International Journal of Hydrogen Energy

Physics-Informed Neural Networks for Parameter Identification of an Equivalent Circuit Model of PEM Electrolyzers

--Manuscript Draft--

Manuscript Number:	
Article Type:	VSI: EHEC 2026 (Sánchez)_Full length article
Section/Category:	Electrolysis / Electrolyzers
Keywords:	Physics-Informed Neural Networks (PINNs) Physics-Informed Long Short-Term Memory (PI-LSTM) Equivalent Circuit Model (ECM) PEM Electrolyzer Parameter Identification
Corresponding Author:	Milad Sabamehr Polytechnic University of Catalonia - Barcelona Tech SPAIN
First Author:	Milad Sabamehr
Order of Authors:	Milad Sabamehr
	Carles Batlle, PhD
	Maria Serra, PhD
Abstract:	Accurate parameter identification of equivalent circuit models (ECMs) is essential for monitoring and control of Proton Exchange Membrane (PEM) electrolyzers. This paper compares Physics- Informed Neural Networks (PINNs) and Physics-Informed Long Short-Term Memory (PI- LSTM) networks for identifying six unknown parameters of a second-order PEM electrolyzer ECM. Both approaches are trained using multiple excitation current profiles, while a sensitivity analysis is performed to select the PI-LSTM architecture. The results demonstrate that PI-LSTM consistently outperforms the conventional PINN, estimating all parameters with errors below 5%, whereas PINN exhibits significantly larger errors for several dynamic parameters. Moreover, PI-LSTM generalizes well to unseen current profiles and maintains high accuracy under noisy measurements, achieving an RMSE of 1.436 mV, an MAE of 1.215 mV, and an 2 of 0.9999. These findings highlight PI-LSTM as a robust and accurate framework for PEM electrolyzer parameter identification.

Dear Editor,

We are pleased to submit our manuscript entitled **"Physics-Informed Neural Networks for Parameter Identification of an Equivalent Circuit Model of PEM Electrolyzers"** for consideration for publication in the ***International Journal of Hydrogen Energy***.

Proton Exchange Membrane (PEM) electrolyzers are key components of future green hydrogen production systems due to their fast dynamic response and compatibility with renewable energy sources. Accurate identification of equivalent circuit model (ECM) parameters is essential for model-based control, monitoring, diagnostics, and digital twin development. In this work, we investigate the application of physics-informed deep learning for parameter identification of PEM electrolyzers by conducting a comprehensive comparison between conventional Physics-Informed Neural Networks (PINNs) and Physics-Informed Long Short-Term Memory (PI-LSTM) networks.

The main contributions of this manuscript are:

- We present a systematic comparison of conventional PINNs and PI-LSTM for identifying the six unknown parameters of a second-order PEM electrolyzer equivalent circuit model.
- We demonstrate that the proposed PI-LSTM framework accurately estimates all circuit parameters with relative errors below 5%, significantly outperforming conventional PINNs in estimating the dynamic RC-network parameters.
- We evaluate the proposed approach using multiple excitation current profiles and demonstrate strong generalization to unseen operating conditions.
- We further investigate the robustness of the proposed framework under noisy voltage measurements, showing that accurate parameter identification is maintained despite measurement uncertainty.

We believe this work is well aligned with the scope of the *International Journal of Hydrogen Energy*, as it combines advances in physics-informed machine learning with hydrogen energy technologies and provides a practical methodology for improving the modeling and control of PEM electrolyzers. The proposed framework has potential applications in real-time monitoring, fault diagnosis, parameter estimation, and digital twin development for hydrogen production systems.

We would also like to inform you that an earlier version of this work was presented as an **oral presentation** at the **European Hydrogen Energy Conference (EHEC 2026)** held in Seville, Spain. The conference publication consisted only of an extended abstract. The present manuscript is a substantially expanded and revised version of that work. It includes a comprehensive description of the proposed methodology, additional theoretical developments, extensive experimental validation, a systematic comparison between PINN and PI-LSTM, sensitivity analysis of the PI-LSTM architecture, evaluation of generalization to unseen current profiles, robustness analysis under noisy measurements, and significantly expanded discussion and conclusions. Accordingly, the present manuscript represents a substantial and original contribution beyond the conference abstract.

This manuscript is original, has not been published previously, and is not under consideration for publication elsewhere. All authors have reviewed and approved the manuscript and agree with its submission to the ***International Journal of Hydrogen Energy***. The authors declare that there are no conflicts of interest regarding this work.

Thank you for your consideration. We appreciate the opportunity to submit our work to the *International Journal of Hydrogen Energy* and look forward to your response.

Sincerely,

Milad Sabamehr (Co-Author)

Institut de Robòtica i Informàtica Industrial (IRI), CSIC-UPC
Barcelona, Spain

Email: milad.sabamehr@upc.edu

Highlights

Physics-Informed Neural Networks for Parameter Identification of an Equivalent Circuit Model of PEM Electrolyzers

Milad Sabamehr, Carles Batlle, Maria Serra

- Comparison of PINN and PI-LSTM for PEM electrolyzer ECM identification.
- PI-LSTM estimates six ECM parameters with errors below 5%
- PI-LSTM generalizes well to unseen inputs and noisy measurements.

Highlights

Physics-Informed Neural Networks for Parameter Identification of an Equivalent Circuit Model of PEM Electrolyzers

Milad Sabamehr, Carles Batlle, Maria Serra

- Comparison of PINN and PI-LSTM for PEM electrolyzer ECM identification.
- PI-LSTM estimates six ECM parameters with errors below 5%.
- PI-LSTM generalizes well to unseen inputs and noisy measurements.

Physics-Informed Neural Networks for Parameter Identification of an Equivalent Circuit Model of PEM Electrolyzers

Milad Sabamehr[a], Carles Batlle[b] and Maria Serra[a]

[a] *Institut de Robòtica i Informàtica Industrial (IRI), CSIC-UPC, Llorens i Artigas 4-6, Barcelona, 08028, Spain*
[b] *Departament de Matemàtiques, Institut d'Organització i Control, EPSEVG, Universitat Politècnica de Catalunya (UPC), Vilanova i la Geltrú, 08800, Spain*

ARTICLE INFO

Keywords:
Physics-Informed Neural Networks (PINNs)
Physics-Informed Long Short-Term Memory (PI-LSTM)
Equivalent Circuit Model (ECM)
PEM Electrolyzer
Parameter Identification

ABSTRACT

Accurate parameter identification of equivalent circuit models (ECMs) is essential for monitoring and control of Proton Exchange Membrane (PEM) electrolyzers. This paper compares Physics-Informed Neural Networks (PINNs) and Physics-Informed Long Short-Term Memory (PI-LSTM) networks for identifying six unknown parameters of a second-order PEM electrolyzer ECM. Both approaches are trained using multiple excitation current profiles, while a sensitivity analysis is performed to select the PI-LSTM architecture. The results demonstrate that PI-LSTM consistently outperforms the conventional PINN, estimating all parameters with errors below 5%, whereas PINN exhibits significantly larger errors for several dynamic parameters. Moreover, PI-LSTM generalizes well to unseen current profiles and maintains high accuracy under noisy measurements, achieving an RMSE of 1.436 mV, an MAE of 1.215 mV, and an R^2 of 0.9999. These findings highlight PI-LSTM as a robust and accurate framework for PEM electrolyzer parameter identification.

1. Introduction

The transition to sustainable energy systems has driven a growing global interest in green hydrogen production as an essential solution for renewable energy storage and improved grid flexibility. Among the different hydrogen production technologies, Proton Exchange Membrane (PEM) electrolyzers stand out as a promising option due to their high efficiency, compact configuration, and fast dynamic response [7, 10]. However, for PEM electrolyzers to be an economically viable solution, some improvements are still necessary. The modeling of PEM electrolyzers is crucial to obtaining most of these improvements, which include ensuring safe and efficient operation, enhancing dynamic performance, minimizing degradation , and enabling monitoring and diagnosis, This is because PEM electrolyzers used for green hydrogen production are typically coupled with renewable energy sources, whose intermittent and fluctuating power output requires accurate dynamic models for system analysis[28]. Modeling enables the simulation of different operating conditions and design configurations, allowing for the development of more efficient and cost-effective systems.

Electrical equivalent circuit models (ECMs) have been widely employed because of representing the main dynamic behavior of PEM electrolyzers while preserving a level of simplicity suitable for parameter identification and real-time implementation [22]. In the study of [3], several ECM structures effectively describe and evaluate proton exchange membrane electrolyzer transient responses across various operating conditions. This provides valuable insight into their control performance and durability under fluctuating loads. Among these models, second-order ECMs, typically composed of two RC branches together with ohmic and reversible voltage terms, have shown a strong capability to reproduce the characteristic transient polarization behavior of PEM electrolyzers exposed to time-varying current profiles [8].

Likewise, the work presented in [13] proposed an ECM to improve hydrogen production efficiency by identifying the optimal operating frequency in pulsed electrolysis. The study demonstrated that ECMs can provide a simplified yet effective representation of the electrochemical and physical phenomena occurring within PEM electrolyzers. Nevertheless, despite their practical advantages, ECMs rely on accurate parameter estimation, and conventional identification methods are often sensitive to measurement noise, which can degrade parameter estimation accuracy and reduce the reliability of the resulting models [19].

Corresponding Author
milad.sabamehr@upc.edu (M. Sabamehr)

Recent studies have increasingly investigated data-driven techniques for modeling the dynamic behavior of PEM electrolyzers. For instance, [27] employed real-time operational data in combination with Long Short-Term Memory (LSTM) and Artificial Neural Network (ANN) models to accurately capture the transient response of the system under varying operating conditions. Likewise, [21] evaluated several machine learning algorithms, including Support Vector Regression (SVR), Extreme Gradient Boosting (XGB), and ANN, for predicting electrolyzer performance and analyzing the sensitivity of transport properties. Their results demonstrated that ANN and XGB consistently outperformed SVR in predicting current density and cell efficiency.

Although such data-driven methods can provide strong predictive performance within the range of the training data, they often exhibit limited extrapolation capability and offer little physical interpretability, especially when applied to unseen operating conditions. To address these challenges, Study of [20] proposed the use of Physics-Informed Neural Networks (PINNs), which integrate governing physical equations directly into the training process. This approach enables the solution of both forward and inverse problems while reducing the dependence on large volumes of experimental data and improving the generalization ability of neural networks.

In PINNs, physical knowledge is incorporated into the training process by embedding the governing differential equations and boundary conditions directly into the loss function alongside observational data. As a result, the network learns solutions that satisfy both the available measurements and the underlying physical principles, improving generalization while reducing the dependence on large training datasets.

Although PINNs have been successfully applied to both forward and inverse problems, recent studies have increasingly explored their use for parameter estimation in electrical circuits. For example, [29] demonstrated that PINNs can accurately estimate electrical parameters, such as resistance and capacitance, in simple circuit models. Similarly, [26] proposed a Generalizable Physics-Informed Fourier Neural Network (GPIFNN) for electrical circuit analysis, addressing both forward and inverse modeling tasks. Their results showed that the proposed framework outperformed conventional PINNs during both training and testing across various circuit configurations while providing more accurate predictions of resistance, inductance, and capacitance. More recently, [24] applied PINNs to identify the parameters of first and second order equivalent circuit models of PEM electrolyzers, demonstrating their effectiveness for parameter estimation in electrochemical systems.

Despite these promising results, conventional PINNs still face challenges when modeling long sequential datasets. Conventional PINNs are typically based on feedforward neural networks and rely on automatic differentiation to enforce the governing differential equations throughout the temporal domain. As the sequence length increases, training becomes computationally more demanding, and the network may struggle to capture the long-term temporal dependencies that characterize dynamic systems [1].

To address these limitations, researchers have explored integrating recurrent neural network architectures with physical constraints. Among different recurrent neural network architectures, Long Short-Term Memory (LSTM) networks have attracted considerable attention for their ability to model sequential and temporal data via dedicated memory cells and gating mechanisms [11]. Several studies have successfully applied physics-informed recurrent architectures to dynamical systems, for instance, study [25] developed a Physics-Informed Recurrent Neural Network (PiRNN) for modeling resistive random-access memory (RRAM) devices, demonstrating that the integration of device physics with recurrent learning, improves the representation of history-dependent electrical behavior and also enhances predictive performance.

The paper [31] proposed PILNet (Physics-Constrained Informer-LSTM Network) to estimate the state of charge (SOC) and internal circuit parameters of lithium-ion batteries using observable battery data. The proposed framework enforced physical consistency through constraints derived from battery physics and outperformed several state-of-the-art methods, including Informer, Mamba-Lithium, and Informer-LSTM, in terms of RMSE and MSE.

Research [18] proposed a Physics-Informed Long Short-Term Memory (PI-LSTM) framework for estimating unknown parameters of nonlinear dynamical systems using limited training data. The proposed method was validated on the Lorenz system under stable, chaotic, and periodic operating regimes, as well as on the hyperchaotic Rössler system. The results demonstrated a high parameter estimation accuracy, with estimation errors generally below 1% for all investigated cases. The proposed framework achieved faster convergence, lower parameter estimation errors, and improved robustness with respect to different initial conditions compared to conventional Physics-Informed Neural Networks (PINNs).

This study presents a comparative evaluation of PINNs and PI-LSTM for parameter estimation of a PEM electrolyzer equivalent circuit model. The primary objective is to assess the parameter estimation performance of both approaches. In addition, the proposed PI-LSTM framework is evaluated in terms of its robustness to noisy measurement data and its ability to generalize to unseen current profiles, providing a comprehensive assessment of its applicability to dynamic PEM electrolyzer modeling. We do not claim that PINN based methods for parameter identification are superior to standard methods. The strength of the proposed framework, besides being able to identify the parameters with reasonable accuracy, lies in the fact that, at the same time, trains the neural network as an effective and fast surrogate for the system, which can be deployed in real time control applications where computational speed is of paramount importance.

The paper is organized as follows. Section 2 describes the equivalent circuit model and its physical interpretation. Section 3 presents the PINN framework and the different network architectures that we use, and Section 4 details the application to the ECM model of the PEM electrolizer. Section 5 explains some points about the training and presents an study about the sensibility of the results with respect to the size of the neural network. Section 6 contains the results comparing the PINN and PI-LSTM variants, and the generalization capacity of the later, as well as the effect of noise. We state our conclusions in 7. Finally, Appendix A describes the metrics used to evaluate the performance of the model.

2. Model

To represent the dynamic behavior of the PEM electrolyzer, this work employs the second-order equivalent circuit model presented in study of [16], illustrated in Figure 1. This model has demonstrated good capability in reproducing the transient voltage behavior of PEM electrolyzers under time-dependent current inputs, and it is therefore adopted in the present study.

The equivalent circuit is composed of a reversible voltage source E_{rev}, an ohmic resistance R_{ohm}, and two parallel RC branches, namely (R_1, C_1) and (R_2, C_2). The ohmic resistance accounts for voltage losses associated with the membrane and electrical contacts, while the two RC branches describe dynamic processes occurring at different time scales inside the PEM electrolyzer. In this formulation, the state variables are the voltages across the two capacitive branches, denoted by $V_1(t)$ and $V_2(t)$. The input to the model is the applied current $I(t)$, and the output is the cell terminal voltage $V(t)$. The circuit parameters are taken directly from the literature [16] and are assumed to remain constant throughout this study.

The terminal voltage is given by

$$V = E_{\text{rev}} + R_{\text{ohm}} I + V_1 + V_2 \tag{1}$$

The instantaneous electrical power delivered to the electrolyzer is expressed as $I(t)V(t)$. Nevertheless, due to the different loss mechanisms present in the system, not all of this supplied power is effectively converted into useful power for hydrogen generation. The numerical values of the equivalent circuit parameters are adopted from the identification study presented in [16] and are treated as fixed known constants in the simulations carried out in this work. The parameter values used are listed in the table 1.

The existence of two distinct time constants highlights the multi-time-scale nature of the electrolyzer dynamics. The branch associated with the larger time constant captures slower polarization phenomena, whereas the branch with the smaller time constant represents faster transient effects. This separation of dynamics allows the model to reproduce the voltage evolution more accurately under different operating scenarios, especially when the applied current changes rapidly.

3. Methods

3.1. Physics Informed Neural Networks (PINNs) For Inverse Problem

Physics-Informed Neural Networks (PINNs) provide a learning framework in which neural networks are trained by combining available measurement data with the governing differential equations of a physical system. In forward problems, the physical parameters of the model are assumed to be known, and the neural network is mainly used to approximate the solution of the governing equations. However, in inverse problems, some model parameters are unknown and must be identified from the available data. Therefore, the PINN framework can be extended by treating unknown physical parameters as trainable variables that are optimized together with the weights and biases of the neural network [12]. Figure 2 shows a schema of the PINN framework for inverse problems in the control system considered.

We are considering a control system given by:

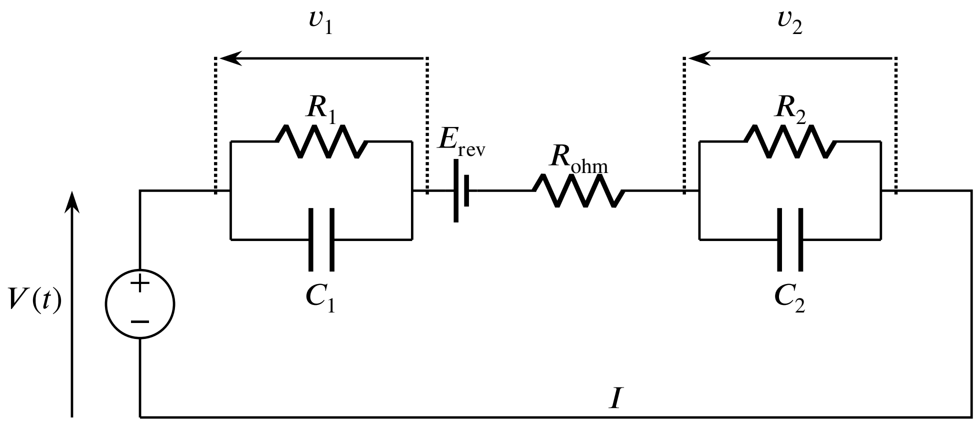

Figure 1: Second-order equivalent circuit model of the PEM electrolyzer.

Table 1
Equivalent circuit model parameters of the PEM electrolyzer, taken from [16].

Parameter	Value	Unit
Reversible voltage, E_{rev}	1.24	V
Ohmic resistance, R_{ohm}	0.002275	Ω
Resistance of RC branch 1, R_1	0.001488	Ω
Capacitance of RC branch 1, C_1	407.4	F
Resistance of RC branch 2, R_2	0.001319	Ω
Capacitance of RC branch 2, C_2	45.94	F
Time constant of RC branch 1, $\tau_1 = R_1C_1$	0.6062112	s
Time constant of RC branch 2, $\tau_2 = R_2C_2$	0.06059486	s

$$\frac{dy}{dt} = f(y, u(t), \phi) \tag{2}$$

where $u(t)$ is the control input and ϕ are the parameters of the system.

The output of the neural network y_θ, which depends on the parameters θ of the neural network, is identified as a surrogate of the state of the system by using it in a loss function that computes to what extent y_θ satisfies the differential equation of the control system.

To enforce the governing dynamics in discrete time, the physics residual is computed from the finite-difference approximation of the state evolution between two consecutive time steps. The physics loss is then obtained by minimizing the squared norm of these residuals over all time values indexed by k:

$$\mathrm{RES}_{\mathrm{phys},k} = \frac{\mathbf{y}_{\theta,k+1} - \mathbf{y}_{\theta,k}}{\Delta t} - f\left(\mathbf{y}_{\theta,k}, u_k, \boldsymbol{\phi}\right). \tag{3}$$

The PINN training objective is to minimize a total loss function composed of several contributions, including the physics residual loss, the data loss. The physics loss is formulated as :

$$\mathcal{L}_{\mathrm{physics}} = \frac{1}{N} \sum_{k=0}^{N-1} \left\| \mathrm{RES}_{\mathrm{phys},k} \right\|^2, \tag{4}$$

where N is the number of points where the differential equation is evaluated.

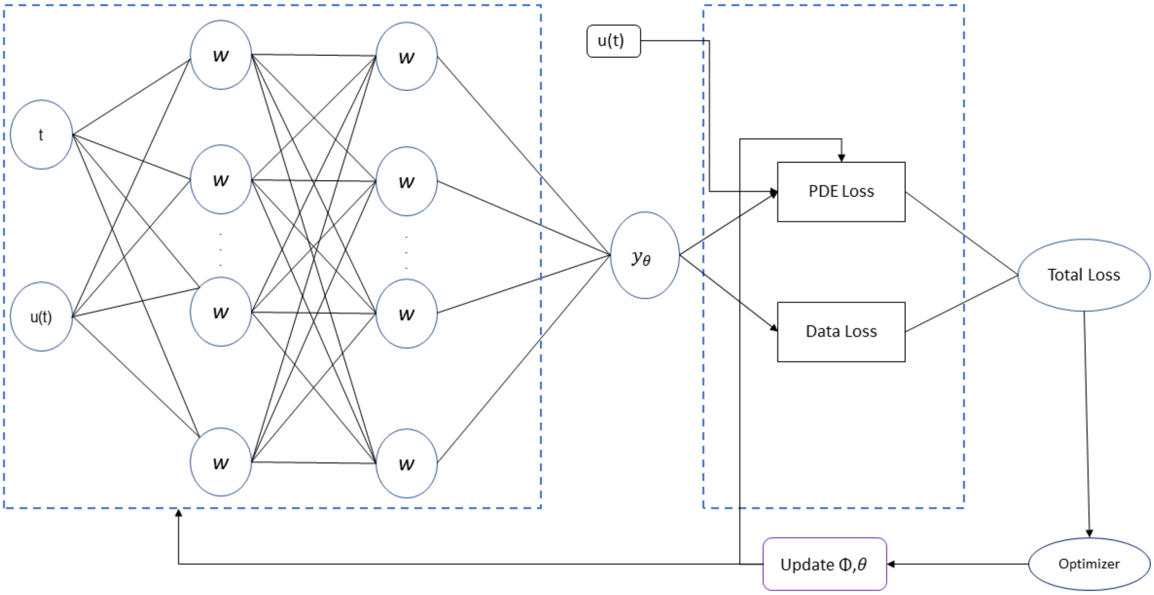

Figure 2: This figure shows the architecture of the Physics-Informed Neural Network (PINN) for inverse problem in control system. The neural network receives time t and input signal $u(t)$ as inputs and predicts the output $\mathbf{y}_\theta$. The predicted output is used to compute both the physics-based loss and the data loss, which are combined into the total loss function. The optimizer updates the neural network weights (θ) and the physical parameters (ϕ) iteratively during training.

When measurement data are available, the discrepancy between the predicted outputs and the observed measurements at each discrete time is minimized through the data loss term:

$$\mathcal{L}_{\text{data}} = \frac{1}{N}\sum_{k=0}^{N-1} \left\| \mathbf{y}_{\theta,k} - \mathbf{y}_{\text{data},k} \right\|^2, \tag{5}$$

where N describes the number of available data samples.

In this study, the loss of the initial condition is not considered because the data points at $k = 0$ are included in the data loss term.

The total loss function used to train the PINN is defined as the weighted sum of the physics loss and the data loss:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{phys}}\mathcal{L}_{\text{physics}} + \lambda_{\text{data}}\mathcal{L}_{\text{data}}, \tag{6}$$

where λ_{phys} and λ_{data} are weighting coefficients that balance the contribution of physics-based constraints and measurement data during training.

In this inverse problem, the six parameters are initialized as trainable variables. During training, the neural network predicts the latent voltages V_1 and V_2, while physical parameters are simultaneously updated through backpropagation by minimizing both the physical and data losses. Unlike the forward problem, the governing equations are not solved using known parameters; instead, the parameters themselves are estimated from the measured voltage and current data.

Although the PINN framework is independent of the underlying neural network architecture, its performance depends on the choice of the function approximator. In this study, two different neural network architectures are investigated within the physics-informed framework: a Feedforward Neural Network (FFNN) and a Long Short-Term Memory (LSTM) network. The FFNN serves as the conventional baseline architecture commonly employed in PINNs, whereas the LSTM is adopted to better capture temporal dependencies in sequential data. The background of these two architectures are briefly reviewed in the following subsections.

3.1.1. Feed Forward Neural Network

A Feed Forward Neural Network (FFNN) is one of the most fundamental architectures in artificial neural networks. In an FFNN, information propagates in a single direction from the input layer through one or more hidden layers to the output layer without any feedback or recurrent connections. Each neuron performs a weighted summation of its inputs followed by a nonlinear activation function, enabling the network to approximate complex nonlinear mappings between the input and output variables [9, 6].

For a network with (L) layers, the output of the (I)th hidden layer is computed as

$$\sigma\left(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}\right), \qquad l = 1, 2, \ldots, L-1, \tag{7}$$

where $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ denote the weight matrix and bias vector of the l-th layer, respectively, $\sigma(\cdot)$ represents a nonlinear activation function, and $\mathbf{h}^{(0)} = \mathbf{x}$ corresponds to the input vector.

The final network output is computed as

$$\mathbf{W}^{(L)}\mathbf{h}^{(L-1)} + \mathbf{b}^{(L)}, \tag{8}$$

where ($\mathbf{y}_\theta$) denotes the predicted output of the network. The trainable parameters of the FFNN are defined as

$$\Theta = \left\{\mathbf{W}^{(l)}, \mathbf{b}^{(l)}\right\}_{l=1}^{L} \tag{9}$$

During the training, these parameters are optimized by minimizing a loss function using gradient-based optimization algorithms such as stochastic gradient descent(SGD) or Adam.

3.1.2. Long Short-Term Memory Neural Network

Long Short-Term Memory (LSTM) networks are a specialized class of Recurrent Neural Networks (RNNs) designed to capture long-term temporal dependencies in sequential data. Conventional RNNs often suffer from vanishing and exploding gradient problems during training, limiting their ability to learn relationships over long sequences. LSTMs address these issues by introducing a memory cell and a set of gating mechanisms that regulate the flow of information through time [2, 11]. As illustrated in Fig. 3, an LSTM maintains two internal states: the hidden state $\mathbf{h}_t$, which represents the current output, and the cell state $\mathbf{C}_t$, which serves as a long-term memory.

LSTM Cell Architecture: At each time step t, the LSTM receives the current input vector $\mathbf{x}_t \in \mathbb{R}^d$ together with the previous hidden state $\mathbf{h}_{t-1} \in \mathbb{R}^h$ and the previous cell state $\mathbf{C}_{t-1} \in \mathbb{R}^h$, where d denotes the input dimension and h is the number of hidden units. The internal memory is updated through three gating mechanisms: the forget gate, the input gate, and the output gate.

Forget Gate: The forget gate determines which information from the previous cell state should be retained or discarded. It is defined as

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_f\mathbf{x}_t + \mathbf{U}_f\mathbf{h}_{t-1} + \mathbf{b}_f\right), \tag{10}$$

where $\sigma(\cdot)$ denotes the sigmoid activation function, and $\mathbf{W}_f \in \mathbb{R}^{h\times d}$, $\mathbf{U}_f \in \mathbb{R}^{h\times h}$, and $\mathbf{b}_f \in \mathbb{R}^h$ are learnable parameters. Values of $\mathbf{f}_t$ close to one preserve information, whereas values close to zero remove it from the memory.

Input Gate: The input gate controls how much new information is incorporated into the cell state. It operates together with a candidate memory vector computed as

$$\mathbf{i}_t = \sigma\left(\mathbf{W}_i\mathbf{x}_t + \mathbf{U}_i\mathbf{h}_{t-1} + \mathbf{b}_i\right), \tag{11}$$

$$\tilde{\mathbf{C}}_t = \tanh\left(\mathbf{W}_c\mathbf{x}_t + \mathbf{U}_c\mathbf{h}_{t-1} + \mathbf{b}_c\right), \tag{12}$$

where $\mathbf{W}_i$, $\mathbf{W}_c \in \mathbb{R}^{h\times d}$, $\mathbf{U}_i$, $\mathbf{U}_c \in \mathbb{R}^{h\times h}$, and $\mathbf{b}_i$, $\mathbf{b}_c \in \mathbb{R}^h$ are trainable parameters.

Cell State Update: The new cell state is obtained by combining the retained information from the previous memory with the newly generated candidate information:

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t, \tag{13}$$

where $\odot$ denotes element-wise multiplication. This additive update mechanism enables efficient propagation of information across long sequences and significantly alleviates the vanishing gradient problem.

Output Gate: The output gate determines which components of the updated cell state are exposed as the hidden state. It is computed as

$$\mathbf{o}_t = \sigma\left(\mathbf{W}_o\mathbf{x}_t + \mathbf{U}_o\mathbf{h}_{t-1} + \mathbf{b}_o\right), \tag{14}$$

followed by the hidden state update

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{C}_t). \tag{15}$$

Here, $\mathbf{W}_o \in \mathbb{R}^{h\times d}$, $\mathbf{U}_o \in \mathbb{R}^{h\times h}$, and $\mathbf{b}_o \in \mathbb{R}^{h}$ are learnable parameters. The hyperbolic tangent function maps the cell state to the interval $(-1, 1)$ before it is filtered by the output gate to produce the hidden representation [5].

Gradient Flow and Long-Term Dependency Learning: A key advantage of LSTMs over conventional RNNs is the preservation of gradients through the cell state during backpropagation through time (BPTT). Specifically, the gradient with respect to an earlier cell state can be expressed as

$$\frac{\partial \mathcal{L}}{\partial \mathbf{C}_{t-k}} = \frac{\partial \mathcal{L}}{\partial \mathbf{C}_t} \prod_{j=1}^{k} \mathbf{f}_{t-j+1}, \tag{16}$$

where $\mathcal{L}$ denotes the training loss. Since the gradient propagates primarily through element-wise multiplication by the forget gate values, information can be preserved over many time steps when $\mathbf{f}_t$ remains close to one. This mechanism enables LSTMs to effectively model long-range temporal dependencies [4].

Learnable Parameters: For an LSTM layer with input dimension d and hidden dimension h, the total number of learnable parameters is

$$\theta_{\text{LSTM}} = 4\left(hd + h^2 + h\right) = 4h(d + h + 1), \tag{17}$$

where the factor of four accounts for the parameter sets associated with the forget gate, input gate, candidate cell state, and output gate.

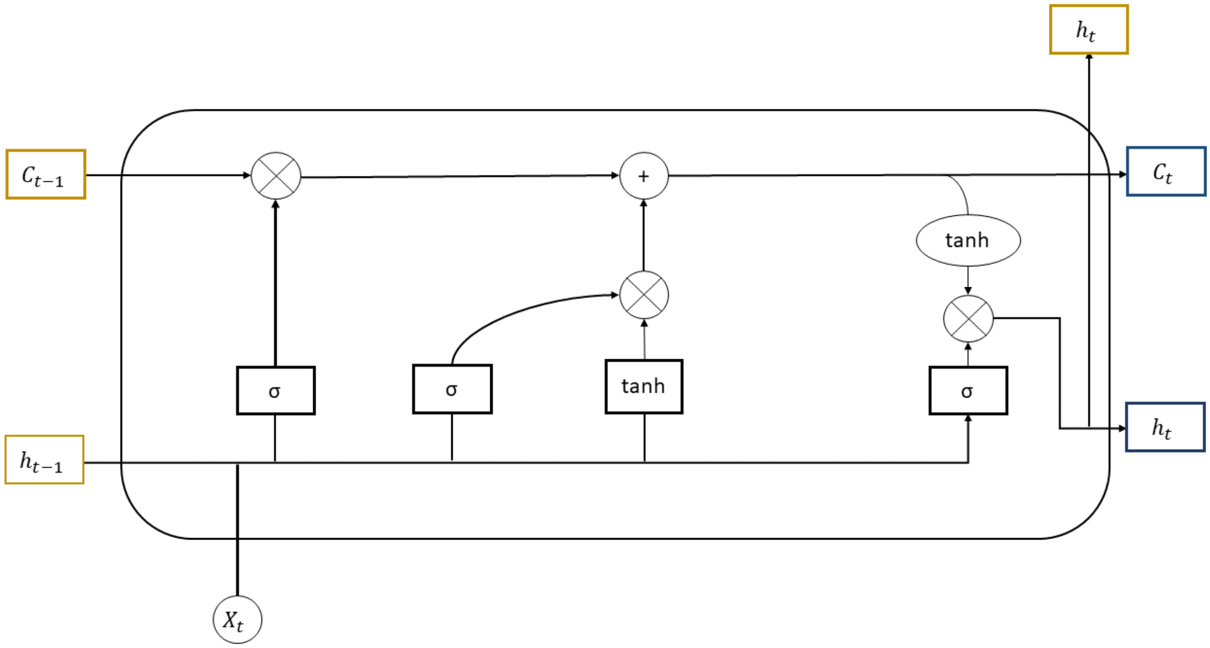

Figure 3: Architecture of a standard Long Short-Term Memory (LSTM) cell showing the forget, input, and output gates, candidate cell state generation, cell state update, and hidden state computation.

4. Application of PINN for Equivalent Electrical Circuit

The PEM electrolyzer is represented by a second-order equivalent electrical circuit (ECM) with two RC branches, an ohmic resistance, and a reversible voltage source represented in the section 2. The terminal voltage is

$$V(t) = E_{\text{rev}} + R_{\text{ohm}}\, I(t) + V_1(t) + V_2(t), \tag{18}$$

where $I(t)$ is the known applied current and $V_1(t)$, $V_2(t)$ are the voltages across the two RC branches. The branch dynamics are

$$\dot{V}_1 = \frac{I(t) - V_1/R_1}{C_1}, \tag{19}$$

$$\dot{V}_2 = \frac{I(t) - V_2/R_2}{C_2}, \tag{20}$$

with initial conditions $V_1(0) = V_{10}$ and $V_2(0) = V_{20}$. Introducing the time constants $\tau_i = R_i C_i$, (19)–(20) can be written in the first-order form

$$\tau_i \dot{V}_i + V_i = R_i I(t), \qquad i \in \{1, 2\}. \tag{21}$$

The inverse problem is to estimate

$$\boldsymbol{\phi} = \left(E_{\text{rev}},\, R_{\text{ohm}},\, R_1,\, C_1,\, R_2,\, C_2\right) \tag{22}$$

from measured terminal voltages $\{V_k^{\text{meas}}\}_{k=0}^{N-1}$ on a known time grid $\{t_k\}_{k=0}^{N-1}$ under one or more current profiles $\{I_k\}_{k=0}^{N-1}$. The internal branch voltages are not measured in the realistic identification setting.

4.1. Neural surrogate for latent branch states

Inverse PINN uses a shared point-wise Neural Network (NN) that maps normalized time and current to the latent states at each sample:

$$\mathcal{N}_\theta : (\tilde{t}_k, \tilde{I}_k) \mapsto \left(\hat{V}_{1,k}, \hat{V}_{2,k}\right), \qquad k = 0, \dots, N-1, \tag{23}$$

with

$$\tilde{t}_k = 2\,\frac{t_k}{T} - 1, \tag{24}$$

$$\tilde{I}_k = 2\,\frac{I_k - I_{\min}}{I_{\max} - I_{\min}} - 1. \tag{25}$$

The network has architecture [2, 128, 128, 2] with tanh activations in the hidden layers. The outputs are bounded by

$$\hat{V}_{i,k} = s_i \tanh\!\left(z_{i,k}\right), \qquad i \in \{1, 2\}, \tag{26}$$

where s_1 and s_2 are fixed state scales. This keeps the latent voltages in a physically plausible range during training.

The predicted terminal voltage is then assembled from the trainable ECM parameters and the NN states:

$$\hat{V}_k = \hat{E}_{\mathrm{rev}} + \hat{R}_{\mathrm{ohm}}\, I_k + \hat{V}_{1,k} + \hat{V}_{2,k}. \tag{27}$$

4.2. Trainable physical parameters with identifiability constraints

All ECM parameters are learned in log-space to enforce positivity:

$$\hat{E}_{\mathrm{rev}} = e^{\xi_E}, \quad \hat{R}_{\mathrm{ohm}} = e^{\xi_{R_{\mathrm{ohm}}}}, \quad \hat{R}_1 = e^{\xi_{R_1}}, \quad \hat{R}_2 = e^{\xi_{R_2}}, \tag{28}$$

where $\{\xi_j\}$ are unconstrained learnable scalars. To avoid branch swapping between the two RC pairs, the time constants are parameterized as

$$\hat{\tau}_2 = e^{\xi_{\tau_2}}, \qquad \hat{\tau}_1 = \hat{\tau}_2 + \Delta\tau_{\min} + \mathrm{softplus}(\xi_{\Delta\tau}), \tag{29}$$

with a minimum gap $\Delta\tau_{\min} > 0$, and the capacitances are recovered as $\hat{C}_i = \hat{\tau}_i / \hat{R}_i$. This enforces $\hat{\tau}_1 > \hat{\tau}_2$ and improves identifiability of the slow and fast branches.

The same physical parameters $\boldsymbol{\phi}$ are shared across all current-profile experiments. Only the measured voltage trajectories differ between experiments.

4.3. Discrete-time physics residuals

This method enforces the ECM dynamics through *discrete-time residuals* rather than automatic-differentiation time derivatives. For each RC branch $i \in \{1, 2\}$ and each interval $[t_k, t_{k+1}]$ with $\Delta t_k = t_{k+1} - t_k$, the default midpoint residual is

$$r_{i,k} = \hat{\tau}_i \frac{\hat{V}_{i,k+1} - \hat{V}_{i,k}}{\Delta t_k} - \hat{R}_i I_{k+1/2} + \hat{V}_{i,k+1/2}, \tag{30}$$

where

$$\hat{V}_{i,k+1/2} = \tfrac{1}{2}(\hat{V}_{i,k+1} + \hat{V}_{i,k}), \qquad I_{k+1/2} = \tfrac{1}{2}(I_{k+1} + I_k). \tag{31}$$

The reason for selecting a finite-differences implementation of the time derivatives over automatic differentiation is that automatic differentiation provides exact derivatives of the learned surrogate, but these derivatives are not necessarily accurate approximations of the derivatives of the true physical solution. In inverse problems, errors in these derivative terms can directly bias the estimated physical parameters. Finite-difference derivatives, on the other hand, approximate the temporal or spatial variation of the predicted solution over neighboring grid points. This couples adjacent predictions and enforces the governing equation over the discrete trajectory rather than only through the local differential properties of the neural network. This is particularly suitable for recurrent architectures such as LSTMs, which naturally model the solution as a sequence of states. Therefore, the finite-difference residual is better aligned with the discrete-time representation learned by the LSTM but this should not be interpreted as finite differences being generally more accurate than automatic differentiation.

4.4. Loss function

For M current-profile experiments indexed by $m = 1, \ldots, M$, the total loss is

$$\mathcal{L} = \lambda_{\text{data}}\,\mathcal{L}_{\text{data}} + \lambda_{\text{phys}}\,\mathcal{L}_{\text{phys}} + \lambda_{\text{ic}}\,\mathcal{L}_{\text{ic}} + \lambda_{\text{jump}}\,\mathcal{L}_{\text{jump}}, \tag{32}$$

The data term fits the measured terminal voltage:

$$\mathcal{L}_{\text{data}} = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{N_m}\sum_{k=0}^{N_m-1}\left(\hat{V}_k^{(m)} - V_k^{\text{meas},(m)}\right)^2. \tag{33}$$

The physics term penalizes the discrete ECM residuals, normalized by the state scales s_i:

$$\mathcal{L}_{\text{phys}} = \frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{N_m-1}\sum_{k=0}^{N_m-2}\left[\left(\frac{r_{1,k}^{(m)}}{s_1}\right)^2 + \left(\frac{r_{2,k}^{(m)}}{s_2}\right)^2\right] + \frac{1}{N_m}\sum_{k=0}^{N_m-1}\left(r_{V,k}^{(m)}\right)^2.\right] \tag{34}$$

The initial-condition term anchors the latent states at t_0:

$$\mathcal{L}_{\text{ic}} = \frac{1}{M}\sum_{m=1}^{M}\left[\left(\frac{\hat{V}_{1,0}^{(m)} - V_{10}^{(m)}}{s_1}\right)^2 + \left(\frac{\hat{V}_{2,0}^{(m)} - V_{20}^{(m)}}{s_2}\right)^2\right]. \tag{35}$$

An optional ohmic-resistance hint is obtained from large current jumps via $\hat{R}_{\text{ohm}}^{\text{hint}} \approx \text{median}(\Delta V/\Delta I)$, and penalized by

$$\mathcal{L}_{\text{jump}} = \left(\frac{\hat{R}_{\text{ohm}} - \hat{R}_{\text{ohm}}^{\text{hint}}}{\hat{R}_{\text{ohm}}^{\text{hint}}}\right)^2. \tag{36}$$

This helps identify R_{ohm} from sharp transients. The weight values are used in this study for training are, $\lambda_{\text{data}} = 15$, $\lambda_{\text{phys}} = 1$, $\lambda_{\text{ic}} = 50$, and $\lambda_{\text{jump}} = 2$.

4.5. Multi-profile training and optimization

Training uses five diverse current profiles over $t \in [0, T]$ with sampling step $\Delta t = 5 \times 10^{-3}$ s: composite, fast PRBS, slow PRBS, multisine, and steps/ramps. Each profile provides rich transient content:

- zero-current rest segments to anchor E_{rev} and initial states;
- sharp current jumps to identify R_{ohm};
- dwell times shorter than, comparable to, and longer than both τ_1 and τ_2;
- frequency content around the nominal corner frequencies.

Optimization proceeds in two stages on full trajectories:

1. **Adam** Adam was employed with a learning rate of 10^{-3} for 2501 epochs. During each iteration, the total loss function, consisting of data, physics, initial-condition, and parameter regularization terms, was minimized through backpropagation. To improve numerical stability and prevent exploding gradients, gradient clipping with a maximum norm of 100 was applied.
2. **L-BFGS** L-BFGS was initialized using the parameters obtained from the Adam phase. The optimizer was configured with a learning rate of 0.1, a history size of 50, and the strong wolf line-search strategy. During each iteration, a closure function was used to repeatedly evaluate the PINN loss and its gradients as required by the L-BFGS algorithm.

If L-BFGS produces non-physical or numerically unstable parameters, the solution reverts to the post-Adam checkpoint.

Table 2
Training and generalization excitation current profiles used for parameter identification of the PEM electrolyzer equivalent circuit model.

Profile	Main Characteristics	Identification Objective
Composite	Mixed excitation	Broad-band excitation
Fast PRBS	Hold time = 75 ms	Excitation of fast dynamics
Slow PRBS	Hold time = 500 ms	Excitation of slow dynamics
Multi-sine	Frequency range: 0.15–8 Hz	Frequency-domain excitation
Steps/Ramps	Structured transient excitation	Steady-state and transient response analysis
Chirp	Frequency sweep: 0.05–10 Hz	testing and generalization assessment

5. Training and Sensitivity Analysis

The excitation signals used for training and testing are shown in Fig 4 and summarized in Table 2. The training dataset consists of a composite profile, fast and slow PRBS excitations, a multi-sine signal, and a step/ramp profile. These signals were selected to excite the PEM electrolyzer dynamics over a broad range of frequencies and operating conditions. In particular, the fast and slow PRBS signals target the short-term and long-term dynamic modes of the equivalent circuit, respectively, while the multi-sine excitation provides broadband frequency content. The step/ramp profile evaluates transient and quasi-steady-state responses, and the composite profile combines multiple excitation patterns within a single experiment.

Figure 6 illustrates the evolution of the different loss components during the training of the proposed PI-LSTM framework. The total loss exhibits a steady decrease throughout the optimization process, indicating stable convergence of the training procedure. Similarly, the data loss and physics loss decrease consistently as the network learns to simultaneously fit the measured voltage data while satisfying the governing physical equations. The initial-condition and ohmic-resistance regularization terms also converge to very small values, demonstrating that the estimated states and parameters remain physically consistent during optimization.

For testing, a chirp signal with a continuously increasing frequency as shown in 8 (a) was used. Since this profile was not included during training, it provides an independent assessment of the generalization capability of the identified model.

Sensitivity Analysis: To determine the an appropriate architecture for the PI-LSTM model, a sensitivity analysis was performed by varying the number of hidden layers (2,3 and 4) and the hidden size (16,32,64 and 128 neurons). To account for the stochastic nature of the neural network training, each architecture was trained independently 10 times using different random initializations, and the mean relative parameter estimation error over the 10 runs was used as the evaluation metric. As shown in Fig. 5, increasing the hidden size generally improved identification accuracy, whereas increasing the number of hidden layers did not consistently reduce error. Among all of these tests configurations, the architecture with 2 hidden layers and 128 neurons demonstrated the lowest mean relative estimation error (approximately 1.07%). Consequently, this architectural configuration was selected for all subsequent experiments.

6. Results

This section presents a comprehensive comparison between the Physics-Informed Neural Network and the Physics-Informed Long Short-Term Memory (PI-LSTM) approaches for the parameter identification and voltage prediction of the PEM electrolyzer equivalent circuit model. In the following sections, the performance of both PINN and PI-LSTM will be analyzed for generalization and robustness to noise to evaluate their applicability in realistic scenarios. The evaluation metrics used in this study are presented in the Appendix.

6.1. Parameter Identification

The primary objective of this study is to estimate the unknown parameters of the PEM electrolyzer equivalent circuit shown in the section 2, using measured voltage and current data. To evaluate the performance of the proposed approaches, identified parameters by PINN and PI-LSTM are compared with the reference values used to generate the

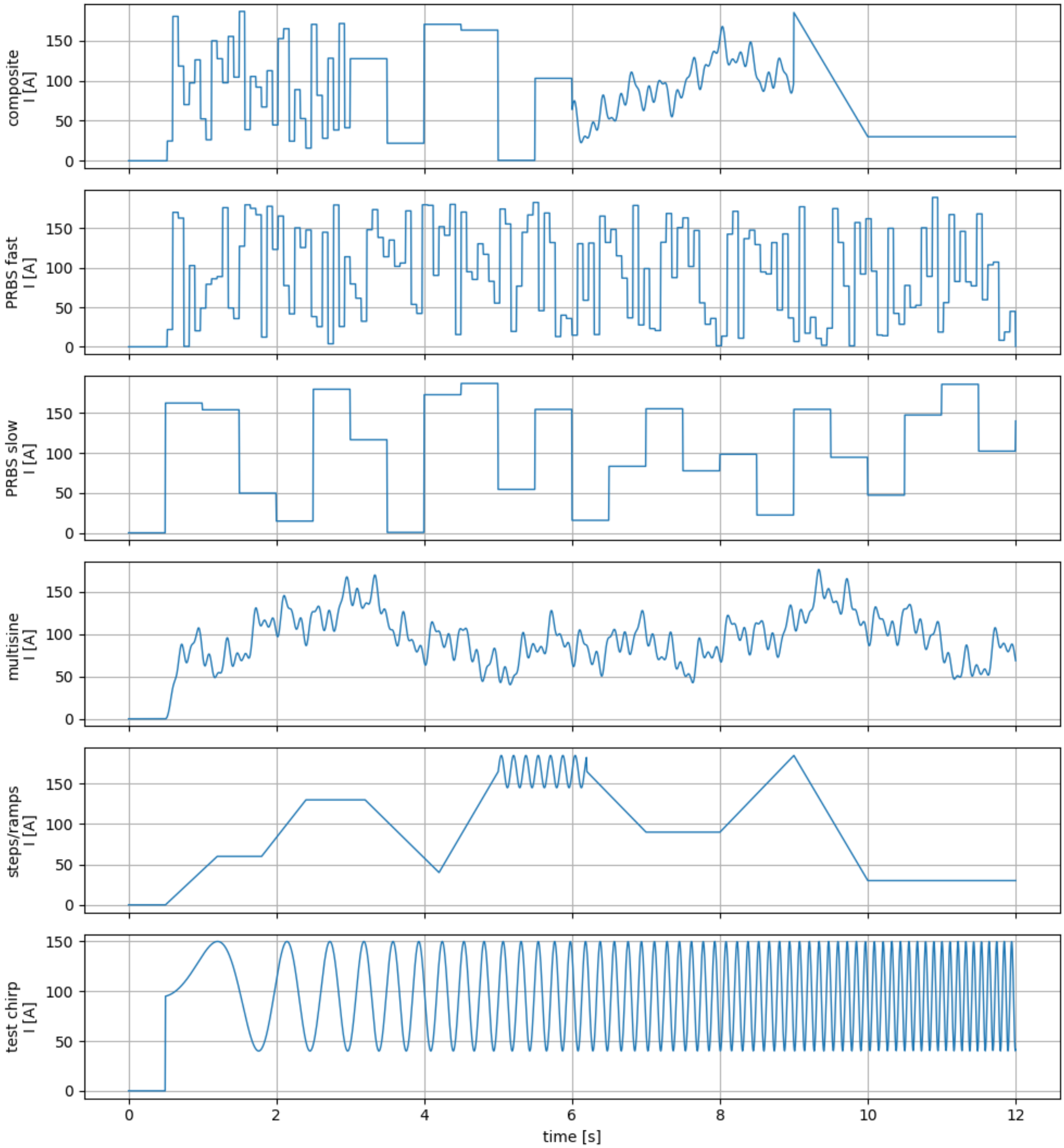

Figure 4: Excitation current profiles used for training and generalization of the PINN and PI-LSTM parameter identification. The training dataset consists of a composite profile, fast PRBS, slow PRBS, multi-sine, and step/ramp excitations, while the chirp signal is reserved for generalization assessment.

simulation data. Table 3 summarizes the true parameter values, the estimated values obtained by each method, and the corresponding relative errors.

Table 3 presents the parameter identification results obtained using PINN and PI-LSTM. The PI-LSTM model accurately estimates the parameters with an error below 5% for every parameter considered. In contrast, PINN could

Figure 5: The architecture with two hidden layers and 128 neurons was chosen because it provided the best trade-off between model complexity and parameter identification accuracy, achieving the lowest average relative error among all tested configurations.

Table 3
Parameter identification results obtained using PINN and PI-LSTM.

Parameter	True Value	PINN	Error (%)	PI-LSTM	Error (%)
E_{rev}	1.2400	1.326411	6.969	1.239155	-0.068
R_{ohm}	0.002275	0.002297732	0.999	0.002279825	0.212
R_1	0.001488	0.0001564422	-89.486	0.001481260	-0.453
C_1	407.4	4549.286	1016.663	424.1094	4.101
R_2	0.001319	0.001449492	9.893	0.001329944	0.830
C_2	45.94	3.200101	-93.034	47.46867	3.328
τ_1	0.6062112	0.7117003	17.401	0.6282165	3.630
τ_2	0.06059486	0.004638519	-92.345	0.06313069	4.185

successfully identified the reversible voltage and ohmic resistance; considerable errors were observed for the RC-network parameters. Estimated RC branches parameters of R_1, C_1, C_2 deviated substantially from their true values, that results for a large errors in the associated time constants.

These results show that PI-LSTM outperforms the PINN in the parameter identification performance for the considered PEM electrolyzer equivalent circuit model. The recurrent structure of PI-LSTM enables the model to capture temporal dependencies more effectively, leading to a more accurate estimation of the dynamic parameters that govern the transient behavior of the system, while the PINN model showed challenges in parameter identifiability. That R_{ohm} and, to a lesser extent, E_{rev} display a relatively small error for PINN can be explained by the fact that they appear directly in the total voltage, without being affected by the values of the hidden states V_1 and V_2.

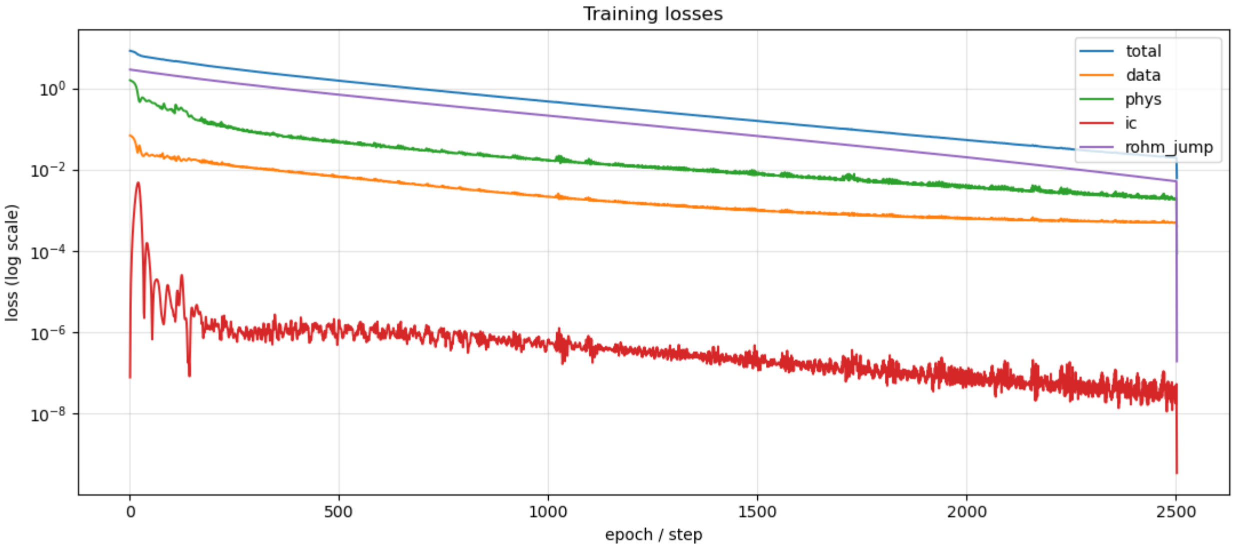

Figure 6: Evolution of the different loss components during PI-LSTM training. The total loss, data loss, physics loss, initial-condition loss, and ohmic-resistance regularization term decrease progressively throughout optimization, demonstrating stable convergence and effective learning of both the measurement data and the underlying physical constraints.

6.2. Voltage Prediction

The voltage prediction performance of the PI-LSTM model was evaluated using different excitation current profiles, namely composite, PRBS-fast, PRBS-slow, multi-sine, and step/ramp current signals. Figure 7 compares the predicted voltage with the reference voltage generated by the equivalent circuit to the PEM electrolyzer.

As shown in Fig. 7, the predicted voltage closely follows the reference response for all excitation profiles. The agreement remains excellent during both steady-state and transient operating conditions, demonstrating the capability of the PEM electrolyzer. In particular, the model accurately reproduces the rapid voltage variations observed under the PRBS and composite current profiles, which contain frequent changes in operating conditions.

Generalization To evaluate the generalization capacity of the proposed PI-LSTM model, an unseen current profile shown in the fig 8 (a) employed during the testing. Test profile consists of a sinusoidal current with continuously increasing frequency, providing a challenging scenario for assessing the model's ability to extrapolate to new operating conditions.

Figure 8 compares the voltage response generated using the identified parameters shown in Table 3, with the reference voltage obtained from the true model parameters. The results show excellent agreement between the predicted and reference voltage trajectories throughout the entire simulation horizon. Despite the progressively increasing excitation frequency, the model successfully captures both the transient and steady-state voltage dynamics.

The prediction remains small during over the entire test interval, as shown in the panel c of Fig. 8. Quantitatively, the model achieved an RMSE of 0.798 mV, an MAE of 0.662 mV, and an R^2 of 0.999969. These results illustrate the PI-LSTM's ability to accurately reproduce the PEM electrolyzer dynamics even under operating conditions not encountered during training. This performance strength demonstrates that proposed PI-LSTM framework does not merely memorize the training data but instead learns the underlying physical behavior of the system.

6.3. Noise Robustness

In practical PEM electrolyzer systems, voltage measurements are inevitably affected by sensor inaccuracies, environmental disturbances, and data acquisition noise. Therefore, it is important to evaluate the robustness of the proposed PI-LSTM framework when the training data are corrupted by measurement noise. To investigate this aspect, zero-mean Gaussian white noise with a standard deviation of 0.05 V was added to the voltage measurements according to

$$V_{\text{meas}} = V_{\text{clean}} + \mathcal{N}(0, \sigma^2) \tag{37}$$

Table 4
Parameter identification results obtained using noisy voltage measurements.

Parameter	True Value	Estimated Value (Error %)
E_{rev}	1.2400	1.240014 (0.001)
R_{ohm}	0.002275	0.002268003 (-0.308)
R_1	0.001488	0.001468457 (-1.313)
C_1	407.4	419.2634 (2.912)
R_2	0.001319	0.001344699 (1.948)
C_2	45.94	47.30492 (2.971)
τ_1	0.6062112	0.6156702 (1.560)
τ_2	0.06059486	0.06361088 (4.977)

where $\sigma = 0.05$ V.

Table 4 shows the identified parameters obtained using the noisy measurements. Despite the existence of measurement noise, the accuracy of the proposed PI-LSTM model enabled the recovery of all circuit parameters. The relative estimation errors remained below 5% for all identified parameters, with errors of only 0.001% for E_{rev}, -0.308% for R_{ohm}, -1.313% for R_1, 2.912% for C_1, 1.948% for R_2, and 2.971% for C_2. Similarly, the estimated time constants τ_1 and τ_2 exhibited errors of 1.560% and 4.977%, respectively. These results show that dynamic characteristic of the pem electrolyzer were successfully identified by PI-LSTM even in the presence of the measurement noise.

Figure 9 illustrates the voltage predictions and reconstructed internal states obtained from the identified noisy-data model for different excitation current profile explained in the section 5.The predicted voltage remains in excellent agreement with the reference voltage across all operating conditions. The overlap between the predicted and reference voltage trajectories demonstrate the ability of the PI-LSTM to captures both steady-state and transient behavior with noisy measurements used during the training. To further evaluate the robustness of the proposed PI-LSTM framework, the model trained with noisy voltage measurements was tested using the unseen chirp current profile introduced in Section 5. Figure 10 shows the corresponding generalization results. Despite the presence of measurement noise during training, the predicted voltage remains in excellent agreement with the reference response throughout the entire simulation interval. To further evaluate the robustness of the proposed PI-LSTM framework, the model trained with noisy voltage measurements was tested using the unseen chirp current profile introduced in Section 5. Figure 10 shows the corresponding testing results. Despite the presence of measurement noise during training, the predicted voltage remains in excellent agreement with the reference response throughout the entire simulation interval.

The model achieved an RMSE of 1.436 mV, an MAE of 1.215 mV, an MSE of 2.06×10^{-6} V^2, and an R^2 score of 0.9999. These results demonstrate that the proposed PI-LSTM maintains strong generalization capability and accurately reproduces the PEM electrolyzer dynamics under previously unseen operating conditions, even when trained with noisy data.

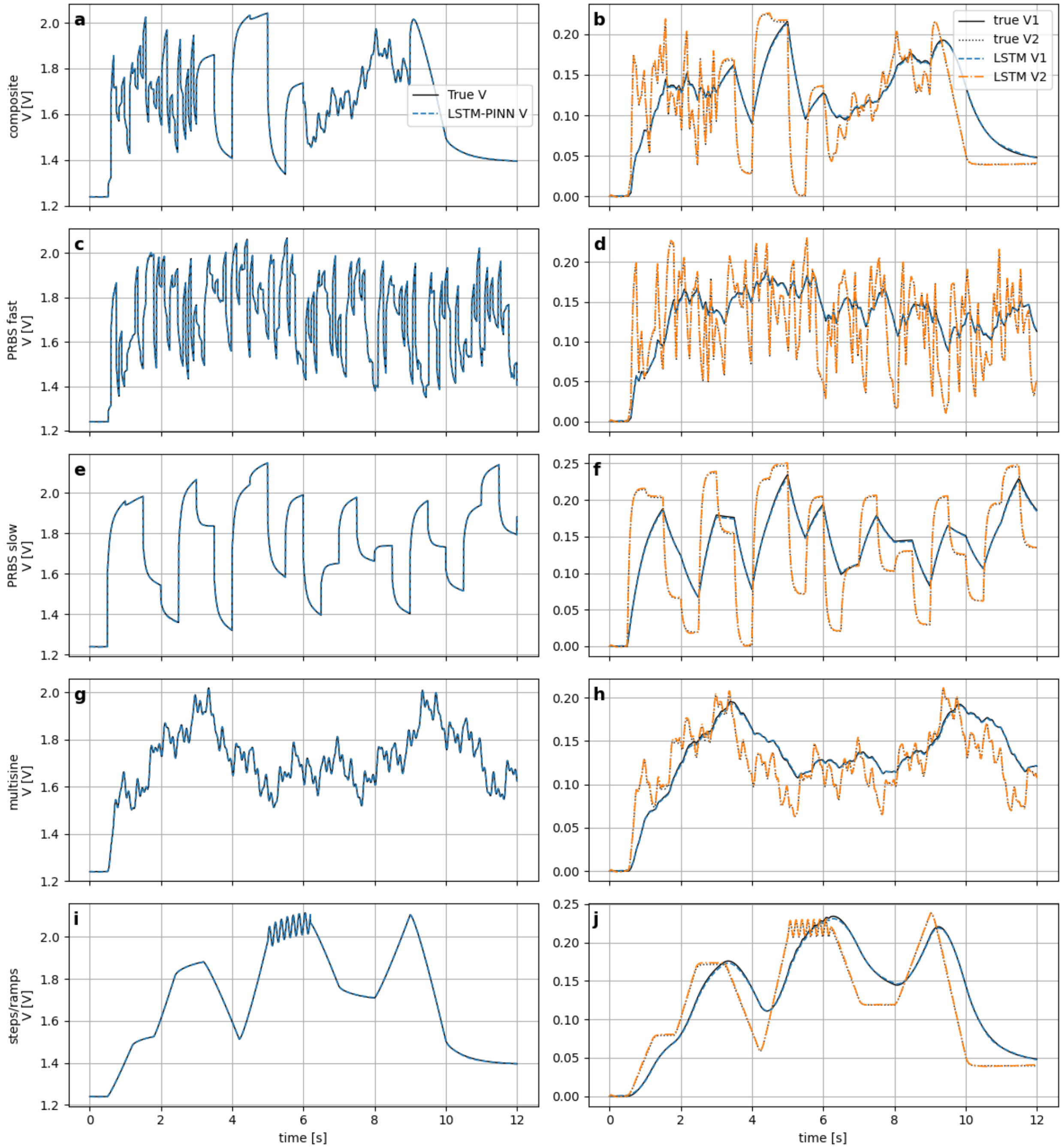

Figure 7: Comparison between the reference and PI-LSTM predictions for different excitation profiles. The left column (a,c,e,g,i) shows the terminal voltage V , while the right column (b,d,f,h,j) presents the reconstructed internal RC branch voltages V_1,V_2 The proposed PI-LSTM accurately reproduces the terminal voltage and captures the underlying state dynamics across all considered operating conditions.

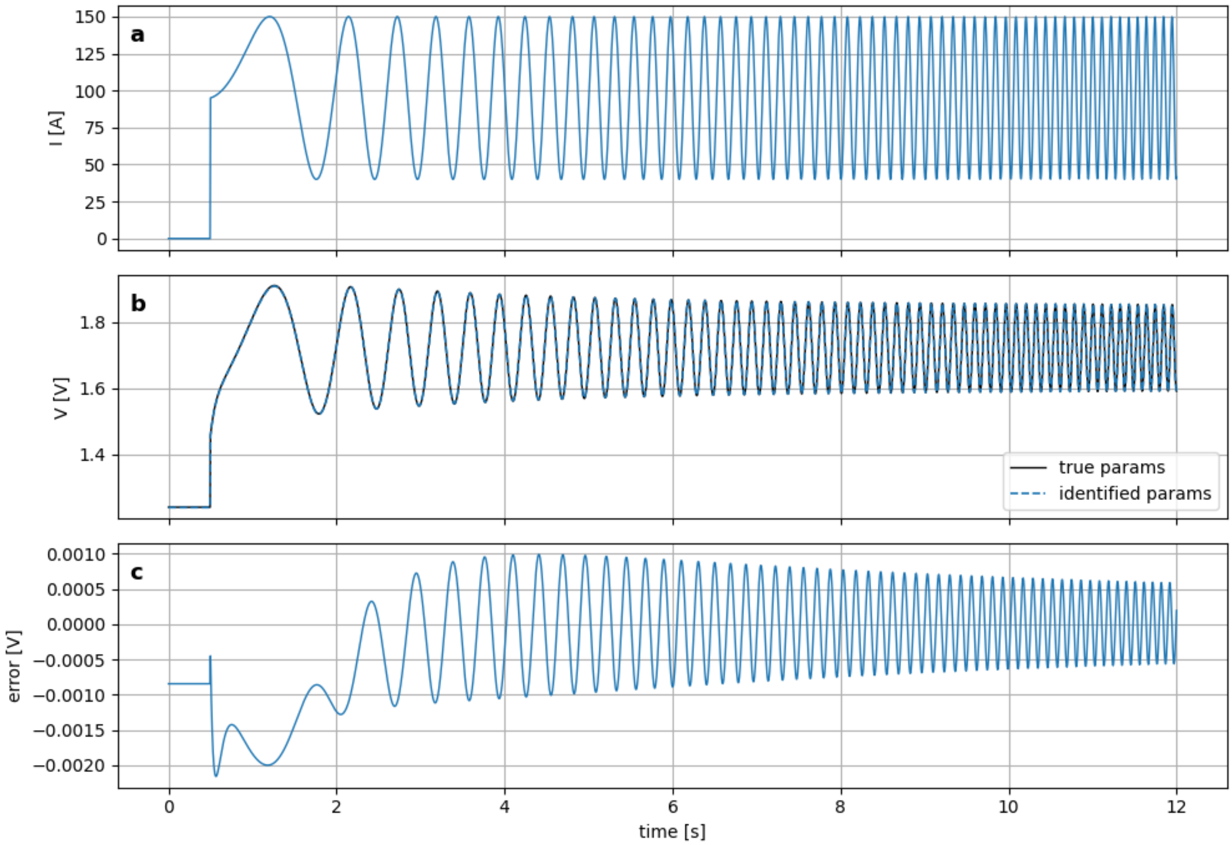

Figure 8: Generalization performance of the proposed PI-LSTM model on an unseen chirp excitation profile. (a) Test current with continuously increasing frequency. (b) Comparison between the reference terminal voltage and the voltage reconstructed using the identified parameters. (c) Voltage prediction error, demonstrating millivolt-level accuracy throughout the test interval.

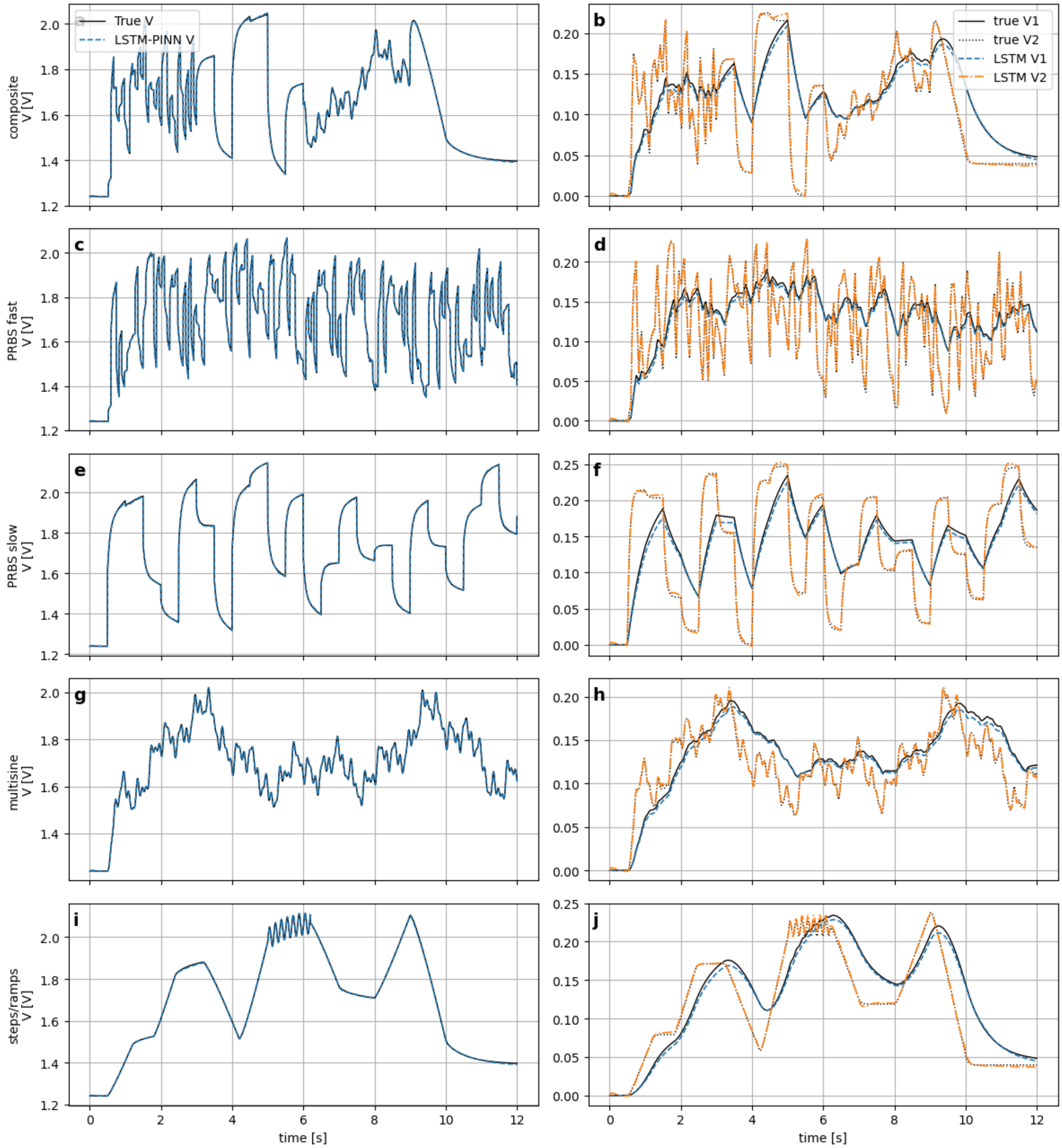

Figure 9: Comparison between the reference trajectories and PI-LSTM predictions obtained after training with noisy voltage measurements. The left column (a, c, e, g, i) shows the terminal voltage responses for the composite, fast PRBS, slow PRBS, multisine, and step/ramp excitation profiles, while the right column (b, d, f, h, j) presents the corresponding reconstructed internal RC branch voltages (V_1 and V_2). The close agreement demonstrates the robustness of the proposed PI-LSTM framework to measurement noise.

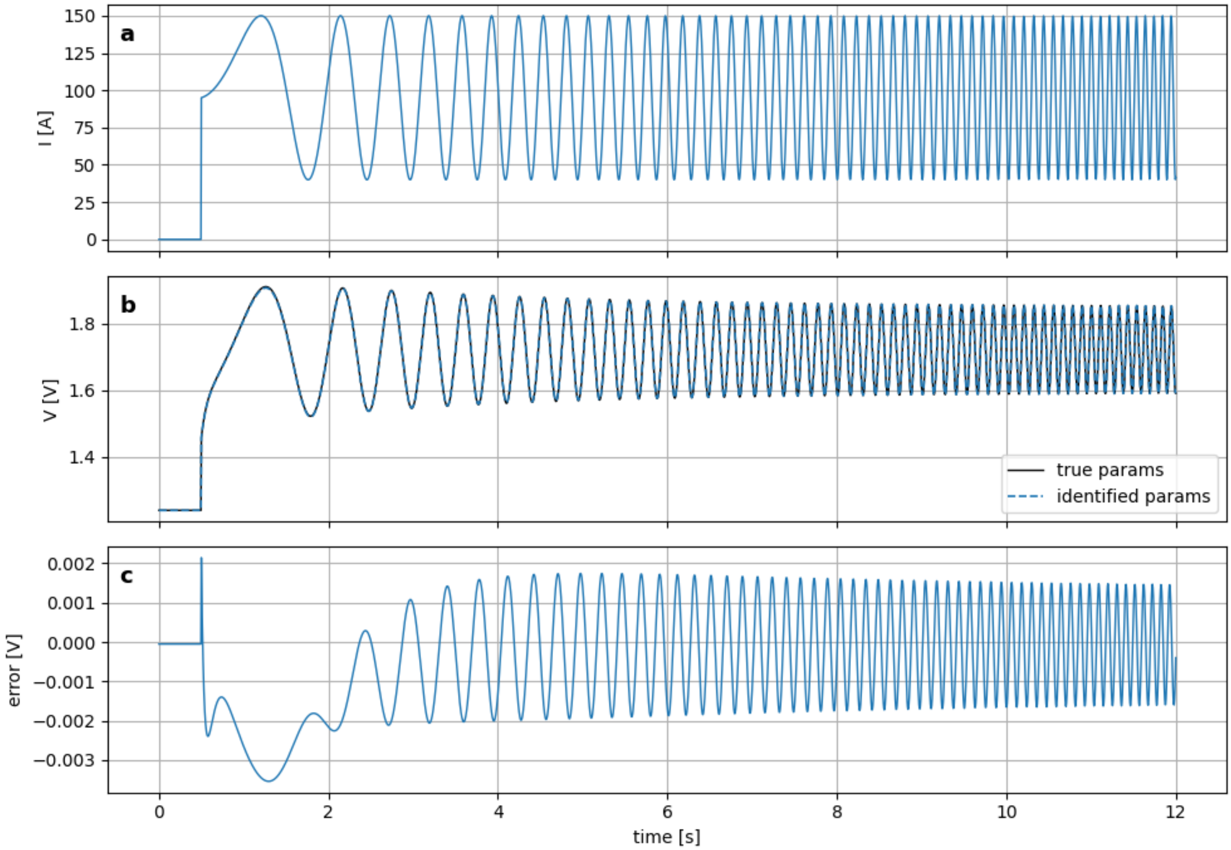

Figure 10: Generalization performance of the proposed PI-LSTM trained with noisy voltage measurements. (a) Unseen chirp current profile used for generalization. (b) Comparison between the reference terminal voltage and the voltage reconstructed using the identified parameters. (c) Corresponding voltage prediction error, demonstrating that the proposed framework maintains high prediction accuracy despite being trained with noisy data.

7. Conclusion

This study presented a comparative evaluation of Physics-Informed Neural Networks (PINNs) and Physics-Informed Long Short-Term Memory (PI-LSTM) networks for the parameter identification of a second-order equivalent circuit model (ECM) of a PEM electrolyzer. The objective was to estimate the six unknown circuit parameters using physics-informed neural networks while assessing the identification accuracy, generalization capability, and robustness of the two approaches.

During the training phase, both of the proposed approaches trained using six different excitation current profiles covering a wide range of dynamic operating conditions. The obtained results demonstrate that the PI-LSTM significantly outperforms the conventional PINN in the parameter identification task. PI-LSTM successfully estimated six circuit unknown parameters with relative errors below 5%, whereas PINN demonstrated considerably larger errors for several dynamic parameters of the equivalent circuit model. These findings indicate that incorporating recurrent memory enables the model to capture the dynamics of the PEM electrolyzer more effectively, resulting in improved parameter estimation accuracy.

The reason why the PI-LSTM architecture outperforms the feed-forward PINN is because the inverse problem is governed by temporal dependencies. While the feed-forward network approximates the solution as a static mapping from space-time coordinates to state variables, the LSTM explicitly models sequential correlations through its hidden state. This enables it to capture transient dynamics, suppress noise, and provide more informative gradients for estimating the unknown physical parameters. Therefore, the LSTM architecture is better aligned with the structure of time-dependent parameter-identification problems. This is not to mean that the PI-LSTM approach is universally superior to other architectures for PINN, but that it is better matched to the temporal character of the data that is needed to identify parameters. This is particularly relevant when, as is our case, there are states in the system for which there is not direct data.

In addition, the proposed PI-LSTM framework exhibited excellent generalization capability when evaluated on previously unseen current profiles and also kept high prediction accuracy in the presence of noisy measurements. The model showed very low RMSE and MAE values together with R^2 score close to one, illustrating generalization capacity and the dynamic voltage behavior of the equivalent circuit model under diverse operating conditions.

Overall, the results suggest that Physics-Informed LSTM constitutes a promising and robust alternative to conventional PINNs for parameter identification of PEM electrolyzer models. Its strength in parameter identification and generalization make it a strong candidate for applications in model-based monitoring, diagnostics, control, and digital twin development for hydrogen energy systems.

Future work will address the application of the method to more complex models of PEM electrolyzers, both higher order and non-linear, as well as the use of experimental data instead of synthetic one. We will also explore the use of the LSTM architecture for forward problems, as well as its integration in a neural ODE framework.

8. Acknowledgment

This work has been supported by:

- DECODER project (PID2024-158394OB-C22), funded by the Spanish Ministry of Science, Innovation and Universities, the Spanish National Research Agency and the European Regional Development Fund (MICIU/AEI/10.13039/501100011033/FEDER UE).
- AVANTE project (PDC2025-165204-C22), funded by the Spanish Ministry of Science, Innovation and Universities and the Spanish National Research Agency (MICIU/AEI/10.13039/501100011033).
- Siemens Energy AI Chair: Energy Sustainability for a Decarbonized Society 5.0 (TSI-100930-2023-5), funded by the Secretary of State for Digitalization and Artificial Intelligence through the ENIA 2022 Chairs call, and co-funded by the European Union-Next Generation EU.

A. Performance Evaluation Metrics

To quantitatively assess the performance of the PINN and PI-LSTM models, several complementary error metrics are employed. These metrics evaluate the discrepancy between the model predictions and a reference solution obtained from either an analytical formulation or a high-fidelity numerical simulation. The evaluation is performed using independent test datasets that are not utilized during training, enabling the assessment of both interpolation capability within the training domain and generalization performance under unseen operating conditions.

Let $t_i{}_{i=1}^{N}$ denote the discrete time instances of the test dataset, where N represents the total number of evaluation samples. For each time instant t_i, $y(t_i)$ denotes the reference value, while $y_\theta(t_i)$ represents the corresponding prediction generated by the PINN or PI-LSTM model.

A.1. Mean Squared Error (MSE)

The Mean Squared Error (MSE) is defined as

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y(t_i) - y_\theta(t_i)\right)^2 \tag{38}$$

MSE quantifies the average squared difference between the predicted and reference values. Owing to the squaring operation, larger prediction errors contribute disproportionately to the metric, making it particularly sensitive to significant deviations. Furthermore, positive and negative errors are treated equally, and the resulting objective function remains smooth and differentiable, which is advantageous for optimization procedures [15].

For dynamic system modeling, MSE is useful for identifying error accumulation during long-term predictions and evaluating model robustness under out-of-distribution operating conditions, such as current excitations beyond the training range. Lower MSE values indicate improved predictive accuracy and stronger generalization capability in previously unseen scenarios.

A.2. Root Mean Squared Error (RMSE)

The Root Mean Squared Error (RMSE) is computed as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y(t_i) - y_\theta(t_i)\right)^2} \tag{39}$$

RMSE represents the standard magnitude of prediction errors and has the same physical units as the target variable. Consequently, it provides a more intuitive interpretation of model performance from an engineering perspective. Unlike MSE, which is expressed in squared units, RMSE directly reflects the typical prediction error magnitude in the original measurement units [30].

A.3. Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) is defined as

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y(t_i) - y_\theta(t_i)\right| \tag{40}$$

MAE measures the average absolute difference between the predicted and reference values. Since the errors are not squared, MAE is less sensitive to outliers than MSE and RMSE while maintaining the same units as the target variable. As a result, it provides an easily interpretable measure of the typical prediction accuracy achieved by the model [23].

In the context of electrolyzer modeling, accurate voltage prediction is essential for reliable monitoring, control, and optimization. Therefore, MAE serves as a practical indicator of the average prediction error that may influence operational performance and control decisions.

A.4. Coefficient of Determination (R^2)

The coefficient of determination is given by

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y(t_i) - y_\theta(t_i)\right)^2}{\sum_{i=1}^{N}\left(y(t_i) - \bar{y}\right)^2}, \tag{41}$$

where

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y(t_i) \tag{42}$$

denotes the mean value of the reference solution.

The R^2 metric evaluates the proportion of variance in the reference data that is explained by the model predictions. A value of $R^2 = 1$ corresponds to perfect agreement between the predicted and reference responses, whereas values approaching zero indicate limited predictive capability. Negative values may occur when the model performs worse than a simple mean-value predictor.

In this study, the R^2 score is employed as a complementary global performance indicator to assess the ability of PINN and PI-LSTM models to reproduce the dynamic voltage behavior of the electrolyzer system [14, 17].

References

[1] Antonelo, E.A., Camponogara, E., Seman, L.O., Jordanou, J.P., Souza, E.R.d., Hübner, J.F., 2024. Physics-informed neural nets for control of dynamical systems. Neurocomputing 579, 127419. doi:10.1016/j.neucom.2024.127419.

[2] Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks 5, 157–166. doi:10.1109/72.279181.

[3] Dang, J., Yang, F., Li, Y., Deng, X., Ouyang, M., 2022. Transient behaviors and mathematical model of proton exchange membrane electrolyzer. Journal of Power Sources 542, 231757. doi:10.1016/j.jpowsour.2022.231757.

[4] Fadziso, T., 2020. Overcoming the vanishing gradient problem during learning recurrent neural nets (rnn). Asian Journal of Applied Science and Engineering 9, 197–208. doi:10.18034/ajase.v9i1.41.

[5] Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: Continual prediction with lstm. Neural Computation 12, 2451–2471. doi:10.1162/089976600300015015.

[6] Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep learning. volume 1. MIT press Cambridge.

[7] Gosu, D.L., Durvasula, A., Annamalai, J., Chowdhary, A.R., Freire, G.S., 2024. Modelling of pem electrolyzer dynamics for green hydrogen production. Computer Aided Chemical Engineering , 505–510doi:10.1016/b978-0-443-28824-1.50085-5.

[8] Guilbert, D., Vitale, G., 2019. Dynamic emulation of a pem electrolyzer by time constant based exponential model. Energies 12, 750. doi:10.3390/en12040750.

[9] Haykin, S., 2009. Neural networks and learning machines, 3/E. Pearson education india.

[10] Hidouri, D., Marouani, R., Chérif, A., 2024. Modeling and simulation of a renewable energy pv/pem with green hydrogen storage. Engineering Technology Applied Science Research 14, 12543–12548. doi:10.48084/etasr.6492.

[11] Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735.

[12] Kim, D., Lee, J., 2024. A review of physics informed neural networks for multiscale analysis and inverse problems. Multiscale Science and Engineering 6, 1–11. doi:10.1007/s42493-024-00106-w.

[13] Kim, J., Oh, C., Kim, K., Lee, J., Kim, T., 2022. Parameter identification of electrical equivalent circuits including mass transfer parameters for the selection of the operating frequencies of pulsed pem water electrolysis. Energies 15, 9303. doi:10.3390/en15249303.

[14] Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research 35, 233–241. doi:10.1029/1998wr900018.

[15] Manoj, S.O., Ananth, J.P., Rohini, M., Dhanka, B., Pooranam, N., Arumugam, S.R., 2022. Fws-dl: forecasting wind speed based on deep learning algorithms. Artificial Intelligence for Renewable Energy Systems , 353–374doi:10.1016/b978-0-323-90396-7.00007-9.

[16] Mao, X., Tian, Y., Yang, A., Zhang, G., 2024. Identification of equivalent circuit parameters for proton exchange membrane (pem) electrolyzer engineering models. IEEE Access 12, 15509–15524.

[17] Nash, J., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part i — a discussion of principles. Journal of Hydrology 10, 282–290. doi:10.1016/0022-1694(70)90255-6.

[18] Nathasarma, R., Roy, B.K., 2023. Physics-informed long-short-term memory neural network for parameters estimation of nonlinear systems. IEEE Transactions on Industry Applications 59, 5376–5384. doi:10.1109/TIA.2023.3280896.

[19] Odungat, A., Chanda, S., 2025. Data-driven model for predicting the electro-thermal behavior of lithium-ion pouch cell under varying ambient temperature conditions. Applied Thermal Engineering 265, 125367. doi:10.1016/j.applthermaleng.2024.125367.

[20] Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics 378, 686–707. doi:10.1016/j.jcp.2018.10.045.

[21] Raman, K.A., Hammacher, L., Kungl, H., Karl, A., Jodat, E., Eichel, R., Karyofylli, V., 2025. Data-driven surrogate modeling for performance prediction and sensitivity analysis of transport properties in proton exchange membrane water electrolyzers. Applied Energy 386, 125529. doi:10.1016/j.apenergy.2025.125529.

[22] Ratib, M.K., Muttaqi, K.M., Islam, M.R., Sutanto, D., Agalgaonkar, A.P., 2024. Electrical circuit modeling of proton exchange membrane electrolyzer: The state-of-the-art, current challenges, and recommendations. International Journal of Hydrogen Energy 49, 625–645. doi:10.1016/j.ijhydene.2023.08.319.

[23] Robeson, S.M., Willmott, C.J., 2023. Decomposition of the mean absolute error (mae) into systematic and unsystematic components. Plos One 18, e0279774. doi:10.1371/journal.pone.0279774.

[24] Sabamehr, M., Batlle, C., Serra-Prat, M., 2026. Pinn for parameter identification of equivalent circuit models of pem electrolyzer. Conference abstract, European Hydrogen Conference 2026.
[25] Sha, Y., Lan, J., Li, Y., Chen, Q., 2023. A physics-informed recurrent neural network for rram modeling. Electronics 12, 2906.
[26] Shahbaz, I., Abdel-Rahman, M.J., Hammad, E., 2025. G-pifnn: A generalizable physics-informed fourier neural network framework for electrical circuits. arXiv preprint arXiv:2512.02712 .
[27] Silva-Vera, E.D., Valdez-Reséndiz, J.E., Alejo-Reyes, A., Rumbo-Morales, J.Y., Rosas-Caro, J.C., Sanchez, V., 2025. Data-driven modeling of proton-exchange membrane fuel cell stacks. International Journal of Hydrogen Energy 141, 1050–1060. doi:10.1016/j.ijhydene.2025.01.157.
[28] Sood, S., Prakash, O., Boukerdja, M., Dieulot, J., Ould-Bouamama, B., Bressel, M., Gehin, A., 2020. Generic dynamical model of pem electrolyser under intermittent sources. Energies 13, 6556. doi:10.3390/en13246556.
[29] Taj, R., 2024. Physics-informed neural networks for electrical circuit analysis: Applications in dielectric material modeling. arXiv preprint arXiv:2411.10483 .
[30] Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. Climate Research 30, 79–82. doi:10.3354/cr030079.
[31] Xie, Y., Han, Z., Zhang, J., Liu, L., 2025. A physics-constrained informer-lstm network for battery state-of-charge estimation. Journal of Energy Storage 136, 118381. doi:10.1016/j.est.2025.118381.

Declaration of interests

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

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Milad Sabamehr reports financial support was provided by Siemens Energy Inc. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.