

Multi-physics-resolved digital twin of proton exchange membrane fuel cells with a data-driven surrogate model

Bowen Wang^{a,b}, Guobin Zhang^a, Huizhi Wang^c, Jin Xuan^{b,*}, Kui Jiao^{a,*}

^a State Key Laboratory of Engines, Tianjin University, 135 Yaguan Road, Tianjin 300350, China

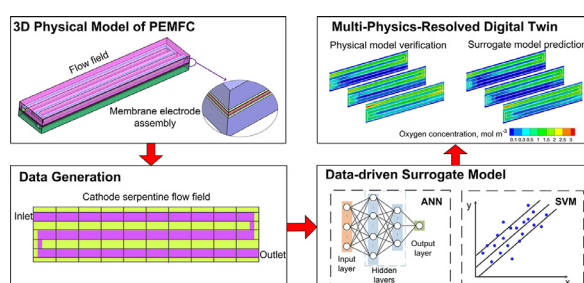
^b Department of Chemical Engineering, Loughborough University, Loughborough, UK

^c Department of Mechanical Engineering, Imperial College London, London, UK

HIGHLIGHTS

- Develops a high-accuracy data-driven surrogate model of PEMFCs.
- Realizes the efficient multi-physics-resolved digital twin of PEMFCs.
- Demonstrates the PEMFC healthy operation envelope and state map.
- Proposes application prospects of combining physical models and data-driven methods.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 20 January 2020

Received in revised form 25 March 2020

Accepted 29 March 2020

Available online 22 April 2020

Keywords:

Proton exchange membrane fuel cell

Digital twin

Data-driven surrogate model

Three-dimensional physical model

Model-based design

ABSTRACT

The development of multi-physics-resolved digital twins of proton exchange membrane fuel cells (PEMFCs) is significant for the advancement of this technology. Here, to solve this scientific issue, a surrogate modelling method that combines a state-of-the-art three-dimensional PEMFC physical model and data-driven model is proposed. The surrogate modelling prediction results demonstrate that the test-set relative root mean square errors (rRMSEs) of the multi-physics fields range from 3.88% to 24.80% and can mirror the multi-physics field distribution characteristics well. In summary, for multi-physics field prediction, the data-driven surrogate model has a comparable accuracy to the comprehensive 3D physical model; however, it considerably reduces the cost of computation and time and achieves the efficient multi-physics-resolved digital-twin. Two model-based designs based on the as-developed digital twin framework, i.e. the PEMFC healthy operation envelope and the PEMFC state map, are demonstrated. This study highlights the potential of combining data-driven approaches and comprehensive physical models to develop the digital twin of complex systems, such as PEMFCs.

1. Introduction

A proton exchange membrane fuel cell (PEMFC) is a highly efficient and zero-emission energy conversion device. The high commercial potential of PEMFCs in vehicle powertrain has been extensively recognised [1–3]. Numerous leading automobile factories, including Toyota, Honda, Hyundai, and Saic, have established remarkable plans for the development programs of PEMFCs and fuel cell vehicles (FCVs) and have launched some milestone FCV products [4–6]. However, the

PEMFC technology at present still needs significant development for reducing the cost, enhancing the performance, and improving the durability to satisfy the requirements of large-scale commercialisation [6–8]. Currently, the concept of digital twins is emerging rapidly with the objectives of constructing a virtual mirror of a physical entity and extensive applications in complex systems of different scales, including cities, factories, and aircrafts. In fact, digital twins have been already used in the industry to increase efficiency and solve problems and been extensively considered as a revolutionising technology [9,10]. Complex physical and chemical processes occur in a PEMFC. Developing a digital twin of a PEMFC, i.e. constructing a precise and virtual copy, to efficiently determine and monitor its multi-physics field state, will be of tremendous significance in cell design and control operation [11–13]. Until now, the

* Corresponding authors.

E-mail addresses: j.xuan@lboro.ac.uk (J. Xuan), kjiao@tju.edu.cn (K. Jiao).

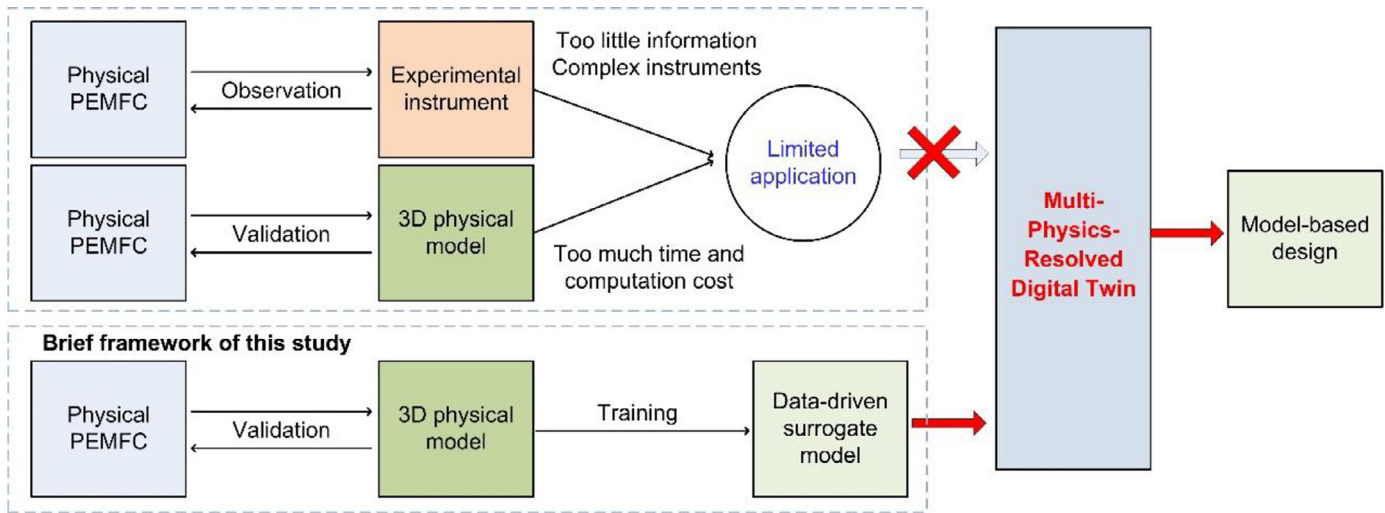


Fig. 1. The Motivation and brief framework of this study.

relevant studies have still not realised an efficient multi-physics digital twin of a PEMFC appropriately.

As proposed by Tao and Qi, data difficulty is the primary challenge in digital twin development [9]. For a PEMFC, the multi-physics fields, including gas reactant, water content, and temperature, directly determine its performance. However, the spatial distributions of these multi-physics fields are difficult to be in-situ measured by typical experimental methods, and their measurement is even impossible to conduct in real-time in an in-situ automotive PEMFC [14–18]. As an emerging research method, machine learning has been used by the fuel cell and other energy systems communities for performance prediction, lifetime prediction, and fault diagnosis [19–25]. However, data collection is still the major challenge, because the observation of multi-physics fields generally needs the usage of specific and expensive instruments. Despite this, the experiments typically provide insufficient information to construct databanks.

Simulation is also an important approach for multi-physics fields prediction, and a few researchers consider that any three-dimensional (3D) physical model can provide a digital twin [8,26–28]. Various studies compared simulation results of the 3D modelling of PEMFCs with experimental results, including the overall characteristics (polarisation curves and ohmic loss) and the spatial distribution of multi-physics fields (current density, temperature, and water saturation), and achieved a good agreement [29–32]. Three-dimensional modelling also highly contributes to promoting the PEMFC technology, which is called as ‘model-based design.’ Moreover, the state-of-the-art 3D models have the ability to well simulate a PEMFC stack, full-scale automotive PEMFC, and PEMFC with complex flow field designs [33–37]. However, conducting highly precise simulation by constructing complex 3D physical model is extremely expensive in terms of the computation and time, which severely restricts its application.

An efficient multi-physics digital twin of a PEMFC is still difficult to realise by the existing experimental observation and traditional simulation tools. In this study, we propose a framework of surrogate modelling which combines a state-of-the-art 3D PEMFC physical model and data-driven models, to achieve an efficient multi-physics-resolved digital twin of a PEMFC. This framework is displayed in Fig. 1. A state-of-the-art well-validated 3D two-phase computational fluid dynamics (CFD) model of a PEMFC is implemented in various scenarios of operating conditions, and the simulation results of various multi-physics fields are used to construct a dataset. Machine learning tools are utilised to train a data-driven surrogate model based on the dataset. The developed surrogate model is expected to have an accuracy comparable to that of a physical model for multi-physics fields prediction. It can efficiently

mirror the multi-physics fields in any operating condition scenario and serve as a multi-physics-resolved digital twin of the physical PEMFC. A detailed PEMFC is selected as a case study to demonstrate the surrogate modelling process, and its schematic is presented in Fig. 2(a). A dataset is generated from the simulation results of the multi-physics fields in 100 different operating condition scenarios. Two machine learning algorithms are employed for the surrogate model development, which are an artificial neural network (ANN)-based model and a support vector machine (SVM)-based model, and their prediction performance is compared. The relative root mean square errors (rRMSE) in the test set for all the multi-physics fields range from 3.88% to 24.80%. The proposed surrogate model has an acceptable accuracy for multi-physics field prediction to achieve a multi-physics-resolved digital twin of a PEMFC. Furthermore, some model-based designs based on the digital twin are presented and demonstrated.

2. Methods

As displayed in Fig. 2(a), the simulated PEMFC consists of an anode parallel flow field, a cathode serpentine flow field, and a membrane electrode assembly (MEA) (including a membrane, gas diffusion layers (GDLs), micro-porous layers (MPLs), and catalyst layers (CLs)). The anode and cathode flows are in opposite directions, i.e. there is counter-flow arrangement. The other detailed dimensions, operating conditions, and material and kinetics properties of the simulated PEMFC are listed in Table 1.

2.1. 3D physical model of PEMFC

A comprehensive 3D PEMFC physical model which couples the heat and mass transport processes and the electrochemical kinetics is employed in this study [36]. The main governing equations solved in the model are as follows:

Mass of the gas mixture:

$$\frac{\partial}{\partial t} (\epsilon(1-s)\rho_g) + \nabla \cdot (\rho_g \mathbf{u}_g) = S_m. \quad (1)$$

Momentum of the gas mixture:

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{\rho_g \mathbf{u}_g}{\epsilon(1-s)} \right) + \nabla \cdot \left(\frac{\rho_g \mathbf{u}_g \mathbf{u}_g}{\epsilon^2(1-s)^2} \right) = & -\nabla P_g + \mu_g \nabla \cdot \left(\nabla \left(\frac{\mathbf{u}_g}{\epsilon(1-s)} \right) + \nabla \left(\frac{\mathbf{u}_g^T}{\epsilon(1-s)} \right) \right) \\ & - \frac{2}{3} \mu_g \nabla \cdot \left(\nabla \cdot \left(\frac{\mathbf{u}_g}{\epsilon(1-s)} \right) \right) + S_u \end{aligned} \quad (2)$$

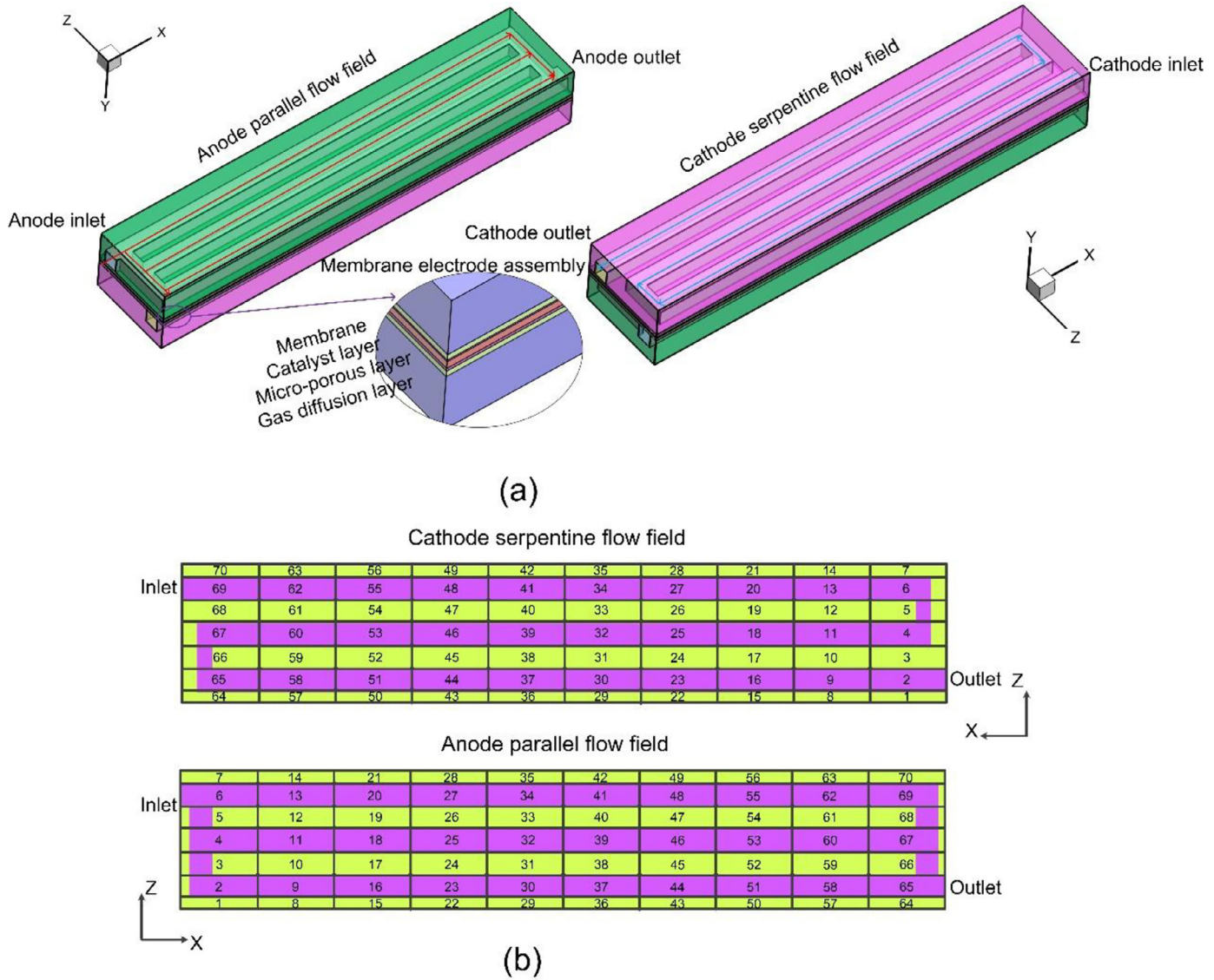


Fig. 2. (a) Schematics of the simulated PEMFC. (b) Schematics of the cathode serpentine flow field and anode parallel flow field, and node partition for data generation of multi-physics fields. The activation planes (X-Z planes) of the PEMFC are divided into 70 nodes according to under- channel or under- shoulder, and upstream and downstream of flow.

Gas species:

$$\frac{\partial}{\partial t} (\epsilon(1-s)\rho_g Y_i) + \nabla \cdot (\rho_g u_g Y_i) = \nabla \cdot (\rho_g D_i^{\text{eff}} \nabla Y_i) + S_i. \quad (3)$$

Liquid pressure in a porous electrode:

$$\frac{\partial}{\partial t} (\epsilon s \rho_l) = \nabla \cdot \left(\rho_l \frac{K k_l}{\mu_l} \nabla P_l \right) + S_l. \quad (4)$$

Membrane water content:

$$\frac{\rho_{\text{mem}}}{EW} \frac{\partial}{\partial t} (\omega \lambda) + \nabla \cdot \left(n_d \frac{J_{\text{ion}}}{F} \right) = \frac{\rho_{\text{mem}}}{EW} \nabla \cdot (D_{\text{mw}}^{\text{eff}} \nabla \lambda) + S_{\text{mw}}. \quad (5)$$

Electronic potential:

$$0 = \nabla \cdot (\kappa_e^{\text{eff}} \nabla \varphi_e) + S_e. \quad (6)$$

Ionic potential:

$$0 = \nabla \cdot (\kappa_{\text{ion}}^{\text{eff}} \nabla \varphi_{\text{ion}}) + S_{\text{ion}}. \quad (7)$$

Energy:

$$\begin{aligned} & \frac{\partial}{\partial t} (\epsilon s \rho_l C_{p,l} T + \epsilon(1-s) \rho_g C_{p,g} T) + \nabla \cdot (\epsilon s \rho_l C_{p,l} u_l T + \epsilon(1-s) \rho_g C_{p,g} u_g T) \\ & = \nabla \cdot (k^{\text{eff}} \nabla T) + S_T. \end{aligned} \quad (8)$$

The liquid saturation in a porous electrode can be calculated using the solved liquid pressure. The agglomerate model of the CLs is embedded, which can better simulate the oxygen transport in a porous electrode than a homogeneous model. A constant-voltage mode is implemented in the simulation by defining the electronic potentials at the external surfaces of the anode and cathode flow field plates as zero and as the difference of the reversible potential and the output voltage, respectively. The gas supply rates at the flow field inlets are calculated using the current density and the stoichiometry ratio, and the back pressure is defined at the flow field outlets. The thermal boundary conditions are defined as a constant temperature at the external surfaces of the PEMFC. Refer to our previous study for more details on the model validation, model assumptions, transport parameters, and electrochemical kinetics [36].

The mesh comprises 331,800 structured volume grid cells, and there are 4200 surface grid cells in the activation plane. The numerical solution of the physical model is obtained by a commercial CFD software, Fluent, coupled with in-house user-defined functions (UDFs). The simulation results of the present physical model, including the polarisation curves, ohmic loss, and current density spatial distributions, for vari-

Table 1
Dimensions, operating conditions, material and kinetics parameters of the simulated PEMFC.

Parameter	Value	Unit
Length and width of the PEMFC	50, 6	mm
Width and depth of channels	1, 1	mm
Thickness of GDLs, MPLs, the membrane	200, 20, 25.4	μm
Temperature	353	K
Porosity of GDLs, MPLs	0.6, 0.5	
Contact angle of GDLs, MPLs, CLs	120, 120, 95	$^{\circ}$
Intrinsic permeability of GDLs, MPLs, CLs, the membrane	2.0e^{-12} , 1.0e^{-12} , 1.0e^{-13} , 2.0e^{-20}	m^2
Equivalent weight of the membrane	1.1	kg mol^{-1}
Thickness of CLs	Anode: 6 Cathode: 10	μm
Pt loading in CLs	Anode: 0.1 Cathode: 0.3	mg cm^{-2}
Pt/Carbon ratio in CLs	Anode: 0.2 Cathode: 0.4	
Ionomer/Carbon ratio in CLs	Anode: 0.7 Cathode: 0.9	
Electrical conductivity of flow field plates, GDLs, MPLs, CLs	20,000, 8000, 5000, 5000	S m^{-1}
Thermal conductivity of flow field plates, MPLs, CLs, the membrane	20, 1, 1, 0.95	$\text{W m}^{-1} \text{K}^{-1}$
Thermal conductivity of GDLs	In-plane: 21 Through-plane: 1.7	$\text{W m}^{-1} \text{K}^{-1}$
Transfer coefficient	0.5	
Reference gas reactant concentration	Anode: 56.4 Cathode: 3.39	mol m^{-3}
Reference exchange current density	Anode: 3.0 Cathode: 3.0e^{-4}	A m^{-2}

Table 2

Six operating condition parameters and their common ranges. The values of the six parameters will only be the specific values as below.

Operating conditions parameters	Ranges	Specific values
Output voltage, V	0.35 – 0.75	0.35, 0.4, ..., 0.7, 0.75
Back pressure, atm	1.0 – 2.2	1.0, 1.2, ..., 2.0, 2.2
Anode stoichiometry ratio	1.2 – 1.6	1.2, 1.3, 1.4, 1.5, 1.6
Cathode stoichiometry ratio	1.6 – 2.4	1.6, 1.8, 2.0, 2.2, 2.4
Anode relative humidity	0.5 – 1.0	0.5, 0.6, ..., 0.9, 1.0
Cathode relative humidity	0.5 – 1.0	0.5, 0.6, ..., 0.9, 1.0

ous scenarios of operating conditions and cell parameters were compared with the experimental results obtained in our previous work, and a good agreement was achieved [36,38–40]. The computations for the data generation are performed on a workstation (CPU: Intel (R) Xeon (R) E5-2650 v4 2.20 GHz 24 threads) by a four-process parallel computing. The computing time of each case is approximately 8–12 h.

2.2. Data generation

A typical dataset includes a series of input and output variables. The input variables are the global operating parameters: the voltage, back pressure (the same in the anode and the cathode), anode stoichiometry ratio, cathode stoichiometry ratio, anode relative humidity, and cathode relative humidity. Their ranges and specific values are listed in Table 2. The corresponding output variables are the distributions of the molar concentrations of the reacting gases (hydrogen at the anode and oxygen at the cathode), water content, and temperature, which are the most important physical quantities of the PEMFC. Here, relative humidity of the anode and the liquid saturation of the cathode are used to represent the water content [11]. Datasets are generated on six representative X–Z planes (activation planes) inside the PEMFC, i.e. the interfaces between the anode GDL and the anode MPL (gdl_mpl_a), the anode MPL and the anode CL (mpl_cl_a), the membrane and the anode CL (mem_cl_a), the cathode GDL and the cathode MPL (gdl_mpl_c), the cathode MPL and the cathode CL (mpl_cl_c), and the membrane and the cathode CL (mem_cl_c). Each plane is divided into 70 nodes (7×10) based on the shape of the flow fields, as displayed in Fig. 2(b). The data points are

generated on each local node, of which the output variables are derived by area-weighted averaging within the grids. Each node additionally contains its coordinates as a part of the dataset. The physical model is implemented in random operating condition scenarios to generate 100 physical model samples. The total dataset is randomly split into two parts: 3/4 (75 physical model samples) as the training set and 1/4 (25 physical model samples) as the test set.

2.3. Data-driven surrogate model development

According to the aforementioned data features of the multi-physics fields under the different operating conditions, here, two extensively used machine learning algorithms are employed aiming at different data topologic structures, i.e. ANN and SVM. Each algorithm has its advantages and typical application. ANN can well fit the highly complex and non-linear relation between inputs and outputs with a large data volume, and SVM can achieve a good compromise between model complexity and generalisation with a limited sample size.

2.3.1. ANN-based model

In this model, an ANN is employed to predict the distributions of the physical quantities within the selected X–Z planes. One ANN model is set for each physical quantity. The node coordinates at the X–Z planes along with the six operating condition parameters are treated as the input. There are totally eight input parameters and one output parameter, 5250 data points in the training set, and 1750 data points in the test set for one ANN. The development of each ANN is implemented by the MATLAB neural network toolbox [41]. The feed-forward multi-layer perceptron is employed, and the Levenberg–Marquardt algorithm is used as the training function. For data pre-processing, both the inputs and outputs are normalised between 0 and 1. In the training processes, the training data are further divided into three parts: training (70%), validation (15%), and testing (15%). The training process stops when the prediction performance, mean squared error (MSE) of validation, reaches the desired goal or the error of validation does not drop anymore with further training, which will increase the risk of overfitting. Searching the optimal topology of an ANN is critical for model development, including the hidden layer number and hidden neuron number,

and this process is generally performed by manual regulation. By a preliminary test, we select a two-hidden-layer ANN, which presents a better fitting performance than a single-hidden-layer ANN. Typically, the number of data points for training should be several times larger than the number of connected weights. Therefore, the ranges of the neuron number of the first and second hidden layers are adopted as 10–30 and 3–15, respectively, and we manually search the optimal topology within these ranges.

2.3.2. SVM based model

A different data structure is developed for the SVM-based model. Here, one SVM model is developed for one physical quantity at a single node independently. Therefore, 70 models are made parallel for the display of the physical quantity distributions within the selected X–Z planes with 70 nodes. The development of the SVM models is implemented by LIBSVM (Chih-Chung Chang and Chih-Jen Lin) in MATLAB [42]. The data pre-processing is the same as above. ϵ -SVR is one function for regression and employed in this study. The radial basis function (RBF) kernel is selected, and a five-fold cross-validation is applied to search the optimal values of the cost parameter and gamma of the RBF kernel.

It should be noted that for both the methods, the test set is fully independent of the model development, and it is only used to assess the generalisation following model development. The training, prediction, and data post-processing are conducted on a typical laptop (Lenovo X1 Yoga, CPU: Intel(R) Core i7-6600 U 2.60 GHz).

3. Results and discussion

3.1. Prediction performance

Prediction results of the SVM-based model are reorganised into series of two-dimensional (2D) physical fields which have the same format as that in the ANN-based model. The root mean square error (RMSE) and rRMSE are introduced to measure the prediction performance across the entire dataset [23].

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad \text{and} \quad (9)$$

$$\text{rRMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}} \times 100\%, \quad (10)$$

where N is the number of data points, y_i denotes the simulated values by the physical model, $\hat{y}_i$ represents the predicted values by the surrogate model, and $\bar{y}$ is the mean of the simulated values by the physical model.

The prediction performance of the two models (i.e. the ANN and SVM-based models) for the multi-physics-resolved digital twin of the PEMFC are quantified and compared using rRMSE over both the test sets (25 samples) and training sets (75 samples), as listed in Table 3. Although there is a lack of similar studies on machine learning prediction for PEMFC multi-physics, we attempt to determine the relative error (rRMSE) based on other machine learning prediction studies. Sun et al. used a convolutional neural network (CNN) to predict the photovoltaic (PV) output based on an image of the sky, and the rRMSE in the test set ranged from 26.0% to 30.2% for two solar systems [23]. In this study, the test set rRMSEs of the multi-physics fields range from 3.88% to 24.80%. Therefore, we can consider that both the models typically achieve the acceptable accuracy for multi-physics prediction, and there is no notable overfitting problem. For the ANN-based model, the rRMSE in the training set is considerably higher than that in the test set for all the physical fields, which is expected for the application of ANNs. An ANN can well fit complex non-linear relations between an input and an output; however, its generalisation is difficult to ensure. For the SVM-based model, the rRMSE in the test set is similar to or even smaller than

that in the training set. We examine the dataset and exclude data leakage, and infer that this scenario is caused by the differences in the distributions of the training and test sets when randomly cutting a dataset of a small data volume (100 operating condition scenarios). As presented in the following discussion, the acceptable prediction result is achieved based on this database. Therefore, we do not further increase the data volume. Furthermore, the prediction performance of the different physical quantities is also very different, even when obtained by a single modelling method. However, the prediction performance of one physical quantity on different planes is relatively similar, specifically when obtained by the SVM-based model. Because the data feature difference of the different physical quantities leads to a significant difference in the regression effects, the data features of one physical quantity on different planes are relatively similar.

Therefore, we can choose optimal modelling methods for different physical quantities. The prediction performance across the entire dataset of the ANN-based model is significantly better than that of the SVM-based model for hydrogen, whereas the prediction performance in the test set of the SVM-based model is generally better than that of the ANN-based model for other physical quantities to an extent. Considering the lower risk of overfitting of the SVM-based model, we choose the ANN-based model for the prediction of the hydrogen distribution and use the SVM-based model to predict the distributions of the relative humidity, oxygen, liquid saturation, and temperature. Following the selection of the optimal modelling method, the prediction errors measured by the RMSE of all the physical fields in the test set are obtained, which are listed in Table 4. The comparison of data points of the multi-physics fields of the physical model verification and the surrogate model prediction in the test set are also intuitively presented in Fig. 3. We can notice that the regression effects of the hydrogen and temperature distributions are better than that of the relative humidity, oxygen, and liquid saturation distributions. Furthermore, note that the values of the oxygen concentration lie in a relatively large range (approximately 0–10 mol m⁻³). Moreover, the surrogate model prediction might result in the predicted values beyond the reasonable range of physical relevance, i.e. slightly negative oxygen concentration. In this case, the values are corrected as zero. In summary, when the surrogate model is developed to predict a new physical quantity field, the optimal modelling method and prediction performance should still be determined and assessed.

3.2. Digital twin presentation

The digital twin multi-physics fields are illustrated and compared with the results from the verified physical models in Fig. 4. Here, the representative case is selected from one of the test set samples. It can be found that the digital twin presented by the surrogate model reproduces the multi-physics fields and resolves the important local phenomena well. Figs. 4(a) and (b) illustrate the hydrogen and relative humidity fields in the anode. The trends of the hydrogen consumption and humidity accumulation along the through-plane and flow directions are captured well. Figs. 4(c) and (f) depict the temperature fields in the anode and the cathode, respectively. Local hot spots (e.g. high-temperature region in the under-middle-channel area) are observed and reflected in the digital twin. Fig. 4(d) illustrates the oxygen fields in the cathode. The difference in the oxygen distributions in the under-channel and under-shoulder areas are notable. The internal oxygen-starving regions (e.g. under-shoulder area at downstream and the near-border area) can be effectively located, which provides insight and guidance for the design and operation of the PEMFC. Fig. 4(e) presents the liquid saturation fields in the cathode. High-liquid saturation is found in the under-shoulder area and the near-border area, indicating a high risk in water management.

In addition to the prediction performance, the larger computation and time cost saving on applying the surrogate model than those with the physical model should be also highlighted here. When the surrogate model is well-trained, the computational time of the surrogate model to predict the multi-physics fields in one operating condition scenario

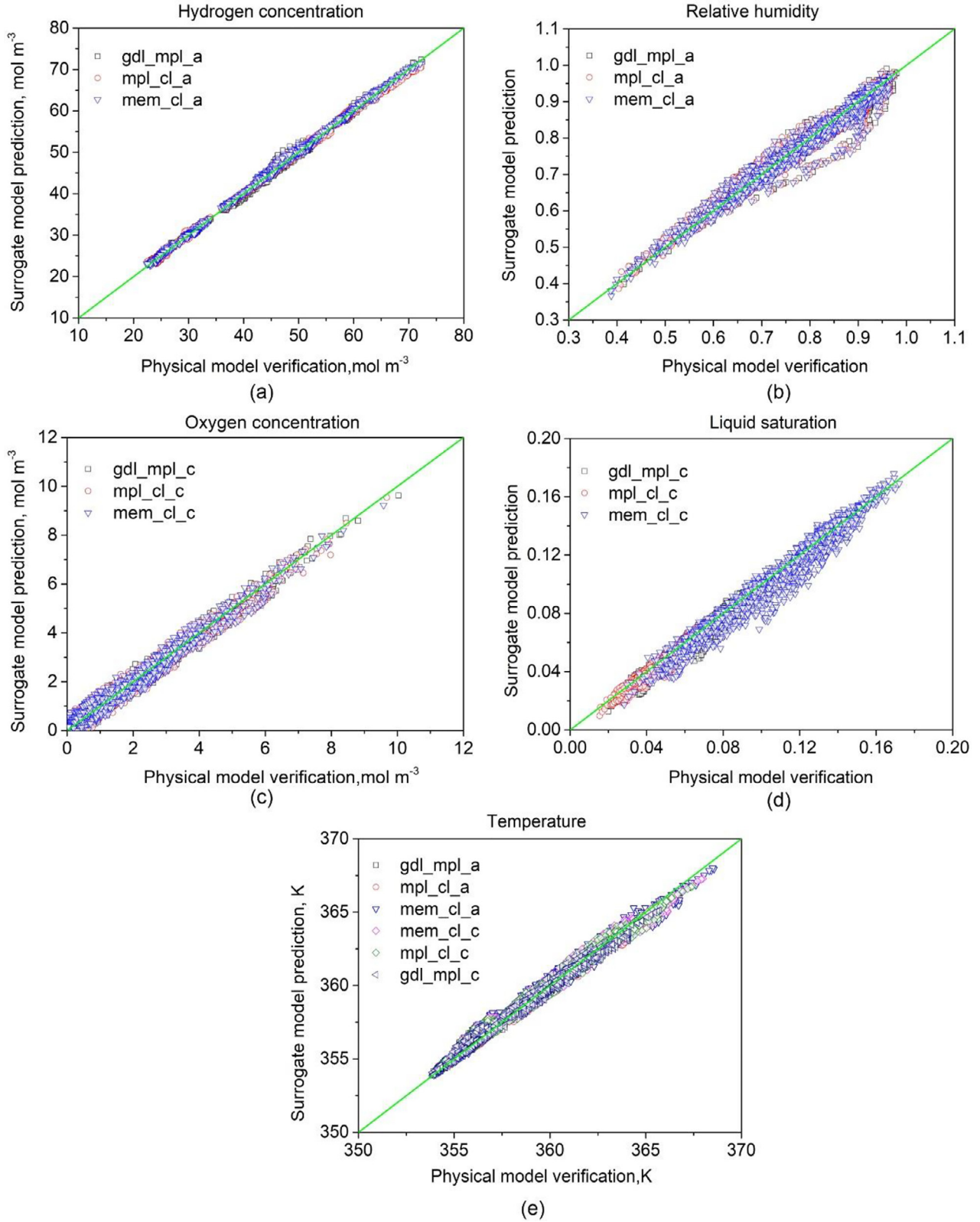


Fig. 3. Prediction performance. Physical model verified and surrogate model predicted values of multi-physics fields in the test set. (a) Hydrogen concentration (b) Relative humidity (c) Oxygen concentration (d) Liquid saturation (e) Temperature. Owing to the fact that the value ranges of one physical quantity at different planes are generally similar, we integrate the results of one physical quantity at different planes into one graph. Each model has 1750 data points in the test set. (a–d) contains results of three 2D physical fields, and there are totally 5250 data points in one graph. (e) contains the results of six 2D physical fields, and there totally 10,500 data points in the graph. Furthermore, note that hydrogen is predicted by the ANN based model and other variables are predicted by the SVM based model.

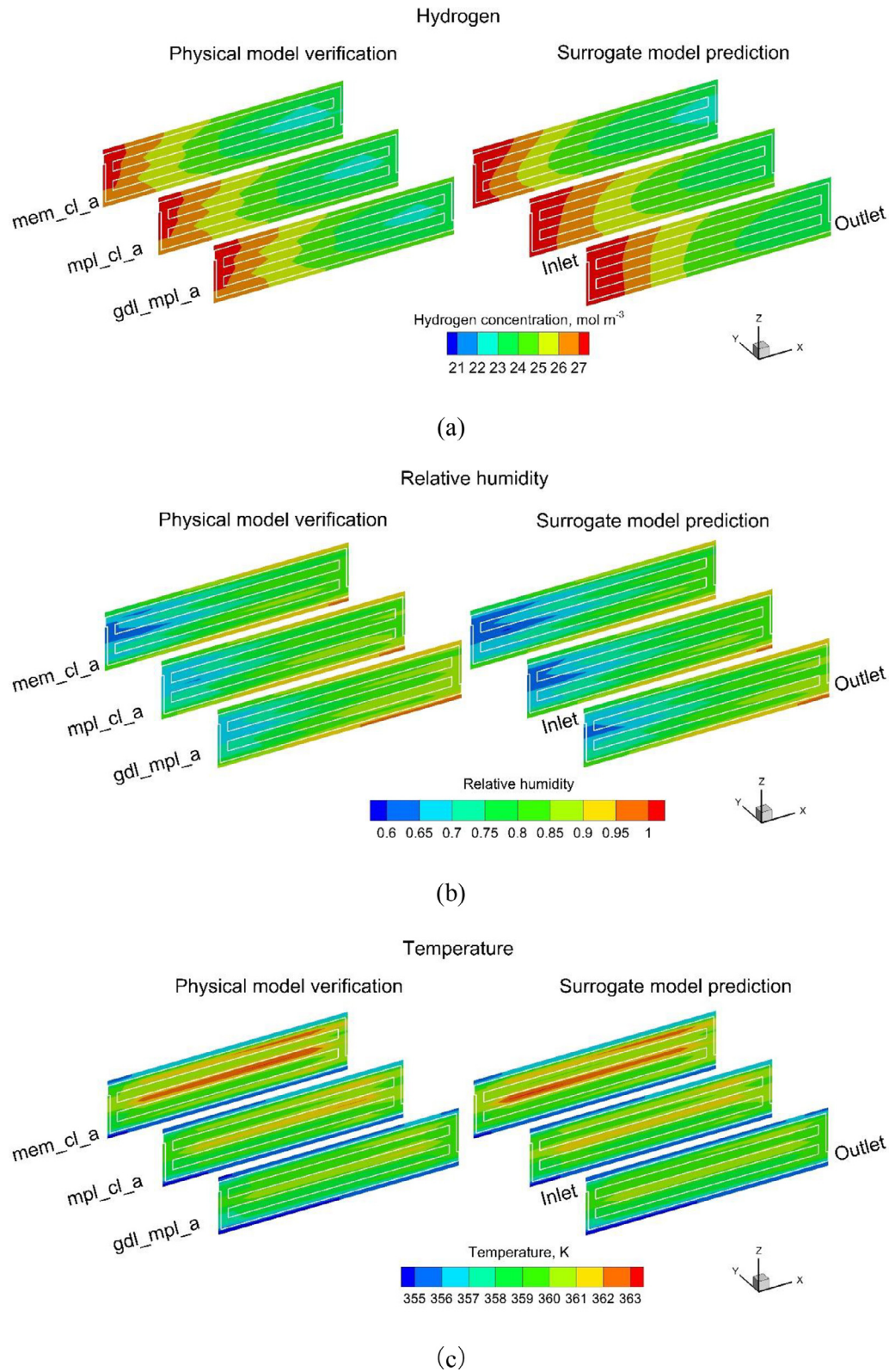
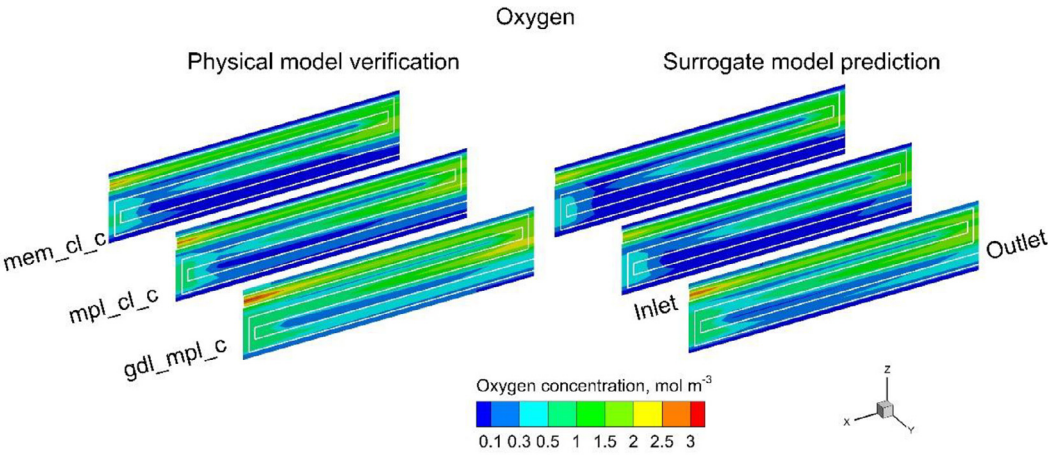
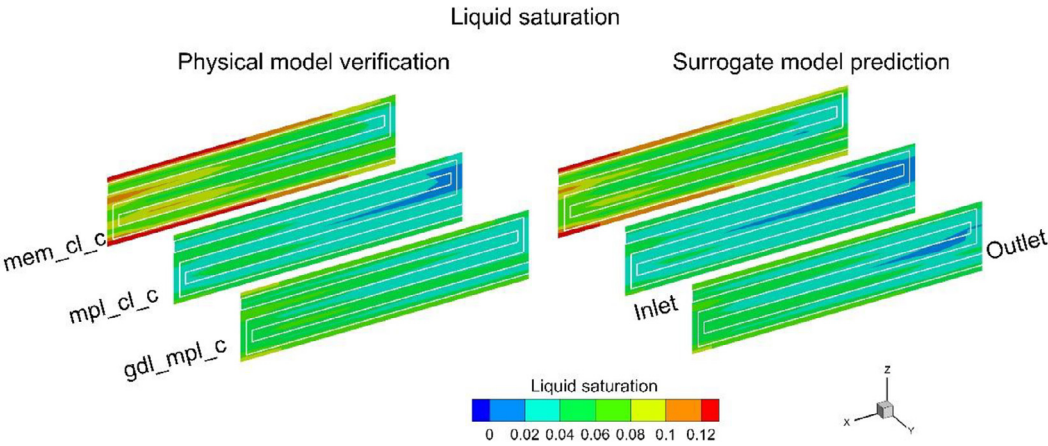


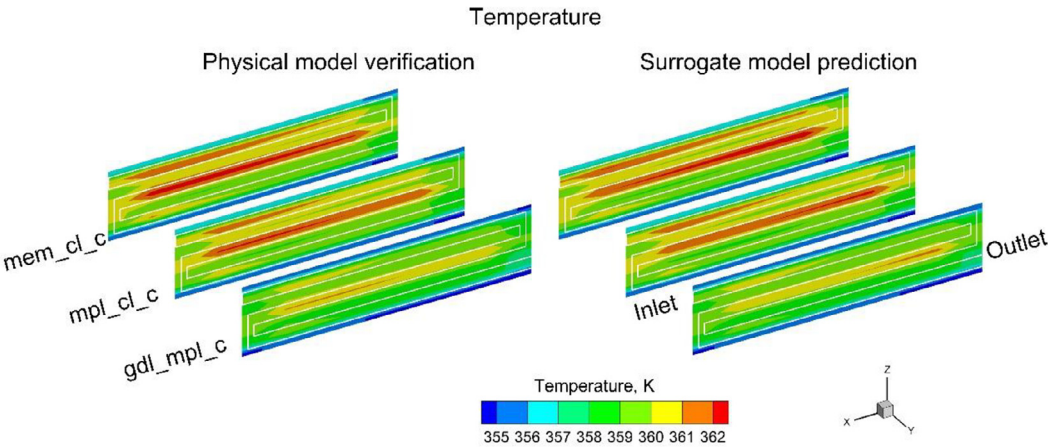
Fig. 4. Digital twin presentation. Physical model verified and surrogate model predicted multi-physics field contours which is selected from one of 25 test set samples. (a) Hydrogen in the anode (b) Relative humidity in the anode (c) Temperature in the anode (d) Oxygen in the cathode (e) Liquid saturation in the cathode (f) Temperature in the cathode. To demonstrate distribution characteristics of physical fields, the shapes of the anode parallel flow field and cathode serpentine flow field are displayed in the anode and cathode, respectively. Moreover, note that hydrogen is predicted by the ANN based model, and other variables are predicted by the SVM based model.



(d)



(e)



(f)

Fig. 4. Continued

Table 3

Comparison of rRMSE of multi-physics field prediction by the ANN based model and SVM based model in the training set and test set.

Physical quantities	Planes of physical fields	ANN based model rRMSE,%		SVM based model rRMSE,%	
		Training	Test	Training	Test
Hydrogen	gdl_mpl_a	2.59	4.69	9.71	10.86
	mpl_cl_a	2.13	3.88	9.71	10.86
	mem_cl_a	2.59	3.96	9.68	10.87
Relative humidity	gdl_mpl_a	5.71	23.19	18.99	18.78
	mpl_cl_a	10.83	22.66	19.12	18.64
	mem_cl_a	8.05	24.11	19.44	18.06
Oxygen	gdl_mpl_c	10.65	19.25	14.13	13.43
	mpl_cl_c	10.49	18.62	15.17	14.11
	mem_cl_c	10.39	19.96	15.56	14.11
Liquid saturation	gdl_mpl_c	9.21	20.07	20.46	24.16
	mpl_cl_c	9.29	15.85	20.96	24.80
	mem_cl_c	8.93	26.42	19.01	22.95
Temperature	gdl_mpl_a	6.88	24.19	16.05	10.54
	mpl_cl_a	7.62	19.50	16.30	10.56
	mem_cl_a	6.89	24.10	16.69	10.67
	gdl_mpl_c	6.83	17.18	16.47	10.62
	mpl_cl_c	7.18	19.95	16.72	10.98
	mem_cl_c	6.96	17.77	16.77	10.87

Table 4

The summary of prediction error measured by RMSE of multi-physics field prediction in the test set.

	Hydrogen RMSE, mol m ⁻³	Relative humidity RMSE	Temperature RMSE, K
gdl_mpl_a	0.5527	0.0209	0.2740
mpl_cl_a	0.5304	0.0210	0.3003
mem_cl_a	0.4656	0.0214	0.3421
	Oxygen RMSE, mol m ⁻³	Liquid saturation RMSE	Temperature RMSE, K
gdl_mpl_c	0.2169	0.0044	0.2815
mpl_cl_c	0.2184	0.0033	0.3257
mem_cl_c	0.2160	0.0070	0.3337

Note: Hydrogen is predicted by the ANN based model, and other variables are predicted by the SVM based model.

is within 1 s, whereas the that of one physical model case is approximately 8–12 h by a four-processes parallel calculation. Overall, it is notable that the digital twin framework provides comparable prediction results to the verified physical model, but with negligible computation and time compared to the implementation of the physical model. This offers possibilities for model-based designs. Moreover, two specific model-based designs are demonstrated to present the applications of the as-developed digital twin and data-driven surrogate modelling framework.

3.3. PEMFC healthy operation envelope

In real operations, it is preferred that the multi-physics fields of PEMFCs, including the reactant concentration, temperature, and water content, lie in a certain range and avoid the local problems, such as local gas starvation, hot spots, or liquid flooding, which is called as healthy operation [43]. Understanding the healthy operation ranges of a PEMFC in advance is helpful to ensure high performance, high efficiency, and low degradation in actual operation. However, as discussed, the high-definition and spatial distribution of multi-physics observations generally provides insufficient information for databank construction. Concurrently, conducting simulations in all the operating conditions by a complex physical model would consume significant computation and time. Here, we use the efficient digital twin framework to generate the PEMFC ‘healthy operation envelope.’

Tracking the extreme values (the maximum or minimum value) in the physical field is a rapid and simple approach to determine whether a PEMFC lies in the healthy operation range. Fig. 5 displays the healthy operation envelope of the PEMFC with regards to four physical quantities, i.e. temperature, oxygen, relative humidity, and liquid saturation,

within all the operating conditions. For visualisation, Fig. 5 only plots the healthy operation envelope with regards to the three most important variables only, which are back pressure, cathode stoichiometry ratio, and cathode relative humidity. Other operating condition parameters are kept as constant. To demonstrate the concept of PEMFC healthy operation envelope, a series of ISO surfaces is extracted, which divides the entire operating condition region into two parts. The grey region represents the healthy operating condition region, and the red or blue region represents the higher- or lower-value regions, which may reflect that a local hot spot, oxygen starvation, drying, or liquid flooding exists. Furthermore, note that to the best knowledge of the authors, there is no extensively recognised definition of the specific threshold value of a healthy operation in the previous literature, and the values of the ISO surfaces are defined according to value ranges of the physical quantity. The maximum values of the temperature fields and the minimum values of the relative humidity fields are generally in a relatively small range for the simulated PEMFC, which indicates that they are not very sensitive to the operating conditions. The minimum values of the oxygen concentration fields maintain a low level (0–0.2829 mol m⁻³), which indicates that oxygen starvation is an unavoidable problem for the simulated PEMFC. Benefiting from the healthy operation envelope, we can pre-set the optimal operating conditions region by a comprehensive consideration of multi-physics.

3.4. PEMFC state map

Real-time condition monitoring is significant for the operation of PEMFCs, which could provide guidance for the PEMFC control system. Most existing experimental methods of multi-physics fields observation are only suitable for laboratory testing and are difficult to con-

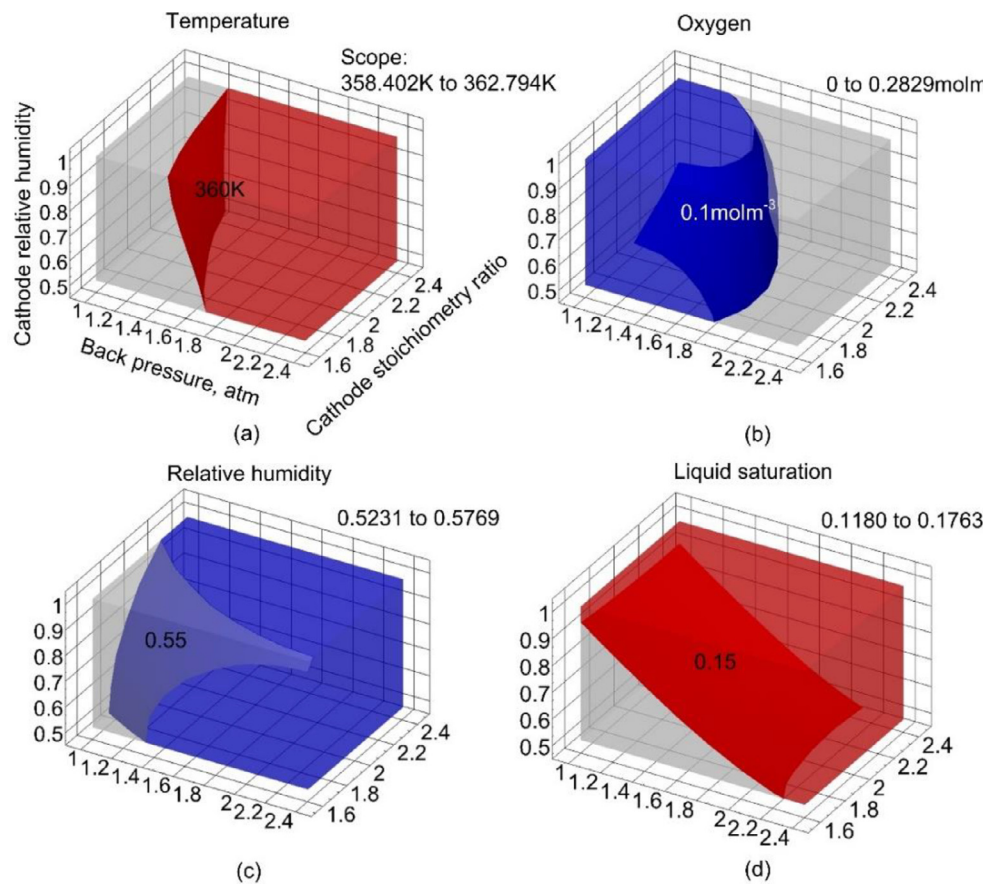


Fig. 5. PEMFC healthy operation envelope. (a) Maximum local values of temperature fields at mem_cl_a and mem_cl_c (b) Minimum local values of oxygen concentration filed at mem_cl_c (c) Minimum local values of relative humidity filed at mem_cl_a (d) Maximum local values of liquid saturation filed at mem_cl_c. Three most important operating condition parameters are chosen as variables for the healthy operation envelope, i.e. output voltage, anode stoichiometry ratio and anode relative humidity, are fixed at 0.6 V, 1.2, and 0.6, respectively. Series of ISO surfaces are extracted and the whole operating condition region is divided into two parts, and the grey region represents the healthy operating condition region, and the red region represents the higher-value region, and the blue region represents the lower-value region. The scopes of the physical quantities in the whole operating conditions are indicated.

duct in an in-situ automotive PEMFC. Control-oriented models should have the advantages of both high efficiency and high accuracy in physical state prediction, and numerous control-oriented models were developed, which are generally simplified physical models, including one-dimensional or quasi- two-dimensional models [44–46]. However, these simplified models cannot accurately reflect the multi-physics distributions and track the local problems, and serve as a digital twin. Here, the PEMFC state map based on the efficient digital twin framework is presented, which can offer a possibility of applying the simulation results of a complex model to the control-oriented model.

Fig. 6 shows the PEMFC state maps of three physical quantities at four representative nodes, which are node 1 (under-shoulder near the cathode outlet), node 25 (under-channel in midstream), node 38 (under-shoulder in midstream), and node 66 (integration of under-channel and under-shoulder) (as illustrated in Fig. 2(b)) in the entire operating conditions. For visualisation, Fig. 6 only plots three operating condition parameters as the variables which would influence the physical fields maximally by the theoretical analysis in each map, and the other parameters are kept constant, as listed in Table 5. Similar to the healthy operation envelope, a series of ISO surfaces is extracted in the PEMFC state map based on the value ranges in the entire operating conditions.

As depicted in Fig. 6(a), the oxygen concentration at the four nodes will increase with the back pressure and cathode stoichiometry ratio and decrease with the cathode relative humidity. According to the ranges of the oxygen concentration in the entire operating conditions at different nodes, the oxygen concentration at node 1 always keeps a very low level (the maximum is only $0.2466 \text{ mol m}^{-3}$). This suggests that increasing the cathode pressure and stoichiometry ratio only have a little influence on improving the oxygen starvation at node 1. Node 38 also has a very low region in the map, but its below 0.1 mol m^{-3} region

Table 5

The operating condition cases for PEMFC state maps in Fig. 5.

	Oxygen	Relative humidity	Liquid saturation
Output voltage, V	0.6	0.6	0.6
Back pressure, atm	–	2.0	2.0
Anode stoichiometry ratio	1.2	–	1.2
Cathode stoichiometry ratio	–	2.0	–
Anode relative humidity	0.6	–	–
Cathode relative humidity	–	–	–

Note: “–” represents the variable operating condition parameters used in the corresponding PEMFC state map, and their ranges are shown in Fig. 5.

is much smaller than that in node 1. Furthermore, the oxygen concentration at node 38 will be higher than 0.5 mol m^{-3} (the maximum is $0.8461 \text{ mol m}^{-3}$), which indicates that increasing the cathode pressure and stoichiometry ratio or decreasing relative humidity has a relative effective influence on improving the oxygen starvation at node 38. The oxygen concentrations at node 25 and 66 are relatively sufficient in the entire operating conditions. In the map, the upper left region of the ISO surface is generally lower than the value of the ISO surface, and the lower right region is higher than the value of the ISO surface. As shown in Fig. 6(b), the relative humidity at node 1 has a significant positive correlation with the anode inlet relative humidity and is slightly influenced by the anode stoichiometry ratio and the cathode relative humidity. However, the relative humidity at node 25 is relatively more influenced by the anode stoichiometry compared to that at node 1, because node 25 is an under-channel and node 1 is an under-shoulder. The values of the relative humidity at node 38 and 66 are both in a very limited range, which suggests that the relative humidity at these two nodes are insensitive in the entire operating conditions. There is no

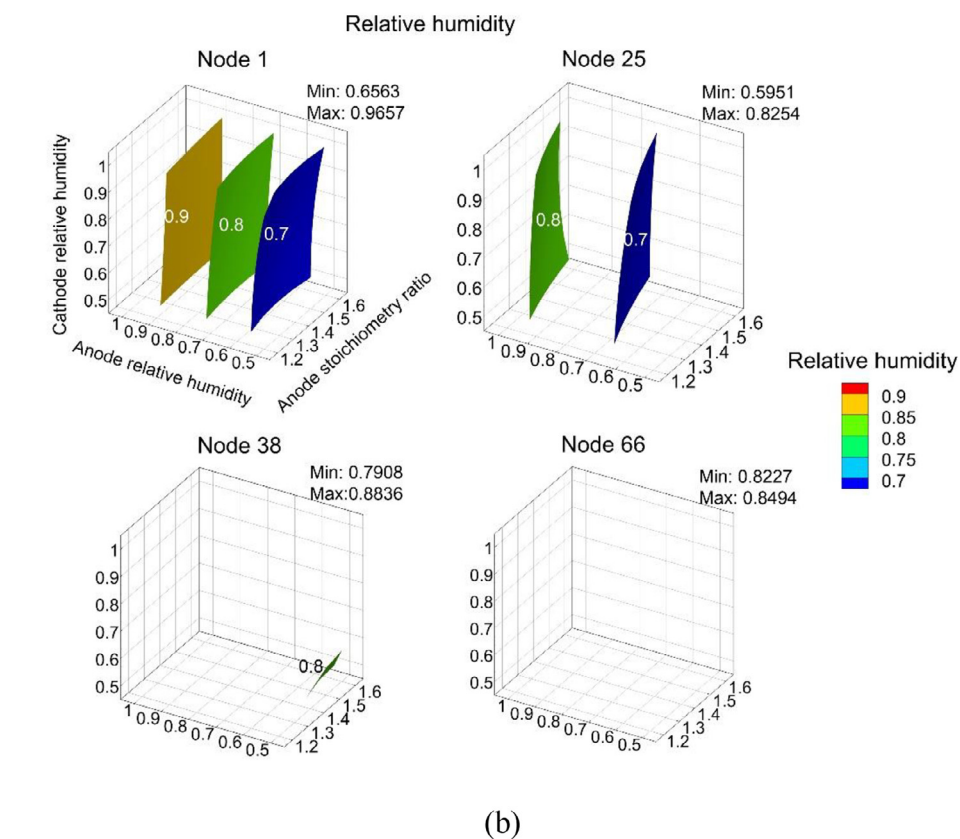
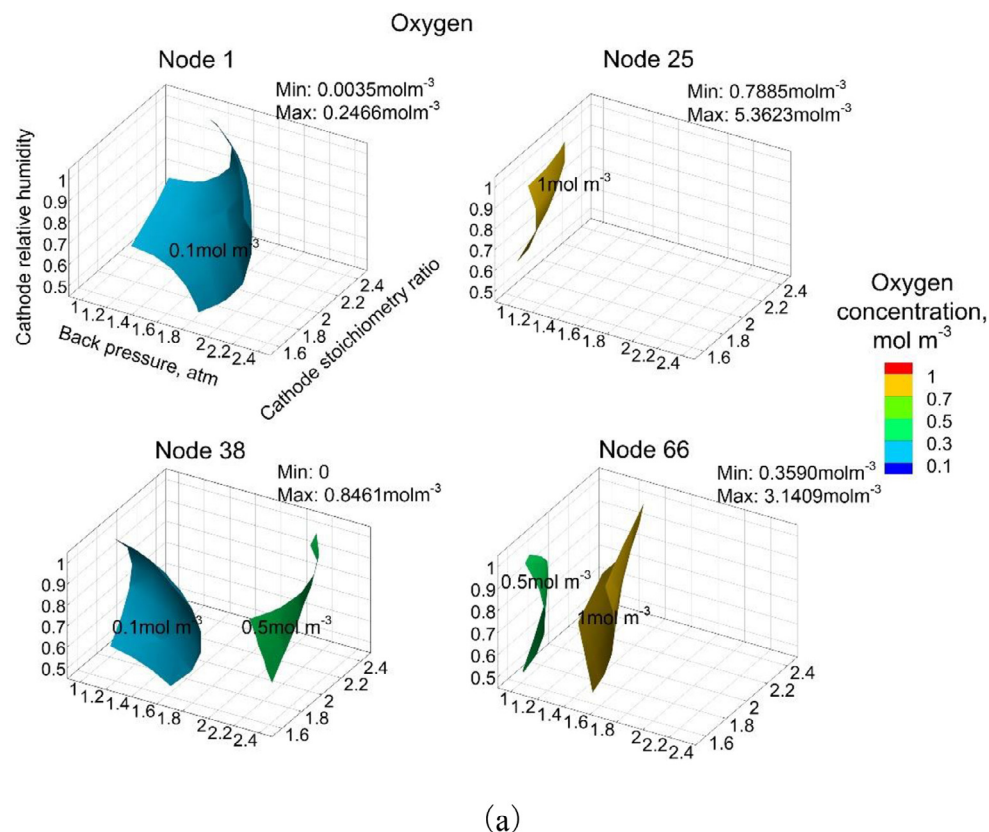


Fig. 6. PEMFC state maps. (a) Oxygen in gdl_cl_c (b) Relative humidity in gdl_cl_a (c) Liquid saturation in gdl_cl_c at four representative nodes in the whole operating conditions. Only three operating condition variables can be employed maximally in the map for visualisation and other operating condition variables are fixed values (see Table 5). Series of ISO surfaces are extracted for map visualisation.

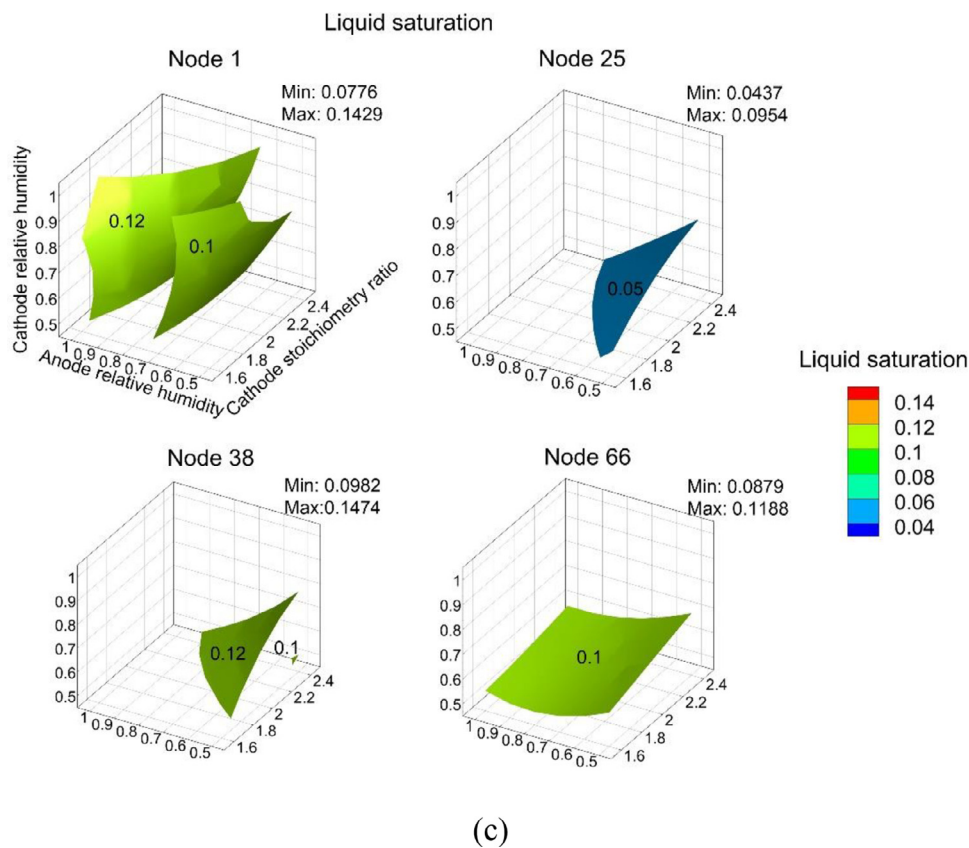


Fig. 6. Continued

ISO surface shown in the map of node 66 because the value scope may be even smaller than the prediction error. As depicted in Fig. 6(c), the low-liquid-saturation region is generally at the lower right end of the map, i.e. low cathode relative humidity, low anode relative humidity, and high cathode stoichiometry ratio. The liquid saturation at node 1 and 38 (under-shoulder) are higher than that at node 66 (integration of under-channels and under-shoulders) and node 25 (under-channel). Although the ranges of the liquid saturation at node 1 and 38 are similar, we can find that the high-liquid-saturation region (higher than 0.1 or 0.12) of node 38 is much larger than that of node 1, which suggests that the liquid flooding will be severer at node 38. Based on the shapes of the ISO surfaces, we can also find that the liquid saturation at node 25 and 38 are sensitive to the cathode stoichiometry ratio, that at node 1 is sensitive to the anode relative humidity, and that at node 66 is sensitive to the cathode relative humidity, respectively. This is probably because node 25 and 38 are located at the central flow area, node 1 is in the anode upstream, and node 66 is in the cathode upstream. Based on the PEMFC state map, the multi-physics distribution could be easily constructed in the entire operating conditions and applied as the control-orientated model in the PEMFC control system.

3.5. Limitations and outlooks

Although the 3D CFD modelling of PEMFCs is widely recognised as an effective method to simulate the physical PEMFC, and the physical model employed in this study is well-validated, the relevance between the physical model and the physical PEMFC is also expected to be enhanced. This will be helpful in realising a higher level of digital twins. Specifically, described below are the recommended approaches, and these would be our future research. First, the state-of-the-art 3D physical models still include some modelling assumptions, which make the simulated PEMFC an ideal one. For instance, a physical PEMFC may possess a poor assembly and its performance would

be worse than that of an ideal PEMFC. These problems are not considered in the present physical model, and thus, cannot be reflected by the present digital twin framework. Therefore, the proposed physical model is still expected to be improved in the future. Although the increasing model complexity will increase the cost of the computation and time, the data-driven surrogate model proposed in this study would provide support. Second, multiple easy-to-observe physical signals collected from the physical PEMFC are expected to be used in the physical model. Only six operating condition parameters are considered in this study. Third, an in-depth comparison with the observed results from the physical PEMFC to validate the physical model is relevant. There are some modelling parameters which are difficult to be provided by experiments. Extensive sensitivity analyses have proved that different modelling parameters can fit the same PEMFC characteristics in some operating conditions; however, these might become invalid in a new operating condition [47]. Using machine learning and optimisation algorithms to determine the uncertain parameters in the physical model and enhance the relevance between the physical model and the physical entity would be a potential prospect in physical model development.

4. Conclusion

In this study, we propose a data-driven surrogate modelling framework to achieve the multi-physics-resolved digital twin of PEMFCs. Simulation results of a state-of-the-art 3D physical model of PEMFCs construct a dataset and are applied to train the surrogate model. Two data-driven modelling methods with different data structures using ANN and SVM, respectively, are presented and compared. The prediction results show that the rRMSE in the test set of all the multi-physics fields range from 3.88% to 24.80%, and the prediction performance of the different physical quantities differs highly. The digital twin is represented in the form of multi-physics field contours and compared with the simu-

lation results of the physical model. The results demonstrate that the predicted multi-physics fields well mirror the main distribution characteristics of the simulation results. In summary, the data-driven surrogate model has a comparable accuracy for multi-physics field prediction to that of the comprehensive 3D physical model, whereas the computing time is within 1 s using an ordinary laptop. Moreover, its computing efficiency is thousands of times higher than that of the physical model. The proposed framework realises an efficient multi-physics-resolved digital-twin. Furthermore, two model-based designs, the PEMFC healthy operation envelope and PEMFC state map, are demonstrated which could be used to understand the healthy operation ranges in advance and run as the digital twin embedded into the PEMFC control system. Discussing extensively, this study points out the developing method and application prospects of combining data-driven approaches and simulation results of comprehensive physical models or other high-accuracy data to construct the digital twin, or other applications, of complex systems.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

This research is supported by the China-UK International Cooperation and Exchange Project (Newton Advanced Fellowship) jointly supported by the National Natural Science Foundation of China (grant No. 51861130359) and the UK Royal Society (grant No. NAF\R1\180146), and the Natural Science Foundation of Tianjin (China) for Distinguished Young Scholars (Grant No. 18JCJC46700).

Reference

- [1] Pollet BG, Kocha SS, Staffell I. Current status of automotive fuel cells for sustainable. *Curr Opin Electrochem*. 2019;16:90–5.
- [2] Cano ZP, Banham D, Ye S, Hintennach A, Lu J, Fowler M, Chen Z. Batteries and fuel cells for emerging electric vehicle markets. *Nat Energy* 2018;3(4):279.
- [3] Zheng W, Wang L, Deng F, Giles SA, Prasad AK, Advani SG, Yan Y, Vlachos DG. Durable and self-hydrating tungsten carbide-based composite polymer electrolyte membrane fuel cells. *Nat Commun* 2017;8:418.
- [4] Kojima K, Fukazawa K. Current status and future outlook of fuel cell vehicle development in Toyota. *ECS Trans* 2015;69(17):213–19.
- [5] Wang Y, Diaz DFR, Chen KS, Wang Z, Adroher XC. Materials, technological status, and fundamentals of PEM fuel cells – A review. *Mater Today* 2020;32:178–203.
- [6] Wang G, Yu Y, Liu H, Gong C, Wen S, Wang X, Tu Z. Progress on design and development of polymer electrolyte membrane fuel cell systems for vehicle applications: a review. *Fuel Process Technol* 2018;179:203–28.
- [7] Wang B, Deng H, Jiao K. Purge strategy optimization of proton exchange membrane fuel cell with anode recirculation. *Appl Energy* 2018;225:1–13.
- [8] Wang B, Chen W, Pan F, Wu S, Zhang G, Park JW, Xie B, Yin Y, Jiao K. A dot matrix and sloping baffle cathode flow field of proton exchange membrane fuel cell. *J Power Sources* 2019;434:226741.
- [9] Tao F, Qi Q. Make more digital twins. *Nature* 2019;572:490–1.
- [10] Lu Y, Liu C, Wang KI, Huang H, Xu X. Digital twin-driven smart manufacturing: connotation, reference model, applications and research issues. *Robot Comput Integr Manuf* 2020;61:101837.
- [11] Jiao K, Li X. Water transport in polymer electrolyte membrane fuel cells. *Prog Energy Combust Sci* 2011;37(3):221–91.
- [12] Lopes T, Beruski O, Manthanwar AM, Korkischko I, Pugliesi R, Stanojev MA, Meneghini JR. Spatially resolved oxygen reaction, water, and temperature distribution: experimental results as a function of flow field and implications for polymer electrolyte fuel cell operation. *Appl Energy* 2019;252:113421.
- [13] Zhao Y, Li X. A review of polymer electrolyte membrane fuel cell durability for vehicular applications: degradation modes and experimental techniques. *Energy Convers Manag* 2019;199:112022.
- [14] Shao H, Qiu D, Peng L, Yi P, Lai X. In-situ measurement of temperature and humidity distribution in gas channels for commercial-size proton exchange membrane fuel cells. *J Power Sources* 2019;412:717–24.
- [15] Kotaka T, Tabuchi Y, Pasaogullari U, Wang CY. Investigation of interfacial water transport at the gas diffusion media by neutron radiography. *ECS Trans* 2014;64(3):839–51.
- [16] Martens I, Vamvakeros A, Chattot R, Blanco MV, Rasola M, Pusa J, Heidemann S. X-ray transparent proton-exchange membrane fuel cell design for in situ wide and small angle scattering tomography. *J Power Sources* 2019;437:226906.
- [17] Tang YQ, Fang WZ, Lin H, Tao WQ. Thin film thermocouple fabrication and its application for real-time temperature measurement inside PEMFC. *Int J Heat Mass Transf* 2019;141:1152–8.
- [18] Siegel JB, Bohac SV, Stefanopoulou AG, Yesilyurt S. Nitrogen front evolution in purged polymer electrolyte membrane fuel cell with dead-ended anode. *J Electrochem Soc* 2010;157(7):B1081–93.
- [19] Wang B, Xie B, Xuan J, Jiao K. AI-based optimization of PEM fuel cell catalyst layers for maximum power density via data-driven surrogate modeling. *Energy Conversion Manag* 2020;205:112460.
- [20] Ma R, Yang T, Breaz E, Li Z, Briois P, Gao F. Data-driven proton exchange membrane fuel cell degradation prediction through deep learning method. *Appl Energy* 2018;231:102–15.
- [21] Jeppesen C, Araya SS, Sahlin SL, Thomas S, Andreasen SJ, Kær SK. Fault detection and isolation of high temperature proton exchange membrane fuel cell stack under the influence of degradation. *J Power Sources* 2017;359:37–47.
- [22] Bejan A. AI and freedom for evolution in energy science. *Energy and AI* 2020;1:100001.
- [23] Li W, Zhu J, Xia Y, Gorji MB, Wierzbicki T. Data-Driven safety envelope of lithium-ion batteries for electric vehicles. *Joule* 2019;3(11):2703–15.
- [24] Sun Y, Szűcs G, Brandt AR. Solar pv output prediction from video streams using convolutional neural networks. *Energy Environ Sci* 2018;11(7):1811–18.
- [25] Rio-Chanona EA, Wagner JL, Ali H, Fiorelli F, Zhang D, Hellgardt K. Deep learning-based surrogate modeling and optimization for microalgal biofuel production and photobioreactor design. *AIChE J* 2019;65(3):915–23.
- [26] Berning T, Lu DM, Djilali N. Three-dimensional computational analysis of transport phenomena in a PEM fuel cell. *J Power Sources* 2002;106(1–2):284–94.
- [27] Jiao K, Li X. Three-dimensional multiphase modeling of cold start processes in polymer electrolyte membrane fuel cells. *Electrochim Acta* 2009;54(27):6876–91.
- [28] Um S, Wang CY. Three-dimensional analysis of transport and electrochemical reactions in polymer electrolyte fuel cells. *J Power Sources* 2004;125(1):40–51.
- [29] Carnes B, Spornjak D, Luo G, Hao L, Chen KS, Wang CY, Borup RL. Validation of a two-phase multidimensional polymer electrolyte membrane fuel cell computational model using current distribution measurements. *J Power Sources* 2013;236:126–37.
- [30] Hao L, Moriyama K, Gu W, Wang CY. Three dimensional and experimental comparisons for a large-scale proton exchange membrane fuel cell. *J Electrochem Soc* 2016;163(7):F744–51.
- [31] Iranzo A, Boillat P, Rosa F. Validation of a three dimensional PEM fuel cell CFD model using local liquid water distributions measured with neutron imaging. *Int J Hydrog Energy* 2014;39(13):7089–99.
- [32] Wang YK, Chen KS. Through-plane water distribution in a polymer electrolyte fuel cell: comparison of numerical prediction with neutron radiography data. *J Electrochem Soc* 2010;157(12):B1878–86.
- [33] Yoshida T, Kojima K. Toyota mirai fuel cell vehicle and progress toward a future hydrogen society. *Electrochim Soc Interface* 2015;24(2):45–9.
- [34] Macedo-Valencia J, Sierra JM, Figueroa-Ramírez SJ, Díaz SE, Meza M. 3D CFD modeling of a PEM fuel cell stack. *Int J Hydrog Energy* 2016;41(48):23425–33.
- [35] Zhang G, Xie B, Du Q, Jiao K. Large-scale multi-phase simulation of proton exchange membrane fuel cell. *Int J Heat Mass Transf* 2019;130:555–63.
- [36] Zhang G, Xie B, Bao Z, Niu Z, Jiao K. Multi-phase simulation of proton exchange membrane fuel cell with 3D fine mesh flow field. *Int J Energy Res* 2018;42(15):4697–709.
- [37] Carton JG, Olabi AG. Three-dimensional proton exchange membrane fuel cell model: comparison of double channel and open pore cellular foam flow plates. *Energy* 2017;136:185–95.
- [38] Zhang G, Wu J, Wang Y, Yin Y, Jiao K. Investigation of current density spatial distribution in PEM fuel cells using a comprehensively validated multi-phase non-isothermal model. *Int J Heat Mass Transf* 2020;150:119294.
- [39] Ozen DN, Timurkutluk B, Altinisik K. Effects of operation temperature and reactant gas humidity levels on performance of PEM fuel cells. *Renew Sustain Energy Rev* 2016;59:1298–306.
- [40] Carnes B, Spornjak D, Luo G, Hao L, Chen KS, Wang CY, Borup RL. Validation of a two-phase multidimensional polymer electrolyte membrane fuel cell computational model using current distribution measurements. *J Power Sources* 2013;236:126–37.
- [41] Mathworks, feedforwardnet. <https://uk.mathworks.com/help/deeplearning/ref/feedforwardnet.html>.
- [42] Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol (TIST)* 2011;2(3):27.
- [43] Wang R, Zhang G, Hou Z, Wang K, Zhao Y, Jiao K. Comfort index evaluating the water and thermal characteristics of proton exchange membrane fuel cell. *Energy Convers Manag* 2019;185:496–507.
- [44] Wang B, Wu K, Yang Z, Jiao K. A quasi-2D transient model of proton exchange membrane fuel cell with anode recirculation. *Energy Convers Manag* 2018;171:1463–75.
- [45] Xu L, Fang C, Li J, Ouyang M, Lehnert W. Nonlinear dynamic mechanism modeling of a polymer electrolyte membrane fuel cell with dead-ended anode considering mass transport and actuator properties. *Appl Energy* 2018;230:106–21.
- [46] Gao F, Blunier B, Simoes MG, Miraoui A. PEM fuel cell stack modeling for real-time emulation in hardware-in-the-loop applications. *IEEE Trans Energy Convers* 2010;26(1):184–94.
- [47] Wang XD, Xu JL, Lee DJ. Parameter sensitivity examination for a complete three-dimensional, two-phase, non-isothermal model of polymer electrolyte membrane fuel cell. *Int J Hydrog Energy* 2012;37:15766–77.