

Original research article



Vaccination compartmental epidemiological models for the delta and omicron SARS-CoV-2 variants

J. Cuevas-Maraver ^{a,b,*}, P.G. Kevrekidis ^c, Q.Y. Chen ^c, G.A. Kevrekidis ^{d,e,f}, Y. Drossinos ^g

^a Grupo de Física No Lineal, Departamento de Física Aplicada I, Universidad de Sevilla. Escuela Politécnica Superior, C/ Virgen de África, 7, 41011 Sevilla, Spain

^b Instituto de Matemáticas de la Universidad de Sevilla (IMUS), Edificio Celestino Mutis. Avda. Reina Mercedes s/n, 41012 Sevilla, Spain

^c Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01003, USA

^d Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218, USA

^e Los Alamos National Laboratory, Los Alamos, NM, USA

^f Mathematical Institute for Data Science, Johns Hopkins University, Baltimore MD, USA

^g Thermal Hydraulics & Multiphase Flow Laboratory, Institute of Nuclear & Radiological Sciences and Technology, Energy & Safety, N.C.S.R. "Demokritos", GR 15341, Agia Paraskevi, Greece

ARTICLE INFO

Keywords:

Epidemiological model

SARS-CoV-2

Vaccination

Delta variant

Omicron variant

ABSTRACT

We explore the inclusion of vaccination in compartmental epidemiological models concerning the delta and omicron variants of the SARS-CoV-2 virus that caused the COVID-19 pandemic. We expand on our earlier compartmental-model work by incorporating vaccinated populations. We present two classes of models that differ depending on the immunological properties of the variant. The first one is for the delta variant, where we do not follow the dynamics of the vaccinated individuals since infections of vaccinated individuals were rare. The second one for the far more contagious omicron variant incorporates the evolution of the infections within the vaccinated cohort. We explore comparisons with available data involving two possible classes of counts, fatalities and hospitalizations. We present our results for two regions, Andalusia and Switzerland (including the Principality of Liechtenstein), where the necessary data are available. In the majority of the considered cases, the models are found to yield good agreement with the data and have a reasonable predictive capability beyond their training window, rendering them potentially useful tools for the interpretation of the COVID-19 and further pandemic waves, and for the design of intervention strategies during these waves.

1. Introduction

Over the last three and a half years, the COVID-19 pandemic has been deemed responsible for 770 million confirmed cases, and nearly 7 million deaths worldwide, as of this writing and according to the World Health Organization COVID-19 dashboard. As such, its emergence wreaked havoc in life as we knew it throughout the world and forced a dramatic modification of our social and economic activities during this interval. At the same time, it triggered a global mobilization of the scientific community to produce vaccines rapidly, especially through (thankfully, by that time, fairly mature) technology of mRNA-based methods. This effort led to the remarkable result of having a vaccine against SARS-CoV-2 within a year of its emergence. Nevertheless, this was far from the end of the story, as new variants of the SARS-CoV-2 virus kept emerging within 2020 and 2021. The so-called delta variant appeared in India in late 2020 and it had spread to 179 countries by November 2021. Subsequently, the delta variant was

superseded by the so-called omicron variant that was reported in South Africa on November 2021, which rapidly became the predominant variant of SARS-CoV-2 thereafter.

The theoretical and mathematical modeling of infectious diseases such as COVID-19 has a long and time-honored history since the classic work of Kermack and McKendrick [1]. Relevant efforts have been summarized in numerous venues in recent years, such as, e.g., [2–4], to mention only a few. The urgency and severity of the COVID-19 pandemic brought about an intense effort on the side of the mathematical and physical communities to develop analytical models and computational tools that could be used to examine the unprecedented volume of available data regarding the temporal (and spatial) evolution of the pandemic and to make predictions for the weeks (or in some cases month(s)) ahead. A notable example of comparison of such efforts can be seen in, e.g., websites such as [5]. Relevant modeling efforts have now been summarized in a number of reviews such as [6,7],

* Corresponding author at: Grupo de Física No Lineal, Departamento de Física Aplicada I, Universidad de Sevilla. Escuela Politécnica Superior, C/ Virgen de África, 7, 41011 Sevilla, Spain.

E-mail address: jcuevas@us.es (J. Cuevas-Maraver).

<https://doi.org/10.1016/j.mbs.2023.109109>

Received 2 May 2023; Received in revised form 14 October 2023; Accepted 14 November 2023

Available online 18 November 2023

0025-5564/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

including ones of specialized modeling aspects such as the study of metapopulation network models [8], while other works summarized the challenges and difficulties of associated modeling [9,10].

Over the last two years, a large portion of the focus of the modeling efforts has shifted towards the inclusion of vaccination in epidemiological models. Alternative, and complementary, modeling efforts include game-theoretical approaches to evaluate decision strategies of players involved in vaccine and drug development. The players usually considered include individuals, drug companies that develop vaccines and drugs, their sponsors, and nature's response to disease contraction (with or without vaccination), recovery or death [11,12]. While a lot of information is available regarding the effectiveness and efficacy of vaccines [13] (see also websites such as [14]) mathematical models can still be quite useful in a number of ways, including in guiding and informing distribution strategies thereof [15,16]. It is in that light that numerous compartmental epidemiological models with vaccination strategies have arisen in the literature [17,18], including some specific to different geographical locations [19] and to different social infrastructures, such as nursing homes [20]. While the relevant models feature different levels of complexity starting from SIRV (Susceptible–Infected–Recovered–Vaccinated) extensions of the classic SIR (Susceptible–Infected–Recovered) [17] model and progressively extending to multicomponent models such as [21], our aim here is to build systematically on the earlier modeling attempt of [22] by considering SARS-CoV-2 variants that differently affect the vaccinated population. While the previously discussed models, and the ones developed herein, are at the population level, within-host dynamics have also been extensively studied, as in the modeling of influenza A dynamics and its interaction with the immune system [23].

More concretely, our aim is to present a model of the omicron variant (model A and its two implementations A1 and A2), in which its highly contagious nature allows for so-called breakthrough infections, whereby vaccinated individuals may still be infected. In that light, we account for the standard populations of our earlier work [22], including exposed, presymptomatics, and asymptomatics [24], as well as hospitalizations, recoveries and fatalities. It is worth noting here that the relevance of including presymptomatic populations [25] has been argued in various models [26] and reviews [27] on the basis of their different viral load and viral shedding in comparison to asymptomatics. The latter individuals do not have a “time zero” (time of symptom appearance), while presymptomatics do, leading in that way to a change of their contacts and social norms. In addition, we consider such populations in *both* the unvaccinated and vaccinated portions of the population and their interactions. The primary aim of the associated study is to explore the dynamical evolution of the omicron variant from the end of 2021 to early spring 2022. We also present a simpler model (model B) for the evolution of the earlier delta variant during the fall of 2021. In that case, vaccination was deemed to protect individuals from being infected, and the fraction of breakthrough infections was quite small, even in groups such as the potentially highly exposed group of healthcare workers [28]. Accordingly, we assume that the vaccinated population may be effectively removed from the susceptible compartment. More generally, our intention in the present work is to offer an approach which is tuned to both the concrete features of different variants (their timing and infectivity, the efficacy of vaccines, the relevance of breakthrough infections, etc.) and the availability of concrete datasets that can be used to assess the corresponding models. We hope that this can help further consideration of such settings, including in the undesirable, but not unlikely possibility of future variants.

We examine two versions of the proposed omicron model in Section 2 (models A1 and A2), their difference motivated by data availability. The population flows in Model A1 terminate at the fatalities compartment: as such, the model considers that fatalities in both the unvaccinated and vaccinated populations provide the most reliable data. Model A2 is motivated by the existence of systematic data for the

total number of hospitalizations (conventional and in Intensive Care Units (ICU)): here, population flows terminate at the hospitalizations compartment, i.e., they do not branch further to the fatalities compartment as in model A1. This is for a number of reasons: *in primis*, reporting of fatalities occasionally occurs retroactively (and less reliably). Our models are applied to two regions with similar populations (approximately 8 million inhabitants): Andalusia and Switzerland (including the Principality of Liechtenstein). The motivation for this choice arises, once again, from the availability of suitably stratified data, whereby both fatalities and hospitalizations are available for vaccinated and unvaccinated individuals. In Section 3 we propose a model for the delta variant, model B, in addition to the more detailed model for the omicron variant presented in Section 2. We use model B in the same spatial regions. We typically find that numerical results compare favorably to available data, both in terms of the comparison of the regression results and also in connection to testing beyond the end of the training period for the model parameters. Finally, in Section 4, we summarize our findings and present our conclusions. In the Appendix, we consider the question of structural identifiability of the models developed herein from a mathematical perspective.

2. Omicron variant

2.1. Model A1: Branches terminate at fatalities

The first model for the omicron variant, model A1, extends our compartmental epidemiological model used to examine the COVID-19 pandemic evolution in Mexico [22]. Accordingly, the susceptible population S can turn to exposed (E) through interactions with either symptomatically infected (J), presymptomatic (P), or asymptomatic (A) individuals. The exposed population E , in turn, can convert to either P , within a time scale $1/\sigma_1$, leading to different clinical stages of the disease or to A , a compartment that has been recognized to play a key role in the dynamical evolution of COVID-19 [24]. The asymptomatic population A can only lead to undisclosed recoveries (denoted as U), over a time scale $1/\mu$. On the other hand, the presymptomatic individuals P turn to infected with clinical symptoms I over a time scale $1/\sigma_2$. The addition of the latent period $1/\sigma_1$ and the preclinical period $1/\sigma_2$ constitute the incubation time scale of the disease, $\tau_{\text{inc}} = \sigma_1^{-1} + \sigma_2^{-1}$. Subsequently, the symptomatically infected can either turn to hospitalized H at a rate γ_h , while the rest may recover (R) at a rate γ_r . Finally, those in the hospitalized population of H can, again, branch into two populations: they either recover at a rate κ_r , or they lead to fatalities (D) at a rate κ_d .

While these populations were also present in our earlier work [22], it is relevant to highlight the differences of the omicron-variant modeling. For the period under consideration (fall 2021 to spring 2022), vaccines had been deployed extensively in Andalusia and Switzerland. More importantly, breakthrough infections due to the omicron variant were substantial within the vaccinated population (contrary to the case of the delta variant considered in Section 3). In light of that, we formulated *two* sets of populations: one representing unvaccinated individuals, denoted by (the subscript) u , and the other representing the substantial population of vaccinated individuals, denoted by (the subscript) v . Each population subset had its own set of parameters.

$$\frac{dS_u}{dt} = -\beta^{uu} S_u (I_u + A_u + P_u) - \beta^{uv} S_u (I_v + A_v + P_v) - \theta V(t), \quad (1a)$$

$$\frac{dE_u}{dt} = -\sigma_1 E_u + \beta^{uu} S_u (I_u + A_u + P_u) + \beta^{uv} S_u (I_v + A_v + P_v), \quad (1b)$$

$$\frac{dP_u}{dt} = (1 - \phi_u) \sigma_1 E_u - \sigma_2 P_u, \quad (1c)$$

$$\frac{dA_u}{dt} = \phi_u \sigma_1 E_u - \mu_u A_u, \quad (1d)$$

$$\frac{dU_u}{dt} = \mu_u A_u, \quad (1e)$$

$$\frac{dI_u}{dt} = \sigma_2 P_u - (\gamma_{r,u} + \gamma_{h,u}) I_u, \quad (1f)$$

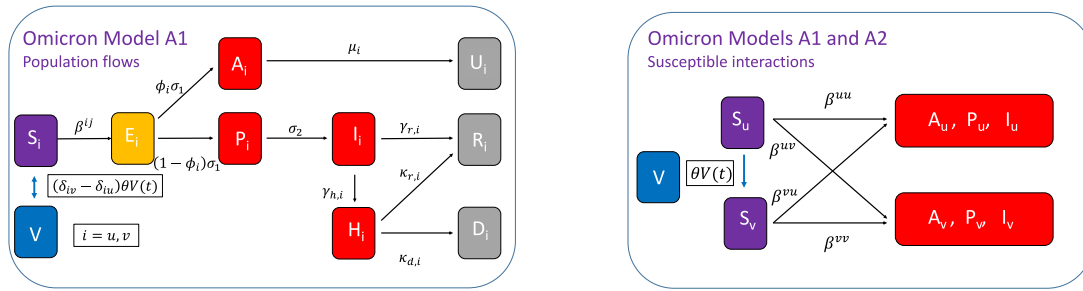


Fig. 1. Schematic diagram of population flows according to model A1 (left panel) and susceptible interactions with other population compartments, for both models A1 and A2 (right panel). The symbol δ_{ij} with $i, j = u, v$ is the Kronecker delta.

$$\frac{dH_u}{dt} = \gamma_{h,u} I_u - (\kappa_{r,u} + \kappa_{d,u}) H_u, \tag{1g}$$

$$\frac{dR_u}{dt} = \gamma_{r,u} I_u + \kappa_{r,u} H_u, \tag{1h}$$

$$\frac{dD_u}{dt} = \kappa_{d,u} H_u, \tag{1i}$$

$$\frac{dS_v}{dt} = -\beta^{vv} S_v (I_v + A_v + P_v) - \beta^{vu} S_v (I_u + A_u + P_u) + \theta V(t), \tag{1j}$$

$$\frac{dE_v}{dt} = -\sigma_1 E_v + \beta^{vv} S_v (I_v + A_v + P_v) + \beta^{vu} S_v (I_u + A_u + P_u), \tag{1k}$$

$$\frac{dP_v}{dt} = (1 - \phi_v) \sigma_1 E_v - \sigma_2 P_v, \tag{1l}$$

$$\frac{dA_v}{dt} = \phi_v \sigma_1 E_v - \mu_v A_v, \tag{1m}$$

$$\frac{dU_v}{dt} = \mu_v A_v, \tag{1n}$$

$$\frac{dI_v}{dt} = \sigma_2 P_v - (\gamma_{r,v} + \gamma_{h,v}) I_v, \tag{1o}$$

$$\frac{dH_v}{dt} = \gamma_{r,v} I_v - (\kappa_{r,v} + \kappa_{d,v}) H_v, \tag{1p}$$

$$\frac{dR_v}{dt} = \gamma_{h,v} I_v + \kappa_{r,v} H_v, \tag{1q}$$

$$\frac{dD_v}{dt} = \kappa_{d,v} H_v. \tag{1r}$$

This division of the whole population into two almost independent subgroups (that interact via contacts and through the vaccinated time series) is reflected in the presentation of the model schematic in Fig. 1. In fact, Fig. 1 summarizes population flows and interactions: the left panel illustrates the main population compartments and the corresponding flows, while the right panel shows the interactions of susceptibles with individuals in other compartments, both for vaccinated and unvaccinated populations. The model equations are reported as Eqs. (1).

We made a number of simplifying assumptions to reduce the number of parameters and enhance the identifiability of the model (see also the relevant analysis in the Appendix). We consider a model with only four transmission rates β^{ij} ($i, j = u, v$). We assumed that infectious contacts could only occur between four groups: between unvaccinated individuals (unvaccinated–unvaccinated contacts denoted by the superscript uu), between vaccinated and vaccinated individuals (denoted by the superscript vv) and across these two groups (denoted by uv for vaccinated transmitting to unvaccinated and vu for the reverse path of infection). Notice that uv and vu are not *a priori* assumed to be equivalent. Within each subgroup of infection transmission (uu, vv, uv , and vu), infections induced by the three infectious compartments P, A and I are assumed to occur at the same rate, i.e., the transmission rate is considered to be independent of whether the infectious individual exhibits symptoms (I) or not (A, P). While we expect these transmission rates to differ (in fact, we know that even within a given population the $S \rightarrow I$ transmission rate differs from the $S \rightarrow A$ transmission rate, see for example, Ref. [22]), the identifiability analysis based on the available time series suggests that they would not be independently computable in a definitive way. In addition, we introduced a single

constraint that requires that the incubation period of the disease $\tau_{inc} = \sigma_1^{-1} + \sigma_2^{-1}$ be a value randomly sampled from a normal distribution with mean 3.42 and standard deviation 0.2755. This leverages information about the (shorter) incubation period associated with the omicron variant [29].

Vaccine efficiency is introduced via the parameter θ , whose variation bounds were set in the range 75%–95%. This factor multiplied by the time series of vaccinations $V(t)$ effectively “transfers” individuals from the unvaccinated susceptible population to the vaccinated susceptible compartment.

It is important to remark that H measures both conventional and critical (i.e., ICU) hospitalizations together. While we recognize the relevance of the ongoing debate of distinguishing deaths “from COVID” vs. “with COVID” [30], unfortunately the data available herein do not allow for a definitive distinction between the two.

We obtained the best-fit parameters and initial conditions by minimizing an appropriately chosen norm. For both regions of interest, the time period used for the fits was from November 15, 2021 to March 1, 2022. The identification of the date a particular variant appeared in a geographical location is fraught with uncertainties. The choice of November 15, 2021 as the initial day of fittings stems from a number of indirect indications: Ref. [31] reports a surge of cases in Germany at the beginning of November; Ref. [32] mentions that an omicron-variant case was reported on November 19; and the WHO site [33] mentions that in South Africa the first confirmed infection was reported on November 24, although arising in the sequencing of a sample collected on November 9. Additionally, inspection of the data shows a gradual increase starting at November 15, after a plateau. The effective parameter training period indicated above (till March 1, 2022), is followed by a prediction period (with the optimal parameters and initial conditions fixed, as determined in the training period). The predicted time series that terminates on March 29, 2022 is then compared to the reported data. Predictions do not go beyond that date because the measurement strategy in Andalusia changed, thereby rendering our fixed parameters of limited relevance to the new data. Moreover, around that time Spanish public policy also changed, and face masks were no longer required. In Switzerland some restrictions were removed in the middle of February. More details are presented in the appropriate results sections.

We perform two separate fits, i.e., we use two different norms to compare predictions to reported numbers depending on data availability. First, we fit the predicted total number of fatalities to the reported number by minimizing the norm \mathcal{N} (i.e., the loss function)

$$\mathcal{N} = \frac{1}{n} \sum_{i=1}^n \left\{ \log [D_{u,num}(t_i) + D_{v,num}(t_i)] - \log [D_{obs}(t_i)] \right\}^2, \tag{2}$$

where the subscript “num” refers to predicted (calculated) numbers and “obs” to observation (reported numbers), and n refers to the number (days) of observations. We use the mean square logarithmic error as the norm since it is less prone to outliers and, additionally, when multiple error contributions (which may be of different size) are independently included in the minimization procedure, the norm treats them on

broadly the same scale (due to the properties of the logarithm). This loss function effectively does not *distinguish* the compartmental origin of the fatalities, i.e., whether they arise from the u or v compartments: it only accounts for the cumulative number of fatalities. This will, inevitably, result in the determination of some parameters between the unvaccinated and the vaccinated populations being not necessarily epidemiologically meaningful (as we will see in the detailed comparisons of our predictions for Andalusia and Switzerland).

Whenever we used norm (2), we also included the waning effect of the vaccines and booster vaccination effects, in addition to those fully vaccinated. It is well-documented that different vaccines have different waning immunities (see, for instance, the detailed analysis of Ref. [34]). However, to be able to account for these effects without adding a large number of additional coefficients, we assumed that vaccines are roughly effective for an interval of about 180 days. Consequently, we define $V(t)$ as:

$$V(t) = V_{fv}([t]) - V_{fv}([t - 180]) + V_b([t]), \tag{3}$$

where the subscripts fv refers to fully vaccinated, and b to booster.

The above optimization via norm (2) is a point estimator, that is, a single set of parameters and initial conditions is obtained. To calculate their confidence intervals, and consequently the confidence interval of the predictions, we follow the bootstrapping method described in [35]. The first step is to generate 250 random, synthetic, time series for the fatalities based on the reported data. To accomplish this, we first apply the optimization to find the best fit to the original data set: we refer to that optimization of the reported fatalities data as the “numerical truth”. In this first optimization, we also included $I_u(0)$, $I_v(0)$, $E_u(0)$, $E_v(0)$, $A_u(0)$, $A_v(0)$, $H_u(0)$ and $H_v(0)$ as fitting parameters: these parameters were fixed in the subsequent bootstrapping steps. Then, random noise of a prescribed level, empirically chosen to be 5%, was added to the “numerical truth” to generate 250 “polluted” (i.e., noisy) time series for the fatalities. Knowledge of the error in data collection may be helpful to select an appropriate noise level. In the second step, we apply the optimization procedure to find the best fit to each of the 250 synthetic fatalities time series to obtain 250 sets of parameters from which the confidence intervals for the parameters and predictions can be computed. The same bootstrapping procedure was used for the hospitalization time series.

We also fitted separately, if the reported data allowed us, the vaccinated and unvaccinated fatalities time series using them as separate inputs to our minimization objective. In that case, the relevant norm is

$$\mathcal{N} = \frac{1}{n} \sum_{i=1}^n \left\{ \log[D_{u,num}(t_i)] - \log[D_{u,obs}(t_i)] \right\}^2 + \left\{ \log[D_{v,num}(t_i)] - \log[D_{v,obs}(t_i)] \right\}^2. \tag{4}$$

With this norm, we are genuinely treating the vaccinated compartment separately: we expect its fraction of fatalities (proportionally to the corresponding susceptible population) to be reflected in the obtained parameters. It should be added that in this case, given the way that the data are obtained, the vaccinated status corresponds to people who had received the full doses, independently of antibodies waning, boosting, or efficacy of vaccines. Consequently, θ was fixed to 1 in every fit, and $V(t)$ is defined as

$$V(t) = V_{fv}([t]). \tag{5}$$

We mention here that an important consideration pertinent to the model concerns the identifiability of its coefficients (and initial conditions). This pertains to whether, based on the time series given, the unknown model parameters can be uniquely identified [36]. In addition to the question whether all parameters can be uniquely identified (global identifiability) or some may have multiple possible values (local identifiability), there are also practical issues concerning whether different sets of parameters lead to similar (although not necessarily identical) observations; see, e.g., the discussion of [37]. Here, following

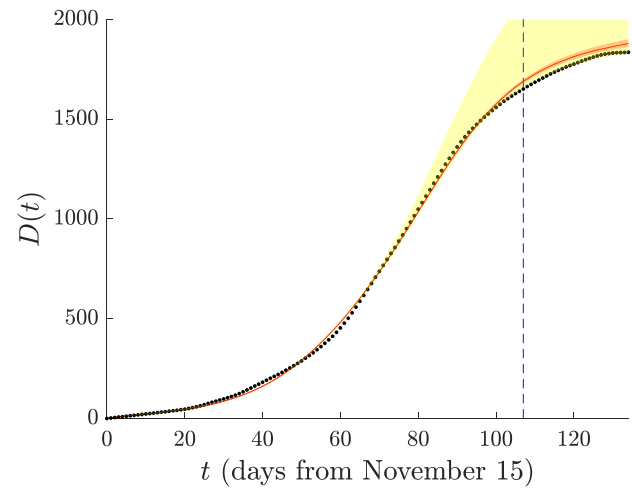


Fig. 2. Omicron-variant model A1: Fit and prediction for the total number of fatalities in Andalusia (norm of Eq. (2)). The calculated curve is plotted in red, while the yellow shade corresponds to the 95% confidence interval comprised between the 2.5 and 97.5 percentiles. Reported data for the total number of fatalities are given by the black points. The vertical line, the beginning of the prediction interval, is March 1, 2022.

the approach presented in the Appendix (see also [38–41]), we find that all the parameters are globally identifiable, except

$$\gamma_{h,u}, \gamma_{r,u}, \gamma_{h,v}, \gamma_{r,v}, \kappa_{d,u}, \kappa_{r,u}, \kappa_{d,v}, \kappa_{r,v}.$$

For these eight parameters, the following combinations are globally identifiable:

$$\gamma_{h,u} + \gamma_{r,u}, \gamma_{h,v} + \gamma_{r,v}, \kappa_{d,u} + \kappa_{r,u}, \kappa_{d,v} + \kappa_{r,v}, \gamma_{h,u}\kappa_{d,u}, \gamma_{h,v}\kappa_{d,v}.$$

As concerns the initial conditions, $H_u(0)$ and $H_v(0)$ are not identifiable, in addition to the initial conditions for the terminal compartments U_u, U_v, R_u, R_v .

The identifiability analysis suggests that ϕ_u, ϕ_v and many other parameters are globally identifiable, i.e., they have a unique value given the functions $D_u(t)$ and $D_v(t)$. However, these theoretical-analysis results do not exactly transfer to numerical calculations for various reasons. A globally identifiable parameter may not necessarily have a sharp estimate due to the potential sloppiness [42] of the model. When the output functions are not sensitive to a particular parameter, a sharp estimate will not be expected, even though the parameter may be globally identifiable. Many other factors, e.g., the reliability and accuracy of the reported time series, may exacerbate the situation. The identifiability analysis assumes that both functions $D_u(t)$ and $D_v(t)$ are outputs, which includes much more information than the simple loss term Eq. (4) can provide. It is an open question how close these estimates are to the actual parameters. Interestingly, however, we find that these estimates still give fairly accurate predictions, even though certain parameter estimates may not be as sharp as desired. When the total death, $D_u(t) + D_v(t)$, is the only output, we cannot obtain any identifiability results. The implications of these remarks on model identifiability are further elaborated in our comments of best-fit parameters.

2.1.1. Andalusia

As mentioned, the fitting time window we used to obtain the optimized parameters and initial conditions was from November 15, 2021 to March 1, 2022. The prediction interval ended on March 29, 2022. The time series for Andalusia is available from the Spanish Health Ministry, but we used the series compiled at [43]. Note that in Andalusia the reported values of $D(t)$, and the total number of hospitalizations $J(t)$, the latter discussed in Section 2.2, correspond to

Table 1

Optimal parameters and initial conditions for the omicron-variant models in Andalusia: Model fits to the total number of fatalities, discussed in Section 2.1.1 (model A1, third column) and to the total number of hospitalizations, discussed in Section 2.2.1 (model A2, right column). Population $N_{\text{And}} = 8.4M$.

Parameter	Symbol	Median (interquartile range)	
		Fit to total number of deaths [Model A1, norm Eq. (2)]	Fit to total number of hospitalizations [Model A2, norm Eq. (7)]
Transmission rate uu [per day]	β^{uu}	0.6437 (0.6313–0.6532)	0.1279 (0.1134–0.1491)
Transmission rate uv [per day]	β^{uv}	0.3687 (0.3040–0.4044)	0.0209 (0.0166–0.0270)
Transmission rate vu [per day]	β^{vu}	0.0679 (0.0591–0.0822)	0.2400 (0.1775–0.2855)
Transmission rate vv [per day]	β^{vv}	0.1939 (0.1839–0.2168)	0.4516 (0.4381–0.4637)
Latent period [days]	$1/\sigma_1$	1.7923 (1.6999–1.8935)	1.8143 (1.7266–1.9164)
Preclinical period [days]	$1/\sigma_2$	1.6467 (1.5513–1.7496)	1.6300 (1.5038–1.7509)
A_u/P_u partitioning	ϕ_u	0.3689 (0.3639–0.3832)	0.3592 (0.3486–0.3743)
A_v/P_v partitioning	ϕ_v	0.3363 (0.3311–0.3457)	0.3933 (0.3816–0.4063)
Infectivity period (A_u) [days]	$1/\mu_u$	2.9185 (2.8497–2.9433)	3.1540 (3.0764–3.2333)
Recovery rate $I_u \rightarrow R_u$ [per day]	$\gamma_{r,u}$	0.1999 (0.1958–0.2103)	0.1998 (0.1871–0.2082)
Transition rate $I_u \rightarrow H_u$ [per day]	$\gamma_{h,u}$	0.0056 (0.0050–0.0060)	0.0068 (0.0062–0.0072)
Infectivity period (A_v) [days]	$1/\mu_v$	3.2038 (3.1581–3.2467)	3.3018 (3.1663–3.6227)
Recovery rate $I_v \rightarrow R_v$ [per day]	$\gamma_{r,v}$	0.1728 (0.1692–0.1802)	0.2088 (0.2023–0.2153)
Transition rate $I_v \rightarrow H_v$ [per day]	$\gamma_{h,v}$	0.0057 (0.0051–0.0064)	0.0015 (0.0014–0.0017)
Recovery rate $H_u \rightarrow R_u$ [per day]	$\kappa_{r,u}$	0.3436 (0.3314–0.3576)	–
Death rate $H_u \rightarrow D_u$ [per day]	$\kappa_{d,u}$	0.0090 (0.0080–0.0106)	–
Recovery rate $H_v \rightarrow R_v$ [per day]	$\kappa_{r,v}$	0.3354 (0.3198–0.3506)	–
Death rate $H_v \rightarrow D_v$ [per day]	$\kappa_{d,v}$	0.0097 (0.0092–0.0101)	–
Vaccine efficiency [-]	θ	0.8825 (0.8767–0.8885)	0.8510 (0.8436–0.8598)
Initial condition			
Initial unvaccinated Exposed (E_u) population [#]	$E_u(0)$	641	1485
Initial unvaccinated Presymptomatic (P_u) population [#]	$P_u(0)$	2217	1541
Initial unvaccinated Asymptomatic (A_u) population [#]	$A_u(0)$	2273	1124
Initial unvaccinated symptomatically Infected (I_u) population [#]	$I_u(0)$	2677	2438
Initial unvaccinated Hospitalized (H_u) population [#]	$H_u(0)$	60	–
Initial vaccinated Exposed (E_v) population [#]	$E_v(0)$	840	3218
Initial vaccinated Presymptomatic (P_v) population [#]	$P_v(0)$	2902	3339
Initial vaccinated Asymptomatic (A_v) population [#]	$A_v(0)$	2976	2435
Initial vaccinated symptomatically Infected (I_v) population [#]	$I_v(0)$	3504	5281
Initial vaccinated Hospitalized (H_v) population [#]	$H_v(0)$	370	–

the event day, whereas for Switzerland they correspond to the report day.

The vaccination data were extracted from the Regional Government of Andalusia (Junta de Andalucía, Ref. [44]). We ignored the vaccinations for kids under 12 years old, as there were many data anomalies, resulting in a time series that appears to be problematic. Irrespective of that, this population segment corresponds to only $\lesssim 4\%$ of the total vaccinations. The fatalities time series we used did not report how many fatalities could be attributed to vaccinated or unvaccinated individuals. Therefore, for the region of Andalusia we used only norm (2), coupled to the modified vaccination time series as described in Eq. (3), to perform the optimizations.

Fig. 2 shows the calculated fatalities time series (both fitting and prediction intervals) and the reported numbers. Table 1 (model A1 in column 3) presents the optimized parameters and initial conditions. The reported interquantile range arises from 250 fits in the bootstrapping step, as discussed above.

We observe that the overall trend of the fatalities seems to be reasonably well captured by the model within its prediction intervals (and their associated uncertainty). We do note, however, a slight over-prediction towards the end of the time series, during March 2022. The transmission rates β^{ij} ($i, j = u, v$) with at least one member of the unvaccinated population uu and uv are clearly higher than the vv rate between members of the vaccinated population. In fact, β^{uu} is more than three times higher than β^{vv} . The lowest transmission rate is predicted to be β^{vu} , even lower than β^{vv} . At this point it is important to recall our discussion about the use of the total number of deaths and the resulting inability to identify definitively the model parameters. Hence, the above numbers, even when they appear to be intuitively relevant, should be taken with a grain of salt. The latent period is approximately 3.5 days, as imposed by our constraint, and in agreement with [29]. The role of asymptomatics, as reflected by the fraction ϕ_i of exposed who become asymptomatics, is considerable, approximately

$1/3$ and independent of whether the population is vaccinated or not. The calculated fraction of asymptomatics is in reasonable agreement with Ref. [45] who reported a pooled fraction of asymptomatics for the omicron variant of 25.5% (95% confidence interval 17.0%–38.2%). The vaccinated and unvaccinated infectivity period for asymptomatic infections $1/\mu_i$ ($i = u, v$) is approximately constant, at about three days, again independent of whether the u or v compartment is considered.

We also find that some parameters are more difficult to justify. Specifically, we find that the recovery rates of symptomatically infected individuals $I_i \rightarrow R_i$ and that of the hospitalized individuals $H_i \rightarrow R_i$ are almost independent of whether the population is vaccinated or not. The same holds for the transition rates $I_i \rightarrow H_i$ and the death rates $H_i \rightarrow D_i$; all four of them are found to have weak variations. The independence of these rates on the administration of the vaccine might be related to the norm we used that does not distinguish between fatalities of vaccinated or unvaccinated individuals. We will return to this point in our analysis of the Switzerland data.

2.1.2. Switzerland

We chose to perform model calculations for a territory with a population similar in number to that of Andalusia, and for which adequate data are available. As such, we chose a region that contains Switzerland and the Principality of Liechtenstein (data are jointly reported) since the total population of this aggregate territory is 8.7M, (compared to 8.4M for Andalusia). Overall, we followed a procedure very similar to what we used for Andalusia, with a few minor changes. Identical fitting and prediction intervals are used as those for Andalusia. We do note, however, that starting February 17, 2022 most restrictions were lifted in Switzerland. We believe this is one of the reasons we observe a model under-prediction of the number of fatalities in Figs. 3 and 4.

Case reporting was slightly different. The Swiss government through the Federal Office of Public Health provides daily the status (vaccinated, unvaccinated or unknown) of each hospitalized/deceased person

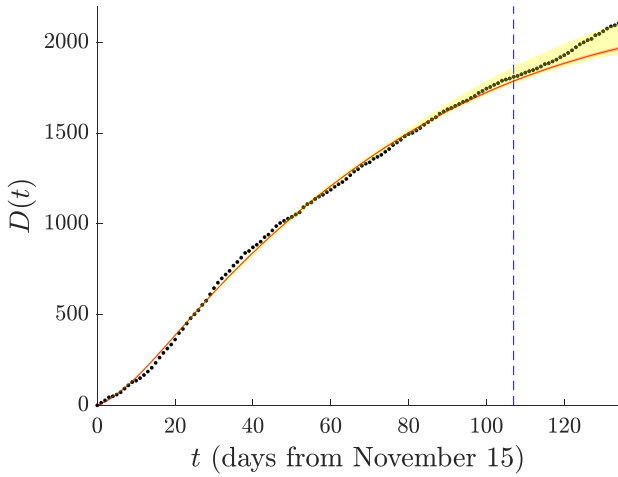


Fig. 3. Omicron-variant model A1: Fit and prediction for the total number of fatalities in Switzerland (norm Eq. (2)). The calculated curve is plotted in red, while the yellow shade corresponds to the 95% confidence interval comprised between the 2.5 and 97.5 percentiles. Reported data for the total number of fatalities are given by the black points. The vertical line, the beginning of the prediction interval, is March 1, 2022.

at Ref. [46]. In the absence of a concrete metric on how to partition unknown fatalities to vaccinated and unvaccinated individuals, we used the following procedure to convert these three time series into two, one associated with vaccinated $D_v(t_i)$ and the other to unvaccinated individuals $D_u(t_i)$. Let $\bar{d}_v(t_i)$, $\bar{d}_u(t_i)$ and $\bar{d}_*(t_i)$ denote the number of daily reported fatalities with vaccinated, unvaccinated, and unknown state, respectively. We randomly sample an integer number $\delta_i \in [0, \bar{d}_*(t_i)]$, following a uniform distribution, and then we define the daily number of vaccinated/unvaccinated deceased as $d_v(t_i) = \bar{d}_v(t_i) + \delta_i$ and $d_u(t_i) = \bar{d}_u(t_i) + [\bar{d}_*(t_i) - \delta_i]$. The total number of deaths is the cumulative sum, i.e. $D_v(t_i) = \sum_{j=1}^i d_v(t_j)$ and $D_u(t_i) = \sum_{j=1}^i d_u(t_j)$. Note that we followed the same procedure to generate the hospitalizations J_v and J_u used in model A2, in Section 2.2.2. As mentioned earlier, $D(t)$ and $J(t)$ for Switzerland correspond to the report day.

Given the reconstructed time series $D_i(t)$ we used norm (4), in addition to the norm (2) used in the case of Andalusia, to fit and predict the fatalities time series for the territory of Switzerland (and the Principality of Liechtenstein). We attempted to fit separately the vaccinated and unvaccinated deceased, using them as separate inputs to our minimization objective. As mentioned earlier, since we consider that vaccinated individuals have received the full dose (neglecting immunity waning, boosting, of vaccine efficiency) we take $\theta = 1$ in every fit, and $V(t)$ is defined as described in Eq. (5). Our results for the fit to the total number of fatalities are shown in Fig. 3, whereas those for the separate fits to vaccinated and unvaccinated deaths are presented in Fig. 4. Table 2, columns two and four, presents the fitting parameters and initial conditions.

We can see a clear model under-prediction of the fatalities (within the prediction interval), for both optimizations (norm (2) and (4)), despite an accurate following of the time-series trend throughout the period over which regression is performed. The under-prediction is more severe in the case of the total number of fatalities, Fig. 3, and in the vaccinated death time series of the right panel in Fig. 4. As mentioned earlier, we attribute the under-prediction to the fact that after the end of the fitting period, restrictions were considerably relaxed leading to more cases, and eventually more fatalities, a feature that was not explicitly factored in the model.

As regards the parameters of the model, we observe very similar trends to what we obtained for Andalusia. A notable exception is that in the total-deaths fit β^{uv} is the highest transmission rate, retaining however $\beta^{uu} \gg \beta^{vv}$ in the case of norm (2). The latent period is well reproduced (as expected due to the constraint and Ref. [29]),

and the fraction of asymptomatics is approximately 25% (again in agreement with [45]) irrespective of vaccination or not. The remaining parameters follow similar trends as reported in Table 1 for Andalusia. It is noteworthy that in both cases recovery, transmission, and death rates seem to depend relatively weakly on whether the vaccine had been administered or not.

A comparison of the parameters predicted by the two optimization is in order. When the two distinct populations are used in the regression, we observe, fourth column in Table 2, that the transition rate β^{uv} becomes the largest one with β^{uu} being the smallest. While a calculated higher viral transmissivity of vaccinated individuals could, in principle, be attributed to taking fewer measures to limit pathogen transmission via behavioral changes, e.g., higher contact rates, negligence to use face masks, etc, it is not obvious that such an attribution is meaningful, rather than the potential outcome of the sloppiness of the model. Another surprising feature is that we do not find a significant dependence of the parameters on the norm used (apart from the noted difference in the transmission rates). The asymptomatic fraction is predicted to be slightly larger, approximately 30%, the $H_u \rightarrow R_u$ ($\kappa_{r,u}$) is slightly smaller, and the death rate is slightly larger ($\kappa_{d,v}$).

2.2. Model A2: Branches terminate at hospitalizations

We now consider model A2, an omicron-variant model similar to A1, but where the population branches terminate at the total number of hospitalizations (see Fig. 5). The total hospitalization data appear, in our gauge, to be more reliable than fatalities, as the latter (at least in Andalusia) include deceased by any cause that may have recently generated a positive test. That is to say, we believe that numerous fatalities were attributed to COVID even though the primary reason for these events had not been COVID [47]. By considering the reported (total, namely conventional and critical) hospitalizations, this possible misattribution of fatalities to COVID-19 may be diminished. Model A2 is the same as model A1 described by Eqs. (1), differing only in the terminal compartments of hospitalizations. This implies that the ODEs Eqs. (1a)–(1f) and Eqs. (1j)–(1o) form part of the model A2 equations, as well. However, Eqs. (1g)–(1i) and Eqs. (1p)–(1r) are to be replaced by

$$\begin{aligned} \frac{dJ_i}{dt} &= \gamma_{h,i} I_i, \quad i = u, v, \\ \frac{dR_i}{dt} &= \gamma_{r,i} I_i, \quad i = u, v \end{aligned} \tag{6}$$

where $J(t)$ is the total number of hospitalizations, and γ_{ij} is the rate symptomatically infected individuals u, v become hospitalized $\gamma_{h,i}$ or recovered $\gamma_{r,i}$.

The optimal parameters (and initial conditions) are obtained by a procedure similar to what we used in model A1 with $J(t)$ playing the role of $D(t)$. Accordingly, the norms change: Eq. (2) becomes

$$\mathcal{N} = \frac{1}{n} \sum_{i=1}^n \left\{ \log [J_{u,num}(t_i) + J_{v,num}(t_i)] - \log [J_{obs}(t_i)] \right\}^2, \tag{7}$$

while Eq. (4) becomes

$$\begin{aligned} \mathcal{N} = \frac{1}{n} \sum_{i=1}^n \left\{ \log [J_{u,num}(t_i)] - \log [J_{u,obs}(t_i)] \right\}^2 \\ + \left\{ \log [J_{v,num}(t_i)] - \log [J_{v,obs}(t_i)] \right\}^2. \end{aligned} \tag{8}$$

2.2.1. Andalusia

The results of fitting the total hospitalizations (arising from both vaccinated plus unvaccinated populations using norm (7)) are shown in Fig. 6. The optimal parameters and initial conditions are summarized in Table 1, last column.

As in the case of the cumulative optimization of both the vaccinated and the unvaccinated fatalities, the results of Fig. 6 appear quite accurate, including the forward prediction for the month of March

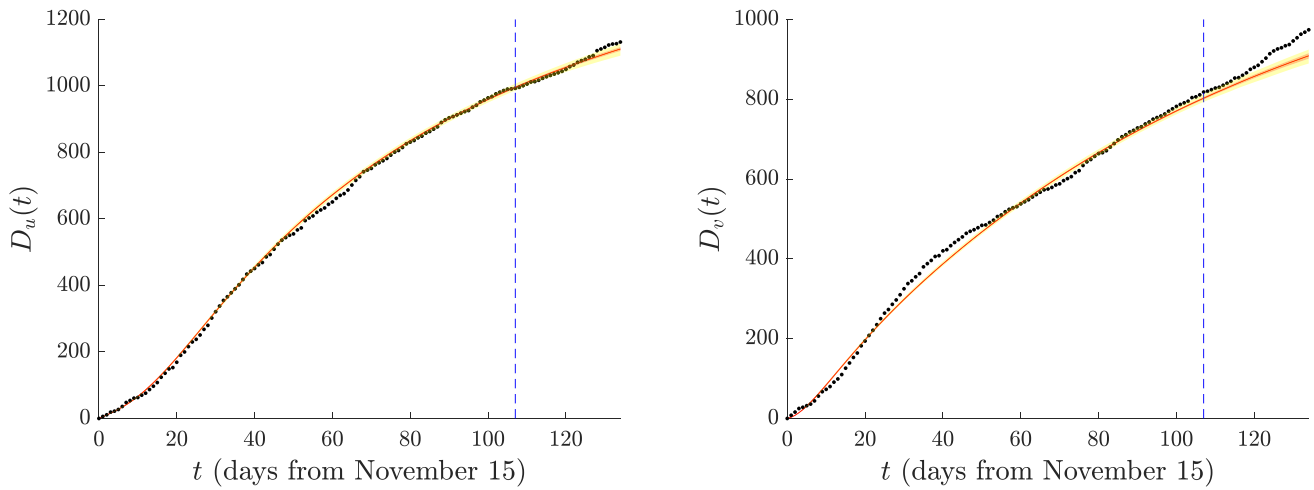


Fig. 4. Omicron-variant model A1: Separate fits of unvaccinated and vaccinated fatalities in Switzerland (norm Eq. (4)). Left panel: Fatalities of unvaccinated individuals. Right panel: Fatalities of vaccinated individuals. Calculated curves are plotted in red, whereas the yellow shade presents the 95% confidence interval comprised between the 2.5 and 97.5 percentiles. Reported data for the total number of vaccinated and unvaccinated fatalities are given by the black points. The vertical line, the beginning of the prediction interval, is March 1, 2022.

Table 2

Optimal parameters and initial conditions for the omicron-variant models in Switzerland: Model fits to (a) the total number of fatalities (model A1, second column from the left), (b) the total number of hospitalizations (model A2, third column), (c) to the u, v total number of fatalities separately (model A1, fourth column), and (d) to the u, v total number of hospitalizations separately (model A2, fifth column). Population $N_{\text{CHL}} = 8.7M$. Parameters descriptions are as defined in Table 1.

Parameter	Median (interquartile range) Fit: total number of deaths [Model A1, norm Eq. (2)]	Median (interquartile range) Fit: total hospitalizations [Model A2, norm Eq. (7)]	Median (interquartile range) Fit: u, v deaths separately [Model A1, norm Eq. (4)]	Median (interquartile range) Fit: u, v hospitalizations separately [Model A2, norm Eq. (8)]
β^{uu}	0.1682 (0.1491–0.1915)	0.2563 (0.2241–0.2855)	0.1607 (0.1339–0.1862)	0.3418 (0.3300–0.3539)
β^{uv}	0.2450 (0.2088–0.2840)	0.2084 (0.1798–0.2328)	0.2023 (0.1840–0.2235)	0.1509 (0.1216–0.1846)
β^{vu}	0.0780 (0.0541–0.0979)	0.0709 (0.0357–0.0995)	0.0261 (0.0158–0.0399)	0.0500 (0.0453–0.0551)
β^{vv}	0.0977 (0.0790–0.1144)	0.1422 (0.1123–0.1711)	0.2266 (0.1917–0.2625)	0.1885 (0.1737–0.2037)
$1/\sigma_1$	1.7558 (1.6627–1.8545)	1.7605 (1.6190–1.9650)	1.7811 (1.6837–1.8945)	1.7169 (1.6315–1.8140)
$1/\sigma_2$	1.6798 (1.5660–1.8040)	1.6690 (1.4572–1.8302)	1.6384 (1.5075–1.7495)	1.7085 (1.5905–1.8255)
ϕ_u	0.2469 (0.2333–0.2588)	0.2659 (0.2472–0.2865)	0.3125 (0.2885–0.3366)	0.3360 (0.2880–0.3803)
ϕ_v	0.2551 (0.2465–0.2617)	0.2680 (0.2546–0.2834)	0.3346 (0.3063–0.3535)	0.3161 (0.2936–0.3337)
$1/\mu_u$	3.1794 (3.0941–3.2628)	3.1836 (3.0829–3.3157)	3.2599 (3.1700–3.4231)	3.4953 (3.2164–3.9040)
$\gamma_{r,u}$	0.1509 (0.1335–0.1661)	0.1845 (0.1678–0.2006)	0.1386 (0.1263–0.1511)	0.1731 (0.1562–0.1867)
$\gamma_{h,u}$	0.0115 (0.0091–0.0133)	0.0037 (0.0034–0.0042)	0.0107 (0.0093–0.0121)	0.0039 (0.0038–0.0041)
$1/\mu_v$	3.2920 (3.2141–3.4560)	3.3012 (3.1761–3.4442)	3.0606 (2.8636–3.2685)	3.2347 (3.0419–3.4284)
$\gamma_{r,v}$	0.1607 (0.1431–0.1707)	0.1679 (0.1357–0.1892)	0.1627 (0.1469–0.1892)	0.1424 (0.1315–0.1571)
$\gamma_{h,v}$	0.0106 (0.0095–0.0120)	0.0028 (0.0025–0.0031)	0.0105 (0.0095–0.0118)	0.0029 (0.0028–0.0029)
$\kappa_{r,u}$	0.2521 (0.2216–0.2852)	–	0.1882 (0.1515–0.2290)	–
$\kappa_{d,u}$	0.0185 (0.0162–0.0207)	–	0.0186 (0.0178–0.0193)	–
$\kappa_{r,v}$	0.1886 (0.1314–0.2738)	–	0.1688 (0.1576–0.1808)	–
$\kappa_{d,v}$	0.0071 (0.0060–0.0084)	–	0.0109 (0.0102–0.0116)	–
θ	0.8483 (0.8417–0.8552)	0.8662 (0.8529–0.8871)	1	1
Initial condition				
$E_u(0)$	5648	8234	2690	6483
$P_u(0)$	4767	4340	2677	5530
$A_u(0)$	3609	3762	988	3110
$I_u(0)$	11 318	12 759	5189	12 655
$H_u(0)$	295	–	278	–
$E_v(0)$	9401	8642	10 330	6477
$P_v(0)$	7936	4555	10 278	5524
$A_v(0)$	6007	3949	3795	3107
$I_v(0)$	18 839	13 392	19 925	12 644
$H_v(0)$	381	–	200	–

(despite a slight under-prediction of hospitalizations). Note that model A1 predictions (see Fig. 2) were slightly above reported fatalities. However, a more careful inspection of the obtained parameters suggests that some of them may not be epidemiologically realistic. Inspection of the optimized transmission rates shows that the vv rate is larger than the uu rate (specifically, we note $\beta^{vv} > \beta^{uu}$, $\beta^{vu} > \beta^{uv}$, and $\beta^{vu} > \beta^{uu}$). Whereas these inequalities may be related to changes in the behavior of vaccinated individuals, (for example, vaccinated individuals may take fewer precautions and socialize more) we believe instead that this

aspect points to the non-identifiability of the model. It is also likely that the origin of these transmission rates stems from our regression's inability to expressly distinguish between the two u, v compartments. More concretely, and similarly to model A1, the only available output is the cumulative number of hospitalizations $J_u(t) + J_v(t)$. However, for such an output we could not obtain any identifiability results following the procedure described in the Appendix. The model is too complex and beyond the capability of the current identifiability-analysis packages.

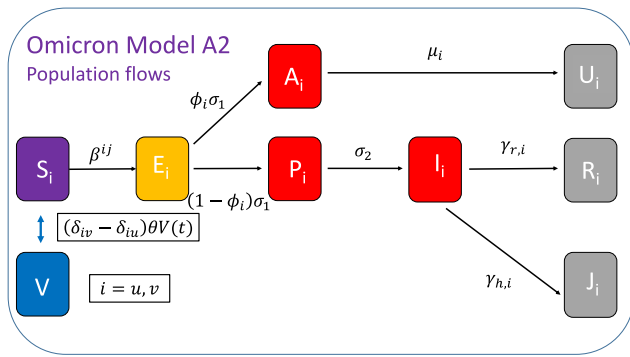


Fig. 5. Schematic diagram of the population flows according to model A2. The susceptible interactions are as in model A1, shown in the right panel of Fig. 1. The symbol δ_{ij} with $i, j = u, v$ is the Kronecker delta.

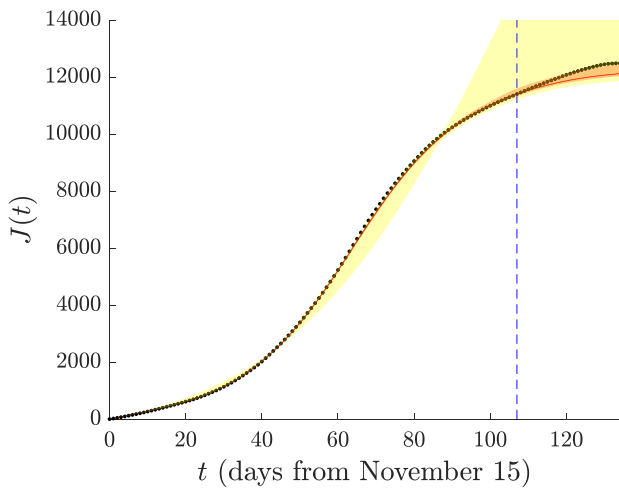


Fig. 6. Omicron-variant model A2: Fit and prediction for the total number of hospitalizations in Andalusia (norm of Eq. (7)). The calculated curve is plotted in red, whereas yellow shade corresponds to the 95% confidence interval comprised between the 2.5 and 97.5 percentiles. Reported data for the total number of hospitalizations are given by the black points. The vertical line, the beginning of the prediction interval, is March 1, 2022.

2.2.2. Switzerland

We followed the same procedure to generate the model A2 fits and predictions for Switzerland (including the Principality of Liechtenstein). As for model A1, we fitted both the total number of hospitalizations via norm (7) and the two vaccination-identified compartments via norm (8). We followed the same procedure as that used to determine $D_u(t)$ and $D_v(t)$ to obtain estimates for $J_u(t)$ and $J_v(t)$.

Fig. 7 presents our results for the fit to total number of hospitalizations. Fig. 8, instead, corresponds to the vaccinated and unvaccinated populations considered separately in the regression. Table 2, third and fifth column, summarizes all the fitting parameters and initial conditions.

We can see a clear model under-prediction of hospitalizations in the prediction interval, despite an accurate following of the time-series throughout the period over which regression is performed. The under-prediction is more pronounced in the case of the fit to the total number of hospitalizations. In the case of the separate fittings, the hospitalizations of the vaccinated population are more under-predicted than the hospitalizations of unvaccinated individuals. We attribute this to the fact that, as also discussed in the context of fatalities, towards the end of the fitting period, restrictions were considerably relaxed leading to more cases, and eventually more fatalities. As regards the parameters of the model, we find the transmission rates to be more in

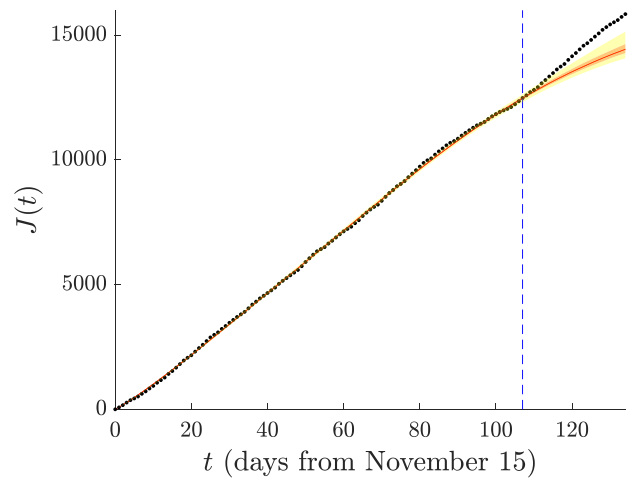


Fig. 7. Omicron-variant model A2: Fit and prediction of the total number of hospitalizations in Switzerland (norm of Eq. (7)). The calculated curve is plotted in red, whereas the yellow shade presents the 95% confidence interval comprised between the 2.5 and 97.5 percentiles. Reported data for the total number of hospitalizations are given by the black points. The vertical line, the beginning of the prediction interval, is March 1, 2022.

line with what one might typically expect. In particular, $\beta^{uu} > \beta^{vv}$ for both fittings (with the two different norms), but $\beta^{uv} > \beta^{vv}$ for the fitting to the total hospitalization, whereas the reverse is true for the fitting to the two separate populations (although in both cases, the rates are fairly similar, even more so when considering the interquartile ranges). The fraction of asymptomatics varies from approximately 27% to 33%, again in reasonable agreement with [45]. The remaining parameters do not seem to significantly depend on the norm chosen.

In our identifiability analysis of model A2 with the two vaccination compartments treated separately we considered, as in the case of model A1, that $J_u(t)$ and $J_v(t)$ are separately and continuously known. Moreover, we took $\theta = 1$. Following the procedure described in the Appendix, we can show that all parameters and initial conditions are globally identifiable. This implies that, in principle, all initial conditions and parameters can be determined from the output $J_u(t)$ and $J_v(t)$. Practically speaking, however, only the discrete time series $J_u(t)$ and $J_v(t)$ are known: we do not have the full information of the continuous changes of $J_u(t)$ and $J_v(t)$ as the identifiability analysis supposes. Hence, the loss function used in the parameter estimation is based on a discrete time series reflecting a finite number of observations. Given the complexity of the model and the large number of parameters involved, the optimization package often fails to find a minimum. To alleviate the situation, we chose to fix certain initial conditions, even though such a choice is inconsistent with being globally identifiable. Combined with the possible sloppiness of the model, the result of such a choice may be that the estimates for some of the parameters may not be as sharp. We indicate the above to mitigate a potential impression (to the reader) that the mathematically obtained global identifiability of the model should be expected to translate into the most definitive model results.

3. Delta variant

Having explored the more elaborate model setting of the omicron variant we now turn to the simpler case of the delta variant. What simplifies the model considerably is that it is sufficient to consider a single susceptible population since infection of vaccinated individuals was rare. The SARS-CoV-2 vaccines were highly effective against the delta variant, leading to rather few breakthrough infections. Consequently, susceptible individuals who are vaccinated are added to a “withdrawn” population. An alternative option is to add them to the

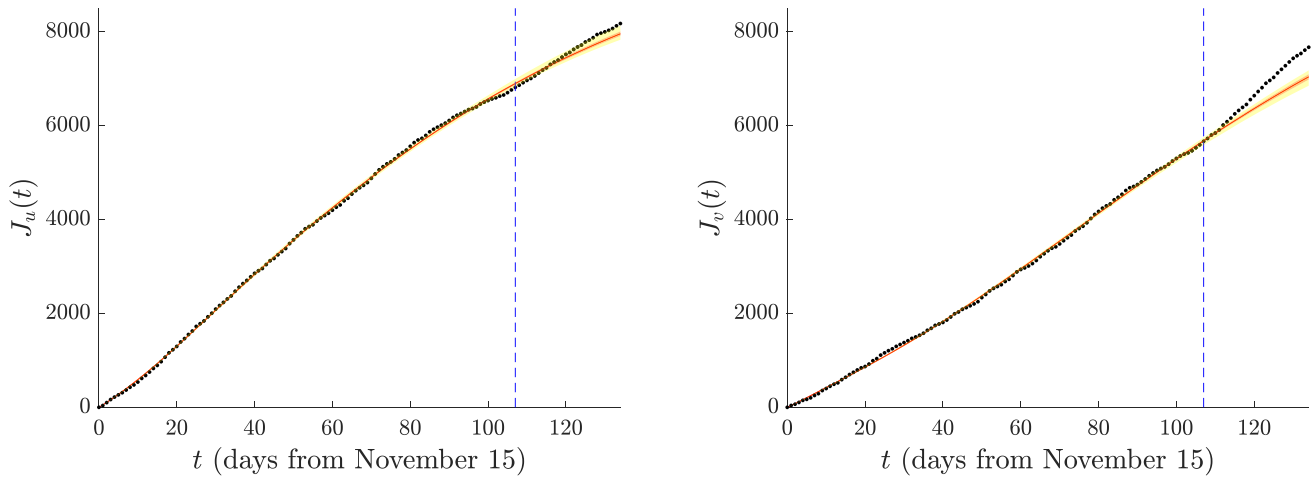


Fig. 8. Omicron-variant model A2: Separate fits of unvaccinated and vaccinated hospitalizations in Switzerland (norm Eq. (8)). Left panel: Hospitalizations of unvaccinated individuals. Right panel: Hospitalizations of vaccinated individuals. Calculated curves are plotted in red, whereas the yellow shade presents the 95% confidence interval comprised between the 2.5 and 97.5 percentiles. Reported data for the total number of vaccinated and unvaccinated hospitalizations are given by the black points. The vertical line, the beginning of the prediction interval, is March 1, 2022.

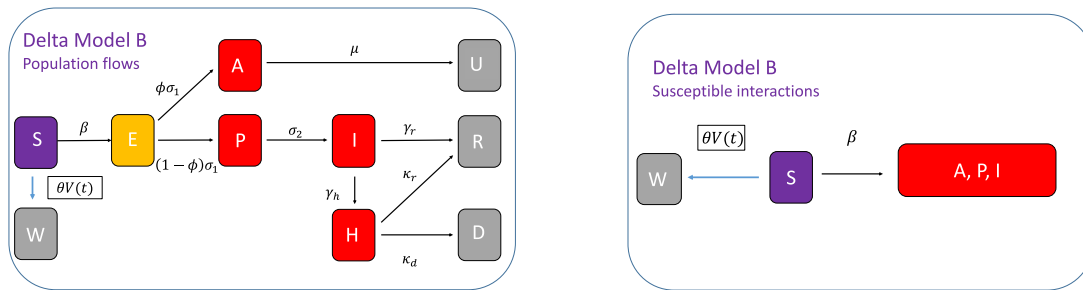


Fig. 9. Schematic diagram of population flows according to the delta-variant model B (left panel) and susceptible interactions with other population compartments (right panel). A single population is modeled, as we consider that the waning immunity time scale (either due to vaccine immunity or to recovery) is much longer than the time scale of prevalence of the delta variant. Vaccinated individuals W are permanently removed from the susceptible compartment, their population becoming a terminal compartment of the model.

recovered population in the sense that this is a *terminal* compartment of the model. For the time period of 3–4 months for which the delta variant was dominant, the potential waning of immunity (either from recovery or from the vaccine) is not considered sufficient to allow these individuals to replenish the susceptible compartment. For more persistent variants, replenishing the susceptible population may be relevant. The small number of breakthrough infections removes the need for a detailed modeling of compartments within the vaccinated category. Accordingly, the relevant model with the same (but single-component) populations as before and with the addition of the withdrawn (W) compartment reads as reported in Eqs. (9). A schematic of the population flows (left panel) and the susceptible interactions is shown in Fig. 9.

The initial conditions are taken in a similar fashion as in the omicron variant, except for $S(0)$, which is taken as a fitting parameter. We chose to render it a fitting parameter since susceptibles who became infected with previous variants are immune to the delta variant. However, their number is not definitively known. Moreover, as in our modeling of the omicron variant via models A1 and A2, we supposed that the transmission rate β is the same for all three infectious compartments: asymptomatics, presymptomatics, and for the symptomatically infected population. In addition, as in the case of the omicron-variant models, we imposed the single constraint on the incubation period $\tau_{inc} = \sigma_1^{-1} + \sigma_2^{-1}$ to be equal to a value randomly sampled following a normal distribution whose mean now is 4.41 and standard deviation

0.3291, in line with what is reported in [29].

$$\begin{aligned}
 \frac{dS}{dt} &= -\beta S(I + A + P) - \theta V(t), \\
 \frac{dE}{dt} &= -\sigma_1 E + \beta S(I + A + P), \\
 \frac{dP}{dt} &= (1 - \phi)\sigma_1 E - \sigma_2 P, \\
 \frac{dA}{dt} &= \phi\sigma_1 E - \mu A, \\
 \frac{dU}{dt} &= \mu A, \\
 \frac{dI}{dt} &= \sigma_2 P - (\gamma_r + \gamma_h)I, \\
 \frac{dH}{dt} &= \gamma_h I - (\kappa_r + \kappa_d)H, \\
 \frac{dR}{dt} &= \gamma_r I + \kappa_r H, \\
 \frac{dD}{dt} &= \kappa_d H, \\
 \frac{dW}{dt} &= \theta V(t).
 \end{aligned} \tag{9}$$

The norm associated with model B, and minimized during the optimization procedure, is:

$$\mathcal{N} = \frac{1}{n} \sum_{i=1}^n \left\{ \log[D_{num}(t_i)] - \log[D_{obs}(t_i)] \right\}^2. \tag{10}$$

Again, following the approach presented in the Appendix, we find that the three parameters σ_1, ϕ, μ and the following combinations

$$(\gamma_h + \gamma_r) + \sigma_2, \quad (\gamma_h + \gamma_r) \cdot \sigma_2, \quad \kappa_d + \kappa_r, \quad \frac{\beta(\gamma_h + \gamma_r)}{\gamma_h \kappa_d}, \quad \frac{\gamma_h}{\gamma_h + \gamma_r} \kappa_d \theta,$$

are globally identifiable. Thus, σ_2 and the sum ($\gamma_h + \gamma_r$) are locally identifiable. Initial conditions other than $D(0)$, which is explicitly available through the fatality time series, are not identifiable. As discussed in the identifiability analyses of the two omicron models, this is just the theoretical result, based on the analysis of [39]. As we have empirically observed, when we fix certain initial conditions or choose a bound for the parameters to be optimized, the identifiability properties of the model may change. In that light, the relevant parameter identifications should be considered with the associated practical “word of caution” indicated above.

3.1. Andalusia and Switzerland

In our simulations, the fitting window for the Andalusia calculation started on June 15, 2021, whereas it started on July 1, 2021 for the Switzerland simulations. The choice of the initial time was determined from the existence of a plateau in the associated time series. The fitting period ended on October 1, 2021 for both regions, and the prediction interval terminated on November 1, 2021. As mentioned in Section 2, the omicron variant appeared in November 2021.

Fig. 10 shows the results of our model B simulations, both for the fitting and the prediction intervals. The left panel presents results for Andalusia, whereas the right panel does it for Switzerland. The best fitting parameters and initial conditions for both countries corresponding to the model B ODEs are presented in Table 3. In the case of Andalusia, we observe a high quality fit, not only for the regression interval but also for the prediction interval. Nevertheless, some parameters do not seem to be in agreement with current knowledge of the epidemiology of the delta variant of SARS-CoV-2. We believe, that the primary reason is the lack of identifiability of the model (both the local aspects thereof theoretically, as well as the practical aspect highlighted above in connection to data and initial condition choices). For example, the recovery time of asymptomatics, $1/\mu$ is found to be ≈ 3.5 days, and an upper bound to the recovery time of symptomatically infected is $1/\gamma_r \approx 5.4$ days. It may be expected that both time scales are likely to be longer than these predictions, although these numbers are in reasonable correspondence with findings, e.g., such as the ones of [48] for the delta variant. On the other hand, the fraction of asymptomatics, 8%, is in agreement with the review and analysis of [45] who found a considerably smaller fraction of asymptomatics associated with the delta than the omicron variants, again in agreement with our calculations.

Our model calculations for Switzerland in Fig. 10 provide a reasonable fit throughout the training interval: calculations initially under-predict and later over-predict. Nevertheless, the predicted time series considerably under-predicts the number of fatalities over the prediction period. Some optimized parameters for this territory differ significantly from those obtained for Andalusia. In particular, the transmission rate for the Switzerland data is higher than that for the Andalusia data, as are the $H \rightarrow R$ recovery rates. The Switzerland parameters, however, for the death rate and the vaccine efficiency are predicted to be lower than those for Andalusia. In this case, we do not have a definitive attribution of the relevant result (i.e., the under-prediction of fatalities) in the case of Switzerland. The only change in policy that we could identify was that from September 13, 2021, access to most indoor public spaces like restaurants, bars, museums or fitness centers was permitted with a valid COVID certificate in Switzerland. No other restrictions were enforced on fully vaccinated and boosted people.

4. Conclusions and future challenges

In this work we presented a new class of compartmental epidemiological models that was motivated by the immunological properties of the delta and omicron variants of SARS-CoV-2. More generally, our aim was to present possibilities for settings where variants are highly transmissible (and hence relevant to consider vaccinated individuals and their epidemiological characteristics) as in the case of models A1–A2 for

the omicron variant, as well as ones where breakthrough infections are more rare, and hence vaccination is tantamount to withdrawal from the susceptible population as in the case of model B for the delta variant. Therefore, we constructed model B with the stipulation that vaccinated individuals were permanently withdrawn from the susceptible population based on the vaccination records and vaccine coverage rate. On the other hand, the epidemiology of the omicron variant suggests a substantial number of breakthrough infections, namely infections of vaccinated individuals. Accordingly, we developed models for both vaccinated and unvaccinated populations and analyzed their pairwise interaction and overall time evolution. Indeed, two classes of regression results were given. In the first (and more crude) regression, only the cumulative number of fatalities was accounted for in the optimization objective. This was done when the data did not allow the partitioning of fatalities (or hospitalizations) to vaccinated and unvaccinated components. In the second, more refined approach, fatalities (or hospitalizations) stemming from the two different (vaccinated or not) groups were separately considered. More broadly, our motivation is to contribute towards a modeling platform that can incorporate the specific features of past or/and future variants, and the available datasets (such as fatalities or/and hospitalizations), with the aim of a quantitative characterization of different components of the involved population (in line, also, with the corresponding identifiability analysis).

We addressed the identifiability of the various models and considered mathematical issues (e.g., parameters globally and locally identifiable, given particular time series), we raised some practical considerations due to the finite nature of the available observations, and we considered the compatibility of the selection of some initial conditions. In the case where the time series associated with vaccinated and unvaccinated individuals are required, we identified the issue of how to handle the so-called “unknown” deaths if the individual vaccination status remains undeclared. We proposed a concrete approach to address such disparities, yet clearly these topics merit further investigation.

In our presentation, we focused on the region of Andalusia in Spain and the country of Switzerland (which included data from the Principality of Liechtenstein). These two territories have similar populations. In each territory, we presented studies of a regression effort involving the fatalities (model A1), as well as one terminating at the compartment of total (i.e., conventional plus critical) hospitalizations (model A2). Our models gave generally good agreement with the corresponding training sets, but also reasonable predictions within the prediction interval of periods of about a month beyond the end of the training period (up to which the optimization is performed). In the cases where deviations from the predictions were more significant, plausible explanations were offered on the basis of, e.g., the relaxation of measures or other changes of policies.

Naturally, these models offer a starting point for further considerations and are intended as a stepping stone for further studies. On the one hand, it would be quite relevant to seek additional sources of data and other approaches to parameter estimation (than the regression and bootstrapping methodologies used here), to incorporate more accurately the measurement uncertainty and to improve the adequacy of the parameter estimation, in line with our expectations stemming from the analysis of the model identifiability. From a more mathematical perspective, it may also be quite relevant to seek to identify plausible vaccination function patterns $V(t)$. Then, one can use these to provide a systematic analysis of the evolution, stability features, and asymptotic behavior of the system. An especially appealing feature of the model in that regard is its mathematically non-autonomous nature which renders such an analysis a less explored, but quite worthwhile direction. Another important direction is to add the spatial dimension to the proposed well-mixed ODE models, to incorporate the mobility of vaccinated individuals. This can be done either at the level of metapopulation models [8,49,50] or at that of PDE approaches [51–54]. Finally, numerous additional dimensions of such modeling of vaccinations are relevant to consider such as, e.g., the age stratification of such effects [2,22,55]. These directions are currently under consideration and will be reported in future publications.

Table 3
Optimal parameters and initial conditions for the delta-variant model, model B, in Andalusia (third column) and Switzerland (fourth column). Model fits to the total number of fatalities, norm (10).

Parameter	Symbol	Median (interquartile range)	
		Andalusia Model B, norm (10)	Switzerland Model B, norm (10)
Transmission rate [per day]	β	0.4526 (0.4210–0.4703)	0.5445 (0.5320–0.5570)
Latent period [days]	$1/\sigma_1$	2.2001 (2.0896–2.2982)	2.2027 (2.1113–2.3021)
Preclinical period [days]	$1/\sigma_2$	2.1590 (2.0518–2.2789)	2.2056 (2.1112–2.3055)
A/P partitioning [-]	ϕ	0.0801 (0.0780–0.0827)	0.0800 (0.0787–0.0820)
Infectivity period (A) [days]	$1/\mu$	3.4493 (3.3551–3.5658)	3.4708 (3.3879–3.6194)
Recovery rate $I \rightarrow R$ [per day]	γ_r	0.1852 (0.1643–0.1982)	0.1888 (0.1842–0.1959)
Transition rate $I \rightarrow H$ [per day]	γ_h	0.0017 (0.0016–0.0020)	0.0026 (0.0024–0.0027)
Recovery rate $H \rightarrow R$ [per day]	κ_r	0.0629 (0.0523–0.0831)	0.2480 (0.2355–0.2637)
Death rate $H \rightarrow D$ [per day]	κ_d	0.0462 (0.0449–0.0481)	0.0069 (0.0066–0.0072)
Vaccine efficiency [-]	θ	0.7956 (0.7802–0.8240)	0.6162 (0.6047–0.6189)
Initial ratio [#]	$S(0)/N(0)$	0.5824 (0.5673–0.5951)	0.5278 (0.5178–0.5377)
Initial conditions			
Initial exposed population [#]	$E(0)$	1322	669
Initial presymptomatic population [#]	$P(0)$	351	591
Initial asymptomatic population [#]	$A(0)$	205	332
Initial symptomatically infected population [#]	$I(0)$	1914	1441
Initial hospitalized population [#]	$H(0)$	80	57

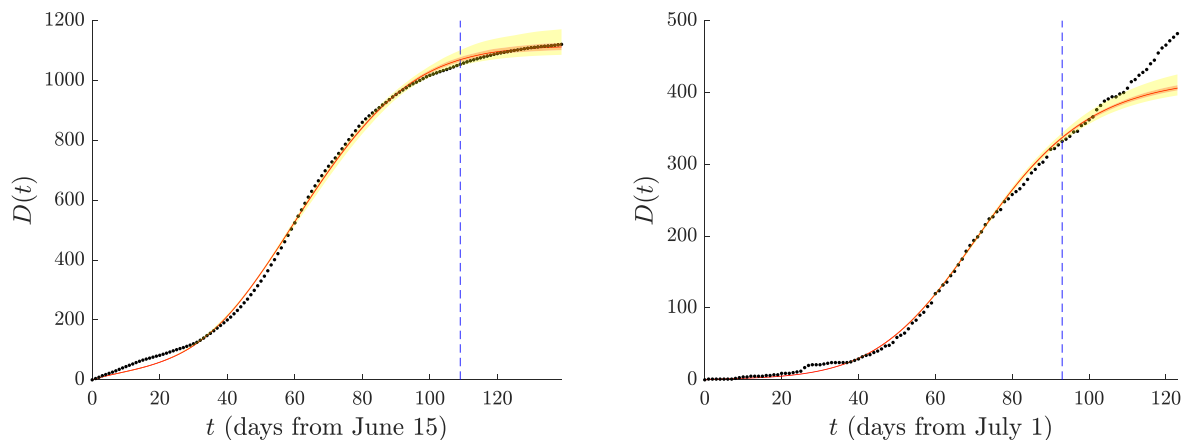


Fig. 10. Delta-variant model B: Fit and prediction of the total number of fatalities in Andalusia (left panel) and Switzerland (right panel). Norm (10) was used. The vertical line, the beginning of the prediction interval, is October 1, 2022. Note the significant difference in the number of fatalities. Line and symbol colors, as well as ranges, are as described in the previous figures.

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.

Acknowledgments

JC-M acknowledges support from EU (FEDER program 2014–2020) through MCIN/AEI/10.13039/501100011033 under the project PID2020-112620GB-I00. PGK and GAK acknowledge support through the C3.ai Inc. and Microsoft Corporation. The authors thank Gleb Pogudin for his invaluable help in the identifiability analysis work.

Appendix. Identifiability analysis

As defined in [38] (see also references therein), a parameter or an initial condition of a differential model is called **globally identifiable** if it can be uniquely determined from the input and output of the system. On the other hand, they are called **locally identifiable** if there is only a finite number of possible values. Due to many issues, such as parameter sensitivity, a globally identifiable parameter does *not necessarily* lead to a very sharp estimate, although it is guaranteed, in theory, to have a unique value. At the same time, only locally,

but not globally, identifiable parameters often have poor estimates. Nevertheless, as pointed out in [42] and elsewhere, the poor estimates of these parameters may still yield reasonable predictions of the state variables. Overall, identifiability analysis sheds important insights on the quality of the parameter estimation and the prediction of the state variables, and consequently it should be carried out before any parameter estimation is done.

Following the differential algebraic approach in [36], one needs to rewrite the whole system as a single high-order differential equation of the observable (i.e., the data). Inevitably, a moderate system leads to an equation with a huge number of terms which is very likely beyond the capability of symbolic mathematics software like Mathematica. The difficulty is due to the nonlinear terms in the original system. Our approach is to leave some original equation(s) untouched and rewrite the rest as a high-order differential equation of the observable. Basically, we explicitly carry out as many derivations as possible, and stop short of writing the original system into a single ODE, as the last step(s) may lead to an exceedingly complex equation. Then we apply the identifiability analysis package *SIAN* [38,39] and *StructuralIdentifiability* [40,41] to the new system. In what follows, we explain how it is done for the model given by Eqs. (1), model A1. The other two models, models A2 and B, are analyzed in the exactly same way. Note that this approach is not limited to the models presented in this work.

In addition, we remark that results could not be obtained for all the cases.

The subscript u or v will be dropped whenever there is no obvious ambiguity. We introduce intermediate parameters $\gamma_s = \gamma_r + \gamma_h$ and $\kappa_s = \kappa_r + \kappa_d$. The idea is to rewrite Eqs. (1) as a system of equations for the number of asymptomatics A and a high-order ODE for the number of fatalities D 's. The D 's need to be kept because they are the output (i.e., the observable).

For the unvaccinated state variables, Eqs. ((1i), (1g), (1f), (1c)) lead to

$$\begin{aligned} H &= \frac{1}{\kappa_d} D', \\ I &= \frac{1}{\gamma_h} (H' + \kappa_s H) = \frac{1}{\gamma_h \kappa_d} (D'' + \kappa_s D'), \\ P &= \frac{1}{\sigma_2} (I' + \gamma_s I) = \frac{1}{\sigma_2 \gamma_h \kappa_d} (D''' + D''(\kappa_s + \gamma_s) + D'(\gamma_s \kappa_s)), \\ E &= \frac{P' + \sigma_2 P}{(1 - \phi)\sigma_1} = \frac{1}{(1 - \phi)\sigma_1 \sigma_2 \gamma_h \kappa_d} \\ &\quad \times [D^{(4)} + D'''(\kappa_s + \gamma_s + \sigma_2) + D''(\kappa_s \gamma_s + \kappa_s \sigma_2 + \gamma_s \sigma_2) + D'(\kappa_s \gamma_s \sigma_2)]. \end{aligned}$$

Here all subscripts u are dropped: identical equations for vaccinated state variables should also be considered. From the last equation, we compute the following (since it resembles certain terms in Eq. (1b)),

$$\begin{aligned} E + \sigma_1 \int_{t_0}^t E(\tau) d\tau &= \frac{1}{(1 - \phi)\sigma_1 \sigma_2 \gamma_h \kappa_d} [D^{(4)} + k_3 D''' + k_2 D'' + k_1 D' + k_0 D + \tilde{\alpha}], \quad (A.1) \end{aligned}$$

where $\tilde{\alpha}$ is an integration constant (a new parameter) and

$$\begin{aligned} k_3 &= \kappa_s + \gamma_s + \sigma_1 + \sigma_2, \quad k_2 = \kappa_s \gamma_s + \kappa_s \sigma_1 + \kappa_s \sigma_2 + \gamma_s \sigma_1 + \gamma_s \sigma_2 + \sigma_1 \sigma_2, \\ k_1 &= \kappa_s \gamma_s \sigma_1 + \kappa_s \gamma_s \sigma_2 + \kappa_s \sigma_1 \sigma_2 + \gamma_s \sigma_1 \sigma_2, \quad k_0 = \kappa_s \gamma_s \sigma_1 \sigma_2. \end{aligned}$$

With Eq. (A.1), we now add Eqs. (1a) and (1b), and then integrate to obtain

$$\begin{aligned} (S' + E') &= -\sigma_1 E - \theta V(t) \Rightarrow \\ S &= \frac{-1}{(1 - \phi)\sigma_1 \sigma_2 \gamma_h \kappa_d} [D^{(4)} + k_3 D''' + k_2 D'' + k_1 D' + k_0 D + \alpha + \tilde{\theta} \tilde{V}(t)], \end{aligned}$$

where α is another integration constant (different from $\tilde{\alpha}$), and

$$\tilde{V}(t) = \int_{t_0}^t V(\tau) d\tau, \quad \tilde{\theta} = (1 - \phi)\sigma_1 \sigma_2 \gamma_h \kappa_d \theta.$$

So far, the state variables for unvaccinated population H, I, P, E, S are expressed as functions of D and its derivatives. We have identical formulas for the vaccinated populations H_v, I_v, P_v, E_v, S_v , except that for S_v a negative sign should be added in front of $\tilde{\theta}$.

The equation for E_u , Eq. (1b), multiplied by $(1 - \phi_u)\sigma_1 \sigma_2 \gamma_{h,u} \kappa_{d,u}$ becomes

$$\begin{aligned} D_u^{(5)} &= -[k_3 D_u^{(4)} + k_2 D_u''' + k_1 D_u'' + k_0 D_u'] \\ &\quad - [D_u^{(4)} + k_3 D_u''' + k_2 D_u'' + k_1 D_u' + k_0 D_u + \alpha_u + \tilde{\theta}_u V(t)] \\ &\quad \cdot [\beta^{uu}(I_u + A_u + P_u) + \beta^{uv}(I_v + A_v + P_v)]. \end{aligned}$$

Note that the subscript u for the parameters k_3, k_2, k_1, k_0 is still omitted. But for this equation they should be computed from the parameters associated with the unvaccinated population, whereas for the equation of $D_v^{(5)}$ they should be computed from the parameters associated with the vaccinated population.

Furthermore,

$$\begin{aligned} \beta^{uu}(I_u + A_u + P_u) &= \beta^{uu} \left[\frac{D_u'' + \kappa_{s,u} D_u'}{\gamma_{h,u} \kappa_{d,u}} + A_u + \frac{D_u''' + D_u''(\kappa_{s,u} + \gamma_{s,u}) + (\kappa_{s,u} \gamma_{s,u}) D_u'}{\sigma_2 \gamma_{h,u} \kappa_{d,u}} \right] \\ &= \frac{\beta^{uu} [D_u'' + \kappa_{s,u} D_u']}{\gamma_{h,u} \kappa_{d,u}} + \frac{\beta^{uu} [D_u''' + D_u''(\kappa_{s,u} + \gamma_{s,u}) + (\kappa_{s,u} \gamma_{s,u}) D_u']}{\sigma_2 \gamma_{h,u} \kappa_{d,u}} \\ &\quad + \beta^{uu} A_u. \end{aligned}$$

Similarly,

$$\begin{aligned} \beta^{uv}(I_v + A_v + P_v) &= \frac{\beta^{uv} [D_v'' + \kappa_{s,v} D_v']}{\gamma_{h,v} \kappa_{d,v}} \\ &\quad + \frac{\beta^{uv} [D_v''' + D_v''(\kappa_{s,v} + \gamma_{s,v}) + (\kappa_{s,v} \gamma_{s,v}) D_v']}{\sigma_2 \gamma_{h,v} \kappa_{d,v}} + \beta^{uv} A_v, \\ \beta^{vu}(I_u + A_u + P_u) &= \frac{\beta^{vu} [D_u'' + \kappa_{s,u} D_u']}{\gamma_{h,u} \kappa_{d,u}} \\ &\quad + \frac{\beta^{vu} [D_u''' + D_u''(\kappa_{s,u} + \gamma_{s,u}) + (\kappa_{s,u} \gamma_{s,u}) D_u']}{\sigma_2 \gamma_{h,u} \kappa_{d,u}} + \beta^{vu} A_u, \\ \beta^{vv}(I_v + A_v + P_v) &= \frac{\beta^{vv} [D_v'' + \kappa_{s,v} D_v']}{\gamma_{h,v} \kappa_{d,v}} \\ &\quad + \frac{\beta^{vv} [D_v''' + D_v''(\kappa_{s,v} + \gamma_{s,v}) + (\kappa_{s,v} \gamma_{s,v}) D_v']}{\sigma_2 \gamma_{h,v} \kappa_{d,v}} + \beta^{vv} A_v. \end{aligned}$$

The next step is to scale variables as follows

$$\begin{aligned} \gamma_{h,u} \kappa_{d,u} A_u \rightarrow A_u, \gamma_{h,v} \kappa_{d,v} A_v \rightarrow A_v, \quad \frac{\beta^{uu}}{\gamma_{h,u} \kappa_{d,u}} \rightarrow \beta^{uu}, \frac{\beta^{uv}}{\gamma_{h,v} \kappa_{d,v}} \\ \rightarrow \beta^{uv}, \frac{\beta^{vu}}{\gamma_{h,u} \kappa_{d,u}} \rightarrow \beta^{vu}, \frac{\beta^{vv}}{\gamma_{h,v} \kappa_{d,v}} \rightarrow \beta^{vv}. \quad (A.2) \end{aligned}$$

The above equation for $D_u^{(5)}$ becomes

$$\begin{aligned} D_u^{(5)} &= -[k_3 D_u^{(4)} + k_2 D_u''' + k_1 D_u'' + k_0 D_u'] \\ &\quad - [D_u^{(4)} + k_3 D_u''' + k_2 D_u'' + k_1 D_u' + k_0 D_u + \alpha_u + \tilde{\theta}_u V(t)] \\ &\quad \times \left(\beta^{uu} [D_u'' + \kappa_{s,u} D_u'] \right. \\ &\quad + \frac{\beta^{uu}}{\sigma_2} [D_u''' + D_u''(\kappa_{s,u} + \gamma_{s,u}) + (\kappa_{s,u} \gamma_{s,u}) D_u'] + \beta^{uu} A_u \\ &\quad + \beta^{uv} [D_v'' + \kappa_{s,v} D_v'] \\ &\quad \left. + \frac{\beta^{uv}}{\sigma_2} [D_v''' + D_v''(\kappa_{s,v} + \gamma_{s,v}) + (\kappa_{s,v} \gamma_{s,v}) D_v'] + \beta^{uv} A_v \right). \end{aligned}$$

The equation for A_u , Eq. (1d), takes the form

$$\begin{aligned} A_u' &= \frac{\phi_u}{(1 - \phi_u)\sigma_2} [D_u^{(4)} + D_u'''(\kappa_{s,u} + \gamma_{s,u} + \sigma_2) \\ &\quad + D_u''(\kappa_{s,u} \gamma_{s,u} + \kappa_{s,u} \sigma_2 + \gamma_{s,u} \sigma_2) + D_u'(\kappa_{s,u} \gamma_{s,u} \sigma_2)] - \mu_u A_u. \end{aligned}$$

Similarly, the equation of A_v and the high order equation for D_v are:

$$\begin{aligned} D_v^{(5)} &= -[k_3 D_v^{(4)} + k_2 D_v''' + k_1 D_v'' + k_0 D_v'] \\ &\quad - [D_v^{(4)} + k_3 D_v''' + k_2 D_v'' + k_1 D_v' + k_0 D_v + \alpha_v + \tilde{\theta}_v V(t)] \\ &\quad \times \left(\beta^{vu} [D_u'' + \kappa_{s,u} D_u'] \right. \\ &\quad + \frac{\beta^{vu}}{\sigma_2} [D_u''' + D_u''(\kappa_{s,u} + \gamma_{s,u}) + (\kappa_{s,u} \gamma_{s,u}) D_u'] + \beta^{vu} A_u \\ &\quad + \beta^{vv} [D_v'' + \kappa_{s,v} D_v'] \\ &\quad \left. + \frac{\beta^{vv}}{\sigma_2} [D_v''' + D_v''(\kappa_{s,v} + \gamma_{s,v}) + (\kappa_{s,v} \gamma_{s,v}) D_v'] + \beta^{vv} A_v \right), \end{aligned}$$

$$\begin{aligned} A_v' &= \frac{\phi_v}{(1 - \phi_v)\sigma_2} [D_v^{(4)} + D_v'''(\kappa_{s,v} + \gamma_{s,v} + \sigma_2) \\ &\quad + D_v''(\kappa_{s,v} \gamma_{s,v} + \kappa_{s,v} \sigma_2 + \gamma_{s,v} \sigma_2) + D_v'(\kappa_{s,v} \gamma_{s,v} \sigma_2)] - \mu_v A_v. \end{aligned}$$

Up to now, we rewrote the original system as a system of $D_u^{(5)}, D_v^{(5)}, A_u, A_v$. It can be written as a first-order system (by using $D, D', D'', D^{(3)}, D^{(4)}$) so that the identifiability analysis package *StructuralIdentifiability* can be applied to find the identifiability property of the 18 parameters of this new system:

$$\beta^{uu}, \beta^{uv}, \beta^{vu}, \beta^{vv}, \sigma_1, \sigma_2, \phi_u, \phi_v, \gamma_{s,u}, \gamma_{s,v}, \mu_u, \mu_v, \kappa_{s,u}, \kappa_{s,v}, \tilde{\theta}_u, \tilde{\theta}_v, \alpha_u, \alpha_v$$

Then, the identifiability property of the 19 parameters of the original system

$$\beta^{uu}, \beta^{uv}, \beta^{vu}, \beta^{vv}, \sigma_1, \sigma_2, \theta, \phi_u, \phi_v, \mu_u, \mu_v, \gamma_{r,u}, \gamma_{r,v}, \gamma_{h,u}, \gamma_{h,v}, \kappa_{r,u}, \kappa_{r,v}, \kappa_{d,u}, \kappa_{d,v}$$

can be derived. Furthermore, one may use the *SIAN Webapp*, with the globally identifiable parameters (from *StructuralIdentifiability*) as extra outputs, to find the identifiability property of the initial conditions. Identifiability results obtained from these calculations are reported in appropriate sections in the main text.

References

- [1] W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics – I, *Bull. Math. Biol.* 53 (1991) 33–55, <http://dx.doi.org/10.1007/BF02464423>.
- [2] H.W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.* 42 (2000) 599–653, <http://dx.doi.org/10.1137/S0036144500371907>.
- [3] F. Brauer, C. Castillo-Chávez, *Mathematical Models in Population Biology and Epidemiology*, Springer-Verlag, 2012.
- [4] D. Chen, Modeling the spread of infectious diseases: A review, in: D. Chen, B. Moulin, J. Wu (Eds.), *Analyzing and Modeling Spatial and Temporal Dynamics of Infectious Diseases*, John Wiley & Sons, 2014, p. 19.
- [5] The COVID-19 forecast hub. <https://covid19forecasthub.org>.
- [6] L. Cao, Q. Liu, COVID-19 Modeling: A Review, <http://dx.doi.org/10.1101/2022.08.22.22279022>, <https://arxiv.org/abs/2104.12556>.
- [7] S.M. Shakeel, N.S. Kumar, P.P. Madalli, R. Srinivasiah, D.R. Swamy, COVID-19 prediction models: A systematic literature review, *Osong Public Health Res. Perspect.* 12 (2021) 215–229, <http://dx.doi.org/10.24171/j.phrp.2021.0100>.
- [8] D. Calvetti, A.P. Hoover, J. Rose, E. Somersalo, Metapopulation network models for understanding, predicting, and managing the coronavirus disease COVID-19, *Front. Phys.* 8 (2020) 261, <http://dx.doi.org/10.3389/fphy.2020.00261>.
- [9] A.L. Bertozzi, E. Franco, G. Mohler, M.B. Short, D. Sledge, The challenges of modeling and forecasting the spread of COVID-19, *P. Natl. Acad. Sci.* 117 (2020) 16732–16738, <https://www.pnas.org/doi/abs/10.1073/pnas.2006520117>.
- [10] I. Holmdahl, C. Buckee, Wrong but useful- what COVID-19 epidemiological models can and cannot tell us, *New Engl. J. Med.* 383 (2020) 303–305, <http://dx.doi.org/10.1056/NEJMp2016822>.
- [11] K. Hausken, M. Mcube, Decisions of persons, the pharmaceutical industry, and donors in disease contraction and recovery assuming virus mutation, *Health Econ. Rev.* 11 (2021) 1–14, <http://dx.doi.org/10.1186/s13561-021-00320-4>.
- [12] K. Hausken, M. Mcube, A game theoretic analysis of competition between vaccine and drug companies during disease contraction and recovery, *Med. Decis. Mak.* 42 (2022) 571–586, <http://dx.doi.org/10.1177/0272989X211053563>.
- [13] J.S. Tregoning, K.E. Flight, S.L. Higham, Z. Wang, B.F. Pierce, Progress of the COVID-19 vaccine effort: Viruses, vaccines and variants versus efficacy, effectiveness and escape, *Nat. Rev. Immunol.* 21 (2021) 626, <http://dx.doi.org/10.1038/s41577-021-00592-1-636>.
- [14] COVID-19 vaccines. <https://www.uptodate.com/contents/covid-19-vaccines>.
- [15] C.E. Wagner, C.M. Saad-Roy, B.T. Grenfell, Modelling vaccination strategies for COVID-19, *Nat. Rev. Immunol.* 22 (2022) 139, <http://dx.doi.org/10.1038/s41577-022-00687-3141>.
- [16] G. Angelov, R. Kovacevic, N.I. Stilianakis, V.M. Veliov, Optimal vaccination strategies using a distributed model applied to COVID-19, *Cent. Eur. J. Oper. Res.* 31 (2022) 499–521, <http://dx.doi.org/10.1007/s10100-022-00819-z>.
- [17] T.T. Marinov, R.S. Marinova, Adaptive SIR model with vaccination: Simultaneous identification of rates and functions illustrated with COVID-19, *Sci. Rep.* 12 (2022) 15688, <http://dx.doi.org/10.1038/s41598-022-20276-7>.
- [18] T. Usherwood, Z. LaJoie, V. Srivastava, A model and predictions for COVID-19 considering population behavior and vaccination, *Sci. Rep.* 11 (2021) 12051, <http://dx.doi.org/10.1038/s41598-021-91514-7>.
- [19] C.R. MacIntyre, V. Costantino, M. Trent, Modelling of COVID-19 vaccination strategies and herd immunity, in scenarios of limited and full vaccine supply in NSW, *Australia Vaccine* 40 (2022) 2506–2513, <http://dx.doi.org/10.1016/j.vaccine.2021.04.042>.
- [20] R. Kahn, I. Holmdahl, S. Reddy, J. Jernigan, M.J. Mina, R.B. Slayton, Mathematical modeling to inform vaccination strategies and testing approaches for coronavirus disease 2019 (COVID-19) in nursing homes, *Clin. Infect. Dis.* 74 (2021) 597–603, <http://dx.doi.org/10.1093/cid/ciab517>.
- [21] J. Rychtar, M.L. Diagne, H. Rwezaura, S.Y. Tchoumi, J.M. Tchuente, A mathematical model of COVID-19 with vaccination and treatment, *Comput. Math. Methods Med.* (2021) 1250129, <http://dx.doi.org/10.1155/2021/1250129>.
- [22] J. Cuevas-Maraver, P.G. Kevrekidis, Q.Y. Chen, G.A. Kevrekidis, Z. Rapti, Y. Drossinos, Lockdown measures and their impact on single- and two-age-structured epidemic model for the COVID-19 outbreak in Mexico, *Math. Biosci.* 336 (2021) 108590, <http://dx.doi.org/10.1016/j.mbs.2021.108590>.
- [23] J.F. Moxnse, K. Hausken, Mathematical modelling of acute virus influenza a infections, *Math. Comput. Model. Dyn. Syst.* 18 (2022) 521–538, <http://dx.doi.org/10.1080/13873954.2012.669387>.
- [24] Q. Ma, J. Liu, Q. Liu, L. Kang, R. Liu, W. Jing, Y. Wu, M. Liu, Global percentage of asymptomatic SARS-CoV-2 infections among the tested population and individuals with confirmed COVID-19 diagnosis: A systematic review and meta-analysis, *JAMA Netw. Open* 4 (2021) e2137257, <http://dx.doi.org/10.1001/jamanetworkopen.2021.37257>.
- [25] P. Wu, et al., Assessing asymptomatic, presymptomatic, and symptomatic transmission risk of severe acute respiratory syndrome Coronavirus 2, *Clin. Infect. Dis.* 73 (6) (2021) e1314–e1320, <http://dx.doi.org/10.1093/cid/ciab271>.
- [26] X. Hao, S. Cheng, D. Wu, et al., Reconstruction of the full transmission dynamics of COVID-19 in Wuhan, *Nature* 584 (2020) 420–424, <http://dx.doi.org/10.1038/s41586-020-2554-8>.
- [27] C. Savvides, R. Siegel, Asymptomatic and presymptomatic transmission of SARS-CoV-2: A systematic review, 2020, <http://dx.doi.org/10.1101/2020.06.11.20129072>, medRxiv [Preprint].
- [28] S.E. Waldman, T. Buehring, D.J. Escobar, S.K. Gohil, R. Gonzales, S.S. Huang, K. Olenzler, K.K. Prabaker, T. Sandoval, J. Yim, D.S. Yokoe, S.H. Cohen, Secondary cases of delta-variant COVID-19 among vaccinated healthcare workers with breakthrough infections is rare, *Clin. Infect. Dis.* 75 (2021) e895, <https://researcherprofiles.org/profile/238157167>.
- [29] Y. Wu, L. Kang, Z. Guo, J. Liu, M. Liu, W. Liang, Incubation period of COVID-19 caused by unique SARS-CoV-2 strains: A systematic review and meta-analysis, *JAMA Netw. Open* 5 (2022) e2228008, <http://dx.doi.org/10.1001/jamanetworkopen.2022.28008>.
- [30] T.A. Slater, S. Straw, M. Drodz, S. Kamalathanan, A. Cowley, K.K. Witte, Dying ‘due to’ or ‘with’ COVID-19: A cause of death analysis in hospitalised patients, *Clin. Med.* 20 (2020) e189, <http://dx.doi.org/10.7861/clinmed.2020-0440>.
- [31] Los angeles times, 2021, <https://www.latimes.com/world-nation/story/2021-11-12/europe-germany-covid-hospot>.
- [32] The guardian, 2021, <https://www.theguardian.com/world/2021/nov/30/omicron-covid-variant-present-in-europe-at-least-10-days-ago>.
- [33] World Health organization, Classification of omicron (B.1.1.529): SARS-CoV-2 variant of concern, [https://www.who.int/news/item/26-11-2021-classification-of-omicron-\(b.1.1.529\)-sars-cov-2-variant-of-concern](https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern).
- [34] B. Reiner, COVID-19 model update: Omicron and waning immunity, 2021, <http://www.healthdata.org/special-analysis/omicron-and-waning-immunity>.
- [35] G. Chowell, Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts, *Infect. Dis. Model.* 2 (2017) 379–398, <http://dx.doi.org/10.1016/j.idm.2017.08.001>.
- [36] J.H. Tien, M.C. Eisenberg, S.L. Robertson, Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease, *J. Theor. Biol.* 324 (2013) 84, <http://dx.doi.org/10.1016/j.jtbi.2012.12.021>.
- [37] H. Hu, C.M. Kennedy, P.G. Kevrekidis, H.-K. Zhang, A modified PINN approach for identifiable compartmental models in epidemiology with application to COVID-19, *Viruses* 14 (2022) 2464, <http://dx.doi.org/10.3390/v14112464>.
- [38] H. Hong, A. Ovchinnikov, G. Pogudin, C. Yap, Global identifiability of differential models, *Comm. Pure Appl. Math.* 73 (9) (2020) 1831–1879.
- [39] H. Hong, A. Ovchinnikov, G. Pogudin, C. Yap, SIAN: Software for structural identifiability analysis of ODE models, *Bioinformatics* 35 (2019) 2873, <http://dx.doi.org/10.1002/cpa.21921>.
- [40] A. Ovchinnikov, A. Pillay, G. Pogudin, T. Scanlon, Computing all identifiable functions of parameters for ODE models, *Syst. Control Lett.* 157 (2021) 105030, <http://dx.doi.org/10.1016/j.sysconle.2021.105030>.
- [41] R. Dong, C. Goodbrake, H. Harrington, G. Pogudin, Differential elimination for dynamical models via projections with applications to structural identifiability, *SIAM J. Appl. Algebra Geom.* (2023) <http://dx.doi.org/10.1137/22M1469067>.
- [42] R.N. Gutenkunst, J.J. Waterfall, F.P. Casey, K.S. Brown, C.R. Myers, J.P. Sethna, Universally sloppy parameter sensitivities in systems biology models, *PLoS Comput. Biol.* 3 (2007) e189, <http://dx.doi.org/10.1371/journal.pcbi.0030189>.
- [43] https://github.com/montera34/escovid19data/blob/master/data/output/covid19-ccaa-spain_consolidated.csv.
- [44] The data used for this analysis can be found at the link https://personal.us.es/jcuevas/covid/Vaccines_Andalusia.xlsx. While these data were originally retrieved from the official website <https://www.juntadeandalucia.es/institutodeestadisticaycartografia/salud/vacunas-COVID19.html>, the relevant dataset has been modified therein and appears to no longer be available in the form of the dataset used herein.
- [45] W. Yu, Y. Guo, S. Zhang, Y. Kong, Z. Shen, J. Zhang, Proportion of asymptomatic infection and nonsevere disease caused by SARS-CoV-2 omicron variant: A systematic review and analysis, *J. Med. Virol.* 94 (2022) 5790, <http://dx.doi.org/10.1002/jmv.28066>.
- [46] <https://opendata.swiss/en/dataset/covid-19-schweiz>.
- [47] Vozpopuli, Al menos 8 CCAA inflan sus cifras reales de muertos por COVID-19, <https://www.vozpopuli.com/sanidad/cifras-muertos-covid.html>.
- [48] N. Kumar, S. Quadri, A.I. AlAwadhi, M. AlQahtani, COVID-19 recovery patterns across Alpha (B.1.1.7) and Delta (B.1.617.2) variants of SARS-CoV-2, *Front. Immunol.* 3 (2022) 812606, <http://dx.doi.org/10.3389/fimmu.2022.812606>.
- [49] V. Colizza, A. Vespignani, Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations, *J. Theor. Biol.* 251 (2008) 450, <http://dx.doi.org/10.1016/j.jtbi.2007.11.028>.
- [50] Z. Rapti, J. Cuevas-Maraver, E. Kontou, S. Liu, Y. Drossinos, P.G. Kevrekidis, G.A. Kevrekidis, M. Barmann, Q.-Y. Chen, The role of mobility in the dynamics of the COVID-19 epidemic in Andalusia, *Bull. Math. Biol.* 85 (2023) 54, <http://dx.doi.org/10.1007/s11538-023-01152-5>.

- [51] Y. Mammeri, A reaction–diffusion system to better comprehend the unlockdown: Application of SEIR-type model with diffusion to the spatial spread of COVID-19 in France, *Comput. Math. Biophys.* 8 (2020) 102, <http://dx.doi.org/10.1515/cmb-2020-0104>.
- [52] A. Viguerie, G. Lorenzo, F. Auricchio, D. Baroli, T.J.R. Hughes, A. Patton, A. Reali, T.E. Yankeelov, A. Veneziani, Simulating the spread of COVID-19 via a spatially-resolved susceptible–exposed–infected–recovered–deceased (SEIRD) model with heterogeneous diffusion, *Appl. Math. Lett.* 111 (2021) 106617, <http://dx.doi.org/10.1016/j.aml.2020.106617>.
- [53] P.G. Kevrekidis, J. Cuevas-Maraver, Y. Drossinos, Z. Rapti, G.A. Kevrekidis, Reaction–diffusion spatial modeling of COVID-19: Greece and Andalusia as case examples, *Phys. Rev. E* 104 (2021) 024412, <http://dx.doi.org/10.1103/PhysRevE.104.024412>.
- [54] A. Vaziry, T. Kolokolnikov, P.G. Kevrekidis, Modelling of spatial infection spread through heterogeneous population: From lattice to partial differential equation models, *R. Soc. Open Sci.* 9 (2022) 220064, <http://dx.doi.org/10.1098/rsos.220064>.
- [55] V. Ram, L.P. Schaposnik, A modified age-structured SIR model for COVID-19 type viruses, *Sci. Rep.* 11 (2021) 15194, <http://dx.doi.org/10.1038/s41598-021-94609-3>.