

# 空间统计学在食品污染物分布研究中的应用

王欣梅<sup>1</sup> 肖革新<sup>2</sup> 梁进军<sup>1</sup> 郭丽霞<sup>3</sup> 刘杨<sup>3,4</sup>

<sup>1</sup>湖南省疾病预防控制中心食品安全监测科,长沙 410005; <sup>2</sup>国家食品安全风险评估中心风险监测部,北京 100022; <sup>3</sup>国家食品安全风险评估中心风险交流部,北京 100022;

<sup>4</sup>贵州科学院博士后工作站,贵阳 550001

通信作者:肖革新, Email:xiaogexin@cfsa.net.cn

**【摘要】** 目的 以2017年某省食品安全监测大米中砷含量数据为例,探讨空间统计学方法在食品污染物分析中的应用价值。**方法** 采用空间点模式估计、核密度分析,全局以及局部自相关性分析等空间统计学方法,在县级空间尺度下,对某省大米中砷含量进行探索性空间数据分析。**结果** 空间点模式分布图显示,该省大米砷污染的空间分布比较分散,核密度分析结果显示污染热点区域主要在该省中东部地区。全局自相关Moran's *I*指数值为0.11,有统计学意义,大米样品中砷污染呈现出低度空间聚集性。有1个“高-高”聚集区,2个典型的“低-低”聚集区。**结论** 空间统计学运用于食物污染物分布研究上,可以很好地可视化展示、识别污染分布规律、热点地区和聚集区,为基于问题的监测工作的开展提供技术支持。

**【关键词】** 空间统计学; 大米; 砷

**基金项目:**国家重点研发项目(2017YFC1602002)

DOI: 10.3760/cma.j.issn.0254-6450.2019.02.022

## Application of spatial statistics in studying the distribution of food contamination

Wang Xinmei<sup>1</sup>, Xiao Gexin<sup>2</sup>, Liang Jinjun<sup>1</sup>, Guo Lixia<sup>3</sup>, Liu Yang<sup>3,4</sup>

<sup>1</sup>Food Safety Monitoring Section, Hunan Provincial Center for Disease Control and Prevention, Changsha 410005, China; <sup>2</sup>Risk Monitoring Department, China National Center for Food Safety Risk Assessment, Beijing 100022, China; <sup>3</sup>Risk Communication Department, China National Center for Food Safety Risk Assessment, Beijing 100022, China; <sup>4</sup>Post-doctoral Station, Guizhou Academy of Sciences, Guiyang 550001, China

Corresponding author: Xiao Gexin, Email: xiaogexin@cfsa.net.cn

**【Abstract】 Objective** Based on data related to arsenic contents in paddy rice, as part of the food safety monitoring programs in 2017, to discuss and explore the application of spatial analysis used for food safety risk assessment. **Methods** One province was chosen to study the spatial visualization, spatial point model estimation, and kernel density estimation. Moran's *I* statistic of spatial autocorrelation methods was used to analyze the spatial distribution at the county level. **Results** Data concerning the spatial point model estimation showed that the spatial distribution of pollution appeared relatively dispersive. From the kernel density estimation, we found that the hot spots of pollution were mainly located in the central and eastern regions. The global Moran's *I* values appeared as 0.11 which presented low spatial aggregation to the rice arsenic contamination and with statistically significant differences. One “high-high” and two typical “low-low” clustering were seen in this study. **Conclusion** Results from our study provided good visual demonstration, identification of pollution distribution rules, hot spots and aggregation areas for research on the distribution of food pollutants. Spatial statistics can provide technical support for the implementation of issue-based monitoring programs.

**【Key words】** Spatial statistics; Rice; Arsenic

**Fund program:** National Key Research and Development Program of China (2017YFC1602002)

DOI: 10.3760/cma.j.issn.0254-6450.2019.02.022

建立健全食品污染物监测体系是食品安全管理的核心环节,有助于系统了解国家或地区层面食品

安全的整体状况及变化趋势,为基于风险的食品安全标准的制修订提供科学数据,为降低食品安全和

公共卫生风险提供科学信息<sup>[1]</sup>。近年来随着信息化与食品安全的深度融合,我国食品安全数据呈现出体量大、传播快、种类多、价值高等典型的大数据4V (volume, velocity, variety, value)特征<sup>[2]</sup>,如何从这些碎片化的数据资源中提炼出有效信息加以分析利用,更好地服务于食品安全监管与人民健康保障是目前食品安全监测信息收集部门面临的最大挑战<sup>[3]</sup>。

空间统计学是以地理实体为研究对象,以空间统计学模型为工具,将非结构化的空间数据和结构化的属性数据有机结合用于分析空间格局、关系、时空变化规律,进而揭示成因,还可有效补充经典统计学的时空尺度单一、污染状况的呈现缺乏空间直观感受的缺点<sup>[4-6]</sup>,已经广泛应用于地理地质<sup>[7-8]</sup>、生态环境<sup>[9-11]</sup>、生产生活等诸多领域的研究<sup>[12-13]</sup>,并逐步向医疗卫生行业渗透融合,已应用于疾病空间分布规律和特征、时空预警、病因探索等方面<sup>[14-15]</sup>,而在食品中污染物分布的空间研究上应用较少。因此为充分了解食品污染物的空间分布模式,基于问题有的放矢的展开监测,本研究基于空间统计学分析方法,以某省大米中砷污染状况分析为例,从点模式分析(空间点模式分析和核密度分析)和面状数据分析(空间自相关分析)两个层次分析该省整体和局部的污染状况和分布规律,展示空间分析在食品安全风险监测工作中的应用。

## 资料与方法

1. 资料来源:本研究数据来自某省2017年开展的食品安全风险专项调查工作中收集的种植环节的大米样品及其砷含量检测信息,共4 840条数据。

2. 数据分析方法:应用SPSS 18.0和Arcgis 10.2软件对2017年某省大米中的砷含量分布进行描述性和探索性分析。

(1)地理编码:地理编码指按照一定的规则赋予个体地域唯一、可识别的编码,建立个体地域与标准地址、空间坐标的映射关系,从而可将地址与空间坐标进行自动转换<sup>[16]</sup>。本研究以监测样品数据中食品样品的采样地点信息为研究对象,根据样品中采样地点的详细地址,通过调用百度地图开放平台的JavaScript API接口,将文本地址批量转换为经纬度坐标。之后将经纬度数据导入某省电子地图的数据库中,建立某省大米样品信息的地理信息数据库,采用Arcgis 10.2软件进行地图绘制,实现样品采样地点的点空间位置信息可视化描述。

(2)核密度分析:核密度估计法是一种非参数估

计,假定探测变量在空间点密度大的区域含量高,在空间点密度低的区域含量低。其原理是以 $P$ 点为圆心, $r$ 为半径,统计以 $P$ 、 $r$ 为圆的范围内探测变量含量值,并除以圆的面积。一般以 $P$ 点为圆心的探测变量的核密度为:

$$P(x) = \frac{1}{nh} \sum_{i=1}^n \left\{ K \left[ \frac{d(x, x_i)}{h} \right] \right\}$$

式中, $n$ 表示距离尺度范围内所包含的探测变量的数量; $K$ 为核密度函数; $h$ 表示带宽,即核密度估计法的尺度; $d(x, x_i)$ 表示两点之间的欧式距离。

核密度估计法可在Arcgis平台的密度分析工具中实现。该工具有两个重要的参数,分别为Population字段和搜索半径。Population字段探测变量在特定地点的含量,即在整个区域内的权重。

(3)全局空间自相关分析:全局空间自相关分析是从整体上反映探测变量在研究区域有无空间分布聚集性,常用Moran's  $I$ 指数来衡量,然后通过 $Z$ 检验验证是否存在聚集性。

Moran's  $I$ 指数计算公式:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\left( \sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \sum_{i=1}^n (x_i - \bar{x})^2}$$

其中, $n$ 表示研究对象的值空间的区域数; $x_i$ 表示第 $i$ 个区域内的属性值, $x_j$ 表示第 $j$ 个区域内的属性值, $\bar{x}$ 表示所研究区域的属性值的平均值; $W_{ij}$ 表示空间权重矩阵,一般为对称矩阵。 $Z$ 检验公式: $Z = \frac{I - \mu}{\sigma}$ ,

其中, $\mu$ 为999次随机分布 $I$ 的均值; $\sigma$ 为标准差。只有当 $|Z| > 1.96$ 、 $P < 0.05$ 时,表明研究对象的存在显著的空间自相关,此时Moran's  $I > 0$ ,表明所研究区域存在空间正相关,且取值越接近1,表明空间正自相关性越强,研究对象呈聚集分布;当Moran's  $I < 0$ 时,表明所研究区域存在空间负相关,取值越接近-1,表明空间负自相关性越强,研究对象呈离散分布;当Moran's  $I = 0$ ,研究对象呈随机分布,不存在自相关性。

(4)局部自相关分析:局部自相关分析是为了识别探测变量在研究区域的局部聚集区域位置及聚集类型状态,常用局部Moran's  $I$ 指数(LISA值)来衡量每个区域与空间相邻近观察值之间的聚集性,然后通过 $Z$ 检验来验证其显著性。

对于第 $i$ 个区域单元来说,Moran's  $I$ 的LISA定义:

$$I_i = \frac{x_i - \bar{x}}{S^2} \sum_{j=1}^n w_{ij}(x_j - \bar{x}), \text{且 } i \neq j$$

其中,  $n$  表示研究对象的值空间的区域数,  $x_i$  表示第  $i$  个区域内的属性值,  $x_j$  表示第  $j$  个区域内的属性值,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  表示所研究区域属性值的平均值,  $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  表示标准差的平方,  $w_{ij}$  为空间权重矩阵。

同样利用  $Z$  检验对 Moran's  $I$  的 LISA 统计量 (每个  $I_i$  都要做  $Z$  检验) 进行假设检验, 同全局空间自相关分析类似, 只有当  $|Z| > 1.96$  时,  $P < 0.05$ , 认为研究对象的值存在显著的空间自相关, 此时  $LISA > 0$ , 表明研究区域局部空间单元与相邻空间单元之间存在空间正相关, 表现为“高-高”或“低-低”聚集; 当  $LISA < 0$  时, 研究区域局部空间单元与相邻空间单元之间存在空间负相关, 表现为“低-高”或“高-低”聚集。

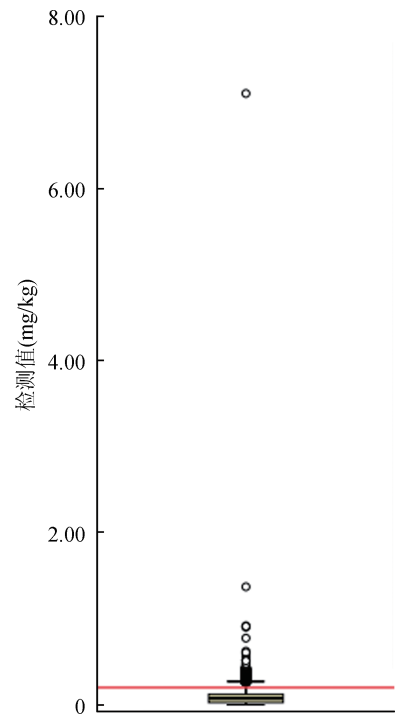
## 结 果

1. 基本概况: 目前大米尚无砷相关的国家标准, 本研究依据 GB 2762—2017《食品中污染物限量》对大米中的无机砷的限量标准(0.2 mg/kg)进行评价。2017年某省监测的 4 840 份大米样品中, 砷元素检出率为 85.33%, 平均值为 0.08 mg/kg, 最大值为 7.10 mg/kg(图1), 211 份大米的砷含量  $> 0.2$  mg/kg, 超标率为 4.36%。从大米中砷含量分布箱式图看, 数据呈现“右偏态”, 以“未超标”数据为主, 存在部分离群值(图2A), 将检测值进行对数转换后, 数据分布接近正态分布(图2B)。

2. 地理编码结果: 对 2017 年某省食品安全监测数据中的大米样品的采样地点进行地理编码, 绘制出采样地点专题地图(图3)。采样地点基本覆盖全省, 分布较为均匀。

3. 空间点模式分析: 空间点模式分布图显示 2017 年该省砷含量超标的大米样品的空间分布比较分散, 主要集中在北部的 AX 县、WL 县、中部的 XX 县、SS 县、XAT 县、YT 县、SF 县; 南部的 YZ 县、DX 县; 西部的 HT 县、HC 县、MY 县、ZJ 县、YUL 县等, 见图4。

4. 核密度分析: 根据 2017 年某省大米样品中的砷含量和分布的空间位置, 用核密度分析法分析了默认搜索半径下的污染物空间分布情况, 探测污染热点地区。2017 年大米砷污染热点区域主要在



注: 红线为 GB 2762—2017《食品中污染物限量》中对大米中的无机砷的限量标准(0.2 mg/kg)。

图1 2017年某省大米中砷含量箱式图

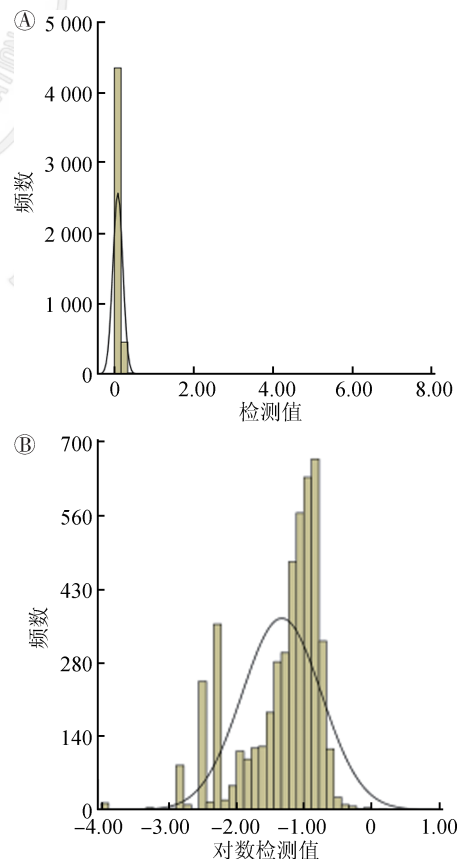


图2 2017年某省大米中砷含量频数分布

该省中东部地区(北起 KF 县, 南到 TYQ 县, 西到 YH 县, 东到 SFQ 县), 此外, TJ 县与 HSQ 县交接区、



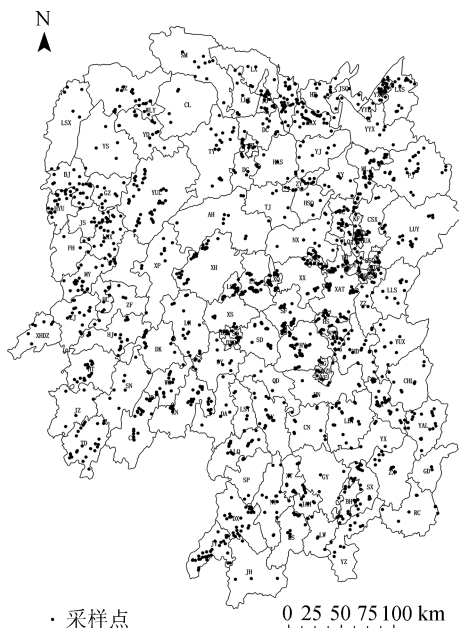


图3 2017年某省大米采样地点的空间分布

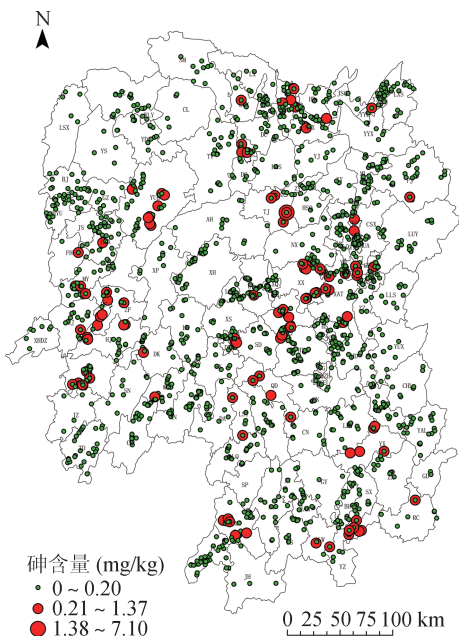


图4 2017年某省大米中砷含量的空间点模式分布

LSX县、YS县和BJ县区域,HN县、LEY县和CN县区域形成2个典型的“低-低”聚集区域,灰色地区表示无聚集性(图7)。

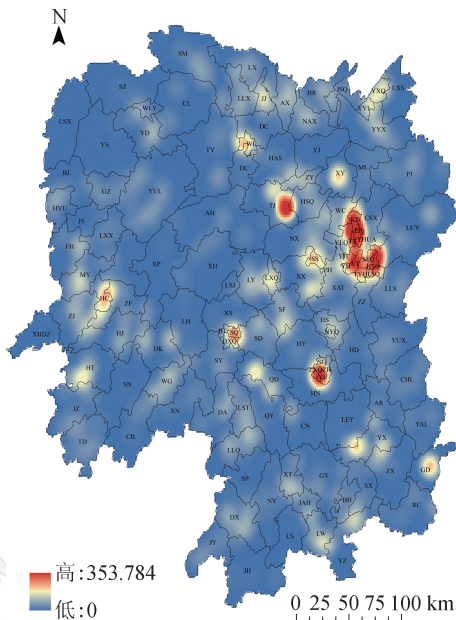


图5 2017年某省大米中砷含量的核密度分析

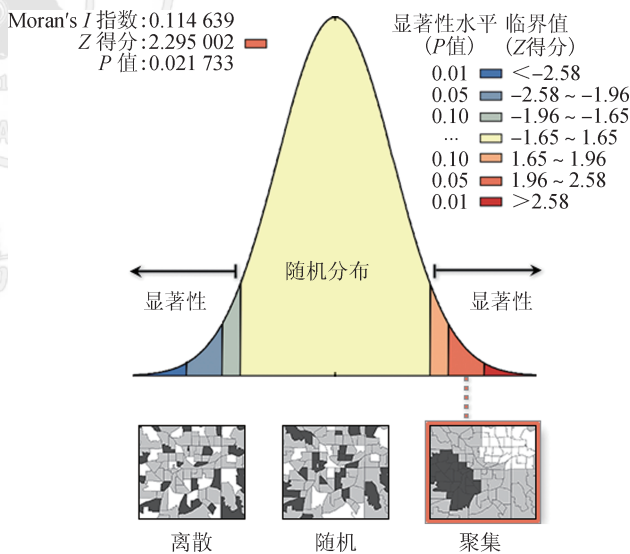


图6 2017年某省大米中砷含量全局自相关性

ZXQ县、ZH县和YF县交接区也存在大米砷污染高风险(图5)。

5. 空间自相关分析:2017年某省大米中砷含量的全局自相关Moran's *I*指数值为0.11,  $P < 0.05$ ,且  $Z > 1.96$ ,表明该省的大米样品中砷含量呈现出低度空间聚集性(图6)。

通过计算LISA值发现,HC县形成一个典型的“高-高”聚集区域,为大米中砷污染的热点区域;

### 讨 论

我国自1981年加入全球环境监测系统食品项目(global environmental monitoring system, GEMS Food)组织,1992年开始进行食品中污染物部分数据的收集,到2000年已经实现全国范围内连续性、规模化的食品中污染物监测体系<sup>[17]</sup>。既往研究中多是利用传统的统计学方法对食品中的污染物进行单一维度属性的数字化分析展示,比如按食品种类、采



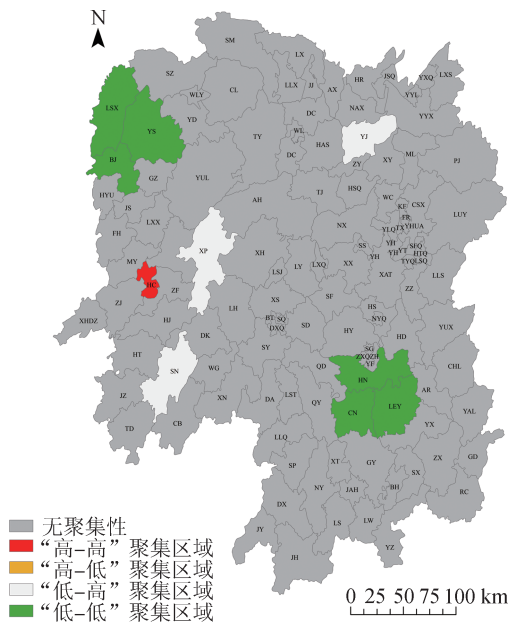


图7 2017年某省大米中砷含量局部自相关性

样时间和地点等分类,用直方图、折线图和三线表等描述污染物分布态势<sup>[18]</sup>。对数据进一步挖掘性分析变量间的差异性、相关性、聚集性,甚至污染溯源时,也多用 $P$ 值等平面数据来表现,针对的时空尺度较为单一,污染状况的呈现缺乏空间直观感受<sup>[5-6, 19]</sup>。此外,从“农田到餐桌”的各个环节数据都携带着空间地理属性,而经典统计学方法侧重于单一属性的随机变量,往往会湮没空间数据的区域化特征,引起偏移<sup>[20]</sup>。

本研究首先通过频数分布图和箱式图初步描述了2017年某省大米中砷含量的总体水平、数据的分布类型以及离群值情况;运用空间统计学方法,从点、面两个层面多维度分析该省大米中砷含量的空间分布。空间点模式分析结果显示,大米中砷污染的空间分布相对分散,在该省中、南、西部均有超标位点;进行核密度点模式分析发现,中、东部地区的3个热点地区存在大米砷污染高风险;进一步采用空间自相关分析定量揭示空间数据的空间关联类型和高危聚集地区,从面状空间模式分析结果看,该省大米砷污染存在低度空间聚集性,其中HC县呈现“高-高”聚集现象,在核密度分析图中也显示该地区为污染高风险地区,提示应针对这些地区开展专题调查,一方面要核实污染状况,调查影响因素,进行污染溯源,采取相应措施进行污染修复,控制甚至切断污染源,另一方面可以结合本地膳食消费量开展人群健康风险评估,了解当前污染状况下的人群暴露危害大小,尤其是小孩和老人等敏感人群。

在“互联网+”的大环境下,不断完善食品污染物监测网络,搭建数据收集云平台的同时,更重要的是提取、分析和利用碎片化信息,由问题型监测向服务型监测转变。空间统计学方法不仅能集成空间属性实现分析结果的可视化展示,还能有效补充和完善传统统计学方法无法有效解决空间样本点的选取、空间估值和空间数据间的关系等问题,挖掘食品污染物的空间关系和分布规律,探索其成因与发展,进行跟踪追溯管理,提示区域性食品安全风险,为政府精准治理提供科学依据<sup>[3]</sup>。

综上所述,本研究以2017年某省大米中砷含量数据为例,应用宏观与微观的空间统计学方法,从点数据和面状数据两个层面进行污染物含量分析和探测高风险地区,为今后食品污染物监测相关分析提供方法学上的借鉴。

利益冲突 所有作者均声明不存在利益冲突

### 参 考 文 献

- [1] 付文丽,陶婉亭,李宁,等. 借鉴国际经验完善我国食品安全风险监测制度的探讨[J]. 中国食品卫生杂志, 2015, 27(3): 271-276. DOI: 10.13590/j.cjfh.2015.03.011.  
Fu WL, Tao WT, Li N, et al. Improve the food safety risk monitoring system of China from international experience [J]. Chin J Food Hyg, 2015, 27(3): 271-276. DOI: 10.13590/j.cjfh.2015.03.011.
- [2] 肖革新,肖辉. 基于空间统计的食品安全风险研究现状与展望[J]. 中国食品卫生杂志, 2016, 28(4): 409-414. DOI: 10.13590/j.cjfh.2016.04.001.  
Xiao GX, Xiao H. Current status and prospect of spatial statistics in food safety[J]. Chin J Food Hyg, 2016, 28(4): 409-414. DOI: 10.13590/j.cjfh.2016.04.001.
- [3] 肖辉,任鹏程,肖革新,等. 食品安全健康大数据平台构建[J]. 医学信息学杂志, 2016, 37(5): 28-31. DOI: 10.3969/j.issn.1673-6036.2016.05.006.  
Xiao H, Ren PC, Xiao GX, et al. Construction of the platform for big data of food safety and health[J]. J Med Intell, 2016, 37(5): 28-31. DOI: 10.3969/j.issn.1673-6036.2016.05.006.
- [4] 肖革新. 空间统计实战[M]. 北京: 科学出版社, 2018.  
Xiao GX. Operational space statistics[M]. Beijing: Science Press, 2018.
- [5] 王艳莉,郭宝福,祝白春,等. 2011—2015年南京市食品中化学污染物及有害因素监测分析[J]. 职业与健康, 2016, 32(23): 3225-3229. DOI: 10.13329/j.cnki.zyyjk.2016.1012.  
Wang YL, Guo BF, Zhu BC, et al. Monitoring and analysis on chemical pollutants and harmful factors in foods in Nanjing city from 2011-2015[J]. Occupat Health, 2016, 32(23): 3225-3229. DOI: 10.13329/j.cnki.zyyjk.2016.1012.
- [6] Paltseva A, Cheng ZQ, Deeb M, et al. Accumulation of arsenic and lead in garden-grown vegetables: factors and mitigation strategies[J]. Sci Total Environ, 2018, 640/641: 273-283. DOI: 10.1016/j.scitotenv.2018.05.296.
- [7] Thanh NT. 区域地球化学元素空间分布模式研究[D]. 北京: 中国地质大学, 2014.  
Thanh NT. Study on spatial distribution patterns of regional

