



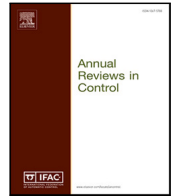
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

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



Contents lists available at ScienceDirect

Annual Reviews in Control

journal homepage: www.elsevier.com/locate/arcontrol

Tutorial Article

Structural identifiability and observability of compartmental models of the COVID-19 pandemic[☆]Gemma Massonis, Julio R. Banga, Alejandro F. Villaverde^{*}*BioProcess Engineering Group, IIM-CSIC, Vigo 36208, Galicia, Spain*

ARTICLE INFO

Keywords:

Identifiability
 Observability
 Dynamic modelling
 Epidemiology
 COVID-19

ABSTRACT

The recent coronavirus disease (COVID-19) outbreak has dramatically increased the public awareness and appreciation of the utility of dynamic models. At the same time, the dissemination of contradictory model predictions has highlighted their limitations. If some parameters and/or state variables of a model cannot be determined from output measurements, its ability to yield correct insights – as well as the possibility of controlling the system – may be compromised. Epidemic dynamics are commonly analysed using compartmental models, and many variations of such models have been used for analysing and predicting the evolution of the COVID-19 pandemic. In this paper we survey the different models proposed in the literature, assembling a list of 36 model structures and assessing their ability to provide reliable information. We address the problem using the control theoretic concepts of structural identifiability and observability. Since some parameters can vary during the course of an epidemic, we consider both the constant and time-varying parameter assumptions. We analyse the structural identifiability and observability of all of the models, considering all plausible choices of outputs and time-varying parameters, which leads us to analyse 255 different model versions. We classify the models according to their structural identifiability and observability under the different assumptions and discuss the implications of the results. We also illustrate with an example several alternative ways of remedying the lack of observability of a model. Our analyses provide guidelines for choosing the most informative model for each purpose, taking into account the available knowledge and measurements.

1. Introduction

The current coronavirus disease (COVID-19) pandemic, caused by the SARS-CoV-2 virus, continues to wreak unparalleled havoc across the world. Public health authorities can use mathematical models to answer critical questions related with the dynamics of an epidemic (severity and time course of infected people), its impact on the healthcare system, and the design and effectiveness of different interventions (Adam, 2020; Currie et al., 2020; Li, 2018; Lofgren et al., 2014). Mathematical modelling of infectious diseases has a long history (Brauer, van den Driessche, & Wu, 2008; Martcheva, 2015). Modelling efforts are particularly important in the context of COVID-19 because its dynamics can be particularly complex and counter-intuitive due to the uncertainty in the transmission mechanisms, possible seasonal variation in both susceptibility and transmission, and their variation within subpopulations (Cobey, 2020). The media has given extensive coverage to analyses and forecasts using COVID-19 models,

with increased attention to cases of conflicting conclusions, giving the impression that epidemiological models are unreliable or flawed. However, a closer look reveals that these modelling studies were following different approaches, handling uncertainty differently, and ultimately addressing different questions on different time-scales (Holmdahl & Buckee, 2020).

Broadly speaking, data-driven models (using statistical regression or machine learning) can be used for short-term forecasts (one or a few weeks). Mechanistic models based on assumptions about transmission and immunity try to mimic how the virus spreads, and can be used to formalize current knowledge and explore long-term outcomes of the pandemic and the effectiveness of different interventions. However, the accuracy of mechanistic models is constrained by the uncertainties in our knowledge, which creates uncertainties in model parameters and even in the model structure (Holmdahl & Buckee, 2020). Further, the uncertainty in the COVID-19 data and the exponential spread

[☆] This research has received funding from the Spanish Ministry of Science, Innovation and Universities and the European Union FEDER under project grant SYNBIOTROL (DPI2017-82896-C2-2-R) and the CSIC, Spain intramural project grant MOEBIUS (PIE 202070E062). The funding bodies played no role in the design of the study, the collection and analysis of the data or in the writing of the manuscript.

^{*} Corresponding author.

E-mail addresses: gmassonis@iim.csic.es (G. Massonis), julio@iim.csic.es (J.R. Banga), avillaverde@iim.csic.es (A.F. Villaverde).

<https://doi.org/10.1016/j.arcontrol.2020.12.001>

Received 26 June 2020; Received in revised form 24 September 2020; Accepted 1 December 2020

Available online 21 December 2020

1367-5788/© 2020 Elsevier Ltd. All rights reserved.

of the virus amplify the uncertainty in the predictions. Predictability studies (Scarpino & Petri, 2019) seek the characterization of the fundamental limits to outbreak prediction and their impact on decision-making. Despite the vast literature on mathematical epidemiology in general, and modelling of COVID-19 in particular, comparatively few authors have considered the predictability of infectious disease outbreaks (Castro, Ares, Cuesta, & Manrubia, 2020; Scarpino & Petri, 2019). Uncertainty quantification (Arriola & Hyman, 2009) is an interconnected concept that is also key for the reliability of a model, and that has received similarly scant attention (Faranda et al., 2020; Raimúndez et al., 2020).

In addition to predictability and uncertainty quantification approaches, identifiability is a related property whose absence can severely limit the usefulness of a mechanistic model (Chowell, 2017). A model is identifiable if we can determine the values of its parameters from knowledge of its inputs and outputs. Likewise, the related control theoretic property of observability describes if we can infer the model states from knowledge of its inputs and outputs. If a model is non-identifiable (or non-observable) different sets of parameters (or states) can produce the same predictions or fit to data. The implications can be enormous: in the context of the COVID-19 outbreak in Wuhan, non-identifiability in model calibrations was identified as the main reason for wide variations in model predictions (Roda, Varughese, Han, & Li, 2020).

Reliable models can be used in combination with optimization and optimal control methods to find the best intervention strategies, such as lock-downs with minimum economic impact (Acemoglu, Chernozhukov, Werning, & Whinston, 2020; Alvarez, Argente, & Lippi, 2020). Further, they can be used to explore the feasibility of model-based real-time control of the pandemic (Casella, 2020; Köhler, Schwenkel, Koch, Berberich, Pauli, & Allgöwer, 2020). However, using calibrated models with non-identifiability or non-observability issues can result in bad or even dangerous intervention and control strategies.

It is common to distinguish between structural and practical identifiability. Structural non-identifiability may be due to the model and measurement (input–output) structure. Practical non-identifiability is due to lack of information in the considered data-sets. Non-identifiability results in incorrect parameter estimates and bad uncertainty quantification (Chowell, 2017; Kao & Eisenberg, 2018), i.e. a misleading calibrated model which should not be used to analyse epidemiological data, test hypothesis, or design interventions. The structural identifiability of several epidemic mechanistic models has been studied e.g. in Brouwer, Meza, and Eisenberg (2019), Chapman and Evans (2009), Eisenberg, Robertson, and Tien (2013), Evans, Chapman, Chappell, and Godfrey (2002), Evans, White, Chapman, Godfrey, and Chappell (2005), Prague, Wittkop, Clairon, Dutartre, Thiebaut, and Hejblum (2020). Other recent studies have mostly focused on practical identifiability, such as Alahmadi et al. (2020), Chowell (2017), Kao and Eisenberg (2018), Roosa and Chowell (2019), Tuncer and Le (2018).

Since checking structural identifiability can be hard, one might think that directly checking practical identifiability could be a better alternative. However, it has been shown Janzén et al. (2016) that this practice entails at least two dangers: (i) checking practical identifiability can exhibit numerical issues that lead to wrong results, and (ii) if structural identifiability is unknown, the source of uncertainty in the parameter estimates can be either the data or the model structure, so it will not be clear how to improve the situation. In contrast, if we can verify structural identifiability, we can then focus on the data uncertainty to resolve possible practical identifiability issues. In other words, practical identifiability analysis cannot be used to deduce structural identifiability, and ignoring the latter can lead to modelling artefacts.

In this paper we assess the structural identifiability and observability of a large set of COVID-19 mechanistic models described by deterministic ordinary differential equations, derived by different authors using the compartmental modelling framework (Brauer, 2008).

Compartmental models are widely used in epidemiology because they are tractable and powerful despite their simplicity. We collect 36 different compartmental models, of which we consider several variations, making up a total of 255 different model versions. Our aim is to characterize their ability to provide insights about their unknown parameters – i.e. their structural identifiability – and unmeasured states – i.e. their observability. To this end we adopt a differential geometry approach that considers structural identifiability as a particular case of nonlinear observability, allowing to analyse both properties jointly. We define the relevant concepts and describe the methods used in Section 2. Then we provide an overview of the different types of compartmental models found in the literature in Section 3. We analyse their structural identifiability and observability and discuss the results in Section 4, where we also show different ways of remedying lack of observability using an illustrative model. Finally, we conclude our study with some key remarks in Section 5.

2. Methods

2.1. Notation, models, and properties

We consider models defined by systems of ordinary differential equations with the following notation:

$$\mathcal{M} = \begin{cases} \dot{x}(t) = f(x(t), \theta, u(t), w(t)) & (a) \\ y(t) = h(x(t), \theta, u(t), w(t)) & (b) \end{cases} \quad (1)$$

where f and h are analytical (generally nonlinear) functions of the states $x(t) \in \mathbb{R}^n$, known inputs $u(t) \in \mathbb{R}^u$, unknown constant parameters $\theta \in \mathbb{R}^n_\theta$, and unknown inputs or time-varying parameters $w(t) \in \mathbb{R}^w$. The output $y(t) \in \mathbb{R}^y$ represents the measurable functions of model variables. The expressions (1)a–(1)b are sufficiently general to represent a wide range of model structures, of which compartmental models are a particular case.

Definition 1 (*Structurally Locally Identifiable* (DiStefano III, 2015)). A parameter θ_i of model \mathcal{M} is structurally locally identifiable (s.l.i.) if for almost any parameter vector $\theta^* \in \mathbb{R}^n_\theta$ there is a neighbourhood $\mathcal{N}(\theta^*)$ in which the following relationship holds:

$$\hat{\theta} \in \mathcal{N}(\theta^*) \text{ and } y(t, \hat{\theta}) = y(t, \theta^*) \Rightarrow \hat{\theta}_i = \theta_i^* \quad (2)$$

Otherwise, θ_i is *structurally unidentifiable* (s.u.). If all model parameters are s.l.i. the model is s.l.i. If there is at least one s.u. parameter, the model is s.u.

Likewise, a state $x_i(\tau)$ is *observable* if it can be distinguished from any other states in a neighbourhood from observations of the model output $y(t)$ and input $u(t)$ in the interval $t_0 \leq \tau \leq t_f$, for a finite t_f . Otherwise, $x_i(\tau)$ is *unobservable*. A model is called observable if all its states are observable. We also say that \mathcal{M} is *invertible* if it is possible to infer its unknown inputs $w(t)$, and we say that $w(t)$ is reconstructible in this case.

Structural identifiability can be seen as a particular case of observability (Sedoglavic, 2002; Tunali & Tarn, 1987; Villaverde, 2019), by augmenting the state vector with the unknown parameters θ , which are now considered as state variables with zero dynamics, $\tilde{x} = (x^T, \theta^T)^T$. The reconstructibility of unknown inputs $w(t)$, which is also known as input observability, can also be cast in a similar way, although in this case their derivatives may be nonzero. To this end, let us augment the state vector further with w as additional states, as well as their derivatives up to some non-negative integer l :

$$\tilde{x} = \begin{pmatrix} x^T & \theta^T & w^T & \dots & w^{(l)T} \end{pmatrix}^T, \quad (3)$$

The l -augmented dynamics is:

$$\begin{aligned} \dot{\tilde{x}}(t) &= f^l(x(t), u(t), w^{(l+1)}(t)) \\ &= \begin{pmatrix} f(x(t), u(t))^T & 0_{1 \times n_p} & w(t)^T & \dots & w^{(l+1)}(t)^T \end{pmatrix}^T, \end{aligned}$$

leading to the l -augmented system:

$$\mathcal{M}^l \begin{cases} \dot{\tilde{x}}(t) = f^l(x(t), u(t), w^{(l+1)}(t)) \\ y(t) = h(x(t), u(t)) \end{cases} \quad (4)$$

Remark 1 (*Unknown Inputs, Disturbances, or Time-Varying Parameters*). In Section 4, when reporting the results of the structural identifiability and observability analyses, we will explicitly consider some parameters as time-varying. In the model structure defined in Eqs. (1)a–(1)b the unknown parameter vector θ is assumed to be constant. To consider an unknown parameter as time-varying we include it in the “unknown input” vector $w(t)$. Thus, changing the consideration of a parameter from constant to time-varying entails removing it from θ and including it in $w(t)$. The elements of $w(t)$ can be interpreted as unmeasured disturbances or inputs of unknown magnitude or, equivalently, as time-varying parameters. Regardless of the interpretation, they are assumed to change smoothly, i.e. they are infinitely differentiable functions of time. For the analysis of some models it is necessary, or at least convenient, to introduce the mild assumption that the derivatives of $w(t)$ vanish for a certain non-negative integer s (possibly $s = +\infty$), i.e. $w^{(s)}(t) \neq 0$ and $w^{(i)}(t) = 0$ for all $i > s$. This assumption is equivalent to assuming that the disturbances are polynomial functions of time, with maximum degree equal to s (Villaverde, Tsiantis, & Banga, 2019).

Definition 2 (*Full Input-State-Parameter Observability, FISPO* (Villaverde et al., 2019)). Let us consider a model \mathcal{M} given by (1)a–(1)b. We augment its state vector as $z(t) = (x(t)^T \quad \theta^T \quad w(t)^T)^T$ (3), which leads to its augmented form (4). We say that \mathcal{M} has the FISPO property if, for every $t_0 \in I$, every model unknown $z_i(t_0)$ can be inferred from $y(t)$ and $u(t)$ in a finite time interval $[t_0, t_f] \subset I$. Thus, \mathcal{M} is FISPO if, for every $z(t_0)$ and for almost any vector $z^*(t_0)$, there is a neighbourhood $\mathcal{N}(z^*(t_0))$ such that, for all $\hat{z}(t_0) \in \mathcal{N}(z^*(t_0))$, the following property is fulfilled:

$$y(t, \hat{z}(t_0)) = y(t, z^*(t_0)) \Rightarrow \hat{z}_i(t_0) = z_i^*(t_0), \quad 1 \leq i \leq n_x + n_\theta + n_w.$$

2.2. Structural identifiability and observability analysis

In this paper we analyse input, state, and parameter observability – that is, the FISPO property defined above – using a differential geometry framework. Such analyses are *structural* and *local*. By structural we refer to properties that are entirely determined by the model equations; thus we do not consider possible deficiencies due to insufficient or noise-corrupted data. By local we refer to the ability to distinguish between neighbouring states (similarly, parameters or unmeasured inputs), even though they may not be distinguishable from other distant states. This is usually sufficient, since in most (although not all, see e.g. Thomaseth and Saccomani (2018)) applications local observability entails global observability. This specific type of observability has sometimes been called local weak observability (Hermann & Krener, 1977).

This approach assesses structural identifiability and observability by calculating the rank of a matrix that is constructed with Lie derivatives. The corresponding definitions are as follows (in the remainder of this section we omit the dependency on time to simplify the notation):

Definition 3 (*Extended Lie Derivative* (Karlsson, Anguelova, & Jirstrand, 2012)). Consider the system \mathcal{M} (1)a–(1)b with augmented state vector (3) and augmented dynamics (4). Assuming that the inputs u are analytical functions, the extended Lie derivative of the output along $\tilde{f} = \tilde{f}(\cdot, u)$ is:

$$L_{\tilde{f}}^h(\tilde{x}, u) = \frac{\partial h}{\partial \tilde{x}}(\tilde{x}, u) \tilde{f}(\tilde{x}, u) + \frac{\partial h}{\partial u}(\tilde{x}, u) \dot{u}.$$

The zero-order derivative is $L_{\tilde{f}}^0 h = h$, and the i -order extended Lie derivatives can be recursively calculated as:

$$L_{\tilde{f}}^i h(\tilde{x}, u) = \frac{\partial L_{\tilde{f}}^{i-1} h}{\partial \tilde{x}}(\tilde{x}, u) \tilde{f}(\tilde{x}, u) + \sum_{j=0}^{i-1} \frac{\partial L_{\tilde{f}}^{i-1} h}{\partial u^{(j)}}(\tilde{x}, u) u^{(j+1)}, \quad i \geq 1.$$

Definition 4 (*Observability–identifiability Matrix* (Villaverde et al., 2019)). The Observability–identifiability matrix of the system \mathcal{M} (1)a–(1)b with augmented state vector (3), augmented dynamics (4), and analytical inputs u is the following $mn_{\tilde{x}} \times n_{\tilde{x}}$ matrix,

$$\mathcal{O}_I(\tilde{x}, u) = \frac{\partial}{\partial \tilde{x}} \left(L_{\tilde{f}}^0 h(\tilde{x}, u)^T \quad L_{\tilde{f}}^1 h(\tilde{x}, u)^T \quad L_{\tilde{f}}^2 h(\tilde{x}, u)^T \quad \dots \quad L_{\tilde{f}}^{n_{\tilde{x}}-1} h(\tilde{x}, u)^T \right)^T, \quad (5)$$

The FISPO property of \mathcal{M} can be analysed by calculating the rank of the Observability–identifiability matrix:

Theorem 1 (*Observability–identifiability Condition, OIC* (Karlsson et al., 2012)). If the identifiability–observability matrix of a model \mathcal{M} satisfies $\text{rank}(\mathcal{O}_I(\tilde{x}_0, u)) = n_{\tilde{x}} = n_x + n_\theta + n_w$, with \tilde{x}_0 being a (possibly generic) point in the augmented state space, then the system is structurally locally observable and structurally locally identifiable.

2.2.1. Analysis tools

In this paper we generally check the OIC criterion of 1 using STRIKE-GOLDD, an open source MATLAB toolbox (Villaverde, Barreiro, & Papachristodoulou, 2016). Alternatively, for some models we use the Maple code ObservabilityTest, which implements a procedure that avoids the symbolic calculation of the Lie derivatives and is hence computationally efficient (Sedoglavic, 2002). A number of other software tools are available, including GenSSI2 (Ligon, Fröhlich, Chiş, Banga, Balsa-Canto, & Hasenauer, 2018) in MATLAB, IdentifiabilityAnalysis in Mathematica (Karlsson et al., 2012), DAISY in REDUCE (Saccomani, Bellu, Audoly, & d’Angio, 2019), SIAN in Maple (Hong, Ovchinnikov, Pogudin, & Yap, 2019), and the web app COMBOS (Meshkat, Kuo, & DiStefano III, 2014). It should be taken into account that in the present work we are interested in assessing structural identifiability and observability both with constant and continuous time-varying model parameters (or equivalently, with unknown inputs), as explained in Remark 1. Ideally, the method of choice should provide a convenient way of analysing models with this type of parameters (inputs). It is always possible to perform this type of analysis by assuming that the time dependency of the parameters is of a particular form, e.g. a polynomial function of a certain maximum degree.

3. Models

In this article we review compartmental models, which are one of the most widely used families of models in epidemiology. They divide the population into homogeneous compartments, each of which corresponds to a state variable that quantifies the number of individuals that are at a certain disease stage. The dynamics of these compartments are governed by ordinary differential equations, usually with unknown parameters that describe the rates at which individuals move among different stages of disease. The total size of the population is known and commonly denoted as N (Roosa & Chowell, 2019). The basic compartmental model used for describing a transmission disease is the SIR model, in which the population is divided into three classes:

- Susceptible: individuals who have no immunity and may become infected if exposed.
- Infected and infectious: an exposed individual becomes infected after contracting the disease. Since an infected individual has the ability to transmit the disease, he/she is also infectious.
- Recovered: individuals who are immune to the disease and do not affect its transmission.

Another class of models, called SEIR, include an additional compartment to account for the existence of a latent period after the transmission:

- Exposed: individuals vulnerable to contracting the disease when they come into contact with it.

Table 1
List of SIR models and their main features.

ID	Ref.	States	Parameters	Output	ICS	Input	Equations
6	Zheng (2020)	S, I, R	β, γ, K	(1) I (2) KI	100000 100 0		$\dot{S}(t) = -\beta SI/N$ $\dot{I}(t) = \beta SI/N - \gamma I$ $\dot{R}(t) = \gamma I$
7	Zheng (2020)	S, I, R, Q	β, γ, δ	(1) Q (2) I, R, Q	1000 10 0 0		$\dot{S}(t) = -\beta SI/N$ $\dot{I}(t) = \beta SI/N - \gamma I - \delta I$ $\dot{R}(t) = \gamma I$ $\dot{Q}(t) = \delta I$
13	Maier and Brockmann (2020)	S, I, R, X	$\beta, \alpha, k, k_0, C(t_0)$	NX	$1 - I(t_0) - X(t_0)$ $I_0 C(t_0)/(X_0 N)$ $C(t_0)/N$ R_0		$\dot{S}(t) = -\alpha SI - k_0 S$ $\dot{I}(t) = \alpha SI - (\beta + k + k_0)I$ $\dot{X}(t) = (k + k_0)I$ $\dot{R}(t) = \beta I + k_0 S$
15	Roda et al. (2020)	S, I, R	$\beta, \rho, \mu, \tau, d, I_0$	$\rho I + \tau$	S_0 I_0 R_0		$\dot{S}(t) = -\beta SI$ $\dot{I}(t) = \beta SI - (\rho + \mu)I$ $\dot{R}(t) = \rho I - dR$
20	Giordano et al. (2020)	S, I, D, A, R, T, H, E	$\alpha, \beta, \rho, \mu, \tau$ $\epsilon, \eta, \xi, \lambda, \sigma, \kappa,$ $\theta, \nu, \gamma, \delta, \zeta$	D, R, T	S_0 200/60e8 20/60e8 1/60e8 2/60e8 0 0 0		$\dot{S}(t) = -S(\alpha I + \beta D + \gamma A + \delta R)$ $\dot{I}(t) = S(\alpha I + \beta D + \gamma A + \delta R) - (\epsilon + \zeta + \lambda)I$ $\dot{D}(t) = \epsilon I - (\eta + \rho)D$ $\dot{A}(t) = \zeta I - (\theta + \mu + \kappa)A$ $\dot{R}(t) = \eta D + \theta A - (\nu + \xi)R$ $\dot{T}(t) = \mu A + \nu R - (\sigma + \tau)T$ $\dot{H}(t) = \lambda I + \rho D + \kappa A + \xi R + \sigma T$ $\dot{E}(t) = \tau T$
21	Castro et al. (2020)	S, I, D, C, R	p, q, r β, μ	(1) I (2) C (3) C, D			$\dot{S}(t) = -\beta SI/N - pS + qC$ $\dot{I}(t) = \beta SI/N - (r + \mu)I$ $\dot{R}(t) = rI$ $\dot{C}(t) = pS - qC$ $\dot{D}(t) = \mu I$
19	Gaeta (2020)	S, I, J, R, U	β, α, η, ξ	(1) R (2) I, R	S_0 I_0 $(1 - \xi)I_0/\xi$ R_0 U_0		$\dot{S}(t) = -\alpha(I + J)S$ $\dot{I}(t) = \alpha \xi S(I + J) - \beta I$ $\dot{J}(t) = \alpha(1 - \xi)S(I + J) - \eta J$ $\dot{R}(t) = \beta I$ $\dot{U}(t) = \eta J$
24	Kim and Milner (1995)	S, I, R	μ, γ, ϕ c, K	KI		$\sigma(t)$	$\dot{S}(t) = A - \mu S - c\phi \frac{SI}{S+I}$ $\dot{I}(t) = -\mu I + c\phi \frac{SI}{S+I} - \gamma I - I \frac{\sigma(S+I)}{S+I}$ $\dot{R}(t) = -\mu R + \gamma I + I \frac{\sigma(S+I)}{S+I}$
25	Lourenco et al. (2020)	y, z, A	$\beta, \sigma,$ ρ, θ	A	y_0 z_0 A_0	z_d	$\dot{y}(t) = \beta y(1 - z) - \sigma y$ $\dot{z}(t) = \beta y(1 - z)$ $\dot{A}(t) = N \rho \theta z_d$
26	Gallina (2012)	S, I, R, A, Q, J	$\mu_1, \mu_2, d_1, d_2,$ $d_3, d_4, d_5, d_6,$ $k_1, k_2, \lambda, \gamma_1,$ $\gamma_2, \epsilon_a, \epsilon_q, \epsilon_j$	Q, J			$\dot{S}(t) = bN - S(I\lambda + \lambda Q \epsilon_a \epsilon_q + \lambda \epsilon_a A + \lambda \epsilon_j J + d_1)$ $\dot{I}(t) = k_1 A - (\gamma_1 + \mu_2 + d_2)I$ $\dot{R}(t) = \gamma_1 I + \gamma_2 J - d_3 R$ $\dot{A}(t) = S(I\lambda + \lambda Q \epsilon_a \epsilon_q + \lambda \epsilon_a A + \lambda \epsilon_j J) - (k_1 + \mu_1 + d_4)A$ $\dot{Q}(t) = \mu_1 A - (k_2 + d_5)Q$ $\dot{J}(t) = k_2 Q + \mu_2 I - (\gamma_2 + d_6)J$
22	Hethcote, Zhien, and Shengbing (2002)	S, I, Q, R	d, ϵ, β $\gamma, \delta, \alpha_1, \alpha_2$	Q			$\dot{S}(t) = A - \beta SI - dS$ $\dot{I}(t) = \beta SI - I(\gamma + d + \delta + \alpha_1)$ $\dot{Q}(t) = \delta I - (\epsilon + d + \alpha_2)Q$ $\dot{R}(t) = \gamma I + \epsilon Q - R$

(continued on next page)

These idealized models differ from the reality. Contact tracing, screening, or changes in habits are some differences that are not considered in basic SIR or SEIR models, but are important for evaluating the

effects of an intervention. Furthermore, it is not only important to enrich the information about the behaviour of the population; the characteristics of the disease must also be taken into account. These additional details can be incorporated to the model as new parameters, functions,

Table 1 (continued).

ID Ref.	States	Parameters	Output	ICS	InputEquations
27 Raimúndez et al. (2020)	S, A, I, R, D	$\gamma, \delta, k,$ ζ_0, β_0	(1) I, R (2) R, D (3) I, R, D	S_0	$\dot{S}(t) = -\beta_0 gSI/N - \zeta_0 gSA/N$
				A_0	$\dot{A}(t) = \beta_0 gSI/N + \zeta_0 gSA/N - kA$
				I_0	$\dot{I}(t) = kA - (\gamma + \delta)I$
				R_0	$\dot{R}(t) = \gamma I$
28 Raimúndez et al. (2020)	S, A, I, R, RR, D	$\gamma, \delta, k,$ v, ζ_0, β_0	(1) I, R (2) R, D (3) I, R, D	D_0	$\dot{D}(t) = \delta I$
				S_0	$\dot{S}(t) = -\beta_0 gSI/N - \zeta_0 gSA/N$
				A_0	$\dot{A}(t) = \beta_0 gSI/N + \zeta_0 gSA/N - kA$
				I_0	$\dot{I}(t) = kA - (\gamma + \delta)I$
29 Franco (2020)	s, i	R_0	i	R_0	$\dot{R}(t) = \gamma I$
				R_0	$\dot{R}(t) = \gamma I$
				R_0	$\dot{R}(t) = \gamma I$
				R_0	$\dot{R}(t) = \gamma I$
30 Franco (2020)	s, i	R_0, κ	i	$N - 2/e6$	$\dot{s}(t) = -R_0 si$
				$2/e6$	$\dot{i}(t) = (R_0 s - 1)i$
				$N - 2/e6$	$\dot{s}(t) = -\frac{R_0 si}{1 + \kappa i}$
				$2/e6$	$\dot{i}(t) = \left(\frac{R_0 s}{1 + \kappa i} - 1 \right) i$
35 Fosu, Opong, and Appati (2020)	S, L, I, Q, R	$\gamma, \beta, \eta,$ $\delta, \theta_1, \alpha_1,$ α_2	Q, L	$\dot{S}(t) = \mu N - \beta SI - (\gamma + \eta)S + \delta L$	
				$\dot{L}(t) = \eta S - (\gamma + \delta)L$	
				$\dot{I}(t) = \beta SI - (\gamma + \theta_1 + \alpha_1)I$	
				$\dot{Q}(t) = \theta_1 I - (\gamma + \alpha_2)Q$	
37 Gevertz, Greene, Hixahuary Sanchez Tapia, and Sontag (2020)	Sd, Sn, Ad, An, I, R	$\epsilon_s, \epsilon_i, f,$ $h_1, h_2, \beta_1, \delta,$ $\beta_a, \gamma_{ai}, \gamma_{ir}$	Sd, I	0	$\dot{S}d(t) = -\epsilon_s \beta_a (A_n + e_a A_d) S_d - h_1 S_d + h_2 S_n - \epsilon_s \beta_1 S_d I$
				$1 - 10^{-5}$	$\dot{S}n(t) = -\beta_1 S_n I - \beta_a (A_n + e_a A_d) S_n + h_1 S_d - h_2 S_n$
				0	$\dot{A}d(t) = \epsilon_s \beta_1 S_d I + \epsilon_s \beta_a (A_n + e_a A_d) S_d + h_2 A_n - \gamma_{ai} A_d - h_1 A_d$
				0	$\dot{A}n(t) = \beta_1 S_n I + \beta_a (A_n + e_a A_d) S_n + h_1 A_d - \gamma_{ai} A_n - h_2 A_n$
37 Gevertz, Greene, Hixahuary Sanchez Tapia, and Sontag (2020)	Sd, Sn, Ad, An, I, R	$\epsilon_s, \epsilon_i, f,$ $h_1, h_2, \beta_1, \delta,$ $\beta_a, \gamma_{ai}, \gamma_{ir}$	Sd, I	10^{-5}	$\dot{I}(t) = f \gamma_{ai} (A_d + A_n) - \delta I - \gamma_{ir} I$
				0	$\dot{R}(t) = (1 - f) \gamma_{ai} (A_d + A_n) + \gamma_{ir} I$
				0	$\dot{R}(t) = (1 - f) \gamma_{ai} (A_d + A_n) + \gamma_{ir} I$
				0	$\dot{R}(t) = (1 - f) \gamma_{ai} (A_d + A_n) + \gamma_{ir} I$

or extra compartments. Compartments such as asymptomatic, quarantined, isolated, and hospitalized have been widely used in COVID-19 models. From 29 articles, most of which are very recent (Castro et al., 2020; Chatterjee, Chatterjee, Kumar, & Shankar, 2020; Dohare, 2020; Eikenberry et al., 2020; Fosu et al., 2020; Franco, 2020; Gaeta, 2020; Gevertz et al., 2020; Giordano et al., 2020; Hubbs, 2020; Jia et al., 2020; Liangrong, Yang, Zhang, Zhuge, & Hong, 2020; Lopez & Rodo, 2020; Lourenco et al., 2020; Maier & Brockmann, 2020; McGee, 2020; Pribylova & Hajnova, 2020; Rahman & Kuddus, 2020; Raimúndez et al., 2020; Roda et al., 2020; Sameni, 2020; Shi, Cao, & Feng, 2020; Zha et al., 2020; Zheng, 2020), we have collected 36 models. Depending on whether they have an exposed compartment or not, they can be broadly classified as belonging to the SIR or SEIR families. However, most of these models include additional compartments.

3.1. SIR models

Susceptible individuals become infected with an incidence of:

$$-\beta SI$$

where $\beta = pc$ is the transmission rate, c is the contact rate and p the probability that a contact with a susceptible individual results in a transmission (Martcheva, 2015). Individuals who recover leave the infectious class at rate γ , where $1/\gamma$ is the average infectious period. The set of differential equations describing the basic SIR model is given by:

$$\begin{aligned}\dot{S}(t) &= -\beta S(t)I(t) \\ \dot{I}(t) &= \beta S(t)I(t) - \gamma R(t) \\ \dot{R}(t) &= \gamma R(t)\end{aligned}\quad (6)$$

As mentioned above, compartmental models can be extended to consider further details. We have found models that incorporate the

following features: asymptomatic individuals, births and deaths, delay-time, lock-down, quarantine, isolation, social distancing, and screening. Fig. 1 shows a classification of the SIR models reviewed in this article, and Table 1 lists them along with their equations. Multiple output choices have been considered in the study of the structural identifiability and observability of some models. In such cases the observations are listed in the *Output* column.

3.2. SEIR models

Individuals in the SEIR model are divided in four compartments: Susceptible (S), Exposed (E), Infected (I) and Recovered (R). Compared to the SIR models, the additional compartment E allows for a more accurate description of diseases in which the incubation period and the latent period do not coincide, i.e. the period between which an infected becomes infectious. This is why SEIR models are in principle best suited to epidemics with a long incubation period such as COVID-19 (Franco, 2020).

Susceptible individuals move to the exposed class at a rate $\beta I(t)$, where β is the transmission rate parameter. Exposed individuals become infected at rate κ , where $1/\kappa$ is the average latent period. Infected individuals recover at rate γ , where $1/\gamma$ is the average infectious period.

Thus, the set of differential equations describing the basic SEIR model is:

$$\begin{aligned}\dot{S}(t) &= -\beta S(t)I(t) \\ \dot{E}(t) &= \beta S(t)I(t) - \kappa E(t) \\ \dot{I}(t) &= \kappa E(t) - \gamma I(t) \\ \dot{R}(t) &= \gamma I(t)\end{aligned}\quad (7)$$

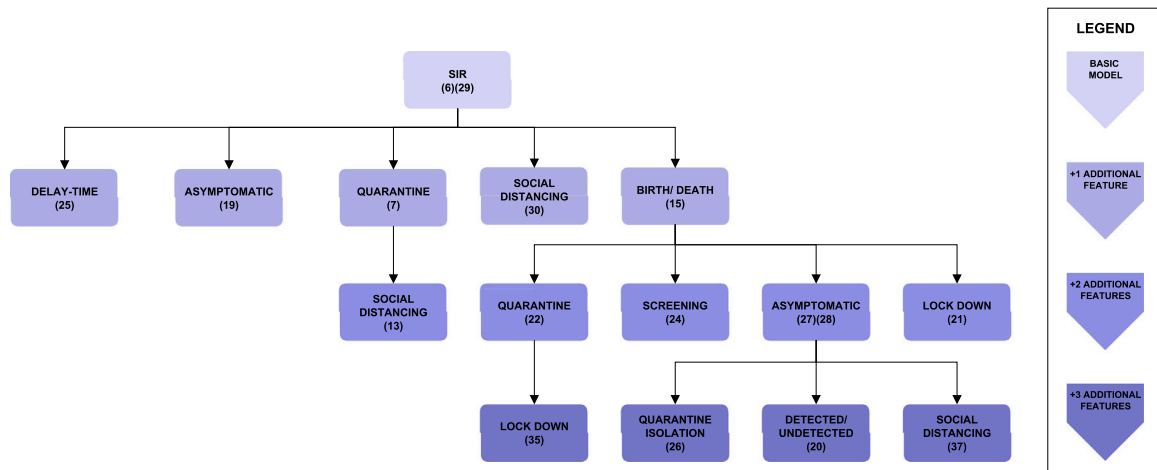


Fig. 1. Classification of SIR models. Each block represents a model structure. The basic, three-compartment SIR model structure is on top of the tree. Every additional block is labelled with the additional feature that it contains with respect to its parent block. The darkness of the shade indicates the number of additional features with respect to the basic SIR model.

Existing extensions of SEIR models may incorporate some of the following features: asymptomatic individuals, births and deaths, hospitalization, quarantine, isolation, social distancing, screening and lock-down. Fig. 2 shows a classification of the models found in the literature; Table 2 lists them along with their equations.

4. Results and discussion

We analysed the structural identifiability and observability of the 17 SIR model structures (a total of 98 model versions considering the different output configurations and time-varying parameter assumptions) and 19 SEIR models (with a total of 157 model versions) listed in Tables 1 and 2. The detailed results for each model are given in Appendix, which reports the structural identifiability of each parameter and the observability of each state, for every model version. In the remainder of this section we provide an overview of the main results.

4.1. General patterns

The general patterns regarding state observability are as follows. The recovered state (R) is almost never observable unless it is directly measured (D.M.) as output; the only exceptions are two SEIR models, 31 and 38, for which R is observable under the assumption of time-varying parameters. The susceptible state (S), in contrast, is observable in roughly two thirds of the models (SIR: 65/98, SEIR: 103/157); this is also true for the exposed state (E) in the SEIR models. The infected state (I) is included in most studies among the outputs, either directly (D.M.) or indirectly measured (as part of a parameterized measurement function). When it is not considered in this way, its observability is generally similar to that of S (in 18/157 model versions I is not an output and it is observable in 13/18).

The transmission and recovery rates (β , γ) are the two parameters common to all SIR models. The transmission rate is identifiable in 59/98 model versions, and γ in 51/98 and its derivatives in 12/98. SEIR models have a third parameter in common, the latent period (κ). It is identifiable in most of the models (145/157), as well as the recovery rate (111/157). The transmission rate is identifiable in 101/157 model versions, but it is not identifiable in any SEIR model version that accounts for social distancing (numbers 34 and 61); we found no clear pattern in the other models.

4.2. The effect of time-varying parameters

The transmission rate β , the recovery rate γ , and in SEIR models the latent period κ , can vary during an epidemic as a result of changes in the population's behaviour (Chen, Lu, Chang, & Liu, 2020; Liangrong et al., 2020), the introduction of new drugs or new medical equipment (Liangrong et al., 2020), or the reduction of the period duration as a result of high temperatures (Li & Zhao, 2019). To account for such variations, the present study has considered both the constant and the time-varying cases, by including the corresponding variables either in the constant parameter vector θ or in the unknown input vector $w(t)$, respectively, as described in Remark 1. Changing a parameter from constant to time-varying naturally influences structural identifiability and observability. This effect is graphically summarized in Figs. 3–7, which represent classes of models in tree form and classify them according to their observability. Each model is shaded with a colour, according to the observability of the parameter studied. Some models include different rates for different population groups: for example, they may consider two different transmission rates for symptomatic and asymptomatic individuals. For those models, each rate may have different observability properties when considered time-varying parameters; in such cases the model is depicted between two colour blocks (see for example the SIR 20 model in Fig. 3).

Changing β from a constant to a time-varying parameter (or equivalently an unknown input) does not change its observability nor that of the other variables in SIR models. In contrast, this is not the case with the recovery rate γ , for which a somewhat counter-intuitive result may be obtained: by changing γ from a constant to a continuous function of time with at least a non-zero derivative, its model can become more observable and identifiable — despite the fact that it is an unknown function. An example of this is the SIR model 15: if γ is constant the model has only one identifiable parameter, τ , and no observable states; if γ is time-varying with at least one non-zero derivative, two parameters become identifiable (β , μ), two states become observable (I, S), and γ itself is observable. In the other models, when γ is not identifiable as a constant nor observable as an unknown input, its successive derivatives are observable.

For the SEIR models, the consideration of the β parameter as an unknown input function follows a similar trend to that of the SIR models with the exception of model 38, which gains both observability and identifiability and becomes FISPO. Considering the recovery rate γ (Fig. 7) or the latent period κ (Fig. 6) individually as time-varying parameters generally leads to greater observability, except for model 31(1). As an example, in model 39(2) one of the unknown inputs becomes observable, three states become observable (S, E, I), and three

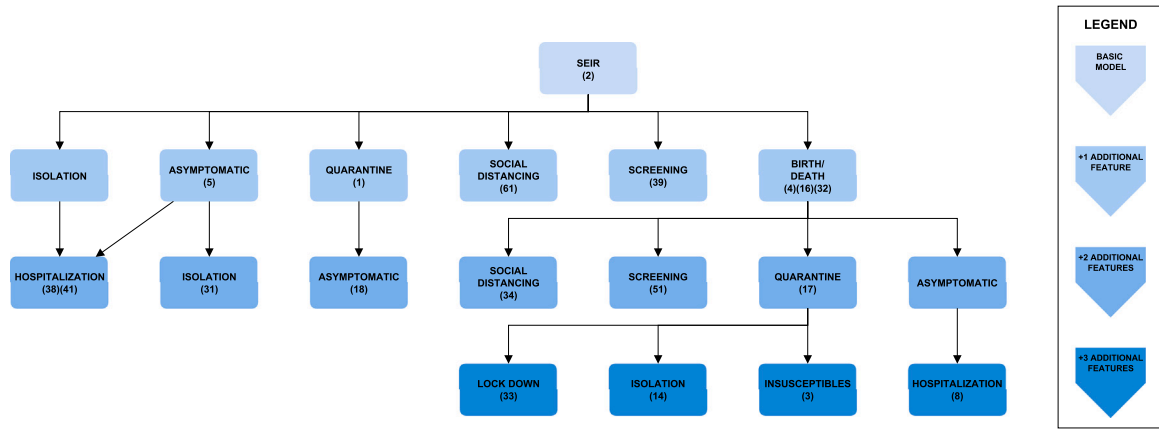


Fig. 2. Classification of SEIR models. Each block represents a model structure. The basic, four-compartment SEIR model structure is on top of the tree. Every additional block is labelled with the additional feature that it contains with respect to its parent block. The darkness of the shade indicates the number of additional features with respect to the basic SEIR model.

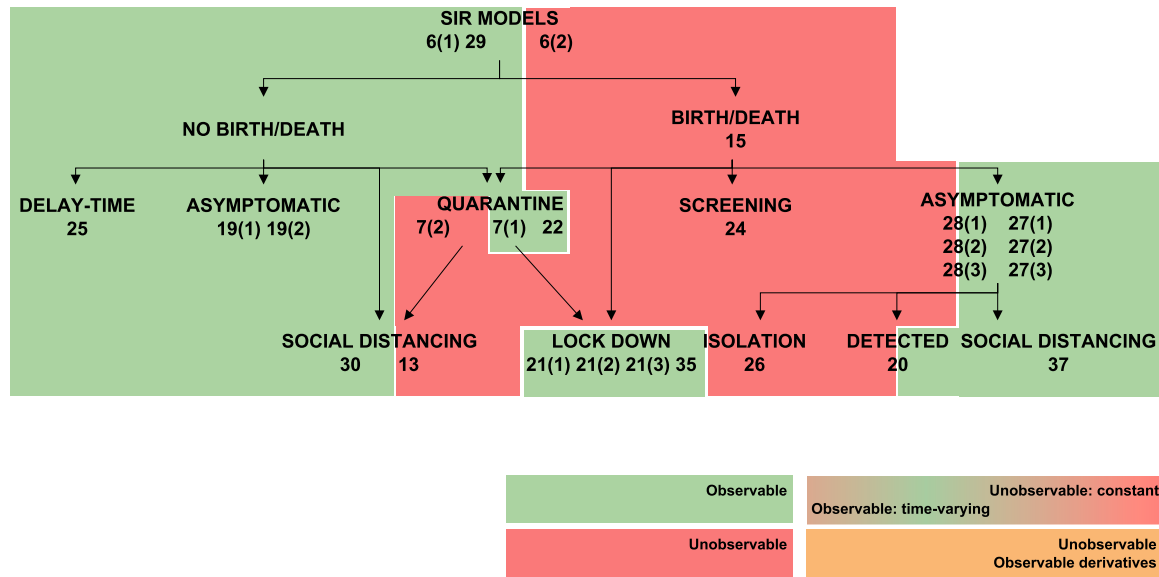


Fig. 3. Observability of β (transmission rate) in SIR models. Models in which β is observable are shown in green, and non-observable in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

parameters become identifiable (γ , μ_i , β); or the 16(2) model, in which both its input and three states (S, E, I) become observable and two parameters (μ , β) become identifiable.

Besides the transmission rate, latent period, and recovery rate, other rates (screening, disease-related deaths, and isolation) have also been considered as time-varying parameters in some studies. The observability of most models is not modified if these parameters are allowed to change in time; the exception being 8 models which gain observability. An example is the SEIR model 41(1), which has seven parameters, seven states, and one output. Assuming constant parameters, five of them are structurally identifiable (κ , α , β , γ_1 , γ_2) and two are unidentifiable (q , ρ), while there are three observable states (I, J, C) and four unobservable states (S, E, A, R) (Roosa & Chowell, 2019). However, when the parameter ρ (which describes the proportion of exposed/latent individuals who become clinically infectious) is considered time-varying, all parameters become identifiable (including ρ) and six states become observable (all except R, which is never observable unless it can be directly measured, as we have already mentioned).

The fact that allowing an unknown quantity to change in time can improve its observability – and also the observability of other variables in a model – may seem paradoxical. An intuitive explanation can be

obtained from the study of the symmetry in the model structure. The existence of Lie symmetries amounts to the possibility of transforming parameters and state variables while leaving the output unchanged, i.e. their existence amounts to lack of structural identifiability and/or observability (Yates, Evans, & Chappell, 2009). The STRIKE-GOLDD toolbox used in this paper includes procedures for finding Lie symmetries (Massonis & F. Villaverde, 2020). Let us use the SIR 15 model as an example. This model has five parameters (τ , β , ρ , μ , d), of which only τ is identifiable if assumed constant. The model contains the following symmetry:

$$\mu^* = \mu + \epsilon \rho,$$

where ϵ is the parameter of the Lie group of transformations. Thus, there is a symmetry between ρ and μ that makes them unidentifiable: changes in one parameter can be compensated by changes in the other one. However, if ρ is time-varying and μ is constant, the latter cannot compensate the changes of the former, and the symmetry is broken. Indeed, if ρ is considered time-varying the model gains identifiability (not only μ , but also τ and β become identifiable) and observability (S, I and ρ become observable).

Table 2
List of SEIR models and their main features.

ID	Ref.	States	Parameters	Output	ICS	Input	Equations
2	Roosa and Chowell (2019)	S, E, I, R	β, γ, k, K	(1) C (2) KC			$\dot{S}(t) = -\beta SI/N$ $\dot{E}(t) = \beta SI/N - kE$ $\dot{I}(t) = kE - \gamma I$ $\dot{R}(t) = \gamma I$ $\dot{C}(t) = kE$
34	Chitnis (2017)	S, E, I, R	$\beta, \gamma, \mu,$ ϵ, τ, K	KI			$\dot{S}(t) = \Lambda - r\beta SI/N - \mu S$ $\dot{E}(t) = \beta SI/N - \epsilon E - \mu E$ $\dot{I}(t) = \epsilon E - \gamma I - \mu I$ $\dot{R}(t) = \gamma I - \mu R$
16	Roda et al. (2020)	S, E, I, R	$\beta, \rho, \mu, d,$ ϵ, τ, E_0, I_0	(1) $\rho I + \tau$ (2) μI	S_0 E_0 I_0 R_0		$\dot{S}(t) = -\beta SI$ $\dot{E}(t) = \beta SI - \epsilon E$ $\dot{I}(t) = \epsilon E - (\rho + \mu)I$ $\dot{R}(t) = \rho I - dR$
51	McGee (2020)	S, E, I, De, Di, R, F	$q, \mu_1, \mu_d,$ $\sigma, \sigma_d, \theta_e, \theta_i,$ $\gamma, \gamma_d, \beta, \beta_d,$ ϕ_i, ϕ_e, μ_0	(1) De, Di, F (2) Di, F			$\dot{S}(t) = -\beta SI/N - q\beta_d SDi/N + vN - \mu_0 S$ $\dot{E}(t) = \beta SI/N + q\beta_d SDi/N - \sigma E - \theta_e \phi_e E - \mu_0 E$ $\dot{I}(t) = \sigma E - \gamma I - \mu_1 I - \theta_i \phi_i I - \mu_0 I$ $\dot{De}(t) = \theta_e \phi_e E - \sigma_d De - \mu_0 De$ $\dot{Di}(t) = \theta_i \phi_i I + \sigma_d De - \gamma_d Di - \mu_d Di - \mu_0 Di$ $\dot{R}(t) = \gamma I + \gamma_d Di - \mu_0 R$ $\dot{F}(t) = \mu_1 I + \mu_d Di$
14	Lopez and Rodo (2020)	S, E, I, Q, R, D, C	$\beta, \alpha, \gamma,$ δ, τ	C, Q, D		$k(t), \lambda(t)$	$\dot{S}(t) = \mu N - \alpha S - \beta SIN - \mu S$ $\dot{E}(t) = \beta SIN - \mu E - \gamma E$ $\dot{I}(t) = \gamma E - \delta I - \mu I \mu S$ $\dot{Q}(t) = \delta I - \lambda Q - kQ - \mu Q$ $\dot{R}(t) = \lambda Q - \mu S$ $\dot{D}(t) = kQ$ $\dot{C}(t) = \alpha S - \mu C - \tau C$
61	Hubbs (2020)	S, E, I, R	$p, \alpha, \beta, \gamma, K$	KI			$\dot{S}(t) = -p\beta SI$ $\dot{E}(t) = p\beta SI - \alpha E$ $\dot{I}(t) = \alpha E - \gamma I$ $\dot{R}(t) = \gamma I$
5	Pribylova and Hajnova (2020)	S, E, I, R, A	$\mu_1, \mu_2, \gamma, \beta, p$	I			$\dot{S}(t) = -\beta S(I + A)$ $\dot{E}(t) = \beta S(I + A) - \gamma E$ $\dot{I}(t) = \gamma pE - \mu_1 I$ $\dot{A}(t) = \gamma(1 - p)E - \mu_2 A$ $\dot{R}(t) = \mu_1 I + \mu_2 A$
1	Zha et al. (2020)	S, E, I, R, Q	ϕ, β, γ, w	Q			$\dot{S}(t) = -\beta SI$ $\dot{E}(t) = \beta SI - wE$ $\dot{I}(t) = wE - \phi I - (1 - \phi)\gamma I$ $\dot{R}(t) = \gamma Q + (1 - \phi)\gamma I$ $\dot{Q}(t) = -\gamma Q + \phi I$
3	Liangrong et al. (2020)	S, E, I, R, Q, D, P	$\alpha, \beta, \gamma, \delta$ E_0, I_0	Q, R, D	S_0 E_0 I_0 Q_0 R_0 D_0 P_0	$\lambda(t), \kappa(t)$	$\dot{S}(t) = -\beta SI/N - \alpha S$ $\dot{E}(t) = \beta SI/N - \gamma E$ $\dot{I}(t) = \gamma E - \delta I$ $\dot{Q}(t) = \delta I - \lambda Q - \kappa Q$ $\dot{R}(t) = \lambda Q$ $\dot{D}(t) = \kappa Q$ $\dot{P}(t) = \alpha S$

(continued on next page)

4.3. Applying the results in practice: an illustrative example

Let us now illustrate how the results of this study may be applied in a realistic scenario. We use as an example the model SIR 26, which has 6 states (S, I, R, A, Q, J) and 16 parameters ($d_1, d_2, d_3, d_4, d_5, d_6, k_1, k_2, \lambda, \gamma_1, \gamma_2, \epsilon_a, \epsilon_q, \epsilon_j, \mu_1, \mu_2$); its equations are shown in Table 1. This model includes the following additional features with respect to the basic SIR model: birth/death, asymptomatic individuals (A), quarantine (Q), and isolation (J). In its original publication two states were measured (Q, J). With these two states as outputs the model

has five identifiable parameters ($d_1, d_5, \epsilon_q, k_2, \mu_1$) and two observable states (A, I); thus, there are two unobservable states (S, R) and ten unidentifiable parameters.

If we are interested in estimating e.g. the number of susceptible individuals (S), this model would not be appropriate. How should we proceed in that scenario?

One way of improving observability could be by including more outputs (option 1). For example, since there is a separate class for asymptomatic individuals (A), the infected compartment (I) considers only individuals with symptoms, and we could assume that they can be

Table 2 (continued).

4	Sameni (2020)	S, E, I, R, P	$\alpha_e, \alpha_i, \rho,$ β, μ, κ, e_0	(1) $S + vs(t)$	$N - e_0$	$vs(t)$	$\dot{S}(t) = -\alpha_e SE - \alpha_i SI$
				(1) $I + vi(t)$	e_0	$vi(t)$	$\dot{E}(t) = \alpha_e SE + \alpha_i SI - \kappa E - \rho E$
				(1) $R + vr(t)$	0	$vr(t)$	$\dot{I}(t) = \kappa E - \beta I - \mu I$
				(1) $P + vp(t)$	0	$vp(t)$	$\dot{R}(t) = \beta I + \rho E$
				(2) P	0		$\dot{P}(t) = \mu I$
8	Eikenberry et al. (2020)	S, E, I, R A, H, D	$\delta, \alpha, \eta,$ $\sigma, \gamma_a, \gamma_i, \gamma_h$ Φ, E_0, R_0, A_0 H_0, D_0	D, H, I	1e6		$\dot{S}(t) = -\beta(I + \eta A)S/N$
					E_0		$\dot{E}(t) = \beta(I + \eta A)S/N - \sigma E$
					50		$\dot{I}(t) = \alpha \sigma E - \Phi I - \gamma_i I$
					A_0	$\beta(t)$	$\dot{A}(t) = (1 - \alpha)\sigma E - \gamma_a A$
					H_0		$\dot{H}(t) = \Phi I - \delta H - \gamma_h H$
38	Shi et al. (2020)	S, E, I, Sq, Eq, H, R	c, q, λ, β $\delta_i, \delta_q, \alpha, \gamma_i$ γ_h, θ	(1) I, R			$\dot{S}(t) = -(c\beta + cq(1 - \beta))S(I + \theta E) + \lambda Sq$
				(2) H, viI, vrR		$\sigma(t)$	$\dot{E}(t) = c\beta(1 - q)S(I + \theta E) - \sigma E$
				(3) Sq, Eq			$\dot{I}(t) = \sigma E - (\delta_i + \alpha + \gamma_i)I$
							$\dot{Sq}(t) = cq(1 - \beta)S(I + \theta E) - \lambda Sq$
							$\dot{Eq}(t) = cq\beta S(I + \theta E) - \delta_q Eq$
41	Roosa and Chowell (2019)	S, E, A, I, J, R, C	β, k, γ_1 $\gamma_2, \alpha, \rho, q$	(1) C			$\dot{H}(t) = \delta_i I + \delta_q Eq - (\alpha + \gamma_h)H$
				(2) J, I			$\dot{R}(t) = \gamma_i I + \gamma_h H$
							$\dot{S}(t) = -\beta SI(I + J + qA)/N$
							$\dot{E}(t) = \beta S(I + J + qA)/N - kE$
							$\dot{A}(t) = k(1 - \rho)E - \gamma_1 A$
39	McGee (2020)	S, E, I, De, Di, R, F	$q, \mu_i, \mu_d,$ $\sigma, \sigma_d, \theta_e, \theta_i,$ $\gamma, \gamma_d, \beta, \beta_d,$ ϕ_i, ϕ_e	(1) F, Di			$\dot{I}(t) = k\rho E - (\alpha + \gamma_1)I$
				(2) F, De, Di			$\dot{J}(t) = \alpha I - \gamma_2 J$
							$\dot{R}(t) = \gamma_1(A + I) + \gamma_2 J$
							$\dot{C}(t) = \alpha I$
							$\dot{S}(t) = -\beta SI/N - q\beta_d SDi/N$
17	Chatterjee et al. (2020)	S, E, I, R, Q, D	$\beta, \epsilon, \gamma, d,$ q, qt	(1) D	249		$\dot{E}(t) = \beta SI/N + q\beta_d SDi/N - \sigma E - \theta_e \phi_e E$
				(2) Q, D	E_0		$\dot{I}(t) = \sigma E - \gamma I - \mu_i I - \theta_i \phi_i I$
				(3) D, I, Q	qQ_0		$\dot{De}(t) = \theta_e \phi_e E - \sigma_d De$
					Q_0		$\dot{Di}(t) = \theta_i \phi_i I + \sigma_d De - \gamma_d Di - \mu_d Di$
					23		$\dot{R}(t) = \gamma I + \gamma_d Di$
18	Jia et al. (2020)	S, E, I, A, R, Q, D	$\beta, \theta, \lambda, \sigma,$ $\rho, \epsilon_a, \epsilon_i, \gamma_i,$ $\gamma_a, \gamma_d, p, d, d_d$	(1) D, I, Q	5		$\dot{F}(t) = \mu_i I + \mu_d Di$
				(2) Q, D	9219849e3		$\dot{S}(t) = -\beta SI(I + \theta A) - pS + \lambda Q$
				(3) D	4142251e2		$\dot{Q}(t) = pS - \lambda Q$
					3207		$\dot{E}(t) = \beta S(I + \theta A) - \sigma E$
					595		$\dot{A}(t) = \sigma(1 - \rho)E - \epsilon_a A - \gamma_a A$
31	Dohare (2020)	S, E, I, A, J, R	$\alpha, \sigma, h,$ $r, q, f, \beta_1,$ $\beta_2, \phi, \gamma, I_0$	(1) I, J	563		$\dot{I}(t) = \sigma\rho E - \gamma_i I - d_i I - \epsilon_i I$
				(2) I	227		$\dot{D}(t) = \epsilon_a A + \epsilon_i I - d_d D - \gamma_d D$
					3		$\dot{R}(t) = \gamma_a A + \gamma_i I + \gamma_d D$
					0.9N		$\dot{S}(t) = -\alpha \frac{E + I + A}{N} S - \sigma S$
					$9(I_0 + A_0)$		$\dot{E}(t) = \alpha \frac{E + I + A}{N} S - \beta_1 E$
17	Chatterjee et al. (2020)	S, E, I, R, Q, D	$\beta, \epsilon, \gamma, d,$ q, qt	(1) D	I_0		$\dot{I}(t) = \beta_1 hE + \beta_2 rA - \phi q I - \gamma(1 - q)I$
				(2) Q, D	$I_0 f$		$\dot{A}(t) = \beta_1(1 - h)E - \beta_2 rA - \gamma(1 - r)A$
				(3) D, I, Q	0		$\dot{J}(t) = \phi q I - \gamma J$
					0		$\dot{R}(t) = \gamma(1 - q)I + \gamma(1 - r)A + \gamma J$
					0		

(continued on next page)

detected. By including 'I' in the output set, the structural identifiability and observability of the model improves: six more parameters are identifiable ($\lambda, \epsilon_a, \epsilon_j, d_4, k_1, \mu_2$) and the state in which we are interested (S) becomes observable.

However, including more outputs is not always realistic. Another possibility would then be to reduce the complexity of the model by decreasing the number of additional features (option 2). For example,

leaving out the asymptomatic compartment leads to the following model:

$$\dot{S}(t) = bN - S(I\lambda + \lambda\epsilon_q Q + \lambda\epsilon_j J + d_1)$$

$$\dot{I}(t) = S(I\lambda + \lambda\epsilon_q Q + \lambda\epsilon_j J) - (\gamma_1 + \mu_1 + \mu_2 + d_2)I$$

$$\dot{R}(t) = \gamma_1 I + \gamma_2 J - d_3 R$$

$$\dot{Q}(t) = \mu_1 I - (k_2 + d_5)Q$$

Table 2 (continued).

32	Rahman and Kuddus (2020)	S, L, I, R	$w, \beta, \alpha,$ γ, μ	wL	$\dot{S}(t) = A - \frac{\beta SI}{1 + \alpha I} - \mu S$ $\dot{L}(t) = \frac{\beta SI}{1 + \alpha I} - (w + \mu)L$ $\dot{I}(t) = wL - (\gamma + \mu)I$ $\dot{R}(t) = \gamma I - \mu R$
33	Fosu et al. (2020)	S, L, E, I, Q, R	$\gamma, \beta_1, \eta,$ $\delta, \xi, \theta_2,$ $\epsilon, \theta_1, \alpha_1, \alpha_2,$	L, Q	$\dot{S}(t) = \mu N - \beta_1 SI - (\gamma + \eta)S + \delta L + \xi E$ $\dot{L}(t) = \eta S - (\gamma + \delta)L$ $\dot{E}(t) = \beta_1 SI - (\gamma + \theta_2 + \epsilon + \xi)E$ $\dot{I}(t) = \epsilon E - (\gamma + \theta_1 + \alpha_2)I$ $\dot{Q}(t) = \theta_1 I + \theta_2 E - (\gamma + \alpha_2)Q$ $\dot{R}(t) = \alpha_1 I + \alpha_2 Q - \gamma R$

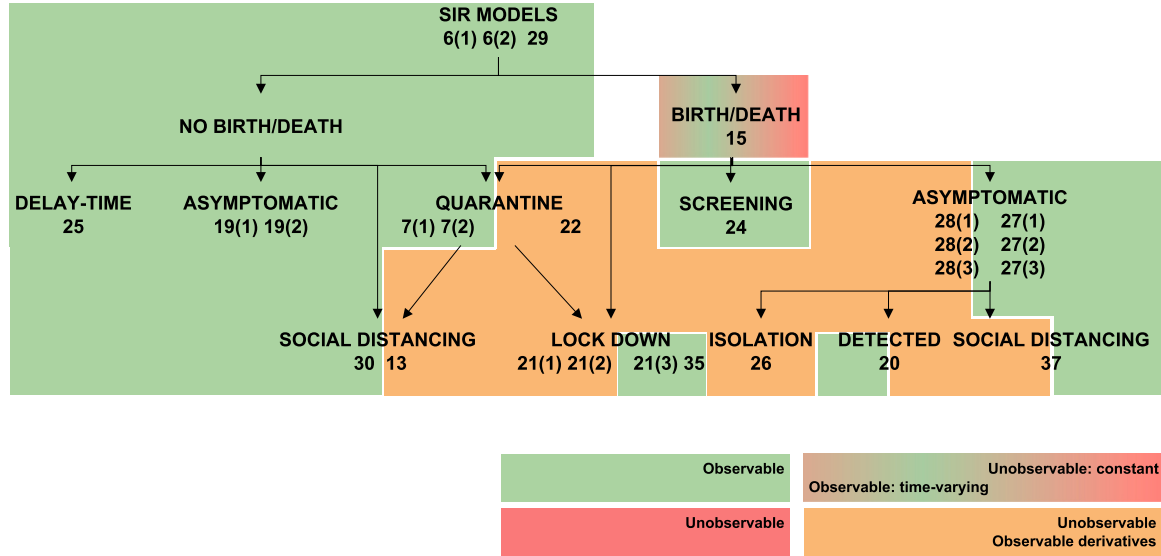


Fig. 4. Observability of γ (recovery rate) in SIR models. Models in which γ is observable are shown in green. Models in which it is unobservable if constant and observable if time-varying are shown in a green-red gradient. Finally, models in which only its time-derivatives are observable are shown in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

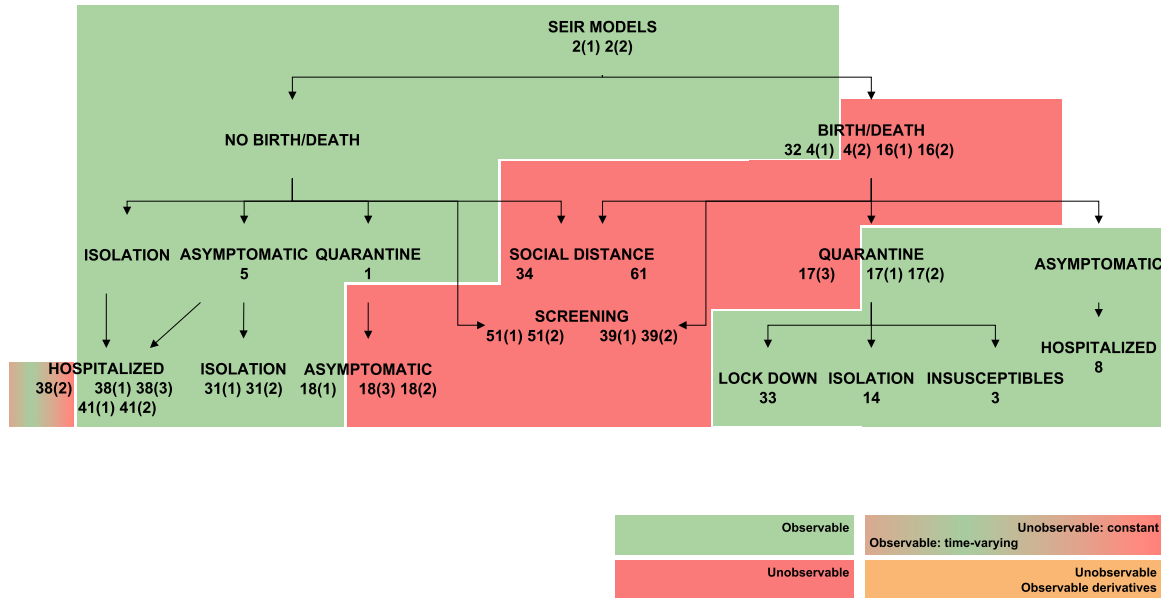


Fig. 5. Observability of β (transmission rate) in SEIR models. Models in which β is observable are shown in green, and non-observable in red. Models in which it is unobservable if constant and observable if time-varying are shown in a green-red gradient. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

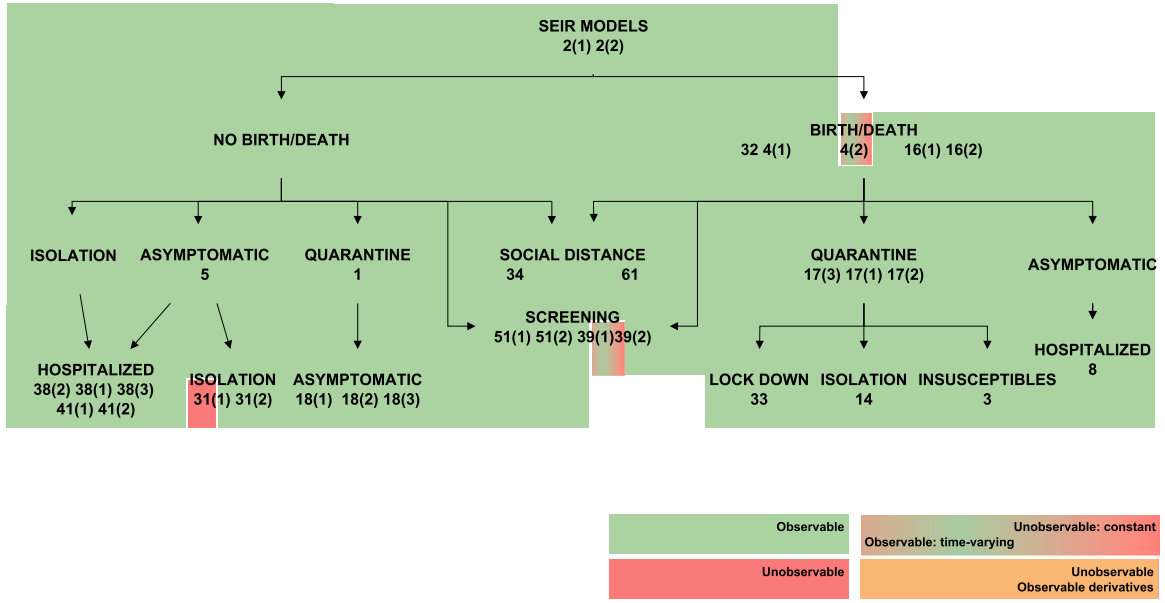


Fig. 6. Observability of κ (latent period) in SEIR models. Models in which κ is observable are shown in green, and non-observable in red. Models in which it is unobservable if constant and observable if time-varying are shown in a green-red gradient. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

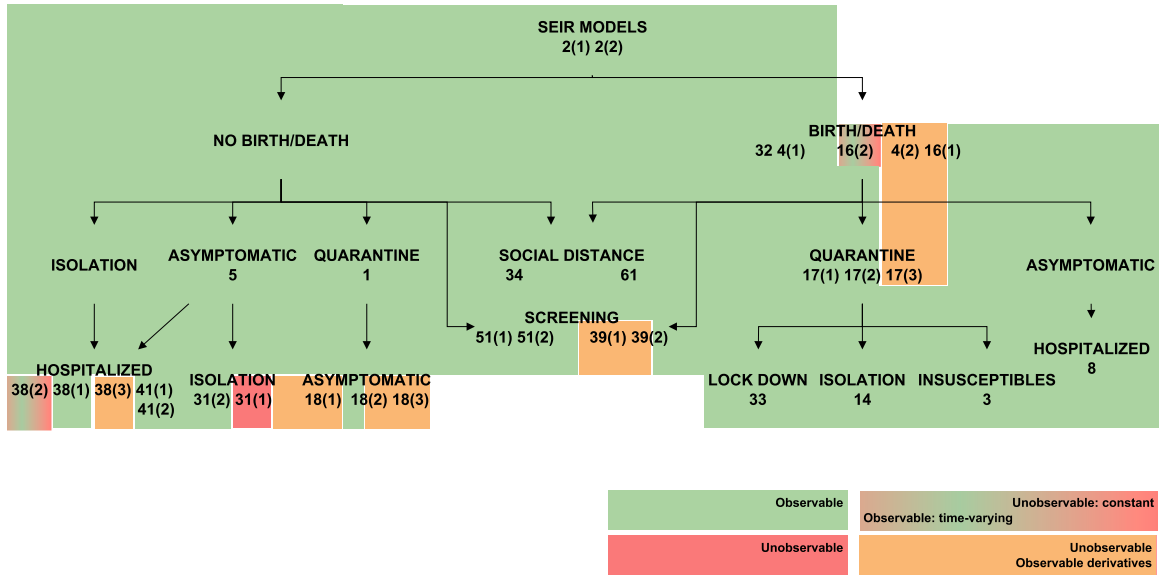


Fig. 7. Observability of γ (recovery rate) in SEIR models. Models in which γ is observable are shown in green, and non-observable in red. Models in which it is unobservable if constant and observable if time-varying are shown in a green-red gradient. Finally, models in which only its time-derivatives are observable are shown in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\dot{J}(t) = k_2 Q + \mu_2 I - (\gamma_2 + d_6) J$$

The output of the model is the same, Q , J . In this case, the model has eight identifiable parameters (λ , ϵ_q , ϵ_j , d_1 , d_5 , μ_1 , μ_2 , k_2) and two observable states (S , I).

A third possibility is to *simplify the parameterization of the model (option 3)*. This model considers a different death rate for every compartment (d_i , $i = 1, \dots, 6$). With some loss of generality, we could consider a specific death rate for infected individuals, $d_I = d_2$, and a general death rate d for all non-infected and asymptomatic

individuals, $d = d_1 = d_3 = d_4 = d_5 = d_6$. This reduction of the number of parameters leads to a better observability to the model: the only unidentifiable parameters are d_2 , γ_1 , and k_1 , and the only non-observable state is R . Thus, this option also allows to identify S .

Finally, a fourth possible course of action is to *reparameterize the model using identifiable combinations*. This solution entails reducing the number of parameters involved in the model, with new variables consisting of e.g. the product or quotient of some parameters and/or states. These new variables are identifiable (observable) combinations,

although the individual variables are not independently identifiable nor observable. There are several methods for finding identifiable combinations, with theoretical bases including Taylor series (Evans & Chappell, 2000), Lie Algebra (Massonis & F. Villaverde, 2020; Merkt, Timmer, & Kaschek, 2015; Sedoglavic, 2002), Gröbner bases (Meshkat et al., 2014), and bootstrap (Hengl, Kreutz, Timmer, & Maiwald, 2007). These methods can be used to find functionally related parameters and/or identifiable combinations of them. This information provides hints for generating identifiable model reparameterizations, although this step requires some user intervention. As an example, consider the simplest SIR model 6, in which the parameter combination β/K is identifiable. Reparameterizing the model in this way makes it identifiable, although it is still impossible to estimate either the transmission rate or the ratio of reported cases. In addition, the three states would also have to be scaled by one of the above-mentioned parameters, thus losing their original meaning.

5. Conclusions

Our analyses have shown that a fraction of the models found in the literature have unidentifiable parameters. Key parameters such as the transmission rate (β), the recovery rate (γ), and the latent period (κ) are structurally identifiable in most, but not all, models. The transmission and recovery rates are identifiable in roughly two thirds of the models, and the latent period in almost all (> 90%) of them. Likewise, the states corresponding to the number of susceptible (S) and exposed (E) individuals are non-observable in roughly one third of the model versions analysed in this paper. The number of infected individuals (I) can usually be directly measured, but it is non-observable in one third of the model versions in which it is not measured. The situation is worse for the number of recovered individuals (R), which is almost never observable unless it is directly measured. Many models include other states in addition to S, E, I, and R, which are not always observable either.

The transmission rate and other parameters may vary during the course of an epidemic, as a result of a number of factors such as changes in public policy, population behaviour, or environmental conditions. To account for these variations, in the present study we have considered both the constant and the time-varying parameter case. Somewhat unexpectedly, we found that allowing for variability in an unknown parameter often improves the observability and/or identifiability of the model. This phenomenon might be explained by the contribution of this variability to the removal of symmetries in the model structure. Structural identifiability and observability depend on which states or functions are measured. The lack of these properties may in principle be surmounted by choosing the right set of outputs (Anguelova, Karlsson, & Jirstrand, 2012), but the required measurements are not always possible to perform in practice. Epidemiological models are a clear example of this; limitations such as lack of testing or the existence of asymptomatic individuals usually make it impossible to have measurements of all states. An alternative to measuring more states is to use a model with fewer compartments and/or a simpler parameterization, thus decreasing the number of states and/or parameters. Reducing the model dimension in this way may achieve observability and identifiability.

Even when it is not possible (or practical) to avoid non-observability or non-identifiability by any means, the model may still be useful, as long as it is only used to infer its observable states

or identifiable parameters. For example, we may be interested in determining the transmission rate β but not the number of recovered individuals R; in such case it is fine to use a model in which β is identifiable even if R is not observable. Of course, this means that, to ensure that a model is properly used, it is necessary to characterize its identifiability and observability in detail, to know if the quantity of interest is observable/identifiable.

The contribution of this work has been to provide such a detailed analysis of the structural identifiability and observability of a large set of compartmental models of COVID-19 presented in the recent literature. The results of our analyses can be used to avoid the pitfalls caused by non-identifiability and non-observability. By classifying the existing models according to these properties, and arranging them in a structured way as a function of the compartments that they include, our study has answered the following question: given the sets of existing models and available measurements, which model is appropriate for inferring the value of some particular parameters, and/or to predict the time course of the states of interest?

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.

Appendix. Detailed structural identifiability and observability results

The tables included in the following pages report the results of the observability and structural identifiability analyses of all the model variants considered in this paper. Each block of rows represents one of the following assumptions:

- All parameters considered constant (i.e. as is usually the case in the original publications).
- Transmission rate β considered time-varying.
- Latent period κ considered time-varying (only in SEIR models; SIR models do not have this parameter).
- Recovery rate γ considered time-varying.
- All parameters considered time-varying.

Within each block, each row provides detailed information about identifiable and non-identifiable parameters, observable and non-observable states, directly measured (D.M.) states, observable and unobservable unknown inputs (and time-varying parameters), known inputs, and number of derivatives of the unknown inputs (and time-varying parameters) assumed to be non-zero (nnDerW). The suffix *_d number* represents the n^{th} derivative of an unknown function (e.g. β_d1 is the first derivative of the time-varying parameter β).

The blank blocks in the tables of the SEIR models numbers 38 and 8 indicate that the corresponding time-varying case is already considered in the original formulation of the model. The SIR models 29 and 30 have only been studied in their original form, i.e. without considering time-varying parameters, because these models do not contain the common parameters of the SIR models; instead they use the R_0 constant.

		25	21				27		
		h=A	h=I	h=C	h=C, D	h=I, R	h=R, D	h=I, R, D	
Constant parameters	Identifiable	β, σ	β, p, q	β, p, q	β, μ, p, q, r	$\beta, \sigma, \gamma, \delta, k$	FISPO	FISPO	
	Non Identifiable	p, θ	r, μ	r, μ					
	Observable states	y, z	C, S	S, I	S, I	A, S			
	Non Observable states		D, R	D, R	R	D			
	D. M.	A	I	C	C, D	I, R			
	Observable unknown inputs								
	Non Observable unknown inputs	z_d							
	Known inputs					g			
Transmission rate as unknown input	Identifiable	σ	p, q	p, q	All	All			
	Non Identifiable	p, θ		r, μ					
	Observable states	y, z	C, S	S, I	S, I	A, S			
	Non Observable states		D, R	D, R	R	D			
	D. M.	A	I	C	C, D	I, R			
	Observable unknown inputs	β	β	β	β	β, σ			
	Non Observable unknown inputs	z_d							
	Known inputs								
Recovery rate as unknown input	nnDerW	1--2	1	1--2	1--2	1			
	Identifiable	β	p, q, β	p, q, β	All	All			
	Non Identifiable	p, θ	μ	μ					
	Observable states	y, z	C, S	S, I	S, I	A, S			
	Non Observable states		D, R	D, R	R	D			
	D. M.	A	I	C	C, D	I, R			
	Observable unknown inputs	σ	r_d1, r_d2	r_d1, r_d2	r	γ			
	Non Observable unknown inputs	z_d	r	r					
Unknown inputs	Known inputs								
	nnDerW	1--2	1--2	1--2	1--2	1--2			
	Identifiable								
	Non Identifiable								
	Observable states	y, z	C, S	S, I	S, I	A, S			
	Non Observable states		D, R	D, R	R	D			
	D. M.	A	I	C	C, D	I, R			
	Observable unknown inputs	β, σ	β, p, q	β, p, q	β, μ, p, q, r	$\beta, \sigma, \gamma, \delta, k$			
Non Observable unknown inputs	p, θ, z_d	r, μ	r, μ						
Known inputs					g				
nnDerW	1	1	1	1	1				

		19		22	15	26	29	30
		h=R	h=I, R	h=Q	h=τ+pl	h=Q, J	h=I	h=I
Constant parameters	Identifiable	α, β, η, ξ	α, β, η, ξ	d, β, δ	τ	d_1, d_5, eq, k_2, μ_1	FISPO	FISPO
	Non Identifiable			e, γ, α_1, α_2	β, p, μ, d	λ, ea, ej, k1, γ_1, γ_2, μ_2, d2, d3, d4, d6		
	Observable states	J, S, I	J, S	I, S		A, S		
	Non Observable states	U	U	R	I, R, S	I, R		
	D. M.	R	I, R	Q		Q, J		
	Observable unknown inputs							
	Non Observable unknown inputs							
	Known inputs							
Transmission rate as unknown input	Identifiable	All	All	d, δ	τ	d_1, d_5, eq, k_2, μ_1		
	Non Identifiable			e, γ, α_1, α_2	p, μ, d	ea, ej, k1, γ_1, γ_2, μ_2, d2, d3, d4, d6		
	Observable states	I, J, S	J, S	I, S		A, S		
	Non Observable states	U	U	R	I, R, S	I, R		
	D. M.	R	R, I	Q		Q, J		
	Observable unknown inputs	α	α	β				
	Non Observable unknown inputs				β	λ		
	Known inputs							
Recovery rate as unknown input	Identifiable	All	All	d, β, δ	τ, β, μ	d_1, d_5, eq, k_2, μ_1		
	Non Identifiable			α_1, α_2	d	λ, ea, ej, k1, μ_2, d2, d3, d4, d6		
	Observable states	I, J, S	J, S	I, S	I, S	S, A		
	Non Observable states	U	U	R	R	I, R		
	D. M.	R	R	Q		Q, J		
	Observable unknown inputs	β, η	β, η	γ_d1, e_d1, γ_d2, e_d2	p	γ_1, d1, γ_2, d1, γ_1, d2, γ_2, d2		
	Non Observable unknown inputs			γ, e		γ_1, γ_2		
	Known inputs							
Unknown inputs	Identifiable	1	1--2	1--2	1--4	1--2		
	Non Identifiable							
	Observable states	I, J, S	J, S	I, S		A, S		
	Non Observable states	U	U	R	I, R, S	I, R		
	D. M.	R	I, R	Q		Q, J		
	Observable unknown inputs	α, β, η, ξ	α, β, η, ξ	d, β, δ	τ	d_1, d_5, eq, k_2, μ_1		
	Non Observable unknown inputs			e, γ, α_1, α_2	β, p, μ, d	λ, ea, ej, k1, γ_1, γ_2, μ_2, d2, d3, d4, d6		
	Known inputs							
nnDerW	1	1	1	1	1			

		2		34	16	14		61
		h=KC	h=C	h=KI	h=μ I	h=ρ I+τ	h=C, Q, D	h=KI
Constant parameters	Identifiable	γ, κ	k, β, γ	K, ε, μ, γ	ε	ε, τ	α, β, γ, δ, τ	α, γ
	Non Identifiable	β, K		r, β	β, ρ, μ, d	β, ρ, μ, d	λ ₀ , λ ₁ , k ₀ , k ₁	p, β, K
	Observable states		E, I, S	E, I, S			S, E, I	
	Non Observable states	S, E, I, R, C	R	R	S, E, I, R	S, E, I, R	R	S, E, I, R
	D. M.		C				D, Q, C	
	Observable unknown inputs							
	Non Observable unknown inputs							
	Known inputs						λ, k	
Transmission rate as unknown input	Identifiable	γ, κ	All	K, ε, μ, γ	ε	ε, τ	α, γ, δ, τ	α, γ
	Non Identifiable	K		r	ρ, μ, d	ρ, μ, d	λ ₀ , λ ₁ , k ₀ , k ₁	p, K
	Observable states		E, I, S	E, I, S			E, I, S	
	Non Observable states	S, E, I, R, C	R	R	S, E, I, R	S, E, I, R	R	S, E, I, R
	D. M.		C				D, Q, C	
	Observable unknown inputs	β	β				β	
	Non Observable unknown inputs			β	β	β		β
	Known inputs						λ, k	
Latent period as unknown input	nnDerW	1--2	1--2	1	1	1--2	1--2	1--2
	Identifiable	γ	All	K, μ, γ		τ	α, β, δ, τ	γ
	Non Identifiable	β, K		r, β	ρ, μ, d, β	β, ρ, μ, d	λ ₀ , λ ₁ , k ₀ , k ₁	p, K
	Observable states		S, E, I	E, I, S			S, E, I	
	Non Observable states	S, E, I, R, C	R	R	S, E, I, R	S, E, I, R	R	S, E, I, R
	D. M.		C				D, Q, C	
	Observable unknown inputs	κ	κ	ε	ε	ε	γ	α
	Non Observable unknown inputs							
Recovery rate as unknown input	Known inputs						κ, λ	
	nnDerW	1--2	1--2	1	1	1--2	1--2	1--2
	Identifiable	κ	All	K, μ, ε	ε	ε, τ, μ, β	α, β, γ, τ	α
	Non Identifiable	β, K		r, β	μ, d, β	d	λ ₀ , λ ₁ , k ₀ , k ₁	p, K
	Observable states		S, E, I	E, I, S		S, E, I	S, E, I	
	Non Observable states	S, E, I, R, C	R	R	S, E, I, R	R	R	S, E, I, R
	D. M.		C				D, Q, C	
	Observable unknown inputs	γ	γ	γ	ρ _{d1} , ρ _{d2}	ρ	δ	γ
Unknown inputs	Non Observable unknown inputs				ρ			
	Known inputs						κ, λ	
	nnDerW	1--2	1--2	1	1--2	1--2	1--2	1--2
	Identifiable						λ ₀ , λ ₁ , k ₀ , k ₁	
	Non Identifiable						S, E, I	
	Observable states		E, I, S	S, E, I				
	Non Observable states	S, E, I, R, C	R	R	S, E, I, R	S, E, I, R	R	S, E, I, R
	D. M.		C				D, Q, C	
Unknown inputs	Observable unknown inputs	γ, κ	k, β, γ	K, ε, μ, γ	ε	ε, τ	α, β, γ, δ, τ	γ, α
	Non Observable unknown inputs	β, K		r, β	β, ρ, μ, d	β, ρ, μ, d		p, β, K
	Known inputs						λ, k	
	nnDerW	1--2	1	1	1	1	1	1

		38		41	
		h=I, R	h=H, vi I, vr R	h=C	h=J, I
Constant parameters	Identifiable	FISPO	$\theta, \lambda, \delta, q$	$\alpha, \beta, \gamma_1, \gamma_2$	$k, \alpha, \beta, \gamma_1, \gamma_2$
	Non Identifiable		$c, q, \delta, \alpha, \beta, \theta, \lambda, \delta, q$	$\alpha, \beta, \gamma_1, \gamma_2$	$k, \alpha, \beta, \gamma_1, \gamma_2$
	Observable states		$\gamma_i, \gamma_h, \alpha, \delta_i$	q, p	q, p
	Non Observable states		S, E, I	I, J	
	D. M.		H, R	A, E, R, S	A, C, E, R, S
	Observable unknown inputs		Sq, Eq	C	J, I
	Non Observable unknown inputs		σ		
Transmission rate as unknown input	Identifiable	FISPO	$\theta, \lambda, \delta, q, c, q$	$\alpha, \gamma_1, \gamma_2, k$	$\alpha, \gamma_1, \gamma_2, k$
	Non Identifiable		$\delta_i, \alpha, \gamma_h, \gamma_i$	q, p	q, p
	Observable states		S, E, I	I, J	
	Non Observable states		H, R	A, E, R, S	A, C, E, R, S
	D. M.		Sq, Eq	C	J, I
	Observable unknown inputs		β, σ	β	β
	Non Observable unknown inputs				
Latent period as unknown input	Identifiable			$\alpha, \gamma_1, \gamma_2, \beta$	$\alpha, \gamma_1, \gamma_2, \beta$
	Non Identifiable			q, p	q, p
	Observable states			I, J	
	Non Observable states			A, E, R, S	A, C, E, R, S
	D. M.			C	J, I
	Observable unknown inputs			k	k
	Non Observable unknown inputs				
Recovery rate as unknown input	Identifiable	FISPO	$\beta, \delta, q, \lambda, \theta, c, q$	β, α, k	β, α, k
	Non Identifiable		δ_i, α	q, p	q, p
	Observable states		S, E, I	I, J	
	Non Observable states		H, R	A, E, R, S	A, C, E, R, S
	D. M.		Sq, Eq	C	J, I
	Observable unknown inputs		$\sigma, \gamma_i, d1$	γ_1, γ_2	γ_1, γ_2
	Non Observable unknown inputs		γ_h, γ_i		
Unknown inputs	Identifiable				
	Non Identifiable				
	Observable states		S, E, I	I, J, S, A, E	S, A, E
	Non Observable states		H, R	R	C, R
	D. M.		Sq, Eq	C	I, J
	Observable unknown inputs		$\alpha, c, q, \beta, \theta, \lambda, \delta, q$	$k, \alpha, \beta, \gamma_1, \gamma_2, q, p$	$k, \alpha, \beta, \gamma_1, \gamma_2, q, p$
	Non Observable unknown inputs		$\gamma_i, \gamma_h, \alpha, \delta_i$		
	Known inputs				
	nnDerW	1--2	1	1	1

		5	1	3	4	8
		h=l	h=Q	h=Q, R, D	h=S+vs, I+vi, R+vr, P+vp	h=P
Constant parameters	Identifiable	$p, \beta, \gamma, \mu_1, \mu_2$	v, β, γ, ψ	$k, \alpha, \beta, \delta, \gamma$	$\alpha_e, \alpha_i, p, \beta, \mu, \kappa$	$\delta, \sigma, \gamma, h, \gamma_i, \gamma_h, \psi$
	Non Identifiable					α, η
	Observable states	A, S, E	E, I, S	S, E, I	E, I, S	
	Non Observable states	R	R	P	R, P	E, I, S, R
	D. M.	I	Q	Q, R, D		I, H, D
	Observable unknown inputs			k	vi, vs, vr_d1, vp_d1	β
	Non Observable unknown inputs			λ	vr, vp	
Transmission rate as unknown input	Identifiable	p, γ, μ_1, μ_2	All	All		
	Non Identifiable					p, β, μ, κ
	Observable states	A, S, E	E, I, S	S, E, I	S, E, I	
	Non Observable states	R	R	P	P, R	E, I, S, R
	D. M.	I	Q	Q, R, D		P
	Observable unknown inputs	β	β	k, β	vi, vs, $\alpha_e, \alpha_i, vr_d1, vp_d1$	
	Non Observable unknown inputs			λ	vr, vp	α_e, α_i
Latent period as unknown input	Identifiable	p, β, μ_1, μ_2 (All)	All	All		
	Non Identifiable					$\psi, \delta, \gamma_0, \gamma_i, \gamma_h$
	Observable states	A, S, E	S, E, I	S, E, I	S, E, I	α, η
	Non Observable states	R	R	P	R, P	E, I, S, R
	D. M.	I	Q	Q, R, D		A, E, R, S
	Observable unknown inputs	γ	v	γ, k	vi, vs, k, vr_d1, vp_d1, vr_d2, vp_d2	D, H, I
	Non Observable unknown inputs			λ	vp, vr	β, σ
Recovery rate as unknown input	Identifiable	p, β, γ (All)	All	All		
	Non Identifiable					δ, ψ, σ
	Observable states	A, S, E	S, E, I	S, E, I	S, E, I	α, η
	Non Observable states	R	R	P	R, P	E, I, S, R
	D. M.	I	Q	Q, R, D		A, E, R, S
	Observable unknown inputs	μ_1, μ_2	ψ	δ, k	vi, vs, p, $\beta, vr_d1, vr_d2, vp_d1, vp_d2$	D, H, I
	Non Observable unknown inputs			λ	vp, vr	$\beta, \gamma_i, \gamma_0, \gamma_h$
Unknown inputs	Identifiable					
	Non Identifiable					$\alpha_e, \alpha_i, \mu, \kappa$
	Observable states	A, S, E	E, I, S	S, E, I	S, E, I	α, η
	Non Observable states	R	R	P	R, P	E, I, S, R
	D. M.	I	Q	Q, R, D		A, E, R, S
	Observable unknown inputs	$p, \beta, \gamma, \mu_1, \mu_2$	v, β, γ, ψ	$k, \alpha, \beta, \delta, \gamma$	$\alpha_e, \alpha_i, p, \beta, \mu, \kappa, vi, vs, vr_d1, vp_d1$	$\beta, \delta, \gamma_0, \gamma_i, \gamma_h, \psi, \sigma$
	Non Observable unknown inputs			λ	vr, vp	α, η

		39	51
		h=F, Di, De	h=Di, F
Constant parameters	Identifiable	$\mu_d, \sigma_d, \gamma_d$	$\mu_d, \sigma_d, \gamma_d$
	Non Identifiable	$q, \mu_i, \sigma, \theta_e, \theta_i, \gamma, \beta, \delta, \psi, \psi_e$	$q, \mu_i, \sigma, \theta_e, \theta_i, \gamma, \beta, \delta, \psi, \psi_e$
	Observable states	E, I, R, S	S, E, I
	Non Observable states	R	R
	D. M.	De, Di, F	Di, F
	Observable unknown inputs		
	Non Observable unknown inputs		
Transmission rate as unknown input	Identifiable	$\mu_d, \sigma_d, \gamma_d$	$\mu_d, \sigma_d, \gamma_d$
	Non Identifiable	$q, \mu_i, \sigma, \theta_e, \theta_i, \gamma, \psi, \psi_e$	$q, \mu_i, \sigma, \theta_e, \theta_i, \gamma, \psi, \psi_e$
	Observable states	De	De
	Non Observable states	R	R
	D. M.	De, Di, F	Di, F
	Observable unknown inputs	β, β_d	β, β_d
	Non Observable unknown inputs		
Latent period as unknown input	Identifiable	$\gamma_d, \mu_d, \beta, \mu_i, \gamma$	$\gamma_d, \mu_d, \gamma, \mu_i, \beta$
	Non Identifiable	$q, \theta_e, \theta_i, \beta, \delta, \psi, \psi_e$	$q, \theta_e, \theta_i, \beta, \delta, \psi, \psi_e$
	Observable states	E, I, S	S, E, I, De
	Non Observable states	R	R
	D. M.	De, Di, F	Di, F
	Observable unknown inputs	σ_d, σ	σ, σ_d
	Non Observable unknown inputs		
Recovery rate as unknown input	Identifiable	μ_d, σ_d	μ_d, σ_d
	Non Identifiable	$q, \mu_i, \sigma, \theta_e, \theta_i, \beta, \beta_d, \psi, \psi_e, \sigma$	$q, \mu_i, \sigma, \theta_e, \theta_i, \beta, \beta_d, \psi, \psi_e, \sigma$
	Observable states	De	De
	Non Observable states	R	R
	D. M.	De, Di, F	Di, F
	Observable unknown inputs	$\gamma_d, \gamma_{d1}, \gamma_{d2}$	$\gamma_d, \gamma_{d1}, \gamma_{d2}$
	Non Observable unknown inputs	γ	γ
Unknown inputs	Identifiable		
	Non Identifiable		
	Observable states	S, E, I	S, E, I
	Non Observable states	R	R
	D. M.	De, Di, F	Di, F
	Observable unknown inputs	$\mu_d, \sigma_d, \gamma_d, \mu_i, \sigma, \gamma, \beta$	$\mu_d, \sigma_d, \gamma_d, \mu_i, \sigma, \gamma, \beta$
	Non Observable unknown inputs	$q, \theta_e, \theta_i, \beta, \delta, \psi, \psi_e$	$q, \theta_e, \theta_i, \beta, \delta, \psi, \psi_e$

		17			32	33
		h=D	h=Q, D	h=D, I, Q	h=WL	h=L, Q
Constant parameters	Identifiable	q, ϵ	d, q, qt, β , ϵ , γ	d, q, qt, β , ϵ , γ	W, β , α , μ , γ	γ , $\beta 1$, η , δ , ξ , $\theta 2$, ϵ , $\theta 1$, $\alpha 1$, $\alpha 2$
	Non Identifiable	β , γ , d, qt				
	Observable states		E, I, S	E, S	I, S, L	E, I, S
	Non Observable states	E, I, R, Q, S	R	R	R	R
	D. M.	D	Q, D	Q, D, I		L, Q
	Observable unknown inputs					
	Non Observable unknown inputs					
	Known inputs					
Transmission rate as unknown input	Identifiable	q, ϵ	All	All	All	All
	Non Identifiable	γ , d, qt				
	Observable states		E, I, S	E, S	I, S, L	E, I, S
	Non Observable states	E, I, R, Q, S	R	R	R	R
	D. M.	D	Q, D	Q, D, I		L, Q
	Observable unknown inputs		β	β	β	β
	Non Observable unknown inputs	β				
	Known inputs					
Latent period as unknown input	nnDerW	1	1	1	1	1
	Identifiable	q	All	All	All	All
	Non Identifiable	β , γ , d, qt				
	Observable states		E, I, S	E, S	I, S, L	E, I, S
	Non Observable states	E, I, R, Q, S	R	R	R	R
	D. M.	D	Q, D	Q, D, I		L, Q
	Observable unknown inputs	ϵ	ϵ	ϵ	W	ϵ
	Non Observable unknown inputs					
Recovery rate as unknown input	Known inputs					
	nnDerW	1--2	1--2	1--2	1	1
	Identifiable	q, ϵ	All	All	All	All
	Non Identifiable	β , d				
	Observable states		E, I, S	E, S	I, S, L	E, I, S
	Non Observable states	E, I, R, Q, S	R	R	R	R
	D. M.	D	Q, D	Q, D, I		L, Q
	Observable unknown inputs	$\gamma_d 1$, qt_d1, $\gamma_d 2$, qt_d2	γ , qt	γ , qt	γ	$\alpha 1$, $\alpha 2$
Non Observable unknown inputs	γ , qt					
Unknown inputs	Known inputs					
	nnDerW	1--2	1--2	1--2	1	1
	Identifiable					
	Non Identifiable					
	Observable states		E, I, S	E, S		E, I, S
	Non Observable states	E, I, R, Q, S	R	R	I, S, L	R
	D. M.	D	Q, D	Q, D, I	R	L, Q
	Observable unknown inputs	q, ϵ	d, q, qt, β , ϵ , γ	d, q, qt, β , ϵ , γ	W, β , α , μ , γ	γ , $\beta 1$, η , δ , ξ , $\theta 2$, ϵ , $\theta 1$, $\alpha 1$, $\alpha 2$

References

- Acemoglu, D., Chernozhukov, V., Werning, I., & Whinston, M. D. (2020). A multi-risk SIR model with optimally targeted lockdown. National Bureau of Economic Research.
- Adam, D. (2020). Special report: The simulations driving the world's response to COVID-19. *Nature*, 580(7803), 316.
- Alahmadi, A., Belet, S., Black, A., Cromer, D., Flegg, J. A., House, T., et al. (2020). Influencing public health policy with data-informed mathematical models of infectious diseases: Recent developments and new challenges. *Epidemics*, Article 100393.
- Alvarez, F. E., Argente, D., & Lippi, F. (2020). A simple planning problem for COVID-19 lockdown. National Bureau of Economic Research.
- Angelova, M., Karlsson, J., & Jirstrand, M. (2012). Minimal output sets for identifiability. *Mathematical Biosciences*, 239(1), 139–153.
- Arriola, L., & Hyman, J. M. (2009). Sensitivity analysis for uncertainty quantification in mathematical models. In *Mathematical and statistical estimation approaches in epidemiology* (pp. 195–247). Springer.
- Brauer, F. (2008). Compartmental models in epidemiology. In *Mathematical epidemiology* (pp. 19–79). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Brauer, F., van den Driessche, P., & Wu, J. (Eds.). (2008). *Mathematical epidemiology*. Springer Berlin Heidelberg.
- Brouwer, A. F., Meza, R., & Eisenberg, M. C. (2019). Integrating measures of viral prevalence and seroprevalence: a mechanistic modelling approach to explaining cohort patterns of human papillomavirus in women in the USA. *Philosophical Transactions of the Royal Society B*, 374(1773), Article 20180297.
- Casella, F. (2020). Can the COVID-19 epidemic be controlled on the basis of daily test reports? arXiv preprint arXiv:2003.06967.
- Castro, M., Ares, S., Cuesta, J. A., & Manrubia, S. (2020). Predictability: Can the turning point and end of an expanding epidemic be precisely forecast? arXiv:2004.08842.
- Chapman, J. D., & Evans, N. D. (2009). The structural identifiability of susceptible-infective-recovered type epidemic models with incomplete immunity and birth targeted vaccination. *Biomedical Signal Processing and Control*, 4(4), 278–284.
- Chatterjee, K., Chatterjee, K., Kumar, A., & Shankar, S. (2020). Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. *Medical Journal Armed Forces India*, 76(2), 147–155.
- Chen, Y.-C., Lu, P.-E., Chang, C.-S., & Liu, T.-H. (2020). A time-dependent SIR model for COVID-19 with undetectable infected persons. arXiv:2003.00122.
- Chitnis, N. (2017). Introduction to SEIR models. In *Workshop on mathematical models of climate variability, environmental change and infectious diseases, Trieste, Italy*.
- Chowell, G. (2017). Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3), 379–398.
- Cobey, S. (2020). Modeling infectious disease dynamics. *Science*, 368(6492), 713–714.
- Currie, C. S., Fowler, J. W., Kotiadis, K., Monks, T., Onggo, B. S., Robertson, D. A., et al. (2020). How simulation modelling can help reduce the impact of COVID-19. *Journal of Simulation*, 1–15.
- DiStefano III, J. (2015). *Dynamic systems biology modeling and simulation*. Academic Press.
- Dohare, R. (2020). Mathematical model of transmission dynamics with mitigation and health measures for SARS-CoV-2 infection in European countries. *Research Square*.
- Eikenberry, S. E., Mancuso, M., Iboi, E., Phan, T., Eikenberry, K., Kuang, Y., et al. (2020). To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infectious Disease Modelling*, 5, 293–308.
- Eisenberg, M. C., Robertson, S. L., & Tien, J. H. (2013). Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease. *Journal of Theoretical Biology*, 324, 84–102.
- Evans, N., Chapman, M., Chappell, M., & Godfrey, K. (2002). The structural identifiability of a general epidemic (SIR) model with seasonal forcing. *IFAC Proceedings Volumes*, 35(1), 109–114.
- Evans, N. D., & Chappell, M. J. (2000). Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems. *Mathematical Biosciences*, 168(2), 137–159.
- Evans, N. D., White, L. J., Chapman, M. J., Godfrey, K. R., & Chappell, M. J. (2005). The structural identifiability of the susceptible infected recovered model with seasonal forcing. *Mathematical Biosciences*, 194(2), 175–197.
- Faranda, D., Castillo, I. P., Hulme, O., Jezequel, A., Lamb, J. S., Sato, Y., et al. (2020). Asymptotic estimates of SARS-CoV-2 infection counts and their sensitivity to stochastic perturbation. *Chaos. An Interdisciplinary Journal of Nonlinear Science*, 30(5), Article 051107.
- Fosu, G., Opong, J., & Appati, J. (2020). Construction of compartmental models for COVID-19 with quarantine, lockdown and vaccine interventions. *SSRN Electronic Journal*.
- Franco, E. (2020). A feedback SIR (fSIR) model highlights advantages and limitations of infection-based social distancing. arXiv:2004.13216.
- Gaeta, G. (2020). A simple SIR model with a large set of asymptomatic infectives. arXiv:2003.08720.
- Gallina, R. (2012). Dynamic models for the analysis of epidemic spreads. [Ph.D. thesis], Università degli studi di Padova.
- Gevertz, J., Greene, J., Hixahuary Sanchez Tapia, C., & Sontag, E. D. (2020). A novel COVID-19 epidemiological model with explicit susceptible and asymptomatic isolation compartments reveals unexpected consequences of timing social distancing. medRxiv, Cold Spring Harbor Laboratory Press.
- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., et al. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*.
- Hengl, S., Kreutz, C., Timmer, J., & Maiwald, T. (2007). Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics*, 23(19), 2612–2618.
- Hermann, R., & Krener, A. J. (1977). Nonlinear controllability and observability. *IEEE Transactions on Automatic Control*, 22(5), 728–740.
- Hethcote, H., Zhién, M., & Shengbing, L. (2002). Effects of quarantine in six endemic models for infectious diseases. *Mathematical Biosciences*, 180(1), 141–160.
- Holmdahl, I., & Buckee, C. (2020). Wrong but useful—what COVID-19 epidemiologic models can and cannot tell us. *New England Journal of Medicine*.
- Hong, H., Ovchinnikov, A., Pogudin, G., & Yap, C. (2019). SIAN: software for structural identifiability analysis of ode models. *Bioinformatics*, 35(16), 2873–2874.
- Hubbs, C. (2020). Social distancing to slow the coronavirus. <https://towardsdatascience.com/social-distancing-to-slow-the-coronavirus-768292f04296>.
- Janzén, D. L., Bergenholm, L., Jirstrand, M., Parkinson, J., Yates, J., Evans, N. D., et al. (2016). Parameter identifiability of fundamental pharmacodynamic models. *Frontiers in Physiology*, 7, 590.
- Jia, J., Ding, J., Liu, S., Liao, G., Li, J., Duan, B., et al. (2020). Modeling the control of COVID-19: impact of policy interventions and meteorological factors. *Electronic Journal of Differential Equations*, 2020, 1–24.
- Kao, Y.-H., & Eisenberg, M. C. (2018). Practical unidentifiability of a simple vector-borne disease model: Implications for parameter estimation and intervention assessment. *Epidemics*, 25, 89–100.
- Karlsson, J., Angelova, M., & Jirstrand, M. (2012). An efficient method for structural identifiability analysis of large dynamic systems. In *16th IFAC symposium on system identification*, vol. 16 (pp. 941–946).
- Kim, M., & Milner, F. (1995). A mathematical model of epidemics with screening and variable infectivity. *Mathematical and Computer Modelling*, 21(7), 29–42.
- Köhler, J., Schwenkel, L., Koch, A., Berberich, J., Pauli, P., & Allgöwer, F. (2020). Robust and optimal predictive control of the COVID-19 outbreak. arXiv preprint arXiv:2005.03580.
- Li, M. Y. (2018). *An introduction to mathematical modeling of infectious diseases*, vol. 2. Springer.
- Li, F., & Zhao, X.-Q. (2019). A periodic SEIRS epidemic model with a time-dependent latent period. *Journal of Mathematical Biology*, 78.
- Liangrong, P., Yang, W., Zhang, D., Zhuge, C., & Hong, L. (2020). Epidemic analysis of COVID-19 in China by dynamical modeling. arXiv:2002.06563.
- Ligon, T. S., Fröhlich, F., Chiş, O. T., Banga, J. R., Balsa-Canto, E., & Hasenauer, J. (2018). Gensi 2.0: multi-experiment structural identifiability analysis of SBML models. *Bioinformatics*, 34(8), 1421–1423.
- Lofgren, E. T., Halloran, M. E., Rivers, C. M., Drake, J. M., Porco, T. C., Lewis, B., et al. (2014). Opinion: Mathematical models: A key tool for outbreak response. *Proceedings of the National Academy of Sciences*, 111(51), 18095–18096.
- Lopez, L. R., & Rodo, X. (2020). A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics. Cold Spring Harbor Laboratory Press.
- Lourenco, J., Paton, R., Ghafari, M., Kraemer, M., Thompson, C., Simmonds, P., et al. (2020). Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic.
- Maier, B. F., & Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. 368 (6492) 742–746, American Association for the Advancement of Science.
- Martcheva, M. (2015). *An introduction to mathematical epidemiology*, vol. 61. Springer.
- Massonis, G., & F. Villaverde, A. (2020). Finding and breaking Lie symmetries: Implications for structural identifiability and observability in biological modelling. *Symmetry*, 12, 469.
- McGee, R. S. (2020). Models of SEIRS epidemic dynamics with extensions, including network-structured populations, testing, contact tracing, and social distancing. *GitHub repository*. GitHub, <https://github.com/ryansmcgee/seirplus>.
- Merk, B., Timmer, J., & Kaschek, D. (2015). Higher-order Lie symmetries in identifiability and predictability analysis of dynamic models. *Physical Review E*, 92(1), Article 012920.
- Meshkat, N., Kuo, C. E.-z., & DiStefano III, J. (2014). On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and COMBOS: A novel web implementation. *PLoS One*, 9(10).
- Prague, M., Wittkop, L., Clairon, Q., Dutartre, D., Thiebaut, R., & Hejblum, B. P. (2020). Population modeling of early COVID-19 epidemic dynamics in french regions and estimation of the lockdown impact on infection rate. medRxiv.
- Pribylova, L., & Hajnova, V. (2020). SEIAR model with asymptomatic cohort and consequences to efficiency of quarantine government measures in COVID-19 epidemic. arXiv:2004.02601.

- Rahman, A., & Kuddus, M. A. (2020). Modelling the transmission dynamics of COVID-19 in six high burden countries. Cold Spring Harbor Laboratory Press.
- Raimúndez, E., Dudkin, E., Vanhoefer, J., Alamoudi, E., Merkt, S., Fuhrmann, L., et al. (2020). COVID-19 outbreak in wuhan demonstrates the limitations of publicly available case numbers for epidemiological modelling. *medRxiv*.
- Roda, W. C., Varughese, M. B., Han, D., & Li, M. Y. (2020). Why is it difficult to accurately predict the COVID-19 epidemic?. *Infectious Disease Modelling*, 5, 271–281.
- Roosa, K., & Chowell, G. (2019). Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. *Theoretical Biology and Medical Modelling*, 16(1), 1.
- Saccomani, M., Bellu, G., Audoly, S., & d'Angio, L. (2019). A new version of DAISY to test structural identifiability of biological models. In *International conference on computational methods in systems biology* (pp. 329–334). Springer.
- Sameni, R. (2020). Mathematical modeling of epidemic diseases; a case study of the COVID-19 coronavirus. *arXiv preprint arXiv:2003.11371*.
- Scarpino, S. V., & Petri, G. (2019). On the predictability of infectious disease outbreaks. *Nature Communications*, 10(1), 1–8.
- Sedoglavic, A. (2002). A probabilistic algorithm to test local algebraic observability in polynomial time. *Journal of Symbolic Computation*, 33, 735–755.
- Shi, P., Cao, S., & Feng, P. (2020). SEIR transmission dynamics model of 2019 nCoV coronavirus with considering the weak infectious ability and changes in latency duration. Cold Spring Harbor Laboratory Press.
- Thomaseth, K., & Saccomani, M. P. (2018). Local identifiability analysis of nonlinear ODE models: how to determine all candidate solutions. *IFAC-PapersOnLine*, 51(2), 529–534.
- Tunali, E. T., & Tarn, T.-J. (1987). New results for identifiability of nonlinear systems. *IEEE Transactions on Automatic Control*, 32(2), 146–154.
- Tuncer, N., & Le, T. T. (2018). Structural and practical identifiability analysis of outbreak models. *Mathematical Biosciences*, 299, 1–18.
- Villaverde, A. F. (2019). Observability and structural identifiability of nonlinear biological systems. *Complexity*, 2019, Article 8497093.
- Villaverde, A. F., Barreiro, A., & Papachristodoulou, A. (2016). Structural identifiability of dynamic systems biology models. *PLoS Computational Biology*, 12(10), Article e1005153.
- Villaverde, A. F., Tsiantis, N., & Banga, J. R. (2019). Full observability and estimation of unknown inputs, states, and parameters of nonlinear biological models. *Journal of the Royal Society Interface*, 16, Article 20190043.
- Yates, J. W., Evans, N. D., & Chappell, M. J. (2009). Structural identifiability analysis via symmetries of differential equations. *Automatica*, 45(11), 2585–2591.
- Zha, W.-t., Pang, F.-r., Zhou, N., Wu, B., Liu, Y., Du, Y.-b., et al. (2020). Research about the optimal strategies for prevention and control of varicella outbreak in a school in a central city of China: based on an SEIR dynamic model. *Epidemiology and Infection*, 148.
- Zheng, W. (2020). Total variation regularization for compartmental epidemic models with time-varying dynamics. *arXiv:2004.00412*.