

空间信息分析技术

王劲峰, 武继磊, 孙英君, 李连发, 孟 斌

(中国科学院地理科学与资源研究所, 北京 100101)

摘要: 在 GIS 技术日趋成熟和空间数据极大丰富的今天, 通过分析空间数据探索空间过程机理正变得日益迫切。空间信息分析技术至少包括以下六个主要方面: (1) 空间数据获取和预处理; (2) 属性数据空间化和空间尺度转换; (3) 空间信息探索分析; (4) 地统计; (5) 格数据分析; (6) 复杂信息反演和预报。本文提出了解决具体应用问题一般的空间数据分析计算、结果解释和反馈程序。认为空间过程的一般共性和作为共同的研究对象, 各种不同的方法技术最终可能导致空间数学 (spatial mathematics) 的产生, 同时发展鲁棒的空间分析软件包对于普及空间数学是必要的。

关 键 词: GIS; 空间信息; 空间过程; 空间分析; 机理解释

文章编号: 1000 0585(2005) 03 0464 09

1 引言

随着对地观测、社会经济调查、计算机网络和格网信息处理能力的迅速提高, 空间数据正在以指数方式急速增加。通用和专用的 (时) 空间数据结构^[1]、应用于具体事物的管理信息系统、以及对这些海量空间数据进行深加工以获得高附加值的信息产品的空间信息分析技术^[2] 成为空间信息三大领域。

数据分析通常用统计学方法, 当今流行软件包 SPSS、MATLAB 等大大地促进了数据分析深加工及其在各领域的应用 (徐建华, 2002)。

但是, 空间数据通常具有非独立性, 这与经典统计学基本假设相驳。因此, 专门的空间信息分析理论和技术正在迅速发展^[3~5], 已在遥感、生态环境、地球科学、社会经济等领域得到诸多成功的应用, 展现出广阔的应用潜力。

2 研究动态

统计分析是常规数据分析的主要手段。然而, 传统统计学在分析空间数据时存在致命的缺陷, 这种缺陷是由空间数据的本质特征和传统的统计学方法的基本假设共同造成的。传统的统计学方法是建立在样本独立与大样本两个基本假设之上的, 对于空间数据, 这两个基本假设前提通常都得不到满足。空间上分布的对象与事件在空间上的相互依赖性普遍存在的, 这使大部分空间数据样本间不独立, 即不满足传统统计分析的样本独立性前提, 因而不适于进行经典统计分析^[6]; 另一方面, 有些空间数据采样困难, 如某些气象台站分布在稀疏的山区, 导致样本点太少而不能满足传统统计分析方法大样本的前提; 地学

收稿日期: 2004 10 16; 修订日期: 2005 04 28

基金项目: 国家自然科学基金课题 (40471111)、863 项目 (2002AA135230 1) 和 973 项目 (2001CB5103) 支持。

作者简介: 王劲峰 (1965), 上海市人, 研究员, 博士生导师。主要从事地球信息科学研究。

Email: wangjf@gsnrr.ac.cn

空间数据通常的不可重复性进一步造成了空间数据分析的特殊性。因此，专门的空间信息分析理论、方法和技术自 20 世纪 60 年代末开始得到认识并研究。

空间数据大体上可分为空间离散或连续型数据（可互相转化），以及多边形数据两大类。自然科学多涉及前者，而社会经济科学多涉及后者。随着人地一体化研究趋势的发展，对两类数据进行综合分析的趋势日益显露。

目前空间数据分析软件包已有不少，主要来自两大学科领域：地理学和地质学。由于地理学和地质学研究对象不同，所涉及的数据特点和分析方法不同，造成两大流派在软件功能、结构、风格上的不同。

在欧美，60 年代经历了地理学计量革命，其主流是试图将社会经济时空格局和过程数学公式化。在 GIS 趋于成熟和空间数据迅速膨胀的技术条件下，当时的学者成为现今地理信息科学的主要推动者，造成地理学者所研发的空间信息分析软件包多带有处理多边形数据（社会经济统计单元，遥感像元数据亦属此类）的特点。相反，源于地质学的空间分析软件包一般适用于分析离散和连续的数据。

空间信息分析理论和技术较为复杂，对于一般科研人员而言掌握难度大，耗费精力多。为此，美英两国都已建立了相应的国家级研发计划，更有甚者，全球销量最大的 GIS 基础软件的研发者美国环境系统研究所在 ARCGIS 8.1 及以上版本已集成了地统计学模块^[7]，目前正在积极争取集成美国国家基金会资助完成的 Lattice 数据分析模块，该功能可以被几乎所有从事地理信息管理的用户所使用，进行空间信息的分析和深加工。

3 技术体系

空间信息分析包括数据获取和预处理（模块 M1）；当进行多源异构数据综合分析时，需要进行属性数据空间化和尺度转换处理（模块 M2）；然后进行空间数据探索分析（模块 M3）；之后根据属性数据的空间存在格式（空间连续分布或离散分布、规则或不规则多边形分布、多源复杂异构信息），分别引导建立地统计模型（模块 M4）、格数据分析模型（模块 M5）、复杂信息分解和预报模型（模块 M6）、以及其他（见图 1）。M1 至 M3 属数据预处理范畴，M4 至 M6 属问题建模范畴，对建立的模型 M4 - M6 需要求解，机理解释及预报。

图中各模块可独立或联合运行，这取决于所研究的问题和对象（所需输出），以及研究人员当前所处的分析阶段（输入）。具体研究内容如下：

模块 1. 空间数据获取和预处理

空间数据采集与尽可能完备化是所有工作重要的第一步，主要利用采样、插值、缺值方法。

（1）空间信息抽样估计 在地学空间应用领域，如环境监测、资源调查、社会统计等，空间信息抽样估计问题大量存在^[8,9]。空间抽样针对地学对象普遍存在的空间关联性和先验信息，从样本选取方式、空间关联性及精度衡量三方面对空间信息获取提供符合统计假设的新的解决思路。

全国耕地面积动态调查，原先穷尽总体的方法（卫片和航片全覆盖），既耗时又花费巨大，报告形成缓慢；而采用 Sandwich 空间采样模型^[10]后，只要利用先验信息并辅以少量样本，就可以同时形成全国 2700 多个县的分县报告，既节省大量时间和开支又能满足精度要求。在此基础上建立空间抽样优化决策通用集成模型，解决地学分析及资源调查

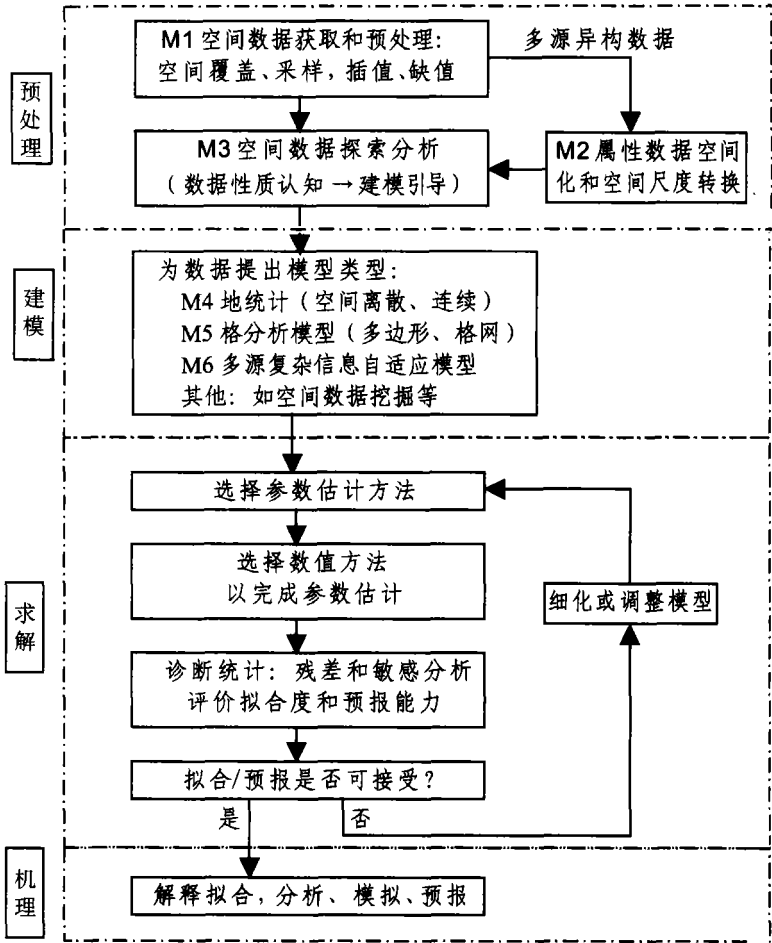


图 1 空间数据分析技术和计算程序

Fig 1 Spatial data analysis techniques and algorithm

中空间抽样问题。软件实现上，基于 Microsoft Windows 平台，Visual 系列软件组件对象方法系统集成^[11]。

(2) 空间数据插值与缺值 根据已知空间样本点（例如气象台站、野外调查）数据进行插值^[12, 13]或推理^[14]来生成面状数据或估计未测点数值是地学研究经常遇到的问题。理解初始假设和使用的的方法是空间插值过程的一个关键，为不同空间过程选择不同插值方法^[15, 16]。

缺值问题与插值类似，有部分交叉，以 Bayes 先验概率为其特征。插值有点、面之分，对于面插值，经过预处理（如去除趋势特征等）可以进行缺值分析；对于点插值，经过预处理（如构建泰森多边形再去除趋势特征等）也可以使用缺值方法。对于缺值的补整，如果具备某些时空特征，则完全可以使用插值方法补整。

模块 2 属性数据空间化与空间转换

自然与人文要素结合研究的时代正在到来。美国宇航局（NASA）将地表人类活动模式加入全球变化模型中从本质上提高了预报精度。又如，自然灾害的灾情空间预报需要自

然要素和社会经济要素相互作用的计算，自然要素可以通过遥感获取，社经要素需要根据统计数据空间细化。地球生态环境以及社会经济数据通常是具有不同形状和尺度的地理空间单元，需要建立属性数据空间化及空间尺度转换技术，其核心是非空间信息或更大空间单元的属性数据在（较小）空间上表达的理论和方法，或称可变面元问题（Modifiable areal unit problem）（MAUP），包括三类技术：

GIS 方法。实现地理空间单元间属性数据的转化，包括聚集、拆分和空间建模（Spatial modeling）。聚集主要解决从小区域（点）向大区域（面）转化问题；拆分则考虑从大区域向小区域转化问题。前者可利用空间采样技术实现。

尺度转化理论及方法。不同的时间和空间尺度限制了信息被观测、描述、分析和表达的详细程度。尺度转化存在“自上而下 scaling down”和“自下而上 scaling up”两种基本方式。不同尺度水平上，尤其是 Downscaling 的转化，存在其区域边界规则与地理边界规则的差异、数据转化的控制和实施形式。

小区域统计学理论及方法。所谓“小区域”，本质上是指区域内样本点较少，因此在统计分析过程中，需要从相关区域“借力”来获得详细的信息，其核心是建立相关区域（数据）的联系模型，实现属性数据空间表达。

以上内容着眼点有所不同，但都可归结为属性数据从大区域向小区域转化，最终完成综合的属性数据空间化方法和软件包设计与实现，并对表达结果的精确性和不确定性做定量分析和解释。

模块 3 空间信息探索分析

探索数据分析（Exploratory Data Analysis，EDA）目标是最大化分析者对数据集及其隐含结构的洞察，揭示数据属性，用以引导选择合适的数据分析模型。EDA 主要采用图形技术进行以下分析：（1）最大化对数据集的洞察；（2）揭示数据隐含的结构；（3）提取重要的变量；（4）检测异常点；（5）检验隐含的假设条件；（6）发展精简的模型；（7）决定因素的最优设置。

空间数据探索分析（Exploratory Spatial Data Analysis，ESDA）是 EDA 的扩展，用来对具有空间定位信息的属性进行分析，包括：探索数据的空间模式^[17]，对假设数据模型、模型基础和数据的地理性质进行阐述，评价空间模型等。ESDA 技术同样要求可视和健壮，强调把数字和图形技术与地图联系。对“某些事件在地图上的什么位置”^[18、19]、“这部分地图的属性值在概括统计中处于什么位置”、“地图的哪些区域满足特定的属性要求”等问题的回答是很重要的。本模块具体内容包括：

探索工具。软件实现当前已经有 ESDA 技术，并且实现表（Table）、地图（Map）和图形（Figure）各窗口链接，支持同时打开四个窗口：数据表窗口（显示当前数据和在过程中创建的变量），地图窗口，图形窗口和文本输出窗口（返回 ESDA 输出）。

区划（regionalization）。空间单元有时很小，比如一个村或乡镇，如果以此进行自然灾害发生率或疾病发生率等小概率事件，统计对这种小的统计单元很敏感。由于空间信息分析要求各空间单元内的信息变差具有相对一致性，因此，经常需要对空间单元的大小进行调整，通过模型计算，实现统计单元的均质性（homogeneity）、平等性（equality）和空间紧凑性（compactness）。

模块 4 地统计模块

点状数据或空间连续分布数据（直接获得或由点状数据通过空间插值或趋势面模拟获

得)是空间数据的一种主要存在形式(例如气象台站、地质物探、生态环境等数据)^[20]。主要包括以下几方面:

变异函数(variogram)。变异函数可以说是整个地统计学的基础,但以前人们只注重了点的变异函数。随着尺度问题的提出,变异函数的正则化(regulization)问题提上了日程。

克里格方法(Kriging)。以往把Kriging法看成是地统计学的全部。通过它人们找到了一种最优的、无偏的估值方法。长期以来,人们仅仅把它看作是一种插值方法。其实不仅如此,Kriging还可以给出每一点的不确定性、利用多种附属信息填补采样不足的缺陷等。

模拟方法(Simulation)。这是当今地统计领域中最为活跃的一种方法。由于Kriging方法具有平滑作用,即将极大值偏小估计、极小值偏大估计,使得整幅图像趋于平缓。但有时人们需要的恰恰正是那种极值,例如在水污染中,就是要通过污染金属所表现的极值来发现污染源。此外,Simulation通过多种实现(realization)系统的表现,研究问题的各种情况,具有较好的统计效用。

模块5 格数据分析模块

多边形数据(社会经济统计数据格式),包括规则格网数据(遥感),是空间数据的又一种主要存在形式。通过空间自相关和协相关模型,找出研究对象在空间布局上的联系与差异,以及空间多元解释变量。例如土地利用变化的环境和人文经济驱动因子识别,环境污染的空间动态建模,为预报和调控提供科学依据。空间局域统计分析技术,则可找出空间热点(hotspots)问题区,应用于疾病、犯罪等空间格局的热点诊断和预报。

空间回归模型。空间回归分析技术可以用于探讨估计值的空间关系,地学对象的空间依赖性和空间异质性使一般回归方法不适应在空间上的分析。空间回归分析有三种特殊形式:(1)联立自回归模型(Simultaneous Autoregressive Model, SAR);(2)空间移动平均模型(Spatial Moving Average Model, SMA);(3)条件自回归模型(Conditional Autoregressive Model, CAR)。

局域统计模型。局域统计分析模型对研究区域内距某一目标单元一定距离的空间范围内所有点的值进行分析。计算指定距离内的空间关联度,从而监测空间内的热点区域,并通过G检验。

模块6 多源复杂时空信息的分解、融合、预报模块

环境、地学、社会、经济、交通、疾病、遥感监测等时空信息大多是多源复杂因素综合作用的结果。通过观测信息反演地学过程机理是地球科学,特别是地球空间信息科学的基本任务之一,目前可用的数学方法有统计^[21]、神经网络、小波分析、遗传算法、元胞自动机^[22]等,也是目前数据挖掘的基本手段。但这些方法只能进行单一成因要素(团)提取和简单过程预报,要实现基于时空信息复合过程的科学预报,不仅需要实现多源复杂因子的分解,更需要将多种分解模式重新组构融合,形成一个新的整体模式。

另外,气象台站、生态网络、污染监测、野外采样、社会经济调查等等,均是空间分散的点状数据。根据已知点的值推断未知点的值,以及由这些点推断整个区域状况是地学研究的经常性任务。目前使用的是空间几何插值、简单加权平均的办法,和新近的考虑全局和局域关联信息的空间统计方法。实际上,前人的研究已经获得了很多半结构化的地学机理和规律,这些先验信息与空间统计的结合有望从本质上提高空间估计预报的精度。借

助贝叶斯方法建立统计和机理结合的空间扩展模型是一种可能的新途径。

总而言之, 空间信息分析是指对地球空间信息进行分析模拟和预测, 以及根据空间信息进行时空运筹的技术、方法和理论。下表给出了从输入输出角度理解什么是空间信息分析模型及其举例。

表 1 空间信息分析模型的输入和输出

Tab. 1 Input and output of spatial information analysis model

举 例		输出信息属性	
		空间	非空间
输入信息属性	空间	热点探测、插值	空间采样模型 (Ups caling)
	非空间	属性数据空间化 (Dow n scaling)	经典统计学

4 若干创新和前沿

(1) 空间抽样模型 目前空间抽样模型是针对空间连续分布对象(如坡面降水)提出的, 对于如大区域耕地问题等空间离散分布地物调查, 缺乏专门的离散空间抽样模型, 虽然可以借助面积成数进行换算; 另外, 对小样本问题、流动地物(海洋)、多维地物(降雨、温度、土壤含水量等多要素)联合调查和动态空间(植被变化)采样问题需要发展相应的空间抽样模型。

(2) 属性数据空间化 属性数据空间化的系统研究在国际上刚刚开始, 有影响的研究成果较少, 主要限于 GIS 的空间分析、尺度变化研究领域。小区域统计学与 GIS 的结合最近开始得到广泛的关注。

(3) 空间尺度转换 Statistical Dow nscaling 基于不同尺度空间变量的统计关系来建立, 在全球环流模型(GCM)中有广泛应用。该方法假定变量之间的统计关系不随尺度变化而改变。Dynamical Dow nscaling 也集中在 GCMs 领域。因研究对象不同, 需要分别设计模型, 在社会经济数据方面研究则较少。

(4) 随机模拟 除 Kriging 法外, 现在的地统计学还可以作为评估未知值不确定性的工具, 这主要通过模拟来完成。即通过条件模拟产生多幅模拟图象用以再现空间依赖的分布格局等统计特性。模拟图像是地统计学领域中最具潜力的领域。

(5) 格数据模型 遥感和社会经济属性信息均以规则和不规则多边形为其载体。随着计算技术迅速发展起来的格数据分析技术(Lattice data analysis)^[23, 24], 在美国国家基金支持下, 自 1999 年开始也设立了五年的研究计划, 由 Sergio Rey 和 Luc Anselin 牵头, 其核心就是空间计量经济学模型分析软件包的构建, 在该研究计划中, 设定了软件包开发计划, 同 ESRI 公司合作, 开发成熟的、具有友好用户界面的商业化空间统计与空间分析软件包。

(6) 复合信息模型 我们经常希望从得到的空间信息反演地学过程的各种影响因子, 例如遥感影像的像元值经常是地表混合物体以及地物多种物理化学几何属性综合作用的结果; 全球气候变暖受自然和人文因子的双重作用, 孰重孰轻分解清楚对于制订科学的环境政策至关重要; 生态环境的空间动态变化受温度、降水、人类活动空间格局变化的多重影响, 准确分解之将有助于对生态环境空间动态进行预报。

功率谱和小波可以分解出不同尺度的结构特征;多元回归通常是全局性的;神经网络对于短记忆非线性较为有效;遗传规划可以对多个给定模式进行优化组合。但是,空间数据经常受多种因子影响,在缺少对其生成机理了解之前,不应人为地套用某个单一模式,各种信息都可能隐含在其中,应该建立多模式融合模型,让数据自己说话。

5 应用领域举例

(1) 灾情损失快速准确评估 重大灾情快速评估涉及受灾地区自然和社会经济情况的综合信息^[25]。遥感等技术为获得受灾地区的致灾因子信息提供了有效的途径,但社会经济统计数据更新周期较长,不能满足实时需要,并且是以行政单元为存在形式,与灾害的地域单元不一致。属性数据空间化技术与抽样技术结合有望弥补当前灾情评估中社会经济数据难以快速和准确获得的缺陷,为灾情损失评估提供技术保障。

(2) 人类健康和疾病研究领域 通过遥感和地理信息系统手段对截面数据如疾病分布、空间扩散^[26]、时空演变^[27]与自然和经济环境状况的分布进行分析。公众的健康问题很大程度与当地的环境因素有关,对于某些关键的疾病以及医疗设施的分布研究也可通过空间数据的分析和处理进行。社会学家、经济学家、人类学家以及环境工作者也将公众健康安全与有毒废物、低收入人口乃至少数人种分布联系起来。

(3) 公共安全领域 犯罪发生率的分析,开始应用空间统计与分析手段,例如,根据美国的犯罪数据分析,盗窃、抢劫以及其他类别的暴力犯罪与当地的城市中低收入、高失业率以及人群混杂的区域高度相关。利用空间统计和空间分析等手段,可以对犯罪率高发区集中控制以及进行社会经济调节,维护公共安全。

(4) 其他 统计分析在气候变化、环境退化、区域可持续发展、人类活动与环境及气候变化的关系研究、社会经济空间格局动态变化^[28]、城市扩张、商业、迁移、旅游、交通和通讯网络分析中广泛的应用,空间分析技术有望挖掘更多的信息和知识。

由于空间信息分析技术的复杂性,实际运用基本限于商业软件所提供的功能,因此适应于各种实际问题的空间信息分析理论和软件的研发是受欢迎的,例如 SpaceStat, GesLib, CrimeStat, 和由国家 863 计划支持于 2004 年完成的国产空间分析软件包原型 Simple0.1^[29]。

6 结论和讨论

空间信息来源于地球科学、社会经济、生命科学、资源和能源、军事等等广泛领域,亦是认识和调控这些系统的基本根据。数据分析的理论是概率论、数理统计和数学,但是由于空间过程广泛存在的相互关联性以及空间信息在空间上特有的多种存在形式与已有经典数学的假设条件常常产生矛盾,或者在理论上没有被后者覆盖,因此产生了专门用于空间信息分析的理论研究^[30]。当今空间信息分析技术至少包括以下六个主要方面:(1) 空间数据获取和预处理;(2) 属性数据空间化和空间尺度转换;(3) 空间信息探索分析;(4) 地统计;(5) 格数据分析;(6) 复杂信息反演和预报模块。应用需求、计算技术进步、空间数据激增、学科交叉是空间信息分析技术和理论发展的主要动力。空间信息分析技术的目的是建立研究对象的(统计)数学模型、因子识别、预测和调控。

数据挖掘或者机器学习技术首先产生于计算机科学和生物信息分析^[31],以概率统计为理论根据,迅速被地球信息分析所借鉴,发展出来的生物信息科学和地球信息科学在数

学方法上大体有 70% 的重合, 虽然研究对象完全不同。数据挖掘技术的目的是通过信息发现研究对象的知识、规则、关联和模式。

当空间过程的机理比较清楚, 并且可以据此建立数学模型时, 例如大气环流模式、海洋物理场、地球动力学等, 统计类理论就降低到次要地位。但是由于真实世界的复杂性和或多或少的不确定性, 需要确定性的数学模型和不确定性的统计模型的结合。

确定性数学模型、空间信息分析技术和数据挖掘技术是分析空间过程的三种不同的手段。对于传染病时空过程建模, 既可以用确定性的微分方程, 也可以用不确定的随机过程, 当后者取极限时将逼近前者, 达到形式统一; 空间信息分析技术用于识别传染病传播的影响因子, 而数据挖掘技术被用于反演传染病参数。三种方法灵活使用, 互相补充, 可以揭示空间过程的多方面特征。

虽然出发点不同, 但是研究对象都是空间过程, 所以空间过程数学模型、空间信息分析技术和数据挖掘技术等发展最终融合形成统一的空间数学 (spatial mathematics) 是有可能的。Matlab, SAS, Spss 等软件包的开发大大推动了数学在各领域的广泛应用, 空间信息分析技术的广泛使用亦需要开发空间分析软件包。

参考文献:

- [1] Buliung R, Kanaroglou P. On design and implementation of an object relational spatial database for activity/travel behavior research. *Journal of Geographical System*, 2004, 6: 237~262
- [2] 王劲峰, 李连发, 葛咏, 等. 地理信息空间分析的理论体系. *地理学报*, 2000, 55(1): 92~103
- [3] Fischer Manfred, Getis Arthur. *Recent Developments in Spatial Analysis*. Berlin: Springer, 1997.
- [4] Haining Robert. *Spatial Data Analysis, Theory and Practice*. Cambridge University Press, 2003
- [5] Anselin Luc. *Spatial Econometrics: Methods and Models*. Dordrecht: Kuwer Academic, 1988
- [6] 柏延臣, 李新, 冯学智. 空间数据分析与空间模型. *地理研究*, 1999, 18(2): 185~190
- [7] ESRI, <http://www.esri.com/software/arcgis/extensions/geostatistical/index.html>
- [8] 梁进社, 张华. 土地利用变化遥感监测精度评价系统——以随机抽样为基础的方法. *地理研究*, 2004, 23(1): 29~37.
- [9] 郑袁明, 陈同斌. “北京市土壤和蔬菜重金属污染与风险评价”取得重要进展. *地理研究*, 2004, 23(5): 455
- [10] Wang J, Liu J, Zhuang D, *et al*. Spatial sampling design for monitoring cultivated land. *International Journal of Remote Sensing*, 2002, 23(2): 263~284.
- [11] 李连发, 王劲峰, 刘纪远. 国土遥感调查的空间抽样优化决策. *中国科学 D 辑*, 2004, 34(10): 975~982
- [12] 郑红星, 刘昌明, 王中根, 等. 黄河典型流域分布式水文过程模拟. *地理研究*, 2004, 23(4): 447~454
- [13] 廖顺宝, 李泽辉. 积温数据栅格化方法的实验. *地理研究*, 2004, 23(5): 633~640
- [14] 周斌, 王繁. 基于决策树模型的土壤性质空间推断. *土壤通报*, 2004, 35(4): 385~390
- [15] 朱会义, 刘述林, 贾绍凤. 自然地理要素空间插值的几个问题. *地理研究*, 2004, 23(4): 425~432
- [16] 李新, 程国栋, 卢玲. 空间内插方法比较. *地球科学进展*, 2000, 15(3): 260~265
- [17] 邵全琴, 戎恺, 马巍巍, 等. 西北太平洋柔鱼中心渔场分布模式. *地理研究*, 2004, 23(1): 1~9
- [18] 杨林生, 陈如桂, 王五一, 等. 1840 年以来我国鼠疫的时空分布规律. *地理研究*, 2000, 19(3): 43~248
- [19] WHO. *Disease Mapping and Risk Assessment for Public Health Decision making*. EUR/ICP/EHRO 02 03 01, 1997.
- [20] 朱华忠, 罗天祥, Christopher Daly. 中国高分辨率温度和降水模拟数据的验证. *地理研究*, 2003, 22(3): 349~359
- [21] 徐建华. *现代地理学中的数学方法*(第二版). 北京: 高等教育出版社, 2002
- [22] 黎夏, 叶嘉安. 知识发现及地理元胞自动机. *中国科学 D 辑*, 2004, 34(9): 865~872

- [23] Ying Longgeng. Understanding China's recent growth experience: a spatial econometric perspective. *Ann. Reg. Sci.*, 2003, 37: 613~628
- [24] Chen Shupeng, Chenghu Zhou. New generation of grid mapping. In: *Proceedings of the International Conference on Geoinformatics and Modelling Geographical Systems*. Beijing, 2004. 15~23
- [25] Wang J, Wise S, Haining R. An integrated regionalization of earthquake, flood and drought hazards in China. *Transactions in GIS*, 1997, 2(1): 25~44
- [26] 颜维安, 冯建良, 杭德荣, 等. 用地理信息系统评价垃圾填埋场对环境污染及危害的研究. *中国卫生工程学*, 2004, 3(3): 130~132
- [27] Dankmar Bohning, Ekkehart Dietz, Peter Schlattmann. Space time mixture modelling of public health data. *Statistics in Medicine*, 2000, 19: 2333~2344
- [28] 刘纪远, 张增祥, 庄大方, 等. 20 世纪 90 年代中国土地利用变化时空特征及其成因分析. *地理研究*, 2003, 22(1): 1~12
- [29] Simple 0.1. <http://autolib.homebj.com/>
- [30] Michael F Goodchild, Robert P Haining. GIS and spatial data analysis: converging perspectives. *Papers Reg. Sci.*, 2004, 83: 363~385
- [31] Pierre Baldi, Soren Brunak. *Bioinformatics: The Machine Learning Approach*. MIT, 2001.

Techniques of spatial data analysis

WANG Jin feng, WU Ji lei, SUN Ying jun, LI Lian fa, MENG Bin

(State Key Lab of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China)

Abstract: GIS is a geo-referenced data based system for manipulating digital spatial and thematic data. Because of its unique properties of spatial data (spatial dependence and spatial heterogeneity), there is an urgent need to investigate the mechanisms of spatial processes through exploring spatial data. The techniques of spatial analysis can be categorized into six groups: (1) generation of spatial data and its preliminary treatment; (2) spatialization of attribute data and scale transformation of spatial data; (3) visualization and exploration of spatial data; (4) geo-statistics; (5) analysis of lattice data; and (6) reversing information from complicated spatial data. To meet the real world solutions, a procedural diagram of spatial data analysis is suggested as below: (1) preliminary data treatment; (2) modeling; (3) solution; and (4) interpretation. Various techniques are being employed to investigate spatial processes, the identical study objective and some common characteristics of the study object may lead to the integration and convergence of the various approaches, and further the introduction of a general science: spatial mathematics. A same important issue is to develop robust spatial analysis computer packages for public uses.

Key words: GIS; spatial data; spatial process; spatial analysis; implications of mechanism