
Epidemic Modelling via the Physics of the Kinetics of Crystallization

PH3110: Theoretical Project

Author: Anany Tripathi



Candidate Number: 2208048

Supervisor: Dr. Andrew Ho
Department of Physics,
Royal Holloway, University of London,
United Kingdom

11/03/2022

|

Abstract

In this computational Project, I Plan to model the spread of COVID-19 using Mathematica. I'll be using data from The UK, England and Runnymede(which is the surrounding area around Royal Holloway) and different time periods (The first wave and the Omicron wave).

1 Introduction

In This report, I have covered the spread of COVID-19 using data available from the UK government website. I plan to fit the data both exponentially and using the avrami fitting. I'll analyse the data and explain both their pros and cons and which one is better to predict and if avrami fitting can be used by the Government and hospitals for future hypothetical pandemic situations to predict the severity and spread of diseases.

1.1 – Aims And Objectives

- Exponentially fit the data available and draw conclusions
- Fit the data using avrami fitting and draw conclusions
- Understand which model is better for epidemiology

1.2 – COVID-19

The COVID-19 pandemic, also known as the coronavirus pandemic, is a global coronavirus disease 2019 (COVID-19) pandemic caused by coronavirus 2 (SARS-CoV-2) that causes severe acute respiratory syndrome. In December 2019, the unique virus was discovered in an epidemic in the Chinese city of Wuhan, and attempts to contain it there failed, allowing it to spread throughout the world. On January 30, 2020, the World Health Organization (WHO) proclaimed a Public Health Emergency of International Concern and a Pandemic on March 11, 2020. The pandemic has produced over 443 million illnesses and 5.99 million fatalities as of March 5, 2022, making it one of the worst in history.

Spread of COVID-19

Current evidence suggests that the virus spreads mainly between people who are in close contact with each other, for example at a conversational distance. The virus can spread from an infected person's mouth or nose in small liquid particles when they cough, sneeze, speak, sing or breathe. Another person can then contract the virus when infectious particles that pass through the air are inhaled at short range (this is often called short-range aerosol or short-range airborne transmission) or if infectious particles come into direct contact with the eyes, nose, or mouth (droplet transmission).

The virus can also spread in poorly ventilated and/or crowded indoor settings, where people tend to spend longer periods of time. This is because aerosols can remain suspended in the air or travel farther than conversational distance (this is often called long-range aerosol or long-range airborne transmission).

People may also become infected when touching their eyes, nose or mouth after touching surfaces or objects that have been contaminated by the virus.

We know that the disease is caused by the SARS-CoV-2 virus, which spreads between people in several different ways.

All of the data has been collected from the UK government Website - <https://coronavirus.data.gov.uk/details/cases> which monitors everything related to COVID-19 in the UK.

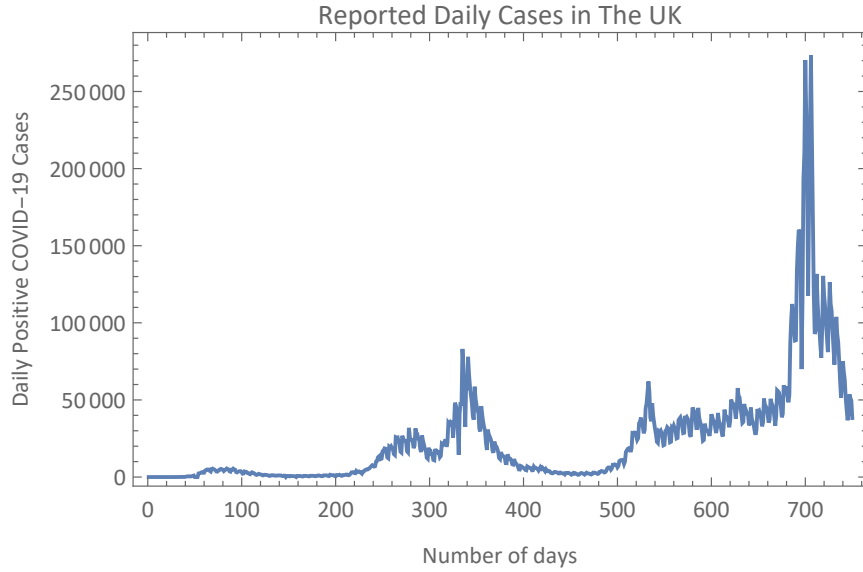


Figure 1: Plot of the daily reported cases in the UK over 750 days

In figure 1 which is the daily case plot, We can see the trajectory of the number of cases going up and down which shows us that the precautions taken by the people and the restrictions imposed by the government did slow the virus down. However, We can see the large spike in the end which is the omicron variant whose growth I will also cover. Short time-frames of this plot might be useful to show exponential growth but a larger time-frame will best be represented by the cumulative plot.

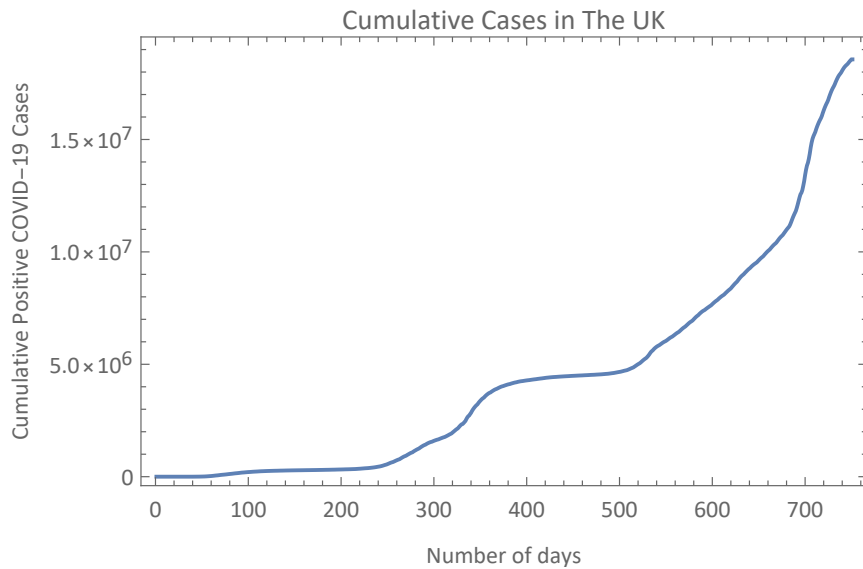


Figure 2: Plot of the cumulative cases in the UK over 750 days

The cumulative plot is a much better plot to use for data fitting as it accounts for all the people who have been infected with the virus. It's a much better indicator to understand if the virus is still growing exponentially or not, even for larger time-frames.

2 Exponential Fitting

2.1 – Methodology

Exponential growth is when there's a rapid growth of quantity of the subject over time. It occurs when the instantaneous rate of change (that is, the derivative) of a quantity with respect to time is proportional to the quantity itself. A nuclear chain reaction and compound interest are 2 very good examples of Exponential growth and radioactive decay is a very good example of exponential decay. The exponential model can fit the initial transmission of the COVID-19 very well as there's a sudden outburst of the virus and there is a rapid spread as there are no measures in the first few weeks to contain the outbreak.

There are various ways to model the data exponentially. The method I have used is to use the equation -

$$A = a * e^{(t-t_0)/\tau} \quad (1)$$

where 'A' is the people affected after time 't', 'a' are the people affected at time 't₀' and 'τ' is the time constant.

```
1
2   defw = de[[41 ;; 63]];
3   cefw = ce[[41 ;; 63]];
4   tefw = Table[j, {j, 41, 63}];
5   dailyefw = Transpose[{tefw, defw}];
6   dailyefwplot = ListPlot[dailyefw, Joined -> True];
7
8   exp = a*Exp[(T - 41)/tau];
9
10  dailyefwfit = FindFit[dailyefw, exp, {a, tau}, T]
11  dailyefwfitplot =
12  Plot[exp /. dailyefwfit, {T, 41, 63}, PlotStyle -> Red];
13  Show[{dailyefwplot, dailyefwfitplot}]
14
15  cumulefw = Transpose[{tefw, cefw}];
16  cumulefwplot = ListPlot[cumulefw, Joined -> True];
17  cumulefwfit = FindFit[cumulefw, exp, {a, tau}, T]
18  cumulefwfitplot =
19  Plot[exp /. cumulefwfit, {T, 41, 63}, PlotStyle -> Yellow,
20  PlotLabel -> cumulative exponential];
21  Show[cumulefwplot, cumulefwfitplot]
22
```

Listing 1: Code written to achieve Exponential Fitting of the data in England

Exponential fitting is by far the most easiest and convenient method to fit the data as it helps us to calculate the most important thing for an epidemic which is the doubling time. The doubling time is vital to understand the severity of the epidemic as it's an indication of how fast the virus is spreading among the population.

Using Equation 1, we can find the formula for doubling time, Let $A=2a$ (because of the doubling) and we can set t_0 as 0 for day 0.

$$A = a * e^{(t-t_0)/\tau} \rightarrow 2a = a * e^{(t-0)/\tau} \rightarrow 2 = e^{t/\tau}$$

$$\text{Now taking log on both sides} \rightarrow \log 2 = t/\tau \rightarrow \log 2 * \tau = t$$

Using my Code, I can find τ of the data provided and multiplying it by $\log 2$ which gives us the doubling time.

2.2 – Results and Data Analysis

Reported Cases of COVID-19 in England

The plot in figure 3 is for the exponential fitting of the reported daily cases between 10th march and 1st april. We can clearly see the fluctuations in the blue line which is the actual data. The daily data is very inconsistent as it was during the early days of pandemic when testing was not readily available. The red line is the exponential fit which very accurately predicts the growth of the pandemic and even though the data is inconsistent, it still fits very accurately. $a = 363.98$, $\tau = 8.97168$, so the doubling time is 6.1 days which is very severe.

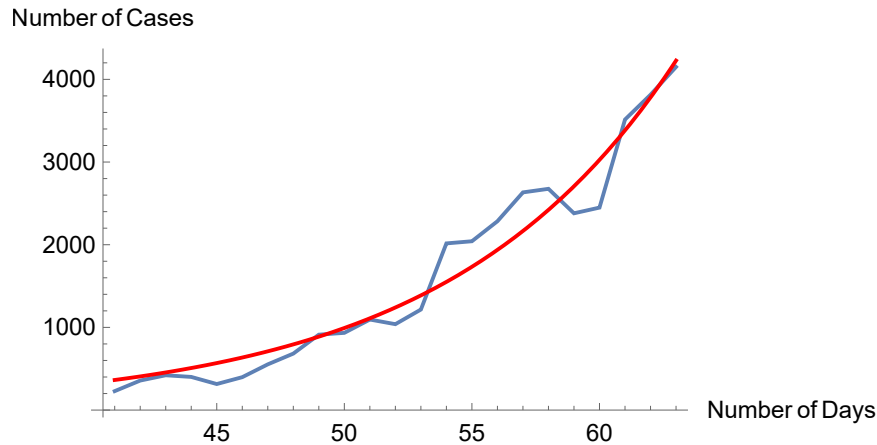


Figure 3: Daily COVID-19 Cases in England between 10th March and 1st April,2020

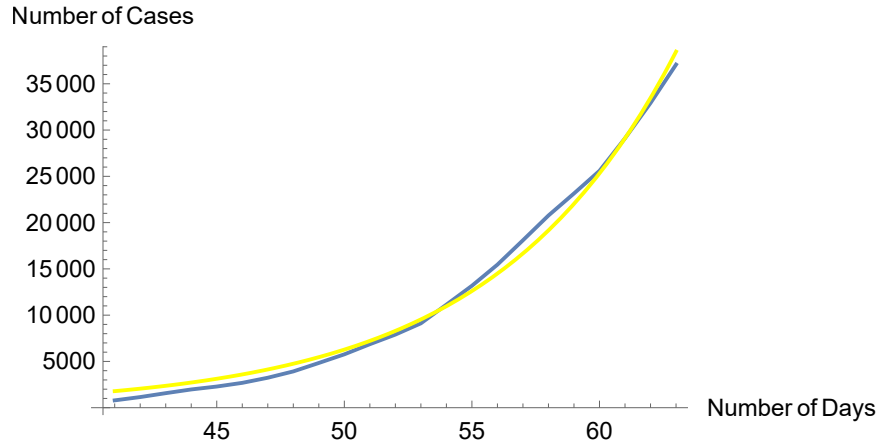


Figure 4: Cumulative COVID-19 Cases in England between 10th March and 1st April

$a = 1788.79$, $\tau = 7.16966$, which means that the doubling time is 4.92 or approx 5 days. However, the doubling time should be the same as daily data, it is not because of the fluctuations in reporting of the data. The cumulative data is a better way to approach the fitting as it evens the data curve. Even in the early days, people started to take precautions themselves before the UK government stepped in with the official restrictions such as masks and lock-downs. This caused the daily cases to be fluctuating so much as there was a drop in numbers during the weekends and numbers picked up again on the weekdays. Thus to smoothen the data and accurately predict better, it's smarter to use cumulative data which gives a far better fit.

Reported Deaths through COVID-19 in England

Deaths are a far more accurate representation of the pandemic because of the lack of testing in the early days. People can refuse to get tested but if their case is severe enough, they'll need to go to the hospital. However, COVID-19 had a very low mortality rate, the large number of people still yield a fairly large number of deaths.

In the future, when testing became readily available, we got even more accurate numbers which gave us an very idea of the severity of the virus.

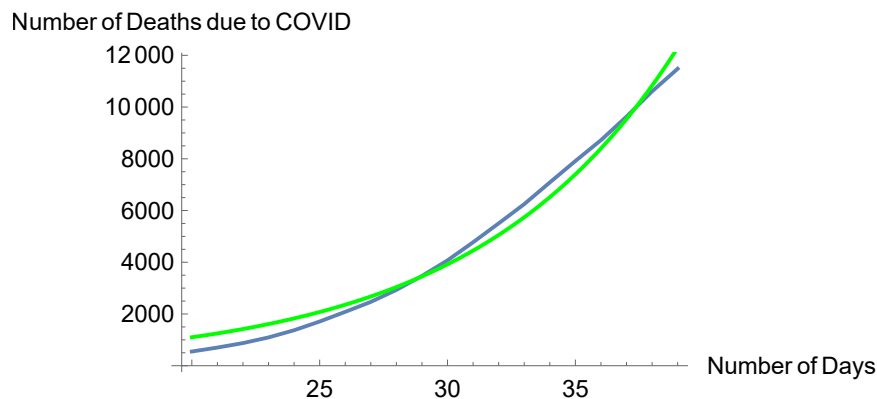


Figure 5: Deaths Due to COVID-19 in England between 21st March and 9th April, 2020

$a = 1100.63$, $\tau = 7.87334$, Doubling time = 5.4 days.

Compared to the cumulative cases, there is going to be a delay in the reporting in the deaths because it takes a few days for the symptoms to appear and then a few more days for them

to get severe enough to cause the unfortunate demise of a patient. However we can see clearly that the death rate had started to increase exponentially even before the caserate because of the lack of testing in the early days. So, we can see that even on the 21st March which is the first day on the plot, there had been 546 deaths due to COVID-19 which were only going to rise from thereafter proving that there was a lack of tests in the early days.

Reported Cases of COVID-19 during the First Wave in Runnymede

Exponential Fitting also works very well for smaller regions and is able to predict the growth accurately. I can choose a larger time frame in the smaller region and the fit is only slightly affected. I've chosen to stop at this date because lockdown was imposed on dsd and its affect start to get visible after 7-10 days. $a = 89.2667$, $\tau = 18.6973$, doubling time = 12 days

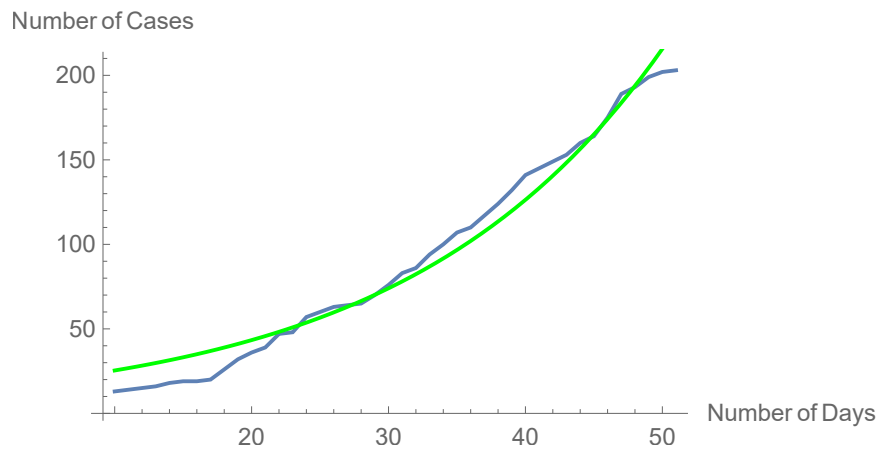


Figure 6: Cumulative COVID-19 Cases in Runnymede between 18th March and 28th April,2020

2.3 – Robustness of the Fits

$a = 0.0289401$, $\tau = 3.34569$, doubling time = 2.3 days $a = 1041.32$, $\tau = 5.27213$, doubling

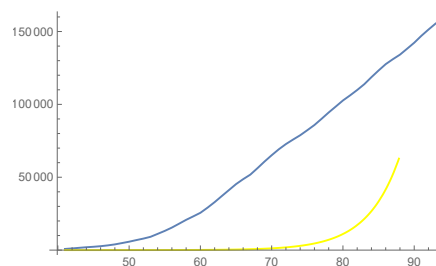


Figure 7: Exponential fitting over 52 days

time = 3.6 days

It is clear that robustness of fits depends on range of days. This proves that Exponential fitting is suitable for smaller range of days (anything less than or around 30 days) and not for larger ranges.

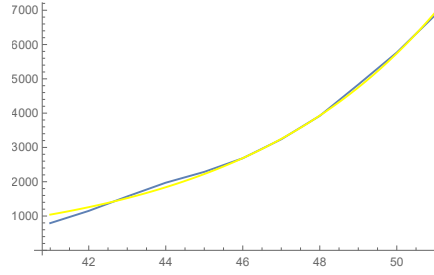


Figure 8: Exponential fitting over 10 days

Limitations of Exponential Fitting

Exponential growth models of physical events can only be applied in certain locations since unbounded development is physically impossible. Although growth may appear to be exponential at first, the modelled phenomenon will eventually enter a region in which previously ignored negative feedback factors become significant (leading to a logistic growth model), the limit of people in this case as the population in England or even in the world is limited so the spread of the virus will have to stop after a point in time. A logistic growth model has an 'S' shape curve which is also called the sigmoidal curve. The Exponential model fails to explain the dynamics of the process of the spread. This is where we need the Avrami Model which is one of the few physical models to actually be able to model epidemics. As materials science models describe successfully the kinetics of overall phase transitions, we can use this in epidemiology too.

3 Avrami Equation

The Avrami equation describes how solids transform from one phase to another at constant temperature. It can specifically describe the kinetics of crystallisation, can be applied generally to other changes of phase in materials, like chemical reaction rates, and can even be meaningful in analyses of ecological systems. The avrami fitting has a sigmoidal curve which is a possible realistic fit for our data.

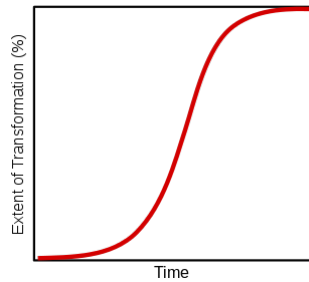


Figure 9: Sigmoidal Curve

The avrami equation is derived when Nucleation takes place at a growth rate. The nucleation has growth front (The nucleation grows with a wavefront). We consider the case when the concentration N_0/N_{total} of the initial number of susceptible cells N_0 exceeds the percolation threshold. The disturbance disseminates along a network of contacts related to the geographical distribution of the population and to the mobility patterns. These involve a combination of local transmissions and long-distance contacts. [reference]

The Assumptions made during the derivation are that -

- Nucleation happens at random and uniformly over the whole untransformed section of the material.
- The pace of growth is unaffected by the degree of transformation.
- Growth takes place at the same pace in all directions.

When these conditions are met, we come to the conclusion that

$$\ln(-\ln[1 - Y(t)]) = \ln(k) + n \ln(t) \quad (2)$$

which allows the determination of the constants n and k from a plot of $\ln \frac{1}{1-Y} \ln \frac{1}{1-Y}$ vs $\ln t$.

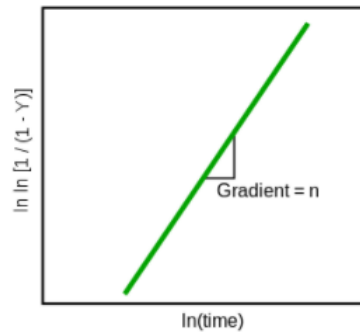


Figure 10: If the transformation follows the Avrami equation, this yields a straight line with slope n and intercept $\ln K \ln K$.

3.1 – Methodology

Avrami equation works best for crystal growth where there is a very small start and particles around it are in liquid form. The growth is very rapid but comes to a halt because of the physical limit imposed by the size of the crystal. That is why when we are talking about epidemiology we need to consider 'Y' as to be the $N0/N_{tot}$ which is the number of cumulative people affected/ population of England. We know that using 2 and using Mathematica, we can plot a loglog-plot which takes log of both x and y axis and plots $\ln(-\ln(1 - Y))$ against $\ln(t)$. I use that to accurately and easily plot the data available to me. Using the function FindFit, all the 3 parameters t_0, s and m are found which are very useful to explain the fitting and can predict the model even further

I take only a very small percentage of the population to accurately fit the data.

```

1  cefw = ce[[41 ;; 63]];
2  logcdata = -Log[1 - (cefw/pe)];
3  cumulelogfw = Transpose[{tefw, logcdata}];
4  cumulelogfwplot = ListLogLogPlot[cumulelogfw];
5  Clear[t, t0, s, m]
6  avramieq = ((t - t0)/s)^m
7  cumulelogfwfit = FindFit[cumulelogfw, avramieq, {s, m, t0}, t]
8  cumulelogfwfitplot = LogLogPlot[avramieq /. cumulelogfwfit, {t,
9  41, 63}];
10 Show[cumulelogfwplot, cumulelogfwfitplot]
```

11
12

Listing 2: Code written to achieve avrami fitting for the first wave in England

3.2 – Results and data analysis

In our equation 2, The most important parameter is 'm' which is related to space. 'm' is the effective dimension in which the spread takes place. Originally, 'm' was assumed to have an integer value between 1 and 4, reflecting the nature of the transformation under consideration. The number 4 may be considered to include contributions from three growth dimensions and one indicating a constant nucleation rate. There are alternative derivations when 'm' has a different value. If the nuclei are prefabricated, and therefore all existing from the start, the transition is exclusively due to the nucleus's 3-dimensional development, and 'm' has a value of 3. When nucleation occurs on certain places (such as grain boundaries or impurities), the transformation quickly saturates. Nucleation may be random at first, and development may be unrestrained, resulting in large values for 'm'. (3 or 4). The creation of new particles will stop if the nucleation sites are depleted. Furthermore, if the distribution of nucleation sites is not random, development may be limited to one or two dimensions. For surface, edge, and point sites, site saturation may result in 'm' values of 1, 2, or 3. 'm' has a very important part to play in our understanding of the spread.

Avrami Fitting of the Reported Cases and deaths in the First Wave of COVID-19 in England

In this plot,

$s \rightarrow 95.1248, m \rightarrow 4.0554, t_0 \rightarrow 27.5433$, We can see that Avrami fits the data perfectly over a range of 20 days. As it is the early days, there are no slowing of slowing down as there is still a large number of population still uninfected by COVID.

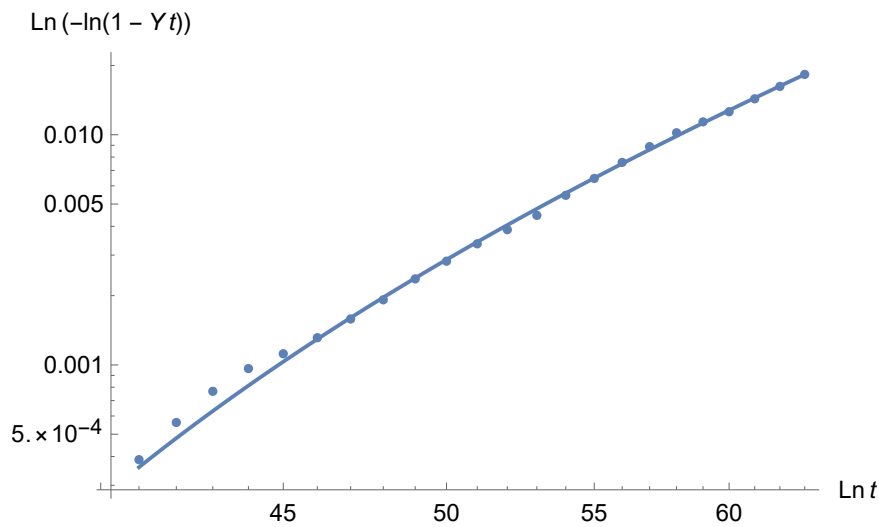


Figure 11: Avrami fitting for COVID-19 cases in England between 10th March and 1st April, 2020

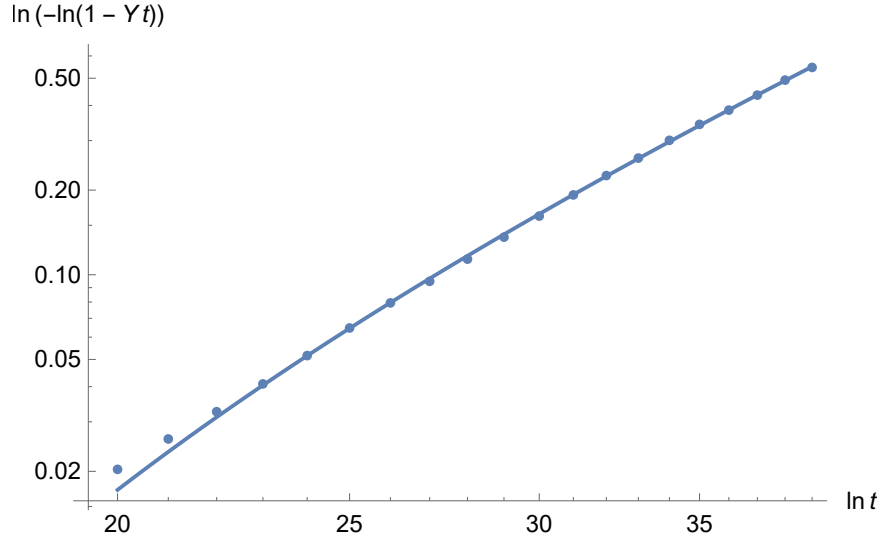


Figure 12: Avrami Fitting of the Reported Deaths in the First Wave of COVID-19 in England between 21st March and 9th April, 2021

In this plot $s \rightarrow 34.7614$, $m \rightarrow 3.23021$, $t_0 \rightarrow 10.1209$. This plot also fits the data perfectly.

3.3 – Robustness of Fits

$s = 658.789$, $m = 2.27555$, $t_0 = 14.294$ when I use 30 percent of the population.

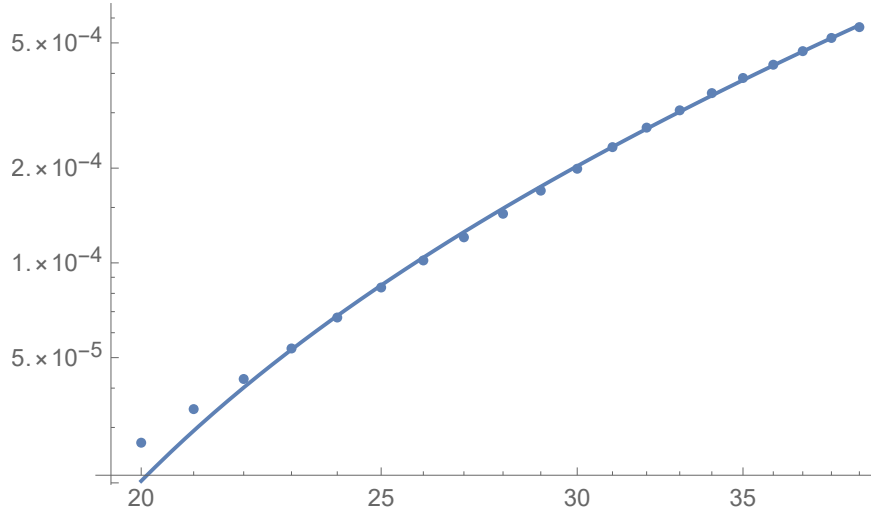


Figure 13: Plot when 30 percent population is taken into account

$s = 34.7183$, $m = 4.1048$, $t_0 = 5.87297$ When I use 0.03 percent of the population.

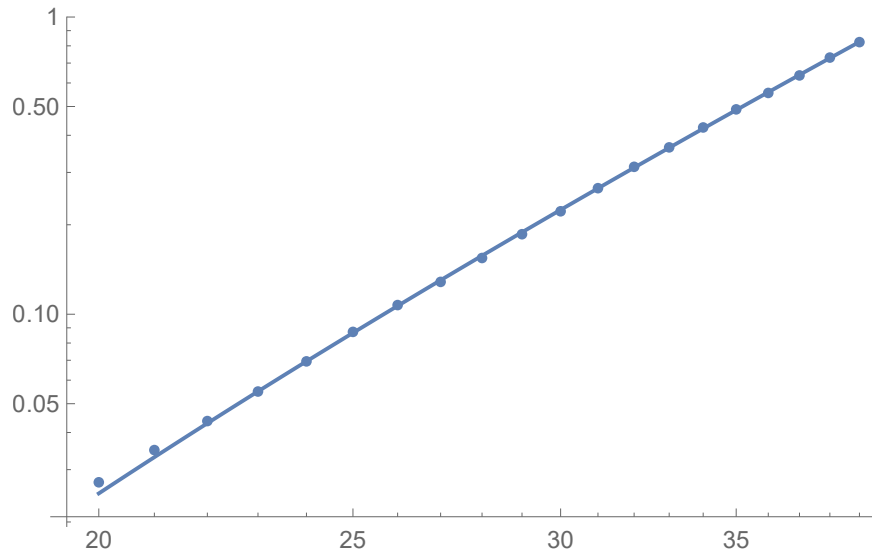


Figure 14: Plot when 0.03 percent of population is taken into account

We can see that Avrami Fitting is Robust even when we make huge changes in the number of population. However, Above 30 percent the plot starts to behave weirdly and produces no results.

4 COVID-19 Omicron Variant

$a = 14275.6$, $\tau = 5.67098$, Doubling time = 3.88 days

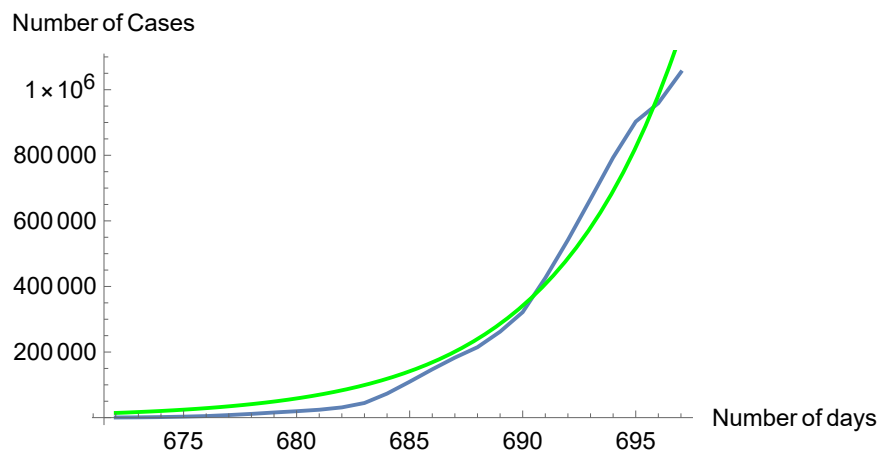


Figure 15: Exponential fitting for the omicron variant in England between 1st December and 26th December

We can see that the doubling time is 3.88 days for Omicron which shows how fast the virus can spread compared to the first wave. Here is when we can use exponential model to understand the rapid spread of the virus.

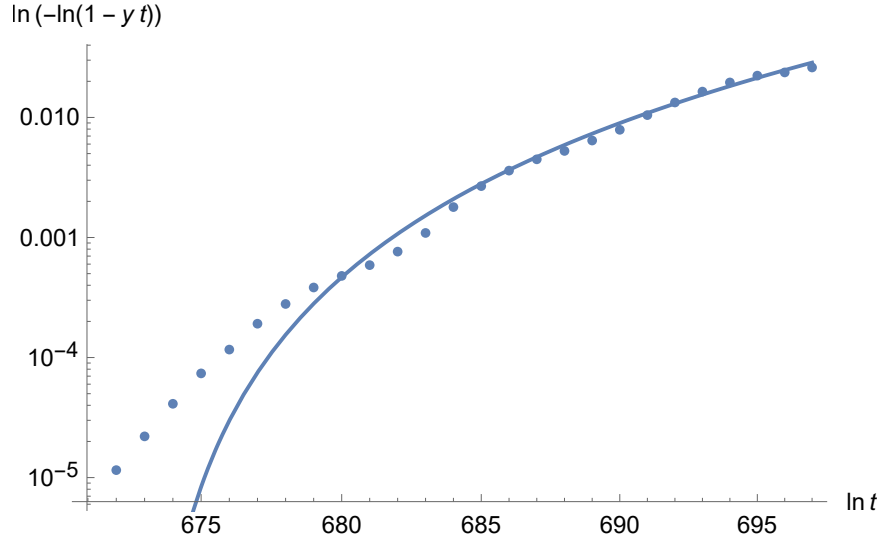


Figure 16: Avrami fitting for the omicron variant in England between 1st December and 26th December

$s = 68.9156$, $m = 3.39651$, $t_0 = 672.795$. The large value of 'm' shows the quick spread of the virus and we can see how it almost starts to flatten at the top indicating that it is very close to affect almost all of the population which can be susceptible to get infected.

5 Conclusion

We can see that in this plot, Avrami fits much more accurately. Exponential Modelling is easier

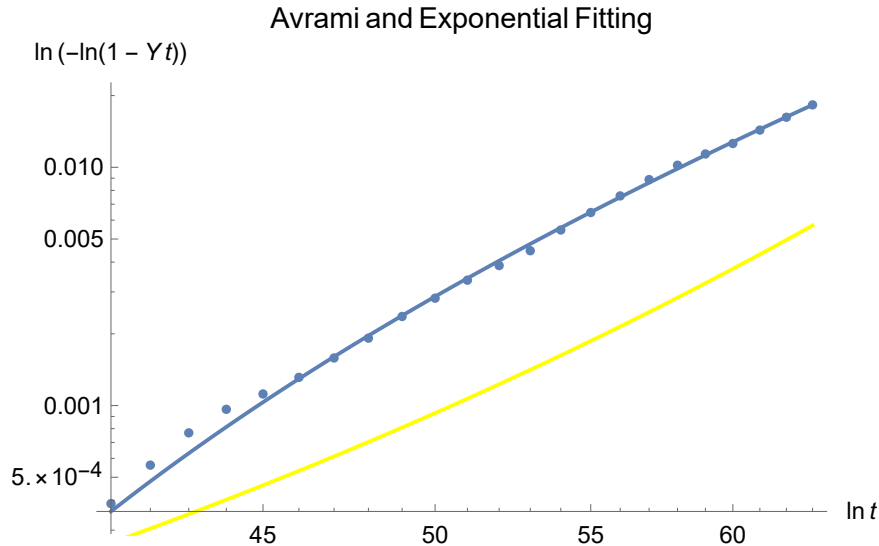


Figure 17: Avrami(Blue) and Exponential(Yellow) fitting for the first wave in England between 10th March and 1st April

to perform while Avrami is harder to do. Exponential Modelling is very good to predict the rapid growth of the virus and is simple data modelling that someone who's not interested to go deeper in the subject will also find easy to understand. However, as there are physical restrictions in the world, the virus has to stop growing at some point. Initial exponential growth often

does not last forever, instead slowing down eventually due to upper limits caused by external factors and turning into logistic growth. A logistic growth model has an ‘S’ shaped curve. This is very well reflected by the Avrami model which uses the same theory as crystallisation of solids. Avrami can accurately predict the spread using not just local transmission but also long distance infection and also takes into the regard the spatial arrangement and offers an analytical solution to the problems. For short time-frames in the initial days, it is better to consider the exponential model as it accurately predicts. However, In the long term and to trul understand the spread, Avrami Modelling might be key.

Furthermore, Data modelling using Avrami might also help us understand when and where some places have provided fake or manipulated data in their reporting of COVID numbers.

6 References

1. <https://www.science.org/doi/10.1126/science.abb4557>
2. https://en.wikipedia.org/wiki/COVID-19_pandemic_in_the_United_Kingdom_First_wave
3. <https://coronavirus.data.gov.uk/details/cases>
4. https://en.wikipedia.org/wiki/Avrami_equation
5. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/615620
6. https://en.wikipedia.org/wiki/Exponential_growth
7. https://en.wikipedia.org/wiki/Logistic_function
8. Avramov, I. (2007). Kinetics of distribution of infections in networks. *Physica A: Statistical Mechanics and its Applications*, 379(2), pp.615–620.