

重金属污染的卡尔曼预测及可视化系统研究

王 硕¹, 蒲丽丹², 梁雪勇¹, 王金康³, 梁永超^{1*}

¹贵州大学大数据与信息工程学院, 贵州 贵阳

²贵州大学外国语学院, 贵州 贵阳

³贵州大学药学院, 贵州 贵阳

收稿日期: 2022年6月5日; 录用日期: 2022年7月1日; 发布日期: 2022年7月8日

摘 要

传统的重金属污染预测是基于数学统计进行回归分析预测, 由于现代人类生产活动、非自然地质演变等要素所带来的噪声影响, 传统的数学统计方法预测效果欠佳, 为了解决这一问题, 针对基于卡尔曼滤波算法的重金属污染预测模型进行了研究, 并采用皮尔逊相关系数检验和Wilcoxon符号秩检验来验证该预测模型的准确性, 实验结果表明通过卡尔曼滤波算法建立的重金属污染预测模型具有良好的稳定性和准确性。此外, 重点就重金属污染防治可视化进行了研究探讨, 并基于Django架构设计的城市规划平台成功将区域重金属污染状况在地图上进行数据渲染和动态监测, 对重金属污染起到一定的预防作用。

关键词

Django, 重金属污染, 卡尔曼预测, 数据可视化, 动态监测

Research on Kalman Prediction and Visualization System for Heavy Metal Pollution

Shuo Wang¹, Lidan Pu², Xueyong Liang¹, Jinkang Wang³, Yongchao Liang^{1*}

¹College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

²College of Foreign Languages, Guizhou University, Guiyang Guizhou

³School of Pharmaceutical Sciences, Guizhou University, Guiyang Guizhou

Received: Jun. 5th, 2022; accepted: Jul. 1st, 2022; published: Jul. 8th, 2022

*通讯作者。

文章引用: 王硕, 蒲丽丹, 梁雪勇, 王金康, 梁永超. 重金属污染的卡尔曼预测及可视化系统研究[J]. 计算机科学与应用, 2022, 12(7): 1719-1727. DOI: 10.12677/csa.2022.127172

Abstract

Traditional heavy metal pollution prediction is based on mathematical statistics for regression analysis prediction, due to the noise influence brought by modern human production activities, unnatural geological evolution and other elements, the traditional mathematical statistics method prediction effect is poor, in order to solve this problem, for the heavy metal pollution prediction model based on Kalman filter algorithm was studied, and the Pearson correlation coefficient test and Wilcoxon The experimental results show that the heavy metal pollution prediction model established by Kalman filtering algorithm has good stability and accuracy. In addition, the research is focused on the visualization of heavy metal pollution prevention and control, and the urban planning platform designed based on Django architecture successfully renders the regional heavy metal pollution status on the map for data rendering and dynamic monitoring, which plays a certain role in the prevention of heavy metal pollution.

Keywords

Django, Heavy Metal Pollution, Kalman Prediction, Data Visualization, Dynamic Monitoring

Copyright © 2022 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

城市是一个复杂综合的集成系统,研究城市重金属污染对保障当地生态环境及制定相应的预防决策具有重大的现实意义[1]。有关重金属污染预测模型的研究,前人已做了大量研究工作,孙少龙等人提出了基于灰色-灾变与回归预测的组合模型分析了城市地表层土壤重金属浓度的变化规律,以预测未来城市土壤重金属的灾变日期和相对应的日期数值[2];骆永明就我国土壤污染临界负荷进行了研究探讨,并指出在未来研究中需要加强对临界浓度制定和重金属污染动态模型构建的研究[3];以姚琳和廖欣峰为代表的杭州市环境监测中心站提出了建立大气重金属污染预测预警系统的构想,并首次采用建立重金属危害指数的方法,解析和评估大气颗粒物重金属的环境危害[4]。此外还有其他学者提出了很多稳定性良好、准确性高的预测模型,但大多是基于数学统计的方法对重金属污染进行回归分析以达到一定的预测作用,国内有关重金属污染动态预测模型的研究仍然处于待发展阶段,大部分学者就重金属污染动态预测模型的研究仅仅只提出了一些构想和概念设计,并没有深入实际的实现。目前有关重金属污染的可视化研究,只是从理论角度阐述了可视化的实现方法及可视化平台的概念设计,并没有给出具体的实现过程。

此外,现代城市重金属污染的变化趋势是一个非线性的动态过程,人类活动、工业污染、地质变化等外在不稳定因素都使得重金属污染的变化趋势不能简单地以传统的数学统计方法来描述。为了准确地预测区域重金属的污染状况,本文结合卡尔曼滤波算法,采用时间变化协和量模拟重金属污染的非自然演变误差,对多个时空状态下的重金属污染状况进行多次迭代更新,从动态系统中通过稳定的状态方程描述重金属污染的变化趋势,借助不断迭代更新的当前状态观测值去修正下一时态下重金属污染的预测值,实现重金属污染的动态预测,以达到重金属污染的预防作用。本文基于 Django 架构设计了城市规划平台,重点研究重金属污染防治可视化的具体实现过程,对城市重金属污染的动态监测及预防起到一定的作用。

2. 系统开发的应用框架

本文通过前后端分离的架构模式开发了城市规划平台。项目前端基于 vue 渐进式框架设计，后端以 Django 架构搭建，基于 MVC (Model (模型) + View (视图) + Controller (控制器))设计模式[5]，实现程序的动态设计。

如图 1 所示，项目的整体运行逻辑如下：首先由 Django 框架接收用户在前端发送的 POST 请求，接着自动匹配 URL 转发给视图系统(view.py)处理，视图系统是开发框架中的核心，负责接收前端请求、访问数据库数据、返回后端处理结果。当后端处理完前端的请求后，最后由视图系统将处理得到的结果返回给用户。

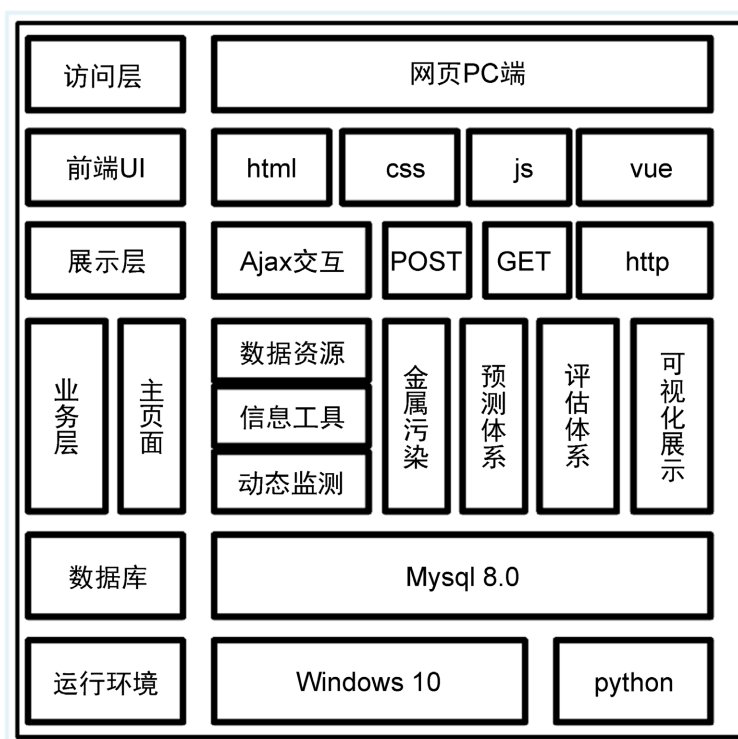


Figure 1. Urban planning platform system architecture
图 1. 城市规划平台系统架构

此外，系统的业务层包括金属污染的地图渲染、城市规划影响指标的热力图可视化展示、城市规划发展水平预测及评估；展示层提供浏览器访问页面时的 AJAX (异步的 JavaScript 和 XML)请求；前端界面采用最新的 vue.js 框架，使得整个页面美观大方，并能让用户在使用平台时有较好的体验；顶层的访问系统采用 HTTP (超文本传输协议) [6]，实现实时响应功能。

有关数据库方面，该平台采用体积小、速度快、使用成本低的关系型数据库，即 MySQL 数据库。按照业务逻辑关系的不同，数据库分别建立用户的注册表和管理各功能模块的数据表，其中功能表的 Value 属性表示对应经纬度下某一指标的属性值，比如在此经纬度下重金属的含量。为了避免数据冗余以及数据的操作问题，所有数据库中表的设计严格遵守第三范式。此外，本项目中的数据主要来源于贵阳市政府数据开放平台与贵州大学资源与环境工程学院。

当用户登录成功后，页面跳转到主页面。如图 2 所示，主页面集成了多个功能模块的入口，包括数据汇总、金属污染、可视化展示、预测体系、评估体系以及平台信息。除此之外，还包括一些图形化小

组件来丰富主页面的内容，如新闻资讯、国家有害重金属污染标准以及平台设计的相关介绍。简约、大方的商业化界面风格能够使用户在使用该平台时拥有更出色的体验。



Figure 2. Main interface display page

图 2. 主界面展示页面

后台管理页面基于 vue 框架编写而成，实现了后台用户访问量的统计、代办事务的统计、主要数据指标的数据可视化报表显示，其后台管理平台由默认的超级管理员来管理。首先由数据库中 Date 属性检索当日数据库中存储的各项数据指标，接着由后端通过 SQL 查询语句从数据库中获取该数据并封装为 json 类型的数据集，然后通过 post 方法传递到前端页面，最后由前端页面解析该 json 类型的数据集，并根据解析后的贵阳市城市规划数据指标生成数据可视化报表，从而更直观、简洁地反映贵阳市的城市实时状况，以帮助管理员制定城市规划的运营决策。后台管理系统只是就城市区域内主要重金属含量的实时情况做数据可视化报表，以实现重金属污染状况的动态监测和实时反馈，接下来将对重金属污染的数据可视化做一个更加深入的研究。

3. 重金属污染的可视化研究

本文基于高德地图组件的热力图实现重金属污染在地图上的数据渲染。首先，需要高德地图共享平台中申请个人使用的 key 键，再使用 vue-amap 在 vue 中引入高德地图，最后在 main.js 配置文件中引入高德地图插件，便可以在前端实现地图功能。

在完成上述配置后，即可进入数据渲染页面。当进入数据可视化功能模块时，设置 1000 ms 的缓冲时间，这是为了避免高德地图等第三方插件不能及时加载完毕而导致部分功能失效。所有插件加载完毕，地图主视角默认定位在当前 ip 地址的地理位置，根据时间、地理位置等条件检索所需要渲染的数据，此类数据封装为 json 类型，单条数据格式为 {lng: 经度值, lat: 纬度值, value: 渲染程度}。检索好所需要渲染的数据后，将会对 heatmap 对象进行初始化，依次设置热力图渲染的样式和数据集，调用创建热力图的函数便可在地图上实现数据可视化功能。如果后续有新增数据，可通过 ajax 方法依次加载新增加的数据，通过对页面的局部刷新达到页面的动态更新效果。

如图 3 所示，按照中国重金属污染风险评价标准进行风险程度划分[7]，在地图上以不同层次的颜色直观反映重金属污染的风险程度，采用 ajax 方法实时更新数据能有效实现重金属污染的动态渲染，实现重金属污染的动态监测，对城市重金属污染起到一定的监管和预防作用。

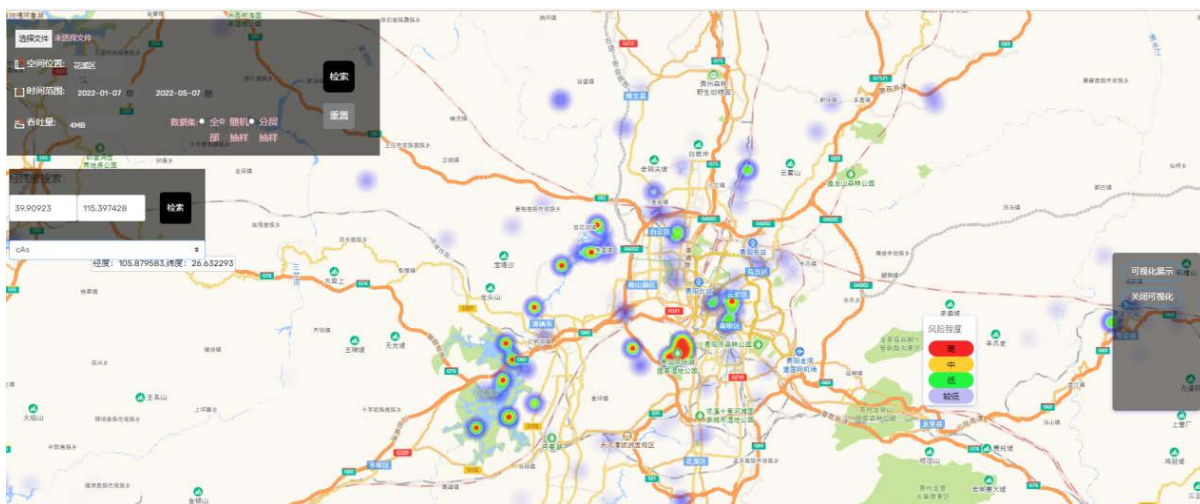


Figure 3. Data rendering page display

图 3. 数据渲染页面展示

4. 重金属污染指标的卡尔曼插值预测模型

由监测站点采集到的数据为离散采样数据，且重金属污染数据具有动态性，会随着时空变化在一定范围内波动。此外，在对贵阳市重金属污染数据进行采集的过程中由于人工操作不当或仪器精度等原因产生一定的过程误差和观测误差。为了更好地预测重金属污染数据，并准确反映这种波动，本文采用卡尔曼预测方法进行重金属污染浓度的预测。卡尔曼预测的实质是采用递归思想，通过均方差最小准则得到最优估计值。由已知测量数据的值得到状态矢量的最佳初值和状态协方差矩阵初值，获得最优估计的一系列数值，从动态系统中通过稳定的状态方程描述趋势，起到预测作用[8]。

本文通过以下的状态方程和测量方程模拟了一个非线性重金属污染指标预测系统模型[8] (以贵阳市土壤中重金属 Pb 含量为例):

$$X_k = G(X_{k-1}, U_k) + W_k \quad (1)$$

$$Y_k = h(X_k) + \delta_k^2 \quad (2)$$

其中 X_k 表示在离散时间参数 k 时的系统状态向量， $G(X_{k-1}, U_k)$ 表示系统函数在 $k-1$ 时刻的状态函数,其中 U_k 表示在 k 时刻下的输入， W_k 表示 k 时刻下的过程噪声，即状态方程的时间变化协方差量。 Y_k 表示 k 时刻系统输出，即观测量， $h(X_k)$ 是系统函数在 k 时刻的状态， δ_k^2 为观测噪声。由于在数据采集过程中，采集仪器及操作人员是固定的，故将过程噪声 W_k 和观测噪声 δ_k^2 为一个定量。

将 $G(X_{k-1}, U_k)$ 与 $h(X_k)$ 分别进行线性化展开[9]，模型可简化为：

$$X_k = AX_{k-1} + BU_k + W_k \quad (3)$$

$$Y_k = CX_k + \delta_k^2 \quad (4)$$

其中 A 为状态转移矩阵， B 为输入控制矩阵， C 为输出系数矩阵。

具体算法思想如下：

1) 由已知的 $k-1$ 时刻的重金属指标数据的最优估计 X_{k-1} 去预测 k 时刻系统的先验估计 \hat{X}_k 。

$$\hat{X}_k = AX_{k-1} + BU_k + W_k \quad (5)$$

2) 由先验估计 \hat{X}_k 的表达式求得先验估计的协方差方程，由于 W_k 本身就是过程噪声，故 W_k 的协方差就是自身，其中 U_k 表示 k 时刻的输入而不是随机变量，准确输入的协方差为 0，故预测协方差方程为

$$\text{cov}(\hat{X}, \hat{X}) = \text{cov}(AX_{k-1} + BU_k + W_k, AX_{k-1} + BU_k + W_k) \quad (6)$$

$$\text{cov}(\hat{X}, \hat{X}) = A \text{cov}(X_{k-1}, AX_{k-1}) A^T + W_k \quad (7)$$

3) 由上一次的误差协方差 P_{k-1} 和过程噪声 W_k 来计算下一状态的预测协方差 \hat{P}_k ，其中

$$\hat{P}_k = AP_{k-1}A^T + W_k \quad (8)$$

4) 计算卡尔曼增益

$$K_k = \frac{\hat{P}_k C^T}{C \hat{P}_k C^T + \delta_k^2} \quad (9)$$

5) 计算当前最优估计值

$$X_k = \hat{X}_k + K_k (Y_k - C \hat{X}_k) \quad (10)$$

6) 更新协方差矩阵，以迭代更新求得下一状态下重金属 Pb 指标的最优估计值。

$$P_k = (1 - K_k C) \hat{P}_k \quad (11)$$

卡尔曼预测是一个递归的过程，可以结合上一时刻状态的估计值与当前时刻状态的观测值，同时使用观测值修正当前值，得到最优结果。本文以 2021 年 11 月当中的 28 天内贵阳市花溪区十里河滩区域内重金属 Pb 的含量为例，结合三次样条插值方法将数据曲线进行拟合插值，绘制出贵阳市花溪区十里河滩区域内重金属 Pb 的月平均含量的插值模拟图，模拟十里河滩区域内重金属 Pb 的空间分布及局部污染特征，并拟合地理区域内每个经纬度网格对应的重金属 Pb 的含量浓度，最终检验拟合值与实际值的差异，以验证本方法的可行性与准确性。

由于本文预测的目标数据为一维数据，故初始化 $A = 1, B = 0, C = 1$ ，由于 W_k 和 δ_k^2 无论取何值都会使模型最终收敛，估这两个参数的取值对最终的预测模型影响不大[10]。经过多次调整超参数可得：本实验状态矢量的最佳初值 $X_0 = 5$ ，状态协方差矩阵初值 $P_0 = 20$ 。

下面以贵阳市花溪区十里河滩区域 2021 年 11 月 1 日至 2021 年 11 月 28 日的重金属 Pb 含量为例验证卡尔曼预测模型的准确性，本实验假定观测值大小等同于真实数据基础上叠加一个高斯噪声(0, δ_k^2)，其中 $\delta_k^2 = 0.25$ ，过程噪声 $W_k = 1$ 。如图 4 所示，其中 11 月 4 日，11 月 12 日，11 月 14 日等日卡尔曼最

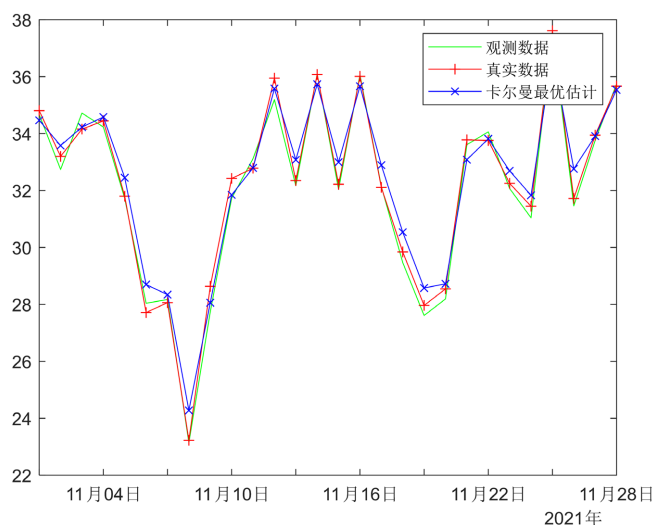


Figure 4. Comparison of predicted and actual heavy metal Pb content in Shilihe Beach, Guiyang
图 4. 贵阳市十里河滩重金属 Pb 含量预测与实际对比图

优估计预测精确度高达 94%以上, 与真实数据偏离最大的 11 月 19 日预测精确度也达到了 87%, 卡尔曼最优估计的预测效果整体良好, 能较好地反映其真实数据的变化趋势。

$$R = \frac{\sum_i^N (M_i - \overline{M_i})(O_i - \overline{O_i})}{\sqrt{\sum_i^N (M_i - \overline{M_i})^2 \cdot \sum_i^N (O_i - \overline{O_i})^2}} \quad (12)$$

为了验证卡尔曼插值模型的准确性, 本文采用皮尔逊相关系数检验来验证卡尔曼预测值与真实数据之间的关联性。如公式 12 所示, R 为相关系数, M_i 为预测值, O_i 为实际观测值, $\overline{M_i}$ 和 $\overline{O_i}$ 分别为预测值与实际观测值的均值, 而相关系数 R 是衡量两个变量之间相互关系的参数。两个变量之间的相互关系大体上可以分为 3 种, 正相关、负相关和无关。一般说来, 相关系数 R 的绝对值大于 0.8 就认为两个变量有很强的相关性[11]。经计算可得 $R = 0.87$, 说明基于卡尔曼预测模型所得到的预测值与实际观测值呈强正相关。在此基础上, 分别将预测值与实际观测值样本在 SPSS 中做 Wilcoxon 带符号秩检验, 通过分析配对样本, 验证基于卡尔曼滤波预测的重金属污染含量和实际监测值的一致性。如表 1 所示, 发现在剔除一些明显的异常值后, 监测得到的重金属污染含量值双侧渐进显著性概率为 0.802, 大于显著性水平 $\alpha = 0.05$ [12], 即认为重金属 Pb 含量的预测值和监测值没有明显差异, 确定了卡尔曼预测模型在重金属指标预测方面的可用性和可靠性。

Table 1. Wilcoxon test statistical results graph
表 1. Wilcoxon 检验统计结果图

	卡尔曼预测值 - 真实检测值
z	-0.250 ^a
渐近显著性(双尾)	0.802

(a. 基于正秩)。

5. 卡尔曼插值模型应用结果

根据贵阳市花溪区地理经纬度位置(东经 106°27'~106°52', 北纬 26°11'~26°34'), 选取花溪区行政区域内的地理空间进行研究。应用本文的重金属污染指标的卡尔曼插值预测模型, 通过 ArcGIS 工具进行空间插值, 绘制出了贵阳市花溪区重金属 Pb 污染的插值模拟图, 模拟重金属的空间分布及局部污染特征, 并根据现有阶段的污染状况预测未知区域的重金属 Pb 平均污染浓度, 这使得花溪区的区域面内每个经纬网格上都可以直观地体现出重金属的污染浓度特征值。

由图 5(a), 图 5(b)可以看出贵阳市花溪区重金属 Pb 的污染浓度空间分布特征与污染浓度的频数分布范围。整个贵阳市花溪区重金属 Pb 污染的平均浓度主要分布在 4.98~30.26 mg/kg 之间, 结合当地经济发展状况分析可得: 人类活动频繁, 车流量大, 工厂密集的区域重金属污染浓度较高; 靠近花溪国家城市湿地公园, 受工业污染程度较低的区域重金属污染浓度较低。总的来说, 贵阳市花溪区的重金属含量整体低于贵州省土壤背景值, 区域内水土环境质量较为优良。

6. 结语

本文是 Django 架构在城市规划领域的一次深入研究与应用, 拓展了国内 Django 的应用领域, 文中有关城市规划的数据可视化平台设计与实现的方法及研究思路能给贵阳市的城市规划提供帮助。特别是

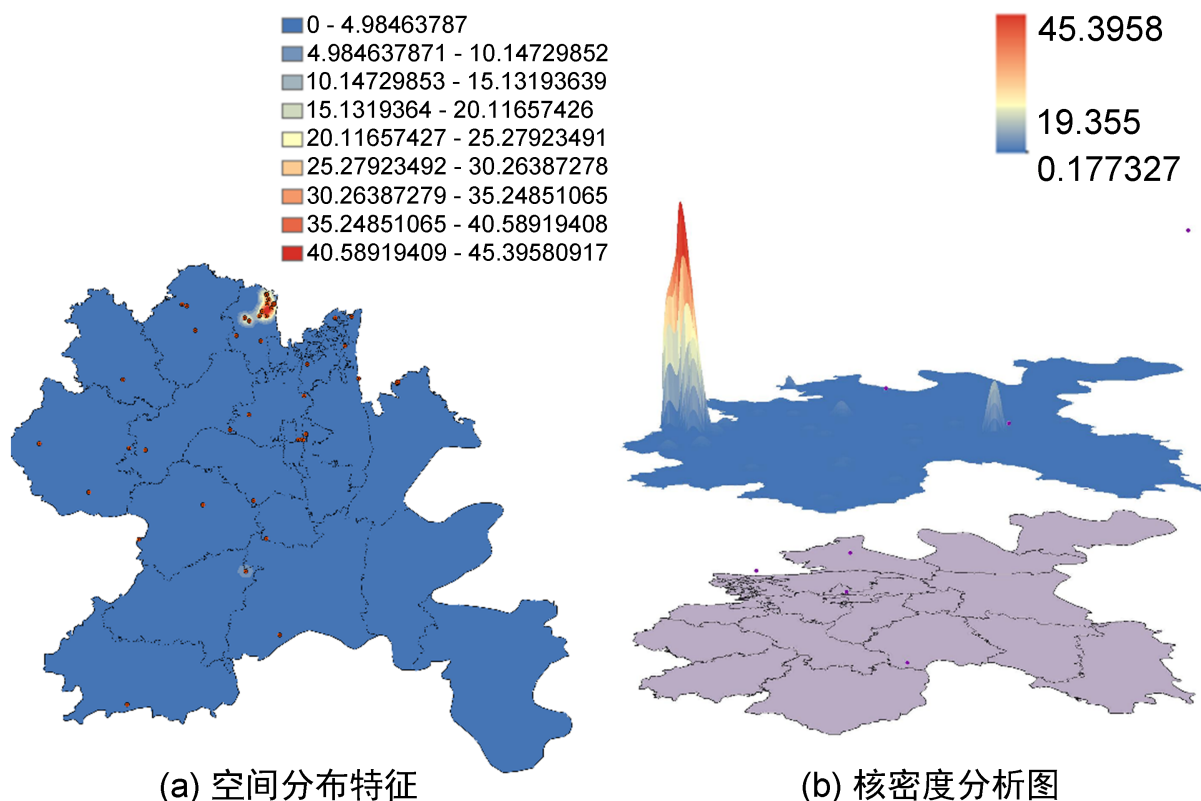


Figure 5. Spatial distribution characteristics and kernel density analysis of heavy metal Pb pollution in Huaxi District
图 5. 花溪区重金属 Pb 污染空间分布特征及核密度分析

在重金属指标预测方面, 本文基于卡尔曼滤波并通过递归的方式进行重金属指标预测, 并重点研究了卡尔曼预测模型在重金属污染预测方面的应用, 为重金属污染预测提供了一种新的方法。结果表明: 卡尔曼插值预测模型在重金属污染预测方面是可行的, 其预测结果的显著性概率为 0.802, 具有较好的预测效果, 并能较好地拟合重金属的空间分布情况。但文中仅能获得贵阳市十里河滩区域在 2021 年 11 月份当中的 28 天内重金属 Pb 的平均含量监测, 其时间与空间涵盖的广度不足, 所拟合的预测模型并不能很好地推广到整个贵阳市区域内。希望在后期能与有关部门展开更多的合作, 实现对整个贵阳市重金属污染的预测。

基金项目

贵州大学大学生创新创业项目资助(S202010657052);
 贵州大学高等教育研究项目(GDGJYT2020011)。

参考文献

- [1] 李嘉蕊. 基于土壤-作物-人体系统的耕地重金属污染评价和健康风险评估[D]: [硕士学位论文]. 杭州: 浙江大学, 2019.
- [2] 孙少龙, 蔡宏强, 王冲. 城市表层土壤重金属污染分析模型[J]. 数学理论与应用, 2012, 32(1): 71-76.
- [3] 胡宁静, 骆永明. 土壤污染物临界负荷研究进展[J]. 土壤, 2006(5): 552-558.
- [4] 姚琳. 杭州市大气颗粒物重金属污染特征与防治对策研究[Z]. 杭州: 杭州市环境监测中心站, 2013-12-28.
- [5] 孙建军, 李琪, 吕强. 浅析 Web 开发工具 Django 的 MVC 架构[J]. 品牌与标准化, 2021(6): 105-106+109.
- [6] 刘艳雯, 吴涛, 沈彬, 杨金栋, 庞达. 基于 Django 的施工数据共享平台设计与实现[J]. 现代计算机, 2022, 28(2):

117-120.

- [7] 高琳琳, 周晓天, 孟源思, 杨庆波, 李丁, 马友华. 农用地土壤重金属污染风险评价标准与风险划分[J]. 安徽农学通报, 2021, 27(8): 122-125+135.
- [8] 王伟, 郑斌, 陈彬林, 安耀明, 姜小明, 李章勇. 基于微区域 PM2.5 浓度卡尔曼插值预测模型的研究[J]. 生物医学工程学杂志, 2018, 35(1): 64-69.
- [9] Aspeel, A., Legay, A., Jungers, R.M. and Macq, B. (2021) Optimal Measurement Budget Allocation for Kalman Prediction over a Finite Time Horizon by Genetic Algorithms. *EURASIP Journal on Advances in Signal Processing*, **2021**, Article No. 39.
- [10] 林旭梅, 刘帅, 石智梁. 基于自适应卡尔曼滤波的多传感器信号降噪[J]. 计算机仿真, 2022, 39(2): 507-511.
- [11] 刘子军. 基于 Pearson 相关系数的低渗透砂岩油藏重复压裂井优选方法[J]. 油气地质与采收率, 2022, 29(2): 140-144.
- [12] 林跃东. 基于威尔科克森符号秩检验用能权核查数据分析[J]. 化学工程与装备, 2021(9): 187-189.