https://doi.org/10.31891/2219-9365-2026-85-47
UDC 004

POHORIELOV Stanislav
National Technical University "Kharkiv Polytechnic Institute"
https://orcid.org/0000-0002-0189-8655
Stanislav.Pohorielov@khpi.edu.ua
BALABA Yaroslav
National Technical University "Kharkiv Polytechnic Institute"
https://orcid.org/0009-0005-9710-2971
Yaroslav.Balaba@cs.khpi.edu.ua

FRACTIONAL SIR MODEL OF COVID-19 WAVES WITH TIME-CHANGING PARAMETERS AND IDENTIFICATION BASED ON NEURAL NETWORK METHODS

The paper considers the task of describing individual waves of the COVID-19 epidemic using SIR-type models on real data. It is shown that the classical SIR model with constant parameters of β transmission and γ extraction is not able to adequately reproduce the form of the epidemic wave, since on real data these parameters are significantly time-dependent and reflect the impact of anti-epidemic measures, changes in population behavior, vaccination, etc. Additionally, the standard SIR model is memoryless (Markovian), while empirical morbidity series show the effects of long memory.

The aim of the work is to combine fractional SIR models with time-varying parameters and neural network identification methods to describe COVID-19 waves in three European countries (Germany, Italy, Ukraine). A hierarchy of five models is built with increasing flexibility and physical structure. As the simplest basic option, the classic SIR model with constant parameters is considered. Next, a window SIR model with piecewise time-constant $\beta(t)$ and $\gamma(t)$ is introduced, which are evaluated by gradient methods on sliding windows. The next step is the window fractional FSIR model, in which the system's memory is taken into account using a power-law kernel with a fractional order of $\alpha \in (0,1)$. The fourth model is the window FSIR-PINN, in which a multilayer neural network approximates the hidden trajectories $S(t), I(t), R(t)$ taking into account discrete FSIR equations in the error functional. Finally, the fifth model is the windowless global FSIR-PINN, which learns on the whole wave at once.

Numerical experiments show that in all three countries, the window FSIR model with $\alpha < 1$ provides a better balance between the error on the training and test parts of the series than the window classic SIR model, and gives smooth and interpreted trajectories of effective transfer and extraction ratios. Window FSIR-PINN achieves comparable accuracy in incidence and additionally provides model-aligned $S(t), I(t), R(t)$ estimates. On the other hand, the global FSIR-PINN, in the presence of only one observed series of incidents, is not able to stably reproduce the waveform, which indicates significant limitations in the identification of such global models on noisy real data. The results obtained confirm the feasibility of using fractional SIR models with time-varying parameters and local PINN architectures as a practical compromise between interpretation and flexibility for retrospective analysis of epidemic waves.

Keywords: COVID-19; FSIR model; fractional differentiation; time-changing parameters; physics-informed neural networks (PINN); parameter identification, time series, machine learning

ПОГОРЄЛОВ Станіслав, БАЛАБА Ярослав
Національний технічний університет «Харківський політехнічний інститут»

ДРОБОВА SIR-МОДЕЛЬ ХВИЛЬ COVID-19 ЗІ ЗМІННИМИ В ЧАСІ ПАРАМЕТРАМИ ТА ІДЕНТИФІКАЦІЄЮ НА ОСНОВІ НЕЙРОМЕРЕЖЕВИХ МЕТОДІВ

У статті розглядається задача опису окремих хвиль епідемії COVID-19 за допомогою моделей типу SIR на основі реальних статистичних даних. Показано, що класична SIR-модель із сталими параметрами передачі β та вилучення γ не здатна адекватно відтворювати форму епідемічної хвилі, оскільки в реальних даних ці параметри істотно змінюються з часом і відображають вплив протиепідемічних заходів, зміну поведінки населення, вакцинацію та інші фактори. Крім того, стандартна SIR-модель є безпам'ятною (марковською), тоді як емпіричні часові ряди захворюваності демонструють ефекти довготривалої пам'яті. Метою роботи є поєднання дробових SIR-моделей із параметрами, що змінюються у часі, та методів ідентифікації на основі нейронних мереж для опису хвиль COVID-19 у трьох європейських країнах (Німеччина, Італія, Україна). Побудовано ієрархію з п'яти моделей із поступовим зростанням гнучкості та збереженням фізичного змісту. Як найпростіший базовий варіант розглядається класична SIR-модель із постійними параметрами. Далі вводиться віконна SIR-модель із кусочно-сталими функціями β(t) та γ(t), які оцінюються градієнтними методами на ковзних часових вікнах. Наступним кроком є віконна дробова модель FSIR, у якій враховується пам'ять системи за допомогою степеневого ядра з дробовим порядком α∈(0,1). Четверта модель — віконна FSIR-PINN, у якій багатошарова нейронна мережа апроксимує приховані траєкторії S(t), I(t), R(t) із урахуванням дискретних рівнянь FSIR у функціоналі похибки. Нарешті, п'ята модель — глобальна безвіконна FSIR-PINN, що навчається одразу на всій хвилі.

Чисельні експерименти показують, що для всіх трьох країн віконна дробова модель FSIR з α<1 забезпечує кращий баланс між похибкою на навчальній та тестовій частинах ряду порівняно з класичною віконною SIR-моделлю, а також формує гладкі та інтерпретовані траєкторії ефективних коефіцієнтів передачі та вилучення. Модель FSIR-PINN досягає порівнянної точності щодо динаміки захворюваності та додатково забезпечує узгоджені з моделлю оцінки траєкторій S(t), I(t), R(t). Водночас глобальна FSIR-PINN за наявності лише одного спостережуваного ряду захворюваності не здатна стабільно відтворити форму хвилі, що свідчить про суттєві обмеження ідентифікації подібних глобальних моделей на шумних реальних даних. Отримані результати підтверджують доцільність використання дробових SIR-моделей зі змінними у часі параметрами та локальних архітектур PINN як практичного компромісу між інтерпретованістю та гнучкістю при ретроспективному аналізі епідемічних хвиль.

Ключові слова: COVID-19; модель FSIR; дробове диференціювання; параметри, що змінюються у часі; фізично інформовані нейронні мережі (PINN); ідентифікація параметрів; часові ряди; машинне навчання.

Стаття надійшла до редакції / Received 22.01.2026
Прийнята до друку / Accepted 24.02.2026
Опубліковано / Published 05.03.2026

INTRODUCTION

For decades, susceptible-infected-removed (SIR) mathematical models have been one of the main tools for analyzing the spread of infectious diseases and supporting decision-making in the field of public health [1]. During the COVID-19 pandemic, SIR and SEIR models were actively used to assess the baseline reproductive number, analyze the impact of anti-epidemic measures (lockdowns, mask regime, mobility restrictions), and explore possible scenarios for strengthening or weakening control [1-4]. Despite their conceptual simplicity and interpretation, classical SIR models face a number of serious limitations when applied to real COVID data, which are noisy, heterogeneous, and significantly non-stationary [2,5]. Mathematical modeling of COVID-19 dynamics for Ukraine based on generalized SIR approaches has been considered in a number of studies [6-8], some works focus on parametric identification for specific regions of Ukraine, in particular Ternopil region [9].

Two standard hypotheses are particularly significant. First, the β transmission rate and the γ extraction rate in the classical formulation are considered stable. In practice, the rates of infection and recovery change over time under the influence of anti-epidemic measures, changes in population behavior, seasonal factors, vaccination, the emergence of new variants of the virus, etc. Significant variability $\beta(t)$ and $\gamma(t)$ is one of the key reasons why a simple SIR model with constant parameters is usually not able to qualitatively reproduce the shape of a real epidemic wave. Secondly, the standard SIR model is Markovian: the future state of the system depends only on the current one, which actually corresponds to the exponential distribution of the time spent in the state of infection and the "short memory" of the process. Instead, empirical series of COVID-19 incidence often show sub-exponential growth, long wave "tails", and long-memory effects [2,5].

In this approach, the coefficients $\beta(t)$ and $\gamma(t)$ are considered as functions of time (for example, piecemeal steel or smooth), which allows for a significant improvement in the model's consistency with the data. The resulting trajectories $\beta(t)$ and $\gamma(t)$ can be associated with the calendar of anti-epidemic measures and other external factors. Another direction is related to fractional differentiation [10-13]. The replacement of classical derivatives in SIR equations with fractional derivatives (in the sense of Caputo, Grunwald-Letnikov, etc.) leads to fractional SIR models (FSIR), in which the order of the fractional derivative $\alpha \in (0,1]$ plays the role of a memory parameter: at $\alpha = 1$, the standard SIR model is restored, and at $\alpha < 1$, a power memory appears in dynamics, which can better describe the slow attenuation and inertial effects.

At the same time, the use of machine learning methods for modeling epidemics is actively developing. Purely data-oriented models, such as autoregressive neural networks, LSTM, multilayer perceptrons, can demonstrate good short-term predictive accuracy, but are often poorly interpreted and do not guarantee compliance with basic epidemiological restrictions. Against this background, the physics-informed neural networks (PINN) approach appeared, in which the neural network is trained not only from the data, but also taking into account the fact that its outputs must satisfy the given differential equations, i.e. the physical (or epidemiological) structure is clearly "sewn" into the model [14].

In previous works, we combined FSIR dynamics with neural network identification methods on synthetic data, where the true trajectories $S(t), I(t), R(t)$, parameters $\beta(t), \gamma(t)$ and order α are known in advance. It was shown that in the absence of noise and with an exact correspondence of the model structure to the data source, it is possible to restore smooth time-changing parameters of transmission and extraction, as well as to estimate the fractional order of α as memory parameter. At the same time, synthetic experiments only partially reflect the real situation, since they do not take into account measurement noise, changing testing modes, administrative artifacts in reporting, and structural discrepancies between the idealized model and the real mechanisms of transmission of infection.

The purpose of this article is to transfer these approaches to real data on COVID-19 and to systematically compare different variants of SIR/FSIR models in combination with machine learning methods. For this purpose, a hierarchy of models is built with increasing flexibility and degree of "physical" structure, which are applied to the same waves of the epidemic in three countries: Germany, Italy and Ukraine. As a naïve starting point, we consider the classic SIR model with constant parameters. we introduce a window SIR model with piecewise-constant $\beta(t)$ and $\gamma(t)$, which are evaluated on sliding windows by gradient optimization. The next step is a window fractional FSIR model with power memory. The fourth model is the window FSIR-PINN, in which the neural network reconstructs the hidden trajectories $S(t), I(t), R(t)$ taking into account the sampled FSIR equations in the loss function. Finally, as a more ambitious, but a potentially less identifiable alternative, consider the global FSIR-PINN model, which learns on the whole wave without being broken into windows.

The main results of the work can be summarized as follows. Firstly, a set of fractional and non-fractional SIR-type models with time-varying parameters on real data on the incidence of COVID-19 in three countries is

implemented and compared. Secondly, it is shown that a simple window FSIR model with a fixed fractional order $\alpha < 1$ in all cases considered provides a better balance between the quality of approximation on the training and test parts of the series, than the window-based classic SIR model, and forms smooth, interpreted trajectories $\beta(t)$ and $\gamma(t)$. Thirdly, it has been demonstrated that the window FSIR-PINN-model makes it possible to reconstruct the trajectories $S(t), I(t), R(t)$ consistent with fractional dynamics, while the global FSIR-PINN-model in the presence of only one series of incidence turns out to be non-identifiable and does not reproduce the waveform. This allows you to formulate practical recommendations on data requirements for more complex global PINN approaches.

DATA AND PRE-PROCESSING

The numerical experiments use open daily data on the incidence of COVID-19 for three European countries: Germany, Italy and Ukraine. The paper uses an open global dataset Our World in Data COVID-19 dataset, which is supported by the research group Our World in Data (University of Oxford). The data was downloaded from the official GitHub repository of the https://github.com/owid/covid-19-data/tree/master/public/data project.

This set brings together daily COVID-19 morbidity and mortality rates for all countries in the world: confirmed cases and deaths are based on data from the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE), and test, vaccination, hospitalization, and other metrics are aggregated from official national sources (ministries of health, national statistical offices, etc.), which are systematically collected and published by the Our World in Data team [15].

For each country, a separate wave of the epidemic is considered, covering the period from October 11, 2020 to June 12, 2021. The selected interval includes the autumn-winter increase in incidence and subsequent decline and is long enough to clearly observe a distinct epidemic wave of incidence. The approach chosen in the paper is consistent with previous studies, where COVID-19 waves in Ukraine were described on the basis of generalized SIR models [6,7,9].

For each country and each day t within the selected wave, the following values are used:

- Calendar date t;
- Country ID (Location field);
- 7-day smoothed number of new confirmed cases $\text{new\_cases\_smoothed}(t)$;
- population N of the country concerned.

Instead of the raw series of daily new cases ($\text{new\_cases}(t)$), which is significantly distorted by the baseline, changes in testing strategy, and other administrative factors, the 7-day smoothed series ($\text{new\_cases\_smoothed}(t)$), provided by the data source is used. On its basis, the daily incidence per 100 000 population is calculated.

Let:

- $\text{new\_cases\_smoothed}(t)$ – smoothed number of new confirmed cases on day t for selected country;
- N – population of this country.

Then the daily incidence per 100 000 population is given by the formula

$$Y_{\text{data}}(t) = \frac{\text{new\_cases\_smoothed}(t)}{N} \cdot 100\,000 \quad (1)$$

That is, it is simply "the proportion of new cases in the population per day" multiplied by 100,000 to obtain the value of "cases per 100 000 population".

It is this series $Y_{\text{data}}(t)$ that is the main observed value that is reproduced by all models considered in the paper. Other components of the state (number of susceptible, infected and removed) in real data are not directly observed and are restored only by models.

For each country, the data is sorted by date, truncated to the selected wave interval and divided into training and test parts. For simplicity and transparency of division, the rule is used: the first 60% of the wave days form a training sample, the remaining 40% form a test sample. The same train/test division is applied to all models, which ensures a correct comparison of their ability:

- be consistent with the data on the educational part of the series;
- summarize to the deferred part that was not used when adjusting the parameters.

For ease of interpretation and numerical stability, an effective population scale $\widetilde{N} = 100\,000$ is introduced. In all SIR-type models below, the variables $S(t), I(t), R(t)$ are interpreted as the number of persons per 100 000 population, and the model incidence $Y_{\text{model}}(t)$ is expressed in the same units as $Y_{\text{data}}(t)$.". All results and RMSEs below are given precisely in units of "cases per 100 000 population" Such renormalization does not change the qualitative dynamics and does not affect the relative values of indicators, but simplifies numerical implementation, selection of parameter scales and interpretation of error functional values for different countries and models.

MODELS

This section describes five models that form a hierarchy with increasing flexibility and complexity. All models work with discrete time $t = 0,1,2,\ldots,T$ (days) and with normalized population size $\widetilde{N} = 100\,000$.

Classic SIR model with constant parameters (Model 1)

The basic continuous SIR model is given by a system of differential equations

$$\frac{dS}{dt} = -\beta\frac{SI}{N}, \qquad \frac{dI}{dt} = \beta\frac{SI}{N} - \gamma I, \qquad \frac{dR}{dt} = \gamma I, \tag{2}$$

where

- $S(t)$ – number (or proportion) of susceptible,
- $I(t)$ – number of infected,
- $R(t)$ – number of removed (recovered/deceased),
- N – total population (in our case it is replaced by $\widetilde{N} = 100\,000$),
- $\beta > 0$ – transmission ratio,
- $\gamma > 0$ – removed rate (recovery/death).

The daily model incidence (the number of new infections per day per 100 000 population) is given as

$$Y_{\text{model}}(t) = \beta\,\frac{S(t)\,I(t)}{\widetilde{N}} \tag{3}$$

That is, we compute incidence as the infection intensity $Y_{\text{model}}(t)$ rather than a raw difference of cumulative cases

For numerical implementation, time sampling is used in increments of 1 day:

$$\begin{gathered} S(t+1) = S(t) - Y_{\text{model}}(t), \\ I(t+1) = I(t) + Y_{\text{model}}(t) - \gamma I(t), \\ R(t+1) = R(t) + \gamma I(t), \end{gathered} \tag{4}$$

with initial conditions

$$S(0) = \widetilde{N} - I_0, \qquad I(0) = I_0, \qquad R(0) = 0, \tag{5}$$

where $I_0 > 0$ – the initial number of infected (per 100 000).

In Model 1, the parameters β, γ, and I_0 are considered constant over the entire wave period. They are selected by minimizing the standard error between the observed and model incidence

$$\text{MSE} = \frac{1}{T_{\text{train}}} \sum_{t\in\text{train}} \left(Y_{\text{model}}(t) - Y_{\text{data}}(t)\right)^2, \tag{6}$$

where the sum is taken by the days of the training part of the wave. Optimization is performed by gradient methods (through the PyTorch implementation). This model serves as a naïve basis for estimating the gain from taking into account the time variability of parameters and memory.

Window SIR model with time-changing parameters (Model 2)

In Model 2, the assumption of constancy of β, γ is weakened: the parameters are considered piecewise-constant on sliding windows.

Let

- L – window length (in days),
- s – stride (offset step),
- k – window index,
- t_k – start of the k-th window.

Then the k-th window covers the days

$$t \in \{t_k, t_k + 1, \ldots, t_k + L - 1\}. \tag{7}$$

On each window, the parameters β and γ are considered constant:

$$\beta(t) = \beta_k, \qquad \gamma(t) = \gamma_k, \qquad t \in [t_k, t_k + L - 1], \tag{8}$$

but can vary from window to window.

For each window, the same discrete SIR sampling is applied as in Model 1:

$$\begin{gathered} Y_{\text{model}}(t) = \beta_k\,\frac{S(t)I(t)}{\widetilde{N}} \\ S(t+1) = S(t) - Y_{\text{model}}(t) \\ I(t+1) = I(t) + Y_{\text{model}}(t) - \gamma_k I(t) \\ R(t+1) = R(t) + \gamma_k I(t) \end{gathered} \tag{9}$$

The initial condition for the first window is taken as

$$S(t_0) = \widetilde{N} - I_0, \quad I(t_0) = I_0, \quad R(t_0) = 0. \tag{10}$$

For subsequent windows, the initial values $S(t_k), I(t_k), R(t_k)$ are taken as the system state at the beginning of the window, derived from the simulations on the previous windows (i.e. the trajectory is integral).

On each window the standard error between $Y_{\text{model}}(t)$ and $Y_{\text{data}}(t)$ on this window is separately minimized (with additional regularizers to prevent unrealistic values). For each window, we perform several dozen Adam steps, initializing the parameters with values from the previous window, which ensures smooth trajectories of $\beta(t)$ and $\gamma(t)$. As a result, we get sequences

$$\{\beta_k\}, \quad \text{i} \quad \{\gamma_k\}, \tag{11}$$

which can be interpreted as estimates of time-changing parameters $\beta(t)$ and $\gamma(t)$, given by stepped time functions.

Thus, Model 2 allows:

- better adapt to the waveform;
- interpret changes of β_k, γ_k in connection with the introduction/relaxation of measures.

Window fractional FSIR model with memory (Model 3)

Model 3 takes into account the long memory of the process by generalizing the SIR to fractional orders $\alpha \in (0,1]$. Instead of a full implementation of the fractional derivative in the Caputo sense, a discrete power-law kernel is used for the "effective number of infected". This approach can be considered as a discrete approximation of the Caputo kernel over a limited memory interval, which retains the main effect of power extinction of the effect of past values, but is much simpler and more stable in numerical implementation on real data.

Enter the weighted amount

$$I_{\text{eff}}(t) = \sum_{\tau=0}^{t} w_\tau \; I(t-\tau), \tag{12}$$

where the weights w_τ are given by the power law

$$w_\tau = \frac{(1+\tau)^{-k(\alpha)}}{\sum_{j=0}^{M-1}(1+j)^{-k(\alpha)}} = \frac{(1+\tau)^{-\alpha}}{C_\alpha}, \tag{13}$$

Where C_α is the normalization constant (or weights can be used without normalization). With $\alpha \to 1$ the contribution of ancient values $I(t-\tau)$ quickly decreases (almost without memory), with smaller ones $\alpha < 1$, the contribution of the distant past decreases more slowly, which corresponds to a longer memory.

Next, the model incidence is given as

$$Y_{\text{model}}(t) = \beta_k \frac{S(t)\, I_{\text{eff}}(t)}{\widetilde{N}}, \qquad t \in [t_k, t_k + L - 1], \tag{14}$$

and the equations for S, I, R remain in discrete form

$$\begin{gathered} S(t+1) = S(t) - Y_{\text{model}}(t), \\ I(t+1) = I(t) + Y_{\text{model}}(t) - \gamma_k I(t), \\ R(t+1) = R(t) + \gamma_k I(t). \end{gathered} \tag{15}$$

As in Model 2, on each window, the parameters β_k, γ_k are considered constant, but the order α is set in advance (the values $\alpha = 0.9$ and $\alpha = 0.7$ are considered in the work). At $\alpha = 1$ model coincides with the window SIR model, at $\alpha < 1$ it is a fractional FSIR model with memory.

Functionality is minimized on each window

$$\text{MSEwindow} = \frac{1}{L}\sum t \in w\mathit{indow}\left(Y_{\text{model}}(t) - Y_{\text{data}}(t)\right)^2 + \mathit{regularization}, \tag{16}$$

where regularization limits the range of β_k, γ_k, and the value $I(t)$ to avoid unrealistic trajectories. The received for all windows $\{\beta_k, \gamma_k\}$ are interpreted as time-changing parameters of the model with memory.

Thus, Model 3 expands Model 2 by fractional order α and allows you to take into account the influence of previous values $I(t)$ on the current intensity of infections.

Window FSIR-PINN Model (Model 4)

Model 4 uses the physics-informed neural networks (PINN) approach for window FSIR dynamics. All experiments used the same multilayer perceptron architecture for PINN with 3-4 hidden layers of 32–64 tanh-activated neurons. The idea is that on each window:

- Multilayer neural network approximates trajectories $S(t), I(t), R(t)$ as a function of time;
- parameters β_k, γ_k (or their logarithms) are also optimized;
- loss function includes both data (incidence) and physical equations of FSIR.

Let t within the window is linearly scaled

$$\tau \in [0,1], \quad t_j = t_k + j, \quad \tau_j = \frac{j}{L-1}. \tag{17}$$

A neural network with parameters θ sets the approximation.

$$\hat{S}(\tau, \theta), \hat{I}(\tau, \theta), \hat{R}(\tau, \theta). \tag{18}$$

At the discrete points τ_j corresponding to the days of the window, we define

$$S_j = \hat{S}(\tau_j, \theta), \quad I_j = \hat{I}(\tau_j, \theta), \quad R_j = \hat{R}(\tau_j, \theta). \tag{19}$$

As in Model 3, we enter the effective number of infected

$$I_{\text{eff},j} = \sum_{l=0}^{j} w_{j-l} I_l, \tag{20}$$

where the weights w_τ are determined by the order of α.

Model incidence in step j:

$$Y_{\text{model},j} = \beta_k \frac{S_j\, I_{\text{eff},j}}{\widetilde{N}}. \tag{21}$$

Next, the equations of discrete dynamics FSIR are given:

$$\begin{gathered} S_{j+1}^{\text{FSIR}} = S_j - Y_{\text{model},j}, \\ I_{j+1}^{\text{FSIR}} = I_j + Y_{\text{model},j} - \gamma_k I_j, \\ R_{j+1}^{\text{FSIR}} = R_j + \gamma_k I_j. \end{gathered} \tag{22}$$

The neural network values $S_{j+1}, I_{j+1}, R_{j+1}$ should be close to the right parts $S_{j+1}^{\text{FSIR}}, I_{j+1}^{\text{FSIR}}, R_{j+1}^{\text{FSIR}}$ of the sampled FSIR model with the historical $I^{(\alpha)}$.

The loss function on one window looks like

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_{\text{phys}}\, \mathcal{L}_{\text{phys}} + \lambda_{\text{reg}}\, \mathcal{L}_{\text{reg}}, \tag{23}$$

where data-loss:

$$\mathcal{L}_{\text{data}} = \frac{1}{L}\sum_j \left(Y_{\text{model},j} - Y_{\text{data},j}\right)^2; \tag{24}$$

and physics-loss:

$$\mathcal{L}_{\text{phys}} = \frac{1}{L}\sum_j \left(\left(S_{j+1} - S_{j+1}^{\text{FSIR}}\right)^2 + \left(I_{j+1} - I_{j+1}^{\text{FSIR}}\right)^2 + \left(R_{j+1} - R_{j+1}^{\text{FSIR}}\right)^2\right); \tag{25}$$

regularization includes penalties for too high values I_j ($I_j \ll \widetilde{N}$) and, if necessary, smoothing trajectories.

FSIR-PINN neural network implementation.

In the FSIR-PINN window model, for each sliding window of length L, unknown trajectories (S_t, I_t, R_t) are approximated by a multilayer perceptron f_θ. Normalized time $\tau \in [0,1]$ is given as input to the network, which linearly corresponds to discrete days within the window. The output of the network are three latent quantities $S(\tau), I(\tau), R(\tau)$ interpreted as the number of susceptible, infected, and removed in an "effective" population of size $N_{\text{eff}} \approx 10^5$.

For approximation (S, I, R) the same architecture is used for all windows and countries: one-dimensional input (normalized time τ), three hidden layers of 32 neurons with activation $tanh$, and an output linear layer of dimension 3:

$$\tau \xrightarrow{Linear+tanh} \mathbb{R}^{32} \xrightarrow{Linear+tanh} \mathbb{R}^{32} \xrightarrow{Linear+tanh} \mathbb{R}^{32} \xrightarrow{Linear} \mathbb{R}^3 \tag{26}$$

To guarantee the non-negative and approximate preservation of the total number, the raw network outputs $\left(\tilde{S}, \tilde{I}, \tilde{R}\right)$ are converted using the Softplus operator and subsequent normalization:

$$\begin{gathered} u = softplus\left(\tilde{S}\right) + \varepsilon, \qquad v = softplus\left(\tilde{I}\right) + \varepsilon, \qquad w = softplus\left(\tilde{R}\right) + \varepsilon \\ S = N_{eff}\frac{u}{u+v+w}, \qquad I = N_{eff}\frac{v}{u+v+w}, \qquad R = N_{eff}\frac{w}{u+v+w} \end{gathered} \tag{27}$$

where $\varepsilon > 0$ is a small numerical constant. Thus, for each moment of time on the window, we get $S, I, R \geq 0$ and $S + I + R \approx N_{\text{eff}}$.

The parameters of the transfer function f_θ as well as the window parameters $\beta_k, \gamma_k > 0$ (for each k window) are evaluated jointly by minimizing the total loss function. Non-negativity of β_k *and* γ_k is provided by parameterization through the Softplus function from trained scalars (specified as $\beta_k = softplus(b_k)$, $\gamma_k = softplus(g_k)$, where b_k and g_k are ordinary numerical parameters that are evaluated together with the weights of the neural network during training). For each window, optimization is done by the Adam method with a learning step 10^{-3}, for up to 1500 epochs, without mini-batches (each iteration uses an entire time window). In the physics-loss part (penalty for not executing discretized FSIR equations with memory), the mean square of the residuals for the three components S, I, R is minimized, and a small regularization term is added to the total loss, which penalizes too large values $I(t)/N_{\text{eff}}$ and prevents the unrealistic "dumping" of the proportion of infected to 1.

The same network architecture and optimization scheme is used in both the FSIR-PINN window model and the global FSIR-PINN variant (Model 5), where instead of individual windows, the entire incidence wave is treated as a single time interval.

As a result of simultaneous optimization of neural network parameters θ and window parameters β_k, γ_k, we get:

- Trajectory estimates on this window $S(t), I(t), R(t)$, consistent with fractional dynamics;
- estimates β_k, γ_k compatible with both data and FSIR equations.

By stitching the results by window, we form global values $S(t), I(t), R(t), \beta(t), \gamma(t)$. Thus, Model 4 is a "neural network add-on" over the window FSIR model, which allows for more flexible recovery of hidden trajectories of states.

In window FSIR models, the parameters β_k, γ_k are not direct outputs of the neural network, but are specified as separate numerical parameters for each window, which are evaluated by gradient methods along with network weights. Thus, time-dependent trajectories $\beta(t)$ and $\gamma(t)$ are described by piecewise-constant functions, where each window has its own values β_k, γ_k, and the neural network is responsible for reconstructing hidden trajectories

$S(t), I(t), R(t)$ within these parameters.

Global FSIR-PINN-Windowless Model (Model 5)

Model 5 is a more ambitious version of FSIR-PINN, in which a single neural network approximates trajectories $S(t), I(t), R(t)$ along the entire wave period without breaking into windows. In this case,:

- neural network $S_\theta(t), I_\theta(t), R_\theta(t)$ is defined for each $t \in [0, T]$;
- parameters $\beta(t), \gamma(t)$ can be set as separate neural network outputs or as smooth time functions;
- the loss function contains:
  - consistency of incidence $Y_{\text{model}}(t)$ with $Y_{\text{data}}(t)$ within the entire interval;
  - total physics-loss for discrete FSIR equations on all days.

Formally, the loss structure is similar to Model 4, but without localization in windows, i.e. one optimization task covers the entire wave. On synthetic data and/or in the presence of additional observable values (e.g., active cases, hospitalizations), such a formulation can be very attractive.

In this paper, the global FSIR-PINN model is used primarily as an experimental test: how realistic it is to identify a full-fledged FSIR model with time-varying $\beta(t), \gamma(t)$ and hidden trajectories $S(t), I(t), R(t)$ on real data based on just one series of incidences.

Numerical experiments: staging and setting up

For each of the three countries (Germany, Italy and Ukraine), one COVID-19 epidemic wave is considered, divided into training and test parts as described in Section 2 and is the same for all models (1-5). All model parameters (including neural network scales for PINN models) are configured exclusively according to the data of the training part; the test part is only used to assess the model's ability to generalize.

In window models (Models 2-4), the time axis is split into overlapping sliding windows of fixed length:

- window length $L = 21\ days$;
- stride (offset step) $s = 7\ days$.

That is, each subsequent window begins 7 days after the previous one, and the windows themselves are closed. For the k-th window, the beginning t_k is defined, and the window covers the space

$$t \in \{t_k, t_k + 1, \dots, t_k + L - 1\} \tag{28}$$

Such a choice of parameters L and s provides a sufficient number of points inside each window for a stable assessment of parameters; allows you to get fairly frequent sampling of trajectories $\beta(t), \gamma(t)$, since the windows overlap; smooths out the impact of random noise by co-optimising parameters over a 21-day period.

In Models 2 and 3 scalar parameters β_k, γ_k are evaluated on each window. In Model 4, local training of the neural network is additionally performed, which restores trajectories $S(t), I(t), R(t)$ within the window.

For fractional models (FSIR), i.e. Model 3 (FSIR_window) and Model 4 (FSIR_PINN_window), order α is treated as a given hyperparameter rather than as an automatically evaluated parameter. The paper analyzes two values: $\alpha = 0.9$ and $\alpha = 0.7$.

The value $\alpha = 1$ corresponds to classical (non-fractional) SIR sampling, and in this sense, Model 2 is a special case of Model 3 at $\alpha = 1$. The value $\alpha = 0.9$ characterizes a moderate memory effect, and $\alpha = 0.7$ – stronger memory with a slower decline in the weights of past values $I(t)$ in the nucleus w_τ.

For each country and each model, separate series of experiments are carried out for $\alpha = 0.9$ and $\alpha = 0.7$, which allows comparing the effect of memory strength on the quality of the wave approximation and the behavior of the evaluated $\beta(t), \gamma(t)$.

All models are implemented in a Python environment using the PyTorch library for automatic differentiation and gradient optimization.

For Model 1 (SIR_const), the optimization of the trio of parameters (β, γ, I_0) is done by the Adam method with a small number of iterations and constraints on the range of parameters (e.g. $\beta, \gamma > 0, I_0 > 0$).

For Models 2 and 3 (SIR_window and FSIR_window) on each window β_k, γ_k are optimized using Adam. The initial approximations for the parameters on the new window are taken from the previous window, which improves the stability and smoothness of the estimates. To prevent unrealistic values, constraints on β_k, γ_k (e.g. due to parameterization in the log scale) and regularization of the value $I(t)$ (so as not to exceed a small fraction of $\widetilde{N}$) are used.

For Model 4 (FSIR_PINN_window), the weights of a multilayer neural network θ are simultaneously optimized on each window, which approximates $S(t), I(t), R(t)$ and parameters β_k, γ_k.

Loss function includes data-loss (coordination $Y_{\text{model}}(t)$ with $Y_{\text{data}}(t)$ within the window); physics-loss (deviation of neural network trajectories from discretized FSIR equations); regularization terms.

The number of training epochs per window is chosen to be sufficient for convergence, but limited to avoid over-retraining on the local piece of data. Limitations and regularization similar to Model 3 are also used for $\theta, \beta_k, \gamma_k$.

Model 5 (FSIR_PINN_global) is optimized similarly to Model 4, but without splitting into windows: one

neural network approximates $S(t), I(t), R(t)$ within the entire wave period. Due to the high complexity of this task and the limited data limitation, we consider the global PINN model as an experimental test for identification, and not as the main working tool.

For quantitative comparisons of models, standard errors (RMSE) between the observed and model incidence are used, separately for the training and test parts of the wave. Let T_{train} - the number of days in the training, T_{test} - in the test. Then:

training RMSE:

$$RMSE_{\text{train}} = \sqrt{\frac{1}{T_{\text{train}}} \sum_{t \in \text{train}} \left(Y_{\text{model}}(t) - Y_{\text{data}}(t)\right)^2}; \quad (29)$$

test RMSE:

$$RMSE_{\text{test}} = \sqrt{\frac{1}{T_{\text{test}}} \sum_{t \in \text{test}} \left(Y_{\text{model}}(t) - Y_{\text{data}}(t)\right)^2}. \quad (30)$$

All RMSEs are calculated in the same units: "cases per 100 000 population". Subsequently, for each country and each α-value , the RMSE tables for all models (1-4) are analyzed, and for Germany, the results of the global FSIR-PINN model (5) are additionally given, which allows one to assess both the approximation to the data and the degree of overtraining or degradation of quality on the test.

RESULTS

This section presents the results of simulations of COVID-19 waves in Germany, Ukraine and Italy for all models considered. For each country, we compare the quality of reproduction of the daily incidence per 100 000 population in terms of standard error (RMSE) on the training and test parts of the series. Separately, we analyze how the results are affected by the choice of fractional order α in FSIR models.

Germany

First, let's look at the results for Germany, which can be considered "benchmark", since it was for this country that we tested all five models, including the global PINN staging. Table 1 contains the RMSE values for the four main models (1–4) at $\alpha = 0.7$ and $\alpha = 0.9$.

Table 1.

RMSE values for Germany

Model	α	$RMSE_{\text{train}}$	$RMSE_{\text{test}}$ ()
SIR_const	–	10.26	13.08
SIR_window	–	3.75	13.76
FSIR_window	0.7	5.98	5.77
FSIR_window	0.9	5.88	5.61
FSIR_PINN_window	0.7	10.97	8.84
FSIR_PINN_window	0.9	11.23	9.38

For Germany, the classic SIR model with constant parameters (SIR_const) gives a rather limited quality: the RMSE on the training sample is about 10.3, on the test sample about 13.1. This means that such a model only roughly "feels" the scale of the wave, but is not able to correctly reproduce its shape, in particular the asymmetry of growth and decline (see Fig. 1).

The transition to a window SIR model with piecewise-constant parameters (SIR_window) sharply reduces the training error to 3.75, but does not give a gain on the test: $RMSE_{test}$ even worsens a little (13.76 versus 13.08). In other words, the model adapts very well to the learning part of the wave, but its generalization ability remains at the level of naive SIR_const (see Fig. 1). This is a typical symptom of local overtraining: the parameters β_k, γ_k adjust the noise in each window too aggressively (see Fig. 2).

The situation changes after the introduction of fractional memory in FSIR_window models. For both values of fractional order ($\alpha = 0.7$ and $\alpha = 0.9$) the learning error increases compared to SIR_window (to about 5.9), but the test error decreases by more than half to 5.8 at $\alpha = 0.7$ and to 5.6 at $\alpha = 0.9$. Thus, the inclusion of memory (through the power-law kernel in the intensity of infections) makes the model more rigid, but significantly improves its ability to generalize (see Fig.1). In a sense, FSIR_window "sacrifices" some of the accuracy on the training to avoid overtraining and better describe the shape of the test part of the wave. For Germany, the value $\alpha = 0.9$ gives a slightly better compromise, but the difference between $\alpha = 0.7$ and $\alpha = 0.9$ is not fundamental.

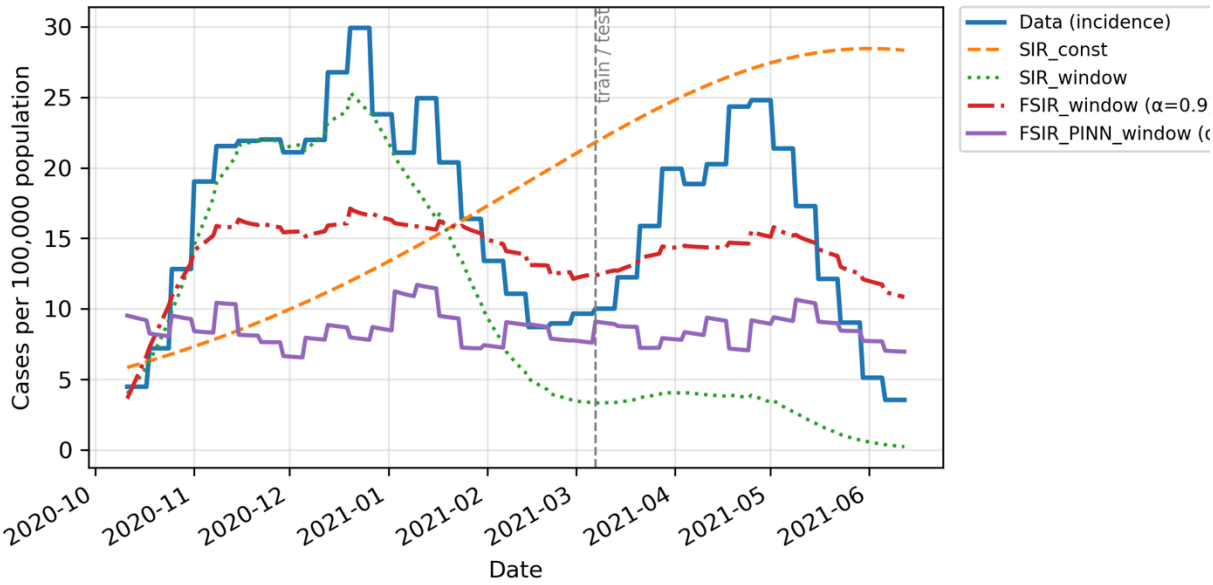

Fig.1. Germany: COVID-19 incidence and models

The window PINN model FSIR_PINN_window shows intermediate results. Its RMSE on the test (about 8.8-9.4) is worse than that of FSIR_window, but better than that of SIR_const and SIR_window. The main advantage of this model is not in minimizing the RMSE, but in the fact that it simultaneously reproduces the incidence and provides trajectories consistent with fractional equations $S(t), I(t), R(t)$ (see Fig.1). For Germany, we observe that FSIR_PINN_window is capable of constructing physically meaningful SIR trajectories, but these additional constraints "cost" a certain increase in error compared to a purely parametric FSIR_window.

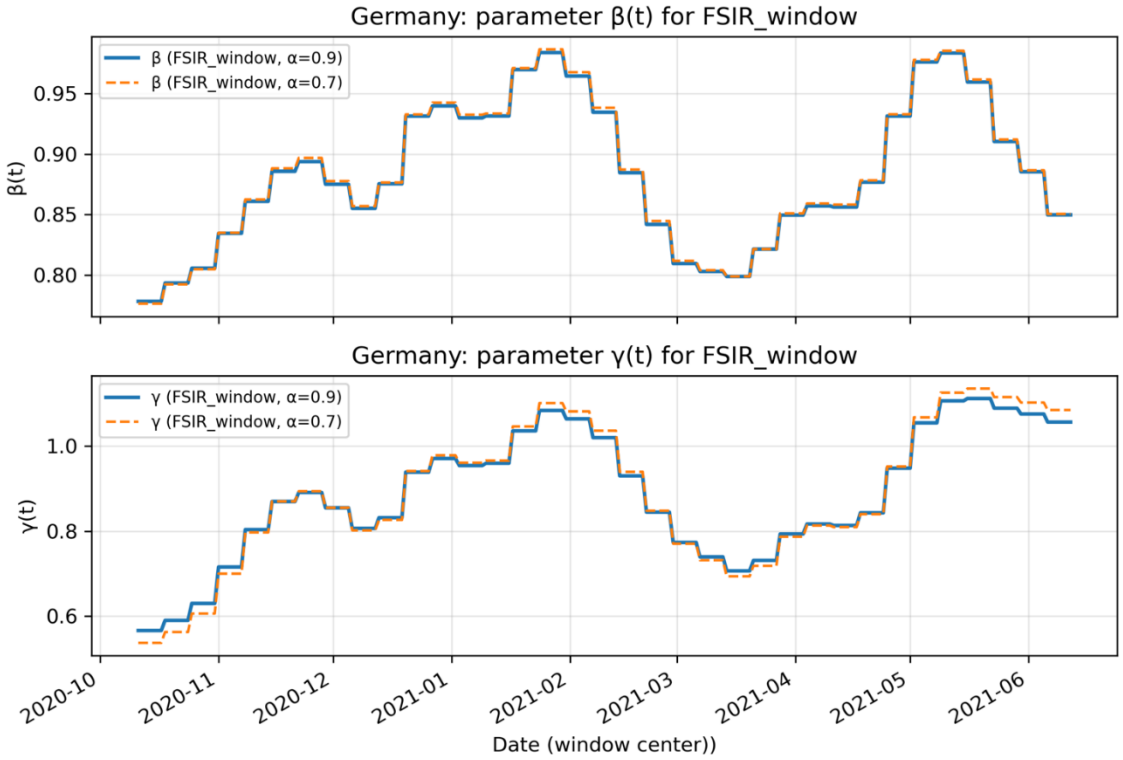

Fig.2. Germany: dependencies of parameters $\beta(t)$ and $\gamma(t)$ for FSIR_window model

To complete the picture, let's look at the global PINN model FSIR_PINN_global, which learns on the entire wave at once (without windows). For Germany, it gives $RMSE_{\text{train}}$≈6.94 and $RMSE_{\text{test}}$≈6.38, that is, formally slightly worse than the window FSIR_window, but not catastrophically. However, the analysis of the graphs shows that in this case the model wave becomes too smoothed, the peaks are "cut off", and the estimated values $\beta(t), \gamma(t)$ almost do

not change in time (see Fig. 3). That is, the model finds a compromise that looks "decent" in terms of RMSE, but epidemiologically unrealistic. This indicates weak identification of the global FSIR-PINN staging in the presence of only one observed series $Y_{\text{data}}(t)$.

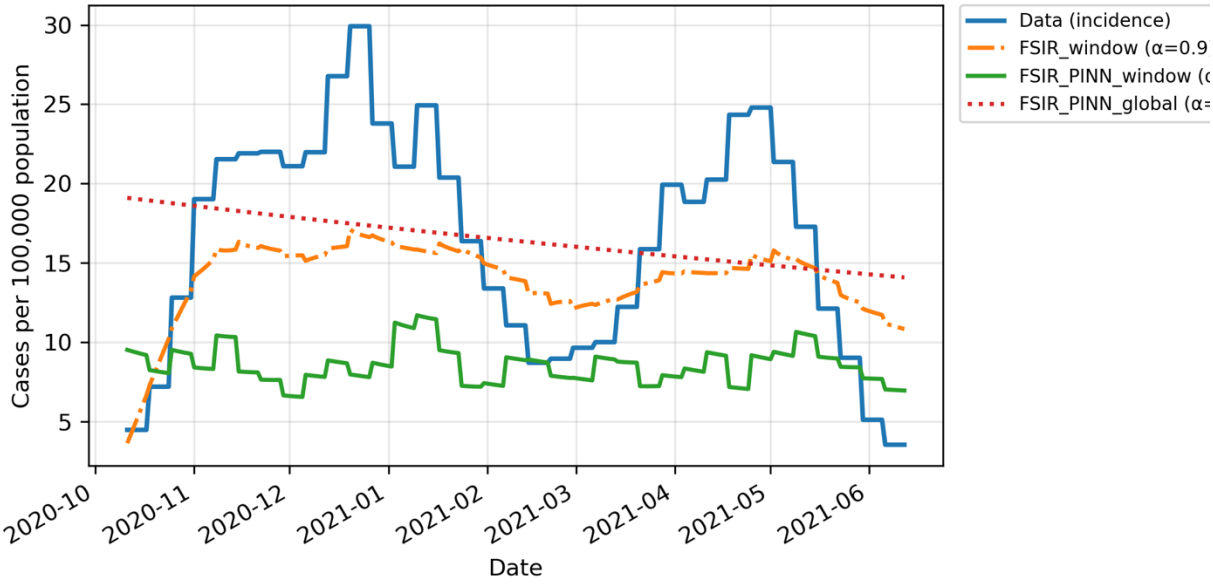

Fig.3. Germany: comparison of FSIR_window, FSIR_PINN_window and FSIR_PINN_global models

Thus, for Germany, the most balanced in terms of approximation quality and interpretation is precisely the window FSIR-model with $\alpha < 1$, while the window FSIR-PINN is useful primarily for the reconstruction of hidden SIR trajectories, and the global PINN model serves as an example of how an "overly ambitious" setup can turn out to be unidentifiable on real data.

Ukraine

For Ukraine, a very similar picture is observed, although the numerical values of the RMSE themselves are slightly different, reflecting both a different scale of the wave and a higher level of noise in the data. Table 2 shows the results for $\alpha = 0.7$ and $\alpha = 0.9$.

Table 2.

RMSE values for Ukraine

Model	α	$RMSE_{\text{train}}$	$RMSE_{\text{test}}$
SIR_const	–	10.34	12.28
SIR_window	–	4.49	21.01
FSIR_window	0.7	6.00	10.85
FSIR_window	0.9	5.67	10.49
FSIR_PINN_window	0.7	12.54	17.55
FSIR_PINN_window	0.9	12.47	16.90

As in the case of Germany, SIR_const with constant parameters gives moderate quality (test RMSE ≈12.3) (see Fig.4).

Switching to SIR_window sharply reduces the error on training (to 4.49), but on the test it leads to a very large deterioration $RMSE_{test}$≈21. This result is consistent with intuition: a local model with piecewise-constant parameters over-adjusts fluctuations in each window and is not formed as a globally consistent SIR dynamics over the entire wave (see Fig. 4).

The window fractional model FSIR_window stabilizes the behavior again: at $\alpha = 0.7$ we get $RMSE_{\text{test}}$≈10.85, and at $\alpha = 0.9$ even better, ≈10.49. Both values are significantly lower than those of SIR_const and SIR_window. As for Germany, FSIR_window manifests itself as a model that "loses" a little in accuracy on training, but significantly wins in generalization, especially on fragments of a wave with a more complex shape (see Fig. 4).

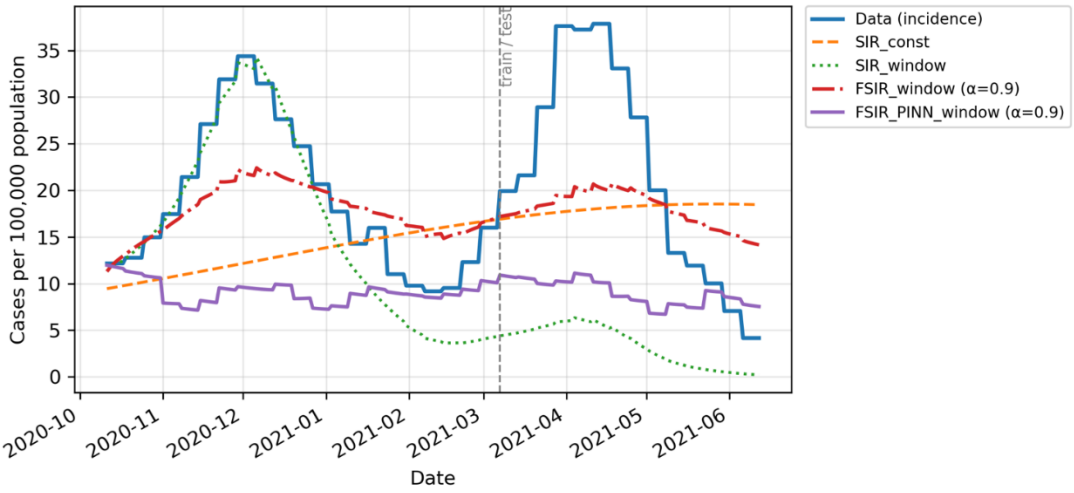

Fig.4. Ukraine: COVID-19 incidence and models

The PINN model FSIR_PINN_window in Ukraine is already noticeably losing in all metrics: for both values α it $RMSE_{\text{train}}$ exceeds 12, and $RMSE_{test}$ exceeds 16-17. This can be explained by the more complex structure of the Ukrainian wave and the greater unevenness of the data: a neural network that simultaneously tries to match the data and the FSIR difference equations often "chooses" compromise trajectories of the SIR, which are acceptable from an epidemiological point of view, but are far from the optimal approximation of the incidence in the sense of RMSE (see Fig.4). However, even in this case, the FSIR_PINN_window is useful as a tool for checking whether it is possible to find a physically meaningful SIR structure with fractional memory inside the data at all.

As a result, for Ukraine, if we focus on the quality of approximation, the best result is given by FSIR_window with $\alpha = 0.9$. At the same time, FSIR_PINN_window here rather plays an auxiliary role for structural analysis.

Italy

For Italy, the situation is even more pronounced: the classic SIR model does not keep up with the wave dynamics, the window SIR is greatly overtrained, and the fractional FSIR_window again turns out to be the most balanced. The corresponding RMSE values are shown in Table 3.

Table 3.

RMSE values for Italy

Model	α	$RMSE_{\text{train}}$	$RMSE_{\text{test}}$
SIR_const	–	20.49	13.65
SIR_window	–	11.20	23.50
FSIR_window	0.7	9.85	9.90
FSIR_window	0.9	9.42	9.02
FSIR_PINN_window	0.7	25.32	16.89
FSIR_PINN_window	0.9	24.45	17.72

The base SIR_const for Italy reproduces the wave very approximately: train-RMSE≈20.5, test-RMSE≈13.7 (see Fig.5). The window SIR model still drastically reduces the error on training (to 11.2), but the test error increases to ~23.5, that is, it turns out to be worse than even in SIR_const.

This suggests that for Italy, "local" parameterization without memory is particularly heavily overtrained, reflecting not only the true change of modes, but also random noise and uneven reporting (see Fig.5).

In this context, the transition to FSIR_window turns out to be particularly successful. For $\alpha = 0.7$ the model reaches $RMSE_{\text{train}}$≈9.85 and $RMSE_{\text{test}}$≈9.90; for $\alpha = 0.9$ $RMSE_{\text{train}}$≈9.42 and $RMSE_{test}$≈9.02 (see Fig. 5). All these values are significantly better for SIR_const and, even more so, for SIR_window. The best overall result (on both training and test) is given by the option with $\alpha = 0.9$, which is consistent with the conclusions for Germany and Ukraine: a moderate fractional order α provides the best compromise between flexibility and regularization.

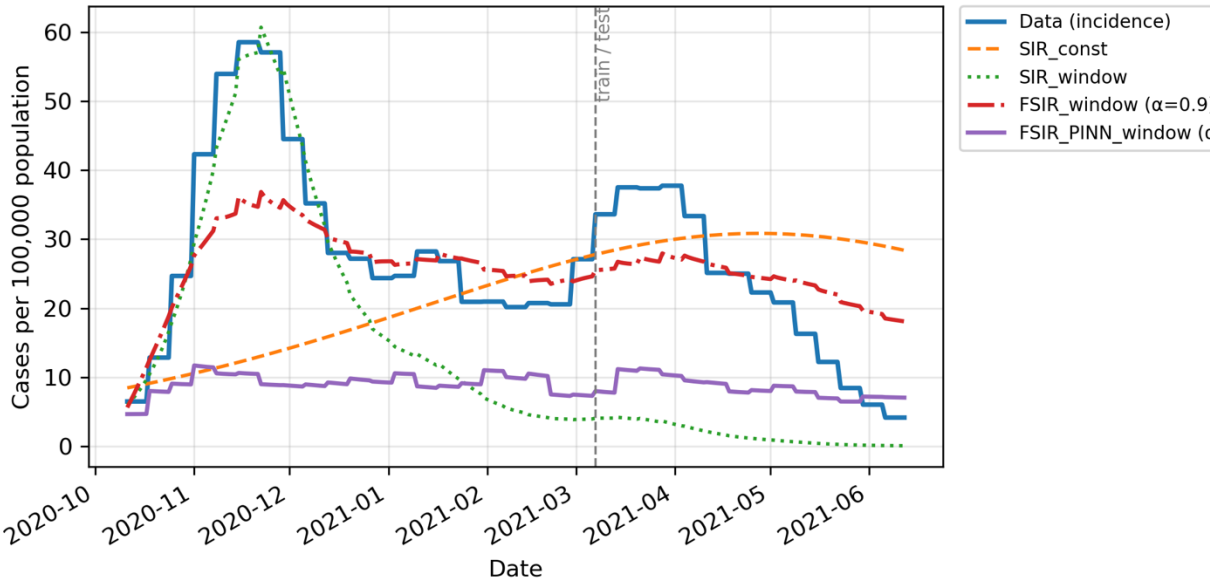

Fig.5. Italy: COVID-19 incidence and models

The PINN model FSIR_PINN_window for Italy, on the contrary, shows very large errors ($RMSE_{\text{train}}$≈24-25, $RMSE_{\text{test}}$≈17-18). Here PINN "costs" the model too much in terms of RMSE, and from a practical point of view, it should not be used as a main model to describe incidence, but as a tool for studying possible SIR structures and parameter sensitivity in fractional formulation (see Fig.5).

GENERALIZATION OF RESULTS

Comparing all three countries, we see the same general picture. The simple SIR_const option provides a limited but stable quality of wave approximation and logically serves as a basic model. Also, the limitations of the classical SIR model in the analysis of real epidemic waves of COVID-19 are confirmed in the works [2,4,5]. The window SIR model SIR_window radically reduces the error on the training part, but on the test it almost always loses even SIR_const, which indicates a strong overtraining of local piecewise-constant parameters without memory.

The fractional window model FSIR_window in all three countries shows the opposite behavior: it "sticks" to training a little worse, but it significantly improves the quality on the test and turns out to be the only model that consistently reduces the test RMSE compared to the base SIR_const. The resulting effect of stabilizing parameters and improving generalization is well consistent with the conclusions of the studies [10-13], where fractional order was also useful for describing the dynamics of COVID-19. This suggests that the introduction of fractional memory into the SIR model, together with local parameter estimation on windows, yields a natural and, importantly, physically interpreted regularization.

The role of PINN architectures in this context is twofold. The window FSIR_PINN_window model, despite the worst RMSEs, is valuable as a tool for reconstructing the trajectories $S(t), I(t), R(t)$ themselves and for testing how the observed incidence series is generally compatible with the assumed fractional SIR dynamics. On the other hand, the global FSIR_PINN_global model, despite acceptable RMSE values, turns out to be structurally unreliable: without additional observable values, it is not able to unambiguously restore either the parameters or the waveform.

In general, it can be concluded that for retrospective analysis of individual waves in real data, the most practical trade-off between approximation quality and interpretation is a window fractional model FSIR_window with moderate fractional order $\alpha < 1$. It is advisable to consider PINN approaches as an addition to structural analysis and study of hidden SIR components, but not as a "first" model for fitting incidence in the presence of only one time series.

DISCUSSION

The results obtained allow us to return to the initial questions that actually stand behind any epidemiological modeling: what exactly do we want to get from the model and what price are we willing to pay for interpretation, flexibility and noise resistance. In our case, the emphasis was not on operational forecasting, but on the retrospective analysis of individual waves of COVID-19 in order to understand whether fractional SIR dynamics with time-changing parameters can adequately explain the shape of real waves and what types of neural network add-ons are appropriate.

The first important conclusion concerns the role of memory in the model. The basic SIR model is Markov and, as the numerical results showed, in all three countries it is able to only roughly reproduce the scale of the wave, but not its shape. The conclusions we have obtained regarding the limitations of classical SIR staging and the

importance of correct identification of parameters are in good agreement with previous work on the analysis of COVID-19 in Ukraine [6, 7, 9]. Window SIR model, which allows parameters β and γ to change over time, supposedly should solve this problem, but in practice it turns out to be highly prone to overlearning: the training error drops sharply, while the test error increases and exceeds even SIR_const. Due to the lack of memory, the model easily "adjusts" local fluctuations without building a coherent coordinated dynamics.

In this sense, the transition to a fractional FSIR_window model with fixed-order $\alpha < 1$ turns out to be a natural way to introduce structural regularization. The power-law kernel in the intensity of infections does not allow the model to react only to the current value $I(t)$, forcing the "history" of the process to be taken into account. The results for all three countries show that such memory, on the one hand, makes the model less flexible in the sense of a formal fitting of the training part, but on the other hand, it significantly improves the quality of reproduction of the delayed fragment of the wave. In other words, fractional order α turns out to be not just a technical parameter, but an actual regulator of the balance between flexibility and generalization.

The second block of questions is related to the role of neural networks and the PINN approach. Our experiments show that window FSIR_window without any neural networks already provides a very good compromise between interpretation and quality of approximation. It clearly sets the structure in the form of a fractional SIR model, has few parameters for each window, and gives a stable gain on the test compared to a non-fractional window SIR model. Against this background, the question naturally arises: why do we need PINN add-ons at all, if they tend to lose to the simplest FSIR_window at RMSE?

The answer, in our opinion, lies in the type of information we want to get. Parametric FSIR_window allows you to reliably estimate time-variables $\beta(t)$ and $\gamma(t)$, but does not give internal trajectories $S(t), I(t), R(t)$ with control of their consistency with equations. Window FSIR_PINN_window, on the contrary, puts in the first place the coordination of the hidden SIR structure with the data: the neural network recovers $S(t), I(t), R(t)$, and the physics-informed part of the losses causes these trajectories to satisfy the discretized FSIR equations. In this formulation, the RMSE ceases to be a single goal by incident, but turns into one of the parts of the compromise between the "physical" and "data-oriented" components. Therefore, the lower accuracy of the FSIR_PINN_window on the test part seems quite logical: the model pays part of the accuracy to adhere to a more rigid structural scheme.

A special place in our work is occupied by the global FSIR_PINN_global model. Formally, it looks the most "harmonious": one neural network describes $S(t), I(t), R(t)$ within the whole wave period, parameters $\beta(t), \gamma(t)$ can be smooth functions of time, and physics-loss is superimposed on the entire trajectory at once. However, it is here that identification problems are most clearly manifested. If there is only one observed series $Y_{\text{data}}(t)$, such a model turns out to be too multidimensional, and different combinations of trajectories $S(t), I(t), R(t)$ and parameters can give similar RMSE values. Our test in Germany showed that the global PINN is able to achieve a formally "decent" error, but at the same time reproduces an artificially smoothed wave with almost constant $\beta(t), \gamma(t)$ values, that is, it loses exactly those features of dynamics that interest us. This negative result, in our opinion, is no less important than the positive ones: it indicates that in order to correctly apply global PINN approaches on real epidemic data, it is necessary either to significantly increase the volume of observations (to include additional series, for example, active cases, hospitalizations, deaths, reinfections), or to significantly limit the space of permissible trajectories.

From a practical point of view, the results obtained support this approach to modeling individual waves. For problems of retrospective analysis and comparison of waves in different countries, the window fractional model FSIR_window with a moderate fractional order α (in our examples $\alpha = 0{,}9$) seems to be the most expedient. It gives a stable gain in test error, is naturally interpreted through SIR variables, allows you to build smooth trajectories $\beta(t), \gamma(t)$, and at the same time does not require a complex neural network infrastructure. Window FSIR_PINN_window can be considered as the next "layer" of analysis, if the researcher is no longer interested in errors by incidence so much as in the form and behavior of hidden $S(t), I(t), R(t)$, and in agreement with fractional dynamics. Global PINN models, in our opinion, make sense only in situations where the volume and quality of data are sufficient to avoid non-identifiability, or if pre-built simpler models have already shown their failure.

Separately, it is worth emphasizing the limitations of the study. In the first place, each country is represented by only one wave of COVID-19, and although this allows for a detailed analysis of specific scenarios, for conclusions of a general nature it is advisable to include more waves, different variants of the virus and different modes of measures. Secondly, the fractional order α in our work was not automatically evaluated, but was set in the form of several fixed values. This approach is justified in the context of the article, but for the completeness of the study, it is advisable to consider, at least on synthetic data, collaborative optimization procedures with $\alpha, \beta(t), \gamma(t)$ or grid search by α. Third, in fractional equations, we used a simplified power-law kernel and finite memory, rather than a full-fledged implementation of the Caputo derivative with classical schemes like ABM. This decision was motivated by numerical stability and the need for adequate computational time on real data.

Despite these limitations, the conducted simulations made it possible to form a holistic picture of how different levels of complexity of SIR-type models behave on real COVID data, and to identify several practically important principles. First, the consideration of memory through fractional order $\alpha < 1$ is not just a cosmetic extension, but a significant step towards a more stable generalization. Second, local window schemes combined with fractional memory can be more efficient than formally more "beautiful" global models. Thirdly, neural network PINN

architectures should be considered not as a universal replacement for classical models, but as a tool that complements them, allowing you to check how the assumed physical structure is really present in the data. It is in this combination – fractional SIR models with time-changing parameters plus local PINNs – that we see the most promising direction for the further development of modeling epidemic processes.

CONCLUSIONS

The paper shows that for a retrospective description of COVID-19 waves on the data of Germany, Italy and Ukraine, the classical SIR model and its window version without memory are insufficient: the first does not reproduce the waveform, the second is subject to strong overtraining. The best balance between accuracy on deferred data and interpretation is provided by the fixed-order $\alpha < 1$ window fractional FSIR_window model, which gives a steady reduction in test RMSE and naturally takes into account memory effects. Neural network FSIR-PINN models should be used as a tool for structural analysis and reconstruction of hidden trajectories $S(t), I(t), R(t)$, while their global variants require additional data or stricter constraints to ensure identification.

References

1. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics // Proceedings of the Royal Society A. 1927. Vol. 115, P. 700–721. https://doi.org/10.1098/rspa.1927.0118
2. Kalachev L., et al. Revisiting classical SIR modelling in light of the COVID-19 pandemic // Infectious Disease Modelling. 2023. https://doi.org/10.1016/j.idm.2022.12.002
3. Nesteruk I. Simulations and Predictions of COVID-19 Pandemic With the SIR Model // Innovative Biosystems and Bioengineering. 2020. Vol. 4(1), P. 20–29. https://doi.org/10.20535/ibb.2020.4.2.204274
4. Mohajan H. Mathematical Analysis of SIR Model for COVID-19 Pandemic // MPRA Paper No. 114390, 2022. https://mpra.ub.uni-muenchen.de/114390/
5. Atkeson A. COVID-19: Epidemiological Models // Annual Review of Financial Economics. 2023. Vol. 15, P. 497–524. https://doi.org/10.1146/annurev-financial-110821-020639
6. Brovchenko I.O. Development of a mathematical model of the spread of COVID-19 in Ukraine. Policy Brief of the National Academy of Sciences of Ukraine, 2020. https://files.nas.gov.ua/PublicMessages/Documents/0/2020/05/200506172747204-403.pdf?utm_source=chatgpt.com
7. Nesteruk I. Simulations of new COVID-19 pandemic waves in Ukraine and the world with the use of generalized SIR model // Journal of the Institute of Applied System Analysis (KPI), 2022. https://doi.org/10.20535/SRIT.2308-8893.2022.2.07
8. Gavrilchyk L.S. Estimation of the effectiveness of the SIRV model for analyzing the spread of COVID-19 // Problems of programming and modeling. 2024. https://doi.org/10.32782/mathematical-modelling/2024-7-1-3
9. Martsenyuk V.P. Identification of parameters in SIR models based on the data of the COVID-19 pandemic in the Ternopil region // Infectious diseases. 2020. https://doi.org/10.11603/1681-2727.2020.2.11282
10. Kozioł K., Latosińska J.N., Kozioł M. Fractional-Order SIR Epidemic Model for Transmission of COVID-19 // Applied Sciences. 2020. Vol. 10(23), 8316. https://doi.org/10.3390/app10238316
11. Banerjee R., Bhattacharyya R., Roy T.K. Fractional optimal control of compartmental SIR model for COVID-19 with Caputo's fractional derivative // IFAC-PapersOnLine. 2022. Vol. 55(2), P. 349–354. https://doi.org/10.1016/j.ifacol.2022.04.101
12. Alshomrani A.S., Khan A., Ahmad I., Baleanu D. Caputo SIR model for COVID-19 under optimized fractional order // Chaos, Solitons & Fractals. 2021. Vol. 143, 110619. https://doi.org/10.1186/s13662-021-03345-5
13. Danane J., Hammouch Z., Allali K., Rashid S., Singh J. A fractional-order model of coronavirus disease 2019 (COVID-19) with governmental action and individual reaction // Mathematical Methods in the Applied Sciences. 2021. P. 1–14. https://doi.org/10.1002/mma.7759
14. Raissi M., Perdikaris P., Karniadakis G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations // Journal of Computational Physics. 2019. Vol. 378, P. 686–707. https://doi.org/10.1016/j.jcp.2018.10.045
15. Our World in Data COVID-19 dataset [OWID, 2020–2024]. https://github.com/owid/covid-19-data/tree/master/public/data