

## 基于改进遗传算法的 COVID-19 传染病模型参数反演

李佳琴, 李虎飞

北方民族大学数学与信息科学学院 宁夏银川

**【摘要】**本文研究了 COVID-19 传染病模型的参数反演问题, 首先建立 COVID-19 无症状感染者传染病的 SEAIRD 模型, 其次, 利用反问题法将模型转化为目标函数的最小问题, 利用改进的遗传算法获取模型参数, 最后对参数进行了敏感性分析. 结果表明, 改进的遗传算法对新冠肺炎疫情模型的参数反演效果良好.

**【关键词】** COVID-19; 传染病模型; 无症状感染者; 参数反演; GA

**【基金项目】** 北方民族大学研究生创新项目 (YCX22103)

**【收稿日期】** 2023 年 1 月 26 日 **【出刊日期】** 2023 年 3 月 21 日 **【DOI】** 10.12208/j.aics.20230018

### Parameter inversion of COVID-19 infectious disease model based on improved genetic algorithm

Jiaqin Li, Hufei Li

School of Mathematics and Information Science, North Minzu University, Yinchuan, Ningxia

**【Abstract】** In this paper, parameter inversion of COVID-19 infectious disease model is studied. Firstly, the SEAIRD model of infectious disease in asymptomatic COVID-19 infected persons was established. Secondly, the inverse problem method is used to transform the model into the minimum problem of the objective function, and the improved genetic algorithm is used to obtain the model parameters. Finally, the sensitivity of the parameters was analyzed. The results show that the improved genetic algorithm is effective in parameter inversion of the novel coronavirus epidemic model.

**【Keywords】** Infectious Disease Model; Asymptomatic Infected Persons; Parameter Inversion; GA

### 1 引言

自 2019 年底以来, 新型冠状病毒肺炎 (COVID-19) 疫情在世界许多国家和地区肆虐<sup>[1,2]</sup>. 截至 2023 年 2 月 6 日, 全球有 7.54 亿确诊病例, 681 万人死亡. 疫情的蔓延严重影响了社会的正常发展和人民生活. 显然, 新冠肺炎疫情对全世界都造成了灾难性的影响.

流行病学, 包括新冠肺炎、艾滋病、疟疾等, 近年来这些疾病的动态演变受到越来越多的关注<sup>[3,4]</sup>. 为了预测 COVID-19 的全球传播趋势, 学者们采用了不同的模型结构, 分析了疫情传播的基本特征, 帮助政府做出应对病毒传播的决策. Carcione 等<sup>[5]</sup>使用改进的 SEIR 模型获得了一种意大利流行病的情景. Ivorra 等<sup>[6]</sup>构建了一种适应 COVID-19 特征的新的  $\theta$ -SEIHRD 模型, 考虑了未发现的感染者和 COVID-19 控制措施的效果. Rapolu 等<sup>[7]</sup>提出了 SEIR-D 模型, 得到了印度部分地区、意大利伦巴第和俄罗斯莫

斯科的时变参数计算结果. 随着对新型冠状病毒疫苗的深入研究和易感人群免疫力的提高, 在感染者中出现的无症状感染越来越多. 参考文献<sup>[8,9]</sup>显示, 对无症状人群的监测检测是预防 COVID-19 传播的关键策略. 杜克大学实施了 SARS-CoV-2 监测项目, 使用汇总检测监测无症状患者, 并对有症状者<sup>[10]</sup>进行个人检测. 因此, 评估无症状 COVID-19 在无症状感染者中的流行率和传染性对于了解疾病的总体流行率至关重要<sup>[11,12]</sup>, 有必要建立更多针对无症状感染者的传染病模型, 并深入开展传染病模型的参数研究.

传染病模型中有许多参数, 如感染者的隔离率、感染者的恢复率、感染的概率等, 这些参数有时难以测量或直接获得. 因此, 当利用传染病动力学的数学模型来反映各种疾病的传播机制, 确定该模型的参数是非常重要的. 反问题<sup>[13,14]</sup>是模型参数控制的关键问题. 研究反问题的方法可以成功地应用于传染病的研究.

究,使流行病学家和公共卫生专家能够对流行病进行预测。Stevens 等<sup>[15]</sup>对复杂 COVID-19 模型控制的逆问题进行了全面的计算研究。这些研究为具体的 COVID-19 逆问题提供了重要的见解,并为数学生物学和其他不同应用领域的逆问题的类似分析提供了指导。通过求解反问题得到的模型参数称为参数反演<sup>[16,17]</sup>。Liu 等<sup>[18]</sup>提出了考虑中国疫情防治措施的 QSEIR 模型,该模型的参数采用统计方法进行估计,然后根据中国的疫情预防策略对疫情情况进行预测。Marinov 等<sup>[19]</sup>解决了 SIR 传染病模型中传播率和去除率随时间变化的反演问题。Godio 等<sup>[20]</sup>建立了应用于 SARS-CoV-2 传播的广义 SEIR 模型,并采用粒子群优化方法寻找模型参数,以提高 30 天预测的可靠性,从以上文献可以看出,关于无症状感染者 COVID-19 传染病模型参数反演的研究仍较少。因此,本文研究了考虑无症状感染者的 COVID-19 传染病模型的参数反演问题。

### 2 COVID-19 传染病模型建立

大部分关于传染病的研究都是基于经典的传染病模型, Krivorotko<sup>[21]</sup>基于 SEIR 模型提出了 SEIR-D 模型,

$$\begin{cases} \frac{dS}{dt} = -c(t-\tau)\left(\frac{\alpha_I S(t)I(t)}{N} + \frac{\alpha_E S(t)E(t)}{N}\right) + \gamma R(t), \\ \frac{dE}{dt} = c(t-\tau)\left(\frac{\alpha_I S(t)I(t)}{N} + \frac{\alpha_E S(t)E(t)}{N}\right) - (\kappa + \rho)E(t), \\ \frac{dI}{dt} = \kappa E(t) - \beta I(t) - \mu I(t), \\ \frac{dR}{dt} = \beta I(t) + \rho E(t) - \gamma R(t), \\ \frac{dD}{dt} = \mu I(t), \end{cases}$$

其中  $S(t)$ 、 $E(t)$ 、 $I(t)$ 、 $R(t)$  和  $D(t)$  分别表示  $t$  时刻易感人数、暴露人数、感染人数、恢复人数和死亡人数,  $\alpha_I$  表示感染者与易感个体之间的感染传播,  $\alpha_E$  表示潜伏期人群与易感人群之间的感染传播,  $\kappa$  表示感染率,  $\gamma$  表示再感染率,  $\rho$  表示潜伏期的恢复率,  $\beta$  表示感染病例的治愈率,  $\mu$  为致死率。

目前已证实 COVID-19 传染病对全人群易感,传播速度和致病性较强。我们知道无症状感染因素可以是感染源,并刺激身体的免疫反应。为了预测和控制传染病模型的准确性,基于 SEIR-D 建立了无症状感染者的 SEAIRD 模型,该模型主要由 6 个非线性常微

分方程组成,其中  $A(t)$  表示  $t$  时刻无症状感染者,参数及符号取值范围与描述如表 1 所示。

$$\begin{cases} \frac{dS}{dt} = -\frac{\alpha_I S(t)I(t)}{N} - \frac{\alpha_A S(t)A(t)}{N} - \frac{\alpha_E S(t)E(t)}{N} + \theta R(t), \\ \frac{dE}{dt} = \frac{\alpha_I S(t)I(t)}{N} + \frac{\alpha_A S(t)A(t)}{N} + \frac{\alpha_E S(t)E(t)}{N} - p\beta_A E(t) - (1-p)\beta_I E(t), \\ \frac{dA}{dt} = p\beta_A E(t) - \gamma_A A(t), \\ \frac{dI}{dt} = (1-p)\beta_I E(t) - \gamma_I I(t) - \mu I(t), \\ \frac{dR}{dt} = \gamma_A A(t) + \gamma_I I(t) - \theta R(t), \\ \frac{dD}{dt} = \mu I(t), \end{cases}$$

### 3 基于精英保留策略的 GA 的参数反演

GA<sup>[22]</sup>最早于 1975 年由荷兰提出,从生物世界的进化规律演化而来的一种随机搜索方法,以自动获得和引导优化后的搜索空间,从而自适应地调整搜索方向。然而,当适应度函数选择不当时,该算法可能收敛于局部最优而不是全局最优。因此,加入了精英保留策略<sup>[23]</sup>,将进化过程中出现的最佳个体直接复制到下一代,而无需配对交叉。现有的研究从理论上证明了具有精英保留的标准遗传算法是全局收敛的。因此,本文选择了具有精英保留策略的遗传算法来解决目标函数的最小值问题,从而弥补了适应度函数不能达到全局最优的问题。

传染病动力学的研究对象是性质随时间变化的非平衡动态系统。系统在一般由微分方程表示的状态空间中按照一定的规律演化。传染病的数学模型可以总结如下:

$$\begin{cases} Y'_i(t) = F_i(t, Y_1, Y_2, \dots, Y_n, q_1, q_2, \dots, q_m) \\ Y_i(t_0) = Y_{i0} \end{cases} \quad (1)$$

将模型写成向量形式:

$$\begin{aligned} Y'(t) &= F(t, Y, \vec{q}), Y(t_0) = Y_0, Y \in R^6, q \in R^9, \\ Y &= (Y_1(t), Y_2(t), \dots, Y_6(t))^T, F = (F_1, F_2, \dots, F_6)^T, \\ Y_0 &= (Y_{10}, Y_{20}, \dots, Y_{60})^T, \vec{q} = (q_1, q_2, \dots, q_9)^T. \end{aligned} \quad (2)$$

本文总假定函数  $F(t, Y)$  在  $R$  上连续,且关于  $Y$  满足 Lipschitz 条件,即存在一个常数  $C > 0$ ,使得  $|F(t, Y) - F(t, \bar{Y})| \leq C |Y - \bar{Y}|$ 。根据常微分方程理论可知,方程 (1) 的解存在且唯一。

表 1 参数描述

参数	描述	取值范围
$\alpha_I$	有症状感染的感染概率	(0, 1)
$\alpha_A$	无症状感染的感染概率	(0, 1)
$\beta_I$	暴露者向感染人群转变的概率	(0, 1)
$\beta_A$	暴露者向无症状感染人群转变的概率	(0, 1)
$\gamma_I$	感染者恢复概率	(0, 1)
$\gamma_A$	无症状感染者恢复概率	(0, 1)
$\mu$	病死率	(0, 1)
$\theta$	恢复者再次成为易感者概率	(0, 1)
$p$	无症状感染者比例	0.58

根据常微分方程的理论, 得到了等式的解 (1) 是存在的和唯一的。为了克服“不适定”的困难, 需要在反问题求解中需要附加相关的解信息。因此, 反问题的任务是根据所提供的相关信息, 尽可能多地恢复原问题, 然后得到了反问题的稳定近似解。为了求得未知参数向量  $\bar{q}$ , 给定一个附加测量:  $j$  时间点上的解, 即  $(Y_1(t_j), \dots, Y_6(t_j))^T, j=1, 2, \dots, r$ 。在这种情况下, 参数反演问题被简化为最小化目标函数问题。我们所需要的就是最小化以下的目标函数, 此时将参数反问题化为最小化目标函数问题, 即求得以下目标函数的最小值:

$$J(q) = \sum_{i=1}^6 \sum_{j=1}^r [Y_i(t_j, \bar{q}) - Y_i^{\text{exp}}(t_j)]^2, \quad (3)$$

其中  $Y_i(t_j, \bar{q})$  是以  $\bar{q}$  为参数求解原问题获得的相应计算值, 然后利用优化后的遗传算法得到目标函数方程式 (3) 的最优解。具体的步骤如图 1 所示。

#### 4 模型实证与疫情分析

本节使用世界卫生组织正式发布的每日新确诊病例数、累计确诊病例数、每日新增死亡病例数和累计死亡病例数。选择俄罗斯地区作为代表, 给出了俄罗斯的初步数据:

$$S_0 = N - I_0 - q_9 - q_{10} - q_{11}, E_0 = q_9, A_0 = q_{10}, I_0 = 1795059, R_0 = q_{11}, D_0 = 35321,$$

其中, 俄罗斯总人口为  $N = 145912022$ , 所需的参数向量

$$q_k = (\alpha_I, \alpha_A, \beta_I, \beta_A, \gamma_I, \gamma_A, \theta, \mu, E_0, A_0, R_0) \in R^{11}.$$

参数反演数据来自于 2022.02.01-2022.04.30 时期的俄罗斯公共资源。通过 Matlab 对 2022.05.01-2022.06.01 期间的智能算法下的预测值与确诊病例和死亡病例的真实值进行比较。

图 2 为每天新病例的拟合值、预测值和真实值的比较。从图 2 中可以看出, 改进后的遗传算法对确诊病例的拟合相对稳定, 初值和峰值的变化趋势与真实值基本一致。从过去 30 天的预测来看, 改进的 PSO 算法和遗传算法的预测曲线是叠加的。图 3 为总案例的拟合值、预测值和真实值的比较。结果表明, 改进后的遗传算法和 PSO 算法对总体情况具有良好的拟

合效果。但从近 30 天的预测结果来看, 改进后的遗传算法的预测曲线与实际值基本一致。累计死亡病例的拟合值、预测值和真实值的比较如图 4 所示。从图 4 中不难看出, 遗传算法的曲线与实值曲线更一致, 说明改进后的遗传算法的拟合和预测明显优于 PSO 算法。在图 5 中还可以发现, 随着迭代次数的增加, 改进后的遗传算法的自适应值比 PSO 算法更早达到最低值。同时, 结果表明, 改进后的遗传算法比 PSO 算法在拟合时间和预测时间上都更接近真实值。

最后, 两种优化方法下的重构参数(取平均值后)如表 2 所示。通过拟合和预测数据集, 分析了 SEAIRD 参数的行为。与其他参数相比, 可以看出改进的遗传算法反演后的恢复概率接近于俄罗斯的实际恢复率。结合图 2-4, 也可以看出, 改进后的 GA 对 COVID-19 传染病模型的参数更加准确。

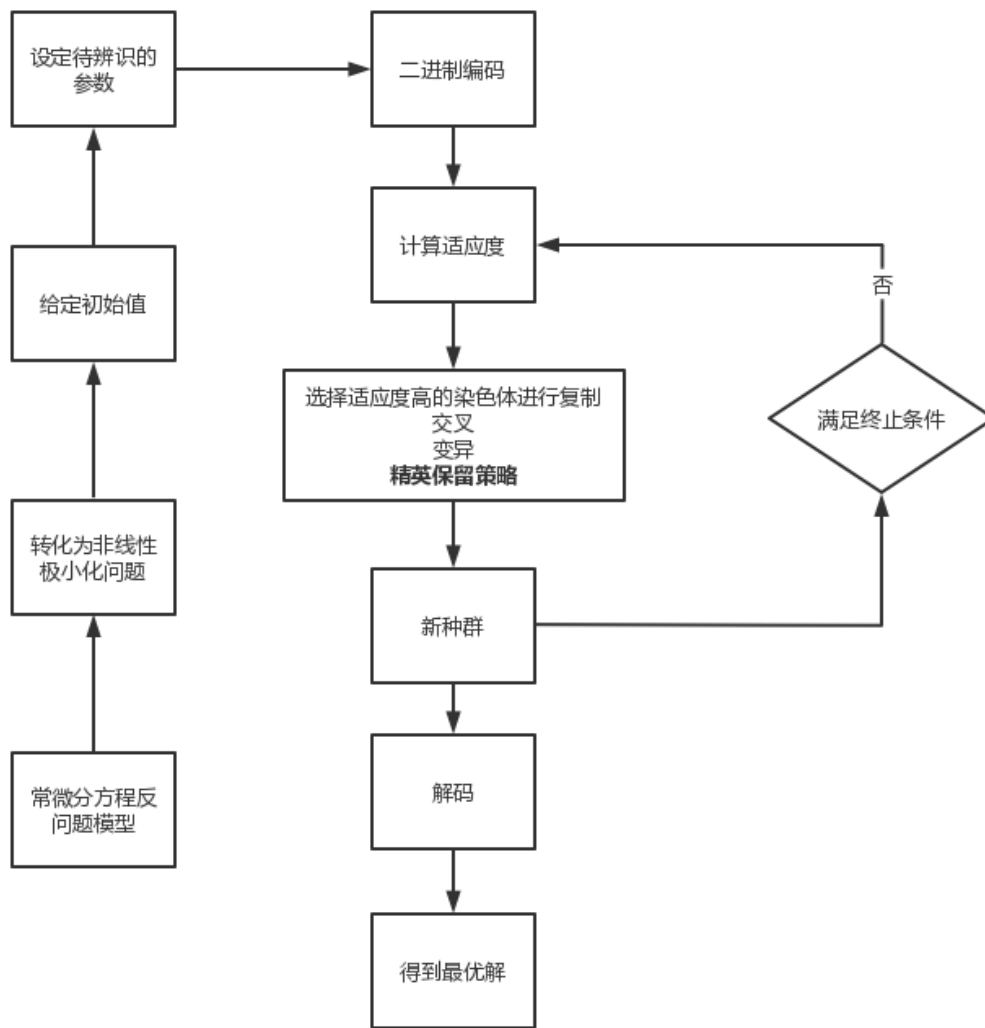


图 1 算法流程图

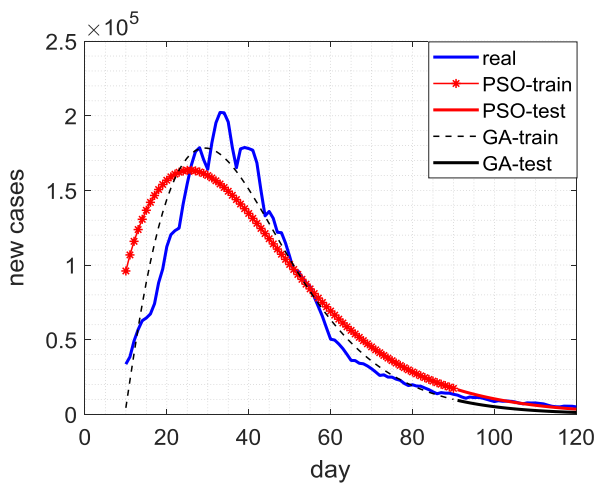


图 2 每日最新确诊病例的拟合值、预测值与真实值的对比图

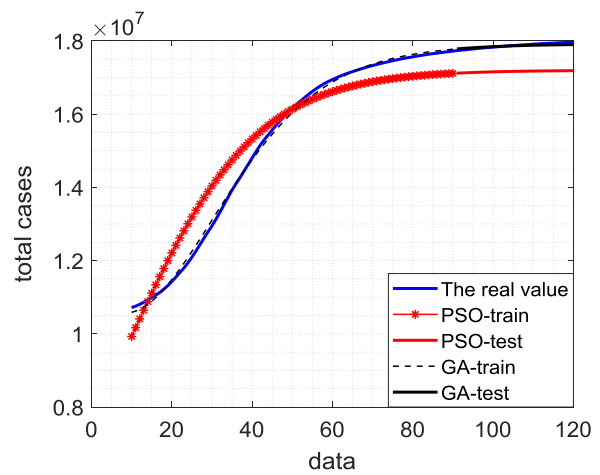


图 3 累计确诊病例的拟合值、预测值与真实值的对比图

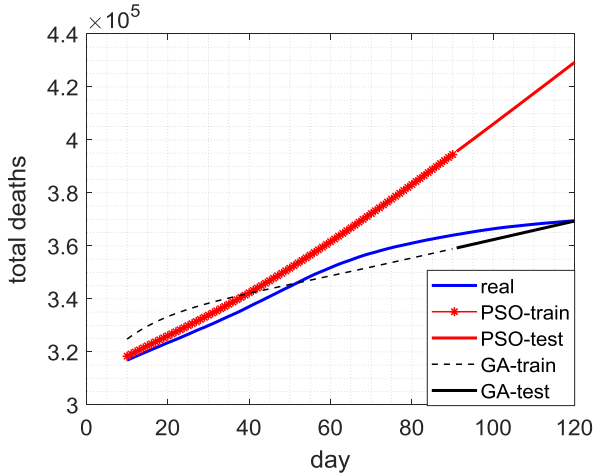


图 4 累计死亡病例的拟合值、预测值与真实值的对比图

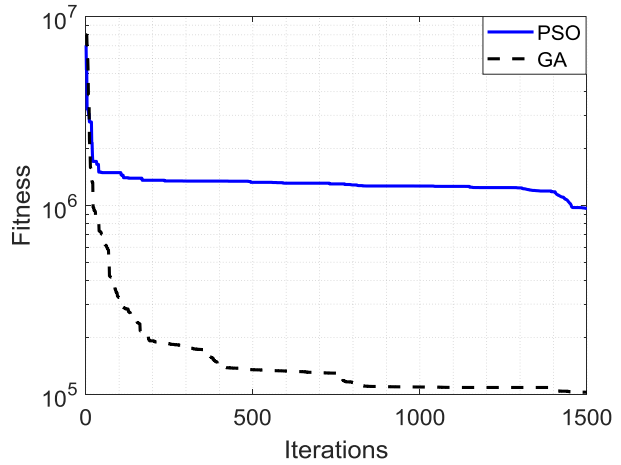


图 5 GA 与 PSO 的迭代次数与适应度值关系图

表 2 改进的遗传算法和 PSO 算法下的重构参数表

	$\beta_I$	$\beta_A$	$\gamma_I$	$\gamma_A$	$\mu$	$\theta$
GA	0.0116	0.052	0.5362	0.1623	0.055	0.2035
PSO	0.0834	0.1512	0.8236	0.1571	0.1094	0.5739

### 5 参数敏感性分析

参数敏感性分析<sup>[24]</sup>是通过适当地改变参数对模型进行仿真, 并观察数值结果对目标函数的影响, 目标函数的变化越大, 该参数就越敏感, 反之亦然。本文通过计算半相对灵敏度的方法对参数进行灵敏度分析。为了控制反问题求解的稳定性和预测的质量, 估计参数  $q$  对模型函数的敏感性是很重要的<sup>[25-26]</sup>。

为了计算其半相对灵敏度, 我们需要确定参数向量的观测模型量  $Y = (A, I)^T$  (其中, 模型解  $\bar{Y} = (S, E, A, I, R, D)^T$ ) 的相对灵敏度, 模型解对参数  $q_k$  的半相对灵敏度由表达式:  $\frac{\partial Y_i(t)}{\partial q_k} q_k$  定义, 并通

过计算 ODE 模型的微分形式  $\begin{cases} \frac{d\bar{Y}}{dt} = f(t, \bar{Y}; q) \\ \bar{Y}(0) = \bar{Y}_0 \end{cases}$  来定义。

因此, 我们得到灵敏度函数  $\bar{Y}_q(t; q) = \frac{\partial \bar{Y}(t; q)}{\partial q}$  的一个微分方程系统  $\frac{d}{dt} \left( \frac{\partial \bar{Y}(t)}{\partial q} \right) = \frac{\partial f}{\partial \bar{Y}} \frac{\partial \bar{Y}(t)}{\partial q} + \frac{\partial f}{\partial q}$ , 其中,  $\frac{\partial f}{\partial \bar{Y}}$  是雅可比矩阵,  $\frac{\partial f}{\partial q}$  是上式方程右侧参数的导数。初始条件为  $\frac{\partial \bar{Y}(0)}{\partial q} = \frac{\partial \bar{Y}_0}{\partial q} = 0$ 。这意味着任何初始条件下的零

矩阵都不是所有估计参数的函数。这就给出了关于敏感性作为时间的函数的信息。对于参数组合,

我们在  $t$  中 (在  $L_2$  空间中) 取一个范数来确定最敏感的参数 (见表 3) :

一般来说,  $\left\| \frac{\partial x_i(t)}{\partial q_k} q_k \right\|_2$  越小, 参数  $q_k$  对变量的影响越小。从表中可以看出,  $\beta_A$  和  $\gamma_A$  的变化对变量  $A$  和  $I$  的影响较大, 结果表明, 暴露人群转化为无症状感染者概率和无症状感染者的康复情况对确诊病例的影响较大。

参数敏感性分析表明, 该模型的参数  $\beta_A$  和  $\gamma_A$  对数据变化的敏感性最低。也就是说, 随着反问题的解决, 这些参数的确定更加稳定。参数  $\gamma_I$  和  $\theta$  对数据变化最为敏感。因此, 有必要开发一种正则化算法来控制敏感参数的质量。

表 3 参数敏感性

变量	参数	$\left\  \frac{\partial x_i(t)}{\partial q_k} q_k \right\ _2$
A+I	$\beta_I$	98.87
A+I	$\beta_A$	184.96
A+I	$\gamma_I$	78.18
A+I	$\gamma_A$	241.16
A+I	$\mu$	104.72
A+I	$\theta$	54.48

## 6 结论

随着传染病的发展,对传染病模型的参数反演引起了人们极大地关注。本文对新冠肺炎传染病 COVID-19 进行了两类研究,一类是传染病数学模型的建立,另一类是对模型进行参数反问题的研究。

首先,考虑无症状感染者可能成为感染源并且可以刺激身体的免疫反应,建立带有无症状感染者的传染病模型对流行病发展的研究是合理的、必要的。其次,利用改进的智能算法对所提的模型进行参数反演。对比 2022 年 2 月 1 日至 4 月 30 日俄罗斯真实确诊病例和死亡数据与改进 GA 下的拟合值,很明显可以看出拟合效果良好。对比 2022 年 5 月 1 日至 6 月 1 日所得参数解的预测值与真实值趋势基本一致。同样也说明参数解对模型的预测有意义。最后,对参数进行了敏感性分析,进一步验证了参数对模型的重要性作用。相信传染病反演得到的参数解对未来传染病预测以及传播机理的研究会起到良好的作用。

## 参考文献

- [1] Ciotti M, Ciccozzi M, Terroni A, et al. The COVID-19 pandemic[J]. *Critical reviews in clinical laboratory sciences*, 2020, 57(6): 365-388.
- [2] Lipsitch M, Swerdlow D L, Finelli L. Defining the epidemiology of Covid-19—studies needed[J]. *New England journal of medicine*, 2020, 382(13): 1194-1196.
- [3] Bulut C, Kato Y. Epidemiology of COVID-19[J]. *Turkish journal of medical sciences*, 2020, 50(9): 563-570.
- [4] Mandal M, Jana S, Nandi S K, et al. A model based study on the dynamics of COVID-19: Prediction and control[J]. *Chaos, Solitons & Fractals*, 2020, 136: 109889.
- [5] Carcione J M, Santos J E, Bagaini C, et al. A simulation of a COVID-19 epidemic based on a deterministic SEIR model[J]. *Frontiers in public health*, 2020, 8: 230.
- [6] Ivorra B, Ferrández M R, Vela-Pérez M, et al. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China[J]. *Communications in nonlinear science and numerical simulation*, 2020, 88: 105303.
- [7] Rapolu T, Nutakki B, Rani T S, et al. A time-dependent seird model for forecasting the COVID-19 transmission dynamics[J]. *medRxiv*, 2020: 2020.05. 29.20113571.
- [8] Oran D P, Topol E J. Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review[J]. *Annals of internal medicine*, 2020, 173(5): 362-367.
- [9] Paltiel A D, Zheng A, Walensky R P. Assessment of SARS-CoV-2 screening strategies to permit the safe reopening of college campuses in the United States[J]. *JAMA network open*, 2020, 3(7): e2016818-e2016818.
- [10] Denny T N, Andrews L, Bonsignori M, et al. Implementation of a pooled surveillance testing program for asymptomatic SARS-CoV-2 infections on a college campus—Duke University, Durham, North Carolina, August 2–October 11, 2020[J]. *Morbidity and Mortality Weekly Report*, 2020, 69(46): 1743.
- [11] Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)[J]. *Science*, 2020, 368(6490): 489-493.
- [12] Fracchia J A, Motta J, Miller L S, et al. Evaluation of asymptomatic microhematuria[J]. *Urology*, 1995, 46(4): 484-489.
- [13] Lassas M, Siltanen S. Foreword: inverse problems in biology[J]. *Journal of mathematical biology*, 2013, 67: 1-1.
- [14] Jameson A, Kreindler E. Inverse problem of linear optimal control[J]. *SIAM Journal on Control*, 1973, 11(1): 1-19.
- [15] Stevens M, Sunseri I, Alexanderian A. Hyper-differential sensitivity analysis for inverse problems governed by ODEs with application to COVID-19 modeling[J]. *Mathematical Biosciences*, 2022, 351: 108887.
- [16] Vogel C R. *Computational methods for inverse problems*[M]. Society for Industrial and Applied Mathematics, 2002.
- [17] Kaipio J, Somersalo E. *Statistical and computational inverse problems*[M]. Springer Science & Business Media, 2006.
- [18] Liu X, Hewings G, Wang S, et al. Modelling the situation of COVID-19 and effects of different containment strategies in China with dynamic differential equations and parameters estimation[J]. *MedRxiv*, 2020: 2020.03. 09.20033498.
- [19] Marinov T T, Marinova R S. Dynamics of COVID-19 using inverse problem for coefficient identification in SIR epidemic models[J]. *Chaos, Solitons & Fractals: X*, 2020, 5: 100041.
- [20] Godio A, Pace F, Vergnano A. SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence[J]. *International journal of environmental research and public health*, 2020, 17(10): 3535.
- [21] Krivorot'ko O I, Kabanikhin S I, Zyat'kov N Y, et al. Mathematical modeling and forecasting of COVID-19 in

- Moscow and Novosibirsk region[J]. Numerical Analysis and Applications, 2020, 13: 332-348.
- [22] Katoch S, Chauhan S S, Kumar V. A review on genetic algorithm: past, present, and future[J]. Multimedia Tools and Applications, 2021, 80: 8091-8126.
- [23] Wei X, Duan X, Chen L, et al. Optimal multi-objective distributed generation location and capacity by elite strategy genetic algorithm considering timing characteristics[C]//E3S Web of Conferences. EDP Sciences, 2020, 185: 01018.
- [24] Raue A, Becker V, Klingmüller U, et al. Identifiability and observability analysis for experimental design in nonlinear dynamical models[J]. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2010, 20(4): 045105.
- [25] Krivorotko O I, Andomaya D V, Kabanikhin S I. Sensitivity analysis and practical identifiability of some mathematical models in biology[J]. Journal of Applied and Industrial Mathematics, 2020, 14: 115-130.
- [26] Long J, Khaliq A Q M, Furati K M. Identification and prediction of time-varying parameters of COVID-19 model: a data-driven deep learning approach[J]. International journal of computer mathematics, 2021, 98(8): 1617-1632.

**版权声明:** ©2023 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

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



**OPEN ACCESS**