

Prediction of Soil Heavy Metal Concentration—Using a Random Forest Machine Learning Algorithm

Yanyan Liu^{1,2*}, Hanhu Liu¹, Zhiyong Hu², Johan Libens²

¹Chengdu University of Technology, Chengdu Sichuan

²University of West Florida, Pensacola Florida

Email: *lly950929@163.com

Received: May 17th, 2020; accepted: Jun. 8th, 2020; published: Jun. 15th, 2020

Abstract

The potential hazard of heavy metals in soil has been attracted more and more attention. Prediction of a continuous heavy metal concentration surface helps identify hotspot problem areas for environmental management and remediation. In recent years, forest-based algorithm has become popular in solving many kinds of problems in different fields, including geoscience. Random Forest Regression is one of them works well in prediction. The paper represents an application of Random Forest Regression algorithm in predict concentration surface using soil sample data collected the research project by Liebens *et al.* (2012) and other covariates data. Such as “sewer waste points”, “dry cleaner”, “Traffic data (Named AADT) and EPA TRI (toxic release inventory) in this study relate with the main way about how soil heavy metal spread. It also includes Florida DEP “major/minor emitters” and “DEP emitters” data. In this paper, we use random forest regression to predict continuous heavy metal concentration focus on Pb and Zn. The point's data were divided into different kinds to calculate the density surfaces. Joined these density data with soil properties data and sampling data to create two datasets: prediction points and training points. Both are in the study area and have the same set of explanatory variables. Then run a Random Forest Regression model to get the prediction results. Finally, assess prediction accuracy with RMSE. The results show that RMSE value decreases when the number of decision trees increases in the random forest regression model.

Keywords

GIS Application, Machine Learning, Soil Heavy Metals, Environmental Science, Soil Science

基于随机森林的土壤重金属浓度预测

刘艳艳^{1,2*}, 刘汉湖¹, 胡志勇², Johan Libens²

*通讯作者。

¹成都理工大学, 四川 成都²西佛罗里达大学, 佛罗里达 彭萨科拉

Email: lyy950929@163.com

收稿日期: 2020年5月17日; 录用日期: 2020年6月8日; 发布日期: 2020年6月15日

摘要

土壤中重金属的潜在危害越来越受到人们的重视。预测连续的重金属浓度面有助于确定环境管理和修复的热点问题区域。近年来, 基于森林的算法在解决包括地球科学在内的各个领域的各种问题中得到了广泛的应用。随机森林回归是一种有效的预测方法。本文利用土壤样本数据和其他协变量数据, 将随机森林回归算法应用于预测土壤重金属浓度栅格图。本研究中的“污水排放点”、“干洗店”、“交通数据”(命名为AADT)和EPA TRI (有毒物质释放清单)与土壤重金属如何传播的主要途径有关。它还包括佛罗里达州“主要/次要排放”和“副排放”的数据。本文采用随机森林回归方法预测了铅、锌的连续重金属浓度。将点数据分成不同的种类进行密度面计算。将这些密度数据与土壤性质数据和采样数据结合起来, 创建两个数据集: 预测点和训练点。两者都在研究范围内, 具有相同的解释变量集。然后利用随机森林回归模型得到预测结果。最后, 使用RMSE评估预测精度。结果表明, 随机森林回归模型中决策树数量的增加, RMSE值降低。

关键词

GIS应用, 机器学习, 土壤重金属, 环境科学, 土壤科学

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>

Open Access

1. 引言

土壤中重金属残余过量会造成一些自然灾害并危害人类健康。例如, 过量的锌(Zn)会降低土壤酶活性, 也会减少土壤中的细菌数量, 进而会降低农业产量[1]。小麦最容易受到土壤中过量锌的不利影响。铅(Pb)是另一种毒性最大的重金属, 它会影响我们的脑细胞, 尤其是儿童的脑细胞[2]。

国内外的土壤学家、环境保护者、统计学家和环境科学专家对土壤重金属的空间分析进行了大量的研究。目前, 进行土壤重金属的空间分析的方法分为三类: 地质统计插补、非地质统计插补和组合插补, 主要依赖于利用地理样本进行空间插值[3]。虽然已有众多空间分析方法应用于土壤重金属空间特征分析, 且取得了不错的效果, 但这些研究在以下几个方面还有一定的局限性。

1) 土壤重金属空间分析的研究思路基本相似, 大多使用多种空间插值技术中的一种来生成简单连续的表面图, 从而描述土壤重金属浓度的空间分布并分析评估土壤重金属的潜在来源(以地统计学插值法为主), 空间分析方法过于单一。

2) 大多研究在插值时只使用了样本浓度值和位置, 并未包含导致重金属空间浓度变化的协变量, 研究要素的完整性有待提高。

3) 大多研究集中在农田、耕地[4] [5], 矿区等工业区[6] [7] [8] [9]。很少有研究兼顾所有功能区, 将完整城市作为研究区, 进行整体研究。

4) 大多数学者对城市重金属的研究主要集中在大城市[10] [11] [12] [13] [14], 如北京、上海、香港、曼谷和巴勒莫, 对中小城市的关注度较少。

针对以上四点考虑, 本文拟对佛罗里达州西北部彭萨科拉大都市区采用随机森林方法且考虑协变量因素开展所有功能区土壤重金属空间分析。本文选择的研究区分布有为数不少的污染性企业, 且 2010 年曾发生原油泄漏重大污染事件, 但至今州、县政府相关部门并未对当地土壤重金属进行调查分析, 因此, 开展该项工作有助于当地土壤重金属污染防治。

2. 研究区概况

如图 1, 研究区位于佛罗里达州沿海平原的狭长地带, 处于北纬 $30^{\circ}18'$ 至 31° , 西经 $86^{\circ}38'$ 至 $87^{\circ}38'$ 之间, 幅员面积约为 3310 Km^2 。研究区地势整体平坦, 仅在桑塔罗萨县西部高地处呈缓坡状, 缓坡西向西南。在东北部、东部和东南部, 它是平缓地倾斜到倾斜。研究区的海拔高度较低, 西北部、南部和中部的地势几乎与海平面平齐, 海拔高度从海平面到海拔 0.762 m 不等。区内的海拔最高点为阳光山丘(Sunny Hill), 高度为 62.484 m 。土壤是该研究区主要自然资源之一, 主要由松散的沙子、淤泥和粘土组成, 海湾沿岸低地的土壤的排水性质由好到差, 呈水平分布。

研究区南部靠近墨西哥湾, 沙滩海岸线约为 70.81 Km , 海湾水湾共 9 个, 河流及划行航道共 7 个。其中彭萨科拉地区的海湾是墨西哥湾沿岸最深的海湾; 位于两县交界的埃斯堪比亚河是该地区最大的河流。研究区内的河流大多向南或向西南流, 如塘溪, 金水溪, 朱尼伯河、黑水河、黄河, 东河等。如图 1 所示, 整个研究区被无数的河流切割, 南部的低洼地区相对未被分割的、几乎水平的平原组成。低洼地区由平坦的森林和混合的硬木沼泽组成。

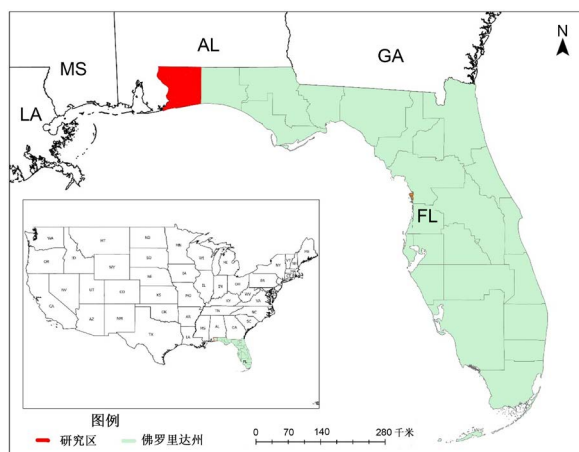


Figure 1. Study area
图 1. 研究区位置示意图

3. 数据和方法

工业及生活废水也可能导致重金属释放到土壤中[15] [16], 虽然废水中的金属浓度较低, 但长期灌溉会导致重金属在土壤中积累。本文研究区内最为常见的商业污染源为干洗店。为了清理污渍, 干洗店会用大量的干洗溶剂去清洗衣物, 而干洗店主要的清洁剂是在石油中提取出来的。大量清洁剂的使用会产生重金属污染, 排放到水体及土壤中。干洗店在清洗结束后会集中晾晒衣物, 这一过程会将残留在干洗衣物上的化学物挥发到空气中, 形成的污染又可沉降到土壤表层。综合分析, 污水排放点、干洗店站点是影响土壤重金属的变量之一, 因此本文选择污水排放点、干洗店站点为空间分析变量。

城市地区最常见的重金属来源不仅包括工业活动,还包括车辆交通和道路基础设施[17][18][19][20]。交通重金属排放也是土壤生态系统污染的重要来源。周卫红等(2017)人在研究中发现工业源、交通源和农业源是土壤中重金属浓度的首要原因[21]。日常的通勤机动车排放出大量含有重金属的尾气,含四乙基铅汽油燃烧排放的铅,在空气沉降作用下,重金属颗粒下沉在地表,造成土壤重金属污染,增加了城市和乡村地区主要道路附近土壤中铅的含量。汽车轮胎(磨损)添加剂中含有的重金属元素如锌(Zn)和镉(Cr),同样加剧了道路附近的土壤污染。还有研究发现(Zhang *et al.*, 2012),土壤中 Zn、Pb 含量与交通流量显著相关,虽然这一发现和实际情况比较符合,但由于不同时间段内,道路上的机动车流量是不同的,难以计算和控制,此方面的研究尚不多见[22]。本文收集到了研究区年均日交通量(AADT),从而将其作为土壤重金属来源的协变量之一,用于模型检测分析。

有些重金属可以被水移动,但在不透水层的阻挡下,该移动会被削弱,重金属会停留在某一表面并被水稀释[23]。而对于不透水层强度不够的土壤,其中的重金属会随地下水从表层向底层转。例如,我们可以在 A 层(A 层为表土层,依据美国土壤分类系统 Soil Taxonomy, Soil Survey Staff, 1999, 土壤分为 O、A、E、B、C、R 六层)发现重金属污染,但在 E(诊断化育层)层没有。研究证明,不透水层数据(impervious)会直接或间接影响土壤中重金属的浓度,有必要作为协变量加入到模型中[24]。

土壤的分类类别和性质可以影响重金属的浓度,如晶体结构、pH 值和颗粒大小[25]。有研究表明,土壤中颗粒态有机质及其占总有机碳的比例随重金属积累的增加而增加;而微生物量碳占总碳的比例却随土壤重金属污染水平的提高而下降[26]。有学者发现酸化可能会增加土壤中重金属的生物利用度和毒性[27]。在整个佛罗里达地区,土壤酸碱性因土壤的组成成分不同而存在差异,例如松林等林地土壤酸性相对较强,由富钙的物质形成的土壤往往是碱性。本文研究区北部为酸性土壤,南部为碱性土壤。因此土壤属性数据也作为协变量加入到研究之中。本文土壤属性数据包括土壤分类大类组(great group)、将土壤粒径(particle size)以及生物炭(soc_5)。该研究中土壤粒径(Particle Size)采用的是 12 级分类法(USDA: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_054261)。不同土壤深度所含重金属种类以及浓度存在差异,土壤表层土往往含有较多的土壤重金属,因此在检测中,选取 0~5 cm 的表层土壤有机碳数据(soc0_5)作为协变量加入到模型中。

本文采用了随机森林模型对佛罗里达州西北部的土壤中重金属进行空间分析,并进一步分析其污染源以及不同影响因子的重要性。根据土壤重金属污染的主要来源(包括垃圾填埋场、交通运输产生的铅油、合成化学品的使用、制造业工业废水、城市污泥以及空气污染等)选取了 14 个协变量(土壤重金属浓度影响因子)的数据。将土壤采样数据、土地利用与土地覆被分类等协变量数据加载到随机森林模型中得到佛罗里达西北部地区土壤重金属浓度栅格图并对其进行分析,这将有助于确定区域内环境管理和修复的热点问题。

3.1. 数据来源

3.1.1. 土壤数据

本文的土壤样本选取构建了包含 270 个潜在样本的采样点地图。对于研究区内位于非城区的公共场,则将其全部保留为潜在的采样点;对于城区内的公共场所,则是通过地理信息系统人工选择潜在的采样点,使得潜在采样点在空间中分布均匀。最终获得了研究区内的 129 个土壤采样点,收集了铅(Pb)、和锌(Zn)的重金属数据。

3.1.2. 地理信息系统(GIS)数据

1) 研究区行政区划数据:获取佛罗里达州县级行政边界,选取埃塞坎比亚县及桑塔罗萨县作为研究区的边界范围。该行政数据来自佛罗里达州地理数据库。

2) 土壤分类及土壤属性数据: 该数据来自美国农业局, 选取该数据库内分类最为详细的县级土壤类别及土壤属性数据。该数据包含不同级别的土壤分类(土纲, 亚纲, 土类, 亚类, 土属, 亚属, 土种); 土壤粒径分级, 该研究中土壤粒径(Particle Size)采用的美国农业局(USDA)的分类方法(研究区内含有 8 类: coarse-loamy, fine, fine-loamy, fine-loamy over, fine-silty, loamy, not used, sandy), 土壤有机碳(SOC)含量(6 层: 0~5 cm, 5~20 cm, 20~50 cm, 50~100 cm, 100~150 cm, 150~150+ cm)。

3) 土地利用及土地覆被分类数据: 该数据来自美国多分辨率土地特征(MRLC)联盟, 本文选取全美 2011 年 30 m 分辨率的土地利用与土地覆被分类数据, 该数据包含 9 个土地利用及土地覆被类型, 其下又分出亚类, 共 20 个类型。

4) 不透水层数据: 本文选取 MRLC 2011 年 30 m 分辨率的不透水层数据。

5) 道路交通数据: 年均日交通量(AADT: Annual Average Daily Traffic)数据来自(FGDL)。它是一条公路或公路一年的车辆总流量除以 365 天后得到的数据。AADT 是一种简单但有用的道路繁忙程度的测量方法。

6) 有毒物质来源相关数据: 其中包括 6 个影响土壤重金属聚集的潜在影响因素, 记录均为矢量数据, 可以是点源污染。相关数据。

3.2. 数据处理

3.2.1. 数据处理流程

本研究中的数据最终全部转化为地理信息系统可识别的文件添加到 ArcGIS Pro 中进行处理, 采用的投影坐标为 NAD 1983 UTM Zone 16N, 并对所有层进行了重新设计。根据采样数据将不同的特征加入到属性中, 使用 Arc GIS 中的样本, 得到 400 多万个预测点。将所有特征层合并到随机森林回归模型中, 然后运行该模型, 计算 RMSE 等参数以了解该预测的准确性, 最后获得预测结果。

3.2.2. 协变量数据处理

本研究所涉及的 14 个协变量数据有栅格影像数据也有矢量点源数据, 其中点源污染的矢量数据中又分为含排放量的矢量数据和不含排放量的点源数据。为实现佛罗里达西北部重金属含量空间分析及源分析, 并保证预测及分析结果的准确可靠, 对研究所涉及的初级数据进行相应的数据预处理。

1) 有排放量线数据: 年均日交通量(AAFT), 采用线密度工具进行分析, 获得关于年均日交通量的面数据。彭萨科拉市区附近交通流最大, 研究区北部的交通流较少。

2) 有排放量点源数据: 根据佛罗里达环保局污染物排放点数值记录(DEP_Emission), 使用默认的“搜索半径”, 计算“核密度(Kernel Density)”曲面。对于有毒物质释放清单(TRI), 利用“核密度”计算了, 空气中的污染物质“Total_air”的密度表面。

3) 无排放量点源数据: 垃圾填埋场站点(superfund), 干洗店站点(dry clean), 有害固体废物站点(solid waste), 被污染土地地点(brownfield), 污水处理站点(sewer treatment), 主要次要排放点(major_minor_emitters), 此类数据无加权, 根据点的丰度计算得出对应的“核密度”面。

3.3. 创建准确性评估数据集

为了验证模型结果本研究将模型中的实际土壤采样点作为评价结果准确性的数据集。即该研究中的训练样本也为研究中的准确性评价样本。选取保留了 20% 的数据作为参与评估训练数据的样本(% of Training Data Excluded for Validation, 保留为测试数据集以进行验证的百分比), 即有 20% 的数据作为依据对整体样本进行准确性评价, 该参数可以设置的区间范围为 10% 到 50% 之间, 模型默认数值为 10%。模型将在没有这些随机数据子集的情况下进行训练, 并将这些特征的观测值与预测值进行比较默认是 10%。

原本的实际采样点共 121 个,在前期测验模型时,将全部样本点加入模型中,重金属铅(Pb)与重金属锌(Zn)得出的预测结果的 RMSE 分别为 29.51 与 29.59。预测结果因异常值的存在而不是很理想,即模型在运行过程中受异常值(outlier)影响过大,使得模型的鲁棒性(Robustness)较差。因此,需要剔除异常值(计算公式如公式 1 及 2 所示)来提高模型预测的景区行。

$$Outlier_{left} \leq Q_1 - 1.5(Q_3 - Q_1) \quad (1)$$

$$Outlier_{right} \geq Q_3 + 1.5(Q_3 - Q_1) \quad (2)$$

模型中实际用到的采样点数值 $\in (Outlier_{left}, Outlier_{right})$

其中, Q_1 (Quartile1)为数据顺序排列的四分之一处的数值, Q_3 (Quartile3)为数据顺序排列的四分之三处的数值。

对于重金属铅(Pb),计算出来的异常值分别为-10.5 和 30.3。通过以上剔除异常值的计算后得到 105 个实际采样点作为训练样本和结果准确性评价样本加入到检测模型中(图 2 中呈黄色)。

对于重金属锌(Zn),计算出来的异常值分别为-17.475 和 47.475。通过以上剔除异常值的计算后得到 102 个实际采样点作为训练样本和结果准确性评价样本加入到检测模型中(图 2 中呈红色)。

图 2 为模型中训练样本集及准确性评估的数据集。对于研究区内的重金属铅(Pb)以及重金属锌(Zn),其准确性评估数据所在的地理位置大部份是相同的(图 2 中呈橘黄色)。

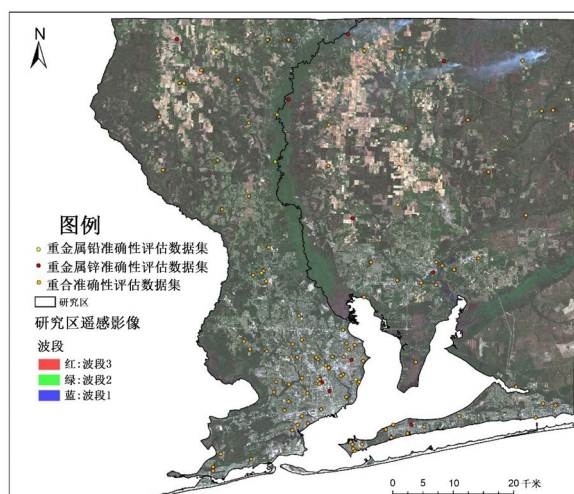


Figure 2. Accuracy evaluation samples

图 2. 准确性评估的数据集

4. 结果分析

4.1. 土壤重金属浓度空间分析

图 3 为重金属铅(Pb)在研究区内的空间分布图,绿色表示土壤中重金属铅(Pb)的浓度较低,红色表示土壤中重金属铅(Pb)的浓度较高区,其浓度数值范围在 3.61~20.97 mg/kg 之间(此处保留两位小数)。研究区内的重金属铅(Pb)主要集中在河流流域、河流两岸以及湿地沼泽,彭萨科拉(Pensacola)城市圈以及北部地区的含量次之,海滩附近的含量最少。

空间分布图表明,研究区内严重污染的场地较为集中。河流流域中重金属铅(Pb)的比例较高,推断它的主要的污染来源属于人为来源。研究区南部靠近墨西哥湾,曾有一段时间当地居民的饮水资源受到

污染，人民只能通过购买纯净水来满足日常生活需求。这主要源于 2010 年墨西哥湾发生石油泄漏事件，大规模的石油泄漏对该墨西哥湾附近的区域的水资源造成严重污染，虽然当地政府以及联邦机构曾采取行动进行治理，但石油在风力以及水波推力的作用下逐渐扩散，泄露的石油蔓延到了研究区所在地，并通过渗透作用迁移至当地土壤加剧了当地土壤重金属铅(Pb)的污染。

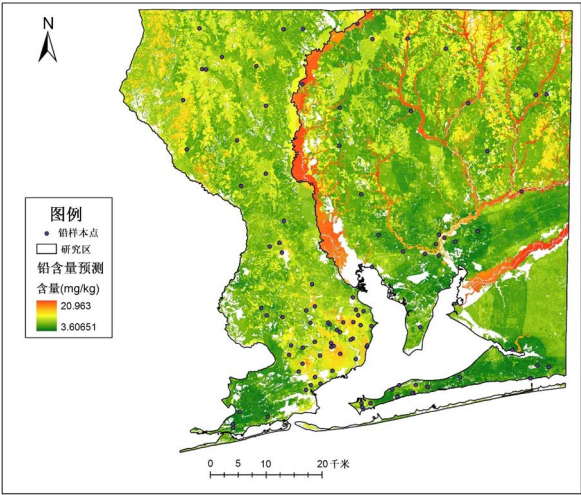


Figure 3. Lead concentration map
图 3. 重金属铅(Pb)空间分布

在彭萨科拉城区，由于人口较为密集，相较于其它地区来说其交通流量较大。而含四乙基铅汽油在燃烧时会向空中排放的铅，而重金属铅(Pb)在高温处理过程中会挥发，从而导致这些金属转化为氧化物并凝结成微粒，沉降在土壤表面难以降解，污染附近的土壤。这可能为城市地区和主要道路附近土壤重金属铅(Pb)浓度较高的原因。

研究区的北部主要是 Palafox 工业走廊地带，重金属铅(Pb)的含量也相对较高，这是由于之前工业废弃物残留的污染。由于重金属对土壤的污染难以降解，因此工业区留下的铅(Pb)污染会积累沉淀，难以快速消除，对附近土壤产生长久影响。

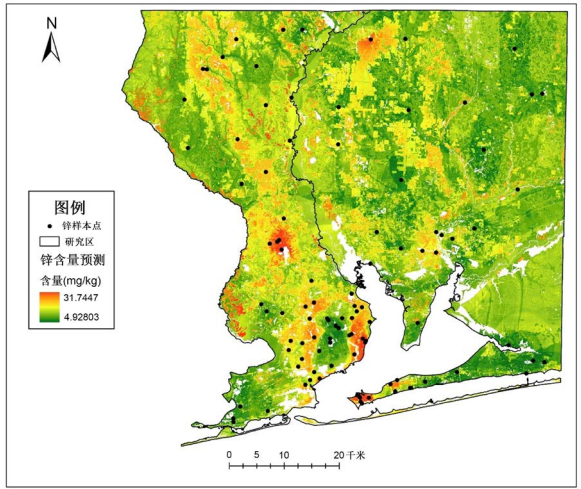


Figure 4. Zinc concentration map
图 4. 重金属锌(Zn) 空间分布

图4为重金属锌(Zn)在研究区内的空间分布图,绿色表示土壤中重金属锌(Zn)的浓度较低,红色表示土壤中重金属锌(Zn)的浓度较高区,其浓度数值范围在4.93~31.74mg/kg之间(此处保留两位小数)。研究区内的重金属锌(Zn)主要集中在彭萨科拉(Pensacola)城市圈处以及北部老工业区和海滩附近。

空间分布图表明,研究区内严重污染的场地较为集中,尤其是在彭萨科拉(Pensacola)城市圈、北部 Palafox 老工业走廊和海滩附近,农业用地重金属锌(Zn)含量次之。

彭萨科拉(Pensacola)城市圈、北部 Palafox 老工业走廊和海滩的共同特点表现为人流量大,车辆流通密集。汽车轮胎和润滑油含有大量的重金属锌(Zn),因此车辆在行驶中容易产生重金属锌(Zn)并转移至附近土壤表层,造成这些区域重金属锌(Zn)含量高于其他地区。

4.2. 土壤重金属含量检测模型影响因子重要性分析

图5表示的是对研究区内重金属铅(Pb)聚集产生影响的14个因子的重要性数值(自上往下,重要性依次降低)。根据数据,土壤粒径(Particle Size)、不透水层(Improves)以及年均日交通量(AADT)对研究区重金属铅(Pb)的影响较大,其重要性总占比超过了整体的 $\frac{1}{3}$ 。

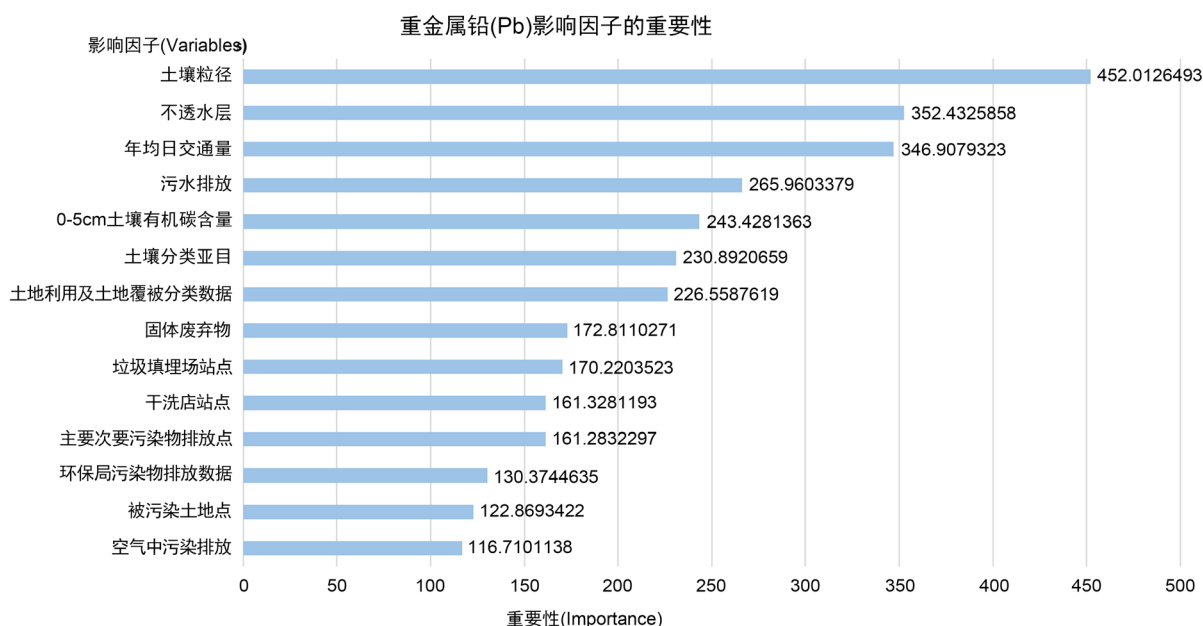


Figure 5. Importance of Lead influencing factor

图5. 重金属铅(Pb)影响因子重要性

根据重要性与上一小节的空间分布图相结合分析可得知,该区域土壤重金属铅(Pb)聚集主要受土壤粒径(Particle Size)大小的影响,其中土壤颗粒直径会影响重金属铅(Pb)在土壤中的吸附作用。吴婷(2017)等人的研究报告指出,重金属铅(Pb)在不同土壤粒径条件下对土壤的污染程度不同,不同粒径的土壤颗粒具有不同的累积因子、污染系数和活性系数,这三者会随着颗粒物粒径的减小而增大。随着土壤粒径的减小,重金属铅(Pb)含量升高。土壤粒径与重金属铅(Pb)污染程度的关系表现为:粉粒组(0.01~0.05 mm) > 细砂粒(0.05~0.25mm) > 粗粘粒(0.001~0.002 mm) > 细粘粒(<0.001 mm),而研究区内的土壤以由淤泥(粒径 0.05~0.002mm)、沙子(粒径 0.063~2 mm)和少量粘土(粒径<0.002 mm)组成的壤土(lomay)为主(https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/nedc/training/soil/?cid=nrcs142p2_054262),尤其是在河流流域以及湿地沼泽地带,这可能为以上区域重金属铅(Pb)更为聚集的原因。

研究区的重金属铅(Pb)污染主要来自人类活动,其中 2010 年墨西哥湾石油泄漏事件对该区域的影响依旧存在/被释放到水体的重金属,以不同的方式被土壤截留固定。重金属铅(Pb)一旦进入土壤就很难被分解分离出来,修复土壤重金属污染是一个长期过程,从源头加强管理防范是保护土壤免受过量重金属污染的有效措施。

图 6 表示的是对研究区内重金属铅(Pb)聚集产生影响的 14 个因子的重要性数值(自上往下,重要性依次降低)。根据表格 4-5 中的数据,年均日交通量(AADT)、0~5 cm 土壤有机碳含量(soc05)以及干洗店(dry_clean)对研究区重金属锌(Zn)的影响较大,其重要性总占比超过了整体的 $\frac{1}{3}$ 。

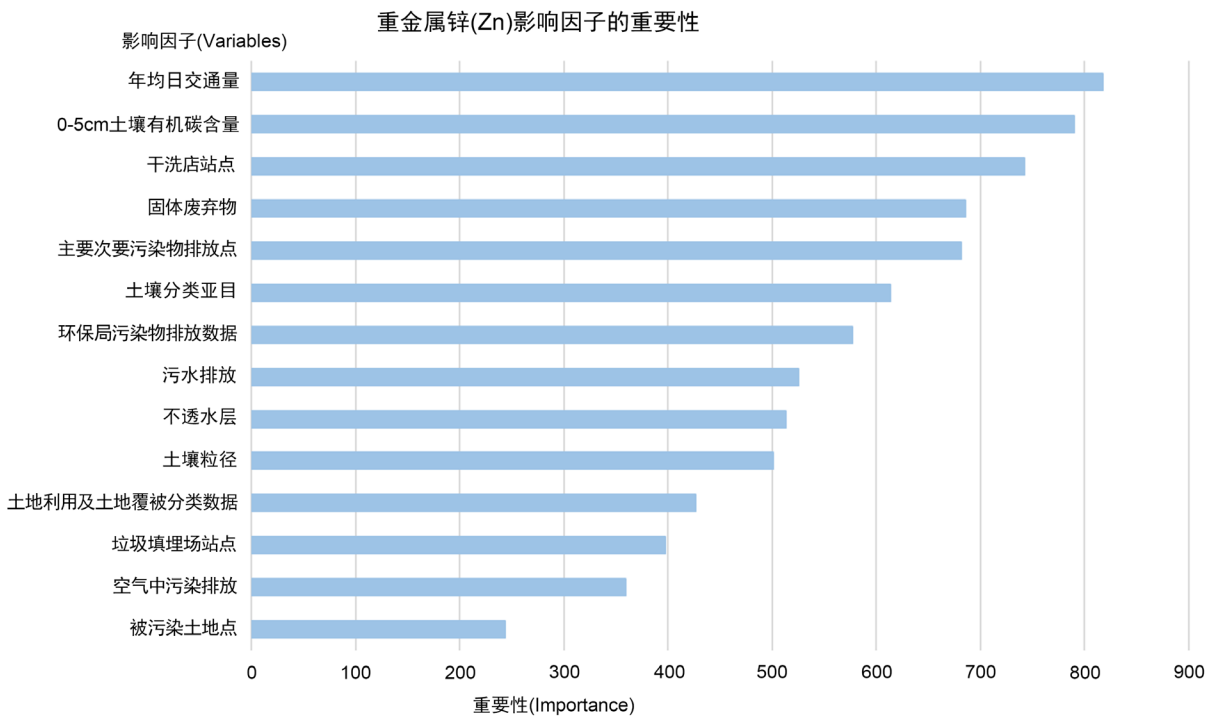


Figure 6. Importance of Zinc influencing factor
图 6. 重金属锌(Zn)影响因子重要性

根据重要性与上一小节的空间分布图相结合分析可得知,该区域土壤重金属锌(Zn)的污染也主要为人为来源,且受交通运输(年均日交通量 AADT)的影响最大。交通运输中轮胎和润滑油含有大量重金属(Zn),机动车在行驶过程中与路面摩擦时,排放尾气,将重金属锌(Zn)释放到周围环境中,并在道路附近的土壤沉淀。

此外有机碳含量也能明显反应土壤重金属锌(Zn)的浓度。研究表明(王秀丽,徐建民等,2002),重金属锌(Zn)和铜(Cu)的浓度与土壤微生物生物量碳的浓度密切相关。土壤有机碳含量与有重金属锌(Zn)的含量呈显著负相关。本研究仅得出了土壤有机碳含量在预测土壤重金属锌(Zn)浓度时有明显作用但并未对其二者间具体的回归关系做出研究。

对比重金属铅(Pb)与重金属锌(Zn),只有交通运输(年均日交通量 AADT)的重要性均超过全部影响因子 $\frac{1}{10}$,这与美国环保局(EPA)研究结果相近(US Government, 2001)。美国环保局(EPA)研究结果指出了 21 种来自交通污染的主要重金属有毒物,其中就包括铅(Pb)以及锌(Zn),且这两类重金属有毒物质分别列居

第一位和第三位。分析铅(Pb)、和锌(Zn)在公路两侧土壤中的分布格局,为公路的规划建设、公路沿线的农业生产布局及公路交通导致的重金属污染的防治提供科学依据。

5. 结论

1) 获得土壤重金属浓度预测图:研究区内的重金属铅(Pb)主要集中在河流流域、河流两岸以及湿地沼泽,彭萨科拉(Pensacola)城市圈以及北部地区的含量次之,海滩附近的含量最少。研究区内的重金属锌(Zn)主要集中在彭萨科拉 (Pensacola)城市圈处以及北部老工业区和海滩附近。

2) 影响土壤重金属含量因子的重要性:在选取输入的 14 个因子中,土壤粒径(Particle Size)、不透水层(Improves)以及年均日交通量(AADT)对研究区重金属铅(Pb)的影响较大。年均日交通量(AADT)、0~5 cm 土壤有机碳含量(soc)05以及干洗店(dry_clean)对研究区重金属锌(Zn)的影响较大。对比重金属铅(Pb)与重金属锌(Zn),只有交通运输(年均日交通量 AADT)的重要性均超过全部影响因子 1/10,这与美国环保局(EPA)研究结果相近(US Government, 2001)。美国环保局(EPA)研究结果指出了 21 种来自交通污染的主要重金属有毒物,其中就包括铅(Pb)以及锌(Zn),且这两类重金属有毒物质分别列居第一位和第三位。

6. 总结与展望

1) 获得土壤重金属浓度预测图:研究区内的重金属铅(Pb)主要集中在河流流域、河流两岸以及湿地沼泽,彭萨科拉(Pensacola)城市圈以及北部地区的含量次之,海滩附近的含量最少。研究区内的重金属锌(Zn)主要集中在彭萨科拉 (Pensacola)城市圈处以及北部老工业区和海滩附近。

2) 影响土壤重金属含量因子的重要性:在选取输入的 14 个因子中,土壤粒径 (Particle Size)、不透水层(Improves)以及年均日交通量(AADT)对研究区重金属铅(Pb)的影响较大。年均日交通量(AADT)、0~5 cm 土壤有机碳含量(soc)05 以及干洗店(dry_clean)对研究区重金属锌(Zn)的影响较大。对比重金属铅(Pb)与重金属锌(Zn),只有交通运输(年均日交通量 AADT)的重要性均超过全部影响因子 1/10,这与美国环保局(EPA)研究结果相近(US Government, 2001)。美国环保局(EPA)研究结果指出了 21 种来自交通污染的主要重金属有毒物,其中就包括铅(Pb)以及锌(Zn),且这两类重金属有毒物质分别列居第一位和第三位。

3) 重金属铅(Pb)与重金属锌(Zn)预测结果的 R2 值分别为 0.35 和 0.38,两者均介于接近于 0,表明该模型在预测研究区重金属铅 (Pb)以及重金属锌(Zn)浓度方面有待提高。笔者改变模型中决策树的数量以及随机森林模型运行的迭代次数,发现 R2 值会随决策树数量的增加而增加。因此,笔者认为通过大幅度增加模型中决策树的数量来优化预测结果。在未来土壤重金属浓度空间时,笔者还将采用插值技术中的经验贝叶斯克里格 EBK (Empirical Bayesian Kriging),将所得结果与基于随机森林的回归结果进行对比分析,找出更适合研究区土壤重金属浓度检测的回归方法。此外,宋申辉(2019)等在其研究中先对土壤重金属浓度进行级别划分,然后加入到模型中进行土壤重金属浓度的分析。这一做法使得原本连续的数值数据变为分类数据,有效率减少了模型预测时数值进制对结果精确度的影响。因此,笔者在后续的研究中,会尝试先将土壤重金属浓度进行分级划分,然后转为基于随机森林的分类分析,检验这一方法是否对本研究区的土壤重金属浓度空间分析有所帮助。

致 谢

感谢 Dr. Liebens,感谢您为本文提供的土壤样本数据,有了您的土壤数据,才会有本文的产生。正是您的《土壤科学》课程,让我对土壤产生兴趣,让我了解到土壤重金属污染不仅是中国面临的严峻问题在环境优美的佛罗里达地区也有这类污染的困扰。感谢 Dr. Hu 在本文模型选取上的指导,以及在模型中影响因子选取预筛选上的建议。感谢我的导师刘汉湖副教授对我学业上关心和帮助。

参考文献

- [1] Cai, L., Xu, Z., Ren, M., Guo, Q., Hu, X., Hu, G., Wan, H. and Peng, P. (2012) Source Identification of Eight Hazardous Heavy Metals in Agricultural Soils of Huizhou, Guangdong Province, China. *Ecotoxicology and Environmental Safety*, **78**, 2-8. <https://doi.org/10.1016/j.ecoenv.2011.07.004>
- [2] Dong, B., Zhang, R., Gan, Y., et al. (2019) Multiple Methods for the Identification of Heavy Metal Sources in Crop-land Soils from a Resource-Based Region. *Science of the Total Environment*, **651**, 3127-3138. <https://doi.org/10.1016/j.scitotenv.2018.10.130>
- [3] 王春红, 阮进生, 蒋超群, 等. 基于 GIS 和地统计学的土壤重金属污染评价研究进展[J]. 安徽农业科学, 2012, 40(20): 10414-10418.
- [4] 纪冬丽, 曾琬晴, 张新波, 张竞, 王庆国, 张万军, 邓令, 杨光辉, 吴思睿. 天津近郊农田土壤重金属风险评价及空间主成分分析[J/OL]. 环境化学, 1-11.
- [5] 徐蕾, 肖昕, 马玉, 韩筱璇. 徐州农田土壤重金属空间分布及来源分析[J]. 生态与农村环境学报, 2019, 35(11): 1453-1459.
- [6] Chen, Y.X., Jiang, X.S., Wang, Y., et al. (2018) Assessment of Ecological Environment and Human Health of Heavy Metals in Mining Area Based on GIS. *Acta Scientiae Circumstantiae*, **38**, 1642-1652.
- [7] 刘硕, 吴泉源, 曹学江, 王集宁, 张龙龙, 蔡东全, 周历媛, 刘娜. 龙口煤矿区土壤重金属污染评价与空间分布特征[J]. 环境科学, 2016, 37(1): 270-279.
- [8] Zhao, H., He, B., Wang, T.Y., et al. (2019) Pollution Characteristics of Heavy Metals and Source-Sink Relationship in Typical City of the South China. *Acta Scientiae Circumstantiae*, **39**, 2231-2239.
- [9] Soffianian, A.R., Bapeer Bakir, H. and Khodakarami, L. (2015) Evaluation of Heavy Metals Concentration in Soil Using GIS, RS and Geostatistics. *IOSR Journal of Environmental Science*, **9**, 61.
- [10] Li, F., Cai, Y. and Zhang, J. (2018) Spatial Characteristics, Health Risk Assessment and Sustainable Management of Heavy Metals and Metalloids in Soils from Central China. *Sustainability*, **10**, 91. <https://doi.org/10.3390/su10010091>
- [11] Fei, X., Christakos, G., Xiao, R., Ren, Z., Liu, Y. and Lv, X. (2019) Improved Heavy Metal Mapping and Pollution Source Apportionment in Shanghai City Soils Using Auxiliary Information. *Science of the Total Environment*, **661**, 168-177. <https://doi.org/10.1016/j.scitotenv.2019.01.149>
- [12] Sergeev, A.P., Buevich, A.G., Baglaeva, E.M. and Shichkin, A.V. (2019) Combining Spatial Autocorrelation with Machine Learning Increases Prediction Accuracy of Soil Heavy Metals. *Catena*, **174**, 425-435. <https://doi.org/10.1016/j.catena.2018.11.037>
- [13] Cai, L.M., Wang, Q.S., Wen, H.H., et al. (2019) Heavy Metals in Agricultural Soils from a Typical Township in Guangdong Province, China: Occurrences and Spatial Distribution. *Ecotoxicology and Environmental Safety*, **168**, 184-191. <https://doi.org/10.1016/j.ecoenv.2018.10.092>
- [14] Huang, S., Shao, G., Wang, L., et al. (2019) Spatial Distribution and Potential Sources of Five Heavy Metals and One Metalloid in the Soils of Xiamen City, China. *Bulletin of Environmental Contamination and Toxicology*, **103**, 308-315. <https://doi.org/10.1007/s00128-019-02639-5>
- [15] Hou, D., O'Connor, D., Nathanail, P., Tian, L. and Ma, Y. (2017) Integrated GIS and Multivariate Statistical Analysis for Regional Scale Assessment of Heavy Metal Soil Contamination: A Critical Review. *Environmental Pollution*, **231**, 1188-1200. <https://doi.org/10.1016/j.envpol.2017.07.021>
- [16] Santos-Francés, F., Martínez-Graña, A., Zarza, C.Á., Sánchez, A.G. and Rojo, P.A. (2017) Spatial Distribution of Heavy Metals and the Environmental Quality of Soil in the Northern Plateau of Spain by Geostatistical Methods. *International Journal of Environmental Research and Public Health*, **14**, 568. <https://doi.org/10.3390/ijerph14060568>
- [17] Zhou, J. and Liu, H. (2013) A Spatial-Homogeneity-Based Interpolation Algorithm for Soil Properties. *International Conference on Geoinformatics*, Kaifeng, 20-22 June 2013, 1-4. <https://doi.org/10.1109/Geoinformatics.2013.6626171>
- [18] Yang, J., Cao, L., Wang, J., Liu, C., Huang, C., Cai, W., Fang, H. and Peng, X. (2014) Speciation of Metals and Assessment of Contamination in Surface Sediments from Daya Bay, South China Sea. *Sustainability*, **6**, 9096-9113. <https://doi.org/10.3390/su6129096>
- [19] Agca, N. (2015) Spatial Distribution of Heavy Metal Content in Soils around an Industrial Area in Southern Turkey. *Arabian Journal of Geosciences*, **8**, 1111-1123. <https://doi.org/10.1007/s12517-013-1240-7>
- [20] 周卫红, 张静静, 邹萌萌, 等. 土壤重金属有效态含量检测与监测现状、问题及展望[J]. 中国生态农业学报, 2017, 25(4): 605-615.
- [21] Yang, J., Lv, F., Zhou, J., Song, Y. and Li, F. (2017) Health Risk Assessment of Vegetables Grown on the Contaminated Soils in Daye City of Hubei Province, China. *Sustainability*, **9**, 2141. <https://doi.org/10.3390/su9112141>

-
- [22] Technica, G. and City, S.T. (2019) Characterization of Heat Waves: A Case Study for Peninsular Malaysia. *Geographia Technica*, **14**, 146-156. https://doi.org/10.21163/GT_2019.141.11
- [23] Liebens, J., Mohrherr, C.J. and Rao, K.R. (2012) Trace Metal Assessment in Soils in a Small City and Its Rural Surroundings, Pensacola, FL, USA. *Environmental Earth Sciences*, **65**, 1781-1793. <https://doi.org/10.1007/s12665-011-1158-x>
- [24] Shaheen, A. and Iqbal, J. (2018) Spatial Distribution and Mobility Assessment of Carcinogenic Heavy Metals in Soil Profiles Using Geostatistics and Random Forest, Boruta Algorithm. *Sustainability (Switzerland)*, **10**, 799. <https://doi.org/10.3390/su10030799>
- [25] Bhunia, G.S., Shit, P.K. and Maiti, R. (2018) Comparison of GIS-Based Interpolation Methods for Spatial Distribution of Soil Organic Carbon (SOC). *Journal of the Saudi Society of Agricultural Sciences*, **17**, 114-126. <https://doi.org/10.1016/j.jssas.2016.02.001>
- [26] Göl, C., Bulut, S. and Bolat, F. (2017) Comparison of Different Interpolation Methods for Spatial Distribution of Soil Organic Carbon and Some Soil Properties in the Black Sea Backward Region of Turkey. *Journal of African Earth Sciences*, **134**, 85-91. <https://doi.org/10.1016/j.jafrearsci.2017.06.014>
- [27] Fifi, U., Winiarski, T. and Emmanuel, E. (2013) Assessing the Mobility of Lead, Coppermore Are Cadmium in a Calcareous Soil of Port-au-Prince, Haiti. *International Journal of Environmental Research and Public Health*, **10**, 5830-5843. <https://doi.org/10.3390/ijerph10115830>