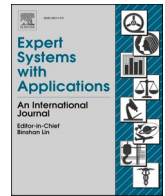




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Analysis of the COVID-19 pandemic using a compartmental model with time-varying parameters fitted by a genetic algorithm

Yuri Zelenkov<sup>\*</sup>, Ivan Reshetsov<sup>\*</sup>

HSE Graduate School of Business, HSE University, 109028, 11 Pokrovsky blv., Moscow, Russian Federation

## ARTICLE INFO

### Keywords:

COVID-19 pandemic modeling  
Compartmental model  
SEIR model extension  
SEIR model with time-varying parameters  
Actual number of infectious  
Effectiveness of vaccines

## ABSTRACT

Analyzing the COVID-19 pandemic is a critical factor in developing effective policies to deal with similar challenges in the future. However, many parameters (e.g., the actual number of infected people, the effectiveness of vaccination) are still subject to considerable debate because they are unobservable. To model a pandemic and estimate unobserved parameters, researchers use compartmental models. Most often, in such models, the transition rates are considered as constants, which allows simulating only one epidemiological wave. However, multiple waves have been reported for COVID-19 caused by different strains of the virus. This paper presents an approach based on the reconstruction of real distributions of transition rates using genetic algorithms, which makes it possible to create a model that describes several pandemic peaks. The model is fitted on registered COVID-19 cases in four countries with different pandemic control strategies (Germany, Sweden, UK, and US). Mean absolute percentage error (MAPE) was chosen as the objective function, the MAPE values of 2.168%, 2.096%, 1.208% and 1.703% were achieved for the listed countries, respectively. Simulation results are consistent with the empirical statistics of medical studies, which confirms the quality of the model. In addition to observables such as registered infected, the output of the model contains variables that cannot be measured directly. Among them are the proportion of the population protected by vaccines, the size of the exposed compartment, and the number of unregistered cases of COVID-19. According to the results, at the peak of the pandemic, between 14% (Sweden) and 25% (the UK) of the population were infected. At the same time, the number of unregistered cases exceeds the number of registered cases by 17 and 3.4 times, respectively. The average duration of the vaccine induced immune period is shorter than claimed by vaccine manufacturers, and the effectiveness of vaccination has declined sharply since the appearance of the Delta and Omicron strains. However, on average, vaccination reduces the risk of infection by about 65–70%.

## 1. Introduction

In December 2019 a new virus causing severe acute respiratory syndrome was identified in Wuhan, China. Within a short time, cases of new disease were detected in other countries. The growing threat has entailed the worldwide response and reaction. In March 2020, the World Health Organization (WHO) upgraded the COVID-19 outbreak to a pandemic (CRS, 2022). Due to the increasing risk of the pandemic, governments have begun to develop certain policies to slow the spread of infection, reduce the workload of the public healthcare system and reduce the mortality rate, while trying not to halt the economic development and helping people to get through temporary suspending of certain processes and services (GRT, 2022; Miikkulainen et al., 2021). It is worth noting that most of these policies were introduced in response

to the certain changes of the course of the pandemic, namely “waves” characterized by the significant outbreaks of the number of new infections, hospitalizations, or deaths.

Since the pandemic began, pharmaceutical companies and medical institutions around the world have started developing vaccines against the new disease. Notwithstanding widespread speculation and indecision by much of the population about the vaccine, by March 31, 2022, more than 5 billion people worldwide had been vaccinated. In addition, in the fall of 2021, a revaccination campaign was initiated, triggered by a waning immunity, and involving the administration of a booster dose of vaccine at intervals ranging from a few months to a year (Assefa et al., 2022).

At the time of preparing this work (late 2022), the pandemic can be considered over, but there is still considerable debate about its actual

<sup>\*</sup> Corresponding authors.

E-mail address: [yuri.zelenkov@gmail.com](mailto:yuri.zelenkov@gmail.com) (Y. Zelenkov).

<https://doi.org/10.1016/j.eswa.2023.120034>

Received 13 December 2022; Received in revised form 13 March 2023; Accepted 1 April 2023

Available online 5 April 2023

0957-4174/© 2023 Elsevier Ltd. All rights reserved.

scope (i.e., the number of cases), the effectiveness of pandemic controls (in particular, vaccine effectiveness), etc. These assessments are important for developing policies to respond to similar challenges that may arise in the future.

However, the problem in obtaining such an assessment is that official statistics contain data only on registered COVID-19 cases and deaths. Therefore, models are needed to extract information on the dynamics of other subpopulations affected by the pandemic (unregistered infected, vaccinated, etc.) from the available data to study the effectiveness of various pharmaceutical and non-pharmaceutical interventions. These challenges have attracted the attention of many researchers, not only health professionals, but also specialists in mathematical and computer modelling (Chang et al., 2021; Miikkulainen et al., 2021).

An area of particular interest is the compartmental models that describe the development of infection (Bjørnstad et al., 2020). The classical SIR model (Kermack and McKendrick, 1927; 1933) divides a population into three groups: susceptible ( $S$ ), infectious ( $I$ ) and recovered ( $R$ ), and defines coefficients determining the transition rate between compartments. As we noted above, the observed data do not contain information about the dynamics of all groups, most often only the number of registered cases is available. However, because the model is deterministic, it allows us to reconstruct transition rates from the data and thus obtain an estimate of all groups (Bjørnstad et al., 2020; Otunuga, 2021). This is the main advantage of the compartmental models. Thus, the problem is to develop a model that corresponds to reality as closely as possible, that is, to identify the necessary compartments and the links between them.

For COVID-19, various modifications of the SIR model have been proposed, including additional groups such as exposed, vaccinated (Acuña-Zegarra et al., 2021; Schlickeiser and Kröger, 2021), hospitalized (Arik et al., 2020; Capistran et al., 2021), asymptomatic or unregistered cases (Ivorra et al., 2020; Liu et al., 2020), etc. Fitting variables of these models to the observed values provides important insights into the unobserved groups, as well as assessing the effectiveness of various epidemic control methods (Wibbens et al., 2020; Feng et al., 2021) and economic and social impacts (Boissay et al., 2020; Karin et al., 2020).

The model is usually fitted by numerically solving corresponding differential equations and optimizing its parameters (Brewer et al., 2008). In most cases the model parameters, i.e., the transition rates between compartments, are assumed to be constant (not varying with time). However, in real life, many parameters of the system vary in time. Government policies to limit contact affect the infection rate (Boissay et al., 2020; Wibbens et al., 2020), and the rate of vaccination varies with the availability of vaccines and efforts to promote vaccination (Jentsch et al., 2021; Bruxvoort et al., 2021; Rella et al., 2021; Feikin et al., 2022; Yu et al., 2022). In addition, the development of medical protocols affects COVID-caused death rate (Knight et al., 2020). Moreover, the probability of registering the COVID cases is highly influenced by the testing volume and the pandemic handling strategy (Ivorra et al., 2020; Rippinger et al., 2021). Additionally, the emergence of new strains of the virus seriously affects the nature of the outbreak (Lopez Bernal et al., 2021; Shah and Woo, 2022). Strictly speaking, we should consider these events as external to the system under study but affecting its behavior.

Moreover, compartmental models with constant parameters can describe the spread of infection as a single epidemiological wave with a single peak (Comunian et al., 2020); however, at least three peaks were observed for the COVID-19 pandemic. This is why the dependence of parameters on time is crucial for the analysis. It should be noted that some authors assume that the parameters of the system can change over time but introduce additional constraints. For example, Schlickeiser and Kröger (2021) assume that transition rate ratios are constant; Ivorra et al. (2020) establish controls in the form of functional dependencies. We can state that fitting a compartmental model with time-varying transition rates is a difficult task and, to our knowledge, there is no generally accepted method today.

To overcome these limitations, we propose an extended compartmental model, in which transition rates are not constant. This approach allows us to extract time dependencies from the data and find more realistic distributions. The system parameters are estimated using a genetic algorithm. Thus, the contributions of our work can be summarized as follows:

1. We introduce an extended compartment model that allows us to extract the actual number of infected individuals and the actual population with immunity (both natural and vaccine-induced) from empirical data containing only registered COVID-19 cases and deaths.
2. We treat model parameters as time-varying, model is fitted using genetic algorithms. This allows us to get a pandemic simulation with multiple peaks, which is more consistent with reality.
3. The results obtained for four countries (Germany, United Kingdom, United States and Sweden) are compared with each other and provide important insights into pandemic spread.

It should be noted, however, that by using this approach we sacrifice predictive capabilities because, by acting in this way, we can reconstruct the time series based on available past observations. The further dynamics of the transition rates remain unknown; therefore, multivariate time series must be predicted to predict the behavior of the system. On the other hand, because this approach allows post factum estimation of many parameters that cannot be measured directly, simulation results, together with observed indicators such as infection mortality rates or latency period length, provide insight into the spread of the pandemic and control methods, which is very important for analysis and policy making.

The rest of papers is organized as follows. Section 2 contains the description of the compartmental approach and the analysis of the recent papers devoted to the COVID-19 pandemic modelling. Section 3 describes the applied epidemic model and data used for the fitting. Section 4 presents the obtained results and their discussion. Finally, all the conclusions are summarized in Section 5 where the opportunities for further research are also discussed.

## 2. Related research

### 2.1. Compartmental model

Here we introduce some definitions that will be required for further discussion. The general modelling technique of infectious diseases is a compartmental model. The classical SIR model (Kermack and McKendrick, 1927; 1933) divides population on three compartments (groups): susceptible ( $S$ ), infectious ( $I$ ) and recovered ( $R$ ), with a total population size  $N = S + I + R$ . The amount of people assigned with each compartment varies over time, the progress of individuals between compartments can be presented as a flow diagram and a corresponding differential equations.

The movement of individuals from one compartment to another is determined by the constant transition rate. It is assumed that the time individual spent in the state  $X$  before to progress to state  $Y$  is a random variable with an exponential distribution and rate parameter  $\theta_{XY} > 0$ . So, the  $1/\theta_{XY}$  is the mean period that individual spent in  $X$  and  $\theta_{XY}$  itself is the number of people moved from  $X$  to  $Y$  in the time interval  $dt$ .

The infection rate is often considered a consequence of the number of contacts between individuals of susceptible and infected groups. Then it can be presented as  $\beta SI/N$ , where  $\beta$  is the proportion of contacts leading to infection.

Further, we will view all compartmental variables as normalized by the total population  $N$ , so for SIR model  $S + I + R = 1$ . Thus, the corresponding ordinary differential equations (ODE) are

$$\frac{dS}{dt} = -\beta \frac{SI}{N}$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I,$$

where  $\gamma$  is the transition rate from infectious to recovered compartments.

However, there is a latent period between being infected and becoming infectious for most infectious diseases: the exposed group ( $E$ ). This extension is considered in SEIR model (Bjørnstad et al., 2020). The time spent in the exposed state is also assumed to be a random variable with exponential distribution and average value  $\sigma$ . Since immunity after recovery is temporary for many infections, individuals in group  $R$  will lose immunity and return to  $S$  with transition rate  $\omega$ . SEIR model also considers mortality that contributes to flows between groups. Death due to infection will cause a loss of individuals from the  $I$  group at a rate  $\alpha$ , and all groups will experience background death from other causes at a rate  $\mu$ . For the time periods, when total population does not change significantly, it can be assumed constant  $N(t) = N$ , so birth and natural death rates in the SEIRS model are the same and represented by  $\mu$ .

## 2.2. COVID-19 epidemic modelling

In this section we will turn specifically to the research that are aimed at the modelling of COVID-19 pandemic. All epidemic simulation studies can be classified in two ways. First taxonomy bases on the structure of the compartmental model. Many authors use the traditional SEIR model, while others expand it by introducing additional compartments. This allows them to consider subpopulations (most often unobserved) that are important in the context of a particular study. Second, studies can be divided by the way in which model parameters (transition rates) are considered. These rates can be treated as constant or variable. As already noted, a model with constant parameters allows modeling only one peak epidemic. However, finding variable parameters is a difficult task and there is no generally accepted way to do it. Table 1 presents taxonomy of all the publications discussed below.

Authors who use the classical SEIR model with constant transition rates (SEIR-CTR) focus mainly on assessing the effectiveness of epidemic control methods such as lockdown. For example, Rawson et al. (2020) applied the SEIR-CTR model to estimate the efficiency of two possible lockdown exit strategies in the UK. The first strategy supposed the gradual canceling of restrictions, introduced by the government at the start of pandemic, while the second one included full elimination of restrictions but with their temporary reintroduction in case of new outbreaks.

Karin et al. (2020) used SEIR-CTR models to estimate important parameters affecting outbreak duration and severity and then predicted changes in their values caused by various restrictive policies. The results of the study suggested a new approach to an exit strategy, which is to alternate a 4-day work period with a 10-day isolation period in succession. The study also shows that the SEIR model is very sensitive to the

level of infection and the latency period of the virus.

Feng et al. (2021) used SEIR-CTR to estimate cumulative number of registered infectious in Wuhan to assess the impact of strict lockdown. The authors also train Artificial Neural Networks (ANN) to predict development of epidemic in other parts of China.

Another line of research focuses on subpopulations that are important for pandemic control (vaccinated, hospitalized, asymptomatic cases, etc.) but are not considered in the traditional SEIR model. To this end, the authors extend the SEIR model by adding compartments to track these subpopulations. To analyze COVID-19 vaccination policies, Acuña-Zegarra et al. (2021) proposed a model that expands SEIRS considering the temporal immunity after vaccination, vaccine imperfectness and two groups of infected individuals: symptomatic and asymptomatic. According to authors hypotheses, the vaccine has effects only on susceptible individuals. Thus, susceptible individuals move to  $V$  (vaccinated) group at  $v$  rate. By analogy with the loss of immunity of recovered individuals, the mean time of immunity waning after vaccination is also a random variable with an exponential distribution and mean  $\lambda$ . The following proposition is that the vaccine is imperfect. Thus, a fraction of individuals in  $V$  may also become infected, however, with a lower probability than those in the  $S$  class. This probability is determined by vaccine efficacy  $\epsilon$ . Finally, in Acuña-Zegarra et al. (2021) model, exposed individuals remain in class  $E$  until they become infectious and move to either symptomatic or asymptomatic class. Asymptomatic patients can be detected and recorded, for example, using medical tests or postmortem investigation. At the same time, symptomatic patients may not seek medical attention and therefore avoid registration.

Arik et al. (2020) presented a study using the extended model with inclusion hospitalized and asymptomatic compartments. The presence of the latter group is particularly important because it allows one to determine the real number of people infected with COVID-19 since the significant portion of population may not show any symptoms while being infectious. The model also extends the hospitalized group to those who are under intensive care or on a ventilator.

Capistran et al. (2021) implemented a model which is aimed to forecast a hospitalization rate during the outbreaks of COVID-19 pandemic. The epidemic model proposed in the article divides the hospitalized compartment to the patients with mild symptoms, those who require intensive care unit (ICU) such as respirators and the patients in the critical state. Additionally, infected members are also split into registered and unregistered ones with the different infection rate for both groups. The article also estimates the efficiency of various governmental policies against the pandemic such as setting lockdowns or changing the duration of quarantine for those who were in contact with the infected people. The proposed model was applied on the daily mortality and hospitalization data reported from 32 states of Mexico during March – July 2020. As a result, the accuracy of predicting the hospital bed occupancy varied from 0.8 to 0.95 between the states. Besides, another result of the study is the suggestion of the periods for reducing the restrictive policies and lockdown due to the decrease of COVID-19 infections. However, the research did not consider the change in the hospital residence times, therefore it can't be applied for a long-term period due to erratic behavior of pandemic. Authors also assume this limitation to be a significant obstacle for estimation of lockdown exit strategies.

Yarsky (2021) extends the SEIR model by adding a new compartments  $A$  to track asymptomatic individuals and  $T$  to predict the rate at which diagnostic tests are performed and yield positive results. The author uses genetic algorithms for estimation of transition rates, where genome comprises a list of coefficients which do not vary over time. Fitness function is computed by comparing model results to the reported daily numbers of infections and deaths.

As we noted above, the traditional compartmental model with constant transition rates allowing to simulate only one peak in the spread of infection, so many authors use various techniques to account for changes in model parameters over time. Bouchnita et al. (2021) adapted SEIR

**Table 1**  
Compartmental models in the COVID-19 analysis literature.

		Model structure	
		SEIR	Extended
Model transition rates	Constant	Feng et al. (2021) Karin et al. (2020) Rawson et al. (2020)	Acuña-Zegarra et al. (2021) Arik et al. (2020) Capistran et al. (2021) Yarsky (2021)
	Varying	Bouchnita et al. (2021) Eryarsoy et al. (2021) Vega et al. (2022)	Ivorra et al. (2020) Otunuga (2021) Schlickeiser and Kröger (2021)

model to review the restrictions introduced by Vietnamese government to stop the first wave of COVID-19 pandemic. Authors divided the observed timespan into two periods. The first one starts with the first appeared COVID-19 case in Vietnam, while the second period goes from the introduction of the nationwide lockdown till the easing of the restrictions. Therefore, despite the parameters of the model are still constant throughout each period, changes of their values between the periods can be observed.

Vega et al. (2022) integrated machine learning (ML) into the SIR model to predict the number of people infected with COVID-19 over a 1–4 week period in Canada and the United States. At each iteration, the ML model was used to estimate whether infection rates would change. If such a change is expected, new SIR model parameters are estimated.

Eryarsoy et al. (2021) focused on predicting the number of cases and deaths using the SIR model. The authors based their work on the assumption that the transition rates in the SIR model can be represented by diffusion models (S-curves) commonly used in business research.

There are also several papers based on different techniques for accounting for the variability of transition rates in extended SEIR models. Otunuga (2021) proposed two improvements; first, the author extended the SEIR model by dividing the infected class into asymptomatic and symptomatic classes, and second, he suggested that transmission, symptomatic recovery, and immunity rates are functions of time, but the asymptomatic recovery rate is constant. He also assumed that the transmission rate and symptomatic recovery rate are stochastic variables whose fluctuations are modeled with Gaussian white noise. The author then determined the unknown parameters using a generalized method of moments on daily case data of infected and recovered individuals over a one-year period in the United States.

Ivorra et al. (2020) presented a model including unregistered cases, hospitalization, and mortality rates, originally applied to the Ebola pandemic analysis, and adapted it for COVID-19. A distinctive feature of the proposed model was the division of the hospitalized subpopulation into those who would recover and those who would die. In addition, the effect of various public policies on controlling outbreak severity, such as isolation, quarantine, and increased medical resources, was analyzed. This was done by introducing time-varying functions representing the efficiency of the control measures applied to the corresponding compartments. The study used data on infections reported in China, and parameters were estimated using multi-objective optimization. According to the results, unreported cases may have caused about 52 percent of the total number of infections in the first two months of the pandemic. In addition, increasing population testing has a huge impact on the detection rate and can reduce the outbreak, as can implementing restriction policies. However, the effect that these measures have on infection rates is delayed by about two weeks.

Schlickeiser and Kröger (2021) proposed an extended model that accounts for vaccination of the susceptible population. However, this model considers immunity acquired after vaccination as permanent, which is not valid for many infections. Another exciting feature is that the authors believe that the parameters of the equations depend on time. However, in the following discussion, they assume that the ratios of the coefficients are constant. It allows them to study many model features, but they do not investigate the impact of changing parameters.

This is a review of a limited number of papers from a stream of hundreds of publications, however, it provides a general view of the use of compartmental models in COVID-19 pandemic research. It can be concluded that compartmental modelling is applied for a variety of purposes: predicting the total number of infected, hospitalized and in need of ICU; assessing infection-handling policies, economic impact and so on. However, in most cases the authors propose models with constant transition rates, which, as noted in the Introduction, do not correspond to practice.

The classical SIR model and its variations treat transition rates as constants, assuming the time an individual spends in a particular group is a random variable with an exponential distribution. Thus, such a

model describes the spread of infection as one epidemiological wave with a single peak, whereas for COVID-19 at least three waves caused by different virus strains have been reported. In addition, different actions to control the pandemic also affect the transition rates. Therefore, it is necessary to find more realistic distributions of the system parameters. However, many authors considering time-varying transition rates introduce additional constraints that implicitly or explicitly constrain the distributions.

### 3. Proposed method

This section presents the compartmental model we developed and the proposed method for fitting it, both solutions aimed at eliminating the gaps identified above. The model (Section 3.1) includes additional groups that allow us to account for subpopulations of interest for the analysis of pandemic experience (vaccinated individuals and unregistered cases). This section also provides a rationale for selecting transition rates that vary over time. Section 3.2 describes how to fit the model.

#### 3.1. Compartmental model

The following section focuses on the description of the developed approach for modelling the outbreak of COVID-19 pandemic. Proposed extension of SEIRV (Susceptible – Exposed – Infectious – Recovered – Vaccinated) model is presented in Fig. 1.

Firstly, we consider division of the population on six compartments: vaccinated ( $V$ ), susceptible ( $S$ ), exposed ( $E$ ), infected ( $I$ ), dead ( $D$ ) and recovered ( $R$ ). According to the concept of compartmental epidemic modelling, all the individuals initially belong to susceptible compartment but then move on to other groups with the start of the pandemic. The introduction of the vaccinated compartment is crucial to analyze the effects of the vaccination strategies implemented in different countries, determine the efficiency and effectiveness of various vaccines used to eradicate the disease and evaluate the immunity induced by vaccine. Various publications show that vaccination induces higher immunity level than prior infection (Yu et al., 2022). Moreover, WHO recommends people to get the jab even if they were previously infected with COVID-19. Therefore, it is highly important to distinguish vaccinated persons into a separate group.

According to many studies (e.g., Acuña-Zegarra et al., 2021; Milne et al., 2021) vaccine has effect only on susceptible individuals. Thus, susceptible individuals move to  $V$  group at  $\nu$  rate. Moreover, although approved vaccines provide a high level of protection against severe disease and death, no vaccine extends 100% protection against possible infection (which is mostly asymptomatic or with mild symptoms). In other words, individuals protected by vaccine can also get infected with COVID-19 but with significantly lower probability than persons from  $S$  compartment. Consequently, it is safe to assume that a vaccine has a level of efficacy  $\varepsilon$  by which the possibility of being infected can be determined (Acuña-Zegarra et al., 2021). At the same time, despite producing strong antibody response, vaccines can't provide permanent protection against the infection. Although, some researchers suggest that the decrease in level of antibodies does not lead to the declining protection, there is evidence of immunity waning several months after the vaccination (Baraniuk, 2021). Therefore, it is required to introduce the possibility for individuals to move from vaccinated to susceptible compartment. In terms of the proposed model, this implies introducing the mean time of immunity waning after vaccination which is a random variable with an exponential distribution and mean  $\lambda$ . It should also be noted that  $V(t)$  in our model relates to the number of people protected by the vaccine at time  $t$ . Published vaccinated data represents total number of vaccinated cumulatively, regardless the waning of the vaccine-induced immunity, which is inconsistent with our definition of  $V(t)$ . Thus, we treat  $V(t)$  as unobserved variable.

The important feature of the proposed model is an introduction of exposed compartment as an intermediate between susceptible and

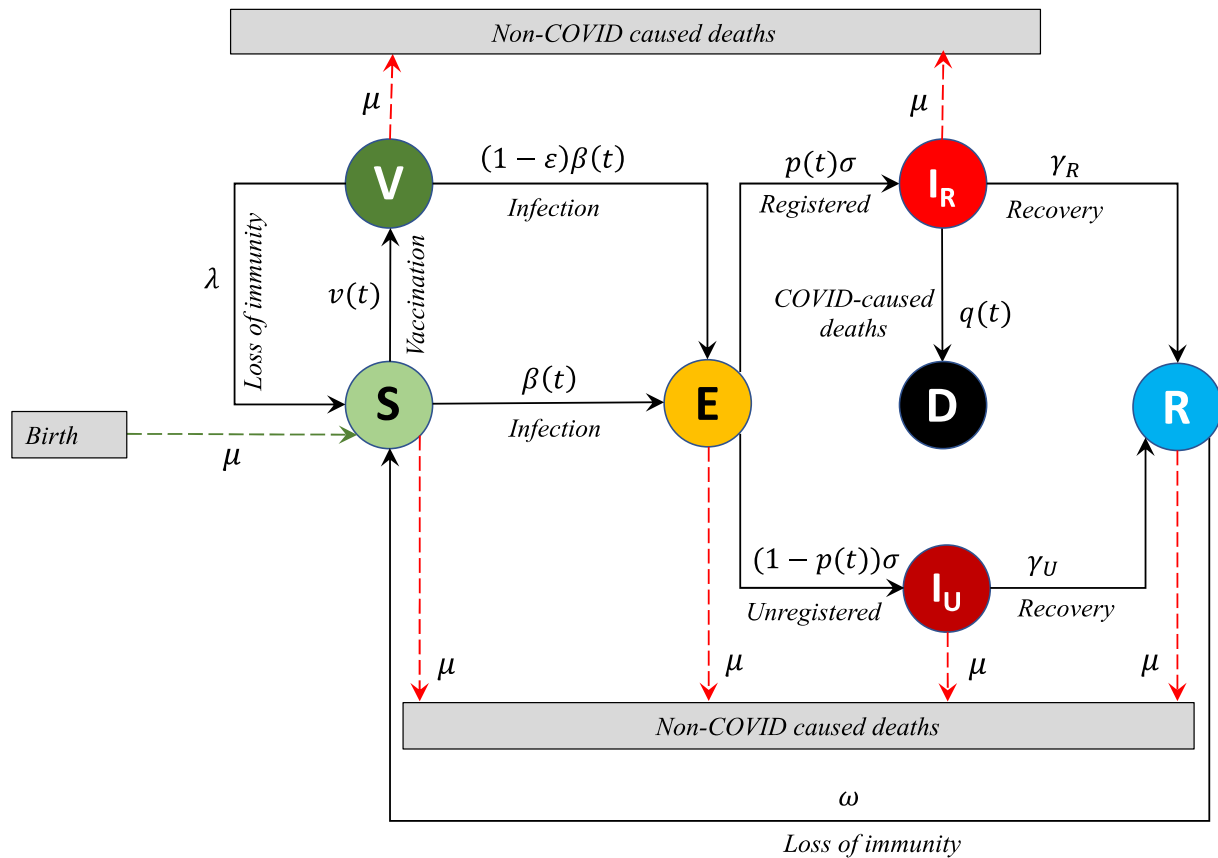


Fig. 1. SEIRV model.

infected individuals. As mentioned above, COVID-19 is characterized by the existence of incubation (latency) period during which a person is infected but not infectious. The presence of such a compartment in the model provides an opportunity to estimate the infection rate more precisely as well as consider the differences between the different strains of the virus. Moreover, the duration of the latency period is a vital index in epidemiology that helps to understand the spread of the disease and apply specific handling restrictions and policies for different areas and countries at various times (Cheng et al., 2021). The last point is an extremely huge advantage in view of the time-dependent character of the suggested model. As a result, the individuals transfer from susceptible and vaccinated compartments to the exposed group with the certain probability that in case of vaccinated population also considers the vaccine efficacy. This probability  $\beta(t)$  represents the infection rate which is the crucial parameter of the epidemic model as it refers to the speed at which individuals get infected. Nevertheless, the infected compartment of population in the developed approach is significantly different from classical SIR models. We follow the technique used in contemporary research (Liu et al., 2020; Acuña-Zegarra et al., 2021) which states that exposed individuals remain in class E until they become infectious and move to either symptomatic or asymptomatic class.

Due to the nature of virus-related diseases the number of cases reported by healthcare institutions is usually underestimated. Generally, the difficulty of calculation is due insufficient testing, data depression of mild or asymptomatic patients, limited awareness of the virus in the general population and a time-lag bias (Rippinger et al., 2021). Therefore, the accurate data can be provided only for the certain part of the infected population which is released into a separate class used further on as observable variable in the model. Another class, for its part, provides useful information about the overall magnitude of the pandemic and shows the cases that are not considered in the official reports and

statistics. However, instead of focusing on the level of the symptoms, we call these classes  $I_R$  (infected registered) and  $I_U$  (infected unregistered) following Liu et al. (2020), since this more accurately reflects reality. Asymptomatic patients can be detected and recorded, for example, using medical tests or postmortem investigation. Compulsory testing introduced in several countries such as Hong Kong or Austria resulted in the identification of the significant number of asymptomatic cases which were included in the COVID-19 statements. At the same time, symptomatic patients may not seek medical attention and therefore avoid registration. Therefore, a considerable share of patients with symptoms can be not registered in the reports. Let  $p(t)$  be the probability that an exposed individual will be registered, thus the corresponding transition rate to  $I_R$  is  $p(t)\sigma$  and transition rate from E to  $I_U$  is  $(1 - p(t))\sigma$ , where  $1/\sigma$  corresponds to the latency period.

Another important aspect of the infected compartment is the nature of the infection rate  $\beta(t)$  considering the separation of the group on registered and unregistered classes. Notwithstanding the existence of the papers where infection rate varies for these classes (e.g., Liu et al., 2020), we assume that the proportion of contacts leading to infection (i. e.,  $\beta(t)$ ) remains the same for both types of contacts. This assumption is based on the fact that both symptomatic and asymptomatic individuals are presented in each class and on the challenging estimation of the asymptomatic people' contribution to the spread of the virus. Additionally, a single value of  $\beta(t)$  allows to explicitly include the contact frequency in the model and is hence beneficial for the analysis and interpretation of its results and outcomes. Therefore, the transition from class S to class E is determined by the frequency of contacts of susceptible people with registered and unregistered infectious,  $\beta I_R/N$  and  $\beta I_U/N$  correspondingly.

It should be also noticed that the output  $I_R$  compartment is split into three outgoing flows: the deaths caused by COVID-19, the deaths caused by other reasons and the recovered compartment. By contrast, the

individuals from  $I_U$  class can move only to recovered group or to the non-COVID caused deaths. Although unregistered patients can also die from the decease, it is complicated to estimate the exact number of them due to the lack of the common approach of tracking this measure. For example, monitoring excess mortality based on postmortem investigation is the most widespread method for calculating the precise number of COVID-induced deaths. However, long-term effects of COVID-19 more frequently reported by the recovered individuals as well as so called post-COVID-19 symptoms may also be the cause of deaths of a large number of individuals (Davis et al., 2020). Presence of these factors creates serious obstacles for estimation of the overall number of deaths caused by COVID-19. Therefore, the  $D$  in the proposed compartment includes only those people, who were officially registered as COVID-19 positive patients and their death was directly caused by the infection.

By analogy with the loss of immunity of vaccinated individuals described above, the mean time of immunity waning after recovery is also a random variable with an exponential distribution and mean  $\omega$ . The following proposition is that the natural immunity is imperfect and does not provide lifelong protection against the decease. It is also worth mentioning that birth and death rate is included in the model and is represented by the parameter  $\mu$ , which appears in each transition phase between the compartments and relates to all the vital dynamics that is not related with COVID-19. The introduction of this coefficient makes the model more realistic because it considers change in the size of the initial population.

As noted above, the classical SIR model and its variations treat transition rates as constants, and such models describe the spread of infection as one epidemiological wave with a single peak (Comunian et al., 2020), whereas for COVID-19 at least three waves caused by different virus strains have been reported. Therefore, we consider that many parameters of the system vary in time. Government policies to limit contact affect the  $\beta(t)$ , and the rate  $v(t)$  varies with the availability of vaccines and efforts to promote vaccination. In addition, the development of medical protocols affects the probability  $p(t)$  and rate  $q(t)$ . Moreover, the probability of registering the COVID cases is highly influenced by the testing volume and the pandemic handling strategy. Additionally, the emergence of new strains of the virus seriously affects the nature of the outbreak, that's why time-dependency of the parameters is crucial to analyze. Therefore, introducing time-dependency of the parameters helps to fulfil the task of the study which is to estimate the effect of the governmental policies on the course of the epidemics. All other parameters can be considered constant for the simplification purposes.

Let  $N(t) = V(t) + S(t) + E(t) + I_R(t) + I_U(t) + R(t)$  and  $N(t) + D(t) = 1$ . Thus, corresponding system of ordinary differential equations (ODE) is

$$\begin{aligned} \frac{dV(t)}{dt} &= v(t)S(t) - (\mu + \lambda)V(t) - (1 - \varepsilon)\beta(t)V(t) \frac{I_R(t) + I_U(t)}{N(t)} \\ \frac{dS(t)}{dt} &= \mu N(t) + \lambda V(t) - [\mu + v(t)]S(t) - \beta(t)S(t) \frac{I_R(t) + I_U(t)}{N(t)} \\ \frac{dE}{dt} &= \beta(t)[S(t) + (1 - \varepsilon)V(t)] \frac{I_R(t) + I_U(t)}{N(t)} - (\mu + \sigma)E(t) \\ \frac{dI_R(t)}{dt} &= p(t)\sigma E - [\mu + q(t) + \gamma_R]I_R(t) \\ \frac{dI_U(t)}{dt} &= [1 - p(t)]\sigma E - (\mu + \gamma_U)I_U(t) \\ \frac{dR(t)}{dt} &= \gamma_R I_R(t) + \gamma_U I_U(t) - (\mu + \omega)R \\ \frac{dD(t)}{dt} &= q(t)I_R(t) \end{aligned} \tag{1}$$

Tables 2 and 3 list the variables and parameters of the system correspondingly.

**Table 2**  
Variables of SEIRV model.

Variable	Description	Observable
$N(t)$	Total number of individuals in the population	Yes
$S(t)$	Number of susceptible individuals	No
$V(t)$	Number of individuals protected by vaccine	No
$E(t)$	Number of exposed individuals	No
$I_R(t)$	Number of individuals registered as COVID-19 positive patients, regardless of symptoms	Yes
$I_U(t)$	Number of individuals infected with COVID-19 but not registered in the official reports	No
$D(t)$	Number of individuals officially registered as infected with COVID-19 whose death was caused by the virus	Yes
$R(t)$	Number of recovered individuals	No

**Table 3**  
Parameters of SEIRV model.

Parameter	Description	Time dependent
$\mu$	Natural death / birth rate	No
$v(t)$	Vaccination rate	Yes
$\lambda$	Waning rate of vaccine, $1/\lambda$ is the average time to lose vaccine-induced immunity	No
$\varepsilon$	Vaccine efficacy	No
$\beta(t)$	Fraction of contacts leading to infection	Yes
$\sigma$	Latency rate, $1/\sigma$ is the average latency period	No
$p(t)$	Exposed individuals' fraction who become registered infectious	Yes
$\gamma_R$	Recovery rate of registered infected individuals, $1/\gamma_R$ is the average time which registered individuals leave being infectious and contagious	No
$\gamma_U$	Recovery rate of unregistered infected individuals, $1/\gamma_U$ is the average time which unregistered individuals leave being infectious and contagious	No
$q(t)$	Infection-induced death rate	Yes
$\omega$	Rate of loss of natural immunity, $1/\omega$ is the average time to lose natural immunity	No

### 3.2. Parameters estimation and optimization

As ODE are most widely used for describing the temporal evolution of a large range of systems, the problem of fitting parameters has attracted significant attention of researchers. Estimating model parameters from data requires two components (Brewer et al., 2008). The first one is an error function  $E_D(\theta)$  that quantifies the difference between output of model with parameters  $\theta$  and the data  $D$ . The second component is the optimization method that finds the value of  $\theta$  that minimizes  $E_D(\theta)$ .

Here we will focus on so-called solution-based approaches, which require the analytical solution of Eq. (1), or the approximate solution obtained numerically if an analytical one does not exist. In the last case, the approximate solution is

$$\frac{d\hat{\mathbf{x}}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), \theta, t)$$

which more often is accompanied by the least-squares error function, as it should be based on the difference between the reconstructed and true values. The most widespread method of solving the ODE and finding  $d\hat{\mathbf{x}}(t)/dt$  is the Runge-Kutta fourth-order method with adaptive step-size control.

The considerable advantage of this approach is that it suitable for use, notwithstanding the fact that some of the components of the vector  $\hat{\mathbf{x}}(t)$  can not be measured. This is particularly useful for the epidemic data as some of the compartments are non-measurable (e.g., exposed, or infected population without symptoms). In such case, the calculation of the error function is just taken over the measurable components.

Analyzing the flow of research literature, we can identify three classes of optimization methods used to find  $\theta$  that minimizes  $E_D(\theta)$ . The

first one is based on deterministic single-objective optimization techniques like the Nelder-Mead method (Brewer et al., 2008). This procedure allows finding only point estimations of  $\theta$ . The second class implements the Bayesian approach for learning the model parameters from data (Huang et al., 2020), for example, Markov Chain Monte Carlo (MCMC) and its various adaptations. The undoubted advantage of this approach is that it allows estimating the distributions of the parameters  $\theta$ . It is widely used for fitting epidemiological models (Baguelin et al., 2013; Chatzilela et al., 2019; Acuña-Zegarra et al., 2021). The third class includes nature-inspired metaheuristics methods (Boroujeni & Pashaei, 2021), in particular genetic algorithms (Katare et al., 2004). This approach is also widely used in compartmental models (Shah et al., 2007, Ivorra et al., 2020). We will also use genetic algorithms as it is the only way to fit time-varying parameters.

Thus, our task is to find such parameters of the SEIRV model, which provide the minimum deviation from the observed data for the given initial conditions. The observed data, in this case, is a time series  $I_R(t)$ ,  $D(t)$ . The initial conditions correspond to the moment before the start of the epidemic, so we can set them as

$$S(0) = N(0) = 1, V(0) = E(0) = I_R(0) = I_D(0) = D(0) = R(0) = 0.$$

Many authors have already solved a similar problem for various compartmental models. Still, in most cases, they considered all the model parameters as constant and independent of time. However, we assume that, at a minimum, the infection force  $\beta(t)$ , the probability of registering an infected individual  $p(t)$ , the rate of death of an infected person  $q(t)$  and the rate of vaccination  $v(t)$  vary with time (Table 3).

We also use combination of three techniques, namely:

1. Using difference equations instead of differential equations,
2. Using genetic algorithms to present time-dependent coefficients,
3. Fit equations using mean absolute percentage error (MAPE).

The first point concerns the use of time-dependent parameters when integrating ODE by the Runge-Kutta method. This problem arises since the integration step should be smaller than the time interval presented in observed data. The usual solution is a linear extrapolation within the interval given by the observed values. Thus, the value of time-dependent parameter  $\theta(t)$  corresponding to the current integration step can be obtained as  $\theta_{i+h} = h(\theta_{i+1} - \theta_i)/m$ , where  $m$  is the number of integration steps in the time interval;  $h$  is the number of the current point.

However, in our case, this approach did not provide the required accuracy. Therefore, we transformed ODEs equation into first-order difference equations (here  $i = 1, \dots, M$  is the observation id and  $M$  is the number of observations):

$$\begin{aligned} V_{i+1} &= V_i + v_i S_i - (\mu + \lambda) V_i - (1 - \epsilon) \beta_i V_i \frac{I_i^R + I_i^U}{N_i} \\ S_{i+1} &= S_i + \mu N_i + \lambda V_i - [\mu + v_i] S_i - \beta_i S_i \frac{I_i^R + I_i^U}{N_i} \\ E_{i+1} &= E_i + \beta_i [S_i + (1 - \epsilon) V_i] \frac{I_i^R + I_i^U}{N_i} - (\mu + \sigma) E_i \\ I_{i+1}^R &= I_i^R + \sigma p_i E_i - (\mu + \gamma_R + q_i) I_i^R \\ I_{i+1}^U &= I_i^U + [1 - p_i] \sigma E_i - (\mu + \gamma_U) I_i^U \\ R_{i+1} &= R_i + \gamma_R I_i^R + \gamma_U I_i^U - (\mu + \omega) R_i \\ D_{i+1} &= D_i + q_i I_i^R \\ N_i &= V_i + S_i + E_i + I_i^R + I_i^U + R_i \\ N_i + D_i &= 1 \end{aligned} \tag{2}$$

According to the equation regarding  $D_{i+1}$ ,  $q_i = \frac{D_{i+1} - D_i}{I_i^R}$ , both  $D$  and  $I^R$  are observable variables. Thus, we can exclude  $D_i$  and  $q_i$  from the system and we are left with only one observable variable  $I^R$  for model fitting.

(continued)

---

Algorithm 1. Fitting the SEIRV model from data

---

**Input:** initial conditions of SEIRV model (Eq. (2)), observed values of  $I_R(t)$  population size, mutation and crossover probability, max generations

10 Generate initial population  $\theta$

20 **While** stop condition is not satisfied:

30 Obtain  $\widehat{I}_{R_i}(\theta)$  solving Eq. (2)

40 Compute  $E_D(\theta)$ , Eq. (3)

50 Generate new population  $\theta$  using selection, mutation, and crossover operations

60 **End while**

**Output:** fitted SEIRV model

---

To solve the system of equations, we construct a genetic algorithm (Algorithm 1). The model fitting bases on matrix of observation  $I^R[M \times 1]$ , where  $M$  is the number of observations. So, each time-dependent rate should have  $M$  values which correspond to observations of  $I^R$ . Thus, the gene (Fig. 2) includes five subgroups: four strings with length  $M$  for the rates  $\beta(t), p(t), q(t), v(t)$  filled with real numbers and seven real numbers for the rest rates. Total gene length is  $4M + 7$ .

The core operations of genetic algorithms are selection, mutation, and crossover. While for selection, we use the standard roulette wheel rule, mutation and crossover should be modified according to gene structure. The mutation is performed simultaneously in each five gene subgroups ( $\beta(t), p(t), q(t), v(t)$ , and rest constant seven rates) with probability conducted separately for each group. Since each time-dependent rate varies in its range, the crossover is also performed separately for the subgroups, with the probability calculated independently. We use a two-point crossover in each subgroup, i.e., items exchange a subset of genes of the same length.

When fitting the compartment models using GA, most researchers use mean least squared error as a fitness function. However, analysis of the infectious data shows that they vary over a scale. Therefore, we use MAPE, which is scale-independent, as a fitness function (Tofallis, 2015):

$$E_D(\theta) = \frac{100\%}{M} \sum_{i=1}^M \left| \frac{I_i^R - \widehat{I}_i^R(\theta)}{I_i^R} \right| \tag{3}$$

here  $I_i^R$  is an observed values of the individuals infected with COVID-19 at a given time which can be obtained from official reports;  $\widehat{I}_i^R(\theta)$  approximated values obtained by solving the system of difference equations with current parameter set  $\theta$ .

### 3.3. Epidemiological data

The main objective of our study is to analyze and compare various handling strategies and approaches realized in different countries. Therefore, it is required to apply the proposed model for the data across countries with different and even opposite strategies for combating COVID-19 outbreaks. However, many researchers indicate the lack of trustfulness in the COVID-19 statistical reports issued by the healthcare and governmental institutions (Silva & Figueiredo Filho, 2021). Thus, before the selection of the countries we evaluated the reliability of COVID-19 figures produced by the compartmental epidemic models.

Balashov et al. (2021) used the Benford's Law (BL) to investigate the credibility of the data regarding total cases and the number of deaths caused by COVID-19 for 185 countries affected by the pandemic. The research revealed the violations of the BL for approximately one-third of countries. Moreover, they compared the level of deviations with the overall development of the country (based on such socioeconomic indicators as GDP, Human Development Index (HDI) and several other healthcare performance indexes). It was found that the most reliable data is provided by the countries with more developed economic systems especially regarding the death toll. Based on these results, we decided to analyze four countries, namely Germany, United Kingdom, Sweden, and the USA. These countries implemented different strategies

Algorithm 1. Fitting the SEIRV model from data

(continued on next column)



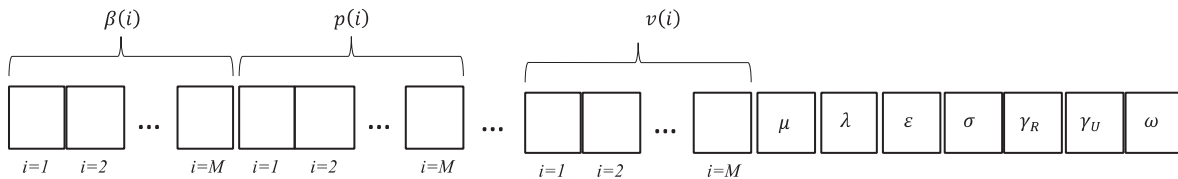


Fig. 2. Chromosome structure.

for handling the spread of the epidemic, therefore it can be beneficial to take them for comparative analysis. Besides, the effect of different vaccines against COVID-19 may be observed as the listed countries used various products for vaccination purposes.

Although proposed model requires only one observable variable  $I^R$  (which stands for the number of registered active COVID-19 cases), we also use certain other figures for analytical purposes. For example, we need the time series representing the number of people vaccinated with the first and second doze (or fully vaccinated in case of single-shot vaccine) to compare them with the variable  $V$  which shows the number of people protected by vaccine. Comparison of these values can provide the useful insights about the waning of vaccine-induced immunity and the efficacy of the current vaccines against the new strains of COVID-19. Furthermore, it is important to consider the number of confirmed deaths caused by the infection to estimate, inter alia, the excess mortality and calculate the death rate of the COVID-19. All the above-mentioned indicators were taken from the reports issued by John Hopkins University which collects and integrates the best data and expert guidance regarding the COVID-19 pandemics (CRS, 2022).

The data refers to the period between the February 16, 2020 (the first available date in CRS reports) and the January 9, 2022. It was decided to fit the model on weekly data, so the indicated period includes 100 observations.

#### 4. Results and discussion

In this section we discuss the results obtained after fitting the model to the observed data. Section 4.1 proves the accuracy of the model fit. Section 4.2 discusses compartment dynamics, which shows the spillover of individuals from susceptible group to recovered and protected by the vaccine. Also in section 4.2, we begin our discussion of the ratio of unregistered to registered cases, which continues in section 4.3. According to our results, the number of unregistered cases at the peak of the pandemic was higher than the number of registered cases. In Section 4.3 we also analyze the probability  $p(t)$  that exposed individuals become registered patients. Section 4.4 discusses the infection rate  $\beta(t)$  (the proportion of contacts between susceptible and infected individuals that result in infection) and the infection-induced mortality rate  $q(t)$ . Section 4.5 discusses the effectiveness of vaccination. We find a dramatic decrease in the effectiveness of vaccine-induced protection in the second half of 2021. Section 4.6 discusses parameters that are commonly used in pandemic analyses (time to loss of vaccine-induced and natural immunity, latency period, etc.).

##### 4.1. Model fitting

The most important issue in model fitting is the finding global minimum of the error function  $E_D(\theta)$ . Since the model is fitted only to the number of registered infected individuals  $I_t^R$ , given the large number of model parameters, it is very likely that the parameters cannot be determined uniquely. The genetic algorithm belongs to the class of heuristic algorithms, hence different runs of the same algorithm with the same initial conditions can produce different approximations to the optimal solutions. Therefore, we chose an approach based on multiple runs of the Algorithm 1. After each run, we calculated the average values of the parameters based on all runs performed. The process was

continued until the change in the averages became statistically insignificant. Our experiments have shown that the optimal number of runs is 10. Each run was performed with the following hyperparameters: 300 individuals in the population, 2000 generations, mutation and cross-over probabilities equal to 0.2 and 0.7, respectively.

Fig. 3 presents the fitted value  $\hat{I}_t^R(\theta)$  resulting from 10 runs of Algorithm 1 compared to the observed  $I_t^R$  values for all countries analyzed over the entire study period. The resulting MAPEs are 2.168, 2.096, 1.208 and 1.703 for Germany, Sweden, the United Kingdom, and the United States respectively. All the graphs contain also the 95% confidence interval. It is important to note that the actual time series of all countries fit into the confidence interval, which indicates the high quality and accuracy of the implemented model.

Slight discrepancy between the actual and fitted data starting from December 2021 can be explained by the appearance of the new Omicron variant which is characterized by enhanced transmissibility and higher resistance to the existing vaccines and the natural immunity (Shah & Woo, 2022). The studied period contains only the starting weeks of the Omicron wave, so the model may have produced the results of lower accuracy for December 2021 that, nevertheless, get into the 95% confidence interval. In addition to the peak caused by Omicron, all the countries faced another peak of the number of registered cases in January 2021. This wave was probably caused by the Alpha variant of COVID-19 that emerged in December 2020 in UK and spread quickly in the following months.

The overall share of infected registered compartment  $I_R(t)$  is significantly higher for the USA and UK (up to 0.06 at the peak) than for Germany and Sweden (up to 0.02 at the highest peaks of the pandemic).

##### 4.2. The dynamics of compartments

Fig. 4 depicts the evolution of the compartments for the analyzed countries starting from the first week after the launching of vaccination campaign (January 2021). The general pattern for every country is the gradual decline of the share of susceptible compartment from the highest values between 0.6 and 0.8 at the start of the given period to the lowest values between 0.05 and 0.1 at the end. In contrast, the share of recovered individuals and those protected with vaccine is constantly increasing and make up most population by the start of 2022. Altogether these two compartments constitute up to 80% of the total population of the examined countries by the end of the observation period. The share of recovered individuals varies between the countries, being the highest in Sweden and the lowest in Germany. These differences mainly arise from the vaccine efficacy, vaccination rate and rate of loss of natural immunity – the parameters that will be thoroughly analyzed later.

The compartment of exposed individuals is the third-largest group in all countries except Sweden which attributes to the difference in latency rates between the countries. The number of unregistered infected individuals exceeds the number of registered cases in all countries. It should be also noted that this number is the lowest for Germany than for other three countries. In addition, it is worth noting the difference in the proportions of registered and unregistered infected. The USA and UK have more registered cases in the total amount of infected individuals than Sweden and Germany.

It can be explained by the differences in the testing volumes in these countries (Hasell et al., 2020) presented in Table 4. Testing rate in

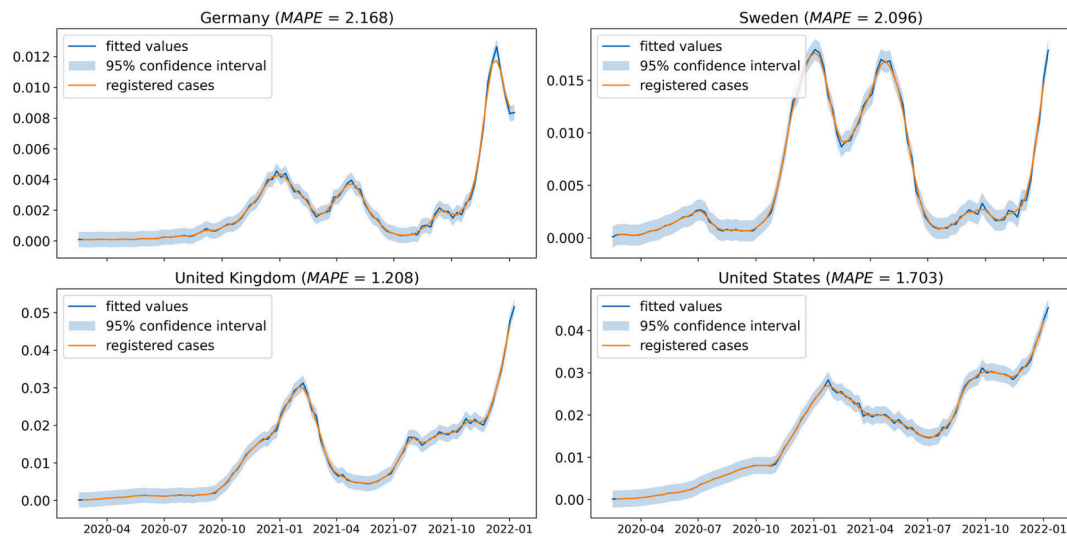


Fig. 3. Fitted vs observed values of the registered infected individuals.

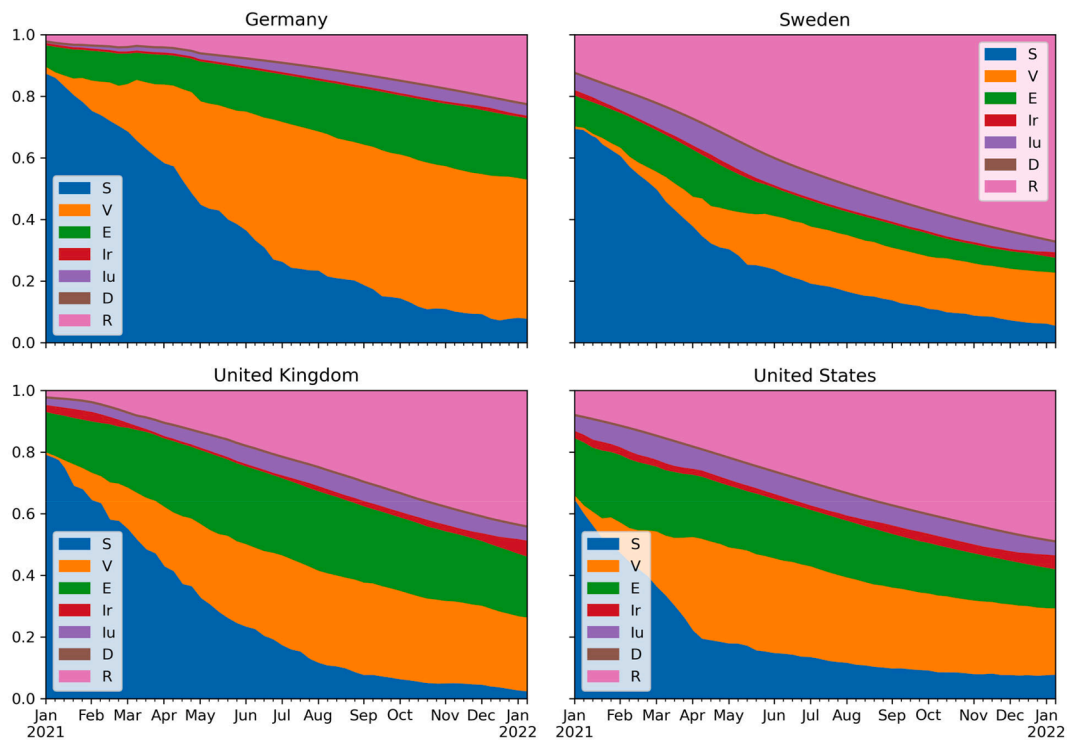


Fig. 4. The evolution of the compartments' sizes for different countries in 2021.

Table 4

Rate of COVID-19 tests performed in the analyzed countries as of April 19, 2022 (per 100 000 population).

Country	Testing rate per 100 000 population	Mean probability of registering an infected person $E[p(t)]$
Germany	145 177	0.179
Sweden	209 032	0.195
UK	747 378	0.240
USA	297 967	0.209

Germany is 2 times lower than in the United States and five times lower than in the United Kingdom. Similarly, the testing rate in Sweden is approximately 1.5 times smaller than in the USA and 3.5 times smaller

than in the UK. It can be assumed, that active testing in the UK and the USA may lead to the detection of the higher number of asymptomatic cases, that can get into the unregistered compartment in Germany or Sweden. However, we can also see an increase in the share of registered compartment in Germany, starting with November 2021. This could have been caused by the rise of performed tests due to the spread of Omicron variant of COVID-19.

Note also that testing intensity is positively correlated with the average probability of detecting and registering infected individuals (Table 4), represented in our model by the parameter  $p(t)$ .

#### 4.3. Registered and unregistered infected individuals

As it can be hard to get the insights about the evolution of certain

compartment from the stacked area chart, Fig. 5 compares the changing of the share of registered and unregistered infected people in analyzed countries. From the data presented, we can see that the maximum proportion of unregistered infected individuals was up to 9.5% of the population in Sweden, 8.4% in the United States, 4.4% in Germany and 6.6% in the United Kingdom in the Spring of 2021. The sum of both compartments  $I_R$  and  $I_U$ , i.e., the total share of infectious individuals in population is 11%, 10.7%, 4.7%, and 8.2% respectively.

In general, data for every country follows similar pattern. Regarding the number of registered cases, we can see the upward trend with several peaks in January, April or October 2021 and the following increase which is probably caused by the spread of Omicron strain. Presumably, the appearance of new potentially dangerous variants of COVID-19 motivates the governments to promote the testing campaign more active which leads to the increase of the share of registered cases in total number of infections.

As for the share of unregistered cases, an upward trend can be observed till summer 2021, which then changes into a downward trend. Furthermore, overall decline of the infected compartment's size may be caused by the high number of vaccinated and recovered individuals. It is also worth mentioning that significant discrepancies between the graphs representing the number of registered and unregistered infected individuals started in January 2021 with the spread of Alpha variant. This suggests the high impact that every new strain of COVID-19 has on the evolution of pandemic.

In turn, Fig. 6 represents the line graphs that shows the changing of the share of exposed individuals in analyzed countries throughout pandemics. According to the graphs, the countries can be divided into two groups. The first group which contains the US and UK is characterized by the ascending trend of the share of exposed individuals which turns into descending trend in summer 2021. It may be caused by the decreasing of the latency period for the new variants of COVID-19 that spread in these countries after June 2021 (e.g., Delta and Omicron). Besides, the vigorous vaccination campaign or the rise of the number of recovered individuals could basically reduce the size of the compartment of people susceptible to the virus. The share of exposed individuals

in the Germany and Sweden generally follows the same evolutionary pattern with the difference that the decline of the exposed compartment's size in Sweden starts in spring 2021 and goes more sharply than for the US and UK. It should also be highlighted that the share of exposed individuals is as well as the number of registered cases generally higher in the USA and UK rising above 0.2 and 0.25 respectively at the peak of pandemics.

Fig. 7 presents the evolution of the exposed individuals' fraction who become registered as COVID-19 positive patients. It should be noted that the value of this ratio is influenced not only by the accuracy of tests that are used to detect COVID-19 but also by the efficacy of the testing campaign that should target the large groups of population susceptible to infection. Therefore, due to the higher number of performed tests (Table 4) the average value of  $p(t)$  is higher for UK and the USA.

We can see that the values mostly fluctuate between 0.2 and 0.4 and 2–3 peaks above these numbers. The first peak in April – July 2020 is probably caused by the high degree of caution in the first months of pandemics. People performed tests when the first signs of the symptoms appeared and that contributed to the high level of infection's detection. Another peak values between October 2021 and January 2022 can be explained by the spread of the Omicron variant that led to the increase in the testing volumes. It should be also highlighted that the values of the ratio for UK hit a peak around January 2021. This considerable increase was probably caused by the appearance and spread of Alpha variant. It is worth reminding that the UK was the first country where this strain emerged, so that healthcare and governmental institutions attempted to detect as many new cases of Alpha as possible to stop or slow down the spread of this variant.

#### 4.4. Infection rate and mortality

Fig. 8 shows the evolution of the infection rate  $\beta(t)$  (i.e., the proportion of contacts between susceptible and infected individuals leading to infection) throughout pandemic for individuals not protected by vaccine. It is necessary to highlight the high noise level in data that makes its interpretation more complicated. Therefore, we decompose

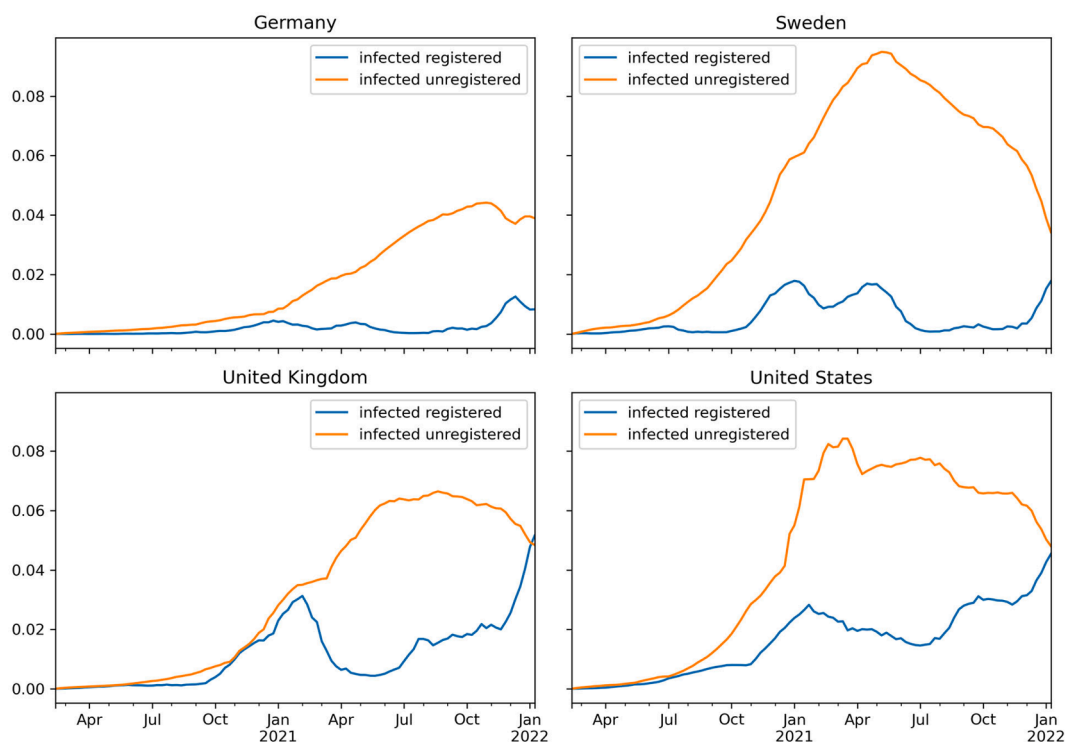


Fig. 5. Share of registered and unregistered infected individuals in total population.

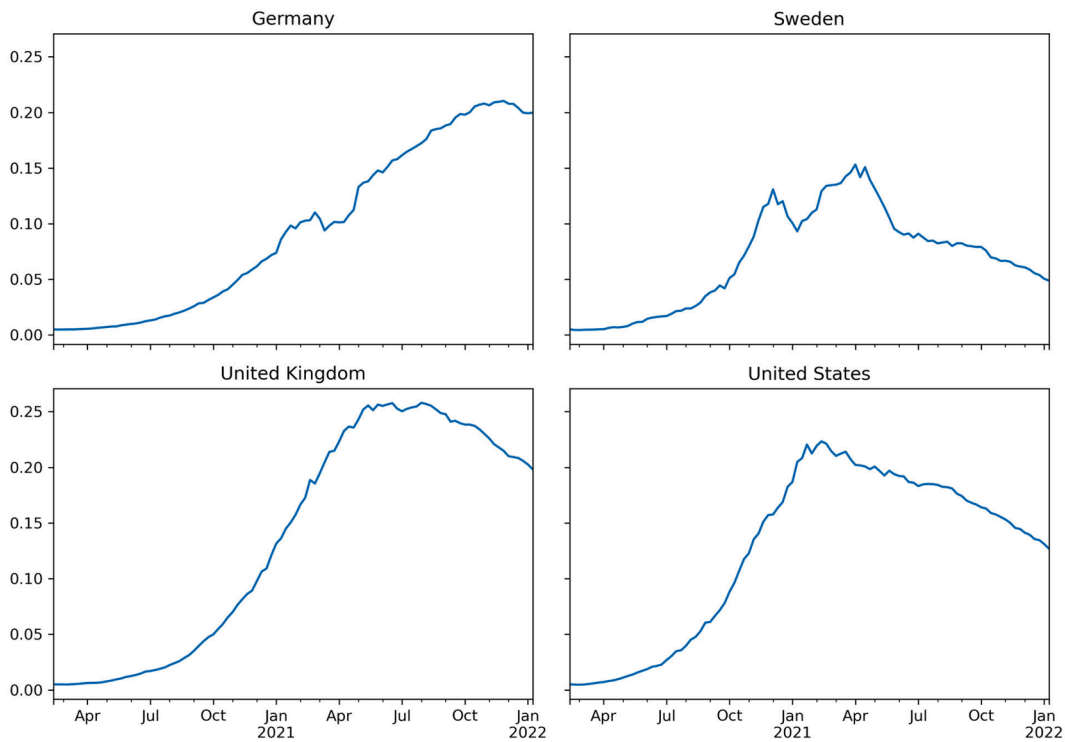


Fig. 6. Share of exposed individuals in total population.

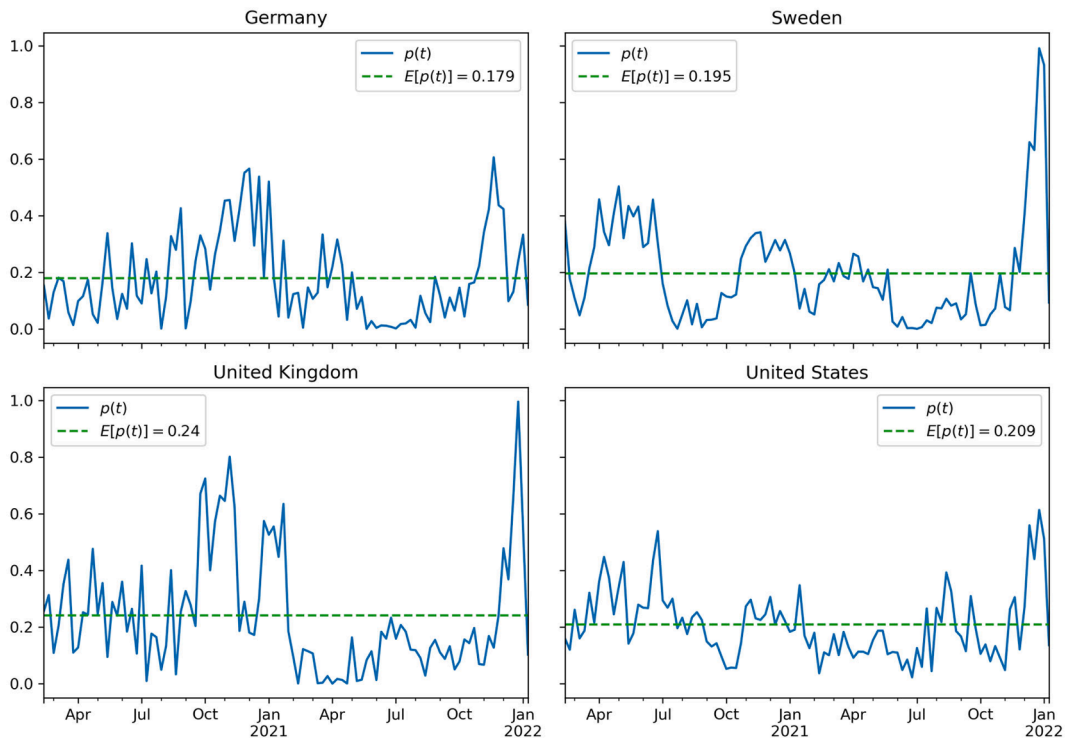


Fig. 7. Exposed individuals' fraction who become registered infectious,  $p(t)$ .

the time series with the help of Singular Spectrum Analysis (SSA). SSA was developed from principal component analysis (Golyandina et al., 2001). The use of the SSA algorithm does not require limiting assumptions about the structure and properties of the time series (for example, stationarity), and display the good performance on short time series. Moreover, Golyandina (2011) emphasized such an advantage of SSA as resistance to noise, and, consequently, the effectiveness of this method

when working with real data, and highly appreciated the possibility of visual interpretation of the components of the analyzed system. Fig. 8 also shows the first component of the SSA decomposition (with the window length of 4), labelled SSA-1, which presents the general trend of the infection rate. All other components of the SSA decomposition oscillate around zero with small amplitude.

Fig. 9 shows the densities of the  $\beta(t)$  distributions. The vertical

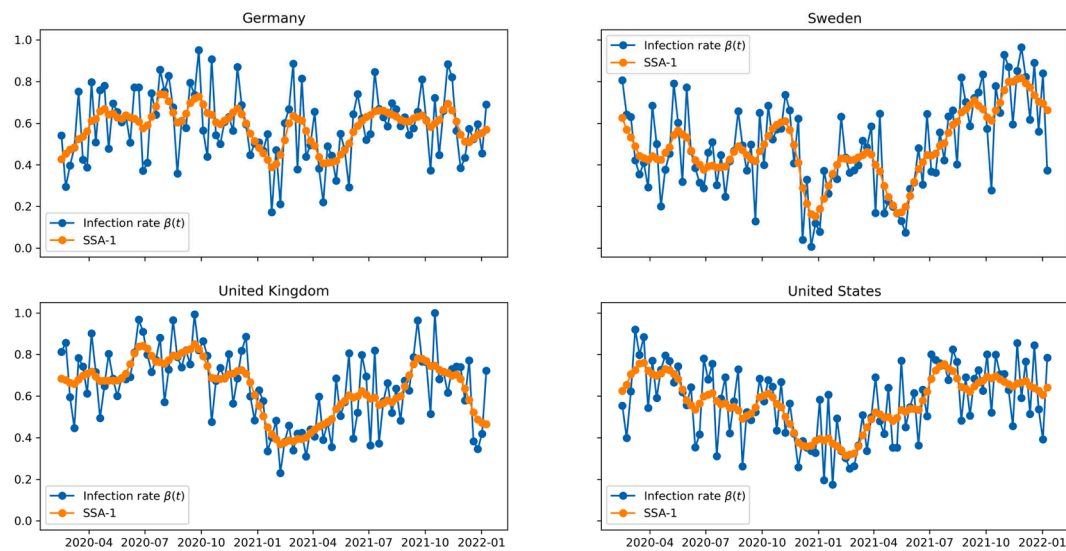


Fig. 8. Infection rate for unvaccinated individuals and its trend presented by first SSA component.

dashed line corresponds to the mean value, the solid blue line represents a smooth density estimate obtained using a Gaussian kernel, the black solid line corresponds to a normal distribution fitted to parameters extracted from the data. As can be seen, the  $\beta(t)$  distribution for Germany and Sweden is very close to normal, as confirmed by the Andersen-Darling and Shapiro-Wilk tests. The actual  $\beta(t)$  distribution for the UK and USA does not follow a normal law. These differences can be explained by a combination of external factors influencing the spread of the disease: effectiveness of vaccination, government policies, etc.

Table 5 presents statistics of  $\beta(t)$  extracted from data including the results from Andersen-Darling and Shapiro-Wilk tests. It also lists the value of Hurst exponent, which is a useful statistical technique for classifying time series according to their long-term memory. The values of the Hurst exponent lie in the range from 0 to 1. The value of 0.5 represents a true random process (Brownian motion or random walk). That means the lack of correlation between the past and future points of the time series. The value above 0.5 means the strong trend and persistence. Such a time series can be characterized as those with positive autocorrelation, so the increase or decrease in values will likely to be complimented by the further increase or decrease in the future. Lastly, the value of Hurst exponent between 0 and 0.5 indicates the negative autocorrelation or the anti-persistence of the time series. As can be seen, the value of the Hurst exponents for all countries corresponds mean-reverting motion.

As shown in Table 5 and Fig. 9, the average infection rate ranges from 0.48 in Sweden to 0.64 in the UK. This means that not all contact between susceptible and infected individuals lead to the spread of infection. However, it reaches a value close to 1 at certain times (especially in the UK in September 2020 and November 2021). What these extreme values are related to (measurement errors, model inaccuracy, external factors) remains to be seen.

Fig. 10 shows the evolution of the infection-induced death rate  $q(t)$ , its mean values and first component of SSA decomposition (trend). The values in Germany in Sweden are generally higher than in the USA and UK. In the USA and UK, there is a sharp increase in pandemic-related deaths in winter 2021, but the number is declining in the second half of the year. In Germany and Sweden, the picture is different - a short spike in winter 2021 is followed by a decline, which is quickly followed by a rise in mortality that continues until the end of the study period. Again, this may be explained by the Alpha variant that emerged in December 2020 in UK and spread quickly in the following months. It is also worth mentioning the existence of other medical, social, and biological factors that may have impact on the death rate and cannot be

estimated in our study.

#### 4.5. Vaccination efficacy

Fig. 11 presents the share of population which is protected by vaccine alongside with the number of people that are fully vaccinated or vaccinated at least with one doze. As it was mentioned above, the variable  $v(t)$  considers the waning of vaccine-induced immunity, therefore it provides more accurate data than the cumulative values of vaccinated individuals which are published in the official reports. The line charts for all four countries follow the same pattern. With the start of vaccination campaign in January 2021 only one doze is already sufficient for gaining the immunity to COVID-19, as the chart for  $v(t)$  lies between the chart for vaccinated and fully vaccinated individuals. However, during the period between April and July 2021, the share of  $v(t)$  compartment stops rising despite the increase in the number of people getting vaccines and even start decreasing from autumn 2021. These results may indicate the higher level of vaccines' efficacy against the initial strains of SARS-CoV-2 and Alpha variant than against Delta and Omicron variants.

Indeed, there are many researchers (e.g., Feikin et al., 2022), proving that existing vaccines provide lower level of protection for Delta and Omicron strains. Moreover, the researchers also claim that one doze of mRNA vaccines (those produced by Moderna and Pfizer-BioNTech) gives a high level of protection against the decease (Bruxvoort et al., 2021; Lopez Bernal et al., 2021).

Additionally, the higher share of  $V$  compartment in Germany in comparison with other three countries should be also highlighted. This feature can be explained by the later start of the active phase of vaccination campaign. While UK and Sweden vaccinated around 50 per cent of the population with at least one doze by April 2021, Germany provided vaccine only for 10 per cent of the individuals. Thus, the increase in the share of  $v(t)$  compartment was achieved mainly by the high level of vaccination which delayed the immunity waning to July 2021. Regarding the USA, low numbers of people protected by vaccine may be caused by the active use of Moderna vaccine which, in fact, was restricted in Germany as well as in Sweden for the alleged side effects (Bruxvoort et al., 2021).

In our model, the effectiveness of the vaccine is determined by the parameter  $\epsilon$ , the value  $1 - \epsilon$  determining how many times less likely the vaccinated person is to be infected by contact with the disease than the susceptible one. The  $\epsilon$  values are 0.689, 0.696, 0.723, and 0.652 for Germany, Sweden, UK, and USA, respectively. Thus, we can conclude that, on average, vaccination reduces the risk of infection by about

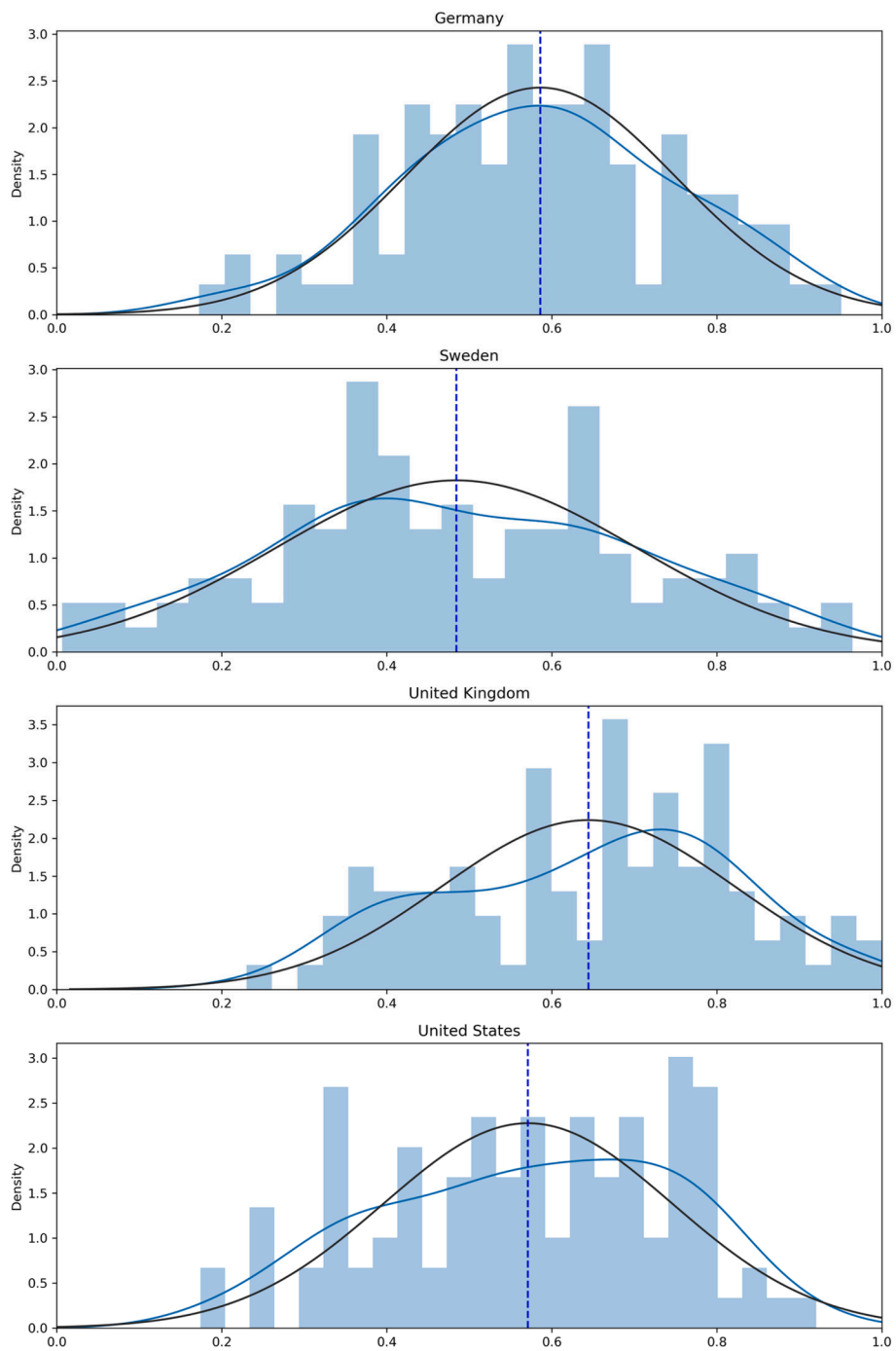


Fig. 9. Distribution of the infection rate  $\beta$ .

Table 5  
Parameters of infection rate distribution.

Country	Mean	Std. dev.	Skewness	Kurtosis	Normality test	Hurst exponent
Germany	0.59	0.16	-0.07	-0.35	yes	0.05
Sweden	0.48	0.22	0.04	-0.64	yes	0.14
UK	0.64	0.18	-0.18	-0.78	no	0.19
USA	0.57	0.18	-0.23	-0.87	no	0.18

65–70%.

#### 4.6. Average values of pandemic parameters

Table 6 shows the average values of the parameters that are commonly used for analyzing epidemics. The difference in the average

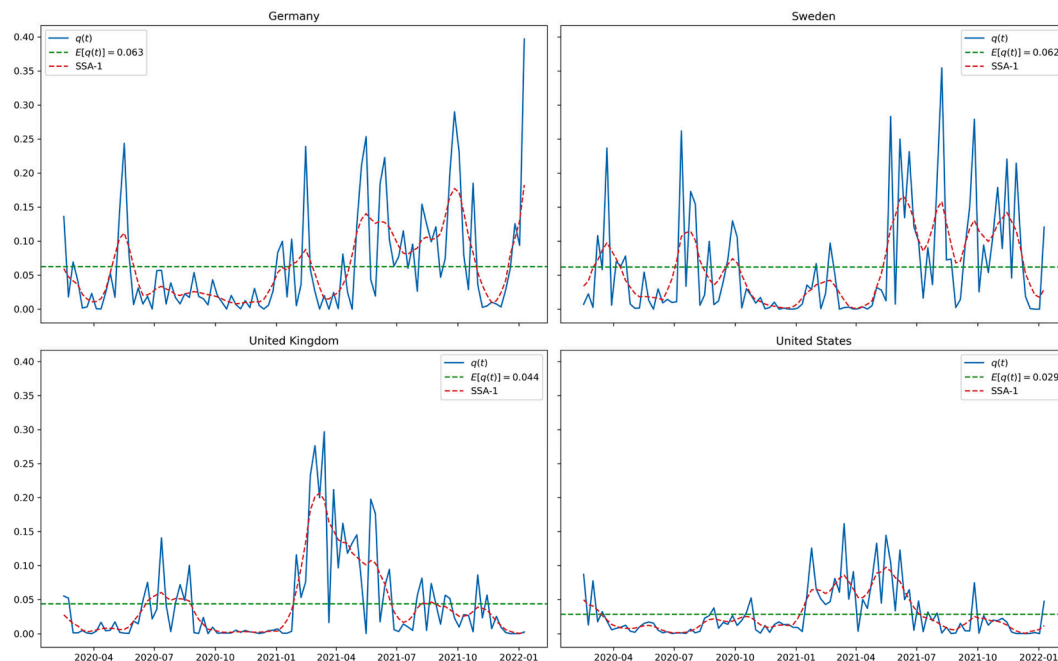


Fig. 10. Infection-induced death rate  $q(t)$ , its average value and first component of SSA decomposition throughout the pandemic.

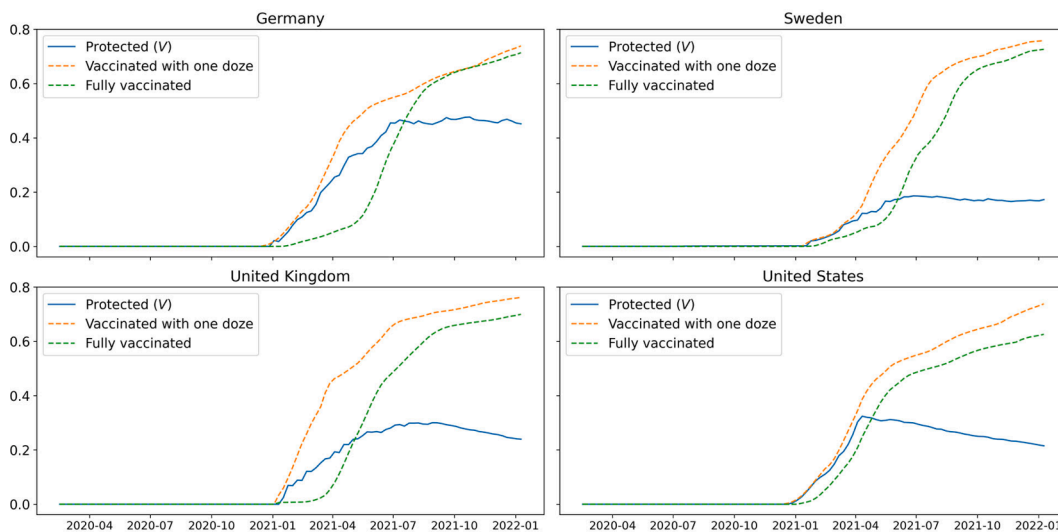


Fig. 11. Share of individuals protected by vaccine vs the share of vaccinated and fully vaccinated individuals.

Table 6

Average values of pandemic parameters (days).

Parameter	Value	Country			
		Germany	Sweden	UK	USA
Average time to lose vaccine-induced immunity	$1/\lambda$	154	151	188	127
Average time to lose natural immunity	$1/\omega$	256	263	233	238
Average time which registered individuals leave being infectious	$1/\gamma_R$	6.26	5.67	6.33	6.43
Average time which unregistered individuals leave being infectious	$1/\gamma_U$	5.54	5.31	5.88	5.81
Average latency period	$1/\sigma$	5.33	5.29	5.66	5.44

times between the analyzed countries can be explained by the variety of treatment protocols used by their healthcare institutions and by other factors that are not considered in the implemented model.

The average time to lose vaccine-induced immunity for all the analyzed countries slightly exceeds five months, although vaccine manufacturers guarantee immunity against COVID-19 for at least six months after the full vaccination (Zhang et al., 2021). However, these times were evaluated on the initial variant of SARS-CoV-2 while the new strains such as Alpha, Delta and Omicron can overcome the protection provided by the existing vaccines. Therefore, since the summer 2021 the healthcare systems of the world countries started promoting booster shots that ensure the immunity against the new variants of the virus. The period that should pass between the primary and booster vaccination varies depending on the current epidemiological situation and can be also shorter for the most vulnerable parts of the population. For example, CDC recommends the booster shot four months after the prior vaccination to people over the age of 65 and to those who are over 50

with certain medical conditions increasing the chance of severe illness. Nevertheless, all the vaccines maintain their high efficiency against severe disease and hospitalization. Therefore, it is required to calculate these values to be able to assess the performance of the vaccines, but this task is not the part of our study.

According to the values of the duration of natural and vaccine-induced immunity, recovering from the infection provides longer and sustainable immunity lasting around 8 months. This is inconsistent with Yu et al. (2022) who suggest that vaccination induces higher immunity level than prior infection.

Regarding the average time to lose vaccine-induced immunity, similar values for Germany in Sweden should be highlighted. Both countries used Pfizer-BioNTech vaccine as the main one in their vaccination campaign, so that this ensured almost identical waning of the immunity. The low value for the USA can be again explained by the active use of Moderna vaccine. Additionally, it is worth mentioning the longest duration of the vaccine-induced immunity for the UK. This may relate to the usage of viral vector AstraZeneca vaccine in the immunization campaign.

It is also important to analyze and compare the average time which registered and unregistered individuals leave being infectious. As it can be seen in the Table 6, this time is higher for registered cases. This is because the largest part of registered cases includes symptomatic and severe diseases that require more time for treatment and recovery. The lowest values of these parameters for Sweden should also be noted. Nevertheless, we cannot thoroughly estimate the values without considering medical protocols approved in the analyzed countries. Regarding the latency period, it varies between 5.29 and 5.66 days which is aligned with the estimations performed by other researchers (Xin et al., 2021; Lauer et al., 2020).

It should also be stated that the parameters shown in Table 6 were not supposed to be time dependent in the proposed model (Table 2). However, in real life their values may change mainly because of the spread of different new strains of COVID-19. There are many studies showing that the duration of immunity against infection as well as latency and infectious period vary for different variants of SARS-CoV-2. For example, Rella et al. (2021) concluded that vaccine-resistant strains can significantly reduce the duration of immunity as well as decrease the efficacy of the vaccines. The same is applicable for the natural immunity which has a higher waning rate for the vaccine-resistant strains. Therefore, the stationarity of the above-mentioned parameters that was introduced for simplification purposes can be considered as a limitation of the implemented model.

## 5. Conclusion

The proposed approach, based on reconstruction of real distributions of transition rates from data using genetic algorithms, allows fitting a model describing several pandemic waves. The model was fitted with MAPEs of 2.168, 2.096, 1.208 and 1.703 for Germany, Sweden, the UK, and the US respectively. The results are consistent with general empirical statistics from medical studies (e.g., incubation period, latency period, etc.), confirming the quality of the model. Using the model, we investigated some features of the pandemic in four countries.

The main results, which open the way for further research, are as follows:

- The mean infection rate  $\beta(t)$  i.e., the proportion of contacts between susceptible and infected individuals leading to infection is 0.48, 0.57, 0.58 and 0.64 for Sweden, the USA, Germany, and the UK respectively. However, at certain times it will reach a value close to 1 (e.g., in the UK in September 2020 and November 2021). A possible explanation is that these extremes are associated with new, more infectious strains of the virus, but more research is needed on the causal relationships of the resulting time series  $\beta(t)$  with a broader list of external factors (government policies, data quality, etc.).

- The maximum proportion of exposed individuals in the entire population ranges from 0.14 in Sweden to 0.25 in the UK. This allows us to estimate the spread of the pandemic. It is worth noting that in all countries except Germany, there is a decreasing trend in the number of exposed persons from the first quarter of 2021.
- The proportion of registered infections among exposed individuals generally ranges between 0.2 and 0.4, with several peaks above these figures. Potentially interesting to find out how this relates to testing and other non-pharmacological interventions.
- According to our model, the average duration of vaccine induced immunity is shorter than the manufacturers claim. Before the appearance of Delta variant of SARS-CoV-2, even one shot of the vaccine provided the required protection against the virus, while with the appearance of new variants of the virus the vaccine-induced immunity became less effective. However, on average, vaccination reduces the risk of infection by about 65–70%.
- There is variation in the duration of vaccine-induced immunity, which may be due to different vaccines being used in countries. The UK displays the longest duration of vaccine-induced immunity and the highest reduction in the probability of infection of the vaccinated through contact with the disease ( $\epsilon = 0.723$ ). However, group V (individuals protected by vaccine) reaches the largest size in Germany.
- The average time of loss of natural immunity is higher in Germany and Sweden than in the UK and USA. To find out why this is the case, a comparative study of the policies of pandemic-handling in these countries is needed.

The advantages of the proposed approach, as discussed above, are the ability to model multiple pandemic waves and hence to estimate unobserved subpopulations more accurately. However, the focus on reconstructing the time series representing the transition rate from observed data limits the predictive ability of the model. In the case of constant transition rates, predictions can be obtained by simply integrating the corresponding ODEs. To make predictions based on our approach, a predictive model for multivariate time series  $\beta(t), p(t), q(t), v(t)$  is additionally needed. This extension is the goal of future research.

There is another scope for further improvement. The data on Omicron variant can be taken for the analysis and the results can be compared to those for other strains of the virus. The data for other countries can also be considered for the analysis to provide a more holistic picture.

Note also that the calculation of real distributions of transition rates opens new possibilities for analyzing pandemic handling policies. Typically, researchers in this field consider population distributions across compartments as endogenous variables. Consideration of how different interventions affect transition rates can provide additional insights.

## CRedit authorship contribution statement

**Yuri Zelenkov:** Conceptualization, Methodology, Validation, Writing – review & editing. **Ivan Reshetsov:** Methodology, Software, Validation, Writing – original draft.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.



## References

- Acuña-Zegarra, M. A., Díaz-Infante, S., Baca-Carrasco, D., & Olmos-Liceaga, D. (2021). COVID-19 optimal vaccination policies: A modeling study on efficacy, natural and vaccine-induced immunity responses. *Mathematical Biosciences*, 337, Article 108614. <https://doi.org/10.1016/j.mbs.2021.108614>
- Arik, S., et al. (2020). Interpretable sequence learning for COVID-19 forecasting. *Advances in Neural Information Processing Systems*, 33, 18807–18818.
- Assefa, Y., Gilks, C. F., Reid, S., et al. (2022). Analysis of the COVID-19 pandemic: Lessons towards a more effective response to public health emergencies. *Globalization and Health*, 18, 10. <https://doi.org/10.1186/s12992-022-00805-9>
- Baguélin, M., Flasche, S., Camacho, A., Demiris, N., Miller, E., & Edmunds, W. J. (2013). Assessing optimal target populations for influenza vaccination programmes: An evidence synthesis and modelling study. *PLoS Medicine*, 10(10), e1001527.
- Balashov, V. S., Yan, Y., & Zhu, X. (2021). Using the Newcomb-Benford law to study the association between a country's COVID-19 reporting accuracy and its development. *Scientific Reports*, 11, 22914. <https://doi.org/10.1038/s41598-021-02367-z>
- Baraniuk, C. (2021). How long does covid-19 immunity last? *The BMJ*, 373, Article n1605. <https://doi.org/10.1136/bmj.n1605>
- Bjørnstad, O. N., Shea, K., Krzywinski, M., & Altman, N. (2020). The SEIRS model for infectious disease dynamics. *Nature Methods*, 17, 557–558. <https://doi.org/10.1038/s41592-020-0856-2>
- Boissay, F., Rees, D., & Rungcharoenkitkul, R. (2020). Dealing with COVID-19: understanding the policy choices, *BIS Bulletins* 19, Bank for International Settlements.
- Boroujeni, S. P. H., & Pashaei, E. (2021). Data clustering using chimp optimization algorithm. In 11th *IEEE International Conference on Computer Engineering and Knowledge (ICCKE)*, pp. 296–301. 10.1109/ICCKE54056.2021.9721483.
- Bouchnita, A., Chekroun, A., & Jebrane, A. (2021). Mathematical modeling predicts that strict social distancing measures would be needed to shorten the duration of waves of COVID-19 infections in Vietnam. *Frontiers in Public Health*, 12(8), Article 559693. <https://doi.org/10.3389/fpubh.2020.559693>
- Brewer, D., Barenco, M., Callard, R., Hubank, M., & Stark, J. (2008). Fitting ordinary differential equations to short time course data. *Philosophical Transactions of the Royal Society A*, 366, 519–544. <https://doi.org/10.1098/rsta.2007.2108>
- Bruxvoort, K. J. et al. (2021) Effectiveness of mRNA-1273 against Delta, Mu, and other emerging variants of SARS-CoV-2: test negative case-control study. *BMJ*, 375: e068848. doi:10.1136/bmj-2021-068848.
- Capistran, M. A., Capella, A., & Christen, J. (2021). Forecasting hospital demand in metropolitan areas during the current COVID-19 pandemic and estimates of lockdown-induced 2nd waves. *PLoS One*, 16(1), e0245669.
- Chang, S., Pierson, E., Koh, P. W., et al. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589, 82–87. <https://doi.org/10.1038/s41586-020-2923-3>
- Chatzilena, A., Van Leeuwen, E., Ratmann, O., Baguelin, M., & Demiris, N. (2019). Contemporary statistical inference for infectious disease models using Stan. *Epidemics*, 29, Article 100367. <https://doi.org/10.1016/j.epidem.2019.100367>
- Cheng, C., et al. (2021). The incubation period of COVID-19: A global meta-analysis of 53 studies and a Chinese observation study of 11 545 patients. *Infectious Diseases Poverty*, 10, 119. <https://doi.org/10.1186/s40249-021-00901-9>
- Comunian, A., Gaburro, R., & Giudici, M. (2020). Inversion of a SIR-based model: A critical analysis about the application to COVID-19 epidemic. *Physica D: Nonlinear Phenomena*, 413, Article 132674. <https://doi.org/10.1016/j.physd.2020.132674>
- CRS (2022). Coronavirus Resource Center. Johns Hopkins University. Available at: <https://coronavirus.jhu.edu> Accessed: June 20, 2022.
- Davis, H. et al. (2020). Characterizing Long COVID in an International Cohort: 7 Months of Symptoms and Their Impact. *medRxiv*. 10.1101/2020.12.24.20248802.
- Eryarsoy, E., Delen, D., Davazdahemami, B., & Topuz, K. (2021). A novel diffusion-based model for estimating cases, and fatalities in epidemics: The case of COVID-19. *Journal of Business Research*, 124, 163–178. <https://doi.org/10.1016/j.jbusres.2020.11.054>
- Feikin, D. R., et al. (2022). Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: Results of a systematic review and meta-regression. *The Lancet*, 399(10328), 924–944. [https://doi.org/10.1016/S0140-6736\(22\)00152-0](https://doi.org/10.1016/S0140-6736(22)00152-0)
- Feng, S., Feng, Z., Ling, C., Chang, C., & Feng, Z. (2021). Prediction of the COVID-19 epidemic trends based on SEIR and AI models. *PLoS One*, 16(1), e0245101.
- Golyandina, N., Nekrutkin, V., & Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*. New York - London: Chapman & Hall/CRC.
- Golyandina, N. (2011). *On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods*. New York - London: Chapman & Hall/CRC.
- GRT (2022). COVID-19 Government Response Tracker. University of Oxford. Available at: <https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker>. Accessed: June, 20, 2022.
- Hasell, J., et al. (2020). A cross-country database of COVID-19 testing. *Scientific Data*, 7, 345. <https://doi.org/10.1038/s41597-020-00688-8>
- Huang, H., Handel, A., & Song, X. (2020). A Bayesian approach to estimate parameters of ordinary differential equation. *Computational Statistics*, 35, 1481–1499. <https://doi.org/10.1007/s00180-020-00962-8>
- Ivorra, B., Ferrández, M. R., Vela-Pérez, M., & Ramos, A. M. (2020). Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Communications in Nonlinear Science & Numerical Simulation*, 88, Article 105303. <https://doi.org/10.1016/j.cnsns.2020.105303>
- Jentsch, P. C., Anand, M., & Bauch, C. T. (2021). Prioritising COVID-19 vaccination in changing social and epidemiological landscapes: A mathematical modelling study. *The Lancet Infectious Diseases*, 21(8), 1097–1106. [https://doi.org/10.1016/S1473-3099\(21\)00057-8](https://doi.org/10.1016/S1473-3099(21)00057-8)
- Karin, O. et al. (2020). Adaptive cyclic exit strategies from lockdown to suppress COVID-19 and allow economic activity. *medRxiv*. 10.1101/2020.04.04.20053579.
- Katara, S., Bhan, A., Caruthers, J. M., Delgass, W. N., & Venkatasubramanian, V. (2004). A hybrid genetic algorithm for efficient parameter estimation of large kinetic models. *Computers & Chemical Engineering*, 28(12), 2569–2581. <https://doi.org/10.1016/j.compchemeng.2004.07.002>
- Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of Royal Society of London*, A115700–A115721.
- Kermack, W. O. & McKendrick, A. G. (1933). Contributions to the mathematical theory of epidemics, iii - further studies of the problem of endemicity. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematics*. 141 94–122.
- Knight, S. R., et al. (2020). Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: Development and validation of the 4C mortality score. *BMJ*, 370, Article m3339. <https://doi.org/10.1136/bmj.m3339>
- Lauer, S. A., et al. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172(9), 577–582. <https://doi.org/10.7326/M20-0504>
- Liu, Z., Magal, P., Seydi, O., & Webb, G. (2020). A COVID-19 epidemic model with latency period. *Infectious Disease Modelling*, 5, 323–337. <https://doi.org/10.1016/j.idm.2020.03.003>
- Lopez Bernal, J., et al. (2021). Effectiveness of Covid-19 vaccines against the B.1.617.2 (Delta) variant. *The New England Journal of Medicine*, 385(7), 585–594. <https://doi.org/10.1056/NEJMoa2108891>
- Miikkulainen, R., Francon, O., Meyerson, E., Qiu, X., Sargent, D., Canzani, E., et al. (2021). From prediction to prescription: Evolutionary optimization of nonpharmaceutical interventions in the COVID-19 pandemic. *IEEE Transactions on Evolutionary Computation*, 25(2), 386–401. <https://doi.org/10.1109/TEVC.2021.3063217>
- Milne, G., et al. (2021). Does infection with or vaccination against SARS-CoV-2 lead to lasting immunity? *The Lancet Respiratory Medicine*, 9(12), 1450–1466. [https://doi.org/10.1016/S2213-2600\(21\)00407-0](https://doi.org/10.1016/S2213-2600(21)00407-0)
- Otunuga, O. (2021). Estimation of epidemiological parameters for COVID-19 cases using a stochastic SEIRS epidemic model with vital dynamics. *Results in Physics*, 28, Article 104664. <https://doi.org/10.1016/j.rinp.2021.104664>
- Rawson, T., Brewer, T., Veltcheva, D., Huntingford, C., & Bonsall, M. B. (2020). How and when to end the COVID-19 lockdown: An optimization approach. *Frontiers in Public Health*, 8, 262. <https://doi.org/10.3389/fpubh.2020.00262>
- Rella, S. A., Kulikova, Y. A., Dermitzakis, E. T., & Kondrachov, F. A. (2021). Rates of SARS-CoV-2 transmission and vaccination impact the fate of vaccine-resistant strains. *Scientific Reports*, 11, 15729. <https://doi.org/10.1038/s41598-021-95025-3>
- Rippering, C., et al. (2021). Evaluation of undetected cases during the COVID-19 epidemic in Austria. *BMC Infectious Diseases*, 21, 70. <https://doi.org/10.1186/s12879-020-05737-6>
- Schlickeiser, R., & Kröger, M. (2021). Analytical modeling of the temporal evolution of epidemics outbreaks accounting for vaccinations. *Physics*, 3, 386–426. <https://doi.org/10.3390/physics3020028>
- Shah, M., & Woo, H. G. (2022). Omicron: A heavily mutated SARS-CoV-2 variant exhibits stronger binding to ACE2 and potentially escapes approved COVID-19 therapeutic antibodies. *Frontiers in Immunology*, 12. <https://doi.org/10.3389/fimmu.2021.830527>
- Shah, N. A., Moffitt, R. A., & Wang, M. D. (2007). Modified genetic algorithm for parameter selection of compartmental models. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 143–146). <https://doi.org/10.1109/IEMBS.2007.4352243>
- Silva, L., & Figueiredo Filho, D. (2021). Using Benford's law to assess the quality of COVID-19 register data in Brazil. *Journal of Public Health*, 43(1), 107–110. <https://doi.org/10.1093/pubmed/fdaa193>
- Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66(8), 1352–1362. <https://doi.org/10.1057/jors.2014.103>
- Vega, R., Flores, L., & Greiner, R. (2022). SIMLR: Machine learning inside the SIR Model for COVID-19 forecasting. *Forecasting*, 4(1), 72–94. <https://doi.org/10.3390/forecast4010005>
- Wibbens, P., Koo, W., & McGahan, A. (2020). Which COVID policies are most effective? A Bayesian analysis of COVID-19 by jurisdiction. *PLoS One*, 15(12), e0244177.
- Xin, H., et al. (2021). Estimating the latent period of coronavirus disease 2019 (COVID-19). *Clinical Infectious Diseases*, 74(9), 1678–1681. <https://doi.org/10.1093/cid/ciab746>
- Yarsky, P. (2021). Using a genetic algorithm to fit parameters of a COVID-19 SEIR model for US states. *Mathematics and Computers in Simulation*, 185, 687–695. <https://doi.org/10.1016/j.matcom.2021.01.022>
- Yu, Y., et al. (2022). mRNA vaccine-induced antibodies more effective than natural immunity in neutralizing SARS-CoV-2 and its high affinity variants. *Scientific Reports*, 12, 2628. <https://doi.org/10.1038/s41598-022-06629-2>
- Zhang, Y., Ndzouboukou, J. L. B., Gan, M., Lin, X., & Fan, X. (2021). Immune evasive effects of SARS-CoV-2 variants to COVID-19 emergency used vaccines. *Frontiers in Immunology*, 12, Article 771242. <https://doi.org/10.3389/fimmu.2021.771242>