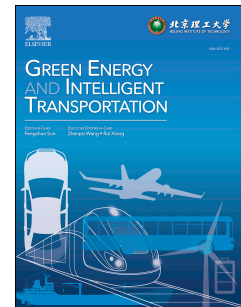


Journal Pre-proof

Uncertainty Quantification-based Framework for Predicting Degradation Trends of Proton Exchange Membrane Fuel Cell

Bingxin Guo, Changjun Xie, Wenchao Zhu, Yang Yang, Hao Li, Yang Li, Hangyu Wu



PII: S2773-1537(25)00047-7

DOI: <https://doi.org/10.1016/j.geits.2025.100297>

Reference: GEITS 100297

To appear in: *Green Energy and Intelligent Transportation*

Received Date: 1 June 2024

Revised Date: 10 September 2024

Accepted Date: 23 September 2024

Please cite this article as: Guo B, Xie C, Zhu W, Yang Y, Li H, Li Y, Wu H, Uncertainty Quantification-based Framework for Predicting Degradation Trends of Proton Exchange Membrane Fuel Cell, *Green Energy and Intelligent Transportation*, <https://doi.org/10.1016/j.geits.2025.100297>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 The Author(s). Published by Elsevier Ltd on behalf of Beijing Institute of Technology Press Co., Ltd.

Uncertainty Quantification-based Framework for Predicting Degradation Trends of Proton Exchange Membrane Fuel Cell

Bingxin Guo¹, Changjun Xie^{1,2,3}, Wenchao Zhu^{1,2,*}, Yang Yang^{1,2}, Hao Li^{1,*},
Yang Li⁴, Hangyu Wu¹

¹ School of Automation, Wuhan University of Technology, Wuhan 430070, China.

² Hubei Key Laboratory of Advanced Technology for Automotive Components, Wuhan University of Technology, Wuhan 430070, China.

³ Modern Industry College of Artificial Intelligence and New Energy Vehicles, Wuhan University of Technology, Wuhan 430070, China.

⁴ Department of Electrical Engineering, Gothenburg, 41296, Sweden.

* Corresponding author

E-mail address: zhuwenchao@whut.edu.cn (W. Zhu); lihao@whut.edu.cn (H. Li)

Author contributions

Bingxin Guo: Data curation, Investigation, Methodology, Software, Validation, Visualization, Writing-original draft, Writing-review & editing

Changjun Xie: Funding acquisition, Resources, Supervision

Wenchao Zhu: Supervision

Yang Yang: Supervision

Hao Li: Supervision

Yang Li: Supervision

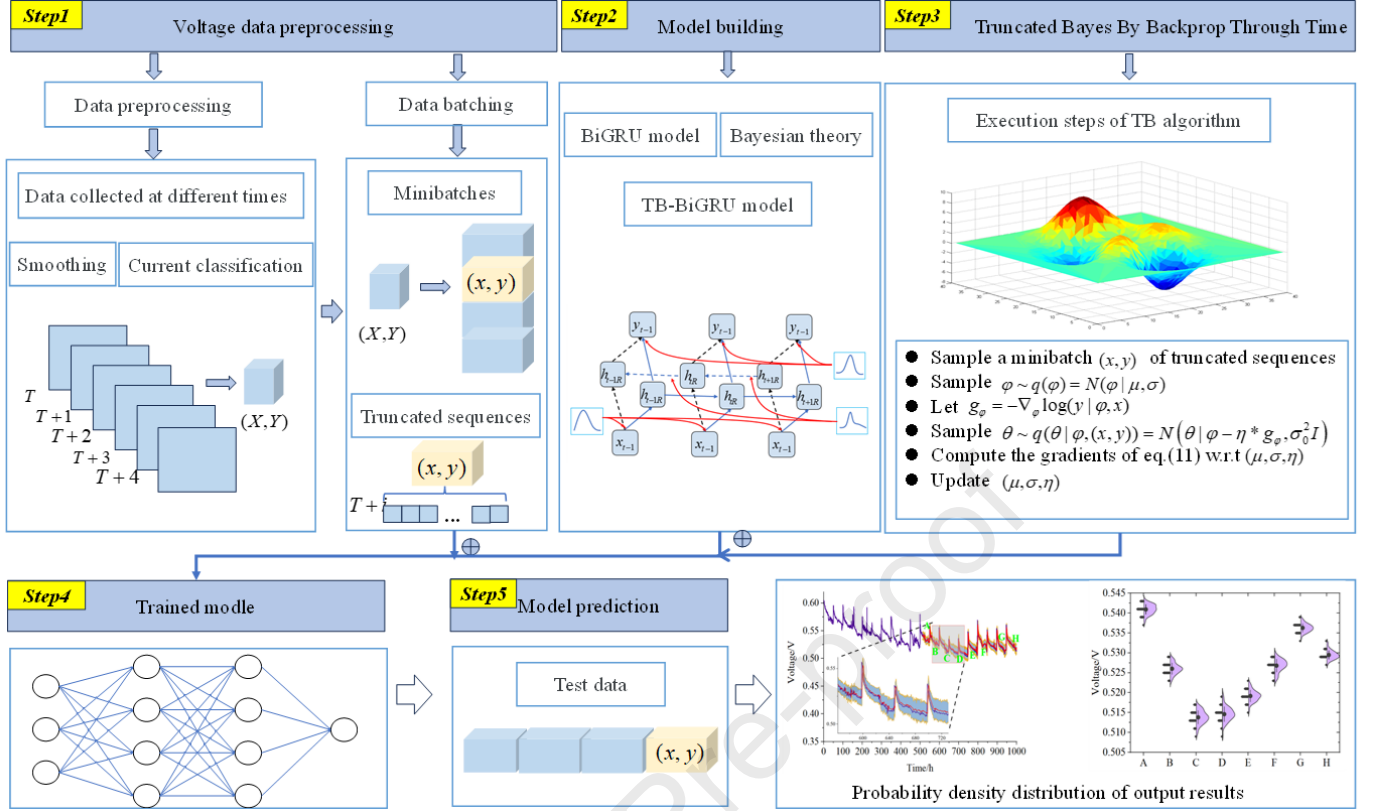
Hangyu Wu: Investigation

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2020YFB1506802) and the Key R&D Plan of Hubei Province (2023BAB114).

Data Availability

The data and materials used to support the findings of this study are available from the corresponding author upon reasonable request.



Uncertainty Quantification-based Framework for Predicting Degradation Trends of Proton Exchange Membrane Fuel Cell

Abstract:

Accurately predicting the degradation trends of proton exchange membrane fuel cells (PEMFCs) can provide a solid basis for optimizing the control of vehicles and stations based on PEMFCs. However, most prediction methods do not consider factors such as measurement errors from experimental environments and the inherent cognitive uncertainty of the model. These methods can only offer point estimates, lacking credibility. This paper introduces a deep learning prediction framework that combines a bidirectional gated recurrent unit (BiGRU) model with a truncated Bayes by backpropagation through time (TB) algorithm. The TB algorithm reconstructs fixed parameters in the model into probability density distributions, transforming the output from point estimation to interval estimation with probability density distributions. Under dynamic conditions, the TB-BiGRU (truncated Bayes-based bidirectional gated recurrent unit) improves the mean absolute error (MAE) and root mean square error (RMSE) by 37.28% and 36.09%, respectively, compared to the TB-GRU (truncated Bayes-based gated recurrent unit). Compared with TB-GRU and B-GRU (Bayesian gated recurrent unit), TB-BiGRU has significantly improved uncertainty quantification ability. Under different working conditions and noise levels, the prediction accuracy of TB-BiGRU is superior to that of the other seven models, and it exhibits better noise resistance and stability. This method holds greater practical significance compared to other prediction approaches. Additionally, the paper proposes four effective evaluation

metrics for uncertainty quantification, providing higher reference value in effectively characterizing the model's prediction accuracy and uncertainty quantification capability.

Keywords:

Uncertainty quantification; Truncated bayes by backprop through time (TB); Bidirectional gated recurrent unit (BiGRU); Proton exchange membrane fuel cell (PEMFC)

1. Introduction:

1.1 Background and Literature Review

Proton exchange membrane fuel cell (PEMFC) is characterized by its high power density, light weight, abundant resources, and environmental friendliness [1-2]. It has been widely applied in various modes of transportation, including new energy vehicles [3], heavy-duty trucks [4], and trains [5], emerging as one of the most promising power generation equipment [6]. However, the high cost and limited durability of PEMFC have hindered its large-scale commercialization [7-8].

In order to enhance durability and save development costs, it is essential to predict the degradation trend and monitor the health status of PEMFC [9]. Currently, PEMFC degradation prediction methods include model-driven, data-driven, and hybrid models [10]. Due to the complex nature of PEMFC systems involving multi-physics, multi-scale interactions of delayed electrochemical reactions and proton transfer, establishing a precise physical model to describe the dynamic behavior of PEMFC is highly challenging [11]. For instance, Khan et al. [12] introduced a dynamic semi-empirical model designed to forecast the degradation of PEMFC. Kandidayeni et al. [13] introduced a technique for estimating time-varying parameters, aiming to enhance the predictive precision of the PEMFC semi-empirical model. Xing et al. [14] introduced an adaptive estimation approach that incorporates unspecified parameters and formulated a nonlinear dynamic model for PEMFC based on this methodology. These model-driven prediction methods always rely on complex, high-precision models, which are currently unattainable [15].

With the progress of computer science, data-driven methods represented by deep learning has showcased impressive scalability and the ability to generalize effectively when dealing with extensive and intricate datasets [16]. The application of deep

learning in health prediction has garnered growing interest both in academia and industry [17-18]. Zuo et al. [10] combines the attention mechanism with the recurrent neural network (RNN) to precisely predict the degradation trend of PEMFC output voltage through dynamic durability test data. Wang et al. [19] introduced a symbolic Long Short-Term Memory network (LSTM), which predicts the performance degradation trend of PEMFC by reconstructing the data into a sequence of symbols. In [15], a data-driven model for predicting PEMFC performance was obtained using deep belief networks, which was trained with data from a 3D PEMFC numerical model. Ma et al. [20] developed an online data-driven method based on G-LSTM. By learning underlying relationships within degradation data, data-driven methods effectively the challenges faced by complex physical models and provide more accurate predictive results.

Due to the complex and diverse operating conditions of fuel cells [21-22], as well as the noise contained in the data, which can seriously affect the prediction of fuel cell degradation trends, in order to improve the hardness of the prediction results, the above deep learning methods will use filtering methods or signal decomposition to filter out data noise. However, noise filtering often leads to information loss and a lack of credibility in prediction results, which may result in erroneous control decisions and potentially pose safety issues, making it difficult to apply in practice. More importantly, these prediction methods do not take into account the impact of various uncertainty factors in health prediction, such as measurement uncertainty, model uncertainty related to deep learning models, and prediction uncertainty caused by the randomness of future environmental and operating conditions [23-24], which may lead to unreliable prediction results. Addressing the objective reality of uncertainty in health prediction and applying scientific methods to predict the performance degradation of fuel cells is an urgent problem that needs to be solved.

In 2015, Ghahramani [25] first categorized uncertainty into epistemic uncertainty and aleatoric uncertainty in *Nature*. Epistemic uncertainty arises from the prediction model and is often referred to as model uncertainty. Aleatoric uncertainty measures the inherent noise in dataset, stemming from the method of data collection. Ghahramani [25] also emphasized the usefulness of Bayesian methods in handling uncertainty in machine learning. Since then, machine learning algorithms based on Bayesian methods have been employed to address uncertainty quantification problems in various fields. Wang et al. [23] introduced a Bayesian Neural Network (BNN) to quantify uncertainty

in diesel engines. He delved into the structure and parameters of the BNN model and visualized the model parameters to present the probability density distribution they follow. This model is only a tentative combination of Bayesian theory and artificial neural networks, and its effectiveness in other engineering fields needs further exploration. Peng et al. [24] introduced an uncertainty quantification method based on Bayesian Deep Learning (BDL), effectively quantifying uncertainty in the prediction process in two typical applications: ball bearing and turbocharger scenarios. BDL is a deep extension of the BNN model. Although it achieves uncertainty quantification in multivariate feature regression, it performs poorly in predicting the degradation trend of PEMFC under complex and diverse operating conditions. Zhu et al. [26] combined Bayesian theory with Gated Recurrent Unit (GRU) to propose a B-GRU model capable of uncertainty quantification, which excels in both single-point estimation prediction and uncertainty quantification. This model is the first time our team has combined Bayesian theory with GRU model and validated its effectiveness under specific PEMFC operating conditions, providing a feasible path for us to combine Bayesian theory with other deep learning prediction models. Xie et al. [27] combined singular spectrum analysis (SSA) and deep Gaussian process (DGP) to predict remaining useful life (RUL). The model can accurately predict the nonlinear details of PEMFC performance degradation, and quantifies the uncertainty to provide reliable confidence intervals for prediction results. But the prediction accuracy and uncertainty quantification ability of this model will be greatly affected by the kernel density function. Jia et al. [28] applied the multi head self attention mechanism to the BiGRU model to predict the degradation trend of PEMFC. Although the model did not consider uncertainty quantification, its research results highlighted that the BiGRU model has better predictive performance and generalization ability. In addition, Li et al. [29] combined quantile regression with BiGRU model and used sparrow optimization algorithm for hyperparameter optimization to achieve interval prediction of PEMFC degradation trend. The authors compared and analyzed BiGRU with GRU, LSTM and other models in detail, and the results showed that BiGRU model has higher prediction accuracy. However, this article did not provide a detailed analysis of uncertainty quantification. Although the above-mentioned researchers have made certain contributions to the quantification of uncertainty, there is still a common problem that they have not proposed corresponding evaluation indicators for uncertainty quantification. Therefore, how to more intuitively evaluate the uncertainty quantification of evaluation models is still an urgent problem

to be solved.

1.2 Research Gap and Contributions

In summary, applying Bayesian theory to neural networks is a potentially viable approach that allows for uncertainty quantification and enhances the credibility of model predictions. Crucially, through this improvement, performance predictions along with uncertainty quantification can be more effectively applied in practical applications.

Considering that the BiGRU model is more suitable for solving the degradation trend prediction problem of PEMFC, this paper developed a TB algorithm and combined it with the BiGRU model to establish the TB-BiGRU model, in order to improve the accuracy and credibility of degradation trend prediction. Additionally, four uncertainty quantification evaluation metrics are proposed to characterize the model's capability in uncertainty quantification. The newly proposed model exhibits higher prediction accuracy, and the new evaluation metrics accurately represent the model's capability in uncertainty quantification, making it more suitable for handling dynamic test sets of PEMFC. The methods presented in this paper bring forth the following novel and innovative elements:

- 1) Compared with TB-GRU, TB-BiGRU has improved to varying degrees under all conditions, and the improvement is more significant under dynamic conditions.
- 2) Compared with other deep learning models, the TB BiGRU model exhibits superior noise resistance, higher accuracy in qualitative analysis, and more reliable uncertainty quantification ability under different working conditions and noise levels
- 3) Four evaluation metrics are proposed to measure the accuracy and effectiveness of the model, which can well characterize the uncertainty quantification ability of the model.
- 4) The proposed framework is also suitable for other deep learning models, which has migration capabilities and scalability.

The paper proceeds with the following organization:

In Section 2, the prediction framework based on TB for BiGRU models is presented, including the fundamental principles of the TB algorithm and the configuration of internal model parameters. Section 3 introduces experiment and datasets. Section 4 introduces four uncertainty quantitative evaluation metrics. Section 5 demonstrates the superiority of the new model from various perspectives and validates the reliability of the proposed evaluation metrics. Finally, Section 6 concludes

the paper.

2. Methodology

In order to enhance the accuracy and credibility of PEMFC degradation trend prediction, this section proposes a prediction framework based on truncated bayes by backprop through time algorithm, and elucidates the fundamental theory of the related models. We construct the TB-BiGRU models based on the bayesian theory.

2.1 The Prediction Framework Based on TB-GRU

Combining the TB algorithm with the BiGRU model specifically refers to reconstructing the internal parameters of the BiGRU model using this algorithm. In other words, the originally fixed parameters in the model are reconstructed to follow a probability density distribution of Gaussian distribution. The prediction framework mainly consists of the following steps: voltage data preprocessing, TB-BiGRU model establishment, TB algorithm, model training based on the TB algorithm, and model prediction.

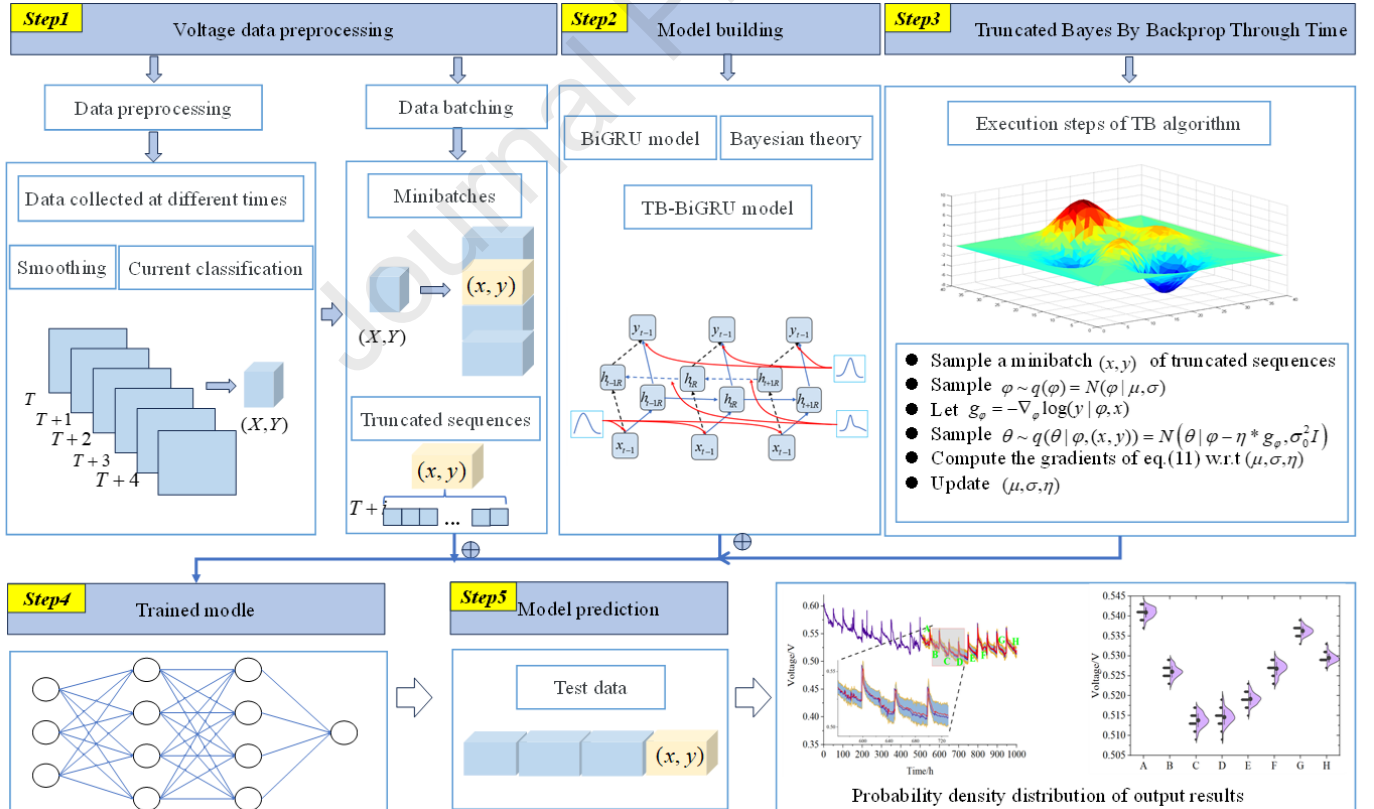


Figure 1. The Degradation trend prediction framework of PEMFC based on TB-BiGRU model

Firstly, historical data of the fuel cell is collected and processed through smoothing and batching to prepare for model training. Then, the TB algorithm is applied to

reconstruct the model parameters, transforming the original fixed parameters into probability density distributions. This quantifies uncertainty through probability density distributions. Next, the model is trained according to the execution flow of the TB algorithm, utilizing the Adam optimization algorithm. Finally, combining historical and operational data yields uncertainty-quantified prediction results.

With this prediction framework, results that incorporate uncertainty quantification can offer more insightful decision recommendations. The probability density distributions provided by the prediction results serve as crucial reference standards for practical applications. Further details about the mathematical model are provided below.

2.2 BiGRU

BiGRU consists of forward GRU and reverse GRU. The calculation principle of each link in GRU is as follows:

$$z_t = \sigma(W_z \bullet [h_{t-1}, x_t]) \quad (1)$$

$$r_t = \sigma(W_r \bullet [h_{t-1}, x_t]) \quad (2)$$

$$h_t = \tanh(W_h [r_t \square h_{t-1}, x_t]) \quad (3)$$

$$h_t = (1 - z_t) \square h_{t-1} + z_t \square h_t \quad (4)$$

In the formula, W_z , W_r and W_h are the network weights for the update gate, reset gate, and candidate state. σ and \tanh are the activation functions corresponding to the update gate and reset gate, z_t, r_t and h_t are the outputs of the update gate, reset gate, and hidden state, \square is the element wise product of a matrix, \bullet is a matrix multiplication operation.

In unidirectional GRU, the reset gate combines the hidden state h_{t-1} of the previous neuron with the input x_t of the current neuron to obtain the candidate hidden state h_t . Equation (4) indicates that the update gate controls the ratio of the hidden state h_{t-1} of the previous neuron to the candidate hidden state h_t of the current neuron, obtaining a new hidden state h_t and subsequently obtaining the GRU prediction result. In this process, the hidden state always passes from front to back, which is not conducive to feature extraction work. In the process of voltage state changes in PEMFC, the voltage states before and after are highly correlated, and it is necessary to consider this characteristic.

The hidden state of BiGRU is determined by the combined states of the positive

and negative GRUs, as shown in Figure 2(a). At this point, the generation process of BiGRU hidden state can be represented as:

$$h_{Bi} = Bi(h_{iF}, h_{iR}) \quad (5)$$

In the formula, h_{Bi} represents the hidden state of BiGRU at the current time, while h_{iF} and h_{iR} represent the hidden states of forward and reverse GRU. This bidirectional structure enables BiGRU to better extract voltage variation characteristics and improve the prediction performance of nonlinear components.

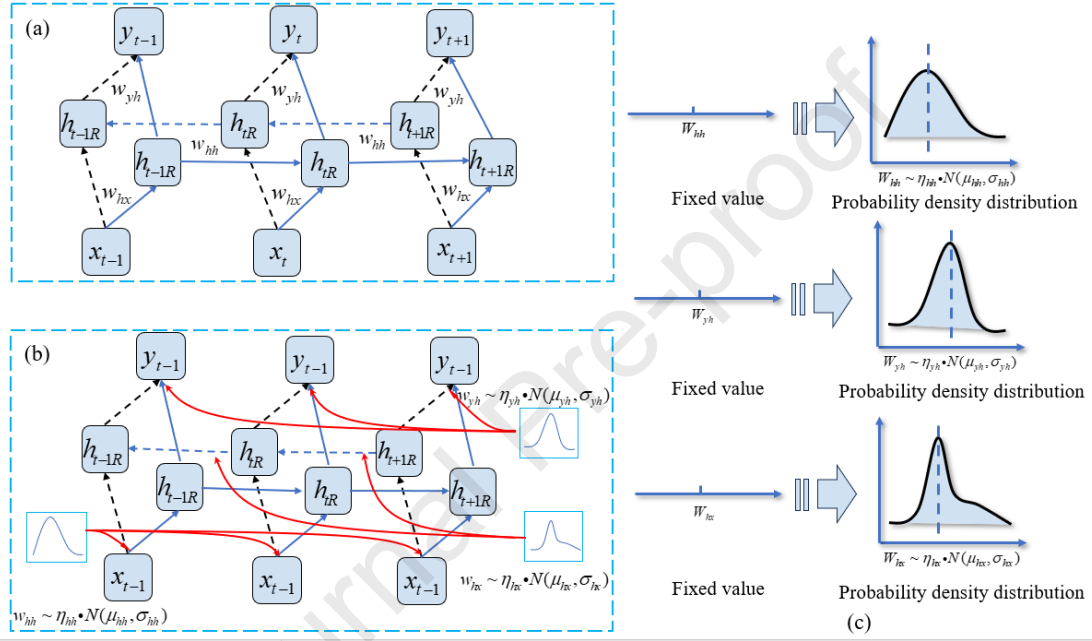


Figure 2. Comparison of BiGRU structure based on TB algorithm and original BiGRU structure. (a) The original BiGRU structure diagram; (b) The BiGRU Structure based on TB Algorithm; (c) Fixed weights replaced with probability density distributions

2.3 Bayes By Backprop

Bayes By Backprop (BBB) employs variational inference to estimate the posterior distribution of weights $\theta \in R^d$ in a neural network. This posterior distribution is usually considered to be a Gaussian distribution, denoted as $N(\theta | \mu, \sigma^2)$, where μ, σ represent the mean and standard deviation of the Gaussian distribution, satisfying $\mu \in R^d, \sigma \in R^d$, and d is the dimensionality of the network parameters.

Let $\log p(y | \theta, x)$ be the log-likelihood of the model, then train the model by minimizing the variational free energy of (6):

$$L(\theta) = E_{q(\theta)} \left[\log \frac{q(\theta)}{p(y | \theta, x) p(\theta)} \right] \quad (6)$$

Where, $p(\theta)$ is the prior distribution of the parameters.

Minimizing the variational free energy (6) is equivalent to maximizing the log-likelihood $\log p(y|\theta, x)$ of the network parameters with the KL complexity term acting as a regularizer:

$$L(\theta) = -E_{q(\theta)}[\log p(y|\theta, x)] + KL[q(\theta)||p(\theta)] \quad (7)$$

In the case of a Gaussian prior with zero mean, the KL can be viewed as a form of mean parameter weight decay.

2.4 BiGRU based on Truncated Bayes By Backprop Through Time (TB-BiGRU)

Applying TB to reconstruct BiGRU is depicted in Fig. 2(b). Compared to the fixed parameters in the original BiGRU structure, the model parameters of TB- BiGRU approximately obey a Gaussian distribution, which realizes the quantification of uncertainty factors through the form of probability density distribution of parameters and the weight matrices of BiGRU are obtained from a distribution learned by TB. The variational free energy of BiGRU on a sequence of length T is:

$$L(\theta) = -E_{q(\theta)}[\log p(y_{1:T}|\theta, x_{1:T})] + KL[q(\theta)||p(\theta)] \quad (8)$$

Where, $\log p(y_{1:T}|\theta, x_{1:T})$ is the likelihood of the sequence generated when the state f_T is input to an appropriate probability distribution, and θ is the distribution that the network parameters follow. Due to the length of the time series, in practice, it is necessary to use mini-batch training with truncated sequences to train the BiGRU model.

Let B be the batch size and C be the number of truncated sequences. Then, (6) can be written as:

$$L(\theta) = -E_{q(\theta)}[\log \sum_{b=1}^B \sum_{c=1}^C p(y^{(b,c)}|\theta, x^{(b,c)})] + KL[q(\theta)||p(\theta)] \quad (9)$$

Where, b represents the b th mini-batch and c represents the c th truncated sequence within the mini-batch. Therefore, the variational free energy of the b th mini-batch and the c th truncated sequence can be written as:

$$L_{(b,c)}(\theta) = -E_{q(\theta)}[\log p(y^{(b,c)}|\theta, x^{(b,c)}, s_{\text{prev}}^{(b,c)})] + w_{KL}^{(b,c)} KL[q(\theta)||p(\theta)] \quad (10)$$

Where, $w_{KL}^{(b,c)}$ allocates the KL cost to minibatches and truncated sequences, and $\sum_{b=1}^B \sum_{c=1}^C w_{KL}^{(b,c)} = 1$. $s_{\text{prev}}^{(b,c)}$ represents the initial state of the BiGRU for the minibatch. In practice, we set $w_{KL}^{(b,c)} = 1/BC$, so the KL penalty is evenly distributed across all mini-batches and truncated sequences. Each subsequent mini-batch's truncated sequence is selected in order, hence $s_{\text{prev}}^{(b,c)}$ is set as the last state of the $x^{(b,c-1)}$ th BiGRU.

For the variational posterior $q(\theta)$, it can be strengthened by adding side information to make the posterior of parameters more accurate. For the given sample data (x, y) , construct $q(\theta | (x, y))$:

$$q(\theta | (x, y)) = \int q(\theta | \varphi, (x, y)) q(\varphi) d\varphi \quad (11)$$

Where $q(\varphi) = N(\varphi | \mu, \sigma)$, $\mu, \sigma \in R^d$, and $*$ represents element-wise product, we have:

$$q(\theta | \varphi, (x, y)) = N(\theta | \varphi - \eta * g_\varphi, \sigma_0^2 I) \quad (12)$$

Where, $\eta \in R^d$ is a learnable hyperparameter, σ_0 is a scalar hyperparameter, and I is the identity matrix.

During the training process, we optimize the error by obtaining $\theta \sim q(\theta | (x, y))$ through ancestral sampling:

$$L(\mu, \sigma, \eta) = E_{(x, y)} [E_{q(\varphi)q(\theta | \varphi, (x, y))} [L(x, y, \theta, \varphi | \mu, \sigma, \eta)]] \quad (13)$$

Where,

$$L(x, y, \theta, \varphi | \mu, \sigma, \eta) = -\log p(y | \theta, x) + KL[q(\theta | \varphi, (x, y)) \| p(\theta | \varphi)] + (1/C) KL[q(\varphi) \| p(\varphi)] \quad (14)$$

Where, μ, σ, η represents the d -dimensional model parameters, p is the prior defined by q , and C is a constant denoting the number of truncated sequences. From here, the parameter optimization of the model changes from θ to μ, σ, η . Fig1. illustrates how Truncated Bayes by Backprop Through Time is implemented to quantify uncertainty in practice.

3. Data and Experimental Preparation

3.1 Introduction to Experiments and Datasets

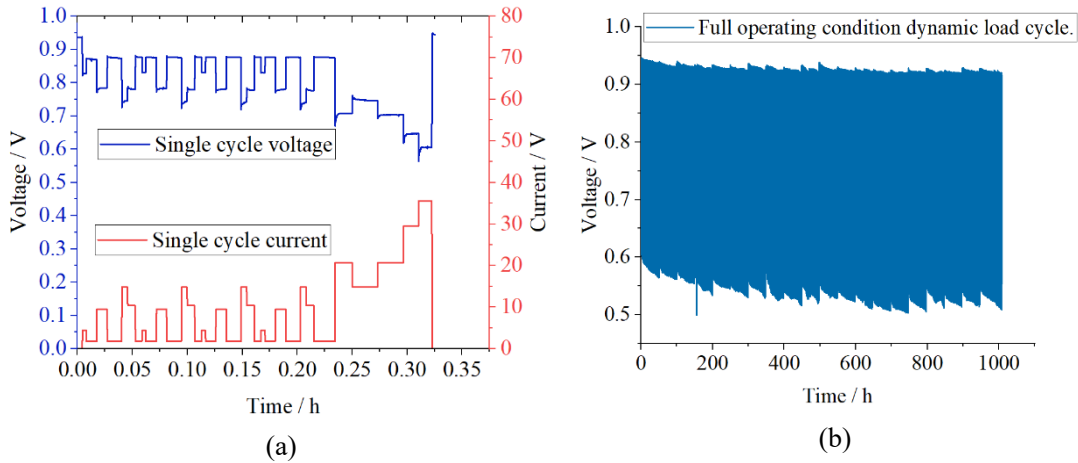


Figure 3. The output voltage of the fuel cell

The method proposed in this paper focuses on the results of durability test for individual automotive PEMFC cells [30], referring to the New European Driving Cycle (NEDC) [31] as a reference. The key technical parameters involved in the experiments are listed in Table 1. The inlet pressures of the cathode/anode, relative humidity, and operating temperature were regulated through external cooling water systems and built-in humidifiers. The entire aging experiment consisted of 3076 dynamic load cycles (FC-DLC), totaling approximately 1008 hours, covering nine different load currents as shown in Table 1. Each FC-DLC cycle comprised 35 current load steps, as illustrated in Fig. 3(a). The output voltage of the entire aging test is depicted in Fig. 3(b).

Table 1. Dynamic aging test conditions

Relevant parameters	value				
Activation Area. / cm^2	25				
Inlet pressure of H_2 / kPa	110				
Inlet pressure of air/ kPa	110				
Operating temperature. / $^{\circ}C$	85				
Relative humidity of H_2 / %	50				
Relative humidity of air / %	80				
Full load current. / A	35.6				
Load currents involved in the dynamic load	0	1.76	4.42	9.48	10.37
cycle / A	14.81	20.70	29.58	35.53	

3.2 Data Set Preprocessing

Traditional smoothing methods are inadequate for meeting the preprocessing needs of raw data, requiring specialized data processing techniques. Through the analysis of raw data within each current load step of the FC-DLC cycle, it is observed that the data located in the middle segment of each current load is closer to actual working conditions. Therefore, this paper adopts the following data smoothing rule for the raw data: for each current load step of the FC-DLC cycle, the count of the fifth-last voltage value is chosen as the final value of this step [19]. The data after this specific smoothing process is used as the FC-DLC dynamic operating condition dataset.

This paper separates the voltages under different current loads to serve as the steady-state load dataset for performance degradation prediction [32]. Firstly, specific voltage data corresponding to a specific current load is selected from the FC-DLC full operating condition dataset. Then, we smooth the dataset and use it as the FC-DLC steady-state operating condition dataset.

3.3 Model Training and Hyperparameter Selection

Model training and hyperparameter selection are key steps in machine learning and deep learning processes [21], which have a crucial impact on the performance and practical applications of the final model. On the one hand, appropriate hyperparameter configuration can reduce training time, accelerate model convergence speed, and obtain high-performance models in a shorter period of time, which is particularly important for scenarios that require processing large-scale datasets or real-time applications; On the other hand, reasonable selection of hyperparameters can improve the stability and robustness of the model, avoid training instability or model performance fluctuations caused by improper hyperparameter settings, and ensure consistent performance of the model in different datasets and application environments. Therefore, this article explores the two hyperparameters that can most affect the performance of the model, namely the number of hidden layers and the number of neurons corresponding to the hidden layers.

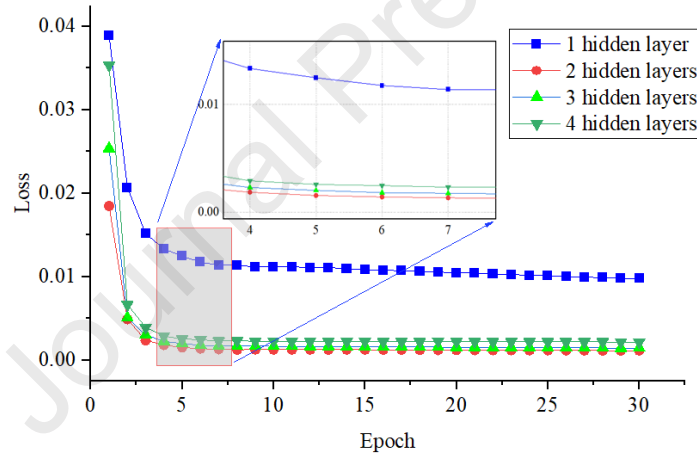


Figure 4. The variation of losses corresponding to different hidden layers of TB-BiGRU with iteration times

For neural networks and deep learning, the more layers built, the better the effect. Too many hidden layers will greatly reduce the convergence speed of the model and affect its prediction accuracy. In this section, while ensuring that the number of neurons in each hidden layer is consistent, the number of TB-BiGRU hidden layers is set to 1, 2, 3, and 4, respectively, and the down sampling loss during the model training process is calculated according to equation (14) to characterize the convergence effect of the model. The model training process uses an ADAM optimizer to update and iterate the internal parameters of the model. It calculates the gradient of the objective function (14) and corrects, updates, and iterates the first-order and second-order moments of the

gradient to optimize the model parameters. The parameter settings of the ADAM optimizer are shown in Table 3. The convergence of the losses corresponding to different hidden layers with the number of iterations is shown in the figure 4:

From the figure, it can be seen that the TB-BiGRU model has a fast convergence speed, and the model has basically reached convergence at 5 iterations. As the number of iterations increases, the convergence results of the loss corresponding to different hidden layers vary. When the hidden layer is 2, its convergence result is the smallest. When the hidden layers are set to 3, 4, or even more, the corresponding convergence results actually increase. On the one hand, this indicates that the more complex the model structure, the better. On the other hand, choosing the appropriate number of hidden layers can help improve the convergence accuracy of the model. For the prediction of voltage degradation in PEMFC, this paper sets the number of hidden layers in the TB-BiGRU model to 2. After determining the number of hidden layers, it is still necessary to further determine the number of neurons in each layer. In this paper, a grid search method is adopted to calculate the convergence results of the model corresponding to different combinations of neurons. The three-dimensional convergence diagram is shown in the figure 5:

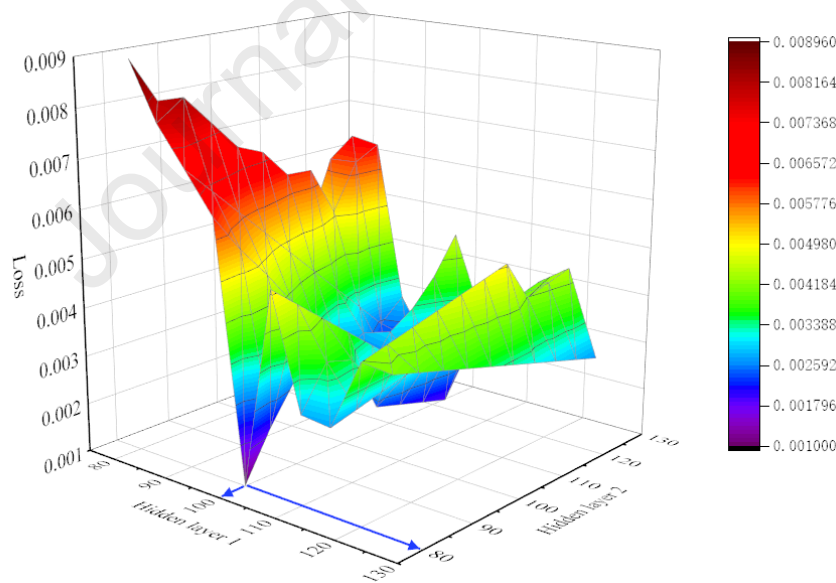


Figure 5. Three dimensional convergence diagram of TB-BiGRU under different combinations of neurons

According to the color mapping of the three-dimensional convergence graph, it can be seen that when the number of neurons in hidden layer 1 and hidden layer 2 is set to (100, 80) respectively, the convergence result of the TB-BiGRU model is optimal. Therefore, the number of neurons in the two hidden layers is set to 100 and 80

respectively. For the input layer of the model, the number of neurons is generally related to the shape of the input data, which will not be discussed here. At the same time, in order to facilitate comparison with the GRU and TB-GRU models, the hyperparameters of the GRU and TB-GRU models were kept consistent with those of the TB BiGRU model, and the model parameter settings are shown in Table 2. Here, "50/100/100" indicates that this layer has 50 neurons, with each neuron having 100 weight parameters and 100 bias parameters.

Table 2. Internal structure parameter settings of the model

Model	Sampling times	Input layer	Hidden layer	Hidden layer	Output layer
GRU	-	50/100/100	100/80/80	80/10/10	10/1/1
TB-GRU	50	50/100/100	100/80/80	80/10/10	10/1/1
TB-BiGRU	50	50/100/100	100/80/80	80/10/10	10/1/1

Table 3. ADAM optimization algorithm parameter settings

Parameters	Physical meaning	Value
m_0	Initial value of first moment estimate	0
v_0	Initial value of second moment estimate	0
β_1	Exponential decay rate for the first moment estimate	0.9
β_2	Exponential decay rate for the second moment estimate	0.999
α	Learning rate	0.1
ε	A positive constant	10^{-8}

3.4 Visualization of Network Parameters

According to the model introduction in Chapter 2, the parameters of the TB-BiGRU model exhibit an approximate Gaussian distribution compared to the BiGRU model. This is achieved by replacing the weight parameter w_{hx}, w_{yh}, w_{hh} of the BiGRU model, which exists in point form, with a Gaussian distribution controlled by μ, σ, η . In order to highlight the unique features of the TB-BiGRU model compared to the BiGRU model, this section visualizes some parameters of the TB-BiGRU model to facilitate understanding of the parameter form of the new model proposed in this article. The parameter visualization results are shown in figure 6:

The visualization results in figure 6 show that after the model training is completed, the GRU model parameters are fixed values, while the TB BiGRU model parameters exhibit a probability density distribution. For weight w_{34} in the graph, it is $w_{34} = 0.486$ in the GRU model. In the TB-BiGRU model, w_{34} follows a Gaussian

distribution controlled by three parameters, namely $\mu_{34} = 0.481$, $\sigma_{34}^2 = 0.642$, and $\eta_{34} = 0.831$. During the model prediction process, since the parameters of the GRU model are fixed, the corresponding prediction results are theoretically determined. However, the parameters of the TB-BiGRU model are approximately Gaussian distributed, and their corresponding prediction results should follow a probability density distribution. The special structure of the TB-BiGRU model can not only improve the model's anti-interference ability and stability, but also provide important reference for practical control decisions due to its inherent uncertainty quantification ability.

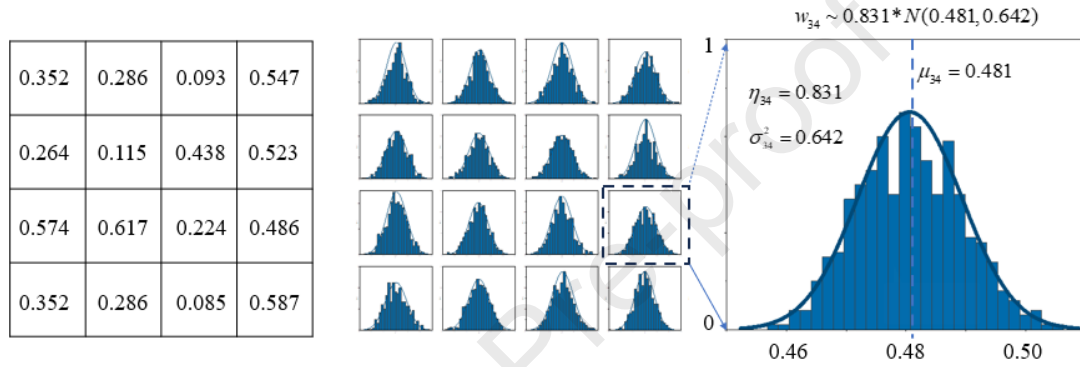


Figure 6. Visualization results of some parameters of GRU and TB-BiGRU models

3.5 The mechanism for quantifying uncertainty

The uncertainty quantification ability of the TB-BiGRU model mainly comes from its internal parameter structure. In section 3.4, we visualized some parameters of the TB-BiGRU model, and the visualization results showed that the model parameters of TB-BiGRU follow an approximate Gaussian distribution rather than fixed point form parameters. After the data is input into the model, the model will perform parallel sampling on the probability density distribution followed by the parameters to determine the model weights, in order to achieve the propagation process from input data to output results. The specific implementation process is shown in Figure 7:

Figure 7 divides the uncertainty quantification process of the model into three steps: inputting historical data, parallel parameter sampling, and outputting prediction results. Specifically, a sliding window is used to divide historical data containing uncertain factors according to the length of the time window (Figure B shows the principle of uncertainty quantification, which only divides eight time windows A, B, C, D, E, F, G, and H, each containing data from 10 historical moments); Secondly, the internal parameters of the TB-BiGRU model are sampled multiple times through parallel sampling to obtain multiple sets of parameter matrices, each corresponding to

a fixed model; Finally, when inputting time window A, the data within the time window will be subjected to matrix operations with the sampled parameter matrix to obtain the output result. Each set of parameter matrices corresponds to an output result, and when the number of samples is sufficient, the output result will present a probability density distribution, thereby achieving uncertainty quantification.

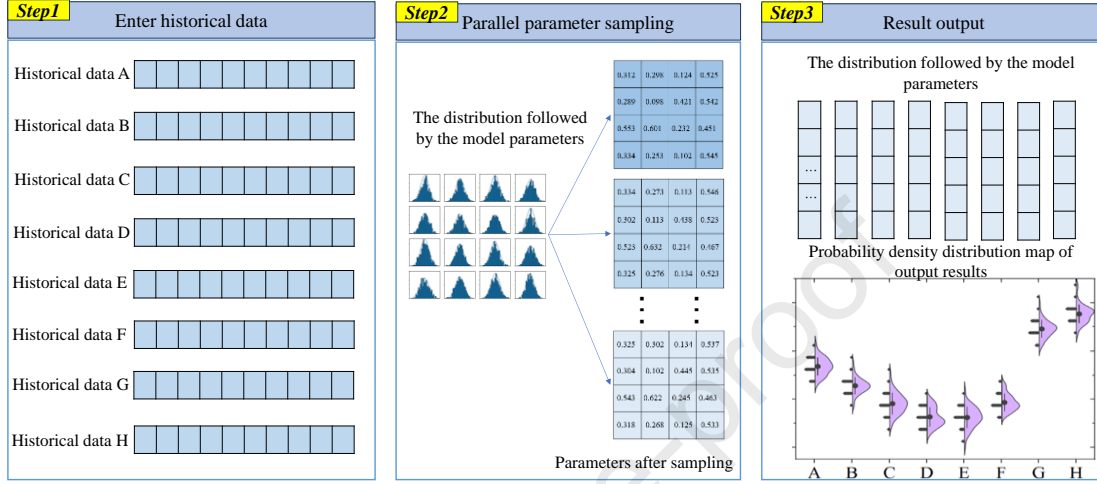


Figure 7 The principle diagram of uncertainty quantification

The quantification of uncertainty is characterized by the probability density distribution of the predicted results. When there are fewer uncertain factors such as noise and measurement errors in the data, the probability density distribution corresponding to the parameters of the TB-BiGRU model will become tighter after training, and the probability density distribution corresponding to the predicted results of the model will also become denser; When the uncertainty factors in the data increase significantly, the probability density distribution corresponding to the trained parameters of the model will become sparse, and the probability density distribution of the predicted results will also become sparse. This article comprehensively analyzes the stability, accuracy, and uncertainty quantification ability of the proposed TB-BiGRU model in Chapter 5. Compared with other models, TB-BiGRU has a wider range of application scenarios and value.

4. Degradation trend prediction evaluation metrics

Evaluation metrics can effectively and intuitively measure the quality of model predictions, and good evaluation metrics are essential for practical applications. This section introduces the evaluation metrics used in this paper, including commonly used point estimation metrics and the proposed evaluation metrics for quantifying uncertainty representation.

4.1 Point Estimation Evaluation Metrics

Point estimation evaluation metrics are one of the most commonly used standards for measuring model accuracy. In this paper, we introduce the mean absolute error (MAE) and the root mean square error (RMSE) as evaluation metrics. For these metrics, smaller values indicate higher prediction accuracy. The relevant calculation expressions are as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

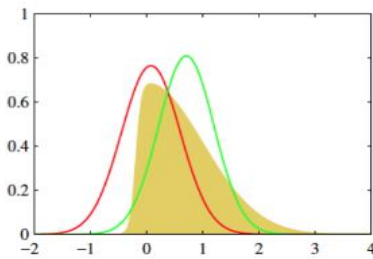
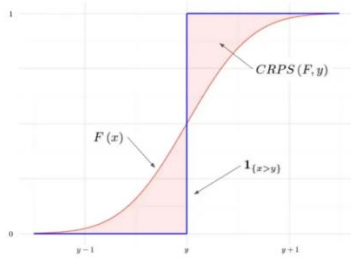
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (16)$$

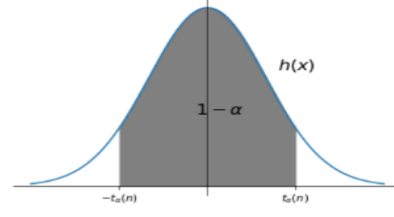
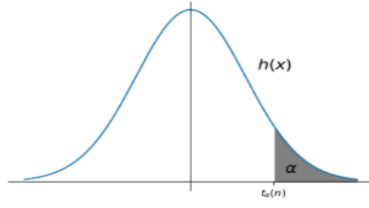
Where, y_i represents the actual result, \hat{y}_i is the predicted result, and n is the total number of predicted results.

4.2 Uncertainty Quantification Evaluation Metrics

As this paper involves uncertainty quantification research, the above-mentioned evaluation metrics may not fully meet the comparison requirements. Thus, we introduce four more practical uncertainty quantification evaluation metrics: Gaussian Negative Log Likelihood (G-NLL), Continuous Ranked Probability Score (CRPS), *check_score* and *Interval_score*. The descriptions of these evaluation metrics can be found in Table 4, and their corresponding calculation expressions are as follows:

Table 4. The related descriptions of uncertainty quantification evaluation metrics.

G-NLL	CRPS
	
<p>G-NLL characterizes the accuracy of the model by calculating the cumulative error between the model parameter distribution and the real sample data distribution. The smaller the value, the more accurate the model is.</p>	<p>CRPS characterizes the accuracy of the prediction results by calculating the area between the cumulative distribution function of the prediction results and the cumulative distribution function of the actual samples.</p>
<i>check_score</i>	<i>Interval_score</i>



check_score represents the probability that the actual result falls outside the confidence interval by calculating the sum of the differences between the cumulative distribution function and the actual result at different quantiles.

Interval_score represents the accuracy of the confidence interval of the prediction result by calculating the sum of the differences between the upper and lower quantiles of the cumulative distribution function and the actual results. The smaller the value, the more accurate and credible the confidence interval.

1) G-NLL

Assuming at time step i , the distribution followed by the predicted results of the model is $1/\sqrt{2\pi}\sigma_i \exp[-(y_i - \mu_i)^2 / 2\sigma_i^2]$. The *NLL* of predicted result can be expressed as:

$$\begin{aligned} NLL &= -\sum_{i=1}^n \log\{1/\sqrt{2\pi}\sigma_i \exp[-(y_i - \mu_i)^2 / 2\sigma_i^2]\} \\ &= \sum_{i=1}^n \frac{1}{2} [\log \sigma_i^2 + \log 2\pi + (y_i - \mu_i)^2 / \sigma_i^2] \end{aligned} \quad (17)$$

Where n is the total number of predicted results, y_i is the actual result corresponding to the time step i , μ_i, σ_i are the mean and standard deviation corresponding to the probability distribution of the predicted results at the time step i .

2) CRPS

The purpose of CRPS is to measure the difference between the probability density distribution of the predicted results and the actual observed values. When the distribution of the predicted results follows a Gaussian distribution, CRPS can be expressed as:

$$CRPS = \sum_{i=1}^n \sigma_i \left(\frac{y_i - \mu_i}{\sigma_i} (2\Phi(\frac{y_i - \mu_i}{\sigma_i}) - 1) + 2\phi(\frac{y_i - \mu_i}{\sigma_i}) - \frac{1}{\sqrt{\pi}} \right) \quad (18)$$

Normally, the distribution of predicted results is not a standard normal distribution, and it is necessary to standardize the difference between the actual value y_i and the predicted mean μ_i to obtain the standardized difference $(y_i - \mu_i) / \sigma_i$. At the same time, it is also necessary to standardize the cumulative distribution function (CDF) and

probability density function (PDF) of the prediction results. The standardized CDF and PDF are $2\Phi((y_i - \mu_i) / \sigma_i) - 1$ and $2\phi((y_i - \mu_i) / \sigma_i)$. Where $\Phi(\bullet)$ is the cumulative distribution function of the standard normal distribution corresponding to the time step i , and $\phi(\bullet)$ represents the standard normal distribution at the time step i . The purpose of constant term $1/\sqrt{\pi}$ is to adjust the value of CRPS to better meet the evaluation criteria in practical applications.

3) *check_score*

check_score is used to evaluate the predictive accuracy of the model at different quantiles. By calculating the errors at each quantile and combining them with mask values, the performance of the model can be more comprehensively measured. The calculation formula is as follows:

$$check_score = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{99} \sum_{q=1}^{99} \left(\phi_{q/100} \left(\frac{y_i - \mu_i}{\sigma_i} \right) - y_i \right) * mask_{iq} \right] \quad (19)$$

For the convenience of calculation, this formula also needs to standardize the difference between the actual value y_i and the predicted mean μ_i to obtain the standardized difference $(y_i - \mu_i) / \sigma_i$. $\phi_{q/100}((y_i - \mu_i) / \sigma_i)$ is the standard normal distribution function value at quantile $q/100$ at the time step i . Where $q/100$ is the quantile, $0 < q/100 < 1$, $mask_q$ is the coefficient corresponding to the q -th quantile at the time step i . The function of the mask is to selectively weight the error of each observation value, and they satisfy the following equation:

$$mask_{iq} = Q - \frac{q}{100}, Q = \begin{cases} 1, \phi_{q/100} \left(\frac{y_i - \mu_i}{\sigma_i} \right) - y_i \geq 0 \\ 0, others \end{cases} \quad (20)$$

4) *Interval_score*

Interval_score is mainly used to evaluate the accuracy and coverage of the model's prediction interval, measuring the quality of the prediction interval by considering errors within and outside the interval. The calculation formula is as follows:

$$\begin{aligned} Interval_score = & \frac{1}{n} \sum_{i=1}^n \frac{1}{99} \sum_{q=1}^{99} \left\{ \phi_{hq} \left(\frac{y_i - \mu_i}{\sigma_i} \right) - \phi_{lq} \left(\frac{y_i - \mu_i}{\sigma_i} \right) \right. \\ & + \frac{2}{1 - q/100} \left[\phi_{lq} \left(\frac{y_i - \mu_i}{\sigma_i} \right) - y_i \right] * below_l_q \\ & \left. + \frac{2}{1 - q/100} \left[y_i - \phi_{hq} \left(\frac{y_i - \mu_i}{\sigma_i} \right) \right] * above_u_q \right\} \end{aligned} \quad (21)$$

The different parts of the formula handle error within and outside the interval

separately, and adjust the coefficients to ensure the rationality and accuracy of the scoring. The error calculation within the interval is represented as $\phi_{hq}((y_i - \mu_i)/\sigma_i) - \phi_{lq}((y_i - \mu_i)/\sigma_i)$, which is the difference between the standard normal distribution function value corresponding to the right percentile and the standard normal distribution function value corresponding to the left percentile. For actual values outside the lower and upper limits of the prediction interval, calculate the relative error adjustment terms separately and multiply them by the adjustment coefficient $2/(1-q/100)$. Where, $q/100$ is the quantile, $lq = 0.5 - q/200$ and $hq = 0.5 + q/200$ are symmetric left and right quantiles, and they satisfy $0 < lq < hq < 1$. $below_l_q, above_u_q$ are the left and right quantile coefficients corresponding to the q -th quantile at the time step i , and they satisfy the following equation:

$$below_l_q = \begin{cases} 1 \cdot \phi_{lq/99}(\frac{y_i - \mu_i}{\sigma_i}) - y_i > 0 \\ 0, others \end{cases} \quad (22)$$

$$above_u_q = \begin{cases} 1, y_i - \phi_{hq/99}(\frac{y_i - \mu_i}{\sigma_i}) > 0 \\ 0, others \end{cases} \quad (23)$$

5. Performance Analysis of TB-BiGRU Model

This section provides an in-depth analysis of the prediction outcomes generated by the proposed model. Two different models, TB-BiGRU and TB-GRU, were developed using the dataset from the dynamic durability test of fuel cells introduced in Section 2. This section includes the performance comparison between TB-GRU and TB-BiGRU, the uncertainty quantification performance analysis of TB-BiGRU, efficiency and reliability analysis of TB-BiGRU model, and comparison of multiple uncertainty quantification models.

5.1 Performance Comparison between TB-GRU and TB-BiGRU

In this section, we developed the TB-BiGRU and TB-GRU models and applied them to predict the degradation trend of PEMFCs. The parameter settings for the models are shown in Tables 2 and 3. The dynamic and steady-state datasets in this study were divided in a 5:5 ratio. The input feature is the stack voltage, and the output is also voltage. The prediction results of TB-GRU and TB-BiGRU based on dynamic and steady-state conditions are shown in Table 5.

Table 5. The prediction results of TB-GRU and TB-BiGRU

Model	TB-GRU		TB-BiGRU	
	MAE	RMSE	MAE	RMSE
Working conditions				
dynamic	0.032017	0.040528	0.020081	0.025903
0A	0.001563	0.004106	0.001523	0.003961
1.76A	0.003155	0.00366	0.002653	0.003322
4.42A	0.003899	0.00527	0.003353	0.004565
9.48A	0.003034	0.005307	0.002716	0.005098
10.37A	0.002383	0.005457	0.002363	0.004765
14.81A	0.004821	0.007083	0.00268	0.005274
20.70A	0.003769	0.006023	0.003433	0.005906
29.58A	0.003743	0.006851	0.002468	0.005303
35.53A	0.003688	0.006545	0.002105	0.004926

Table 5 shows the errors in the prediction results of TB-GRU and TB-BiGRU models under dynamic and nine different steady-state conditions, in order to more comprehensively analyze the advantages and disadvantages of the TB-BiGRU model compared to the TB-GRU model. From the prediction results in Table 5, it can be seen that compared with the TB-GRU model, the TB-BiGRU model has improved in terms of MAE and RMSE. Under dynamic conditions, the MAE and RMSE of the TB-BiGRU model are 0.020081 and 0.025903, respectively, which are 37.28% and 36.09% higher than those of the TB-GRU model. This is mainly due to its bidirectional access to data information model structure and robust data processing capabilities, making the TB-BiGRU model more suitable for dynamic condition prediction. Under steady-state conditions, the TB-BiGRU model showed the most significant improvement in prediction accuracy at 14.81A. Compared with the TB-GRU model, the MAE and RMSE increased by 44.41% and 25.54%, respectively. For other operating conditions, the prediction accuracy of the TB-BiGRU model has also been improved to varying degrees. For ease of observation, we use a bar chart to visualize the prediction error of the model under different operating conditions, as shown in Figure 8.

Figure 8 shows ten operating conditions of PEMFC, each of which was predicted using eight models including TB-GRU and TB-BiGRU models (the other six models are TB-LSTM, MLP, LSTM, GRU, TCN, and BiGRU). The model with the best prediction results for each operating condition is labeled in the figure. From the annotation results, both MAE and RMSE show that the TB-BiGRU model proposed in

this article performs the best. In addition, based on the prediction results of each operating condition, there is basically a pattern of model prediction accuracy, namely TB-BiGRU>TB-GRU>TB-LSTM>BiGRU>GRU>TCN>MLP. From this, it can be seen that the three models combined with the TB algorithm have all improved compared to their original models, mainly due to the TB algorithm's ability to reconstruct the model parameters and quantify the uncertainty obtained by the model, thereby improving the model's ability to handle noisy data.

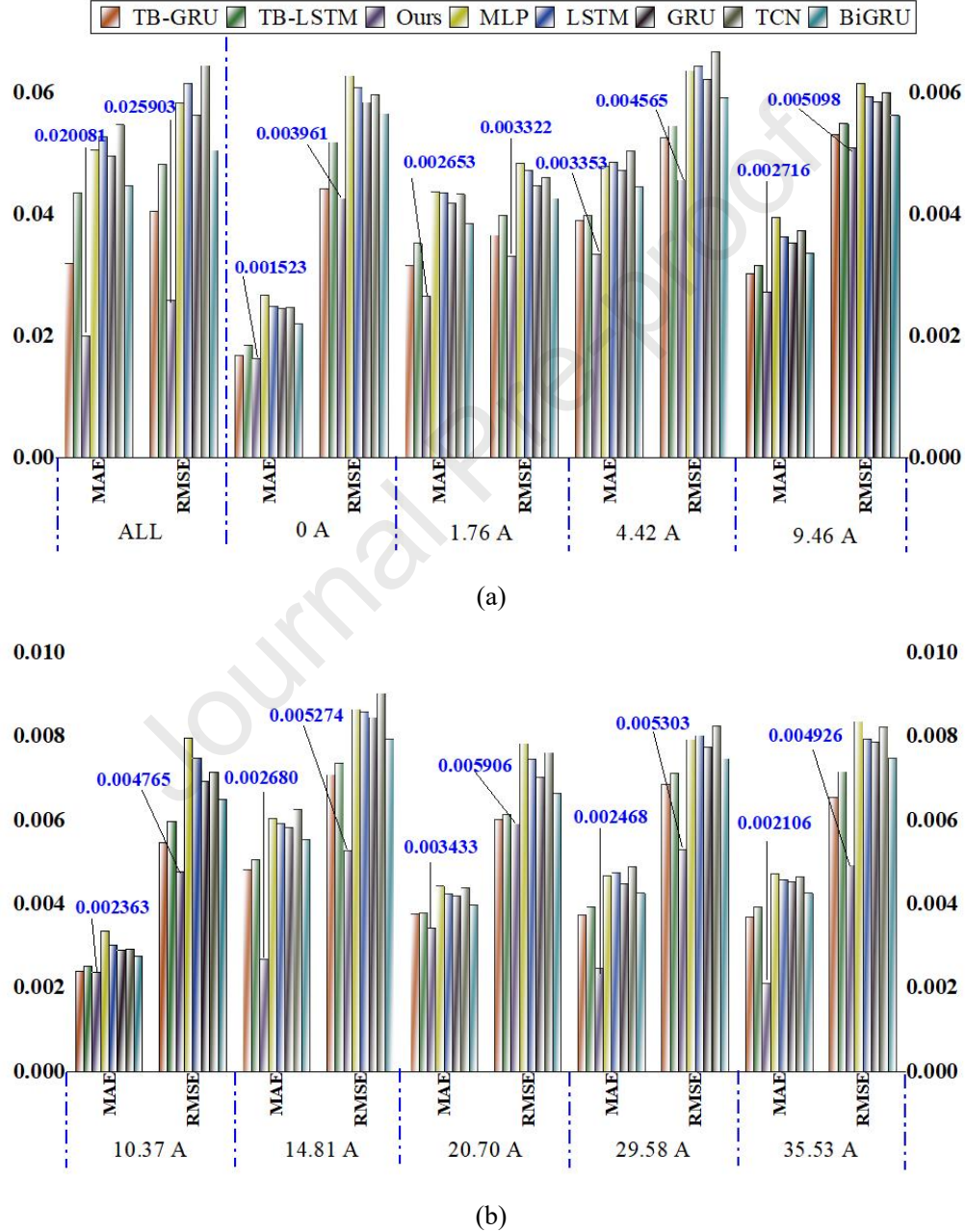


Figure 8. Comparison of prediction errors of multiple learning models under different operating conditions of PEMFC

In order to further demonstrate the performance improvement brought by uncertainty quantification to the model, this section uses exponential weighted averaging to perform varying degrees of noise reduction on the data collected from dynamic and partially static operating conditions. Exponential weighted average achieves data smoothing by controlling the weighting coefficient. The smaller the coefficient, the smoother the processed data and the less uncertainty factors it contains. In order to preserve some uncertain factors in the data, the weighting coefficients should not be set very small. During the experiment, five groups were set, namely 0.5, 0.6, 0.7, 0.8, and 0.9. After sending the corresponding smoothed data for each group to different models for training, calculate the MAE and RMSE of each model's prediction results, and plot the calculation results as Figure 9.

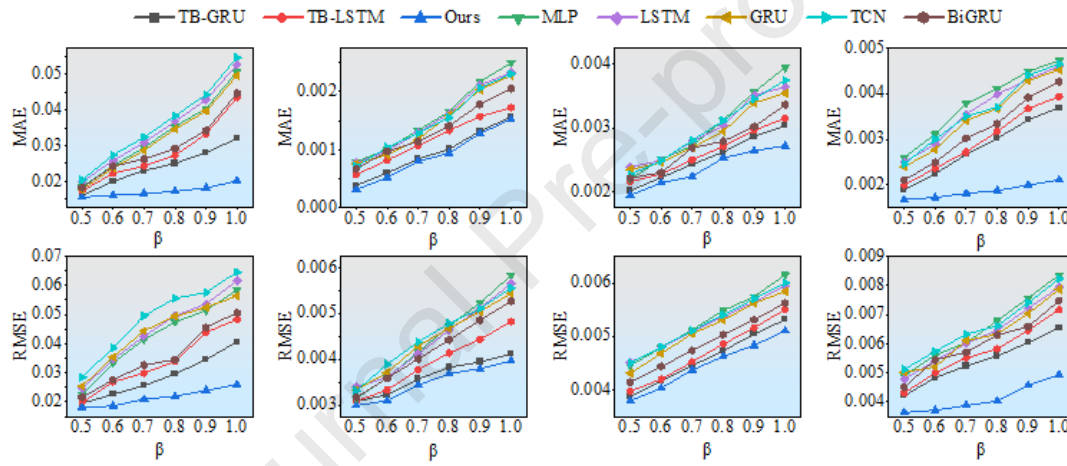


Figure 9 shows the variation trend of prediction errors of eight models under dynamic conditions, 0A, 9.46A, and 35.53A

Figure 9 provides a detailed comparison of the prediction results of eight models under four different operating conditions. As the MAE and RMSE increase, both show a significant upward trend. This is because the larger the MAE, the more noise factors are included in the data, which can interfere with the prediction process of the model and lead to a decrease in the accuracy of the prediction results. Although uncertain factors such as data noise have a significant impact on the prediction accuracy of deep learning models, the TB-BiGRU model proposed in this paper demonstrates better noise resistance performance compared to the other seven models. Under the dynamic operating conditions of PEMFC, with the increase of noise factors, the upward trend of MAE and RMSE predicted by TB-BiGRU is significantly slower than other models; Under the operating conditions of load currents of 0A, 9.46A, and 34.43A, the error of the predicted results also shows the same trend, which is particularly evident under the 35.53A condition. This experimental result indicates that the TB-BiGRU model

proposed in this paper has better noise resistance performance under various PEMFC operating conditions. This is due to its bidirectional access to data information model structure and robust data processing capabilities. On the other hand, the model parameters that follow Gaussian distribution can greatly improve the model's ability to process noisy data compared to point based model parameters.

It is worth mentioning that the TB-BiGRU model has lower prediction errors than the other seven models under different noise levels in different operating conditions, further demonstrating the stability and accuracy of the TB-BiGRU model. In practical application scenarios, the collected data inevitably contains uncertain factors such as noise. However, the TB-BiGRU model has better noise resistance and good uncertainty quantification ability, and has greater application value compared to other deep learning models.

5.2 Interval estimation performance analysis of TB-BiGRU

Interval estimation is an important aspect of uncertainty quantification. This section analyzes the interval estimation performance of the TB-BiGRU model based on various operating conditions, and verifies the effectiveness of four proposed uncertainty quantification indicators. The interval estimation of the TB-BiGRU model is shown in Fig. 10 and 11.

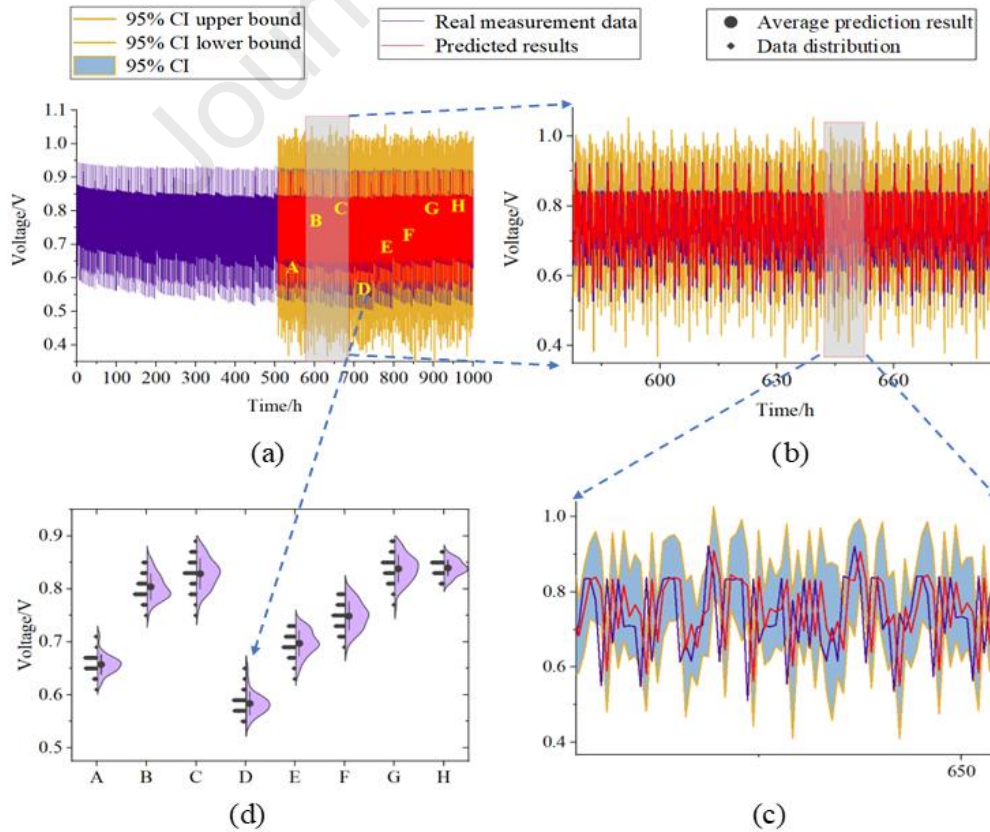


Figure 10. Predictive effect diagram of TB-BiGRU under dynamic operating conditions. 10(a) Overall effect prediction diagram of TB-BiGRU based on dynamic operating conditions; 10(b) Partial enlargement of 11(a); 10(c) Partial enlargement of 10(b); 10(d) Probability density distribution diagram of predicted results at points A-H in 10(a)

Due to the dynamic operating conditions encompassing the operational data of all conditions in FC-DLC, this paper zooms in on Fig. 10(a) twice to obtain Fig. 10(c). From Fig. 10(c), it can be observed that the mean of TB-BiGRU prediction results closely follows the actual results, and the actual results mainly fall within the 95% confidence interval of the TB-BiGRU predictions. This indicates that the TB-BiGRU model possesses good interval estimation capabilities. In order to visually demonstrate the interval estimation capabilities of the TB-BiGRU model, this paper performs equidistant sampling on the predicted results and presents the probability distribution of the sampled prediction results, as shown in Fig. 10(d). A-H represent the positions of the sampled points, with each sampled point corresponding to a probability distribution. It can be observed from the figure that the probability distribution of the predicted results basically follows a Gaussian distribution, which corresponds to the structure of the model. Through the probability density, one can more intuitively see the distribution of the predicted results and the deviation between the actual results and the predicted mean. The results in Fig. 10(d) demonstrate that the TB-BiGRU model exhibits excellent interval estimation capabilities under dynamic operating conditions.

Fig. 11 includes predictive results for various individual operating conditions. From the Fig. 6, it can be observed that for most operating conditions, the confidence intervals of TB-BiGRU model predictions cover the actual results, such as conditions 0A, 14.81A, and 20.70A. This result aligns with the evaluation metrics in Table 6, where smaller values of NLL, CRPS, *check_score* and *Interval_score* indicate better uncertainty quantification. When the condition is 0A, the smallest values for NLL, CRPS, *check_score* and *Interval_score* are 3.484554, 0.001279, 0.000643, and 0.009593, respectively. Conversely, when the condition is 29.58A, the largest values for NLL, CRPS, *check_score* and *Interval_score* are 13.88969, 0.004294, 0.002157, and 0.036082, respectively. Based on the evaluation metrics, the uncertainty quantification performance for condition 0A is significantly superior to that of condition 29.58A. Looking at Fig. 11(a) and 11(d), it can be observed that the coverage of the confidence intervals in Fig. 11(a) is better than that in Figure 11(d), indicating that the uncertainty quantification performance for condition 0A is better.

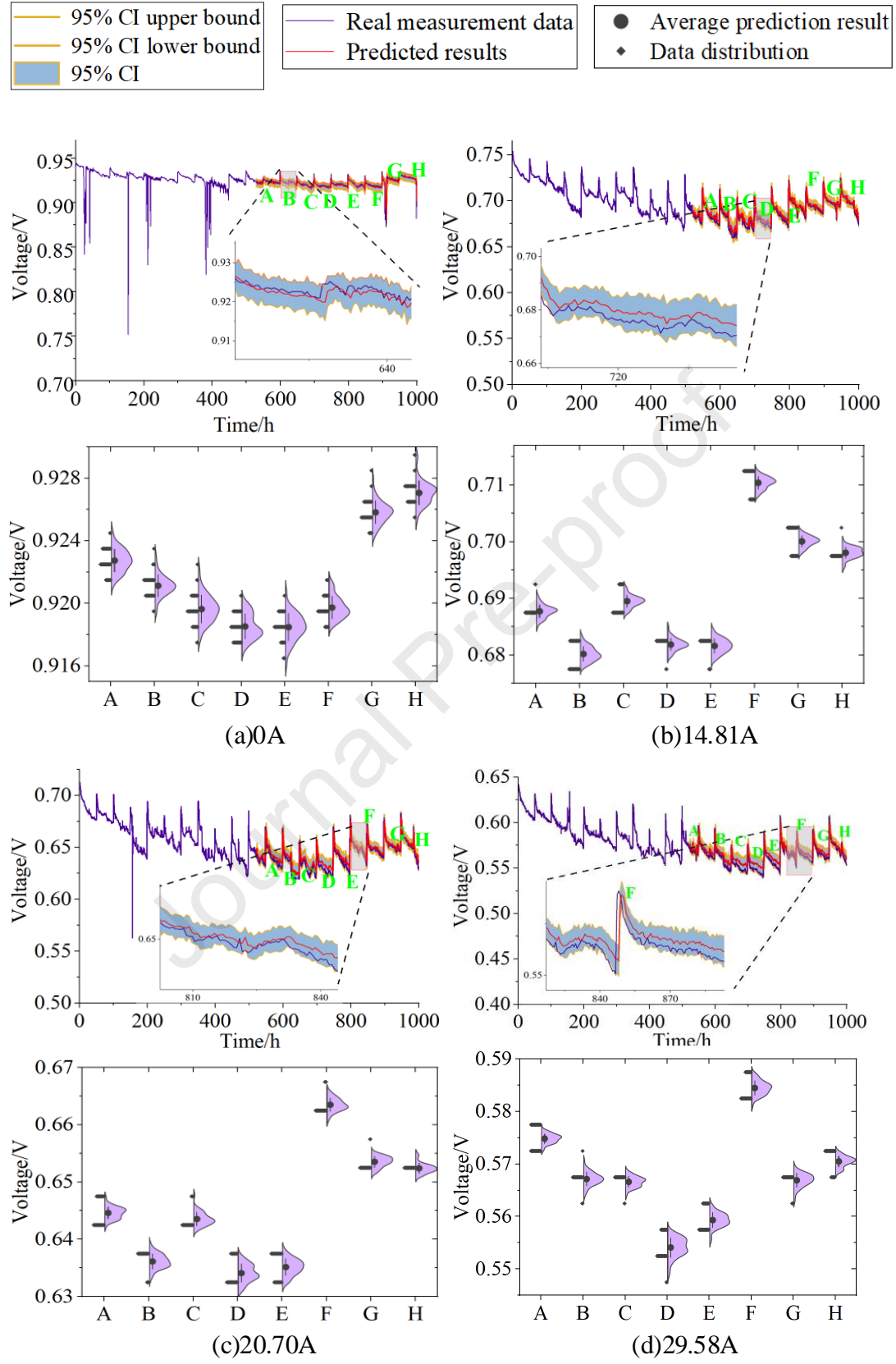


Figure 11. Predictive effect diagram of TB-BiGRU based on different steady-state operating conditions

Simultaneously, the predictive effect diagrams for other operating conditions correspond to their respective evaluation metric results. This further demonstrates that the proposed model in this paper exhibits excellent uncertainty quantification capabilities under both dynamic and various steady-state operating conditions. It is more suitable for practical applications with a high degree of uncertainty. Additionally, the uncertainty quantification evaluation metrics proposed in this paper can effectively capture the differences in predictive performance under different operating conditions, providing a good representation of the model's uncertainty quantification capabilities.

Table 6. The uncertainty quantification results of TB-BiGRU

Working conditions	NLL	CRPS	<i>check _ score</i>	<i>Interval _ score</i>
All	18.3282	0.101696	0.051055	0.883734
0A	3.484554	0.001279	0.000643	0.009593
1.76A	9.221416	0.002362	0.001186	0.019859
4.42A	13.51943	0.003553	0.001783	0.031383
9.48A	4.244665	0.002226	0.00112	0.016078
10.37A	4.093864	0.002014	0.001013	0.015698
14.81A	4.350689	0.003094	0.001556	0.022578
20.70A	6.525078	0.002886	0.001451	0.021872
29.58A	13.88969	0.004294	0.002157	0.036082
35.53A	4.920571	0.003033	0.001526	0.021706

5.3 Efficiency and Reliability analysis of TB-BiGRU model

Based on the findings presented in Section 5.1 and 5.2, it is clear that the TB-BiGRU model demonstrates a notable enhancement in performance when compared to the initial TB-GRU model, and it demonstrates good uncertainty quantification performance. This section discusses the computational efficiency of the TB-BiGRU model under different sampling times and the reliability of the model prediction results.

For most uncertainty quantification models, setting the number of samples is essential. The more samples taken, the more accurate the probability density distribution of the prediction results becomes, and the more reliable the confidence intervals for the uncertainty quantification results are. However, this significantly increases the model's reasoning time, which is detrimental to real-time predictions of fuel cell degradation trends. To enhance the real-time performance and practical significance of the TB-BiGRU model, we utilized the multithreading capabilities of multi-core CPUs and GPUs to switch the model from sequential sampling to parallel

sampling without changing the hardware. This greatly improved the reasoning time during the model's prediction process. The relationship between reasoning time and the number of samples for sequential and parallel sampling under dynamic operating conditions of fuel cells is shown in Fig. 12(a). Additionally, we plotted the relationship between the RMSE of the prediction results and the reasoning time as the number of samples increases, as shown in Fig. 12(b).

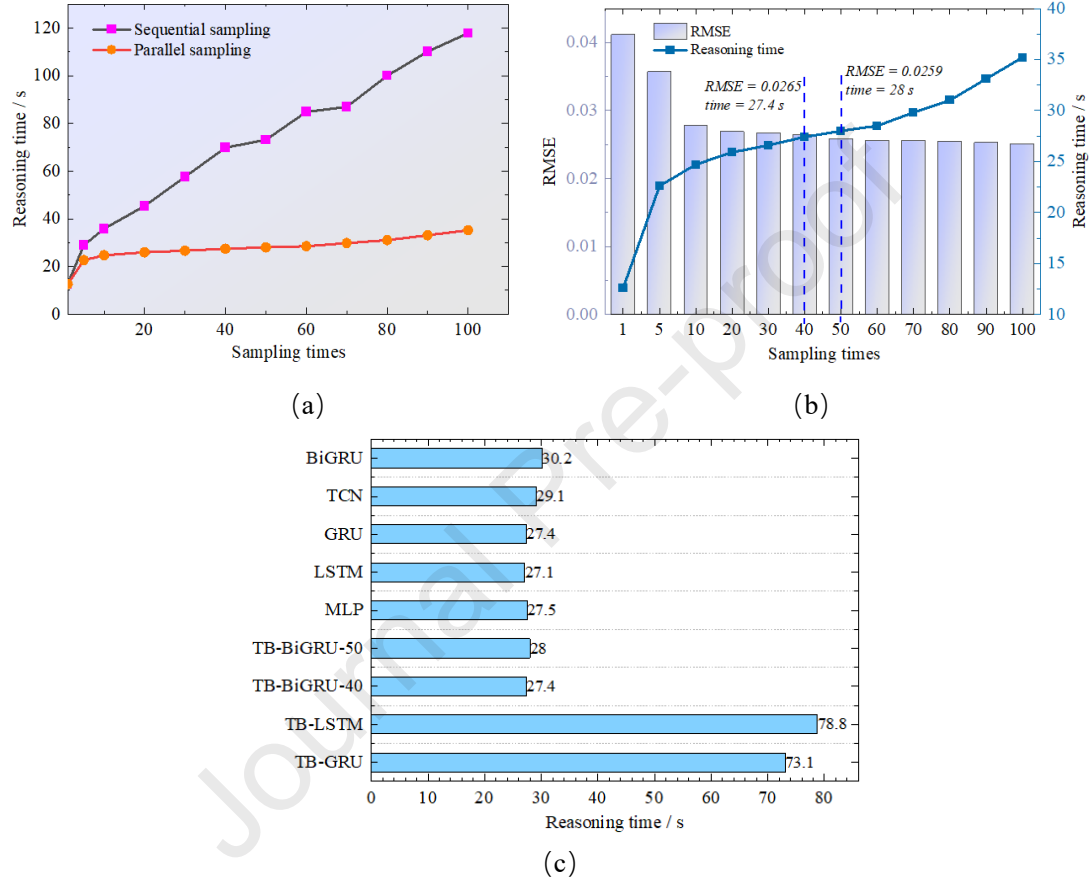


Figure 12 Analysis of the computational efficiency of the TB-BiGRU model under dynamic operating conditions

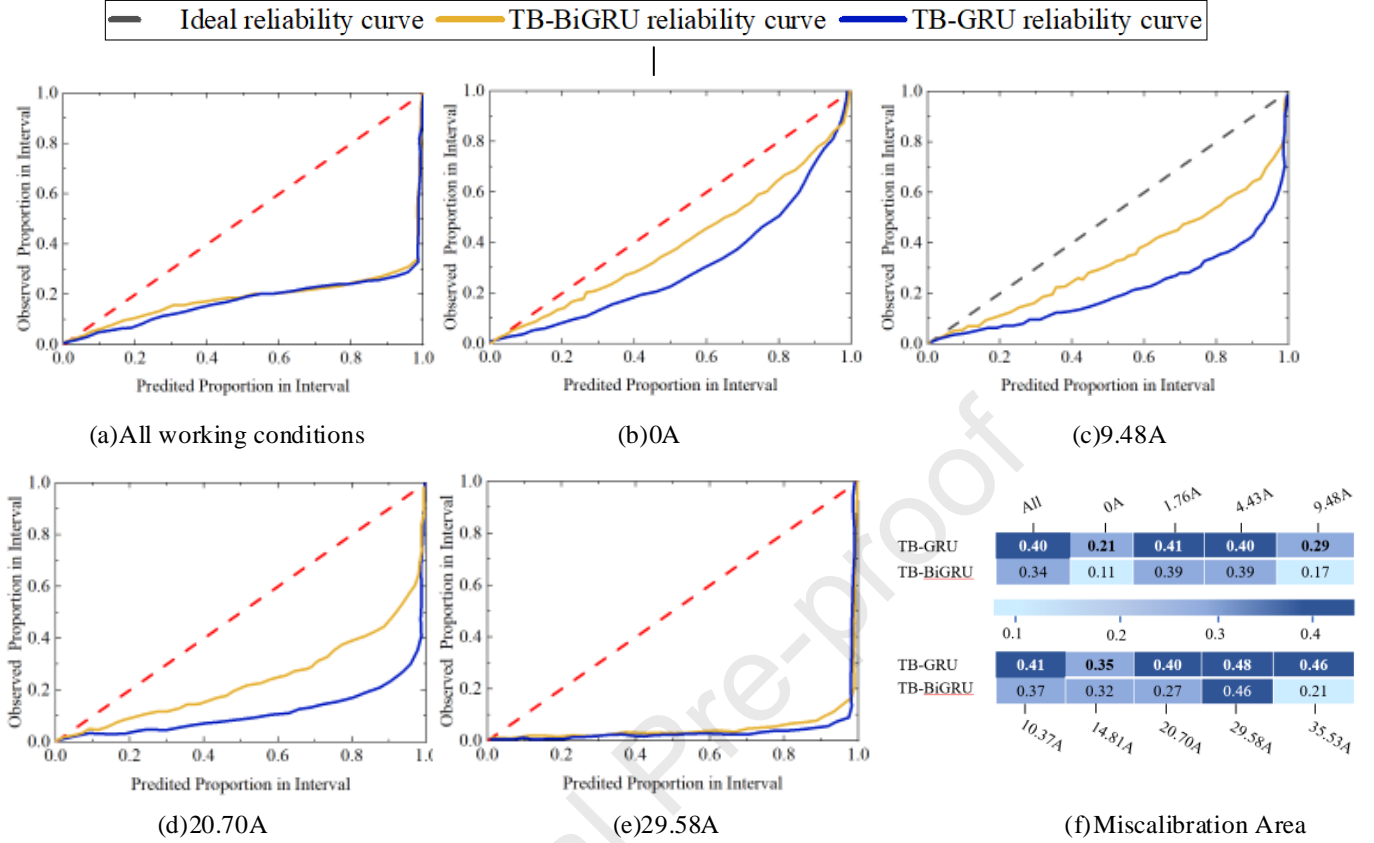
As shown in Fig. 12(a), when the number of samples exceeds five, the reasoning time of sequential sampling increases exponentially with the number of samples, significantly higher than that of parallel sampling. This indicates that parallel sampling has a clear advantage in reducing the reasoning time of the TB-BiGRU model, thereby effectively enhancing the model's real-time performance in predicting fuel cell degradation trends. Meanwhile, Fig. 12(b) shows that as the number of samples increases, the RMSE of the prediction results exhibits a noticeable downward trend. When the number of samples exceeds 40, this downward trend slows down, as the probability density distribution of the prediction results becomes more complete and the corresponding mean stabilizes. However, as the number of samples increases, the

model's reasoning time also grows accordingly. To balance real-time performance and prediction accuracy, the number of samples must be selected carefully. The intersection region of the RMSE bar chart and the reasoning time curve in Fig. 12(b) represents an optimal choice, with the number of samples set to 40 or 50. Fig. 12(c) further compares the reasoning time of the TB-BiGRU model with other data-driven models at 40 and 50 samples. The results show that the reasoning times of the TB-BiGRU model at these two sample sizes are 27.4 seconds and 28 seconds, respectively, which are significantly lower than those of the TB-LSTM and TB-GRU models, and comparable to the reasoning times of traditional LSTM and GRU models. This further demonstrates that the TB-BiGRU model not only maintains strong real-time performance but also ensures prediction accuracy, making it highly promising for practical applications.

While balancing real-time performance and accuracy, the reliability of the model should also be a key consideration. Therefore, we evaluated the uncertainty quantification performance of the TB-BiGRU model using reliability diagrams and compared it with the TB-GRU model. In a short, reliability curves indicate the ability of a machine learning model to quantify uncertainty on a relevant data set. A model that provides precise predictions of uncertainty at every confidence level will exhibit a reliability curve nearing $y = x$. To better assess the uncertainty quantification performance of TB-GRU and TB-BiGRU models, this section continues to make predictions based on dynamic and different steady-state operating conditions. In order to take into account the accuracy and real-time performance of the TB-BiGRU model, the calculated reliability curves are obtained by averaging the predictions from 50 samples in the prediction process. For enhanced comparative analysis, this paper calculates the area enclosed between the reliability curve and the diagonal line, referred to as the Miscalibration Area (MA), as shown in Fig. 13.

In Fig. 13, the orange curve represents the reliability curve for TB-BiGRU, while the blue curve represents the reliability curve for TB-GRU. The figure illustrates a consistent trend where the reliability curve of the TB-GRU model consistently lags behind that of the TB-BiGRU model, and its MA results are consistently higher than those of TB-BiGRU. Overall, whether in the full operating conditions or in individual operating conditions, TB-GRU model's metrics such as MAE, RMSE, NLL, CRPS, *check_score* and *Interval_score* are almost always smaller than those of the TB-BiGRU model. This result indicates that the TB BiGRU model with bidirectional structure is more suitable for quantifying the uncertainty in the degradation trend of

PEMFC; On the other hand, the reliability curve and the results of MA further validate



the effectiveness of the proposed uncertainty quantification evaluation metrics.

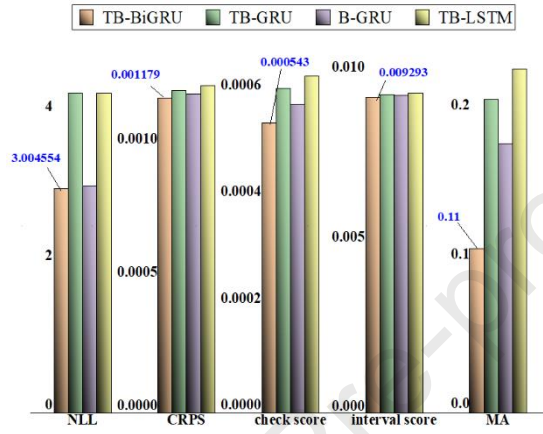
Figure 13. Reliability curves and miscalibration area of TB-BiGRU and TB-GRU

5.4 Comparison of multiple uncertainty quantification models

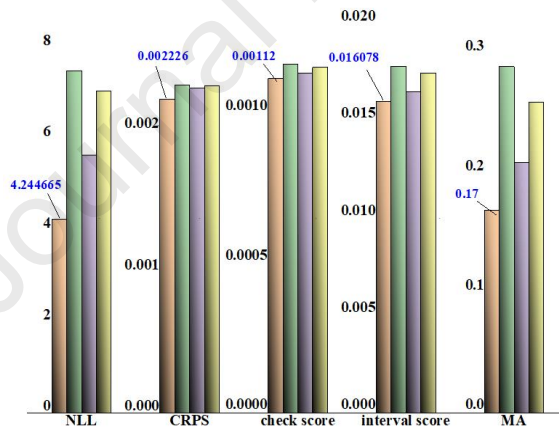
The results of Section 5.2 validate the effectiveness of the uncertainty quantification indicators proposed in this article. In this section, we calculated the uncertainty quantification indicators for four models: TB-BiGRU, TB-LSTM, TB-GRU, and B-GRU, under PEMFC no-load (0 A), light load (9.38 A), and heavy load (35.53 A) operating conditions. These results are shown in Fig. 13 for comparison.

From Fig.14, it can be observed that the uncertainty quantification index of the TB-BiGRU model is smaller than that of the other three models under three different operating conditions, followed by the B-GRU model. Further observation reveals that the uncertainty quantification metrics of the PEMFC are smaller under no-load and heavy-load conditions compared to light-load operation. This is primarily due to the accidental uncertainty during PEMFC operation. Accidental uncertainty and cognitive uncertainty are the main sources of uncertainty in uncertainty quantification work. Under the same model structure, cognitive uncertainty factors remain unchanged, but

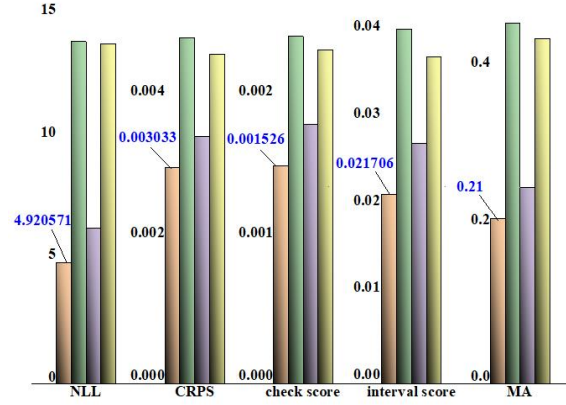
the accidental uncertainty factors present in data collected under different working conditions vary. When the accidental uncertainty factors in the operating conditions are reduced, the uncertainty quantification index of the prediction results will be smaller, resulting in a more compact interval estimation. Conversely, when there are more accidental uncertainty factors in the data, the uncertainty quantification index of the prediction results increases, leading to a sparser interval estimation and more uncertainty in the prediction results.



(a) 0 A



(b) 9.38 A



(c) 35.53 A

Figure 14 Comparison of uncertainty quantification capabilities of various uncertainty quantification models under different operating conditions

6. Conclusion

To address the performance degradation of fuel cells under dynamic operating conditions, this paper proposes a performance prediction method that implements uncertainty quantification based on TB algorithm. This method combines TB with BiGRU models, modifying the internal parameter structure of the model by replacing fixed internal parameters with random variables following a Gaussian distribution. This achieves uncertainty quantification of the prediction results, enhancing the credibility of the predictions, and demonstrating practical application value. The proposed approach's efficacy has been confirmed through the utilization of a dynamic load cycling dataset. Based on a comparative analysis of the prediction outcomes, the following deductions can be made:

1) The accuracy of the TB-BiGRU model shows a significant improvement compared to the TB-GRU. Under dynamic operating conditions, TB-BiGRU's MAE and RMSE are enhanced by 37.28% and 36.09%, compared to TB-GRU. Under different steady-state operating conditions, compared with other deep learning models, TB-BiGRU exhibits better noise performance and stability under different operating conditions and noise levels.

2) Interval estimation represents the range of prediction results through probability density distribution, which not only improves the accuracy of point estimation, but also improves the confidence of the results.

3) The four proposed evaluation indicators can accurately capture the uncertainty quantification ability of the TB-BiGRU model under different working conditions, and

have more practical application value.

4) Compared with other uncertainty quantification models, TB-BiGRU performs well under different operating conditions of PEMFC. And the proposed prediction framework has scalability and is suitable for other deep models.

Although the TB-BiGRU model can provide more accurate and reliable uncertainty quantification results for predicting the degradation trend of fuel cells, the model structure is more complex compared to the original BiGRU, and the establishment process involves variational inference, which may increase the computational resources consumed in model training and prediction. To achieve this, it is necessary to minimize the complexity of the model as much as possible without affecting its prediction accuracy and credibility. Based on the research in this paper, we will further study a more lightweight uncertainty quantification model to achieve accuracy, credibility, and real-time prediction of fuel cell degradation trends. In addition, most deep learning models currently lack interpretability, and the TB-BiGRU model is no exception. We are currently researching an interpretable uncertainty quantification model to expand its interpretability.

References:

- [1] Xinyu Lu, Banghua Du, Wenchao Zhu, Changjun Xie, et al. Thermodynamic and dynamic analysis of a hybrid PEMFC-ORC combined heat and power (CHP) system. *Energy Convers and Manag* 2023;297:117408.
- [2] Xu Hao, Yuebo Yuan, Hewu Wang, Minggao Ouyang. Plug-in hybrid electric vehicle utility factor in China cities: influencing factors, empirical research, and energy and environmental application. *eTransportation* 2021;10:100138.
- [3] Jianxing Liu, Yabin Gao, Xiaojie Su, et al. Disturbance-observer-based control for air management of PEM fuel cell systems via sliding mode technique. *IEEE Trans Control Syst Technol* May 2018;27(3):1129-1138,
- [4] Zunyan Hu, Liangfei Xu, Jianqiu Li, Minggao Ouyang, et al. A reconstructed fuel cell life prediction model for a fuel cell hybrid city bus. *Energy Convers Manag* 2018;156(14):723-732.
- [5] Hujun Peng, Zhu Chen, Kai Deng, et al. A comparison of various universally applicable power distribution strategies for fuel cell hybrid trains utilizing component modeling at different levels of detail: from simulation to test bench measurement. *eTransportation* 2021;9:100120.
- [6] Dario R. Dekel. Review of cell performance in anion exchange membrane fuel

- cells. *J Power Sources* 2018;375(31):158-169.
- [7] Jiawei Liu, Qi Li, Weirong Chen, et al. A discrete hidden Markov model fault diagnosis strategy based on K-means clustering dedicated to PEM fuel cell systems of tramways. *Int J Hydrogen Energy* 2018;43(27):12428-12441.
- [8] Chu Wang, Manfeng Dou, Zhongliang Li, et al. A fusion prognostics strategy for fuel cells operating under dynamic conditions. *eTransportation* 2022;12: 10016.
- [9] Marine. Jouin, Rafael Gouriveau, Daniel Hissel, et al. Prognostics of PEM fuel cell in a particle filtering framework. *Int J Hydrogen Energy* 2014;39(1):481-494.
- [10] Jian Zuo, Hong Lv, Daming Zhou, et al. Deep learning based prognostic framework towards proton exchange membrane fuel cell for automotive application. *Applied Energy* 2021;281: 115937.
- [11] Yashan Xing, Jing Na, Ramon Costa-Castello. Real-time adaptive parameter estimation for a polymer electrolyte membrane fuel cell. *IEEE Trans. Ind. Inform.* Nov 2019;15(11):6048–6057.
- [12] Saad Saleem Khan, Hussian Shareef, Mohsen Kandidayeni, et al. Dynamic semiempirical PEMFC model for prognostics and fault diagnosis. *IEEE Access* 2021;9:10217–10227.
- [13] Mohsen Kandidayeni, Hicham Chaoui, Sousso. Kelouwani, et al. Online system identification of a fuel cell stack with guaranteed stability for energy management applications. *IEEE Trans. Energy Convers.* Dec 2021;36(4) 2714–2723.
- [14] Yashan Xing, Jing Na, Mingrui Chen, et al. Adaptive nonlinear parameter estimation for a proton exchange membrane fuel cell. *IEEE Trans. Power Electron.* Aug 2022;37(8):9012–9023.
- [15] Hongwei Li, Boshi Xu, Changhe Du, et al. Performance prediction and power density maximization of a proton exchange membrane fuel cell based on deep belief network. *J. Power Sources* 2020;461:228154.
- [16] Hao Yang, Penglei Wang, Yabin An, et al. Remaining useful life prediction based on denoising technique and deep neural network for lithium-ion capacitors. *eTransportation* 2020;5:10078.
- [17] Rui Zhao, Zhenghua Chen, Kezhi Mao, et al. Deep learning and its applications to machine health monitoring. *Mech Syst Signal Process* 2019;115(15):213-237.
- [18] Rui Ding, Shiqiao Zhang, Yawen Chen, et al. Application of machine learning in optimizing proton exchange membrane fuel cells: a review. *Energy and AI* 2022;9:100170.

- [19] Chu Wang, Zhongliang Li, Rachid Outbib, et al. Symbolic deep learning based prognostics for dynamic operating proton exchange membrane fuel cells. *Appl Energy* 2022;305: 117918.
- [20] Rui Ma, Tao Yang, Elena Breaz, et al. Data-driven proton exchange membrane fuel cell degradation predication through deep learning method. *Appl Energy* 2018;231(1):102-115.
- [21] Wenxin Wan, Yang Yang, Yang Li, Changjun Xie, et al. Operating conditions combination analysis method of optimal water management state for PEM fuel cell. *Green Energy and Intelligent Transportation*;2023,2(4):100105.
- [22] Caixia Liu, Yong Chen, Renzong Xu, et al. Co-optimization of Energy Management and Eco-Driving Considering Fuel Cell Degradation Via Improved Hierarchical Model Predictive Control. *Green Energy and Intelligent Transportation*; 2024:100176.
- [23] Ruihan Wang, Hui Chen, Cong Guan. A Bayesian inference-based approach for performance prognostics towards uncertainty quantification and its applications on the marine diesel engine. *ISA (Instrum Soc Am) Trans* 2021;118:159–173.
- [24] Weiwen Peng, Zhisheng Ye, Nan Chen. Bayesian deep learning based health prognostics towards prognostics uncertainty. *IEEE Trans Ind Electron* 2019;2019(67):2283-2293.
- [25] Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature* 2015;521(7553):452-459.
- [26] Wenchao Zhu, Bingxin Guo, Yang Li, et al. Uncertainty Quantification of Proton-Exchange-Membrane Fuel Cells Degradation Prediction Based on Bayesian-Gated Recurrent Unit. *eTransportation* 2023; 16:100230.
- [27] Yuchen Xie, Jianxiao Zou, Chao Peng, et al. A novel PEM fuel cell remaining useful life prediction method based on singular spectrum analysis and deep Gaussian processes. *Int J Hydro Energy* 2020;45(55):30942-30956.
- [28] Ruihan Wang, Hui Chen, Cong Guan. A Bayesian inference-based approach for performance prognostics towards uncertainty quantification and its applications on the marine diesel engine. *ISA (Instrum Soc Am) Trans* 2021;118:159–173.
- [29] Weiwen Peng, Zhisheng Ye, Nan Chen. Bayesian deep learning based health prognostics towards prognostics uncertainty. *IEEE Trans Ind Electron* 2019;2019(67):2283-2293.
- [30] Jian Zuo, Hong Lv, Daming Zhou, et al. Long-term dynamic durability test datasets

for single proton exchange membrane fuel cell. Data Brief 35:106775.

- [31] Jian Zhao, Xionguo Li. A review of polymer electrolyte membrane fuel cell durability for vehicular applications: degradation modes and experimental techniques. Energy Convers Manage 2019;199:112022.
- [32] Mohammad Mahdi, Mehdi Ghatee. A systematic review on overfitting control in shallow and deep neural networks. Artificial Intelligence Review 2021;54: 6391-6438.

- A framework for predicting fuel cell degradation trends incorporating uncertainty quantification.
- Application of a Bayesian theory-based TB algorithm to a BiGRU model to achieve probability density distribution of model structural parameters.
- Introduction of four indicators to measure the model's uncertainty quantification capability.
- Provision of both point estimates and interval estimates with probability density distribution.
- The proposed model has better noise resistance, stability, and uncertainty quantification ability

Declaration of interests

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

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