RESEARCH

A hybrid compartmental model with a case study of COVID-19 in Great Britain and Israel

Greta Malaspina[1*] , Stevo Racković[2] and Filipa Valdeira[3]

*Correspondence:
greta.malaspina@dmi.uns.ac.rs
[1]Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia
Full list of author information is available at the end of the article

Abstract

Given the severe impact of COVID-19 on several societal levels, it is of crucial importance to model the impact of restriction measures on the pandemic evolution, so that governments are able to make informed decisions. Even though there have been countless attempts to propose diverse models since the rise of the outbreak, the increase in data availability and start of vaccination campaigns calls for updated models and studies. Furthermore, most of the works are focused on a very particular place or application and we strive to attain a more general model, resorting to data from different countries. In particular, we compare Great Britain and Israel, two highly different scenarios in terms of vaccination plans and social structure. We build a network-based model, complex enough to model different scenarios of government-mandated restrictions, but generic enough to be applied to any population. To ease the computational load we propose a decomposition strategy for our model.

Keywords: COVID 19; SIR; Social Networks

1 Introduction

When facing a severe pandemic, as we have seen recently with COVID-19, it is fundamental for governments to take actions to prevent or contain its spread, while taking into account their impact from the socio-economic point of view. In this context, modelling the evolution of the spreading and the impact of different kinds of containment measures becomes an indispensable tool. The development of vaccines and the early start of vaccination plans is also a crucial factor in the evolution of the pandemic and it was not available in the initial studies, calling for new models that may lead to different conclusions regarding what could be the best course of action.

Mathematical models where one can conduct simulations considering different scenarios regarding implemented measures are therefore of the utmost importance in order to make informed decisions. Consequently, since the beginning of the pandemic, there has been a vast number of papers and simulations concerning specifically the outbreak of COVID-19 (see, for instance, review paper [1] and references therein). The approaches differ both in regard to properties of the population taken into account — age groups, density, size and connectivity — but also in the considered government-imposed restrictions — lockdown, closing of schools and workplaces or restrictions in public transport,

the mandatory wearing of masks and, finally, medical procedures, including mass vaccination. Among the most common approaches are compartmental [2] and network-based models [3].

Compartmental models   In compartmental models, the population is partitioned into epidemiologically relevant classes. Then, a system of differential equations regulates how the cardinalities of such classes change over time and how they influence one another. A typical model in this class is the Kermack-McKendrick SIR model [2, 4], given by the following equations

$$\begin{cases} \frac{d}{dt}S(t) = -\frac{\beta}{N}S(t)I(t), \\ \frac{d}{dt}I(t) = \frac{\beta}{N}S(t)I(t) - \gamma I(t), \\ \frac{d}{dt}R(t) = \gamma I(t), \end{cases} \tag{1}$$

where $S(t)$, $I(t)$, and $R(t)$ represent the number of susceptible (S), infected (I), and recovered (R) individuals at time t, while parameters β and γ denote the infection and recovery rate, respectively.

Methods of this kind are of great interest because their simple structure makes them suitable for a wide set of problems, and their deterministic nature allows some theoretical considerations regarding the asymptotic behavior of the number of infected individuals, as well as the existence of equilibrium points of the system. Furthermore, they can be easily modified to introduce additional classes or more complex equations, in order to get a more comprehensive description of the situation. For the COVID-19 outbreak, several extensions of this model have been proposed in the literature, taking into consideration, for instance, the presence of a latency period of the virus [5], the adoption of masks [6] and vaccination campaigns [7], or restrictive measures and human behavior [8].

However, these methods usually rely on several assumptions regarding the population under study and the disease itself. Typical hypotheses are that the population is large and closed (i.e. there are no external influences on the system) and that the natural birth and death rates can be disregarded (e.g. because the considered time interval is short enough). Moreover the coefficients involved in the equations, e.g. β and γ in (1), are usually assumed to be deterministic, and the individuals are assumed to be uniformly distributed in space, so that in a given time interval each individual has an equal probability of meeting any other. Overall, to provide an accurate estimation, these models require good knowledge of the disease-related parameters and of the implications of restrictive measures, such as social distancing and travel limitations, over the average number of contacts between individuals. In practice, though, great variability has been observed in the transmission rate and other important parameters of COVID-19 [9], so it may be difficult to estimate them accurately. At the same time, the inherent complexity of interactions between individuals and the large variety of restrictive measures enforced by governments at different times makes the assumption about the uniform distribution of the population impractical.

Network-based models   To overcome this last issue, several network-based models [10] have been developed. In this setting, the population and the direct interactions between individuals are represented by a graph, where the infection spreads. The nature of the

equations behind these methods is usually analogous to those behind compartmental models. The fundamental difference is that at each time step, the number of contacts between individuals of different classes does not simply depend on the size of said classes (e.g. the term $S(t)I(t)/N$ in (1)) but also on the connections between nodes and the states that neighboring nodes occupy. The use of a network to model direct interactions between individuals provides a significantly more precise estimate of the number of contacts at each time-step, which can lead to better predictions of the spread of the infection even when accurate estimations of the disease-related parameters are not available. On the other hand, as these approaches require working with networks with as many nodes as the number of individuals in the population, they can become prohibitively expensive when considering large populations such as large regions or entire countries. Therefore, they are usually employed to model the infection at the city or community level [3, 11]. Furthermore, individual data is seldomly available and expensive to obtain, while compartmental models can be applied to easily accessible data.

Our approach   The model that we consider in this paper comes from an extended SIR network-based approach. Regarding the compartmental component, besides the three standard states (S, I and R) in which a node can be at a given time, we include several additional states and inter-states that take into account factors like the incubation period of the virus and quarantine periods. These are crucial in the evolution of COVID-19. We also use different parameters for individuals belonging to different states, such as different susceptibility to symptoms and different recovery and death rates based on the current state. The parameters describing the virus behaviour are chosen within the range proposed in the literature and according to the provided data.

The population is approximated with a set of graphs where each graph corresponds to a densely populated area (city) and each node represents an individual person.[1] The connections between nodes change over time according to the imposed governmental restrictions. Moreover, a SIR-like approach (analogous to the first equation of (1), with small β) is used to model the interaction between different graphs. This can be interpreted as having the population partitioned into densely populated cities and assuming that the number of interactions between individuals belonging to different cities is considerably smaller than the number of interactions between individuals coming from the same city.

Furthermore, we include in our model data from imposed measures, allowing us to include vaccination programs, contact tracing and testing, as well as movement restrictions. The latter is modeled by carefully selecting the topology of the network, while the former is introduced by altering the transitions and states of the existing models.

When considering large populations (e.g. entire countries) there is usually a trade-off between accuracy and computational time: running a model with a larger population size (that is, closer to the actual size of the country) gives more accurate results but increases the computational cost. As the use of the graphs allows us to get a more accurate model of the local contacts between individuals and the strategy used to estimate the number of contacts between individuals in different cities is inexpensive, this partitioned approach helps us to increase the population size and improve the accuracy while keeping the computational load moderate.

---

[1]Note that, since we take only a fraction of population, in our simulation, one node actually stands for more than one person.

1.1 Related work

Model   Common extensions of the SIR model are SEIR [12], with an additional exposed (*E*) state, accounting for the period of exposure or virus incubation before developing the disease, and SEIRD [13], with an additional *D* state, accounting for deceased individuals. However, we base our model on a further extension, SEIRS+ [11], where hospitalization and quarantine are also considered. These states are crucial for the particular COVID-19 outbreak, as they allow us to compare the number of hospitalized people and to model the effects of quarantine mitigation measures.

Networks   One common approach for modeling the virus spread in literature is to use multiplex networks [14, 15], where each overlapping network describes a type of contact (household, workplace). In these cases, each agent usually has associated a series of individual properties (location, sex, age) and each day must be simulated in detail, by having the agent at each specific location and verifying the contacts in place. In this paper, the distinction is only made between children and adults, as the former will not get vaccinated, but this has no impact on the network. Besides, we just change the topology of a single network according to the combination of restrictions in place, avoiding the need to model each kind of specific contact and in this sense our work is closely related to [3]. In [3] the authors propose a SEIR network-based model with four different networks representing the population depending on the restrictions enforced at a given time (in particular, they distinguish between the cases when no restrictions are in place, big events are forbidden, and social distancing or lockdown is enforced). In our model, we consider more than 4 different networks, thus allowing for better distinction of the measures in place. Furthermore, the underlying state model is SEIRS+, instead of SEIR.

Vaccination   Since the start of vaccination campaigns, several models have been proposed in the literature that take into account the effect of vaccines on the spread of the virus [7, 16–18]. Most of these works focus specifically on vaccinations, analysing how different vaccination rates impact the number of cases and studying the relationship between vaccine efficacy and the number of administered vaccines, in order to determine if herd immunity can be achieved given a set of policies. In this paper, we take into account the influence of vaccines on both the probability of contracting the virus and the probability that, once infected, a person develops the symptoms. However, we prefer to keep a broader point of view, and not focus solely on the effects of vaccination, taking into account the interaction between diverse measures.

Transition parameters   The values of transition parameters between each model state are still not well defined for COVID-19. At the beginning of the pandemic there were several efforts to find transmission parameters, especially focusing on China (or more specifically Wuhan) [19–22]. These initial estimates, however, have been continuously updated as detailed statistics became available for different countries and geographical regions. In general, contact rate is considered a decreasing function of time [20] since it is strongly influenced by the public awareness state, hence many papers propose fitting different values for different days of the pandemic [7]. It is also common to consider different parameters depending on the individual's characteristics (parameters calibrated by age are found to be an important factor on a good prediction [23]). In our case, the parameters are either

constant for the entire timeline or data-dependent. We also compare our model for two different countries — Israel and Great Britain — in order to ensure a good generalization of the model. On the other hand, most studies focus on singular particular locations [22, 24–26], thus allowing for finer parameter tuning — however, one should keep in mind that such an approach can easily lead to overfitting.
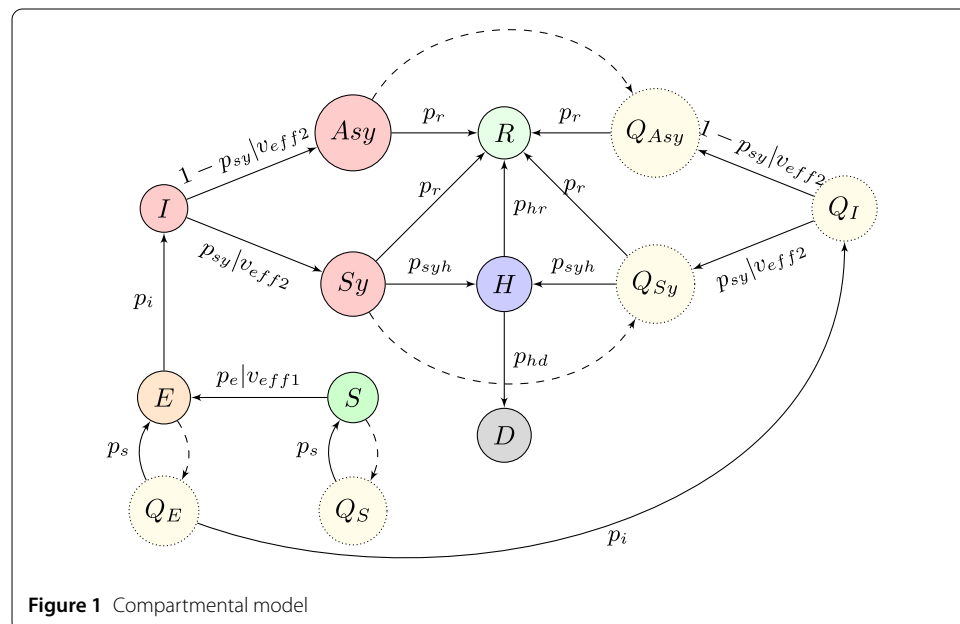
Population   In order to model the diversity of the population and contacts among individuals, some models resort to census data [15, 23, 27, 28] and/or mobile location [15, 29]. In our case, we just require the percentage of the adult population and the measures in place to predict the model evolution, thus leading to an easier adaptation of the model to different settings and populations. While we assume that some of the parameters are country-dependent and they still need to be fitted to the specific country data, we do not require any additional information, such as census data.

1.2 Contributions

The contributions of this work are three-fold. In the first place, we propose a model that combines a typical compartmental model with a graph-based one. While both these models are well known and widely studied in the literature, we noticed that the approach of combining the two is largely overlooked in the prior literature. In this paper, we combine the extended SEIRS+ network with the standard compartmental model, to exploit the respective advantages of each. In this sense, we can go beyond a single city or region and mimic several densely populated areas with a relatively low level of connection. Since the interactions between separate areas are cheap to incorporate, increasing the number of regions (and so the simulated population) leads to only a linear growth of the computational cost. This allows the model to benefit from the higher accuracy given by modeling the population with a network, while avoiding the high computational cost that usually derives from the employment of networks. When suitable data is available, this also allows the modeling of local (e.g. regional) policies, without major alterations to the model here presented. Furthermore, while our choice of the underlying graphs to model the population is motivated in a way similar to [3], we noticed that the authors of [3] did not provide details regarding modelling the transition between graphs. In specific, if the graphs for different levels of restrictions are modelled independently, the transitions from a dense to a sparse graph can produce the counter-intuitive result of a sudden increase in the number of infected nodes. This is due to the set of newly introduced edges. To overcome this problem, we build the graphs on top of each other, starting from the most sparse up to the densest one.

A further contribution lies in the used data. We took into consideration two countries with significantly different structures and contrasting policies of containment, which have a great influence on the local spread of the disease. This allows us to strive for a more general model, staying away from overfitting. While we do have some country-specific parameters, they all regard population behavior, where it is reasonable to require different models. The virus-related parameters are kept the same for both countries enforcing our confidence in their estimates.

Finally, we use the proposed model to investigate a number of hypothetical scenarios. We consider different governmental measures, such as levels of restrictions, vaccination

rates and testing policies, and we analyze how these choices affect the evolution of the contagion, what are the joint effects when measures are combined and how they compare with each other.

1.3 Outline

This document is organized as follows. In Sect. 2 we describe our model and the datasets. In Sect. 3 we describe the process of fitting the model parameters and compare our prediction with ground truth data. In Sect. 4 we test the impact of several measures by setting up a variety of different hypothetical scenarios. Finally, in Sect. 5 we discuss the behaviour of the produced model and obtained results.

2 Data and model

The proposed model consists of several connected graphs, where each node of a graph represents a single individual in the population, and the edges between nodes correspond to contacts between people. At every day of the simulation a network-based SEIRS+ model is applied to each graph. The influence of different graphs on each other is simulated in the manner of a typical differential compartmental model. This section describes in detail how the model was built and how the COVID-19 statistics data was incorporated to fit the model parameters.

2.1 Compartmental model

The design of our compartmental model is depicted in Fig. 1. Circles represent the possible states that each individual can occupy at any given time, and arrows represent possible transitions between states, with the parameters regulating such transitions. Dashed arrows indicate that a given transition between two states is not active during the whole simulation — these are the transitions concerning the quarantine, hence we only activate them for the periods when the government policies demand so. We consider the following states:

Figure 1 Compartmental model

- S: susceptible individuals — people that have not been exposed to the virus but that could contract it upon contact with an infected person;
- E: exposed individuals — people that have been in contact with an infected individual and will become infected after the latency period of the virus has passed;
- $I = Sy \cup Asy$: infected individuals, divided into symptomatic and asymptomatic;
- H: hospitalized individuals;
- R: recovered individuals — people that had the virus in the past, but that are now no longer infected nor susceptible;
- D: deceased individuals;
- Q_S, Q_E, Q_{Asy}, Q_{Sy}: individuals that are susceptible, exposed, asymptomatic or symptomatic and that are currently in a state of quarantine.

For the sake of simplicity, let us first observe the scheme in Fig. 1 under the setting of no imposed quarantine. For each individual in the population we have a Markov chain [30] with a source node S and two sink nodes R and D. Node I is an intermediate state, and as soon as an individual jumps into I it transitions to Sy with probability p_{sy} or to Asy otherwise. To account for vaccination, we introduce two coefficients in our model: v_{eff1} and v_{eff2}, appearing in some of the transition edges (see Sect. 2.2.1). Further, it is assumed that each individual can only get hospitalized if they previously manifested symptoms, and can only die if they were already hospitalized (that is, the only incoming edge to the H state is from Sy, and the only incoming edge to D is from H). If quarantine policies are in place, we see the analogous process but with Q_{Asy} and Q_{Sy}. The following is a detailed list of the transition coefficients.

- p_e: probability that a susceptible person who has been in direct contact with an infected individual becomes exposed. We assume that this parameter changes over time, as to represent the variation in awareness of the population regarding the risks of transmission and the compliance of individuals with the restrictions (see Sect. 2.2.4). If the considered individual has been vaccinated, this probability is altered by the vaccine efficacy parameter v_{eff1} (see Sect. 2.2.1).
- p_i: probability that an exposed individual becomes infected — this can be interpreted as the inverse of the expected latency period of the virus.
- p_{sy}: probability that an infected person develops symptoms. In case the considered individual is vaccinated, this probability is altered by the vaccine efficacy parameter v_{eff2} (see Sect. 2.2.1).
- p_{syh}: probability that a symptomatic individual gets hospitalized.
- p_r: probability that an infected individual recovers — this can be interpreted as the inverse of the average recovery time for people affected by the virus.
- p_{hr}: probability that a hospitalized individual recovers — this can be interpreted as the inverse of the average recovery time for hospitalized people.
- p_{hd}: probability that a hospitalized individual dies.
- p_s: probability that a quarantined person leaves the quarantine.

The transition between a class X and its corresponding quarantined class Q_X is modeled according to different possible testing policies, as well as contact tracing procedures that may be in place (see Sect. 2.2.2).

2.2 Datasets
We used available data [31] for two countries, Great Britain (GBR) and Israel (ISR), over the course of more than a year.

Since the goal is to study the effects of different policies on the evolution of the epidemic, we split the data attributes into two meaningful categories. The first group of attributes can be seen as potential inputs for the predictive model — these are mainly governmental restrictions and policies in place, as well as the number of vaccinations and tests. The second group consists of attributes containing the number of infected, hospitalized, dead individuals, etc. These are potential ground truth (GT) values for the model. From this group, we mainly focus on the number of hospitalized and deceased individuals as GT. The choice lies in the fact that these are the most reliable values, since not all cases of the virus were actually detected, and there is no perfect way to trace recovered people. (We also do not account for particular cases of ICU (Intensive Care Unit) or people undergoing ventilation in our model because the counts are not consistent across the countries, so we do not make this distinction). Even though we do not directly use the count of confirmed cases in our model, we keep in mind that the number of predicted cases should not differ significantly from the available count. Figure 2 shows the GT values for both countries as a percentage of the total population. A cumulative number of cases is presented in the log scale to compensate for the exponential growth of the curve. For GBR we see three distinctive waves, with the second one having two peaks, while for ISR there are four. We further notice that both the numbers of hospitalized and deceased cases are lower for ISR than for GBR.
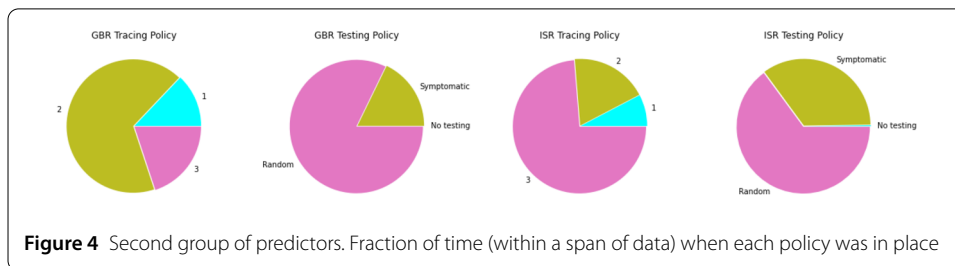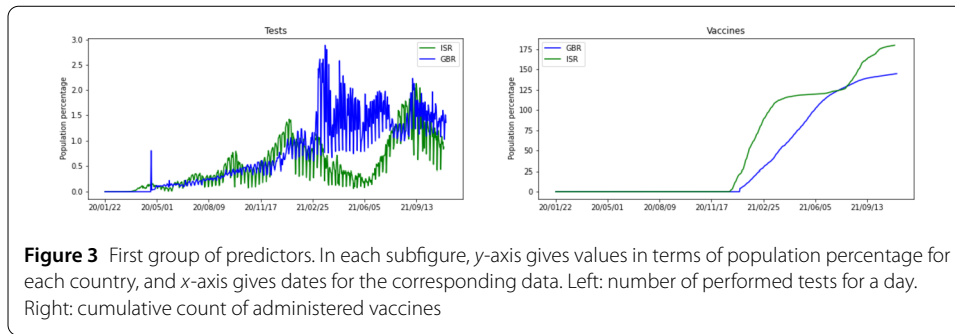
As the predictors for our model we consider the following 3 groups of attributes:

- Vaccination and testing — we have access to the number of administered vaccines and performed tests per day, as shown in Fig. 3. Both countries started vaccination campaigns around the same day, but vaccination in ISR had a faster pace and added up to a higher overall number of doses compared to GBR. On the other hand, GBR had a higher and more consistent number of performed tests.
- Testing and contact tracing policies — these are two ordinal attributes and Fig. 4 illustrates that in both countries a modest level of contact tracing was employed for most of the time.
- Governmental restrictions — in specific *'stay_home_restrictions'*, *'school_closing'* and *'workplace_closing'*, were used to build underlying contact networks and a varying exposure parameter.
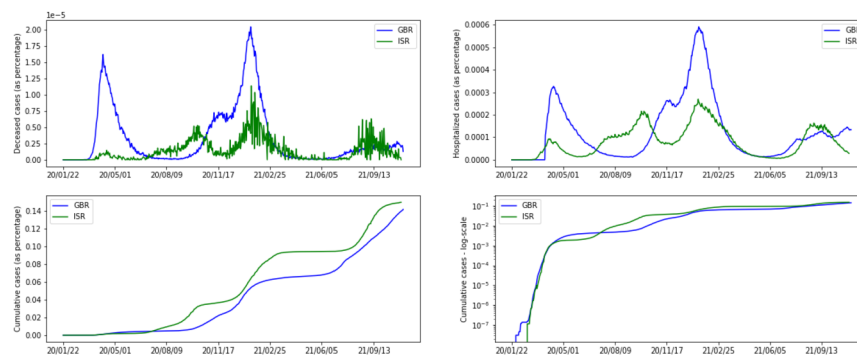
Figure 2 Ground truth attributes. In each subfigure, *y*-axis represents values in terms of population percentage for each country, while *x*-axis provides the dates for the corresponding data. Upper left: new deceased cases at a given day. Upper right: hospitalized cases. Bottom left: cumulative count of confirmed cases. Bottom right: cumulative count of confirmed cases in a log-scale

Figure 3 First group of predictors. In each subfigure, *y*-axis gives values in terms of population percentage for each country, and *x*-axis gives dates for the corresponding data. Left: number of performed tests for a day. Right: cumulative count of administered vaccines

Figure 4 Second group of predictors. Fraction of time (within a span of data) when each policy was in place

2.2.1 Vaccination

We have available the total number of vaccines administered each day for each country. Therefore, we randomly select the corresponding number of nodes and classify them as vaccinated. When choosing the potential nodes to get vaccinated, we exclude deceased, hospitalized, symptomatic, quarantined, already vaccinated, and the nodes representing children (for each graph, we take a specific percentage of nodes representing children). We further use the parameters v_{eff1} and v_{eff2} to account for the different impacts of the vaccine on the virus. When a vaccinated susceptible node comes into contact with an infected neighbor, the probability that it gets exposed is reduced from p_e to $(1 - v_{eff1})p_e$ (in Fig. 1 this effect is denoted with $p_e|v_{eff1}$). Additionally, if a vaccinated node is already infected, the chances that it develops symptoms are reduced from p_{sy} to $(1 - v_{eff2})p_{sy}$ (this means that the probability of being asymptomatic is increased accordingly), and in Fig. 1 this effect is denoted with $p_{sy}|v_{eff2}$ (and $1 - p_{sy}|v_{eff2}$).

2.2.2 Testing and contact tracing

When a testing policy is in place (available in the data along with the number of daily tests — see Figs. 3 and 4), there can be random testing or exclusive testing of symptomatic individuals. In case of random testing, we sample nodes that are not already quarantined, deceased, hospitalized, or recovered. All individuals with a positive test go to quarantine, so they move to either Q_{Asy} or Q_{Sy}, depending on their current state.

Nodes can also transition to quarantine through *'contact_tracing'* - when this policy is active, nodes that were in contact (i.e., share an edge) with a node that tested positive get automatically quarantined. This variable has 3 levels, contemplating: 1 — no contact tracing, 2 — limited contact tracing, or 3 — extensive contact tracing (see Fig. 4). For levels 2 and 3, we consider a parameter p_{ct} that controls the percentage of the contacts that get traced. (Ideally, the value for the third level $p_{ct}(3)$ should be 1, but this is not realistic, and instead we take $0 \ll p_{ct}(2) < p_{ct}(3) < 1$.) When this measure is active, states Q_E and Q_S of our model (Fig. 1) are feasible as well.

2.2.3 Social networks

Every day t and for each region, we consider a network $G(t)$ representing the observed population. Each individual is a single node in the graph, and the edges between nodes correspond to contacts between people. The topology of $G(t)$ depends on the restriction policies that are in place at day t, hence we need to consider different underlying graphs in order to model the population for various restriction scenarios. Here we consider *'stay_home_restrictions'* (a binary flag), *'school_closing'* (four levels of restrictions) and *'workplace_closing'* (four levels of restrictions), and similarly to [3], we introduce a series of lattice and random graphs corresponding to each combination of imposed measures:

- Regular Lattice L_0: each node i is connected to 4 other nodes physically close to i. This network is used to model the population when hard isolation is enforced, which means that all three flags are set to the highest level of restrictions. Theoretically, hard isolation should correspond to separate small connected components (corresponding to different households), but in practice, this representation is known to be more realistic as it takes into account a certain level of unavoidable interaction between individuals in different households [3].

- Small-World Lattice L_α [32]: starting from a regular lattice with degree 4, each edge (i,j) has a (small) probability α of being replaced with (i,k) where k is a random node in the graph. Typically here the rewiring probability α is chosen close to zero, and we use these networks to represent situations where the population is not in lock-down, but there are still strong limitations in place regarding gatherings and movement.

- Random Graph R_β (Erdos-Renyi network) [33]: each possible edge between two nodes is created with probability β. This network, even in the case when β is chosen in such a way that the average degree is 4, provides a much larger level of connectivity than the rewired lattice L_α (unless α is close to 1), because the number of interactions between individuals that are physically far away from each other will be significantly larger. We use these networks when only mild restrictions (or no restrictions at all) are enforced.

By progressively combining the above networks, we create different levels of contact between the population. In order to do so, we start with a four-degree lattice graph L_0 and sequentially compose it with three other small lattice graphs and afterward with four random graphs as to grow the contact between nodes. In total, this yields 8 different networks built in the following way:

$$
\begin{aligned}
L_s &= L_0 \cup L_{p_l}, \\
L_m &= L_s \cup L_{p_l}, \\
L_l &= L_m \cup L_{p_l}, \\
R_{xs} &= L_l \cup R_{p_{rxs}}, \\
R_s &= R_{xs} \cup R_{p_{rs}}, \\
R_m &= R_s \cup R_{p_{rm}}, \\
R_l &= R_m \cup R_{p_{rl}}.
\end{aligned}
\tag{2}
$$

The fact that we build each network on top of the previous one ensures that we do not introduce new edges when transitioning from the denser graph to a more sparse one (when

Table 1 Underlying networks for different combinations of school and workplace closing levels and lockdown restrictions. A lower level entails less restrictions, increasing up to 7, which represents the full lockdown situation

WPCf + SCf + LDf	0	1	2	3	4	5	6	7
Graph	R_l	R_m	R_s	R_{xs}	L_l	L_m	L_s	L_0
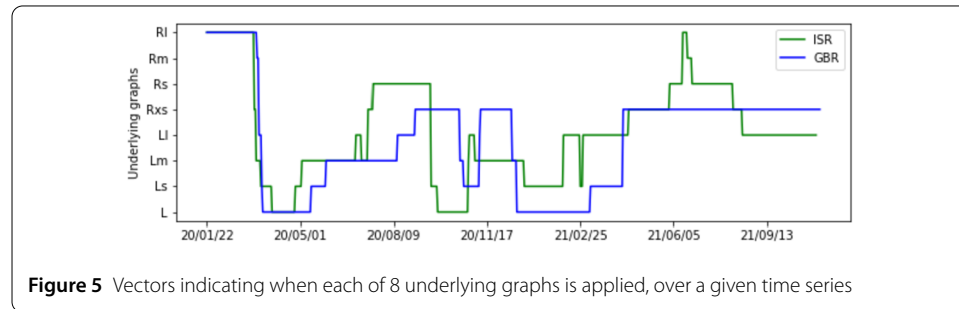
Figure 5 Vectors indicating when each of 8 underlying graphs is applied, over a given time series

the restrictions get more strict). During the simulation, on each day a suitable network is chosen according to the restrictions in place as explained in Table 1. Abbreviations from the table stand for: WPCf: workplace closing flag; SCf: school closing flag; LDf: lockdown flag. Figure 5 shows the temporal positions of the graphs for each country. Note that we leave the network probabilities (p_l, p_{rxs}, p_{rs}, p_{rm}, p_{rl}) as model parameters that should be later optimized. The reasoning is that they represent population behavior with respect to the measures and so should be tuned separately to each country.

2.2.4 Probability of exposure p_e
The probability of being exposed after contact with an infected individual depends on the virus characteristics, but also on what kind of contact people have with each other. For instance, with the progression of the pandemic, people have become more aware of the measures to be taken when in contact with others. We assume that this probability varies with the restrictions in place, and we model it as a function of three parameters ($p_{e_{max}}$, $p_{e_{min}}$ and t) and a lockdown flag. When there are more severe restrictions, people are more careful in their contacts, so this value should be lower and vice-versa. We assume that there is a maximum ($p_{e_{max}}$) and minimum ($p_{e_{min}}$) value that the parameter p_e can take (in extreme cases). Before the first lockdown occurs, p_e keeps its maximum value ($p_{e_{max}}$), and as soon as the lockdown is imposed it drops to its minimum ($p_{e_{min}}$). After lifting the lockdown, p_e starts linearly increasing so that after t months it attains its maximum value again (unless another lockdown is imposed). This behaviour is enforced in the same way any time a new lockdown occurs. The estimated p_e vectors for each country are presented in Fig. 6.

2.3 Decomposition and complete model
Instead of creating a single large graph to represent a country's population, we consider several smaller graphs corresponding to separate cities or densely populated regions that have fewer connections with the rest of the population. While this has the primary goal of enforcing a more realistic simulation setting, it provides an additional advantage in terms of computational costs. When the size N of the considered population is large, working

Figure 6 Evolution of p_e over time
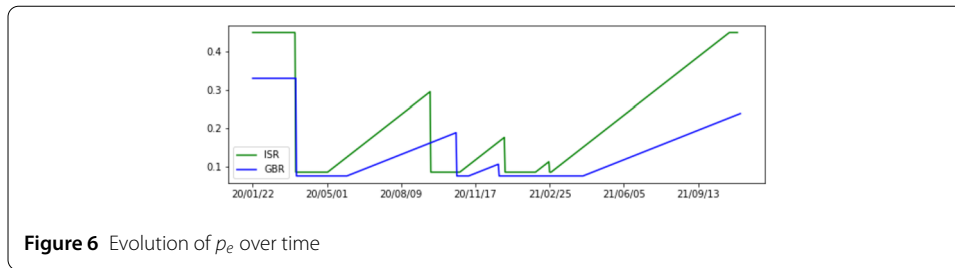
with a network of N nodes becomes too expensive, but by splitting it into m subgraphs of sizes $N_i \ll N$ for $i = 1, \ldots, m$, we reduce this cost considerably and allow for parallel processing. At each time step, for each separate network, we simulate the progression in the spread of the infection according to the model described in the previous subsections. Then, we model the interactions among the networks following a standard (not network-based) compartmental approach. That is, once the states of all nodes have been updated, we take into account the exposures that may arise from contacts between people in different networks. For every network indexed $i = 1, \ldots, m$ we define the additional set of exposed people as

$$\hat{E}_i(t) = \sum_{j \neq i} w_{ij}(t) \frac{p_e}{N} I_i(t) S_j(t), \tag{3}$$

where p_e is the probability of an individual becoming exposed after being in direct contact with an infected person and $w_{ij}(t) \in [0, 1]$ is a parameter that regulates how likely it is for individuals in the region i to enter in contact with individuals of region j. We can see that the above equation is analogous to the one used in the classical SIR model (1): the only significant difference is that here we assume that individuals become exposed before transitioning into the class of infected people, and that we weight the number of expected contacts by multiplying with the coefficient w_{ij}. The choice of the coefficient $w_{ij}(t)$ will depend on the restrictions that are enforced at the considered time t.

This decomposition allows for the application of an approximated network-based approach to the case of populations of entire countries. A further advantage is that the decomposition can be done in networks of equal size or in networks representing the country's administrative regions — an appealing setting if different restrictions are enforced in different areas.

Finally, in order to account for people entering the country while infected, we introduce a parameter p_{int}, active only when there are no international travel restrictions, that represents the probability of introducing a new exposed case in each region. This parameter is set to a low value $p_{int} = 0.01$, since higher values would make the model deviate too far from the closed system setting, and setting it to zero would reduce a realistic representation. With additional information on international travel, one could potentially improve this estimate.

3 Numerical simulations

3.1 General setting

As mentioned before, the population of a country is simulated with an underlying sparse graph. In order to execute simulations on a user-level PC, we cannot work with the exact

population size but reduce it by a factor of 100. Further, to compensate for the fact that the population is not actually homogeneous but consists of several connected components (cities or regions) that have only sporadic contacts, we compose the population graph G as a union of disjoint subgraphs $G_1, \ldots, G_m$. Each subgraph G_i consists of n_i nodes, where $25{,}000 < n_i < 35{,}000$. Since the regions obtained in this way are disjoint, we model the "inter-regional contact" using a differential compartmental model as described in Sect. 2.3. This will introduce additional exposed cases, proportional to the number of currently infected individuals in the neighboring regions, unless internal traveling restrictions are in place. For our simulations, we assume that the parameter w_{ij} of equation (3) is the same for every pair of indices i and j, and equal to r_{mix}. Since this parameter models the number of contacts among individuals in different cities or regions, it is reasonable to keep its value small. We assume in our simulations that $r_{max} \leq 0.1$. It is important to note that this approach to modeling the population is also favorable in terms of computational costs — working on a single connected graph of the same size as G is infeasible on user-level machines.

Since the data shows that the first cases appear after 30 days of recording, we introduce one exposed node per region at day 30 of the simulation. As an approximation for the number of people that should be excluded from vaccination, we take the percentage of children in each country (18% for GBR and 27% for ISR).[2]

Simulations are run over more than 20 months, starting on January 2020, until November 2021. Every simulation with each set of parameter values is executed ten times; this, together with randomness in graph sizes and connections, has shown to be enough to give a good generalization of a scenario.

3.2 Parameter optimization

We distinguish between parameters related to the virus biology and parameters related to the population's particular characteristics or behavior. We consider the first as country-independent and assume that their values should be in a similar range as proposed in the literature. On the other hand, parameters related to the population should be country-dependent (thus, we allow different values for ISR and GBR during the optimization) and cannot be found in the literature, as they are specific to our model and often have very different meanings. Both groups are listed in Tables 2 and 3, with (available) literature ranges and our final estimates.

Regarding the first group of parameters, numerous papers proposed the values for the compartmental models concerning COVID-19; however, most of them focus only on a single country or a region or were estimated early in the pandemic, hence we should be careful to accept them. Still, we might consider that parameter values obtained in different studies give a reliable interval range as a starting point in optimization. Therefore, we take values suggested across the literature to narrow down the grid search space when fitting the values for ISR and GBR.

An additional remark is due regarding vaccination. The information about vaccination present in our data only refers to the total number of doses given per day, without detailing how many doses each individual received, which leads us to consider that one dose is

---

[2]For more accurate modeling, these percentages could be augmented with people that will not get vaccinated due to due to some other reasons (medical, religious, etc.)

Table 2 Country-independent parameters with literature ranges suggested in the literature. The probability of infection is inverse to the number of incubation days (13 days), the probability of recovering is inverse to the average length of recovery (10 days), probability of leaving the quarantine is inverse to the length of quarantine (14 days). Vaccine efficacy averages around 0.9

Parameter	Literature range	Source	Our estimate
(v_{eff1}, v_{eff2})	0.8–0.95	[7, 11]	$(0.95, 0.7)$
p_i	3^{-1}–20^{-1}	[6, 19, 20, 34–36]	0.08
p_{sy}	0.13–0.65	[6, 20, 27, 34, 37]	0.5
p_r	3^{-1}–30^{-1}	[6, 20, 36]	1/29
p_{syh}	0.05–0.15	[6]	0.006
p_{hd}	10^{-5}–10^{-1}	[6, 20, 27, 36, 38]	0.03
p_{hr}	3^{-1}–30^{-1}	[6, 20, 36]	1/11

Table 3 Country-dependent parameters, with initial and final estimates for GBR and ISR

Parameter	Initial range	GBR estimate	ISR estimate
$(p_{e_{min}}, p_{e_{max}})$	$(0.01$–$0.2, 0.2$–$0.5)$	$(0.075, 0.33)$	$(0.085, 0.45)$
t	3–18	11	8
$(p_{ct}(2), p_{ct}(3))$	$(0.3$–$0.7, 0.7$–$1)$	$(0.65, 0.8)$	$(0.65, 0.8)$
p_l	0.004–0.008	0.006	0.004
$(p_{rxs}, p_{rs}, p_{rm}, p_{rl})$	0.5–4.5	$(1.5, 0.8, 1.5, 0.8)$	$(0.9, 1.5, 1, 0.8)$
r_{mix}	0.05–0.15	0.065	0.065

enough for a person to be vaccinated. However, this means that we should allow for lower values of v_{eff1} than the ones found in the literature.

3.2.1 Parameter tuning

We use two metrics in the optimization process: sMAPE [39] (symmetric Mean Absolute Percentage Error) and Pearson correlation [40]. Both functions are altered to fit our data-series structure, and performed over smoothed curves, in the following way. Given some ground truth vector $y \in \mathbb{R}^n$ and the corresponding predicted vector y^*, we first compute the moving sum for both of them, given as $\tilde{y}$ and $\tilde{y}^*$, respectively. The two modified metrics are then computed as

$$sMAPE(y, y^*) = \frac{1}{n} \sum_{i=1}^{n} \frac{2|\tilde{y}^* - \tilde{y}|}{|\tilde{y}| - |\tilde{y}^*|} \times 100 \tag{4}$$

and

$$pearson(y, y^*) = \frac{\sum_{i=1}^{n} (\tilde{y}_i^* - \tilde{y}_M^*)(\tilde{y}_i - \tilde{y}_M)}{\sqrt{\sum_{i=1}^{n} (\tilde{y}_i^* - \tilde{y}_M^*)^2 \sum_{i=1}^{n} (\tilde{y}_i - \tilde{y}_M)^2}}, \tag{5}$$

where $\tilde{y}_M$ and $\tilde{y}_M^*$ are the means of $\tilde{y}$ and $\tilde{y}^*$, respectively. Both metrics are computed only for values of $\tilde{y}$ different than zero.

To evaluate the model we take y as the number of hospitalized (N_H) and deceased (N_D) individuals per day. As mentioned in Sect. 2.2, we consider these two attributes to be the best indicators of model accuracy. Consequently, we have 4 different metrics $sMAPE(N_H, N_H^*)$, $sMAPE(N_D, N_D^*)$, $pearson(N_H, N_H^*)$ and $pearson(N_D, N_D^*)$. We further enforce a constraint based on the cumulative number of cases, to ensure that the optimization of the remaining metrics is not achieved at the cost of an unreasonable number of

cases (even if we do not wish to optimize it, due to its unreliability). Therefore, if the total cumulative number of predicted cases is below or above a given threshold dependent on the true value, the combination of parameters is deemed unacceptable. The combination is then excluded if any of 10 runs violates the constraints. This additionally ensures that our model does not overfit.

The tuning of the parameters with respect to the previous metrics was conducted in two main stages that we now describe. Let us denote with θ the vector containing all the parameters to be optimized (as described in Sect. 3.2). The goal of the first stage is to select an initial guess for the optimization stage. We consider a set of randomly generated values of θ and run the simulation for each given value. The results were then compared and the value of θ giving the best results was selected as starting point for the subsequent stage. During the second stage, the optimization was performed by grid search. To ease the computational load, we split the parameters into subsets, and optimize the model with respect to the parameters in each subset, in a cyclical fashion. At each iteration of the process, the optimization is performed as follows. Let us assume that $\{\theta_j\}_{j \in J_k}$ are the parameters to be optimized at the k-th iteration. For each $j \in J_k$ we consider m equidistant values of θ_j over a given interval $I_j^k \subset \mathbb{R}$ and we define a set of $m^{|J_k|}$ instances of the vector θ, obtained by combining the different values of θ_j for $j \in J_k$. We run the simulation for each of the obtained vectors θ and select the combination that provides the best results. After this, the intervals I_j^k are refined and a new subset of parameters is selected for optimization.

3.3 Final predictions

Simulations with the optimal set of parameter values were run 100 times, and the results for both countries are presented in Fig. 7. The mean over predictions is presented in solid lines, while the shaded area corresponds to the standard deviation (divided by two). An important remark is that both the mean and standard deviation reached a stable solution after only 10 trials, without noticeable changes in the subsequent 90.

Figure 7 Final estimated solution. Left column: GBR; right column: ISR. Ground truth data is presented by a solid black line, while estimated numbers are presented in blue for GBR and in green for ISR — solid lines indicate the mean over 100 simulations, and the shaded area shows standard deviation
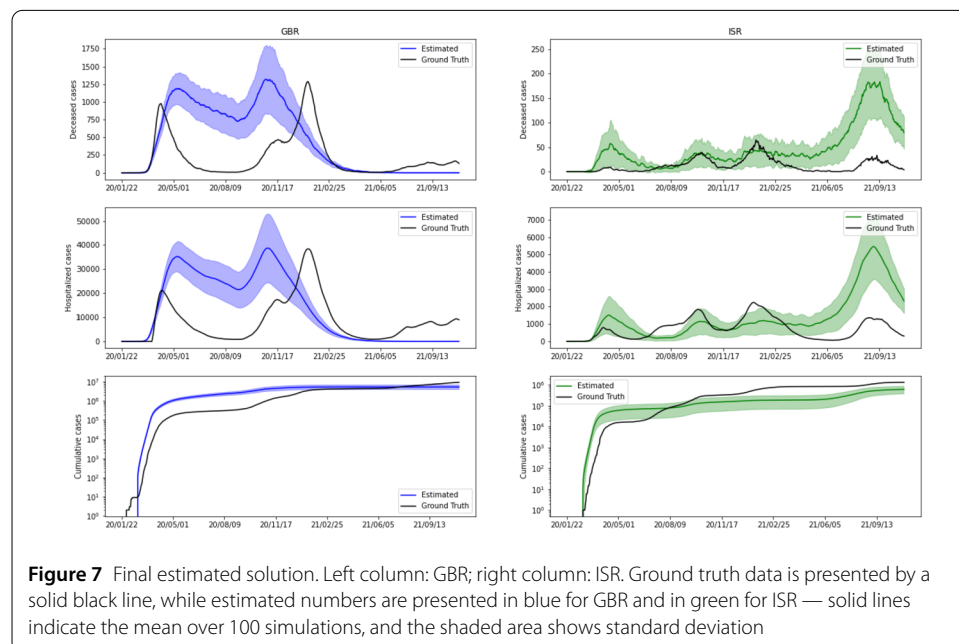
Even if we do not observe a perfect fit for the curves, it is noticeable that the main waves are fairly well predicted in terms of time and strength — the main miss fit is at the final stage. For that period, our simulations are not able to predict the third wave in GBR and produce an overestimation of cases for ISR. A probable explanation lies in the long-term fitting that we are taking into account (almost two years). A more realistic model could consider possible loss of immunity for recovered and vaccinated nodes, or even natural demographic changes. This is a possible line of research to be addressed in future work.

It is pertinent to compare the estimated values for country-dependent parameters in Table 3. A larger $p_{e_{max}}$ value for ISR suggests more contacts in the absence of restrictions for this country. In addition, a lower value of t suggests that after a lockdown ISR is faster to return to the usual state of contact levels. Interestingly, the percentage of contact tracing, even if left free to be individually tuned, was found to take similar values, which suggests levels 2 and 3 have the same meaning for both countries. Finally, looking at the network parameters we see that p_l takes a lower value for ISR, possibly entailing that the restrictions closer to lockdown are enforced more strictly there. For the remaining topologies, given the considered ranges, we find that the estimated parameters are in general similar, especially for the most relaxed networks (i.e. closer to a non-restricted scenario). This is reasonable, since $p_{e_{max}}$ should already be expressing any differences regarding this situation.
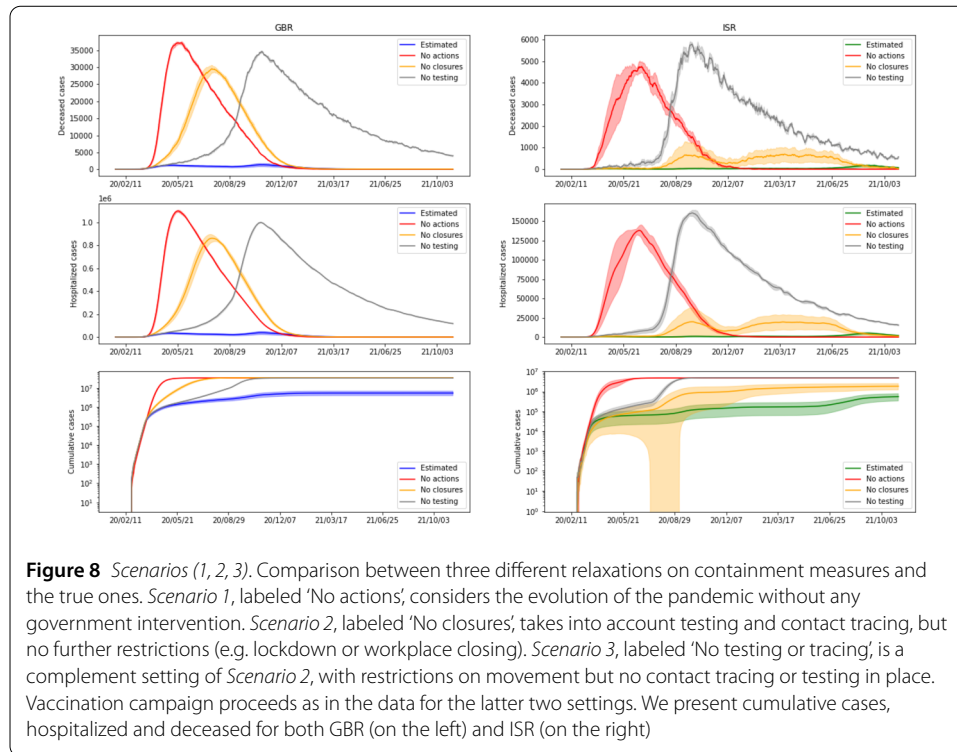
4 Impact of different measures

In this section we use the proposed model to study different hypothetical scenarios. This can be useful in order to understand the impact of different policies and restrictions and therefore in providing helpful information that may guide the decision process. Unless otherwise stated, the parameters used in the simulations are the ones obtained in the previous section. Moreover, since the goal is to evaluate the behaviour of the model when different measures are in place, we compare the scenarios with the original setting when pertinent.

4.1 Restriction measures

In *Scenario 1*, we consider the case where no action at all is taken to try to contain the spread of the disease. We assume here that no stay-at-home policy is mandated by the government and that schools and workplaces are open without any restrictions. Moreover, we assume that no tests are performed to try and detect infected individuals and that no contact tracing strategy is in place (in particular this implies that there is no mandatory quarantine for anybody). We also assume that the vaccination campaign never starts.

In *Scenario 2* we consider the case where no generalized restrictions are mandated by the government, but all the other policies such as testing, contact tracing (with quarantine) and vaccinations are still in place and are the same as in the data.

In *Scenario 3* we consider the case where stay-at-home and work/study-from-home policies are enforced according to data, but neither testing nor contact tracing policies are employed, hence no infected individual gets quarantined, regardless of their state. We also assume that the vaccination campaign proceeds as in the data. In Fig. 8 we plot the predictions for these three scenarios. As we can see, they all lead to a spread of the disease that is much larger than the one given by the actual policies, suggesting that, as expected, none of the three approaches represented by the described scenarios are viable in practice. In particular, we can notice that limitations on daily activities such as working and

Figure 8 *Scenarios (1, 2, 3). Comparison between three different relaxations on containment measures and the true ones. Scenario 1, labeled 'No actions', considers the evolution of the pandemic without any government intervention. Scenario 2, labeled 'No closures', takes into account testing and contact tracing, but no further restrictions (e.g. lockdown or workplace closing). Scenario 3, labeled 'No testing or tracing', is a complement setting of Scenario 2, with restrictions on movement but no contact tracing or testing in place. Vaccination campaign proceeds as in the data for the latter two settings. We present cumulative cases, hospitalized and deceased for both GBR (on the left) and ISR (on the right)*

studying from home policies are essential to contain the spread of the virus, as well as testing procedures and planned quarantine time for the infected individual. Lifting any of these containment measures seems to lead to statistics comparable with the case where no action was taken. In Fig. 8, we see that in all three scenarios we have a single mode, i.e. the virus quickly reaches the point of saturation, except in the scenario of no imposed closures in ISR. This last case is in accordance with the scenario that we will consider next, and indicates that the stay home policy was enforced more strictly in ISR than in GBR. From the same results we can also conclude that, while both restrictions and testing are essential, neither of these policies alone is enough to sufficiently limit the spread of the disease.

We now study milder relaxations of the restrictions, with respect to the data. The obtained results are plotted in Fig. 9. In *Scenario 4*, we consider the case where no stay-at-home policy is enforced, but the closing of schools and workplaces (as well as all other measures such as quarantine and vaccinations) are mandated as in the original data. In *Scenario 5*, we assume that whenever a lockdown is in place, schools and workplaces are also closed but that no limitations on school and work attendance are imposed when stay-at-home policies are not enforced.

As we can see from Fig. 9, these two choices lead to a number of cases that is similar to our estimate for the original data. The results for *Scenario 5* suggest that limiting the physical attendance of schools and workplaces only for periods of time when stay-at-home policies are also enforced may be a viable option. This is particularly interesting because it could result in a significantly shorter amount of time during which students and employees are forced to work from home, without leading to major changes in the spread of the virus. At the same time, we see that as long as schools and workplaces are subject to restrictions, stay-at-home policies seem to be inessential: for the two countries that we

Figure 9 *Scenarios (4, 5)*. Comparison between milder relaxations on restriction policies, confronted with the real ones. *Scenario 4*, labeled 'No lockdown', considers closing of schools and workplaces, but absence of stay-at-home policy. All other measures remain unaltered. *Scenario 5*, labeled 'Strict lockdown', considers that whenever a lockdown is in place, schools and workplaces are also closed, but than no limitations on schools and work attendance are imposed when stay-at-home policies are not enforced. We present cumulative cases, hospitalized and deceased for both GBR (on the left) and ISR (on the right)
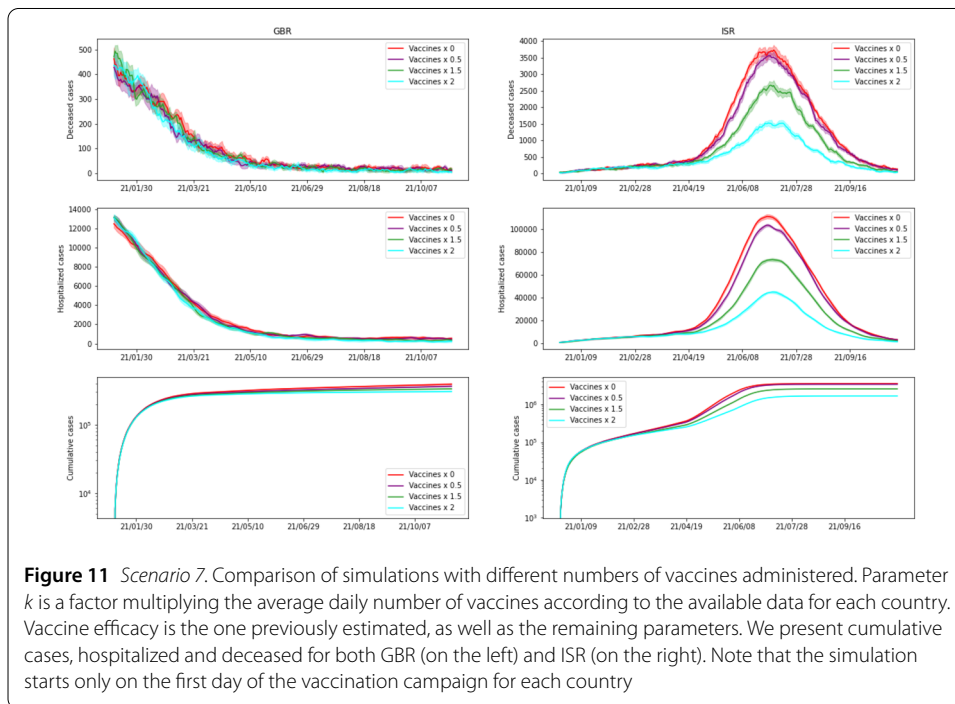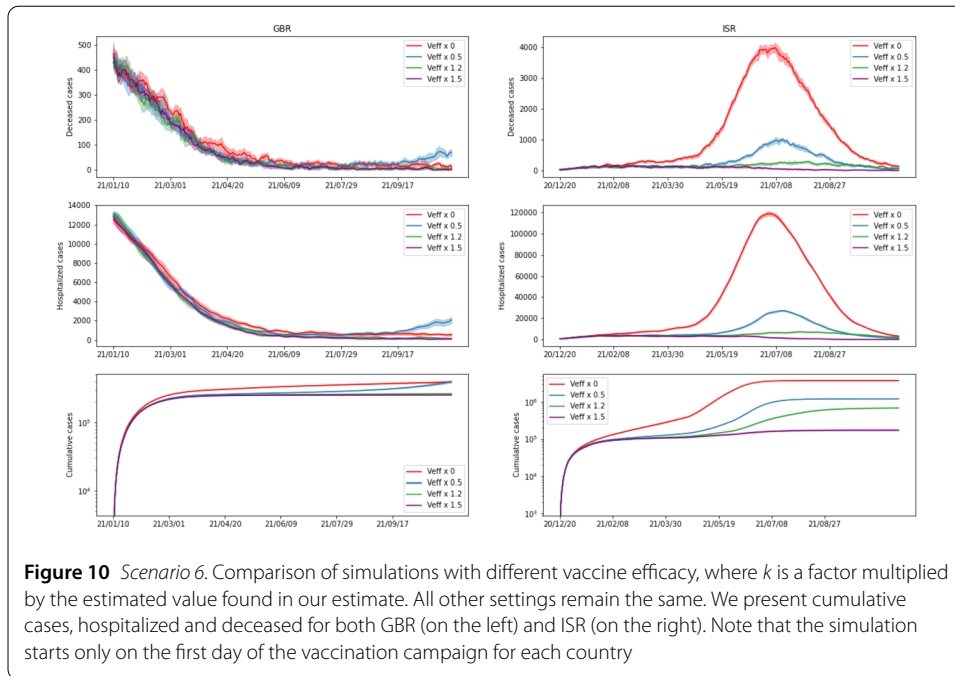
took in the examination this seems to make a significant difference only for the end of the considered time-frame and only for Israel. The reason why GBR seems to be less affected by the changes in *Scenario 4* and *Scenario 5* with respect to the original data is that, for the vast majority of the considered days, stay-at-home and work/study-from-home policies were mandated and lifted at the same time. In ISR, on the other hand, several different combinations of restrictions were active on different days during the considered time interval.

4.2 Vaccines

In *Scenario 6* we study how the spread of the virus among the population is influenced by vaccine efficacy. We assume that the vaccination campaign proceeds in the two countries as in the considered data, but we consider the case where the efficacy of the vaccine is given by a factor k times the efficacy found in the previous section, for $k \in \{0, 0.5, 1.2, 1.5\}$. In *Scenario 7* we investigate the consequences of different vaccination rates on the evolution of the contagion. We assume that the vaccine efficacy is the same as in the previous section and that the initial day of the vaccination campaigns in the two countries is the same as in the data, but we vary the amount of administered vaccines. Specifically we consider the case where the number of vaccines distributed to the population per day is k times the real number of daily vaccines with $k \in \{0, 0.5, 1.5, 2\}$. We remark that in both *Scenario 6* and *Scenario 7*, a factor $k = 0$ naturally corresponds to the case where no vaccines are administered at all.

In Figs. 10 (*Scenario 6*) and 11 (*Scenario 7*), we plot the obtained results. In order to have a fair picture of how vaccines affect the outcome, we need to have the same starting point. We accomplish this by starting simulations on the initial day of the campaign (day 333 for ISR and day 354 for GBR), with the same initial state. That is, we start with a graph whose nodes are assigned to corresponding states according to the average of our initial

Figure 10 *Scenario 6*. Comparison of simulations with different vaccine efficacy, where *k* is a factor multiplied by the estimated value found in our estimate. All other settings remain the same. We present cumulative cases, hospitalized and deceased for both GBR (on the left) and ISR (on the right). Note that the simulation starts only on the first day of the vaccination campaign for each country

Figure 11 *Scenario 7*. Comparison of simulations with different numbers of vaccines administered. Parameter *k* is a factor multiplying the average daily number of vaccines according to the available data for each country. Vaccine efficacy is the one previously estimated, as well as the remaining parameters. We present cumulative cases, hospitalized and deceased for both GBR (on the left) and ISR (on the right). Note that the simulation starts only on the first day of the vaccination campaign for each country

estimate for the considered day. For completeness, we include both countries, but since in the considered timeframe the vaccination campaign of GBR started relatively late, we can notice that different vaccination policies and efficacy have no significant impact on the evolution of the spread there. Nonetheless, we point out that for vaccine efficacy reduced by a factor of 2 there is a start of a second peak, which is also present in the real data. This may originate from our simplified vaccine model, where we consider one person fully vaccinated after a single dose. From Figs. 10 and 11, we can see that, as expected,

not vaccinating the population leads to a significantly higher number of hospitalized and deceased people, compared to any other vaccination campaign. Moreover, comparing the two Figures, we can see that vaccine efficacy has a larger impact on the number of cases and hospitalized people when compared to the vaccination rate. This suggests that, while having promptly available vaccines and proceeding rapidly in the distribution to the population certainly helps to reduce the number of cases and in particular the number of people seriously affected by the disease, the quality of the vaccines plays a major role.

4.3 Testing and contact tracing

To study the influence of the testing and contact tracing procedures on the spread of the disease, we consider three different scenarios:

- varying the number of tests;
- varying the initial date of the testing campaign;
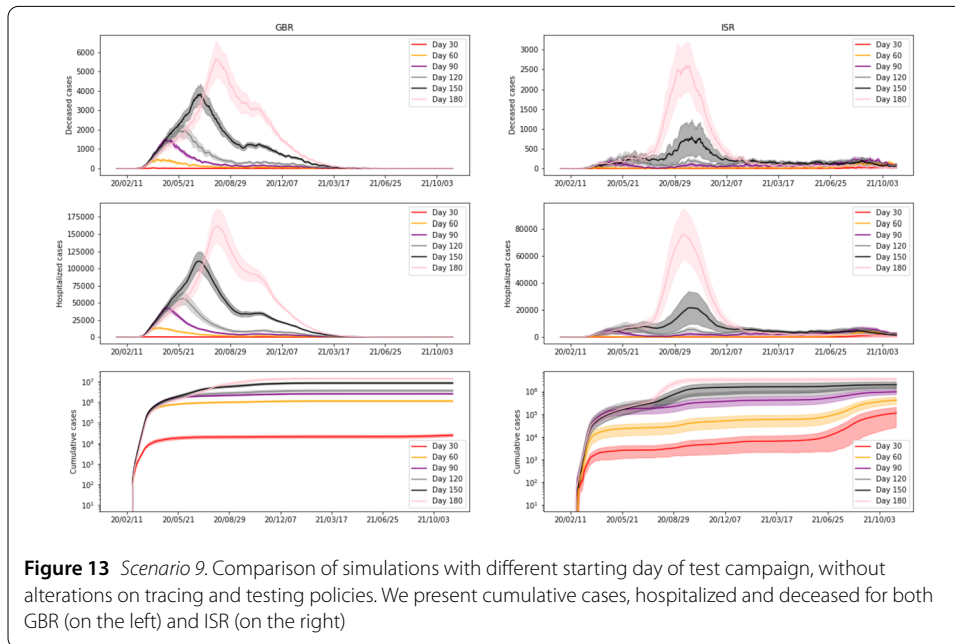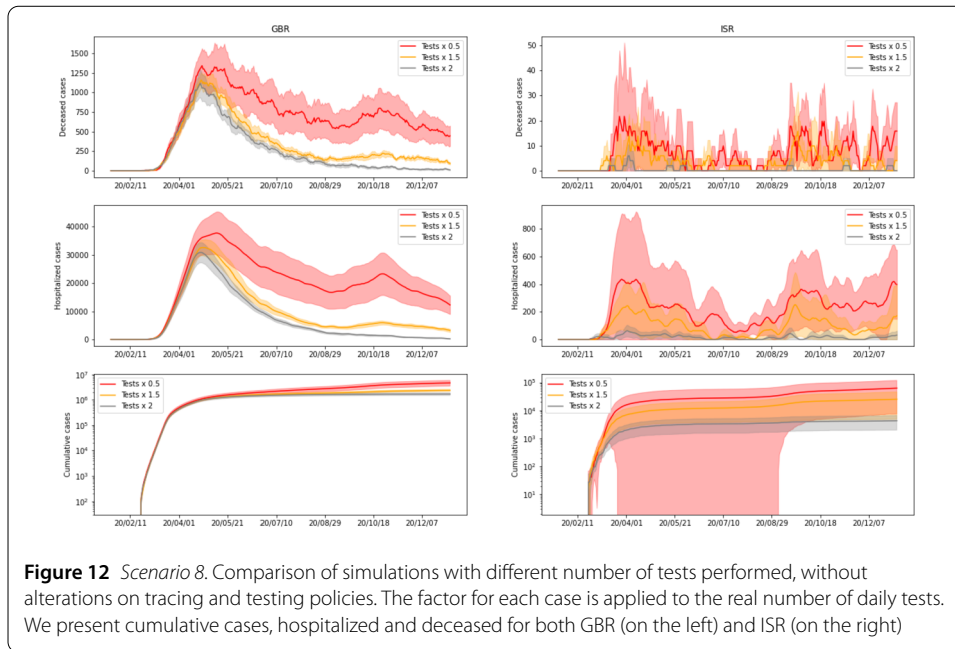- varying the accuracy of the contact tracing policy.

In general, the more tests are carried out, among both the people who manifest symptoms and the general population, the more infected individuals are detected and therefore sent to quarantine, which should significantly limit the spread of the virus. A similar remark can be made for contact tracing policies — the more accurate the contact tracing procedure (so, the higher the number of direct contacts that are quarantined), the smaller the risk of potentially infected individuals spreading the virus.

In *Scenario 8*, we assume that the contact tracing and testing policies are the same as in the data, but that the number of tests is different. Specifically, we consider the case where the number of tests performed at each day of the simulation is k times the real number of daily tests, with $k \in \{0.5, 1.5, 2\}$. In *Scenario 9*, we study how the starting date of the testing procedure influences the spread of the virus. We assume that the testing campaign begins at day t_0 and we consider several values of t_0 between 30 and 180. To simplify the comparison, we assume that at each day t after t_0, a fixed number of tests (equal to the daily average of tests performed by the country over the considered timeframe) is performed over the population.

Results for *Scenario 8* can be found in Fig. 12, and for *Scenario 9* in Fig. 13. As expected, the increase in daily tests leads to a decrease in the number of cases. We further note that the case of no tests (i.e. a factor of 0) can be found in Fig. 8 under the label 'No testing or tracing' and follows the same pattern. When it comes to the starting day, the differences are visible from the beginning, and the simulations with a late start of testing exhibit extremely large peaks for deaths and hospitalizations. In the final days, all the simulations go down towards zero, so in Fig. 13 we show only the first 300 days, for the sake of clarity. The main conclusion to be drawn is that a prompt start of testing campaigns may be of great help in significantly limiting and potentially stopping the spread of the virus. It is also pertinent to compare the two scenarios — the variation achieved in Fig. 12 for different factors is almost negligible, when compared to the one in Fig. 13. This suggests that the latter is more important in controlling the number of cases.
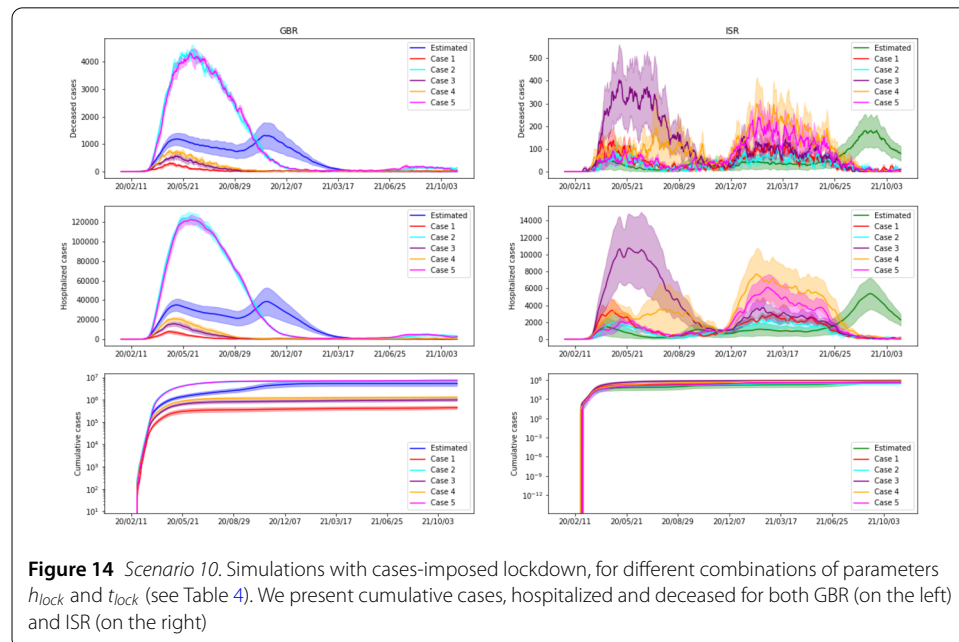
4.4 Cases-driven lockdown

Finally, we test *Scenario 10*, where a lockdown is activated when a given number of hospitalized cases is reached. This threshold h_{lock} is set as a percentage of the total population and we test different values for it. The lockdown is then kept until the cases stay below the

Figure 12 *Scenario 8*. Comparison of simulations with different number of tests performed, without alterations on tracing and testing policies. The factor for each case is applied to the real number of daily tests. We present cumulative cases, hospitalized and deceased for both GBR (on the left) and ISR (on the right)

Figure 13 *Scenario 9*. Comparison of simulations with different starting day of test campaign, without alterations on tracing and testing policies. We present cumulative cases, hospitalized and deceased for both GBR (on the left) and ISR (on the right)

same threshold for a given period t_{lock}, also tested with different values. We show 5 different combinations of these two parameters, described in Table 4. Note that the threshold value was generally chosen to be smaller for GBR, since a value of h_{lock} larger than 7.5^{-5} leads to an uncontrollable explosion in cases and so does not offer much insight. In Fig. 14, we see that cases 2 and 5 yield a high number of deaths for GBR, while the remaining options perform better than the original scenario. Another interesting insight is provided in Table 5 showing the average number of days with imposed lockdown. We note that, for GBR, it is possible to reduce the number of causalities while enforcing lockdown restrictions of an inferior duration than what Fig. 14 might suggest. For ISR, we can notice

Table 4 Parameters for different cases of *Scenario 9*

Case	t_{lock}	h_{lock} (GBR)	h_{lock} (ISR)
1	7	5×10^{-6}	5×10^{-5}
2	7	10^{-5}	10^{-4}
3	14	10^{-5}	10^{-4}
4	7	7.5×10^{-5}	7.5×10^{-5}
5	14	7.5×10^{-5}	7.5×10^{-5}

Figure 14 *Scenario 10*. Simulations with cases-imposed lockdown, for different combinations of parameters h_{lock} and t_{lock} (see Table 4). We present cumulative cases, hospitalized and deceased for both GBR (on the left) and ISR (on the right)

Table 5 Average number of days with imposed quarantine for each case in *Scenario 9*

Case	1	2	3	4	5
GBR	336	315	333	336	307
ISR	423	358	427	428	409

two distinctive peaks for any considered scenario, which explains the longer lockdown duration in Table 5. Unlike GBR, we do not achieve an improvement for any of the proposed settings,[3] so it would be relevant to consider a broader range of scenarios in order to potentially reach an effective strategy for both countries.

5 Discussion

In this paper, we proposed an extended SIR model that blends a standard compartmental (differential) model and a graph-based approach, fitted over two countries, in order to ensure better generalization properties. Estimated solutions show good behavior, correctly predicting positions and magnitudes of the peaks, except for the very last months of the time series. In order to increase the accuracy of the final phase, additional assumptions should be considered, including loss of immunity induced by time, and perhaps booster doses of the vaccine — we leave this as a direction for further research.

---

[3] We note a further reduction in the threshold value is not feasible in our experiments due to the simulated population size. This is an interesting line for future research.

Furthermore, we have investigated the effects of different containment measures w.r.t. the spread and the impact of the virus, by simulating alternative hypothetical scenarios with diverse combinations of the input parameters. We have found that testing plays a significant role in containing the outbreak, suggesting that an early start and frequent occurrence are fundamental. As expected, strategically imposing different levels of closure results in a decreasing (or increasing) number of cases. However, somewhat surprisingly, stay-at-home restrictions as an independent measure (i.e. without closing of schools and workplaces), do not show a great impact – particularly for GBR, where the simulated scenario hardly differs from the original estimate.

Another interesting scenario is imposing (and lifting) the lockdown based on the number of hospitalized people at the moment. This strategy showed favorable results for GBR. For ISR, the considered scenarios were probably not diverse enough to achieve improvement in the number of cases. Besides a broader range of parameters, other possibilities of restrictions could be investigated in the future.

The influence of the vaccination campaign shows clear and predictable patterns in the case of ISR — if either the efficacy or the number of administered vaccines is increased, the number of cases decreases, and vice versa. For GBR, on the other hand, we observe almost negligible variations, which may be explained by their lower pace of vaccination and overall percentage of vaccinated people in the time frame that we considered.

Abbreviations
COVID-19, COrona VIrus Disease 2019; GBR, Great Britain; ISR, Israel; SEIRD model, Susceptible, Exposed, Infectious, Recovered, Dead model; GT, ground truth.

Availability of data and materials
The datasets generated and/or analysed during the current study are available in the Covid-19 Data Hub repository, https://covid19datahub.io/index.html [31]. Code is available in the github repository https://github.com/stevorackovic/A-hybrid-compartmental-model-with-a-case-study-of-COVID19-in-Great-Britain-and-Israel.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author contribution
All authors evenly contributed to the whole work. All authors read and approved the final manuscript.

Author details
[1]Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia. [2]Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal. [3]Department of Environmental Science and Policy, Università degli Studi di Milano, Milan, Italy.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Cao L, Liu Q. COVID-19 modeling: a review. Pre-print arXiv. 2021; .
2. Kermack W, McKendrick A. Contributions to the mathematical theory of epidemics – I. Bull Math Biol. 1991;53(1–2):33–55.
3. Small M, Cavanagh D. Modelling strong control measures for epidemic propagation with networks - a COVID-19 case study. IEEE Access. 2020;8:109719–31.
4. Belik V, Geisel T, Brockmann D. Natural human mobility patterns and spatial spread of infectious diseases. Phys Rev X. 2011;1:011001.
5. Adak D, Majumder A, Bairagi N. Mathematical perspective of Covid-19 pandemic: disease extinction criteria in deterministic and stochastic models. Chaos Solitons Fractals. 2021;142:110381.
6. Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, Kuang Y et al. To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infect Dis Model. 2020;5:293–308.
7. Giordano G, Colaneri M et al. Modeling vaccination rollouts, SARS-CoV-2 variants and the requirement for non-pharmaceutical interventions in Italy. Nat Med. 2021;27:993–8.
8. Gounane S, Barkouch Y et al. An adaptive social distancing SIR model for COVID-19 disease spreading and forecasting. Epidemiol Methods. 2021;10(s1):20200044.
9. Manski CF, Molinari F. Estimating the COVID-19 infection rate: anatomy of an inference problem. J Econom. 2021;220(1):181–92.
10. Lorig F, Johansson E, Davidsson P. Agent-based social simulation of the Covid-19 pandemic: a systematic review. J Artif Soc Soc Simul. 2021;24(3):5.
11. McGee RS, Homburger JR, Williams HE, Bergstrom CT, Zhou AY. Model-driven mitigation measures for reopening schools during the COVID-19 pandemic. medRxiv. 2021;10(s1).
12. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton: Princeton University Press; 2011.
13. Korolev I. Identification and estimation of the SEIRD epidemic model for COVID-19. J Econom. 2021;220:63–85.
14. Chung NN, Chew LY. Modelling Singapore COVID-19 pandemic with a SEIR multiplex network model. Sci Rep. 2021;11:10122.
15. Aleta A, Martín-Corral D, Pastore y Piontti A et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nat Hum Behav. 2020;4:964–71.
16. Ramos AM, Vela-Pérez M, Ferrández MR, Kubik AB, Ivorra B. Modeling the impact of SARS-CoV-2 variants and vaccines on the spread of COVID-19. Commun Nonlinear Sci Numer Simul. 2021;102:105937.
17. Usherwood T, LaJoie Z, Srivastava V. A model and predictions for COVID-19 considering population behavior and vaccination. Sci Rep. 2021;11:12051.
18. Tetteh JNA et al. Network models to evaluate vaccine strategies towards herd immunity in COVID-19. J Theor Biol. 2021;531:110894.
19. Read J, et al. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. MedRxiv. 2020;347.
20. Tang B et al. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). Infect Dis Model. 2020;5:248–55.
21. Langfeld K. Dynamics of epidemic diseases without guaranteed immunity. J Math Ind. 2021;11:5.
22. Maier B, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science. 2020;368(6492):742–6.
23. Chang SL, Harding N, Zachreson C et al. Modelling transmission and control of the COVID-19 pandemic in Australia. Nat Commun. 2020;11:5710.
24. Tarrataca L, Dias CM, Haddad DB, De Arruda EF. Flattening the curves: on-off lock-down strategies for COVID-19 with an application to Brazil. J Math Ind. 2021;11:2.
25. Wu J, Tang B, Bragazzi NL, Nah K, McCarthy Z. Quantifying the role of social distancing, personal protection and case detection in mitigating COVID-19 outbreak in Ontario, Canada. J Math Ind. 2020;10(1):15.
26. Götz T, Heidrich P. Early stage COVID-19 disease dynamics in Germany: models and parameter identification. J Math Ind. 2020;10:20.
27. Ferguson N, et al. Report 9: impact of non-pharmaceutical interventions. (NPIs) to reduce COVID19 mortality and healthcare demand. 2020.
28. Bicher M, Rippinger C, Urach C, Brunmeir D, Siebert U, Popper N. Agent-based simulation for evaluation of contact-tracing policies against the spread of SARS-CoV-2. Med Decis Mak. 2021;41(8):1017–32.
29. Schlosser F, Maier B, Jack O, Hinrichs D, Zachariae A, Brockmann D. COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. Proc Natl Acad Sci USA. 2020;117:32883–90.
30. Meyn SP, Tweedie RL. Markov chains and stochastic stability. Berlin: Springer; 2012.
31. Guidotti E, Ardia D. DataHub. J Open Sour Softw. 2020;5(51):2376.
32. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998;393:440–2.
33. Erdos P, Rényi A. On random graphs I. Publ Math (Debr). 1959;6:290–7.
34. Li R et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science. 2020;368(6490):489–93.
35. Lauer SA et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med. 2020;172(9):577–82.
36. Fei Z et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020;395(10229):1054–62.
37. Moriarty LF et al. Public health responses to COVID-19 outbreaks on cruise ships – worldwide, February–March 2020. Morb Mort Wkly Rep. 2020;69(12):347–52.
38. Verity R et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis. 2020;20:669–77.
39. Makridakis S. Accuracy measures: theoretical and practical concerns. Int J Forecast. 1993;9(4):527–9.
40. Pearson K. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. Philos Trans R Soc A, Math Phys Eng Sci. 1896;187:253–318.