



**HAL**  
open science

# Exploring and optimising infectious disease policies with a stylised agent-based model

Jeonghwa Kang, Juste Raimbault

## ► To cite this version:

Jeonghwa Kang, Juste Raimbault. Exploring and optimising infectious disease policies with a stylised agent-based model. French Regional Conference on Complex Systems, May 2023, Le Havre, France. pp.179-196, 10.5281/zenodo.7957531 . hal-04428487

**HAL Id: hal-04428487**

**<https://hal.science/hal-04428487v1>**

Submitted on 31 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Exploring and optimising infectious disease policies with a stylised agent-based model

Jeonghwa Kang · Juste Raimbault

**Abstract** The quantitative study of the spread of infectious diseases is a crucial aspect to design health policies and foster responsiveness, as the recent COVID-19 pandemic showed at an unprecedented scale. In-between abstract theoretical models and large-scale data driven microsimulation models lie a broad set of modelling tools, which may suffer from various issues such as parameter uncertainties or the lack of data. We introduce in this paper a stylised ABM for infectious disease spreading, based on the SIRV compartmental model. We account for a certain level of geographical detail, including commuting modes and workplaces. We apply to it a set of model validation methods, including global sensitivity analysis, surrogates, and multi-objective optimisation. This shows how such methods could be a new tool for more robust design and optimisation of infectious disease policies.

**Keywords** Infectious disease · Agent-based modelling · Model exploration and validation · Multi-objective optimisation

## 1 Introduction

The impact of one of the fatal epidemics in history, the Black Death, has led to the death of almost one-third of the population of Europe, and the most recent epidemic in the world today, COVID-19, is currently costing millions of valuable human lives [1]. Before the introduction of mathematical and computational modelling, humans had little knowledge on how to effectively control and minimise the spread of infectious diseases. Furthermore, the importance of

---

J. Kang  
CASA, UCL

J. Raimbault  
LASTIG, IGN-ENSG  
E-mail: [juste.raimbault@ign.fr](mailto:juste.raimbault@ign.fr)

forecasting the spread of diseases in recent years has become even more crucial as the world is rapidly globalising with increasing travel rates and distances [2]. Beyond the health burden of epidemics, many dimensions of social systems are affected, with for example broad economic impacts [3]. The current world is a highly complex system in which numerous dimensions play a role in decision-making [4]. Designing disease control strategies imply accounting for various interacting factors and finding a compromise between multiple objectives [5].

The first modelling approach to disease spreading goes back to the 17th century as John Graunt conducted an empirical study on infections affecting individuals in various regions in Britain [6]. Bernouilli introduced in the 18th century an equation-based approach to studying the outbreak of the smallpox epidemic in Europe [7]. Most recent modelling approaches range from elaborated mathematical models to data-driven microsimulation models. The recent pandemic has shown the usefulness of quantitative modelling to evaluate policies, such as how to vaccinate the population [8] or to evaluate which degree of lockdown is necessary.

Agent-based modelling are proposed by some researchers as a powerful tool to explore practical scenarios for decision-making when managing the spread of infectious diseases [9]. The classical epidemiological modelling relies on compartmental models, such as the SIR model (Susceptible, Infected, Recovered), with differential equations governing the evolution of the different compartments, or in some cases discrete-time equations [10]. Recent extensions, such as the SEIRD model [11], include a more detailed description of disease stages. [12] describe an agent-based model based on similar states for agents. The advantage of the ABM approach is that one can define for each agent characteristics such as location, age, and many other factors believed necessary in the model. Once the agents are defined with their own set of characteristics, they proceed through a set of rules which enable them to interact with other groups of agents within the simulation. Age groups have contact probabilities to simulate the interactions and infections. More details can be included in an ABM, such as wearing face masks or respecting social distancing [13]. Furthermore, the infectious disease may interact with other chronic conditions, what can be accounted for in an ABM [14].

We propose in this paper to explore the potential of model validation methods to design and optimise epidemiological policies. We introduce a stylised model for the spread of infectious diseases, which is hybrid in the sense that it lies halfway large data-driven models and abstract theoretical models. We account for geographical factors, in particular commuting and workplaces. We focus, as an illustration, for two policy factors, which are social distancing and face masks [15], and vaccination [16]. Our main contribution is the application of state-of-the-art model sensitivity analysis and validation techniques [17] to such a stylised model, opening research perspectives to cases where data is lacking, model parametrisation is uncertain, or actors have not the resources to implement a large scale model, but where accounting for spatial and social structure remains crucial. We do not provide directly actionable policy insights, but rather provide a proof-of-concept of the use of such tech-

niques to extend model capabilities: since our model lies between simple but scalable compartment models and very complicated data-driven microsimulation models, it could in principle not be used to investigate scenarios and decision-making. Our approach allows reducing uncertainty on parameters and providing multi-objective optima for example, improving thus the applicability of the model.

A few examples of the application of such techniques to epidemiological models can be found in the literature. [18] utilise a multi-objective optimisation algorithm for the distribution of limited vaccines resources. [19] review sensitivity analysis of infectious disease models. No systematic application of an umbrella of model validation methods can however be found.

The rest of this paper is organised as follows: we first describe the stylised agent-based model, we then apply model exploration and validation methods to it, and finally discuss the implications of our results.

## 2 Model description

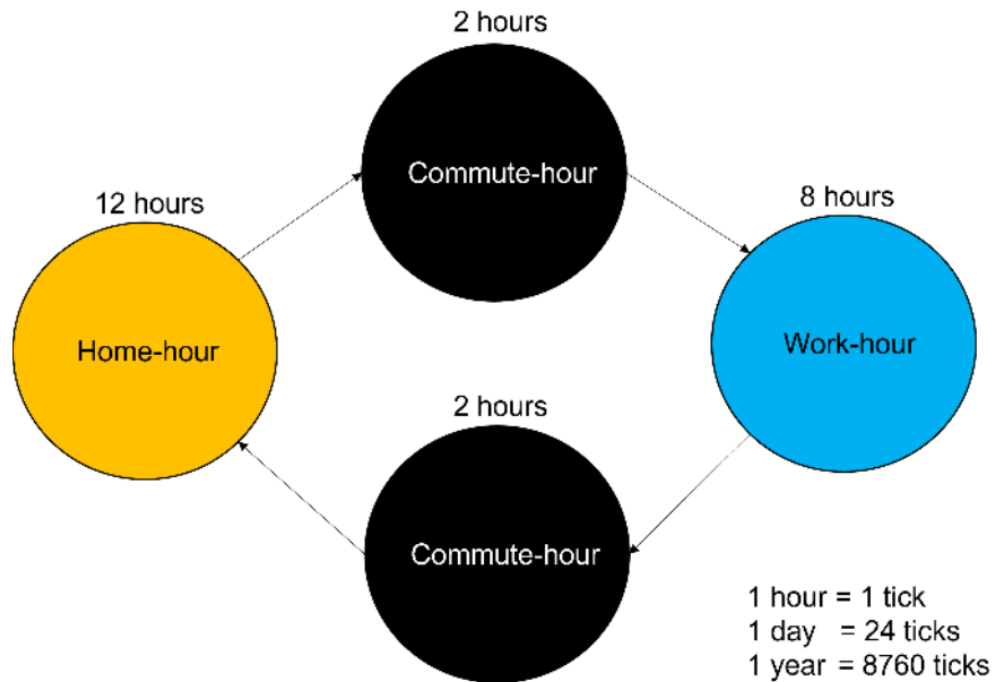
### 2.1 Rationale

Our simulation model projects a real-life situation where agents commute from home to their designated workplaces via different modes of transportation on a fixed daily schedule. At the beginning of the simulation, the entire population is set to susceptible. Then, an infected agent is introduced daily for the first 30 days because if none of the infected agents gets introduced into the environment, the infection process does not happen. The infection then spreads through localised interactions between agents.

### 2.2 Model setup and dynamics

One main assumption of the model is that most human interactions occur while people travel and work (while home infections play a role in disease transmission, our model is built to focus on parameters linked to social distancing in transport and at work). Hence, the scenario is primarily about people commuting from home to their designated workplaces via different modes of transportation, closely depicting ordinary people's daily routine. The simulation of this model considers several essential assumptions. The entire duration of the simulation is set to 365 days. Each day is equivalent to 24 ticks. Hence, each tick is equivalent to an hour. A single day comprises three pre-defined phases: the 'Home-hour', 'Commute-hour' and 'Work-hour' phases, as show in Fig. 1.

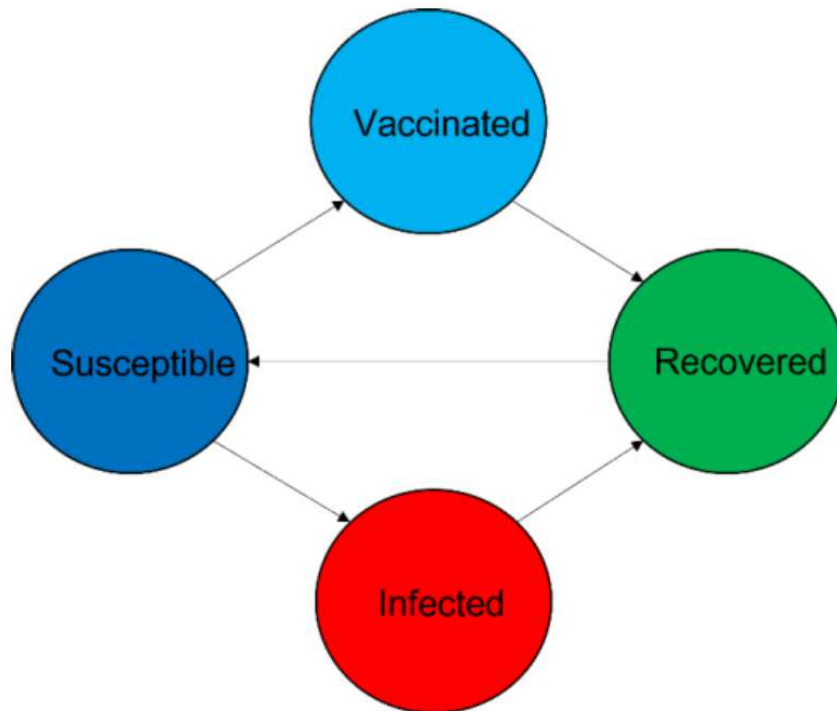
During these phases, the movement of agents is completely randomised. No learning behaviour alternates the agents' movement in the simulation since it would be difficult for an individual to be fully aware of another agent's infection status. During the interaction, agents fall into one of the four states defined by



**Fig. 1** Transitions between phases in the model.

the SIRV compartmental model. The SIRV model is an extended SIR model that accounts for the vaccination of the susceptible population [20]. The agent states are shown in Fig. 2.

Agents can interact with each other based on the state in which they fall. However, the interaction and infection processes are set to occur between the agents within the same patch only. The infection does not happen during the 'Home-hour' phase as this model does not consider social activities or household infections. During the 'Commute-hour' phase, an infected agent can only infect the susceptible agents that use the same mode of transportation as itself. It would be logically sensible that an infected agent commuting via train can only interact with and infect other susceptible agents commuting via the same mode of transportation. This same logic applies during the 'Work-hour' phase, where an infected agent can only interact with and infect other susceptible agents working in the same company. We include four modes of transportation in the model (metro, train, walking, car). The stylised way transportation is described allows avoiding the complexity of transport microsimulation (see e.g. [21] for an adaptation of the MATSim framework to epidemiological modeling) while still accounting for transport processes. We furthermore consider stylised company size distributions (7 large companies, 4 medium sized, 4 small). There are total of fifteen companies which are randomly categorised into small, mid, and big-sized companies. The central assumption is that the bigger the company size, the more the human interactions involved. Hence, the agents working in a big-sized companies are more likely to get infected than the



**Fig. 2** Agent states in the model

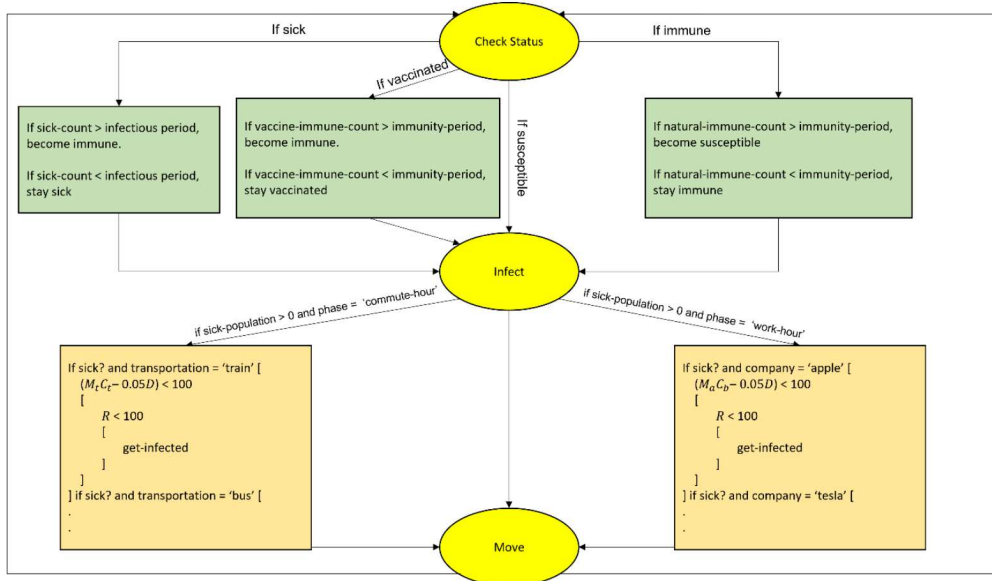
agents working in mid and small-sized companies (in practice, this assumption could be refined, as infections also depends on the size of the premises and on ventilation; simulating density and real-time interaction remains however out of the scope of our stylised approach).

At the beginning of the simulation,  $N = 1500$  susceptible agents are randomly distributed in the simulation environment. The variable values for the agents, such as the transportation and company, are randomly assigned during the setup stage. An infected agent is then constantly introduced at a random location for the simulation's first month (30 days), to take into account possible transient dynamics. Every day, a random percentage value from zero to one percent of the total population gets replaced. This is based on the assumption that the population is dynamic (considering thus an open system, consistently with the intermediate complexity of our model). A certain number of people leave the area, and others come to stay in the area. Hence, the total population is not fixed to the initial population throughout the simulation.

Model dynamics are summarised in Fig. 3.

### 2.3 Model parametrisation

We give in this section more details on the parametrisation of the stylised model.



**Fig. 3** Description of model dynamics. Processes are described using the NetLogo syntax, as used in the model implementation.

### 2.3.1 Global variables

Infection rates are calculated based on global and local parameters. The variables that define contact rates for each mode of transportation and company are pre-defined with specific values.

For transportation contact rates, we take stylised values decreasing with the density one can expect in each transport mode:  $c_{bus} = 40\%$ ,  $c_{metro} = 35\%$ ,  $c_{train} = 30\%$ ,  $c_{walk} = 10\%$ ,  $c_{car} = 0\%$ .

No discrete values define the accurate contact rates for different modes of transportation, but the general assumption was made based on the understanding that the lower the capacity of transportation, the higher the density; hence, the higher contact rates between the passengers. The contact rate specifically for the car is set to zero because it is assumed that an agent who commutes by a car makes no contact with other agents during the 'Commuter-hour' phase. The same logic was applied when defining the contact rates for different companies. The assumption was based on the understanding that the bigger the company size, the more the agents interact.

For contact rates on work site, we therefore take the following values:  $c_{Big} = 40\%$ ,  $c_{Medium} = 35\%$ ,  $c_{Small} = 30\%$ . These values do not have an empirical basis, and should be integrated in the sensitivity analysis, would the model be applied to a real-world decision making problem (in the following, we provide only a proof-of-concept with few other parameters).

### 2.3.2 Agent variables

The importance of wearing a mask while interacting with an infected agent can significantly reduce the chance of infection. Furthermore, patients with cardiovascular diseases are possibly with a higher risk of getting an infection [22] (this aspect could also capture individuals with a deficient immune system for example). Due to these reasons, we have pre-defined values for infection risk rates based on the combination of these two variables: No mask & Cardio = 50% risk rate; Mask & Cardio = 45%; No mask & No cardio = 15%; Mask & No Cardio = 10%. We do not consider the fatality of the disease, but these would follow similar patterns.

### 2.4 Model parameters

We list below main model parameters which will vary during numerical experiments, corresponding to the processes on which the model focuses.

- Infectious period, ranging from 2 to 4 weeks; the longevity of the infectious period of an agent is believed to play an essential role in the spread of diseases because the longer the infected agent stays contagious, the more the chance of infection during an extended amount of time [23].
- Immunity period, ranging from 9 to 11 months. This is selected as one of the models input parameters because it is believed to play a vital role in protecting the agents from infection [24]. Hence, we are interested to know to what extent the immunity period significantly affects the spread of diseases.
- Vaccination rate: as discussed in the literature review, vaccines are the most effective preventive measure which can trigger a biological immune response to fight disease-causing organisms: beside strongly reducing the fatal outcomes, they also induce a decrease in infection rates (variable depending on diseases and vaccines, for example some estimates of effectiveness against infection on particular Covid-19 variants and vaccines were mainly above 50% 30 days after injection [25]). The value ranges from 0 to 0.5 %, meaning from 0 to 0.5 % of the susceptible population is set to be vaccinated daily. A high vaccination rate is strongly believed to be the key to a successful control of epidemics; hence, finding out the impacts of different vaccination rates on the overall infection process is meaningful.
- Social distancing level: a certain level of social distancing is being applied by governments worldwide, hoping to stop the spread of diseases like COVID-19. However, the effect of social distancing is not yet well known. Due to this, it would be an interesting experiment to find out the effectiveness of such measures on disease transmission dynamics. The level of social distancing ranges from 1 to 3. Each level refers to a percentage value that reduces the contact rate by 5. Hence, the actual percentage value of the social distancing ranges from 5 to 15 percent. The higher the intensity of social distancing, the lesser the chance people contact each other. In



practice, quantifying social distancing is complicated [26], thus this stylised mechanism which should be refined in case of a practical policy application of our model.

## 2.5 Model indicators

As mentioned previously, the ABM is constructed following a structure similar to the SIRV compartmental model. Therefore, the indicators consist of proportional numbers of agents in each compartment, including susceptible, infected, recovered, and vaccinated.

## 2.6 Model implementation and exploration

The model is implemented in NetLogo, which is a reasonable choice for such hybrid models in which visualisation is important. It is integrated into the OpenMOLE platform for model exploration [27] to carry out the numerical experiments described below. Source code of the model and exploration scripts are openly available on a git repository at <https://github.com/henry-kang-7/CASA0004>.

The pre-defined input parameter values of the model are the following: infectious-period  $\in \{2, 3, 4\}$  (weeks); immunity-period  $\in \{4, 5, 6\}$  (months); vaccination-rate  $\in \{0, 0.25, 0.5\}$  (%); social-distancing-levels  $\in \{1, 2, 3\}$  (levels).

This results in 81 different combinations of the input parameter values for a grid sampling. The model is set to run over 125 times for every combination of the input parameter values, and the output parameters' average values are obtained at the end. The size of the 95% confidence interval around the estimated average of a normal distribution of standard deviation  $\sigma$ , with  $n$  replications, is given by  $2 \cdot 1.96 \cdot \sigma / \sqrt{n}$ , and therefore  $n = 125$  ensures that the error on indicators is more than ten times smaller than the standard deviation.

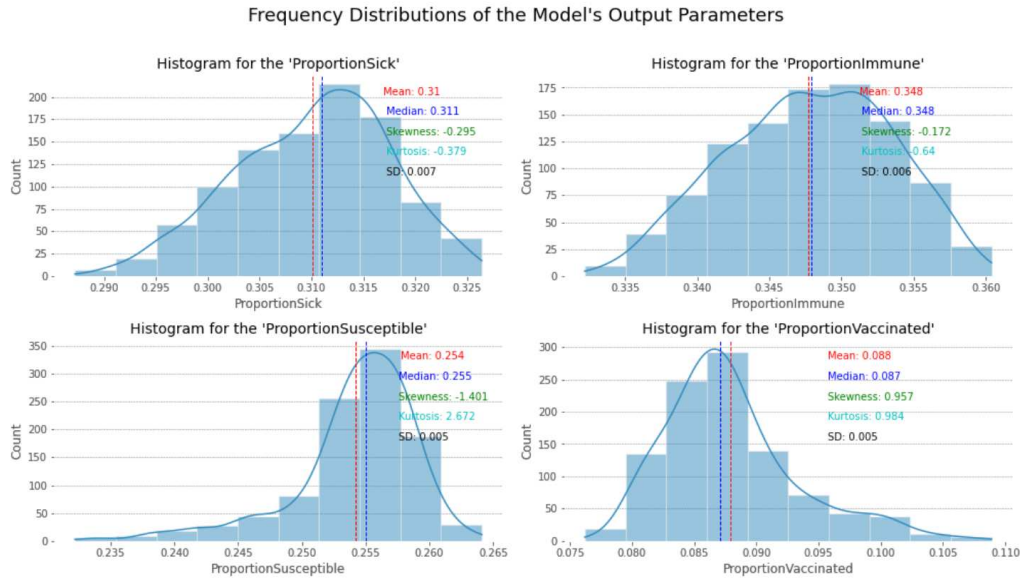
## 3 Results

We describe now numerical results obtained through the extensive exploration with the OpenMOLE platform.

### 3.1 Statistical distribution of indicators

The LHS (Latin hypercube Sampling) method was used to explore the input space of the model. A total of 10,000 samples were drawn during the process (corresponding to 4 days and 4h of CPU time for model execution), used to study statistical properties of model indicators.

The histograms in Fig. 4 illustrate the four different output indicators of the model. The histograms in the upper row comprise the 'ProportionSick' and the



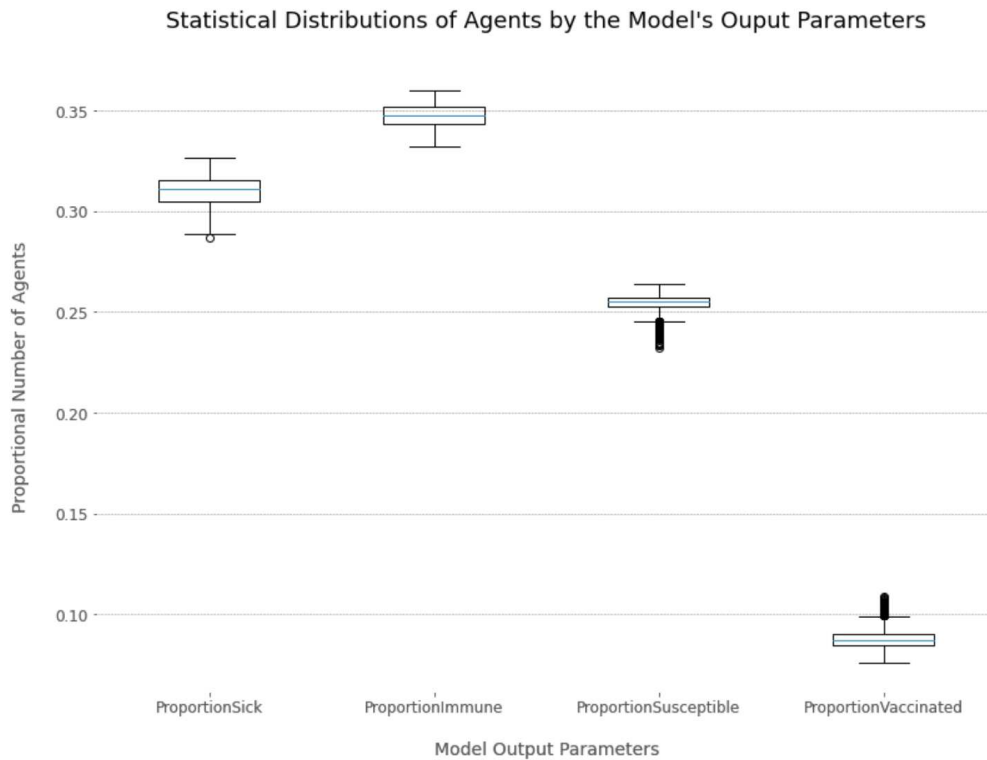
**Fig. 4** Histograms of model outputs.

'ProportionImmune' histograms, which appear symmetrical, but the skewness values indicate the traits of negative skewness on each histogram plot. Both kurtosis values are negative, suggesting that the distributions are thin-tailed. From this finding, we can assume that they generally follow characteristics of the platykurtic distribution. The 'ProportionSusceptible' and the 'ProportionVaccinated' histograms in the lower row appear negatively and positively skewed, respectively. The 'ProportionSusceptible' histogram is highly skewed to the left, which is indicated by the skewness value lesser than -1, while the 'ProportionVaccinated' histogram appears moderately skewed to the right as its skewness value is lower than 1. Furthermore, the kurtosis values of the two histograms are positive, which indicates that they both have heavier tails and are likely to follow the shape of the leptokurtic distribution.

The boxplots presented in Fig. 5 show different data ranges and medians of the four different output parameters of the model. The boxplot representing the 'ProportionImmune' appears to have the highest median among the groups, followed by the 'ProportionSick', 'ProportionSusceptible' and 'ProportionVaccinated', respectively. Outliers in the boxplots are present regarding the 'ProportionSusceptible' and 'ProportionVaccinated'. Based on the histograms shown in Fig. 4, these distributions have positive kurtosis values and are disproportionately distributed, indicating low standard deviation; hence, high chance of producing outliers.

### 3.2 Grid exploration

The total combination of the model's input parameters with predefined values is 81. The model run for each combination of the inputs is set to 125 times as

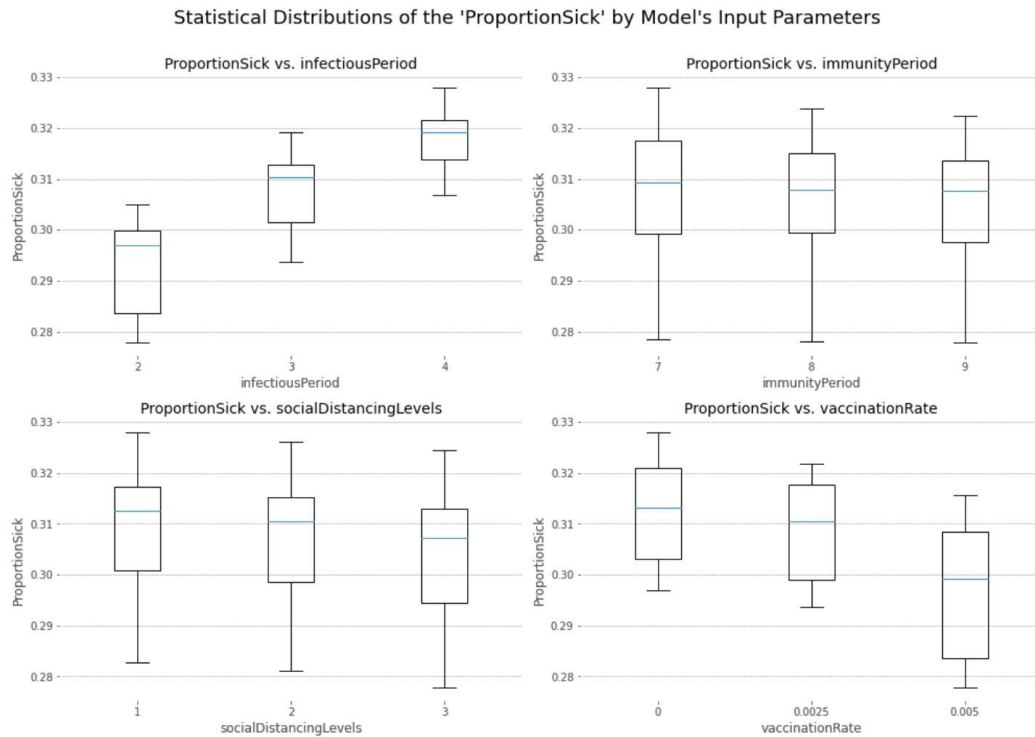


**Fig. 5** Box plots of model outputs

explained previously. This results in 10,125 model runs, for a total execution time of 4 days and 5 hours.

The box plots in Fig. 6 show statistical distributions of the 'ProportionSick', which are grouped by each input parameter value of the model. The range of each input parameter is indicated on the horizontal axis of the graphs. The first graph represents the grouped values of the 'ProportionSick' by the values of the 'infectiousPeriod'. It indicates that a unit increase in the 'infectiousPeriod' value increases the boxplot's overall median and max values. The boxplots related to the 'immunityPeriod' and the 'socialDistancingLevels' show a marginal decrease in the median of the boxplots per unit increase. The last graph is related to the 'vaccinationRate', which indicates that a 0.0025 (0.25 %) increase in the rate decreases the boxplots overall values. It indicates that vaccines have an apparent effect on successfully controlling the spread of diseases.

The correlation matrix illustrated in Fig. 7 displays the correlation between different model parameters. The 'ProportionSick' and the 'infectiousPeriod' appear to be highly correlated. According to the box plots shown in Fig. 6, every unit increase in the 'infectiousPeriod' also increased the 'ProportionSick' values. The rest of the input parameters, including the 'immunityPeriod', 'socialDistancingLevels' and 'vaccinationRate' are negatively correlated with the 'ProportionSick' at different intensity levels.



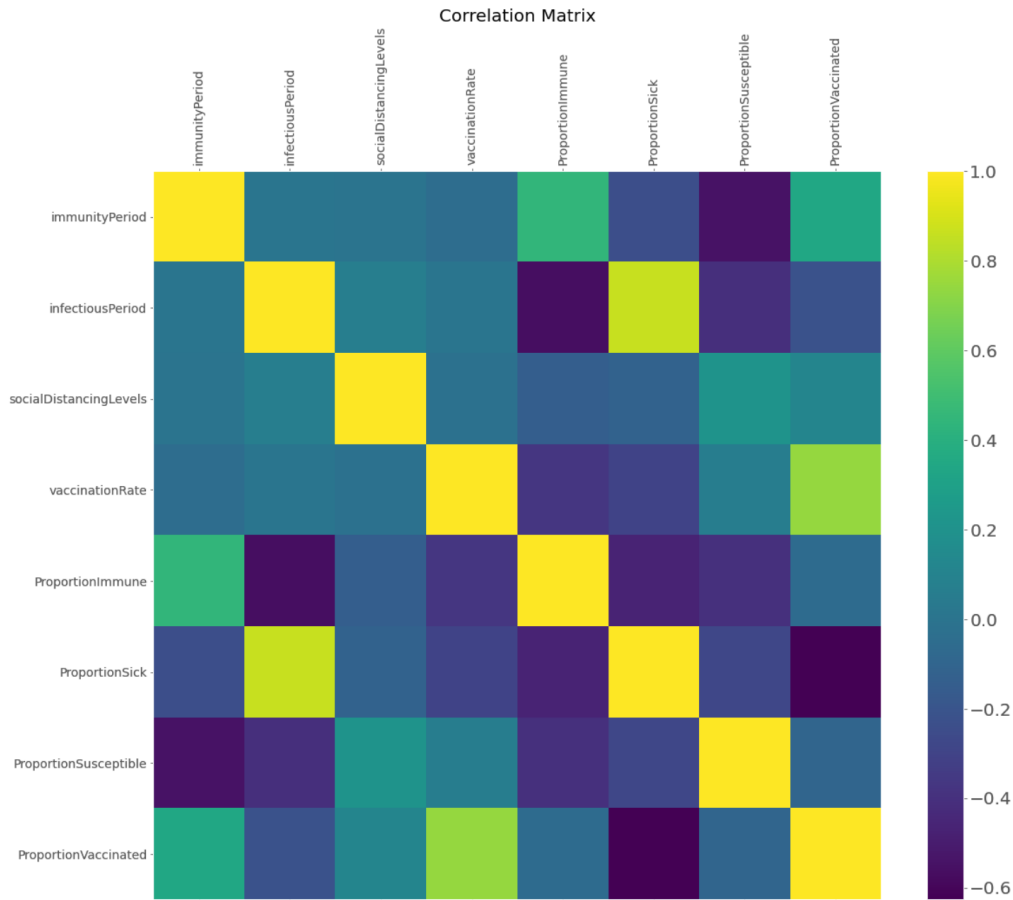
**Fig. 6** Box plots showing aggregate values of model outputs based on each value of inputs.

An OLS regression with the simulated data shows that the spread of infection mainly depends on the longevity of the infectious period. The longevity of the infectious period plays a vital role in disease transmission because a more extended infectious period allows an infected agent to interact with a more significant number of susceptible agents for an extended time. For example, an infected agent who is infectious for a single day is unlikely to infect a more significant number of susceptible agents than an infected agent who is infectious for a week. A more extended infectious period is, therefore, a dangerous factor that may rapidly increase the chances of infections, often creating a mass infection wave.

### 3.3 Model surrogate

Then, a random forest model is trained with the original data obtained by the LHS method during the first step of the analysis to forecast the proportional number of sick agents. This supervised machine learning method is versatile, easy to use and train, but also provides interpretable results on the relative importance of features, and we use it for these reasons.

In order to train the random forest model to predict the values of the 'ProportionSick', the original data is categorised into train, test, and validation data in the ratio of 7:1.5:1.5. Before fitting the model with the train data, various numbers of trees with their corresponding accuracy scores are calculated.

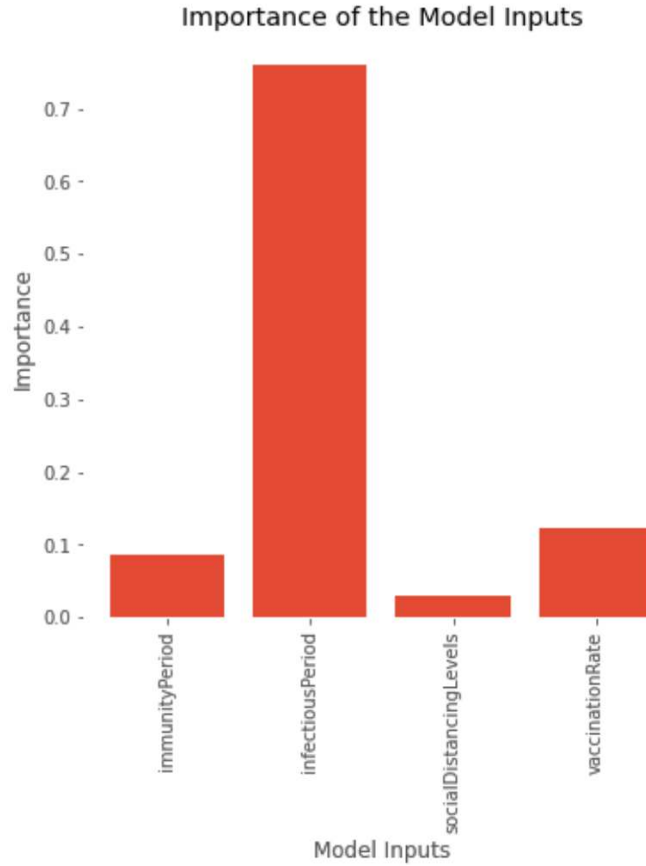


**Fig. 7** Correlation matrix of model parameters and indicators.

Then, this hyper-parameter for the model was optimised using the grid search method for a given set of estimator values, including 800, 900, 950, 1000 and 1050. As a result, 900 is the best value for optimising the random forest model. The random forest with the 900 decision trees was developed by fitting the training data to predict the values of the 'ProportionSick' and including all four input parameters of the model.

We obtain a MAE (how big of an error we can expect from the forecast on average) of 0.001, which implies that the average errors between the predicted and actual values are marginal; hence, the predictions are considered highly accurate. Moving on to the MAPE, the value of 0.363 indicates that, on average, the forecast is off by 0.363%. This means that the accuracy of the forecast is as high as 99.637%. Lastly, the value of the RMSE is also 0.001, which indicates that the difference between the predicted and actual values is marginal.

The random forest model is furthermore useful to quantify the importance of the model's input parameters, shown in Fig. 8. The 'infectiousPeriod' is by far the most significant input parameter of the model, with a value of approximately 0.75, followed by the 'vaccinationRate' with a value of slightly over 0.1.



**Fig. 8** Feature importance of input parameters obtained with the random forest surrogate.

**Table 1** Saltelli total order indices

| Outputs               | infectiousPeriod | immunityPeriod | vaccinationRate | socialDistancingLevels |
|-----------------------|------------------|----------------|-----------------|------------------------|
| ProportionImmune      | 0.72             | 0.333          | 0.665           | 0.7                    |
| ProportionSick        | 0.692            | 0.09           | 0.365           | 0.143                  |
| ProportionVaccinated  | 0.516            | 0.319          | 0.587           | 0.337                  |
| ProportionSusceptible | 0.341            | 0.066          | 0.92            | 0.277                  |

The 'immunityPeriod' and the 'socialDistancingLevels' have an importance rate below 0.1, showing less importance than the previous two parameters.

### 3.4 Global sensitivity analysis

The previous results of parameter importance can be compared with an other method to quantify parameter's influence on indicators. The global sensitivity analysis method, introduced by [28], gives a broad summary of such an influence. In order to generate an accurate result for the sensitivity analysis, a total of 60,000 samples were run (25 days 7 hours of model runtime).

The total-order indices shown above indicate that the 'infectiousPeriod' has the most significant impact on the values of the 'ProportionSick' followed

by the 'vaccinationRate', 'socialDistancingLevels' and 'immunityPeriod' respectively. We find a very low effect of the immunity period on the proportion of sick and susceptible, meaning that uncertainties on this value will have a limited impact. The vaccination rate is also a critical input parameter of the model that plays a vital role in controlling the spread of diseases because the vaccine is a significant source of protective measures that can effectively stop viruses from transmitting one agent to another [29].

### 3.5 Multi-objective model optimisation

In attempting to control the spread of epidemics, many problems involve multiple objectives which cannot simply be described as the more, the better or the lesser, the better; instead, each objective has an ideal target value, and the main goal is to be as close as possible of the targeted value. To optimise the values of the input parameters, we define two objectives that are expected to be minimised. Firstly, we expect the proportional number of sick agents to be optimised, which is directly linked to the main interest of our research question. Secondly, we expect the proportional number of vaccinated agents to be optimised. Vaccination is a primary preventive measure that plays a vital role in successfully controlling the spread of diseases and even has the potential to achieve herd immunity, while other measures such as social distancing can only slow down the process of disease spreading instead of stopping them [16]. However, vaccines are not always available for multiple reasons, such as the high manufacturing and transportation costs [30]. This is where decision-makers often face a dilemma because attempting to control epidemics will cost money and the best option which they might have at this point is to find a good balance between the use of vaccines and the proportional number of sick agents, respecting the economic budget.

The way to find the results that provide a good approximation of the Pareto frontier with acceptable trade-offs between the identified objectives is by using the NSGAI algorithm. The NSGAI is a commonly used type of multi-objective optimisation algorithm which includes a fast non-dominated sorting, crowding distance assignment and a sorting procedure [31].

The Pareto Front consisting of non-dominated solutions relative to the 'ProportionSick' and 'ProportionVaccinated' is drawn after running a total of 10,000 generations of the OpenMOLE implementation of the algorithm.

The Pareto Front shown in Fig. 9 illustrates the optimal solution set of the 'ProportionSick' and the 'vaccinationRate' as two objectives. The value of the 'ProportionSick' appears to decrease rapidly as the value of the 'vaccinationRate' increases. The slope of the line is generally steep throughout the graph. Plus, the difference between the values of the 'ProportionSick' with and without the vaccination is clear. The non-linearity of the front is an interesting feature, witnessing the underlying complexity when exploring policy trade-offs.

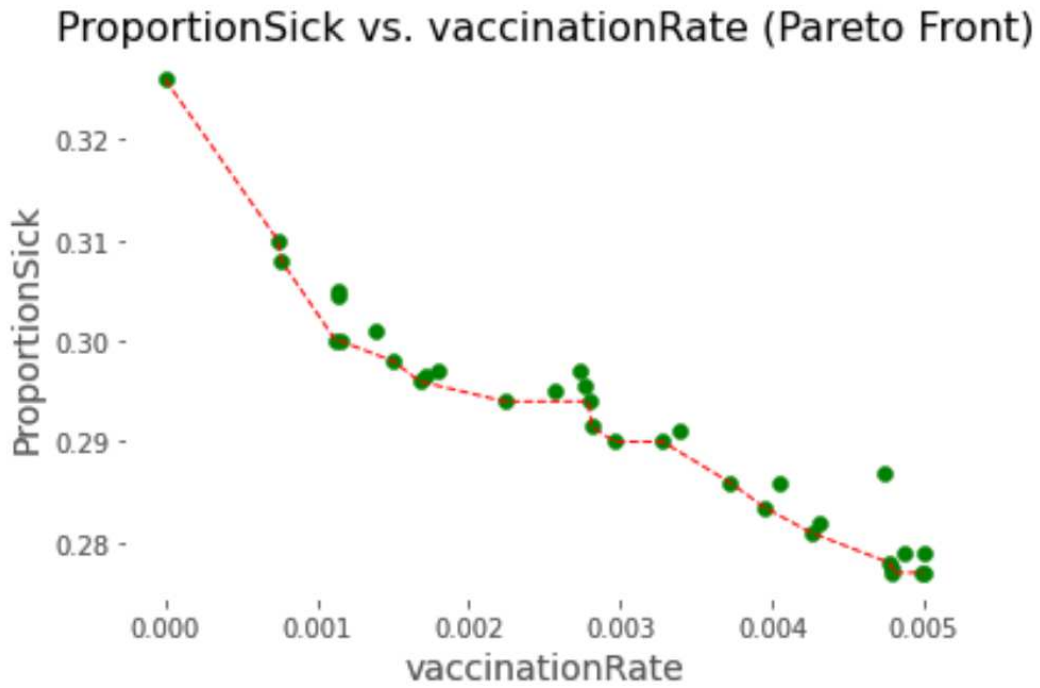


Fig. 9 Pareto front of proportion of sick vs. vaccination rate.

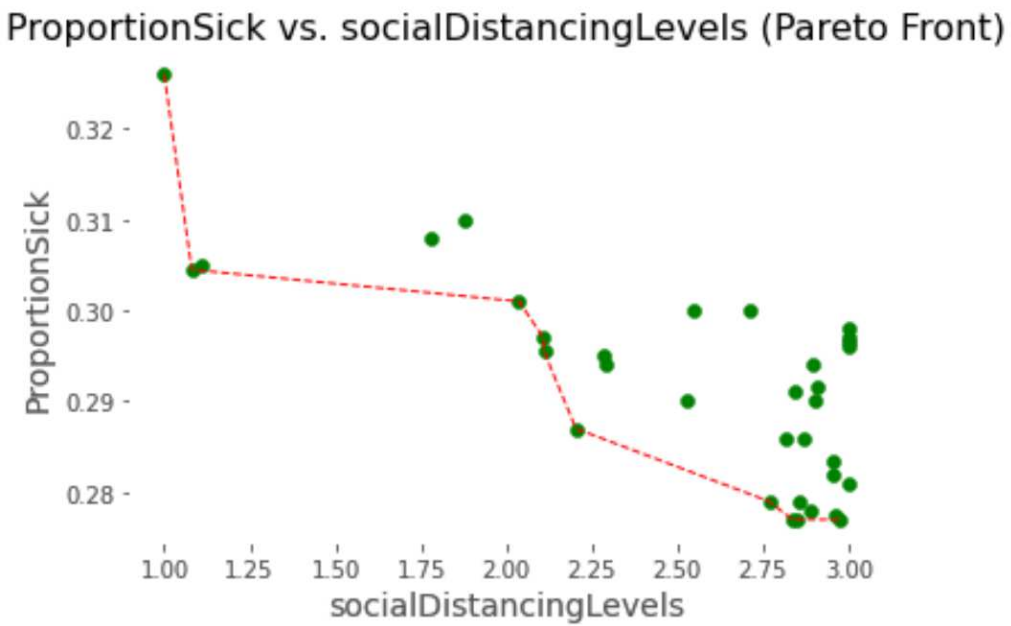


Fig. 10 Pareto front of proportion of sick vs. social distancing levels.



The Pareto Front, visualised in Fig. 10, is formed with the optimal values of the 'ProportionSick' and the 'socialDistancingLevels' as two objectives. It indicates a clear sign of a decrease in the value of the 'ProportionSick' as the value of the 'socialDistancingLevels' increases. In some regions of the front, a quasi-vertical increase means that much control may be gained at a very low social cost, what has implications when finding the compromise between the epidemic impact and the social impact of lockdowns. The social distancing level is indeed an essential input parameter that acts as a preventive measure against infectious diseases in our model. Social distancing must be implemented with care because if the intensity is too high, it prohibits necessary human interactions, which reduces the productivity rate, negatively impacting the economy in general [32]. Due to this, finding a good balance between the social distancing level and the proportional number of sick agents would be essential in ensuring a healthy economy while keeping the number of infected cases low.

## 4 Discussion

The diverse numerical experiments, applying various model exploration and validation methods, can be used to provide useful policy insights in this stylised case. They are not directly applicable to real world situations, as a slightly more data-driven approach would be needed, but this already suggests how such methods can compensate for model simplifications. Thus, we suggest that our work is a proof-of-concept of how such systematic model exploration may hinder issues due to parameter uncertainty, or the lack of data, for similar epidemiological models.

Some limitations of this work can be given. First, the population size of the simulation had to be limited to a small number, mainly due to the expensive computational cost and time. In order to fasten the process, these executions have been distributed using OpenMOLE, but the limitation in the level of detail and granularity is a recurring issue with ABMs. We still need to investigate if the trade-off chosen here is reasonable.

The current number of input parameters is limited to only four parameters. However, there can be more numbers of input parameters introduced in our model to generate more dynamic results which can better reflect real-life situations. Furthermore, the current model can be extended with more compartments that can describe various states of an agent in the infection process, such as death or exposure [33]. The model is also flexible in that it can add many more characteristics of agents, such as age, occupation, vaccination record and gender, which can be considered in the process of interaction and infection.

As a future work, regarding the stylised parametrisation taken here for several aspects of model setup, a more advanced sensitivity analysis relaxing these parameters could be done. More particularly, quantifying the role of geography (transport geography, but also company size distribution and locations),

is an interesting prospect, as new methods to achieve this have recently been developed [34].

## 5 Conclusion

This study proposed various model assessment methods to explore results obtained with a stylised ABM we introduced, based on the SIRV compartmental model to study the spread of diseases. Through the model exploration, this study analysed the impacts of the pre-defined input parameters, which included the infectious period, immunity period, daily vaccination rate and social distancing level, on the spread of infectious diseases. We provide thus a proof-of-concept of how model exploration and validation methods can be a powerful tool to design and optimise health policies.

## References

1. K.A. Glatter, P. Finkelman, *The American journal of medicine* **134**(2), 176 (2021)
2. L. Saker, K. Lee, B. Cannito, A. Gilmore, D.H. Campbell-Lendrum, et al., (2004)
3. R. Boucekchine, A. Carvajal, S. Chakraborty, A. Goenka, *Journal of Mathematical Economics* **93**, 102498 (2021)
4. S.J. Bickley, H.F. Chan, A. Skali, D. Stadelmann, B. Torgler, *Globalization and health* **17**(1), 57 (2021)
5. J. Kaszowska-Mojša, P. Włodarczyk, A. Szymańska, *Entropy* **24**(1), 126 (2022)
6. A. Morabia, *Epidemiology (Cambridge, Mass.)* **24**(2), 179 (2013)
7. K. Dietz, J. Heesterbeek, *Mathematical biosciences* **180**(1-2), 1 (2002)
8. P. Bosetti, C. Tran Kiem, A. Andronico, V. Colizza, Y. Yazdanpanah, A. Fontanet, D. Benamouzig, S. Cauchemez, *BMC medicine* **20**(1), 33 (2022)
9. F. Miksch, B. Jahn, K.J. Espinosa, J. Chhatwal, U. Siebert, N. Popper, *PloS one* **14**(8), e0221564 (2019)
10. A. Ramani, A. Carstea, R. Willox, B. Grammaticos, *Physica A: Statistical Mechanics and its Applications* **333**, 278 (2004)
11. G. Chowell, M. Miller, C. Viboud, *Epidemiology & Infection* **136**(6), 852 (2008)
12. D. Chumachenko, V. Dobriak, M. Mazorchuk, I. Meniailov, K. Bazilevych, in *2018 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)* (IEEE, 2018), pp. 192–195
13. J.M.A. Minoza, V.P. Bongolan, J.F. Rayo, arXiv preprint arXiv:2101.11400 (2021)
14. G. Fekadu, F. Bekele, T. Tolossa, G. Fetensa, E. Turi, M. Getachew, E. Abdisa, L. Assefa, M. Afeta, W. Demisew, et al., *International Journal of Physiology, Pathophysiology and Pharmacology* **13**(3), 86 (2021)
15. S. Kwon, A.D. Joshi, C.H. Lo, D.A. Drew, L.H. Nguyen, C.G. Guo, W. Ma, R.S. Mehta, F.M. Shebl, E.T. Warner, et al., *Nature Communications* **12**(1), 3737 (2021)
16. M. Bicher, C. Ripplinger, G. Schneckenreither, N. Weibrecht, C. Urach, M. Zechmeister, D. Brunmeir, W. Huf, N. Popper, *Scientific Reports* **12**(1), 2872 (2022)
17. J. Raimbault, D. Pumain, *Geographical Modeling: Cities and Territories* **2**, 125 (2019)
18. E.G. Baquela, A.C. Olivera, in *Humanitarian Logistics from the Disaster Risk Reduction Perspective: Theory and Applications* (Springer, 2022), pp. 273–291
19. J. Wu, R. Dhingra, M. Gambhir, J.V. Remais, *Journal of The Royal Society Interface* **10**(86), 20121018 (2013)
20. R.C. Poonia, A.K.J. Saudagar, A. Altameem, M. Alkhatami, M.B. Khan, M.H.A. Hasanat, *Life* **12**(5), 647 (2022)
21. K.W. Axhausen, in *Institute of Space and Earth Information Science (ISEIS) seminar at the Chinese University of Hong Kong* (IVT, ETH Zurich, 2021)

22. N. Ielapi, N. Licastro, M. Provenzano, M. Andreucci, S.d. Franciscis, R. Serra. Cardiovascular disease as a biomarker for an increased risk of covid-19 infection and related poor prognosis (2020)
23. R.R. Wilkinson, K.J. Sharkey, *Physical Review E* **97**(5), 052403 (2018)
24. J. Reyes-Silveyra, A.R. Mikler, *Theoretical Biology and Medical Modelling* **13**, 1 (2016)
25. I. Mohammed, A. Nauman, P. Paul, S. Ganesan, K.H. Chen, S.M.S. Jalil, S.H. Jaouni, H. Kawas, W.A. Khan, A.L. Vattoth, et al., *Human vaccines & immunotherapeutics* **18**(1), 2027160 (2022)
26. P. Caley, D.J. Philp, K. McCracken, *Journal of the Royal Society Interface* **5**(23), 631 (2008)
27. R. Reuillon, M. Leclaire, S. Rey-Coyrehourcq, *Future Generation Computer Systems* **29**(8), 1981 (2013)
28. A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, *Global sensitivity analysis: the primer* (John Wiley & Sons, 2008)
29. C.B. Storlie, B.D. Pollock, R.L. Rojas, G.O. Demuth, P.W. Johnson, P.M. Wilson, E.P. Heinzen, H. Liu, R.E. Carter, E.B. Habermann, et al., in *Mayo Clinic Proceedings*, vol. 96 (Elsevier, 2021), vol. 96, pp. 1890–1895
30. S. Plotkin, J.M. Robinson, G. Cunningham, R. Iqbal, S. Larsen, *Vaccine* **35**(33), 4064 (2017)
31. K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, *IEEE transactions on evolutionary computation* **6**(2), 182 (2002)
32. S. DeLuca, N. Papageorge, E. Kalish, Johns Hopkins coronavirus resource Center (2020)
33. B. Reyn e, N. Saby, M.T. Sofonea, *Anaesthesia, Critical Care & Pain Medicine* **41**(1), 101017 (2022)
34. J. Raimbault, C. Cottineau, M. Le Texier, F. Le Nechet, R. Reuillon, *Journal of Artificial Societies and Social Simulation* **22**(4) (2019)