

长春新冠病毒疫情分析

田泽禹, 鲁荣欣, 林治豪

2022 年 5 月 18 日

声明: 本文仅从数学建模角度研究长春新冠病毒疫情数据, 对疫情判断与防控措施不含有任何立场.

摘要

2022 年 3 月, Omicron 毒株的新冠病毒肺炎疫情席卷了长春. 由于 Omicron 病毒极强的传播能力, 这次疫情的严重程度史无前例. 经过了五十余天的全民攻坚, 这轮疫情目前已经得到了全面的控制. 本文旨在基于长春各县市区的疫情数据, 结合长春疫情防控的实际情况, 利用 SERI 和 SERI-ARIMA 方法, 对 Omicron 病毒在长春市的传播进行数学建模. 以期探索疫情的动力学原因, 并希望未来遭遇类似的病毒时, 能给予一定程度的帮助和指导.

关键词: 新冠疫情; 长春疫情; 传染病预测.

目录	I
----	---

目录

1	问题背景	1
2	问题分析	1
2.1	病毒动力学特征	2
2.2	防疫措施	2
3	符号说明	3
4	模型建立	4
4.1	数据分析与处理	4
4.1.1	九台区	7
4.1.2	宽城区	8
4.1.3	长春市整体及其它区	8
4.2	传染病模型 SEIR	10
4.2.1	经典 SEIR 模型	10
4.2.2	暴露者具有感染性	11
4.3	基于 ARIMA 的 SEIR-ARIMA 疫情预测模型	13
4.3.1	经典 ARIMA 模型	13
4.3.2	基于 ARIMA 的 SEIR-ARIMA 疫情预测模型	15
5	灵敏度分析	17
5.1	个体措施	17
5.1.1	SEIR 模型	17
5.2	政策影响	19
5.2.1	SEIR 模型	19
5.2.2	基于 ARIMA 的 SEIR-ARIMA 疫情预测模型	21

目录	II
6 模型的评价与推广	22
6.1 模型优点	22
6.2 模型不足与推广思路	22
6.3 数据问题	23
参考文献	24
A 附录	25
A.1 代码	25

1 问题背景

2019 年末以来, 新冠病毒疫情逐渐由局部爆发传播至全球肆虐. 我国 3 年以来不同地区在不同程度上也经历了新冠疫情的爆发. 至 2022 年 5 月, 新冠病毒已在全球变异进化出至少 11 种主要毒株, 而每一大类中又有许多小分支难以统计. 我国目前主要经历了 Beta, Delta, Omicron 等流行毒株导致的疫情的爆发.

Omicron 毒株最早于 2021 年 11 月 9 日在南非首次检测到, 2021 年 11 月 26 日, 世界卫生组织将其定义为第五种“关切变异株”; 11 月 29 日, 世卫组织称, 新冠病毒奥密克戎变异毒株在全球总体风险评估为“非常高”, 其极高的传染能力, 可能使其在世界广泛传播.

2022 年 3 月以来, 国内多地本土疫情大爆发, 多地出现零星病例, 突发聚集性疫情. 本次流行毒株为 Omicron B.A.2, 由于其极高的传播能力和非活体附着力, 我国部分地区疫情严重爆发. 其中上海市, 吉林省长春市, 吉林省吉林市遭遇严重爆发, 超过以往疫情严重程度的最高纪录.

截至 2022 年 5 月中旬, 吉林省疫情已经得到全面控制. 从 3 月初至 5 月初, 吉林省不同地区都采取了不同的干预措施以控制疫情的蔓延. 本文将以太安市这一时间段内的疫情数据为切入点, 结合长春疫情防控阶段的实际情况, 通过数学建模的手段, 对疫情的动力学原因进行探索并给出其它情况下的数据模拟.

2 问题分析

影响病毒传播或疫情数据的因素主要由病毒本身参数, 传播环境与宿主活动决定. 其中病毒参数包括病毒的传染力, 病毒的非活体存活力, 病毒的潜伏期, 发病的死亡率或自愈周期等; 传播环境主要指病毒是否适应客观气候环境并进行传播; 宿主活动主要指易感者的物种种类, 易感者为避免感染所采取的措施等. 由于目前没有数据表明新冠病毒疫情的传播会受到

气候的明显影响, 因此我们忽略气候等条件对于病毒传播的影响.

由于 Omicron B.A.2 的病理学特征不会对疫情传播产生本质的影响, 因此我们主要考虑 Omicron BA.2 的传播动力学特征. 同时, 目前没有数据表明人类以外物种会对新冠病毒传播和疫情发展产生本质影响, 因此我们忽略其它物种在本次疫情传播中的作用; 在宿主方面, 我们将考虑人为可能的防疫措施对新冠病毒疫情的影响.

2.1 病毒动力学特征

我们将列举一些常见的病毒动力学特征, 建模时我们将给出相关参数的详细符号与定义. 病毒动力学特征包括: 病毒传染率, 感染的潜伏期, 感染的病死率, 治愈率, 康复是否永久免疫等;

据估计, Omicron BA.2 的基础传播系数 R_0 高达 12 [1], 意味着在无约束的自然传播下, 平均一人可以传播给 12 人. 而据统计平均数值显示, Omicron BA.2 毒株的潜伏期为 3-5 天 [2], 其中大部分感染呈现为无症状感染, 无症状感染期仍具有传染性. Omicron BA.2 感染的病死率根据不同年龄, 人种, 疫苗品类与接种情况会呈现较大的不同, 本次长春疫情爆发没有死亡病例的数据, 因此本次建模中可假设感染病毒的死亡率为 0. 不同系统中病毒治愈率不同, 可以根据疫情数据计算同时间段内疾病的治愈率. 数据表明, 在治愈后, 病人仍有几率重新感染病毒, 因此 Omicron BA.2 毒株不属于终生免疫型病毒.

2.2 防疫措施

防疫措施能够分为个体和群体两个方面. 个体层面上, 个人可以通过戴口罩, 减少聚集, 保持距离, 注意卫生等方面减少病毒感染的概率. 群体层面的措施包括加强病毒检测筛查提高识别率, 及时收治病人, 科学的隔离措施, 限制人员密度和流动以减少接触, 加强疫苗接种等方法来降低传播风险.

在部分模型中, 物流系统和冷链系统被认为是病毒传播的一种途径. 限制物品的交换和加强消毒成为控制病毒的另一种途径.

3 符号说明

- 易感者 (**Susceptible**): 指缺乏免疫能力健康人, 与感染者接触后容易受到感染.
- 暴露者 (**Exposed**): 暴露者, 指接触过感染者但不存在传染性的人, 可用于存在潜伏期的传染病.
- 感染者 (**Infectious**): 指有传染性的病人, 可以传播给 S, 将其变为 E 或 I .
- 康复者 (**Recovered**): 指病愈后具有免疫力的人, 如是终身免疫性传染病, 则不可被重新变为 S, E 或 I , 如果免疫期有限, 就可以重新变为 S 类, 进而被感染.

表 1: SEIR 模型符号说明

t	时间变量
$S(t)$	t 时刻的易感者人数
$E(t)$	t 时刻的暴露者人数
$I(t)$	t 时刻的感染者人数
$R(t)$	t 时刻的康复者人数
β	即每个患病者每天有效接触的易感者的平均人数
σ	即每天发病成为患病者的暴露者占暴露者总数的比率
γ	即每天被治愈的患病者人数占病人总数的比率
$\frac{1}{\gamma}$	平均传染期, 即从患病到治愈的天数
$\sigma = \frac{\beta}{\gamma}$	即每个患病者在整个传染期 $\frac{1}{\gamma}$ 天内, 有效接触的易感者人数

表 2: 无症状具有感染性模型符号说明

r_1	患病者 I 接触的人的个数
b_1	每次接触患病者将易感者传染的概率
r_2	暴露者 E 接触的人的个数
b_2	每次接触暴露者将易感者传染的概率
a	暴露者转为患病者的概率
c_1	患病者康复的概率
c_2	暴露者康复的概率
p	易感者感染后变成暴露者的概率

4 模型建立

在本模型中, 我们仅考虑病毒传播动力学和人为干预的影响, 且不考虑病毒的病死率. 建模数据为长春市 3 月 11 日至 5 月 1 日的公开数据. 由于此时间段长春已处于管控阶段, 因此我们的系统不为整个长春市, 而是考虑以区划甚至以社区为单位的建模, 并将其按照时序线性相加.

4.1 数据分析与处理

根据百度新型冠状病毒大数据所拟合出的曲线, 我们可以明显地看出本次长春新冠疫情的情况十分复杂, 多重因素共同影响决定了最终的数据曲线. 由百度给出的长春疫情曲线图 [3], 其数据多尖峰, 多波动, 无明显趋向. 长春疫情与防疫阶段大致分为如下阶段

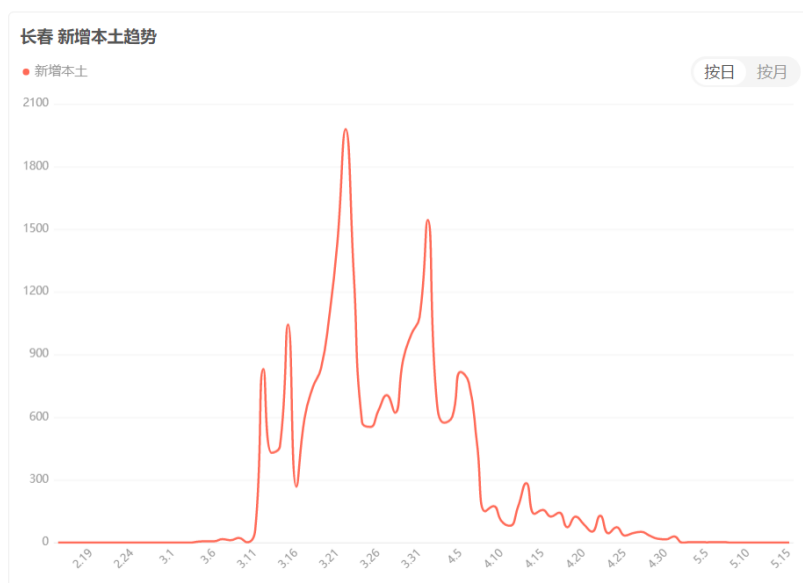


图 1: 百度新冠疫情大数据-长春

- 3月初, 疫情开始暗处传播, 可能有吉林市关联疫情.
- 3月12日左右, 长春市众多高校开始封校.
- 3月15日左右, 长春市实行交通管制.
- 3月20日, 长春市启动封城.
- 4月4日, 长春市启动“5日主城区清零”计划.
- 4月14日, 长春市宣布社会面基本动态清零, 此后新增感染均为管控区内上报.

分析长春疫情曲线可能的影响因素, 大致分为以下几点:

- 传播区域: 长春市疫情是分区传播, 直观数据为不同区域的数据加和, 因此数据失去趋向; 由于交通管制与社区管制的变化, 传播逐渐由大区域变为零星的社区传播.

- 传播途径: 日常接触传播, 密集群体传播, 物品交换引起传播, 集体核酸检测时接触传播.
- 检测效果: 核酸检测结果并非完全准确, 可能出现”假阴性”情况; 物资和人力会导致一定的检测能力限制; 资源不足情况下可能出现数据波动或延迟的情况.
- 政策力度: 不同政策可能在检测筛查, 病人转运诊治, 管控限制上产生较大影响, 进而影响数据.

根据吉林省卫生健康委员会与长春市疫情发布会公布的数据 [4], 我们可以获得长春市 3 月 17 日至 4 月 13 日各区划分区的疫情数据情况, 将各区分开考虑将获得

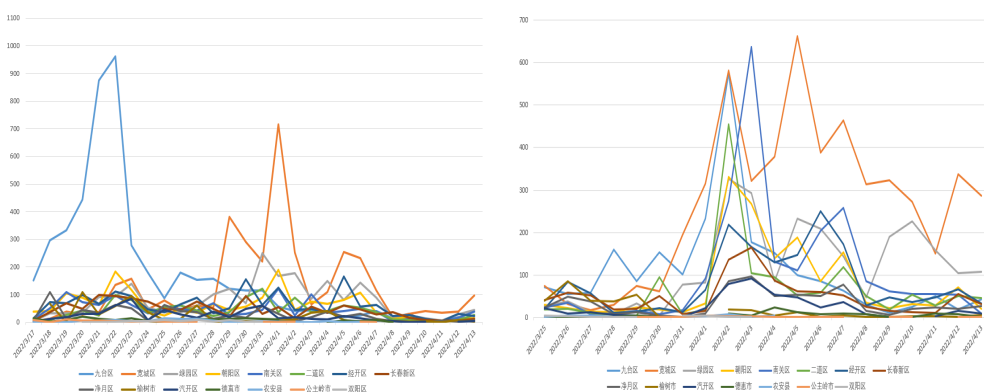


图 2: 长春分区划新增确诊病例与无症状感染者

在本次疫情整体的传播过程中, 根据疫情控制程度我们可以将这些区划分为四类

- 严重爆发: 九台区, 宽城区
- 中度爆发: 绿园区, 朝阳区, 南关区, 二道区, 经开区, 长春新区
- 中低程度爆发: 净月区, 榆树市, 汽开区

- 低疫情地区: 德惠市, 农安县, 公主岭市, 双阳区

除九台区与宽城区外, 其它地区的新增确诊曲线与新增无症状感染者曲线大致趋势相同. 因此我们需单独拟合这两个分区, 其余区划可以单独考虑.

4.1.1 九台区

由 Gaussian 型曲线拟合, 我们可以得到九台区的疫情曲线. 其函数表达式为:

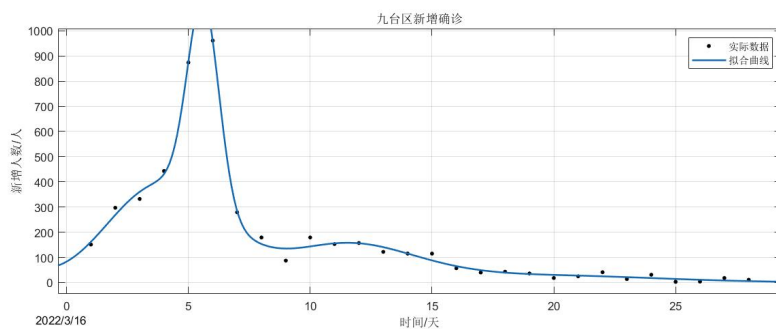


图 3: 九台区新增确诊病例

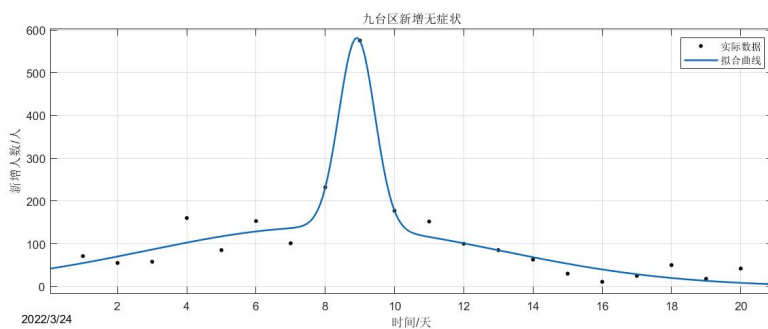


图 4: 九台区新增无症状感染者

4.1.2 宽城区

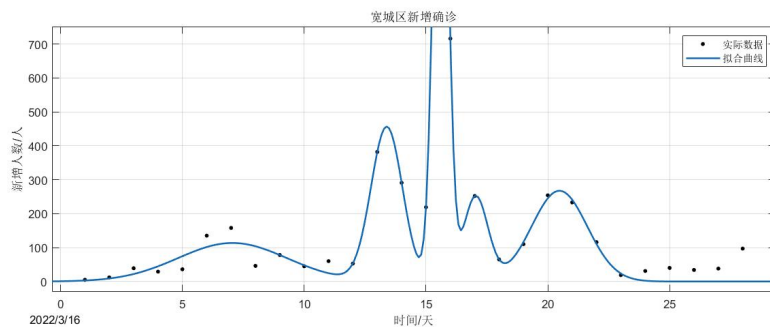


图 5: 宽城区新增确诊病例

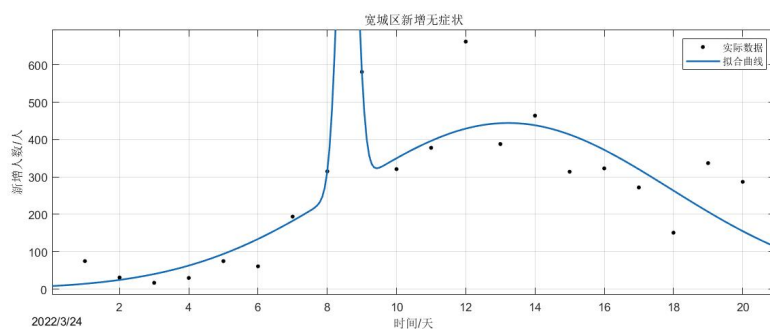


图 6: 宽城区新增无症状感染者

4.1.3 长春市整体及其它区

长春市及其它区的曲线具有一定的不光滑性, 这个现象与核酸采样的间隔以及数据录入及公布具有一定的关系. 政策强度对数据也具有一定的影响.

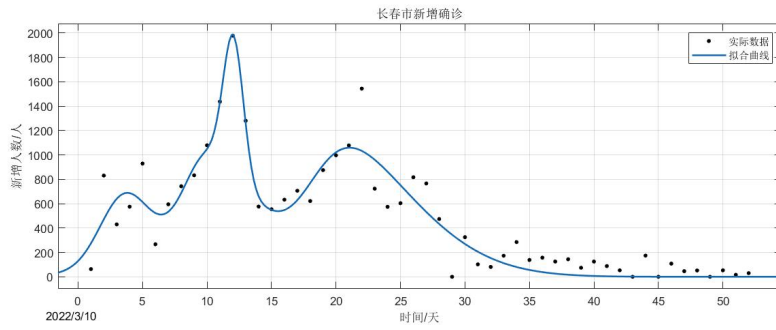


图 7: 长春市新增确诊病例

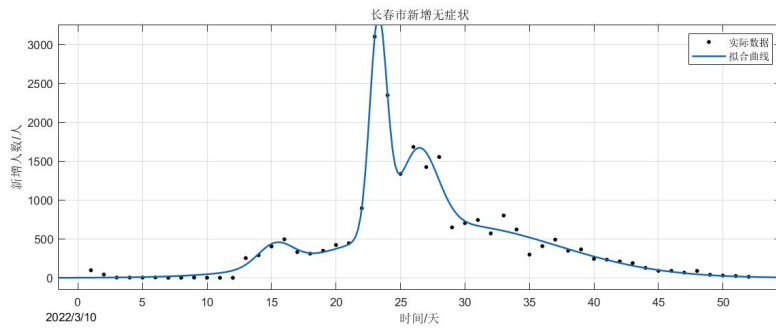


图 8: 长春市新增无症状感染者

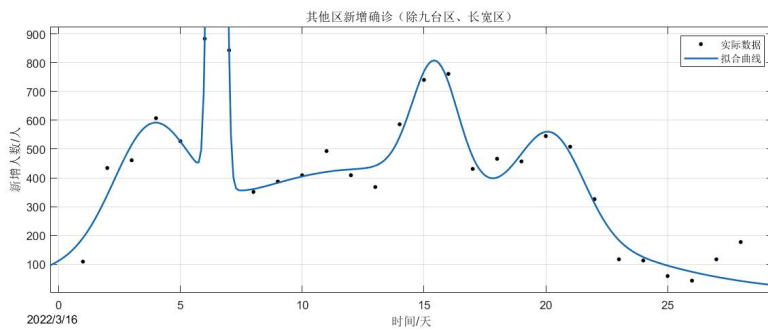


图 9: 其他区总新增确诊病例

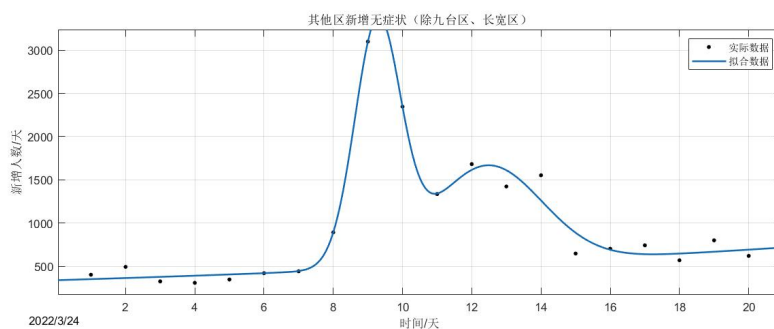


图 10: 其他区总新增无症状感染者

4.2 传染病模型 SEIR

4.2.1 经典 SEIR 模型

在新冠病毒传播的过程中, 病毒存在潜伏期, 确诊后病人可以通过治疗恢复健康因此我们可以使用 SEIR 模型. 而虽然病愈后存在再次感染的几率, 但出院病人一般都仍会处在一定的医疗观察之中, 不会再参与模型的运动过程, 因此我们可以不考虑 R 再转化为 S 的过程. 在没有出生或死亡的封闭人口中, SEIR 模型变为:

$$\begin{aligned}\frac{dS(t)}{dt} &= -\frac{\beta S(t)I(t)}{N}, \\ \frac{dE(t)}{dt} &= \frac{\beta S(t)I(t)}{N} - \sigma E(t), \\ \frac{dI(t)}{dt} &= \sigma E(t) - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t).\end{aligned}$$

在此意义下, 传播参数都为社会学意义下的平均参数, 并非自然传播过程中的参数.

4.2.2 暴露者具有感染性

Omicron BA.2 的平均潜伏期为 3-5 天, 在此段无症状感染时期, 一般第 3 天左右患者就已经具有感染性, 因此我们认为此时患者具有低于患病者的传染性. 潜伏者仅转变为感染者. 四类人群符号同上, 其他参数如符号约定处相同.

由此我们可以列出微分方程组:

$$\begin{aligned}\frac{dS}{dt} &= -r_1 b_1 SI - r_2 b_2 SE \\ \frac{dE}{dt} &= r_1 b_1 SI - aE + r_2 b_2 SE \\ \frac{dI}{dt} &= aE - c_1 I \\ \frac{dR}{dt} &= c_1 I\end{aligned}$$

如果我们考虑在无症状时期就进行治疗干预, 并且无症状暴露者可以直接转为康复者, 感染后有按概率变为感染者或暴露者. 新增参数如符号约定处, 我们可以列出以下的微分方程组.

$$\begin{aligned}\frac{dS}{dt} &= -r_1 b_1 SI - r_2 b_2 SE \\ \frac{dE}{dt} &= p(r_1 b_1 SI + r_2 b_2 SE) - aE - c_2 E \\ \frac{dI}{dt} &= aE - c_1 I + (1 - p)(r_1 b_1 SI + r_2 b_2 SE) \\ \frac{dR}{dt} &= c_1 I + c_2 E\end{aligned}$$

取长春整体的防疫数据, 根据实际情况设置平均初始值, 取 $r_1 = 5$, $b_1 = 0.9$, $r_2 = 10$, $b_2 = 0.3$, $a = 0.1$, $c_1 = 0.05$, $c_2 = 0.1$, $N = 55000$, $init = 200$, $p = 0.9$. 其中 $init$ 指初始时刻确诊人数. 其中 c_1 , c_2 的治疗率低

于常识, 是因为其代表单位时间内的治愈概率, 而我国诊疗标准对于治愈的判定较为严格, 因此将数值稍微调小. r_1, r_2 分别代表患病者与暴露者接触人的平均值, 由于 Omicron BA.2 具有一定的物传性质, 因此数值稍微调高. 由于涉及新冠病毒的人群一般都会被观察, 由整体的接触观察人群在 45000 人左右, 推断疫情的系统整体涉及 55000 人. 在此初始值下, 使用微分方程组数值解的手段, 获得理论曲线.

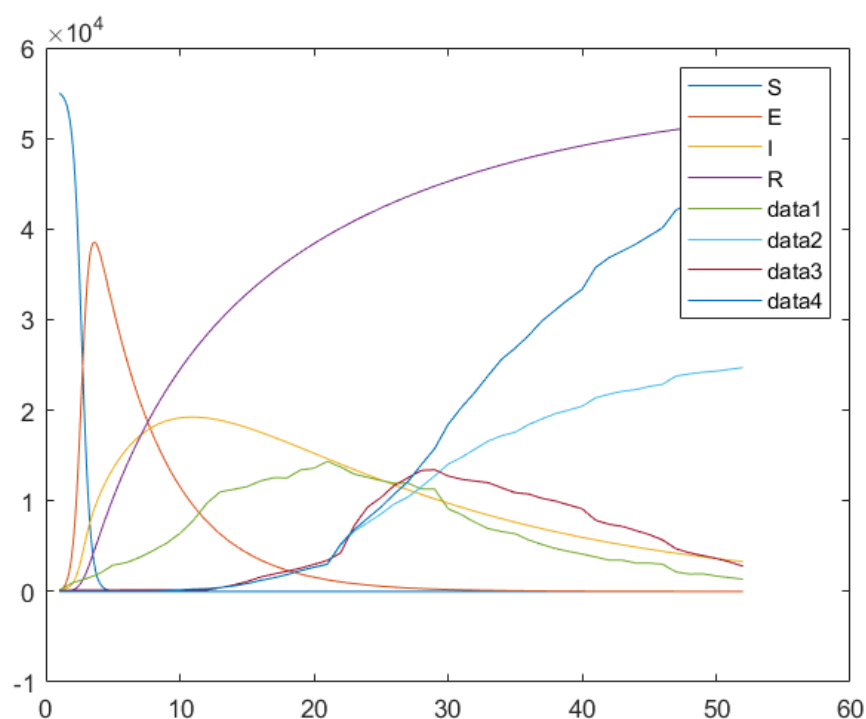


图 11: 长春市新冠疫情 SEIR 模型理论曲线

其中图例中“S”, “E”, “I”, “R” 均为理论预测; data 所指的四条曲线分别为: 长春市新冠现确诊病例, 由确诊病例康复, 长春市现无症状病例或医学观察者, 由确诊或无症状医学观察等康复或解除观察的总数. 由于前期仅有九台区一区发生了新冠病毒疫情爆发, 涉疫人员少, 因此实际值远低于预

测值. 后期长春市中部区域都有中规模的疫情爆发, 没有极值数据存在, 且疫情控制在特定人员的区域内 (即模型假设范围), 因此后段曲线拟合较为贴合. 康复者的曲线具有时滞, 是因为康复者在第一次满足康复特征后仍会保持一段时间的医学观察期, 而 data 的数据来源为每日康复出院的数据, 因此在时滞的观点下具有一定的合理性.

而暴露者即无症状感染者的数据出现了很大偏差. 在 SEIR 模型中, 暴露者总会先感染者达到峰值, 因此在 SEIR 模型中长春疫情曲线是无法同时拟合暴露者的. 但根据我国多地区的疫情曲线数据, 无症状感染者的高峰时间上同样先于确诊者, 且数据上远远多于确诊人数. 据哈佛大学研究, Omicron BA.2 的传播大部分为无症状感染者 [5], 据因此可以推断长春市无症状感染者数据是不符合疫情的动力学规律的, 这样的数据可能由多种因素导致.

通过使用最小二乘法拟合累计确诊人数的数据参数, 发现求解过程在广域内搜索会出现发散的情况, 这与长春市前期疫情数据干扰, 后期疫情数据非自然因素的噪点太多的原因相关.

4.3 基于 ARIMA 的 SEIR-ARIMA 疫情预测模型

4.3.1 经典 ARIMA 模型

ARIMA 预测模型 (Autoregressive Integrated Moving Average Model) 是自回归移动平均模型. ARIMA 模型的基本思想是: 将预测对象随时间推移而形成数据序列视为一个随机序列, 用一定的数学模型来近似描述这个序列. 这个模型一旦被识别后就可以从时间序列的过去值及现在值来预测未来值.

ARIMA(p, d, q) 模型可以表示为:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

其中 L 是滞后算子 (Lag operator),

$$d \in \mathbb{Z}, d > 0.$$

ARIMA 模型含有三个参数: p, d, q .

p - 代表预测模型中采用的时序数据本身的滞后数 (lags), 称作 AR/Auto-Regressive 项.

d - 代表时序数据需要进行几阶差分才是稳定的, 称作 Integrated 项.

q - 代表预测模型中采用的预测误差的滞后数 (lags), 称作 MA/Moving Average 项.

ARIMA 模型的基本程序:

(一) 根据时间序列的散点图, 自相关函数和偏自相关函数以 ADF 单位根检验其方差, 趋势及其季节性变化规律, 对序列的平稳性进行识别. 一般来讲, 经济运行的时间序列都不是平稳序列.

(二) 对非平稳序列进行平稳化处理. 如果数据序列是非平稳的, 并存在一定的增长或下降趋势, 则需要对数据进行差分处理, 如果数据存在异方差, 则需对数据进行技术处理, 直到处理后的数据自相关函数值和偏相关函数值无显著地异于零.

(三) 根据时间序列模型的识别规律, 建立相应的模型. 若平稳序列的偏相关函数是截尾的, 而自相关函数是拖尾的, 可断定序列适合 AR 模型; 若平稳序列的偏相关函数是拖尾的, 而自相关函数是截尾的, 则可断定序列适合 MA 模型; 若平稳序列的偏相关函数和自相关函数均是拖尾的, 则序列式适合 ARMA 模型.

(四) 进行参数估计, 检验是否具有统计意义. 对模型的参数进行估计的方法通常有相关矩估计法, 最小二乘估计以及极大似然估计等.

(五) 进行假设检验, 诊断残差序列是否为白噪声.

(六) 利用已通过检验的模型进行预测分析.

具体流程如下图所示:

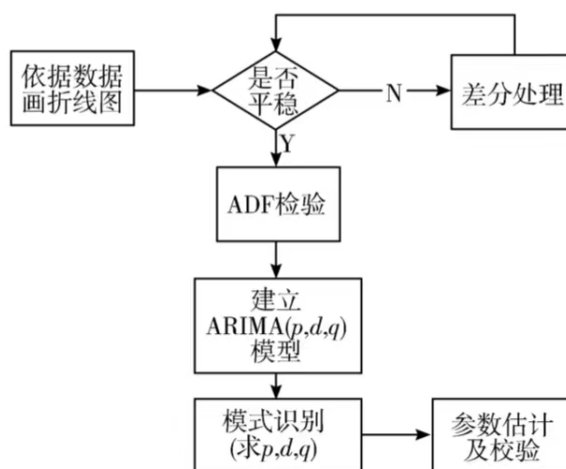


图 12: ARIMA 模型流程图

4.3.2 基于 ARIMA 的 SEIR-ARIMA 疫情预测模型

SEIR 模型可以较好地预测疫情的发展趋势,但是对于初始的传染概率等外在因素也只是估算的,在国家采取疫情防控的措施下,传染概率也会随之降低,因此 SEIR 无法根据历史的波动数据,进行模型的自我调节. ARIMA 模型虽然是基于时间序列预测的模型之一,但是具有由内生变量所决定不受外界因素干预的优点,很好地弥补了 SEIR 模型的弊端. 因此,为了更加准确地预测疫情的发展趋势,通过对 SEIR 模型预测结果和 ARIMA 模型预测结果进行线性回归建模,形成了 SEIR-ARIMA 混合模型. SEIR-ARIMA 混合模型的优势是既学习了一般传染病的发展趋势,又避免了后期因管控措施,病毒变异等影响,造成外生变量的改变,而进行的分阶段划分分析,可以根据历史数据的发展趋势,统筹地预测接下来疫情发展的优点. SEIR-ARIMA 模型是通过 SEIR 模型和 ARIMA 模型预测结果线性回归建模,而线性回归是研究因变量与自变量之间关系,在线性回归中,自变量可以是连续的,也可以是离散的. 通过一元线性方程或者多元线性方程,找到 SEIR 模型预测值与 ARIMA 模型预测值之间的内部关系,形成最佳的线性拟合直线. SEIR-ARIMA 模型进行预测步骤如下:

1) SEIR 模型预测.

通过 2022 年 3 月 11 日至 2022 年 5 月 1 日疫情数据进行分析, 结合相关研究以及参数的优化, 得到其预测值 f .

2) ARIMA 模型预测.

将疫情历史数据作为输入值, 检验时间序列数据是否平稳化, 采用 ADF 验证假设. 若不平稳就通过一元差分, 二元差分等方法, 直到数据平稳化. 通过计算 ACF 和 PACF 得到 p, q 的值, 建立 ARIMA 模型. 训练样本集, 优化模型, 得到预测值 e .

3) SEIR-ARIMA 模型构建.

SEIR-ARIMA 模型构建以上述所得的 2 种模型的预测值 f, e 为输入值, 真实数据为目标值, 采用线性回归建模, 计算 SEIR-ARIMA 混合模型的预测值.

$$\hat{y} = kx + b$$

其中 \hat{y} 表示 SEIR-ARIMA 混合模型的预测值, b 表示偏移量, k 表示特征矩阵的系数, x 表示特征值 $[f, e]$, 即表示 SEIR 和 ARIMA 预测值.

求解系数 k 的过程:

$$k = \frac{1}{n} \sum_{i=1}^n [\hat{y}_i / y]$$

即预测值与真实值的倍速, 求解偏移数据 b ,

$$b = \frac{1}{n} \sum_{i=1}^n \hat{y}_i - kx$$

经过 MATLAB 程序运算, 得到如下的结果:

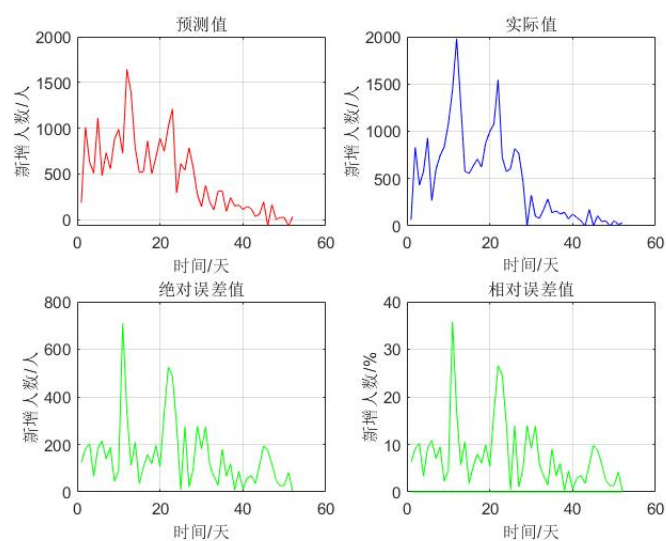


图 13: SEIR-ARIMA 模型结果

5 灵敏度分析

5.1 个体措施

5.1.1 SEIR 模型

由于长春疫情系统中, 3 月 11 日至 5 月 1 日并非自然传播的系统, 个人的接触人员数已经因政策限制而限制到最小, 因此个体措施仅能改变传播率. 在政策封控条件下, 我们代入两组不同的初值进行对比.

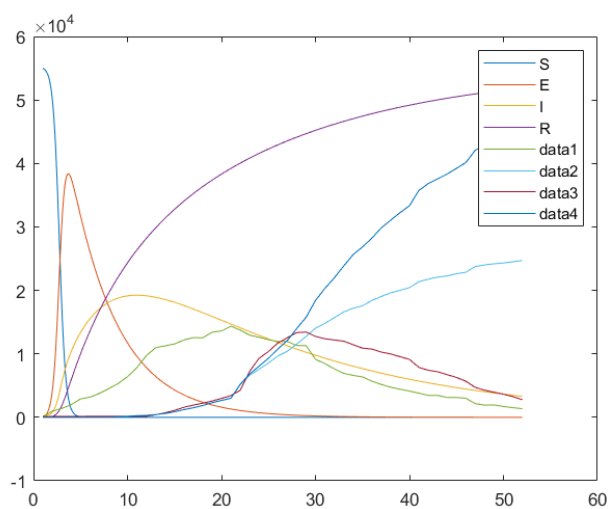


图 14: $r_1 = 5$, $b_1 = 0.8$, $r_2 = 10$, $b_2 = 0.3$, $a = 0.1$, $c_1 = 0.05$, $c_2 = 0.1$,
 $N = 55000$, $init = 200$, $p = 0.9$

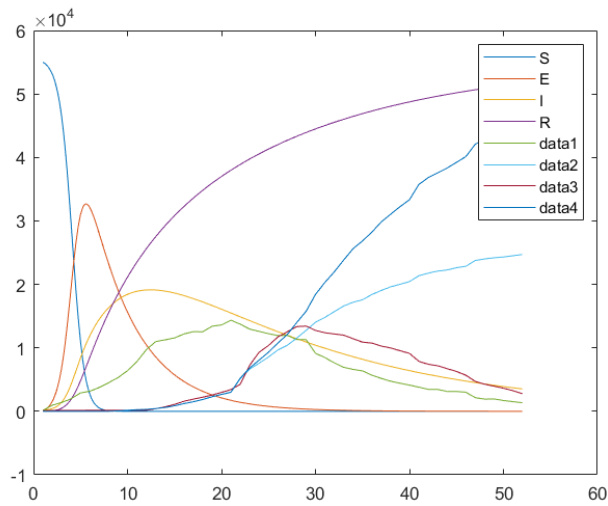


图 15: $r_1 = 5$, $b_1 = 0.9$, $r_2 = 10$, $b_2 = 0.1$, $a = 0.1$, $c_1 = 0.05$, $c_2 = 0.1$,
 $N = 55000$, $init = 200$, $p = 0.9$

可以见到, 除了对于无症状感染者的峰值有少许影响外, 整体的疫情理

论值都没有发生什么太大变化. 因此可以判定, 严格的封控已经大幅消除了个体主观对于整体曲线的影响.

5.2 政策影响

5.2.1 SEIR 模型

由于数据均来自封控阶段, 不存在封控时间平移的影响, 因此我们考虑封控严格程度所带来的系统规模的变化引起的疫情数据变化.

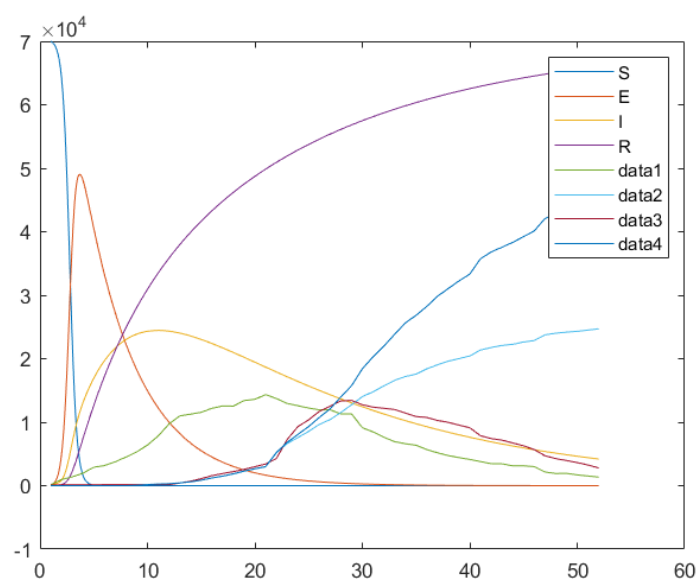


图 16: $r_1 = 5$, $b_1 = 0.9$, $r_2 = 10$, $b_2 = 0.3$, $a = 0.1$, $c_1 = 0.05$, $c_2 = 0.1$, $N = 70000$, $init = 200$, $p = 0.9$

若系统规模增加至涉及 70000 人, 则无症状感染者峰值增加 1000 人左右, 确诊人数增加 500 人左右, 分别为 66% 和 33% 的波动.

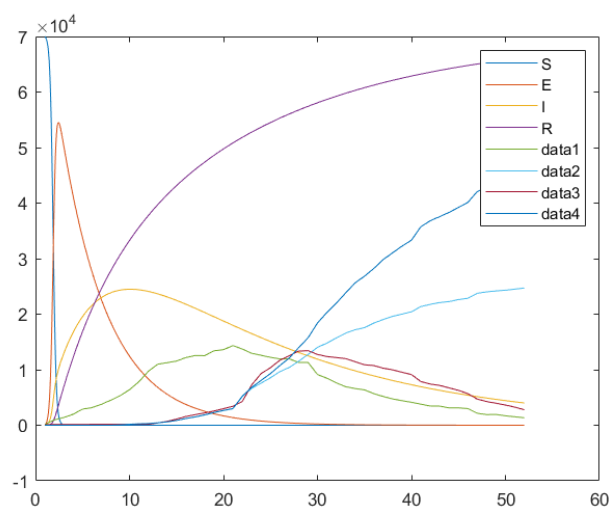


图 17: $r_1 = 10$, $b_1 = 0.9$, $r_2 = 20$, $b_2 = 0.1$, $a = 0.1$, $c_1 = 0.05$, $c_2 = 0.1$,
 $N = 70000$, $init = 200$, $p = 0.9$

若系统规模增加至涉及 70000 人, 与此同时局部限制放松, 接触人员数翻倍, 则无症状感染者峰值再增加 500 人左右, 其它趋势影响较小.

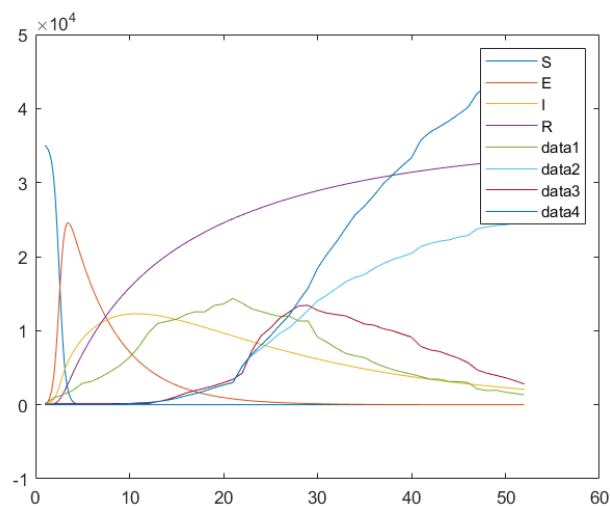


图 18: $r_1 = 5$, $b_1 = 0.9$, $r_2 = 10$, $b_2 = 0.1$, $a = 0.1$, $c_1 = 0.05$, $c_2 = 0.1$,
 $N = 35000$, $init = 200$, $p = 0.9$

若疫情系统规模能短时间内控制在 35000 人内, 则无症状感染者峰值减少 2000 人左右, 确诊峰值减少 1000 人左右. 但当初始确诊人数为 200 的情况下, 将波及人群限制在 35000 人以内是十分困难的.

5.2.2 基于 ARIMA 的 SEIR-ARIMA 疫情预测模型

为了探究政策对长春市疫情的影响, 针对本 SEIR-ARIMA 疫情预测模型, 我们主要分析了绝对误差值与相对误差值的变化, 以此来分析模型在不同政策下的影响效果.

在 SEIR-ARIMA 模型的程序运算下, 绝对误差值与相对误差值的变化情况如下图所示:

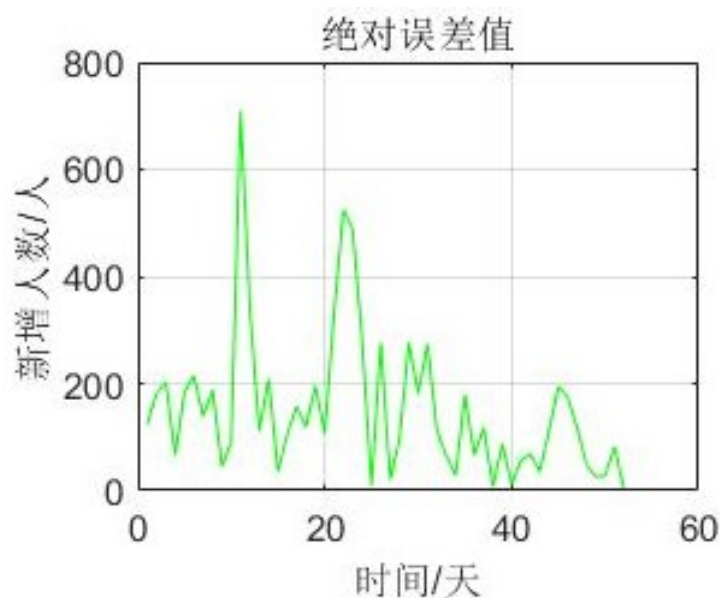


图 19: 绝对误差值

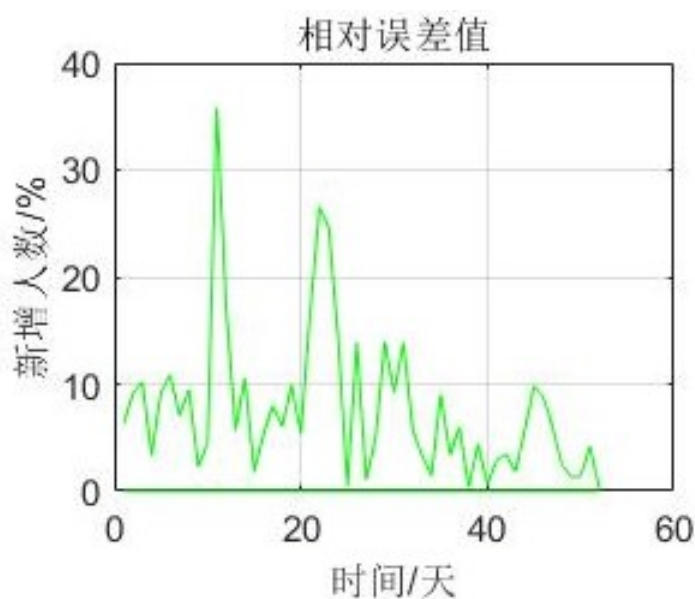


图 20: 相对误差值

通过两图的比较分析可知, 绝对误差值与相对误差值总体上保持稳定且逐渐减小, 在长春市制定政策后, 两误差值发生明显的增加. 我们可以据此分析得, SEIR-ARIMA 模型对具体政策的反映较为灵敏.

6 模型的评价与推广

6.1 模型优点

- SEIR 模型具有一般性和扩展性, 揭示了自然条件下传染病的基本要素.
- 其通解没有表达式, 但可以通过简单的方程数值解模拟参数影响.

6.2 模型不足与推广思路

SEIR 模型中参数具有一般性, 使用单一模型会出现以下问题:

- 单一模型中难以考虑不同年龄段, 疫苗接种, 地区等因素产生的差异.
- 无法或难以考虑不同区域因政策不同而产生的差异.
- 疫情的传播可以看作连续系统, 但病毒的筛查, 病人的收治以天为周期, 将对连续系统中变量产生影响, 产生时滞等现象.

因此在考虑政策的强力影响下, 若仍使用整体疫情传播的观点, 可以使用离散系统建模, 或使用人群密度蒙特卡洛模拟, 统计模型; 若考虑局部, 可采用随机游走主体建模.

6.3 数据问题

在数据的收集与处理时, 有以下多个方面会影响建模过程

- 时间因素的数据缺失: 在分区划的数据中, 长春市 3 月 17 日前的数据不存在行政区划的划分, 因此难以进行此前时间的区域建模.
- 区域因素的数据缺失: 公布的数据中, 不同区划的数据仅包含新增确诊病例数以及新增无症状感染者数, 而不含治愈人数和解除医疗观察人数; 且无症状转为确诊病例的人数仅有长春市总数的统计, 不同区划的疫情数据难以准确得出.
- 数据收集不均: 由于新冠病毒核酸检测为周期性检测, 因此每日的新增病例数据会出现周期性波动的情况, 数据收集不均.
- 数据公布不均: Omicron BA.2 的传播特征为以无症状为主, 但在疫情前期却出现全部为” 确诊病例” 的现象, 而后期” 无症状感染者” 的数据又出现集中井喷的现象. 因此不排除病例归类标准变化, 和数据延迟公布的情况.

参考文献

- [1] Pharma Times Online *Omicron BA.2 variant triggers COVID-19 surge in England* [EB/OL]. 17th March 2022. <https://www.pharmatimes.com/>.
- [2] Diario AS. Oliver Povey. *What is the incubation period for Omicron BA.2 and how many days are you contagious with this variant?* [EB/OL]. April 5th, 2022 13:00 EDT. <https://en.as.com/>
- [3] 百度: 新型冠状病毒肺炎疫情实时大数据报告 [EB/OL]. <https://voice.baidu.com/act/newpneumonia/newpneumonia/>.
- [4] 吉林省卫生健康委员会: 吉林省卫生健康委关于新型冠状病毒肺炎疫情情况通报 [EB/OL]. <http://wsjkw.jl.gov.cn/xwzx/xwzx/>.
- [5] Harvard Health Publishing: *Most COVID-19 cases are spread by people without symptoms* [EB/OL]. www.health.harvard.edu.
- [6] 董章功, 宋波, 孟友新. 基于 SEIR-ARIMA 混合模型的新冠肺炎预测 [J]. 计算机与现代化, 2022(02):1-6.
- [7] 鲍昕, 谭智一, 鲍秉坤, 徐常胜. 基于时空注意力机制的新冠疫情预测模型 [J]. 北京航空航天大学学报:1-11.
- [8] 任晓龙, 李忠, 申天恩, 毛亦鹏, 宋俊杰. Logistic 回归模型对美国新冠疫情预测研究 [J]. 福建电脑, 2021, 37(04):47-49.

A 附录

A.1 代码

文件 test_err.m 最小二乘法, 结果发散

```
1      function Parameter=test_err(k)
2      OriginalDATA = SEIR_readi_swi(k);
3      Parameter=zeros(8);
4      testDATA =zeros(52,1);
5      for r1=5:1:20
6      for b1=0.4:0.05:0.5
7      for r2=r1:1:30
8      for b2=0.4:0.05:b1
9      for a=0.3:0.05:0.6
10     for c = 0.05:0.05:0.2
11     for N=20000:2000:80000
12     [t,x] = SEIR_model3(r1,b1,r2,b2,a,c,N,OriginalDATA(1));
13     for m=1:1:52
14     testDATA(m)=x(find_nearest(m,x(:,3)),3);
15     end
16     delta = errfunc(testDATA,OriginalDATA);
17     if (Parameter(8)==0 || delta < Parameter(8))
18     Parameter = [r1,b1,r2,b2,a,c,N,delta];
19     end
20     end
21     end
22     end
23     end
24     end
25     end
26     end
27     end
28
```

```
29     function delta = errfunc(x,y)
30     delta = norm(x-y,2);
31     end
```

文件 SEIR_model3.m 使用 ODE 数值解拟合配合最小二乘法

```
1     % 感染者接触的人数为r1
2     % 传染的概率为b1
3     % 潜伏患病的概率为a
4     % 潜伏接触人数为r2
5     % 潜伏传染概率为b2
6     % 康复概率为c
7     function [t,x] = SEIR_model3(r1,b1,r2,b2,a,c,N,init);
8     [t,x]=ode45(@trans,[1,52],[N,0,init,0]);
9     plot(t,x);
10    legend('S','E','I','R');
11    function dx=trans(t,x);
12    dx = [-r1*b1*x(1)*x(3)/N-r2*b2*x(1)*x(2)/N;
13    r1*b1*x(1)*x(3)/N+r2*b2*x(1)*x(2)/N-a*x(2);
14    a*x(2)-c*x(3);
15    c*x(3)];
16    end
17    end
```

```
1     function index = find_nearest(c,x)
2     [~,index] = min(abs(x-c));
3     end
```

```
1     function DATA = SEIR_readi_swi(k)
2     [SEIR_I,~,~]=xlsread('SEIR_I.xlsx');
```

```
3     DATA = SEIR_I(:,k);  
4     end
```

文件 SEIR_model5.m 用于 SEIR 模型拟合长春市数据

```
1     % 感染者接触的人数为 r1  
2     % 传染的概率为 b1  
3     % 潜伏患病的概率为 a  
4     % 潜伏接触人数为 r2  
5     % 潜伏传染概率为 b2  
6     % 康复概率为 c  
7     % 概率 p 是无症状  
8     function [t,x] = SEIR_model5(r1,b1,r2,b2,a,c1,c2,N,init  
9         ,p);  
10  
11     [SEIR_I,~,~]=xlsread('SEIR_IA.xlsx');  
12     DATA = SEIR_I(:,1:4);  
13  
14     [t,x]=ode45(@trans,[1,52],[N,0,init,0]);  
15     plot(t,x);  
16     legend('S','E','I','R');  
17     hold;  
18     plot([1:52],DATA);  
19     function dx=trans(t,x)  
20     dx = [-r1*b1*x(1)*x(3)/N-r2*b2*x(1)*x(2)/N;  
21     p*(r1*b1*x(1)*x(3)/N+r2*b2*x(1)*x(2)/N)-a*x(2)-c2*x(2);  
22     a*x(2)-c1*x(3)+(1-p)*(r1*b1*x(1)*x(3)/N+r2*b2*x(1)*x(2)  
23     /N];  
24     c1*x(3)+c2*x(2)];  
25     end  
26     end
```

文件 AMRIMA.m 用于 SEIR-AMRIMA 预测长春市数据

```
1
2      %%%%%%%%%%原始数据:
3      y=xlsread('incre_num.xlsx');
4
5
6      step=6;%后6个测试
7      %step=23 ;%后23 个测试
8      TempData=SourceData;
9      TempData=detrend(TempData);%去趋势线
10     TrendData=SourceData-TempData;%趋势函数
11     %-----差分, 平稳化时间序列-----
12     H=adftest(TempData);
13     difftime=0;
14     SaveDiffData=[];
15     while ~H
16         SaveDiffData=[SaveDiffData, TempData(1,1)];
17         TempData=diff(TempData);%差分, 平稳化时间序列
18         difftime=difftime+1;%差分次数
19         H=adftest(TempData);%adf检验, 判断时间序列是否平稳化
20     end
21     %-----模型定阶或识别-----
22     u = iddata(TempData);
23     test = [];
24     %for p = 1:5%自回归对应PACF, 给定滞后长度上限p和q, 一般
        取为T/10, ln(T)或 $T^{(1/2)}$ , 这里取T/10=12
25     for p = 1:5%自回归对应PACF, 给定滞后长度上限p和q, 一般取
        为T/10, ln(T)或 $T^{(1/2)}$ , 这里取T/10=12
26     %for q = 1:5%移动平均对应ACF
27     for q = 1:5%移动平均对应ACF
28         m = armax(u, [p q]);
29         AIC = aic(m);%arimax(p,q), 计算AIC
30     test = [test;p q AIC];
```

```
31     end
32     end
33     test1=[];
34
35
36     for p=1:5;%%%%%%%%%%%%%%
37     for q=1:5;%%%%%%%%%%%%%%
38     for k = 1:size(test,1)
39     if test(k,3) == min(test(:,3)) %选择AIC值最小的模型
40     p_test = test(k,1);
41     q_test = test(k,2);
42
43     m = armax(u,[p q]);%
44     AIC = aic(m);
45     m1 = armax(u,[p_test q_test]);%
46     BIC = aic(m1);
47     test1=[test1;p q p_test q_test AIC BIC ];%
48         %%%%%%%%%%%%%%%
49
50     break;
51     end
52     end
53     end %%%%%%%%%%%
54     end %%%%%%%%%%%
55
56
57
58
59
60     %-----1阶预测-----
61
62
```



```
63     TempData=[TempData;zeros(step,1)];
64     n=iddata(TempData);
65     %p_test=1;q_test=2;
66     m = armax(u,[p_test q_test]);
67     %m = armax(u(1:ls),[p_test q_test]);? ?? ???%armax(p,q)
        , [p_test q_test]对应AIC值最小, 自动回归滑动平均模型
68     P1=predict(m,n,1);
69     PreR=P1.OutputData;
70     PreR=PreR';
71     %-----还原差分-----
72     if size(SaveDiffData,2)~=0
73     for index=size(SaveDiffData,2):-1:1
74     PreR=cumsum([SaveDiffData(index),PreR]);
75     end
76     end
77     %-----预测趋势并返回结果-----
78     mp1=polyfit([1:size(TrendData',2)],TrendData',1);
79     xt=[];
80     for j=1:step
81     xt=[xt,size(TrendData',2)+j];
82     end
83     TrendResult=polyval(mp1,xt);
84     PreData=TrendResult+PreR(size(SourceData',2)+1:size(
        PreR,2));
85     tempx=[TrendData',TrendResult]+PreR;% tempx为预测结果
86
87
88
89
90     figure(1)
91     %plot(tempx,'r');%红线预测值
92     subplot(2,2,1),plot(tempx,'r');%红线预测值;
93     hold on
```

```
94     grid;
95     title('预测值');
96     xlabel('时间/天');ylabel('新增人数/人');
97
98
99
100    %figure(2)
101    subplot(2,2,2),plot(Data,'b');%蓝色线实际值
102    hold on
103    grid;
104    title('实际值');
105    xlabel('时间/天');ylabel('新增人数/人');
106
107
108
109
110    %figure(3)
111    t=tempx';
112    erro=abs(Data-t);
113    subplot(2,2,3),plot(erro,'g');%误差值
114    hold on
115    grid;
116    title('绝对误差值');
117    xlabel('时间/天');ylabel('新增人数/人');
118
119
120    h = 0.001;    % 步长，取不同的值，结果会有变化
121    x = -30:h:30;
122    z1 = abs(Data-t);
123    zmax1 = max(z1); % 最大绝对误差的值
124    zpjun1=mean(z1) ;% 平均绝对误差的值
125
126
```

```
127
128     %figure(4)
129     t=tempx';
130     erro=abs(Data-t);
131     Erro=(erro/Data)*100;
132     subplot(2,2,4),plot(Erro,'g');%误差值
133     hold on
134     grid;
135     title('相对误差值');
136     xlabel('时间/天');ylabel('新增人数/%');
137
138
139     h = 0.001; % 步长，取不同的值，结果会有变化
140     x = h:1;
141     z2 = abs((Data-t)/Data);
142     zmax2 = max(z2) ; % 最大相对误差的值
143     zpjun2=mean(z2); % 平均相对误差的值
144
145
146
147
148
149     %plot(tempx,'r');%红线预测值
150     %hold on
151     %plot(Data,'b');%蓝色线实际值
152     %hold on;
153     %t=tempx';
154     %erro=Data-t;
155
156     %plot(erro,'g');
157     %hold on;
```