



---

*Research article*

## **Augmented data and neural networks for robust epidemic forecasting: Application to COVID-19 in Italy**

**Giacomo Dimarco<sup>1</sup>, Federica Ferrarese<sup>1,\*</sup> and Lorenzo Pareschi<sup>1,2</sup>**

<sup>1</sup> Department of Mathematics and Computer Science & Center for Modeling, Computing and Statistics (CMCS), University of Ferrara, via Machiavelli 30, 44121 Ferrara, Italy

<sup>2</sup> Maxwell Institute and Department of Mathematics School of Mathematical and Computer Sciences Heriot-Watt University, Edinburgh, UK

\* **Correspondence:** Email: federica.ferrarese@unife.it.

**Abstract:** In this work, we propose a data augmentation strategy aimed at improving the training phase of neural networks and, consequently, the accuracy of their predictions. Our approach relies on generating synthetic data through a suitable compartmental model combined with the incorporation of uncertainty. Available data are used to calibrate the model, which is further integrated with deep learning techniques to produce additional synthetic data for training. The results show that neural networks trained on these augmented datasets exhibit significantly improved predictive performances. In particular, we focus on two different neural network architectures: Physics-Informed Neural Networks (PINNs) and Nonlinear Autoregressive (NAR) models. The NAR approach proves especially effective for short-term forecasting, thereby providing accurate quantitative estimates by directly learning the dynamics from data and avoiding the additional computational cost of embedding physical constraints into the training. In contrast, PINNs yield less accurate quantitative predictions but capture the qualitative long-term behavior of the system, thus making them more suitable to explore broader dynamical trends. Numerical simulations of the second phase of the COVID-19 pandemic in the Lombardy region (Italy) validate the effectiveness of the proposed approach.

**Keywords:** mathematical epidemiology; physics-informed neural networks; non-linear autoregressive deep learning; COVID-19 data

---

### **1. Introduction**

Throughout history, pandemics have had a profound impact on global health, economies, and everyday life [1–4]. From the Spanish flu in 1918 to more recent outbreaks such as the H1N1 influenza in 2009 and the COVID-19 pandemic, the repeated emergence of infectious diseases has emphasized

the urgent need for timely and accurate response strategies [5–9]. In this context, mathematical modeling plays a crucial role in predicting disease transmission, assessing the effectiveness of intervention measures, and informing public health decisions. Among the most commonly used approaches are traditional compartmental models, such as the Susceptible-Infected-Recovered (SIR) framework, which are valued for their simplicity and interpretability [10–12]. These models classify the population into key compartments: Susceptible individuals, who are at risk of infection; Infected individuals, who can spread the disease; and Recovered individuals, who have either recovered or died and are no longer infectious. The basic SIR model can be extended to incorporate additional compartments [13, 14]. For example, the SIAR model introduces an asymptomatic class, especially relevant for diseases like COVID-19, where asymptomatic individuals significantly contribute to transmission [15, 16]. Despite their simplicity, such models often fall short in capturing the full complexity and variability of real-world epidemic dynamics, especially when dealing with incomplete or uncertain data. In the early stages of an outbreak, underreporting of infections is common, thus making it essential to incorporate uncertainty in model parameters or initial conditions to achieve more realistic scenarios [17–19]. Additionally, the simplifying assumptions underlying traditional models can limit their ability to reflect population heterogeneity and the evolving nature of outbreaks. To address these limitations, more sophisticated models have been proposed that allow for time and state dependent transmission rates [20–22]. These enhancements enable a better representation of intervention measures—such as lockdowns, which, in the case of COVID-19, played a significant role in reducing transmission. Moreover, accounting for individual heterogeneity is essential to capture the varying behaviors observed among individuals from different groups, such as distinct age classes [23–25].

In recent years, the use of neural networks for epidemic predictions has emerged as a promising complement to traditional modeling approaches [26–30]. In particular, neural networks for time series forecasting have shown strong potential. While purely data-driven models such as feed-forward networks can be useful for interpolation, they often lack physical consistency [31–33]. In contrast, Physics-Informed Neural Networks (PINNs) incorporate the governing equations into the training [26, 34, 35], thus leading to more realistic solutions. However, the process involves automatic differentiation and other intrinsic computations, which may lead to very computationally expensive simulations. Alternative architectures, such as Recurrent Neural Networks (RNNs), offer a different approach by directly learning temporal dependencies without explicitly embedding the physical model into the training process [36–38]. A notable example is the Nonlinear Autoregressive (NAR) network, which effectively captures time dynamics by using past observations to predict future values [39]. These networks offer a data-driven and computationally efficient alternative to PINNs for short-term forecasting, thereby delivering a quantitative description of disease progression that can be particularly valuable for monitoring and managing the pandemic within hospitals. In contrast, PINNs provide a more qualitative representation of the disease dynamics, which can be especially useful to investigate strategies aimed at mitigating epidemic peaks during a pandemic.

In this work, we consider a specific compartmental model that has been shown to effectively capture the time evolution of epidemic spread in the presence of uncertain parameters [22]. This model extends a SIAR-type framework by incorporating age-structured dynamics to better reflect the pandemic's impact across different demographic groups, as well as the influence of lockdown measures. Starting from this depicted model, we use real-world data to calibrate the model parameters under uncertainty [19, 33].

Then, we introduce two different neural network architectures, namely PINNs and NAR networks, and investigate how data augmentation strategies can enhance their predictive performance. In our previous work [40], we explored such strategies in a simplified context of deterministic epidemic models by augmenting training datasets with synthetic data generated from model simulations. Here, we extend this analysis to models that contain uncertainty.

Through a series of numerical experiments, we demonstrate that NAR networks can accurately capture epidemic dynamics in both interpolation and extrapolation tasks. In short-term forecasting, they outperform PINNs, particularly when trained on augmented datasets. Additionally, we highlight the computational advantages of NAR networks: unlike PINNs, which require evaluating the underlying differential equations during training, NAR networks solely rely on data, thus leading to faster training and lower computational costs. In contrast, for long-term forecasting, PINNs provide more reliable predictions, particularly in capturing the epidemic peaks.

The rest of the paper is organized as follows: in Section 2, we introduce the mathematical models, thereby starting from the classical SIR, and we introduce a recent compartmental model which takes social behavior, age-structure, and uncertainty into account [22]; in Section 3, we detail the parameter estimation procedure under uncertainty using real world data which permits to match the model evolution with the COVID-19 epidemic spread; Section 4 presents the neural network architectures used for epidemic prediction-PINNs and NAR networks trained with data and models outcome; in Section 5, we report different numerical experiments to evaluate the networks performance, thereby making a comparison between PINNs and NARs in short and long term forecasting; and finally, Section 6 outlines possible future research directions.

## 2. Model setting: compartments, social behavior, age structure and uncertainty

In our analysis, we will consider a suitable extension of the classical SIR model [10, 12]:

$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta \frac{S(t)I(t)}{N}, \\ \frac{dI(t)}{dt} &= \beta \frac{S(t)I(t)}{N} - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t),\end{aligned}\tag{1}$$

which typically describes the spread of an infectious disease in a population of size  $N$  by partitioning it into three compartments: Susceptible (S), Infected (I), and Recovered (R). While this model has been widely adopted in the past due to its simplicity, it has been proven to fall short in capturing complex epidemic dynamics [17, 41], as it assumes, among other simplifying assumptions, the transmission and recovery rates, namely the parameters  $\beta > 0$  and  $\gamma > 0$ , to be constant. Some significant modifications were recently proposed in [20], where the authors, thereby starting from a microscopic interaction dynamics, derived a compartmental model that accounts for the role of social contacts among individuals in the spread of an epidemic. By adding an additional variable which characterizes the number of social contacts  $x \geq 0$  among individuals and by denoting with  $f_S(x, t)$ ,  $f_I(x, t)$  and  $f_R(x, t)$ , the distributions at time  $t > 0$  of the number of social contacts of the population of susceptible, infected, and recovered individuals, upon renormalization, one can fix the total distribution of social contacts to be

a probability density for all times  $t \geq 0$ :

$$\int_{\mathbb{R}_+} f(x, t) dx = 1.$$

Then, the model combines the epidemic process with the contact dynamics, which gives the following system:

$$\begin{aligned} \frac{\partial f_S(x, t)}{\partial t} &= -K(f_S, f_I)(x, t) + Q_S(f_S)(x, t), \\ \frac{\partial f_I(x, t)}{\partial t} &= K(f_S, f_I)(x, t) - \gamma f_I(x, t) + Q_I(f_I)(x, t), \\ \frac{\partial f_R(x, t)}{\partial t} &= \gamma f_I(x, t) + Q_R(f_R)(x, t), \end{aligned} \quad (2)$$

where the transmission of the infection is governed by the function  $K(f_S, f_I)$ , namely the local incidence rate, which is expressed by the following:

$$K(f_S, f_I)(x, t) = f_S(x, t) \int_{\mathbb{R}_+} \kappa(x, y) f_I(y, t) dy, \quad (3)$$

where  $\kappa(x, y)$  is called the contact function, which is a nonnegative function that grows with respect to the number of contacts  $x$  and  $y$  of the populations of susceptible and infected, respectively. A leading example reads as follows:

$$\kappa(x, y) = \beta x^\alpha y^\alpha,$$

where  $\alpha, \beta$  are positive constants, that is, by taking the incidence rate dependent on the product of the number of contacts of susceptible and infected people. The operators  $Q_J$ ,  $J \in \{S, I, R\}$  are integral operators that modify the distribution of contacts  $f_J(x, t)$ ,  $J \in \{S, I, R\}$  through repeated interactions among individuals [20]. Now, by integrating over the number of social contact and under suitable hypothesis on the operators  $Q_J$ ,  $J \in \{S, I, R\}$  that characterize the distribution of social contacts at equilibrium, one can observe that the evolution of the mass fractions obeys to a SIR-type model as follows:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta S(t)I(t)H(I(t)), \\ \frac{dI(t)}{dt} &= \beta S(t)I(t)H(I(t)) - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t). \end{aligned} \quad (4)$$

Here, for simplicity, the population size  $N$  is absorbed into the coefficient  $\beta$ , while the function  $H(I(t))$  represents the average behavior induced by microscopic interactions (2). It denotes a macroscopic incidence rate that captures time-dependent modifications to the transmission dynamics, thus reflecting behavioral responses and public health interventions such as lockdowns. The inclusion of a state-dependent transmission rate enables the model to account for a broader range of epidemic scenarios as shown in [20].

Moving forward into the modeling, we consider an additional extension to the framework (4), capable of addressing more complex scenarios, as shown in [22], that is, we include an additional compartment,  $A$ , to account for asymptomatic individuals, and more important, we incorporate uncertainty

into the model parameters. These extensions are particularly relevant for diseases such as COVID-19, where asymptomatic transmission has shown to play a crucial role [17]. Moreover, incorporating uncertainty provides a more realistic representation of data limitations, especially during the early stages of the pandemic, when the true number of infections was often significantly underreported. This permits us to enhance the forecasting capabilities of the Neural Networks, as shown later in the article. The resulting model is given by the following:

$$\begin{aligned}\frac{\partial S(t, z)}{\partial t} &= -\Lambda(t, z), \\ \frac{\partial I(t, z)}{\partial t} &= \xi(z)\Lambda(t, z) - \gamma_I(z)I(t, z), \\ \frac{\partial A(t, z)}{\partial t} &= (1 - \xi(z))\Lambda(t, z) - \gamma_A(z)A(t, z), \\ \frac{\partial R(t, z)}{\partial t} &= \gamma_I(z)I(t, z) + \gamma_A(z)A(t, z),\end{aligned}\tag{5}$$

being

$$\Lambda(t, z) = \beta(z)S(t, z)H_S(I(t, z))(H_I(I(t, z))I(t, z) + H_A(I(t, z))A(t, z)),\tag{6}$$

where  $\beta(z), \gamma_I(z), \gamma_A(z) > 0$  are the transmission and recovery rates, and the macroscopic incidence rates  $H_J(\cdot)$  for  $J \in \{S, I, A\}$  are given by the following:

$$H_S(r) = \frac{\mu(z)}{\sqrt{1 + \nu(z)r}}, \quad H_A(r) = H_S(r), \quad H_I(r) = kH_S(r),\tag{7}$$

with  $\mu(z), \nu(z) > 0$ . These incidence rates model how, both the susceptible and asymptomatic populations tend to reduce their average number of daily social contacts in a similar manner in response to the spread of the disease, while the contact rate of the infected population is further reduced by an additional factor  $k \in [0, 1]$ . The parameter  $\xi(z)$  represents the percentage of asymptomatic individuals in the population and takes values in the interval  $[0, 1]$ . The parameter  $z$  represents the uncertainty and is distributed according to a given distribution  $p(z)$  which will be specified later on.

A further extension of the model in (5) consists of incorporating an age structure. In this case, the population is divided into subclasses corresponding to different age groups. As before, we consider four compartments: Susceptible (S), Infected (I), Asymptomatic (A), and Recovered (R). Each variable now depends not only on time and the uncertainty parameter  $z$ , but also on the age variable  $x \in \mathcal{A} = (0, 100)$  representing the age class. Incorporating an age structure introduces additional heterogeneity into the system, which is essential to accurately capture the dynamics of an epidemic. Individuals belonging to different age groups typically display distinct contact patterns and social behaviors, which results in varying transmission dynamics across the population. Under these assumptions, the model can be written as follows:

$$\begin{aligned}\frac{\partial S(x, t, z)}{\partial t} &= -\Lambda(x, t, z), \\ \frac{\partial I(x, t, z)}{\partial t} &= \xi(x, z)\Lambda(x, t, z) - \gamma_I(x, z)I(x, t, z), \\ \frac{\partial A(x, t, z)}{\partial t} &= (1 - \xi(x, z))\Lambda(x, t, z) - \gamma_A(x, z)A(x, t, z), \\ \frac{\partial R(x, t, z)}{\partial t} &= \gamma_I(x, z)I(x, t, z) + \gamma_A(x, z)A(x, t, z),\end{aligned}\tag{8}$$

being

$$\Lambda(t, z, x) = \beta(x, z)S(x, t, z)H_S(I(x, t, z)) \int_{\mathcal{A}} (H_I(I(y, t, z))I(y, t, z) + H_A(I(y, t, z))A(y, t, z))dy, \quad (9)$$

where  $\beta(x, z), \gamma_I(x, z), \gamma_A(x, z) > 0$  are the transmission and recovery rates, which are now also dependent on the age class  $x$ , and the macroscopic incidence rates  $H_J(\cdot)$  for  $J \in \{S, I, A\}$  are given by the following:

$$H_S(r) = \frac{\mu(x, z)}{\sqrt{1 + \nu(x, z)r}}, \quad H_A(r) = H_S(r), \quad H_I(r) = kH_S(r), \quad (10)$$

where the age-dependent parameters  $\mu(x, z), \nu(x, z)$  are supposed to be positive for  $k \in [0, 1]$ .

### 3. Parameters estimation

The goal of this section is to illustrate the calibration of the model parameters from available data. The procedure is described for the age-structured social SIAR model (8), but similar results can be obtained for the social SIAR model (5) without age structure by integrating over the age classes. We focus on the second wave of COVID-19 in Italy, particularly in the Lombardy region. Following the methodology outlined in [22], we first estimate the recovery rates  $\gamma_I$  and  $\gamma_A$ . These rates are derived from the Time of Viral Clearance (TVC), which measures the duration between the first positive test and recovery/death. As reported in [22], the recovery times follow a beta distribution, which allows the parameters to be defined as follows:

$$\frac{1}{\gamma_I(x, z_1, z_2)} = \begin{cases} h_{1,1} + h_{1,2}z_1, & \text{if } x \leq 50, \\ h_{2,1} + h_{2,2}z_2, & \text{if } x > 50, \end{cases} \quad \gamma_A(x, z_1, z_2) = 2\gamma_I(x, z_1, z_2), \quad (11)$$

being  $h_{1,1} = h_{2,1} = 5$ ,  $h_{1,2} = 32$ ,  $h_{2,2} = 40$ , and with  $z_1 \sim B(\alpha_1, \beta_1)$ ,  $z_2 \sim B(\alpha_2, \beta_2)$  beta distributions with parameters  $\alpha_1 = 2.1$ ,  $\beta_1 = 5.1$ , and  $\alpha_2 = 1.8$ ,  $\beta_2 = 3.9$ . Next, the transmission rate  $\beta(x, z)$  and the parameter  $\xi(x, z)$ , which represents the initial proportion of asymptomatic individuals, are estimated using data over the time interval  $[t_0, t_L]$ , being  $t_0$  as the initial time (e.g., October 8th, 2020) and  $t_L$  the final time (e.g., October 20th, 2020). In this period of time, there were no specific restrictions to the mobility and life on the individual. To count for uncertainty, we generate samples  $\{\gamma_I\}_{m=1}^M$  from  $\gamma_I$  and  $\{\gamma_A\}_{m=1}^M$  from  $\gamma_A$ . These samples are constructed using a collocation approach based on Gauss-Jacobi polynomials with  $M = 5$  nodes. For any  $m = 1, \dots, M$ , we solve the following optimization problem:

$$\min_{\beta_m(x), \xi_m(x)} \int_{\mathcal{A}} \mathcal{J}(I(y, t, z_m), \hat{I}(y, t), R(y, t, z_m), \hat{R}(y, t))dy, \quad \text{for } t \in [t_0, t_L], \quad (12)$$

being

$$\mathcal{J}(\cdot) = p\|I(y, t, z_m) - \hat{I}(y, t)\|_2 + (1 - p)\|R(y, t, z_m) - \hat{R}(y, t)\|_2, \quad (13)$$

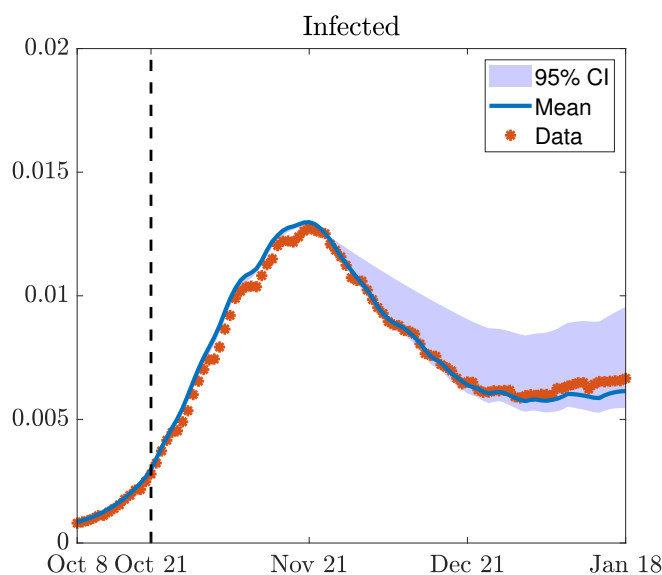
for  $p \in [0, 1]$ , where  $I(\cdot), R(\cdot)$  are the numerical solutions to the underline model assuming  $H_S = 1$ , meaning that no specific restrictions are set during this period of time, while  $\hat{I}(\cdot), \hat{R}(\cdot)$  are the available data.

Once the epidemic parameters have been estimated, the second optimization procedure consists of identifying the optimal functions  $H(x, t_j, z)$ , which mimic the implementation of governmental interventions such as mask mandates and mobility restrictions, and  $\xi(x, t_j, z)$ , which is the number of

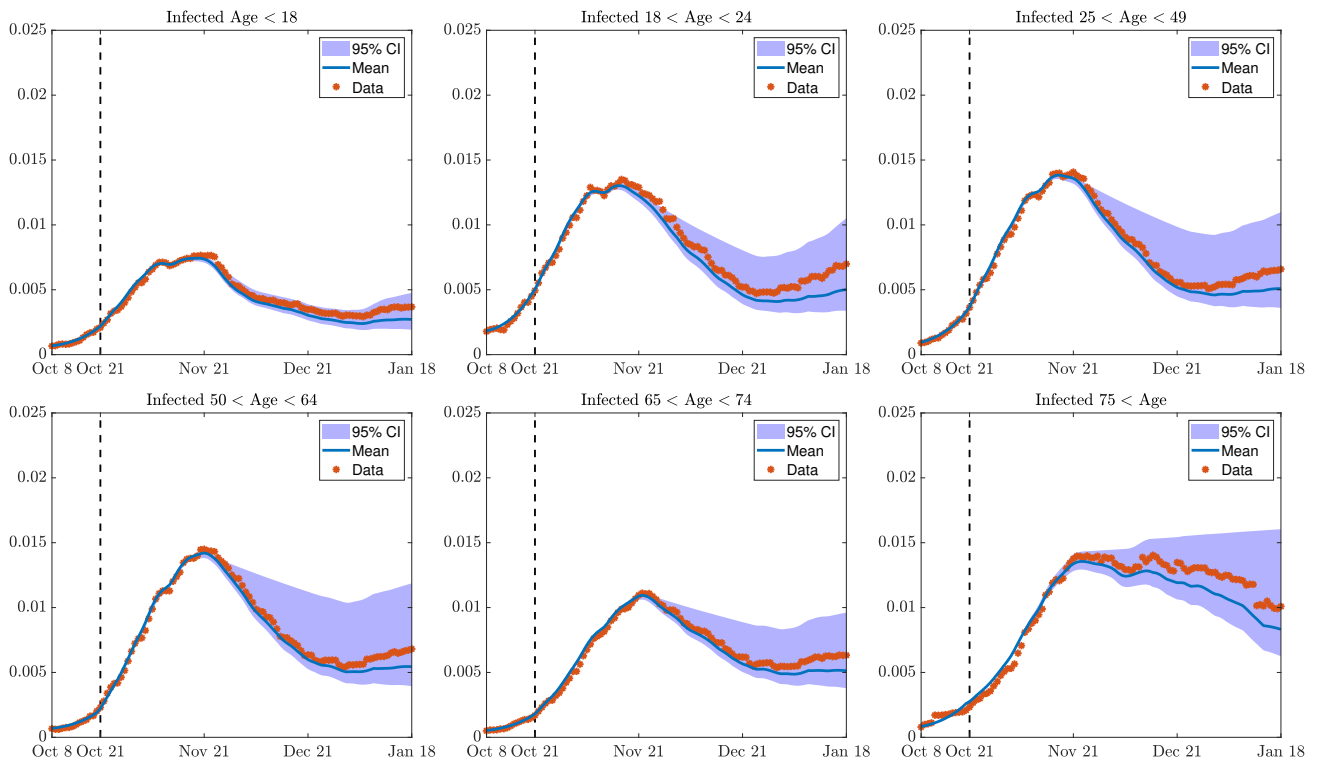
unknown asymptomatic for  $t_j \in [t_j - k_l, t_j + k_r]$  with  $j = 1, \dots, N_t$ , where  $N_t$  is the total number of considered time steps,  $k_l = 3$  and  $k_r = 4$ . This choice corresponds to averaging the epidemic data over a one-week window for each value of  $z$  representing uncertainty and for each age class  $x$ . The data used for this second optimization correspond to the subsequent phase of the second wave of the pandemic, specifically the interval  $[t_L, T]$ , with  $t_L$  marking the start date (October 21st, 2020) and  $T$  marking the end date (January 18th, 2021). As in the previous optimization, we sample the recovery rates from beta distributions and solve the following optimization problem for each  $m = 1, \dots, M$  and for  $t \in [t_j - k_l, t_j + k_r]$ :

$$\min_{H_m(x, t_j), \xi_m(x, t_j)} \int_{\mathcal{A}} \mathcal{J}(I(y, t, z_m), \hat{I}(y, t), R(y, t, z_m), \hat{R}(y, t)) dy, \quad (14)$$

where  $\mathcal{J}(\cdot)$  is defined as in (13). Both optimization problems (12)–(14) have been solved using the Matlab function `fmincon` combined with an RK4 integration method of the systems of ordinary differential equations (ODEs). Figure 1 shows the comparison of the above described optimization strategy for the social SIAR model (5) in terms of the number of infected individuals. In the figure, the mean trajectory and the 95% confidence interval, together with the available reported data on infected cases, are shown. The black dashed line distinguishes between the first and second phases of the pandemic, which correspond to the two optimization procedures. Figure 2 shows the corresponding results for the age-structured model (8). As before, the mean trajectory and the 95% confidence interval are displayed, together with the reported data on infected cases. The black dashed line separates the first and second phases of the pandemic. Each image corresponds to a different age group.



**Figure 1.** Dynamics of the infected population obtained by solving the calibrated social-SIAR model (5) and compared with experimental data. The figure shows the mean epidemic trajectory with the 95% confidence interval (shaded area), alongside the observed data. The black dashed line separates the two epidemic phases.



**Figure 2.** Dynamics of the infected population obtained by solving the social-SIAR model (8) and compared with experimental data. The plots show the mean epidemic trajectory with the 95% confidence interval (shaded area), alongside the observed data. The black dashed line separates the two epidemic phases. Each image corresponds to a different age group.

#### 4. Predictions of COVID-19 dynamics using neural networks

Building upon the methodology introduced in our previous work [40], we aim to train neural networks on augmented datasets to enhance the accuracy of their solution both in terms of interpolations and predictions. In our earlier study, we considered deterministic synthetic data that arose from the solution of a simplified epidemic model, namely the social-SIR model (4). Since the observational data were recorded on a daily basis (with time step  $h = 1$ ), we generated additional data over the same interval by solving the system of ODEs with a finer resolution ( $h = 0.2$ ). In this work, we extend this strategy to more complex models (5)–(8), with the important distinction that we now incorporate uncertainty in the model parameters, as outlined in the previous sections. Our goal is to train neural networks that are well-suited for predictive tasks, such as PINNs, and NAR networks. Both PINNs and NAR networks are implemented using a Feed-Forward neural network architecture with  $L + 1$  layers as follows:

$$\begin{aligned} x^1 &= W^1 x + b^1, & x^l &= \sigma \circ (W^l x^{l-1} + b^l), & l &= 2, \dots, L-1, \\ f^{NN}(x; \mathbf{b}, \mathbf{W}) &= W^L z^{L-1} + b^L, \end{aligned} \quad (15)$$



where  $x = (x^1, \dots, x^d) \in \mathbb{R}^d$  is the input,  $\sigma$  is an activation function to convert the input signals to output signals, and  $\mathbf{W} = (W^1, \dots, W^L)$  and  $\mathbf{b} = (b^1, \dots, b^L)$  represent the weights and biases, respectively. The weights and biases associated with these connections serve as the parameters of the network, which are iteratively adjusted through techniques such as gradient descent or the Adam method during the training phase to minimize prediction errors and enhance the model's performance. To find the optimal parameters  $\theta^* = (W^1, b^1, \dots, W^L, b^L)$ , we solve the following minimization problem:

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\theta), \quad (16)$$

being

$$\mathcal{L}(\theta) = \omega_d \mathcal{L}_d(\theta) + \omega_p \mathcal{L}_p(\theta), \quad (17)$$

the loss function, for  $\omega_d \geq 0$ ,  $\omega_p \geq 0$ . Specifically, in (17), the function  $\mathcal{L}_d(\theta)$  measures the discrepancy between the neural network solution and the data, while  $\mathcal{L}_p(\theta)$  encodes the physics of the problem. In the following, we present PINNs and NAR networks in the context of the age-structured model (8). A similar formulation applies to the simpler model (5), where integration is performed over all age classes  $x \in \mathcal{A}$ , with  $\mathcal{A}$  denoting the age-classes.

#### 4.1. Physics informed neural networks

PINNs offer a physics-driven learning framework where the network is trained not only to fit the data but also to satisfy the underlying differential equations governing the system dynamics. This makes them particularly appealing in scenarios where data are scarce or noisy, as the physical constraints serve as an effective regularizer. In our context, we define the PINN using the epidemic model equations, thereby incorporating uncertain parameters  $z$  sampled as described earlier. The architecture consists of the following:

- **Input Layer:** Takes both the age-time variables  $x, t$  and the uncertainty parameter  $z$  as the input.
- **Hidden Layers:** Several fully connected layers with non-linear activation functions (e.g., tanh) are used to approximate the solution.
- **Output Layer:** Returns the approximated solution values  $f_i^{NN}(x, t, z)$  for  $i \in \mathcal{I} = \{S, I, A, R\}$  for any  $x, t$  and  $z$ .

To determine the optimal parameters, we solve the minimization problem in (16), thereby defining the loss function as in (17) with  $\omega_d, \omega_p > 0$ . Specifically, the loss function that encodes the data constraints reads as follows:

$$\mathcal{L}_d(\theta) = \sum_{i \in \mathcal{I}} \sum_{n=1}^{N_c} \int_{\mathcal{A}} \left( \bar{f}_i^{NN}(x, t_n; \theta) - \bar{f}_i(x, t_n) \right)^2 dx, \quad (18)$$

being

$$\bar{f}_i^{NN}(x, t_n; \theta) = \sum_{m=1}^M f_i^{NN}(x, t_n, z_m; \theta) w_m, \quad \bar{f}_i(x, t_n; \theta) = \sum_{m=1}^M \hat{f}_i(x, t_n, z_m; \theta) w_m, \quad (19)$$

where  $w_m$  are the weights associated to the uncertainty values  $z_m$ , and  $N_c, M$  are the total number of samples that we consider in time and in the uncertainty space, respectively, while  $\hat{f}_i$  are the data that represent the number of Susceptible, Infected, Asymptomatic, and Recovered individuals.

Then, we compute the derivative via automatic differentiation with respect to time of the quantities  $f_i^{NN}(x, t, z; \theta)$  for any  $i \in \mathcal{I}$ , and we define the physical loss as follows:

$$\mathcal{L}_p(\theta) = \sum_{i \in \mathcal{I}} \sum_{n=1}^{N_c} \int_{\mathcal{A}} \bar{\mathcal{R}}_i^2(x, t_n; \theta) dx, \quad (20)$$

being

$$\bar{\mathcal{R}}_i(x, t_n; \theta) = \sum_{m=1}^M \mathcal{R}_i(x, t_n, z_m; \theta) w_m, \quad (21)$$

with  $w_m$  weights associated with the Gauss-Jacobi nodes, and

$$\begin{aligned} \mathcal{R}_S &= \partial_t f_S^{NN}(x, t_n, z_m; \theta) + \Lambda(x, t_n, z_m; \theta), \\ \mathcal{R}_I &= \partial_t f_I^{NN}(x, t_n, z_m; \theta) - \xi(x, t_n, z_m) \Lambda(x, t_n, z_m; \theta) + \gamma_I(x, z_m) f_I^{NN}(x, t_n, z_m; \theta), \\ \mathcal{R}_A &= \partial_t f_A^{NN}(x, t_n, z_m; \theta) - (1 - \xi(x, t_n, z_m)) \Lambda(x, t_n, z_m; \theta) + \gamma_A(x, z_m) f_A^{NN}(x, t_n, z_m; \theta), \\ \mathcal{R}_R &= \partial_t f_R^{NN}(x, t_n, z_m; \theta) - \gamma_I(x, z_m) f_I^{NN}(x, t_n, z_m; \theta) - \gamma_A(x, z_m) f_A^{NN}(x, t_n, z_m; \theta), \end{aligned} \quad (22)$$

where

$$\Lambda = \beta(x, z_m) f_S^{NN}(x, t_n, z_m; \theta) H(x, t_n, z_m) \int_{\mathcal{A}} H(y, t_n, z_m) (k f_I^{NN}(y, t_n, z_m; \theta) + f_A^{NN}(y, t_n, z_m; \theta)) dy, \quad (23)$$

for  $k \in [0, 1]$ .

#### 4.2. Non-linear autoregressive networks

NAR networks leverage past time series values to forecast future states, making them particularly effective for time-dependent predictions, especially in short term forecasting. The NAR architecture includes the following components:

- **Input Layer:** Composed of infected population values at previous time steps,  $I(x, t - d, z), \dots, I(x, t - 1, z)$ , for a chosen delay  $d \geq 1$  and for any value of the uncertain variable  $z$  and of the age variable  $x$ .
- **Hidden Layers:** A set of fully connected layers equipped with nonlinear activation functions (e.g., ReLU or tanh) to model complex temporal dependencies.
- **Output Layer:** Produces the predicted value  $I(x, t, z)$ , which represents the number of infected at time  $t$  for each realization of  $z$  and of the age variable  $x$ .

Once trained, the NAR network can recursively predict the epidemic evolution by feeding its own predictions as inputs for future time steps (closed-loop strategy). The NAR approach is particularly effective for qualitative short-term forecasting, as it can directly learn epidemic dynamics from data without relying on the underlying model equations. This data-driven nature allows the NAR network to outperform PINNs in the short-term setting, both in terms of accuracy and computational efficiency. However, when it comes to long-term forecasting, PINNs provide a more reliable framework, as their physics-informed structure enables them to incorporate the underlying epidemic dynamics and capture long-term trends that the NAR network can not.

The training process involves the minimization of the loss function (17), where we suppose  $\omega_d > 0$ ,  $\omega_p = 0$ , and

$$\mathcal{L}_d(\theta) = \sum_{n=1}^{N_c} \sum_{m=1}^M \int_{\mathcal{A}} \left( I^{NN}(x, t_n, z_m; \theta) - \hat{I}(x, t_n, z_m) \right)^2 dx, \quad (24)$$

being  $I^{NN}$  the solution computed by the neural network and  $\hat{I}$  the data, with  $N_c$  denoting the number of samples in time, and  $M$  is the total number of samples in the uncertainty space.

**Remark 1.** In the case of PINNs, it is essential to include constraints for all population compartments in the loss function to accurately capture the dynamics prescribed by the social SIR model, even if the primary objective is to only reproduce the infected population. In contrast, for the NAR network, the dynamics are not explicitly enforced in the loss function, but the calibrated model is instead used to produce an augmented data set used to feed the network.

## 5. Numerical experiments

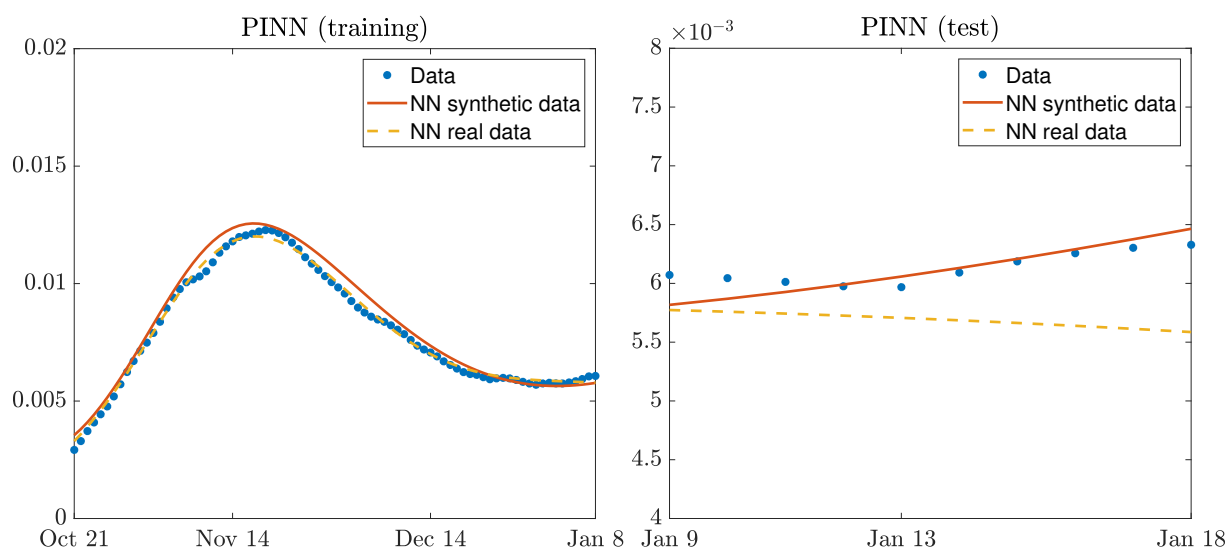
Now, we proceed with numerical experiments to validate our methodology. In particular, we aim to show that neural networks trained on larger synthetic datasets achieve a higher accuracy in both prediction and extrapolation tasks. In addition, the goal is to demonstrate that NAR networks provide a viable alternative to PINNs in the context of short term predictions, especially when trained on synthetic data. The synthetic datasets are generated by solving systems (5)–(8) over the time interval  $[t_0, T]$ , with  $t_0 = 15$  and  $T = 105$ , using a time step of  $h = 0.2$ . This time window corresponds to the second phase of the pandemic in Italy, specifically from October 21st, 2020 to January 18th, 2021. We empathize that data are generated over the entire time horizon to guarantee sufficient samples for both short- and long-term forecasting. However, the neural networks are exclusively trained on data within the specified training window, without any use of information from the test period. Model parameters are treated as uncertain, as previously described, and uncertainty is represented by sampling  $M = 5$  Gauss-Jacobi nodes  $z_m$ . The macroscopic incidence rates  $H_J(\cdot)$  for  $J \in \{S, I, A\}$  and the number of asymptomatic  $\xi(\cdot)$  are reconstructed from the discrete dataset, and computed by solving the minimization problem (14), is selected the value corresponding to the time index immediately preceding the desired time  $t$  is selected. For the sake of comparison, we also train the same neural networks only using the real data at disposal. We will mainly focus on the short term forecast, thereby splitting the datasets into training and test sets: the training set covers the period from October 21st to January 8th, while the test set spans the period from January 9th to January 18th. At the end of the section, we compare the performances of the PINNs and NARs networks by focusing on the short and long term forecasting. For the long term forecasting, we split the dataset into a training set (from October 21st to November 19th) and a test set (from November 20th to January 3rd), to capture the peak of the pandemic.

### 5.1. Physics informed neural network

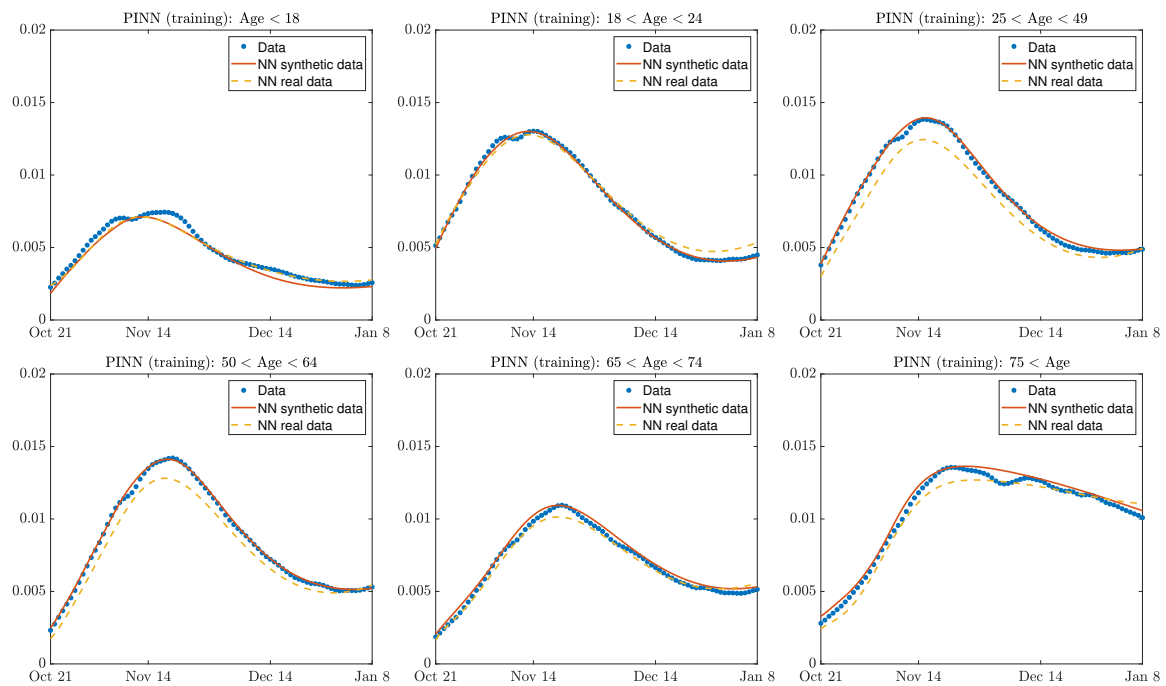
We start by defining two Feed-Forward network architectures with width 32 and 3 hidden layers, using  $\tanh$  as activation function. The networks are designed to approximate the solutions of the social SIAR model and the age-structured social SIAR model by means of PINNs. Both networks are trained on real and synthetic data, assuming  $\omega_d = \omega_p = 1$  in the loss function (17) to account both for the

data and the underline physics. For real data, in the case of the non-age structured model, the trained network takes the time  $t$  as the input and produces the corresponding value  $f_i(t)$  for any compartment  $i \in \{S, I, A, R\}$  at time  $t$  as the output, for any  $t$  corresponding to real data. In the case of synthetic data, the input consists of both the time  $t$  and the uncertainty variable  $z$ , and the output is the value  $f_i(t, z)$  for each compartment  $i \in \{S, I, A, R\}$ , for any  $t$  corresponding to synthetic data. In the age-structured setting, the network input is the triplet  $(x, t, z)$  and the output is  $f_i(x, t, z)$ , where  $x \in \mathcal{A}$ , with  $\mathcal{A}$  denoting the different age classes. The networks are trained for 50,000 epochs using the Adam optimizer with a learning rate  $10^{-2}$ . Following previously described procedure, we ensure that both the data and the physics of the system are learned by the network.

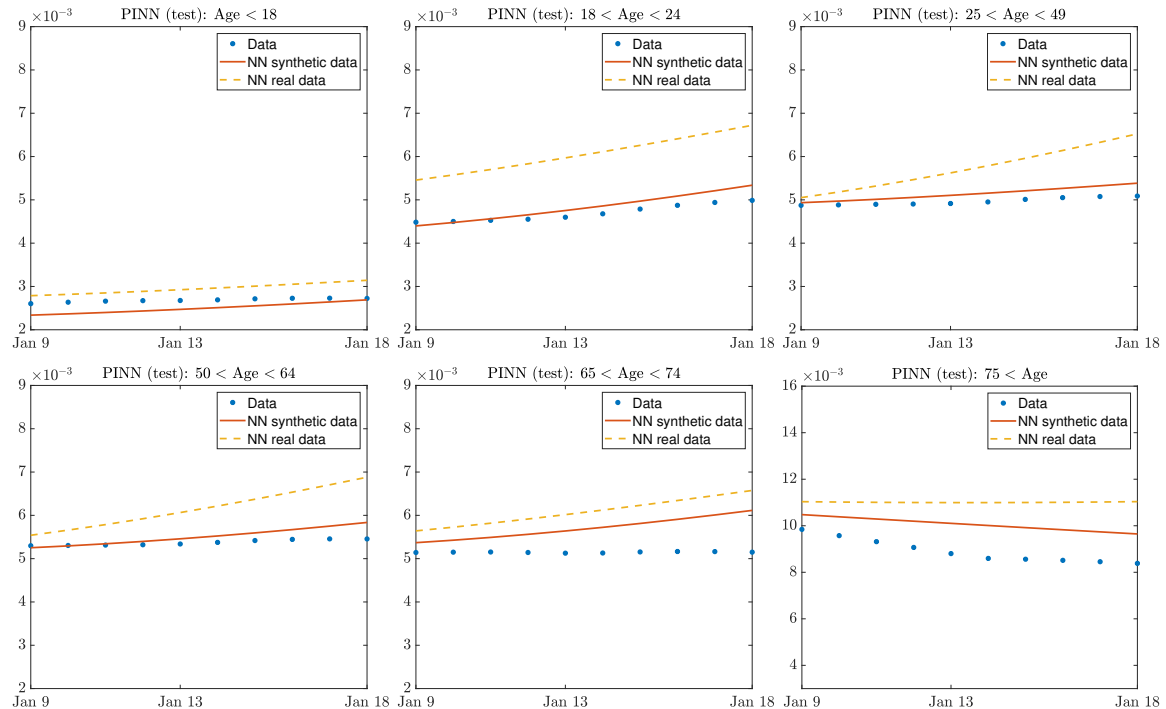
In Figure 3, a comparison between the PINN solutions and the available data in the case of the social SIAR model (5) is reported. On the left, the image shows the part relative to the training set, while the image on the right refers to the test set. We clearly see that the network trained on real data achieves a better fit on the training set, whereas the network trained on synthetic data produces a more accurate solution on the test set; thus, it is able to better forecast the time evolution of the epidemic. Figure 4 shows the comparison between the PINN solutions and the available training data in the case of the age-structured social SIAR model (8) for the training part. Figure 5 shows the same results but computed on the test set. In most of the cases, despite of the complexity of the solution, the networks achieve a good approximation in interpolation; however, they still face challenges in prediction accuracy, even when synthetic data are used to enhance their forecasting capabilities.



**Figure 3.** Physics informed neural network for the social SIAR model (5). Solutions obtained by training a PINN network on both real and synthetic data, compared to the available data. On the left, the solution computed on the training set. On the right, the solution computed on the test set.



**Figure 4.** Physics informed neural network (training set) for the age-structured social SIAR model (8). Solutions obtained by training a PINN network on both real and synthetic data, compared to the available data. Each plot corresponds to a different age class.



**Figure 5.** Physics informed neural network (test set) for the age-structured social SIAR model (8). Solutions obtained by training a PINN network on both real and synthetic data, compared to the available data. Each plot corresponds to a different age class.

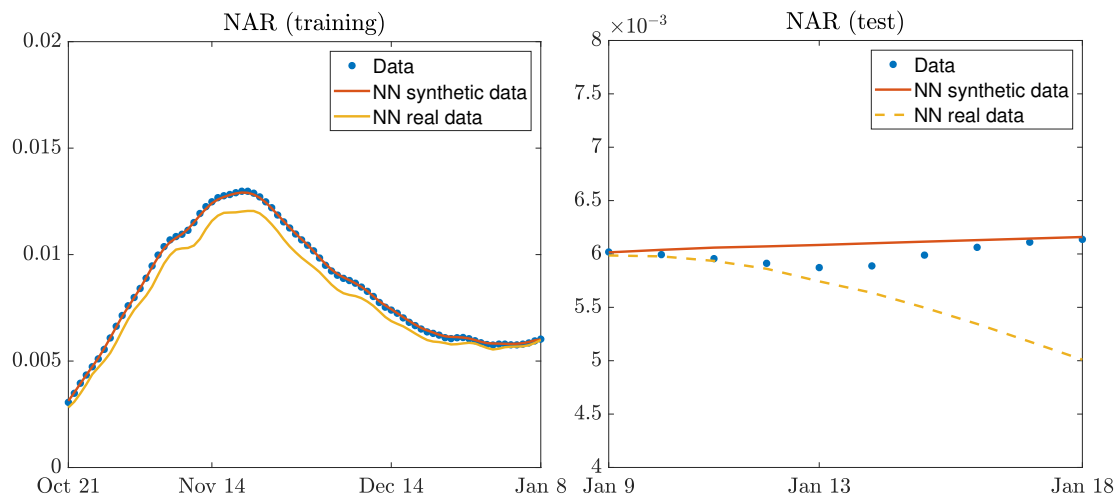
## 5.2. Non-linear autoregressive network

Now, we consider the NAR networks. We define two additional Feed-Forward network architectures with a width of 32 and 3 hidden layers, using ReLu as activation function. One network is designed to approximate the solution of the social SIAR model (5) and the other for the age-structured SIAR model (8). As for the PINNs, both networks are trained on real and synthetic datasets, assuming  $\omega_d = 1$  and  $\omega_p = 0$  in the loss function (17). In the case of real data, and for the non-age structured model, the network takes the number of infected individuals at times  $t - d, \dots, t - 1$  within the training set as the input and predicts the number of infected at time  $t$ , for any  $t$  corresponding to real data. In the case of synthetic data, the input includes the number of infected at time  $t - d, \dots, t - 1$  for any value of  $z$ , and the output is the number of infected for any value of  $z$  at time  $t$ , for any  $t$  corresponding to synthetic data. To select the lag parameter  $d$ , we performed a series of preliminary tests on the simpler model in (5), thereby evaluating the accuracy of the predicted solutions. Specifically, we set  $d$  to vary in the range  $d = 3, \dots, 10$ , and we measure the Root Square Mean Errors (RMSE) between the available data and the neural network predictions, defined as the following:

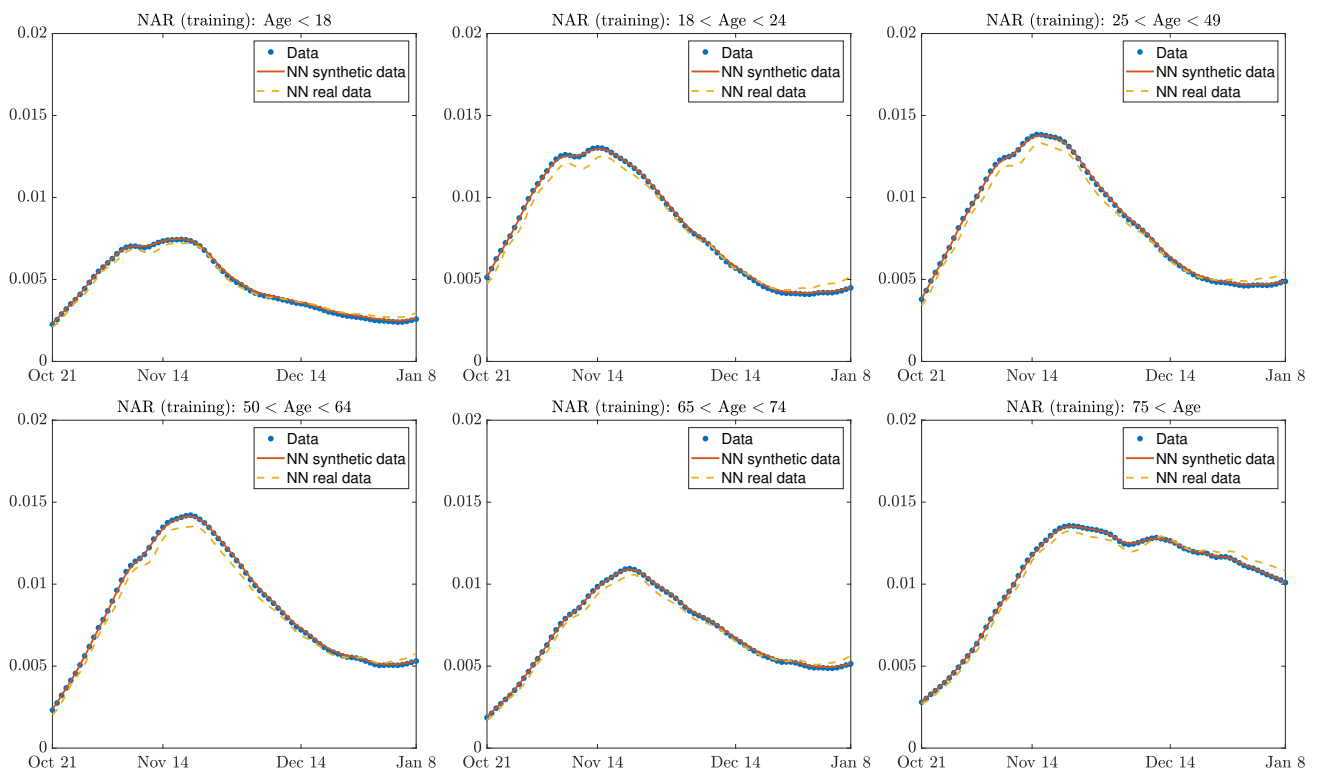
$$RMSE(t) = \sqrt{\sum_{m=1}^M (\hat{I}(t, z_m) - I_i^{\text{NN}}(t, z_m; \theta_*))^2 w_m}, \quad (25)$$

where  $\hat{I}(t, z_m)$  denotes the observed data and  $I_i^{\text{NN}}(t, z_m; \theta_*)$  is the output of the neural network trained on real or synthetic data, and  $w_m$  are the nodes associated to the uncertain parameters  $z_m$ . The results show that, for both synthetic and real data, changing the value of  $d$  does not lead to a significant improvement in the accuracy of the neural network predictions, with mean RMSE being of the order of  $10^{-4}$  for synthetic data and of the order of  $10^{-3}$  for real data. Therefore, in all subsequent experiments, we set  $d = 5$ . For the age-structured model, both the input and output include the infected number for all age classes. The networks are trained for 20,000 epochs using the Adam optimizer with a learning rate of  $10^{-2}$ . To assess their performance, we first compute the solutions on the training set. Then, using a closed-loop strategy, where the network predictions are recursively used as inputs—we compute the solutions on the test set.

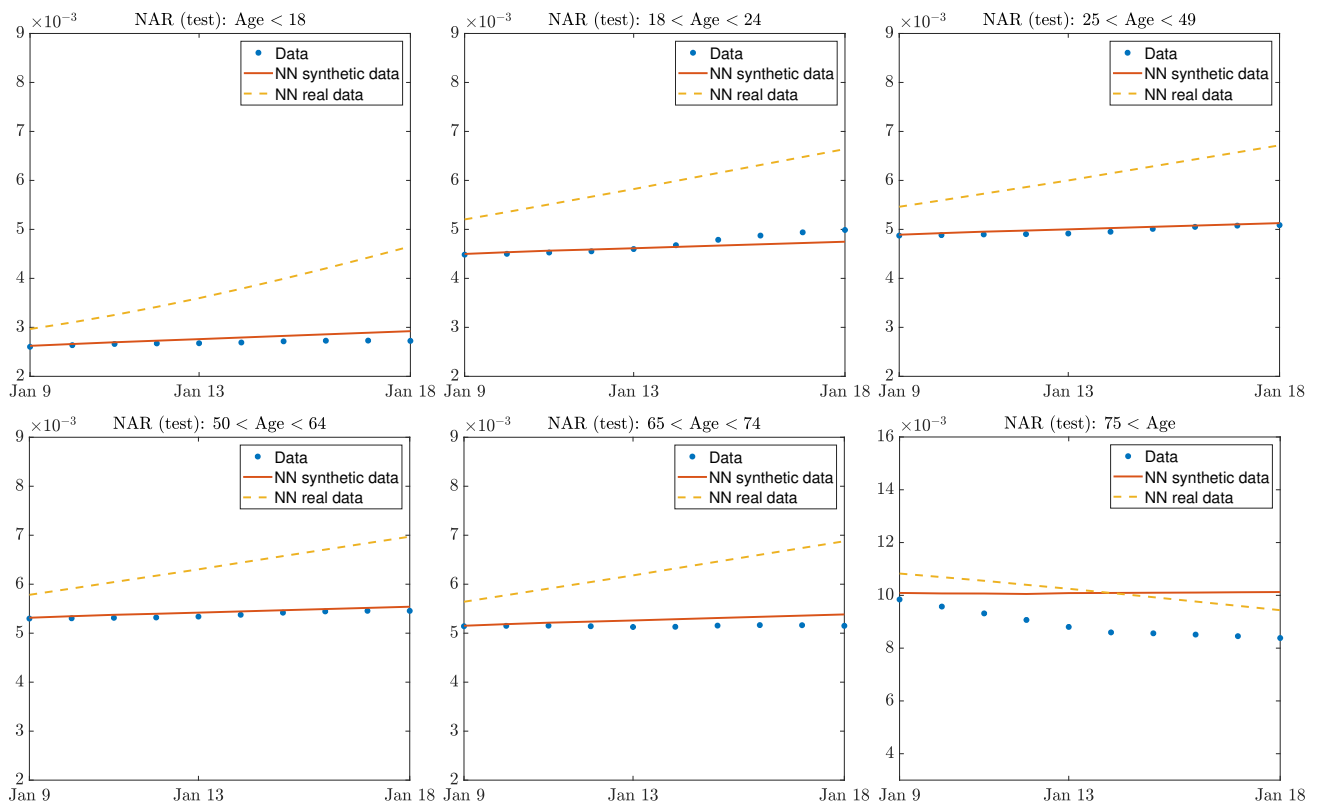
Figure 6 shows the comparison between the NAR network predictions and the available data in the case of the social SIAR model (5). For synthetic data, we plot the mean solution with respect to the uncertainty, computed as in (19). The image on the left corresponds to the training set, while the results on the test set are shown on the right. The network trained on synthetic data yields more accurate solutions, demonstrating qualitative improved performances in both interpolation and prediction with respect to the PINNs. Figure 7 presents the comparison between the neural network predictions and the available data for the age-structured social SIAR model (8) computed on the training set, while Figure 8 shows the corresponding results on the test set. Each plot corresponds to a different age class. In most of the cases, the network trained on synthetic data produces qualitatively more accurate results with respect to the one trained on real data and to PINNs, both in terms of interpolation and prediction.



**Figure 6.** Nonlinear autoregressive network for the social SIAR model (5). Solutions obtained by training a NAR network on both real and synthetic data, compared to the available data. On the left, the solution computed on the training set. On the right, the solution computed on the test set.



**Figure 7.** Nonlinear autoregressive network (training set) for the age-structured social SIAR model (8). Solutions obtained by training a NAR network on both real and synthetic data, compared to the available data. Each image corresponds to a different age class.



**Figure 8.** Nonlinear autoregressive network (test set) for the age-structured social SIAR model (8). Solutions obtained by training a NAR network on both real and synthetic data, compared to the available data. Each image corresponds to a different age class.

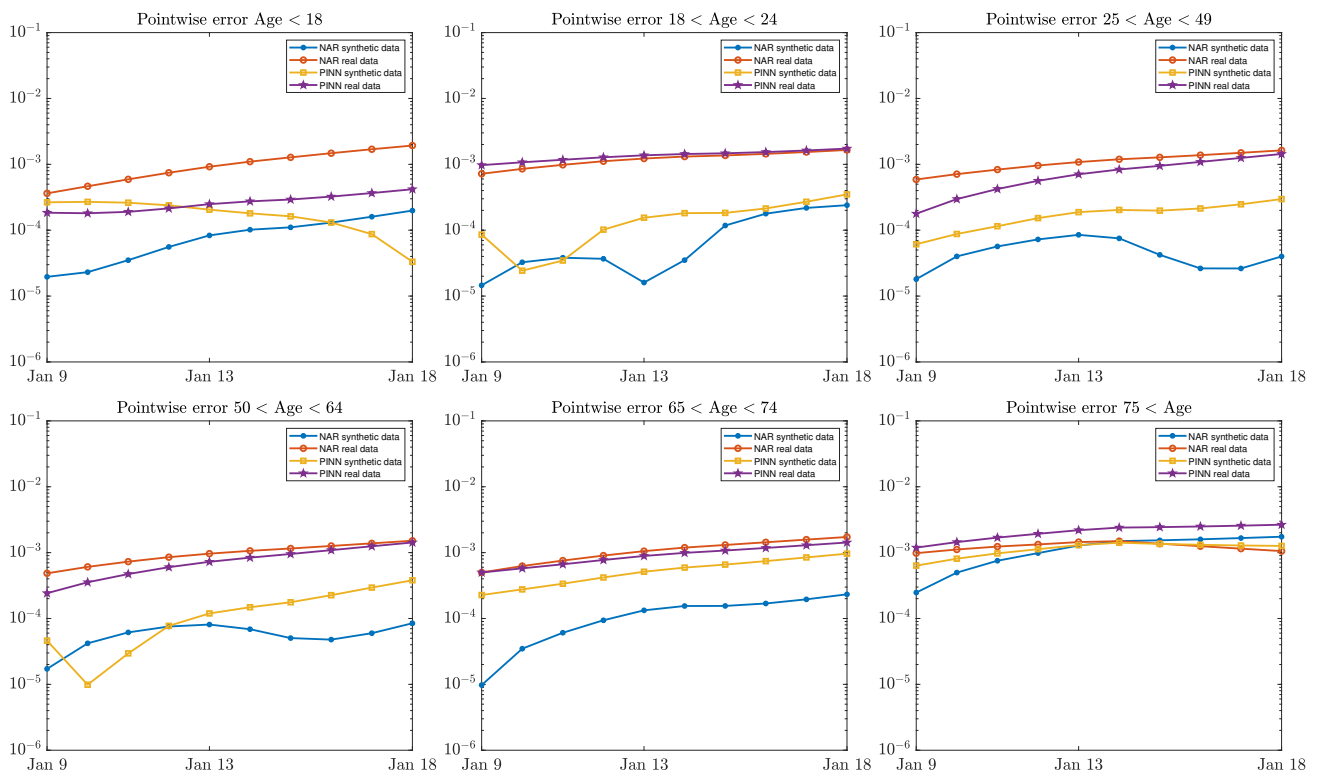
### 5.3. Comparison between PINNs and NARs

Now, we compare the performance of the NAR network and the PINN. We focus on both short term forecasting. First, we compare the trained networks in terms of the computational cost. For comparison purposes, we assume to train the PINNs and NAR networks, both on synthetic and real data, over 50,000 epochs. Training a PINN involves significantly higher computational costs with respect to training NAR networks, particularly in the case of the age-structured model. Indeed, automatic differentiation is needed to enforce the physical constraints imposed by the system of ODEs. The results are confirmed by Table 1, which reports the training time, in seconds, required to complete 50,000 epochs referred to both models (5)–(8).

**Table 1.** Training time in seconds required to complete 50,000 epochs in the cases of NAR and PINN trained both on synthetic and real data. First row: social SIAR model (5). Second row: age-structured social SIAR model (8).

	NAR (synthetic)	NAR (real)	PINN (synthetic)	PINN (real)
Non-aged model	32 s	26 s	307 s	205 s
Age-structured model	51 s	34 s	468 s	234 s





**Figure 9.** Comparison between NAR and PINN in terms of accuracy over the test set referred to the age-structured social SIAR model (8). Short term forecasting. The error between available data and neural network solutions in time is computed as in (26). Markers have been added just to denote different lines.

**Short term forecasting.** In short-term forecasting, NAR networks offer a competitive alternative to physics-informed approaches, particularly when used in combination with data augmentation strategies. By directly learning the epidemic dynamics from available data, NAR networks provide accurate quantitative estimates that can be highly valuable to monitor the progression of the pandemic and support decision-making aimed at mitigating its impact. Additionally, their data-driven nature allows for faster training and computational efficiency, making them particularly suitable for timely analyses in rapidly evolving scenarios. The results are confirmed by Figure 9, which displays the error between the real data and the neural network solutions computed as follows:

$$\mathcal{E}(x, t) = |\hat{I}(x, t) - I_i^{\text{NN}}(x, t; \theta_*)|, \quad (26)$$

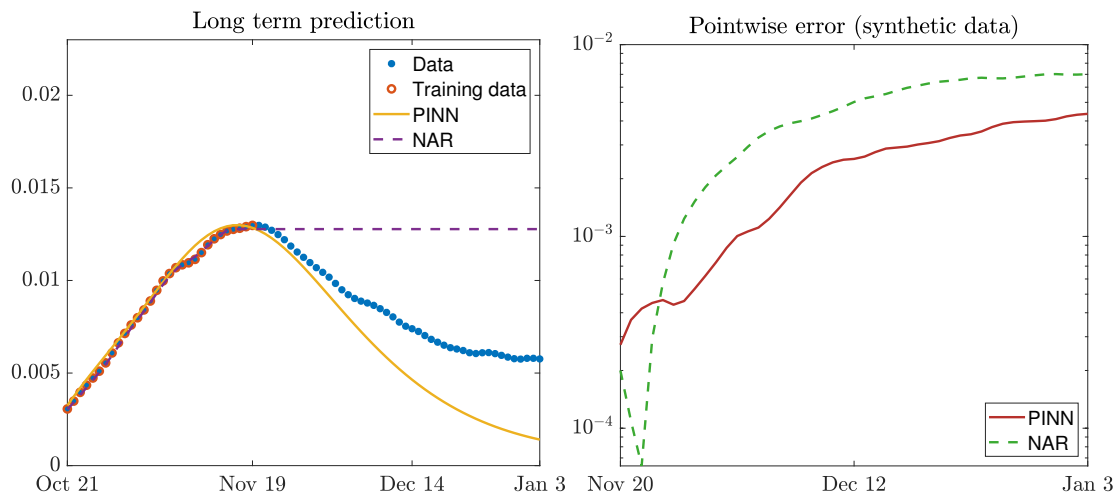
where  $\hat{I}(x, t)$  denotes the observed data, and  $I_i^{\text{NN}}(x, t; \theta_*)$  is the output of the neural network trained on real or synthetic data. In the case of synthetic data, the network output corresponds to the mean solution with respect to the uncertainty, as in (19). Training either a PINN or a NAR network on synthetic data provides clear benefits in terms of accuracy. In particular, NAR networks trained on synthetic data proved to be effective in nearly all cases. We present the results only for the age-structured model (8); however, similar findings hold for the simpler model (5) as well. These results are further supported by Table 2, which reports the maximum error (26) over time between real data and the different neural

network solutions. In the case of model (5), we integrate over  $x \in \mathcal{A}$ , where  $\mathcal{A}$  represents the age-classes.

**Table 2.** Accuracy of the neural network solutions w.r.t. the available data computed as the maximum of  $\mathcal{E}(t)$  over time. Short term forecasting. First row: social SIAR model (5). Second-Seventh rows: age-structured social SIAR model (8).

	NAR (synthetic)	NAR (real)	PINN (synthetic)	PINN (real)
Non-aged model	$2.1 \times 10^{-4}$	$1.2 \times 10^{-3}$	$10^{-3}$	$1.2 \times 10^{-3}$
Age < 18	$2 \times 10^{-4}$	$1.9 \times 10^{-3}$	$3 \times 10^{-4}$	$4 \times 10^{-4}$
19 < Age < 24	$2 \times 10^{-4}$	$1.7 \times 10^{-3}$	$3 \times 10^{-4}$	$1.7 \times 10^{-3}$
25 < Age < 49	$10^{-4}$	$1.6 \times 10^{-3}$	$3 \times 10^{-4}$	$1.4 \times 10^{-3}$
50 < Age < 64	$10^{-4}$	$1.5 \times 10^{-3}$	$4 \times 10^{-4}$	$1.4 \times 10^{-3}$
65 < Age < 74	$2 \times 10^{-4}$	$1.7 \times 10^{-3}$	$10^{-3}$	$1.4 \times 10^{-3}$
75 < Age	$1.7 \times 10^{-3}$	$1.5 \times 10^{-3}$	$1.4 \times 10^{-3}$	$2.7 \times 10^{-3}$

Finally, let us mention that NAR networks achieve, in general, good accuracy after 20,000 training epochs, whereas PINNs require at least 50,000 epochs to produce good results. This is valid for both the social SIAR model and the one with age-structure.



**Figure 10.** Long term forecasting for the social SIAR model (5). On the left, the solution obtained by training a PINN and a NAR network on synthetic data, compared to the available data. On the right, the pointwise error over the test set computed as in (26).

**Long term forecasting.** In long-term forecasting, PINNs prove to be more accurate than NARs, by offering a qualitative understanding of the epidemic dynamics. While their quantitative accuracy may be lower than that of NAR networks in the short term, PINNs are particularly useful to explore strategies to mitigate epidemic peaks and understand broader trends in disease progression. By embedding the underlying physical and epidemiological laws into the learning process, they can capture the overall behavior of the system over extended periods, which helps to inform public health interventions and long-term planning. Since the main goal of this test is to illustrate the capability of PINNs to repro-

duce the epidemic peak, we restrict our analysis to the simpler model in (5). Capturing the additional complexity introduced by age-structured dynamics would likely require a significantly greater training effort from the network. Furthermore, as we have already shown that the augmented data strategy improves the accuracy of neural network solutions, we exclusively train both PINNs and NARs on augmented datasets, without reporting results obtained from training on real data.

Figure 10 (left) compares the NAR and PINN networks trained on synthetic data for long-term forecasting. The PINNs achieve a higher accuracy than the NARs, thus successfully reproducing the epidemic peak and capturing the qualitative behavior of the solution. These findings are further supported by the error plot, computed as in (26) on the test set, as shown in the right panel of Figure 10.

All the numerical experiments were run on a Desktop Computer equipped with an Intel(R) Core(TM) i7-8700 CPU processor and 32 GB RAM.

#### 5.4. Effectiveness of the data augmentation strategy

To conclude our study and assess the effectiveness of the proposed data augmentation strategy, we perform an additional test that focuses on short-term forecasting, referring for simplicity to the model introduced in (5). We assume the recovery parameters to follow beta distributions defined as follows:

$$\gamma_I^1 = 0.04 + 0.05z_1, \quad \gamma_I^2 = 0.04 + 0.05z_2, \quad (27)$$

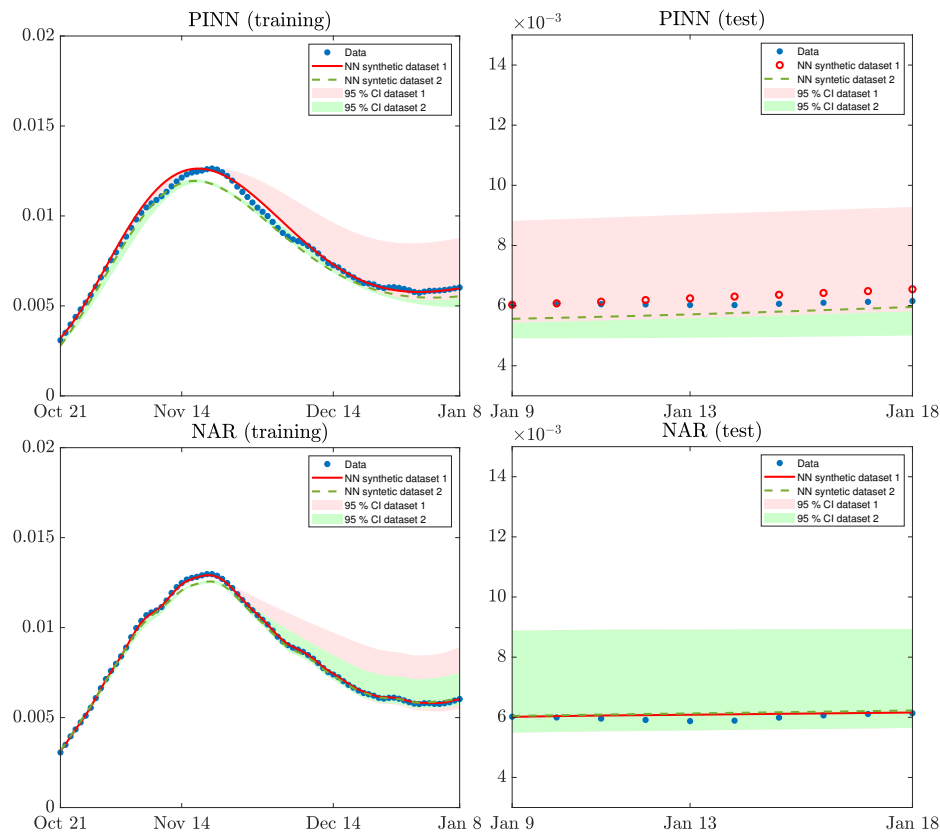
where  $z_1 \sim \text{Beta}(1.95, 4.95)$  and  $z_2 \sim \text{Beta}(2, 6)$ . Following the procedure described in Section 3, we set  $\gamma_A^i = 2\gamma_I^i$  for  $i = 1, 2$  and determine the transmission parameter  $\beta$ , the unknown number of asymptomatic  $\xi(\cdot)$ , and the macroscopic incidence rates  $H_J(\cdot)$  for  $J \in \{S, I, A\}$  by solving the corresponding minimization problems. As in the previous experiments, we generate synthetic data by solving the ODE system in (5) and evaluate the performance of both the PINN and NAR networks on this dataset. Specifically, we select the period from October 21st to January 8th as the training set, and the interval from January 9th to 18th as the test set. Figure 11 illustrates the solutions computed by the two networks trained on the respective datasets for the different uncertain parameters. Specifically, dataset one and two correspond to  $z_1 \sim \text{Beta}(1.95, 4.95)$  and  $z_2 \sim \text{Beta}(2, 6)$ , respectively. The plots display the mean predicted solutions, with the confidence intervals represented by the shaded area. The left panel corresponds to the training phase, while the right panel to the testing phase. The first row plots the results for the NAR network, and the second row for the PINN. In terms of mean the solution, the results are comparable.

Finally, we compute the accuracy of the solution on the test set according to different error metrics. Specifically, we consider the RMSE computed as in (25), as well as the Mean Absolute Error (MAE) and the coefficient of determination ( $R^2$ ) as follows:

$$\begin{aligned} MAE(t) &= \frac{1}{N_c} \sum_{m=1}^M |\hat{I}(t, z_m) - I_i^{\text{NN}}(t, z_m; \theta_*)| w_m, \\ R^2(t) &= 1 - \frac{\sum_{m=1}^M (\hat{I}(t, z_m) - I_i^{\text{NN}}(t, z_m; \theta_*))^2 w_m}{M\hat{\sigma}^2}, \end{aligned} \quad (28)$$

where  $\hat{I}(t, z_m)$  denotes the observed data,  $I_i^{\text{NN}}(t, z_m; \theta_*)$  is the output of the neural network trained on real or synthetic data,  $N_c$  is the number of considered time points,  $\hat{\sigma}^2$  is the variance of the observed

data with respect to the uncertainty parameter  $z_m$ , and  $w_m$  are the nodes associated with  $z_m$  for  $m = 1, \dots, M$ . Table 3 reports the mean value with respect to the time  $t$  of the quantities computed in (25)–(28). The coefficient of determination  $R^2$  indicates that the model provides an excellent fit to the data, a conclusion which is further supported by the other error metrics. Moreover, the neural networks demonstrate robustness, as the errors remain comparable across different uncertainty levels. Once again, the results confirm that the NAR network yields more accurate approximations than the PINN for short-term predictions.



**Figure 11.** NAR network and PINN for the social SIAR model (5). Comparison between two different datasets of uncertain parameters computed as in (11) and following the procedure described in Section 3. First row: PINN. Second row: NAR networks. On the left, the solution computed on the training set. On the right, the solution computed on the test set. Dataset one and two correspond to  $z_1 \sim \text{Beta}(1.95, 4.95)$  and  $z_2 \sim \text{Beta}(2, 6)$ , respectively.

**Table 3.** Accuracy of the neural network solutions to the social SIAR model (5) w.r.t. the available data computed as in (25)–(28). Short term forecasting.

	NAR (dataset 1)	PINN (dataset 1)	NAR (dataset 2)	PINN (dataset 2)
<i>RMSE</i>	$2 \times 10^{-4}$	$6.2 \times 10^{-4}$	$10^{-4}$	$6 \times 10^{-4}$
<i>MAE</i>	$1.8 \times 10^{-4}$	$10^{-3}$	$2.2 \times 10^{-4}$	$7 \times 10^{-4}$
$R^2$	0.99	0.99	0.99	0.99

## 6. Conclusions

In this work, we explored a data augmentation strategy aimed at enhancing the performance of neural networks for epidemic modeling, thereby focusing on both interpolation and prediction tasks. By combining real-world data with synthetic data generated through simulations of socially structured SIAR models that incorporate parameter uncertainty and age dependence, we demonstrated the ability of neural networks to effectively capture complex epidemic dynamics when additional data from suitable models are included into the training. In addition to PINNs, we considered an emerging class of models known as NAR networks. These networks, which use past time-step solutions to predict future dynamics, have shown particular effectiveness in capturing temporal dependencies, especially in the short time forecasting. Unlike PINNs, which require embedding the governing equations directly into the training process, NAR networks adopt a purely data-driven approach that has proven to be both more accurate and significantly less computationally expensive. A key finding of this study is that incorporating uncertainty into model parameters not only enables the generation of more realistic synthetic datasets but also improves the predictive accuracy of neural networks.

Future research will explore alternative neural network architectures within the class of recurrent models, such as Long Short-Term Memory (LSTM) networks, which may offer further advantages for time-series forecasting. Moreover, we plan to extend our framework to spatially dependent epidemic models under uncertainty. This direction is motivated by the fact that disease transmission is often spatially heterogeneous. For example, urban areas typically exhibit a faster spread than rural regions, and spatially targeted interventions such as localized lockdowns or travel restrictions introduce additional complexity. Incorporating spatial structure and uncertainty will enable a more realistic and localized understanding of epidemic dynamics, thus supporting more effective and adaptive containment strategies.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

This work has been written within the activities of GNCS and GNFM groups of INdAM (Italian National Institute of High Mathematics). G.D. has been partially funded by the European Union — NextGenerationEU, MUR–PRIN 2022 PNRR Project No. P2022JC95T “Datadriven discovery and control of multi-scale interacting artificial agent systems”. G.D. and F.F. thank the Italian Ministry of University and Research (MUR) through the PRIN 2020 project (No. 2020JLWP23) “Integrated Mathematical Approaches to Socio–Epidemiological Dynamics”. L.P. has been partially funded by the European Union– NextGenerationEU under the program “Future Artificial Intelligence– FAIR” (code PE0000013), MUR PNRR, Project “Advanced MATHematical methods for Artificial Intelligence– MATH4AI”. L.P. acknowledges the support by the Royal Society under the Wolfson Fellowship “Uncertainty quantification, datadriven simulations and learning of multiscale complex systems governed by PDEs” and by MIUR-PRIN 2022 Project (No. 2022KKJP4X), “Advanced numerical methods for time dependent parametric partial differential equations with applications”. The partial support by

ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union – NextGenerationEU is also acknowledged.

## Conflict of interest

The authors declare there is no conflict of interest.

## References

1. G. Chowell, C. Ammon, N. W. Hengartner, J. Hyman, Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions, *J. Theor. Biol.*, **241** (2006), 193–204. <https://doi.org/10.1016/j.jtbi.2005.11.026>
2. G. Dimarco, L. Pareschi, G. Toscani, M. Zanella, Wealth distribution under the spread of infectious diseases, *Phys. Rev. E*, **102** (2020), 022303. <https://doi.org/10.1103/PhysRevE.102.022303>
3. A. Lunelli, A. Pugliese, C. Rizzo, Epidemic patch models applied to pandemic influenza: Contact matrix, stochasticity, robustness of predictions, *Math. Biosci.*, **220** (2009), 24–33. <https://doi.org/10.1016/j.mbs.2009.03.008>
4. G. Toscani, A multi-agent approach to the impact of epidemic spreading on commercial activities, *Math. Models Methods Appl. Sci.*, **32** (2022), 1931–1948. <https://doi.org/10.1142/S0218202522500440>
5. G. Albi, L. Pareschi, M. Zanella, Control with uncertain data of socially structured compartmental epidemic models, *J. Math. Biol.*, **82** (2021), 63. <https://doi.org/10.1007/s00285-021-01617-y>
6. G. Albi, L. Pareschi, M. Zanella, Modelling lockdown measures in epidemic outbreaks using selective socio-economic containment with uncertainty, *Math. Biosci. Eng.*, **18** (2021), 7161–7191. <https://doi.org/10.3934/mbe.2021355>
7. L. Bolzoni, E. Bonacini, R. D. Marca, M. Groppi, Optimal control of epidemic size and duration with limited resources, *Math. Biosci.*, **315** (2019), 108232. <https://doi.org/10.1016/j.mbs.2019.108232>
8. S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, et al., Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, *Nature*, **584** (2020), 257–261. <https://doi.org/10.1038/s41586-020-2405-7>
9. S. Lee, G. Chowell, C. Castillo-Chávez, Optimal control for pandemic influenza: The role of limited antiviral treatment and isolation, *J. Theor. Biol.*, **265** (2010), 136–150. <https://doi.org/10.1016/j.jtbi.2010.04.003>
10. V. Capasso, G. Serio, A generalization of the Kermack-McKendrick deterministic epidemic model, *Math. Biosci.*, **42** (1978), 43–61. [https://doi.org/10.1016/0025-5564\(78\)90006-8](https://doi.org/10.1016/0025-5564(78)90006-8)
11. H. W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.*, **42** (2000), 599–653. <https://doi.org/10.1137/S0036144500371907>
12. W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. London, Ser. A*, **115** (1927), 700–721. <https://doi.org/10.1098/rspa.1927.0118>

13. 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** (2018), 172341. <https://doi.org/10.1098/rsos.172341>
14. C. Anastassopoulou, L. Russo, A. Tsakris, C. Siettos, Data-based analysis, modelling and forecasting of the COVID-19 outbreak, *PloS One*, **15** (2020), e0230405. <https://doi.org/10.1371/journal.pone.0230405>
15. K. Y. Leung, P. Trapman, T. Britton, Who is the infector? Epidemic models with symptomatic and asymptomatic cases, *Math. Biosci.*, **301** (2018), 190–198. <https://doi.org/10.1016/j.mbs.2018.04.002>
16. 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** (2020), 2000180. <https://doi.org/10.2807/1560-7917.ES.2020.25.40.2010082>
17. G. Albi, G. Bertaglia, W. Boscheri, G. Dimarco, L. Pareschi, G. Toscani, et al., Kinetic modelling of epidemic dynamics: Social contacts, control with uncertain data, and multiscale spatial dynamics, in *Predicting Pandemics in a Globally Connected World: Toward a Multiscale, Multidisciplinary Framework through Modeling and Simulation*, Springer, **18** (2022), 43–108. [https://doi.org/10.1007/978-3-030-96562-4\\_3](https://doi.org/10.1007/978-3-030-96562-4_3)
18. G. Bertaglia, W. Boscheri, G. Dimarco, L. Pareschi, Spatial spread of COVID-19 outbreak in Italy using multiscale kinetic transport equations with uncertainty, *Math. Biosci. Eng.*, **18** (2021), 7028–7059. <https://doi.org/10.3934/mbe.2021350>
19. A. Capaldi, S. Behrend, B. Berman, J. Smith, J. Wright, A. L. Lloyd, Parameter estimation and uncertainty quantification for an epidemic model, *Math. Biosci. Eng.*, **9** (2012), 553–576. <https://doi.org/10.3934/mbe.2012.9.553>
20. G. Dimarco, B. Perthame, G. Toscani, M. Zanella, Kinetic models for epidemic dynamics with social heterogeneity, *J. Math. Biol.*, **83** (2021), 4. <https://doi.org/10.1007/s00285-021-01630-1>
21. A. Korobeinikov, P. K. Maini, Non-linear incidence and stability of infectious disease models, *Math. Med. Biol.*, **22** (2005), 113–128. <https://doi.org/10.1093/imammb/dqi001>
22. M. Zanella, C. Bardelli, G. Dimarco, S. Deandrea, P. Perotti, M. Azzi, et al., A data-driven epidemic model with social structure for understanding the COVID-19 infection on a heavily affected Italian province, *Math. Models Methods Appl. Sci.*, **31** (2021), 2533–2570. <https://doi.org/10.1142/S021820252150055X>
23. C. Castillo-Chavez, H. W. Hethcote, V. Andreasen, S. A. Levin, W. M. Liu, Epidemiological models with age structure, proportionate mixing, and cross-immunity, *J. Math. Biol.*, **27** (1989), 233–258. <https://doi.org/10.1007/BF00275810>
24. A. Franceschetti, A. Pugliese, Threshold behaviour of a SIR epidemic model with age structure and immigration, *J. Math. Biol.*, **57** (2008), 1–27. <https://doi.org/10.1007/s00285-007-0143-1>
25. I. Voinsky, G. Baristaite, D. Gurwitz, Effects of age and sex on recovery from COVID-19: Analysis of 5769 Israeli patients, *J. Infect.*, **81** (2020), e102–e103. <https://doi.org/10.1016/j.jinf.2020.05.026>

26. S. Han, L. Stelz, H. Stoecker, L. Wang, K. Zhou, Approaching epidemiological dynamics of COVID-19 with physics-informed neural networks, *J. Franklin Inst.*, **361** (2024), 106671.
27. C. Lu, X. Zhu, Bifidelity data-assisted neural networks in nonintrusive reduced-order modeling, *J. Sci. Comput.*, **87** (2021), 8. <https://doi.org/10.1007/s10915-020-01403-w>
28. G. Bertaglia, C. Lu, L. Pareschi, X. Zhu, Asymptotic-preserving neural networks for multiscale hyperbolic models of epidemic spread, *Math. Models Methods Appl. Sci.*, **32** (2022), 1949–1985. <https://doi.org/10.1142/S0218202522500452>
29. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
30. C. Millevoi, D. Pasetto, M. Ferronato, A Physics-Informed Neural Network approach for compartmental epidemiological models, *PLoS Comput. Biol.*, **20** (2024), e1012387. <https://doi.org/10.1371/journal.pcbi.1012387>
31. A. McInerney, K. Burke, A statistical modelling approach to feedforward neural network model selection, *Stat. Model.*, **25** (2025), 323–342. <https://doi.org/10.1177/1471082X241258261>
32. E. Pashaei, E. Pashaei, Training feedforward neural network using enhanced black hole algorithm: A case study on COVID-19 related ACE2 gene expression classification, *Arabian J. Sci. Eng.*, **46** (2021), 3807–3828. <https://doi.org/10.1007/s13369-020-05217-8>
33. C. L. Chen, D. B. Kaber, P. G. Dempsey, A new approach to applying feedforward neural networks to the prediction of musculoskeletal disorder risk, *Appl. Ergon.*, **31** (2000), 269–282. [https://doi.org/10.1016/S0003-6870\(99\)00055-1](https://doi.org/10.1016/S0003-6870(99)00055-1)
34. T. De Ryck, S. Mishra, Numerical analysis of physics-informed neural networks and related models in physics-informed machine learning, *Acta Numer.*, **33** (2024), 633–713. <https://doi.org/10.1017/S0962492923000089>
35. S. Berkhahn, M. Ehrhardt, A physics-informed neural network to model COVID-19 infection and hospitalization scenarios, *Adv. Contin. Discrete Models*, **2022** (2022), 61. <https://doi.org/10.1186/s13662-022-03733-5>
36. I. D. Mienye, T. G. Swart, G. Obaido, Recurrent neural networks: A comprehensive review of architectures, variants, and applications, *Information*, **15** (2024), 517. <https://doi.org/10.3390/info15090517>
37. Z. Li, Y. Zheng, J. Xin, G. Zhou, A Recurrent Neural Network and differential equation based spatiotemporal infectious disease model with application to COVID-19, in *12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, **23** (2020).
38. A. B. Amendolara, D. Sant, H. G. Rotstein, E. Fortune, LSTM-based recurrent neural network provides effective short term flu forecasting, *BMC Public Health*, **23** (2023), 1788. <https://doi.org/10.1186/s12889-023-16720-6>
39. R. Sarkar, S. Julai, S. Hossain, W. T. Chong, M. Rahman, A comparative study of activation functions of NAR and NARX neural network for long-term wind speed forecasting in Malaysia, *Math. Probl. Eng.*, **2019** (2019), 6403081. <https://doi.org/10.1155/2019/9605393>



40. M. Awais, A. S. Ali, G. Dimarco, F. Ferrarese, L. Pareschi, A data augmentation strategy for deep neural networks with application to epidemic modelling, *Boll. Unione Mat. Ital.*, **2025** (2025), 1–17. <https://doi.org/10.1007/s40574-025-00486-3>
41. M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, et al., Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures, *Proc. Natl. Acad. Sci. U.S.A.*, **117** (2020), 10484–10491. <https://doi.org/10.1073/pnas.2004978117>



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)