Identifiability and predictability of integer- and fractional-order epidemiological models using physics-informed neural networks

Ehsan Kharazmi^{1,6}, Min Cai^{1,2,6}, Xiaoning Zheng^{1,3}, Zhen Zhang¹, Guang Lin⁴ and George Em Karniadakis⁵

We analyze a plurality of epidemiological models through the lens of physics-informed neural networks (PINNs) that enable us to identify time-dependent parameters and data-driven fractional differential operators. In particular, we consider several variations of the classical susceptible-infectious-removed (SIR) model by introducing more compartments and fractional-order and time-delay models. We report the results for the spread of COVID-19 in New York City, Rhode Island and Michigan states and Italy, by simultaneously inferring the unknown parameters and the unobserved dynamics. For integer-order and time-delay models, we fit the available data by identifying time-dependent parameters, which are represented by neural networks. In contrast, for fractional differential models, we fit the data by determining different time-dependent derivative orders for each compartment, which we represent by neural networks. We investigate the structural and practical identifiability of these unknown functions for different datasets, and quantify the uncertainty associated with neural networks and with control measures in forecasting the pandemic.

Since the outbreak of the novel coronavirus at the beginning of 2020, over 100,000 papers have been published related to COVID-19, with a substantial portion of them focusing on epidemiological models that describe the observed and unobserved dynamics, primarily in postdictive mode, although some models are also used for short-term forecasting. These models represent the lumped dynamics of a big city, a state or a country, but suffer from large uncertainties¹, resulting primarily from the lack of identifiability as well as the noise in the available sparse data. This lack of identifiability is related to modeling assumptions (structural identifiability), data availability and the complex biological characteristics of virus transmission, which are largely unknown and hard to measure² and include various emerging mutations of the virus³. The true scientific challenge is to recognize the large limitations of these (potentially useful) models, identify the multiple sources of uncertainty and suggest flexible models that can deal with seasonal variation in susceptibility, time delays, noisy data, under-determined systems, non-Markovian behavior and inherent stochasticity⁴. In addition to seasonal variation in transmission, for example, due to weather or mobility, for a given model the data uncertainty is usually propagated into the model parameters, rendering them as random variables/processes with an underlying probability distribution. The uncertainties in the input parameters affect the model predictability adversely, leaving many of these models inadequate for any decision-making, as they lack robustness, which is a measure of the extent to which the forward solvers amplify uncertainties from the input to the output⁵. In general, quantification of parametric input uncertainty is only based on a single given model, hence ignoring the bigger source of uncertainty associated with the model structure. A clear example of uncertainty associated with several different models in analyzing and predicting the dynamics of this com-

plex system can be found in the COVID-19 Forecast Hub⁶, which is a public online server that serves as a central repository of forecasts and predictions from over 50 international research groups. Even with a 95% confidence interval, one can see that the predictions of these different models can vary drastically (a snapshot of vastly different predictions is shown in Supplementary Section 2).

A parameterized differential equation model is structurally identifiable if the undetermined parameters can be uniquely identified when the exact forms of certain state variables are available. If a model is not structurally identifiable, there will be large uncertainty in the estimated parameter values and the inferred state variable values, even when the model is fitted to noiseless data. Practical identifiability refers to the non-uniqueness issue that occurs when the model is fitted to noisy data on discrete time points. Practical identifiability is a sufficient but not necessary condition of structural identifiability. Parameter identifiability relates to model uncertainty for a specific model, but here we want to pose questions beyond this uncertainty notion, and aim to investigate two fundamentally different approaches to epidemiological modeling. In the classical approach, the models are obtained by reducing the complex system into several compartments governed by a (nonlinear coupled) system of ordinary differential equations (ODEs), where we adjust the (possibly time-dependent) parameters of the model to fit the available data. This is of course all we can do because of the constraint of using pre-fabricated integer-order differential operators that represent fixed rate dynamics at all times. A variation of this is to introduce time delays into the ODEs to account for some memory effects. An opposite way of constructing a model is to tweak the differential operators instead of tweaking the parameters, using, for example, fractional derivatives with their orders (and hence the rate of dynamics) determined directly by the data, while we use simple

¹Division of Applied Mathematics, Brown University, Providence, RI, USA. ²Department of Mathematics, Shanghai University, Shanghai, China.

³Department of Mathematics, College of Information Science and Technology, Jinan University, Guangzhou, China. ⁴Department of Mathematics and School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA. ⁵Division of Applied Mathematics and School of Engineering, Brown University, Providence, RI, USA. ⁶These authors contributed equally: Ehsan Kharazmi, Min Cai. ✉e-mail: george_karniadakis@brown.edu

parameters with some nominal values in front of the operators. By considering these two classes, we expand both the structural and practical identifiability as well as predictability analyses in a broader perspective that allows us to investigate a plurality of epidemiological models and test their performance with diverse COVID-19 datasets for cities, states and countries. In particular, we consider several different integer-order, fractional-order and time-delay models, each expressed as a system of ODEs, where the parameters are either biologically determined and fixed or time-dependent. The integer-order and time-delay models fall into the first class of modeling with time-dependent parameters and unknown time-delay duration to fit the data. The fractional-order models are different paradigms of modeling, where we tune the time-dependent operator to fit the available data. Our goal is not to pick a winner out of the multiple available models, but rather to provide a framework to screen which of these models are possible good candidates to fit the data and make reasonable predictions with bounded levels of uncertainty. Specifically, we investigate why we need time-dependent parameters in the lumped epidemiological models, how to extrapolate and forecast beyond the training period, and which is the most robust model for short- and long-term prediction.

In this broader framework, where the unknowns to be determined in the integer-order and time-delay models or the derivative orders in the fractional-order models are functions of time, the inverse problem to solve is particularly challenging. Several numerical methods have been developed to solve the inverse problem of inferring model (mostly constant) parameters from available data. They typically convert the problem into an optimization problem and then formulate a suitable estimator by minimizing an objective function^{7–15}. Here we overcome this difficulty by developing physics-informed neural networks (PINNs) both for integer-order and fractional-order models¹⁶. PINNs provide a flexible computational tool that encodes the ODEs into the neural networks to satisfy the equations while accurately fitting the data. In other words, parameter inference and simulation of the observed and unobserved dynamics take place simultaneously. The PINN formulation allows the parameters to be time-dependent functions, represented by separate neural networks, and this makes the inverse problem of inferring the parameters and unobserved dynamics from available data readily applicable to several different models. The structural and practical identifiability of integer-order models have been studied recently by Zhang and colleagues¹⁷. However, this analysis cannot be applied to the fractional models or even to integer-order models with parameters, which are continuous functions of time. The unique determination of the derivatives' variable order in the case of a time-fractional linear diffusion equation on a general multidimensional domain was proved by Zheng and colleagues¹⁸. To the best of our knowledge, the existence and uniqueness of the solution have not been proven for the inverse problem of identifying variable order for the nonlinear system of equations. Through systematic numerical experiments, in this Article we study and discuss the structural and practical identifiability of different time-dependent fractional orders of fractional models by using different amounts of data. If limited to only the reported available data, the inference is not very robust and the inferred fractional orders may exhibit relatively large uncertainty. However, if sufficient data are available for training, the PINN formulation can accurately infer different time-dependent fractional orders for each state of the modified susceptible-infectious-removed (SIR) models. The following steps show the implementation of our formulation:

1. Pre-processing of the available data.
2. For each epidemiological model, studying the identifiability.
3. Using PINNs to infer the unknown parameters of the model and discover the unobserved dynamics.
4. Using the inferred model to forecast the pandemic.

5. Propagating the parameter uncertainty to the predicted trajectories of the dynamics.

Results

Model plurality. The pluralization of epidemiological models strives for either developing a variety of compartmental models with time-dependent parameters or employing new fractional operators with different time-dependent fractional orders of the derivatives. This systematic analysis of different models provides a measure of their identifiability, predictability and uncertainty.

A general framework used in modeling the spread of infectious diseases among a fixed population is compartmental modeling, in which individuals are categorized into different epidemiological classes (compartments) according to their disease-related status. The dynamics of the virus spread through the population is governed by a nonlinear system of ODEs. The classical prototypical compartmental model is the SIR model. This simple model can be thought of as a lumped-compartment model, where the main compartments can be decomposed into various other compartments to form a variant of the SIR model to study specific state dynamics in more detail. These variations of the SIR model are usually developed by appending additional compartments to the basic SIR model. The choice of the model is subjective to the purpose of the study and the available data to calibrate the model parameters, and at the present time it is done in an ad hoc manner.

We consider different integer-order, fractional-order and time-delay models for different variations of SIR model. The integer-order models use simple derivative operators that do not account for the effect of memory in the evolution of dynamics, but instead the model parameters are treated as time-varying variables. However, many works have confirmed the power-law scaling feature of the dynamics of COVID-19 transmission^{19,20}, which indicates the notion of memory effects in the spread dynamics. In the classical SIR model, the current number of infectious individuals has an exponential distribution in the infectious period, that is, $I(t) = I_0 e^{-\gamma t} + \int_0^t \beta S(u) I(u) e^{-\gamma(t-u)} du$, where s and β are number of susceptible individuals and transmission rate, respectively. This can be interpreted as the probability for an infected subject to remain infected at time t being $e^{-\gamma t}$, where $\frac{1}{\gamma}$ refers to the mean recovery period^{21,22}. Thus, the number of patients has an exponential distribution in the mean infection period. In a general setting, we can extend the SIR model into $I(t) = I_0 e^{-\gamma t} + \int_0^t \beta S(u) I(u) \phi(t-u) du$, where ϕ is a probability function (non-increasing in t , $\phi(0) = 1$) and the mean infection period $\int_0^{+\infty} \phi(t) dt$ (refs. ^{23,24}). If we choose ϕ to be a step function then we arrive at a constant time-delay model. If we choose ϕ to be an exponential function then we recover a classical SIR model²². If we choose ϕ to be a power-law waiting time distribution then we obtain a fractional SIR model²⁵. Other epidemiological models characterized by partial differential equations (PDEs) can be viewed as variations of the standard ODE models that include more variables such as spatial effects, health status and age^{26–28}. Although they may provide a more realistic picture of disease via spatial diffusion and higher level parameterization, the PDE models lack structural identifiability. More importantly, the spatial data incompleteness and heterogeneity become a major issue for such models, making them impractical at this juncture. In this Article, we consider a total of nine different ODE models. In a general setting, if we let $\mathbf{U}(t)$ be the vector of all epidemiological classes considered in a model, then the coupled system of differential equation governing the dynamics of that model can be written as

$$\mathcal{L}\mathbf{U}(t) = \mathcal{F}(\mathbf{U}, t; \lambda), \quad (1)$$

Fig. 1 | A plurality of epidemiological models. Shown are sketches of the compartments as well as their mathematical definitions. Each gray box represents an epidemiological compartment and the arrows show the direction of population flow. Panels 1–3 are integer-order models, panels 4–6 are fractional-order models, and panel 7 is a time-delay model. Model $\mathbb{I}_1$, integer-order SEIJDHR; model $\mathbb{I}_2$, integer-order SEPIJDHR; model $\mathbb{I}_3$, integer-order SEPIJDHQR; model $\mathbb{F}_1$, fractional-order SIR; model $\mathbb{F}_2$, fractional-order SIDR; model $\mathbb{F}_3$, fractional-order SIHDR; model $\mathbb{T}_1$, time-delay SIJHDR. Details of the mathematical definitions are given in Supplementary Section 3.

where $\mathcal{L}$ is either an integer-order or a fractional-order temporal differential operator, $\mathcal{F}$ is a nonlinear operator and λ is the set of known/unknown model parameters. The epidemiological classes in the models include the number of individuals in susceptible (S), exposed (E), pre-symptomatic (P), quarantined (Q), infectious (I), asymptomatic (J), hospitalized (H), death (D) and recovered/removed (R) compartments. We consider the auxiliary cumulative compartments (I^c and H^c) to help fitting the available data. The three crucial time-dependent parameters¹⁷ are the community transmission rate ($\beta_c(t)$), the proportion of disease-related death from the H class ($q(t)$) and the proportion of hospitalized individuals ($p(t)$). The transmission rate of a disease, $\beta_c(t)$, is the per capita rate of infection when a contact occurs. Control measures implemented in different regions and the subsequent relaxation of restrictions can directly impact the incidence curve in different ways. Thus, estimating this value accurately is critical because it represents the effects of public health policies. The percentage of disease-related deaths, $q(t)$, changes over the course of the outbreak mainly due to various public health policies and the discovery of better therapies and treatments, and other parameters take different values²⁹. Similarly, the hospitalization ratio, $p(t)$, also varies due to increased resources being channeled to the healthcare system in the city^{30,31}. In Supplementary Fig. 4, we carry out a systematic study to demonstrate the need for considering the aforementioned three parameters as time-dependent instead of fixed parameters.

Seven of the nine models are shown in Fig. 1 (the other two models are discussed in Supplementary Section 3). The first row in Fig. 1 shows the three integer-order variations of the classic SIR model and the transition graphs between the epidemiological classes in these models. The second row in Fig. 1 shows the three fractional models with different fixed/time-variable fractional orders. In particular, we use the Caputo fractional derivative^{32–34} of variable order $\kappa(t) \in (0, 1)$. Let $u(t) \in C^1$ be a differentiable function, the first derivative of which is continuous, and Γ the Gamma function, then the Caputo fractional derivative is given as

$${}_0^C \mathcal{D}_t^{\kappa(t)} u(t) = \frac{1}{\Gamma(1 - \kappa(t))} \int_0^{\kappa(t)} (t - s)^{-\kappa(t)} u'(s) ds, \quad (2)$$

where the variable order $\kappa(t)$ allows us to adjust the effect of non-locality as the dynamics evolves. A similar definition holds if the order is constant. The third row in Fig. 1 shows the time-delay model and different epidemiological classes. A detailed definition of the corresponding differential equations for all of these models and the complete list of parameters are provided in Supplementary Section 3.

Experimental set-up. We carried out our analysis based on different datasets for COVID-19 from New York City (NYC)^{31,35}, Michigan state (MI)^{36,37}, Rhode Island state (RI)³⁸ and Italy³⁹. In each case, the datasets were pre-processed by applying a moving averaging window of seven days to smooth the weekday–weekend fluctuations in the reporting of the outbreak. Because not all of the epidemiological states are tractable in practice, the reported data of individuals are restricted to only a few compartments. Each dataset reports different epidemiological classes, leading to different formulations of PINNs (a detailed explanation of the different datasets for the PINN formulation is provided in Supplementary Section 6). For example, NYC data report the cumulative infectious I^c , hospitalized H^c and death D^c cases. The MI data instead report the current values of hospitalized individuals H , while the RI data report both the current H and the cumulative hospitalized individuals H^c . The Italy dataset reports the current I , R and D . Typically, the D compartment in the epidemiological models does not have an outflow and therefore we can have $D^c = D$. Given the cumulative values, we can obtain the daily increase in (new) values for infectious, hospitalized and deaths cases by $I^{\text{new}}(t) = I^c(t) - I^c(t-1)$, $H^{\text{new}}(t) = H^c(t) - H^c(t-1)$ and $D^{\text{new}}(t) = D^c(t) - D^c(t-1)$.

In the following, we show different aspects of our formulation and the results based on the NYC and Italy datasets. We provide the results based on other datasets in Supplementary Section 4. We consider the effect of vaccination in integer-order and time-delay models by adding the extra connection from S to R with rate $\frac{V(t)S(t)}{N}$. Here, $V(t)$ denotes the number of effective vaccinated individuals at time t , which is calculated by $V(t) = 0.52(D_1(t) - D_2(t)) + 0.95D_2(t)$ (ref. 17). The terms D_1 and D_2 denote the number of individuals that have received the first and second doses of vaccine, respectively. The first dose of COVID-19 vaccine in the United States was administered on 14 December 2020. Accounting for two weeks delay in building immunity after the first dose of vaccination, we assume that the number of effective vaccinated individuals to be zero before 1 January 2021.

Simultaneous inference of unknown parameters and dynamics. The PINN formulation encodes the mathematical model into the network and forms the loss function, comprised of two parts. The first part is the mismatch between the network output and the available data, and the second part comprises the ODE residuals. By minimizing the loss function, we optimize the network parameters to learn the data and simultaneously infer the unobserved dynamics by satisfying the ODEs. The success of PINN in this setting lies in its flexibility in allowing the model parameters to be time-dependent functions represented by separate neural networks. For the integer-order models we carry out structural and practical identifiability analyses in Supplementary Section 5. We note that, in the fractional models, we infer the time-dependent fractional orders as we fix the parameters and let the fractional operators change in time.

We train PINNs for integer-order models $\mathbb{I}_1$, $\mathbb{I}_2$ and $\mathbb{I}_3$ and time-delay model $\mathbb{T}_1$ based on the NYC data³¹. In Fig. 2a we show the fitting of PINN formulations for the four models to the available data. The trained network can accurately fit the different fluctuations in

Fig. 2 | PINNs inference using the integer-order models $\mathbb{I}_1$, $\mathbb{I}_2$ and $\mathbb{I}_3$ and time-delay model $\mathbb{T}_1$ for NYC. **a, Fitting (black line) to the available data (red dots) of daily infectious, hospitalized and death cases. **b**, Inference of time-dependent parameters ($\beta_I(t)$, $p(t)$, $q(t)$) for each model. Here the inferred delay is $d=3.10$ for the time-delay model $\mathbb{T}_1$. **c**, Inference of unobserved dynamics. Each graph shows the evolution of one epidemiological compartment. The **Q** and **P** compartments only exist in model $\mathbb{I}_3$ and models $\mathbb{I}_2$ and $\mathbb{I}_3$, respectively. **d**, Prediction and uncertainty quantification of daily infectious, hospitalized and death cases for a four-month time window using model $\mathbb{I}_3$. The vertical line divides the fitting and prediction window. The two curves and shaded area in the prediction window correspond to 10% and 20% standard deviations in β_I . We have included the new available data (plus symbols) for the prediction period that were not used in the fitting to show that the predictions are correct.**

the data. In Fig. 2b,c we show the inferred time-dependent parameters ($\beta_I(t)$, $p(t)$, $q(t)$) and the inferred unobserved compartments for the four models, respectively. The plots demonstrate an almost

similar trend throughout the spread of the virus. Some of the models have additional compartments, for example, the **Q** compartment in model $\mathbb{I}_3$.

Forecasting with uncertainty. The inferred model from PINNs yields a predictive model that we use to forecast future dynamics by considering different control measures for each model. In particular, we can predict the evolution of dynamics using the integer-order models $\mathbb{I}_1$, $\mathbb{I}_2$ and $\mathbb{I}_3$ and the time-delay model $\mathbb{T}_1$. Figure 2d shows the prediction of new infectious, hospitalized and death cases for NYC data based on model $\mathbb{I}_3$. In the prediction window, we fix the parameters $\beta_i(t)$, $p(t)$ and $q(t)$ at their final values from the training window. However, we postulate that the effect of several different control measures in the community transmission rate $\beta_i(t)$ can be captured by adding uncertainty bounds of 15% and 30% to its mean value in the prediction window. By using standard forward propagation techniques, that is, Monte Carlo⁴⁰ and probabilistic collocation methods⁴¹, we propagate the uncertainty into the prediction window. We also consider different vaccination scenarios by considering different vaccination rates (Vac. rate) in calculating the effective vaccination per day ($V(t)$) in the prediction window by

In particular, for the NYC dataset³⁵, we set the starting day of the effective vaccination as $t_{\text{Vac.start}} = 300$ and we cap (saturate) the effective value of vaccination by $\text{Vac. cap} = 30,000$. We consider two vaccination rates and adjust $t_{\text{Vac.cap}}$; for example, $\text{Vac. rate} = 600$ and $t_{\text{Vac.cap}} = 350$.

Fractional models. The structure of PINNs for fractional models is usually more complex than for the other models. This is mainly because of the fractional operators, for which the automatic differentiation technology is not applicable, and we hence need to resort to numerical discretization, such as the L1 scheme^{42,43}. This formulation, namely fractional PINNs (fPINNs), was first developed by Pang et al.⁴⁴ and we extend it to apply to our problem. The other source of the complexity of the PINN structure for fractional models is the time-dependent fractional orders. The PINN formulation uses a separate neural network to represent each fractional order $\kappa_i(t)$. In particular, Fig. 3 shows the fPINNs schematic for the fractional model $\mathbb{F}_3$. We use a (relatively large) neural network to represent the states $\mathbf{U}^{(\mathbb{F}_3)}(t)$ (the green-shaded box) and five separate neural networks to represent the fractional orders $\kappa_i(t)$, $i=1, \dots, 5$ (the red-shaded boxes). We discuss our methodology in the Methods and provide the formulation in Supplementary Section 6. The PINN formulation makes the inverse problem more efficient as it uses the network to represent the time-dependent parameters and integrates data and differential equations through the loss function. Therefore, it provides a flexible computational tool to perform inference in fractional models.

In Fig. 4a we only use the available data ($\mathbf{I}^{\text{new}}$, $\mathbf{H}^{\text{new}}$, $\mathbf{D}^{\text{new}}$) to train the network. We can observe that the inferred dynamics and

Fig. 4 | Identifiability study for PINNs for fractional-order model $\mathbb{F}_3$. Instead of time-dependent parameters, here we aim to infer different time-dependent fractional operators for $\kappa_i(t)$, $i = 1, \dots, 5$. **a–c**, In the first row are plotted the inferred unobserved dynamics from fPINNs against the corresponding simulation results of the integer-order SIHDR model. In the second row, we show the inferred corresponding time-dependent fractional orders for different available data for NYC. In **a**, we employ only the available data on I^{new} , H^{new} and D^{new} for NYC. In **b**, in addition to the measures in **a**, we also assume that we have additional simulated data on I and H . In **c**, we have additional simulated data on S , I , H , D and R . The uncertainty bands shown reflect the robustness of each case. In each panel in the top rows, black curves show the results for model $\mathbb{F}_3$ and red the results for the corresponding integer-order model.

the corresponding fractional order have relatively large uncertainty bounds and are different from the results of the integer-order model. We note, however, that the **D** compartment has the smallest uncertainty bound. This is because the **D** compartment does not have any outflow and hence $\mathbf{D} = \mathbf{D}^c$. It is counterintuitive, however, that its corresponding fractional order has the largest uncertainty bounds. This is due to coupling of the dynamics of the **D** compartment with other **H** and **R** compartments that contain large uncertainty bounds. In Fig. 4b, in addition to the data in Fig. 4a, we use the simulated data of the **I** and **H** compartments. These values are not available in practice for the NYC data, so we obtain them from the corresponding integer-order model and use them to investigate the identifiability of fractional model $\mathbb{F}_3$. We can see that the inference can be improved by including the additional simulated data. In Fig. 4c we show that, if we add extra data we can further reduce the uncertainty bounds in the inferred time-dependent parameters. It is interesting to observe in Fig. 4c that a fractional model with time-dependent fractional order can resemble the dynamics of its integer-order counterparts

with time-dependent parameters. This example supports the two alternate paradigms of modeling philosophy, discussed at the beginning of this Article.

We note that, although the fractional models are capable of accounting for memory effects via non-local operators, and their application has been motivated in the literature to a great extent, the crucial but less-studied element is the development of a robust computational framework to implement the inverse problem of inferring the time-dependent fractional orders. For example, Jahanshahi et al.¹⁹ suggest the use of time-dependent variable fractional orders; however, in their implementation, the fractional orders are parameterized with a few constant parameters as piecewise constant functions to render the inverse problem feasible in their formulation. The under-parameterization of fractional orders may lead to inaccurate results. More importantly, when different fractional orders are considered, the total population is not conserved automatically by the system of ODEs and should be enforced in the simulation (more details are provided in the Discussion). We show here that our formulation has the

Fig. 5 | PINNs results of the fractional-order model $\mathbb{F}_2$ based on Italy data. a, fPINN fitting (black) to the available data for **S**, **I**, **R** and **D** (red). In each graph, we also show the fPINN inference of the corresponding time-dependent fractional order (blue) and compare it with the piecewise fractional orders (dashed blue) from ref. ¹⁹. **b**, Comparison of the piecewise parameter $r(t)$ (blue) and the fPINN inference of the time-dependent parameter $r(t)$ (black). The piecewise fractional orders and parameters are provided in ref. ¹⁹. The other parameters of the model are fixed as $a = 0.0215$ and $b = 0.0048$ (ref. ¹⁹) in the simulation.

capacity to efficiently extend the parameterization of time-dependent fractional orders via neural networks, and we implement an efficient computational framework to infer them from data. For comparison, we consider the identifiability of fractional model $\mathbb{F}_2$ based on the Italy data. This is a practical example of the case discussed in Fig. 4c. The Italy data report the current values **I**, **R** and **D** and $S = N - I - D - R$, which comprise the four compartments in the fractional model $\mathbb{F}_2$ with time-dependent fractional orders. Because we have access to a large amount of data in this case, we do not fix the infection rate parameter ($r(t)$, Supplementary Tables 1 and 3) and let it be represented by a separate neural network. The left column in Fig. 5a shows the accurate fitting to the available data and the right column shows the inferred time-varying parameters $\kappa_i(t)$, $i = 1, 2, 3, 4$, and $r(t)$. These results demonstrate that, even in the case with access to data for all compartments, the inference of fractional orders is not very robust and still exhibits some uncertainties, which is a reflection of the lack of identifiability of the model. The choice of piecewise constant parameters in the work of Jahanshahi et al.¹⁹ simplifies the problem but leads to erroneous results.

Discussion

One of the main drawbacks of lumped-compartment epidemiological models is the lack of uniqueness of the parameters, especially when mathematical models with time-dependent parameters are used to explore possible future scenarios for control measures, evaluate retrospectively the efficacy of specific interventions and identify prospective strategies⁴⁵. At the present time there are no rigorous a priori identifiability analyses for integer-order models with time-dependent parameters or time-dependent fractional orders. Although this imposes a limitation on the underlying mathematical formulation, in the present work we resorted to systematic numerical experimentation to provide a posteriori some measures of identifiability using uncertainty quantification. We have learned some useful lessons by fitting the epidemic data using nine different epidemiological models (five integer-order, one time-delay and three fractional-order models; Supplementary Section 3). In the following, we summarize and discuss some observations related to our findings.

The different sources of uncertainty are the primary limitations to the efficiency of epidemiological models. The novelty of

Fig. 6 | The main challenges for epidemiological models. **a**, Number of daily infectious cases (I^{new}) for NYC, with blue representing the raw reported data and red representing the seven-day average. **b**, Additional available data of current hospitalized individuals for MI to fit the parameter $q(t)$ expressing the proportion of disease-related deaths from the **H** class, which is only available in some states. **c**, The inherent time delay in the only available data for NYC. **d**, Inferred time-varying parameters ($\beta(t)$, $p(t)$, $q(t)$) for the integer-order and delay models $\mathbb{I}_1$, $\mathbb{I}_2$, $\mathbb{I}_3$ and $\mathbb{T}_1$ based on NYC data. **e**, Although there are a priori identifiability analyses for integer-order models, there is no identifiability methodology for fractional models. Shown is a comparison of the inferred dynamics of models $\mathbb{F}_1$ and $\mathbb{I}_1$ for NYC data. The shaded region is the uncertainty of the fractional model. **f**, Inferred different time-dependent fractional orders of model $\mathbb{F}_3$ using simulated data based on the integer-order model for NYC. **g**, Left: plot of the inferred dynamics from integer-order and time-delay models, which show differences but consistent trends. Right: inferred dynamics of model $\mathbb{F}_1$, which shows an erroneous trend due to the lack of identifiability (window in yellow). The SIHDR model in **e** is defined as model $\mathbb{I}_5$ in Supplementary Section 3.4.

coronavirus naturally leads to many uncertainties¹, there being many unknown factors, including the biological features of transmission, the different mutations of the virus and pathogen behavior and concentrations. The most systematic source of uncertainty that can adversely affect almost all mathematical models' outputs is that the exact number of infected individuals is unknown. The onset of an outbreak always precedes the start of data reporting, so there is a lack of initial data at the beginning of the outbreak and inaccurate parameter estimation in all models during that initial period, especially for those models with memory effects. For example, if we use the parameters estimated in the paper by Hao et al.⁴⁶ for fitting data up to the present time, the results are not entirely accurate. The reporting practices for COVID-19 case data are not consistent among different populations and the reported data contain intermittent artifacts as a result of weekday/weekend effects (Fig. 6a). For example, data reporting in NYC differs from that in RI, despite the geographic proximity, and the same model may perform differently if applied to datasets from those two sources. Accordingly, because of the multitude of uncertainties, quantifying the parametric input uncertainty is not sufficient⁵. Consequently, model selection becomes the major issue, in our opinion, not only in terms of how to best fit the data but also in relation to how robust the model is, given these uncertainties.

The use of fixed parameters in lumped-compartment epidemiological models is not appropriate because of the long duration of the pandemic and the associated multi-rate dynamics (Supplementary Fig. 4). Indeed, as the disease symptoms, concentration and behavior of the pathogen are changing throughout the course of the disease^{47,48}, the models should accommodate time-dependent parameters. In our formulation, the integer-order models account for this time variability (Fig. 6d). Additionally, the fractional mod-

els include different memory effects for each epidemiological class by considering different time-dependent fractional orders (Fig. 6f). In Supplementary Fig. 4 we show that we can easily infer the fixed parameters by fitting the data; however, due to the non-uniqueness of the problem, running the models forward with the inferred parameters leads to erroneous dynamics. This error is due to the identifiability issue for fractional models. The identifiability of integer-order models with piecewise constant parameters has been systematically investigated in the work by Zhang and colleagues¹⁷. However, such an analysis is not available for the fractional models, which may exhibit higher uncertainty (Fig. 6e). Here, we have investigated the robustness of fractional-order models and we have concluded that, due to their fragility, if any fractional model should be considered, its identifiability must be investigated at least numerically. Even if fitted accurately to the available data, a non-identifiable model may still show some serious artifacts that are not intuitively consistent (Fig. 6g). It is also important to comment on dimensional consistency in fractional models. In general, the dynamics of each compartment in the fractional epidemiological model can be obtained by

$$\frac{(\Delta t)^{\kappa-1}}{\Gamma(1+\kappa)} {}^C D_t^\kappa u(t) = \text{inflow} - \text{outflow}$$

where the timescale Δt denotes a characteristic time of observation, which amounts to a built-in scale effect. The scale effect goes away as $\kappa \rightarrow 1$, $\Delta t^{\kappa-1} \rightarrow \Delta t^0 \rightarrow 1$, and thus the equation recovers the classical continuity equation. This scaling factor is also useful in practice as it ensures the dimensional consistency of the units in the fractional model (more details are provided in Supplementary Section 3.3). Another consideration for the fractional models is the

conservation of total population. In most integer-order models, the conservation of the total population is satisfied automatically by the system of ODEs. This is not the case for the fractional models when different fractional orders are considered. We have observed that, in many published studies using the fractional modeling, this fact has been overlooked and the population conservation has not been included in the system of equations. This can result in an erroneous output of the model, especially because compartmental models require this assumption.

The purely statistical approaches are generally not well suited for long-term predictions of epidemiological dynamics (for example, Supplementary Fig. 1). They do not have the ability to account for how the transmission occurs, so most of the forecasting models limit their projections to one week or a few weeks ahead⁴⁹. The mechanistic models, if their time-dependent parameters are inferred correctly from data, are often helpful to quantify different sources of uncertainty and examine the implications of various assumptions and control measures about a highly nonlinear process that is hard to predict using only intuition or statistical models. However, they are constrained by their limitations, what we assume and what we do not know. For example, the proportion of hospitalized individuals, $p(t)$, may reach an asymptotic value after some time, but the community transmission rate, $\beta_c(t)$, is highly influenced by control measures, mobility, mutant variations and so on, and hence we specify it as a random variable for future predictions (Fig. 2d). More specifically, we have investigated three different approaches to extrapolate for short- and long-term predictions (Supplementary Figs. 5 and 6), and we have found that by using PINNs we can extrapolate the time-dependent parameters, at least for the short term, and this would lead to more accurate predictions compared to other approaches. Furthermore, we have evaluated the integer-order models, which seem to be more robust than the fractional-order models, and among them the model $\mathbb{I}_3$ gives the best prediction (Supplementary Fig. 6) based on the new available data from March to June, which we have not used in our previous training.

Given the discussed observations and limitations, we recommend the use of multiple models with time-dependent parameters tailored to a specific region and specific data availability, and quantify both parametric uncertainty and model uncertainty. The use of both structural and practical identifiability analyses could help to determine the proper parameter ranges and the consistency of the inferred values.

Methods

PINNs were first introduced in the work of Raissi and colleagues¹⁶. Since then, PINNs have been successfully applied in solving forward and inverse problems in many practical applications^{50–57}. An analysis and convergence study of PINNs was carried out by Shin and others⁵⁸. Figure 3 presents a schematic of such PINNs that is almost the same for all the models considered in this Article. The depicted structure in Fig. 3 is specifically related to the fractional-order model $\mathbb{I}_3$ with five time-dependent parameters.

In the PINN formulation, we use separate deep neural networks with input t to represent the states $\mathbf{U}(t)$ and (time-dependent) parameters. Each network is parameterized by a set of parameters Θ as weights and biases of the network. Thus, we let $\mathbf{U}(t) \approx \mathbf{U}_{\text{NN}}(t; \Theta_U)$ (green-shaded box, Fig. 3). For integer-order models $\mathbb{I}_1$, $\mathbb{I}_2$ and $\mathbb{I}_3$ and time-delay model $\mathbb{T}_1$, we let $\beta_I(t) \approx \beta_{\text{INN}}(t; \Theta_\beta)$, $p(t) \approx p_{\text{NN}}(t; \Theta_p)$ and $q(t) \approx q_{\text{NN}}(t; \Theta_q)$. For fractional-order models $\mathbb{F}_1$, $\mathbb{F}_2$ and $\mathbb{F}_3$, we let $\kappa_i(t) \approx \kappa_{\text{INN}}(t; \Theta_{\kappa_i})$ for $i = 1, \dots, 5$ (red-shaded boxes, Fig. 3). Using equation (1), we define the residual of equation as $\mathcal{R}_{\text{NN}}(t) = \mathcal{L}\mathbf{U}_{\text{NN}}(t) - \mathcal{F}(\mathbf{U}_{\text{NN}}(t); \lambda)$ (orange-shaded box, Fig. 3). This residual is the measure of the approximation $\mathbf{U}_{\text{NN}}$ satisfying the ODEs, and ideally the exact solution is recovered when the residual is identically zero. We define two finite sets of training points $\{t_u\}_{j=1}^{N_u}$ and residual points $\{t_r\}_{j=1}^{N_r}$. The training points are the points where data are available and the residual points are the points where the residual $\mathcal{R}_{\text{NN}}(t)$ is satisfied and they are freely available all over the computational domain. Therefore, we define the loss function of PINN as

$$L(\Theta, \lambda) = \omega_u \text{m.s.e.}_u + \omega_r \text{m.s.e.}_r \\ = \omega_u \frac{1}{N_u} \sum_{j=1}^{N_u} \left| \mathbf{U}_{\text{NN}}(t_u^j) - \mathbf{U}(t_u^j) \right|^2 + \omega_r \frac{1}{N_r} \sum_{j=1}^{N_r} \left| \mathcal{R}_{\text{NN}}(t_r^j) \right|^2, \quad (4)$$

where m.s.e. is the mean squared error. The loss function of PINNs is composed of two terms—m.s.e._u and m.s.e._r, which are defined by the last part of the above equation. Parameters $\{\omega_u, \omega_r\}$ denote the weight coefficients in the loss function that can balance the optimization effort between learning the data and satisfying the ODEs. They may be user-specified or tuned manually or automatically, for example, in practice based on the numerical experiment in each problem⁵⁹.

In all cases, the derivatives of the network with respect to the input t and network parameters Θ are computed by applying the chain rule for differentiating compositions of functions using the automatic differentiation⁶⁰. In particular, we use TensorFlow programming⁶¹ for automatic differentiation and deep learning computations. Detailed derivations of the loss functions for different models and different datasets are provided in Supplementary Section 6.

Data availability

All datasets used in this paper are publicly available. NYC data can be downloaded from refs. ³¹ and ³⁵. MI data can be downloaded from refs. ³⁶ and ³⁷. RI data can be downloaded from ref. ³⁸. Italy data can be downloaded from ref. ³⁹. Source data are provided with this paper.

Code availability

The version of PINN-COVID and the data that were used to generate our results are publicly available^{62,63}.

Received: 5 April 2021; Accepted: 12 October 2021;

Published online: 22 November 2021

References

- Kreps, S. E. & Kriner, D. L. Model uncertainty, political contestation and public trust in science: evidence from the COVID-19 pandemic. *Sci. Adv.* **6**, eabd4563 (2020).
- Holmdahl, I. & Buckee, C. Wrong but useful—what COVID-19 epidemiologic models can and cannot tell us. *N. Engl. J. Med.* **383**, 303–305 (2020).
- Science Brief: Emerging SARS-CoV-2 Variants (CDC, 2019); <https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html>
- Cobey, S. Modeling infectious disease dynamics. *Science* **368**, 713–714 (2020).
- Edeling, W. et al. The impact of uncertainty on predictions of the CovidSim epidemiological code. *Nat. Comput. Sci.* **1**, 128–135 (2021).
- Cramer, E. et al. reichlab/covid19-forecast-hub: release for Zenodo, 20210816 <https://doi.org/10.5281/zenodo.5208210> (2020).
- Chakraborty, P., Meerschaert, M. M. & Lim, C. Y. Parameter estimation for fractional transport: a particle-tracking approach. *Water Resources Res.* **45**, W10415 (2009).
- Cho, Y., Kim, I. & Sheen, D. A fractional-order model for MINMOD millennium. *Math. Biosci.* **262**, 36–45 (2015).
- Kelly, J. F., Bolster, D., Meerschaert, M. M., Drummond, J. D. & Packman, A. I. FracFit: a robust parameter estimation tool for fractional calculus models. *Water Resources Res.* **53**, 2559–2567 (2017).
- Lim, C. Y., Meerschaert, M. M. & Scheffler, H. P. Parameter estimation for operator scaling random fields. *J. Multivariate Anal.* **123**, 172–183 (2014).
- Ghazizadeh, H. R., Azimi, A. & Maerefat, M. An inverse problem to estimate relaxation parameter and order of fractionality in fractional single-phase-lag heat equation. *Int. J. Heat Mass Transfer* **55**, 2095–2101 (2012).
- Yu, B., Jiang, X. Y. & Qi, H. T. Numerical method for the estimation of the fractional parameters in the fractional mobile/immobile advection-diffusion model. *Int. J. Comput. Math.* **95**, 1131–1150 (2018).
- Suzuki, J. L. & Zayernouri, M. A self-singularity-capturing scheme for fractional differential equations. *Int. J. Comput. Math.* **98.5**, 933–960 (2021).
- Pang, G. F., Perdikaris, P., Cai, W. & Karniadakis, G. E. Discovering variable fractional orders of advection-dispersion equations from field data using multi-fidelity Bayesian optimization. *J. Comput. Phys.* **348**, 694–714 (2017).
- Kharazmi, E. & Zayernouri, M. Fractional sensitivity equation method: application to fractional model construction. *J. Sci. Comput.* **80**, 110–140 (2019).
- Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
- Zhang, S., Ponce, J., Zhang, Z., Lin, G. & Karniadakis, G. An integrated framework for building trustworthy data-driven epidemiological models: application to the COVID-19 outbreak in New York City. *PLoS Comput. Biol.* **17**, e1009334 (2021).
- Zheng, X. C. & Wang, H. Uniquely identifying the variable order of time-fractional partial differential equations on general multi-dimensional domains. *Inverse Problems Sci. Eng.* **29**, 1401–1411 (2020).
- Jahanshahi, H., Munoz-Pacheco, J. M., Bekiros, S. & Alotaibi, N. D. A fractional-order SIRD model with time-dependent memory indexes for encompassing the multi-fractional characteristics of the COVID-19. *Chaos Solitons Fractals* **143**, 110632 (2021).

20. Taghvaei, A., Georgiou, T. T., Norton, L. & Tannenbaum, A. Fractional SIR epidemiological models. *Sci. Rep.* **10**, 20882 (2020).
21. Ma, Z. E. & Jin, Z. The stability of an SIR epidemic model with time delays. *Math. Biosci. Eng.* **3**, 101–109 (2012).
22. Ma, Z. E., Zhou, Y. C. & Wu, J. H. *Modeling and Dynamics of Infectious Diseases* (World Scientific, 2009).
23. Beretta, E. & Takeuchi, Y. Global stability of an SIR epidemic model with time delays. *J. Math. Biol.* **33**, 250–260 (1995).
24. Beretta, E., Hara, T., Ma, W. & Takeuchi, Y. Global asymptotic stability of an SIR epidemic model with distributed time delay. *Nonlinear Anal.* **47**, 4107–4115 (2001).
25. Angstmann, C. N., Henry, B. I. & McGann, A. V. A fractional order recovery SIR model from a stochastic process. *Bull. Math. Biol.* **78**, 468–499 (2016).
26. Jha, P. K., Cao, L. & Oden, J. T. Bayesian-based predictions of COVID-19 evolution in Texas using multispecies mixture-theoretic continuum models. *Comput. Mech.* **66**, 1055–1068 (2020).
27. Christakos, G., Zhang, C. T. & He, J. Y. A traveling epidemic model of space-time disease spread. *Stochastic Environ. Res. Risk Assess.* **31**, 305–314 (2017).
28. Lotfi, E. M., Maziane, M., Hattaf, K. & Yousfi, N. Partial differential equations of an epidemic model with spatial diffusion. *Int. J. Partial Differ. Eqn.* **2014**, 186437 (2014).
29. Horwitz, L. et al. Trends in COVID-19 risk-adjusted mortality rates in a single health system. Preprint at <https://doi.org/10.1101/2020.08.11.20172775> (2020).
30. Petrilli, C. M. et al. Factors associated with hospitalization and critical illness among 4,103 patients with COVID-19 disease in New York City. Preprint at <https://doi.org/10.1101/2020.04.08.20057794> (2020).
31. NYC Coronavirus Disease 2019 (COVID-19) Data. <https://github.com/nychealth/coronavirus-data>
32. Cai, M. & Li, C. P. Numerical approaches to fractional integrals and derivatives: a review. *Mathematics* **8**, 43 (2020).
33. Li, C. P. & Cai, M. *Theory and Numerical Approximations of Fractional Integrals and Derivatives* (SIAM, 2019).
34. Li, C. P. & Zeng, F. H. *Numerical Methods for Fractional Calculus* (CRC Press, 2015).
35. NYC Health. COVID-19 data: vaccines. <https://www1.nyc.gov/site/doh/covid/covid-19-data-vaccines.page>
36. The COVID Tracking Project. Michigan overview. <https://covidtracking.com/data/state/michigan>
37. Michigan.gov. COVID-19 vaccine dashboard. https://www.michigan.gov/coronavirus/0,9753,7-406-98178_103214-547150-,00.html
38. COVID-19 Rhode Island data.
39. COVID-19 data repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. GitHub <https://github.com/CSSEGISandData/COVID-19>
40. Fishman, G. S. *Monte Carlo: Concepts, Algorithms and Applications* (Springer, 1996).
41. Xiu, D. B. & Hesthaven, J. S. High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.* **27**, 1118–1139 (2005).
42. Sun, Z. Z. & Wu, X. N. A fully discrete difference scheme for a diffusion-wave system. *Appl. Numer. Math.* **56**, 193–209 (2006).
43. Lin, Y. M. & Xu, C. J. Finite difference/spectral approximations for the time-fractional diffusion equation. *J. Comput. Phys.* **225**, 1533–1552 (2007).
44. Pang, G. F., Lu, L. & Karniadakis, G. E. fPINNs: fractional physics-informed neural networks. *SIAM J. Sci. Comput.* **41**, A2603–A2626 (2019).
45. Metcalf, C. J. E., Morris, D. H. & Park, S. W. Mathematical models to guide pandemic response. *Science* **369**, 368–369 (2020).
46. Hao, X. J. et al. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* **584**, 420–424 (2020).
47. Fine, P., Eames, K. & Heymann, D. L. ‘Herd immunity’: a rough guide. *Clin. Infect. Dis.* **52**, 911–916 (2011).
48. Peak, C. M., Childs, L. M., Grad, Y. H. & Buckee, C. O. Comparing nonpharmaceutical interventions for containing emerging epidemics. *Proc. Natl Acad. Sci. USA* **114**, 4023–4028 (2017).
49. Jewell, N. P., Lewnard, J. A. & Jewell, B. L. Caution warranted: using the institute for health metrics and evaluation model for predicting the course of the COVID-19 pandemic. *Ann. Intern. Med.* **173**, 226–227 (2020).
50. Raissi, M., Babaee, H. & Givi, P. Deep learning of turbulent scalar mixing. *Phys. Rev. Fluids* **4**, 124501 (2019).
51. Mao, Z. P., Jagtap, A. D. & Karniadakis, G. E. Physics-informed neural networks for high-speed flows. *Comput. Methods Appl. Mech. Eng.* **360**, 672112789 (2020).
52. Yang, L., Zhang, D. K. & Karniadakis, G. E. Physics-informed generative adversarial networks for stochastic differential equations. *SIAM J. Sci. Comput.* **42**, A292–A317 (2020).
53. Kharazmi, E., Zhang, Z. Q. & Karniadakis, G. E. Variational physics-informed neural networks for solving partial differential equations. Preprint at <https://arxiv.org/abs/1912.00873v1> (2019).
54. Kharazmi, E., Zhang, Z. Q. & Karniadakis, G. E. hp-VPINNs: variational physics-informed neural networks with domain decomposition. *Comput. Methods Appl. Mech. Eng.* **374**, 113547 (2021).
55. Jagtap, A. D., Kawaguchi, K. & Karniadakis, G. E. Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. *J. Comput. Phys.* **404**, 109136 (2019).
56. Jagtap, A. D., Kawaguchi, K. & Karniadakis, G. E. Locally adaptive activation functions with slope recovery term for deep and physics-informed neural networks. *Proc. R. Soc. A* **476**, <https://doi.org/10.1098/rspa.2020.0334> (2020).
57. Jagtap, A. D., Kharazmi, E. & Karniadakis, G. E. Conservative physics-informed neural networks on discrete domains for conservation laws: applications to forward and inverse problems. *Comput. Methods Appl. Mech. Eng.* **365**, 113028 (2020).
58. Shin, Y., Darbon, J. & Karniadakis, G. E. On the convergence and generalization of physics informed neural networks. Preprint at <https://arxiv.org/abs/2004.01806> (2020).
59. Wang, S. F., Teng, Y. J. & Perdikaris, P. Understanding and mitigating gradient pathologies in physics-informed neural networks. Preprint at <https://arxiv.org/abs/2001.04536> (2020).
60. Baydin, A., Pearlmutter, B. A., Radul, A. A. & Siskind, J. M. Automatic differentiation in machine learning: a survey. *J. Mach. Learn. Res.* **18**, 5595–5637 (2017).
61. Abadi, M. et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. Preprint at <https://arxiv.org/abs/1603.04467> (2016).
62. Kharazmi, E. et al. Identifiability and predictability of integer- and fractional-order epidemiological models using physics-informed neural networks. Zenodo <https://doi.org/10.5281/zenodo.5565308> (2021).
63. Kharazmi, E. & Cai, M. PINN-COVID. GitHub <https://github.com/ehsankharazmi/PINN-COVID> (2021).

Acknowledgements

M.C. acknowledges support by the China Scholarship Council (no. 201906890040). G.E.K. and G.L. acknowledge support by the MURI/ARO (W911NF-15-1-0562).

Author contributions

G.E.K., G.L. and E.K. designed the study and supervised the project. E.K. and M.C. developed the integer and fractional models and ran the simulations. X.Z. developed the time-delay model and ran the simulations. All authors analyzed the results and contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43588-021-00158-0>.

Correspondence and requests for materials should be addressed to George Em Karniadakis.

Reprints and permissions information is available at www.nature.com/reprints.

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

Reviewer recognition statement: *Nature Computational Science* thanks Amirhossein Taghvaei, Sara Grundel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Handling editor: Fernando Chirigati, in collaboration with the *Nature Computational Science* team.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021