Three-dimensional multi-field digital twin technology for proton exchange membrane fuel cells

Fan Bai [a], Hong-Bing Quan [a], Ren-Jie Yin [a], Zhuo Zhang [a], Shu-Qi Jin [a], Pu He [a], Yu-Tong Mu [b], Xiao-Ming Gong [a], Wen-Quan Tao [a,*]

[a] *Key Laboratory of Thermo-Fluid Science & Engineering of MOE, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, PR China*
[b] *School of Human Settlements and Civil Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, PR China*

HIGHLIGHTS

- A novel 3D multi-physics field digital twin model for PEMFCs is proposed.
- The computational fluid dynamic technique is integrated in the digital twin model.
- The model is demonstrated within twenty randomly selected working conditions.
- The proposed model can predict PEMFC physics field characteristics in 0.913 s.

ARTICLE INFO

ABSTRACT

In times of the commercialization process of proton exchange membrane fuel cells (PEMFCs), a full knowledge of in-situ state in PEMFCs is of critical significance to the in-situ operational process and the evaluation of material stage and potential damage. The conventional experimental observation and in-situ prediction models can only obtain very limited information while the computational fluid dynamics approach takes too long time to get the detailed information. To reach a full knowledge of PEMFC real-time state, a novel 3D multi-physics digital twin model for PEMFCs is proposed based on the proper orthogonal decomposition (POD) method. In the model, firstly, for one kind of PEMFC, 139 ex-situ snapshots are designed and simulated based on the three-dimensional two-phase non-isothermal numerical model with the assumption of liquid pressure continuity in the whole membrane electrode assembly. Then the modes of each field in snapshots are extracted by singular value decomposition method using Jacobi algorithm. Finally, the coefficients in the POD prediction equation are obtained by using the multivariate adaptive regression splines. The digital twin results of voltage, temperature, membrane water content and liquid water saturation fields are exhibited and analyzed. Results suggest that for the studied PEMFC, the digital twin technique can capture the global values and the local distribution characteristics of each above physical fields well in 0.913 s. The mean global deviations of the above four fields of 20 groups of random conditions within wide current density and operational condition ranges are 5.7 %, 1.3 %, 8.9 % and 12.0 % respectively. Even though the practical results can only be applied for the studied PEMFC, the proposed methodology has its general application range.

1. Introduction

Nowadays, with the accelerating climate and carbon neutrality agenda [1,2] around the world, energy industry is being transformed from fossil fuel-dominated structure to renewable and sustainable energy-dominated one with an ever-increasing rate. However, the intermittency of renewable energies restricts its further popularization and application [3,4]. Hydrogen is regarded as an important energy storage method hopefully to resolve this issue [4]. As an important hydrogen energy utilization equipment, proton exchange membrane fuel cell (PEMFC) will be one of the most promising widespread power generating devices in the future. Although challenges still exist [5–12], very recently, the commercialization of PEMFCs has been developed rapidly around the world [13–18].

A full knowledge of real-time state in PEMFCs (for instance, flooding

---

Nomenclature

Symbol		θ	Jacobi rotation angle/rad
F	Physical field matrix	λ	Lagrange multiplier
J	Functional	σ	Single value
J	Jacobi matrix	$\boldsymbol{\psi}$	Analytical or vector expression of mode
R	Kernel matrix/function	ψ	Discrete expression of mode
St	Stoichiometric ratio	ψ'	Non-normalization mode
U	Left single vectors		
V	Right single vectors	*Abbreviations*	
b	Weight coefficient	3D	Three-dimensional
ƒ	Analytical or vector expression of physical field	BP	Bipolar plate
f	Discrete expression of physical field	CCL	Cathode catalyst layer
J	Current density/A·m^{-2}	CFD	Computational fluid dynamics
l	Truncation order	CGDL	Cathode gas diffusion layer
n	Snapshot number	CH	Gas flow channel
Δ	Truncation error	CMPL	Cathode micro pore layer
Σ	Singular value matrix	MEM	Proton exchange membrane
Ω	Three-dimensional Euclidean space	PEMFC	Proton exchange membrane fuel cell
δ	Kronecker symbol	POD	Proper orthogonal decomposition
		SVD	Singular value decomposition
		UDF	User-defined function

---

[19] and membrane dehydration [20], etc.) is of critical significance to the in-situ operational process [21] and the evaluation of lifetime. Moreover, proper and timely adjustments of the abnormal states for water and thermal managements benefit to a longer mileage and lifetime [22]. Due to the complex micron scale structures, the in-situ measurements of complete physical fields in PEMFCs are hardly possible by the conventional experimental methods [23]. Therefore, a series of in-situ prediction models are proposed, including mechanism model [24–29], semi-empirical model [30–36] and experimental data model [21,37–42]. The mechanism model and experimental data model are established based on mathematical and physical mechanism (represented by partial differential equations) and data driven model (represented by machine learning technology), respectively. The semi-empirical model seeks balance between them. However, the predicted physical quantities in the above models are generally very limited. As indicated by Rubio and Agila [21] very recently, in the operation of PEMFCs, designing a robust model considering the distribution of water in each layers as a function of inlet gas humidity, heat transfer, gas diffusion, etc., is an important further research work.

So far, researchers have proposed lots of conventional computational fluid dynamics (CFD) models to numerically predict three-dimensional (3D) multi-physics fields (for instance, gas velocity, pressure, concentrations of gas species, voltage, liquid water, membrane water content, etc.) in PEMFCs [43–58], among which 3D multi-phase non-isothermal model is the most comprehensive numerical model [56,57]. In general, the conventional CFD simulation for one current density point on a polarization curve will take several hours to days even in the investigation of the simplest flow channel model (single straight channel model). However, during the practical PEMFC real-time operation, a quick online response, as little as several seconds or less, is highly required. Therefore, the conventional CFD method can hardly be applied directly in in-situ operational process.

The concept of digital twin is firstly proposed by Professor Grieves in 2003 and is comprehensively defined by NASA as an aircraft oriented [59,60]. Digital twin is extensively considered as a revolutionizing technology in the industry [61]. The concept of digital twin covers lots of aspects, virtualization-multi-physics modeling, full life cycle management (from product design to manufacturing stage) and integration with other technologies, etc. [60] This paper focuses on the first aspect (i.e., visualization-multi-physics modeling) and paved the way for further applications, diagnosis, in-situ operations and lifetime predictions. To the authors' knowledge, digital twin is firstly introduced

into PEMFC research by Wang et al. [23] in 2020 with a data-driven surrogate model. In their research, 75 random ex-situ solutions are simulated previously as a training-set. After training by machine learning algorithms, the predictions of multi-physics fields under in-situ operational conditions can be very fast and the relative root mean square errors of the multi-physics fields for the test-set range from 3.88 % to 24.80 %. However, although solutions in the training-set are simulated by a state-of-the-art three-dimensional numerical model, their training models are still applied on limiting two-dimensional planes. Moreover, the ranges of the studied operational conditions are limited. Later, several researchers focus on digital twins in the aspects of prognostics method [62], fault diagnosis [63], the hydrogen powertrain and the auxiliary systems [64]. However, due to the difficulty caused by the large amount (or order) of data and the complex non-linear coupled characteristics, there are few studies on the digital twin in the aspect of complete 3D virtualization-multi-physics modeling in PEMFCs.

The critical points to resolve this issue are how to find the low-dimensional space that describes the original physical problem and how to establish its reduced-order model in the low-dimensional space. Proper orthogonal decomposition (POD) technology is a feasible reduced-order model for the fast and accurate predictions of velocity and temperature fields in the fluid flow and heat transfer problems [65]. POD was firstly introduced into fluid mechanics by Lumley et al. [66] for the analysis of coherent structure in turbulent flow. Till now, POD has been successfully applied in the dynamical analysis of micro-electromechanical systems [67,68], the extraction of ocean surface wave fields [69], the analysis of satellite and radar observations for the performance enhancement of weather forecast [70], etc. POD is a reduced order technology extracting the characteristic functions (called modes) from a large number of ex-situ numerical or experimental data. The in-situ physical fields can be quickly predicted (generally less than 1 s) from the linear combination of the modes. Furthermore, it is interesting to note that although POD technology is based on the above linear combination of modes, it is also feasible in the analysis of non-linear problems [71] (for example, Navier-Stokes equations in fluid mechanics). Therefore, POD is a potential solution to deal with the above-referred difficulties in the 3D digital twin of PEMFCs.

In summary, in the researches about PEMFC in-situ experimental observations, the observed quantities are very limited. Although the computational fluid dynamic approach can obtain 3D multi-physics fields in PEMFCs, the simulation time is too long to satisfy the in-situ prediction needs. In the literatures about PEMFC multi-physics digital

twin, the training models are applied only on two-dimensional planes. To reach a full knowledge of PEMFC real-time state, this paper pertains to realize a complete 3D virtualization-multi-physics digital twin for PEMFCs (abbreviated as digital twin hereafter this paper) by POD technology. In the remainder of this paper, the basic principles and algorithms of the digital twin process are briefly introduced in Section 2 and the results are exhibited and discussed in Section 3. Finally, some conclusions are given in Section 4.

2. Methodology

2.1. Brief introduction to POD technology

The principle of POD is similar to Fourier decomposition method better known by researchers. POD is a reduced-order technique which expands physical fields as a linear equation of the modes extracted from a series of beforehand numerical or experimental results, called snapshots, as is shown in Eq. (1).

$$f_k(\Omega) = \sum_j \psi_j(\Omega) b_{j,k} \tag{1}$$

where Ω represents the 3D Euclidean space, f_k is the analytical expression (or vector) of one physical field in the kth snapshot, ψ_j is the normalization modes defined in the Euclidean space, $b_{j,k}$ is the weight coefficient of the jth mode for the studied physical field in the kth snapshot. An example of the above analytical expression of one physical field or one mode is $f(\Omega) = xyz^2$, where x, y and z are the rectangular Cartesian axes.

In order to guarantee the centralization of the information for the convenience of further order reduction, the selection of modes ψ_j require:

$$\begin{cases} \max: & \langle |(f_k, \psi)|^2 \rangle \\ s.t. & \|\psi\| = 1 \end{cases} \tag{2}$$

where $(\cdot, \cdot)$, $\langle \cdot \rangle$ and $\|\cdot\|$ represent the inner product, the mean value and the L2 norm respectively. Eq. (2) is a variational problem of functional with constraint. According to Lagrange multiplier method, Eq. (2) can be transformed into a variational problem without constraint, as shown in Eq. (3) [72].

$$\max: J[\psi] = \langle |(f_k, \psi)|^2 \rangle - \lambda (\|\psi\|^2 - 1) \tag{3}$$

where λ is the Lagrange multiplier. According to the variational principle, ψ requires

$$\frac{\delta}{\delta \psi} J[\psi] = 0 \tag{4}$$

Solving Eq. (4), yielding following eigenvalue problem.

$$(\mathbf{R}, \psi) = \lambda \psi \tag{5}$$

where $\mathbf{R}$ is the kernel function which can be calculated by Eq. (6).

$$\mathbf{R}(\Omega, \Omega^{\cdot}) = \langle f_k^*(\Omega^{\cdot}) \otimes f_k(\Omega) \rangle \tag{6}$$

where $\cdot^*$ and $\cdot \otimes \cdot$ represent the transpose and the averaged autocorrelation function (or tensor product) respectively. Detailed derivation processes can be referred to the supplementary material of this paper. ψ_j can be obtained by solving the eigenvalue problem in Eq. (5). Due to the high-order characteristic of $\mathbf{R}$ (node number order after discretization), the solution number of ψ in Eq. (5) is also huge. Furthermore, the solutions of ψ are pairwise orthogonal, as shown in Eq. (7).

$$(\psi_j, \psi_k) = \delta_{j,k} \tag{7}$$

where $\delta_{j,k}$ is the Kronecker symbol taking 1 and 0 when $j = k$ and $j \neq k$ respectively. Under one working condition, once the weight coefficients

are determined, the physical field can be predicted based on Eq. (1).

The above-mentioned technology is the classical POD. However, the order of the eigenvalue problem in Eq. (5) is still very high, leading to a huge storage requirement for the kernel function $\mathbf{R}$. Note that the modes can also be expressed as the formation of linear combinations of snapshots, as shown in Eq. (8).

$$\psi'_k(\Omega) = \sum_j f_j(\Omega) V_{j,k} \tag{8}$$

where $\psi'_k(\Omega)$ is the non-normalization mode which is the multiple transformation of $\psi_k(\Omega)$. Substituting Eq. (8) into the classical POD technique, the solution of the high order eigenvalue problem shown in Eq. (5) can be transformed into a snapshot number order eigenvalue problem solving the characteristic vectors $V_{j,k}$. This technology is called snapshot method and is applied in this paper.

In the numerical results of PEMFCs, physical fields appear as the collection of discrete data points. Eqs. (1), (5), (6), (7) and (8) can be transformed into discrete form, as shown in Eqs. (9), (10), (11), (12) and (13) respectively. For conciseness, all formulas hereafter in this paper follow Einstein's summation convention.

$$f_{i,k} = \psi_{i,j} b_{j,k} \tag{9}$$

$$R_{i,j} \psi_{j,k} = \lambda_k \psi_{i,k} \tag{10}$$

$$R_{i,j} = \frac{1}{n} f_{i,k} f_{j,k} \tag{11}$$

$$\psi_{k,i} \psi_{k,j} = \delta_{i,j} \tag{12}$$

$$\psi'_{i,j} = f_{i,k} V_{k,j} \tag{13}$$

where the first indexes of f and ψ represent the index of discrete grid node, n represents the snapshot number. Matrix $V_{k,j}$ can be selected as the right singular vector of the snapshot matrix $f_{i,j}$, detailed derivation processes of which can be referred to the supplementary material.

It should be noted that the three-dimension characteristics of a physical field are not reflected directly in Eq. (9). Each column of $f_{i,k}$ stores the same type of physical quantities (such as voltage) on all the grid nodes in the computational domain of each snapshot. For the numerical solutions using one pre-generated grid system, each node itself contains the geometric position information. Therefore the discrete expression $f_{i,j}$ will not reduce the dimension of the physical space itself compared with the based numerical model.

In summary, the centerpiece of POD is the linear transformation relationship between snapshot space and weight coefficient space, as shown in Eqs. (9) and (13). With that in mind, the general algorithm for the digital twin technology is as follows. (a) Design of snapshot configuration. (b) Generation of snapshots. (c) Extraction of the modes. (d) Prediction of the weight coefficients under in-situ conditions. (e) Construction of digital twin results by modal superposition. In the following these steps will be described in order.

2.2. Snapshot configuration design

It is necessary to design their simulating conditions before the generation of the snapshots. There are two critical concepts in the snapshot configuration design process. Factors represent all the variable parameters among the snapshots. Levels represent all the specific values of each factor. In this paper, a given PEMFC with straight gas flow channel is studied. Only current density and seven operational conditions, including the pressures, stoichiometric ratios, relative humidities at the inlets of anode and cathode gas flow channels, and the operational temperature are varied. They are selected as factors. The levels of each factor are shown in Table 1. There are a variety of test configuration design methods, such as full factorial design [73], orthogonal

Table 1
Designing levels for each factor.

Factor Unit Level	Current density A/m^2 –	Temperature °C Anode, cathode and terminals	Pressure atm Anode	Cathode	Relative humidity % Anode	Cathode	Stoichiometric ratio – Anode	Cathode
1	1000	30	1	1	0	0	1.2	1.2
2	2000	40	2	2	10	10	1.5	1.5
3	3000	50	3	3	20	20	1.7	1.7
4	5000	60	4	4	30	30	1.9	1.9
5	7000	70		5	40	40	2.1	2.1
6	9000	78			50	50	2.4	2.4
7	11,000	82			60	60	2.7	2.7
8	12,000	86			70	70	3	3
9	13,000	90			80	80		
10	14,000				90	90		
11	15,000				100	100		

experimental design [74], pairwise testing [75], uniform design [76], etc. Orthogonal experimental design method is a famous test configuration design algorithm. However, when the factor number or level number is large, it is very difficult to find a software to implement orthogonal experimental design. In this paper, pairwise testing is applied and a test configuration design tool developed by Microsoft, Pairwise Independent Combinatorial Testing (PICT), is adopted which can partially cope with the above-mentioned issue. After excluding four groups of abnormal working conditions (represented by the condition in which the setting current density is higher than the limiting current density), 139 groups of operational parameters are finally selected from the design results of PICT, as shown in Table S1.
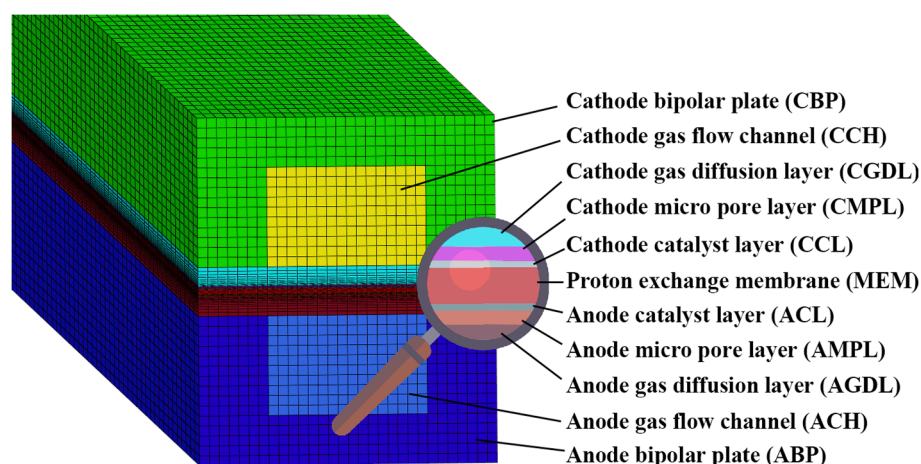
2.3. Generation of snapshots

The physical fields in the PEMFCs defined in Table S1 are simulated using the three-dimensional multi-phase non-isothermal numerical model with the assumption of liquid pressure continuity. The geometric diagrammatic sketch and grid system of the computational domain are shown in Fig. 1. Total node number is 482,143 (channel length: 150 mm). The governing equations are summarized in Table 2. Details about corresponding source terms, boundary conditions, physical property parameters and coupling relationships can be referred to [57]. The governing equations are solved by the finite volume method in ANSYS FLUENT 14.0 with self-developed user-defined function (UDF). The numerical model in this paper is verified by comparing with Ozen's experimental polarization curve [77], as shown in Fig. 2.

The physical fields of PEMFCs under 139 groups of working conditions in Table S1 are adopted as snapshots. One single simulation follows the strategy of calculating current density with voltage boundary con-

dition. However, in Table S1, current density is the preset condition. Therefore, the output voltage is adjusted until the current density reaches the preset value. There are lots of snapshots to be simulated. It will take plenty of time if the current density of each snapshot is accurately adjusted to the set value. Fortunately, the accurate adjustment is not necessary. In this paper, when the simulated current density is adjusted within ± 1 % compared with the preset value, it is considered as convergence. In the numerical model, the preset current density and stoichiometric ratios only play a part in the calculation of inlet gas flow rate. Therefore, the actual simulated working conditions are not strictly coincident with the preset ones. Only the current density and stoichiometric ratios will slightly deviate from the preset values. The actual current density of each snapshot is the current density that the simulation converges to. The actual stoichiometric ratios can be converted according to Eq. (14).

$$St_{\text{real}} = \frac{St_{\text{preset}} J_{\text{preset}}}{J_{\text{real}}} \tag{14}$$

where St and J (A·m^{-2}) represent stoichiometric ratio and current density respectively. The actual convergent working conditions and output voltage of each snapshot are shown in Table S1. In the following analyses and calculations related to the snapshot current densities and operational conditions, the actual convergent ones are adopted. It should be noted that this treatment can completely eliminate the possible error caused by the slight deviation of the current density. Finally, after extracting the simulation results using 'Tecplot' format and the position and topology information in the mesh file (suffix: '.msh'), the snapshot matrix for each physical field can be obtained and the statistics for the physical fields in each layer or surface component can be freely accessed.

Cathode bipolar plate (CBP)
Cathode gas flow channel (CCH)
Cathode gas diffusion layer (CGDL)
Cathode micro pore layer (CMPL)
Cathode catalyst layer (CCL)
Proton exchange membrane (MEM)
Anode catalyst layer (ACL)
Anode micro pore layer (AMPL)
Anode gas diffusion layer (AGDL)
Anode gas flow channel (ACH)
Anode bipolar plate (ABP)

Fig. 1. The geometric diagrammatic sketch and grid system of the single straight channel PEMFC model.

Table 2
Governing equations of the 3D multi-phase non-isothermal model with the assumption of liquid pressure continuity.

Physical quantity	Solution zone	Governing equation
Density (ρ)	CHs, GDLs, MPLs, CLs	$\dfrac{\partial(\rho_g u_i)}{\partial x_i} = S_m$
Velocity (u_i)	CHs, GDLs, MPLs, CLs	$\dfrac{\partial}{\partial x_k}\left(\dfrac{\rho_g u_k u_j}{\varepsilon^2(1-s_{lq})^2}\right) = -\dfrac{\partial p_g}{\partial x_j} + \dfrac{\partial}{\partial x_i}\left(\mu_e\dfrac{\partial}{\partial x_i}\left(\dfrac{u_j}{\varepsilon(1-s_{lq})}\right)\right) + \dfrac{\partial}{\partial x_i}\left(\mu_e\dfrac{\partial}{\partial x_j}\left(\dfrac{u_i}{\varepsilon(1-s_{lq})}\right)\right)$ $-\dfrac{2}{3}\dfrac{\partial}{\partial x_j}\left(\mu_e\dfrac{\partial}{\partial x_i}\left(\dfrac{u_i}{\varepsilon(1-s_{lq})}\right)\right) + S_{uj}$
Species (Y_i)	CHs, GDLs, MPLs, CLs	$\dfrac{\partial(\rho_g u_k Y_j)}{\partial x_k} = \dfrac{\partial}{\partial x_i}\left(\rho_g D_j^{eff}\dfrac{\partial Y_j}{\partial x_i}\right) + S_{Y_j}$
Energy (T)	Entire domains	$\dfrac{\partial(\varepsilon s_{lq}\rho_{lq} c_{p,lq} u_{lq,j}T + \varepsilon(1-s_{lq})\rho_g c_{p,g} u_j T)}{\partial x_j} = \dfrac{\partial}{\partial x_i}\left(k^{eff}\dfrac{\partial T}{\partial x_i}\right) + S_E$
Electronic potential (ϕ_{ele})	BPs, GDLs, MPLs, CLs	$0 = \dfrac{\partial}{\partial x_i}\left(\kappa_{ele}^{eff}\dfrac{\partial\phi_{ele}}{\partial x_i}\right) + S_{ele}$
Ionic potential (ϕ_{ion})	CLs, MEM	$0 = \dfrac{\partial}{\partial x_i}\left(\kappa_{ion}^{eff}\dfrac{\partial\phi_{ion}}{\partial x_i}\right) + S_{ion}$
Membrane water content (λ)	CLs, MEM	$\dfrac{\partial}{\partial x_j}\left(\dfrac{n_d}{F}J_{ion,j}\right) = \dfrac{\rho_{MEM}}{EW}\dfrac{\partial}{\partial x_i}\left(D_{mw}^{eff}\dfrac{\partial\lambda}{\partial x_i}\right) + S_{mw}$
Liquid water pressure (p_{lq})	GDLs, MPLs, CLs	$0 = \dfrac{\partial}{\partial x_i}\left(\rho_{lq}\dfrac{Kk_{lq}}{\mu_{lq}}\dfrac{\partial p_{lq}}{\partial x_i}\right) + S_{lq}$
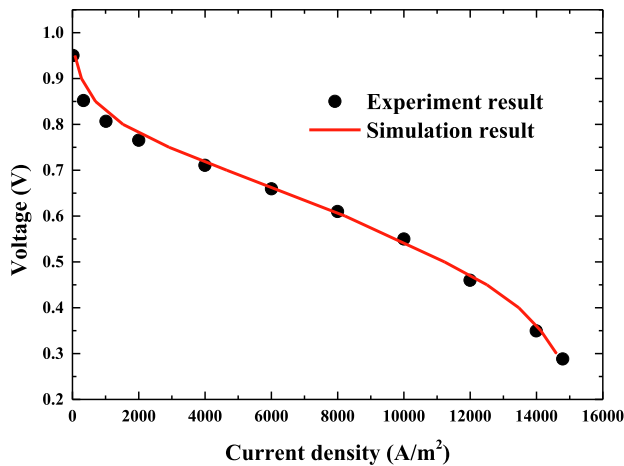
Fig. 2. Verification of the 3D multi-phase non-isothermal numerical model compared with the experimental polarization curve in Ref. [77].

2.4. Extraction of the modes

In general, there are three methods to extract modes, Karhunen–Loève decomposition (KLD), principal component analysis (PCA) and singular value decomposition (SVD) [67]. Due to its wide application in signal processing [78], SVD method is adopted in this paper. The centerpiece of SVD is the decomposition of the snapshot matrix **F** as Eq. (15).

$$\mathbf{F} = \mathbf{U\Sigma V}^* \tag{15}$$

where **U** and **V** are unitary matrixes, called left singular vectors and right singular vectors respectively, and **Σ** is the singular value matrix which is a non-square diagonal matrix.

SVD can be implemented by various of algorithms, for instance, QR algorithm [79], QD algorithm [80], Jacobi algorithm [81], etc. In this paper, one-sided Jacobi algorithm [81] is applied and realized using C++ to calculate the right singular vectors. The key of one-sided Jacobi algorithm is to decompose right singular vector to a series of Jacobi transformation, as shown in Eqs. (16) and (17).

$$\mathbf{V} = \mathbf{J}_0\mathbf{J}_1\mathbf{J}_2\cdots\mathbf{J}_N \tag{16}$$

$$\mathbf{J}_k = \left\{\begin{matrix} 1 & & & & & \\ & \ddots & & & & \\ & & \cos\theta_k & & -\sin\theta_k & \\ & & & \ddots & & \\ & & \sin\theta_k & & \cos\theta_k & \\ & & & & & \ddots \\ & & & & & 1 \end{matrix}\right\} \tag{17}$$

According to Eqs. (13) and (15), after a series of Jacobi transformations, the kernel matrix $\mathbf{R} = \mathbf{F}^*\mathbf{F}$ can be transformed into a diagonal matrix, as shown in Eq. (18).

$$\mathbf{\Psi}^{'*}\mathbf{\Psi}^{'} = \mathbf{V}^*\mathbf{F}^*\mathbf{F}\mathbf{V} = \mathbf{J}_N^*\cdots\mathbf{J}_2^*\mathbf{J}_1^*\mathbf{J}_0^*\mathbf{F}^*\mathbf{F}\mathbf{J}_0\mathbf{J}_1\mathbf{J}_2\ldots\mathbf{J}_N = \mathbf{\Sigma}^*\mathbf{\Sigma} = \mathbf{\Sigma}^2 \tag{18}$$

The implementation steps of one-sided Jacobi algorithm are as follows:

(a) Kernel matrix initialization

$$\mathbf{R}^k = \mathbf{F}^*\mathbf{F}, k = 0 \tag{19}$$

(b) Loop all pairs $1 \leq p < q \leq n$. Extract the second order principal submatrix with the indexes of p and q in the present kernel matrix $\mathbf{R}^k$. Execute steps (c) and (d).

$$\begin{pmatrix} R_{p,p}^k & R_{p,q}^k \\ R_{q,p}^k & R_{q,q}^k \end{pmatrix} \tag{20}$$

(c) Single Jacobi matrix calculation.

$$\cot 2\theta_k = \dfrac{R_{p,p}^k - R_{q,q}^k}{R_{p,q}^k + R_{q,p}^k} \tag{21}$$

After one single Jacobi transformation, The second order principal submatrix are transformed as diagonal matrix, as shown in Eq. (22).

$$\begin{pmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{pmatrix}\begin{pmatrix} R_{p,p}^k & R_{p,q}^k \\ R_{q,p}^k & R_{q,q}^k \end{pmatrix}\begin{pmatrix} \cos\theta_k & -\sin\theta_k \\ \sin\theta_k & \cos\theta_k \end{pmatrix} = \begin{pmatrix} R_{p,p}^{k+1} & 0 \\ 0 & R_{q,q}^{k+1} \end{pmatrix} \tag{22}$$

(d) Update the core matrix by the present Jacobi transform, as shown in Eq. (23).

$$\mathbf{R}^{k+1} = \mathbf{J}_k^*\mathbf{R}^k\mathbf{J}_k \tag{23}$$

(e) Repeat step (b) until the whole kernel matrix is transformed into a diagonal matrix. All the Jacobi matrixes are obtained.

(f) Non-normalization mode calculation: substitute Eq. (16) into Eq. (13).

(g) Normalization of each mode and calculation of singular value according to Eqs. (24) and (18).

$$\psi_i = \frac{\psi'_i}{\|\psi'_i\|} \tag{24}$$

It is interesting to note that each singular value σ_i represents the amount of information captured by the corresponding mode. Eq. (18) can be transformed into the column vector form, as shown in Eq. (25).

$$\sigma_i = \sqrt{(\mathbf{FV}_i)^*(\mathbf{FV}_i)} = \left\| (f_1 \, f_2 \, \cdots f_n) \begin{pmatrix} V_{1,i} \\ V_{2,i} \\ \vdots \\ V_{n,i} \end{pmatrix} \right\| = \|\psi'_i\| \tag{25}$$

Therefore, the singular value of each mode can be understood as the coefficient divided in the normalization step (g). Considering that a series of Jacobi rotation transformations $\mathbf{J}_k$ don't change the magnitude of the information, the bigger the coefficient σ_i is, the more important the corresponding mode to the original physical field is. Therefore, singular value can be understood as the amount of information captured by a mode.

2.5. Weight coefficient calculation

The weight coefficients at the snapshot conditions can be calculated by projecting snapshots on the mode space, as shown in Eq. (26).

$$b_{i,j} = \psi_{k,i} f_{k,j} \tag{26}$$

Furthermore, derived from Eqs. (9) and (26), obviously, the snapshot space f_j and the weight coefficient space b_j are isomorphic according to the definition. Therefore, the weight coefficient space can completely represent the original PEMFC problem without reducing the amount of information contained in the snapshot data.

After sorting the modes by the descending order of their respective singular values, the order of the problem can be further reduced. Due to the fact that the information concentration of the mode can be guaranteed in POD, as shown in Eq. (2), with the increase of the mode order, the amount of information will be greatly reduced. High order modes containing small amount of information can be ignored. The number of the modes considered and the proportion of the total singular values captured by the ignored modes are called truncation order l and truncation error Δ respectively, as shown in Eqs. (27) and (28).

$$f_j = \sum_{i=1}^{n} b_{i,j} \psi_i \approx \sum_{i=1}^{l} b_{i,j} \psi_i \tag{27}$$

$$\Delta = \frac{\sum_{i=l+1}^{n} \sigma_i}{\sum_{i=1}^{n} \sigma_i} \tag{28}$$

In practice, there are many cases in which the working conditions are within the ranges prespecified in Table 1. None of a working condition is exactly the same as any of the design snapshots in Table S1. They are called off-design conditions. Using the same modes extracted from the snapshots, the physical fields under off-design working conditions can also be quickly and accurately predicted, called extrapolation. It is this remarkable characteristic that makes POD can be applied in PEMFC engineering for digital twin. It should be noted that even in the ranges of the design working conditions shown in Table 1, it is also difficult to predict the weight coefficients for the solution desired directly according to the weight coefficients of the original snapshots. This is because, on the one hand, the non-linear characteristics of the numerical model. For example, the source terms of the voltage field in catalyst layers are non-

linearly relative to lots of other physical fields in Table 2 [57]. On the other hand, the simulated snapshots are finite while the number of the studied factors is up to 8, resulting in a prediction error.

The extrapolation process can be abstracted to the following mathematical models: giving a series of snapshot vectors b_i under their respective input vectors x_i, how to predict the vector b' under an off-design vector x'. In the conventional POD technology in the community of thermofluidic engineering, interpolation method and Galerkin projection method are widely applied to predict b' [65]. However, POD interpolation method requires a full factorial snapshot configuration design method. According to the design levels in Table 1, 15.33 million snapshots need to be simulated, which will take about 28 millennium·cores. The required computation resource is unrealistic. With regard to the Galerkin projection method [82], the extremely complex coupling characteristics of the governing equations in Table 2 lead to a great difficulty.

In mathematics, regression is also an effective method to resolve this issue. Regression guarantees that b' in the previous paragraph can be predicted based on a relatively small number of snapshots, which is also the reason for the discussion of the test configuration design method in Section 2.2. In the regression method, the relationship between input and output vectors $b = g(x)$ is modelled according to the snapshot set firstly. And then the vector b' can be predicted by substituting the corresponding input vector x' into the relationship g. It is very difficult to model g by conventional mathematical regression methods due to a high dimension of input vector x. In the past decade, machine learning technology has been developed rapidly to resolve the prediction problem in lots of domains. In this paper, a machine learning model, the multiple adaptive regression spline (MARS) [83], is applied by self-programming based on Python toolkit.

Finally, after the prediction of b', the physical field can be extrapolated by Eq. (29).

$$\widehat{f}'_i \approx \sum_{j=1}^{l} \widehat{b}'_j \psi_{i,j} \tag{29}$$

where $\widehat{}$ represents the predicted value.

From the above discussion, it can be seen that there are three types of prediction methods in this paper: CFD simulation, reconstruction and digital twin. The general purposes of these methods are coincident, to predict the 3D multi-physics fields in PEMFCs. However, their algorithms and application scopes are quite different. The conventional simulation method is implemented by the discretization of the governing equations and the solution of the linear simultaneous equations with the order of control volume number. It is time consuming thus can only be applied directly in ex-situ analyses. Simulation provides snapshots as the training set of the other two methods. The reconstruction is implemented by Eq. (27). The modes and weight coefficients are all known beforehand in the prediction process. The prediction time of the reconstruction is the least. It can only be applied in the snapshot conditions shown in Table S1. The digital twin is implemented by Eq. (29). The modes are known beforehand. The weight coefficients are predicted by the pre-trained machine leaning model with the order of truncation order (less than snapshot number). Thus its prediction time is a little longer than the reconstruction. It can be applied in most of the conditions within the ranges prespecified in Table 1 (for instance: current density 1000–15,000 A·m^{-2}, etc.).

3. Results and discussion

After implementing the algorithm in Section 2 for each physical field, digital twin can finally be realized. In this section, the mode extraction, reconstruction and digital twin results of various physical fields, voltage, temperature, membrane water content, liquid water pressure and liquid water saturation, are exhibited and analyzed. In the following, modes of each physical field will firstly be extracted in Section 3.1, followed by

the reconstruction of physical fields under the snapshot conditions in Section 3.2, and finally the digital twin will be presented for randomly selected current densities and operational conditions in Section 3.3.

3.1. Characteristics of extracted modes

This subsection exhibits the mode extraction results. In POD, one mode may capture most of the information. In this situation, a potential small relative deviation for the prediction of the weight coefficient of this mode will result in a large deviation for the predicted physical field, which may reduce or even submerge the influence of other higher order modes on the physical field. For example, the temperature difference in one PEMFC is usually several degrees centigrade while the range of the operational temperature is 60 °C. The above-referred mode capturing the datum values will appear if temperature fields are directly adopted to generate snapshot matrix in POD. However, it is not necessary to extract this information from the snapshots because it is known beforehand. With that in mind, the datum values of temperature, anode operational pressure and cathode operational pressure in each snapshot are subtracted from the temperature, anode liquid water pressure and cathode liquid water pressure fields respectively, which are selected as snapshot matrices.

The orthogonality of modes is firstly validated. According to the extracted modes, the mean value of the principal diagonal elements, the standard deviation of the principal diagonal elements and the L1 norm of the non-principal diagonal elements in Eq. (12) are shown in Table 3. We can see that the orthogonality deviations of the modes are tiny.

The distribution contours of the first four modes for the voltage field are presented in Fig. 3. As shown, the first mode represents the general magnitude of the voltage field in most layers. The second mode mainly capture the information of the voltage variety inconsistency in the GDLs, MPLs and CLs. Other modes correct the voltage field in several local regions, represented by the region near CH inlet and outlet surfaces. The modes for other physical fields show similar distribution patterns, as shown in the supplementary material, and will not be presented here.

The information proportion captured by each mode and the information accumulation proportion captured by the first several modes are shown in Fig. 4 and Fig. 5 respectively. Each curve in Fig. 4 shows the information proportion of each mode as a function of mode order. Generally speaking, with the increase of the mode order, the amount of information captured by the mode decreases rapidly, which is consistent with the original intention for further order reduction. The truncation orders to reach different truncation errors for each physical field are shown in Table 4. For example, if it is needed to reach 0.01 % truncation error for the predicted voltage, then 24 modes should be included in Eqs. (27) and (29).

Come here to discuss the mode-captured information for each physical field shown in Fig. 4.

For voltage field, the first mode captures a tremendous amount of the

information. It alone captures 98.6 % of the total information. This means that the first mode represents the most important characteristic of voltage field. Therefore, the accurate prediction of the first order weight coefficient is of critical importance to the prediction of output voltage and voltage field. The first two modes capture 99.7 % of the total information. The later modes capture much smaller amount of information. For example, the third and sixth modes only capture 0.16 % and 0.02 % of the total information respectively. Although this, not all of the third to 139th modes can be neglected. This is because that these modes represent the contribution of the non-linear characteristics (such as source term) or local characteristics to the total information. An arbitrary truncation of modes may lead to the loss of local field characteristics.

For relative temperature field, the first mode captures 70.2 % of the total information, which is much less than that in the voltage field. This is because that in most of the regions in PEMFC, the Laplace's equation is the governing equation of voltage, leading to a low complexity of the voltage field. Therefore, the information is centralized in the first two modes. However, a convection diffusion equation with complex nonlinear source terms is solved for thermal simulation in most of the regions. More characteristics are necessary for the accurate description of the temperature field. Therefore, the information proportion captured by the first mode is lower.

For liquid water saturation field and relative liquid water pressure field, the first modes capture 65.0 % and 65.4 % of the total information respectively, which is also much less than that in the voltage field. The information proportions of the same order modes in the two fields are nearly identical. This is because that there is an obvious conversion relationship between the two fields, as shown in Eqs. (30) and (31).

$$p_{ca} = p_g - p_{lq} = \sigma\cos\theta\left(\frac{\varepsilon}{K_0}\right)^{0.5} G(s_{lq}) \tag{30}$$

$$G(s_{lq}) = \begin{cases} 1.42(1 - s_{lq}) - 2.12(1 - s_{lq})^2 + 1.26(1 - s_{lq})^3 & \theta < 90° \\ 1.42s_{lq} - 2.12s_{lq}^2 + 1.26s_{lq}^3 & \theta > 90° \end{cases} \tag{31}$$

where p_{ca} (Pa), p_g, p_{lq}, σ (N·m^{-1}), θ, ε, K_0 (m^2) and s_{lq} are capillary pressure, gas pressure, liquid water pressure, surface tension, contact angle, porosity, permeability and liquid water saturation respectively. The two physical fields show different characteristics. For example, the liquid water saturation field jumps across the interface between two layers while the liquid water pressure field is continuous. However, they are essentially the same, representing liquid water field. Considering the intuitiveness, the liquid water saturation field is selected hereafter in this paper.

For membrane water content field, the first mode captures 68.4 % of the total information, which is close to that in the liquid water field. This is because it is membrane water that carries protons across the membrane while liquid water is generated from the cathode electrochemical reaction. For a given working condition, the membrane water required to carry protons and the liquid water generated are both closely related to the electrochemical reaction rate. Furthermore, there is a direct phase change relationship between membrane water and liquid water. Therefore, the information proportion curves of the two fields are close to each other.

3.2. Reconstruction of cases identical to snapshot

Physical fields under the snapshot conditions can be reconstructed according to Eq. (27) after selecting different truncation orders. For each physical field in PEMFCs, on the one hand, there are a series of representative local values, for instance, output voltage, maximum temperature, maximum membrane water content and maximum liquid water saturation. On the other hand, it is necessary to have a good prediction

Table 3
Mode orthogonality validation for each physical field.

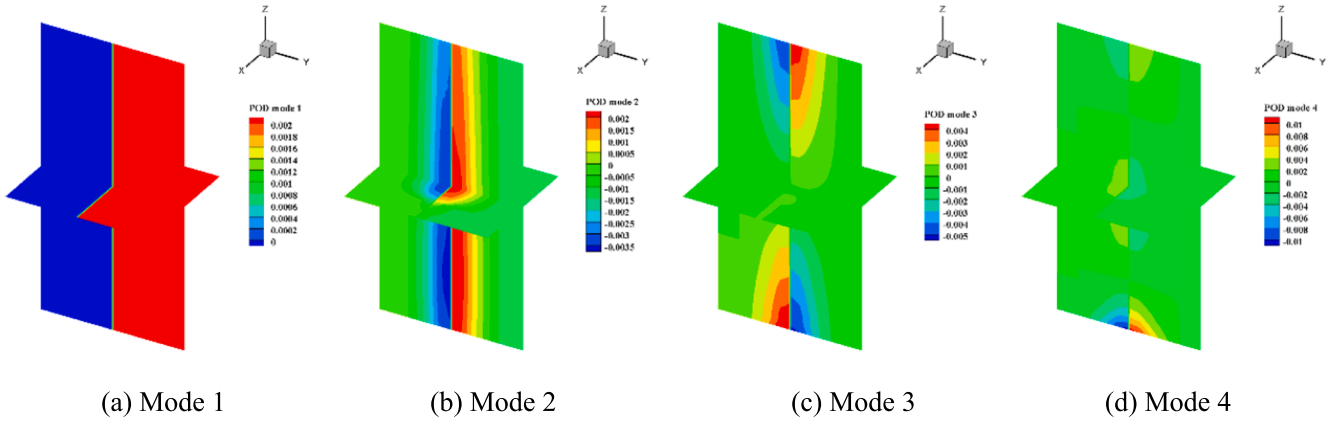|	Mean value of the principal diagonal elements	Standard deviation of the principal diagonal elements	L1 norm of the non-principal diagonal elements
Voltage	1.000	2.793×10^{-10}	6.875×10^{-11}
Relative temperature	1.000	2.160×10^{-10}	5.991×10^{-11}
Membrane water content	1.000	3.339×10^{-10}	9.450×10^{-11}
Relative liquid water pressure	1.000	2.732×10^{-10}	6.981×10^{-11}
Liquid water saturation	1.000	2.569×10^{-10}	7.222×10^{-11}

(a) Mode 1 (b) Mode 2 (c) Mode 3 (d) Mode 4

Fig. 3. The first four modes for voltage field. (In each mode, the left and right parts represent anode and cathode respectively. The channel inlet surfaces are located on the bottom. Intermediate cross sections perpendicular to the x and z axes are selected as typical surfaces to present the 3D fields. The scale factor in the channel length direction, z axis, is 0.0465 for visualization. The similar presentation pattern is adopted hereafter in this paper.).
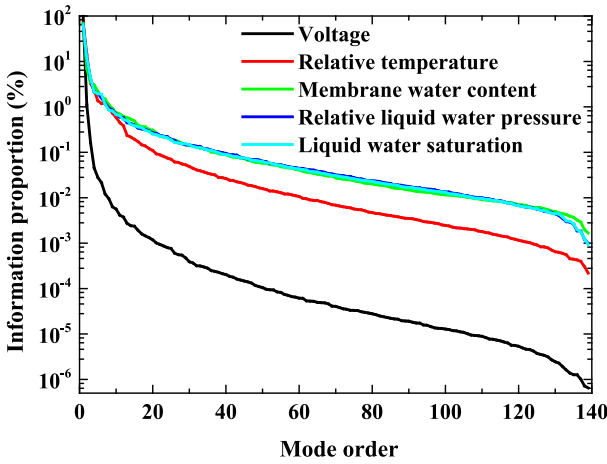

Fig. 4. The information proportion of each mode as a function of mode order for each physical field.
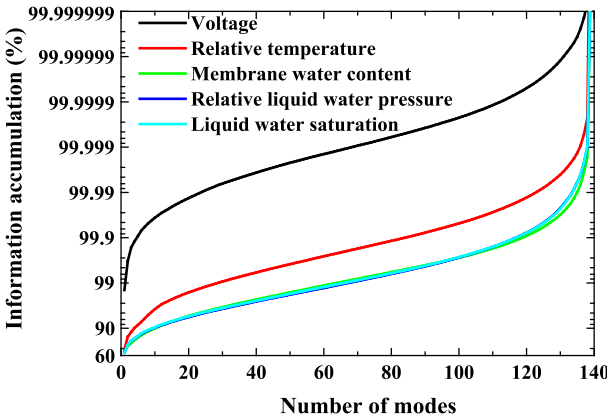

Fig. 5. The information accumulation proportion of the first i modes as a function of mode number i for each physical field.

of the whole field. The prediction deviations of them are called local deviation and global deviation respectively. The global relative deviation of each snapshot can be defined by L2 norm, as shown in Eq. (32).

Table 4
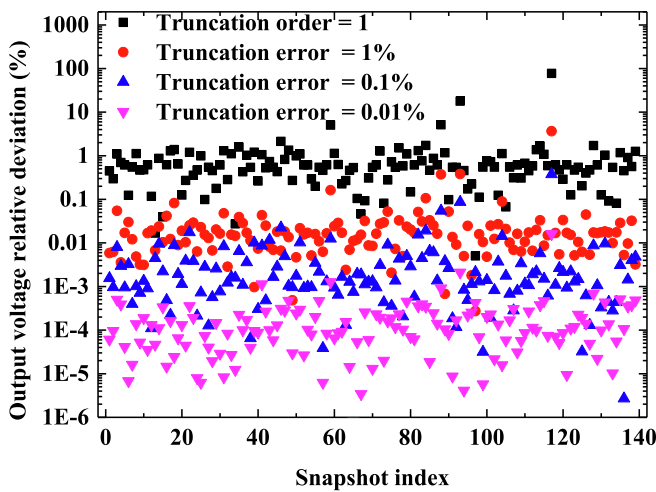Truncation orders to reach different truncation errors for each physical field.

Truncation error	1%	0.1%	0.01%
Physical Field	Truncation order		
Voltage	2	5	24
Relative temperature	29	84	123
Membrane water content	64	119	135
Relative liquid water pressure	66	118	134
Liquid water saturation	68	117	134

$$\delta_j = \sqrt{\frac{\sum_i \left(f_{i,j} - \widehat{f}_{i,j}\right)^2}{\sum_i f_{i,j}^2}} \tag{32}$$

Selections of one mode (1st mode) and numbers of modes with truncation errors of 1 %, 0.1 % and 0.01 % are applied for the snapshot reconstruction. The local and global relative deviations of the four studied fields are shown in Figs. 6–9 respectively. From Fig. 6 (a), it can be seen that for voltage field, when the mode number is selected to reach a truncation error of 1 % (2 modes as shown in Table 4), then all the snapshots have the deviations around 1 % shown by the black points. As shown in the figures, generally speaking, with the reduction of the truncation error, the local and global relative deviations decrease significantly by several orders of magnitude. Compared with the local relative deviation, the global relative deviation is more concentrated among different snapshots. For example, the red points in Fig. 8 (a) are scattered while those in Fig. 8 (b) are concentrated. This is because POD technology is based on the characteristic extraction of the whole field, while the local values represent the extreme points of the physical fields. They are obtained according to the statistical results of the predicted fields. Therefore, more potential prediction uncertainties among snapshots are introduced for the local values.

For voltage field, as shown in Fig. 6, when truncation order is equal to 1, the reconstructed output voltages for most of the snapshots reach a relative deviation of within 5.2 %. This is also because the first POD mode captures the most dominant information, as analyzed above. However, large relative deviations will be introduced for several snapshots. For example, the output voltage relative deviations of the 93rd and 117th snapshots are 17.9 % and 77.3 % respectively. This is because the working conditions of the cathode sides in the two snapshots are disgusting. Their current densities are up to 1.5 and 1.3 A/cm^2, while the operational pressures are only 1 atm, the temperatures are high (82 °C) and the cathode stoichiometric ratios are low (1.5 and 1.2). On the one hand, under these working conditions, the output voltages themselves
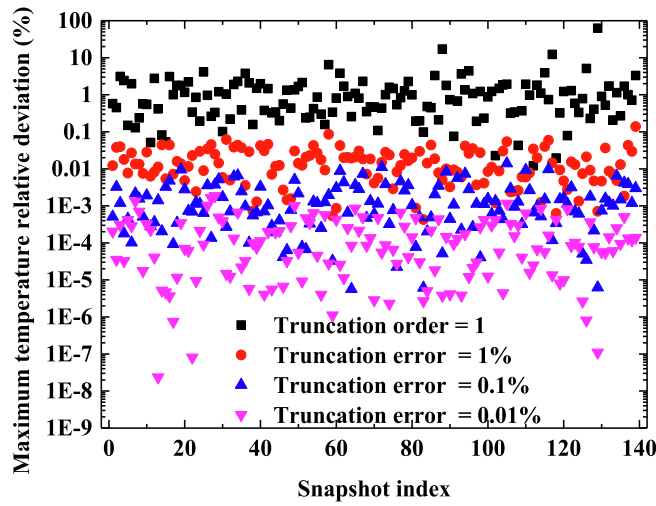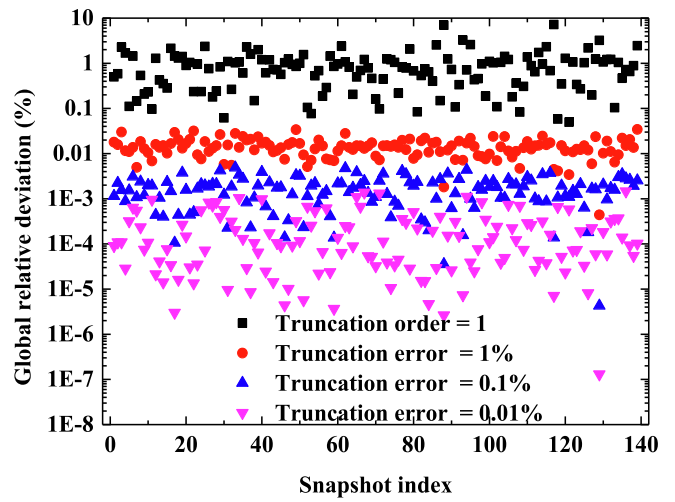
(a) Output voltage relative deviation
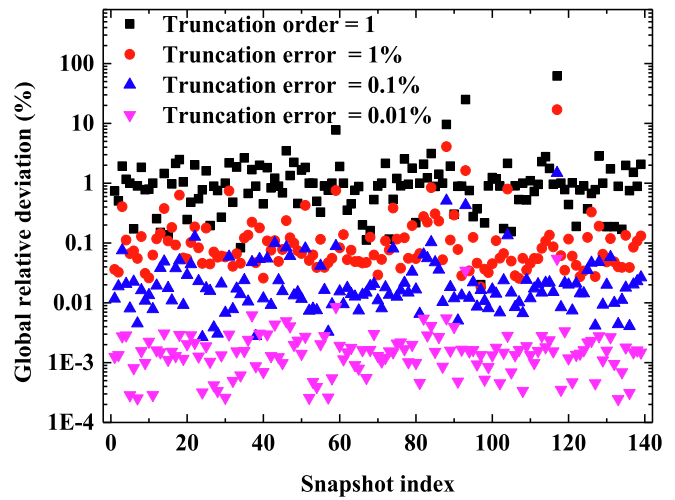
(b) Global relative deviation

Fig. 6. Reconstruction relative deviation of voltage field for each snapshot under different truncation order or truncation error. (a) Output voltage relative deviation. (b) Global relative deviation.

(a) Maximum temperature relative deviation

(b) Global relative deviation

Fig. 7. Reconstruction relative deviation of temperature field for each snapshot under different truncation order or truncation error. (a) Maximum temperature relative deviation. (b) Global relative deviation.

are small, resulting in the situation that information captured by other modes reaches the same order of magnitude as the first one. For example, the weight coefficients of the first and second modes in the 117th snapshot are 11.01 and 8.37 respectively. Therefore, truncation at the second mode will lead to a high relative deviation. On the other hand, at the above working condition, the fuel cell is working at the part of the polarization where concentration is dominant. Therefore, in the regions with strongly nonlinear characteristics, more nonlinear characteristics captured by higher order modes are necessary in order to accurately predict the voltage field.

For temperature field, as shown in Fig. 7, when truncation order is equal to 1, the global relative deviations are slightly higher than that of the voltage field. This is consistent with the information accumulation proportion shown in Fig. 5. For membrane water content and liquid water saturation fields, as shown in Fig. 8 and Fig. 9, respectively, when truncation order is equal to 1, the relative deviations are higher than the above two fields, up to more than 10 %. More modes are necessary in

order to accurately predict these fields. According to the above analyses and Figs. 6-9, the truncation error is selected as 0.01 % in order to comprehensively guarantee the reconstruction accuracy of each reconstructed field under various working conditions as much as possible.

Due to their importance in the output voltage, heat and water managements in PEMFCs, the reconstructed 3D contours of the voltage, temperature and liquid water saturation fields are discussed taking the 117th snapshot as an example. The comparison between the simulation results and reconstruction results are shown in Figs. 10-12 respectively. We can see that when truncation error equal to 0.01 %, the reconstructed snapshot results are nearly identical with the simulation results, verifying the accuracy of the reconstruction. Furthermore, as shown in Fig. 11, the local temperature in this snapshot reaches 130 °C. This is because that this case has a high current density but low output voltage, the efficiency of the PEMFC is very low, leading to a great amount of energy dissipation in the form of heat.
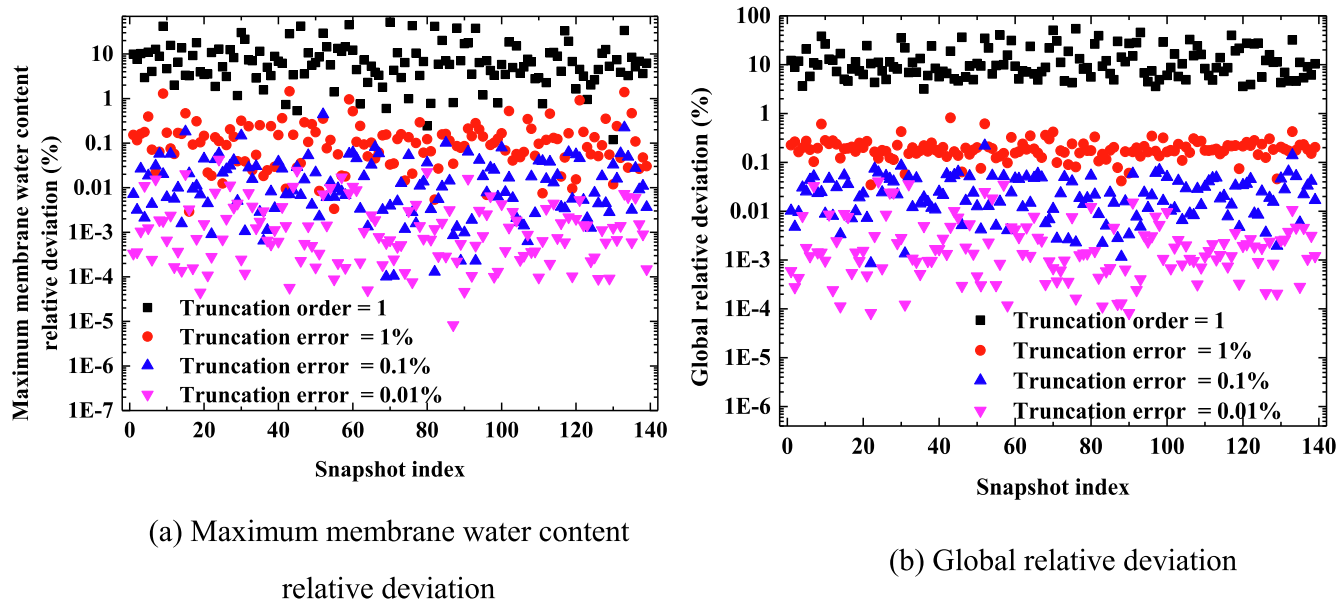
(a) Maximum membrane water content

relative deviation

(b) Global relative deviation

Fig. 8. Reconstruction relative deviation of membrane water content field for each snapshot under different truncation order or truncation error. (a) Maximum membrane water content relative deviation. (b) Global relative deviation.
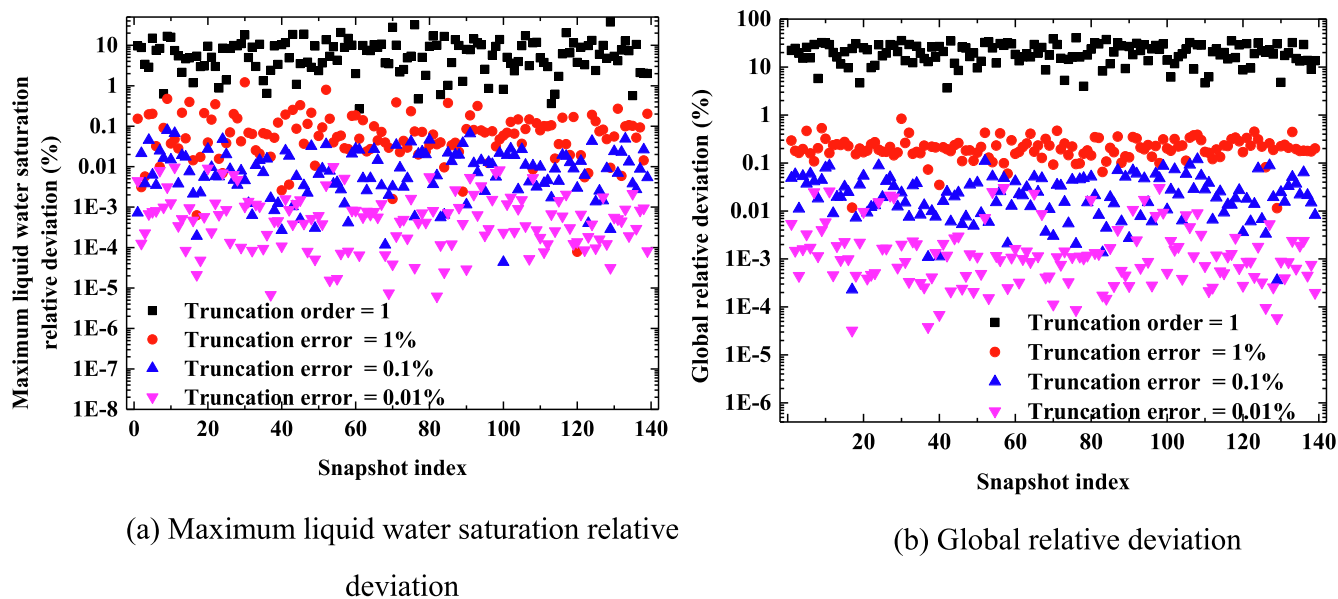

(a) Maximum liquid water saturation relative

deviation

(b) Global relative deviation

Fig. 9. Reconstruction relative deviation of liquid water saturation for each snapshot under different truncation order or truncation error. (a) Maximum liquid water saturation relative deviation. (b) Global relative deviation.

3.3. Digital twin for 20 cases of random working conditions

This subsection shows the comparison between the digital twin and simulation results under random off-design working conditions. Twenty cases of working conditions are randomly generated within the parameter ranges shown in Table 1. The design and actual convergent working conditions are shown in Table S2.
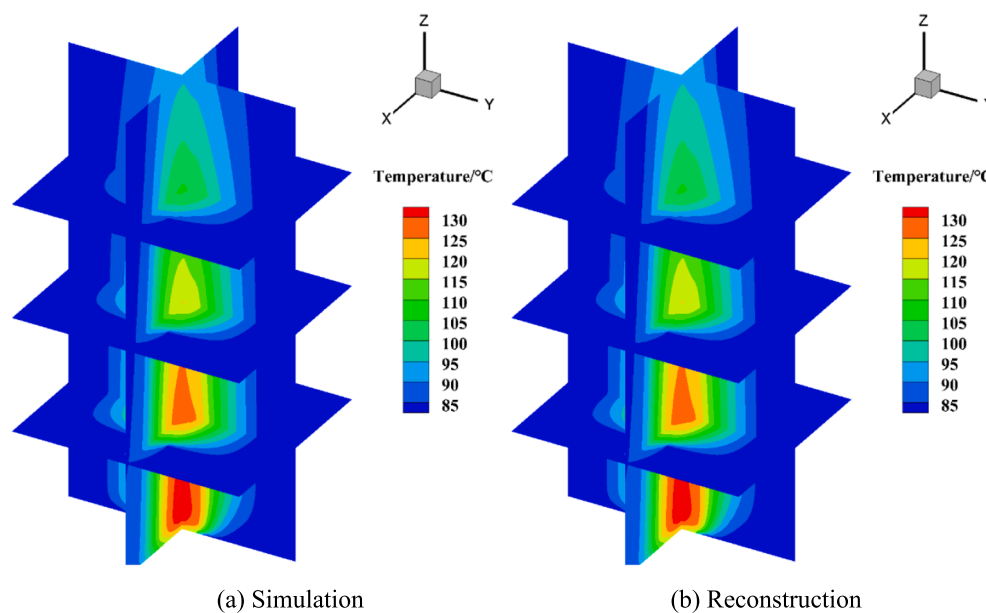
The local and global relative deviations for each physical field are shown in Figs. 13 (a) and (b) respectively. The mean deviations of the 20 cases for the output voltage and global voltage field are 5.8 % and 5.7 % respectively. Except the fourth case (the fourth red point in Fig. 13 (a), 16.8 %), the output voltage relative deviations under other conditions are all within 10.8 %. In Case 4, the working condition is disgusting. The stronger nonlinear characteristics lead to a lower digital twin accuracy. The output voltage relative deviations under Cases 5, 9, 17, 18 and 20

are about 9 %, larger than that under other conditions. This is due to relatively higher current densities. Furthermore, their working conditions are not so disgusting as Case 4. For example, their stoichiometric ratios are higher, leading to lower relative deviations. It should be noted that there are also several cases with large current densities but high prediction accuracy (about 2 %), including Cases 7, 11, 12 and 16. This may because that these cases have working conditions closer to the snapshot conditions.

The global relative deviation of the predicted temperature field under each working condition is within 3.9 % while the mean value is 1.3 %. The relative deviations of field maximum temperatures are within 10.3 %. The deviations of other physical fields are relatively higher. For the liquid water saturation field, the mean global relative deviation is 12.0 %. Except four cases with relative deviations of about 20 %, the global relative deviations of other cases are all within 15 %. For the

(a) Simulation　　　　　　　　(b) Reconstruction
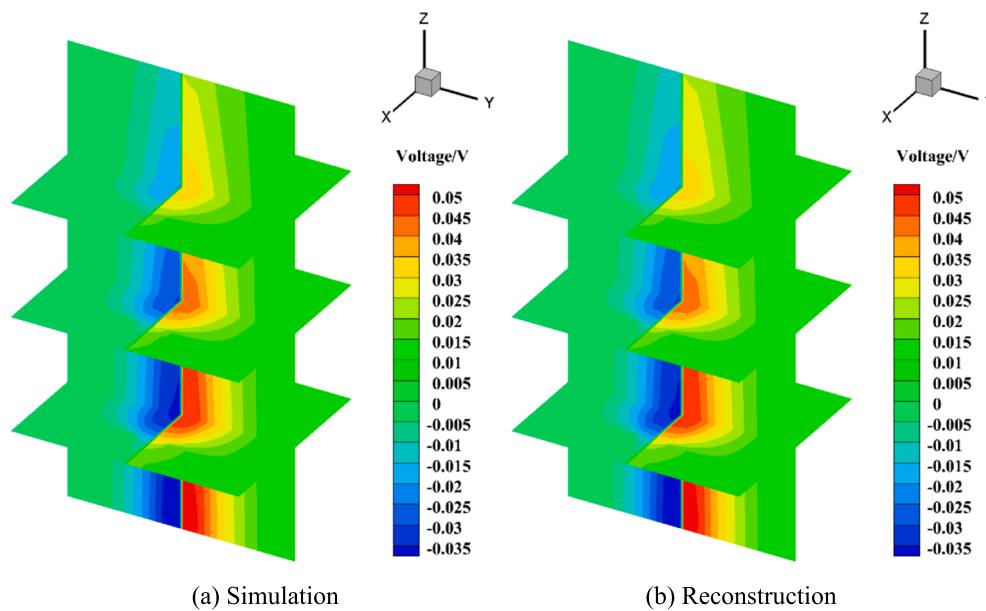
Fig. 10. Validation of voltage field reconstructed result for the 117th snapshot compared with the simulation result. (a) Simulation. (b) Reconstruction.

(a) Simulation　　　　　　　　(b) Reconstruction

Fig. 11. Validation of temperature field reconstructed result for the 117th snapshot compared with the simulation result. (a) Simulation. (b) Reconstruction.

membrane water content field, the global relative deviations for most cases are within 15 % and the mean global deviation is 8.9 %. In general, the global relative deviations of all the physical fields in the test set range within 0.11–24.01 %.

The reason why the digital twin deviations are higher than the deviations of the reconstructed results by POD are as follows. On the one hand, due to the large amount of computational workload, the snapshot number is limited. On the other hand, the coupling characteristics of physical fields in PEMFCs are complex, resulting in a regression difficulty for weight coefficients under random working conditions within wide ranges.

The 10th case is selected as an example to exhibit the contours of voltage, temperature and liquid water saturation fields obtained by digital twin technology compared with numerical results, as shown in Figs. 14-16 respectively. For voltage field, from Fig. 14 we can see that the digital twin voltage field is very close to the simulation one. The

digital twin results can capture the output voltage and voltage distribution information well. For temperature and liquid water saturation fields, from Fig. 15 and Fig. 16 we can see that the digital twin technology can predict the overall distribution regular patterns and various local distribution characteristics well, for instance, higher temperatures in the membrane under middle posterior segment of the channel, the jump phenomenon of liquid water saturation across the interface between two layers, the accumulation phenomenon of liquid water under the rib, etc. The effectiveness of the digital twin technology is validated. However, the digital twin deviation of the maximum temperature is relatively larger. This is because that temperature field is affected by various factors represented by irreversible heat, activation heat, ohmic heat and latent heat source term. Therefore, the accurate prediction of temperature field is more difficult compared with voltage field.

It is interesting to compare the prediction time required for the simulation and the digital twin. For the PEMFC studied in this paper, the
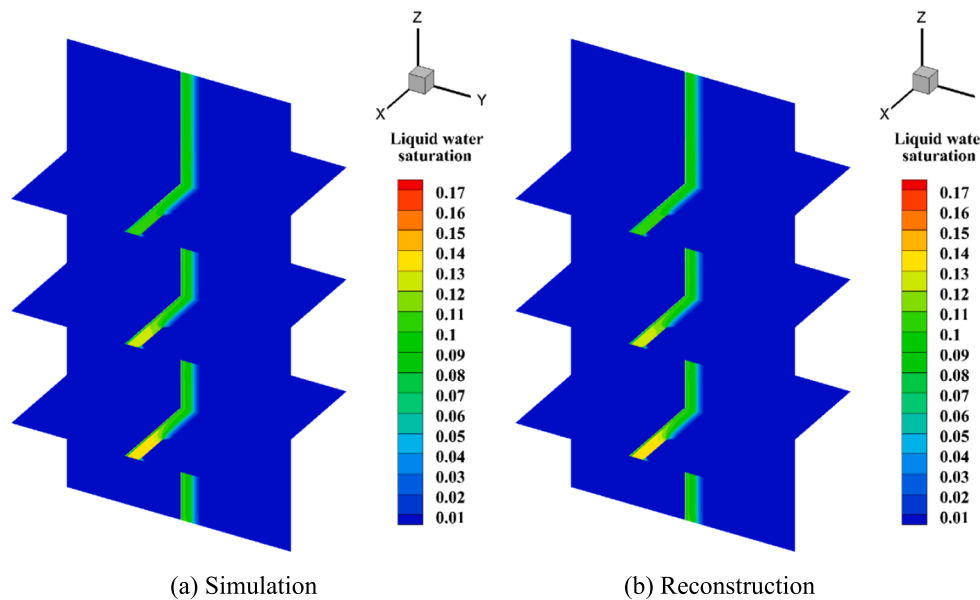
(a) Simulation

(b) Reconstruction

Fig. 12. Validation of liquid water saturation field reconstructed result for the 117th snapshot compared with the simulation result. (a) Simulation. (b) Reconstruction.
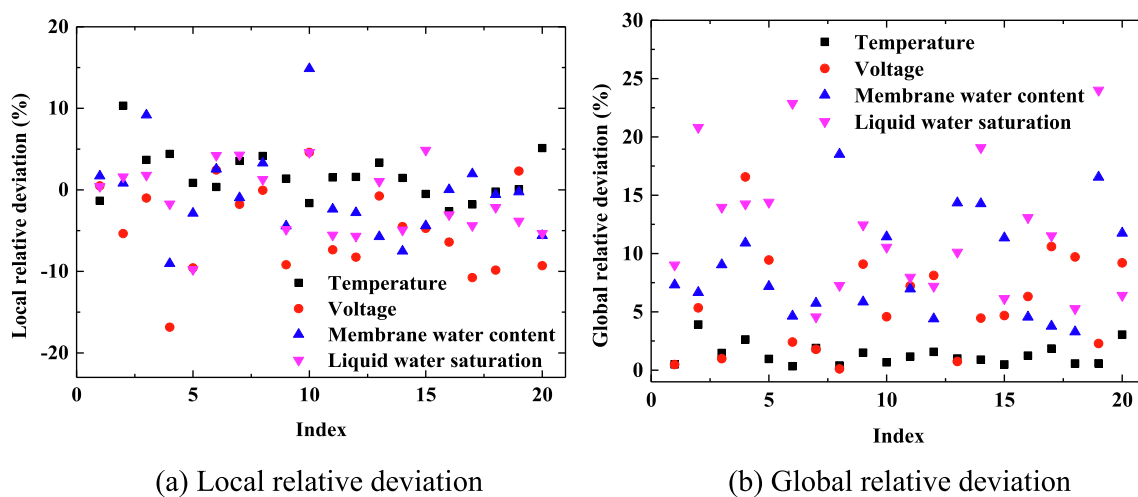

(a) Local relative deviation

(b) Global relative deviation

Fig. 13. Digital twin result relative deviations of each physical field for 20 cases of random working conditions compared with simulation results. (a) Local relative deviation. (b) Global relative deviation. (The local relative deviation of each physical field is defined the same as that in the reconstruction process inFigs. 6-9.).

simulation times of the snapshots differ a lot due to the different non-linear characteristics among snapshots. The simulation time is about several hours to days even by 16-core parallel server, Intel(R) Xeon(R) E5-2670 @ 2.60 GHz. As for the digital twin, the prediction times of different conditions are identical due to the same calculation complexity. In the digital twin technology, there are mainly five steps: (1) reading modes into Random Access Memory (RAM); (2) predicting weight coefficients by MARS; (3) constructing physical fields by Eq. (29); (4) calculating field local or average values; (5) outputting physical fields into the hard disk. For the prediction of one single physical field in one case, in average, the time consumptions of the five steps are 39.774 s, 0.018 s, 0.894 s, 0.001 s (neglected) and 2.088 s respectively by 1-core personal computer, Intel(R) Core(TM) i7-10750H @ 2.60 GHz. The entire time consumption is 42.775 s. Furthermore, in consecutive in-situ digital twins, step (1) only needs to be executed once in the initialization process. One single digital twin and output times can be regarded as 0.913 s and 2.088 s respectively.

Finally, it should be emphasized that the digital twin technology is a data-driven machine learning prediction method rather than an

accelerating convergence technique for CFD simulation. A large number of snapshots need to be conducted by ex-situ CFD which is also time-consuming. For example, the simulation of all the snapshots in Table S1 costs about two months. It is the quick response that is highly required for PEMFC in-situ prediction, which is at the expense of lots of ex-situ simulation time.

4. Conclusion

In this research, a complete 3D virtualization-multi-physics digital twin for PEMFCs is realized by POD technology. 139 snapshots are designed using Pairwise Independent Combinatorial Testing, and simulated based on the three-dimensional two-phase non-isothermal numerical model with the assumption of liquid pressure continuity. Modes for voltage, temperature, membrane water, liquid water pressure and liquid water saturation fields are extracted. 20 cases of in-situ off-design working conditions are randomly generated within entire current density range and the effectiveness of the digital twin technology is validated in the aspects of voltage, temperature, liquid water saturation
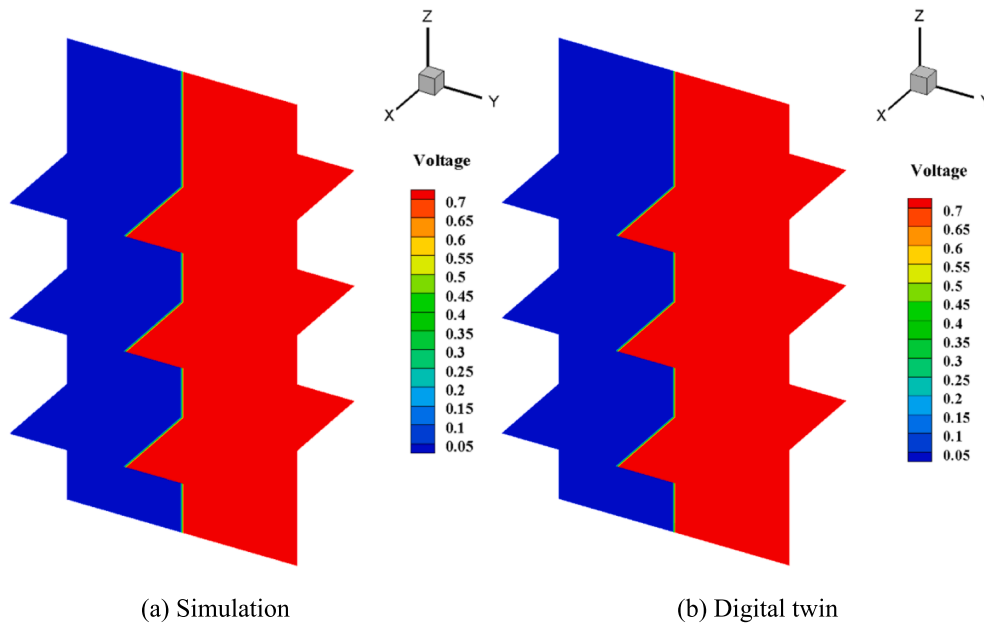
(a) Simulation   (b) Digital twin

Fig. 14. Validation of the voltage field digital twin result for the 10th case compared with the simulation result. (a) Simulation. (b) Digital twin.

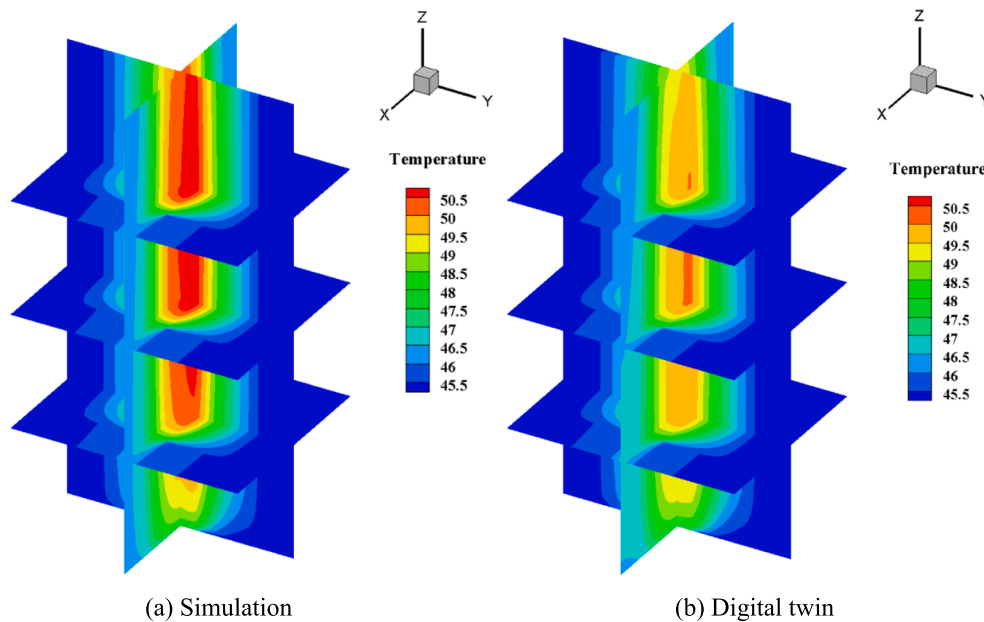(a) Simulation   (b) Digital twin

Fig. 15. Validation of the temperature field digital twin result for the 10th case compared with the simulation result. (a) Simulation. (b) Digital twin.

and membrane water content fields. The main conclusions of this paper are summarized as follows:

(1) The modes extracted from the snapshots are pairwise orthogonal with deviations of $5.991 \times 10^{-11} \sim 3.339 \times 10^{-10}$. The mode capturing maximum information represents the global magnitude of the field, while higher order modes represent the nonlinear or local characteristics of the field. The closer the correlation between two physical fields is, the more identical the information proportion curves of the modes are.

(2) Within the scope of the studied snapshots, only 24 modes are necessary for voltage field prediction to reach a truncation error of 0.01 %, while most of the modes should be applied for other physical fields.

(3) With the reduction of the truncation error, snapshot reconstruction relative deviations decrease significantly by several orders of magnitude. The reconstruction physical fields are nearly identical with the simulation results when truncation error equals to 0.01 %.

(4) Within the scope of the studied conditions and fields, the global deviations of digital twin technology range within 0.11–24.01 %. For output voltage, mean relative deviation is 5.8 % while most of the deviations are within 10.8 %. For temperature field, global relative deviations are within 3.9 %. For liquid water saturation field, the mean global relative deviation is 12.0 %. Except four cases with relative deviations of about 20 %, the global relative deviations under other cases are all within 15 %.

(5) For the fuel cell studied, the digital twin technology successfully captures the overall distribution regular patterns and various
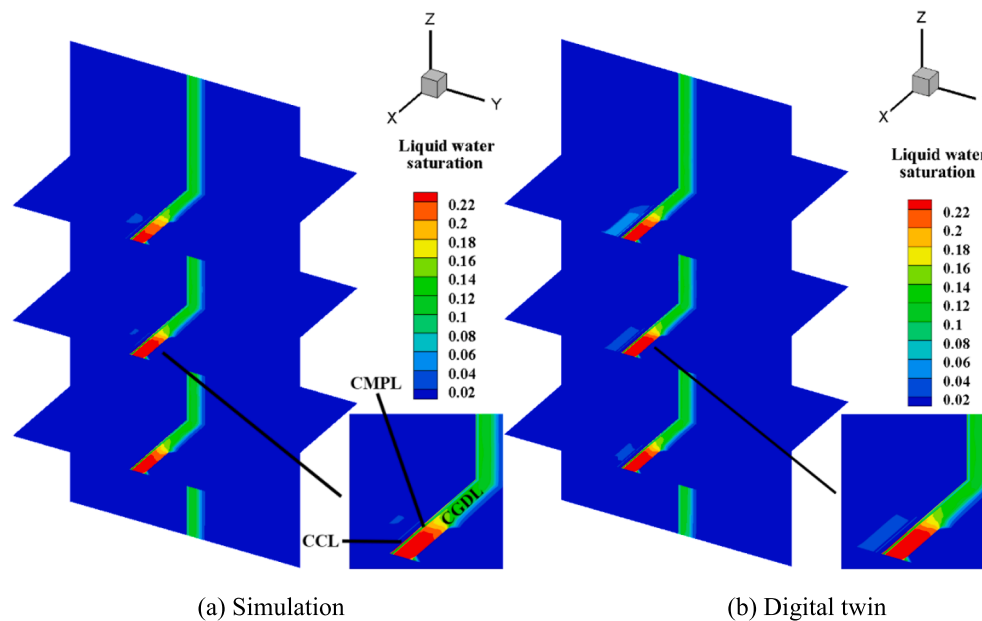
(a) Simulation   (b) Digital twin

Fig. 16. Validation of the liquid water saturation field digital twin result for the 10th case compared with the simulation result. (a) Simulation. (b) Digital twin.

local distribution characteristics in 0.913 s, which can realize the intention of in-situ operation control.

CRediT authorship contribution statement

Fan Bai: Investigation, Writing – original draft, Writing – review & editing, Visualization. **Hong-Bing Quan:** Investigation. **Ren-Jie Yin:** Investigation. **Zhuo Zhang:** Investigation. **Shu-Qi Jin:** Investigation. **Pu He:** Investigation. **Yu-Tong Mu:** Investigation. **Xiao-Ming Gong:** Investigation. **Wen-Quan Tao:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgment

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apenergy.2022.119763.

References

[1] Möhren S, Meyer J, Krause H, Saars L. A multiperiod approach for waste heat and renewable energy integration of industrial sites. Renew Sustain Energy Rev 2021; 148:111232. https://doi.org/10.1016/j.rser.2021.111232.

[2] Xi JP. An important speech delivered by Jin-Ping Xi at the general debate of the 75th UN General Assembly (In Chinese), http://www.gov.cn/xinwen/2020-09/22/content_5546168.htm; 2020 [accessed 22 September 2020].

[3] Gowrisankaran G, Reynolds SS, Samano M. Intermittency and the value of renewable energy. J Political Econ 2016;124(4):1187–234. https://doi.org/10.1086/686733.

[4] Suberu MY, Mustafa MW, Bashir N. Energy storage systems for renewable energy power sector integration and mitigation of intermittency. Renew Sustain Energy Rev 2014;35:499–514. https://doi.org/10.1016/j.rser.2014.04.009.

[5] Pei P, Li Y, Xu H, Wu Z. A review on water fault diagnosis of PEMFC associated with the pressure drop. Appl Energy 2016;173:366–85. https://doi.org/10.1016/j.apenergy.2016.04.064.

[6] Hua Z, Zheng Z, Péra M-C, Gao F. Remaining useful life prediction of PEMFC systems based on the multi-input echo state network. Appl Energy 2020;265: 114791. https://doi.org/10.1016/j.apenergy.2020.114791.

[7] Jiao K, Xuan J, Du Q, Bao ZM, Xie B, Wang B, et al. Designing the next generation of proton-exchange membrane fuel cells. Nature 2021;595(7867):361–9. https://www.nature.com/articles/s41586-021-03482-7.

[8] Bai F, Lei L, Zhang Z, Chen L, Chen L, Tao WQ. Application of similarity theory in the study of proton exchange membrane fuel cells: A comprehensive review of recent developments and future research requirements. Energy Storage Sav 2021;1 (1):3–21. https://doi.org/10.1016/j.enss.2021.09.001.

[9] Bai F, Lei Le, Zhang Z, Li H, Yan J, Chen Li, et al. Application of similarity theory in modeling the output characteristics of proton exchange membrane fuel cell. Int J Hydrogen Energy 2021;46(74):36940–53. https://doi.org/10.1016/j.ijhydene.2021.08.205.

[10] Lin R, Zhong Di, Lan S, Guo R, Ma Y, Cai X. Experimental validation for enhancement of PEMFC cold start performance: Based on the optimization of micro porous layer. Appl Energy 2021;300:117306. https://doi.org/10.1016/j.apenergy.2021.117306.

[11] Rahman MA, Sarker M, Mojica F, Chuang P-Y. A physics-based 1-D PEMFC model for simulating two-phase water transport in the electrode and gas diffusion media. Appl Energy 2022;316:119101. https://doi.org/10.1016/j.apenergy.2022.119101.

[12] Keller N, von Unwerth T. Advanced parametric model for analysis of the influence of channel cross section dimensions and clamping pressure on current density distribution in PEMFC. Appl Energy 2022;307:118132. https://doi.org/10.1016/j.apenergy.2021.118132.

[13] Eudy L, Post M, Jeffers M. Zero Emission Bay Area (ZEBA) Fuel Cell Bus Demonstration Results: Fifth Report. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2016.

[14] Eudy L, Post M, Jeffers M. American Fuel Cell Bus Project Evaluation: Third Report. National Renewable Energy Laboratory (US), 2017.

[15] Eudy L, Post M. Fuel cell buses in us transit fleets: Current status 2017. National Renewable Energy Laboratory (US), 2017.

[16] Construction of renewable energy demonstration zone in Zhangjiakou Hebei - "hydrogen" dance, zero carbon emission (In Chinese). Guangxi Energy Conserv 2021;03:27.

[17] Amy A. Railway Gazette International. Cummins to build railway fuel cell factory, https://www.railwaygazette.com/business/cummins-to-build-railway-fuel-cell-factory/57781.article?adredir=1; 2020 [accessed 13 November 2020].

[18] Toshiba Energy Systems & Solutions Corporation Kawasaki Heavy Industries, Ltd. Demonstration Project Begins for Commercialization of Vessels Equipped with High-power Fuel Cells, https://www.nyk.com/english/news/2020/20200901_01.html; 2020 [accessed 01 September 2020].

[19] Wang Y, Basu S, Wang CY. Modeling two-phase flow in PEM fuel cell channels. J Power Sources 2008;179(2):603–17. https://doi.org/10.1016/j.jpowsour.2008.01.047.

[20] Lim BH, Majlan EH, Daud WRW, Husaini T, Rosli MI. Effects of flow field design on water management and reactant distribution in PEMFC: a review. Ionics 2016;22 (3):301–16. https://doi.org/10.1007/s11581-016-1644-y.

[21] Rubio GA, Agila WE. A Fuzzy Model to Manage Water in Polymer Electrolyte Membrane Fuel Cells. Processes 2021;9(6):904. https://doi.org/10.3390/pr9060904.

[22] Li Y, Yang J, Song J. Structure models and nano energy system design for proton exchange membrane fuel cells in electric energy vehicles. Renew Sustain Energy Rev 2017;67:160–72. https://doi.org/10.1016/j.rser.2016.09.030.

[23] Wang B, Zhang G, Wang H, Xuan J, Jiao K. Multi-physics-resolved digital twin of proton exchange membrane fuel cells with a data-driven surrogate model. Energy and AI 2020;1:100004. https://doi.org/10.1016/j.egyai.2020.100004.

[24] Bernardi DM, Verbrugge MW. Mathematical model of a gas diffusion electrode bonded to a polymer electrolyte. AIChE J 1991;37(8):1151–63. https://doi.org/10.1002/aic.690370805.

[25] Bernardi DM, Verbrugge MW. A mathematical model of the solid-polymer-electrolyte fuel cell. J Electrochem Soc 1992;139(9):2477–91. https://doi.org/10.1149/1.2221251.

[26] Springer TE, Zawodzinski TA, Gottesfeld S. Polymer electrolyte fuel cell model. J Electrochem Soc 1991;138(8):2334–42. https://doi.org/10.1149/1.2085971.

[27] Bao C, Bessler WG. Two-dimensional modeling of a polymer electrolyte membrane fuel cell with long flow channel. Part I. Model development. J Power Sources 2015;275:922–34. https://doi.org/10.1016/j.jpowsour.2014.11.058.

[28] Bao C, Bessler WG. Two-dimensional modeling of a polymer electrolyte membrane fuel cell with long flow channel. Part II. Physics-based electrochemical impedance analysis. J Power Sources 2015;278:675–82. https://doi.org/10.1016/j.jpowsour.2014.12.045.

[29] Goshtasbi A, Pence BL, Ersal T. Computationally efficient pseudo-2D non-isothermal modeling of polymer electrolyte membrane fuel cells with two-phase phenomena. J Electrochem Soc 2016;163(13):F1412–32. https://doi.org/10.1149/2.0871613jes.

[30] Amphlett JC, Baumert RM, Mann RF, Peppley BA, Roberge PR, Harris TJ. Performance modeling of the Ballard Mark IV solid polymer electrolyte fuel cell: I. Mechanistic model development. J Electrochem Soc 1995;142(1):1–8. https://doi.org/10.1149/1.2043866.

[31] Amphlett JC, Baumert RM, Mann RF, Peppley BA, Roberge PR, Harris TJ. Performance modeling of the Ballard Mark IV solid polymer electrolyte fuel cell: II. Empirical model development. J Electrochem Soc 1995;142(1):9–15. https://doi.org/10.1149/1.2043959.

[32] Larminie J, Dicks A, McDonald MS. Fuel Cell Systems Explained. 2rd ed. Chichester, UK: John Wiley & Sons; 2003. https://sv.20file.org/up1/482_0.pdf.

[33] Dicks AL, Rand DAJ. Fuel Cell Systems Explained. 3rd ed. Glasgow: John Wiley & Sons; 2018. https://www.wiley.com/en-us/Fuel+Cell+Systems+Explained%2C+3rd+Edition-p-9781118706961.

[34] Kim J, Lee SM, Srinivasan S, Chamberlin CE. Modeling of proton exchange membrane fuel cell performance with an empirical equation. J Electrochem Soc 1995;142(8):2670–4. https://doi.org/10.1149/1.2050072.

[35] Saadi A, Becherif M, Aboubou A, Ayad MY. Comparison of proton exchange membrane fuel cell static models. Renew Energy 2013;56:64–71. https://doi.org/10.1016/j.renene.2012.10.012.

[36] Moreira MV, Da Silva GE. A practical model for evaluating the performance of proton exchange membrane fuel cells. Renew Energy 2009;34(7):1734–41. https://doi.org/10.1016/j.renene.2009.01.002.

[37] Mammar K, Laribi S. Application of adaptive neuro-fuzzy inference system techniques to predict water activity in proton exchange membrane fuel cell. J Electrochem Energy Convers Storage 2018;15:041009. https://doi.org/10.1115/1.4040058.

[38] Mammar K, Saadaoui F, Laribi S. Design of a PEM fuel cell model for flooding and drying diagnosis using fuzzy logic clustering. Renew Energy Focus 2019;30:123–30. https://doi.org/10.1016/j.ref.2019.06.001.

[39] Jemeï S, Hissel D, Péra MC, Kauffmann M. On-board fuel cell power supply modeling on the basis of neural network methodology. J Power Sources 2003;124 (2):479–86. https://doi.org/10.1016/S0378-7753(03)00799-7.

[40] Nanadegani FS, Lay EN, Iranzo A, Salva JA, Sunden B. On neural network modeling to maximize the power output of PEMFCs. Electrochim Acta 2020;348:136345. https://doi.org/10.1016/j.electacta.2020.136345.

[41] Tian Y, Zou Q, Han J. Data-Driven Fault Diagnosis for Automotive PEMFC Systems Based on the Steady-State Identification. Energies 2021;14(7):1918. https://doi.org/10.3390/en14071918.

[42] Derbeli M, Napole C, Barambones O. Machine learning approach for modeling and control of a commercial heliocentris FC50 PEM fuel cell system. Math 2021;9(17):2068. https://doi.org/10.3390/math9172068.

[43] Sivertsen BR, Djilali N. CFD-based modelling of proton exchange membrane fuel cells. J Power Sources 2005;141(1):65–78. https://doi.org/10.1016/j.jpowsour.2004.08.054.

[44] Tao WQ, Min CH, Liu XL, He YL, Yin BH, Jiang W. Parameter sensitivity examination and discussion of PEM fuel cell simulation model validation: Part I. Current status of modeling research and model development. J Power Sources 2006;160(1):359–73. https://doi.org/10.1016/j.jpowsour.2006.01.078.

[45] Min CH, He YL, Liu XL, Yin BH, Jiang W, Tao WQ. Parameter sensitivity examination and discussion of PEM fuel cell simulation model validation: Part II: Results of sensitivity analysis and validation of the model. J Power Sources 2006;160(1):374–85. https://doi.org/10.1016/j.jpowsour.2006.01.080.

[46] Wu HW, Ku HW. The optimal parameters estimation for rectangular cylinders installed transversely in the flow channel of PEMFC from a three-dimensional PEMFC model and the Taguchi method. Appl Energy 2011;88(12):4879–90. https://doi.org/10.1016/j.apenergy.2011.06.034.

[47] Kulikovsky AA. A physically–based analytical polarization curve of a PEM fuel cell. J Electrochem Soc 2013;161(3):F263. https://doi.org/10.1149/2.028403jes.

[48] Jiao K, Bachman J, Zhou Y, Park JW. Effect of induced cross flow on flow pattern and performance of proton exchange membrane fuel cell. Appl Energy 2014;115:75–82. https://doi.org/10.1016/j.apenergy.2013.10.026.

[49] Kang S. Quasi-three dimensional dynamic modeling of a proton exchange membrane fuel cell with consideration of two-phase water transport through a gas diffusion layer. Energy 2015;90:1388–400. https://doi.org/10.1016/j.energy.2015.06.076.

[50] Perng SW, Wu HW. A three-dimensional numerical investigation of trapezoid baffles effect on non-isothermal reactant transport and cell net power in a PEMFC. Appl Energy 2015;143:81–95. https://doi.org/10.1016/j.apenergy.2014.12.059.

[51] Jeon SW, Cha D, Kim HS, Kim Y. Analysis of the system efficiency of an intermediate temperature proton exchange membrane fuel cell at elevated temperature and relative humidity conditions. Appl Energy 2016;166:165–73. https://doi.org/10.1016/j.apenergy.2015.12.123.

[52] Hu J, Li J, Xu L, Huang F, Ouyang M. Analytical calculation and evaluation of water transport through a proton exchange membrane fuel cell based on a one-dimensional model. Energy 2016;111:869–83. https://doi.org/10.1016/j.energy.2016.06.020.

[53] Jahnke T, Futter G, Latz A, Malkow T, Papakonstantinou G, Tsotridis G, et al. Performance and degradation of Proton Exchange Membrane Fuel Cells: State of the art in modeling from atomistic to system scale. J Power Sources 2016;304:207–33. https://doi.org/10.1016/j.jpowsour.2015.11.041.

[54] Li W, Zhang Q, Wang C, Yan X, Shen S, Xia G, et al. Experimental and numerical analysis of a three-dimensional flow field for PEMFCs. Appl Energy 2017;195:278–88. https://doi.org/10.1016/j.apenergy.2017.03.008.

[55] Liu JX, Guo H, Ye F, Ma CF. Two-dimensional analytical model of a proton exchange membrane fuel cell. Energy 2017;119:299–308. https://doi.org/10.1016/j.energy.2016.12.075.

[56] Fan L, Zhang G, Jiao K. Characteristics of PEMFC operating at high current density with low external humidification. Energy Convers Manage 2017;150:763–74. https://doi.org/10.1016/j.enconman.2017.08.034.

[57] He P, Mu YT, Park JW, Tao WQ. Modeling of the effects of cathode catalyst layer design parameters on performance of polymer electrolyte membrane fuel cell. Appl Energy 2020;277:115555. https://doi.org/10.1016/j.apenergy.2020.115555.

[58] Xia L, Ni M, Xu Q, Xu H, Zheng K. Optimization of catalyst layer thickness for achieving high performance and low cost of high temperature proton exchange membrane fuel cell. Appl Energy 2021;294:117012. https://doi.org/10.1016/j.apenergy.2021.117012.

[59] Tuegel EJ, Ingraffea AR, Eason TG, Spottswood SM. Reengineering aircraft structural life prediction using a digtial twin. Int J Aerosp Eng 2011;2011:154798. https://doi.org/10.1155/2011/154798.

[60] Wu J, Yang Y, Cheng XUN, Zuo H, Cheng Z. The development of digital twin technology review. In: 2020 Chinese Automation Congress (CAC). IEEE; 2020. p. 4901–6. https://doi.org/10.1109/CAC51589.2020.9327756.

[61] Tao F, Qi Q. Make more digital twins. Nature 2019;573:490–1. https://doi.org/10.1038/d41586-019-02849-1.

[62] Meraghni S, Terrissa LS, Yue M, Ma J, Jemei S, Zerhouni N. A data-driven digital-twin prognostics method for proton exchange membrane fuel cell remaining useful life prediction. Int J Hydrogen Energy 2021;46(2):2555–64. https://doi.org/10.1016/j.ijhydene.2020.10.108.

[63] Zhao J, Zhu J. A digital twin approach for fault diagnosis in PEM fuel cell systems. In: 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI); 2021. p. 168–71. https://doi.org/10.1109/DTPI52967.2021.9540157.

[64] Bartolucci L, Cennamo E, Cordiner S, Mulone V, Pasqualini F, Boot MA. Digital Twin of Fuel Cell Hybrid Electric Vehicle: a detailed modelling approach of the hydrogen powertrain and the auxiliary systems. In: E3S Web of Conferences EDP. Sciences, vol. 334; 2022. p. 06003. https://doi.org/10.1051/e3sconf/202233406003.

[65] Ding P, Wu XH, He YL, Tao WQ. A fast and efficient method for predicting fluid flow and heat transfer problems. J Heat Transfer 2008;130(3):032502. https://doi.org/10.1115/1.2804935.

[66] Berkooz G, Holmes P, Lumley JL. The proper orthogonal decomposition in the analysis of turbulent flows. Annu Rev Fluid Mech 1993;25(1):539–75. https://doi.org/10.1146/annurev.fl.25.010193.002543.

[67] Liang YC, Lee HP, Lim SP, Lin WZ, Lee KH, Wu CG. Proper orthogonal decomposition and its applications—Part I: Theory. J Sound Vib 2002;252(3):527–44. https://doi.org/10.1006/jsvi.2001.4041.

[68] Liang YC, Lin WZ, Lee HP, Lim SP, Lee KH, Sun H. Proper orthogonal decomposition and its applications–part II: Model reduction for MEMS dynamical analysis. J Sound Vib 2002;256(3):515–32. https://doi.org/10.1006/jsvi.2002.5007.

[69] Kammerer AJ, Hackett EE. Use of proper orthogonal decomposition for extraction of ocean surface wave fields from X-band radar measurements of the sea surface. Remote Sens 2017;9(9):881. https://doi.org/10.3390/rs9090881.

[70] Wang J, Zhang L, Guan J, Zhang M. Evaluation of Combined Satellite and Radar Data Assimilation with POD-4DEnVar Method on Rainfall Forecast. Appl Sci 2020;10(16):5493. https://doi.org/10.3390/app10165493.

[71] Park HM, Cho DH. The use of the Karhunen-Loeve decomposition for the modeling of distributed parameter systems. Chem Eng Sci 1996;51(1):81–98. https://doi.org/10.1016/0009-2509(95)00230-8.

[72] LeGresley PA. Application of proper orthogonal decomposition (POD) to design decomposition methods. to design decomposition methods. Stanford University; 2006. https://www.proquest.com/pagepdf/304977522?accountid=16294.

[73] Mukerjee R, Wu CF. A modern theory of factorial design. New York: Springer; 2006. https://link.springer.com/book/10.1007/0-387-37344-6.

[74] Liu RJ, Zhang YW, Wen CW, Tang J. Study on the design and analysis methods of orthogonal experiment (In Chinese). Exp Technol Manage 2010;9:52–5. http://doi.org/10.16791/j.cnki.sjg.2010.09.016.

[75] Tai KC, Lei Y. A test generation strategy for pairwise testing. IEEE Trans Softw Eng 2002;28(1):109–11. https://doi.org/10.1109/32.979992.

[76] Fang KT, Lin DK, Winker P, Zhang Y. Uniform design: theory and application. Technometrics 2000;42(3):237–48. https://doi.org/10.1080/00401706.2000.10486045.

[77] 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. https://doi.org/10.1016/j.rser.2016.01.040.

[78] Moonen M, De Moor B. SVD and Signal Processing, III: Algorithms, Architectures and Applications. Netherlands: Elsevier; 1995. https://www.elsevier.com/books/svd-and-signal-processing-iii/moonen/978-0-444-82107-2.

[79] Businger PA, Golub GH. Algorithm 358: Singular value decomposition of a complex matrix [f1, 4, 5]. Commun ACM 1969;12(10):562–6. https://doi.org/10.1145/363235.363249.

[80] Fernando KV, Parlett BN. Accurate singular values and differential qd algorithms. Numer Math 1994;67(2):191–229. https://doi.org/10.1007/s002110050024.

[81] Demmel J, Veselić K. Jacobi's method is more accurate than QR. SIAM J on Matrix Anal Appl 1992;13(4):1204–45. https://doi.org/10.1137/0613074.

[82] Ravindran SS. Reduced-order adaptive controllers for fluid flows using POD. J Sci comput 2000;15(4):457–78. https://doi.org/10.1023/A:1011184714898.

[83] Friedman JH. Multivariate adaptive regression splines. Ann Statist 1991;19(1):1–67. https://doi.org/10.1214/aos/1176347963.