A data-driven method for multi-step-ahead ~~degradation~~ prediction and long-term prognostics of proton exchange membrane fuel cell ~~using conditional dilated convolutional neural network~~

Khaled Benaggoune[a], Meiling Yue[b,1], Samir Jemei[c], Noureddine Zerhouni[a]

[a] *FEMTO-ST Institute, Univ. Bourgogne Franche-Comté, CNRS, Besançon, France*
[b] *School of mechanical, electronic and control engineering, Beijing Jiaotong University, Beijing, China*
[c] *FEMTO-ST Institute, FCLAB, Univ. Bourgogne Franche-Comté, CNRS, Belfort, France*

Abstract

Fuel cell technology has been rapidly developed in the last decade owing to its clean characteristic and high efficiency. Proton exchange membrane fuel cells (PEMFCs) are increasingly used in transportation applications and small stationary applications; however, the cost and the unsatisfying durability of the PEMFC stack have limited their successful commercialization and market penetration. In recent years, thanks to the availability and the quality of emerging data of PEMFCs, ~~digitalization~~ digitization is happening to offer possibilities to increase the productivity and the flexibility in fuel cell applications. Therefore, it is crucial to clarify the potential of ~~digitalization~~ digitization measures, how and where they can be applied, and their benefits. This paper focuses on the degradation performance of the PEMFC stacks and develops a data-driven intelligent method to predict both the short-term and long-term degradation. The dilated convolutional neural network is for the first time applied for predicting the time-dependent fuel cell performance and is proved to be more efficient than other recurrent networks. To deal with the long-term performance uncertainty, a conditional neural network is proposed. Results have shown that the proposed method can predict not only the degradation tendency, but also contain the degradation behavior dynamics.

Keywords: convolutional neural network, prognostics, proton membrane exchange fuel cell, time series prediction

Nomenclature

$ARIMA$	auto regressive integrated moving average
$BILSTM$	bi-directional long short term memory
CNN	convolutional neural network
CO_2	carbon dioxide

[1] Corresponding author. E-mail address: meiling.yue@femto-st.fr

EOL end-of-life

ESN echo state network

$GELU$ Gaussian error linear unit

$LSTM$ long short term memory

$MAPE$ mean absolute percentage error

MLP multilayer perceptron

$PEMFC$ proton exchange membrane fuel cell

PHM prognostics and health management

$RMSE$ root mean square error

RNN recurrent neural network

RUL remaining useful life

SOH state of health

1. Introduction

Electric vehicles are playing an important role in the third energy revolution and the health of our planet benefit significantly from the new "clean sources" with the reduction of carbon dioxide (CO_2) emissions in the atmosphere. Compared with the lithium battery industry chain, the hydrogen energy and fuel cell industry chain is more extensive, more complex, and has greater economic value [1]. Fuel cell electric vehicles benefit from their fast-fueling characteristic, long driving range and can eliminate the problem caused by battery recycling. Developing fuel cell electric vehicles can also promote the comprehensive utilization of hydrogen energy. ~~From practical applications of hydrogen energy and fuel cells, fuel cell electric vehicles are the most effective way to use hydrogen energy distributively and efficiently.~~ The current hydrogen energy industry chain has begun to take shape, and the performance of fuel cell systems ~~has~~ tends to meet the commercial requirements. However, its economy and practicability still restrict the large-scale commercialization. Improvements in fuel cell stack power and power density, cold start temperature, stack durability and efficiency have become key research directions in the field of fuel cell scientific research, ~~Improvements in power, performance, lifespan, and cost have become the key research directions in the field of fuel cell scientific research,~~ among which the fuel cell durability attracts increasing attention pertaining to satisfy the requirements of transportation applications [2]. Researches have been conducted for fault diagnostic and tolerant control in the last decade, however, degradation has never drawn a particular attention concerning the control design [3]. This is because the ageing phenomenon is complex and is far from being fully understood so that it is rarely predicted and the prediction accuracy is not satisfying for performing control actions.

The evolution of the Industry 4.0 revolution implies a transition towards intelligent energy

sources and their management. Technologies like digital twins, the internet of things, etc., have been adopted in the energy sector to control the distributed energy sources and visualize their operation, as shown in Figure 1 [4]. The information transfer in industrial ~~digitalization~~ digitization is close and in real-time, while the process should be understood, predicted, optimized, and controlled based on a model in the virtual space. Towards the realization of ~~digitalization~~ digitization, prognostics and health management (PHM) as a novel discipline has been engaged in implementing health supervision and developing control and management strategies. It is one of the main protagonists of the Industry 4.0 revolution. PHM has been developed in the energy sector, as well as in the domain of fuel cells and fuel cell systems [5]. By collecting the data information from the system, it aims to monitor and evaluate the health of the system, predict its failure before the system fails, and integrate the prediction information on different time scales to provide a series of control, management, and maintenance decisions with the help of information technology.

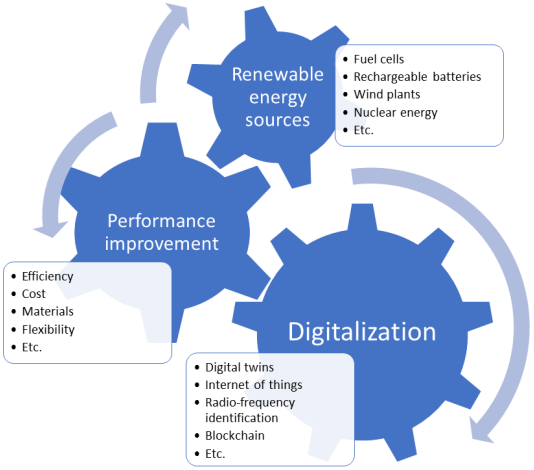


Figure 1: ~~digitalization~~ digitization in the development of energy sector

Owing to the full development of modern information technology, artificial intelligence, and big data are becoming the development direction of the ~~digitalization~~ digitization era. The advantages of data-driven and digital methods of learning are brought into play and applied to various energy systems. As fuel cell degradation has not yet been fully discovered on the level of physics, developing data-driven methods has become an effective means to solve the problems of fuel cell energy systems with multiple temporal and spatial scales, time-varying features, and complex dynamics. It can improve the deficiencies of many traditional analysis and control methods and make progress of the fuel cell performance prediction for vehicle applications [6].

Data-driven methods have been widely applied in fuel cell degradation prognostics [7]. Compared to model-based methods, which highly depend on the model complexity and parameter selection, data-driven methods are more flexible. They do not need to integrate physical knowledge of the system so that they are competitive methods, mainly when applied to multi-physics and multi-time scale systems. For example, recurrent neural networks (RNNs) are the most popular method for time series data processing as they can model the time series data holistically. Various RNN architectures have been investigated in the literature. Ma et al. have used the long-short term memory (LSTM) network to model the fuel cell degradation behaviour and have proposed a grid LSTM structure to improve the degradation prediction accuracy [7]. ~~LSTM keeps the short-term memory for long periods so that it can avoid gradient exploding and vanishing problems.~~ Unlike

traditional RNNs, LSTM introduces a forgetting mechanism. Rather than predicting the future state only based on the recent state, it can select the states that have more impacts on the current state and forget the insignificant ones. LSTM is often used to improve the fuel cell degradation prediction performance because the forgetting gate in LSTM can be learned from the neural network, to a certain extent, it can avoid the distant gradient from decaying. Wang et al. have proposed to add an attention mechanism to a bi-directional LSTM network, which is capable to select the critical outputs of earlier layers to the following and has been firstly applied to fuel cell degradation problem [8]. Echo state network (ESN), as a reservoir computing method, is also extensively applied in prognostics problems owing to its simplicity and fast-training characteristic. Li et al. have proposed an ensemble ESN prognostics method, in which the ESNs are trained with different initialization scenarios [9]. A multi-reservoir ESN has been proposed in [10] to optimize the ESN parameters, e.g., spectral radius, etc.

However, the RNN tends to capture the time dependence from the beginning to the end of the time series. In other words, in a RNN paradigm, each point in the time series depends on the previous time states. The matrix operation parallelization cannot be exploited fully because the current time step waits on the last step; therefore, RNNs are slow to train and infer. Besides, gradient vanishing and exploding problems can be encountered with some deep RNNs. The deficiency in speed and gradient problems have motivated the researchers to use convolutional neural networks (CNNs) to model the time series. CNNs perform predictions as well as the RNNs but are computationally cheaper than RNNs [11]. For example, Cannizzaro et al. have used CNN to predict the solar irradiance in PV applications for both short-term and long-term predictions [12]. The CNN has also been used in [13] to learn the local nonlinear spatial-temporal features in an industrial hydrocracking process. Efforts have been made to enhance the CNN time series prediction performance. For example, a temporal convolutional network has been used in [14] to calculate the activation values of each layer from the earlier values of the previous layer, and a dilated convolution scheme has been used select the values of the previous layer contributing to the next layer in a wider perspective using a dilated rate. In fuel cell applications, Ma et al. have developed a CNN model to predict the impedance parameters of the fuel cell [15] and Huo et al. have used CNN to obtain the polarization curve based on a vast number of V-I datasets and therefore, to predict the fuel cell performance [16]. However, CNN has not yet found its place in fuel cell degradation prediction due to the complexity in its degradation phenomena: changing variance, irregular seasonality, recovery phenomena, high residuals, etc. Current research has started to apply CNNs for the fuel cell internal parameter modeling. Few investigations have been conducted to deploy CNN's prediction nature in fuel cell degradation prognostics.

Therefore, to benefit from the advantages of CNN time series prediction and to introduce it into the domain of fuel cell prognostics, this paper proposes a novel dilated CNN structure with an attention mechanism to perform multi-step-ahead fuel cell degradation prediction. The results are compared with other state-of-the-art RNN structures. Moreover, since the long-term prognostics may be influenced by the fuel cell degradation recovery phenomenon, leading to the failure of the neural network to track the actual degradation trend, a condition-based solution using the previous proposed dilated CNN structure is developed. The efficient and accurate degradation prediction method can predict the fuel cell degradation multiple steps ahead, which allows the control engineers to modify fuel cell load current reference taking the fuel cell state of health into account. When the fuel cell is unable to supply a certain amount of required power, it is important to show its remaining useful life for which an algorithm of prognostics should be deployed. The

proposed method is towards a reliable and spendless fuel cell lifetime prediction solution and will accelerate the integration of the digital model in the operation optimization of the future fuel cell systems. The main innovations of this paper are as follows:

1. Towards industrial ~~digitalization~~ digitization, a fuel cell degradation prediction algorithm is proposed and validated through rough experimental data, which can be used for online performance-based communication.
2. CNN as an encouraging time series prediction method has been first proposed for the ~~multi-step ahead~~ multi-step-ahead fuel cell degradation prediction, combined with a dilated structure and an attention block. The prediction performance is compared with three other RNN prediction tools. The result proves that the CNN-based approach is superior in short-term prediction.
3. A conditional CNN framework has been proposed to improve the long-term prognostics performance.

The following paper is arranged as: Section 2 introduces the PEMFC and its degradation, as well as two ~~aging~~ ageing experiments, which have generated three datasets used for the validation of the proposed prediction methods. Section 3 proposes the dilated CNN with attention mechanism and compares the multi-step-ahead prediction results with other methods. Section 4 analyses the problem of long-term prognostics and proposes the conditional CNN structure. Finally, Section 5 concludes the paper.

2. PEMFC degradation and performance prediction

Fuel cells generate electricity through chemical reactions without burning, which convert hydrogen and oxygen into water and generate electricity during the process. Among the different types of fuel cells, the PEMFC can quickly start and work at low temperatures. It has no electrolyte loss, and it is non-corrosive. Therefore, the PEMFC is the main fuel cell technology for transportation and small stationary applications. This section introduces the PEMFC and its degradation. Two ~~aging~~ ageing experiments are implemented, and the obtained degradation data are introduced.

2.1. Stack structure of PEMFC

Figure 2 shows the operation principle of a PEMFC cell [1]. The hydrogen gas is sent to the anode plate. Reacted with the catalyst, the hydrogen atom becomes a hydrogen ion and an electron. As the electrons cannot pass the electrolyte, they pass through the external circuit and generate an electric current. When reaching the cathode, they combine with hydrogen ions and oxygen atoms to form water. The reactions on the anode and the cathode, as well as the overall reaction, are written as:

$$Anode : H_2 \longrightarrow 2H^+ + 2e^- \tag{1}$$

$$Cathode : \frac{1}{2}O_2 + 2H^+ + 2e^- \longrightarrow H_2O \tag{2}$$

$$Overall : \frac{1}{2}O_2 + H_2 \longrightarrow H_2O + electricity + heat \tag{3}$$

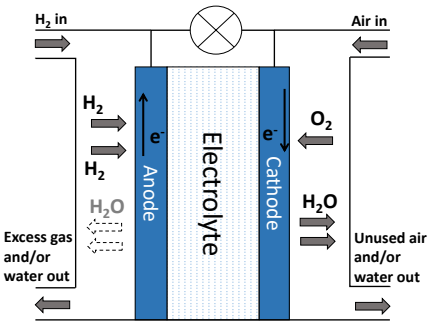


Figure 2: PEMFC operation principle

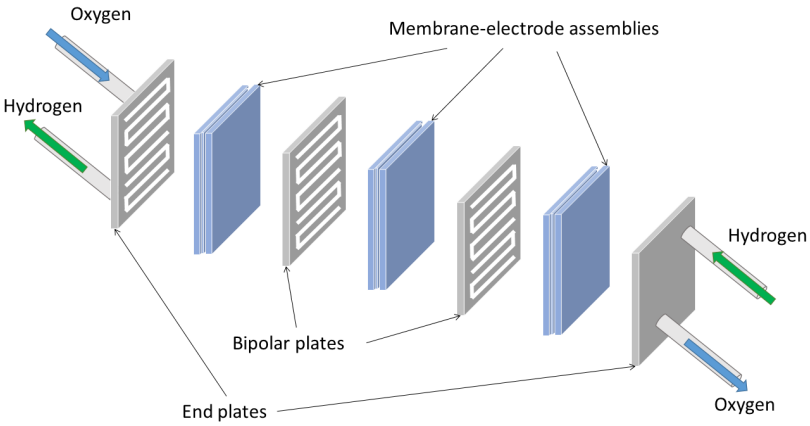


Figure 3: PEMFC stack struture

A single PEMFC cell cannot be used to power the devices because the single cell voltage is only around 1 V at open circuit and 0.6 V - 0.7 V at full load [17]. Therefore, to increase the operation voltage, the stack structure is very common in PEMFC applications. The PEMFC stack is the place where the electrochemical reaction occurs and it is the core part of the fuel cell power system. It is alternatively stacked by the bipolar plates and the membrane-electrode assemblies, and the seals are embedded between them, as shown in Figure 3. They are tightened by the front and rear plates and then fastened with screws. During the reaction, the hydrogen and the oxygen enter separately from their inlet and are distributed to the bipolar plates through the gas diffusion layer. Then, the reaction gases are evenly distributed to the electrode through the bipolar plate guide, and the electrode support is in contact with the catalyst for the electrochemical reaction.

2.2. PEMFC stack degradation

In vehicle applications, PEMFCs usually work as the main power supply and are connected to an energy storage system, which is used to supply the transient power demand and recover the braking energy. Thanks to the energy storage system, the PEMFC can be operated under rather constant conditions with minimal power changes, which can also help to save the fuel cell's life. However, along with the operation, the fuel cell can no longer remain the nominal power output due to various processes and parameter changes, e.g., the fuel and oxygen impurities, stack starvation, pressure, temperature, and hydration variation, peak power demand, current ripples, open circuit voltage operation, etc. However, the potential effects of all these possible degradation causes all result in the voltage drop. Therefore, the degradation of PEMFCs is usually considered at the stack level and is monitored by the stack voltage drop. This is also due to the following two reasons: (1) the degradation of different cells may be different depending on the cell position; (2) the degradation between components is difficult to model, as summarized in Table 1 [18]. Although the phenomena are complex, the degradation trend can be presented by the stack voltage drop versus time. Besides, as the future operation cannot be predicted, using the past information to predict the short-term and long-term degradation of the PEMFC is based on the hypothesis that the impact of the operation conditions is kept the same on the PEMFC stack [19]. Following ~~aging~~ ageing experiments were conducted on different PEMFC stacks under various loads, i.e., constant load and dynamic load. The generated data are used to validate the proposed degradation prediction and prognostics method.

Table 1: PEMFC degradation mechanisms on different components [18]

Component	Degradation mechanism
Membrane	Mechanical degradation due to non-uniform press pressure and inadequate humidification or penetration of the catalyst; thermal degradation due to thermal stress and cycles; chemical degradation due to contamination and radical attack, etc.
Electrodes	Activation loss due to sintering or dealloying of electrocatalyst; conduction loss due to the corrosion of electrocatalyst support; mass transport rate decrease due to mechanical stress; water management ability loss due to change in hydrophobicity of materials due to Nafion dissolution, etc.
Gas diffusion layer	Structural decrease due to the degradation of backing material and carbon corrosion; conduction loss due to corrosion, etc.
Bipolar plates	Deformation due to mechanical stress; conduction loss due to corrosion, etc.
Sealing gaskets	Deformation due to mechanical stress; corrosion, etc.

2.3. ~~Aging~~ Ageing experiments of PEMFC stack

Two ageing experiments were conducted in the laboratory FCLAB Research Federation (FR CNRS 3539) [2], France, on different PEMFC stacks. The global view of the test bench is shown in Figure 4. The first tested two stacks are 5-cell stack modules offered by Proton Motor Fuel Cell, which have an active area of 100 cm^2, denoted as 'PEMFC1' and 'PEMFC2'. They are tested with a constant current and a current with small ripples, respectively. Then, another 15-cell stack of PRAGMA Industries with an active area of 33.625 cm^2, denoted as 'PEMFC3', is tested with a constant current. The ~~aging~~ ageing experiment details for those stacks are described as follows.

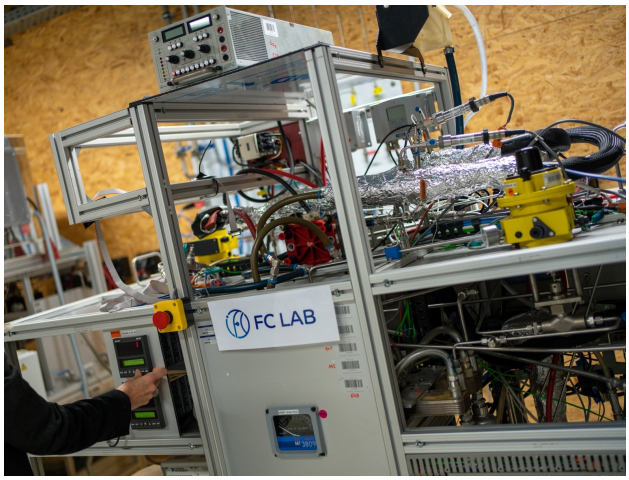


Figure 4: Global view of the test bench for PEMFC ageing experiments

[2] http://www.fclab.fr/

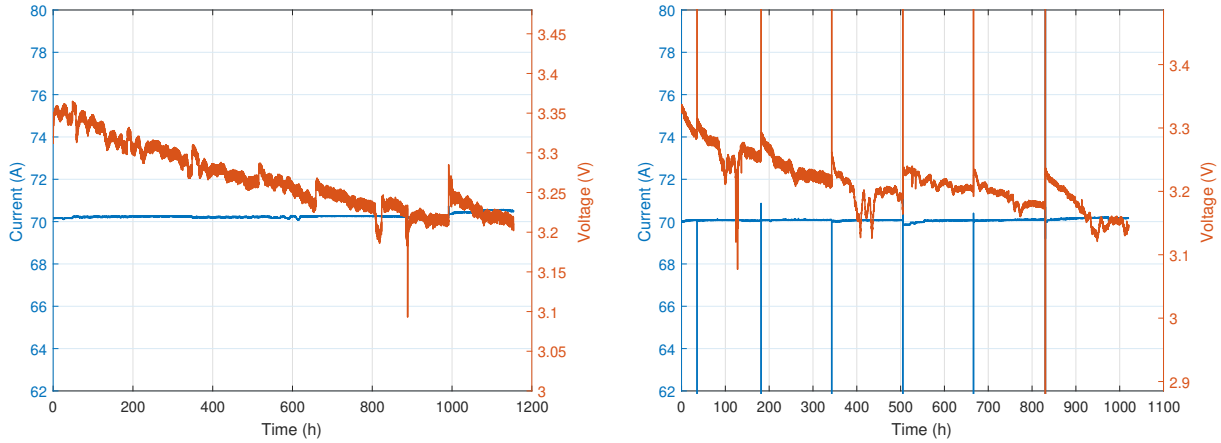
2.3.1. PEMFC1 and PEMFC2

The test bench is adapted for the fuel cell ~~aging~~ ageing experiments with a power up to 1 kW. Some critical parameters of the studied fuel cell stack module are listed in Table 2.

Table 2: Parameters of the tested PEMFC1 and PEMFC2

Parameter	Value
Active surface	100 cm^2
Number of cells	5
Gas pressure	0-2 bar
Maximum operating temperature	80 °C
Maximum current	100 A (1 A/cm^2)

PEMFC1 was operated with approximately nominal operating conditions at a current density of 0.70 A/cm^2. PEMFC2 was operated with high frequency current ripples, i.e. the same current density as that of PEMFC1 but with a ripple current of 0.14 A/cm^2 oscillating at the frequency of 5 kHz. The measured output current and voltage of PEMFC1 and PEMFC2 are shown in Figure 5a and Figure 5b. Characterizations of the stack, i.e. polarization curves, are conducted each week.



(a) Output current and voltage of PEMFC1

(b) Output current and voltage of PEMFC2

Figure 5: Output current and voltage of PEMFC1 and PEMFC2

2.3.2. PEMFC3

A third stack with different configurations is tested to generate another dataset. It is designed with an open cathode and dead-end anode structure, and a 24 Vdc air fan is integrated with the stack for air supply and temperature regulation. Some critical parameters of the studied fuel cell stack module are listed in Table 3.

Table 3: Parameters of the tested PEMFC3

Parameter	Value
Active surface	33.625 cm^2
Number of cells	15
Nominal pressure at hydrogen inlet	0.35 bar
Maximum operating temperature	75 °C
Maximum current	13.45 A (0.4 A/cm^2)

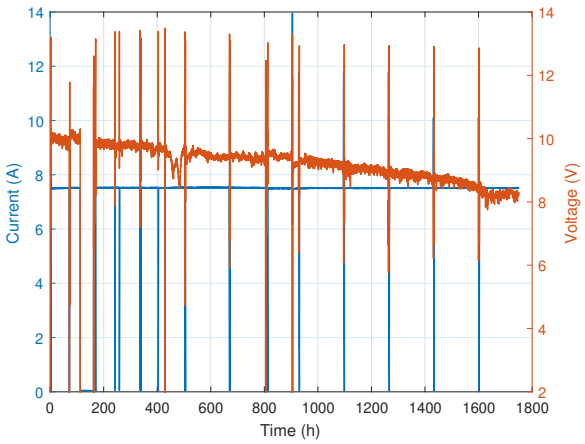


Figure 6: Measured output current and voltage of PEMFC3

The stack module is tested under constant current load. The measured output current and voltage of PEMFC3 are shown in Figure 6. Characterizations of the stack and the unintended peripheral interrupts have caused the voltage peaks and a stop happening at around 110 hours. The stop period will be deleted when developing prediction algorithms under the hypothesis that the stack remains in its previous state during the stop.

2.4. PEMFC performance prediction

As the degradation of the PEMFC can be observed for the output voltage drop, many studies have developed data-driven methods to predict the degradation performance of the fuel cells directly with the voltage signal. The so-called "prediction" refers to two definitions in this paper: multi-step-ahead prediction and long-term prognostics, as shown in Figure 7. Multi-step-ahead prediction predicts only several values ahead by learning the previous values, and after each prediction, the real measurements are fed into the model. The model is updated to predict the next few steps. The prediction is short-term, and after each prediction, modifications to the system control and management strategies can be conducted according to the predicted results. However, long-term prognostics does not need actual measurements in the prediction phase, i.e., when the prediction starts, only the predicted values are fed into the next step to obtain the following sequence. The long-term prognostics allow us to predict the end of service life of the fuel cell stack, and it transfers more general information, i.e., the RUL. The RUL can be used to schedule the system maintenance, optimize the operating efficiency, and avoid unplanned downtime. Predicting the RUL is more complicated than multi-step-ahead prediction because the future is unknown and full

of uncertainties. In order to master the behavior of the fuel cell, the prediction method should be capable of dealing with the both cases. In this paper, an efficient CNN has been developed for the multi-step-ahead prediction, and the CNN with conditioning has been proposed for the long-term prognostics, as presented in the following sections.

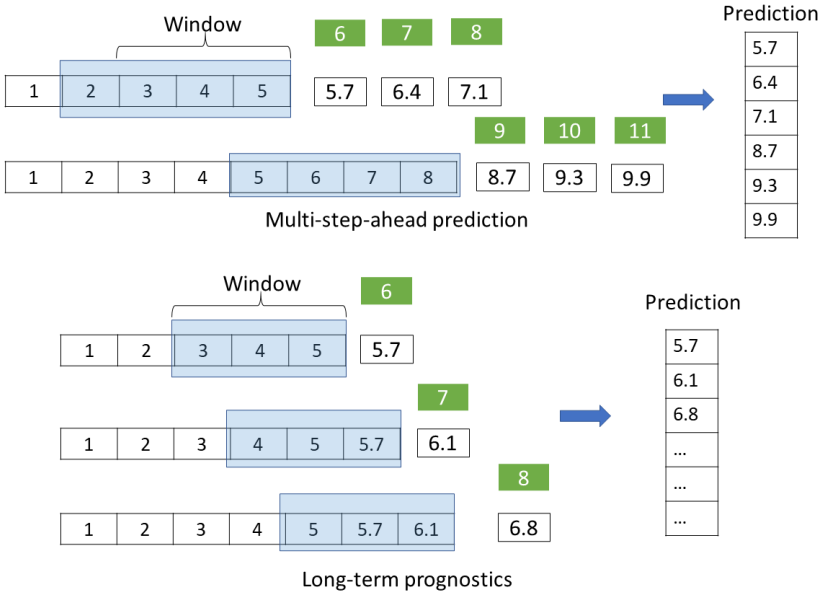


Figure 7: Multi-step-ahead prediction and long-term prognostics demonstration

3. Multi-step-ahead prediction based on dilated CNN

This section describes the proposed methodology for the control-oriented multi-step-ahead PEMFC degradation prediction using dilated CNN with attention mechanism. First, the proposed neural network is trained by the training dataset, and then n upcoming steps are predicted at each step in the testing phase. To predict the next step, the actual values are fed back into the network. The principle of implementing CNN for time series prediction is presented in Section 3.1. Then, the dilated convolution adapted for time series forecasting and the attention mechanism are described in Section 3.2 and Section 3.3, respectively. The global framework of the proposed method is presented in Section 3.4. Finally, Section 3.5 compares and discusses the prediction results with other state-of-the-art methods using the datasets from the above-described ageing experiments.

3.1. CNN

CNN is a type of neural network that can be well adapted for time series modeling and forecasting. This is because the CNN can filter the input to obtain its feature map through a series of convolution operations. Using CNN to deal with time series problems, the convolutions should be casual, i.e., there should be no information leakage from the future to the past, and the output sequence has the same length as the input. Figure 8 shows the casual convolutions with three operation layers. It has the output of the same length as the input by applying zero padding, i.e., zeros are added to keep the subsequent layers having the same length as the previous layers.

Each node in the output layer (the top layer) is determined by a horizon of 4, showing that the multi-layer CNN can be conducted for time series forecasting problems as the current output is determined not only by the current input but also the historical values, in other words, the casual CNN is of temporal dependencies.

To implement CNN, the convolution layer learns the feature from the input time series using a predefined kernel through the convolution operations. The convolution kernel has two parameters, i.e., size and depth, which correspond to the moving window length and the number of channels. The multi-layer convolution operation can be written as:

$$\mathbf{x}_{l,j}^{conv} = \sum_{i=1}^{p} \mathbf{w}_{l,j} \otimes \mathbf{x}_{l-1,i}^{activate} + b_{l,j} \tag{4}$$

where $\mathbf{x}_{l,j}^{conv}$ is the convolution result of the j-th channel in the l-th convolution layer, $\mathbf{x} \in R^p$ is the time series, $\mathbf{w} \in R^k$ is the kernel and $k << p$. $b_{l,j}$ denotes the bias.

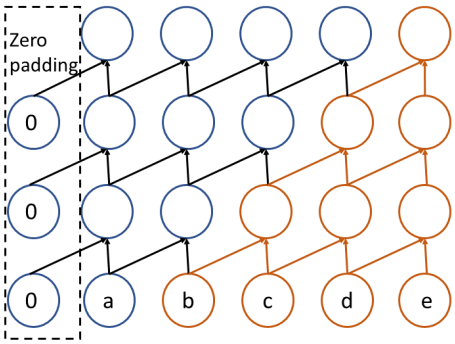


Figure 8: Casual convolution with three convolution layers

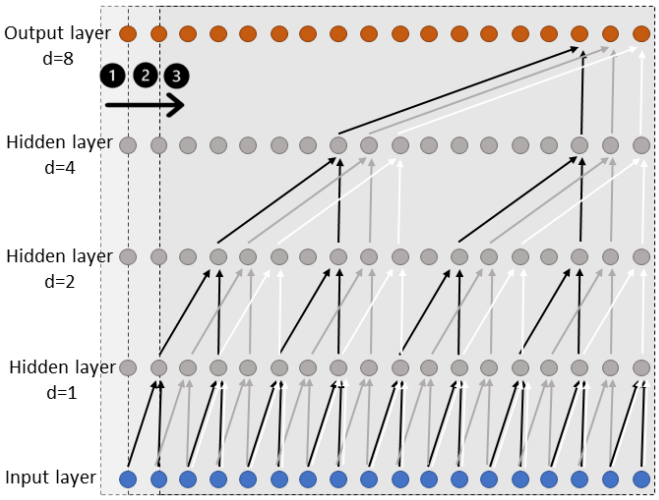


Figure 9: Multi-layer dilated convolution with different dilated factors

Some CNNs have a pooling layer to summarize the convolution values by calculating the maximum or mean values, which are the most common pooling methods. The pooling layer can reduce the input size, speed up the computing and avoid over-fitting problems. Then, a non-linear activation function is used to implement the prediction projection. The Gaussian error linear unit (GELU) activation function is applied in this paper, which can be regarded as a smoother rectified linear unit (RELU) function and gives the weights to the inputs by their percentile, approximated as

$$GELU(\mathbf{x}_{l,j}^{conv}) = 0.5 \cdot \mathbf{x} \cdot (1 + tanh(\sqrt{2/\pi} \cdot (\mathbf{x} + 0.044715 \cdot \mathbf{x}^3))) \tag{5}$$

Finally, the output is generated using the fully connected layer, which is also called the dense layer. The data used for the fully connected layer are the flattened features smoothed by the convolution layers and the pooling layers [14].

However, a simple casual CNN, like the one shown in Figure 8, can only look back on a limited size of time series, which puts up a challenge on time series prediction when longer historical data

needs to be considered. One of the solutions to include a larger receptive field is to apply dilated convolutions [20], described in the next section.

3.2. Dilated convolution

To enlarge the horizon that the CNN can look back, a dilation operation is used to select the values in the previous layer that are kept to generate the values in the next layer, as shown in Figure 9. The objective to use the dilation operation is to consider the long-term dependencies in the time series when forecasting the following values. The dilated convolution is calculated as:

$$\mathbf{x}_{l,j}^{conv} = \sum_{i=1}^{p} \mathbf{w}_{l,j} \otimes_d \mathbf{x}_{l-1,d\cdot i}^{activate} + b_{l,j} \tag{6}$$

where d is the dilation factor, an example of a four-layer dilated CNN is shown in Figure 9, where the kernel has a size of 2 and the L layers of dilated convolutions have the dilation factor increasing by a factor of two: $d \in [2^0, 2^1, ..., 2^{L-1}]$. Thus, the horizon of the output is influenced by the kernel size and the number of layers, written as:

$$H = 2^{L-1} \cdot k \tag{7}$$

Therefore, the output of the network shown in Figure 9 is given by a horizon of 16 inputs.

3.3. Attention mechanism

The attention mechanism is newly developed in recent years, and it simulates the auto-selection mechanism in the human brain, which automatically allocates more information on the desired target [21]. The attention mechanism is used in time series prediction to avoid the problem causing by the large time step, which may include redundant information that does not contribute to the forecasting result.

Here, the vector of output features from the dilated CNN is projected into a hidden state g, while the hidden state of the last window is h. Then, the attention weights at at the current stamp t are calculated as :

$$at_{t,i} = \frac{e^{z_i}}{\sum_{i=1}^{p} e^{z_i}}, \; z_i = g_t^T W h_i \tag{8}$$

where W is the trainable weight matrix of attention Luong's multiplicative style score [22].

Then, a context vector c_t is calculated as the weighted sum of the attention weights and the last state h as: $c_t = \sum at_{t,i} h_i$. Finally, the final output of temporal attention layer is the concatenation of the context vector c_t and the hidden state g, along with a non-linear projection operation as:

$$a_t = f(c_t, g_t) = tanh(W_c[c_t; g_t]) \tag{9}$$

where W_c is a trainable weight matrix.

3.4. Proposed structure

The global framework of the proposed multi-step-ahead prediction method based on a stacked dilated CNN with attention mechanism is shown in Figure 10. The inputs are injected into the CNN but also bypass the network, which formulates a residual block. This is because sometimes shadow networks may have better performance than deeper networks with more layers. The residual block

actually learns the residual of the input and the output so that it can benefit from the elimination of the singularities causing by the linear dependence [23]. To train the output, the result of the residual block goes through a dropout layer and an attention block to improve the importance of the training input. The dropout layer consists of ignoring some neurons during the training phase of a set of randomly selected neurons. The dropout technique is applied when the neural network is too big, the training is too long and the training samples are few to avoid over-fitting [24]. During the training phase, certain neurons are deactivated at each update with a probability value and the in-going and out-going edges of the nodes are removed. The network is forced to improve its generalization by learning different neurons and connections. After the dropout layer, the self-attention block is over the output layer, which is used to select the crucial values of the last layer from the input sequence, as described in Section 3.3.

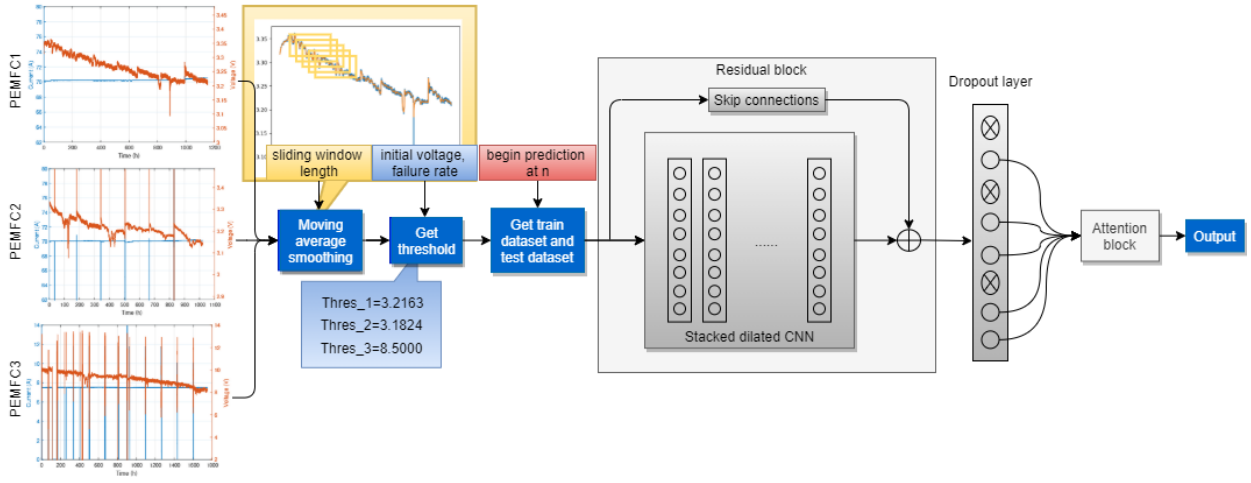


Figure 10: Proposed multi-step-ahead prediction structure based on dilated CNN with attention mechanism

Based on the framework proposed above, the implementation of the multi-step-ahead prediction is summarized as follows:

1. Collect the ~~aging~~ ageing data of the PEMFCs, described in Section 2. The stack voltage is used directly for the implementation of the multi-step-ahead prediction. Preprocess the data by removing the peaks and smoothing using the moving average smoothing method. Split the data into training datasets and testing datasets.
2. Decide the structure of the proposed method. The configuration of the proposed dilated CNN with an attention block is listed in Table 4, which contains two convolution layers, one skip connection layer and an attention block. The configuration optimization is considered during the model construction through grid search.
3. Determine the optimizer ($Adam$) and the loss function (mse) for each layer.
4. In the testing phase, the trained model is used to output the predicted values for each prediction step. The predicted values are compared with the true measurements and the performance evaluation criteria are calculated.

Table 4: Configuration of the proposed dilated CNN with attention block

	Input layer	Convolution layer 1				Convolution layer 2			
	Input size	Channels	Kernel size	Activation	Dilation rate	Channels	Kernel size	Activation	Dilation rate
PEMFC1	17	10	2	gelu	1	32	2	gelu	2
PEMFC2	20	10	2	gelu	1	64	2	gelu	2
PEMFC3	25	10	2	gelu	1	32	2	gelu	2
	Attention layer	Skip connection							
	Output size	Channels	Kernel size	Activation	Dilation rate				
PEMFC1	12	32	1	linear	1				
PEMFC2	32	64	1	linear	1				
PEMFC3	12	32	1	linear	1				

3.5. Multi-step-ahead prediction performance

3.5.1. Performance evaluation criteria

Two widely used time series prediction evaluation criteria are used in this paper: the mean absolute percentage error (MAPE) and the root mean square error (RMSE), written as:

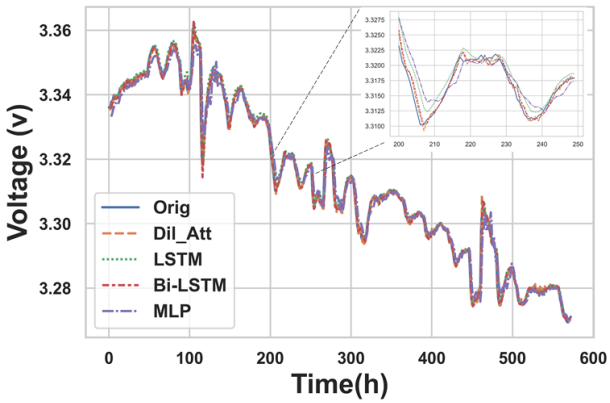
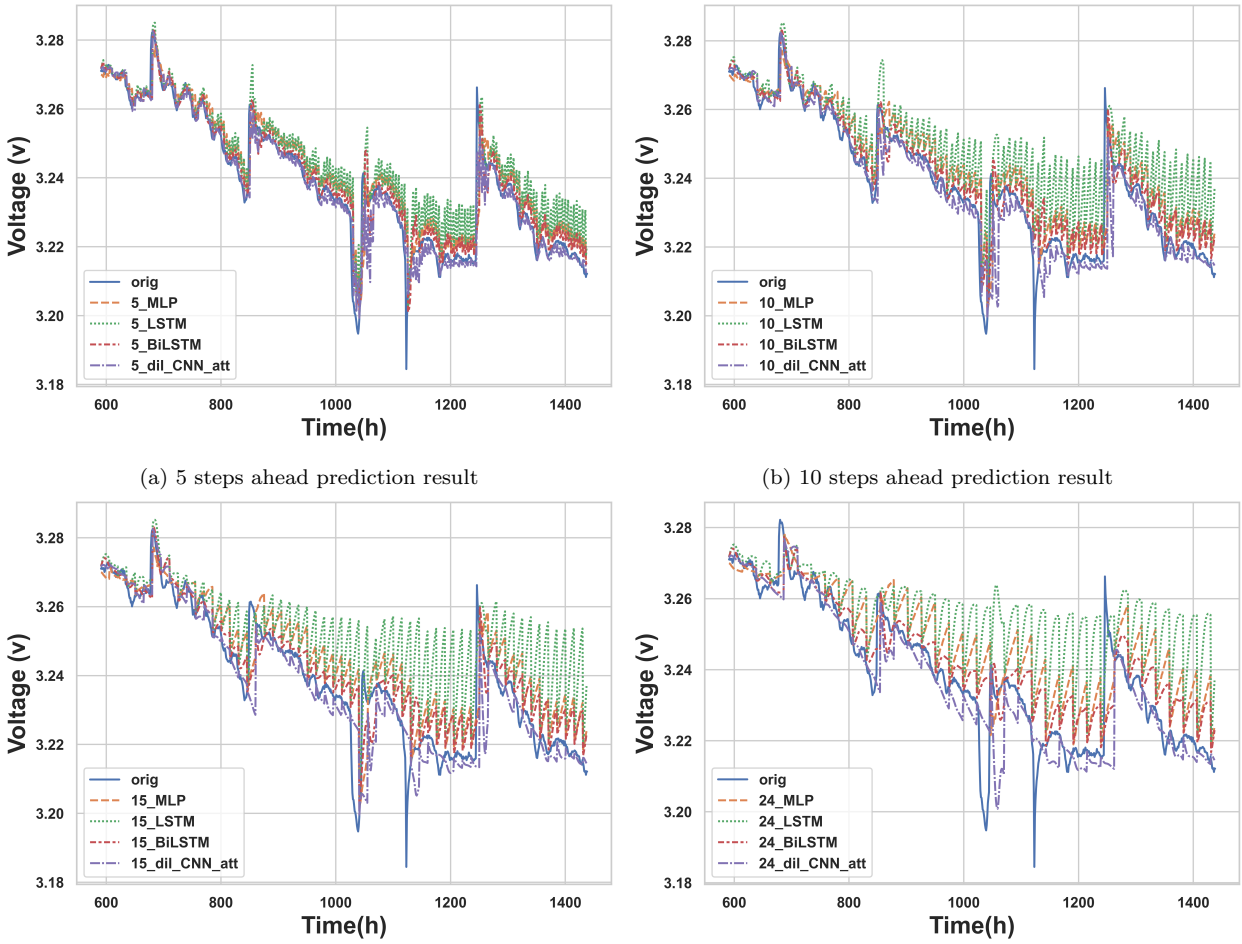


Figure 11: Voltage drop prediction on the training set of PEMFC1

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{x(t)-\hat{x}(t)}{x(t)}\right| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\frac{|x(t)-\hat{x}(t)|}{|x(t)|}} \tag{11}$$

where $x(t)$ is the actual voltage, $\hat{x}(t)$ is the predicted voltage, t is the prediction time, and N is the total number of test data. The two metrics are used to evaluate the prediction accuracy, and from the last two equations, the values of MAPE and RMSE closing to zero are regarded as accurate predictions.



(a) 5 steps ahead prediction result

(b) 10 steps ahead prediction result

(c) 15 steps ahead prediction result

(d) 24 steps ahead prediction result

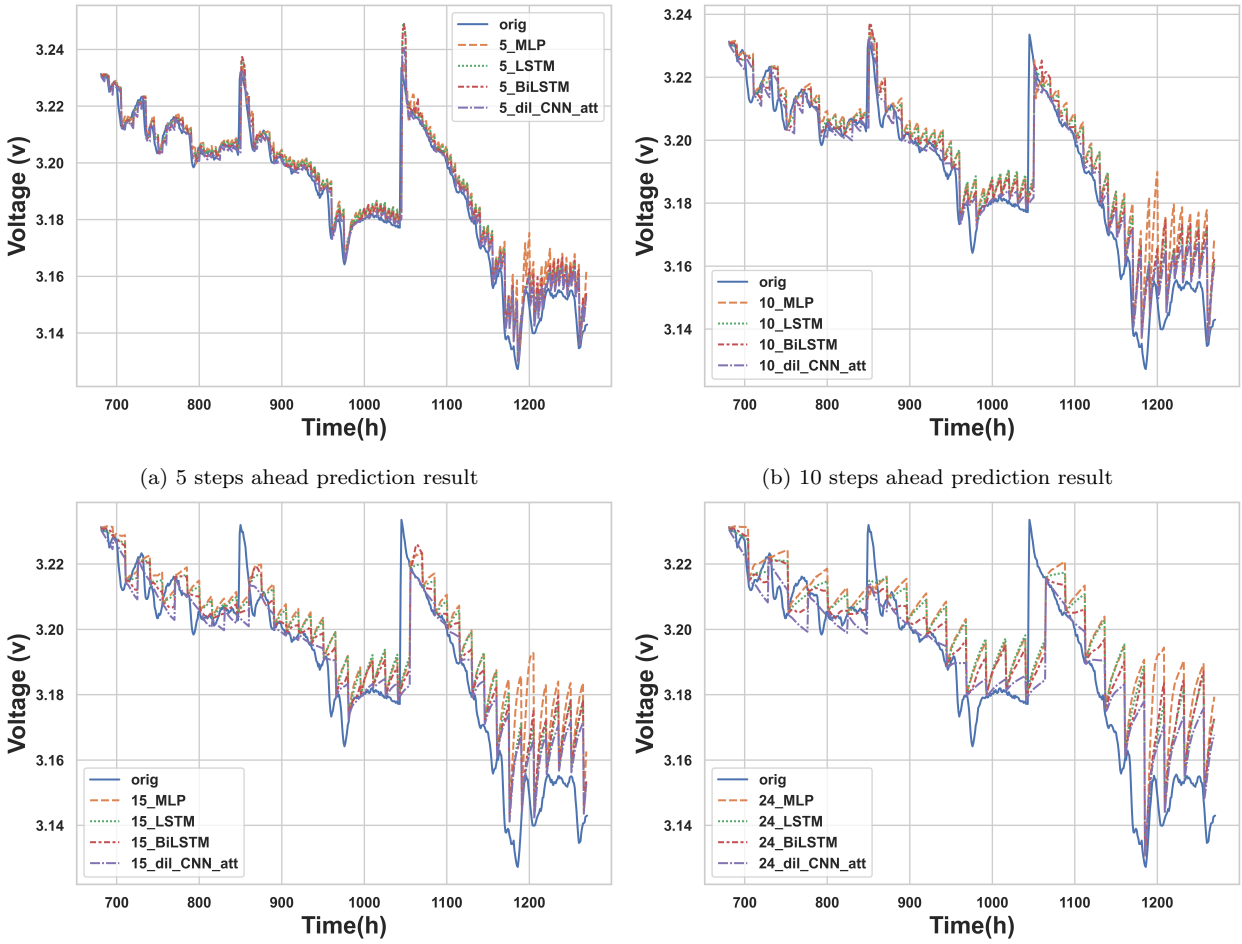
Figure 12: Multi-step-ahead prediction for PEMFC1

3.5.2. Result comparison

The performance of the proposed dilated CNN with attention has been compared with other three recurrent machine learning techniques in the literature: multilayer perceptron (MLP) [25], LSTM [7] and bi-directional LSTM (BI-LSTM) [8]. As an example, Figure 11 shows the training results of PEMFC1. All the compared algorithms can follow the voltage variation of the fuel cell. The proposed dilated CNN has the best results, with 0.0015 and 0.0248 for RMSE and MAPE, respectively. Also, the Bi-LSTM model performs well with RMSE of 0.0017 and MAPE of 0.0273, while LSTM is less accurate with 0.0023 and 0.0414, and MLP achieves the worst results with 0.003 and 0.0588 for RMSE and MAPE.

Then, the prediction results with different horizons (n steps ahead, $n = 5, 10, 15, 24$) are compared for the three tested fuel cells. During the testing phase, n steps ahead are predicted at each prediction step and actual values are fed to the model for the next prediction. The visual comparison is shown in Figure 12, Figure 13 and 14 for PEMFC1, PEMFC2 and PEMFC3, respectively and the error comparison is presented in Table 5.

As shown in Figure 12, the prediction starts at 591st hour until reaching the end of the dataset.



(a) 5 steps ahead prediction result

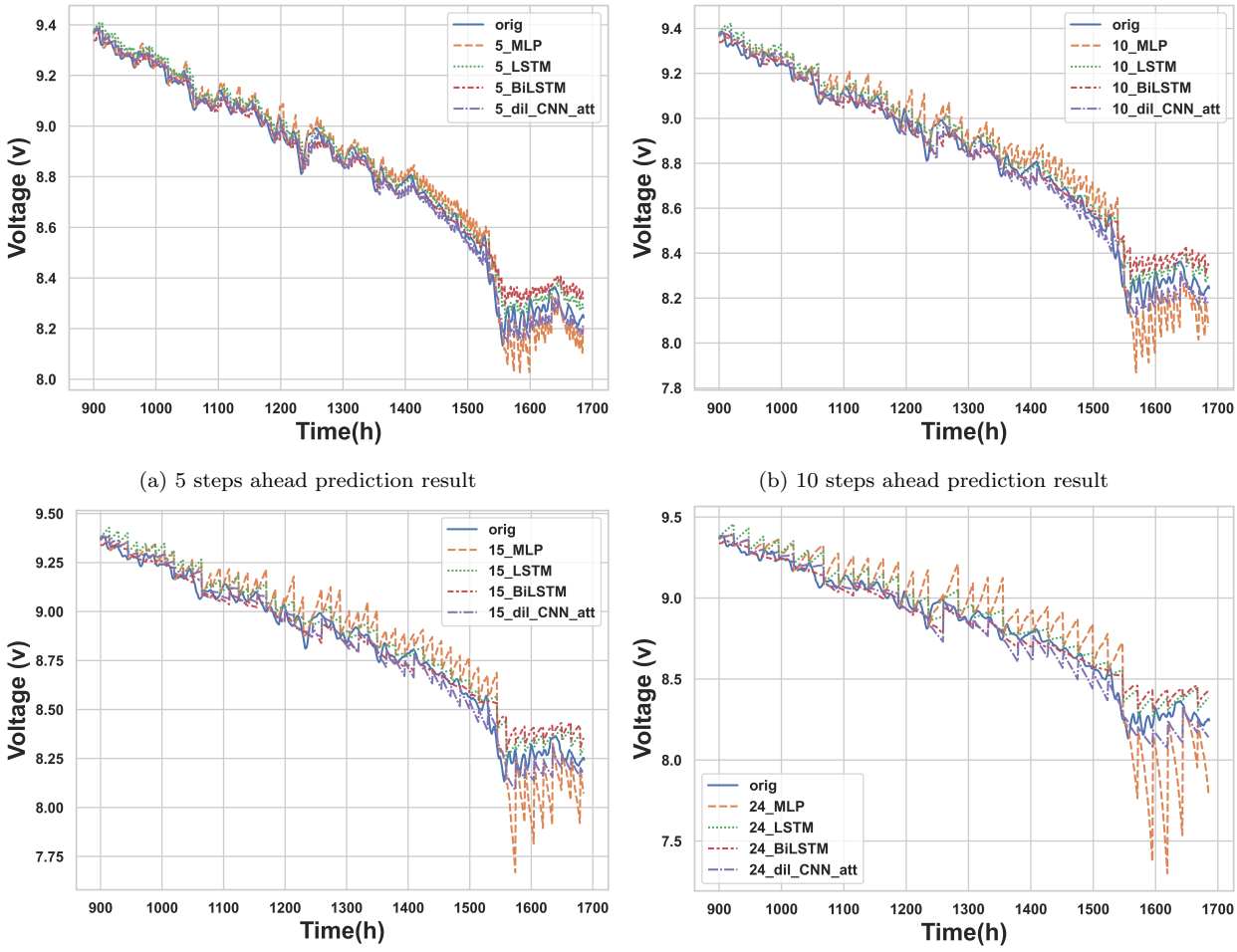
(b) 10 steps ahead prediction result

(c) 15 steps ahead prediction result

(d) 24 steps ahead prediction result

Figure 13: Multi-step-ahead prediction for PEMFC2

The proposed prediction method can follow the degradation performance of the PEMFC1, and the predictions are mostly "early" predictions, which means that the algorithm can look before the fuel cell degrades. During the experiment, some external interruptions have caused irregular degradation trends, which are due to the stops for characterizations. However, the proposed prediction method can well converge the abnormality into the model and provides relatively accurate multi-step-ahead prediction results. The RMSE values for 5 steps, 10 steps, 15 steps and 24 steps ahead prediction are 0.0044, 0.0069, 0.0086 and 0.0092, respectively, and the MAPE values are 0.0702, 0.1092, 0.1479 and 0.1625, respectively, as shown in Table 5. The error evaluation results of the proposed method have been almost the best except for the RMSE of 10 steps ahead prediction using BI-LSTM, which is slightly higher. When comparing the proposed CNN and the powerful recurrent BI-LSTM, although BI-LSTM generates accurate predictions, as shown in Figure 12, those are "late" predictions, which means that the predictions are later than the happening of the actual degradation. When solving control-oriented ~~aging~~ ageing prediction problems, it is far more important to have early predictions because they can give more time margin for the operators to control and manage the system to adapt for the more appropriate performance.



(a) 5 steps ahead prediction result

(b) 10 steps ahead prediction result

(c) 15 steps ahead prediction result

(d) 24 steps ahead prediction result

Figure 14: Multi-step-ahead prediction for PEMFC3

For the PEMFC2, as shown in Figure 13, the prediction starts at 681st hour, and the degradation can be early predicted until around 900 hours, while after 900 hours, the predictions are afterward the real values due to the large dynamic in the voltage variation. The RMSE values for 5 steps, 10 steps, 15 steps and 24 steps ahead prediction are 0.0055, 0.0090, 0.0118 and 0.0132, respectively, and the MAPE values are 0.0973, 0.1657, 0.2243 and 0.2661, respectively, as shown in Table 5. To deal with the large dynamics that influence the correct degradation prediction, a conditional neural network will be introduced, which trains the neural network with a condition signal obtained from autoregression. The conditional signal cannot simulate the dynamic character in the degrading voltage; however, the degradation tendency gives a reference for the long-term time series prediction and will drive the predictions to the corrected direction. It will be introduced in the next section.

For the PEMFC3, the proposed method has obtained the best prediction results. The prediction starts at 900th hour and the RMSE values for 5 steps, 10 steps, 15 steps and 24 steps ahead prediction are 0.0369, 0.0505, 0.0594 and 0.0780, respectively, and the MAPE values are 0.3235, 0.4490, 0.5260 and 0.6753, respectively, as shown in Table 5. In this case, the BI-LSTM algorithm

Table 5: Multi-step-ahead prediction results for the tested PEMFC1, PEMFC2 and PEMFC3

PEMFC1		**Steps ahead**			
		5	**10**	**15**	**24**
MLP	RMSE	0.0064	0.0084	0.0109	0.0134
	MAPE	0.1398	0.1962	0.2524	0.3252
LSTM	RMSE	0.0078	0.0127	0.0172	0.0210
	MAPE	0.1834	0.3027	0.4053	0.5144
BI-LSTM	RMSE	0.0052	**0.0067**	0.0089	0.0100
	MAPE	0.0979	0.1349	0.1862	0.2203
Dil-CNN-att	RMSE	**0.0044**	0.0069	**0.0086**	**0.0092**
	MAPE	**0.0702**	**0.1092**	**0.1479**	**0.1625**
PEMFC2		**Steps ahead**			
		5	**10**	**15**	**24**
MLP	RMSE	0.0072	0.0123	0.0164	0.0187
	MAPE	0.1491	0.2592	0.3619	0.4367
LSTM	RMSE	0.0066	0.0103	0.0136	0.0161
	MAPE	0.1248	0.2159	0.2991	0.3719
BI-LSTM	RMSE	0.0061	0.0095	0.0126	0.0143
	MAPE	0.1197	0.1965	0.277	0.3169
Dil-CNN-att	RMSE	**0.0055**	**0.0090**	**0.0118**	**0.0132**
	MAPE	**0.0973**	**0.1657**	**0.2243**	**0.2661**
PEMFC3		**Steps ahead**			
		5	**10**	**15**	**24**
MLP	RMSE	0.0541	0.0923	0.1251	0.1978
	MAPE	0.4471	0.7729	1.0495	1.5644
LSTM	RMSE	0.0484	0.0613	0.0739	0.0898
	MAPE	0.4359	0.548	0.6612	0.8168
BI-LSTM	RMSE	0.0543	0.0644	0.0726	0.0835
	MAPE	0.4388	0.5188	0.5762	0.6824
Dil-CNN-att	RMSE	**0.0369**	**0.0505**	**0.0594**	**0.0780**
	MAPE	**0.3235**	**0.4490**	**0.5260**	**0.6753**

also follows the degradation of the fuel cell voltage; however, as shown in Figure 14, after 1550 hours, when there is a recovery in the voltage signal, BI-LSTM has lost its advantage, and it cannot continuously learn the degradation tendency. The proposed dilated CNN benefits from its large horizon when the time series is strongly time-dependent. Even though there is a sudden recovery in the signal, it rapidly adapts to the variation.

Multi-step-ahead prediction can be adapted to the fault-tolerant control and management of the PEMFC system. However, if the steps ahead become larger, the prediction algorithm has no longer the competitive performance. This is because the neural network takes the capability to learn the dynamics in the signal; yet, it can be trapped into over-fitting problems, in which the degradation tendency makes no difference and cannot be followed. To deal with this problem, it is essential to consider the degradation tendency during the long-term prediction so that the neural network will not lose the tendency and the dynamic performance of the fuel cell. To solve this problem, a conditional dilated CNN predicts the time series based on another related series, as described in the next section.

4. Prognostics based on conditional CNN

In smart manufacturing, prognostics is performed to estimate the operating time before failure, and the risk of existence or later appearance of one or more failure modes [26]. The objective is to estimate the RUL and its probability, which are used for further ~~aging~~ ageing tolerant control and management. Data-driven prognostics methods are model-free and can capture the non-linearity

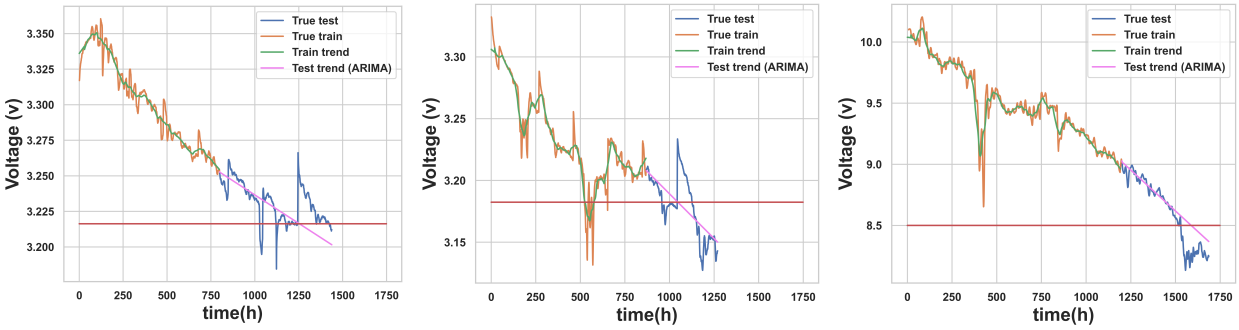
of the degradation of complex systems, like PEMFCs. However, as described above, the voltage evolution of PEMFC contains a lot of dynamics, which lead to over-fitting problems when applying neural networks. In this section, a conditional dilated CNN based on the structure presented in the last section is proposed to address this problem. The PEMFC degradation tendency is generated using an autoregressive method, which is then used as a condition for the long-term degradation prognostics. The following of this section is arranged as: Section 4.1 introduces an autoregressive method used in this paper to generate the PEMFC degradation tendency online. The implementation of the proposed conditional CNN is described in Section 4.2 and the performance of the proposed prognostics method is evaluated and discussed in Section 4.3.

4.1. ARIMA

ARIMA is a statistical method that has been widely used for time series forecasting. Ma et al. have used a data fusion method based on ARIMA and LSTM to predict the fuel cell degradation by updating the fusion rate [27]. Wang et al. have implemented an ARIMA with exogenous variables offline to generate a navigation sequence, which is added to the training data for online prognostics. To implement ARIMA, it first applies a difference method to the time series, and then the stationary time series can be represented by its previous values and random disturbance terms, written as:

$$\mathbf{x}'_i = c + \alpha_1 \mathbf{x}'_{i-1} + \alpha_2 \mathbf{x}'_{i-2} + ... + \alpha_p \mathbf{x}'_{i-p} + \epsilon_i + \beta_1 \epsilon_{i-1} + \beta_2 \epsilon_{i-2} + ... + \beta_q \epsilon_{i-q} + \epsilon_i \tag{12}$$

where i is the current time instant, p is the order of the autoregressive model and q is the order of the moving average model. Another parameter d represents the difference's order, which differentiates the original time series if it is not stationary. ARIMA model can be used to describe the low-volatility component of the data, while the high-volatility component should be obtained using a more complicated neural network [28]. Examples of using the ARIMA model to predict the degradation tendency of PEMFC1, PEMFC2, and PEMFC3 are shown in Figure 15. The training data is filtered and used to fit the ARIMA model. As shown in Figure 15, the prediction is nearly linear, which contains no dynamic characteristics of the PEMFC degradation. However, it well follows the degradation tendency and can be used as a navigated condition in the behavior model trained by the CNN.



(a) ARIMA result of PEMFC1 (b) ARIMA result of PEMFC2 (c) ARIMA result of PEMFC3

Figure 15: Long-term prediction result using ARIMA model

4.2. Conditional CNN implementation

The time series prognostics using conditional CNN can maximize the conditional likelihood of the target predictions on the condition series. In this paper, the condition series is composed of two parts, which correspond to the training part and the testing part. The condition series of the training part is the trend decomposition of the preprocessing training dataset, and the condition series of the testing part is the ARIMA model output, as shown in Figure 16. The conditional likelihood is written as:

$$p(x|x',\theta) = \prod_{i=0}^{N} p(x_{i+1}|x_0,...,x_i,x'_0,...,x'_i,\theta) \tag{13}$$

It can be obtained by modifying the activation function of the CNN by adding another path, in which the condition series is used as the input, as shown in Figure 16. The activation function is then calculated as:

$$GELU(\mathbf{x}_{l,j}^{conv}) + GELU(\mathbf{cond}_{l,j}^{conv}) \tag{14}$$

The conditioning path has exactly the same configuration as the input path, as described in Table 4. As explained in Section 3, the stacked dilated CNN ensures that the condition can be extended to a wide horizon to increase the time dependency. The skip connections layer allows eliminating those conditions that may not improve the prediction result [29].

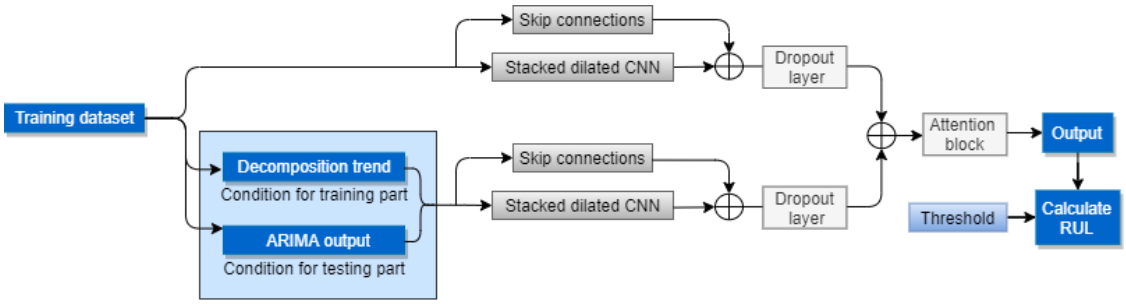


Figure 16: Proposed prognostics method based on conditional CNN

4.3. Prognostics performance

4.3.1. RUL calculation

Supposing that the prediction starts at t_λ, and the end-of-life (EOL) failure threshold of the PEMFC is reached at t_{EOL}, then the actual RUL is calculated as the time difference between t_{EOL} and t_λ. The predicted RUL depends on the time instant when the degrading voltage reaches the failure threshold, denoted as $\hat{t}_{EOL}$. The predicted RUL is calculated as:

$$\hat{RUL} = \hat{t}_{EOL} - t_\lambda \tag{15}$$

4.3.2. Result comparison

The long-term prognostics is implemented based on the scenario proposed in Figure 7. The prediction step is iterated by connecting the predicted value of the last step to the current sequence and this process is continuously effectuated until the predicted value reaching the defined EOL thresholds. The prognostics results using the proposed conditional CNN are shown in Figure

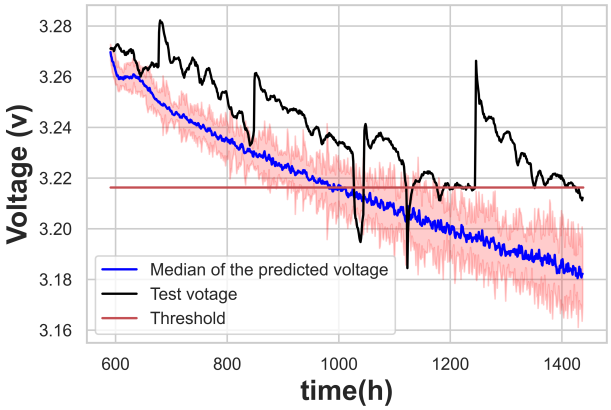
17. Figure 17a, Figure 17c and Figure 17e show the implementation of prognostics on PEMFC1, PEMFC2 and PEMFC3 at the split point 591th hour, 681th hour and 900th hour, respectively. Monte Carlo dropout technique has been used to investigate the uncertainty of the prediction results. 50 experiments have been repeated for each fuel cell and the median value of the predictions are calculated and plotted as the blue lines. According to the length of the tests, the EOL thresholds of the three datasets are defined as reaching 96%, 96% and 85% voltage loss, respectively, ~~By defining the failure threshold,~~ as shown in Figure 10. The actual time reaching the threshold is found: $t_{EOL} = 1028h$ for PEMFC1, $t_{EOL} = 958h$ for PEMFC2 and $t_{EOL} = 1519h$ for PEMFC3. Then, the RULs are calculated using (15). The prognostics have been implemented at different time split points and the box plots are used to represent the prediction results and their uncertainties, as shown in Figure 17b, Figure 17d and Figure 17f.

For PEMFC1, during the training phase, the training data is quasi-linear. The neural network is trained with the decomposition result and the residuals so that the prognostic results are less dynamic. Therefore, the divergence of the repeated experiments is less than that of PEMFC2. The mean values of the predicted RULs at 591th, 691th, 791th, 891th, and 991th hour are $[364, 288, 191, 167, 10]$ hours, respectively. The errors comparing with the actual RULs are $[73, 49, 46, 30, 27]$ hours. The maximum error is an early prediction of 16.0% happening at 591th hour. All the predictions are early predictions so that the early notifications can conduct corresponding actions to avoid failure.

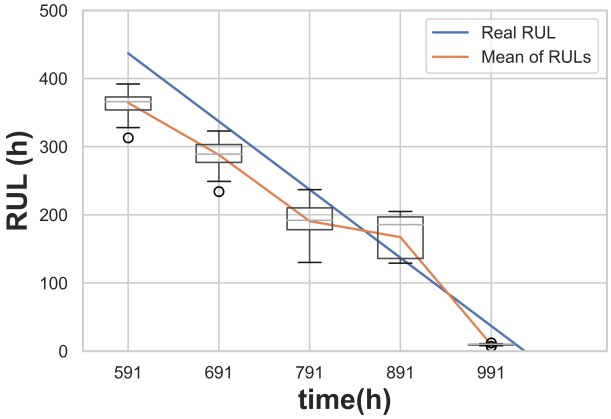
The degradation of PEMFC2 is observed to have more dynamic variations than that of PEMFC1. The training phase learns the trend and the residuals of the training data, and in the prediction phase, the residuals are predicted based on the conditioning. The median value of the predictions can well follow the degradation trend of the real measurements, as well as the dynamics. Although the uncertainties increase along with the accumulation of prediction steps due to the large variations, the median value adapts well to the degradation trend and does not diverge. The predicted RUL converges to the real RUL as the time accumulated. The mean values of the predicted RULs at 681th, 731th, 781th, 831th, 881th, and 931th hour are $[227, 179, 157, 108, 58, 16]$ hours, respectively. The errors comparing with the actual RULs are $[50, 48, 20, 19, 19, 11]$ hours. The maximum error is an early prediction of 18.0% happening at 681th hour, while the minimum error is happening at 931th hour. However, if the predicted RUL at 931th hour is compared with the actual RUL, the error ratio is great; therefore, for the predictions close to the EOL, it is more favorable to use the multi-step-ahead prediction to get more precise results.

For PEMFC3, the prognostics result is consistent to the real measurements, as shown in Figure 17e. This is because the stack voltage degrades linearly and does not consist of irregular variations. Even at its EOL, where there is a change in its degradation trend, the proposed prognostics method can well follow this change thanks to the conditioning. The mean values of the predicted RULs at 900th, 1000th, 1100th, 1200th, 1300th, and 1400th hour are $[552, 437, 422, 304, 231, 65]$ hours, respectively. The errors comparing with the actual RULs are $[67, 82, 3, 15, 12, 54]$ hours. The maximum error is an early prediction of 15.8% happening at 1000th hour, while the minimum error is 0.7% happening at 1100th hour.

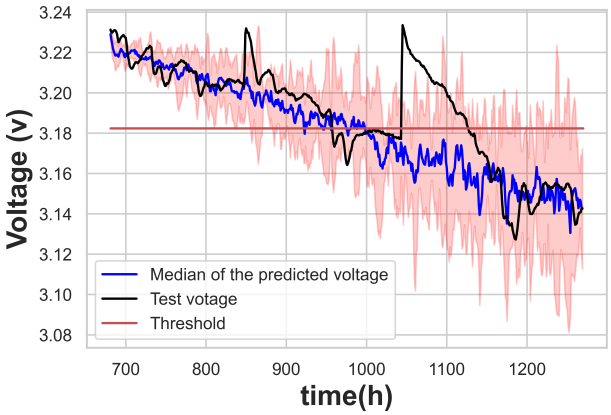
The developed conditional CNN prognostics method is also compared with non-conditional dilated CNN proposed in Section 3 and other RUL prediction methods in the literature [9, 30]. The RUL prediction results of implementing the long-term prognostics at different starting points are listed in Table 6, in which the tests have been performed 50 times to calculate the median values. The results show that the proposed conditional CNN has obtained the most accurate



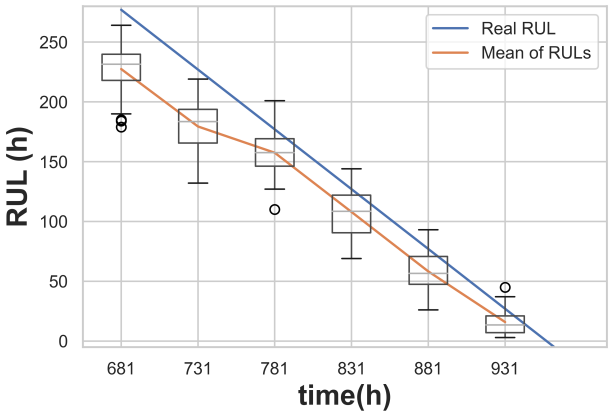
(a) Prognostics result of PEMFC1



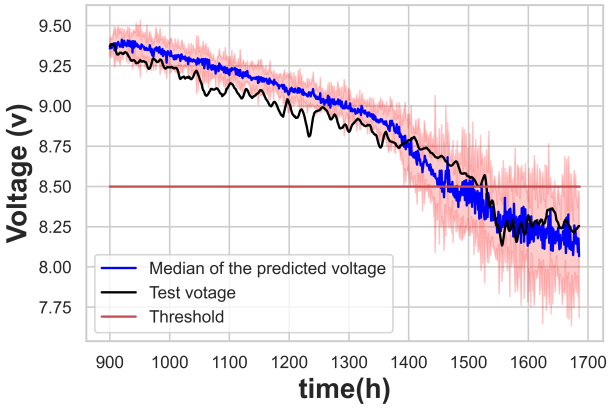
(b) RUL estimation of PEMFC1



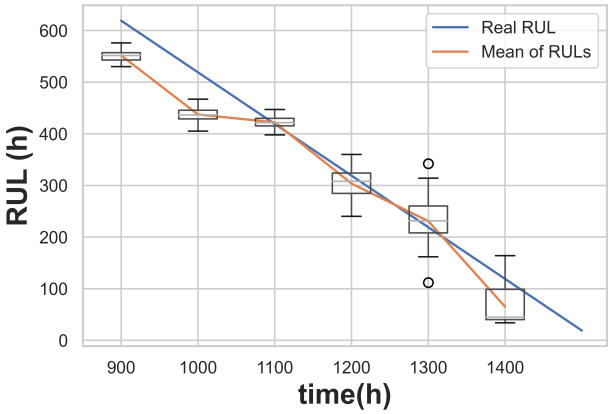
(c) Prognostics result of PEMFC2



(d) RUL estimation of PEMFC2



(e) Prognostics result of PEMFC3



(f) RUL estimation of PEMFC3

Figure 17: Prognostics results with the proposed conditional CNN

RUL prediction results (bolded results in Table 6) compared to other methods. As it can be observed from the table, certain tests cannot obtain RUL values. This is because the predictions diverge and cannot reach the setting EOL threshold so that the RULs cannot be calculated. The proposed conditional CNN has well addressed this problem as the regression result of ARIMA can navigate the prediction of the dilated CNN. Moreover, the dilated CNN itself can also capture more dynamics in the voltage degradation, therefore, the RUL predictions are more favourable compared with other methods.

Table 6: RUL comparison of the long-term prognostics results using conditional CNN (ConCNN), non-conditional dilated CNN (DilCNN), ESN and LSTM

PEMFC1	**Prognostics starting at (h)**					
	591	**691**	**791**	**891**	**991**	
True RUL	437	337	237	137	37	
ConCNN	**364**	**288**	**191**	**167**	10	
DilCNN	90	115	77	221	54	
ESN	371	-	292	485	211	
LSTM	141	209	89	61	**37**	
PEMFC2	**Prognostics starting at (h)**					
	681	**731**	**781**	**831**	**881**	**931**
True RUL	277	227	177	127	77	27
ConCNN	**227**	**179**	**157**	**108**	**58**	**16**
DilCNN	196	286	112	188	48	41
ESN	330	167	222	177	127	-
LSTM	190	-	14	91	26	14
PEMFC3	**Prognostics starting at (h)**					
	900	**1000**	**1100**	**1200**	**1300**	
True RUL	619	519	419	319	219	
ConCNN	**552**	**437**	**422**	**304**	231	
DilCNN	426	289	338	228	163	
ESN	492	339	345	279	**213**	
LSTM	-	-	-	-	-	

Moreover, the implementation time of the four compared methods are recorded and listed in Table 7. Although the implementation time of prognostics depends on the configuration and the performance of the neural networks, it can be seen from the results that the training time of CNN is superior to that of RNN, like LSTM network, while the prediction time is very close. However, the training time of ESN is the shortest. This is because, unlike CNN and RNN using backpropagation, ESN is a reservoir computing network and it randomly allocates weights to the states, which is the major advantage of it. As prognostics is used for predicting the remaining useful life of the fuel cell and it is implemented based on relatively large timescale, the high-speed calculation is not the first requirement. Considering the prognostics performance, the proposed conditional CNN is desirable in both speed and accuracy.

5. Conclusion

This paper has developed a novel multi-step-ahead and long-term prognostics method for the ~~PEMFC~~ proton exchange membrane fuel cell (PEMFC) based on a temporal ~~CNN~~ convolutional neural network (CNN). The multi-step-ahead prediction has been conducted with a dilated CNN structure with attention mechanism to ensure the horizon of the time dependence, and the long-term prognostics has been implemented by the conditional CNN, which uses the trend of degradation as the guidance of the neural network. Results showed that the proposed prediction method is

Table 7: Training time and Prediction time comparison among conditional CNN (ConCNN), non-conditional dilated CNN (DilCNN), ESN and LSTM

PEMFC1	Training time (s)				Prediction time (s)			
Prognostics starting at (h)	**ConCNN**	**DilCNN**	**ESN**	**LSTM**	**ConCNN**	**DilCNN**	**ESN**	**LSTM**
591	6.47	6.39	0.09	25.2	25.82	24.22	0.15	25.4
691	7.72	7.68	0.1	31.9	23.01	21.58	0.11	22.4
791	8.89	8.76	0.13	34.8	20.07	18.79	0.08	20.2
891	9.81	9.68	0.15	37.7	17.93	16.3	0.06	17.3
991	10.92	10.74	0.16	47.5	13.62	12.99	0.05	13.6
PEMFC2	**Training time (s)**				**Prediction time (s)**			
Prognostics starting at (h)	**ConCNN**	**DilCNN**	**ESN**	**LSTM**	**ConCNN**	**DilCNN**	**ESN**	**LSTM**
681	7.51	7.44	0.1	30.2	26.59	25.19	0.07	18.2
731	8.33	8.2	0.11	32.6	18.11	16.025	0.06	16.7
781	9.08	8.99	0.12	35.8	16.72	14.22	0.06	15.2
831	9.21	9.1	0.12	37.5	15.76	13.88	0.05	14.3
881	9.55	9.41	0.13	40.6	12.23	10.74	0.05	12.4
931	10.17	9.98	0.14	41.9	11.44	9.96	0.03	11.2
PEMFC3	**Training time (s)**				**Prediction time (s)**			
Prognostics starting at (h)	**ConCNN**	**DilCNN**	**ESN**	**LSTM**	**ConCNN**	**DilCNN**	**ESN**	**LSTM**
900	9.57	9.54	0.12	43.4	25.13	22.96	0.08	25.6
1000	10.71	10.65	0.14	44.5	21.32	19.7	0.08	21.6
1100	11.76	11.66	0.15	50.1	19.22	17.88	0.07	18.1
1200	12.69	12.57	0.16	54.2	16.91	14.87	0.05	15.7
1300	14.18	13.93	0.16	56.9	12.84	11.71	0.04	12.2

superior to other ~~RNN~~ recurrent neural network algorithms when dealing with the PEMFC degradation prediction problems. Furthermore, the conditional CNN developed for long-term prognostics can follow the degradation trend and, at the same time, take into consideration the degradation dynamics. When comparing with other methods, the proposed conditional CNN method is desirable in both speed and accuracy for fuel cell prognostics.

The integration and penetration of digital tools into the field of transportation and fuel cells are the requirement of the energy revolution. The utilization, analysis, and reasoning of the digitized data bring challenges but also opportunities. The data-driven prediction and prognostics method proposed in this paper is towards a smarter decision-making process, e.g., fault-tolerant control, ~~aging~~ ageing tolerant energy management, predictive maintenance, etc., that can protect and mitigate the degradation of the fuel cell systems by making full use of the lifetime prediction results.

Acknowledgment

This work is supported by the EIPHI Graduate School (contract ANR-17-EURE-0002), Bourgogne Franche-Comté, France.

Appendix A. Data preparation

Data used for long-term prognostics should be able to reveal the underlying trend in the data, however, the raw datasets obtained from the test bench are noisy and consist of many peaks due to sensor sensitivity, data acquisition devices, profile vibration, etc. Data processing is an essential prerequisite for the prognostics task. In this paper, the raw datasets are processed by removing the peaks and smoothing using moving average algorithm. The data volume of datasets PEMFC1,

PEMFC2 and PEMFC3 are 143863, 127372 and 6071395, respectively and the corresponding window widths are chosen as 100, 100, and 25000. The process can be explained by:

$$\bar{y_t} = \frac{y_t + y_{t-1} + ... + y_{t-n-1}}{n} \tag{A.1}$$

where n is the window width of the moving average algorithm, y_t is the current value and $\bar{y_t}$ is the smoothed value. The smoothing result of the three datasets are shown in Figure A.18, Figure A.19 and Figure A.20.

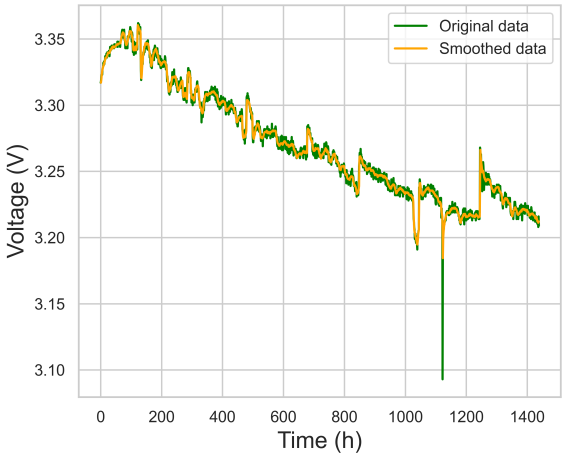


Figure A.18: Smoothing result of PEMFC1

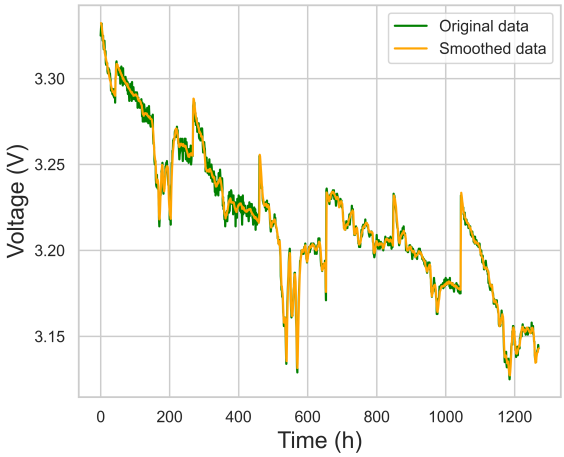


Figure A.19: Smoothing result of PEMFC2

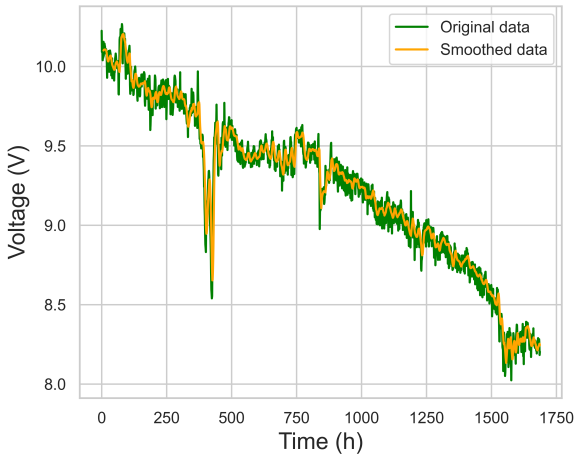


Figure A.20: Smoothing result of PEMFC3

References

[1] M. Yue, H. Lambert, E. Pahon, R. Roche, S. Jemei, D. Hissel, Hydrogen energy systems: A critical review of technologies, applications, trends and challenges, Renewable and Sustainable Energy Reviews 146 (2021) 111180. doi:https://doi.org/10.1016/j.rser.2021.111180.

[2] H. Zhang, X. Li, X. Liu, J. Yan, Enhancing fuel cell durability for fuel cell plug-in hybrid electric vehicles through strategic power management, Applied Energy 241 (2019) 483–490. doi:https://doi.org/10.1016/j.apenergy.2019.02.040.

[3] M. Bressel, M. Hilairet, D. Hissel, B. Ould Bouamama, Model-based aging tolerant control with power loss prediction of proton exchange membrane fuel cell, International Journal of Hydrogen Energy 45 (19) (2020) 11242–11254. doi:https://doi.org/10.1016/j.ijhydene.2018.11.219.

[4] P. F. Borowski, Digitization, digital twins, blockchain, and industry 4.0 as elements of management process in enterprises in the energy sector, Energies 14 (7) (2021). doi:10.3390/en14071885.

[5] M. Jouin, M. Bressel, S. Morando, R. Gouriveau, D. Hissel, M.-C. Péra, N. Zerhouni, S. Jemei, M. Hilairet, B. Ould Bouamama, Estimating the end-of-life of pem fuel cells: Guidelines and metrics, Applied Energy 177 (2016) 87–97. doi:https://doi.org/10.1016/j.apenergy.2016.05.076.

[6] S. Meraghni, L. S. Terrissa, M. Yue, J. Ma, S. Jemei, N. Zerhouni, A data-driven digital-twin prognostics method for proton exchange membrane fuel cell remaining useful life prediction, International Journal of Hydrogen Energy 46 (2) (2021) 2555–2564. doi:https://doi.org/10.1016/j.ijhydene.2020.10.108.

[7] R. Ma, T. Yang, E. Breaz, Z. Li, P. Briois, F. Gao, Data-driven proton exchange membrane fuel cell degradation predication through deep learning method, Applied Energy 231 (2018) 102–115. doi:https://doi.org/10.1016/j.apenergy.2018.09.111.

[8] F.-K. Wang, T. Mamo, X.-B. Cheng, Bi-directional long short-term memory recurrent neural network with attention for stack voltage degradation from proton exchange membrane fuel cells, Journal of Power Sources 461 (2020) 228170. doi:https://doi.org/10.1016/j.jpowsour.2020.228170.

[9] Z. Li, Z. Zheng, R. Outbib, Adaptive prognostic of fuel cells by implementing ensemble echo state networks in time-varying model space, IEEE Transactions on Industrial Electronics 67 (1) (2020) 379–389. doi:10.1109/TIE.2019.2893827.

[10] R. Mezzi, S. Morando, N. Y. Steiner, M. C. Péra, D. Hissel, L. Larger, Multi-reservoir echo state network for proton exchange membrane fuel cell remaining useful life prediction, in: IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, 2018, pp. 1872–1877. doi:10.1109/IECON.2018.8591345.

[11] S. Bai, J. Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, CoRR abs/1803.01271 (2018). arXiv:1803.01271.

[12] D. Cannizzaro, A. Aliberti, L. Bottaccioli, E. Macii, A. Acquaviva, E. Patti, Solar radiation forecasting based

on convolutional neural network and ensemble learning, Expert Systems with Applications 181 (2021) 115167. doi:https://doi.org/10.1016/j.eswa.2021.115167.

[13] X. Yuan, S. Qi, Y. Wang, H. Xia, A dynamic cnn for nonlinear dynamic feature learning in soft sensor modeling of industrial process data, Control Engineering Practice 104 (2020) 104614. doi:https://doi.org/10.1016/j.conengprac.2020.104614.

[14] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, A. Troncoso, Deep learning for time series forecasting: A survey, Big Data 9 (1) (2021) 3–21. doi:10.1089/big.2020.0159.

[15] T. Ma, Z. Zhang, W. Lin, M. Cong, Y. Yang, Impedance prediction model based on convolutional neural networks methodology for proton exchange membrane fuel cell, International Journal of Hydrogen Energy 46 (35) (2021) 18534–18545. doi:https://doi.org/10.1016/j.ijhydene.2021.02.204.

[16] W. Huo, W. Li, Z. Zhang, C. Sun, F. Zhou, G. Gong, Performance prediction of proton-exchange membrane fuel cell based on convolutional neural network and random forest feature selection, Energy Conversion and Management 243 (2021) 114367. doi:https://doi.org/10.1016/j.enconman.2021.114367.

[17] R. Linares, S. Raël, K. Berger, M. Hinaje, J. Lévêque, Pem single fuel cell as a dedicated power source for high-inductive superconducting coils, International Journal of Hydrogen Energy 43 (11) (2018) 5913–5921. doi:https://doi.org/10.1016/j.ijhydene.2017.09.013.

[18] M. Yue, S. Jemei, R. Gouriveau, N. Zerhouni, Review on health-conscious energy management strategies for fuel cell hybrid electric vehicles: Degradation models and strategies, International Journal of Hydrogen Energy 44 (13) (2019) 6844–6861. doi:https://doi.org/10.1016/j.ijhydene.2019.01.190.

[19] R. Mezzi, N. Yousfi-Steiner, M. C. Péra, D. Hissel, L. Larger, An echo state network for fuel cell lifetime prediction under a dynamic micro-cogeneration load profile, Applied Energy 283 (2021) 116297. doi:https://doi.org/10.1016/j.apenergy.2020.116297.

[20] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, Wavenet: A generative model for raw audio (2016). arXiv:1609.03499.

[21] J. Zuo, H. Lv, D. Zhou, Q. Xue, L. Jin, W. Zhou, D. Yang, C. Zhang, Deep learning based prognostic framework towards proton exchange membrane fuel cell for automotive application, Applied Energy 281 (2021) 115937. doi:https://doi.org/10.1016/j.apenergy.2020.115937.

[22] M.-T. Luong, H. Pham, C. D. Manning, Effective approaches to attention-based neural machine translation, arXiv preprint arXiv:1508.04025 (2015).

[23] A. E. Orhan, Skip connections as effective symmetry-breaking, CoRR abs/1701.09175 (2017).

[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, Journal of Machine Learning Research 15 (56) (2014) 1929–1958. URL `http://jmlr.org/papers/v15/srivastava14a.html`

[25] M. Shiblee, P. K. Kalra, B. Chandra, Time series prediction with multilayer perceptron (mlp): A new generalized error based approach, in: M. Köppen, N. Kasabov, G. Coghill (Eds.), Advances in Neuro-Information Processing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 37–44.

[26] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, N. Zerhouni, Prognostics and health management of pemfc – state of the art and remaining challenges, International Journal of Hydrogen Energy 38 (35) (2013) 15307–15317. doi:https://doi.org/10.1016/j.ijhydene.2013.09.051.

[27] R. Ma, Z. Li, E. Breaz, C. Liu, H. Bai, P. Briois, F. Gao, Data-fusion prognostics of proton exchange membrane fuel cell degradation, IEEE Transactions on Industry Applications 55 (4) (2019) 4321–4331. doi:10.1109/TIA.2019.2911846.

[28] S.-L. YU, Z. Li, Stock price prediction based on arima-rnn combined model, DEStech Transactions on Social Science, Education and Human Science (03 2018). doi:10.12783/dtssehs/icss2017/19384.

[29] A. Borovykh, S. Bohté, C. Oosterlee, Conditional time series forecasting with convolutional neural networks, arXiv: Machine Learning (2017).

[30] C. Wang, Z. Li, R. Outbib, D. Zhao, M. Dou, Proton exchange membrane fuel cells prognostic strategy based on navigation sequence driven long short-term memory networks, in: IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, 2020, pp. 3969–3974. doi:10.1109/IECON43393.2020.9255373.