SPA-SVM	$c=886.02, g=14.00$	0.985 5	32.86	0.946 2	61.69	3.09
	SPA-BPNN	Nodes=10	0.968 7	48.83	0.974 4	55.03	4.44
pH 7.0	SPA-PLS	主因子数=8	0.981 9	39.85	0.980 2	48.84	5.05
	SPA-SVM	$c=995.90, g=13.92$	0.996 2	20.25	0.975 7	33.04	4.56
	SPA-BPNN	Nodes=12	0.995 7	20.64	0.931 7	47.82	2.75
pH 8.5	SPA-PLS	主因子数=10	0.997 7	14.52	0.994 6	22.70	9.63
	SPA-SVM	$c=746.64, g=7.40$	0.990 2	28.03	0.978 4	43.70	4.83
	SPA-BPNN	Nodes=15	0.992 5	23.94	0.986 8	35.40	6.17

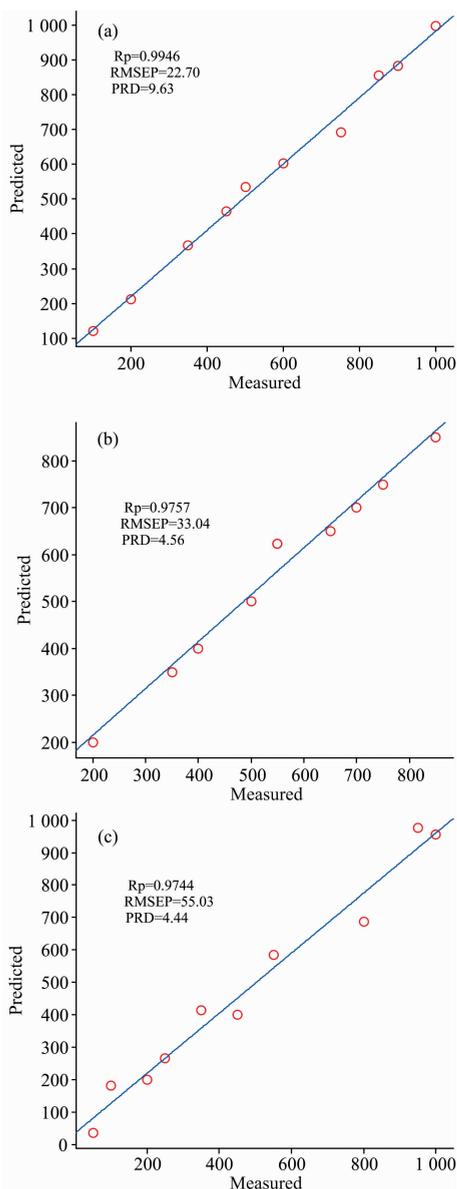


图 3 三种 pH 条件下样品的最佳模型预测图

(a): pH 8.5; (b): pH 7.0; (c): pH 5.5

Fig. 3 The best prediction results

(a): pH 8.5; (b): pH 7.0; (c): pH 5.5

和 SPA-BPNN 的预测模型, 其中 SPA-PLS 的最佳主因子个数, SPA-SVM 的最优  $c$  和  $g$ , SPA-BPNN 的最优隐含层节点数以及三种模型的预测结果如表 2 所示。可以看出, 在经过 SPA 选择特征频率后的建模效果普遍比全光谱效果好, 这表明 SPA 算法选择的特征频率不仅含有原始光谱的有效信息, 能代替原始光谱进行建模, 而且减少了数据量, 达到简化模型并提高模型的精度以及稳健性的目的。其中, pH 8.5 的样品模型预测结果中, 效果最好的为 SPA-PLS 模型,  $R_c$ ,  $R_p$ , RMSEC, RMSEP 和 RPD 分别为 0.997 7, 0.994 6, 14.52  $\text{mg} \cdot \text{kg}^{-1}$ , 22.70  $\text{mg} \cdot \text{kg}^{-1}$  和 9.63。pH 7.0 的样品模型预测结果较全光谱有了很大提升, 其中效果最好的为 SPA-SVM 模型,  $R_c$ ,  $R_p$ , RMSEC, RMSEP 和 RPD 分别为 0.996 2, 0.975 7, 20.25  $\text{mg} \cdot \text{kg}^{-1}$ , 33.04  $\text{mg} \cdot \text{kg}^{-1}$  和 4.56; pH 5.5 样品 SPA-PLS 模型中的预测效果相比于全光谱虽有所提升, 但 RMSEP 仍然高于样品的梯度值 50  $\text{mg} \cdot \text{kg}^{-1}$ , 说明 pH 5.5 条件下的样品数据不适合类似于 PLS 的线性预测模型。但在非线性的预测模型 SPA-SVM 和 SPA-BPNN 中, 得到了不错的预测效果, 其中预测效果最好的是 SPA-BPNN 模型,  $R_c$ ,  $R_p$ , RMSEC, RMSEP 和 RPD 分别为 0.968 7, 0.974 4, 48.83  $\text{mg} \cdot \text{kg}^{-1}$ , 55.03  $\text{mg} \cdot \text{kg}^{-1}$  和 4.44。基于同一种模型作比较, 还可以得出 3 种 pH 条件下的样品的预测效果高低排序依次为 pH 8.5 > pH 7.0 > pH 5.5, 可能是由于铅在进入碱性土壤后, 更容易与土壤产生络合反应, 形成影响太赫兹光谱曲线的化学结合态, 而具体是哪种形态铅影响太赫兹光谱的吸收, 目前尚且不清楚, 有待后续研究。其中三种 pH 条件下样品的最佳模型预测图如图 3 所示。

### 3 结论

应用太赫兹光谱技术对不同 pH 土壤中铅含量的最佳反演预测模型进行了探索性研究。制备了 3 种不同 pH 的含铅土壤样品, 采集样品的太赫兹光谱并做预处理。对预处理后的光谱数据进行了特征频率选择和建模分析, 并通过模型评价参数对不同模型进行比较。实验结果表明: (1) 在经过 SPA 选择特征频率后的建模效果普遍比全光谱的效果好; (2) 其中 pH 8.5 的样品最佳预测模型为 SPA-PLS,  $R_c$ ,  $R_p$ ,

RMSEC, RMSEP 和 RPD 分别为 0.997 7, 0.994 6, 14.52  $\text{mg} \cdot \text{kg}^{-1}$ , 22.70  $\text{mg} \cdot \text{kg}^{-1}$  和 9.63; (3) pH 7.0 的样品最佳预测模型为 SPA-SVM,  $R_c$ ,  $R_p$ , RMSEC, RMSEP 和 RPD 分别为 0.996 2, 0.975 7, 20.25  $\text{mg} \cdot \text{kg}^{-1}$ , 33.04  $\text{mg} \cdot \text{kg}^{-1}$  和 4.56; (4) pH 5.5 的样品最佳预测模型为 SPA-BPNN,  $R_c$ ,  $R_p$ , RMSEC, RMSEP 和 RPD 分别为 0.968 7, 0.974 4, 48.83  $\text{mg} \cdot \text{kg}^{-1}$ , 55.03  $\text{mg} \cdot \text{kg}^{-1}$  和 4.44。(5)建

模效果高低排序依次为  $\text{pH } 8.5 > \text{pH } 7.0 > \text{pH } 5.5$ 。本研究为不同 pH 土壤中铅含量的反演预测提供了一种新思路,但土壤取样具有区域局限性,重金属类型单一,且样本数量也有待进一步增加。在将来的研究工作中将进一步完善实验的不足之处,探索更加精准的土壤铅及其他重金属含量的反演预测模型。

## References

- [ 1 ] XU Xi-bo, LÜ Jian-shu, XU Ru-ru(徐夕博, 吕建树, 徐汝汝). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2018, 34(9): 216.
- [ 2 ] XIE Long-tao, PAN Jian-jun, BAI Hao-ran, et al(谢龙涛, 潘剑军, 白浩然, 等). Acta Pedologica Sinica (土壤学报), <http://kns.cnki.net/kcms/detail/32.1119.P.20190103.0927.002.html>.
- [ 3 ] YANG Xiu-min, REN Guang-meng, LI Li-xin, et al(杨秀敏, 任广萌, 李立新, 等). China Mining Magazine(中国矿业), 2017, 26(6): 79.
- [ 4 ] Liu Y L, Chen Y Y. Soil and Sediment Contamination: An International Journal, 2012, 21(8): 951.
- [ 5 ] CHEN Yuan-peng, ZHANG Shi-wen, LUO Ming, et al(陈元鹏, 张世文, 罗明, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2019, 50(1): 170.
- [ 6 ] Sun W C, Zhang X. International Journal of Applied Earth Observation and Geoinformation, 2017, 58: 126.
- [ 7 ] ZHANG Qiu-xia, ZHANG He-bing, LIU Wen-kai, et al(张秋霞, 张合兵, 刘文楷, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2017, 33(12): 230.
- [ 8 ] Sun W C, Zhang X, Sun X J, et al. Geoderma, 2018, 327: 25.
- [ 9 ] Shi T Z, Wang J J, Chen Y Y, et al. International Journal of Applied Earth Observation and Geoinformation, 2016, 52: 95.
- [ 10 ] LI Bin, ZHAO Chun-jiang(李斌, 赵春江). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2016, 47(S1): 291.
- [ 11 ] GB 15618—2018 National Standards of the People's Republic of China(中华人民共和国国家标准). Soil Environmental Quality Risk Control Standard for Soil Contamination of Agricultural Land(土壤环境质量农用地土壤污染风险管控标准), 2018.
- [ 12 ] FENG Hai-kuan, LI Zhen-hai, JIN Xiu-liang, et al(冯海宽, 李振海, 金秀良, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2016, 32(12): 165.
- [ 13 ] Liu W, Liu C H, Yu J J, et al. Food Chemistry, 2018, 251: 86.
- [ 14 ] Li M L, Dai G B, Chang T Y, et al. Applied Sciences, 2017, 7(2): 172.

## Terahertz Spectrum Inversion Modeling of Lead Content in Different pH Soils

LI Chao<sup>1,2</sup>, LI Bin<sup>2,3\*</sup>, ZHANG Li-qiong<sup>2</sup>, YE Da-peng<sup>1\*</sup>, ZHENG Shu-he<sup>1</sup>

1. College of Mechanical and Electrical Engineering, Fujian Agriculture and Forestry University, Fuzhou 350012, China

2. Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China

3. Key Laboratory of Quantitative Remote Sensing in Agriculture, Ministry of Agriculture, Beijing 100097, China

**Abstract** Aiming at the quantitative determination of heavy metal lead in soils, the optimal inversion prediction model of lead content in soils at different pH was studied based on terahertz spectroscopy. Lead-containing soil samples with pH of 8.5, 7.0 and 5.5 were prepared. Terahertz time-domain spectroscopy system TERA K15 was used to collect the Terahertz spectra of the samples, and multivariate scattering correction (MSC), baseline correction and Savoitzy-Golay smoothing were used to pre-process the spectra. For the spectral data of pre-treatment, successive projection algorithm (SPA) was used to select the sensitive frequencies of spectral data. Based on the selected sensitive frequencies, partial least squares (PLS), support vector machine (SVM) and back propagation neural network (BPNN) was used to establish inversion prediction models of lead content in the soil. The correlation coefficient of calibration ( $R_c$ ), root mean square error of calibration (RMSEC), the correlation coefficient of prediction ( $R_p$ ), root mean square error of prediction (RMSEP) and residual predictive deviation (RPD) were used as model

evaluation parameters to evaluate the performance of the model, and to determine the best prediction model of leadship in different pH soils. The experimental results show that the modeling effect after SPA choosing sensitive frequencies is generally better than that of full spectrum. Among them, the best prediction models for the samples with pH 8.5 were SPA-PLS,  $R_c$ ,  $R_p$ , RMSEC, RMSEP and RPD were 0.9977, 0.9946, 14.52 mg · kg<sup>-1</sup>, 22.70 mg · kg<sup>-1</sup> and 9.63, respectively; the best prediction models for the samples with pH 7.0 were SPA-SVM,  $R_c$ ,  $R_p$ , RMSEC, RMSEP and RPD were 0.9962, 0.9757, 20.25 mg · kg<sup>-1</sup>, 33.04 mg · kg<sup>-1</sup> and 4.56, respectively; and the samples with pH 5.5 were the best. The prediction models are SPA-BPNN,  $R_c$ ,  $R_p$ , RMSEC, RMSEP and RPD are 0.9687, 0.9744, 48.83 mg · kg<sup>-1</sup>, 55.03 mg · kg<sup>-1</sup> and 4.44, respectively. The results provide a new idea for inversion prediction of lead content in different pH soils, and also provide theoretical methods and technical support for other heavy metals inversion prediction models in different pH soils.

**Keywords** Soil; Lead; Terahertz spectrum; PLS; SVM; BPNN

(Received Jul. 16, 2019; accepted Nov. 28, 2019)

\* Corresponding authors

## 《光谱学与光谱分析》期刊社决定采用 ScholarOne Manuscripts 在线投稿审稿系统

《光谱学与光谱分析》期刊社与汤森路透集团签约,自 2010 年 12 月 1 日起《光谱学与光谱分析》决定采用 Thomson Reuters 旗下的 ScholarOne Manuscripts 在线投稿审稿系统。

- ScholarOne Manuscripts, 该系统不仅能轻松处理稿件,而且能提速科技交流。
- 全球已有 360 多家学会和出版社的 3 800 多种期刊选用了 ScholarOne Manuscripts 系统作为在线投稿、审稿平台,全球拥有超过 1 350 万的注册用户,代表着全球学术期刊在线投审稿的一流水平。
- ScholarOne Manuscripts 与 EndNote, Web of Science 无缝链接和整合;使科研探索、论文评阅和信息传播效率大为提高。
- ScholarOne Manuscripts 是汤森路透科技集团的一个业务部门,拥有丰富的学术期刊业务经验,为学术期刊提供综合管理工作流程系统,使期刊更有效管理投稿、同行评审、加工和发表过程,提高作者心中的专业形象,缩短论文发表时间,削减管理成本,帮助期刊提高科研绩效和实现学术创新。

《光谱学与光谱分析》采用“全球学术期刊首选的在线投稿审稿系统—ScholarOne Manuscripts”,势必对 2010 年 11 月 30 日以前向本刊投稿的作者在查阅稿件信息时,会带来某些不便,在此深表歉意!为了推进本刊的网络化、数字化、国际化进程,以实现与国际先进出版系统对接;为了不断提高期刊质量,加快网络化、数字化建设,加快与国际接轨的进程,希望能得到广大作者、读者们的支持与理解,对您的理解和配合深表感激。这是一件新事物,肯定有不周全、不完善的地方,让我们共同努力,不断改进和完善起来。

《光谱学与光谱分析》期刊社

2010 年 12 月 1 日