Degradation trajectories prognosis for PEM fuel cell systems based on Gaussian process regression

Deng, Huiwen; Hu, Weihao; Cao, Di; Chen, Weirong; Huang, Qi; Chen, Zhe; Blaabjerg, Frede

Published in:
Energy

DOI (link to publication from Publisher):
10.1016/j.energy.2021.122569

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

Degradation trajectories prognosis for PEM fuel cell systems based on Gaussian process regression

Huiwen Deng[1], Weihao Hu[1,*], Di Cao[1], Weirong Chen[2], Qi Huang[1], Zhe Chen[3], Frede Blaabjerg[3]

[1] School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China

[2] School of Electrical Engineering, Southwest Jiaotong University, Chengdu, 610031, China

[3] Department of Energy Technology, Aalborg University, Aalborg, Denmark

[*] Author to whom correspondence should be addressed; E-mail: whu@uestc.edu.cn

Abstract: The aging trajectory prognosis is an effective tool to prolong the lifespan and lower the cost of proton exchange membrane fuel cell (PEMFC) systems. In this paper, Gaussian process regression modeling frameworks based on sparse pseudo-input Gaussian process (SPGP) and variational auto-encoded deep Gaussian process (VAE-DGP) are proposed to predict the degradation trend and cope with the model uncertainty for PEMFCs. The optimal hyper parameters and pseudo-input locations are obtained with conjugate gradient by maximizing the marginal likelihood. Besides, the variational parameters and closed-form variational lower bound are optimized through variable inference. Radial basis function (RBF) kernel is utilized to determine the priori distribution of Gaussian process. Then stack voltage and output power are extracted as health indicators (HIs). To fully demonstrate the prediction performance, long-term experimental validations with both static and dynamic aging testing profiles are performed, single-input and multi-input structures are respectively constructed in SPGP and VAE-DGP for comparison with existing models. The results show that the proposed methods outperform other data-driven methods, SPGP is more suitable for large data regime and VAE-DGP operates better with small data regime. Finally, the performance evolution is also presented with a 95% confidence interval to validate the mapping accuracy further.

Keywords: Proton exchange membrane fuel cells (PEMFCs); Sparse pseudo-input Gaussian process (SPGP); Variational auto-encoded deep Gaussian process (VAE-DGP); Data-driven model; Degradation prognosis.

1. Introduction

Fuel cell technology has been considered as one of the most attractive power sources for chemical energy into electrical energy for the future, which is very environmentally friendly in power generation [1] [2]. Among all types of fuel cells, proton exchange membrane fuel cell (PEMFC), with the merits of high power density, rapid start up and low operating temperature, has great potential in portable devices, backup power and transportations [3]-[5] et al. However, the short lifespan and high cost of PEMFC are still critical barriers today for large scale application at a worldwide level [6]. In the practical engineering, the PEMFC is subjected to rough working conditions when it is integrated to the transportations, such as hybrid trams [5] and vehicles [7]. Its performance is strongly affected by the mechanical, thermal and electrochemical

degradations. The compression of gas diffusion layer (GDL), carbon corrosion, polymer membrane degradation due to drying, flooding and gas crossover, and platinum dissolution in catalyst layer are the major degradation factors, especially for the long-term operating mode and frequent load change [8]. Therefore, to prolong the limited lifetime of PEMFC and realize efficient control, future degradation behavior should be predicted with precisely forecasting methods [9] [10].

Recently, a growing number of methods on degradation prognosis have been developed, which can be generally divided into three categories: model-driven, data-driven and fusion approach [8]. The model-based methods predict the aging tendency by building a physical, empirical or semi-empirical degradation model from the physical or empirical point of view. In [1] [11] [12], extended Kalman filter (EKF) and unscented Kalman filter (UKF) were utilized to estimate the fuel cell state of health. A multi-physical degradation model with particle filter (PF) approach was presented in [10] to predict different degradation phenomena in the fuel cell. Ulteriorly, a global prediction framework composed of parameter extraction and joint PF structure was applied to achieve the power behavior forecasting, less data were used while good prognostics result was offered as well [13]. A reconstructed empirical fuel cell model based on the output voltage was proposed to observe the voltage degradation, the physical parameters corresponding to the voltage drop were estimated to represent the fade tendency during the aging process [14]. In addition, an improved prediction method based on the error correction gray model (GM) was introduced in [15], achieving at least 11.7% accuracy improvement against the traditional model. Although the model-based prediction approaches can get rid of large amount of experimental data, the fully understanding of the degradation mechanisms is impossible. Thus, it is difficult to establish a complete and accurate mathematical model in multi-time scale, multi-space scale and multiple physical domains.

With the development of artificial intelligence (AI) technology, the data-based approaches gain popularity to implement the degradation prediction and RUL estimation for PEMFCs [16]-[28]. The advantages of simplicity and no need to thoroughly study the aging mechanisms help them to ignore the physical or empirical models, only the historical monitoring data is utilized to make prediction. According to [16], a self-adaptive relevance vector machine (RVM) method with a modified design matrix was developed to obtain an accurate prediction of the aging mechanisms for PEMFCs. Afterwards, the modified RVM in [17] was further used to enhance the accuracy and robustness of the presented model. The on-line model-free method based on the summation wavelet-extreme learning machine (SW-ELM) algorithm was developed in [18], ensemble and incremental learning were used in the aging model. Based on the deep learning, the deep belief network (DBN) and ELM were integrated to deal with the highly nonlinear statistical voltage data for PEMFC [19]. The adaptive neuro-fuzzy inference system with fuzzy-c-means (ANFIS-FCM) demonstrated in [7] showed satisfied short-term prognostics performance. In recent years, the long short-term memory (LSTM) recurrent neutral network (RNN) is often

applied in PEMFC aging prediction [20]-[23]. Thereinto, the improved grid LSTM architecture [21], stacked LSTM model [22], and bi-directional LSTM with attention mechanism (BILSTM-AT) structure [23] were suitable for the time series forecasting, which outperformed the traditional LSTM. As a new paradigm, the echo state network (ESN) was first implemented by Morando et.al in [24] to estimate the RUL of PEMFC. Later, the authors optimized the ESN structure with Hurst coefficient and wavelet filter in [25]. Furthermore, the multi-reservoir ESN (MR-ESN) [26], ensemble ESN [3], multiple inputs and multiple outputs ESN (MIMO-ESN) [27] were also implemented to enhance the prediction accuracy for PEMFC stack. Moreover, some other discrete wavelet transforms (DWTs) were utilized to deal with the nonstationary signals, and prognostics was made on a regression model [28]. However, most of the data-based methods cited above choose the stack voltage as the HI, in fact the power signal also includes the degradation process, and it is rarely considered in the literatures. Moreover, the recovery phenomena contained in the HIs exhibit high nonlinear and nonstationary characteristics, these features are difficult to be fully extracted. Furthermore, most of prognostic methods using different kinds of filters to pre-process the experimental data that contains fluctuation and noise, directly using the data including spikes is infrequent. Last but not least, deterministic estimations of stack voltage are achieved in those literatures, confidence level of the prediction of HIs is less considered as well.

Regarding fusion approaches, a novel data fusion approach based on moving window technique was illustrated to capture the long-range fade trend and short-term nonlinear characteristics [29]. Later, a hybrid prediction approach based on empirical model and physical aging model (PAM) was further given in [30], and the time delay neural network (TDNN) was used to obtain the final prediction result. The evolutionary algorithm and ANFIS were combined in [31] and auto-regressive integrated moving average (ARIMA) was cooperated with LSTM architecture in [32], where both were used to achieve the same purpose. The fusion models still relay on the accurate practical models of PEMFC, and the complex structures of multiple models and training data lead to heavy computing burden.

Considering the shortcomings above, this paper presents novel data-based methodologies for PEMFC prognostic. Due to the advantages of highly flexible to capture the features of the original data and intrinsical uncertainty prediction for HIs, the Gaussian process (GP) for regression provides strong nonlinear mapping ability in nonlinear time series prediction problem [33] [34]. In this paper, we propose VAE-DGP and SPGP modeling frameworks to handle the limitations. At present, the use of VAE-DGP and SPGP for PEMFC degradation trajectories prognosis is original explored in this study. The inherent property of providing quantified uncertainty makes them very suitable for using them in safety-critical health diagnostic. By means of a recognition model, the VAE-DGP can avoid the challenges of variational parameter initialization and simplify the inference. The optimal model hyper parameters and pseudo-input locations are obtained using conjugate gradient by

maximizing the marginal likelihood, moreover, the variational parameters and closed-form variational lower bound of VAE-DGP are optimized by assuming a variational posterior distribution for latent variables. Due to inversion of a kernel function RBF, the training complexity of each layer in DGP leads to large computational scale of VAE-DGP. Hence, as a sparse regression method, SPGP with a particular parameterized nonstationary covariance function can reduce the training and prediction cost in each test case. In order to verify the effectiveness of the two methods, comparisons with the existing prognostics are performed by using the experimental data obtained from a 1kW PEMFC platform. And the novelty and contributions are shown as follows:

(1) Different from many prognostic approaches, the proposed VAE-DGP and SPGP modeling frameworks deal with the experimental data without removing the spikes or fluctuations. Thus, the nonlinear and nonstationary features in the recovery phenomena of HIs are reserved.

(2) The integrated HIs, including stack voltage and output power are defined as the indicators of PEMFC in this study. In addition, the future aging trend and the uncertainty of HIs can be simultaneously predicted and quantified based on the presented GP based models, the generalization performance and disturbance sensitivity are improved.

(3) The SPGP shows superiority in large data regime, as for VAE-DGP only small quantity of data is sufficient for performing accurate prediction. Meanwhile, the confidence interval of stack voltage is narrower using VAE-DGP, in terms of the output power prognostic, the predicted confidence interval gotten from SPGP is much more reliable than that of VAE-DGP.

The paper is arranged as follows: Section 2 introduces mathematical principles of standard GPR and the VAE-DGP and SPGP modeling frameworks for aging trend prediction. In Section 3, the experimental PEMFC stacks are illustrated and the implementations of VAE-DGP and SPGP are described. Comparative experimental results are given in Section 4. Finally, conclusions are made in Section 5.

2. Gaussian process regression based modeling

The GP is a new kind of machine learning on the basis of Bayesian theory and statistical learning theory. In this section, the necessary background on GP and GP-based modeling is illustrated.

2.1. Standard GP modeling

The standard GP is a non-parametric model for function estimation. In the probabilistic reasoning framework, a set of experimental training input-output pairs are obtained from particular PEMFCs and stored in $\boldsymbol{X} \in \Re^{N\times Q}$ and $\boldsymbol{Y} \in \Re^{N\times D}$ respectively. That means the N input vectors $\boldsymbol{X}=\{x_n\}_{n=1}^{N}$ and output vectors $\boldsymbol{Y}=\{y_n\}_{n=1}^{N}$ are of dimension Q and D. Then

the unobserved latent function $f = f(\boldsymbol{X})$ responsible for generating $\boldsymbol{Y}$ when giving $\boldsymbol{X}$. In this setting, we place a zero mean and a covariance (or kernel) function k Gaussian process nonparametric prior distribution $f(\boldsymbol{X}) \sim \mathcal{GP}(\boldsymbol{0}, k(\boldsymbol{X}, \boldsymbol{X}'))$ on latent function f. More formally, each output y_n is influenced by noise, then y_n is produced by the corresponding $f(x_n)$ with independent Gaussian noise [35], i.e.:

$$y_n = f(x_n) + \varepsilon_n, \quad \varepsilon_n \sim \mathcal{N}(0, \sigma_\varepsilon^2 \boldsymbol{I}) \tag{1}$$

where $\boldsymbol{I}$ is the $n \times n$ identify matrix.

Denote the latent functions instantiations by $\boldsymbol{F} = \{f_n\}_n^N$, which is normally distributed. Then the conditional distribution of training outputs $\boldsymbol{Y}$ with respect to the training inputs $\boldsymbol{X}$ takes the form:

$$p(\boldsymbol{Y} \mid \boldsymbol{X}) = \int p(\boldsymbol{Y} \mid \boldsymbol{F}) p(\boldsymbol{F} \mid \boldsymbol{X}) d\boldsymbol{F} = \mathcal{N}(\boldsymbol{Y} \mid \boldsymbol{0}, \mathbf{K}_{NN} + \sigma_\varepsilon^2 \boldsymbol{I}) \tag{2}$$

where $\int p(\boldsymbol{Y} \mid \boldsymbol{F}) p(\boldsymbol{F} \mid \boldsymbol{X}) d\boldsymbol{F}$ can also be detailed with the form $\int \prod_{n=1}^{N} p(y_n \mid f_n) p(f_n \mid x_n) d\boldsymbol{F}$, and $\mathbf{K}_{NN} = k(\boldsymbol{X}, \boldsymbol{X})$.

Considering a set of newly acquired sample $\boldsymbol{X}^* \in \Re^{M \times Q}$ and future data point $\boldsymbol{Y}^* \in \Re^{M \times D}$, the relationship between the set of noiseless disturbance prediction output $f(\boldsymbol{X}^*)$ and the corresponding input vectors can be shown by a testing set $\mathcal{T}^* = (\boldsymbol{X}^*, f(\boldsymbol{X}^*))$, where $f(\boldsymbol{X}^*)$ is unknown. The joint distribution of $f(\boldsymbol{X}^*)$ and $\boldsymbol{Y}$ can be written as:

$$\begin{pmatrix} \boldsymbol{Y} \\ f(\boldsymbol{X}^*) \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} k(\boldsymbol{X}, \boldsymbol{X}) + \sigma_\varepsilon^2 \boldsymbol{I} & k(\boldsymbol{X}, \boldsymbol{X}^*) \\ k(\boldsymbol{X}^*, \boldsymbol{X}) & k(\boldsymbol{X}^*, \boldsymbol{X}^*) \end{pmatrix} \right] \tag{3}$$

then the conditional distribution of prediction output $\boldsymbol{Y}^*$ given $\boldsymbol{X}^*$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ is obtained as Gaussian with the predictive mean and variance, which are shown as follows:

$$(\boldsymbol{Y}^* \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{X}^*) \sim \mathcal{N}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \tag{4}$$

with

$$\boldsymbol{\mu}^* = k(\boldsymbol{X}, \boldsymbol{X}^*)(k(\boldsymbol{X}, \boldsymbol{X}) + \sigma_\varepsilon^2 \boldsymbol{I})^{-1} \boldsymbol{Y} \tag{5}$$

$$\boldsymbol{\Sigma}^* = k(\boldsymbol{X}^*, \boldsymbol{X}^*) - k(\boldsymbol{X}^*, \boldsymbol{X})(k(\boldsymbol{X}, \boldsymbol{X}) + \sigma_\varepsilon^2 \boldsymbol{I})^{-1} k(\boldsymbol{X}, \boldsymbol{X}^*) + \sigma_\varepsilon^2 \boldsymbol{I} \tag{6}$$

where the hyper parameters in k and σ_ε^2 can be obtained by maximizing the log-likelihood function, moreover, the PEMFC aging can be predicted when the kernel function $k(\boldsymbol{X}, \boldsymbol{X}')$, prediction input $\boldsymbol{X}^*$, training pairs $\boldsymbol{X}$, $\boldsymbol{Y}$ and noise

perturbation σ_ε are known.

However, the assumption on joint normal distribution of the latent functions may limit the model applicability. To address the limitation, deep Gaussian process (DGP) employed process composition has come into being [36].

2.2. VAE-DGP mathematical implementation

The DGP architecture consists of multiple layers and a recognition model, the mapping beween the consecutive layers is governed by Gaussian process. Observed outputs (Leaves) are the measured signals, which are placed in the bottom layer, and the observed inputs $\boldsymbol{Z} = \boldsymbol{X}_H \in \Re^{N\times Q_z}$ are placed on the top level. Latent nodes $\boldsymbol{X}_l \in \Re^{N\times Q_l}$, where $l = 1, 2, \cdots, H-1$, H is the number of layers. Then the constructed DGP model is shown as:

$$\boldsymbol{Y} = f^{\boldsymbol{Y}}(\boldsymbol{X}_1) + \varepsilon_1 \quad \varepsilon_1 \sim \mathcal{N}(\boldsymbol{0}, \sigma_1^2 \boldsymbol{I}) \tag{7}$$

$$\boldsymbol{X}_l = f^{\boldsymbol{X}_l}(X_{l+1}) + \varepsilon_l^{\boldsymbol{X}_l} \quad \varepsilon_l^{\boldsymbol{X}_l} \sim \mathcal{N}(\boldsymbol{0}, \left(\sigma_l^{\boldsymbol{X}_l}\right)^2 \boldsymbol{I}) \tag{8}$$

where the functions $f^{\boldsymbol{Y}}$, $f^{\boldsymbol{X}_l}$ are obtained from the GP with kernal functions k_Y, k_l. The intermediate nodes invoving in two GPs searve as an input and output respectively, then node $\boldsymbol{X}_1$ takes part in the GPs of input $f^{\boldsymbol{X}_1} \sim \mathcal{GP}(\boldsymbol{0}, k_1(\boldsymbol{X}_1, \boldsymbol{X}_1'))$ and output $f^{\boldsymbol{Y}} \sim \mathcal{GP}(\boldsymbol{0}, k_Y(\boldsymbol{X}, \boldsymbol{X}'))$.

In the supervised learning scenario, the inputs of top layer $\boldsymbol{Z}$ can be observed with traing data $\boldsymbol{X}$ and testing data $\boldsymbol{X}^*$. The bottom layer $\boldsymbol{Y}$ outputs the training set $\boldsymbol{Y}$ and prediction $\boldsymbol{Y}^*$. However, the latent variables $\boldsymbol{X}_l$ in DGP have to be integrated out to establish the Gaussian process. Therefore, the variational distributions $q(\boldsymbol{X}_1)$, $q(\boldsymbol{X}_2), \cdots, q(\boldsymbol{X}_{H-1})$ are adopted to approximate fit posterior distributions $p(\boldsymbol{X}_1 \mid \boldsymbol{Y})$, $p(\boldsymbol{X}_2 \mid \boldsymbol{X}_1), \ldots, p(\boldsymbol{X}_{H-1} \mid \boldsymbol{X}_{H-2})$.The viariational inference can realize the integration of $\boldsymbol{X}_l$ by maximize the lower bound of variation. The marginal likelihood over the measurable data is given as:

$$p(\boldsymbol{Y}) = \int p(\boldsymbol{Y} \mid \boldsymbol{X}_1) \prod_{l=2}^{H} p(\boldsymbol{X}_{l-1} \mid \boldsymbol{X}_l) p(\boldsymbol{X}_H) d\boldsymbol{X}_1 \cdots d\boldsymbol{X}_H . \tag{9}$$

Define $KL(q(\boldsymbol{X}_1)q(\boldsymbol{X}_2)\cdots q(\boldsymbol{X}_{H-1}) \,\|\, p(\boldsymbol{X}_1 \mid \boldsymbol{Y})p(\boldsymbol{X}_2 \mid \boldsymbol{X}_1)\cdots\, p(\boldsymbol{X}_{H-1} \mid \boldsymbol{X}_{H-2}))$ the Kullback-Leibler divergence. To reduce the differene between the distributions, target function $\hat{q} = \arg\min_q KL(\prod_{l=1}^{H-1} q(\boldsymbol{X}_l) \,\|\, p(\boldsymbol{X}_1 \mid \boldsymbol{Y}) \prod_{l=1}^{H-1} p(\boldsymbol{X}_l \mid \boldsymbol{X}_{l-1}))$ should be obtained, futher $KL(\cdot) = p(\boldsymbol{Y}) - \mathcal{L}$, $\mathcal{L}$ is the lower bound of log marginal distribution, which is derived as:

$$\mathcal{L} = \sum_{l=1}^{H-1} \mathcal{F}_l + \sum_{l=1}^{H-2} H(q(\boldsymbol{X}_l)) - KL(q(\boldsymbol{X}_{H-1}) \,||\, p(\boldsymbol{X}_{H-1})) \tag{10}$$

where $\mathcal{F}_1$ and $\mathcal{F}_l$ are free energy of indicicual layers with $l = 2, \cdots, H-1$, $H(q(\boldsymbol{X}_l))$ is the entropy of $q(\boldsymbol{X}_l)$. The expressions are shown as follows:

$$\begin{cases} \mathcal{F}_1 = \left\langle \log p(\boldsymbol{Y} \mid \boldsymbol{X}_1) \right\rangle_{q(\boldsymbol{X}_1)} \\ \mathcal{F}_l = \left\langle \log p(\boldsymbol{X}_{l-1} \mid \boldsymbol{X}_l) \right\rangle_{q(\boldsymbol{X}_{l-1})q(\boldsymbol{X}_l)} \end{cases} \tag{11}$$

$$H(q(\boldsymbol{X}_{l-1})) = -\int q(\boldsymbol{X}_{l-1}) \log q(\boldsymbol{X}_{l-1}) d\boldsymbol{X}_{l-1} \tag{12}$$

$$KL(q(\boldsymbol{X}_{H-1}) \,||\, p(\boldsymbol{X}_{H-1})) = \int q(\boldsymbol{X}_{H-1}) \log \frac{q(\boldsymbol{X}_{H-1})}{p(\boldsymbol{X}_{H-1})} d\boldsymbol{X}_{H-1} \tag{13}$$

the traget function can be transformed as:

$$\hat{q} = \arg\min_q KL(\cdot) = \arg\max_q \mathcal{L}. \tag{14}$$

The $\boldsymbol{X}_l$ can be parameterized by Gaussian distribution $q(\boldsymbol{X}_l) = \prod_{n=1}^{N} \mathcal{N}(x_l^{(n)} \mid \boldsymbol{\mu}_l^{(n)}, \boldsymbol{\Sigma}_l^{(n)})$, $l = 1, 2, \cdots, H-1$. The gradients of $q(\boldsymbol{X}_l)$ are back-propagated through the model. Supposing the prior distribution as $p(\boldsymbol{X}_{H-1}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, then the Eq.(10) is analytic. The variational parameters are $\boldsymbol{\theta} = \left\{ \left\{ \boldsymbol{\mu}_l^{(n)} \right\}_{n=1}^{N}, \left\{ \boldsymbol{\Sigma}_l^{(n)} \right\}_{n=1}^{N} \right\}_{l=1}^{H-1}$, specifically, $\boldsymbol{\mu}_1^{(n)} = g_1(y^n)$ and $\boldsymbol{\mu}_l^{(n)} = g_l(\cdots g_1(y^n))$. The transformation functions g_1 and g_l are parameterized by multi-layer perceptrons (MLPs), and $\boldsymbol{\Sigma}_l^{(n)}$ are supposed to be diagonal with the same datapoint.

As for the vatiational lower bound in Eq.(10), the free energy terms are obtained by a variational approximation technique, denoting the noise-free observations of the observed layer as $\boldsymbol{F}_1 \in \Re^{N \times D}$, inducing inputs as $\boldsymbol{U}_1 \in \Re^{M \times D}$ and $\boldsymbol{Z}_1 \in \Re^{M \times D}$. The free energy of layer $\boldsymbol{Y}$ and hidden layers $\boldsymbol{X}_l$ can be lower bounded sepearately by $\mathcal{F}_1 \geq \left\langle \log p(\boldsymbol{Y} \mid \boldsymbol{F}_1) - KL(q(\boldsymbol{U}_1) \,||\, p(\boldsymbol{U}_1)) \right\rangle_{p_1(\cdot)}$ and $\mathcal{F}_l \geq \left\langle \log p(\boldsymbol{X}_{l-1} \mid \boldsymbol{F}_l) - KL(q(\boldsymbol{U}_l \mid \boldsymbol{X}_{l-1}) \,||\, p(\boldsymbol{U}_l)) \right\rangle_{p_l(\cdot)}$ with $p_1(\cdot) = p(\boldsymbol{F}_1 | \boldsymbol{U}_1, \boldsymbol{X}_1) q(\boldsymbol{U}_1) q(\boldsymbol{X}_1)$, $p_l(\cdot) = p(\boldsymbol{F}_l | \boldsymbol{U}_l, \boldsymbol{X}_l)\ q(\boldsymbol{U}_l | \boldsymbol{X}_{l\text{-}1}) q(\boldsymbol{X}_{l-1}) q(\boldsymbol{X}_l)$, $l = 2, \cdots, H-1$. Combing with Eq.(10), a closed form variational lower bound can be obtained. The conjugate gradient based method is utilized to optimize the variational variables and model hyper parameters. However, due to the high traning complexity of VAE-DGP in each layer, the sparse Gaussian process using pseudo-inputs is introduced to reduce the computation complexity.

2.3. SPGP modeling approach

The sparse Gaussian process is regarded as a Bayesian regression model, the covariance function is parameterized by the locations of M pseudo-inputs with $M \ll N$. An active set found through a continuous optimization is not constrained in a subset of data, on the basis of the advantage, the precise locations of SPGP can be fine-tuned to improve the fit quality. It can be considered as the standard GP with a specific non-stationary covariance function parameterized by pseudo-input [37].

Denote the pseudo data set $\bar{\mathcal{D}}$ of size M, pseudo-inputs $\bar{\boldsymbol{X}} = \{\bar{\boldsymbol{x}}_m\}_{m=1}^{M}$ and pseudo targets $\bar{\boldsymbol{f}} = \{\bar{f}_m\}_{m=1}^{M}$. The pseudo targets $\bar{\boldsymbol{f}}$ are not real observations, different from the actual observed values that should be assumed noisy, it is trivial to include noise variances for the targets $\bar{\boldsymbol{f}}$. Therefore, the single data point likelihood takes the form:

$$p(y \mid \boldsymbol{x}, \bar{\boldsymbol{X}}, \bar{\boldsymbol{f}}) = \mathcal{N}(y \mid \boldsymbol{k}_{\boldsymbol{x}}^T \boldsymbol{K}_M^{-1} \bar{\boldsymbol{f}}, K_{\boldsymbol{xx}} - \boldsymbol{k}_{\boldsymbol{x}}^T \boldsymbol{K}_M^{-1} \boldsymbol{k}_{\boldsymbol{x}} + \sigma_\varepsilon^2) \tag{15}$$

where $\left[\boldsymbol{K}_M\right]_{mm'} = K(\bar{\boldsymbol{x}}_m, \bar{\boldsymbol{x}}_{m'})$, $\left[\boldsymbol{k}_{\boldsymbol{x}}\right]_m = K(\bar{\boldsymbol{x}}_m, \boldsymbol{x})$ for $m = 1, 2 \cdots, M$ with new input data $\boldsymbol{x}$.

With parameterized mean function and input dependent noise model in Eq. (15), the data likelihood can be given as:

$$p(\boldsymbol{Y} \mid \boldsymbol{X}, \bar{\boldsymbol{X}}, \bar{\boldsymbol{f}}) = \prod_{n=1}^{N} p(y_n \mid x_n, \bar{\boldsymbol{X}}, \bar{\boldsymbol{f}}) = \mathcal{N}(\boldsymbol{Y} \mid \mathbf{K}_{NM} \mathbf{K}_M^{-1} \bar{\boldsymbol{f}}, \boldsymbol{\Lambda} + \sigma_\varepsilon^2 \boldsymbol{I}) \tag{16}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda})$, $\lambda_n = K_{nn} - \boldsymbol{k}_n^T \boldsymbol{K}_M^{-1} \boldsymbol{k}_n$, and $\left[\boldsymbol{K}_{NM}\right]_{nm} = K(\boldsymbol{x}_n, \bar{\boldsymbol{x}}_m)$.

Learning in the model includes finding an appropriate set of pseudo data to well describe the real data. Rather than maximize the likelihood in Eq. (16), the pseudo targets $\bar{\boldsymbol{f}}$ can be integrated out with a Gaussian prior, which can be given as:

$$p(\bar{\boldsymbol{f}}, \bar{\boldsymbol{X}}) = \mathcal{N}(\bar{\boldsymbol{f}} \mid \boldsymbol{0}, \boldsymbol{K}_M). \tag{17}$$

It is reasonable for the pseudo data to be distributed in the same way with the real data, but placing a prior on the pseudo-inputs $\bar{\boldsymbol{X}}$ is difficult. So maximum likelihood is applied to handle this. Based on Eq. (16) and (17), the posterior distribution over $\bar{\boldsymbol{f}}$ is described as follows:

$$p(\bar{\boldsymbol{f}} \mid \bar{\mathcal{D}}, \bar{\boldsymbol{X}}) = \mathcal{N}(\bar{\boldsymbol{f}} \mid \boldsymbol{K}_M \boldsymbol{Q}_M^{-1} \boldsymbol{K}_{MN} (\boldsymbol{\Lambda} + \sigma_\varepsilon^2 \boldsymbol{I}) \boldsymbol{y}, \boldsymbol{K}_M \boldsymbol{Q}_M^{-1} \boldsymbol{K}_M) \tag{18}$$

where $\boldsymbol{Q}_M = \boldsymbol{K}_M + \boldsymbol{K}_{MN} (\boldsymbol{\Lambda} + \sigma_\varepsilon^2 \boldsymbol{I})^{-1} \boldsymbol{K}_{NM}$. Then with the new input $\boldsymbol{X}^*$ and combing the likelihood and posterior in Eq. (15) and (18), the predictive distribution is obtained by:

$$p(y_* \mid \boldsymbol{X}^*, \bar{\mathcal{D}}, \bar{\boldsymbol{X}}) = \int d\bar{\boldsymbol{f}}\, p(y_* \mid \boldsymbol{x}_*, \bar{\boldsymbol{X}}, \bar{\boldsymbol{f}}) p(\bar{\boldsymbol{f}} \mid \bar{\mathcal{D}}, \bar{\boldsymbol{X}}) = \mathcal{N}(y_* \mid \boldsymbol{\mu}_*, \boldsymbol{\sigma}_*^2) \tag{19}$$

with

$$\boldsymbol{\mu}_* = \boldsymbol{k}_*^T \boldsymbol{Q}_M^{-1} \boldsymbol{K}_{MN} (\boldsymbol{\Lambda} + \sigma_\varepsilon^2 \boldsymbol{I}) \boldsymbol{y} \tag{20}$$

$$\boldsymbol{\sigma}_*^2 = K_{**} - \boldsymbol{k}_*^T (\mathbf{K}_M^{-1} - \mathbf{Q}_M^{-1}) \boldsymbol{k}_* + \sigma_\varepsilon^2 . \tag{21}$$

The computational cost is determined by $\boldsymbol{K}_{MN}(\boldsymbol{\Lambda} + \sigma_\varepsilon^2 \boldsymbol{I})\boldsymbol{K}_{NM}$ during the calculation of $\boldsymbol{Q}_M$, then the sparse regression method has $\mathcal{O}=(M)$ for the mean and $\mathcal{O}=(M^2)$ for the variance during the prediction per test case.

As for the pseudo-input locations $\bar{\boldsymbol{X}}$ and hyper parameters $\boldsymbol{\Theta}$, they can be obtained with the marginal likelihood from Eq. (16) and (17), which is shown as follows:

$$p(\boldsymbol{Y} \mid \boldsymbol{X}, \bar{\boldsymbol{X}}, \boldsymbol{\Theta}) = \int d\bar{\boldsymbol{f}}\, p(\boldsymbol{Y} \mid \boldsymbol{X}, \bar{\boldsymbol{X}}, \bar{\boldsymbol{f}}) p(\bar{\boldsymbol{f}} \mid \boldsymbol{X}) = \mathcal{N}(\boldsymbol{Y} \mid \mathbf{0}, \mathbf{K}_{NM}\mathbf{K}_M^{-1}\mathbf{K}_{MN} + \boldsymbol{\Lambda} + \sigma_\varepsilon^2 \boldsymbol{I}) . \tag{22}$$

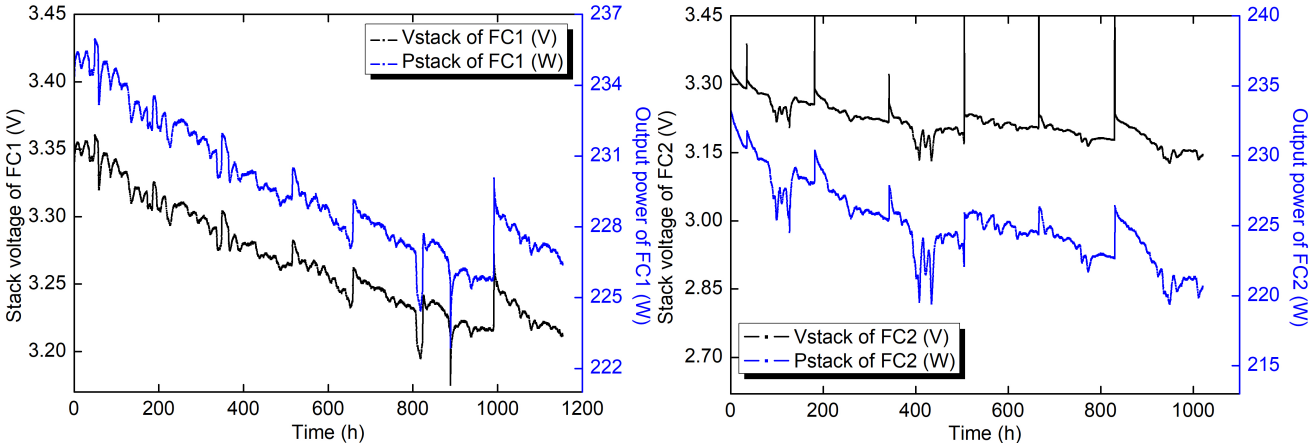
Then the marginal likelihood can be maximized in terms of all the parameters $\{\bar{\boldsymbol{X}}, \boldsymbol{\Theta}\}$ by gradient ascent.

3. The long term prognostics for PEMFCs

In this section, the experimental data and the proposed aging performance prediction methods are introduced.

3.1. Experiment Settings

The experiments for PEMFC considered in this paper is obtained from FCLAB Research Federation [38]. A 1kW PEMFC platform was built for the degradation experiments. The stack has 5 cells with active area of 100 cm^2 in each cell. During the experiments, the operating temperature of the stack is controlled about 55°C, the gas pressure is around 1.3 bar, the relative humidity maintains at 52%, and the stoichiometry ratios at anode and cathode are within 1.5-2. Two tests are conducted to obtain the degradation data, the first test in the stationary regime is realized on FC1, the load current was set as 70 A, and the second test operated on FC2 is under dynamic condition with a high frequency triangular ripples in the constant current (70A), the ripples have 14A peak-to-peak value at the frequency of 5 kHz. And the accelerated aging test is used to emulate the effect of power converters on the output of stack, more experimental configuration can be found in [7] [19] [34]. Moreover, the stack voltage and output power are utilized for prognostics in this paper, where the stack voltage is an admitted HI for PEMFC in many studies [7] [21]-[23] [27], but the stack current always treated as the scheduling variable also contains degradation information, prediction of output power can ensure sufficient power supply anytime. Furthermore, the raw experimental data including 143862 data sets for FC1 and 127370 data sets for FC2 are obtained, before conducting the prediction tests with this data, the noise is removed, spikes and nonlinearity are reserved in HIs, in addition, the HIs are resampled every half an hour, which are shown in Fig. 1.



(a) (b)

Fig. 1. Degradation trend of health indicators. (a) In stationary regime. (b) Under dynamic situation.

As we can see, the aging trend of HIs under dynamic current (FC2) in Fig.1 (b) are much more serious than that of FC1 in Fig.1 (a). Moreover, the nonlinearity and periodicity are reserved due to the degradation recovery phenomenon in HIs to further demonstrate the effectiveness of presented methods.

3.2. Prognostic Implementation

Based on the PEMFC experimental test bench, the prediction data of HIs can be obtained by the proposed models shown in Fig. 2 (a). As we know, the deeper structures can improve the training accuracy to some extent. Unfortunately for the VAE-DGP when the hidden layers are greater than two, the accuracy improvement is not significant [39]. Meanwhile, the increase of number of hidden layers can increase the complexity of model parameters and the computational burden linearly. Therefore, in this paper a two-level structure with one hidden layer is adopted in VAE-DGP model. According to Section 2, although the prior of VAE-DGP cannot be simply interpreted as a GP, the input $\{x_1, x_2, \cdots, x_n\}$ and output pairs $\{y_1, y_2, \cdots, y_n\}$ are mapped by the relationship between the layers, as the layers are connected by the GP models $f^{X_1}(\cdot)$ and $f^{Y}(\cdot)$, the $\{X_1^1, X_1^2, \cdots X_1^n\}$ is the output of the upper GP model (also the input of the next GP model), the variable parameter $\boldsymbol{Z}_1$ and the inducing variable $\boldsymbol{U}_1$ are treated as the inducing inputs in each layer to extend the probability space of each model. As for the SPGP model, a set of latent variables $\{\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \cdots, \overline{\boldsymbol{x}}_m\}$ is referred as the pseudo inputs to find the sparse approximation of original covariance matrix, the latent variables are obtained from the input locations, so $\{\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \cdots, \overline{\boldsymbol{x}}_m\}$ also follows the joint Gaussian distribution. By using Bayes rule, the posterior distribution over pseudo outputs $\{\overline{f}_1, \overline{f}_2, \cdots, \overline{f}_m\}$ can be obtained, then the predictive distribution is obtained by integrating the posterior in Eq. (18) with the likelihood in Eq. (15).

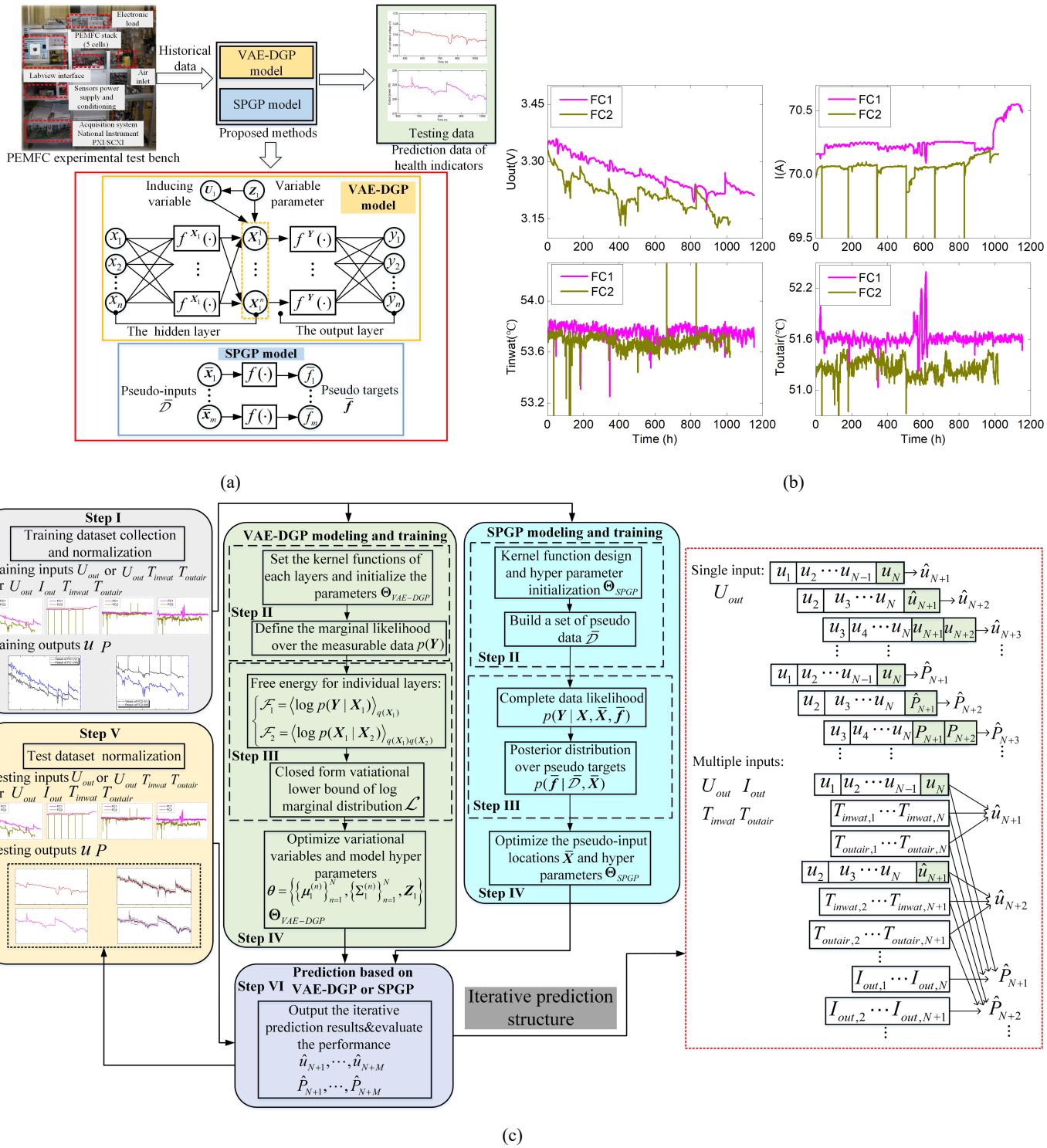


Fig. 2. (a) Implementation architecture of DP-based methods. (b) Operating parameters for PEMFC prediction. (c) Prediction flow chart of PEMFC aging tendency based on VAE-DGP and SPGP.

What is more important is that the kernel function design may affect the performance of the DP-based models. Denote the covariance function in the hidden layers of VAE-DGP and SPGP as RBF kernel, which can be described by:

$$k(\boldsymbol{x},\boldsymbol{x}^{'})=\sigma_f^2\exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{x}^{'}\right)^T M^{-1}\left(\boldsymbol{x}-\boldsymbol{x}^{'}\right)\right) \tag{23}$$

where σ_f^2 is the signal variance, M=diag(l_f^2) , l_f is the scale of variance. The hyper parameter vectors $\Theta_{VAE-DGP}=\left\{\left\{\sigma_f^{X_1},\sigma_f^{Y}\right\},\left\{M_{X_1},M_{Y}\right\},\left\{\sigma_1^{X_1},\sigma_1\right\}\right\}$ and $\Theta_{SPGP}=\left\{\sigma_f,M,\sigma_\varepsilon\right\}$ that can affect the ability of data expressing of kernel functions are optimized by the marginal likelihood method respectively. Based on conjugat egradient, taking partial derivation of the negative log likelihood functions with respect to $\Theta_{VAE-DGP}$ and Θ_{SPGP}, the optimization of hyper parameters can be acquired by minimizing the partial derivatives. Meanwhile, the predicted mean value (point estimation) of HIs and covariance (measure of uncertainty) of the test dataset can be further obtained.

Considering enormous raw data obtained from the monitoring system in this study, but not every operating parameter can affect the degradation performance of PEMFC [36]. Therefore, the forecasting of HIs using VAE-DGP and SPGP methods are conducted with single input and multiple inputs operating parameters. Naturally, the stack voltage is regarded as the single input variable both in FC1 and FC2. With regard to the selection of multiple input parameters, on the basis of the conclusions in [27], the single stack voltage U_{out}, the stack current I_{out}, the inlet temperature of water T_{inwat} and the outlet temperature of air T_{outair} shown in Fig. 2 (b) are included.

Specifically, the flow chart for PEMFC aging tendency prediction based on the proposed VAE-DGP and SPGP is shown in Fig.2 (c), the entire calculation process of VAE-DGP is described as follows:

Step I: Collect and normalize the training data $\left\{u_n\right\}_{n=1}^{N}$, $\left\{I_{out,n}\right\}_{n=1}^{N}$, $\left\{T_{inwat,n}\right\}_{n=1}^{N}$ and $\left\{T_{outair,n}\right\}_{n=1}^{N}$ in terms of the sampling time $\left\{t_n\right\}_{n=1}^{N}$. The training inputs are divided into three groups, namely the single U_{out}; the U_{out}, T_{inwat} and T_{outair}; and the U_{out}, I_{out}, T_{inwat} and T_{outair}.

Step II: Set the kernel functions of VAE-DGP in Eq. (23), and initialize the hyper parameters and noise perturbation in $\Theta_{VAE-DGP}$. Then define the marginal likelihood $p(\boldsymbol{Y})$ in Eq. (9).

Step III: Set the free energy for every individual layers in Eq. (11), the variational inference is used to obtain vatiational lower bound $\mathcal{L}$ in Eq.(10).

Step IV: The conjugate gradient based method is utilized to optimize the variational variables and hyper parameters,i.e. $\boldsymbol{\theta}=\left\{\left\{\boldsymbol{\mu}_1^{(n)}\right\}_{n=1}^{N},\left\{\boldsymbol{\Sigma}_1^{(n)}\right\}_{n=1}^{N},\boldsymbol{Z}_1\right\}$ and $\Theta_{VAE-DGP}$.

Step V: Send the test data into the optimized VAE-DGP model, where the three groups of testing inputs are normalized.

Step VI: With the input test dataset and operating parameters, the aging tendency of HIs in the form of mean and

covariance based on VAE-DGP model can be obtained by the iterative prediction structure.

With respect to the SPGP model, the calculation process of **Step I**, **Step V** and **Step VI** are almost the same with VAE-DGP, therefore, the detailed description is omitted, and the rest of the steps for SPGP modeling and training are given as:

Step II: Design the kernel function in Eq. (23) and initialize the hyper parameters $\boldsymbol{\Theta}_{SPGP}$ of SPGP. Then select a set of pseudo data $\bar{\mathcal{D}}$ to reduce the sample size of training sets.

Step III: Define the complete data likelihood in Eq. (16) and posterior distribution over the pseudo targets $\bar{\boldsymbol{f}}$ in Eq. (18).

Step IV: Maximize the marginal likelihood in terms of parameters $\{\bar{\boldsymbol{X}}, \boldsymbol{\Theta}_{SPGP}\}$ by gradient ascent.

Moreover, in Fig. 2 (c) the iterative prediction is utilized, forcing the predicted value as the new input value. Taking the stack voltage for example, the sampling data $u_1, \cdots, u_N$ is trained to predict $\hat{u}_{N+1}$, then $u_2, \cdots, \hat{u}_{N+1}$ is used to obtain $\hat{u}_{N+2}$, after two time point values prediction, the input voltage values including the prediction values are replaced by the real values to predict $\hat{u}_{N+3}$, and the sequence is the same for $\hat{u}_{N+4}$, $\hat{u}_{N+5}$ …etc., this action is repeated for $\hat{P}_{N+1}$, $\hat{P}_{N+2}$, $\hat{P}_{N+3}$ … etc. in the case of single input until the prediction horizon is reached. As for multiple inputs, the operating parameters are also utilized to infer the future evolution of HIs.

4. Experimental results

The prognostic performance of the proposed models is conducted in this section, where the stack voltage and output power are considered as degradation indicator to reflect the aging status of PEMFC. Moreover, the degradation estimation of BPNN, LSTM, Stacked LSTM algorithms are also applied to verify the superiority of DP-based models.

4.1. Parameter configuration and criteria definition

In the experiments of FC1 and FC2, before the training process, the initial parameter configuration is firstly introduced. The inputs and outputs of the models are normalized within [0, 1]. For FC1, the number of hidden layer for VAE-DGP is 1, the number of latent dimension is 20, inducing point number is chosen as 30, the mean value of latent input is initialized randomly based on the standard nominal distribution, the deviation of latent function is set as 0.01, the inducing output is initialized randomly with a deviation of 0.01 according to the nominal distribution, and the maximum iterations are 800 and 1500 for VAE-DGP. Besides, the inducing point number is chosen as 30 for SPGP, the maximum iteration are selected as 15 and 5 in the model. For FC2, the latent dimension number for VAE-DGP is changed to 35, the inducing point number is 70. As for SPGP, inducing point number is 60, and maximum iteration are 150 and 50. Moreover, all the kernel variance and

length scale are set as 2. Then the parameters are optimized during the training process.

In the testing process, the commonly used evaluation criteria of forecasting accuracy, i.e. room mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE), are adopted to study the prediction performance, the calculation formulas are shown as follows:

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(y_i(n) - \hat{y}_i(n)\right)^2} \tag{24}$$

$$MAPE = \frac{1}{M}\left(\sum_{i=1}^{M}\left|\frac{y_i(n) - \hat{y}_i(n)}{y_i(n)}\right|\right) \tag{25}$$

$$MAE = \frac{1}{M}\left(\sum_{i=1}^{M}\left|\hat{y}_i(n) - y_i(n)\right|\right) \tag{26}$$

where M is the prediction horizon, $\hat{y}_i(n)$ is the forecasted value and $y_i(n)$ is the original signal.

In order to further demonstrate the effectiveness of the proposed models, three metrics of performance improvement ratio are described as follows:

$$P_{RMSE} = \frac{RMSE_{other} - RMSE_{proposed}}{RMSE_{other}} \times 100 \tag{27}$$

$$P_{MAPE} = \frac{MAPE_{other} - MAPE_{proposed}}{MAPE_{other}} \times 100 \tag{28}$$

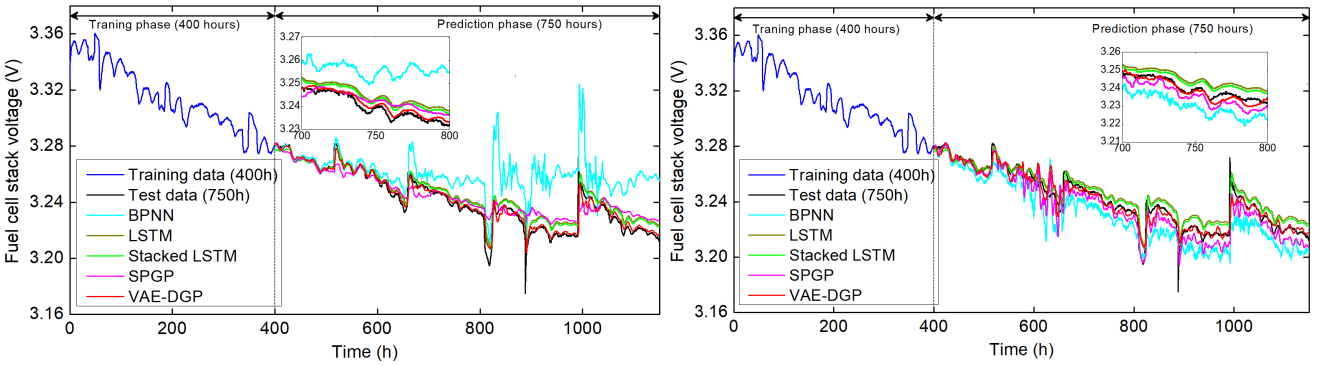
$$P_{MAE} = \frac{MAE_{other} - MAE_{proposed}}{MAE_{other}} \times 100 . \tag{29}$$

The subscript "*other*" represent the index of compared models, while subscript "*proposed*" stands for the index of proposed models in this paper. In the quantitative metrics, RMSE can make more intuitive evaluation on the order of magnitude; MAPE considers the ratio between the predicted value error and the real data; MAE is able to well reflect the predicted value error in actual situation; P_{RMSE}, P_{MAPE} and P_{MAE} are used to reflect the percentage improvement in the model accuracy between the proposed models and the other models.

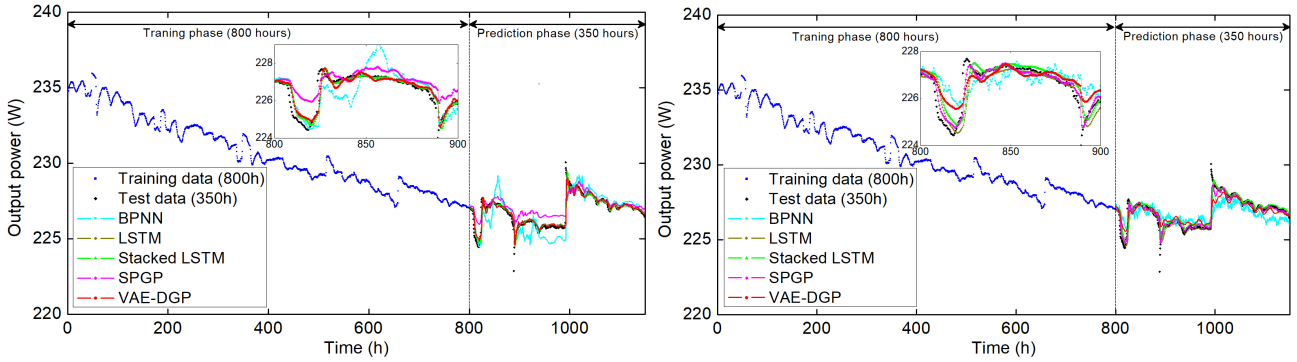
4.2. Prognostic under static operation mode

In static current mode, the total operation time is 1050h, two cases are performed under the training phase (TP) of 400 hours and 800 hours, Case 1 is the single input situation and Case 2 is the multiple-input situation. In the meanwhile, the degradation indexes are predicted separately. The aging prediction of stack voltage and output power for FC1 are presented

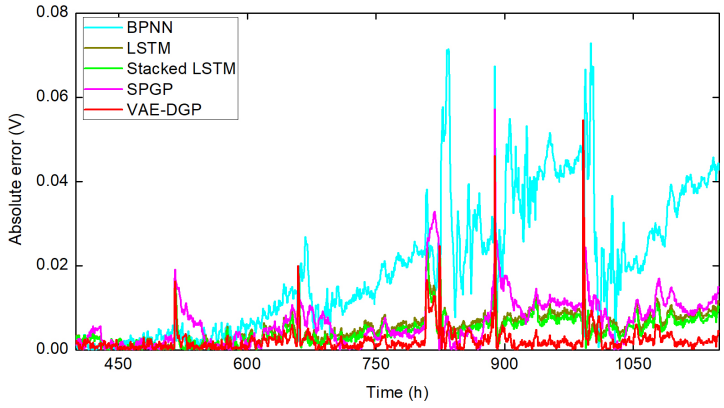
in Fig. 3, and the accuracy of the model is compared in Table 1 and 2. As we can see from Fig. 3 (a) and (b), the first 400h among the 1150h data are applied to train the model, and the rest 750h are used for demonstrating the predictive performance, as for Fig. 3 (c) and (d), the testing duration is 350h. It is obvious that in Fig. 3 (a) and (c), the performance of VAE-DGP is better than Stacked LSTM, LSTM, SPGP and BPNN with single input. Under multiple-input situation, the Fig. 3 (b) shows that the VAE-DGP still performs the best in stack voltage degradation forecasting. However, when the training data (including the length of TP and the input parameter number) increases in Fig.3 (d), the SPGP outperforms the other four methods, the value of RMSE, MAPE and MAE are obtained as 0.30956, 0.00061 and 0.14026 accordingly, and the VAE-DGP results in poor performance, which is even worse than that of Stacked LSTM and LSTM.



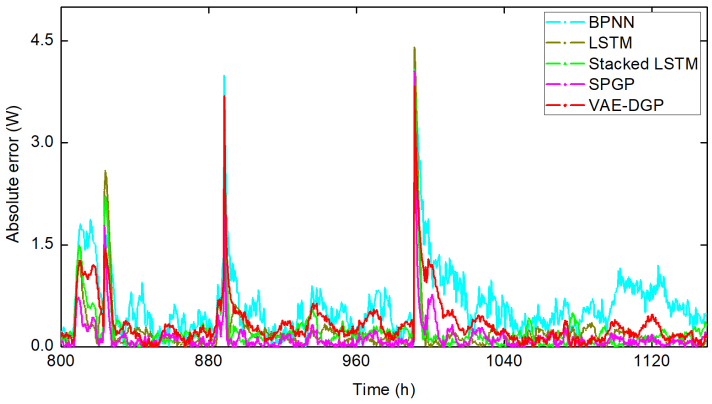
(a) (b)



(c) (d)



(e)



(f)

Fig. 3. (a) Stack voltage degradation prediction of FC1 when TP=400h with single input. (b) Stack voltage degradation prediction of FC1 when TP=400h with multiple-input. (c) Output power degradation prediction of FC1 when TP= 800h with single input. (d) Output power degradation prediction of FC1 when TP= 800h with multiple-input. (e) Absolute error comparison of (a). (f) Absolute error comparison of (d).

Moreover, Fig. 3 (e) and (f) depict the prediction absolute error curves based on the five methods in Fig.3 (a) and (d), it should be noted that the absolute error of stack voltage prediction with VAE-DGP is minimum in Fig.3 (e). However, in Fig.3 (f) the absolute errors of output power with Staked LSTM and SPGP are smaller, even though during 998h-1001h the absolute error of SPGP method is slightly higher than that of Staked LSTM method, it is still the smallest during the whole forecasting process.

Table 1 The comparative prediction results of FC1 under two cases when HI is stack voltage.

Methods	Case 1: Single input U_{out}						Case 2: Multiple inputs U_{out}, T_{inwat}, T_{outair}					
	Training phase (400h)			Training phase (800h)			Training phase (400h)			Training phase (800h)		
	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE
BPNN	0.02334	0.00536	0.01731	0.00868	0.00205	0.00663	0.01349	0.00361	0.01173	0.00960	0.00235	0.00759
LSTM	0.00683	0.00168	0.00543	0.00497	0.00078	0.00253	0.00696	0.00173	0.00561	0.00554	0.00081	0.00261
Stacked LSTM	0.00574	0.00134	0.00436	0.00493	0.00073	0.00235	0.00598	0.00142	0.00460	0.00522	0.00072	0.00233
SPGP	0.01007	0.00239	0.00774	0.00528	0.00093	0.00312	0.00753	0.00182	0.00591	**0.00484**	**0.00070**	**0.00229**
VAE-DGP	**0.00408**	**0.00072**	**0.00222**	**0.00469**	**0.00061**	**0.00198**	**0.00552**	**0.00105**	**0.00342**	0.00587	0.00111	0.00358

Table 2 The comparative prediction results of FC1 under two cases when HI is output power.

Methods	Case 1: Single input U_{out}						Case 2: Multiple inputs U_{out}, I_{out}, T_{inwat}, T_{outair}					
	Training phase (400h)			Training phase (800h)			Training phase (400h)			Training phase (800h)		
	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE
BPNN	1.41032	0.00440	0.99953	0.73471	0.00241	0.54669	0.91247	0.02964	0.67516	0.78968	0.00270	0.61324
LSTM	0.47443	0.00162	0.36852	0.36666	0.00075	0.17032	0.52188	0.00152	0.34571	0.46067	0.00108	0.23986
Stacked LSTM	0.44626	0.00149	0.33892	0.36611	0.00065	0.14779	0.49903	0.00144	0.32802	0.45116	0.00104	0.23577
SPGP	0.63475	0.00192	0.43798	0.62080	0.00172	0.39038	**0.39760**	**0.00118**	**0.26875**	**0.30956**	**0.00061**	**0.14026**
VAE-DGP	**0.30084**	**0.00067**	**0.15248**	**0.35240**	**0.00063**	**0.14574**	0.52877	0.00155	0.35342	0.53973	0.00160	0.36336

According to Table 1 and 2, the VAE-DGP shows a good prediction accuracy in Case 1, the RMSE, MAPE and MAE are 0.00408, 0.00072, 0.00222 for stack voltage prediction when TP is 400h, and 0.35240, 0.00063, 0.14574 for output power prediction when TP is 800h. In Case 2, the SPGP can realize the lowest value in terms of the criteria except when the training

phase is 400h for stack voltage prediction. Under the static operation situation, the VAE-DGP shows superiority in stack voltage prediction, the estimation of output power gets worse when the training data increases. Moreover, in order to further investigate the performance of the DP-based models, the improved percentages of BPNN, Stacked LSTM, LSTM and VAE-DGP (or SPGP) compared with SPGP (or VAE-DGP) in FC1 are shown in Fig. 4, it is clear that the VAE-DGP obtains significantly higher accuracy than BPNN, Stacked LSTM, LSTM and SPGP in Case 1 when TP is 800h and 400h respectively. In Case 2, when the TP is changed to 800h for stack voltage forecasting and 400h for output power forecasting, the SPGP improves the P_{RMSE}, P_{MAPE} and P_{MAE} compared to other models.

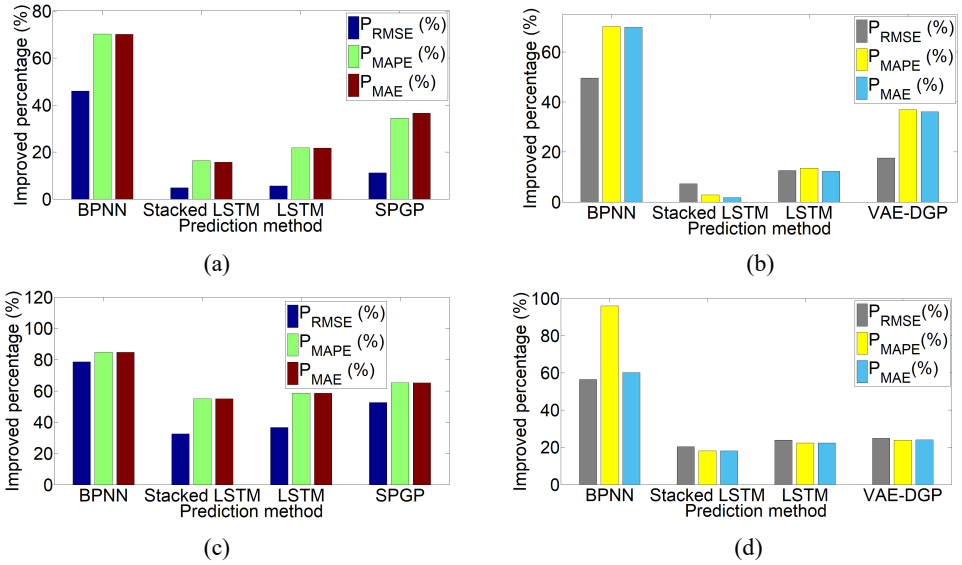


Fig. 4. Accuracy improvement of DP-based models in FC1. (a) TP=800h in Case 1 for stack voltage prediction. (b) TP=800h in Case 2 for stack voltage prediction. (c) TP=400h in Case 1 for output power prediction. (d) TP=400h in Case 2 for output power prediction.

Table 3 The improved percentage of FC1 under two cases when HI is stack voltage.

Comparison models	Case 1: Single input U_{out}						Case 2: Multiple inputs U_{out}, T_{inwat}, T_{outair}					
	Training phase (400h)			Training phase (800h)			Training phase (400h)			Training phase (800h)		
	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}
BPNN→SPGP	56.85	55.41	55.28	39.17	54.63	52.94	44.18	49.58	49.61	**49.58**	**70.21**	**69.82**
LSTM→SPGP	-47.43	-42.26	-42.54	-6.23	-19.23	-23.32	-8.18	-5.20	-5.34	**12.63**	**13.58**	**12.26**
Stacked LSTM →SPGP	-75.43	-78.35	-77.52	-7.09	-27.39	-32.76	-25.91	-28.16	-28.47	**7.27**	**2.78**	**1.72**
VAE-DGP→SPGP	-146.81	-231.94	-248.64	-12.57	-52.45	-57.57	-36.41	-73.33	-72.80	**17.54**	**36.93**	**36.03**
BPNN→VAE-DGP	**82.51**	**86.56**	**87.17**	**45.96**	**70.24**	**70.13**	**59.08**	**70.91**	**70.84**	38.85	52.76	52.83
LSTM→VAE-DGP	**40.26**	**57.14**	**59.11**	**5.63**	**21.79**	**21.73**	**20.68**	**39.30**	**39.03**	-5.96	-37.04	-37.16
Stacked LSTM →VAE-DGP	**28.91**	**46.26**	**49.08**	**4.86**	**16.43**	**15.74**	**7.69**	**26.05**	**25.65**	-12.45	-54.16	-53.64
SPGP→VAE-DGP	**59.48**	**69.87**	**71.31**	**11.17**	**34.40**	**36.53**	**26.69**	**42.30**	**42.13**	-21.28	-58.57	-56.33

Table 4 The improved percentage of FC1 under two cases when HI is output power.

Comparison models	Case 1: Single input U_{out}						Case 2: Multiple inputs U_{out}, I_{out}, T_{inwat}, T_{outair}					
	Training phase (400h)			Training phase (800h)			Training phase (400h)			Training phase (800h)		
	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}
BPNN→SPGP	54.99	56.36	56.18	15.50	28.63	28.59	**56.42**	**96.01**	**60.19**	**60.79**	**77.40**	**77.12**
LSTM→SPGP	-33.79	-18.51	-18.84	-69.31	-129.33	-129.20	**23.81**	**22.36**	**22.26**	**32.80**	**43.51**	**41.52**
Stacked LSTM →SPGP	-42.23	-28.85	-29.22	-69.56	-164.61	-164.14	**20.32**	**18.05**	**18.06**	**31.38**	**41.34**	**40.50**
VAE-DGP→SPGP	-110.99	-186.56	-187.23	-76.16	-173.01	-167.86	**24.80**	**23.87**	**23.95**	**42.64**	**61.87**	**61.39**
BPNN→VAE-DGP	**78.66**	**84.77**	**84.74**	**52.03**	**73.85**	**73.34**	42.05	94.77	47.65	31.65	40.74	40.74
LSTM→VAE-DGP	**36.58**	**58.64**	**58.62**	**3.89**	**15.99**	**14.43**	-1.32	-1.97	-2.23	-17.16	-48.14	-51.48
Stacked LSTM →VAE-DGP	**32.58**	**55.03**	**55.01**	**3.74**	**3.07**	**1.39**	-5.95	-7.63	-7.74	-19.63	-53.84	-54.11
SPGP→VAE-DGP	**52.60**	**65.10**	**65.18**	**43.23**	**63.37**	**62.66**	-32.99	-31.35	-31.50	-74.35	-162.29	-159.06

Moreover, the integrated comparison are described in Tables 3 and 4. In the tables, the symbol "→" means the comparison of two models, for example, "BPNN→SPGP" represents the improved percentage of SPGP with respect to BPNN, when the result is negative, the SPGP reduces the prediction accuracy, and in contrast the model performance is improved by SPGP. According to Table 3 and 4, the VAE-DGP maintains its superiority in Case 1, for instance, the VAE-DGP can reduce the RMSE, MAPE and MAE by 59.48%, 69.87% and 71.31% in terms of "SPGP→VAE-DGP" in Table 3 when TP=400h. However, as the input number of parameters increases, the SPGP can achieve better prediction. Considering "VAE-DGP→ SPGP" in Table 4, the improved percentages are obtained as 42.64%, 61.87% and 61.39% accordingly when TP=800h.

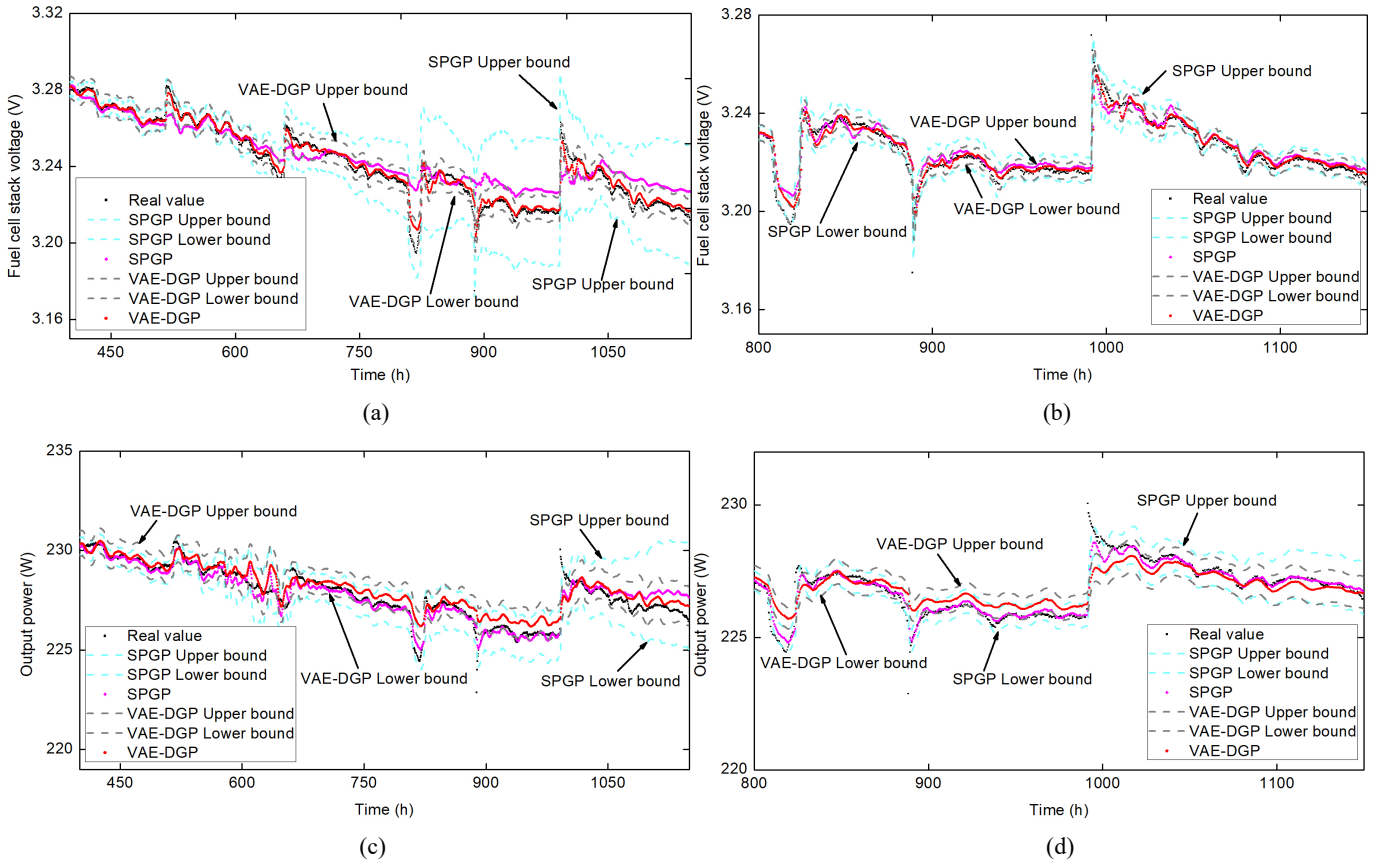


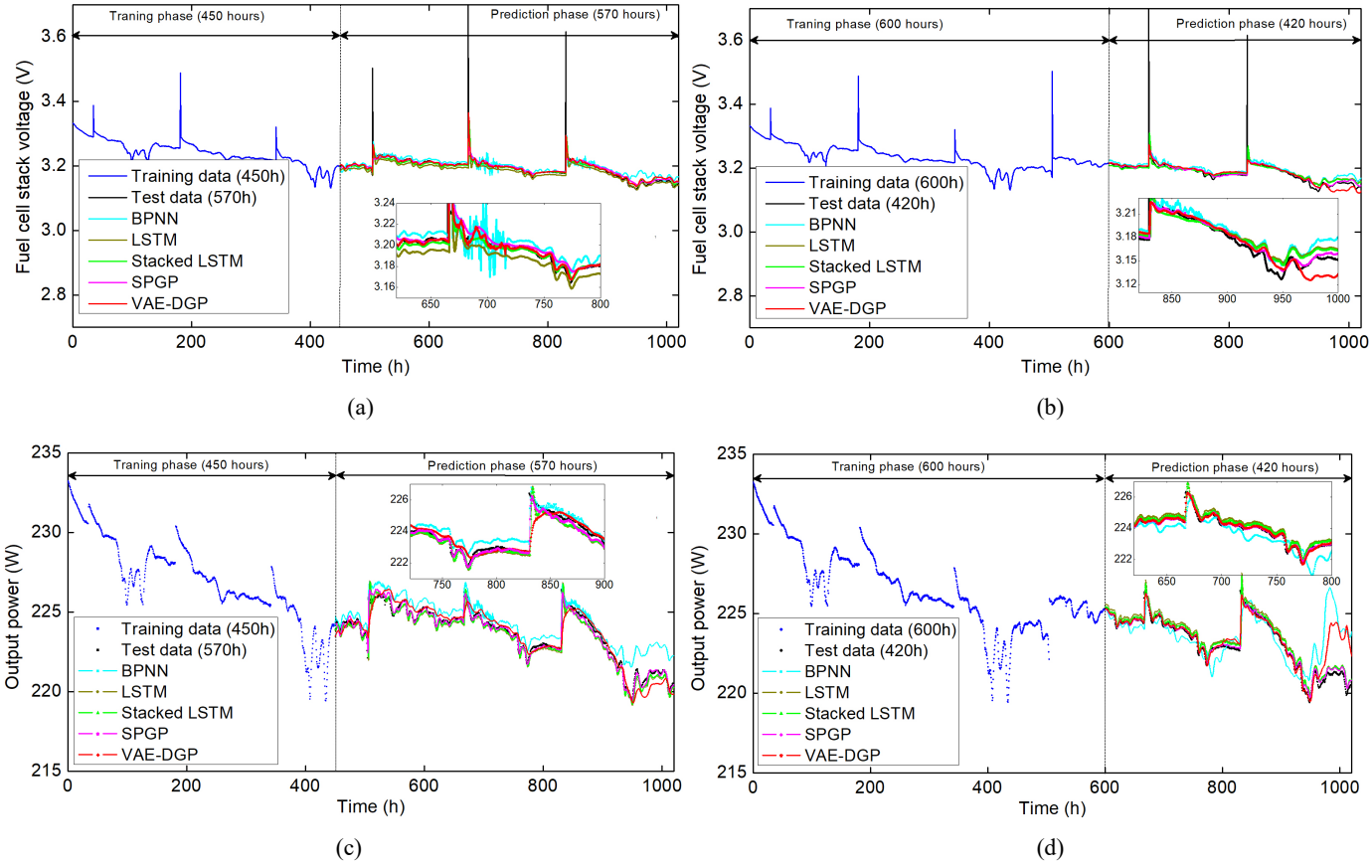
Fig. 5. (a) Stack voltage estimation of FC1 with 95% confidence interval in Case 1 when TP=400h. (b) Stack voltage estimation of FC1 with 95%

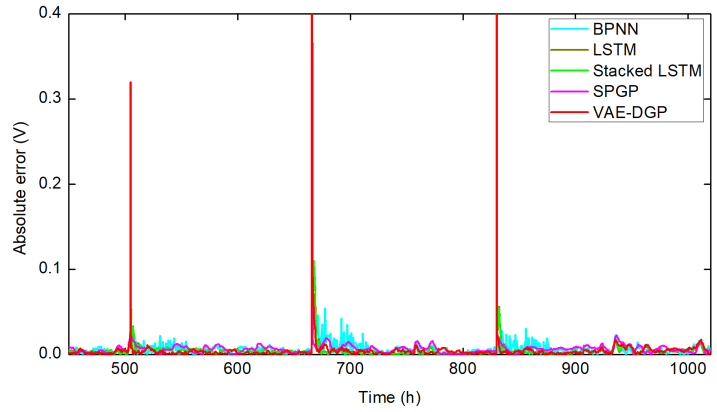
confidence interval in Case 1 when TP=800h. (c) Output power estimation of FC1 with 95% confidence interval in Case 2 when TP=400h. (d) Output power estimation of FC1 with 95% confidence interval in Case 2 when TP=800h.

Fig. 5 depicts some of the estimation of stack voltage and output power with 95% confidence interval for FC1 based on SPGP and VAE-DGP in Case1 and Case 2. It is obvious that the mean values of stack voltage in Fig. 5 (a) and (b) are within the estimated interval with the VAE-DGP model, the narrow interval width leads to reliable prognostics. However, even though the SPGP model can roughly describe the degradation trend of stack voltage, the prediction accuracy is rather poor. Therefore, a conclusion is drawn that the prediction of VAE-DGP has low uncertainty in the stack voltage prediction. As we can see from Fig. 5 (c) and (d), the estimated confidence intervals of output power almost include the actual voltage points for SPGP, in terms of the VAE-DGP, when the stack voltage curve drawn in black line changes rapidly, especially at the 817h, 890h, 940h, 990h, etc., the estimated bounds cannot contain the actual voltage curve, so it may be difficult for VAE-DGP model to provide reliable confidence interval for output power estimation.

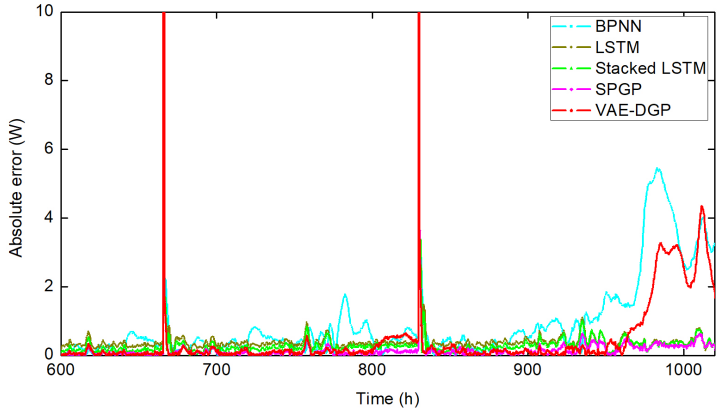
4.3. Prognostic under dynamic operation mode

The degradation of PEMFC under dynamic current load is more complex and rapid, the oscillations in HIs are more serious in the degradation curves because the stack current is no longer constant. In order to discuss the prediction accuracy and applicability of DP-based models further, in the dynamic operation task, the TPs are set as 450h and 600h during 1020h





(e)



(f)

Fig. 6. Degradation prediction of FC2. (a) Stack voltage prediction when TP=450h with single input. (b) Stack voltage prediction when TP=600h with multiple-input. (c) Output power prediction when TP=450h with multiple-input. (d) Output power prediction when TP=600h with single input. (e) Absolute error comparison of (a). (f) Absolute error comparison of (d).

duration, respectively. Accordingly, the prediction phases are 570h and 420h. The prediction results for HIs under different TPs with single input and multiple inputs are demonstrated in Fig. 6. It's clear that in Fig. 6 (a), the VAE-DGP can capture the spikes in the stack voltage degradation curve and well present the nonlinearity. However, there are estimated deviations in BPNN, Stacked LSTM and LSTM, and the performance curve based on SPGP is close to that of VAE-DGP. In Fig.6 (b) the SPGP shows its superiority in the case of more training data, which outperforms the other prediction models in terms of RMSE (0.02732), MAPE (0.00174) and MAE (0.00574) in our study. More obviously, when the HI is output power in Fig.6 (c) and (d), the SPGP can effectively track the variations of output power drop trend and forecast the future aging trend. However, the VAE-DGP shows a relatively poor performance, the criteria of prediction accuracy are 0.52955, 0.00166, 0.37290 and 0.96383, 0.00199, 0.44269, respectively, which are even worse than that of the Staked LSTM and LSTM. From the absolute error curves of the five methods in Fig.6 (e) and (f), we can see that under the dynamic load situation, the absolute error margins of VAE-DGP and SPGP are small in Fig.6 (e). In Fig.6 (f) before 800h, the predicted absolute error of output power for VAE-DGP is smaller than that of BPNN, LSTM and Stacked LSTM, after 800h the absolute error of

VAE-DGP becomes larger, but in the whole prediction phase, the absolute error of SPGP method is minimum.

Table 5 The comparative prediction results of FC2 under two cases when HI is stack voltage.

Methods	Case 1: Single input U_{out}						Case 2: Multiple inputs U_{out}, T_{inwat}, T_{outair}					
	Training phase (450h)			Training phase (600h)			Training phase (450h)			Training phase (600h)		
	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE
BPNN	0.02835	0.00369	0.01192	0.02820	0.00225	0.00737	0.02655	0.00245	0.00799	0.02925	0.00369	0.01191
LSTM	0.02777	0.00310	0.01009	0.02824	0.00246	0.00763	0.02628	0.00170	0.00561	0.02839	0.00243	0.00831
Stacked LSTM	0.02620	0.00171	0.00563	0.02807	0.00245	0.00765	0.02620	0.00163	0.00538	0.02826	0.00237	0.00775
SPGP	0.02613	0.00169	0.00557	**0.02750**	**0.00172**	**0.00568**	**0.02556**	**0.00147**	**0.00486**	**0.02732**	**0.00174**	**0.00574**
VAE-DGP	**0.02549**	**0.00160**	**0.00526**	0.02780	0.00186	0.00611	0.02567	0.00221	0.00730	0.02804	0.00197	0.00807

Table 6 The comparative prediction results of FC2 under two cases when HI is output power.

Methods	Case 1: Single input U_{out}						Case 2: Multiple inputs U_{out}, I_{out}, T_{inwat}, T_{outair}					
	Training phase (450h)			Training phase (600h)			Training phase (450h)			Training phase (600h)		
	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE
BPNN	2.03854	0.00515	1.14107	1.51381	0.00421	0.93557	0.91454	0.00325	0.72525	0.87374	0.00261	0.58035
LSTM	0.60025	0.00218	0.48778	0.43473	0.00168	0.37618	0.41727	0.00134	0.29969	0.38475	0.00091	0.20333
Stacked LSTM	0.45763	0.00142	0.31838	0.36766	0.00127	0.28442	0.37264	0.00108	0.24259	0.37631	0.00089	0.20076
SPGP	0.59340	0.00180	0.40193	**0.24213**	**0.00063**	**0.14196**	**0.23626**	**0.00068**	**0.15271**	**0.21558**	**0.00065**	**0.15182**
VAE-DGP	**0.33317**	**0.00080**	**0.18024**	0.96383	0.00199	0.44269	0.52955	0.00166	0.37290	0.66119	0.00203	0.45300

Moreover, comprehensive results are shown in Table 5 and 6, which include the results of other scenarios, for example, the TP is 450h with multiple inputs in Table 5. It is should be noted that under the dynamic operation situation, only in a small data regime in Case 1 with TP=450h, the VAE-DGP can outperform other methods in both tables, fortunately the SPGP can make up for the shortcoming when in large data regime. This finding suggests that different HIs and training phases should match different prediction models due to the difference in degradation characteristics.

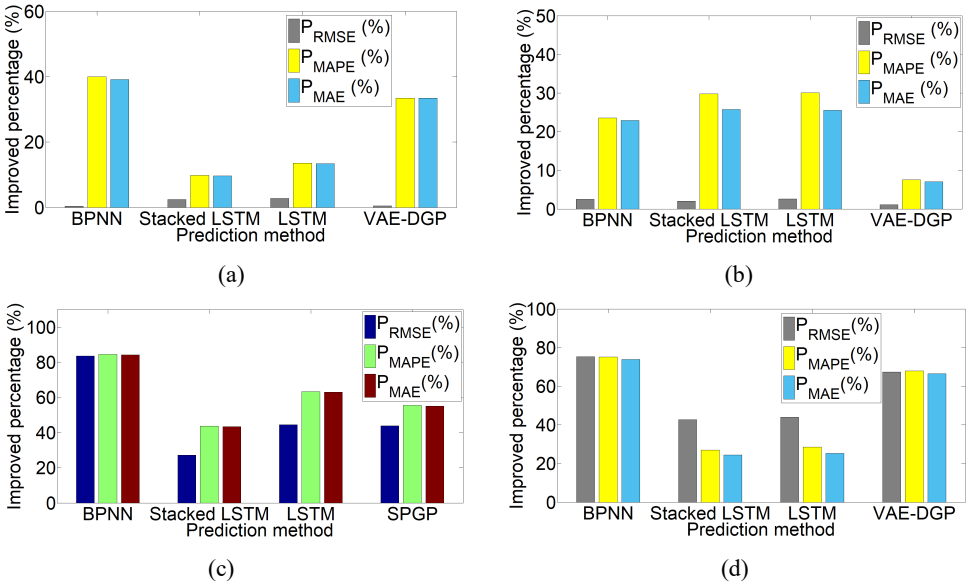


Fig. 7. Accuracy improvement of DP-based models in FC2. (a) TP=450h in Case 2 for stack voltage prediction. (b) TP=600h in Case 1 for stack voltage prediction. (c) TP=450h in Case 1 for output power prediction. (d) TP=600h in Case 2 for output power prediction.

The further study on the percentage improvement of proposed models are illustrated in Fig.7 with different scenarios from

Fig. 6 (a)-(d) We can see that in Fig.7 (a), (b) and (d), the P_{RMSE}, P_{MAPE} and P_{MAE} are positive, which means the SPGP can improve the model accuracy compared to BPNN, Stacked LSTM, LSTM and VAE-DGP. However, due to many spikes in the stack voltage, the increase of P_{RMSE} is less than 5% in Fig. 7 (a) and (b), we conclude that the performance of output power prediction is much better than that of voltage aging forecasting in FC2. And in Fig. 7 (c), the VAE-DGP outperforms the other methods, the model accuracy is improved greatly.

Table 7 The improved percentage of FC2 under two cases when HI is stack voltage.

Comparison models	Case 1: Single input U_{out}						Case 2: Multiple inputs U_{out}, T_{inwat}, T_{outair}					
	Training phase (450h)			Training phase (600h)			Training phase (450h)			Training phase (600h)		
	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}
BPNN→SPGP	7.83	54.20	53.27	**2.48**	**23.55**	**22.93**	**0.37**	**40.00**	**39.17**	**6.59**	**52.84**	**51.80**
LSTM→SPGP	5.91	45.48	44.79	**2.62**	**30.08**	**25.55**	**2.73**	**13.52**	**13.36**	**3.77**	**28.39**	**30.92**
Stacked LSTM →SPGP	0.27	1.17	1.06	**2.03**	**29.79**	**25.75**	**2.44**	**9.81**	**9.66**	**3.32**	**26.58**	**25.93**
VAE-DGP→SPGP	-2.51	-5.62	-5.89	**1.07**	**7.52**	**7.03**	**0.43**	**33.48**	**33.42**	**2.57**	**11.67**	**28.87**
BPNN→VAE-DGP	**10.08**	**56.63**	**55.87**	1.42	17.33	17.09	3.31	9.79	8.63	4.14	46.61	32.24
LSTM→VAE-DGP	**8.21**	**48.38**	**47.86**	1.56	24.39	19.92	2.32	-30.00	-30.12	1.23	18.93	2.89
Stacked LSTM →VAE-DGP	**2.71**	**6.43**	**6.57**	0.96	24.08	20.13	2.02	-35.58	-35.68	0.78	16.87	-4.12
SPGP→VAE-DGP	**2.45**	**5.33**	**5.56**	-1.09	-8.13	-7.57	-0.43	-50.34	-50.20	-2.63	-13.21	-40.59

Table 8 The improved percentage of FC2 under two cases when HI is output power.

Comparison models	Case 1: Single input U_{out}						Case 2: Multiple inputs U_{out}, I_{out}, T_{inwat}, T_{outair}					
	Training phase (450h)			Training phase (600h)			Training phase (450h)			Training phase (600h)		
	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}	P_{RMSE}	P_{MAPE}	P_{MAE}
BPNN→SPGP	70.89	65.04	64.77	**84.00**	**85.03**	**84.82**	**74.16**	**79.07**	**78.94**	**75.32**	**75.09**	**73.83**
LSTM→SPGP	1.14	17.43	17.60	**44.30**	**62.50**	**62.26**	**43.37**	**49.25**	**49.04**	**43.96**	**28.57**	**25.33**
Stacked LSTM →SPGP	-29.66	-26.76	-26.24	**34.14**	**50.39**	**50.08**	**36.59**	**37.03**	**37.05**	**42.71**	**26.96**	**24.37**
VAE-DGP→SPGP	-78.11	-125.00	-122.99	**74.87**	**68.34**	**67.93**	**55.38**	**59.03**	**59.04**	**67.39**	**67.98**	**66.48**
BPNN→VAE-DGP	**83.65**	**84.46**	**84.20**	36.33	52.73	52.68	42.09	48.92	48.58	24.32	22.22	21.94
LSTM→VAE-DGP	**44.49**	**63.30**	**63.04**	-121.70	-18.45	-17.68	-26.90	-23.88	-24.42	-71.84	-123.07	-122.79
Stacked LSTM →VAE-DGP	**27.19**	**43.66**	**43.38**	-162.15	-56.69	-55.64	-42.10	-53.88	-53.71	-75.70	-128.08	-125.64
SPGP→VAE-DGP	**43.85**	**55.55**	**55.15**	-298.06	-215.87	-211.84	-124.13	-144.11	-144.18	-206.70	-212.30	-198.37

Specifically, the quantified performance improvement ratio with VAE-DGP and SPGP are presented in Table 7 and 8. In the eight tests, the SPGP obtains significantly higher prediction accuracy, except when the TP is 450h with single input in both tables. For example, in Table 7 when TP is 450h in Case 1, compared with Stacked LSTM, the VAE-DGP reduces the RMSE by 2.71%; reduces MAPE by 6.43%; and reduces MAE by 6.57%. Similarly, in Table 8 when TP is 600h in Case 2, compared with Stacked LSTM, the SPGP reduces the RMSE by 42.71%; reduces MAPE by 26.96%; and reduces MAE by 24.37%.

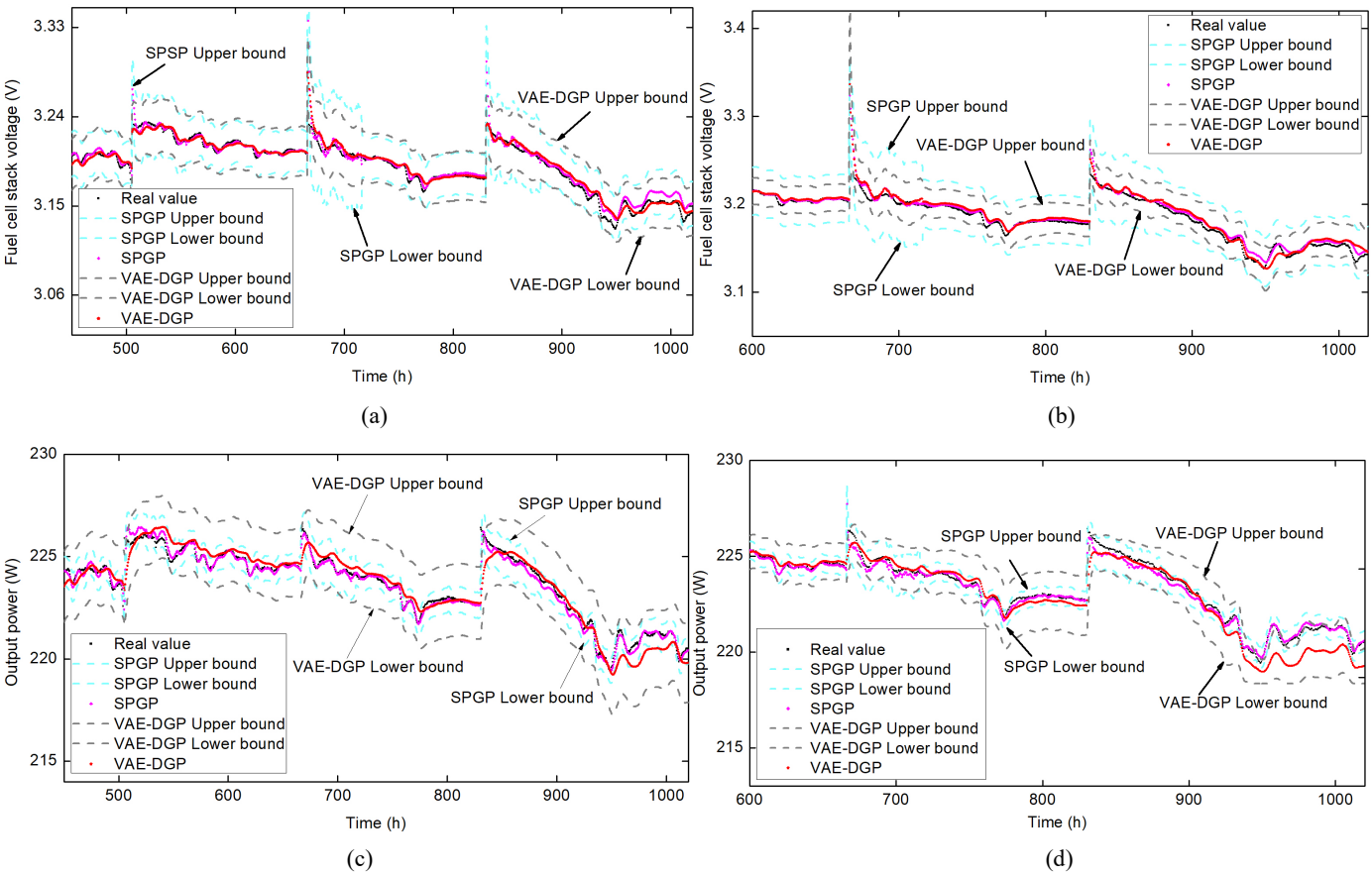


Fig. 8. (a) Stack voltage estimation of FC2 with 95% confidence interval in Case 1 when TP=450h. (b) Stack voltage estimation of FC2 with 95% confidence interval in Case 1 when TP=600h. (c) Output power estimation of FC2 with 95% confidence interval in Case 2 when TP=450h. (b) Output power estimation of FC2 with 95% confidence interval in Case 2 when TP=600h.

As shown in Fig. 8, the uncertainties of prognostic results of FC2 are described by the 95% confidence interval on the basis of VAE-DGP and SPGP in Case 1 and Case 2 using 450h and 600h learning phase, respectively. According to Fig.8 (a) and (b), the actual stack voltage points are scattered between the estimated upper and lower bounds of SPGP and VAE-DGP, however, the confidence interval of VAE-DGP is narrow, a narrower confidence interval may lead to more accurate and reliable prediction results. Besides, the SPGP model is modified to forecast the aging trend based on the historical data, due to the latent states updating, the irreversible phenomenon occurs at the 667h and 830h in Fig.8 (a) and (b), thus the uncertainties increase, the inferred bounds can still contain the actual voltage curve. As for the output power uncertainties estimation in Fig.8 (c) and (d), the actual power points are within the estimated bounds based on the SPGP. However, the bounds of VAE-DGP are much wider, there are still some actual power points during 965h-993h below the lower bound. Therefore, the SPGP is more suitable for the output power prognostic.

5. Conclusion

In this paper, data-driven methods based on VAE-DGP and SPGP are proposed to forecast the degradation trajectories of

PEMFC. The stack voltage and output power are considered as the degradation indexes. Due to the probabilistic property, the DP-based models in our study can avoid overfitting and generalization capability is strong, moreover, through confidence interval, the prediction reliability is improved. On the basis of long-term fuel cell aging experimental data, the model prediction is validated under sixteen different test scenarios of FC1 and FC2. Meanwhile, different existing methods including BPNN, Stacked LSTM and LSTM are compared with the proposed models. The following conclusions are made:

(1) When the HIs are in a small fluctuation range in the static operation task of FC1, the VAE-DGP can effectively achieve higher prediction accuracy in most scenarios. Regarding more complicated condition in FC2, the SPGP with moving pseudo inputs is flexible to model the non-stationary process in HIs.

(2) The stack voltage aging trend prediction accuracy can be improved with VAE-DGP method, the confidence interval of stack voltage estimation is much narrower; the SPGP has advantage in output power prediction, the output power uncertainties estimation can contain almost all the actual power points compared to VAE-DGP, which can provide reliable predictions.

(3) The accuracy comparison and absolute error analysis demonstrate the effectiveness of the proposed models, it is clear that a relatively large amount of training data should be learned by SPGP to achieve a better performance, on the contrary, the VAE-DGP is very suitable for the small sample prediction, which means when the measurement data for model training is less in some situations, the VAE-SGP can provide satisfied aging prediction results.

Acknowledgment

This work was supported by the Sichuan Science and Technology Program (Sichuan Distinguished Young Scholars) under Grant 2020JDJQ0037. The authors would like to thank the FCLAB Federation for the data used in this research.

References

[1] M. Bressel, M. Hilairet, D. Hissel, and B. Ould Bouamama, Remaining useful life prediction and uncertainty quantification of proton exchange membrane fuel cell under variable load, IEEE Trans. Ind. Electron., 63 (2016) 2569-2577.

[2] H. Deng, Q. Li, W. Chen, and G. Zhang, High-order sliding mode observer-based OER control for PEM fuel cell air-feed system, IEEE Trans. Energy Convers., 33 (2018) 232-244.

[3] Z. Li, Z. Zheng and R. Outbib, Adaptive prognostic of fuel cells by implementing ensemble echo state networks in time-varying model space, IEEE Trans. Ind. Electron., 67 (2020) 379-389.

[4] H. Deng, W. Chen, D. Cao, J. Chen, and W. Hu, Uncertainty analysis and robust control of fuel delivery systems considering nitrogen crossover phenomenon, Int. J. Hydrogen Energy, 45 (2020) 32367-32387.

[5] Y. Yan, Q. Li, W. Chen, W. Huang, and J. Liu, Hierarchical management control based on equivalent fitting circle and equivalent energy consumption method for multiple fuel cells hybrid power system, IEEE Trans. Ind. Electron., 67 (2020) 2786-2797.

[6] H. Deng, Q. Li, Y. Cui, Y. Zhu, and W. Chen, Nonlinear controller design based on cascade adaptive sliding mode control for PEM fuel cell air supply systems, Int. J. Hydrogen Energy, 44 (2019) 19375-19369.

[7] H. Liu, J. Chen, D. Hissel, and H. Su, Short-term prognostics of PEM fuel cells: A comparative and improvement study, IEEE Trans. Ind. Electron., 66 (2019) 6077-6086.

[8] T. Sutharssan, D. Montalvao, Y. Chen, W. Wang, C. Pisac, and H. Elemara, A review on prognostics and health monitoring of proton exchange membrane fuel cell, Renew. Sust. Energ. Rev., 75 (2017) 440-450.

[9] M. Jouin, R. Gouriveau, D. Hissel, M. Péra, and N. Zerhouni, Degradations analysis and aging modeling for health assessment and prognostics of PEMFC, Reliab. Eng. Syst. Saf., 148 (2016) 78-95.

[10] D. Zhou, Y. Wu, F. Gao, E. Breaz, A. Ravey, and A. Miraoui, Degradation prediction of PEM fuel cell stack based on multiphysical aging Model with particle filter approach, IEEE Trans. Ind. Appl., 53 (2017) 4041-4052.

[11] M. Bressel, M. Hilairet, D. Hissel, and B. Belkacem Ould, Extended Kalman filter for prognostic of proton exchange membrane fuel cell, Appl. Energy, 164 (2016) 220-227.

[12] H. Li, A. Ravey, A. N' Diaye, and A. Djerdir, Online adaptive equivalent consumption minimization strategy for fuel cell hybrid electric vehicle considering power sources degradation, Energy Convers. Manage., 192 (2019) 133-149.

[13] M. Jouin, R. Gouriveau, D. Hissel, M. Péra, and N. Zerhouni, Joint particle filters prognostics for proton exchange membrane fuel cell power prediction at constant current solicitation, IEEE Trans. Reliab., 65 (2016) 336-349.

[14] Z. Hu, L. Xu, J. Li, M. Ouyang, Z. Song, and H. Huang, A reconstructed fuel cell life-prediction model for a fuel cell hybrid city bus, Energy Convers. Manage., 156 (2018) 723-732.

[15] D. Zhou, A. Al-Durra, K. Zhang, A. Ravey, and F. Gao, A robust prognostic indicator for renewable energy technologies: A novel error correction grey prediction model, IEEE Trans. Ind. Electron., 66 (2019) 9312-9325.

[16] Y. Wu, E. Breaz, F. Gao, D. Paire, and A. Miraoui, Nonlinear performance degradation prediction of proton exchange membrane fuel cells using relevance vector machine, IEEE Trans. Energy Convers., 31 (2016) 1570-1582.

[17] Y. Wu, E. Breaz, F. Gao, and A. Miraoui, A modified relevance vector machine for PEM fuel-cell stack aging prediction, IEEE Trans. Ind. Appl., 52 (2016) 2573-2581.

[18] K. Javed, R. Gouriveau, N. Zerhouni, and D. Hissel, Prognostics of proton exchange membrane fuel cells stack using an ensemble of constraints based connectionist networks, J. Power Sources, 324 (2016) 745-757.

[19] Y. Xie, J. Zou, Z. Li, F. Gao, and C. Peng, A novel deep belief network and extreme learning machine based performance degradation prediction method for proton exchange membrane fuel cell, IEEE Access, 8 (2020) 176661-176675.

[20] J. Liu J, Q. Li, W. Chen, Y. Yan, Y. Qiu, and T. Cao, Remaining useful life prediction of PEMFC based on long short-term memory recurrent neural networks, Int. J. Hydrogen Energy, 44 (2019) 5470–5480.

[21] R. Ma, T. Yang, E. Breaz, Z. Li, P. Briois, and F. Gao, Data-driven proton exchange membrane fuel cell degradation predication through deep learning method, Appl. Energy, 231 (2018) 102-115.

[22] F. K. Wang, X. B. Cheng, and K. C. Hsiao, Stacked long short-term memory model for proton exchange membrane fuel cell systems

degradation, J. Power Sources, 448 (2020) 227591.

[23] F. K. Wang, T. Mamo, and X. B. Cheng, Bi-directional long short-term memory recurrent neural network with attention for stack voltage degradation from proton exchange membrane fuel cells, J. Power Sources, 461 (2020) 228170.

[24] S. Morando, S. Jemei, R. Gouriveau, N. Zerhouni, and D. Hissel, Fuel cells prognostics using echo state network, in Proc. 39th Annu. IEEE IECON, 2013 1632-1637.

[25] S. Morando, S. Jemei, D. Hissel, R. Gouriveau, and N. Zerhouni, Proton exchange membrane fuel cell ageing forecasting algorithm based on echo state network, Int. J. Hydrogen Energy, 42 (2017) 1472-1480.

[26] R. Mezzi, S. Morando, N. Y. Steiner, M. C. Péra, D. Hissel, and L. Larger, Multi-reservoir echo state network for proton exchange membrane fuel cell remaining useful life prediction, in Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc., 2018 1872-1877.

[27] Z. Hua, Z. Zheng, M. C. Péra, F. Gao, Remaining useful life prediction of PEMFC systems based on the multi-input echo state network, Appl. Energy, 265 (2020) 114791.

[28] M. Ibrahim, N. Y. Steiner, S. Jemei, and D. Hissel, Wavelet-based approach for online fuel cell remaining useful lifetime prediction, IEEE Trans. Ind. Electron., 63 (2016) 5057-5068.

[29] D. Zhou, F. Gao, E. Breaz, A. Ravey, and A. Miraoui, Degradation prediction of PEM fuel cell using a moving window based hybrid prognostic approach, Energy, 138 (2017) 1175-1186.

[30] D. Zhou, A. AI-Durra, K. Zhang, A. Ravey, and F. Gao, Online remaining useful lifetime prediction of proton exchange membrane fuel cells using a novel robust methodology, J. Power Sources, 399 (2018) 314-328.

[31] H. Liu, J. Chen, D. Hissel, H. Su, Remaining useful life estimation for proton exchange membrane fuel cells using a hybrid method, Appl. Energy, 237 (2019) 910-919.

[32] R. Ma, Z. Li, E. Breaz, C. Liu, H. Bai, P. Briois, and F. Gao, Data-fusion prognostics of proton exchange membrane fuel cell degradation, IEEE Trans. Ind. Appl., 55 (2019) 4321-4331.

[33] L. Zhu and J. Chen, Prognostics of PEM fuel cells based on Gaussian process state space models, Energy, 149 (2018) 63-73.

[34] Y. Xie, J. Zou, C. Peng, Y. Zhu, and F. Gao, A novel PEM fuel cell remaining useful life prediction method based on singular spectrum analysis and deep Gaussian processes, Int. J. Hydrogen Energy, 45 (2020) 30942–30956.

[35] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning. Cambridge, MA, USA: MIT Press, 2006.

[36] C. L. C. Mattos, Z. Dai, A. Damianou, G. A. Barreto, and N. D. Lawrence, Deep recurrent Gaussian processes for outlier-robust system identification, Journal Process Contr., 60 (2017) 82–94.

[37] E. Snelson and Z. Ghahramani, Sparse Gaussian processes using pseudo-inputs, in Proc. Adv. Neural Inf. Process. Syst., 2005 1257–1264.

[38] 2014 IEEE PHM Data Challenge, FCLAB Research Federation, France, 2014.

[39] H. Salimbeni and M. Deisenroth. Doubly stochastic variational inference for deep Gaussian processes. in *31st conference on Neural Inf. Process. Syst.*, Long Beach, USA, 2017 1-12.