



**北京航空航天大学**  
B E I H A N G U N I V E R S I T Y

**第三十届“冯如杯”学生学术科技作品竞赛项目论文**

**新型冠状病毒肺炎疫情预测与评估**

——基于 SIR 模型

2020 年 4 月

## 摘要

自 2019 年 12 月至 2020 年 4 月，受武汉新型冠状病毒影响，全国各地陆续限制交通，限制公共场所开放，累计八万余人确诊新型冠状病毒肺炎，对社会的发展生产造成极大阻碍。本文希望建立在现代流行病管理体系下的流行病模型，对新型冠状病毒肺炎在湖北省的发展情况用数学模型进行描述并根据在其中发现的问题提出相关建议。

本文将湖北疫情发展分为 2019-nCoV 病毒自由传播阶段与 2019-nCoV 病毒疫情管控隔离两个阶段，并对阶段 II 进行细致分析。阶段 I，新型冠状病毒肺炎疫情自由传播阶段，从 2019 年 12 月 31 日至 2020 年 1 月 22 日；阶段 II，新型冠状病毒肺炎疫情隔离管控阶段，从 2020 年 1 月 23 日湖北封城至 2020 年 4 月 3 日，视为湖北新型冠状病毒肺炎疫情隔离管控阶段。

本文对于阶段 II，考虑到新型冠状病毒肺炎传染性强、潜伏期长；政府管控力度高等方面的因素，对经典的 SIR 模型进行修正，建立起综合考虑发病延迟、模型参数随时间改变的动态 SIR 模型。

本文通过建立 II 阶段湖北新型冠状病毒肺炎的动态发展模型。加深人们对流行性传染病的理解，并为疫情防控提出有效建议。

**关键词：**新型冠状病毒，新型冠状病毒肺炎，预测，SIR 模型，有效疫情防控

## Abstract

Suffering the adverse impact of 2019 novel coronavirus, from December,2019 to April,2020, governments restrained transport among cities, and the opening of public places. Until April,3rd,2020, the quantity of cumulative confirmed cases has exceeded 80 thousand cases, which causes continuous impediment towards social production and development. This article aims at establishing an epidemic model under modern containing system to describe the development of the epidemic in Hubei Province.

The development of 2019-nCoV in Hubei Province is divided into two phases and analyze phase II delicately: Phase I, from December 31st,2019 to January 22rd, 2020, which is regarded as the free transmission phase of the epidemics; Phase II, from January 23th, 2020 (the time when Hubei Province launched the policy to lockdown cities)to April 3rd, which is considered to be the phase of isolation and government's containment.

For Phase I, considering public's lacking awareness towards virus disease and government's shortage of effective containment. This phase can be regarded as virus's free transmission period, which can use Logistic model to describe.

For Phase II, considering 2019-nCoV's infectivity and long incubation and government's complete and effective containments, this article modifies typical SIR model and establish a dynamic SIR model which comprehends dynamic parameters and the delay of confirming cases.

This article established the dynamic model for each phase, aiming to enhance public's comprehension towards epidemics and offer effective instruction for epidemic prevention and containment.

### **Keywords**

**2019-nCoV, Corona Virus Disease, prediction, SIR model, effective epidemic prevention and containment**

# 目录

摘要.....	i
Abstract .....	ii
目录.....	iii
第一章 绪论.....	1
1.1 项目背景 .....	1
1.1.1 新型冠状病毒肺炎流行特征介绍 .....	1
1.1.2 新型冠状病毒肺炎疫情过程中湖北省疫情防控关键时间 .....	1
1.1.3 经典 SIR 模型简介 .....	3
1.2 项目思路 .....	3
第二章 模型设计.....	4
2.1 第 II 阶段的 SIR 动态模型 .....	4
2.2 梯度下降方法求模型最优参数 .....	6
第三章 实证分析.....	8
3.1 对重要时间段进行疫情分析预测，并给出防控建议 .....	8
3.2 从 1 月 23 日开始至 3 月 6 日的预测曲线 .....	12
3.2.1 每日确诊数量预测及批判性分析 .....	12
3.2.2 模型参数的预测曲线及分析 .....	13
总结与展望.....	16
参考文献.....	17
附录 A.....	20

# 第一章 绪论

## 1.1 项目背景

自 2019 年 12 月至 2020 年 4 月，受武汉新型冠状病毒影响，全国各地陆续限制交通，限制公共场所开放，累计八万余人确诊新型冠状病毒肺炎，对社会的发展生产造成极大阻碍。本文希望对经典的 SIR 模型进行修正，以建立在现代流行病管理体系下的流行病模型，对新型冠状病毒肺炎在湖北省的发展情况用数学模型进行描述。下面分别对新型冠状病毒肺炎流行特征、湖北省疫情防控关键点、经典 SIR 模型进行介绍。

### 1.1.1 新型冠状病毒肺炎流行特征介绍

1.学名与病原体：2019-nCoV 感染导致的肺炎被命名为新型冠状病毒肺炎（Corona Virus Disease, COVID-19）<sup>[1]</sup>。（下文中称将该传染病称为 COVID-19），其病原体为新型冠状病毒。（下文中称该病毒为 2019-nCoV）

2.主要传染源：在疫情持续一段时间后，主要传染源为 COVID-19 感染者与 COVID-19 潜伏者。

3.易感人群：COVID-19 为新发传染病，人群普遍缺乏免疫能力。

4.主要传播方式：2019-nCoV 主要经飞沫传播和接触传播。此外，2019-nCoV 也可以通过暴露于空气中的眼结膜与口腔黏膜传播<sup>[2]</sup>。

5.潜伏期：COVID-19 的潜伏期平均为 5.8 天<sup>[6]</sup>，短潜伏期患者在感染 2-3 天后发病，长潜伏期患者一般在 10-12 天后发病<sup>[3]</sup>。

### 1.1.2 新型冠状病毒肺炎疫情过程中湖北省疫情防控关键时间

2020 年 1 月 5 日，经武汉市政府确认，最早一例的新型冠状病毒肺炎发病时间为 2019 年 12 月 12 日<sup>[7]</sup>。

2020 年 1 月 23 日 10 时起，武汉全市公交、地铁、轮渡、长途客运暂停运营；市民无特殊原因不得离开武汉<sup>[4]</sup>。

2020 年 3 月 25 日 0 时起，湖北省武汉市以外地区解除道路管控，有序恢复湖北与其它省份之间的交通<sup>[5]</sup>。

2020 年 4 月 8 日 0 时起，武汉市解除道路管控，并有序恢复与其它地区之间的交通<sup>[5]</sup>。

2020 年 2 月 27 日，湖北省疫情基本得到控制，此日以后每日新增确诊患者小于 200 人并不断减少。下面两幅图展示了 2020 年 1 月 12 日以来湖北省每日新增病例人数。由于 2 月 3 日~2 月 4 日新增病例为 0 例，而 2 月 4 日~2 月 5 日新增病例陡增至 6143 例；2 月 11 日~2 月 12 日新增病例为 0 例，而 2 月 12 日~2 月 13 日新增病例陡增为 14840 例。因此，在绘制折线图时将这 4 个特殊日期排除，以更好地表现疫情发展趋势。

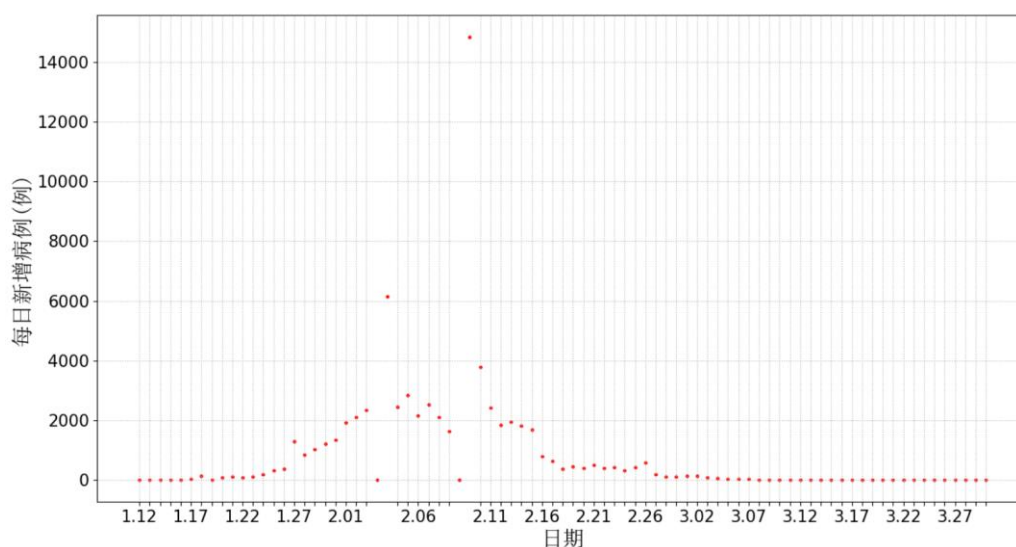


图 1 1 月 12 日以来湖北省每日 COVID-19 新增病例人数散点图

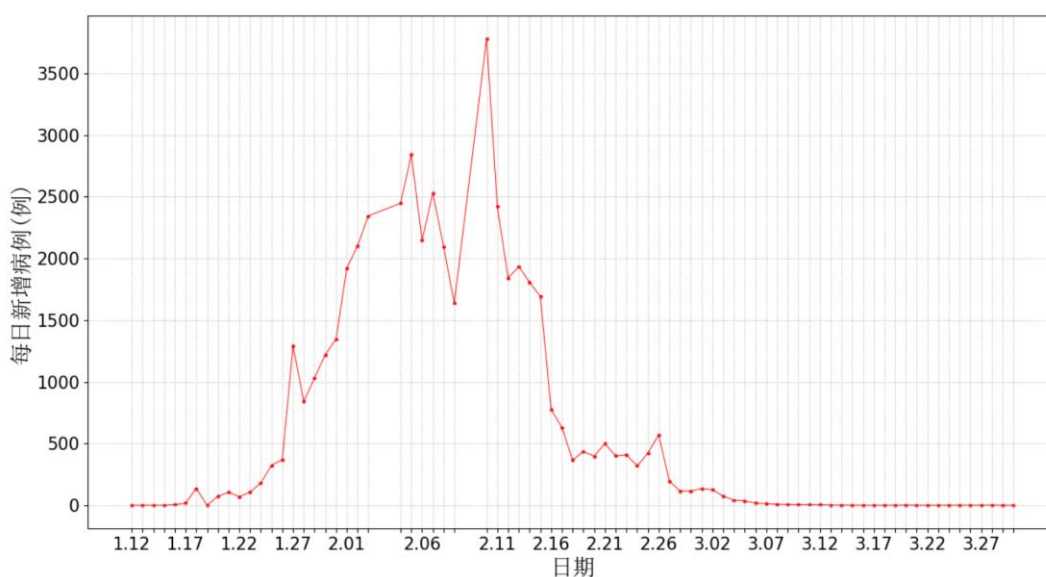


图 2 1 月 12 日以来湖北省每日 COVID-19 新增病例人数折线图

### 1.1.3 经典 SIR 模型简介

SIR 模型是一种传染病动力学模型，最先由 Kermack 与 McKendrick 在研究伦敦黑死病时提出。SIR 模型的 3 个假设条件为：

1.一定范围内的人群可被分类为易感染者(Susceptible)、染病者(Infective)、移出者(Removed)。其中易感染者不会传播病毒，但有可能成为染病者；染病者携带病毒，并且会将病毒传染给其他人；移出者为死亡或者被治愈的人，不再被感染，且不会再患病。

2.在研究时间段内没有新增人口、死亡人口、人口流动等动力学因素。人口总数  $N$  始终保持一个常数，也即：

$$N(t) \equiv K \quad (1)$$

若将  $t$  时刻易感染者、染病者、移出者数量分别记为  $S(t)$ 、 $I(t)$ 、 $R(t)$ ，得：

$$\begin{cases} S(t) + I(t) + R(t) = N(t) \\ N(t) \equiv K \end{cases} \quad (2)$$

3.从  $t$  时刻，单位时间内一个染病者能够传染的易感染者数量与该范围内的易感染者总数成正比，该比例称为感染系数为  $\beta$ ；从染病者中移出的人数与染病者数量成正比，该比例称为恢复系数为  $\gamma$ 。根据上述关系列出方程：

$$\begin{cases} dI(t) = \beta I(t)S(t)dt - \gamma I(t)dt \\ dR(t) = \gamma I(t)dt \end{cases} \quad (3)$$

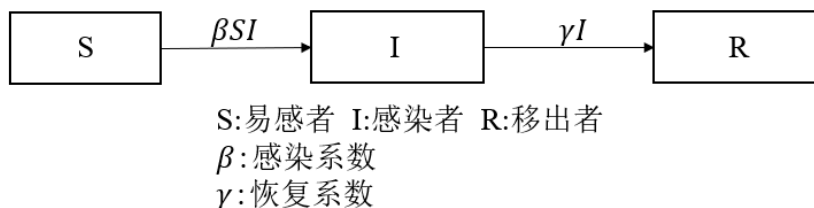


图 3 经典的 SIR 模型示意图

## 1.2 项目思路

本文将湖北疫情发展分为 2019-nCoV 病毒自由传播阶段与 2019-nCoV 病毒疫情管控隔离两个阶段。阶段 I，2019-nCoV 疫情自由传播阶段，从 2019 年 12 月 12 日至 2020 年 1 月 22 日；阶段 II，2019-nCoV 疫情隔离管控阶段，从 2020 年 1 月 23 日湖北封城至 2020 年 4 月 3 日，视为湖北 2019-nCoV 疫情隔离管控

阶段。本文专门对第 II 阶段进行研究。

湖北 COVID-19 疫情的第 II 阶段与以往如 SARS 等大型传染病相比，在流行特征上存在明显区别；政府管控力度更强，湖北省各市政府采取了最高级防控措施——封城，确诊者一经发现立刻进入隔离诊治状态。而随着疫情的发展，防控力度不断加大；群众防范意识提升，绝大多数群众佩戴口罩，遏制病毒的传播。这些特征都是建立模型时不可被忽略的因素。本文根据上述特征对经典的 SIR 模型进行修正。根据不同因素对模型所进行的修正如下表所示。

**表 1 COVID-19 疫情重要特征与建模需考虑的因素对应表**

特征	建立模型时考虑的因素
潜伏期长	确诊者数量增加需要考虑延迟效应
潜伏期患者仍会散播病毒	在传染期内易被感染的人群数量增加
确诊者被迅速隔离，无法再散播病毒	确诊者不再是传染源的一部分
湖北省各城市采取“封城”措施	假设湖北省与其它省份之间没有人口流动
政府防控力度不断加大，群众防范意识不断提升	SIR 模型参数随时间推移而变化

## 第二章 模型设计

### 2.1 第 II 阶段的 SIR 动态模型

根据表 1，并参考以往的研究<sup>[10~13]</sup>，我们对 SIR 模型进行修正得到 SIR 动态模型，描述如下：

1. 将这次疫情中的湖北民众分为，易感者(Susceptible)，潜伏携带者(Incubation carrier)，确诊者(Confirmed)。

**表 2 易感者、潜伏携带者、确诊者的名词解释**

类别	解释
易感者	不散播病毒但曾进入被感染范围却没有充分防护的人群
潜伏携带者	携带并散播病毒，但尚未被确诊隔离的人群
确诊者	携带病毒，但因被确诊隔离而不再散播病毒的人群

附注 1：从传染病学角度，虽然人群对 2019-nCoV 普遍易感，但并不是所有人都是易感



者。除去潜伏携带者与确诊者，仍一部分人对于疫情由相当完备的防护措施，不容易被感染。但这部分人只要有一次松懈，让自己在没有充分防护的时候暴露于潜伏携带者散播病毒的范围之中，那么将成为易感者。

2.为了方便描述，我们将“第  $n$  天”定义为 2020 年 1 月 23 日以来第  $m$  天以后的第  $n$  天。(例如 2020 年 1 月 31 日为 2020 年 1 月 23 日以来第 4 天以后的第 4 天。此时  $m=4$ ,  $n=4$ )第  $n$  天时易感者、潜伏者、确诊者数量分别记为  $S_n$ 、 $I_n$ 、 $C_n$ 。病毒潜伏期为  $\tau$  天。

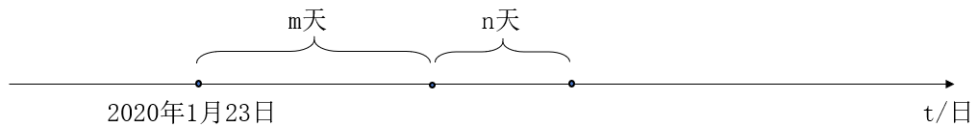


图 3 对“第  $n$  天”定义的解释

3.潜伏携带者在其潜伏期内出入各种公共场所，使得没有充分防护准备的人暴露于易被病毒感染环境之中，成为易感者。在第  $n$  天，记每位潜伏携带者在其潜伏期时间内平均使  $k(n)$  人成为易感者。 $k(n)$  称作易感再生数。

4.易感者曾暴露在被感染范围内，有可能成为潜伏携带者。在第  $n$  天，设每位易感者成为潜伏携带者的概率为  $u_1(n)$ 。 $u_1(n)$  称作感染强度； $k(n)u_1(n)$  则称作有效感染数。

由条件 2、条件 3：

$$S_{n+1} - S_{n-\tau} = k(n)(C_{n+1} - C_n) - u_1(n)S_n \quad (4)$$

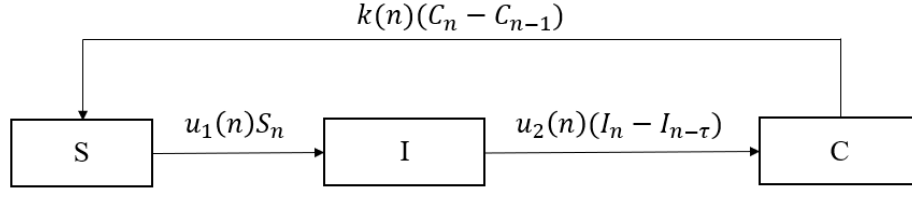
5.潜伏携带者在潜伏期内发病，或在发病前直接被检测出携带 2019-nCoV 而被确诊隔离，成为确诊者。在第  $n$  天，设每位潜伏携带者在潜伏期  $\tau$  天内发病率为  $u_2(n)$ 。 $u_2(n)$  称为潜伏发病率。

由条件 5：

$$I_{n+1} - I_n = u_1(n)S_n - u_2(n)(I_n - I_{n-\tau}) \quad (5)$$

6.每位确诊者的数据将被记录，且每日湖北省政府将公开累计确诊者数目。将  $C_{n+1} - C_n$  称为每日新增确诊病例数。

下图为 SIR 动态模型示意图，说明了上述条件中各个变量之间的关系。



S:易感者 I:潜伏感染者 R:确诊者  
 $k$ :易感再生数  
 $u_1$ :感染强度  
 $u_2$ :潜伏发病率

图 4 动态 SIR 模型示意图

基于上述描述，每日模型迭代过程中，SIR 动态模型各部分关系如下：

$$\begin{cases} S_{n+1} - S_{n-\tau} = k(n)(C_{n+1} - C_n) - u_1(n)S_n \\ I_{n+1} - I_n = u_1(n)S_n - u_2(n)(I_n - I_{n-\tau}) \\ C_{n+1} - C_n = u_2(n)(I_n - I_{n-\tau}) \end{cases} \quad (6)$$

为求解（6）式，需要知道 $(k, u_1, u_2)$ 三个模型参数，以及 $C_n$ 、 $S_n$ 的初值。其中， $C_0$ 及第 0 天前的确诊数可根据湖北省新型冠状病毒感染肺炎疫情防控指挥部提供的数据获知，而 $k$ 、 $u_1$ 、 $u_2$ 与 $S_n$ 的初值则无法直接得知，可以将其作为模型参数。

$$C_n = D_{k,u_1,u_2,S_1}(n) \quad (7)$$

## 2.2 梯度下降方法求模型最优参数

为求解（6）式，需要知道 $(k, u_1, u_2)$ 三个模型参数，以及 $C_n$ 、 $S_n$ 的初值。其中， $C_0$ 可根据湖北省新型冠状病毒感染肺炎疫情防控指挥部提供的数据获知，而三个模型参数与 $S_n$ 的初值则无法直接得知。

本文通过给定向量 $(k, u_1, u_2, S_0, S_{-6}, I_0, I_{-6})$ 初始值 $\theta_0$ ，通过定义损失函数将求解结果与实际数据进行比较，再利用梯度下降算法，对模型参数进行多次迭代后找出最符合实际情况的模型参数。

本文采用基于 Adabound 自动调参方法的随机梯度下降方法来找到动态 SIR 模型的最优参数<sup>[8]</sup>。

---

### Algorithm 1 SGD

---

**Input:** loss function  $J(\theta, n)$ , initial parameter vector  $\theta_0$ , initial step size  $\alpha$ ,  $\beta_1$ ,  $\beta_2$ , lower bound function  $\eta_l$ , upper bound function  $\eta_u$ , iteration times  $Titer$ , sample number  $m$

```

1: Set  $m_0=0$ ,  $v_0=0$ 
2: function decreaseDirection (current iteration times  $t$ , gradient  $\mathbf{g}_t$ )
3:    $\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{g}_t$ 
4:    $\widehat{\mathbf{m}}_t = \frac{\mathbf{m}_t}{1 - \beta_1^t}$ 
5:   return  $\widehat{\mathbf{m}}_t$ 
6: function amendmentFactor (current iteration times  $t$ , gradient  $\mathbf{g}_t$ )
7:    $v_t = \beta_2 v_{t-1} + (1 - \beta_2) \|\mathbf{g}_t\|^2$ 
8:    $\widehat{v}_t = \frac{v_t}{1 - \beta_2^t}$ 
9:   return  $\frac{\alpha}{\sqrt{\widehat{v}_t}}$ 
10: function Clip ( current lower bound  $\eta_l$ , current upper bound  $\eta_u$ , learningRate  $\alpha$ )
11:   if  $\alpha < \eta_l$  then
12:     return  $\eta_l$ 
13:   else if  $\alpha < \eta_u$  then
14:     return  $\eta_u$ 
15:   else
16:     return  $\alpha$ 
17: for  $i = 1$  to  $Titer$  do
18:    $X\_new, y\_new = \text{RANDOMSORT}(X, y)$ 
19:   #RANDOMSORT( $X, y$ ) is to sort original sample randomly.
20:   for  $j = 1$  to  $m$  do
21:      $t = (i - 1) * m + j$ 
22:      $n = t$ 
23:     #RANDINT( $1, m$ ) is to generate an integer from  $[1, m]$ 
24:      $\mathbf{g}_t = \nabla J(\theta_t, n)$ 

```

- 25:  $m_t = \text{decreaseDirection}(t, g_t)$
- 26:  $\alpha_t = \text{Clip}(\eta_l(t), \eta_u(t), \text{amendmentFactor}(t, g_t))$
- 27:  $\theta_{t+1} = \theta_t - \alpha_t m_t$

$$J(\theta, n) = (\ln(1 + D_\theta(n)) - \ln(1 + a_n))^2 \quad (8)$$

$$\frac{\partial J(\theta, n)}{\partial \theta_j} = 2 \times \frac{\partial D_\theta(n)}{\partial \theta_j} \times \frac{\ln(1 + D_\theta(n)) - \ln(1 + a_n)}{D_\theta(n) + 1} \quad (9)$$

### 第三章 实证分析

#### 3.1 对重要时间段进行疫情分析预测，并给出防控建议

下面我们将用上述模型对几个时间段的易感再生数，潜伏发病率，感染强度，及相应感染人数初始值进行计算与分析，并对相应时间段内的确诊人数进行估计。同时，在验证预测结果准确性后本文将为每一时间段的疫情防控提供相关建议。

我们采用均方根对数误差 RMSLE 计算模型预测结果精度。

$$RMSLE = \sqrt{\frac{1}{n} \times \sum_{i=1}^n (\ln(p_i + 1) - \ln(a_i + 1))^2} \quad (10)$$

下面我们将对几个时间段进行预测并进行相应的分析：

(1) 2020 年 1 月 23 日~2020 年 1 月 29 日：下图给出了该时间段中疫情实际数据（蓝色）与疫情预测数据（红色）的曲线以及在每一天的相应值。

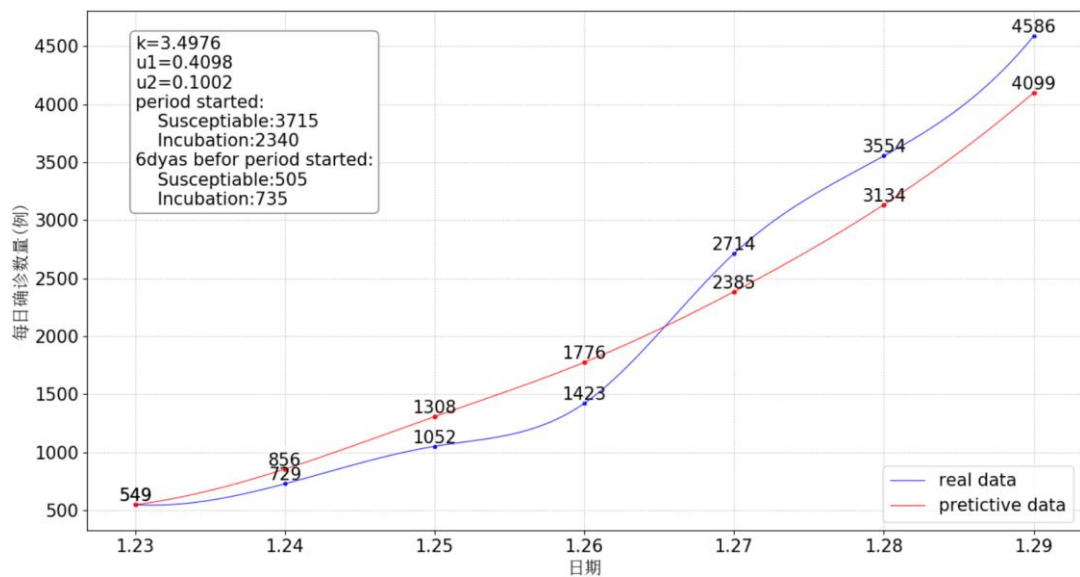


图 5 2020-01-23 至 2020-01-29 相关参数与疫情预测

**预测结果准确性分析：**

考虑到研究对象是一整个湖北省（约 5927 万人<sup>[9]</sup>），而最高偏差却不足 500 人，经计算 RMSLE 值为 0.167，准确度较高。从曲线拟合情况上看，预测曲线与实际曲线基本吻合。

**表 3 2020-01-23 至 2020-01-29 相关参数与疫情预测**

模型参数	参数值
易感再生数 $k$	3.4976
感染强度 $u_1$	0.4098
潜伏发病率 $u_2$	0.1002
<b>1 月 23 日疫情数据</b>	<b>人数</b>
易感者数量	3715
潜伏携带者数量	2340
<b>1 月 17 日疫情数据</b>	<b>人数</b>
易感者数量	505
潜伏携带者数量	735

**重要参数分析：**

潜伏携带者数量：1 月 17 日至 1 月 23 日潜伏携带者数量从 735 人急速上升至 2340 人，可见这次病毒的传染性之强，而且这个趋势也与接下 1 月 23 日至 1 月 29 日确诊病例数量以千人/每日的速度急速上升吻合。

易感再生数：1 月 23 日至 1 月 29 日时间段易感再生数平均为 3.49 人，这意味着每一个感染病毒的人每天将导致 3.49 个人感染病毒，这会导致感染病毒的人数指数级上升。

**疫情分析及防控建议：**

1.病毒的传染性强，该阶段易感再生数为3.49人，从传染速度上看感染病毒的人数将以指数级上升。但应当注意到只有易感者有可能会感染病毒，如果能够控制易感者数量，那么即便病毒传染速度快，也只有有限的人会被感染，病毒感染者数量也不会以指数上升的速度增长。因此我们建议，广大群众防范应当提高防范意识，减少出行，及时佩戴口罩。

2.1 月 17 日确诊病例数仅 62 例，而此时根据预测已约有 735 人感染。可见若仅凭确诊病例数，想要在这次疫情在发展初期就采取相应有效措施是相对困难的。若想要从疫情发展之处就采取相应的措施，那么医院门诊在发现呼吸科，发热科人数突然出现异常时就应当及时上报并组织相关小组并及时开展对传染源的研究，认识病毒的流行特性。根据病毒的流行特性提出相关的建议，如果由该病毒引起的疾病可能会导致严重后果，应当及时上报。

3.部分国家政府“戴口罩对防止病毒传播没有实际作用”的论断是错误的，佩戴口罩有利于减少被病毒感染的概率，降低易感人群数量，能够限制病毒的流行传播。

(2) 2020 年 1 月 30 日~2020 年 2 月 6 日：下图给出了该时间段中疫情实际数据（蓝色）与疫情预测数据（红色）的曲线以及在每一天的相应值。

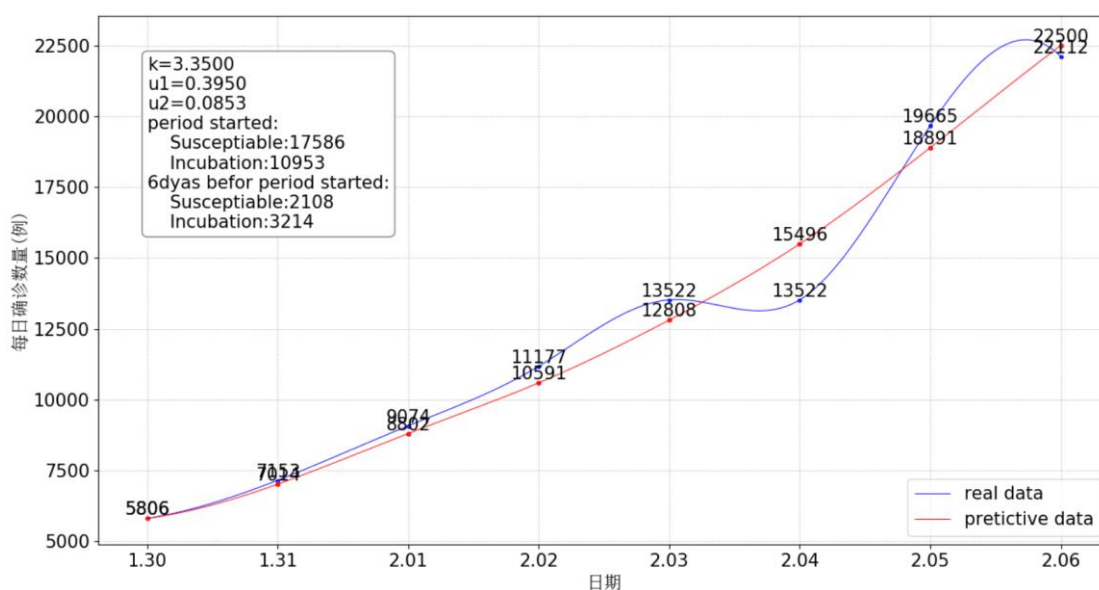


图 5 2020-01-30 至 2020-02-06 相关参数与疫情预测

#### 预测结果准确性分析：

该模型 RMSLE 值为 0.0627，准确度相当高。准确度与上一时间段(2020-01-23 至 2020-01-29)相比大幅度上升。这是由于 SIR 动态模型要求“无人口流动”的条件，而自 2020-01-23 武汉封城以来，人们外出活动减少，且湖北省各市陆续封城，不接受外省游客，这使得实际情况越来越符合 SIR 动态模型的假设。因此，在该时间段模型准确性很高，能够地很好描述疫情发展。

表 5 2020-01-30 至 2020-02-06 相关参数与疫情预测

模型参数	参数值
易感再生数 $k$	3.3500
感染强度 $u_1$	0.3950
潜伏发病率 $u_2$	0.0853
<b>1 月 23 日疫情数据</b>	<b>人数</b>
易感者数量	17586
潜伏携带者数量	10953
<b>1 月 17 日疫情数据</b>	<b>人数</b>
易感者数量	2108
潜伏携带者数量	3214

### 疫情分析及防控建议：

1. 由于该模型在该时间段的高准确性，本文对 2020 年 2 月 4 号湖北省政府公布的疫情确诊病例数据（13522 人）提出质疑：

首先给出 2020 年 2 月 03 日至 2020 年 2 月 6 日的预测如下图：

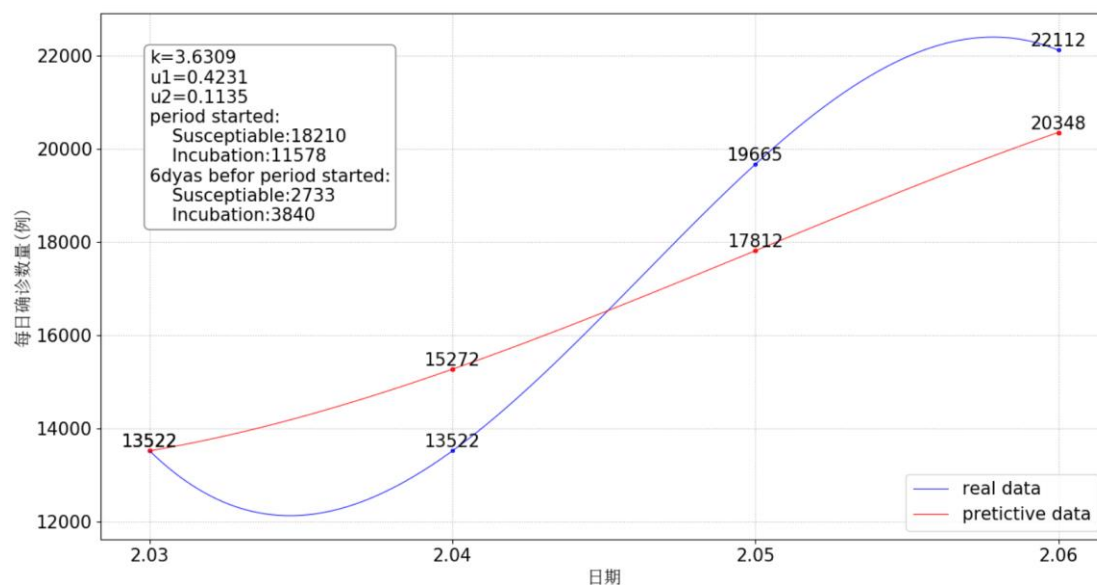


图 6 2020 年 2 月 03 日至 2020 年 2 月 06 日相关参数与疫情预测

让人怀疑的地方在于，2020 年 02 月 04 日的新增确诊病例数为 0 而仅过一天确诊病例人数就激增 6152 人，为 2 月 5 日至 2 月 6 日增加人数的两倍。基于这样的不平衡，本文对此数据的真实性提出质疑。并预估 2020 年 02 月 04 日累计确诊病例数的真实值在 15272 例左右。

因此，我们建议政府应当及时公布疫情信息，并尽可能确保信息的准确度，以免产生错误的群众引导。

## 3.2 从 1 月 23 日开始至 3 月 6 日的预测曲线

### 3.2.1 每日确诊数量预测及批判性分析

下面给出 2020 年 1 月 23 日至 2020 年 3 月 5 日，利用动态 SIR 模型计算得到的每日确诊数量预测曲线（横坐标为 0 代表 2020 年 1 月 23 日，以后横坐标每增加 1 代表过去 1 天）由于数据密度较大，在图中不标注具体的值；具体预测数据将在附录 A 中给出。

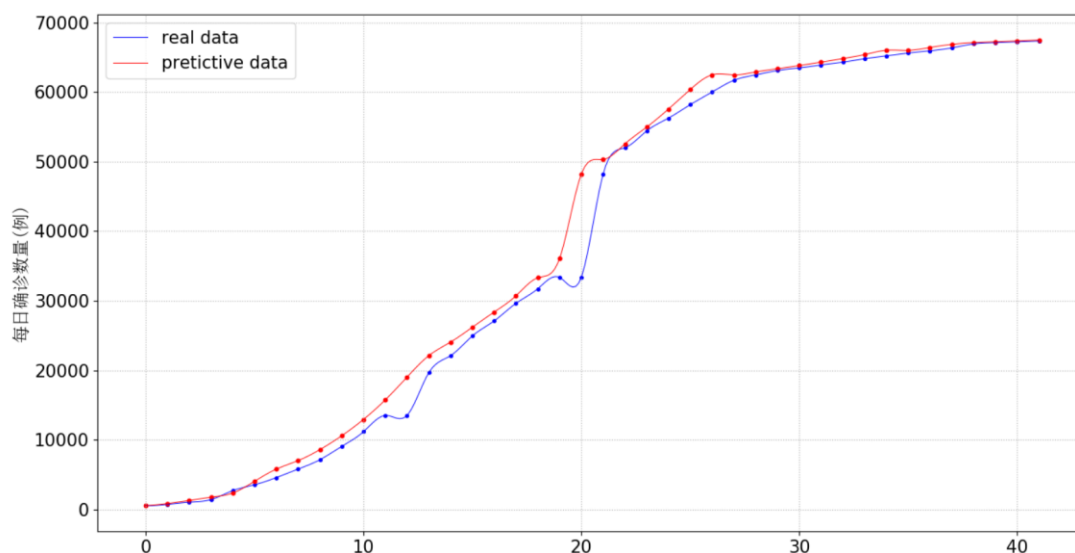


图 7 2020 年 1 月 23 日至 2020 年 3 月 5 日每日确诊数量预测曲线

从曲线整体上看拟合情况是相当不错的，RMSLE 值为 0.1249。而实际上在具体计算的过程中，部分时间段的 RMSLE 的准确度远高于整体情况。下面给出不同时间段的 RMSLE 值

表 6 在预测过程中划分的不同时间段的 RMSLE 值

时间段	RMSLE
2020-01-23 至 2020-01-29	0.167162
2020-01-30 至 2020-02-05	0.073320
2020-02-06 至 2020-02-12	0.043555
2020-02-13 至 2020-02-19	0.025000
2020-02-20 至 2020-02-26	0.001936



2020-02-27 至 2020-03-01	0.000973
2020-03-02 至 2020-03-05	0.000055

可以看出模型精度在不断升高,这说明人口流动不断减小使得实际情况与模型假设逐渐符合。而这也与 SIR 动态模型的延迟因素密切相关,随着时间的推移,潜伏感染者数量不断减少,确诊人数增长量也就不断减小;这个过程使得实际确诊数量增长较为缓和,适合进行较为准确的模型预测。

同时观察预测曲线与实际曲线,我们发现一般情况下模型预测值大于实际值。但是,这不能简单理解为无法在每一时间准确上报确诊人数(虽然这也是导致这个结果的因素之一)。应当注意到,动态 SIR 模型中一直有相当数量的潜伏感染者与易感者,而使得易感者数量减少的唯一方式是“易感者变成潜伏感染者”,这与实际情况是相违背的,所以这也是该模型在设计之处没有考虑到的因素之一。基于此,我们应当再增加一个有关于“易感者退出易感人群进入不易感人群”的参数(这里再次强调本文对易感者的定义,见表 2)

### 3.2.2 模型参数的预测曲线及分析

下图给出利用动态 SIR 模型得到的不同阶段的模型参数预估值,其中 2020-01-23 点表示 1 月 23 日至 1 月 30 日的预估值,2020-01-30 点表示 1 月 30 日至 2 月 06 日的预估值,以此类推。

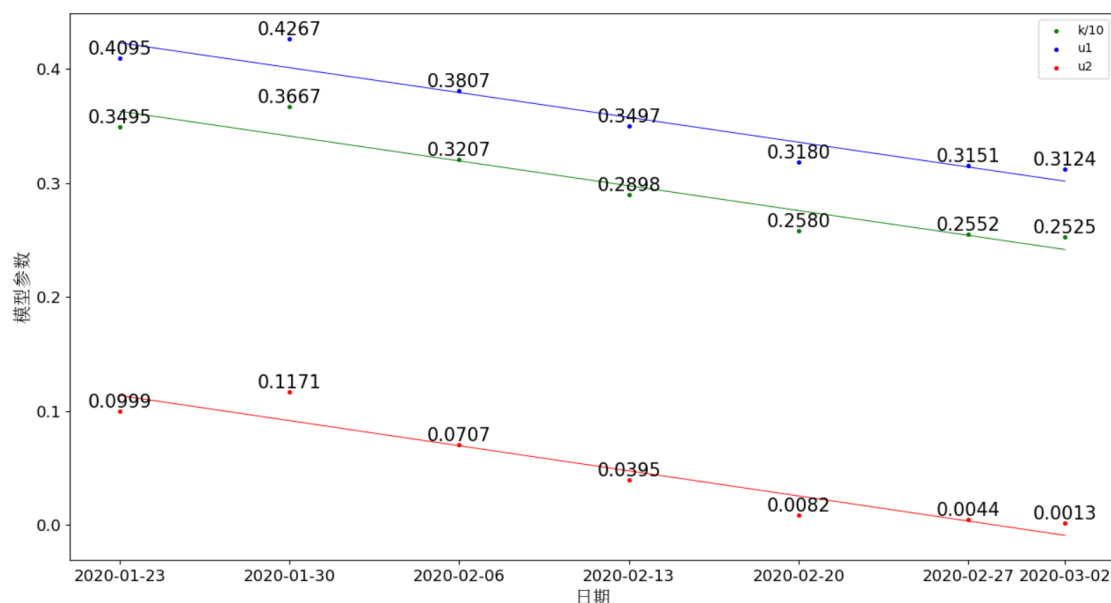


图 8 各个时间段的模型参数预估

附注 2: 为了更清晰的显示参数,将参数 $k$ 缩小为 $k/10$ 后再进行描点

经过反复尝试，本文考虑之间用线性回归方法进行拟合。且三条直线的线性相关系数都在 0.9 以上。

表 7 用直线方程 $y = ax + b$ 拟合的模型参数

参数名	系数 $a$	系数 $b$
$u_1$	0.42288	-0.003111
$u_2$	0.1134	-0.003144
$k$	3.6286	-0.031095

下面对参数的变化进行分析：

1. 从预测图像上可以看出 2020 年 1 月 30 日左右三个模型参数突然增加，本文在从数据角度对此现象进行细致无果后查阅多方新闻报道，发现 2020 年 1 月 27 日至 2020 年 1 月 28 日两天多地医疗工作者驰援湖北。这使得病毒检测能力大大提高，从而导致了明显的参数波动。

2. 有效再生数(易感再生数 $k$ 与感染强度 $u_1$ 的乘积)逐日下降并在 2020 年 02 月 13 日接近 1，并不断下降。这意味着在此之后疫情将得到有效控制。(有效再生数为 1 意味着每增加一个病患，那么只会有 1 个人会再被感染)

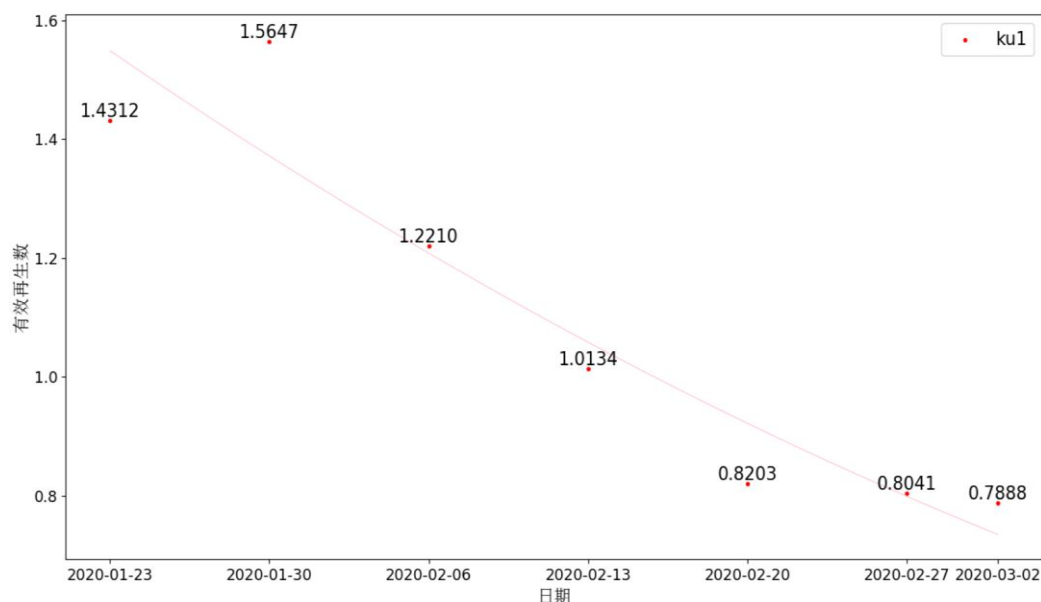


图 9 各个时间段的有效再生数

应当注意的是，“疫情得到有效控制”与每日新增确诊病例数为 0 不同。有效再生数为 0，仅仅说明新增病毒感染者数为 0。此时还有一定规模的潜伏感染者会成为确诊病人。但是可以确定的是，在 2020 年 02 月 13 日以后，潜伏感染

者数将不断减小并向 0 接近。

3.易感再生数逐日下降说明了政府的管控措施的有效。感染强度的下降则说明了人们的自我防范意识提高，外出活动时间与疫情刚开始相比减少约 40%。

## 总结与展望

本文基于动态 SIR 模型研究了 2019-nCoV 病毒在湖北省的传播情况(从 2019 年 12 月 31 日至 2020 年 4 月)。新型冠状病毒由以下 7 个变量决定:感染强度、潜伏发病率、易感再生数、初期潜伏感染者数、初期易感患者数、初期确诊患者数;其中,感染强度由病毒本身的传染性、人们的自我防范意识决定。易感再生数由政府的管控措施与潜伏感染者的人口流动决定。从实时性上看,模型参数根据湖北省政府的疫情实时公开数据得到,因而该模型能够反应不同阶段政府管控措施所带来的效果。从预测角度上看,该动力学模型本身就具有一定的不确定性,模型参数也是基于梯度下降算法得到的最优解,并不代表完全准确;而且随着政府的防控措施的推进与人们自我防护意识的提高,模型参数不是固定的。为了得到尽可能准确合理的数据,而是根据每日的新增数据构成一个又一个新的时间段,分段研究每一天的模型参数及其变化规律,得到动态的模型参数,使模型尽可能与实际情况接近,也对现实中的各种持续变化的因素有着更好的预测能力。

本文基于经典的 SIR 模型构建的动态 SIR 模型对疫情数据的预测结果有较高的准确度。而且,当大批医护人员驰援武汉时(2020-01-27 至 2020-01-28)模型参数迅速上升也说明该模型能够很好的反映社会的变化。不过也因此可以说明该模型对参数的敏感度很高,需要反复调参才能够找到合适的参数。

在本文之前,笔者也曾尝试在不考虑任何模型的基础上,仅仅使用多项式逻辑回归的方法对湖北省 COVID 疫情发展进行预测,却屡屡出现 2019-nCoV 病毒爆炸式扩散或是疫情迅速消退等与事实明显不符的结果。因此,笔者建议在对流行病问题进行研究前,应当对流行病的特点与政府措施进行深入调研,并针对不同流行病的特点构建不同的流行病动力学模型。切不可抱着“数据就是一切”的态度盲目开展研究工作,唯有将数据与事实紧密联系起来,才能够构建出结果与事实紧密联系的数学模型。

## 参考文献

- [1] 国家卫生健康委. 国家卫生健康委关于修订新型冠状病毒肺炎英文命名事宜的通知[EB/OL]. (2020-02-22) [2020-04-08]. [http://www.gov.cn/zhengce/zhengceku/2020-02/08/content\\_5476248.html](http://www.gov.cn/zhengce/zhengceku/2020-02/08/content_5476248.html)
- [2] Lu CW, Liu XF, Jia ZF. 2019-nCoV transmission through the ocular surface must not be ignored[J]. Lancet. [published online ahead of print February 06, 2020]. DOI:10.1016/s0140-6736(20)30313-5
- [3] 国家中医药管理局办公室. 关于印发新型冠状病毒感染的肺炎诊疗方案（试行第五版）的通知 [EB/OL] . (2020-02-05) [2020-04-08]. <http://www.nhc.gov.cn/yzygj/s7653p/202002/3b09b894ac9b4204a79db5b8912d4440.shtml>
- [4] 武汉市新型冠状病毒感染的肺炎疫情防控指挥部. 武汉市新型冠状病毒感染的肺炎疫情防控指挥部通告（第1号）[EB/OL] . (2020-01-23) [2020-4-12]. [http://www.gov.cn/xinwen/2020-01/23/content\\_5471751.html](http://www.gov.cn/xinwen/2020-01/23/content_5471751.html)
- [5] 湖北省新型冠状病毒感染肺炎疫情防控指挥部. 湖北省新型冠状病毒感染肺炎疫情防控指挥部通告 [EB/OL] . (2020-03-24) [2020-4-12]. [http://www.hubei.gov.cn/zhuant/2020/gzxxgzbd/zxtb/202003/t20200324\\_2189256.shtml](http://www.hubei.gov.cn/zhuant/2020/gzxxgzbd/zxtb/202003/t20200324_2189256.shtml)
- [6] Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. The New England journal of medicine 2020. [published online first: January 30, 2020]. DOI:10.1056/NEJMoa2001316

[7] 武汉市卫生健康委员会. 武汉市卫生健康委员会关于不明原因的病毒性肺炎情况通报 [EB/OL]. (2020-01-05) [2020-4-18]

<http://wjw.wuhan.gov.cn/front/web/showDetail/2020010509020>

[8] Liangchen Luo, Yuanhao Xiong, Yan Liu, Xu Sun. Adaptive Gradient Methods with Dynamic Bound of Learning Rate[J/OL]. [published online first: May 7, 2019]. <https://openreview.net/forum?id=Bkg3g2R9FX>

[9] 湖北省统计局. 2019 年湖北省国民经济和社会发展统计公报[EB/OL]. (2020—03-23) [2020-5-3] [http://news.cnhubei.com/content/2020-03/23/content\\_12883985.html](http://news.cnhubei.com/content/2020-03/23/content_12883985.html)

[10] 夏程遗, 孙世温, 刘忠信, 等. 考虑感染时延的局域世界复杂网络上疾病的传播行为[J]. 辽宁工程技术大学学报, 2009, 28(4): 615-618

[11] 郁孜, 吕忠全, 张贵清, 刘庆珍. 基于时变参数-SIR 模型的 2019-nCov 疫情评估和预测 [J/OL]. 电子科技大学学报. (2020-02-02) [2020-4-15] <http://kns.cnki.net/kcms/detail/51.1207.t.20200207.1546.002.html>

[12] 杨慧, 唐明, 徐伯铭. 自适应网络中的流行病传播动力学研究综述[J]. 复杂系统与复杂性科学, 2020(04):67-87.

[13] Jonathan M Read, Derek AT Cummings, et al. Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions [EB/OL]. (2020-01-23) [2020-4-15]

[https://www.researchgate.net/publication/338815978\\_Novel\\_coronavirus\\_2019-nCoV\\_early\\_estimation\\_of\\_epidemiological\\_parameters\\_and\\_epidemic\\_predicti](https://www.researchgate.net/publication/338815978_Novel_coronavirus_2019-nCoV_early_estimation_of_epidemiological_parameters_and_epidemic_predicti)

[ons](#)

[14] 徐宝春. 基于 SIR 模型的 SARS 传染病研究 [D]. 2019.

<http://cdmd.cnki.com.cn/Article/CDMD-10422-1019081450.htm>

## 附录 A

日期	实际确诊人数	预测确诊人数
2020-01-23	549	549
2020-01-24	729	854
2020-01-25	1052	1301
2020-01-26	1423	1764
2020-01-27	2714	2367
2020-01-28	3554	4060
2020-01-29	4586	5806
2020-01-30	5806	7024
2020-01-31	7153	8603
2020-02-01	9074	10592
2020-02-02	11177	12939
2020-02-03	13522	15746
2020-02-04	13522	19027
2020-02-05	19665	22112
2020-02-06	22112	24064
2020-02-07	24953	26196
2020-02-08	27100	28378
2020-02-09	29631	30714
2020-02-10	31728	33306
2020-02-11	33366	36053
2020-02-12	33366	48206
2020-02-13	48206	50289
2020-02-14	51986	52544
2020-02-15	54406	54955
2020-02-16	56249	57552



2020-02-17	58182	60364
2020-02-18	59989	62468
2020-02-19	61682	62457
2020-02-20	62457	62882
2020-02-21	63088	63331
2020-02-22	63454	63798
2020-02-23	63889	64285
2020-02-24	64287	64803
2020-02-25	64786	65374
2020-02-26	65187	65996
2020-02-27	65596	65980
2020-02-28	65914	66399
2020-02-29	66337	66842
2020-03-01	66907	67103
2020-03-02	67103	67215
2020-03-04	67217	67338
2020-03-04	67332	67467