

文章编号: 1007-6301 (2003) 03-0219-010

# 空间数据分析技术在公共卫生领域的应用

武继磊<sup>1</sup>, 王劲峰<sup>1</sup>, 郑晓瑛<sup>2</sup>, 宋新明<sup>2</sup>, 孟 斌<sup>1</sup>, 张科利<sup>3</sup>

(1 中国科学院地理科学与资源研究所, 北京 100101;

2 北京大学人口研究所, 北京 100871;

3 北京师范大学资源与环境科学系, 北京 100875)

**摘 要:** 空间数据分析技术是分析具有空间属性的事物之间相互关系, 对空间信息进行认知、解释、预测及调控等, 而公共卫生研究领域集中在疾病时空规律的认识、成因分析以及干预和防御等领域。本文从公共卫生研究的方向出发, 论述了空间数据分析技术在公共卫生领域的应用, 以及现阶段空间数据分析软件的发展状况, 并对空间数据分析技术在公共卫生领域的应用前景作了展望。

**关 键 词:** 空间数据分析; 公共卫生

**中图分类号:** R 188; N 945. 1

1853 年, 一场严重的霍乱传染病在伦敦爆发, 当时连霍乱疾病的传染模式乃至传染体都不清楚, 更无从谈起有效的治疗方法。然而, John Snow 通过细心分析所获得的数据, 绘制其所知的每例疾病发生的地点, 结果发现, 病例发生地均集中在 Broad Street 街道的一个公用饮水井四周。接下来便是公众卫生研究历史上记载的事情, 水井被移除, 霍乱病也迅速消失了<sup>[1]</sup>。

该案例现已成为公众卫生与空间数据分析技术综合研究的典型案例。这一研究结果表明, 在缺乏对疾病完全了解的情况下, 周密检测发病之间的状况和细心的疾病研究也可以成功地预防疾病。John Snow 成功地揭示了现在对于疾病检测的三个要素, 即: 人、地点、时间之间的有机联系, 尤其是地点要素。

基于一定环境下的人群及疾病分布是公共卫生的主要研究对象, 而人的生活环境和活动范围都在一定的环境条件制约之下, 人群疾病的发生更是受周围环境的影响。因此, 公共卫生领域的研究本身是和地理环境因素紧密结合在一起的。本文拟从空间数据分析技术体系出发, 分别对其在公共卫生领域的应用加以讨论, 同时也论述了可用于公共卫生领域的空间数据分析软件包的开发现状, 并对其应用前景进行了总结和展望。

收稿日期: 2003-02; 修订日期: 2003-03

基金项目: 本研究是在国家 973 项目 (2001CB5103)、863 课题以及中国科学院地理科学与资源研究所知识创新工程和国家自然科学基金支持下进行的。

作者简介: 武继磊 (1976-), 男, 山东菏泽人, 博士生, 研究方向为 lattice 统计与分析, GIS 与空间建模等。Email: wujl@lreis.ac.cn

## 1 空间数据分析技术和公共卫生研究领域

### 1.1 空间数据分析技术

随着信息技术的发展,人们对空间数据的获取和处理能力已经得到了极大的丰富和加强,利用这些数据及其相应的分析技术,能更好地认识和把握自然、社会及环境的运动规律,进行过程的重现、科学预测以及调控。所谓的空间分析,是指用于分析具有空间属性事物的一系列技术,其分析的结果依赖于事件的分布<sup>[2]</sup>,面向公众和私人最终用户,其主要目的在于:空间信息的认知,即有效获取和科学描述空间数据;空间信息的解释,即对空间过程进行理解和合理解释;空间信息预测;空间上发生事件的科学调控等<sup>[3]</sup>。

### 1.2 公共卫生研究的领域

公共卫生的概念是在医学研究人员长期对疾病分析和认识的过程中形成的。它是运用医学、工程学和社会科学等领域的各种成就,用以改善和保障人群的健康、预防疾病的一门学科。就公共卫生研究的领域,通常来说,包括以下三个方面:

#### (1) 人群疾病的时空分布规律

即认识人群疾病发生的时间和空间的分布规律,或者疾病的发生风险以及在时间和空间上的分布状态。公共卫生研究人群中的疾病现象和健康状态,人群是环境的组成要素,人群的疾病与健康现象不可避免地与环境紧密相连。环境要素在空间上表现出不同的区域差异,一方面表现为对疾病发生风险的影响,即各种人群疾病的发生与扩散;另一方面,也对人群本身的免疫抵抗能力产生影响。人群疾病的时空规律分析,要分别从疾病本身的发病机制、传播媒介和人群的暴露强度、持续时间等两个角度来认识。

#### (2) 疾病成因分析

疾病成因分析,即分析导致疾病发生的危险因子或者病因。随着人们对健康概念认识的不断深化,健康与疾病是由多因子作用所决定已成共识。所谓疾病的危险因子是使疾病的发生可能性增加的因素;而病因则是危险因子的数量增多,性质恶化到一定程度的产物。因此,疾病的危险因子是疾病成因研究的一个最重要的内容。

需要指出,导致疾病发生危险因子的分析和识别,一种方式是在医学实验室内可控条件下,逐一因子试验来完成;而另一种方法则需要通过对人群及其环境各要素等多种因子综合分析,这是一般实验条件下所不能完成的,尤其是在复杂环境下,甚至一些疾病本身的发病机制不清楚的条件下,采取有效的办法进行疾病成因分析是公共卫生领域中研究的难点问题。

#### (3) 疾病的预防和控制

疾病的预防和控制则是指在认识人群疾病发生的时空规律或者确定疾病发生的危险因素后,通过风险人群的回避、疫苗接种、患者隔离、医疗资源有效配给等手段进行疾病的防御。其中,疾病发生的时空趋势分析是疾病预防和控制的前提,而医疗资源的有效分配则是疾病预防和控制的保障。近年来,医疗资源的有效分配和地理公平性也成为了公共卫生领域研究的热点之一。

## 2 空间数据分析技术在公共卫生领域的应用

空间数据分析研究的对象覆盖领域广泛, 从空间数据的可视化与探索, 到空间统计及空间计量经济学, 其技术体系特点在于探索、证明空间观测对象之间存在的依赖性。从其研究对象来说, 包括三个大类, 面域空间属性值分析, 即地统计分析; 观测点属性值分析, 如流行病学中的案例分析等, 称为点模式分析; 格网属性值分析, 即附属于研究区域属性值分析。展开空间数据分析技术在公共卫生领域应用的讨论, 还需要从公共卫生研究的主要领域出发, 分别详细论述。

### 2.1 人群疾病时空分布规律认识

地理要素在空间上表现出不同的区域差异, 这些差异会引起环境状态的不同, 如地质变异会引起土壤矿物质含量变化; 土壤的湿度和渗透性等差异会造成土壤的地化学性质不同, 有害物质的扩散也不同; 同时也影响到农业生产方式变化甚至带来的社会关系变异等等。这样, 疾病在不同环境下发生, 也会因其影响而产生空间分布的不同与变异。王敬华对于山西省山阴和应县两地从环境水文地球化学角度分析当地高砷、氟水的原因, 研究表明岩石、土壤、淤泥质沉积和地貌、构造、古地理及水文地质环境分别是当地高砷、氟水必要的物质来源和关键的形成条件, 进而从地理要素上分析砷、氟中毒引起的疾病空间分布<sup>[4]</sup>。

除了疾病在空间上表现出的各种空间差异性之外, 疾病还具有发生、扩散、流行、人类干预和消除的时间延续过程。因此, 对于疾病的分析, 还需要利用 GIS 和空间数据分析技术, 从时间和空间两个角度上进行分析, 找出疾病的时空发展规律。在此方面, Christina Frank 等人对莱姆关节炎(Lyme disease)进行了空间分析, 得出了美国 Maryland 地区 1993 ~ 1998 年, 按照邮政区编码分布的该种疾病的时空分布趋势<sup>[5]</sup>。李海蓉<sup>[6]</sup>, 杨林生<sup>[7]</sup>等人也分别就中国鼠疫的时空分布特点做了分析, 在此基础上, 确立了鼠疫病情的综合评价指标, 为鼠疫的有效控制提供了条件。

从另一个角度讲, 疾病在时间、空间上的分布差异, 而易受疾病感染的人群即风险人群的分布在空间上也具有差异性。F. Cipriani 等人对意大利 1980~ 1990 年期间与饮酒有关的死亡人口进行对照分析, 结果表明, 饮酒引起的死亡率在各个地区之间存在明显的差异, 从而证明了各地区人群免疫性的差别<sup>[8]</sup>。B. Solca 等人对不同区域之间、城市和乡村之间的人口碘摄取量的差异进行分析, 确定碘缺乏疾病的风险人群, 结果发现乡村和城市之间差异甚少, 而区域间的差异甚大, 从而根据区域划定出碘缺乏疾病易发生人群的地域分布<sup>[9]</sup>。武克恭等也对内蒙古地区的砷中毒病区的病人分布差异做了分析比较, 得出该地区风险人群的范围<sup>[10]</sup>。更甚者, Geoffrey T. F 等人则对加利福尼亚地区普鲁氏菌病的时空聚类分布进行了分析。通过对 1973 至 1992 年长达 20 年的数据进行空间分析, 一方面确定了疾病的空间分布, 另一方面也得出疾病与人群的关联性, 即西班牙后裔牛奶消费的生活习惯与该种疾病发生风险的关系<sup>[11]</sup>。

总体说来, 空间数据分析技术在人群疾病时空分布规律的认识上目前已经得到广泛的应用, 但定量刻画疾病的空间分布规律还有待于进一步的深入研究和应用。在此方面, 空间数据分析中关于时空分布规律探索的方法如 Moran I 统计<sup>[12]</sup>、G\* 统计<sup>[13]</sup>等可以很好的

刻画出研究对象在空间上分布的聚集状态和相关程度。

对于 Moran I 统计, 它可以反映出研究对象之间的空间相关关系, 其表述如下:

$$I = \frac{n \cdot \sum_{i=1}^n \sum_{j=1}^n w_{ij} \cdot (y_i - \bar{y})(y_j - \bar{y})}{\left( \sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \cdot \sum_{i=1}^n (y_i - \bar{y})^2}$$

这里,  $n$  表示研究对象的数目,  $y_i$  为观测值,  $w_{ij}$  为研究对象  $i$ 、 $j$  之间的权重, 其中  $w_{ii} = 0$ , 对于权重的确定, 一般有两种方式, 即 1) 二值相邻矩阵, 当  $i$ 、 $j$  相邻,  $w_{ij} = 1$ , 否则  $w_{ij} = 0$ ; 2) 距离矩阵,  $w_{ij} = (d_{ij})^{-r}$ , 一般情况下, 权重矩阵为对称矩阵。对于  $I$  的取值, 一般在 -1 到 1 之间, 当  $I > 0$  时, 表示研究对象之间存在空间上的正向自相关关系, 即空间上研究对象之间的取值, 高高相邻;  $I < 0$  时, 则研究对象之间存在空间上的负向自相关关系, 即空间上研究对象的取值分布存在高低相邻的情况。

而  $G^*$  统计和 Moran I 统计不同, 通过  $G^*$  统计, 可以很好地反映出研究对象在局域空间上的热点区域分布, 其统计模型为:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij}(d) \cdot y_j - W_i \cdot \bar{y}}{S \cdot \{ [nS_{ii}^* - W_i^*] / (n-1) \}^{1/2}}$$

这里,  $S$  是观测值  $y$  的标准方差; 当  $j$  距  $i$  为  $d$  之内, 权重矩阵  $w_{ij}(d) = 1$ , 否则  $w_{ij} = 0$ ;  $S_{ii}^* = S \cdot W_{ij}$ ,  $W_i^* = S \cdot W_{ij}(d)$ ;  $G_i^*$  所得值越高, 则表示在给定距离下, 该点所产生的影响力越大, 由此可以成为研究事物发生的热点区域。

## 2.2 疾病的成因分析

作为环境组成部分, 人类自身的健康状态时刻受到周围环境的影响。任何环境的异常, 包括自然环境和社会经济环境, 如环境中的矿物质或者某一类化学元素含量缺乏或者过多, 以及社会经济等因素导致的心理压力过大或者不平衡, 将会引起人体的不适应或者应激反应, 当超出人体承受限度时, 则表现为疾病的产生。通过空间数据分析技术来对人群生存环境进行分析, 可以在一定程度上有效揭示某些疾病发生的原因或者危险因子。如王宝德通过研究地质环境和生命科学及人体健康之间的关系, 建议进行 1: 20 万尺度下地球化学扫面资料的空间分析, 确定抵制异常区域, 以期为医学工作者在研究区域性疾病的发病机理时提供部分空间分析思路<sup>[14]</sup>。而社会经济等因素对人的健康影响, 如呼吸道疾病、肌肉萎缩、心脏病和关节炎等慢性疾病 (limiting long-term illness, LLTD) 等就与失业率、居住空间、单亲家庭、社会等级以及人口密度等因素有关<sup>[15]</sup>。

定性分析环境与人群健康之间的关系已经作了很多工作, 如在上节疾病时空分布规律的认识以及一些疾病成因的分析等, 但定量刻画环境要素对人群健康之间的关系模型尚处于起始阶段。尽管如此, 定量分析环境致病因子与人群健康之间的关系研究能够弥补实验室缺乏对自然环境条件下多因子对人体健康状况影响分析的不足, 具有综合分析优势。地理信息系统的发展, 大量空间信息和人群健康信息可以集成到一个平台上, 进而为量化分析疾病的成因提供了依据, 在缺乏明确的病因网络关系的状况下, 空间数据分析技术中的空间回归分析方法可以在一定程度上揭示环境致病因子与人群健康之间的定量关系。

由于空间要素之间的关联性, 传统的统计分析中要求的样本独立条件不成立, 因此传

统的统计分析中的回归方法不能够有效、无偏、一致地分析环境致病因子与人群健康之间的关系。空间回归分析起源于空间计量经济学,其发展的核心在于考虑到空间效应条件下的分析方法,空间效应包括空间自相关和空间异质性,因此其工作步骤包括以下四点:制定具有空间效应的计量经济学模型;模型估计其空间效应;考虑空间效应条件下模型的检验;以及模型预测。在空间效应中,考虑更多的是空间自相关性分析。

空间自相关性是指每一个研究对象和其邻近对象之间的属性值具有一定的相关性。可以协方差表达:

$$\text{cov}[y_i, y_j] = E[y_i y_j] - E[y_i] \cdot E[y_j] = 0, \quad \text{for } i \neq j$$

这里,  $i, j$  分别是研究对象的观测点,  $y_{i(j)}$  是观测点获取的值, 当观测点的值即样本点独立时, 协方差应该为 0; 而观测点及样本空间存在空间结构时, 协方差不等于 0。由此对于空间关系的衡量进一步表现为空间自回归 (spatial autoregressive, SAR) 和空间移动平均 (spatial moving average, SMA) 两个过程。即

$$(y - \mu) = \beta W (y - \mu) + \epsilon \quad \text{or} \quad (y - \mu) = (I - \beta W)^{-1} \epsilon$$

和

$$y = \lambda W y + \epsilon \quad \text{or} \quad y = (I + \lambda W)^{-1} \epsilon$$

这里,  $I$  是单位矩阵,  $i$  是  $N \times 1$  维单位向量,  $\mu$  为观测点的均值,  $\epsilon$  为独立、符合正态分布且方差为  $\sigma^2$  的误差项,  $W$  为空间衡量空间关系的权重矩阵,  $\rho$  和  $\lambda$  分别表示了空间自回归和空间移动平均过程的参数<sup>[16]</sup>。

以空间自回归和空间移动平均为特征的空间回归分析已在社会经济领域的分析上得到了广泛的应用, 在公共卫生领域的应用业已逐步开展, 可以定量刻画环境要素与人群健康状况之间的关系。

### 2.3 长期疾病的监测与干预的医疗资源配置

对于人群疾病发生的规律认识或者疾病成因的分析, 其最终目的是人群疾病的预防与干预。一般情况下, 人群中发生疾病的监测与干预存在长期性等特点, 而医疗设施等医疗资源不能够满足充分干预和治疗的目的。由此也就产生了针对人群疾病空间分布和风险人群分布进行医疗资源空间优化配置等问题。如 Vincent J. DeI V asino 对于健康服务空间分布的均衡性进行了讨论, 指出疾病发病区之间均衡配置医疗资源具有重要意义<sup>[17]</sup>, JP Twigger 等也对小区域的医院分布和时间距离之间的关系进行了分析<sup>[18]</sup>, 这些研究均提出了医疗资源优化配置的问题。H.L. Holley 则论述了在财政限制和健康管理分散的情况下, 区域健康需求和服务、医疗设施利用与治疗的不均衡性。指出利用地理研究中的方法, 作区域性研究对于人群健康具有重要意义, 同时也指出了地理分析技术在医疗资源平衡分配中的潜在应用意义<sup>[19]</sup>。

就目前医疗资源的配置方案来说, 其出发点包括两个方面, 第一, 地理可达度的公平性 (geographically fair), 即在给定时间内, 一定区域的人获得有效医疗的条件是平等的; 第二, 干预区域的侧重性 (regional emphases), 一般情况下, 区域人群的发病具有一定的特点, 在有效医疗资源的条件下, 不同的区域对于医疗资源的投入应该是有重点地投入到最大化消除人群疾病、取得最佳健康效益的领域。既要使得人群获得医疗资源分配的公正性, 又要优先保障疾病高发区的医疗资源配给, 因此, 必须采取多目标空间资源优化配置模型进行医疗资源的有效配置。在此方面, 王劲峰等就水资源的空间分配做了深入研究<sup>[20, 21]</sup>, 而

对于医疗资源的分配,同样可以采取制定目标函数、层层优化、变量的逐步释放以及 Lagrange 条件极值的算法,得出空间上的最佳医疗配置方案,表述如下:

设定在区域 ( $i$ )、最小干预时间 ( $j$ ) 和干预疾病种类 ( $k$ ) 之间医疗资源的投入总量为  $R$ ,  $\beta(i, k, R)$  为不同疾病种类在各个区域之间投入一段时间后的健康恢复弹性函数,则医疗资源的最优化分配为

$$HR = \int_{i=0}^{reg} \int_{j=0}^T \int_{k=0}^K \beta(i, k, R) dR(i, j, k) dj di$$

这里,  $reg$  为疾病干预的区域,  $T$  为干预时间。

令  $A$  为医疗资源分配的目标泛函,则医疗资源分配决策的过程即为寻找达到  $A$  的极大值的  $R(i, j, k)$ 。则根据医疗资源分配的出发点,存在不同的医疗资源分配目标。

目标 1: 疾病干预期间,医疗资源分配取得效果最佳,即

$$A = \max_{R(i, j, k)} \int_{i=0}^{reg} \int_{j=0}^T \int_{k=0}^K \beta(i, k, R) dR(i, j, k) dj di$$

目标 2: 区域医疗资源地理公平性,即医疗资源的可达度对人群分布一致,即

$$A = \min_{R(i, j, k)} \left\{ \int_{i=0}^{reg} \int_{j=0}^T \int_{k=0}^K \beta(i, k, R) dR(i, j, k) dj - m \int di \right\}$$

这里

$$m = \int_{i=0}^{reg} \int_{j=0}^T \int_{k=0}^K \beta(i, k, R) dR(i, j, k) dj di$$

目标 3: 对疾病高风险区进行重点干预或者个别疾病种类重点防御,即

$$A = \max_{R(i, j, k)} \left\{ \min_i \int_{j=0}^T \int_{k=0}^K \beta(i, k, R) dR(i, j, k) di, \forall i \right\}$$

当然,对于公共卫生的服务目标多种多样,因此医疗资源的分配也不局限于以上的一个目标,但最终的医疗资源分配可以通过多个目标的组合的优化过程而获得,至于目标方程的求解,可以参考<sup>[21]</sup>。

### 3 公共卫生空间数据分析软件包的开发

空间数据分析技术方法的应用,促进了大量空间数据分析软件及软件包的开发,其中不少软件和软件包在公共卫生领域的疾病分析和干预方面得到了进一步的应用。就这些可应用于公共卫生领域的软件和软件包的开发源头来看,主要包括以下三类:

第一类,以现有的统计分析软件为基础,构造商业化的边缘应用软件(Commercial off-the-shelf, CoTs)。这类软件本身依靠原有商业软件的强大分析功能,可以进行专业的时序统计分析,具有很好的经典统计分析理论基础。同时,原有的公共卫生领域中的卫生统计本身没有考虑到空间要素前,这些软件就得到了广泛的应用,因此结合空间数据分析技术方法特点,对原有统计分析功能模块进行扩充,形成空间数据分析软件包,在公共卫生领域的应用占有较高的比例。如基于经典统计分析大型软件 SAS 系统,结合地理信息系统软件,形成了 SAS/GIS 空间分析软件包,进行了广泛的流行病研究<sup>[22]</sup>;此外,著名的统计分

析软件 S-Plus 也扩展了其空间分析能力, 形成了 EnvironmentalStats for S-PLUS、S-PLUS for ArcView GIS 及 S+ SpatialStats 等空间数据分析模块<sup>[23]</sup>。这类软件的突出特点是基于成熟的商业软件进行扩充和开发, 系统稳定, 但在空间数据获取和空间表达能力上则有诸多欠缺之处。

第二类, 从现有成熟的地理信息系统软件出发, 利用软件本身提供的二次开发语言或者其他开发语言, 进行空间数据分析软件包的开发。现有的商业地理信息系统软件本身一般具有二次开发语言或者和其他软件开发环境相通的控件或接口, 用于功能的扩展。如 ArcInfo 软件提供的 AML 语言、ArcView 提供的 Avenue 语言以及 MapInfo 提供的 MapBasic 语言等, 此外这类软件还提供 ArcObject、MapObject 控件等, 允许用户进行二次开发。剑桥大学的 Robert Haining 等人在 ArcInfo 地理信息系统环境下, 开发出了专门用于人群健康分析的软件包 SAGE, 取得了一定的应用成果<sup>[24]</sup>。Patrick A. Wall 等人则结合地理信息系统软件 MapInfo 的二次开发以及 Delphi 集成开发环境, 开发出空间数据分析的空间关联和聚类分析的软件包 MapStat, 对俄亥俄州的肺癌死亡率进行了空间分析<sup>[25]</sup>。这类软件基于以空间信息管理为显著功能的地理信息系统平台, 因此在空间数据的获取和空间表达能力上具有突出优势, 但由于系统自带二次开发语言过于简单, 对于复杂的统计分析功能实现来说, 需要大力加强。

第三类, 独立开发空间数据分析软件包。这类软件包的开发本身借助基本的软件开发语言或者其他集成开发环境, 形成脱离地理信息系统环境, 同时可从地理信息系统软件中获取空间数据并输出能够被地理信息系统识别和演示数据的软件包。如 Getis & Chen 等开发的空间点模式分析统计软件包 PPA<sup>[26]</sup>, 以及 Anselin 的通用空间分析软件 SpaceStat<sup>[27]</sup>等, 它们分别应用 C 语言开发环境以及 GAUSS 集成开发环境等开发, 融入最新的空间数据分析理论、技术和方法, 能够突出反映空间信息的本质是这类软件的特点, 但就目前这类软件的发展来看, 由于其应用多在科研领域, 没有形成大规模的开发需求, 因此软件在可操作性上需要大力改善, 以上两个软件包为例, 其运行基本还处于 DOS 的工作平台下, 与现阶段的视窗操作 Windows 环境还有一定的距离。

从现有的这些可应用于公共卫生领域的各种软件和软件包的发展来看, 公共卫生空间数据分析软件包也面临着非常大的市场需求。但就此类软件包的应用和推广, 还需要以下几个方面的技术进展:

第一, 研制更加通用的空间分析软件, 使其具备友好用户界面, 降低入门台阶, 鼓励和支持用户利用空间数据进行公共卫生应用领域的研究分析。

第二, 在通用大规模空间数据分析软件形成规模之前, 较高的兼容性也是空间数据分析软件需要注意的一个重要方面, 提高软件处理大数据集的能力, 提供多种数据融合与共享以满足用户需求, 促进空间数据分析软件的发展。

第三, Linux 系统的发展说明开源代码和系统框架, 具有极大的发展优越性。因此, 采取开放式结构、模块化设计、支持重用性, 允许空间数据分析方法的扩充与重构, 从而使得用户可以混合和重用不同的空间数据分析模块, 乃至源代码公开, 允许研究者可以在前人的研究基础上对系统再开发, 这对于空间数据分析软件的发展也具有重要意义。

## 4 结论和展望

数字地球战略的实施和空间信息基础设施的建设,使我们能够快速获取和掌握大量的自然、生态、环境及社会经济等大量数据信息,从而客观上促进了空间数据分析技术的发展及其在各个领域中的应用。其中,空间数据分析已经在公共卫生领域得到了初步成功的应用,展示出了良好的发展态势和应用前景。

然而,从整体上看,空间数据分析技术在公共卫生领域的应用还处于初步阶段,其主要研究领域主要集中在地方病和个别媒介传染疾病的分布监测上,还远没有深入到亟需空间分析的公共卫生诸多领域之中,究其限制原因,主要有以下两点。

第一、空间数据分析技术的有效依赖平台——地理信息系统是在 20 世纪六十年代以来出现并发展起来的,而空间数据分析技术则是近二十年内才获得较快的发展,一些空间数据分析的模型、方法和技术本身尚存在许多不完善之处。尽管空间数据分析在社会和环境分析中已经初步显示出其极大的应用潜力,但地理信息系统和空间数据分析技术本身仍需要拓展其在各领域中的应用范围,并结合各领域自身特点,进一步完善其功能。

第二、空间数据分析技术和方法在公共卫生领域的应用受到数据源的限制。公共卫生研究的发展始于 17 世纪欧洲工业革命时期,当时大量农民向城市集中,带来的环境恶化导致霍乱、痢疾和肺结核等疾病蔓延,居民健康状况下降,在采取环境卫生措施和对疾病作斗争的过程中公共卫生研究才逐步完善起来。公共卫生的发展过程中缺乏相应的空间理论、方法和技术手段,因此,长期以来积累的数据缺乏空间属性,限制了空间数据分析在公共卫生领域的展开。

但是,“地理信息系统可以处理海量数据,和空间数据分析结合,决定了其在分析疾病地理分布模式和社会、自然环境条件的关系中的核心作用,作为疾病的预防和干预决策支持系统的核心,其潜力是不可替代的<sup>[28]</sup>”。因此,随着信息技术的发展,GIS 空间技术的进一步广泛应用和空间数据分析理论、技术和方法体系的完善,空间数据分析将在公共卫生领域做出突出的、不可替代的贡献。

## 参考文献

- [1] Terracini B. Environmental Epidemiology: a historical perspective. In: Elliott P, Cuzick J, English D (eds). *Geographical & Environmental Epidemiology Methods for Small Area Studies*. Oxford, UK: Oxford University Press, 1992: 253~ 263.
- [2] Haining R. *GIS and Spatial Analysis*, Beijing: Lecture series in Chinese Academy of Sciences, 1999.
- [3] 王劲峰等. 地理信息空间分析的理论体系探讨. *地理学报*, 2000, 55(1): 92~ 103.
- [4] 王敬华. 山阴与应县地区高砷高氟水的形成环境研究. *内蒙古预防医学*, 1997, 22(4): 145~ 147.
- [5] Christina Frank, Alan D. Fix, Cesar A. Pena and G. Tomas Strickland, Mapping Lyme Disease Incidence for Diagnostic and Preventive Decisions. *Maryland. Emerging Infectious Diseases*, 2002, 8(4): 427~ 429.
- [6] 李海蓉, 杨林生, 王五一等. 150 年来中国鼠疫的医学地理评估. *地理科学进展*, 2001, 20(1): 73~ 80.
- [7] 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): 243~ 248.
- [8] Cipriani D, Balzi B, Sorso S and Buiatti A. Alcohol-related mortality in Italy. *Public Health*, 1998, 112: 183~ 188.
- [9] B Solca, SE Jaeggi-Groisman, V Saglini and H Gerber, Short Communication Iodine Supply in Different

- Geographical Areas of Switzerland: Comparison Between Rural and Urban Populations in the Berne and the Ticino Regions, *European Journal of Clinical Nutrition*, 1999, 53: 754~ 755.
- [10] 武克恭, 马恒之, 夏雅娟等. 我区中西部地方性砷中毒病区病人分布差异的比较研究. *内蒙古地方病防治研究*, 1995, 20(1): 20~ 22.
- [11] Geoffrey T. Fosgate, Tim E. Carpenter, Bruno B. Chomel etc. Time-Space Clustering of Human Brucellosis, California, 1973~ 1992. *Emerging Infectious Diseases*, 2002, 8(7): 672~ 678.
- [12] R. J. Haining. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge, U. K., 1989.
- [13] Getis A., Ord J. K., 1992, The Analysis of Spatial Association by Use of Distance Statistics, *Geographical Analysis*, 1992, 24: 189~ 206.
- [14] 王宝德. 地质科学与生命科学的关系. *石家庄经济学院学报*, 2001, 24(3): 315~ 324.
- [15] A. Stewart Fotheringham et al., *Quantitative Geography*, SAGE publication, London 2000: 114~ 128.
- [16] Luc Anselin, *Spatial econometrics: methods and models*, Kluwer Academic Publishers, 1989.
- [17] Vincent J. Del Vasino, Jr. Healthier Geographies: Mediating the Gaps between the Needs of People Living with HIV and AIDS and Health Care in Chiang Mai, Thailand, *Professional Geographer*, 2001, 53: 407~ 421.
- [18] JP Twigger and EG Jessop, Small Area Variation in Hospital Admission: Random or Systematic?, *Public Health*, 2000, 114: 328~ 329.
- [19] H. L. Holley, Geography and mental health: a review, *Soc Psychiatry Psychiatr Epidemiol*, 1998, 33: 535~ 542.
- [20] 王劲峰等. 水资源空间配置的边际效益均衡模型, *中国科学(D)*, 2001, 31(5): 421~ 427.
- [21] 王劲峰等. 区域调水时空优化配置理论模型探讨, *水利学报*, 2001, (4): 7~ 14.
- [22] *Working with Spatial Data Using SAS/GIS Software*, SAS Publishing, 2000.
- [23] <http://www.insightful.com/products/addons.asp>.
- [24] Haining R. P., et al. The design of a software system for interactive spatial statistical analysis linked to a GIS. *Computational Statistics*. 1996, 11: 449~ 466.
- [25] Patrick A. Wall, et al. Interactive Analysis of the Spatial Distribution of Disease Using a Geographical Information System, *Journal of Geographical Systems*, 2000, (2): 243~ 256.
- [26] <http://www.geog.ucsb.edu/~dongmei/ppa/ppa.html>.
- [27] Luc Anselin, Computing environments for spatial data analysis, *Journal of Geographical Systems*, 2000, 2: 201~ 220.
- [28] Gerard Rushton, et al. Considerations for Improving Geographic Information System Research in Public Health, *URISA Journal*, 2000, 12(2): 31~ 49.

## A Review on Application of Spatial Data Analysis Technology in Public Health

WU Jilei<sup>1</sup>, WANG Jinfeng<sup>1</sup>, ZHENG Xiaoying<sup>2</sup>,  
SONG Xinming<sup>2</sup>, MENG Bin<sup>1</sup>, ZHANG Keli<sup>3</sup>

(1) Institute of Geography Science and Nature Resource Research, CAS, Beijing 100101, China

2 Institute of Population Research, Peking University, Beijing 100871, China

3 Department of Resource and Environment Science, Beijing Normal University, Beijing 100875, China)

**Abstract:** Spatial data analysis (SDA) technology mostly is concerned with data, which are

not independent spatially and temporally, and the spatial correlation and spatial association among the studied objects are of interest. Spatial correlation mostly refers to the relation among the objects themselves while the spatial association considers the neighborhood effects between the objects. After the correlation and association among objects with spatial attribution are analyzed with SDA technology, spatial pattern will be perceived and explained. Then we can predict and coordinate the spatial processes further.

In public health study area, most data acquired have spatial and temporal attributions, so advanced technology are needed to analysis the pattern of diseases distribution, this is a preliminary and important step to identify the risk factors in environment to the diseases. With the understanding of the patterns and risk factors, the prevention and intervening counter measures to the disease would be easily acquired. In this paper, the author gave a review on application of SDA technology in public health, including the pattern identification, hotspots detection, environmental risk factors identification, and public health resource distribution as well. Moreover, the developments of the software which are useful to public health spatial analysis were discussed also.

Though the progress of SDA technology applied into public health meets some encumbrance, there are more advantages to public health with spatial analysis. And with the fast developments of information technology and geographical information system software's maturation in applications in many fields, spatial data will be enriched and give potential requirement of SDA technology including public health study; then this will take a rapid change in public health study.

**Key words:** spatial data analysis; public health