

# 工业污染场地 土壤污染空间分布研究

刘庚 著



科学出版社

# 工业污染场地

## 土壤污染空间分布研究

刘庚 著



科学出版社  
北京

## 内 容 简 介

大型工业污染场地是一个全球性环境问题，工业污染场地不仅破坏区域土壤环境和生态系统的平衡，土壤中高含量的污染物质还会对人体健康造成严重威胁。在现有工业污染场地土壤污染物空间分布预测、风险评估和修复治理过程中，由于缺乏对数据空间统计特征、热点区识别及空间分布预测模型等方面的深入分析手段，造成环境调查结果有很大的不确定性和误差。本书在数据科学解读及空间分布预测等方面，建立了一套适合污染场地有机污染物空间分布预测及不确定性评价的科学方法体系，具有重要的理论和实践意义。

本书可作为高等院校、科研院所相关专业的教学用书，也可供环境科学和环境工程领域的科研工作者和技术人员参考。

### 图书在版编目(CIP)数据

工业污染场地：土壤污染空间分布研究/刘庚著. —北京：科学出版社，2017

ISBN 978-7-03-053276-3

I. ①工… II. ①刘… III. ①工业污染防治—研究 IV. ①X322

中国版本图书馆 CIP 数据核字 (2017) 第 128648 号

责任编辑：周艳萍 刘文军 / 责任校对：陶丽荣  
责任印制：吕春珉 / 封面设计：王子艾工作室

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

<http://www.sciencep.com>

三河市骏杰印刷有限公司印刷

科学出版社发行 各地新华书店经销

\*

2017 年 11 月第 一 版 开本：B5 (720×1000)

2017 年 11 月第一次印刷 印张：9 3/4

字数：226 000

定价：68.00 元

(如有印装质量问题，我社负责调换〈骏杰〉)

销售部电话 010-62136230 编辑部电话 010-62151061

版权所有，侵权必究

举报电话：010-64030229；010-64034315；13501151303

# 前 言

大型工业污染场地是一个全球性环境问题，不仅破坏区域土壤环境和生态系统的平衡，土壤中高含量的污染物质还会对人体健康产生严重威胁。近年来，污染场地引起的危害事件和带来的环境污染问题已引起广泛关注，基于风险的污染场地管理和修复治理工作已显得极为迫切和重要。另外，污染场地的存在对社会经济发展、相关规范的制定以及土地利用类型的转变都具有重要影响，是目前各国共同关注的环境问题。

在现有工业污染场地污染物空间分布预测、风险评估、修复治理过程中，由于缺少对数据统计特征、空间特征以及热点区识别的深入分析手段，造成环境调查结果有很大的不确定性和误差。国内针对大型工业污染场地开展了广泛研究，但研究重点主要集中在污染场地风险评估与控制、污染场地修复治理措施及相关技术等方面，对于污染场地土壤中污染物空间分布表征及不确定性研究，目前还未有相关资料进行系统性报道。

本书以我国多个典型工业污染场地为研究案例，立足于污染场地管理方面的需求，基于经典统计学、环境科学、地统计学等理论方法体系，拟在数据解读方面针对污染场地污染物样点含量数据提供一套科学的数据分析流程和手段。本书综合运用多种方法学手段，系统地阐述了大型工业污染场地土壤污染物空间分布表征及不确定性评价的流程体系，内容丰富，层次清晰，具有技术性、全面性、创新性和专一性的特点。

本书是作者对过去相关研究成果的总结，包括国家自然科学基金项目“顾及污染物向异性特征的土壤有机污染物三维空间分布预测方法研究”（编号：41401236）、山西省自然科学基金项目“典型工业污染场地地下水健康与生态风险评估研究”（编号：2015021166）、山西省人文社科重点基地项目“汾河流域典型区域土壤污染识别及控制对策”（编号：2016330）、环境保护部科研专项子课题“污染场地中典型污染物迁移过程与表征方法研究”等。本书得以出版，作者要感谢山西农业大学毕如田教授、中国环境科学研究院郭观林研究员、太原师范学院王

尚义教授、牛俊杰教授、郭文炯教授对本书编写提出的宝贵建议和指导；同时感谢科学出版社周艳萍老师所付出的辛劳。

鉴于作者水平有限，书中缺点和不当之处在所难免，恳请广大读者批评指正。

作者

2016年10月

# 目 录

第 1 章 绪论	1
第 2 章 工业污染场地土壤污染及环境风险	5
2.1 污染场地土壤污染空间分布的不确定性	5
2.2 污染场地土壤污染环境风险及空间分布表征技术	7
2.2.1 污染场地数据统计分析及空间插值技术	7
2.2.2 污染场地高偏倚性数据处理及热点区识别	13
2.2.3 污染场地采样调查过程中的不确定性分析	15
2.2.4 污染场地管理现状及其环境风险	16
第 3 章 案例污染场地特征与样点数据采集	27
3.1 焦化污染场地特征描述	27
3.2 土壤样点数据采集与分析	30
3.2.1 土壤样点数据采集	30
3.2.2 土壤样点数据分析	32
3.3 污染场地概念模型	33
3.4 场地调查与初步污染识别结果	34
第 4 章 场地土壤中 PAHs 污染数据的统计特征分析	36
4.1 研究方法	37
4.1.1 多元统计分析方法	37
4.1.2 全局趋势分析方法	41
4.1.3 局部变异分析方法	41
4.2 统计特征分析	42
4.2.1 描述性统计特征	42

4.2.2	多元统计分析	46
4.2.3	样点含量数据异常值识别	51
4.2.4	空间分布趋势分析	54
4.2.5	局部空间变异分析	57
<b>第 5 章</b>	<b>场地土壤中 PAHs 污染的热点区识别</b>	<b>62</b>
5.1	研究方法	63
5.1.1	全局 Moran's I 及局部空间自相关	63
5.1.2	空间权重矩阵	65
5.1.3	全局空间自相关和局部空间自相关适用性对比分析	65
5.2	全局空间自相关性分析	66
5.2.1	全局空间自相关的散点图分析	66
5.2.2	全局空间自相关统计结果	69
5.3	污染热点区识别	70
<b>第 6 章</b>	<b>场地土壤中 PAHs 空间分布的分异性特征</b>	<b>74</b>
6.1	研究方法	75
6.2	阈值设定及指示半变异函数的拟合	76
6.2.1	阈值设定	76
6.2.2	指示半变异函数的拟合	76
6.3	污染场地 PAHs 分布的指示克里格分析	77
<b>第 7 章</b>	<b>场地土壤中 PAHs 空间分布范围预测的不确定性分析</b>	<b>81</b>
7.1	研究方法	82
7.1.1	数据正态变换方法	82
7.1.2	插值模型	83
7.1.3	插值精度评价方法	86
7.2	正态变换方法对普通克里格模型预测结果的不确定性影响	87
7.2.1	数据基本统计特征分析	87
7.2.2	最优理论半变异函数拟合	89
7.2.3	普通克里格模型预测结果分析	90
7.3	不同插值模型预测结果的不确定性分析	93
7.3.1	不同插值模型对均值、极大值和极小值的预测精度	93

7.3.2	不同插值模型预测的污染范围分析	94
7.3.3	不同插值模型预测结果的不确定性分析	96
7.4	某铅酸蓄电池污染场地 Pb 空间分布预测	98
7.4.1	场地特征及样品采集分析	98
7.4.2	场地土壤中 Pb 含量统计及空间变异特征	99
7.4.3	不同插值方法对土壤中 Pb 的污染插值结果比较	101
<b>第 8 章</b>	<b>场地土壤中污染物三维空间插值研究</b>	<b>106</b>
8.1	研究方法	107
8.2	不同三维插值模型对土壤 Bap 预测结果的影响	111
8.2.1	钻孔数据的采集与分析	111
8.2.2	土壤钻孔样点含量数据的统计特征分析	112
8.2.3	污染场地的三维地层建模及分布表征	113
8.2.4	场地土壤中 PAHs 污染范围界定及土方量估算	115
8.3	顾及污染物向异性特征的土壤 Pb 三维分布预测	117
8.3.1	钻孔数据采集分析与插值参数设定	117
8.3.2	场地不同地层土壤中 Pb 含量的统计分析	119
8.3.3	不同模型估算场地不同地层受污染土方量分析	120
8.3.4	不同模型界定的 Pb 含量三维污染范围评价比较	121
<b>第 9 章</b>	<b>非规则地形污染场地土壤中 PAHs 污染模式及分布特征</b>	<b>125</b>
9.1	场地概况与样品采集分析	126
9.2	场地土壤中污染物含量特征	127
9.2.1	不同分区土壤中 PAHs 含量	127
9.2.2	场地不同分区及不同土层土壤中 PAHs 残留特征	129
9.3	场地土壤中污染物污染模式及空间分布规律	130
9.3.1	不同分区中污染物样点分布的主成分分析	130
9.3.2	场地土壤中特征污染物空间分布特征	131
<b>第 10 章</b>	<b>工业污染场地土壤污染空间分布规律</b>	<b>134</b>
	参考文献	137



# 第 1 章

## 绪 论

污染场地是一个全球性的环境问题，尤其是大型工业污染场地，更是许多国家环境监管和风险管理的重点对象。近年来，随着我国经济的迅速发展和产业结构调整步伐的加快，某些经济发达城市的工业污染企业已普遍关停或者搬迁。遗留的工业污染场地数量及污染程度呈现上升趋势，工业污染场地引起的环境问题日益突出，对生态环境和人类健康造成了严重威胁，极大地制约了我国城市土地资源的可持续发展。

区域土壤环境是整个生态环境系统的重要组成部分，工业污染场地不仅破坏了区域土壤环境原有的生态功能与系统平衡，赋存在土壤中的各种污染物还会通过迁移转化、富集，对生物和人体健康产生危害。

污染场地的存在对国家的经济可持续发展、土地资源再利用、法律法规和相关行业标准的制定等都有重要的影响，已引起了我国政府部门的高度关注。与发达国家相比，目前我国对污染场地环境风险管理还缺少完善的法律和管理体系，在理论基础和技术储备方面尚待完善，因此针对我国污染场地管理和应用方面的需求，有必要加强新的理论和技术方面的研究。

污染场地中污染物的空间分布信息和污染分布表征是场地风险评价及修复治理决策制定的基础。受人类强烈干扰和污染累积释放因素的影响，污染场地的污染特征不同于一般面源污染，在局部区域存在污染热点区。目前常用的空间分布预测插值方法都是以对样本总体含量特征的最佳预测为目标，在对污染场地污染土壤样点含量数据进行空间分布预测和表征时，会使污染热点区域产生平滑现象，空间插值模型对污染场地中污染物空间分布的准确界定有较大影响。为准确界定污染场地中污染物空间分布范围，评价不同插值模型对预测结果的不确定性影响，分析不确定性产生的原因和预测不确定区域的分布，本书选择某大型工业污染场地为研究对象，运用多元统计分析、趋势分析理论和空间局部变异理论揭示污染土壤样点含量数据的统计特征；采用空间自相关理论分析样点在场地中的聚集特征并识别污染热点区域；结合非参数地统计学中指示克里格模型、反距离加权模

型、Johnson 数据正态变换+普通克里格模型以及分块组合预测模型预测污染场地中特征污染物空间分布范围并进行不确定性评价；基于地质建模系统及三维表征技术对污染物在不同地层上的分布进行三维插值和可视化表征。

本书以工业污染场地为研究对象，综合运用上述多种方法学模型，系统揭示了场地中土壤污染物空间分布规律及不确定性，主要探讨领域和研究内容如下。

### 1. 污染物样点含量数据的统计与空间特征分析

1) 分析目标污染场地中污染物样点含量数据的描述性统计特征，获取样点含量数据的频率分布规律并识别样点数据中的异常真实高值点。

2) 利用经典多元统计分析方法中的主成分分析、相关性分析和聚类分析方法，分析污染物样点含量数据的多元统计规律，获取污染物的来源、成因以及污染特征与场地生产工艺的关系。

3) 从三维透视角度来分析采样点数据在整个场地中的分布情况，比较不同污染物的空间分布趋势，判别其在土壤中累积的影响因素，结合场地历史生产和车间布局情况进行叠加分析，揭示场地中污染物的整体分布趋势和污染成因。

4) 采用 Voronoi 方法，分析样点与相邻样点的相似性，判别土壤中污染物含量的局部变化特征。

### 2. 场地土壤中污染物的热点区识别

1) 基于空间自相关理论，分析污染物的全局空间自相关特性，描述整体分布状况，判断在空间上是否有聚集特征。

2) 分析污染物的局部空间自相关特性，描述区域单元与周围区域单元之间的空间聚集程度和空间上的阶层性分布特征。

3) 基于全局和局部空间自相关特征，分析污染物在目标污染场地中的热点区和冷点区的分布特征和规律。

### 3. 污染物空间分布的分异性特征研究

1) 采用非参数地统计学中指示克里格预测模型，对污染物在场地中的空间分布进行概率预测，绘制污染概率分布图。

2) 获得各概率区间的污染分布面积，确定目标场地中高概率污染的重点区域以及污染概率分布规律。

3) 将污染概率图与采样点和原厂区生产平面布置图叠加分析，描述污染物空间分布的分异性特征。

#### 4. 污染物空间分布范围预测的不确定性分析

1) 分析污染物含量数据的统计特征及不同空间插值模型的适用原理, 研究如何对大型工业污染场地中具有异常真实高值点的含量数据进行分布预测。

2) 针对本场地样点含量数据集具有严重偏斜性的特征, 比较3种典型数据正态变换方法 (Box-Cox、Johnson、Normal Score) 对普通克里格模型预测结果的影响, 评价不同正态变换模型的插值精度。

3) 比较常用插值模型中的反距离加权方法、Johnson 数据正态变换+普通克里格模型以及数据拆分后的组合预测模型对特征污染物预测的结果, 分析不同预测模型的总体精度及对数据统计特征的预测精度, 揭示误差产生的原因, 并评价预测结果的不确定性和不确定性产生的原因。

4) 建立适合表征具有严重偏斜数据特征的最佳空间插值模型, 基于空间预测结果, 创建预测标准误差图, 界定出在判断规则条件下预测的污染范围、未污染范围和污染的不确定性区域。

#### 5. 污染物的三维可视化表征及污染土方量估算

1) 基于地质建模系统和土壤钻孔样点数据, 通过构建场地的地层三维模型, 模拟场地中地层在真三维环境中的分布, 获取不同地层的分布状况和分布规律, 加强对场地地质特征的认识, 辅助分析污染物在不同地层的分布和迁移规律。

2) 采用不同的三维插值方法及综合污染物向异性结构特征, 对场地特征污染物进行三维插值预测, 揭示污染物在不同地层中污染分布状况和分布规律; 通过交叉验证结果, 比较不同三维插值模型的精度, 建立适合本场地特征的最佳三维预测模型。

3) 基于设定的污染阈值, 采用预测精度最高的三维预测模型, 界定特征污染物在不同地层中的污染范围, 并估算受污染土壤的土方量。

#### 6. 顾及地形特征的污染场地土壤多环芳烃污染模式及分布特征研究

1) 根据该场地地形特征, 结合原厂区车间布局, 将该场地划分为西部、西南、东北3个分区, 在每个分区内根据污染源分布, 按照判断布点的原则进行布点并采集样品。

2) 采用三维克里格模型表征场地特征污染物空间分布规律, 进行三维插值计算, 综合场地地形特征, 即将整个地层数据视为一个整体和考虑地表高程差两种情况进行计算。

3) 采用二维和三维克里格插值模型, 对比分析研究不同模型在考虑和不考虑地表高程差对界定污染范围和受污染土方量的影响。

4) 采用多变量多元统计分析方法, 揭示不同地形区域内样点的空间组合及分离特征, 综合分析场地特征污染物的空间分布模式。

在现有工业污染场地中污染物空间分布预测、风险评估、修复治理过程中, 由于缺少对数据统计特征、空间特征以及热点区识别的深入分析手段, 造成环境调查结果有很大的不确定性和误差。本研究拟在数据解读方面针对污染场地有机污染物样点含量数据提供一套科学的数据分析流程和手段。在数据分析结果和比较不同插值模型基础上, 采用数据拆分的二维插值模型及顾及污染物向异性特征的三维空间插值模型, 建立了适合表征具有很大偏斜度和极大异常值数据特征的空间插值方法, 并能够评价污染空间分布预测中的不确定性区域及产生不确定性的因素。

## 第 2 章

### 工业污染场地土壤污染及环境风险

#### 2.1 污染场地土壤污染空间分布的不确定性

污染场地,尤其是大型工业污染场地是一个全球性环境问题(黄瑾辉等,2012; Forslund et al., 2010),在西方一些发达国家工业污染土壤比例高达 20%以上,同时这些国家长期以来在污染场地环境管理方面开展了大量工作(Agostini et al., 2012; Lemming et al., 2012; Agnew et al., 2011; Baciocchi et al., 2010)。工业污染场地不仅破坏区域土壤环境和生态系统的平衡,土壤中高含量的污染物质还会对人体健康产生严重威胁(Szabová et al., 2008; Colombo et al., 2006),污染场地引起的危害事件和带来的环境污染问题已引起广泛关注,基于风险的污染场地管理和修复治理工作已显得极为迫切和重要(Sinha et al., 2007)。另外,污染场地的存在对社会经济发展、相关规范的制定以及土地利用类型的转变都具有重要影响,是目前各国共同关注的环境问题。

近年来,伴随着我国经济的高速发展和城市规模的扩大,污染场地及其严重危害事件时有报道,由污染场地所引起的生态环境破坏、食品安全降低、人体健康危害等环境问题日益突出,污染场地分布与污染程度呈现上升趋势(骆永明, 2011; 2009)。由于污染场地对人体和环境产生了重大危害,因此为适应我国在污染场地管理方面巨大的技术需求,迫切需要开展污染场地风险评价、管理等方面的技术研究。与发达国家相比,目前我国对污染场地环境风险管理还缺少完善的法律和管理体系,在理论基础和技术储备方面尚待完善(郭观林等, 2010; 2009),极大地制约了我国污染场地环境监管的水平和力度。为了提高我国污染场地科学管理的水平,控制污染场地的泛滥,修复治理已有污染场地的危害,必须加快我国工业污染场地环境监测、风险评估与功能修复技术的研究,提升我国污染场地环境管理决策水平及应对环境风险的能力。

在我国一些经济发达或快速发展城市,由房地产等行业引起的对某些土地利用类型的用途改变,促使污染场地修复工作陆续开展(胡新涛等, 2012; 陶锟等,

2012)。污染场地中污染物的空间分布信息和污染分布表征是场地风险评价及修复治理决策制定的基础。受土壤空间异质性及污染调查手段的影响,在实际污染场地环境调查工作中,污染物空间分布预测结果与实际情况相差较大,导致场地中污染物的污染评价存在一定的误差和不确定性。降低场地环境调查的不确定性,提高污染物空间分布预测结果的精度,对场地的环境管理以及修复治理具有重要的理论及现实意义。污染场地中污染物含量受历史生产过程中人类强烈干扰和累积释放因素的影响,在场地土壤中具有极强的空间异质性和空间分布的不连续性,依据现有的采样布点方案,很难采集具有足够代表性的样点,采样布点方案及空间预测插值模型是导致污染场地环境调查结果不确定性的主要因素(姜成晟等,2009)。受采样成本的限制,很难进行大规模样点采集,在实际工作过程中,通常基于现有采样点,通过比较不同插值模型的适用范围和预测精度来提高调查结果的准确性(Panagopoulos et al., 2006)。因此,空间插值模型的选择对污染物空间分布预测结果的精度有重要影响。目前常用的空间分布预测插值方法都是对样本总体含量特征的最佳估计,在应用于污染场地污染物空间分布预测时,可能会对污染热点区域产生平滑现象,在对污染场地中污染物进行空间分布预测时会出现低值被高估和高值被低估的现象。另外,污染场地属于典型的点源污染,在人类强烈干扰和污染成因等因素影响下,其有机污染物含量在局部地区存在异常真实高值现象,样点含量数据具有很大的偏斜度和空间不连续性,也难以直接用现有空间插值模型进行空间分布预测计算。如何依据污染场地有机污染物的性质,构建一套数据特征结构分析和空间分布预测的科学方法体系,是目前实际工作中面临的亟须解决的问题。

由于污染场地土地利用类型的变化,目前已有大量场地进入调查、评估和修复过程,所有过程均以采样点数据为基础,来揭示污染物空间分布和累积释放效应。污染场地土壤污染调查数据大多都是离散的点状数据,在污染物空间分布、风险评估、修复范围确定等工作中需要统计分析这些点状数据以揭示污染物的时空演化规律。因此基于采样调查点状数据,对其进行统计特征和空间特征分析是研究污染物空间分布规律、分异性特征及不确定性的基础。同时,污染场地中污染物空间分布的准确界定也是污染场地环境调查和风险评估等相关工作的基础,对修复范围的确定、修复土方量的估算以及修复治理成本具有直接重要影响。

受人类强烈干扰和点源污染的影响(Rawlins et al., 2005; Lark, 2002),污染场地的污染特征不同于一般面源污染。样点数据特征影响空间预测和风险评价计算模型的选择,会对结果产生很大的不确定性(Wu et al., 2011; Sinha et al., 2007)。在实际工作过程中经常不考虑数据统计规律和空间特征,缺乏数据的有效深入分析,从而降低了调查结果的精度,进而影响到污染物空间分布的准确界定。

因此,本研究立足于污染场地管理方面的需求,基于管理学、经典统计学、环境科学、地统计学等理论方法体系,利用多元统计分析、空间变异分析、趋势分析等模型分析污染数据结构特征、空间变异及空间连续性的特点,采用非参数地统计学中指示概率预测模型研究污染土壤中污染物的空间分异性特征;比较常用插值模型中的反距离加权方法、Johnson 数据正态变换+普通克里格模型以及数据拆分后的组合预测模型对特征污染物预测的精度,提出适合具有较大变异性 and 较高偏斜度数据特征的污染预测方法思路,评价污染预测过程中的不确定性及产生不确定性的原因;运用三维建模系统模拟场地的地层分布状况,并对污染物在不同地层中的分布进行三维可视化表达。本研究从数据科学解读及空间分布预测等方面建立了一套适合污染场地有机污染物空间分布预测及不确定性评价的科学方法体系,具有重要的理论和实践意义。

## 2.2 污染场地土壤污染环境风险及空间分布表征技术

### 2.2.1 污染场地数据统计分析及空间插值技术

#### 1. 采样调查数据处理程序与基本原理

污染场地的环境调查、风险评价及空间分布预测等工作过程均以土壤采样点数据为基础,污染场地土壤采样数据同其他地学数据获取类似,要经过采样、预处理和化学分析等一系列步骤(裴韬等,1998),以前受采样仪器、采样方法及化学分析实验条件的影响,会产生不同过程的误差,使最终获取的数据中含有一定的“噪声”,常采用移动平均法、傅里叶滤波法和小波滤波等方法进行去噪处理(秦前清等,1994; Aminzadeh, 1991; 胡以镛, 1991)。随着科学研究的不断发展,采样仪器和采样方法也都有很大的进步,样品分析化验的精度也有了极大提高,现阶段土壤样品中污染物含量的分析精度已经达到 ppb 级 ( $10^{-9}$ ),经采样分析化验的样点含量数据质量和精度能够满足现有应用要求。污染场地采样点含量数据由于包含空间坐标信息,因此是一种特殊类型的空间数据(李德仁等,2003; 1998)。除传统统计方法学手段外,目前已有多种空间数据挖掘技术和探索性空间分析技术引入地学数据或环境调查数据的研究中来。

传统统计学中的描述性统计分析方法能够从数理统计学的角度反映样点含量数据集的特征,在区域土壤环境调查中常采用描述性统计分析方法来分析采样点数据集的统计学规律(谢云峰等,2010; 刘江生等,2008)。描述性统计分析可以描述样点含量数据的各种特征及所代表的总体特征,便于发现数据集的内在规律。描述性统计分析要对目标数据集所有变量的相关数据进行统计性描述,主要有数

据的频度分析、集中趋势分析、数据离散度分析等。常用的描述性统计指标包括样点含量数据的范围、均值、标准差、方差、中值、四分位数、偏度、峰度、变异系数等。集中趋势分析是指数据集向某一中心值靠近的程度，反映数据集中心点所在的位置，常用众数、中位数、均值等指标来表征。数据离散程度反映了各变量值远离数据集中心值的程度，也从另一个侧面说明了集中趋势测度值的代表程度，测度指标主要有全距、四分位差、平均差、标准差、方差和离散系数。

经典统计学中的多元统计分析方法已广泛应用于区域土壤环境采样调查数据中。相关性分析、主成分分析和聚类分析等多元统计分析方法可以简化数据，用综合指标代替一类相关性较高的数据，从而反映数据之间的关联（吕建树等，2012）。多元统计分析方法在农业、生物、环境等领域有着广泛的应用（Dominick et al., 2012；陈修康等，2012；Arrouays et al., 2011；Selvarasu et al., 2010；Saby et al., 2009），在城市土壤（陈景辉等，2011；Zhang, 2006；Li et al., 2004），尤其农业土壤（Cai et al., 2012；石宁宁等，2010；Franco et al., 2009；Micó et al., 2006；Martín et al., 2006）重金属源解析方面，取得了较好的效果。但多元统计方法应用于大型工业污染场地样点数据统计分析的报道还不多见。相关性分析可以检验数据集之间的近似性，对土壤污染中各种污染物含量数据进行相关性分析，可以判别土壤污染物的污染成因及识别不同的污染来源（Chen et al., 1997）。主成分分析可以同时分析多个要素，将多个元素构成的数据矩阵在多维空间的变异归纳为几种少数互不相关的潜在因素的影响。在土壤污染调查中，主成分分析方法可以用于区域人类活动和自然背景对土壤污染物累积的贡献（Idris, 2008；Facchinelli et al., 2001）。主成分分析在区分土壤元素主要来源方面具有优势，即不需要对元素形态进行细致分析，对数据量没有特别要求，也不需与历史数据对比，即可判断出哪些土壤元素含量受到人为因素的影响明显（赵彦锋等，2008）。聚类分析可以根据样本自身的属性，用数学方法比较各要素之间的性质，依据某些相似性或差异性指标，来定量描述样本之间的亲疏关系，并按这种关系的亲疏程度对元素进行聚类（程荣进等，2009），在土壤污染调查中，聚类分析方法可以判别污染物的相似程度。多元统计分析方法虽然没有本质差异，但不同方法的分析结果可以相互验证（Yalcin et al., 2008）。

随着空间信息技术的发展，趋势分析理论被引入环境领域应用中（马民涛等，2010；齐鑫山等，2000），利用趋势分析能够从不同角度分析样点含量数据的全局趋势分布，有助于判别污染成因和在土壤中累积的影响因素。空间趋势分析主要表达空间物体在空间区域上变化的显著特征，反映了空间物体的总体规律，而忽略局部的变化情况。全局趋势分析根据抽样调查数据，拟合一个数学曲面，用数



学曲面来表达空间分布的变异情况。如果能较好模拟出样点含量在整个区域中的趋势分布规律,进而能更准确地模拟短程随机变异,则可以从生成的趋势面分析透视图中看出全局趋势。

空间自相关是区域化变量的基本属性之一。空间自相关统计可用于检测研究区域内变量的分布是否具有空间结构(张朝生等,1995),在区域土壤环境变量的相关研究中也经常用到空间自相关方法。土壤中各种元素以及赋存于土壤中的污染物虽具有一定的空间变异性,但同时土壤在空间上是一个自然连续体,土壤空间上的这种连续性的存在使得邻近样点的相似性比距离较远的样点高,并且在空间自相关范围以内的采样也并非完全独立。用传统的数理统计方法对其进行研究,由于其不能满足随机条件而可能产生偏差(Zhang et al., 1997);而空间自相关分析方法是通过对区域变量空间分布相近样点间的相关程度进行检验(Martin, 1996),通过检验反映出空间上某样点值与相近空间点上的样点值的显著相关程度(Waser et al., 1990)。空间自相关方法揭示了在空间上变量样点位置越靠近,其属性值越相似的空间分布特征。当某一变量值高而其周围样点属性同样是高值时,称为空间正相关性,相反,则称为空间负相关性。空间自相关方法在土壤微量元素(郭旭东等,2000)、土壤有机质(黄智刚等,2006)、土壤重金属(霍霄妮等,2009)等领域都有较好的应用效果。空间自相关分析使用全局和局部两种指标进行描述。全局空间自相关描述样点变量在整体空间上的分布状况,判断在空间上是否存在聚集特征,用单一值来反映该区域的自相关程度,但不能明确指出在哪些区域。局部空间自相关描述每一个空间单元与邻近单元就某一属性的相关程度(张松林等,2007)。将全局性不同空间间隔的空间自相关统计量依次排序,绘制空间自相关系数图,可以发现该现象在空间上是否有阶层性分布。

数据离群值识别、局部空间变异分析和多数据集协变分析也是常用的区域土壤环境变量数据处理分析方法。数据离群值是指数据中某个或某几个数值与其他数值差异较大,包括全局离群值和局部离群值两类。全局离群值是指对于全体数据来说,个别数据值高于或低于其他数据值很多;局部离群值是指整个数据集中的数据都处于正常范围,但个别数据与其相邻数据差异较大。离群值可能是由于化验分析等处理不当产生的,也有可能是真实异常值。如果离群值由人为因素引起,则在数据集中要将其修改或删除,否则将影响半变异建模和邻域分析的取值。在土壤污染调查中,离群值是研究和诊断污染的重点。在单个变量、样本量较小的情况下, $t$ 检验、Dixon检验、Grubbs检验及Walsh检验具有较好的识别效果(赵慧等,2003;陶澍,1994)。局部空间变异分析可以揭示土壤采样点与周围采样点之间的关系,通过分析某个样点与其相邻点的相似程度,可以识别土壤中污染物的局部变化情况、空间离散特性及局部区域受点源污染影响的特征,有助于判别