MDPI

Article

A Hybrid Method for Performance Degradation Probability Prediction of Proton Exchange Membrane Fuel Cell

Yanyan Hu [1,2,*], Li Zhang [1], Yunpeng Jiang [3], Kaixiang Peng [4] and Zengwang Jin [5,6]

1   School of Intelligence Science and Technology, University of Science and Technology Beijing, Beijing 100083, China
2   Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China
3   SPIC Digital Technology Co., Ltd., Beijing 100080, China
4   School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
5   School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710072, China
6   National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, Northwestern Polytechnical University, Xi'an 710072, China
*   Correspondence: huyanyan@ustb.edu.cn

Abstract: The proton exchange membrane fuel cell (PEMFC) is a promising power source, but the short lifespan and high maintenance cost restrict its development and widespread application. Performance degradation prediction is an effective technique to extend the lifespan and reduce the maintenance cost of PEMFC. This paper proposed a novel hybrid method for the performance degradation prediction of PEMFC. Firstly, considering the randomness of PEMFC degradation, a Wiener process model is established to describe the degradation of the aging factor. Secondly, the unscented Kalman filter algorithm is used to estimate the degradation state of the aging factor from monitoring voltage. Then, in order to predict the degradation state of PEMFC, the transformer structure is used to capture the data characteristics and fluctuations of the aging factor. To quantify the uncertainty of the predicted results, we also add the Monte Carlo dropout technology to the transformer to obtain the confidence interval of the predicted result. Finally, the effectiveness and superiority of the proposed method are verified on the experimental datasets.

1. Introduction

The proton exchange membrane fuel cell (PEMFC) has many excellent properties: renewable, no pollution, high energy conversion rate, fast start-up speed, and low operating temperature [1,2]. It is regarded as a promising power conversion device, widely used in the military, vehicle transportation, portable power supply and many other fields [3–5]. Nevertheless, the short lifespan is one of the key barriers to its large-scale commercialization [6]. Performance degradation prediction has the ability to predict the future performance degradation trend and the failure time of PEMFC. Based on the prediction result of performance degradation, appropriate maintenance measures could be taken in advance to extend the lifetime. Consequently, performance degradation prediction is an effective way to guarantee the safe and reliable operation of PEMFC, and also directly promotes its commercial application.

The performance degradation prediction of PEMFC has yielded many fruitful results. The existing PEMFC prediction methods can be generally divided into three categories: model-based methods, data-driven methods and model and data hybrid-driven methods.

The model-based methods considered from the inside of the system usually establish mechanism models or empirical models according to the degradation mechanisms of the

equipment or the degradation trends of the monitoring data, and describe the changes in the inside of the systems with specific mathematical expressions in order to predict the failure time. Such mechanism models can complete the degradation prediction task without a large amount of online data, such as the method based on particle filter (PF) in [7]. However, due to the complexity of the internal structures of the systems, it is often difficult to establish such mechanism models of the systems. Later, scholars generally improved physical models as empirical models, where model parameters were processed through fuzzy or estimated online based on monitoring data. In [8], the unscented Kalman filter (UKF) algorithm was combined with three empirical models, including linear, logarithmic and exponential models, to approximate the aging process and thus predict the performance degradation trend. Model-based methods have the advantage of requiring less data and being able to generate analytical predictions for remaining useful life (RUL). Nevertheless, the majority of machines undergo a gradual degradation process leading up to failure, which can pose a challenge for model-based methods to accurately predict due to the intricate degradation mechanisms involved. In addition, these methods can only obtain the general trend of PEMFC degradation and cannot capture the local characteristics of degradation information well.

Different from model-based methods, data-driven methods do not need an in-depth analysis of the mechanism and law of equipment degradation but can directly predict the degradation trend of PEMFC by machine learning methods based on monitoring data, such as relevance vector machine (RVM) [9,10], echo state network (ESN) [11,12], extreme learning machine (ELM) [13,14], etc. In recent years, with the continuous development of deep learning theory, neural network models have been gradually introduced into the prediction of the degradation trend of equipment, for example, convolutional neural network (CNN) [15], long- and short-term memory (LSTM) [16–18], gated recurrent unit (GRU) [19], etc. With the continuous enhancement of the data mining ability of deep learning models, these kinds of methods gradually give up the dependence on expert knowledge, and can directly extract health indicators from the original data as the input of the neural network to learn the degradation law of the equipment and realize the degradation prediction. Although these methods can use the abundant monitoring data obtained by sensors and quickly mine the degradation information of the equipment health status, it can only obtain the point prediction result, and the uncertainty in the prediction process is often ignored.

To deal with the aforementioned constraints and integrate the strengths of both model-based and data-driven prediction methods, some scholars have proposed hybrid prediction approaches. Cheng et al. [20] proposed a hybrid prediction method combining least square support vector (LSSVM) and regularized particle filter (RPF). Liu et al. [21] used adaptive neuro-fuzzy inference system (ANFIS) to predict the long-term degradation trend of stack voltage, and then estimated the aging state of PEMFC based on the predicted results using an adaptive unscented Kalman filter (AUKF) algorithm, ultimately determining the RUL. Xie et al. [22] proposed a fusion prognostic method based on PF and LSTM; PF and LSTM independently realize the prediction of RUL, and then fuse under the given fusion principle to obtain more accurate prediction results. Ma et al. [23] integrated the extended Kalman filter (EKF) with LSTM algorithms to improve the accuracy of prediction result. EKF was used to estimate the system state to further obtain the estimated voltage, while the voltage result predicted by LSTM was used as the new observations for EKF to continue to estimate the system state. In [24], a decomposition prediction framework was proposed with consideration of the voltage recovery phenomenon. The EKF algorithm was used to predict the overall trend of voltage degradation, and LSTM was used to predict the local recovery information of voltage. Most of the existing prediction methods usually adopt the LSTM to learn the degradation trend of PEMFC. As a particular kind of recurrent neural network (RNN), LSTM has obvious superiority in time series processing and is therefore widely used in the field of degradation prediction of PEMFC. When processing time series problems, the training of LSTM is iterative. The output of LSTM at the next moment

depends on the output of the previous moment. This mode operates in a sequential manner, which will significantly impede the computational efficiency of the model. Additionally, LSTM relies on recurrent connections; it is prone to the issue of gradient vanishing, which causes the network to forget important information and become less effective at capturing long-term dependencies. The ability of the model to retain and utilize previous information may deteriorate over time, leading to suboptimal performance on lengthy input sequences.

Transformer [25] is a recently proposed network based on self-attention mechanism. This model can not only deal with the change of long-term dependence relationship over time effectively but also improve the operation efficiency through parallel computing. Compared to RNN, transformer breaks the limitation that RNN models cannot be computed in parallel. Compared with CNN, transformer can capture remote features since its field of view is not limited by the size of convolution kernel when processing temporal features. The transformer network has been widely used in computer vision (CV) and natural language processing (NLP), but it is rarely used in PEMFC degradation prediction. In view of the above problem, this paper proposes the prediction model based on transformer as the data-driven method to predict the degradation trend of PEMFC. In addition, the attention mechanism of transformer makes the neural network more focused on important time steps and has better characteristics in global information capture and long-term dependence.

With regard to the performance degradation prediction problem, as we all known that the aging process of the PEMFC is affected by many uncertain factors due to the comprehensive effect of internal structure and external environment, such as the temporal variability, measurement variability, and unit-to-unit variability. So it is more reasonable to use a stochastic process model to model the degradation of PEMFC. The Wiener process can model the non-monotone degradation process and has a clear physical explanation, which is widely used in bearing, turbofan engine and other fields. Therefore, in order to describe the uncertainty of the whole degradation process, a Wiener process model is built in this paper.

Another problem with performance degradation prediction is that although data-driven methods can accurately predict the degradation trend, the uncertainty quantification and reliability of degradation prediction are mostly ignored, and only the point prediction result can be obtained. Point prediction is only a deterministic value, which is difficult to provide sufficient guidance for equipment maintenance in practice. At present, the methods of quantitative uncertainty prediction mainly include Bayesian estimation, Gaussian process and hidden Markov method. The Bayesian neural network introduces the probability distribution to network parameters and trains the model by approximating the posterior distribution based on variational inference technique. Although this method can obtain the uncertainty of the network model, it needs to optimize the mean and standard error of the probability distribution at the same time. Many extra parameters are introduced that increase the training time [26]. Applying dropout to neural networks is equivalent to Bayesian variational inference, and an approximate uncertainty representation similar to Bayesian variational inference can be obtained by sampling techniques [27]. Therefore, the uncertainty of prediction result is quantified using the Monte Carlo dropout (MC-dropout) technique in this paper.

According to the above descriptions, a hybrid driven method for performance degradation prediction of PEMFC is proposed in this paper. Firstly, a linear Wiener process considering multiple variability sources is constructed to adaptively estimate the degradation states based on the UKF algorithm and expectation maximization (EM) algorithm. Secondly, the transformer network is used to learn the degradation characteristics and predict the long-term degradation trend of PEMFC. The MC-dropout technology is added to the network to quantify the uncertainty of the predicted results and obtain a 95% confidence interval. Finally, the proposed method is verified on experimental datasets under both static and dynamic conditions and compared with the classic LSTM method. Experimental results show that the proposed method can achieve satisfactory prediction performance under both scenarios and has fewer prediction errors compared with LSTM, which illustrates

the effectiveness and superiority of the proposed method. The main contributions of this paper are summarized as follows:

(1) With consideration of the uncertainties in the degradation process, a Wiener process model is established to describe the overall degradation trend of PEMFC and multiple kinds of variability sources are adequately considered in the model.

(2) To overcome the disadvantage of LSTM in parallel processing, a degradation prediction model is established by transformer, which is used to predict the degradation trend and capture the local fluctuation information.

(3) The MC-dropout is added in transformer network to quantify the uncertainty of the prediction results in order to provide more effective decision support for practical engineering applications.

The organization of this paper is as follows. The Wiener process model is established in Section 2. In Section 3, the degradation state estimation and model parameter estimation methods are introduced. Section 4 introduces the transformer network with MC-dropout technology. Section 5 provides the experimental results to demonstrate the proposed method. Section 6 concludes this paper.

2. Degradation Modeling of PEMFC

The voltage of PEMFC is one of the quantities that can be easily measured during its long-term operation and a semi-empirical voltage model [28] of PEMFC can be expressed as

$$V_{st} = n_{cell}\left(E_0 - iR - AT\ln\left(\frac{i}{i_0}\right) - BT\ln\left(1 - \frac{i}{i_L}\right)\right) \tag{1}$$

where V_{st} is the output stack voltage, n_{cell} is the cell number, A is the Tafel constant and B is the concentration constant, respectively. The stack current i and operating temperature T can be measured by sensors during the operation of PEMFC. The open circuit voltage E_0 and the exchange current i_0 can be identified by genetic algorithm [29] and Levenberg–Marquardt method [30].

It is demonstrated in [30,31] that the resistance R and the current density i_L are the only two parameters which have significant changes during the degradation of PEMFC, no matter whether the load is constant or dynamic. E_0 and i_0 can be considered constant without obvious changes. In addition, studies have found that the variations of R and i_L are quite similar and highly correlated with each other [28]. In this paper, a single variable α is adopted to describe the change trend of aging parameters R and i_L:

$$R(t) = R_0(1 + \alpha(t)) \tag{2}$$

$$i_L(t) = i_{L_0}(1 - \alpha(t)) \tag{3}$$

where R_0 and i_{L_0} are initial values. α represents the change of aging parameters, which can reflect the inner degradation state of PEMFC. The larger the value of α, the more severe the degradation of PEMFC.

Considering that the degradation of PEMFC is stochastic, we use a Wiener process to model the degradation trend of aging factor α:

$$\alpha(t) = \alpha(0) + \theta t + \sigma B(t) \tag{4}$$

where θ is the drift coefficient denoting the degradation rate of α, $\sigma > 0$ is the diffusion coefficient. Standard Brownian motion $\{B(t), t > 0\}$ describes the stochastic dynamics of degradation process and the temporal variability of degradation process with time. $\alpha(0)$ is the initial degradation state. Here, θ is taken as a stochastic variable to describe the unit-to-unit variability of the degradation process with $\theta \sim N(\mu_\theta, \delta_\theta^2)$, where μ_θ is the averaged degradation rate with δ_θ^2 as its variance. In order to describe the change of drift coefficient over time, its updating mechanism over time is established as follows:

$$\theta(t) = \theta(t-1) + v_t \tag{5}$$

where $v_t \sim N(0, v^2)$.

3. State of Health Estimation and Parameter Estimation

Based on the above description, the aging state estimation of PEMFC can be expressed as a non-linear system:

$$\begin{cases} \mathbf{X}_k = \mathbf{A}\mathbf{X}_{k-1} + \mathbf{W}_k \\ V_k = h(\mathbf{X}_k) + \Gamma_k \end{cases} \tag{6}$$

$$\mathbf{X}_k = \begin{bmatrix} \alpha_k \\ \theta_k \end{bmatrix}, \mathbf{A} = \begin{bmatrix} 1 & t_k - t_{k-1} \\ 0 & 1 \end{bmatrix},$$

$$\mathbf{W}_k = \begin{bmatrix} \sigma[B(t_k) - B(t_{k-1})] \\ v_k \end{bmatrix}, \mathbf{\Phi}_k = \begin{bmatrix} \sigma^2(t_k - t_{k-1}) & 0 \\ 0 & v^2 \end{bmatrix}.$$

$$h(\mathbf{X}_k) = n_{cell}(E_0 - i_k R_0(1 + \alpha_k) - AT\ln(\frac{i_k}{i_0}) - BT\ln(1 - \frac{i_k}{i_{L0}(1 - \alpha_k)})) \tag{7}$$

where $\mathbf{X}_k$ represents the aging state of PEMFC, V_k represents the stack voltage, $\mathbf{W}_k$ represents the process noise with the variances of $\mathbf{\Phi}_k$, and $\mathbf{\Gamma}_k$ represents the measurement noise with the variances of R that can reflect the measurement variability.

UKF is a non-linear filtering algorithm [32]. According to the symmetric sampling rule, UKF obtains a set of sigma points with different weights by using unscented transform to approximate the probability density function of the Gaussian distribution state, which reflects the probability density distribution of the transformation state of the system. UKF avoids the process of finding the inverse matrix. Compared with EKF, it improves the estimation accuracy and reduces the calculation time. According to (6), the degradation state of PEMFC is estimated by UKF algorithm in this paper. The procedures of the UKF algorithm are shown in Algorithm 1.

Algorithm 1 The procedures of the UKF algorithm

---

1. Initialization (k = 0) :

$\hat{\mathbf{X}}_0 = \mathbf{E}(\mathbf{X_0})$, $\mathbf{P}_0 = \mathbf{E}[(\mathbf{X}_0 - \hat{\mathbf{X}}_0)(\mathbf{X}_0 - \hat{\mathbf{X}}_0)^T]$

2. Time update:

$\hat{\mathbf{X}}_{k|k-1} = \mathbf{A}\hat{\mathbf{X}}_{k-1|k-1}$, $\quad \mathbf{P}_{k|k-1} = \mathbf{A}\mathbf{P}_{k-1|k-1}\mathbf{A}^T + \mathbf{\Phi}$

3. Sigma points and weights calculation:

(1) $\{\mathbf{X}_{k|k-1}^{(0)} = \hat{\mathbf{X}}_{k|k-1}, i = 0; \mathbf{X}_{k|k-1}^{(i)} = \hat{\mathbf{X}}_{k|k-1} + (\sqrt{(n+\lambda)\mathbf{P}})_i, i = 1\ldots n; \mathbf{X}_{k|k-1}^{(i)} = \hat{\mathbf{X}}_{k|k-1} - (\sqrt{(n+\lambda)\mathbf{P}})_i, i = n+1\ldots 2n$

where $(\mathbf{P})^T(\mathbf{P}) = \mathbf{P}$, $(\mathbf{P})_i$ is the ith column of the square root of the matrix $\mathbf{P}$.

(2) $w_m^{(0)} = \lambda / n + \lambda; w_c^{(0)} = \lambda / n + \lambda + (1 - \tau^2 + \beta); w_m^{(i)} = w_c^{(i)} = \lambda / 2(n + \lambda), i = 1 \ldots 2n$, where $\lambda = \tau^2(n + \kappa)$ is the scaling parameter, and the other parameters are generally set to $\tau^2 = 0.01, \beta = 0, (n + \lambda)\mathbf{P} = 3$

4. Measurement update

$V_{k|k-1}^{(i)} = h[\mathbf{X}_{k|k-1}^{(i)}]$

$\bar{V}_{k|k-1} = \sum_{i=0}^{2n} w_m^{(i)} V_{k|k-1}^{(i)}$

$P_{V_k V_k} = \sum_{i=0}^{2n} w_c^{(i)} [V_{k|k-1}^{(i)} - \bar{V}_{k|k-1}][V_{k|k-1}^{(i)} - \bar{V}_{k|k-1}]^T + R$

$\mathbf{P}_{\mathbf{X}_k V_k} = \sum_{i=0}^{2n} w_c^{(i)} [\mathbf{X}_{k|k-1}^{(i)} - \hat{\mathbf{X}}_{k|k-1}][V_{k|k-1}^{(i)} - \bar{V}_{k|k-1}]^T$

$\mathbf{K}_k = \mathbf{P}_{\mathbf{X}_k V_k} P_{V_k V_k}^{-1}$

$\hat{\mathbf{X}}_{k|k} = \hat{\mathbf{X}}_{k|k-1} + \mathbf{K}_k[V_k - \bar{V}_k]$

$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - P_{V_k V_k}\mathbf{K}_k\mathbf{K}_k^T$

The degradation model is initialized after model parameter μ_θ, δ_θ^2, σ^2, v^2, R are estimated. Aiming at the problem of unknown parameter estimation in the adaptive state space model based on UKF, the EM algorithm is used to estimate these parameters by using the voltage data from the initial time to the current time. The parameters are denoted by $\Theta = (\mu_\theta, \delta_\theta^2, \sigma^2, v^2, R)$. It should be noted that the essence of the EM algorithm is to approximate the maximum likelihood parameters estimation by maximizing the joint likelihood function under the premise of the posterior probability. The unknown parameters of the model can be estimated by referring to the method in reference [33].

4. Method of Degradation Prediction

4.1. Problem Description

The degradation of PEMFC is a long-term time accumulation process, so it is necessary to extract the long-term dependence of degradation data to predict the performance degradation trend of PEMFC. The transformer network solves the problem of long-term RNN dependency and can handle long sequences of inputs. In order to predict the degradation trend of PEMFC, we directly use the aging factor α filtered by UKF as a health index and predict it to evaluate the degradation degree.

The performance degradation prediction problem can be formally defined as follows. The input is $\alpha_t \in \Re^k$, $t = (1, 2, \ldots, T)$, where T is the length of α. The output is the label Y_t for each time step prediction. The purpose is to establish a mapping relationship between input data α and label Y_t to predict a real-time label, which can be expressed as

$$Y_t = f(\alpha_t) \tag{8}$$

where f represents the mapping function, and Y_t represents the real-time label predicted by f. In this paper, we use the transformer structure to establish the mapping.

4.2. The Transformer Structure

The network structure of transformer is composed of the attention mechanism. Instead of the sequential input of RNN, transformer feeds all of a sequence of data into the network at once for calculation, which brings good parallelism to the network model. Therefore, considering this advantage of transformer, we try to realize the performance degradation prediction based on the transformer network. The network structure is shown in Figure 1. The network consists of input, encoder, decoder and output.
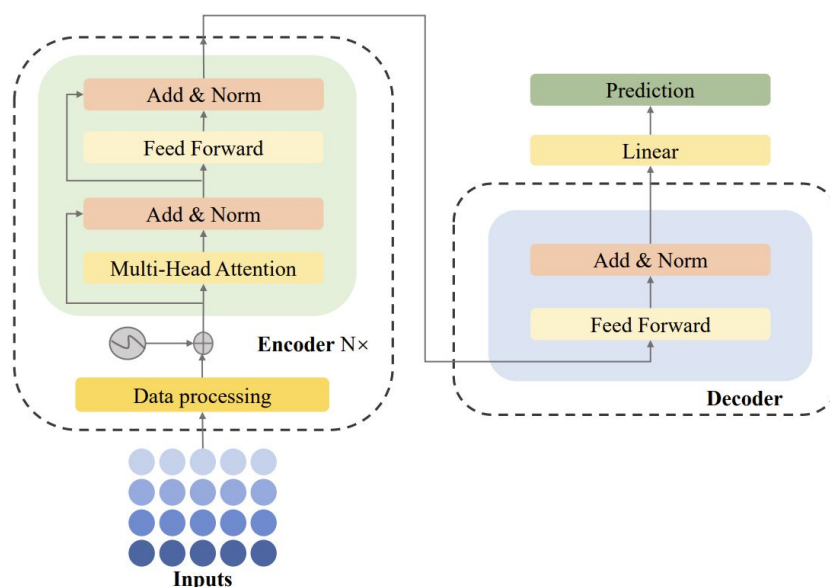

Figure 1. The structure of transformer network.

4.2.1. Data Input

Since transformer abandons the RNN structure and cannot adequately capture the sequential information of the feature sequences, position encoding is added before input. Therefore, the relative position information is added to the feature sequence by the position encoding technique and then is sent to an encoder to learn how the data points relate to each other and how important each time step is, resulting in more global characteristics. This paper adopts the most commonly used position encoding methods in transformers, namely sine position encoding and cosine position encoding [25]:

$$
\begin{cases}
P(t, 2i) = sin\dfrac{t}{10{,}000^{\frac{2i}{d}}}, \\
P(t, 2i + 1) = cos\dfrac{t}{10{,}000^{\frac{2i}{d}}}
\end{cases}
\tag{9}
$$

4.2.2. Encoder

The encoder module consists of N encoders. Each encoder consists of three parts, including multi-head attention mechanism, feedforward neural network, residual connection and normalization layer.

The encoder is responsible for projecting the input feature sequence into a vector space that contains information about the input features through an embedding layer. The core of encoder is the self-attention mechanism. The self-attention mechanism learns the correlation between each sample in the input sequence and all other samples in the sequence, and calculates the similarity between vectors to capture internal dependencies, thus solving the problem of capturing long-term dependencies. The purpose of the self-attention mechanism is to screen out important information from the input feature sequence and assign greater weight to indicate the importance of capturing information, allowing the model to focus on more important information and improve the quality of output.

The self-attention mechanism uses the scaled dot product attention to calculate the attention value of the feature matrix. The query matrix and the key matrix are dotted with scaling and softmax normalization to calculate the weight of each value matrix, and the output attention is a weighted summation of the value matrix:

$$
Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = soft(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}})\mathbf{V}
\tag{10}
$$

$$
\begin{cases}
\mathbf{Q} = \alpha_t \mathbf{W}_Q \\
\mathbf{K} = \alpha_t \mathbf{W}_K \\
\mathbf{V} = \alpha_t \mathbf{W}_V
\end{cases}
\tag{11}
$$

where the $\mathbf{Q}$ is query matrix, $\mathbf{K}$ is the key matrix, and $\mathbf{V}$ is the value matrix. They are obtained by multiplying the input characteristic matrix α_t with the corresponding weight matrix $\mathbf{W}_Q$, $\mathbf{W}_K$ and $\mathbf{W}_V$ respectively, and d is the dimension of $\mathbf{Q}$, $\mathbf{K}$ and $\mathbf{V}$.
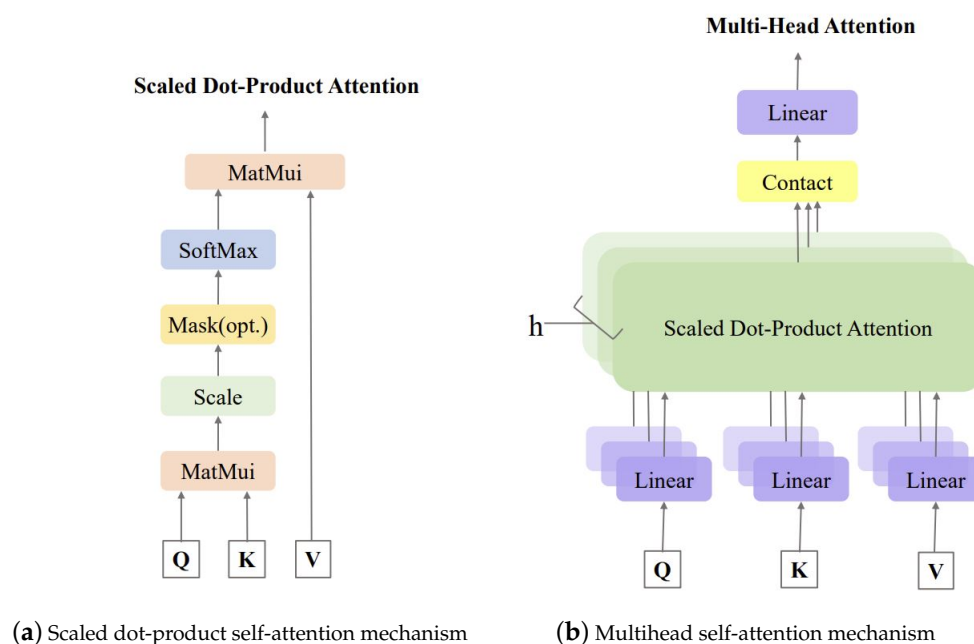
The self-attention mechanism can enable the network model to focus more on important feature information in the input sequence, while a single attention mechanism can only learn related information in one subspace. The multi-head self-attention mechanism can learn and focus on information from multiple subspaces, thereby obtaining richer features and information contained in the input sequence. The multi-head self-attention mechanism consists of several parallel and independent self-attention mechanisms. By concatenating and linearly transforming multiple attention values, the final attention value can be obtained, achieving multi-level and multi-perspective capture of the input sequence features. Figure 2 visualizes multihead self-attention, which can be expressed as

$$
MultiHead(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = concat(head_1, head_2, \ldots, head_h)\mathbf{W}
\tag{12}
$$

$$
head_i = Attention(\alpha_t \mathbf{W}_i^Q, \alpha_t \mathbf{W}_i^K, \alpha_t \mathbf{W}_i^V)
\tag{13}
$$

where $\mathbf{W}_i^Q$, $\mathbf{W}_i^K$, $\mathbf{W}_i^V$ are the weight matrix of the *i*th self-attention head, *h* is the number of heads of self-attention, and the contact function is used to concatenate the output value of each head's attention calculated.

The feedforward neural network reduces the effect of long time sequence on the fitting degree of attention mechanism. The addition of residual structure can be used to prevent network degradation. Adding layerNorm can make the input of neurons in each layer have the same mean and variance, accelerate convergence and improve running speed.

Scaled Dot-Product Attention

Multi-Head Attention

(**a**) Scaled dot-product self-attention mechanism

(**b**) Multihead self-attention mechanism

Figure 2. Process of multihead self-attention.

4.2.3. Decoder

The decoder is responsible for mapping the intermediate vector output by the encoder into an output sequence. The decoder module also consists of *N* decoders, each of which should include the mask multi-head self-attention mechanism, multi-head attention mechanism, feedforward neural network, residual connection and normalization layer. Unlike machine translation tasks, time series prediction does not require a layer of masked multiple attention, so the decoder layer is changed to a fully connected layer.

4.2.4. Output

Since the input is numerical data, there is no need for a softmax function. After, the decoder directly outputs the prediction result.

4.3. The Method of MC-Dropout

The method based on deep learning can only provide the point estimation of degradation prediction as the prediction result, and cannot provide the confidence interval. The neural network will give a good prediction result in most cases after training, but will occasionally give a poor result. This kind of point estimate cannot quantify the deviation between the predicted value and the real value, which is of little significance to the application of PEMFC in practical engineering and has limited support for maintenance decisions. Therefore, it is not enough to calculate only the point estimation results of the degradation prediction, so it is necessary to obtain the probability interval of the prediction results.

There are two uncertainties in deep learning algorithms, cognitive uncertainty and accidental uncertainty. Cognitive uncertainty refers to the inability to obtain accurate models due to limited observations in practical applications. Accidental uncertainty usually refers to the uncertainty of the sensor measurement. With the consideration of these

uncertainty sources, MC-dropout was proposed by Gal et al. [27]. The method of MC-dropout means that if the dropout is turned on during the test, the uncertain prediction results can be obtained by using the method of multiple Monte Carlo sampling. The number of neurons will be randomly discarded or adjusted according to the weight parameter, and the prediction results of each time will be different, so the probability distribution of the predicted results can be obtained.

Specifically, the Bayesian neural network is approximated by the dropout, and the probability distribution is approximated by Monte Carlo sampling to establish the uncertainty of prediction. During the training process, the data of α are selected to train the transformer network, and the dropout is turned on to effectively avoid the occurrence of the over-fitting phenomenon. Turn on the dropout during the prediction process, and the neurons in the hidden layer are temporarily dropped from the network according to a certain probability, making the predicted results different each time. A number of different point prediction results can be obtained by Monte Carlo sampling. The neural network model trained with dropout is iterated n times to obtain the mean and standard error of the predicted results n times. By utilizing the mean and variance, an approximation of the probability distribution for the uncertainty in degradation prediction can be obtained.

4.4. The Hybrid Prediction Method for Performance Degradation

The hybrid method for the performance degradation prediction of PEMFC are shown in Figure 3. It includes three parts: data acquisition, degradation modeling based on the model-based method and degradation prediction based on the data-driven method. Firstly, the data monitored by the sensor are processed to obtain the voltage data. Then, a Wiener process model for aging factor α is established to characterize the degradation of PEMFC. The degradation state of PEMFC is estimated by the UKF algorithm and EM algorithm, and the estimation results of the unknown parameters of the model are obtained. Finally, the aging factor α obtained after filtering is divided into the training set and test set, and the degradation trend of PEMFC is predicted in the long term through the transformer network.
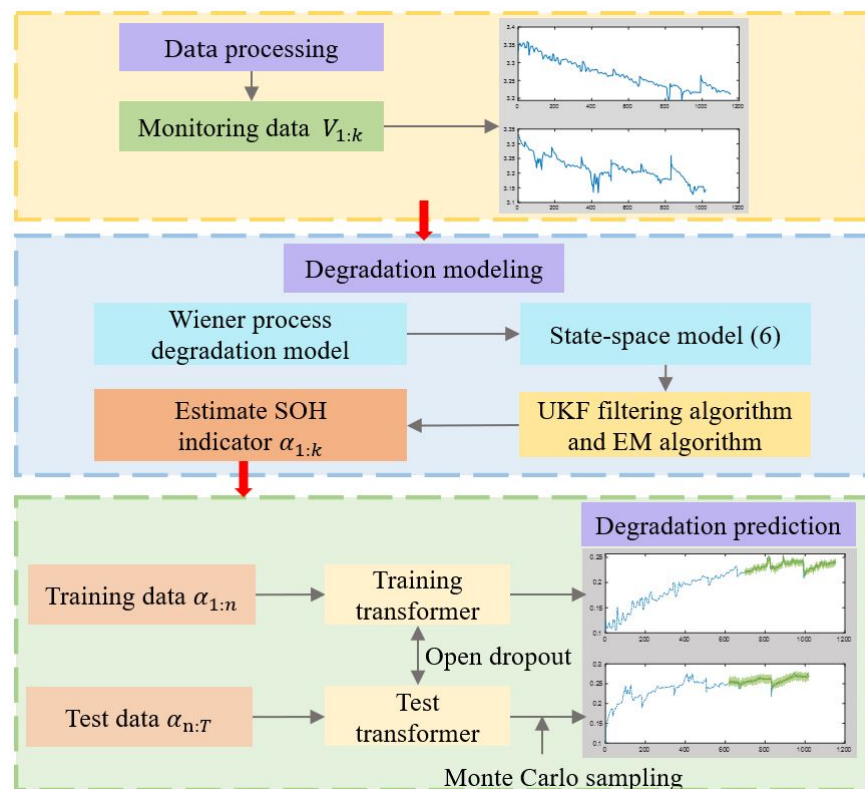

Figure 3. The hybrid method for degradation of PEMFC.

On the premise of knowing the voltage failure threshold of PEMFC, the failure threshold of the aging factor corresponding to the failure time can be determined, and the current health state of the PEMFC can be evaluated based on the prediction results of the neural network. The RUL also can be estimated by evaluating the time until the aging factor reaches the failure threshold. According to the estimated RUL, the appropriate maintenance measures can be taken in advance to reduce operating and maintenance costs and extend the lifespan of the PEMFC. At the same time, the RUL prediction results can serve as an important reference for system design and optimization.

5. Experimental Study

5.1. Experimental Dataset

The experimental datasets are obtained from the IEEE PHM Data Challenge 2014 provided by the French Federation for Fuel Cell Research FCLAB [34]. The data consist of two parts: FC1 and FC2, which are obtained under static and dynamic operating conditions, respectively. In order to ensure the stability of PEMFC operation, the flow rate of hydrogen and oxygen inlet and outlet are monitored in real time during testing, in order to avoid the instability of the output caused by too low or too high flow rates. The polarization curve test and electrochemical impedance spectroscopy (EIS) test of FC1 and FC2 are also conducted weekly to evaluate the performance status and ensure the stable operation of FC1 and FC2. More detailed information can be found in the description file of this dataset [34].

The proposed method is applied to the above experimental data, both the static dataset FC1 and the dynamic dataset FC2. The parameters of the transformer network are listed in Table 1.

Table 1. Parameters of transformer.

Parameters Setting	Value
Window length	10
Batch size	50
Epochs	50
Dropout	0.1
Multi-Head	10
Learning rate	0.001

5.2. State Estimation

The initial values of the state $\mathbf{X}_0$, covariance matrix $\mathbf{P}_0$, process error covariance matrix $\mathbf{\Phi}$, and measurement error covariance matrix R are set as follows: $\mathbf{X}_0 = [0.1, 0.002]^T$, $\mathbf{P}_0 = [10^{(-4)}, 0; 0, 10^{(-4)}]$, $\mathbf{\Phi} = [10^{(-8)}, 0; 0, 10^{(-8)}]$, $R = 10^{(-3)}$. The estimated results of aging factor α and drift rate θ of FC1 and FC2 are shown in the sub-figures in Figure 4a,b, respectively. The parameter α represents the degradation state, and a higher value means a more serious degradation degree of PEMFC. The parameter θ represents the degradation rate of PEMFC, and a larger value means a faster degradation rate. It can be seen that the aging factor α increases with time, and the drift rate θ fluctuates slightly.

Figure 5 shows the voltage fitting results after taking the aging factor α filtered by UKF into the semi-empirical aging model. The fitting voltage curves can accurately track the voltage degradation trend of PEMFC.
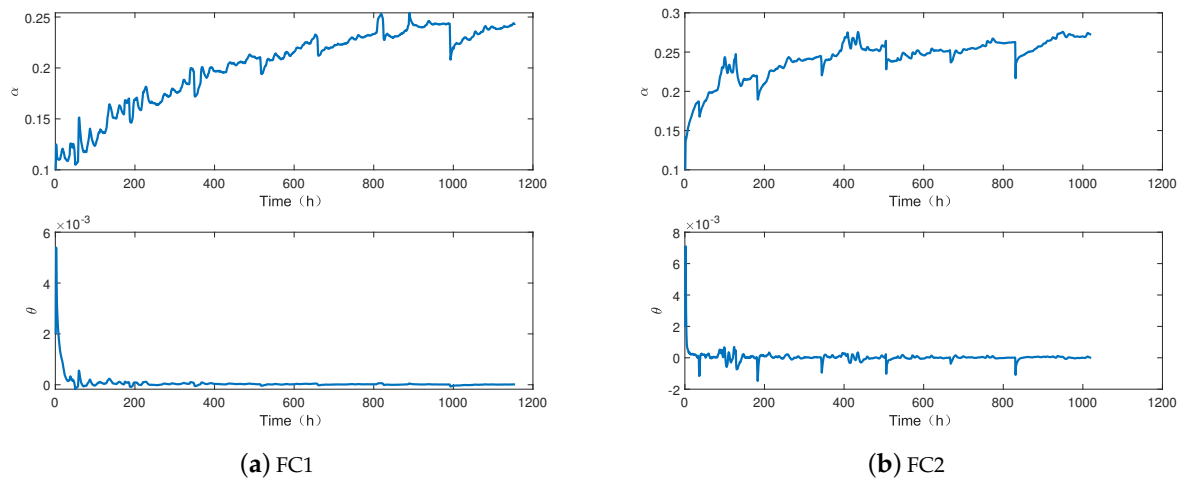
(**a**) FC1   (**b**) FC2

Figure 4. The results of state estimation.

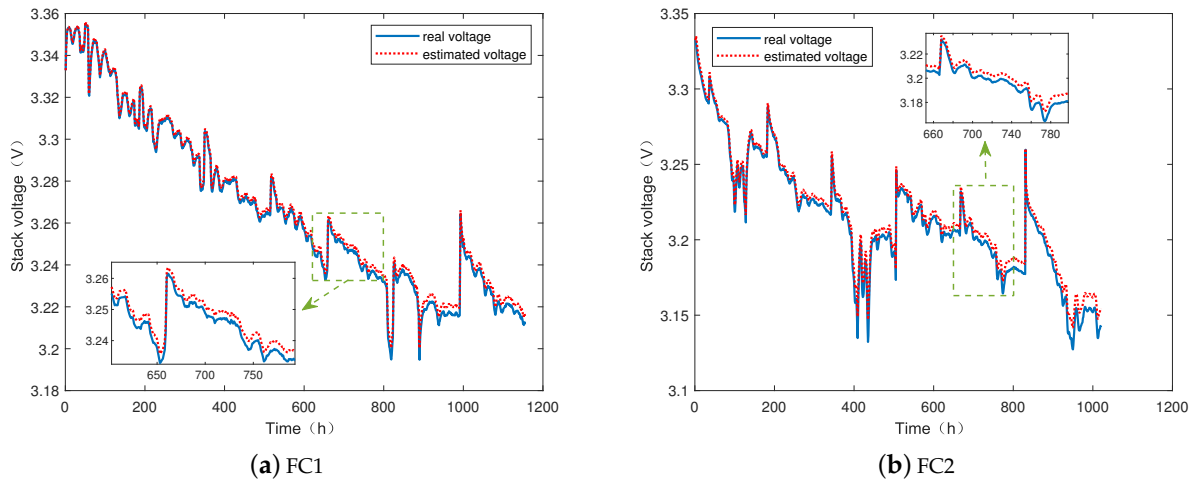(**a**) FC1   (**b**) FC2

Figure 5. The fitting voltage curves.

5.3. Performance Degradation Prediction Results

After estimating the aging state of PEMFC, the performance degradation prediction can be carried out by transformer with MC-dropout. To evaluate the efficacy of the MC-dropout approach, evaluation metrics, such as mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE) are used:

$$MAPE = \frac{1}{k} \sum_{1}^{k} \frac{|Y_t - \hat{Y}_t|}{|Y_t|} \tag{14}$$

$$RMSE = \sqrt{\frac{\sum_{1}^{k}(Y_t - \hat{Y}_t)^2}{k}} \tag{15}$$

$$MAE = \frac{\sum_{1}^{k}|Y_t - \hat{Y}_t|}{k} \tag{16}$$

Figure 6 shows the degradation prediction results of FC1 with different training set sizes (40%, 50%, and 60%). The output predicted value is the average of 100 iterations of the MC-dropout method in the optimal model after 50 epochs training. At the same time, the confidence interval made by the standard error obtained from MC-dropout is given.

Figure 7 and Table 2 show the prediction performance of FC1 under different training set sizes. It can be seen from the figure that with the increase in training set data, the prediction performance gradually becomes better, and the three performance evaluation indexes MAPE, MAE and RMSE also decrease gradually. Since the prediction network can extract more information from more training data, it helps the subsequent network prediction task.
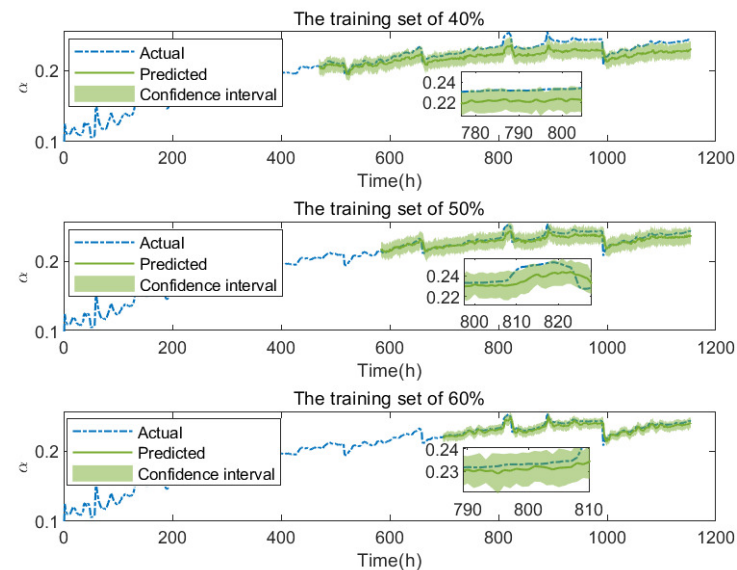
Figure 6. The degradation prediction of FC1 with different training set sizes.
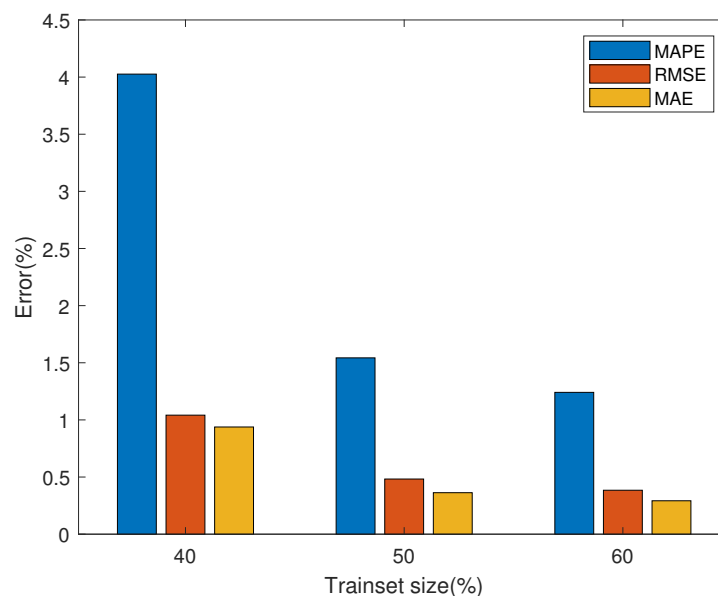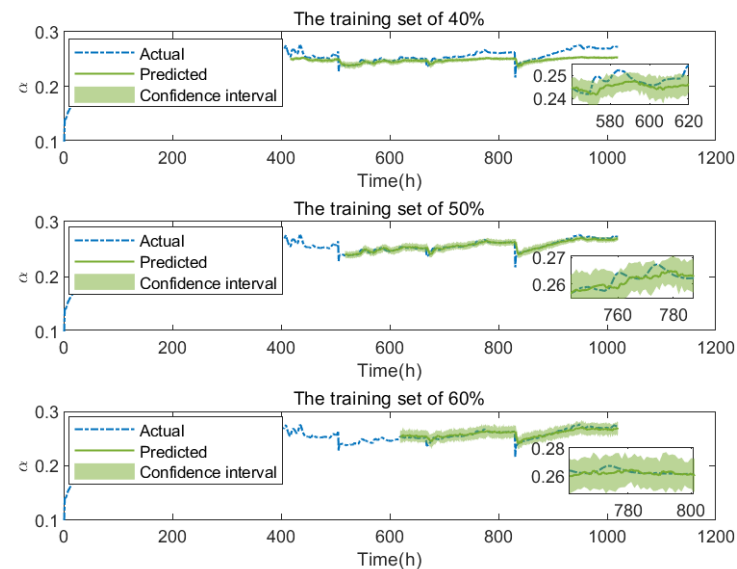

Figure 7. Error of FC1 for degradation prediction with different training set sizes.

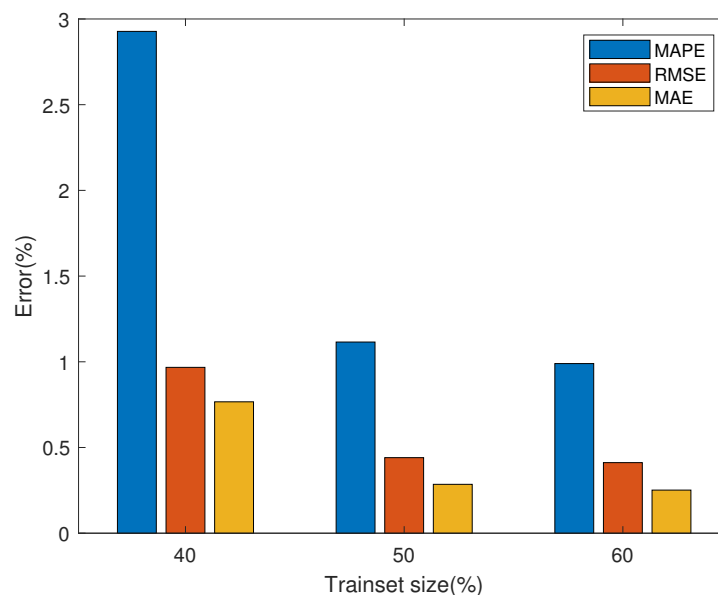**Table 2.** Prediction performance with different training set sizes for FC1.

Training Set	MAPE (%)	RMSE (%)	MAE (%)
40%	4.0261	1.0416	0.9380
50%	1.5429	0.4826	0.3630
60%	1.2410	0.3845	0.2923

The prediction results of FC2 under dynamic conditions are shown in Figure 8. Figure 9 and Table 3 give three evaluation indexes under different training set sizes. The predicted

mean also fits the real value fairly well. It turns out that the average of all predictions provides a reliable estimate of the final prediction, whereas the standard error can serve as an indicator of the uncertainty of the model with respect to the input data.

Figure 8. The degradation prediction of FC2 with different training set sizes.

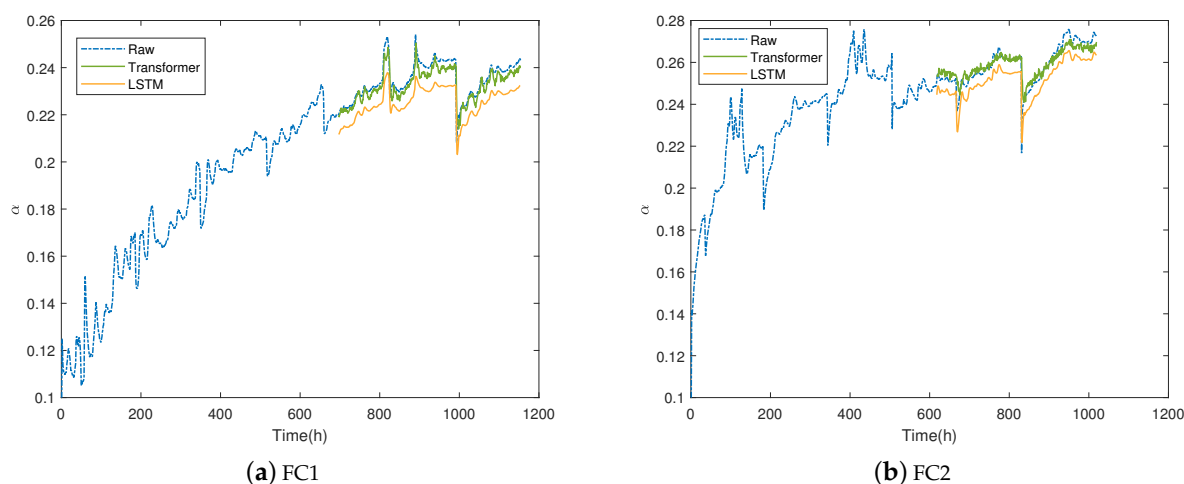**Figure 9.** Error of FC2 for degradation prediction with different training set sizes.

Table 3. Prediction performance with different training set sizes for FC2.

Training Set	MAPE (%)	RMSE (%)	MAE (%)
40%	2.9276	0.9675	0.7665
50%	1.1153	0.4408	0.2851
60%	0.9896	0.4117	0.2515

5.4. Verification with LSTM

The LSTM network is a classical method in time series data processing. In order to prove the superiority of the model proposed in this paper, LSTM model is also established to predict degradation trend. LSTM network uses two hidden layers, and the number

of neurons in the first layer and the second layer is set to 100. The other parameters are the same as the transformer settings, and the training set size is 60%. The comparison results of FC1 and FC2 are shown in Figure 10. When 50 epochs are trained, it can be roughly estimated from the figure that the prediction error of LSTM is larger than that of transformer. When the transformer training is complete, the LSTM network structure has not been trained to its best. This is because transformer can process time series data in parallel and can quickly extract global information, whereas LSTM training is iterative.

(**a**) FC1   (**b**) FC2

Figure 10. The result of PEMFC for degradation prediction with proposed method and LSTM.

6. Conclusions

In this work, a hybrid prediction framework is developed for the real-time performance degradation prediction of PEMFC, in which MC-dropout, combined with the transformer network, predicts the long-term degradation trend, and an adaptive Wiener process model with multi kinds of variability sources is established to model the degradation of PEMFC. The effectiveness and advantages of the framework are illustrated by experimental verification. While transformer network have advantages in the parallel processing of input data, the timing characteristics of time series are ignored when dealing with time series problems. Transformer only uses position encoding to determine the order of input data, whereas LSTM regards position information as the sequential input. Transformer does not fully achieve the effect of the sequential structure of LSTM in this respect. In the future, we will further consider combining these two network structures to improve the effectiveness of RUL prediction.

Author Contributions: Conceptualization, Y.H., Y.J. and K.P.; Methodology, Y.H., L.Z., Y.J., K.P. and Z.J.; Software, L.Z.; Writing—original draft, L.Z.; Writing—review & editing, Y.H. and Z.J. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhou, S.; Zhang, G.; Fan, L.; Gao, J.H.; Pei, F.L. Scenario-oriented stacks allocation optimization for multi-stack fuel cell systems. *Appl. Energy* **2022**, *308*, 118328. [CrossRef]
2. Hua, Z.G.; Zheng, Z.X.; Pahon, E.; Pera, M.C.; Gao, F. Remaining useful life prediction of PEMFC systems under dynamic operating conditions. *Energy Convers. Manag.* **2021**, *231*, 113825. [CrossRef]

3. Wang, F.K.; Huang, C.Y.; Mamo, T.; Cheng, X.B. Ensemble model for the degradation prediction of proton exchange membrane fuel cell stacks. *Qual. Reliab. Eng. Int.* **2021**, *37*, 34–46. [CrossRef]
4. Li, Q.; Yang, W.Y.; Yin, L.Z.; Chen, W.R. Real-time implementation of maximum net power strategy based on sliding mode variable structure control for proton-exchange membrane fuel cell system. *IEEE Trans. Transp. Electrif.* **2020**, *6*, 288–297. [CrossRef]
5. Manohar, M.; Kim, D. Advantageous of hybrid fuel cell operation under self-humidification for energy efficient bipolar membrane. *Acs Sustain. Chem. Eng.* **2019**, *7*, 16493–16500. [CrossRef]
6. Hua, Z.G.; Zheng, Z.X.; Pera, M.C.; Gao, F. Remaining useful life prediction of PEMFC systems based on the multi-input echo state network. *Appl. Energy* **2020**, *265*, 114791. [CrossRef]
7. Zhou, D.M.; Wu, Y.M.; Gao, F.; Breaz, E.; Ravey, A.; Miraoui, A. Degradation prediction of PEM fuel cell stack based on multiphysical aging model with particle filter approach. *IEEE Trans. Ind. Appl.* **2017**, *53*, 4041–4052. [CrossRef]
8. Chen, K.; Laghrouche, S.; Djerdir, A. Fuel cell health prognosis using Unscented Kalman Filter: Postal fuel cell electric vehicles case study. *Int. J. Hydrogen Energy* **2019**, *44*, 1930–1939. [CrossRef]
9. Wu, Y.M.; Breaz, E.; Gao, F.; Miraoui, A. A modified relevance vector machine for PEM fuel-cell stack aging prediction. *IEEE Trans. Ind. Appl.* **2016**, *52*, 2573–2581. [CrossRef]
10. Wu, Y.M.; Breaz, E.; Gao, F.; Paire, D.; Miraoui, A. Nonlinear performance degradation prediction of proton exchange membrane fuel cells using relevance vector machine. *IEEE Trans. Energy Convers.* **2016**, *31*, 1570–1582. [CrossRef]
11. Hua, Z.G.; Zheng, Z.X.; Pera, M.C.; Gao, F. Data-driven prognostics for PEMFC systems by different echo state network prediction structures. In Proceedings of the 2020 IEEE Transportation Electrification Conference & Expo (ITEC), Chicago, IL, USA, 23–26 June 2020.
12. Mezzi, R.; Yousfi-Steiner, N.; Pera, M.C.; Hissel, D.; Larger, L. An echo state network for fuel cell lifetime prediction under a dynamic micro-cogeneration load profile. *Appl. Energy* **2021**, *283*, 116297. [CrossRef]
13. Zhang, X.X.; Yu, Z.X.; Chen, W.R. Life prediction based on D-S ELM for PEMFC. *Energies* **2019**, *12*, 3752. [CrossRef]
14. Chen, K.; Laghrouche, S.; Djerdir, A. Degradation model of proton exchange membrane fuel cell based on a novel hybrid method. *Appl. Energy* **2019**, *252*, 113439. [CrossRef]
15. Peng, Y.L.; Chen, T.; Xiao, F.; Zhang, S.J. Remaining useful lifetime prediction methods of proton exchange membrane fuel cell based on convolutional neural network-long short-term memory and convolutional neural network-bidirectional long short-term memory. *Fuel Cells* **2022**, *23*, 75–87. [CrossRef]
16. Ma, R.; Yang, T.; Breaz, E.; Li, Z.; Briois, P.; Gao, F. Data-driven proton exchange membrane fuel cell degradation predication through deep learning method. *Appl. Energy* **2018**, *231*, 102–115. [CrossRef]
17. Ma, R.; Breaz, E.; Liu, C.; Bai, H.; Briois, P.; Gao, F. Data-driven prognostics for pem fuel cell degradation by long short-term memory network. In Proceedings of the 2018 IEEE Transportation and Electrification Conference and Expo(ITEC), Long Beach, CA, USA, 13–15 June 2018.
18. Liu, J.W.; Li, Q.; Chen, W.R.; Yan, Y.; Qiu, Y.B.; Gao, T.Q. Remaining useful life prediction of PEMFC based on long short-term memory recurrent neural networks. *Int. J. Hydrogen Energy* **2019**, *44*, 5470–5480. [CrossRef]
19. Zhang, R.F.; Chen, T.; Xiao, F.; Luo, J.L. Bi-directional gated recurrent unit recurrent neural networks for failure prognosis of proton exchange membrane fuel cells. *Int. J. Hydrogen Energy* **2022**, *47*, 33027–33038. [CrossRef]
20. Cheng, Y.J.; Zerhouni, N.; Lu, C. A hybrid remaining useful life prognostic method for proton exchange membrane fuel cell. *Int. J. Hydrogen Energy* **2018**, 43, 12314–12327. [CrossRef]
21. Liu, H.; Chen, J.; Hissel, D.; Su, H. Remaining useful life estimation for proton exchange membrane fuel cells using a hybrid method. *Appl. Energy* **2019**, *237*, 910–919. [CrossRef]
22. Xie, R.Y.; Ma, R.; Pu, S.C.; Xu, L.C.; Zhao, D.D.; Huangfu, Y.G. Prognostic for fuel cell based on particle filter and recurrent neural network fusion structure. *Energy AI* **2020**, *2*, 100017. [CrossRef]
23. Ma, R.; Xie, R.Y.; Xu, L.C.; Huangfu, Y.G.; Li, Y.R. A hybrid prognostic method for PEMFC with aging parameter prediction. *IEEE Trans. Transp. Electrif.* **2021**, *7*, 2318–2331. [CrossRef]
24. Xia, Z.T.; Wang, Y.N.; Ma, L.H.; Zhu, Y.; Li, Y.J.; Tao, J.L.; Tian, G.Z. A hybrid prognostic method for proton-exchange-membrane fuel cell with decomposition forecasting framework based on AEKF and LSTM. *Sensors* **2023**, *23*, 166. [CrossRef] [PubMed]
25. Ashish, V.; Noam, S.; Niki, P.; Jakob, U.; Llion, J.; Aidan, N.G.; Lukasz, K.; Illia, P. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
26. Wei, M.; Gu, H.R.; Ye, M.; Wang, Q.; Xu, X.X.; Wu, C.G. Remaining useful life prediction of lithium-ion batteries based on Monte Carlo Dropout and gated recurrent unit. *Energy Rep.* **2021**, *7*, 2862–2871. [CrossRef]
27. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016.
28. Larminie, J.; Dicks, A. *Fuel Cell Systems Explained*, 2nd ed.; John Wiley and Sons: Chichester, UK, 2003.
29. Pan, R.; Yang, D.; Wang, Y.J.; Chen, Z.H. Performance degradation prediction of proton exchange membrane fuel cell using a hybrid prognostic approach. *Int. J. Hydrogen Energy* **2020**, *45*, 30994–31008. [CrossRef]
30. Jha, M.S.; Bressel, M.; Ould-Bouamama, B.; Dauphin-Tanguy, G. Particle filter based hybrid prognostics of proton exchange membrane fuel cell in bond graph framework. *Comput. Chem. Eng.* **2016**, *95*, 216–230. [CrossRef]
31. Bressel, M.; Hilairet, M.; Hissel, D.; Bouamama, B.O. Extended Kalman Filter for prognostic of Proton Exchange Membrane Fuel Cell. *Appl. Energy* **2016**, *164*, 220–227. [CrossRef]

32. Julier, S.J.; Uhlmann, J.K.; Durrant-Whyte, H.F. A new approach for filtering nonlinear systems. In Proceedings of the 1995 American Control Conference—ACC'95, Seattle, WA, USA, 21–23 June 1955.

33. Zhang, J.S.; Jiang, Y.C.; Li, X.; Huo, M.Y.; Luo, H.; Yin, S. An adaptive remaining useful life prediction approach for single battery with unlabeled small sample data and parameter uncertainty. *Reliab. Eng. Syst. Saf.* **2022**, *222*, 108357. [CrossRef]

34. FCLAB Research. IEEE PHM 2014 Data Challenge. 2014. Available online: http://eng.fclab.fr/ieeephm-2014-data-challenge/ (accessed on 16 May 2014).