

文章编号 1000-5013(2004)04-0435-05

地统计学的 GIS 空间分析功能扩展

孙英君 王劲峰 柏延臣 李道波

(中国科学院地理科学与资源研究所, 北京 100101; 清华大学环境科学与工程系, 北京 100084;

山东聊城中华发电厂, 山东 聊城 252033)

摘要 随着各行各业对 GIS 的广泛应用, 人们越来越希望在一个统一的 GIS 软件包界面下同时完成各种各样的空间分析, 因此将传统的空间分析方法与 GIS 相结合一直受到关注. 文中利用 ArcView 下的编程语言 Avenue, 实现地统计学分析方法与 GIS 软件的链接. GIS 软件包扩展模块的地统计学分析方法, 除经典的克里格系列方法外还包括一定的模拟方法, 如概率场模拟、高斯序列模拟和模拟退火等方法.

关键词 地统计学, GIS, 空间分析, 模拟, Avenue 语言

中图分类号 TP 79 P 628+.2 TP 39 P 208

文献标识码 A

空间分析功能是 GIS 区别于其它系统的本质特征之一. 当前 GIS 技术的发展已经走过了以空间数据为中心的应用阶段, 逐步进入以决策支持为特征的应用时期. GIS 的技术优势, 不应只局限于空间数据采集、存储、管理、显示与输出一体的数据流程, 更应体现在其信息综合、空间分析、预测预报和辅助决策能力上^[1]. 将地统计学的分析方法与 GIS 软件包相链接, 从而将地统计学方法融合为 GIS 的空间分析方法, 是本文要讨论的重点.

1 地统计学简介

地统计学起源于世纪 60 年代, 主要用于研究地质学现象的空间结构和进行空间估值. Matheron 将其定义为随机函数形式对自然现象进行勘察与估计的应用^[2]. 地统计学技术基于采样点建立空间相互关联随机变量模型, 并用该模型估计未采样点的属性值.

近年来地统计学在国际上发展迅猛, 特别是 GIS 的发展, 对空间分析功能提出一个新的要求. 它使得地统计学成为多个学科重视的焦点. 目前地统计理论热点, 如正则化、模拟 (Simulation) 等已相对成熟, 但其软件却相对落后. 现有软件包如 Geovariance, GsLib, Gstat, GS+ 等, 涉及到 Simulation 分析的仅限于极少数, 且大都基于 Dos 或者 Unix 系统, 不具有通用性. ESRI 推出的 ArcGIS 系列软件, 从 8.0 版本开始, 将地统计学作为一个扩展模块加入进来, 但它仅仅包含了克里格插值的系列方法. 随机模拟方法在地统计分析中, 也占有重要地位.

在国内, 地统计学的研究工作开始于 1978 年^[3]. 最初应用专业基本限制在矿业、石油部门, 到 80 年代末期, 已在生态、水利、土壤学领域的应用逐渐推广开来. 1987 年, 清华大学水利系资源组将地质统计学应用于水科学中. 1989 年, 沈思渊用地统计学研究土壤空间的变异结构^[4]. 1994 年, 李天生、周国法利用地统计学对马尾松毛虫空间分布进行建模^[5]. 1998 年, 李哈滨、王政权等对阔叶红松景观中林型和土壤类型的空间异质进行变异函数的定量研究^[6]. 与此同时, GIS 技术得到了飞速的发展, 各行各业的科学家在可视化表达的基础上, 提出了同时进行空间统计分析等更高的要求. 因此, 地统计学与 GIS 软件

收稿日期 2004-01-29

作者简介 孙英君 (1976-), 女, 博士研究生, 主要从事 GIS 空间分析及遥感尺度扩展方面的研究. E-mail: qingdaosyj@yahoo.com.cn

基金项目 国家高技术研究发展计划 (863 计划) 的空间信息分析技术及原型研发基金资助项目 (2002AA135230-1)

包的发展,可以说是相辅相成、相互促进的.目前可以供给不同领域研究人员使用的地统计学软件包,典型的为 ArcGIS 内含的地统计学扩展模块.但正如上文所述,它仅仅包含了克里格插值方法,因此,本文推出了功能相对完整的地统计学软件包,将各种模拟方法也包含在扩展模块内,并实现模块与地理信息系统软件包的无缝链接.它已成为不同空间领域广大科研人员得力的分析工具.

2 软件包功能介绍

本软件包采用 Fortran 语言作为底层开发语言.同时,采用 Visual Basic 的组件式技术与 GIS 软件包——ArcView 3.2 实现无缝链接.软件包共包括 3 个模块,分别是变异函数模块、克里格插值模块、随机模拟模块.

2.1 模块与 GIS 软件包的结合方式

GIS 与空间分析模型结合是 GIS 发展的一个新动向.1997 年,王学军指出两者的结合方式在当时应采取 GIS 外挂空间分析模块为主,逐渐向嵌入式过渡^[7]. 本文即是基于组件式技术,将地统计学分析软件包嵌入到 GIS 软件包内部,实现两者的无缝链接.组件式技术是继面向对象思想后,更适应分布式软件模型的开发方法.即将复杂程序设计成一些小的、功能单一的组件模块,这些组件模块可以运行在 1 台或者多台计算机上^[8]. 本文首先根据功能区分,将整个地统计学软件包划分为 3 大类之内的若干子块,逐一进行编辑调试.各模块间相互独立,互不依赖,如图 1 所示.针对每一具体模块,本文采用 Visual

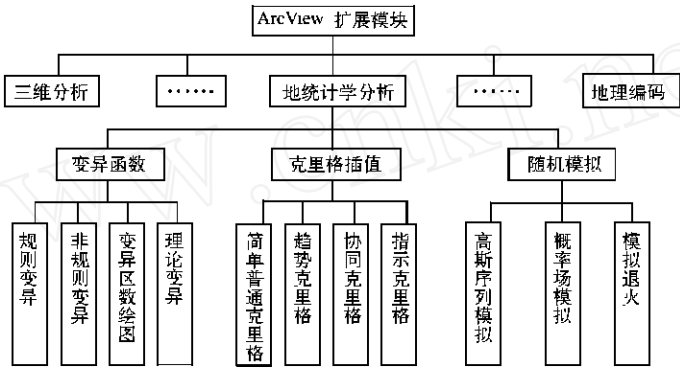


图 1 地统计学分析模块组成

Basic 技术进行参数输入的可视化界面设计.利用 Aven ue 语言编写的 Script 程序,在 ArcView 内部进行地统计学分析.将生成的文本文件结果,转化为 ArcView 自带的. shp 文件格式.整个地统计学软件包作为 ArcView 的 1 个扩展模块,由于没有同项目文件捆在一起,因而非常方便.当用户载入该扩展模块时,当前 ArcView 部分的所有项目文件都增加了相应的功能(图 2).

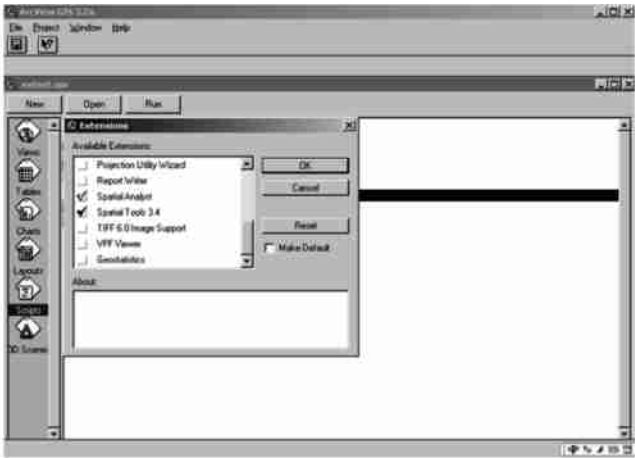


图 2 Extension 中的地统计学分析模块

2.2 软件包模块简介

地统计学的核心是空间异质性,通过变异函数对研究对象的空间分布进行建模.然后,利用估计

(Estimation) 或模拟 (Simulation) 方法进行未知点的解算. 这 3 个子模块, 对于地统计学而言是独立的, 因此软件包主要包括 3 部分. 其中, 估计方法包括二维克里格插值 (普通克里格、简单克里格)、趋势克里格、指示克里格及协同克里格方法. 模拟方法主要包括高斯序列模拟、 ρ -Field 模拟和退火模拟方法. 本文的软件包首先进行数据的标准化转换, 然后调用 Visual Basic 编写的可视化参数录入界面进行地统计学的相关分析, 最后通过程序的内部转换, 将结果以 .shp 文件格式显示在 ArcView 中. 整个分析过程都是在 ArcView 软件包内部进行, 实现了传统的地统计分析方法与 GIS 软件包的无缝链接.

2.2.1 输入数据模块 考虑到软件包的通用性, 规定输入数据采用最简单的文本文件列方式. 其基本

要素包括坐标 (x, y, z) 和属性值. 当所要分析的数据对象不是这种形式, 而是 GIS 软件包

拥有的形 2.2.2 变异函数模块 文件或 (点) 层文件时, 利用 Avenue 语言编写的数据转换

模块, 即可以实现数据格式的标准化转换. 转换方法主要是依次调用数据对象的属性文件,

读取坐标、属性值等有用信息. 然后, 写到一个符合标准格式的文本文件中, 如图 3 所示. 变

异函数是地统计分析中一个核心的组成部分. 考虑到研究对象空间分布形状的不同, 软件包

提供了规则变异函数建模和非规则变异函数建模两个模块. 规则变异函数模块主要针对规则几何形状的研究对象, 如格网、遥感图像等. 这种数据对象

的变异方向, 往往可以通过坐标系及单元网格的大小来进行描述. 如图 4 所示, 利用 Offsetx, Offsety 和

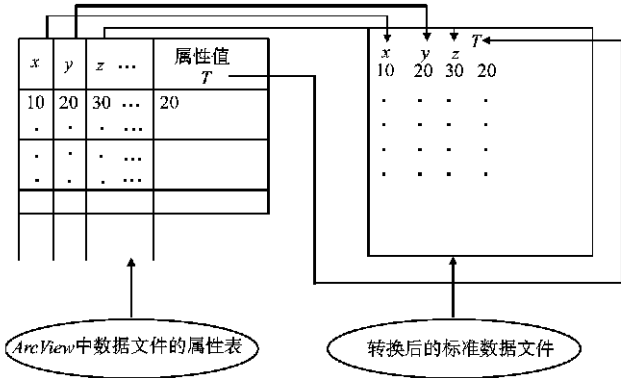


图 3 输入文件格式转换

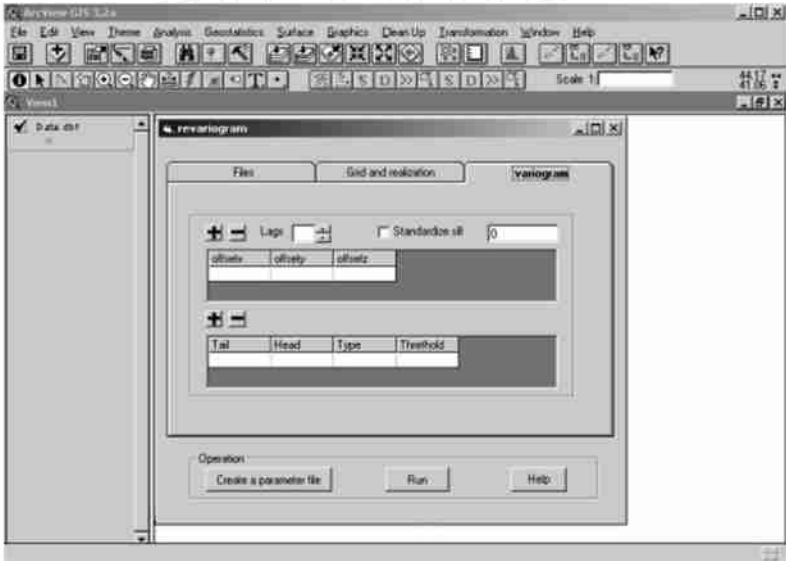


图 4 变异函数模块示例中的规则变异函数建模

Offsetz 分别表示三维坐标. 诸如, $(1, 0, 0)$ 代表沿 x 轴方向, $(1, 1, 0)$ 代表从正北方向的 y 轴开始向 x 轴方向旋转 45° 的方向. 对于非规则数据对象, 如离散分布的采样点, 则可利用非规则变异函数模块进行建模. 这时研究对象的变异方向需要通过具体的角度来进行描述. 此外, 通过变异函数绘图模块, 可利用经验变异函数的图形来进一步确定所要采用的理论变异函数. 而通过理论变异函数建模, 结合变异函数的可视化绘图, 可以通过图形的方式直观的比较理论与经验变异函数模型之间吻合的程度. 软件包提供了 5 种基本的理论变异函数, 即球状模型、指数模型、高斯模型、幂指数模型和孔洞模型. 这 5 种模型可单成一体, 也可结合起来共同描述研究对象的空间变异规律.

2.2.3 克里格插值模块 克里格插值是地统计学分析中比较传统的方法. 根据在局部范围内, 研究对象均值是否已知, 本软件包包含了简单克里格插值、普通克里格插值两种方法. 如果研究对象明显的含有趋势, 可选用软件包的趋势克里格插值方法. 因为克里格插值方法处理的对象是随机变量, 而不是确

定性变量.通过相减将趋势去除,然后再对剩余的残差项进行克里格分析,最后再将趋势加回去.研究对象由于某种原因采样点较少,而与其相关的另一变量却比较容易得到,这时需采用协同克里格插值方法进行分析.

2.2.4 随机模拟模块 随机模拟是地统计学分析中相对较新的一种方法,它主要克服克里格插值引起的平滑效应.因此它更加注重整体态势的表达,是一种研究对象波动性的方法.如前所述,现行地统计学软件包,包含随机模拟功能的免费软件包多为 DOS 界面.与 GIS 软件包相结合的地统计分析模块,如 ArcGIS 系列软件只是含有基本的克里格插值方法.所以,本软件包将随机模拟作为 ArcView 扩展模块的组成部分.同时,利用 Visual Basic 语言实现了参数的可视化输入界面(图 5).随机模拟最基本的方法是序列高斯模拟.它主要是通过将前面已经模拟过的结果,作为下一个要模拟点的已知条件进行条件模拟

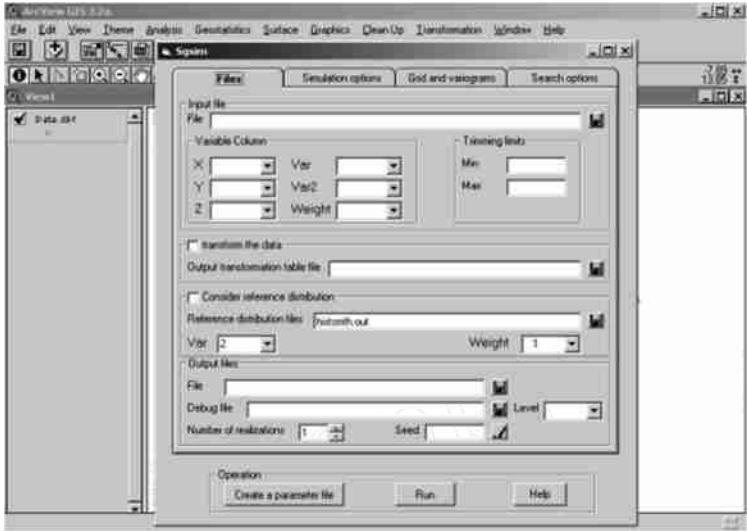


图 5 随机模拟示例中可视化的序列高斯模拟参数输入界面

来表现整体趋势的.为了提高大数据量模拟运算的效率,本软件包还包括了概率场模拟和模拟退火方法.概率场模拟首先进行无条件模拟,然后再利用已有的概率分布进行条件插值.这样,就不需要在每 1 个点处都要进行 n 维方程组的求解.模拟退火方法是将数学中的模拟退火方法和研究对象的空间特性相结合,具体表现在其目标函数的组成.它包括频率分布函数、变异函数与第二变量相关的系数等.

2.3 实例

本文以某地铜采样观测数据为实验数据.在 ArcView 环境下,利用克里格插值模块,进行该范围内铜分布曲面进行研究.该实验区面积(m^2)大小为 50×50 ,可获得的铜采样点空间分布(图 6).利用铜采

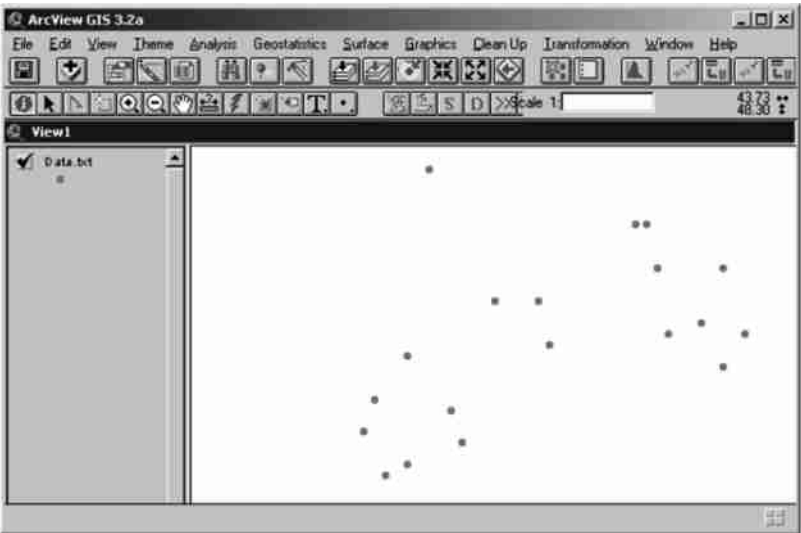


图 6 实验区铜元素的空间分布

样点数据,可知该分布的变异函数为 $\gamma(h) = 32 \times Sph(h/10)$.同时确定要拟合的铜曲面为 $50 \times 50 \times 1$,

即单元格尺度为 1 m, 整个实验区范围内共有 2 500 个网格组成. 将这些信息作为参数输入, 最后的插值结果, 如图 7 所示. 根据该曲面, 可以确定每一点 (指每一单元格) 的铜含量. 整个插值过程包括原始数据、克里格参数录入、最后结果显示, 都在 ArcView 中进行. 从而, 无缝地实现 GIS 软件的空间分析——地统计分析功能.

3 结束语

本文利用组件式开发技术, 将地统计学的空间分析技术的变异函数建模、系列克里格插值及随机模拟方法与 GIS 结合, 扩展了 GIS 的空间分析功能. 需要指出的是 ArcView 仅仅是早期 ArcInfo 系列的一个软件包. 本文选择它作为实验平台, 主要考虑其短小精悍做一个初步的尝试. 作为主流的 GIS 软件包, ArcGIS 利用 AO 技术实现包括随机模拟在内的地统计分析模块, 应该指日可待. 此外, 本文所包括的方法仅限于最基本的插值、模拟方法, 还有很多分析方法需要补充、完善. 比如, 因子克里格方法、用于分析分类数据的指示模拟方法, 等等.

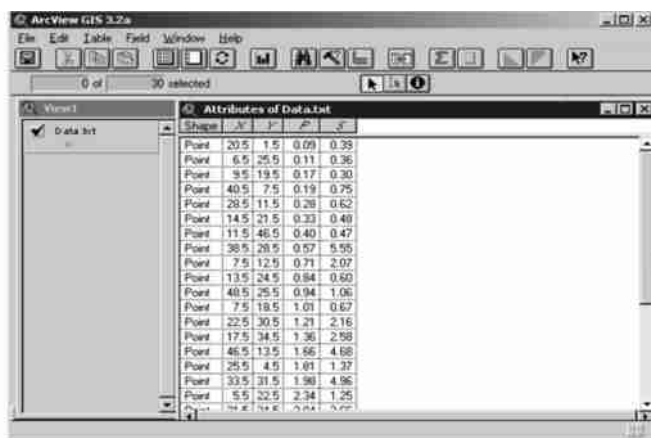


图 7 克里格插值得到的铜曲面

参 考 文 献

- 1 冯克忠, 万 庆, 励惠国. 基于组件技术的 GIS 广义空间分析[J]. 地球信息科学, 2003, 5(1): 62~66
- 2 Matheron B. The theory of Regionalized variables [M]. Fontainebleau: Centre de Morphologie Mathématique, 1971. 1~200
- 3 侯景儒. 中国地质统计学 (空间信息统计学) 发展的回顾与前景[J]. 地质与勘探, 1997, 33(1): 53~58
- 4 沈思渊. 土壤空间变异研究中地统计学的应用及其展望[J]. 土壤学年进展, 1989, 17(3): 11~24
- 5 李天生, 周国法. 空间自相关与分布型指数研究[J]. 生态学报, 1994, 14(3): 327~331
- 6 李哈滨, 王政权, 王庆成. 空间异质性定量研究理论与方法[J]. 应用生态学报, 1998, 9(6): 651~657
- 7 王学军. 空间分析技术与地理信息系统的结合[J]. 地理研究, 1997, 16(3): 70~74
- 8 张文江, 陈秀万, 李 京等. 基于 COM 组件技术的 GIS 空间模型库研究[J]. 中国图象图形学报, 2003, 8(1): 110~114

Extending the Function of Spatial Analysis in Geographic Information System Based on Geostatistics

Sun Yingjun Wang Jinfeng Bo Yanchen Li Daobo

(Inst. of Geograph Sci. & Natural Resources Res., ACS, 100101, Beijing, China;

Dept. of Environ. Sci. & Eng., Tsinghua Univ., 100084, Beijing, China;

Shandong Liaocheng Power Plant, 252033, Liaocheng, China)

Abstract With the wide applications of geographic information system (GIS) to all fields, people more and more expect to accomplish various spatial analyses simultaneously at the interface of an integrated GIS software package. So the combination of conventional method of spatial analysis and GIS is the focus of people's attention all along. By taking advantage of programming language Avenue under Arc view, the authors realize the inter linkage of the method of statistical analysis and GIS software. As the extended module of GIS software package, the method of geostatistic analysis includes some simulation methods such as probability field simulation, Gaussian sequential simulation and simulated annealing method in addition to the method of classical Kriging series.

Keywords geostatistics, geographic information system (GIS), spatial analysis, simulation, Avenue language