



THE UNIVERSITY of EDINBURGH
School of Physics
and Astronomy

Senior Honours Project

Exploring the Effectiveness of Compartmental Models
to Predict Epidemiological Trends

Craig W. Miller

Abstract

Throughout human history, we have regularly been susceptible to waves of infection that significantly impact our global population and general way of life. In the century between the Spanish flu and Coronavirus pandemics, the field of epidemiology has seen great improvements such that we can create a range of mathematical models to predict outcomes and mechanisms to aid in reducing impacts such as not overwhelming our healthcare systems and protecting those most vulnerable. This report emphasises the importance of using different compartmental models to protect our society. Using data gathered by the Office for National Statistics (ONS)^[1] and The World Health Organisation (WHO)^[2], I explore the continued effects of COVID-19 in the years following the initial outbreak and seek to reproduce the unnatural sustained periodic behaviour seen in Omicron case numbers between June 2022 and June 2023, through the implementation of various compartmental models. The most successful is a spatial SIRS model incorporating cellular automata methods, which estimates the Basic Reproduction number, R_0 , for the Omicron variant to lie between 2.4 and 3.

Declaration: I declare that this project and report is my own work.

Signature:

Date: 29/03/2024

Supervisor: Professor Graeme J. Ackland

10 Weeks

Contents

1	Introduction	2
2	Epidemics	2
2.1	The COVID-19 Pandemic	2
2.2	Importance of Modelling	4
3	Compartmental Models	5
3.1	SIR	5
3.2	SEIR	6
3.3	SIRS	6
3.4	Gamma Distribution	7
3.5	Further model variations	7
4	The Omicron Oscillation Problem	8
5	Implementation	9
5.1	Basic Models	10
5.2	SEIRS Model	10
5.3	Spatial SIRS Model	11
6	Results	12
6.1	SIR and SEIR Models	12
6.2	SIRS Model	12
6.3	SEIRS Model	12
6.4	Spatial SIRS Model	13
7	Discussion	14
8	Conclusions	16
8.1	Potential Improvements	17
8.2	Expanded Model	17

1 Introduction

Four years after the onset of the Coronavirus outbreak in December 2019, we continue to see its lasting effects on the UK population. The World Health Organisation (WHO) recorded a total of 1402 people infected in the week beginning the 3rd of March 2024, this is far from the peak of 1.32 million weekly infections seen with the initial Omicron variant in January 2022 [2]. Naturally, this statistical decline has meant that media prominence and hospital admissions have reduced along with all forms of social restrictions. Consequently, these changes have led to a decrease in the regularity of testing and the frequency of data publication. To combat this issue and ensure data remained consistent, The Office for National Statistics (ONS) carried out a testing survey in which they drew candidates from a cross-section of the UK population to continue with regularly scheduled tests [3]. Amongst the figures collected, a notable trend arises in the form of sustained oscillations in infections and fatalities over a three-month cycle [4][2]. These trends appear unnatural and thus raise the question: How can sustained waves occur even in the absence of new restrictions?

In search of answers, I explore the use of compartmental models in the field of epidemiology to mathematically predict the spread of diseases. I then implement a series of SIR-based models with the aim to recreate the trends shown in the ONS data, whilst simultaneously showcasing the effectiveness of these models and how they can continue to aid in mitigating global impacts for future generations.

2 Epidemics

With references from Hippocrates dating back more than 2.5 millennia, the word "Epidemic" has evolved from the Greek words: *epi*(on) and *dēmos*(people) [5]. It is defined in Miquel's "A Dictionary of Epidemiology" as "*The occurrence in a community or region of cases of an illness, specific health-related behaviour, or other health-related events clearly in excess of normal expectancy.*" [6]. It can also refer to the "first invasion" of a disease that had not previously been recognised in a region. This is expanded with the requirement of specifying the area in which the spread is contained, a "Pandemic" occurs where the epidemic breaches international boundaries and infects a large number of individuals[6][7].

2.1 The COVID-19 Pandemic

In December 2019 the first case of the Coronavirus outbreak was recorded in Wuhan, China, beginning a period that would wreak havoc on our modern civilisation from which we are still recovering 4 years later. The virus responsible named SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus - 2), is closely related to SARS-CoV-1 which caused the 2003 SARS outbreak that infected over 8,000 people globally [8]. In early 2020, a virus identified in horseshoe bats (*Rhinolophus affinis*), RaTG13, was shown to be a 96.1% genome match with the SARS-CoV-2 strain suggesting that the virus likely made a transitional leap from bats to humans upon the origin of the COVID-19 outbreak [9][10]. On the scale of genome diversity, there is still a relatively substantial difference between the two strains. For example, our DNA differs by as little as 1-3% from that of chimpanzees [11], this suggests that the coronavirus transmission did not occur directly and may have instead done so through a yet-to-be-identified intermediate host [12].

SARS-CoV-2 is a positive-sense, single-strand RNA virus (+ssRNA)[10]. Notably, its genome

contains the ORF1ab polyprotein, in addition to the Nucleocapsid (N) and Spike (S) proteins, the presence and/or absence of which can be used to differentiate between mutated variants and define some of their key characteristics. Genome testing is carried out to investigate the appearance of these proteins, in which a sample is only considered positive where more than one is present, individually they are insufficient to diagnose an individual with COVID-19 [1].

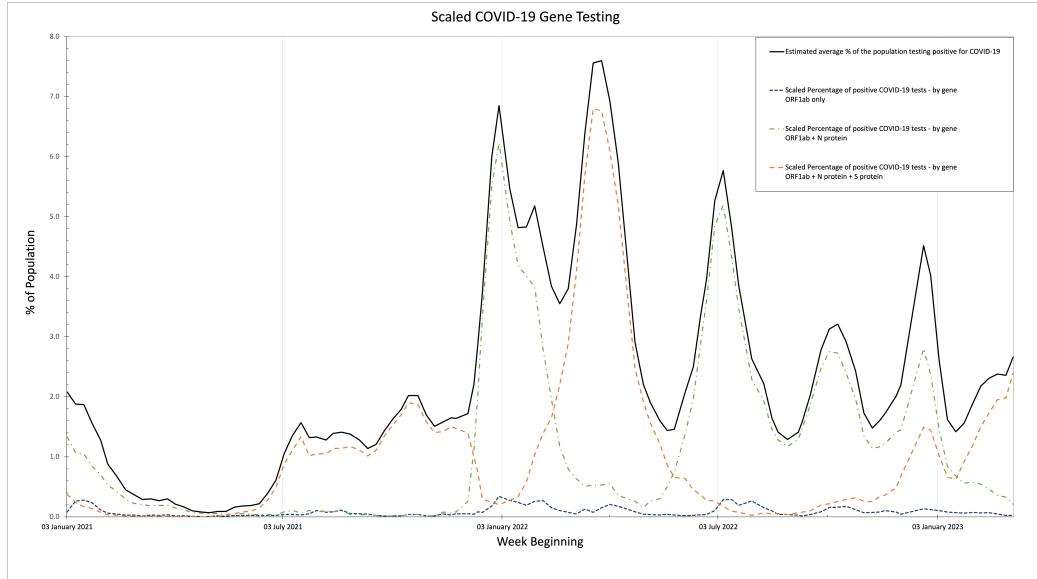


Figure 1: Plot of the estimated percentage of the UK population testing positive for COVID-19 between January 2021 and March 2023. Gene testing results carried out in parallel are scaled against the positive test figures, Strains included are based on the ORF1ab, N and S proteins. The data shows a range of oscillations with a period of roughly 3 months between January 2022 and 2023. Raw data collected by the Office for National Statistics (ONS) [1]

Figures 1 and 2 show the prominence of key variants within the United Kingdom, these periods of dominance are summarised below [4]:

- 'Alpha: B.1.1.7' consists of the ORF1ab and N genes and circulated from mid-November 2020 to November 2021.
- 'Beta: B.1.351 and Delta: B.1.617.2' both contain the ORF1ab, S and N genes - They propagated up until mid-January 2022 with the majority of cases being attributed to the Delta strain.
- 'Omicron: BA.1' consists of the ORF1ab and N genes and was prominent between December 2021 and May 2022.
- 'Omicron: BA.1, BA.4, BA.5' each contain both the ORF1ab and N genes and notably spread between early and mid-June 2022.
- 'Omicron: BA.4, BA.5 and their sub-lineages' compose the ORF1ab and N genes and circulated from late June 2022 onwards.
- 'Omicron: BA.2' consists of the ORF1ab, S and N genes and began circulating in mid-January 2022.

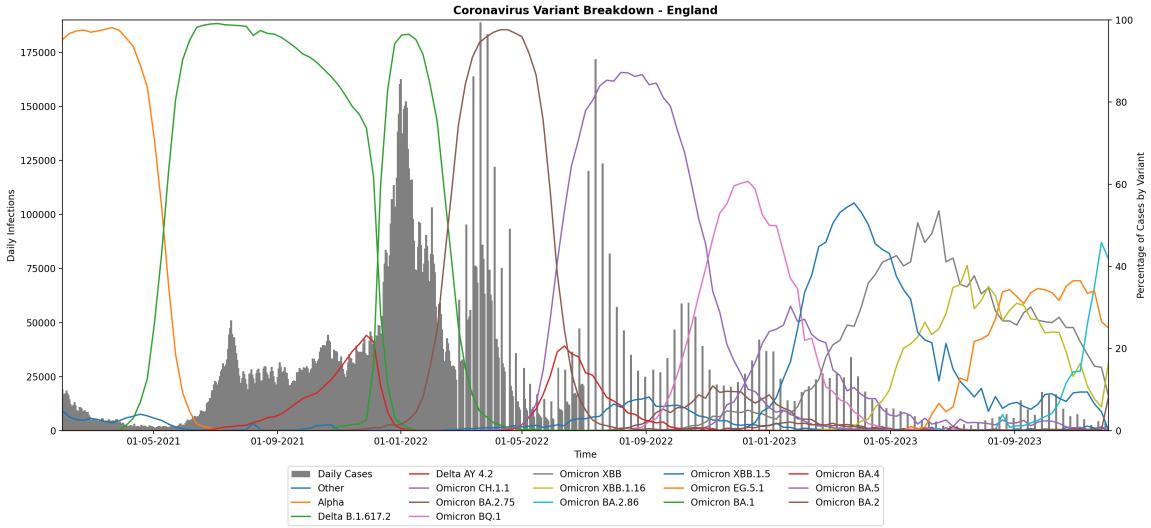


Figure 2: The line data describes a breakdown of the individual Sars-CoV-2 (COVID-19) strains in England and their proportion amongst positive cases between February 2021 and December 2023. The bar plot provides insight into the nation’s daily cases and combined they reveal a strong correlation between new strains emerging and large spikes in infection. This relationship is more pronounced for the early Alpha and Delta variants and fades as the pandemic progresses [4].

2.2 Importance of Modelling

Modern epidemiological models allow us to identify trends in data that can influence how we act to slow infections and mitigate unnecessary societal impact. In March 2020, a team of collaborators from The Imperial College COVID-19 Response Team published their results and predictions on how the pandemic could affect the UK population [13]. This report acted as a key resource for scientific advice that was supplied to the British government to aid in enforcing decisions on public guidance and social restrictions. The study focused on a range of suppression and mitigation strategies detailing the results of using different non-pharmaceutical interventions (NPIs) that could aid in preventing the NHS from becoming overwhelmed based on the estimated surge capacity of ICU beds available.

Suppression tactics use NPIs (and vaccinations once available) to rapidly decrease the rate of reproduction to below 1 and keep measures in place for the duration of the spread. Here, the reproduction number, R , is the average secondary infections generated by infected individuals within a population, this value is localised to regions and changes over time [14]. R is the term frequently used by news outlets during the pandemic to convey the current state of the spread.

Mitigation methods instead attempt to slow infection through the use of NPIs (and again, vaccinations once available) whilst keeping $R > 1$ to build public immunity ultimately leading to an eventual rapid decline in transmission due to a lack of available hosts [13].

The team ran their model for a simulation period of two years and predicted that in the complete absence of interventions, COVID-19 could kill over half a million UK residents [13]. Their results showed that suppression strategies provided the best-case scenario leading to the least anticipated fatalities. This plan included school and university closures, household quarantine for infected individuals and social distancing on a national scale. These measures were shown to be most effective when being triggered on and off based on weekly ICU admissions [13]. However, some studies have

shown that the closure of Schools and Universities led to more fatalities spread over the course of the pandemic which was chosen in order to reduce the peak demand on critical care centres [15].

Throughout the pandemic, COVID-19 has infected as many as 774 million people across the globe with 7.04 million of those ultimately succumbing to the disease (Death Rate: 0.91% as of 15th March 2024) [2]. Although less lethal than other notable pandemics (Bubonic Plague: Estimated as 30%-60% [16]) the scale of infections has reached almost 1-in-10 people globally, thus if a virus of that magnitude had been any more fatal it could have proved catastrophic to society. This shows the importance of modelling the spread of epidemics such that we can act quickly to intervene and reduce preventable fatalities and other global impacts.

3 Compartmental Models

As seen from the study by Imperial College [13], compartmental models are a key mechanism in predicting the effects of epidemics and other outbreaks. There are many different models out there varying in degrees of complexity, with the foundation being the Susceptible, Infected & Resistant model, commonly known as SIR.

3.1 SIR

Introduced by Kermack and McKendrick in 1927, the SIR model is the initial building block of all epidemiological compartmental methods used to predict the propagation of infection within a system [17]. At each time step the model routinely updates each compartment by assessing the following set of ordinary differential equations (ODEs):

$$\frac{\delta S}{\delta t} = -\beta SI \quad (1)$$

$$\frac{\delta I}{\delta t} = \beta SI - \gamma I \quad (2)$$

$$\frac{\delta R}{\delta t} = \gamma I \quad (3)$$

Here, the total population has been normalised ($S + I + R = 1$) such that we are dealing with fractional values of a population and thus knowledge of its exact magnitude is unnecessary, thereby helping to generalise the model to both small and large scales.

β represents the rate at which susceptible people become infected and γ is the rate of recovery for an infected person. By taking the ratio of these two values we obtain the basic reproduction number of a disease, R_0 , which is the number of secondary infections generated by an individual at the start of an epidemic amongst a completely susceptible population. This value is specific to the virus itself and estimations are frequently used in epidemiology to configure models [14][18]:

$$R_0 = \frac{\beta}{\gamma} \quad (4)$$

This is often confused with R , the reproduction number (as introduced in Section 2.2), which unlike R_0 , can accurately be measured locally to describe the current state of viral transmission [14]. When $R \geq 1$, the infection spreads exponentially through the population and the virus can officially be classed as an epidemic [7].

3.2 SEIR

A key aspect of compartmental models is that they can always be expanded to resemble more complex and lifelike systems, an example of this is SEIR which adds an "Exposed" section. This represents the time between contracting an illness and becoming infectious, this is referred to as the latency period, not to be mistaken with the incubation period which refers to the time before the emergence of clinical symptoms [19].

This approach improves on SIR by effectively splitting the infectious group in half to create a distribution curve instead of an immediate exponential decay in terms of recovery probability. Essentially, it models that a person will deteriorate before getting better and that they will tend to be poorly for several days. However, some people will take longer/shorter times to recover based on other factors such as age and immune health. The model is defined by the following system of equations:

$$\frac{\delta S}{\delta t} = -\beta SI \quad (5)$$

$$\frac{\delta E}{\delta t} = \beta SI - \alpha E \quad (6)$$

$$\frac{\delta I}{\delta t} = \alpha E - \gamma I \quad (7)$$

$$\frac{\delta R}{\delta t} = \gamma I \quad (8)$$

Where the new parameter α represents the latency rate such that $1/\alpha$ is the period of time for an individual to transition $E \rightarrow I$.

3.3 SIRS

A disadvantage of these two methods is they assume permanent immunity once someone has recovered, this is true for some viruses such as Chicken Pox and Measles however, it is not a reality for all diseases. Thus, SIRS invokes the idea that once someone has recovered from the virus they are not 'immune' forever and will eventually return to the susceptible group. This is implemented by adding a flow between the resistant and susceptible compartments:

$$\frac{\delta S}{\delta t} = -\beta SI + \mu R \quad (9)$$

$$\frac{\delta I}{\delta t} = \beta SI - \gamma I \quad (10)$$

$$\frac{\delta R}{\delta t} = \gamma I - \mu R \quad (11)$$

Where μ is the rate of immunity loss. A notable feature of this model is that it can ultimately reach an endemic state wherein the overall transitions between compartments become equal and thus the number of individuals in each group reaches a constant value.

3.4 Gamma Distribution

By analysing the compartmental transitions in traditional SIR-based models it becomes apparent that they follow the form of a negative exponential distribution in terms of time to progress, this however gives a maximum probability of recovery on day 0. In reality, a person will tend to be infected for several days before recovering, although this can differ amongst a population depending on factors such as age, sex and immune response. It can instead be modelled more realistically implementing a Gamma (or log-normal) distribution as this provides a peak in the probability at an arbitrary position and has a long tail which accurately describes the variation in recovery time. An example of this is shown in figure 3.

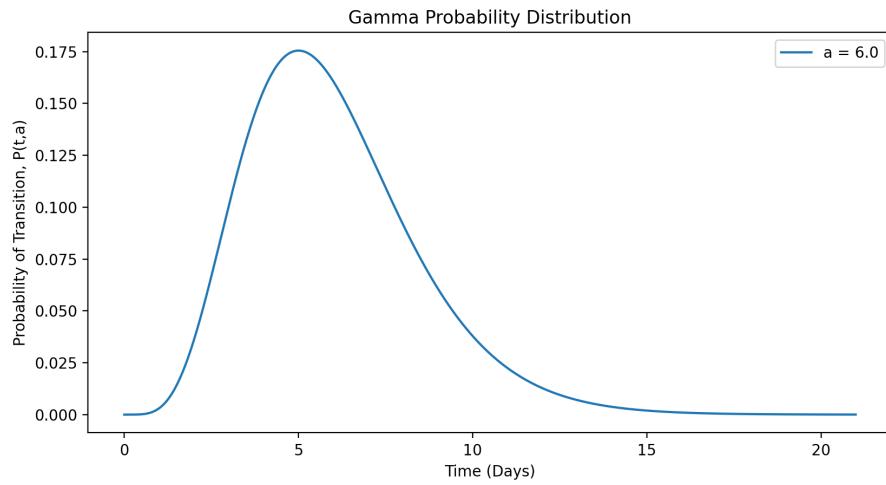


Figure 3: This plot shows a Gamma Probability Distribution function with its peak at $a - 1 = 5$ days (for $a = 6$ days).

The function is defined as follows:

$$p(t, a) = \frac{t^{a-1} e^{-t}}{\Gamma(a)} \quad (12)$$

Where p is the probability a person will transition out of that compartment on a given day, a is the shape parameter where $a - 1$ is the peak of the distribution and $\Gamma(a)$ is used to normalise the distribution:

$$\Gamma(a) = \int_0^{+\infty} t^{a-1} e^{-t} dt \quad (13)$$

The only issue with the gamma distribution is that it asymptotically approaches 0 at $+\infty$. This can be solved by only considering t values within a given confidence range (e.g. 95% or 99%).

3.5 Further model variations

As seen by the evolution of SIR to SIRS an advantage of using compartmental models is that they are relatively straightforward to expand and adapt to situations of greater complexity. This section

discusses several features that could be implemented to enhance the performance of a large-scale model in terms of its effectiveness in simulating real-world situations.

Diseases are known to have varied impacts across demographics such as varying symptom severity and recovery periods, thus it is not entirely accurate to carry out a model that treats the entire population as equal. An example of this observed during COVID-19 arose following large televised sporting events which led to a general increase of infections amongst adult males which in turn led to a spike in females shortly after, followed by children and finally the elderly. This trickle-down effect through the population could be described by splitting the system into its constituent classes such as age or gender. These separate sub-systems each contain the SIR compartments using the unique parameters associated with the groups itself such as a low hospitalisation rate for children. These separated systems can be interlinked through a contact matrix to ensure transmission is properly spread throughout the system and the population is conserved.

As discussed in Section 2, throughout an epidemic, viruses mutate and adapt to form new strains, these new emerging variants lead to a reduction in public immunity. This idea is displayed in Figure 2 where daily COVID-19 infections spike upon the emergence of a novel strain. Thus, it is a key factor to consider when implementing a realistic model.

Another potential factor is the inclusion of social restrictions which can be triggered on and off based on the number of patients in critical care. This idea was used in the Imperial College study to show the effects different measures and trigger values have on the number of estimated fatalities [13]. However, social behaviours can differ amongst people due to many factors including education, religious beliefs and cost of living. Therefore, some may be ignorant of these restrictions, many will listen, yet others will eventually become complacent. Thus, returning to the idea of separating a model into its constituent demographic classes.

In general, basic compartmental models neglect to incorporate a spatial factor consequently meaning that transmission can occur equally between any infected-susceptible pair. Whereas in reality, a person in the UK cannot physically infect someone halfway across the planet, to implement this idea the concept of cellular automata can be explored. It must be noted that diseases differ in transmission mechanisms hence depending on this factor, the model could treat this proximity measure in a variety of ways.

A perfect approach would include all of the above social, biological and geographical features, however, in practice, this would prove computationally expensive and the benefit to the overall results would likely taper off quickly. Hence, a method should focus on the main aspects that provide a larger impact on the progression of infection and rely less on the more redundant features as doing so can still return a highly accurate representation of epidemic transmission without over-fitting the known data.

4 The Omicron Oscillation Problem

As first identified in Figure 1, data gathered by the ONS displays a clear set of sustained oscillations in COVID-19 infections attributed to the Omicron sub-variants between 2022 and 2023. The near-constant period between waves of three months appears unnatural at first and requires further investigation. Unfortunately, the ONS terminated their publishing of data in March 2023 making it unclear if the trend continues beyond this time frame.

However, focusing on the period of these oscillations, data obtained from WHO (displayed in Figure 4) shows a comparison of deaths and infections in the United Kingdom with weekly cases

taking the form of a damped oscillator whereas fatalities hold relatively consistent in their amplitudes. This contrasts with the basic intuition that fewer cases would suggest fewer deaths. The decrease in the amplitude of cases can likely be attributed to a decrease in testing frequency after the government scaled back on free Lateral Flow Testing (LFT) for low-risk individuals beginning on April 1st 2022 [20]. Official deaths do not have this issue as they are registered daily thus staying constant.

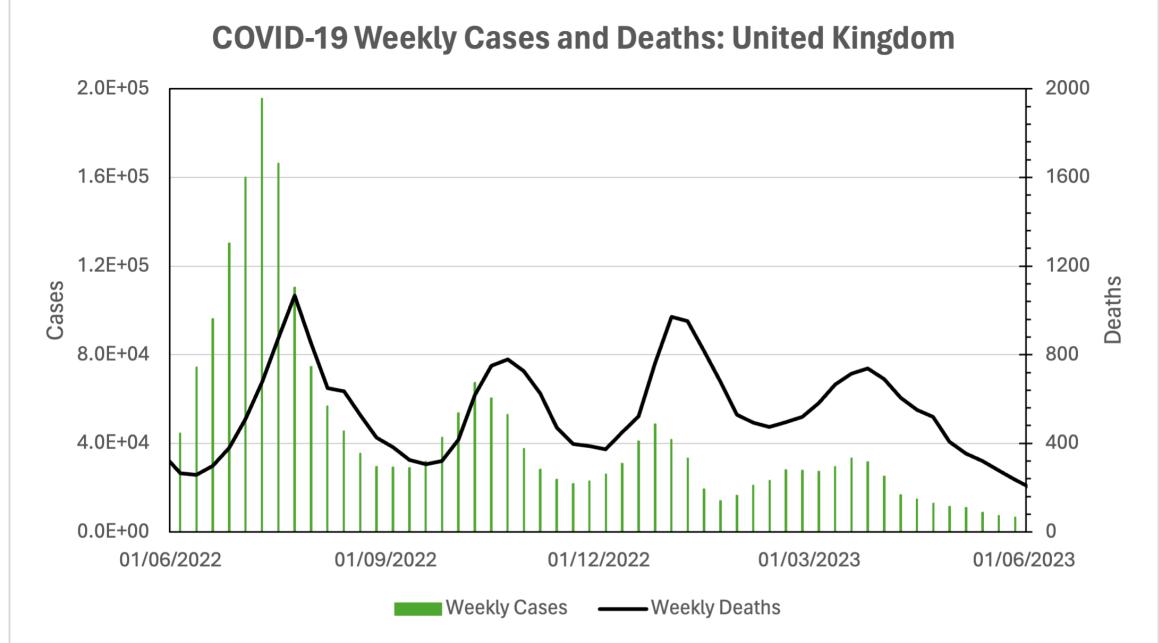


Figure 4: Weekly COVID-19 infection and death data among the UK population from June 2022 to June 2023. This timescale follows after the main Alpha, Beta and Delta waves and focuses on the later Omicron variants. A clear oscillation pattern arises in both sets of data with death tolls peaking roughly 2 weeks after infections which comes as expected as people will not pass away immediately. The period of these oscillations is roughly 3 months and each aligns well with the emergence of new omicron sub-variants (see Figure 5)

Reflecting on Figure 2 it was shown that the peaks in infections occurred in close correlation with the emergence of new viral mutations, this is likely due to a reduction in public immunity to the new strain, providing more susceptible hosts to infection. Contrary to this, Figure 5 shows less of a direct relationship between new strains when only focusing on the time frame of the oscillations, suggesting there could be another factor driving this behaviour.

5 Implementation

The main goal to reproduce the oscillatory trends in Omicron data (as displayed in Figure 4) has been carried out through the implementation of a series of SIR-based Python models. This section

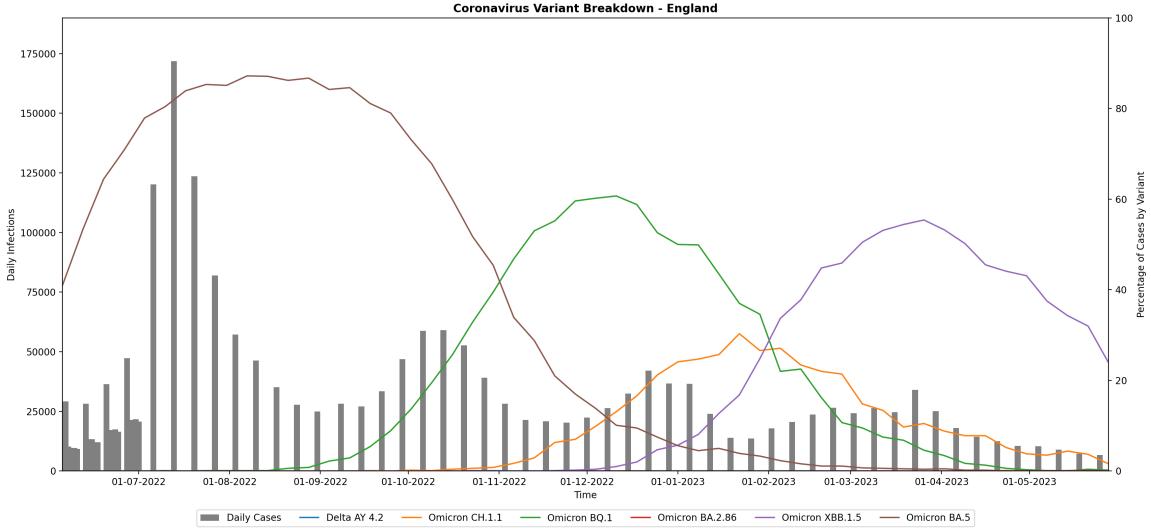


Figure 5: Zoomed and enhanced Figure 2 for Daily COVID-19 infection and variant prominence data among the UK population but focused on the months between June 2022 and 2023. Other present variants with negligible impact have been removed for clarity (For the full set of data see Figure 1)

focuses on the key logic and design choices that were made to produce the most effective and efficient results possible.

5.1 Basic Models

To effectively conclude which model best replicates the data, it was decided that it would be beneficial to implement each of the models described in Section 3. The inclusion of these would allow for further comparison and the ability to track the effects of implementing each improvement upon the prior method. The SIR, SEIR and SIRS models were constructed using their respective sets of ODEs as outlined in Section 3.

5.2 SEIRS Model

Building on the previous incarnations incorporating a combination of the SEIR and SIRS methods, the SEIRS model was also carried out [21]. This approach used the following family of ODEs to update the state of the system upon each iteration:

$$\frac{dS}{dt} = \mu - \beta SI + \omega R - \mu S \quad (14)$$

$$\frac{dE}{dt} = \beta SI - \sigma E - \mu E \quad (15)$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \mu I \quad (16)$$

$$\frac{dR}{dt} = \gamma I - \omega R - \mu R \quad (17)$$

Where the transition parameters are defined as [21]:

- β : Infection Rate $\Rightarrow \frac{1}{\beta}$: Time to transition $S \rightarrow E$
- σ : Latency Rate $\Rightarrow \frac{1}{\sigma}$: Time to transition $E \rightarrow I$
- γ : Recovery Rate $\Rightarrow \frac{1}{\gamma}$: Time to transition $I \rightarrow R$
- ω : Immunity Loss Rate $\Rightarrow \frac{1}{\omega}$: Time to transition $R \rightarrow S$
- μ : Birth and Death rate

Using these values the basic reproduction number, R_0 , can be determined using the following relation [21]:

$$R_0 = \frac{\sigma}{(\sigma + \mu)} \cdot \frac{\beta}{(\gamma + \mu)} \quad (18)$$

This model was successfully implemented in Python using Numpy arrays to store the fractional value of the population within each compartment at any given time step [22].

5.3 Spatial SIRS Model

This final method was performed by applying the concepts of cellular automata to update the system using the SIRS basis. Within the model, an NxN system (Numpy array) is initialised such that each node has four nearest neighbours (Top, Bottom, Left, Right). It was decided that periodic boundary conditions must be put in place such that nodes at edges are treated as the neighbours of those at the opposite side, thus eliminating discontinuities and creating a seamless system. Each individual is represented by a 0, 1 or 2 in the array, signifying the three possible states: Susceptible, Infected and Resistant. The initial distribution of these states is set to be equal in quantity, yet random in position (i.e. Values for each node are randomly assigned with a 33.3% probability of being 0, 1 or 2).

Updates are carried out stochastically over many sweeps, where a sweep is defined as N^2 iterations. Upon each iteration, the model selects a random position to update, this is done so that each sweep provides all individual nodes with an equal chance of being selected however, some may be chosen more or less than once. Once a node is chosen, it is assessed on whether to transition to the next state based on the following set of rules:

- Susceptible nodes (0) update with probability, p_1 , if at least one of its nearest neighbours is infected.
- Infected nodes (1) will recover with probability, p_2 .
- Resistant nodes (2) will lose immunity with probability, p_3 .

Here, the respective transition probabilities are given upon initialisation of the model ($p_1 = P(S \rightarrow I)$, $p_2 = P(I \rightarrow R)$, $p_3 = P(R \rightarrow S)$).

6 Results

With all models successfully implemented in Python, they can be used to test and compare their effectiveness at reproducing sustained oscillations (The Omicron Oscillation Problem). An important idea to ensure close and easy comparison was kept between the models and data, was that each method must be run using parameters accurate to literature-based estimations of COVID-19 characteristics [13][14]. The values of these vary slightly from model to model to provide a deeper insight into the uncertainty of these estimations [14].

6.1 SIR and SEIR Models

The SIR and SEIR models cannot produce oscillations as to do so individuals must cycle through the population repeatedly for a wave-like state to propagate. Figure 6 shows the results for both of the models using the same COVID-19 parameters as the SEIRS approach.

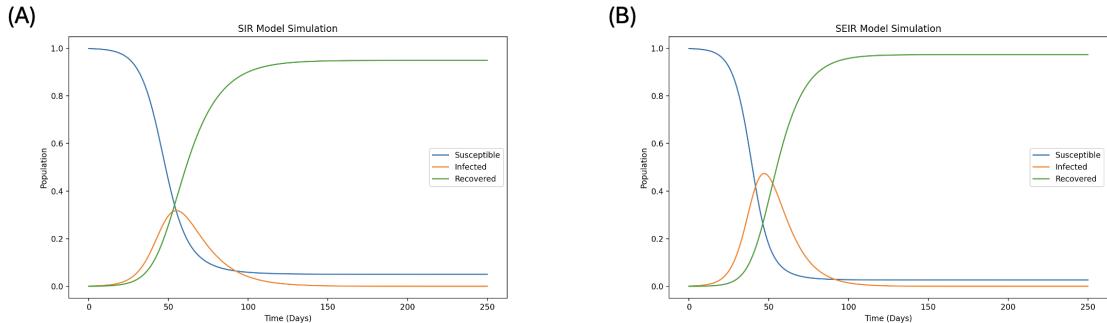


Figure 6: The results for the (A)SIR and (B)SEIR models are displayed both of which produce a single peak in infections at around 50 days. The parameters used for SIR were: $P(S \rightarrow I) = 0.218/\text{day}$; $P(I \rightarrow R) = 0.071/\text{day}$. SEIR used the following: $P(S \rightarrow E) = 0.218/\text{day}$; $P(E \rightarrow I) = 0.182/\text{day}$; $P(I \rightarrow R) = 0.091/\text{day}$

6.2 SIRS Model

Unlike the previous two models, SIRS can generate oscillations in the form of damped waves, shown in Figure 7, the spread rapidly reaches an endemic state after only 2 vastly different infection peaks. Due to an insufficient number of peaks and the wide disparity in their amplitudes, it is not necessary to compare this result with the ONS data.

6.3 SEIRS Model

Optimal parameter values must be chosen to reflect the unique characteristics of COVID-19. Imperial College's Report 9 estimated the basic reproduction number to be around 2.4 [13]. The total latency and recovery periods vary and were chosen to be 5.5 and 11 days respectively [23]. For ease, birth and death rates were chosen to be equal to ensure the population was conserved, this value was set as the average life expectancy in the UK (80.9 years [24]). Applying these estimated values to equation 18 we can solve to find $\beta \simeq 0.218/\text{day}$.

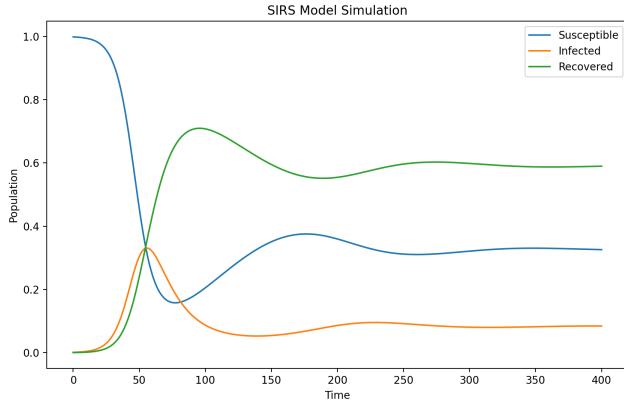


Figure 7: The results for the SIRS model display the behaviour of a damped oscillator that quickly reaches an endemic state after only 2 oscillations thus there is insufficient data to estimate an accurate period for the wave. The parameters used were: $P(S \rightarrow I) = 0.218/\text{day}$; $P(I \rightarrow R) = 0.071/\text{day}$; $P(R \rightarrow S) = 0.01/\text{day}$

The SEIRS model was then executed using the above parameters producing the results shown in Figure 8. The plot shows that the model is successful in producing wave-like trends in the data however these functions correspond to a damped oscillator and eventually reach an endemic state with a total of 2.54% of the population living with the disease. The period of these oscillations is around 420 days. Although the model has been able to produce oscillations they are not sustained and the magnitude of the period is far beyond the expected value of 3 months ($\simeq 90$ days). This shows that the SEIRS model lacks the capacity to replicate real infection data to an acceptable accuracy.

6.4 Spatial SIRS Model

The spatial model can produce three distinct behaviours: These are Absorbing, Wave and Dynamic equilibrium states:

1. In an absorbing state, any infection will die out rapidly
2. In a wave state, continuous oscillations will be produced
3. In dynamic equilibrium, the system reaches an endemic state where the transition rates between compartments are equal.

To optimise the model and produce the best results possible, it was run for a fixed value of $p_2 = 0.1$ which corresponds to an infection period of 10 days and all combinations of p_1 and p_3 in the range of 0 to 0.35. The model ran each parameter set for 5000 sweeps (days) and calculated the average infected fraction of the system across that time frame. This produced the heat map shown in Figure 9, where the parameters associated with each of the three states are denoted. By running the spatial model for parameters lying within each defined section of the heat-map plot, results displaying the three distinct behaviours can be produced.

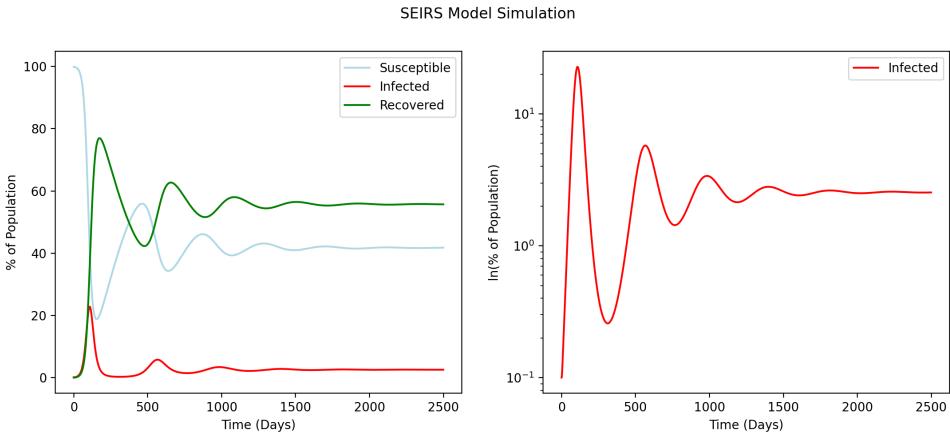


Figure 8: Results for the SEIRS model are plotted against time. The Exposed and Infected results have been combined here to improve the clarity of the oscillations. The plot on the right uses a log scale to focus on the small-scale oscillations of the Infected compartment. The parameters used in this run were: $\beta = 0.218/\text{day}$; $1/\gamma = 11 \text{ days}$; $1/\mu = 80.9 \text{ years}$; $1/\omega = 1 \text{ year}$; $1/\sigma = 5.5 \text{ days}$

With aid from the results in Figure 9, the model was then executed again for fixed values of $p_2 = 0.1$ and $p_3 = 0.01$ while varying p_1 to reflect a range of R_0 values ($R_0 = \{2.4, 3, 4, 5, 8\}$) (Based on Alpha and Omicron estimates [13][25][14]). These values lie on the boundary of region II (Figure 9) where we can expect the production of wave states in the system and refer to minimum recovery and immunity periods of 10 and 100 days respectively [26]. Both recovery (10-120 days [27]) and immunity (3 months - 5 years [28]) rates have been shown to vary greatly due to differences in the severity of cases and strength of an individual's immune response, thus the minimum value was chosen for both to keep consistency. The results of the model for these parameter sets are shown in (Figure 10) accompanied by key frames from the animated progression of the system in the case of $R_0 = 8$ (Figure 11).

7 Discussion

As expected the basic SIR and SEIR models lack the complexity needed to produce any oscillatory behaviour. SIRS is the first method that is shown to produce damped waves this is then improved upon in the SEIRS approach where more prominent peaks in infection are apparent and it takes much longer to reach an endemic state (enhanced on a log scale in Figure 8). The spatial method has proven to be the only approach capable of producing the sustained wave-like behaviour seen from the Omicron cases and it has even been able to do so for a wide range of basic reproduction estimates.

Estimations of the COVID-19 R_0 value tend to lie in the range of 2-3 [13][25] but have been predicted as high as 4.9 [14]. Figure 10 shows that the amplitude of waves increases and the period decreases as we raise the R-number (R_0) in the spatial model. The trends seen in the Omicron data (Figure 4) show a consistent period of around 3 months (≈ 90 days) affecting between 3 and 8% of the population at a time (Figure 11). These values are likely higher due to reduced testing consistency and frequency, thus it is assumed that the maximum bound of 8% is a more accurate representation of the virus' spread [20].

Table 1 shows the analysis of the resultant data from the Spatial SIRS model (Figure 10). It is

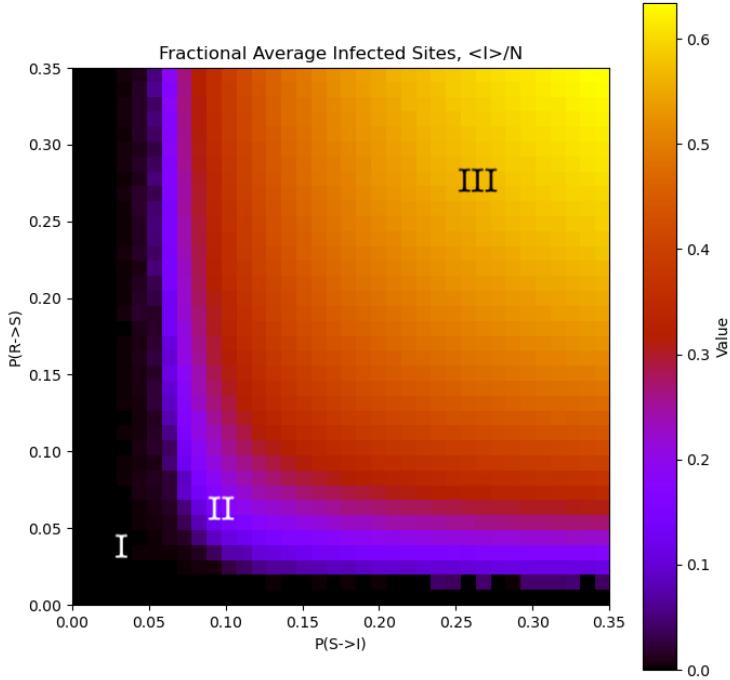


Figure 9: This heat-map displays the average infected fraction of nodes in the spatial SIRS model for varying values of p_1 and p_3 with p_2 fixed at 0.1 ($p_1 = P(S \rightarrow I)$; $p_2 = P(I \rightarrow R)$; $p_3 = P(R \rightarrow S)$). Three distinct sections are denoted by: I, II and III - representing the Absorbing, Wave and Dynamic Equilibrium states respectively.

apparent that the higher the R-number is raised the more prominent and defined the infection peaks will be. The peak values for $R_0 = 2.4 \& 3$ lie within a 1σ and 2.38σ range respectively of the expected value of 8%, there is a huge variation in the period from $R_0 = 2.4$ to 3 suggesting that the true value may lie closer to 3. However the period of these oscillations is much longer than 90 days, this can be ignored for now as, to reduce this period the R-number would either have to be increased far above 8 or the other parameters in the system would need to be further optimised. Thus we can make the statement that an estimate of the true R_0 value for the Omicron variant lies near 3 although further testing and model improvements should be implemented to reduce this range of uncertainty. This idea that the R-number is larger than originally thought can be reinforced by the fact that the ONS survey data frequently recorded more cases than the official government records at the time which suggested a larger R_0 than was widely assumed [14].

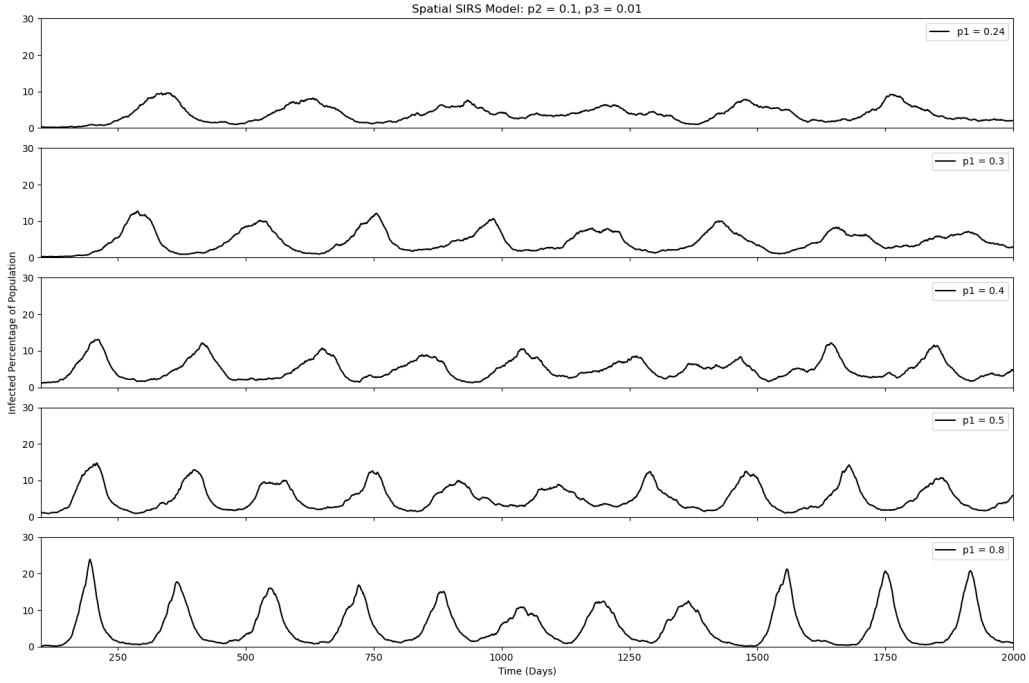


Figure 10: Combined plots of the Infected fraction of the population over time for varied $p_1 = P(S \rightarrow I) = \{0.24, 0.3, 0.4, 0.5, 0.8\}$ to reflect the performance of the Spatial SIRS model on a range of R_0 values. This result was produced for a parameter set of: $N = 100; p_2 = P(I \rightarrow R) = 0.1; p_3 = P(R \rightarrow S) = 0.01$. The data shows that sustained oscillations are produced for each of the parameter sets and displays that as the infection rate is increased the period of these oscillations decreases and their amplitude increases.

8 Conclusions

The effectiveness of compartmental modes has been demonstrated by the research team at Imperial College where their contributions and results allowed government interventions to save over 275,000 lives in the UK alone throughout the COVID-19 pandemic (233,791 confirmed fatalities [4] compared to the 510,000 initially predicted for an unhindered pandemic [13]). Without the use of these compartmental models to predict impacts, the suppression strategies used would have never been as successful and would certainly have led to countless avoidable deaths.

The approaches used in this report to reproduce the real data seen beyond the 2-year range of Imperial's simulation reinforce the value of this methodology and its capability of adapting to many situations [13]. The cellular automata-based spatial SIRS model has successfully demonstrated the oscillatory behaviour shown during the Omicron waves of COVID-19 and has also been able to estimate that the R_0 value for this family of sub-strains lies higher than previously thought. It has been shown to a 2.38σ range that the R_0 lies at 3, thus the method has been effective at reproducing the real ONS data further underlining the importance of using compartmental models to paint a clearer picture of the spread of epidemics and increasing our preparedness in order to reduce mortality and further global impacts.

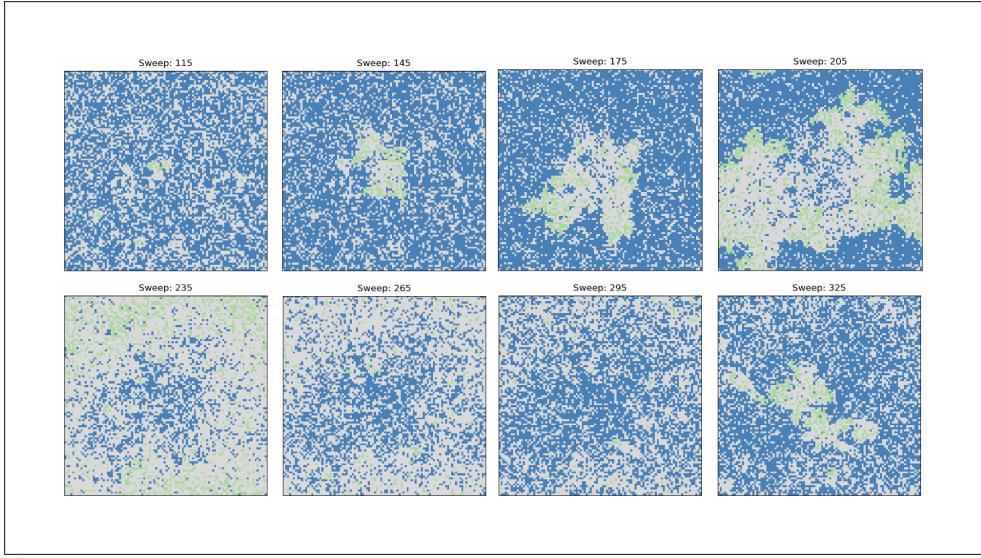


Figure 11: The plots show a complete cycle of a wave of infection spreading through a system which repeats leading to periodic peaks in infection data (See Figure 10). The blue, green and grey nodes represent Susceptible, Infected and Resistant individuals respectively. This result was produced for a parameter set of: $N = 100$; $p_1 = P(S \rightarrow I) = 0.8$; $p_2 = P(I \rightarrow R) = 0.1$; $p_3 = P(R \rightarrow S) = 0.01$. The cycle starts with a mainly susceptible system with a couple of small areas of infection, these areas slowly develop into large waves and propagate across the population. The infection eventually slows down due to a lack of susceptible hosts causing resistance to rise, eventually, infection dissipates and leads the model back towards its initial heavily susceptible state.

8.1 Potential Improvements

The main downside of the spatial SIRS model is that the period of oscillations does not match with real data. To reduce this discrepancy, the spread of infection could be increased, one approach to this is raising the number of nearest neighbours that the system checks. Currently, the model assesses the four nearest neighbours of a node, by instead considering eight (3x3 grid with the chosen node at the centre) along with scaling the probability of infection by the number of infected neighbours so faster propagation could be achieved. These changes would likely reduce the oscillatory periods closer to the real value of 3 months without having to change the initial R_0 of the system.

8.2 Expanded Model

Using the logic gained from the progression of SIR to SEIRS we could continue to expand the capacity of the model to resemble a more realistic situation, this is frequently used in epidemiology [14][13]. Figure 12 shows a potential approach to this, where there are seven different compartments and stages of the disease all of which are interlinked using unique probabilistic parameters. A further possible improvement to this would incorporate both approaches using the spatial characteristic from the cellular automata SIRS method and the gamma distribution-based transition rules defined in Section 3. These adaptations would improve the realism of the results obtained leading to a more accurate way of predicting impacts and outcomes of future outbreaks.

R_0	Mean Period (Days)	Mean Peak Infected Population (%)	Peak Error (%)
2.4	316.7	8.00	0.43
3	237.5	9.71	0.72
4	202.2	10.48	0.59
5	185.8	11.69	0.61
8	172.7	16.79	1.27

Table 1: Analysis of results from Figure 10 where R_0 is calculated from $\frac{p_1}{p_2}$. The values were calculated by taking the time value at each peak and averaging over the number of full oscillations in the time frame. A single oscillation was determined to end where its tails become flat before rising again. The error in the peak values is taken to be the standard error on the mean.

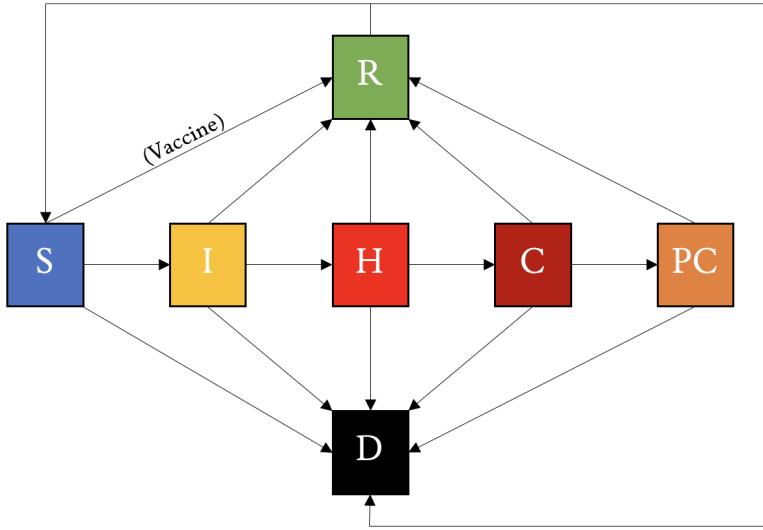


Figure 12: This Figure shows an expanded version of the SIR model. The compartments shown are as follows: S - Susceptible, I - Infected, H - Hospitalised, C - Critical Care (ICU), PC - Post Critical, R - Resistant, and D - Deceased. The arrows show the possible transitions between each compartment each having a unique probability.

References

- [1] E. Fordham and M. Bracher. COVID-19 Infection Survey Dataset: UK. *Office for National Statistics*, 2023.
- [2] World Health Organization 2023. WHO Coronavirus (COVID-19) dashboard. Date Accessed: 15th March 2024.
- [3] K. Steel and E. Fordham. Coronavirus (COVID-19) Infection Survey: methods and further information, 2023.

- [4] UK Health Security Agency. Coronavirus (COVID-19) in the UK. Date Accessed: 24th January 2024.
- [5] M. V. Martin and E. Martin-Granel. 2,500-year Evolution of the Term Epidemic. *Emerging Infectious Diseases*, 12:976–980, 2006.
- [6] M. A. Porta. *A Dictionary of Epidemiology (5th ed.)*. Oxford University Press, 2008.
- [7] R. C. Dicker et al. *Principles of Epidemiology in Public Health Practice (3rd ed.) - An Introduction to Applied Epidemiology and Biostatistics*. Centers for Disease Control and Prevention (CDC), 2012.
- [8] J. W. LeDuc and M. A. Barry. SARS, the First Pandemic of the 21st Century. *Emerging Infectious Diseases*, 10:11, 2004.
- [9] P. Zhou et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579:270–273, 2020.
- [10] P. V'kovski et al. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*, 19:155–170, 2021.
- [11] M. V. Suntsova and A. A. Buzdin. Differences between human and chimpanzee genomes and their implications in gene expression, protein functions and biochemical properties of the two species. *BMC Genomics*, 21:535, 2020.
- [12] D. Cyranoski. Mystery deepens over animal source of coronavirus. *Nature*, 579:18–19, 2020.
- [13] N. M. Ferguson et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College London*, 2020.
- [14] R. Anderson and C. Donnelly et al. Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation. *The Royal Society*, 2020.
- [15] K. Rice et al. Effect of school closures on mortality from coronavirus disease 2019: old and new predictions. *British Medical Journal*, 371, 2020.
- [16] World Health Organization 2023. Plague - Key facts. Date Accessed: 15th March 2024.
- [17] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, 115:700–721, 1927.
- [18] N. Imai et al. Report 3: Transmissibility of 2019-nCoV. *Imperial College London*, 2020.
- [19] L. Rimbaud and S. Dallot et al. Assessing the Mismatch Between Incubation and Latent Periods for Vector-Borne Diseases: The Case of Shaka. *Phytopathology*, 105:1408–1416, 2015.
- [20] UK Health Security Agency. Guidance COVID-19: testing from 1 April 2024, 2024.
- [21] O. N. Bjørnstad et al. The SEIRS model for infectious disease dynamics. *Nature Methods*, 17:557–558, 2020.
- [22] C. R. Harris et al. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020.

- [23] H. Xin et al. Estimating the Latent Period of Coronavirus Disease 2019 (COVID-19). *Clinical Infectious Diseases*, 74:1678–1681, 2022.
- [24] J. Buxton. National life tables – life expectancy in the UK: 2020 to 2022, 2024.
- [25] G. J. Ackland et al. Fitting the reproduction number from UK coronavirus case data and why it is close to 1. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 380:2233, 2022.
- [26] M. S. Al-Thaqafy et al. Factors Affecting Confirmed COVID-19 Patient's Recovery Time at King Abdulaziz Medical City, Jeddah. *Cureus*, 15, 2023.
- [27] B. Liu and D. Jayasundara et al. Whole of population-based cohort study of recovery time from COVID-19 in new south wales australia. *The Lancet Regional Health - Western Pacific*, 12:100193, 2021.
- [28] J. P. Townsend and H. B. Hassler et al. The durability of immunity against reinfection by SARS-CoV-2: a comparative evolutionary study. *The Lancet Microbe*, 2, 2021.

Acknowledgements

I would like to thank my supervisor Prof. Graeme Ackland for his advice throughout the project and notably his recommendation of delving into the data trends in Omicron case numbers which made up one of the key aspects of this research project.

I would also like to acknowledge Prof. Davide Marenduzzo, the Course Organiser for the Senior Honours course 'Modelling and Visualisation in Physics' at the University of Edinburgh where inspiration for the Spatial SIRS model used in this project originated.

Appendix A

The code and all other associated files for this project are available to view at: https://github.com/craigmiller24/SH_Project

The datasets used in this project can be viewed and downloaded from the following pages:

1. Office for National Statistics COVID-19 Infection Survey [1]: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionssurveyleadlineresultsuk/2022>
2. World Health Organisation COVID-19 Data [2]: <https://data.who.int/dashboards/covid19/data>
3. UK Health Security Agency [4]: <https://coronavirus.data.gov.uk/details/cases?areaType=nation&areaName=England>

Appendix B

Figure 13 matches the data shown in Figure 8 but this time plotted with the Exposed and Infected compartments separated to give further detail on the model's results.

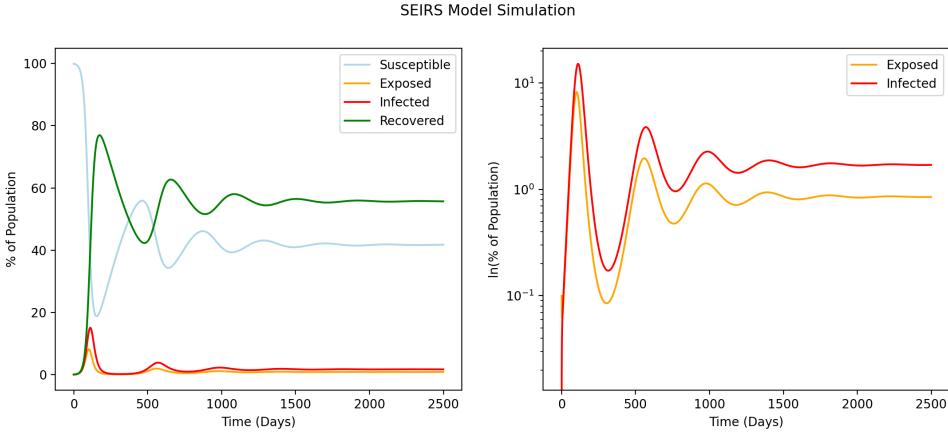


Figure 13: Results for the SEIRS model are plotted against time for the same data used in Figure 8. However, the Exposed and Infected compartments have been separated for further insight. The parameters used in this run were: $\beta = 0.218/\text{day}$; $1/\gamma = 11 \text{ days}$; $1/\mu = 80.9 \text{ years}$; $1/\omega = 1 \text{ year}$; $1/\sigma = 5.5 \text{ days}$

Appendix C

The standard error on the mean calculated to produce the results in table I were determined using the following relations:

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{x})^2} \quad (19)$$

Where s is the sample standard deviation, x_i is a single value from the sample and \hat{x} is the sample mean

The standard error (σ) is then calculated:

$$\sigma = \frac{s}{\sqrt{N}} \quad (20)$$