

## · SARS 传播动力学模型 ·

## 北京市 2003 年 SARS 疫情的多维分布及其影响因素分析

王劲峰 孟斌 郑晓瑛 刘纪远 韩卫国 武继磊 刘旭华 李小红 宋新明

**【摘要】** 目的 采用多维可视化分析研究北京市 2003 年严重急性呼吸综合征(SARS)疫情的扩散过程。方法 以北京市 SARS 疫情数据,结合北京市地理信息系统,基于数据驱动和模型驱动的理论和技术,利用热点分析、空间过程分析和因子识别等数据探索分析方法,根据传染病的多维传播特性,同时利用遗传规划和模拟退火相结合的算法对易感-感染-移出(SIR)模型求解,从模型直接求取 SARS 流行病学参数。结果 SARS 密切接触者在城市内部呈现出大尺度上沿交通线聚集和在小尺度上随机分布态势;在不同发展阶段,北京市 SARS 发病趋势存在显著的空间聚集和扩散变化特征;地理位置、人口以及医院和医生数量是 SARS 空间传播的重要影响因子;通过对 SIR 模型直接求解,反演传染病参数,对 SARS 确诊病例数可进行早期预报性分析。结论 对 SARS 密切接触者的集聚探测揭示了北京市人群存在两个空间尺度上的流动接触过程,为人-人接触性传染病的预防和控制策略制定提供重要依据;多个环境和人文因子对 SARS 传播起作用,其统计显著性随时间变化;有效的算法可以对 SIR 直接求解,使其可以用于传染病参数反演和早期预测。

**【关键词】** 严重急性呼吸综合征;空间数据分析;模型;易感-感染-移出

**Analysis on the multi-distribution and the major influencing factors on severe acute respiratory syndrome in Beijing** WANG Jin-feng\*, MENG Bin, ZHENG Xiao-ying, LIU Ji-yuan, HAN Wei-guo, WU Ji-lei, LIU Xu-hua, LI Xiao-wen, SONG Xin-ming. \*Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

**【Abstract】 Objective** To analyse the multi-dimension nature of severe acute respiratory syndrome (SARS) transmission. **Methods** Based on the data of SARS in 2003 and the geographic information system of Beijing, as well as under the broad range of the theorems and techniques of data-driven and model-driven knowledge mining, hierarchical techniques were used to test the hot spots. Wavelet technique was also used to decompose Moran's I frequency to survey the spatial clustering process of SARS. For factors analysis, BW test was used to distinguish factors which influencing SARS process. In temporal aspects, susceptible-infective-removal model (SIR) without Taylor expansion was solved by a genetic-simulated annealing algorithm, that directly provided a new approach to obtain epidemic parameters from the SIR model. **Results** Different order of spatial hot spots were noticed and the clustering were relevant with the means of transportation. Diffusion dynamics were changed along with the temporal process of SARS. Regarding factor analysis, geographic relationship, population density, the amount of doctors and hospitals appeared to be the key elements influencing the transmission of SARS. The predictable number of SARS cases evolving with time were also calculated. **Conclusions** Cluster detection of close contacts of SARS infective in Beijing revealed the spatial characters of urban population flow and having important implications in the prevention and control of this communicable diseases. Some human and physical environment factors played statistical significant roles in different periods during SARS epidemics. An efficient algorithm was developed to solve SIR model directly, enabling the estimation of epidemic parameters from SIR and early forecast.

**【Key words】** Severe acute respiratory syndrome; Spatial data analysis; Susceptive-infective-removal model

基金项目:国家自然科学基金(JJ03000101、49871064);国家“863”(2002AA135230-1)、“973”(2001CB5103)高技术研究发展计划专项基金和中国医学科学院知识创新工程资助项目

作者单位:100101 北京,中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室(王劲峰、孟斌、刘纪远、韩卫国、武继磊、刘旭华);北京大学人口研究所(郑晓瑛、宋新明);中国科学院遥感应用研究所(李小红)

对传染病环境因素的分析是充分认识传播规律和机制不可或缺的部分<sup>[1]</sup>。传染病的传播具有多维性,表现在疫情随时间、空间、年龄、职业、交通工具、社会活动、人口流动、营养等的变化。我们应用空间分析理论和技术,分析了可用于传染病传播的数据,构建传染病多维分析的方法,并以 2003 年北京市严重急性呼吸综合征(SARS)的实际数据验证了该方法的可行性和可靠性。

## 基本原理

1. 空间数据探索分析:空间数据探索(exploratory spatial data analysis, ESDA)主要关注如何利用可视化技术和简单的统计指标从尽可能多的角度对数据进行描述和洞察。ESDA 主要围绕以下一个或多个方面展开:①对数据异常值或错误值的识别;②数据分布模式的识别;③通过数据形成假设和④模型的检验。ESDA 一般可以进行可视化表达,即用图像、曲线、二维图形、三维体和动画来显示数据,并可对其模式和相互关系进行可视化分析<sup>[2]</sup>。在流行病学研究中,ESDA 主要内容包括数据探索分析、传播不同阶段的影响因子识别、传播源探测、动态危险区划和参数库建设。

(1)数据探索分析:利用各种统计图表(如散点图、直方图)对患者的职业、年龄、发病时间和地点等进行分析 and 统计分布检验及对疫情扩散的重要因子的计算,如北京市各级医院可接近度、北京市各点的人口密集度和交通可接近度等。数据探索分析的结果可以为进一步建模和分析提供服务。

(2)热点探测:热点(hot spots)是病例在时间和空间上大量集中的表现。在研究北京市 SARS 空间特征时,采用空间聚集分析探测,主要分析发生病例在空间上的热点区域。该方法类似聚类分析,根据某种规则(如邻近距离),通过定义一个“聚集单元”(cluster)及其“极限距离或阈值”,然后比较聚集单元与每一点的距离,当某一点与其距离小于该极限距离,该点被计入聚集单元,据此将原始点数据聚类为若干区域,称为一阶(first order)热点区;对一阶热点区利用同样方法,聚类得到二阶(second order)热点区,以此类推更高阶热点区。此外,可以利用空间信息多维区划和复杂时空信息结构自适应模型进行区域风险估计预测和动态空间区划<sup>[3,4]</sup>。

(3)空间过程分析:在 SARS 研究中,最重要的一个问题就是判断疫情在空间上是扩散还是在积聚

状态。采用空间自相关指标 Moran's I 检测 SARS 的积聚或扩散状态,公式:

$$I = \frac{n \cdot \sum_i \sum_j w_{ij} \cdot (y_i - \bar{y})(y_j - \bar{y})}{(\sum_i \sum_j w_{ij}) \cdot \sum_i (y_i - \bar{y})^2} \quad (1)$$

式中: $n$  为空间对象数目, $y_i, y_j$  为观测值, $w_{ij}$  为研究对象  $i, j$  之间的权重,其中  $w_{ij} = 0$ ,权重矩阵的确定,一般有两种方式,即相邻矩阵和距离矩阵。Moran's I 的取值在  $-1 \sim 1$  之间,理论上,  $I$  的期望值为  $-1/(n-1)$ ,其检验一般采取得分检验方式(score test),即  $Z = [I - E(I)]/std(I)$ ,当  $|Z| > 1.96$  时,可以认定为小概率,拒绝  $H_0$  假设( $P < 0.05$ ),可以认定为存在空间相关性。Moran's I 检验存在空间相关性时,表明疾病在空间存在显著的积聚状态<sup>[5]</sup>。

(4)传播不同阶段的影响因子识别:传染病从起源、初期、快速发展期、顶峰期、下降期、结束期等疫情过程的不同阶段,不同的因子对于传染病的空间扩散起的作用是动态变化的。SARS 的多维传播受到多种因子影响,对这些因子的提取及其作用的识别对于政府制定防控措施具有方向性的参考价值。本研究中采用空间统计模型中“连接数目统计”(join-count statistics)方法来检验不同影响因子的假设。

在连接数目统计中,对每个研究单元(区)按其是否发生病情分别标识为 Black(B)或 White(W)。如果相互连接的两个区域有且只有一个区域发生疫情,则称为 BW 连接。通过比较计算出的 BW 实际数目和基于不存在空间相关的零假设的理论 BW 数目,并进行显著性检验。其检验公式:

$$Z(BW) = N_{bw} - E(N_{bw}) / \sqrt{Var(BW)} \quad (2)$$

式中: $N_{bw}$  为实际的 BW 连接的数目, $E(N_{bw})$  为不存在空间相关的零假设下的理论 BW 连接数目。 $Var(BW)$  为 BW 连接的数目的标准差。当  $|Z(BW)| > 1.96$ ,表明在 95% 的显著性水平下可以推翻不存在空间相关的零假设,即疫情存在空间集聚特点。通过定义不同的连接方式,可以检验不同因素对疫情空间扩散的影响效果<sup>[6]</sup>。

2. 模型与模拟分析:传染病的时空传播模型包括空间统计学、数学模型和智能计算技术以及时空传播动力学模型和宿主模型。

(1)空间统计学、数学模型和智能计算技术:在空间统计学中,可以利用 Kulldorff's Scan<sup>[7]</sup>、空间回

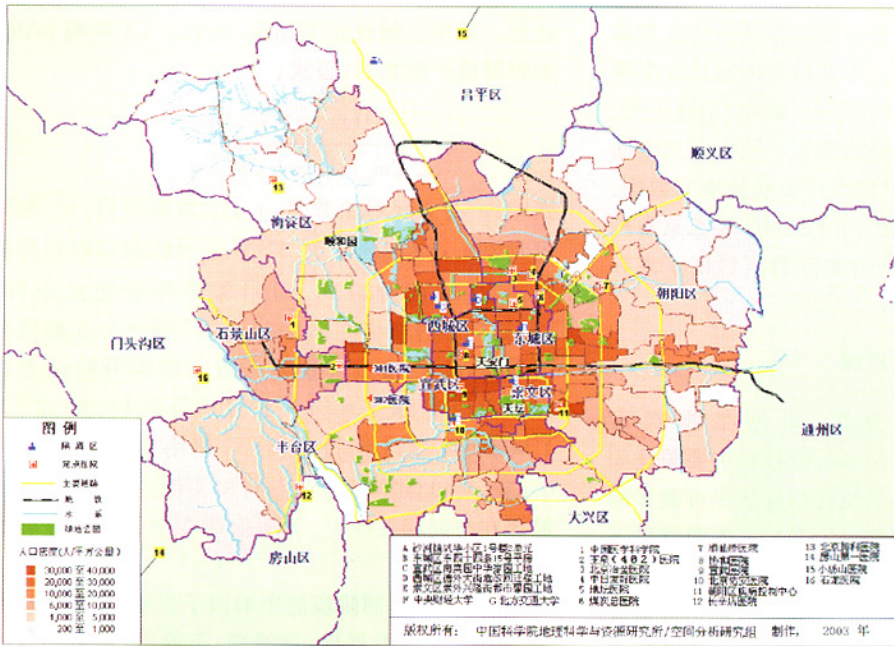


图1 北京市 2003 年 SARS 期间定点医院、人口密度、主要道路及公园绿地的分布

我们利用遗传算法和退火算法结合, 可以对方程式(3-6)的进行直接求解, 并且求解得到的参数具有明确的物理意义, 在初期就可以进行预报。

(2)时空传播动力学模型: 目前传染病随时间和空间传播的数学描述是基于质量、能量守恒的扩散方程和传热量方程来构建, 但这明显与传染病传播机制不符, 为此有必要建立符合传染病传播机理的时

空传播模型<sup>[11]</sup>。归分析如 SAR、MA、CAR 模型等研究传染病时空传播特征<sup>[8,9]</sup>。成熟的数学模型包括确定性 SIR (susceptive-infective-removal) 和 SEIR 模型<sup>[10]</sup>、随机 Greenwood 和 Reed-Frost 模型<sup>[11]</sup>。时间走线的 SIR 一般方程本身已经成熟, 但是由于该方程具有刚性, 在数学上, 除个别特殊形式, 刚性方程尚无一般解法。因此, 到目前为止, SIR 模型的求解均是泰勒级数展开的简化近似解, 但在其简化形式中参数物理含义不明, 因此, 要用 SIR 模型进行预报, 需要实际走线大体成型后经曲线拟合方可实现。SIR 模型具体形式:

$$\frac{dS}{dt} = -kS(t)I(t) \quad (3-1)$$

$$\frac{dI}{dt} = kS(t)I(t) - lI(t) \quad (3-2)$$

$$\frac{dR}{dt} = lI(t) \quad (3-3)$$

式中  $S(t)$ 、 $I(t)$  和  $R(t)$  分别表示  $t$  时刻易感人数、感染例数和移出人数, 满足式(3-4):

$$S(t) + I(t) + R(t) = N \quad (3-4)$$

设  $S_0$  为初始易感人数, 假定  $R(0) = 0$ , 由式(3-1)、(3-2)、(3-3)和(3-4)得出

$$\frac{dR}{dt} = l[N - R(t) - S_0 \exp(-\frac{R(t)}{\rho})] \quad (3-5)$$

由式(3-3)和(3-5)得到  $I(t)$  的表达式

$$I(t) = N - R(t) - S_0 \exp(-\frac{R(t)}{\rho}) \quad (3-6)$$

空传播模型<sup>[11]</sup>。

### 实例分析

1. 研究对象与数据来源: 本研究采用 2003 年北京市 SARS 疫情的实际数据, 包括: ①SARS 确诊病例、疑似病例和密切接触者的位置信息; ②北京市电子地图数据(人口分布、航空、铁路、公路、城市交通); ③北京市各级医院分布(地点、性质、医生人数、床位数); ④全国和北京市 SARS 疫情大事记, 含政策措施、疫情报道、传播链分析等。图 1 为 2003 年北京市 SARS 发病期间收治 SARS 患者的定点医院、各街道的人口密度、主要道路及公园绿地的分布情况。

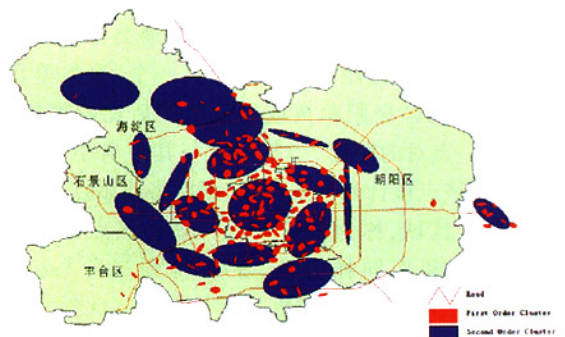


图2 北京市 2003 年 SARS 疫情中密切接触者的空间热点探测

### 2. SARS 疫情的热点探测: 利用层次分析

(hierarchical techniques)技术,对北京市 SARS 疫情中密切接触者的空间分布进行热点探测, SARS 疫情的热点区域具有以下特点(图 2):①一级热点区(红色)集中在三环路以内,且呈空间随机分布;②二级热点区(蓝色)则呈明显的环状分布,向西和西北方向延伸;③在东边的通州区内,有一些“飞点”形式的一级热点区,并形成二级热点区。

3. SARS 疫情的空间过程分析:根据北京市 18 区县 2003 年 4 月 27 日至 5 月 18 日 SARS 新确诊病例公布数据,利用公式(1),计算出 Moran's I (图 3 中最上部分),表明在此期间,北京市 SARS 疫情从存在显著的空间积聚特性向空间随机演变。为更加清楚地揭示不同阶段 SARS 的空间趋势,对 Moran's I 的结果进行小波分解。小波分解主要是将数据的总体趋势和其噪声(即随机项)区分开来,其基本过程就是将原始数据分解为总体趋势和表示细节或噪声的两个组成部分,根据需要,可以进一步对分解后的总体趋势部分做类似的分解。对北京市 SARS 疫情 Moran's I 分解层次为 4(图 3),其中  $a_4$  表示 SARS 扩展的空间聚集和发散的总体变化趋势,  $d_1 \sim d_4$  表示不同尺度的聚集或发散的震荡幅度:①2003 年 4 月 27-30 日, Moran's I 逐渐增加,表明北京市 SARS 疫情空间积聚状态逐渐增强;②2003 年 5 月 1-15 日, Moran's I 逐渐减小,表明北京市 SARS 疫情空间积聚状态逐渐减弱,呈现空间扩散状态;③2003 年 5 月 16-18 日, Moran's I 接近于零,表明北京市 SARS 疫情的空间格局呈现无序状态,且如果其他情况没有变化,这种无序会持续到疫情结束。

4. 多因子识别:根据 SARS 空间传播影响因子的各种猜测以及可量化识别的能力,我们列出了以下候选因子:医院和医生数量、区域毗邻、人口规模、人口密度、交通方便度等,并以这些因子构建连接数目统计中的区域连接方式,根据北京市 18 区县 2003 年 4 月 27 日至 5 月 23 日 SARS 新确诊病例公布数据,采用公式(2)进行分析。结果表明:①“区际相邻性”在北京 SARS 新增病例高发期(2003 年 4 月 29 日前后),对 SARS 的空间传播作用显著;②“空间可接近度”对北京市 SARS 空间传播的影响与“区际相邻性”作用相似,这可能与北京市棋盘型空间结构有关;③“区域人口总数”对北京市 SARS 空间传播影响不明显;④“区域人口密度”在北京市 SARS 新增病例高发期,是 SARS 空间传播的重要

因子;⑤“区域医生总数”在北京市 SARS 新增病例高发期,是 SARS 空间传播的重要因子,反映医生通过家庭、工作过程传染导致传播;⑥“区域医院总数”对北京市 SARS 区际传播影响不明显,反映医院对疫情的区际传播作用很小。关于“医院成为新的污染源造成疫情的(区内)传播”的观点需有更加细致的空间数据经检验后进行判断;⑦“城乡关系”表明 5 月 13-19 日有一次显著的扩散过程。

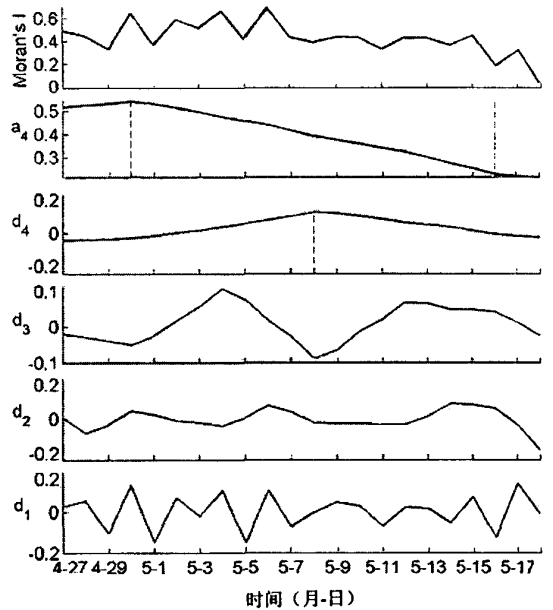
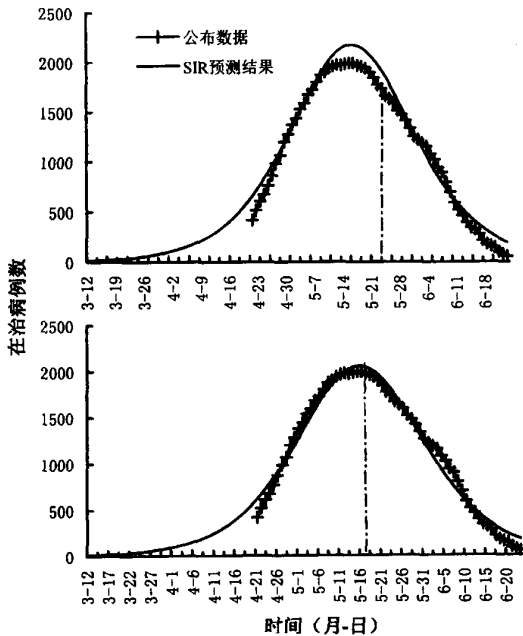


图3 Moran's I 及其小波分解

5. SIR 模型时间维预测:利用遗传算法和退火算法结合调试成功对 SIR 方程的直接求解,求解得到的参数具有明确的物理意义,可以在初期就进行预报。用此技术在北京市 SARS 初期(2003 年 4 月 21 日至 5 月 8 日),利用公式(3-6)对 SARS 确诊病例全程走线进行了预测(图 4 上部)。因为在北京市 SARS 上升期,新加坡和我国香港的疫情已接近尾声,该两地具有物理含义的传染病参数已有不少研究<sup>[12,13]</sup>,并且三地均是城市环境,人口密度和人间接触具有相似性。因此,考虑到 SARS 传播过程的时延性,采用 SIR 模型对香港和新加坡两地的数据进行计算,所得参数用于北京市 SARS 传播趋势的预测。图 4 下部是利用 4 月 21 日至 5 月 18 日的实测数据回归 SIR 模型参数,再次进行拟合预报的结果,这时,实际走线已基本呈形,拟合和预报精度自然较初期预报有所提高。

讨 论

对 SARS 的密切接触者的空间热点探测表明,



上部分为 2003 年 5 月 8 日绘制; 下部分为 2003 年 5 月 19 日绘制

图4 北京市 2003 年 SARS 确诊病例趋势预测

“一级热点”揭示了个体接触史,主要集中在三环路以内,具有空间随机性,表明北京市人口分布及社会、经济活动的密集区域还是在城市中心区域,该区域应该是传染病监测与预防的重点地区;图 2 可以发现二级热点区分布与北京市环状交通路网格局和城市轻轨交通有相当大的视觉关联度,表明“二级热点”在一级热点区的基础上,进一步揭示了北京市人群流动的空间特征,也从一个侧面证明北京市采取的“早报告、早隔离”政策的有效性。

对 SARS 疫情的空间集聚特征的分析表明, SARS 疫情在空间上具有显著的相关性,说明疫情的空间扩散过程非常明显。同时,随着对疫情的控制措施的加强, SARS 疫情在空间上的相关性也在逐步下降,且演变为不具有空间相关性。我们的分析表明,从 2003 年 5 月 15 日开始,空间相关性的趋势发生显著变化。而北京市 SARS 疫情的实际趋势(图 3,4 所绘 SARS 实际数据也在 5 月 15 日左右到达高峰)与此分析存在一致性。这种一致性背后的原因还需要进一步探讨,但是也提示空间格局分析对传染病的空间传播研究具有重要的参考价值。

采用空间分析方法,对影响传染病扩散的因子进行研究,为了解传染病,尤其是类似 SARS 的新型传染病在空间和人群中的传播途径提供了有力的帮助。研究表明,地理上的邻近和人群交往,是疾病传播的一种重要方式,而人口密度也较人口数量对 SARS 的传播更

具有特别重要的意义。同时,在 SARS 疫情中,医院和医护人员也存在重要的影响,我们的研究也证实这一点。因此,通过对传播因子的识别和分析,可以提高控制和预防疾病的能力与水平。

在传染病模型研究中,除了基于微分方程的动力学模型如 SIR,基于统计学的概率模型也相当成熟。后者可以模拟传染病在个体至人群中的逐步传播过程,并对走势进行同步的机理解释。在有翔实的传染参数的条件下,这是一个相当理想的数学手段。前者对传染病走线全过程有一个系统的数学描述,但是模型参数需要是常值,或者是需要采用级数展开来简化模型,削弱了动力学模型探视传播过程的能力。我们利用遗传规划和模拟退火相结合的算法对 SIR 模型求解,为从模型直接求取传染病的参数提供了新途径。在疫情初期便具有相当现实和比较准确的预报能力。根据 SIR 模型预测发病情况与实际发展存在一定距离,主要原因在于 SARS 后期,大量行之有效的人工干预手段的合理使用,有效地遏制了疫情的进一步发展。

参 考 文 献

- Carlsonl C, Eberle M, Kruglyak L, et al. Mapping complex disease loci in whole-genome association studies. *Nature*, 2004, 429:446-452.
- 黄志澄. 数据可视化技术及其应用展望. *电子展望与决策*, 1999, 8:3-9.
- Wang JF, Wise S, Haining R. An integrated regionalization of earthquake, flood and drought hazards in China. *Transactions in GIS*, 1997, 2:25-44.
- Wang JF, Li QL. Adaptive structure model of earthquake trend regionalization. *Earthquake Research in China*, 1996, 12: 78-88.
- Haining R. *Spatial data analysis in social and environmental sciences*. Cambridge University Press, 1990.
- Haggett, P. Hybridizing alternative models of an epidemic diffusion process. *Economic Geography*, 1976, 52:136-146.
- Kulldorff M. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 1997, 26: 1481-1496.
- Haining R. *Spatial data analysis: theory and practice*. Cambridge University Press, 2003.
- Kermarck WO, McKendrick AG. Contributions to the mathematical theory of epidemics. *Proc Roy Soc A*, 1927, 115: 700-721; 1932, 138: 55-83.
- Daley D, Gani J. *Epidemic modelling*. Cambridge University Press, 2001.
- Wang JF. Modeling epidemic in space and time. *Annual Conference of the Association of American Geographers*, Los Angles, US, 2002.
- Riley S, Christophe F, Christl AD, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science*, www. sciencexpress. org/23 May 2003/Page 5/10. 1126/science. 1086478, 2003
- Christ AD, Azra CG, Gabriel ML, et al. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. <http://image. the lancet. com/extras/ 03art4453web. pdf> 2003

(收稿日期:2004-07-29)

(本文编辑:张林东)