[文章编号] 1005-6661(2011)02-0217-04 。综述。

空间统计分析方法及相关软件在传染病研究中的应用

冯军,吴晓华,李石柱,周晓农*

[摘要] 本文就空间统计分析方法及其相关的分析软件在传染病研究中的应用作一综述,旨在加深对空间统计分析 及相关软件在传染病研究中应用的认识,为传染病研究及预防控制提供参考。

[关键词] 传染病;空间统计分析;软件

[中图分类号] R51 [文献标识码] A

Application of spatial statistical analysis methods and related analytic soft. wares in research of infectious diseases

Feng Jun Wu Xiao hua Li Shi zhu Zhou Xiao nong*

National Institute of Parasitic Diseases Chinese Center for Disease Control and Pievention Shanghai 200025 China

* Corresponding author

[Abstract This paper summarizes the application of several spatial statistical analysism ethods and their related softwares in the research of infectious diseases in order to improve the prevention and control of infectious diseases

Infectious disease Spatial statistical analysis Software

目前,由于经济的全球化趋势、生态环境及人类行为、生活 方式等改变的影响,新发传染病不断增加,如艾滋病、军团菌 病、莱姆病、各种病毒性出血热、Qaa霍乱、雅氏综合征和新型肝 炎等,因而传染病仍是导致人类死亡的一个重要因素[12]。据 估计,80%的传染病流行病学研究资料都具有空间属性,如人 群和动物的发病感染情况、宿主媒介的分布、温度、湿度、降雨、 土壤、卫生设施等都与地理位置有关。 然而, 传统统计分析方 法在分析空间数据时存在致命缺陷,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 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,这使得 大部分空间数据样本并不独立,即不满足传统统计分析的样本 独立性前提,因而不适用于传统统计分析。随着信息技术的发 展,空间统计分析方法及相关软件应运而生,从而为研究者更 好地认识和把握自然、社会及环境的运动规律提供了条件。 在 传染病学研究中,各种空间统计分析方法的运用为疾病的空间 聚集性、影响因素的探索及疾病的预防控制提供了更有力的研 究方法和工具。本文就空间统计分析的发展历程及其分析方 法,以及相关分析软件在传染病研究中的应用作一综述。

1 空间统计分析的发展历程

空间统计分析, 即地统计学(Geostatistics), 亦称地质统计 学,是以区域化变量理论 (Theory of regionalized variable) 为基

国家科技支撑计划 (2007 BAC03A02); 国家重大科技专 [基金项目] 项 (2008ZX10004-011)

[作者单位] 中国疾病预防控制中心寄生虫病预防控制所 (上海

*通信作者 E-mail Pddpouxi@ sh163 net (C)1994-2021 China Academic Journal Electronic Publishin

础,以变异函数(Variogram)为基本工具,研究分布于空间并呈 现出一定的随机性和结构性的自然现象的科学,是现代计量地 理学中一个快速发展的方向和领域。该方法可研究某些变量 (或特征)的空间分布特性,并对其进行最优估计。 在模拟所研 究对象的离散性、波动性或其他性质时,也可应用空间统计分 析的理论与方法[3]。 20世纪 60年代, 地理学计量革命中的部 分模型初步考虑了空间信息的关联问题,成为当今空间数据分 析模型的萌芽。法国统计学家 Matherorl⁴]在此基础上通过大 量理论研究,总结并提出了地统计学的概念,或称 Krging方 法。 1973年、Clif等 [5] 成功构建了空间自相关指标、从而为测 度地理空间单元之间的相互联系提供了有效的统计学手段。 1979年, 地理学家 Toble 指出了空间自相关性的存在, 并在此 基础上提出地理学第一定律,即"空间上分布的事物是相互联 系的, 但近距离事物之间相似性大于远距离事物之间的相似 性"[6]。随后、Anselin Griffith Cressie等[79]学者在这一领域的 诸多学术研究,则从根本上将理论地理学和区域科学中的研究 重点从传统的空间分析导向了空间数据的统计分析。 1995年. Kulldorf等[10]提出了一种可对某些疾病的空间聚集性进行观 察和推断的方法,即扫描统计量(Scan statistic)。2000年、王劲 峰等 $^{[11]}$ 、W ang等 $^{[12]}$ 首次在国内对空间统计分析进行了系统的 介绍及探讨。

2 空间统计分析及其在传染病中的应用

空间统计分析的方法除一般的统计图表分析、密度分析之 外,还包括空间自相关分析、空间分类分析(空间聚类分析、空 间聚合分析和判别分析)、多变量统计分析(含主成分分析、主 因子分析、变量聚类分析和采样点聚类分析)、空间插值分析、 空间结构分析以及空间模拟等,其核心是认识与地理位置相关

的数据间的空间关联或空间自相关,通过空间位置建立数据间

的统计关系[13]。以下主要介绍在传染病研究中常用的方法。

21 空间自相关分析 空间自相关是指空间位置上越靠近的事物或者现象越相似,即事物或现象具有对空间位置的依赖关系。空间自相关的概念源于时间自相关,但比后者复杂。主要是因为时间是一维函数,而空间是多维函数。 因此,在度量空间自相关时,还需要解决地理空间结构的数学表达,定义空间对象的相互邻接关系 [14]。 与经典统计学中的相关相比,空间自相关的差别在于处理的数据资料不同。空间自相关是同一属性值在不同空间位置上的相互关系;而经典的相关是指两个或多个属性变量之间的相互关系及其密切程度。计算空间自相关的方法通常分为全局型和局部性两种类型,最为常用方法是 $Moran s \downarrow Geary s \in Getisly$ 及半变异图与空间自相关系数图等 [5.15]。

杨国静等[17]曾用 Moran s I统计量对血吸虫病患者与钉螺分布的空间自相关性进行了研究。范新生等[18]运用 Moran s 和 Getis统计量对我国 2003年暴发的 SARS疫情的省级空间分布格局进行了分析,结果表明,SARS疫情的空间分布在疫情发源阶段以及新暴发中心形成阶段具有很强的空间自相关性。 K itxon等[19]以 Moran s 为指标对美国威斯康星州莱姆病的空间分布进行了研究,发现人感染莱姆病的概率与该地区蜱的分布呈空间正相关,并且两者均与春、秋季的校正植被指数值 (NDVI)呈正相关。 Hu等[20]采用 Moran s I等指标,对昆士兰州已上报的登革热病例资料进行了空间自相关性及空间分布估计,结果表明,登革热的空间分布在某些时期呈现聚集性,且在不同地区聚集程度不同。

2 1. 2 局部空间自相关分析 全局空间自相关是对整个研究空间的总体描述,仅对同质的空间过程有效。然而,由于环境和社会因素等外界条件的不同,空间自相关的大小在整个研究空间,特别是在范围较大的研究空间上并不一定是均匀同质的,其可能随空间位置的不同有所变化,甚至可能在一些空间位置发现正空间自相关,而在另一些空间位置发现负空间自相关。这种现象称为空间异质性(SPatal heterogeneity),在全局空间自相关分析中是无法发现的,而局部空间自相关统计量则可对其进行识别。局部空间自相关分析的指标主要包括局部 Morair's I统计量(Local Morair's I统计量(Local Morair's I统计量(Local Morair's I统计量(Local Morair's I)、局部空间自相关统计量(Local Getis)[2]等。

不同发展阶段的空间聚集中心,并提供了这一中心随时间的空间转移轨迹。蒋敏等 [22] 对四川省 H IV/A IDS的发生与空间位置之间的关系进行了研究,并采用局部空间自相关 Getis统计量探测了聚集位置与属性。曹志东等 [23] 在对广州 SARS流行的空间风险因子与空间相关性特征的研究中,计算了广州市 SARS发病率的 LISA,由此得到了该研究区域的局部空间聚集特征。裴小琴等 [24] 采用 R Getid 等软件对内蒙古自治区 2004-2005年布鲁氏菌病资料进行的空间分布模式等分析,其局域系数统计显示,布鲁氏菌病发病率和空间自相关系数大小存在一定的空间统一性和差异性。

2 2 空间插值分析 空间插值分析是指根据一组已知的离散 数据或分区数据,按照某种数学关系推求其他未知点或未知区 域的数据分析过程[25],可分为整体插值和局部插值方法两类。 整体插值方法用于研究区所有采样点的数据, 以进行全区特征 拟合,主要包括边界内插方法、趋势面分析、变换函数插值等。 局部插值方法则是用邻近的数据点来估计未知点的值,能弥补 整体插值方法的缺陷,可用于局部异常值,且不受插值表面上 其他点的内插值影响。主要包括泰森多边形 (又称 Dirichle或 Voronoi多边形)、距离倒数插值(Inverse distance weighted)、样 条函数插值方法(SPlinemethod)及克里格插值法(Kriging)[26] 等。目前,距离倒数插值、趋势面分析及克里格插值法等在传 染病研究中应用较多,可绘制疾病或媒介的空间分布图,从而 描述疾病及其传播媒介的空间分布特征,并估计其变化趋势, 以供卫生决策参考[27]。国外研究者运用空间插值分析有关传 染病及其媒介的研究较早。 K leinsclm id 等 [28] 在非洲马里地区 利用克里格法进行疟疾研究, 构建了马里地区疟疾流行的预测 图, 为当地疟疾防治提供了依据。 Saka 等 [29] 运用克里格法对 日本流感样病症流行的时空变化趋势展开研究,建立了流感样 病症发病的等值线图。苏永强等[30]运用空间局部内插分析方 法对海南省疟疾空间分布特征进行了研究,绘制了海南省 1995 - 1999年疟疾空间分布图。易彬樘等[31]利用 ArcGIS 8 1空间 分析软件,对广东省 1995年各县 (市)登革热的发病资料及同 期伊蚊媒介监测资料进行的协同克里格(CoKriging)分析显 示, 登革热流行、种群媒介及标化植被植树 (NDVI) 三者之间的 空间分布呈现地域一致性。张治英等[3]运用普通克里格的变 异函数对江宁县江滩钉螺分布的空间自相关性进行了分析,结 果显示 2000年江宁县江滩钉螺分布呈空间自相关性,其变异 函数为球型模型,且当距离< 0.030 1时,钉螺空间分布变异与 距离有关,并以此为基础用普通克立格法建立了江宁县江滩钉 螺分布 预测图。周水森等[33]运用克里格插值法对我国黄淮流 域疟疾空间分布特征进行了研究,结果显示该地区的疟疾在空 间分布上与距离有关, 并呈现出两个明显的聚集中心, 且其空 间分布并非与行政区划分类完全一致。

2 3 空间回归分析 空间回归分析技术是地理相关性研究及生态学分析的主要方法,其可用于探讨估计值的空间关系,而地学对象中的空间依赖性和空间异质性使得一般回归方法不宜用于空间分析 [34]。 其有联立自回归模型 (Sinultaneous Autoregressive Model), 空间移动平均模型 (SPatial Moving Average

· 范新生等 [18] 采用局部 Getis统计量确定了我国 SARS疫情 ... Model)和条件自回归模型 (Conditional Autoregressive Model)等 (C) 1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

3种特殊形式[型]。以联立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得到了广泛的应用,其在传染病学研究中的应用也已逐步开展。此分析方法主要从地理(或生态学)的角度研究疾病发病(或患病、死亡等)的空间分布与环境因素(如空气、水、土壤等)、社会经济因素间的关系。

3 常用空间统计分析软件

空间数据分析技术方法的应用,促进了空间数据分析软件 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传染病研究和干预方面得到了进一步应用。

- 3.1 专业空间统计分析软件 如通用空间统计分析软件 (Geodata Analysis Software, GeoDa)[35]、空间点模式分析统计软 件包(Point Pattern Analysis)[36]、以及地统计学库(Geostatistical Software Library GSLIB, [37]、SaScant [38] 及 GS+环境科学空间统 计分析软件 (Geospatistics for the Environmental Sciences GS +)[39]等空间统计分析软件。此类软件的优点在于其开发本 身借助基本的软件开发语言及其他集成开发环境,并融入最新 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: 但目前其应用多在科研领域, 未形成大规模的开发需求, 因 此在空间建模能力和可操作性上需要进一步加强。GeoDa和 SaScar在传染病研究中的应用较其他同类软件多。 Dreesman 等[40]运用 SatScan软件中的扫描统计量,识别了德国下萨克森 州的 16种法定传染病的空间聚集性及趋势, 为进一步分析这 些传染病的流行病学特征奠定了基础,同时也为疾病监测提供 了技术支持。 Fang等[4]分别运用 GeoDa和 SatScan软件对中 国肾综合征出血热(HFRS)高流行省份进行了疾病的空间自相 关分析及空间聚集性分析,为定量阐明 HFRS高流行地区的疾 病负担,以及更深入地研究环境所致疾病相关因素提供了基 础。 Fische等[42]利用 SaScan软件中的空间扫描统计量对孟 加拉国 4个村庄麻风病的空间聚集性进行了研究,提示对于地 方病空间聚集性的分析应在较大区域内进行。
- 3 2 以现有统计分析软件为基础结合 G LS软件的空间统计分析软件包 如基于大型统计分析软件 S AS 结合 G LS软件形成的 S AS G LS空间统计分析软件包 E AS 结合 G LS软件形成的 E AreV E iews E GS E S

了可视化时间与分析时间两个概念, 用户可以利用 ArcGIS 10 0 创建、管理并可视化时空数据, 并动态显示时空数据集, 还可发布、查询时空地图服务, 并能识别随时间变化的数据, 从而使得用户能够开展更深入的空间统计分析。此类软件的优点是其基于空间信息管理为功能的 GIS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。随着此类软件空间统计分析功能的不断加强, 其在传染病研究中的运用也越来越多多[4849]。方立群等[50]在 ArcGIS软件的支持下, 利用空间统计分析模块对 1995—1998年全国 41个 HFRS监测点人群发病资料进行了反距离权重插值分析。 Frank等[51]利用 ArcGIS软件对阿散蒂地区的霍乱进行了空间统计分析, 发现在库马西市的中部地区存在空间聚集性。

除此之外,还有一些具有空间统计分析功能的开放源代码软件,如 Gstat R语言等 [\$2-53]。 Gsta软件具有多变量空间统计建模、预测、模拟等功能;而 R语言也具有许多空间统计分析功能 [54]。 此类软件由于采取开放式结构、模块化设计,且支持空间数据分析方法的扩充与重构,因而可在已有研究基础之上对系统进行再开发,从而对于空间数据分析统计软件的发展具有重要意义。

4 小结

目前,空间统计分析在传染病研究中的应用主要集中于血吸虫病、疟疾、鼠疫等疾病,其他方面尚不深入。究其原因,一是由于空间统计分析是近 50年内发展起来的,某些空间数据分析模型、方法和技术本身尚存在许多不完善之处。尽管空间统计分析在传染病研究中已经显示出了极大的应用潜力,但其仍需结合传染病研究领域的自身特点,以进一步完善其功能;二是由于长期以来在我国的传染病研究中缺乏相应的空间理论、方法和技术手段,长期以来积累的数据缺乏空间属性,因而限制了空间统计分析在相关传染病研究中的深入运用。随着空间信息基础设施的建设及发展,快速获取和掌握大量的自然、生态、环境及社会经济等数据信息已经成为可能;加之一些传染病(如血吸虫病、疟疾、鼠疫等)空间数据库的构建,这些都从客观上促进了空间统计分析在传染病研究中的应用,同时也为传染病的理论研究及预防控制提供了有力的工具及技术支持。

[参考文献]

- [1] Cohen MI, Changing patterns of infectious disease J. Nature, 2000, 406 (6797): 762-767.
- [2] Jones KE, Pate NG, Levy MA, et al. Global trends in emerging in fectious diseases J. Nature, 2008, 451(7181): 990-993
- [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 [...]. 地质与勘探, 1997, 33(1): 53-58.
- [4] Matheron G. Principles of geostatistics J. Econ Geo.], 1963, 58 (8):
- [5] CliffAD, Ord JK, Spatjal autocorne lation Mj. London, Pion, 1973, 40-56
- [6] Chantel DŞ Eric JD, Xun Ş et al Ecogeographic genetic epidemiolo.

 gy J. Genet Epidemiol 2009, 33(4): 281-289
- 全局多项式插值法、局部多项式插值法等: 同时此版本还提出 [7] Anselin L. Local indicators of spatial association [1] Geog Anal (C) 1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

- 1995, 27(2): 93-115
- [8] Griffith DA. Toward a theory of spatial statistics another step forward $\mbox{[} \mbox{ J. } \mbox{Geog} \mbox{ Ana,l } \mbox{1987} \mbox{ 19} \mbox{ (1): } \mbox{69-82}$
- [9] Cressie N. Statistics for spatial data J. Terra Nova 1992, 4 (5):
- [10] KulldorffM, Nagawalla N. Spatjal disease cluster detection and interference J. Stat Med 1995, 14(8): 799-810.
- [11] 王劲峰, 王智勇. 地理信息空间分析的理论体系探讨[J]. 地理学报, 2005, 5(1); 92-103.
- [12] Wang JF, Li LF, GeY, et al. A theoretic framework for geographic information analysis J. Acta Geog Sin. 2000, 55 (1): 318-328.
- [13] 冯益明, 唐守正, 李增元, 等. 空间统计分析在林业中的应用[]. 林业科学, 2004, 40(3): 149-154.
- [14] 张学良. 探索性空间数据分析模型研究 [J]. 当代经济管理, 2007, 29(2): 26-29
- [15] Getis A, Ord JK, The analysis of spatial association by use of distance statistics J. Geog Anal 1992, 24(3): 189-206
- [16] Getis A. Reflections on spatial autocorrelation []. Reg Sci U than $E^{con}, \ 2007, \ 37(4); \ \ 491-496$
- [17] 杨国静, 周晓农, 汪天平, 等. 安徽、江西及江苏 3省血吸虫病患者与钉螺分布的空间自相关分析 [基]. 中国寄生虫学与寄生虫病杂志, 2002, 20(1): 6-9.
- [18] 范新生, 应龙根. 中国 SAR S疫情的探索性空间数据分析 []. 地理科学进展, 2005, 20(3): 6-9
- [19] Kitron U, Kazmierczak JJ Spatial analysis of the distribution of Lyme disease in Wisconsin J. Am JEP idemiol 1997, 145 (6): 558-566.
- [20] HuW, Clements A, Williams G, et al. Spatial analysis of motified dengue fever infections J. Epidem io Infect, 2010, 15(4): 1000-1017.
- [21] Ord K, Getis A. Local spatial autocorrelation statistics Distributional issues and application J. Geog Anal 1995, 27(4): 286-306
- [22] 蒋敏, 李晓松, 冯子健, 等. 四川省 HN/ADS空间自相关分析 [.]. 现代预防医学, 2008, 35(22): 4329-4331
- [23] 曹志冬,王劲峰,曾光,等.广州 SARS流行的空间风险因子与空间相关性特征[j. 地理学报, 2008, 63(9); 981-993
- [24] 裴小琴, 冯子健, 马家奇, 等. 2004—2005年内蒙古自治区布鲁氏 菌发病的空间自相关分析 [J]. 中华疾病控制杂志, 2009, 13(6): 654-658.
- [25] 吴立新, 史文中. 地理信息系统原理与算法 [M]. 北京: 科学出版 社, 2003, 183-201.
- [26] 乌伦,刘瑜,张晶,等.地理信息系统——原理、方法和应用[M]. 北京:科学出版社, 2000, 178-191.
- [27] Lance AW, Carol AG. Applied spatial statistics for Public health data [M]. New Jersey Wiley 2004, 274-313.
- [28] Kleinschmidt J. Bagayoko M. Clarke GPY, et al. A spatial statistical approach to malaria mapping J. Int JEP idemio, 2000, 29(2): 355-361.
- [29] SakaiŢ SuzukiḤ SasakiĄ et al Geographic and temporal trends in influenza like illness, Japan, 1992—1999[J. Emerg InfectDis, 2004, 10(10): 1822-1826.
- [30] 苏永强, 张治英, 徐德忠, 等. 运用空间局部内插研究海南省疟疾空间分布特征[1]. 中华流行病学杂志, 2003, 24(4): 269-271
- [31] 易彬樘, 张治英, 徐德忠, 等. 协同克立格和标化植被指数在广东省 登革热分布特征研究中的应用[]. 西安交通大学学报. 医学版,

- [32] 张治英, 徐德忠, 彭华, 等. 普通克里格法预测江宁县 江滩钉螺分布 [J. 中国寄生虫学与寄生虫病杂志, 2004, 22(3): 170-172
- [33] 周水森, 黄芳, 汤林华, 等. 运用 Krējag法对我国黄淮流域疟疾空间分布特征的研究[3]. 中国病原生物学杂志, 2007, 2(3): 204-206
- [34]王劲峰,武继磊,孙英君,等.空间信息分析技术[J. 地理研究, 2005, 24(3): 464-472.
- [35] Anse lin L. Computing environments for spatial data analysis J . $G^{eogr}\,S^{yst}, \ 2000, \ 2(\ 3); \ \ 201\text{-}220$
- [36] Point Pattern ana Jusis (PPA) [OL]. http://www.geog.ucsh.edu/-dongmei/ppa/ppa/html
- [37] Reservoir modeling with GSLB[OL]. http://www.statios.com/ Training/index.htm.
- [38] Software for the spatia, tempora, and space time scan statistics OI). http://www.satscan.org/
- [39] Geostatistics for the environment sciences OI]. http://www.gam-mades.gn.com/
- [40] Dreesman J Scharlach H Spatial statistical analysis of infectious discase notification data in lower saxony J. Gesundheitswesen, 2004, 66(12): 783-789.
- [41] Fang L, Yan L, Liang S, et al. Spatial analysis of homorihagic fever with renal syndrome in China J. BMC Infect Dis 2006 6 77.
- [42] Fischer EAJ Pahan D Chowdhury SK, etal The spatial distribution of leprosy in four villages in Bangladesh; an observational study J. BMC Infect Dis 2008 8 125
- [43] Bailey TC Spatial statistical methods in heal of J. Cad Saude Publica, 2001, 17(5): 1083-1098.
- [44] Products and solution/SAS/GIS[OI]. http://www.sas.com/products/gis/
- [45] Procducts/S-splus[OL]. http://www.insightful.com/products/addons asp
- [46] Modern apllied statistics with S. Software OLJ. http://www.stats.ox.ac.uk/pub/MASS4/Software.html
- [47] Patrick AW, Owen JD. Interactive analysis of the spatial distribution of disease using a geographical information system. J. J Geogr Syst 2000, 2(3): 243-256.
- [48] 孙英君, 王劲峰, 柏延臣, 等. 地统计学的 G S空间分析功能扩展 [J]. 华侨大学学报. 自然科学版, 2004, 25(4): 435-439.
- [49] 王艳妮, 谢金梅, 郭祥. ArG B中的地统计克里格插值法及其应用 [J. 软件导刊, 2008, 7(12); 36-38.
- [50] 方立群,曹务春,陈化新,等. 应用地理信息系统分析中国肾综合征出血热的空间分布[J]. 中华流行病学杂志, 2003, 24(4): 265-268.
- [51] Osei FB, Duker AA, Spatial and demographic patterns of Cholera in Ashanti region Ghana J. Int J Health Geogr 2008, 7, 44.
- [52] The Gstat R Package and S. Puls library multivariable geostatistics
 [OI]. http://www.gstat.org/s.hm.l
- [53] Cran task views and Vsis of spatial data OI₃. http://cran.rprojectorg/web/views/Spatial.htm.
- [54] 江何, 周华春. 开放源代码软件的应用研究[1]. 计算机系统应用, 2002, 6, 77-79.

[收稿日期] 2011-01-21 [编辑] 邓瑶