

22. 基于 1stOpt 新冠疫情的数学模型与预测-以上海为例

上海本次爆发的新冠疫情非常严峻，其走势如何？万众瞩目。基于 1stOpt 软件平台，仅从数学模型的角度进行简单的分析及预测，同时也展示 1stOpt 在流行病模型方面的应用潜力。

22.1 数据

数据如下表，全部采自官方发布的从 3 月 1 日至 4 月 10 日上海疫情数据，所有数据为确诊病例和无症状感染病例之和。

表 22.1: 上海 3 月 1 日至 4 月 10 日官方疫情数据

月份	日期	序号	每日感染人数	累积感染人数
3	1	1	2	2
	2	2	8	10
	3	3	16	26
	4	4	19	45
	5	5	28	73
	6	6	48	121
	7	7	55	176
	8	8	65	241
	9	9	80	321
	10	10	75	396
	11	11	83	479
	12	12	65	544
	13	13	169	713
	14	14	138	851
	15	15	202	1053
	16	16	158	1211
	17	17	260	1471
	18	18	374	1845
	19	19	509	2354
	20	20	758	3112
	21	21	896	4008
	22	22	981	4989
	23	23	983	5972
	24	24	1609	7581
	25	25	2269	9850
	26	26	2676	12526
	27	27	3500	16026
	28	28	4477	20503
	29	29	5982	26485
	30	30	5635	32120
	31	31	4502	36622
4	1	32	6311	42933
	2	33	8226	51159

3	34	9006	60165
4	35	13354	73519
5	36	17077	90596
6	37	19982	110578
7	38	21222	131800
8	39	23624	155424
9	40	24943	180367
10	41	26087	206454

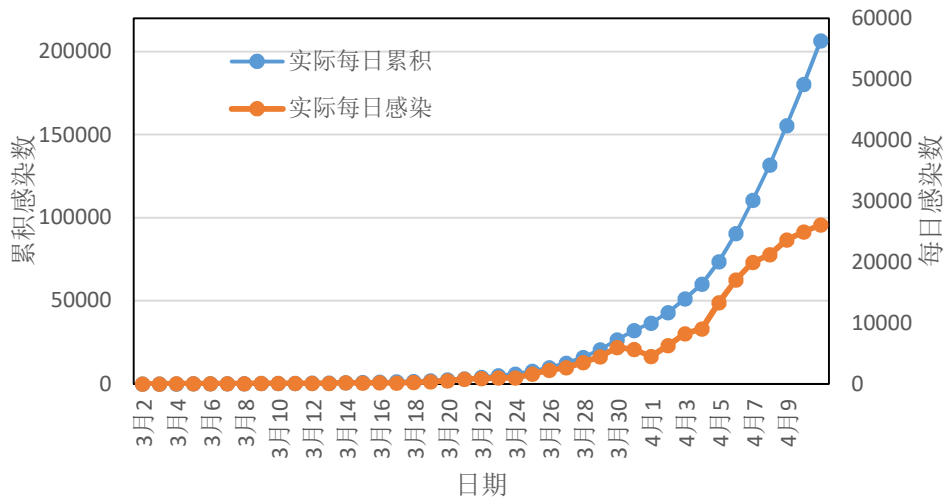


图 22-1 实际每日及累积感染数

22.2 数学模型

采用两种模型，第一种是 Richards 模型，为简单的 S 型增长数学模型，如公式(22-1)；第二种为传染病动力学 SIR 模型，如公式(22-2)。

Richards 模型:

$$y = a \cdot \left(1 + (d - 1) \cdot \exp(-k \cdot (t - xc))\right)^{\frac{1}{1-d}} \quad (22-1)$$

其中，t 为时间，自变量；y 为累积感染数，因变量；a、b、xc 和 d 为模型 4 个参数。

SIR 模型:

$$\begin{cases} \frac{dS}{dt} = -\beta \cdot S \cdot I \\ \frac{dI}{dt} = \beta \cdot S \cdot I - \gamma \cdot I \\ \frac{dR}{dt} = \gamma \cdot I \end{cases} \quad (22-2)$$

其中:

S: 易感者人群，为在传染病传播期间的 t 时刻还未被感染，但存在被传染风险的群体；

I: 感染者人群，为在传染病传播期间的 t 时刻已被感染，带有该病症状与传染力的群体；

R: 移出者人群, 为在传染病传播期间的 t 时刻已从 I 中移出, 包括因该传染病死亡和已被治愈的群体。

β 和 γ : 模型参数。

22.3 模型求解

Richards 模型求解实际上就是一简单的非线性拟合计算问题, 在 1stOpt 中极其容易实现, 代码及结果如下

Richards 模型求解代码

```
Variable t,y;
Function y=a*(1+(d-1)*exp(-k*(t-xc))^(1/(1-d)));
Data;
t=1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41;
y=2,10,26,45,73,121,176,241,321,396,479,544,713,851,1053,1211,1471,1845,2354,3112,4008,4989,5972,7581,9850,12526,16026,20503,26485,32120,36622,42933,51159,60165,73519,90596,110578,131800,155424,180367,206454;
```

结果

```
Sum Squared Error (SSE): 46459320.7136365
Root of Mean Square Error (RMSE): 1064.49714125837
Correlation Coef. (R): 0.999808120081112
R-Square: 0.999616276980126
Adjusted R-Square: 0.999596081031712
Determination Coef. (DC): 0.999592045078837
Chi-Square: 3044.83383810736
F-Statistic: 30067.1814475196
```

Parameter	Best Estimate
a	280195.690213548
d	4.0677466385129
k	0.594136202071256
xc	39.8607991580899

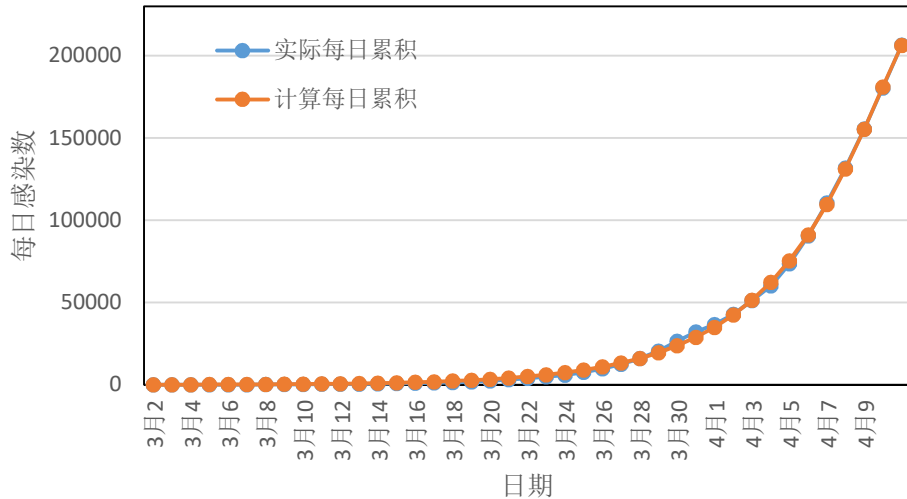


图 22-2 Richards 模型实际及计算每日累积感染数

Sir 模型实际为一微分方程优化拟合求解问题，求解代码及结果如下，初始条件取总人口数为 2500 万，感染者人群 I 和移出者人群 R 初始值分别取 2 和 0。为防止出错，注意代码中两个参数 β 和 γ 分别用 beta1 和 gamma1 表示，因为 beta 和 gamma 是 1stOpt 的缺省内置数学函数。

Sir 模型求解代码

```

Constant N=25000000;
Parameter beta1=[0,],gamma1=[0,];
InitialODEValue t=1,S=N,I=2,R=0;
Variable t, I;
ODEFunction S'=-beta1*S*I;
           I'=beta1*S*I-gamma1*I;
           R'=gamma1*I;

Data;
t=1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41;
I=2,8,16,19,28,48,55,65,80,75,83,65,169,138,202,158,260,374,509,758,896,981,983,1609,2269,2676,3500,4477,5982,5635,4502,6311,8226,9006,13354,17077,19982,21222,23624,24943,26087;

```

结果

Sum Squared Error (SSE): 25624597.5960203
 Root of Mean Square Error (RMSE): 790.563207636675
 Correlation Coef. (R): 0.970953797631053
 R-Square: 0.942751277134164
 Adjusted R-Square: 0.96581791327684
 Determination Coef. (DC): 0.965373855204167
 F-Statistic: 3705.1729412648

Parameter	Best Estimate
beta1	2.37773900379385E-7

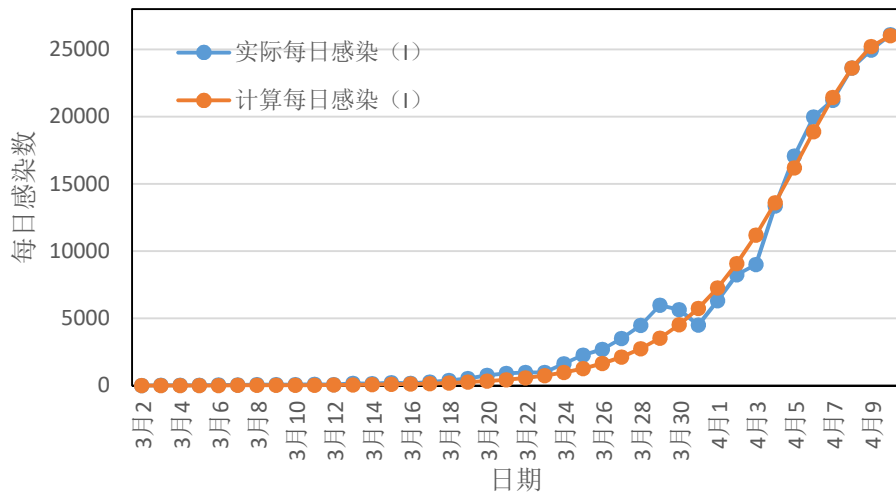


图 22-3 Sir 模型实际及计算每日感染数

详细计算数值结果见表 22.2.

从上面计算结果看，两种模型吻合度都比较高，Richards 模型相关系数高达 0.999。

22.4 模型预测

基于前述计算所得模型参数结果，用两种模型对 4 月 11 日至 4 月 29 日进行预测，详细结果见表 22.2 及图 22-4

表 22.2: 上海 3 月 1 日至 4 月 29 日计算及预测数据（红色为预测数据）

月份	日期	每日实际感染人数	Sir 模型计算及预测每日感染人数	Richards 模型计算及预测每日感染人数
3	1	2		
	2	8	3	22
	3	16	3	27
	4	19	4	33
	5	28	6	40
	6	48	8	49
	7	55	10	59
	8	65	13	72
	9	80	17	87
	10	75	23	105
	11	83	30	128
	12	65	39	155
	13	169	51	188
	14	138	66	229
	15	202	87	278

	16	158	114	337
	17	260	149	409
	18	374	195	496
	19	509	254	602
	20	758	333	731
	21	896	435	887
	22	981	568	1077
	23	983	741	1307
	24	1609	965	1586
	25	2269	1257	1925
	26	2676	1633	2337
	27	3500	2116	2836
	28	4477	2735	3441
	29	5982	3522	4176
	30	5635	4511	5066
	31	4502	5743	6144
	1	6311	7251	7446
	2	8226	9062	9015
	3	9006	11182	10893
	4	13354	13584	13113
	5	17077	16193	15686
	6	19982	18873	18550
	7	21222	21430	21508
	8	23624	23627	24127
	9	24943	25223	25681
	10	26087	26020	25316
	11		25917	22582
	12		24933	17994
	13		23203	12862
	14		20938	8404
4	15		18379	5146
	16		15748	3020
	17		13217	1727
	18		10903	972
	19		8868	543
	20		7129	301
	21		5679	167
	22		4490	92
	23		3529	51
	24		2762	28
	25		2153	16
	26		1674	9
	27		1299	5
	28		1006	3
	29		778	1
合计			390681	280087

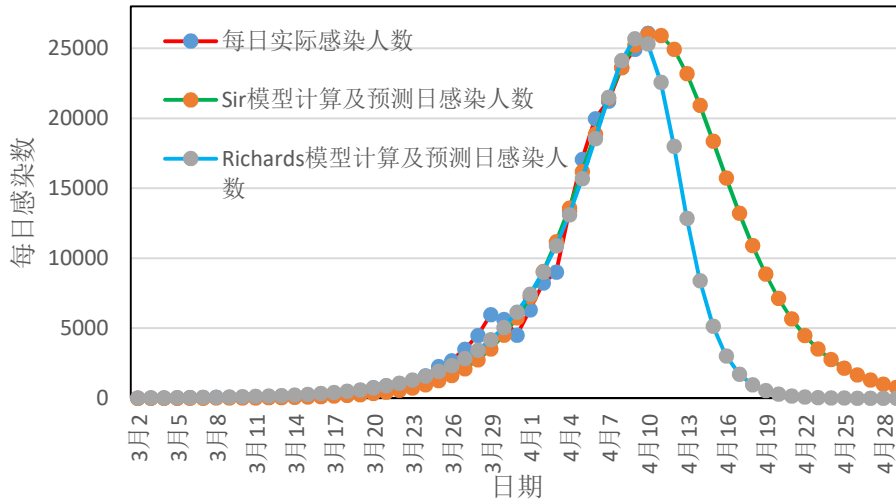


图 22-4 预测 4 月 11 日至 29 日每日感染数

22.5 讨论

不论是 Richards 模型还是 Sir 模型，从预测数值看，上海每日感染数即将达到峰值，也即拐点即将出现，最终感染总数 Sir 模型为约 40 万，Richards 则为 30 万，从图 22-4 看，衰减阶段 Richards 模型更为陡峭，Sir 模型相对平缓；到 4 月底 Sir 模型预测每日感染数降到千位数之下，Richards 模型则更是降到个位数。

之前新闻媒体中经常说到疫情“指数级增长”或“指数级增长拐点已到”等，不论从数学角度还是万物真实世界看，都不是很严谨或根本不存在。

指数模型公式如下

$$y = a \cdot \exp(b \cdot t)$$

仍以上海疫情数据为例，计算可得该指数模型参数 $a = 155.6410732$ ， $b = 0.176280462$ ，相关系数高达 0.999144，即

$$y = 155.6410732 \cdot \exp(0.176280462 \cdot t)$$

仅从数学角度看，该模型因变量（累积感染数 y ）随自变量（时间 t ）是单调增加的，不存在数学意义上的“拐点”一说；另外如果以该模型进行预测上海 10 天、20 天及 30 天后（对应 4 月 21 日、5 月 1 日及 5 月 11 日）总感染人数，将分别达到 125 万、728 万和 4243 万，显然是极其不合理和错误的。真实世界任何事物从小到大或从弱到强的发展基本服从 S 型增长曲线，该类曲线也存在真正的拐点。

22.6 结论

不论是对简单的非线性拟合问题还是相对复杂的微分方程拟合问题，1stOpt 都可以轻松处理并获得满意的结果。

本文模型所得任何疫情预测相关数据，都是仅从数学考虑和角度出发，无任何其它实际意义。

最后希望疫情早日结束，国泰民安！