

文章编号:1005-3085(2003)07-0045-08

非典数学模型的建立与分析

王议锋, 田 一, 杨 倩

指导老师:尚寿亭

(哈尔滨工业大学, 哈尔滨 150001)

编者按:这是一篇公布在<http://www.hit.edu.cn/hmcm/20030626/a1.doc>上的论文,完成于 2003 年 6 月 9 日。2003 年高教社杯全国大学生数学建模竞赛中,有些做 A 题的队部分地引用了这篇论文,也发现一些队明显地抄袭了它的结果,未将该文列入参考文献,也未在论文中指出,违反了他们自己签名的保证书,全国组委会已进行了严肃、认真的处理。现在我们刊登该文,以供参考。由于篇幅的限制,请作者进行了压缩和删节,简化了文字的叙述。

摘 要:本文以 2003 年 6 月以前的有关数据为资料,在传统的 SEIR 传染病模型的基础上,对人群作了合理的分类,建立了控制前传播模型和控制后传播模型,通过合理估计、曲线拟合和概率平均的方法得到了各个参数。重点分析了控后模型,用龙格-库塔法求解了方程,并对北京、内蒙古、广东、香港四个 SARS 重点疫区的疫情作了具体的分析,最后评价了模型的合理性、实用性,提出了模型的改进方向和思路。

关键词:微分方程;概率平均;龙格-库塔方法;曲线拟合

分类号:AMS(2000) 34B08;41A20

中图分类号:O241.81

文献标识码:A

1 问题的提出(略)

2 数学模型的分析与建立

2.1 分析与假设

SARS 爆发初期,政府和公众对其重视程度远远不够;当被感染者大幅度增加时,政府才开始采取多种措施以控制 SARS 的进一步蔓延。所以 SARS 的传播可以分为三个阶段:

- a) **控制前**,接近于自然传播时的传播模式。
- b) **过渡期**,在公众开始意识到 SARS 的严重性到政府采取得力措施前的一段时间内。
- c) **控制后**,在介入人为因素之后的传播模式。

我们统一将所有地区的 SARS 传播规律用“控制前”和“控制后”两个时期来分析。

1) 总体假设

1. 假设一个 SARS 康复者不会二度感染,他们已退出传染体系,因此将其归为“退出者”。

2. 不考虑这段时间内的自然出生率和死亡率由 SARS 引起的死亡人数, 归为“退出者”。

3. 假设潜伏期为一常数 $\tau = 5$ 天。^[1]

4. 根据国家卫生部资料可知处于潜伏期的 SARS 病人不具有传染性。

2) 控制前(包括控制力度不大的阶段)的传播模型的相关假设

1. 直接接触。

2. 在疾病传播期内所考察的地区的总人数 N 视为常数。

3. 设每个病人单位时间有效接触的人数可视为常数。

4. 流入和流出的人群中的带菌者处于潜伏期。

5. 将人群分为四类:

健康者(易受感染者): 用 S 表示健康者在人群中的比例。

处于潜伏期者: 用 E 表示他们在人群中的比率。

病人(已受感染者): 用 I 表示病人在人群中的比例。

退出者(包括“被治愈者”和“死亡者”): 用 R 表示退出者在人群中的比例。

3) 控制后的传播模型的相关假设

1. 由于对人口流动加以了限制, 假设此阶段无病源的输入和输出。

2. 设每个病人单位时间有效接触的人数 λ_2 可视为常数。

3. 在控制后阶段, 因与非典传染源或疑似非典传染源接触而被隔离的人群视作健康者。

这部分人在隔离期限过去后又重新进行正常的社会活动, 相当于又进入了传染链中, 故可将他们作为健康者处理。

4. 考虑到采集到的数据, 将人群分为五类:

健康者(易受感染者): 用 S 表示健康者在人群中的比例。

病人(已受感染者): 用 I 表示病人在人群中的比例。

退出者(包括“被治愈者”和“死亡者”): 用 R 表示退出者在人群中的比例。

自由带菌者: 不可控的病毒携带者。用 M 来表示这部分人在人群中的比例。

疑似者: 所有被疑似为非典病的非健康者。包括已出现有关症状但未确诊的被隔离者, 未出现症状但已疑似带菌的被隔离者: 用 Y 表示疑似者在人群中所占比例。

2.2 模型的建立

1) 控前模型的建立

① 参数设定

1) λ_1 ——每个病人平均每天有效接触(足以使被解除者感染)的人数。

2) q ——退出率, 为 SARS 患者的日死亡率和日治愈率之和。

3) l ——(流入)流出人口占本地总人口的比率。

4) ϵ_1 ——处于潜伏期的病人的日发病率。

5) p ——流入人口中带菌者所占的比例。

② 控前方程的建立

根据变量的分析^{[2][3]}, 结合实际的疫情的传播规律, 可以建立如下的方程组(I):

$$\begin{cases} \frac{dS}{dt} = -\lambda_1 IS & (1) \\ \frac{dE}{dt} = \lambda_1 IE - \epsilon_1 E + lP - lE & (2) \\ \frac{dI}{dt} = \epsilon_1 E - qI & (3) \\ \frac{dR}{dt} = qI & (4) \\ S_0 > 0, E_0 > 0, I_0 > 0, R_0 > 0 & (\text{初值}) \end{cases}$$

方程组(I)

③ 参数的确定

1) λ_1 ——根据医学资料和有关数据推导而得。

2) q ——由该城市的医疗水平和已知的统计分析,求其统计平均值。

3) l ——由经济发达程度和交通状况决定。

4) ϵ_1 ——根据医学研究和调查的有关结果和该城市的疫情发展状况可得。

5) p ——由流入该城市人群的地区分布情况和各其他地区的疫情决定。

2) 控后模型的建立

① 补充参数设定

1) y_1 ——疑似中每日被排除人数占疑似人数的比例;

2) y_2 ——疑似者中每日确诊的人数占疑似人数的比例;

3) ϵ ——每个自由带菌者转化为病人的日转化率;

4) λ_2 ——每个自由带菌者发病后被收治前平均每天感染的有效人数;

5) α ——被自由带菌者有效感染的人中可以控制的比率。

6) β ——接触病源的人的发病率

② 控后方程的建立

经分析得到如下的方程组(II)

$$\begin{cases} \frac{dS}{dt} = Yy_1 - \lambda_2 MS & (5) \\ \frac{dI}{dt} = \epsilon M - qI + y_2 Y & (6) \\ \frac{dR}{dt} = qI & (7) \\ \frac{dY}{dt} = -y_1 Y - y_2 Y + \lambda_2 MS\alpha & (8) \\ \frac{dM}{dt} = \lambda_2 MS(1 - \alpha) - \epsilon M & (9) \\ S_0, I_0, R_0, Y_0, M_0 & (\text{初值}) \end{cases}$$

方程组(II)

此模型的优点:(1)明确了疑似者所指的范围;(2)可从现有数据中分析出所需的参数和变量初值;(3)将 λ_2 定义为“有效接触人数”既有利于数据的分析也可减少未知参数的数量。

③ 参数的确定

以北京为例来说明参数分析方法。

a) y_1 ——疑似者的日排除比例:

$$\text{计算公式: } y_1 = \frac{\text{每天新增的疑似排除人数}}{\text{当天疑似病例累计人数}}$$

首先根据卫生部“每日疫情公布”所公布的数据求出相应的 y_1 , 用 Matlab 画图, 如下图 1 所示:

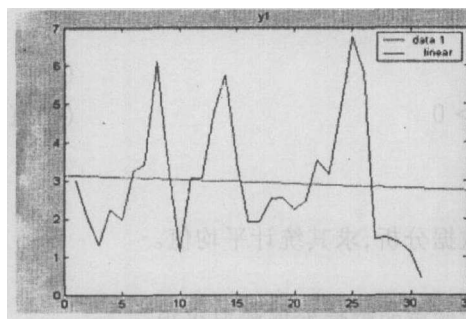


图 1 北京地区 SARS 疑似者的日排除比率 y_1 分布

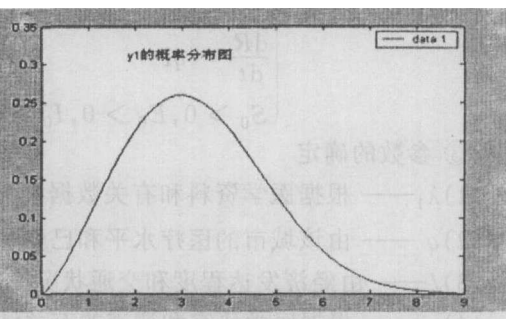


图 2 用 cubic 拟合 y_1

初步用曲线拟合处理一下原始数据, 如图 2 所示(光滑的线为 cubic 拟合曲线)。

可以看出 y_1 有两个峰值, 第一个峰值是由于政府措施很得力, 加之强化控制初期市民的恐慌心理, 导致疑似病例中非感染者所占比例升高; 第二个峰值则是由于大部分带病的疑似者转化为确诊后, 实际未带菌的疑似者相对比例增大造成的。

三阶拟合一定程度上反映了 y_1 的变化规律, 但经分析发现: 如果用原始图来分析误差就会特别大, 不妨去除几个偏离太大的点, 得到图 3, 其中, 平直的线为 linear 拟合直线。

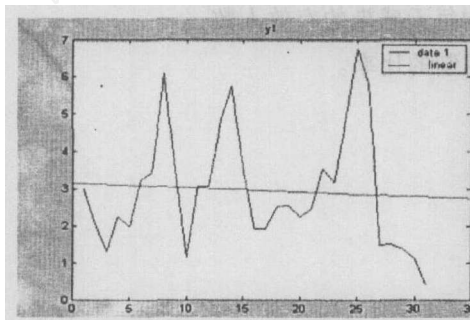


图 3 经处理后的 y_1 的图像以及 linear 拟合曲线

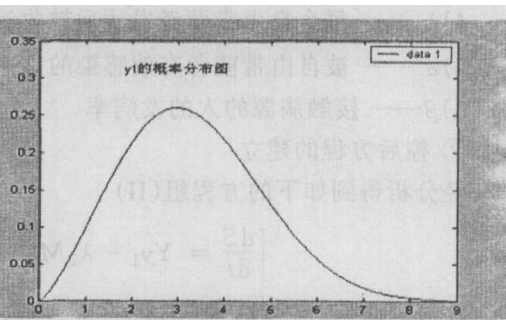


图 4 y_1 的威布尔分布图

现用威布尔分布^[4]观察一下处理后的 y_1 的值的分布情况, 如图 4 所示。可见 y_1 主要分布在 2%—4.5% 之间, 其中概率最大的取值为: 3.51%, 故取 3.51% 为 y_1 的概率平均值。

b) y_2 ——疑似转化为病例的日转化比:

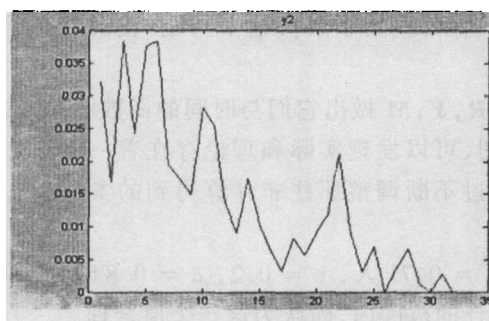
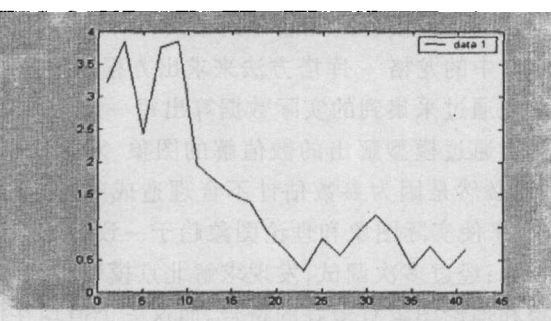
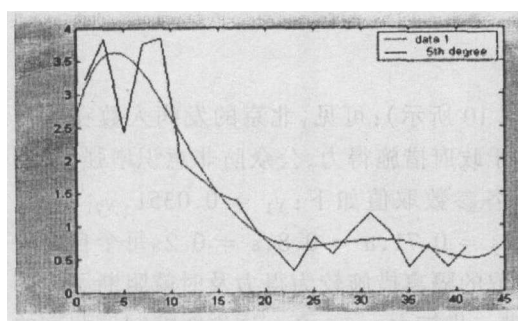
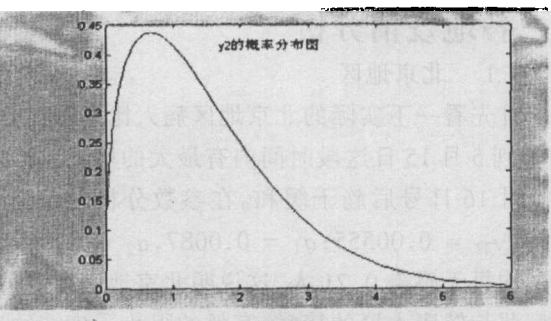
$$\text{计算公式: } y_2 = \frac{\text{每天新增的疑似转化为确诊的人数}}{\text{当天疑似累计人数}}$$

由已知数据求得的每天 y_2 的变化趋势如图 5 所示。去除一些偏离太大的点得到图 6。

从原始数据看到 y_2 总的趋势是下降的。先用曲线拟合处理得到如图 7 所示的图形(光滑的线是 y_2 的五阶拟合线)。

因初期自由带菌者较多, y_2 会有一较大的峰值, 疫情得到重视后 y_2 总趋势是下降的。

最后用威布尔分布来观察一下 y_2 值的分布情况, 如图 8 所示。可以看出 y_2 的值主要分布在 0.05%—2% 之间, 但 y_2 不同于 y_1 分布的那么均匀, 所以我们不能简单的用一个有效

图5 北京地区SARS疑似转化为病例 y_2 图图6 经处理后的 y_2 的图像图7 y_2 的五阶拟合曲线图8 y_2 的威布尔分布图

值来取代 y_2 的值。经过仔细分析,把 y_2 的值等效为两个阶段值:2.229% 和 0.59%,如图9所示 y_2 的两个有效值分布在中直线的两侧。

从对 y_1 与 y_2 的分析来看,可以将强化控制后的这段时间分为两个阶段:过渡期和平稳期;这两个阶段的产生与非典自身的特性有关;同时由于强化控制初期,非典控制力度不够,造成很多自由带菌者与健康者自由接触;加之各项措施从颁布到实行有一段缓冲期等等诸多因素造成了过渡期的形成,其中过渡期特征为: y_2 较大, q (退出率)较小。

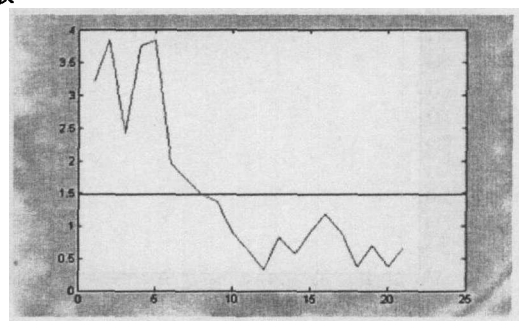
c) q 的计算公式 = $\frac{\text{每天新增的治愈和死亡的人数}}{\text{当天病人累计人数}}$

从 q 的原始数据可看出, q 的值也存在阶段性。5.16日以前, q 的值大概在1%左右摆动,不存在较大的波动;而5.16日以后, q 的值基本都在1%以上。 q 的定义中包括了治愈率与死亡率两部分,在过渡期,由于发病人数较多,治愈率相对较低;当进入平稳期后,发病人数减少,治愈率必然增高。故这与我们上面对于过渡期和平稳期的假设是吻合的。

d) ϵ ——从数据可推算出其值在12%—30%之间,经过分析取 $\epsilon = 20\%$ 。

e) α ——与城市的人口密度、生活习惯等因素有关,由于在强化控制阶段对人员流动控制得相当严格,还采取了诸如封校、小区隔离、公共场合的关闭、减少聚集活动等有效措施,故我们可估计 $\alpha = 70\% \sim 90\%$

4) 模型的求解:

图9 y_2 等效为两个有效值

从建立的模型是无法得到 S, I, R, Y, M 的解析解的。为了解决这个问题,我们求助于 matlab 中的龙格 - 库塔方法来求出方程的数值解。

先通过采集到的实际数据算出每一天的 S, I, R, Y, M 做出它们与时间的函数图象,然后画出通过模型解出的数值解的图象。对比两组图,可以发现实际和理论存在着一定的差异。这必然是因为参数估计不合理造成的。必须通过不断调整那些非计算得到的参数($\lambda_2, \epsilon, \alpha$)来使实际图象和理论图象趋于一致。

如:经过多次调试,发现求解北京模型时,取 $\lambda_2 = 0.71$ 人, $\epsilon = 0.2$, $\alpha = 0.8$ 时,实际图象和理论图象有最好的符合。而这三个值均在估计的范围内,即认为这三个值合理。

3 各地疫情分析

3.1 北京地区

首先看一下实际的北京地区病人比率图(如图 10 所示):可见,北京的发病人数在 4 月 29 日到 5 月 15 日这段时间内有最大的增长率;由于政府措施得力,公众防非意识增强,疫情从 5 月 16 日号后趋于缓和。在参数分析中,北京各参数取值如下: $y_{11} = 0.0351, y_{21} = 0.0299, y_{22} = 0.00555; q_1 = 0.0087, q_2 = 0.025, \lambda_2 = 0.71, \alpha = 0.8, \epsilon = 0.2$ 。每个自由病人平均每天感染 0.71 人,这说明北京地区政府采取的隔离措施较为得力及时就阻断了大部分病源与健康人群的接触,有效的阻止了疫情的进一步扩大; $q_2 = 3q_1$ 这说明控制后北京的治疗水平有较大提高。

3.2 内蒙古地区

图 11 为根据原始数据画出的内蒙古累计病人比率的统计图:由于内蒙古疫情始期较晚,其时全国对非典已较为重视,因而疫情的控制比北京难度要小。数据取定如下: $y_{11} = 0.0271, y_{12} = 0.06, y_{21} = 0.0456, y_{22} = 0.0078; q_1 = 0.014, q_2 = 0.030, \lambda_2 = 0.75, \alpha = 0.78, \epsilon = 0.2$ 。 λ_2 取的是 0.75,比北京的稍高,这是由于内蒙人口比较分散、自由带菌者不易控制; $q_2 = 3q_1$,说明内蒙古卫生部门控后的医疗水平有较大提高。

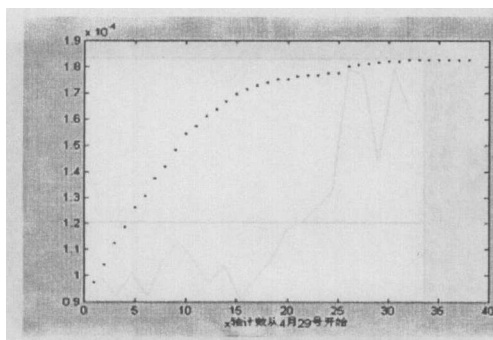


图 10 北京地区原始 I-T 图

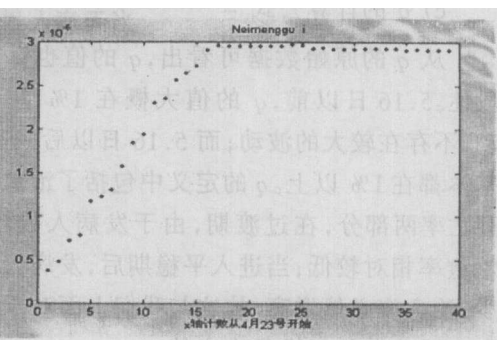


图 11 内蒙古地区原始 I-T 图

3.3 广东地区

广东地区的病人累计比率统计图(如图 12 所示):参数取定: $y_{11} = 0.0694, y_{12} = 0.01, y_{21} = 0.25, y_{22} = 0.001; q_1 = 0.006, q_2 = 0.0065, \lambda_2 = 0.8, \alpha_1 = 0.8, \alpha_2 = 0.85, \epsilon = 0.2$ 。 $q_2 \approx q_1$,可见广东的卫生部门采取的措施开始不是很得力,疫情初期没有得到政府的充分重视,从而导致了疫情的持续,快速增长,一直到国家下达严防命令后才采取有效的措

施,故广东的疫情缓解也是从5月10号左右才开始的。

3.4 香港

图13是根据香港原始数据画出的病人累计统计图。可以看出:香港采取的措施比较早,效果较好。但是由于人口流动很快,导致不可控的病毒传播者比较多,因此,疫情一直呈现出上升趋势,在近期内病人比率还不会会有较大幅度的下降,因此香港的非典工作需加大力度。

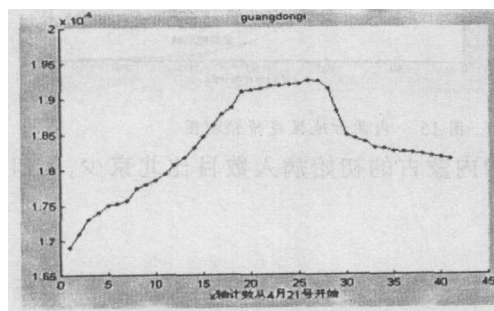


图12 广东地区原始I-T图

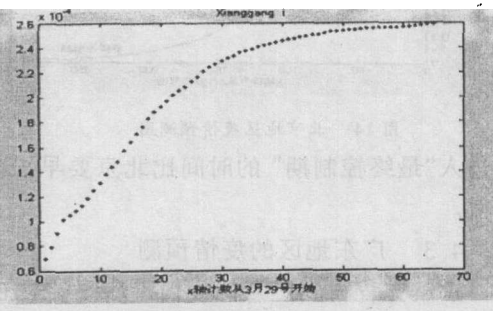


图13 香港地区原始I-T图

4 各地疫情预测

总体的分析与假设:

- 1) 根据现有的控制程度进行预测。如果随着疫情的缓解人们的警惕心理下降,政府的措施不再得力,将会导致疫情的传播死灰复燃,引起病人的比率又出现小的升高。
- 2) 根据现在的医疗水平进行预测。随着医学研究的深入和SARS治疗方案的完善,治愈率可能会有较大的提高,同时,预防措施也会更加有效(据有关资料,在不久的将来可能会研制出非典预防针),因此实际疫情很可能会比我们预测的提前结束。

4.1 北京地区的疫情预测

利用控后模型对北京地区的SARS发病人数进行预测,得到图14(病人比率图,其中的离散的点是实际住院病人占的比率)。可见:

- 1) 病人数目在4月24日到5月3号左右增长最快,即疫情达到“高峰期”。5月3号以后虽然病人很多,但增长趋势已趋于平缓。此时疫情进入“高平台期”,虽然表面看来病人数目高居不下,实际上疫情已经得到了控制;
- 2) 疫情大约在5月30号之后趋向缓解。从5月末到7月末病人下降速率最快。说明在这段时间内将有大量的病人出院,这段时期可称为“缓解期”。
- 3) 发病者的比率将在11月以后逐渐趋向于0,即进入疫情的“最终控制期”。需要指出的是,预测是依据现在的医疗水平进行的,实际的SARS最终控制期可能会有所提前。

4.2 内蒙古地区的疫情预测

利用已知数据求解控后模型,得到内蒙古的疫情预测图(图15,注意图中 y 轴值的数量级为 $\times 10^{-5}$)。其中离散的点是实际住院病人比率。可以看出:

- 1) 4月23号到5月8号左右是疫情的“高潮期”;
- 2) 发病人的比率在 $x = 22$ 即5月14~15号左右出现最大值,达到“高平台区”。
- 3) 疫情大约在5月30号逐渐趋向缓解。从5月末到7月末是疫情的“缓解期”;
- 4) 发病者的比率将在10月以后逐渐趋向于0,即进入疫情的“最终控制期”。易见内蒙

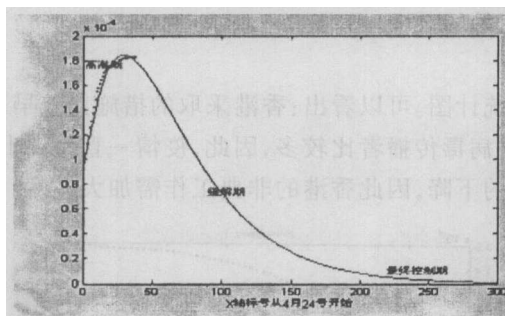


图14 北京地区疫情预测图

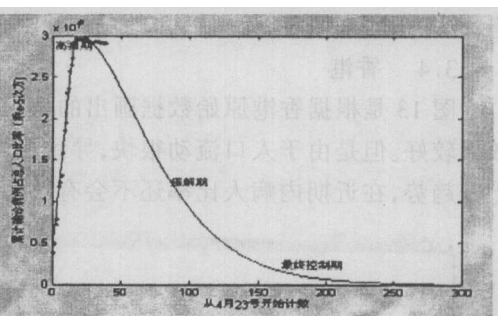


图15 内蒙古地区疫情预测图

古进入“最终控制期”的时间比北京要早,这是因为内蒙古的初始病人数目比北京少,人口分散。

4.3 广东地区的疫情预测

求解方程得到广东的疫情预测图(图16),其中离散的点是实际住院病人比率,可以看出:

- 1) 广东省病人数目的初值较大,且在5月初以前一直以一个较快的速率上升;
- 2) 病情在5月10号左右达到“高平台期”;
- 3) 5月中旬到明年2月为疫情的“缓解期”。随后病人比率趋近于0,疫情进入“完全控制期”。

比较三地疫情预测,广东地区疫情的“最终控制期”偏晚。这是因为:从数据来看,这段时期广东地区的病人治愈人数一直比较少,按现在的治愈率来预测疫情维持时间必然比实际的长。还需要指出的是,采集到的数据是从4月21号才开始的,而广东的疫情是从2002年11月开始的,因此数据的不完备可能会给预测带来较大的误差。

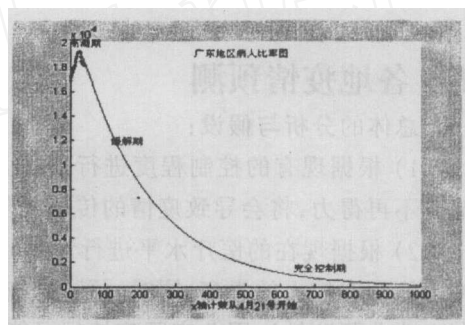


图16 香港地区疫情预测图

5 模型优缺点分析与改进方向(略)

本论文的模型因查到的统计数据不完善、对非典的传播规律了解不够,预测结果有较大偏差。相信随着人们对SARS的进一步认识,随着社会各界的深入研究,从数学角度看,模型将更加完善,预测结果将更准确,从医学角度看,SARS将有更好的治疗方案和防控措施,疫期将进一步缩短。

参考文献:

- [1] 中华人民共和国卫生部网站, <http://www.moh.gov.cn/>
- [2] 王树禾著. 常微分方程模型与混沌[M]. 中国科学技术大学出版社, 1999年2月第1版
- [3] 朱道元编著. 数学建模精品案例[M]. 东南大学出版社, 1999年8月第1版
- [4] 茆诗松编著. 贝叶斯统计[M]. 中国统计出版社, 1999年10月第1版
- [5] 吕本富. 北京SARS防御措施的效果分析[J]. <http://www.blogchina.com/new/display/8023.html>
2003;9:24

(下转62页)

当大家害怕得藏在家里,见到亲人朋友连握手都担心的时候,是因为我们对传染病的传播途径不了解,所以恐慌。

当政府说明已采取了许多措施,而大家看到医院每日收治病人仍然不断攀升的时候,是因为我们不了解传染病的发展规律,所以恐慌。

这一切的恐慌,都是因为我们不了解疫情而带来的。而要全面了解、控制疫情,除了相关的医学知识外,重要的问题之一是建立恰当的数学模型。

如果在疫情一开始,我们对传染病的常用数学模型有一定的认识,了解如何确定一种传染病的危害程度,何时进行控制,如何用科学的观点,详细的数据向民众解释,我们就不会因为盲目而恐慌。

如果我们了解基本传染病传播规律以及相应模型的决定因素,那么我们在生活中也可以通过控制这些因素以有效控制疾病,而不必恐慌。

如果我们知道传染病的发病曲线,我们也不会为新闻发布新的各种疫情数字而恐慌。

今天,在网上搜索一下“SARS 数学模型”,我们可以找到近千篇文章,十几种基本的传染病数学模型及其改进的模型,也许这些模型不一定很符合这次 SARS 传播的实际情况,也许我们对这些模型中的各种深奥的公式与定理不太了解。但是,模型得到的一些结论如:高峰期的到达时间,控制因素的效果等,至少可以让我们心安,不必盲目恐慌。而更精确的模型,以利于我们更好地制定相应的应急措施,有效的医疗保障系统等。

SARS 曾让我们惊慌失措,也让我们相信科学,认识与了解科学。拥有医学与数学等科学知识,我们应该有信心,流行病并不可怕。

(上接 52 页)

The Mathematical Modeling and Analysis of SARS

WANG Yi-feng, TIAN Yi, YANG Qian

Advisor: SHANG Shou-ting

(Harbin Institute of Technology, Harbin 150001)

Abstract: In this paper, two mathematical models for the spreading of SARS before and after the control respectively were established based on the data displayed on the Internet before June 2003 and the classic SEIR model for infectious disease, in which people are defined into several groups reasonably. Each parameters was solved by reasonable estimation, method of curve fitting and method of statistical theory. Then the primary analysis is laid on the after-control-model, solving the equations by Runge-Kutta method, analyzing the specific situations in Peking, Inner Mongolia, Guangdong and Hongkong where the SARS was rampant. Finally, the evaluation was made about the rationality and practicability of the models, suggesting the method for further revising of the model.

Keywords: differential equation; Runge-Kutta method; method of curve fitting; statistical theory