

# 基于SIR模型预测新型冠状病毒在广东省的传播

张 帅, 吕浩嘉, 邹怀川, 张荣培\*

广东工业大学先进制造学院, 广东 揭阳

收稿日期: 2024年4月29日; 录用日期: 2024年5月22日; 发布日期: 2024年5月31日

## 摘 要

自2019年底以来,全球爆发的COVID-19疫情对全球公共卫生系统带来了前所未有的挑战。传染病模型,尤其是SIR模型,成为疾病传播和防控研究的关键工具。本文采用数据驱动的建模方法研究新型冠状病毒在广东省的传播,首先采集广东省疫情初期的数据,然后通过采用有限差分方法和数据拟合求解得到SIR模型的两个重要参数:传染率和康复率,最后应用龙格库塔方法对SIR模型进行数值求解。本文的模型可以更好地理解疫情传播过程,为决策者提供科学依据,有助于政府和卫生机构制定更合理的防控措施和资源分配策略。

## 关键词

新型冠状病毒, SIR, 龙格库塔方法, 有限差分

# SIR Model-Based Prediction of Infected Population of Corona Virus in Guangdong Province

Shuai Zhang, Haojia Lv, Huaichuan Zou, Rongpei Zhang\*

School of Advanced Manufacturing, Guangdong University of Technology, Jieyang Guangdong

Received: Apr. 29<sup>th</sup>, 2024; accepted: May 22<sup>nd</sup>, 2024; published: May 31<sup>st</sup>, 2024

## Abstract

Since the end of 2019, the global outbreak of the COVID-19 pandemic has presented unprece-  
\*通讯作者。

mented challenges to the global public health system. The SIR model have become essential tools for researching disease transmission and control. This paper employs a data-driven modeling approach to study the spread of the novel coronavirus in Guangdong Province. Initially, data from the early stages of the pandemic in Guangdong Province is collected. Subsequently, by utilizing finite difference methods and data fitting, two crucial parameters of the SIR model, namely the transmission rate and recovery rate, are obtained. Finally, the Runge-Kutta method is applied for numerical solutions of the SIR model. The model presented in this paper offers a better understanding of the pandemic transmission process, providing decision-makers with a scientific foundation. This contributes to the formulation of more rational prevention and control measures, as well as resource allocation strategies by governments and health organizations.

## Keywords

COVID-19, SIR, Runge-Kutta Method, Finite Difference Method

Copyright © 2024 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

2019 年底以来, 全球爆发的冠状病毒疫情, 即 COVID-19, 给全球各地的公共卫生系统带来了前所未有的巨大挑战。在这样的背景下, 使用传染病模型成为了疾病传播和防控研究的重要工具之一, 其中 SIR 模型(Susceptible-Infectious-Removed)是常用的一种[1]。SIR 模型可以通过建立基本的传染病传播方程, 预测病毒在人群中的传播趋势。通过对模型参数的调整, 可以预测感染者数量、康复者数量以及潜在易感者的变化, 从而帮助决策者制定合理的防控措施。本文通过收集广东省公开的疫情数据, 使用 SIR 模型对广东省疫情状况进行初步拟合并预测疫情走向。这对于更好地理解病毒性质以及提高对未来疫情的准备具有重要意义。

SIR 模型是一种常见的传染病传播模型, 用于描述人群中传染病的传播过程。其基本假设是将人群分为三个类别: 易感染者(Susceptible)、感染者(Infectious)和康复者或免疫者(Recovered)。易感染者是没有接触过病毒或者没有获得免疫力, 感染者是指已经被冠状病毒感染的人群, 康复者或免疫者是指已经康复并获得免疫力的人群。SIR 模型基于这三个类别之间的人群转移过程, 通过一组微分方程描述了疾病在人群中的传播动态。

本文选用 SIR 模型对广东省疫情进行参数估计及感染人数预测。SIR 模型涉及的变量有: 区域总人数  $N$ 、感染人数  $I$ 、易感人数  $S$ 、治愈及死亡总人数  $R$ 、时间  $t$ , 它们有以下关系:

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{IS}{N} \\ \frac{dI}{dt} = \beta \frac{IS}{N} - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (1)$$

其中参数  $\beta$  (传染率)和  $\gamma$  (康复率)是常数, 反应了疫情的传播特征。

在过去的几年里, 国内外的学者们在传染病扩散模型和分析方法方面做出了许多重要的工作。文献

[2]进行了新冠病毒感染在湖北省扩散的 SIR 模型预测,文献[3]基于 SIR 传染病模型的印度新冠疫情波及影响分析,文献[4]研究了 SIR 传染病模型在新冠疫情分析中的应用,文献[5]对新冠疫情的扩散进行了数学建模与分析,文献[6]进行了新冠肺炎传染病模型的建立及算法研究, SIR 模型在遇到传染病引发的疫情等方面都有着广泛的应用。需要指出的是,上述的部分文献在求解 SIR 模型参数时大多都使用经验参数进行求解,缺乏一定的事实依据,可能使模型预测能力不足以满足工作。所以本文采用由数据对参数进行拟合的方法,提高准确性,与以上文献相比较更具有事实依据。

本文采用 SIR 模型,为了求出基于实际数据拟合出的参数,首先通过有限差分方法求出数据的近似一阶导数,将其代入公式求出传染率和康复率的实际数据的离散值,再利用非线性最小二乘法拟合求解出传染率和康复率关于时间的变化趋势。然后利用龙格库塔方法求解微分方程,进行未来疫情的预测,这是一种典型的数据驱动的建模方法。这种方法的优势在于简单而直观,同时通过实际数据的拟合,可以更好地反映疫情的实际情况,更具有事实依据。通过本文给出的方法,可以更好地理解疫情的传播过程,为实际应对提供科学依据。有助于政府和卫生机构做出更合理的决策,合理分配资源。

## 2. 基于实际数据的模型参数估计

在方程组(1)中,传染率  $\beta$  及康复率  $\gamma$  是未知的参数,本文对参数  $\beta$  和  $\gamma$  进行拟合,得到  $\beta$  和  $\gamma$  关于时间的函数,从而得到完整的 SIR 模型。通过收集广东省早期的公开疫情数据,包括感染者,康复者人数,以及死亡人数,可以计算出模型所需的感染率  $\beta$  和治愈率  $\gamma$ 。

由方程组(1)式可得  $\beta$  和  $\gamma$  的表达式:

$$\begin{cases} \beta = -\frac{N}{IS} \frac{dS}{dt} \\ \gamma = \frac{1}{I} \frac{dR}{dt} \end{cases} \quad (2)$$

通过对感染者  $I$  和康复者  $R$  关于时间求导数,可得出  $\beta$  和  $\gamma$  的近似值,基于已有的数据,对感染者  $I$  和康复者  $R$  进行有限差分进行推导。

设函数  $X(t)$  在  $t_0, t_1, \dots, t_n$  (取等间距  $\Delta t$ ) 的函数值为  $X_0, X_1, \dots, X_n$ , 则可以推出函数  $S(t)$  在个点的导数近似值为:

$$X'(t_k) = \frac{X_{k+1} - X_{k-1}}{2}, k = 1, 2, \dots, n-1$$

对于  $t_0$  和  $t_n$  有:

$$X'(t_0) = \frac{-3x_0 + 4x_1 - x_2}{2}, X'(t_n) = \frac{x_{n-2} - 4x_{n-1} + 3x_n}{2}$$

又因为实际数据中时间间隔为一天,因此取  $\Delta t = 1$ 。由以上推导得:

$$\begin{cases} S'(t_k) = \frac{S_{k+1} - S_{k-1}}{2}, k = 1, 2, \dots, n-1 \\ R'(t_k) = \frac{R_{k+1} - R_{k-1}}{2}, k = 1, 2, \dots, n-1 \end{cases} \quad (3)$$

同理,对于  $t_0$  和  $t_n$  有:

$$S'(t_0) = \frac{-3s_0 + 4s_1 - s_2}{2}, S'(t_n) = \frac{s_{n-2} - 4s_{n-1} + 3s_n}{2}$$

$$R'(t_0) = \frac{-3R_0 + 4R_1 - R_2}{2}, R'(t_n) = \frac{R_{n-2} - 4R_{n-1} + 3R_n}{2}$$

以上即对易感者  $S$  和康复者  $R$  进行有限差分得出近似一阶导数, 将式(3)代入方程(2)可以求出  $\beta$  和  $\gamma$  的实际离散值。图 1 以样本量为 40 天的数据, 通过差分求解得到关于时间的变化趋势。

由于后续进行的操作中需要避开数值为零的数据, 所以提前将值为零的数据进行预处理。由于零值较少, 本文使用线性插值作为预处理方法, 即将零值替换为相邻非零值的平均值。

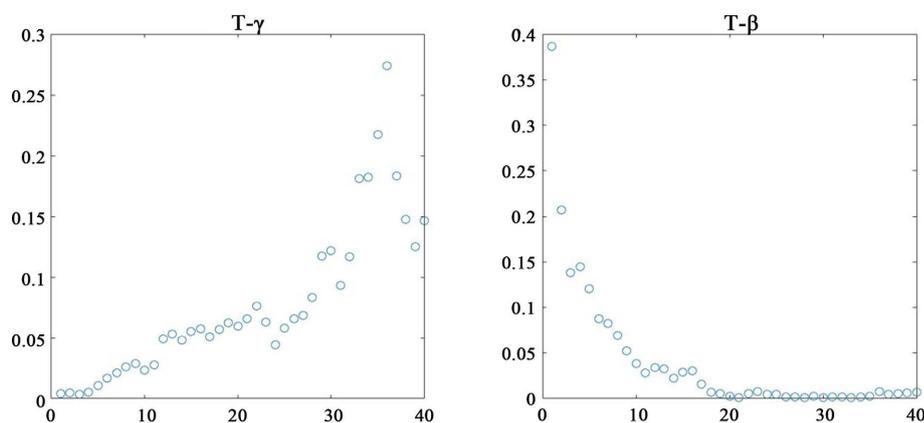


Figure 1. Trends of  $\beta$  and  $\gamma$  over time  
图 1.  $\beta$  和  $\gamma$  关于时间的变化趋势

在求解得到  $\beta$  和  $\gamma$  离散数据后, 下面将利用数据拟合的方法, 求出  $\beta$  和  $\gamma$  的函数表达式。通过观察两个图像(图 2), 可以看出  $\gamma$  和  $\beta$  是关于时间的指数型函数。假设  $\beta$  和  $\gamma$  关于时间的函数满足以下关系:

$$\begin{cases} \beta = a_1 e^{b_1 t} \\ \gamma = a_2 e^{b_2 t} \end{cases} \quad (4)$$

在 Matlab 中选择使用 `lsqcurvefit` 指令求得(3)式  $a_1, b_1, a_2, b_2$ 。该指令需要初始的参数估计作为初始值, 对(3)式两边同取对数后, 即可使用线性最小二乘法得到一个相对准确的参数作为 `lsqcurvefit` 的初始值。本文为了更精确的结果, 将第一次拟合的结果作为初始值再进行一次迭代从而最终得出结果, 如图 2 所示。指令的返回值即为所求的  $a_1, b_1, a_2, b_2$ 。

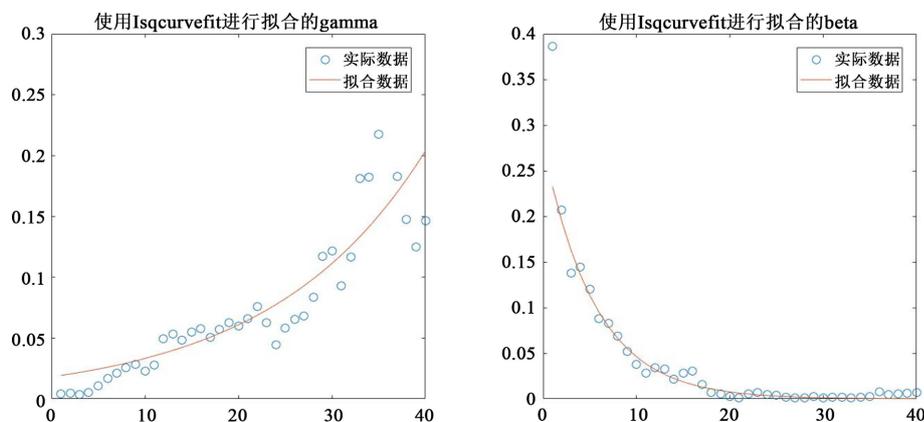


Figure 2. The fitting results and real data of parameter  $\beta$  and  $\gamma$   
图 2.  $\beta$  和  $\gamma$  的拟合结果与实际数据对比

### 3. SIR 模型的求解

将上一部分所求的  $a_1$ ,  $b_1$ ,  $a_2$ ,  $b_2$  代入式(3), 再将其代入 SIR 模型的微分方程(1)中, 这样就得到一组非线性常微分方程, 其中  $\beta$  和  $\gamma$  都是拟合得出的函数。为了求解这个非线性常微分方程组, 可以使用数值方法, 如龙格库塔方法。命令 `ode45` 是一个高效的工具, 用于求解非刚性的常微分方程初值问题。通过输入相应的微分方程和初始条件, `ode45` 可以近似地计算出  $S$ 、 $I$  和  $R$  随时间的变化。

图 3(a)是使用 35 天的样本数据为例, 并使用得到的模型来预测未来 50 天的疾病发展情况得出的预测与实际数据的对比图。根据 SIR 模型中模拟, 可以看到疫情的发展过程。

初始阶段, 感染者的数量逐渐增加, 易感染者的数量逐渐减少。随着时间的推移, 感染者数量逐渐达到峰值, 然后开始下降, 康复者的数量逐渐增加。最终, 感染者数量趋于稳定, 康复者或免疫者的数量增加到一定程度。为例提高模型的准确性, 我们又使用 42 天的样本数据重新计算, 如图 3(b)所示。通过两个图对比, 可以明确地看到, 随着样本量的增加, 参数估计的准确性有了显著的提升。

这里的预测结果可以看出, 实际数据符合 SIR 模型, 更长时间的样本拟合的参数可以使结果的拟合得更精确。

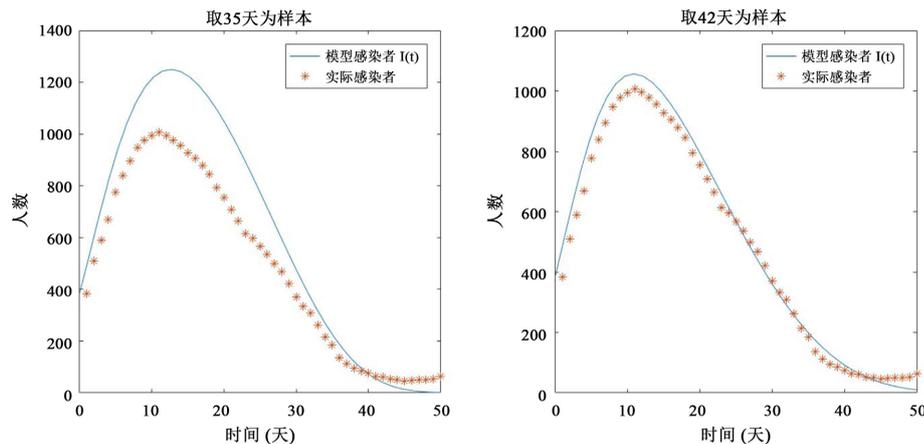


Figure 3. The comparison between real data and fitting results of 35 and 42 days

图 3. 样本为 35 和 42 天的拟合结果与实际数据的对比图

### 4. 结论

在广东省的疫情爆发后所实施的疫情防控措施可以看作是通过人为调控  $\beta$  和  $\gamma$  来抑制疫情的传播。为了降低  $\beta$  的值, 可以通过加强对病例追踪和封控措施, 减少人与人之间的接触, 从而减少感染者的数量。同时, 为了增加  $\gamma$  的值, 通过加强医疗资源配置, 提高治愈率, 从而增加康复者的数量。

本文通过收集广东省疫情初期的公开数据, 拟合出 SIR 模型的所需参数并通过求解模型进行预测, 最后可以发现 SIR 模型预测与实际数据并无较大偏差, 并成功预测了疫情的峰值和拐点, 在以后面对类似于此次大规模的传染性疾病时可以拟合出参数并运用 SIR 模型去预测疫情的变化趋势, 通过预测值得出疫情的拐点和发展趋势, 利于政府实时监控疫情走向, 可以更好地控制疫情的传播。

与其他使用经验参数的方法相比较本文的优势在于利用函数拟合的方法求解参数, 这种方法得出的模型基于实际数据, 更加具有事实依据, 有利于决策者更好地实施政策以控制疫情。

然而, 需要注意的是, 本文所用的 SIR 模型是一个比较简单的模型, 实际疫情的发展过程受到各种复杂因素的影响, 因此在实际应用中需要结合具体情况进行调整和完善。在以后的工作中, 我们将考虑

多种因素，建立更为复杂的模型进行精确的模拟。

## 基金项目

2022 年校级“本科教学工程”项目(跨校区环境下《数学建模》课程的教学改革与实践——以揭阳校区培养拔尖人才为例)；

2023 年校级“大学生创新创业”项目(xj2023118451026：研究分数阶微分方程在传染病模型建立中的应用)。

## 参考文献

- [1] Balderrama, R., Peressutti, J., Pinasco, J.P., *et al.* (2022) Optimal Control for a SIR Epidemic Model with Limited Quarantine. *Scientific Reports*, **12**, 12583. <https://doi.org/10.1038/s41598-022-16619-z>
- [2] 张钧, 王立洪, 王骥. 新型冠状病毒感染在湖北省扩散的 SIR 模型预测[EB/OL]. <http://dx.doi.org/10.13140/RG.2.2.35244.00648>, 2020-02-24.
- [3] 丁黄艳, 铁禧玥. 基于 SIR 传染病模型的印度新冠疫情波及影响分析[J]. 重庆工商大学学报(自然科学版), 2022, 39(5): 70-71.
- [4] 刘小扬. SIR 传染病模型在新冠疫情分析中的应用[D]: [硕士学位论文]. 青岛: 山东大学, 2022.
- [5] 卜兴业. 新冠肺炎疫情扩散的数学建模与分析[D]: [硕士学位论文]. 合肥: 中国科学技术大学, 2022.
- [6] 李猛, 谷乐川. 新冠肺炎传染病模型的建立及算法研究[J]. 许昌学院学报, 2023, 42(2): 15-20.