Practice Challenge: Data fitting to SIR model

Arsha Sherly

September 16, 2022

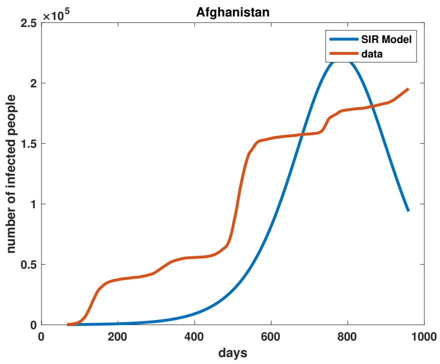
1 SIR model

The SIR model is a deterministic compartmental model used to simulate disease spread and is given by the following set of ordinary differetial equations where S, I and R denotes the compartments into which the population is assigned. S demotes the total number of susceptible people, I denotes the infected people and R that group of the population who are removed from the disease cycle either by recovery or death.

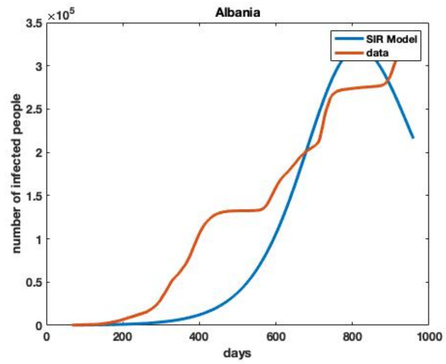
$$\begin{aligned}
\frac{dS}{dt} &= -\beta I \frac{S}{N} \\
\frac{dI}{dt} &= \beta I \frac{S}{N} - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned}$$

The paramters involved are infection rate β and recovery rate γ. A data for the evolution of covid-19 in different countries that was given in COVID 19 Data from John Hopkins University is used to find the optimal value for β and γ to obtain a solution for the SIR model that fits close enough to the data. We have done this study on the data for the countries Afghanistan, Albania, Algeria, Andorra. The data was extracted using the extraction method given in Reading and Visualising COVID-19-data. Since the initial values where mostly $I = 0$ and this denotes the disease equilibrium the simulations were started from $t = 67$. The code has averaged out the total confirmed cases for these countries by 7 days. So the data denotes the 7-day average of the confirmed covid cases for each day.

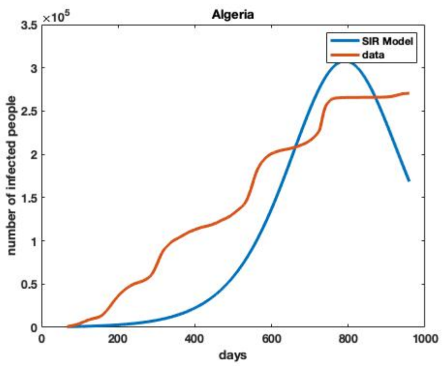
The data has been fitted using fminsearch function in MATLAB. The parameters were obtained and a comparative study has been done by plotting the real data to the solution curves from the model using these estimated parameters. The results are given in the figure 1. The resulting β and γ values are listed in the table 1



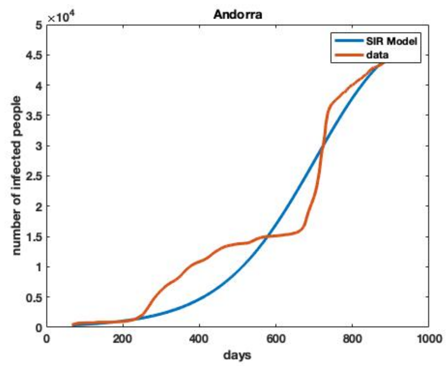
(a)



(b)



(c)



(d)

Figure 1: SIR model vs Data

Country	β	γ
Afghanistan	0.1146	0.1026
Albania	0.0273	0.0155
Algeria	0.0886	0.0784
Andorra	0.0088	0.0011

Table 1: Parameters in each patch

2 Observation and ways to improve the fit

From the parameter estimation and comparison of the model to the real data, it is observed that the fit is poor. Moreover the maximum number of infected people of the model is occuring much earlier than that of the data. A solution to fix this problem was found in two papers. In the basic paramter estimation, we are seeking for two constant parameters for β and γ. At this place if we seek for a time-dependent β and γ this can be sorted. This is computationally achieved either by making use of deep neaural networks as in the following paper or making use of the data itself to find a time dependent infected rate and recovery rate as in a Frontiners article

In the Frontiners article they are using confirmed, recovered and death case data to find out the active cases. This data is used to fit the model. From this article we have understood that the model fitting in the previous section which was done on the confirmed cases may be a problem and that for making use of the active cases we should also make use of the data of recovered and death cases for these countries. Also they have calculated a time dependent contagion rate and recovery rate using the information from the data of confirmed, recovered and death cases. With the available data, we may use a similar method for incorporating a time dependent β to the model and try to fit the data by estimating γ using fminsearch. For doing this, a simple explicit Euler scheme was used to solve the problem involving the upgraded SIR model. While using ODE43 there was a problem with indexing. As a quick fix ODE43 was replaced with an explicit Euler solver. The new model looks like the follows

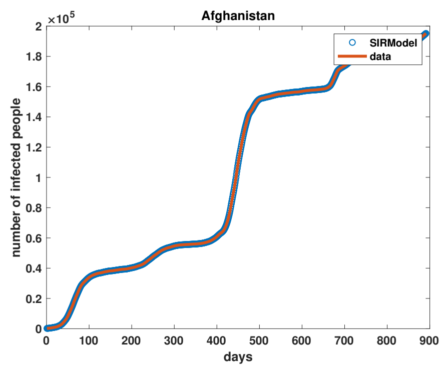
$$\begin{aligned}\frac{dS}{dt} &= -\beta(t)I\frac{S}{N}\\ \frac{dI}{dt} &= \beta(t)I\frac{S}{N} - \gamma I\\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

where $\beta(t)$ is approximated from the data as follows.

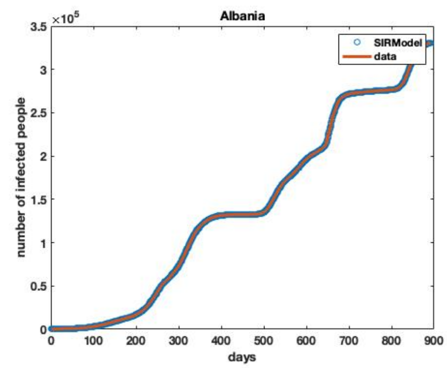
$$\beta_k = \frac{N}{S(t_k)}\left(\frac{1}{data_k}\frac{\Delta data_k}{\Delta t} - \gamma\right)$$

where the suffix k represents the k th data point. Afterwards we use fminsearch to estimate the parameter γ This method gives a better fit compared to the SIR model. The results are given in figure 2

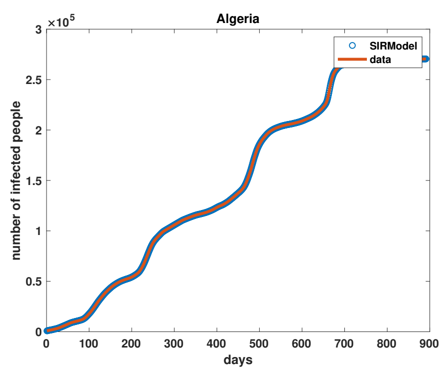
The second method is to assume β and γ as time dependent parameters and use deep neural networks to find the best estimate. This might need a bit more insight into deep neural networks. (I may need a little extra time to work on this. Right now due to time constraints I am not able to try this)



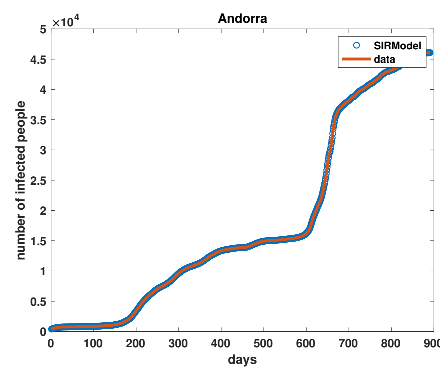
(a)



(b)



(c)



(d)

Figure 2: Comparison of real data and the model shows that the time dependent β and the γ estimated using fminsearch gave a better fit than the regular SIR model

3 Coupled model

How to formulate the coupled model which gives a single model and incorporates the information from all four places? For this we may use the multi-patch model where all four countries may be seen as four patches and assume that each patch has its own SIR model and that they are coupled through a matrix.

One major method is incorporating the concept of migration in the basic SIR model. The system of ODEs is given below where n is the total number of patches

$$\begin{aligned}
\frac{dS_i}{dt} &= \beta_i S_i \sum_{k=1}^{n} A_{ij} I_j \\
\frac{dI_i}{dt} &= -\gamma_i I_i + \beta_i S_i \sum_{k=1}^{n} A_{ij} I_j \\
\frac{dR_i}{dt} &= \gamma_i I_i
\end{aligned}$$

where A is the adjacency matrix which takes values 0 or 1 respectively where $A_{ij} = 0$ means that there is no migration between the patches i and j and the value being 1 when there is a migration between the patches.

Another approach is to incorporate the commuting behaviour through a residence time budgeting matrix P. A multi-patch model can be formulated from the SIR model as follows.

$$\begin{aligned}
\frac{d\mathcal{S}_i}{dt} &= -\sum_{j=1}^{n} \beta_j \sum_{k=1}^{n} p_{kj} N_k \mathcal{I}_k \frac{p_{ij}\mathcal{S}_i}{\sum_{k=1}^{n} p_{kj} N_k} \\
\frac{dI_i}{dt} &= -\gamma_i \mathcal{I}_i + \sum_{j=1}^{n} \beta_j \sum_{k=1}^{n} p_{kj} N_k \mathcal{I}_k \frac{p_{ij}\mathcal{S}_i}{\sum_{k=1}^{n} p_{kj} N_k} \\
\frac{d\mathcal{R}_i}{dt} &= \gamma_i \mathcal{I}_i
\end{aligned}$$

where p_{ij} denotes the average time fraction that people in patch i spend in patch j and $0 \leq p_{ij} \leq 1$ and $\sum_{j=1}^{n} p_{ij} = 1$. Here $\mathcal{S}_i, \mathcal{I}_i$ and $\mathcal{R}_i$ denotes the ratio of susceptibles, infected and recovered people in each patch. The flux part of this system of ODEs also takes into account the total number of encounters happening between the susceptible people of patch i with effective number of infected people in patch j and this term involves the infected people coming from all patches to patches j

If we have the data for P or M matrix we can use the data from all the countries and again use an ODE solver like ODE43 and fminsearch to find the paramters for each patches. (At present I have not added a study for fitting paramters using the coupled models.).