COVID19
How many people will be affected and when?

A Preprint

Beril Sirmacek
bsirmacek@gmail.com

March 22, 2020

Abstract

In this short article, the author discusses the potential spread factors and offers some ideas which could be beneficial in order to model the spread and predict the number of upcoming new cases. The author discusses the spread factors in three stages which could give the parameters of a modeling function. If there could be ways to use such modeling function, the local regions could predict the future numbers of the people who are going to be affected at a certain time. Due to the on going pandemic around the time when this article is written, the author had not much chance to conduct experiments, determine globalization of the hypothesis and validate the modeling idea. Therefore, the author wishes that the proposed hypothesis might contribute to the studies of the other researchers in globe, in order to find better ways to do predictions. If such models can predict the number of people who are going to be affected locally, there might be chances for health services, governments, economists, education institutions and local market suppliers to take precautions in order to reduce the amount of challenges during the pandemic.

***K*eywords** COVID19 · machine learning · prediction models · epidemic spread of disease

1 Introduction

Predicting number of infected population is one of the biggest challenges that researchers have been encountered with. The reason is that there could be many factors that we are not able to take into account.

The author believes that if predictions could be done accurately, the healthcare system could increase the capacities in an early period in order to be able to serve to population. Besides, the economic market, food suppliers and many other sectors including education and transport, might have chances to be prepared in order to reduce the undesirable effects of the pandemic in the local areas.

The author suggest three stages of the epidemic spread in order to determine how to fit a prediction model for estimating number of people who are going to be affected at a certain time.

2 Stage 1 - The country is exposed to the virus

In many countries the first and later on the highest number of cases are seen at the cities which have an international airports and seaports. An example map of south Sweden is represented in Figure 1. In the figure, the main airports are seen with stars and the seaport is seen with a blue dot (in Stockholm). The number of infected population (based on http://www.STV.se on March 15 2020) is represented as a heat map (reddish colormap) on the same figure. The correlation between the airport locations and the number of infected people appears to be high, however there are also many other high infected cities which do not have airports in their close neighbourhood.

The similar pattern can be seen in the map of the Netherlands in Figure 2. The start cases and the highest cases have been the airport cities (Eindhoven and Amsterdam), however further developments have been seen later in the other cities which is suggested as Stage-2. **It is assumed that the Stage-1 determines when to start seeing effects in a certain region (considering the spatial relation to the airport or seaport like international connection points).**

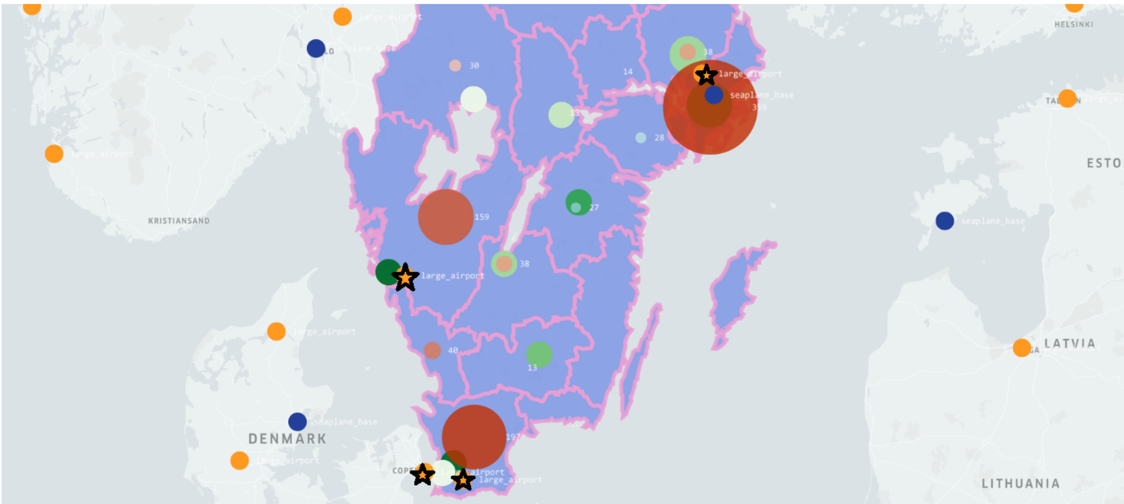

Figure 1: Stars are the airports, the reddish colormap and the numbers are the positive cases (on March 15th, 2020). The green colormap indicates the university populations (seems uncorrelated).

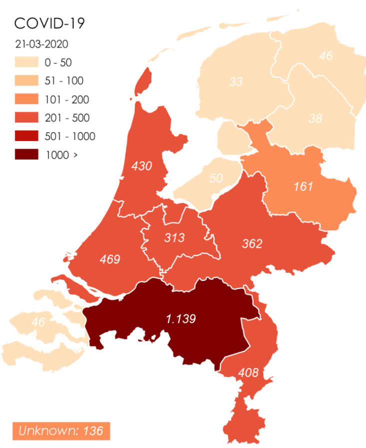

Figure 2: COVID19 cases in the Netherlands on March 22nd 2020, based on Wikipedia.org

3 Stage 2 - The virus is travelling to more cities in the country

At the Stage-2 of the spread, the author first considered the international universities with high populations as a potential impact factor to the number of affected people. Figure 3 shows the populations of the major universities with green color. The university locations did not directly correspond to the cities which do not have airports but high number of positive cases. Therefore, the author could not consider the international university distribution as a high impact role player of the epidemic spread.

Considering the high impact events in Germany and in the Netherlands (having high number of cases in a city suddenly) [1], [2], the steepest case inclines appear at the locations where carnival like public events (indoor or outdoor) happen. **Therefore, in order to model the steepness of the incline (related to the variance), the local event calendar and number of large social activities could be considered. (Proportion of the social events to the population of the area could be inverse related to the variance parameter.)**

To summarize what is discussed so far, a region starts seeing cases (Stage-1) depending on the spatial position regarding the seaport and airports. Afterwards, the region starts seeing high incline (Stage-2) depending on proportion of the number of people who are engaged with large social events.

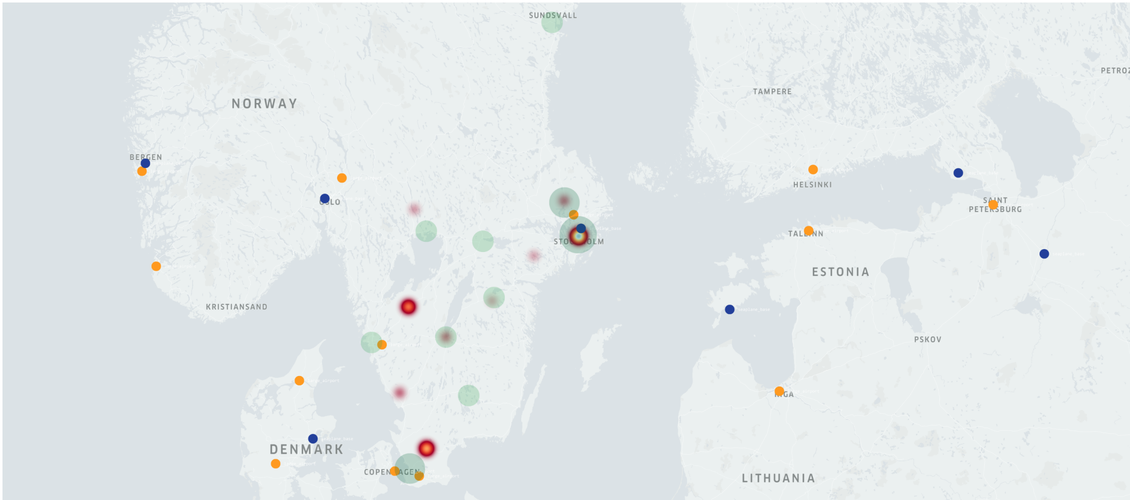

Figure 3: South Sweden map including the major international universities. University populations are seen with green dots where the dot size indicates the population of each university. Orange dots: main airport locations. Blue dots: main seaport locations. Heat map: number of positive Corona cases in March 15 2020.

4 Stage 3 - In each city, the virus is spreading wider within the community

In the Stage-3, the author assumes that many of the people in all around the country are already carriers of the virus. At this stage, children at school, work colleagues and other people who interact in the social environment start transitioning the virus to each other. At this stage (Stage-3), most of the time a sudden decline of the cases are seen if the country agrees with self-isolation. The decline of the model relies on the incubation time of the virus (and of course depends on the maintaining the social isolation). In the right side of Figure 5, the steep decline of the cases in Republic of Korea are seen after the national lock down decision [3].

As suggested by Centers for Disease Control and Prevention (CDC) [4], timeline of the affected people numbers depends on the incubation period of the virus. For a virus with a short incubation time (hours to days), the model looks steep however short. The reason is probably that, since the sick people are recognized quickly, the healthy people can protect themselves (left image in Fig 4). However, for a virus which has longer incubation time (weeks), the spread continues in an extended period since the sick people are not recognized and taken into social activities. The curve which is representing a long incubation time virus effect (right image in Fig 4) can be modeled by an F-distribution function (such as the green color curve in Figure 5 [5] which is often used in statistical modeling of cases which are influenced by many parameters. The good parameter estimation of the F-distrubution function relies on determining factors influencing the parameter d_1 and d_2;

$$f(x, d_1, d2) = \sqrt{d_1(x)^{d_1} d_2^{d_2}} / xB(d_1/2, d_2/2) \tag{1}$$

In this equation $B(.)$ represents the beta function [5]. Mean value $\mu = d_2/(d_2 - 2)$ seems highly related to the time when the self-isolation decision is taken. Therefore if such a decision time exists, the mean value would be known. Knowing the mean value could simply parameter estimation (since d_2 can be calculated already). Besides, looking at the on going skewness of the collected cases d_1 value can be predicted in a more reliable way (since the skewness value is determined by d_1 and d_2).

5 Future work suggestion

The suggested model is currently been tested in the regions where the pandemic started much earlier and further consequences (in Stage-3) are already seen. We should however keep in mind that the reported cases are results of the laboratory tests and these tests are not always accessible to every individual in each country.

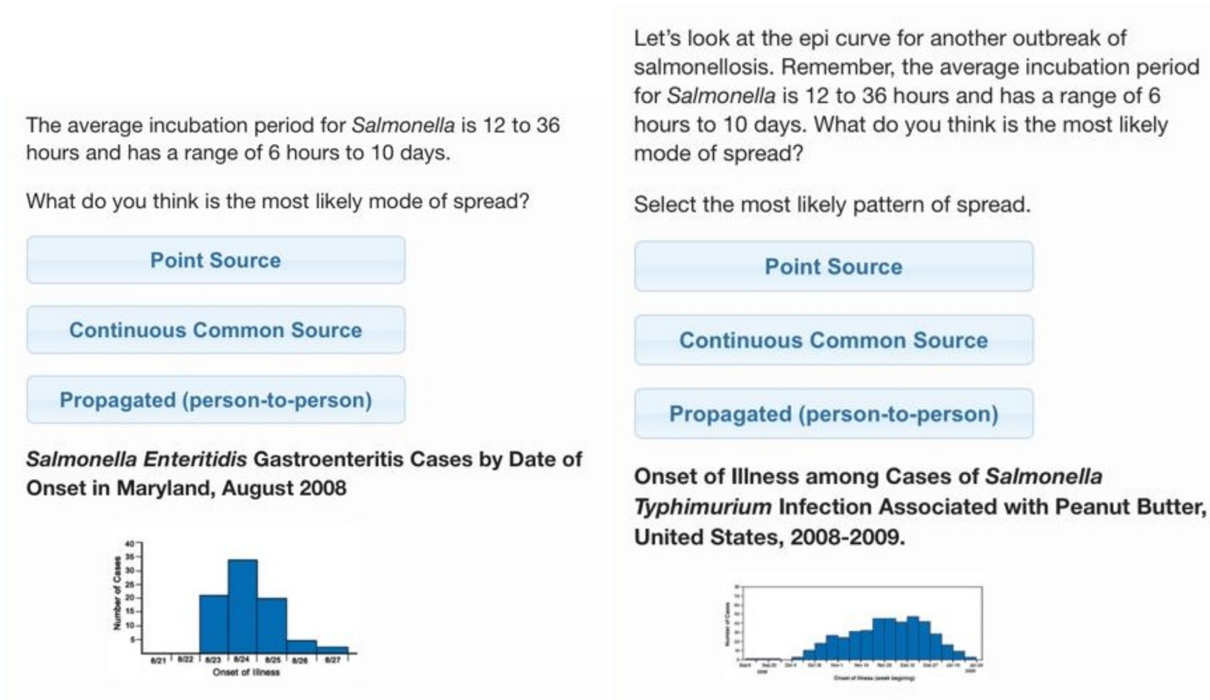

Figure 4: Left: Spread timeline of a virus which has shorter incubation time. Right: Spread timeline of a virus which has longer incubation time.

ACKNOWLEDGMENT

Since the pandemic is currently on-going, the proposed models are not validated. The stages are defined by the author who is from the computer science domain. The author does not have any medical or social domain expertise.

References

[1] `https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_the_Netherlands`. [Online; accessed 22-03-2020].

[2] `https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Germany`. [Online; accessed 22-03-2020].

[3] `https://experience.arcgis.com/experience/685d0ace521648f8a5beeeee1b9125cd`. [Online; accessed 22-03-2020].

[4] `https://www.cdc.gov/`. [Online; accessed 22-03-2020].

[5] `https://www.itl.nist.gov/div898/handbook/eda/section3/eda3665.htm`. [Online; accessed 22-03-2020].

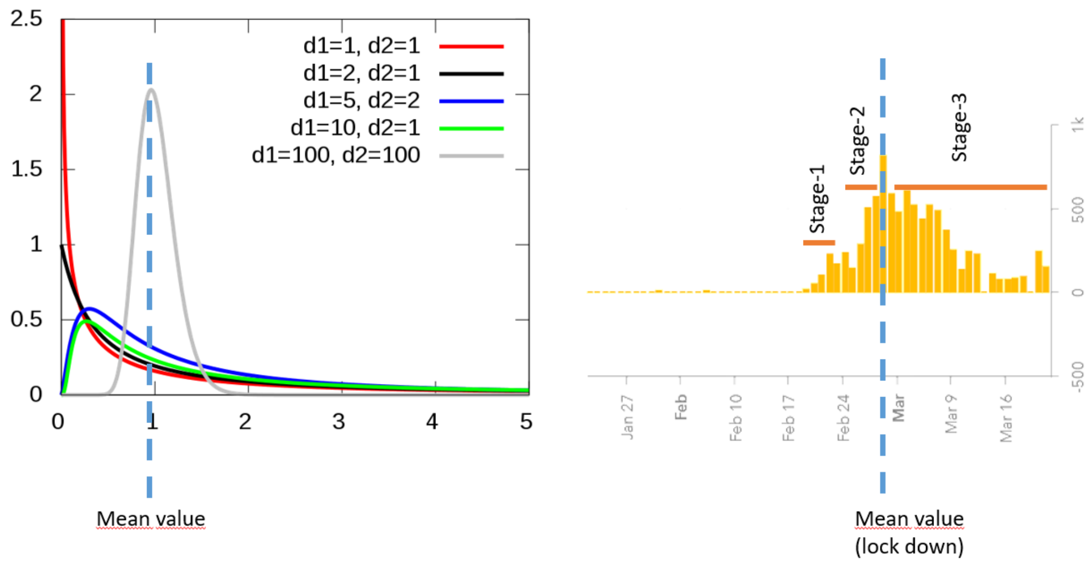

Figure 5: F-distribution which could model number of affected people in the upcoming time frame. The case numbers of Republic of Korea at the right side shows some validation that the function could be used for modelling the number of people who are going to be affected in time.