NETWORK SCIENCE

Lack of practical identifiability may hamper reliable predictions in COVID-19 epidemic models

Luca Gallo[1,2], Mattia Frasca[3,4]*, Vito Latora[1,2,5,6], Giovanni Russo[7]

Compartmental models are widely adopted to describe and predict the spreading of infectious diseases. The unknown parameters of these models need to be estimated from the data. Furthermore, when some of the model variables are not empirically accessible, as in the case of asymptomatic carriers of coronavirus disease 2019 (COVID-19), they have to be obtained as an outcome of the model. Here, we introduce a framework to quantify how the uncertainty in the data affects the determination of the parameters and the evolution of the unmeasured variables of a given model. We illustrate how the method is able to characterize different regimes of identifiability, even in models with few compartments. Last, we discuss how the lack of identifiability in a realistic model for COVID-19 may prevent reliable predictions of the epidemic dynamics.

INTRODUCTION

The pandemic caused by severe acute respiratory syndrome coronavirus-2 is challenging humanity in an unprecedented way (*1*), with the disease, which in a few months has spread around the world, affecting large parts of the population (*2, 3*) and often requiring hospitalization or even intensive care (*4, 5*). Mitigating the impact of coronavirus disease 2019 (COVID-19) urges synergistic efforts to understand, predict, and control the many, often elusive, facets of the complex phenomenon of the spreading of a previously unknown virus, from RNA sequencing to the study of the virus pathogenicity and transmissibility (*6, 7*) to the definition of suitable epidemic spreading models (*8*) and the investigation of nonpharmaceutical intervention policies and containment measures (*9–12*). In particular, a large number of epidemic models have recently been proposed to describe the evolution of COVID-19 and evaluate the effectiveness of different counteracting measures, including social distancing, testing, and contact tracing (*13–19*). However, even the adoption of well-consolidated modeling techniques, such as the use of mechanistic models at the population level based on compartments, poses fundamental problems. First of all, the very same choice of the dynamical variables to use in a compartmental model is crucial; as such, variables should adequately capture the spreading mechanisms and need to be tailored to the specific disease. This step is not straightforward, especially when the spreading mechanisms of the disease are still unknown or only partially identified. In addition, some of the variables considered might be difficult to measure and track as, for instance, in the case of COVID-19, it occurred in the number of individuals showing mild or no symptoms. Second, compartmental models, usually, involve a number of parameters, including the initial values of the unmeasured variables, which are not known and need to be estimated from data.

Having at disposal a large amount of data, unfortunately, does not simplify the problem of parameter estimation and prediction of unmeasured states. Once a model is formulated, it may occur that some of its unknown parameters are intrinsically impossible to determine from the measured variables or that they are numerically very sensitive to the measurements themselves. In the first case, it is the very same structure of the model to hamper parameter estimation as the system admits infinitely many sets of parameters that fit the data equally well; for this reason, this problem is referred to as structural identifiability (*20, 21*). In the second case, although under ideal conditions (i.e., noise-free data and error-free models) the problem of parameter estimation can be uniquely solved for some trajectories, it may be numerically ill conditioned, such that from a practical point of view, the parameters cannot be determined with precision even if the model is structurally identifiable (*22*). This situation typically occurs when large changes in the parameters entail a small variation of the measured variables, such that two similar trajectories may correspond to very different parameters (*23*). The term practical identifiability is adopted in this case.

Identifiability in general represents an important property of a dynamical system, as in a nonidentifiable system, different sets of parameters can produce the same or very similar fits of the data. Consequently, predictions from a nonidentifiable system become unreliable. In the context of epidemics forecasting, this means that even if the model considered is able to reproduce the measured variables, a large uncertainty may affect the estimated values of the parameters and the predicted evolution of the unmeasured variables (*24*).

The problem of practical identifiability of model parameters has been investigated using different methodologies based on Fisher's information theory (*25, 26*), profile likelihood (*27*), Monte Carlo simulations (*28*), and other computational approaches (*29*). However, the lack of practical identifiability can also affect the reliability of the prediction of the unmeasured variables dynamics (*27*), an issue of utmost importance in the context of COVID-19, which nevertheless still requires a systematic investigation. In particular, an approach to simultaneously characterize the problem of sensitivity to parameters and that of the reliability of predictions of unmeasured variables is still missing.

In more detail, in this paper, we investigate the problem of the practical identifiability of dynamical systems whose state includes

[1]Department of Physics and Astronomy, University of Catania, Catania 95125, Italy. [2]INFN Sezione di Catania, Via S. Sofia, 64, Catania 95125, Italy. [3]Department of Electrical, Electronics and Computer Science Engineering, University of Catania, Catania 95125, Italy. [4]Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti," Consiglio Nazionale delle Ricerche (IASI-CNR), 00185 Roma 00185, Italy. [5]School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, UK [6]Complexity Science Hub Vienna, A-1080 Vienna, Austria. [7]Department of Mathematics and Computer Science, University of Catania, Catania 95125, Italy.
*Corresponding author. Email: mattia.frasca@dieei.unict.it

not only measurable but also hidden variables, as is the case of compartment models for COVID-19 epidemic. We present a general framework to quantify not only the sensitivity of the measured variables of a given model on its parameters but also the sensitivity of the unmeasured variables on the parameters and on the measured variables. This will allow us to introduce the notion of practical identifiability of the hidden variables of a model. As a relevant and timely application, we show the variety of different regimes and levels of identifiability that can appear in epidemic models, even in the simplest case of a four compartment system. Last, we study the actual effects of the lack of practical identifiability in more sophisticated models introduced for COVID-19.

RESULTS
Dynamical systems with hidden variables
Consider the n-dimensional dynamical system described by the following equations

$$\dot{\mathbf{m}} = \mathbf{f}(\mathbf{m}, \mathbf{h}, \mathbf{q}),$$
$$\dot{\mathbf{h}} = \mathbf{g}(\mathbf{m}, \mathbf{h}, \mathbf{q}) \tag{1}$$

where we have partitioned the state variables into two sets: the variables $\mathbf{m} \in \mathbb{R}^{n_m}$ that can be empirically accessed (measurable variables) and those $\mathbf{h} \in \mathbb{R}^{n_h}$, with $n_m + n_h = n$, that cannot be measured (hidden).

The dynamics of the system is governed by the two Lipschitz-continuous functions $\mathbf{f}$ and $\mathbf{g}$, which also depend on a vector of structural parameters $\mathbf{q} \in \Omega_q \subset \mathbb{R}^{n_q}$. The trajectories $\mathbf{m}(t)$ and $\mathbf{h}(t)$ of system in eq. 1 are uniquely determined by the structural parameters $\mathbf{q}$ and by the initial conditions $\mathbf{m}(0) = \mathbf{m_0}$, $\mathbf{h}(0) = \mathbf{h_0}$. Here, we assume that some of the quantities $\mathbf{q}$ are known, while the others are not known and need to be determined by fitting the trajectories of measurable variables $\mathbf{m}(t)$.

We denote by $\mathbf{p} \in \Omega_p \subset \mathbb{R}^{n_p}$, the set of unknown parameters that identify the trajectories, which comprises the unknown terms of $\mathbf{q}$ and the unknown initial conditions $\mathbf{h_0}$. The initial values of the hidden variables are not known and act indeed as parameters for the trajectories generated by system in eq. 1. The initial conditions of the measurable variables $\mathbf{m_0}$ may be considered fitting parameters as well.

System in eq. 1 is said to be structurally identifiable when the measured variables satisfy [21]

$$\mathbf{m}(t, \hat{\mathbf{p}}) = \mathbf{m}(t, \mathbf{p}), \forall t \geq 0 \Rightarrow \hat{\mathbf{p}} = \mathbf{p} \tag{2}$$

for almost any $\mathbf{p} \in \Omega_p$. Notice that, as a consequence of the existence and uniqueness theorem for the initial value problem, if system in eq. 1 is structurally identifiable, the hidden variables can also be uniquely determined.

Structural identifiability guarantees that two different sets of parameters do not lead to the same time course for the measured variables. When this condition is not met, one cannot uniquely associate a data fit to a specific set of parameters or, equivalently, recover the parameters from the measured variables [23].

Assessing the practical identifiability of a model
Structural identifiability, however, is a necessary but not sufficient condition for parameters estimation, so that when it comes to use a dynamical system as a model of a real phenomenon, it is fundamental to quantify the practical identifiability of the dynamical system.

To do this, we consider a solution, $\overline{\mathbf{m}}(t) = \mathbf{m}(t, \overline{\mathbf{p}})$ and $\overline{\mathbf{h}}(t) = \mathbf{h}(t, \overline{\mathbf{p}})$, obtained from parameters $\mathbf{p} = \overline{\mathbf{p}}$, and we explore how much the functions $\mathbf{m}(t)$ and $\mathbf{h}(t)$ change as we vary the parameters $\overline{\mathbf{p}}$ by a small amount $\boldsymbol{\delta p}$. To first order approximation in the perturbation of the parameters, we have $\boldsymbol{\delta m} = \frac{\partial \mathbf{m}}{\partial \mathbf{p}} \boldsymbol{\delta p} + \mathcal{O}(\|\boldsymbol{\delta p}\|^2)$ and $\boldsymbol{\delta h} = \frac{\partial \mathbf{h}}{\partial \mathbf{p}} \boldsymbol{\delta p} + \mathcal{O}(\|\boldsymbol{\delta p}\|^2)$.

Hence, by dropping the higher order terms, we have $\|\boldsymbol{\delta m}\|^2 = \int_0^\infty |\boldsymbol{\delta m}|^2 dt = \boldsymbol{\delta p}^T \mathrm{M} \boldsymbol{\delta p}$ and $\|\boldsymbol{\delta h}\|^2 = \int_0^\infty |\boldsymbol{\delta h}|^2 dt = \boldsymbol{\delta p}^T \mathrm{H} \boldsymbol{\delta p}$, where the entries of the sensitivity matrices $\mathrm{M} = \mathrm{M}(\overline{\mathbf{p}}) \in \mathbb{R}^{n_p \times n_p}$ and $\mathrm{H} = \mathrm{H}(\overline{\mathbf{p}}) \in \mathbb{R}^{n_p \times n_p}$ for the measured and unmeasured variables are defined as

$$(\mathrm{M})_{ij} = \int_0^\infty \frac{\partial \mathbf{m}^T}{\partial p_i} \frac{\partial \mathbf{m}}{\partial p_j} dt; (\mathrm{H})_{ij} = \int_0^\infty \frac{\partial \mathbf{h}^T}{\partial p_i} \frac{\partial \mathbf{h}}{\partial p_j} dt \tag{3}$$

Note that these matrices are positive semidefinite by construction. The smallest change in the measured variables $\mathbf{m}(t)$ will take place if $\boldsymbol{\delta p}$ is aligned along the eigenvector $\mathbf{v_1}$ of M corresponding to the smallest eigenvalue $\lambda_1(\mathrm{M})$. Hence, we can consider $\sigma = \sqrt{\lambda_1(\mathrm{M})}$ to quantify the sensitivity of the measured variables to the parameters. Practical identifiability requires high values of σ as these indicate cases where small changes in the parameters may produce considerable variations of the measurable variables, and therefore, the estimation of the model parameters from fitting is more reliable.

Suppose now we consider a perturbation, $\boldsymbol{\delta p_1}$, of the parameters aligned along the direction of $\mathbf{v_1}$. We can evaluate the change in $\mathbf{h}(t)$ due to this perturbation by

$$\eta^2 = \frac{\boldsymbol{\delta p_1}^T \mathrm{H} \boldsymbol{\delta p_1}}{\boldsymbol{\delta p_1}^T \boldsymbol{\delta p_1}} \tag{4}$$

The value of η quantifies the sensitivity of the hidden variables to the parameters of the model, when these parameters are estimated from the fitting of the observed variables since $\|\boldsymbol{\delta h}\| = \eta \|\boldsymbol{\delta p_1}\|$.

Notice that in this case, and differently from σ, lower values of η are desirable because they imply a better prediction on the hidden variables.

Last, with the help of the sensitivity matrices defined above, we can also evaluate the sensitivity of the hidden variables to the measured variables as

$$\mu^2 = \max_{\|\boldsymbol{\delta p}\|=1} \frac{\boldsymbol{\delta p}^T \mathrm{H} \boldsymbol{\delta p}}{\boldsymbol{\delta p}^T \mathrm{M} \boldsymbol{\delta p}} \tag{5}$$

This parameter is of particular relevance here, since it provides a bound on how the uncertainty on the measured variables affects the evolution of the hidden variables. In addition, the parameter μ^2 can be efficiently computed as it corresponds to the maximum generalized eigenvalue of matrices (H, M), as shown in Materials and Methods.

The sensitivity matrices are useful in studying the effect of changing the number of hidden variables and unknown parameters on the practical identifiability of a model. Assume that we have access to one more variable, thus effectively increasing the size of the set of measured variables to $n_m' = n_m + 1$ and, correspondingly, reducing that of the unmeasured variables to $n_h' = n_h - 1$. This corresponds to considering new variables $\mathbf{m}'$ and $\mathbf{h}'$. From the definition in Eq. 3, the new sensitivity matrix can be written as $\mathrm{M}' = \mathrm{M} + \mathrm{M}_1$, where M_1 is the sensitivity matrix for the newly measured variable. Given Weyl's inequality [page 239 of [30]], we have that

$\lambda_1(M') \geq \lambda_1(M) + \lambda_1(M_1)$ and since M_1 is also positive semidefinite, $\lambda_1(M') \geq \lambda_1(M)$. This means that measuring one further variable (or more than one) of the system increases the practical identifiability of a model, as expected. As $H' = H - M_1$, it is also possible to demonstrate that $\mu(M') \leq \mu(M)$ (see Materials and Methods). Let us now consider a different scenario: Suppose we have a priori knowledge of one of the model parameters so that we do not need to estimate its value by fitting the model to the data. In this case, we can define new sensitivity matrices $\widetilde{M}, \widetilde{H} \in \mathbb{R}^{(n_p-1)\times(n_p-1)}$ for the measured and unmeasured variables, respectively. Given the Cauchy's interlacing theorem [page 242 of (30)], we have that $\lambda_1(\widetilde{M}) \geq \lambda_1(M)$, which implies that practical identifiability is improved by acquiring a priori information on some of the model parameters. For instance, in the context of COVID-19 models, one may decide to fix some of the parameters, such as the rate of recovery, to values derived from medical and biological knowledge (24, 31–33) and to determine from fitting the more elusive parameters, such as the percentage of asymptomatic individuals or the rates of transmission.

The sensitivity measures we have introduced point out that prior knowledge of some of the parameters, or a larger set of measurable variables, reduces the sensitivity of the measured variables to the parameters and that of the hidden variables to a variation in the measured ones. However, gathering further knowledge can be difficult or even not possible so that these results, which should not be interpreted as an oversimplified solution to the problem of identifiability, have to be considered in the light of practical issues that might arise in the measurement of the model variables and parameters.

The sensitivity measures reveal different regimes of identifiability

As a first application, we study the practical identifiability of a four compartment mean-field epidemic model (34) in the class of SIAR models (35), developed to assess the impact of asymptomatic carriers of COVID-19 (8, 36, 37) and other diseases (38–40). In such a model (Fig. 1), a susceptible individual (S) can be infected by an infectious individual who can either be symptomatic (I) or asymptomatic (A). The newly infected individual can either be symptomatic ($S \rightarrow I$) or asymptomatic ($S \rightarrow A$). Furthermore, we also consider the possibility that asymptomatic individuals develop symptoms ($A \rightarrow I$), thus accounting for the cases in which an individual can infect before and after the onset of the symptoms (41). Last, we suppose that individuals cannot be reinfected as they acquire a permanent immunity (R).

One of the crucial aspects of COVID-19 is the presence of asymptomatic individuals who are difficult to trace as the individuals themselves could be unaware about their state. Consequently, we assume that the fraction of asymptomatic individuals, $a(t)$, is not measurable, while the fractions of symptomatic, $\iota(t)$, and recovered, $r(t)$, are measured variables, that is, $\mathbf{m} \equiv [\iota, r]$ and $\mathbf{h} \equiv [s, a]$. As mentioned above, practical identifiability is a property of the trajectories of the system, which are uniquely determined by the values of the unknown parameters $\mathbf{p}$. Here, we illustrate how the sensitivity of both measured and unmeasured variables changes with the probability γ that a newly infected individual shows no symptoms when all the other parameters of the model are fixed (to the values reported in Materials and Methods). Concerning the choice of vector $\mathbf{p}$, here, we consider the following case. First, as the number of symptomatic infectious and recovered individuals are supposed
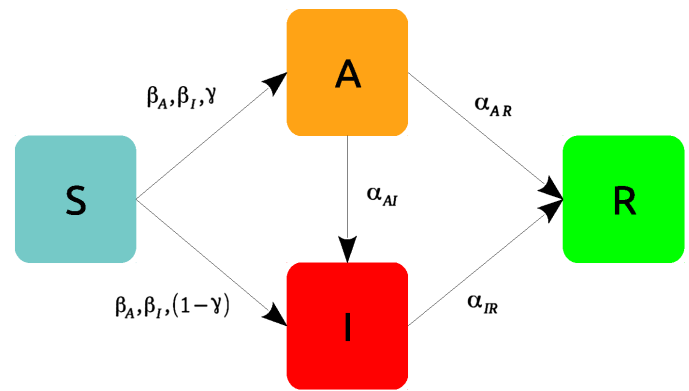

Fig. 1. Graphical representation of a SIAR model in which infectious individuals can either be symptomatic (I) or asymptomatic (A) (see also Eq. 21 in Materials and Methods).

to be measurable, we have assumed the initial conditions $\iota(0)$, $r(0)$, and the recovery rate of the symptomatic individuals, i.e., α_{IR}, to be known quantities. Second, we assume to be able to measure, for instance, through backward contact tracing, the rate at which asymptomatic individuals develop symptoms, i.e., α_{AI}. Hence, the parameters to be determined are the remaining ones, i.e., $\mathbf{p} = [a(0), \beta_I, \beta_A, \gamma, \alpha_{AR}]$.

Figure 2A shows a nontrivial nonmonotonic dependence of our sensitivity measures, σ and η, on γ. The value of σ has a peak at $\gamma = 0.51$, in correspondence of which η takes its minimum value. This represents an optimal condition for practical identifiability, as the sensitivity to parameters of the measured variables is high, while that of the unmeasured ones is low, and this implies that the unknown quantities of the system (both the model parameters and the hidden variables) can be estimated with small uncertainty. On the contrary, for $\gamma = 0.86$, we observe a relatively small value of σ and a large value of η, meaning that the measured variables are poorly identifiable, and the unmeasured variables are sensitive to a variation of parameters. This is the worst situation in which the estimated parameters may substantially differ from the real values, and the hidden variables may experience large variations even for small changes in the parameters. Furthermore, the quantity μ, which measures the sensitivity of the hidden variables to the measured ones, reported in Fig. 2B, exhibits a large peak at the value of γ for which σ is minimal. This is due to the fact that the vector that determines μ is almost aligned with $\mathbf{v}_1$. When this holds, we have that $\mu = \eta/\sigma$, which explains the presence of the spike in the μ curve. Similarly, the sensitivity μ takes its minimum almost in correspondence of the maximum of σ. The behavior of the model for $\gamma = 0.86$ is further illustrated in Fig. 2C, where the trajectories obtained in correspondence to the unperturbed values of the parameters, i.e., $\mathbf{m}(t, \mathbf{p})$ and $\mathbf{h}(t, \mathbf{p})$ (solid lines), are compared with the dynamics observed when $\mathbf{p}$ undergoes a perturbation with $\|\boldsymbol{\delta p}\| = 0.3\|\mathbf{p}\|$ along $\mathbf{v}_1$ (dashed lines). The small sensitivity σ of the measured variables $\iota(t, \mathbf{p})$ and $r(t, \mathbf{p})$ to parameters is reflected into perturbed trajectories that remain close to the unperturbed ones, whereas the large sensitivity η of the unmeasured variables $s(t, \mathbf{p})$ and $a(t, \mathbf{p})$ yields perturbed trajectories that significantly deviate from the unperturbed ones.

We now illustrate the different levels of identifiability that appear in the SIAR model for diverse settings of the parameters. Its analysis, in fact, fully depicts the more complete perspective on the
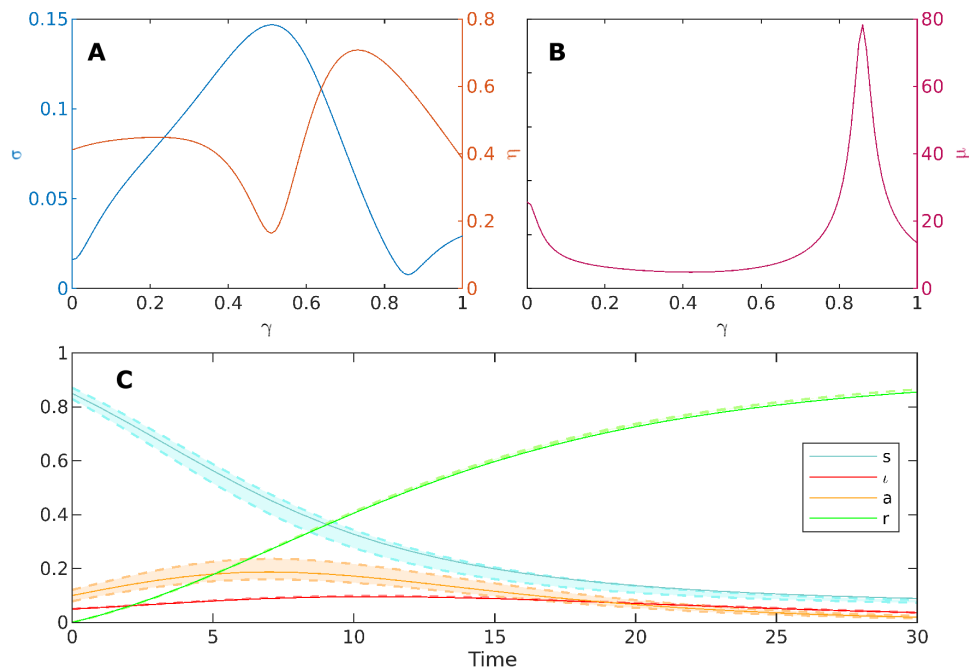
Fig. 2. Practical identifiability of the SIAR model in Fig. 1 as a function of the fraction γ of asymptomatic new infectious individuals. (**A**) Sensitivity σ and η of measured and hidden variables, respectively, to the parameters of the model. (**B**) Sensitivity μ of the hidden variables to the measured ones. (**C**) State variables for unperturbed values of parameters (with γ = 0.86, solid line) and for a perturbation with $\|\boldsymbol{\delta p}\| = 0.3\,\|\mathbf{p}\|$ along the first eigenvector of M (dashed lines).

problem of practical identifiability offered by simultaneously inspecting the sensitivity measures, σ and η. As the two sensitivity measures are not necessarily correlated, there can be cases for which a high identifiability of the measured variables to the parameters, i.e., large values of σ, corresponds to either a low or a high identifiability of the hidden variables to the parameters. Analogously, for other system configurations, in correspondence of small values of σ, namely, to nonidentifiable parameters, one may find large values of η, meaning that the hidden variables are nonidentifiable as well or, on the contrary, small values of η, indicating that the hidden variables are poorly sensitive to parameter perturbations. Together, four distinct scenarios of identifiability can occur, and all of them effectively appear in the SIAR model (Fig. 3): (A) low identifiability of the model parameters p and high identifiability of the hidden variables **h**, (B) high identifiability of both **p** and **h**, (C) low identifiability of both **p** and **h**, and (D) high identifiability of **p** and low identifiability of **h**. To illustrate them, we have considered four distinct configurations of the model (with parameters as given in Table 3 and illustrated in Materials and Methods) and, for each case study, compared the unperturbed trajectories to the perturbed ones, with the vector of parameters undergoing a variation $\|\boldsymbol{\delta p}\| = 0.3\,\|\mathbf{p}\|$ along $\mathbf{v}_1$. As regard cases (A) and (C), we have considered the vector of parameters to determine to be $\mathbf{p} = [\iota(t), a(0), r(t), \beta_I, \beta_A, \gamma, \alpha_{IR}, \alpha_{AR}, \alpha_{AI}]$, while for the cases (B) and (D), we have $\mathbf{p} = [a(0), \beta_I, \beta_A, \gamma, \alpha_{AR}]$, which is the same choice of **p** adopted in Fig. 2. Figure 3 shows the results obtained for each parameter configuration. In each panel, the solid lines represent the unperturbed trajectories, while the dashed lines correspond to the perturbed dynamics. In cases (A) and (B), we see that, under the variation $\boldsymbol{\delta p}$, the perturbed trajectories of
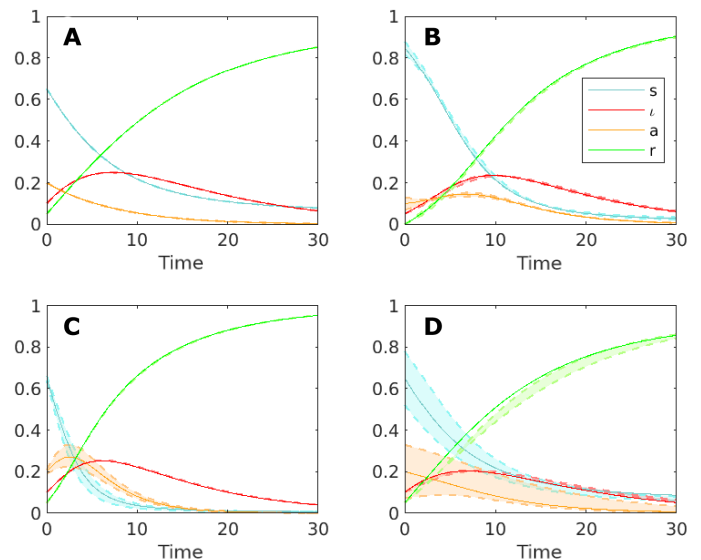
Fig. 3. Four scenarios of identifiability for the SIAR model of Fig. 1. All panels show the system dynamics (solid line) and the evolution of the system when the vector of parameters undergoes a variation δp such that $\|\boldsymbol{\delta p}\| = 0.3\,\|\mathbf{p}\|$ along the first eigenvector of M (dashed lines). (**A**) and (**C**) display configurations for which the observed variables (ι, r) are not sensitive to the variation, i.e., the model parameters are not identifiable, while (**B**) and (**D**) show the opposite case. Furthermore, (A) and (B) present scenarios for which the unobserved variables (s, a) are insensitive to the variation, meaning that they are predictable; vice-versa, (C) and (D) show the case in which the variables s and a are sensitive.

the hidden variables remain close to the unperturbed dynamics. Hence, the hidden variables are highly identifiable. Conversely, in cases (C) and (D), the perturbed trajectories substantially differ from the unperturbed dynamics, meaning that the hidden variables are poorly identifiable as they are sensitive to a variation of the model parameters. As concerns the measured variables, in cases (A) and (C), the perturbed trajectories slightly differ from the unperturbed dynamics. Therefore, as the measured variables are insensitive to the perturbation δ**p**, the model parameters have a low degree of identifiability. On the other hand, in cases (B) and (D), the perturbation of the parameters significantly affects the trajectories of the measured variables, meaning that the set of parameters reproducing the observed data is more identifiable.

Last, Table 1 illustrates the values of the sensitivity measures σ, η, and μ for each case. In particular, case (C) represents the worst scenario as the value of σ is relatively small, meaning that the model parameters **p** are poorly identifiable, and the value of η is large, indicating a high sensitivity of the hidden variables to the parameters. Conversely, the best scenario is represented by case (B), for which both the model parameters and the hidden variables are highly identifiable as the value of σ is large compared to the other cases, while the value of η remains relatively small.

Poorly identifiable models may provide unreliable predictions when the parameters are estimated from data

So far, we have illustrated how variations on the parameters affect the trajectories of measurable and hidden variables under different degrees of identifiability. However, a high sensitivity of the hidden variables to measured ones has relevant practical consequences, especially when the parameters are unknown and need to be fitted from data. In these conditions, a small uncertainty in the measurable variables due to the presence of noise in the data can propagate markedly and make the prediction of the hidden variables unreliable. Hence, in this section, we study how the lack of practical identifiability

Table 1. Values of σ, η, and μ for the four configurations of the SIAR model shown in Fig. 3.

	Case A	Case B	Case C	Case D
σ	0.0096	0.15	0.013	0.091
η	0.012	0.16	0.36	1.4
μ	34	5.2	29	15
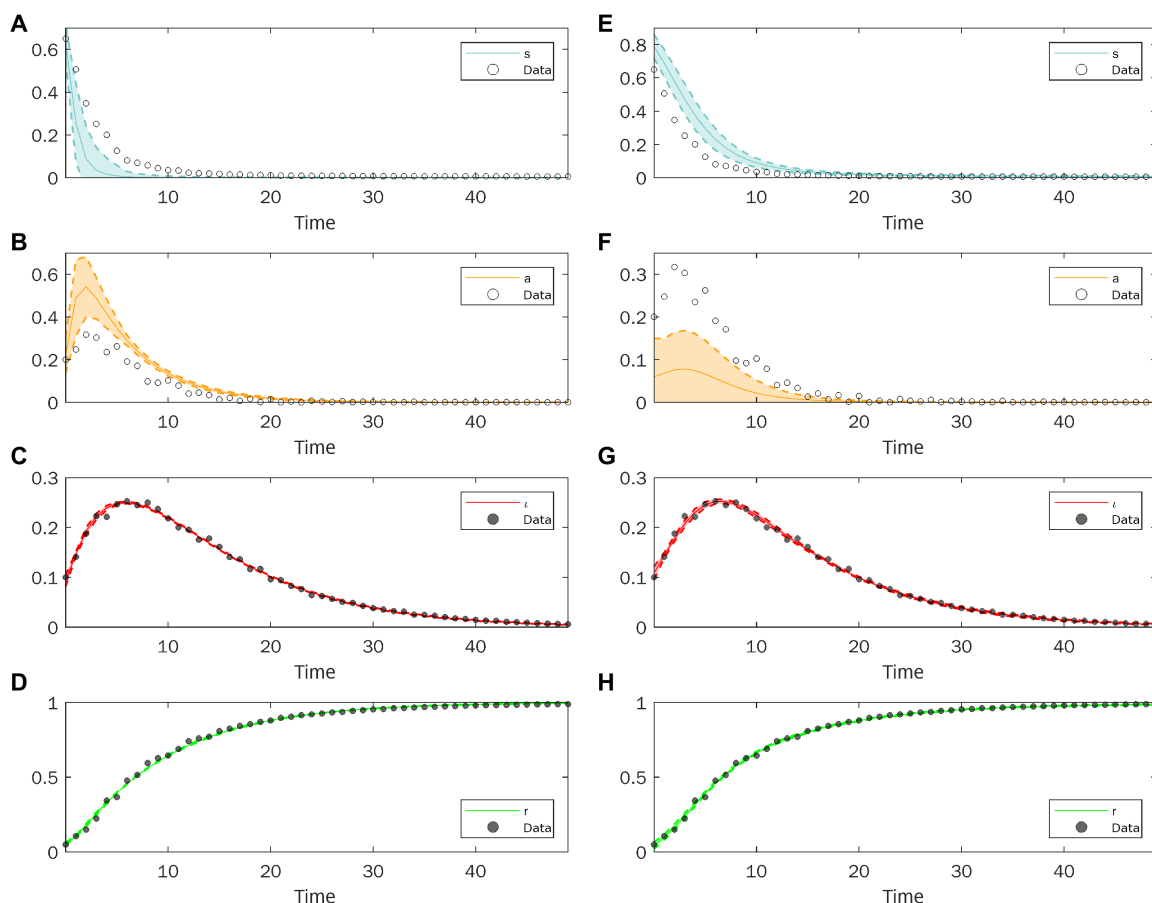
Fig. 4. Dynamics of the SIAR model when two different methods to estimate the model parameters from data are used. (A to D) display the results obtained by a least square error minimization procedure, while (**E** to **H**) show the outcome of Bayesian inference. The time evolution of both measured and hidden variables (solid lines) is reported together with the data. Data are shown with different markers if they pertain to measured variables (full circles, used for fitting) or unmeasured ones (empty circles, not used for fitting).
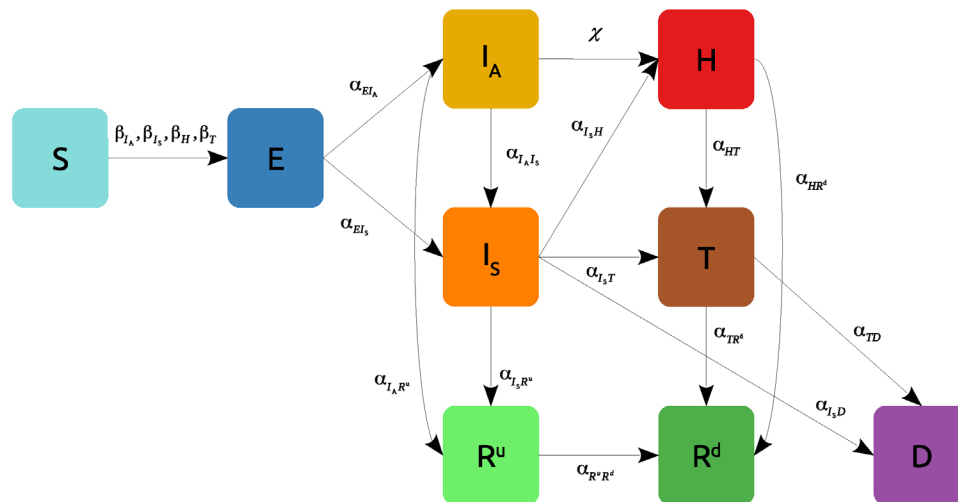
Fig. 5. Graphical representation of a nine-compartment model for the propagation of COVID-19 (see also Eq. 24 in Materials and Methods).

can affect the predictions of the SIAR model when this is fitted to empirical data. The reliability of the model predictions has been investigated by means of two different fitting techniques: a least square error minimization procedure (*42*), which provides a point estimate of the parameters, and a Bayesian inference approach (*43*), which, conversely, gives an estimate of the probability distribution in the parameter space. To carry out the numerical analysis, we consider the model under the same settings (i.e., fixing the values of the six parameters and the initial values of three variables) as those adopted in case (C) in the previous section, which correspond to the case of low identifiability of both the parameters and the hidden variables. We then generate from such a model a synthetic dataset of trajectories, which we fit using the two approaches mentioned above (see Materials and Methods for further details). All the model parameters are considered unknown and thus need to be determined through the fit.

First, we consider the least square error minimization approach. To show how, because of the lack of identifiability of the model, significant variations in the dynamics of the hidden variables can be obtained when fitting the measured variables, we have performed the following analysis. As the estimation procedure (based on a nonlinear optimization algorithm; see Materials and Methods) starts from an initial guess of the fitting parameters, indicated as $\mathbf{p_0}$, instead of fitting a single set of values, we have repeated the procedure under the very same conditions of the algorithm, for 500 runs, randomly selecting $\mathbf{p_0}$ from a Gaussian distribution centered on a fixed point of the parameter space and with variance equal to 0.25. We then discarded those runs yielding a fitting error $d > 0.015$, which corresponds to a relative error of 2.5%, thus keeping a total of 65 sets of parameters fitting the measured variables with a similar value of the error. The fact that different sets of parameters are obtained in this way may indicate that the error function has several local minima. Figure 4 (A to D) displays the average trajectories (over the 65 sets of parameters) of the four state variables (solid lines) and the respective regions where 95% of the trajectories lie (shadowed area). While the dynamics of the measured variables (Fig. 4, C and D) produced by the SIAR model are in very good agreement with the data, the same is not true for the temporal evolution of the hidden variables (Fig. 4, A and B). In particular, the

number of newly asymptomatic infected individuals is substantially overestimated.

We now consider the Bayesian inference approach. To implement it, we used the Delayed Rejection Adaptive Metropolis (DRAM) (*44*), a Markov Chain Monte Carlo (MCMC) algorithm, with settings as described in Materials and Methods. In particular, since we have here assumed to have no a priori information on the value of the model parameters, we have considered uniform prior probability distributions, representing the less informative conditions for the model (further details are given in Materials and Methods). Figure 4 (E to H) shows the temporal evolution of the SIAR variables. Solid lines represent the average trajectory obtained by sampling 500 sets of parameters from the posterior distributions, while the shadowed areas indicate the regions where 95% of the trajectories lie.

Similarly, to the case of the least square minimization, while the dynamics of the measured variables (Fig. 4, G and H) is in a good agreement with the synthetic data, the prediction of the hidden variables (Fig. 4, E and F) is not. At variance with the previous example, the number of newly asymptomatic infected individuals is here largely underestimated. Hence, these results indicate that the lack of identifiability can lead to unreliable results even when a Bayesian approach is adopted.

In the analysis presented in this section, we have assumed to have no a priori information on the values of the model parameters. Therefore, when performing the least square error minimization, we have extracted all the parameters from fitting, while, following the same reasoning, we have chosen a uniform prior probability distribution in the Bayesian approach. When instead we have strong a priori knowledge of a disease, this can be used to inform the models, by fixing the values of certain parameters while estimating the others, in the case of the least square error method, or by considering more informative prior distributions, in the case of Bayesian inference. When a priori information on the values of the parameters can be obtained, for instance, through medical and biological studies, the model predictions are expected to become less affected by uncertainty. The analysis of the sensitivity matrices (see also Materials and Methods) confirms this expectation as we have demonstrated that additional knowledge of the parameters, or
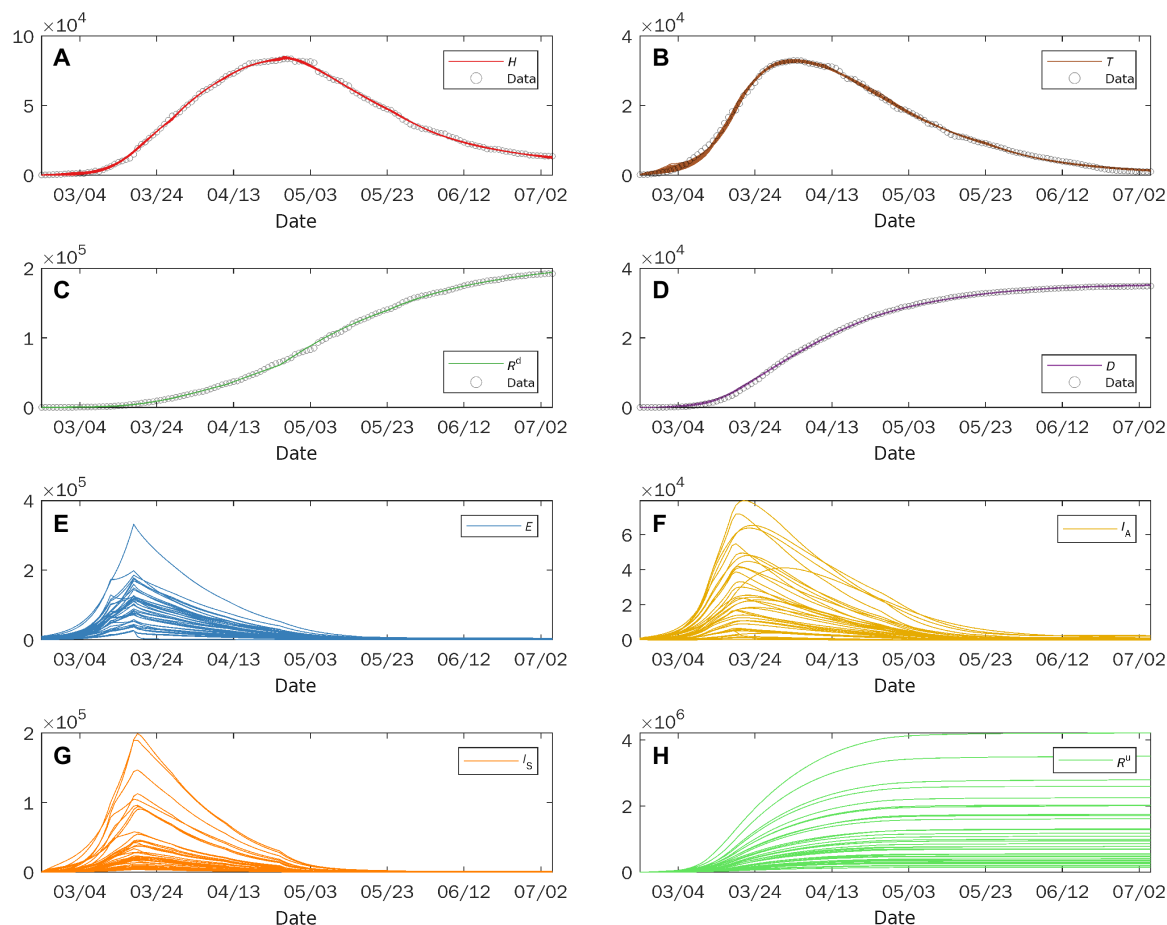
Fig. 6. Modeling the COVID-19 outbreak in Italy. The evolution of both measured (**A** to **D**) and hidden variables (**E** to **H**) of the model in Fig. 5 (solid lines) is reported together with the official data from the Civil Protection Department (circles).

the measurement of a hidden variable, can reduce the sensitivity to the measured variables, thus improving the reliability of the prediction. However, there are cases in which a priori knowledge is not available, and the analysis of identifiability becomes crucial. As an example, one could estimate the percentage of asymptomatic individuals according to serological surveys or to longitudinal studies. However, these data can be unavailable at an early stage of an epidemic outbreak (45, 46), preventing their use to inform the epidemiological models. These considerations hallmark once again the need for a synergistic approach to study newly discovered infectious diseases and stress the importance of assessing the reliability of mathematical modeling when the amount of available information is limited.

Lack of identifiability in COVID-19 modeling prevents reliable predictions

As a second application, we show the relevance of the problem of practical identifiability in the context of COVID-19 pandemic modeling. We consider a realistic model (Fig. 5) of the disease propagation, that is a variant of the SIDARTHE model (16) and is characterized by nine compartments accounting respectively for susceptible (S), exposed (E), undetected asymptomatic (I_A), undetected symptomatic (I_S), home isolated (H), treated in hospital (T), undetected recovered (R^u), detected recovered (R^d), and deceased

(D) individuals. Following the study of Giordano *et al.* (16), to account for the different nonpharmaceutical interventions and testing strategies issued during the COVID-19 outbreak in Italy (47, 48), the model parameters have been considered piece-wise constant and estimated using nonlinear optimization by fitting of the official data provided by the Civil Protection Department (49). As the dataset provides the evolution in time of the daily number of home isolated, hospitalized, detected recovered, and deceased individuals, we have considered four measured and five hidden variables in the model, namely, $\mathbf{m} \equiv [H, T, R^d, D]$ and $\mathbf{h} \equiv [S, E, I_A, I_S, R^u]$.

It is here worth discussing an important issue that concerns the model parameters. We have considered that different policy strategies affect, according to their nature, only specific parameters. In particular, we have assumed that a change in the containment strategy leads to a variation in the transmission rates β, while an adjustment in the testing strategy affects the values of the parameters $\alpha_{I_S H}$, α_{HT}, and α_{HR^d} (further details are reported in Materials and Methods). As both the containment and the testing strategies in Italy have frequently changed during the pandemic, most of the parameters to estimate consist of transmission and detection rates. While other parameters, such as the death or the recovery rates, can be derived from the current literature (50), very limited information is available on the transmission and detection rates that are difficult to measure directly and, therefore, need to be estimated by fitting available

data. Hence, as in (16), we have assumed all the model parameters to be unknown.

To show how significant variations in the evolution of the hidden variables can arise when fitting the measured variables, we have performed a numerical analysis similar to the one of the previous section. Again, rather than fitting a single set of values, we have repeated the minimization procedure under the same conditions of the algorithm, for 500 runs, randomly selecting the initial guess $\mathbf{p_0}$ from a Gaussian distribution centered on a fixed point of the parameter space, with variance equal to 0.25. We discarded the runs yielding a fitting error $e > 900$, corresponding to a relative error of 1.4%, thus keeping a total of 40 sets of parameters. Figure 6 shows the dynamical trajectories that we have obtained for each of the 40 sets of parameters (solid lines). Both measured (Fig. 6, A to D) and hidden (Fig. 6, E to H) variables are reported. While the time evolution of the measured variables produced by the model is in very good agreement with the empirical data, reported as circles in Fig. 6, significant differences in the trend of the hidden variables appear. A large variability is observed, confirming that the lack of identifiability yields a high sensitivity of the hidden variables to the measured one.

These findings have relevant implications. The large uncertainty on the size of the asymptomatic population makes questionable the use of the model as a tool to decide the policies to adopt.

DISCUSSION
The practical identifiability of a dynamical model is a critical, but often neglected, issue in determining the reliability of its predictions. In this paper, we have introduced a novel framework to quantify: (i) the sensitivity of the dynamical variables of a given model to its parameters, even in the presence of variables that are difficult to access empirically and (ii) how changes in the measured variables affect the evolution of the unmeasured ones.

The measures we have proposed are easy to compute and enable to assess, for instance, if and when the model predictions on the unmeasured variables are reliable or not, even in the cases in which the parameters of the model can be fitted with high accuracy from the available data.

As we have shown with a series of case studies, practical identifiability can critically affect the predictions of even very refined epidemic models introduced for the description of COVID-19, where dynamical variables, such as the population of asymptomatic individuals, are impossible or difficult to measure. This by no means should question the importance of these models—in that they enable a scenario analysis, otherwise impossible to carry out, and a deeper understanding of the spreading mechanisms of a novel disease—but should hallmark the relevance of a critical analysis of the results that takes into account sensitivity measures. It also highlights the importance of cross-disciplinary efforts that can provide a priori information on some of the parameters, ultimately improving the reliability of a model (8, 24).

A problem related to the one studied in our paper is that of observability, which investigates how to reconstruct the internal state of a system from measurements on the input and output, under the hypothesis that the model and its parameters are known (51, 52). Techniques based on the observability problem are clearly extremely important and may be applied, for instance, to derive the time evolution of asymptomatic individuals from measurements on

infected and recovered individuals, when it is possible to develop a fully observable model with known parameters.

MATERIALS AND METHODS
The sensitivity matrices and their properties
The sensitivity matrices considered in this paper are given by

$$\mathrm{M}_{ij} = \int_0^\infty \frac{\partial \mathbf{m}^T}{\partial p_i} \frac{\partial \mathbf{m}}{\partial p_j} dt; \ \mathrm{H}_{ij} = \int_0^\infty \frac{\partial \mathbf{h}^T}{\partial p_i} \frac{\partial \mathbf{h}}{\partial p_j} dt \tag{6}$$

where the vector functions $\mathbf{m} = \mathbf{m}(t, \mathbf{p})$ and $\mathbf{h} = \mathbf{h}(t, \mathbf{p})$ are obtained integrating system in eq. 1. The derivative of measurable and hidden variables with respect to the parameters $\mathbf{p}$

$$\mathbf{m}_i \equiv \partial \mathbf{m}/\partial p_i, \ \mathbf{h}_i \equiv \partial \mathbf{h}/\partial p_i$$

can be obtained by integrating the system

$$\begin{aligned} \frac{d\mathbf{m}_i}{dt} &= \frac{\partial \mathbf{f}}{\partial \mathbf{m}} \cdot \mathbf{m}_i + \frac{\partial \mathbf{f}}{\partial \mathbf{h}} \cdot \mathbf{h_i} + \frac{\partial \mathbf{f}}{\partial p_i} \\ \frac{d\mathbf{h}_i}{dt} &= \frac{\partial \mathbf{g}}{\partial \mathbf{m}} \cdot \mathbf{m}_i + \frac{\partial \mathbf{g}}{\partial \mathbf{h}} \cdot \mathbf{h_i} + \frac{\partial \mathbf{g}}{\partial p_i} \end{aligned} \tag{7}$$

where $i = 1, \dots n_p$.

The numerical evaluation of the sensitivity matrices is carried out first by integrating system in eq. 7 (for this step we use a fourth-order Runge-Kutta solver with adaptive step size control), resampling the trajectories with a sampling period of 1 day, and then performing a discrete summation over the sampled trajectories. Moreover, integration is carried out over a finite time interval $[0, \tau]$, with large enough τ. In the context of our work, as we have considered SIR (susceptible infected removed)-like epidemic models, we set the value of τ such that the system has reached a stationary state, i.e., the epidemic outbreak has ended, as every infected individual has eventually recovered (or dead, depending on the model).

We now present an important property of the sensitivity matrices. We will only take into account the set of measured variables $\mathbf{m}$, as similar considerations can be made for the hidden variables. Let us assume to be able to measure only a single variable, so that the vector $\mathbf{m}$ collapses into a scalar function, which we call $m_1(t)$. In this case, the element M_{ij} of the sensitivity matrix would be simply given by

$$(\mathrm{M})_{ij} = \int_0^\infty \frac{\partial m_1}{\partial p_i} \frac{\partial m_1}{\partial p_j} dt \tag{8}$$

Let us call this sensitivity matrix M_1.

Consider now a larger set of measured variables $\mathbf{m} = (m_1, m_2, \dots, m_{n_m})$. The quantity $\partial \mathbf{m}^T/\partial p_i \partial \mathbf{m}/\partial p_j$ in Eq. 6 is given by

$$\frac{\partial \mathbf{m}^T}{\partial p_i} \frac{\partial \mathbf{m}}{\partial p_j} = \frac{\partial m_1}{\partial p_i} \frac{\partial m_1}{\partial p_j} + \frac{\partial m_2}{\partial p_i} \frac{\partial m_2}{\partial p_j} + \dots + \frac{\partial m_{n_m}}{\partial p_i} \frac{\partial m_{n_m}}{\partial p_j} \tag{9}$$

Therefore, integrating over time in the interval $[0, \infty]$ and given the linearity property of the integrals, we find that the sensitivity matrix M of the set of the measured variables is given by the sum of the sensitivity matrices of the single measured variables. Formally, we have that

$$M = M_1 + M_2 + \ldots + M_{n_m} \tag{10}$$

This property of the sensitivity matrices is useful to demonstrating how measuring a further variable affects the sensitivity measures σ and μ, as discussed in the following subsection and in Results. Last, because matrices M and H are positive semidefinite, their eigenvalues are nonnegative. For any positive semidefinite matrix A of order m, we shall denote its eigenvalues as $0 \leq \lambda_1(A) \leq \lambda_2(A) \leq \ldots \leq \lambda_m(A)$.

Sensitivity measures and their properties

Here, we discuss in more detail the sensitivity measures introduced in Results. First, we want to propose a measure to quantify the practical identifiability of the model parameters given the measured variables. To do this, we need to evaluate the sensitivity of the trajectories of the measured variables to a variation of the model parameters. If this sensitivity is small, then different sets of parameters will produce very similar trajectories of the measured variables, meaning that the parameters themselves are poorly identifiable. In particular, as a measure of the parameters identifiability, we can consider the worst scenario, namely, the case in which the perturbation of the parameters minimizes the change in the measured variables. This happens when the variation of the model parameters δp is aligned along the eigenvector v_1 of M corresponding to the minimum eigenvalue $\lambda_1(M)$. Given the definition of M, we have that $\| \delta m \| = \sqrt{\lambda_1(M)} \| \delta p \|$; hence, we can consider the quantity

$$\sigma = \sqrt{\lambda_1(M)} \tag{11}$$

as an estimate of the sensitivity of the measured variables to the parameters. Note that, here and in the rest of the paper, $\|v\|$ denotes the Euclidean norm of a finite dimensional vector v, $\|v\|^2 = v \cdot v$, while for a function $u(t)$, $\|u\|$ denotes the L^2 norm of u in $[0, \infty]$, i.e., $\|u\|^2 = \int_0^\infty u \cdot u \, dt$.

Let us now focus on the hidden variables h. In general, as the hidden variables are not directly associated to empirical data, the largest uncertainty on the hidden variables is obtained in correspondence of a variation of the parameters along the eigenvector of H associated to the largest eigenvalue, namely, $\lambda_{n_p}(H)$.

Hence, to quantify the sensitivity of the hidden variables to the parameters, one may consider

$$\eta_{MAX} = \sqrt{\lambda_{n_p}(H)} \tag{12}$$

However, it is crucial to note that the hidden variables ultimately depend on the parameters of the model, which are estimated by fitting data that are available for the measured variables only. As a consequence, it is reasonable to consider a quantity that evaluates how the uncertainty on the model parameters (determined by the uncertainty of the measured variables and by their sensitivity to the parameters) affects the identifiability of the hidden variables. Therefore, as a measure of the sensitivity of the hidden variables to the parameters, we consider

$$\eta^2 = \frac{\delta p_1^T H \delta p_1}{\delta p_1^T \delta p_1} \tag{13}$$

where δp_1 is a perturbation of the parameters along the eigenvector v_1 of M corresponding to the minimum eigenvalue $\lambda_1(M)$. Note that, when v_1 and the eigenvector of H corresponding to the largest eigenvalue $\lambda_{n_p}(H)$ are aligned, by definition, we have $\eta = \eta_{MAX}$.

Last, we want to define a quantity to estimate how much the hidden variables are perturbed given a variation of the measured ones. In particular, as a measure of the sensitivity of the hidden variables to the measured variables, we consider the maximum perturbation of the hidden variables given the minimum variation of the measured ones, which is

$$\mu^2 = \max_{\|\delta p\|=1} \frac{\delta p^T H \delta p}{\delta p^T M \delta p} \tag{14}$$

Note that μ^2 can be computed considering the following generalized eigenvalue problem

$$H u_k = \lambda_k M u_k \tag{15}$$

where H and M are the sensitivity matrices for the hidden and the observed variables respectively, and $\lambda_k = \lambda_k(M, H)$ denotes the k-th generalized eigenvalue of matrices M and H. We will denote by λ_{n_p} the largest generalized eigenvalue and u the corresponding generalized eigenvector. Note that, since both matrices are symmetric, if u is a right eigenvector, then u^T is a left eigenvector. Multiplying each member of the equation by u^T and dividing by $u^T M u$, we obtain

$$\lambda_{n_p} = \frac{u^T H u}{u^T M u} = \max_{\|v\|=1} \frac{v^T H v}{v^T M v} \tag{16}$$

where one can recognize the definition of μ^2 provided in Eq. 14. In other words, μ^2 represents the largest eigenvalue of the matrix $M^{-1}H$.

It is worth noting two aspects about the sensitivity measure μ. First, given the definitions in Eqs. 11 and 12, for any δp with $\|\delta p\| = 1$, we have that $\delta p^T H \delta p \leq \eta_{MAX}^2$ and $\delta p^T M \delta p \geq \sigma^2$. As a consequence, we have that

$$\mu^2 \leq \frac{\eta_{MAX}^2}{\sigma^2} \tag{17}$$

Second, when the vector δp that determines μ is aligned with the eigenvector v_1 of M, it is possible to express μ in terms of the sensitivity measures σ and η. When $\delta p = \| \delta p \| v_1 = v_1$, recalling definitions in Eqs. 11 and 13, one obtains $v_1^T M v_1 = \sigma^2$, while $v_1^T H v_1 = \eta^2$, from which it follows

$$\mu = \frac{\eta}{\sigma} \tag{18}$$

In addition, we note that if v_1 and the eigenvector of H corresponding to its largest eigenvalue are aligned, one obtains that $\mu = \eta_{MAX}/\sigma$, which is the maximum value for the sensitivity measure μ.

We now demonstrate that the sensitivity of the hidden variables to the measured ones, μ^2, decreases as we measure one further variable. Let us assume now that we are able to measure one further variable, thus increasing the size of the set of measured variables to $n_m' = n_m + 1$ and, correspondingly, reducing that of the unmeasured variables to $n_h' = n_h - 1$. Given the property in Eq. 10, the new sensitivity matrices can be written as $M' = M + M_1$ and $H' = H - M_1$, where by M_1, we denote the sensitivity matrix for the newly measured variable. The new generalized eigenvalue problem is

$$H'\mathbf{u}' = \lambda'M'\mathbf{u}' \Leftrightarrow (H - M_1)\mathbf{u}' = \lambda'(M + M_1)\mathbf{u}' \qquad (19)$$

where, for simplicity, we have denoted by λ' the largest generalized eigenvalue of matrices M' and H'.

Left multiplying by $\mathbf{u}'^T$ and dividing by $\mathbf{u}'^T M\mathbf{u}'$, we obtain

$$\lambda_{n_p} = \frac{\mathbf{u}^T H\mathbf{u}}{\mathbf{u}^T M\mathbf{u}} \geq \frac{\mathbf{u}'^T H\mathbf{u}'}{\mathbf{u}'^T M\mathbf{u}'} = \frac{\mathbf{u}'^T H'\mathbf{u}' + \mathbf{u}'^T M_1\mathbf{u}'}{\mathbf{u}'^T M\mathbf{u}' - \mathbf{u}'^T M_1\mathbf{u}'} \geq$$
$$\frac{\mathbf{u}'^T H'\mathbf{u}'}{\mathbf{u}'^T M'\mathbf{u}'} = \lambda' \qquad (20)$$

where the first inequality comes from the definition of λ_{n_p}, while the second comes from the fact that H, M, H', M', and M_1 are positive semidefinite. In short, we find that $\lambda_{n_p} \geq \lambda'$, meaning that, by measuring one variable, the sensitivity of the hidden variables to the measured ones decreases.

SIAR model and setup for numerical analysis

The SIAR model of Fig. 1 is described by the following equations

$$\begin{cases} \dot{s} = -s(\beta_I \iota + \beta_A a) \\ \dot{\iota} = (1 - \gamma)s(\beta_I \iota + \beta_A a) + \alpha_{AI} a - \alpha_{IR}\iota \\ \dot{a} = \gamma s(\beta_I \iota + \beta_A a) - (\alpha_{AI} + \alpha_{AR})a \\ \dot{r} = \alpha_{IR}\iota + \alpha_{AR} a \end{cases} \qquad (21)$$

where $s(t)$, $\iota(t)$, $a(t)$, and $r(t)$ represent population densities, i.e., $s(t) = S(t)/N$, $\iota(t) = I(t)/N$, $a(t) = A(t)/N$, and $r(t) = R(t)/N$, where $S(t)$, $I(t)$, $A(t)$, and $R(t)$ represent the number of susceptible, infectious, asymptomatic, and recovered individuals, and N is the size of the population, so that $s(t) + \iota(t) + a(t) + r(t) = 1$. Here, β_I and β_A are the transmission rates for the symptomatic and the asymptomatic individuals, respectively, γ is the probability for newly infected individuals to show no symptoms, α_{AI} is the rate at which asymptomatic individuals become symptomatic, and α_{IR} and α_{AR} are the recovery rates for the two infectious populations. Note that all these parameters are positive quantities.

Asymptomatic individuals are difficult to trace as the individuals themselves could be unaware about their state. As a consequence, we assume that the density of asymptomatic individuals is not measurable, while the densities of symptomatic and recovered individuals are measured variables. According to the notation introduced in Eq. 1, we therefore have that $\mathbf{m} \equiv [\iota, r]$ and $\mathbf{h} \equiv [s, a]$. Note that, as a first approximation, here, we assume to be able to trace the asymptomatic individuals once they recover.

The results presented in Fig. 2 have been obtained considering the following setup. As the number of symptomatic infectious and recovered individuals are considered measurable, we have assumed that the initial conditions $\iota(0)$, $r(0)$, and the rate of recovery α_{IR} are known parameters. Second, we have supposed to be able to measure, for instance, through backward contact tracing the rate at which asymptomatic individuals develop symptoms, i.e., α_{AI}. Hence, the vector of parameters to determine by calibrating the model is given by $\mathbf{p} = [a(0), \beta_I, \beta_A, \gamma, \alpha_{AR}]$. Table 2 displays the value of the model parameters used to obtain the results shown in Fig. 2.

For the analysis of the four scenarios considered in Fig. 3, the values of the model parameters have been set as given in Table 3. Furthermore, to better contrast the results arising in the different case studies, in (A) and (C), we have considered $\mathbf{p} = [\iota(0), a(0),$

Table 2. Values of the model parameters used for the case study in Fig. 2.

ι_0	a_0	r_0	β_I	β_A	α_{IR}	α_{AR}	α_{AI}
0.05	0.1	0	0.6	0.3	0.1	0.2	0.03

Table 3. Values of the model parameters used for the case study in Fig. 3.

	ι_0	a_0	r_0	β_I	β_A	γ	α_{IR}	α_{AR}	α_{AI}
Case A	0.1	0.2	0.05	0.3	0.4	0.26	0.1	0.2	0.03
Case B	0.05	0.1	0	0.6	0.3	0.51	0.1	0.2	0.03
Case C	0.1	0.2	0.05	0.6	0.8	0.77	0.1	0.2	0.1
Case D	0.1	0.2	0.05	0.3	0.4	0.53	0.1	0.2	0.03

$r(0)$, β_I, β_A, γ, α_{IR}, α_{AR}, $\alpha_{AI}]$, while in (B) and (D), we have set $\mathbf{p} = [a(0), \beta_I, \beta_A, \gamma, \alpha_{AR}]$.

Estimating the parameters of the SIAR model

The SIAR model parameters have been estimated from data by adopting two different approaches, i.e., a nonlinear least square error minimization and a Bayesian inference. To generate a synthetic dataset, we have integrated the deterministic model in Eq. 21. To mimic measurement errors, we have adopted the following procedure. First, we compute $\bar{r}(t)$ by adding to $r(t)$ a uniform noise in the interval $(-\delta_t/2, \delta_t/2)$, where $\delta_t = |r(t + 1) - r(t)|$, checking that the synthetic time series remains monotonically nondecreasing. We have then generated the data $\bar{s}(t)$ in a similar fashion, this time controlling that the synthetic time series remains monotonically nonincreasing. To generate the data $\bar{\iota}(t)$, we have added a Gaussian noise with zero mean and standard deviation (SD) equals to 3% to the time series, making sure that $\bar{s}(t) + \bar{\iota}(t) + \bar{r}(t) \leq 1$. Last, the data $\bar{a}(t)$ have been evaluated using the fact that $\bar{s}(t) + \bar{\iota}(t) + \bar{a}(t) + \bar{r}(t) = 1$.

The integration of Eq. 21 has been carried out by using the lsoda ordinary differential equation (ODE) solver (53, 54) and then resampling the data with a sampling period of one time unit. We assumed that the density of asymptomatic individuals a is not measurable, while the densities of symptomatic and recovered individuals, i.e., ι and r, are measured variables.

As regard to the least square error minimization approach, the model parameters have been estimated using a nonlinear optimization procedure (implemented via the function fmincon in MATLAB) with the following objective function to minimize

$$d = \sqrt{\frac{1}{2\tau}\sum_{k=1}^{\tau}\left((\iota(k) - \bar{\iota}(k))^2 + (r(k) - \bar{r}(k))^2\right)} \qquad (22)$$

where $\bar{\iota}(k)$ and $\bar{r}(k)$ with $k = 1, \ldots, \tau$ (with $\tau = 50$) represent the noisy synthetic time series of the densities of infectious and recovered individuals, respectively, while $\iota(k)$ and $r(k)$ are the values of the corresponding variables obtained from the integration of Eq. 21.

The core idea of Bayesian inference is to provide an a posteriori probability distribution for the model parameter vector, $\mathbf{p}$, given an a priori probability distribution on the value of $\mathbf{p}$ and a likelihood function, which quantifies the goodness of a model in reproducing empirical data $\mathcal{D}$. The relationship between these is given by the Bayes' theorem, which reads

$$\pi(\mathbf{p}\,|\,\mathcal{D}) = \frac{\mathcal{L}(\mathcal{D}\,|\,\mathbf{p})\,\pi(\mathbf{p})}{\int_{\mathbf{p}}\mathcal{L}(\mathcal{D}\,|\,\mathbf{p})\,\pi(\mathbf{p})\,d\mathbf{p}} \qquad (23)$$

where $\pi(\mathbf{p})$ indicates the prior distribution, $\mathcal{L}(\mathcal{D}\,|\,\mathbf{p})$ the likelihood, and $\pi(\mathbf{p}\,|\,\mathcal{D})$ the posterior distribution. Usually, it is not possible to evaluate analytically the integral appearing in the denominator, especially when a large number of parameters are considered. Therefore, one relies on MCMC algorithms, which allow one to approximate of the posterior distribution. The MCMC algorithm we used to implement the Bayesian inference is the DRAM (44). As the likelihood function $\mathcal{L}(\mathcal{D}\,|\,\mathbf{p})$, we have considered the root mean square error, evaluated on the measurable variables only, i.e. $\iota(t)$ and $r(t)$, which corresponds to Eq. 22, namely, to the objective function of the nonlinear optimization procedure. As we have assumed to have no a priori knowledge of the values of the model parameters, for the Bayesian inference, we have considered uniform prior probability distributions, which are the simplest and least informative choice (55, 56). Flat priors do not require any additional information apart from setting the interval of possible parameter values. These intervals have been defined taking into account the following considerations. On the one hand, we have the initial conditions of the dynamical variables. As in the SIAR model, these represent population densities, we can assume the uniform prior distribution for their initial conditions to be defined in the interval [0,1]. Similarly, the parameter γ, which indicates the fraction of newly infected individuals not developing symptoms, can be assumed to be defined in the same interval. As regard the remaining parameters, since we have assumed to not have any other information except for the fact that they are positive quantities, we can consider the uniform distribution to be extended in the interval [0, ∞].

Nine-compartment model for COVID-19

The nine-compartment model of Fig. 5 can be considered as a variant of the SIDARTHE model (16). It is characterized by the presence of an incubation state, in which the individuals have been exposed to the virus (E) but are not yet infectious, and by infectious individuals, that, in addition to being symptomatic or asymptomatic, can be either detected or undetected. The model, therefore, includes four classes of infectious individuals: undetected asymptomatic (I_A), undetected symptomatic and pauci-symptomatic (I_S), home isolated (H, corresponding to detected asymptomatic and pauci-symptomatic), and treated in hospital (T, corresponding to detected symptomatic). Last, removed individuals can be undetected (R^u), detected (R^d), or deceased (D).

The model dynamics is described by the following equations

$$\begin{cases} \dot{S} = -S(\beta_{I_A}I_A + \beta_{I_S}I_S + \beta_H H + \beta_T T)/N \\ \dot{E} = S(\beta_{I_A}I_A + \beta_{I_S}I_S + \beta_H H + \beta_T T)/N \\ \quad -(\alpha_{EI_A} + \alpha_{EI_S})E \\ \dot{I}_A = \alpha_{EI_A}E - (\alpha_{I_A I_S} + \alpha_{I_A R^u})I_A - \chi I_A \\ \dot{I}_S = \alpha_{EI_S}E + \alpha_{I_A I_S}I_A \\ \quad -(\alpha_{I_S H} + \alpha_{I_S T} + \alpha_{I_S R^u} + \alpha_{I_S D})I_S \\ \dot{H} = \alpha_{I_S H}I_S + \chi I_A - (\alpha_{HT} + \alpha_{HR^d})H \\ \dot{T} = \alpha_{I_S T}I_S + \alpha_{HT}H - (\alpha_{TR^d} + \alpha_{TD})T \\ \dot{R}^u = \alpha_{I_A R^u}I_A + \alpha_{I_S R^u}I_S \\ \dot{R}^d = \alpha_{HR^d}H + \alpha_{TR^d}T \\ \dot{D} = \alpha_{I_S D}I_S + \alpha_{TD}T \end{cases} \qquad (24)$$

where the state variables represent the number of individuals in each compartment, $N = 60 \cdot 10^6$ and $S + E + I_A + I_S + H + T + R^u + R^d + D = N$. The official data on the spreading of COVID-19 in Italy made available by the Civil Protection Department [Dipartimento della Protezione Civile, (49)] provide information only on four of the nine compartments of the model, namely, the home isolated (H), hospitalized (T), detected recovered (R^d), and deceased individuals (D). These compartments constitute the set of the measured variables, while the other variables have to be considered as hidden, i.e., $\mathbf{m} \equiv [H, T, R^d, D]$ and $\mathbf{h} \equiv [S, E, I_A, I_S, R^u]$.

All the parameters appearing in (24) are considered unknown; thus, they need to be determined through fitting the model to the available data. It should also be noted that, as many nonpharmaceutical interventions have been issued/lifted, and the testing strategy has been changed several times over the course of the epidemics (47, 48), not all parameters can be considered constant in the whole period used for the fitting. Hence, similarly to (16), we have divided the whole period of investigation (which in our case ranges from 24 February to 06 July 2020) into different windows, within each of which the parameters are assumed to be constant. In each time window, one allows only some parameters to vary according to what is reasonable to assume will be influenced by the government intervention during that time window.

We distinguish two kinds of events that may require an adaptation of the model parameters. On the one hand, there are the nonpharmaceutical containment policies aimed at reducing the disease transmission. When these interventions are issued, the value of the parameters β may vary. On the other hand, the testing strategy, which affects the probability of detecting infected individuals, was also not uniform in the investigated period. When the testing policy changes, the value of the parameters $\alpha_{I_S H}$, α_{HT}, and α_{HR^d} may vary. Here, we notice two important points. First, the value of $\alpha_{I_S T}$ is assumed to be constant in the whole period, as we suppose that there are no changes in how the symptomatic individuals requiring hospitalization are detected. Second, as a change in the sole parameter $\alpha_{I_S H}$ would affect too much the average time an individual remains infected, then α_{HT} and α_{HR^d} also have to be included in the set of parameters that may change. On the basis of these considerations, the intervals in which each parameter remains constant or may change are identified. This defines the specific piece-wise waveform assumed for each of the parameters appearing in the model and, consequently, the effective number of values that need to be estimated for each parameter.

Hereafter, we summarize the events defining the different windows in which the whole period of investigation is partitioned:

1) On 02 March, a policy limiting screening only to symptomatic individuals is introduced.

2) On 12 March, a partial lockdown is issued.

3) On 18 March, a stricter lockdown, which further limits nonessential activities, is imposed.

4) On 29 March, a wider testing campaign is launched. Starting from this date, as the number of tests has constantly increased while the number of new infections was decreasing, the parameters are allowed to change every 14 or 28 days, namely, on 11 April, 25 April, and 23 May.

5) On 04 May, a partial lockdown lift is proclaimed.

6) On 18 May, further restrictions are relaxed.

7) On 03 June, interregional mobility is allowed. This is the last time the model parameters are changed.

Note that, for the time period until 5 April, we have followed the same time partition used in (*16*).

The model parameters have been estimated using a nonlinear optimization procedure (implemented via the function fmincon in MATLAB) with the following objective function to minimize

$$e = \sqrt{\frac{1}{4\tau}\sum_{k=1}^{\tau}\left((H(k)-\bar{H}(k))^2 + (T(k)-\bar{T}(k))^2 + \left(R^d(k)-\bar{R}^d(k)\right)^2 + (D(k)-\bar{D}(k))^2\right)} \tag{25}$$

where $\bar{H}(k)$, $\bar{T}(k)$, $\bar{R}^d(k)$, and $\bar{D}(k)$ with $k = 1, \ldots, \tau$ ($\tau = 134$ days) represent the time series of daily data for isolated, hospitalized, detected recovered, and deceased individuals provided by the Civil Protection Department (*49*), and $H(k)$, $T(k)$, $R^d(k)$, and $D(k)$ are the values of the corresponding variables obtained from the integration of Eq. 24. The integration of Eq. 24 has been carried out by using a suitable ODE solver with maximum integration step size equal to 10^{-2} days and then resampling the data with a sampling period of 1 day.

REFERENCES AND NOTES

1. R. M. Anderson, H. Heesterbeek, D. Klinkenberg, T. D. Hollingsworth, How will country-based mitigation measures influence the course of the covid-19 epidemic? *Lancet* **395**, 931–934 (2020).
2. World Health Organization (WHO), Coronavirus disease (covid-19): Weekly epidemiological update (2020); https://who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports[accessed 15 October 2020].
3. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track covid-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
4. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
5. N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet* **395**, 507–513 (2020).
6. W. J. Wiersinga, A. Rhodes, A. C. Cheng, S. J. Peacock, H. C. Prescott, Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): A review. *JAMA* **324**, 782–793 (2020).
7. L. Wang, Y. Wang, D. Ye, Q. Liu, Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence. *Int. J. Antimicrob. Agents* **55**, 105948 (2020).
8. E. Estrada, COVID-19 and SARS-CoV-2. Modeling the present, looking at the future. *Phys. Rep.* **869**, 1–51 (2020).
9. M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini Jr., A. Vespignani, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
10. K. Leung, J. T. Wu, D. Liu, G. M. Leung, First-wave covid-19 transmissibility and severity in China outside hubei after control measures, and second-wave scenario planning: A modelling impact assessment. *Lancet* **395**, 1382–1393 (2020).
11. P. Castorina, A. Iorio, D. Lanteri, Data analysis on coronavirus spreading by macroscopic growth laws. *Int. J. Mod. Phys. C* **31**, 2050103 (2020).
12. D. Lanteri, D. Carcò, P. Castorina, M. Ceccarelli, B. Cacopardo, Containment effort reduction and regrowth patterns of the COVID-19 spreading. arXiv:2004.14701 (2020).
13. D. Fanelli, F. Piazza, Analysis and forecast of covid-19 spreading in china, italy and france. *Chaos Solitons Fractals* **134**, 109761 (2020).
14. A. Arenas, W. Cota, J. Gomez-Gardeñes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano, B. Steinegger, A mathematical model for the spatiotemporal epidemic spreading of COVID-19. *medRxiv* (2020).
15. A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, N. Davies, A. Gimma, K. van Zandvoort, H. Gibbs, J. Hellewell, C. I. Jarvis, S. Clifford, B. J. Quilty, N. I. Bosse, S. Abbott, P. Klepac, S. Flasche, Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *Lancet Infect. Dis.* **20**, 553–558 (2020).
16. G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* **26**, 855–860 (2020).
17. A. Aleta, D. Martín-Corral, A. Pastore y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini Jr., S. Merler, A. Pentland, A. Vespignani, E. Moro,
Y. Moreno, Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **4**, 964–971 (2020).
18. F. Della Rossa, D. Salzano, A. Di Meglio, F. De Lellis, M. Coraggio, C. Calabrese, A. Guarino, R. Cardona-Rivera, P. De Lellis, D. Liuzza, F. L. Iudice, G. Russo, M. di Bernardo, A network model of italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic. *Nat. Commun.* **11**, 5106 (2020).
19. D. Proverbio, F. Kemp, S. Magni, A. Husch, A. Aalto, L. Mombaerts, A. Skupin, J. Gonçalves, J. Ameijeiras-Alonso, C. Ley, Dynamical SPQEIR model assesses the effectiveness of non-pharmaceutical interventions against COVID-19 epidemic outbreaks. *PLOS ONE* **16**, e0252019 (2021).
20. T. Heinemann, A. Raue, Model calibration and uncertainty analysis in signaling networks. *Curr. Opin. Biotechnol.* **39**, 143–149 (2016).
21. A. F. Villaverde, A. Barreiro, A. Papachristodoulou, Structural identifiability of dynamic systems biology models. *PLOS Comput. Biol.* **12**, e1005153 (2016).
22. G. Chowell, Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infect. Dis. Model.* **2**, 379–398 (2017).
23. T. Quaiser, M. Mönnigmann, Systematic identifiability testing for unambiguous mechanistic modeling–application to JAK-STAT, MAP kinase, and NF-κ B signaling pathway models. *BMC Syst. Biol.* **3**, 50 (2009).
24. W. C. Roda, M. B. Varughese, D. Han, M. Y. Li, Why is it difficult to accurately predict the COVID-19 epidemic? *Infect. Dis. Model.* **5**, 271–281 (2020).
25. M. C. Eisenberg, S. L. Robertson, J. H. Tien, Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease. *J. Theor. Biol.* **324**, 84–102 (2013).
26. N. Tuncer, T. T. Le, Structural and practical identifiability analysis of outbreak models. *Math. Biosci.* **299**, 1–18 (2018).
27. A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, J. Timmer, Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923–1929 (2009).
28. N. Tuncer, M. Marcheva, S. LaBarre, S. Payoute, Structural and practical identifiability analysis of zika epidemiological models. *Bull. Math. Biol.* **80**, 2209–2241 (2018).
29. K. Roosa, G. Chowell, Assessing parameter identifiability in compartmental dynamic models using a computational approach: Application to infectious disease transmission models. *Theor. Biol. Med. Model.* **16**, 1–15 (2019).
30. R. A. Horn, C. R. Johnson, *Matrix Analysis* (Cambridge Univ. press, 2012).
31. K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the diamond princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* **25**, 2000180 (2020).
32. Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, T. Feng, Epidemiology and transmission of covid-19 in shenzhen china: Analysis of 391 cases and 1,286 of their close contacts. *medRxiv* (2020).
33. E. Lavezzo, E. Franchin, C. Ciavarella, G. Cuomo-Dannenburg, L. Barzon, C. Del Vecchio, L. Rossi, R. Manganelli, A. Loregian, N. Navarin, D. Abate, M. Sciro, S. Merigliano, E. Decanale, M. C. Vanuzzo, F. Saluzzo, F. Onelia, M. Pacenti, S. Parisi, G. Carretta, D. Donato, L. Flor, S. Cocchio, G. Masi, A. Sperduti, L. Cattarino, R. Salvador, K. A. M. Gaythorpe; Imperial College London COVID-Response Team, A. R. Brazzale, S. Toppo, M. Trevisan, V. Baldo, C. A. Donnelly, N. M. Ferguson, I. Dorigatti, A. Crisanti, Suppression of COVID-19 outbreak in the municipality of Vo, Italy. *medRxiv* (2020).
34. C. Liu, X. Wu, R. Niu, X. Wu, R. Fan, A new SAIR model on complex networks for analysing the 2019 novel coronavirus (COVID-19). *Nonlinear Dyn.*, 1–11 (2020).
35. R. H. Chisholm, P. T. Campbell, Y. Wu, S. Y. Tong, J. McVernon, N. Geard, Implications of asymptomatic carriers for infectious disease transmission and control. *R. Soc. Open Sci.* **5**, 172341 (2018).
36. L. Pribylova, V. Hajnova, SEIAR model with asymptomatic cohort and consequences to efficiency of quarantine government measures in COVID-19 epidemic. arXiv:2004.02601 (2020).
37. J. B. Aguilar, J. S. Faust, L. M. Westafer, J. B. Gutierrez, Investigating the impact of asymptomatic carriers on COVID-19 transmission. *medRxiv* (2020).
38. M. Robinson, N. I. Stilianakis, A model for the emergence of drug resistance in the presence of asymptomatic infections. *Math. Biosci.* **243**, 163–177 (2013).
39. D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, A. Vespignani, Seasonal transmission potential and activity peaks of the new influenza A(H1N1): A Monte Carlo likelihood analysis based on human mobility. *BMC Med.* **7**, 45 (2009).
40. D. Balcan, B. Gonçalves, H. Hu, J. J. Ramasco, V. Colizza, A. Vespignani, Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *J. Comput. Sci.* **1**, 132–145 (2010).
41. X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang,

B. J. Cowling, F. Li, G. M. Leung, Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).

42. H. T. Banks, M. Davidian, J. R. Samuels, K. L. Sutton, G. Chowell, M. Hyman, L. M. A. Bettencourt, C. Castillo-Chavez, Chapter 11: Mathematical and Statistical Estimation Approaches in Epidemiology, in *Mathematical and Statistical Estimation Approaches in Epidemiology*, G. Chowell, J. M. Hyman, L. M. A. Bettencourt, and C. Castillo-Chavez, Eds. (Springer Netherlands, 2009) pp. 249–302.

43. T. Toni, D. Welch, N. Strelkowa, A. Ipsen, M. P. H. Stumpf, Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187–202 (2009).

44. H. Haario, M. Laine, A. Mira, E. Saksman, DRAM: Efficient adaptive MCMC. *Stat. Comput.* **16**, 339–354 (2006).

45. A. Vespignani, H. Tian, C. Dye, J. O. Lloyd-Smith, R. M. Eggo, M. Shrestha, S. V. Scarpino, B. Gutierrez, M. U. G. Kraemer, J. Wu, K. Leung, G. M. Leung, Modelling COVID-19. *Nat. Rev. Phys.* **2**, 279–281 (2020).

46. J. Lourenço, R. Paton, M. Ghafari, M. Kraemer, C. Thompson, P. Simmonds, P. Klenerman, S. Gupta, Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. *MedRxiv* 2020.03.24.20042291 (2020).

47. Presidenza del Consiglio dei Ministri (Presidency of the Council of Ministers), Governmental containment policies; http://governo.it/it/coronavirus-misure-del-governo[accessed 6 July 2020].

48. Presidenza del Consiglio dei Ministri (Presidency of the Council of Ministers), Legislation issued in response to covid-19 epidemic; http://governo.it/it/coronavirus-normativa[accessed 6 July 2020].

49. Dipartimento della Protezione Civile (Civil protection department), Data on the national trend; https://github.com/pcm-dpc/COVID-19/tree/master/dati-andamento-nazionale[accessed 06 July 2020].

50. R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. F. John, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, N. M. Ferguson, Estimates of the severity of coronavirus disease 2019: A model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).

51. E. W. Griffith, K. S. P. Kumar, On the observability of nonlinear systems: I. *J. Math. Anal. Appl.* **35**, 135–147 (1971).

52. W. S. Gray, J. P. Mesko, Observability functions for linear and nonlinear systems. *Syst. Control Lett.* **38**, 99–113 (1999).

53. A. C. Hindmarsh, Odepack, a systematized collection of ODE solvers. *Sci. Comput.* **1**, 55–64 (1983).

54. L. Petzold, Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM J. Sci. Stat. Comput.* **4**, 136–148 (1983).

55. F. Kemp, D. Proverbio, A. Aalto, L. Mombaerts, A. Fouquier d'Hérouël, A. Husch, C. Ley, J. Gonçalves, A. Skupin, S. Magni, Modelling COVID-19 dynamics and potential for herd immunity by vaccination in Austria, Luxembourg and Sweden. *J. Theor. Biol.* **530**, 110874 (2021).

56. G. E. P. Box, G. C. Tiao, *Bayesian Inference in Statistical Analysis*, vol. 40 (John Wiley & Sons, 2011).

Science Advances

Lack of practical identifiability may hamper reliable predictions in COVID-19 epidemic models

Luca Gallo, Mattia Frasca, Vito Latora, and Giovanni Russo

View the article online
https://www.science.org/doi/10.1126/sciadv.abg5234
Permissions
https://www.science.org/help/reprints-and-permissions

Use of this article is subject to the Terms of service