



Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational
Mathematics (IMACM)

Preprint BUW-IMACM 26/11

Achraf Zinihi, Matthias Ehrhardt and Moulay Rchid Sidi Ammi

**A Koopman-PINN Framework for Epidemic Models:
Parameter Inference and Forecasting**

June 13, 2026

<http://www.imacm.uni-wuppertal.de>

A Koopman-PINN Framework for Epidemic Models: Parameter Inference and Forecasting

Achraf Zinihi^{a,b}, Matthias Ehrhardt^{a,*}, Moulay Rchid Sidi Ammi^b

^a*University of Wuppertal, Applied and Computational Mathematics,
Gaußstrasse 20, 42119 Wuppertal, Germany*

^b*Department of Mathematics, AMNEA Group, Faculty of Sciences and Techniques,
Moulay Ismail University of Meknes, Errachidia 52000, Morocco*

Abstract

We propose a Koopman-enhanced physics-informed neural network (K-PINN) framework for parameter inference and forecasting in nonlinear epidemic models. This method combines Koopman operator theory and physics-informed learning. It maps epidemic states into a latent observable space where the dynamics evolve approximately linearly while satisfying the governing epidemic equations through automatic differentiation. This integration improves interpretability, parameter identifiability, and long-term predictive stability.

We apply the proposed framework to a normalized SEIRSD epidemic model and evaluate it using synthetic monkeypox (Mpox) data and real-world datasets from Germany, Morocco, and Sweden for the SARS-CoV-2 virus. Synthetic trajectories are generated using a structure-preserving, nonstandard finite difference scheme to ensure reliable training data. Numerical results demonstrate that K-PINN achieves more accurate parameter estimation, trajectory reconstruction, and long-term forecasting than classical PINNs and Koopman-EDMD approaches.

These results suggest that K-PINN is an effective machine learning framework for epidemic modeling that can be extended to more complex systems.

Keywords: Epidemic modeling, Koopman operator, Physics-informed neural networks, Nonstandard finite difference method, Parameter estimation

2020 Mathematics Subject Classification: 92D30, 68T07, 92-10.

1. Introduction

Mathematical epidemiology is fundamental to understanding, predicting, and controlling the spread of infectious diseases. Classical compartmental models have been widely used to describe the transmission dynamics of diseases such as influenza, Ebola, and SARS-CoV-2. Due to their interpretability and mathematical structure, these models offer valuable insights into epidemic propagation, stability analysis, intervention strategies, and long-term forecasting. However, practical epidemiological modeling is challenging due to incomplete observations, noisy and sparse data, underreporting, and the nonlinear nature of epidemic dynamics. These difficulties often complicate parameter estimation, hidden-state reconstruction, and reliable forecasting, especially when real epidemiological data is limited or uncertain.

In recent years, physics-informed neural networks (PINNs) have emerged as a powerful machine learning framework in physics for solving forward and inverse problems governed by

*Corresponding author

Email addresses: a.zinihi@edu.umi.ac.ma (Achraf Zinihi), ehrhardt@uni-wuppertal.de (Matthias Ehrhardt), rachidsidiammi@yahoo.fr (Moulay Rchid Sidi Ammi)

differential equations. PINNs were introduced by Raissi, Perdikaris, and Karniadakis in 2019 [1] and integrate physical laws directly into neural network training through automatic differentiation. This allows the learned solutions to satisfy the governing differential equations while simultaneously fitting observational data. This approach has demonstrated strong capabilities in parameter estimation, hidden-state identification, and forecasting for nonlinear dynamical systems. In epidemiological applications, PINNs have shown promising performance in inferring epidemic parameters and in data-driven reconstruction from sparse observations. However, classical PINN frameworks may suffer from training instability, parameter identifiability issues, and reduced long-term predictive accuracy when epidemic dynamics become highly nonlinear or observational data is scarce.

In parallel with these developments, Koopman operator theory has attracted increasing attention in the fields of dynamical systems and data-driven modeling. Originally introduced by Koopman [2] and further developed by Mezić [3], the Koopman framework provides a linear operator-theoretic representation of nonlinear dynamical systems by acting on observable functions instead of state variables directly. This enables nonlinear systems to be analyzed using linear spectral techniques and operator theory. Data-driven Koopman approaches, including *dynamic mode decomposition* (DMD) and *extended dynamic mode decomposition* (EDMD), have been widely applied to prediction, model reduction, and system identification [4, 5]. Recently, deep learning architectures have been introduced to automatically learn Koopman-invariant latent representations and nonlinear observable mappings [6–9]. However, many existing approaches rely on handcrafted observables or purely data-driven embeddings that may not preserve the physical and epidemiological structure of the underlying system.

Several recent works have explored advanced mathematical and computational methodologies for epidemic modeling, motivated by these challenges. These works include [10–15]. For instance, the authors’ studies [12, 13] have developed advanced computational frameworks for analyzing complex epidemic dynamics in spatially distributed and interconnected population systems. In [12], a constraint-aware PINN approach was proposed for reaction-diffusion epidemic models. In another study [13], a hybrid multi-city epidemic modeling strategy was introduced, combining topology-based transportation networks, self-consistent coupling parameters, and distance-dependent temporal migration delays to capture the spread of infectious diseases across interconnected urban centers. In [14], a PINN framework was developed for fractional-order epidemic models where the Caputo fractional derivative was incorporated into the learning process to estimate epidemiological parameters and the fractional memory order simultaneously. In another study conducted by Zinihi, Ehrhardt, and Sidi Ammi [15], an epidemic model incorporating waning immunity and disease-induced mortality was developed and analyzed within the Koopman operator framework. The study demonstrated the ability of Koopman-based approaches to capture nonlinear epidemic dynamics and accurately predict important outbreak features, such as peak infection behavior, using EDMD together with epidemiologically informed observables.

Based on these developments, this paper proposes a *Koopman-enhanced physics-informed neural network* (K-PINN) framework for nonlinear epidemic systems. To the best of our knowledge, this is the first attempt to combine Koopman-based latent linearization and physics-informed learning for epidemic compartmental models. This methodology leverages the advantages of Koopman analysis, such as operator theory, and the capabilities of PINNs, such as equation enforcement, through a neural lifting architecture. This architecture maps epidemic states into a Koopman-compatible latent space with approximately linear evolution. The governing epidemic equations are simultaneously enforced through a physics-informed loss function based on automatic differentiation. This framework improves parameter identifiability, latent stability, and long-term forecasting while preserving the mechanistic structure of epidemic dy-

namics.

The main contributions of this work are as follows: First, we propose a Koopman-enhanced physics-informed neural network (K-PINN) framework for nonlinear epidemic systems. This framework uses a neural lifting architecture to automatically learn Koopman observables and map epidemic dynamics into a latent representation with approximately linear evolution. Meanwhile, the governing epidemic equations are enforced through a physics-informed optimization procedure based on automatic differentiation. This enables a robust inverse-learning strategy for parameter estimation and forecasting under sparse and noisy observations. We further benchmark the proposed methodology against classical Koopman-EDMD (K-EDMD) and standard PINN approaches and validate it on SEIRSD epidemic dynamics using synthetic and real epidemiological datasets.

In addition, structure-preserving numerical schemes, such as *nonstandard finite difference* (NSFD) methods, provide stable and positivity-preserving epidemic simulations. This makes them particularly suitable for generating reliable synthetic training data [16, 17].

The remainder of this paper is organized as follows. Section 2 introduces the proposed epidemic model together with its formulation, normalization, and well-posedness analysis. Section 3 presents the proposed K-PINN architecture and the associated physics-informed loss functions. Section 4 describes the numerical methodology, training procedure, and synthetic data generation strategy. Section 5 presents numerical experiments and forecasting results using both synthetic and real epidemiological datasets. Finally, Section 6 concludes the paper and discusses future research directions.

2. Proposed Model and Mathematical Formulation

We consider a compartmental epidemic model that divides the population into five epidemiological classes: susceptible $S(t)$, exposed $E(t)$, infected $I(t)$, recovered $R(t)$, and deceased $D(t)$. Let

$$N_L(t) = S(t) + E(t) + I(t) + R(t),$$

denote the *total living population* at time t . The disease transmission dynamics are described by the following nonlinear SEIRSD system

$$\begin{cases} \dot{S}(t) = -\beta \frac{S(t)I(t)}{N_L(t)} + \alpha R(t), \\ \dot{E}(t) = \beta \frac{S(t)I(t)}{N_L(t)} - \sigma E(t), \\ \dot{I}(t) = \sigma E(t) - (\gamma + \mu)I(t), \\ \dot{R}(t) = \gamma I(t) - \alpha R(t), \\ \dot{D}(t) = \mu I(t), \end{cases} \quad (1)$$

supplied with the initial conditions

$$S(0) = S_0 > 0, \quad E(0) = E_0 > 0, \quad I(0) = I_0 > 0, \quad R(0) = R_0 \geq 0, \quad D(0) = D_0 \geq 0. \quad (2)$$

The parameter $\beta > 0$ represents the effective transmission rate, $\sigma > 0$ denotes the progression rate from the exposed to the infected class, $\gamma > 0$ is the recovery rate, $\mu > 0$ corresponds to the disease-induced mortality rate, and $\alpha \geq 0$ characterizes the loss of acquired immunity. Recovered individuals return to the susceptible compartment when immunity wanes. Permanent immunity is achieved in the special case $\alpha = 0$.

The infection process is governed by the incidence term $\beta S(t)I(t)/N_L(t)$, which assumes frequency-dependent transmission. Susceptible individuals become exposed through effective contact with infected individuals. Exposed individuals progress to the infected stage at a rate σ , and infected individuals either recover or die from the disease at rates γ and μ , respectively. Recovered individuals may lose immunity and re-enter the susceptible class at rate α , while the deceased compartment accumulates disease-induced fatalities and does not participate in disease transmission. Figure 1 presents a schematic representation of these transitions. The nonlinear nature of the model (1) originates from the bilinear transmission term and the dependence of the incidence rate on the time-varying living population $N_L(t)$. Consequently, the model exhibits richer dynamic behavior than classical epidemic models and provides a realistic framework for studying outbreaks in the presence of disease-induced mortality and temporary immunity.

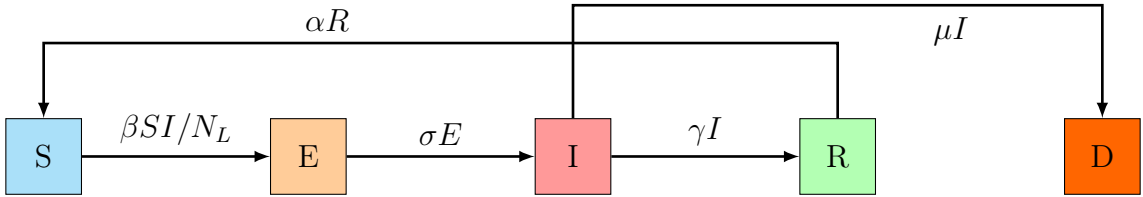


Figure 1: Flow diagram of the proposed SEIRSD model.

Remark 1. *The model (1) does not incorporate vital dynamics, such as natural births and natural deaths. This simplification is justified because epidemic outbreaks generally occur over time horizons that are significantly shorter than those associated with demographic evolution. Under this assumption, population changes are driven solely by disease-related transitions, while the total population remains constant throughout the study period. This formulation allows the analysis to focus solely on the mechanisms governing disease transmission, progression, recovery, and mortality.*

To establish the basic dynamical properties of (1), we first consider the *total population* $N(t) = S(t) + E(t) + I(t) + R(t) + D(t)$. Summing all the equations of (1) yields

$$\dot{N}(t) = \dot{S}(t) + \dot{E}(t) + \dot{I}(t) + \dot{R}(t) + \dot{D}(t) \stackrel{(1)}{=} 0. \quad (3)$$

Therefore, the total population remains constant throughout the epidemic evolution, i.e. $N(t) = N_0$. As a consequence, all trajectories of system (1) remain bounded in the positively invariant set

$$\Omega = \left\{ (S, E, I, R, D) \in \mathbb{R}_+^5 : S + E + I + R + D = N_0 \right\}.$$

Furthermore, $N_L(t) = N(t) - D(t)$, so differentiation yields $\dot{N}_L(t) = -\mu I(t)$. Since $I(t) \geq 0$ and $\mu > 0$, it follows that $\dot{N}_L(t) \leq 0$, showing that the living population decreases over time. However, since $N_L(0) > 0$ and deaths occur continuously, the quantity $N_L(t)$ remains strictly positive for every finite time interval. Standard existence and uniqueness results apply because the right-hand side of system (1) is locally Lipschitz in the biologically feasible region where $N_L(t) > 0$. Combining local existence with the boundedness and positivity properties established above guarantees global solvability. A detailed proof of the existence, uniqueness, positivity, and boundedness of solutions can be obtained by following the methodology developed in [14, pp. 4–6] and [15, pp. 3–5]. Therefore, only the main result is stated here. For completeness, the main results are summarized below.

Corollary 1. *For any admissible parameter set $\beta, \sigma, \gamma, \mu > 0$, and $\alpha \geq 0$, together with initial conditions that satisfy (2), system (1) admits a unique global solution that remains nonnegative and bounded for all $t \geq 0$.*

Since the deceased compartment does not directly affect the dynamics of the remaining epidemiological classes, except through the living population, it is convenient to reformulate the model in terms of population fractions. We therefore introduce the following dimensionless variables

$$s(t) = \frac{S(t)}{N}, \quad e(t) = \frac{E(t)}{N}, \quad i(t) = \frac{I(t)}{N}, \quad r(t) = \frac{R(t)}{N}, \quad d(t) = \frac{D(t)}{N}.$$

Using the conservation property (3), the living population can be expressed as

$$N_L(t) = S(t) + E(t) + I(t) + R(t) = N - D(t) = N(1 - d(t)),$$

Substituting the normalized variables into (1) and dividing by N yields

$$\begin{cases} \dot{s}(t) = -\beta \frac{s(t)i(t)}{1-d(t)} + \alpha r(t), \\ \dot{e}(t) = \beta \frac{s(t)i(t)}{1-d(t)} - \sigma e(t), \\ \dot{i}(t) = \sigma e(t) - (\gamma + \mu)i(t), \\ \dot{r}(t) = \gamma i(t) - \alpha r(t), \end{cases} \quad (4)$$

with

$$s(0) = s_0, \quad e(0) = e_0, \quad i(0) = i_0, \quad r(0) = r_0, \quad d(0) = d_0, \quad (5)$$

where

$$d(t) = 1 - s(t) - e(t) - i(t) - r(t). \quad (6)$$

Since $\dot{s} + \dot{e} + \dot{i} + \dot{r} + \dot{d} = 0$, showing that the simplex

$$\Delta = \left\{ (s, e, i, r, d) \in \mathbb{R}_+^5 : s + e + i + r + d = 1 \text{ and } d < 1 \right\},$$

is forward invariant. Therefore, if the initial conditions belong to Δ , then all state variables remain nonnegative and bounded by 1.

The normalized system (4)–(5) serves as the reference model for the subsequent Koopman analysis and the development of the proposed K–PINN framework. This formulation offers several important advantages: (i) The state variables $s(t)$, $e(t)$, $i(t)$, $r(t)$, and $d(t)$ represent population proportions and are therefore naturally bounded in the interval $[0, 1]$. This improves numerical conditioning and facilitates the enforcement of positivity and conservation constraints. (ii) The use of dimensionless variables eliminates the dependence on the total population size. This allows epidemiological parameters and dynamical behaviors to be interpreted and compared more consistently across different populations and datasets. (iii) From a Koopman operator perspective, normalized variables reduce scaling disparities among observables and provide a more homogeneous representation of the dynamics. This improves the construction of Koopman-invariant latent spaces and the learning of linear representations of the underlying nonlinear dynamics. (iv) The factor $(1 - d(t))^{-1}$ explicitly incorporates the effect of disease-induced mortality through the time-varying living population. This preserves the epidemiological realism of the transmission process while maintaining a scale-free formulation. For these reasons, all subsequent theoretical developments, Koopman formulations, and K–PINN constructions are based on the normalized system (4)–(5).

3. Koopman-Enhanced Physics-Informed Neural Networks

The normalized model introduced in Section 2 provides a biologically meaningful and mathematically consistent description of epidemic dynamics. However, it remains difficult to estimate unknown epidemiological parameters and forecast long-term behavior from sparse or noisy observations. The dynamics are nonlinear, only partially observed, and sensitive to parameter uncertainty. Classical numerical fitting methods may capture observed data locally, yet they fail to preserve the underlying epidemic structure. Standard neural networks can approximate trajectories but do not respect differential equations.

PINNs improve this situation by embedding the governing equations into the learning process. Nevertheless, for strongly nonlinear epidemic systems, PINNs may suffer from optimization stiffness, identifiability issues, and limited extrapolation ability. Koopman operator theory offers a complementary perspective by lifting the nonlinear dynamics into a space of observables, in which the evolution is approximately linear. This linear latent evolution can improve interpretability and long-term prediction.

We combine these ideas through a K-PINN framework. The proposed method uses a single, epidemiologically-informed observable representation; evolves the lifted variables under a finite-dimensional Koopman operator; reconstructs the original epidemic state via a decoder; and enforces the normalized SEIRSD equations through physics-based regularization. This produces a unified inverse-learning model that improves accuracy, leverages domain knowledge, and enables robust epidemic reconstruction, parameter inference, and forecasting.

3.1. Proposed Observable Dictionary

Let

$$\mathbf{x}(t) = [s(t), e(t), i(t), r(t), d(t)]^\top \in \mathbb{R}^5,$$

be the normalized epidemic state. To represent the nonlinear structure of the SEIRSD system (4)–(6) in a lifted space, we define a single observable dictionary

$$\Phi(\mathbf{x}) = [\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \dots, \phi_m(\mathbf{x})]^\top \in \mathbb{R}^m.$$

A suitable epidemiological dictionary may include the basic compartments together with selected nonlinear interaction terms, such as

$$\Phi(\mathbf{x}) = [s, e, i, r, d, si, se, sr, ei, er, ir, \frac{si}{1-d}, s^2, e^2, i^2, r^2, d^2]^\top.$$

The goal is to provide a compact lifted representation that retains interpretability while capturing the essential nonlinear couplings of the epidemic process.

In the K-PINN framework, the lifting function $\Phi(\mathbf{x})$ determines the observable space in which the nonlinear epidemic dynamics are approximated by a finite-dimensional Koopman operator. For the SEIRSD model (4), we augment the set of observables beyond the bilinear transmission term si by introducing several quadratic and cross-product observables, such as $sr, ei, er, ir, s^2, e^2, i^2, r^2$, and d^2 . While only a subset of these observables admits a direct epidemiological interpretation, such as the infection incidence term si , the other terms are included to enrich the functional basis and improve the approximation of the underlying Koopman-invariant subspace. These higher-order observables provide additional flexibility in representing nonlinear interactions and are widely used in reduced-order and nonlinear dynamical system modeling. Furthermore, quadratic terms, such as s^2, e^2 , and i^2 can be interpreted as capturing density-dependent or feedback effects in epidemiological dynamics [15]. Unlike multi-dictionary K-EDMD studies, the present framework relies on one coherent observable set. This is sufficient because the physics-informed loss and the latent linear dynamics jointly guide the learning process.

3.2. Latent Koopman Dynamics

The observable vector is mapped to a latent Koopman representation by

$$\mathbf{z}(t) = \Phi_\theta(\mathbf{x}(t)),$$

where Φ_θ denotes either a fixed dictionary map or a trainable neural lifting operator. The latent state $\mathbf{z}(t) \in \mathbb{R}^m$ is assumed to evolve according to

$$\frac{d\mathbf{z}(t)}{dt} = K\mathbf{z}(t),$$

where $K \in \mathbb{R}^{m \times m}$ is the Koopman matrix. This yields the formal solution

$$\mathbf{z}(t) = e^{Kt}\mathbf{z}(0).$$

The linear latent representation provides a structured reduced-order description of the epidemic dynamics and enables stable forecasting beyond the observation interval.

3.3. State Reconstruction

To recover the epidemic variables from the lifted coordinates, we introduce a decoder network

$$\hat{\mathbf{x}}(t) = \Psi_\omega(\mathbf{z}(t)),$$

where $\Psi_\omega: \mathbb{R}^m \rightarrow \mathbb{R}^5$ is a neural network with parameters ω . The full lifted reconstruction is therefore

$$\hat{\mathbf{x}}(t) = \Psi_\omega(\Phi_\theta(\mathbf{x}(t))).$$

The decoder maps the latent observables back to physically meaningful epidemic compartments.

3.4. Physics-Informed Residuals

Let

$$\hat{\mathbf{x}}(t) = [\hat{s}(t), \hat{e}(t), \hat{i}(t), \hat{r}(t), \hat{d}(t)]^\top,$$

denote the reconstructed solution. The normalized SEIRSD residual operator is defined by

$$\begin{aligned} \mathcal{R}(t) &:= \mathcal{R}(\hat{\mathbf{x}}(t)) = [\mathcal{R}_s(t), \mathcal{R}_e(t), \mathcal{R}_i(t), \mathcal{R}_r(t), \mathcal{R}_d(t)]^\top, \\ &= \frac{d\hat{\mathbf{x}}}{dt} - F(\hat{\mathbf{x}}), \end{aligned}$$

where F denotes the right hand side of the system (4). These residuals are computed using automatic differentiation and are used to penalize deviations from the governing SEIRSD equations during training.

3.5. Composite Loss Function

The K-PINN training objective is given by

$$\mathcal{L} = \lambda_d \mathcal{L}_{\text{data}} + \lambda_k \mathcal{L}_{\text{koop}} + \lambda_p \mathcal{L}_{\text{phys}} + \lambda_c \mathcal{L}_{\text{cons}},$$

where $\lambda_d, \lambda_k, \lambda_p, \lambda_c > 0$. Let N_d be the number of observed epidemic snapshots, and N_r be the number of collocation points used to evaluate the residuals of the governing equations. Thus,

- The data reconstruction loss is given by

$$\mathcal{L}_{\text{data}} = \frac{1}{N_d} \sum_{j=1}^{N_d} \|\mathbf{x}_j - \hat{\mathbf{x}}_j\|_2^2.$$

- The Koopman consistency loss is given by

$$\mathcal{L}_{\text{koop}} = \frac{1}{N_d} \sum_{j=1}^{N_d} \left\| \frac{d\mathbf{z}_j}{dt} - K\mathbf{z}_j \right\|_2^2.$$

- The physics loss is given by

$$\mathcal{L}_{\text{phys}} = \frac{1}{N_r} \sum_{j=1}^{N_r} \left\| \mathcal{R}(t_j) \right\|_2^2.$$

- The constraint term is given by

$$\mathcal{L}_{\text{cons}} = \mathcal{L}_{\text{pos}} + \mathcal{L}_{\text{param}},$$

where \mathcal{L}_{pos} penalizes negative population values and $\mathcal{L}_{\text{param}}$ regularizes the epidemiological parameters.

3.6. Epidemiological Constraints

Since the state variables are normalized population fractions, they satisfy

$$0 \leq s(t), e(t), i(t), r(t), d(t) \leq 1 \quad \text{and} \quad s(t) + e(t) + i(t) + r(t) + d(t) = 1.$$

To ensure admissible parameter values, we can parameterize

$$\beta = \exp(\tilde{\beta}), \quad \delta = \exp(\tilde{\delta}), \quad \gamma = \exp(\tilde{\gamma}), \quad \text{and} \quad \mu = \exp(\tilde{\mu}),$$

with unconstrained variables $\tilde{\beta}, \tilde{\delta}, \tilde{\gamma}, \tilde{\mu} \in \mathbb{R}$.

3.7. Inverse Learning Problem

Let the full set of trainable parameters be

$$\Theta = \{\theta, \omega, K, \beta, \delta, \gamma, \mu\}.$$

The inverse problem is formulated as

$$\Theta^* = \arg \min_{\Theta} \mathcal{L}(\Theta).$$

This single optimization simultaneously learns the observable map, the Koopman operator, the decoder, and the epidemiological parameters.

4. Numerical Methodology

The numerical experiments are designed to evaluate the K-PINN proposed for the system under different observation regimes. The main objectives are to reconstruct epidemic trajectories, estimate unknown parameters, and forecast future dynamics based on sparse or noisy observations. Depending on the experiment, the model may use full-state data, partial observations, or perturbed measurements. To ensure consistency between the mathematical model, the learning framework, and the numerical experiments, the same normalized formulation is used throughout.

4.1. Synthetic Data Generation via a Nonstandard Finite Difference Scheme

Synthetic trajectories are generated using an NSFD scheme, which is particularly suitable for epidemic models because it preserves positivity, boundedness, structural consistency, and dynamical consistency. Unlike standard time discretizations, which may introduce nonphysical behavior such as negative compartment values, NSFD methods are designed to avoid such artifacts.

Let $\{t_0, t_1, \dots, t_M\}$ be a uniform time grid with a step size $k = \Delta t$. In the NSFD framework, the derivative of a generic variable $u(t)$ is approximated by

$$\dot{u}(t_n) \approx \frac{u^{n+1} - u^n}{\varphi(k)},$$

where $\varphi(k)$ is a nontrivial denominator function that satisfies

$$\varphi(k) > 0, \quad \text{and} \quad \varphi(k) = k + \mathcal{O}(k^2).$$

A common choice is

$$\varphi(k) = \frac{e^{\eta k} - 1}{\eta},$$

where η is a stabilization parameter obtained by comparing with the asymptotic behavior of the total population [18]. In the absence of natural mortality, one may simply set $\eta = 0$, yielding $\varphi(k) = k$.

Let $s^n \approx s(t_n)$, $e^n \approx e(t_n)$, $i^n \approx i(t_n)$, $r^n \approx r(t_n)$, and $d^n \approx d(t_n)$. The NSFD discretization of the SEIRSD system (4) is written as

$$\begin{cases} \frac{s^{n+1} - s^n}{\varphi(k)} = -\beta \frac{s^{n+1} i^n}{1 - d^n} + \alpha r^n, \\ \frac{e^{n+1} - e^n}{\varphi(k)} = \beta \frac{s^{n+1} i^n}{1 - d^n} - \sigma e^{n+1}, \\ \frac{i^{n+1} - i^n}{\varphi(k)} = \sigma e^{n+1} - (\gamma + \mu) i^{n+1}, \\ \frac{r^{n+1} - r^n}{\varphi(k)} = \gamma i^{n+1} - \alpha r^{n+1}. \end{cases}$$

Afterwards,

$$\begin{cases} s^{n+1} = \frac{s^n + \alpha \varphi(k) r^n}{1 + \beta \varphi(k) \frac{i^n}{1 - d^n}}, \\ e^{n+1} = \frac{e^n + \beta \varphi(k) \frac{s^{n+1} i^n}{1 - d^n}}{1 + \sigma \varphi(k)}, \\ i^{n+1} = \frac{i^n + \sigma \varphi(k) e^{n+1}}{1 + (\gamma + \mu) \varphi(k)}, \\ r^{n+1} = \frac{r^n + \gamma \varphi(k) i^{n+1}}{1 + \alpha \varphi(k)}, \end{cases} \quad (7)$$

with the deceased proportion updated consistently by

$$d^{n+1} = 1 - s^{n+1} - e^{n+1} - i^{n+1} - r^{n+1}. \quad (8)$$

This discretization adheres to standard NSFD design principles. First, the nonlinear incidence term $\beta \frac{si}{1-d}$ is treated semi-implicitly by evaluating s at the new time level and i at

the current time level. This preserves positivity while ensuring that the scheme is solvable sequentially. Second, the recovery and progression terms are written to maintain the correct flow between compartments. Third, the update for d^{n+1} is obtained from the population conservation relation, ensuring that the discrete total population remains equal to one. For more details about the NSFD analysis, we encourage readers to consult [16, 17] and the references therein.

The scheme is designed to preserve the nonnegativity of all compartments, keeping the solution inside the invariant simplex Δ . This makes NSFD trajectories suitable for use as reliable synthetic data for training the proposed K-PINN and for validating its reconstruction and forecasting performance. In the numerical experiments, the time step k is chosen small enough to accurately capture the epidemic dynamics accurately while maintaining computational efficiency. Due to the limited availability of real data, we set $\alpha = 0$ in Section 5.

4.2. Data Collection and Koopman Matrix Estimation

The K-PINN framework requires state trajectory data. This data can be obtained either from: numerical simulations (synthetic data) or empirical epidemiological time series, after appropriate normalization. We discretize the trajectory at uniform sampling times $t_n = nk = n\Delta t$, with $n = 0, 1, \dots, M$. This produces a discrete sequence of state snapshots

$$\mathbf{x}_n = [s(t_n), e(t_n), i(t_n), r(t_n), d(t_n)]^\top.$$

To construct the Koopman representation, each snapshot is mapped into an observable space through a selected set of functions

$$\mathbf{y}_n = [\psi_1(\mathbf{x}_n), \psi_2(\mathbf{x}_n), \dots, \psi_m(\mathbf{x}_n)]^\top,$$

where $\{\psi_j\}_{j=1}^m$ denotes the proposed observable dictionary. In the present work, this dictionary may be chosen to reflect the epidemic structure by including the compartment variables and selected nonlinear interaction terms that capture infection and coupling effects, as described in Section 3.1. The lifted snapshot matrices are then defined as

$$Y = (\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{M-1}) \in \mathbb{R}^{m \times M}, \quad \text{and} \quad Y' = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M) \in \mathbb{R}^{m \times M}.$$

In the lifted space, the Koopman evolution is approximated by

$$\mathbf{y}_{k+1} \approx K\mathbf{y}_k,$$

where $K \in \mathbb{R}^{m \times m}$ is the finite-dimensional Koopman matrix. Consequently, the data matrices satisfy

$$Y' \approx KY.$$

The Koopman matrix is identified by the least-squares problem

$$K = \arg \min_{\tilde{K} \in \mathbb{R}^{m \times m}} \|Y' - \tilde{K}Y\|_F,$$

whose closed-form solution is given by

$$K = Y'Y^*,$$

where Y^* denotes the Moore-Penrose pseudoinverse of Y .

The eigen-decomposition of K provides approximations of the Koopman eigenvalues and modes, which describe the dominant growth, decay, and oscillatory patterns of the epidemic dynamics in the lifted space. These quantities are useful for analyzing the reduced-order structure of the system and for comparing the learned latent dynamics with the proposed K-PINN representation. After computing the Koopman matrix, one can propagate future trajectories in the observable space. This provides an additional numerical benchmark for the model's forecasting performance.

4.3. Training Strategy and Forecasting Procedure

The proposed K-PINN is trained by minimizing the composite loss defined in Section 3. In practice, training begins with a short pretraining phase based on the reconstruction term. This phase helps initialize the encoder, decoder, and latent Koopman operator in a stable regime. Following this initialization stage, the full objective is optimized to enforce data fidelity, physics residuals, Koopman consistency, and admissibility constraints simultaneously.

The training data consist of two sets of points. The first set contains the observed epidemic snapshots used for reconstruction loss. The second set contains collocation points sampled across the temporal domain for evaluating the residuals. Then, automatic differentiation is used to compute the time derivatives of the network outputs, which allows the physics-informed terms to be evaluated without discretization error.

Adam is used for optimization in the early stage because it is robust to noisy gradients and helps the network explore the parameter space. Once the loss has stabilized, L-BFGSS¹ is applied to refine the solution and improve convergence accuracy. This two-stage strategy is effective for PINN-type models, especially when the latent Koopman matrix and epidemiological parameters are learned simultaneously.

After training, the learned latent dynamics are used for forecasting. Starting from the encoded initial condition, the Koopman operator propagates the observables forward in time, and the decoder reconstructs the epidemic compartments at future times. This produces a structured forecasting mechanism that combines linear latent evolution with nonlinear epidemic constraints. Figure 2 summarizes the complete K-PINN pipeline, from NSFD data generation and observable lifting to joint optimization and forecasting.

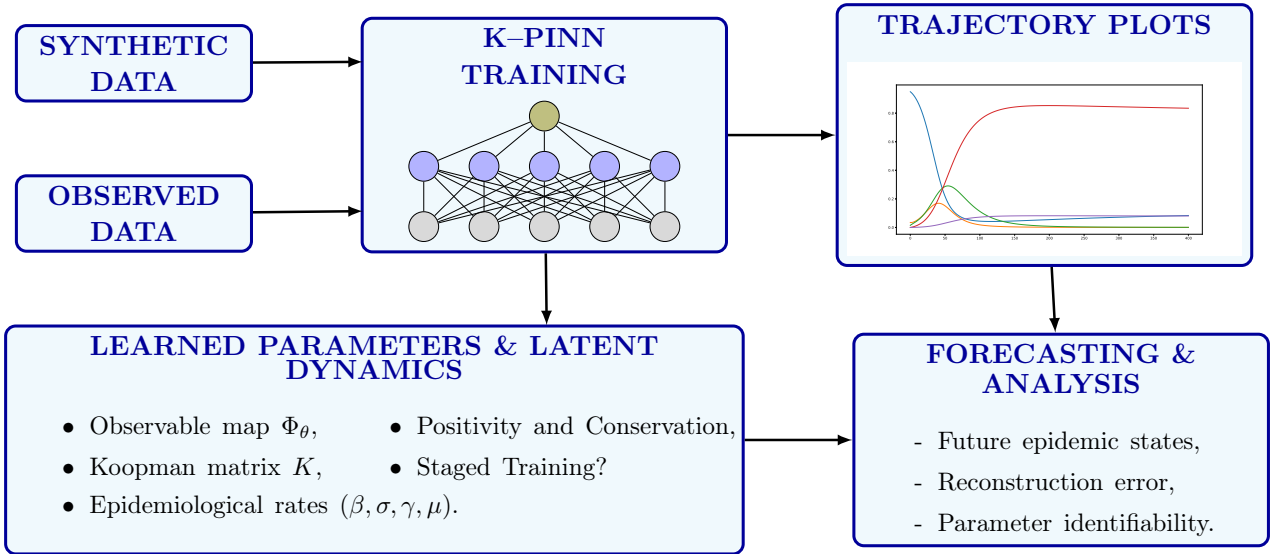


Figure 2: Workflow of the proposed K-PINN framework for epidemic reconstruction and forecasting.

Algorithm 1 outlines the step-by-step training and prediction procedure used in the numerical experiments.

¹A memory-efficient quasi-Newton optimization method used to refine neural network training and accelerate convergence near the optimum

Algorithm 1 Training and Forecasting Procedure for the Proposed K-PINN

- 1: *Input*: Epidemic trajectories, observed snapshots, collocation points, initial parameters.
- 2: *Initialization*: Initialize the encoder Φ_θ , decoder Ψ_ω , Koopman matrix K , and epidemiological parameters $(\beta, \sigma, \gamma, \mu)$.
- 3: *Data preparation*: Sample observation points $\{t_j\}_{j=1}^{N_d}$ and collocation points $\{\tau_j\}_{j=1}^{N_r}$ over the time interval.
- 4: *Pretraining stage*: Minimize the reconstruction loss $\mathcal{L}_{\text{data}}$ using Adam to stabilize the encoder and decoder.
- 5: *Joint training stage*: Minimize the full composite loss

$$\mathcal{L} = \lambda_d \mathcal{L}_{\text{data}} + \lambda_k \mathcal{L}_{\text{koop}} + \lambda_p \mathcal{L}_{\text{phys}} + \lambda_c \mathcal{L}_{\text{cons}}$$

using Adam.

- 6: *Physics evaluation*: Use automatic differentiation to compute the residuals of the normalized system at collocation points.
 - 7: *Koopman update*: Update the latent operator K jointly with the neural parameters.
 - 8: *Refinement stage*: Switch to L-BFGS and continue optimization until convergence.
 - 9: *Forecasting*: Propagate the latent state forward using $\mathbf{z}_{k+1} = K\mathbf{z}_k$, then reconstruct future epidemic states using $\hat{\mathbf{x}}_{k+1} = \Psi_\omega(\mathbf{z}_{k+1})$.
 - 10: *Output*: Reconstructed trajectories, estimated parameters, and future predictions.
-

4.4. Error Metrics

Several standard error metrics are employed to evaluate the accuracy of the proposed framework. For an individual compartment, e.g. the susceptible population, the root mean square error (RMSE) is defined as follows

$$\text{RMSE}(s) = \sqrt{\frac{1}{M+1} \sum_{n=0}^M (s_n - \hat{s}_n)^2},$$

while the mean absolute error (MAE) is given by

$$\text{MAE}(s) = \frac{1}{M+1} \sum_{n=0}^M |s_n - \hat{s}_n|.$$

In addition, the relative error is computed as

$$\text{RelErr}(s) = \frac{\|s - \hat{s}\|_2}{\|s\|_2} \times 100\%.$$

Similar expressions are used for the remaining compartments. To evaluate the reconstruction accuracy of the entire epidemic state, we consider the discrete L^2 error

$$E_{L^2} = \sqrt{\frac{1}{M+1} \sum_{n=0}^M \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|_2^2},$$

and the maximum error

$$E_{L^\infty} = \max_{0 \leq n \leq M} \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|_\infty.$$

These metrics are used to compare the K-PINN with the baseline methods in the numerical results section.

5. Numerical Results

We conducted numerical experiments on both synthetic monkeypox (Mpox) data and real-world SARS-CoV-2 (Covid-19) epidemic data to evaluate the proposed K-PINN framework. Throughout the simulations, we set $\alpha = 0$.

First, we consider synthetic epidemic trajectories generated to emulate Mpox dynamics, where the true parameters $(\beta, \sigma, \gamma, \mu)$ are known. This allows us to rigorously assess the framework’s ability to recover hidden epidemiological parameters from partial and noisy observations using the training procedure described in Section 4. We also compare the K-PINN approach with classical PINNs and K-EDMD methods to benchmark its performance in terms of identifiability and stability of the learned dynamics. After validating the methodology on synthetic data, we apply it to a real-world SARS-CoV-2 dataset from Sweden, focusing on confirmed, recovered, and deceased cases.

Section 5.1 presents results for the synthetic experiments, where data are generated using the NSFD scheme (7)–(8). Section 5.2 reports the results on real data, emphasizing predictive accuracy, parameter inference, and epidemiological interpretability.

5.1. Synthetic Mpox Data Experiments

The epidemiological parameters were selected within biologically plausible ranges reported in recent Mpox studies, as summarized in Table 1, particularly for clade Ib outbreaks during 2024–s2025. We adopted representative values within these intervals to generate the data: $\beta = 0.25$, $\sigma = 0.13$, $\gamma = 0.052$, and $\mu = 0.005$.

Table 1: Epidemiological parameters for the proposed model applied to Mpox (clade Ib, 2024–2025).

Epidemic	β	σ	γ	μ	References
Mpox	0.1 – 0.3	0.077 – 0.2	0.036 – 0.071	0.001 – 0.03	[19, 20]

Figure 3 presents the synthetic trajectories of the susceptible, exposed, infected, recovered, and deceased compartments over the considered time horizon. These trajectories serve as the ground-truth dataset used throughout the numerical experiments.

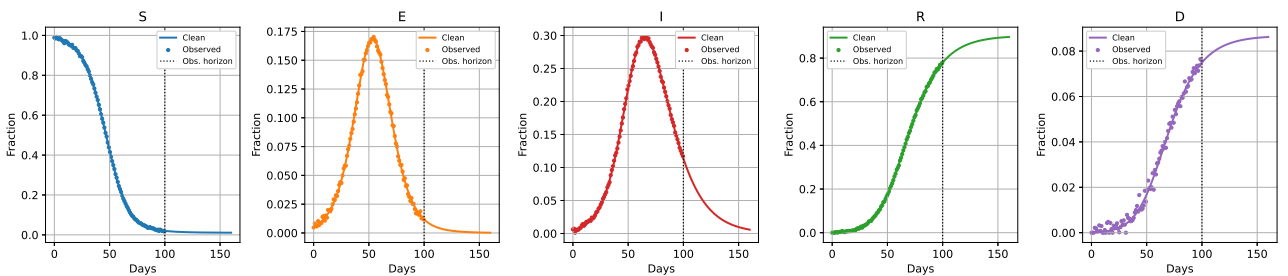


Figure 3: Synthetic epidemic trajectories generated using the NSFD scheme (7)–(8).

As a first benchmark, the epidemic dynamics were approximated using the classical EDMD approach. The observable dictionary introduced in Section 3.1 was employed to create a finite-dimensional approximation of the Koopman operator. Figure 4 compares the reconstructed trajectories and future forecasts obtained through K-EDMD with the reference synthetic data. Although K-EDMD captures the overall qualitative behavior of the epidemic evolution, discrepancies become more pronounced during long-term forecasting due to the lack of explicit enforcement of the governing epidemic equations. Diagnostic quantities associated with this approach are presented in Figure 14.

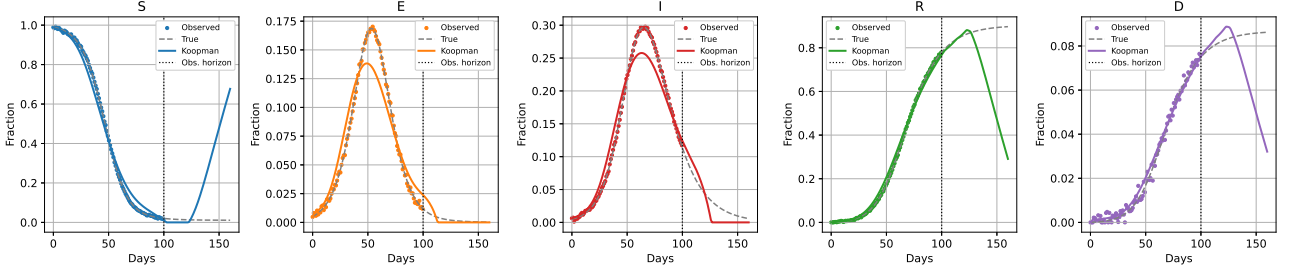


Figure 4: State reconstruction and forecasting using the K-EDMD framework for the synthetic dataset.

Next, we consider a standard PINN implementation in which the normalized equations are enforced via a physics-based loss function, but Koopman latent dynamics are not incorporated. Figure 5 illustrates the reconstructed epidemic trajectories and forecasting results. The PINN framework successfully reconstructs the observed trajectories and provides accurate estimates of the epidemiological parameters, as shown in Figures 15 and 16. However, the forecasting performance may deteriorate over extended prediction horizons due to the difficulty of learning strongly nonlinear epidemic dynamics directly in the original state space.

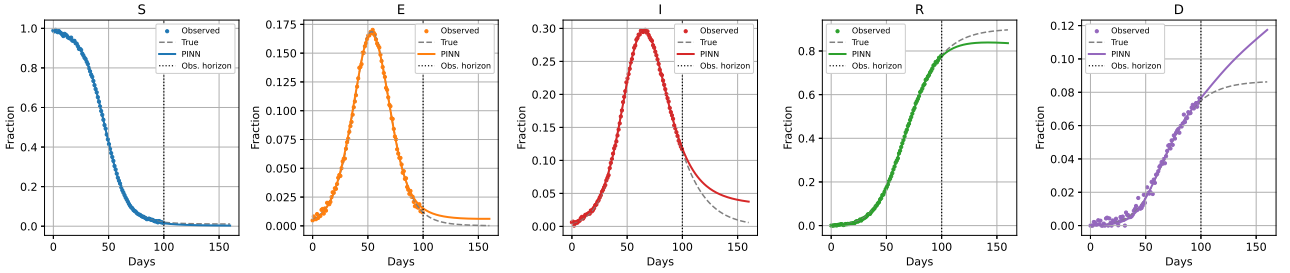


Figure 5: Epidemic state reconstruction and forecasting using the PINN framework for the synthetic dataset.

Finally, the proposed K-PINN framework is applied to the same synthetic dataset. Figure 6 presents the reconstructed epidemic trajectories together with long-horizon forecasts. Figure 7 displays the convergence behavior of the composite loss function, demonstrating the stable optimization of the data, physics, Koopman, and constraint terms. Figure 8 illustrates the evolution of the estimated epidemiological parameters throughout training. The results show that K-PINN accurately reconstructs all epidemic compartments while simultaneously providing reliable parameter estimates and improved forecasting capability. The incorporation of Koopman latent dynamics significantly enhances stability and predictive performance compared with both the standalone K-EDMD and PINN approaches.

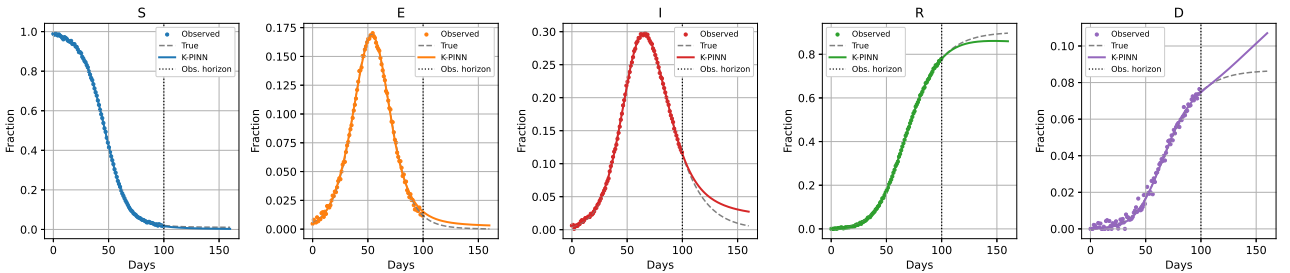


Figure 6: Epidemic state reconstruction and forecasting using the proposed K-PINN framework for the synthetic dataset.

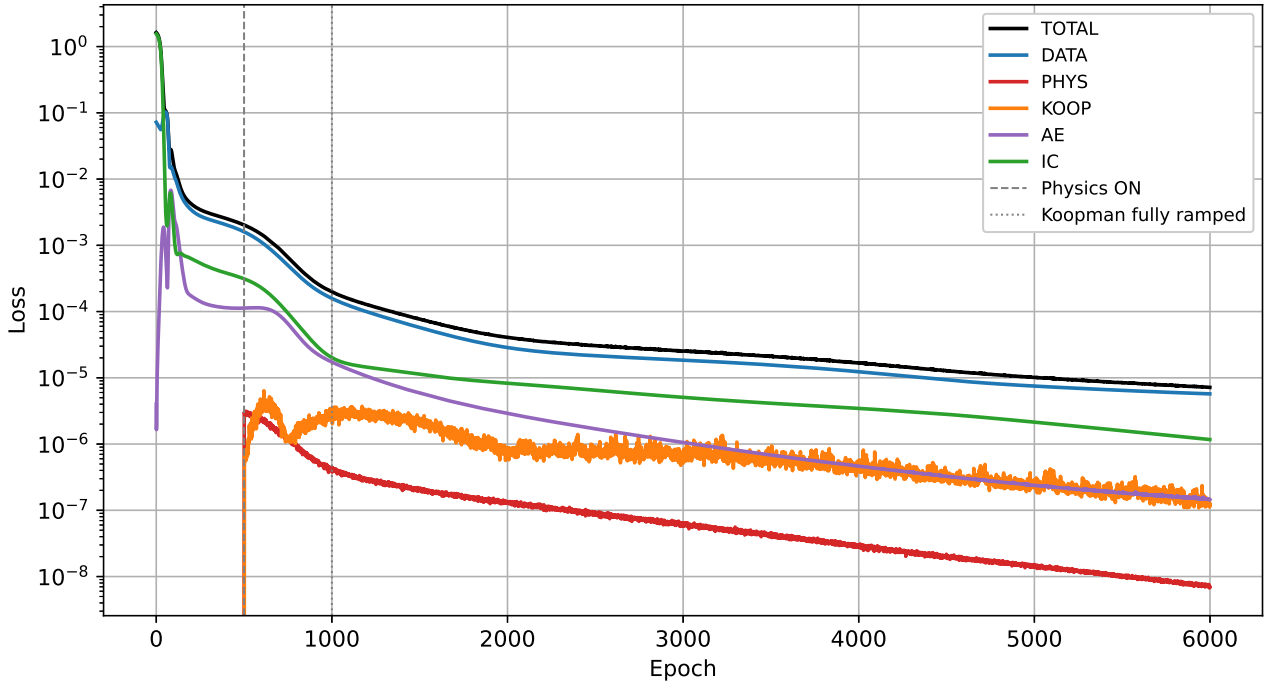


Figure 7: Evolution of the composite K-PINN loss function during training for the synthetic dataset.

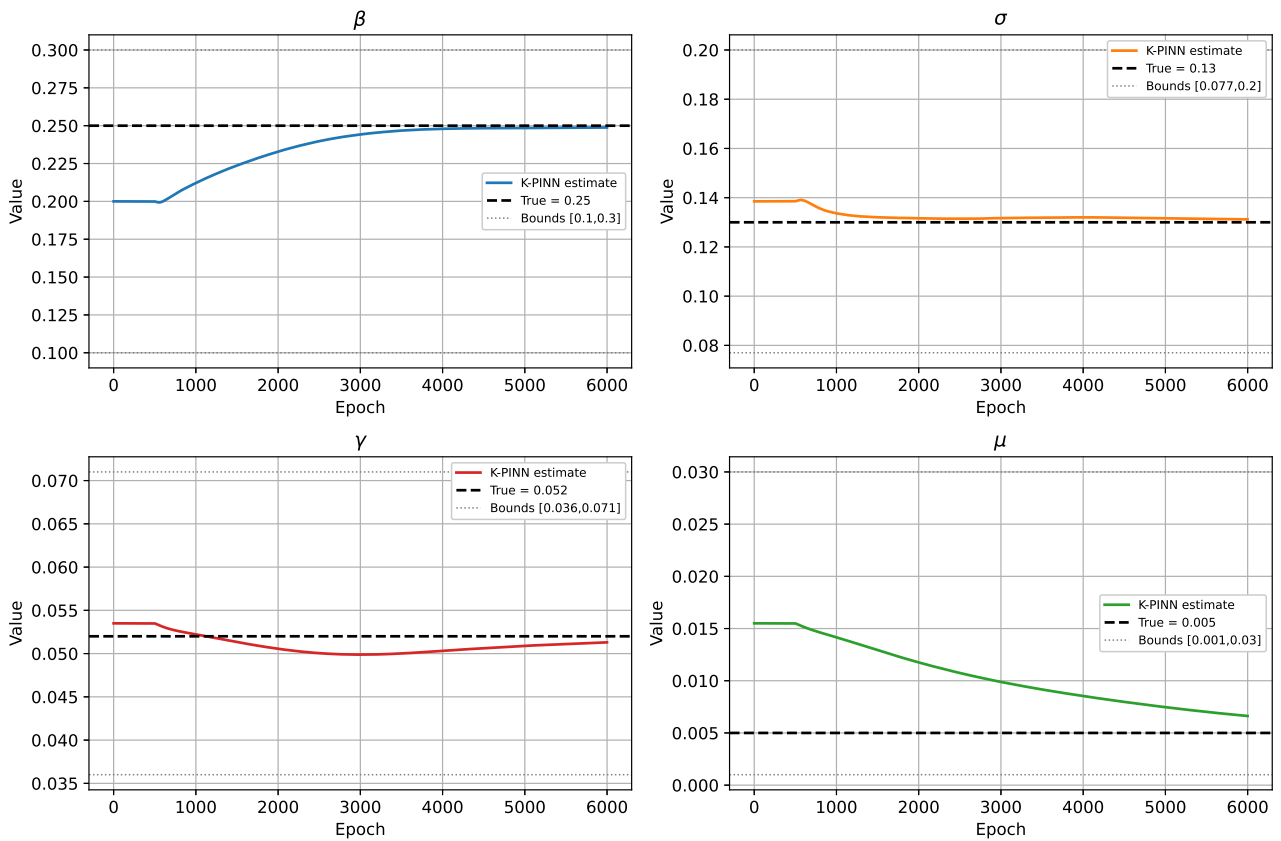


Figure 8: Convergence of epidemiological parameter estimates obtained by the proposed K-PINN framework for the synthetic dataset.

For clarity of presentation, Figure 13 displays the collective reconstruction and forecasting results obtained with the proposed K-PINN framework. Detailed, compartment-wise compar-

isons for all epidemic variables and all competing approaches are provided in Appendix A. Table 2 reports the reconstruction and forecasting errors obtained by the three competing approaches, providing a quantitative assessment of the proposed methodology. Table 3 summarizes the parameter estimation results and compares the recovered epidemiological parameters with their reference values.

Table 2: Comparison of reconstruction and forecasting errors obtained using K-EDMD, PINN, and the proposed K-PINN framework for the synthetic MpoX dataset.

Method	E_{L^2}	E_{L^∞}	RMSE(S)	RMSE(E)	RMSE(I)	RMSE(R)	RMSE(D)
K-EDMD	0.24420	0.66615	0.18123	0.01317	0.01939	0.16138	0.01398
PINN	0.03130	0.06097	0.00456	0.00343	0.01468	0.02469	0.01106
K-PINN	0.01839	0.03722	0.00422	0.00242	0.00966	0.01353	0.00618

Table 3: True and estimated epidemiological parameters for the synthetic SEIRSD experiment.

Parameters	β	σ	γ	μ
True value	0.2500	0.1300	0.0520	0.0050
PINN estimate	0.2479	0.1312	0.0512	0.0067
K-PINN estimate	0.2488	0.1310	0.0513	0.0066

5.2. Real-World COVID-19 Data Experiments

To further evaluate the applicability of the proposed K-PINN framework, we examine real-world epidemic data from Germany, Morocco, and Sweden during the COVID-19 outbreaks. We selected these countries because they have reliable daily case and mortality reports and distinct epidemic dynamics. The corresponding epidemiological parameters are constrained within biologically plausible ranges, as summarized in Table 4.

Table 4: Epidemiological parameters for the proposed model applied to COVID-19 in Europe (2024).

Epidemic	β	σ	γ	μ	References
COVID-19	0.2 – 0.4	0.1 – 0.3	0.05 – 0.1	0.001 – 0.01	[21, 22]

The daily number of confirmed cases and deaths comes from publicly available epidemiological databases [23–25]. These cumulative records are transformed into compartmental trajectories consistent with the SEIRD framework. Specifically, the infected compartment (I) is reconstructed from active case estimates, the deceased compartment (D) is reconstructed from reported mortality data, and the recovered compartment (R) is reconstructed by aggregating recovered and deceased cases when recovery information is available. Since the exposed compartment (E) is not directly observable in the data, it is treated as a latent state variable to be inferred by the proposed model.

Figure 9 presents the real-world trajectories of the spread of COVID-19 in Germany, Morocco, and Sweden, which were used in the numerical experiments. The compartments representing the susceptible, exposed, infected, recovered, and deceased illustrate the different epidemic patterns in these countries and provide a benchmark for evaluating the reconstruction, parameter estimation, and forecasting capabilities of the proposed framework.

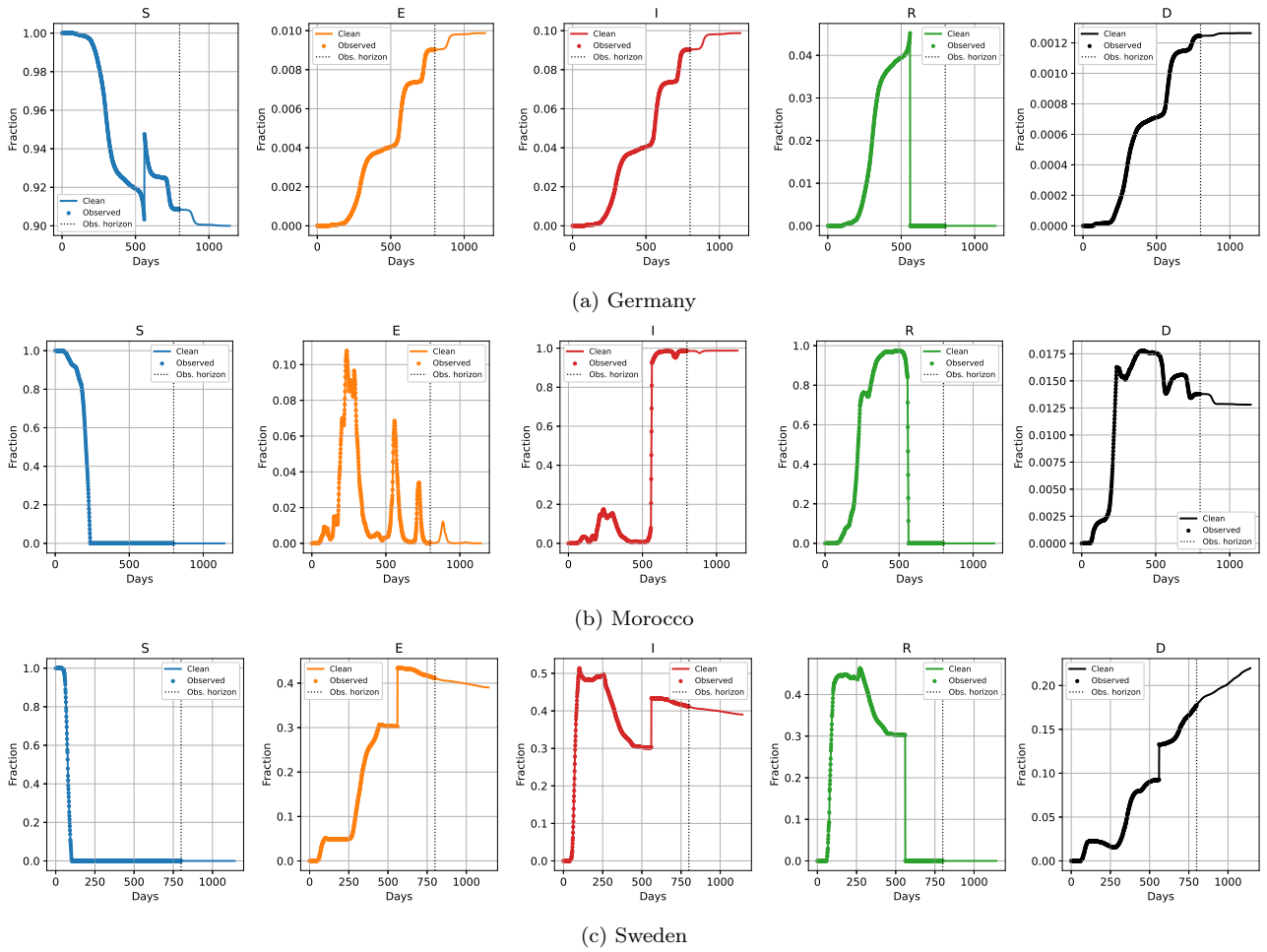


Figure 9: Real-world COVID-19 epidemic trajectories for Germany, Morocco, and Sweden.

Figure 10 presents the reconstruction and forecasting results obtained using the K-EDMD approach for the Germany, Morocco, and Sweden datasets. Although the Koopman-based representation captures the overall qualitative trends of the epidemic evolution, noticeable discrepancies remain in several compartments, especially during the forecasting interval. These results suggest that the finite-dimensional Koopman approximation alone is insufficient for accurately representing the complex nonlinear interactions that govern the epidemic dynamics. The corresponding quantitative errors reported in Table 5 confirm the limited predictive performance of the standalone K-EDMD approach.

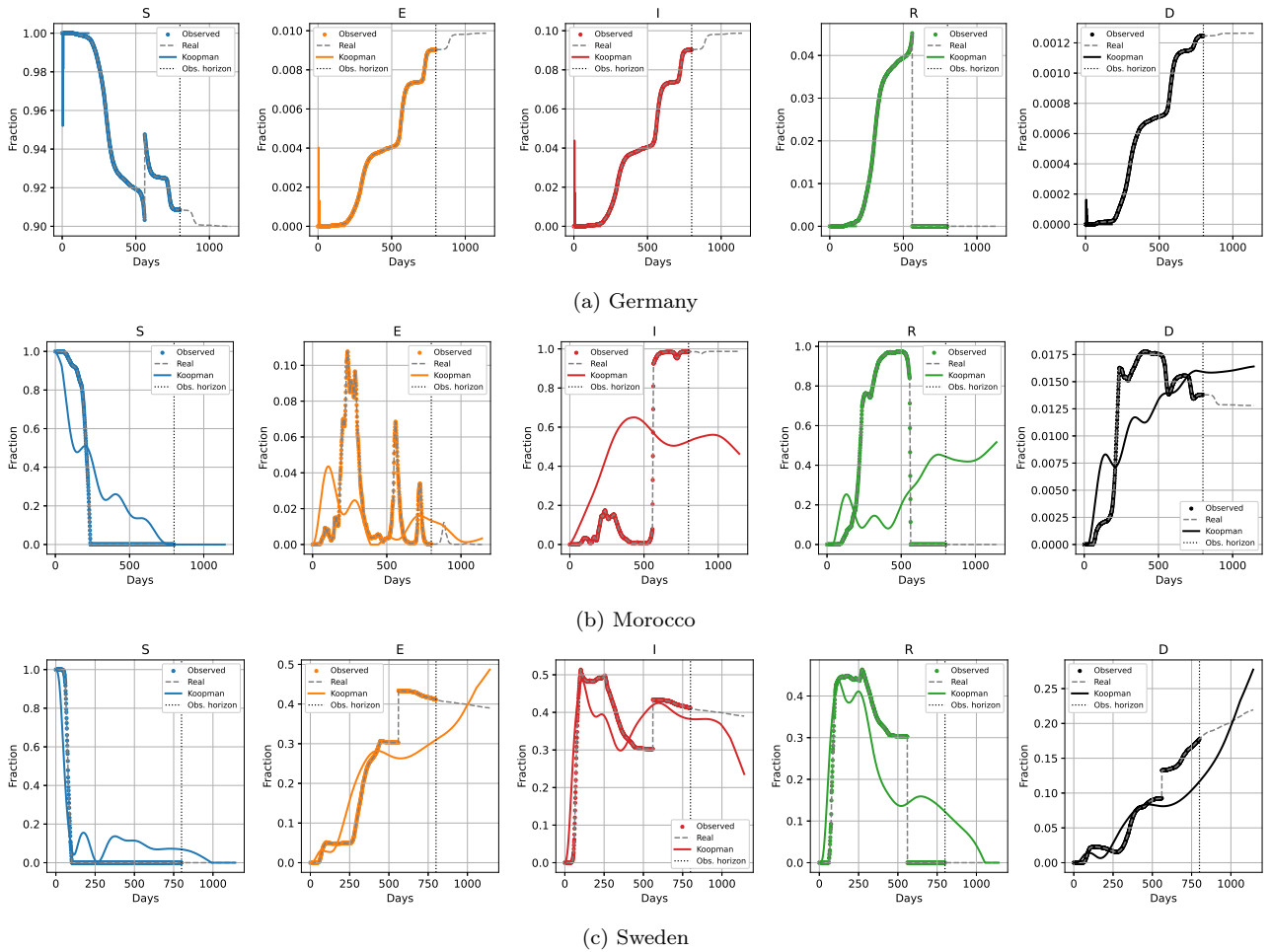


Figure 10: Comparison between the observed epidemiological trajectories and the K-EDMD reconstructions for the Germany, Morocco, and Sweden datasets.

Figure 11 illustrates the reconstruction and forecasting performance of the PINN framework for the three real-world datasets. By incorporating the governing epidemic equations into the learning process, the PINN framework significantly improves the agreement between the predicted and observed trajectories compared with the K-EDMD approach. The reconstructed solutions remain stable throughout both the training and forecasting intervals, accurately capturing the main epidemic trends. However, small deviations are still visible in some compartments, particularly during long-term forecasting, reflecting the challenges of learning highly nonlinear epidemic dynamics from limited observations. The corresponding quantitative errors are reported in Table 5.

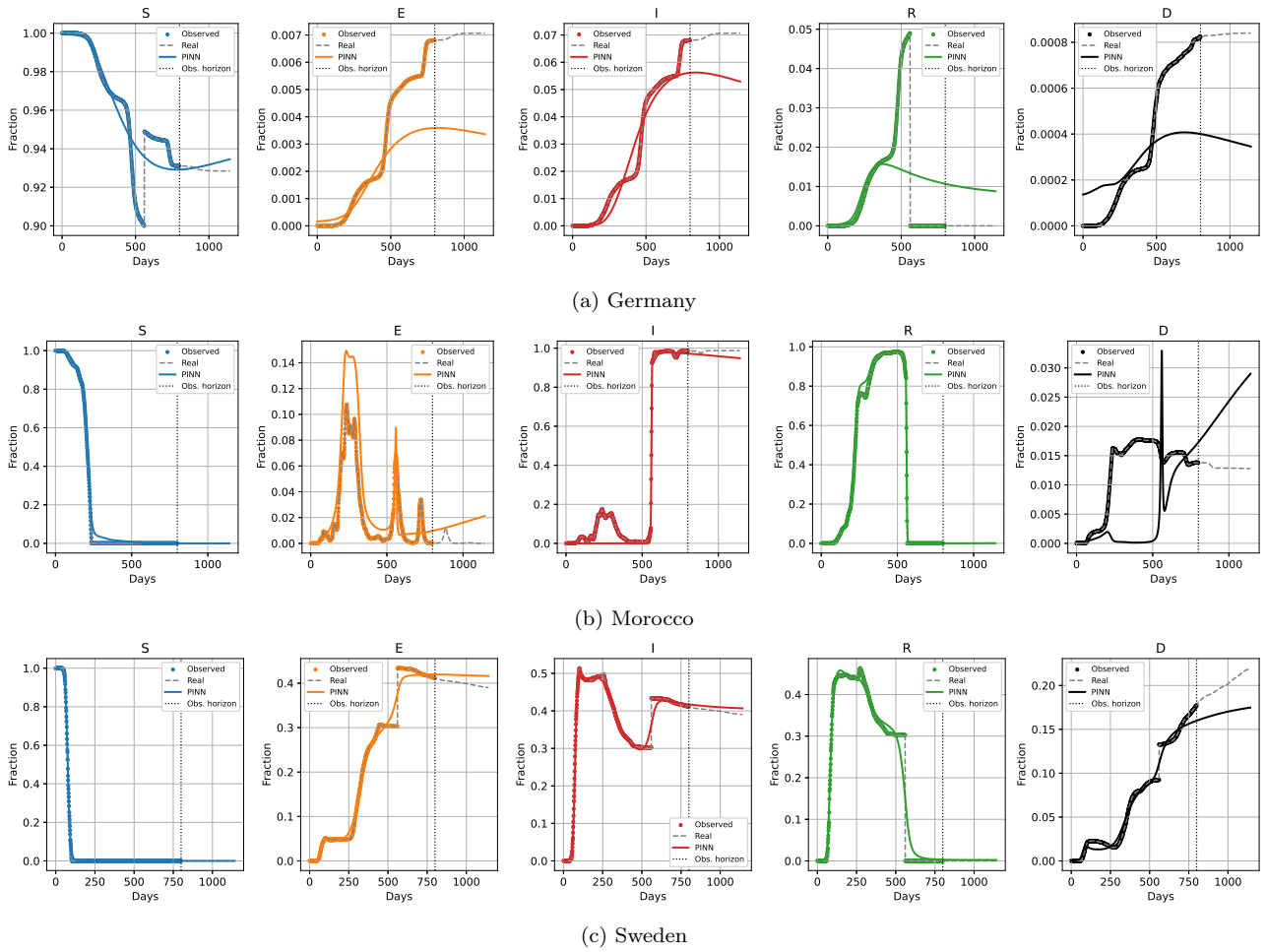


Figure 11: Comparison between the observed epidemiological trajectories and the PINN reconstructions for the Germany, Morocco, and Sweden datasets.

Figure 12 presents the reconstruction and forecasting results obtained using the proposed K-PINN framework. The predicted trajectories show excellent alignment with the observed epidemic data across all compartments and datasets. Compared with both K-EDMD and PINN, the proposed approach provides more accurate reconstructions and improved long-term forecasting performance. These results demonstrate that the combining Koopman latent linearization and physics-informed learning effectively captures the underlying epidemic dynamics while preserving the model’s mechanistic structure.

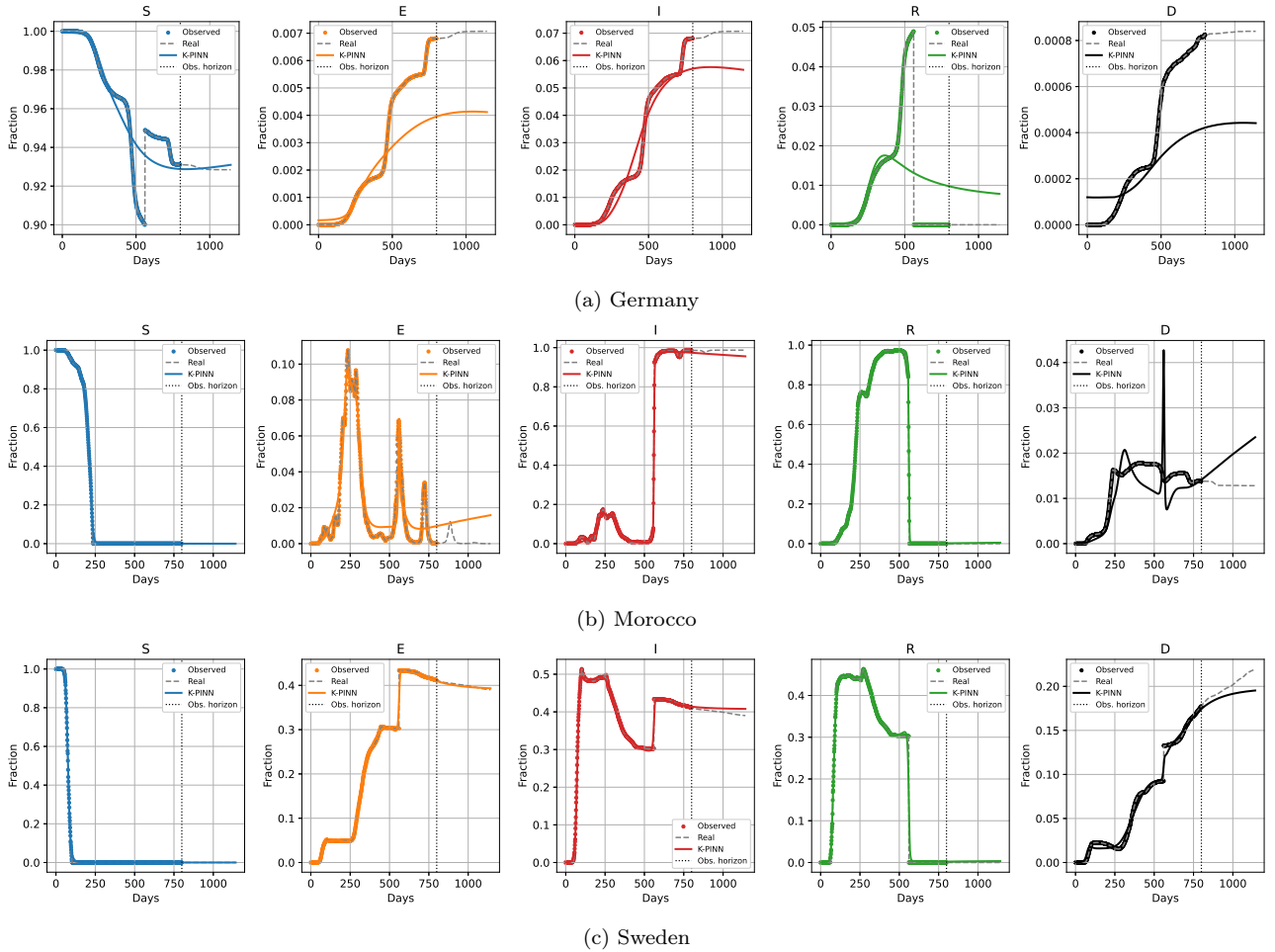


Figure 12: Comparison between the observed epidemiological trajectories and the proposed K-PINN reconstructions for the Germany, Morocco, and Sweden datasets.

The quantitative results reported in Tables 5 and 6 further confirm the advantages of the proposed K-PINN framework. For all three datasets, the K-PINN framework achieves the lowest global reconstruction errors, outperforming both the K-EDMD and the classical PINN approach. This improvement is particularly noticeable for the Morocco and Sweden datasets, where the nonlinear epidemic dynamics are more challenging to capture. These results demonstrate that integrating Koopman-based latent representations with physics-informed learning results in more precise state reconstruction, enhanced forecasting capabilities, and reliable epidemiological parameter estimation.

Table 5: Comparison of reconstruction and forecasting errors obtained using K-EDMD, PINN, and the proposed K-PINN framework for the Germany, Morocco, and Sweden datasets.

Dataset	Method	E_{L^2}	E_{L^∞}	RMSE(S)	RMSE(E)	RMSE(I)	RMSE(R)	RMSE(D)
Germany	K-EDMD	0.89466	0.99964	0.89319	0.00528	0.04902	0.01421	0.00060
	PINN	0.01830	0.03597	0.01067	0.00225	0.00927	0.01140	0.00031
	K-PINN	0.01731	0.03584	0.01055	0.00194	0.00800	0.01098	0.00029
Morocco	K-EDMD	0.70428	0.88771	0.19012	0.02544	0.44309	0.51271	0.00387
	PINN	0.06419	0.17455	0.01948	0.02161	0.05218	0.02084	0.01083
	K-PINN	0.02278	0.11854	0.00727	0.00986	0.01589	0.00949	0.00513
Sweden	K-EDMD	0.19763	0.53042	0.11379	0.08303	0.07483	0.11130	0.03505
	PINN	0.03824	0.15156	0.00592	0.01684	0.01303	0.02507	0.01858
	K-PINN	0.01812	0.14256	0.00591	0.00528	0.00865	0.01111	0.00822

Table 6: Estimated epidemiological parameters obtained using PINN and the proposed K-PINN framework for the real-world datasets.

Dataset	Method	β	σ	γ	μ
Germany	PINN	0.2207	0.2595	0.0576	0.0039
	K-PINN	0.2254	0.2596	0.0589	0.0041
Morocco	PINN	0.3096	0.1454	0.0547	0.0019
	K-PINN	0.2593	0.1744	0.0547	0.0019
Sweden	PINN	0.2583	0.1484	0.0620	0.0075
	K-PINN	0.2257	0.1200	0.0544	0.0078

For clarity of presentation, Figures 17–19 display the collective reconstruction and forecasting results obtained with the proposed K-PINN framework. Detailed compartment-wise comparisons for all epidemic variables and all competing approaches are provided in Appendix B.

6. Conclusion and Future Work

In this paper, we proposed a K-PINN framework for nonlinear epidemic systems. This methodology combines the strengths of Koopman analysis, an operator-theoretic approach, with the capabilities of PINNs, which enforce equations to provide a robust, interpretable, and data-efficient framework for epidemic reconstruction, parameter inference, and forecasting. The approach elevates the epidemic dynamics to a latent observable space, where evolution is approximately governed by a finite-dimensional Koopman operator. Through a physics-informed loss function, the framework simultaneously enforces the governing epidemic equations and integrates data-driven learning with mechanistic epidemiological modeling.

We developed and analyzed the framework using an SEIRSD epidemic model that incorporates disease-induced mortality and temporary immunity. We presented theoretical considerations on the model’s well-posedness and invariance properties, providing a mathematically consistent foundation for the subsequent Koopman-based learning framework. Next, we introduced a unified K-PINN architecture consisting of a neural lifting operator, latent Koopman dynamics, a reconstruction network, and physics-informed regularization terms that enforce the governing epidemic equations and epidemiological constraints.

A series of numerical experiments involving both synthetic and real-world epidemiological datasets demonstrated the effectiveness of the proposed methodology. The synthetic experiments confirmed the framework’s ability to accurately reconstruct epidemic trajectories and identify unknown epidemiological parameters. Additional validation using COVID-19 datasets from Germany, Morocco, and Sweden further demonstrated the applicability of the proposed approach to real epidemic observations. In all scenarios considered, the K-PINN framework achieved improved reconstruction accuracy, enhanced parameter identifiability, and superior forecasting performance compared to the standalone K-EDMD approach and the classical PINN framework. These results suggest that combining Koopman latent linearization with physics-informed learning is a powerful way to capture complex nonlinear epidemic dynamics while preserving the underlying epidemiological structure.

The proposed framework offers several important advantages. First, the latent Koopman representation provides a structured and interpretable description of nonlinear dynamics, thereby improving long-term forecasting capabilities. Second, physics-informed constraints ensure consistency with the governing epidemic model and reduce the risk of physically unrealistic predictions. Third, the unified inverse-learning formulation enables simultaneous state reconstruction and parameter estimation from limited observations. This makes the approach particularly attractive for real-world epidemic monitoring and decision support.

The present work yields several promising directions for future research. From an epidemiological perspective, for example, the proposed K–PINN framework can be extended to more sophisticated disease transmission models. These models can include age-structured, multi-group, multi-strain, and metapopulation systems. This would enable the analysis of heterogeneous populations, spatial mobility, and complex contact networks. Another important avenue is incorporating stochastic effects to account for environmental variability, demographic randomness, and measurement uncertainties. Additionally, the methodology can be generalized to fractional-order epidemic models, enabling memory and hereditary effects to be explicitly represented within the learning framework.

The proposed K–PINN methodology has applications beyond epidemiology and can be used with a broad range of nonlinear dynamical systems. Potential applications include ecological population dynamics, predator-prey systems, tumor growth models, biological networks, reaction-diffusion processes in chemistry, fluid dynamics, climate and environmental modeling, and engineering systems governed by ordinary, fractional, stochastic, or partial differential equations. These extensions demonstrate the versatility of the proposed framework as a general machine-learning methodology for complex nonlinear systems, combining latent linear representations with physics-informed learning.

Appendix A Additional and Supplementary Synthetic Data Experiments

This appendix provides a detailed, compartment-by-compartment comparison of the reconstruction and forecasting results obtained using K–EDMD, PINN, and the proposed K–PINN framework. Figure 13 compares the reference synthetic trajectory with the corresponding predictions generated by the three approaches for each epidemiological compartment over the training and forecasting intervals. These supplementary results provide insight into the strengths and limitations of each methodology, further highlighting the improved accuracy and long-term predictive capability of the K–PINN framework.

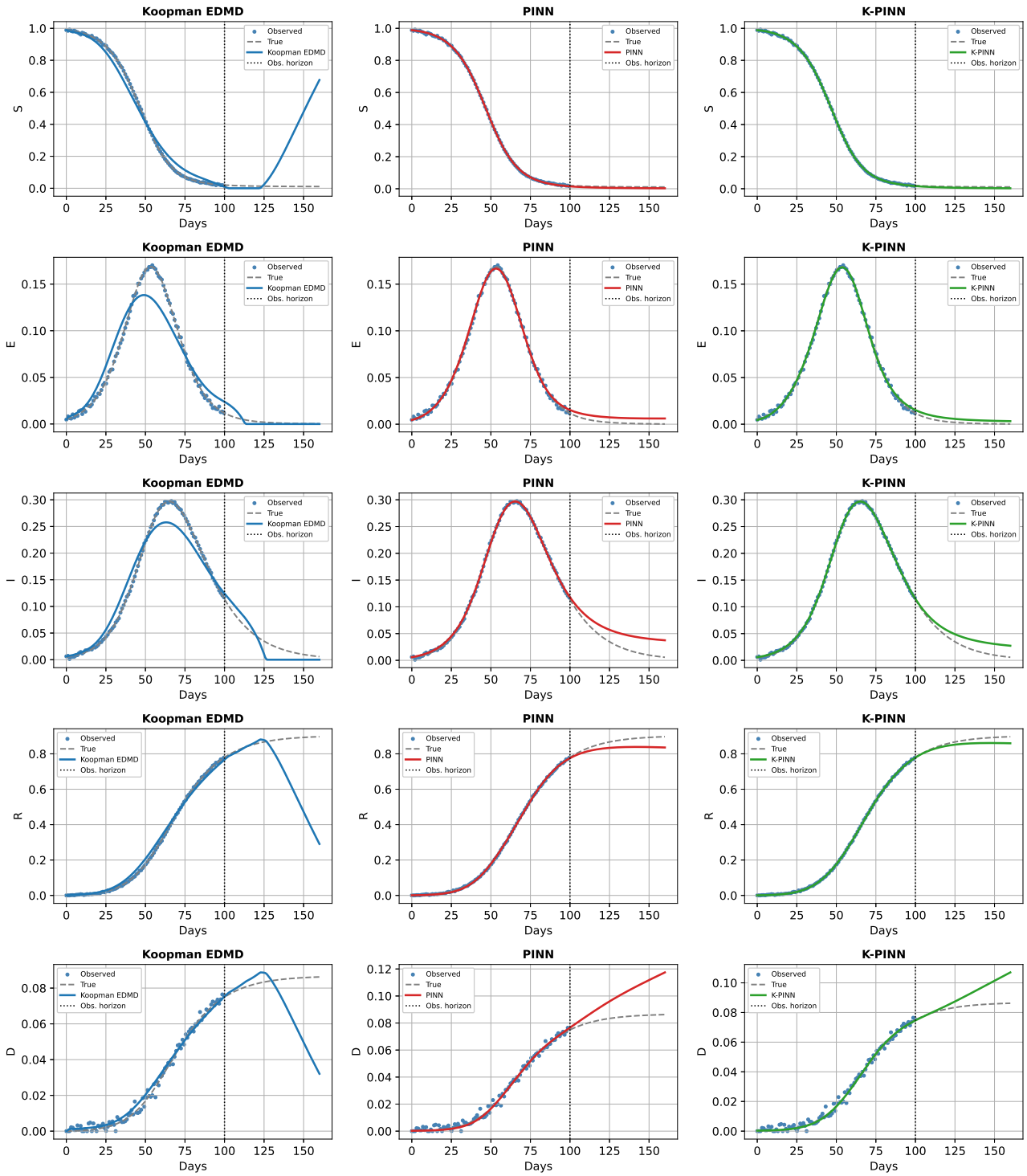


Figure 13: Compartment-wise comparison of K-EDMD, PINN, and K-PINN predictions against the synthetic trajectories.

Figure 14 reports the diagnostic quantities associated with the Koopman approximation, including the latent-space prediction errors and the reconstruction diagnostics.

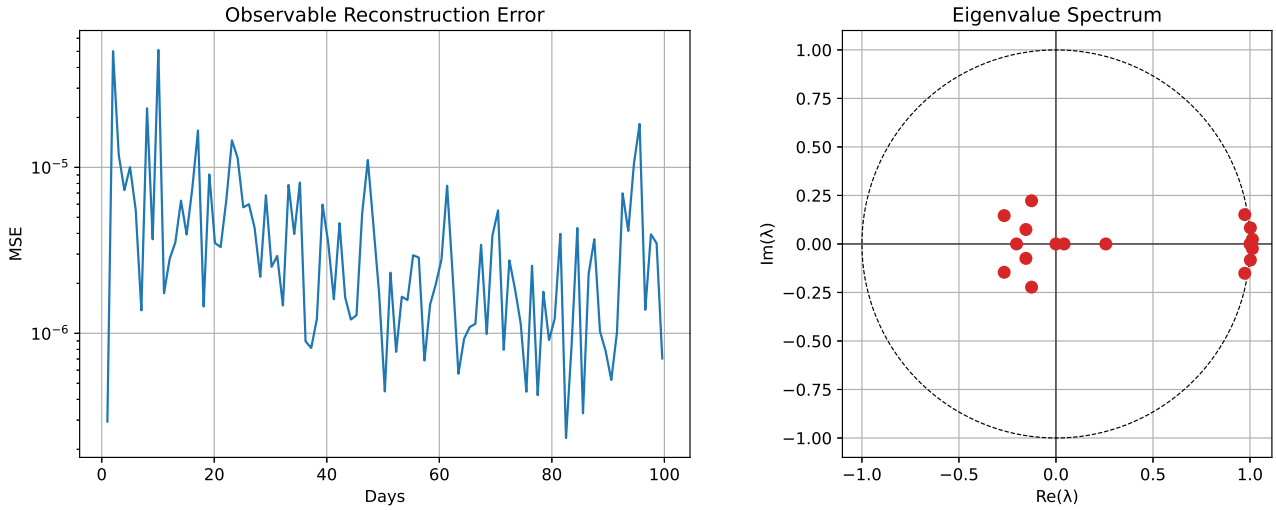


Figure 14: K-EDMD diagnostic analysis and latent-space prediction errors for the synthetic dataset.

Figure 15 shows the convergence history of the PINN optimization process, and Figure 16 shows the evolution of the estimated epidemiological parameters during training.

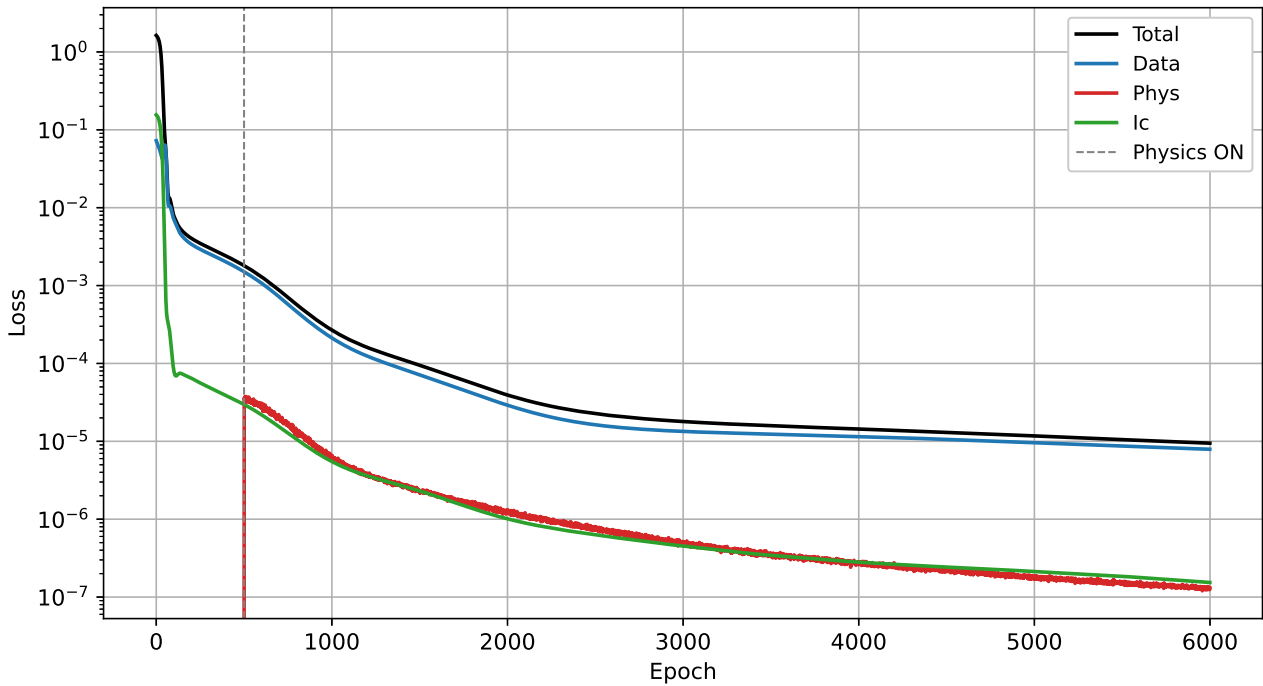


Figure 15: Evolution of the composite loss function during PINN training for the synthetic dataset.

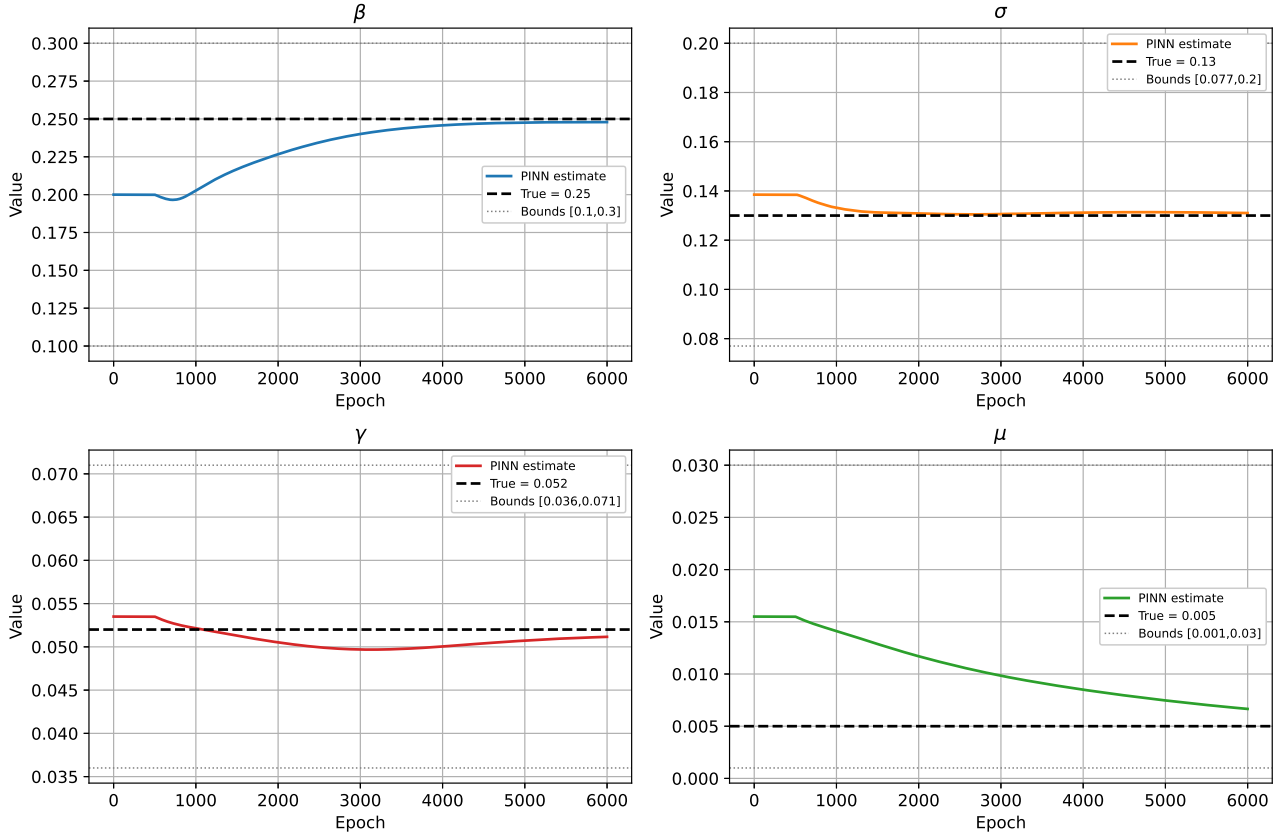


Figure 16: Convergence of epidemiological parameter estimates obtained by PINN for the synthetic dataset.

Appendix B Additional and Supplementary Real-World Data Experiments

This appendix provides a detailed, compartment-by-compartment comparison of the reconstruction and forecasting results obtained using K-EDMD, PINN, and the proposed K-PINN framework for the Germany, Morocco, and Sweden datasets. Figures 17–19 compare the observed epidemic trajectories with the corresponding predictions generated by the three approaches over the training and forecasting intervals for each epidemiological compartment. These supplementary results provide insight into the strengths and limitations of each methodology, further highlighting the improved accuracy, robustness, and long-term predictive capability of the proposed K-PINN framework.

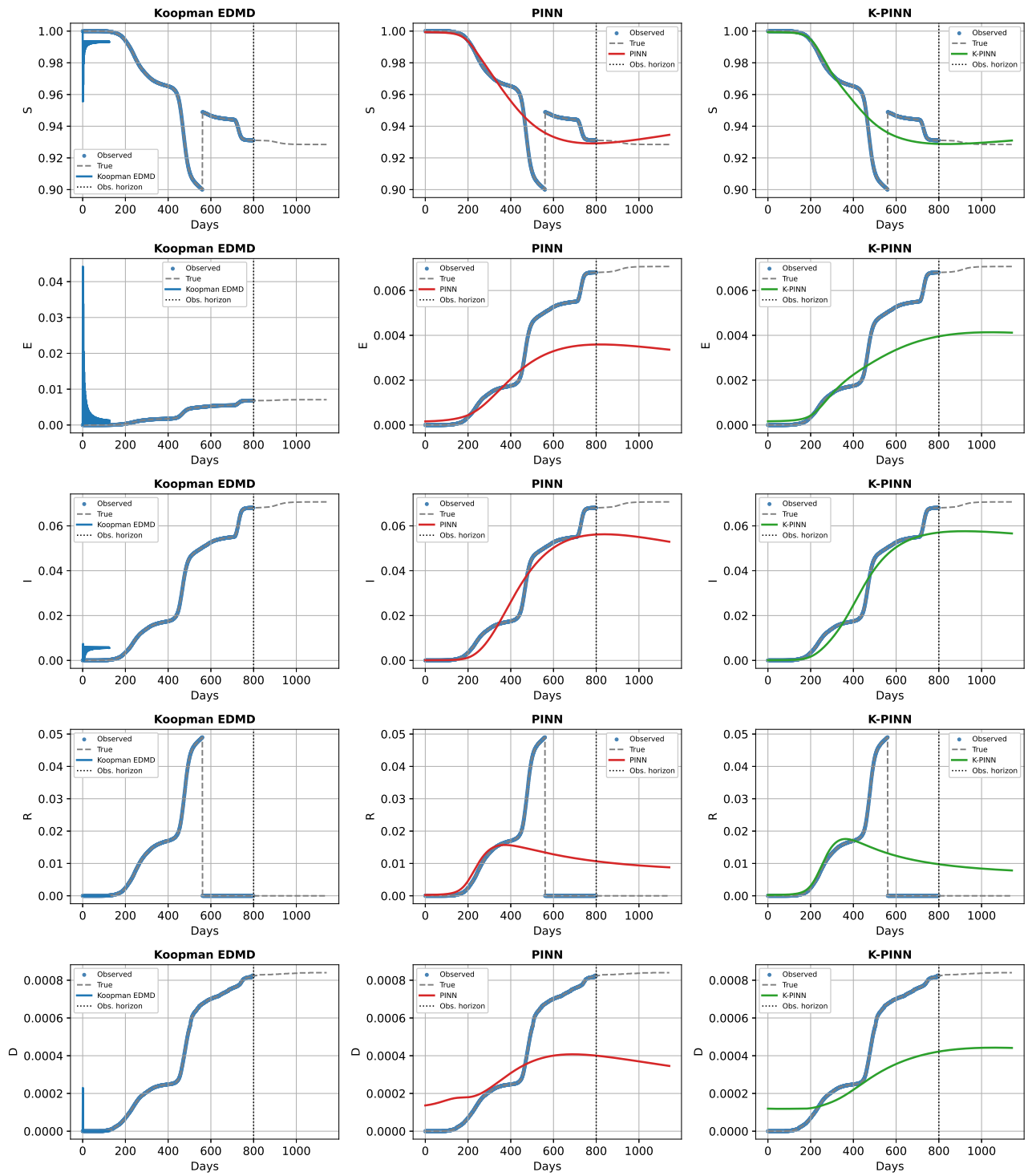


Figure 17: Compartment-wise comparison of K-EDMD, PINN, and K-PINN predictions for the COVID-19 Germany dataset.

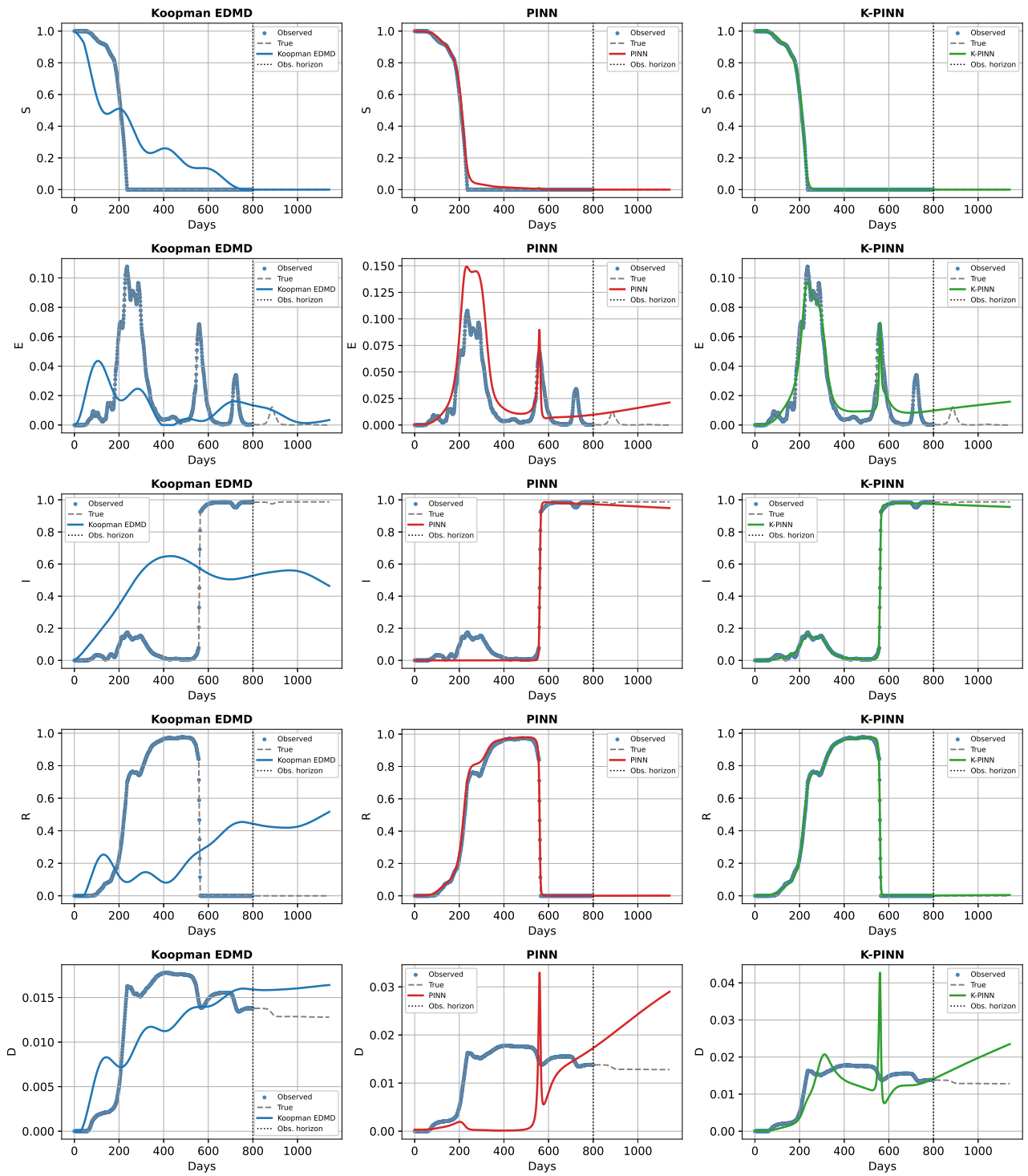


Figure 18: Compartment-wise comparison of K-EDMD, PINN, and K-PINN predictions for the COVID-19 Morocco dataset.

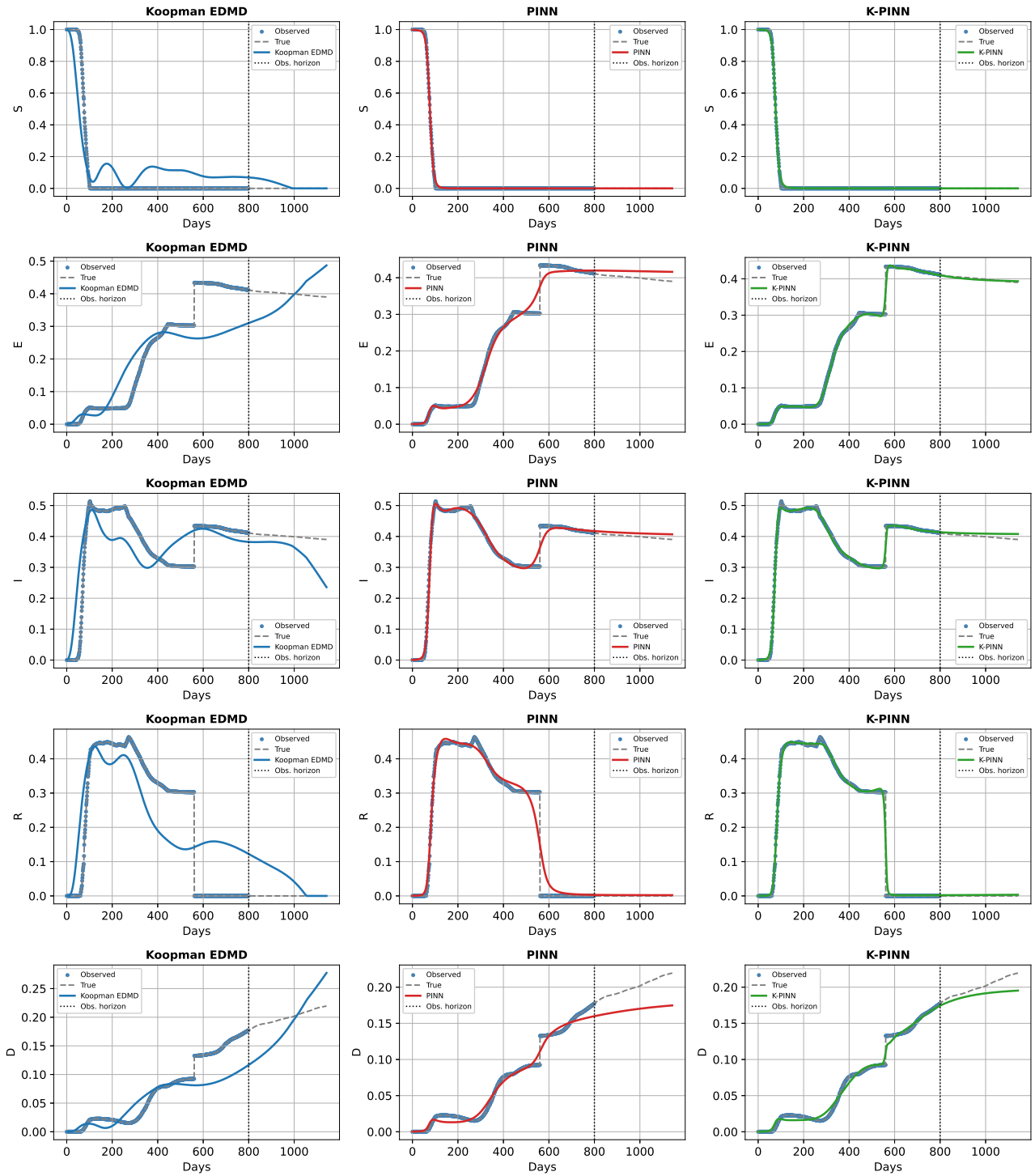


Figure 19: Compartment-wise comparison of K-EDMD, PINN, and K-PINN predictions for the COVID-19 Sweden dataset.

Figure 20 depicts the evolution of the composite K-PINN loss function during training. A consistent decrease of the loss is observed for all datasets, indicating stable optimization of the data reconstruction, Koopman consistency, and physics-informed residual terms.

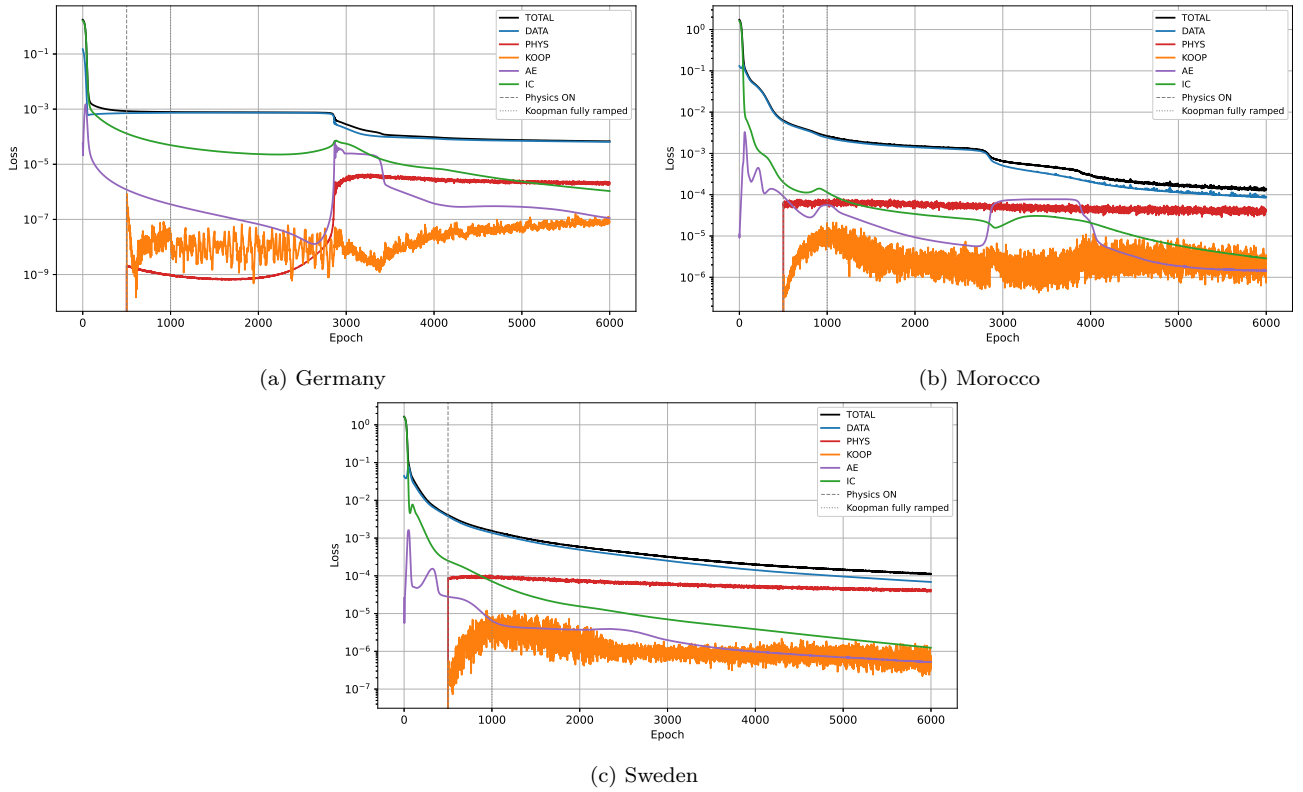
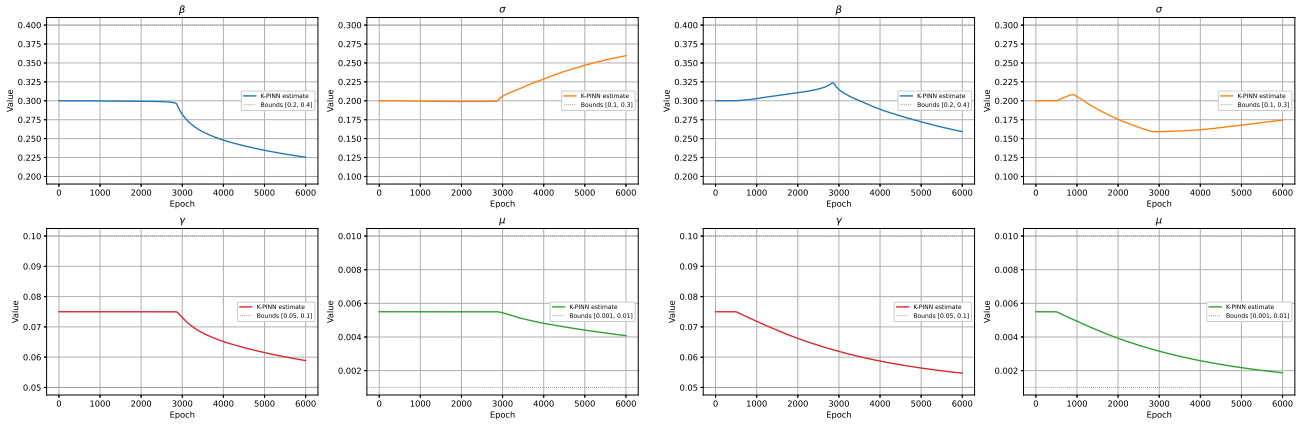


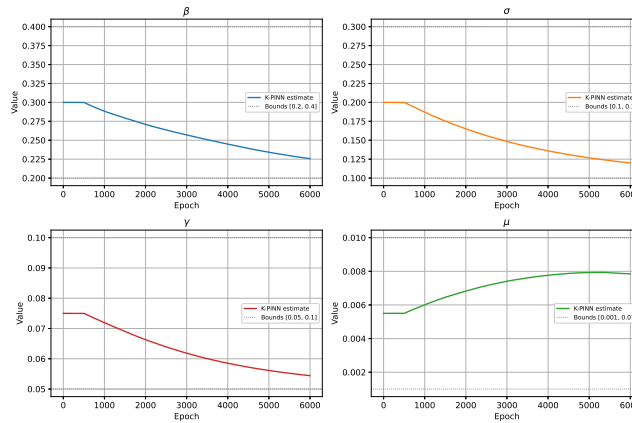
Figure 20: Evolution of the composite K-PINN loss function during training for the Germany, Morocco, and Sweden datasets.

Figure 21 displays the evolution of the epidemiological parameter estimates learned by the proposed K-PINN framework. The estimated parameters smoothly converge toward stable values, which demonstrates the robustness of the inverse-learning strategy. The resulting parameter estimates are summarized in Table 6 and are consistent with biologically plausible epidemic regimes, supporting the improved predictive performance of the proposed methodology.



(a) Germany

(b) Morocco



(c) Sweden

Figure 21: Convergence histories of the epidemiological parameter estimates obtained by the proposed K–PINN framework for the Germany, Morocco, and Sweden datasets.

Declarations

Conflict of Interest

The authors declared that they have no conflict of interest.

Author Contributions

A. Zinihi: Conceptualization, Software, Methodology, Validation, Formal Analysis, Investigation, Writing-Original Draft, Writing-Review and Editing, Visualization.

M. Ehrhardt: Conceptualization, Supervision, Methodology, Formal Analysis, Investigation, Writing-Review and Editing.

M. R. Sidi Ammi: Conceptualization, Supervision, Methodology, Formal Analysis.

Computational Setup

All numerical simulations for the K–EDMD, PINN, and K–PINN frameworks were conducted on a workstation running Microsoft Windows 11 Home. The system features an Intel Core Ultra 9 275HX CPU (24 cores, 24 threads) with 32 GB of RAM.

Computations were accelerated using a hybrid GPU setup, including an integrated Intel GPU and an NVIDIA GeForce RTX 5070 Ti Laptop GPU (12 GB VRAM).

Software Environment

The simulations were implemented in *Python 3.13* using PyTorch.

Data Availability

This study uses both synthetic and real epidemic datasets. Real datasets were obtained from the following publicly available source: [19–25].

References

- [1] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019.
- [2] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proc. Nat. Acad. Sci.*, 17(5):315–318, 1931.
- [3] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlin. Dynam.*, 41(1-3):309–325, 2005.
- [4] M. Korda and I. Mezić. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149–160, 2018.
- [5] J. L. Proctor, S. L. Brunton, and J. N. Kutz. Dynamic mode decomposition with control. *SIAM J. Appl. Dynam. Syst.*, 15(1):142–161, 2016.
- [6] B. Lusch, J. N. Kutz, and S. L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Commun.*, 9(1), 2018.
- [7] N. Takeishi, Y. Kawahara, and T. Yairi. Learning Koopman invariant subspaces for dynamic mode decomposition. *Adv. Neural Inform. Process. Syst.*, 30, 2017.
- [8] S. E. Otto and C. W. Rowley. Linearly recurrent autoencoder networks for learning dynamics. *SIAM J. Appl. Dynam. Syst.*, 18(1):558–593, 2019.
- [9] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton. Data-driven discovery of coordinates and governing equations. *Proc. Nat. Acad. Sci.*, 116(45):22445–22451, 2019.
- [10] A. Zinihi, M. R. Sidi Ammi, and D. F. M. Torres. Fractional differential equations of a reaction-diffusion SIR model involving the Caputo-fractional time-derivative and a nonlinear diffusion operator. *Evol. Eqs. Contr. Theo.*, 14(5):944–967, 2025.
- [11] A. Zinihi, M. Ehrhardt, and M. R. Sidi Ammi. Spatiotemporal SEIQR epidemic modeling with optimal control for vaccination, treatment, and social measures, 2025.
- [12] A. Zinihi and M. Ehrhardt. Constraint-aware physics-informed neural networks for SEIR reaction-diffusion epidemic models with vital dynamics, 2026. arXiv 2605.19886.
- [13] A. Zinihi, M. R. Sidi Ammi, and A. Bachir. Multi-city modeling of epidemics using a topology-based SIR model: Neural network-enhanced SAIRD model. *J. Comput. Sci.*, 92:102721, 2025.
- [14] A. Zinihi. Identifying memory effects in epidemics via a fractional SEIRD model and physics-informed neural networks, 2025. arXiv 2509.22760.
- [15] A. Zinihi, M. Ehrhardt, and M. R. Sidi Ammi. A Koopman operator framework for nonlinear epidemic dynamics: Application to an SIRSD model. *Nonlin. Sci.*, 6:100095, 2026.

- [16] M. Ehrhardt and R. E. Mickens. A nonstandard finite difference scheme for convection-diffusion equations having constant coefficients. *Appl. Math. Comput.*, 219(12):6591–6604, 2013.
- [17] A. Zinihi, M. Ehrhardt, and M. R. Sidi Ammi. A nonstandard finite difference scheme for an SEIQR epidemiological PDE model. *Appl. Math. Comput.*, 520:129953, 2026.
- [18] M. H. Maamar, M. Ehrhardt, and L. Tabharit. A nonstandard finite difference scheme for a time-fractional model of Zika virus transmission. *Math. Biosci. Engrg.*, 21(1):924–962, 2024.
- [19] World Health Organization. Mpox: Multi-country external situation report no.44; https://www.who.int/docs/default-source/coronaviruse/situation-reports/20241223_mpxo-external-sitrep_-44.pdf, Accessed on: June 4, 2026.
- [20] The Lancet Global Health. Modelling vaccination approaches for mpox containment; [https://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(24\)00384-X/fulltext](https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(24)00384-X/fulltext), Accessed on: June 4, 2026.
- [21] UK Health Security Agency. Epidemiology of COVID-19 in England: January 2020 to December 2024; <https://www.gov.uk/government/publications/epidemiology-of-covid-19-in-england/epidemiology-of-covid-19-in-england-january-2020-to-december-2024>, Accessed on: September 20, 2025.
- [22] World Health Organization. COVID-19 epidemiological update, 24 December 2024; <https://www.who.int/publications/m/item/covid-19-epidemiological-update---24-december-2024>, Accessed on: September 20, 2025.
- [23] J. P. Gehrcke. COVID-19 case numbers for Germany; https://github.com/jgehrcke/covid-19-germany-gae?utm_source, Accessed on: June 6, 2026.
- [24] Our World in Data. Estimated cumulative excess deaths during COVID-19, Jan 1, 2020 to Aug 24, 2025; <https://ourworldindata.org/explorers/covid?Metric=Excess+mortality>, Accessed on: June 6, 2026.
- [25] Worldometers.info. Sweden COVID-Coronavirus statistics; <https://www.worldometers.info/coronavirus/country/sweden/>, Accessed on: June 6, 2026.