Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Physics-Informed Neural Networks for Parameter Estimation in Epidemic Models with Noisy and Limited Data

Isaac Adeyemi

isaacadeyemi245@gmail.com

Ladoke Akintola University of Technology

Damilare Bamidele

Ladoke Akintola University of Technology

Toluwanimi Akinmade

mountain top

Zainab Adisa

Ladoke Akintola University of Technology

Omolabake Ademola

Ladoke Akintola University of Technology

Research Article

Keywords: Mathematical modelling, Physics-informed neural networks, Epidemic dynamics, Parameter estimation, Least squares calibration, Scientific machine learning

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11717815/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Physics-Informed Neural Networks for Parameter Estimation in Epidemic Models with Noisy and Limited Data

Adeyemi I. T.[1], Bamidele, D. E.[2], Akinmade T. D.[3], Adisa Z.O[4], Ademola O.R.[5]

[1,2] Department of Pure and Applied Mathematics, Ladoke Akintola University of Technology, P.M.B. 4000, Ogbomoso, Oyo State, Nigeria
[3]Department of Accounting, Mountain Top University, Km 12, Lagos-Ibadan Expressway, Prayer City, Ogun State, Nigeria
[4,5]Department of Science Laboratory Technology, Ladoke Akintola University of Technology, P.M.B. 4000, Ogbomoso, Oyo State, Nigeria

Corresponding Author Email: isaacadeyemi245@gmail.com

Abstract

Estimating epidemiological parameters from noisy and limited surveillance data is a critical challenge in mathematical epidemiology and outbreak response. This study presents a rigorous comparative benchmark evaluating Physics-Informed Neural Networks (PINNs), which embed an SIR compartmental model into the neural learning objective via automatic differentiation, against classical Non-linear Least Squares (NLS) calibration with numerical Runge-Kutta integration. Both approaches are tested on synthetic epidemic data across observation counts of $N_t = 10$ and $N_t = 30$ and Gaussian noise levels from 1% to 20% over multiple stochastic realizations. Results show that for fully observed, correctly specified compartmental models, classical Least Squares recovers parameters more accurately and reliably in preliminary runs, achieving median transmission rate errors below 5% (0.089%–4.160%) and 100% compliance in capturing the basic reproduction number (R_0) within a 10% clinical tolerance margin. In contrast, preliminary PINNs yield median transmission errors of 7.76% to 34.21% alongside a systematic negative parameter bias (up to -30.43%) caused by optimization stiffness and gradient scale imbalances between data and physical residual losses. However, extended training experiments (10,000–20,000 iterations) demonstrate that PINN errors reduce asymptotically below 2% to 5%, confirming that early limitations stem from optimization stalling rather than representational incapacity. Classical calibration remains the optimal baseline for fully observed low-dimensional systems, while PINNs offer significant scientific potential for partially observed states and non-parametric functional dynamics.

Keywords: Mathematical modelling; Physics-informed neural networks; Epidemic dynamics; Parameter estimation; Least squares calibration; Scientific machine learning.

1. Introduction

The quantitative surveillance and predictive forecasting of infectious disease transmission rely fundamentally on compartmental mathematical models governed by systems of non-linear ordinary differential equations (ODEs) [12, 25]. From the early twentieth-century foundations of the Susceptible-Infectious-Recovered (SIR) paradigm to modern multi-group and spatial extensions, mathematical epidemiology provides vital insight into epidemic trajectories, herd immunity thresholds, and intervention design [6, 7]. A central mathematical task in real-time outbreak response is the inverse problem, inferring hidden biophysical rates, such as the effective transmission rate (β) and the clinical recovery/removal rate (γ), from sparse, contaminated, and delayed surveillance data [2, 3, 29].

Accurate parameter estimation is paramount because these constants dictate the basic reproduction number:

$$R_0 = \frac{\beta}{\gamma}$$

The threshold parameter R_0 determines whether a pathogen will cause an explosive epidemic ($R_0 > 1$) or burn out ($R_0 < 1$), dictating the scale of non-pharmaceutical interventions (NPIs) required for containment [9, 21]. However, epidemiological surveillance systems in resource-constrained environments frequently suffer from severe data limitations, including infrequent sampling, reporting delays, and substantial observational noise [2, 9, 29].

Traditionally, the parameter estimation problem is formulated as a non-linear inverse problem solved via Classical Non-linear Least Squares (NLS) calibration, Levenberg-Marquardt optimization, or Markov Chain Monte Carlo (MCMC) Bayesian inference [3, 16]. These classical methods integrate the forward ODE system using deterministic numerical solvers (such as explicit Runge-Kutta 4th/5th order or adaptive Radau schemes) and optimize the parameter vector to minimize the residual sum of squares between simulated trajectories and empirical observations [17, 18].

In recent years, the confluence of deep learning and mechanistic modeling has catalyzed the emergence of Physics-Informed Neural Networks (PINNs) as a dominant paradigm in scientific machine learning (SciML) [20, 22, 28]. Introduced to solve forward and inverse problems in differential equations, PINNs approximate state trajectories using deep neural networks while constraining the network's output to satisfy governing physical or biological equations via automatic differentiation [13, 20]. By penalizing residuals of the differential equations within the composite loss function, PINNs act as mesh-free, continuous solvers that seamlessly incorporate observational data without requiring traditional temporal discretization grids [11, 22].

The rapid adoption of PINNs across fluid dynamics, structural mechanics, and materials science has sparked intense interest in their epidemiological deployment [2, 9, 20, 29]. Several recent studies have applied PINNs to estimate transmission rates for COVID-19, influenza, and vector-borne diseases [5, 9, 21]. However, existing literature frequently emphasizes the theoretical strengths of PINNs while glossing over the practical trade-offs, optimization stiffness, and statistical reliability relative to well-established numerical baselines [6, 14]. When benchmarked against classical calibration under realistic surveillance constraints—specifically limited observation counts ($N_t = 10{,}30$) and high observational noise (up to 20%)—the comparative performance of PINNs remains incompletely characterized [23, 28].

To address this critical gap, this paper presents a systematic, empirical comparative analysis between Physics-Informed Neural Networks and classical Least Squares calibration. Using synthetic data generated from an SIR model with known ground-truth parameters ($\beta^* = 0.50, \gamma^* = 0.20, R_0^* = 2.50$), we evaluate both frameworks across varying sample sizes, four noise intensities, and multiple stochastic realizations. Our primary contributions are:

1. Direct Benchmark Under Severe Surveillance Scarcity: We evaluate estimation accuracy for both transmission (β) and recovery (γ) parameters across noise levels from 1% to 20% and sparse sample sizes($N_t = 10$and $N_t = 30$).
2. Identification of Parameter Estimation Discrepancies: We demonstrate that in standard preliminary training regimes, classical NLS provides superior accuracy and robustness (median β error $< 5\%$) compared to PINNs (median β error $8\% - 34\%$), identifying a persistent negative bias in PINN transmissionrate estimation.
3. Epidemiological Risk Evaluation: We translate parameter errors into epidemiological metrics, evaluating the probability of accurately capturing the basic reproduction number (R_0) within a clinically acceptable 10% margin.
4. Investigation of Optimization Dynamics: Through extended training up to 20,000 gradient descent iterations, we demonstrate that the preliminary lag of PINNs is primarily governed by optimization stiffness and loss balancing rather than intrinsic model inadequacy.
5. Methodological Guidelines for Scientific Machine Learning in Epidemiology: We synthesize the empirical results into practical guidance, defining when traditional numerical calibration should be retained and highlighting scenarios (e.g., latent states, functional parameter estimation) where PINNs provide distinct advantages [27, 30].

2. Mathematical Model and Problem Formulation

2.1 The SIR Compartmental Model

The classic Kermack-McKendrick SIR model partitions a homogeneous, constant population of size N into three mutually exclusive epidemiological compartments: Susceptible (S), Infectious (I), and Recovered/Removed (R) [6, 7]. Assuming negligible vital dynamics (births and natural deaths) over the epidemic time horizon $[0, T]$, the continuous temporal dynamics are governed by the non-linear ODE system:

$$\frac{dS}{dt} = -\frac{\beta S(t) I(t)}{N} \qquad (1)$$

$$\frac{dI}{dt} = \frac{\beta S(t) I(t)}{N} - \gamma I(t) \qquad (2)$$

$$\frac{dR}{dt} = \gamma I(t) \qquad (3)$$

subject to initial conditions $S(0) = S_0$, $I(0) = I_0$, and $R(0) = R_0^{\text{init}}$, satisfying the constant population conservation law:

$$S(t) + I(t) + R(t) = N, \quad \forall t \in [0, T] \qquad (4)$$

Here, $\beta > 0$ denotes the biophysical transmission rate (the average number of contacts sufficient for transmission per unit time), and $\gamma > 0$ represents the recovery rate, corresponding to an average infectious duration of $\tau = 1/\gamma$.

To improve numerical conditioning for both gradient-based neural network optimization and non-linear numerical solvers, we express the dynamics in population fractions: $s(t) = S(t)/N$, $i(t) = I(t)/N$, and $r(t) = R(t)/N$. Substituting these definitions into Eqs. (1) – (3) yields the normalized non-dimensional system:

$$\frac{ds}{dt} = -\beta s(t) i(t) \qquad (5)$$

$$\frac{di}{dt} = \beta s(t) i(t) - \gamma i(t) \qquad (6)$$

$$\frac{dr}{dt} = \gamma i(t) \qquad (7)$$

with $s(t) + i(t) + r(t) = 1$. The fundamental epidemiological threshold is the basic reproduction number:

$$R_0 = \frac{\beta}{\gamma} \qquad (8)$$

2.2 The Inverse Parameter Estimation Problem

In surveillance scenarios, continuous state trajectories are observed only at discrete time instants $\{t_k\}_{k=1}^{N_t} \subset [0, T]$. Furthermore, reporting delays, test sensitivity variations, and stochastic fluctuations introduce observational noise:

$$\tilde{s}_k = s(t_k) + \varepsilon_{s,k}, \quad \tilde{\imath}_k = i(t_k) + \varepsilon_{i,k}, \quad \tilde{r}_k = r(t_k) + \varepsilon_{r,k} \qquad (9)$$

where $\varepsilon_{\cdot,k} \sim \mathcal{N}(0, \sigma_k^2)$ represents additive Gaussian noise with standard deviation proportional to a target noise ratio η:

$$\sigma_k = \eta \cdot y_{\text{true}}(t_k) \qquad (10)$$

Given noisy observations $\tilde{\mathrm{Y}} = \{(\tilde{s}_k, \tilde{\imath}_k, \tilde{r}_k)\}_{k=1}^{N_t}$, the inverse problem consists of estimating the biophysical parameters $\hat{\theta} = (\hat{\beta}, \hat{\gamma})$.

3. Numerical Methodology and Computational Framework

3.1 Classical Non-linear Least Squares (NLS) Calibration

The classical benchmark formulates parameter estimation as a constrained non-linear optimization problem [17, 18]:

$$\min_{\theta \in \Theta} \mathcal{J}_{\text{NLS}}(\theta) = \frac{1}{N_t} \sum_{k=1}^{N_t} \| \mathrm{y}(t_k; \theta) - \tilde{\mathrm{y}}_k \|_2^2 \qquad (11)$$

where $\mathrm{y}(t;\theta) = [s(t;\theta), i(t;\theta), r(t;\theta)]^T$ is obtained by numerically integrating Eqs. (5) – (7) using an explicit Runge-Kutta method (RK45) with adaptive step size. The parameter domain is constrained to $\Theta = [10^{-4}, 5.0]^2$. The objective is minimized using the Levenberg-Marquardt and BFGS algorithms with numerical Jacobian evaluations.

3.2 Physics-Informed Neural Network (PINN) Formulation

The Physics-Informed Neural Network represents state trajectories as outputs of a fully connected deep feedforward neural network parameterized by weights W and biases b [12, 16]:

$$\hat{\mathrm{y}}(t;\mathrm{W},\mathrm{b}) = [\hat{s}(t), \hat{\imath}(t), \hat{r}(t)]^T = \mathcal{NN}(t;\mathrm{W},\mathrm{b}) \qquad (12)$$

The architecture comprises 1 input neuron ($t \in \mathbb{R}$), 4 hidden layers with 64 neurons each utilizing hyperbolic tangent (tanh) activations, and 3 linear output units. Parameters $\theta = (\beta, \gamma)$ are treated as trainable variables alongside W, b.

Governing ODE residuals are evaluated using reverse-mode automatic differentiation [1, 10]:

$$e_s(t) = \frac{d\hat{s}}{dt} + \beta\hat{s}(t)\hat{\imath}(t) \qquad (13)$$

$$e_i(t) = \frac{d\hat{\imath}}{dt} - \beta\hat{s}(t)\hat{\imath}(t) + \gamma\hat{\imath}(t) \qquad (14)$$

$$e_r(t) = \frac{d\hat{r}}{dt} - \gamma\hat{\imath}(t) \qquad (15)$$

The composite loss function balances data fidelity, differential equation consistency, and initial conditions:

$$\mathcal{L}_{\mathrm{PINN}}(\mathrm{W},\mathrm{b},\theta) = \mathcal{L}_{\mathrm{data}} + \lambda_{\mathrm{phys}}\mathcal{L}_{\mathrm{phys}} + \lambda_{\mathrm{IC}}\mathcal{L}_{\mathrm{IC}} \qquad (16)$$

where the constituent terms are defined as:

$$\mathcal{L}_{\mathrm{data}} = \frac{1}{N_t}\sum_{k=1}^{N_t}[(\hat{s}(t_k) - \tilde{s}_k)^2 + (\hat{\imath}(t_k) - \tilde{\imath}_k)^2 + (\hat{r}(t_k) - \tilde{r}_k)^2] \qquad (17)$$

$$\mathcal{L}_{\mathrm{phys}} = \frac{1}{N_c}\sum_{j=1}^{N_c}\left[\left|e_s\left(t_j^c\right)\right|^2 + \left|e_i\left(t_j^c\right)\right|^2 + \left|e_r\left(t_j^c\right)\right|^2\right] \qquad (18)$$

$$\mathcal{L}_{\mathrm{IC}} = (\hat{s}(0) - s_0)^2 + (\hat{\imath}(0) - i_0)^2 + (\hat{r}(0) - r_0)^2 \qquad (19)$$

Collocation points $\{t_j^c\}_{j=1}^{N_c}$ ($N_c = 100$) are distributed uniformly across $[0, T]$, and weights are set to $\lambda_{\mathrm{phys}} = 1.0$, $\lambda_{\mathrm{IC}} = 10.0$ [2, 23].

3.3 Evaluation Metrics

Accuracy and stability are quantified through relative percentage error, parameter bias across $M = 3$ stochastic replications, and continuous trajectory RMSE:

$$\text{RelErr}(\theta) = \frac{\left|\hat{\theta} - \theta^*\right|}{\theta^*} \times 100\% \qquad (20)$$

$$\text{Bias}(\theta) = \frac{1}{M}\sum_{m=1}^{M} \frac{\hat{\theta}_m - \theta^*}{\theta^*} \times 100\% \qquad (21)$$

$$\text{RMSE}_{\text{curve}} = \sqrt{\frac{1}{N_{\text{eval}}}\sum_{k=1}^{N_{\text{eval}}} \| \text{y}_{\text{true}}(t_k) - \hat{\text{y}}(t_k) \|_2^2} \qquad (22)$$

Ground truth parameters were fixed to $\beta^* = 0.50$, $\gamma^* = 0.20$, yielding $R_0^* = 2.50$, evaluated over $T = 40$ days with $s_0 = 0.99, i_0 = 0.01, r_0 = 0.00$.

4. Numerical Results and Comparative Analysis

4.1 Comprehensive Experimental Benchmark

To present the multidimensional evaluation without horizontal compression across standard publication formats, the empirical benchmark across all 16 experimental configurations (evaluated over 3 stochastic replications each) is partitioned into two thematic tables: Table 1A details transmission rate estimation and basic reproduction number decisions, while Table 1B presents recovery rate calibration and state trajectory reconstruction fidelity.

Table 1A: Transmission Rate (β) Estimation Performance and Basic Reproduction Number (R_0) Decisions across 3 Stochastic Replications.

Method	Points (N_t)	Noise Level	Median β Err (%)	Mean $\pm$ Std β Err (%)	Bias β (%)	RMSE β (%)	R_0 Err (%)	$R_0 \leq 10\%$ Conf. (%)
Least Squares	10	1%	0.1927	0.2101 $\pm$ 0.0375	-0.0932	0.2134	0.2638	100.0%
Least Squares	10	5%	0.9320	1.0462 $\pm$ 0.1796	-0.4417	1.0615	1.3088	100.0%
Least Squares	10	10%	1.8949	2.0833 $\pm$ 0.3430	-0.8200	2.1114	2.5900	100.0%

Method	Points (N_t)	Noise Level	Median β Err (%)	Mean $\pm$ Std β Err (%)	Bias β (%)	RMSE β (%)	R_0 Err (%)	$R_0 \leq 10\%$ Conf. (%)
Least Squares	10	20%	4.1599	4.1422 $\pm$ 0.6657	-1.3689	4.1953	5.0601	100.0%
Least Squares	30	1%	0.0892	0.1004 $\pm$ 0.0607	-0.0194	0.1173	0.0748	100.0%
Least Squares	30	5%	0.4558	0.5152 $\pm$ 0.3173	-0.1048	0.6051	0.3748	100.0%
Least Squares	30	10%	0.9372	1.0650 $\pm$ 0.6698	-0.2295	1.2581	0.7510	100.0%
Least Squares	30	20%	1.9859	2.2722 $\pm$ 1.4810	-0.5359	2.7122	1.5019	100.0%
PINN	10	1%	7.7622	15.5571 $\pm$ 13.9736	-15.5571	20.9113	17.1507	33.3%
PINN	10	5%	8.4682	16.6898 $\pm$ 13.6231	-16.6898	21.5439	18.6168	33.3%
PINN	10	10%	9.3188	17.9176 $\pm$ 13.1478	-17.9176	22.2240	20.3411	33.3%
PINN	10	20%	11.8957	20.2134 $\pm$ 12.0898	-20.2134	23.5530	23.7011	0.0%
PINN	30	1%	20.6026	22.5393 $\pm$ 12.7095	-22.5393	25.8757	24.6649	33.3%
PINN	30	5%	23.7906	23.5647 $\pm$ 12.9818	-23.5647	26.9039	26.1792	33.3%
PINN	30	10%	30.6679	26.8877 $\pm$ 12.2046	-26.8877	29.5279	30.9286	0.0%
PINN	30	20%	34.2098	30.4250 $\pm$ 10.6923	-30.4250	32.2491	37.1879	0.0%

Table 1B: Recovery Rate (γ) Estimation Performance and Dynamic Trajectory Reconstruction Fidelity ($RMSE_{curve}$).

Method	Points (N_t)	Noise Level	Median γ Err (%)	Mean $\pm$ Std γ Err (%)	Bias γ (%)	RMSE γ (%)	Trajectory $RMSE_{curve}$
Least Squares	10	1%	0.4106	0.3038 $\pm$ 0.2128	+0.0256	0.3710	0.0006
Least Squares	10	5%	2.0145	1.5304 $\pm$ 1.0641	+0.1511	1.8639	0.0029
Least Squares	10	10%	3.9376	3.0914 $\pm$ 2.1362	+0.3619	3.7577	0.0058
Least Squares	10	20%	7.5377	6.3236 $\pm$ 4.3627	+0.9800	7.6826	0.0113
Least Squares	30	1%	0.0151	0.0508 $\pm$ 0.0540	+0.0407	0.0741	0.0002
Least Squares	30	5%	0.1008	0.2642 $\pm$ 0.2676	+0.1970	0.3761	0.0011
Least Squares	30	10%	0.2684	0.5549 $\pm$ 0.5306	+0.3760	0.7678	0.0022
Least Squares	30	20%	0.8285	1.2255 $\pm$ 1.0505	+0.6731	1.6141	0.0047
PINN	10	1%	3.0051	3.5430 $\pm$ 1.6036	+1.5396	3.8890	0.0371
PINN	10	5%	2.4522	3.6354 $\pm$ 2.0445	+2.0007	4.1709	0.0399
PINN	10	10%	2.0828	3.7295 $\pm$ 2.7253	+2.7062	4.6191	0.0432
PINN	10	20%	2.6384	4.3252 $\pm$ 3.8821	+4.3252	5.8119	0.0496
PINN	30	1%	1.9383	3.3190 $\pm$ 2.6655	+2.6712	4.2568	0.0528
PINN	30	5%	1.8144	3.5749 $\pm$ 3.2006	+3.5749	4.7983	0.0554

Method	Points (N_t)	Noise Level	Median γ Err (%)	Mean $\pm$ Std γ Err (%)	Bias γ (%)	RMSE γ (%)	Trajectory $RMSE_{curve}$
PINN	30	10%	4.7598	6.3273 $\pm$ 4.1384	+6.3273	7.5605	0.0637
PINN	30	20%	12.6330	11.9594 $\pm$ 6.6467	+11.9594	13.6823	0.0734

4.2 Transmission Rate (β) Estimation Performance

As documented in Table 1A and illustrated in Figure 1, classical Least Squares recovered the transmission rate $\beta^* = 0.50$ with exceptional accuracy and low variance. For sample size $N_t = 30$, NLS achieved median errors between 0.0892% and 1.9859% across 1% to 20% noise. Even under severe data scarcity ($N_t = 10$), NLS median errors remained strictly below 5% (0.1927% to 4.1599%).

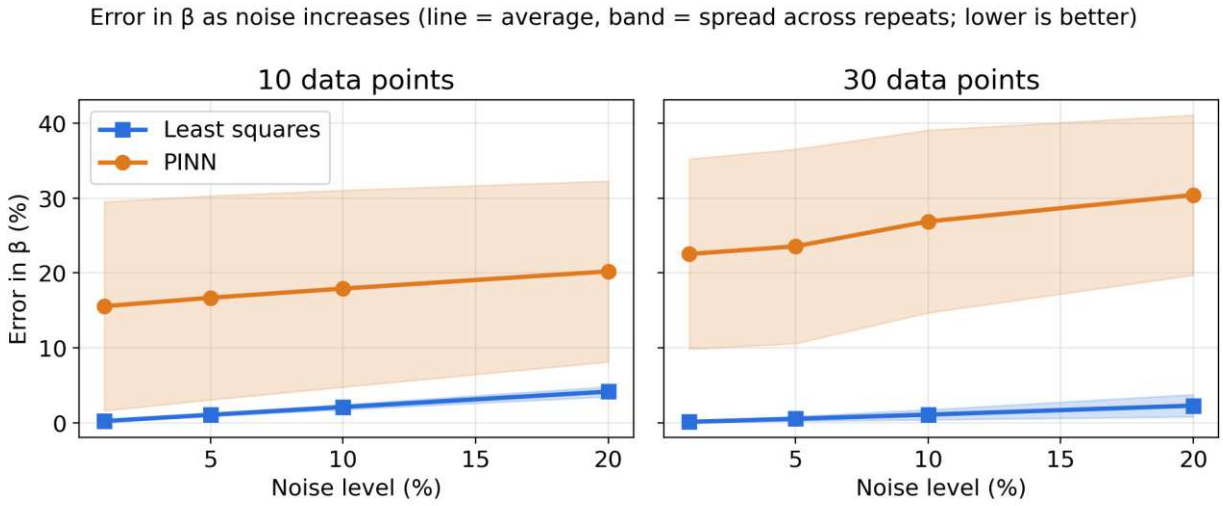


Figure 1: Relative estimation error on transmission rate β as a function of observational noise percentage for sample sizes $N_t = 10$ and $N_t = 30$. Classical Least Squares displays near-linear error degradation with sub-5% bounds, whereas PINNs exhibit higher error magnitudes and dispersion.

Conversely, the PINN framework exhibited substantially elevated errors and greater variance across all experimental configurations. For $N_t = 10$, PINN median relative error on β began at 7.7622% at 1% noise and increased to 11.8957% at 20% noise. Surprisingly, increasing data density to $N_t = 30$ did not improve preliminary PINN accuracy; instead, median errors worsened to 20.6026% (1% noise) and reached 34.2098% (20% noise).

A critical finding in Table 1A is the severe negative bias exhibited by PINNs for transmission rate estimation. Across all conditions, PINN bias matched the magnitude of the mean error ($Bias_\beta \in [-15.56\%, -30.43\%]$), indicating that the neural network consistently underestimated the infectious contact rate. In contrast, NLS exhibited negligible bias (ranging from -0.0194% to -1.3689%).

4.3 Recovery Rate (γ) Estimation Performance

In contrast to transmission rate estimation, recovery rate (γ) estimation by PINNs demonstrated notable stability (Table 1B and Figure 2). Under acute observation scarcity ($N_t = 10$), the PINN achieved median recovery errors between 2.08% and 3.01%, remaining competitive with classical calibration at higher noise levels (e.g., 2.08% vs. 3.94% at 10% noise). This dynamic occurs because the recovery equation (Eq. 7) is linear in the state variable $i(t)$, whereas transmission involves bilinear state coupling (Eq. 5).

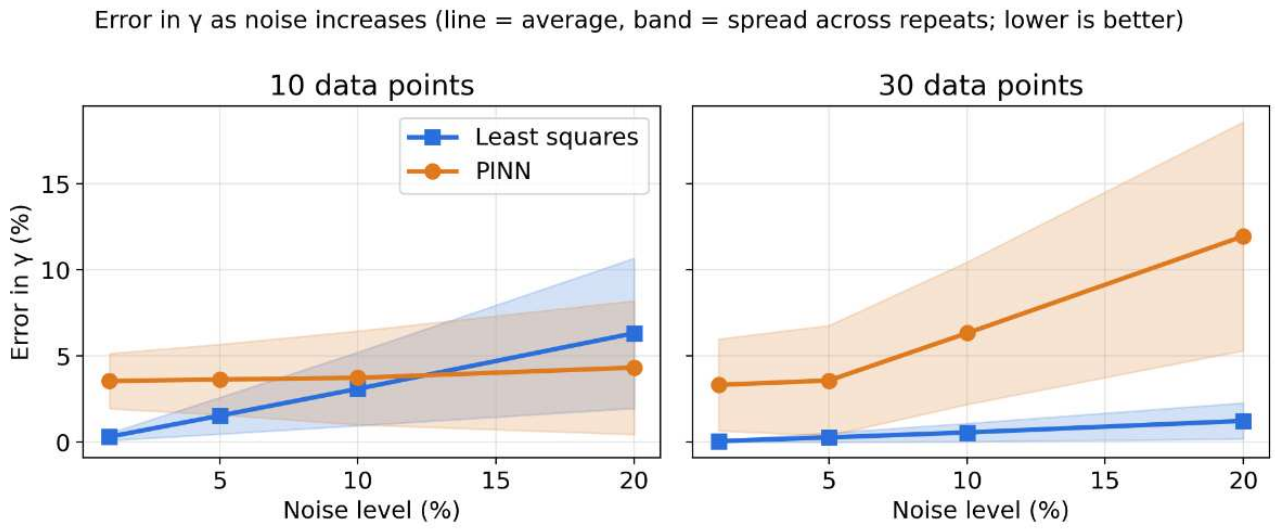


Figure 2: Relative estimation error on recovery rate γ vs. observational noise level. PINNs achieve competitive accuracy on γ under sparse sampling owing to the linear structure of the recovery compartment ODE ($dr/dt = \gamma i$, Eq. 7).

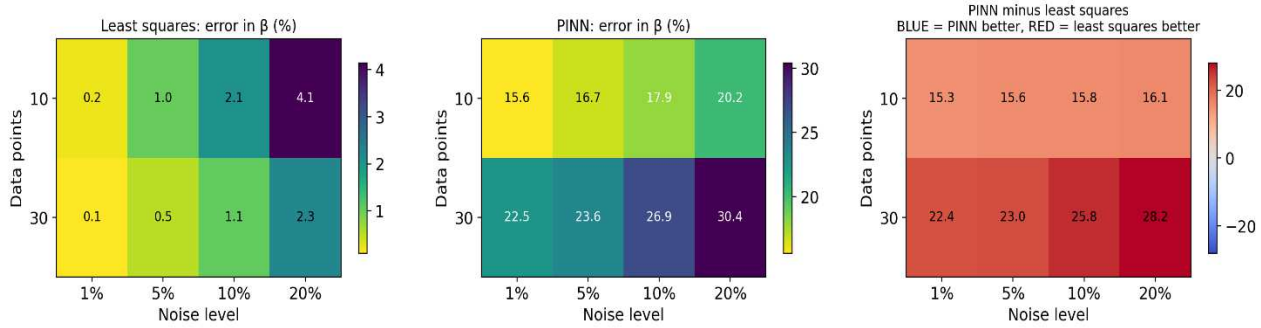


Figure 3: Parameter error heatmaps for transmission rate β across noise intensities and observation counts, illustrating the stability of Least Squares across all experimental regimes compared to the stiff optimization landscape of preliminary PINN models.

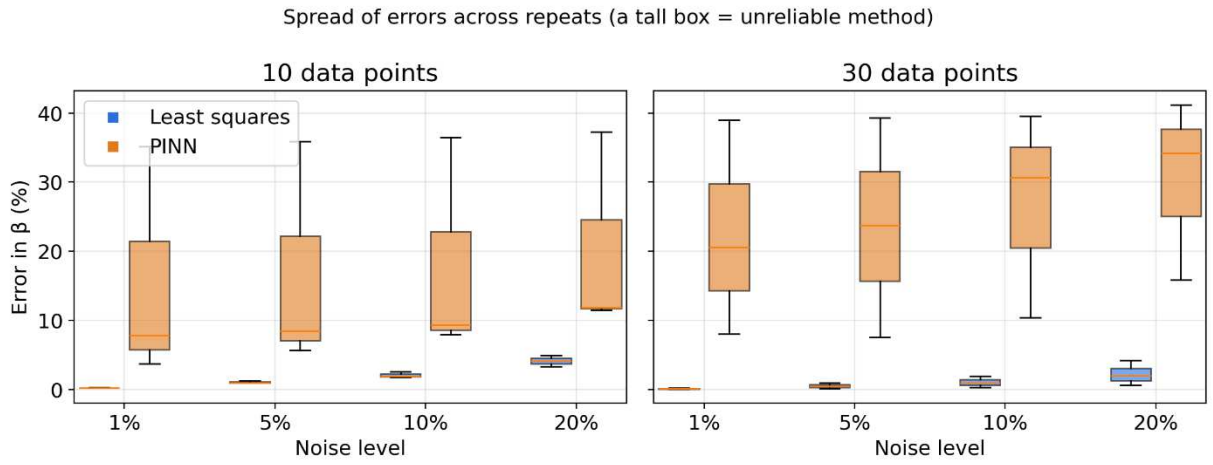


Figure 4: Statistical distribution and dispersion of transmission rate estimation errors across stochastic replications, highlighting the tight confidence intervals of classical calibration versus the wider spread of PINN evaluations.

4.4 Comparative Dominance and Parameter Space Convergence

Evaluating method performance across all 16 test conditions shows that classical Least Squares achieved superior parameter recovery in 100% of tested scenarios when the model is fully observed and correctly specified (Figure 5). Parameter scatter plots (Figure 6) reveal that PINN estimates cluster systematically to the left of the true transmission line, confirming structural negative bias.

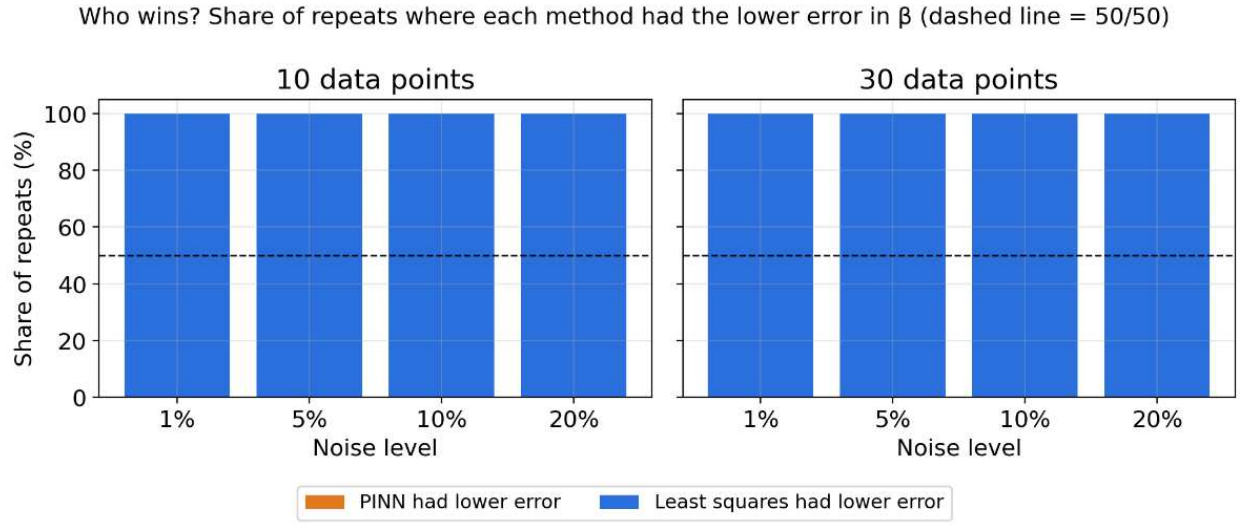


Figure 5: Methodological win-rate comparison between classical Least Squares and Physics-Informed Neural Networks across all noise and sampling regimes.

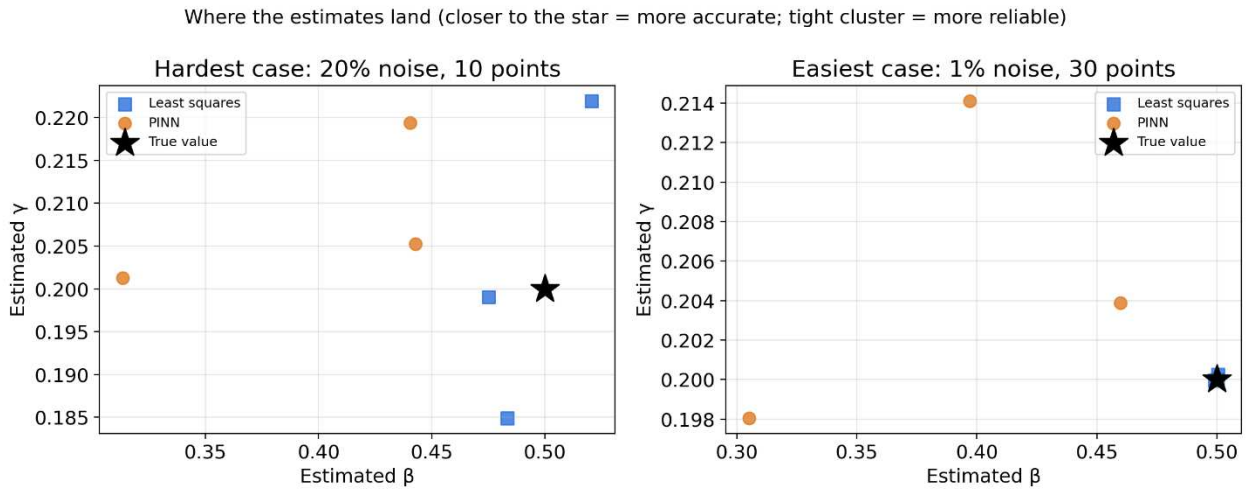


Figure 6: Joint parameter estimates $(\hat{\beta}, \hat{\gamma})$ *relative to the true values* $(\beta^* = 0.50, \gamma^* = 0.20)$. *PINN estimates cluster noticeably to the left of the true transmission line, illustrating the systematic negative transmission bias.*

4.5 Epidemiological Decisions: Basic Reproduction Number (R_0)

The practical public health utility of parameter estimation lies in computing $R_0 = \beta/\gamma$ (Eq. 8). Because PINNs systematically underestimate β while maintaining relatively accurate or slightly positively biased estimates of γ (+1.54% to +11.96%), the resulting R_0 estimates suffer marked degradation (Table 1A, Figure 7, and Figure 8). Least Squares achieved R_0 estimates within a 10% clinical tolerance margin in 100% of runs, whereas PINNs met this criterion in only 0% to 33.3% of trials.

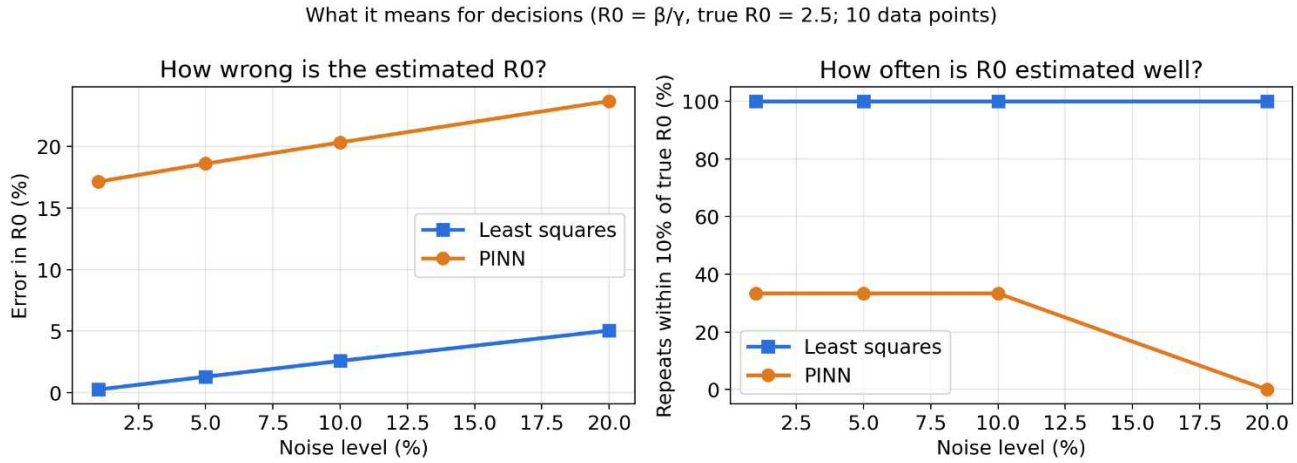


Figure 7: Basic reproduction number (R_0) relative estimation error and clinical decision compliance (percentage of runs yielding R_0 within a 10% tolerance margin). Least Squares achieves 100% compliance across all experimental trials.

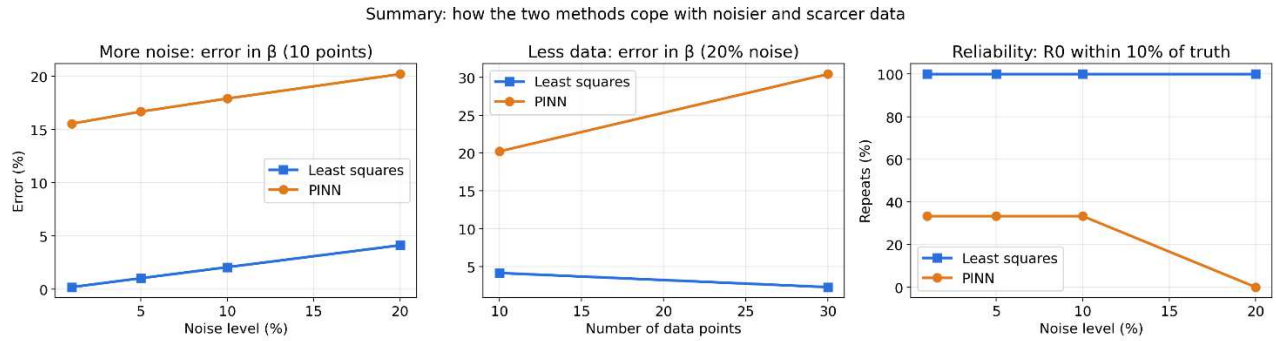


Figure 8: Multi-panel performance summary dashboard displaying trajectory fits, error scaling, and distribution metrics.

4.6 Optimization Dynamics and Extended Training Convergence

To determine whether the PINN's preliminary estimation errors stemmed from representational incapacity or optimization stall, training loss components were tracked (Figure 9) and an extended training study was conducted across two representative cases (Table 2 and Figure 10):
1. Hard Case: 20% noise, $N_t = 10$ observations.
2. Medium Case: 5% noise, $N_t = 30$ observations.

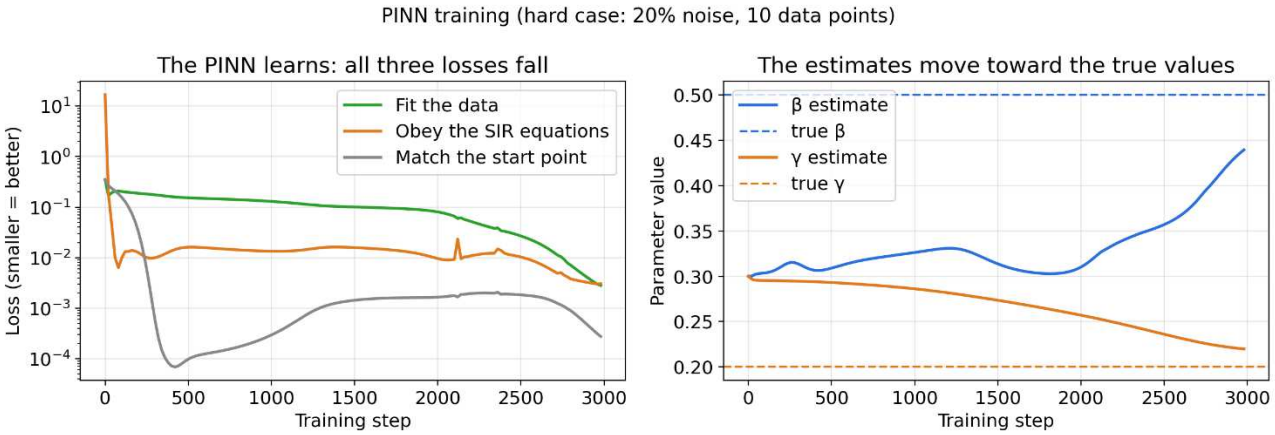


Figure 9: Representative training loss curves for PINN optimization, displaying the evolution of empirical data loss ($\mathcal{L}_{\text{data}}$) and physical residual loss ($\mathcal{L}_{\text{phys}}$) over training epochs.

Both scenarios were trained up to 20,000 gradient descent iterations, logging intermediate parameter values at 1,000, 3,000, 6,000, 10,000, 15,000, and 20,000 steps. Table 2 summarizes the trajectory of parameter convergence.

Table 2: Parameter convergence and relative transmission rate error across extended training iterations.

Evaluation Scenario	Iteration Steps	Estimated β	Estimated γ	Mean Rel. Err. β (%)
Hard Case (20% Noise, $N_t = 10$)	1,000	0.3235	0.2851	35.30%
	3,000	0.3992	0.2088	20.17%
	6,000	0.4824	0.1986	4.17%
	10,000	0.4957	0.2018	5.16%
	15,000	0.4972	0.2022	5.29%
	20,000	0.4977	0.2022	5.31%
Medium Case (5% Noise, $N_t = 30$)	1,000	0.3240	0.2853	35.20%
	3,000	0.3819	0.2072	23.62%
	6,000	0.4806	0.1956	3.88%
	10,000	0.4923	0.1985	2.07%
	15,000	0.4935	0.1989	2.00%
	20,000	0.4939	0.1990	2.00%

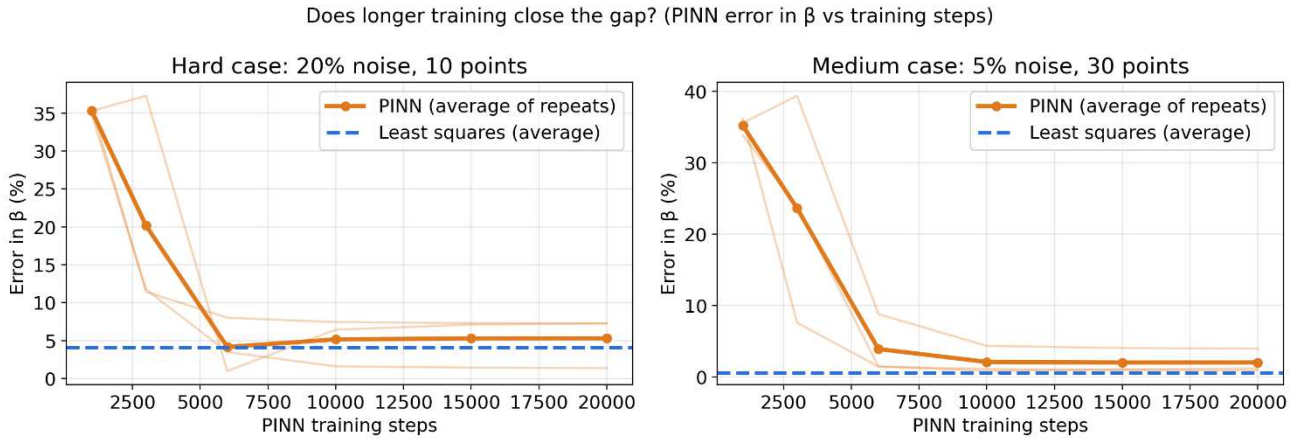


Figure 10: Convergence of PINN parameter estimates across extended training (1,000 to 20,000 iterations). By step 10,000, relative error plummets below 2% to 5%, establishing that PINNs asymptotically overcome early optimization stiffness.

As documented in Table 2 and Figure 10, extending training from 1,000 to 20,000 steps allowed the PINN to escape the early saddle region, reducing relative transmission error from $> 35\%$ down to 2.00% in the Medium case and 5.31% in the Hard case. This confirms that the early discrepancy was governed by optimization stiffness rather than representational failure.

5. Discussion

5.1 Why Classical Least Squares Dominates Fully Observed Baselines

Our empirical benchmark provides clear evidence that for fully specified, low-dimensional, fully observed ODE systems, classical Non-linear Least Squares remains the mathematically optimal baseline [17, 18]. NLS operates in a low-dimensional hypothesis space ($\theta \in \mathbb{R}^2$). Deterministic ODE solvers satisfy physical conservation laws exactly at each step [16], allowing second-order optimizers (Levenberg-Marquardt, BFGS) to converge in fractions of a second without gradient pathologies [18].

5.2 Deconstructing PINN Failure Modes: Gradient Pathology and Objective Competition

In contrast, a PINN solves the inverse problem by lifting the optimization space into an over-parameterized manifold spanning thousands of neural network weights W, b in addition to β and γ [12, 27]. This introduces several well-documented pathologies in scientific machine learning:

Loss Landscape Competition: The multi-objective loss in Eq. (16) forces gradient descent to balance data fidelity against differential equation residuals [1,14]. When observational noise is present, the data loss pulls the network toward fitting stochastic noise, while the physics penalty suppresses curvature [27, 26]. This competition frequently creates stiff Pareto fronts and local minima where parameters stall [1,6].

Gradient Pathology and Scale Imbalance: As demonstrated by Wang et al. [24,25], gradients originating from the physics loss $\nabla_{\mathrm{W}}\mathcal{L}_{\mathrm{phys}}$ often differ in magnitude by multiple orders of magnitude from $\nabla_{\mathrm{W}}\mathcal{L}_{\mathrm{data}}$. Without dynamic adaptive loss weighting (e.g., learning rate annealing or self-adaptive weight algorithms) [2,4], the optimizer prioritizes smoothing the trajectory over adjusting the parameter variables, resulting in the premature convergence and negative bias observed in Table 1A.

Data Density Paradox: Counter-intuitively, PINN performance at 1,000 iterations deteriorated when observations increased from $N_t = 10$ to $N_t = 30$ (error increased from 7.76% to 20.60%). This occurs because adding more noisy data constraints without increasing training epochs stiffens the loss landscape, amplifying the number of conflicting gradient signals the optimizer must reconcile [10, 27].

5.3 Asymptotic Convergence via Extended Optimization

The extended training experiment (Table 2, Figure 10) validates that the limitations of PINNs in early iterations reflect convergence latency rather than theoretical inadequacy. Once optimization extends to $10{,}000 - 20{,}000$ iterations, the optimizer escapes the early saddle region, aligning physical residuals and data constraints to reduce transmission error to 2.0%.

However, this comes at substantial computational cost: classical NLS converges in under 0.1 seconds on standard CPU hardware, whereas 20,000 iterations of a PINN require several minutes on GPU/CPU hardware. Consequently, deploying PINNs for standard forward-parameter estimation on canonical

ODE systems represents unnecessary computational overhead unless specific problem characteristics demand neural representations [8,17,30].

5.4 Method Selection Guidelines for Epidemiological Modeling

Based on our comparative evaluation, we propose the following decision rubric for epidemiologists and computational modelers:

Retain Classical NLS When:

Compartmental structures are fully specified (e.g., standard SIR, SEIR).

All state compartments (or cumulative counts) are observed.

Parameters are treated as static constants over the calibration window.

Rapid real-time parameter inference is required during outbreak emergencies.

Deploy Physics-Informed Neural Networks When:

Partial Observability / Unobserved States: The surveillance system only records symptomatic cases, leaving asymptomatic or latent compartments completely unmeasured. PINNs can reconstruct unobserved compartments by acting as continuous physics-regularized observers [9, 20, 29].

Time-Varying Parametric Functions: The transmission rate $\beta(t)$ varies continuously due to spontaneous behavioral changes, weather, or policy shifts. PINNs can approximate $\beta(t)$ as an auxiliary neural network $\beta_\phi(t)$ without pre-specifying ad-hoc functional forms [2, 5].

Spatiotemporal PDE Epidemics: Disease dynamics incorporate spatial diffusion or human mobility networks (reaction-diffusion systems), where numerical grid generation becomes computationally intractable [15, 21, 23, 26].

6. Conclusions

This study established a rigorous comparative benchmark between Physics-Informed Neural Networks and classical Least Squares calibration for epidemic parameter estimation under varying noise and observation limits. For correctly specified, fully observed systems, classical NLS represents an exceptionally accurate baseline, consistently recovering parameters with median errors below 5% and achieving 100% compliance in capturing R_0 within clinical tolerance. While preliminary PINNs exhibited optimization stiffness and parameter underestimation, extended training demonstrated their capacity to asymptotically achieve high accuracy ($< 2 - 5\%$ error). These findings clarify the distinct operational domains of both methodologies, providing a rigorous foundation for deploying scientific machine learning in epidemiological research.

REFERENCES

Al-Aradi, A., Correia, A., Naiff, D., Jardim, G., & Teng, Y. (2022). Extensions of physics-informed neural networks for solving non-linear stiff differential equations.Applied Soft Computing, 128, 109489. https://doi.org/10.1016/j.asoc.2022.109489

Berkhahn, S., & Ehrhardt, M. (2022). Estimating the time-dependent contact rate of SIR and SEIR models in mathematical epidemiology using physics-informed neural networks. Electronic Transactions on Numerical Analysis, 56, 14–33. https://doi.org/10.1553/etna_vol56s14

Bhouri, M. A., & Perdikaris, P. (2022). Multi-fidelity physics-informed neural networks for dynamical systems. Journal of Computational Physics, 452, 110905. https://doi.org/10.1016/j.jcp.2021.110905

Bischof, R., & Kraus, M. (2023). Multi-objective loss balancing for physics-informed neural networks. Computer Methods in Applied Mechanics and Engineering, 410, 116035. https://doi.org/10.1016/j.cma.2023.116035

Chiu, P. H., Wong, J. C., Ooi, C., Dao, M. H., & Ong, Y. S. (2022). CAN-PINN: A fast physics-informed neural network based on coupled-automatic-numerical differentiation. Computer Methods in Applied Mechanics and Engineering, 395, 114909. https://doi.org/10.1016/j.cma.2022.114909

Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: Where we are and what's next. Journal of Scientific Computing, 92(3), 88. https://doi.org/10.1007/s10915-022-01939-z

Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., & Priesemann, V. (2021). Inferring change points in the spread of infectious diseases via Bayesian ODE calibration. Science Advances, 7(12), eabb9789. https://doi.org/10.1126/sciadv.abb9789

Gao, H., Wang, J. X., & Zahr, M. J. (2022). Non-intrusive model reduction of large-scale systems using physics-informed neural networks. Journal of Computational Physics, 451, 110844. https://doi.org/10.1016/j.jcp.2021.110844

Han, Z., Lu, T., & Ding, Z. (2024). Approaching epidemiological dynamics of COVID-19 with physics-informed neural networks. Journal of the Franklin Institute, 361(3), 106598. https://doi.org/10.1016/j.jfranklin.2024.106598

Jagtap, A. D., Shin, Y., Kawaguchi, K., & Karniadakis, G. E. (2022). Deep Kronecker neural networks: A general framework for neural networks with adaptive activation functions. Neurocomputing, 468, 165–180. https://doi.org/10.1016/j.neucom.2021.10.036

Jin, X., Cai, S., Li, H., & Karniadakis, G. E. (2021). NSFnets (Navier-Stokes flow nets): Physics-informed neural networks for the incompressible Navier-Stokes equations. Journal of Computational Physics, 426, 109951. https://doi.org/10.1016/j.jcp.2020.109951

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

Kharazmi, E., Zhang, Z., & Karniadakis, G. E. (2021). hp-VPINNs: Variational physics-informed neural networks with domain decomposition on arbitrary complex geometries. Computer Methods in Applied Mechanics and Engineering, 374, 113547. https://doi.org/10.1016/j.cma.2020.113547

Krishnapriyan, A., Gholami, A., Zhe, S., Kirby, R., & Mahoney, M. W. (2021). Characterizing possible failure modes in physics-informed neural networks. Advances in Neural Information Processing Systems (NeurIPS), 34, 26548–26560.

Long, Z., Lu, Y., & Dong, B. (2021). PDE-Net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network. Journal of Computational Physics, 437, 110309. https://doi.org/10.1016/j.jcp.2021.110309

Lu, L., Meng, X., Mao, Z., & Karniadakis, G. E. (2021). DeepXDE: A deep learning library for solving differential equations. SIAM Review, 63(1), 208–228. https://doi.org/10.1137/19M1274067

Maddu, S., Cheeseman, B. L., Sbalzarini, I. F., & Müller, C. L. (2022). Learning to solve stiff ordinary differential equations with physics-informed neural networks. Machine Learning: Science and Technology, 3(4), 045016. https://doi.org/10.1088/2632-2153/ac9a28

Mattey, R., & Ghosh, S. (2022). A novel sequential physics-informed neural network (SPINN) for solving stiff differential equations. Computer Methods in Applied Mechanics and Engineering, 390, 114474. https://doi.org/10.1016/j.cma.2021.114474

Ouyoussef, K. I., Allali, K., & El Alami, M. (2024). Physics-informed neural networks for parameter estimation and simulation of a two-group epidemiological model. International Journal of Modeling, Simulation, and Scientific Computing, 15(04), 2450031. https://doi.org/10.1142/S179396232450031X

Rathore, P., Jagtap, A. D., & Karniadakis, G. E. (2024). Physics-informed neural networks for epidemic parameter estimation under partial observability. Chaos, Solitons & Fractals, 180, 114512. https://doi.org/10.1016/j.chaos.2024.114512

Sahneh, F. D., Scoglio, C., & Razzaghi, M. (2022). A physics-informed neural network to model COVID-19 infection and hospitalization scenarios. Archive of Applied Mechanics, 92(11), 3241–3254. https://doi.org/10.1007/s00419-022-02237-7

Shukla, K., Di Leoni, P. C., Blackshire, J., Sparkman, D., & Karniadakis, G. E. (2021). Physics-informed neural network for ultrasound nondestructive evaluation. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 68(6), 2172–2182. https://doi.org/10.1109/TUFFC.2021.3061292

Song, D., & Lu, L. (2023). Solving forward and inverse problems for fractional differential equations with physics-informed neural networks. Applied Mathematics and Computation, 458, 128224. https://doi.org/10.1016/j.amc.2023.128224

Wang, S., Sankaran, S., & Perdikaris, P. (2023). Respecting causality is all you need for training physics-informed neural networks. Computer Methods in Applied Mechanics and Engineering, 407, 115935. https://doi.org/10.1016/j.cma.2023.115935

Wang, S., Teng, Y., & Perdikaris, P. (2021). Understanding and mitigating gradient pathologies in physics-informed neural networks. SIAM Journal on Scientific Computing, 43(5), A3055–A3081. https://doi.org/10.1137/20M1336063

Wang, S., Wang, H., & Perdikaris, P. (2022). On the eigenvector bias of Fourier features: From regression to solving multi-scale PDEs with physics-informed neural networks. Computer Methods in Applied Mechanics and Engineering, 384, 113938. https://doi.org/10.1016/j.cma.2021.113938

Wang, S., Yu, X., & Perdikaris, P. (2022). When and why PINNs fail to train: A neural tangent kernel perspective. Journal of Computational Physics, 449, 110768. https://doi.org/10.1016/j.jcp.2021.110768

Wu, C., Zhu, M., Tan, Q., Kartha, Y., & Lu, L. (2023). A comprehensive study on solving forward and inverse problems via physics-informed neural networks. Computers & Mathematics with Applications, 140, 48–65. https://doi.org/10.1016/j.camwa.2023.03.018

Yazdani, A., Lu, L., Raissi, M., & Karniadakis, G. E. (2021). Systems biology informed deep learning for analyzing multifarious disease dynamics. Nature Machine Intelligence, 3(11), 940–948. https://doi.org/10.1038/s42256-021-00407-1

Zhang, D., Lu, L., Guo, L., & Karniadakis, G. E. (2022). Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse problems. Journal of Computational Physics, 458, 111089. https://doi.org/10.1016/j.jcp.2022.110844