

A model learning framework for inferring the dynamics of transmission rate depending on exogenous variables for epidemic forecasts

Giovanni Ziarelli ^{ID}*, Stefano Pagani ^{ID}, Nicola Parolini ^{ID}, Francesco Regazzoni ^{ID}, Marco Verani ^{ID}

MOX, Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano, 20133, Italy

ARTICLE INFO

Dataset link: <https://github.com/giovanziarelli/inferringTRdynamicsML/>

Keywords:

Scientific machine learning
Model learning
Data assimilation
Epidemiology
Forecasts
Neural differential equations
Hidden dynamics

ABSTRACT

In this work, we aim to formalize a novel scientific machine learning framework to reconstruct the hidden dynamics of the transmission rate, whose inaccurate extrapolation can significantly impair the quality of the epidemic forecasts, by incorporating the influence of exogenous variables (such as environmental conditions and strain-specific characteristics). We propose a hybrid model that blends a data-driven layer with a physics-based one. The data-driven layer is based on a neural ordinary differential equation that learns the dynamics of the transmission rate, conditioned on the meteorological data and wave-specific latent parameters. The physics-based layer, instead, consists of a standard *SEIR* compartmental model, wherein the transmission rate represents an input. The learning strategy follows an end-to-end approach: the loss function quantifies the mismatch between the actual numbers of infections and its numerical prediction obtained from the *SEIR* model incorporating as an input the transmission rate predicted by the neural ordinary differential equation. We apply this original approach to both a synthetic test case and a realistic test case based on meteorological data (temperature and humidity) and influenza data from Italy between 2010 and 2020. In both scenarios, we achieve low generalization error on the test set and observe strong alignment between the reconstructed model and established findings on the influence of meteorological factors on epidemic spread. Finally, we implement a data assimilation strategy to adapt the neural equation to the specific characteristics of an epidemic wave under investigation, and we conduct sensitivity tests on the network's hyperparameters.

1. Introduction

Mathematical models have been extensively employed in epidemiology in order to answer to the most common questions arising from both policy-makers and the scientific community. As the recent SARS-CoV-2 pandemic has highlighted, key-problems in this sense are the allocation of pharmaceutical [1–3] and non-pharmaceutical control interventions [4–6] and, especially, the forecasting of infection trends [7,8]. Long-term scenario analyses are valuable for making strategic decisions regarding, e.g., treatment facilities placement, non-pharmaceutical interventions allocation, and the social and economical burden. On the other hand, short-term forecasts, ranging from days to weeks, help in predicting the immediate need for resources such as protective gear, ventilators,

* Corresponding author.

E-mail address: giovanni.ziarelli@polimi.it (G. Ziarelli).

<https://doi.org/10.1016/j.cma.2025.117796>

Received 22 October 2024; Received in revised form 1 January 2025; Accepted 24 January 2025

Available online 4 February 2025

0045-7825/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

hospital beds and vaccinations. However, these forecasts pose significant challenges due to the many uncertainties that arise over extended time horizons.

Over the years, different modeling strategies have been proposed to tackle the problem of epidemic forecasting. For instance, compartmental models are a popular choice and can be tailored for taking into account complex dynamics, e.g. transmission mechanisms [9], concurring variants and illness-specific peculiarities [10], mobility [11] or geographic dependencies of infections' diffusion [12]. An alternative paradigm which does not involve adopting equation-based models is represented by Agent-Based Models [13,14]. ABMs follow a bottom-up approach for evaluating various interventions [15], like home-schooling [16], mobility-restrictions and so forth. Recently, machine-learning-based methods have become widespread and reliable tools for making predictions and scenario analyses based on available data [17,18]. In some cases, the demand for accurate epidemic surrogate models has led to the development of innovative architectures, such as Asymptotic Preserving Neural Networks (APINN) [19]. These networks are designed to address the challenges posed by multiscale hyperbolic PDE problems by incorporating heterogeneities, such as geographic features, into the model. Additionally, drawing inspiration from fields like meteorology, forecasts generated by various techniques have recently been aggregated into forecasting hubs [20,21]. By combining different models using standardized formats, these ensemble predictions have shown greater accuracy than most individual models.

In general, the reliability of epidemic forecasts is intertwined with the choice of the model, the calibration strategy, and the method used to extrapolate the estimated parameters beyond the calibration interval. A common approach for extrapolation is to keep parameters constant over time. This strategy turns out to be effective when applied to parameters that can be determined through specific clinical trials and cohort studies, as they are related to the virological characteristics of the disease, such as the average recovery rate or incubation period. Alternatively, linear, polynomial or exponential extrapolations are commonly used techniques, but suffer when the parameters exhibit a complex time-dependent behavior. Among the many parameters to be calibrated and extrapolated in an epidemiological model, a special role is played by the transmission rate, which is a time dependent parameter crucially influencing the overall transmission mechanism. A poor extrapolation strategy of the transmission rate can seriously harm the quality of the overall forecasting process [20].

In recent years, different approaches have been proposed to extract from available data the time evolution behavior of the transmission rate. For instance, in [22] the authors propose a new approach for training PINNs tracking temporal changes in epidemiological data and reconstructing the transmission rate. Instead, a new architecture called Transmission-Dynamics-Informed Neural Network (TDINN) has been formalized for simulating the COVID19 epidemic in five Chinese cities, combining scattered data and *SIR*-like models in a physics-informed neural network (PINN) fashion [23]. A more straightforward approach, adopted in [24], consists in prescribing a time dependent law for the transmission rate based on exponential decay models [25], which need to be fitted by available data on infections.

Actually, the time-dependent law derived from the aforementioned approaches does not fully account for the fact that the transmission rate depends on varying external factors. By definition, the transmission rate represents the number of infected individuals per unit of time per susceptible individual necessary to spread the disease. Thus, its value is influenced by both the intrinsic transmissibility, which remains constant over short time horizons without the emergence of new variants, and by external, or exogenous, factors such as environmental conditions (e.g., climate) [26], social dynamics shaped by mobility patterns [27], the enforcement of non-pharmaceutical interventions and the distribution of vaccine doses across different age groups. Consequently, incorporating these factors into the dynamic evolution of the transmission rate is crucial to improve the reliability of the forecasts.

It is worth mentioning that calibrating time-dependent coefficients influenced by exogenous variables presents a significant technical challenge across many scientific and engineering fields, ranging from epidemiology [28,29], to cardiac applications [30,31], climate modeling [32], and control systems [33].

1.1. Original contributions

In this work, we propose an innovative neural network architecture, in a model learning fashion (see, e.g., [34]), designed to learn the evolution of the transmission rate associated to a prescribed compartmental model. A schematic overview of the architecture is provided in Fig. 1. The proposed architecture is hybrid in nature, as it consists of a data-driven layer (learning the transmission rate dynamics) coupled with a physics-based layer (predicting the epidemic scenario). The data-driven differential model for the transmission rate is governed by a neural network that takes also exogenous variables as input. The output is then integrated using a standard numerical integration scheme to generate the full time series of the transmission rate. This time series is subsequently inserted into a physics-based epidemiological ODE model to predict disease incidence. For the ease of presentation, we employ the classical *SEIR* compartmental model, commonly used for epidemic diseases with an incubation period, but the framework is general enough to be employed with other different compartmental (or not compartmental) models. All training variables are calibrated, in an end-to-end training, by minimizing an error metric based on infection data, specifically by reducing the mean squared error between the observed number of new cases (weekly incidence) and the corresponding values reconstructed by the hybrid model. Additionally, we introduce a data assimilation strategy to recover, during both training and testing phases, a latent parameter that affects the transmission rate. This parameter is interpreted as a measure of the intrinsic transmissibility of the illness, influenced by various factors such as disease strains, the immunization profile and individuals' behavior in response to the outbreak, and it characterizes each wave of the same illness independently.

The proposed approach enables learning and exploration of the hidden dynamics of the transmission rate in relation to external factors, while providing reliable forecasts through the *SEIR* model without depending on common, often inaccurate, extrapolation techniques.

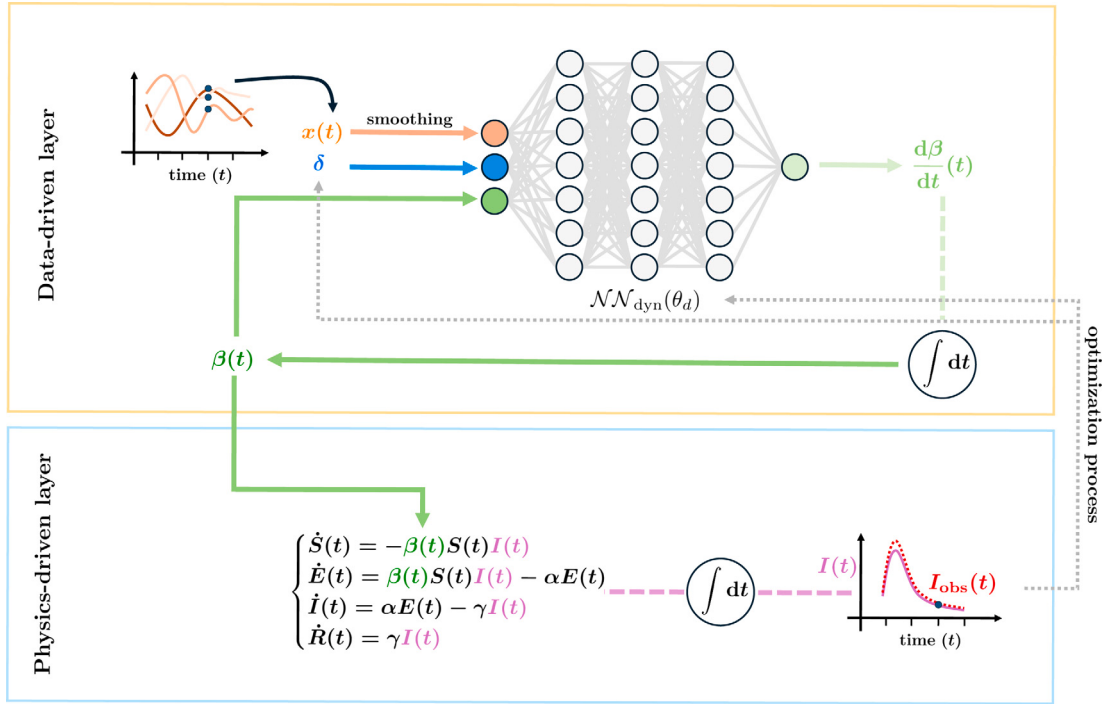


Fig. 1. Schematic representation of the architecture. In the data-driven layer we approximate the unknown time-dependent behavior of the transmission rate $\beta(t)$, depending on the prescribed input exogenous quantities, by learning the right-hand side of its dynamics with a neural network. The resulting transmission rate, obtained by numerically integrating in time the ODE, feeds the chosen epidemiological model. During training, the mismatch between observations on measurable quantities of interest and the hybrid model combining the data-driven and physics-based layer is minimized in order to retrieve optimal values of the trainable variables of the network and the latent parameters δ associated with each trajectory. The latter parameters account for potential exogenous quantities beyond those explicitly included.

1.2. Outline

This paper is structured as follows. In Section 2, we introduce our strategy, focusing specifically on how we addressed the training and testing stages, as well as the notation that will be used throughout the work. In Section 3, we carry out an extensive numerical testing campaign to evaluate the sensitivity of the architecture to the neural network's topology hyperparameters in a synthetic test case (see Section 3.1), where the differential law of the transmission rate is prescribed *a priori*, and temperature is treated as an exogenous variable. We also discuss the validity of this approach in a real-world scenario using data from influenza epidemics in Italy between 2010 and 2020 (see Section 3.2), investigating the role of meteorological factors such as temperature and relative humidity on the differential law governing the transmission rate (the impact of both factors on the transmission mechanism has been largely studied in epidemiology). Finally, in Section 4, we summarize the main contributions of this work and highlight potential future directions and extensions.

2. Methods

In this section, we introduce the notation used in this work and illustrate our proposed method for making reliable forecasts by analyzing the evolving dynamics of the transmission rate. We define the vector of exogenous variables as,

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} : [t_0, T] \subset \mathbb{R}^+ \rightarrow \mathbb{R}^n. \quad (1)$$

Each $x_i(t)$, $\forall i = 1, 2, \dots, n$ represents a different exogenous quantity on which we assume the transmission rate depends. The choice of x_i variables depends on the conditions that modellers want to embody in the analysis and on the disease at stake. In the following, we will consider the case of infectious respiratory diseases as influenza, that shows a marked wintertime seasonality [35]. In this respect, some of the factors driving seasonality dependence are: variations in the host's ability to withstand immune system stress caused by extremes in temperature, as measured by melatonin and vitamin D levels; environmental factors, such as temperature itself, humidity, UV irradiation, and the direction of ambient air movement, which influence spread of the disease and seasonal

variations in the host's behavior. Another key quantity which impacts on the transmission rate is the relative transmissibility of the influenza virus itself, depending on the major lineages and on the virological composition of the spreading disease (cf. e.g., [36,37]). More precisely, in *Test case 1* (cf. Section 3.1) we choose $x_1(t)$ as $T_p : [t_0, T] \rightarrow \mathbb{R}$ i.e. the temperature measured in Celsius degree, taken as a suitable index for seasonality, whilst a scalar δ can be thought as an unknown parameter taking into account for unmodeled quantities, including lineages variability. Furthermore, in *Test case 2* (cf. Section 3.2) we consider an additional exogenous variable $x_2(t)$ as $U : [t_0, T] \rightarrow \mathbb{R}$, which is the average relative humidity value. We postpone to Section 3.2 the discussion on the motivations behind our choices. Additionally, in both test cases we assume that transmission rate depends on a *latent parameter* $\delta \in \mathbb{R}$, which is a possible unknown exogenous variable (to be learned).

Hence, our goal is to learn the function f governing the unknown dynamics of the transmission rate:

$$\begin{cases} \dot{\beta}(t) = f(\beta(t); \mathbf{x}(t), \delta) & \forall t \in (t_0, T] \\ \beta(t_0) = \beta_0. \end{cases} \quad (2)$$

The identification of f is further complicated by the lack of direct measures of β , with the only available data being the effect on the epidemic (like, e.g., the number of weekly new infected patients). This problem can be reformulated as a statistical learning problem based on the accessible and observable data by properly choosing the functional space $\mathcal{F}$ containing f .

Problem 1. Let $\{\tilde{\mathbf{x}}^{(s)}, \tilde{\mathbf{y}}^{(s)}\}_{s \leq N}$ the sequence of pairs of observable data, where $\tilde{\mathbf{x}}^{(s)} \in \mathcal{X}_d = \{\mathbf{x} : [t_0, T] \rightarrow \mathbb{R}^n\}$ represents an input exogenous quantity and $\tilde{\mathbf{y}}^{(s)} \in \mathcal{Y}_d = \{\mathbf{y} : [t_0, T] \rightarrow \mathbb{R}\}$ is the associated output functions. Given $\tilde{\mathbf{x}}^{(s)}$, we define $\beta^{(s)}$ as the corresponding solution of (2) for a certain $f \in \mathcal{F}$. Each $\beta^{(s)}$ is uniquely determined once the couple of associated initial value and latent parameter $\{\beta_0^{(s)}, \delta^{(s)}\}$ is picked. Furthermore, we assume that an epidemic model, mapping the transmission rate to the output quantity $y^{(s)}(t) = y(\beta^{(s)})(t)$, is prescribed. Our learning problem reads as follows:

$$f^*, \{\beta_0^{(s)*}\}_{s=1}^N, \{\delta^{(s)*}\}_{s=1}^N = \underset{f \in \mathcal{F}, \{\beta_0^{(s)}\}_{s=1}^N \in \mathbb{R}, \{\delta^{(s)}\}_{s=1}^N \in \mathbb{R}}{\operatorname{argmin}} \mathcal{L}(\mathbf{y}^{(s)}, \tilde{\mathbf{y}}^{(s)}), \quad (3)$$

where $\mathcal{L} : \mathcal{Y}_d \times \mathcal{Y}_d \rightarrow \mathbb{R}$ is a properly defined discrepancy measure or error metric to be minimized (see Eqs. (7) and (8) for precise definitions in our context).

Inspired by the recently introduced Latent Dynamics Neural Networks (LDNets) [38], we choose $\mathcal{F}$ equal to the space of neural network functions and we assume that the map from the transmission rate to the output is given by a classical compartmental model. A schematic representation of our methodology is depicted in Fig. 1. Additionally, we propose a computational strategy to estimate online the δ variable for both training and testing stages.

Specifically, we set

$$\begin{aligned} \mathcal{F} = \left\{ f_{nn} : \mathbb{R} \times \mathbb{R}^{n+1} \rightarrow \mathbb{R} \text{ s.t. } f_{nn}(\beta, \mathbf{x}, \delta; \theta_d) = W_L z^L + b_L, \right. \\ \left. \text{where } z^l = \sigma(W_{l-1} z^{l-1} + b_{l-1}) \text{ for } l = 1, 2, \dots, L, z^0 = \begin{bmatrix} \beta \\ \mathbf{x} \\ \delta \end{bmatrix} \right\}, \end{aligned} \quad (4)$$

i.e. the set of neural network functions with L layers and N_l neurons per layer. We denote by the generic θ_d the set of all trainable network parameters, namely the matrices W_l and vectors b_l for $l = 0, \dots, L$.

We frame Problem 1 in the specific context of this work. We define a map from the input $[x_1, x_2, \dots, x_n]^T \in \mathcal{X} \subset [L^\infty([t_0, T])]^n$ to the vector $\mathbf{y} = [S, E, I, R]^T \in \mathcal{Y} \subset [C^0([t_0, T])]^4$ by solving the following system of ODEs:

$$\begin{cases} \dot{\beta}(t) = f_{nn}(\beta(t), \mathbf{x}(t), \delta; \theta_d) & \forall t \in (t_0, T] \\ \beta(t_0) = \beta_0 \\ \dot{\mathbf{y}}(t) = \text{SEIR}(\mathbf{y}(t), \beta(t)) & \forall t \in (t_0, T] \\ \mathbf{y}(t_0) = \mathbf{y}_0, \end{cases} \quad (5)$$

where the SEIR epidemic model [39–43] reads as:

$$\begin{cases} \dot{S}(t) = -\beta(t)S(t)I(t), & \forall t \in (t_0, T], \\ \dot{E}(t) = \beta(t)S(t)I(t) - \alpha E(t), & \forall t \in (t_0, T], \\ \dot{I}(t) = \alpha E(t) - \gamma I(t), & \forall t \in (t_0, T], \\ \dot{R}(t) = \gamma I(t), & \forall t \in (t_0, T], \\ [S(t_0), E(t_0), I(t_0), R(t_0)]^T = [S_0, E_0, I_0, R_0]^T. \end{cases} \quad (6)$$

The states $S, E, I, R : [t_0, T] \rightarrow \mathbb{R}^+$ represent the relative amount of susceptible, exposed, infected and recovered individuals with respect to the illness under investigation. In this work we assume that a suitable calibration stage for initial conditions $[S_0, E_0, I_0, R_0]^T$ has been already undertaken before the simulated scenarios. The parameter $\alpha = \frac{1}{T_I} > 0$ is the inverse of the average incubation time. Instead, $\gamma = \frac{1}{T_R} > 0$ is the recovery rate, where T_R represents the recovery time, i.e. the mean time infected individuals remain infected and infectious. Both parameters can be estimated starting from contact tracing studies during

the epidemic waves, through cohort studies following groups of people exposed to the disease, or through statistical and Bayesian inference, e.g. exploiting Kaplan–Meier estimation as in [44].

Before addressing the solution of Problem 1, let us mention that for integrating (5) we employ the Forward Euler method with a constant time step, Δt , selecting the step size to ensure both stability and accurate representation of the system's dynamics. For this purpose, it may be necessary to integrate the differential equations on a finer grid than the observation grid. In our case we perform a suitable resampling of each $\mathbf{x}^{(s)}(t)$ exploiting piecewise linear interpolation.

We are now ready to describe the computational strategy for addressing Problem 1. In doing this we provide details about the training and test phases. During the training phase we look for the optimal values of the network trainable variables (i.e. weights and biases), the initial conditions for the transmission rate and the corresponding latent parameter for each training trajectory. The testing phase on unseen inputs is, in turn, subdivided in two stages, as it is schematically synthesized in Fig. 2: (a) estimation stage, combined with data-assimilation techniques, for retrieving initial transmission rate and latent parameter of each testing trajectory; (b) prediction stage to guess the trend of the epidemic model for each input by numerically solving the transmission and epidemic ODEs using the corresponding estimated parameters. In the following, we detail the numerical procedure of each stage:

1. **Training:** We employ a two-step optimization routine, where the two steps differ in the choice of the cost functional to be minimized: the output of the first step is employed as initial guess for the second step. Specifically, we first optimize $\mathcal{L}_{\text{train}, \mathcal{E}_1}$ and then $\mathcal{L}_{\text{train}, \mathcal{E}_2}$, where

$$\begin{aligned} \mathcal{L}_{\text{train}, \mathcal{E}_1}(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N) = & \sum_{i=1, \dots, N} \sum_{\omega \in \mathcal{W}} \mathcal{E}_1(\{\mathbf{y}^{(i)}(\tau)\}_{\tau \in (\omega, \omega+1]}, \overline{\Delta I}^{(i)}(\omega)) + \alpha_{\text{reg}} \mathcal{R}(\theta_d) \\ & + B(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N) + D(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N) \\ & + I(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N) + \mathcal{A}(\{\delta^{(i)}\}_{i=1}^N), \end{aligned} \quad (7)$$

and

$$\begin{aligned} \mathcal{L}_{\text{train}, \mathcal{E}_2}(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N) = & \sum_{i=1, \dots, N} \sum_{\omega \in \mathcal{W}} \mathcal{E}_2(\{\mathbf{y}^{(i)}(\tau)\}_{\tau \in (\omega, \omega+1]}, \overline{\Delta I}^{(i)}(\omega)) + \alpha_{\text{reg}} \mathcal{R}(\theta_d) \\ & + B(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N) + D(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N) \\ & + I(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N) + \mathcal{A}(\{\delta^{(i)}\}_{i=1}^N), \end{aligned} \quad (8)$$

with $\mathcal{W}$ representing the set of weeks considered in each scenario. In each loss function we also consider other regularization terms (described in detail in Appendix A in the Supplementary Materials) in order to help the training process. Each term has its own specific weight factor to be tuned (α_*). The numerical values corresponding to each test case will be provided in their respective sections. All the optimization problems are solved numerically through two iterative schemes applied in sequence: the first order Adam and the second order BFGS (for additional information see, e.g., [38]). To ensure BFGS convergence, we perform a sufficient number of Adam iterations, following manual selection when the cost functional reaches a plateau. In contrast, BFGS terminates based on an early stopping criterion.

The two loss functions differ for the discrepancy metrics ($\mathcal{E}_*$) employed. Indeed, for the first $N_{ep,1}$ iterations, we minimize $\mathcal{L}_{\text{train}, \mathcal{E}_1}$ with $\mathcal{E}_1$:

$$\mathcal{E}_1(\{\mathbf{y}^{(i)}(\tau)\}_{\tau \in (\omega, \omega+1]}, \overline{\Delta I}^{(i)}(\omega), \delta^{(i)}) = \alpha_{\mathcal{E},1} \frac{(\Delta I^{(i)}(\omega) - \overline{\Delta I}^{(i)}(\omega))^2}{\overline{\Delta I}^{(i)}(\omega)^2}. \quad (9)$$

This metric measures the relative discrepancies among weekly new cases, the computed ($\Delta I^{(i)}(\omega)$) and the target ($\overline{\Delta I}^{(i)}(\omega)$) ones. The choice of considering the weekly datum is guided by the way available data about infected individuals are commonly delivered for epidemic outbreaks (e.g. influenza's new infections are collected weekly in a public repository as it is described in Section 3.2).

Afterwards, we focus on the second optimization process, which is crucial for accurately determining peaks, a primary objective in epidemiology. Indeed, in the first optimization process, we prioritized learning the start and end times of epidemic wave phases by minimizing relative discrepancies ($\mathcal{E}_1$). As a result, at the end of the first stage, we often obtain inaccurate trajectories in terms of peak values. Therefore, after completing $N_{ep,1}$ iterations, we initiate the second optimization process, using the current values of the trainable variables, and aiming to minimize $\mathcal{L}_{\text{train}, \mathcal{E}_2}$ over the remaining $N_{ep,2}$ iterations. The metric $\mathcal{E}_2$ is defined as:

$$\mathcal{E}_2(\{\mathbf{y}^{(i)}(\tau)\}_{\tau \in (\omega, \omega+1]}, \overline{\Delta I}^{(i)}(\omega), \delta^{(i)}) = \alpha_{\mathcal{E},2} \left(\Delta I^{(i)}(\omega) - \overline{\Delta I}^{(i)}(\omega) \right)^2. \quad (10)$$

Note that, with an abuse of notation, we omit the explicit dependencies of the transmission rate, $\beta(t) = \beta(t; \theta_d, \delta)$ and the term $\Delta I(\omega) = \Delta I(\omega; \theta_d, \delta)$. The training optimization routine reads as follows:

Routine. Initialize $(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N)$. Then, solve

$$(\theta_{d,1}^*, \{\beta_{0,1}^{i,*}\}_{i=1}^N, \{\delta_1^{i,*}\}_{i=1}^N) = \underset{\theta_d, \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N}{\text{argmin}} \mathcal{L}_{\text{train}, \mathcal{E}_1}(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N). \quad (11)$$

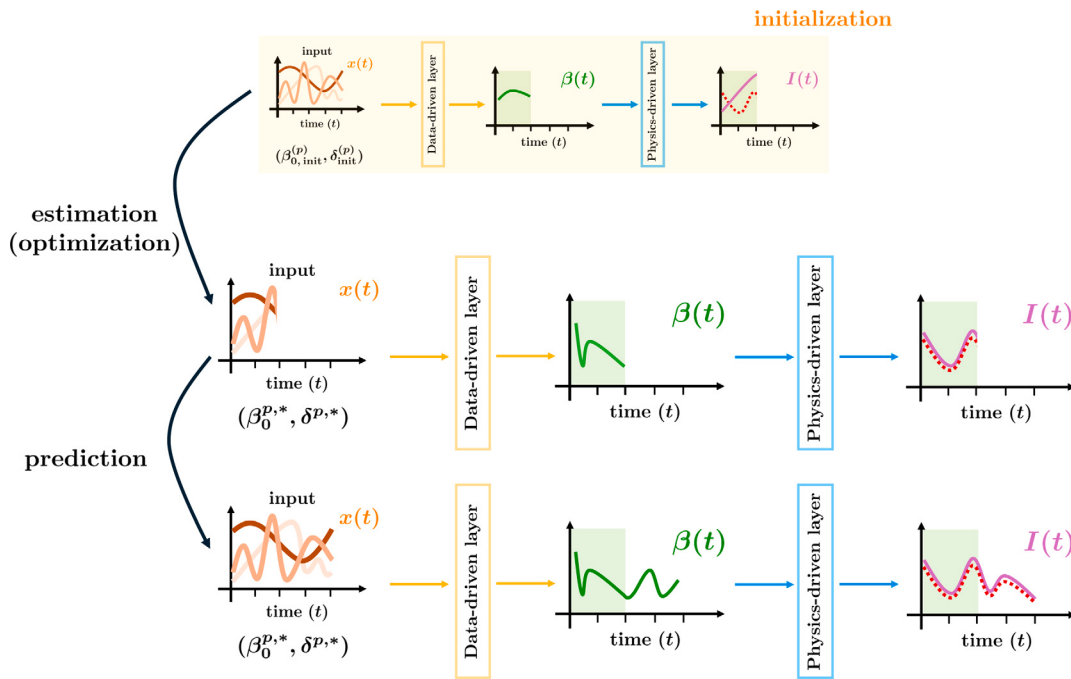


Fig. 2. Schematic representation of the testing stage, which is subdivided in estimation and prediction, for the specific trajectory of an epidemic event. The dashed red line represents the target trajectories of infected individuals (right pictures). The green line indicate the reconstructed transmission rate, while the pink lines depict the reconstructed infected individuals. The green shaded area corresponds the observation window $[t_0, T_{\text{obs}}]$ for estimation. The dashed red data in the observation windows enter in the loss function to minimize during the estimation phase. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Starting from $(\theta_{d,1}^*, \{\beta_{0,1}^{i,*}\}_{i=1}^N, \{\delta_{1,1}^{i,*}\}_{i=1}^N)$, we solve

$$(\theta_d^*, \{\beta_0^{i,*}\}_{i=1}^N, \{\delta^{i,*}\}_{i=1}^N) = \underset{\theta_d, \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N}{\operatorname{argmin}} \mathcal{L}_{\text{train}, \mathcal{E}_2}(\theta_d; \{\beta_0^{(i)}\}_{i=1}^N, \{\delta^{(i)}\}_{i=1}^N). \quad (12)$$

2. **Estimation:** Once a testing sample is drawn, its associated $\delta^{(p)}$ needs to be estimated, together with the initial value for the transmission rate $\beta_0^{(p)}$. For this purpose, we rely on the approach proposed in [45] which is based on data-assimilation techniques.

Starting from given initial guesses $(\beta_{0,\text{init}}^{p,*}, \delta_{\text{init}}^{p,*})$, we solve another optimization problem for determining the unknown couple of coefficients $(\beta_0^{(p)}, \delta^{(p)})$, while keeping frozen the parameters of the trained model. In the interval $[0, T_{\text{obs}}]$, where $T_{\text{obs}} \leq T$ is a prescribed observation time, we minimize the following estimation loss:

$$\begin{aligned} \mathcal{L}_{\text{estim}}(\beta_0^{(p)}, \delta^{(p)}) = & \sum_{\omega \in \mathcal{W}, \omega \leq T_{\text{obs}}} \mathcal{E}_2(\{y^{(p)}(\tau)\}_{\tau \in (\omega, \omega+1]}, \overline{\Delta I}^{(p)}(\omega)) \\ & + \alpha_A \frac{(\delta^{(p)} - \bar{\delta})^2}{C_\delta} + \alpha_{\beta_0} \frac{(\beta_0^{(p)} - \bar{\beta}_0)^2}{C_\beta}, \end{aligned} \quad (13)$$

where $(\alpha_A, \alpha_{\beta_0})$ are the weights of the *a priori* regularization. In $\mathcal{L}_{\text{estim}}$ we have prescribed an empirical gaussian-like prior distribution for the (disjoint) pair (δ, β_0) based on the observed values $\{\delta^{(i)}, \beta_0^{(i)}\}_{i=1}^N$ after the training stage. In this case $(\bar{\delta}, \bar{\beta}_0)$ are the sample mean values of the reconstructed parameters after training $\{\delta^{(i)}, \beta_0^{(i)}\}_{i=1}^N$ and (C_δ, C_β) the sample covariances. We remark that the model for the transmission rate has already been determined, *i.e.* weights and biases of the network do not belong to the set of trainable variables. Hence, the estimation problem reads as

$$(\beta_0^{p,*}, \delta^{p,*}) = \underset{\beta_0^{(p)}, \delta^{(p)}}{\operatorname{argmin}} \mathcal{L}_{\text{estim}}(\beta_0^{(p)}, \delta^{(p)}). \quad (14)$$

3. **Prediction:** Once the trainable variables have been optimized and the sample-dependent couple of parameters has been estimated, we solve the forward system (5) to predict the new sample evolution in the whole interval $[t_0, T]$.

Without loss of generality, in the following section we keep $t_0 = 0$.

2.1. Computational demand of the numerical procedure

Pseudo-algorithm 1 Training, estimation and prediction pseudo-algorithms

Training

inputs: exogenous variables $\{\tilde{x}^{(s)}(t)\}_{s \leq N}$ for $t \in [t_0, T]$.

outputs: trainable variables of the neural network θ_d , initial transmission rate and latent parameters for each training sample $\{\beta_0^{s,*}, \delta^{s,*}\}_{s \leq N}$.

1: Initialization: $\theta_d, \{\beta_0^{(s)}, \delta^{(s)}\}_{s \leq N}$

2: Minimization (Adam + BFGS) of $\mathcal{L}_{\text{train}, \mathcal{E}_1}$ for $N_{ep,1}$ iterations;

3: Minimization (Adam + BFGS) of $\mathcal{L}_{\text{train}, \mathcal{E}_2}$ for $N_{ep,2}$ iterations;

Estimation

inputs: exogenous variables $\{\tilde{x}^{(p)}(t)\}_{p \leq N_p}$ for $t \in [t_0, T]$.

outputs: initial transmission rate and latent parameters for each test sample $\{\beta_0^{p,*}, \delta^{p,*}\}_{p \leq N_p}$.

1: Initialization: $\{\beta_0^{(p)}, \delta^{(p)}\}_{p \leq N_p}$

2: Minimization (Adam + BFGS) of $\mathcal{L}_{\text{estim}}$ for N_{estim} iterations;

Prediction

inputs: exogenous variables $\{\tilde{x}^{(p)}(t)\}_{p \leq N_p}$ for $t \in [t_0, T_{\text{obs}}]$, initial transmission rate and latent parameters for each test sample $\{\beta_0^{p,*}, \delta^{p,*}\}_{p \leq N_p}$ retrieved during estimation;

outputs: prediction on each test sample.

1: Forward evaluation of the architecture with $\{\beta_0^{p,*}, \delta^{p,*}\}_{p \leq N_p}$.

In this subsection, we discuss the computational demand of the different stages required to train and test the proposed architecture. A schematic workflow of these different steps is reported in pseudo-algorithm 1.

Addressing the different minimization problems can be computationally intensive, especially for long-term simulations or large datasets. In particular, the training stage involves solving two minimization problems, with different loss functions (cf. Eqs. (7) and (8)), to ensure convergence to a local minimum. In contrast, the estimation stage requires solving a single optimization problem, with fewer parameters to learn (2 for each prediction sample). Both stages use a combination of Adam and BFGS optimization schemes, which encompass computing the gradient of the neural network function with respect to weights, biases, and the latent parameters to be optimized (through automatic differentiation). Adam serves as a proxy scheme to initialize the BFGS iterations close to the minimum.

The computational time is inherently influenced by the network architecture, which, in our case, is designed to be exceptionally lightweight. This design choice reduces both the computational time and the complexity of the optimization procedure by preferring more regular solutions, thereby mitigating the risk of overfitting. Consequently, the highest computational cost arises from the time-integration process and the choice of the time step.

On an 8-core Intel i7 machine, the training stage takes less than 14 min for a dataset of size 50 and a 2-layer, 4-neuron per layer, neural network (17% of the total time is dedicated to Adam iterations, while the remaining time is occupied by BFGS). The estimation stage, with 50 samples, takes approximately 70 s (12% of the total time is spent on Adam iterations, while the remaining 88% is used by BFGS). Instead, prediction on the same 50 samples occurs in less than 10 s.

Although we achieve highly accurate results in a computationally efficient manner, further strategies to reduce the computational burden could be investigated. For instance, good practices to explore are (a) data distillation techniques to shrink the training dataset [46,47], (b) using first-order Adam optimization with a reduce-on-plateau scheduler [48], (c) maintaining a low neuron count, (d) batching [49], and (e) adopting a coarser time step for the surrogate model's transmission rate. In the last approach, care must be taken to avoid overly coarse time steps, which can compromise accuracy.

3. Results

In this section we present some numerical results in order to validate our novel approach. In particular, in Section 3.1 we test our approach in an artificial scenario in which we aim at recovering the prescribed differential dynamics of the transmission rate which depends on an exogenous variable (temperature). In this context we also show the ability of the architecture to reconstruct hidden dynamics of the transmission rate in presence of data affected by noise. Additionally, we perform an extensive campaign of numerical simulations in order to assess the sensitivity and robustness of our learning procedure with respect to some network hyperparameters, to the width of the observation window during the estimation stage and to the increase in the amount of training samples. We assess the performance of our approach by comparing different cases based on the test error, measured on a fixed batch of previously unseen samples.

In Section 3.2 we consider a realistic scenario investigating on the differential model relating meteorological quantities, *i.e.* temperature and relative humidity, to the recovered transmission rate for the case of influenza spreading in Italy during 2010–2020, with waves of average length of 28 weeks. Moreover, we propose to study the equilibrium points of the reconstructed dynamical system associated with the transmission rate, considering different values of the input exogenous variables.

Table 1
Table of the parameters for generating synthetic data (Test case 1).

Parameter	Distribution	Interval	Unit of measurement
T_m	Uniform	[10, 15]	°C
A	Uniform	[5, 10]	°C
f	Constant	$\frac{1}{365}$	$[\frac{1}{\text{days}}]$
ϕ	Uniform	$[\frac{1}{6} \frac{\pi}{365}, \frac{3\pi}{365}]$	[-]
ϵ	Uniform	[0.1, 3]	[-]
β_0	Uniform	[0, 0.1]	$[\frac{1}{\text{days}}]$

3.1. Test case 1: Artificial scenario

3.1.1. Data generation

We generate a training dataset of temperatures taking seasonal variations into account. In particular, we consider each input time history belonging to the following family:

$$T^{(i)}(t) = T_m + A^{(i)} \sin(2\pi f t + \phi^{(i)}), \quad (15)$$

where T_m represents a medium value of temperature, $A^{(i)}$ is the amplitude of the seasonal effect, f is the seasonal wave frequency and $\phi^{(i)}$ its phase. The i th superscript indicates that the quantities are associated with the i th sample. Moreover, we assume that the transmission rate satisfies a logistic equation with fluctuations driven by temperature:

$$\dot{\beta}^{(i)}(t) = \left(\frac{T^{(i)}(t)}{T_m} - 1 \right) \left(\frac{\beta^{(i)}(t)^2}{\epsilon^{(i)} \beta_r} - \beta^{(i)}(t) \right), \quad (16)$$

where β_r is a scaling factor for the transmission rate, and $\epsilon^{(i)} > 0$ represents a suitable scaling factor embodying the specific infectivity of different epidemic waves. The differential law (16) has been retrieved starting from the modeling choices of [50] as explained in Appendix B (Supplementary Materials). We remark that during both training and testing phases we will estimate the latent parameter δ for each sample, that we expect to be linked to the lineage parameter ϵ – the only other factor driving our prescribed synthetic dynamics.

In order to retrieve the target values corresponding to each couple $\{(T^{(i)}(t), \epsilon^{(i)})\}_i$ we solve the *SEIR* model (6), by keeping fixed the parameters:

$$[S_0, E_0, I_0, R_0]^T = [0.97, 0.01, 0.02, 0]^T, \quad \alpha = 1/10, \quad \text{and} \quad \gamma = 1/20.$$

We simulate each phase until the final time $T = 364$, *i.e.* considering a time window of length one year. Unless otherwise specified, we deal with training datasets of size 50 and with a fixed test dataset of size 50. Both datasets are generated by randomly sampling the input parameters following the probability distributions in Table 1. Instead, the reference value for the transmission rate has been computed through the following algebraic expression:

$$\beta_r = \gamma \frac{\log\left(\frac{S_0}{S_\infty}\right)}{1 - S_\infty}, \quad (17)$$

which is derived from a standard *SEIR* model with constant coefficients by imposing a final size of Susceptible S_∞ . In our case we keep $S_\infty = 0.80$, therefore $\beta_r \approx 0.061$.

The hyperparameters associated with the optimization loops and relative weights of the different terms involved in the loss are reported in Appendix C (Supplementary Materials).

3.1.2. Results: Impact of uncertain data

In this section we provide numerical results dealing with noisy data, to test robustness of the proposed approach. This analysis has a crucial impact in scenarios embodying real infectious data, whose quality is typically affected by under-reporting or reporting delays, potentially causing misinformation to public health authorities [51]. Moreover, infections are difficult to be isolated during epidemic waves, and the proxies that we use, such as cases and deaths, often provide noisy approximations of the unknown real amount of infections.

To this aim, we train the model adding artificial noise in a multiplicative way to the target amount of new cases:

$$\widetilde{\Delta I}^{(i)}(\omega) = \widetilde{\Delta I}^{(i)}(\omega)(1 + u\kappa), \quad (18)$$

where $u > 0$ is the absolute entity of uncertainty, $\widetilde{\Delta I}^{(i)}(\omega)$ is the detected amount of new cases and $\kappa \sim \mathcal{N}(0, 1)$ is a standard Gaussian error. In this synthetic case, we retrieve each $\widetilde{\Delta I}^{(i)}(\omega)$ by numerically solving the *SEIR* model coupled with (16). According to (18), we assume higher probability of detection errors in the neighborhood of peak values, since, due to limited detection resources, it is often more difficult to detect the correct amount of infected when the illness has already largely spread. In Fig. 3, the evolution of

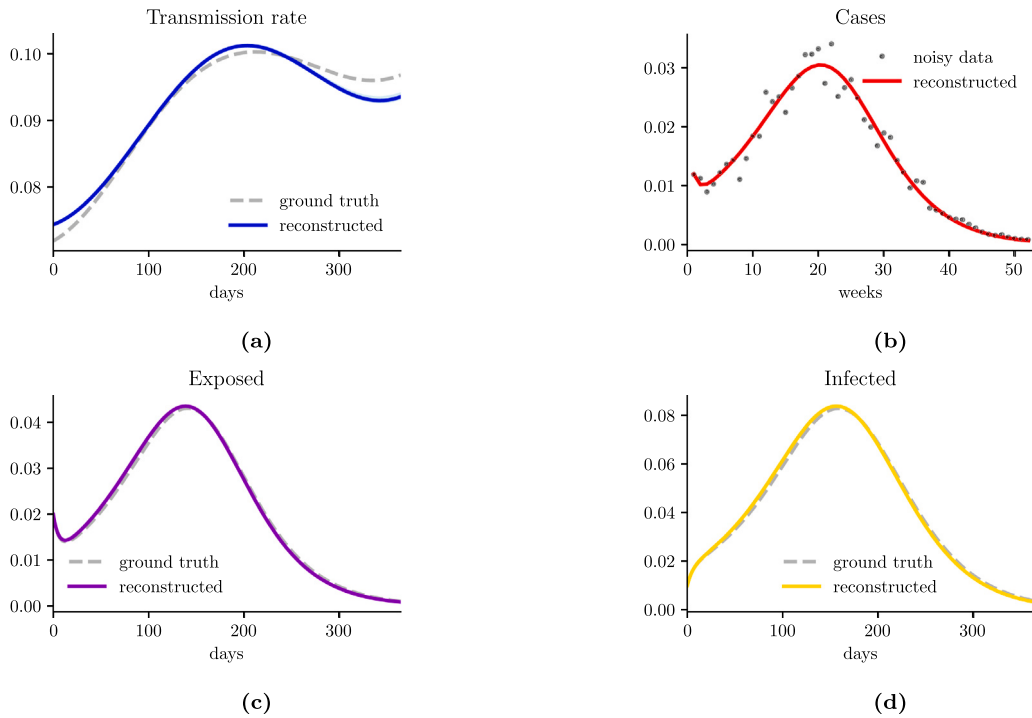


Fig. 3. Reconstructed training sample with uncertainty level $u = 0.1$. In particular we report the respective (a) Transmission Rate, (b) Weekly cases, (c) Exposed and (d) Infected, alongside the ground truth values, represented by gray dashed lines and dots.

the reconstructed compartments of infected and exposed and the amount of cases is presented for one simulation obtained with a random initializations of weights and biases (cf. Appendix C in the Supplementary Materials for more details on the initialization). The reconstructed transmission rate, as a function of temperature as in Eq. (15), has been recovered with a mean relative error in the whole reconstruction $e_r \approx 1.3\%$, even if the amount of cases are affected by uncertainty of module $u = 0.1$. With the reconstructed transmission rate, infected and exposed compartments overlap with the denoised original data.

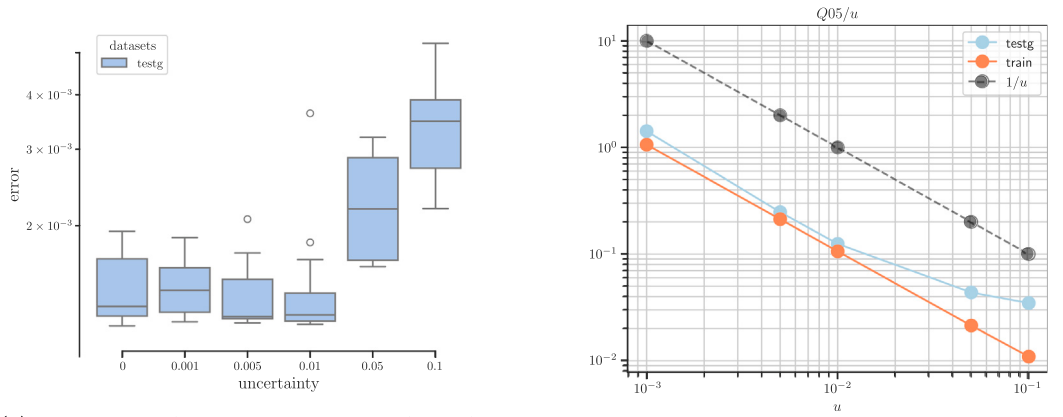
We then consider various levels of uncertainty

$$u \in \{0, 0.001, 0.005, 0.01, 0.05, 0.1\}, \quad (19)$$

and, after the training process, we evaluate the prediction error (10) in the test dataset specified in Section 3.1.1. The original denoised test set is altered in each case with the same level of uncertainty error of the training set: for obtaining each test dataset for a given level of uncertainty u , we start from the denoised 50 trajectories of the neat test dataset described in Section 3.1.1 and we apply (18) corresponding to the respective u .

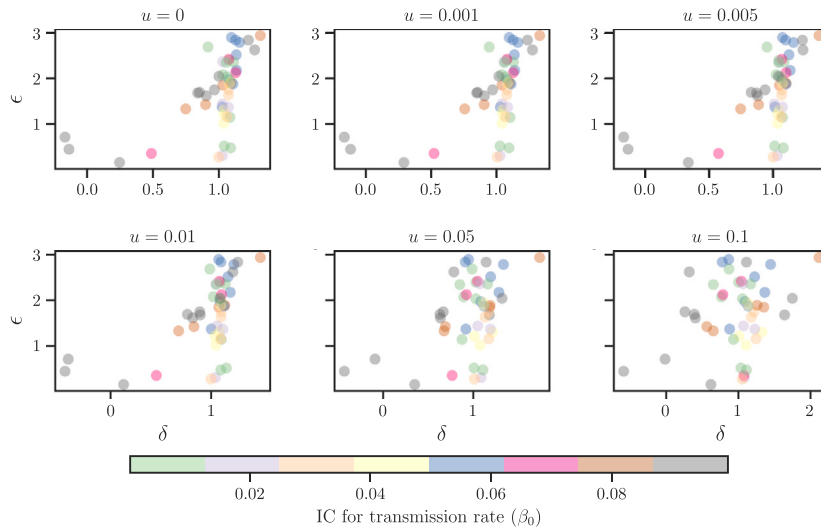
In Fig. 4(a) we considered 20 different runs with varying initializations of the trainable variables, presenting test errors in boxplot form for the difference between the reconstructed new infection cases and the denoised ground truth values. The IQR of the boxplots widens across the five cases as uncertainty increases, and the median error value gets higher. However, in Fig. 4(b) we observe that the median test error ($Q05$), when scaled by the average magnitude of uncertainty (u), decreases as uncertainty grows. Additionally, the median training error remains constant despite uncertainty rises.

In Fig. 4(c), we examine the latent parameter distribution on the test set ($\delta^{(p)}$). The x-axis represents the values of the lineage parameter ϵ used to generate the dataset through (16), while the y-axis shows the mean value of the reconstructed latent parameter across simulations with different initializations of the trainable variables. As noted in [45], the reconstruction value of the latent parameter through the data-assimilation technique introduced above is not unique, but for different target values the reconstructions should lie on a parametric curve. For similar values of $\beta_0^{(p)}$, the initial value for the transmission rate dynamics of each testing sample, a clear trend is expected. The network should increasingly recognize these trends as more training data becomes available and noise decreases, as it happens, e.g. for the black and brown latent parameters in the first row of Fig. 4(c). However, in some cases the parameter reconstructions for similar $\beta_0^{(p)}$ values present a cloudy shape, e.g. the green latent parameters in Fig. 4(c). Actually, these values correspond to declining epidemics with low initial transmission rates, when the epidemic wave does not outbreak, and it evolves independently on the transmission rate dynamics. In these contexts the curve $\delta - \epsilon$ is hardly identifiable. We also observe that increasing uncertainty in the target data impacts the latent parameters identification, causing their reconstructions to become more diffuse with respect to the lineage parameter as uncertainty increases within the same range of initial transmission rates. We remark that $\delta^{(p)}$ is allowed to take negative values, as it is not constrained to match the values of $\epsilon^{(p)}$. In general, when working



(a) Test errors with varying uncertainties, obtained from 20 different runs for each case considering different initializations of weights and biases.

(b) Median testing (blue), training (orange) errors over amplitude of uncertainty level.



(c) Lineage parameters generating the test set versus the mean value of the reconstructed latent parameters with different uncertainties. Different colors correspond to different values of the initial condition (IC) for the transmission rate dynamics according with the colorbar.

Fig. 4. Analysis of uncertainty.

with real data (as demonstrated in Test Case 2), $\delta^{(p)}$ accounts for various factors. Therefore, it should be interpreted as a label, where similar values indicate comparable waves, rather than in absolute value like the lineage parameter, which carries a physical interpretation.

In Fig. 5 we reported six samples, one per row, of transmission rates and the respective weekly amount of cases belonging to the test dataset for three different levels of uncertainty. The dot points correspond to noised observed data, whilst the dashed line is the unknown original (denoised) target trajectory. The uncertainty bands correspond to the 0.1–0.9 quantiles of the trajectories considering 20 different models trained starting with different random initializations of weights and biases. We note that the learned evolutive model for the transmission rate is able to capture the frequency and phase of the harmonic signal inferred by temperature. We observe that the transmission rate is coherent with the ground truth, even if, as one can expect, noisy measurements influence the reconstructed dynamics. After the epidemic peak is reached, the reconstructed transmission rates tend to be less accurate and to be more stationary than the target curve. This behavior is explained by the fact that, with the compartment of infected individuals nearly depleted, constant non-zero transmission rates have little effect on the epidemic, which is already in decline. As a result, the trained model tends to be more accurate during the early stages of the epidemic wave, less in the long-time horizons.

3.1.3. Results: Sensitivity analysis

We perform a sensitivity analysis with synthetic data in order to motivate our practical choices for the involved hyperparameters. For this purpose we consider the following hyperparameters and evaluate their impact in terms of testing and training error over

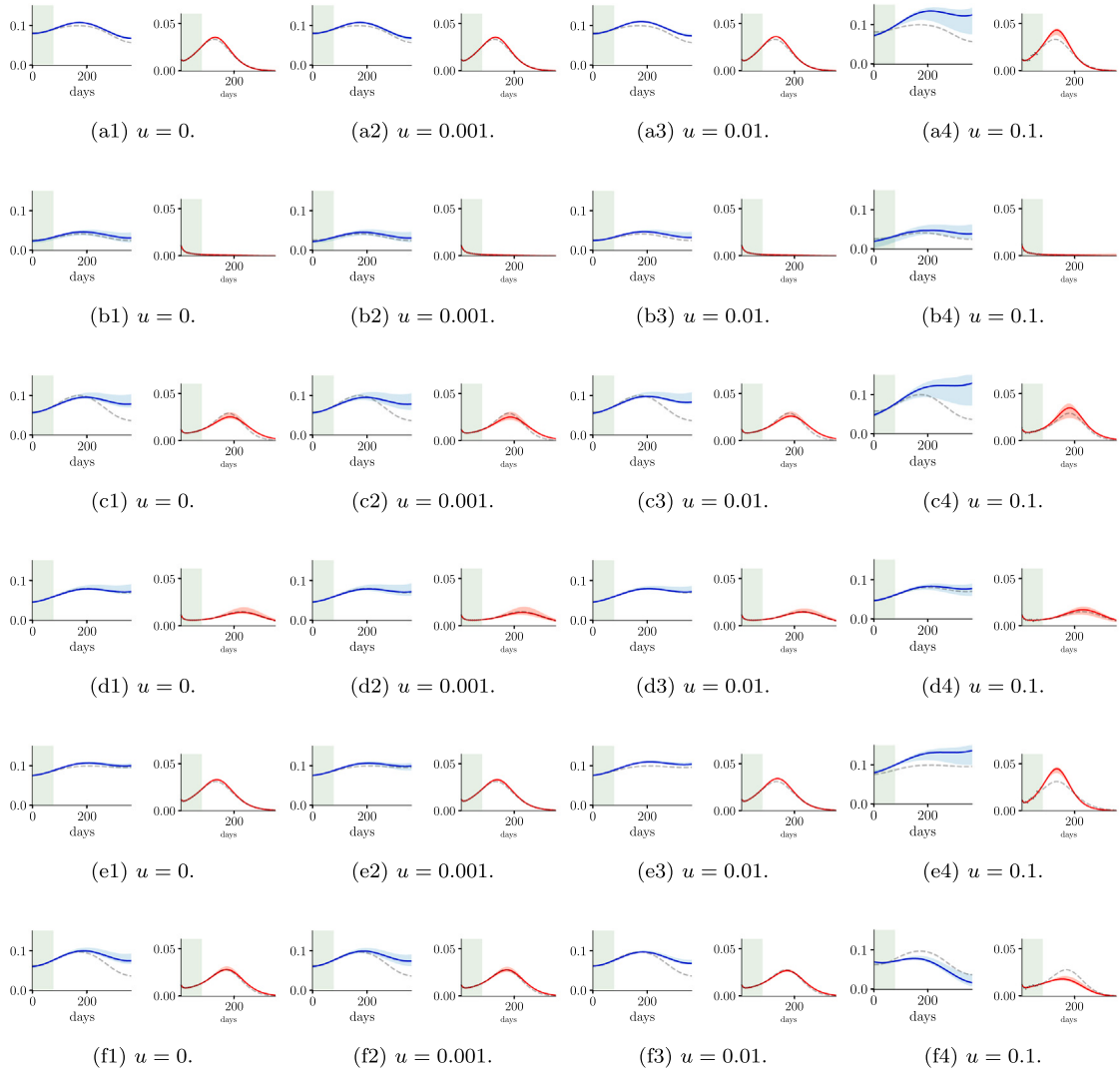


Fig. 5. Testing transmission rates (blue) and cases (red) with data with different uncertainties. Each row shows the same sample with growing uncertainty: $u \in \{0, 0.001, 0.01, 0.1\}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

20 different runs with random initializations of weights: (i) number of neurons of the neural network; (ii) number of layers of the neural network; (iii) number of training samples; (iv) observation window width (only on test error).

In Fig. 6(a) we gathered the results in terms of amount of neurons with a two-layers neural network. Following the *Occam's razor principle*, reducing the number of neurons, for example to 4, leads to higher median accuracy by preventing overfitting of the training trajectories. Similarly, increasing the model's nonlinearity by adding more layers does not contribute to a reduction in test error. Indeed, in Fig. 6(b) the median error of the two-layers neural network with 4 neurons is lower with respect to the other considered cases.

Furthermore, increasing the training size enhances the possibilities of achieving higher test accuracy, since the neural network has more samples to learn from. This behavior is evident from the median trends of test errors in Fig. 6(c), although the training process becomes more time consuming (see Table 2).

At last, we consider the impact of the width of the observation window for estimating the couples of parameters when a new trajectory has to be simulated. From Fig. 6(d) we deduce that prediction accuracy suffers when windows which are not large enough. On the other hand, after almost 11 weeks, corresponding to of one-fourth of the total simulation time, the prediction error tends to saturate, thus motivating our choice for the synthetic scenario. Similar conclusions can be drawn from Fig. 7, where we can note that the learnt latent parameters, corresponding to similar values of initial transmission rates, identify one-dimensional curves starting from an observation time $T_{\text{obs}} = 77$ days. This behavior is not observed with shorter time windows.

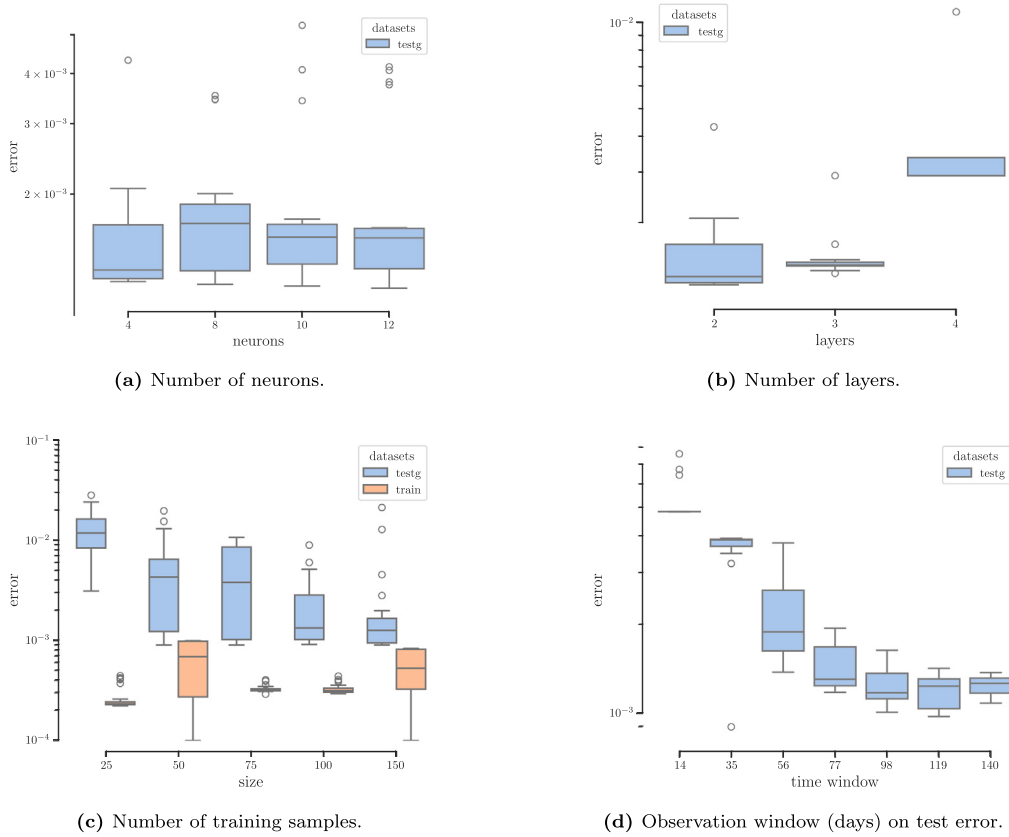


Fig. 6. Sensitivity analyses on hyperparameters indicated in the captions. Blue boxplots correspond to the test error, orange ones to the training error obtained with 20 different runs with different initializations of weights and biases.

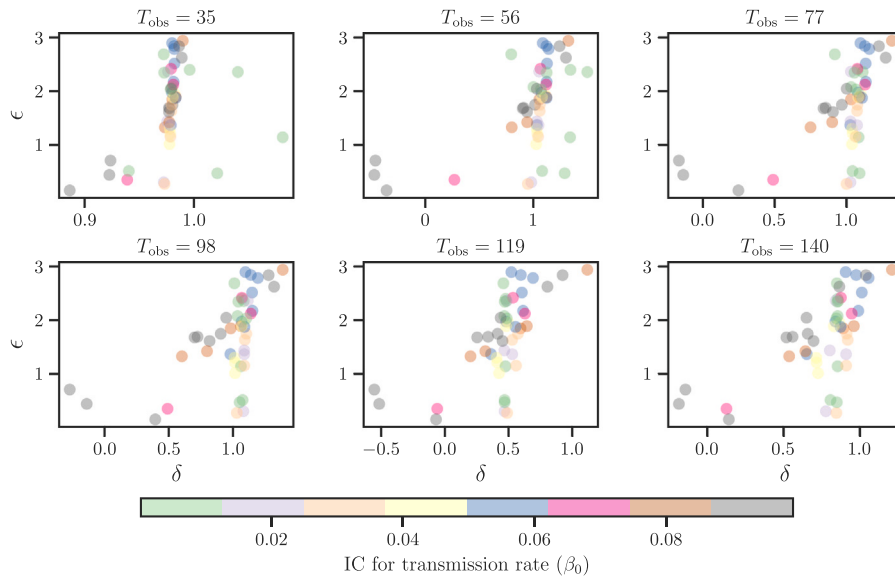


Fig. 7. Lineage parameter versus the reconstructed latent parameters varying the observation windows. Different colors correspond to different values of the initial condition (IC) for the transmission rate dynamics according with the colorbar. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Median over 20 runs of the computational times for training the architecture with different training sizes on a 8-core machine Intel i7.

Size	Median training time [s]	Q01 training time [s]	Q09 training time [s]
25	629.7	579.8	724.3
50	780.7	536.4	802.8
75	887.5	820.6	900.6
100	957.3	946.2	973.3
150	1229	1179	1235



Fig. 8. Repartition of Italy in 15 areas in order to retrieve weighted average national temperature and relative humidity. Indeed, only meteorological data from the largest cities were available. Each color corresponds to a specific area, which has a different weight depending on the population of the region. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Test case 2: Realistic scenario dealing with influenza waves

Seasonal influenza is a contagious, infectious disease of the respiratory tract, which is estimated to cause tens of millions of infections and linked respiratory illnesses every year, as well as nearly 250.000–500.000 deaths worldwide [52]. Early detection and reliable predictions of disease evolution, when followed by a rapid response, can drastically reduce consequences of both seasonal and pandemic influenza, and encourage prevention measures such as vaccinations.

To address these challenges, the following subsections present the application of the proposed novel architecture within this specific epidemic context.

3.2.1. Data generation

We exploit the data reported by the Italian epidemiological and virological influenza surveillance system from 2010–2011 to 2019–2020, available in the influnet open access Github repository.¹ Every season almost 2000 samples have been collected, with a proportion of positive specimens of around 34%. In [53] it was estimated the surveillance system average detects from 18.4% to 29.3% of actually infected by influenza. Therefore, we divided total new cases by a mean factor $u_{\text{inf}} = 0.23$, in order to be coherent with a *SEIR* mathematical model which does not take into account for undetected infected. Additionally, it was estimated that the average duration of the incubation period for influenza is 1.5 days ($\frac{1}{\alpha}$) [54] and the infectious period 1.2 days ($\frac{1}{\gamma}$), in order to have a total generation time of 2.7 days, in accordance with existing literature [55]. Concerning the initial conditions for the *SEIR* model, we assume as in [53] to prescribe a small seed corresponding to 10^{-6} initial percentage of infected individuals, and

$$E_0 = \frac{\overline{\Delta I(1)}}{7\alpha}, \quad (20)$$

i.e. considering a plausible value of initial exposed individuals in order to approximately generate in 7 days the initial amount of new cases.

¹ <https://github.com/fbranda/influnet>

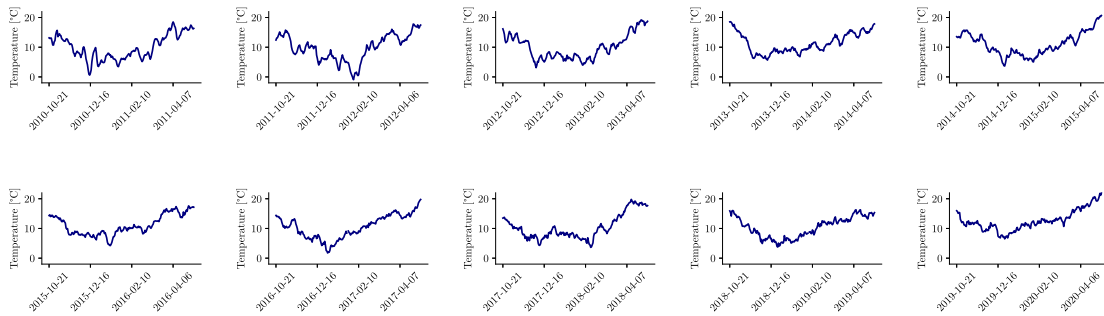


Fig. 9. Aggregated national data of temperature used as exogenous quantity in Test case 2.

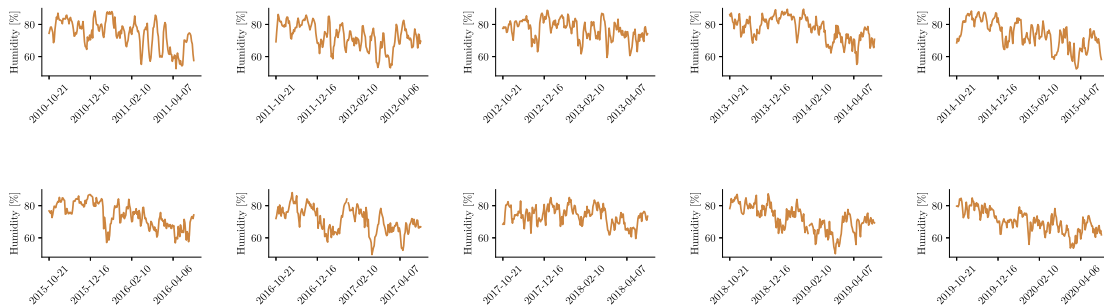


Fig. 10. Aggregated national data of relative humidity used as exogenous quantity in Test case 2.

In this Test case, we take into account meteorological data, *i.e.* temperature and relative humidity, as exogenous quantities, following the analysis of [56,57], which studied the coupled interplay between relative humidity and temperature in the spreading of infectious diseases. These data are available only for some cities in different administrative regions. Therefore, in order to obtain the national data, they have been averaged after their extraction from a popular Italian meteorological website² as detailed below.

We dispose of meteorological data coming from 15 different cities belonging to different administrative Italian regions. The remaining 5 regions (Abruzzo, Basilicata, Molise, Valle d'Aosta, Umbria), where meteorological data were not available for more than four of the ten considered years, have been paired with neighboring regions of similar latitude (cf. Fig. 8). Then, the national averages for temperature and humidity were then calculated using a weighted mean, where each area's weight was determined by its relative population size.

Furthermore, the average data are quite noisy, as illustrated in Figs. 9 and 10. In order to filter out higher frequencies, we apply to both temperature and relative humidity time series a Savitzky–Golay filter of order 2 [58]. In this way we retrieve in each point a smoother least square approximation of these data which is more prone for training and testing the proposed approach (see Fig. 11).

Unlike other pandemic events over the past five centuries that occurred unexpectedly, influenza exhibits a certain cyclicity attributed to various genetic mechanisms responsible for triggering new outbreak waves. The causative agent of influenza is not seen as a single, stable strain persisting over the years, but rather as a heterogeneous collection of viral evolutionary events that, while highly similar, are not identical [59]. This justifies the need to estimate the parameter δ during training and testing stages. Differences in transmissibility, mortality rates, and severity can be identified for previous influenza waves, largely due to genetic heterogeneity, which has been extensively studied [60]. Influenza circulating in Italy during the last decade is mainly constituted by three strains: influenza A/H1N1pdm09, influenza B and influenza A/H3N2. The reproduction number for A/H1N1pdm09 is similar to that of seasonal influenza, even though this virus is highly transmissible, leading to a global pandemic in 2009. Its spread has been largely boosted by lacking in pre-existing immunity in the population. The A/H3N2 subtype often dominates during epidemic seasons, causing more frequent and intense outbreaks, in particular in older adults. It has undergone significant antigenic drifts over the years, leading to more severe symptoms with respect to those of seasonal flu. Instead, influenza B is generally less transmissible with respect to A wildtypes, affecting predominantly children and young adults. However, it can lead to severe outcomes in older adults, especially in presence of comorbidities.

We aim at catching differences across waves of different seasons with the use of the latent parameter δ , which will be confronted with the average transmissibility of the virus' lineages circulating across the years as studied by [53]. From this study inferring

² <https://www.ilmeteo.it/portale/archivio-meteo>

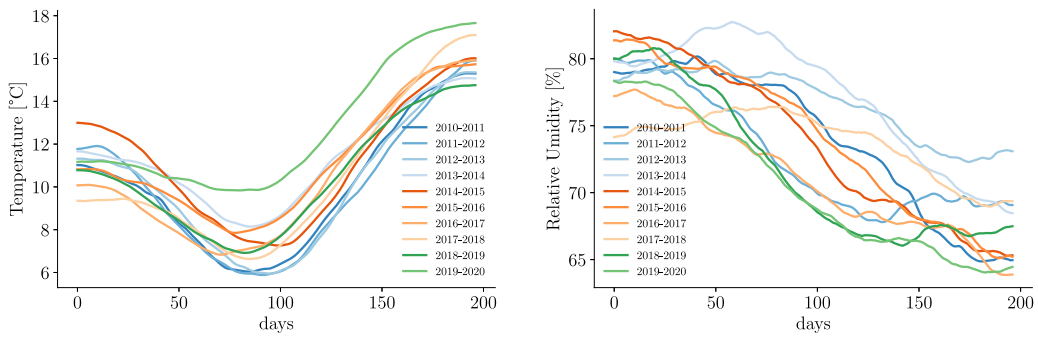


Fig. 11. Input time series of national temperature and relative humidity after applying the Savitzky-Golay filter.

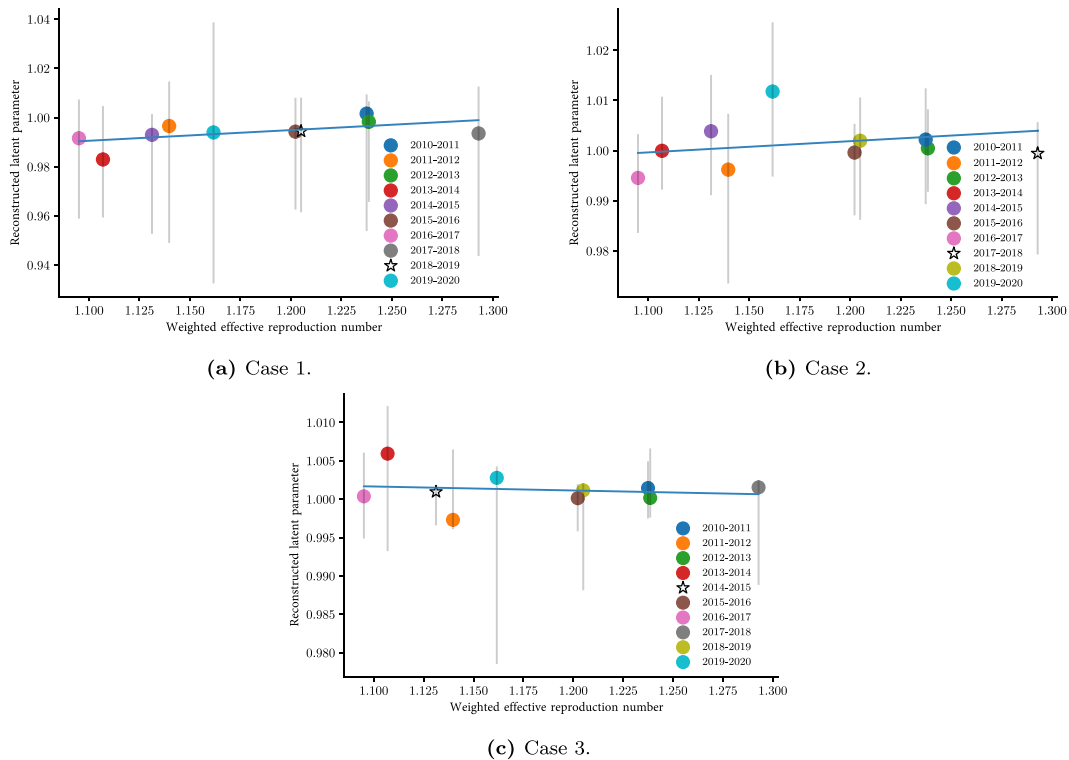


Fig. 12. Effective reproduction number of influenza waves weighted by lineage composition retrieved starting from [53] (x-axis) versus reconstructed latent parameters δ (y-axis). In each Figure, star-shaped point is associated with the testing sample. The blue line corresponds to the linear regression of training median values. Each gray bar corresponds to 0.1–0.9 percentile values of the respective reconstructed parameter (considering various random initializations). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

on effective reproduction numbers of the different strains in each season, estimated from the serological analyses of ILI cases in Lombardy during each year for determining strain-compositions, we derive a mean effective reproduction number that we compare with the reconstructed δ (see Fig. 12).

We explicitly incorporate temperature and humidity into the model for the transmission rate, while other exogenous factors are implicitly included in factor δ .

3.2.2. Results

Among the 10 available influenza waves, we randomly select 9 samples for the training set and 1 for the testing set (leave-one-out approach); then, we run 20 different simulations with the same network topology described in Appendix D (Supplementary Materials), with different initializations of weights and biases. In the same section, the list of hyperparameters associated are reported. We present three representative cases of different scenarios by considering varying training and testing data:

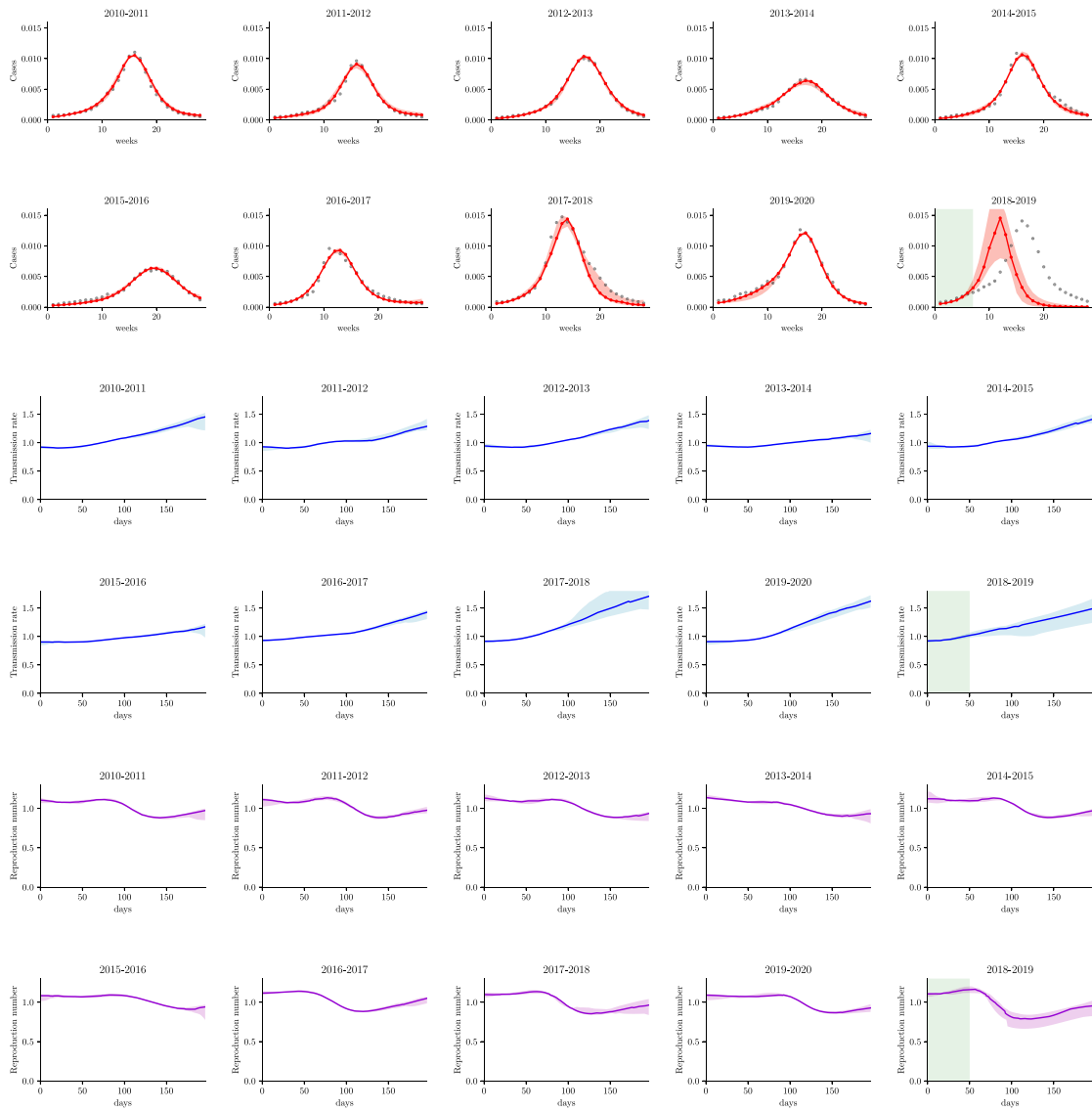


Fig. 13. Case 1. Reconstructions of cases, transmission rate, reproduction number exploiting influenza data. The training set is constituted by all epidemic waves from 2010–2011 to 2019–2020 except for 2018–2019, which belongs to the test set. The green window corresponds to the 49 days of observability in which we estimate the latent parameter for the testing sample. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- Case 1: testing wave corresponds to 2018–2019;
- Case 2: testing wave corresponds to 2017–2018;
- Case 3: testing wave corresponds to 2014–2015.

In Fig. 12 we collect the mean retrieved values for the median latent parameter δ for each influenza wave in the three cases, together with 0.1–0.9 percentile bands. Those quantities have been plotted against the weighted effective reproduction number for the different influenza waves, which depends on the dominant variants' composition as obtained in [53]. In all three cases, the linear correlation coefficients are, in absolute value, less than 0.5, indicating that the recovered unknown latent parameter cannot be linked to the weighted effective reproduction number through a linear transformation. In general, when working with realistic data, the latent parameter is expected to be linked to virological and epidemiological characteristics. These characteristics often vary across distinct strains and also across different waves of the same strain. For example, the latent parameter may be influenced by variations in incubation time, infectious period, attack rate, mutation rate, pathogen shedding factor, case fatality rate (CFR), infection fatality rate (IFR), herd immunity threshold, and contact rate [36]. A key challenge in epidemiology lies in disentangling and isolating the differences in these parameters, as they reflect distinct effects of the disease and have varying impacts on the

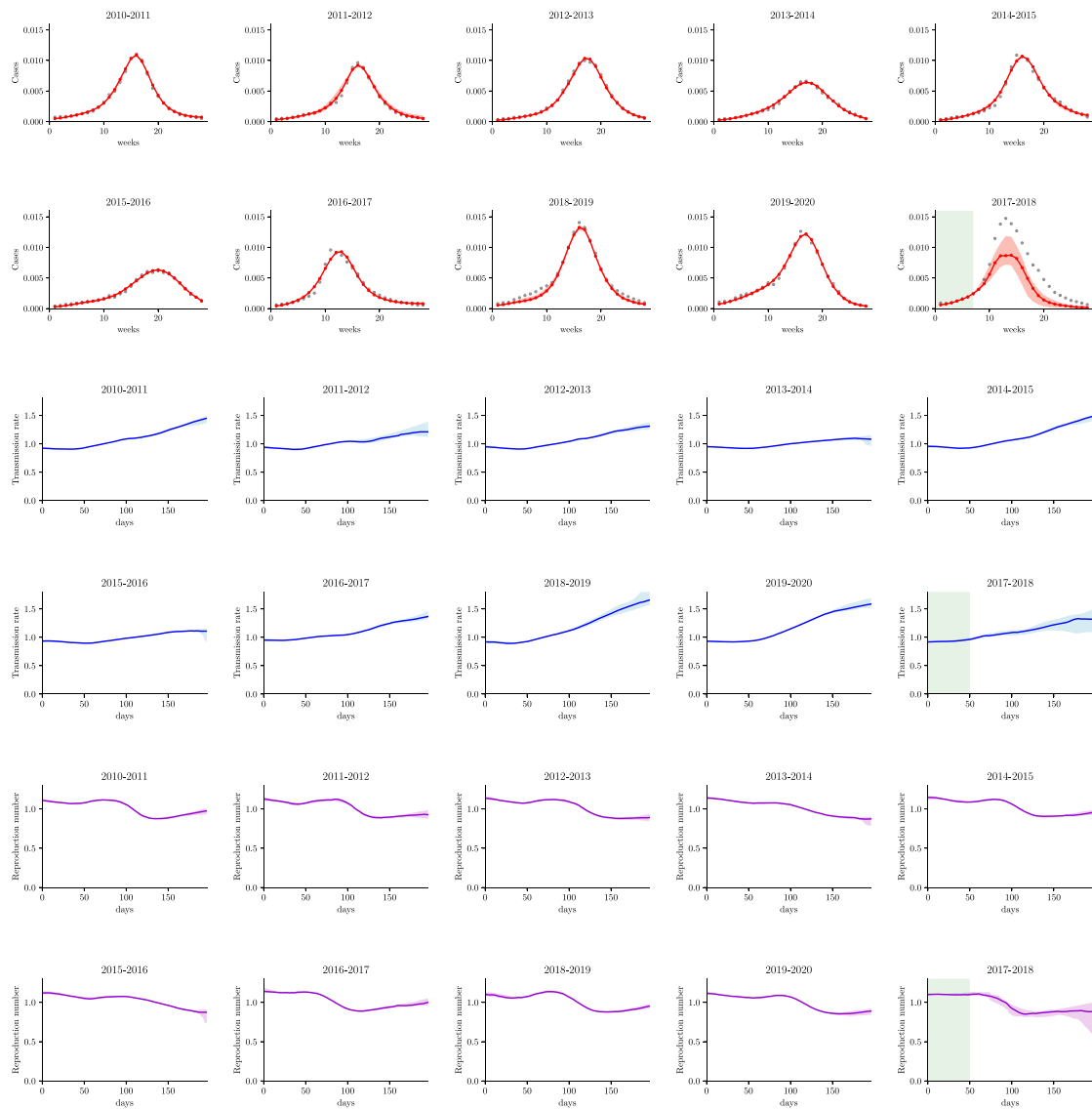


Fig. 14. Case 2. Reconstructions of cases, transmission rate, reproduction number exploiting influenza data. The training set is constituted by all epidemic waves from 2010–2011 to 2019–2020 except for 2017–2018, which belongs to the test set. The green window corresponds to the 49 days of observability in which we estimate the latent parameter for the testing sample. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

transmission mechanism. Consequently, it is not surprising that latent parameters represent an averaged effect of these quantities, along with other exogenous factors, and cannot be ascribed to one single underlying variable. We conclude that the evolution of transmission rate cannot be solely attributed to the intrinsic reproducibility of the current strain; rather, other exogenous factors (whose effects are embedded in the latent parameters) significantly influence each single wave. However, since the median values of δ for the same wave remain similar across the three different cases, training is robust when using the leave-one-out approach. The clustering of δ values near 1 is also associated with a relatively high weight of the corresponding regularization term in the training and testing loss.

Figs. 13–15 represent the new cases, transmission rate and reproducing number for both training and testing samples in the three cases respectively. The number of new cases obtained on the training samples is always caught in terms of absolute values and, furthermore, the time at which the peak happens coincide with the ground truth. In Case 1 the absolute value of the peak for the testing wave is retrieved, even though the peak is reached almost three weeks in advance. In this case increasing the time-frame of observability (green region in the figure) for the estimation phase would help in making predictions more adherent with the amount of cases actually counted in 2018–2019 (see Fig. 16). Besides, in Case 2 the time at which the peak occurs is accurately predicted, although the median amount of cases is underestimated with respect of the ground truth. Indeed, the attained values at

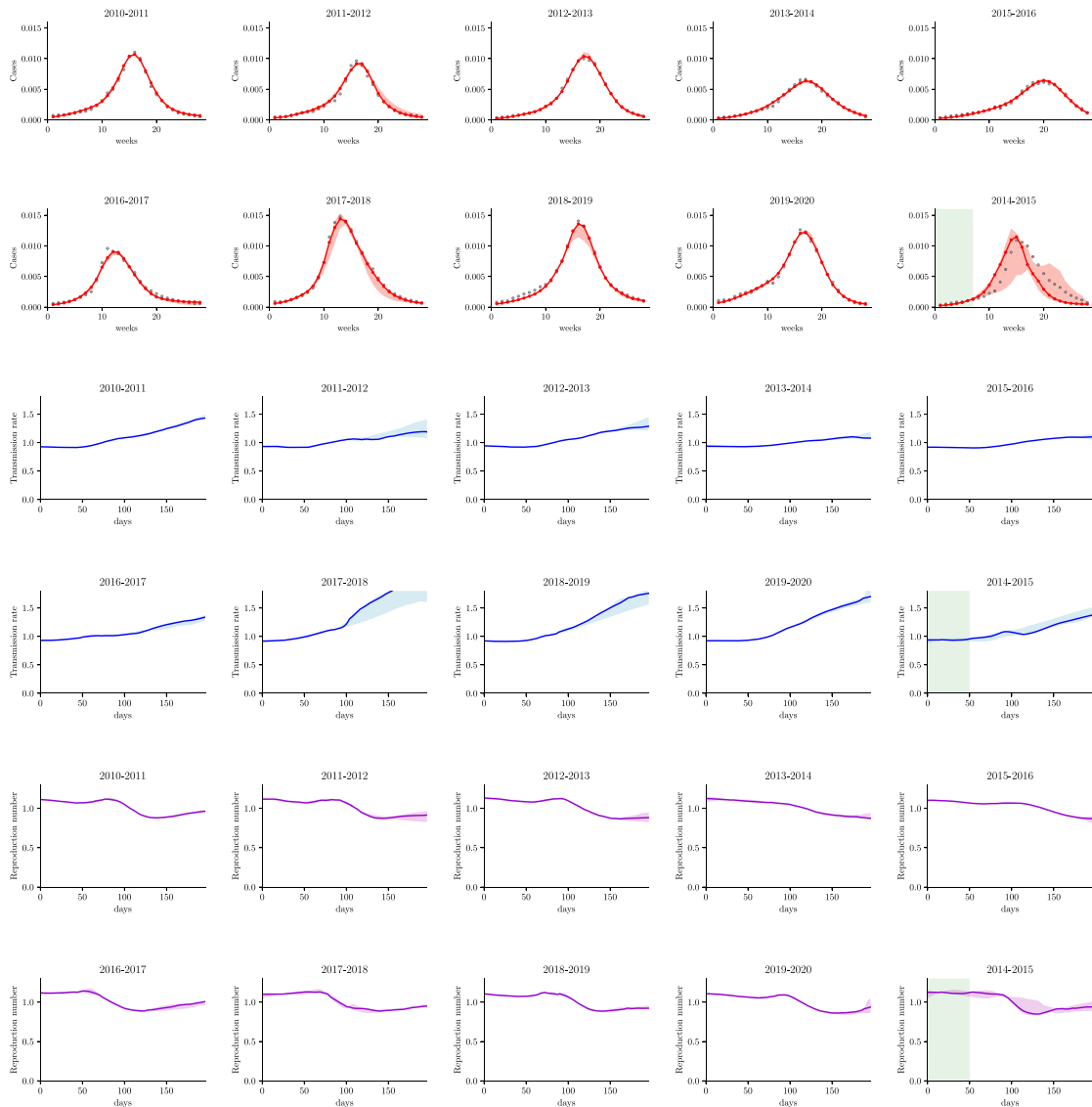


Fig. 15. Case 3. Reconstructions of cases, transmission rate, reproduction number exploiting influenza data. The training set is constituted by all epidemic waves from 2010–2011 to 2019–2020 except for 2014–2015, which belongs to the test set. The green window corresponds to the 49 days of observability in which we estimate the latent parameter for the testing sample. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the peak for this epidemic wave are higher with respect to the other training waves. Peak values this high have not been learnt by the model and, hence, they seem to be hardly predictable. Finally, the median trajectory of new cases in Case 3 successfully catches the behavior of real data for the epidemic wave of 2014–2015. The difficulties of the proposed architecture during the estimation and, consequently, during the testing phase can be ascribed to the limited amount of data coming from these epidemic waves, combined with high variability and similarities of input data as one can deduce from Fig. 11. The latter impact on the prior used for determining the new δ parameter during the estimation phase.

From the transmission rate reconstructions, for which we do not dispose of a ground truth to compare our results with, we can infer some qualitative dependencies on the input variables: in accordance with literature [56,57] we observe an inverse growth of the transmission rate with respect to relative humidity (see, e.g. [26] where it emerges that drier environments enhance transmission spread, similarly to air pollution [61]). On the other hand, the transmission rate on temperature is not obvious in our case, since temperature's time series correspond to the same seasonal periods in different waves, and therefore the input functions can hardly be distinguished. It is worth remarking that in our results peaks always happen during the colder months of wintertime. Finally, the reconstructed transmission rate continues to grow in the long-term horizon, even the outbreak is declining. This growing trend in the transmission rate was also observed in *Test case 1* (cf. Fig. 5). Note that no explicit relationship between the behavior (whether

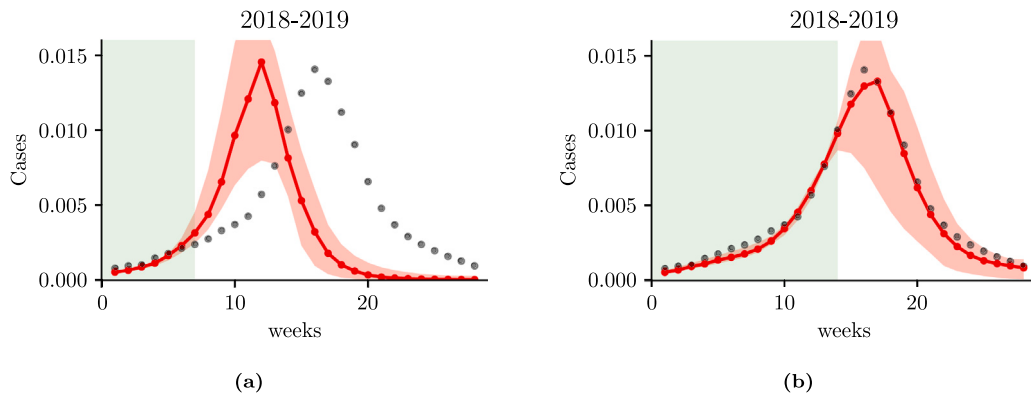


Fig. 16. Case 1. Reconstruction of new cases of the wave corresponding to 2018–2019 (test set) with two widths of the observation window (green areas), respectively (a) $T_{\text{obs}} = 49$ days and (b) $T_{\text{obs}} = 98$ days. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

increasing or decreasing) of the transmission rate and the behavior of the infected population is imposed in the loss function. However, we could explicitly enforce a long-term decline in the transmission rate within the loss function when the epidemic wave reaches its tail to prevent any unexpected upward trends.

The estimated reproduction number R_t always lies in the confidence interval estimated by [53] for each influenza epidemic. After the amount of new cases has reached its peak, the reproduction number always registers a sudden drop under the bifurcation value $R_t = 1$, indicating that the epidemic is proceeding to the equilibrium depleting infectious. Larger uncertainty bands characterize long-time behavior for both transmission rate and, consequently, reproduction number.

To gain more insight about the interpretability of our learned model, we analyze the phase-plane of the reconstructed transmission rate model. Specifically, we focus on the median model of Case 3, and we analyze the value of the surrogate right-hand side (f_{nn} , cf. (5)) at different values of the transmission rate. In this case, we keep two of the three input parameters (T , U , δ) cyclically fixed, while the third varies generating a different curve in the phase plane (β , f_{nn}), cf. Equation (B.5) in the Supplementary Materials. In this way we explore how the unstable equilibrium point behaves in relation to different constant values of this latter parameter (see Fig. 17). Hence, as humidity decreases, the unstable equilibrium shifts to higher values, and a similar effect is observed when temperature increases (cf Section 3.2.1). This behavior aligns with the literature (cf. [56,57]), as the range of values where the transmission rate has negative derivative expands with rising temperatures and decreasing humidity. In contrast, the latent parameter, which differentiates between waves, shifts the equilibrium point further as its value decreases.

4. Conclusions

In this work, we propose an innovative scientific machine learning method to infer the unknown differential evolution governing the transmission rate during epidemics, depending on exogenous parameters, which are known to influence transmission mechanisms. By learning a neural dynamical model of the transmission rate, we aim to address the critical issue of extrapolating this parameter's values in order to make reliable forecasts beyond the observation interval. Our framework incorporates a data assimilation approach to estimate, online, a latent parameter that distinguishes different waves of the same disease. This latent parameter embodies the averaged effects of external factors that are not considered explicitly in the chosen set of exogenous factors, but still influence the transmission rate.

We present an extensive overview of numerical results for two distinct test cases. In a synthetic scenario, we assess performance in terms of test error and the ability to recover the prescribed hidden dynamics. The results prove that the proposed architecture achieves good accuracy, even when handling noisy data. Additionally, we investigate the recovered model for transmission rates influenced by temperature and relative humidity, using real data from influenza waves in Italy between 2010 and 2020. Despite data scarcity, the leave-one-out results indicate accurate peak predictions. Furthermore, the results show that this tool is effective especially for short-term forecasts and, in some cases, successfully captures the entire yearly dynamics of new cases. We explore the interpretability of the reconstructed model and confirm that it aligns with literature findings on the influence of increasing temperature and relative humidity on transmission rates. However, as with many other computational frameworks, we remark that long-term predictions are less reliable and can sometimes be inaccurate. Notably, it is impossible to obtain direct observations of the transmission rate's evolution, and the epidemic data used for calibration are generally less affected by monitoring uncertainties near the infection peak compared to later stages of the outbreak. Moreover, in our context this limitation arises from the few data available for training and the modest set of exogenous variables considered. Extending this work to include additional time series is straightforward and will be the focus of future studies.

Possible future developments of this work include expanding the model to incorporate a broader set of key exogenous variables, such as non-pharmaceutical interventions, vaccinations, and immunity profiles. Additionally, the data assimilation approach can be

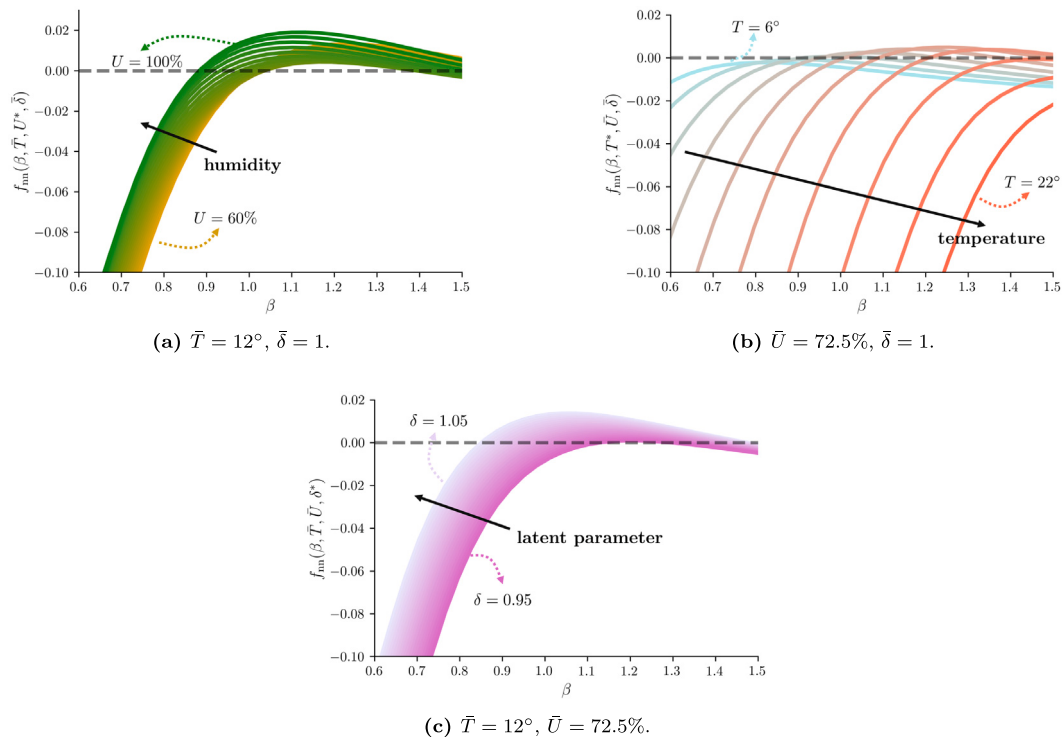


Fig. 17. Analysis of the reconstructed model for the transmission rate: plot of the right-hand side of the transmission rate (f_m) in terms of β where two input exogenous parameters are cyclically kept fixed, while the third varies. Each curve corresponds to different values of this latter varying parameter: relative humidity in (a), temperature in (b) and the reconstructed latent parameter in (c).

enhanced by using real-time moving windows to better estimate the latent parameter as more data becomes available. Finally, the initial conditions of the epidemic, which were artificially fixed in both cases, could be learned online using the same approach. This latter extension necessitates further investigation from a numerical standpoint.

We emphasize that this approach is general and can be extended to other models aimed at discovering the hidden dynamics of key parameters, even beyond the field of epidemiology.

CRediT authorship contribution statement

Giovanni Ziarelli: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stefano Pagani:** Writing – review & editing, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Nicola Parolini:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Francesco Regazzoni:** Writing – review & editing, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Marco Verani:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

NP, SP, FR, MV and GZ are members of INdAM-GNCS. The present research is part of the activities of “Dipartimento di Eccellenza 2023–2027”, MUR, Italy. FR and SP have received support from the project PRIN2022, MUR, Italy, 2023–2025, P2022N5ZNP “SIDDMS: shape-informed data-driven models for parametrized PDEs, with application to computational cardiology”, funded by the European Union (Next Generation EU, Mission 4 Component 2).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.cma.2025.117796>.

Data availability

The computational framework is available on GitHub together with the numerical results of this work: <https://github.com/giovanniziarelli/inferringTRdynamicsML/>.

References

- [1] J.C. Lemaître, D. Pasetto, M. Zanon, E. Bertuzzo, L. Mari, S. Miccoli, R. Casagrandi, M. Gatto, A. Rinaldo, Optimal control of the spatial allocation of COVID-19 vaccines: Italy as a case study, *PLoS Comput. Biol.* 18 (7) (2022) e1010237.
- [2] L. Souto Ferreira, O. Canton, R.L.P. Da Silva, S. Poloni, V. Sudbrack, M.E. Borges, C. Franco, F.M.D. Marquitti, J.C. de Moraes, M.A.d.M. Veras, et al., Assessing the best time interval between doses in a two-dose vaccination regimen to reduce the number of deaths in an ongoing epidemic of SARS-CoV-2, *PLoS Comput. Biol.* 18 (3) (2022) e1009978.
- [3] G. Ziarelli, L. Dedè, N. Parolini, M. Verani, A. Quarteroni, Optimized numerical solutions of SIRDVW multiage model controlling SARS-CoV-2 vaccine roll out: An application to the Italian scenario, *Infect. Dis. Model.* 8 (3) (2023) 672–703.
- [4] G. Giordano, M. Colaneri, A. Di Filippo, F. Blanchini, P. Bolzern, G. De Nicolao, P. Sacchi, P. Colaneri, R. Bruno, Modeling vaccination rollouts, SARS-CoV-2 variants and the requirement for non-pharmaceutical interventions in Italy, *Nature Med.* 27 (6) (2021) 993–998.
- [5] Q. Richard, S. Alizon, M. Choisy, M.T. Sofonea, R. Djidjou-Demasse, Age-structured non-pharmaceutical interventions for optimal control of COVID-19 epidemic, *PLoS Comput. Biol.* 17 (3) (2021) e1008776.
- [6] R. Hinch, W.J. Probert, A. Nurtay, M. Kendall, C. Wymant, M. Hall, K. Lythgoe, A. Bulas Cruz, L. Zhao, A. Stewart, et al., OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing, *PLoS Comput. Biol.* 17 (7) (2021) e1009146.
- [7] D. Pasetto, F. Finger, A. Camacho, F. Grandesso, S. Cohuet, J.C. Lemaître, A.S. Azman, F.J. Luquero, E. Bertuzzo, A. Rinaldo, Near real-time forecasting for cholera decision making in Haiti after Hurricane Matthew, *PLoS Comput. Biol.* 14 (5) (2018) e1006127.
- [8] N. Parolini, L. Dedè, P.F. Antonietti, G. Ardenghi, A. Manzoni, E. Miglio, A. Pugliese, M. Verani, A. Quarteroni, SUIHTER: A new mathematical model for COVID-19. Application to the analysis of the second epidemic outbreak in Italy, *Proc. R. Soc. A* 477 (2253) (2021) 20210027.
- [9] V. Marziano, G. Guzzetta, A. Mammone, F. Riccardo, P. Poletti, F. Trentini, M. Manica, A. Siddu, A. Bella, P. Stefanelli, et al., The effect of COVID-19 vaccination in Italy and perspectives for living with the virus, *Nat. Commun.* 12 (1) (2021) 7272.
- [10] N. Parolini, L. Dedè, G. Ardenghi, A. Quarteroni, Modelling the COVID-19 epidemic and the vaccination campaign in Italy by the SUIHTER model, *Infect. Dis. Model.* 7 (2) (2022) 45–63.
- [11] K. Linka, M. Peirlinck, F. Sahli Costabal, E. Kuhl, Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions, *Comput. Methods Biomech. Biomed. Eng.* 23 (11) (2020) 710–717.
- [12] E. Bertuzzo, L. Mari, D. Pasetto, S. Miccoli, R. Casagrandi, M. Gatto, A. Rinaldo, The geography of COVID-19 spread in Italy and implications for the relaxation of confinement measures, *Nat. Commun.* 11 (1) (2020) 42–64.
- [13] N. Hoertel, M. Blachier, C. Blanco, M. Olsson, M. Massetti, M.S. Rico, F. Limosin, H. Leleu, A stochastic agent-based model of the SARS-CoV-2 epidemic in France, *Nature Med.* 26 (9) (2020) 1417–1421.
- [14] L.L. Lima, A.P. Atman, Impact of mobility restriction in COVID-19 superspreading events using agent-based model, *Plos One* 16 (3) (2021) e0248708.
- [15] C.C. Kerr, R.M. Stuart, D. Mistry, R.G. Abeyesuriya, K. Rosenfeld, G.R. Hart, R.C. Núñez, J.A. Cohen, P. Selvaraj, B. Hagedorn, et al., Covasim: an agent-based model of COVID-19 dynamics and interventions, *PLoS Comput. Biol.* 17 (7) (2021) e1009149.
- [16] J. Lasser, J. Sorger, L. Richter, S. Thurner, D. Schmid, P. Klimek, Assessing the impact of SARS-CoV-2 prevention measures in Austrian schools using agent-based simulations and cluster tracing data, *Nat. Commun.* 13 (1) (2022) 554.
- [17] K.D. Olumoyin, A.M. Khaliq, K.M. Furati, Data-driven deep-learning algorithm for asymptomatic COVID-19 model with varying mitigation measures and transmission rate, *Epidemiologia* 2 (4) (2021) 471–489.
- [18] S. Shaier, M. Raissi, P. Seshaiyer, Data-driven approaches for predicting spread of infectious diseases through DINNs: Disease informed neural networks, *Lett. Biomath.* 9 (1) (2022) 71–105.
- [19] 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 (10) (2022) 1949–1985.
- [20] K. Sherratt, H. Gruson, H. Johnson, R. Niehus, B. Prasse, F. Sandmann, J. Deuschel, D. Wolfram, S. Abbott, A. Ullrich, et al., Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations, *Elife* 12 (2023) e81916.
- [21] S. Fiandrino, A. Bizzotto, G. Guzzetta, S. Merler, F. Baldo, E. Valdano, A. Mateo-Urdiales, A. Bella, F. Celino, L. Zino, et al., Collaborative forecasting of influenza-like illness in Italy: the infLucast experience, 2024, *MedRxiv* 2024-2009.
- [22] C. Millevoi, D. Pasetto, M. Ferronato, A physics-informed neural network approach for compartmental epidemiological models, *PLoS Comput. Biol.* 20 (9) (2024) e1012387.
- [23] M. He, S. Tang, Y. Xiao, Combining the dynamic model and deep neural networks to identify the intensity of interventions during COVID-19 pandemic, *PLoS Comput. Biol.* 19 (10) (2023) e1011535.
- [24] M. Jing, K.Y. Ng, B. Mac Namee, P. Biglarbeigi, R. Brisk, R. Bond, D. Finlay, J. McLaughlin, COVID-19 modelling by time-varying transmission rate associated with mobility trend of driving via Apple Maps, *J. Biomed. Informatics* 122 (2021) 103905.
- [25] G. Chowell, C. Viboud, L. Simonsen, S.M. Moghadas, Characterizing the reproduction number of epidemics with early subexponential growth dynamics, *J. R. Soc. Interface* 13 (123) (2016) 20160659.
- [26] P. Mecenias, R.T.d.R.M. Bastos, A.C.R. Vallinoto, D. Normando, Effects of temperature and humidity on the spread of COVID-19: A systematic review, *PLoS One* 15 (9) (2020) e0238339.
- [27] M. Chinazzi, J.T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, et al., The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, *Science* 368 (6489) (2020) 395–400.
- [28] C.M. Hazelbag, J. Dushoff, E.M. Dominic, Z.E. Mthombathi, W. Delva, Calibration of individual-based models to epidemiological data: A systematic review, *PLoS Comput. Biol.* 16 (5) (2020) e1007893.
- [29] J.P. Gleeson, T. Brendan Murphy, J.D. O'Brien, N. Friel, N. Bargary, D.J. O'Sullivan, Calibrating COVID-19 susceptible-exposed-infected-removed models with time-varying effective contact rates, *Phil. Trans. R. Soc. A* 380 (2214) (2022) 20210120.
- [30] J. Sainte-Marie, D. Chapelle, R. Cimrman, M. Sorine, Modeling and estimation of the cardiac electromechanical activity, *Comput. Struct.* 84 (28) (2006) 1743–1759.
- [31] F. Regazzoni, L. Dedè, A. Quarteroni, Biophysically detailed mathematical models of multiscale cardiac active mechanics, *PLoS Comput. Biol.* 16 (10) (2020) e1008294.
- [32] B. Sansó, C. Forest, Statistical calibration of climate system properties, *J. R. Stat. Soc. Ser. C. Appl. Stat.* 58 (4) (2009) 485–503.
- [33] P. Saiteja, B. Ashok, A.S. Wagh, M.E. Farrag, Critical review on optimal regenerative braking control system architecture, calibration parameters and development challenges for EVs, *Int. J. Energy Res.* 46 (14) (2022) 20146–20179.

- [34] F. Regazzoni, L. Dedè, A. Quarteroni, Machine learning for fast and reliable solution of time-dependent differential equations, *J. Comput. Phys.* 397 (2019) 108852.
- [35] A.C. Lowen, J. Steel, Roles of humidity and temperature in shaping influenza seasonality, *J. Virol.* 88 (14) (2014) 7692–7695.
- [36] E. Kuhl, *Computational Epidemiology*, Springer, 2021.
- [37] M. Martcheva, *An Introduction to Mathematical Epidemiology*, vol. 61, Springer, 2015.
- [38] F. Regazzoni, S. Pagani, M. Salvador, L. Dedè, A. Quarteroni, Learning the intrinsic dynamics of spatio-temporal processes through latent dynamics networks, *Nat. Commun.* 15 (1) (2024) 1834.
- [39] V. Dukic, H.F. Lopes, N.G. Polson, Tracking epidemics with Google flu trends data and a state-space SEIR model, *J. Amer. Statist. Assoc.* 107 (500) (2012) 1410–1426.
- [40] P.E. Lekone, B.F. Finkenstädt, Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study, *Biometrics* 62 (4) (2006) 1170–1177.
- [41] B. Buonomo, A simple analysis of vaccination strategies for rubella, *Math. Biosci. Eng.* 8 (3) (2011) 677–687.
- [42] S. Kumar, R. Kumar, M.S. Osman, B. Samet, A wavelet based numerical scheme for fractional order SEIR epidemic of measles by using Genocchi polynomials, *Numer. Methods Partial Differential Equations* 37 (2) (2021) 1250–1268.
- [43] K. Linka, M. Peirlinck, E. Kuhl, The reproduction number of COVID-19 and its correlation with public health interventions, *Comput. Mech.* 66 (2020) 1035–1050.
- [44] V. Virlogeux, M. Li, T.K. Tsang, L. Feng, V.J. Fang, H. Jiang, P. Wu, J. Zheng, E.H. Lau, Y. Cao, et al., Estimating the distribution of the incubation periods of human avian influenza A (H7N9) virus infections, *Am. J. Epidemiol.* 182 (8) (2015) 723–729.
- [45] F. Regazzoni, D. Chapelle, P. Moireau, Combining data assimilation and machine learning to build data-driven models for unknown long time dynamics—applications in cardiovascular modeling, *Int. J. Numer. Methods Biomed. Eng.* 37 (7) (2021) e3471.
- [46] T. Nguyen, R. Novak, L. Xiao, J. Lee, Dataset distillation with infinitely wide convolutional networks, *Adv. Neural Inf. Process. Syst.* 34 (2021) 5186–5198.
- [47] N. Loo, R. Hasani, A. Amini, D. Rus, Efficient dataset distillation using random feature approximation, *Adv. Neural Inf. Process. Syst.* 35 (2022) 13877–13891.
- [48] E. Centofanti, M. Ghiotto, L.F. Pavarino, Learning the Hodgkin–Huxley model with operator learning techniques, *Comput. Methods Appl. Mech. Engrg.* 432 (2024) 117381.
- [49] R. Bollapragada, J. Nocedal, D. Mudigere, H.-J. Shi, P.T.P. Tang, A progressive batching L-BFGS method for machine learning, in: *International Conference on Machine Learning*, PMLR, 2018, pp. 620–629.
- [50] M.G. Johnsen, L.E. Christiansen, K. Græsbøll, Seasonal variation in the transmission rate of covid-19 in a temperate climate can be implemented in epidemic population models by using daily average temperature as a proxy for seasonal changes in transmission rate, *Microb. Risk Anal.* 22 (2022) 100235.
- [51] L.F. White, M. Pagano, Reporting errors in infectious disease outbreaks, with an application to Pandemic Influenza A/H1N1, *Epidemiologic Perspect. Innov.* 7 (2010) 1–12.
- [52] W.H. Organization, et al., Influenza fact sheet: Overview, *Wkly. Epidemiological Record= Relev. é é Pidé Miologique Hebd.* 78 (11) (2003) 77–80.
- [53] F. Trentini, E. Pariani, A. Bella, G. Diurno, L. Crottogini, C. Rizzo, S. Merler, M. Ajelli, Characterizing the transmission patterns of seasonal influenza in Italy: lessons from the last decade, *BMC Public Health* 22 (1) (2022) 19.
- [54] J. Lessler, N.G. Reich, R. Brookmeyer, T.M. Perl, K.E. Nelson, D.A. Cummings, Incubation periods of acute respiratory viral infections: a systematic review, *Lancet Infect. Dis.* 9 (5) (2009) 291–300.
- [55] J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers, *Proc. R. Soc. B: Biological Sci.* 274 (1609) (2007) 599–604.
- [56] A.C. Lowen, S. Mubareka, J. Steel, P. Palese, Influenza virus transmission is dependent on relative humidity and temperature, *PLoS Pathog.* 3 (10) (2007) e151.
- [57] A. Mubayi, A. Pandey, C. Brasic, A. Mubayi, P. Ghosh, A. Ghosh, Analytical estimation of data-motivated time-dependent disease transmission rate: An application to ebola and selected public health problems, *Trop. Med. Infect. Dis.* 6 (3) (2021) 141.
- [58] S.R. Krishnan, C.S. Seelamantula, On the selection of optimum Savitzky–Golay filters, *IEEE Trans. Signal Process.* 61 (2) (2012) 380–391.
- [59] D.M. Morens, J.K. Taubenberger, Pandemic influenza: certain uncertainties, *Rev. Med. Virol.* 21 (5) (2011) 262–284.
- [60] M. Biggerstaff, S. Cauchemez, C. Reed, M. Gambhir, L. Finelli, Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature, *BMC Infect. Dis.* 14 (1) (2014) 1–20.
- [61] A.J. Cohen, M. Brauer, R. Burnett, H.R. Anderson, J. Frostad, K. Estep, K. Balakrishnan, B. Brunekreef, L. Dandona, R. Dandona, et al., Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015, *Lancet* 389 (10082) (2017) 1907–1918.