scientific reports

Check for updates

OPEN

Novel reinforcement learning technique based parameter estimation for proton exchange membrane fuel cell model

Nermin M. Salem[2], Mohamed A. M. Shaheen[2] & Hany M. Hasanien[1,2]✉

Proton Exchange Membrane Fuel Cells (PEMFCs) offer a clean and sustainable alternative to traditional engines. PEMFCs play a vital role in progressing hydrogen-based energy solutions. Accurate modeling of PEMFC performance is essential for enhancing their efficiency. This paper introduces a novel reinforcement learning (RL) approach for estimating PEMFC parameters, addressing the challenges of the complex and nonlinear dynamics of the PEMFCs. The proposed RL method minimizes the sum of squared errors between measured and simulated voltages and provides an adaptive and self-improving RL-based Estimation that learns continuously from system feedback. The RL-based approach demonstrates superior accuracy and performance compared with traditional metaheuristic techniques. It has been validated through theoretical and experimental comparisons and tested on commercial PEMFCs, including the Temasek 1 kW, the 6 kW Nedstack PS6, and the Horizon H-12 12 W. The dataset used in this study comes from experimental data. This research contributes to the precise modeling of PEMFCs, improving their efficiency, and developing wider adoption of PEMFCs in sustainable energy solutions.

Keywords Clean energy, PEMFC, Reinforcement learning

The environmental concerns about fossil fuels have accelerated the search for clean and sustainable energy sources[1]. Fuel cells are one of the most promising options since they can directly transform chemical energy into electrical energy through electrochemical processes[2]. They are efficient and can generate power with low gas emissions[3]. Several fuel cell types have been used, such as MCFCs[4], SOFCs[5], and PEMFCs[6]. Although SOFCs and MCFCs function at higher temperatures[7], PEMFCs are well-known for their mobility[8], making them suitable for energy supply in residential[9], commercial[10], and industrial applications[11].

PEMFC is supposed to play a vital role in a cleaner and more sustainable future[12]. This is because of its advantages, including low emissions and high efficiency[13]. PEMFC has the potential to contribute to solving the problems related to pollution and dependence on fossil fuels[14,15]. Research studies focus on various applications for PEMFCs, such as microgrids[16,17]. Accurate models are necessary to analyze PEMFC performance and validate software simulations using experimental data.

Numerous approaches have been recommended for modeling PEMFCs. These approaches include analytical[18], empirical[19], and theoretical methods[20,21]. The Analytical methods use mathematical equations and physical principles to describe the behavior of PEMFC. The empirical methods depend on data to get relationships between input and output variables. The Theoretical methods develop mathematical models that represent the physical processes of a PEMFC. Theoretical models, both conventional and metaheuristic, are commonly used in research on PEMFC parameters extraction. The methodology introduced in[22] uses a hybrid analytical with a Computational Fluid Dynamics model to optimize the thermos fluid performance of fuel cells. Reference[23] uses a semi-empirical model to analyze voltage performance in PEMFC stacks, considering friction losses. Experimental data also validate the study of that research. In[24], the PEMFC modeling is based on a semi-empirical approach where electrical, thermal, and degradation models simulate PEMFC performance. The approach of such reference is designed to optimize efficiency and lifetime for naval applications. With notable advancements in computing, various meta-heuristic algorithms can be used to solve this problem such as chaos game optimization algorithm[25], Walrus optimization algorithm[26], Coot Bird Algorithm[27], Sunflower

[1]Electrical Power and Machines Department, Faculty of Engineering, Ain Shams University, Cairo 11517, Egypt. [2]Faculty of Engineering and Technology, Future University in Egypt, Cairo 11835, Egypt. ✉email: hanyhasanien@ieee.org

Optimization Algorithm[28], and Transient Search Algorithm[29,30]. The WOA approach is validated in[31] by comparing the model estimated results with experimental data from various PEMFC systems under different conditions.

Recently, there has been interest in applying RL methods to PEMFC modeling. RL offers a promising alternative to traditional methods, as it can learn from data and adapt to changing conditions[32,33]. RL is a subset of machine learning that performs superiorly in environments where direct solutions are hard to compute due to complex, nonlinear relationships[34,35]. RL and other data-driven methods have demonstrated significant potential in computing prediction methodology for various applications[36,37], including Fuel Cell Performance Prediction, providing a promising alternative to the traditional approach. Instead of relying on predefined rules, an RL agent learns through interaction with the system and receives feedback (rewards or penalties) based on its actions[38]. Despite its promising performance, RL remains underutilized in the energy sector, presenting a clear need for further exploration and research[39]. However, recent progress in RL, especially in gradient-based methods like actor-critic algorithms, has significantly improved the ability to learn effective performance prediction model for complex systems.

RL can be classified into several categories based on how the learning process is defined and the types of environments the agent interacts with. For the complex nature of PEMFC optimization, Actor-Critic methods are particularly well-suited for tasks with continuous action spaces, as they can directly learn policies that output continuous actions[40]; it involves tuning continuous nonlinear parameters. PEMFCs are governed by multiple nonlinear parameters, which affect the system's behaviour in intricate ways. This makes them an ideal candidate for RL approaches that balance value-based and policy-based learning.

Actor-critic algorithms are a class of reinforcement learning methods that combine two key networks: the actor $\pi_\theta(s,a)$, which maps states s to a probability distribution over actions a, which determines the best action to take in each state, and the critic, value function V(s), which evaluates the action by estimating the value function. The actor updates its policy based on feedback from the critic, while the critic continuously improves its value estimations based on the rewards received. This actor-critic acts as a two-player game that allows for more stable and efficient learning than standalone policy-based or value-based methods. Actor-critic methods are particularly effective and well-suited in environments with complex, continuous state-action spaces in which decisions involve multiple variables and nonlinear dynamics.

Indeed, the fuel cell optimization problem is considered challenging and focuses on unlocking new opportunities for applying RL. In[41] The authors introduced an RL model that combines the Proximal Policy Optimization (PPO) algorithm with the REINFORCE update rule for optimizing both the design and prediction of the robotic environment, outperforming other methodologies used in[42] and[43]. In[44], the authors adopted a combined policy gradient with a model for optimizing photovoltaic and battery.

In this work, a PPO agent, a type of actor-critic RL method, is employed as an advanced reinforcement learning tool to optimize PEMFCs by minimizing the squared error between measured and simulated terminal voltage, which is governed by seven nonlinear parameters. PPO, an on-policy gradient method, accomplishes a balance between exploring new actions and exploiting known information by constraining the update step size based on the most recent experiences collected during training, which avoids divergence issues while training and guarantees adaptability to change environments[45]. The PPO agent iteratively interacts with the PEMFC system, adjusting the nonlinear parameters based on the reward feedback related to the voltage error reduction. By gradually refining its policy through this interaction, the agent learns an optimal prediction strategy to align the simulated voltage closely with the actual measurements.

The proposed PPO model is also a model-free reinforcement learning approach as it does not require a predefined mathematical model of the environment[46]. It learns directly from environmental interactions, making it versatile and applicable to PEMFCs without an explicit system model.

The proposed on-policy, model-free PPO's advantage lies in its ability to handle complex, continuous prediction problems, making it well-suited for optimizing the intricate dynamics of PEMFCs. This approach offers a robust solution for enhancing fuel cell performance, offering improvements over traditional performance prediction models.

The main contribution of this paper is to present a novel reinforcement learning-based approach for optimizing PEMFC nonlinear parameters and improving the accuracy of PEMFC model under different operating conditions of temperature and pressure for different types of PEMFCs. The remaining sections of the paper are presented as follows. Section 2 provides the mathematical representation of the PEMFC. Section 3 provides the objective function and the constraints. Section 4 provides the methodology. Section 5 includes a discussion of the simulation results. Section 5 presents conclusions and future work for this study.

Mathematical Modelling of PEMFC and objective function formulation

A detailed PEMFC model can be found in[47]. The PEMFC voltage is computed as shown in Eq. (1). In this Equation, N_{cells} is the number of series cells. The total of the voltages across all the individual cells is the voltage of the stack.

$$\boldsymbol{V_{stack}} = \boldsymbol{N_{cells}} \cdot (\boldsymbol{E_{nernst}} - \boldsymbol{v_{act}} - \boldsymbol{v_{\Omega}} - \boldsymbol{v_{conc}}) \tag{1}$$

Here, E_{nernst} refers to the Nernst potential. It computes the PEMFC OC voltage, which is calculated as shown in Eq. (2). v_{act} denotes the activation overpotential. v_{act} denotes the V_{loss} caused by the resistance. v_{conc} denotes the concentration overpotential. Mathematically, v_{act}, v_{Ω}, and v_{conc} are computed as in Eqs. (3)-(5).

$$\boldsymbol{E_{Nernst}} = 1.229 - 0.85 \times 10^{-3}(\boldsymbol{T_{fc}} - 298.15) + 4.3085 \times 10^{-5}\boldsymbol{T_{fc}} ln\left(\boldsymbol{P_{H_2}}\sqrt{\boldsymbol{P_{O_2}}}\right) \tag{2}$$

where, T_{fc} is the temperature in Kelvin. P_{H_2} and P_{O_2} are the partial pressures of hydrogen and oxygen.

$$\boldsymbol{v_{act}} = -\left[\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2 \boldsymbol{T_{fc}} + \boldsymbol{\xi}_3 \boldsymbol{T_{fc}} \boldsymbol{ln}\left(\boldsymbol{C_{O_2}}\right) + \boldsymbol{\xi}_4 \boldsymbol{T_{fc}} \boldsymbol{ln}\left(\boldsymbol{I_{fc}}\right)\right] \tag{3}$$

where $\boldsymbol{C_{O_2}} = \frac{\boldsymbol{P_{O_2}}}{5.08.10^6}.\boldsymbol{exp}(498/\boldsymbol{T_{fc}})$

$$\boldsymbol{v_{\Omega}} = \boldsymbol{I_{fc}}\left(\boldsymbol{R_m} + \boldsymbol{R_c}\right);\ \boldsymbol{R_m} = \frac{\boldsymbol{\rho}_{\boldsymbol{m}} \boldsymbol{l}}{\boldsymbol{M_A}} \tag{4}$$

where $\boldsymbol{\rho}_{\boldsymbol{m}} = \frac{181.6\left[1+0.03\left(\frac{\boldsymbol{I_{fc}}}{\boldsymbol{M_A}}\right)+0.062\left(\frac{\boldsymbol{I_{fc}}}{303}\right)^2\left(\frac{\boldsymbol{I_{fc}}}{\boldsymbol{M_A}}\right)^{2.5}\right]}{\left[\boldsymbol{\lambda}-0.634-3\left(\frac{\boldsymbol{I_{fc}}}{\boldsymbol{M_A}}\right)\right].\boldsymbol{exp}\left[4.18\left(\frac{\boldsymbol{T_{fc}}-303}{\boldsymbol{T_{fc}}}\right)\right]}$

$$\boldsymbol{v_{conc}} = -\boldsymbol{\beta}.\boldsymbol{ln}\left(1 - \frac{\boldsymbol{J}}{\boldsymbol{J_{max}}}\right) \tag{5}$$

Here, ξ_1 to ξ_4 as well as β represent empirical coefficients. C_{O_2} denotes the oxygen concentration. I_{fc} denotes the current. R_m and R_c denote the membrane and contact resistances. ρ_m denotes the membrane resistivity. l denotes the thickness. M_A denotes the surface area. J denotes the current density. λ also represents a design variable. The defined design variables in this study are (ξ_1, ξ_2, ξ_3, ξ_4, λ, R_c, and β). The PEMFC is represented by seven nonlinear parameters directly influencing the fuel cell output voltage. The goal of the optimization task is to minimize the SSE between the measured terminal voltage and the simulated voltage, modeled by the equations governing fuel cell behavior[48]. Equation (6) provides the goal function's mathematical formulation.

$$\boldsymbol{SSE} = \sum_{\boldsymbol{m}=1}^{\boldsymbol{N_{samples}}} \left[\boldsymbol{V_{FC,exp}}\left(\boldsymbol{m}\right) - \boldsymbol{V_{FC,est}}\left(\boldsymbol{m}\right)\right]^2 \tag{6}$$

where, $\mathrm{N_{samples}}$ is the number of experimental readings. $\mathrm{V_{FC,exp}}$ is the experimental voltage. $\mathrm{V_{FC,est}}$ denotes the estimated voltage. Limits, expressed as inequality constraints, apply to the design variables. To optimize the goal, the suggested RL method was conducted using Google Colab notebooks.

Reinforcement learning for PEMFC Parameter Estimation

This section outlines the proposed RL approach for system prediction and then details the customization made to enable learning system designs. The methodology involves designing a custom environment in accordance with the PEMFC model and applying the PPO algorithm to optimize seven nonlinear parameters iteratively to minimize the sum of squared errors (SSE) between measured and estimated fuel cell voltages. Three different cells were tested, and our proposed RL model outperforms other traditional methods.

The problem is simulated as a reinforcement learning task, where the agent's objective is to minimize the SSE between the actual and simulated voltage by adjusting the seven nonlinear parameters.

Environment design

A custom reinforcement learning environment was developed using the gymnasium interface, specifically tailored to the PEMFC system. The software used in this study was created in Python, and the development was done in a GPU-based Google Colab notebook. We relied on the Stable Baselines3 library to build the reinforcement learning agent utilizing the Proximal Policy Optimization (PPO) algorithm. The agent's interactions occurred within a custom environment set up and managed using Gymnasium. As the agent made decisions and received feedback, NumPy handled all the necessary numerical calculations and data processing. Finally, to better understand the results and monitor the learning process, Matplotlib was used to generate visual plots.

The environment defines the state space, action space, and reward structure as follows:

- State Space s: The state space consists of the seven nonlinear parameters influencing the fuel cell's voltage output. Based on empirical data, these parameters are initialized randomly within predefined bounds.
- Action Space a: The action space is continuous and represented by a seven-dimensional vector, where each action corresponds to an adjustment of one of the nonlinear parameters within a normalized range of [-1, 1].
- Reward Function R: The reward is calculated as the negative of the SSE between the measured and simulated terminal voltages. When the SSE falls below a threshold value, an additional reward bonus of 1000 is provided, encouraging the agent to achieve an accurate simulation.
- Termination Criteria: The episode terminates when the SSE falls below a threshold value, indicating sufficient optimization of the parameters or when a predefined maximum number of steps is reached.

The threshold value is defined based on the type of cell tested; it is the minimum SSE achieved for each cell based on the literature review. This threshold value encourages the agent to achieve approximately equal voltage values for the cells tested.

PPO agent training

The PPO algorithm from the `stable-baselines3` library was selected for training the agent due to its robustness in handling continuous action spaces and efficiency in large-scale optimization tasks. The agent's policy was modeled using a multi-layer perceptron (MLP) network. Two networks are created, one for the actor and one for the critic. The actor-network is responsible for outputting a probability distribution over actions. At the same time, the critic network estimates the current state's value, which helps evaluate how good the chosen action was, providing feedback to the actor. Both MLP networks consist of fully connected layers with nonlinear

activation functions, and a Rectified Linear Unit (ReLU) is applied after each fully connected layer to introduce nonlinearity. This nonlinearity is crucial for the network to learn complex representations and relationships within the data.

The PPO agent was trained using the following hyperparameters:

- Learning Rate: 0.0001.
- n_steps: 2048 steps per iteration.
- Batch Size: 64.
- n_epochs: 10.
- Gamma (discount factor): 0.99.
- Clip Range: 0.2.
- Entropy Coefficient: 0.01.

These hyperparameters were selected to balance exploration and exploitation while ensuring stable convergence during training.

Training process

The training process involved initializing the environment with random values of the nonlinear parameters and then allowing the PPO agent to interact with the environment by adjusting the parameters iteratively. At each step, the agent modified the parameters and received feedback as a reward based on the calculated SSE. The agent aimed to minimize the SSE by learning an optimal policy for parameter adjustment. The training was conducted over total_timesteps = 10,000, corresponding to ten epochs, each consisting of an entire episode during which the agent was allowed to interact with the environment until termination. After each iteration, the best-performing set of parameters was recorded based on the lowest SSE achieved. The training lasted for 5 min on a GPU (T4) -based Google Colab Notebook.

Algorithm 1 outlines the optimization process. It begins with initializing the design distribution, facilitating an extensive exploration of designs. Throughout the training process, the framework fine-tunes the policy parameters θ and design parameters ϕ to gradually phase out less effective designs. This enables the policy to specialize and concentrate on a narrowing set of promising designs. Consequently, the variance within the design distribution decreases, steering the system towards converging on an optimal design. The policy is updated using the clipped surrogate objective: $L_{clip} = \min(\mathrm{r}(\theta) * \mathrm{A_t}, \mathrm{clip}(\mathrm{r}(\theta), 1-\epsilon, 1+\epsilon) * \mathrm{A_t}$, where $\mathrm{r}(\theta) = \frac{\pi_\theta(\mathrm{a_t}/\mathrm{s_t})}{\pi_{\theta\_old}(\mathrm{a_t}/\mathrm{s_t})}$ is the ratio between the new and old policies and A_t is the estimated advantage. The value function is updated by minimizing the squared error between the predicted value and the actual return $L_{value}(\theta) = (R_t - V(s_t))^2$, where is the discounted return. The entropy is used to encourage exploration $L_{entropy}(\theta) = -\mathbb{E}[Entropy(\pi_\theta)]$, The entropy coefficient determines how much weight is given to exploration. The total loss function combines the clipped policy objective, value function loss, and entropy bonus $L_{total} = -L_{clip} + c_1 L_{value} - c_2 L_{entropy}$, where c_1 and c_2 are scaling factors that balance the different components.

Algorithm 1

Initialize policy (actor) network $\pi_\theta(s, a)$ and value (critic) network $V_\theta(s)$
Initialize environment-specific parameters for PEMFC cell.
Set PPO hyperparameters (learning rate, batch size, clip range, etc.)

repeat
• **Sample experience tuples** $\{s_t,\ a_t,\ r_t,\ s_{t+1}\}$over n_steps from the environment using π_θ .
• **Compute advantages** A_t and **discounted rewards** R_t:
$A_t = R_t - V(s_t)$
$R_t = r_t + \gamma * V(s_t + 1)$
• **Update value network** V_θ by minimizing the loss:
$\mathrm{L(V)} = \frac{1}{N}\sum (R_t - V(s_t))^2$
• **Update policy network** π_θ using PPO objective with gradient descent:
$\text{Compute ratio } r(\theta) = \frac{\pi_\theta(a_t/s_t)}{\pi_{\theta\_old}(a_t/s_t)}$
$L_{clip} = \min(r(\theta) * A_t,\ clip(r(\theta),\ 1-\epsilon,\ 1+\epsilon) * A_t$
Final loss for policy update
$L_{total} = -L_{clip} + c_1 * L_{value} - c_2 * Entropy(\pi_\theta)$
• **Backpropagate and update** policy and value networks.

Until the End of training

Results, discussion, and insights

The performance of the PPO agent was evaluated by tracking the progression of SSE over time and recording the best set of parameters that minimized the voltage error. In the testing phase of our RL model, we performed a comprehensive evaluation to assess the performance of the trained agent. The raw data collected during testing were used in the visualization phase to gain insights into the model's behavioral pattern and its effectiveness in minimizing the error metrics. The analysis involved plotting the values of SSE, reward, and Zeta parameters for each iteration. All agents were trained for total_timesteps = 10,000, corresponding to ten epochs.

The partial pressure of both oxygen and hydrogen was maintained at 0.5 atm for all experiments. Additionally, the operating temperature was consistently set at 50 °C for all cases.

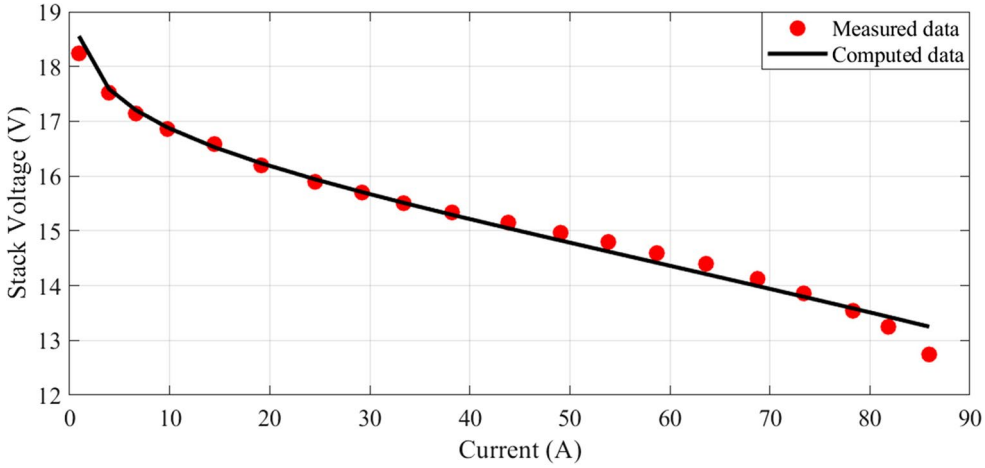

Figure 1. I-V Curves for Case 1.

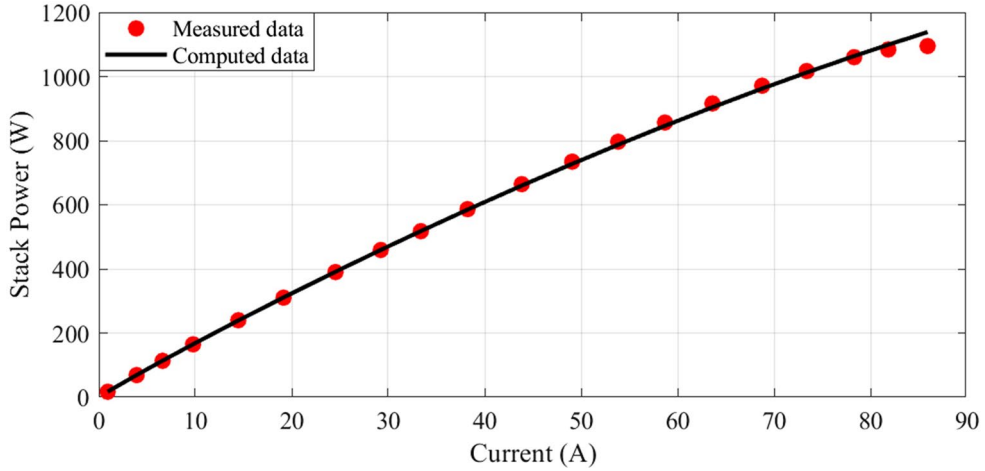

Figure 2. I-P Curves for Case 1.

Case study 1: Temasek 1 kW PEMFC

In this scenario, there are 20 series-connected cells with a total active area of 150 cm^2. The maximum current density is 1.5 A/cm^2. The I-V characteristics are depicted in Fig. 1, and a comparison between the calculated and measured voltages is presented. Additionally, Fig. 2 illustrates the relationship between current and power.

Table 1 compares optimization techniques with the RL approach. It displays the best possible candidate solutions for a range of design variables and provides the best results for reducing the SSE between the observed and estimated terminal voltages achieved by each approach.

The characteristics at different temperatures are examined in the next two figures. I-V waveforms at 50, 70, and 85 °C are shown in Fig. 3. The plot shows that the voltage for a given current increase when the temperature rises. Additionally, Fig. 4 shows how the I-P curves compare at different temperatures. The power curves demonstrate how a little temperature variation affects the output power.

The simulations are run at a constant temperature and with varying pressures. The I-V and I-P graphs are shown in Figs. 5 and 6. They visually represent the outcomes of these simulations. Analyzing these figures indicates a noticeable increase in voltage with pressure.

Figure 7 illustrates the convergence of the design variables over 250 iterations in case 1. The seven design variables are denoted by zeta 0 – zeta 6 in Fig. 7. Each line represents a different design variable value, and the plot shows how these values stabilize as the simulation progresses. Ideally, the curves should remain relatively stable if the agent has found an effective parameter set during training. Fluctuations could suggest that the agent is still exploring slightly during testing or that the environment introduces variability that the agent must respond to by adjusting the parameters. The number of iterations (from 0 to 250) during testing is displayed on the X-axis (Iterations). The smallest SSE was achieved in the 60th iteration.

In contrast, the values of the seven ζ parameters during these iterations are displayed on the Y-axis. The majority of the ζ parameters (Zeta 0, 1, 2, 3, 5, and 6) exhibit stability with minimal variation during testing. This suggests that the PPO agent has effectively learned the optimal values for these parameters during training,

Parameter	RL	EWO[49]	KOA[49]	MPA[49]	HHO[49]
ξ_1	-1.0789201	-0.881369628	-0.8731	-0.9777	-0.8532
ξ_2	0.003271523	0.002988173	2.7642	3.424	0.002329774
ξ_3	5.19432E-05	7.44626E-05	6.13E-05	4.97E-05	3.60E-05
ξ_4	-9.54E-05	-0.0000954	-9.5	-23.6873	-0.0000954
λ	10	13	13	10	13
R_c	1E-04	0.0001	0.0001	0.0001	0.0008
β	0.13017318	0.163327329	0.1619	0.0225	0.0136
SSE	0.559348311	0.578753177	0.590467	0.7559	0.825511853

Table 1. Design variables for case 1.

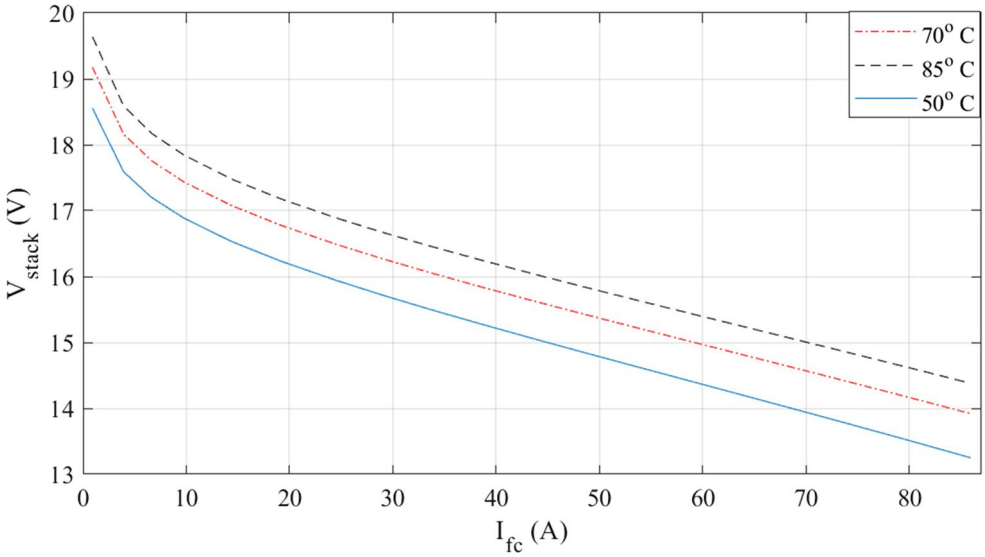

Figure 3. I-V curves at various temperatures for Case 1.

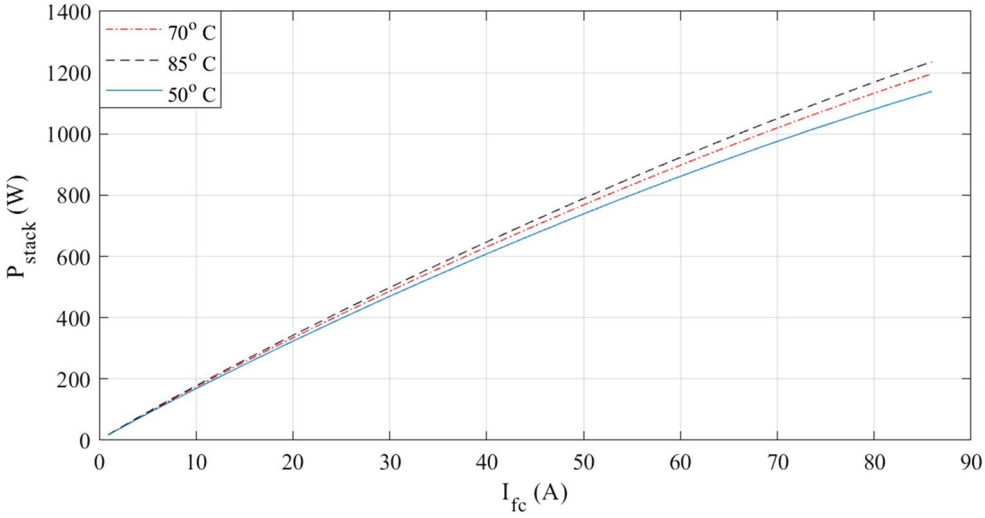

Figure 4. I-P curves at various temperatures for Case 1.

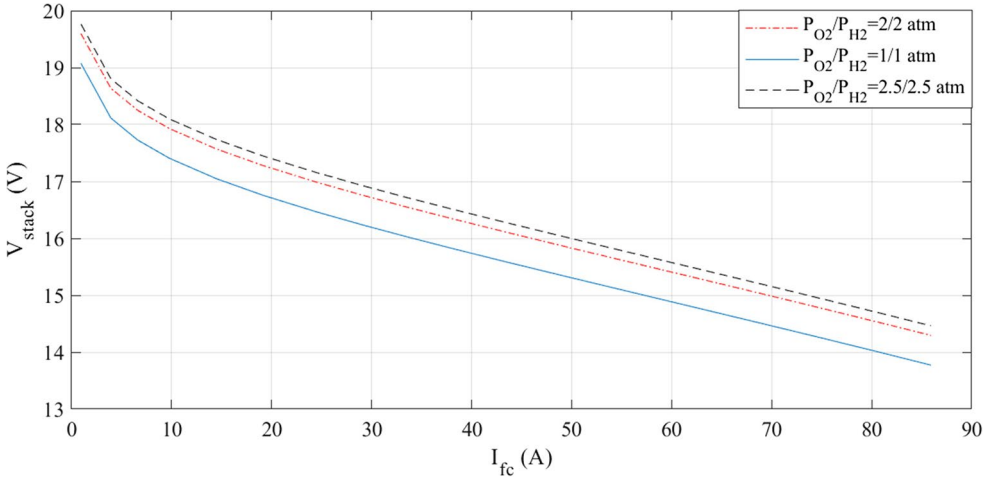

Figure 5. I-V curves at various pressures for Case 1.

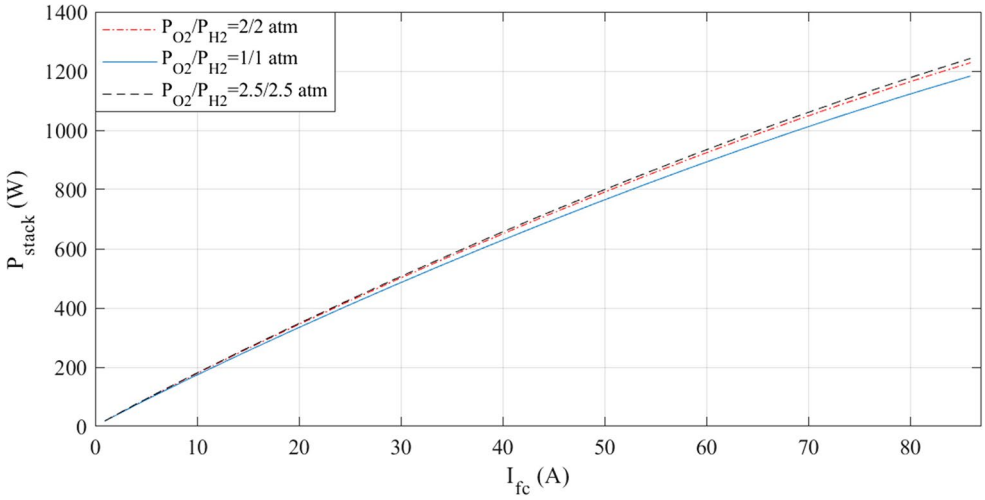

Figure 6. I-P curves at various pressures for Case 1.

and they remain relatively constant in the testing phase. Zeta 4 is the exception, showing significant fluctuations across the testing iterations. This could indicate that this parameter is more challenging for the agent to optimize or that it is highly sensitive to changes in the environment for the three tested cells.

Figure 8 illustrates the variation of the reward function during the testing in Case 1. The x-axis represents the iterations, while the y-axis shows the reward value. The testing reward curve reflects the agent's ability to generalize its learned parameters to new situations.

The SSE curves, Figs. 9 and 18, and Fig. 27, appear to fluctuate within a certain range dependent on each cell tested, indicating that the model is making predictions with varying levels of accuracy throughout the iterations to reach the optimized parameters with the lowest SSE value for each cell. The reward curve will similarly fluctuate as it is often closely tied to the SSE. Fluctuations during testing are a normal part of reinforcement learning, particularly in complex environments with some degree of stochasticity or variability, as in our case.

Case study 2: 6 kW Nedstack PS6 PEMFC

The Nedstack PS6 has a power output of 6 kW, a membrane thickness of 1.78 mm, and 65 cells connected in series. The active area is 240 cm^2, with the highest current density being five A/cm^2. The best values obtained by the RL are compared to those acquired by the algorithms in Table 2. The SSE for the RL findings is lower.

The I-V characteristics are illustrated in Fig. 10, comparing estimated and measured values. A strong correlation between the calculated and measured values indicates a good match between the observed values and the model predictions. Similarly, the I-P curves are depicted in Fig. 11, with the calculated curve closely aligning with the observed data points, demonstrating a strong connection between the estimated values and the actual measurements.

The characteristics at different temperatures are examined in the next two figures. An I-V curve comparison at 50, 70, and 85 °C is shown in Fig. 12. The voltage increases with temperature, as the curves demonstrate.

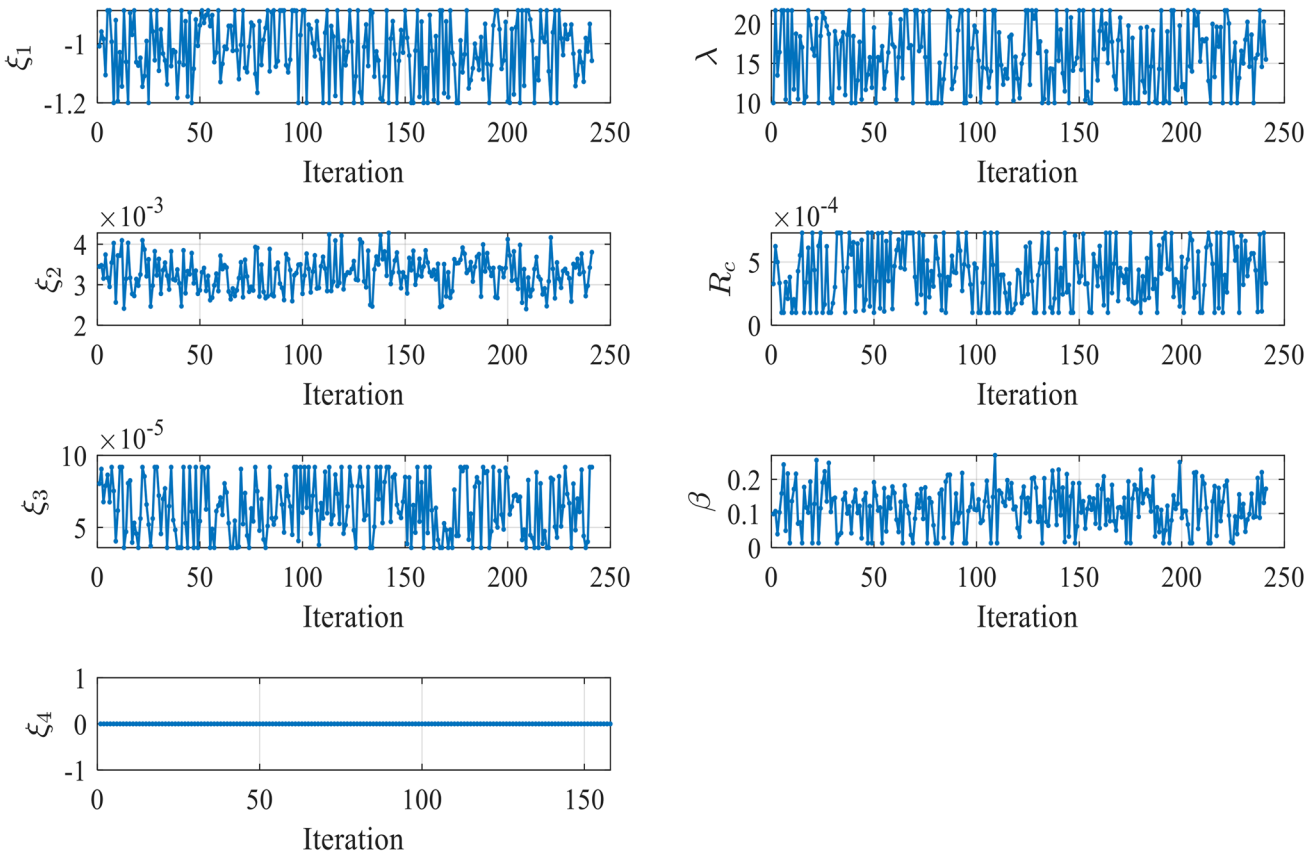

Figure 7. Values of Design Variables Used in Testing in Case 1.

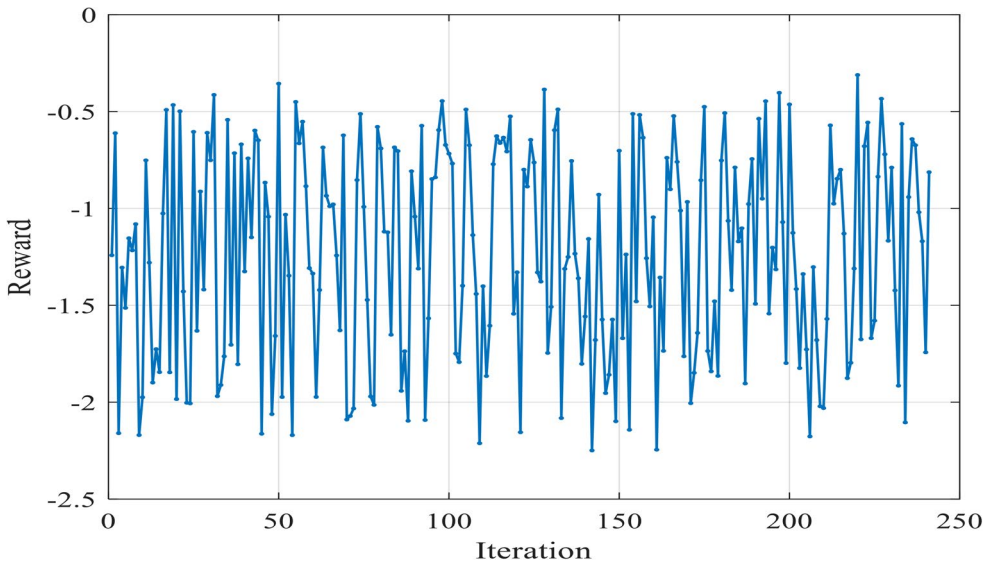

Figure 8. Reward Function Variation during Testing in Case 1.

Additionally, the comparison of I-P curves is shown in Fig. 13. As the temperature varies, so does the output power.

The simulations are run at different pressures while keeping the temperature constant. Figures 14 and 15 present a graphic illustration of the results. An increase in voltage accompanies a rise in pressure.

Figure 16 illustrates the convergence of the design variables over 17 iterations in case 2. Figure 17 illustrates the variation of the reward function during the testing in Case 2. The x-axis represents the iterations, while the y-axis shows the reward value. The smallest SSE was achieved in the 5th iteration as shown in Fig. 18. The seven

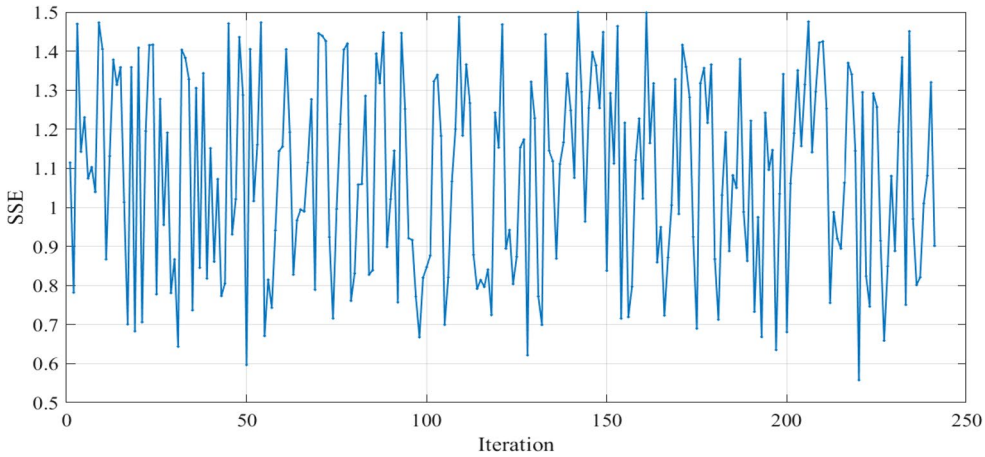

Figure 9. SSE computed through Testing in Case 1.

Parameter	RL	NNA[48]	SSO[48]	TSO[47]
ξ_1	-0.89999998	-0.8535	-0.9719	-0.8532
ξ_2	0.0028	2.4316	3.3487	2.461745
ξ_3	0.000054	3.7545	7.9111	3.94
ξ_4	-9.54E-05	-9.54	-9.5435	-9.54
λ	13.015454	13.0802	13	14.1357
R_c	1E-04	0.1	0.1	0.109423
β	0.0136	0.0136	0.0534	0.1139157
SSE	1.955545929	2.14487	2.18067	2.219

Table 2. Design variables for case 2.

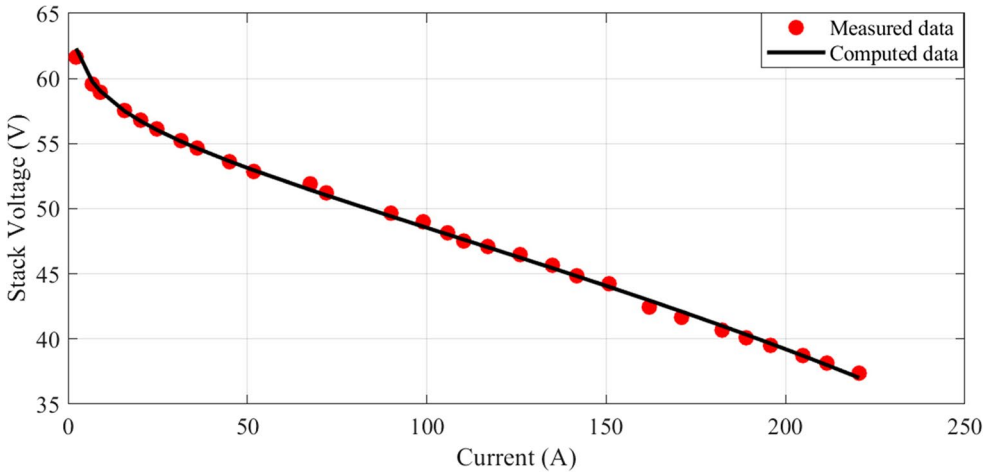

Figure 10. I-V Curves for Case 2.

design variables are denoted by zeta 0—zeta 6 in Fig. 16. Each line represents a different design variable value, and the plot shows how these values stabilize as the simulation progresses.

Figure 17 illustrates the variation of the reward function during the testing in Case 2. The x-axis represents the iterations, while the y-axis shows the reward value.

Case study 3: Horizon H-12, 12 W PEMFC

The Horizon H-12 is a 12 W stack with 13 cells arranged in series and a 25 μm membrane thickness[48]. It has a maximum current density of 0.86 A/cm² and an active area of 8.1 cm². The best values obtained by the RL are

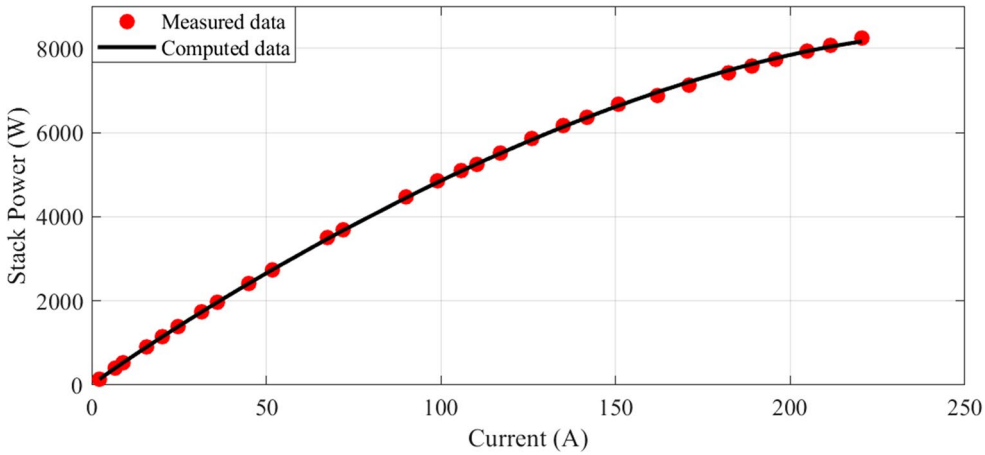

Figure 11. I-P Curves for Case 2.

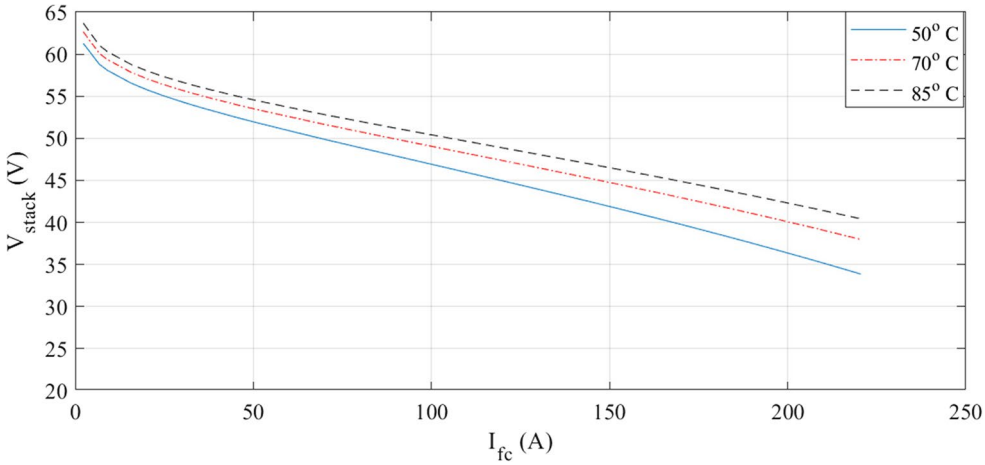

Figure 12. I-V curves at various temperatures for Case 2.

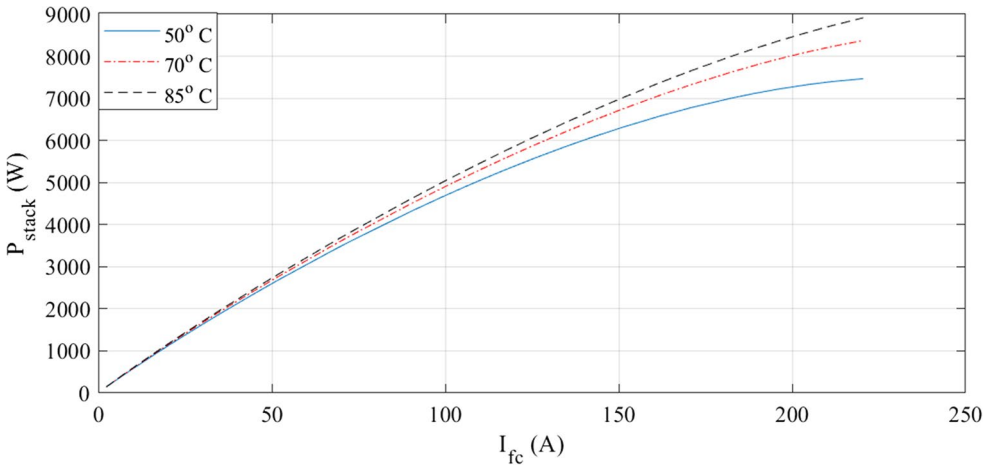

Figure 13. I-P curves at various temperatures for Case 2.

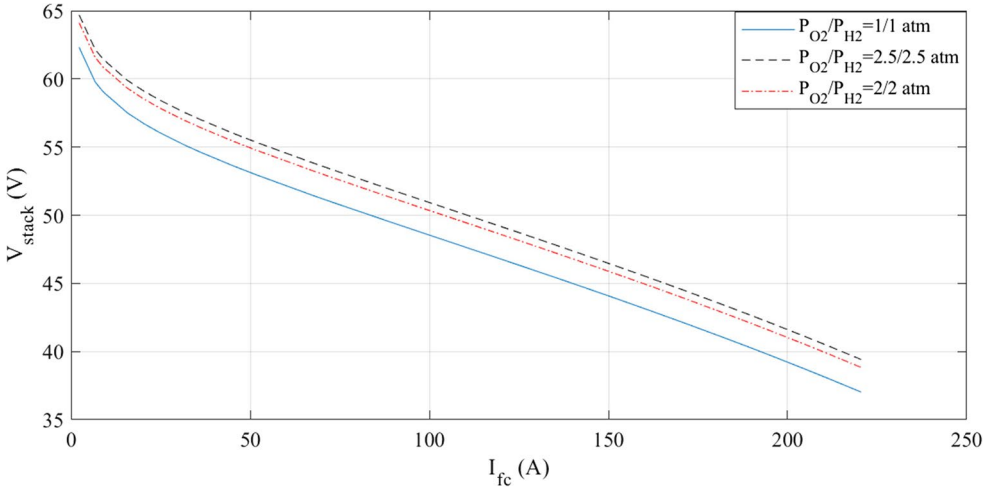

Figure 14. I-V curves at various pressures for Case 2.

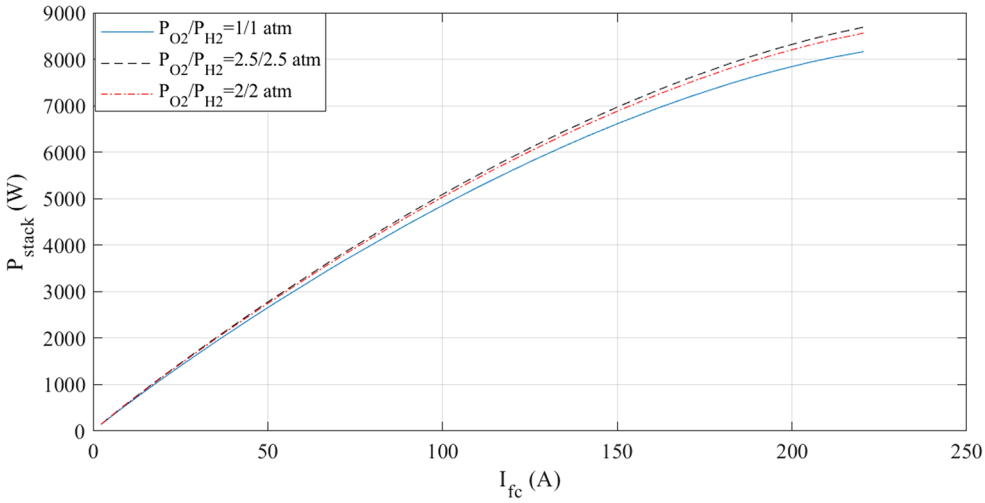

Figure 15. I-P curves at various pressures for Case 2.

listed in Table 3 and compared with those obtained by the competing algorithms. A lower SSE in the RL findings indicates more precise optimization.

The I-V characteristics are shown in Fig. 19, where observed values are compared with estimated values obtained from the model. The observed and calculated values show a high connection. Similarly, the I-P curves for the same PEMFC are demonstrated in Fig. 20. It indicates that there is a significant relationship between the estimated values and the actual measurements.

The properties at different temperatures are examined in the next two figures. I-V curves at temperatures of 50, 70, and 85° C are compared in Fig. 21. The voltage increases with temperature, as the curves demonstrate. Additionally, the comparison of I-P curves is shown in Fig. 22. A slight shift in output power is observed with a rise in temperature.

At a fixed temperature, the simulations are run at different pressures. Figures 23 and 24 provide illustrations of the findings. An increase in voltage accompanies a rise in pressure.

Figure 25 illustrates the convergence of the design variables over 40 iterations in case 3. Figure 26 illustrates the variation of the reward function during the testing in Case 3. The x-axis represents the iterations, while the y-axis shows the reward value. The smallest SSE was achieved in 14th iteration as shown in Fig. 27. The seven design variables are denoted by zeta 0 – zeta 6 in Fig. 25. Each line represents a different design variable value, and the plot shows how these values stabilize as the simulation progresses.

Figure 26 illustrates the variation of the reward function during the testing in Case 3. The x-axis represents the iterations, while the y-axis shows the reward value.

Statistical analysis

To further assess the performance of the RL-based parameter estimation approach for each PEMFC type, a comprehensive statistical analysis was conducted. The following metrics were employed:

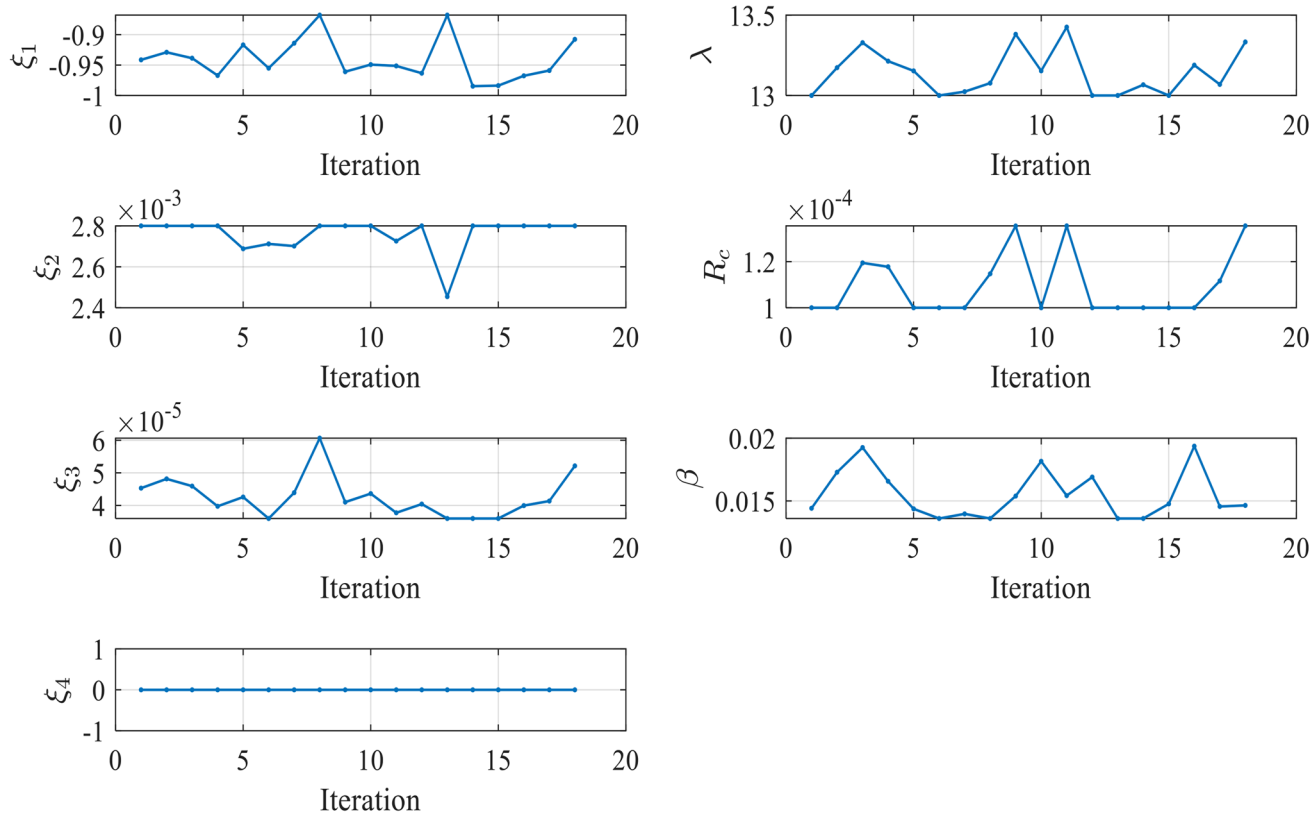

Figure 16. Values of Design Variables Used in Testing in Case 2.

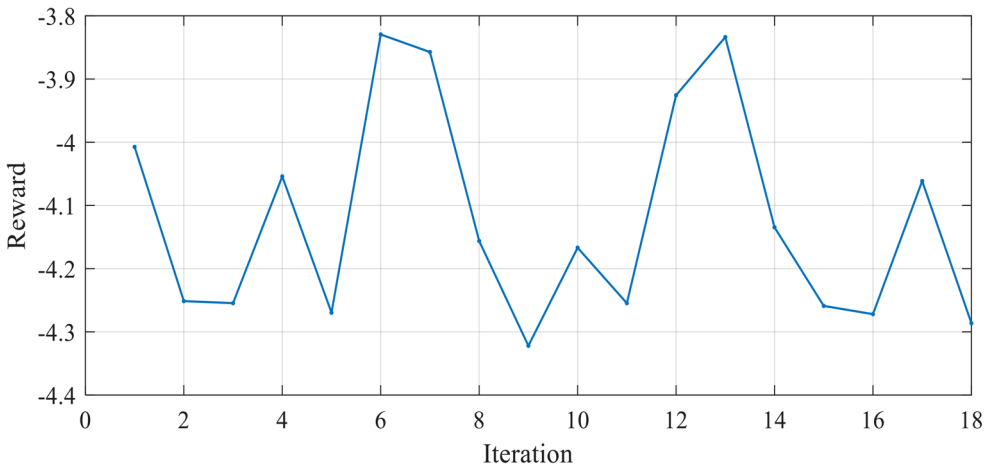

Figure 17. Reward Function Variation during Testing in Case 2.

- MSE: Measures the average squared difference between the estimated and measured voltages.
- MAE: Measures the average absolute difference between the estimated and measured voltages.
- RMSE: The square root of the MSE, providing another measure of the magnitude of errors.
- R-squared: Indicates the proportion of variance in the measured voltage explained by the estimated voltage.

The statistical metrics for the three PEMFC types are presented in Table 4.

The statistical analysis indicates that the RL-based approach effectively estimated parameters for all three PEMFC types. The R-squared values for all cases were high. This indicates that the estimated voltage closely follows the measured voltage. The MSE, MAE, and RMSE values varied across the cell types.

Furthermore, to assess the variability in the performance of the RL-based approach across multiple independent runs, the SSE was calculated for each case. The results are summarized in Table 5. It demonstrates some performance variability. The standard deviations ranging from 0.0012 to 0.0183. The overall performance

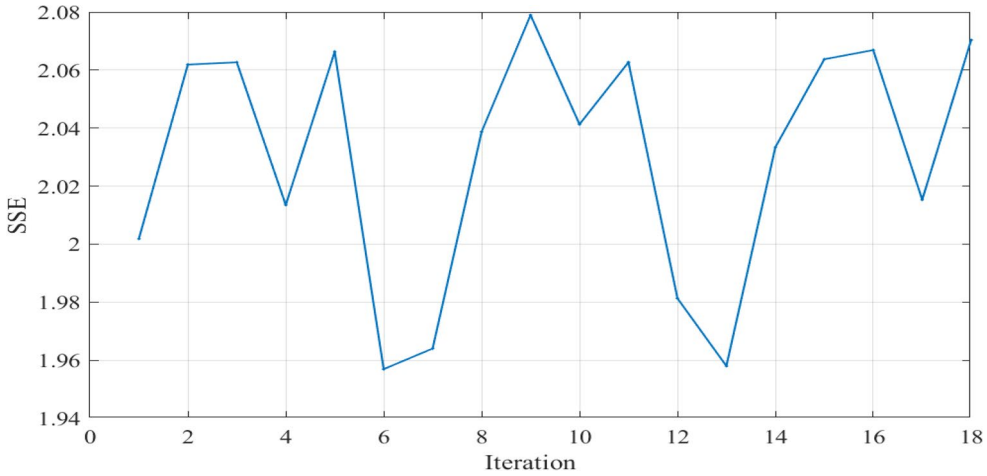

Figure 18. SSE computed through Testing in Case 2.

Parameter	RL	PSO[47]	TSO[47]	WOA[47]
ξ_1	-0.88785005	-1.0347536	-0.8532	-1.187
ξ_2	0.001859869	2.5449	1.571852	2.6697
ξ_3	4.87516E-05	6.32	3.61	3.6
ξ_4	-9.54E-05	-9.54	-9.54	-9.54
λ	14.682545	23	13.0243709	13.824
R_c	0.000173404	0.8	0.327874	0.8
β	0.17679477	0.1827039	0.17527388	0.1598
SSE	0.096572414	0.09658	0.09685	0.116

Table 3. Design variables for case 3.

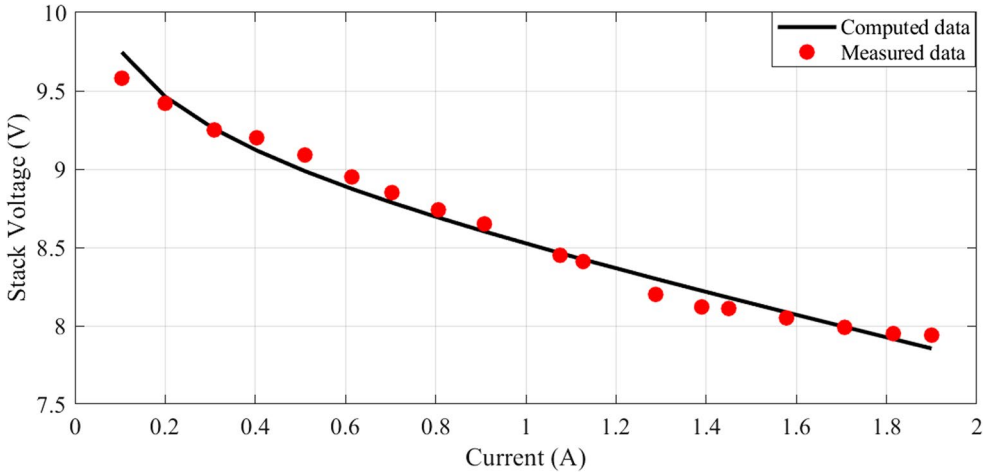

Figure 19. I-V Curves for Case 3.

of the RL-based approach remains consistent across the different cases. The mean SSE values for all cases are relatively low. This indicates that the agent can achieve accurate parameter estimation. These findings highlight the potential of the RL-based approach for reliable parameter estimation in PEMFCs.

Conclusions

The proposed PPO-based reinforcement learning approach successfully optimized prediction strategies for three different PEMFC cells, achieving the goal of developing a theoretical model that closely matches measured

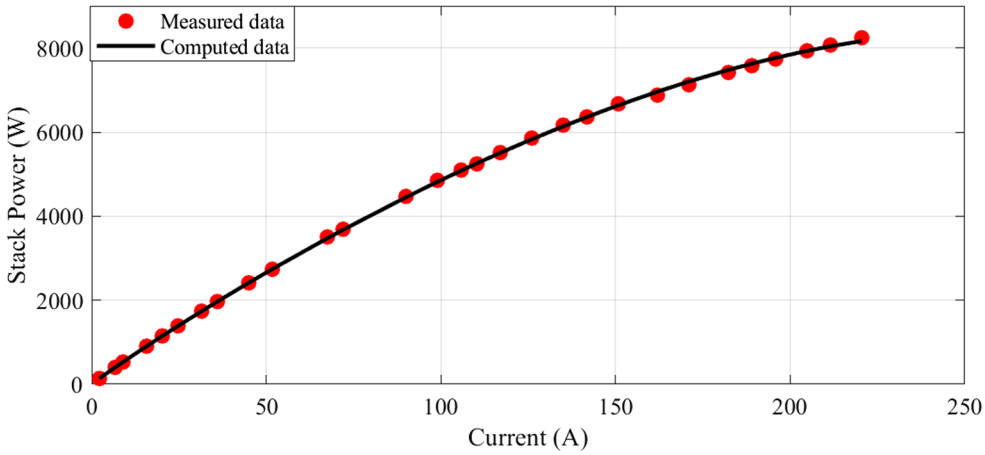

Figure 20. I-P Curves for Case 3.

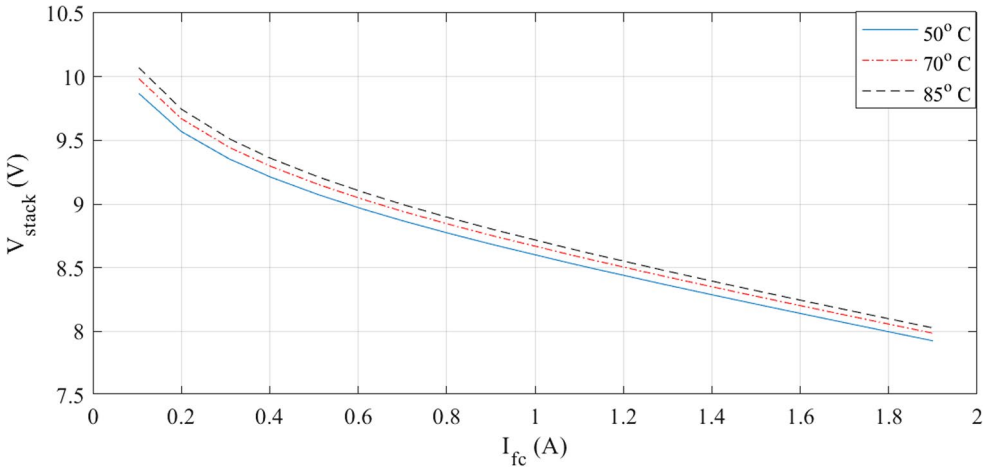

Figure 21. I-V curves at various temperatures for Case 3.

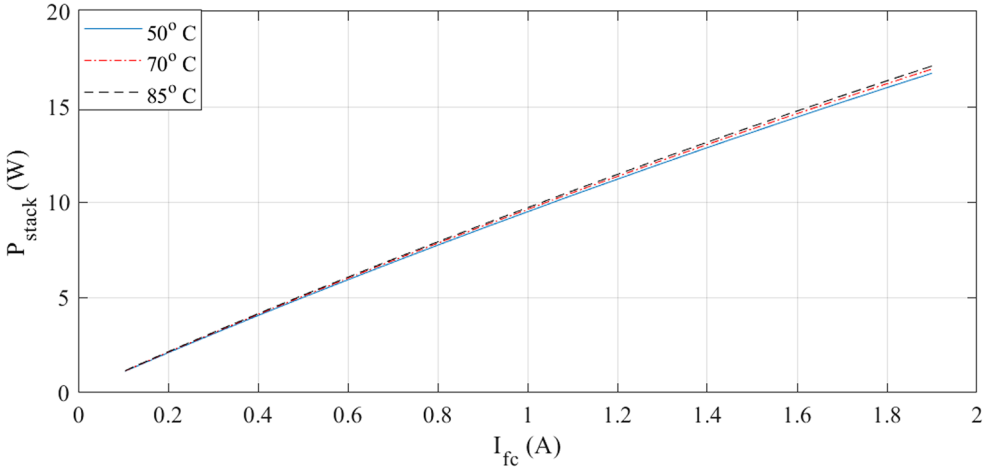

Figure 22. I-P curves at various temperatures for Case 3.

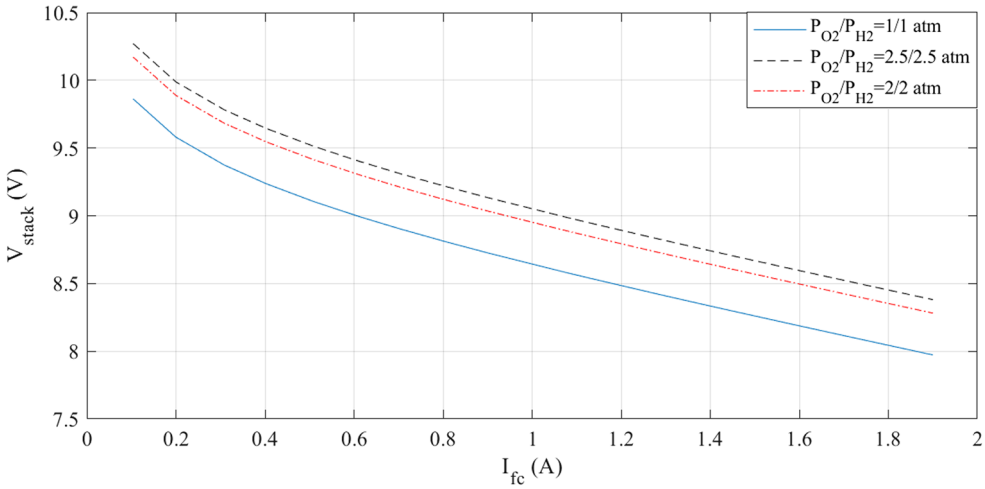

Figure 23. I-V curves at various pressures for Case 3.

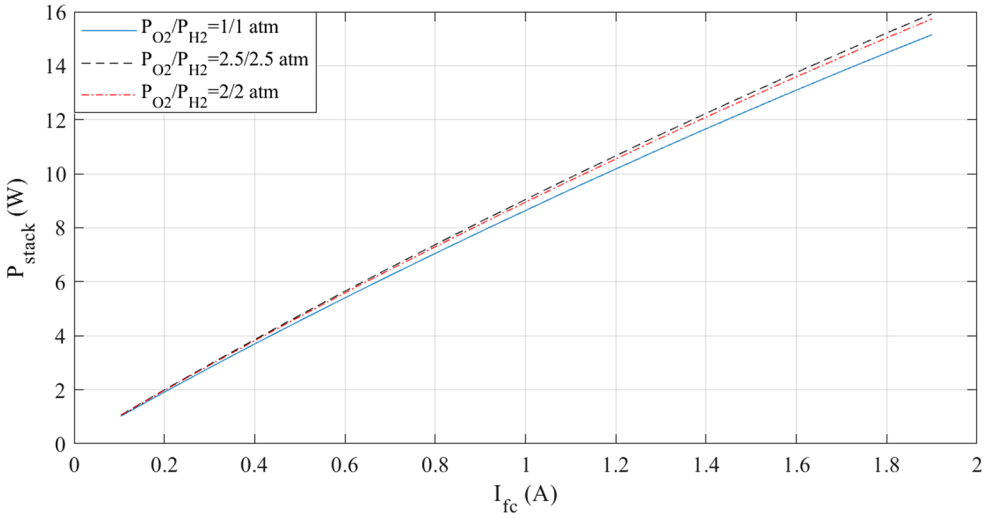

Figure 24. I-P curves at various pressures for Case 3.

data. This article presented a parameter estimation for a PEMFC model, verified under a range of pressure and temperature conditions. The accuracy of the model was evaluated against experimental data and tested on commercial PEMFCs, including the Temasek 1 kW, the 6 kW Nedstack PS6, and the Horizon H-12 12 W. While the performance varied between cells, the agent was able to find optimal design variables for each, minimizing the SSE and improving voltage estimation. The proposed approach achieved an improvement in accuracy ranging from 3 to 48% in case 1, 10–23% in case 2, and up to 23% in case 3. To the knowledge of the authors, the use of reinforcement learning in PEMFC modeling has not been previously explored in the literature, making this study a novel contribution to the field. Fluctuations in the reward and SSE curves are expected due to the complexity and stochastic nature of the environment, but overall, the approach proved to be effective. Further work could focus on improving generalization across cells and refining the agent's performance through targeted hyperparameter tuning and domain adaptation strategies.

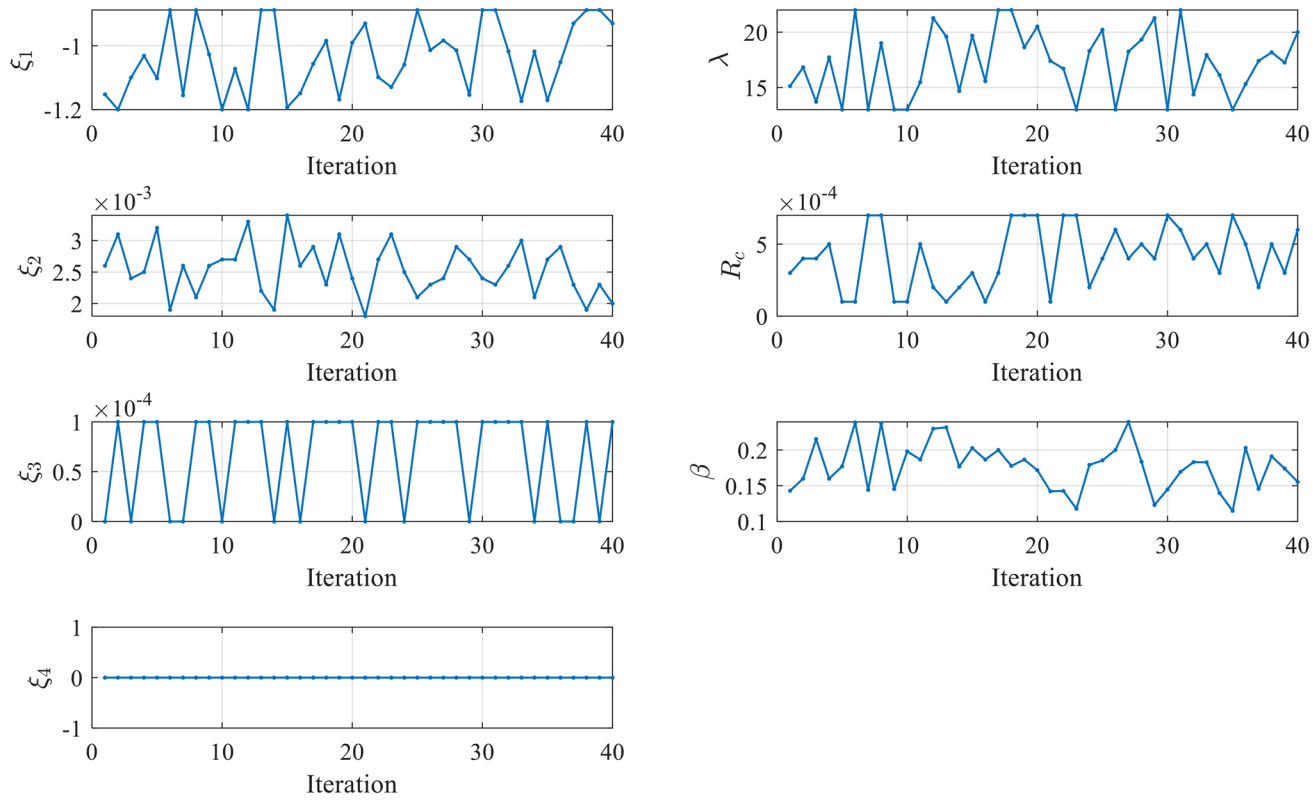

Figure 25. Values of Design Variables Used in Testing in Case 3.

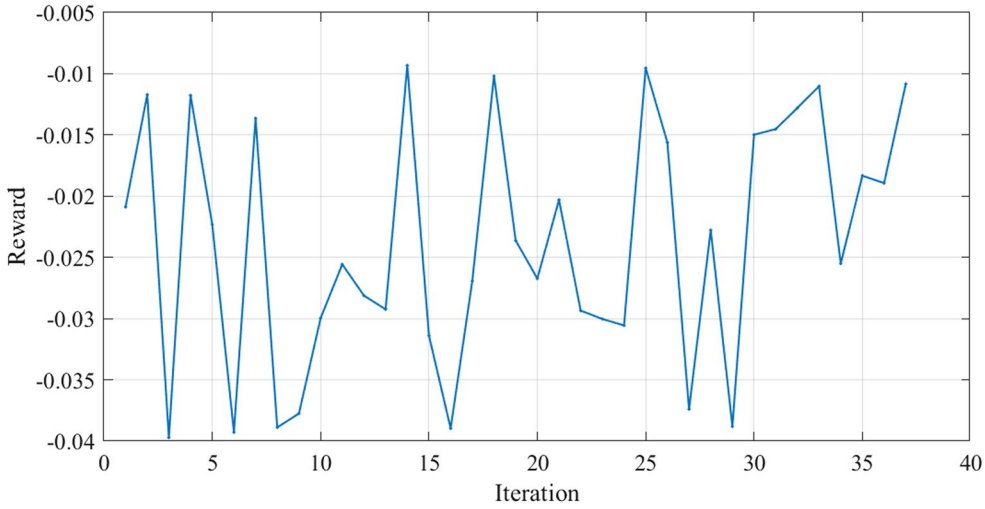

Figure 26. Reward Function Variation during Testing in Case 3.

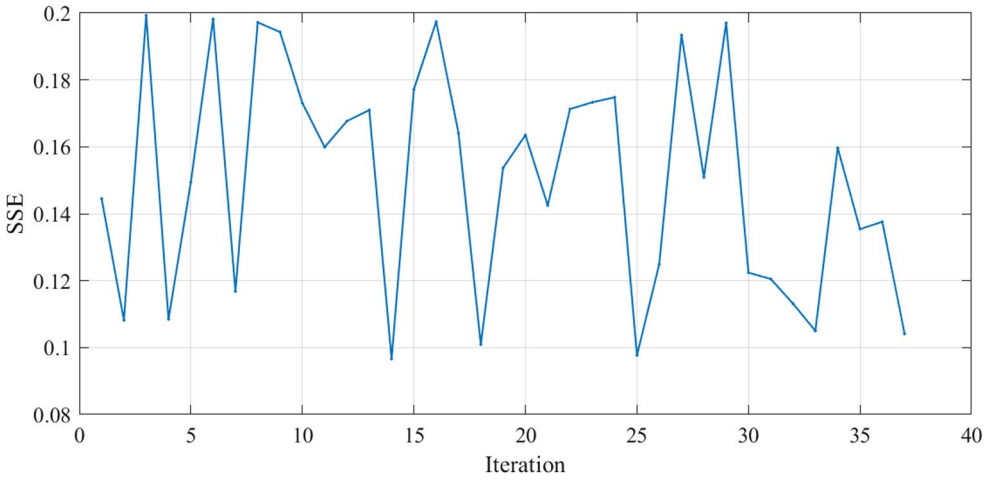

Figure 27. SSE Variation during Testing in Case 3.

	Case 1	Case 2	Case 3
Mean Squared Error (MSE)	0.0375	0.0674	0.005365
Mean Absolute Error (MAE)	0.122	0.201	0.0608
Root Mean Squared Error (RMSE)	0.194	0.26	0.0732
R-squared	0.998	0.999	0.981

Table 4. Statistical Metrics for the three PEMFC types.

Metric	Case 1	Case 2	Case 3
Best SSE	0.5583	1.9555	0.0965
Worst SSE	0.611	1.9591	0.1457
Mean SSE	0.5826	1.9574	0.1034
Standard Deviation of SSE	0.0161	0.0012	0.0183

Table 5. Summary of SSE Statistics for Multiple Independent Runs.

Data availability

Dataset generated during the current study are available from the corresponding author on reasonable request.

Received: 19 September 2024; Accepted: 28 October 2024
Published online: 11 November 2024

References

1. Dincer, I. & Aydin, M. I. New paradigms in sustainable energy systems with hydrogen. *Energy Convers. Manag.* **283**. https://doi.org/10.1016/j.enconman.2023.116950 (2023).
2. Pramuanjaroenkij, A. & Kakaç, S. The fuel cell electric vehicles: the highlight review. *Int. J. Hydrogen Energy*. **48**(25), 9401–9425. https://doi.org/10.1016/j.ijhydene.2022.11.103 (2023).
3. Hassan, Q., Azzawi, I. D. J., Sameen, A. Z. & Salman, H. M. Hydrogen Fuel Cell vehicles: opportunities and challenges. *Sustain. (Switzerland)*. **15**(15). https://doi.org/10.3390/su151511501 (2023).
4. Dybiński, O., Milewski, J., Szabłowski, A., Szczęśniak & Martinchyk, A. Methanol, ethanol, propanol, butanol and glycerol as hydrogen carriers for direct utilization in molten carbonate fuel cells. *Int. J. Hydrogen Energy*. **48**(96), 37637–37653. https://doi.org/10.1016/j.ijhydene.2023.05.091 (2023).
5. Mehran, M. T. *et al.* A comprehensive review on durability improvement of solid oxide fuel cells for commercial stationary power generation systems. *Appl. Energy*. **352**. https://doi.org/10.1016/j.apenergy.2023.121864 (2023).
6. Kahraman, H. & Akın, Y. Recent studies on proton exchange membrane fuel cell components, review of the literature. *Energy Convers. Manag.* **304**. https://doi.org/10.1016/j.enconman.2024.118244 (2024).
7. Pérez-Trujillo, J. P. *et al.* Thermoeconomic comparison of a molten carbonate fuel cell and a solid oxide fuel cell system coupled with a micro gas turbine as hybrid plants. *Energy Convers. Manag.* **276**. https://doi.org/10.1016/j.enconman.2022.116533 (2023).
8. Chakraborty, S. *et al.* A Review on the Numerical Studies on the Performance of Proton Exchange Membrane Fuel Cell (PEMFC) Flow Channel Designs for Automotive Applications. *Energies 2022* **15** (24), 9520. https://doi.org/10.3390/EN15249520 (2022).
9. Lyu, X., Yuan, Y., Ning, W., Chen, L. & Tao, W. Q. Investigation and optimization of PEMFC-CHP systems based on Chinese residential thermal and electrical consumption data. *Appl. Energy*. **356**. https://doi.org/10.1016/j.apenergy.2023.122337 (2024).

10. Zhao, J., Tu, Z. & Chan, S. H. Carbon corrosion mechanism and mitigation strategies in a proton exchange membrane fuel cell (PEMFC): a review. *J. Power Sources*. **488**, 229434. https://doi.org/10.1016/J.JPOWSOUR.2020.229434 (2021).
11. Baroutaji, A. *et al*. PEMFC Poly-Generation systems: developments, merits, and challenges. *Sustain. 2021*. **13**(21), 11696. https://doi.org/10.3390/SU132111696 (2021).
12. Acar, C., Beskese, A. & Temur, G. T. Comparative fuel cell sustainability assessment with a novel approach. *Int. J. Hydrogen Energy*. **47**(1), 575–594. https://doi.org/10.1016/j.ijhydene.2021.10.034 (2022).
13. Sun, D. & Liu, Z. Performance and economic study of a novel high-efficiency PEMFC vehicle thermal management system applied for cold conditions. *Energy*. **305**. https://doi.org/10.1016/j.energy.2024.132415 (2024).
14. Rivarolo, M., Rattazzi, D., Lamberti, T. & Magistri, L. Clean energy production by PEM fuel cells on tourist ships: a time-dependent analysis. *Int. J. Hydrogen Energy*. **45**, 25747–25757. https://doi.org/10.1016/j.ijhydene.2019.12.086 (2020).
15. Ebid, A. M., Abdel-Kader, M. Y., Mahdi, I. M. & Abdel-Rasheed, I. Ant colony optimization based algorithm to determine the optimum route for overhead power transmission lines. *Ain Shams Eng. J.* **15**(1). https://doi.org/10.1016/j.asej.2023.102344 (2024).
16. Samal, K. B., Pati, S. & Sharma, R. A review of FCs integration with microgrid and their control strategies. *Int. J. Hydrogen Energy*. **48**, 35661–35684. https://doi.org/10.1016/j.ijhydene.2023.05.287 (2023).
17. Hussien, A. M., Hasanien, H. M., Qais, M. H. & Alghuwainem, S. Adaptive-width generalized Correntropy Diffusion Algorithm for Robust Control Strategy of Microgrid Autonomous Operation. *IEEE Access*. **11**, 91312–91323. https://doi.org/10.1109/ACCESS.2023.3308039 (2023).
18. Kulikovsky, A. Analytical model for PEM fuel cell concentration impedance. *J. Electroanal. Chem*. **899**. https://doi.org/10.1016/j.jelechem.2021.115672 (2021).
19. Zhao, Y., Luo, M., Yang, J., Chen, B. & Sui, P. C. Numerical analysis of PEMFC stack performance degradation using an empirical approach. *Int. J. Hydrogen Energy*. **56**, 147–163. https://doi.org/10.1016/j.ijhydene.2023.12.096 (2024).
20. Abdel-Kader, M. Y., Ebid, A. M., Onyelowe, K. C., Mahdi, I. M. & Abdel-Rasheed, I. (AI) in infrastructure projects—gap study. *Infrastruct. (Basel)*. **7**(10). https://doi.org/10.3390/infrastructures7100137 (2022).
21. Pan, M. *et al*. Design and modeling of PEM fuel cell based on different flow fields. *Energy*. **207**. https://doi.org/10.1016/j.energy.2020.118331 (2020).
22. Berasategi, J. *et al*. A hybrid 1D-CFD numerical framework for the thermofluidic assessment and design of PEM fuel cell and electrolysers. *Int. J. Hydrogen Energy*. **52**, 1062–1075. https://doi.org/10.1016/j.ijhydene.2023.06.082 (2024).
23. Jiang, Y., Zhang, X. & Huang, L. Analysis on pressure anomaly within PEMFC stack based on semi-empirical and flow network models. *Int. J. Hydrogen Energy*. **48**(8), 3188–3203. https://doi.org/10.1016/j.ijhydene.2022.10.037 (2023).
24. Igourzal, A., Auger, F., Olivier, J. C. & Retière, C. Electrical, thermal and degradation modelling of PEMFCs for naval applications. *Math. Comput. Simul.* **224**, 34–49. https://doi.org/10.1016/j.matcom.2023.04.026 (2024).
25. Shaheen, M. A. M., Hasanien, H. M., Mekhamer, S. F. & Talaat, H. E. A. A chaos game optimization algorithm-based optimal control strategy for performance enhancement of offshore wind farms. *Renew. Energy Focus*. **49**. https://doi.org/10.1016/j.ref.2024.100578 (2024).
26. Shaheen, M. A. M., Hasanien, H. M., Mekhamer, S. F. & Talaat, H. E. A. Walrus optimizer-based optimal fractional order PID control for performance enhancement of offshore wind farms. *Sci. Rep.* **14**(1). https://doi.org/10.1038/s41598-024-67581-x (2024).
27. Hussien, A. M. *et al*. Coot bird algorithms-based tuning PI Controller for Optimal Microgrid Autonomous Operation. *IEEE Access.***10**, 6442–6458. https://doi.org/10.1109/ACCESS.2022.3142742 (2022).
28. Hussien, A. M., Hasanien, H. M. & Mekhamer, S. F. Sunflower optimization algorithm-based optimal PI control for enhancing the performance of an autonomous operation of a microgrid. *Ain Shams Eng. J.* **12**(2), 1883–1893. https://doi.org/10.1016/J.ASEJ.2020.10.020 (2021).
29. Shaheen, M. A. M. *et al*. Enhanced transient search optimization algorithm-based optimal reactive power dispatch including electric vehicles. *Energy*. **277**, 127711. https://doi.org/10.1016/J.ENERGY.2023.127711 (2023).
30. Hussien, A. M., Hasanien, H. M., Qais, M. H. & Alghuwainem, S. Hybrid transient search algorithm with Levy Flight for optimal PI controllers of Islanded Microgrids. *IEEE Access*. **12**, 15075–15092. https://doi.org/10.1109/ACCESS.2024.3357741 (2024).
31. El-Fergany, A. A., Hasanien, H. M. & Agwa, A. M. Semi-empirical PEM fuel cells model using whale optimization algorithm. *Energy Convers. Manag.* **201**. https://doi.org/10.1016/j.enconman.2019.112197 (2019).
32. Milad, R. *et al*. Estimating the stress distribution within MERO joint using (FEM-ANN) hybrid technique. *J. Comput. Sci.* **79**. https://doi.org/10.1016/j.jocs.2024.102294 (2024).
33. Shaheen, M. A. M. *et al*. Probabilistic Optimal Power Flow Solution using a Novel Hybrid Metaheuristic and Machine Learning Algorithm. *Mathematics*. **10**(17). https://doi.org/10.3390/math10173036 (2022).
34. Rashad, A. *et al*. Developing preliminary cost estimates for foundation systems of high-rise buildings. *Int. J. Constr. Manage.* https://doi.org/10.1080/15623599.2024.2352180 (2024).
35. Maher, S. M., Ebrahim, G. A., Hosny, S. & Salah, M. M. A cache-enabled device-to-device Approach Based on Deep Learning. *IEEE Access*. **11**, 76953–76963. https://doi.org/10.1109/ACCESS.2023.3297280 (2023).
36. Perera, A. T. D., Wickramasinghe, P. U., Nik, V. M. & Scartezzini, J. L. Introducing reinforcement learning to the energy system design process. *Appl. Energy*. **262**. https://doi.org/10.1016/j.apenergy.2020.114580 (2020).
37. Quest, H. *et al*. A 3D indicator for guiding AI applications in the energy sector. *Energy AI*. **9**. https://doi.org/10.1016/j.egyai.2022.100167 (2022).
38. François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G. & Pineau, J. An introduction to deep reinforcement learning. *Found. Trends Mach. Learn.* **11**, 3–4. https://doi.org/10.1561/2200000071 (2018).
39. Perera, A. T. D. & Kamalaruban, P. Applications of reinforcement learning in energy systems. https://doi.org/10.1016/j.rser.2020.110618 (2021).
40. Sutton, R. S. & Barto, A. G. Reinforcement learning: an introduction. *IEEE Trans. Neural Netw.* **9**(5). https://doi.org/10.1109/tnn.1998.712192 (2005).
41. Schaff, C., Yunis, D., Chakrabarti, A. & Walter, M. R. Jointly learning to construct and control agents using deep reinforcement learning, in *Proceedings - IEEE International Conference on Robotics and Automation*. https://doi.org/10.1109/ICRA.2019.8793537 (2019).
42. Ha, D. Reinforcement learning for improving agent design. *Artif. Life*. **25**(4). https://doi.org/10.1162/artl_a_00301 (2019).
43. Bhatia, J. S., Jackson, H., Tian, Y., Xu, J. & Matusik, W. Evolution Gym: a large-scale benchmark for Evolving Soft Robots. in *Adv. Neural. Inf. Process. Syst.*, (2021).
44. Cauz, M. *et al*. Reinforcement Learning for Joint Design and Control of Battery-PV Systems, in *36th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems, ECOS 2023*. https://doi.org/10.52202/069564-0281 (2023)
45. Zeng, S., Huang, C., Wang, F., Li, X. & Chen, M. A policy optimization-based deep reinforcement learning method for data-driven output voltage control of grid connected solid oxide fuel cell considering operation constraints. *Energy Rep.* **10**. https://doi.org/10.1016/j.egyr.2023.07.036 (2023).
46. Yuan, H., Sun, Z., Wang, Y. & Chen, Z. Deep reinforcement learning Algorithm based on Fusion optimization for fuel cell gas supply System Control. *World Electr. Veh. J.* **14**(2). https://doi.org/10.3390/wevj14020050 (2023).
47. Hasanien, H. M. *et al*. Precise modeling of PEM fuel cell using a novel enhanced transient search optimization algorithm. *Energy*. **247**, 123530. https://doi.org/10.1016/J.ENERGY.2022.123530 (2022).

48. Selem, S. I., Hasanien, H. M. & El-Fergany, A. A. Parameters extraction of PEMFC's model using manta rays foraging optimizer. *Int. J. Energy Res.* **44**(6), 4629–4640. https://doi.org/10.1002/er.5244 (2020).
49. Alqahtani, A. H., Hasanien, H. M., Alharbi, M. & Chuanyu, S. Parameters estimation of Proton Exchange membrane fuel cell model based on an Improved Walrus optimization Algorithm. *IEEE Access.* **12**, 74979–74992. https://doi.org/10.1109/ACCESS.2024.3404641 (2024).

Author contributions

Nermin M. Salem: Concept, formulation, methodology, investigation, writing the paper. Mohamed A. M. Shaheen: Concept, formal analysis, methodology; validation, writing the paper. Hany M. Hasanien: Concept, validation, visualization, review, supervision.

Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-78001-5.

Correspondence and requests for materials should be addressed to H.M.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.