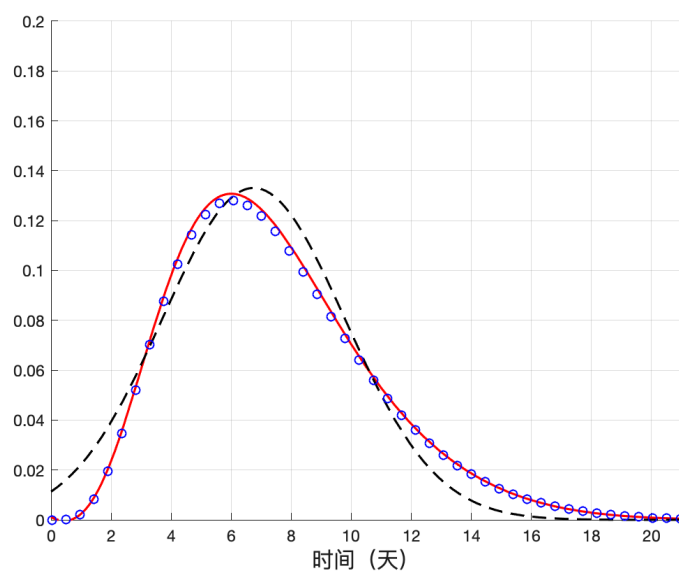




上海市现代应用数学重点实验室研究报告 Research Report Series of SKLCAM

(2020 年第二期)

关于新型冠状病毒肺炎一类基于 CCDC 统计 数据的随机时滞动力学模型



上海市现代应用数学重点实验室
Shanghai Key Laboratory of Contemporary Applied Mathematics

关于新型冠状病毒肺炎一类基于 CCDC 统计数据的随机时滞动力学模型

邵年¹, 陈瑜², 程晋^{2,3}, 陈文斌^{3*}

¹ 复旦大学数学科学学院, 上海市, 200433

² 上海财经大学数学学院, 上海市, 200433

³ 复旦大学数学科学学院, 上海市现代应用数学重点实验室, 上海市, 200433

2020 年 2 月 15 日

摘要

科学地预测新型冠状病毒肺炎疫情发展趋势对疫情防控至关重要. 本文对中国疾病预防控制中心 (CCDC) 发布的 [9] 的数据进行了分析, 给出了关于新型冠状病毒肺炎的一些统计模型: 传播链中连续病例的发病时间间隔分布、感染至发病的时间间隔分布和发病至住院的间隔时间三个分布, 并形成了感染至确诊的时间间隔分布表达. 结合 CCDC 统计数据和程晋团队的时滞动力学模型 (TDD-NCP 模型), 我们发展了新的随机时滞动力学模型 (TDD-NCP-CCDC 模型), 并给出了参数反演结果与疫情分析.

关键词 新型冠状病毒肺炎, 时滞动力学模型, 随机时滞动力学模型, 疫情预测.

Keywords Novel Coronavirus Pneumonia, time delay dynamic model, statistical time delay dynamic model, Outbreak prediction.

This paper analyzes the data released by the Centers for Disease Control and Prevention (CCDC) and obtain some statistical models of novel coronavirus pneumonia: the explicit expressions on the time from infection to illness onset, the time from illness onset in successive cases in a transmission chain, distributions of times from illness onset to hospital admission, and also formulate the distribution on the time from infection to hospital admission. Combining the time-delay model (TDD-NCP) developed by Jin Cheng's group with the statistical data from CCDC, we develop one statistical time-delay model (TDD-NCP-CCDC), and present some numerical results and outbreak predictions.

* 通讯作者, Email: wbchen@fudan.edu.cn

1 背景介绍

2019 年武汉发生的新型冠状病毒肺炎 (Novel Coronavirus Pneumonia, 简称 NCP) 疫情牵动了亿万中华儿女的心, 1 月 30 日一篇发表在著名医学期刊《新英格兰医学期刊》(The New England Journal of Medicine) 上的论文《新型冠状病毒感染的肺炎在中国武汉的早期传播动力学》[9] 引发了强烈的反响和争议 [14], 也引起了我们的关注. 我们回顾一下医学界几篇重要的研究报告:

- 1 月 24 日《柳叶刀》刊登的《武汉市 2019 年新型冠状病毒感染患者的临床特征》[6], 是本次新冠肺炎疫情期间最早披露的临床特征数据. 研究将疫情早期 (截至 1 月 2 日) 首批患病的 41 名患者作为研究对象, 将发病症状与死亡率量化并与其它传染病对比. 《科学》(Science) 杂志随后 (1 月 26 日) 刊文认为, 华南海鲜市场或许并非新冠病毒的唯一发源地.
- 1 月 29 日同样在《柳叶刀》上发表的论文《武汉市 2019 例新型冠状病毒肺炎 99 例流行病学和临床特征》[2] 将临床病例拓展至 99 例 (截至 1 月 20 日, 样本选自武汉金银潭医院).
- 1 月 30 日由《新英格兰医学期刊》刊登的《新型冠状病毒感染的肺炎在中国武汉的早期传播动力学》[9], 披露了迄今为止最完整的临床诊治数据, 共计 425 份确诊病例 (确诊时间截至 1 月 22 日). 相比此前两份研究, 这篇论文最重要的结论在于: 2019 年 12 月中旬以来, 新冠病毒已开始人际传播. 同时他们测算认为, 现有患者平均将病毒传给了另外 2.2 人, 估计平均潜伏期为 5.2 天.

钟南山的团队 2 月 6 日在 medRxiv 公开共享平台上传了一篇论文《中国 2019 年新型冠状病毒感染的临床特征》(Clinical characteristics of 2019 novel coronavirus infection in China) [5], 论文通过对截至 1 月 29 日的 1099 例确诊患者数据的进行回顾性研究, 发现新冠病毒潜伏期中位数缩短为 3 天, 但最长可达 24 天, 致死率相对较低.

复旦大学和上海财经大学程晋团队在分析传染病传播的数学模型如 SIR 模型 [1, 10, 11], SEIR 模型 [10] 和 SEIJR 模型 [8] 等模型的基础上, 考虑了潜伏期对于传播的时间滞后效应和数据滞后效应, 建立了一种新的时滞动力学模型 (TDD-NCP)[4, 13, 7, 3]. 西安交通大学马知恩 [10] 等的著作《传染病动力学的数学建模与研究》第二章中提出了对淋病和肺结核等疾病的带有时滞的数学建模. 但是在研究 SARS 时, 参考 CCDC 中心曾光等专家的工作 (中国疾病预防控制中心流行病学首席科学家曾光在他主持的“传染性非典型肺炎传播链分析”科研项目中给出了 SARS 传播五大特点, 指出“未发现潜伏期非典病人具有传染性”) 认为潜伏期是否有传染性难以确定 (见 [10] 第 97 页), 所以不考虑潜伏期的传染. 而程晋团队提出

的 TDD-NCP 模型明确地考虑了潜伏期的传染, 通过应用公开数据进行拟合计算与反演, 发现新的模型对实际数据的预测是比较理想.

中国疾控中心等单位对新型冠状病毒肺炎的研究成果 [9] 中已经有了一些统计数据, 特别是给出了传播链中连续病例的发病时间间隔、感染至发病的时间间隔、发病至住院的间隔时间等图像. CCDC 的这些数据对于传染病的动力学研究是很有价值的, 可以看作是一种新型冠状病毒肺炎的随机动力学模型 (类似于 SARS 研究中的工作 [12]). 而在 TDD-NCP 模型中, 核函数的选择是不确定的. 本文尝试把中国疾控中心得到的一些经验分布和随机动力行为数据 [9] 融入到 TDD-NCP 模型中, 建立有医学数据支持的 TDD-NCP 模型, 我们把这一类模型称为 TDD-NCP-CCDC 模型. 下面在第二节我们介绍了 TDD-NCP 模型和其缩减表示形式 TDD-NCP-r 模型. 在第三节我们对 [9] 的数据进行了处理, 得到了关于新型冠状病毒肺炎 (NCP) 的一些模型: 传播链中连续病例的发病时间间隔分布、感染至发病的时间间隔分布、和发病至住院的间隔时间三个分布的具体表达式和可能的候选, 并形成了感染至确诊的时间间隔分布. 在第四节结合 CCDC 统计数据和 TDD-NCP 模型, 我们发展了一系列的随机时滞动力学模型 (TDD-NCP-CCDC 模型), 然后再第五节给出了参数反演与疫情分析, 最后我们在结论部分做了一些讨论.

2 武汉新型冠状病毒时滞动力学模型 (TDD-NCP)

在通常的传染病模型中, 通常将人群分为易感染者 S , 潜伏者 E , 传染者 I , 确诊者 J 与康复者 R , 通过人员之间的转换关系, 从而可以建立起所谓的仓室模型 (SIR 和 SEIR 模型等)[10], 虽然现在已经有丰富的传染病模型, 并且有许多的深入的数学分析, 但是在实际操作与应用中, 模型的建立和变量的选择仍然是一个有挑战性的问题. 在 [4, 13] 中, 在复旦大学丁光宏团队关于 SARS 传播机制研究 [8] 的基础上, 程晋团队对武汉新型冠状病毒 (2019-nCoV, 现统称为 NCP) 引入了一类新的时滞动力学模型 (TDD-NCP), 这个模型研究感染者、确诊者和隔离者之间的关系. 下面先简单介绍一下这个模型, 这里约定记号如下:

- $I(t)$: t 时刻感染者的累计总数;
- $J(t)$: t 时刻确诊者的累计总数;
- $G(t)$: t 时刻已感染且未确诊, 但已经隔离的人群的实时总数.

考虑到潜伏期的影响, 在 TDD-NCP 模型中做了如下假设:

1. 感染者在出现明显症状前会经历 τ_1 天的潜伏期, 一旦出现症状, 感染者将寻求治疗, 从而转为确诊病例;

2. 由于政府干预控制措施, 某些感染者在潜伏期内尚未出现症状时已被隔离, 在被隔离了平均 τ_1' 天后出现症状成为确诊病例, 隔离不影响发病时间.

由此, 我们可以得到如下时滞动力学模型 (TDD-NCP model)[7]:

$$\frac{dI}{dt} = \beta(I(t) - J(t) - G(t)), \quad (1)$$

$$\frac{dJ}{dt} = \gamma \int_0^t h_1(t - \tau_1, t') \beta(I(t') - J(t') - G(t')) dt', \quad (2)$$

$$\frac{dG}{dt} = \ell(I(t) - J(t) - G(t)) - \ell \int_0^t h_2(t - \tau_1', t')(I(t') - J(t') - G(t')) dt'. \quad (3)$$

如果进一步考虑 $R(t)$: t 时刻康复者的累计总数, 假设确诊者在 τ_2 天后结束治疗, 有 κ 的比例被治愈, $1 - \kappa$ 的比例死亡. 可以考虑如下的关系:

$$\frac{dR}{dt} = \kappa \int_0^t h_3(t - \tau_1 - \tau_2, t') \beta(I(t') - J(t') - G(t')) dt'. \quad (4)$$

下面对这个模型的动机稍做解释:

1. β 定义为传染率, 代表每个传染者在单位时间内平均传染人数. 我们假设确诊者与隔离者均无传染性, 因此, 在时刻 t 能引起进一步传染的人数是 $I(t) - J(t) - G(t)$;
2. 系数 γ 是致病率参数. 累计确诊人数 $J(t)$ 均来自于潜伏期 τ_1 天内的感染人数;
3. 累计确诊人数 $J(t)$ 取决于潜伏期 τ_1 天内的感染人数.
4. 隔离者人数 $G(t)$ 的变化由以下两个因素决定:
 - 具有传染性的人群因政府防控措施而被隔离, 其中隔离率为 ℓ .
 - 隔离者在隔离了 τ_1' 天后被确诊入院. 时滞项表示隔离人群收到的历史数据影响.

在政府每日公布的疫情数据中, 可以得到累计确诊人数 $J(t)$ 和累计治愈人数 $R(t)$ 的信息.

5. $h_i(\hat{t}, t')$ ($i = 1, 2, 3$) 是延迟天数的概率分布, 满足归一化条件:

$$\int_0^t h_i(\hat{t}, t') dt' = 1, \quad t' \in (0, t), \quad i = 1, 2, 3$$

这个模型在实际的反演过程中, 产生了不同的版本 [13, 4, 7], 这里我们主要参考 [7] 版本.

2.1 TDD-NCP-r 模型

如果我们进一步对 TDD-NCP 模型进行分析, 我们发现实际上这个模型可以进一步简化: 引入变量

$$I_0(t) = I(t) - J(t) - G(t). \quad (5)$$

那么我们可以用一个方程来描述 I_0 :

$$\frac{dI_0(t)}{dt} = \beta I_0(t) - \gamma_\beta \int_0^t h_1(t - \tau_1, t') I_0(t') dt' - \ell \left(I_0(t) - \int_0^t h_2(t - \tau'_1, t') I_0(t') dt' \right). \quad (6)$$

这里 $\gamma_\beta = \gamma * \beta$. 一旦 I_0 确定, 我们可以进一步求解 J 、 G 和 R . 这个模型我们称为 TDD-NCP-r 模型. 特别地, 如果这里的延迟核函数取为某种 δ 函数, 我们可以得到如下的离散 d-TDD-NCP-r 模型:

$$I_0(t+1) = (1 + \beta - \ell) I_0(t) - \gamma_\beta I_0(t - \tau_1) + \ell I_0(t - \tau'_1), \quad (7)$$

$$J(t+1) = J(t) + \gamma I_0(t - \tau_1), \quad (8)$$

$$G(t+1) = G(t) + \ell(I_0(t) - I_0(t - \tau'_1)). \quad (9)$$

这里的 τ_1 和 τ'_1 可以与 TDD-NCP-r 模型不同, 通常可考虑传染病随机模型的最大似然估计得到的时刻. 这个模型可以以天为单位, 利用公布的数据来反演参数.

3 CCDC 的随机动力学模型

在中国疾病预防控制中心 (以下简称 CCDC) 的论文 [9] 中, 他们对数据进行了一些统计分析, 得到了一种随机动力学模型 [12]. 他们论文中提到, 他们检查了 10 例确诊病例的暴露数据, 估计平均潜伏期为 5.2 天 (95% 置信区间), 绝大部分在 12.5 天以下 (95% 置信区间), 截至 2020 年 1 月 4 日的流行曲线, 流行增长率为每天 0.10, 病例倍增时间为 7.4 天, 估计基本传染数 R_0 为 2.2 (95% 置信区间, 范围在 1.4 到 3.9), 平均一个病人能传染 2.2 个人. 据估计, 1 月 1 日前发病的 45 名患者从发病到首次就诊时间按平均为 5.8 天, 1 月 1 日至 11 日发病的 207 名患者平均为 4.6 天. 44 例在 1 月 1 日前发病的患者, 发病到入院的平均时间为 12.5 天, 1 月 1 日到 11 日发病的 189 例患者的平均时间为 9.1 天. CCDC 在论文 [9] 图 2 中给出了这些数据的图形. 他们认为从 2019 年 12 月中旬以来, 人与人之间的传播主要在密切接触者中发生, 并在随后的一个月逐步向外传播.

由于我们没有一手的数据资料, 我们从这个图的图像中抽取数据, 并进行了分析和拟合, 通过 Matlab 的拟合工具得到有关新型冠状病毒的从感染至发病的时间间隔、传播链中连续病例的发病时间间隔和发病至住院的间隔时间的分布信息如下:

- **传播链中连续病例的发病时间间隔:** 传播链中连续病例的发病时间间隔分布 $f_1(t)$ 为参数为 $(\mu, \sigma^2) = (5.171, 7.715), (5.213, 23.42), (1.357, 3.578)$ 的正态分布的线性组合, 见图 3. 具体表达为:

$$f_1(t) = 0.04186e^{-\left(\frac{t-5.717}{3.928}\right)^2} + 0.09397e^{-\left(\frac{t-5.213}{6.844}\right)^2} - 0.07311e^{-\left(\frac{t-1.357}{2.675}\right)^2}. \quad (10)$$

如果要使用一个正态分布拟合, 可得到如下表达式:

$$f_1(t) = 0.1331e^{-\left(\frac{t-6.747}{4.295}\right)^2}. \quad (11)$$

其参数为 $(\mu, \sigma^2) = (6.474, 9.224)$.

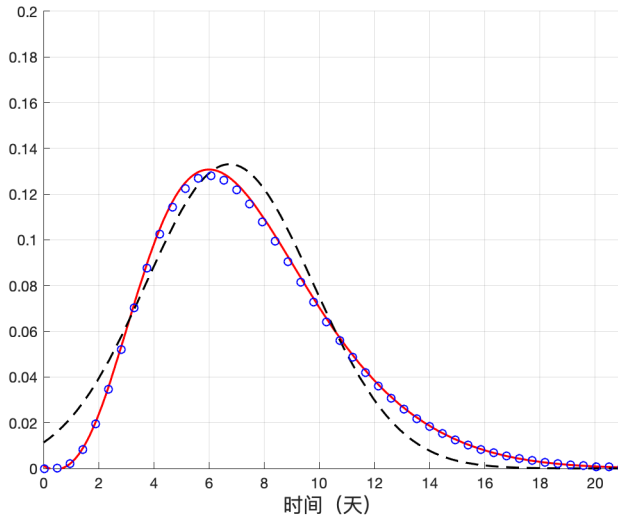


图 1: 传播链中连续病例的发病时间间隔: 实线表示用三个正态分布拟合, 虚线表示用一个正态分布拟合, 数据点用圈表示, 采样自 [9] 图 2 中子图 B.

- **感染至发病的时间间隔:** 感染至发病的时间间隔分布 $f_2(t)$ 为参数为 $(\mu, \sigma^2) = (1.417, 0.4525)$ 的对数正态分布, 见图 2. 具体表达为:

$$f_2(t) = \frac{0.5977}{t} e^{-1.105(\ln(t)-1.417)^2}. \quad (12)$$

- **发病至住院的间隔时间:** 发病至住院的间隔时间分布 $f_3(t)$ 为参数为 $(\lambda, k) = (179.9, 2.641)$ 的 Weibull 分布, 见图 3. 具体表达为:

$$f_3(t) = 0.005559t^{1.641}e^{-0.002105t^{2.641}}. \quad (13)$$

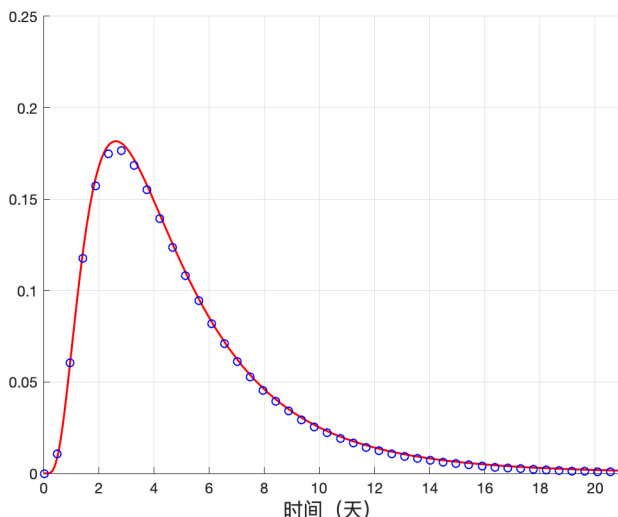


图 2: 感染至发病的时间间隔: 实线为拟合图, 数据点用圈表示, 采样自 [9] 图 2 中子图 A.

事实上, 如果我们把从感染到确诊分成感染到发病和发病到确诊, 则可得到如下感染到确诊时间分布.

感染至确诊的时间间隔: 感染至确诊 (或住院) 的时间间隔分布由两个参数为 $(\mu, \sigma^2) = (10.87, 14.46), (15.97, 23.42)$ 的正态分布组合而成, 其具体表达式为:

$$f_4(t) = f_2 * f_3(t) = 0.06244e^{-\left(\frac{t-10.87}{5.378}\right)^2} + 0.03322e^{-\left(\frac{t-15.97}{23.8}\right)^2}. \quad (14)$$

如果要使用一个正态分布拟合, 也可得到如下表达式:

$$f_4(t) = f_2 * f_3(t) = 0.08355e^{-\left(\frac{t-12.44}{6.785}\right)^2}. \quad (15)$$

其参数为 $(\mu, \sigma^2) = (12.44, 23.02)$.

我们也注意到钟南山团队 [5] 通过对截至 1 月 29 日的 1099 例确诊患者数据的回顾性研究, 发现新冠病毒中位数潜伏期缩短为 3 天, 但最长可达 24 天. 由于我们可以采样得到 CCDC 的数据, 所以我们基本上以 CCDC 的数据为参考模型. 因此我们建议用如下的随机动力学模型建模:

- 传播链中连续病例的发病时间间隔: 待定期望和方差的正态分布或多个正态分布的线性组合;
- 感染至发病的时间间隔: 待定期望和方差的的对数正态分布;
- 发病至住院的间隔时间: 待定比例和形状参数的 Weibull 分布;

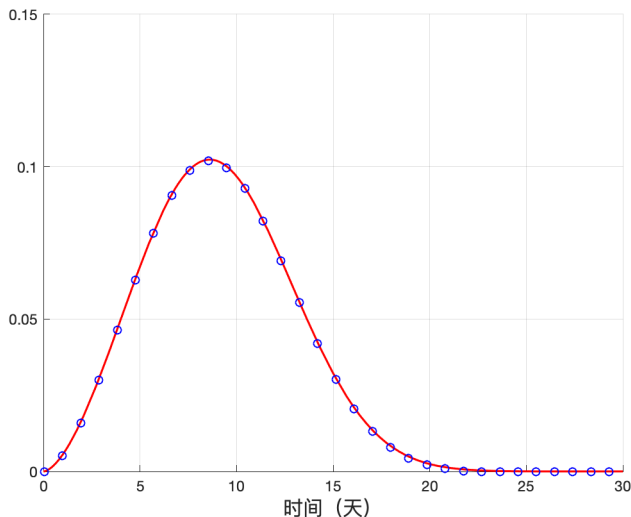


图 3: 发病至住院的间隔时间: 实线为拟合图, 数据点用圈表示, 采样自 [9] 图 2 中子图 D.

- 感染至确诊的时间间隔: 待定期望和方差的正态分布或多个正态分布的线性组合.

这些参数的估计可以通过一些实际病例得到. 在实际模拟中我们采用 CCDC 的数据得到的表达式.

4 TDD-NCP-CCDC 模型的建立

在本节中, 在本节我们结合 TDD-NCP 模型和中国疾控中心的随机动力学模型, 提出了下面四种随机动力学模型:

- 假设治病率参数 $\gamma = 1$, 也就是我们这里不引入治病率参数.
- 将从感染到确诊分成两个过程: 从感染到发病、从发病到确诊.

在下述模型中, 我们仍然定义 $I_0(t) = I(t) - J(t) - G(t)$.

假设在 t 时刻, 被感染的人数是 $\beta I_0(t)$, 考虑感染到发病的时间间隔分布 $f_2(t)$, 那么在 t 时刻发病的人数是 $f_2(t)$ 与 $\beta I_0(t)$ 的卷积 (没有隔离), 然后再考虑发病到住院的时间间隔分布 $f_3(t)$, 那么从感染到住院的人数为

$$\beta f_2(t) * f_3(t) * I_0(t) = \beta \int_{-\infty}^t f_3(t-s) \int_{-\infty}^s f_2(s-u) I_0(u) du ds. \quad (16)$$

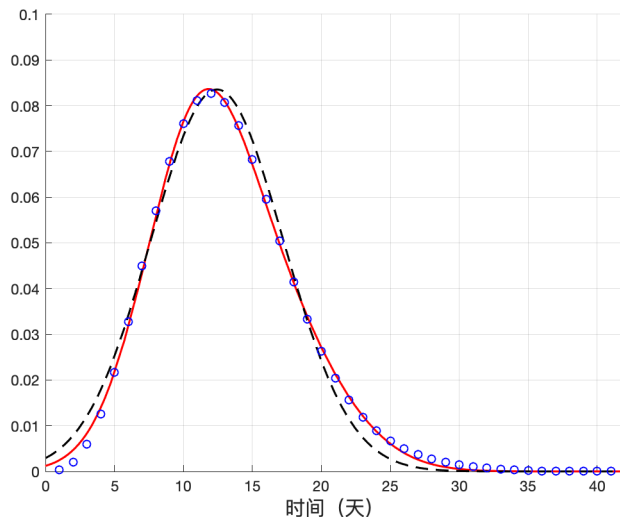


图 4: 感染至住院的间隔时间: 由 f_2 和 f_3 卷积得到, 这里实线表示用两个正态分布拟合, 虚线表示用一个正态分布拟合, 数据点用圈表示, 采样自 [9] 图 2 中子图 A 和子 D.

这里 $*$ 表示卷积. 也可以直接用感染到住院的时间间隔分布 $f_4(t)$ 来表示:

$$\beta f_4(t) * I_0(t) = \beta \int_{-\infty}^t f_4(t-s) I_0(s) ds. \quad (17)$$

对于隔离人数的分析可以类似进行. 同时我们对隔离发生的时刻做假设得到了如下两个模型: TDD-NCP-CCDC-1 模型和 TDD-NCP-CCDC-2 模型.

4.1 TDD-NCP-CCDC-1 模型 (隔离发生在患者被感染的时刻)

将感染至住院的间隔时间分布 f_4 代入 TDD-NCP 模型, 并假设

- 所有感染者具有相同的传染能力;
- 隔离发生在患者被感染的时刻.

我们有如下模型 TDD-NCP-CCDC-1 模型 (隔离发生在患者被感染的时刻):

$$\frac{dI}{dt} = \beta I_0(t), \quad (18)$$

$$\frac{dJ}{dt} = \beta \int_{-\infty}^t f_4(t-s) I_0(s) ds, \quad (19)$$

$$\frac{dG}{dt} = \ell I_0(t) - \ell \int_{-\infty}^t f_4(t-s) I_0(s) ds. \quad (20)$$

类似地, 我们可以引入如下的离散模型:

$$I(t+1) = I(t) + \beta I_0(t), \quad (21)$$

$$J(t+1) = J(t) + \beta \sum_{s \leq t} f_4(t-s) I_0(s), \quad (22)$$

$$G(t+1) = G(t) + \ell I_0(t) - \ell \sum_{s \leq t} f_4(t-s) I_0(s). \quad (23)$$

下面对这个模型做简单的解释:

- 第二个方程表示 s 时刻感染的人在 t 时刻被隔离.
- 第三个方程表示隔离者会因为不断有人被隔离而增多, 而隔离者被确诊也会导致隔离者减少.

4.2 TDD-NCP-CCDC-2 模型 (假设隔离发生在患者发病时)

与模型一类似, 模型二仍然假设所有感染者具有相同的传染能力, 但此时假设隔离发生在患者发病时, 我们有如下模型 TDD-NCP-CCDC-2 模型:

$$\frac{dI}{dt} = \beta I_0(t), \quad (24)$$

$$\frac{dJ}{dt} = \beta \int_{-\infty}^t f_4(t-s) I_0(s) ds, \quad (25)$$

$$\frac{dG}{dt} = \ell \int_{-\infty}^t f_2(t-s) I_0(s) ds - \ell \int_{-\infty}^t f_4(t-s) I_0(s) ds. \quad (26)$$

类似地, 我们可以引入如下的离散模型:

$$I(t+1) = I(t) + \beta I_0(t), \quad (27)$$

$$J(t+1) = J(t) + \beta \sum_{s \leq t} f_4(t-s) I_0(s), \quad (28)$$

$$G(t+1) = G(t) + \ell \sum_{s \leq t} f_2(t-s) I_0(s) - \ell \sum_{s \leq t} f_4(t-s) I_0(s). \quad (29)$$

最后一个方程通过 f_2 的描述了患者从被感染到发病的过程.

4.3 TDD-NCP-CCDC-3 模型 (考虑传播时滞和隔离发生在患者被感染的时刻)

如果我们进一步考虑传染的时滞性, 我们可以考虑利用传播链中连续病例的发病时间间隔分布 $f_1(t)$, 我们可以修正关于 $I(t)$ 的方程, 同时将所有感染者按照被感染的时间分类, 我们提出了 TDD-NCP-CCDC-3 模型 (考虑传播时滞和隔离发生在

患者被感染的时刻) 和 TDD-NCP-CCDC-4 模型 (考虑传播时滞和隔离发生在发病时刻). 在模型三中, 对隔离采取与模型一相同的假设, 即: 隔离发生在被感染的时刻. 我们有如下模型:

$$\frac{dI}{dt} = \beta \int_{\{s|I(s)>J(t)+G(t)\}} f_1(t-s) \frac{dI}{dt}(s) ds, \quad (30)$$

$$\frac{dJ}{dt} = \int_{-\infty}^t f_4(t-s) \frac{dI}{dt}(s) ds, \quad (31)$$

$$\frac{dG}{dt} = \ell I_0(t) - \ell \int_{-\infty}^t f_4(t-s) I_0(s) ds. \quad (32)$$

类似地, 我们可以引入如下的离散模型:

$$I(t+1) = I(t) + \beta \sum_s f_1(t-s) (\max\{I(s+1) - \max\{I(s), I(t) - I_0(t)\}, 0\}), \quad (33)$$

$$J(t+1) = J(t) + \sum_s f_4(t-s) (I(s+1) - I(s)), \quad (34)$$

$$G(t+1) = G(t) + \ell I_0(t) - \ell \sum_s f_4(t-s) I_0(s). \quad (35)$$

这里面第一个方程表示在 s 日被感染的到 t 日仍没有被隔离的感染者在 t 日的传染能力.

4.4 TDD-NCP-CCDC-4 模型 (考虑传播时滞和隔离发生在发病时刻)

模型四可以看做是模型二与模型三的结合, 即: 隔离发生在发病时且对感染者按被感染时间分类.

$$\frac{dI}{dt} = \beta \int_{\{s|I(s)>J(t)+G(t)\}} f_1(t-s) \frac{dI}{dt}(s) ds, \quad (36)$$

$$\frac{dJ}{dt} = \int_{-\infty}^t f_4(t-s) \frac{dI}{dt}(s) ds, \quad (37)$$

$$\frac{dG}{dt} = \ell \int_{-\infty}^t f_2(t-s) I_0(s) ds - \ell \int_{-\infty}^t f_4(t-s) I_0(s) ds. \quad (38)$$

类似地, 我们可以引入如下的离散模型:

$$I(t+1) = I(t) + \beta \sum_{s \leq t} f_1(t-s)(\max\{I(s+1) - \max\{I(s), I(t) - I_0(t)\}, 0\}), \quad (39)$$

$$J(t+1) = J(t) + \sum_{s \leq t} f_4(t-s)(I(s+1) - I(s)), \quad (40)$$

$$G(t+1) = G(t) + \ell \sum_{s \leq t} f_2(t-s)I_0(s) - \ell \sum_{s \leq t} f_4(t-s)I_0(s). \quad (41)$$

需要注意的是, 在实际推演过程中, 由于变量的实际含义, 往往还要加上非负性等限制条件, 这些条件就不写在这里了.

5 参数反演与疫情分析

在本节中, 我们对 TDD-NCP-CCDC-1 模型和 TDD-NCP-CCDC-2 模型进行参数反演和数值求解, 首先我们增加如下假设条件:

1. 设置初始感染人数 $I_0 = 4$ (在 [9] 中最早发现的新型病原体数目是 4), 隔离与确诊人数均为 0.
2. 在初始阶段隔离率 $\ell = 0$, 在 d 日起, 政府开始采取隔离措施, ℓ 从此日起为某待定参数并保持不变.
3. 由于医疗资源有限, 因此过去的确诊数可能不够准确. 在 2 月 13 日湖北省尤其是武汉市改良了核算确诊病例的方法, 大大提高了可靠性, 由此我们假设 2 月 13 日上报的累计数据是准确的.
4. 人口不发生迁移, 即: 没有人口的流入流出.

由于全国疫情的严重情况差异较大, 我们根据具体情况的不同, 分成三个区域处理:

1. 武汉地区是本次疫情的爆发地, 属于中部特大城市, 医疗资源较为丰富.
2. 湖北省除武汉地区外, 疫情也非常严重, 但医疗资源相对紧张.
3. 全国除湖北地区外, 疫情相对不严重, 且医疗资源充足.

参照万得宏观经济数据库上每日公开的 1 月 16 日至 2 月 11 日的累计确诊数据 (在仿真图中用圆圈表示), 选取前 k 日数据作为真实数据, 求解如下优化问题

$$\min \|J(\beta, \ell, d) - J_{\text{实际}}\|_2.$$

对每一个 k 均可得到每一组 (β, ℓ, d) , 带入模型可以得到一个 2 月 12 日的预测值, 选取预测结果与 2 月 12 日真实值最接近的 (β, ℓ, d) 作为模型参数可得如下结果:

1. TDD-NCP-CCDC-1 模型:
- 武汉: $[\beta, \ell, d] = [0.322, 0.229, 28]$.
 - 湖北（除武汉地区）: $[\beta, \ell, d] = [0.489, 0.533, 17]$.
 - 全国（除湖北地区）: $[\beta, \ell, d] = [0.698, 0.824, 12]$.

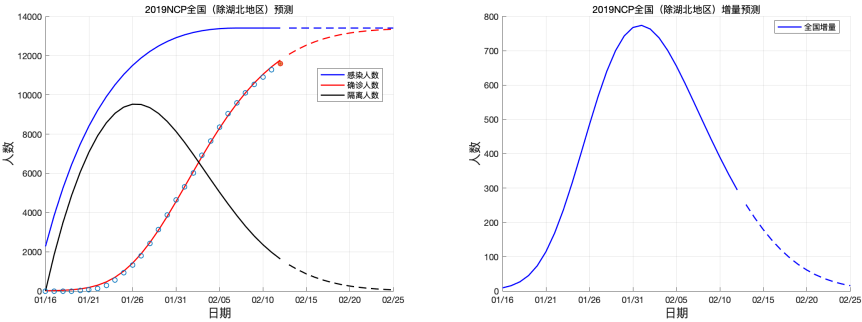


图 5: 全国（除湖北地区）预测, 累计确诊公布信息用圆表示, 实线表示拟合得到的结果, 虚线表示预测信息.

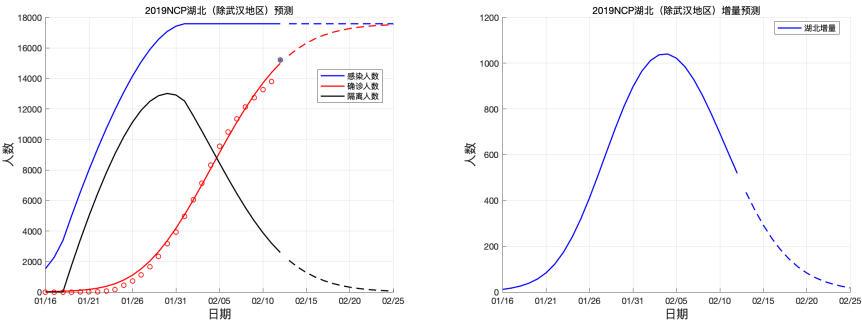


图 6: 湖北（除武汉地区）预测, 累计确诊公布信息用圆表示, 实线表示拟合得到的结果, 虚线表示预测信息.

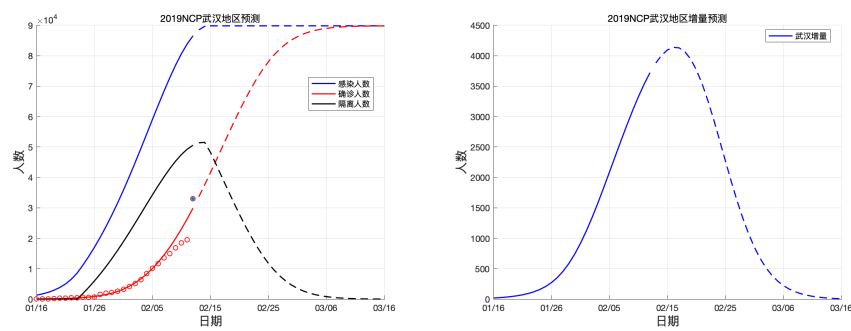


图 7: 武汉地区预测, 累计确诊公布信息用圆表示, 实线表示拟合得到的结果, 虚线表示预测信息.

2. TDD-NCP-CCDC-2 模型:

- 武汉: $[\beta, \ell, d] = [0.268, 3.86, 37]$.
- 湖北 (除武汉地区) : $[\beta, \ell, d] = [0.297, 9.22, 31]$.
- 全国 (除湖北地区) : $[\beta, \ell, d] = [0.273, 3.91, 29]$.

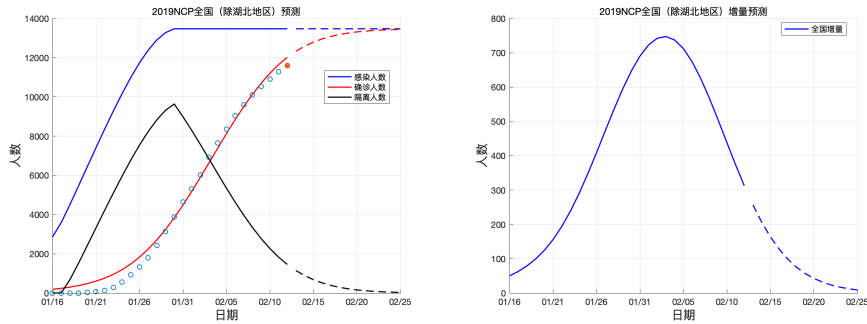


图 8: 全国 (除湖北地区) 预测, 累计确诊公布信息用圆表示, 实线表示拟合得到的结果, 虚线表示预测信息.

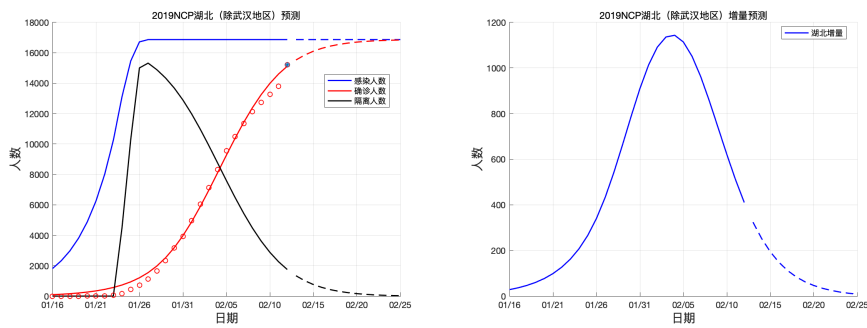


图 9: 湖北（除武汉地区）预测, 累计确诊公布信息用圆表示, 实线表示拟合得到的结果, 虚线表示预测信息.

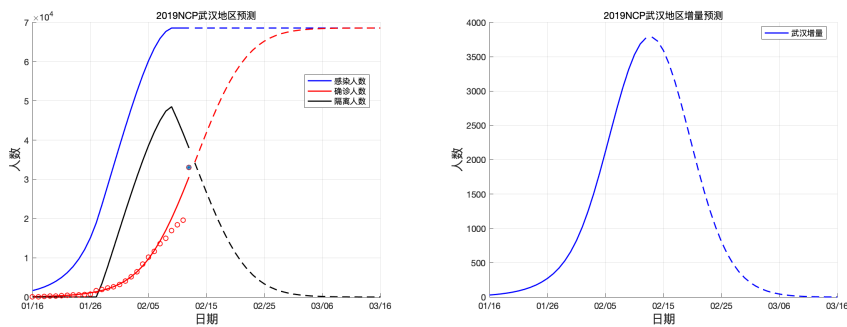


图 10: 武汉地区预测, 累计确诊公布信息用圆表示, 实线表示拟合得到的结果, 虚线表示预测信息.

我们对反演结果进行简要分析:

1. 除湖北地区外, 由于医疗资源充足, 且隔离措施迅速有效, 因此疫情很快将被控制.
2. 湖北其他地区尽管隔离率低, 但由于人口密度相对较低, 因此传染速度较慢, 一段时间后也将得到控制.
3. 武汉地区由于隔离措施采取时间较晚, 且人口基数大, 因此疫情将持续较长时间.

6 结论

在本文, 我们利用 CCDC 公布的一些统计信息和分布, 发展了一套随机时滞动力学模型, 这个想法可能对于促进中国传染病模型的研究是有一定意义的: 如何充分利用现有的数据来发展模型, 进而有意识地去采集一些特定的数据来进行模型预测, 这对于实际操作和政府决策是非常重要的. 在研究过程中, 上网和资料的查询也非常不方便, 许多的文献都没有涉及到, 这里需要抱歉. 新型冠状病毒肺炎有关的模型和仿真研究还在继续发展和研究中, 即使是如何来命名新型冠状病毒肺炎, 也没有统一, 我们这里采用中国自己的命名 NCP(在论文成文时, 2020 年 1 月 12 日世界卫生组织命名为 COVID-19). 这次新型冠状病毒肺炎对于数学界也提出了一些新问题和新的思路, 也有不少的工作. 在本文成文过程中, 通过 1 月 10 日数学会的公众号知道: 陕西师范大学唐三一教授联合陕西师范大学生物数学团队、西安交通大学生物数学团队以及加拿大吴建宏教授团队、北京建筑大学崔景安教授团队北京师范大学范猛教授团队、北京大学公共卫生学院生物统计系主任周晓华教授都对新型冠状病毒肺炎有了一些成果, 我们感觉由衷的高兴, 希望传染病模型不只是数学建模中的一个常规问题, 而成为对付传染病的一个有力的工具.

致谢: 我们这里衷心感谢团队各个成员提供的无私的帮助, 特别是江渝副教授、严阅博士和许伯熹博士以及复旦大学陆帅教授提供的帮助, 感谢 Mathworks 公司提供的帮助, 特别是曹新康先生, 感谢万得开放数据, 也非常感谢家人对我们工作的理解和支持. 本文写作过程中也参考了许多的媒体报道和资料, 这里不一一感谢. 程晋感谢自然科学基金项目 (11971121) 的支持, 陈文斌感谢自然科学基金项目 (11671098, 91630309) 和 111 项目 (B08018) 的支持, 也感谢上海财经大学计算科学与金融数据研究中心提供的帮助.

参考文献

- [1] B. Cantó, C. Coll, E. Sánchez. Estimation of parameters in a structured SIR model. *Advances in differences equations*, 2017,33:<https://link.springer.com/article/10.1186/s13662-017-1078-5>.
- [2] N Chen, M. Zhou, X. Dong, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020. doi: 10.1016/S0140-6736(20)30211-7.
- [3] Yu Chen, Jin Cheng, Yu Jiang and Keji Liu. A time delay dynamic system with external source for the local outbreak of 2019-nCoV, 2020[EB/OL], arXiv:2002.02590, 2020.

- [4] Y. Chen, J. Cheng, Y.Jiang and K. Liu. A time delay dynamical model for outbreak of 2019-nCov and the pramameter identification[EB/OL], arXiv:2002.00418,2020.
- [5] Wei-jie Guan, Zheng-yi Ni, Yu Hu, Wen-hua Liang, Chun-quan Ou, Jian-xing He, Lei Liu, Hong Shan, Chun-liang Lei, David SC Hui, Bin Du, Lan-juan Li, Guang Zeng, Kowk-Yung Yuen, Ru-chong Chen, Chun-li Tang, Tao Wang, Ping-yan Chen, Jie Xiang, Shi-yue Li, Jin-lin Wang, Zi-jing Liang, Yi-xiang Peng, Li Wei, Yong Liu, Ya-hua Hu, Peng Peng, Jian-ming Wang, Ji-yang Liu, Zhong Chen, Gang Li, Zhi-jian Zheng, Shao-qin Qiu, Jie Luo, Chang-jiang Ye, Shao-yong Zhu, Nan-shan Zhong. Clinical characteristics of 2019 novel coronavirus infection in China medRxiv 2020.02.06.20020974.
- [6] Huang C, Wang Y, Li X, et al. Clinical features of patients with 2019 novel coronavirus in Wuhan, China. Lancet. 2020; doi: 10.1016/S0140-6736(20)30183-5.
- [7] 江渝, 刘可^①, 陈瑜, 严阅, 许伯熹, 陈文斌, 陆帅, 徐翔, 程晋. 基于时滞动力系统新冠肺炎传播模型的若干预测分析, 上海市现代应用数学实验室,2020.
- [8] 刘畅, 丁光宏, 龚剑秋等. SARS 爆发预测和预警的数学模型研究 [J], 科学通报, 49(21):2245-2251,2004.
- [9] Li Qun, Guan Xuhua, Wu Peng, Wang Xiaoye, Zhou Lei, Tong Yeqing, Ren Ruiqi, Leung Kathy S.M., Lau Eric H.Y., Wong Jessica Y., Xing Xuesen, Xiang Nijuan, Wu Yang, Li Chao, Chen Qi, Li Dan, Liu Tian, Zhao Jing, Liu Man, Tu Wenxiao, Chen Chuding, Jin Lianmei, Yang Rui, Wang Qi, Zhou Suhua, Wang Rui, Liu Hui, Luo Yinbo, Liu Yuan, Shao Ge, Li Huan, Tao Zhongfa, Yang Yang, Deng Zhiqiang, Liu Boxi, Ma Zhitao, Zhang Yanping, Shi Guoqing, Lam Tommy T.Y., Wu Joseph T., Gao George F., Cowling Benjamin J., Yang Bo, Leung Gabriel M., Feng Zijian. (2020) Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N Engl J Med DOI: 10.1056/NEJMoa2001316.
- [10] 马知恩, 周义仓, 王稳地, 靳祯. 传染病动力学的数学建模与研究, 科学出版社, 2004.
- [11] Zhien Ma and Jia Li. Dynamical modeling and analysis of epidemics, World Scientific, 2009.
- [12] Yaolin Shi. Stochastic dynamic model of SARS spreading, Chinese Sci. Bull., 2003, 13:1287-1292.

- [13] 严闯, 陈瑜, 刘可[㊞], 罗心悦, 许伯熹, 江渝, 程晋. 基于一类时滞动力学系统对新型冠状病毒肺炎疫情的建模和预测 [J/OL], DOI: 10.1360/SSM-2020-0026, 2020.
- [14] 赵鹿鸣. 三篇涉及新冠肺炎临床诊治数据的热点论文详解: 病例从哪来? 作者有哪些人? 上观新闻, 2020 年 1 月 31 日, 网址: <https://www.jfdaily.com/news/detail?id=205271>.