

# 基于 SEIR 模型原理对 2019-nCoV 发展的拟合与预测

**摘要:** 2019 年 12 月, 中国武汉爆发了由新型冠状病毒 (2019-nCoV) 引起的肺炎。时值春节期间, 人口的大规模流动更是直接扩大了病毒的传播范围, 目前疫情已波及全国。本文基于修正的 SEIR 传播动力学模型思想分别从蝙蝠 (Bat)、中间宿主 (Host)、华南海鲜市场 (Market)、人 (People) 四个方面考虑整个病毒传播系统, 同时进一步考虑感染患者有无表现出明显特征以及追踪隔离干预措施的效果等因素提出了优化后的 SEIR 模型并对模型中的参数进行动态化处理, 通过这一模型对新型冠状病毒肺炎的发展轨迹进行拟合并对其未来发展做出了预测。本文发现基础再传播人数  $R_0$  最高时约为 3.17, 全国确诊感染人数将超过 70000 人, 预测峰值将会在 2020 年 2 月 25 日前出现, 上述预测结果表明 2019-nCoV 的传播方式包括人传人, 且传播速度非常快, 波及范围非常大。但是随着加强筛查、监测和控制等工作的进行, 特别是对于武汉以及其他交通枢纽的控制,  $R_0$  在 2020 年 2 月中旬降至 0.5 以下, 4 月初感染人数降至 2000 以下, 政府干预措施效果显著, 但是复工复产等工作的进行在一定程度上会影响疫情的发展, 保守预计疫情会在 4 月份中下旬会被基本控制。

**关键词:** 2019-nCoV, SEIR 模型, 动态化处理, 基础再传播人数  $R_0$

中国分类号 C330.21 文献标志码 A 文章编号

## 1 引言

2019 年 12 月 31 日, 世界卫生组织获悉中国武汉市出现了一组病因不明的肺炎病例。不久, 中国当局确认了一种新型冠状病毒 (2019-nCoV) 是此次疫情爆发的病原体。事实上, 早在 2019 年 12 月 8 日, 湖北省武汉市就出现了第一例新型冠状病毒患者, 病毒真正开始引起公众大规模注意始于武汉华南海鲜市场的聚集型患病、一次或多次人畜共患病传播事件, 截止到 2020 年 2 月 12 日, 全国累计发现 59882 例确诊病例, 单日新增病例 15153 例, 疫情并未呈现出好转的趋势, 而且目前仍未开发出疫苗, 疫情拐点尚未到来, 疫情防控形势依然严峻, 在与疫情斗争的关键阶段, 充分了解 2019-nCoV 的传播模式和持续性人际传播的潜力, 对 2019-nCoV 进行流行传播风险预测是至关重要的。有关病毒传播特征的信息将有助于理解当前的筛查和控制战略, 为此次疫情是否构成国际关注的突发公共卫生事件 (PHEIC) 的决策做出合理解释。

## 2 模型

### 2.1 模型说明

本文中的模型主要参考来自 <https://doi.org/10.1101/2020.01.19.911669> 文章的 BHRP 模型，该文章从蝙蝠，其他野生动物（病毒的宿主），华南海鲜市场，人群 4 个方面提出了一个数学模型，如图 1 所示，大致思想为：蝙蝠有很大的可能是新型冠状病毒的源头，果子狸、穿山甲等野生动物大概率是病毒的中间宿主，根据最新研究华南海鲜市场即使不是病毒的发源地，但是华南海鲜市场贩卖带有 2019-nCoV 的野生动物，同时其还有着人员密度大，人员流动快等特点，所以华南海鲜市场极大地推动了病毒的快速传播。当病毒进入人群，恰逢春节，人员流动较平常日期频繁，病毒会通过“人-人传播”这种方式快速蔓延，与 SARS 不同的是，本次的 2019-nCoV 潜伏期较长，而且还有约 17% 无症状感染者等，这些特点导致了感染者不易被发现，更直接促进了病毒的快速传播。基于上述事实，本文基于 SEIR 模型中的核心思想，并在 SEIR 模型的基础上考虑到了自然出生率、自然死亡率、病毒在不同物种之间的传播、病毒在华南海鲜市场的流通等因素提出一个模型用于拟合 2020 年 2 月 13 日之前的病毒发展，并预测之后的趋势。

表 1 BHMP 模型中参数及其说明

| 模型参数     | 参数说明              |
|----------|-------------------|
| $\beta$  | 传染概率              |
| $\mu$    | 病毒的生命周期           |
| $\omega$ | 潜伏者转化成感染者的速率      |
| $\gamma$ | 感染者恢复健康的速率        |
| $\delta$ | 潜伏者转化成无症状患者的概率    |
| $\xi$    | 无症状感染者将病毒携带到市场的概率 |
| $\alpha$ | 感染病毒后的死亡率         |
| S        | 处于易感染状态群体的数量      |
| E        | 处于病毒潜伏时期群体的数量     |
| D        | 死亡患者人数            |

|       |              |
|-------|--------------|
| $I_P$ | 有症状的已感染者     |
| $A_P$ | 无症状的已感染者     |
| $R$   | 处于已康复状态群体的数量 |
| $N$   | 某种生物的总数      |
| $m$   | 自然死亡率        |
| $n$   | 自然出生率        |
| $a$   | 野生动物的零售购买率   |

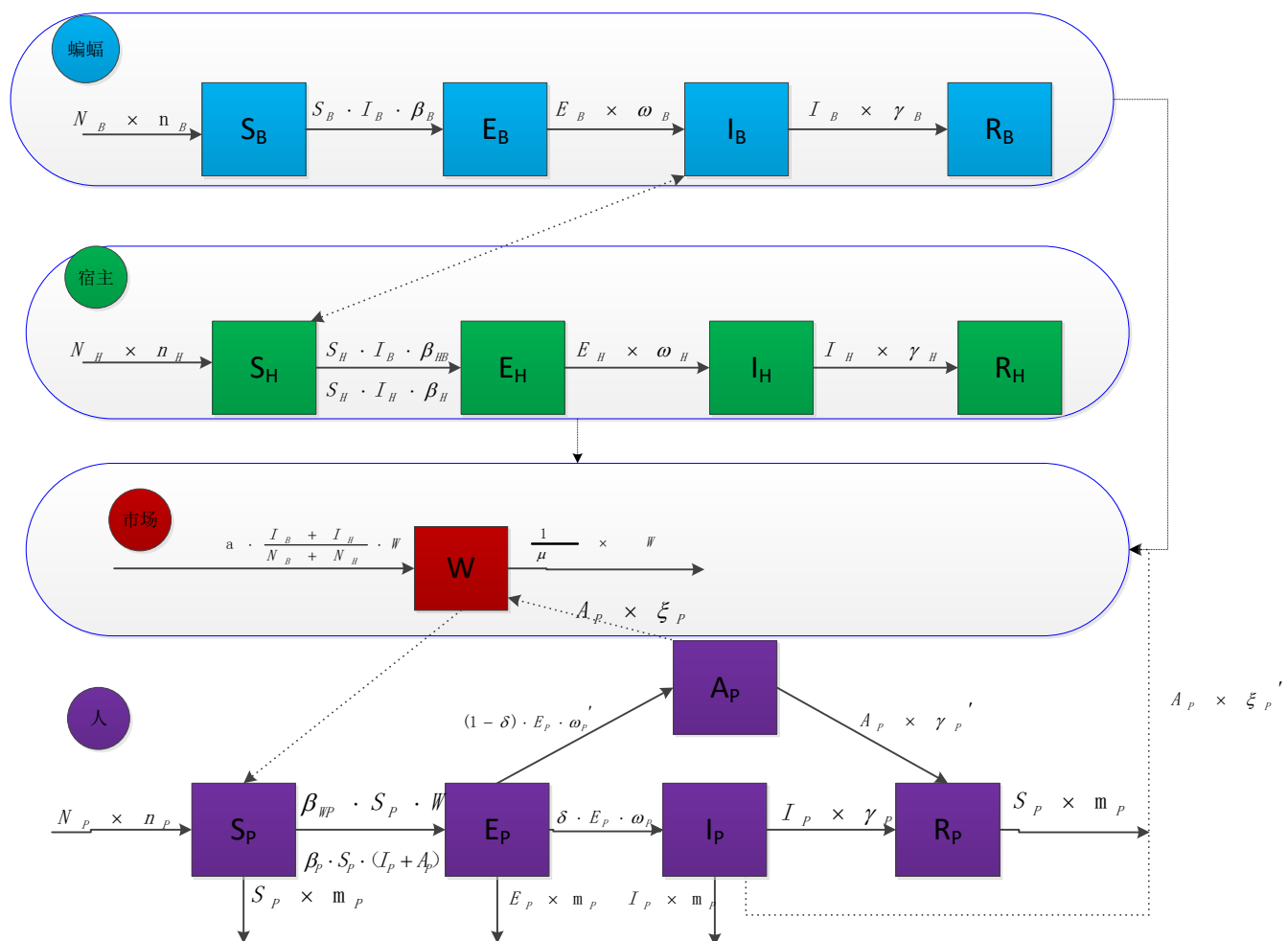


图 1 BHMP 系统模型

## 2.2 模型改进

本文在参考模型的基础上首先将其中 2 个重要的参数由静态改为动态，因为在疫

情发展期间，公众对于疫情的了解会逐渐深入，国家也会采取一定的措施，各种形式的隔离，防护措施的加强会使人人与人之间的接触率降低，病毒在人群中的传播率降低等。其次本文考虑到了生物与生物、生物与地点之间的联系，比如：考虑到了不同野生动物之间的交叉感染，华南海鲜市场除蝙蝠以外还贩卖其他携带病毒的野生动物，人群在感染病毒以后会表现出有症状或者无症状两种形式，无症状患者约占 17% 等。最后本文将 People 部分从整个系统模型中分离出来，因为当疫情开始引起公众的重视，病毒再从野生动物，海鲜市场传播到人群中的概率极小，因此为了简化模型，本文将模型的重点放在了拟合并且预测 2019-nCoV 在人群中发展情况。

(1) 在 BHMP 模型中，病毒在蝙蝠和野生宿主动物的传播可以用优化后的 SEIR 模型表示：即增加了出生率和死亡率要素，以求在尽可能接近自然环境的条件下推测疫情爆发的前期情况。以下为蝙蝠在整个病毒传播系统模型中抽象出的常微分方程：

$$\begin{cases} \frac{dS_B}{dt} = N_B n_B - S_B m_B - S_B I_B \beta_B \\ \frac{dE_B}{dt} = S_B I_B \beta_B - E_B (m_B + \omega_B) \\ \frac{dI_B}{dt} = E_B \omega_B - I_B (m_B + \gamma_B) \\ \frac{dR_B}{dt} = I_B \gamma_B - R_B m_B \end{cases}$$

野生宿主动物与蝙蝠的模型相似，唯一不同的地方在于病毒即会在蝙蝠于宿主动物之间传播，又会在宿主动物之间传播。变动的方程式为：

$$\frac{dE_H}{dt} = \beta_{BH} S_H I_B + \beta_H S_H I_H - E_H (\omega_H + M_H)$$

(2) 武汉华南海鲜市场作为病毒的大规模所在地，其来源绝大多数来自于贩卖的被病毒寄生的野生动物上，少数来自于已感染患者对于华南海鲜市场的病毒输入，根据病毒的自身特性，病毒均存在一定的生命周期，所以还有极少数的病毒因为生命周期到头而丧失毒性。这一过程的常微分方程式为：

$$\frac{dW}{dt} = a \frac{I_B + I_H}{N_B + N_H} W + A_P \xi_P + I_P \xi_P' - \frac{1}{\mu} W$$

(3) 特别需要注意的是，根据最新的研究发现，携带病毒的生物绝不仅仅是两三种野生动物，甚至病毒的起源可能不在中国，本文研究的模型的目的是为了拟合病毒在人群中的发展，并预测其未来走势，为管理措施提供理论支持，因此本文

重点研究的方向是 BHMP 系统模型中的 People 部分，本文在基于 SEIR 模型的基础上，将患者细分为有症状患者和无症状患者，同时增加了感染者死亡病例。为了更加切合实际的拟合并预测疫情发展，本文考虑到了疫情发展过程中人与人接触率的变化以及随着病毒周期等因素的变化导致病毒感染率的变化。

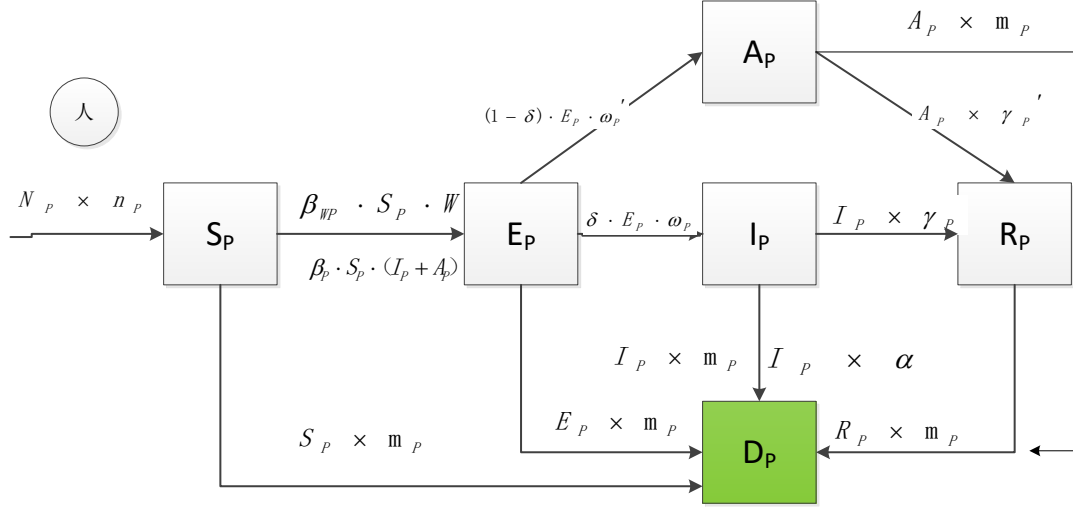


图 2 新型冠状病毒在人群中的传播

图 2 表示的简化的模型中的常微分方程如下：

$$\begin{cases} \frac{dS_p}{dt} = N_p \cdot n_p - S_p \cdot m_p - \beta_{HP} S_p W - \beta_p S_p (I_p + A_p) \\ \frac{dE_p}{dt} = \beta_{HP} S_p W + \beta_p S_p (I_p + A_p) - E_p m_p - (1 - \delta) E_p \omega_p' - \delta E_p \omega_p \\ \frac{dI_p}{dt} = \delta E_p \omega_p - I_p m_p - I_p \gamma_p \\ \frac{dA_p}{dt} = (1 - \delta) E_p \omega_p' - A_p \gamma_p' \\ \frac{dR_p}{dt} = A_p \gamma_p' + I_p \gamma_p - R_p m_p \\ \frac{dD_p}{dt} = (S_p + E_p + A_p + I_p + R_p) m_p + I_p \alpha \\ N_p = S_p + E_p + A_p + I_p + R_p \end{cases}$$

### 2.3 模型假设

(1) 人口总数不变：由于国际航空禁运，严格的居家隔离和肺炎较低的死亡率，这个假设基本成立

(2) 假设有症状患者的平均恢复期  $\gamma = 14$  天，恢复速率  $\frac{1}{\gamma} = \frac{1}{14}$

(3) 假设无症状患者的平均恢复期为  $\gamma' = 7$  天，恢复速率为  $\frac{1}{\gamma'} = \frac{1}{7}$

- (4) 假设病毒潜伏期为 5 天，并且各种症状潜伏期一致
- (5) 假设潜伏期未发病的人数是疑似病例的 4~5 倍
- (6) 假设感染 2019-nCoV 后的死亡率为 2%

#### 2.4 模型中两个动态参数的确定

(1) 当人群中出现了“零号超级感染者”后，病毒通过“人传人”这种方式迅速传播，在疫情爆发前期，公众对于病毒了解、防护知识、防护意识的不足，会导致病毒在人群中快速且广泛的传播，随着疫情的爆发，无论是国家的政策颁发，还是公众意识的增强，各种形式的隔离则会导致病毒的传播率大幅降低。本文通过对 2019-nCoV 爆发的前期数据（第一例患者确诊到武汉封城后的一个星期）进行深度学习，不断平衡权重，以此来确定 2019-nCoV 在人群中每天的传播率，这将被用于动态 SEIR 模型中。

(2) 在传统的 SEIR 模型中，传染率  $\beta$  是一个常数，但是针对 2019-nCoV 病毒，感染率  $\beta$  是会随着时间不断变化的，因为早期由于公众意识不足和病毒的变异，在疫情前期感染率  $\beta$  会增加，随着疫情的不断爆发，国家出台严格的政策、公众认识到疫情的严重性等，此时感染率  $\beta$  会不断减少，所以本文假设  $\beta$  遵循 logistic 函数的斜率。

$$\beta(t) = \frac{C \cdot e^{-\partial t + \Delta b}}{(1 + e^{-\partial t + \Delta b})^2}$$

通过它本文可以得到每天  $\beta$  的估计值，这将被用于动态 SEIR 模型中。

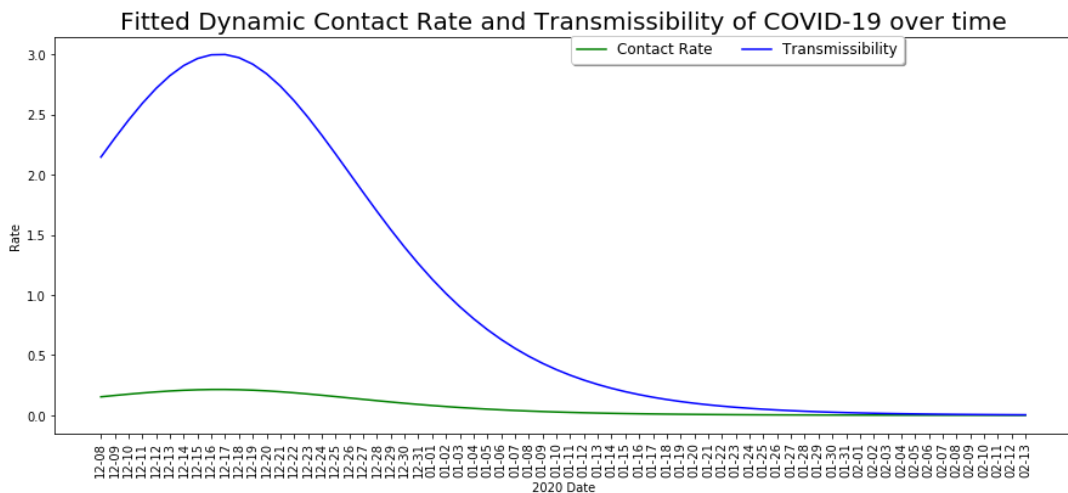


图 3 新型冠状病毒发展过程中的接触率和感染能力( $R_0$ )

图 3 中，Transmissibility 表示的是  $R_0$  基础在传播人数 ( $R_0 = \beta/\gamma$ )， $R_0$  大约在 2019 年 12 月 17 日到达峰值，约为 3.17（封闭华南海鲜市场的前 5 天，这时疫

情在武汉当地已经引起重视), 在二月初降至 0.5 以下, 封城、隔离使得人群不便接触, 切断了病毒传播的主要方式, 建设雷神山、火神山两座医院等措施效果显著。

2.5 模型实现

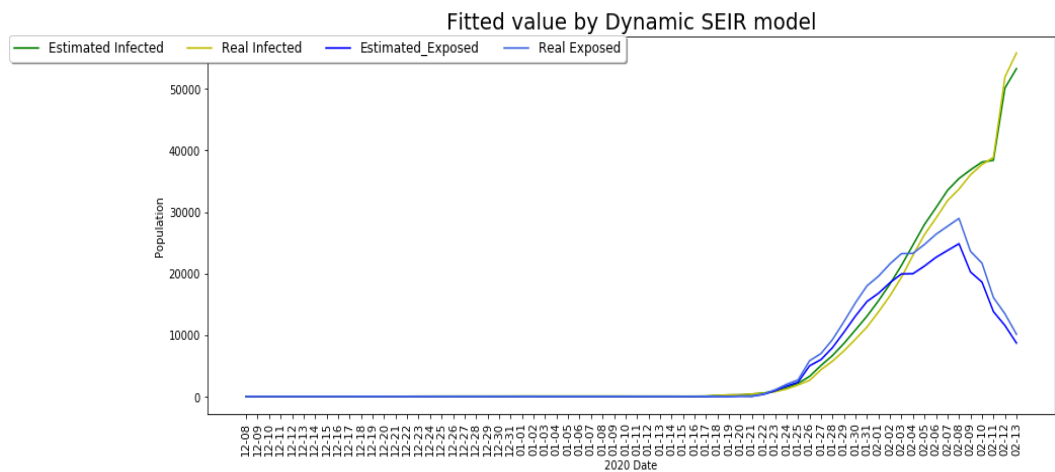


图 4 模型对于 2020-02-13 之前数据的拟合

图 4 中表达的是预测的潜伏患者与感染患者与现实中处于两种状态人的对比, 可以看出模型对于数据的拟合效果较好, 主要体现在模型预测值与实际官方的报告值趋势大体一致, 峰值出现的时间相差无几。

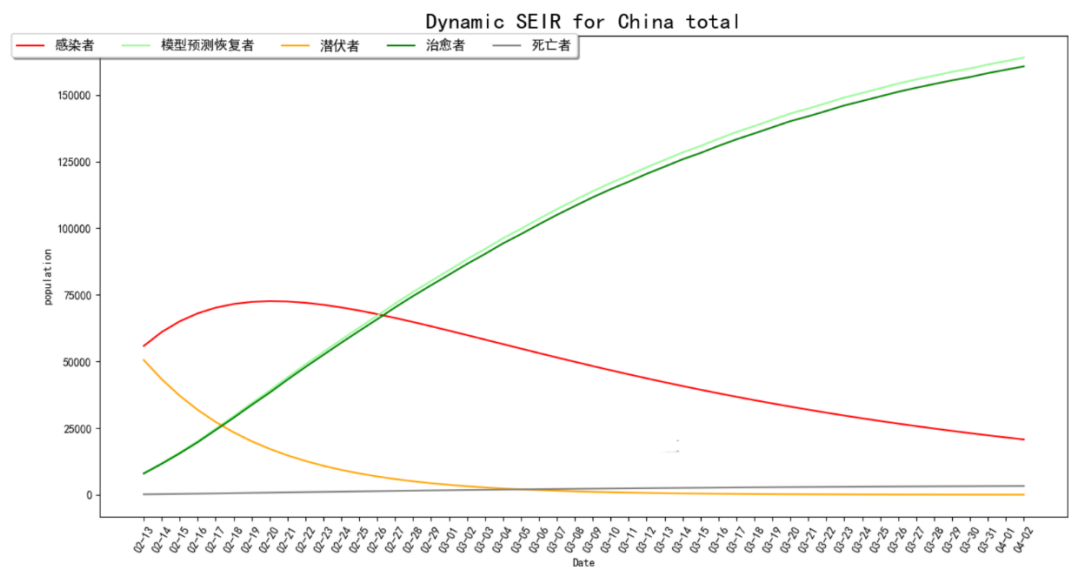


图 3 2020 年 2 月 13 日之后的疫情发展情况预测

从图 5 可以看出, 在 2020 年 2 月 20 日左右, 全国确诊人数将在 75000 例左右, 估计确诊人数峰值将在 2020 年 2 月 25 日之前到来, 到四月初全国的总感染人数将会降至 20000 以下。

### 3 灵敏度分析

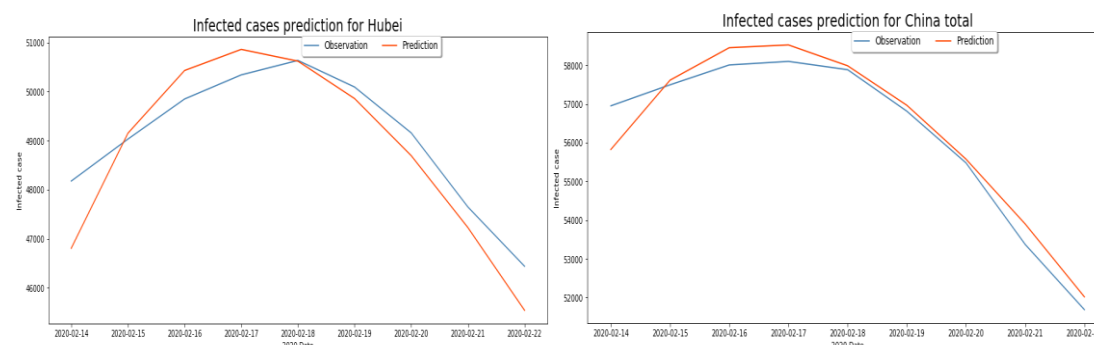


图 4 模型预测湖北以及全国确诊人数与实际人数之间的对比

根据图 6，从总体来看，通过模型预测的峰值时间和总体趋势与实际数据没有太大出入，为了使模型根据有说服力，本文使用平均绝对百分误差（MAPE）来衡量模型与实际数据之间的差异，运算公式为：

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_{ir} - y_{ip}|}{|y_{ir}|}$$

根据图 6 中反映出的日期的数据求得  $MAPE \approx 0.0067$

### 4 模型局限性

首先，预测对政策(出行限制、强制人们呆在家里、武汉为 2019-nCoV 新建 3 所医院等)高度敏感，这些政策都对时间线有很大的影响。本文对未来的预测没有将 2020 年 2 月 13 日之后出台的新政策产生的影响考虑进去。

其次，很难获得实时的正确信息(中国官方数据可能被低估)，尤其是武汉地区。

第三，本文对这种新疾病了解有限。例如，本文没有正确地测试所有 2019-nCoV 患者(症状不明，例如：2019-nCoV 患者无症状是否会传播感染)。

第四，在 SEIR 模型中，潜伏期的人前期不具有传染性。然而新型冠状病毒肺炎感染者在潜伏期就有较高的传染率。

第五，对病毒本身生命周期的考虑不足，同时未考虑天气回暖等因素对病毒传染性的影响。

### 5 主要结论(根据 2019-12-08 至 2020-02-13 官方数据)

- (1) 确诊患者人数峰值有望在 2 月 25 日之前到来（大致会在 2020 年 2 月 20 日到达峰值），预测峰值会在 75000 左右。
- (2) 目前传播速率已经有效得到控制，从最初的  $R_0 > 3$  降至 0.5 以下。
- (3) 四月初全国的总体感染人数会下降到 20000 以下，政府干预措施效果显



著，预计疫情将于四月中下旬得到良好控制。

## 6 评价

- (1) 对模型效果：本文采用 2019 年 12 月 8 日（武汉首例新冠肺炎患者确诊）到 2020 年 2 月 13 日之间的数据作为训练模型的数据，整个时间段模型的预测值和实际值之间的  $MAPE \approx 0.079$ ，预测水平较高。但是通过模型预测得到感染人数的峰值略低于 75000，死亡人数不断增加大致为 3000 例左右，而实际情况截至 2020 年 2 月 28 日，全国累计有 78961 例确诊病例，死亡病例 2791 例，通过模型的得到的峰值日期较早，人数较少，本文认为这与最近的复工、公众思想懈怠有关，仅从疫情角度来看，大规模的复工，公共场所聚众势必会造成人员接触率增高，感染病毒的风险提高等，因此造成了确诊数量仍不断上升，峰值延期。对此，本文认为复工复产是必须的，但与此同时国家应该做好协调政策，比如分期复工复产，做好员工每日的身体检查，及时消毒等，公众更要做好防护，虽然确诊病例每天的新增速率下降，但是疫情时有反复，公众仍应减少外出，做好居家隔离。
- (2) 对模型未来的展望：首先，本模型采用了动态参数，通过对 logistic 函数进行求导来反映  $R_0$  的动态变化，为其他动态参数的研究提供了一种思路。其次，本模型考虑到了将确诊患者分为无症状和有症状两种，对于其他会造成不同症状的传染病也有一定的意义。最后，本模型对预测区间政策的影响效果没有考虑，导致了模型存在一定的偏差，而且，疫情形势变化很快，想要做出一个好的模型需要考虑多方面因素，大致分为：病毒本身特性的周期性变化，公众对于疫情的了解情况以及重视程度，国家对于疫情防控的政策三个方面，模型对每个方面的因素都很敏感，因此模型的建立还应该考虑更多的动态因素来更好的模型现实情况

## 参考文献

- [1]张海峰,张文耀,孙桂全,等.信息的滞后性诱导传染病的周期爆发[J].中国科学(物理学 力学 天文学),2012,42(6):631-638.  
DOI:10.1360/132011-240.
- [2] Tianmu Chen, Qiupeng Wang ,Zeyu Zhao. A mathematical model for simulating the transmission of Wuhan novel Coronavirus[DB/OL]. BioRxiv. DOI: <https://doi.org/10.1101/2020.01.19.911669>
- [3] Julien Riou, Christian L. Althaus. Pattern of early human-to-human transmission of Wuhan 2019-nCoV[DB/OL]. BioRxiv. DOI: <https://doi.org/10.1101/2020.01.23.917351>
- [4] Jonathan M Read, View ORCID ProfileJessica RE Bridgen, Derek AT Cummings. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions[DB/OL].MedRxiv. DOI: <https://doi.org/10.1101/2020.01.23.20018549>

- [5] Natsuko Imai, Ilaria Dorigatti, Anne Cori. Estimating the potential total number of novel Coronavirus cases in Wuhan City, China[N]. Imperial College London, UK. 17 January 2020
- [6] Victor M Corman, Olfert Landt. Detection of 2019 novel coronavirus(2019-nCov)by real-time RT-PCR[J]. Eurosurveillance, 25, 2000045 (2020), <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>
- [7]Yiran Jing. Covid-19 estimation and forecast using statistical model [DB/OL].  
<https://github.com/YiranJing/Coronavirus-Epidemic-2019-nCov>
- [8] Shi Zhao, Qianyin Lin, Jinjun Ran, Salihu S Musa, Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak[J]. International Journal of Infectious Diseases doi: 10.1016/j.ijid.2020.01.050