Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Application of Physics-Informed Neural Networks (PINN) in the Kermack-McKendrick Approach to Epidemiological Models

Ícaro V. Freire[1]
UFRB, Amargosa, BA
Michael M. Diniz[2]
IFSP, São Paulo, SP
Moiseis dos S. Cecconello[3]
UFMT, Cuiabá, MT
Rodney C. Bassanezi[4]
UNICAMP, Campinas, SP

Abstract. This work proposes the use of *Physics-Informed Neural Networks* (PINNs) to solve an inverse problem in the Kermack-McKendrick epidemiological model, fitting the parameters of the theoretical curve to the weekly mortality data from the bubonic plague epidemic in Bombay (1905-1906). The original model was revisited using PINNs, which integrate differential equations and machine learning to estimate parameters, minimizing the original mean squared error (MSE) on the data. Compared to the original fit (MSE ≈ 4273.42) and the Least Squares Method (MSE $\approx$ 3208.53), the PINN achieved an MSE of approximately 3204.00, reducing the error by about 25 % compared to the classical model. The implementation combined data normalization, hyperparameter optimization, and automatic differentiation, enabling greater numerical stability and accuracy in parameter estimation. The results highlight the potential of PINNs in epidemic modeling, with potential future applications to COVID-19 data in Brazil.

Keywords. Epidemiology, Machine Learning, PINN, Forward Problems, Inverse Problems

1 Introduction

The curve (in blue) shown in Figure 1 represents the solution of an epidemiological model proposed by Kermack and McKendrick to fit a dataset (in orange) of weekly mortality from a contagious disease that struck Bombay, India, between 1905 and 1906. This result was an application of the theory developed by the authors in the seminal paper that established the foundations of compartmental epidemiological modeling [7].

The resulting curve became a landmark in biomathematics and appears in various references on epidemiological modeling, differential equations, and the history of epidemiology [1]. Since then, several researchers have revisited the original paper, proposing different methods to approximate the curve to the weekly mortality data [1, 4]. Despite these contributions, to date, there are no records of works using *Physics-Informed Neural Networks* (PINNs) for this purpose.

The PINN technique, introduced by Raissi, Perdikaris, and Karniadakis [9], has been widely studied [5] due to its ability to solve problems governed by differential equations, both in the

[1]icarofreire@ufrb.edu.br
[2]michael.diniz@ifsp.edu.br
[3]moiseis.cecconello@ufmt.br
[4]rodneyb@unicamp.br

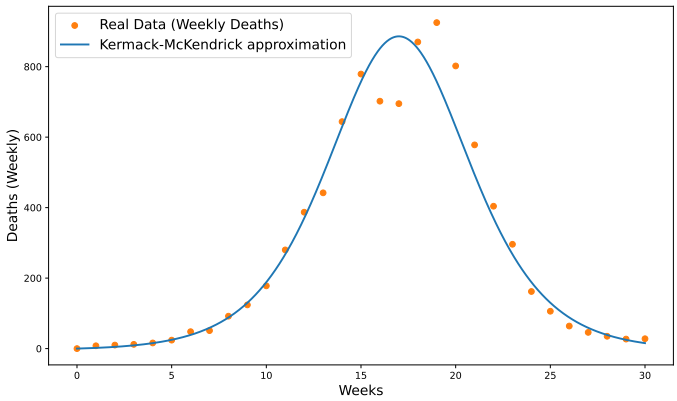


Figure 1: Solution proposed by Kermack-McKendrick (in blue) for the dataset of weekly deaths (in orange) from the bubonic plague epidemic in Bombay, between 1905 and 1906. Source: Authors'.

context of *forward problems*, where the solution of the system is sought given a set of parameters; and in *inverse problems*, where the model parameters are estimated from observational data. In the context of epidemiological modeling, PINNs have been used to approximate epidemic curves [6, 10], but have not yet been applied specifically to the Bombay mortality data modeled by Kermack and McKendrick.

In this work, we propose using the PINN technique to solve an *inverse problem*, estimating the parameters of the Kermack-McKendrick model to minimize the mean squared error between the observed data and the generated curve. Our hypothesis is that this approach will yield a solution better fitted to the data than originally proposed by the authors and superior to traditional methods, such as the Least Squares Method (LSM).

2 Theoretical Framework

2.1 Physics-Informed Neural Networks

The PINN technique, applied to compartmental epidemiological modeling, combines machine learning and differential equations to integrate theoretical knowledge with the fitting of observational data. Unlike traditional neural networks, which learn exclusively from data, PINNs also incorporate constraints imposed by the underlying mathematical model. Thus, optimization occurs both in the adjustable parameters of the system and in the parameters of the neural network.

The training of the PINN minimizes a loss function, loss, typically measured by the mean squared error (MSE), composed of two main contributions: $\mathsf{loss}_{\mathsf{DATA}}$, which quantifies the error between predicted values and observed data; and $\mathsf{loss}_{\mathsf{RESIDUAL}}$, which imposes consistency with the system of differential equations. These terms are balanced by the positive hyperparameters α_1 and α_2, as shown in Equation 1.

$$\mathsf{loss} = \alpha_1 \cdot \mathsf{loss}_{\mathsf{DATA}} + \alpha_2 \cdot \mathsf{loss}_{\mathsf{RESIDUAL}}, \tag{1}$$

For an initial value problem describing epidemiological dynamics, we have

$$\dot{u}(t) = F(u(t), \theta), \qquad u(t_0) = u_0, \tag{2}$$

where θ represents the set of model parameters. The neural network $\mathsf{NN}_\theta^{w,b} : \mathbb{R} \to \mathbb{R}^n$, parametrized by weights w and biases b, approximates the solution $u(t)$. That is, the function $\hat{u}(t) = \mathsf{NN}_\theta^{w,b}(t)$ must satisfy (2).

DOI: 10.5540/03.2026.012.01.0255

The approach proposed by PINNs allows solving both forward problems, where all parameters θ are known, and inverse problems, where a subset θ_d of the parameters needs to be estimated. We define $\theta = \theta_f \cup \theta_d$, where θ_f contains the fixed parameters and θ_d represents the unknown parameters. If $\theta_d \neq \varnothing$, we are dealing with an inverse problem; otherwise, we have a predictive problem.

The training uses points $t_{data} = \{ t_0, t_1, \ldots, t_{r-1} \}$ from the observed data and *collocation points* $t_{phy} = \left\{ t_0^*, \ldots, t_{s-1}^* \right\}$, uniformly distributed in the study interval, with $t_{phy} \supset t_{data}$ in general.

Adherence to epidemiological modeling is imposed by the residuals of the differential equations, defined as

$$\operatorname{res}(\hat{u}(\tau_j), \theta) = \frac{d\hat{u}(\tau_j)}{dt} - F(\hat{u}(\tau_j), \theta), \quad \tau_j \in t_{phy}.$$

We emphasize that the derivative, $\frac{d\hat{u}}{dt}$, is computed exactly via automatic differentiation, and the ODE residual is evaluated at the collocation points [2].

Thus, the total loss function is given by:

$$loss = \alpha_1 \cdot \frac{1}{r} \sum_{i=0}^{r-1} \left(\hat{u}(t_i) - u(t_i) \right)^2 + \alpha_2 \cdot \frac{1}{s} \sum_{j=0}^{s-1} \left(\operatorname{res}(\hat{u}(\tau_j), \theta) \right)^2.$$

2.2 Kermack-McKendrick Model

The SIR model divides the population into susceptible (S), infected (I), and recovered (R) individuals. It has been studied in various ways, such as in the Fuzzy approach, which deals with uncertainties [3]; in the fractional perspective, which incorporates nonlinear dynamics; and with the use of *machine learning* for more accurate predictions; etc. [11]. The epidemiological model used in the inverse problem, with PINNs, was originally obtained by Kermack and McKendrick after significant biological and mathematical simplifications of the SIR model. The authors reduced the system of differential equations

$$\dot{x} = -\kappa x y \qquad \dot{y} = \kappa x y - l y \qquad \dot{z} = l y$$

where x, y, and z represent, respectively, the number of susceptible, infectious, and removed (recovered or deceased) individuals, such that $x + y + z = N$ (with N being the initial population density); to a single equation:

$$\dot{z}(t) = a \operatorname{sech}^2(\omega t - \phi), \tag{3}$$

where a, ω, and ϕ are parameters related to the constants and initial conditions [7, p. 714]. By applying these considerations to a specific period (30 weeks) of a bubonic plague epidemic that struck Bombay, they considered $z(0) = 0$, finding the constants: $a = 890, \omega = 0.2,$ and $\phi = 3.4$ The solution of Equation (3) returns the cumulative number of deaths, $z(t)$; however, it is notable that the original approximation by the authors used the weekly number of deaths, $\dot{z}(t)$. Therefore, in the calculation of the residual in the PINN, the derivative of expression (3) is necessary:

$$\ddot{z}(t) = -2a\omega \operatorname{sech}^2(\omega t - \phi) \tanh(\omega t - \phi), \tag{4}$$

3 Methodology

The data were manually collected from Table IX, on page 753, available in the digitization by the Advisory Committee of the Royal Society [12]. After selecting the corresponding weeks, we produced a `.csv` file for testing. Figure 1 shows the considered points (in orange). The analyzed period was from the week of December 17, 1905, to the week of July 21, 1906, totaling 30 weeks.

DOI: 10.5540/03.2026.012.01.0255

The approximation of the parameters of Equation (3) by the Least Squares Method was considered for an initial comparison, as it is a traditional and widely used method in epidemiological modeling problems. We used the `curve_fit()` function from the SciPy library in Python, adopting as initial parameters those suggested by Kermack-McKendrick.

Regarding the neural network architecture, we considered one *input*, corresponding to time; one *output*, representing the weekly number of deaths; and one *hidden layer* with 128 neurons. The chosen activation function was $\tanh$, applied in the *hidden layer*. For training, we used 600×10^3 *epochs*, a *learning rate* of 1×10^{-5}; the mean squared error as the metric for calculating loss; and Adam [8] as the optimization algorithm. Additionally, the training set included the entire dataset, including the initial condition, totaling 31 observations. The framework used was PyTorch, and to ensure reproducibility, we fixed the random seed with `torch.manual_seed(7)`.

Following the recommendations of [13], we kept time and the weekly number of deaths on the same scale by normalizing them, dividing each set by its maximum value, respectively. The *collocation points* were defined in the interval $[0, 1]$, with the same number of points as the training set, uniformly distributed.

Considering $\{\psi(t_i)\}_{i=0}^{30}$ as the normalized set of weekly deaths ($\psi \equiv \dot{z}$ normalized); $\hat{\psi}(t)$ as the neural network with parameters θ_d for estimation; and the normalization of Equation (4), we have:

$$\mathrm{res}(\hat{\psi}(\tau_j), \theta) = \frac{d\hat{\psi}(\tau_j)}{dt} - \Big(-2 \cdot a \cdot \omega \cdot (T/M) \cdot \mathrm{sech}^2\left(\omega \cdot T \cdot \tau_j - \phi\right) \tanh(\omega \cdot T \cdot \tau_j - \phi) \Big),$$

where τ_j is in the *collocation point*; T is the maximum value of the time instant from the real data, and M is the maximum element of the weekly deaths dataset. Thus, we can define:

$$\mathsf{loss} = \alpha_{\mathsf{week}_1} \cdot \frac{1}{31} \sum_{i=0}^{30} \Big(\hat{\psi}(t_i) - \psi(t_i) \Big)^2 + \alpha_{\mathsf{week}_2} \cdot \frac{1}{31} \sum_{j=0}^{30} \big(\mathrm{res}(\hat{\psi}(\tau_j), \theta) \big)^2$$

The balancing hyperparameters, α_{week_1} and α_{week_2}, were inserted directly into the neural network for estimation, with the exponential function applied to each to ensure positive values. Since we had already calculated the approximations of the constants using LSM, a_{mmq}, ω_{mmq}, and ϕ_{mmq}, we chose to estimate the constants a, ω, and ϕ, associated with the weekly data, as follows: $p_{k+1} = p_{\mathsf{mmq}} + \rho * \tanh(p_k)$, where p_k is one of the mentioned parameters, and p_{mmq} is the parameter estimated by LSM associated with it. We proceeded in this manner because the estimate made by LSM significantly reduced the initial error; thus, it could be a good starting point for the PINN approximations. And, indeed, since $-1 \leq \tanh(x) \leq 1$, the new parameter p_{k+1} will be the value estimated by LSM plus or minus a small proposed variation: in the case of a, we chose $\rho = 0.8$; and for the constants ω and ϕ, we considered $\rho = 0.05$. In these settings, there is no overfitting in the neural network's approximation, and the parameters are properly estimated.

In both cases — with the original and the PINN-estimated parameters — the ODEs are numerically solved using LSODA, an adaptive method for stiff and non-stiff systems.

4 Results and Discussion

The approximations for the constants a, ω and ϕ, made by the PINN, are shown in Table 1. It is noted that all constants were reduced compared to the original values proposed by Kermack-McKendrick.

DOI: 10.5540/03.2026.012.01.0255

Table 1: Estimates produced by the PINN with weekly data.

Parameter	Value
a	873.982 482 910 156 2
ω	0.190 465 152 263 641 36
ϕ	3.322 405 815 124 511 7

Figure 2 shows some of the histories analyzed during the training of the neural network, considering the ordinate axis always in logarithmic scale. Note, in Figure 2a, which displays the history of the loss function, that it is decreasing, showing a steep decline until approximately 300×10^3 *epochs*; and then, it becomes smoother. The nearly linear shape of this graph in the first part indicates that the original decrease is approximately exponential, i.e., it decays rapidly at the beginning, stabilizing later. Figures 2b, 2c and 2d show, respectively, the history of the estimated constants: a, ω, and ϕ. It is observed that all seem to converge and stabilize after a considerable number of iterations.

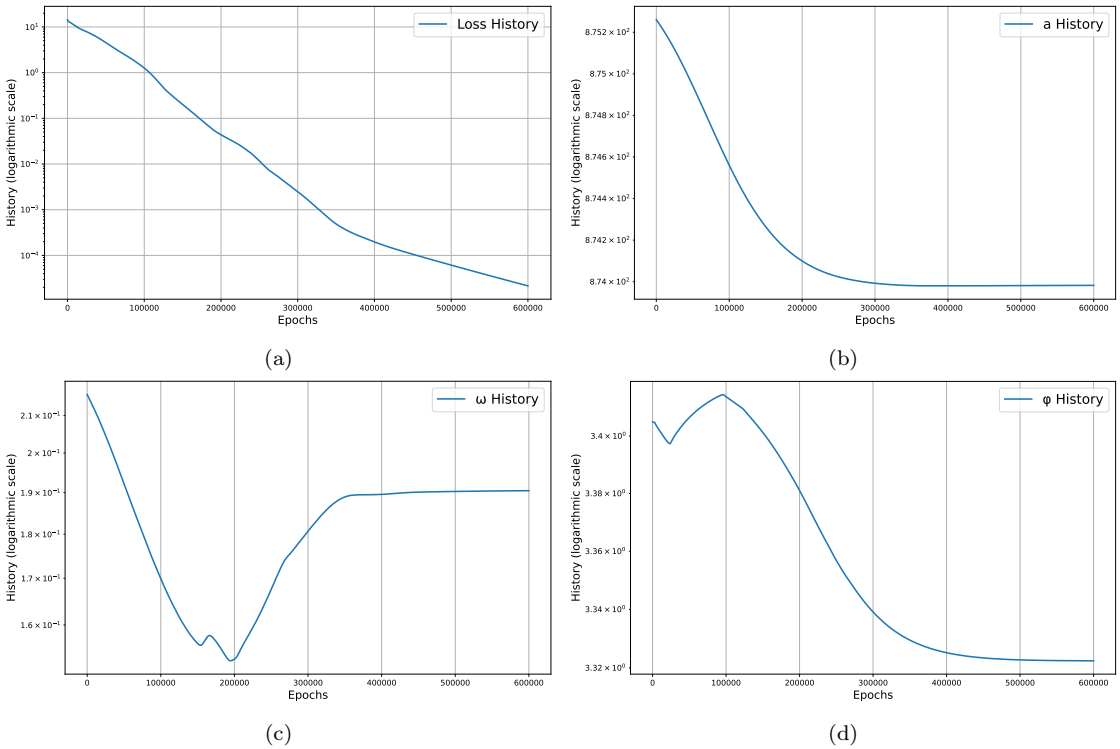
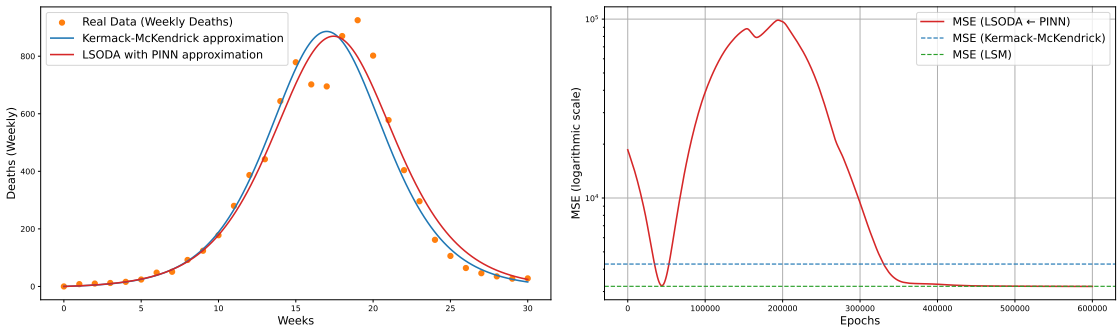


Figure 2: Some of the histories considered during the training of the neural network: loss history, in Figure 2a; a history, in Figure 2b; ω history, in Figure 2c; and, ϕ history, in Figure 2d. Source: Authors'.

This will influence, as we will see, the displacement of the estimated curve: slightly to the right of the original approximation, reducing the difference between the estimated points and some of the higher values (which directly influences the minimization of the MSE). Additionally, at the end of the observation period, the difference between the estimated points and the real values is greater compared to the approximation made by Kermack-McKendrick; however, since the real values in this interval are low (low mortality), these terms do not significantly increase the final MSE calculation.

From the above, the MSE produced by the LSODA approximation, with the parameters estimated by the PINN, was approximately 3204.0021. Relative to the error produced by the approximation made by Kermack-McKendrick, this is a reduction of about 25.02 %; also achieving a lower MSE than that produced by LSM (3208.5309), as shown in Figure 3b. Just out of curiosity, if the training continued until 800×10^3 *epochs*, the mean squared error decreases to 3199.5450, i.e., it remains lower than the other errors considered earlier, suggesting some stabilization.

Finally, Figure 3a shows the compared approximations: in blue, the original approximation by Kermack-McKendrick; and in red, the approximation using the LSODA method with the values of the constants estimated by the PINN.



(a) (b)

Figure 3: **(a)**: The blue curve represents the approximation with the parameters of Kermack-McKendrick (real data in orange). The red curve corresponds to the PINN estimate. **(b)**: History of the approximation with the PINN (red), comparing it with the MSE of the Kermack-McKendrick model (blue) and LSM (green). Source: Authors'.

5 Final Considerations

In this work, we solved an inverse modeling problem using the PINN technique, demonstrating its effectiveness in parameter estimation for epidemiological models. The classical model proposed by Kermack and McKendrick was originally used to approximate the evolution of weekly deaths during a bubonic plague epidemic in Bombay, India, over a 30-week period between 1905 and 1906. In it, the parameters were adjusted to minimize the error between the theoretical epidemiological curve and the observed data, resulting in a mean squared error of approximately 4273.42.

By applying the PINN technique, we were able to reduce this error by about 25 %, achieving a value around 3204.00, also lower than the error produced by traditional numerical methods, such as LSM, whose mean squared error was 3208.53. These results highlight the superiority of the PINN method in solving inverse problems in epidemiology, emphasizing its potential as a robust tool for parameter estimation in compartmental models.

As a follow-up to this study, we intend to explore the use of PINNs for both inverse and predictive problems, including applications to COVID-19 data in Brazil. The flexibility and precision of this method pave the way for new approaches in epidemic modeling and forecasting, contributing significantly to the understanding and control of infectious diseases.

DOI: 10.5540/03.2026.012.01.0255

References

[1] N. Bacaër. "The model of Kermack and McKendrick for the plague epidemic in Bombay and the type reproduction number with seasonality". In: **Journal of Mathematical Biology** 64 (2012), pp. 403–422. DOI: 10.1007/s00285-011-0417-5.

[2] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind. "Automatic differentiation in machine learning: a survey". In: **Journal of Machine Learning Research** 18.153 (2018). Accessed on 2025-02-05, http://jmlr.org/papers/v18/16-107.html, pp. 1–43.

[3] M. S. Cecconello, M. M. Diniz, and R. C. Bassanezi. "Analyzing the Sars-Cov-2 Pandemic Outbreak Using Fuzzy Sets and the SIR Model". In: **Applications of Fuzzy Techniques**. Vol. 500. Springer International Publishing, 2023, pp. 40–51. DOI: 10.1007/978-3-031-16038-7_5.

[4] D. J. D. Earn, S. W. Park, and B. M. Bolker. "Fitting Epidemic Models to Data: A Tutorial in Memory of Fred Brauer". In: **Bulletin of Mathematical Biology** 86.109 (2024), pp. 1–32. DOI: 10.1007/s11538-024-01326-9.

[5] A. Farea, O. Yli-Harja, and F. Emmert-Streib. "Understanding Physics-Informed Neural Networks: Techniques, Applications, Trends, and Challenges". In: **AI** 5.3 (2024), pp. 1534–1557. DOI: 10.3390/ai5030074.

[6] S. Han, L. Stelz, H. Stoecker, L. Wang, and K. Zhou. "Approaching epidemiological dynamics of COVID-19 with physics-informed neural networks". In: **Journal of the Franklin Institute** 361.6 (2024), pp. 1–29. DOI: 10.1016/j.jfranklin.2024.106671.

[7] W. O. Kermack and A. G. McKendrick. "A contribution to the mathematical theory of epidemics". In: **Proceedings of the Royal Society of London. Series A** 115 (1927), pp. 700–721. DOI: 10.1098/rspa.1927.0118.

[8] D. P. Kingma and J. Ba. **Adam: A Method for Stochastic Optimization**. Accessed on 2025-02-05, https://arxiv.org/abs/1412.6980. 2017.

[9] 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". In: **Journal of Computational Physics** 378 (2019), pp. 686–707. DOI: 10.1016/j.jcp.2018.10.045.

[10] M. Raissi, N. Ramezani, and P. Seshaiyer. "On parameter estimation approaches for predicting disease transmission through optimization, deep learning and statistical inference methods". In: **Letters in Biomathematics** 6.2 (2019), pp. 1–26. DOI: 10.1080/23737867.2019.1676172.

[11] C. I. Siettos and L. Russo. "Mathematical modeling of infectious disease dynamics". In: **Virulence** 4.4 (2013), pp. 295–306. DOI: 10.4161/viru.24041.

[12] The Advisory Committee Appointed by the Secretary of State for India, the Royal Society and the Lister Institute. "Reports on Plague Investigations in India". In: **The Journal of Hygiene** 7.6 (1907). Accessed on 2025-01-16, http://www.jstor.org/stable/4619420, pp. 693–985.

[13] S. Wang, Y. Teng, and P. Perdikaris. "Understanding and Mitigating Gradient Flow Pathologies in Physics-Informed Neural Networks". In: **SIAM Journal on Scientific Computing** 43.5 (2021), pp. 1534–1557. DOI: 10.1137/20M1318043.

