Multi-perspective Analysis on a Variant of SEIR Model[*]

Haotian Shen Kai Cao Yujia Yang
Department of Computer Science
University of Cambridge
{hs831, kc645, yy541}@cam.ac.uk

Abstract

The recent Covid-19 epidemic has caused millions of deaths and huge economic losses, prompting people to model it for comprehensive analysis. In this project, we first analyse how human intervention can affect the epidemic: we propose a SEIRSD model, a variant of the SEIR model, and use the model to investigate the relationship between a lockdown policy and the peak infected population caused by the epidemic. Then we analyse how the change of the inherent property of an epidemic (infection rate in our example) will influence the peak infected population. Finally, we provide a use case of our SEIRSD model: we propose a cost function to be optimized to help make an optimal lockdown policy.

1 Introduction

While it has been nearly two years since the Covid-19 outbreak, new variants such as Omicron [1] and XBB.1.5 [2] have constantly been emerging. These variants usually come with new characteristics (for example, the reproduction rate *R0*), and the antibody gained by previous infection may be ineffective in resisting the new variants. Therefore, studying the epidemic under the new situation is still essential. As the basis of this project, two existing pandemic simulators will be briefly described below.

Modelling herd immunity Thomas House proposes a model of herd immunity in his blog post using a set of differential equations [3]. As an SEIR model, it divides the population into compartments: Susceptible, Exposed, Infectious and Recovered. While it is a helpful tool for studying the epidemic trend, it does not capture important factors such as mortality rate and immune evasion. As a disease that has caused more than 6 million total deaths (as reported by WHO [4]), it is crucial to minimise the mortality. Besides, the original model suggests that recovered people would not be infected again, but this assumption can be doubtful because the new Covid-19 variants may evade immunity and causing reinfections. To incorporate these factors, we make improvements based on the original model, of which the details will be given in Section 2.1. The updated model is then used in the sensitivity analysis (Section 2) and lockdown policy guidance (Section 4).

Gillespie simulation The Emukit package includes another SEIR model[2] where Gillespie simulation [5] is chosen to capture the dynamics of the transitions between compartments. This simulation can produce results which are more precise than the former, but it also requires more time to run due to its complexity. Therefore, this simulation is only used in the multi-fidelity modelling part (see Section 3), where a low-fidelity simulation is incorporated to reduce the data points needed from this high-fidelity model.

[*] Code link: `https://colab.research.google.com/drive/1m8AzAPwP8KomtFGsoeem1RSnpDVmikBJ?usp=sharing`

[2] `https://github.com/EmuKit/emukit/blob/main/emukit/examples/spread_of_disease-with_seir_model/Emukit-task-spread-of-disease-with-seir-model.ipynb`

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

This project can be divided into three case studies. In the first case study (see Section 2), we modify Thomas's model to incorporate new factors and performe sensitivity analysis on the three parameters related to lockdown. In the second case study (see Section 3), we combine the Gillespie simulation with a low-fidelity model similar to Thomas's and form a multi-fidelity model for predicting the peak infected population given new infection rates. In the third case study (see Section 4), we design a cost function which balances between different targets and perform Bayesian optimisation to minimise the cost. The resulting lockdown variables can be used to guide policy-making.

2 Case study 1: Emulator design and sensitivity analysis

In this case study, we perform a sensitivity analysis on lockdown start time ($t0ran$), lockdown duration ($tend$) and reproduction number during lockdown (Rt). These parameters are all related to lockdown, and their resulting effects on the peak infected population can be used to guide the policy regarding lockdown [6]. We first modify Thomas's model to incorporate new factors, build an emulator on top of the simulation, and perform sensitivity analysis based on the emulator.

2.1 Improvements on Thomas's model

As described in Section 1, Thomas's "Modelling Herd Immunity" model lacks the effects of mortality and immune evasion. To incorporate these two factors, we first modify the differential equations. The new set of equations is:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta S(I_1 + I_2) + \theta R \tag{1}$$

$$\frac{\mathrm{d}E_1}{\mathrm{d}t} = \beta S(I_1 + I_2) - \sigma E_1 \tag{2}$$

$$\frac{\mathrm{d}E_2}{\mathrm{d}t} = \sigma E_1 - \sigma E_2 \tag{3}$$

$$\frac{\mathrm{d}I_1}{\mathrm{d}t} = \sigma E_2 - \gamma I_1 \tag{4}$$

$$\frac{\mathrm{d}I_2}{\mathrm{d}t} = \gamma I_1 - \gamma I_2 - \mu I_2 \tag{5}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I_2 - \theta R \tag{6}$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \mu I_2, \tag{7}$$

where S, E_1, E_2, I_1, I_2, R are the original compartments (Exposed and Infectious are each split into two compartments), β denotes the infection rate and is defined to be the reproduction number *R0* divided by the length of infectious period, σ denotes the conversion rate from Exposed to Infectious and is defined to be 2 divided by the length of latent period, γ denotes the recovery rate and is defined to be 2 divided by the length of infectious period. Out of these parameters, β can be affected by lockdown during which the reproduction rate *R0* is reduced to Rt. The remaining are new parameters we added and will be discussed below.

Mortality In addition to the six compartments in the original model, we create a new compartment called "Deceased"(D). We add a transition from the second Infectious compartment to the Deceased, whose rate is controlled by the parameter μ. These changes are reflected in Equations 5 and 7.

Immune evasion As time progresses, new variants of the virus may appear which evade the immunity, so the recovered people may become susceptible again. A transition from the Recovered compartment to the Susceptible is added to model the loss of immunity. The rate of this transition is controlled by θ. These changes are reflected in Equations 1 and 6.

2.2 Emulator design

To avoid running the simulation for each set of input parameters, an emulator is built to capture the relationship between the input parameters ($t0ran$, $tend$, Rt) and the peak infected population based

Table 1: Constant model parameters

N	$i0$	$tlast$	$latent\_period$	$infectious\_period$	$R0$	θ	μ
6.7×10^7	1×10^{-4}	100	5	7	2.5	1/120	1/100

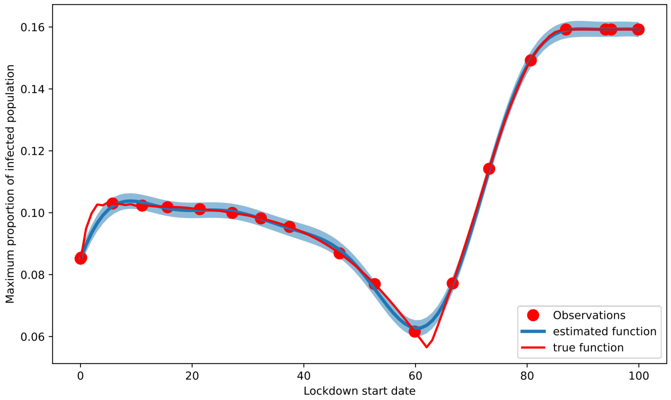


Figure 1: Gaussian process fit of height of peak with respect to lockdown start date ($t0ran$). The other two parameters are set as $tend = 21$ and $Rt = 0.75$.

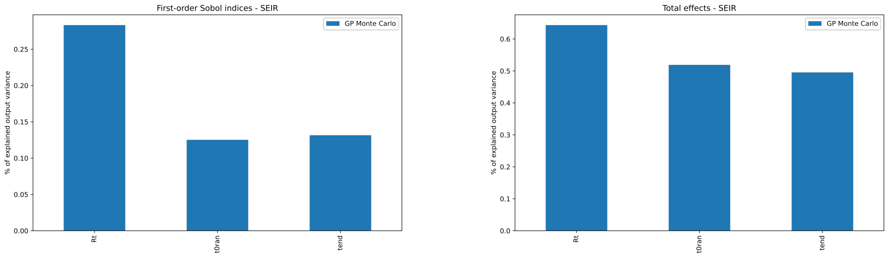
on a Gaussian Process (GP). We first model each parameter individually while keeping the other two constant to verify the concept. For each parameter, we randomly select 5 points in its input space and then run an experimental design to select a further 15 points based on model variance. These data points, together with their corresponding height of peak returned from the simulation, are used to train the Gaussian Process. The kernels of the Gaussian Processes are selected to be Exponentiated Quadratic, given the smoothness of the functions. The kernel parameters (variance and lengthscale) are optimized based on the 20 data points. During the optimization, the noise variance is kept at a low value (10^{-6}) since our simulation model is deterministic and should not contain noise.

During the experiment, we notice that the GP fitted poorly on the function of $t0ran$ (lockdown start date) because there is a sharp drop around $t0ran = 60$, and for $t0ran > 80$, the function plateaus. We therefore reduce the range of $t0ran$ to $0 - 100$ and total length of the simulation $tlast$ to 100, based on the assumption that a lockdown enforced too late with respect to the disease outbreak would have little impact on the peak infected population and thus should be neglected in this study.

The model parameters which are kept constant throughout the study are summarised in Table 1, where N is the total population, $i0$ is half of the proportion of infected people at the start of the simulation, and the other parameters are explained in the previous sections. As an illustration, the fitted Gaussian process for $t0ran$ is shown in Figure 1. Given that only a small number of data points (20) are used, the fit can be considered satisfactory.

2.3 Sensitivity analysis

Based on the previous step, we build another Gaussian process to model the 4D function which takes all three parameters as input variables. The kernel for this GP is initialized as a 3D kernel with variance 1 and lengthscale 5 and optimized later on the obtained data points. Considering the difficulty of training the GP in the 3D parameter space, we increase the number of training data points to 15 initial points and a further 45 through experimental design. The first-order and total sensitivity are calculated on $10,000$ Monte Carlo samples from the GP.



(a) Main effects.

(b) Total effects.

Figure 2: Main and total effects of the three parameters: Rt, $t0ran$, $tend$.

2.4 Results

The main and total effects are illustrated in Figure 2. For main effects, the effect of Rt is around 0.28 which doubles the effect of $t0ran$ (0.13) and $tend$ (0.12). The total effect of Rt (0.64) is also the largest among the three. This seems to suggest that the reproduction rate during the lockdown, in other words, how strict the rules are during lockdown and how faithfully people follow them, may have a more significant impact on the peak infected population than the start date and duration of the lockdown. Nonetheless, it should be noted that the range of Rt was set to $[0.1, R0]$ in the parameter space. According to the statistics from UK Health Security Agency [7], the reproduction rate of Covid-19 has been in the range $[0.6, 1.4]$ since the outbreak. Therefore, one might doubt if it is feasible to suppress Rt to 0.1 in the real world. Despite being a reasonable doubt, it is also important to note that the base reproduction rate (*R0* $= 2.5$), along with the other parameters, may not accurately reflect the real world as well. In addition, the target function in this study only concerns the peak infected population and ignores other factors such as the number of deceased populations. Consequently, the results in this section should be mainly interpreted in this toy set-up. To better model the real-world situation, we incorporate a high-fidelity model in Section 3 and design a more realistic cost function in Section 4.

3 Case study 2: Multi-fidelity Analysis

Most of the simulation process is converted into a mathematical differential equation solving problem with a discretized domain [8]. Although reducing the stepsize could help obtain a more accurate deterministic result, the simulator is inherently unable to model the randomness. Multi-fidelity analysis (MFA) is an approach to harmonizing both high and low-fidelity data to minimize the cost and maximize the model accuracy [9]. In our experiment, both linear and non-linear MFA is applied to the SEIR model to investigate the relationship between infection rate and peak infected population.

3.1 High-Fidelity: Gillespie simulation

The dynamics of the simulator are caused by random state changes of the individuals. The high-fidelity model approximates the dynamic changes of the SEIR model by using the Gillespie simulation (GS). Each run simulates the possible dynamic changes of 100 individuals in the SEIR model. For each individual, the current state is used to calculate the rates of the three possible state changes(S->E, E->I, I->R) and to randomly select the next state according to the probability of this rate. The mean value is gathered for each individual 1000 times to obtain the peak infected population at each infection rate. In addition, the randomness is also represented by the time interval between state changes as it is subject to exponential distribution. Thus, a stochastic and general model is obtained through the process of GS. However, since each simulation requires a thousand state changes for hundred individuals, GS is significantly slower than the ordinary differential equation calculation.

In order to simplify the calculation, we define α (infection rate) as the ratio of infection and recovery rate and β as the ratio of incubation rate and recovery rate. For experimental design, We use the Emukit

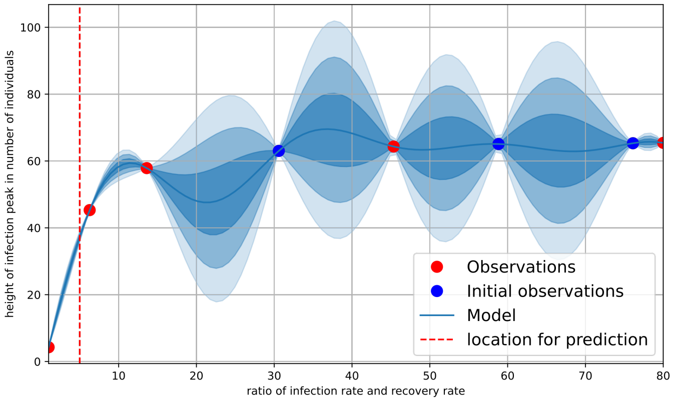


Figure 3: High-fidelity analysis on the relationship of infection rate and peak infected population

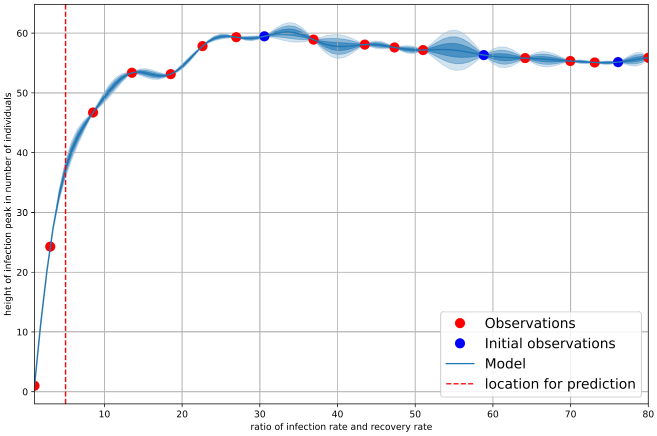


Figure 4: Low-fidelity analysis on the relationship of infection rate and peak infected population

experimental design package from the emukit example `spread-of-disease-with-seir-model`[3]. For the emulator model, we use GPy since GPy model wrappers already exist in Emukit. We fix the incubation rate to $\beta = 4$ and call the function `f_height_of_peak` to compute the peak infected population. One call of `f-height-of-peak` will run 1000 Gillespie runs. After that, we plot the relationship between α and the peak infected population 3.

3.2 Low-Fidelity: Ordinary differential equation (ODE) solver

The low fidelity is fitted using the `sol-ivp` method for numerically integrating a system of ode functions with initial values and performing the Gaussian process. The stochastic nature of the simulation is lost in this process due to the fixedness of the integral solution. The experiment design and code implementation are similar to the previous 2.2 and the plot is presented above4.

[3] `https://github.com/EmuKit/emukit/tree/main/emukit/examples/spread_of_disease-with_seir_model`

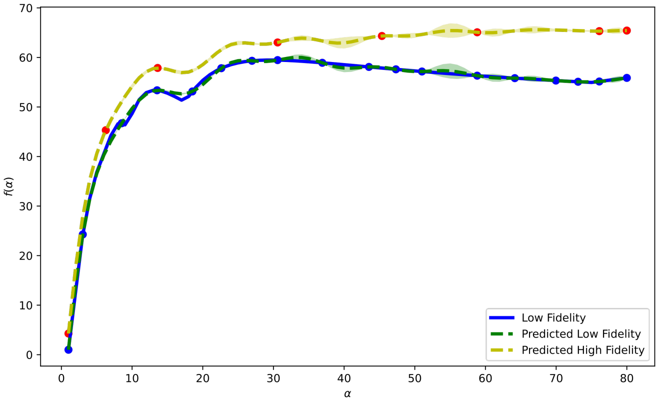


Figure 5: Linear Multi-fidelity analysis on the relationship of infection rate and peak infected population.

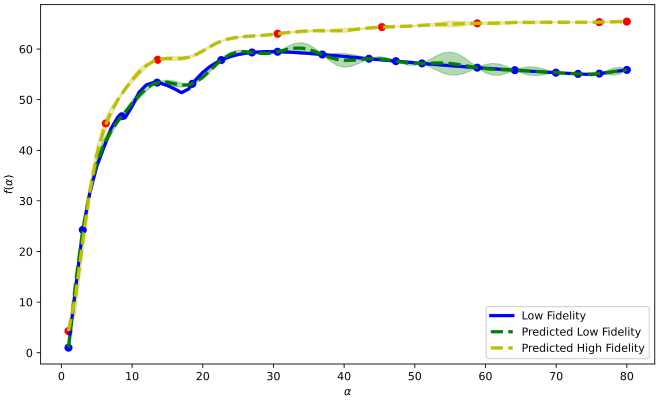


Figure 6: Non-linear Multi-fidelity analysis on the relationship of infection rate and peak infected population

3.3 Multi-Fidelity

Above all, using MFA on the SEIR simulator allows the randomness and accuracy of the high-fidelity model to be combined with the low cost and fast speed of the low-fidelity model. We separately run and get the Linear Multi Fidelity Analysis (see Figure 5)and Non-linear Multi Fidelity Analysis (see Figure 6).

3.4 Result

For the high-fidelity analysis3, the fit of the model is not very accurate. The variance is large. This result is caused by lacking sample points. In our implementation, only three initial points and five experimental design points were selected. Based on the fact that Gillespie simulation cost significant computational resources, adding more initial points will cause the program to run slowly.
For the low-fidelity analysis4, the fit of the gaussian process is accurate and fast, but it is not accurate enough to answer the general question of our study.
Comparing the result of linear5 and non-linear multi-Fidelity6 analysis, both of them share the same trend, and the plots are smooth and accurate. Thus the linear model is complex enough to analyze this problem.

4 Case study 3: Lockdown policy guidance

The SEIRSD model we propose can assist real-world decision-making. One example is its guidance to the lockdown policy. A typical lockdown policy includes three main parts: lockdown start time, duration, and reproduction number during the lockdown. Adopting different lockdown policies will change the peak infected population and the total death toll. Usually, a small peak infected population during the epidemic is preferred as it will put less pressure on medical resources. Also, a smaller total death toll can alleviate the public panic about the epidemic. However, although long lockdown duration and strict lockdown regulations are necessary to obtain small peak infections and less death, it may cause resentment in society [10]. Therefore, proposing an appropriate lockdown policy to balance different factors by minimizing the total cost is essential. We propose a solution to a constrained optimization problem to guide the lockdown policy.

The objective cost function to be minimized has three variables: lockdown start time, duration, and reproduction number during the lockdown. Lockdown start time is within the range [0, $tlast - maxt$], where *tlast* is the pre-defined lasting time of the epidemic. Lockdown duration is at least 0, and at most *maxt*, a user-defined upper limit. The reproduction number during lockdown is within the range [0.6, *R0*], where *R0* is the pre-defined basic reproduction number of the epidemic in the absence of lockdown. Notice that the smaller value means the less likely spread of the epidemic, which indicates a stricter lockdown. In this example, we set *tlast* = 730, *maxt* = 30, and *R0* = 2.5, but the policymaker can change these values to match any epidemic. For ease of solving the cost function, we normalize the three variables to the range of 0 to 1. We name the above three variables x_1, x_2, and x_3, respectively.

To solve the cost function, we also need the peak infected population and the total death toll for the given variables. These can be calculated using the SEIRSD simulator by solving differential equations discusses in section 2.1. Notice that apart from the variables x_1, x_2 and x_3, the simulator also requires other inputs, including the latent period of the epidemic. However, those inputs are related to the epidemic itself and assumed to be pre-defined by the policymaker. Therefore, we only focus on the three mentioned variables. We name the peak infected population and total death toll y_1 and y_2, respectively.

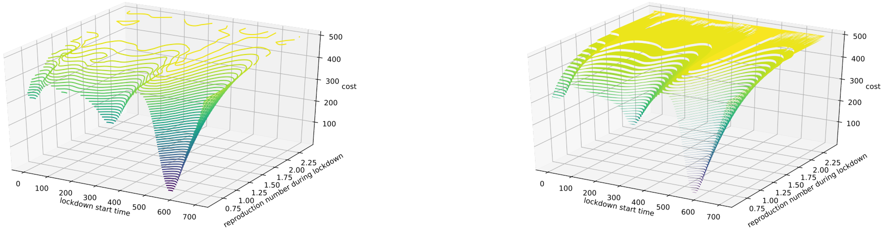
The objective cost function is defined as follows:

$$Cost = tmp + n_2 \times (y_2 - min_d) + n_3 \times x_2 + n_4 \times (R0 - x_3) \tag{8}$$

$$tmp = \begin{cases} n_1 \times (y_1 - min_i) & r < min_i \\ n_1 \times (y_1 - r) & min_i \leq r < y_1 \\ 0 & otherwise \end{cases}$$

$$\text{s.t.} \quad 0 \leq x_j \leq 1, \quad j = 1, 2, 3$$

Notice that the cost is the addition of four weighted factors: peak infected population, total death toll, lockdown duration, and reproduction number during the lockdown. The policymaker can assign different weights to them to determine their importance. However, these four factors have different orders of magnitude: the total death toll may reach millions given a large number of susceptible populations, whereas the reproduction number during the lockdown is at most 2.5. This will cause confusion when assigning weights. Therefore, we normalize the weight by dividing the user-assigned weight by the range of each factor value. The normalized weight of lockdown duration and reproduction number during the lockdown are easily computed as their pre-defined ranges do not change irrespective of different variables. We name their normalized weights n_3 and n_4, respectively. To compute the range of the peak infected population and the total death toll for a given set of variables, we need to use the SEIRSD simulator. The strictest lockdown is required to obtain the minimum peak infected population and total death toll, so lockdown duration should be set to *tmax* and the reproduction number during the lockdown to 0.6. On the contrary, no lockdown should be imposed to obtain the maximum values. We get these values by running the gaussian process and Bayesian optimization. The ranges of the peak infected population and total death toll are then computed for weight normalization. We name the normalized peak infected population and total death toll weight as n_1 and n_2, respectively. Notice that we name min_i as the minimum peak infected population and min_d as the minimum total death toll since they are used in the cost function.



(a) cost optimizer prediction

(b) Ground truth

Figure 7: Comparison between cost optimizer prediction and ground truth

One final thing to mention is that the peak infected population does not affect the cost as simply as other factors: it first compares with the maximum medical resources (denoted by *r*) and then makes corresponding addition to the cost. The cost function above specifies this factor's cost for different medical resources in detail.

We use the Gaussian process and Bayesian optimization to find the minimum cost. Specifically, we sample 200 sets of variables and obtain the corresponding cost as the initial input for the Gaussian process. The faster isotropic RBF kernel instead of the ARD kernel is used for the task thanks to the normalization of the variables. Then we run the Bayesian optimization with the expected improvement acquisition function for 200 steps to obtain the minimum cost and its corresponding variables' value. The policymaker can use the variables that lead to the minimum cost to formulate the lockdown start time, duration, and reproduction number during the lockdown.

To show the correctness of our optimizer, we suppose a case where the policymaker really wants to reduce the total death toll. He assigns 500 as the weight to the total death toll factor and 1 to all other factors. The optimized cost is 1.18, and the corresponding variables: lockdown start time, duration, and reproduction number during the lockdown, are 596.5, 30, and 0.6, respectively. This result is reasonable as the lockdown duration needs to be maximized (in this example, 30), and the lockdown needs to be the strictest (in this example, 0.6) to reduce mortality. We also calculate the cost from the set of variables (573.8, 30, 0.6) which leads to the minimum total death toll. The cost is 1.41, a little higher than the former. Therefore, the optimizer can balance the importance of different factors and obtain a set of variables that can lead to a desirably low cost.

To visualize the Bayesian optimization of the cost function to further demonstrate the effectiveness of our cost function optimizer, we reduce the three variables to two variables while keeping all other setups the same. Specifically, we suppose x_2 is inversely proportional to x_3 and use x_3 to represent x_2. Similar to the above, we sample 200 points to fit the Gaussian process and run Bayesian optimization for 100 steps. We use these 300 points to draw a plot to check which variables can reach the minimum cost (Figure 7a). We also sample 80,000 points using grid sampling to plot the ground truth (Figure 7b. By comparing the two plots, one can find that the optimal set of variables generated by the cost optimizer is very close to actual optimal lockdown-related variables. This again shows the robustness of the cost function optimizer.

5 Conclusion

Our experiment extends the original SEIR model to the SEIRSD model, which considers the immune evasion and total death toll. By running the sensitivity analysis in SEIRSD model, Rt is the critical variable to the peak infected population among the three lockdown related factors. We chose the Gillespie simulation for high-fidelity analysis, which can provide a more accurate result. In multi-fidelity analysis, the linear model is accurate enough for the SEIR model. In addition, we also propose a cost function to offer guidelines to lockdown policies. After testing, it is found to be proper enough.

The future study could work on introducing other factors, such as vaccination and natural death rates, on forming a more complex and realistic model. In addition, subsequent research can add flexibility to variables including *R0* which is fixed in each simulation. For example, we can set up a function for *R0* such that *R0* can vary with respect to time.

References

[1] Centers for Disease Control and Prevention (CDC). *Variants of the Virus*. `https://www.cdc.gov/coronavirus/2019-ncov/variants/index.html`. Accessed: 2023-01-17.

[2] UK Health Security Agency. *COVID-19 variants identified in the UK – latest updates*. `https://www.gov.uk/government/news/covid-19-variants-identified-in-the-uk-latest-updates`. Accessed: 2023-01-17.

[3] Thomas House. *Modelling Herd Immunity*. `https://personalpages.manchester.ac.uk/staff/thomas.house/posts/modelling-herd-immunity.html`. Accessed: 2023-01-17.

[4] World Health Organization (WHO). *WHO Coronavirus (COVID-19) Dashboard*. `https://covid19.who.int/`. Accessed: 2023-01-17.

[5] Daniel T Gillespie. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of Computational Physics* 22.4 (1976), pp. 403–434. ISSN: 0021-9991. DOI: `https://doi.org/10.1016/0021-9991(76)90041-3`. URL: `https://www.sciencedirect.com/science/article/pii/0021999176900413`.

[6] The DELVE Initiative. *Test, Trace, Isolate*. Tech. rep. 2020. URL: `https://rs-delve.github.io/reports/2020/05/27/test-trace-isolate.html`.

[7] UK Health Security Agency. *The R value and growth rate*. `https://www.gov.uk/guidance/the-r-value-and-growth-rate`. Accessed: 2023-01-17.

[8] Pierre Thodoroff et al. *Multi-fidelity experimental design for ice-sheet simulation*. 2022. URL: `https://neurips.cc/Conferences/2022/ScheduleMultitrack?event=56484`.

[9] Wikipedia contributors. *Multifidelity simulation — Wikipedia, The Free Encyclopedia*. [Online; accessed 16-January-2023]. 2023. URL: `https://en.wikipedia.org/w/index.php?title=Multifidelity_simulation&oldid=1131482837`.

[10] Ahmed Sharif et al. "A pragmatic approach to COVID-19 vaccine passport". In: *BMJ Global Health* 6.10 (2021), e006956.