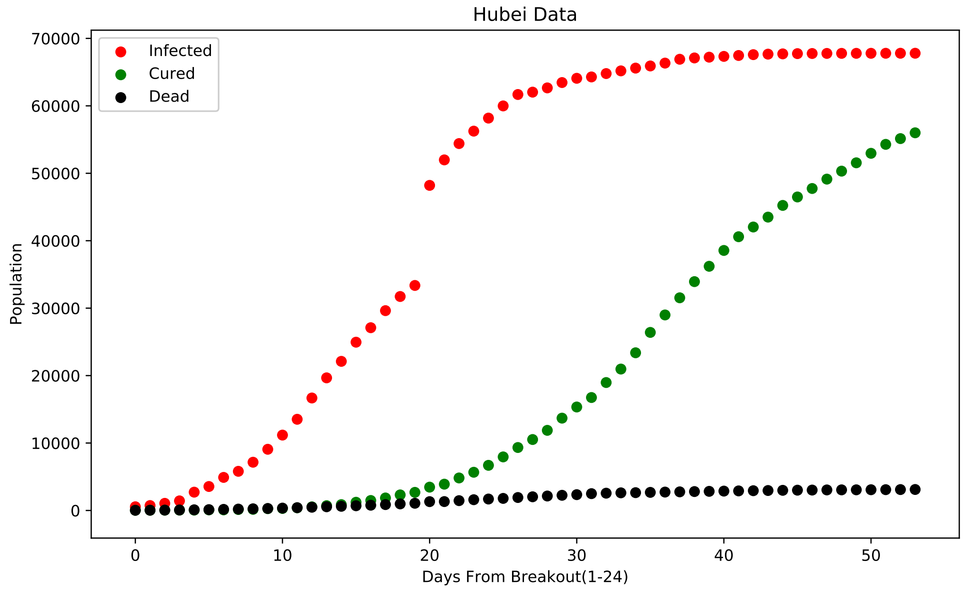
Takehome Project Covid-19

Eric Qu (zq32)

March 21, 2020

1 Source of Data

The data of infected, recoved and dead populaion comes from a github repository[1]. This respository contains data from all over the word. Take the data of Hubei provience for example. It has three variables: accumulated infected, accumulated recovered, accumulated death:



2 SIR Model

The SIR model divided the population into three parts, S for Susceptible, I for Infective, R for Removal. Firstly, the whole population is almost all S with a small amount of I. Then, the S is turning to I, and I is turning to R. It should be noted that the dead and cured populations are all R, since they can not be infected again or have the ability to infect others. This can be represented by a chemical reaction:

$$S + I \xrightarrow{r} 2I$$

$$I \xrightarrow{a} R$$

This first reaction represents the process of infection with the rate r. Then section reaction is the process of death and recovery with the rate a. With rate law, we can write the following ODE:

$$\frac{dS}{dt} = -rSI$$

$$\frac{dI}{dt} = rSI - aI$$

$$\frac{dR}{dt} = aI$$

[1]IDM: https://github.com/InstituteforDiseaseModeling/COVID-19

There is also an important parameter in the SIR model, which is the basic reproduction rate R_0. It represents how many people a single people can infect. This difference between R_0 and 1 is the By the last ODE above, we know that the probability of one patient still alive or uncured is e^{-at}. Therefore, the new population infected by this patient in t is rS_0e^{-at}. So,

$$R_0 = \int_0^\infty rS_0e^{-at}dt = \frac{rS_0}{a}$$

There are some basic assumptions in SIR model:

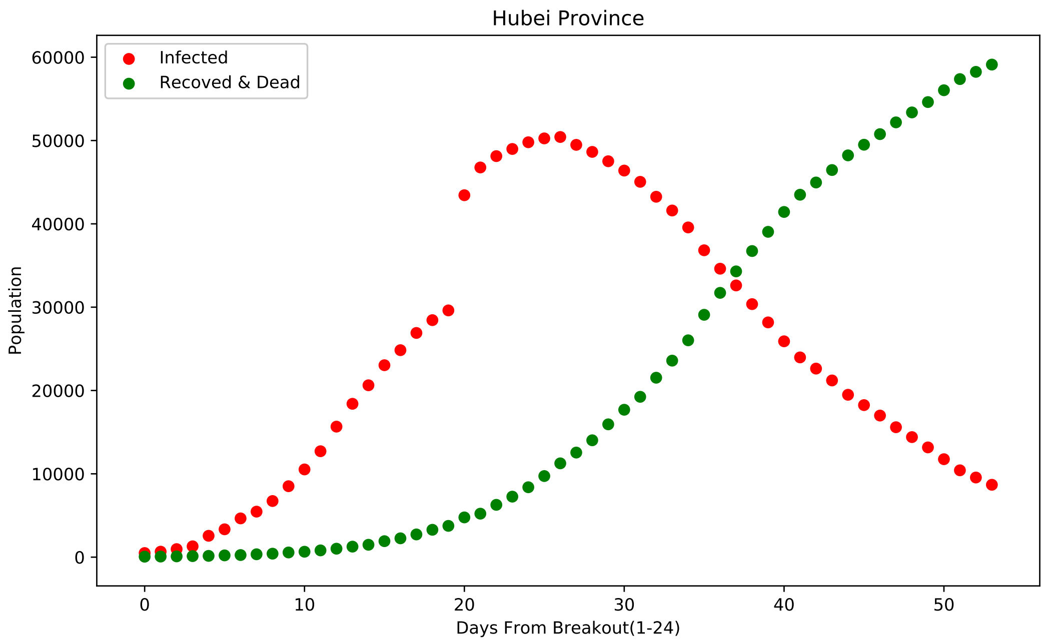
1. Population remains the same in a period of time. No nature death, birth or migration.

2. Everyone can access the population with the same probability.

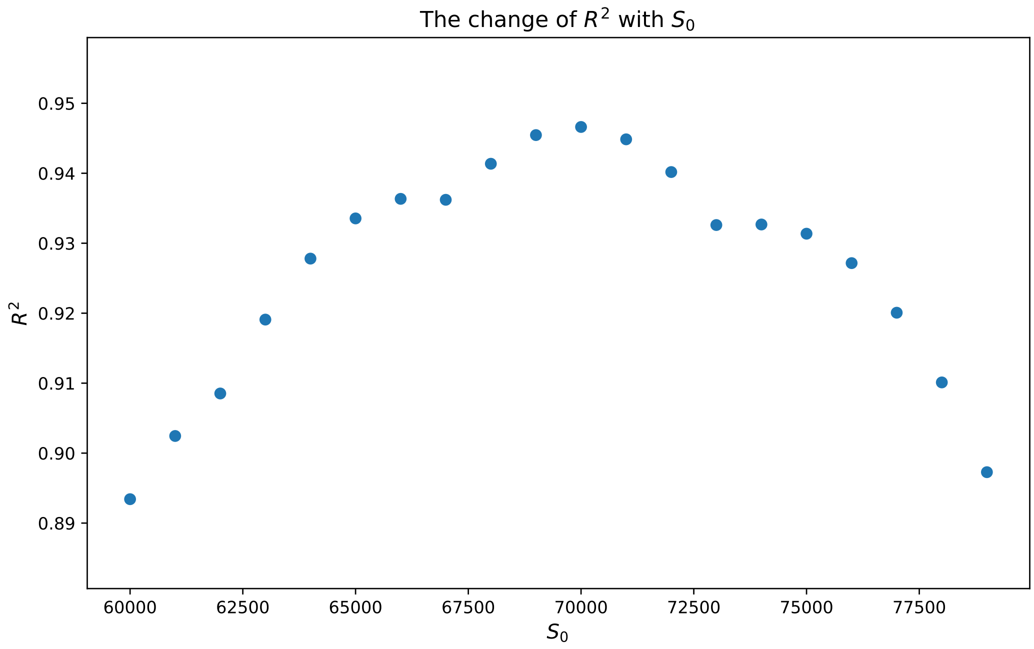
3. The removal rate and infection rate are constant and do not change over time.

3 Fitting the Model in Hubei

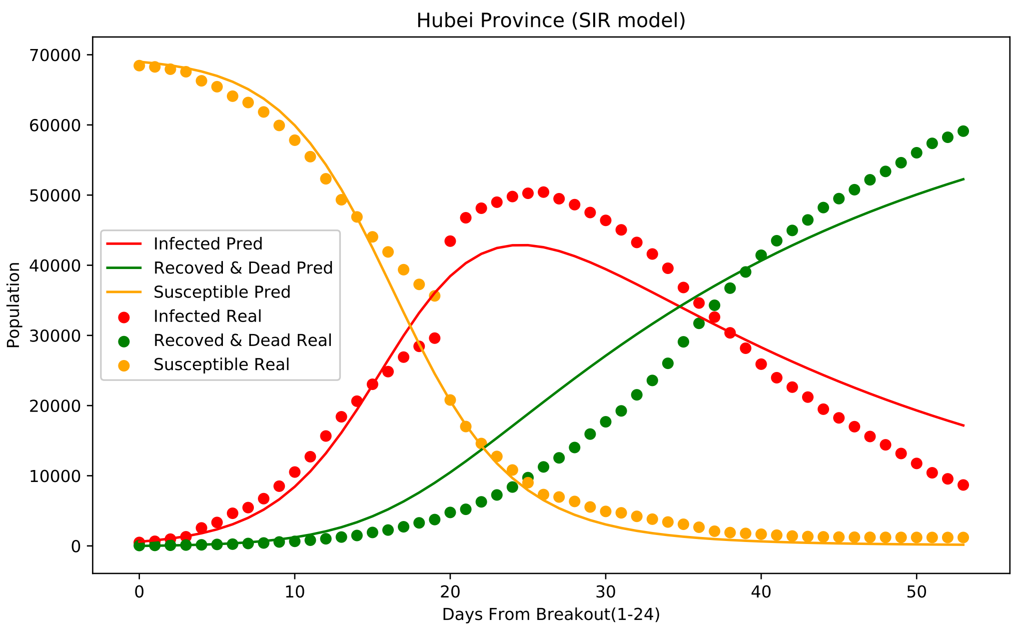
With the assumption 2, we better fit our model in a small region. I choose the data from the outbreak region of the virus. The Infective population is the accumulated infected - accumulated dead - accumulated cured, the Removal population is accumulated dead + accumulated cured. The I and R population is our known data and I fitted my model accroding to that.



Now, there are four parameters need to be fitted, S_0, I_0, r, a. The I_0 could just be set to the initial infected population, but the detection accuracy of the first period time might not be so high. Since the ODE system is really sensitve to the initial values, it is better to set up the I_0 a bit larger. To calcuate the accruacy of the model, I used mean squared error and R^2. For S_0, r, a, I just write a triple loop to try out all the value. I used a large range at first, then made it smaller around the best value in the large range. For example,



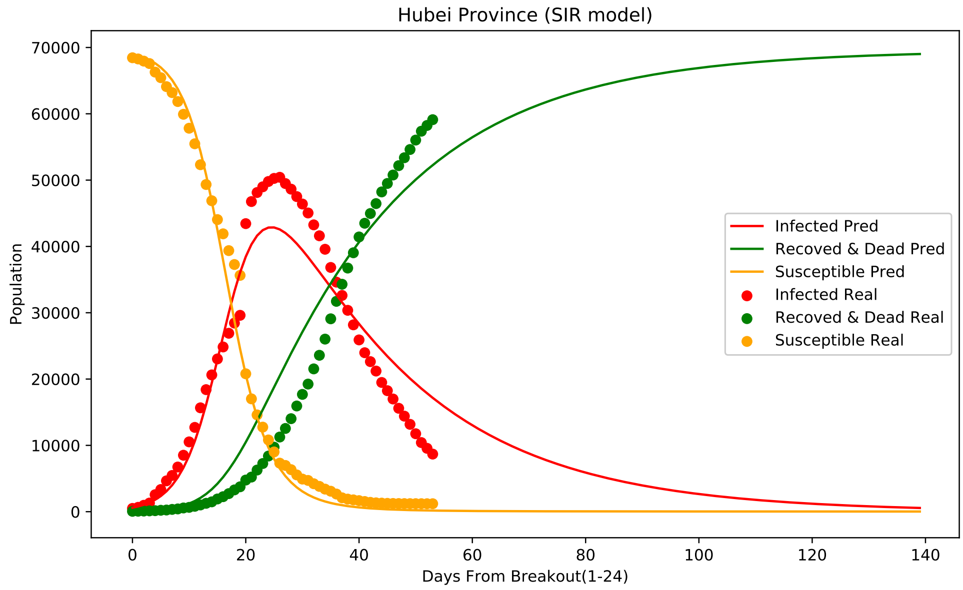
We can see that this range crossed the extreme point, which shows that it is a workable range. The final result:



The R^2 is 0.945. The model is not so accruate because the sudden rise up about 20 days from breakout. The parameters are: $a = 4.6 \times 10^{-6}, r = 0.04, S_0 = 69000, R_0 = 7.94$

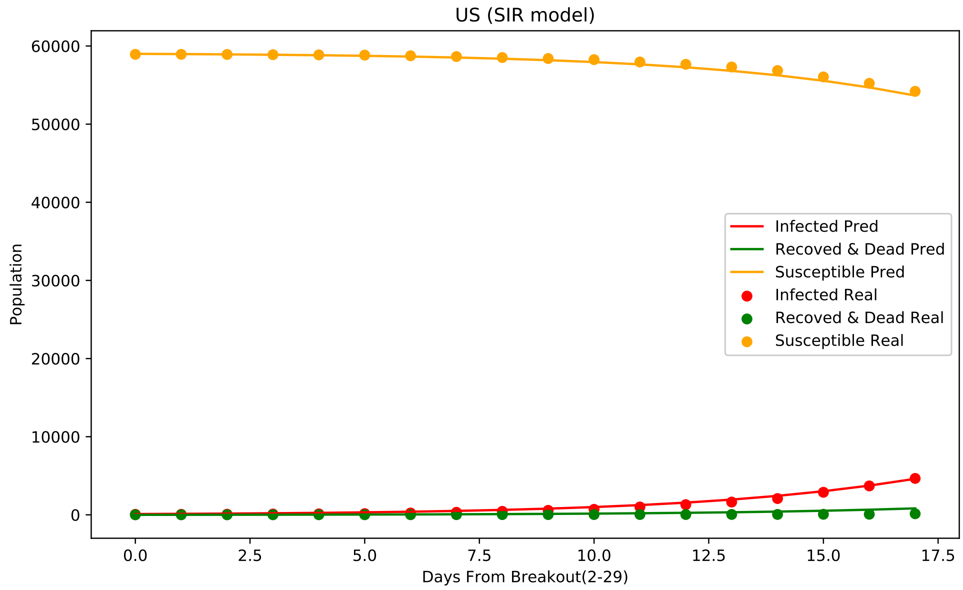
Prediction:

The predicted end of the virus is about 3 months later. But it should be much shorter by the trend of the real data.



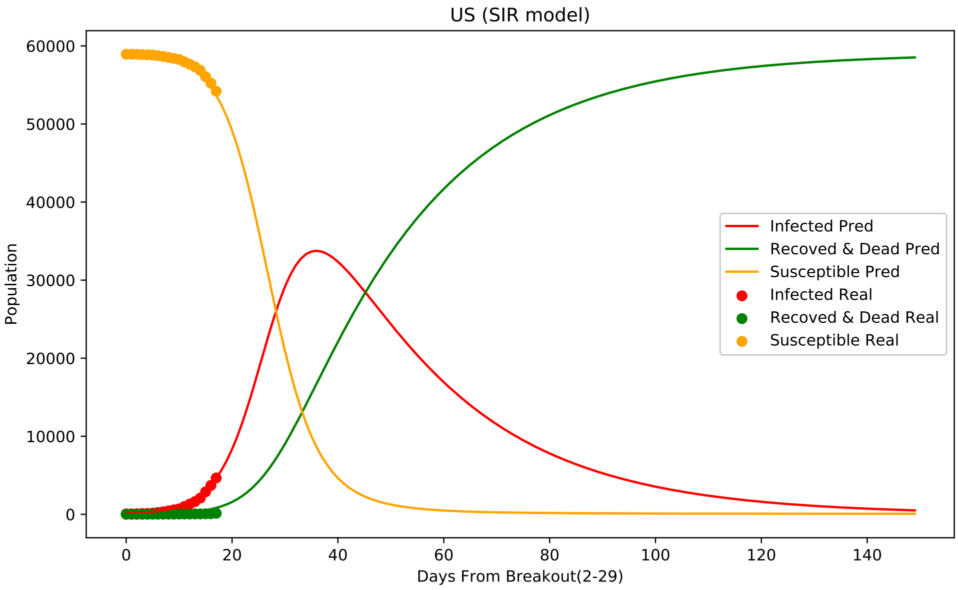
4 Fitting the model in United States

I used the same a for all countries and countinue fitting other parameters. Results:



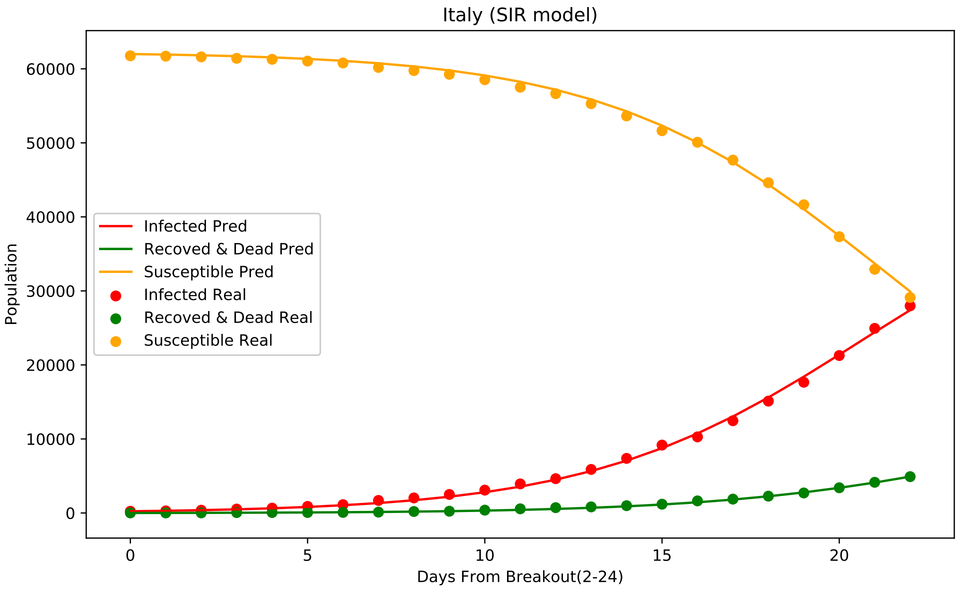
The R^2 is 0.974. The parameters are: $a = 4.6 \times 10^{-6}, r = 0.04, S_0 = 59000, R_0 = 6.78$. Predictions:

It shows that the peak of infection is in about 20 days with about 35000 people and it will die out about 4 months later. We can see that the R_0 and S_0 is smaller than China. This might because the better medical force in the US.



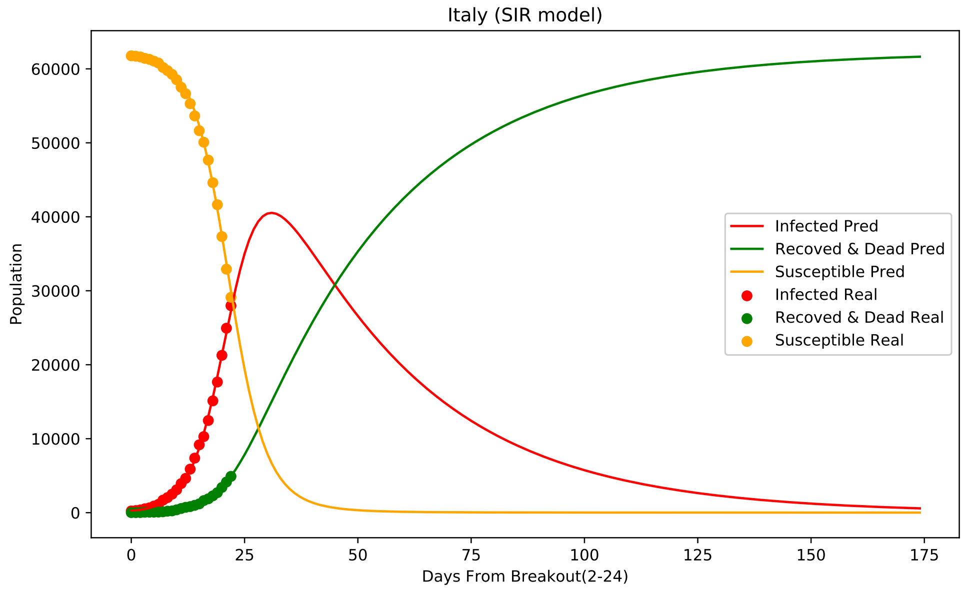
5 Fitting the model in Italy

Results:



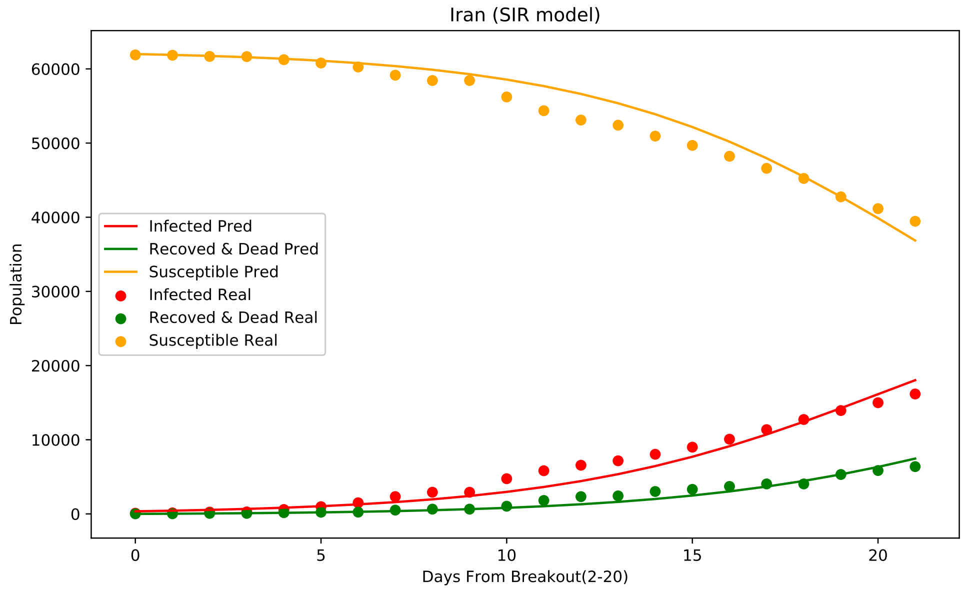
Surprisingly, the R^2 is 0.99. The parameters are: $a = 4.6 \times 10^{-6}, r = 0.031, S_0 = 62000, R_0 = 9.2$.
Predictions:

It shows that the peak is near, in about 5 days with about 40000 people, and it will die out about 5 months later. We can see that the R_0 is really large, which is not surprising.

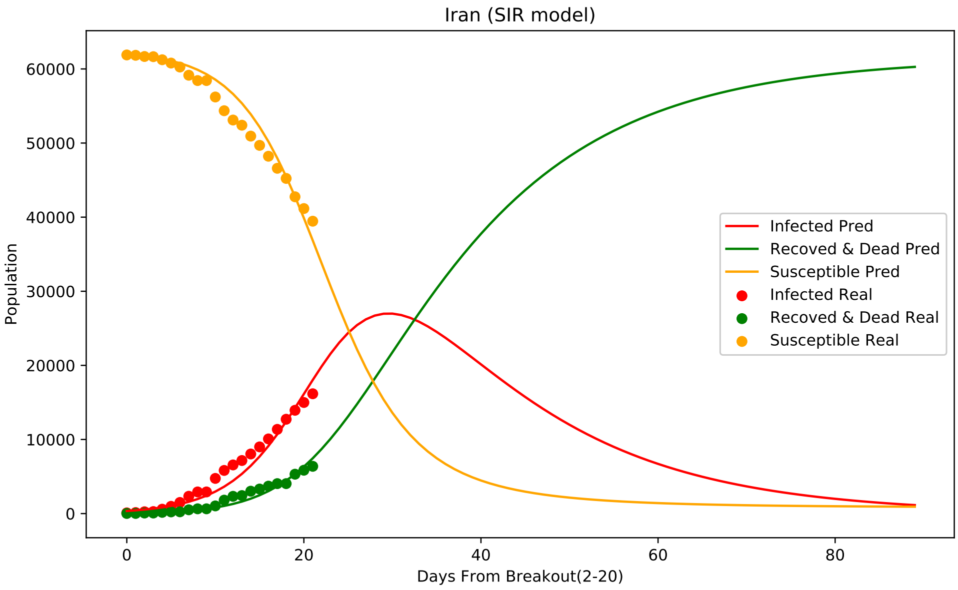


6 Fitting the model in Iran

Results:

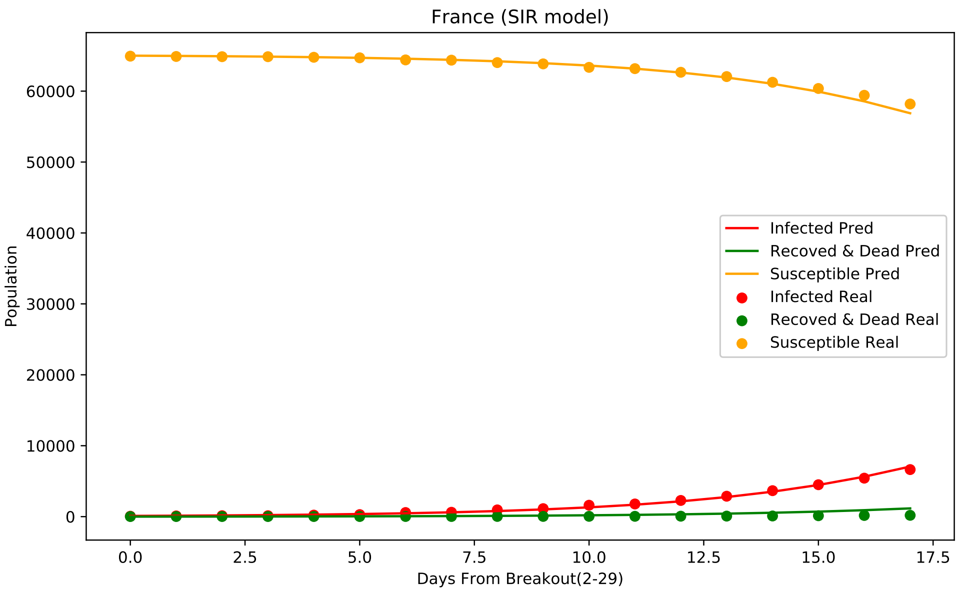


The R^2 is 0.94. The parameters are: $a = 4.6\times10^{-6}, r = 0.066, S_0 = 62000, R_0 = 4.3$. Predictions: It shows that the peak is in about 10 days with about 30000 people, and it will die out about 2 months later. The data has some strange vibrations, which might caused by the poor collections process.



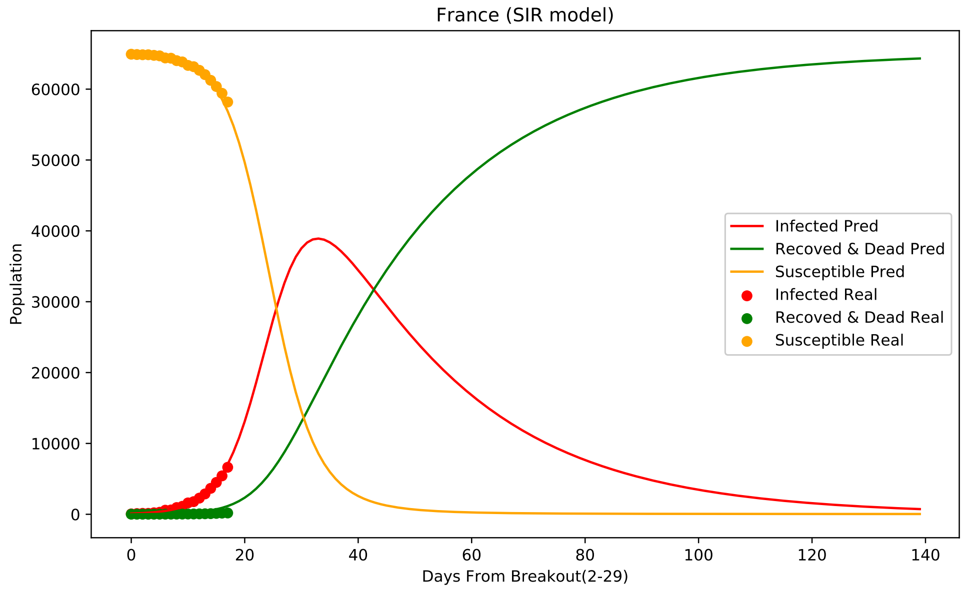
7 Fitting the model in France

Results:



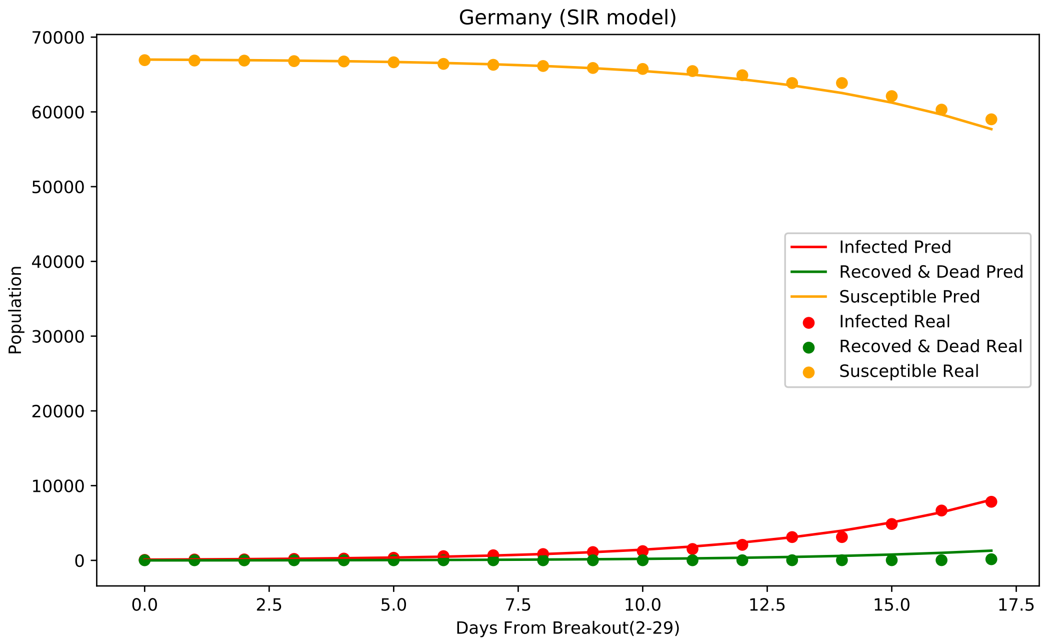
The R^2 is 0.982. The parameters are: $a = 4.6 \times 10^{-6}, r = 0.04, S_0 = 65000, R_0 = 7.47$. Predictions:

It shows that the peak is in about 15 days with about 40000 people, and it will die out about 3 months later.



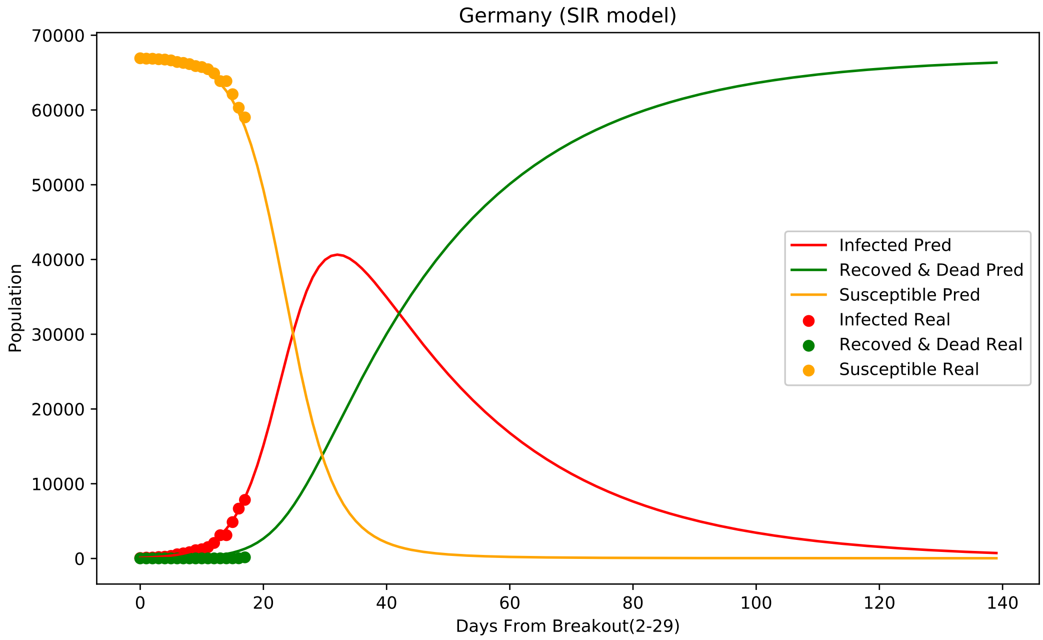
8 Fitting the model in Germany

Results:



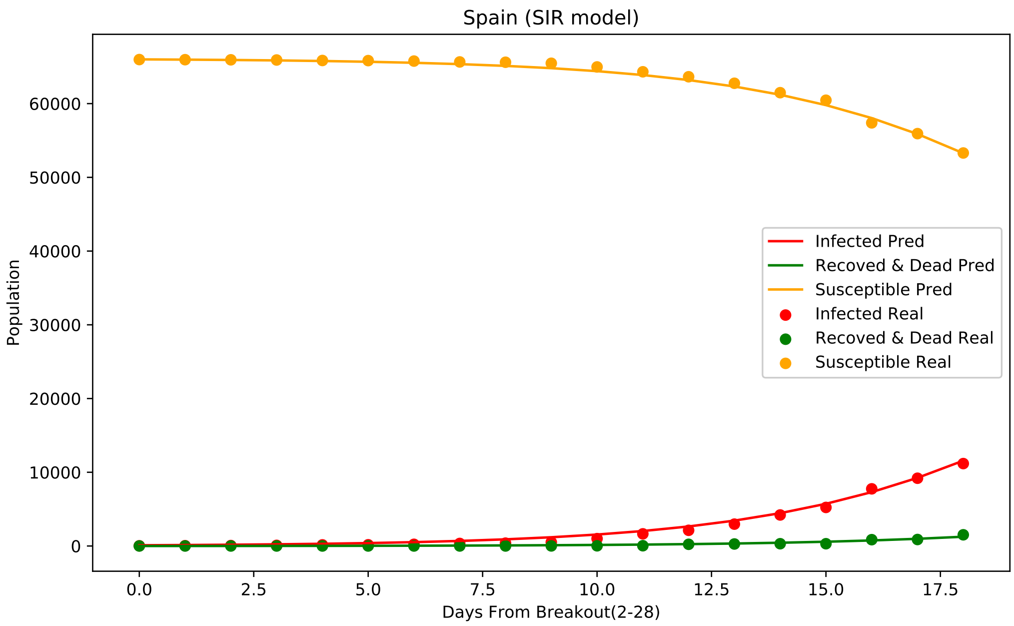
The R^2 is 0.954. The parameters are: $a = 4.6 \times 10^{-6}, r = 0.027, S_0 = 65000, R_0 = 11.24$.
Predictions:

It shows that the peak is in about 15 days with about 40000 people, and it will die out about 3 months later.



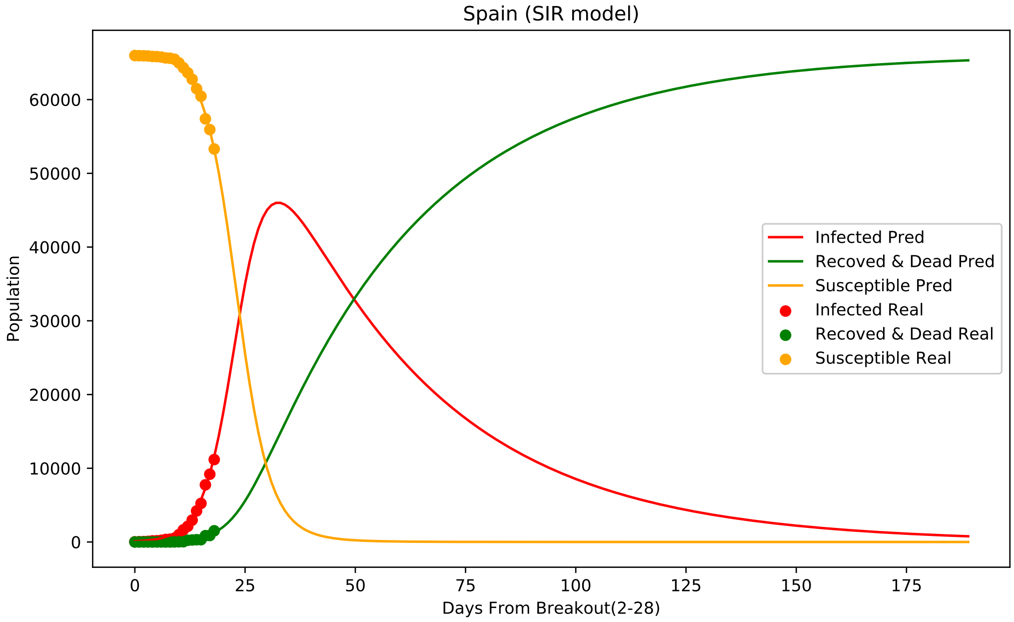
9 Fitting the model in Spain

Results:



The R^2 is 0.83. The parameters are: $a = 4.6\times10^{-6}, r = 0.04, S_0 = 67000, R_0 = 7.70$. Predictions:

It shows that the peak is in about 20 days with about 45000 people, and it will die out about 5 months later.



10 Summary

Continent	Country	R^2	a	r	S_0	R_0
Asia	China	0.945	4.60E-06	0.04	69000	7.94
	Iran	0.94	4.60E-06	0.066	62000	4.3
North America	United States	0.974	4.60E-06	0.04	59000	6.78
Europe	Italy	0.99	4.60E-06	0.031	62000	9.2
	France	0.982	4.60E-06	0.04	65000	7.47
	Germany	0.954	4.60E-06	0.027	65000	11.24
	Spain	0.83	4.60E-06	0.04	67000	7.7

11 Possible Sources of Error

Firstly, this model has many assumptions that contradict the real world. A better and realistic model might be SEIR model that introduced incubation. The data collected in undeveloped countries (such as Iran) might not be as accurate as the developed contries. Also, the government might cover some data, which will also make our predictions inaccurate.