Lecture 11: Modeling in Epidemiology

Radu Balan

Department of Mathematics, NWC
University of Maryland, College Park, MD

Version: March 3, 2024

Models in Epidemiology

So far we analyzed three deterministic models, and their agent-based simulation implementations:
We focus on three deterministic models:

1. SI (Susceptible-Infected) Model
2. SIR (Susceptible-Infected-Removed) Model
3. SEIR (Susceptile-Exposed-Infectious-Removed) Model

We discussed the models, their behavior, and how to calibrate them.

The SI Model

The nonlinear form

The SI model assumes that transmissions occur only when susceptible people get in contact with infected individuals. Such a process is modeled by a rate proportional to the product SI and not just S:

$$\left\{ \begin{array}{lllll} \frac{dS}{dt} & = & -\beta S \frac{I}{N} & , & S(0) = S_0 \\ \frac{dI}{dt} & = & \beta S \frac{I}{N} & , & I(0) = I_0 \end{array} \right. \qquad \text{(SI Model)}$$

where $N = S(0) + I(0)$ is total population (constant over time) and $\beta \geq 0$ is the transmission parameter.
In normalized form, with $s(t) = \frac{S(t)}{N}$ and $i(t) = \frac{I(t)}{N}$:

$$\left\{ \begin{array}{lllll} \frac{ds}{dt} & = & -\beta si & , & s(0) = \frac{S_0}{N} \\ \frac{di}{dt} & = & \beta si & , & i(0) = \frac{I_0}{N} \end{array} \right.$$

The SI Model

Closed form solution and typical behavior

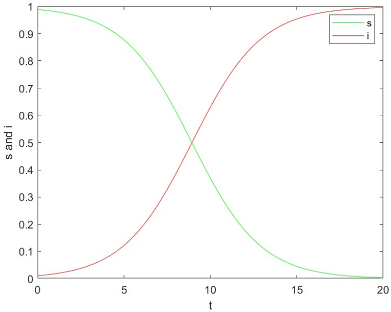
Total population is constant, $S(t) + I(t) = N$ or $s(t) + i(t) = 1$, and

$$I(t) = \frac{NI(0)}{I(0) + (N - I(0))e^{-\beta t}} \quad , \quad S(t) = N - I(t)$$

Simulation of the non-linear SI model:

$$\beta = 0.5 \; , \; S(0) = 2000 \; , \; I(0) = 23$$

Note: The SI model is more suitable for modeling *cumulative* number of infections, not the daily rates.



The SIR Model

Assume a system with three compartments: 'Susceptible' (S), 'Infected' (I) and 'Removed' or 'Recovered' (R). At time $t_0 = 0$ the system has a total of N individuals (initial total population). Most of them are susceptible $S(0)$, but some are infected, $I(0)$ and possibly some are in the recovered state, $R(0)$. Our intention is to model the time evolution of these populations. The previous SI model is modified into:

$$\left\{ \begin{array}{lcl} \frac{dS}{dt} & = & -\beta S \frac{I}{N} \quad , \quad S(0) \\ \frac{dI}{dt} & = & \beta S \frac{I}{N} - \alpha I \quad , \quad I(0) \\ \frac{dR}{dt} & = & \alpha I \quad , \quad R(0) \end{array} \right.$$

where $\alpha \geq 0$ and $\beta \geq 0$ are parameters. Interpretation: $\frac{1}{\alpha} = \mathbb{E}[\text{Infection Time}]$ is the average time of infection. Furthermore, we model the cumulative deaths $Y(t)$ by $Y(t) = \gamma R(t)$, for some $\gamma \geq 0$, the probability of death conditioned on getting infected.

The SIR Model

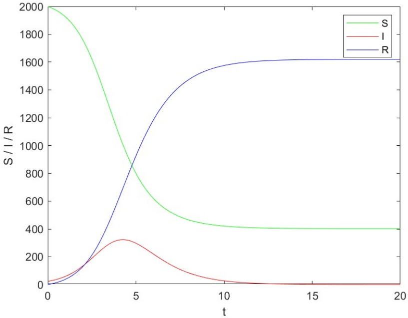
Deterministic simulations

Simulation of the SIR model:

$$\beta = 2 \ , \ \alpha = 1 \ , \ S(0) = 2000 \ , \ I(0) = 23 \ , \ R(0) = 0$$

Results were obtained with an Euler scheme with step size $h = 0.01$.

Note: The infected population $I(t)$ first increases and then decreases eventually to 0. The susceptible population decreases, but converges to some limiting value $S(\infty) > 0$. The removed population is monotone increasing and converges to some value $R(\infty) < N$. Some of the susceptible population who do not get infected are protected by the recovered population surrounding them. This is known as *herd immunity*.



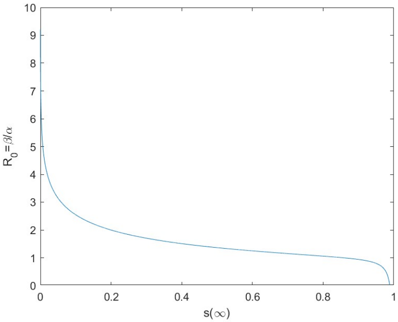
The SIR Model

Herd immunity

Herd immunity occurs when $S(\infty) \gg 0$. The *contact number* (or, the *basic reproduction ratio*) $R_0 = \frac{\beta}{\alpha}$ is related to $s(\infty)$ through:

$$R_0 = \frac{log(s(\infty)) - log(s(0))}{s(\infty) - 1}.$$

For $s(0) = \frac{S(0)}{N} \approx 1$ the equation simplifies to $R_0 = \frac{log(s(\infty))}{s(\infty)-1}$. For $R_0 < 1$, it follows that a significant number of susceptible individual do not get infected.



SIR Model Calibration

Pre-processing steps

Input data: the time series of *cumulative detected infections*, $\{V(0), \cdots, V(T_{max})\}$, and the time series of *cumulative deaths*, $\{Y(0), \cdots, Y(T_{max})\}$.

Step 1: t_0. First detect the onset of infections, and reset the time origin to match this starting time t_0.

Step 2: $I(t)$. The time series of cumulative detected infections $\{V(0), \cdots, V(T_{max})\}$ should be turned into a rate of infections. The daily rate may not be relevant to infection transmissions. Instead use a time window to convert the cumulative count into a rate:

$$I(t) = V(t + t_0 + \tau_0) - V(t + t_0 - \tau_0) \quad , \quad t = 0, 1, 2, \cdots, T_{max}$$

where $\tau_0 > 0$ is chosen so that τ_0 accounts for average infection period. For Covid-19 it is somewhere between 5 days and 10 days. A possible value is $\tau_0 = 7$. Thus $I(t)$ measures the number of invections in a 2-week period centered around t.

Step 3: $Y_{measured}$. Align the time series of cumulative death with t_0:

$$Y_{measured}(t) = Y(t + t_0) \quad , \quad t = 0, 1, 2, \cdots, T_{max}$$

Optimization based on Forward Model Simulation

The objective function

The objective function J_p is chosen to measure residuals $I - \rho I_{sim}$ and $Y_{measured} - Y_{sim}$. Fix cost coefficients c_I and c_Y. For $1 \leq p < \infty$ define

$$J_p(\alpha, \beta, \gamma, \rho) = c_I \sum_{t=0}^{T_{max}} |I(t) - \rho I_{sim}(t)|^p + c_Y \sum_{t=0}^{T_{max}} |Y_{measured}(t) - \gamma R_{sim}(t)|^p$$

For $p = \infty$ define

$$J_\infty(\alpha, \beta, \gamma, \rho) = c_I \max_{0 \leq t \leq T_{max}} |I(t) - \rho I_{sim}(t)| + c_Y \max_{0 \leq t \leq T_{max}} |Y_{measured}(t) - \gamma R_{sim}(t)|$$

Typical choices: $p = 1, 2, \infty$ and $(c_I, c_Y) = (0, 1)$ (if the cumulative detected infections are unreliable) or $(c_I, c_Y) = (1, 1)$ (if detected infections is a reliable measure).

The Calibration Algorithm for SIR Models

Algorithm (Meta-Algorithm for SIR Calibration)

Inputs: *Time series* $\{V(0), \cdots, V(T)\}$, $\{Y(0), \cdots, Y(T)\}$. *Parameters:* V_{min} *(default,* $V_{min} = 5$*),* τ_0 *(default,* $\tau_0 = 7$*),* $N = Population$, $p \in [1, \infty]$ *(default,* $p = 2$*),* $c_I, c_Y > 0$. *Search set* Ω.

1. *Detect the onset of the infection* t_0 *as the first time so that* $V(t_0) \geq V_{min}$, *reset the time origin, and create the time series of infection rates* $I(t) = V(t + t_0 + \tau_0) - V(t + t_0 - \tau_0)$, $0 \leq t \leq T_{max}$, *and aligned cumulative death* $Y_{measured}(t) = Y(t + t_0)$, $0 \leq t \leq T_{max}$.
2. *For each* $(\alpha, \beta) \in \Omega$ *repeat:*
   1. *Simulate a SIR model with parameters* (α, β) *and initial condition* $S_{sim}(0) = N$, $I_{sim}(0) = I(0)$, $R_{sim}(0) = 0$, *and obtain* *daily* *time series* $(S_{sim}, I_{sim}, R_{sim})$.
   2. *Solve* $\hat{\gamma} = argmin_{\gamma} \| Y_{measured} - \gamma R_{sim} \|_p$.
   3. *Solve* $\hat{\rho} = argmin_{\rho} \| I - \rho I_{sim} \|_p$.
   4. *Compute the objective function* $J = J_p(\alpha, \beta, \hat{\gamma}, \hat{\rho})$.
3. *Determine the minimum and the minimizer of J.*

Outputs: *Estimated* $\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\rho}$ *and minimum value* J_{min}.

Numerical results (1)

Analysis of DC data for 2020

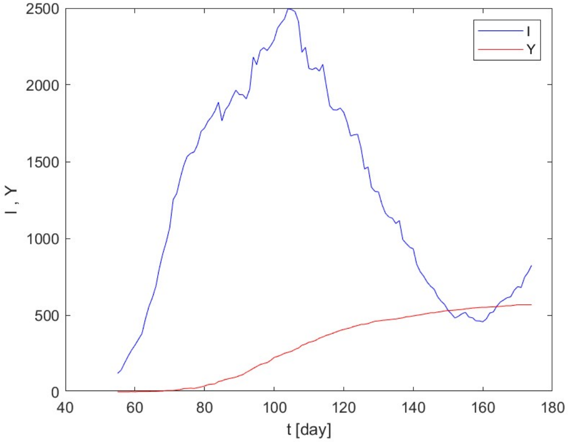
$V_{min} = 5$, $\tau_0 = 7$ [days]. Onset $t_0 = 55$ (Monday: March 16, 2020), $T_{max} = 119$, $N = 705749$.

Notable dates:

Local dips: day=152 (Sunday: 6/21/2020), day=159 (Sunday: 6/28/2020).

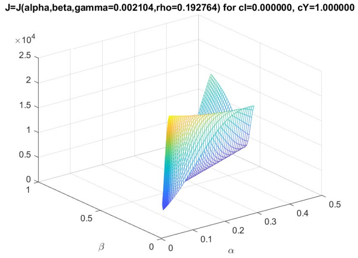
Memorial weekend: Sunday May 24, 2020: day = 124 .

4th of July: Saturday July 4, 2020: day = 165.



Numerical results (2)

Analysis of DC data for 2020. $p = 1$, $c_I = 0$, $c_Y = 1$



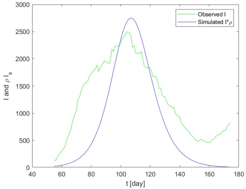
Results:
$\hat{\alpha} = 0.5$, $\hat{\beta} = 0.62$,

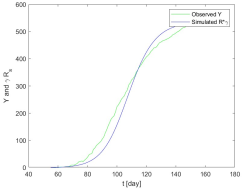
$\hat{R}_0 = 1.24$,

$\hat{\gamma} = 0.21\%$.

$\hat{\rho} = 19\%$

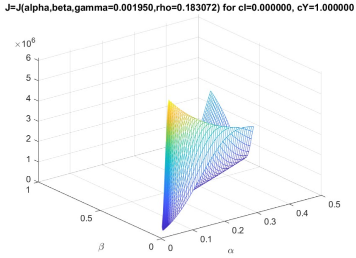
$J_{optim} = 2690$.





Numerical results (3)

Analysis of DC data for 2020. $p = 2,\ c_I = 0,\ c_Y = 1$



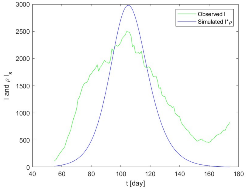
Results:
$\hat{\alpha} = 0.49,\ \hat{\beta} = 0.617$

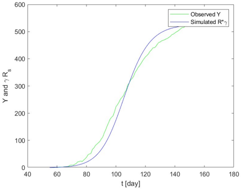
$\hat{R}_0 = 1.26,$

$\hat{\gamma} = 0.195\%.$

$\hat{\rho} = 18\%$

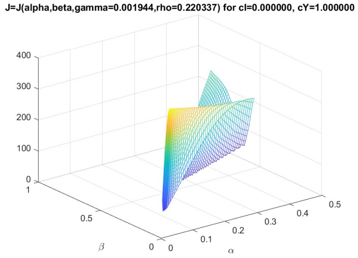
$J_{optim} = 89365.$





Numerical results (4)

Analysis of DC data for 2020. $p = \infty$, $c_I = 0$, $c_Y = 1$



Results:

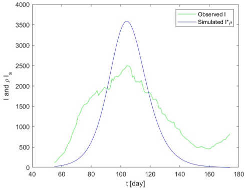
$\hat{\alpha} = 0.5$, $\hat{\beta} = 0.63$,

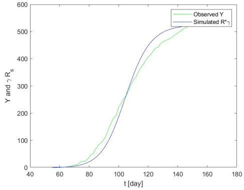
$\hat{R}_0 = 1.26$,

$\hat{\gamma} = 0.19\%$.

$\hat{\rho} = 22\%$

$J_{optim} = 44.98$.





Numerical results (5)

Analysis of DC data for 2020. $p = 1$, $c_I = 1$, $c_Y = 1$

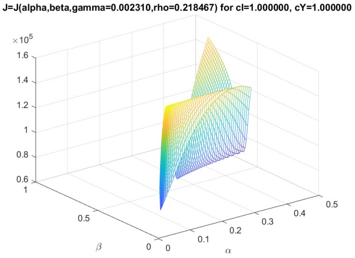
Results:
$\hat{\alpha} = 0.5$, $\hat{\beta} = 0.61$,

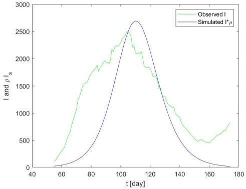
$\hat{R}_0 = 1.22$,

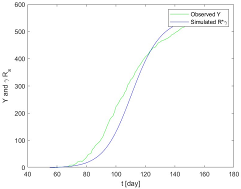
$\hat{\gamma} = 0.23\%$.

$\hat{\rho} = 22\%$

$J_{optim} = 67250$.







Numerical results (6)

Analysis of DC data for 2020. $p = 2$, $c_I = 1$, $c_Y = 1$

Results:

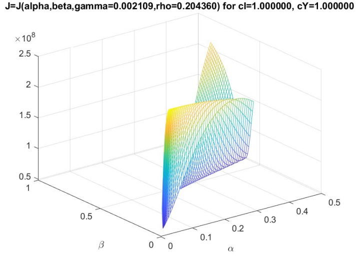
$\hat{\alpha} = 0.5$, $\hat{\beta} = 0.62$

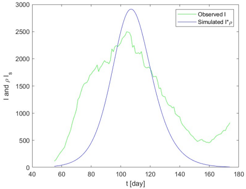
$\hat{R}_0 = 1.24$,

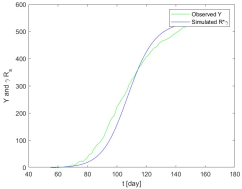
$\hat{\gamma} = 0.21\%$.

$\hat{\rho} = 20\%$

$J_{optim} = 48340173$.

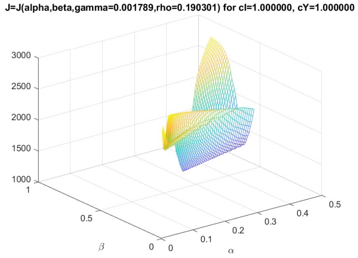






Numerical results (7)

Analysis of DC data for 2020. $p = \infty$, $c_I = 1$, $c_Y = 1$



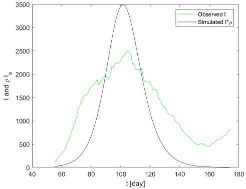
Results:
$\hat{\alpha} = 0.5$, $\hat{\beta} = 0.64$,

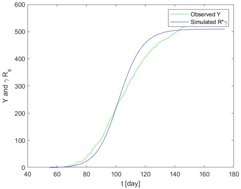
$\hat{R}_0 = 1.28$,

$\hat{\gamma} = 0.18\%$.

$\hat{\rho} = 19\%$

$J_{optim} = 1232$.





The Susceptible-Exposed-Infectious-Removed (SEIR) Model

The Susceptible-Exposed-Infectious-Removed (SEIR) model is obtained from SIR by introducing a compartment between Susceptible and Infected of population that has been exposed to the virus but are not yet contagious:

$$\left\{\begin{array}{rcl} \frac{dS}{dt} & = & -\beta \frac{SI}{N} \quad , \quad S(0) \\ \frac{dE}{dt} & = & \beta \frac{SI}{N} - \delta E \quad , \quad E(0) \\ \frac{dI}{dt} & = & \delta E - \alpha I \quad , \quad I(0) \\ \frac{dR}{dt} & = & \alpha I \quad , \quad R(0) \end{array}\right. \qquad \text{(SEIR Model)}$$

where $\delta \geq 0$ is the rate of transition from exposed to infected. Its reciprocal $1/\delta$ represents the *average incubation period*.
If data is selected from the onset of infections, a natural initial condition is: $R(0) = N \gg E(0), I(0)$, $R(0) = 0$. The initial exposed population $E(0)$ may be set to $I(0)$, or can be fine tuned to fit the data. Assuming $E(0) = I(0)$ is known, the parameters that need to be calibrated are: $\alpha, \beta, \gamma, \delta, \rho$.

Simulations

Here are a few numerical results with the Euler's scheme:

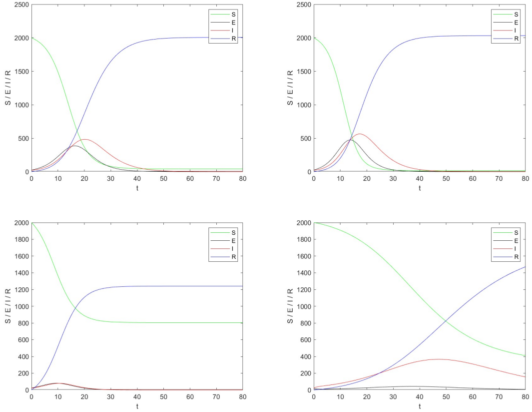


Figure: Top left: $\alpha = 0.2$, $\beta = 0.8$, $\delta = 0.3$; Top right: $\alpha = 0.2$, $\beta = 1.0$, $\delta = 0.3$; Bottom left: $\alpha = 1.0$, $\beta = 1.5$, $\delta = 1.0$; Bottom right: $\alpha = 0.08$, $\beta = 0.176$, $\delta = 0.8$

SEIR Model Calibration

Pre-processing steps

Input data: the time series of *cumulative detected infections*, $\{V(0), \cdots, V(T_{max})\}$, and the time series of *cumulative deaths*, $\{Y(0), \cdots, Y(T_{max})\}$.

Step 1: t_0. First detect the onset of infections, and reset the time origin to match this starting time t_0.

Step 2: $I(t)$. The time series of cumulative detected infections $\{V(0), \cdots, V(T_{max})\}$ should be turned into a rate of infections. The daily rate may not be relevant to infection transmissions. Instead use a time window to convert the cumulative count into a rate:

$$I(t) = V(t + t_0 + \tau_0) - V(t + t_0 - \tau_0) \ , \quad t = 0, 1, 2, \cdots, T_{max}$$

where $\tau_0 > 0$ is chosen so that τ_0 accounts for average infection period. For Covid-19 it is somewhere between 5 days and 10 days. A possible value is $\tau_0 = 7$. Thus $I(t)$ measures the number of invections in a 2-week period centered around t.

Step 3: $Y_{measured}$. Align the time series of cumulative death with t_0:

$$Y_{measured}(t) = Y(t + t_0) \ , \quad t = 0, 1, 2, \cdots, T_{max}$$

Optimization based on Forward Model Simulation

The Meta Loop

A natural choice for initial conditions is given by: $S(0) = N$ and $R(0) = 0$, where we assumed $E(0) + I(0) \ll N$. Another choise is $E(0) = I(0)$ since we do not know the undetected number of infections at time 0 (perhaps $E(0) > I(0)$ is closer to the truth):

$$\left\{ \begin{array}{rcl} \frac{dS}{dt} & = & -\beta S \frac{I}{N} \quad , \quad S(0) = N \\ \frac{dE}{dt} & = & \beta \frac{SI}{N} - \delta E \quad , \quad E(0) \\ \frac{dI}{dt} & = & \delta E - \alpha I \quad , \quad I(0) \\ \frac{dR}{dt} & = & \alpha I \quad , \quad R(0) = 0 \quad , \quad Y = \gamma R \end{array} \right.$$

At this point, the parameters that need to be estimated are: $\{\alpha, \beta, \delta, \gamma, \rho\}$.
The Forward Model based calibration works like this:

1. Construct a search set Ω of the "free" parameters (α, β, δ)
2. For each $(\alpha, \beta, \delta) \in \Omega$: (i) run a SIR simulator using the Euler scheme that produces $(S_{sim}, E_{sim}, I_{sim}, R_{sim})$. (ii) Fit $\hat{\gamma}$ to match the observed time series Y.(iii) Fit $\hat{\rho}$ that accounts the undercounting of actual infections in the observed time series I. (iv) Compute the value of the objective function $J_p(\alpha, \beta, \delta, \hat{\gamma}, \hat{\rho})$.
3. Select $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\gamma}, \hat{\rho})$ that minimize J.

Optimization based on Forward Model Simulation

The objective function

The objective function J_p can be chosen to measure a norm of residuals $I - \rho I_{sim}$ and $Y_{measured} - \gamma R_{sim}$.
To do so, you need to choose and fix meta-parameters p, c_I and c_Y. For $1 \leq p < \infty$ define

$$J_p(\alpha,\beta,\delta,\gamma,\rho) = c_I \sum_{t=0}^{T_{max}} |I(t) - \rho I_{sim}(t)|^p + c_Y \sum_{t=0}^{T_{max}} |Y_{measured}(t) - \gamma R_{sim}(t)|^p$$

For $p = \infty$ define

$$J_\infty(\alpha,\beta,\delta,\gamma,\rho) = c_I \max_{0 \leq t \leq T_{max}} |I(t) - \rho I_{sim}(t)| + c_Y \max_{0 \leq t \leq T_{max}} |Y_{measured}(t) - \gamma R_{s}$$

Typical choices: $p = 1, 2, \infty$ and $(c_I, c_Y) = (0, 1)$ (if the cumulative detected infections are unreliable) or $(c_I, c_Y) = (1, 1)$ (if detected infections is a reliable measure).

The Calibration Algorithm for SEIR Models

Algorithm (Meta-Algorithm for SEIR Calibration)

Inputs: Time series $\{V(0), \cdots, V(T)\}$, $\{Y(0), \cdots, Y(T)\}$. Parameters: V_{min} (default, $V_{min} = 5$), τ_0 (default, $\tau_0 = 7$), N = Population, $E(0)$ (default $E(0) = I(0)$), $p \in [1, \infty]$ (default, $p = 2$), $c_I, c_Y > 0$. Search set Ω.

1. Detect the onset of the infection t_0 as the first time so that $V(t_0) \geq V_{min}$, reset the time origin, and create the time series of infection rates $I(t) = V(t + t_0 + \tau_0) - V(t + t_0 - \tau_0)$, and aligned cumulative death $Y_{measured}(t) = Y(t + t_0)$, $0 \leq t \leq T_{max}$.
2. For each $(\alpha, \beta, \delta) \in \Omega$ repeat:
   1. Simulate a SEIR model with parameters (α, β, δ) and initial condition $S_{sim}(0) = N$, $E_{sim}(0) = E(0)$, $I_{sim}(0) = I(0)$, $R_{sim}(0) = 0$, and obtain daily time series $(S_{sim}, E_{sim}, I_{sim}, R_{sim})$.
   2. Solve $\hat{\gamma} = argmin_{\gamma} \|Y_{measured} - \gamma R_{sim}\|_p$.
   3. Solve $\hat{\rho} = argmin_{\rho} \|I - \rho I_{sim}\|_p$.
   4. Compute the objective function $J = J_p(\alpha, \beta, \delta, \hat{\gamma}, \hat{\rho})$.
3. Determine the minimum and the minimizer of J.

Outputs: Estimated $\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\gamma}, \hat{\rho}$ and minimum value of the objective function J_{min}.

Numerical results (1)

Analysis of DC data for 2020

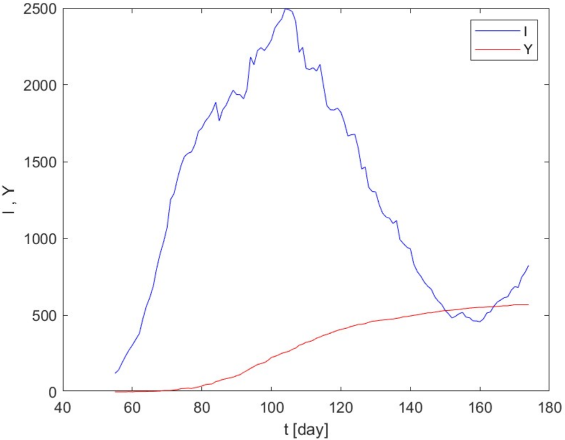
$V_{min} = 5$, $\tau_0 = 7$ [days]. Onset $t_0 = 55$ (Monday: March 16, 2020), $T_{max} = 119$, $N = 705749$.

Notable dates:

Local dips: day=152 (Sunday: 6/21/2020), day=159 (Sunday: 6/28/2020).

Memorial weekend: Sunday May 24, 2020: day = 124 .

4th of July: Saturday July 4, 2020: day = 165.



Numerical results (2)

Analysis of DC data for 2020. $p = 1$, $c_I = 0$, $c_Y = 1$

Results:

$\hat{\alpha} = 0.5$, $\delta = 0.8$,

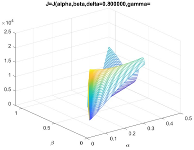
$\hat{\beta} = 0.73$,

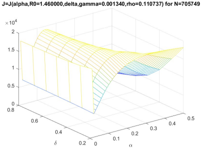
$\hat{R}_0 = 1.46$,

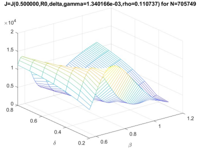
$\hat{\gamma} = 0.134\%$.

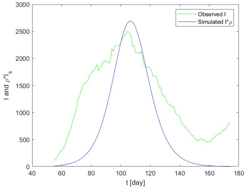
$\hat{\rho} = 11\%$

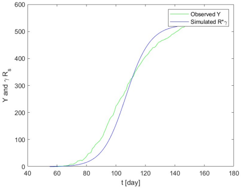
$J_{optim} = 3103$.











Numerical results (3)

Analysis of DC data for 2020. $p = 2$, $c_I = 0$, $c_Y = 1$

Results:

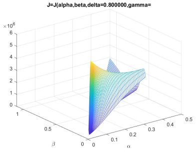
$\hat{\alpha} = 0.5$, $\delta = 0.8$,

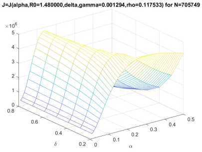
$\hat{\beta} = 0.74$, $\hat{R}_0 = 1.48$,

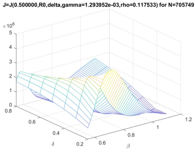
$\hat{\gamma} = 0.13\%$.

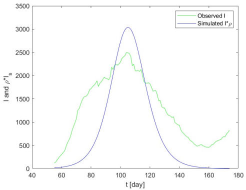
$\hat{\rho} = 11.75\%$.

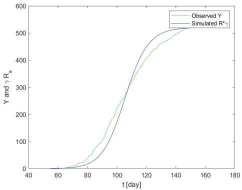
$J_{optim} = 117406$.





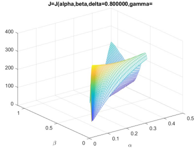


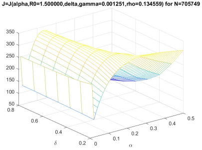


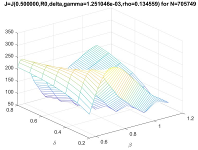


Numerical results (4)

Analysis of DC data for 2020. $p=\infty$, $c_I=0$, $c_Y=1$







Results:

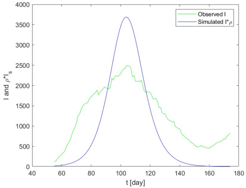
$\hat{\alpha}=0.5$, $\delta=0.8$,

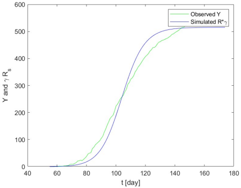
$\hat{\beta}=0.75$, $\hat{R}_0=1.5$,

$\hat{\gamma}=0.125\%$.

$\hat{\rho}=13.45\%$.

$J_{optim}=52.7$.





Numerical results (5)

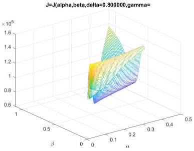
Analysis of DC data for 2020. $p = 1$, $c_I = 1$, $c_Y = 1$

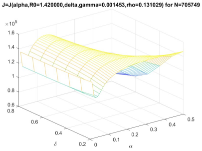
Results:
$\hat{\alpha} = 0.5$, $\delta = 0.8$,

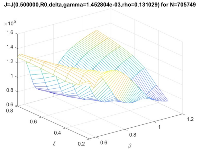
$\hat{\beta} = 0.71$,
$\hat{R}_0 = 1.42$,

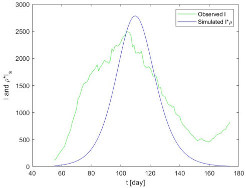
$\hat{\gamma} = 0.146\%$.

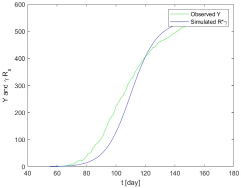
$\hat{\rho} = 13.1\%$
$J_{optim} = 74151$.











Numerical results (6)

Analysis of DC data for 2020. $p = 2,\ c_I = 1,\ c_Y = 1$

Results:

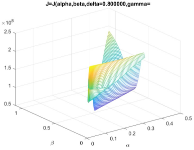
$\hat{\alpha} = 0.5,\ \delta = 0.8,$

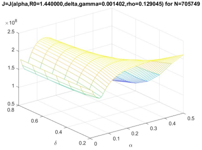
$\hat{\beta} = 0.72,$
$\hat{R}_0 = 1.44,$

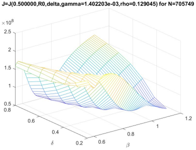
$\hat{\gamma} = 0.14\%.$

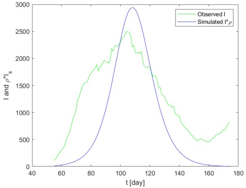
$\hat{\rho} = 12.9\%.$

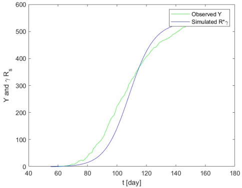
$J_{optim} = 57030309.$











Numerical results (7)

Analysis of DC data for 2020. $p = \infty$, $c_I = 1$, $c_Y = 1$

Results:

$\hat{\alpha} = 0.5$, $\delta = 0.8$,

$\hat{\beta} = 0.75$, $\hat{R}_0 = 1.5$,

$\hat{\gamma} = 0.125\%$.

$\hat{\rho} = 13.45\%$.

$J_{optim} = 1308$.