**AI-based optimization of PEM fuel cell catalyst layers for
maximum power density via data-driven surrogate modeling**

Bowen Wang^a, Biao Xie^a, 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, United Kingdom

*Corresponding author: Email: j.xuan@lboro.ac.uk; kjiao@tju.edu.cn

Abstract

Catalyst layer (CL) is the core electrochemical reaction region of proton exchange membrane fuel cells (PEMFCs). Its composition directly determines PEMFC output performance. Existing experimental or modeling methods are still insufficient on the deep optimization of CL composition. This work develops a novel artificial intelligence (AI) framework combining a data-driven surrogate model and a stochastic optimization algorithm to achieve multi-variables global optimization for improving the maximum power density of PEMFCs. Simulation results of a three-dimensional computational fluid dynamics (CFD) PEMFC model coupled with the CL agglomerate model constitutes the database, which is then used to train the data-driven surrogate model based on Support Vector Machine (SVM), a typical AI algorithm. Prediction performance shows that the squared correlation coefficient (R-square) and mean percentage error in the test set are 0.9908 and 3.3375%, respectively. The surrogate model has demonstrated comparable accuracy to the physical model, but with much greater computation-resource efficiency: the

calculation of one polarization curve will be within one second by the surrogate model, while it may cost hundreds of processor-hours by the physical CFD model. The surrogate model is then fed into a Genetic Algorithm (GA) to obtain the optimal solution of CL composition. For verification, the optimal CL composition is returned to the physical model, and the percentage error between the surrogate model predicted and physical model simulated maximum power densities under the optimal CL composition is only 1.3950%. The results indicate that the proposed framework can guide the multi-variables optimization of complex systems.

Keywords:

proton exchange membrane fuel cell; catalyst layer composition; agglomerate model; data-driven surrogate model; stochastic optimization algorithm.

Nomenclature

CL	Catalyst layer
EW	Equivalent weight of ionomer(kg mol^{-1})
I	Current density (A cm^{-2})
m_C	Carbon particles loading (mg cm^{-2})
m_{Im}	Ionomer loading (mg cm^{-2})
m_{Pt}	Pt loading (mg cm^{-2})
$m_{\text{Pt,a}}$	Pt loading of anode CL (mg cm^{-2})
$m_{\text{Pt,c}}$	Pt loading of cathode CL (mg cm^{-2})
$m_{\text{Pt-C}}$	Carbon-supported Pt loading (mg cm^{-2})

M_w	Molecular weight of liquid water (kg mol ⁻¹)
S	Function
V	Voltage (V)
$\zeta_{Pt/C}$	Pt percentage of carbon-supported Pt
$\zeta_{Pt/C,a}$	Pt percentage of carbon-supported Pt of anode CL
$\zeta_{Pt/C,c}$	Pt percentage of carbon-supported Pt of cathode CL
$\zeta_{Im/C}$	Ionomer to carbon ratio
$\zeta_{Im/C,a}$	Ionomer to carbon ratio of anode CL
$\zeta_{Im/C,c}$	Ionomer to carbon ratio of cathode CL
ε_s	Volume fractions of solid (carbon and platinum)
ε_{CL}	Porosity of CLs
ω	Volume fractions of ionomer
δ_{CL}	Thickness of CLs (m)
λ	Water content
ρ_{Pt}	Density of platinum(kg m ⁻³)
ρ_C	Density of carbon(kg m ⁻³)
ρ_{Im}	Density of ionomer (kg m ⁻³)
ρ_w	Density of liquid water (kg m ⁻³)

34 1. Introduction

35 As an efficient and clean energy conversion device, proton exchange membrane
36 fuel cell (PEMFC) is frequently highlighted in the worldwide energy applications in

recent years [1-3]. PEMFC technology made remarkable progress in the past decade, and its commercialization is also gradually realized [4-6]. However, there still exist challenges before the large-scale commercialization, including the high cost, insufficient durability and low output performance of the present PEMFC [7-9].

Catalyst layer (CL) is the region where electrochemical reactions take place, and output performance of PEMFCs is closely related to the structure of CLs. CLs simultaneously undertake functions of catalyzing electrochemical reactions, conducting protons and electrons, transporting gas reactants and removing product water, and thus CLs should usually consist of three phases, i.e. solid phase for catalysis and electron transport, electrolyte phase for proton transport and pore phase (void region) for mass transport [10]. The common solid and electrolyte phase materials at present are carbon-supported platinum (Pt) particles and ionomer, respectively [11]. The loading and composition ratio of these materials strongly determine the structure and performance of CLs. Due to the randomness and uncontrollability of CL structural parameters in the CL preparation, optimization of CL composition for improving PEMFC output performance is primarily demanded.

To optimize the CL, many researchers investigated the effects of CL composition on the structure and performance of CLs via experiments [12-14]. Zhao et al. [12] measured and compared the effective diffusion coefficient of electrodes in four different Pt loadings (i.e. 0.1, 0.2, 0.3, and 0.4 mg cm⁻²) via the modified Loschmidt cell. The results showed that a higher Pt loading would cause smaller porosity and

mean pore size, and therefore decreases the effective diffusion coefficient generally. Shahgald et al. [13] investigated the kinetics and mass transport effects of CLs and PEMFC overall performance by using two catalysts (20% and 60% Pt to carbon ratio). The results showed that the CL of 20% Pt to carbon ratio catalyst had better kinetics than that of 60% Pt to carbon ratio catalyst due to the less Pt agglomeration, better Pt dispersion and higher active area. However, for the same Pt loading, the CL of 20% Pt to carbon ratio catalyst had higher diffusion resistivity than that of 60% Pt to carbon ratio catalyst which resulted from the difference of CL thickness and effective diffusion coefficient. Overall performance of the PEMFC using 60% Pt to carbon ratio catalyst was better than that using 20% Pt to carbon ratio catalyst. Generally, experiments focused on the observation of transport mechanism inside CLs based on the existing catalysts which could provide very limited guidance on CL composition optimization.

For simulation works, pore-scale models were presented to understand the transport process in CLs and therefore optimize CL composition [15-17]. Deng et al. [16] conducted the pore-scale investigation of CLs via a three-dimensional (3D) single-phase lattice Boltzmann model and suggested that increasing the ionomer content and Pt loading would increase the active area, but meanwhile increase oxygen transport resistivity and limit the utilization rate of catalyst. Hou et al. [17] also developed a pore-scale model of CLs via the lattice Boltzmann method and compared the trade-off effect between the active area and oxygen transport resistivity under

79 different Pt to catalyst ratios and ionomer to catalyst ratios. They proposed the
80 sufficient active area had priority for improving CL performance. Besides the
81 pore-scale model, the full cell model was also used to predict PEMFC overall
82 performance and internal multi-physics under different CL parameters. In some
83 modeling work, the CL is considered as homogenous structure which is described by
84 specified volume fractions of the three phases, and this kind of CL model is called as
85 the homogenous model. The CL homogenous model does not contain the most critical
86 composition parameters of catalyst and ionomer, and it generally underestimates the
87 oxygen transport resistivity in the cathode CL [18]. To better reflect the structure of
88 CLs and performance of PEMFCs, the CL agglomerate model which considers
89 spherical agglomerates of catalyst particles coated by ionomer film is proposed and
90 coupled in the macroscopic full cell model. Plenty of computational fluid dynamics
91 (CFD) full cell model of PEMFC coupled with the CL agglomerate model was
92 developed and used to investigate the effects of CL composition on PEMFC
93 performance, and thus presented the optimal solution [19-23]. However, conducting
94 these simulation, including pore-scale models and macroscopic models, cost much
95 time and computational resources, and thus the corresponding investigations were
96 generally based on the single-variable optimization. In summary, there still exist
97 many limitations on the global optimization of multi-variables of CL composition via
98 the existing experimental and numerical methods.

As one of the most exciting technologies in the 21st century, artificial intelligence (AI) can be widely integrated with most fields to solve the problems, and the “AI Plus” approach has shown great potential to achieve revolutionary progresses for technology development [24-26]. To address the challenge of the CL optimization, here we propose an AI-based, data-driven surrogate modeling and optimization framework to achieve the multi-variables global optimization of CL composition. Firstly, a state-of-the-art 3D CFD full cell model coupled with the CL agglomerate model is employed as the basic of the surrogate model, which can effectively simulate PEMFC output performance under different CL composition. The data-driven surrogate model is developed based on Support Vector Machine (SVM), a typical AI algorithm, and then trained by the simulation results from the physical CFD model. Finally, the validated surrogate model with much higher computation-resource efficiency is implemented in the Genetic Algorithm (GA) to optimize CL composition for improving PEMFC output performance.

2. Methods

2.1 Optimization objective and Framework

Figure 1 shows a typical polarization curve of PEMFC from our simulation results, including the output voltage and power density, and there exists a maximum power density in one polarization curve. Generally, the polarization curve and maximum power density are the main factors to measure output performance of a

119 PEMFC. In this work, the maximum power density is chosen as the objective to be
120 optimized. The framework of this work can be summarized as the following steps:

- 121 1. Develop and implement the physical model, i.e. the 3D CFD full cell model
122 coupled with the CL agglomerate model to simulate current densities under different
123 output voltage and CL composition, and the simulation results constitute the database.
- 124 2. Use the database to train the surrogate model through the data-driven method
125 and test the validity of the developed surrogate model. The well-validated surrogate
126 model can therefore replace the physical model and efficiently predict the current
127 density in a new case. Power density can be calculated by output voltage and the
128 predicted current density, and there exists a maximum power density in one case of
129 CL composition.
- 130 3. Search the optimal CL composition for maximizing the maximum power
131 density by using the surrogate model and stochastic optimization algorithm.
- 132 4. Verify the optimal solution through comparing the power densities under the
133 optimal CL composition by surrogate model prediction and physical model
134 verification.

Figure 1 Polarization characteristic and maximum power density of PEMFCs.

To demonstrate the surrogate modeling and optimization procedure, a case study of a specific PEMFC is presented in the paper. Figure 2(a) shows the schematic of the simulated PEMFC. The activation area is 50mm × 6mm. Anode and cathode flow fields are parallel and serpentine, respectively, and the flow arrangement is counter-flow. The electrode of each side consists of the gas diffusion layer (GDL), micro-porous layer (MPL) and CL. Detailed properties of the simulated PEMFC are listed in Table 1.

Figure 2 Schematic of (a) the simulated PEMFC and computational domain of the physical model and (b) the CL single agglomerate.

2.2 Physical model and Data generation

A previously developed 3D CFD full cell model coupled with the CL agglomerate model is employed in this work [27]. The conservation equations of the model are listed in Table 2 [27]. The schematic of the CL single agglomerate is shown in Figure 2(b), and the CL composition is represented by three parameters, the Pt loading (m_{Pt} , mg cm⁻²), the Pt percentage of carbon-supported Pt (Pt percentage, $\zeta_{Pt/C}$), and the ionomer to carbon ratio (IC ratio, $\zeta_{Im/C}$) which are expressed:

$$\zeta_{Pt/C} = \frac{m_{Pt}}{m_{Pt-C}} \quad (1)$$

$$\zeta_{Im/C} = \frac{m_{Im}}{m_C} \quad (2)$$

where m_{Pt-C} (mg cm⁻²) is the carbon-supported Pt loading; m_{Im} (mg cm⁻²) is the ionomer loading and m_C (mg cm⁻²) is the carbon particles loading. And volume

fractions of the three phases in CLs are also determined by CL composition

parameters:

$$\varepsilon_s = \frac{m_{Pt}}{\delta_{CL}} \left(\frac{1}{\rho_{Pt}} + \left(\frac{1}{\zeta_{Pt/C}} - 1 \right) \frac{1}{\rho_C} \right) \quad (3)$$

$$\omega = \frac{\zeta_{Im/C} m_{Pt}}{\delta_{CL} \rho_{Im}} \left(\frac{1}{\zeta_{Pt/C}} - 1 \right) \left(1 + \frac{M_w \rho_{Im}}{\rho_w EW} \lambda \right) \quad (4)$$

$$\varepsilon_{CL} = 1 - \varepsilon_s - \omega \quad (5)$$

where ε_s , ω and ε_{CL} are the volume fractions of solid (carbon and platinum), electrolyte (ionomer) and pores (porosity), respectively; δ_{CL} (m) is the thickness of CLs; ρ_{Pt} , ρ_C , ρ_{Im} and ρ_w (kg m⁻³) are the densities of platinum, carbon, ionomer and liquid water, respectively; M_w (kg mol⁻¹) is the molecular weight of liquid water; EW (kg mol⁻¹) is the equivalent weight of ionomer and λ is the water content. Our previous work compared polarization curves in different Pt loadings predicted by the present model with that measured by the experiments, and a good agreement was achieved, and the comparison result is shown in Figure 3 [27,28]. The computational domain is shown in Figure 2(a), and the grid numbers along the X, Y and Z directions are 70, 79 and 60, respectively. A commercial CFD software, Fluent, is used for the numerical computation of the model. Other details of the modeling processes can be found in our previous work [27].

Figure 3 Comparison of polarization curves between simulation results by the present physical model [27] and experimental results [28] under different cathode Pt loadings.

Constant voltage mode is used in the simulation and the current density can be therefore predicted. To construct the database, we randomly generate 65 cases of different output voltage and anode and cathode CL composition which are simulated by the physical model. To guarantee the physical meaning of CLs, like volume fractions of the three phases in CLs must be more than zero and less than 1.0 and a too low fraction of any phase is obviously bad for PEMFC performance and computational stability, the restrictive ranges of output voltage, Pt loading, Pt percentage, IC ratio and volume fractions of the three phases, in the case generation are given in Table 3 [23,27]. And the given ranges generally fit the common values of these parameters in the previous literature. Also note that this database is only valid for the optimization of this PEMFC.

2.3 Surrogate model development

In this section, a machine learning algorithm is introduced for development of the surrogate model. Support vector machine (SVM) is a widely-used machine learning algorithm which is based on Vapnik-Chervonenkis dimension and structural risk minimization in the statistical theory. SVM aims to construct the hyperplane in a high dimensional space to maximize the distinction of sample points (classification) or minimize the fitting of sample points (regression). The advantage of SVM is that it has good compromise between model complexity (prediction accuracy) and generalization by using a limited sample size which well fits our current work. Therefore, SVM is selected as the machine learning algorithm to develop the surrogate model.

The whole database is randomly split into the training set and test set which are used to develop the model and assess the generalization of the developed model, respectively. Note that the test set is never used in model development. Surrogate model development is implemented by LIBSVM (developed by Nation Taiwan University) in MATLAB [29]. Several formulations are provided by LIBSVM for classification and regression, and ε -SVR for regression is used in this work. We choose radial basis function (RBF) kernel as the kernel function through the preliminary test. The cost parameter (c) and gamma in RBF kernel (g) are the two hyperparameters remained to be optimized in modeling development, and thus five-fold cross-validation is conducted to search the optimal values of c and g for

hyperparameters tuning. More details about SVM and implementation procedure of LIBSVM can be found in Reference [29]. The developed surrogate model constructs the relationship between input and output variables which can be expressed as:

$$I = S(V, m_{\text{Pt,a}}, \zeta_{\text{Pt/C,a}}, \zeta_{\text{Im/C,a}}, m_{\text{Pt,c}}, \zeta_{\text{Pt/C,c}}, \zeta_{\text{Im/C,c}}) \quad (6)$$

where I (A cm^{-2}) is the current density; V (V) is output voltage; $m_{\text{Pt,a}}$, $\zeta_{\text{Pt/C,a}}$, $\zeta_{\text{Im/C,a}}$, $m_{\text{Pt,c}}$, $\zeta_{\text{Pt/C,c}}$, $\zeta_{\text{Im/C,c}}$ are the Pt loading, Pt percentage and IC ratio of the anode and cathode CLs, respectively; S is the function between the current density and output voltage and CL composition constructed by the surrogate model. As previously discussed, there exist a maximum power density in one polarization curve, i.e. one CL composition for our issue. Maximum power density can be therefore expressed as:

$$p_{\max}(m_{\text{Pt,a}}, \zeta_{\text{Pt/C,a}}, \zeta_{\text{Im/C,a}}, m_{\text{Pt,c}}, \zeta_{\text{Pt/C,c}}, \zeta_{\text{Im/C,c}}) = \max\{VI\}, V \in (0.4, 0.9) \quad (7)$$

where $p_{\max}$ (W cm^{-2}) is the maximum power density which is the function of CL composition. The range of output voltage is 0.4V to 0.9V which is the same as the range in surrogate model development (Table 3), and the maximum power density of the simulated PEMFC always lies in this range through the preliminary test.

2.4 Stochastic optimization algorithm

A stochastic optimization algorithm, genetic algorithm (GA), is used to search the optimal CL composition for improving the maximum power density in this work. GA is a method to search for the approximate optimal solution inspired by the natural evolutionary process. Figure 4 shows the workflow of GA in this work. Six

231 parameters of CL composition, i.e. $m_{Pt,a}$, $\zeta_{Pt/C,a}$, $\zeta_{Im/C,a}$, $m_{Pt,c}$, $\zeta_{Pt/C,c}$, $\zeta_{Im/C,c}$ are the
 232 variables which are expected to be optimized. We use the Genetic Algorithm
 233 Optimization Toolbox (GAOT, developed by North Carolina State University)
 234 implemented in MATLAB to conduct the optimization process in the current work
 235 [30]. Fitness function is used to evaluate the fitness of an individual, and individuals
 236 with the higher fitness have the greater possibility to reproduce the next generation in
 237 the selection. The objective of this work is maximizing the maximum power density,
 238 therefore the predicted maximum power density by the surrogate model is set as the
 239 fitness function. To guarantee the physical meaning of CL composition and the
 240 prediction accuracy of the surrogate model, the optimal solution of CL composition
 241 and the corresponding volume fractions of the three phases should be within the
 242 restrictive ranges in Table 3, namely the variables exist constraint conditions.
 243 Therefore, the penalty term is added in the fitness function to increase the elimination
 244 possibility of individuals which don't meet the constraints in the selection. The fitness
 245 function can be expressed as:

$$246 \quad F = \begin{cases} p_{\max} \left(m_{Pt,a}, \zeta_{Pt/C,a}, \zeta_{Im/C,a}, m_{Pt,c}, \zeta_{Pt/C,c}, \zeta_{Im/C,c} \right) & \text{if } \begin{cases} 0.1 \leq m_{Pt,a}, m_{Pt,c} \leq 0.5 \\ 0.1 \leq \zeta_{Pt/C,a}, \zeta_{Pt/C,c} \leq 0.4 \\ 0.4 \leq \zeta_{Im/C,a}, \zeta_{Im/C,c} \leq 1.2 \\ 0.12 \leq \omega_a, \omega_c \leq 0.3 \\ 0.3 \leq \varepsilon_{CL,a}, \varepsilon_{CL,c} \leq 0.5 \end{cases} \\ 0.01 \times p_{\max} \left(m_{Pt,a}, \zeta_{Pt/C,a}, \zeta_{Im/C,a}, m_{Pt,c}, \zeta_{Pt/C,c}, \zeta_{Im/C,c} \right) & \text{else} \end{cases} \quad (8)$$

247 where F (W cm^{-2}) is the fitness function; $\omega_a, \omega_c, \varepsilon_{CL,a}, \varepsilon_{CL,c}$ are volume fractions of
 248 ionomer and pore of the anode and cathode CLs, respectively; 0.01 is the penalty term.

249 End condition of the optimization process is reaching the given maximum generation.

250 More details about GA and implementation procedure of GAOT can be found in

251 Reference 30 and 31.

252

253 Figure 4 Workflow of genetic algorithm.

254 3. Results and discussion

255 3.1 Prediction performance

256 In the current work, 50 data points randomly selected from the 65 cases

257 constitute the training set and other 15 data points are the test set. After the surrogate

258 model is developed, prediction performance is assessed firstly. Root mean squared

error (RMSE), squared correlation coefficient (R-square), mean percentage error and maximum percentage error are introduced to evaluate model performance:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (I_i - I'_i)^2} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (I_i - I'_i)^2}{\sum_{i=1}^n (I_i - \bar{I}_i)^2} \quad (10)$$

$$\delta_{\text{mean}} = \frac{1}{n} \sum_{i=1}^n \frac{|I_i - I'_i|}{I_i} \times 100 \quad (11)$$

$$\delta_{\text{max}} = \max \left\{ \frac{|I_i - I'_i|}{I_i} \right\} \times 100 \quad (12)$$

where RMSE (A cm^{-2}) is the root mean squared error; R^2 is the squared correlation coefficient; δ_{mean} is the mean percentage error; δ_{max} is the maximum percentage error; n is the number of data points; I_i (A cm^{-2}) is the physical model simulated current density; I'_i (A cm^{-2}) is the surrogate model predicted current density; $\bar{I}_i$ (A cm^{-2}) is the mean value of physical model simulated current densities in the dataset.

Figure 5 shows surrogate model predicted and physical model simulated current densities in the training set and test set. R-square in the training set and test set are 0.9996 and 0.9908, respectively, which are both in a very high level (Table 4). There does not exist the overfitting problem of the developed surrogate model. And other prediction errors in the training set and test set are also shown in Table 4. It can be found that prediction error in the test set is generally higher than that in the training set which is expected in the statistical modeling application. Mean and maximum

percentage error in the test set are 3.3375% and 7.7770%, respectively. Therefore, we can think that the surrogate model has a comparable accuracy with the physical model and it can be used to predict the current density under different CL composition.

(a)

(b)

Figure 5 Surrogate model predicted and physical model simulated current densities in the (a) training set and (b) test set.

3.2 Optimization and verification

As introduced in Section 2.4, the developed surrogate model is employed in the fitness function of GA. Due to the random generation mechanism of initial population, the optimal solution obtained by the stochastic optimization algorithm also exists a certain randomness, and the number of initial population is a critical factor which may influence the optimization result. Through the preliminary test, the numbers of initial population and maximum generation are set as 200 and 1000, respectively. We continuously conduct the GA optimization for five times, and the best fitnesses at the end of maximum generation in the five times are 1.2454 to 1.2647W cm⁻², and the percentage difference is only 1.5497%. It demonstrates that even if the GA optimization is conducted by several times without any change of hyper-parameters (the initial population is randomly generated in each time), the optimized objective (maximum power density) is still in a small scope. Therefore, the given numbers of initial population and maximum generation are feasible. We choose the best fitness of 1.2647W cm⁻² as the final solution, and the best fitness evolution during generations is shown in Figure 6, and the best fitness becomes very stable after 400 generations. And the corresponding optimal CL composition are listed in Table 5.

Figure 6 Optimization result of the maximum power density by GA.

The present optimal CL composition and optimized maximum power density are predicted by the surrogate model. For verification, the optimal CL composition is returned to the physical model to simulate the polarization curve. Figure 7 shows the comparison of polarization curves between the physical model simulation and the surrogate model prediction, and a good agreement is achieved. Output voltage at the maximum power density point predicted by the surrogate model is 0.59V which generally fits the polarization curve simulated by the physical model. Table 6 shows the comparison between maximum power densities of surrogate model prediction and physical model verification under the optimal CL composition, and the percentage error between surrogate model prediction and physical model verification is only

1.3950% which can demonstrate the validity of the optimization results and the proposed framework.

Figure 7 Surrogate model predicted and physical model simulated polarization curves under the optimal CL composition.

4. Conclusion

In this work, an AI-based, data-driven surrogate modeling and optimization framework is presented to realize the multi-variables global optimization of catalyst layer (CL) composition for improving the maximum power density of proton exchange membrane fuel cells (PEMFCs). A case study of a detailed PEMFC is presented in the paper to demonstrate the framework procedure. The simulated current densities under different output voltage and CL composition by a state-of-the-art three-dimensional (3D) computational fluid dynamics (CFD) full cell model coupled

with the CL agglomerate model construct the database. Then, the database is used to train the support vector machine (SVM) model and develop the surrogate model. Prediction performance shows that the squared correlation coefficient (R-square) in the training and test set are 0.9996 and 0.9908, respectively, and the mean and maximum percentage error in the test set are 3.3375% and 7.7770%, respectively. It can be concluded that the efficient surrogate model has a comparable accuracy with the physical model and it can replace the physical model to predict the current density under different CL composition in the optimization process. The data-driven surrogate model has demonstrated 10,000 times higher computation-resource efficiency, which can complete the calculation of one polarization curve within one second, while it may cost hundreds of processor-hours by the physical CFD model. Genetic algorithm (GA) is employed to search the optimal solution of CL composition for maximizing the maximum power density, and the optimal CL composition is returned to the physical model to verify the predicted maximum power density by the surrogate model. The surrogate model predicted and physical model simulated maximum power densities under the optimal CL composition are 1.2647 and 1.2473 W cm⁻², respectively, the percentage error is only 1.3950% which demonstrates the optimization validity of the proposed framework. The results implies that the proposed framework combining the efficient surrogate modeling and stochastic optimization algorithm could be widely applied in the multi-variables optimization of complex systems.

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 National Natural Science Foundation of Tianjin (China) for Distinguished Young Scholars (Grant No. 18JCJQJC46700).

Reference

- [1] A.J. Steinbach, J.S. Allen, R.L. Borup et al., Anode-design strategies for improved performance of polymer-electrolyte fuel cells with ultra-thin electrodes, *Joule* 2 (2018) 1297-1312.
- [2] C. Guo, J. Lu, Z. Tian et al., Optimization of critical parameters of PEM fuel cell using TLBO-DE based on Elman neural network, *Energ. Convers. Manage.* 183 (2019) 149-158.
- [3] Y. Rao, Z. Shao, A.H. Ahangarnejad et al., Shark Smell Optimizer applied to identify the optimal parameters of the proton exchange membrane fuel cell model, *Energ. Convers. Manage.* 182 (2019) 1-8.
- [4] T. Yoshida, K. Kojima, Toyota MIRAI fuel cell vehicle and progress toward a future hydrogen society, *Electrochem. Soc. Interface* 24 (2015) 45-49.
- [5] P. Yi, X. Li, L. Yao, et al., A lifetime prediction model for coated metallic bipolar plates in proton exchange membrane fuel cells, *Energ. Convers. Manage.* 183 (2019) 65-72.

370 [6] Y. Qin, Q. Du, M. Fan, et al., Study on the operating pressure effect on the
371 performance of a proton exchange membrane fuel cell power system, *Energ. Convers.*
372 *Manage.* 142 (2017) 357-365.

373 [7] J. Zhao, X. Li, A review of polymer electrolyte membrane fuel cell durability for
374 vehicular applications: Degradation modes and experimental techniques, *Energ.*
375 *Convers. Manage.* 199 (2019) 112022.

376 [8] H. Chen, X. Zhao, T. Zhang et al., The reactant starvation of the proton exchange
377 membrane fuel cells for vehicular applications: A review, *Energ. Convers. Manage.*
378 182 (2019) 282-298.

379 [9] B. Wang, W. Chen, F. Pan et al., A dot matrix and sloping baffle cathode flow
380 field of proton exchange membrane fuel cell, *J. Power Sources* 434 (2019) 226741.

381 [10] K. Jiao, X. Li, Water transport in polymer electrolyte membrane fuel cells, *Prog.*
382 *Energy. Combust. Sci.* 37 (2011) 221-291.

383 [11] Chang Y, Liu J, Li R et al., Effect of humidity and thermal cycling on the
384 catalyst layer structural changes in polymer electrolyte membrane fuel cells, *Energ.*
385 *Convers. Manage.* 189 (2019) 24-32.

386 [12] J. Zhao, S. Shahgaldi, I. Alaefour et al., Pore structure and effective diffusion
387 coefficient of catalyzed electrodes in polymer electrolyte membrane fuel cells, *Int. J.*
388 *Hydrogen Energy* 43 (2018) 3776-3785.

389 [13] S. Shahgaldi, J. Zhao, I. Alaefour et al., Investigation of catalytic vs reactant
390 transport effect of catalyst layers on proton exchange membrane fuel cell performance,
391 Fuel 208 (2017) 321-328.

392 [14] S. Shahgaldi, A. Ozden, X. Li et al., Cathode catalyst layer design with gradients
393 of ionomer distribution for proton exchange membrane fuel cells, Energ. Convers.
394 Manage. 171 (2018) 1476-1486.

395 [15] L. Chen, Q. Kang, W. Tao, Pore-scale study of reactive transport processes in
396 catalyst layer agglomerates of proton exchange membrane fuel cells, Electrochim.
397 Acta 306 (2019) 454-465.

398 [16] H. Deng, Y. Hou, W. Chen et al., Lattice Boltzmann simulation of oxygen
399 diffusion and electrochemical reaction inside catalyst layer of PEM fuel cells, Int. J.
400 Heat Mass Transf. 143 (2019) 118538.

401 [17] Y. Hou, H. Deng, F. Pan et al., Pore-scale investigation of catalyst layer
402 ingredient and structure effect in proton exchange membrane fuel cell, Appl. Energy
403 253 (2019) 113561.

404 [18] G. Zhang, K. Jiao, Multi-phase models for water and thermal management of
405 proton exchange membrane fuel cell: A review, J Power Sources 391 (2018) 120-133.

406 [19] L. Xing, W. Shi, P.K. Das et al., Inhomogeneous distribution of platinum and
407 ionomer in the porous cathode to maximize the performance of a PEM fuel cell,
408 AIChE J. 63 (2017) 4895-4910.

409 [20] L. Xing, Y. Wang, P.K. Das et al., Homogenization of current density of PEM
410 fuel cells by in-plane graded distributions of platinum loading and GDL porosity,
411 Chem. Eng. Sci. 192 (2018) 699-713.

412 [21] L. Xing, X. Liu, T. Alaje et al., A two-phase flow and non-isothermal
413 agglomerate model for a proton exchange membrane (PEM) fuel cell, Energy 73
414 (2014) 618-634.

415 [22] L. Xing, M. Mamlouk, R. Kumar et al., Numerical investigation of the optimal
416 Nafion® ionomer content in cathode catalyst layer: an agglomerate two-phase flow
417 modelling, Int. J. Hydrogen Energy 39 (2014) 9087-9104.

418 [23] L. Xing, P.K. Das, X. Song et al., Numerical analysis of the optimum
419 membrane/ionomer water content of PEMFCs: The interaction of Nafion® ionomer
420 content and cathode relative humidity, Appl. Energy 138 (2015) 242-257.

421 [24] F. Tao, Q. Qi, Make more digital twins, Nature 572 (2019) 490-491.

422 [25] Y. Sun, G. Szűcs, A. R. Brandt, Solar PV output prediction from video streams
423 using convolutional neural networks, Energy & Environmental Science 11(7) (2018)
424 1811-1818.

425 [26] K. A. Severson, P. M. Attia, N. Jin et al., Data-driven prediction of battery cycle
426 life before capacity degradation, Nature Energy 4(5) (2019) 383.

427 [27] B. Xie, G. Zhang, J. Xuan et al., Three-dimensional multi-phase model of PEM
428 fuel cell coupled with improved agglomerate sub-model of catalyst layer, Energ.
429 Convers. Manage. 199 (2019) 112051.

[28] J.P. Owejan, J.E. Owejan, W. Gu, Impact of platinum loading and catalyst layer structure on PEMFC performance, J. Electrochem. Soc. 160 (2013) F824–33.

[29] C. C. Chang, C.J. Lin. LIBSVM: A library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2 (2011) 27.

[30] C.R. Houck, J. Joines, M.G. Kay, A genetic algorithm for function optimization: a Matlab implementation, Ncsu-ie tr 95(1995) 1-10.

[31] https://en.wikipedia.org/wiki/Genetic_algorithm, excess on 2019.11.9.

Figure and table captions

Figure 1 Polarization characteristic and maximum power density of PEMFCs.

Figure 2 Schematic of the simulated PEMFC and computational domain of the physical model.

Figure 3 Comparison of polarization curves between simulation results by the present physical model [27] and experimental results [28] under different cathode Pt loadings.

Figure 4 Workflow of genetic algorithm.

Figure 5 Surrogate model predicted and physical model simulated current densities in the (a) training set and (b) test set.

Figure 6 Optimization result of the maximum power density by GA.

Figure 7 Surrogate model predicted and physical model simulated polarization curves under the optimal CL composition.

Table 1 Operating conditions, material and electro-chemical kinetics properties of the simulated PEMFC.

451 Table 2 Conservation equations of the 3D CFD full cell model [27].

452 Table 3 Restrictive ranges of parameters in the case generation [23,27].

453 Table 4 Prediction error measured by the RMSE, mean percentage error and

454 maximum percentage error in the training set and test set.

455 Table 5 Optimal anode and cathode CL composition and its output voltage

456 corresponding to the maximum power density.

457 Table 6 Maximum power density of surrogate model prediction and physical model

458 verification under the optimal CL composition and its percentage error.

Table 1 Operating conditions, material and electro-chemical kinetics properties of the simulated PEMFC.

Parameter	Value	Unit
Temperature	353	K
Back pressure	Anode: 2.0 Cathode: 2.0	atm
Stoichiometry ratio	Anode: 1.5 Cathode: 2.0	
Inlet relative humidity	Anode: 0.5 Cathode: 0.5	
Thickness of GDLs, MPLs, membrane	200, 20, 25.4	μm
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
Carbon particle radius	25	nm
Electrical conductivity of flow field plates, GDLs, MPLs, CLs	20000, 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 Conservation equations of the 3D CFD full cell model [27].

Solved variable	Equations
Electronic potential	$0 = \nabla \cdot (\kappa_e^{\text{eff}} \nabla \varphi_e) + S_e$
Ionic potential	$0 = \nabla \cdot (\kappa_{\text{ion}}^{\text{eff}} \nabla \varphi_{\text{ion}}) + S_{\text{ion}}$
Mass of gas mixture	$\frac{\partial}{\partial t} (\varepsilon (1-s) \rho_g) + \nabla \cdot (\rho_g u_g) = S_m$
Momentum of gas mixture	$\frac{\partial}{\partial t} \left(\frac{\rho_g u_g}{\varepsilon (1-s)} \right) + \nabla \cdot \left(\frac{\rho_g u_g u_g}{\varepsilon^2 (1-s)^2} \right) = -\nabla P_g + \mu_g \nabla \cdot \left(\nabla \left(\frac{u_g}{\varepsilon (1-s)} \right) + \nabla \left(\frac{u_g^T}{\varepsilon (1-s)} \right) \right) - \frac{2}{3} u_g \nabla \cdot \left(\nabla \cdot \left(\frac{u_g}{\varepsilon (1-s)} \right) \right) + S_u$
Gas species	$\frac{\partial}{\partial t} (\varepsilon (1-s) \rho_g Y_i) + \nabla \cdot (\rho_g u_g Y_i) = \nabla \cdot (\rho_g D_i^{\text{eff}} \nabla Y_i) + S_i$
Liquid pressure	$\frac{\partial}{\partial t} (\varepsilon s \rho_l) = \nabla \cdot \left(\rho_l \frac{K k_l}{\mu_l} \nabla P_l \right) + S_l$
Membrane water	$\frac{\rho_{\text{mem}}}{\text{EW}} \frac{\partial}{\partial t} (\omega \lambda) + \nabla \cdot \left(n_d \frac{J_{\text{ion}}}{F} \right) = \frac{\rho_{\text{mem}}}{\text{EW}} \nabla \cdot (D_{\text{mw}}^{\text{eff}} \nabla \lambda) + S_{\text{mw}}$
Energy	$\frac{\partial}{\partial t} (\varepsilon s \rho_l C_{p,l} T + \varepsilon (1-s) \rho_g C_{p,g} T) + \nabla \cdot (\varepsilon s \rho_l C_{p,l} u_l T + \varepsilon (1-s) \rho_g C_{p,g} u_g T) = \nabla \cdot (k^{\text{eff}} \nabla T) + S_T$

Table 3 Restrictive ranges of parameters in the case generation [23,27].

	Lower bound	Upper bound
Output voltage, V	0.4	0.9
Pt loading, mg cm ⁻²	0.1	0.5
Pt percentage	0.1	0.4
IC ratio	0.4	1.2
Volume fractions of electrolyte	0.12	0.3
Volume fraction of pores	0.3	0.5

Table 4 Prediction error measured by the RMSE, mean percentage error and maximum percentage error in the training set and test set.

	Training set	Test set
R-square	0.9996	0.9908
RMSE, A cm ⁻²	0.0133	0.0612
Mean percentage error, %	0.5719	3.3375
Maximum percentage error, %	3.5641	7.7770

Table 5 Optimal anode and cathode CL composition and its output voltage corresponding to the maximum power density.

	Anode CL composition	Cathode CL composition
Pt loading, mg cm ⁻²	0.1976	0.4064
Pt percentage	0.3139	0.4
IC ratio	0.5624	0.6639
Output voltage, V	0.59	

Table 6 Maximum power density of surrogate model prediction and physical model verification under the optimal CL composition and its percentage error.

	Maximum power density, W cm ⁻²		Percentage error, %
	Surrogate model prediction	Physical model verification	
Value	1.2647	1.2473	1.3950

Figure 1 Polarization characteristic and maximum power density of PEMFCs.

Figure 2 Schematic of (a) the simulated PEMFC and computational domain of the physical model and (b) the CL single agglomerate.

Figure 3 Comparison of polarization curves between simulation results by the present physical model [27] and experimental results [28] under different cathode Pt loadings.

Figure 4 Workflow of genetic algorithm.

(a)

(b)

Figure 5 Surrogate model predicted and physical model simulated current densities in the (a) training set and (b) test set.

Figure 6 Optimization result of the maximum power density by GA.

Figure 7 Surrogate model predicted and physical model simulated polarization curves under the optimal CL composition.