

23. 基于 1stOpt 新冠疫情的数学模型与预测之二 -以上海为例

前篇公众号文章中，基于上海至 4 月 10 日为止官方发布的疫情数据，通过 Richards 模型和 SIR 经典传染病模型，对上海疫情及其发展走势进行了建模及预测分析，至今一周时间已过，预测结果与最新实际数据相比准确度如何？模型有无必要改进？最近一周实际疫情数据呈锯齿跳跃状，这种数据是否需要进行处理？针对这些问题进行一些相关分析及解答。

23.1 数据

最新原始疫情数据起始于 3 月 1 日，截止于 4 月 17 日，如表 23.1，数据值为确诊病例和无症状感染者之和。

表 23.1: 上海 3 月 1 日至 4 月 17 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	月份	日期	序号	每日感染人数	累积感染人数
3	1	1	2	2	3	26	26	2676	12526
	2	2	8	10		27	27	3500	16026
	3	3	16	26		28	28	4477	20503
	4	4	19	45		29	29	5982	26485
	5	5	28	73		30	30	5635	32120
	6	6	48	121		31	31	4502	36622
	7	7	55	176	4	1	32	6311	42933
	8	8	65	241		2	33	8226	51159
	9	9	80	321		3	34	9006	60165
	10	10	75	396		4	35	13354	73519
	11	11	83	479		5	36	17077	90596
	12	12	65	544		6	37	19982	110578
	13	13	169	713		7	38	21222	131800
	14	14	138	851		8	39	23624	155424
	15	15	202	1053		9	40	24943	180367
	16	16	158	1211		10	41	26087	206454
	17	17	260	1471		11	42	23342	229796
	18	18	374	1845		12	43	26330	256126
	19	19	509	2354		13	44	27719	283845
	20	20	758	3112		14	45	23072	306917
	21	21	896	4008		15	46	23513	330430
	22	22	981	4989		16	47	24820	355250
	23	23	983	5972		17	48	22248	377498
	24	24	1609	7581					
	25	25	2269	9850					

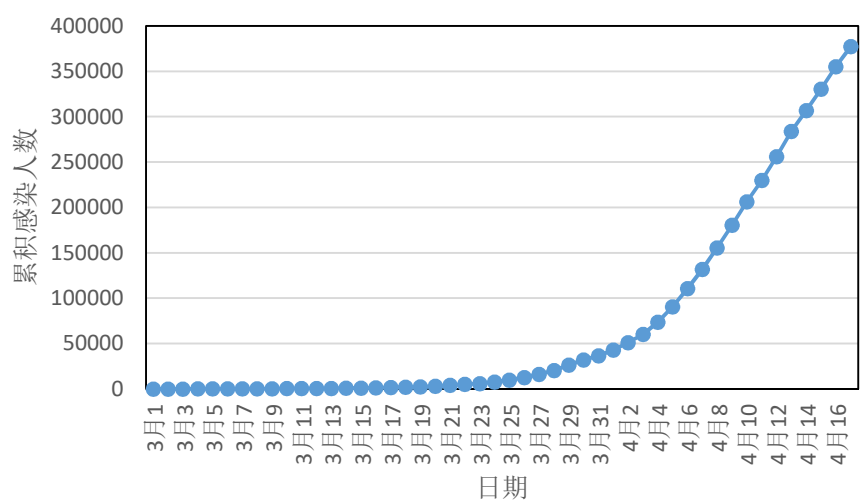


图 23-1 3月1日至4月17日累积感染数

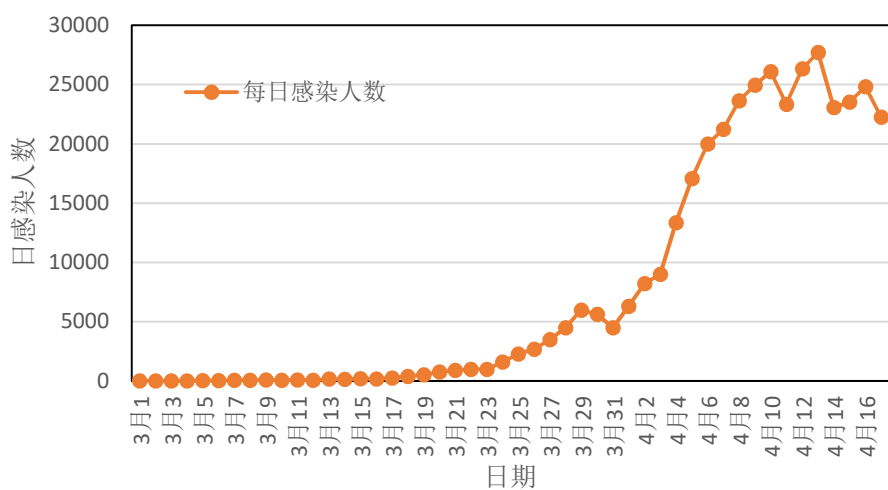


图 23-2 3月1日至4月17日每日感染数

下表 23.2 及图 23-3 为上篇公众号基于 4 月 10 日为止的数据对 4 月 11 日至 17 日做的预测，可以看出误差还是比较大的，预测值普遍偏低。

表 23.2: 上海 4 月 11 日至 4 月 17 日计算及预测数据（基于 4 月 10 日为止的数据）

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
	11	23342	25917	22582
	12	26330	24933	17994
	13	27719	23203	12862
	14	23072	20938	8404
	15	23513	18379	5146
	16	24820	15748	3020
	17	22248	13217	1727

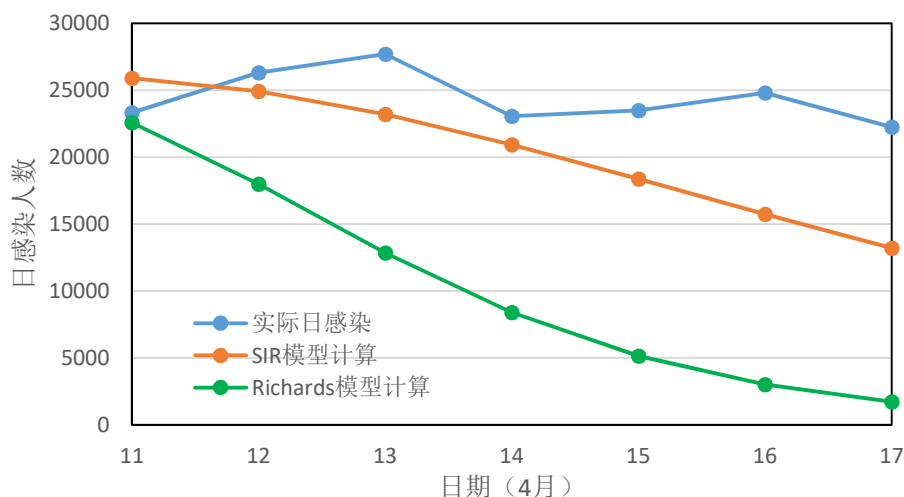


图 23-3 4月1日至4月17日每日感染数预测及实际值

23.2 模型再计算

采用上述表 23.1 最新数据（截至 4 月 17 日），对 Richards 模型和 SIR 模型重新计算。

Richards 模型求解代码

```
Variable t,y;
Function y=a*(1+(d-1)*exp(-k*(t-xc)))^(1/(1-d));/+p1;
Data;
t=1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48;
y=2,10,26,45,73,121,176,241,321,396,479,544,713,851,1053,1211,1471,1845,2354,3112,4008,4989,5972,7581,9850,12526,16026,20503,26485,32120,36622,42933,51159,60165,73519,90596,110578,131800,155424,180367,206454,229796,256126,283845,306917,330430,355250,377498;
```

结果

```
Sum Squared Error (SSE): 109570552.533752
Root of Mean Square Error (RMSE): 1510.8672491166
Correlation Coef. (R): 0.999905605487001
R-Square: 0.999811219884326
Adjusted R-Square: 0.999802829656963
Determination Coef. (DC): 0.999811109819955
Chi-Square: 1164.37589073453
F-Statistic: 77659.9706095233
```

Parameter Best Estimate

```
-----
a      482565.321049775
d      2.01501990178507
k      0.224566919908364
xc     42.4434208751356
```

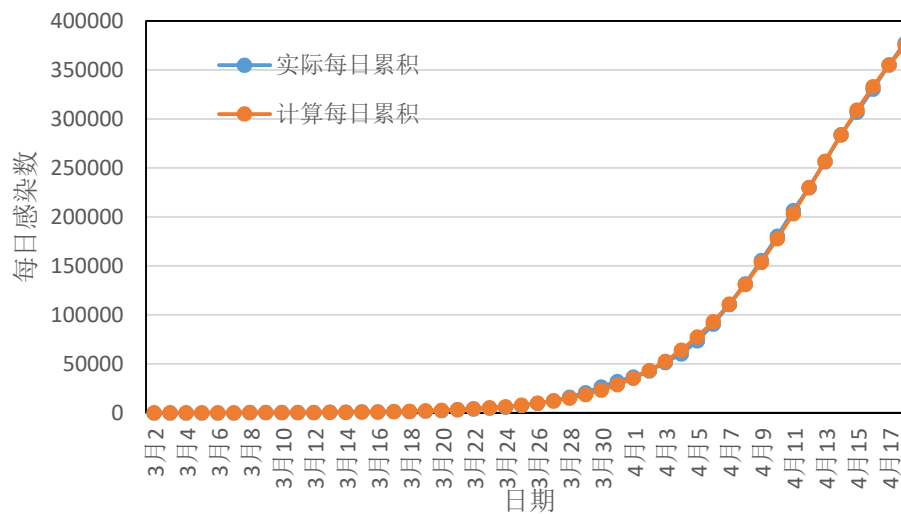


图 23-4 Richards 模型模拟计算累积感染对比图

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,
40,41,42,43,44,45,46,47,48;
y=2,8,16,19,28,48,55,65,80,75,83,65,169,138,202,158,260,374,509,758,896,981,983,1609,2269,2676,3500,447
7,5982,5635,4502,6311,8226,9006,13354,17077,19982,21222,23624,24943,26087,23342,26330,27719,23072,
23513,24820,22248;
```

结果

Sum Squared Error (SSE): 110009586.664722
Root of Mean Square Error (RMSE): 1513.8911416771
Correlation Coef. (R): 0.968835735136445
R-Square: 0.938642681677375
Adjusted R-Square: 0.957679581468557
Determination Coef. (DC): 0.956143013758168
F-Statistic: 1978.25662513264

Parameter	Best Estimate
beta1	2.23120523234151E-7
gamma1	5.31455419167934

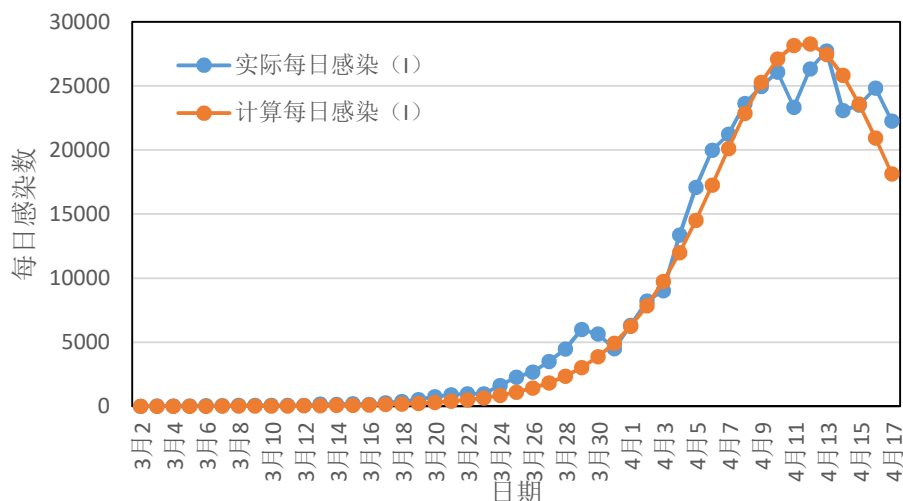


图 23-5 SIR 模型模拟计算每日感染对比图

23.3 模型修改

将原 Richards 模型增加一项常数项系数 y_0 ，模型变为：

$$y = y_0 + a \cdot \left(1 + (d - 1) \cdot \exp(-k \cdot (t - xc))\right)^{\frac{1}{1-d}} \quad (23-1)$$

将原 SIR 模型中的感染数项 I 增加一个指数项参数，修改后 SIR 模型：

$$\begin{cases} \frac{dS}{dt} = -\beta \cdot S \cdot I^\alpha \\ \frac{dI}{dt} = \beta \cdot S \cdot I^\alpha - \gamma \cdot I^\alpha \\ \frac{dR}{dt} = \gamma \cdot I^\alpha \end{cases} \quad (23-2)$$

修改 Richards 模型求解代码

```
Variable t,y;
Function y=y0+a*(1+(d-1)*exp(-k*(t-xc)))^(1/(1-d));
Data;
t=1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48;
y=2,10,26,45,73,121,176,241,321,396,479,544,713,851,1053,1211,1471,1845,2354,3112,4008,4989,5972,7581,9850,12526,16026,20503,26485,32120,36622,42933,51159,60165,73519,90596,110578,131800,155424,180367,206454,229796,256126,283845,306917,330430,355250,377498;
```

结果

```
Sum Squared Error (SSE): 107523422.051448
Root of Mean Square Error (RMSE): 1496.68677175236
Correlation Coef. (R): 0.999907315152911
R-Square: 0.999814638896303
Adjusted R-Square: 0.999806400625027
Determination Coef. (DC): 0.999814638896303
Chi-Square: 2634.63991355178
F-Statistic: 57984.1571158275
```

Parameter	Best Estimate
y0	505779.177251123
a	-505555.897033348
d	2.23203024048203
k	-0.22676762334859
xc	42.4605762214313

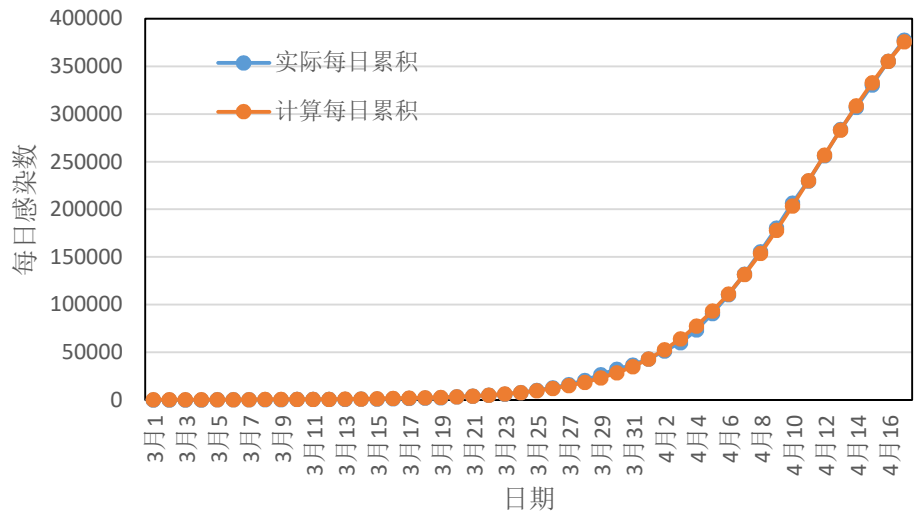


图 23-6 修改 Richards 模型模拟计算累积感染数

修改 SIR 模型求解代码

```

Constant N=25000000;
Parameter beta1=1E-7[0,],gamma1=5.0[0,];
Parameter alpha=[0.5,1.5];
InitialODEValue t=1,S=N,I=2,R=0;
Variable t, I;
ODEFunction S'=-beta1*S*I^alpha;
               I'=beta1*S*I^alpha-gamma1*I^alpha;
               R'=gamma1*I^alpha;

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,
40,41,42,43,44,45,46,47,48;
y=2,8,16,19,28,48,55,65,80,75,83,65,169,138,202,158,260,374,509,758,896,981,983,1609,2269,2676,3500,447
7,5982,5635,4502,6311,8226,9006,13354,17077,19982,21222,23624,24943,26087,23342,26330,27719,23072,
23513,24820,22248;

```

结果

```

Sum Squared Error (SSE): 56331268.3324044
Root of Mean Square Error (RMSE): 1083.31347737321
Correlation Coef. (R): 0.97325996605523
R-Square: 0.947234961525827
Adjusted R-Square: 0.966853543705202
Determination Coef. (DC): 0.967377225316128
F-Statistic: 1813.22965566918

```

Parameter	Best Estimate
beta1	3.06588548268465E-7
gamma1	7.31280041342844
alpha	0.950365856870449

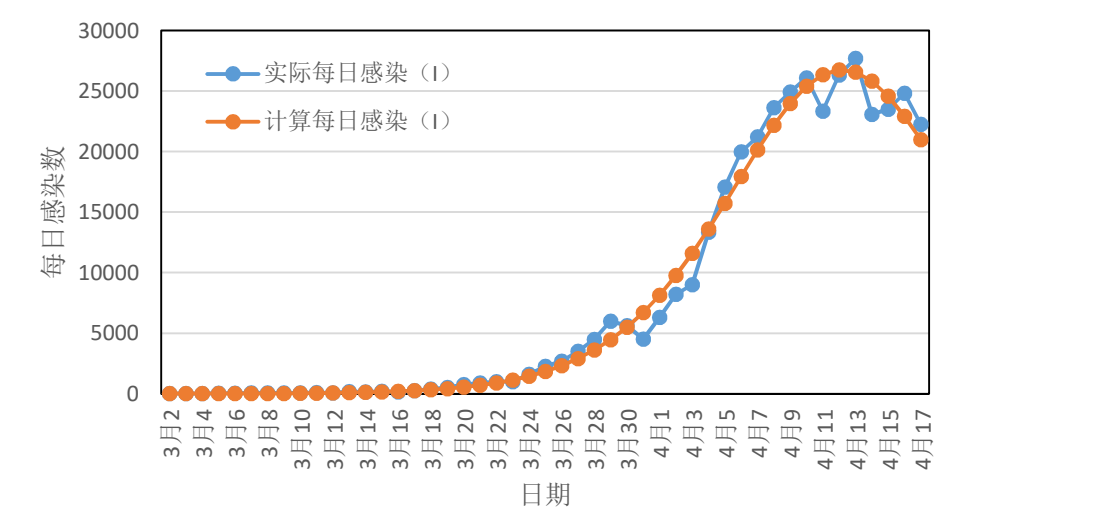


图 23-7 修改 SIR 模型模拟计算每日感染数

两种修改后的模型，相比于原始模型，模拟计算结果都有一定程度的提高。

23.4 增长型 S 函数每日感染数模拟

S 型增长模型，如 Richards 模型等，均是以累积数为模拟计算目标的，因为模拟的是累积数据而非每日数据，因此“累积效应”有可能导致通过计算累积数据转换而来的每日计算数据出现误差，有无可能使用 S 模型，但计算模拟目标值为每时刻（日）而非日累积数据？如同 SIR 模型中的 I 一样为每一时刻的数值。1stOpt 编程模式下直接支持标准的 Fortran、Basic 和 Pascal 等高级语言，可以很容易去实现这一构想。下面以 Fortran 及 Pascal 代码分别实现，其实现原理就是用 t 时刻计算出的累积值减去 t-1 时刻的累积值从而得到对应 t 时刻的日感染数值，代码中的因变量 y 为每日感染实际数据。

修改 Richards 模型求解每日感染数代码（Fortran）

```

Variable x,y;
Parameter y0,a,d,k,xc;
StartProgram [Fortran];
Subroutine MainModel
  integer i
  real(8) temd, tem_sum, tem_sum1, tem_sum2

```

```

do i = 0, DataLength - 1
    tem_sum = y0+a*(1+(d-1)*exp(-k*(x(i)-xc)))^(1/(1-d))/modified Richards
    if (i .eq. 0) then
        tem_sum1 = tem_sum
        tem_sum2 = tem_sum
        temd = tem_sum
    else
        tem_sum1 = tem_sum2
        tem_sum2 = tem_sum
        temd = tem_sum2 - tem_sum1
    end if
    y(i) = temd
end do
End Subroutine
EndProgram;
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
,40,41,42,43,44,45,46,47,48;
y=2,8,16,19,28,48,55,65,80,75,83,65,169,138,202,158,260,374,509,758,896,981,983,1609,2269,2676,3500,447
7,5982,5635,4502,6311,8226,9006,13354,17077,19982,21222,23624,24943,26087,23342,26330,27719,23072,
23513,24820,22248;

```

修改 Richards 模型求解每日感染数代码（Pascal）

```

Variable x,y;
Parameter y0,a,d,k,xc;
StartProgram [Fortran];
Subroutine MainModel
    integer i
    real(8) temd, tem_sum, tem_sum1, tem_sum2
    do i = 0, DataLength - 1
        tem_sum = y0+a*(1+(d-1)*exp(-k*(x(i)-xc)))^(1/(1-d))/modified Richards
        if (i .eq. 0) then
            tem_sum1 = tem_sum
            tem_sum2 = tem_sum
            temd = tem_sum
        else
            tem_sum1 = tem_sum2
            tem_sum2 = tem_sum
            temd = tem_sum2 - tem_sum1
        end if
        y(i) = temd
    end do
End Subroutine
EndProgram;
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
,40,41,42,43,44,45,46,47,48;
y=2,8,16,19,28,48,55,65,80,75,83,65,169,138,202,158,260,374,509,758,896,981,983,1609,2269,2676,3500,447
7,5982,5635,4502,6311,8226,9006,13354,17077,19982,21222,23624,24943,26087,23342,26330,27719,23072,

```

23513,24820,22248;

上面两段代码均可得到下面一致的结果

Sum Squared Error (SSE): 47407131.4308184
Root of Mean Square Error (RMSE): 993.805097328135
Correlation Coef. (R): 0.99509263381025
R-Square: 0.990209349863421
Adjusted R-Square: 0.989774209857351
Determination Coef. (DC): 0.990002086783823
F-Statistic: 1083.81995721627

Parameter	Best Estimate
-----------	---------------

y0	651162.167261287
a	-651170.077310556
d	4.28160656678643
k	-0.268529901146463
xc	42.3099032545047

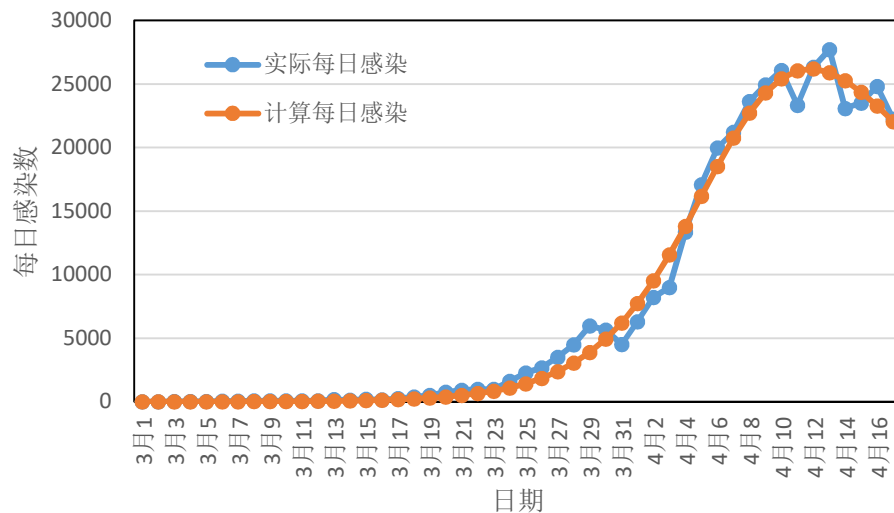


图 23-8 Richards 模型直接模拟计算每日感染数

23.5 数据预处理-移动平均

为了消除人为统计误差而引起的每日实际感染数据上下起伏较大可能导致的对模拟计算精度的影响，取每日实际感染数据 7 天移动平均值作为模拟目标，以修改后 SIR 模型进行仿真模拟，代码中的因变量 y 为每日实际感染数据 7 天移动平均值（少于 7 日的头 6 天取对应天数平均值）。

表 23.3 数据 7 天移动平均预处理

序号	日期	每日感染人数	7 天移动平均	序号	日期	每日感染人数	7 天移动平均
1	3 月 1	2	2	25	3 月 25	2269	1144
2	3 月 2	8	5	26	3 月 26	2676	1453

3	3 月 3	16	9	27	3 月 27	3500	1845
4	3 月 4	19	11	28	3 月 28	4477	2356
5	3 月 5	28	15	29	3 月 29	5982	3071
6	3 月 6	48	20	30	3 月 30	5635	3735
7	3 月 7	55	25	31	3 月 31	4502	4149
8	3 月 8	65	34	32	4 月 1	6311	4726
9	3 月 9	80	44	33	4 月 2	8226	5519
10	3 月 10	75	53	34	4 月 3	9006	6306
11	3 月 11	83	62	35	4 月 4	13354	7574
12	3 月 12	65	67	36	4 月 5	17077	9159
13	3 月 13	169	85	37	4 月 6	19982	11208
14	3 月 14	138	96	38	4 月 7	21222	13597
15	3 月 15	202	116	39	4 月 8	23624	16070
16	3 月 16	158	127	40	4 月 9	24943	18458
17	3 月 17	260	154	41	4 月 10	26087	20898
18	3 月 18	374	195	42	4 月 11	23342	22325
19	3 月 19	509	259	43	4 月 12	26330	23647
20	3 月 20	758	343	44	4 月 13	27719	24752
21	3 月 21	896	451	45	4 月 14	23072	25017
22	3 月 22	981	562	46	4 月 15	23513	25001
23	3 月 23	983	680	47	4 月 16	24820	24983
24	3 月 24	1609	873	48	4 月 17	22248	24435

修改 SIR 模型 7 天移动平均日感染数代码

```

Constant N=25000000;
Parameter beta1=1E-7[0,],gamma1=5.0[0,];
Parameter alpha=[0.5,1.5];
InitialODEValue t=1,S=N,I=2,R=0;
Variable t, I;
ODEFunction S'=-beta1*S*I^alpha;
               I'=beta1*S*I^alpha-gamma1*I^alpha;
               R'=gamma1*I^alpha;

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
,40,41,42,43,44,45,46,47,48;
y=2,5,9,11,15,20,25,34,44,53,62,67,85,96,116,127,154,195,259,343,451,562,680,873,1144,1453,1845,2356,30
71,3735,4149,4726,5519,6306,7574,9159,11208,13597,16070,18458,20898,22325,23647,24752,25017,25001,
24983,24435;

```

结果

```

Sum Squared Error (SSE): 4917704.39659076
Root of Mean Square Error (RMSE): 320.081512944709
Correlation Coef. (R): 0.978551985648012
R-Square: 0.957563988615666
Adjusted R-Square: 0.977881827948126
Determination Coef. (DC): 0.977895761341521
F-Statistic: 17057.151865827

```

Parameter	Best Estimate
beta1	2.5460333481101E-7
gamma1	6.07936754084803
alpha	0.973717016759388

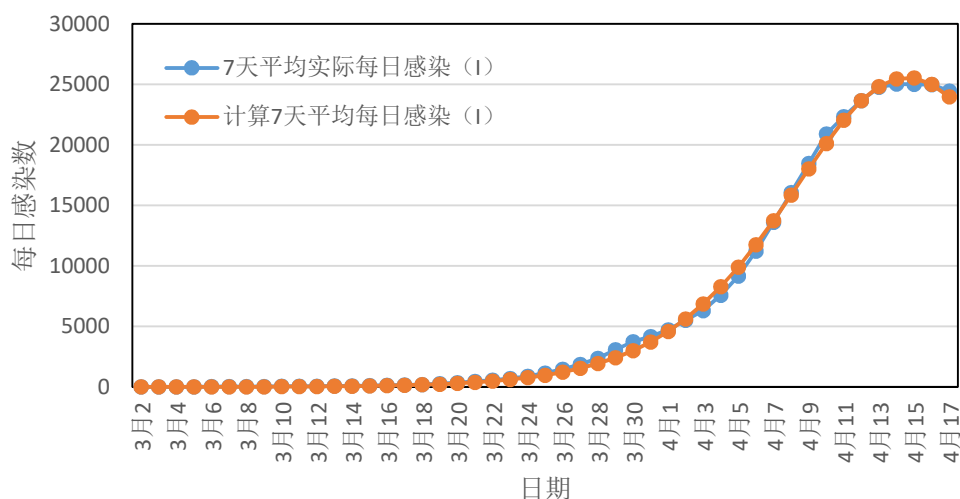


图 23-9 修改 SIR 模型 7 天移动平均日感染数模拟对比

23.6 模型预测

下表 23.4 及图 23-10 为前述 6 种模式计算模拟结果，4 月 18 日至 4 月 31 日为预测结果。

表 23.4 六种模型模式计算及预测结果

日期	Richards 累积模拟 每日	修 改 Richards 累积模拟 每日	修 改 Richards 每日模拟	Sir 模型	修改 Sir 模 型	修改 Sir 模 型 7 天移 动平均	实 际 日 感 染 数
3 月 2	12	11	3	3	3	3	8
3 月 3	15	13	4	3	4	3	16
3 月 4	19	17	5	4	5	5	19
3 月 5	24	21	7	6	7	6	28
3 月 6	30	26	9	7	10	8	48
3 月 7	37	33	12	10	14	10	55
3 月 8	46	41	15	13	19	14	65
3 月 9	58	52	20	16	26	18	80
3 月 10	72	65	26	21	35	23	75
3 月 11	90	82	34	28	46	30	83
3 月 12	112	102	45	36	62	39	65
3 月 13	140	128	58	47	82	50	169
3 月 14	174	161	76	61	109	65	138
3 月 15	217	202	100	80	143	84	202
3 月 16	270	253	131	104	188	108	158
3 月 17	337	317	171	135	246	139	260
3 月 18	420	397	223	176	320	178	374
3 月 19	523	497	291	228	416	228	509
3 月 20	651	622	380	297	538	292	758
3 月 21	810	778	496	386	692	372	896
3 月 22	1007	972	646	501	888	474	981
3 月 23	1250	1213	841	650	1134	603	983
3 月 24	1551	1513	1093	844	1443	764	1609
3 月 25	1922	1884	1417	1093	1828	967	2269
3 月 26	2377	2341	1834	1414	2304	1220	2676
3 月 27	2934	2902	2366	1826	2889	1535	3500
3 月 28	3612	3587	3039	2353	3602	1926	4477
3 月 29	4433	4419	3884	3022	4465	2408	5982
3 月 30	5417	5419	4931	3868	5496	2999	5635
3 月 31	6589	6609	6210	4927	6715	3718	4502

4月1	7965	8009	7741	6237	8136	4587	6311
4月2	9561	9628	9534	7835	9763	5624	8226
4月3	11376	11466	11572	9747	11592	6850	9006
4月4	13396	13503	13810	11980	13598	8278	13354
4月5	15584	15695	16166	14509	15740	9911	17077
4月6	17872	17970	18529	17262	17949	11740	19982
4月7	20163	20227	20761	20108	20135	13739	21222
4月8	22331	22340	22727	22857	22188	15857	23624
4月9	24230	24169	24308	25273	23988	18016	24943
4月10	25711	25578	25426	27108	25412	20115	26087
4月11	26643	26459	26049	28150	26359	22033	23342
4月12	26939	26742	26193	28271	26755	23640	26330
4月13	26568	26417	25911	27461	26572	24815	27719
4月14	25565	25526	25276	25829	25828	25463	23072
4月15	24024	24157	24371	23573	24582	25530	23513
4月16	22078	22430	23275	20934	22932	25012	24820
4月17	19876	20470	22056	18145	20991	23958	22248
4月18	17565	18399	20772	15403	18879	22458	
4月19	15270	16321	19468	12845	16707	20628	
4月20	13086	14314	18175	10556	14570	18594	
4月21	11078	12434	16918	8571	12538	16478	
4月22	9281	10715	15710	6891	10663	14380	
4月23	7708	9173	14563	5497	8973	12379	
4月24	6356	7809	13481	4358	7480	10531	
4月25	5209	6619	12465	3438	6183	8865	
4月26	4248	5589	11517	2701	5073	7396	
4月27	3451	4706	10634	2116	4135	6123	
4月28	2794	3953	9814	1654	3350	5034	
4月29	2256	3314	9054	1290	2699	4115	
4月30	1818	2774	8350	1005	2163	3347	
4月31	1463	2319	7700	782	1726	2711	
合计	476614	493902	560693	434545	491388	456496	

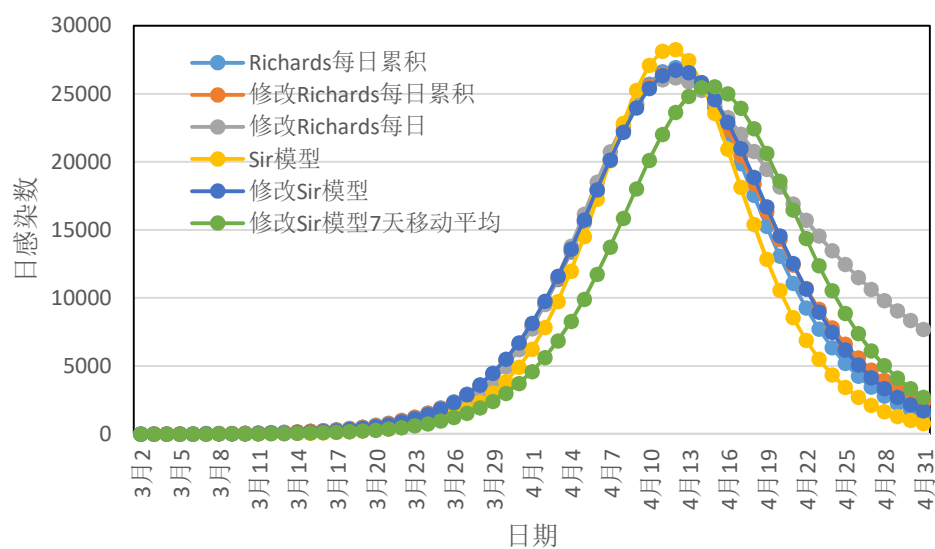


图 23-10 六种模式日感染数模拟计算及预测结果

23.7 结论

不论何种模型模式，从计算结果看，数学意义上的拐点已经出现，在 4 月 11

日至 16 日之间，不同的模型衰减程度有急有缓，也意味着疫情完全消失的时间长短不同。影响疫情走势的原因复杂而多样，很多又是无法在数学模型中量化体现，因此计算结果仅供建模爱好者参考。

1stOpt 在求解这类优化模型中，相比其它诸如 **Matlab**、**OriginPro**、**R** 等软件，拥有巨大的优势，使用这些软件即使多化费 10 倍的时间和精力也很难保证能得到比 **1stOpt** 更好的结果。