Summer Practice Report
Department of Computer Engineering
METU

Furkan TAŞBAŞI

Summer 2020

TABLE OF CONTENTS

1.Introduction

I have done my summer practise in Parallel Computing Laboratory - Department of Computer Engineering, Middle East Technical University. During my summer practise, I have done a project using MATLAB. I was one of the 10 interns in this project but because of the pandemic situation, I was always working alone at home, that is the overall internship was remote. My project subject was Modeling of Dynamic Systems. The specific research area was epidemic diseases. I specifically worked on a recent epidemic disease named Covid-19. The main purpose of our simulations were analyzing and comparing outcomes of our tasks with real life data which is announced by authorities. I have implemented and adapted existing projects for the data which is announced by Turkey Republic Ministry of Health. During my summer practice, I got help from my advisor professor Murat MANGUOĞLU who is a faculty member in Department of Computer Engineering, METU. He guided me on necessary theoretical background and implementations and answered all my questions about my project and helped me to have an idea about the work environment.

In my summer practice report, I will talk about my project in a detailed way. First, I will talk about the analysis phase and design phase. Then, I will tell the general process using step by step progress. Meanwhile, I will tell the major problems that I had and their solutions, methods followed and key points that I have learned. I will also use screenshots and graphics of my project from time to time.

2. Information about project

2.1 Analysis Phase

In the first two weeks of my summer practice, I only researched some tutorials and tried to learn the basics of MATLAB and simulation techniques. Within this period, I mostly used the website called "tutorialspoint" to learn about the MATLAB graphic part of my project and I read theoretical backgrounds about epidemic simulation models on “sarkac.org”. In these two weeks, I learned enough information about MATLAB simulations and graphics together with SIR, SIRD and SEIR models that I have been assigned work on.

At the end of these two weeks, I have searched for some project examples and tried to have an idea of what I will do for my project. Then I decided to do research about data that I will use for Turkey simulations for the models stated above. Firstly, I carefully considered the details before actually starting to simulation. Then I decided to make comparisons between Turkey and Global data to check any misleading or missing information.

2.2 Design Phase

First of all, I started with the installation of MATLAB. After that, I have done some research about existing SIR, SIRD and SEIR libraries. After some research my advisor Professor MANGUOĞLU sent some libraries related to this area. Then I started analyzing these libraries. List of libraries that I used for this purpose was:

1. "https://www.mathworks.com/matlabcentral/fileexchange/74676-fitviruscv19v3-covid-19-sir-model" which was created by Milan BATISTA for data of Italian Covid-19 epidemic that is working SIR Modelling. Second one is
2. "https://www.mathworks.com/matlabcentral/fileexchange/74838-sird-model-for-covid-19-outbreaks" which was created by Diego CACCAVO for Chinese and Italian COVID-19 outbreaks that is working SIRD Modelling.
3. "https://www.mathworks.com/matlabcentral/fileexchange/74545-generalized-seir-epidemic-model-fitting-and-computation" which was created by Etienne CHEYNET for Numerical implementation of an extended SEIR model with time-dependent death and recovery rates that is working SEIR Modelling.

First of all, I will provide the theoretical backgrounds for models named as SIR, SIRD and SEIR.

- Model SIR (***S****usceptible-****I****nfected-****R****emoved or* ***R****ecovered*):

  Let's say that N is the number of people in the overall population who can be affected by the disease in question. Then in SIR model we can name any person as following:

  ➔ Susceptible : People has got sick

- ➔ Infected : People has infected
- ➔ Removed/Recovered people. : Healed, immuned or dead

According to the SIR model, the numbers above will increase but the equilibrium “S+I+R=N” will be ensured.

In the SIR model, we have two stages. In the first stage, healthy individuals (S) are infected at **β**. In the second stage, the affected individuals are recovered (or die) at a rate of **γ**.

β : average number of individuals contacted by each individual per unit time (a) x probability of transmitting the disease (b)

γ : 1 / time duration of the disease

Assumptions taken for SIR Modeling:

1. The population does not change with external influences; population is closed to factors such as birth, death (except COVID-19) and immigration
2. The life of the disease is limited
3. Individuals who survive the disease are fully immune to the disease, that is they will not be affected again. (Not certain yet for Covid-19)
4. The virus has no incubation period. In other words, every individual with the disease can make others sick as soon as they get the disease.(Not true for Covid-19 since it’s announced that it takes at most 14 days of incubation)

We can derive SIR model with the help of equations below:

$$\frac{dS}{dt} = -a\frac{S}{N}\, b\, I$$

$$\frac{dI}{dt} = a\frac{S}{N} b\, I - \gamma\, I$$

$$\frac{dR}{dt} = \gamma\, I$$

According to the first equation, the number of healthy people will decrease as the number of sick people increases. Sick persons (I) make contact with a person per unit time, which can transmit disease, S / N of these contacts become healthy individuals and b likely make them sick. Thus, the rate of decrease of S depends on both I and S.

In the second equation, there are two factors in the rate of change of the number of patients. The first factor is the change in the number of patients, increasing the speed. The increase due to the sickness of these healthy individuals is again due to both S and I. The second term that balances this is reducing the rate of change in the number of patients and the influence of individuals who are now recovering from the disease and cannot make other people sick.

According to the third equation, the rate at which sick people heal is directly proportional to the number of sick people, remember that is a positive number.

Now let's write these equations for the ratios of the S, I and R groups in the population, so that the calculation becomes easier. So let's rearrange it so that s = S / N, i = I / N and r = R / N and s + i + r = 1. = ab, r (t), which is the proportion of people with the rate of transmission of the disease, increases exponentially in the early days. Finally, it approaches a certain saturation rate, again with exponential decay. There is a transition process that connects simple exponential growth to exponential decay at the end. In the process, the proportion of patients first increases exponentially, eventually decreasing exponentially to zero. The proportion of healthy people who have never been sick, on the

other hand, decreases throughout the process and stabilizes at a small rate at the end.

$$\frac{ds}{dt} = -\beta\, s\, i \qquad (1)$$

$$\frac{di}{dt} = \beta\, s\, i - \gamma\, i \qquad (2)$$

$$\frac{dr}{dt} = \gamma\, i \qquad (3)$$

Using equations (1) and (3), we can make a new equation:

$$\frac{1}{s}\frac{ds}{dt} = -\frac{\beta}{\gamma}\frac{dr}{dt}$$

If we take the integral of both sides from a moment (t = 0) that we consider as the beginning of the epidemic, we get the following equation:

$$\ln s(t) - \ln s_0 = -\frac{\beta}{\gamma}[r(t) - r_0]$$

$$s(t) = s_0\, e^{-\frac{\beta r(t)}{\gamma}} \qquad (4)$$

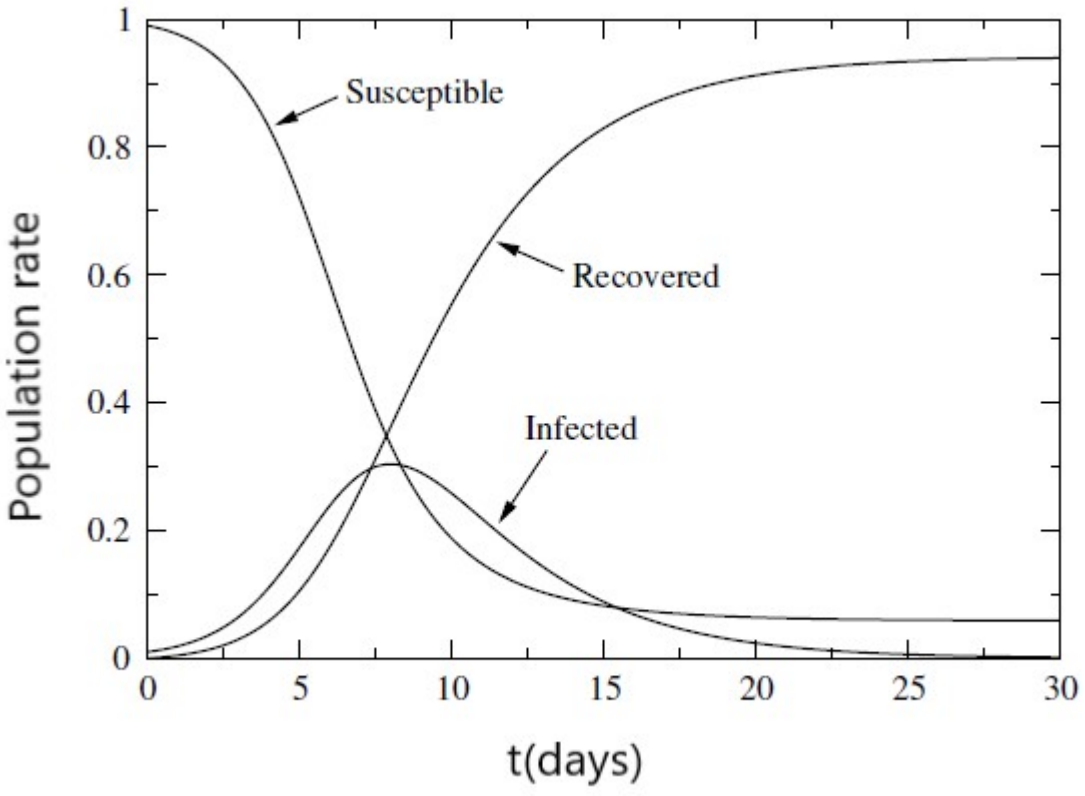
The values at the first moment (at time t = 0) are shown with s_0 and r_0. Here, we assume that there is no one in the r group at first glance. So $r_0 = 0$.

Recall that s+i+r=1, so i=1-r-s. Let's put what we know back into equation (3):

$$\frac{dr}{dt} = \gamma[1 - r(t) - s(t)] = \gamma\left[1 - r(t) - s_0\, e^{-\frac{\beta r(t)}{\gamma}}\right] \qquad (5)$$

We just got an equation for r(t). We know the connection of r and s from equation (4). Since i = 1-r-s, we can find the changes in the number of individuals in all these groups over time.

When we solve these equations in Equation (5) with numerical methods, that is, with the help of a computer, we find solutions for s, i and r as follows.



The graphic shows the changes in the ratio of S, I, R variables to the fixed N in this model over time after the disease occurs. For this graph, the parameters **β** = 1, γ = 0.4, s_0= 0.99, i_0= 0.01 and

r_0= 0 were taken. In other words, only 1% of the population is sick at t = 0.

There are some important points that grabs attention:

- As S individuals transform into I individuals, the proportion of S in the population monotonously decreases and the proportion of Rs in the population increases monotonously. The rate of I's increases a little at the beginning, but decreases over time.
- The rate of S's does not go to 0 because after a certain period of time, it cannot make more individuals sick because there are no sick individuals in the population. In other words, some of the population can bypass the epidemic without getting sick.
- The number of individuals who have had the disease since the beginning of the disease and who have been I at any time period, that is, the balance sheet of the epidemic which is given by overall R.

After a long time, dr/dt = 0, that is, the numbers will not change anymore.

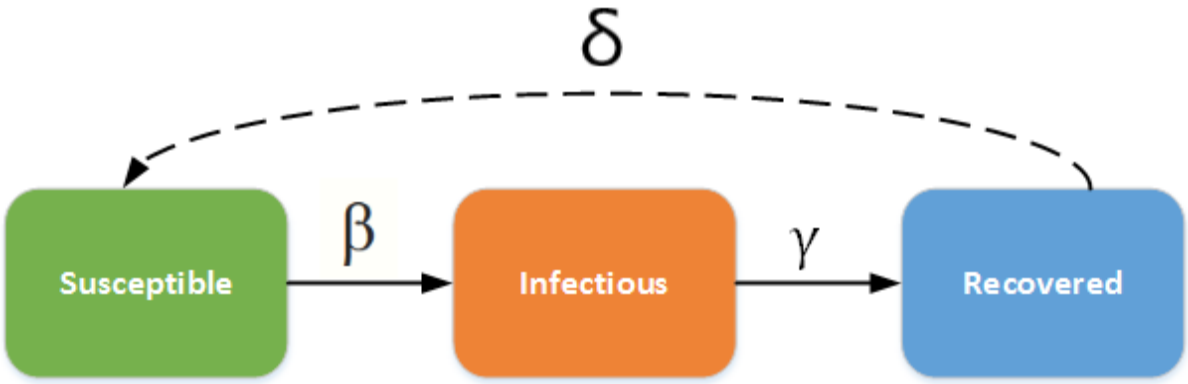
We can find the balance sheet of the epidemic using equation (5):

$$r_B = 1 - s_0\, e^{-\frac{\beta r_B}{\gamma}}$$

For very large populations (like world population) N goes to infinity, while $s_0$will be about 1. Therefore, we can find the following expression for the balance sheet of the epidemic:

$$r_B = 1 - e^{-\frac{\beta r_B}{\gamma}} \qquad (6)$$

- Model SIRS (***S****usceptible-****I****nfected-****R****emoved-****S****usceptible*):



One of the assumptions we made in the SIR model was that the individuals we defined as R would not be sick again, and would become fully immune to the disease after they had the disease. If the individuals who have recovered from the disease can be immunized for a while, the SIRS model is used for such systems. We will define a new **δ** parameter over the parameters we have defined for the SIR model as the rate at which individuals who survive the disease lose their immunity to the disease. Then the SIRS model can be expressed in three differential equations:

$$\frac{ds}{dt} = \delta r - \beta si$$
$$\frac{di}{dt} = \beta si - \gamma i$$
$$\frac{dr}{dt} = \gamma i - \delta r$$

Here S+I+R=1 is still valid.

- According to the first equation, the change in the number of people who are susceptible to the disease increases by individuals who become sick and lose their immunity after surviving the disease, and decreases with the rate of catching the disease.
- The second equation is the same as the SIR model.
- According to the third equation, the change in the number of people who survive the disease increases

with the number of people who recover and decreases with the number of people who lose their immunity.

- Model SIRD (***S**usceptible-**I**nfected-**R**ecovered-**D**ead*):

S, I, R and D stand for the number of susceptible, infected, recovered, and death respectively. The variables, S(t), I(t), R(t), and D(t) show individuals in each component at a specific time.

This model uses the following system of differential equations:

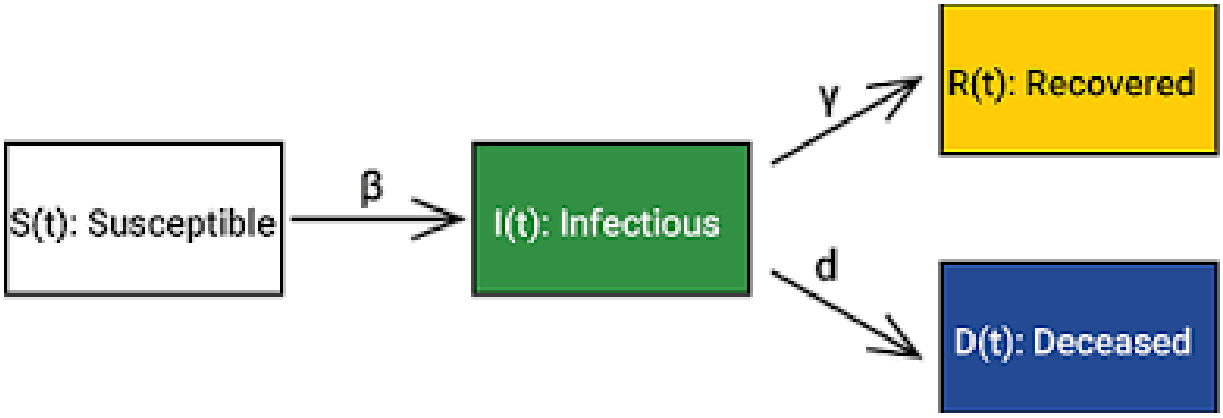
$$\frac{dS}{dt} = -\frac{\beta IS}{N},$$
$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I - \mu I,$$
$$\frac{dR}{dt} = \gamma I,$$
$$\frac{dD}{dt} = \mu I,$$

Where **γ** is the rate of recovery, **β** is the transmission rate, **μ** is the death rate and N is the overall population. The first equation governs the rate of change for the susceptible group, which decreases at a rate proportional to **β** as individuals go from susceptible to infected. Individuals in the infected group move at rates proportional to **μ** and **γ** into either the recovered or deceased group. In addition, R_0 , which is defined as **β**/(**γ**+**d**), is called the basic reproductive ratio. This ratio is derived as the expected number of new infections from a single infection in a population where all subjects are susceptible. It means that the numbers of Infectious tend to decrease when R_0 is less than 1 and increases when R_0 is larger than 1.

The summary of calculation in SIRD is:

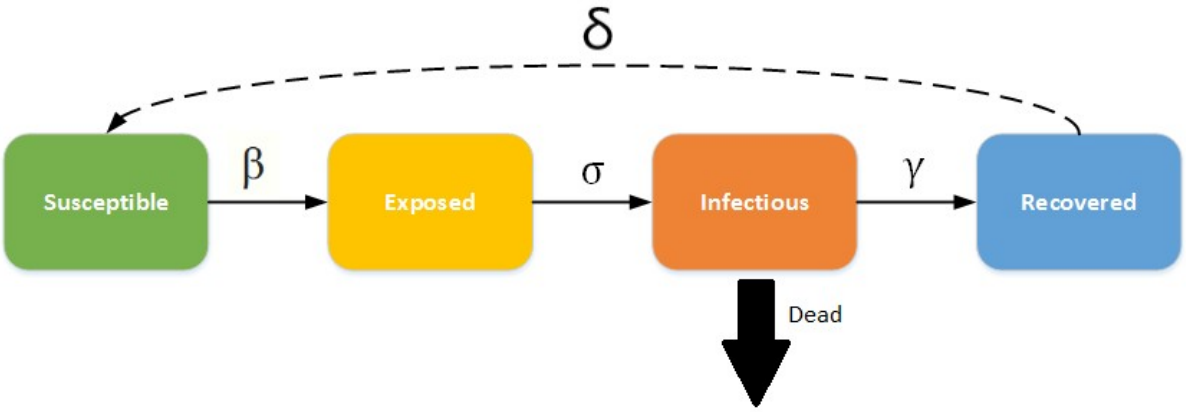
I = N - R - D

The SIRD model scheme is as follows:



- Model SEIR (***S**usceptible-**E**xposed-**I**nfected-**R**ecovered*):

We know that COVID-19 has an incubation period. It is assumed that we cannot transmit the disease to non-sick individuals during this period (Equations : 3,4,5). The SEIR model used in such cases and the course scheme of the disease are as follows:



- ➔ Susceptible : Individuals has got sick
- ➔ Infected : Individuals has infected
- ➔ Exposed : Individuals who have the disease but are not transmitted yet.
- ➔ Recovered : Immune individuals.
- ➔ Dead individuals are leaving the table with the black arrow below.

As seen in the graphic above; after the immunity is lost, individuals will become susceptible to the disease again, that is they will be named as S again.

Once the vaccine is found, we will have to abandon the SEIR model and adopt another model because there will not be any return back to the disease after getting immuned.

2.3 Implementation Phase

2.3.1 SIR Implementation

After analyzing the algorithm Milan BATISTA, I recognized that the data format required by the Matlab algorithm is much more complex and detailed in comparison with data provided by Turkish Ministry of Health.

After searching and comparing with other country data, I decided to use public data provided by The Humanitarian Data Exchange organization.

Then I made some minor changes in the algorithm for parsing data from **humdata** (Humanitarian Data Exchange).

Finally, I ran the project with the input :

- “fitVirusCV19v3("Turkey")”

2.3.2 SIRD Implementation

After analyzing the algorithm Diego CACCAVO, I recognized that the data format required by the Matlab algorithm is much more complex and detailed in comparison with data provided by Turkish Ministry of Health that’s why I decided to use public data provided by The Humanitarian Data Exchange organization again.

Then I made some minor changes in the algorithm for parsing data from **humdata** (Humanitarian Data Exchange) as follows:

confirmed (I) = data_confirmed

recovered (R) = data_recovered

deaths (D) = data_deaths

Finally, I ran the project with the input :

- “sird_covid_run_ita”

2.3.3 SEIR Implementation

After analyzing the algorithm Etienne CHEYNE, I recognized that the data format required by the Matlab algorithm is much more complex and detailed in comparison with data provided by Turkish Ministry of Health that’s why I decided to use public data provided by The G.W.C. Whiting School of Engineering.

Finally, I ran the project with the input after changing parameters and population inputs for Turkey :

- “Example_Country”

2.4 Testing Phase

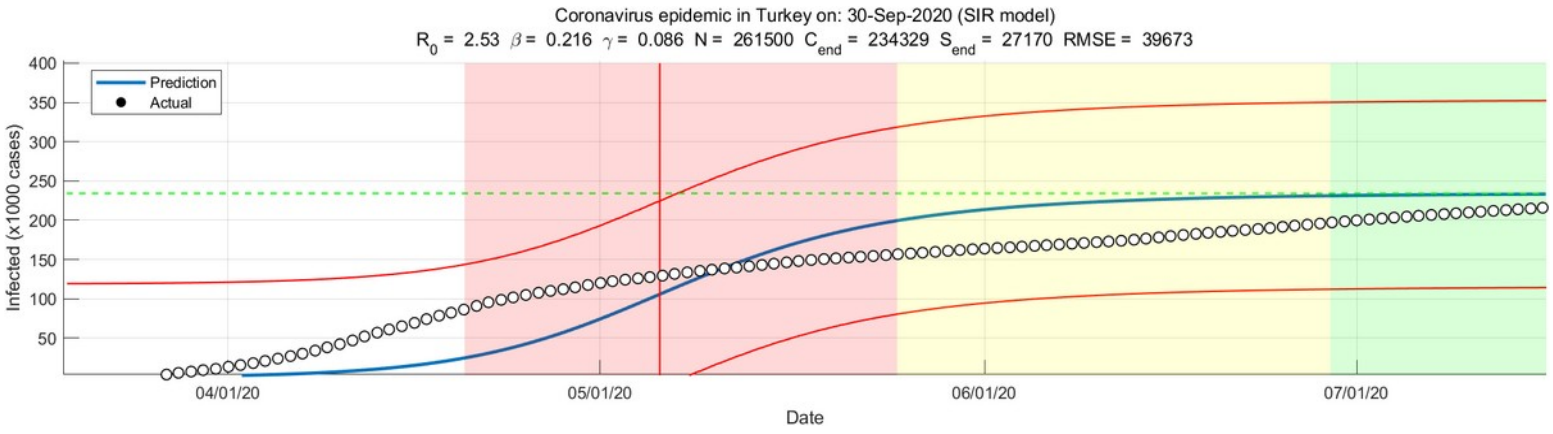
2.4.1 SIR Testing

I ran the project with the input :

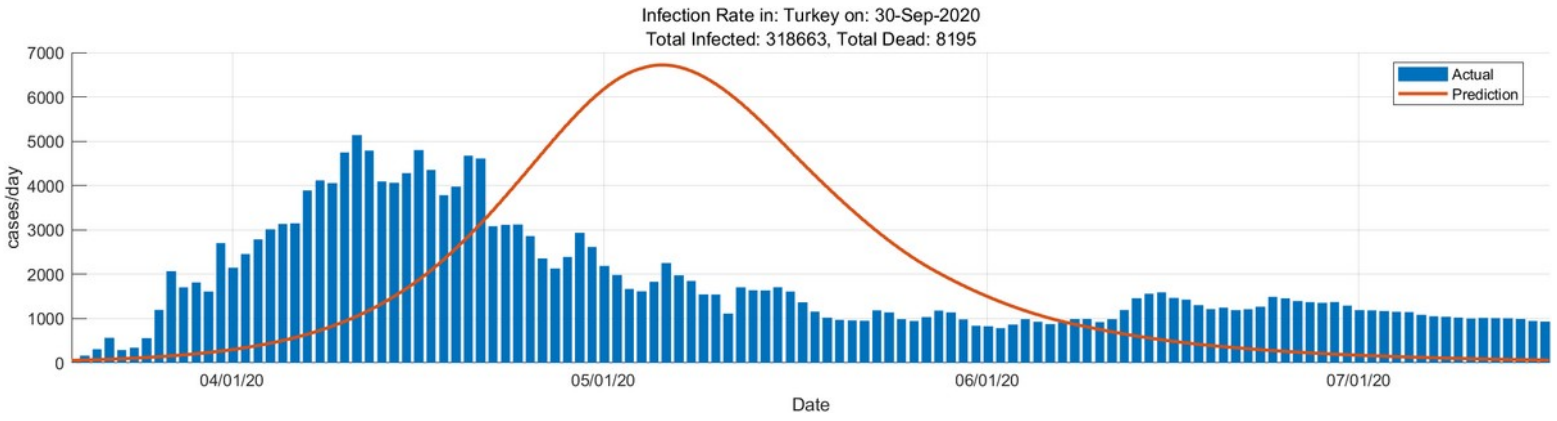
- “fitVirusCV19v3("Turkey")”

The result was :

- ***Warning: R2 = 0.794754



As seen in the figure, the Infected Case number in the graph is approaching the real case values.



On the other hand, this figure shows that Infection Rate is getting more increased in comparison with SIR prediction. The difference between prediction and real infection rates probably occurred because of reduced pandemic measures by Turkish Government.

From the two different figures above we can see that the rate of change has a positive increase after June 2020 which is far more opposite to the prediction data. This result may bond to the releasing pandemic measures.

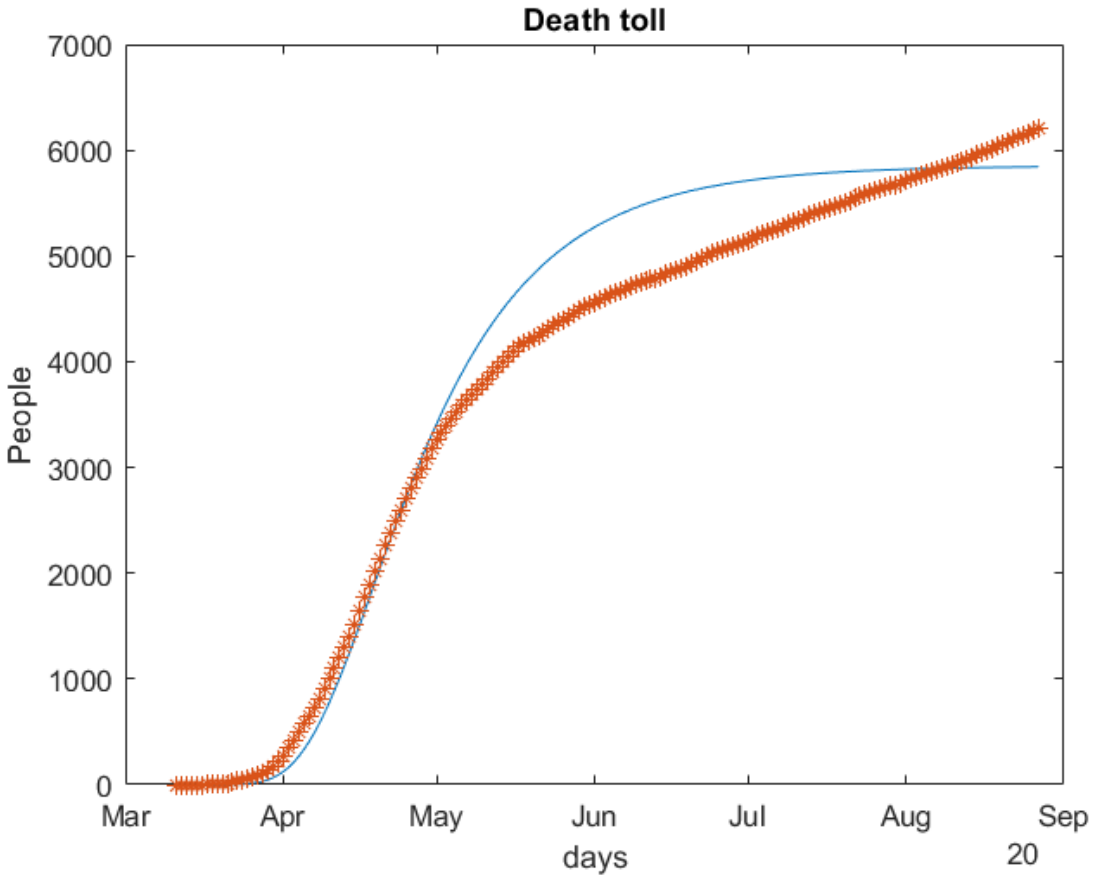
Additionally, a possible skeptical aspect of view may evaluate the first figure as a conspiracy theory which states that the government announces predicted rates instead of real value because the rates are getting closer in the infection graph although the case numbers are exceeding from the predictions.

2.4.2 SIRD Testing

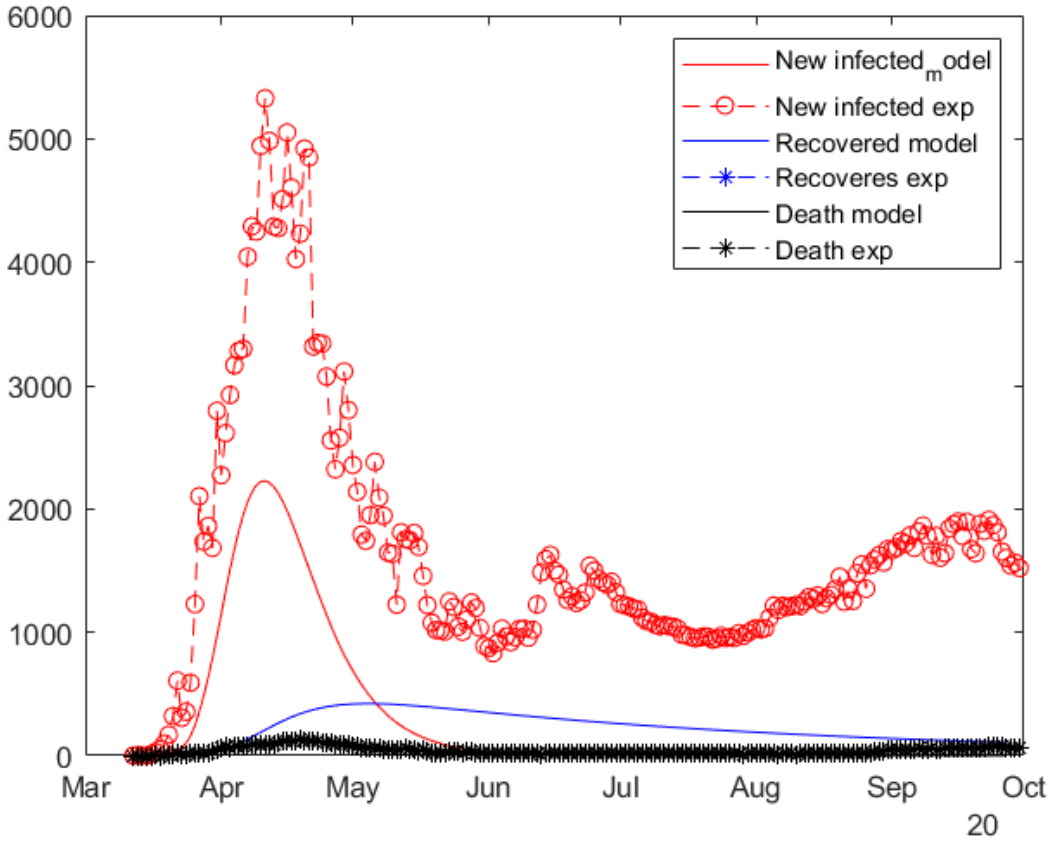
I ran the project with the input :

- “sird_covid_run_ita”

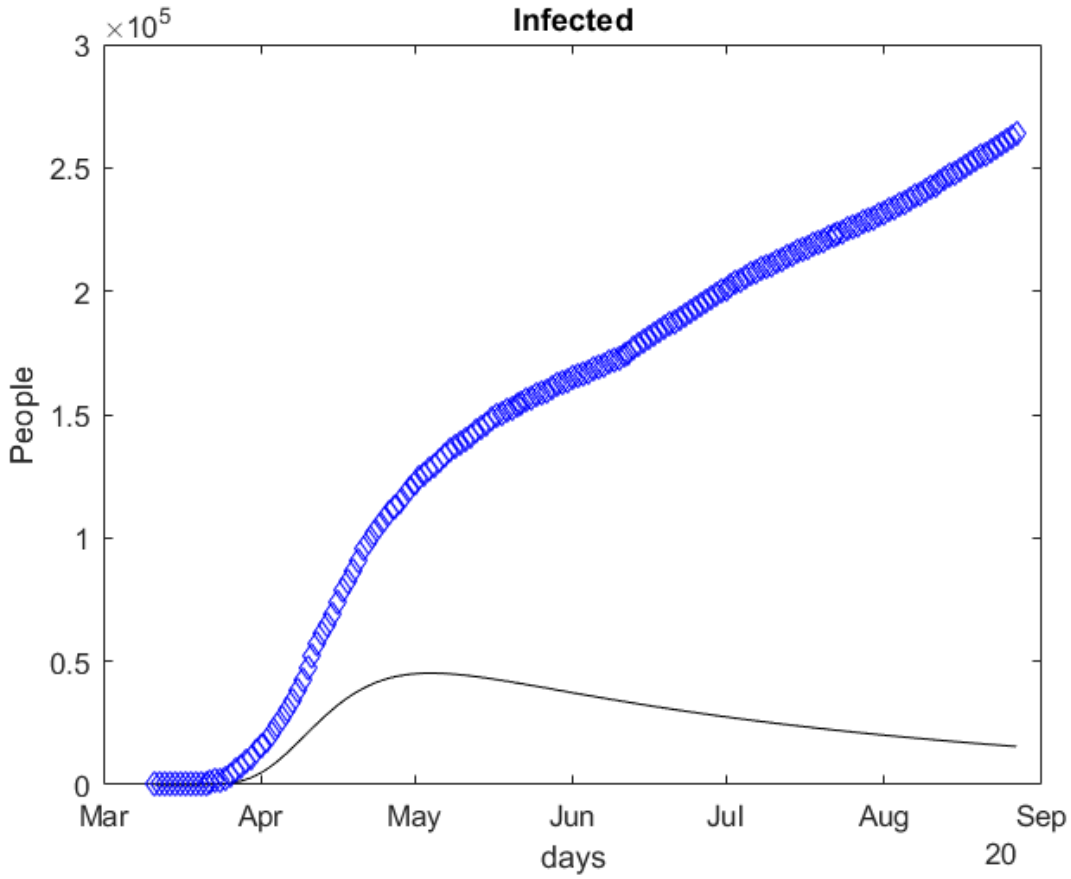
The resulted graphics given below:



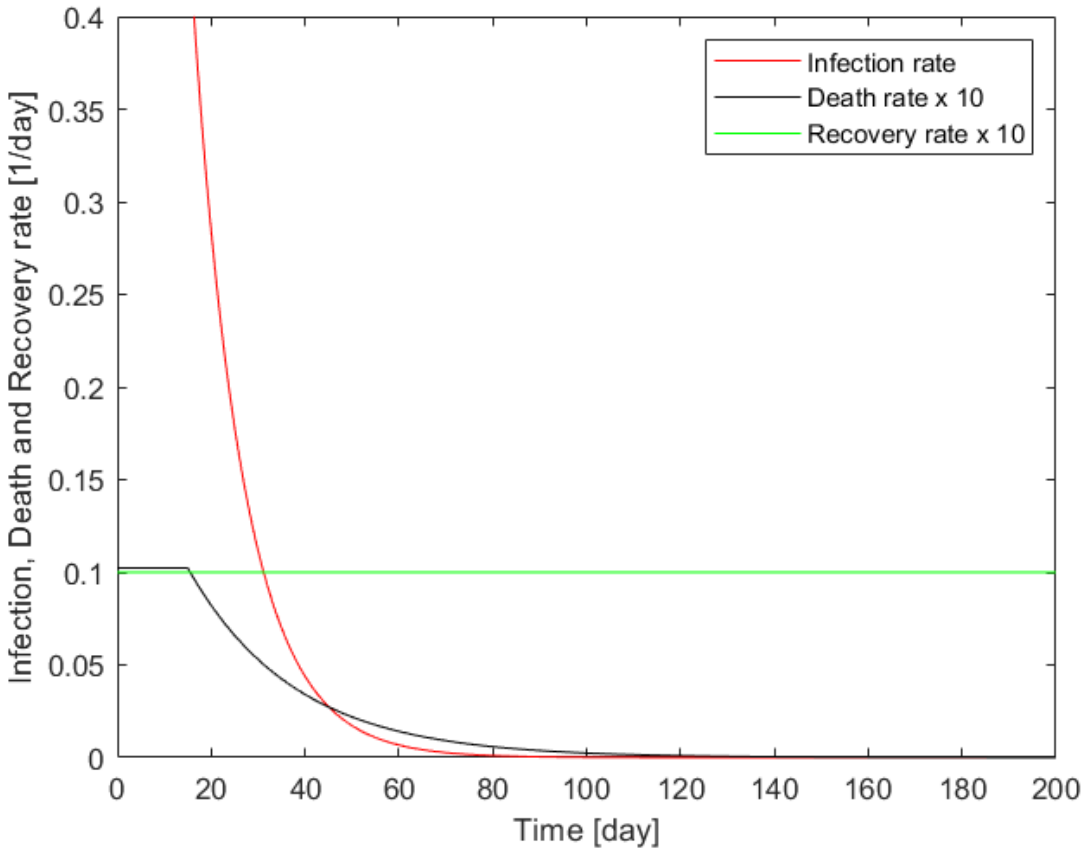
As seen in the graph above, the death toll is below the model between May and August 2020. This is a result spread over time because the infection rates decreased because of measures taken by the government between March and May. But as it is seen from the graph, the rates increase after August since it is also a result of release of measures in the summer term which spreaded over time and seen after August.

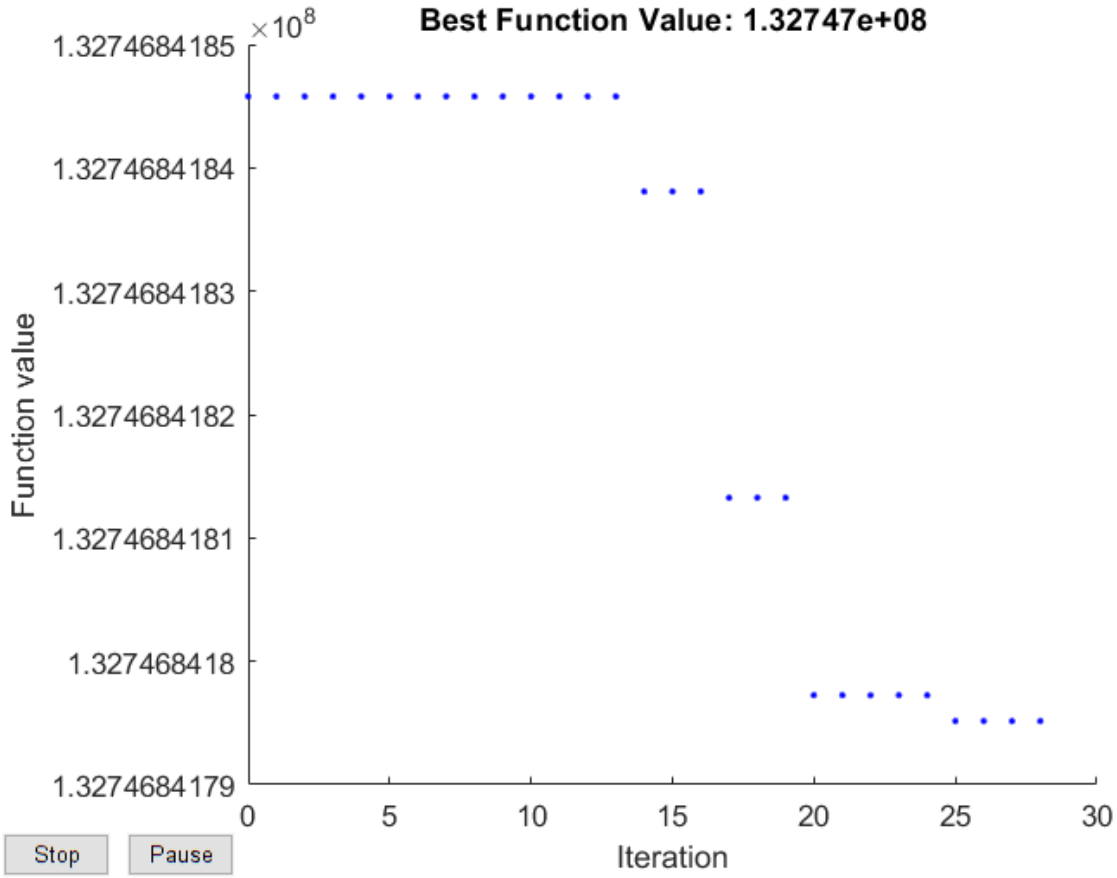


As seen in the graph above the infection numbers are far more than model at all times.

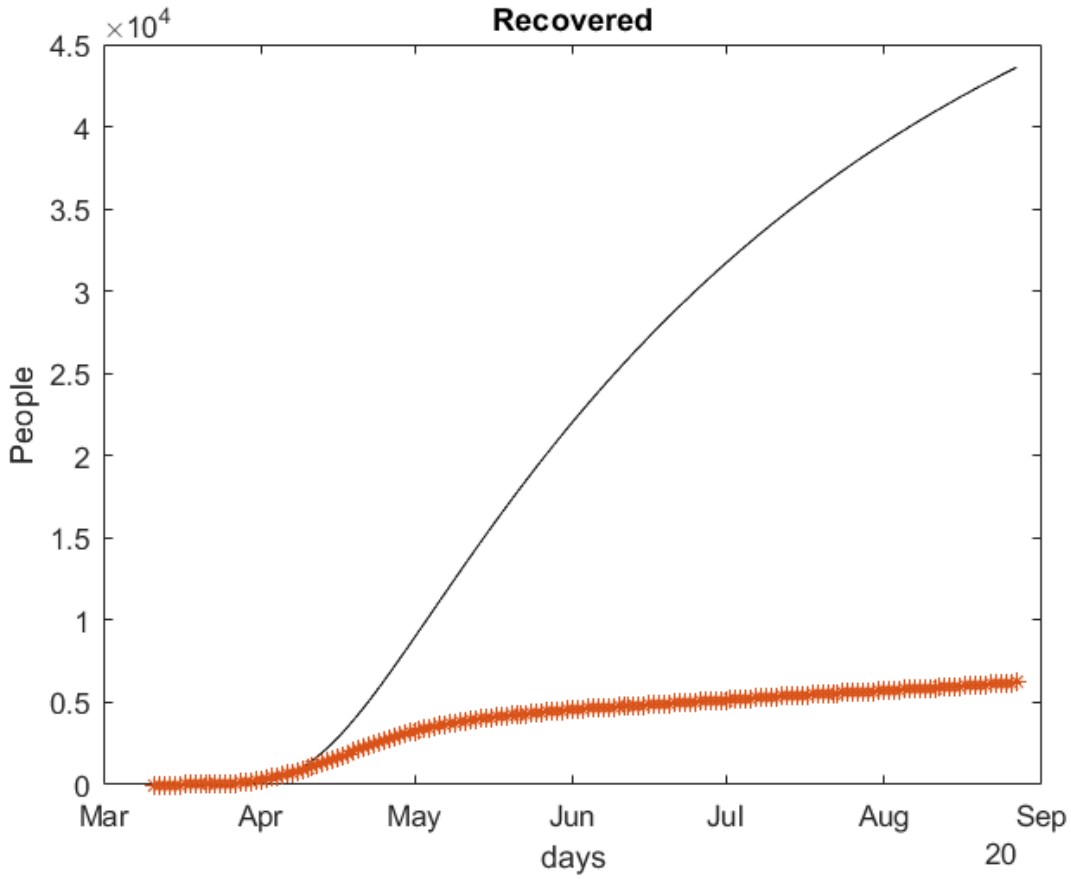

Infected
×10^5
3
2.5
2
1.5
1
0.5
0
People
Mar
Apr
May
Jun
Jul
Aug
Sep
days
20

Infection rates are approaching the peak point in time as clearly observable from the 100th day.





Recovery rate does not fit with the real data because of complex side effects such as chronic illnesses of patients which decreases survival rate.

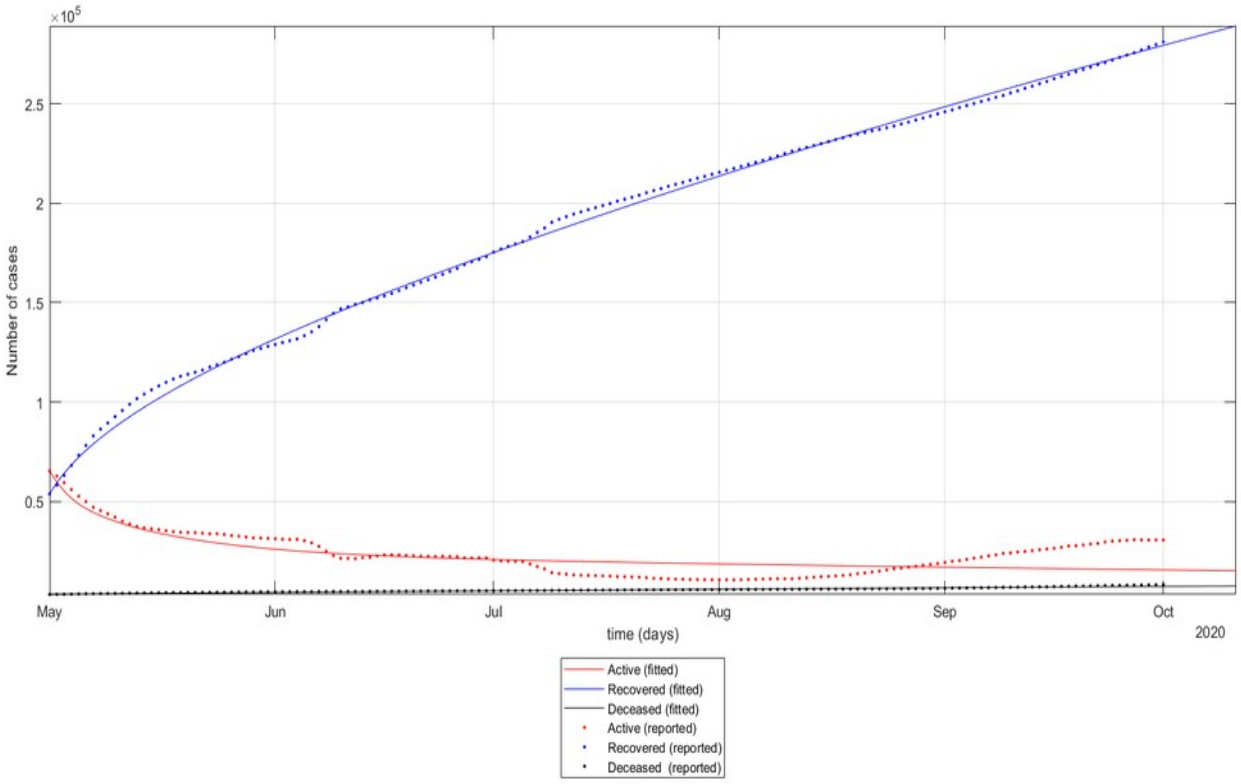


2.4.3 SEIR Testing

I ran the project with the input after changing parameters and population inputs for Turkey :

- “Example_Country”

The results are given below :



In comparison with both SIR and SIRD models, the SEIR model fits more accurately than others as seen above. There is a decay from July to September in real data if we compare with model expectation. But there also exists an increase after September which is the same pattern seen in both SIR and SIRD models.

On the other hand, recovered rates always seem fit with model expectation rates, that is we may conclude that health facilities and actions seem stable (successful) up to this point.

3. Organization

3.1 Organization and structure

The Computer Engineering Department was seeded back in 1967, as a service department to teach courses in computer science discipline to other academic departments in Middle East Technical University (METU). In 1977, the department was integrated into the Faculty of Engineering with the current

name of “Computer Engineering Department”, delegating the new mission of providing degree programs of its own.

Computer Labs

Most of the courses require the students to implement their projects and homeworks on modern (Unix/Linux/Solaris/Windows) operating systems. There are 3 individual laboratories where lab sessions are performed and students can implement their projects and homeworks. All the computers are interconnected via Gigabit Ethernet, which is linked to the METU Campus Backbone via fiber connection. These laboratories include Intel-based all-in-one computers, running Linux (Ubuntu) operating systems. The laboratories are equipped with high performance servers.

Other than software laboratories, a hardware laboratory is used in practical work, concerning courses such as EE 281 Electrical Circuits, CENG 232 Logic Design, CENG 336 Introduction to Embedded Systems Development.

Computing Services

Each student obtains a computing account to be used in the department. This account comes with an email service, a file-storage area which students can use to publish their websites. We also maintain a News Server which provides news channels in various topics, to induce communication and discussion among the students and faculty.

Wireless Network

The students are able to access the department's wireless network. We currently have eleven wireless access points covering the Department's both buildings and also .

High Performance Computing

There is a High Performance Computing cluster in the department, which is used for research and coursework. The system has 46 nodes, with a total of 368 cpu cores, 736 GB main memory and 6 TB of shared storage area. More information on the HPC system can be found at HPC web site

Big Data Servers

There is a BigData server in the department, which is used for research and coursework. The system has 7 nodes with a total of 224 cpu cores, 896 GB main memory, 2*200 SSD system disks and 336 TB of shared storage area. The MapR Converged Data Platform which is the industry's only platform to integrate the enormous power of Hadoop and Spark with global event streaming, real-time database capabilities, and enterprise storage is installed to the servers.

3.2 Methodologies and strategies used in the organization

The goal of the Department of Computer Engineering at Middle East Technical University is to teach, produce and disseminate theory, principles, practice and know-how of computing for the critical analysis, design, evaluation, and improvement of computer-based systems in the contexts of computers and man, computers and the society, computers and the industry and services.

4. Conclusion

As a summary, I have learned the theoretical basics of infectious diseases with some sample modelling on them and implementations on Matlab. I have implemented and reconfigured some models regarding infectious diseases like Covid-19. In this internship I got a chance to analyze infections and their consequences on the community. I was able to monitor all Covid-19 cases in different models and the differences between them. That is I was able to observe the relationship between the health politics of countries and their influences on the community. I also recognized that these politics also affect the modelling parameters and outcomes. Finally, I have learnt the importance of simulation and modelling on pandemic situations because it can easily lead to authorities for possible precautions for all fields of social life in the entire country and all over the world.