Intelligent Control of the Air Compressor (AC) and Back Pressure Valve (BPV) to Improve PEMFC System Dynamic Response and Efficiency in High Altitude Regions

Lei Gao [1] and Xuechao Wang [2,3,*]

[1] School of Mechanical and Vehicular Engineering, Beijing Institute of Technology, Beijing 100811, China; allay.gao@refire.com
[2] China Automotive Engineering Research Institute Co., Ltd., Chongqing 401122, China
[3] School of Vehicle and Mobility, Tsinghua University, Beijing 100190, China
* Correspondence: wangxuechao_mail@163.com

Abstract: Proton exchange membrane fuel cells (PEMFCs), as a clean energy technology, show remarkable potential for a wide range of applications. However, high altitude regions pose significant challenges for PEMFC system operation due to thin air and low oxygen partial pressure. Existing logic judgement-based controls exhibit defects such as poor robustness and poor adaptability, which seriously restrict PEMFC system operation. In order to address this issue, this paper puts forth an intelligent control of a PEMFC system air compressor (AC) and back pressure valve (BPV) using an asynchronous advantage actor-critic (A3C) algorithm and systematically compares it with the logic judgement-based control. The application of an A3C-based control under three distinct high altitude test conditions demonstrated a notable enhancement in dynamic responsiveness, with an improvement of up to 40% compared to the results for the logic judgement-based control. Additionally, an improvement of 5.8% in electrical efficiency was observed. The results demonstrate that the A3C-based control displays significant robustness and control precision in response to altitude alterations.

Keywords: proton exchange membrane fuel cells; air compressor; back pressure valve; high altitude regions; deep reinforcement learning

1. Introduction

As a clean energy technology, PEMFCs display a broad application prospect [1,2]. However, due to the thin air and low oxygen partial pressure in high altitude regions, the operation of PEMFCs is greatly challenged, and air path control is one of the crucial aspects of this scenario [3]. If not properly controlled, it may lead to the oxygen starvation phenomenon, a decrease in oxygen transfer rate, a decrease in cell efficiency, etc., and even seriously affect the normal operation of the PEMFC system. Therefore, real-time and precise control of the air path is a key factor to ensure normal operation of the PEMFCs, which is essential to enhance the adaptability and reliability of the PEMFC system in high altitude regions [4,5].

To date, a considerable number of studies have been conducted on the control of the PEMFCs air path in high altitude regions [6–8]. Zhao et al. [9] proposed a dynamic decoupling strategy for peroxide ratio and pressure control based on fuzzy logic. This strategy allows for the expeditious adjustment of the oxygen evolution reaction (OER), which effectively enhances the output power of PEMFCs. Wang et al. [10] proposed a

synergistic optimal allocation scheme based on the maximum efficiency range of multi-stacked PEMFCs to achieve the best power allocation for various types of PEMFCs therein. The results show that the scheme can effectively improve the efficiency of PEMFCs in high altitude regions. Li et al. [11] proposed an extreme value search strategy for optimizing OER. By incorporating a model-referenced adaptive control based on compact dynamic linearization, the method is able to track the optimal OER in real time, thus improving the output power of the PEMFCs in high altitude regions. Chen et al. [12] designed a regulation method combining a particle swarm optimization feed-forward algorithm and a regimen observation algorithm aimed at optimizing the voltage of the air compressor. The strategy enhances the real-time power optimization capability of the PEMFCs and effectively adapts to the effects present in high altitude regions.

The above research results show that logic judgement-based controls and pole-based search strategies have achieved specific results in regards to PEMFCs' air path control in high altitude regions, but there are still some defects. A long calculation time is needed to optimize the parameters, which affects the real-time performance of the system, unable to respond to environmental changes in a timely manner, which may lead to a decline in the system performance or even failures [13–15]. Sensitive to the initial parameter settings, if the initial parameters are not properly selected, the algorithm may get stuck in a local optimum, preventing the system from achieving the best performance and negatively impacting the power output and efficiency of the PEMFCs [16].

In recent years, an increasing number of scholars have applied artificial intelligence algorithms to PEMFCs' air path control, but there is no precedent for intelligent control of the air path in high altitude regions. Artificial intelligence algorithms with stronger adaptive and intelligent features can more accurately capture the complex dynamic characteristics of the PEMFCs and achieve fine adjustment of the air path control. Its high efficiency and real-time performance enable faster responses to changes in high-altitude environments, ensuring the smooth operation of the PEMFCs. The main innovations of this study include the following:

(1) An intelligent control of PEMFC system's AC and BPV based on A3C is proposed to solve the drawbacks of the weak generalization ability of AI algorithms under high altitude conditions by using A3C parallelized training and the efficient use of multi-core CPUs to achieve the smooth operation of the PEMFCs.

(2) In high altitude regions, the focus is on enhancing the efficiency of PEMFCs. This paper also considers the potential for improving the dynamic response of PEMFCs.

The rest of the paper is comprised of the following: Section 2 introduces an intelligent control strategy for PEMFC system's AC and BPV using the A3C algorithm and analyzes the training results. Section 3 provides a comparative analysis of the control outcomes, while Section 4 presents the main conclusions.

2. Methods

2.1. Intelligent Control Framework

Reinforcement learning adjusts the control in real time by interacting with the environment and learning optimal decision paths [17–19]. In this study, an intelligent control strategy for PEMFC system's AC and BPV based on the A3C algorithm is proposed, as shown in Figure 1. The A3C algorithm, leveraging its parallel computing capability, trains multiple agents asynchronously to accelerate learning, maintain policy diversity, and optimize control behavior. By progressively refining its policy, the intelligent body provides optimal control instructions for the PEMFC system, ensuring robust handling of high-dimensional tasks [20–23].
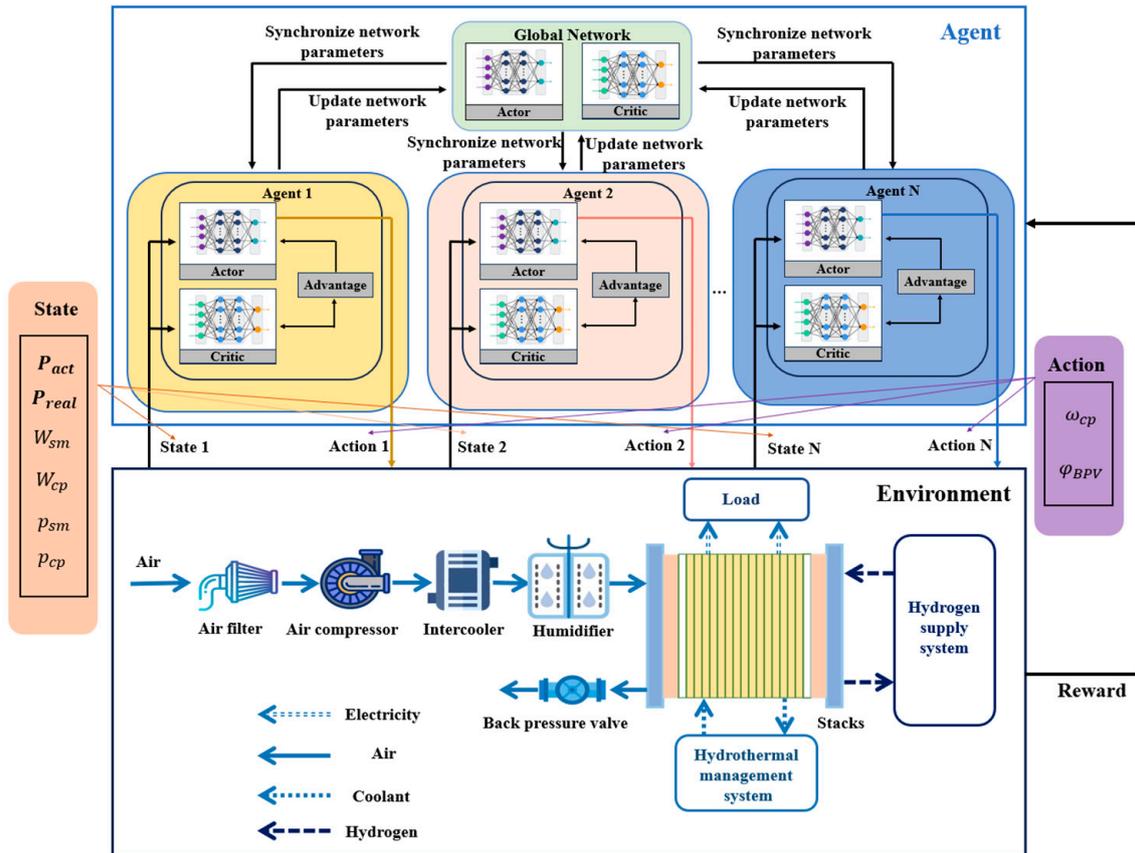
Figure 1. Intelligent control of PEMFCs AC and BPV based on A3C.

2.1.1. Intelligent Body

The key to the A3C algorithm is the dominance function $A(s_t, a_t)$, which is used to measure the advantages and disadvantages of the current action relative to the other actions, helping the intelligent body to better choose the action in the learning process [24]. Its expression is as follows:

$$A(s_t, a_t) = \sum_{i=0}^{n-1} \gamma^i R_t + \gamma^n V(s_{t+n}; \theta_v) - V(s_n; \theta_v) \tag{1}$$

where R_t denotes the immediate reward; γ denotes the discount factor; $\gamma \in [0,1]$ is used to assess the significance of future rewards within the total reward. When $n = 1$, it denotes the single-step reward dominance function, in which the function focuses more on the next reward and favors short-term gains; when $n = k$, it is a multi-step reward dominance function, in which more emphasis is placed on long-term cumulative rewards [25,26].

The loss function expression for the algorithm's strategy network is as follows:

$$L_\pi(\theta_\pi) = \nabla_{\theta_\pi} \ln \pi(a_t | s_t; \theta_\pi) A(s_t, a_t) \tag{2}$$

The value network loss function of this algorithm can be expressed as follows:

$$L_v(\theta_v) = \frac{a(R - V(s_t, \theta_v))^2}{a\theta_v} \tag{3}$$

where R denotes the reward outcome in the current state; $V(s_t, \theta_v)$ refers to the value function in that state.

In the A3C algorithm, multiple threads run in parallel, with each agent following the same algorithmic structure and interacting independently with the environment. The agents retrieve parameters from the global network and update their own networks. Once a terminal state is reached or the iteration limit is exceeded, the agents transmit their updated parameters back to the global network [27,28].

In order to prevent the algorithm from falling into local optimal solutions, the A3C algorithm introduces a policy cross-entropy term in the loss function. This mechanism allows the algorithm to explore the policy space more comprehensively, which enhances the search for globally optimal solutions.

$$L_\pi(\theta_\pi) = \nabla_{\theta_\pi} \ln \pi(a_t|s_t; \theta_\pi)(Q(s,t) - V(s_t, \theta_v)) + \beta \nabla_{\theta'_\pi} H(\pi(s_t, \theta_\pi)) \tag{4}$$

where $H(\pi(s_t, \theta_\pi))$ denotes the strategy cross-entropy, and β denotes the entropy coefficient of the control entropy regularization strength.

The actor network uses a gradient ascent method of the dominance function to update the parameter θ_π. The dominance function is used to assess the degree of dominance an action has over the average. By updating the parameters of the actor network along the positive gradient direction of the dominance function, the aim is to increase the value of the dominance function and thus enhance the performance of the strategy [29].

$$\theta_\pi \leftarrow \theta_\pi + \kappa \sum_n \nabla_{\theta_\pi} \log \pi(a_n|s_n; \theta_\pi) A(s_t, a_t) + \varphi H(\pi(a_n|s_n; \theta_\pi)) \tag{5}$$

where κ denotes the actor network update step.

The critic network uses the temporal difference (TD)-based gradient descent method [30] to update the parameter θ_v. The parameters of the critic network are tuned towards the negative gradient of the TD error, gradually allowing the error to be reduced in the process. In this way, the accuracy of the estimation of the dominance function can be improved, and the optimal solution can be slowly approached.

$$\theta_v \leftarrow \theta_v - \xi \sum_n \nabla_{\theta_v} \left(R_n + \gamma_\gamma V^{\pi(\theta_\pi)}(s_{n+1}; \theta_v) - V_\gamma^{\pi(\theta_\pi)}(s_n; \theta_v) \right) \tag{6}$$

where ξ denotes the update step of the critic network; $R_n + \gamma_\gamma V^{\pi(\theta_\pi)}(s_{n+1}; \theta_v) - V_\gamma^{\pi(\theta_\pi)}(s_n; \theta_v)$ denotes the TD error.

To scale the A3C algorithm for the real and sub-systems of the PEMFC stack, entropy regularization was employed to balance exploration and exploitation. By penalizing overly deterministic policies, entropy loss encourages the algorithm to explore more diverse actions, which is crucial for handling the dynamic and nonlinear behavior of PEMFC sub-systems. Additionally, the policy loss was optimized by minimizing the advantage-weighted log probabilities of the actions, ensuring that the algorithm effectively prioritizes high-reward actions while adapting to real-world uncertainties. This combination of entropy and policy loss adjustments allowed the A3C algorithm to improve its convergence speed, maintain policy stability, and adapt to the specific requirements of PEMFC subsystems.

2.1.2. State and Action Space

In a PEMFC system, the initial stage is the reception of load commands. The system regulates the air path in accordance with the aforementioned commands, thereby ensuring that the requisite air mass flow and pressure are provided to the stack for the generation of the optimal power. The state quantity, designated as s_t, is expressed as follows:

$$s_t = \{P_{act}, P_{real}, W_{sm}, W_{cp}, p_{sm}, p_{cp}\} \tag{7}$$

where P_{act} denotes the demanded power of the PEMFC, P_{real} denotes the actual power of the PEMFC, W_{sm} represents the air mass flow of the inlet pipe, W_{cp} represents the inlet air mass flow of the air compressor, p_{sm} denotes the inlet pipe pressure, and p_{cp} denotes the inlet pressure of the air compressor.

The action a_t is described as a two-dimensional vector, as follows:

$$a_t = \{\omega_{cp}, \varphi_{BPV}\} \tag{8}$$

where ω_{cp} denotes air compressor speed; φ_{BPV} denotes back pressure valve opening.

2.1.3. Bonus Function Setting

The reward function is of paramount importance in reinforcement learning, as it directly influences the efficacy of the training process and the ultimate performance of the model. The reward function R is as follows:

$$R = -\left\{\chi(P_{act}(t) - p_{real}(t))^2 + \delta(1 - \eta)^2 - \varepsilon M_t\right\} \tag{9}$$

where χ, δ, and ε are the weighting coefficients; η is the efficiency of the PEMFC; M_t is the penalty term, which is $M_t = 1$ when the peroxide ratio is $0.5 < \lambda_{O_2} < 1$, $M_t = 2$ when $\lambda_{O_2} < 0.5$, and $M_t = 0$ otherwise.

2.2. Training Result Analysis

In this paper, the high altitude portion of the southern section of the G318 Sichuan–Tibet Highway is selected to generate 300 s of driving conditions as the pre-training data for the A3C algorithm. This driving condition covers a variety of actual driving situations to ensure that the A3C algorithms are effectively trained in diverse processes. The relevant training parameter settings are shown in Table 1.

Table 1. Table of training parameter settings.

Parametric	Numerical Value
Network optimizer	Adam
Learning rate	0.001
Experience pools	1,000,000
Discount factors	0.99
Actor network update step	0.01
Critic network update step	0.001

The A3C algorithm was implemented using Python (Matlab 2022b) on a workstation with an Intel Core i7-10750H CPU, 64 GB DDR4 RAM, and an NVIDIA GeForce GTX 1660 Ti GPU (Lenovo Rescuer, Beijing, China). Training for 1000 episodes took approximately 48 h. The asynchronous parallel mechanism of A3C enhanced computational efficiency by reducing the load compared to that of traditional single-threaded reinforcement learning algorithms, effectively handling high-dimensional tasks. Figure 2 illustrates the average cumulative reward trend during training. Initially, in the first 400 rounds, large fluctuations occur as the algorithm explores. After 400 rounds, the fluctuations decrease, and the average reward stabilizes after the 600th round, indicating that the A3C algorithm's parameter settings and network structure are well-suited to support the learning process.
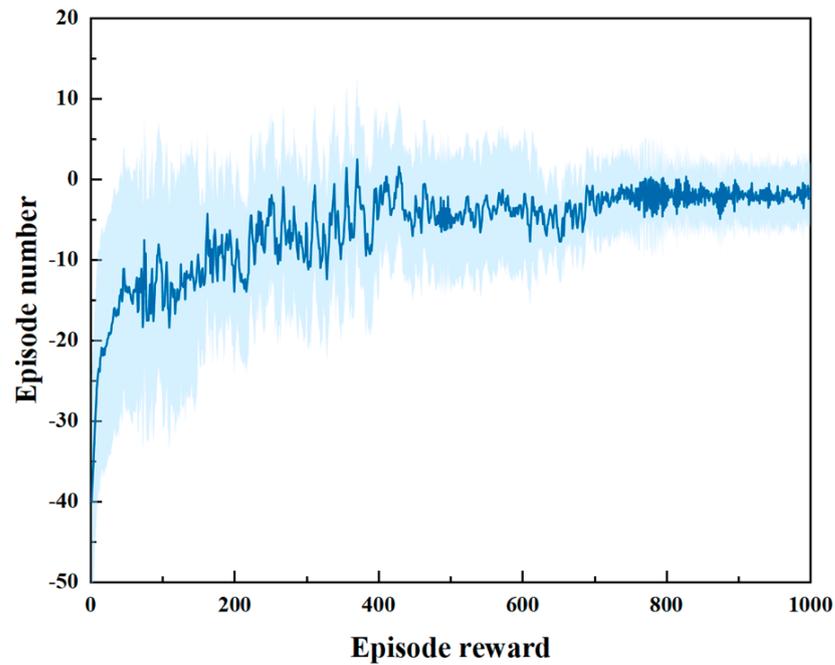
Figure 2. Graph showing the reinforcement learning training results.

3. Results and Discussion

This paper selects the G318 Sichuan–Tibet Highway, including the Chengdu–Ya'an, Kangding–Keduo Pass, and Ranwu–Bomi sections as the test conditions, as shown in Figure 3. The environmental conditions in this section are diverse and challenging, making them very suitable for evaluating the PEMFC system at high altitudes and under dynamic workloads. Table 2 provides a comparison of the air pressure, oxygen partial pressure, and altitude of these areas.
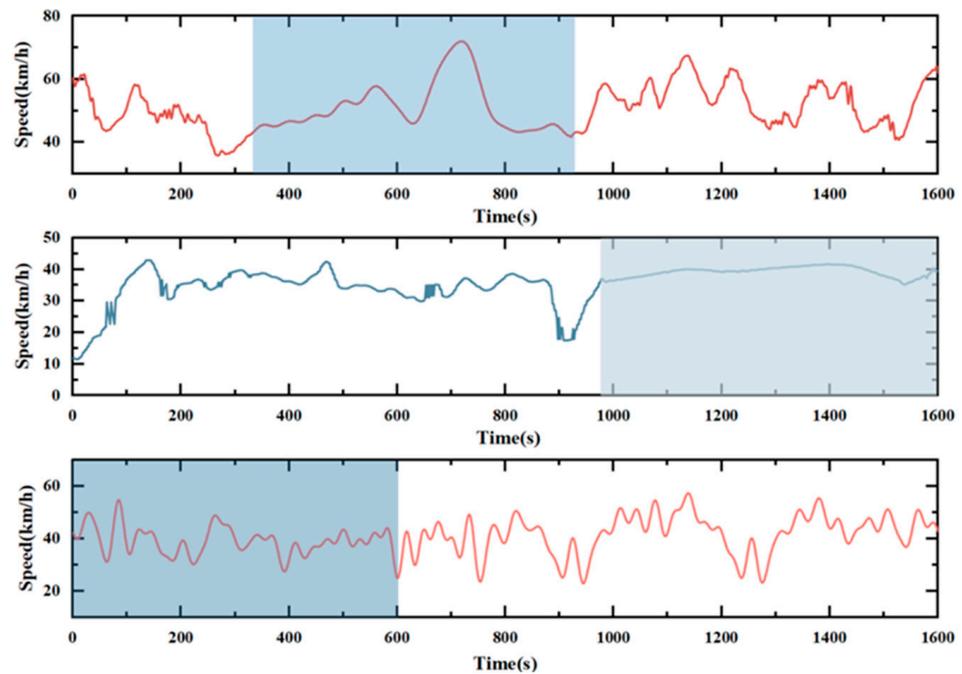
Figure 3. Vehicle speed diagram for testing conditions.

Table 2. Altitude, air pressure, and oxygen partial pressure of three locations.

Region	Altitude (m)	Air Pressure (kPa)	Oxygen Partial Pressure (kPa)
Chengdu–Yaan	500–700	86–92	18–19
Kangding–Keduo Pass	2500–4300	62–79	13–16
Ranwu–Bomi	3850–2700	64–73	13–15

In test condition 1, which comprises a gradually ascending road segment from Chengdu to Yaan at an altitude of 500–700 m, the velocity of the PEMFC vehicle exhibits minimal fluctuations. For the majority of the time, it is in a medium-high speed operational state, with a relatively stable demand for power. However, towards the latter portion of the condition, there is a notable instance of high power operation. In test condition 2, which comprises the section of the route between Kangding and Xuedo Pass at altitudes ranging from 2500 to 4300 m, the speed of the PEMFC vehicle is maintained at approximately 40 km per hour. Furthermore, the power demand of the PEMFC system continues to increase, and the vehicle operates under a full load for an extended period. Test condition 3 is the Ranwu–Bomi section, with an altitude of 3850 m–2700 m above sea level. The speed of the PEMFC vehicle exhibits significant fluctuations, resulting in a rapid change in the fuel cell's demanded power. Consequently, the PEMFC system operates in a state of a continuous variable load.

In accordance with test condition 1, the status of the air compressor and the back pressure valve, in addition to the dynamic response and efficiency of the PEMFC system under A3C-based control, are obtained and compared with those under logic judgement-based control. As illustrated in Figure 4, the A3C-based control exhibits a high degree of correlation between the PEMFC system's power demand and the actual power, thereby effectively maintaining the stable operation of the PEMFC system. As can be observed in the enlarged figure, an increase in altitude results in the logic judgement-based control exhibiting a tendency to overshoot in terms of both the air compressor speed and the back pressure valve opening. In contrast, the A3C-based control is capable of generating commands for the air compressor speed and the back pressure valve opening in a relatively prompt manner, thereby facilitating a smooth and steady-state transition. Additionally, the air compressor power also increases rapidly, which ultimately leads to a reduction in the PEMFC system's electrical efficiency. As the elevation is increased from 560 m to 600 m, the logic judgement-based control requires 1.3 s to reach a steady state, at which point the electrical efficiency of the PEMFC ssytem is 80.3%. In contrast, the A3C-based control reaches a steady state in a mere 0.8 s, with the electrical efficiency increasing to 83.8%. By comparing the two control methods, it can be observed that the A3C-based control significantly improves the dynamic response by 38.4%; in addition, its electrical efficiency is improved by 4.3%.

Logic judgement-based control relies on pre-defined rules and models, which are less adaptable and less likely to cope with complex and dynamically changing high altitude regions. Comparatively, the A3C algorithm is able to dynamically adjust the control parameters by learning and optimizing the strategy to better adapt to load changes.

Figure 5 illustrates the comparative analysis of the impact of the A3C-based control and the logic judgement-based control on the state of the air compressor and the back pressure valve, the dynamic response of the PEMFC system, and the system efficiency under test condition 2. With the A3C-based control, there is a gradual decrease in the fit between the demanded power and the actual power of the PEMFC system with increasing altitude. As the altitude increases rapidly, the logic judgement-based control results in a gradual increase in the volatility of the air compressor speed, the back pressure valve

opening and the PEMFC system's electrical efficiency, which in turn causes instability. In contrast, the A3C-based control is capable of effectively achieving a smooth transition. In particular, the logic judgement-based control requires 1.5 s to reach a steady state when the elevation is rapidly increased from 2600 m to 2840 m. During this period, the electrical efficiency of the PEMFC is 73.8%. In contrast, the A3C-based control requires only 0.9 s to achieve a steady state, with the electrical efficiency of the PEMFC system increasing to 78.1% under these conditions. A comparison of the two control methods reveals that the A3C-based control significantly enhances the dynamic response by 40%, while also improving its electrical efficiency by 5.8%.
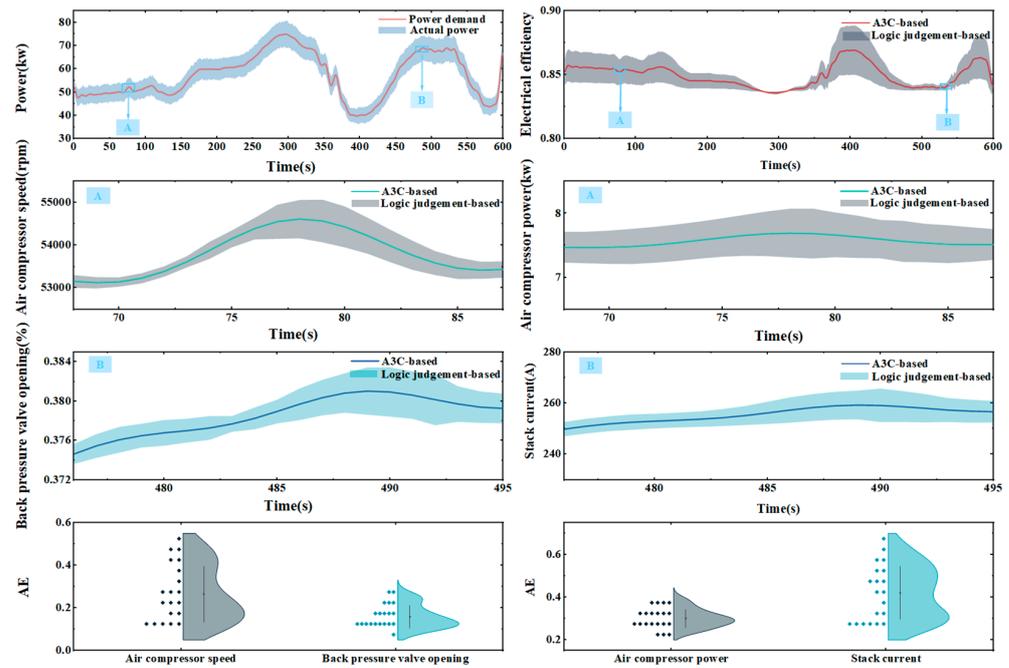
Figure 4. PEMFC system's state and efficiency under test condition 1.
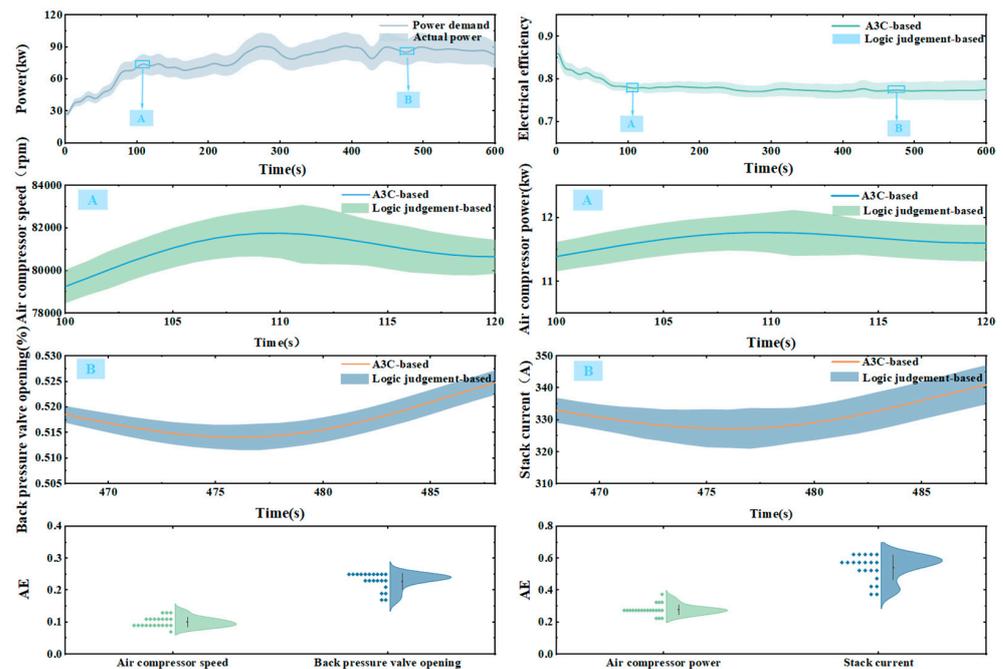
Figure 5. PEMFC system's state and efficiency under test condition 2.

Logic judgement-based control fails to consider the holistic nature of the system and the local optimal solution problem, while the A3C algorithm achieves effective decision making through global optimization. Therefore, the A3C algorithm is able to significantly improve the control effect through effective exploration and utilization in high altitude regions, thus becoming a better choice.

Figure 6 presents a comparative analysis of the A3C-based control and the logic judgement-based control in terms of the states of the air compressor and back pressure valve, the dynamic response of the PEMFC system, and the system efficiency under test condition 3. It is evident that the two control methods exhibit disparate performance in various indices, with the A3C-based control exhibiting a pronounced superiority over the logic judgement-based control in terms of dynamic response speed and system efficiency. As the altitude decreases rapidly, the fluctuations in compressor speed, back pressure valve opening, and PEMFC system electrical efficiency under logic judgement-based control become smaller; however, the control effect remains inferior to that of the A3C-based control. During the rapid elevation increase from 3800 m to 3650 m, it takes 1.6 s for the logic judgement-based control to reach the steady state, during which time the electrical efficiency of the PEMFC system is 69.7%. In contrast, the A3C-based control reaches a steady state in a mere 0.9 s, with the electrical efficiency of the PEMFC system increasing to 73.3%. A comparison of the two control methods reveals that the A3C-based control significantly enhances the dynamic response by 43.7%, while also improving its electrical efficiency by 5.1%.
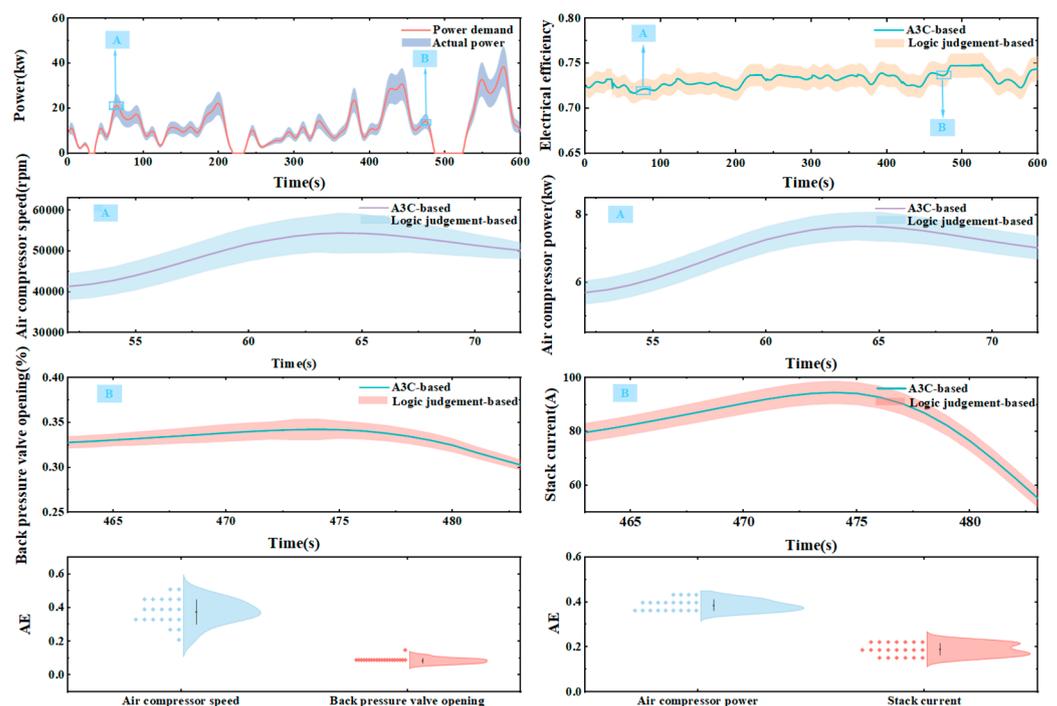
Figure 6. PEMFC system state and efficiency under test condition 3.

The A3C algorithm is capable of autonomously learning and optimizing the control using a substantial quantity of training data, employing multiple threads for parallel updating, which enables the model to converge rapidly and enhances its generalization capacity. In the error analysis, normalization was applied to standardize the data for more consistent evaluation. From the error values of the two control methods, the performance of the PEMFC A3C-based control is more stable with increasing altitude compared to the logic judgment-based control. These findings suggest that the A3C-based control exhibits

superior performance in terms of dynamic response and efficiency when compared to the logic judgment-based control.

However, driver variability was not explicitly considered in the current tests, as the research focused on evaluating the PEMFC system's performance under controlled conditions to minimize external influences. Future studies will incorporate different driving patterns to simulate real-world variability and further validate the robustness of the proposed control strategy.

4. Conclusions

In order to achieve the smooth operation of PEMFC vehicles in high altitude regions, this paper constructs an intelligent control of PEMFC AC and BPV, based on A3C. The specific research conclusions are as follows:

(1) As the altitude increased, the air mass flow and pressure could be increased to maintain smooth PEMFC operation by increasing the air compressor speed and decreasing the back pressure valve opening. The rapid increase in altitude from 2600 to 2840 m resulted in an increase in compressor power of 5.8 kw, with a corresponding 40% increase in dynamic response, but a 6.9% decrease in electrical efficiency.

(2) In this paper, an intelligent control of PEMFC AC and BPV based on A3C is proposed. The method enables the PEMFC system to reach a stable output quickly in high altitude regions. Under three high-altitude test conditions, the A3C-based control reaches a steady state within a maximum of 0.9 s and improves the electrical efficiency up to 78.1%. The A3C-based control significantly improves the dynamic response by 40% and the electrical efficiency by 5.8% compared to the results for the logic judgement-based control. These results show that the proposed optimization strategy offers obvious advantages in significantly improving the operational performance of the PEMFC system.

Author Contributions: Methodology, L.G.; Validation, X.W.; Formal analysis, L.G.; Investigation, L.G.; Writing—review & editing, X.W.; Supervision, X.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: For data quality control purposes, we may decide not to disclose certain data to avoid misuse or misinterpretation.

Conflicts of Interest: Xuechao Wang is affiliated with China Automotive Engineering Research Institute Co., Ltd. The authors declare no conflict of interests.

Glossary

s_t	immediate state
a_t	immediate action
R_t	immediate reward
γ	reward weight
θ_π	actor network parameters
θ_v	critic network parameters
β	entropy coefficient
κ	the actor network update step
ξ	the critic network update step

$A(s_t, a_t)$	advantage function
$\pi(a_t \vert s_t; \theta_\pi)$	current strategy
$V(s_t, \theta_v)$	the value function in that state
$H(\pi(s_t, \theta_\pi))$	the strategy cross-entropy
$R_n + \gamma_\gamma V^{\pi(\theta_\pi)}(s_{n+1}; \theta_v) - V_\gamma^{\pi(\theta_\pi)}(s_n; \theta_v)$	the TD error
P_{act}	the demanded power of the PEMFC system
P_{real}	the actual power of the PEMFC system
W_{sm}	the air mass flow of the inlet pipe
W_{cp}	the inlet air mass flow of the air compressor
p_{sm}	the inlet pipe pressure
p_{cp}	the inlet pressure of the air compressor
ω_{cp}	the air compressor speed
φ_{BPV}	the back pressure valve opening
$\chi, \delta, \varepsilon$	the weighting coefficients
η	the efficiency of the PEMFC system
Mt	the penalty term

References

1. Sazali, N.; Wan Salleh, W.N.; Jamaludin, A.S.; Mhd Razali, M.N. New Perspectives on Fuel Cell Technology: A Brief Review. *Membranes* **2020**, *10*, 99. [CrossRef] [PubMed]

2. Mitra, U.; Arya, A.; Gupta, S. A comprehensive and comparative review on parameter estimation methods for modelling proton exchange membrane fuel cell. *Fuel* **2023**, *335*, 127080. [CrossRef]

3. Martinez-Boggio, S.; Di Blasio, D.; Fletcher, T.; Burke, R.; García, A.; Monsalve-Serrano, J. Optimization of the air loop system in a hydrogen fuel cell for vehicle application. *Energy Conv. Manag.* **2023**, *283*, 116911. [CrossRef]

4. Zhu, J.; Zhang, P.; Li, X.; Jiang, B. Robust oxygen excess ratio control of PEMFC systems using adaptive dynamic programming. *Energy Rep.* **2022**, *8*, 2036–2044. [CrossRef]

5. Bacher-Chong, E. Constraint-Aware and Efficiency-Aware Control of Air-Path in Fuel Cell Vehicles. Ph.D. Thesis, The University of Vermont and State Agricultural College, Burlington, VT, USA, 2022.

6. Zhao, D.D.; Xu, L.C.; Huangfu, Y.G.; Dou, M.F.; Liu, J.X. Semi-physical modeling and control of a centrifugal compressor for the air feeding of a PEM fuel cell. *Energy Conv. Manag.* **2017**, *154*, 380–386. [CrossRef]

7. Sun, L.; Shen, J.; Hua, Q.S.; Lee, K.Y. Data-driven oxygen excess ratio control for proton exchange membrane fuel cell. *Appl. Energy* **2018**, *231*, 866–875. [CrossRef]

8. Wei, J.; Qi, M.; Zhang, H.; Li, X. Investigation on Exhaust Energy Recovery System Using Radial Turbine in High-Power Proton Exchange Membrane Fuel Cells. In Proceedings of the ASME Turbo Expo 2022: Turbomachinery Technical Conference and Exposition, Florence, Italy, 13–17 June 2022; Volume 86052, p. V007T18A018. [CrossRef]

9. Zhao, D.D.; Hua, Z.G.; Dou, M.F.; Huangfu, Y.G. Control oriented modeling and analysis of centrifugal compressor working characteristic at variable altitude. *Aerosp. Sci. Technol.* **2018**, *72*, 174–182. [CrossRef]

10. Wang, Y.M.; Chen, W.R.; Li, Q.; Han, Y.; Guo, A.; Wang, T.H. Coordinated optimal power distribution strategy based on maximum efficiency range of multi-stack fuel cell system for high altitude. *Int. J. Hydrog Energy* **2024**, *50*, 374–387. [CrossRef]

11. Li, S.; Qiu, Y.B.; Yin, L.Z.; Li, R.R.; Gan, R.; Li, Q.; Huangfu, Y.G. Net Power Optimization Based on Extremum Search and Model-Free Adaptive Control of PEMFC Power Generation System for High Altitude. *IEEE Trans. Transp. Electrif.* **2023**, *9*, 5151–5164. [CrossRef]

12. Chen, J.Z.; He, H.W.; Quan, S.W.; Wei, Z.B.; Zhang, Z.D.; Wang, Y.X. Real-time power optimization based on PSO feedforward and perturbation & observation of fuel cell system for high altitude. *Fuel* **2024**, *356*, 129551.

13. Song, W.J.; Chen, H.; Guo, H.; Ye, F.; Li, J.R. Research progress of proton exchange membrane fuel cells utilizing in high altitude environments. *Int. J. Hydrog Energy* **2022**, *47*, 24945–24962. [CrossRef]

14. Hu, D.H.; Wang, Y.T.; Li, J.W.; Wang, J.; Yang, Q. Energy Saving Control of Waste Heat Utilization Subsystem for Fuel Cell Vehicle. *IEEE Trans. Transp. Electrif.* **2023**, *10*, 3192–3205. [CrossRef]

15. Feng, Z.M.; Huang, J.; Jin, S.; Wang, G.Q.; Chen, Y. Artificial intelligence-based multi-objective optimisation for proton exchange membrane fuel cell: A literature review. *J. Power Sources* **2022**, *520*, 230808. [CrossRef]

16. Lu, D.G.; Yi, F.Y.; Hu, D.H.; Li, J.W.; Yang, Q.Q.; Wang, J. Online optimization of energy management strategy for FCV control parameters considering dual power source lifespan decay synergy. *Appl. Energy* **2023**, *348*, 121516. [CrossRef]

17. Hu, D.; Zhang, J.; Huang, J.; Li, J.; Yang, Q.; Wang, J. Speed fluctuation suppression control of super-high-speed electric air compressors considering high-frequency electromagnetic excitation. *IEEE Trans. Power Electron.* **2024**, *39*, 9650–9660. [CrossRef]

18. Lyu, L.; Shen, Y.; Zhang, S.C. The Advance of Reinforcement Learning and Deep Reinforcement Learning. In Proceedings of the IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Changchun, China, 25–27 February 2022; IEEE: Changchun, China, 2022; pp. 644–648.

19. Hu, D.; Wang, Y.; Li, J.; Yang, Q.; Wang, J. Investigation of optimal operating temperature for the PEMFC and its tracking control for energy saving in vehicle applications. *Energy Convers. Manag.* **2021**, *249*, 114842. [CrossRef]

20. Shen, H.; Zhang, K.Q.; Hong, M.Y.; Chen, T.Y. Towards Understanding Asynchronous Advantage Actor-Critic: Convergence and Linear Speedup. *IEEE Trans. Signal Process.* **2023**, *71*, 2579–2594. [CrossRef]

21. Chen, T.; Liu, J.Q.; Li, H.; Wang, S.R.; Niu, W.J.; Tong, E.D.; Chang, L.; Chen, Q.A.; Li, G. Robustness Assessment of Asynchronous Advantage Actor-Critic Based on Dynamic Skewness and Sparseness Computation: A Parallel Computing View. *J. Comput. Sci. Technol.* **2021**, *36*, 1002–1021. [CrossRef]

22. Zhou, J.H.; Xue, Y.; Xu, D.; Li, C.X.; Zhao, W.Z. Self-learning energy management strategy for hybrid electric vehicle via curiosity-inspired asynchronous deep reinforcement learning. *Energy* **2022**, *242*, 122548. [CrossRef]

23. Holliday, J.B. Improving Asynchronous Advantage Actor Critic with a More Intelligent Exploration Strategy. Ph.D. Thesis, University of Arkansas, Fayetteville, AR, USA, 2018.

24. Chen, T.; Liu, J.Q.; Xiang, Y.X.; Niu, W.J.; Tong, E.D.; Wang, S.R.; Li, H.; Chang, L.; Li, G.; Alfred, C.Q. Adversarial retraining attack of asynchronous advantage actor-critic based pathfinding. *Int. J. Intell. Syst.* **2021**, *36*, 2323–2346.

25. Du, J.; Cheng, W.; Lu, G.; Cao, H.; Chu, X.; Zhang, Z.; Wang, J. Resource Pricing and Allocation in MEC Enabled Blockchain Systems: An A3C Deep Reinforcement Learning Approach. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 33–44. [CrossRef]

26. Zhao, X.Y.; Ding, S.F.; An, Y.X.; Jia, W.K. Applications of asynchronous deep reinforcement learning based on dynamic updating weights. *Appl. Intell.* **2019**, *49*, 581–591. [CrossRef]

27. Moghaddasi, K.; Rajabi, S.; Gharehchopogh, F.S. An enhanced asynchronous advantage actor-critic-based algorithm for performance optimization in mobile edge computing-enabled internet of vehicles networks. *Peer-to-Peer Netw. Appl.* **2024**, *17*, 1169–1189. [CrossRef]

28. Leng, J.; Fan, S.Z.; Tang, J.; Mou, H.M.; Xue, J.X.; Li, Q.D. M-A3C: A Mean-Asynchronous Advantage Actor-Critic Reinforcement Learning Method for Real-Time Gait Planning of Biped Robot. *IEEE Access* **2022**, *10*, 76523–76536. [CrossRef]

29. Sharif, A.; Marijan, D. Evaluating the Robustness of Deep Reinforcement Learning for Autonomous Policies in a Multi-Agent Urban Driving Environment. In Proceedings of the 22nd IEEE International Conference on Software Quality, Reliability and Security (QRS), Guangzhou, China, 5–9 December 2022; IEEE Computer Society: Guangzhou, China, 2022; pp. 785–796.

30. Wu, Q.; Wu, J.Q.; Shen, J.; Du, B.; Telikani, A.; Fahmideh, M.; Liang, C. Distributed agent-based deep reinforcement learning for large scale traffic signal control. *Knowl.-Based Syst.* **2022**, *241*, 108304. [CrossRef]