

全国新型冠状病毒疫情分析及预测

瞿颖秋*, 张晓良, 曾莎, 李超

重庆理工大学理学院, 重庆
Email: *yingqiuqu@163.com

收稿日期: 2021年1月19日; 录用日期: 2021年2月15日; 发布日期: 2021年2月22日

摘要

本文对新型冠状病毒感染的确诊人数进行了数据可视化, 然后采用非线性模型、logistic增长函数、SEIR模型三种模型对新型冠状病毒感染的确诊人数数据进行拟合和预测。通过研究表明, 模型分析与疫情发展的真实情况基本相符, 但仍需进一步改进。

关键词

2019_nCoV, 非线性模型, Logistic增长函数, SEIR模型, R软件

Analysis and Forecast of Epidemic Situation in Novel Coronavirus

Yingqiu Qu*, Xiaoliang Zhang, Sha Zeng, Chao Li

College of Science, Chongqing University of Technology, Chongqing
Email: *yingqiuqu@163.com

Received: Jan. 19th, 2021; accepted: Feb. 15th, 2021; published: Feb. 22nd, 2021

Abstract

In this paper, the number of confirmed cases of infection in novel coronavirus was visualized, and then the number of confirmed cases of infection in novel coronavirus was fitted and predicted by three models: nonlinear model, logistic growth function and SEIR model. The research shows that the model analysis is basically consistent with the real situation of epidemic development, but it still needs further improvement.

*通讯作者。

Keywords

2019_nCoV, Nonlinear Model, Logistic Growth Function, SEIR Model, R Software

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

从 2019 年 12 月开始,武汉市出现了未知肺炎的病例,2020 年 1 月 7 日,一种新的冠状病毒发现了,该病毒以飞沫和接触为主,传播速度极快。春运很快到来,新型冠状病毒肺炎(简称“新冠肺炎”)迅速蔓延到全国,多地受新冠病毒影响。自疫情暴发以来,人们对新型冠状病毒的了解逐渐深入,国家对疫情应付手段的调整也随之而变,并且对疫情的信息掌控越来越丰富。在此过程中,学者们对疫情开展了预测研究,付新苗[1]等人基于 2020 年 1 月 21 日至 2 月 14 日的全国及其各省累计病例用 Boltzmann 函数拟合并且估计临界日期(R^2 接近 0.999)。耿辉[2]等人基于 SEIR 模型分析了相关的措施在防控疫情中起到的作用,此外我们的 SEIR 模型将基于耿辉等人的模型对参数进行了改进。金启轩[3]基于 1 月 23 日至 2 月 12 日的全国 30 个省份(不包括湖北省以及港澳台地区)的数据对中国新冠肺炎疫情进行预测建模,我们接下来将使用更新后的数据使用该方法进行建模。

由于前期疫情的诊断治疗流程很繁琐、时间很长,人手,设备和场地也不足,导致每天只能对一定数量的患者进行诊断,因此,湖北省确诊人数的资料反映了目前检测水平,而非实际情况。目前这种情况实际上是医疗能力和诊断标准的问题,本文以相关预测估计中的不足为基础,选择了全国 30 个省(不包括湖北省和港澳台)的资料,综合了国家防控疫情的措施,并对疫情研究的进展作了建模分析,预测了符合目前实际疫情的发展状况,从而预测了疫情发展的进程,对疫情控制有一定的参考价值。

2. 疫情发展现状及特点

疫情数据来源于国家卫健委及其各个省市卫健委。

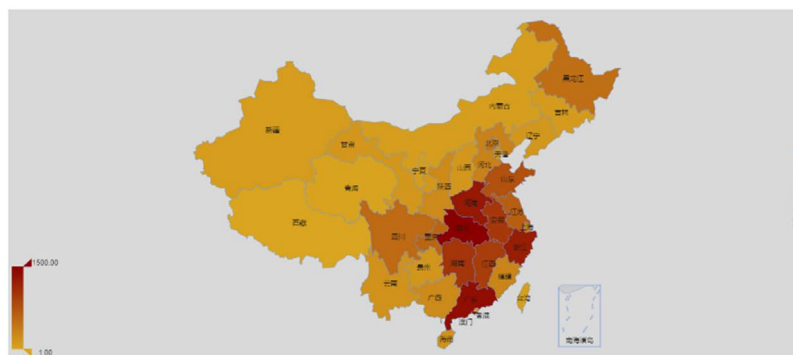


Figure 1. National epidemic map on March 13

图 1. 3 月 13 日全国疫情图

以 2020 年 3 月 13 日的数据为例,图 1 为 3 月 13 日全国疫情图,从该图我们可以看出每个省市的当日的累计确诊病例,也可以看出各个省市的情况。我们看到与湖北临近的河南、安徽、浙江、湖南、广

东等地确诊患病人数非常多，它们的颜色更接近于红色。由黄到红体现了地区患病人数的多少，我们可以清楚的看到这一点。

同样我们可以看到以 3 月 27 日数据为基准的，全球疫情图。由图 2 所示：

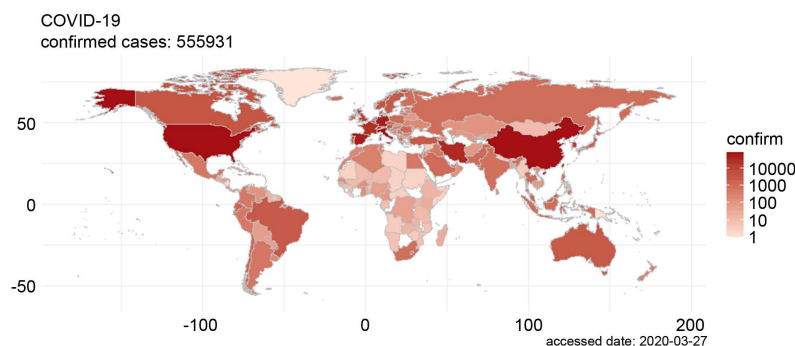


Figure 2. Global epidemic map on March 27

图 2. 3 月 27 日全球疫情图

我们从 1 月 20 号到 3 月 7 号选择了每日确诊的人数，以及湖北省的累计确诊人数。从图 3 中可以看出，从 1 月 20 号到 2 月 16 号左右，新型肺炎的每日确诊病例，虽然略有上升，但总体表现出上升趋势，而累计确诊病例上升趋势为肉眼可见。可观察到 1 月 27 日至之后的一段时期，每天新增的确诊病例有明显的增加，这与病毒检查试剂盒的充分供应有关。2 月 17 日之后，新增病例仍有波动，但我们能看到整体呈下降趋势，而且累积确诊病例上升趋势也逐渐平缓，我们可以认为这种情况已经达到了增速的拐点。由图 3 可以看出，3 月 7 日后每天新增的确诊病例数已经达到个位，累计确诊人数也逐渐平缓。

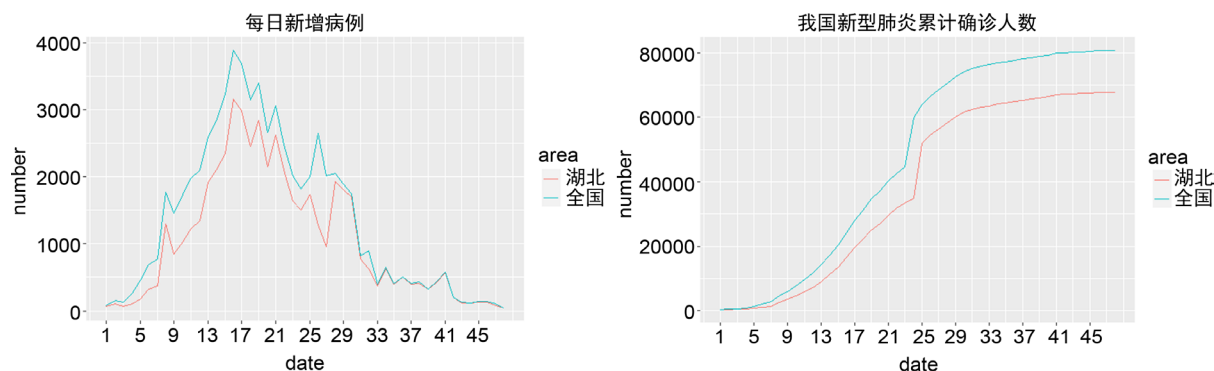


Figure 3. Line chart of daily new cases and cumulative number of confirmed cases in China and Hubei

图 3. 全国与湖北地区每日新增病例和累计确诊人数折线图

3. 模型

3.1. 关于全国 30 个省的相关数据进行建模

我们从 1 月 20 日至 2 月 30 日分别搜集了全国和湖北省及武汉市的每日新增确诊病例和累计确诊病例，数据来自国家卫健委[4]、湖北省卫健委[5]以及武汉市卫健委[6]。

全国 30 省份疫情的性质

对数据观察发现，1 月 23 日前，湖北省之外的其余省份疫情是可控制的。且像武汉市一样的“人传人”现象大都局限在有接触的人之间。之后，国家采取措施，让湖北省内的各市先后禁止开展大型聚集

活动且尽量不出门,这极大的限制了病毒的传播。出于一般性考虑,我们只对全国 30 个省份(除湖北省),因为相对来说,30 个省份的数据可靠、及时,只有人传人,初始基数是确定的。我们用 1 月 23 日之后的增长率进行了统计和建模。此时增长率=每日新增确诊病例/前一天累计确诊病例,我们记 1 月 25 日为第一天,对 1 月 25 日至 2 月 20 日的增长率进行建模。

由散点图我们可以看到它们是非线性的,根据散点图我们可以尝试用三次函数对其进行拟合。此时模型拟合优度 R^2 为 0.959,高度拟合,说明用三次函数来拟合是合适的。此时模型为:

$$\text{增长率} = 0.665 + -9.5 * 10^{-5} * t^3 + 5.317 * 10^{-3} * t^2 - 0.099t。$$

3.2. Logistic 增长函数

病毒在传播初期,由于政府和民众重视程度不够,传染数据呈指数增长,增长不受抑制。在疫情呈现之后,全国各地隔离了确诊和疑似病例,这一措施阻止了部分病毒的传播,因此传染的速度会逐步遭到压制。我们觉得 logistic 增长模型[7]可用于疫情传播前中期的趋势,因此,我们只选取了 1 月 20 日至 2 月 9 日的全国累计确诊病例对其进行建模。logistic 增长函数的一般形式为:

$$N(t) = \frac{KN_0e^{rt}}{K + N_0(e^{rt} - 1)}$$

其中, K 为环境容量, N_0 为初始容量, r 为增长速率。

此时,我们将 1 月 20 日至 2 月 9 日的数据进行拟合,可以得到拟合的模型为:

$$N(t) = \frac{49410 * 606.1 * e^{0.2907t}}{49410 + 606.1 * (e^{0.2907t} - 1)}$$

由拟合的模型我们可以看到确诊人数将会在 5 万左右收敛,而且实际情况远远不止,此时我们要考虑这个模型是否合适。从图 4 我们看到拟合效果是不错的,残差分布接近均匀,模型拟合的结果还是可以的,并且我们对 2 月 10 日至 2 月 15 日的累计确诊病例人数进行了预测,预测结果为:43,555.64, 44,897.10, 45,955.40, 46,779.91, 47,416.02, 47,903.07, 我们可以看到对于 10 日至 12 日的预测与实际的是非常接近的,但是 13 日至 15 日的预测就不那么好了,远小于实际的。因此我们认为此模型对于这些数据来说是不合适的,我们需要考虑到实际的情况,对模型进行修正。

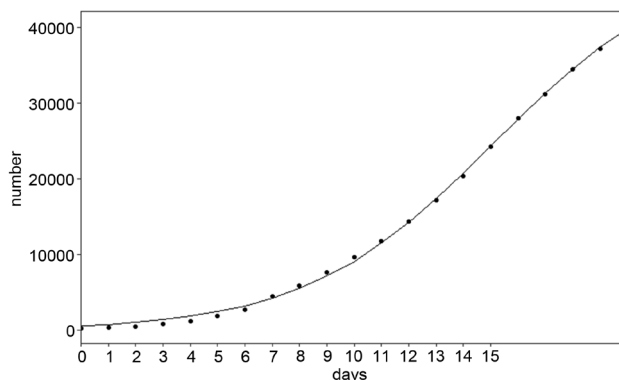


Figure 4. Logistic growth function fitting diagram
图 4. Logistic 增长函数拟合图

3.3. SEIR 模型

此处尝试用 SEIR 模型揭示 2019-nCov 的传播过程,模型涉及 4 类人群:易感人群(S)、潜伏人群(E)、

感染人群(I)、移除人群(R), 假设康复概率是 γ , 易感染者的感染概率是 β , 感染者接触的人数是 r , 潜伏者会传染易感者的概率为 β_2 , 潜伏者每天接触的人数是 r_1 , 另假定易感者的感染概率与潜伏者的传染概率相同, 则微分方程组如下:

$$\begin{aligned} S_i &= S_{i-1} - \frac{r\beta I_{i-1}S_{i-1}}{N} - \frac{r_1\beta_2 E_{i-1}S_{i-1}}{N} \\ E_i &= E_{i-1} - \frac{r\beta I_{i-1}S_{i-1}}{N} - \alpha E_{i-1} + \frac{r_1\beta_2 E_{i-1}S_{i-1}}{N} \\ I_i &= I_{i-1} + \alpha E_{i-1} - \gamma I_{i-1} \\ R_i &= R_{i-1} + \gamma I_{i-1} \end{aligned}$$

我们根据耿辉[2]等人的研究, 对模型的参数进行了一些修正, 此时我们令 $r=12$, 这个时候我们可以将模型看作是在未干预情况下的爆发趋势。

武汉市卫健委在12月12日对第一例患者发病时间进行了调查, 各个领域专家进行了研究探讨, 推断该患者的感染时间大致发生在本月初甚至更早, 因此, 根据SEIR模型, 如图5, 我们可以预测到在未干预的情况下疫情在二月下旬出现高峰。并且我们用在未干预情况下的预测确诊病例与实际相比, 我们可以看到实际的确诊病例数比预测的值要小得多, 特别是越到后面相差越大, 出现这种情况是因为国家出台的相关政策起到了作用。

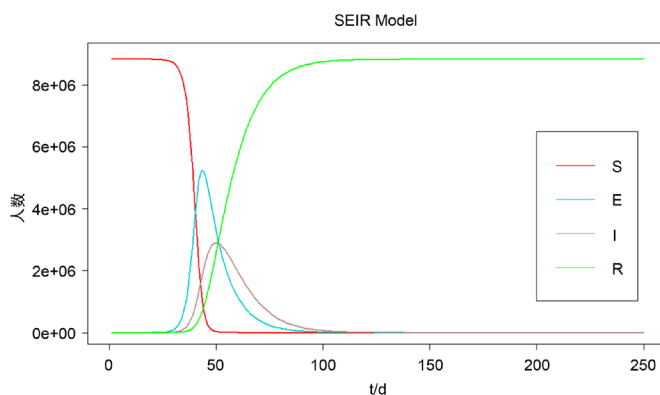


Figure 5. SEIR model fitting diagram

图 5. SEIR 模型拟合图

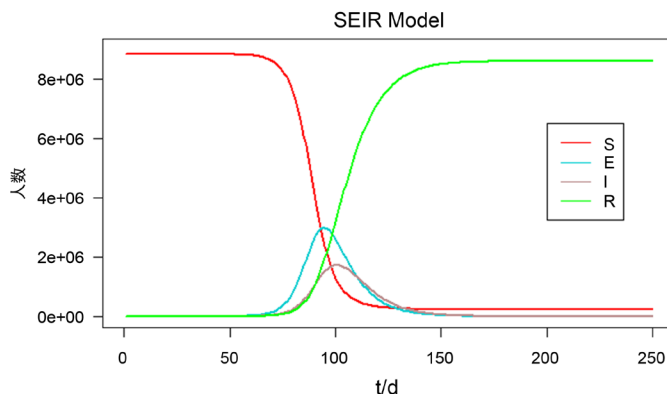


Figure 6. SEIR model fitting diagram with government intervention

图 6. 有政府干预时的 SEIR 模型拟合图

接下来, 我们可以考虑在政府进行干预情况下的模型, 如图 6。此时, 有关的参数我们也要进行修正, 并且在政府进行强制干预后, 感染曲线的减缓趋势明显变得平缓, 感染人群的峰值降低了大约 1/3、潜伏人群的峰值下降了大约 1/2, 我们还发现在时间上两者的峰值数据大致相差了七天, 从中我们也今天观察到政府的干预以及治疗经验和手段的不断丰富, 疫情得到了很大的控制。

4. 结论

由上述拟合我们可以看到, 对于增长率而言, 1 月 23 日之后, 日增长率以三次形式作为模型可高度拟合, 并且增长率逐渐趋于零。此时模型为: $\text{增长率} = 0.665 + -9.5 * 10^{-5} * t^3 + 5.317 * 10^{-3} * t^2 - 0.099t$ 。且可以知道, 增长率将在 3 月上旬为 0, 这与我们的实际情况还是有一些偏差。这是因为国家和整个社会对于病毒的防控起到了作用, 当增长率趋于 0 时, 确诊人数就会达到高峰值, 之后逐渐减少, 最后趋于平稳。我们从 logistic 增长函数仍可以看出这种趋势, 当确诊人数达到一定峰值之后, 就会逐渐下降, 最后趋于平稳。由拟合的结果我们可以看到确诊人数将会在 5 万左右收敛, 而且实际情况远远不止, 累计确诊人数峰值远远高于 5 万, 因此我们认为此模型对于这些数据来说是不合适的, 我们需要考虑到实际的情况, 对模型进行修正。对于国家的防控疫情措施, 我们从 SEIR 模型可以清楚地观察到, 感染人群和潜伏人群的峰值都降低了, 这说明政府的强制干预措施对防控疫情起到了很大的作用。

参考文献

- [1] Fu, X.M., *et al.* (2020) Simulating and Forecasting the Cumulative Confirmed Cases of SARS-CoV-2 in China by Boltzmann Function-Based Regression Analyses. *Journal of Infection*, **80**, 578-606. <https://doi.org/10.1016/j.jinf.2020.02.019>
- [2] 耿辉, 徐安定, 王晓艳, 张勇, 尹小妹, 马茂, 吕军. 基于 SEIR 模型分析相关干预措施在新型冠状病毒肺炎疫情中的作用[J]. 暨南大学学报(自然科学与医学版), 2020, 41(2): 175-180.
- [3] 金启轩. 中国新冠肺炎疫情预测建模与理性评估[J]. 统计与决策, 2020, 36(5): 11-14.
- [4] 中华人民共和国国家卫生健康委员会. 新型冠状病毒感染的肺炎疫情最新情况[EB/OL]. http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml, 2020-02-01.
- [5] 湖北省卫生健康委员会. 湖北省新型冠状病毒感染的肺炎疫情情况[EB/OL]. <http://wjw.hubei.gov.cn/>, 2020-02-01.
- [6] 武汉市卫生健康委员会. 武汉市卫生健康委员会关于新型冠状病毒感染的肺炎疫情通报[EB/OL]. <http://wjw.wuhan.gov.cn/>, 2020-02-01.
- [7] 王双明, 张明军, 樊馨蔓. 一类具时滞的周期 logistic 传染病模型空间动力学研究[J]. 应用数学和力学, 2018, 39(2): 226-238.