

· 监测技术 ·

# 基于地理加权回归模型的葡萄园土壤砷含量高光谱反演

阿依努尔·麦提努日<sup>1</sup>, 麦提吐尔逊·艾则孜<sup>1,2\*</sup>, 李新国<sup>1</sup>

(1. 新疆师范大学地理科学与旅游学院, 新疆 乌鲁木齐 830054;

2. 新疆大学化学与化工学院, 新疆 乌鲁木齐 830046)

**摘要:**以吐鲁番盆地葡萄园土壤中砷(As)为研究对象,分析15种光谱变换下的土壤光谱反射率数据与土壤As含量的相关性,构建土壤As含量预测的偏最小二乘回归(PLSR)模型和地理加权回归(GWR)模型。结果表明:葡萄园土壤原始光谱率( $R$ )经一阶微分(FD)、平方根一阶微分(SRFD)、平方根二阶微分(SRSD)、倒数二阶微分(RTSD)、对数一阶微分(LTFD)、倒对数一阶微分(ATFD)变换对As光谱特征的增强作用最突出。模型预测结果表明,采用基于LTFD变换的GWR模型可有效提高葡萄园土壤As含量的预测精度。

**关键词:**砷;高光谱反演;地理加权回归模型;葡萄园土壤

中图分类号:X833;X87

文献标志码:B

文章编号:1006-2009(2021)04-0045-04

## Hyperspectral Inversion of Arsenic Content in Vineyard Soil Based on Geographically Weighted Regression Model

MATNURI Aynur<sup>1</sup>, EZIZ Mamattursun<sup>1,2\*</sup>, LI Xin-guo<sup>1</sup>

(1. College of Geographical Science and Tourism, Xinjiang Normal University, Urumqi,

Xinjiang 830054, China; 2. College of Chemistry and Chemical Engineering,

Xinjiang University, Urumqi, Xinjiang 830046, China)

**Abstract:** Taking arsenic(As) in vineyard soil in Turpan Basin as the research object, the correlation between soil spectral reflectance data and As content in soil was analyzed under 15 spectral transformations, partial least squares regression(PLSR) model and geography-weighted regression(GWR) model for As content prediction were established. The results showed that the original spectral rate of vineyard soil transformed by the first order differential(FD), square root first order differential(SRFD), square root second order differential(SRSD), reciprocal second order differential(RTSD), logarithmic first order differential(LTFD) and inverse logarithmic first order differential(ATFD) had a significant enhancement effect on the spectral characteristics of As. According to model prediction, GWR model based on LTFD transformation could effectively improve the prediction accuracy of As content in vineyard soil.

**Key words:** Arsenic; Hyperspectral inversion; Geographically weighted regression model; Vineyard soil

农用地土壤中As不仅威胁土壤环境质量,还通过土壤圈-水圈-生物圈间的循环和迁移,危害生态系统安全与生命健康<sup>[1-4]</sup>。我国土壤中As污染点位超标率为2.7%。As由于强烈的生物毒性,对整个生态系统安全影响较大,从而引起了重点关注<sup>[5]</sup>。

由于土壤中As元素含量低,导致传统方法测定效率低,且难以满足大尺度连续动态监测。高光

谱遥感技术通过分析土壤光谱特征构建模型,为获取大尺度和动态连续的重金属含量信息提供新方法<sup>[6-9]</sup>。地理加权回归(GWR)对于空间数据具有

收稿日期:2020-09-30;修订日期:2021-06-21

基金项目:国家自然科学基金资助项目(41867067)

作者简介:阿依努尔·麦提努日(1992—),女,维吾尔族,新疆和田人,硕士,研究方向为绿洲土壤环境安全。

\*通信作者:麦提吐尔逊·艾则孜 E-mail: oasiseco@126.com

较强的局部分析能力,对土壤重金属含量预测精度较高<sup>[10]</sup>。新疆吐鲁番盆地由于特殊的小气候效应,成为我国著名的葡萄主产区。吐鲁番盆地生产的葡萄是国家地理标志产品,葡萄种植业已成为当地的主导产业<sup>[11]</sup>。随着吐鲁番盆地经济的迅速发展,葡萄园土壤环境受到了不同程度的重金属污染威胁<sup>[12]</sup>。今以新疆吐鲁番盆地为研究区,构建基于GWR模型的葡萄园土壤As含量高光谱反演模型,探讨高光谱遥感技术在反演As含量中的适用性,以为葡萄园土壤重金属含量动态监测提供科学依据和技术支撑。

## 1 试验

### 1.1 样品采集与测定

2019年5月在吐鲁番盆地葡萄园进行土壤样品采集,土壤类型以棕漠土、灌耕土和盐土为主<sup>[13]</sup>。采用网格布点,共采集90个土壤样品,采样深度为0 cm~20 cm。所有土壤样品自然风干,过20目尼龙筛后分成两份,一份用来测定As,另一份用来测定土壤光谱反射率。依托新疆维吾尔自治区分析测试研究院,参照文献[14],用BAF-2000型原子荧光光度计测定土壤样品中As含量。测定过程中,加入国家土壤标准参比物质进行质量控制,回收率为93.66%,在允许范围内。

### 1.2 光谱反射率的测定与预处理

使用美国ASD公司的FieldSpec 3型便携式地物光谱仪,选择晴朗无风的天气,在野外测定葡萄园土壤光谱数据,光谱波长范围为350 nm~2 500 nm。为了降低光线散射,在2 m×2 m的黑色卡纸上放40 cm×40 cm的白板进行定标,获取绝对反射率。测定时始终保证光谱仪传感器探头垂直于土壤表面。每个土样光谱数据测定10次,传感器探头每10 min对准白板进行1次优化。利用ViewSpecPro软件均值化处理10条光谱曲线,将其算术平均值作为土壤原始光谱反射率。对均值处理后的所有光谱曲线统一去除350 nm~400 nm、1 350 nm~1 430 nm、1 800 nm~1 950 nm和2 400 nm~2 500 nm范围内的噪声区波段后,总共输出1 772个波段。

### 1.3 土壤光谱数据处理

采用Savitzky-Golay法对葡萄园土壤原始光谱做平滑处理,并剔除原始土壤光谱特征曲线的异常值之后,对土壤光谱反射率( $R$ )数据分别做求一

阶微分(FD)、二阶微分(SD)、平方根(SR)、平方根一阶微分(SRFD)、平方根二阶微分(SRSD)、倒数变换(RT)、倒数的一阶微分(RTFD)、倒数的二阶微分(RTSD)、对数(LT)、对数的一阶微分(LTFD)、对数的二阶微分(LTSD)、倒对数(AT)、倒对数一阶微分(ATFD)与倒对数的二阶微分(ATSD)等14种光谱变换,增强有效波谱信息,选择最佳的光谱处理方法<sup>[15]</sup>。

### 1.4 反演模型建立

建立葡萄园土壤As含量反演的偏最小二乘回归(PLSR)模型。为消除光谱反射率之间的多重共线性,设定原始光谱数据为自变量 $X$ ,As实测值为因变量 $Y$ ,从 $X$ 中提取1个主成分 $t_1$ ,从 $Y$ 中提取1个主成分 $u_1$ ,在满足 $u_1$ 与 $t_1$ 存在最大相关性的基础上建立因变量 $Y$ 和 $t_1$ 的回归方程。考虑到地下水中As含量空间异质性,建立葡萄园土壤As含量反演的GWR模型,并与PLSR模型模拟结果进行对比分析。GWR模型是作为普通二乘回归模型的空间扩展,对不同空间子区域自变量和因变量之间关系随空间变化进行建模的非参数局部空间回归分析方法,在重金属含量预测中效果良好<sup>[10]</sup>。GWR模型的一般形式如下:

$$Y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (i = 1, 2, \dots, n) \quad (1)$$

式中: $(u_i, v_i)$ 是第 $i$ 个样点的空间坐标; $n$ 为验证样点个数; $Y_i$ 和 $x_{ik}$ 分别为因变量 $Y$ 和自变量集 $x_k$ 在空间位置 $(u_i, v_i)$ 处的实测值; $\beta_0(u_i, v_i)$ 为在空间位置 $(u_i, v_i)$ 处的常数项;系数 $\beta_k(u_i, v_i)$ 是连续函数 $\beta_k(u, v)$ 在 $i$ 点的值; $\varepsilon_i$ 为符合 $N(0, \sigma^2)$ 分布的误差项<sup>[10]</sup>。

GWR模型中的 $\beta_k(u_i, v_i)$ 参数估算采用加权最小二乘法实现。采用高斯函数来估算空间权重矩阵,并利用GWR 4.0软件进行GWR模型模拟。采用决定系数( $R^2$ )和均方根误差(RMSE)衡量模型的稳定性和精度。评价参数中, $R^2$ 越接近1,表明模型拟合程度越高;RMSE值越小,表明模型预测能力越高。

## 2 结果与讨论

### 2.1 葡萄园土壤As含量特征

为保证建模样本与验证样本As含量范围一致,将采集的葡萄园土壤样本按As实测值排序,每

隔2个土样选取1个作为验证样本,其中60个土样用于模型建立(校准集),剩下的30个土样用于检验模型预测能力(验证集)。利用SPSS 20.0软件处理90个葡萄园土壤样品中As含量相关数据,结果见表1。由表1可知,吐鲁番盆地葡萄园土壤As的实测值为6.86 mg/kg ~ 13.4 mg/kg,变化幅度较小,平均值为9.70 mg/kg,为新疆灌耕土背景值(9.09 mg/kg)<sup>[16]</sup>的1.07倍,表明As在研究区葡萄园土壤中有一定的积累。校准集和验证集的As平均值、标准差和CV值基本一致,说明校准集所建立的预测模型能较好地适用于建模样本。As测定结果的RSD为17.0%,呈现弱空间变异,变异不明显,表明吐鲁番盆地葡萄园土壤中As含量空间分布较均匀,空间异质性不显著,受外部环境的干扰较小。

表1 葡萄园土壤中As测定结果  
Table 1 Concentrations of As in vineyard soil

样本	实测值 $w/$ ( $\text{mg} \cdot \text{kg}^{-1}$ )	平均值 $w/$ ( $\text{mg} \cdot \text{kg}^{-1}$ )	标准差 $w/$ ( $\text{mg} \cdot \text{kg}^{-1}$ )	RSD/%
校准集	6.86 ~ 12.9	9.67	1.73	17.9
验证集	7.82 ~ 13.4	9.77	1.51	15.5
总样本	6.86 ~ 13.4	9.70	1.65	17.0

## 2.2 葡萄园土壤As含量与光谱反射率相关性

为探讨不同光谱变换形式与As含量之间的关系,对葡萄园土壤As实测值与包括原始光谱反射率在外的15种光谱变换数据进行Pearson相关分析,并通过0.01和0.05显著性检验,在可见光、近红外和短红外波段范围内筛选出显著相关的特征波段。结果表明,葡萄园土壤As实测值与15种变换形式的光谱反射率之间的相关系数均达到0.01的极显著性水平。实测值与土壤光谱反射率通过FD变换的555 nm、2 197 nm与1 125 nm处,SRFD变换的529 nm、2 179 nm与1 125 nm处,SRSD变换的533 nm、1 712 nm、2 376 nm与931 nm处,RTSD变换的528 nm、477 nm、2 376 nm与447 nm处,LTFD变换的449 nm、529 nm、2 197 nm与1 125 nm处,以及ATFD变换的449 nm、2 197 nm与1 125 nm处的光谱值之间相关性较强,其他变换没有或者极少存在与As含量显著相关的特征波段。各种光谱变换对As光谱特征的增强作用不尽相同,其中FD、SRFD、SRSD、RTSD、LTFD和ATFD变换对As光谱特征的增强作用最突出。最大正相

关出现在SRFD变换的529 nm处,相关系数为0.569,位于土壤氧化铁的微弱吸收峰。最大负相关出现在FD变换的2 197 nm处,相关系数为-0.531,位于黏土矿物中金属—OH振动合频产生的吸收峰<sup>[17-18]</sup>。由此可见,土壤中As与铁氧化物、黏土矿物等具有一定的赋存共生关系。经光谱变换后的光谱值与葡萄园土壤中As实测值之间相关系数最大值高于原始反射率与实测值之间的相关系数最大值。

## 2.3 葡萄园土壤As含量PLSR反演模型建立

不同高光谱变换下的葡萄园土壤中As含量PLSR模型精度检验结果见表2。

表2 不同光谱变换下As含量PLSR、GWR模型精度检验  
Table 2 PLSR model of As concentration under different spectral transformation

光谱变换	PLSR模型		GWR模型(校准集)		GWR模型(验证集)	
	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE
FD	0.510	1.201	0.928	0.462	0.704	0.832
SRFD	0.554	1.147	0.958	0.350	0.875	0.552
SRSD	0.402	1.330	0.890	0.605	0.787	0.756
RTSD	0.406	1.325	0.456	1.491	0.736	0.913
LTFD	0.553	1.150	0.954	0.373	0.935	0.402
ATFD	0.548	1.156	0.950	0.383	0.800	0.670

对模型参数进行统计,以30个验证集的As实测值为横坐标,PLSR模型预测值为纵坐标,选取建模效果最好的光谱变换方式,对As含量PLSR模型进行检验。从检验结果来看,研究区葡萄园土壤中As含量实测值与采用FD、SRFD、SRSD、RTSD、LTFD与ATFD变换后的PLSR模型预测值之间的 $R^2$ 为0.131 ~ 0.341, RMSE为1.303 ~ 1.660。从模型稳定性和精确性来看,葡萄园土壤As含量高光谱反演的PLSR模型拟合程度低、预测精度不高,表现不佳。

## 2.4 葡萄园土壤As含量GWR反演模型建立

以60个校准集的As实测值为因变量,各光谱特征值为自变量,利用GWR模型进行回归分析,结果见表2。由表2可知,不同光谱变换形式对As预测精度的提高作用不同,GWR模型的 $R^2$ 整体上比PLSR模型大,其RMSE整体上比PLSR模型小。这表明GWR模型对研究区葡萄园土壤As含量预测效果明显提高。

GWR模型的 $R^2$ 比PLSR模型分别增加41.8%、40.4%、48.8%、5.0%、40.1%、40.2%,除

RTSD的RMSE增加了0.166以外,其余模型的RMSE分别减了0.739、0.797、0.725、0.777、0.773。由此可见,虽然GWR模型对FD、SRFD、SRSD、LTFD与ATFD变换后的模型预测精度改善效果明显,但对RTSD变换后的模型预测精度改善效果不明显。

为验证GWR模型精度并选取最佳预测模型,以30个验证集的As实测值为横坐标,GWR模型预测值为纵坐标,绘制实测值与预测值的对比散点图。结果表明,在FD、SRFD、SRSD、RTSD、LTFD和ATFD变换形式中,GWR模型预测结果的 $R^2$ 分别为0.704、0.875、0.787、0.736、0.935和0.800,分别为PLSR模型的4.12、6.68、2.58、2.16、5.19和4.44倍。RMSE分别为0.832、0.552、0.756、0.913、0.402和0.670,比PLSR模型分别减少了0.706、1.108、0.547、0.412、1.148和0.883,表明As元素含量的GWR预测值与实测值回归方程的 $R^2$ 有了很大的提高,而RMSE则有明显的降低。GWR模型的验证效果进一步证实了GWR模型预测程度较PLSR模型有很大提高。整体上,LTFD变换后采用GWR模型预测研究区葡萄园土壤As含量的预测精度最为理想,模型估算精度比其他变换形式高,比PLSR模型也有明显提高。采用GWR模型能够综合反映各种光谱变换形式的特点,在一定程度上可以增强变量信息,降低模型误差,提高模型精度和稳定性。

### 3 结语

土壤原始光谱率通过FD、SRFD、SRSD、RTSD、LTFD和ATFD后的光谱值与葡萄园土壤As实测值之间的相关性较高且特征波段数量较多。从模型稳定性和精确性来看,在FD、SRFD、SRSD、RTSD、LTFD和ATFD变换下,GWR模型校准集的 $R^2$ 比PLSR模型提高了5.0%~48.8%,相应的RMSE减少了0.166~0.797。在FD、SRFD、SRSD、RTSD、LTFD和ATFD变换下,GWR模型检验集的 $R^2$ 分别为PLSR模型的4.12倍、6.68倍、2.58倍、2.16倍、5.19倍和4.44倍,其RMSE比PLSR模型减少了0.412~1.148。基于LTFD变换的GWR准确度高,为最佳预测模型,能为有效提高葡萄园土壤As含量高光谱反演精度提供科学依据。

#### [参考文献]

[1] 孙超,陈振楼,毕春娟,等.上海市崇明岛农田土壤重金属的

环境质量评价[J].地理学报,2009,64(5):619-628.

- [2] GUO J, YUE T L, LI X T, et al. Heavy metal levels in kiwifruit orchard soils and trees and its potential health risk assessment in Shaanxi, China [J]. Environmental Science and Pollution Research, 2016, 23(14): 14560-14566.
- [3] 吴钟解,陈石泉,蔡泽富,等.海南八门湾沿岸表层沉积物重金属分布特征及污染评价[J].环境监测管理与技术,2019,31(6):28-32.
- [4] 敖明,柴冠群,范成五,等.稻田土壤和稻米中重金属潜在污染风险评估与来源解析[J].农业工程学报,2019,35(6):198-205.
- [5] 袁自然,魏立飞,张杨熙,等.优化CARS结合PSO-SVM算法农田土壤重金属砷含量高光谱反演分析[J].光谱学与光谱分析,2020,40(2):567-573.
- [6] TAN K, YE Y Y, DU P J, et al. Estimation of heavy metal concentrations in reclaimed mining soils using reflectance spectroscopy [J]. Spectroscopy and Spectral Analysis, 2014, 34(12): 3317-3322.
- [7] 叶明亮,杨梦丽,刘纯宇,等.高光谱遥感在土壤重金属污染监测中的应用[J].环境监测管理与技术,2018,30(6):1-5.
- [8] 亚森江·喀哈尔,茹克亚·萨吾提,尼加提·卡斯木,等.优化光谱指数的露天煤矿区土壤重金属含量估算[J].光谱学与光谱分析,2019,39(8):2486-2494.
- [9] SHI T Z, CHEN Y Y, LIU Y L, et al. Visible and near-infrared reflectance spectroscopy—An alternative for monitoring soil contamination by heavy metals [J]. Journal of Hazardous Materials, 2014, 265: 166-176.
- [10] 江振蓝,杨玉盛,沙晋明. GWR模型在土壤重金属高光谱预测中的应用[J].地理学报,2017,72(3):533-544.
- [11] 陈晓丽,陈彤. 特色林果产品消费者行为分析——以新疆为例[J].农村经济,2017(3):55-60.
- [12] 麦尔哈巴·图尔贡,麦提吐尔逊·艾则孜,王维维.吐鲁番盆地葡萄园土壤重金属污染及其潜在健康风险[J].环境与职业医学,2020,37(6):558-565.
- [13] XIANG M, JIANG H L, HUA T Z, et al. Zircon U-Pb shrimp ages from the late Paleozoic Turpan-Hami basin, NW China [J]. Journal of Earth Science, 2014, 25(5): 924-931.
- [14] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会. GB/T 22105.2—2008 土壤质量 总汞、总砷、总铅的测定 第2部分:土壤中总砷的测定[S].北京:中国标准出版社,2008.
- [15] 王学顺,戚大伟,黄安民.基于小波变换的木材近红外光谱去噪研究[J].光谱学与光谱分析,2009,29(8):2059-2062.
- [16] 麦提吐尔逊·艾则孜,阿吉古丽·马木提,艾尼瓦尔·买买提,等.博斯腾湖流域绿洲农田土壤重金属污染及潜在生态风险评价[J].地理学报,2017,72(9):1680-1694.
- [17] 程先锋,宋婷婷,陈玉,等.滇西兰坪铅锌矿区土壤重金属含量的高光谱反演分析[J].岩石矿物学杂志,2017,36(1):60-69.
- [18] 宋婷婷,付秀丽,陈玉,等.云南个旧矿区土壤锌污染遥感反演研究[J].遥感技术与应用,2018,33(1):88-95.