



# 复杂人流网络下的 COVID-19 传播模型

傅家旗<sup>1</sup>, 刘敏<sup>1</sup>, 邓春燕<sup>1</sup>, 黄娟<sup>1</sup>, 江明珠<sup>2</sup>, 郭强<sup>1</sup>, 刘建国<sup>3\*</sup>

(1. 上海理工大学复杂系统科学研究中心 上海 杨浦区 200093; 2. 北京工商大学计算机与信息工程学院 北京 海淀区 100048;  
3. 上海财经大学金融科技研究院 上海 杨浦区 200433)

**【摘要】**考虑复杂人流网络的统计特征和 SI 传播模型, 该文基于实证数据构建了新型冠状病毒肺炎在湖北省和 4 个省传播和扩散的 P-SI 模型。基于实证数据的研究分析表明, P-SI 模型能够根据各省感染 COVID-19 情况拟合每天新增感染人数的变化趋势。此外, P-SI 模型能够根据 4 省 2020 年 1 月 23 日前的感染情况预测 2020 年 1 月 24 日后的每天新增感染人数。该模型有助于理解 COVID-19 的传播特性, 为阻断传播源, 遏制疫情扩散, 制定有效的防控措施提供决策依据。

**关键词** COVID-19; 流动人口; P-SI 模型; 传播模型

中图分类号 TP391 文献标志码 A doi:10.12178/1001-0548.2020109

## Spreading Model of the COVID-19 Based on the Complex Human Mobility

FU Jia-qi<sup>1</sup>, LIU Min<sup>1</sup>, DENG Chun-yan<sup>1</sup>, HUANG Juan<sup>1</sup>, JIANG Ming-zhu<sup>2</sup>, GUO Qiang<sup>1</sup>, and LIU Jian-guo<sup>3\*</sup>

(1. Complex Systems Science Research Center, University of Shanghai for Science and Technology Yangpu Shanghai 200093;  
2. School of Computer and Information Engineering, Beijing Technology and Business University Haidian Beijing 100048;  
3. Institute of Financial Technology Laboratory, Shanghai University of Finance and Economics Yangpu Shanghai 200433)

**Abstract** The statistical characteristics of the human mobility among different cities in the spreading of COVID-19 can be described by SI spreading model. P-SI models are presented to investigate how COVID-19 spreads or diffuses in Hubei and the other 4 provinces. Based on empirical data, some experiments are then conducted under the framework of the P-SI models. The experimental results demonstrate that the P-SI models can describe the number of daily new infected people caused by COVID-19 in Hubei and the other 4 provinces according to the empirical data. In addition, the P-SI models can also predict the number of daily new infected people in the other 4 provinces after 1/24 according to the empirical data before 1/23. The P-SI models are steps toward the understanding of COVID-19's spreading characteristics, which provides supports to isolate the source of COVID-19, to prevent and slow down the spreading of COVID-19, and to roll out effective measures for COVID-19's prevention and control.

**Key words** COVID-19; floating population; P-SI model; spreading model

利用实证数据对流行病传播过程进行建模分析有助于深入分析流行病传播的内在机制, 预测传播范围, 为进行有效的流行病防控提供重要依据<sup>[1-3]</sup>。由于缺乏对新型冠状病毒肺炎 (COVID-19) 传播机制的了解, 且没有针对新型冠状病毒 (SARS-CoV-2) 的疫苗, 自 COVID-19 确诊以来, 确诊规模不断扩大<sup>[4]</sup>。据国家卫生健康委员会官方网站的疫情通报数据显示, 截止至 2020 年 3 月 2 日 24 时, 累计报告确诊病例 80 151 例, 累计死亡病例 2 943 例。

世界卫生组织官网宣布, 将 COVID-19 疫情列为国际关注突发公共卫生事件。疫情的爆发不仅给各地医疗公共卫生机构<sup>[5-7]</sup>, 以及科学界带来了巨大挑战, 还使各行各业直接面临停工停产的严峻考验。尽管在国家和各地政府的努力下, 疫情蔓延的势头在一定程度上受到遏制<sup>[8]</sup>, 但限于对 COVID-19 传播机制的认知, 有关疫情可能持续的期限、近期出现拐点的可能性和可以采取的最有效防控措施等问题仍莫衷一是。因而, 在缺乏确切的治愈手段和疫

收稿日期: 2020-03-04; 修回日期: 2020-04-06; 网络首发日期: 2020-04-15  
基金项目: 国家自然科学基金 (71771152, 617773248); 国家社科重大项目 (18ZDA088)  
作者简介: 傅家旗 (1984-), 男, 博士, 主要从事网络科学、网络重构等方面的研究。  
通信作者: 刘建国, E-mail: liujg004@ustc.edu.cn

苗的情况下,对 COVID-19 的传播机制进行解构分析,能帮助准确研判疫情的传播途径和扩散范围,以便及时干预,科学防治,精准施策,最大限度降低疫情的影响。

引发 COVID-19 的 SARS-CoV-2 是一种新型病毒,各领域的专家学者在短时间内做了大量的研究工作,为深入分析病毒的特性和 COVID-19 的传播机制提供了第一手资料。除针对 COVID-19 病理学及发病机理的研究<sup>[9-10]</sup>外,科研工作者及时追踪每天疫情的进展及病例动态,进而对 COVID-19 疫情的发展态势进行预测和分析。研究工作主要集中在预估新增病例和死亡人数<sup>[11-13]</sup>、COVID-19 传染机制的研究<sup>[14-16]</sup>、人员流动对疫情传播的影响<sup>[17-19]</sup>、隔离和疑似病例对疫情的影响<sup>[20]</sup>、估计疫情在境外的发展趋势<sup>[21-22]</sup>、未来疫情防控工作中对医疗卫生服务的需求<sup>[23]</sup>等。

目前,学者们对 COVID-19 的研究已取得了进展,然而,对于病毒传播过程的模型设计还有待完善。本文试图从实证数据出发,结合 COVID-19 传播的特点,提出了 P-SI 模型。该模型考虑了省际流动人群对每天新增感染 COVID-19 病例的影响,能够刻画一个省内每天新增感染 COVID-19 病例的变化趋势。本文用 2019 年 12 月 31 日-2020 年 1 月 30 日的实证数据拟合出 P-SI 模型,该模型可以描述湖北省 3 个阶段的 COVID-19 传播情况。此外,本文还发现用另一个相同的 P-SI 模型可以描述其他 4 省的 COVID-19 传播过程。用实证数据检验拟合的模型能够描述各省每天新增感染 COVID-19 的人数。在此基础上,以 2020 年 1 月 23 日为分界点,将分界点前的实证数据视为训练集,拟合各省的 P-SI 模型。在修正了拟合模型的参数后,本文发现拟合模型能够对分界点后的每天新增病例人数作辅助预测,其预测出各省感染 COVID-19 的动态情况与实证数据相符。

## 1 P-SI 模型

本文基于实证数据对 COVID-19 的传播过程进行建模,并运用该模型对每天新增感染人数作预测。经典 SI 模型利用传播机制的微分方程模拟传染病的感染爆发过程,在地区总人数不变的情况下设定感染率模拟感染过程<sup>[24-26]</sup>。本文则在经典模型的基础上,结合流动人口的变化情况构建 P-SI 模型对 COVID-19 传播过程进行模拟预测。

根据 COVID-19 传染性较强的特点<sup>[27-28]</sup>,本文

将人群分为两类:易感人群(S)、感染人群(I)。易感人群为暴露在公共场合下的所有人,感染人群为 COVID-19 的确诊病例。鉴于 COVID-19 随流动人口传播,导致相关省份 COVID-19 病例在短时间内急剧增加,故在构建模型时加入流动人口因素。为构建一个简单而有效的模型,先作如下假设:

假设 1:其他 4 省在传播初期流动到湖北省的人群均视作易感人群(实证数据的信息与此假设一致)。在湖北省受到感染人群能流动到其他省,已流动到其他 4 省的感染人群不再进行省际流动。

假设 2:从湖北省流向其他 4 省的人员均视作当天抵达所属省。

### 1.1 基于湖北省实证数据拟合的 P-SI 模型

将 2020 年 12 月 31 日-1 月 23 日的湖北省视为一个开放系统,有客居人口及流动人口这两部分人群。这两部分人群最终未滞留在湖北省,均于 2020 年 1 月 23 日前回到所属省。

客居人口是指原本属于其他 4 省,但又长期客居在湖北省的人口。在湖北省(H)COVID-19 传播期间,定义在第  $t$  天客居人口中易感人群的数量为  $S_H(t)$ ,感染人群的数量为  $I_H(t)$ 。流动人口是指从外省经过湖北省,短期逗留后又离开湖北省的人口。类似地,定义第  $t$  天流动人口中由省  $i$  流动到湖北省的易感人群数量为  $S_{iH}(t)$ ,由湖北省流动到其他省的易感人群数量为  $S_{Ho}(t)$ ,由湖北省流动到其他省的感染人群的数量为  $I_{Ho}(t)$ 。

如果引入人口流动率  $P_{Ho}(t)$  和每天新增感染 COVID-19 人数  $\Delta I_H(t)$ ,便能构建如图 1 所示的传播模型。

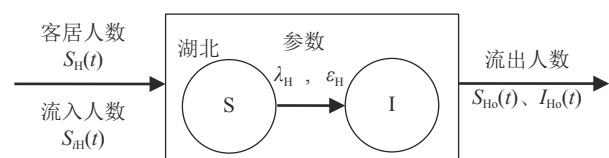


图 1 湖北省 P-SI 模型示意图

$$S_H(t) = \sum_{i=1}^M S_{iH}(t-1) + [1 - P_{Ho}(t)] [S_H(t-1) - \Delta I_H(t)] \quad (1)$$

$$I_H(t) = [1 - P_{Ho}(t)] I_H(t-1) + [1 - P_{Ho}(t-1)] \Delta I_H(t) \quad (2)$$

式中,人口流动率  $P_{Ho}(t)$  刻画了第  $t$  天从湖北流动到外省人口的比率; $M$  是除湖北外其他省的总数。相对应地,第  $t$  天离开湖北省的易感人数  $S_{Ho}(t)$  为  $P_{Ho}(t) [S_H(t-1) - \Delta I_H(t)]$ ,离开湖北省的感染人数

$I_{H_0}(t)$ 为 $P_{H_0}(t)I_H(t-1)+P_{H_0}(t)\Delta I_H(t)$ 。

每天新增感染 COVID-19 人数 $\Delta I_H(t)$ 主要由前一天的感染人数 $I_H(t-1)$ 决定。这是因为 COVID-19 主要通过人与人之间的接触传播<sup>[29]</sup>, COVID-19 的扩散范围主要由病毒携带者的生活圈大小决定。鉴于 COVID-19 有潜伏期, 本文认为在 COVID-19 传播早期, 某些受到感染的人并不会马上发病。因此, 每天新增感染 COVID-19 人数 $\Delta I_H(t)$ 在早期的增长速度应该比较缓慢。此外, 湖北省并非是一个封闭系统, 人员流动也较为频繁, 因而在短期内, 感染人数的增加并不会对湖北省这几个城市的易感人群造成显著影响。基于上述考虑, 本文建立每天新增感染 COVID-19 人数 $\Delta I_H(t)$ 与感染人数 $I_H(t-1)$ 的线性模型:

$$\Delta I_H(t) = \lambda_H I_H(t-1) + \varepsilon_H \quad (3)$$

式中, 参数  $\lambda_H$  是传染率, 表示由感染人群将病毒通过朋友圈传给易感人群, 并导致该省新增 COVID-19 的重要参数;  $\varepsilon_H$  则是一个误差常数。

### 1.2 基于 4 省实证数据拟合的 P-SI 模型

本文将浙江、安徽、陕西和广东这 4 省视作半开放系统, 一旦有人员从湖北省返回其中某个省, 便不再离开。而原本就在 4 省的人员也不离开所属省。因而, 这 4 省有常驻人口以及流入人口这两部分人群。常驻人口包含两部分: 一部分是易感人群  $S_i(t)$ , 另一部分则是感染人群  $I_i(t)$ 。同样地, 本文定义的流入人口也包含两部分: 一部分是易感人群  $S_{Hi}(t)$ , 另一部分则是感染人群  $I_{Hi}(t)$ 。不同省的流入人口总和应与湖北流动到各省的总人数一致:

$$\sum_{i=1}^M S_{Hi}(t) = S_{H_0}(t) \quad (4)$$

$$\sum_{i=1}^M I_{Hi}(t) = I_{H_0}(t) \quad (5)$$

如果引入人口流动率  $P_{iH}(t)$ 和每天新增感染 COVID-19 人数  $\Delta I_i(t)$ , 便能构建如图 2 所示的 COVID-19 模型。

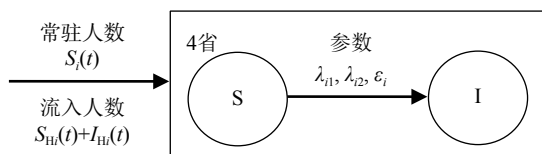


图 2 4 省 P-SI 模型示意图

$$S_i(t) = S_{Hi}(t-1) + [1 - P_{iH}(t-1)][S_i(t-1) - \Delta I_i(t)] \quad (6)$$

$$I_i(t) = I_{Hi}(t-1) + I_i(t-1) + \Delta I_i(t) \quad (7)$$

式中, 人口流动率  $P_{iH}(t)$ 刻画了 COVID-19 感染初期的第  $t$  天由从外省流动到湖北人口的比率。

每天新增感染 COVID-19 人数 $\Delta I_i(t)$ 主要由前一天具有传染性的感染 COVID-19 人数 $I_i^C(t-1)$ 和易感人数 $S_i(t-1)$ 决定。具有传染性的感染人群是指处于开始发病至被确诊的时间范围内的病例。定义这部分群体主要是考虑到湖北外地区的病例处于潜伏期和无症状期的感染能力较弱, 而当病人被确诊后, 往往就直接被医院隔离, 被隔离的病例便不具备传染性。根据掌握的病例情况, 计算出病例由开始发病日期至确诊日期间平均为 6.29 天 (约为 6 天), 与有关文献提供的病例传染的有效性时间相符<sup>[8]</sup>。感染人数 $I_i(t-1)$ 除去 $I_i^C(t-1)$ 后的剩余人数为 $I_i^B(t-1)$ , 这部分感染人群不具备传染性。

考虑到在 COVID-19 传播中后期, 具有传染性的感染人数 $I_i^C(t-1)$ 快速增加, 式 (3) 所描述的新增感染人数缓慢增加的模型已不再适用这一情形。一方面, 与感染人员接触的易感人群很快会受到传染; 另一方面, 易感人群间彼此接触的社交行为也为间接传染带来了可能性。此时, 由于省内的人员不再向外流动, 如果控制不当, COVID-19 会在有限的易感人群中迅速扩散。基于上述考虑, 本文建立每天新增感染 COVID-19 人数 $\Delta I_i(t)$ 与易感人数 $S_i(t-1)$ 和具有传染性的感染人数 $I_i^C(t-1)$ 的模型:

$$\Delta I_i(t) = \lambda_{i1} I_i^C(t-1) S_i(t-1) + \lambda_{i2} S_i^2(t-1) + \varepsilon_i \quad (8)$$

式中, 参数  $\lambda_{i1}$ 和 $\lambda_{i2}$ 均为传染率;  $\lambda_{i1}$ 表示该省由感染人群直接传染给易感人群的传染率;  $\lambda_{i2}$ 是指因易感人群之间彼此接触时可能发生传染的间接传染率;  $\varepsilon_i$  则是一个误差常数。

在进行模型拟合前, 把实证数据中的 573 例感染人员每天的状态进行分类统计。把湖北省及其他 4 个省每天的每 1 例人员分别归入到新增感染 COVID-19 人数 $\Delta I_i$ 、具有传染性的感染人数 $I_i^C$ 、不具有传染性的感染人数 $I_i^B$ 或易感人数 $S_i$ 。从而, 便可统计出每天分布在各省的新增感染 COVID-19 的总人数、已感染的总人数以及仍处于易感状态的总人数。基于上述归类, 将各省的数据分别代入式 (3) 和式 (8), 再用 Matlab 中的 nlinfit 函数求得对应的参数, 完成模型的拟合。

### 1.3 基于 4 省实证数据的 P-SI 辅助预测模型

基于构建的 P-SI 模型, 本文利用 4 省中早期的感染情况来预测后期的感染情况。尽管 COVID-19

的传播行为可以用式 (8) 来描述, 但由于 COVID-19 传播行为的特征在不同时期并不相同, 以中早期的 COVID-19 传播数据拟合的 P-SI 模型显然不适合用于预测后期的 COVID-19 传播行为。于是, 本文考虑对拟合的 P-SI 模型加以修正, 主要修正模型中的参数, 使修正后的 P-SI 模型与实证数据反映的真实情境相符。

本文通过变更方向和设置步长来修正参数。首先, 将 P-SI 模型中的参数列成 1 个 3 维向量的形式:  $\eta = (\varepsilon_i, \lambda_{i1}, \lambda_{i2})^T$ , 接着设置修正的方向和步长。引入 1 个方向矩阵  $\Phi$ :

$$\Phi = \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix} \quad (9)$$

方向矩阵  $\Phi$  中的任意一行均与参数向量  $\eta$  中的元素一一对应, 任取第  $i$  行元素形成向量  $\Phi_i$ , 向量  $\Phi_i$  可视为向量  $\eta$  的修正方向。如向量  $\eta$  能依次按照方向矩阵  $\Phi$  中的所有行向量  $\Phi_i$  变化, 相当于对向量  $\eta$  的所有可能方向做了遍历。有了确切的修正方向后, 还需设置修正步长。将步长的设置分为 2 项: 第 1 项是根据上述参数向量  $\eta$  中每个元素本身的取值大小来定义的 1 个 3 维向量  $\kappa = (\kappa_1, \kappa_2, \kappa_3)$ 。根据拟合的结果, 4 个省拟合得到的参数  $\varepsilon_i$  分布在区间  $(-3, 1)$  内, 参数  $\lambda_{i1}$  和  $\lambda_{i2}$  分布在区间  $(0.0008, 0.02)$  内。基于此, 定义向量元素  $\kappa_1=1$ ,  $\kappa_2$  和  $\kappa_3$  则在区间  $[0.001, 0.005]$  之间变化。另 1 项为调节常数  $\beta, \beta \in (0, 10)$ 。修正的步长便以  $\beta\kappa$  来定义。于是, 按照式 (10) 计算得到调整后的参数向量为:

$$\eta' = \eta + \beta\Phi_i^T \kappa \quad (10)$$

接着, 按照更新后的模型来预测某省后期的感染情况。根据假设 1, 已流动到其他省的感染人群不再进行省际流动, 这段时期内本省的总人数稳定, 变化的是每个人处于易感或感染状态。所以, 可以按照式 (6) 和式 (7) 将训练集中最近的感染人数  $I_i$  和易感人数  $S_i$  更新为模型所需的初始感染人数  $I_i$  和易感人数  $S_i$ 。然后, 将这两个初值和修正后的参数代入式 (11) 计算出每天新增感染人数为:

$$\Delta I_i(t) = \lambda'_{i1} I_i^C(t-1) S_i(t-1) + \lambda'_{i2} S_i^2(t-1) + \varepsilon'_i \quad (11)$$

式中, 参数  $\lambda'_{i1}$  和  $\lambda'_{i2}$  是向量  $\eta'$  中的元素, 均为修正后

的传染率;  $\varepsilon'_i$  也是向量  $\eta'$  中的元素, 是修正后的误差常数。按此思路, 可以依次推测出传播后期每天新增感染人数  $\Delta I_i$ 、感染人数  $I_i$  和易感人数  $S_i$ 。

## 2 实证分析

### 2.1 数据简介

本文选取了在浙江、安徽、广东、陕西 4 省发现的 573 例 COVID-19 感染人员进行研究。数据覆盖的时间从 2019 年 12 月 31 日-2020 年 1 月 30 日, 包含了感染人员进入湖北省的日期, 离开湖北省的日期、发病日期、就诊日期和确诊日期。其中, 有部分人员长期在湖北省居住, 并在 2020 年 1 月 30 日前返回上述 4 省, 累计 163 人。部分人员在返回上述 4 省期间, 曾在湖北省短期停留, 累计 252 人。还有部分人员没有去过湖北省, 一直在上述 4 省, 累计 158 人。本文对这 3 类人员做了统计, 如表 1 所示。本文在使用 P-SI 模型拟合湖北省和 4 省的 COVID-19 传播过程时, 用的是 2020 年 12 月 31 日-1 月 30 日的完整数据。

表 1 实验统计数据

日期	客居数	流动数	留原省数
2019.12.31-2020.01.23	161	242	72
2020.01.24-2020.01.30	2	10	86

实证数据中每个病例都有确切的发病日期, 但个别数据缺少确诊日期或仅登记了就诊日期, 这为准确定义每个病例发病后所具备传染性的有效时间带来了不确定性。为准确估算每个病例传染性的有效时间, 对所有登记了就诊日期和确诊日期的病例做了统计, 发现从就诊日期至确诊日期的平均值为 3.845 2 天 (约为 4 天), 从发病日期至确诊日期的平均值为 6.291 0 天 (约为 6 天)。本文分 2 种情况为缺少确诊日期的病例推算确诊日期, 对登记了就诊日期的病例按延后 4 天推算出确诊日期, 同时对没有登记就诊日期记录的病例按延后 6 天推算出确诊日期。从而可为每个病例定义发病日期至确诊日期为具备传染性的时间段。

以 2020 年 1 月 23 日湖北武汉“封城”作为分界点, 对各省感染的数据进行划分, 将 2019 年 12 月 31 日-2020 年 1 月 23 日的 4 省数据划分为训练集, 将 2020 年 1 月 24 日-1 月 30 日的 4 省数据划分为测试集。

### 2.2 辅助预测模型的评价标准

由于国家卫生健康委员会和各地方政府每天都

会公布有关疫情和感染情况的实时通报, 可以掌握实证数据 573 例感染人员的实时状态。在拟合出 P-SI 辅助预测模型后, 实际可以得到大量与 2019 年 12 月 31 日-2020 年 1 月 23 日期间感染情况相符的模型。为了从中选出最符合实际情况模型, 根据 2020 年 1 月 24 日与 1 月 25 日 2 天的实时通报中公布的新增感染 COVID-19 人数, 仅保留拟合模型中预测得到的 2020 年 1 月 24 日与 1 月 25 日这两天每天新增感染 COVID-19 人数与实际最接近的模型。于是, 本文便能得到有限个较为理想的 P-SI 辅助预测模型, 从而再用这些模型预测 2020 年 1 月 24 日-1 月 30 日 4 省感染 COVID-19 的总体情况。

### 2.3 模型拟合

本文根据掌握的 573 位感染人员的数据, 分别对湖北省及其他 4 个省的 COVID-19 传播模型进行拟合。由于湖北省武汉市于 2020 年 1 月 23 日当日 10:00 开始“封城”, 湖北省其他城市也采取了严格的措施, 大部分客居人员和流动人员大都在“封城”前回到自己所属省。本文便以 2019 年 12 月 31 日-2020 年 1 月 23 日期间在湖北长期客居和短期停留人员的感染情况数据来拟合湖北省的 COVID-19 传播模型。将统计信息代入式 (3), 便可求解得出参数  $\lambda_H=0.5261$  和参数  $\varepsilon_H=0.3499$ 。拟合结果如图 3 所示。

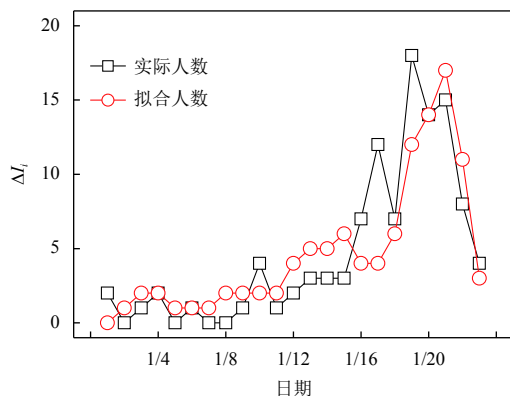


图3 湖北省每天新增感染 COVID-19 人数拟合图

图 3 中空心的方块标记的曲线描述了湖北省每天新增感染 COVID-19 人数, 空心圆圈标记的曲线则是根据拟合的传播模型计算得出每天新增感染 COVID-19 人数。从空心方块标记的曲线变化趋势看, 传播分为 3 个阶段: 第 1 阶段是 2019 年 12 月 31 日-2020 年 1 月 8 日, 为病毒传播的最初期。每天的新增感染人数在该阶段相对较少。第 2 阶段是 2020 年 1 月 9 日-1 月 15 日, 为病毒传播

的平稳期。每天的新增感染人数在该阶段有所上升, 并伴有微小波动。第 3 阶段是 2020 年 1 月 16 日-1 月 23 日, 为病毒传播的爆发期。每天的新增感染人数在该阶段急剧增加, 达到顶峰。后因所有客居人员或流动人员均在湖北省采取严厉措施前回到所属省, 所以湖北省的感染人数随人员的离开而降低了。

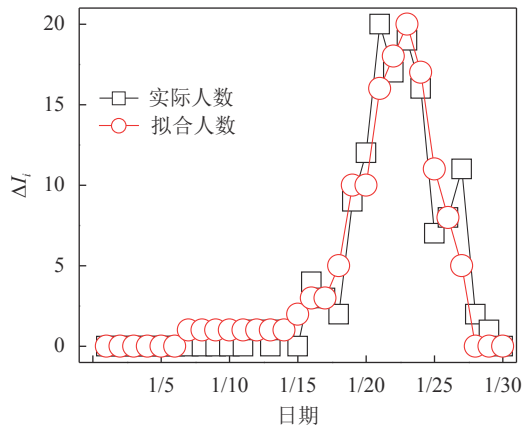
拟合曲线的变化趋势与实际曲线大体相同, 能大致地将每天新增感染 COVID-19 人数描绘出来, 在图中能够明显区分出 3 个阶段的变化。同时拟合曲线能将某些日期上新增感染人数的波动体现出来。如空心方块标记的曲线展示出了 2020 年 1 月 4 日、1 月 9 日、1 月 16 日人数的上升, 以及在 1 月 18 日先下降, 在 1 月 19 日又上升等变化趋势。空心圆圈标记的曲线也都较好地描绘了类似变化趋势, 只是在日期上可能会有 1~2 天的时间差。这主要是因为用于拟合的数据量有限, 会带来一些误差。

以 2019 年 12 月 31 日-2020 年 1 月 30 日间在湖北短期停留人员和没有去过湖北人员的感染情况数据来拟合 4 个省的 COVID-19 传播模型。本文将 4 组统计信息代入式 (8), 便可求解得出对应的 4 组参数。其中浙江省的参数  $\lambda_{i1}=0.0034$ ,  $\lambda_{i2}=0.0023$ , 参数  $\varepsilon_i=-1.6462$ ; 安徽省的参数  $\lambda_{i1}=0.0079$ ,  $\lambda_{i2}=0.0008$ , 参数  $\varepsilon_i=0.2757$ ; 陕西省的参数  $\lambda_{i1}=0.0102$ ,  $\lambda_{i2}=0.0083$ , 参数  $\varepsilon_i=-0.7261$ ; 广东省的参数  $\lambda_{i1}=0.0190$ ,  $\lambda_{i2}=0.0155$ , 参数  $\varepsilon_i=-2.5178$ , 结果如图 4 所示。

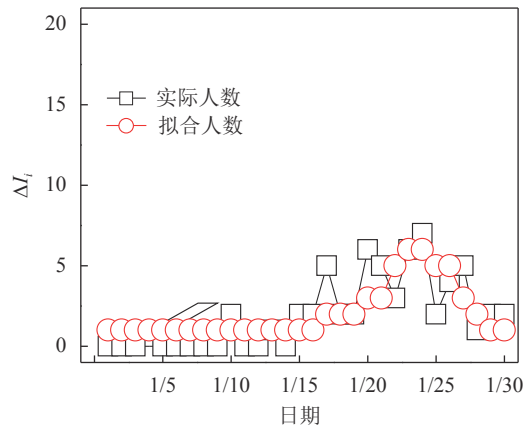
图 4a~4d 中空心的方块标记的曲线依次描述浙江、安徽、陕西、广东 4 个省每天新增 COVID-19 实际感染人数, 空心圆圈标记的曲线则是根据拟合的传播模型计算得出每天的新增感染人数。从空心方块标记的曲线的变化趋势看, 传播的最初期和平稳期内新增感染人数并没有显著变化, 在 2020 年 1 月 16-1 月 18 日开始进入 COVID-19 传播的爆发期。前期感染人数变化不显著的原因是当时感染源主要在湖北武汉, 由少量离开湖北到外省的流动人员携带 SARS-CoV-2 至所属省, 引发 COVID-19 初期传播。由于从湖北省到这 4 个省乘坐高铁, 只需 1 天即可抵达。在武汉“封城”前, 携带 SARS-CoV-2 的感染人员陆续返回所属省, 大量输入病例进入该省, COVID-19 也随之传播。因而, 4 省的爆发期与湖北省的非常接近。在 4 省的爆发期, 新增感染人数达到峰值以后曲线陡然下

降。一方面是因为本文仅追踪了有限的实证数据, 其中涉及的易感人群已大都受到了感染, 将来可能被感染的易感人群成为了极少数, 所以新增感染人

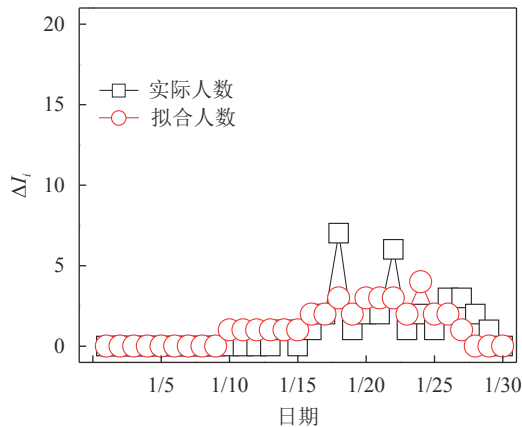
数也接近于 0。另一方面是因为各省均已加大了防控力度, 限制了 COVID-19 的传播渠道和速度, 新增感染人数也相应地得到了一定的控制。



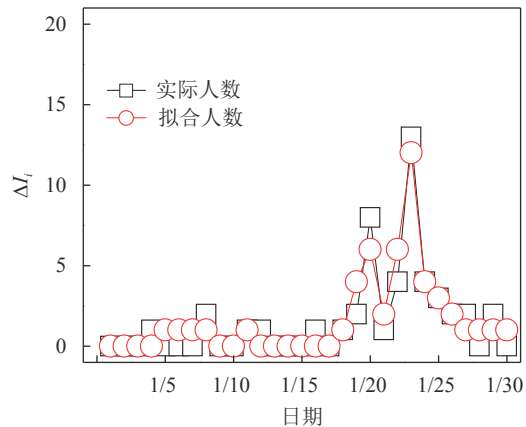
a. 浙江省每天新增感染 COVID-19 人数拟合图



b. 安徽省每天新增感染 COVID-19 人数拟合图



c. 陕西省每天新增感染 COVID-19 人数拟合图



d. 广东省每天新增感染 COVID-19 人数拟合图

图 4 4 个省每天新增感染 COVID-19 人数拟合图

空心圆圈标记的拟合曲线能够大致地刻画每天新增感染人数的变化趋势, 尤其刻画了最初期及平稳期新增感染人数保持平稳的趋势, 而在爆发期人数急剧攀升和人数陡然下滑的变化趋势也与实际情况大体相符。以浙江省为例, 可以观察到, 该省的实际新增感染人数在 2020 年 1 月 15 日前较为平稳, 在 1 月 16 日以后便迅速上升, 在 1 月 21 日到达峰值后, 在 22 日略有下降, 在 1 月 23 日再次攀升后便立即下降。相应地, 空心圆圈标记的拟合曲线同样在 2020 年 1 月 16 日以后迅速上升, 在 1 月 23 日到峰值后便立即下降。该拟合曲线较为平滑, 尽管没能刻画真实情况中的一些细微变化, 如: 局部峰值, 先降后升的变化, 或在日期上存在 1~2 天的偏差等情况, 但总体上能够反映该省每天新增感染 COVID-19 情况。其他 3 个省的拟合曲线总体上也比较符合实际情况。可见, 用同一个模型能够模拟 4 个省的 COVID-19 传播过程。

## 2.4 模型预测

按 1.3、2.1 和 2.2 节的思路将 4 个省的实证数据划分为训练集和测试集, 依次运用各省的训练集数据拟合出 P-SI 模型的相关参数。以 2020 年 1 月 23 日的感染人数  $I_i$  和易感人数  $S_i$  为初始信息, 将其代入式 (8) 推测出 1 月 24 日的新增感染人数  $\Delta I_i$ , 更新当天的感染人数  $I_i$  和易感人数  $S_i$ , 以此类推, 得到 1 月 25 日的每天新增感染人数  $\Delta I_i$ , 感染人数  $I_i$  和易感人数  $S_i$ 。

在产生了多组模型后, 对照 2020 年 1 月 24 日-1 月 25 日两天的新增感染 COVID-19 实际人数, 按每个省分别选取了预测结果最接近实际情况的 4 个模型, 并以之对 2020 年 1 月 24 日-1 月 30 日感染 COVID-19 的情况进行了预测。将预测结果展示在图 5 上, 可以从图 5 看到预测得到新增感染人数  $\Delta I_i$ 、感染人数  $I_i$ 、易感人数  $S_i$  与实际人数之间的差异。同时, 从上述 16 个模型得到预测结果最

接近每天新增感染 COVID-19 人数的 4 个模型的参数, 列于表 2。

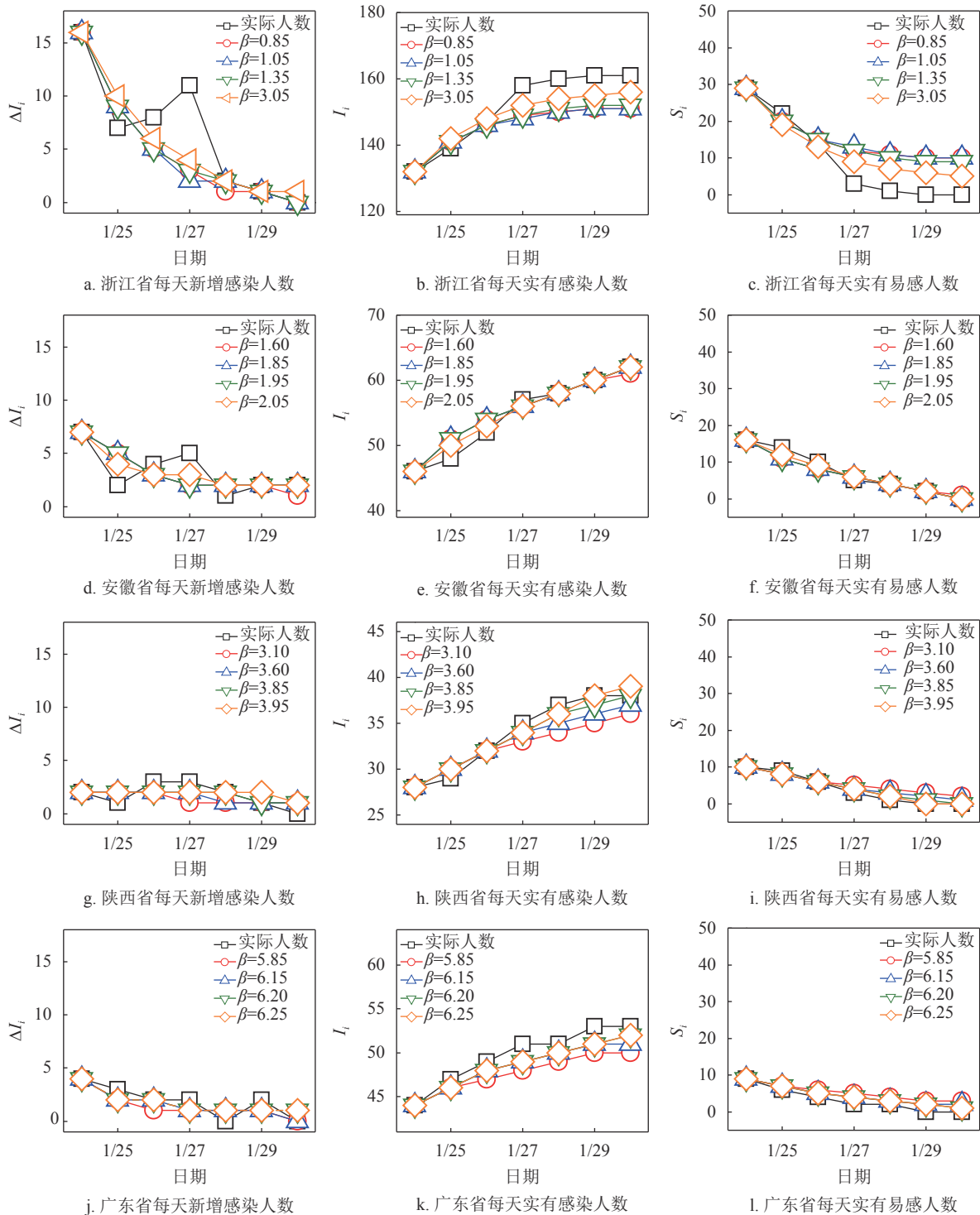


图 5 4 省 2020 年 1 月 24 日以后的预测结果趋势图

表 2 4 省的 P-SI 模型修正后的参数

参数/省份	浙江	安徽	陕西	广东
$\lambda_{i1}$	0.002 9	0.000 2	0.000 4	0.000 4
$\lambda_{i2}$	0.003 2	0.010 6	0.005 2	0.021 4
$\varepsilon_i$	-1.820 5	1.633 4	1.394 0	0.399 1

图 5a~5l 依次展示了模型对浙江、安徽、陕

西和广东 4 省每天新增感染人数  $\Delta I_i$ 、感染人数  $I_i$  和易感人数  $S_i$  的预测情况。以浙江省为例, 图中由空心方块标记的曲线描述了该省每天实际新增感染人数、感染人数和易感人数, 由空心圆圈、空心三角形、空心倒三角形和空心菱形标记的曲线则分别对应参数  $\beta$  取不同值时所得到的预测结果。根据

图 5a, 可以发现由空心方块标记的曲线描绘的每天实际新增感染人数  $\Delta I_i$  在 2020 年 1 月 25 日前和 1 月 28 日后单调下降, 而在 1 月 26 日和 1 月 27 日间有一个先增后降的过程。P-SI 模型预测的每天新增感染人数  $\Delta I_i$  是一条单调下降的曲线, 与实际情况较为接近, 但没有能预测出该省在 2020 年 1 月 26 日和 1 月 27 日间感染人数的波动。图 5b 和图 5c 则是根据预测到的每天新增感染人数  $\Delta I_i$  求得的感染人数  $I_i$  和易感人数  $S_i$  变化趋势。从感染人数  $I_i$  和易感人数  $S_i$  的变化趋势来看, 预测结果本来与实际情况略有偏差, 只因模型未能预测到 2020 年 1 月 26 日和 1 月 27 日新增感染人数的一个上升波动, 加大了预测的易感人数  $S_i$  以及感染人数  $I_i$  与实际人数的偏差量, 但总体的变化趋势较为一致。

根据 4 个省的预测趋势图, 可以看出尽管预测模型对每天新增感染人员的波动情况还不能准确刻画, 而没有掌握 2020 年 1 月 23 日流动人员入省的信息对感染人数  $I_i$  和易感人数  $S_i$  的预测存在一定影响, 如造成了广东省在 2020 年 1 月 24 日预测的新增感染人数与实际情况出现了些许偏差, 且这部分偏差致使后期对当地感染人数的预测偏差逐渐加大。但该辅助预测模型对每个省的感染情况的描述还是大致符合实际的。此外, 预测得到的各省的参数彼此间存在差异, 体现了 COVID-19 在各省传播扩散的速度和程度有所不同。

### 3 结束语

本文以 4 省 573 例感染 COVID-19 的病患信息研究了 COVID-19 在湖北省及 4 省的传播特性。基于 2019 年 12 月 31 日-2020 年 1 月 30 日的实证数据, 本文以湖北省作为研究区域构建了 P-SI 模型。该模型综合考虑了客居人口及流动人口对 COVID-19 传播的影响。本文还构建了另一个 P-SI 模型, 可以同时描述另外 4 省人员每天新增感染 COVID-19 的实际情形。本文分别用实证数据对湖北省及 4 省的模型进行拟合, 拟合的曲线能够反映各省每天新增感染人数, 且能描述各省 COVID-19 在不同时期的情况。在此基础上, 本文还利用 2019 年 12 月 31 日-2020 年 1 月 23 日的实证数据预测了 4 省 2020 年 1 月 24 日以后每天新增感染人数及对应省的总体感染情况。实验结果说明, P-SI 辅助预测模型能够预测出每天新增感染人数, 同时, 能够辅助研判各省的总体感染情况。

尽管本文构建的 P-SI 模型能够刻画 COVID-19 的传播过程, 但现有的模型还并不完善。该模型没有考虑病毒在人体内有潜伏期的因素。本文没有深入分析感染的流动人员对传播过程的影响。此外, 在实际情况下, 每天新增感染人数的变化趋势并不平滑, 该模型对不平滑的情况难以预测。本文的研究仅仅局限于一个小样本, 而各省市的真实病例情况更加复杂, 需要更加精细的模型来客观描述实际情形。这些问题将在以后的工作中再做专门的研究。

感谢上海理工大学李仁德老师, 浙江师范大学胡兆龙老师和苏黎世大学林坚洪博士的交流与讨论。

### 参 考 文 献

- [1] GIRE S K, GOBA A, ANDERSEN K G, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak[J]. *Science*, 2014, 345(6202): 1369-1372.
- [2] RILEY S, FRASER C, DONNELLY C A, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions[J]. *Science*, 2003, 300(5627): 1961-1966.
- [3] 胡兆龙, 刘建国, 任卓明. 基于节点度信息的自愿免疫模型研究[J]. *物理学报*, 2013, 63(21): 218901.  
HU Zhao-long, LIU Jian-guo, REN Zhuo-ming. Analysis of voluntary vaccination model based on the node degree information[J]. *Acta Physica Sinica*, 2013, 63(21): 218901.
- [4] LIU J, ZHENG X, TONG Q X, et al. Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV[J]. *Journal of Medical Virology*, 2020, 92(5): 491-494.
- [5] CORMAN V M, LANDT O, KAISER M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR[J]. *Eurosurveillance*, 2020, 25(3): 2000045.
- [6] WANG G, JIN X. The progress of 2019 novel coronavirus event in China[J]. *Journal of Medical Virology*, 2020, 92(5): 468-472.
- [7] 田怀玉. 2019-nCoV: 来自冠状病毒的新挑战[J]. *中华预防医学杂志*, 2020, 54(3): 233-236.  
TIAN Huai-yu. 2019-nCoV: New challenges from coronavirus[J]. *Chinese Journal of Preventive Medicine*, 2020, 54(3): 233-236.
- [8] CHEN D B, ZHOU T. Control efficacy on COVID-19[EB/OL]. [2020-02-29]. <https://arxiv.org/abs/2003.00305>.
- [9] WANG S, KANG B, MA J L, et al. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)[EB/OL]. [2020-03-30]. <https://doi.org/10.1101/2020.02.14.20023028>.
- [10] WU A P, PENG Y S, HUANG B Y, et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China[J]. *Cell Host &*



- Microbe, 2020, 27(3): 325-328.
- [11] YOU C, DENG Y H, HU W J, et al. Estimation of the time-varying reproduction number of COVID-19 outbreak in China[EB/OL]. [2020-02-19]. <http://dx.doi.org/10.2139/ssrn.3539694>.
- [12] NESTERUK I. Statistics based predictions of coronavirus 2019-nCoV spreading in mainland China[J]. *Innovative Biosystems and Bioengineering*, 2020, 4(1): 13-18.
- [13] 周涛, 刘权辉, 杨紫陌, 等. 新型冠状病毒肺炎基本再生数的初步预测[J]. *中国循证医学杂志*, 2020, 20(3): 359-364.  
ZHOU Tao, LIU Quan-hui, YANG Zi-mo, et al. Preliminary prediction of the basic reproduction number of the novel coronavirus 2019-nCoV[J]. *Chinese Journal of Evidence-Based Medicine*, 2020, 20(3): 359-364.
- [14] WU J T, LEUNG K, LEUNG G M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study[J]. *The Lancet*, 2020, 395(10225): 689-697.
- [15] KUCHARSKI A J, RUSSELL T W, DIAMOND C. Early dynamics of transmission and control of COVID-19: A mathematical modelling study[EB/OL]. [2020-03-11]. [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4).
- [16] 严闯, 陈瑜, 刘可伋, 等. 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]. *中国科学: 数学*, 2020, 50(3): 385-392.  
YAN Yue, CHEN Yu, LIU Ke-ji, et al. Modeling and prediction for the trend of outbreak of NCP based on a time-delay dynamic system[J]. *Scientia Sinica Mathematica*, 2020, 50(3): 385-392.
- [17] ZHAO H X, MAN S, WANG B, et al. Epidemic size of novel coronavirus-infected pneumonia in the epicenter Wuhan: using data of five-countries' evacuation action [EB/OL]. [2020-02-13]. <https://doi.org/10.1101/2020.02.12.20022285>.
- [18] DU Z-W, WANG L, CAUCHEMEZ S, et al. Risk for transportation of 2019 novel coronavirus disease from Wuhan to other cities in China[J]. *Emerging Infectious Diseases*, 2020, 26(5): 1049-1052.
- [19] 许小可, 文成, 张光耀, 等. 新冠肺炎爆发前期武汉外流人口的地理去向分布及影响[J]. *电子科技大学学报*, 2020, DOI: 10.12178/1001-0548.2020033. <http://kns.cnki.net/kcms/detail/51.1207.T.20200213.2002.002.html>  
XU Xiao-ke, WEN Cheng, ZHANG Guang-yao, et al. The geographical destination distribution and effect of outflow population of Wuhan when the outbreak of the 2019-nCoV pneumonia[J]. *Journal of University of Electronic Science and Technology of China*, 2020, DOI: 10.12178/1001-0548.2020033.
- [20] TANG B, XIA F, TANG S Y, et al. The evolution of quarantined and suspected cases determines the final trend of the 2019-nCoV epidemics based on multi-source data analyses[EB/OL]. [2020-02-13]. <http://dx.doi.org/10.2139/ssrn.3537099>.
- [21] BOLDOG P, TEKELI T, VIZI Z, et al. Risk assessment of novel coronavirus 2019-nCoV outbreaks outside China[J]. *Journal of Clinical Medicine*, 2020, 9(2): 571.
- [22] DANON L, BROOKS-POLLOCK E, BAILEY M, et al. A spatial model of CoVID-19 transmission in England and Wales: Early spread and peak timing[EB/OL]. [2020-02-14]. <https://doi.org/10.1101/2020.02.12.20022566>.
- [23] MING W K, HUANG J, ZHANG C J P. Breaking down of healthcare system: mathematical modelling for controlling the novel coronavirus (2019-nCoV) outbreak in Wuhan, China[EB/OL]. [2020-01-30]. <https://doi.org/10.1101/2020.01.27.922443>.
- [24] 原存德, 胡宝安. 具有阶段结构的 SI 传染病模型[J]. *应用数学学报*, 2002, 25(2): 193-203.  
YUAN Cun-de, HU Bao-an. A SI epidemic model with two-stage structure[J]. *Acta Mathematicae Applicatae Sinica*, 2002, 25(2): 193-203.
- [25] CHEN L J, SUN J T. Global stability of an SI epidemic model with feedback controls[J]. *Applied Mathematics Letters*, 2014, 28: 53-55.
- [26] LIU J G, LIN J H, GUO Q, et al. Locating influential nodes via dynamics-sensitive centrality[J]. *Scientific reports*, 2016, 6: 21380.
- [27] 吕亚兰, 刘聪, 周文正, 等. 新型冠状病毒肺炎与 SARS、MERS 的流行病学特征与防控措施比较[J]. *医药导报*, 2020, 39(3): 334-337.  
LÜ Ya-lan, LIU Cong, ZHOU Wen-zheng, et al. Epidemiological characteristics of COVID-19, SARS and MERS and their prevention and control measures[J]. *Herald of Medicine*, 2020, 39(3): 334-337.
- [28] MUNSTER V J, KOOPMANS M, VAN DOREMALEN N, et al. A novel coronavirus emerging in China—key questions for impact assessment[J]. *New England Journal of Medicine*, 2020, 382(8): 692-694.
- [29] GUAN W J, NI Z Y, HU Y, et al. Clinical characteristics of coronavirus disease 2019 in China[EB/OL]. [2020-02-28]. <https://www.nejm.org/doi/full/10.1056/NEJMoa2002032>.