Modelling: How does a Disease Spread Between Countries and the Efficacy of Countermeasures

SCOTT BLYTH

ACM Reference Format:
Scott Blyth. . Modelling: How does a Disease Spread Between Countries and the Efficacy of Countermeasures. 1, 1 (June), 15 pages.

Contents

Author's Contact Information: Scott Blyth.

Manuscript submitted to ACM

1 SPECIFICATION TABLE

Base Model	Susceptible, Infection, Removed (SIR) model, where susceptible move to infected, then removed.
Extension Assumptions	The removed compartment is split into recovered and dead. Recovered individuals lose immunity overtime. Many countries are added, each with SIRD compartments. There is travel between the countries, where a susceptible individual of one countries can move to the susceptible compartment of another country (same for infected and recovered).
Techniques showcased	**Gillespie**: to introduce randomness for studying the variability of epidemic spread. **Genetic Improvement**: to study the properties that make a disease dangerous, in terms of infectivity and mortality.
Research Question 1	How does an infection spread between countries?
Research Question 2	What are the properties of a dangerous disease in terms of spread and mortality?
Research Question 3	What effects do countermeasures have on how the severity of an epidemic?

Table 1. Model Table

2 INTRODUCTION

In the real world, infectious diseases do not often stay within the one country, which is the scenario the SIR model attempts to describe. The COVID-19 pandemic for instance, the first patients were detected in China, on December 12 2019 [1]. Later on January 13 2020, Thailand detected the first case outside of China, a month long gap. I wish to study this phenomena, and in general any other interesting patterns when studying disease spread on a global level using a model.

The other interesting phenomena is the trade offs between mortality and infectivity when considering how dangerous a disease is. For example, if a disease kills an infected individual quickly, they have little time to spread the disease. Conversely, if the infected individual is sick for a week or so, and can move about, they are much more likely to spread the disease. This observation motivates studying the trade offs.

During a pandemic, a country can of course response to reduce the spread of the infection, such as enforcing mask wearing or lock downs. This motivates studying the efficacy of countermeasures on reducing spread will be studied.

The base SIR model is clearly insufficient to model the spread between countries. This motivates a model extension, where the epidemic spread of multiple countries are modelled, with people moving between the countries. That is, an individual (infected or not), has some probability of moving to a country each day. The probability may change on what country they are at, and what country is being considered.

The Gillespie technique is introduced for a couple of reasons. First, solving a coupled ODE of the size that is generated (18 equations for 3 countries) may be difficult and computationally expensive to compute. Secondly, introducing randomness to the model allows studying the variability of the model. For example, on a first run, maybe the disease first spreads to Thailand, where as on the second run, it spreads to Japan first, which is very relevant to the first question “How does an infection spread between countries".

With regards to the flow of people between countries, due to the use of the Gillespie technique, the probabilities can be simply translated to propensities.

Genetic Improvement is introduced to search for a “dangerous" disease, which is in terms of how many people die

during the pandemic. This clearly helps answer the second question “What are the properties of a dangerous disease in terms of spread and mortality?". It also aids in answering the first question, since we would be interested in all sorts of infections when considering infection spread, making searching for the most dangerous infection relevant.

3 MODEL DESCRIPTION

The model is based on the following list of assumptions. New assumptions are in bold. Table 2 shows some properties of the model.

- Infected individuals have some probability of spreading the infection to another susceptible individual.
- Infected individuals either recover or die.
- **Recovered individuals lose immunity overtime**
- **A Susceptible, Infected or Recovered individual has some probability of moving from their country to another country**
- **An infected individual is less likely to go to another country then susceptible or recovered individuals.**
- **A country will go into lockdown after some proportion of the population is infected at the one time.**
- **A lockdown has the effect of reducing the probability of transmitting the infection.**

Numeric or Analytic	**Numeric**
Discrete or Continuous	**Discrete**
Linear or Non Linear	**Linear**
Deterministic or Stochastic	**Stochastic**

Table 2. Caption

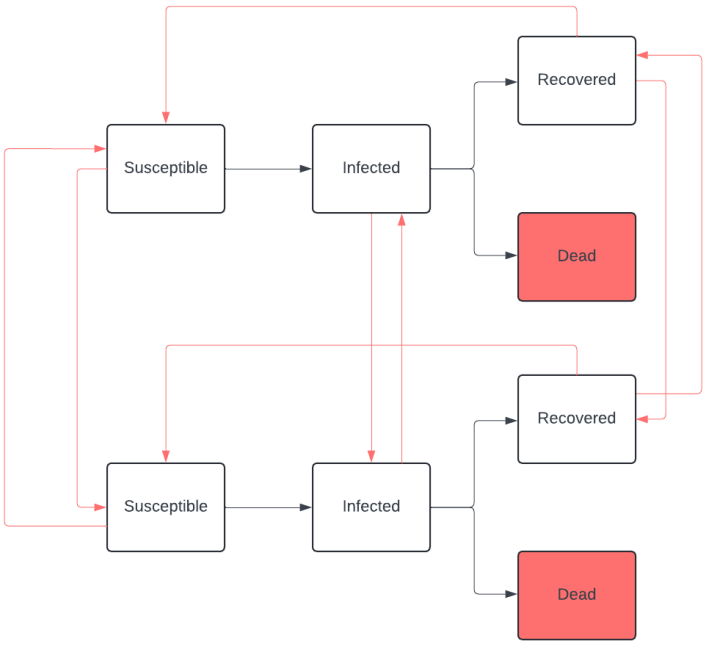


Fig. 1. SIRD Country to Country Spread Model Flow Chart

Figure 1 shows the flow of people between each of the other S,I,R,D compartments. The top boxes are country 1, and the bottom is country 2. This flow chart can be generalised to an arbitrary number of countries. This is done by using a graph.

3.1 Deriving Propensities and Event Consequences

An infection spread is categorised by λ_1, which is the propensity to spread from person to person. The parameter λ_2 is interpreted as the rate at which infected individuals to recover from infection. The third parameter λ_3, is the propensity of infected individuals die from the infection. Finally, λ_4, the rate at which recovered individuals lose their immunity.

In this section, I will describe the precise propensities of each event. As motivation for choosing the propensities for each event, observe that the infectious disease will be ran on different number of countries, for the same $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. An issue that arises where events with the same propensity actually end up occurring less frequently because of the increase in the number of possible events. Therefore, to be able to translate between scenarios with fewer or greater number of propensities, we need to scale the exponential distribution by the number of propensities, P. This way, the occurrence of events have the same frequency no matter how many different events there are.

Event	Propensity
Susceptible to Infected	$\lambda_1 S_i I_i P$
Infected to Recovered	$\lambda_2 I_i P$
Infected to Dead	$\lambda_3 I_i P$
Recovered to Susceptible	$\lambda_4 R_i P$

Table 3. Propensities for Internal Spread country i

Table 3 specifically shows the infection spread within a country, and not the country to country spread. To define the propensities for movement between country to country, we defined a $\lambda_{i \to j}$, which is interpreted as the rate of flow of people from ith country to jth country. We also define μ, which used to reduce the rate at which infected individuals move between countries. The propensity and event table for the flow of people between country i to country j is shown in table 4

Event	Propensity
Susceptible from country i moving to country j	$\lambda_{i \to j} S_i P$
Infected from country i moving to country j	$\frac{1}{\mu} \lambda_{i \to j} I_i P$
Recovered from country i moving to country j	$\lambda_{i \to j} R_i P$

Table 4. Propensities and Events of flow of People from Country i to j

The propensities and events are defined for every country i,j that is "connected", which is all combined for the Gillespie algorithm to choose from. In total, given $|E|$ many edges, and n countries, there are $4n + 3|E|$ many possible events. For a 2x2 grid for example, with edges connecting the countries horizontally and vertically, this gives 40 possible events. The quick growth of the number of events indicates large computation time. This is demonstrated at even at a low number of countries, with a 3x3 grid having an execution time of 185 seconds on one run.

3.2 Simulating Counter Measures

Implementing the lockdown is fairly straight forward. Given a threshold p, and an infection slow down factor s, the infectivity of an infection is slowed by s when the number of deaths proportional to the population exceeds p. More precisely, when $deaths/total\_population \geq p$, the propensity of infection, within a country i, goes from $\lambda_1 S_i I_i P$ to $\mathbf{s} \cdot \lambda_1 S_i I_i P$. The lockdown is then lifted when the infections drop below a second threshold.

3.3 Genetic Improvement

3.3.1 ***Phenotype and Fitness.*** As discussed earlier, each disease, in terms of it's spread, is categorised by the parameters $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. The purpose of the Genetic Improvement algorithm is to therefore search for a set of parameters such that it is dangerous, in terms of the deaths occurred during the spread. Due to performance reasons, the fitness is determined by running the simulation on a single country, where the resulting fitness is the total deaths after the epidemic has finished spreading.

It should be noted, that when finding a solution, we should have some sort of check to determine if the parameters make sense when applied to the real world. For example, a simple best solution would be to put infectivity, λ_1, to 10^{10}, and

the mortality λ_3 to 10^6. This would have an effect of infecting everyone immediately, and then killing them. However this trivial solution is unrealistic. Instead, it would be better to define bounds for each parameter that are reasonable, where if any solution goes outside of this region, the fitness is set to 0. The region is determined loosely by looking at the extremes of each parameter and observing the resulting curves to see if it is "reasonable". In a full paper, looking at real world data may gain insights into what this region should look like.

3.3.2 ***Mutation Operator.*** The mutation operator is simply taking a $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and randomly moving each parameter by a small amount. More specifically, given an infection phenotype $\vec{\lambda}$, and a vector $\vec{u} = (u_1, u_2, u_3, u_4)$, where each component is taken randomly from the interval $[-\alpha_i, \alpha_i]$, the mutated infection $\vec{\lambda}_m$ is given by:

$$\vec{\lambda}_m = \vec{\lambda} + \vec{u}$$

The parameter $\vec{\alpha}$ is a vector, and is used to define how much a parameter can change during a mutation. It is useful for controlling the rate of mutation, where for large changes, the population converges fast to a solution

The advantage of this mutation operator is that it allows for random exploration across the whole behaviour space using small steps in different directions.

The α parameter is useful for controlling the rate of mutation of the population, and be tuned to give better results. For example, a large α is likely initially converge fast, but might take sometime to settle on a single solution. A small α would have a slow convergence rate, however should settle down on a solution after the initial slow convergence.

3.3.3 ***Crossover Operator.*** To motivate the choice of crossover operation, I will use an example. Suppose there is an infection $\vec{\lambda}$, with a high infectivity, but low mortality. Also suppose that there is an infection $\vec{\gamma}$, with a low infectivity and high mortality. Now, only using the parameters already in each of the infections, how would we make a new and more dangerous infectious disease? A good choice might be to choose the infectivity parameter of $\vec{\lambda}$ and the mortality of the of $\vec{\gamma}$, that many people are infected, leading to higher mortality. Using this example as motivation, the crossover operation is defined as follows.

$$crossover\left(\vec{\lambda}, \vec{\gamma}\right) = \left(choose\left(\vec{\lambda}[i], \vec{\gamma}[i]\right)\ for\ i\ in\ 0...3\right)$$

The crossover operator is simply making a new vector where for each parameter, randomly choosing the parameter in $\vec{\lambda}$ or $\vec{\gamma}$.

3.3.4 ***Genetic Algorithm.*** Now that the crossover and mutator operators are defined, we can describe the Genetic Improvement algorithm in full. Some of the specifics of this algorithm is inspired from the book "Advanced Algorithms and Data Structures" by Marcello La Rocca [3].

As with the typical GI implementation, an elite solution is maintained, and is never removed from the population. After each iteration, excluding the elite solution, a random solution is chosen to be removed. The probability of choosing a solution is proportional to their fitness relative to the overall fitness of the group.

For each iteration, two members are chosen at random, proportional to their finesses. The crossover operator is applied depending on some probability (50% chance). If the crossover operation is not applied, the first chosen solution is chosen. From there, the mutation operator is then applied.

The initial population is randomly generated by sampling points approximately within the pre defined search space.

4 RESULTS

4.1 Parameters Used

The solution search space is defined as follows.

$$0.00001 \leq \lambda_1 \leq 0.00003$$

$$0.01 \leq \lambda_2 + \lambda_3 \leq 0.02$$

$$0.0001 \leq \lambda_4 \leq 0.0002$$

Again, these were determined through experimentation, and decided upon whether or not the results looked “reasonable".

The mutation rates for each parameter, $\vec{\alpha}$, is by default:

$$\vec{\alpha} = (0.00001, 0.0001, 0.0001, 0.001)$$

Unless otherwise specified, the movement between countries propensities, $\lambda_{i \to j} = 0.0005$, and the infection movement reduction is such that:

$$\mu = 5$$

4.2 Environment Setup

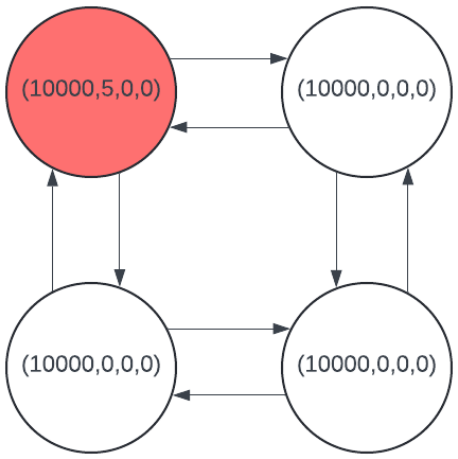


Fig. 2

Figure 2 shows the default environment that the infection spreads through. There are 4 countries, where the arrows show the flow of people. The top left country, coloured in red, is the source country. Each country initially has a population of 10000. The source country initially has 5 infected individuals.

4.3 Research Question One: Spread Trends Globally

The parameters $\vec{\lambda} = (2.3677029021650897 \cdot 10^{-5}, 0.02465276930103708, 0.052665118682770205, 0.0001593990018386284)$ are used by default unless stated otherwise. The movement propensity between countries is set to 0.0005 other stated otherwise.

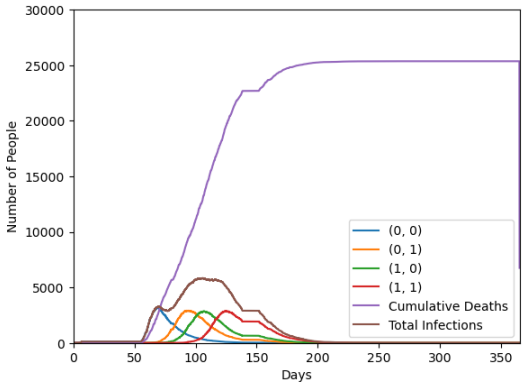
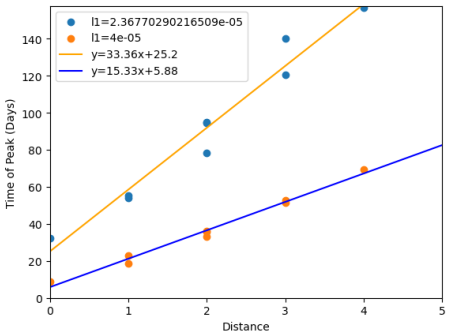
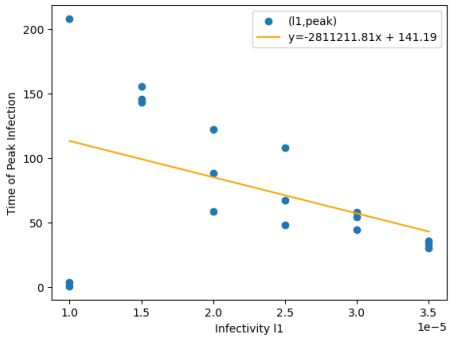


Fig. 3

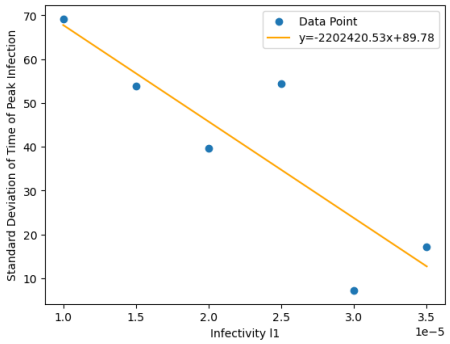
Notice that the furthest country away $(1, 1)$ from the source country $(0, 0)$, is the last country the infections spreads to. Figure 4a also shows this phenomena, where the further away country is, the longer it takes for the infection to spread there. This correlation appears to be a linear one, as figure 4a seems to suggest. Furthermore, the more infectious a disease is, the faster it spreads globally. Figure 4b shows this correlation.
Interestingly, there is a tendency of a diseases ability to spread to have higher variability, as figure 4b appears to show. Figure 5a shows this correlation more convincingly.





(a) Distance from Source Country ($R^2 = 0.96$, $R^2 = 0.99$) ran on a **3x3 grid** with country movement = 0.0005 (b) Ran on a **2x2** grid with country movement = 0.0005 $R^2 = 0.19$

Fig. 4



(a) Variability of Spread Depends on Infectivity ($R^2 = 0.74$ Ran on a 2x2 Grid 3 times)

Fig. 5

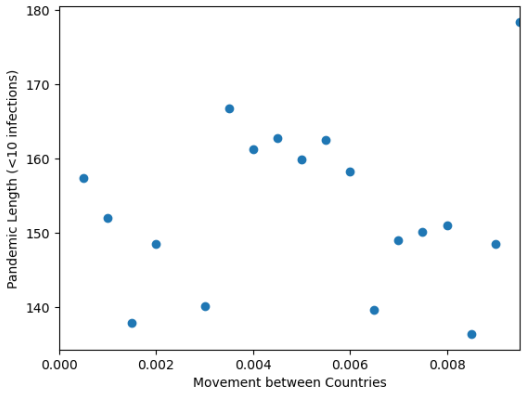


Fig. 6. Pandemic Length vs Movement of People Between Countries

Here, the pandemic length is defined to be the first time after initial spread, that the number of total infections globally are below 10. Figure 6 seems to show little correlation between the movement between countries, and the pandemic length. This suggests that the pandemic length is largely insensitive to the flow of people between the countries. A possible reason for this, is that most of the spread of the disease happens within the country, meaning the “country flow" factor mostly effects whether or not a country is initially infected, and not the overall peak. Indeed, for very low movement (0.00005), the infection has trouble spreading to the other countries and therefore making the pandemic much shorter. Figure 7 shows this, with a pandemic length of 45 days, effecting only the source country.

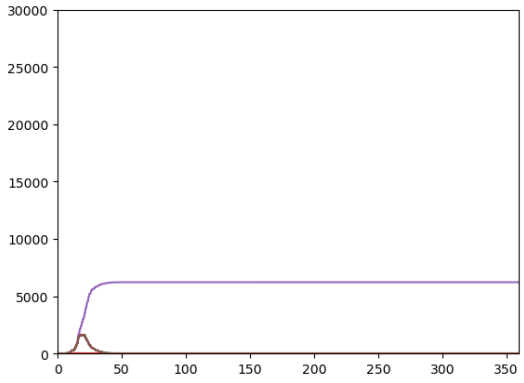
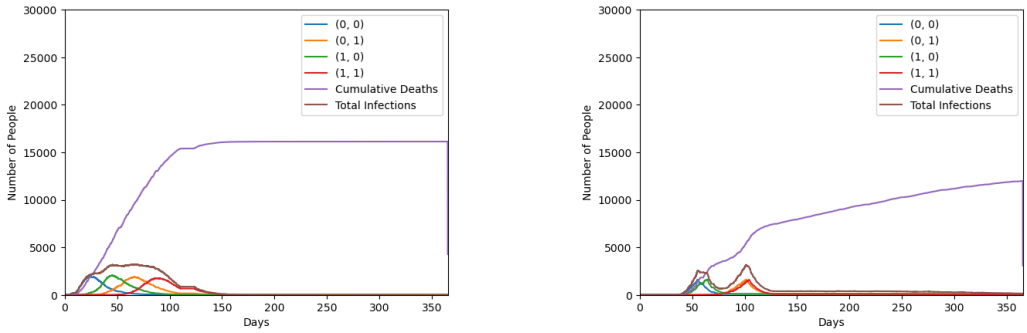


Fig. 7

4.4 Efficacy of Counter Measures



(a) No Counter Measures

(b) Counter Measures Implemented

Fig. 8

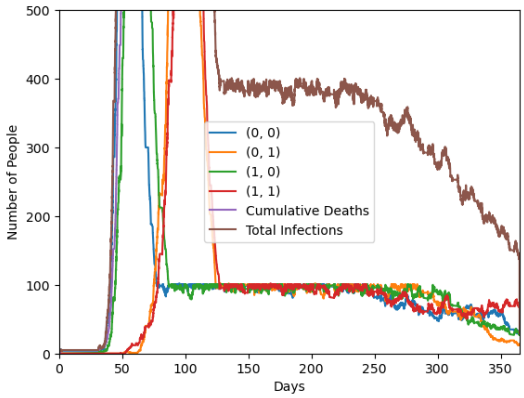
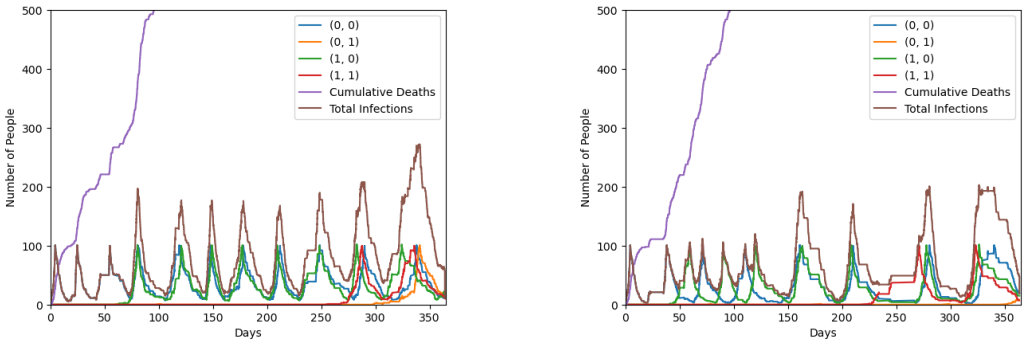


Fig. 9

Figure 8b shows the infection spread when counter measures are used. Specifically, once the proportion of deaths exceed 0.1, the propensity for infection reduces by a factor of 0.01. A country removes the counter measures if the number of infections is less than 1% of the population. Unsurprisingly, there are fewer moralities by the end of the year. However, the counter measures do not end the spread, due to the lifting and re-implementation of the counter measures, as figure 9 shows. This demonstrates that counter measures must be continued beyond the accepted threshold. Figure 10a shows an example of applying this strategy. Here, the trouble of not eliminating the infection remains.

It should be noted that this ineffective strategy also puts the other countries at risk, where the infection eventually spreads to them. However, there is a significant delay in the spread, compared to when no lockdown is implemented (figure 8a and figure 4a). This suggests lock downs may have an impact on the overall global spread, not just locally.

Figure 11 shows how the hardest possible strategy, which is to go into lockdown until there are no infections, is the only effective strategy. This strategy was of course implemented during the Victorian lockdowns [2].



(a) Lockdown Threshold=0.01, Lifting Threshold=0.001=10 People (b) Lockdown Threshold=0.01, Lifting Threshold=0.0005=5 People

Fig. 10. Lockdown Threshold=0.01

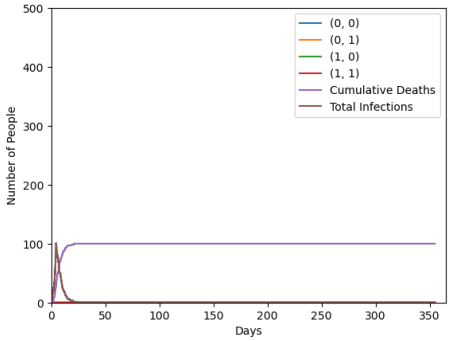
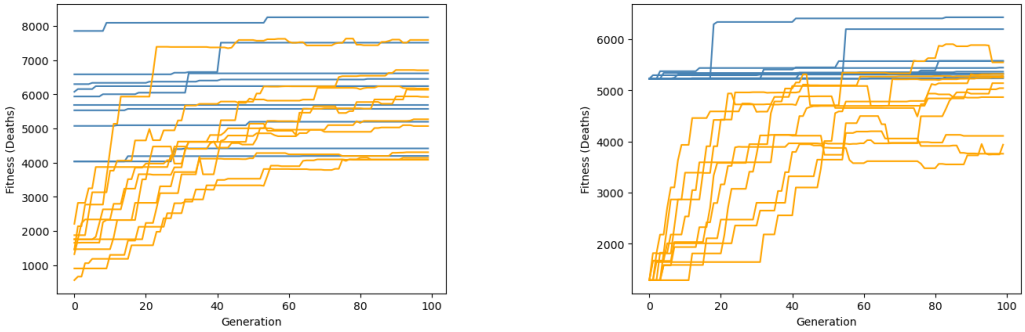


Fig. 11. Lockdown Threshold=0.01, Lifting Threshold=0.0001=1 Person

4.5 What makes an Infection Dangerous



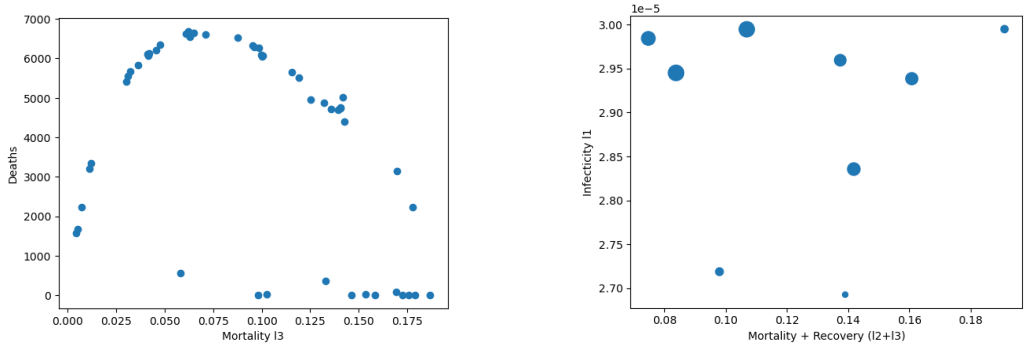
(a) Initial Different Population (10 individuals) (b) Initial Same Population (10 individuals)

Fig. 12

4.5.1 ***Locating Dangerous Infections: Efficacy of Genetic Improvement.*** The genetic improvement algorithm is applied in the environment of a **single** country with an initial population 10000, and 5 infected individuals. The counter measure features is turned off. Figure 12 shows the fitness of the best solution and average fitness of the population for each iteration. The GI algorithm has performed poorly. This can be seen by making the observation that the best fitness curve increases only by small amounts, after many generations. Furthermore, there is significant variability, ranging from approximately 4000 to 8000 fitness. This large range is mainly due to the quality of the initial pool of solutions, as show in figure 12b, where there a reduced variability when using the same initial population for genetic improvement. However, there is still significant variability, with a random of 5000 to 6000 maximum fitness.

The best performing infection found after 100 iterations is given approximately by:

$$\vec{\lambda} = (2.69 \cdot 10^{-5}, 0.0439, 0.095, 0.00014)$$



(a) $\lambda_1 = 2.37 \cdot 10^{-5}$,$\lambda_2 = 0.0246$, $\lambda_4 = 0.00016$

(b) Size = Mortalities

Fig. 13

4.5.2 ***Balancing Mortality with Infectivity.*** Figure 13a shows the mortality rate against the infectivity. We can see the trade off between mortality, where large mortalities negatively impact the spread the infection. Conversely, for low moralities, the infection is simply less deadly. This suggests that infections that lead to higher, quicker mortalities are not necessarily more dangerous, and that there is a balance to be made.
Figure 13 shows solutions plotted found using the Genetic Algorithm. Some observations: the lower recover/mortality propensities have performed better. Furthermore, not unsurprisingly, the more infectious diseases have also performed better. This again confirms the balancing of mortality needed to have a dangerous infection.

4.6 Conclusion

Using a grid structure, where movement of people between vertically and horizontally adjacent countries were simulated, the spread of an infection on a global (multiple countries) level was studied. It was found that time it took for the infection to spread to a country, linearly correlated with the distance from the source country. Interestingly, it was found that the pandemic length was relatively insensitive to the rate at which people moved between countries. It was

suggested that the flow of people between countries mostly effects whether or not the infections spreads to the country, and not the overall spread within the country.

A lockdown was simulated by reducing infectivity once the number of infections reached a threshold, and lifting the lockdown when the number of infections went below a second threshold. Using this, the efficacy of this strategy was studied on a global level. It was found that even leaving 5 infected individuals restarted the epidemic. It was also found that while it increased the time until it spread to other countries, the infection still ended up spreading to them. The only working strategy using this model is to therefore implement the lockdown until there are no infections.

Using Genetic Improvement (GI) to help locate “dangerous" infections, the properties of dangerous infections was studied. It was found that the GI algorithm had high variability, and was often ineffective at finding a good solution. The balancing of mortality rates were studied. It was found that a very high mortality hindered the infections ability to spread, and a very low mortality simply not being dangerous enough. A peak was found, approximately at $\lambda_3 = 0.06$. This shows that a infection with a high mortality is not necessarily the most dangerous infectious disease to humanity.

5 LIST OF ALGORITHMS AND CONCEPTS

This paper’s work builds upon the base SIR model. It relies on the Gillespie algorithm to make the model stochastic. Furthermore, a Genetic Improvement algorithm was implemented to help locate dangerous infections.

REFERENCES

[1] [n. d.]. CDC Museum COVID-19 Timeline. https://www.cdc.gov/museum/timeline/covid19.html
[2] [n. d.]. COVID-19 pandemic in Victoria. https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Victoria
[3] Marcello La Rocca. 2021. *Advanced Algorithms and Data Structures.* Simon and Schuster. 1–27 pages.