

基于模糊集理论的土壤重金属污染空间预测^{*}

檀满枝^{1,2} 陈杰^{1†} 徐方明³ 郑海龙⁴ 张学雷¹

(1 土壤与农业可持续发展国家重点实验室(中国科学院南京土壤研究所), 南京 210008)

(2 中国科学院研究生院, 北京 100039)

(3 成都理工大学核技术及自动化学院, 成都 610008)

(4 香港中文大学地理与资源管理系, 香港)

摘要 基于 54 个土壤表层样品重金属全量浓度实验室测定数据, 应用模糊 ϵ -均值算法对南京城市边缘带化工园附近 20 km² 样区内土壤重金属浓度进行了连续分类, 对样点土壤的隶属度进行空间普通克里格插值, 实现样区土壤重金属浓度和污染状况的空间预测。结果表明, 样区土壤中重金属 Cu、Zn、Cr、As 和 Cd 的浓度均低于国家土壤环境质量标准, 未发生污染; 受工业、交通排放的影响, 样区左侧和南部 Cu、Zn、Cr、As 较高, 个别区域有 Zn、Cr 富集现象。Hg 为样区主要的土壤重金属污染元素, 土壤 Hg 污染主要发生在蔬菜基地及滁河流经区, 较为严重的 Hg 污染土壤集中分布在样区中部的蔬菜基地。与利用样点测定数据直接插值的空间预测方法相比, 基于模糊集理论的土壤污染空间预测方法可获得较好的预测效果。

关键词 模糊集理论; 模糊 ϵ -均值算法; 连续分类; 空间预测; 土壤污染; 重金属

中图分类号 X131.3; O235

文献标识码 A

由各种有害物质进入而导致的土壤污染已经成为全球性的环境问题。重金属在土壤中的积累, 不仅直接影响土壤理化性状、降低土壤生物活性、阻碍养分有效供给, 而且通过植物吸收、淋洗作用进入食物链和地下水, 对生态环境和人类健康构成严重威胁。20 世纪 70 年代以来, 土壤重金属污染日益受到相关学科领域与国际社会的广泛关注^[1]。土壤重金属污染防治与修复逐渐发展为农业资源保护、区域生态环境建设以及食品安全战略的重要组成部分。由于土壤重金属污染在空间上具有分布复杂性、高度可变性以及局部偶然性特征, 准确划定防治区域、科学制订修复方案面临巨大挑战。鉴于此, 拓展土壤重金属污染空间预测(Spatial prediction)理论、探索和改进空间预测技术手段, 具有重要的科学与现实意义。

作为一种土壤空间预测的随机性模型(Stochastic models), 以区域化变量(Regionalized variable)理论为基础的地统计学方法于 20 世纪 80 年代开始广泛应用于土壤空间预测。起初用于揭示单一土壤属性的空间变异, 随之拓展到土壤属性多变量以及土壤

变异多元指数研究^[2], 进而被应用于定性变量(Qualitative variables)和土壤空间变异的多尺度、嵌套尺度研究^[3-5]。地统计学方法的应用必须以获取研究区内土壤对象的变异函数或半方差函数为前提, 并以此为基础进行样区设计和确定采样密度。由于土壤重金属浓度通常具有较高的变异系数, 确定其半方差函数非常困难, 因此导致基于样点数据的插值标准差很高, 土壤重金属污染的空间预测结果具有很大的不确定性。相对于地统计学模型, 基于模糊逻辑(Fuzzy logic)或模糊集理论(Fuzzy set theory)的随机模型, 即模糊分类(Fuzzy classification)在土壤空间预测领域内的应用要晚得多^[6]。在土壤重金属污染空间预测过程中, 地统计学方法通过无偏估计(Unbiased estimates)插值分析输出未测样点土壤的重金属平均预测浓度或超出特定阈值的概率^[7-9]; 而模糊分类方法获取的空间预测结果不再是非此即彼的概率或可能性(Likelihood)陈述, 而是明确表示为未测土壤重金属浓度属性对于多个浓度类别的多重相似性或定量隶属关系。基于模糊分类方法产生的土壤重金属浓度类别是连续的(Continuous classification),

* 国家自然科学基金项目(40571065)、中国科学院知识创新工程重要方向(KZCX3-SW-427)和国家自然科学基金项目(40235054)资助

† 通讯作者, E-mail: jchen@issas.ac.cn

作者简介: 檀满枝(1978~), 女, 安徽望江人。助理研究员, 在职博士研究生。主要从事土壤资源、信息系统研究。已发表论文 10 余篇

收稿日期: 2005-10-17; 收到修改稿日期: 2006-01-06

不仅从基础上解决了土壤重金属浓度在地理空间上的渐变问题,而且能够解决地统计方法无法处理的土壤重金属点源污染的边界突变问题^[1,10]。

实现模糊连续分类的方法主要有两种,即模糊语义输入(Fuzzy semantic input, SI)和模糊聚类(Fuzzy clustering)。其中模糊聚类为无人监督和管理的自动分类方法,模糊 c -均值算法(Fuzzy c -means algorithm, FCM)或模糊 k -均值聚类(一些文献中也被称作模糊 k -均值算法)是土壤科学中应用最广泛的模糊分类方法^[9, 11]。本研究以南京城市边缘带化学工业园邻近的土壤重金属污染调查样区为例,系统介绍模糊 c -均值聚类,即模糊多变量连续分类方法在土壤重金属浓度分布空间预测中的应用,并结合地统计学手段,对样区土壤重金属污染状况进行评价。

1 材料与方法

1.1 样区概况与样品采集

样区位于南京市六合区境内,为以南京化工园

区为中心,向北半径9 km、张角40°的扇形区域,面积约20 km²。样区内土地利用形式多样且变化快速、地表覆盖结构复杂;全区地貌为沿江平原圩区,河流、水面分布较多。本区土壤多发育于不同时期的长江冲积、沉积母质,主要类型为潜育型水稻土下的不同土属、旱耕熟化及厚熟菜园土壤。样区土壤资源利用形式为传统的蔬菜和水稻种植,目前仍为南京市主要的粮食与蔬菜生产基地之一。受土壤资源集约农业利用及周边工业、交通三废物质排放的影响,样区土壤环境质量近年来下降趋势明显。

基于样区 1:50 000 地形图、1:100 000 土壤图及 1:5 000 航空照片,对土壤样点采用复区格网布置,采样过程中根据微域地形、土地利用与覆盖状况、土壤类型空间分布等具体情况对预设样点微调,实际样点以GPS精确定位。研究区共采集土壤样品54个,其中常规样点31个,平均样点间距800 m左右;重点区域(图1中样点密集区)内23个,平均样点间距150 m。土壤样品为0~20 cm 表层土壤,为5 m 对角线5点混合四分法取样。

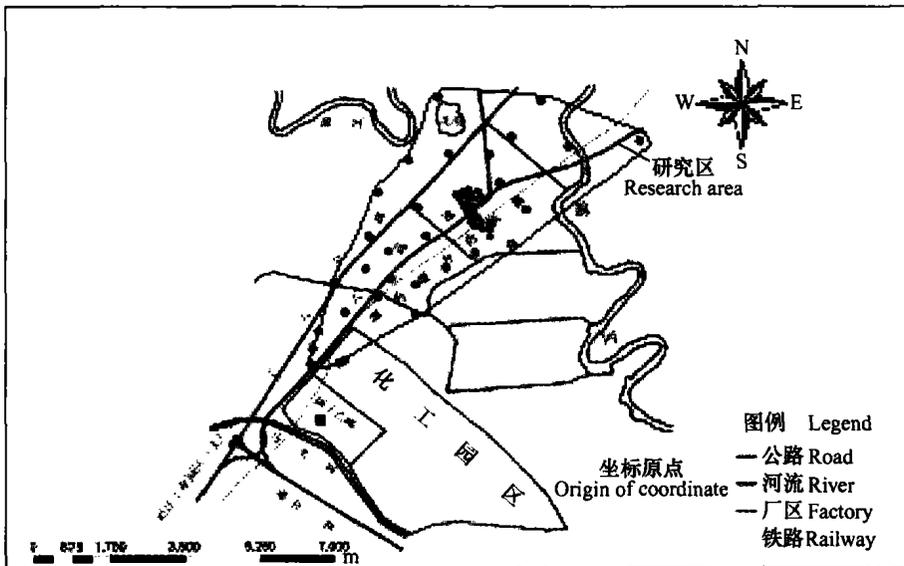


图1 样区位置与样点分布

Fig 1 Locality of the sampling plot and distribution of sampling sites

1.2 样品处理与分析方法

土壤样品在野外剔除与金属采样器械接触部分,放入内衬聚乙烯塑料袋的棉布专用样品袋,避免样品、土壤液体之间接触污染。预处理采用阴凉通风处自然风干,剔除样品中植物根系、有机残渣以及可见侵入体,用木质工具碾碎并用玛瑙研钵研磨,分别过

10目、60目和100目尼龙筛。土壤样品常规理化性状分析参照《土壤农业化学分析方法》^[12]。样品重金属浓度测定项目包括全量Cr、Zn、Cu、Hg、Cd及As,其中Cr、Zn、Cu采用X-荧光法(仪器型号为PHILIPS Magix-PRO),Hg、As采用原子荧光法(仪器型号为AFS820),Cd采用原子吸收石墨炉法(仪器型号为M6)。

1.3 土壤重金属浓度模糊分类

本研究采用模糊 c -均值算法实现样区土壤重金属浓度的模糊聚类。首先定义一个 $n \times p$ 土壤重金属浓度模糊集 X , n ($= 54$) 为空间样点数量, p 为受测重金属元素的数量(即属性变量), X_{ij} 为重金属元素 j 在样点 i 土壤中的浓度 ($i = 1 \dots, 54, j = 1 \dots, p$)。将样区内 54 个样点土壤的属性划分到 c 个重金属浓度类别中, 那么, 用隶属度函数定义的聚类损失函数表示为:

$$J_F(M, c) = \sum_{i=1}^n \sum_{k=1}^c \mu_{ik}^\varphi d_{ik}^2 \quad (1)$$

式中, μ_{ik} 为第 i 个样点的土壤重金属浓度属性对于第 k 个聚类类别的隶属度; d_{ik}^2 为第 i 个样点土壤重金属浓度属性与第 k 个类别中心之间的距离模数; φ 为一个可控制聚类结果和类别交叠程度的常数, 即模糊指数。模糊 c -均值算法要求样本对各个聚类的隶属度之和为 1, 即:

$$\sum_{k=1}^c \mu_{ik} = 1, 1 \leq i \leq n \quad (2)$$

在条件式(2)下用迭代法求解式(3)、式(4)得样本隶属度和聚类类别中心, 并满足准则函数式(1)最小化目标:

$$\mu_{ik} = \frac{d_{ik}^{-2/(\varphi-1)}}{\sum_{k=1}^c d_{ik}^{-2/(\varphi-1)}}, 1 \leq i \leq n, 1 \leq k \leq c \quad (3)$$

$$C_k = \frac{\sum_{i=1}^n \mu_{ik}^\varphi X_i}{\sum_{i=1}^n \mu_{ik}^\varphi}, 1 \leq i \leq n, 1 \leq k \leq c \quad (4)$$

当算法收敛时, 就得到了各类别的聚类中心 C_k (Centroids) 和各个样点土壤对于各聚类类别的隶属度值 μ_{ik} , 从而完成了土壤重金属浓度的模糊聚类, 即样区土壤重金属浓度属性的连续分类。

上文各式中, 模糊类别数量 c 和模糊指数 φ 两个参数的选择和确定直接影响连续分类的输出结果。其中, c 可以根据样区已知土壤重金属浓度信息或主观分类目标人为设定, 也可采用 c 、 φ 多次组合最优取值法。本研究采用后一种方法, 并用 McBratney 和 Moore 提出的模糊性能指数 (Fuzziness performance index, FPI) 对模糊聚类效果进行检验, 使 FPI 最小^[13]:

$$FPI = 1 - \frac{cF - 1}{c - 1}, F = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^2 / n \quad (5)$$

实际运算中, 采用目标式(5)相对于 φ 的派生函数 $-(\partial F / \partial \varphi) c^{0.5}$ 同时确定 c 、 φ 值: 分别对 c 取值, 使 $-(\partial F / \partial \varphi) c^{0.5}$ 峰值最小的 c 值为最优选择; 在选定 c 值的情况下, 使 $-(\partial F / \partial \varphi) c^{0.5}$ 峰值最大的 φ 值为最优选择^[13]。

1.4 连续分类输出结果的空间表达

根据模糊分类结果, 对样点土壤对于不同重金属浓度类别的部分隶属关系 (Partial membership) 进行克里格插值, 获得样区土壤重金属浓度复合隶属关系图。同时, 为进一步明确不同重金属浓度类别在各样点处的交叠程度, 即在样点土壤重金属类别安置中的不确定性 (Uncertainty), 利用混乱指数法 (Confusion index, CI) 对样点隶属关系进行去模糊处理^[14]:

$$CI = 1 - [\mu_{\max i} - \mu_{(\max-1)_i}] \quad (6)$$

式中, $\mu_{\max i}$ 为第 i 个样点土壤对于某重金属浓度类别最大的隶属度值, $\mu_{(\max-1)_i}$ 为第 i 个样点土壤第二大隶属度值。基于 CI 的计算结果, 获得样点重金属浓度类别混乱指数散点图。

2 结果与讨论

2.1 重金属含量描述性统计结果

样区表层土壤重金属浓度测定结果描述性统计分析见表 1。在全部 54 个表层土壤样品中, Cu、As、Cr、Zn 和 Cd 浓度均低于国家土壤环境质量二级标准, 且平均水平小于其土壤地球化学背景值; Cu、As、Cr 在样点土壤中的浓度甚至未超过国家一级标准。统计分析显示, 除 Hg 以外的重金属在土壤中的浓度呈对数正态分布, 样点之间的变异系数小, 表现出空间弱变异特征, 说明样区土壤整体上未发生上述重金属元素导致的单因子污染。而样点土壤中 Hg 浓度较高, 37.0% 的样点土壤 Hg 浓度超过国家二级标准, 属轻度污染; 13.0% 的样点土壤 Hg 浓度超过国家二级标准 2 倍, 属中度污染; 1.9% 的样点土壤 Hg 浓度超过国家二级标准 3 倍以上, 属于严重污染。且样区土壤 Hg 表现出强烈的空间变异属性, 浓度变异系数高达 67.4%, Hg 污染具有点源污染的显著特征。

表1 南京化工园区附近土壤重金属含量描述性统计

Table 1 Descriptive statistics of heavy metal concentrations in the soils around the Nanjing Chemical Industry Park

Table with 9 columns: Metal, Sample number, Minimum, Maximum, Average, SD, CV, Background value, First standard of national value. Rows include Cu, Zn, Cr, Cd, Hg, As.

2.2 土壤重金属浓度模糊分类输出结果

采用模糊 c-均值算法, 利用多变量组合对样点土壤重金属浓度属性进行连续分类。由于样点土壤不同重金属的浓度之间具有相关性, 模糊分类距离采用对角线距离 (Diagonal distance)。

两组土壤重金属模糊集的连续分类输出结果——各聚类类别中心和样点土壤对各聚类类别的隶属度分别列入表 2、表 3。

表2 样区土壤中两组重金属浓度模糊类别中心

Table 2 Class centroids for the Cu, Zn, Cr, As and the Cd, Hg data set (mg kg-1)

Table with 7 columns: Class centroids, Cu, Zn, Cr, As, Cd, Hg. Rows include C1, C2, C3.

表3 样点土壤对于不同重金属浓度类别的隶属度

Table 3 Partial memberships of the sampled soils to classes of heavy metal concentrations

Table with 10 columns: Sample, C1, C2, C3, C1, C2, C3. Rows include samples 1, 2, 3, 4, ..., 54.

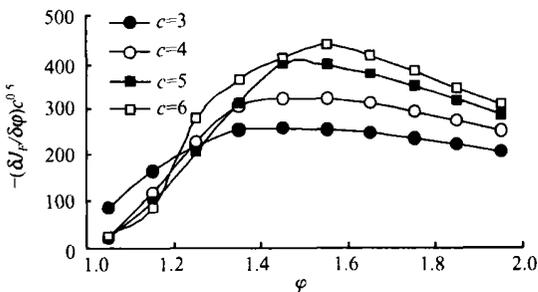


图2 不同模糊类别取值时模糊指数

phi 与 -(delta J_F / delta phi) c^0.5 关系

Fig 2 Fuzziness exponent (phi) against -(delta J_F / delta phi) c^0.5

for different class numbers of the Cu, Zn, Cr, As data set

1) Cu, Zn, Cr, As; 2) Cd, Hg; 3) CI: 混乱指数 Confusion index

从表 2 可以看出, 第一组土壤重金属元素浓度 3 个模糊类别的中心均不超过国家一级标准, 可以确定样区土壤总体上未发生 Cu、Zn、Cr、As 污染; 但 Cr 在 3 个模糊类别的中心均超过其土壤背景值, 表明样区土壤中普遍存在 Cr 富集现象; 第 3 类别中 Zn 的中心超过背景值, 说明 Zn 在隶属于 C_3 的样点土壤中富集。第二组重金属元素中, Hg、Cd 在 C_1 中的中心均不超过国家一级标准, 隶属于 C_1 的样点土壤未发生 Hg、Cd 污染, 但有 Hg 富集现象; Hg 在 C_2 、 C_3 中的中心均大大超过国家标准, 隶属于这两个模糊类别的样点土壤 Hg 污染明显; 而 Cd 在全部 3 个模糊类别的中心都低于国家一级标准和土壤背景值, 表明样区土壤未发生 Cd 污染, 也未出现 Cd 富集。与样点土壤重金属浓度描述统计分析不同, 模糊分类的输出结果不仅包含单一土壤重金属元素的浓度信息, 而且包含一组重金属元素的浓度信息, 这对解释样区土壤重金属复合污染具有重要意义。由于模糊分类将样区 54 个样点土壤划分为 3 个类别, 并通过部分隶属关系建立样点土壤与不同重金属浓度模糊类别之间在属性空间上的联系, 这种属性空间上的联系可以通过样点的地理坐标实现空间表达。譬如, 通过样点土壤的部分隶属关系, 可以在样点分布图上直观地判断样点土壤的重金属浓度和污染状况, 以及属性空间上邻近样点在地理空间上的分布状况。

2.3 空间预测及连续表达

基于样点土壤对于 3 个重金属浓度模糊类别的隶属度计算结果(表 3), 利用普通克里格方法进行插值分析, 获得未采样土壤对于不同模糊类别的隶属度, 从而实现基于模糊集理论的样区土壤重金属浓度和污染状况空间预测。其输出结果的空间连续表达如图 3。

图 3 显示, 样区土壤对于不同重金属浓度模糊类别的隶属关系可以在空间上清晰地识别, 从而有助于对不同区域土壤中受测重金属浓度和污染状况做出更加精确的判断。图 3A 中, 样区中部土壤对于重金属浓度模糊类别 C_1 的隶属度高, 表明这一区域的土壤中 Cu、Zn、Cr 和 As 的浓度最低。样区左侧和南部分别靠近铁路、公路交通线以及南京化工园区的土壤对于 C_2 的隶属度明显高于其他区域, 表明这一区域受工业、交通排放影响, 土壤中 Cu、Zn、Cr 和 As 的浓度较高。综合前文关于土壤重金属模糊类别的解释, 可以判断这一区域存在 Cr 富集的可能。样区绝大部分区域的土壤对于 C_3 的隶属度均

为 0, 表明样区土壤中 Cu、Zn、Cr 和 As 没有出现富集或富集程度较低, 只有入河口处极小面积上的土壤可能出现 Zn 和 Cr 的富集(图 1)。对照 Zn、Cr 两种重金属元素在 C_3 中的浓度中心, 可以判断出这两种重金属元素在入河口尚未形成污染。图 3B 中, 对于高浓度模糊类别的隶属度高的土壤主要分布在样区的中部蔬菜基地及西北部滁河流经区域, 表明这些区域土壤中 Hg 和 Cd 的浓度相对较高。由于全部样点土壤中 Cd 的浓度均未超过其背景值, 而 Hg 在 C_2 、 C_3 中的中心均远超过国家土壤环境质量一级标准, 因此可以判断出土壤 Hg 污染主要发生在蔬菜基地及河流流经区。对于 C_3 隶属度高的土壤主要集中在样区中部蔬菜基地, 表明蔬菜土壤是本区 Hg 污染最严重的土壤。

混乱指数(CI)表征样点模糊隶属关系的复杂程度。CI 值越大, 不同模糊类别在样点土壤上的交叠程度越高、样点土壤隶属关系不确定性越大; 反之, 则表明交叠程度低, 土壤隶属关系趋向明确, 当 $CI=0$ 时, 表明样点土壤在属性空间和地理空间上都具有清晰边界。对比图 4A、图 4B, 在重金属浓度模糊类别数量 c 和模糊指数 φ 已定的情况下, 变量数(重金属元素数量)的增加可能导致样点土壤隶属关系混乱指数增大、属性空间上的距离缩短而难以分离。而在模糊算法输入数集的变量数已定的情况下, 地理坐标相邻的样点土壤的模糊隶属关系均具有很小的混乱指数, 表明样区土壤某些属性在空间分布上具有边界突变特征。图 4B 中显示的混乱指数表明, 样点土壤对于 Hg、Cd 这两种重金属元素的模糊类别的隶属关系趋于明确, 土壤 Hg 具有比较明显的点源污染特征。

3 结 论

1) 样区土壤中重金属 Cu、Zn、Cr 和 As 的浓度具有相似的空间分布特征, 全区未发现上述 4 种重金属污染现象; 样区左侧和南部分别靠近铁路、公路交通线以及南京化工园区的区域受工业、交通排放影响, 土壤中 Cu、Zn、Cr 和 As 的浓度较高, 小面积土壤出现 Cr 富集; 而样区内滁河排灌口处的土壤可能同时出现 Zn 和 Cr 的富集。全区土壤未发现 Cd 污染和富集, 土壤 Hg 污染主要发生在蔬菜基地及滁河流经区, 较为严重的 Hg 污染土壤集中分布在样区中部的蔬菜基地。

2) 模糊 c -均值聚类结合地统计学手段可以有

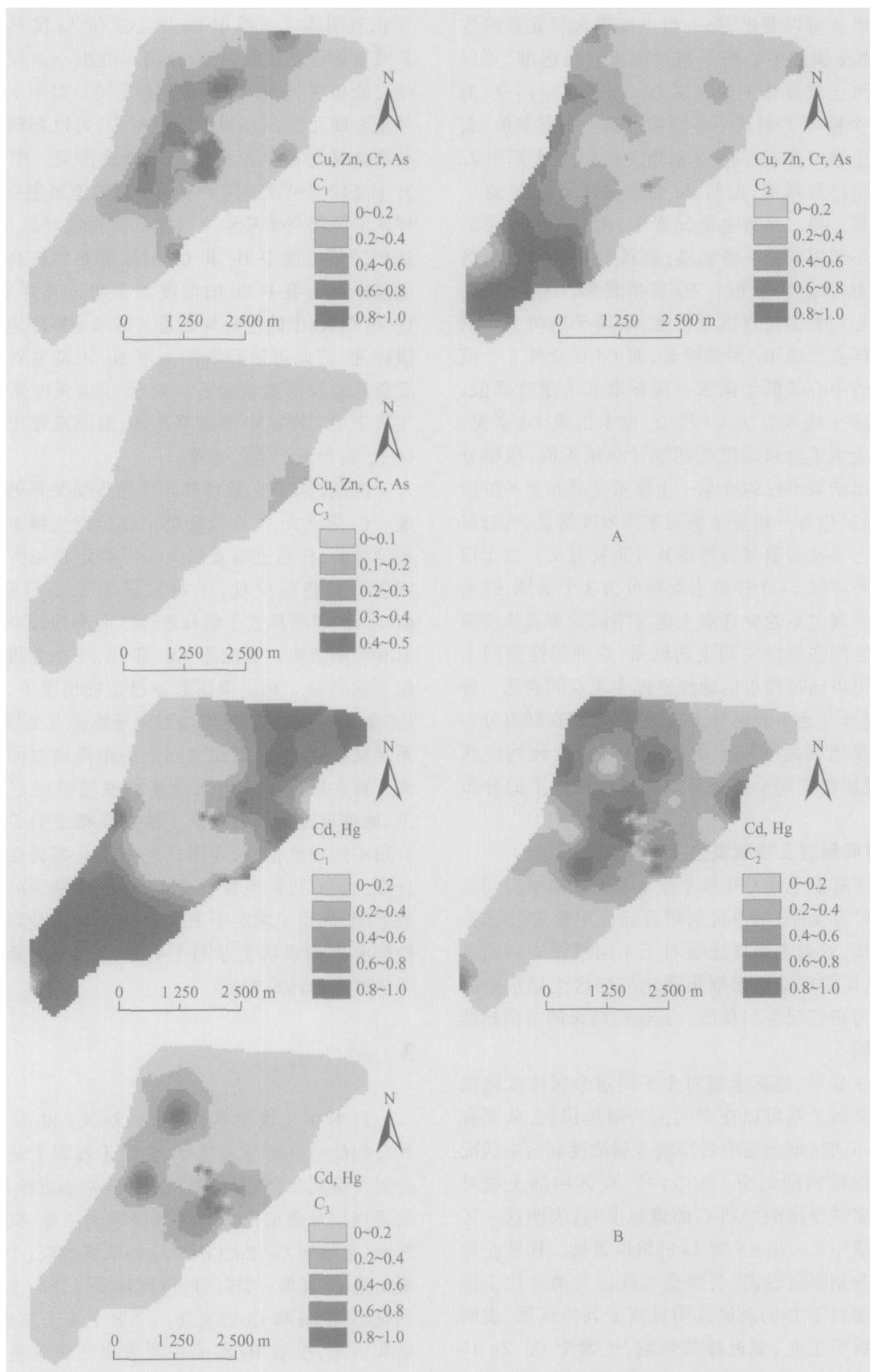


图3 样区土壤重金属浓度隶属度克里格预测

Fig 3 Kriging prediction of partial memberships of soils to individual fuzzy classes of Cu, Zn, Cr, As (A) and Cd, Hg data set (B) (C₁= class 1; C₂= class 2; C₃= class 3)

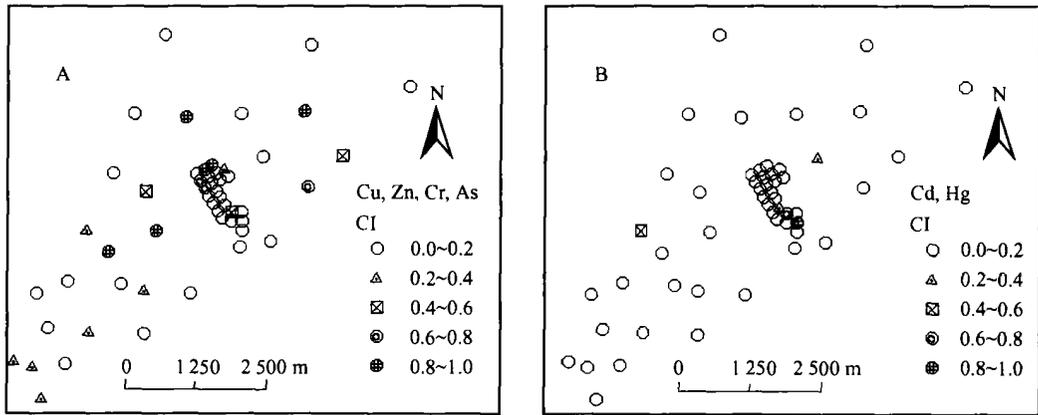


图 4 样点的混乱指数散点图

Fig 4 A scatter plot of confusion indices of soil sampling sites

效降低土壤污染空间预测的不确定性、提高预测精度、丰富样点实测数据所蕴涵的空间信息,强化了土壤空间预测的理论基础。本研究的结果表明,模糊分类通过输出样区土壤重金属浓度属性对于多个模糊浓度类别的多重相似性或定量隶属关系,从根本上解决了土壤重金属浓度在地理空间上的渐变问题,同时有效解决了地统计学方法无法处理的土壤重金属污染源污染的边界突变问题,获得了较常规地统计学空间预测更为理想的效果。

参考文献

- [1] Hendricks H J, Franssen W M, van Eijnsbergen A C, *et al.* Use of spatial prediction techniques and fuzzy classification for mapping soil pollutants. *Geodema*, 1997, 77: 243~ 262
- [2] Webster R, Oliver M A. *Statistical Methods in Soil and Land Resources Survey*. Oxford University Press, 1990
- [3] Bierkens M F P, Burrough P A. The indicator approach to categorical soil data I. Theory. *J. Soil Sci.*, 1993, 44: 361~ 368
- [4] Goovaerts P, Webster R. Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. *Eur. J. Soil Sci.*, 1994, 45: 79~ 95
- [5] Webster R, Atteia O, Dubois J P. Coregionalisation of trace metals in the soil in the Swiss Jura. *Eur. J. Soil Sci.*, 1994, 45: 205~ 218
- [6] Burrough P A. Fuzzy mathematic methods for soil survey and land evaluation. *J. Soil Sci.*, 1989, 40: 447~ 492
- [7] Goovaerts P. *Geostatistics for Natural Resource Evaluation*. New York: Oxford University Press, 1997
- [8] McBratney A B, De Grujter J J, Brus D J. Spatial prediction and mapping soil classes. *Geodema*, 1992, 54: 39~ 64
- [9] De Grujter J J, McBratney A B. A modified fuzzy *k*-means method for predictive classification. In: Bock H H. ed. *Classification and Related Methods of Data Analysis*. Amsterdam: Elsevier, 1988. 97~ 104
- [10] Amini M, Afyuni M, Fathianpour N, *et al.* Continuous soil pollution mapping using fuzzy logic and spatial interpolation. *Geodema*, 2005, 124: 223~ 233
- [11] Gorsevskii P V, Gessler P E, Jankowski P. Integrating a fuzzy *k*-means classification and a Bayesian approach for spatial prediction of landslide hazard. *J. Geograph Syst.*, 2003, 5: 223~ 251
- [12] 鲁如坤. 土壤农业化学分析方法. 北京: 中国农业科技出版社, 2000. 4. Lu R K. *Analytical Method of Soil Agro-chemistry* (In Chinese). Beijing: Chinese Agriculture Science and Technology Press, 2000. 4
- [13] McBratney A B, Moore A W. Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology*, 1985, 35: 165~ 185
- [14] Bragato G. Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain. *Geodema*, 2004, 118: 1~ 16

SPATIAL PREDICTION OF SOIL HEAVY METAL POLLUTION BASED ON FUZZY SET THEORY

Tan Manzhi^{1,2} Chen Jie^{1†} Xu Fangming³ Zheng Hailong⁴ Zhang Xuelei¹

(1 *State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China*)

(2 *Graduate School of the Chinese Academy of Sciences, Beijing 100039, China*)

(3 *College of Nucleus and Automation Technology, Chengdu University of Technology, Chengdu 610008, China*)

(4 *Department of Geography and Resource Management, the Chinese University of Hong Kong, China*)

Abstract Based on the lab data of concentrations of total Cu, Zn, Cr, Cd, Hg, and As obtained through determination of 54 topsoil samples, continuous classification of heavy metal concentrations of the sampled soils was conducted with fuzzy e -mean algorithm (FCM). And, membership values of the sampled soils in each class were interpolated using ordinary kriging to predict spatial distribution of soil heavy metal concentrations and pattern of soil heavy metal pollution of the investigated area. The obtained results show that the concentrations of Cu, Zn, Cr, As and Cd in soils in the area were below the national criteria for soil environment and even their background values, and that no accident of soil heavy metal pollution had ever occur. The concentrations of Cu, Zn, Cr and As, however, in the soils of the west and south parts of the investigated area were higher than that in the rest, and in a fewer plots enrichment of Zn and Cr was found, perhaps demonstrating influences of industrial activities and traffic on the soil environment. It was revealed that Hg was the dominant heavy metal pollutant, and soil Hg pollution occurred mainly in vegetable production bases and land on either side of Chuhe River, especially the vegetable production bases in the central part of the investigated area. Compared to the interpolation method, spatial prediction based on the fuzzy set theory could achieve a continuous output with significantly higher accuracy.

Key words Fuzzy set theory; Fuzzy e -means; Continuous classification; Spatial prediction; Soil pollution; Heavy metal