

# 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

严 阅<sup>1</sup>, 陈 瑜<sup>1</sup>, 刘可伋<sup>1</sup>, 罗心悦<sup>1</sup>, 许伯熹<sup>1</sup>, 江 渝<sup>1</sup>, 程 晋<sup>2\*</sup>

<sup>1</sup>: 上海财经大学数学学院, 上海, 200433

<sup>2</sup>: 复旦数学科学学院, 上海市现代应用数学重点实验室, 上海, 200433

## 摘要

2019 年 12 月, 新型冠状病毒肺炎 (2019-nCoV) 疫情从武汉开始爆发, 几天内迅速传播到全国乃至海外. 科学有效掌控疫情发展对疫情防控至关重要. 本文基于全国各级卫健委每日公布的累计确诊数和治愈数, 提出了一类基于时滞动力学系统的传染病动力学模型. 在模型中引入时滞过程, 用来描述病毒潜伏期与治疗周期. 通过公布的疫情数据, 我们首先准确反演了模型的参数, 其次有效地模拟了目前疫情的发展, 并预测了疫情未来的趋势, 最后分析了各级政府防控措施手段的有效程度, 并发现在现有的高效防控措施下, 疫情将在近期好转.

**关键字:** 新型冠状病毒肺炎, 时滞动力学模型, 参数反演, 疫情预测

## 1 引言

2019 年末, 湖北省省会, 中部最大城市之一的武汉出现一系列新型非典型病毒性肺炎病例, 2019 年 12 月 12 日, 武汉市卫生健康委员会报告了 27 例病毒性肺炎病例 (包含 7 例重症), 此后该疾病伴随春运客流在短时间内迅速蔓延到其他省份乃至周边国, 各地疑似/确诊病例数急速增加. 2020 年 1 月 7 日, 中国确认了病原体为一种新型的冠状病毒. 1 月 12 日, 世界卫生组织 (WHO) 将该新型冠状病毒命名为 2019-nCoV, 并于 1 月 30 日将新型冠状病毒肺炎 (简称新冠肺炎) 疫情列为“国际关注的突发公共卫生事件”, 新冠肺炎疫情的爆发开始引起世界各国的高度重视. 截至 2020 年 2 月 2 日, 武汉市累计确诊病例

---

\*通讯作者, Email: jcheng@fudan.edu.cn.

4000 余例, 全国 31 个省 (自治区、直辖市) 和新疆生产建设兵团报告确诊病例 12000 余例, 死亡 300 余例. 此外, 在日本, 韩国, 新加坡, 美国, 加拿大, 德国, 法国, 英国, 西班牙等多个国家也发现了多起输入性确诊病例.

冠状病毒是具外套膜的正链单股 RNA 病毒, 广泛分布于哺乳动物和鸟类中. 冠状病毒偶尔会导致重大疫情爆发, 如 2003 年中国爆发的“严重急性呼吸综合征”(SARS), 2012 年与 2015 年分别在沙特阿拉伯与韩国爆发的“中东呼吸综合征”(MERS) 等都是由于感染冠状病毒引发的, 在过去 20 年中已累计发病超过 10000 例. 冠状病毒的每一次爆发都对疫情国产生了巨大的经济损失, 同时, 疫情的高死亡率也造成了不同程度的恐慌.

由于新冠肺炎疫情爆发于春节前夕, 大规模的春运人口流动与人群聚集加重了疫情. 疫情爆发后各地政府均采取了一系列前所未有的非药物的干预措施. 例如, 各地相继启动重大突发公共卫生事件 I 级响应, 对疑似/确诊患者进行搜寻并隔离观察或者治疗, 疫情最严重的湖北省实施了严格的出行限制, 湖北乃至全国多个城市关闭了部分乃至全部公共交通, 春节假期被延长等等. 这些干预措施从一定程度上减少了人群的接触, 有助于隔断病毒的疯狂传播, 降低病毒传染率, 同时经过科学的医学治疗, 经过一段时间一定能够有效缓解疫情并最终终结疫情. 但是, 长期的管控给人民群众日常生活带来了诸多不便, 春节后工厂不能按时开工也给经济带来了很大的影响. 因此及时准确衡量干预措施的有效性, 预测措施需要维持的时间对于政府决策具有重要意义. 这些可以通过建立模拟疫情传播行为的数学模型来实现. 通过对相关数学模型的研究, 通过历史数据反演得到的相关参数可以从多方面衡量疫情现状, 从模型数值计算模拟也能预测疫情的发展趋势, 这些都对提出更好更科学的疫情防控方案提供有效参考价值.

由于疫情爆发突然, 直至 2020 年 1 月 23 日后, 全国和各地的卫健委公布的疫情数据才日趋准确, 但至今各地准确的数据还不是很多, 因此新兴的大数据分析手段此时并不实用, 需要利用传统的传染病动力学模型来分析和预测疫情的发展. 描述传染病传播的经典数学模型有 SIR 模型 [1, 3], SEIR 模型 [3] 和 SEIJR 模型 [2] 等, 其主要思想在于将人群分为易感者 S, 潜伏者 E, 传染者 I, 确诊者 J 与康复者 R 等群体, 并通过某一群体转移至另一群体的传播学机制来建立常微分方程组, 从而揭示疫情传播的规律. 这些模型 (及其变形) 被用来研究各类传染病如麻疹, 天花, 狂犬病, Ebola 病毒的传播, 以及人口动力学研究. 2003 年 SARS 疫情过后, 对于 SARS, MERS 等冠状病毒传播规律的数学模型研究逐渐丰富.

2003 年 SARS 疫情爆发后, 复旦大学李大潜院士召集复旦大学相关师生通过讨论班的形式对 SARS 疫情的传播进行了研究. 复旦大学刘畅和丁光宏教授等人利用讨论班的研究结果, 将 SEIJR 模型简化为线性的 SIJR 模型, 计算出模型中封闭系统的精确解, 得到累计病例数与时间的关系, 并通过该关系与累计确诊病例的实际数据进行拟合, 得到了传染率参数  $\beta$  的估计值, 重现了 SARS 的发展趋势; 同时, 模型中引入的隔离率参数  $\ell$  用来描述政府的防控力度, 可以有效分析防控措施对疫情控制成效. 该 SIJR 模型中引入的传染

率  $\beta$  和隔离率  $\ell$  对我们建立新的预测模型具有很大的启发性.

由于大多数传统模型不考虑潜伏期对于传播的时滞影响, 即使考虑了潜伏者的 SEIR 与 SEIRJ 模型也只是假设潜伏者 E 具有弱传染性, 无法刻画此次新型冠状病毒可在潜伏期传播的特性. 此外, 上述模型在反演模型参数时几乎没有考虑公布数据会因为确诊需要时间所带来的数据滞后性. 综上所述, 需要一类具有时滞效应的模型 [6] 来更恰当地描述此次新冠肺炎疫情的传播. 与 SIJR 等传统模型相比, 我们将各类人群的累计总数作为变量 (而非实时人数), 这正好和全国各级卫健委每日公布的数据相匹配.

本文提出的基于时滞动力学系统的传染病动力学模型, 通过公开历史数据对模型参数进行了反演, 参数分析显示了各级政府防控措施的有效程度. 基于这些参数, 我们很好地模拟了目前疫情的发展, 并准确预测了疫情未来的趋势. 模拟结果显示, 在现有防控力度不放松的条件下, 疫情能在较短时间内得到控制并逐渐结束.

本文的其余部分安排如下. 第2节对新模型进行了介绍. 第3节给出了模型中参数反演与疫情预测的算法. 在第4节中, 利用官方发布的实际数据, 进行了若干数值模拟, 验证了带时滞的动力学模型的有效性. 最后, 我们在第5节中给出了总结和防控建议.

## 2 新的时滞动力学模型

在这一节中, 我们介绍新的时滞动力学模型. 研究对象是感染者, 确诊者, 隔离者, 康复者和死亡者, 我们使用如下记号来代表每个人群的人数:

- $I(t)$ :  $t$  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, 仍处于潜伏期, 但已隔离的人群的**实时总数**;
- $R(t)$ :  $t$  时刻康复者的**累计总数**;
- $D(t)$ :  $t$  时刻因病死亡者的**累计总数**.

考虑到潜伏期的影响, 我们假设:

1. 感染者在出现明显症状前会经历  $\tau_1$  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已被隔离, 在被隔离了平均  $\tau_1'$  天后出现症状成为确诊病例.

基于上述假设, 无论确诊者就医前是否被隔离, 在  $t$  时刻的确诊者累计总数  $J(t)$  均由  $t - \tau_1$  时刻的感染者构成. 此外, 我们进一步假设确诊者都在  $\tau_2$  天后结束治疗, 他们中有  $\kappa$  的比例被治愈,  $1 - \kappa$  的比例病重不治.

以上发病至结束治疗的过程可总结为图2.1.

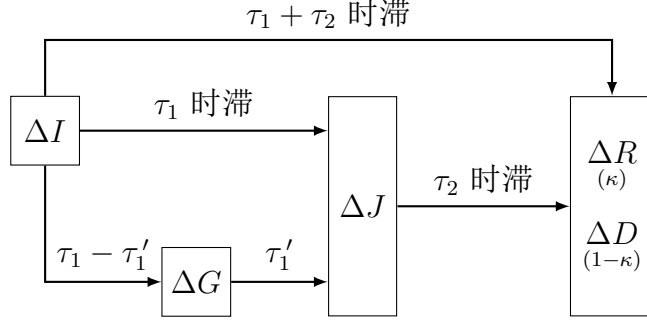


图 2.1: 模型示意图.

同时我们得到了如下时滞动力学模型:

$$\frac{dI}{dt} = \beta(I(t) - J(t) - G(t)), \quad (2.1)$$

$$\frac{dJ}{dt} = \gamma \int_0^t h_1(t - \tau_1, t') \beta(I(t') - J(t') - G(t')) dt', \quad (2.2)$$

$$\frac{dG}{dt} = \ell(I(t) - J(t) - G(t)) - \int_0^t h_2(t - \tau_1', t') G(t') dt', \quad (2.3)$$

$$\frac{dR}{dt} = \kappa \int_0^t h_3(t - \tau_1 - \tau_2, t') \beta(I(t') - J(t') - G(t')) dt'. \quad (2.4)$$

### 模型解释:

1.  $\beta$  被定义为传染率, 它代表每个感染者在单位时间内的平均传染人数. 在就医和隔离期间, 我们认为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  $I(t) - G(t) - J(t)$ .

2. 系数  $\gamma$  是致病率参数. 累计确诊人数  $J(t)$  均来自于潜伏期  $\tau_1$  天内的感染人数.

3. 隔离者人数  $G(t)$  的变化由以下两个因素决定:

(a) 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  $\ell$ .

(b) 隔离者在隔离了  $\tau_1'$  天后被确诊入院. 时滞项  $\int_0^t h_2(t - \tau_1', t') G(t') dt'$  表示  $G(t)$  中新确诊的人群受  $G(t)$  的历史数据影响.

4. 一旦被感染, 需经历  $\tau_1$  天的潜伏期与  $\tau_2$  天的治疗期才能结束治疗.

这里需要指出的是, 在政府每日发布的疫情数据中, 可以获得累计确诊人数  $J(t)$  和累计治愈人数  $R(t)$  的信息, 而  $I(t)$  和  $G(t)$  通常无法获得. 因此实际计算时, 我们采用  $J(t)$  与  $R(t)$  进行模型参数的反演.

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

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

一般地, 我们假设  $h_1(\hat{t}, t')$  是一个正态分布  $h_1(\hat{t}, t') = c_1 e^{-c_2(\hat{t}-t')^2}$ , 其中  $c_1$  与  $c_2$  是常数. 也可取  $h_1$  是一个  $\delta$ -函数, 即  $h_1(\hat{t}, t') := \delta(\hat{t} - t')$ , 这意味着每个感染者都经历了相同的潜伏期和治疗期的延迟天数分布. 同理,  $h_2(\hat{t}, t')$  与  $h_3(\hat{t}, t')$  可取为  $h_2(\hat{t}, t') = c_3 e^{-c_4(\hat{t}-t')^2}$ ,  $h_3(\hat{t}, t') = c_5 e^{-c_6(\hat{t}-t')^2}$ , 其中  $c_3$  至  $c_6$  都为常数.

### 3 参数反演和疫情预测算法

已知参数  $\{\beta, \gamma, \kappa, \ell, \tau_1, \tau_1', \tau_2\}$  的值以及初始条件  $\{I(t_0), G(t_0), J(t_0), R(t_0)\}$ , 利用 Matlab® 程序 **dde23** 我们可以数值求解时滞动力学系统(2.1)-(2.4). 这里  $t_0$  为第一批感染者 (假设为 5 人) 出现的时刻, 并且当天没有被确诊/隔离/康复的人, 即  $I(t_0) = 5$ ,  $G(t_0) = J(t_0) = R(t_0) = 0$ . 我们进一步假设在  $T = t_0 + 15$  前政府没有采取隔离干预措施.

利用全国各级卫健委 2020 年 1 月 23 日至 2020 年 2 月 1 日公开的数据和疫情相关信息, 可以首先大致确定致病率  $\gamma$ , 平均潜伏期  $\tau_1$  与治疗期  $\tau_2$ , 以及隔离后至入院确诊时间段的平均时间  $\tau_1'$ . 根据官方报道, 此次新冠病毒传染上后致病率很高, 由此假设致病率为  $\gamma = 0.99$ . 但同时大部分地区患者都能及时就医并得到有效治疗, 根据公布的死亡案例, 实际死亡率并不高, 因此假设治愈率  $\kappa = 0.97$ . 具体参数设置参见表1.

$\gamma$	$\kappa$	$\tau_1$	$\tau_1'$	$\tau_2$
0.99	0.97	7	4	12

表 1: 参数值.

最后需要根据每天公布的每日累计确诊病例数  $J_{\text{数据}}$  与累计治愈人数  $R_{\text{数据}}$  确定剩下的两个参数: 传染率  $\beta$  和隔离率  $\ell$ . 这两个参数的反演就转化为以下最小二乘问题:

$$\min_{\beta, \ell} \|J(\beta, \ell) - J_{\text{数据}}\|_2. \quad (3.1)$$

参数反演和疫情预测的方法主要分为以下两个步骤

1. 采用 Levenberg-Marquard (L-M) 方法或 Markov chain Monte Carlo (MCMC, 马尔科夫链蒙特卡洛) 方法 [4, 5] 求解最小二乘问题(3.1), 反演得到参数最优解  $\beta^*$  与  $\ell^*$ .

2. 将参数  $\{\beta^*, \ell^*, \gamma, \kappa, \tau_1, \tau_1', \tau_2\}$  代入并数值求解时滞动力系统(2.1)-(2.4), 可以得到  $\{I(t), J(t), G(t), R(t)\}$  历史和未来的演变曲线.

## 4 模型实证与疫情分析

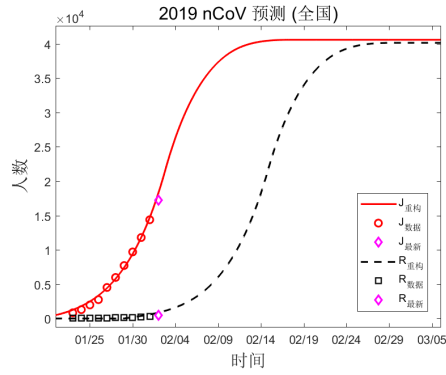
### 4.1 模型参数反演

利用前文提到的参数反演和疫情预测算法, 我们采用 2020 年 1 月 23 日至 2 月 1 日官方公布的累计确诊人数与累计治愈人数, 推断出传染率  $\beta$  与隔离率  $\ell$  的估计值. 我们选取全国 (除港澳台), 武汉市 (疫情源头), 上海市, 和江苏省作为代表. 相应地区的参数的估计值见表2.

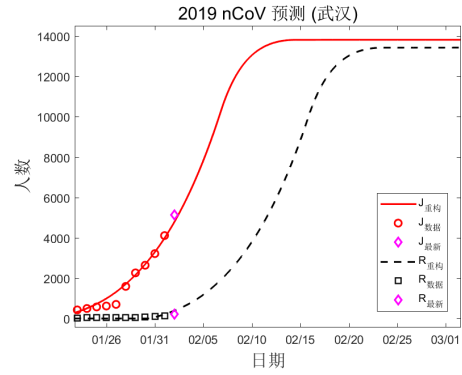
地区	传染率 $\beta$	隔离率 $\ell$
全国 (除港澳台)	0.2320	0.4202
武汉市	0.1957	0.5500
上海市	0.2113	0.5500
江苏省	0.2581	0.5500

表 2: 参数的估计值.

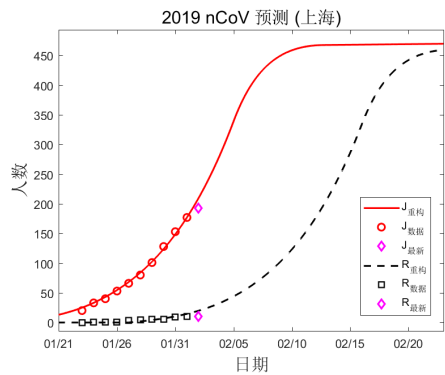
为了验证参数估计的准确性, 我们将表2中反演的参数值代入模型(2.1)-(2.4), 得到了这些地区的疫情发展趋势曲线 (见图4.1(a)-(d)). 通过使用 2020 年 1 月 23 日至 2 月 1 日的官方公布的累计确诊人数 (图中用红色圆圈标记) 与累计治愈人数 (图中用黑色方形标记) 来估计参数, 并采用 2 月 2 日的官方公布数据 (图中用深红菱形标记) 来验证模型. 从图中可以看到, 模型很好地模拟了 2020 年 1 月 23 日至 2 月 1 日的疫情发展, 并成功“预测”了 2 月 2 日的数据, 这说明不仅反演的参数是准确的, 而且新的时滞动力学模型是有效的.



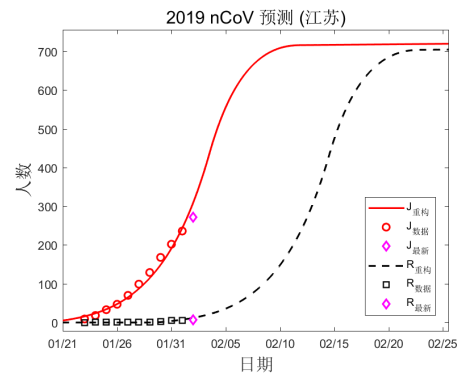
(a) 全国



(b) 武汉市



(c) 上海市



(d) 江苏省

图 4.1: 疫情发展趋势模拟. 横轴为日期, 纵轴为人数,  $J_{数据}$  与  $R_{数据}$  为官方公布的 2020 年 1 月 23 日至 2 月 1 日的累计感染人数与累计治愈人数,  $J_{最新}$  与  $R_{最新}$  为官方公布的 2 月 2 日的累计感染人数与累计治愈人数.  $J_{重构}$  与  $R_{重构}$  为累计感染人数与累计治愈人数的数值模拟结果.

## 4.2 疫情分析

根据图4.1(a)-(d) 显示的疫情演变曲线, 我们可以作出如下分析.

**疫情发展趋势:** 根据图4.1(a)-(d) 中各个地区累计确诊人数的演变曲线, 可以看出现阶段累计确诊人数正在增加, 但增速已经放缓, 特别在 2 月中下旬, 累计确诊人数将趋于不变, 从而疫情得到有效控制. 考虑到逐渐增加的治愈人数, 在院治疗的患者 (尚未治愈或死亡的确证患者) 将在达到高峰后逐渐降低 (参见图4.2).

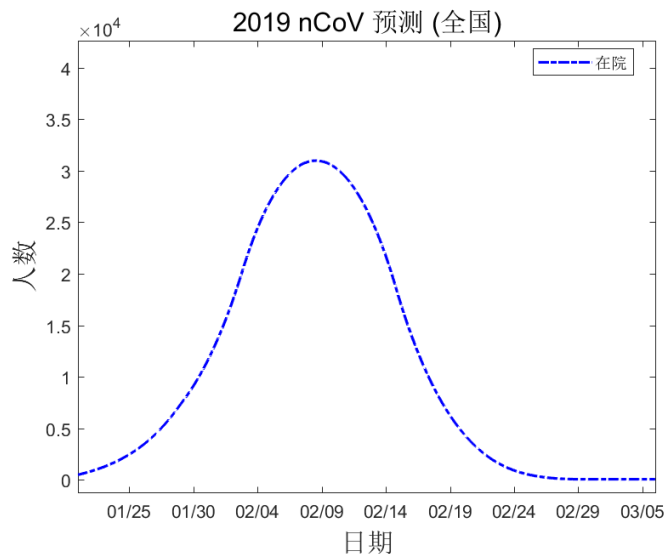


图 4.2: 全国在院治疗患者总人数预测.

此外, 我们将 2 月 3 日至 5 日的  $J(t)$  预测数据列在表3中作为参考. 本预测仅供参考, 由于各地管辖面积、人口基数等尺度大小不一, 预测的数值可能会产生过高或过低的估计.

地区	2 月 3 日	2 月 4 日	2 月 5 日
全国 (除港澳台)	22410	26310	29520
武汉市	5672	6688	7821
上海市	248	291	340
江苏省	389	482	556

表 3: 各地区累计确诊人数预测.

**参数分析:** 根据表2中反演的参数信息, 我们有如下的分析.

1. 传染率  $\beta$ . 处于疫情源头的武汉因为采取了“封城”措施, 其传染率较小. 全国因为平均效应, 传染率也相对较低. 上海市和江苏省距离武汉较远, 防控措施力度基本相似, 但由于它们管辖面积, 人口总量等因素不同, 所以传染率有所区别.

2. 隔离率  $\ell$ . 受春节后的返工潮影响, 全国范围内大规模人员流动还未停止, 所以隔离力度略低. 我们发现隔离率的大小对疫情发展趋势影响显著. 例如, 图4.3显示了当隔离率  $\ell$  分别为 0.55 与 0.25 时, 上海市的疫情预测结果. 我们观察到, 图4.3(a)的疫情发展趋势在 2 月 10 日前后将趋好, 但图4.3(b)的疫情发展趋势在 2 月 10 日后仍不容乐观, 而这个变化正是由于降低隔离率  $\ell$  而产生的. 考虑到 2 月 8 日后将出现返工潮, 政府管控措施将略有放松, 届时疫情在短期内可能会比较严峻.



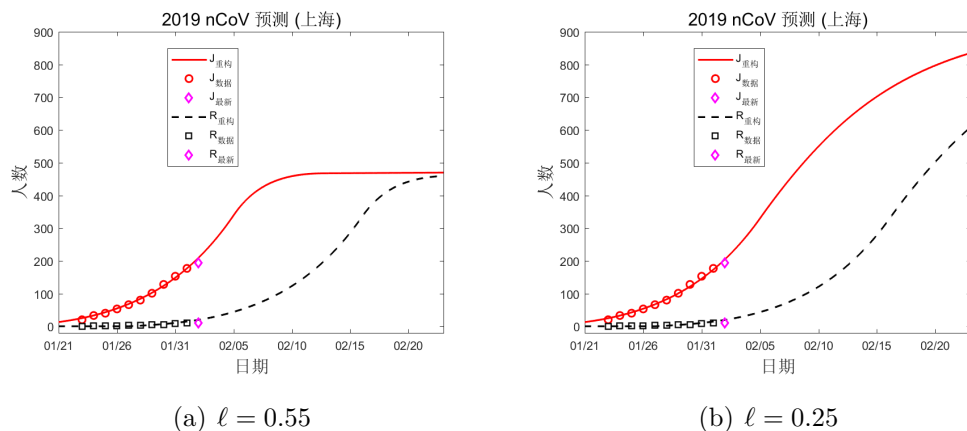


图 4.3: 不同隔离率下上海市的疫情发展趋势.

## 5 总结与疫情防控建议

针对本次新冠病毒肺炎疫情的发展, 我们提出了一类基于时滞动力学系统的传染病动力学模型. 通过该模型, 我们不仅反演出了各地的传染率和隔离率, 而且有效地预测了各地的疫情发展趋势. 数值模拟显示, 在现有防控力度不放松的情况下, 疫情能在较短时间内得到控制并逐步结束.

无论是传染率还是隔离率, 衡量的都是病毒的传染能力而非毒性, 因而阻断传播、加强隔离仍是防控的有效手段之一. 为了防止感染者持续快速增长, 加强对密切接触者的追踪、检疫、隔离等干预措施, 能够有效降低甚至阻隔病毒的进一步传播. 因此, 继续保持目前的防控手段不松懈, 是非常重要的. 由于春运返程潮即将来临, 我们建议乘客们加强自我保护, 注重个人卫生, 佩戴口罩, 正确洗手; 在到达目的地之后, 自觉在家隔离 14 天. 此外, 若治愈率能够进一步提高, 疫情的转变也能提前到来.

本文所采用的数据截至 2020 年 2 月 2 日, 但有报道显示新冠病毒变异很快, 新的传染特性不断显现, 本文考虑不周之处将在今后的研究中修正.

## 致谢

本文由上海财经大学数学学院“大数据融合反演建模”科研创新团队完成. 受国家自然科学基金委员会面上项目 (项目号: 11971121) 和上海科学技术委员会“上海青年科技启明星计划” (项目号: 19QA1403400) 资助. 感谢中泰证券边敬云女士, 上海财经大学数学学院硕士研究生唐梦园, 林御景, 卢俊俊和许天启搜集处理数据. 武汉加油!

## 参考文献

- [1] B. Cantó, C. Coll and E. Sánchez, Estimation of parameters in a structured SIR model[J], *Advances in Difference Equations*, **33**, 2017.
- [2] 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.
- [3] 马知恩, 周义仓, 传染病动力学的数学建模与研究 [M]. 北京: 科学出版社, 2004.
- [4] B. Kaltenbacher, A. Neubauer and O. Scherzer, Iterative regularization methods for nonlinear ill-posed problems[M], Berlin: Walter de Gruyter GmbH & Co. KG, 2008.
- [5] 刘逸侃, 徐定华, 程晋, 统计与计算反问题 (译), 芬兰 Erkki Somersalo, Jari Kaipio 著, 北京: 科学出版社, 2018.
- [6] Y. Chen, J. Cheng, Y. Jiang and K. Liu, A Time Delay Dynamical Model for Outbreak of 2019-nCoV and the Parameter Identification[EB/OL], arXiv:2002.00418, 2020.