

基于 SIRD, SCM 模型的疾病发展趋势预测

——以上海市 Cov-19 数据为例

王铎磊, 袁文璐, 赵 钊

摘要

突发的全球性公共卫生事件会对人类的生产生活造成重大影响, 寻找合适的数学模型来预测分析疾病的发展趋势可以帮助人们控制疾病。本文基于 SI (Susceptible Infected) 模型, 从疾病发展的动力学角度, 对疫情发展进行了建模, 并结合了上海市有关的实际数据对模型进行验证和评估。

此外, 我们扩展了 SIRD 模型, 建立了 SCM (Seven Compartment Model), 充分考虑了被隔离人群 (Quarantined), 无症状感染者 (Asymptomatic), 潜伏期患者 (Exposed) 和确诊患者 (Diagnosed) 的影响, 以更好的适用于真实世界中疾病传播收到了管控政策、疾病潜伏期等因素的影响。

利用有关数据, 通过统计学方法可以得到我们模型所需要的关键参数, 利用这些参数我们对 SIRD 模型进行了理论计算和分析。我们额外设计了 4 种常见的政策和传播情况, 作为 SCM 模型的预测结果, 并与现实世界中的对应情况进行了分析, 结果发现 SCM 模型对于包含了政策导向、医疗条件等现实因素影响的真实世界传播具有优秀的估计效果。

最后, 我们总结了模型的不足和未来可能可以改进的方向。SIRD 模型具有宏观尺度的优秀预测效果, 但是没办法充分考量政策、病毒潜伏期等因素的影响。SCM 模型具有很强的可扩展性, 但是模型需要的参数需要额外的设计和计算, 可能增加 SCM 模型使用的成本。此外, 缺乏优秀的数据库可能也会使得统计学方法学习 SCM 相关参数变得困难。

综上所述, SIRD 模型在少量参数的情况下对于简单的预测具有良好的效果, SCM 模型在提供了有关参数的情况下能适用于多种情况, 具有更高的可塑性。以上模型的合理使用可以有效的预测疾病发展的趋势, 有利于政府在合理的范围内采取有效措施, 进而控制疾病。

Keywords: SI 模型, SIRD 模型, SCM 模型

目录

1 问题重述	1
2 问题分析	1
3 模型假设	1
3.1 SI 模型	1
3.2 SIRD 模型	2
3.3 SCM 模型	2
4 符号说明	3
5 模型的建立与求解	5
5.1 SIRD	5
5.2 SCM	7
6 模型检验与评价	10
6.1 SIRD	10
6.2 SCM	10
6.2.1 优点	10
6.2.2 缺点和评价	11
参考文献	12
7 附录：代码	13

1 问题重述

2020 年 1 月从中国武汉开始，逐步发展到全球流行的新型冠状病毒，简称 COVID-19，对大到全球政治、经济，小到对每个人的日常生活，都造成了极大的影响。我们需要查找疫情相关的数据，并建立病毒传播和预测的模型，对新冠病毒的传播进行预测，并给出相应对策。

2 问题分析

假设病毒的传染率、康复率等参数一定，总人数短时间内不改变。那么以天为单位，每天的感染者、康复者等数量成马尔可夫过程，只与前一天的易感者、感染者、康复者数量相关。那么根据此条件建立微分方程模型，进行拟合与预测。

3 模型假设

3.1 SI 模型

1. 假设预测的过程中，不考虑总人口的自然变化率。
2. 仅仅把人群分为易感人群 (Susceptible) 和易感染 (Infected) 两类，认为他们之间的变化满足以下的微分方程：

$$\begin{cases} S(t) + I(t) = N(t) \\ \frac{dS}{dt} = -\beta \frac{SI}{N} \\ \frac{dI}{dt} = \beta \frac{SI}{N} \end{cases}$$

可以注意到，SI 模型对人群的分类十分粗糙，往往不适用于现实社会中的疾病建模。因此扩展以下因素为 SIRD 模型。

3.2 SIRD 模型

1. 进行短期预测，不考虑总人口的自然变化。
2. 假设人会无意识的暴露在环境中，不存在居家隔离等行为。
3. 假设人口的因病死亡率、感染率、康复率均为定值。
4. 假设康复后不会再感染。

3.3 SCM 模型

1. 预测时间内，不考虑总人口 N 的自然变化。
2. 假设易感人群 S 和隔离人群 Q 在一定比例上相互转化，隔离人群不会受到感染。
3. 假设易感人群 S 与已感染人群，即有症状感染者 I 、无症状感染者 A 接触，会有不同程度的感染病毒的可能。且因接触某类已感染人群造成的感染人数，正比与两种人群的乘积。
4. 假设感染病毒的人群 E 具有一定的潜伏期，潜伏期内不具有传染性，潜伏期后一定比例上转化为有症状感染者 I 和无症状感染者 A 。
5. 假设无症状感染者有分别有一定比例因检测转化为确诊患者 D ，有一定概率转化为康复人群 R 。
6. 假设确诊患者有一定概率转化为康复人群 R 和死亡人群。而无症状感染者人群转化为死亡人群的部分不予考虑。
7. 假设康复人群不会受到二次感染。

根据模型的假设，可以列出下面的微分方程组：

$$\begin{cases} dS = -\beta * S * (I + \theta * A) - p * S + \lambda * Q \\ dQ = p * S - \lambda * Q \\ dE = \beta * S * (I + \theta * A) - \rho * E \\ dA = \sigma * (1 - \rho) * E - \varepsilon_A * A - \gamma_A * A \\ dI = \sigma * \rho * E - \gamma_I * I - d_I * I - \epsilon_I * I \\ dD = \varepsilon_A * A + \epsilon_I * I - (d_D + \gamma_D) * D \\ dR = \gamma_A * A + \gamma_I * I + \gamma_D * D \end{cases}$$

4 符号说明

序号	符号	符号说明
1	σ	病毒感染能力
2	ρ	有症状感染者的感染能力折扣
3	θ	无症状感染者感染能力的折扣
4	β	人群流通能力
5	p	被隔离的可能性
6	λ	解除隔离的可能性
7	γ_A	无症状感染者恢复的比例
8	γ_I	有症状感染者恢复的比例
9	γ_D	确诊患者恢复的比例
10	ε_I	有症状感染者确诊的比例
11	ε_A	无症状感染者确诊的比例
12	d_D	确诊患者的死亡比例
13	d_I	有症状感染者的死亡比例

表 1: SCM 的符号说明

序号	符号	符号说明
1	N'	总人口数量
2	S	易感者数量
3	I	感染者数量
4	R	康复者数量
5	ID	因病死亡数量
6	S_x	第 x 天易感者数量
7	I_x	第 x 天感染者数量
8	R_x	第 x 天康复者数量
9	ID_x	第 x 天因病死亡数量
10	dS_x	第 x 天易感者数量变化量
11	dI_x	第 x 天感染者数量变化量
12	dR_x	第 x 天康复者数量变化量
13	dID_x	第 x 天因病死亡数量变化量
14	\widehat{S}_x	第 x 天易感者数量拟合值/预测值
15	\widehat{I}_x	第 x 天感染者数量拟合值/预测值
16	\widehat{R}_x	第 x 天康复者数量拟合值/预测值
17	\widehat{ID}_x	第 x 天因病死亡数量拟合值/预测值
18	β	感染率
19	γ	康复率
20	δ	因病死亡率
21	β_x	第 x 天感染率
22	γ_x	第 x 天康复率
23	δ_x	第 x 天因病死亡率

表 2: SIRD 的符号说明

5 模型的建立与求解

考虑到 SI 模型忽略了很多因素，只有极其宏观的观测尺度下具有良好的表现，因此本文不对其表现进行进一步的阐述，进而不进行相关模型的解释。下面给出 SIRD 模型和 SCM 模型的详细介绍。

5.1 SIRD

SIRD 模型主要简化了病毒潜伏期、政策等因素的影响。根据前面模型的基本假设，我们可以根据事实可以得出以下条件：

1. 易感者可以转化为感染者，转化速度为：

$$\beta \frac{S \times I}{N'}$$

2. 感染者转化为康复者，转化速度为 γI 。
3. 因病死亡率为 δ ，感染者死亡的速度为 δI 。

因此，我们可以绘制出 SIRD 模型下，不同人群转移的状态图。

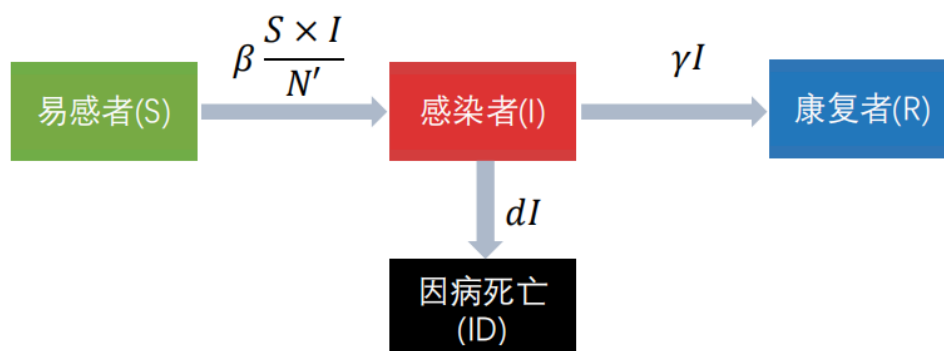


图 1: SIRD 状态转移图

根据 SIRD 模型，我们把转化过程用动力学角度阐释，可以列出一个微分方程组，具体如下：

$$\begin{cases} dS/dt = -\beta \frac{S \times I}{N'} \\ dT/dt = \beta \frac{S \times I}{N'} - \gamma I - \delta I \\ dR/dt = \gamma I \\ dID/dt = \delta I \\ N' = S + I \end{cases}$$

我们额外找到了上海市 2020 年 1 月 20 日到 2020 年 6 月 18 日共 150 天的数据，根据这些数据我们可进行基本的拟合，来得到一些重要参数的参考数值。以下是计算过程：

由微分方程

$$\begin{aligned} \frac{dS}{dt} = S' &= -\beta \frac{S \times I}{N'} \\ &= -\beta \frac{S \times I}{S + I} \end{aligned}$$

代换可以得到参数 β

$$\beta = -S' \frac{S + I}{SI}$$

我们采用参考数据的平均值，作为 β 的估计值 $\hat{\beta}$ ，即

$$\hat{\beta} = \frac{1}{150} \sum_{i=1}^{150} \beta_i = -\frac{1}{150} \sum_{i=1}^{150} dS_i \frac{S_i + I_i}{S_i I_i}$$

带入已有数据计算可得， $\hat{\beta} = 0.07417$ ，同理

$$\hat{\gamma} = \frac{1}{150} \sum_{i=1}^{150} \frac{dR_i}{I_i} = 0.07245$$

$$\hat{\delta} = \frac{1}{150} \sum_{i=1}^{150} \frac{dID_i}{I_i} = 0.00057$$

我们将得到的参考数值 $\hat{\beta} = 0.07417$ ， $\hat{\gamma} = 0.07245$ ， $\hat{\delta} = 0.00057$ 代入 SIRD 模型的微分方程组，可以计算出拟合值：

$$\widehat{S}_x, \widehat{I}_x, \widehat{R}_x, \widehat{ID}_x, \quad (x = 1, 2, \dots, 150)$$

我们可以通过相关系数检验我们的参数的你 and 效果，经过计算发现发现，平均相关系数为 $0.84 > 0.75$ ，拟合效果良好。

表 3: 相关系数

	$(S., \widehat{S.})$	$(I., \widehat{I.})$	$(R., \widehat{R.})$	$(ID., \widehat{ID.})$	AVG
Cov	0.94	0.56	0.97	0.92	0.84

5.2 SCM

考虑 SCM 模型假设中的微分方程组，可以注意到 SCM 模型的表现受到初试参数的影响较大。因此，我们需要根据不同地区政策去指定不同的参数组。首先，我们对 SCM 模型进行简要的分析。

SCM 模型中，易感人群 S 与已感染人群 I ，无症状感染者 A 之间的接触是导致 S 流向 E 的重要因素。而易感人群 S 和被隔离人群 Q 之间的流通情况完全收到参数 λ, p 的影响，而这部分决定了疫情爆发时候的主要人群的流向。

因此，通过调节 λ, p 参数，可以有效的反映出政策的影响。考虑到实际情况的不同，我们给出了五种基本参数，以作为 SCM 模型的初试参数，对应真实世界中的以下情况：

1. $p = 0.99, \lambda = 0.1$. 政府实行高度管控，对于所有人群采取限制流通措施。密接、次密接等被隔离人群需要经过密切观察后才能转化为普通人群（易感染人群）。
2. $p = 0.9, \lambda = 0.3$. 政府对密接、次密接人群实行强力管控，对于密接人群采取强力隔离，检测为阴性后解除管控。
3. $p = 0.7, \lambda = 0.9$. 政府对密接、次密接人群实行一般管控，经过简单的医学检查后解除管控。
4. $p = 0.1, \lambda = 0.9$. 政府实行最低限度的管控，对于密接人群进行医学筛查后解除管控。
5. $p = 0, \lambda = 1.0$. 政府不进行额外管控。

除去上述与政策有关的参数之外，值得注意的是 SCM 模型中对于无症状感染者、有症状感染者向其他人群的转化分别设计了不同参数，为了符合一般的认知，我们认为：无症状感染者 A 转化为确诊患者 D 的概率比起有症状感染者 I 转化的概率更小，而无症状感染者 A 转化为恢复人群 R 的能力会略强于有症状感染者直接转化为恢复人群 R 。另外，考虑到确诊患者 D 往往会受到更好的医疗资源的照顾，其转化为恢复人

群的能力会更强。SCM 模型还考虑了确诊人群 D 和有症状感染者 I 的死亡率 d_D, d_I ，我们认为两类人群的死亡率都很低，但是 D 的死亡率应该略低于有症状感染者 I，因为有症状感染者在转化为确诊患者 D 的过程中可能拥有略差的医疗条件。

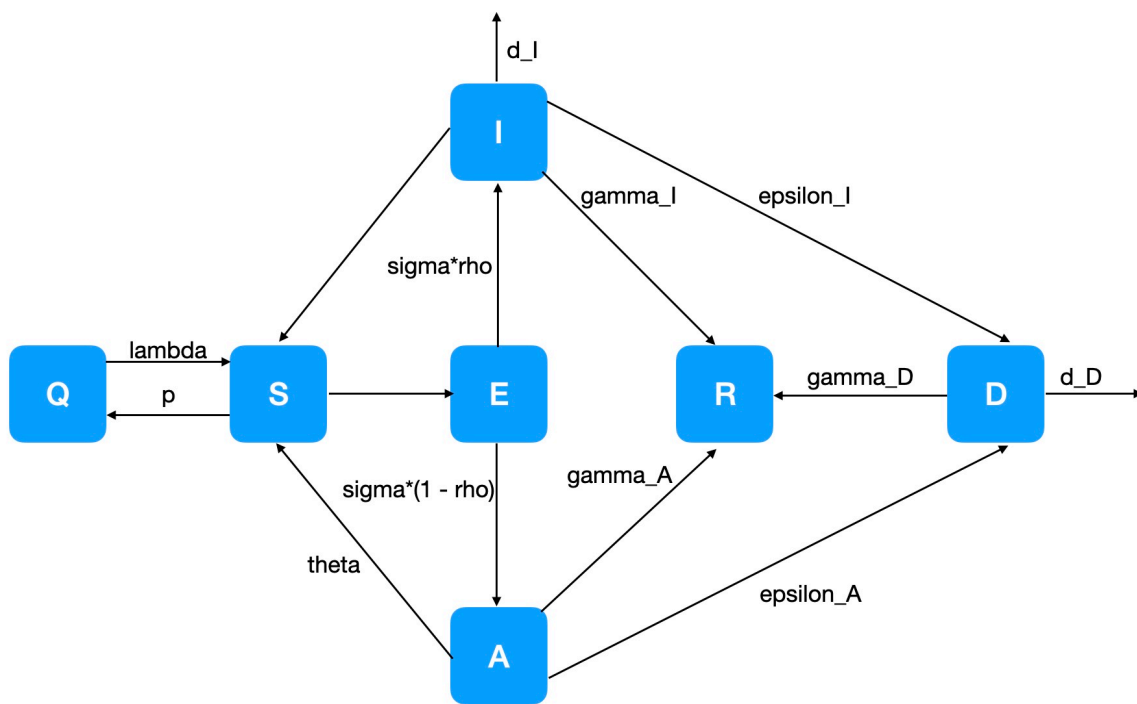


图 2: Caption

下面是基于 **Julia** 语言的 **DifferentialEquations.jl** 库，我们编写的代码 **SCM.jl** 求解的 SCM 预测结果。值得一提的是，**Julia** 语言拥有良好的对数学符号的支持，并且借助宏（macro）定义，可以达到直接书写微分方程表达式级别的简便，在有关领域的计算中具有较高的速度和简易的语法，这也是我们这部分采用 **Julia** 语言的重要原因。额外需要说明的是，有关新冠病毒（Cov-19）的优秀数据库的缺失使得我们难以建立和真实世界中时间的对应，因此在没有相关数据的情况下无法保证对应，以下分析的过程使用“初期”、“早期”、“中期”等相对模糊的时间概念：

首先，对于公共卫生事件而言，最好的结果应当是人类群体产生对病毒的抗性，这个过程可以通过接种疫苗、弱化病毒等手段实行，因此选择合理的策略让群体免疫逐渐实行是十分重要的。如果管控过于松散，容易因疫情早期的医疗资源不足造成诸多恶劣的社会影响。如果管控一直很严厉，开放过晚，不利于经济的发展和人民的正常生产生活。观察上述模型，我们可以注意到以下四种情况：

- 采用强力管控的措施的预测结果为左上角的图形。当大部分人群转化为限制出行的人群 Q，此后限制出行的人群保持长期稳定。这恰好对应了我国早期控制疫情

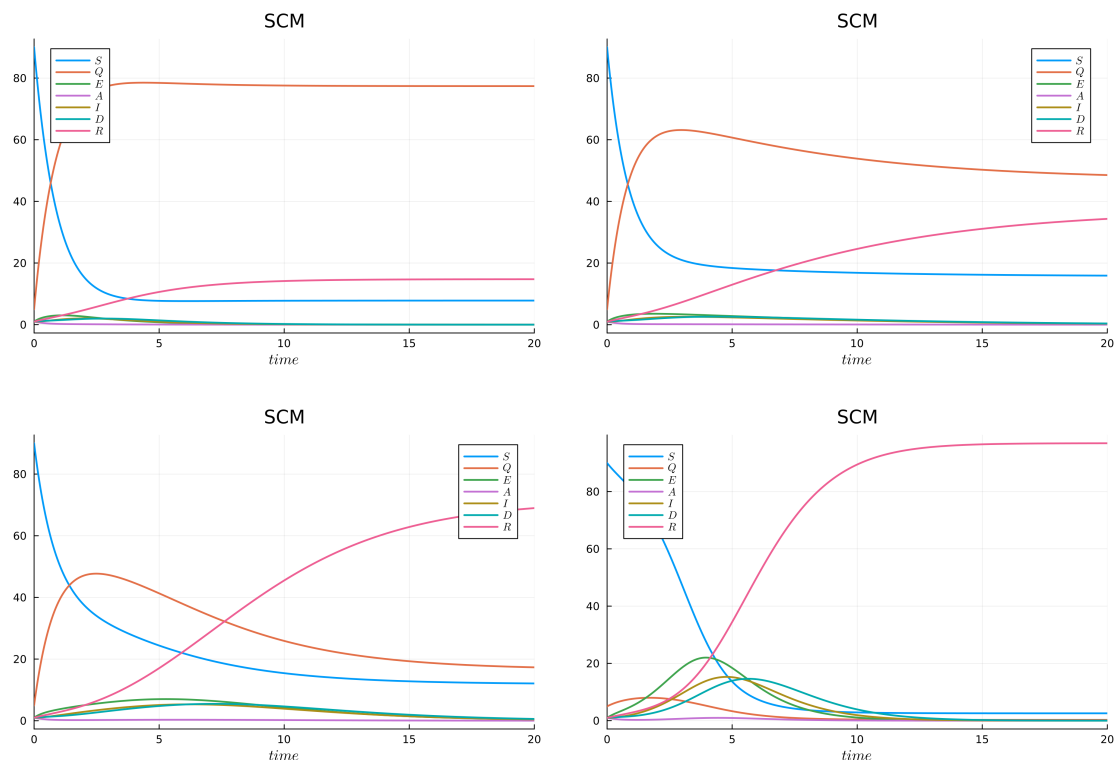


图 3: SCM 在不同参数下的表现

爆发的决心，而效果也是显著的：被感染人群 D，对应图中的绿色线条始终维持在极低的水平。这对于疾病早期医疗资源调动困难的情况是十分有利的，但是相对的代价是大部分人口被限制出行，不利于经济发展。

- 如果政府采取较一般的管控措施，其结果如右上角的图形所示。早期，被限制出行的人群 Q 迅速增加，此后逐步减少，而被感染人群 D 始终维持在一定水平，于是 Q 向 D 再向恢复人群 R 开始逐步转化，总体上会趋于慢慢全部感染。这对应了疫情中期，上海等医疗资源相对充足的城市采用的小区隔离措施，在医疗资源周转顺利的情况下，逐步减少被感染人群。
- 如果政府采取宽松的管制措施，或者说管制手段无法实现真正的管控，其结果就是短期控制效果明显，但是实际上的被感染的群体会迅速增加。对应一些生产力落后的地区尝试采用强力管控但是没有成效，于是短期管控之后仍然造成了大规模扩散。
- 如果政府不额外才去管制措施，可以注意到，疾病的发展几乎和 SI 模型一致。虽然此种措施较早实现了群体免疫，但是发展前中期迅速增加的感染人口极其不利于社会稳定。

综上所述，根据 SCM 模型，当公共卫生事件爆发时，政府可以充分考量自身的医

疗资源的承载能力和管制措施的有效性，面对最终一定要实现群体免疫的结局，如何抉择以让过程对人民的代价最小，是十分重要的命题。

6 模型检验与评价

6.1 SIRD

根据上述模型进行预测，预测 2020/06/24 到 2020/06/28 的 \widehat{I}_x 、 \widehat{R}_x 、 \widehat{ID}_x

表 4: 预测结果/实际数据

	\widehat{I}_x	I_x	\widehat{R}_x	R_x	\widehat{ID}_x	ID_x
2020/06/24	10.7561	26	110.6351	670	0.8704	7
2020/06/25	10.7684	26	111.4148	672	0.8766	7
2020/06/26	10.7808	25	112.1954	674	0.8827	7
2020/06/27	10.7932	25	112.9770	674	0.8888	7
2020/06/28	10.8057	26	113.7594	675	0.8950	7

检验发现拟合的结果有一定的偏差，原因可能包含以下几种：

1. 模型过于简化，不能很好的模拟实际情况。
2. 相对于巨大的人口基数，感染者、康复者和因病死亡数量过少，导致参数过小，容易产生较大的误差。
3. 数据量较小，可能产生误差。

6.2 SCM

6.2.1 优点

可以注意到，SCM 模型有效的考量了政府政策对人群感染的影响。SCM 假设了一个被隔离人群 Q，该人群与被感染者之间没有交互。另外，模型有效的考量了无症状感染者自身抵抗力较强的因素，具体反映在处理与无症状感染者之间的感染情况时，额外乘上了一个参数 θ 。

SCM 模型具有很强的可扩展性，具体体现在，根据政府政策的不同，可以主动调节有关参数。比如，早期病毒具有较强致死率的时候，各国政府多管控、限制出行、实施隔离措施，此时，易感染人群 S 向被隔离人群 Q 的转化率较高，在这种情况下，主要人群处于管控中，最后趋于稳定。

6.2.2 缺点和评价

- SCM 模型参数过多，可能在实际运用的时候需要更多的精细的调节才能获得较好的预测效果。
- SCM 检验的参数是我们根据不同政策进行手动调参的，可能不具有很好的代表性。此外，SCM 的部分参数在实际的数据中很难被学习和计算到，可能需要额外的统计学手段或者借助一些算法才能得到相关参数。
- 注意到，SCM 模型并没有考虑何时投入相应的医疗资源和政策变动的可能。因此可以再进一步的研究中尝试补足此部分。
- SCM 模型的假设过强，但是比起真实世界不足，尤其没有去重点被隔离人群和其他部分的交互，也考虑死亡人群产生的影响。

参考文献

无

7 附录：代码

SIRD.m

```
1  clear all
2
3  N = 1000;
4  i0 = 5;
5  s0 = N - i0;
6  r0 = 0;
7  d0 = 0;
8
9  T = 100;
10
11  beta = 0.4; %感染率
12  gamma = 0.1; %康复率
13  delta = 0.01; %死亡率
14
15  init = [s0 i0 r0 d0];
16  Tspan = 0 : 0.1 : T;
17
18  [T,Y] = ode45(@(t,Y) sird(t,Y,beta,gamma,delta),Tspan,init);
19
20  S = Y(:,1);
21  I = Y(:,2);
22  R = Y(:,3);
23  D = Y(:,4);
24
25  plot(T,S,'k');
26  hold on;
27  grid on;
28  plot(T,I,'m:');
29  plot(T,R,'r--');
30  plot(T,D,'b-.');
31  title(['SIRD model: \beta = ',num2str(beta),', \gamma = ',num2str(gamma),
32        ', \delta = ',num2str(delta)])
33  xlabel('Time')
34  ylabel('Number of Individuals')
35  legend('S','I','R','D','Location','best')
36
37  function dY= sird(t,Y,beta,gamma,delta)
38
39  S = Y(1);
40  I = Y(2);
41  R = Y(3);
42  D = Y(4);
43
44  N1 = S + I;
45
46  dY = zeros(4,1);
47  dY(1) = -beta*S*I/N1 ;
48  dY(2) = beta*S*I/N1 - gamma*I - delta*I;
49  dY(3) = gamma*I;
50  dY(4)= delta*I;
51
52  end
```

SCM.jl

```
1 using DifferentialEquations, ParameterizedFunctions, Plots, Parameters,  
   LaTeXStrings  
2  
3 SCM = @ode_def begin  
4     dS = - beta * S * (I + theta * A) - p * S + lambda * Q  
5     dQ = p * S - lambda * Q  
6     dE = beta * S * (I + theta * A) - rho * E  
7     dA = sigma * (1 - rho) * E - epsilon_A * A - gamma_A * A  
8     dI = sigma * rho * E - gamma_I * I - d_I * I - epsilon_I * I  
9     dD = epsilon_A * A + epsilon_I * I - (d_D + gamma_D) * D  
10    dR = gamma_A * A + gamma_I * I + gamma_D * D  
11 end sigma rho theta beta p lambda gamma_A gamma_I gamma_D epsilon_I  
   epsilon_A d_D d_I  
12  
13 u0 = [90, 5, 1, 1, 1, 1, 1]  
14 tspan = [0, 20]  
15 parameters = [  
16     [.9, .9, .5, 0.05, .99, .1, .99, .05, .99, .99, .95, 0.001, 0.002],  
17     [.9, .9, .5, 0.05, .9, .3, .99, .05, .99, .99, .95, 0.001, 0.002],  
18     [.9, .9, .5, 0.05, .7, .9, .99, .05, .99, .99, .95, 0.001, 0.002],  
19     [.9, .9, .5, 0.05, .1, .9, .99, .05, .99, .99, .95, 0.001, 0.002],  
20     [.9, .9, .5, 0.05, 0, 1, .99, .05, .99, .99, .95, 0.001, 0.002],  
21 ]  
22  
23 for i in range(1, length(parameters))  
24     prob = ODEProblem(SCM, u0, tspan, parameters[i])  
25     solution = solve(prob, Tsit5())  
26     plot(solution,  
27         labels = [L"S" L"Q" L"E" L"A" L"I" L"D" L"R"],  
28         xlabel = L"time",  
29         title = "SCM",  
30         linewidth = 1,  
31         dpi = 300  
32     )  
33  
34     savefig("SCM"*string(i)*".png")  
35 end
```