

DOI: 10.13930/j.cnki.cjea.180131

李梦洁, 张曼胤, 崔丽娟, 王贺年, 郭子良, 李伟, 魏圆云, 杨思, 龙颂元. 基于连续小波变换和随机森林的芦苇叶片汞含量反演[J]. 中国生态农业学报, 2018, 26(11): 1730-1738

LI M J, ZHANG M Y, CUI L J, WANG H N, GUO Z L, LI W, WEI Y Y, YANG S, LONG S Y. Inversion of Hg content in reed leaf using continuous wavelet transformation and random forest[J]. Chinese Journal of Eco-Agriculture, 2018, 26(11): 1730-1738

## 基于连续小波变换和随机森林的芦苇叶片汞含量反演\*

李梦洁<sup>1,3</sup>, 张曼胤<sup>1,3\*\*</sup>, 崔丽娟<sup>1,2</sup>, 王贺年<sup>1,3</sup>, 郭子良<sup>1,3</sup>,  
李伟<sup>1,2</sup>, 魏圆云<sup>1,3</sup>, 杨思<sup>1,3</sup>, 龙颂元<sup>1,3</sup>

(1. 中国林业科学研究院湿地研究所/湿地生态功能与恢复北京市重点实验室 北京 100091; 2. 北京汉石桥湿地生态系统国家定位观测研究站 北京 101399; 3. 河北衡水湖湿地生态系统国家定位观测研究站 衡水 053000)

**摘要:** 植物重金属污染是当今世界面临的重大生态环境问题之一, 高光谱技术为快速、大面积监测植被重金属含量提供了可能性。本研究以重金属汞(Hg)和湿地植物芦苇为研究对象, 采用连续小波变换(CWT)和随机森林(RF)算法相结合的方法建立芦苇叶片总汞含量反演模型, 以期寻求一种较为精准的植物汞污染反演模型, 未来可通过高光谱技术建立模型来无损、快速估测湿地植物重金属汞污染情况, 为湿地生态系统的监测提供方法支持。结果表明: 芦苇叶片总汞含量敏感波段主要分布在可见光波段 419~522 nm、664~695 nm 和 724~876 nm 以及近红外波段 1 450~1 558 nm 和 1 972~2 500 nm; 经 CWT 变换后, 小波系数与叶片总汞含量的相关系数绝对值提高 0.04~0.18, 所构建的预测反演模型拟合效果  $R^2$  提高 0.107~0.177, 模型精度 RMSE 提高 0.008~0.013, 其中利用经小波变换的去包络线光谱(CR-CWT)数据建立的 RF 模型对芦苇叶片总汞含量的反演精度和拟合效果最优( $R^2=0.713$ , RMSE=0.127); 同时在土壤总汞含量约为 20 mg·kg<sup>-1</sup> 时, 采用 CR-CWT 数据构建 RF 模型的方法来反演芦苇叶片总汞含量更为准确和可靠( $R^2=0.825$ , RMSE=0.051)。因此, 利用 RF 算法进行植被重金属含量的反演具有一定的现实可行性, 而结合 CWT 后所构建的反演模型对指导植被重金属含量监测更具参考价值, 应用前景广阔。

**关键词:** 连续小波变换; 随机森林; 高光谱; 重金属汞; 芦苇叶片

**中图分类号:** X5 **文献标识码:** A **文章编号:** 1671-3990(2018)11-1730-09

## Inversion of Hg content in reed leaf using continuous wavelet transformation and random forest\*

LI Mengjie<sup>1,3</sup>, ZHANG Manyin<sup>1,3\*\*</sup>, CUI Lijuan<sup>1,2</sup>, WANG Henian<sup>1,3</sup>, GUO Ziliang<sup>1,3</sup>, LI Wei<sup>1,2</sup>,  
WEI Yuanyun<sup>1,3</sup>, YANG Si<sup>1,3</sup>, LONG Songyuan<sup>1,3</sup>

(1. Institute of Wetland Research, Chinese Academy of Forestry / Beijing Key Laboratory of Wetland Services and Restoration, Beijing 100091, China; 2. Hanshiqiao National Wetland Ecosystem Research Station, Beijing 101399, China; 3. Heibei Hengshuihu National Wetland Ecosystem Research Station, Hengshui 053000, China)

**Abstract:** Heavy metal pollution of plants is one of the most important eco-environmental problems in the world. Rapid and large-scale monitoring of heavy metal content in plants has always been an international problem and a key research topic. Due to its

\* 中央级公益性科研院所基本科研业务费专项(CAFINT2014K05)资助

\*\* 通信作者: 张曼胤, 主要研究方向为湿地生态学及湿地景观与规划设计。E-mail: cneco@126.com

李梦洁, 主要研究方向为湿地生态学。E-mail: 993288528@qq.com

收稿日期: 2018-01-31 接受日期: 2018-05-03

\* This study was supported by the Fundamental Research Funds of Central-level Nonprofit Research Institutes of China (CAFINT2014K05).

\*\* Corresponding author, E-mail: cneco@126.com

Received Jan. 31, 2018; accepted May 3, 2018

high resolution, multiple band and abundant data, hyperspectral technology could offer a rapid and accurate determination of heavy metal pollution in plants. It can be used to detect the absorption, reflection and transmission characteristics of spectral bands corresponding to phytochemical components and to quantitatively analyze weak spectral differences for large-scale determination of the growth and health of plants. However, researchers mostly construct sensitive spectral parameters (e.g., vegetation index) through simple spectral transformation techniques and continuous removal methods. Most of the inversion models are of univariate regression, multiple stepwise regression, principal component regression and other empirical or semi-empirical models. There have also been uses of artificial networks and support vector machine models. These models not only require more training sets, but also easily over-fit. Thus continuous wavelet transform (CWT) and Random Forest (RF) algorithms are used as more accurate models for inverting heavy metal pollution in plants. While CWT model can more clearly characterize spectral signals, RF has strong fitting ability and also has shorter iteration time. It has higher calculation efficiency for large datasets such as hyperspectral data and is superior in model construction. The heavy metal mercury (Hg) and the wetland plant reed (*Phragmites communis*) were used in this research to test the effectiveness off the CWT and RF models. CWT was used to decompose continuous wavelength at different scales in the original spectral reflectivity (R), first-order derivative reflectivity (FD) and de-envelope reflectivity (CR). Correlation analysis was used to determine sensitive bands of R, FD, CR, the spectral reflectance by continuous wavelet transform (R-CWT), the first derivative reflectivity by continuous wavelet transform (FD-CWT) and de-envelope reflectivity by continuous wavelet transform based on the correlation with leaf total Hg content. Then the sensitive bands and RF algorithm were used to establish the inversion model of reed leaf total Hg content. The results showed that sensitive bands of leaf total Hg content were mainly distributed in the visible regions of 419–522 nm, 664–695 nm and 724–876 nm, and the near-infrared regions of 1 450–1 558 nm and 1 972–2 500 nm. After CWT transformation, the absolute value of correlation coefficient between wavelet coefficient and leaf total Hg content increased by 0.04–0.18, the fitting effect ( $R^2$ ) of the prediction inversion model increased by 0.107–0.177 and the accuracy (RMSE) of the prediction inversion model increased by 0.008–0.013. The RF model which used continuum removal reflectance after wavelet transformation (CR-CWT) had optimal inversion precision and fitting effect ( $R^2 = 0.713$ , RMSE = 0.127). At the same time, it was more accurate and reliable to use RF model with CR-CWT to retrieve leaf total Hg content when soil total Hg content was about 20 mg·kg<sup>-1</sup> ( $R^2 = 0.825$ , RMSE = 0.051). Therefore, it was feasible to use RF algorithm to retrieve heavy metal content in plants. The inversion model constructed by CWT had a more reference value in terms of monitoring heavy metal content in plants. The model was widely used and provided methodological support for non-destructive and rapid monitoring of heavy metal pollution in ecosystems.

**Keywords:** Continuous wavelet transformation; Random forest; Hyperspectral data; Heavy metal mercury; Reed leaf

植物重金属污染是当今世界面临的重大生态环境问题之一, 具有范围广、对象多的特点。传统的植被重金属污染监测多以现场点采样调查的方法进行, 这种方法需在试验区设置足够密度的采样点, 采集足够的样本进行生物、化学分析以判断污染程度。这不仅费时费力、难以大范围应用, 而且不能及时发现程度较轻的重金属污染, 容易造成重金属累积。因此, 快速、大面积监测植被重金属含量一直是国际性难题和重点研究问题<sup>[1-2]</sup>。

20 世纪 80 年代兴起的高光谱技术为快速、精准地判断植被重金属污染提供了可能性。目前, 国内外已有许多学者致力于研究植被重金属含量的高光谱监测, 通过提取重金属污染下植被高光谱敏感参数, 建立模型来估测植被重金属的污染情况。提取敏感光谱参数的方法有多种。现阶段, 研究者们多通过简单的光谱变换技术、连续统去除法提取可有效指示植被重金属污染信息的“三边”、吸收谷等敏感光谱参数<sup>[3-4]</sup>, 或尝试多种可表征植物叶绿素、水分等生化组分含量的植被指数对特定重金属污染进行指示<sup>[5-7]</sup>; 此外还有学者利用经离散小波变换的小波分形维数、

小波奇异性等数据对镉、锌重金属胁迫下的水稻 (*Oryza sativa*) 叶片及冠层重金属含量进行了估算<sup>[8-9]</sup>。所构建的反演模型多为单变量回归模型、多元逐步回归模型、主成分回归模型、偏最小二乘模型等经验或半经验模型<sup>[10-11]</sup>, 以及人工神经网络及支持向量机模型<sup>[12-15]</sup>。这些模型不仅需要较多的训练集, 而且很容易过度拟合。本研究采用连续小波变换(continuous wavelet transform, CWT)和机器学习算法中的随机森林(random forest, RF)算法相结合的方法进行模型构建。其中, CWT 在表征光谱信号上比离散小波变换更加清晰<sup>[16]</sup>, 而 RF 算法不仅具有较强的拟合能力, 不会产生过度拟合, 还具有较快的训练速度, 对高光谱数据这类大数据集的计算效率较高, 在模型构建方面优于其他算法<sup>[17]</sup>。

本研究以重金属汞和湿地植物芦苇(*Phragmites communis*)为研究对象, 采用一阶导数变换法、包络线去除法以及连续小波变换对芦苇叶片原始光谱反射率(reflectance, R)进行变换, 得到一阶导数光谱(first derivative spectral reflectance, FD)、去包络线光谱(continuum removal reflectance, CR), 根据阈值提

取相关性强的 R、FD、CR 敏感波段和经小波变换的原始光谱反射率(R-CWT)、经小波变换的一阶导数光谱(FD-CWT)、经小波变换的去包络线光谱反射率(CR-CWT)小波系数,利用 RF 算法建立芦苇叶片汞含量反演模型,促进了高光谱技术在植被重金属监测方面的应用。

## 1 研究方法

### 1.1 试验设计

试验设置 8 个汞浓度梯度,分别为 0 mg·kg<sup>-1</sup> (CK)、10 mg·kg<sup>-1</sup>、20 mg·kg<sup>-1</sup>、40 mg·kg<sup>-1</sup>、60 mg·kg<sup>-1</sup>、80 mg·kg<sup>-1</sup>、100 mg·kg<sup>-1</sup>、160 mg·kg<sup>-1</sup>,处理代号为 Hg(CK)、Hg(10)、Hg(20)、Hg(40)、Hg(60)、Hg(80)、Hg(100)、Hg(160),每个浓度梯度 3 次重复。2017 年 4 月 3 日,按设计浓度将 HgCl<sub>2</sub> 溶液注入盛有营养土的玻璃缸中搅拌均匀,同时注水至土壤饱和并覆盖塑料薄膜防止汞蒸气挥发,静置 1 周后每个玻璃缸内种植 20 cm 高芦苇 20 株,保持 5 cm 水层进行植物培养。每培养 1 个月采集 1 次芦

苇叶片光谱信息,测定芦苇叶片总汞含量,共采集 7 次。同时,为保证土壤总汞含量与试验设计基本一致,在试验前后采集土壤并测定土壤总汞含量。

### 1.2 光谱采集及汞含量测定

采用美国 ASD (Analytical Spectral Devices) 公司 FieldSpec 4 便携式地物波谱仪(波谱范围为 350~2 500 nm)测量芦苇叶片高光谱数据,为减少外界因素干扰,选取健康、生长良好的叶片,使用叶片夹进行叶片光谱的测定,测定时保证叶片表面完全展开,避开叶片中部叶脉,至多每 20 min 进行 1 次白板优化。每个样本保存 40 条光谱,求取平均值作为该样本的光谱数据。

对测量过光谱数据的叶片,于实验室内洗净烘干,采用硝酸水浴消解-冷原子荧光光谱法<sup>[18]</sup>测定叶片总汞含量,采用分光光度计测定叶片叶绿素 a 含量。为保证建模样本的数量,将各汞处理浓度下的叶片样本进行整合后分组。选用 Kennard-Stone(K-S)算法计算出各叶片样本总汞含量之间的欧氏距离,筛选建模集样本 101 个,检验集样本 68 个(表 1)。

表 1 叶片总汞含量描述性统计特征

Table 1 Descriptive statistics of leaf total Hg content

样本类型 Type of sample	个数 Number	最小值 Min.(mg·kg <sup>-1</sup> )	最大值 Max.(mg·kg <sup>-1</sup> )	平均值 Mean (mg·kg <sup>-1</sup> )	标准差 Standard deviation (mg·kg <sup>-1</sup> )
所有样本 All samples	169	0.016	0.610 0	0.215	0.154
建模集 Modeling set	101	0.016	0.061 0	0.174	0.151
检验集 Test set	68	0.016	0.061 0	0.274	0.147

### 1.3 光谱变换方法

本文对原始光谱反射率(R)采用的光谱变换方法有一阶导数光谱(FD)、去包络线光谱(CR)以及连续小波变换(CWT)。其中,CWT 是一种线性变换方法,通过小波基函数将叶片高光谱数据分解为不同尺度上的一系列小波系数,其变换公式为<sup>[19]</sup>:

$$Wf(a,b) \leq f; \psi_{a,b} \geq \int_{-\infty}^{+\infty} f(t)\psi_{a,b}(t)dt \quad (1)$$

其中:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (2)$$

式中:  $Wf(a,b)$  为小波变换结果,  $f(t)$  为叶片高光谱反射率数据,  $t$  为光谱波段(350~2 500 nm),  $\psi_{a,b}(t)$  为小波基函数,  $a$  为尺度因子,  $b$  为平移因子。小波系数包含  $i$  和  $j$  两维,分别是分解尺度( $i = 1, 2, \dots, m$ )和波段( $j = 1, 2, \dots, n$ )组成的  $m \times n$  矩阵。由此,CWT 将一维叶片高光谱数据转换为二维小波系数,通过与叶片总汞含量进行相关性分析,即可得出最优分解尺度。

### 1.4 模型构建及精度检验方法

采用随机森林(RF)算法建立芦苇叶片总汞含量反

演模型。RF 是 Breiman<sup>[20]</sup>提出的一种基于分类回归树模型的集成学习算法,具有高效处理大样本数据、不必担心过拟合、可估计某个特征变量以及抗噪音等特点。RF 利用 bootstrap 重抽样方法从原始样本中抽取多个样本,对每个 bootstrap 样本进行决策树建模,最后组合多棵决策树的预测,通过投票得出最终预测结果<sup>[21]</sup>:

$$H(x) = \arg \max_Y \sum_{i=1}^k I [h_i(x) = Y] \quad (3)$$

式中:  $H(x)$  为组合分类模型,  $h_i$  为单个决策树分类模型,  $Y$  为输出变量,  $I$  为示性函数。

模型精度检验选取决定系数  $R^2$ 、均方根误差 RMSE 以及 1:1 线 3 个参数衡量。其中,1:1 线表示实测值和预测值所构成的点偏离  $y=x$  线的程度。

## 2 结果与分析

### 2.1 不同程度汞处理对芦苇叶片总汞和叶绿素 a 含量的影响

对不同程度汞污染下的土壤及芦苇叶片总汞含量进行分析(表 2、表 3)可知:在试验周期中土壤总汞含量没有发生显著变化;芦苇叶片总汞含量随着土壤总

汞含量的增加而升高, 与土壤总汞含量呈极显著正相关; 芦苇叶片叶绿素 a 含量随着土壤总汞含量的增加而降低, 与土壤总汞含量呈极显著负相关。表明不同土壤汞浓度处理引起芦苇叶片总汞含量及叶绿素 a 含量的变化。同时, 芦苇叶片总汞含量与叶绿素 a 含量呈极显著负相关。

表 2 不同程度 Hg 污染水平下土壤总汞含量及芦苇叶片总汞含量和叶绿素 a 含量

Table 2 Total Hg contents of soil and reed leaf and chlorophyll a contents in reed leaf under different levels of Hg pollution

汞浓度 Hg content (mg·kg <sup>-1</sup> )	土壤总汞含量 Soil total Hg content (mg·kg <sup>-1</sup> )	叶片总汞含量 Leaf total Hg content (mg·kg <sup>-1</sup> )	叶片叶绿素 a 含量 Leaf chlorophyll a content (mg·g <sup>-1</sup> )
0	2.380±3.203	0.154±0.136	1.444±0.419
10	9.707±2.599	0.179±0.103	1.439±0.467
20	19.125±7.777	0.180±0.126	1.417±0.404
40	38.496±8.864	0.214±0.144	1.415±0.438
60	57.293±13.655	0.224±0.202	1.374±0.387
80	76.949±18.336	0.231±0.152	1.319±0.602
100	96.997±10.388	0.255±0.160	1.313±0.463
160	157.556±21.678	0.283±0.187	1.292±0.421

表 3 土壤总汞、叶片总汞及叶绿素 a 含量的相关性分析  
Table 3 Correlation analysis of soil total Hg, leaf total Hg and chlorophyll a content

生化指标 Biochemical index	土壤总汞含量 Soil total Hg content	叶片总汞含量 Leaf total Hg content	叶片叶绿素 a 含量 Leaf chlorophyll a content
土壤总汞含量 Soil total Hg content	1.000	0.968**	-0.944**
叶片总汞含量 Leaf total Hg content	0.968**	1.000	-0.933**
叶片叶绿素 a 含量 Leaf chlorophyll a content	-0.944**	-0.933**	1.000

\*\* P<0.01

### 2.2 敏感光谱波段的筛选

将 R、FD 以及 CR 3 种形式的高光谱反射率与芦苇叶片总汞含量进行相关性分析以筛选汞敏感波段, 结果如图 1 所示。分析可知, R 相关曲线较为平缓, 在可见光波段 419~522 nm、664~695 nm 和 724~876 nm 与芦苇叶片总汞含量存在显著正相关 (P<0.01), 正相关系数最大可达 0.43; 在近红外波段 1 450~1 558 nm 和 1 972~2 500 nm 存在显著负相关 (P<0.01), 负相关系数最大可达-0.51。经一阶导数和去包络线处理后, 相关曲线波动较大, 敏感波段增加且分布零散, 光谱数据与叶片总汞含量的正负相关系数均有不同程度提高, 其中 FD 正相关系数最大可达 0.60, 负相关系数最大可达-0.68, CR 正相关系数最大可达 0.40, 负相关系数最大可达-0.67。为保证建模效果, 选取相关系数绝对值大于 0.5 的波段用于模型构建。

### 2.3 敏感小波系数筛选

选取 Mexican Hat 函数作为小波基函数<sup>[19,22]</sup>, 对叶片建模样本的 R、FD、CR 曲线进行 CWT 变换分解, 分解尺度为 2<sup>1</sup>、2<sup>2</sup>、2<sup>3</sup>、2<sup>4</sup>、2<sup>5</sup>、2<sup>6</sup>、2<sup>7</sup>、2<sup>8</sup>、2<sup>9</sup>、2<sup>10</sup>, 即 1~10 个尺度。将变换后的 10 尺度的小波系数与叶片总汞含量进行相关性分析, 得到相关系数图 (图 2), 图中红色区域表示相关性强的波段。经 CWT 处理后, 各尺度小波系数与叶片总汞含量的相关系数均有不同程度的提升, 其中 R-CWT 最大相关系数绝对值可达-0.69, DR-CWT 最大相关系数绝对值可达 0.72, CR-CWT 最大相关系数绝对值可达 0.74。

为更好地筛选敏感小波系数, 进一步绘制决定系数图, 如图 3 所示。从图中可知, R-CWT 与芦苇叶片总汞含量相关性较好的波段主要集中在 772~884 nm、

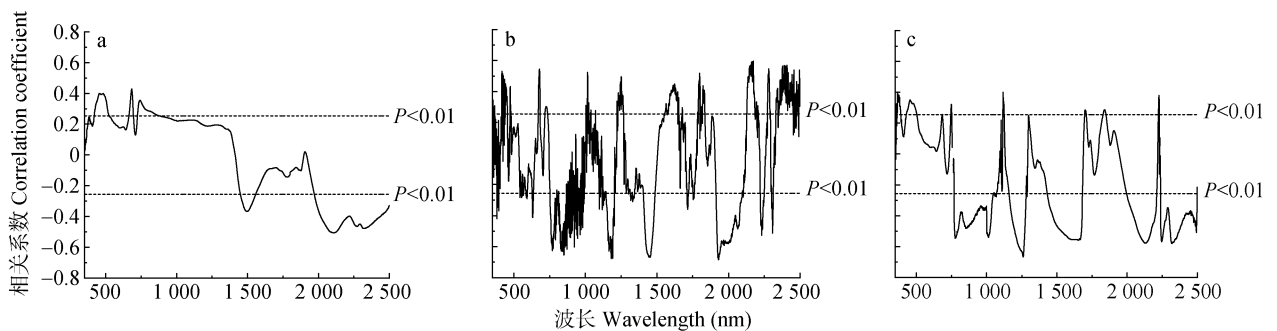


图 1 芦苇叶片总汞含量与原始光谱反射率(a)、一阶导数光谱(b)、去包络线光谱(c)的相关性分析

Fig. 1 Correlation analysis of reed leaf total Hg content with reflectance (a), first derivative spectral reflectance (b) and continuum removal spectral reflectance (c)

1 206~1 235 nm、1 478~1 618 nm、2 042~2 148 nm 以及 2 316~2 361 nm, 并在 821 nm、1 225~1 226 nm、1 508 nm、2 080 nm 以及 2 321 nm 达到最大值, 对应

的尺度分别为第 3、4、2、2 和 4 尺度; FD-CWT 与芦苇叶片总汞含量相关性较好的波段主要集中在 648~650 nm、782~820 nm、824~855 nm、1 377~1 378 nm

以及 2 299~2 499 nm, 并在 649 nm、811 nm、835~838 nm、1 378 nm 以及 2 303 nm 达到最大值, 对应的尺度为第 2、3、4、1 和 4 尺度; CR-CWT 与芦苇叶片总汞含量相关性较好的波段主要集中在 775~834 nm、841~874 nm、1 103~1 143 nm、1 188~1 238 nm、

1 273~1 386 nm 以及 2 286~2 365 nm, 并在 820~821 nm、850~851 nm、1 127 nm、1 225 nm、1 320~1 321 nm、1 363~1 364 nm 以及 2 291 nm 达到最大值, 对应的尺度为第 3、3、6、4、7、6 和 3 尺度。将筛选出的小波系数用于模型构建。

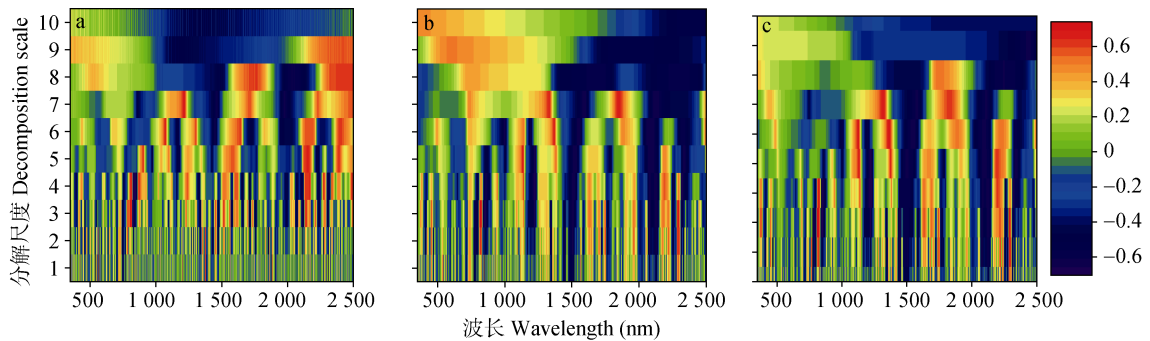


图 2 原始光谱小波系数(a)、一阶导数光谱小波系数(b)、去包络线光谱小波系数(c)与芦苇叶片总汞含量的相关系数图  
Fig. 2 Correlation scalogram between total Hg content and reflectance wavelet coefficient (a), first derivative spectral reflectance wavelet coefficient (b) and continuum removal wavelet coefficient (c) of reed leaf

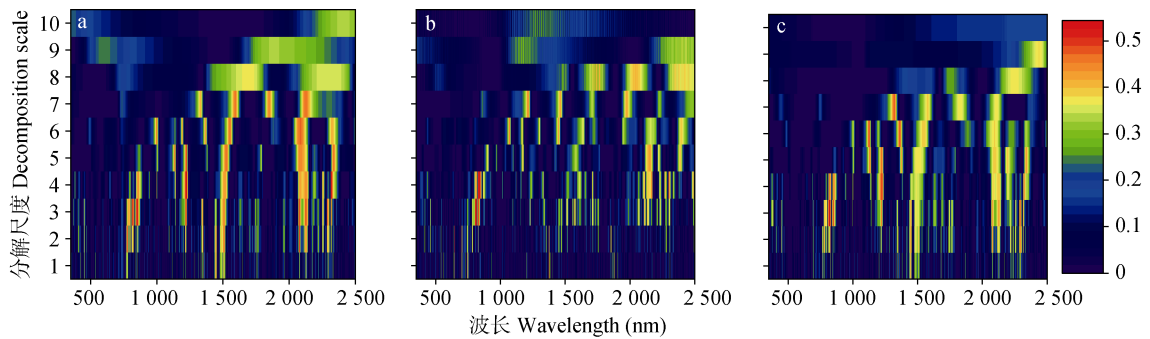


图 3 原始光谱小波系数(a)、一阶导数光谱小波系数(b)、去包络线光谱小波系数(c)小波系数与芦苇叶片总汞含量的决定系数图  
Fig. 3 Correlation of determination scalogram between total Hg content and reflectance wavelet coefficient (a), first derivative spectral reflectance wavelet coefficient (b) and continuum removal wavelet coefficient (c) of reed leaf

### 2.4 芦苇叶片总汞含量反演模型的优选

依据上述分析, 选取 R、FD、CR 与叶片总汞含量显著相关 ( $r>0.05$ ,  $P<0.01$ ) 的敏感波段以及 R-CWT、FD-CWT、CR-CWT 与叶片总汞含量显著相关的小波系数为自变量, 叶片总汞含量为因变量, 采用 RF 方法构建反演模型, 结果如表 4 所示。

表 4 芦苇叶片总汞反演模型的建模集和预测集结果  
Table 4 Calibration and validation results of estimation models for reed leaf total Hg content

模型 Model	建模集 Calibration set		检验集 Validation set	
	$R^2$	RMSE	$R^2$	RMSE
R	0.440	0.123	0.584	0.138
FD	0.585	0.100	0.566	0.129
CR	0.530	0.105	0.536	0.141
R-CWT	0.586	0.096	0.693	0.130
FD-CWT	0.611	0.098	0.673	0.119
CR-CWT	0.625	0.096	0.713	0.127

分析可知, 未经 CWT 变换的光谱数据, FD 模型的建模、预测拟合效果及模型精度均优于 R、CR 模型; 经 CWT 变换的光谱数据, CR-CWT 模型的预测精度明显优于 R-CWT 和 FD-CWT 模型。此外, 是否经过 CWT 变换光谱数据模型分别对比可知, R-CWT、FD-CWT、CR-CWT 模型的建模、预测拟合效果及精度均优于 R、FD、CR 模型, 其中预测集模型拟合效果  $R^2$  提高 0.107~0.177, 模型精度 RMSE 提高 0.008~0.013, 表明经过 CWT 变换后叶片总汞含量反演模型拟合效果及精度有所增加。

为更直观地观察模型对未知样本的适应性, 将 R、FD、CR、R-CWT、FD-CWT 以及 CR-CWT 模型的预测值与实测值进行 1:1 线分析(图 4)。由图可知, FD-CWT 以及 CR-CWT 模型的实测值和预测值样点分布更贴近 1:1 线附近, 其中 CR-CWT 模型 ( $R^2=0.713$ ,  $RMSE=0.127$ ) 拟合效果优于 FD-CWT 模

型, 能够更好地实现叶片总汞含量的反演。因此, CWT 作为一种光谱数据处理方法, 对叶片光谱数据进行变换处理并提取重构小波系数的敏感波段, 不仅模拟了不同尺度上的小波基函数与叶片光谱的相

似性, 而且能够察别到光谱范围内细微的变化情况, 更好地挖掘出微弱的隐藏信息, 与常规的相关系数分析相比反演精度上有所提高<sup>[16]</sup>, 在芦苇叶片总汞含量高光谱定量分析中具有重要意义。

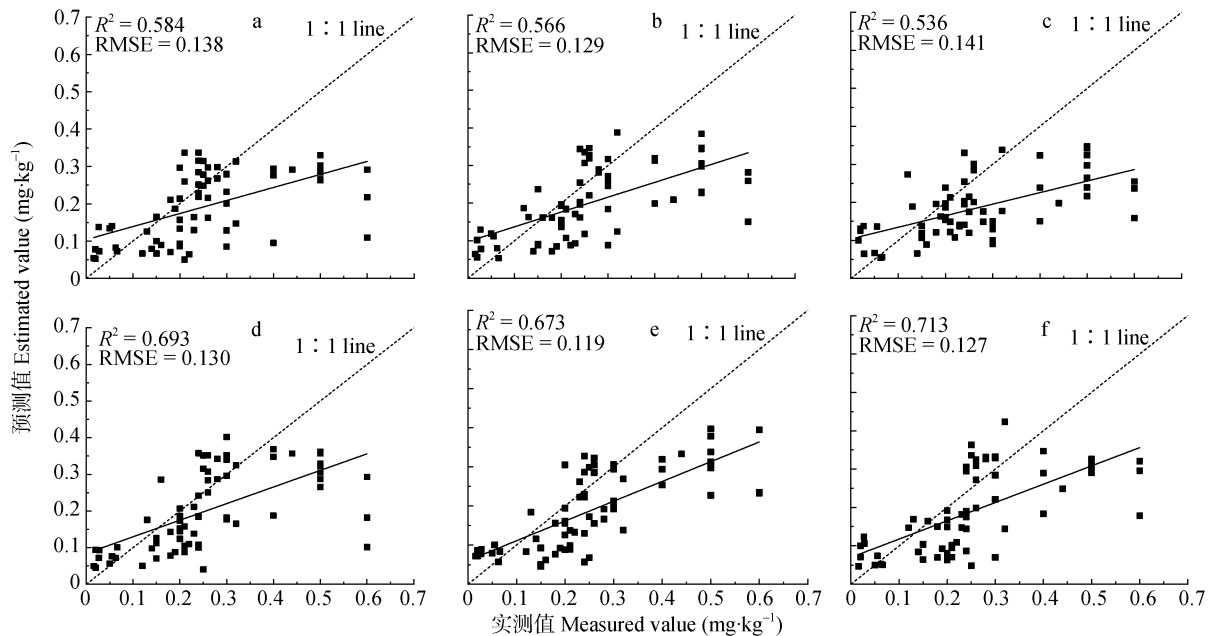


图 4 利用原始光谱(a)、一阶导数光谱(b)、去包络线光谱(c)、原始光谱小波系数(d)、一阶导数光谱小波系数(e)以及去包络线小波系数(f)反演芦苇叶片总汞含量的实测值与预测值比较

Fig. 4 Comparisons of the measured values of reed leaf total Hg content with the estimated values by using reflectance (a), first derivative spectral reflectance (b), continuum removal spectral reflectance (c), reflectance wavelet coefficient (d), first derivative spectral reflectance wavelet coefficient (e) and continuum removal wavelet coefficient (f)

### 3 讨论

在不同土壤汞处理水平下, 随着土壤总汞含量的增加, 芦苇叶片总汞含量逐渐增加, 但叶绿素 a 含量逐渐降低。这是因为重金属汞会抑制叶绿素酸脂还原酶, 并影响氨基- $\gamma$ -戊酮酶的合成<sup>[23]</sup>, 从而影响叶绿素的合成, 降低叶绿素含量<sup>[24-25]</sup>。从一定程度上说明, 不同土壤汞的处理水平对芦苇叶片内总汞含量产生了影响, 使得各处理水平下的芦苇叶片总汞含量存在差异, 可进一步利用高光谱技术进行反演监测。

对敏感波段和敏感小波系数的筛选以及模型的构建发现, 经过连续小波变换的光谱小波系数与芦苇叶片总汞含量的相关性更优, 所构建的模型预测精度也更优。这是因为小波分析可对光谱数据进行时域和频域上的多尺度分解, 通过搜索不同尺度上的最佳信号来预测植被生理生化成分<sup>[26-27]</sup>, 而连续小波变换更可以对光谱数据进行连续分解, 使得分解后的小波系数与原始光谱数据一一对应, 从而更有效提取光谱信息中的细微信号, 在表征光谱信号上比离散小波变换更加清晰<sup>[16,28-29]</sup>。目前, 国内外很

多学者已通过实验证明, 经过小波分析的光谱数据对植被叶绿素含量、水分含量、光合速率及病虫害探测等方面的反演均取得了较高的精度, 且优于传统光谱变换形式下的模型精度<sup>[19,29-33]</sup>, 这与本研究结果极为相近。

随机森林算法选择样本和自变量时具有随机性, 不仅会寻找自变量与因变量的关系, 还注重不同样本和自变量的差异, 使得回归结果既考虑每个样本和自变量的影响, 还不会过分趋近于某个样本<sup>[34]</sup>。李旭青等<sup>[17]</sup>发现利用随机森林算法反演水稻冠层氮含量模型精度可达 0.81; 程立真等<sup>[35]</sup>认为随机森林模型反演苹果(*Malus pumila*)叶片磷含量的精度高达 0.923; 高振东<sup>[36]</sup>曾利用随机森林算法反演水稻叶绿素含量, 并间接用叶绿素含量所处值域判断水稻的重金属污染情况。但直接利用随机森林算法反演植被重金属含量的研究还比较匮乏。本研究中除原始光谱建模集外, 通过随机森林算法建立的芦苇叶片总汞含量的反演模型拟合效果均在 0.5 以上, 表明利用随机森林算法进行植物叶片总汞含量的反演具有一定的可行性。

为了更好地说明所筛选模型的实用性,利用筛选出的 CR-CWT 模型分别对各土壤汞处理水平下的芦苇叶片总汞含量进行模型建立(表 5),检验模型精度。可以发现,利用 CR-CWT 数据所构建的各土壤浓度梯度下的 RF 模型,拟合效果  $R^2$  均在 0.5 以上,说明该模型在反演芦苇叶片总汞含量方面具有一定的稳定性。在所有土壤汞处理水平中, Hg(20)处理下所构建的模型拟合效果和精度最优( $R^2=0.825$ , RMSE=0.051),但各模型拟合效果与土壤总汞含量没有相关性。因此,随机森林算法适用于芦苇叶片重金属总汞含量的反演,且在土壤总汞含量约为 20 mg·kg<sup>-1</sup> 时,采用 CR-CWT 数据构建 RF 模型的方法来反演植物叶片总汞含量更为准确和可靠。

表 5 各土壤汞浓度梯度下去包络线光谱小波系数模型的反演精度

Table 5 Inversion accuracies of continuum removal wavelet coefficient models under various soil Hg contents

汞浓度 Hg content (mg·kg <sup>-1</sup> )	$R^2$	RMSE
0	0.631	0.091
10	0.560	0.079
20	0.825	0.051
40	0.668	0.056
60	0.745	0.090
80	0.616	0.098
100	0.747	0.078
120	0.506	0.118

#### 4 结论

利用随机森林算法进行植被重金属含量的反演具有一定的现实可行性,而结合常用光谱变换方法和连续小波变换后,所构建的反演模型更具有指导植被重金属含量监测的参考价值,应用前景广阔。本文以湿地植物芦苇为例,采用 CWT 对 R、FD、CR 进行不同尺度上连续波段的分解,利用与叶片总汞含量相关性高的 R、FD、CR 敏感波段和 R-CWT、FD-CWT、CR-CWT 小波系数,采用 RF 算法建立芦苇叶片总汞含量反演模型。研究结果表明:

1) 芦苇叶片总汞含量敏感波段主要分布在可见光波段 419~522 nm、664~695 nm 和 724~876 nm 以及近红外波段 1 450~1 558 nm 和 1 972~2 500 nm。

2) 经 CWT 变换后,小波系数与叶片总汞含量的相关系数绝对值提高 0.04~0.18,所构建的预测反演模型拟合效果  $R^2$  提高 0.107~0.177,模型精度 RMSE 提高 0.008~0.013。

3) 利用 CR-CWT 数据建立的 RF 模型对芦苇叶片总汞含量的反演精度和拟合效果最优( $R^2=0.713$ ,

RMSE=0.127)。

4) 在土壤总汞含量约为 20 mg·kg<sup>-1</sup> 时,采用 CR-CWT 数据构建 RF 模型的方法来反演植物叶片总汞含量更为准确和可靠( $R^2=0.825$ , RMSE=0.051)。

#### 参考文献 References

- [1] KEMPER T, SOMMER S. Mapping and monitoring of residual heavy metal contamination and acidification risk after the Aznalcollar mining accident (Andalusia, Spain) using field and airborne hyperspectral data[C]//Proceedings of the 3rd EARSeL Workshop on Imaging Spectroscopy. Herrsching, 2003: 333-343
- [2] REN H Y, ZHUANG D F, PAN J J, et al. Hyper-spectral remote sensing to monitor vegetation stress[J]. Journal of Soils and Sediments, 2008, 8(5): 323-326
- [3] SLONECKER T, HAACK B, PRICE S. Spectroscopic analysis of arsenic uptake in *Pteris* ferns[J]. Remote Sensing, 2009, 1(4): 644-675
- [4] 陈圣波, 周超, 王晋年. 黑龙江多金属矿区植物胁迫光谱及其与金属元素含量关系研究[J]. 光谱学与光谱分析, 2012, 32(5): 1310-1315  
CHEN S B, ZHOU C, WANG J N. Vegetation stress spectra and their relations with the contents of metal elements within the plant leaves in metal mines in Heilongjiang[J]. Spectroscopy and Spectral Analysis, 2012, 32(5): 1310-1315
- [5] DE OLIVEIRA M T G, ROLIM S B A, DE MELLO-FARIAS P C, et al. Industrial pollution of environmental compartments in the Sinos River Valley, RS, Brazil: Geochemical-biogeochemical characterization and remote sensing[J]. Water, Air, and Soil Pollution, 2008, 192(1/4): 183-198
- [6] 顾艳文, 李帅, 高伟, 等. 基于光谱参数对小白菜叶片镉含量的高光谱估算[J]. 生态学报, 2015, 35(13): 4445-4453  
GU Y W, LI S, GAO W, et al. Hyperspectral estimation of the cadmium content in leaves of *Brassica rapa chinensis* based on the spectral parameters[J]. Acta Ecologica Sinica, 2015, 35(13): 4445-4453
- [7] 史钢强, 杨可明, 孙阳阳, 等. 玉米叶片光谱红边位置的铜胁迫响应与污染监测[J]. 湖北农业科学, 2015, 54(13): 3234-3239  
SHI G Q, YANG K M, SUN Y Y, et al. Spectral red edge position responding and pollution monitoring of corn leaves stressed by heavy metal copper[J]. Hubei Agricultural Sciences, 2015, 54(13): 3234-3239
- [8] 刘美玲, 刘湘南, 李婷, 等. 水稻锌污染胁迫的光谱奇异性分析[J]. 农业工程学报, 2010, 26(3): 191-197  
LIU M L, LIU X N, LI T, et al. Analysis of hyperspectral singularity of rice under Zn pollution stress[J]. Transactions of the CSAE, 2010, 26(3): 191-197
- [9] LIU M L, LIU X N, DING W C, et al. Monitoring stress levels on rice with heavy metal pollution from hyperspectral reflectance data using wavelet-fractal analysis[J]. International Journal of Applied Earth Observation and Geoinformation, 2011, 13(2): 246-255
- [10] 邹登巍, 吴昀昭, 马宏瑞. 植物污染胁迫遥感监测研究综

- 述[J]. 遥感技术与应用, 2009, 24(2): 238–245  
WU D W, WU Y Z, MA H R. Review on remote sensing monitoring on contaminated plant[J]. Remote Sensing Technology and Application, 2009, 24(2): 238–245
- [11] 曹仕, 刘湘南, 刘清俊. 利用独立变量分析与高光谱植被指数模型监测成熟期水稻中砷污染[J]. 农业环境科学学报, 2010, 29(5): 881–886  
CAO S, LIU X N, LIU Q J. Monitor arsenic contamination in mature rice by the model based on the independent component analysis and hyperspectral vegetation indices[J]. Journal of Agro-Environment Science, 2010, 29(5): 881–886
- [12] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273–297
- [13] 李蜜, 刘湘南, 刘美玲. 基于模糊神经网络的水稻农田重金属污染水平高光谱预测模型[J]. 环境科学学报, 2010, 30(10): 2108–2115  
LI M, LIU X N, LIU M L. Fuzzy neural network model for predicting stress levels in rice fields polluted with heavy metals using hyperspectral data[J]. Acta Scientiae Circumstantiae, 2010, 30(10): 2108–2115
- [14] 张龙, 潘家荣, 朱诚. 基于近红外光谱的重金属汞、镉和铅污染水稻叶片鉴别[J]. 浙江大学学报: 农业与生命科学版, 2013, 39(1): 50–55  
ZHANG L, PAN J R, ZHU C. Discrimination of mercury, cadmium and lead polluted rice leaves based on near infrared spectroscopy technology[J]. Journal of Zhejiang University: Agriculture & Life Sciences, 2013, 39(1): 50–55
- [15] 许吉仁, 董爱红, 杨源, 等. 基于支持向量机的矿区复垦农田土壤-小麦镉含量高光谱估算[J]. 光子学报, 2014, 43(5): 102–109  
XU J R, DONG Q H, YANG Y H, et al. Support vector machine model for predicting the cadmium concentration of soil-wheat system in mine reclamation farmland using hyperspectral data[J]. Acta Photonica Sinica, 2014, 43(5): 102–109
- [16] 于雷, 洪永胜, 周勇, 等. 连续小波变换高光谱数据的土壤有机质含量反演模型构建[J]. 光谱学与光谱分析, 2016, 36(5): 1428–1433  
YU L, HONG Y S, ZHOU Y, et al. Inversion of soil organic matter content using hyperspectral data based on continuous wavelet transformation[J]. Spectroscopy and Spectral Analysis, 2016, 36(5): 1428–1433
- [17] 李旭青, 刘湘南, 刘美玲, 等. 水稻冠层氮素含量光谱反演的随机森林算法及区域应用[J]. 遥感学报, 2014, 18(4): 923–945  
LI X Q, LIU X N, LIU M L, et al. Random forest algorithm and regional applications of spectral inversion model for estimating canopy nitrogen concentration in rice[J]. Journal of Remote Sensing, 2014, 18(4): 923–945
- [18] 郑伟, 冯新斌, 李广辉, 等. 硝酸水浴消解-冷原子荧光光谱法测定植物中的总汞[J]. 矿物岩石地球化学通报, 2006, 25(3): 285–287  
ZHENG W, FENG X B, LI G H, et al. Determination of total mercury in plants by HNO<sub>3</sub> digestion in the water bath coupled with cold vapor atomic fluorescence spectrometry[J]. Bulletin of Mineralogy, Petrology and Geochemistry, 2006, 25(3): 285–287
- [19] CHENG T, RIVARD B, SÁNCHEZ-AZOFEIFA G A, et al. Continuous wavelet analysis for the detection of green attack damage due to mountain pine beetle infestation[J]. Remote Sensing of Environment, 2010, 114(4): 899–910
- [20] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5–32
- [21] LIAW A, WIENER M. Classification and regression by random forest[J]. R News, 2002, 23(2/3): 18–22
- [22] ZHANG J C, YUAN L, PU R L, et al. Comparison between wavelet spectral features and conventional spectral features in detecting yellow rust for winter wheat[J]. Computers and Electronics in Agriculture, 2014, 100: 79–87
- [23] STOBART A K, GRIFFITHS W T, AMEEN-BUKHARI I, et al. The effect of Cd<sup>2+</sup> on the biosynthesis of chlorophyll in leaves of barley[J]. Physiologia Plantarum, 1985, 63(3): 293–298
- [24] 郁达, 沈宗根, 张恒泽, 等. 汞对萝卜种子发芽及幼苗某些生理特性的影响[J]. 西北植物学报, 2004, 24(2): 231–236  
YU D, SHEN Z G, ZHANG H Z, et al. Effects on some physiological characters of seedling and germination of radish seeds after treated with Hg<sup>2+</sup>[J]. Acta Botanica Boreali-Occidentalia Sinica, 2004, 24(2): 231–236
- [25] GUPTA P, JAIN M, SARANGTHEM J, et al. Inhibition of 5-aminolevulinic acid dehydratase by mercury in excised greening maize leaf segments[J]. Plant Physiology and Biochemistry, 2013, 62: 63–69
- [26] 宋开山, 张柏, 王宗明, 等. 小波分析在大豆叶绿素含量高光谱反演中的应用[J]. 中国农学通报, 2006, 22(9): 101–108  
SONG K S, ZHANG B, WANG Z M, et al. Application of wavelet transformation in *in-situ* measured hyperspectral data for soybean LAI estimation[J]. Chinese Agricultural Science Bulletin, 2006, 22(9): 101–108
- [27] 宋开山, 张柏, 王宗明, 等. 基于小波分析的大豆叶绿素 a 含量高光谱反演模型[J]. 植物生态学报, 2008, 32(1): 152–160  
SONG K S, ZHANG B, WANG Z M, et al. Soybean chlorophyll a concentration estimation models based on wavelet-transformed, *in situ* collected, canopy hyperspectral data[J]. Journal of Plant Ecology, 2008, 32(1): 152–160
- [28] 梁栋, 杨勤英, 黄文江, 等. 基于小波变换与支持向量机回归的冬小麦叶面积指数估算[J]. 红外与激光工程, 2015, 44(1): 335–340  
LIANG D, YANG Q Y, HUANG W J, et al. Estimation of leaf area index based on wavelet transform and support vector machine regression in winter wheat[J]. Infrared and Laser Engineering, 2015, 44(1): 335–340
- [29] 方圣辉, 乐源, 梁琦. 基于连续小波分析的混合植被叶绿素反演[J]. 武汉大学学报: 信息科学版, 2015, 40(3): 296–302  
FANG S H, LE Y, LIANG Q. Retrieval of chlorophyll content using continuous wavelet analysis across a range of vegetation species[J]. Geomatics and Information Science of Wuhan University, 2015, 40(3): 296–302
- [30] CHENG T, RIVARD B, SÁNCHEZ-AZOFEIFA A.



- Spectroscopic determination of leaf water content using continuous wavelet analysis[J]. *Remote Sensing of Environment*, 2011, 115(2): 659–670
- [31] BLACKBURN G A. Wavelet decomposition of hyperspectral data: A novel approach to quantifying pigment concentrations in vegetation[J]. *International Journal of Remote Sensing*, 2007, 28(12): 2831–2855
- [32] BLACKBURN G A, FERWERDA J G. Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis[J]. *Remote Sensing of Environment*, 2008, 112(4): 1614–1632
- [33] 孙少波, 杜华强, 李平衡, 等. 基于小波变换的毛竹叶片净光合速率高光谱遥感反演[J]. *应用生态学报*, 2016, 27(1): 49–58
- SUN S B, DU H Q, LI P H, et al. Retrieval of leaf net photosynthetic rate of moso bamboo forests using hyperspectral remote sensing based on wavelet transform[J]. *Chinese Journal of Applied Ecology*, 2016, 27(1): 49–58
- [34] 王云飞, 庞勇, 舒清杰. 基于随机森林算法的橡胶林地上生物量遥感反演研究——以景洪市为例[J]. *西南林业大学学报*, 2013, 33(6): 38–45
- WANG Y F, PANG Y, SHU Q T. Counter-Estimation on aboveground biomass of *Hevea brasiliensis* plantation by remote sensing with random forest algorithm — A case study of Jinghong[J]. *Journal of Southwest Forestry University*, 2013, 33(6): 38–45
- [35] 程立真, 朱西存, 高璐璐, 等. 基于随机森林模型的苹果叶片磷素含量高光谱估测[J]. *果树学报*, 2016, 33(10): 1219–1229
- CHENG L Z, ZHU X C, GAO L L, et al. Hyperspectral estimation of phosphorus content for apple leaves based on the random forest model[J]. *Journal of Fruit Science*, 2016, 33(10): 1219–1229
- [36] 高振东. 基于水稻叶绿素含量变化的重金属污染胁迫遥感分析与评价[D]. 长春: 东北师范大学, 2015
- GAO Z D. Analysis and evaluation of heavy-metal pollution stress based on chlorophyll content of rice using hyperspectral data[D]. Changchun: Northeast Normal University, 2015

## 欢迎订阅 2019 年《大豆科学》

《大豆科学》是由黑龙江省农业科学院主管主办的大豆专业领域学术性期刊, 也是被国内外多家重要数据库和文摘收录源收录的重点核心期刊。主要刊登有关大豆遗传育种、品种资源、生理生态、耕作栽培、植物保护、营养肥料、生物技术、食品加工、药用功能及工业用途等方面的学术论文、研究报告、研究简报、国内外研究述评、学术活动简讯和新品种介绍等。

《大豆科学》主要面向从事大豆科学研究的科技工作者, 大专院校师生、各级农业技术推广部门的技术人员及科技种田的农民。

《大豆科学》为双月刊, 16 开本, 国内外公开发行, 国内每期定价: 20.00 元, 全年 120.00 元, 邮发代号: 14-95。国外每期定价: 10.00 美元(含邮资), 全年 60.00 美元, 国外代号: Q5587。全国各地邮局均可订阅, 也可向编辑部直接订购。

热忱欢迎广大科研及有关企事业单位刊登广告, 广告经营许可证号: 2301030000004。

地址: 哈尔滨市南岗区学府路 368 号《大豆科学》编辑部

(邮编: 150086)

电话: 0451-86668735

网址: www.haasep.cn

E-mail: ddkxbjb@126.com

欢迎关注微信公众号!

