

新型冠状病毒肺炎早期时空传播特征分析*

王聪¹⁾²⁾ 严洁^{2)3)†} 王旭⁴⁾ 李敏⁵⁾

1) (四川警察学院计算机科学与技术系, 泸州 646000)

2) (四川警察学院, 四川警事科学研究院, 成都 610200)

3) (四川警察学院道路交通管理系, 泸州 646000)

4) (四川师范大学影视与传媒学院, 成都 610068)

5) (四川师范大学计算机科学学院, 成都 610068)

(2020年2月25日收到; 2020年3月11日收到修改稿)

通过最新公布的流行病学数据估计了易感者-感染者模型参数, 结合百度迁徙数据和公开新闻报道, 刻画了疫情前期武汉市人口流动特征, 并代入提出的支持人口流动特征的时域差分方程模型进行动力学模拟, 得到一些推论: 1) 未受干预时传染率在一般环境下以 95% 的置信度位于区间 $[0.2068, 0.2073]$, 拟合优度达到 0.999; 对应地, 基本传染数 R_0 位于区间 $[2.5510, 2.6555]$; 极限环境个案推演的传染率极值为 0.2862, 相应的 R_0 极值为 3.1465; 2) 百度迁徙规模指数与铁路发送旅客人数的 Pearson 相关系数达到 0.9108, 有理由作为人口流动的有效估计; 3) 提出的模型可有效推演疫情蔓延至外省乃至全国的日期, 其中 41.38% 的预测误差 ≤ 1 d, 79.31% 的预测误差 ≤ 3 d, 96.55% 预测误差 ≤ 5 d, 总体平均误差约为 2.14 d.

关键词: 新型冠状病毒肺炎, 流行病学模型, 交通流, 人口迁徙

PACS: 07.05.Tp, 07.05.Kf, 01.75.+m, 45.70.Vn

DOI: 10.7498/aps.69.20200285

1 引言

2020年伊始, 新型冠状病毒肺炎(COVID-19)疫情猝然爆发. 临近春节, 疫情在春运因素的叠加下从九省通衢的武汉三镇向祖国大地迅速蔓延. 目前国内疫情发展已相对趋于平稳, 但国际视角下发展趋势却日益恶化: 截至2020年3月9日, 疫情已出现在全球102个国家和地区, 其中意大利、伊朗和韩国等国家确诊人数均已达到数千人. 如不采取紧急措施, COVID-19的全球流行趋势将日趋严重. 考虑到疫情早期传播过程客观上更能反映其传染特征, 为给国内可能产生的局部复发及国际疫情防控协作提供有意义的决策参考, 利用当前相对丰

富的病例调查数据, 挖掘疫情在爆发初期的传染和域间传播规律仍有必要.

COVID-19的基本数据估计是研究的基础. 来自《新英格兰医学杂志》的研究^[1]对COVID-19早期的425名感染者的统计显示: 疾病平均潜伏期约为5.2 d, 平均传染周期约为7.5 d, 基本传染数(basic reproduction number, R_0)约为2.2. Zhou等^[2]利用公开新闻报道数据代入仓室模型得到了另一个可能的结果, 他们初步估计, COVID-19的 R_0 值约为2.8—3.3, 传播能力接近SARS. 通过对“钻石公主”号邮轮个案的分析, 文献^[3]认为在95%的置信度下, R_0 的值约为2.28(2.06—2.52). 然而该文献对每日感染人数的泊松分布假设并未得到其他文献印证, 因乘客和工作人员陆续离船和核酸检

* 国家自然科学基金(批准号: 61602331)资助的课题.

† 通信作者. E-mail: yan_jie@foxmail.com

测周期带来的时变因素也不能完全忽略. 改进的仓室模型同样被应用于防控效果评价^[4]和拐点预测^[5]. 但传统的仓室模型并不能够模拟疫情在地域间的传播,《科学》杂志的一项研究^[6]利用复杂网络对人类迁徙流动进行了分析,指出疾病的传播与城市间的“等效距离”,而非地理距离密切相关,而等效距离的核心构成因子是城市间的交通流量. 因此,考虑人口迁徙的疫情传播模型得到了广泛关注,《柳叶刀》的一项研究^[7]利用了航空网络人口流量数据和腾讯移动数据,通过一种考虑人口迁移的 SEIR 模型^[6,8]首次对本次疫情的时空传播特征进行建模,得出了 R_0 的估计值为 2.68,并预测疫情在全国的指数级爆发将会在 1—2 周后出现. 然而该研究的主要问题在于,铁路而非航空才是国内春运期间的核心运力;假定武汉至其他城市的流量相等,与真实的域间交通流特征同样有着显著差异. Yang 等^[9]基于百度迁徙^[10]数据和中国疾病预防控制中心 (Chinese Center for Disease Control and Prevention, CCDC) 每日公布的上报病例数^[11],参照 SARS 的流行病学调查,利用深度学习的方法预测了 COVID-19 在国内和几个重点省份的发病曲线,认为疫情在 4 月底前难以完全平息. 该研究的结果紧密依赖于百度迁徙规模指数,而该指数构造方法目前并未公开,与真实人口迁徙的关系并不完全明确. 由于流行病学模型的预测结果通常显著依赖于初始系数的取值,以上研究在参量估值上的差异可能使得使用不同来源的研究结果大相径庭^[12]. 更重要的问题在于,来自 CCDC 的针对截至 2 月 11 日中国内地所有报告病例的调查^[13]证实,前期各项研究采用的上报日期数据与真实发病日期之间在时间和波形上均存在显著差异. 作为真实发病日期的一种近似,确诊和上报日期存在的这种误差对前期各项研究的可靠性也提出了挑战. 考虑到有限的医疗资源、社会承受力和行政成本^[14],各级防控措施的介入时点和力度是另一个亟待研究的问题. 管控介入时点的一个重要参考是疫情首次到达本地时间^[15],Chen 等^[16]建立了全球流行病的城际传播模型,利用航空流量数据对疾病到达航空网中任一城市的时间进行了估计. 该研究将城市间的交通流量以及疾病传染率定义为常数. 而在真实环境下,传染率等关键参数通常是时变^[17,18]的,且对于本次疫情而言,春运因素使得交通流量的常数假定也难以成立. 通过以上分析可知,目前的研究集中

于疫情基本参数的估计和疫情传播防控过程的建模和预测,且受限于早期相对匮乏的数据,进一步的有效性交叉印证和误差校正也有其必要性.

本文通过甲类传染病防控措施尚未启动时病例的回顾数据,首次利用病例的发病日期而非确诊上报日期,给出了未受干预条件下 COVID-19 传染能力估计,并利用一宗极端环境下的个案讨论了疫情的极限传染能力. 在此基础上,首次给出了疫情蔓延至全国各省区的关键时点估计:利用百度迁徙的公开数据和新闻报道数据推断湖北至各省区的真实人口迁徙流量,并与其他来源的公开报道数据进行了交叉校验;利用估计的参量值和人口迁徙流量构造了时域差分方程传播模型,用于推断其他地区首例病人到达时间,并将预测结果与收集得到的全国各省区首达病例时间做对比,验证了模型的有效性. 本文的结论对复工复产复学过程中疫情局部复发的防控工作可能有所助益,也为疫情在域外的传播防控提供了参考.

2 疫情的初始传染情况估计

2.1 基本数据

来自 CCDC 最新文献的统计^[13],截至 2 月 11 日,确诊病例共计 44672 人. 从中可整理 12 月 31 日前病例发病日期统计如表 1.

2020 年 1 月 1 日以后的精确数据文献^[13]并未给出,本文采用两种方法进行估计:

1) 利用 GetData Graph Digitizer 软件抓取文献^[13]公布的每日发病人数统计直方图,将其转化为近似数值;

2) 利用文献^[19]提出的蒙特卡罗方法,从该文献拟合的韦布尔分布中采样得到发病日期与上报日期的时间差,并结合 CCDC 网站^[11]爬取得到的各地上报数据,反推出每日发病人数的估计值.

为校验方法的可靠性,统计两种估计方法得到的每日累积发病人数,依据文献^[13]明确给出的 1 月 10 日,20 日和 31 日数据推算的累积发病人数和对应日期上报 CCDC 的确诊人数如表 2 所列.

由于 CCDC 公布的确诊病例统计自 1 月 3 日开始,表中 12 月 31 日的上报人数以 3 日人数代替. 表 2 有两点值得注意. 第一,文献^[13]数据与上报数据,特别是疫情前期的数据存在显著差距. 其中一个重要原因是,文献^[13]以发病日期统计病

表 1 2019 年累积发病人数

Table 1. The cumulative number of confirmed cases in 2019.

日期	人数	日期	人数	日期	人数	日期	人数	日期	人数
12/09	1	12/17	12	12/21	29	12/25	47	12/29	78
12/11	2	12/18	14	12/22	37	12/26	49	12/30	90
12/12	5	12/19	16	12/23	40	12/27	59	12/31	102
12/15	8	12/20	25	12/24	45	12/28	68		

表 2 2020 年各时间段累积发病人数

Table 2. The cumulative number of confirmed cases in each time slots in 2020.

截至时点	估计发病人数		实际发病人数	上报CCDC人数
	软件抓取法	蒙特卡罗法		
2019/12/31	102	49	102	0
2020/01/10	738	459	738—781	41
2020/01/20	6162	5800	6143—6187	291
2020/01/31	32661	29654	32633—32677	11821
2020/02/11	44692	69163	44672	44730

例, 而 CCDC 网站以各地上报日期统计病例. 这一差别极有可能增大前期部分研究的误差. 考虑到真实发病日期更具参考意义, 在此以文献 [13] 数据作为参数估计的基准数据. 第二, 表 2 的多数数据为某一区间范围. 这是因为文献 [13] 百分比仅保留至小数点后一位, 由此带来的舍入误差在早期病例数量研究中不容忽视, 因此按照四舍五入的原则, 反推出单项数据的上下限. 从表 2 可以看出, 无论是蒙特卡罗法亦或软件抓取法, 所得数据与文献 [13] 给出的当日真实累积发病人数均较为接近, 其参考价值可以互相印证. 比较而言, 软件抓取法的数值均落入依据文献 [13] 反推的区间内; 且从 2 月 11 日的总计数据来看, 软件抓取法获得的累积估计发病人数与截至当日的实际累积发病人数误差仅约 0.04%; 受 2 月 11 日将临床诊断纳入确诊依据这一重大变化的影响, 12 日新增确诊病例高达 15153 人, 导致蒙特卡罗法在 2 月 11 日的估计误差达到了 54.82%, 因此本文采用软件抓取法的数据为基准.

2.2 一般情况下的传染率与基本传染数估计

仓室模型是目前流行病传播研究的主流方法之一. 模型通常包括易感者 (susceptible, S)、感染者 (infected, I)、潜伏期感染者 (exposed, E) 和康复者 (recovered, R) 四类人群, 根据考察目的的不同, 可分别构成 SI, SIS, SIR, SEIR 等模型. 考虑

到疫情初期尚不存在, 或极少存在康复者, 且潜伏期感染者同样存在传染的可能, 简化讨论起见, 将潜伏者与感染者合并计算, 并以 SI 模型模拟疫情初期发病情况. 在此假定下, 潜伏期对模型的核心影响可近似表达为时域上的平移, 不影响其他参数的实质性讨论. 令总人口为 N ; 人群中初始感染者人数为 i_0 ; 传染率, 即每名传染者每天使得易感者致病的概率为 β ; 则在时刻 t , 感染者的数量 i_t 可表达为

$$i_t = \frac{N}{1 + \frac{(N - i_0)}{i_0} e^{-\beta t}}. \quad (1)$$

可见这是一个 Logistic 函数. 将首次出现发病病例的 2019 年 12 月 9 日作为坐标原点, 并考虑到我国在 1 月 21 日将 COVID-19 列为乙类法定传染病, 并采取甲类传染病的防控措施, 利用 2020 年 1 月 21 日及以前的数据进行估计. 将 N 定义为武汉城市长期居住人口, 公开资料显示约为 1418.65 万人. 注意到截至 12 月 31 日疫情虽已传播至湖北部分市州, 然而由于此时病例主要集中在武汉, 且总人口基数足够大, 使用武汉亦或湖北的人口数据为基数对估计结果不产生决定性影响. 将表 1 和表 2 数据代入 (1) 式做回归分析, 可得在不同时间点上, β 的拟合值, 95% 置信区间与对应的拟合优度 (goodness of fit, R^2) 如表 3 所列.

从表 3 中可以看出, 即使在疫情早期数据不够充分, 且受随机因素影响波动较大的情况下, 模型

表 3 传染率 β 可能的取值
Table 3. Possible values of the infection rate.

截止日期	β	置信区间	R^2	截止日期	β	置信区间	R^2
2019/12/31	0.2213	[0.2152, 0.2274]	0.868	2020/01/11	0.2066	[0.2056, 0.2274]	0.990
2020/01/01	0.2171	[0.2116, 0.2225]	0.878	2020/01/12	0.2063	[0.2056, 0.2071]	0.993
2020/01/02	0.2168	[0.2127, 0.2209]	0.923	2020/01/13	0.2060	[0.2054, 0.2067]	0.995
2020/01/03	0.2159	[0.2127, 0.2191]	0.949	2020/01/14	0.2059	[0.2054, 0.2064]	0.997
2020/01/04	0.2155	[0.2130, 0.2179]	0.967	2020/01/15	0.2056	[0.2052, 0.2060]	0.998
2020/01/05	0.2138	[0.2118, 0.2159]	0.973	2020/10/16	0.2058	[0.2054, 0.2061]	0.998
2020/01/06	0.2127	[0.2109, 0.2144]	0.980	2020/01/17	0.2060	[0.2057, 0.2063]	0.999
2020/01/07	0.2109	[0.2093, 0.2126]	0.980	2020/01/18	0.2064	[0.2061, 0.2066]	0.999
2020/01/08	0.2091	[0.2075, 0.2107]	0.979	2020/01/19	0.2065	[0.2063, 0.2067]	0.999
2020/01/09	0.2080	[0.2067, 0.2094]	0.984	2020/01/20	0.2066	[0.2064, 0.2068]	0.999
2020/01/10	0.2067	[0.2054, 0.2080]	0.985	2020/01/21	0.2070	[0.2068, 0.2073]	0.999

的拟合优度最低亦能达到 0.868; 随着日期的推移, 可以获取的数据日渐丰富, 传染率 β 的取值从 0.2213 逐渐收敛到 0.2060—0.2070, 拟合优度也随之上升至 0.999. 可以认为 SI 模型能够有效模拟疫情早期的爆发情况. 抽取表 3 中 β 在几个时间点上, 包括 12 月 31 日, 1 月 3 日, 1 月 9 日和 1 月 20 日的取值对疫情做拟合, 可得图 1.

表 4 重要时间节点的累积发病人数
Table 4. Cumulative confirmed cases in key time nodes.

截至时点	β			
	0.2213	0.2159	0.2080	0.2066
2019/12/31	130	116	97	95
2020/01/10	1190	1001	777	753
2020/01/20	10870	8666	6220	5964

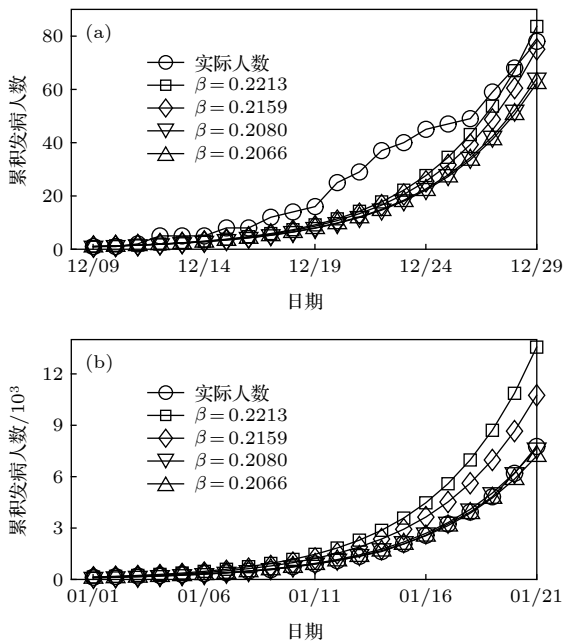


图 1 不同的 β 取值下实际累积发病人数与预测发病人数的对比 (a) 2019 年; (b) 2020 年

Fig. 1. Comparing the actual cumulative number of cases and its estimations according to different β : (a) Year 2019; (b) year 2020.

注意图 1 纵轴取对数. 为便于观察, 将几个重要的时间节点数据列于表 4 中.

由表 4 可见, 早期数据中存在的较大波动可能导致显著的计算误差. 而代入 1 月 10 日和 20 日的日期后, 无论取上限还是下限所受波动均在可接受的范围内. 因此有理由将该表的结尾, 即 1 月 20 日的 β 取值和置信区间作为传染率的有效估计.

一般认为基本传染数 (basic reproduction number) R_0 的一个可行的计算公式为^[20]

$$R_0 = 1 + \beta T_g, \quad (2)$$

其中 T_g 为平均感染天数, 即病例从具备传染能力到康复或死亡而失去传染能力的平均时长. 由此可得到 R_0 在疫情早期的一般估计. 文献 [1] 给出的 T_g 可能的取值为 7.5 d. 《新英格兰医学杂志》最近的研究^[21]对 1099 个病例进行了调查, 显示 COVID-19 潜伏期中位数和四分位距 (interquartile range, IQR) 为 4.0 d (2.0—7.0 d); 自出现症状至发展为肺炎的中位数和 IQR 为 3.0 d (1.0—6.0 d). 由于 COVID-19 在潜伏期同样具有传染性, 且从 IQR 来看, 这两个数据均存在一定程度的右偏, 有理由认为 T_g 的取值为 (4.0+3.0) d + 右偏量, 其后病人应已得到收治. 通过不同文献的相互印证, 可以推定 $T_g = 7.5$ 的取值在当前研

究阶段是有依据的. 将 1 月 21 日估计的 β 取值和置信区间代入 (2) 式, 估计对应的 R_0 为 2.5525, 95% 置信区间为 [2.5510, 2.6555].

2.3 极端环境下的个案分析

截至目前本文的分析仅涉及了一般情况. 全国范围内的复工复产目前已陆续展开, 复学工作也在谨慎研究中. 在特殊情况下, 如环境更封闭 (工厂车间)、人口密度更大 (中小学与大中专院校)、人际接触更频繁 (办公环境与写字楼) 等, 一旦爆发疫情, 后果仍不堪设想. 因此本文专门针对一宗封闭环境下中小人群聚集性疫情的个案进行分析.

据 2 月 21 日新闻报道 [22], 山东任城监狱发生聚集性疫情. 监狱是典型的同时具备封闭环境、高人口密度和人际频繁接触等特征的研究样本, 在此将其作为极端情况进行分析. 相对于文献 [3] 研究的“钻石公主”号邮轮个案, 监狱环境更加封闭, 人口密度更高, 且任城监狱的个案核酸检测周期较短, 可以将“钻石公主”号邮轮因检测周期和乘客在长达 10 日内陆续离船导致的时变因素最大限度地排除. 由于监狱在押人员数量和人员编制不属于政府信息公开, 本文根据公开报道 [23] 进行侧面推算. 该报道称, 截至 2 月 20 日该监狱确诊病例 207 例, 核酸检测人数为 2077 人, 因此本文假定 $N = 2077$, $i_t = 207$; 同时有报道称 2 月初有个人出现咳嗽症状, 因此假定 $i_0 = 1$, 坐标原点为 2 月 1 日, 可得到 $t = 19$. 考虑到缺乏数据的交叉印证, 在此并未沿用文献 [3] 中以泊松分布拟合每日发病人数的方法, 而采用了点估计, 将 $t = 19$ 时的确诊人数 $i_t = 207$ 直接代入 (1) 式进行求解, 可得 $\beta \approx 0.2862$; 代入 (2) 式可得对应的 R_0 达 3.1465, 显著高于一般情况下的传染率. 在此条件下, 患病人数翻倍时间约为 2.42 d. 考虑到 2077 人的数量级与多数中小学或中等规模的工厂接近, 这为我们敲响了警钟, 日常生产生活中的疫情防控工作仍不能放松.

2.4 简要小结

本节估计, 在 95% 置信区间内, 未加人工干预时 COVID-19 的传染率 $\beta \in [0.2068, 0.2073]$, 其极大似然估计为 0.2070; 进而推论 R_0 在 [2.5510, 2.6555] 间波动, 极大似然估计为 2.5525, 与学界对 SARS 较为乐观的估计相近 [24,25]; 作为一个可能

的极限值, 在极端利于病毒传播的环境下, R_0 可能达到 3.1465, 与 SARS 较为悲观的估计相近 [26]. 虽较文献 [2] 的估计略低, 但仍可得到相似的推论, 即 COVID-19 属于传染性中等偏高的疾病.

3 疫情的时空迁徙特征

回顾 SARS 及本次疫情的流行过程, 一个共同的特点都是从一点爆发, 随着人员迁徙流动迅速蔓延至全国. 显然, 如果能够在疫情传播的早期切断域间传染途径, 经济和社会承受的压力会小得多. 因此研究疫情的时空迁徙特征, 特别是爆发初期的特征十分必要. 本节依赖前文的基础参量估计和百度公开的迁徙数据, 构造了简易的时域差分方程模型, 模拟了疫情早期的地域迁徙状况, 并估计了各省级行政区域的首例病人输入时间.

3.1 迁徙流量的基础数据估计

百度迁徙网站 [10] 开放了 2020 年至今全国所有省区和地级市的迁徙规模指数, 以及每个区域的 Top 100 人口流量百分比. 其中迁徙指数和人口流量百分比均分别公布了迁入和迁出数据. 通过对该网站的爬取, 将元旦至春节百度迁徙规模指数与 2019 年同期数据进行对比, 结果如图 2 所示.

考虑历年春运工作均围绕农历新年安排, 因此以农历日期为基准进行统计. 最新研究 [27] 认为虽然在关闭离汉通道前产生短暂的恐慌性迁出人流, 但今年武汉春运期间的迁徙流量与往年相比没有显著的统计学差异, 可以认为离汉人员主体是正常春节返乡人流. 然而由于迁徙指数的构成目前并未公开, 有必要对该指数与迁出人口的对应关系进行验证.

考虑到铁路运力是春运期间域间流动的主体, 且历年春运中铁路运力的占比总体稳定, 从公开的新闻报道中抽取了武汉市三大火车站 (武汉站、汉口站、汉阳站) 春节前的发送旅客人数. 由于文献 [27] 验证了不同年份春运期间人口流量的相似性, 丰富样本起见, 同样抽取了 2019 年春运的报道, 得到统计结果如表 5 所列.

限于信息采集手段, 表 5 的数据缺失了部分日期的发送旅客人数.

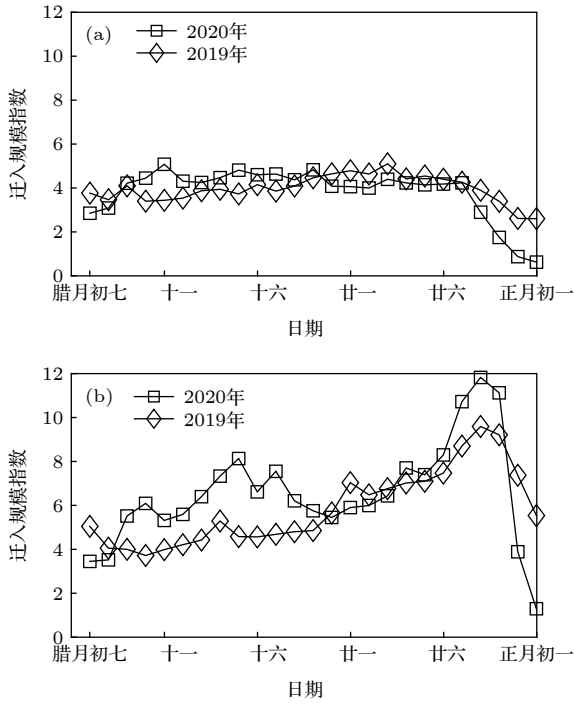


图 2 武汉“封城”前夕迁徙指数与 2019 年的对比 (a) 迁入规模指数; (b) 迁出规模指数
 Fig. 2. Comparing the migration index of Wuhan before the spring festival with the same period of 2019: (a) Inner migration index; (b) outer migration index.

如表 5 所列, 迁徙指数与三大火车站发送旅客人数之间的 Pearson 相关系数达到 0.8408, 二者呈强正相关; 考虑到铁路的极限运力和反应滞后性 (春运期间的车票预订因素), 且 1 月 22 日宣布即将关闭离汉通道可能导致恐慌, 将该点数据作为异常点剔除, 重新计算的 Pearson 相关系数达到 0.9108, 故可以认为二者极强正相关. 因此可以粗略假定区域 j 的迁徙指数 M_j 与实际迁徙人口 p_j 满足线性关系:

$$M_j = kp_j. \quad (3)$$

一个重要的问题是参数 k 的取值. 针对湖北省

的全部市州和直属行政区 (天门、潜江、神农架林区等), 有

$$\sum_{j \in \text{湖北省}} M_{j,\text{out}} = k \sum_{j \in \text{湖北省}} p_{j,\text{out}}, \quad (4)$$

其中 $M_{j,\text{out}}$ 和 $p_{j,\text{out}}$ 分别是区域 i 的迁出指数和实际迁出人数 (万人).

来自《荆州日报》和湖北省春运办的新闻报道称 [28]: “春运首周, ……截至 17 日零时, 全省铁、水、公、空发送旅客 1050.7 万人次”, 可知

$$\sum_{j \in \text{湖北省}} p_{j,\text{out}} \approx 1050.7. \quad (5)$$

统计报道同期湖北省所有直属区域的迁出指数, 可得

$$\sum_{j \in \text{湖北省}} M_{j,\text{out}} = 105.7837644, \quad (6)$$

于是得到比例系数 k 的估计约为 0.100679.

进一步, 武汉市春运开始至封城期间的迁出指数合计 107.116992, 迁入指数合计 56.4331536, 可估计武汉市在此期间人口净流出为 503.4815 万人. 该估计亦可与武汉市政府在 1 月 26 日的新闻发布会上公开的迁离人数相互印证 [29], 可以认为该推论具备一定的参考价值.

3.2 疫情到达时间推演

防控决策时需要回答的重要问题是, 疫情将于或已于何时到达本地. 在此, 提出了一个简易的时域差分模型, 利用迁徙数据, 推导疫情首次到达特定行政区域的可能时间.

首先给出 SI 模型的差分形式:

$$i_t = \left(i_{t-\Delta t} - \frac{\beta i_{t-\Delta t}^2}{N} + \beta i_{t-\Delta t} \right) \Delta t. \quad (7)$$

当相对于 $i_{t-\Delta t}$ 和 β , N 取值极大, 即易感者人数占

表 5 武汉市三大火车站发送旅客人数与迁出指数

Table 5. The Baidu inner migration index and the number of the travelers sent from Wuhan's major railway stations.

日期	迁徙指数	人数/万	日期	迁徙指数	人数/万	日期	迁徙指数	人数/万
2020/01/10	6.6232	27	2020/01/22	11.8403	29.96	2019/01/29	7.0282	27.2
2020/01/11	7.5612	29.8	2019/01/21	4.5718	21.6	2019/01/30	7.1072	27.7
2020/01/12	6.2165	27	2019/01/22	4.6892	21.4	2019/01/31	7.4800	28.1
2020/01/13	5.7620	24.8	2019/01/23	4.8062	23	2019/02/01	8.7140	29.8
2020/01/15	5.9087	26.5	2019/01/24	4.8605	21.7	2019/02/02	9.6043	31.5
2020/01/16	6.0028	27.7	2019/01/26	7.0436	27	2019/02/03	9.2247	29.1
2020/01/19	7.4060	30	2019/01/28	6.7706	26.8			

绝对优势时, 有 $\beta i_{t-\Delta t}^2/N \rightarrow 0$. 便于计算起见, (7) 式可近似为

$$i_t \approx i_{t-\Delta t} + \beta i_{t-\Delta t} \Delta t, \quad (8)$$

其中 Δt 定义为一个自然日, 即满足 $\Delta t = 1$. 当考虑人口迁徙因素时, 假定人口迁徙概率与感染概率条件独立, 对于区域 j , 有

$$i_{j,t} = i_{j,t-1} + \beta_{t-1} i_{j,t-1} + \sum_{n \neq j} \frac{M_{n,\text{out},t-1}}{kN_n} \sigma_{n \rightarrow j,t-1} i_{n,t-1} - \frac{M_{j,\text{out},t-1}}{kN_j} i_{j,t-1}, \quad (9)$$

其中 $i_{j,t}$ 是区域 j 在第 t 天的感染人数; $\sigma_{n \rightarrow j,t-1}$ 为第 $t-1$ 天区域 n 向区域 j 流动的人口占区域 n 当日迁出总人口的比例; $M_{n,\text{out},t-1}$ 为区域 n 在第 $t-1$ 天的迁出指数; N_n 为区域 n 的总人口数. 由表 3 可知, 由于传染率的估计会随着数据的丰富不断变化, 为了拟合这种数据估计上的时变特征, 在此将传染率 β 改写为 β_{t-1} , 即第 $t-1$ 天统计发病人数得到的传染率估计.

根据文献 [13], 截至 2019 年 12 月 31 日湖北省共有 14 个县报告病例. 考虑到仅武汉市辖下县区级行政区域就有 13 个, 可以认为在疫情初期, 绝大部分病例集中出现在武汉市, 因此仅考虑武汉

向外省迁移的情况.

疫情首次到达模拟和判决的伪 Python 代码算法如下:

初始化参量 $i(0), k, \forall N_n$

For $t = 1$ to today:

For Each j :

读入 $M_{n,\text{out},t-1}, \sigma_{n \rightarrow j,t-1}$

估计 β_{t-1}

利用 (9) 式求得 $i_{j,t}$

If $i_{j,t} \geq 1$:

报告区域 j 的疫情到达日期为 t

If 所有区域均已报告, 或数据迭代完毕:

Return.

在此必须考虑, 当疫情前期数据不够丰富时, 传染率 β 的估计值可能不够精确, 但此时却是实施决策的关键时刻, 利用后期得到的传染率代入预测在决策时点上是不做到的. 因此取多个时间点的传染率预测代入算法 1 进行仿真. 本文取 12 月 31 日, 1 月 3 日和 1 月 20 日 β 的估计值代入计算. 取 1 月 3 日的原因是, 利用 12 月 31 日的数值预测, 疫情将于 1 月 4 日蔓延至省外. 因而 3 日可以作为决策前夕的关键时点. 得到的结果如表 6 所列.

表 6 省级区域的首例到达时间
Table 6. The arrival times of each provinces.

省份	β			实际日期	省份	β			实际日期
	0.2213	0.2159	0.2070			0.2213	0.2159	0.2070	
安徽	01/06	01/06	01/07	01/07	辽宁	01/11	01/12	01/13	01/09
北京	01/07	01/07	01/08	01/08*	内蒙古	01/13	01/14	01/16	01/16
福建	01/09	01/09	01/10	01/06	宁夏	01/17	01/18	01/20	01/17
甘肃	01/10	01/11	01/12	01/04	青海	01/19	01/21	01/23	01/21
广东	01/05	01/05	01/06	01/04	山东	01/08	01/08	01/09	01/08
广西	01/08	01/09	01/09	01/13	山西	01/09	01/10	01/10	01/14
贵州	01/08	01/08	01/09	01/06	陕西	01/09	01/09	01/10	01/12
海南	01/10	01/11	01/12	01/13	上海	01/08	01/08	01/09	01/10
河北	01/08	01/09	01/09	01/13	四川	01/08	01/08	01/09	01/07
河南	01/04	01/04	01/05	01/03	天津	01/14	01/15	01/16	01/11
黑龙江	01/12	01/13	01/14	01/12	西藏	>01/23	>01/23	>01/23	01/30
湖南	01/05	01/05	01/06	01/05	新疆	01/12	01/12	01/14	01/17
吉林	01/15	01/15	01/17	01/14	云南	01/09	01/11	01/10	01/07
江苏	01/06	01/07	01/07	01/10	浙江	01/07	01/07	01/08	01/04
江西	01/06	01/07	01/07	01/07	重庆	01/08	01/08	01/09	01/06

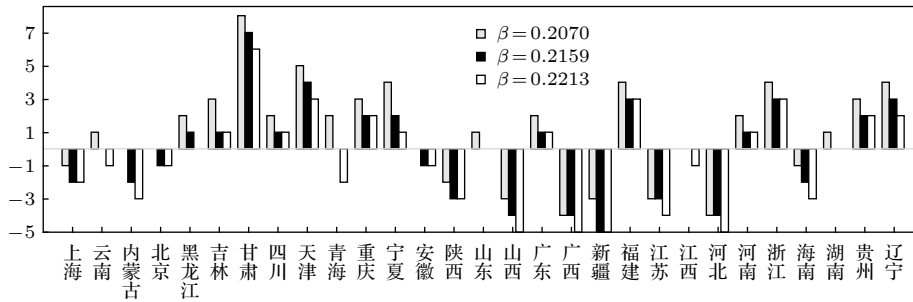


图3 各省区首达病例时间预测误差

Fig. 3. The estimation errors of the first arrival times for each provinces.

搜索各省级区域关于首例确诊病例的新闻报道,并与腾讯网“新冠肺炎确诊病患活动轨迹”专题^[30]提供的病例轨迹数据合并,整理了病例抵达当地的时间填入表6中.表中海南和北京进行了特别标注,原因是:虽然海南第124号确诊病例于1月2日自武汉返琼,但发病日期为2月5日,期间间隔35 d,超过了文献^[1,21]报道的最长潜伏期,不应认定为输入病例,故将该省首达日期认定为第157号病例抵琼日期,即1月13日;北京市病例轨迹数据缺失,但据新闻报道^[31,32],该市至迟1月8日凌晨即存在输入病例,在此取该日期作为首达日期;西藏自治区仅1例病例且到达于1月30日,其统计意义不足,予以剔除.

各省区病例首达时间的预测误差如图3所示.就已搜集的信息来看,模型对疫情向省外的蔓延时点的估计较为准确.基于12月31日的数据预测,疫情将于1月4日蔓延至省外.考虑到蔓延前夕的预测对决策最为关键,因此以1月3日的预测数据为例分析,疫情首先于1月4日蔓延到河南省,首达省份与模型预测一致,首达时间预测误差也仅有1 d.模型估计各省首达时间的平均预测误差为2.14 d,其中有41.38%的预测误差 ≤ 1 d,79.31%的预测误差 ≤ 3 d;除甘肃外,所有省份的首达时间预测误差均 ≤ 5 d,反映了模型的有效性.考虑到各省最早上报病例日期为1月19日,相对在省外蔓延的确切日期滞后16 d,采用模型预测得到的理论预测值显得更有价值.即使择取12月31日和1月20日的模型参量估计也可达到较好效果,其中预测疫情蔓延至省外的时间误差分别为1和2 d,首达省份预测无误;平均预测误差也仅为2.31和2.48 d.表6可能反映的另一个沉痛的现实是,在迅猛发展的疫情面前,强力交通管制措施采取的时间同样显得滞后.如果根据模型的

预测,将强力交通管制措施提前到1月4日,即模型预测的即将蔓延至省外的时间施行,则可有效阻断疫情在除河南外全部省份的蔓延.

4 结论

本文利用最新的流行病学调查,取疫情尚未得到有力干预时的数据代入动力学模型进行拟合,给出了传染率和基本传染数的一般估计.为模拟复工复产过程中可能出现的聚集性疫情特征,利用公开报道的极端环境个案,给出了可能的传染率和基本传染数上限.在此基础上,综合流行病学调查数据、互联网爬取数据和公开新闻报道数据,在佐证百度迁徙规模指数参考价值的前提下,估计了疫情发源地迁至各地的绝对人口数量,代入设计的时域差分模型,模拟了疫情初期在地域间的传播过程,给出了疫情首次到达各省级区域的时间估计,并与新闻报道数据做比对,验证了模型的有效性.

本文的创新在于:首次以流行病学调查中公布的发病时间,而非确诊和报告时间为基准进行疫情推演.相较于两个时间节点间约10 d的错峰,这一选择显然更具合理性;给出了已得到初步应用的百度迁徙规模指数的有效性验证,侧面证实了它与人口迁徙规模的粗略线性关系;利用人口迁徙数据,首次给出了疫情到达全国各省的时间预测.有限度使用新闻报道数据以改善疫情初期科研数据尚不够丰富的现状也是本文的一个特点.

本文还存在以下不足,将在后续工作中改进:本文依据的流行病学调查数据更多依赖于患者个人回忆,可能存在时间偏差,需要更加精确的数据进行进一步验证;本文提出的方法对疫情爆发后强力防控措施介入条件下的推演偏差可能较为显著,应重点思考优化方案;如何对各级力度的疫情防控

措施进行量化分析与效能评价,在有效控制疫情的前提下做到松紧适度,降低对经济社会的负面影响,也是一个值得研究的问题。

参考文献

- [1] Li Q, Guan X H, Wu P, Wang X Y, Zhou L, Tong Y Q, Ren R Q, Leung S, Lau E, Wong J, Xing X S, Xiang N J, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Li M, Tu W X, Chen C D, Jin L M, Yang R, Wang Q, Zhou S H, Wang R, Liu H, Luo Y B, Liu Y, Shao G, Li H, Tao Z F, Yang Y, Deng Z Q, Liu B X, Ma Z T, Zhang Y P, Shi G Q, Lam T, Wu J, Gao G, Cowling B, Yang B, Leung G, Feng Z J 2020 *N. Engl. J. Med.* **382** 1199
- [2] Zhou T, Liu Q H, Yang Z M, Liao J Y, Yang K X, Bai W, Lv X, Zhang W 2020 *J. Evid.-based Med.* **20** 3
- [3] Zhang S, Diao M Y, Yu W B, Pei L, Lin Z F, Chen D C 2020 *Int. J. Infect. Dis.* **93** 201
- [4] Cao S L, Feng P H, Shi P P 2020 *J. Zhejiang Univ.* **49** 1 (in Chinese) [曹盛力, 冯沛华, 时朋朋 2020 *浙江大学学报* **49** 1]
- [5] Fan R G, Wang Y B, Luo M, Zhang Y Q, Zhu C P 2020 *J. Univ. Electron. Sci. & Tech. China* (in press) (in Chinese) [范如国, 王奕博, 罗明, 张应青, 朱超平 2020 *电子科技大学学报* (in press)]
- [6] Dirk B, Helbing D 2013 *Science* **342** 1337
- [7] Wu J T, Leung K, Leung G M 2020 *Lancet* **395** e41
- [8] Wang J, Wang X, Wu J 2018 *Proceedings of the 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* London, United Kingdom, August 19–23, 2018 p2830
- [9] Yang Z F, Zeng Z Q, Wang K, Wong S S, Liang W H, Zanin M, Liu P, Cao X D, Gao Z Q, Mai Z T, Liang J Y, Liu X Q, Li S Y, Li Y M, Ye F, Guan W J, Yang Y F, Li F, Luo S M, Xie Y Q, Liu B, Wang Z L, Zhang S B, Wang Y N, Zhong N S, He J X 2020 *J. Thorac. Dis.* **12** 2077
- [10] Baidu.com. <https://qianxi.baidu.com> [2020-3-10]
- [11] Chinese Center for Disease Control and Prevention. <http://www.chinacdc.cn>[2020-3-10]
- [12] Cohen J 2020 <https://www.sciencemag.org/news/2020/02/scientists-are-racing-model-next-moves-coronavirus-thats-still-hard-predict>[2020-3-10]
- [13] The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team 2020 *Chin. J. Epidemiol.* **41** 145 (in Chinese) [中国疾病预防控制中心新型冠状病毒肺炎应急响应机制流行病学组 2020 *中华流行病学杂志* **41** 145]
- [14] Chen X L, Zhou T S, Feng L, Liang J B, Liljeros F, Havlin S, Hu Y Q 2019 *Phys. Rev. E* **100** 032310
- [15] Wang J, Wang L, Li X 2016 *IEEE Trans. Cybernetics* **46** 2782
- [16] Chen L M, Holzer M, Shapiro A 2018 *Chaos* **28** 013105
- [17] Chowell G, Viboud C, Hyman J, Simonsen L 2015 *PLoS Currents* **7** 1
- [18] Viboud C, Simonsen L, Chowell G 2016 *Epidemics-Neth* **15** 27
- [19] Chen D B, Zhou T 2008 arXiv: 2003.00305 v1[Quantitative Biology]
- [20] Wallinga J, Lipsitch M 2007 *Proc. R. Soc. Lond., B* **274** 599
- [21] Guan W J, Ni Z Y, Hu Y, Liang W H, Ou C Q, He J X, Liu L, Shan S, Lei C L, Hui D, Du B, Li L J, Zeng G, Yuen K, Chen R C, Tang C L, Wang T, Chen P Y, Xiang J, Li S Y, Wang J L, Liang Z J, Peng Y X, Wei L, Liu L, Hu Y H, Peng P, Wang J M, Liu J Y, Chen Z, Li G, Zheng Z J, Qiu S Q, Luo J, Ye C J, Zhu S Y, Zhong N N 2020 *N. Engl. J. Med.* (in press)
- [22] Sina.com. <https://news.sina.com.cn/c/2020-02-21/doc-iimxxstf3200725.shtml> [2020-2-21]
- [23] Huanqiu.com <https://china.huanqiu.com/article/9CaKrnKpveQ> [2020-2-21]
- [24] Lipsitch M, Cohen T, Cooper B, Robins J, Ma S, James L, Gopalakrishna G, Chew S K, Tan C C, Samore M, Fisman D, Murray M 2003 *Science* **300** 1966
- [25] Ho H, Fraser C, Lam T, Ghani C, Leung G, Leung G, Chau Y K, Ho P L, Lo , Abu-Raddad L, Donnelly C, Anderson D, Chan K, Lee K, Lau E, Hedley A, Riley S, Tsang T, Ferguson N, Thach D T 2003 *Science* **300** 1961
- [26] Wallinga J, Teunis P 2004 *Am. J. Epidemiol.* **160** 509
- [27] Xu X K, Wen C, Zhang G Y, Sun H C, Liu B 2020 *J. Univ. Electron. Sci. & Tech. China* (in press) (in Chinese) [许小可, 文成, 张光耀, 孙皓宸, 刘波 2020 *电子科技大学学报* (in press)]
- [28] Cnchu.com. <http://www.cnchu.com/jingxuan/guojiji/345.html> [2020-3-10]
- [29] Thepaper.cn. https://www.thepaper.cn/newsDetail_forward_5642450 [2020-3-10]
- [30] Tenent.com. <https://xw.qq.com/act/fytrace> [2020-3-10]
- [31] Tenent.com. <https://new.qq.com/omn/BJC20200/BJC2020012401027800.html> [2020-3-10]
- [32] Sohu.com. https://www.sohu.com/a/371749528_100030926 [2020-3-10]

Analysis on early spatiotemporal transmission characteristics of COVID-19*

Wang Cong¹⁾²⁾ Yan Jie^{2)3)†} Wang Xu⁴⁾ Li Min⁵⁾

1) (*Department of Computer Science & Technology, Sichuan Police College, Luzhou 646000, China*)

2) (*Institute of Sichuan Police Science, Sichuan Police College, Chengdu 610200, China*)

3) (*Department of Road Traffic Management, Sichuan Police College, Luzhou 646000, China*)

4) (*School of Movie and Media, Sichuan Normal University, Chengdu 610068, China*)

5) (*School of Computer Science, Sichuan Normal University, Chengdu 610068, China*)

(Received 25 February 2020; revised manuscript received 11 March 2020)

Abstract

In this paper, a simple susceptible-infected (SI) model is build for simulating the early phase of COVID-19 transmission process. By using the data collected from the newest epidemiological investigation, the parameters of SI model is estimated and compared with those from some other studies. The population migration data during Spring festival in China are collected from Baidu.com and also extracted from different news sources, the migration characteristic of Wuhan city in the early phase of the epidemic situation is captured, and substituted into a simple difference equation model which is modified from the SI model for supporting migrations. Then several simulations are performed for the spatiotemporal transmission process of COVID-19 in China. Some conclusions are drawn from simulations and experiments below.

1) With 95% confidence, the infection rate of COVID-19 is estimated to be in a range of 0.2068–0.2073 in general situation, and the corresponding basic reproduction number R_0 is estimated to be in a range of 2.5510–2.6555. A case study shows that under an extreme condition, the infection rate and R_0 are estimated to be 0.2862 and 3.1465, respectively.

2) The Pearson correlation coefficient between Baidu migration index and the number of travelers sent by railway is 0.9108, which indicates a strong linear correlation between them, thus it can be deduced that Baidu migration index is an efficient tool for estimating the migration situation.

3) The epidemic arrival times for different provinces in China are estimated via simulations, specifically, no more than 1 day within an estimation error of 41.38%; no more than 3 days within an error of 79.31%, and no more than 5 days with an error of 95.55%. An average estimation error is 2.14 days.

Keywords: COVID-19, epidemic model, traffic flow, migration

PACS: 07.05.Tp, 07.05.Kf, 01.75.+m, 45.70.Vn

DOI: 10.7498/aps.69.20200285

* Project supported by the National Natural Science Foundation of China (Grant No. 61602331).

† Corresponding author. E-mail: yan_jie@foxmail.com