

A Machine Learning Approach for Evaluating Operating Parameters in Proton Exchange Membrane Fuel Cells Using Automotive Dynamic Test Data: Conceptual Consideration for Unmanned Aerial Vehicle Applications

Hasan Çınar^{1*}, Süleyman Tunçel²

¹Department of Aeronautical Engineering, Necmettin Erbakan University, Konya, 42140, Türkiye. (hasan.cinar@erbakan.edu.tr)

²Electronics and Communication Engineering, Turkish Military Academy, National Defence University, Ankara, 06420, Türkiye. (suleyman.tuncel@msu.edu.tr)

Article Info

Received: 25 August 2026
Revised: 26 September 2026
Accepted: 13 October 2026
Published Online: 19 October 2026

Keywords:

Unmanned Aerial Vehicle
Fuel cell management systems
Machine learning
Prognostics and health management
Energy management

Corresponding Author: Hasan Çınar

RESEARCH ARTICLE

<https://doi.org/10.30518/jav.2025203>

Abstract

Owing to their elevated energy density and rapid refuelling capability, proton exchange membrane fuel cells (PEMFCs) hold strong potential as a primary power source for fully electric ground and aerial vehicles, while also generating zero greenhouse gas emissions. Nevertheless, accurately estimating PEMFC state of health (SOH) continues to pose a major obstacle to their broader adoption in electric vehicle applications. Given that PEMFC health is directly linked to essential operating parameters, particularly voltage and power output, continuous monitoring is essential for ensuring safe, dependable vehicle operation. Motivated by this need, the present work implements sophisticated data-driven machine learning techniques to estimate PEMFC output power as a basis for health monitoring. In particular, CatBoost, Decision Tree, and Random Forest algorithms were applied to a publicly accessible dataset obtained under an automotive dynamic driving cycle. Findings indicate that all three algorithms deliver low prediction error when estimating output power. Beyond this, the study provides a thorough comparison of these three machine learning approaches and discusses at a conceptual level how the proposed technique could be incorporated into energy management and trajectory planning systems for PEMFC-driven unmanned aerial vehicles.

1. Introduction

Fossil fuels release significantly harmful gases into the environment and are found in limited quantities in nature (Zhou et al., 2023; C. Yang et al., 2023). In addition, energy demand in the transportation sector is constantly increasing in parallel with the growing world population (Emami Javanmard & Ghaderi, 2023). For these reasons, there is a trend towards renewable energy sources in transportation. Especially, the preference of fossil-based propulsion engines, which are non-renewable, instead of renewable ones has been a frequently encountered situation (Farinha et al., 2023; Ma et al., 2023).

Instead of engines dependent on fossil fuels, batteries and fuel cells are deployed as the main source of propulsion (Nadir et al., 2019). Solar cells can also be utilized as auxiliary sources. Among these energy resources, batteries have relatively high power densities (Y. Wang et al., 2019). However, their shortcomings can be ordered as their low energy density and long charging times (Elouarouar & Medromi, 2022). Fuel cells, on the other hand, provide high energy density compared to batteries. In addition, the high conversion efficiency of fuel cells makes them really attractive for fully electric propulsion (Lei et al., 2022; Gong & Verstraete, 2017). Among the most widely used fuel cell types

are direct methanol fuel cells (DMFC), solid oxide fuel cells (SOFC), and proton exchange membrane fuel cells (PEMFC). PEMFCs in particular have gained considerable attention in recent years within the unmanned aerial vehicle domain, largely owing to their conversion efficiency reaching as high as 60 percent, comparatively high energy density, and low operating temperature requirements (B. Wang et al., 2020; Xiao et al., 2023). Furthermore, projections suggest that PEMFCs will become the dominant propulsion technology for eVTOL vehicles in urban air mobility beyond 2040 (Kiesewetter et al., 2023).

Rather than storing energy, PEMFCs directly transform chemical energy into heat and electrical output. Their operation involves a complex, multifunctional physical process encompassing thermodynamic behavior, heat transfer, and chemical reactions. Hydrogen serves as the fuel within a PEMFC, with heat and water generated as by-products. Structurally, a typical cell is composed of flow channels, a pair of catalyst-coated electrodes, and an electrolyte layer. The power and energy densities of PEMFCs are 300-2000 W/kg and 600-1000 Wh/kg, respectively (Donateo & Çınar, 2022). The theoretical voltage that a PEMFC cell can generate is 1.229 V, which can be reduced to 0.7 V due to activation, ohmic and concentration losses (Abu Kasim et al., 2022; Çınar et al., 2023). The obstacles to commercialization of PEMFC in

the automotive and aerospace areas are issues with hydrogen storage, low dynamic response time, long start-up time, low power density, water management and degradation in long-term operation.

Both the operating conditions and the internal dynamics of the fuel cell influence the degradation behavior of PEMFCs. Factors such as membrane thinning and decomposition, platinum agglomeration and dissolution, carbon corrosion, catalyst poisoning, diffusion stress, and hydrogen impurity all contribute to a reduction in the fuel cell's operational lifespan during use. Exposure to dynamic loading conditions further accelerates this reduction in service life (Pan et al., 2020). This effect is especially pronounced in aviation applications, given that aircraft experience more variable dynamic power demands compared to ground-based vehicles (Mitici et al., 2023). Consequently, degradation prognostics for PEMFCs plays a key role in enhancing their reliability and durability, ultimately extending their operational lifetime.

Small UAVs are an especially significant example in this perspective since they increasingly use hybrid battery/fuel cell propulsion to improve mission endurance beyond the limitations of battery-only systems (Donateo & Çınar, 2022). The requirement for monitoring degradation without affecting operation has made prognostics and health management (PHM) a lively research topic for fuel cell-powered hybrid propulsion (Yue et al., 2021), and it is often incorporated in hierarchical energy management techniques for hydrogen-powered UAVs (Tian et al., 2025). While the current study does not use UAV-specific flight data, Section 4 discusses how the estimation approach presented here can be used for such applications. Note that the dataset is of a liquid cooled single PEMFC cell tested on an automotive driving cycle and hence the UAV related discussion is conceptual and needs validation with UAV specific stack and flight data.

Fuel cell degradation prognostic methods are fundamentally model-based (open-box) and data-based (closed-box). Model-based methods take advantage of the complex and difficult-to-obtain mathematical equations of the fuel cell and do not require experimental data. Their performance highly depends on the accuracy of the mathematical model, and they cannot be implemented in real-time applications due to their high computational burden. Data-based methods are based on measured data from the fuel cell without using any mathematical expressions. In other words, they relate the variables measured by sensors to the degradation of the fuel cell. They are appropriate for real time tasks with the only disadvantage being that they require a lot of data. The data-based methods perform degradation prognostics of PEMFCs using large amounts of data with machine learning algorithms, without considering model characteristics, by use of operational parameters of the PEMFC (Nagulapati et al., 2023).

In recent years, a range of machine learning-based approaches has been proposed for PEMFC degradation prognostics. Ma et al. (2018) applied Grid Long Short-Term Memory (G-LSTM) recurrent neural network (RNN) algorithms to perform degradation prediction across three distinct PEMFC types. Using 5000 hours of automotive data, Vichard et al. (2020) employed a neural network model to forecast PEMFC degradation. In another study, Z. Liu et al. (2022) leveraged machine learning techniques to predict both short-term voltage degradation and long-term durability levels of a PEMFC. Pan et al. (2020), meanwhile, introduced a hybrid degradation prediction method combining model-based and data-driven strategies. Similarly, H. Liu et al. (2019) implemented a machine learning-based hybrid approach to estimate the remaining useful life of a PEMFC, while Xie et

al. (2020) employed novel deep belief network (DBN) and extreme learning machine (ELM) algorithms for the same purpose. A review of existing machine learning-based degradation prediction studies on fuel cells reveals that most rely on the relationship between output voltage and degradation (Ming et al., 2023). Departing from this common approach, the present study instead performs degradation prediction based on output power, a variable that inherently captures both output voltage and output current.

This study aims to perform the degradation prediction of a PEMFC using various machine learning algorithms. In this context, we applied CatBoost, Decision Tree, and Random Forest algorithms on publicly available PEMFC data. By relating the fuel cell's degradation to its power output, we performed output power estimation. The main contributions of this study are:

1. Output power prediction for fuel cell degradation was performed for the first time on this dataset. Previous studies focus on voltage prediction. However, output power is important as it includes both voltage and current dynamic changes.
2. The CatBoost, Decision Tree, and Random Forest algorithms were implemented for the first time for this dataset of PEMFC under dynamic load.
3. This study provides a detailed comparison of machine learning-based prediction algorithms for PEMFC under dynamic load.
4. A conceptual energy management system is put forward, wherein power distribution between the fuel cell and battery is governed by the estimated PEMFC output power; the potential applicability of this system to PEMFC-powered UAVs is discussed at a conceptual level.

This paper's remaining structure is as follows: Section 2 introduces the dataset used, describes the applied data-driven machine learning models, and outlines the evaluation metrics. Section 3 then reports and compares the fuel cell output power prediction outcomes obtained from the CatBoost, Decision Tree, and Random Forest models. In Section 4, the relevance of these outcomes to UAV applications is examined. The study closes with Section 5, which summarizes the key findings and offers suggestions for future research.

2. Materials and Methods

This section first provides definitions of the dataset utilised, followed by a detailed explanation of the machine learning-based prediction method. Then the data partitioning and validation strategy is described and metrics for the performance of these algorithms are presented.

2.1. Dataset Description

The fuel cell dataset employed in this work was obtained from Zuo et al. (2021b), comprising dynamic durability testing outcomes alongside polarisation characterisation results for a single PEMFC unit. The dynamic load test cycle followed a harmonized European testing standard. In terms of durability, the test spans 3076 cycles, corresponding to approximately 1008 hours in total, with each individual cycle lasting 1181 seconds. Figure 1 illustrates the voltage, current, and power values recorded during the first cycle of this dataset. Based on these measurements, the PEMFC cell's dynamic power output ranges from 0 to 24 W.

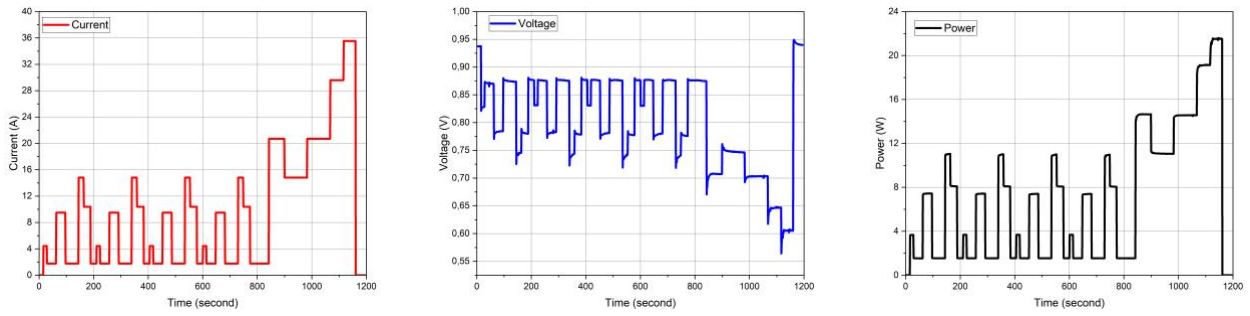


Figure 1. The dynamic load first cycle voltage, current, and power of single-cell PEMFC.

Table 1 shows the test measurement variables of the PEMFC. Accordingly, there are 13 variables, except voltage, current, and power, which are input variables. During the 10 h cycle, each of these variables was recorded. This data was obtained at the Greenlight Innovation, P17-2390, Burnaby, B.C. Canada test stand (Zuo et al., 2021a). The raw data were collected with the G20 test station software at 1 Hz by the study team.

Table 1. PEMFC test measurement variables (adapted from Nagulapati et al., 2023).

Variable	Unit
Input Variable	
Time	second
Pressure anode inlet and outlet	KPa
Pressure cathode inlet and outlet	KPa
Temp. anode inlet and outlet	°C
Temp cathode inlet and outlet	°C
Dewpoint water Temp anode	°C
Dewpoint water Temp cathode	°C
Total stack flow anode	NLPM
Total stack flow cathode	NLPM
Anode endplate Temp	°C
Output Variable	
Voltage	V
Current	A
Power	W

2.2. Data-Driven Machine Learning Algorithms

The machine learning-based fuel cell degradation prediction method consists of 4 cardinal parts: data preprocessing, data splitting, ML-based prediction, and evaluation of results. Figure 2 shows the main framework of these steps. In Nagulapati et al. (2023) and Zuo et al. (2021a), voltage estimation with Artificial Neural Networks, Gaussian Process Regression, Support Vector Machine, Recurrent Neural Network, and Long Short-Term Memory RNN algorithms was performed with the same dataset. In this study, with a different approach, power estimation is performed with machine learning-based CatBoost, Decision Tree, and Random Forest algorithms.

Data pre-processing constitutes the initial stage of the machine learning-based prediction approach. During this stage, the variables are categorized into inputs and outputs, with voltage, current, and power designated as output variables while the remaining data serves as input. Since it captures the combined variation of current and voltage, the fuel cell's output power

was selected as the target for estimation in this study. Following this, the dataset underwent a splitting process into training and testing subsets. Three different training-to-test ratios were applied: 30:70, 50:50, and 70:30, with the goal of examining how this ratio influences the resulting performance.

In this study, we decided to use CatBoost, Decision Tree, and Random Forest algorithms since they have received a lot of attention in recent years (Granado et al., 2022; Raoofi & Yildiz, 2023; Jiao et al., 2023). Then, in the last step, the error rates of the machine learning algorithms are compared.

The settings of the CB, DT, and RF models are given in Table 2. The same settings were used for all training ratios, and no Grid Search, Random Search or Bayesian optimization was applied on the final models, instead Optuna was implemented. The DT and RF models were implemented with the default parameters of the scikit-learn library except for the number of trees in RF that was set to 100. The CB model was also used with default setting and the library set the learning rate automatically to 0.124, 0.134, and 0.141 for 30%, 50%, and 70% training ratios, respectively. The random seed was set to 42 for all models to ensure reproducibility.

Table 2. Hyperparameter settings of the models used for all training ratios.

Hyperparameter	Value
CatBoost (CB)	
iterations	1000
depth	6
learning_rate	0.124 / 0.134 / 0.141
l2_leaf_reg	3
bootstrap_type / subsample	MVS / 0.8
loss_function	RMSE
Decision Tree (DT)	
criterion	squared_error
max_depth	None
min_samples_split	2
min_samples_leaf	1
Random Forest (RF)	
n_estimators	100
bootstrap	True
max_features	1.0
max_depth	None
min_samples_split	2
min_samples_leaf	1
All models	
random_state	42

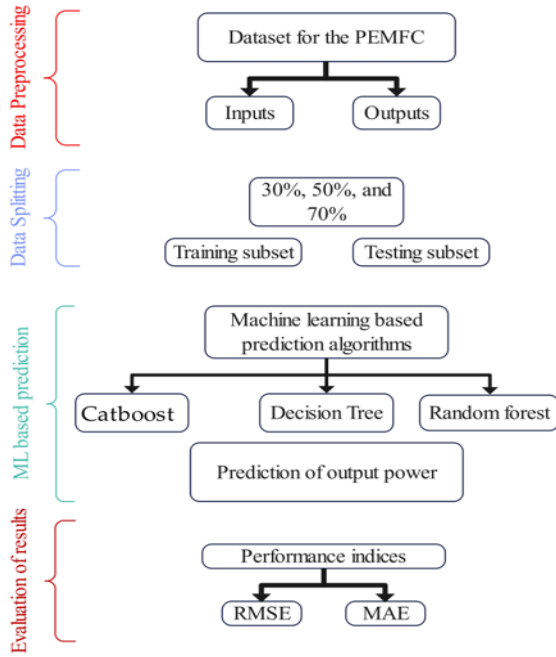


Figure 2. The main framework of machine learning-based estimation.

2.3. Data Partitioning and Validation Strategy

The data from the 20 measurement files (50 h to 1000 h) were concatenated in their acquisition order, while the original row order within each file was maintained. The time variable recorded is reset at the beginning of each file so it was used as an input feature but not for sorting the data. No random shuffling was applied and 30:70, 50:50 and 70:30 training/testing partitions were formed sequentially (i.e. all training samples precede the test samples). In addition, a three-fold expanding-window rolling-origin validation was performed. For this validation, models were trained on the first 30%, 50% and 70% of the data and tested on the following non-overlapping blocks covering 30-50%, 50-70% and 70-100% of the data. Thus, each model was trained solely on data before its test block. This prevents adjacent and highly similar time steps from appearing in both the training and test subsets, which could otherwise lead to data leakage and artificially low errors. The resulting dataset covers 1000 h of operation and contains 3,629,680 samples. Accordingly, the 30:70, 50:50, and 70:30 partitions correspond to 1,088,904/2,540,776, 1,814,840/1,814,840, and 2,540,776/1,088,904 training/test samples, respectively. The model settings in Table 2 were used for all the folds, and no hyperparameter selection was carried out on the test data. Table 3 reports the fold-wise Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) values, their means and standard deviations, and the pooled errors across all test blocks.

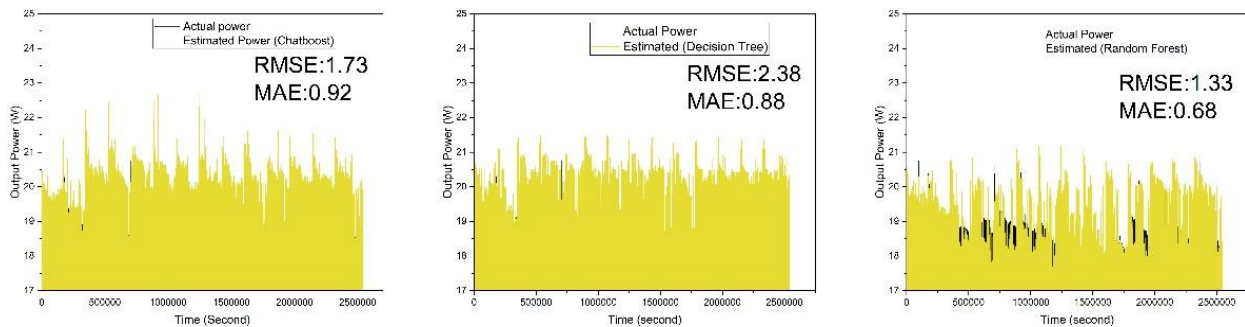


Figure 3. Predicted and observed output power for 30% of training data.

2.4. Performance Criteria

Performance assessment of the machine learning models in this work relied on two error-based indicators: RMSE and MAE. Together, these measures capture the discrepancy separating model predictions from ground-truth values and represent commonly applied benchmarks for gauging prediction accuracy in machine learning contexts. The mathematical definitions of both indicators appear in Equations (1) and (2),

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - y')^2} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - y'| \quad (2)$$

Here, y and y' represent the actual and predicted output power, respectively.

For the rolling-origin validation, the pooled RMSE and MAE were calculated over all test samples of the three folds, as given in Equations (3) and (4).

$$RMSE_{pooled} = \sqrt{\frac{1}{n} \sum_{k=1}^3 \sum_{i=1}^{n_k} (y - y')^2} \quad (1)$$

$$MAE_{pooled} = \frac{1}{n} \sum_{k=1}^3 \sum_{i=1}^{n_k} |y - y'| \quad (2)$$

Here, n_k is the number of test samples in fold k , and n is the total number of test samples in all folds. Unlike the mean of the fold-specific errors, the pooled errors weight each fold by its number of test samples.

3. Results and Discussion

Figure 3 shows the output power estimation results for the case where 30 percent of the data is used as training data. Results are presented for the entire 10-hour test. Accordingly, it was observed that the Random Forest algorithm provided the lowest error rates. The Random Forest algorithm estimated the output power with error rates of RMSE: 1.33 and MAE: 0.68, respectively.

Figure 4 shows the output power estimation results of the Random Forest algorithm for the first and last cycle when 30 percent of the data is used as training data. A cycle is 1181 seconds, where the first cycle is the first 1181 seconds, and the last cycle is the last 1181 seconds of the 10-hour data.

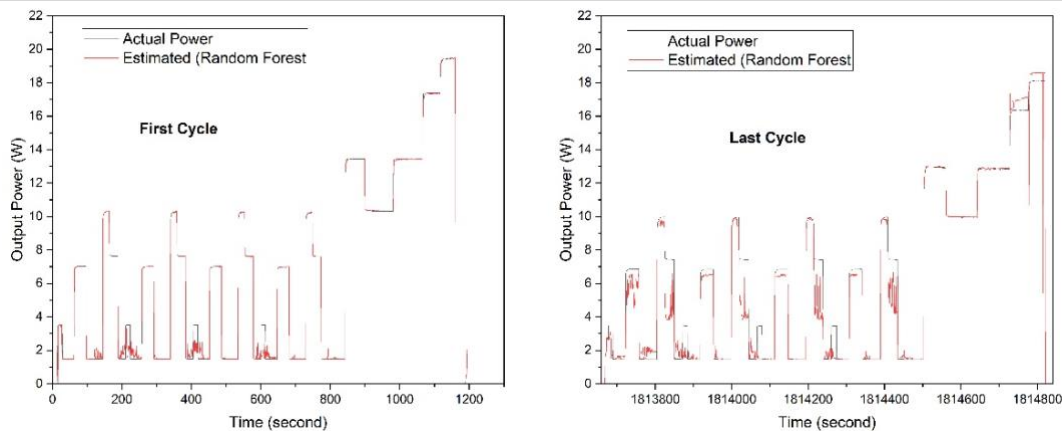


Figure 4. Predicted and observed output power of the Random Forest algorithm for the first and last cycle for 30% of training data.

Figure 5 shows the output power estimation results of the machine learning algorithms when 50 percent of the data is considered as training data. Accordingly, the RMSE error rates of the CatBoost, Decision Tree, and Random Forest algorithms were calculated as 0.59, 1.23, and 0.64, respectively. Compared to the case where 30 percent of the data was used

as training data, a decrease in RMSE error rates was naturally obtained for all 3 algorithms. As a result, the lowest error rates were achieved with the CatBoost algorithm for the case where 50 percent of the data was used as training data.

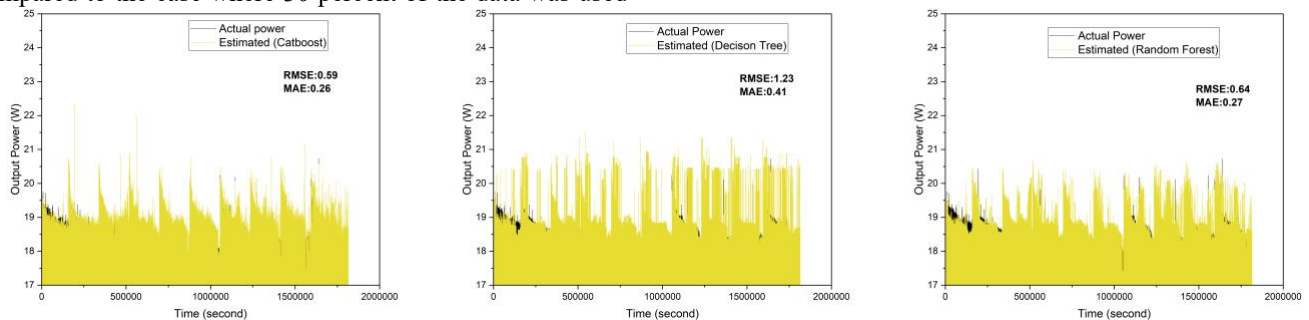


Figure 5. Predicted and observed output power for 50% of training data.

Figure 6 shows the output power estimation results of the CatBoost algorithm for the first and last cycle when 50 percent of the data is used as training data. Accordingly, it can be said that the algorithm's performance is better in the first cycle for output power estimation. Furthermore, it is found that the algorithm performs worse in areas with abrupt load fluctuations, where transient spikes and overshoots occur. This is due to the fact that tree-based models give each area of the input space a constant value, meaning that the prediction is

solely dependent on the current input and not on the PEMFC's recent operating history (Breiman, 2001; Prokhorenkova et al., 2018). Moreover, only a small portion of the dataset consists of transition samples.

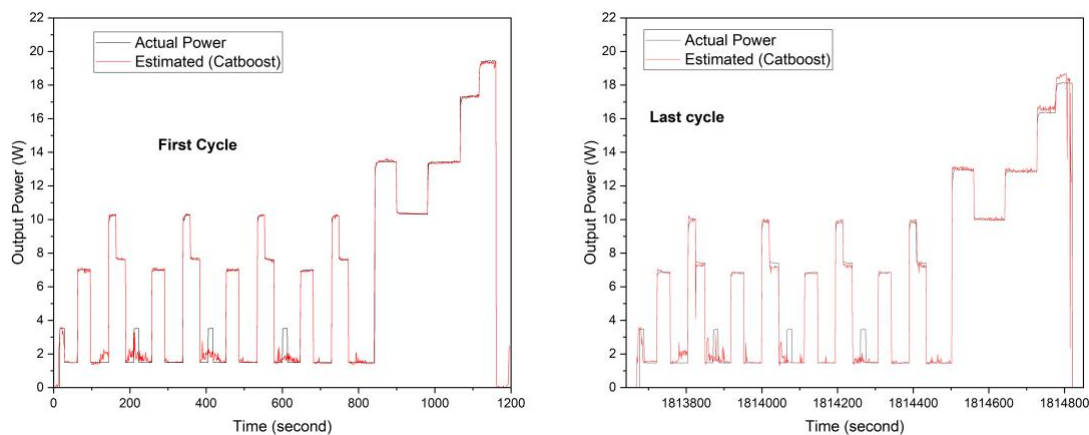


Figure 6. Predicted and observed output power of the CatBoost algorithm for the first and last cycle for 50% of training data.

Figure 7 shows the output power estimation results of the machine learning algorithms when 70 percent of the data is

used as training data. Accordingly, the RMSE error rates of the CatBoost, Decision Tree, and Random Forest algorithms are calculated as 0.64, 1.13, and 0.81, respectively.

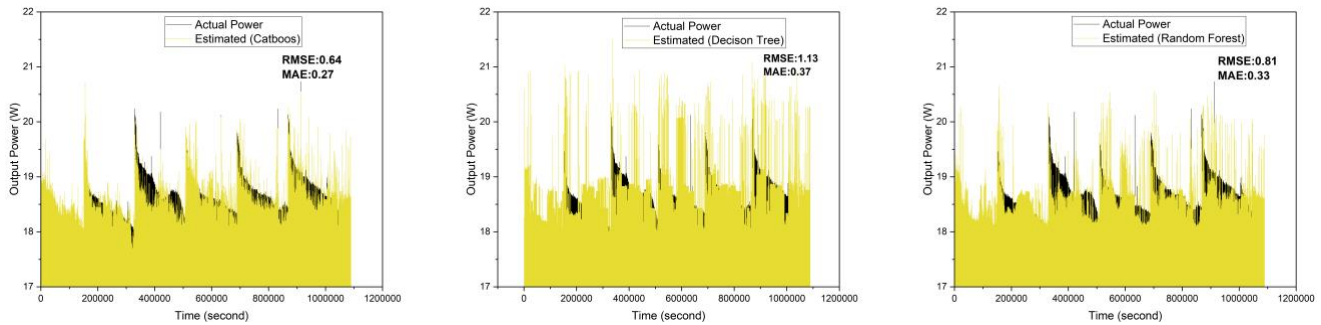


Figure 7. Predicted and observed output power for 70% of training data.

Figure 8 compares the output power estimation results for the first and last cycle for the CatBoost algorithm. Since the lowest error rate was obtained with the CatBoost algorithm when 70 percent of the data was used as training data, its results are shown in Figure 8. Deviations persist at abrupt load changes even though the larger training set lessens spurious peaks. This

may be executed in subsequent research by including rolling-window statistics, time-lagged inputs, and rates of change (like dI/dt) as features, or by fusing tree-based models with sequence models like Long Short-Term Memory (Hochreiter & Schmidhuber, 1997).

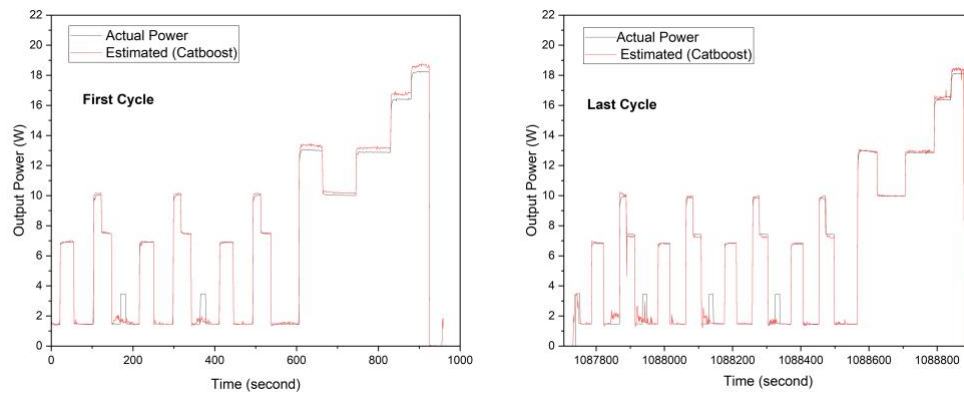


Figure 8. The actual and estimated output power of the CatBoost algorithm for the first and last cycle for 70% of training data.

Figure 9 shows the variation of the error rates of the CB, DT, and RF algorithms with the percentage of training data. Accordingly, the DT and RF algorithms have the lowest error rates when 70 percent of the data is used as training data, whereas for the CatBoost algorithm this was achieved with 50 percent training data. From 50% to 70% training data, the RMSE of CB increases slightly from 0.59 to 0.64. The dataset

was split on a chronological basis and therefore the 70% test subset contains only the latest operating cycles, and the errors for different ratios are not directly comparable. CB was also used with a fixed number of trees and depth (Table 2) and its performance saturates at around 50% training data. No clear overfitting can be concluded, since the error increase is small.

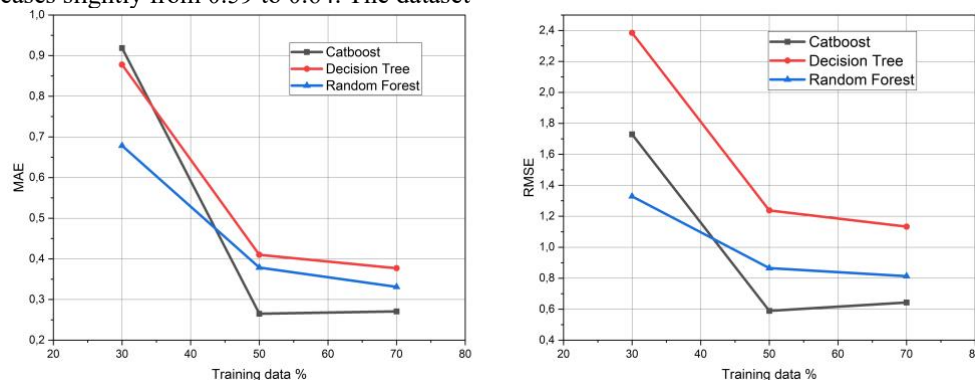


Figure 9. Performance of CB, DT, and RF algorithms on RMSE and MAE values for 30%, 50%, and 70% of the training data.

The output power of the PEMFC decreases with the increase in the number of cycles under dynamic load conditions; in other words, the fuel cell ages. For example, when the test data of the first 50 h in the raw data are analyzed, it is observed that the maximum output power is 21.59 W in the 1st cycle and 20.18 W in the 151st cycle, which is the last cycle. In other

words, the output power of the fuel cell decreased by approximately 6.5% in the first 50 hours. In this context, an opinion can be reached about the aging of the fuel cell by estimating its output power. Therefore, this paper deals with data-driven algorithms for fuel cell output power estimation.

Table 3 presents the results of the three-fold rolling-origin validation, where each model is trained only on data before its test block. In the first fold, all models got the highest errors which is in agreement with the findings in Figure 9 and is justified by the small quantity of training data (30%). As the training data increased, the second and third folds showed a noticeable decrease in errors. CB resulted in the lowest RMSE and MAE in the second and third folds, indicating that this method provides the most accurate estimations when sufficient training data are available. In contrast, RF achieved the lowest pooled errors (RMSE=0.834, MAE=0.364) and the lowest standard deviations, which indicate the most stable performance across the folds. DT had the highest pooled error among the three models. These results confirm that the models can estimate the output power in later operating periods based only on the past data.

Table 3. Results of the three-fold rolling-origin validation. Each model was trained on all data preceding its test block. The lowest value in each row is shown in bold.

Fold (test block)	CB	DT	RF
RMSE			
Fold 1 (30-50%)	1.749	1.869	1.037
Fold 2 (50-70%)	0.467	0.952	0.611
Fold 3 (70-100%)	0.643	1.133	0.812
Pooled	1.055	1.344	0.834
Mean $\pm$ SD	0.953 $\pm$ 0.695	1.318 $\pm$ 0.486	0.820 $\pm$ 0.213
MAE			
Fold 1 (30-50%)	0.829	0.642	0.504
Fold 2 (50-70%)	0.225	0.307	0.275
Fold 3 (70-100%)	0.271	0.376	0.331
Pooled	0.417	0.432	0.364
Mean $\pm$ SD	0.442 $\pm$ 0.336	0.442 $\pm$ 0.177	0.370 $\pm$ 0.120

To assess the computational burden of the models, the average training time with standard deviation, single-sample prediction latency, and model size were measured, as presented in Table 4. The training times were averaged over two to five repeated fits, and the latencies were obtained from 1000 individually timed single-sample predictions per fit after 100 warm-up predictions. All measurements were performed on the same CPU-based computer under identical conditions, using a single thread for prediction. Since the absolute values are hardware-dependent, the results should be interpreted as a relative comparison of the models. DT had the shortest training time and the lowest prediction latency, whereas CB had the smallest model size and a low latency for all training ratios. In contrast, RF required much longer training times and larger model sizes. It should be noted that the RF model with 70% training data was run with file-backed memory due to its high memory requirement; therefore, its timing values include additional storage overhead. These results show that CB provides the best trade-off between accuracy, prediction speed, and model size, whereas RF is not lightweight in this configuration.

Table 4. Training time, single-sample prediction latency, and model size of the models. The lowest value in each row is shown in bold.

Metric	CB	DT	RF
30% training data			
Average training time (s), standard deviation	32.73 $\pm$ 0.38	23.59 $\pm$ 0.26	206.78 $\pm$ 1.77
Mean latency (ms)	0.215	0.032	3.589
Median latency (ms)	0.192	0.028	3.465
95th percentile latency (ms)	0.324	0.051	4.244
Model size (MiB)	1.07	61.38	4080.23
50% training data			
Training time (s)	51.57 $\pm$ 0.55	42.95 $\pm$ 1.08	374.82 $\pm$ 9.21
Mean latency (ms)	0.206	0.032	3.766
Median latency (ms)	0.185	0.028	3.584
95th percentile latency (ms)	0.287	0.048	4.852
Model size (MiB)	1.07	102.02	–
70% training data			
Training time (s)	71.03 $\pm$ 1.09	63.39 $\pm$ 0.65	624.95 $\pm$ 1.20
Mean latency (ms)	0.206	0.033	49.797
Median latency (ms)	0.189	0.029	15.167
95th percentile latency (ms)	0.295	0.050	160.311
Model size (MiB)	1.07	144.37	9654.17

4. Implications and Limitations for UAV Applications

4.1. Motivation

The estimation findings reported in Section 3 demonstrate that tree-based machine learning models can estimate PEMFC output power with low errors during dynamic load situations, although larger deviations occur at sudden load changes. Although the dataset originates from an automotive durability test, both automotive and UAV applications involve frequent and sudden load changes; for example, in UAVs power demand can vary by a factor of three or more between cruise and hover/climb conditions (Donateo & Çınar, 2022). However, the magnitude and time scales of these load changes may differ between the two applications. Similar dynamic-power and energy-management challenges have been reviewed for hydrogen-powered hybrid UAVs more broadly (Çınar et al., 2023).

In recent years, prognostics and health management (PHM) has become an important area of research within fuel cell hybrid propulsion systems, aiming to enhance fuel cell durability via real-time health monitoring and decision-making guided by prognostic information (Yue et al., 2021). Reported UAV energy management strategies typically combine fuel cell-level control, hybrid arbitration, and mission-level planning in a hierarchical structure (Tian et al., 2025), where a predicted/forecasted state variable drives a lower-level power-split controller. This section explores how the PEMFC estimations provided here can be theoretically

incorporated into such a structure, as well as the restrictions that come with it.

4.2. A Conceptual Energy Management Framework

Figure 10 shows a three-layer conceptual EMS that employs the same prediction-decision-actuation logic as hierarchical fuel cell EMS architectures. In contrast to most similar frameworks, which are driven by expected demand or SOC trajectories, the current framework is driven by the estimated health status of the power source, following Yue et al.'s (2021) prognostic-based philosophy.

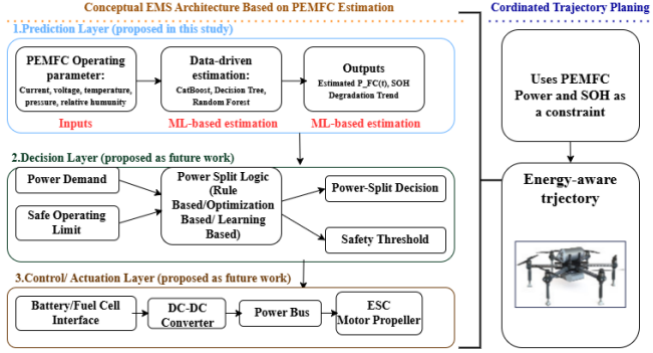


Figure 10. Conceptual three-layer EMS architecture with PEMFC power and SOH estimation.

The prediction layer is the empirically validated component of this study: PEMFC operational parameters feed the CatBoost/Decision Tree/Random Forest models, which estimate instantaneous power and the degradation trend (best result: CatBoost, RMSE=0.59, MAE=0.26, 50:50 split). The decision layer is a suggested extension that compares projected power to UAV power demand using:

$$P_{\text{battery}}(t) = P_{\text{demand}}(t) - \min(P_{\text{FC,predicted}}(t), P_{\text{FC,max,safe}}(t)) \quad (3)$$

where $P_{\text{FC,max,safe}}(t)$ decreases as estimated SOH decreases, conceptually similar to degradation-aware hierarchical EMS designs (e.g., Tümer et al., 2025), shifting load to the battery without interrupting flight for direct SOH measurement, which is otherwise impossible in operation. If the SOH becomes dangerously low, a safety threshold activates an emergency mode. The actuation layer (DC-DC setpoints, FC-battery interface, and UAV power bus) is shown only to demonstrate completeness. Only the prediction layer is empirically validated here; the decision and actuation layers are proposed for future research and are primarily based on the cited literature.

Tree-based models, such as those used here, require significantly lower computational and training resources than deep learning on tabular data (Grinsztajn et al., 2022), including in power prediction tasks specifically, and are well suited to resource-constrained embedded inference — an advantage for onboard UAV flight computers, where MPC- or RL-based EMS approaches would be more computationally demanding.

4.3. Additional Operational Considerations

- Estimate remaining endurance. A simple energy balance using the degradation trend and an assumed cruise power can demonstrate the effect of degradation on flight endurance, similar to the role of SOC/energy reference

trajectories in comparable hierarchical EMS (Howard et al., 2014).

- Pre-flight health check. A rapid, threshold-based SOH calculation could aid in making a go/no-go decision before dispatch, similar to online prognostics-based health management for FCHEVs (B. Yang et al., 2025).
- Trajectory/path-planning coordination. The safe power limit produced by the decision layer could also be shared with a trajectory or path-planning module as a power-budget constraint, following coordinated energy-management, trajectory-control architectures reported for fuel cell/battery hybrid UAVs (H. Liu et al., 2022; Tian et al., 2022; Yao et al., 2023). Rather than embedding path planning within the EMS itself, such a scheme would keep the two functions as parallel, loosely coupled modules.

These applications remain conceptual and are proposed for future validation.

4.4. Limitations

1. **Single-cell vs. stack.** A single PEMFC cell was used, not a multi-cell UAV-representative stack.
2. **Cooling architecture.** The tested cell is liquid-cooled and automotive-oriented, whereas most UAV stacks are air-cooled and open-cathode (Çınar et al., 2023).
3. **Nowcasting, not forecasting.** The models estimate current power/degradation states rather than future ones, in contrast to the prediction horizons used in the hierarchical EMS discussed in Section 4.2. Thus, the present study does not directly estimate the state of health, remaining useful life, future degradation and flight safety of the PEMFC. However, the findings of this study can be used as a basis for such estimates in future studies.
4. **No UAV-specific validation data.** No UAV flight or air-cooled stack data were used; to the authors' knowledge, there is currently no public raw dataset for air-cooled PEMFC stacks under real UAV flight conditions.

These restrictions do not detract from the main contribution, which is low-error PEMFC power estimation with lightweight models, but they clearly mark the distinction between validated results and the notional UAV applications outlined above.

5. Conclusions

PEMFCs have a relatively high conversion rate and high energy density. For these reasons, PEMFC plays an important role as an alternative to batteries in renewable-based mobility. However, PEMFC systems suffer from reliability and durability limitations. In this context, diagnostic approaches to operating parameters are important for the commercialization of PEMFCs in fully electric propulsion.

This study aims to implement a machine learning-based approach for the degradation prediction of a PEMFC under dynamic load operation. In this context, a data-driven machine learning framework is proposed to estimate the output power of a PEMFC. We consider a publicly available dataset of a PEMFC under dynamic load conditions. This dataset consists of 3076 cycles, which is around 1008 hours.

We considered CatBoost, Decision Tree, and Random Forest algorithms for output power estimation of the fuel cell. The dynamic load data of the fuel cell was divided into 30:70,

50:50, and 70:30 (ratio of training to testing) to study the effect of the training-to-testing ratio on the results. Consequently, the best estimation of output power was achieved with the CatBoost algorithm when the ratio of training data to test data was 50:50. For this case, the RMSE and MAE error rates of the CatBoost algorithm were calculated as 0.59 and 0.26, respectively.

In the continuation of this study, the authors plan to conduct feature determination by examining the effect of input variables on the prediction results. Additionally, different machine learning algorithms can be tested by applying domain adaptation to this dataset.

Funding

This research received no external funding.

Data Availability Statement

Data are contained within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- Abu Kasim, A. F. B., Chan, M. S. C., & Marek, E. J. (2022). Performance and failure analysis of a retrofitted Cessna aircraft with a Fuel Cell Power System fuelled with liquid hydrogen. *Journal of Power Sources*, 521, 230987.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Çınar, H., Kandemir, I., & Donateo, T. (2023). Current technologies and future trends of hydrogen propulsion systems in hybrid small unmanned aerial vehicles. In *Hydrogen Electric Vehicles* (pp. 75–109).
- Donateo, T., & Çınar, H. (2022). Conceptual design and sizing optimization based on minimum energy consumption of lift-cruise type eVTOL aircraft powered by battery and fuel cell for urban air mobility. *Journal of Physics: Conference Series*, 2385(1), 012072.
- Elouarouar, S., & Medromi, H. (2022). Multi-rotors unmanned aerial vehicles power supply and energy management. *E3S Web of Conferences*, 336, 00068.
- Emami Javanmard, M., & Ghaderi, S. F. (2023). Energy demand forecasting in seven sectors by an optimization model based on machine learning algorithms. *Sustainable Cities and Society*, 95, 104623.
- Farinha, J., Silva, L. M., Matlock, J., Afonso, F., & Suleman, A. (2023). Hydrogen fuel cell integration and testing in a hybrid-electric propulsion rig. *International Journal of Hydrogen Energy*, 48.
- Gong, A., & Verstraete, D. (2017). Fuel cell propulsion in small fixed-wing unmanned aerial vehicles: Current status and research needs. *International Journal of Hydrogen Energy*, 42(33), 21311–21333.
- Granado, L., Ben-Marzouk, M., Solano Saenz, E., Boukal, Y., & Jugé, S. (2022). Machine learning predictions of lithium-ion battery state-of-health for eVTOL applications. *Journal of Power Sources*, 548.
- Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems*, 35, 507–520.
- Howard, T., Pivtoraiko, M., Knepper, R. A., & Kelly, A. (2014). Model-predictive motion planning: Several key developments for autonomous mobile robots. *IEEE Robotics & Automation Magazine*, 21(1), 64–73.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Jiao, S., Zhang, G., Zhou, M., & Li, G. (2023). A comprehensive review of research hotspots on battery management systems for UAVs. *IEEE Access*, 11, 84636–84650.
- Kiesewetter, L., et al. (2023). A holistic review of the current state of research on aircraft design concepts and consideration for advanced air mobility applications. *Progress in Aerospace Sciences*, 142, 100949.
- Lei, T., Wang, Y., Jin, X., Min, Z., Zhang, X., & Zhang, X. (2022). An optimal fuzzy logic-based energy management strategy for a fuel cell/battery hybrid power unmanned aerial vehicle.
- Liu, H., Chen, J., Hissel, D., & Su, H. (2019). Remaining useful life estimation for proton exchange membrane fuel cells using a hybrid method. *Applied Energy*, 237, 910–919.
- Liu, H., Yao, Y., Wang, J., Qin, Y., & Li, T. (2022). A control architecture to coordinate energy management with trajectory tracking control for fuel cell/battery hybrid unmanned aerial vehicles. *International Journal of Hydrogen Energy*, 47(34), 15236–15253.
- Liu, Z., Xu, S., Zhao, H., & Wang, Y. (2022). Durability estimation and short-term voltage degradation forecasting of vehicle PEMFC system: Development and evaluation of machine learning models. *Applied Energy*, 326, 119975.
- Ma, R., Yang, T., Breaz, E., Li, Z., Briois, P., & Gao, F. (2018). Data-driven proton exchange membrane fuel cell degradation predication through deep learning method. *Applied Energy*, 231, 102–115.
- Ma, R., Song, J., Zhang, Y., Zhang, H., & Yuan, M. (2023). Lifetime-optimized energy management strategy for fuel cell unmanned aircraft vehicle hybrid power system. *IEEE Transactions on Industrial Electronics*, 70(9), 9046–9056.
- Ming, W., et al. (2023). A systematic review of machine learning methods applied to fuel cells in performance evaluation, durability prediction, and application monitoring. *International Journal of Hydrogen Energy*, 48(13), 5197–5228.
- Mitici, M., Hennink, B., Pavel, M., & Dong, J. (2023). Prognostics for lithium-ion batteries for electric Vertical Take-off and Landing aircraft using data-driven machine learning. *Energy and AI*, 12, 100233.
- Nadir, M., Zhou, Z., & Benbouzid, M. (2019). A critical review on unmanned aerial vehicles power supply and energy management: Solutions, strategies, and prospects. 255.
- Nagulapati, V. M., Kumar, S. S., Annadurai, V., & Lim, H. (2023). Machine learning based fault detection and state of health estimation of proton exchange membrane fuel cells. *Energy and AI*, 12, 100237.
- Pan, R., Yang, D., Wang, Y., & Chen, Z. (2020). Performance degradation prediction of proton exchange membrane fuel cell using a hybrid prognostic approach. *International Journal of Hydrogen Energy*, 45(55), 30994–31008.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with

- categorical features. *Advances in Neural Information Processing Systems*, 31.
- Raofi, T., & Yildiz, M. (2023). Comprehensive review of battery state estimation strategies using machine learning for battery management systems of aircraft propulsion batteries. *Journal of Energy Storage*, 59, 106486.
- Tian, W., Zhang, X., & Yang, D. (2022). Double-layer fuzzy adaptive NMPC coordinated control method of energy management and trajectory tracking for hybrid electric fixed wing UAVs. *International Journal of Hydrogen Energy*, 47(92), 39239–39254.
- Tian, W., Zhang, X., Zhou, P., & Guo, R. (2025). Review of energy management technologies for unmanned aerial vehicles powered by hydrogen fuel cell. *Energy*, 323, 135751.
- Tümer, B., Yıldız, D. Ş., & Arkun, Y. (2025). Real-time optimal hierarchical energy and power management system for fuel cell–battery hybrid electric vehicles. *Computers & Chemical Engineering*, 109313.
- Vichard, L., Harel, F., Ravey, A., Venet, P., & Hissel, D. (2020). Degradation prediction of PEM fuel cell based on artificial intelligence. *International Journal of Hydrogen Energy*, 45(29), 14953–14963.
- Wang, B., et al. (2020). Current technologies and challenges of applying fuel cell hybrid propulsion systems in unmanned aerial vehicles. *Progress in Aerospace Sciences*, 116, 100620.
- Wang, Y., Sun, Z., & Chen, Z. (2019). Energy management strategy for battery/supercapacitor/fuel cell hybrid source vehicles based on finite state machine. *Applied Energy*, 254, 113707.
- Xiao, C., Wang, B., Zhao, D., & Wang, C. (2023). Comprehensive investigation on lithium batteries for electric and hybrid-electric unmanned aerial vehicle applications. *Thermal Science and Engineering Progress*, 38, 101677.
- Xie, Y., Zou, J., Li, Z., Gao, F., & Peng, C. (2020). A novel deep belief network and extreme learning machine based performance degradation prediction method for proton exchange membrane fuel cell. *IEEE Access*, 8, 176661–176675.
- Yang, B., Gül, M., & Chen, Y. (2025). Comparative analysis of deep learning and tree-based models in power demand prediction: Accuracy, interpretability, and computational efficiency. *Journal of Building Physics*, 49(1), 127–169.
- Yang, C., Lu, Z., Wang, W., Li, Y., Chen, Y., & Xu, B. (2023). Energy management of hybrid electric propulsion system: Recent progress and a flying car perspective under three-dimensional transportation networks. *Green Energy and Intelligent Transportation*, 2(1), 100061.
- Yao, Y., Wang, J., Zhou, Z., Li, H., Liu, H., & Li, T. (2023). Grey Markov prediction-based hierarchical model predictive control energy management for fuel cell/battery hybrid unmanned aerial vehicles. *Energy*, 262, 125405.
- Yue, M., Al Masry, Z., Jemei, S., & Zerhouni, N. (2021). An online prognostics-based health management strategy for fuel cell hybrid electric vehicles. *International Journal of Hydrogen Energy*, 46(24), 13206–13218.
- Zhou, Y., et al. (2023). Data-driven cost-optimal energy management of postal-delivery fuel cell electric vehicle with intelligent dual-loop battery state-of-charge planner. *Energy*, 290, 130141.
- Zuo, J., et al. (2021a). Deep learning based prognostic framework towards proton exchange membrane fuel cell for automotive application. *Applied Energy*, 281, 115937.
- Zuo, J., et al. (2021b). Long-term dynamic durability test datasets for single proton exchange membrane fuel cell. *Data in Brief*, 35, 106775.

Cite this article: Cinar, H., Tuncel, S. (2026). A Machine Learning Approach for Evaluating Operating Parameters in Proton Exchange Membrane Fuel Cells Using Automotive Dynamic Test Data: Conceptual Consideration for Unmanned Aerial Vehicle Applications. *Journal of Aviation*, 10(3), 470-479.



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License

Copyright © 2026 Journal of Aviation <https://javsci.com> - <https://dergipark.org.tr/en/pub/jav>