

土壤铅污染光谱的 HHT 鉴别及 BC-PLSR 铅含量预测模型

付萍杰^{1,2}, 杨可明^{1,2*}, 程 龙^{1,2}, 王 敏^{1,2,3}

1. 中国矿业大学(北京)煤炭资源与安全开采国家重点实验室, 北京 100083
2. 中国矿业大学(北京)地球科学与测绘工程学院, 北京 100083
3. 华北理工大学, 河北 唐山 063210

摘 要 土壤重金属污染问题一直备受关注, 利用高光谱遥感对其进行研究取得了大量的成果, 主要集中在利用土壤光谱的导数变换、连续统去除等常规方法预测土壤重金属含量上。土壤光谱数据与非线性非平稳的机电信号、医学信号等具有一定的相似性。通过希尔伯特黄变换(Hilbert-Huang transform, HHT), 对土壤铅(Pb)污染光谱进行频率域分析, 实现土壤 Pb 污染光谱的 HHT 鉴别, 并建立土壤 Pb 含量预测模型。首先, 进行土壤 Pb 污染实验, 采集土壤 Pb 污染样品的光谱、含水率及有机质含量; 其次, 通过土壤 Pb 污染样品光谱的 HHT 时频分析和第二个本征模函数(intrinsic mode function, IMF)分量(IMF2)瞬时频率的二阶导数识别土壤 Pb 污染的特征波段; 最后, 选择合适的频率域参数、土壤光谱一阶导数、土壤有机质含量及土壤含水率作为参数, 利用箱形图、聚类分析、偏最小二乘法建立土壤 Pb 含量预测模型。研究表明: 土壤 Pb 污染的 HHT 时频分析图可以鉴别土壤 Pb 污染光谱, 未受污染的土壤光谱 HHT 时频分析图在波段序列为 250~430 之间没有异常信号, Pb 污染土壤的光谱 HHT 时频分析图在波段序列为 250~430 之间存在多个异常信号, 并且随着浓度的升高, 异常信号分布范围越来越广, 当污染浓度达到 $800 \mu\text{g} \cdot \text{g}^{-1}$ 时, 土壤样品的光谱信号在波段序列为 270 处、频率为 0.3 Hz 之前出现了较强的异常信号; 土壤 Pb 污染光谱经验模态分解(empirical mode decomposition, EMD)处理后, 得到的未受污染的土壤光谱 IMF2 的瞬时频率的二阶导数的突变非常微弱, 而 Pb 污染的土壤光谱 IMF2 的瞬时频率的二阶导数存在明显的突变点, 根据突变点及土壤 Pb 污染光谱的 IMF2 的瞬时频率的二阶导数识别的土壤 Pb 污染光谱的特征波段区间为 2 150~2 300 nm; 利用不同浓度 Pb 污染下土壤光谱 Hilbert 能量谱峰值、EMD 能量熵、一阶导数、有机质和含水率, 通过箱形图去除了六组异常样品, 然后利用聚类分析的方法将去除异常样品后的土壤 Pb 污染样品分为两类, 最后将 Hilbert 能量谱峰值、EMD 能量熵、2 134 nm 波段一阶导数、790 nm 波段一阶导数、1 276 nm 波段一阶导数、2 482 nm 波段一阶导数、有机质和含水率作为参数建立两类数据的 BC-PLSR(boxplot cluster-partial least squares regression)模型预测土壤中 Pb 含量, 经验证模型精度较高, 相关系数分别为 0.88 和 0.99。

关键词 土壤 Pb 污染光谱; HHT 分析; 瞬时频率; BC-PLSR 模型

中图分类号: X87 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)05-1543-08

引 言

由于快速的工业化和城市化, 重金属被发现在土壤、岩石、植物, 甚至人或动物体内以各种形式存在。土壤是人类生存环境的重要载体, 并且为农作物的生长提供营养和能

量, 因此, 土壤污染问题一直备受关注, 被认为是对人类健康构成潜在风险的全球性问题。

过去二十年的许多研究探索了利用土壤反射辐射的新的遥感技术^[1]。许多不同性质的土壤具有不同的光谱特征, 这样就提供了利用田间或实验室辐射测量法定量地确定土壤重金属含量的可能性^[2]。Liu 等研究德兴铜矿周围土壤重金属

收稿日期: 2018-03-27, 修订日期: 2018-08-05

基金项目: 国家自然科学基金项目(41271436), 煤炭资源与安全开采国家重点实验室 2017 年开放基金项目(SKLCRSM17KFA09, SKLCRSM17KFB04)资助

作者简介: 付萍杰, 女, 1989 年生, 中国矿业大学(北京)煤炭资源与安全开采国家重点实验室博士研究生 e-mail: fpj890622@163.com

* 通讯联系人 e-mail: ykm69@163.com

污染^[2], 发现有有机质含量与土壤光谱 624 波段和 564 波段的一阶导数的比值具有较强的相关性, 因此, 利用土壤两个波段的光谱数据线性反演土壤中重金属含量。Stazi 等采集图西比亚大学实验农场(意大利中部维泰博)土壤样品进行土壤砷含量预测的高光谱实验, 发现 PLSR 法预测砷含量的精度较高^[3]。Maliki^[4]等采集了澳大利亚五个地区的不同类型的土壤, 除了一个被 Pb 污染的地区的土壤外, 其他四个未见明显的 Pb 污染的地区都加入 $\text{Pb}(\text{CH}_3\text{COO})_2 \cdot 3\text{H}_2\text{O}$, 根据光谱数据建立土壤中 Pb 含量预测模型。Niazi 等在位于沃龙巴的废弃的牛蘸站点采集土壤样品, 对光谱数据进行一阶导数等预处理, 采用 PLSR 法建立土壤中砷含量预测模型^[5]。Gholizadeh 等采集了捷克共和国比纳纳和图西米斯的六个倾倒矿场土壤样本, 研究了七种重金属含量的预测模型^[6]。Gholizadeh 等研究捷克共和国两个矿区的土壤重金属污染^[7], 采集六个重金属污染区的土壤光谱数据, 基于支持向量机建立重金属含量预测模型。Fard 和 Matinfar 采集了伊朗地区重金属污染的土壤样本, 根据变换光谱、有机质与重金属含量的相关性, 分别采用直接方法和间接方法预测重金属含量^[8]。

通过以上叙述可以看出, 土壤重金属污染遥感的研究主要集中在利用光谱导数变换、连续统去除等常规方法进行的含量预测。由于土壤光谱数据具有非线性非平稳的特征, 这与非线性非平稳信号具有一定的相似性。故此拟利用 HHT 变换处理土壤 Pb 污染光谱数据, 实现在频率域中的土壤 Pb 污染光谱的鉴别及土壤 Pb 污染特征波段的识别, 并利用土壤 Pb 污染光谱的频率域参数、一阶导数及土壤含水率和有机质含量, 建立土壤 Pb 含量的 BC-PLSR 预测模型。

1 实验部分

1.1 土壤 Pb 污染实验

不同浓度重金属 Pb 污染土壤样品是这样制备的。选取北京市一片废弃的土地, 采集 0~20 cm 的表层土壤, 在实验室内对土壤进行预处理, 先过 5 mm 筛, 再过 2 mm 筛, 然后将其混合均匀, 平均分成 24 份样品。将 $\text{Pb}(\text{NO}_3)_2$ (分析纯), 浓度配制为 0, 100, 150, 200, 300, 400, 600 和 800 $\mu\text{g} \cdot \text{g}^{-1}$, 每种浓度平行量取 3 份分别混合在预处理后的每份土壤样品中。使用光谱范围为 340~2 514 nm 的 SVC HR-1024I 高性能地物光谱仪采集土壤样品的光谱, 8 种污染梯度 24 份土壤样品分别标记为 CK-1, CK-2, CK-3, ..., Pb(800)-1, Pb(800)-2, Pb(800)-3。最后检测不同浓度 Pb 污染的土壤样品中的 Pb 含量、土壤含水率和土壤有机质含量。

1.2 HHT 变换

在信号的时频分析领域, 有很多传统的数据处理方法^[9-10], 如傅里叶变换、小波变换等, 大多数信号处理方法都受线性或平稳性的束缚。但是, HHT 彻底摆脱了线性和平稳性的束缚, 是目前应用较为广泛的信号时频分析方法之一。对 EMD 得到的 IMF 进行 Hilbert 变换可得信号的时频谱, 从而得到信号 IMF 中具有物理意义的瞬时频率, 还可以

得到 Hilbert 谱, 从而得到信号的频率、幅度和能量随时间变化规律, 国内外专家和学者开展了大量的基于 EMD 的机械故障识别的方法^[11-14]、基于 HHT 变换的电能质量检测^[15]等。HHT 变换主要包含两个部分: EMD 和 Hilbert 变换。具体的计算过程如下:

假设 $\mathbf{X}(t)$ 为一条复杂的原始光谱曲线, 则 EMD 的处理过程为: (1) 找出 $\mathbf{X}(t)$ 所有的极大值点和极小值点, 并用三次样条插值函数拟合形成原光谱曲线的上、下包络线; 然后计算该上、下包络线的均值(记作曲线 m); 再用式 $h = \mathbf{X}(t) - m$ 从 $\mathbf{X}(t)$ 中减去 m , 得到一个新的曲线 h 。如果 h 不满足 IMF 的两个条件, 需要重复上述步骤, 直到分解出第一个分量 IMF1; (2) 从 $\mathbf{X}(t)$ 中减去 IMF1 得到第一阶剩余光谱 r_1 , 把 r_1 作为新的原光谱曲线, 重复步骤(1), 得到 IMF2 与 r_2 ; 以此类推, 可得出 IMF3 与 r_3, \dots, IMF_n 与 r_n , 当 r_n 变为一个单调函数时, 筛选结束。得到 n 个由高频到低频的 IMF。

对 EMD 处理后的原始土壤光谱曲线的 IMF 进行 Hilbert 变换, 可得相应的 Hilbert 谱, 然后, 汇总所有 IMF 的 Hilbert 谱就会得到原始土壤光谱曲线的 Hilbert 谱。Hilbert 变换可以由下式求得

$$Y(t) = a_j(t)e^{i2\pi\int f_j(t)dt} \quad (1)$$

用三维图形表达振幅、频率和波段序列之间的关系, 就可以得到光谱的 Hilbert 谱 $H(f, t)$ 。

Hilbert 能量谱可以测量光谱在每个频率的能量, 具有能量密度的物理意义, 表达了光谱每个频率在整个波段区间内所积累的能量及能量分布情况。其计算公式如式(2)

$$ES(f) = \int_0^T H^2(f, t) dt \quad (2)$$

非平稳信号的频率随时间的变化而不断变化, 因此, 传统意义上的频率已经不能适用于非平稳信号的描述, 而瞬时频率作为一种新的物理量, 可以描述非平稳信号在不同时间中的频率变化。瞬时频率的计算过程如下,

设某个波段区间内土壤光谱的 IMF i 为 $s(t) = a(t)\cos\varphi(t)$, 其中 $a(t)$ 是光谱的幅值, $\varphi(t)$ 为光谱的相位。对 $s(t)$ 进行 Hilbert 变换可得其共轭光谱 $q(t)$

$$q(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{s(\tau)}{t-\tau} d\tau = a(t)\sin\varphi(t) \quad (3)$$

利用光谱 $s(t)$ 和其共轭光谱 $q(t)$, 可得解析光谱

$$z(t) = s(t) + jq(t) = a(t)[\cos\varphi(t) + j\sin\varphi(t)] = a(t)e^{j\varphi(t)} \quad (4)$$

对上述解析光谱进行傅里叶变换可得

$$Z(f) = \int_{-\infty}^{+\infty} z(t)e^{-j2\pi ft} dt = \int_{-\infty}^{+\infty} a(t)e^{j[\varphi(t)-2\pi ft]} dt \quad (5)$$

由驻定相位可得, 上述积分取最大值时, $\frac{d}{dt}[\varphi(t) - 2\pi ft] = 0$, 设 f_s 为驻定相位点, 那么,

$$f_s = \frac{1}{2\pi} \frac{d\varphi(t)}{dt} \quad (6)$$

根据式(6)可得, f_s 可以衡量光谱在频域的能量集中度, 那么瞬时频率可以通过该光谱对应的解析光谱的相位关于时间的导数求得, 即

$$f'(t) = \frac{1}{2\pi} \frac{d}{dt} \arg z(t) \quad (7)$$

1.3 EMD 能量熵

光谱经过 EMD 处理后, 每个 IMF 的能量分别为 E_1, E_2, \dots, E_n , 那么整条光谱的能量 $E = [E_1, E_2, \dots, E_n]$, 因此得到了光谱在频率域中的一种划分方式, EMD 能量熵可以用式(8)和式(9)求得

$$p_i = E_i / E (i = 1, 2, \dots, n) \quad (8)$$

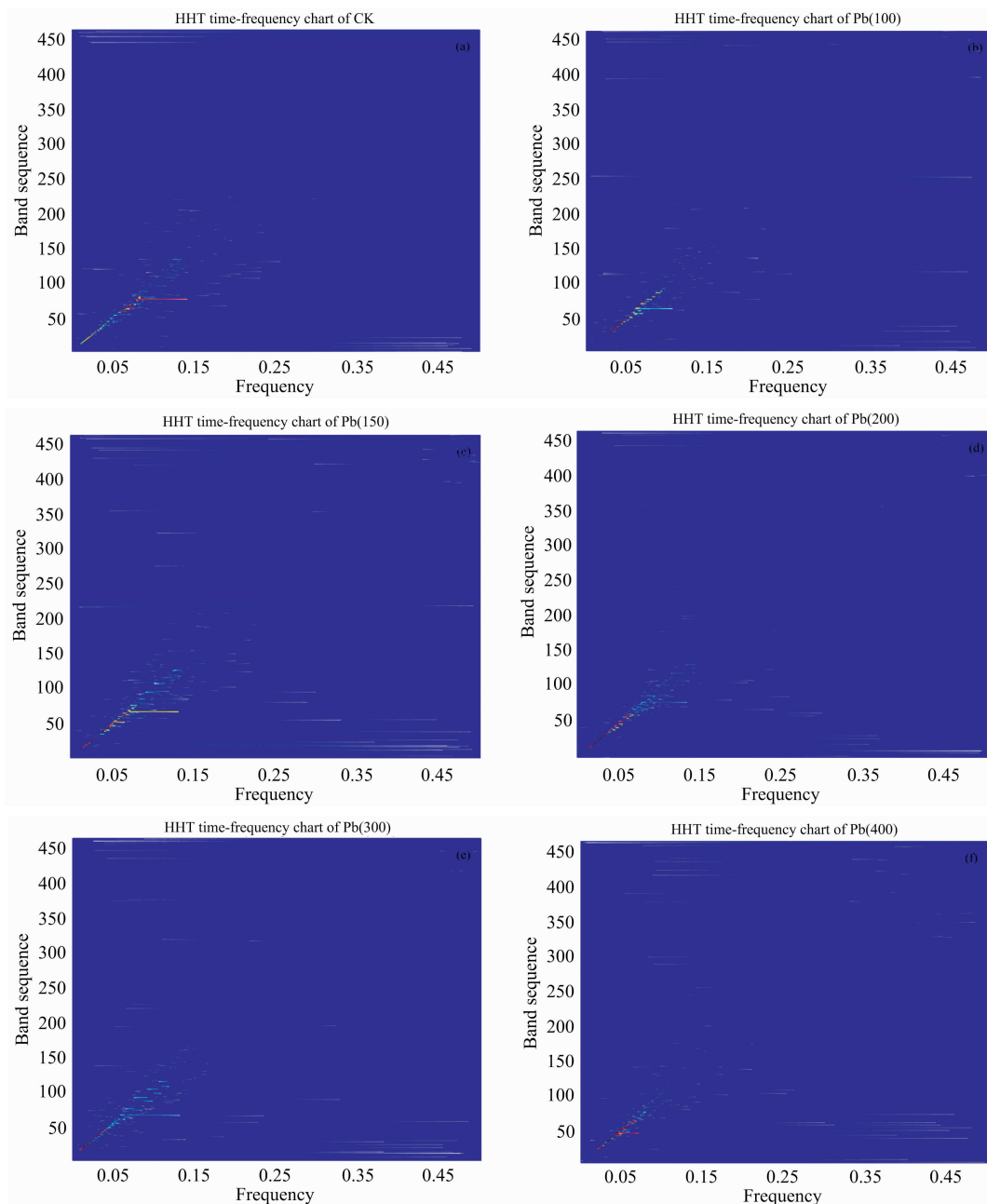
$$E_m = - \sum_{i=1}^n p_i \log p_i \quad (9)$$

其中, E_i 为第 i 个 IMF 能量, p_i 为第 i 个 IMF 能量占整条光谱曲线能量的百分比。

2 结果与讨论

2.1 HHT 时频分析

取同种浓度 Pb 污染下的三条土壤样品光谱的平均值作为该浓度下的 Pb 污染土壤光谱, 分别标记为 CK, Pb(100), ..., Pb(800)。根据以往研究, 与土壤重金属浓度相关性较强的光谱反射率位于 1 800~2 500 nm。根据式(1), 基于 HHT 变换研究该波段区间内不同浓度 Pb 污染土壤光谱的变化规律。从不同土壤样品光谱 HHT 时频分析图中(图 1)可以看出, 所有土壤样品的光谱信号在频率为 0.25 Hz 之前、波段序列 250 之前均有分布, 而 CK 样品在频率为 0.25 Hz 之后、波段序列为 250 之后的 HHT 时频分析图[图 1(a)]与其他浓度 Pb 污染下的土壤光谱的 HHT 时频分析图[图 1(b—h)]具有明显的差异, Pb 污染下的土壤光谱 HHT 时频分析图存在多个异常信号, 且随着浓度的升高, 异常信号分布范围越来越广, 当污染浓度达到 $800 \mu\text{g} \cdot \text{g}^{-1}$ 时, 样品的光谱信号在波段序列为 270 处、频率为 0.3 Hz 之前出现了较强的异



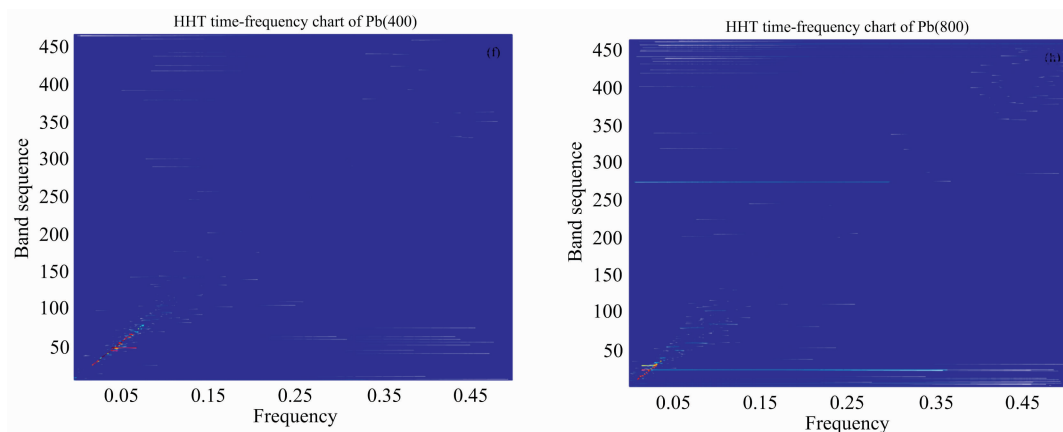


图 1 不同浓度 Pb 污染下土壤光谱 HHT 时频分析图

Fig. 1 The HHT time-frequency charts of soil's spectra under the Pb pollution with different concentrations

常信号。

2.2 IMF2 瞬时频率分析

根据土壤样品光谱 HHT 时频分析图的差异, 利用 EMD 进一步分析 Pb 污染土壤光谱的特征波段。由于光谱噪声成分主要集中在高频段的 IMF 分量上^[16], 为了防止 IMF1

中光谱噪声的存在, 选取 IMF2 作为研究对象, 根据式(3)~式(7)得到土壤样品光谱的 IMF2 分量的瞬时频率(图 2)。从图中可以看出瞬时频率的突变区间大体位于 2 100 nm 之后的波段区间内, 但是曲线波动仍然比较杂乱。

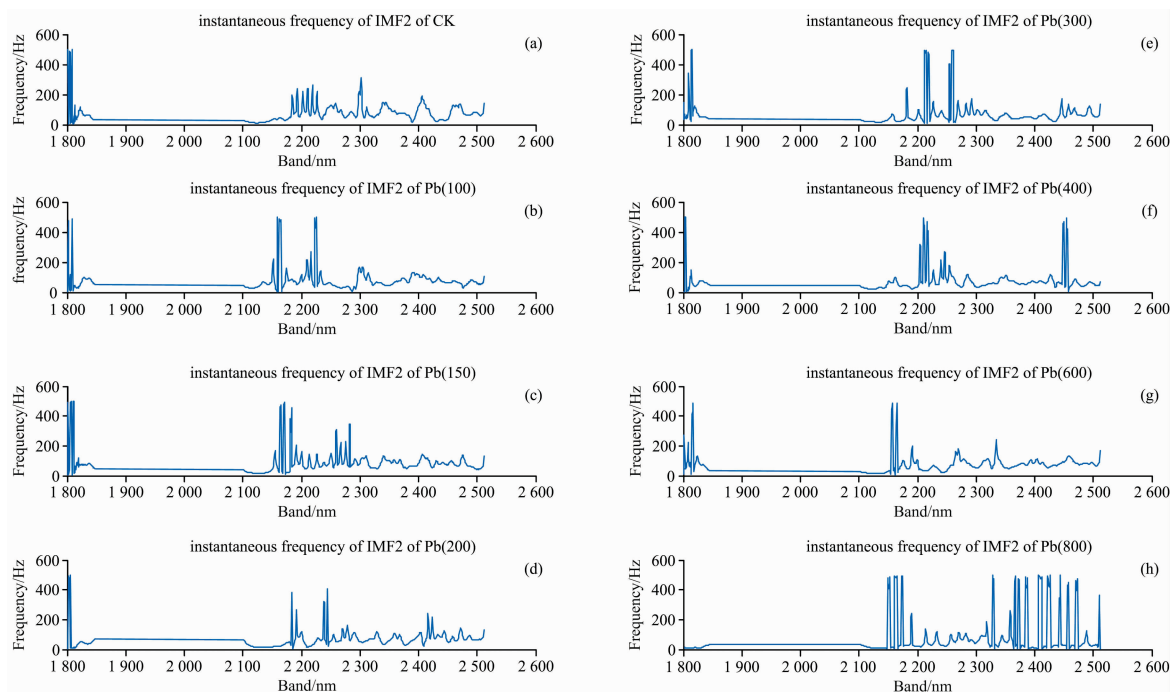


图 2 不同浓度 Pb 污染下土壤光谱 IMF2 瞬时频率

Fig. 2 The IMF2 instantaneous frequency of soil's spectra under the Pb pollution with different concentrations

为了突出不同浓度 Pb 污染下土壤光谱 IMF2 瞬时频率之间的差异, 对 IMF2 瞬时频率进一步做二阶导数处理(图 3), 从图 3 中可以看出, CK 样品光谱的突变基本上非常微弱, 大体呈一条直线分布[图 3(a)], 而其他土壤样品光谱 IMF2 瞬时频率的二阶导数在几个波段上突变显著[图 3(b-h)]; Pb(100)的突变波段为 2 150~2 250 nm, Pb(150)的突

变波段为 2 150~2 300 nm, Pb(200)的突变波段为 2 150~2 250 nm, Pb(300)的突变波段为 2 200~2 300 nm, Pb(400)的突变波段为 2 200~2 250 和 2 400~2 500 nm, Pb(600)的突变波段为 2 150~2 200 nm, Pb(800)的突变波段为 2 150~2 200 和 2 300~2 500 nm。综合分析可得, Pb 污染土壤光谱的特征波段区间为 2 150~2 300 nm 波段。

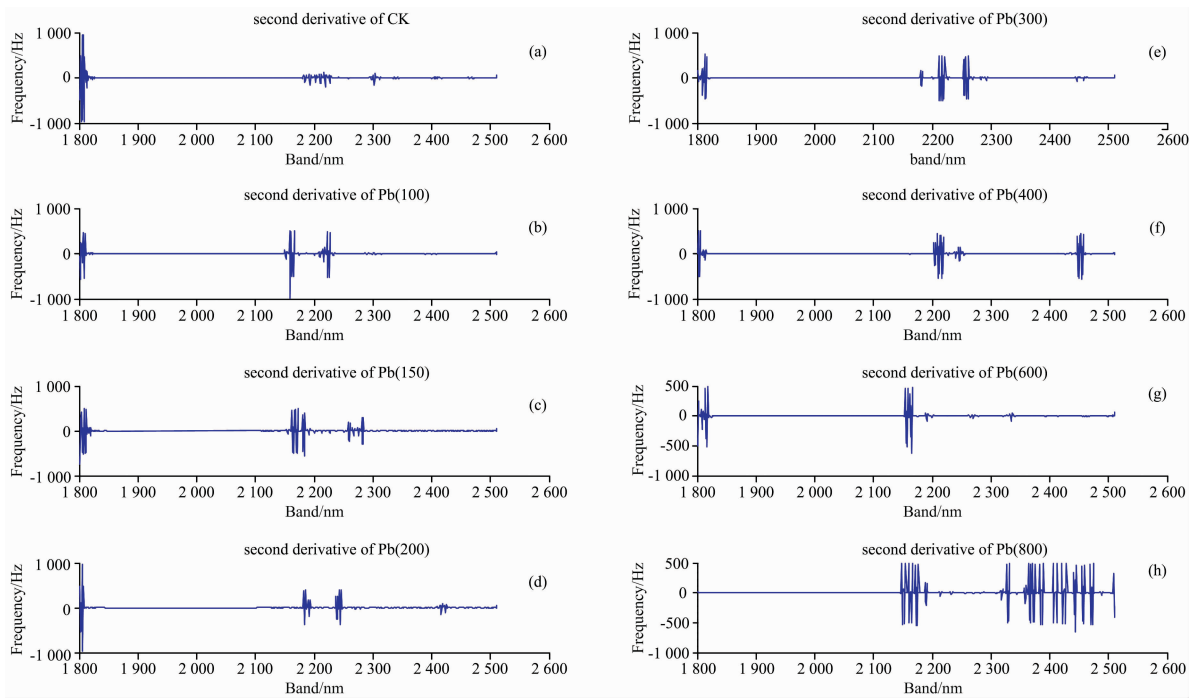


图 3 不同浓度 Pb 污染下土壤光谱 IMF2 瞬时频率二阶导数

Fig. 3 The second derivative of IMF2 instantaneous frequency of soil's spectra under the Pb pollution with different concentrations

2.3 土壤 Pb 含量的 BC-PLSR 预测模型

利用不同浓度 Pb 污染下的土壤光谱的 Hilbert 谱 $H(f, t)$ 的平方对时间的积分, 根据式(2)可以得到不同浓度 Pb 污染下的土壤光谱的 Hilbert 能量谱(图 4), 从图 4 中可以看

出, 不同浓度 Pb 污染下的土壤光谱的 Hilbert 能量主要集中在 0~50 Hz 之间, 并且有一个较高的峰值。对 Hilbert 能量谱峰值进行统计分析发现, 随着 Pb 污染浓度的升高 Hilbert 能量谱峰值呈上升的趋势(图 5)。

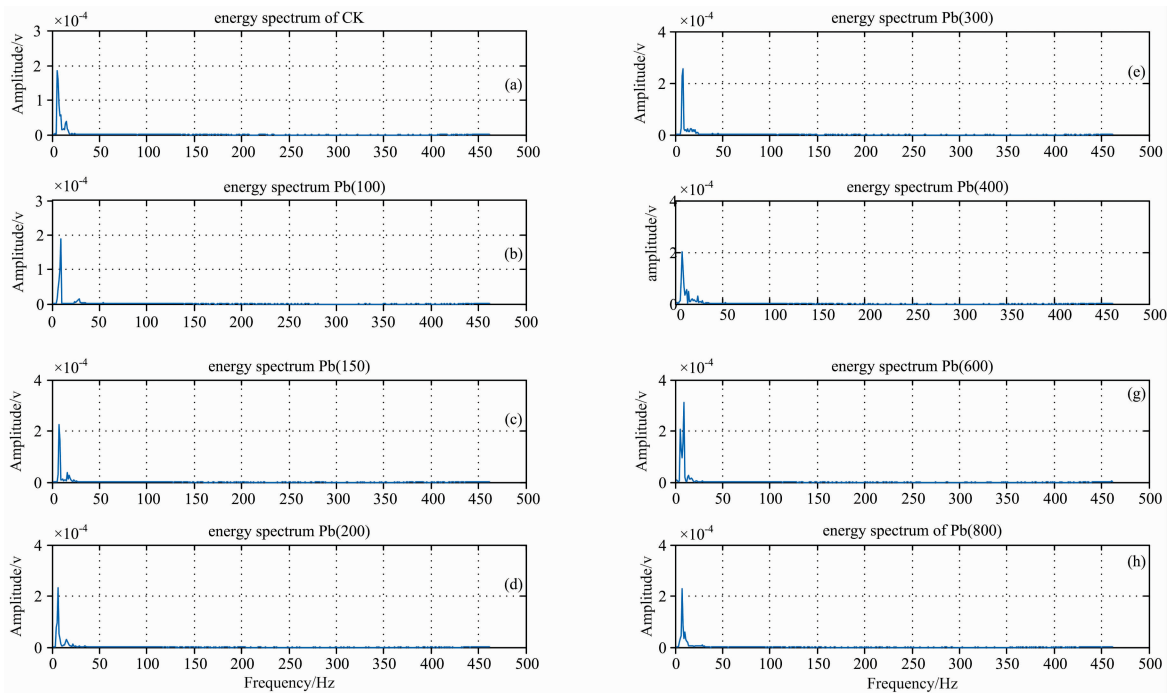


图 4 不同浓度 Pb 污染下土壤光谱 Hilbert 能量谱

Fig. 4 The Hilbert energy spectra of soil's spectra under the Pb pollution with different concentrations

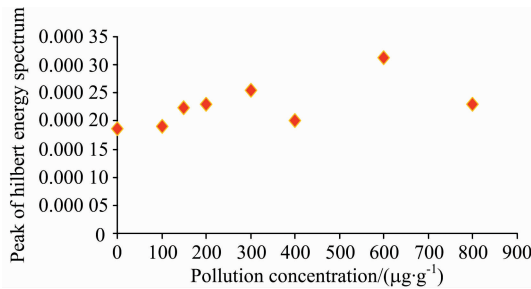


图 5 不同浓度 Pb 污染下土壤光谱 Hilbert 能量谱峰值

Fig. 5 The peaks of Hilbert energy spectra of soil's spectra under the Pb pollution with different concentrations

利用式(8)和式(9)计算 24 份土壤样品的 EMD 能量熵 (E_m), 根据不同浓度 Pb 污染下的土壤光谱的 Hilbert 能量谱峰值 (E_a) 与浓度变化趋势之间的关系, 将 E_m 、 E_a 、有机质、含水率以及与 Pb 含量相关性较高的波段处的反射率一阶导数(FD)作为参数, 建立土壤 Pb 含量的 BC-PLSR 预测模型。

不同浓度土壤 Pb 污染样品光谱的 E_m 和 E_a 可由以上分析计算得出, 分析不同浓度土壤 Pb 污染光谱反射率 FD 与 Pb 浓度的相关性, 可得相关性较高的四个波段为 band 2 134 nm, band 790 nm, band 1 276 nm 和 band 2 482 nm。箱形图可以用于剔除异常样品, 对剔除异常样品后的数据进行建模分析可以提高模型精度, 将不同浓度土壤 Pb 污染样品光谱的 E_m 、 E_a 、FD2134, FD790, FD1276, FD2482, 土壤含水率及有机质含量进行箱形图分析, 可以剔除异常样品(图 6), 从图 6 中可以看出, CK-2, CK-3, Pb(100)-2, Pb(300)-3, Pb(400)-3, Pb(800)-3 为异常数据, 予以剔除。聚类分析可以用于数据分类^[17], 这项技术的目标是将案例分为集群或组, 从而使同一集群成员之间的关联程度较强, 不同集群之间关联程度较弱。对剔除异常数据之后的样品进行系统聚类分析(图 7), 可以将数据分为两组: 第一组为 Pb(100)-3, ..., Pb(200)-3, 第二组为 Pb(200)-2, ..., Pb(600)-1。由于土壤重金属污染实验样品中土壤有机质含量、土壤含水率与土壤中重金属浓度之间关系密切, 分析两组数据的土壤有机质含量与含水率之间的关系可得, 第一组数据土壤有机质含量与含水率整体上呈负相关, 在含水率上升(下降)时有机质下降(上升)[图 8(a)], 第二组数据的有机质与含水率呈正相关, 在含水率上升(下降)时有机质上升(下降)[图 8(b)]。

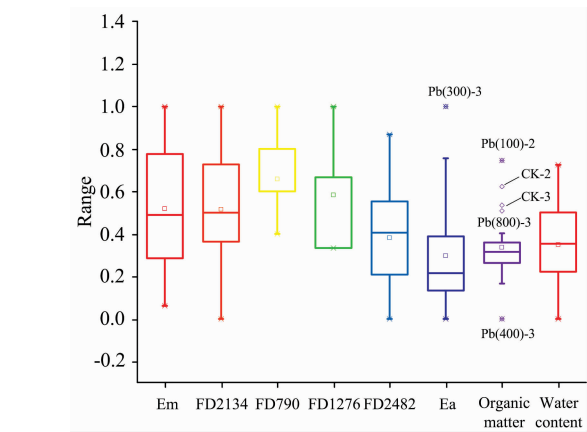
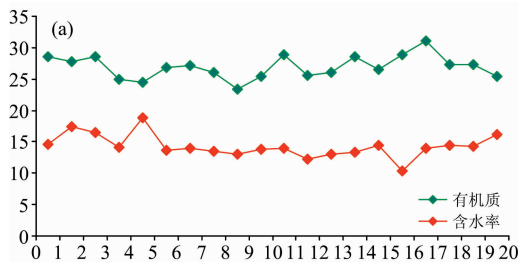


图 6 不同浓度 Pb 污染下土壤光谱 E_m 、 E_a 、一阶导数及有机质, 含水率的箱形图

Fig. 6 The boxplot of organic, water content and the value of E_m , E_a , first derivative spectrum of soil under the Pb pollution with different concentrations

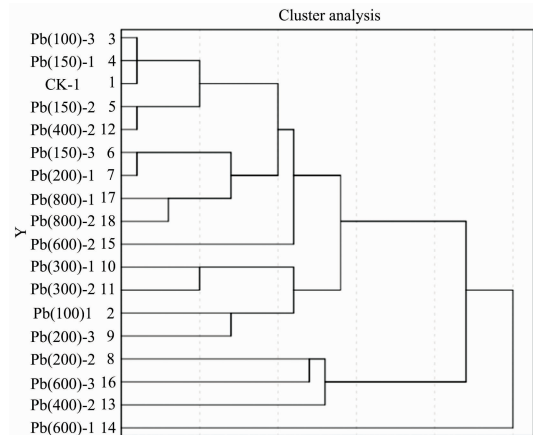


图 7 不同浓度 Pb 污染下土壤样品聚类分析

Fig. 7 The cluster analysis of soil under the Pb pollution with different concentrations

选择土壤样品光谱的 E_m 、 E_a 、FD2134, FD790, FD1276, FD2482, 土壤含水率及有机质含量作为参数, 分别对两组数据建立 BC-PLSR 模型预测土壤中 Pb 含量, 两组数据的自变量系数见图 9(a)和(b), 经验证可得预测值与实测值的相关性较强, r 分别为 0.88 和 0.99[图 10(a), (b)]。

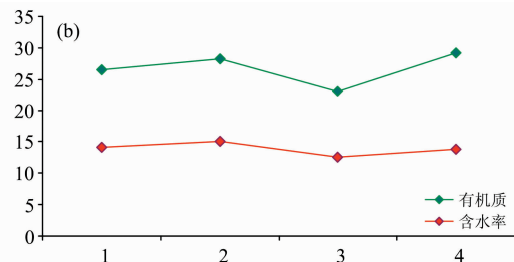


图 8 两组数据有机质与含水率之间的变化趋势

Fig. 8 Trends of organic matter and moisture content of two data sets

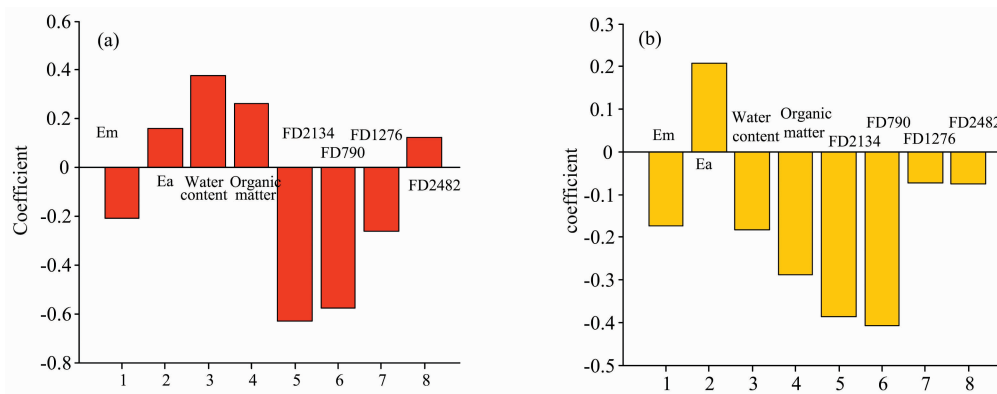


图 9 土壤 Pb 含量 BC-PLSR 预测模型变量系数

Fig. 9 The variable coefficients of the model (BC-PLSR) for predicting the Pb content in soil

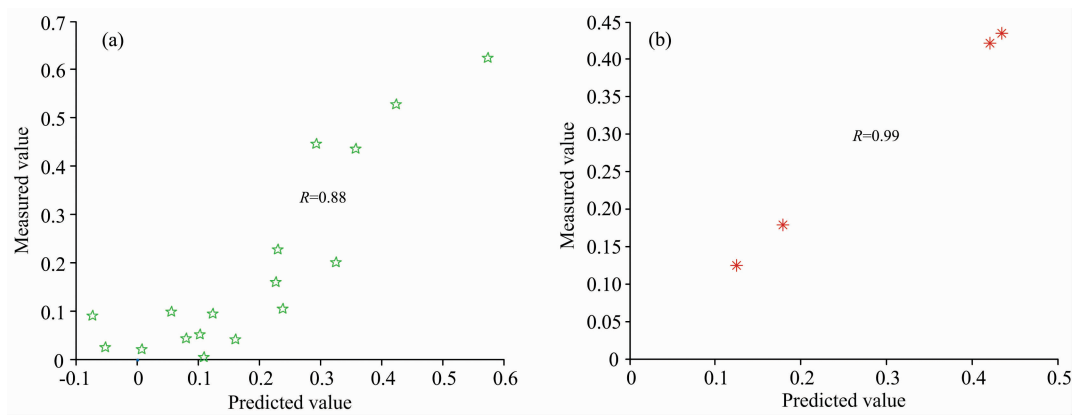


图 10 土壤 Pb 含量 BC-PLSR 预测模型精度

Fig. 10 The precision of the model (BC-PLSR) for predicting the Pb content in soil

3 结 论

采集了不同浓度 Pb 污染下土壤光谱曲线、土壤有机质含量及含水率,通过 HHT 时频分析图和 IMF2 瞬时频率二阶导数建立了土壤 Pb 污染光谱鉴别的 HHT 模型,并识别

了土壤 Pb 污染的特征波段区间为 2 150~2 300 nm;利用不同浓度 Pb 污染下土壤光谱 E_m 、 E_a 、一阶导数、有机质和含水率,通过箱形图、聚类分析和偏最小二乘法建立的 BC-PLSR 模型预测两类土壤样品中 Pb 含量,经验证模型精度较高,相关系数分别为 0.88 和 0.99。

References

[1] Bilgili A V, Van Es H M, Akbas F, et al. Journal of Arid Environments, 2010, 74(2): 229.
 [2] Liu Y, Li W, Wu G, et al. Geo-Spatial Information Science, 2011, 14(1): 10.
 [3] Stazi S R, Antonucci F, Pallottino F, et al. Communications in Soil Science and Plant Analysis, 2014, 45(22): 2911.
 [4] Al Maliki A, Bruce D, Owens G. Environmental Technology & Innovation, 2014, 1: 8.
 [5] Niazi N K, Singh B, Minasny B. International Journal of Environmental Science and Technology, 2015, 12(6): 1965.
 [6] Gholizadeh A, Borůvka L, Vašát R, et al. PloS one, 2015, 10(2), e0117457.
 [7] Gholizadeh A, Borůvka L, Saberioon M M, et al. Soil and Water Research, 2015, 10(4): 218.
 [8] Fard R S, Matinfar H R. Arabian Journal of Geosciences, 2016, 9(20): 745.
 [9] Feng Z, Liang M, Chu F. Mechanical Systems & Signal Processing, 2013, 38(1): 165.
 [10] Bafroui H H, Ohadi A. Neurocomputing, 2014, 133(8): 437.
 [11] Shen Z, Chen X, Zhang X, et al. Measurement, 2012, 45(1): 30.
 [12] Ali J B, Fnaiech N, Saidi L, et al. Applied Acoustics, 2015, 89(3): 16.
 [13] YU Xiao, DING En-je, CHEN Chun-xu, et al(俞 啸, 丁恩杰, 陈春旭, 等). Journal of China Coal Society(煤炭学报), 2015, 40(11): 2587.

- [14] GONG Mao-fa, XIA Wen-hua, ZHANG Xiao-ming, et al(公茂法, 夏文华, 张晓明, 等). Power System Protection and Control(电力系统保护与控制), 2013, 41(22): 64.
- [15] LI Wen-fan, LIU Zhi-gang, SUN Wan-lu(李文帆, 刘志刚, 孙婉璐). Power System Protection and Control(电力系统保护与控制), 2011, 39(23): 123.
- [16] GOU Xing-ming, ZHANG Wen-ying, YUAN Zhi-hui, et al(郭兴明, 张文英, 袁志会, 等). Chinese Journal of Scientific Instrument(仪器仪表学报), 2014, 35(4): 827.
- [17] Li X, Feng L. Atmospheric Environment, 2012, 47: 58.

HHT Identification and BC-PLSR Prediction Model of Soil Lead Pollution Spectrum

FU Ping-jie^{1,2}, YANG Ke-ming^{1,2*}, CHENG Long^{1,2}, WANG Min^{1,2,3}

1. State Key Laboratory Coal Resources and Safe Mining of China University of Mining & Technology (Beijing), Beijing 100083, China
2. College of Geoscience and Surveying Engineering, China University of Mining & Technology (Beijing), Beijing 100083, China
3. North China University of Science & Technology, Tangshan 063210, China

Abstract The problem of soil heavy metal pollution has always attracted attention. Therefore, many results have been achieved in this field by the research on the use of hyperspectral remote sensing, mainly focusing on predicting heavy metal content in soil using conventional methods such as derivative of soil spectra and continuous continuum removal. The soil spectral data showed tremendous similarity with non-linear non-stationary mechatronic signals, medical signals, etc. In this study, HHT was used to analyze the soil's lead (Pb) pollution experimental spectra in the frequency domain. The purpose of the HHT analysis was to achieve the HHT identification of soil's Pb pollution spectra, and establish the model for predicting Pb content in soil. Firstly, the soil Pb pollution experiment was conducted to collect the spectrum, water content, and organic matter content of soil Pb-contaminated samples. Secondly, the HHT time-frequency analysis and the second derivative of the instantaneous frequency of the second intrinsic mode function (IMF2) component of the Pb pollution spectra of soil samples were used to identify the characteristic bands of soil Pb contamination. Finally, the prediction model of soil Pb content that the parameters were appropriate frequency domain parameters, soil first-order derivative, soil organic matter content, and soil water content was established using boxplot, cluster analysis, and partial least squares. The results showed that HHT time-frequency analysis charts of soil Pb-contaminated could identify soil Pb contamination spectra. There was no abnormal signal in the band sequence between 250 and 430 from HHT time-frequency analysis plots of uncontaminated soil spectrum. There were many abnormal signals in the band between 250 and 430 from soil spectral HHT time-frequency analysis plots under Pb contamination, and with the increase of the concentration, the abnormal signal distribution range became wider and wider. When the pollution concentration reached $800 \mu\text{g} \cdot \text{g}^{-1}$, a strong abnormal signal was obtained in the band sequence of 270 and the frequency before 0.3 Hz. The mutation of second derivative of IMF2 instantaneous frequency of uncontaminated soil spectrum was very weak after the EMD, while there were obvious mutation points of Pb-contaminated soil spectrum. The characteristic wavelength band of soil Pb-contaminated soil spectrum was 2150~2300 nm according to the mutation points and second derivative of IMF2 instantaneous frequency of Pb-contaminated soil spectrum. Six groups of abnormal samples were removed from boxplot based on Hilbert energy spectrum peaks, EMD energy entropy, first derivative, organic matter and water content under different Pb concentrations. Then the Pb-contaminated soil samplings were divided into two categories by cluster analysis. Finally, Hilbert energy spectrum peak, EMD energy entropy, spectral first derivative of 2134, 790, 1276 and 2482 band, organic matter and water content were used as parameters. The BC-PLSR (Boxplot Cluster-Partial Least Squares Regression) models for the data of two categories were established to predict Pb content in soil. The accuracy of the validated model was high, and the correlation coefficients were 0.88 and 0.99, respectively.

Keywords Soil Pb pollution spectra; HHT analysis; Instantaneous frequency; BC-PLSR models

* Corresponding author

(Received Mar. 27, 2018; accepted Aug. 5, 2018)