Use of simulated annealing to determine the operational parameters of the SEIR model for the coronavirus for various jurisdictions

Leslie Mulder  lesm@velocity.ca
(Submitted: 16 April 2020 – Published online: 20 April 2020)

Abstract
In this paper simulated annealing is used to derive and optimize the unknown operational parameters of the SEIR model for the current coronavirus pandemic relative to the recorded data for the world and other selected countries.

Introduction
A simple SEIR [1][2] epidemic model is considered for the simulation of the dynamics of the current coronavirus. In this model individuals are assigned to one of the following disease states: Susceptible (S), Exposed (E), Infectious (I) or Recovered (R). The aggregate numbers of individuals in these states are related in the model through a set of coupled first order differential equations.

The SEIR model
The model, with a constant population is:
\[
\frac{dS}{dt} = -\beta SI
\]
\[
\frac{dE}{dt} = \beta SI - \sigma E
\]
\[
\frac{dI}{dt} = \sigma E - \gamma I
\]
\[
\frac{dR}{dt} = \gamma I
\]

where:
\(\beta\): The parameter controlling how often a susceptible-infected contact results in a new exposure;
\(\gamma\): The rate an infected recovers and moves into the resistant phase;
\( \sigma \): The rate at which an exposed person becomes infective.

Of note is that the model is conservative in that if \( N \) represents the initial susceptible population then

a) \( N = S + E + I + R \), and;

b) \( dN = dS + dE + dI + dR = 0 \).

An alternate form, in parameter names only, is also presented in the literature [3]:

\[
\begin{align*}
\frac{dS}{dt} &= -\frac{R}{NT_{inf}} SI \\
\frac{dE}{dt} &= \frac{R}{NT_{inf}} SI - \frac{E}{T_{inc}} \\
\frac{dI}{dt} &= \frac{E}{T_{inc}} - \frac{I}{T_{inf}} \\
\frac{dR}{dt} &= \frac{I}{T_{inf}}
\end{align*}
\]

where

\( R \): the basic reproduction rate of the virus i.e. the number of secondary infections each infected individual creates, with;

Transmission Times:

\( T_{inc} \): the length of incubation period, and;

\( T_{inf} \): the duration for which the patient is infectious.

With respect to the currently available data sets, what appears to be known with some degree of accuracy are the number of Infectious cases, which for the purpose of this paper are taken to be the number of reported active cases, and the number of those cases that have been closed, either through being cured or the patient dying. In the SEIR model there is no distinction between the patient being cured or dying, both are viewed as recovered.

In the current literature the term **active** is alternately used for **infectious** and the term **closed** is an alternate for **recovered** (cured or dead).

Given the current data for active (infectious), cured and dead cases from worldometers.com [5], historical data over the, almost, full history of the corona virus pandemic is available. The data provides for active (infectious) cases and closed (recovered) cases, which is the simple sum of cured and dead cases, on a daily basis.

By evolving the SEIR model using a 4\textsuperscript{th} order Runge-Kutta predictor corrector scheme [4] a comparison was made between the resulting evolution of the SEIR curves over time and the real world empirical
data. An error measure for the comparison was the square root of the sum of the squares of the root mean square differences of the data points for active case versus the models active (infectious) data and the root mean square of the data points for the closed cases versus the models recovered data:

\[
\epsilon = \sqrt{\frac{1}{N} \left( \sum_{k=1}^{N} (A_k - a_k)^2 + \sum_{k=1}^{N} (C_k - c_k)^2 \right)}
\]

where:

\(\epsilon\): is the error measure;

\(A_k\) and \(a_k\): are the empirical data point and the equivalent model data point for time step \(k\) (day \(k\)) for the active (infectious) cases;

\(C_k\) and \(c_k\): are the empirical data point and the equivalent model data point for time step \(k\) (day \(k\)) for the closed (recovered) cases;

\(N\): is the number of time steps (days) in the empirical data set.

Using a simulated annealing algorithm the parameter space of \(S_0\), the initial size of the susceptible population, \(\beta\), \(\gamma\) and \(\sigma\) was explored stochastically accepting unconditionally those parameter sets that yielded a lower error than the previous error and accepting conditionally, with diminishing likelihood, parameter sets that yielded a higher error than the previous set.

In addition the \(\gamma\) parameter can be estimated from the current empirical data through the equation

\[
\frac{dR}{dt} = \gamma I
\]

and this estimation can be used to guide the initial conditions of the annealing.

Finding those parameters that allow the SEIR model to fit the data, allows for the extrapolation of the data set into the future. Also, and perhaps more importantly it allows for the estimation of the size of the of the population, the \(S_0\) parameter, that is going to affect to the level of being considered an active case and then by correlation the number of those cases that will be cured and the number that will die.

The simulated annealing searches were run for several regional data sets; the world, the United States of America, Canada and South Korea.
Results

For the world as of April 16, 2020, the best fit SEIR model is shown as:

Figure 1: Optimized SEIR model for the current corona virus evolution for the world

Associated operational parameters:

\[ S_0 = 7.045658e+06; \]
\[ \beta = 1.095963e-07; \]
\[ \sigma = 2.298148e-02; \]
\[ \gamma = 3.044254e-02; \]

These translate into the more familiar parameters:

\[ R = 25.823036; \]
\[ T_{inc} = 43.513298; \]
\[ T_{inf} = 32.848766; \]

As a measure of quality of the convergence of the annealing algorithm the final error measure was:

\[ \epsilon = 4.420360e+04; \text{ with a relative error } \frac{\epsilon}{S_0} = 6.16E-03 \text{ or } \sim 0.6\% \]
For the United States of America as of April 16, 2020, the best fit SEIR model is shown as:

\[
S_0 = 1.135199 \times 10^6; \\
\beta = 1.364605 \times 10^{-6}; \\
\sigma = 4.011066 \times 10^{-2}; \\
\gamma = 4.357323 \times 10^{-3};
\]

These translate into the more familiar parameters:

\[
R = 355.516072; \\
T_{inc} = 24.931026; \\
T_{inf} = 229.498733;
\]

As an overall measure of quality the final error measure was:

\[
\epsilon = 3.519156 \times 10^3; \text{ with a relative error } \epsilon/S_0 = 3.10E-03 \text{ or } \sim 0.3%
\]
For the Canada as of April 16, 2020, the best fit SEIR model is shown as:

![SEIR Model compared with Canada data](image)

Figure 3: Optimized SEIR model for the current corona virus evolution for Canada

Associated operational parameters:

\[
S_0 = 2.734354\times10^4; \\
\beta = 1.842333\times10^{-05}; \\
\sigma = 8.227016\times10^{-02}; \\
\gamma = 3.386249\times10^{-03}; \\
\]

These translate into the more familiar parameters:

\[
R = 148.766117; \\
T_{inc} = 12.155075; \\
T_{inf} = 295.311992; \\
\]

As an overall measure of quality the final error measure was:

\[
\epsilon = 3.394125\times10^2; \text{ with a relative error } \epsilon/S_0 = 1.24e-02 \text{ or } \sim1.2\%.
\]
For the South Korea as of April 16, 2020, the best fit SEIR model is shown as:

**Figure 4: Optimized SEIR model for the current corona virus evolution for South Korea**

Associated operational parameters:

\[ S_0 = 9.723638\times10^3; \]
\[ \beta = 5.921661\times10^{-5}; \]
\[ \sigma = 4.573711\times10^{-1}; \]
\[ \gamma = 2.785969\times10^{-2}; \]

These translate into the more familiar parameters:

\[ R = 20.667887; \]
\[ T_{inc} = 2.186408; \]
\[ T_{inf} = 35.894155; \]

As an overall measure of quality the final error measure was:

\[ \epsilon = 9.392602\times10^2; \]
with a relative error \( \epsilon/S_0 = 9.66\times10^{-2} \) or \(~9.7\%\).
## Tabulated Parametric Results

<table>
<thead>
<tr>
<th></th>
<th>World</th>
<th>USA</th>
<th>Canada</th>
<th>South Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_0$</td>
<td>7.17E+06</td>
<td>1.14E+06</td>
<td>2.73E+04</td>
<td>9.72E+03</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.10E-07</td>
<td>1.36E-06</td>
<td>1.84E-05</td>
<td>5.92E-05</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>2.30E-02</td>
<td>4.01E-02</td>
<td>8.23E-02</td>
<td>4.57E-01</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>3.04E-02</td>
<td>4.36E-03</td>
<td>3.39E-03</td>
<td>2.79E-02</td>
</tr>
<tr>
<td>$R$</td>
<td>25.82</td>
<td>355.52</td>
<td>148.77</td>
<td>20.67</td>
</tr>
<tr>
<td>$T_{inc}$</td>
<td>43.51</td>
<td>24.93</td>
<td>12.16</td>
<td>2.19</td>
</tr>
<tr>
<td>$T_{inf}$</td>
<td>32.85</td>
<td>229.50</td>
<td>295.31</td>
<td>35.89</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>4.42E+04</td>
<td>3.52E+03</td>
<td>3.39E+02</td>
<td>9.39E+02</td>
</tr>
<tr>
<td>rel. $\epsilon$</td>
<td>6.16E-03</td>
<td>3.10E-03</td>
<td>1.24E-02</td>
<td>9.66E-02</td>
</tr>
</tbody>
</table>
Conclusion

Although still preliminary, the use of simulated annealing to derive the size of the susceptible population for various jurisdictions and the other operational parameters for the SEIR model appears to be a useful tool to determine these values as they relate to a specific region or to the global population.

It is clear that the values determined for the reproduction rate, incubation period and the infectious period lie outside of the ranges normally expected. This is likely due to the way in which the SEIR model is used, that is by starting with data for the number of recorded infected patients and the number of recorded closed cases and opposed to addressing the full population. Running the model this way infers the susceptible pool from the infected pool and, as such, there is an expectation that R would be larger than otherwise and the incubation and infectious periods longer.

The correlation between the recorded data and the model, with the exception of that for South Korea in the long term, is relatively good, with a ~1% or better relative error. This match could prove to provide a useful measure, over time, to allow forward looking decisions to be made regarding the state of the pandemic globally and regionally and its evolution.
References


