卫生健康事业发展70年巡礼

doi:10.3969/j.issn. 1672-5166.2019.05.03

基于地理信息平台的传染病时空聚集性分析与实现

徐 勇[①] 吕 露[①] 张 萌[①] 宋铁[①△] 童昊昕[②]

文章编号：1672-5166（2019)05-0532-06 中图分类号：R-39；R319 文献标志码：A

摘 要 为简化传染病时空聚集性分析工作，基于时空扫描统计、时空重排扫描等基本原理，对扫描统计异常单元的计算、扫描窗口和扫描半径的设置等进行深入研究，逐项解析、编程、调试，实现传染病时空聚集性探测的工程化。利用广东省疾病预防控制中心开发的“基于地理信息平台的突发公共卫生事件监测预警及应急处置技术研究平台”，对2014年度广东省登革热病例数据进行在线时空扫描，与利用satscan软件进行时空扫描的结果相符，无需安装多个软件，可减少人工干预，提高工作效率。

关键词 时空扫描 聚集性 随机数

Analysis and Implementation of Spatio-temporal Aggregation of Infectious Diseases Based on Geographic Information Platform

XU Yong, LV Lu, ZHANG Meng, SONG Tie, TONG Haoxin

Guangdong Provincial Center for Disease Control and Prevention, Guangzhou 551430, Guangdong, China

Abstract In order to simplify the analysis of temporal-spatial aggregation of infectious diseases, based on the basic principles of spatial-temporal scanning statistics and spatial-temporal rearrangement scanning, the calculation of scanning statistical anomaly units, the setting of scanning window and scanning radius are studied in depth, and the detection of temporal-spatial aggregation of infectious diseases is realized by analyzing, programming and debugging item by item. Using this platform, the data of dengue fever cases in Guangdong Province in 2014 were scanned online in time and space, which was consistent with the results of time and space scanning using SaTScan software. With no need to install multiple software, it can reduce manual intervention and improve work efficiency.

Keywords spatio-temporal scanning; aggregation; random number

课题项目：广东省省级科技计划项目（项目编号：2019B020208005）

① 广东省疾病预防控制中心，广东省广州市，551430

② 航天精一（广东）信息科技有限公司，广东省广州市，551430

作者简介：徐勇（1966—），男，硕士，科教与信息部副主任，高级工程师；研究方向：空间流行病学，地理信息系统；E-mail：xuyong@cdcp.org.cn

通信作者：宋铁（1971—），男，硕士，疾控中心副主任，主任医师；研究方向：突发传染病防控；E-mail：tsong@cdcp.org.cn

△ 通信作者

时空聚集性分析在流行病学的应用逐渐增多，但通常每次分析均需经过提前准备病例数据、地理信息数据和人口数据；利用 SAS 等统计分析软件对数据进行处理；在 SaTScan 软件中调用病例、人口和地理信息数据文件，设置参数后进行时空聚集性分析；利用制图软件将分析结果进行地图展示等流程。上述操作过程繁琐，工作量极大，如：数据准备和处理耗时耗力，Satscan 软件在分析大数据量的时候运算力不够，制图结果不够直观，涉及多种软件且学科跨度大。广东省疾病预防控制中心在开发“基于地理信息平台的突发公共卫生事件监测预警及应急处置技术研究平台”时，将时空聚集性分析作为核心功能之一。根据其原始理论、数理方法、统计学意义和疾病制图的方法，采用工程化手段进行逐项解析、编程、调试，在服务器集群上部署，实现传染病在线时空聚集性分析。

1 基本原理和方法

1.1 基本原理

1.1.1 扫描统计量

1965 年，全国空域利用系统 (national airspace utilization system，NAUS) 首次提出扫描统计量，其主要用途在于探测在局部时间或空间上，是否有事件发生数增加的趋势，并检验这种增加趋势是否具有一定的随机性 [1]。扫描统计量计算结果主要与 3 个基本特征密切相关：被扫描区域的几何形状、基于无效假设的概率分布及扫描窗口的形状和大小 [2]。

1.1.2 时空扫描统计

1998 年，Kulldorff 进一步扩展提出了时空扫描统计的方法，旨在探测一定时空范围内的聚集性与随机分布模式比较，是否显著增加，并确定聚集性最可能异常的时空事件集合 [2]。

殷菲等学者根据国家传染病网络直报系统中的麻疹数据，在县级空间尺度上对上海、江苏、浙江 3 省市进行了传染病时空聚集性研究 [3]。徐聪等学者以小空间尺度单元为研究对象，提出采用小空间尺度单元对传染病进行时空聚集性研究 [4]。

1.1.3 时空扫描统计量

对每一个扫描窗口，根据实际发病数和人口数计算出理论发病数后，利用扫描窗口内和扫描窗口外的实际发病数和理论发病数构造检验统计量对数似然比 (log likelihood ratio，LLR)，用 LLR 来评价扫描窗口内发病数的异常程度，并对可能存在的窗口进行蒙特卡罗法 (Monte Carlo method) 随机化检验，以避免多重窗口的假阳性问题 [5-6]。似然比的计算公式如下：

$$\frac{L(z)}{L_0}=\frac{\left(\frac{n_Z}{\mu(Z)}\right)^{n_Z}\left(\frac{n_G-n_Z}{\mu(G)-\mu(Z)}\right)^{n_G-n_Z}}{\left(\frac{n_G}{\mu(G)}\right)^{n_G}}$$

$$\mu(Z)=\frac{n_G}{m_G}\times m_Z$$

$L(Z)$ 为扫描窗口 Z 的似然函数值，L_0 是基于无效假设得到的似然函数值；n_Z 为扫描窗口 Z 中的实际发病数，m_Z 为扫描窗口 Z 中人口数，$\mu(Z)$ 为根据无效假设得到的扫描窗口 Z 中预期发病数，n_G 为所有区域 G 的实际发病数，m_G 为所有区域 G 的人口数。

然后利用蒙特卡罗法产生模拟数据集，对模拟数据集用跟真实数据集相同的方法进行计算，找出发病数异常程度最高的窗口，计算 P 值。

1.1.4 时空重排扫描统计量

时空重排扫描统计量，建模过程中不需要人口数据 [6-7]。时空重排扫描统计量采用 Poisson 广义似然函数 (generalized likelihood ratio，GLR) 来衡量扫描窗口中的发病数是否异常，同样也对可能存在的窗口进行蒙特卡罗法随机化检验。

$$GLR=\left\{\frac{C_A}{\mu_A}\right\}^{C_A}\left\{\frac{C-C_A}{C-\mu_A}\right\}^{C-C_A}$$

$$\mu_A=\sum_{(Z,d)\in A}\mu_{Zd}$$

C 为所有区域在所有时间的总发病例数为，C_A 为每个圆柱 A 中的实际发病数，μ_A 为每个圆柱 A 对应的区域 Z 和 d 天的预期发病数。

1.1.5 前瞻性和回顾性时空扫描统计

用于传染病监测数据的时空扫描统计方法有回顾性

时空扫描和前瞻性时空扫描。前瞻性时空扫描主要用于动态探测正在发生、发展的某类传染病发病数/发病率异常增高的时间点和空间区域，基于到目前为止收集的数据[8]。回顾性时空扫描主要用于探测已经发生过的某类传染病发病数/发病率异常增高的时间点和空间区域，对整个历史数据集进行分析[9]

1.2 扫描统计异常单元的计算

1.2.1 利用整个区域范围的病例计算扫描区域和时间范围的整体均值

令所有区域G的总发病数为n_G，总人口数为m_G，则整体均值为$\frac{n_G}{m_G}$。

1.2.2 计算每个时空扫描单元的预期发病数

（1）当使用非行政区划扫描时，由于无法获取各扫描区域单元的人口数据，故使用时空重排模型建立概率模型，计算每个时空扫描单元的预期发病数。

假设扫描区域Z在d天中的发病数为C_{Zd}，则所有区域在所有时间的总发病例数C为：

$$C=\sum_{Z}\sum_{d}C_{Zd}$$

对每个区域和每天，预期发病数：

$$\mu_{Zd}=\frac{1}{C}\left(\sum_{Z}C_{Zd}\right)\left(\sum_{d}C_{Zd}\right)$$

（2）当使用行政区划扫描时，由于能够准确获取各行政区（市、区县、街镇）的人口数据，故使用Poisson模型建立概率模型，计算每个时空扫描单元的预期发病数。

令n_Z为扫描窗口Z中的实际发病数，m_Z为扫描窗口Z中人口数，$\mu(Z)$为根据无效假设得到的扫描窗口Z中预期发病数，所有区域G的预期发病数为$\mu(G)$：

$$\mu(Z)=\frac{n_G}{m_G}\times m_Z$$

$$\mu(G)=\Sigma\mu(Z)$$

1.2.3 利用求得的扫描单元预期发病数，构建该扫描单元病例的随机分布

利用蒙特卡罗法产生模拟数据集，采用泊松随机数生成算法。具体思路为：利用扫描单元预期发病数，构建该扫描单元的泊松分布随机数。采用累计分布函数相加的方法生成符合泊松分布的随机数：

（1）设λ为扫描单元预期发病数，构造期望为λ的泊松分布概率密度函数。

$$P_k=P(x=k)=\frac{\lambda^k}{k!}e^{-\lambda}\ ,\ k=0,1,2\cdots\cdots$$

（2）利用Math.random()函数产生一个服从（0，1）上均匀分布的随机数y。

（3）设初始值k=0。

（4）将k代入泊松分布的概率密度函数计算概率P_k。

（5）判断概率$P=P_0+P_1+\cdots+P_k$是否小于步骤（2）生成的随机数y。如果$P<y$，则执行$k=k+1$，返回执行步骤（4）；如果$P\geqslant y$，则结束循环，输出k，作为生成的一个服从期望为λ的泊松分布的随机数。

1.2.4 似然比检验

基于似然函数，Kulldorff构造了似然比检验。对每一个地点和大小的扫描窗口，备择假设为：跟窗口外相比，窗口内的发病率增加[2]。

对于某一特定窗口，似然函数为：

$$\left(\frac{c}{n}\right)^{c}\left(\frac{C-c}{C-n}\right)^{C-c}I(c>n)$$

其中C为总例数，c为窗口内的例数，n是基于无效假设由协变量校正过的预期发病数，$I(\cdot)$为0/1指示变量，当窗口内实际发病数高于预期发病数时，$I(\cdot)$取1，反之取0。

对数似然比为：$LLR=c\times(C-c)\log\left(\frac{C}{n}\right)\times\left(\frac{C-c}{C-n}\right)$

对于每一个窗口，都计算似然函数，然后寻找所有地点所有大小的窗口中最大的似然函数值，此处即为最有可能存在聚集性的区域，也是最不可能由随机变异造成的。

对于每一个窗口，根据病例数和人口数按照2.1.3的步骤产生w（至少大于999）个随机数（对于N个扫描窗口，产生$w\times N$个随机数据集）。然后将真实数据集的对数似然比与随机数据集的对数似然比相比较，如果真实数据集的对数似然比的秩为R，则$p=R(1+w)$。

一般情况下，对于P值小于5%（或1%）的时空扫

描窗口，认为该扫描窗口异常存在随机性的概率较小，即该扫描窗口所对应的区域有可能存在传染病的暴发[10-11]。

2 设计与功能实现

2.1 数据上图

自动读取广东省急性传染病监测信息系统里的传染病数据和广东省卫生健康委员会全员人口库数据，实现病例数据和人口数据的地图可视化展示。

2.2 参数设置

时空扫描参数设置，包括扫描类型（前瞻性扫描、回顾性扫描）、疾病类型、开始时间、结束时间、时间步长、空间步长、期望 *P* 值、视角（市级、区县级、街镇级）、人口基数（省级、市级）等。见图 1。

图 1 时空扫描参数设置

2.3 时空扫描

进行时空扫描参数设置后，根据病例数据和每年度各行政区划的人口数据，进行时空聚集性扫描。

2.4 扫描结果可视化展示

扫描结果将根据 LLR 值进行分级展示，并能随着时间的推移，在地图上动态展现传染病聚集性的变化，见图 2—图 5。

2.5 扫描结果列表导出

扫描结果列表导出，前瞻性时空扫描结果列表包括 cluster 序号、发出预警日期、信号包括的天数、实际发

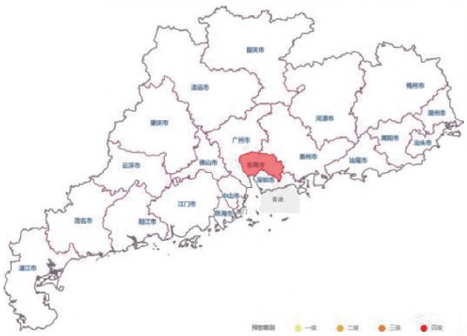

图 2 2019 年 7 月 6 日至 7 月 12 日聚集性结果

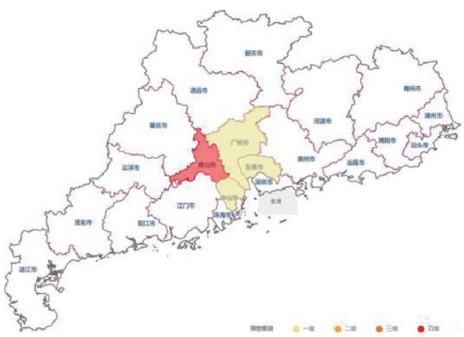

图 3 2019 年 7 月 13 日至 7 月 19 日聚集性结果

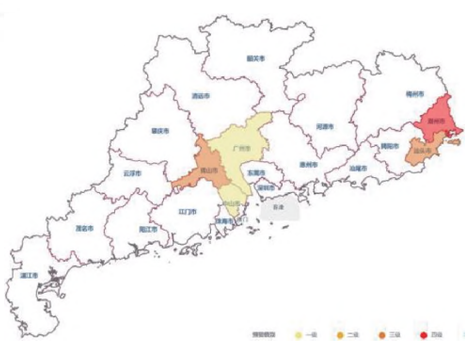

图 4 2019 年 7 月 20 日至 7 月 26 日聚集性结果

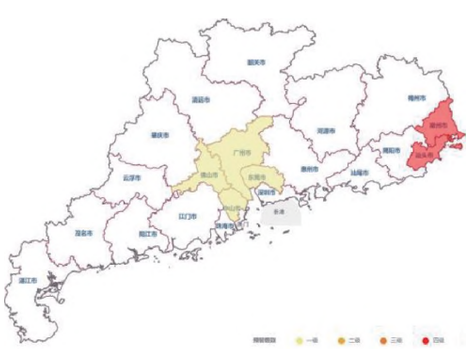

图 5 2019 年 7 月 27 日至 8 月 3 日聚集性结果

表 1　扫描结果列表

登革热疾病专题回顾性扫描时空聚集性探测结果

	cluster序号	地址	空间步长(km)	时间步长	开始时间	结束时间	P	相对危险度（RR）	似然比（LLR）	预期发病数	实际发病数	人口数	发病率(/十万)
1	1121	广东省东莞市	50	7天	2019-07-06	2019-07-13	0.001	2.55556	10.22175	8	23	54175366	0.0425
2	676	广东省佛山市	50	7天	2019-07-13	2019-07-20	0.001	3	16.15312	9	30	60634217	0.0495
3	643	广东省东莞市	50	7天	2019-07-13	2019-07-20	0.002	2.11111	5.76231	8	19	54175366	0.0351
4	747	广东省广州市	50	7天	2019-07-13	2019-07-20	0.002	1.70588	4.97522	16	29	103058977	0.0281
5	535	广东省中山市	50	7天	2019-07-13	2019-07-20	0.008	2	4.34322	4	10	23727571	0.0421
6	836	广东省潮州市	50	7天	2019-07-20	2019-07-27	0.001	5.5	28.12163	3	22	17626651	0.1248
7	972	广东省佛山市	50	7天	2019-07-20	2019-07-27	0.001	3.4	21.91273	9	34	60634217	0.0561
8	1037	广东省汕头市	50	7天	2019-07-20	2019-07-27	0.001	3.57143	19.50747	6	25	35696206	0.07
9	961	广东省广州市	50	7天	2019-07-20	2019-07-27	0.001	2.11765	10.93433	16	36	103058977	0.0349
10	856	广东省中山市	50	7天	2019-07-20	2019-07-27	0.001	2.6	8.58721	4	13	23727571	0.0548
11	234	广东省潮州市	50	7天	2019-07-27	2019-08-03	0.001	20.5	231.09427	3	82	17626651	0.4652
12	264	广东省汕头市	50	7天	2019-07-27	2019-08-03	0.001	13	198.94247	6	91	35696206	0.2549
13	228	广东省佛山市	50	7天	2019-07-27	2019-08-03	0.001	4.2	35.32695	9	42	60634217	0.0693
14	102	广东省广州市	50	7天	2019-07-27	2019-08-03	0.001	2.76471	23.94215	16	47	103058977	0.0456
15	165	广东省东莞市	50	7天	2019-07-27	2019-08-03	0.001	2.77778	12.82882	8	25	54175366	0.0461
16	271	广东省中山市	50	7天	2019-07-27	2019-08-03	0.003	2.4	7.04546	4	12	23727571	0.0506

病数、期望发病数、似然比（LLR）、相对危险度（RR）等；回顾性时空扫描结果列表包括 cluster 序号、开始时间、结束时间、实际发病数、期望发病数、LLR、RR 等，见表 1。

3　结果

通过本系统对广东省 2019 年 1 月 1 日至 8 月 3 日登革热病例数据进行逐周前瞻性时空扫描，与利用 satscan 软件进行时空扫描得到的结果基本一致。限于篇幅，仅列出 7 月 6 日至 8 月 3 日的分析结果。从表 1 可见，7 月 6 日至 7 月 13 日，东莞市登革热实际发病数为 23 例，前瞻性时空扫描统计量计算出的预期发病数为 8 例，LLR 值为 10.22。从图 3 至图 6 可见，聚集首先发生在东莞市，随后扩散到佛山市、广州市、中山市，再逐步蔓延到潮州市、汕头市等地，波及区域达 6 个地市。

4　讨论

本平台能够根据用户自定义的时间周期（如 7 天）对传染病数据进行在线自动前瞻性时空扫描，从而实现传染病疫情的动态聚集性探测，及早干预，有效控制疾病的传播，减轻危害[12]。用户不再需要使用 Satscan 等多个软件实现传染病时空聚集性分析，简化了分析过程，减少了人工干预，提高了工作效率。同时结合地理信息系统，更加直观、全面地展示了发病聚集区域，为以后采取有针对性的预防控制措施，提供了科学的参考依据[13-14]。若在进行时空聚类分析时，以患者的实际活动范围为依据，可能会使其在传染病预警工作中发挥更大的作用，这也是我们今后研究的方向[15]。■

参考文献

[1] 王占宏 . 基于扫描统计方法的上海犯罪时空热点分析 [D]. 华东师范大学 , 2013.

[2] 黄莉 , 许琳 , 李石柱 . 扫描统计量方法在肺结核发病分布中的应用进展 [J]. 实用预防医学 , 2017, 24(6): 132-135.

[3] 杜长慧 , 殷菲 , 尹仲良 , 等 . 2005 年成都市麻疹模拟实时监测与预警研究 [J]. 职业卫生与病伤 , 2010, 25(5): 261-264.

[4] 徐聪 . 基于 GIS 的小空间尺度传染病时空预警技术 [D]. 浙江大学 , 2014.

[5] 周丽君 , 张兴裕 , 马越 , 等 . 前瞻性时空扫描统计量与时空重排扫描统计量在传染病聚集性探测中的适用性探讨 [J]. 现代预防医学 , 2012, 39(5): 1068-1070, 1077.

（下转第551页）

平台的实践，自 2017 年 3 月开始。目前，上海市各级医疗卫生机构系统用户均已实现了基于数字证书的强身份认证方式，登录访问中国疾病预防控制信息系统，强化了信息系统的身份安全，有效提升了信息系统遭受破坏性攻击的防范能力。今后，还将在此实践基础上，进一步推进电子认证服务体系应用，以便更好地保障疾病预防控制领域的网络信息安全。■

参考文献

[1] 李言飞，马家奇 . 中国疾病预防控制信息系统问题分析与对策 [J], 中国卫生信息管理杂志 , 2013,10(6): 230-232.

[2] 马家奇 . 中国疾病预防控制信息系统建设的现状 [J]. 疾病监测 , 2002, 17(2): 69-71.

[3] 任敏，董俊善，周马 . 中国疾病预防控制信息系统运行效果分析 [J]. 预防医学情报杂志 , 2012, 28(7): 558-559.

[4] 张睿，葛辉，杜雪杰，等 . 中国疾病预防控制信息系统电子认证服务建设思路 [J]. 中国卫生信息管理杂志 , 2013, 10(6): 507-510.

[5] 王俊玲，姜韬，张烨 . 中国疾病预防控制信息系统安全性分析 [J]. 中国疫苗和免疫 , 2005, 11(4): 320-321.

[6] 王俊玲，张烨，张睿，等 . 突发公共卫生事件应急机制监测信息系统安全体系设计 [J]. 中国公共卫生管理 , 2005, 21(4): 283-284.

[7] 刘军，韩冬，黄家忠 . 天津市疾病预防控制信息系统数字认证的实现 [J]. 医疗卫生装备 , 2018, 39(2): 56-59.

[8] 张博，朱旋，高炽扬 . 数字证书互信互认技术探讨 [J]. 网络安全技术与应用 , 2012, 12(9): 35-37.

[9] 胡向禹，张洪亮 . 中国疾病预防控制信息系统云认证服务模式的建设与应用 [J]. 信息安全研究杂志 , 2017, 3(5): 554-559.

[10] 王苏灵，杜彪，邓雷升 . 数字证书在线验证系统设计 [J]. 通信技术 , 2017, 50(5): 1079-1083.

[11] 程国青 . 数字证书在信息化项目中一证多用的研究与实现 [J], 中国管理信息化 , 2017, 13(3): 136-138.

[12] 肖晓赟 . 推行组织机构数字证书，实现 " 一证多用 " 的网络身份标识 [J]. 中国电子商务 , 2010, 11(8): 32-33.

[13] 周立兵，周大伟 . 基于数字证书的访问控制研究 [J]. 计算机与数字工程 , 2011, 39(1): 114-116.

[14] 郭亓元，沈宇超，岑荣伟 . 电子政务数字证书互认平台的研究与设计 [J], 信息安全研究 , 2017, 3(6): 548-553.

[15] 赵士洁 . 卫生部印发《卫生系统电子认证服务管理办法 (试行)》通知 [J]. 中国数字医学 , 2010, 5(2): 5.

[收稿日期：2019-08-01　修回日期：2019-09-03]

（上接第536页）

[6] 孙玮璇，廖志林，毛绍菊，等 . 基于前瞻性时空重排扫描统计量的手足口病早期预警方法及其实证研究 [J]. 南昌大学学报 (医学版), 2015, 60(6): 80-83.

[7] 陈飞，李晓松，冯子健，等 . 基于超几何分布的前瞻性时空扫描统计量在疟疾早期预警中的应用 [J]. 中国卫生统计 , 2015, 32(2): 186-189.

[8] 孙玮璇，廖志林，毛绍菊，等 . 基于前瞻性时空重排扫描统计量的手足口病早期预警方法及其实证研究 [J]. 南昌大学学报 (医学版), 2015, 60(6): 80-83.

[9] 董倩楠 . 基于 Kulldorff 扫描统计量的聚类方法研究及应用 [D]. 中国石油大学 , 2016.

[10] 张婷，程昌秀，杨山力，等 . 时空聚集性探测方法在极端高温事件聚集分析中的应用研究 [J]. 地理与地理信息科学 , 2019, 35(3): 51-57.

[11] 高凌云志，宋肖肖，龙华，等 . 基于前瞻性时空扫描统计量的蒙特卡罗重排扫描方法的研究 [J]. 通信技术 , 2018, 51(10): 67-72.

[12] 王显科，胡宇峰，方明金，等 . 前瞻性时空扫描分析在卫生应急指挥决策系统疾病预警模式中的应用研究 [J]. 重庆医学 , 2013, 42(31): 3795-3797.

[13] 童爽，徐慧兰 . 2010 年至 2015 年常德市丙类传染病发病时空扫描分析 [J]. 中国医师杂志 , 2018, 20(3): 394-397.

[14] 钱海坤，杨鹏，张奕，等 . 2005-2010 年北京市猩红热发病时空扫描分析 [J]. 疾病监测 , 2011, 26(6): 435-438.

[15] 安庆玉，吴隽，姚伟 . 传染病自动预警信息系统与时空聚类分析在大连市金州区风疹暴发预警中的效果比较 [J]. 疾病监测 , 2010, 25(7): 577-579.

[收稿日期：2019-08-09　修回日期：2019-08-27]