

文章编号: 1001-4217(2014)04-0061-07

# 空间统计学及其在公共卫生领域中的应用

倪书华

(北京师范大学社会发展与公共政策学院, 北京 100875)

**摘要:** 经典统计学处理公共卫生数据空间属性数据的能力不足, 空间统计学具有优势, 由此得到广泛应用. 目前空间统计学在公共卫生领域主要应用有: 疾病制图、分布格局探测、空间相互关系分析、病因分析以及时空预警等方面.

**关键词:** 空间统计学; 公共卫生; 空间自相关; 空间异质性

中图分类号: R 188 N 945 文献标志码: A

## 0 引言

自从2003年SARS以来, 我国加强了公共卫生信息化建设, 同时新医改提出了宏伟“3521”信息化工程, 积累了大量的疾病与健康相关数据资源, 这些资源是公共卫生宝贵财富, 这些数据资源每日呈指数增长, 标志公共卫生领域已进入了大数据时代; 科学分析这些数据资源有助于掌握疾病的流行规律, 采取有针对性的防控措施, 提高疾病预防控制工作效率, 同时正确的分析结果是进行科学决策的依据; 而这些公共卫生数据大约80%以上具有空间属性, 如果采用传统的统计方法来处理这类数据, 忽略了数据的空间属性; 经典统计分析假设前提是随机独立, 而空间数据往往不满足这些假设, 其结果可能有偏和非最优<sup>[1]</sup>.

## 1 空间统计学概述

在这种背景下, 我们必须思考如何利用空间统计学方法来处理公共卫生疾病监测数据, 充分考虑其空间属性, 使分析的结果更真实、可靠、客观和科学.

### 1.1 空间统计发展历史

1950年Moran首次提出空间自相关测度来研究二维或更高维空间随机分布的现象, 1951年南非学者Krig提出了空间统计学萌芽思想, 后经法国数学家Matheron完善, 于1963年和1967年提出了地统计学和克里金技术. 1973年, Cliff和Ord发表了空间自相关(Spatial Autocorrelation)的分析方法, 1981年出版了Spatial Process: Model and

收稿日期: 2014-08-22

作者简介: 倪书华(1978-), 女, 江苏省张家港市, 经济师, 硕士研究生, 研究方向: 卫生保障.

E-mail: nishuhua@hotmail.com

Application 专著,形成了空间统计理论体系,以及 Getis' G 和 Lisa 提出的空间异质性的局部统计使空间统计理论日趋成熟<sup>[2-3]</sup>. 近年来随着空间分析技术以及空间分析软件(如 GIS、Geoda、SaTScan、Winbugs 等)的迅速发展,与疾病分布有关的空间统计分析也得以较快发展.

### 1.2 空间统计学特征与定义

空间统计具有明显的多学科交叉特征,其显著特点是思想多源、方法多样、技术复杂,并随着相关学科如计算机软硬件技术的发展而发展<sup>[4]</sup>. 空间统计分析是以地理实体为研究对象,空间统计模型为工具,以地理实体空间相关性和空间变异性为出发点,来分析地理对象空间格局、空间关系、时空变化规律,进而揭示其成因的一门新科学. 经典统计学与空间统计学的区别与联系归纳如表 1.

表 1 经典统计学与空间统计学区别与联系

比较	经典统计学	空间统计学
研究的变量	随机变量	区域化变量
假设前提	独立	空间自相关
研究样本	数字特征	空间分布及空间关系
研究结果	没有与 GIS 结合	与 GIS 完美结合
联系	空间统计是经典统计的有效补充	

### 1.3 公共卫生数据的空间特征

空间依赖性和空间异质性是公共卫生数据两个重要的空间属性,公共卫生数据依赖性是指疾病在空间分布上呈现的空间自相关、聚集性及协同的趋势,空间自相关破坏了经典统计当中的样本独立性假设,构成“地理学第一定律”<sup>[5]</sup>. 公共卫生数据异质性是指疾病在空间分布上具有的复杂性、变异性、不均匀性,与地理学第一定律所描述的空间依赖性相对应,Goodchild(2003)将空间异质性总结为“地理学第二定律”. 这两种特殊性质直接影响了空间数据统计分析和建模的方法.

### 1.4 空间统计学主要方法

空间统计学方法很多按照空间数据类型和空间数据特殊性可以分为基于点数据分析方法、面数据分析方法、连续数据分析方法以及空间回归分析等,详见图 1.

## 2 空间统计在公共卫生领域中主要应用

空间统计在公共卫生领域中应用越来越广泛,主要在制作疾病聚集性分析、疾病监测点抽样、发病的估

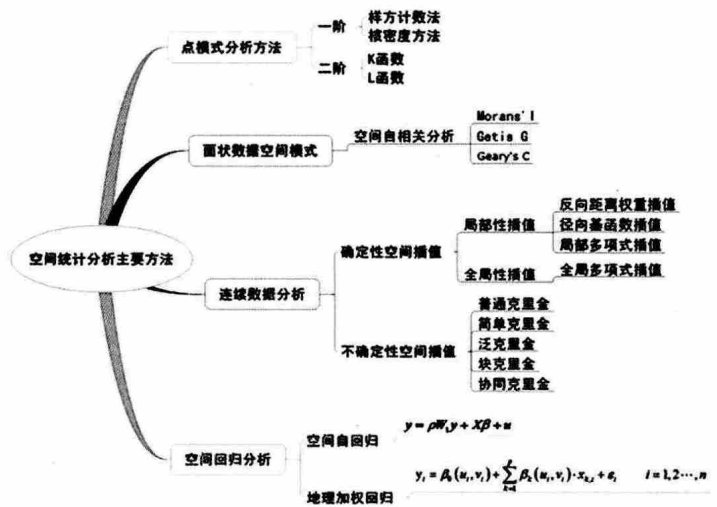


图 1 空间统计分析方法思维导图

算、探讨危险因素、确定疾病或媒介空间特征、预测疾病的时空变化规律等方面。

## 2.1 疾病的空间插值分析

空间插值数据是根据相邻样点的相似原理来生成表面，即用已知的样点的值生成表面来预测整个研究区域内每个位置的值，并评估预测表面的误差和变异性。克里金插值又称之为地统计学，以空间自相关为前提，区域化变量理论为基础，以变异函数为主要工具的一门新学科。其实质是利用区域化变量的原始数据和变异函数的结构特点，对未采样点的区域化变量的取值进行线性无偏、最优估计。半变异函数主要由块金值、变程、基台值、偏基台值几部分组成，半变异函数详解见图 2。

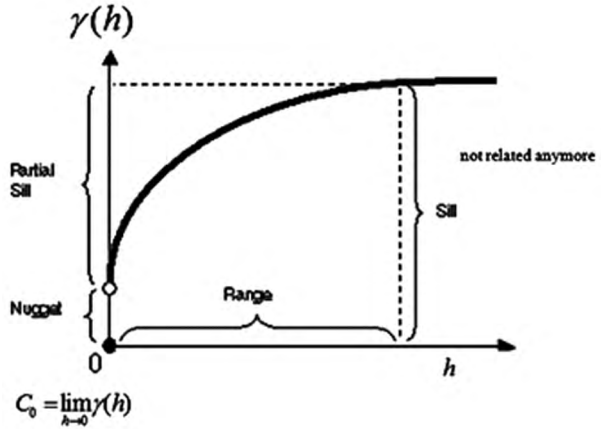


图 2 半变异函数图

块金值(Nugget) $C_0$ 是指疾病空间化变量的随机性，其大小反映随机因素导致的变异部分，是随机变异。基台值(Sill) $C_1$ 表示疾病空间化变量的最大变异取值，即总变异，其大小等于自相关变异和随机变异之和。偏基台值(Partial Sill) $C_1$ 反映疾病空间化变量的结构性，是结构因素引起的变异，其大小表示空间自相关部分引起的变异大小，是空间自相关变异。变程(Range) $a$ 表示疾病空间化变量存在空间自相关的最大空间尺度(距离)，当超过阈值时，疾病空间变异不存在自相关。块金基台比  $[C_0 / C_0 + C_1]$ ，其大小反映空间自相关部分引起疾病空间异质性程度的大小。如果块金基台比较大，说明随机部分引起的疾病空间异质性起主要作用，空间自相关弱。反之，块金基台比较小，说明空间自相关部分引起的疾病空间异质程度起主要作用，空间自相关性较强。描述污染物空间分布是风险评估的一个重要组成部分，Cattle 等人采用指示克里金技术基于已知抽样点污染物浓度去估计非抽样点污染物浓度<sup>[6]</sup>。Asmarian 等人采用泊松克里金插值对 2003–2007 年 336 个县食道癌发病数据进行分析，得出了 Ardebil, Mazandaran and Kordestan 三省与其他省相比有较高的风险<sup>[7]</sup>。Adhikar 等人采用指标和概率克里格方法描述印度德里市 Najafgarh 街区地下水铜、铁、锰污染情况<sup>[8]</sup>。

## 2.2 疾病空间聚集性研究

疾病的聚集性分析目的在于研究潜在危险因素的时空聚集性，从整体上检验疾病的空间分布是随机还是聚集。如果是聚集分布，进一步回答：聚集在什么地方？疾病聚集程度高低及与周边地区关系如何？聚集在多大的空间尺度才有效？这些与地点相关的聚集因素可以是未知的感染因子、地方污染物等。目前我国在手足口病聚集性分析中应用比较多，肖革新等人采用局部 Getis G 热点探测发现了中国大陆 2008–2011 年在县区水平上手足口病空间聚集性及随时间的变化趋势<sup>[9]</sup>。于石成等<sup>[10]</sup>利用时空扫描技术对全国重症手足口病例进行了时空聚集性分析，发现重症病例聚集区域，为进一步研究重症病例成因奠定了基础。Wang 等采用 Satscan 时空扫描工具探测北京 2008–2011 年手足口病

时空聚集模式<sup>[10]</sup>。同时在其他传染病分析中也得到广泛应用,如细菌性痢疾<sup>[11]</sup>、丙型肝炎<sup>[12]</sup>、出血热病<sup>[13]</sup>、HFRS<sup>[14]</sup>、H7N9<sup>[15]</sup>等疾病分析。同样在癌症研究中,研究人员使用空间聚集分析确定地理区域的高危人群,然后筛选人群疾病可以改善癌症控制<sup>[16]</sup>。

### 2.3 疾病时空预警分析

通过时空模型分析不同时期疾病或媒介的空间动态变化,了解疾病随时间的变化规律,对疾病未来的发展趋势做出分析、预测和评估,从而在疾病的预警系统中发挥作用。在这些时空模型中,Kulldorff于1998年提出了时空扫描统计量<sup>[17]</sup>,以及在2001年提出的前瞻性时空重排扫描统计量,在传染病暴发预警中具有较好的应用前景<sup>[18]</sup>。前瞻性时空重排扫描统计量以动态变化的扫描窗口对不同的时间和区域进行扫描,可以有效地对未知的时空聚集性进行探索性分析,达到早期预警的目的。该方法最大的优势在于,由于采用了重排算法,该模型在建模过程中不需要使用人口数据。而基于Poisson分布的时空扫描统计量,在计算过程中需要各区域的人口数据,但确切的人口数据常常很难获得。Mostashari等<sup>[19]</sup>利用死禽数据进行空间统计扫描分析对西尼罗病毒暴发进行早期预警。Mugglin等<sup>[20]</sup>采用贝叶斯时空模型通过对苏格兰流行性感冒病例资料进行分析和预测,可以让医院在应诊能力准备上做得更好。

### 2.4 疾病制图和病因探索

疾病制图是空间统计学的基本功能,其目的是将疾病的危险的空间变异或时空变异在地图上呈现出来,为进一步病因学研究或其他研究提供线索。地理环境相关性研究是研究与环境有关的地理变量(如空气、水体、土壤等)或生活方式等因素与健康之间的相互关系,能够为环境危险因素的研究提供必要的信息<sup>[21]</sup>。自然与社会环境、营养、基因、行为、病媒生物等是许多疾病的致病因子,通常具有空间分布。疾病分布如同自然景观的概念一样,不同疾病的分布结构来自于不同的病因分布结构,这种结构的变化同样影响着疾病的发生、传播和消长的变化。通过分析这些要素和疾病空间分布之间的关系,可以探测是否存在威胁健康的环境危险因素,这些危险因素是否存在交互作用;“地理探测器”(GeoDetector)<sup>[22]</sup>方法基于空间方差分析来探测环境风险因子。地理探测器主要包括四个方面的功能:风险探测器可以指示风险区域;因子探测器可以定量评价不同的环境因子的风险程度;生态探测器可以分析不同的环境因子的影响是否有显著差异;交互探测器可以分析环境因子是独立作用还是多种因子交互作用的结果。自从地理探测器问世以来,在公共卫生领域得到了广泛的应用。Wang等<sup>[23]</sup>利用了地理探测器发现了引起山西省和顺县神经管畸形的环境致病因子以及致病因子之间的相互关系。Wu等<sup>[24]</sup>通过空间探索分析为获得导致出生缺陷的环境致病因子,为进一步分析致病因子奠定了基础。Hu等<sup>[25]</sup>利用地理加权回归分析了气象因子对我国手足口病影响在地理空间上的异质性。Deng等采用空间面板模型分析广东气象因素对手足口病的影响<sup>[26]</sup>。层次贝叶斯时空模型方法在疾病制图及分析疾病的影响因素方面取得较好的效果,获得了疾病危险的空间及时空变异的信息<sup>[27]</sup>。

## 3 结论及展望

空间统计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<sup>[28]</sup>。相比经典统计学更充分考虑疾病与健康的地理空间背景信息，研究结果更加客观、真实、科学，未来将成为当前公共卫生领域用来认识和揭示疾病流行规律的一种重要方法和工具。而且它有效集成了GIS可视化分析功能，有助于推动其他学科知识（如环境学、地理学、生态学、气象学、社会学、经济学等）向公共卫生领域渗透，促进了公共卫生与其他学科协同融合发展。数字地球战略的实施和空间信息基础设施的建设，3S技术不断推广应用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，从客观上促进了空间统计技术发展及其在各个领域中的应用<sup>[29]</sup>。随着公共卫生大数据时代到来，科学研究进入第四范式即数据密集型发现<sup>[30]</sup>，空间统计体现了跨界思维和系统科学，尤其将成为解决位置信息大数据的一种非常重要方法，必将成为大数据挖掘的重要方法和工具，已成为解决宏观公共卫生面临主要难题的一个重要的方法和手段，将成为公共卫生学术界共同努力研究的方向，展示了美好的发展态势和广阔的应用前景。

#### 参考文献:

- [1] 王劲峰, 廖一兰, 刘鑫. 空间数据分析教程 [M]. 北京: 科学出版社, 2009.
- [2] Anselin L. Local indicators of spatial association-LISA [J]. *Geographical Analysis*, 1995, 27(2): 93-115.
- [3] Getis A, Ord J K. The analysis of spatial association by use of distance statistics [J]. *Geographical Analysis*, 1992, 24(3): 189-206.
- [4] 赵永, 王岩松. 空间分析研究进展 [J]. *地理与地理信息科学*, 2011, 27(5): 1-8.
- [5] Harvey J M. Tobler's first law and spatial analysis [J]. *Ann Assoc Am Geographers*, 2004, 94(2): 284-289.
- [6] Cattle J A, McBratney A B, Minasny B. Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination [J]. *Journal of Environmental Quality*, 2002, 31(5): 1576-1588.
- [7] Asmarian N S, Ruzitalab A, Amir K, et al. Area-to-area poisson kriging analysis of mapping of county-level esophageal cancer incidence rates in Iran [J]. *Asian Pacific Journal of Cancer Prevention*, 2013, 14(1): 11-13.
- [8] Adhikary P P, Dash C J, Bej R, et al. Indicator and probability kriging methods for delineating Cu, Fe, and Mn contamination in groundwater of Najafgarh Block, Delhi, India [J]. *Environmental Monitoring and Assessment*, 2011, 176(1/4): 663-676.
- [9] 肖革新, 胡跃华, 马家奇, 等. 中国 2008-2011 年手足口病空间聚集性及变化趋势研究 [J]. *中华流行病学杂志*, 2012, 33(8): 808-812.
- [10] 于石成, 周征奇, 杨芳, 等. 中国重症手足口病的时空聚集性及其变化趋势分析 [J]. *中华流行病学杂志*, 2014, 35(3): 43-47.
- [11] Wang J, Cao Z, Zeng D D, et al. Epidemiological analysis, detection, and comparison of space-time patterns of Beijing hand-foot-mouth disease(2008-2012) [J]. *PLoS One*, 2014, 9(3): e92745.
- [12] 肖辉, 肖革新. 时空扫描统计量在细菌性痢疾监测数据分析中的应用 [J]. *中国食品卫生杂志*, 2014, 26(1): 83-87.
- [13] 孙海泉, 肖革新, 郭莹, 等. 中国 2008—2012 年丙肝流行规律及空间聚集性分析 [J]. *中国公共卫生杂志*, 2014, 30(3): 286-289.

- [14] Lin H L , Liu Q Y , Guo J Q , et al. Analysis of geographic distribution of HFRS in Liaoning Province between 2000 and 2005 [J]. BMC Public Health , 2007(7) : 207.
- [15] Liu W , Yang K , Qi X , et al. Spatial and temporal analysis of human infection with avian influenza A(H7N9) virus in China , 2013 [J]. Euro Surveill , 2013 , 18(47) : pii=20640.
- [16] Sherman R L , Henry K A , Tannenbaum S L , et al. Applying spatial analysis tools in public health : an example using SaTScan to detect geographic targets for colorectal cancer screening interventions [J]. Prev Chronic Dis , 2014 , 11 : E41.
- [17] Kulldorff M , Athas W F , Feuer E J , et al. Evaluating cluster alarms : a space-time scan statistic and brain cancer in Los Alamos , New Mexico [J]. American Journal of Public Health , 1998 , 88 (9) : 1377-1380.
- [18] Kulldorff M , Heffernan R , Hartman J , et al. A space-time permutation scan statistic for disease outbreak detection [J]. PLoS Medicine , 2005 , 2(3) : e59.
- [19] Mostashari F , Kulldorff M , Hartman J J , et al. Dead bird clusters as an early warning system for West Nile virus activity [J]. Emerging Infectious Diseases , 2003 , 9(6) : 641-646.
- [20] Mugglin A S , Cressie N , Gemmell I. Hierarchical statistical modelling of influenza epidemic dynamics in space and time [J] . Statistics in Medicine , 2002 , 21 (18) : 2703-2721.
- [21] 周晓农, 杨国静, 杨坤, 等. 中国空间流行病学的发展历程与发展趋势 [J]. 中华流行病学杂志 , 2011 , 32(9) : 854-858.
- [22] Wang J F , Hu Y. Environmental health risk detection with GeogDetector [J]. Environmental Modelling & Software , 2012 , 33 : 114-115.
- [23] Wang J F , Li X H , Christakos G , et al. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region , China [J]. International Journal of Geographical Information Science , 2010 , 24(1) : 107-127.
- [24] Wu J L , Wang J F , Meng B , et al. Exploratory spatial data analysis for the identification of risk factors to birth defects [J]. BMC Public Health , 2004 , 4 : 23.
- [25] Hu M G , Li Z J , Wang J F , et al. Determinants of the incidence of hand , foot and mouth disease in China using geographically weighted regression models [J]. PLoS ONE , 2012 , 7(6) : e38978.
- [26] Deng T , Huang Y , Yu S , et al. Spatial-temporal clusters and risk factors of hand , foot , and mouth disease at the district level in Guangdong Province , China [J]. PLoS One , 2013 , 8(2) : e56943.
- [27] Choi J , Lawson A B , Cai B , et al. Evaluation of bayesian spatio-temporal latent models in small area health data [J]. Environmetrics , 2011 , 22(8) : 1008-1022.
- [28] Richardson D B , Volkow N D , Kwan M P , et al. Medicine spatial turn in health research [J]. Science , 2013 , 339(6126) : 1390-1392.
- [29] 武继磊, 王劲峰, 郑晓瑛, 等. 空间数据分析技术在公共卫生领域的应用 [J]. 地理科学进展 , 2003 , 22(3) : 219-227.
- [30] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域-大数据的研究现状与科学思考 [J]. 中国科学院刊 , 2012 , 27(6) : 647-656.

## Spatial Statistics and Its Application to the Field of Public Health

NI Shuhua

(The First Affiliated Hospital of Shantou University, Shantou 515041, Guangdong, China)

**Abstract:** Most public health data possess spatial attributes. While the ability of dealing with this type of data using classic statistics is limited, the spatial analytical techniques can make full use of this attribute of the data. The spatial statistics is mainly applied to the field of public health, including the spatial distribution of disease pattern, spatial relationships, causality analysis and spatio-temporal warning analysis, etc.

**Keyword:** public health; spatial statistics; spatial autocorrelation; spatial heterogeneity

---

(上接第 7 页)

[14] Kurzweil H, Stellmacher B. The theory of finite group [M]. New York: Springer-Verlag, 2003.

## Elementary Prime Tight Decomposition of Harmonic Frames and Their Applications

CHEN Caifeng, YANG Shouzhi

(Department of Mathematics, Science College, Shantou University, Shantou 515063, Guangdong, China)

**Abstract:** This paper puts forward the conception of elementary prime tight frames, proving that any finite dimension harmonic frames can be decomposed as unions, direct sums and projection of tensor products of those frames. Conversely, the compound of elementary prime tight frames may not be a harmonic frame, but becomes a general equal normal tight frame (ENTF), which provides a novel method to construct ENTF. Finally, some kinds of ENTF are studied, which have better properties as compared with the former frames.

**Keyword:** harmonic frames; elementary prime tight frames; decomposition of frames