

文章编号: 2096-1618(2021)04-0368-06

# 粒子群算法估计 SEIRD 模型参数的 新冠肺炎疫情预测分析

王 颖, 胡建成

(成都信息工程大学应用数学学院, 四川 成都 610225)

**摘要:**通过建立 SEIRD 模型模拟新型冠状病毒在人群中的传播机制。选取湖北省 2020 年 1 月 23 日到 2 月 16 日疫情数据,建立 SEIRD 模型,根据 SEIRD 模型参数,构建优化模型,应用粒子群算法对参数进行估计,分析和预测湖北省新型冠状病毒肺炎疫情拐点。考虑到政府采取的措施,用修正的 SEQRD 模型,重新对参数进行估计,模型显示在政府的干预下,疫情的拐点提前了,感染人数减少。

**关键词:**新冠肺炎;SEIRD 模型;粒子群算法;SEQRD 模型;预测

**中图分类号:**TP391

**文献标志码:**A

**doi:**10.16836/j.cnki.jcuit.2021.04.003

## 0 引言

新型冠状病毒肺炎(corona virus disease 2019, COVID-19),简称“新冠肺炎”,是指 2019 年新型冠状病毒(SARS-CoV-2)感染所导致的肺炎<sup>[1-2]</sup>。该病发病症状与严重急性呼吸综合征(SARS)类似,但相比 SARS 传染力更强,潜伏期也更长。2019 年 12 月 8 日,武汉市首次发现新冠肺炎患者。随着患者人数越来越多,2020 年 1 月 23 日,武汉市采取“封城”措施,湖北省其他各市也先后采取相应举措,严防疫情扩散。

新冠肺炎在全球爆发后,许多研究员都对其传播和发展趋势做了研究,提出并重新建立了各种关于新冠肺炎的数学模型,为各国应对新冠肺炎疫情的防控发挥了积极作用。盛华雄等<sup>[3]</sup>在疫情的控制阶段运用 SIR 模型和差分递推方法分析和预测武汉市的疫情发展趋势;Ming 等<sup>[4]</sup>根据不同的诊断率和不同的公共卫生干预效果,采用改进的 SIR 模型来预测感染病例的实际数量以及隔离病房和重症监护病房的具体负担;Wan 等<sup>[5]</sup>通过建立 SEIR 模型,研究武汉“封城”后的数据,预测了武汉市的疫情发展趋势;Lin 等<sup>[6]</sup>提出改进的 SEIR 模型在武汉地区疾病传播中的应用,考虑了春运、封锁城市和建造方舱收容所医院等因素,最终有效地捕获了 COVID-19 的演变轨迹;Tang 等<sup>[7]</sup>通过建立 SEIR 模型来模拟并分析湖北武汉采取限行措施对北京疫情情况的影响,证明了采取快速、有效公共卫生措施的重要性;范如国等<sup>[8]</sup>建立 SEIR 模型,考虑新冠肺炎在 3 种不同潜伏期情形下,对中国全境、湖北

省及武汉市的疫情数据进行研究,分析了在不同潜伏期下对感染人数和疫情拐点的影响;Lobato 等<sup>[9]</sup>通过建立 SIDR 模型,并提出一个多目标优化问题,用多种算法估计 SIDR 模型中的参数,模拟出 COVID-19 的动力学过程。

因此,通过分析已有的早期新冠肺炎疫情数据,对已有数据进行建模,预测未来疫情的发展趋势,为政府疫情防控提供理论支持。

## 1 研究方法

### 1.1 SEIRD 模型构建

仓室模型最早由 Kermack 等<sup>[10]</sup>提出,其基本思想是将人群分成几个仓室,每个仓室分别表示处在不同疾病状态的人群,通过数学方法确定这些变量的动力学方程,研究疾病的动力学传播过程。常见的仓室模型有 SI 模型、SIR 模型、SEIR 模型等<sup>[11]</sup>。考虑到此次新型冠状病毒肺炎潜伏期比较长并且死亡人数较多,对 SEIR 模型进行修正,将移出人群分为死亡人群和恢复人群两类,建立包含 5 种人群(易感人群、潜伏期人群、感染人群、康复人群、死亡人群)的 SEIRD 动力学模型来预测此次新型冠状病毒肺炎的传播。

SEIRD 模型把研究人群分成以下 5 类:(1)S 为易感人群(Susceptible),指潜在的感染人群,即在感染前易被其他群体所感染;(2)E 为潜伏期人群(Exposed),指已感染,但无明显感染症状的群体;(3)I 为感染人群(Infected),指表现出有感染症状的群体;(4)R 为康复人群(Recovered),指感染新冠肺炎经过

治疗,痊愈的人群;(5) D 为死亡人群(Dead),指感染新冠肺炎死亡的人群。

SEIRD 模型中有以下假设:(1)易感人群与潜伏期人群接触就有一定的感染可能性;(2)易感人群与感染期人群接触就有一定的感染可能性;(3)康复人群具有免疫性,只要治愈,就不会再次被感染,并且不具备传染性;(4)死亡人群不具备传染性;(5)不考虑在新冠肺炎传播期间,人口的迁入迁出以及自然出生和自然死亡的影响;(6)只考虑人与人之间传播的途径,不考虑环境传播。

根据新冠病毒的特点,各个仓室人群的数量是不断变化的,分别记  $S(t)$ 、 $E(t)$ 、 $I(t)$ 、 $R(t)$ 、 $D(t)$  为时刻  $t$  的易感人群数量、潜伏期人群数量、感染人群数量、康复人群数量、死亡人群数量,显然有

$$S(t) + E(t) + I(t) + R(t) + D(t) = N \quad (1)$$

其中  $N$  为总人数。

SEIRD 模型的动力学过程如图 1 所示。

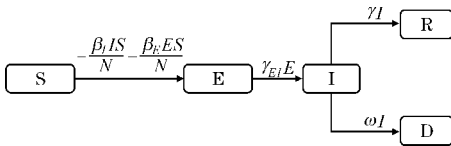


图1 SEIRD 模型动力学过程

该模型微分方程组为

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta_I IS}{N} - \frac{\beta_E ES}{N} \\ \frac{dE}{dt} = \frac{\beta_I IS}{N} + \frac{\beta_E ES}{N} - \gamma_{EI} E \\ \frac{dI}{dt} = \gamma_{EI} E - \gamma I - \omega I \\ \frac{dR}{dt} = \gamma I \\ \frac{dD}{dt} = \omega I \end{cases} \quad (2)$$

其中: $\beta_I$  表示易感人群在单位时间  $t$  内接触感染人群并被传染的概率, $\beta_E$  表示易感人群在单位时间  $t$  内接触潜伏期人群并被传染的概率, $\gamma_{EI}$  表示潜伏期人群在单位时间  $t$  内变成感染人群的概率, $\gamma$  表示感染人群在单位时间  $t$  内痊愈的概率; $\omega$  表示感染人群在单位时间  $t$  内死亡的概率。

## 1.2 参数估计模型

根据建立的 SEIRD 模型,需要对 SEIRD 模型中的参数进行估计。考虑到感染人群、康复人群、死亡人群都有真实数据,在 SEIRD 模型中通过应用四阶龙格库塔法可以拟合出感染人群、康复人群、死亡人群数据。

针对感染人群的误差可以定义为将拟合感染人群数据与真实感染人群数据相减,建立一个误差函数,误差函数定义如下:

$$f_1 = \frac{1}{\max(I_i^p)^2} \sum_{i=1}^N (I_i^p - I_i^t)^2 \quad (3)$$

其中  $I_i^p$  表示拟合的感染数据, $I_i^t$  表示真实的感染数据。

同理可得,针对康复人群、死亡人群的误差函数,定义分别如下:

$$f_2 = \frac{1}{\max(R_i^p)^2} \sum_{i=1}^N (R_i^p - R_i^t)^2 \quad (4)$$

$$f_3 = \frac{1}{\max(D_i^p)^2} \sum_{i=1}^N (D_i^p - D_i^t)^2 \quad (5)$$

其中  $R_i^p$  表示拟合的康复人群数据, $R_i^t$  表示真实的康复人群数据, $D_i^p$  表示拟合的死亡人群数据, $D_i^t$  表示真实的死亡人群数据。

根据 3 种状态下分别建立的误差函数,把 3 种状态的误差函数相加,可以得到最终的目标函数,建立优化问题,即

$$f = \frac{1}{\max(I_i^p)^2} \sum_{i=1}^N (I_i^p - I_i^t)^2 + \frac{1}{\max(R_i^p)^2} \sum_{i=1}^N (R_i^p - R_i^t)^2 + \frac{1}{\max(D_i^p)^2} \sum_{i=1}^N (D_i^p - D_i^t)^2 \quad (6)$$

## 2 粒子群算法

粒子群算法 (particle swarm optimization, PSO) 是在 1995 年由 Kennedy 和 Eberhart 共同提出的<sup>[12]</sup>。粒子群算法是通过模拟鸟类在觅食过程中的迁徙和聚集行为的一种基于群体智能的全局随机搜索算法。在粒子群算法中,每个优化问题的可能解都是搜索空间中的一只鸟,称为“粒子”。所有的粒子都有一个由优化函数决定的适应度值,每个粒子都有一个确定其飞行方向和距离的速度,然后粒子们在解空间中追随当前的最优粒子搜索。

第一个“极值”是粒子自身找到的个体最佳位置,另一个“极值”是整个种群找到的全局最佳位置。

$$v_i(t+1) = \omega v_i(t) + c_1 r_1(t) [p_{b,i}(t) - x_i(t)] + c_2 r_2(t) [p_g(t) - x_i(t)] \quad (7)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (8)$$

并按照式(7)、式(8)更新自己的速度和位置,满足算法终止条件后停止运行,输出优化结果。其中, $v_i$  是粒子速度, $x_i$  是粒子的当前位置, $p_{b,i}$  是粒子自身当前找到的最优位置, $p_g$  是整个粒子群当前找到的最优位置, $x_i$  是粒子的当前位置。 $\omega$  为惯性权重; $c_1$  和  $c_2$  为学习因子,也称加速常数; $r_1$  和  $r_2$  是  $[0,1]$  范围内的均

匀随机数,用于确保粒子飞行随机性。

3 基于粒子群算法估计 SEIRD 模型参数及分析

3.1 湖北地区疫情数据预处理

2020 年 1 月 23 日,武汉市“封城”,湖北省内其他各市也先后采取相应措施,严防疫情扩散。数据选取 1 月 23 日到 2 月 16 日的湖北省疫情数据进行研究,所使用的疫情数据来自湖北省卫健委官网。

在疫情初期监测试剂资源匮乏,医务人员不足,导致湖北地区统计数据波动较大,所以需要对数据进行预处理。1 月 27 日因为检测能力加强,新增确诊病例较多,按比例分配把 1 月 27 日新增确诊病例分配到 1

月 23 日到 26 日。

2 月 12 日,湖北省的病例诊断分类中增加了“临床诊断病例”,将临床诊断病例数纳入确诊例数,尽最大可能发现已感染但未检测出的患者和疑似病例。将 2 月 14 日的累计确诊人数减去 2 月 11 日的累计确诊人数,得到这 3 天的累计确诊人数,并且减去 3 天的每日新增确诊人数,可以得到最终“临床诊断病例”人数。考虑到这些“临床诊断病例”人数往往是之前感染的患者,但是却没能确诊。对数据进行处理,和处理 27 日的数据类似,按比例分配把这几日累计的临床诊断病例分配到 1 月 23 日到 2 月 14 日的数据中。考虑到 2 月 15 日的“临床诊断病例”比较少,就直接列为确诊人数。

预处理前后的湖北省累计确诊、累计治愈及累计死亡人群数据见图 2。

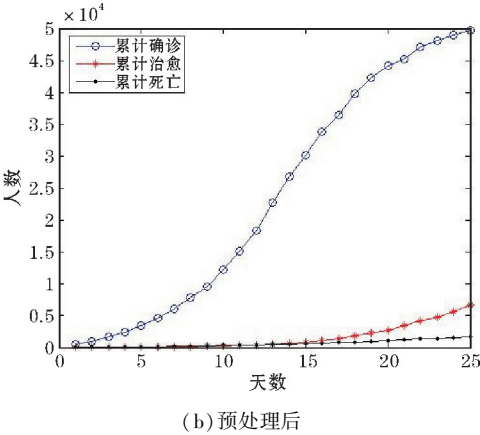
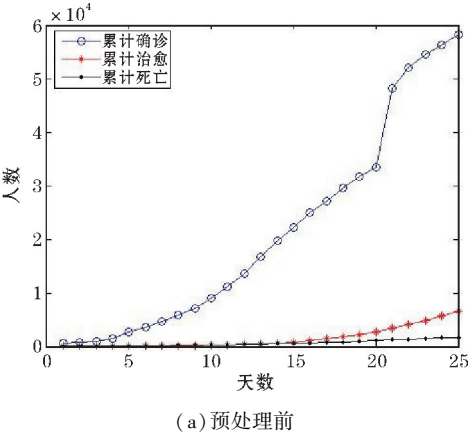


图2 湖北省感染人群数量走势

3.2 SEIRD 模型参数设置

在数据的选取中,选择了 1 月 23 到 2 月 11 日的数据进行研究。对 SEIRD 模型初值设置,由于疫情的传播是在一些地区范围传播的,不可能每个人都是易感者,将湖北省 2020 年 1 月 23 日人群总数估计为 71000 人,易感人群估计为 69465 人,潜伏期人群估计为 400 人。根据对湖北省数据进行预处理后的数据,感染人群设置为 540 人,康复人群设置为 31 人,死亡人群设置为 24 人,SEIRD 模型初值设置如表 1 所示。

表 1 SEIRD 模型初值						
参数	$N$	$S$	$E$	$I$	$R$	$D$
值	71000	69465	400	540	31	24

对 SEIRD 模型的参数设置范围为, $0 \leq \beta_I \leq 0.5$ , $0 \leq \beta_E \leq 0.5$ , $0 \leq \gamma_{EI} \leq 0.6$ , $0 \leq \gamma \leq 0.3$ , $0 \leq \omega \leq 0.5$ 。

3.3 粒子群算法估计参数

在粒子群算法估计 SEIRD 模型参数中,粒子即是

SEIRD 模型中的各个参数,粒子的速度范围即是各个参数的取值范围,适应度函数即是式(6)。

在用粒子群算法求解参数中考虑到种群规模和迭代次数的设置与精度要求和算法设置有关,种群规模越大、迭代次数越多算法精度越高,但相应的运行时间越长,所以设置粒子种群为 50、迭代次数为 100。设置学习因子  $c_1=c_2=1.5$ ,惯性权重  $\omega$  为 1。

用粒子群算法估计参数的伪码如下:

粒子群算法估计参数  
输入粒子群参数:种群规模  $N$ ,迭代次数  $it$  等  
for  $i=1$  to  $N$   
初始化粒子速度  $v_i$  和位置  $x_i$   
计算粒子适应度  
更新粒子最优解  
if  $fit(p_{b,i}) < fit(p_g)$   
 $p_g = p_{b,i}$   
end  
end for

```
for it=1:max it
for i=1 to N
根据式(7)更新粒子速度
根据式(8)更新粒子位置
计算适应度函数
if fit(xi)<fit(pb,i)
pb,i=xi
end
if fit(pb,i)<fit(pg)
```

```
pg=Pb,i
end
end for
print pg,print pb,i
end for
```

经过粒子群算法优化得到  $\beta_I$  为 0.27667,  $\beta_E$  为 0.51637,  $\gamma_{EI}$  为 0.15458,  $\gamma$  为 0.008496,  $\omega$  为 0.0039368,拟合出的疫情预测图,如图 3 所示。

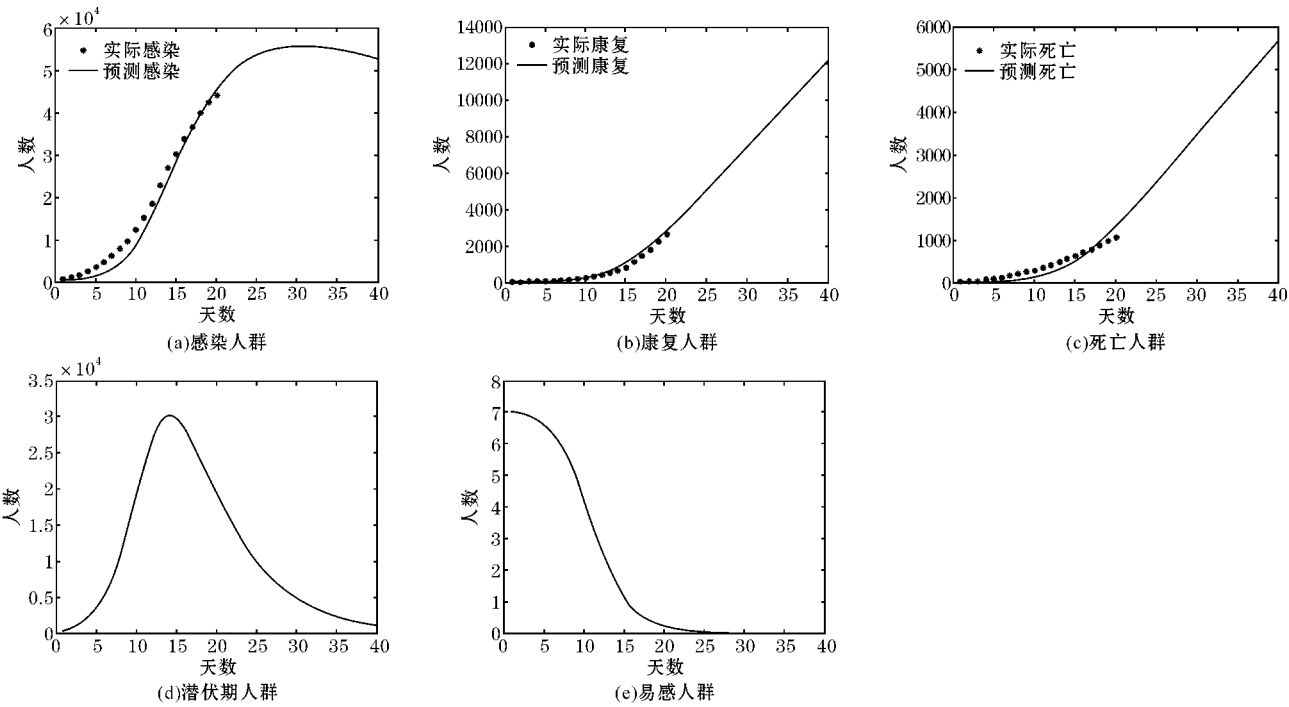


图3 湖北省疫情预测分布图

从图 3 可以得出,潜伏期患者前 5 天增长相对缓慢,5-14 天(2020 年 1 月 27 日到 2 月 5 日)增长非常迅速,并于 2 月 5 日潜伏期人数到达顶峰,预测出潜伏期人数为 30140 人;感染人群在 10-25 天快速增长(2020 年 2 月 1 日到 2 月 16 日),并于 2020 年 2 月 21 日到达顶峰,即拐点在 2 月 21 日,预测感染人数峰值为 55680 人。

3.4 模型改进

随着新冠肺炎疫情的发展,在政府的号召下,中国各个省市的医疗卫生及其他相关部门越来越多的人奔赴湖北省进行支援,并且湖北省建立了比如“火神山”,“雷神山”及其他方舱医院。对数据进行重新选取,选取 2020 年 1 月 27 日到 2 月 16 日的数据进行研究。

考虑到随着医疗环境的增强,只要发现感染者,马上就进行隔离,不会传染给易感人群,把感染人群仓室相当于隔离仓室,对 SEIRD 模型进行修正,得到新的

SEQRD 模型,其中 Q 表示感染人群即隔离人群,修正的模型微分方程组为:

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta_E ES}{N} \\ \frac{dE}{dt} = \frac{\beta_E ES}{N} - \frac{\beta_E ES}{N} - \gamma_{EQ} E \\ \frac{dQ}{dt} = \gamma_{EQ} E - \gamma Q - \omega Q \\ \frac{dR}{dt} = \gamma Q \\ \frac{dD}{dt} = \omega Q \end{cases} \quad (9)$$

根据 SEQRD 模型得到的目标函数,即:

$$f = \frac{1}{\max(Q_i^p)^2} \sum_{i=1}^N (Q_i^p - Q_i^t)^2 + \frac{1}{\max(R_i^p)^2} \sum_{i=1}^N (R_i^p - R_i^t)^2 + \frac{1}{\max(D_i^p)^2} \sum_{i=1}^N (D_i^p - D_i^t)^2 \quad (10)$$

其中  $Q_i^p$  表示拟合的感染人群数据,  $Q_i^t$  表示真实的感



染人群数据,  $R_i^p$  表示拟合的康复人群数据,  $R_i^t$  表示真实的康复人群数据,  $D_i^p$  表示拟合的死亡人群数据,  $D_i^t$  表示真实的死亡人群数据。

在 SEQRD 模型中各个数据的初值如表 2 所示, 参数取值范围如表 3 所示。

表 2 SEQRD 模型初值						
参数	$N$	$S$	$E$	$Q$	$R$	$D$
值	71000	66834	3479	3479	47	100

表 3 SEQRD 参数范围				
参数	$\beta_E$	$\gamma_{EI}$	$\gamma$	$\omega$
取值范围	[0,1]	[0,0.4]	[0,1]	[0,1]

在粒子群的参数设置中, 设置粒子种群为 50、迭代次数为 100。设置学习因子  $c_1=c_2=1.5$ , 惯性权重  $\omega$  为 1, 经过粒子群算法优化得到  $\beta_E$  为 0.72325,  $\gamma_{EI}$  为 0.32196,  $\gamma$  为 0.0097818,  $\omega$  为 0.0030292, 拟合出的疫情预测图, 如图 4 所示。

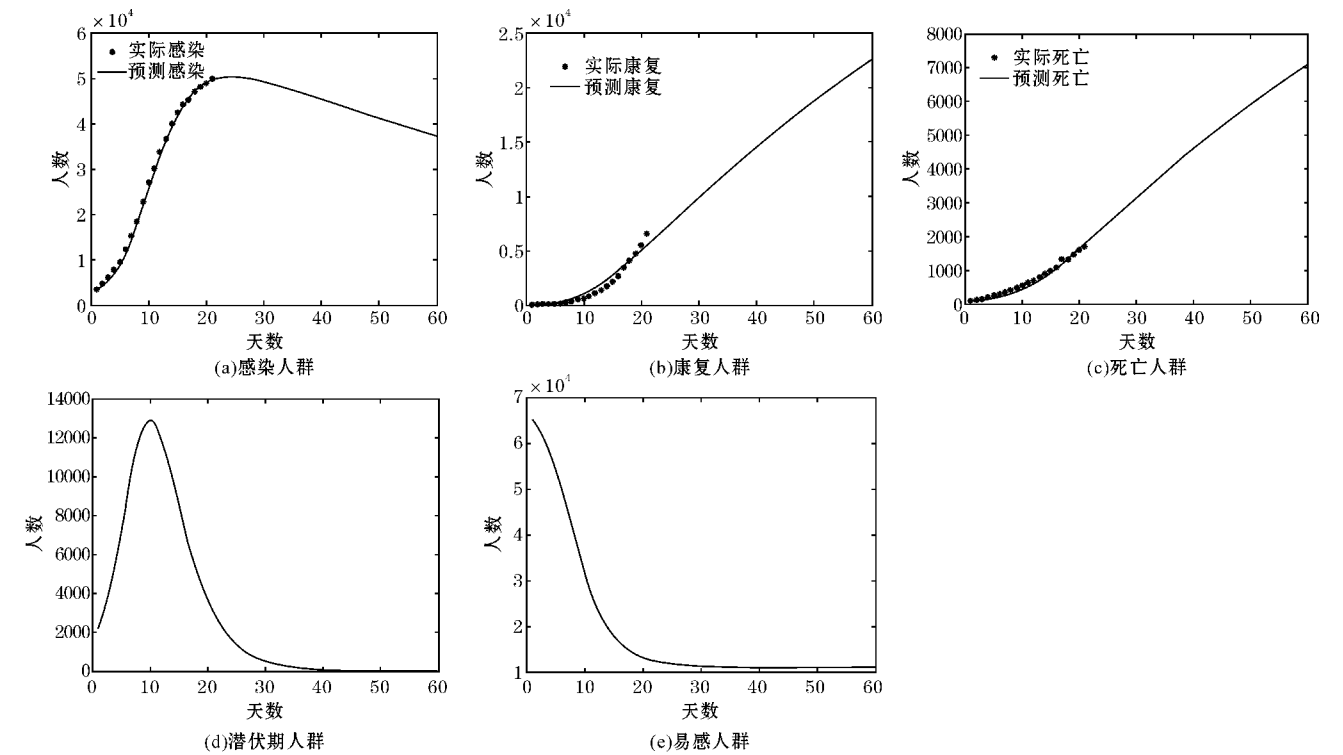


图 4 湖北省疫情预测分布图

从图 4 可以得出, 潜伏期患者前 10 天迅速增长, 并在第 10 天(2020 年 1 月 27 日到 2 月 5 日)达到顶峰, 即是 2 月 5 日潜伏期人数达到拐点, 预测出潜伏期人数为 12900 人; 感染人群在 10~20 天快速增长(2020 年 2 月 5 日到 2 月 15 日), 并在 2020 年 2 月 19 日达到顶峰, 即拐点在 2 月 19 日, 预测感染人数峰值为 50390 人。

相比之前的预测, 感染人数明显减少了很多, 并且拐点还提前, 证明了政府采取的隔离措施是有效的。并且相比之前的预测, 治愈率提高, 死亡率降低。在只考虑了隔离这个因素的前提下, 中国新冠肺炎疫情的控制取得了极大的成果, 那么再考虑增加新冠肺炎疫苗的研发、媒体积极宣传戴口罩、保持社交距离等措施下, 就容易解释中国新冠肺炎疫情从爆发到控制的短时高效。

4 结束语

研究基于 SEIRD 模型, 通过粒子群算法对参数进

行估计, 提高预测效果, 从而较好地拟合了新冠肺炎早期在湖北省的传播, 预测出疫情的拐点, 与真实数据相接近。在 SEQRD 模型中, 通过粒子群算法重新对参数进行估计, 体现了国家采取措施的有效性, 使感染人数降低, 疫情的拐点提前。本文的研究方法, 其他国家和地区一样适用, 可以为新冠肺炎疫情防控提供一定的参考价值。

为使疫情趋势预测模型更为准确, 未来工作将会基于现有模型从以下几点进行改进: (1) 考虑各地区人口的流动情况。(2) 在公共场所戴口罩的情况。(3) 加入政府采取的其他措施, 如疫苗。

参考文献:

[1] Zhou P, Yang X L, Wang X G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin [ J ]. nature, 2020, 579

- (7798): 270–273.
- [2] Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China [J]. *Nature*, 2020, 579(7798): 265–269.
- [3] 盛华雄,吴琳,肖长亮. 新冠肺炎疫情传播建模分析与预测[J]. *系统仿真学报*, 2020, 32(5): 759–766.
- [4] Ming W K, Huang J, Zhang C J P. Breaking down of healthcare system: Mathematical modeling for controlling the novel coronavirus (2019–nCoV) outbreak in Wuhan, China [J/OL]. <https://www.biorxiv.org/CONTENT/10.1101/2020.01.27.922443V2.full>, 2020–01–30.
- [5] Wan K, Chen J, Lu C, et al. When will the battle against novel coronavirus end in Wuhan: A SEIR modeling analysis[J]. *Journal of Global Health*, 2020, 10(1): 1–6.
- [6] Lin Q, Zhao S, Gao D, et al. A conceptual model for the coronavirus disease 2019 (COVID–19) outbreak in Wuhan, China with individual reaction and governmental action[J]. *International journal of infectious diseases*, 2020, 93: 211–216.
- [7] Tang B, Wang X, Li Q, et al. Estimation of the transmission risk of the 2019–nCoV and its implication for public health interventions[J]. *Journal of clinical medicine*, 2020, 9(2): 462.
- [8] 范如国,王奕博,罗明,等. 基于SEIR的新冠肺炎传播模型及拐点预测分析[J]. *电子科技大学学报*, 2020, 49(3): 369–374.
- [9] Lobato F S, Libotte G B, Platt G M. Identification of an Epidemiological Model to Simulate the COVID–19 Epidemic Using Robust Multiobjective Optimization and Stochastic Fractal Search[J]. *Computational and Mathematical Methods in Medicine*, 2020: 1–8.
- [10] Kermack W O, McKendrick A G. Contributions to the mathematical theory of epidemics[J]. *Bulletin of mathematical biology*, 1991, 53(1/2): 33–55.
- [11] Anderson R M, May R M. *Infectious diseases of humans dynamics and control*[M]. Oxford: Oxford University Press, 1991.
- [12] Kennedy J, Eberhart R. Particle swarm optimization[C], *Proceedings of the 4th IEEE International Conference on Neural Networks*, Piscataway: IEEE Service Center, 1995: 1942–1948.

## Prediction and Analysis of COVID-19 Using PSO Algorithm for Estimating SEIRD Model Parameters

WANG Ying, HU Jiancheng

(College of Applied Mathematics, Chengdu University of Information Technology, Chengdu 610225, China)

**Abstract:** SEIRD model was used to simulate the spread of COVID-19 in the crowd. Based on the epidemic data of Hubei Province from January 23 to February 16, 2020, SEIRD model was established. According to the parameters of SEIRD model, the optimization model was constructed, and the parameters were estimated by PSO algorithm to analyze and predict the inflection point of COVID-19 in Hubei Province. Considering the measures taken by the government, the parameters are re estimated by using the modified SEQRD model. The model shows that the inflection point of the epidemic is advanced and the number of infected people is reduced under the intervention of the government.

**Keywords:** COVID-19; SEIRD model; PSO; SEQRD model; predict