



Contents lists available at ScienceDirect



Modeling of pressure drop in reverse osmosis feed channels using multilayer artificial neural networks

Jianghang Gu^{d,1}, Jiu Luo^{e,1}, Mingheng Li^f, Chunyan Huang^g, Yi Heng^{a,b,c,*}

^a School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China

^b Guangdong Province Key Laboratory of Computational Science, Guangzhou 510006, China

^c National Supercomputing Center in Guangzhou (NSCC-GZ), Guangzhou 510006, China

^d School of Chemical Engineering and Technology, Sun Yat-sen University, Zhuhai 519000, China

^e School of Materials Science and Engineering, Sun Yat-sen University, Guangzhou 510275, China

^f Department of Chemical and Materials Engineering, California State Polytechnic University, Pomona, CA 91768, United States

^g School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, China

ARTICLE INFO

Article history:

Received 8 March 2020

Received in revised form 3 April 2020

Accepted 6 April 2020

Available online 4 May 2020

Keywords:

Brackish water reverse osmosis

Pressure-drop model

Multilayer artificial neural networks

Large-scale CFD simulations

High-performance computing

ABSTRACT

This paper presents a data-driven model of pressure drop in spiral wound reverse osmosis membrane channels with the aid of computational fluid dynamics (CFD) simulations and multilayer artificial neural networks (MLN) method. Typical industrial spacer structures are considered in the CFD models to ensure practicability. To generate adequate data for MLN training, three-dimensional (3D) CFD model problems with respect to 500 different design parameter combinations are solved with a high-performance computing (HPC) strategy. With the HPC strategy the efficiency of CFD simulations is improved by about 50 times. The proposed MLN method is then used to correlate the pressure drop considering 7 design parameters, which enables a quantitative description of geometric structures and operation conditions for improvement. The accuracy of the MLN prediction on the test set is 99.734%. The obtained results show that the performance of the predictive model is comparable to alternative approaches in previous literatures. The proposed method has potential applications in intricate PDE-constrained optimization problems involving CFD simulations due to its high computational efficiency.

© 2020 Institution of Chemical Engineers. Published by Elsevier B.V. All rights reserved.

1. Introduction

Nowadays, the situation of water shortage has become more and more serious due to the on-going growth of population, the rapid urbanization, and the pollution caused by indus-

trialization. Facing with threat of usable fresh water scarcity, the development of efficient desalination techniques is stringent. The commonly used desalination techniques can be mainly classified into two categories based on the separation mechanisms: thermal process and membrane-based process

Abbreviations: RO, reverse osmosis; SWM, spiral wound membrane; CP, concentration polarization; CFD, computational fluid dynamics; 1D, 2D, 3D, one-, two-, three-dimensional; BWRO, brackish water reverse osmosis; HTC, high throughput computing; HPC, high-performance computing; MLN, multilayer artificial neural networks; PDE, partial differential equation; PDECO, PDE constrained optimization; MSE, mean square error; RME, relative mean error.

* Corresponding author at: School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China.

E-mail address: hengyi@mail.sysu.edu.cn (Y. Heng).

¹ Authors contributed equally to this work.

<https://doi.org/10.1016/j.cherd.2020.04.019>

0263-8762/© 2020 Institution of Chemical Engineers. Published by Elsevier B.V. All rights reserved.

Nomenclature

L, l_x	length parameters (μm)
W, l_y	width (m)
D	diameter parameters (μm)
h	channel height (μm)
U_0	inlet velocity parameter (m/s)
\mathbf{u}	velocity vector
p	hydraulic pressure (Pa)
x, y, z	cartesian coordinates (μm)
X	input data
$\tilde{X}_{j,k}$	normalized input data
y_{out}	output data
w	synaptic weight of a neuron
IW, LW	weight matrixes
b	bias of a neuron
n	output of a neuron
H	Hessian matrix
I	unit matrix
J	Jacobian matrix
e	value of loss function
 Greek letters	
β	spacer design parameter vector
$\tilde{\beta}$	normalized spacer design parameter vector
α	angle between filaments ($^\circ$)
ρ	density of fluid (kg/m^3)
μ	viscosity of fluid (Pa s)
φ	activation function
θ, Θ	hyperparameter of MLN
$\frac{\Delta p}{\Delta L}$	pressure drop per meter (kPa/m)
$\hat{\Delta p}$	predicted pressure drop per meter (kPa/m)
 Subscript	
j	number of design parameters
k	number of samples
i	number of neurons in a layer
l	number of layers
min	minimum
max	maximum

(Crini and Lichtfouse, 2019). Compared with the thermal process, reverse osmosis (RO) gains more acceptance owing to its distinct advantages such as lower energy consumption, modularity and flexibility.

RO module, or more specifically, the spiral wound membrane (SWM) module, is typically comprised of three parts: the front and back membrane sheets, and the middle feed channel spacer. As an essential structural component of the SWM, feed channel spacers are used to support the semi-permeable membranes while it also obviously impacts the hydrodynamic and mass transfer behaviors of the feed water in flow domain. On the one hand, the spacers, serving as obstacles, increase the pressure drop and thus cause a loss of driving force (Bucs et al., 2014; Li et al., 2016). On the other hand, the spacers can improve mass transfer by promoting turbulence intensity and therefore decrease the concentration polarization (CP) (Bucs et al., 2015; Gu et al., 2017; Kim et al., 2016; Li, 2019; Li et al., 2016).

In recent years, hydrodynamics and mass transportation in SWM channels of RO desalination have been intensively studied using various experimental techniques. For example,

a novel optical coherence tomography was adopted to observe the membrane fouling formation in real time (Gao et al., 2014; West et al., 2016). An in situ non-destructive oxygen imaging method was applied to determine the spatially resolved activity and heterogeneity of the biofilm (Farhat et al., 2016; Staal et al., 2017). A three-dimensional (3D) X-ray computed tomography was used to analyze the precise representation of spacer geometry (Horstmeyer et al., 2018). The particle image velocimetry technique was employed to acquire velocity maps in the spacer-filled channel and an empty channel respectively (Haidari et al., 2016). Their experiments provide reliable data and an insight into the underlying mechanism in SWM. However, experimental research is in general time-consuming, effort-demanding, and restricted to a laboratory scale. By contrast, the use of computational techniques can obtain accurate high-resolution data at a lower cost and risk and it is easier to scale up to industrial-scale. As reported in literature, the results of experiments and the computational fluid dynamics (CFD) simulations are basically consistent and mutually confirmed, which indicates that the micro-scale mechanistic model is useful for practical application (Bucs et al., 2015; Li et al., 2016). Many CFD modeling studies of SWM have been focused on the effect of spacer geometry on flow resistance and mass transfer (Bucs et al., 2014; Gu et al., 2017; Horstmeyer et al., 2018; Kwon et al., 2017; Lee et al., 2018; Mansouri et al., 2019; Sano et al., 2019; Siddiqui et al., 2017; Taamneh and Bataineh, 2017). For example, the relationships between spacer's orientation angle and wall shear stress as well as Nusselt number are discussed (Taamneh and Bataineh, 2017). It is shown that the porosity of spacer-filled feed channels has a linear relationship with the flow velocity and a superlinear effect on the pressure drop, which plays an essential role in membrane performance and biofouling formation (Siddiqui et al., 2017). According to Kwon et al. (2017), the spacer-filled channel has uniform internal resistance over the whole range of velocity, whereas the spacer-less channel shows a dramatic drop in resistance with increasing velocity.

However, most of the efforts are limited to a single or a few factors because of high computational cost involved in CFD simulations. The number of design parameter combinations will increase exponentially with larger number of design parameters. Given that the spacer design involves many parameters, it is unrealistic to search for the optimized structure by enumerating all the combinations. According to recent works, neural networks perform well in modeling nonlinear processes (Chen et al., 2020; Wu et al., 2019a,b). For example, an ensemble of recurrent neural network is utilized to predict nonlinear dynamics of chemical reactors, which shows superior performance than standard state-space model (Wu et al., 2019a,b). In a multiscale thin film deposition process, artificial neural networks are employed for nonlinear model predictive control, where the predictions agree with the results of stochastic multiscale system (Kimaev and Ricardez-Sandoval, 2019). With a focus on CFD simulations, the effort of integration with neural network models has grown over time due to their ideal nonlinear mapping capability (Carlberg et al., 2019; Kutz, 2017; Wu et al., 2019c). This work attempts to constitute a surrogate algebraic pressure-drop model through multilayer artificial neural networks (MLN). The established model is supposed to accurately describe pressure drops in a wide range of design parameters, which can be incorporated in a system-level model for the brackish water reverse osmosis (BWRO). In this way, the multi-dimensional partial differential equation (PDE) constrained optimization (PDECO) problem

can be simplified to a one-dimensional (1D) PDE constrained optimization problem. This circumvents the difficulty of handling packs of optimization variables in combination with the complexities of domain decomposition (Ulbrich and Bloemen Waanders, 2018).

Compared with other surrogate modeling techniques, MLN is chosen in this work because of its universal approximation capacity in fitting high-order nonlinear multivariate functions. Nevertheless, adequate data is necessary to train a precise MLN model due to its instinct reliance on data. In this work, adequate data is obtained by conducting a large number of CFD simulations. In order to ensure high-fidelity of the CFD simulations, typical geometries of commercial feed spacers are employed here rather than directly adopt simplified spacer structure (Johannink et al., 2015). A high-performance computing (HPC) strategy combining CFD parallel computing with high throughput computing (HTC) is employed to ensure computational efficiency. For a unit CFD simulation, it decomposes the entire computational domain into subdomains and simultaneously uses multiple computing nodes to solve the subproblems. Thus, the solution time of the unit CFD problem can be reduced. The HTC approach contemporaneously runs independent unit CFD simulations with different design parameters. In this way, highly efficient executions of the computational tasks can be ensured (Djomehri and Biswas, 2003). Based on the adequate data obtained by running large-scale CFD simulations, the pressure-drop model can then be identified and used to guide optimal experimental design of membranes.

The rest of the paper is organized as follows: Section 2 gives the problem formulation and CFD modeling of SWM. Besides, the proposed HPC strategy is discussed. In Section 3, the pressure-drop model is identified by the MLN. In Section 4, the proposed method is validated by the simulation results. Conclusions and an outlook are given at the end.

2. CFD modeling of the BWRO processes

In our recent work (Luo et al., 2020), a hybrid model integrating a 1D system-level model with a local 3D CFD model was proposed. There, the pressure-drop model, which can quantitatively describe the effect of all design parameters, was not addressed. In this work, to establish the pressure-drop model by coupling the CFD modeling with MLN method is considered as an important step toward the future goal of realizing optimal experimental design of feed spacers. In this work we focus on the discussion of the first two steps (see Fig. 1). In order to prepare sufficient data for building the MLN model, large-scale CFD simulations of the BWRO processes need to be carried out. Thus, the development of an HPC strategy to improve the overall computational efficiency is necessary.

2.1. 3D local hydrodynamics model

The feed channel for SWM consists of spacer units in repeated patterns. In order to ensure the practicability, typical industrial spacers structures are considered in this work. The computational domain includes 5 spacer units (cf. Fig. 2(a)). Zoomed-in membrane channel that includes one spacer unit with length, l_x , width, l_y , and height, h is shown in Fig. 2(b). The spacer geometry is characterized by various parameters: diameters of spacer, $D = (D_{1,A}, D_{2,A}; D_{1,B}, D_{2,B}; D_{tot})$, length, $L = (L_{i,A}, L_{j,B})$ ($i=1, 2, 3, 4, 5; j=1, 2, 3, 4$), and angle between filaments, α

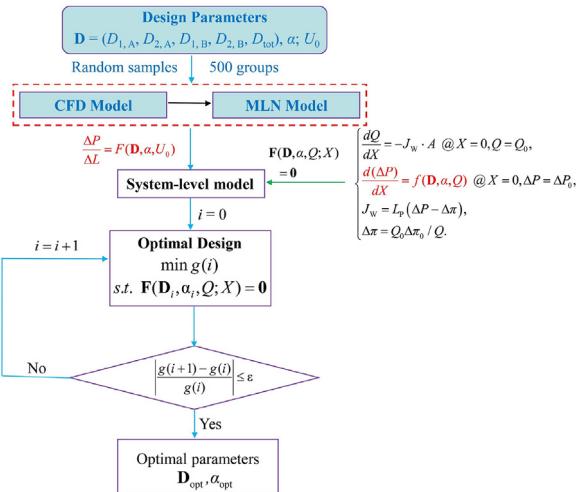


Fig. 1 – Optimal design of the feed spacers based on large-scale 3D CFD simulations, MLN modeling on a system-level.

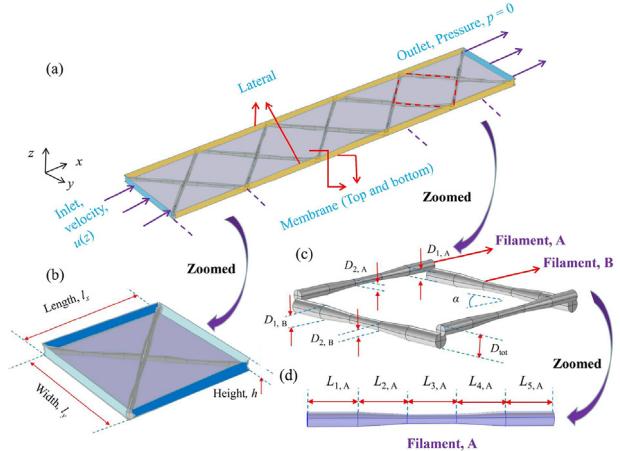


Fig. 2 – (a) 3D Computational domain consists of five spacer units and detailed geometrical characterization for (b) membrane channel with a spacer unit, (c) a spacer unit, and (d) one of spacer filaments.

(cf. Fig. 2(c, d)). In our recent work (Luo et al., 2020), the sensitivity analyses for all above-mentioned parameters under a wide range are carried out, indicating that the hydrodynamics process is more susceptible to D and α , compared with L . $L_{i,A}, L_{j,B}$ are considered as the same constant (cf. Fig. 2(d)) and thus the length of the computational domain L is $5l_x$ (cf. Fig. 2(a)). Therefore, the space of design parameters comprises $\beta = (D_{1,A}, D_{2,A}; D_{1,B}, D_{2,B}; D_{tot}; \alpha; U_0)$. U_0 is considered as a parameter of inlet velocity.

Since CP phenomenon is sufficiently weak to be ignored for brackish water desalination, only the hydrodynamics process is considered in this work (Luo et al., 2020). The Navier-Stokes equation and continuity equation for laminar flow for numerical simulation are considered (Li et al., 2016):

$$\rho(\mathbf{u} \cdot \nabla)\mathbf{u} = \nabla \cdot [-p\mathbf{I} + \mu(\nabla\mathbf{u} + (\nabla\mathbf{u})^T)], \quad (1)$$

$$\rho\nabla \cdot \mathbf{u} = 0, \quad (2)$$

where constant density, ρ (kg/m³), viscosity, μ (Pa·s) of brackish water are applied (Luo et al., 2020). \mathbf{u} , p (Pa) are the velocity

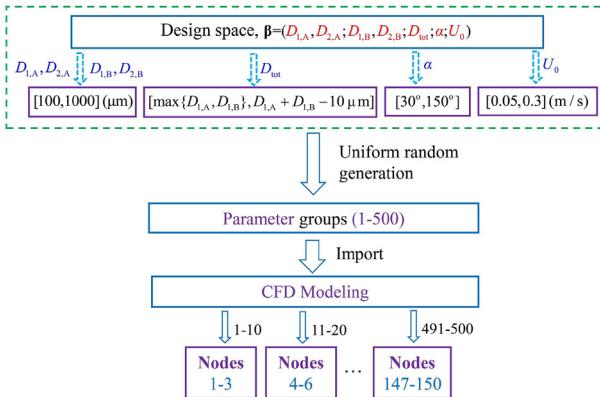


Fig. 3 – The proposed parametric modeling and HPC strategy for the numerical implementation of unit CFD simulations.

vector and hydraulic pressure. Inlet velocity profile (Li et al., 2016) is given by

$$u(z) = \frac{3}{2} U_0 \left[1 - \left(\frac{2z}{H} \right)^2 \right], \quad (3)$$

where U_0 (m/s) as a design parameter is applied to adjust the inlet velocity. Previous works (Li et al., 2002, 2016) showed that the periodically repeated velocity profile in spacer-filled channel is reached within a few cells. Using the parabolic velocity profile may further shorten the length. At outlet, zero-pressure is specified. Periodical boundary conditions are considered on both lateral boundaries (cf. Fig. 2(a)). On the membrane surface (see Fig. 2(a)), an adiabatic wall by setting a zero-flux condition is applied.

2.2. The HPC strategy and numerical implementation

Similar to the case that huge amount of biological DNA sequence data is generated by different base-pair combinations, in this work the design parameter space that contains a large number of parameter combinations is considered. 500 CFD models with respect to different parameter combinations are simulated. The average number of tetrahedral finite elements used for each unit CFD simulation is approximately 2 million. The numerical implementation for all 3D CFD models is performed by COMSOL Multiphysics 5.3a on Tianhe-2 supercomputer. Due to their highly nonlinearity, the solutions will involve high computational cost. Therefore, a series of speedup strategies are developed here. For a 3D BWRO model, the computational time rapidly increases with finer mesh discretization. In order to accelerate the convergence, hydrodynamics results, velocity and pressure distributions are first acquired on a coarse mesh. Then, the approximated results are used as initial estimates to further obtain accurate solutions on a finer mesh. The iterative computing procedure on the fine mesh terminates when relative convergence tolerance is less than 10^{-4} . Furthermore, the HTC strategy is proposed to handle numerous design parameter combinations.

As shown in Fig. 3, the range of parameter values of $D_{1,A}$, $D_{2,A}$, $D_{1,B}$, $D_{2,B}$ are chosen from $100 \mu\text{m}$ to $1000 \mu\text{m}$. Filament A and filament B (cf. Fig. 2(c)) must be overlapped to keep the membrane from top and bottom apart. Therefore, the choice of design parameter, D_{tot} needs to satisfy

$$\max \{ D_{1,A}, D_{1,B} \} \leq D_{\text{tot}} \leq D_{1,A} + D_{1,B} - 10 \mu\text{m}. \quad (4)$$

The maximum and the minimum value of α is 30° and 150° respectively. Apart from geometrical parameters, the considered inlet velocity parameter, U_0 is ranged from 0.05 m/s to 0.3 m/s .

In this work, randomly generated 500 parameter combinations are divided into 50 groups for efficient numerical implementation. The established model for each group includes 10 parameter combinations is solved by using 3 nodes (each node has 24 cores), where scalability tests for the model were minutely discussed in our recent work (Luo et al., 2020). Theoretically, 3600 cores computing scale can be reached, which leads to about 1000 core hours in total. Using the HPC method, the absolute computing time is shortened by nearly 50 times compared with that using a standard sequential computing strategy.

3. Model identification of the pressure drop with MLN

The PDECO problems arise in many fields of industry and science especially in problems of design (Rees et al., 2010). The size and complexity of high-dimensional mesh-dependent PDEs lead to significant computational challenges in contemporary PDECO methods (Ulbrich and Bloemen Waanders, 2018). To simplify the PDECO problems and to circumvent the adjoint problems, it is desirable to substitute the PDEs with a surrogate model. There are various methods that can be used. Compared with other surrogate modeling techniques, for example, linear regression, multivariate adaptive regression splines, Kriging radial basis functions and so on, MLN has better performance in fitting high-order nonlinear multivariate functions. In the mathematical theory of MLN, the universal approximation theorem states that a feed-forward network with a single hidden layer containing a finite number of neurons can approximate continuous functions on compact subsets of \mathbb{R}^n , under mild assumptions on the activation function (Csáji, 2001; Lu et al., 2017). So, the MLN can be considered as a preference to map the relationship between multi-dimensional design space $\beta = (D_{1,A}, D_{2,A}; D_{1,B}, D_{2,B}; D_{\text{tot}}; \alpha; U_0)$ and pressure-drop. Thus, in this work the 3D PDECO problem can be reduced to a 1D PDE constrained optimization problem. The construction procedure of the MLN is shown below. For convenience, $\frac{\Delta p}{\Delta L}$ (kPa/m) denotes the pressure-drop per meter and $\hat{\frac{\Delta p}{\Delta L}}$ (kPa/m) represents its prediction.

3.1. Data preparation

The data preparation is conducted to limit errors and ensure efficient analyses. Among the aforementioned 500 parameter groups, a total of 473 successful CFD simulations are carried out in this work. 7 important variables are selected (see Section 2) from all 16 parameters (Luo et al., 2020). So, the input data X is a 7×1 matrix, including six structure parameters $D_{1,A}$, $D_{2,A}$, $D_{1,B}$, $D_{2,B}$, D_{tot} , α and one operating parameter U_0 . All input data are then normalized as follows, so as to achieve faster convergence during MLN training:

$$\tilde{x}_{j,k} = \frac{x_{j,k} - x_{j,\min}}{x_{j,\max} - x_{j,\min}}; \quad j = 1, 2, \dots, 7; \quad k = 1, 2, \dots, 473, \quad (5)$$

where $\tilde{x}_{j,k}$ is the normalized data, $x_{j,k}$ stands for the j th parameter and k th sample. $x_{j,\min}$, $x_{j,\max}$ represent the minimum and

maximum values of $x_{j,k}$ respectively. The normalized design parameters are defined as $\bar{D}_{1,A}, \bar{D}_{2,A}, \bar{D}_{1,B}, \bar{D}_{2,B}, \bar{D}_{tot}, \bar{\alpha}, \bar{U}_0$.

3.2. Construction procedure of the MLN network

Neurons are the basic information-processing units in the MLN architecture. As shown in Fig. 4(a), a neuron is typically comprised of three parts: connecting links, adder, and activation function. The connecting links connect the input signal X_j with synaptic weight w_j by multiplication. The adder serves as a summing junction for all input signals ($X_j, j = 1, 2, \dots, 7$) and an external bias b . In summary, the output signal of a neuron can be mathematically described as:

$$n_i = \varphi \left(\sum_{j=1}^n w_j X_j + b \right), \quad (6)$$

where φ denotes the activation function.

The basic MLN is constituted by an input layer, an output layer and a hidden layer as depicted in Fig. 4(b). By adding the hidden layers, the MLN is enabled to extract high-order nonlinear relation, whereas the generalization performance will deteriorate. The specific choice of the number of hidden layers is discussed in Section 4. In the feedforward MLN, the preceding layer of nodes projects its result onto its adjacent backward layer of nodes, but not vice versa. The output of the MLN in summary can be formulated as follows:

$$y_{out} = \sum_{j=1}^{nl} w_{i,j} n_i^{last}, \quad (7)$$

$$n_i^{last} = \varphi \left(\sum_{L=1}^{Lm} w_{m,Lm} \varphi \left(\sum_{L=1}^{L...} w_{...,L} \varphi \left(\sum_{L=1}^{L2} w_{2,L2} \varphi \left(\sum_{L=1}^{L1} w_{1,L1} \varphi \left(\sum_{j=1}^7 w_{0,j} X_j + b_0 \right) + b_1 \right) + b_2 \right) + b_{...} \right) + b_m \right).$$

In Eq. (7), $w_{i,j}$ is the synaptic weight of the output layer, n_i^{last} is the output signal of the last hidden layer, nl represents the number of neurons in the last hidden layer, y_{out} stands for the output of the MLN. In Eq. (8), $w_{0,j}$, $w_{1,L1}$ and $w_{2,L2}$ respectively represent the synaptic weight of the input layer, the first hidden layer, the second hidden layer, while b_1, b_2, b_m symbolize the corresponding bias. $w_{...,L}$ and $b_{...}$ are the weight and bias of other hidden layers.

3.3. Network training and method validation

The loss function is the objective of MLN training, which is given by Eq. (9), where N is the number of samples. y_k and $y_{k,out}$ are the actual value and predicted value, e stands for the loss.

$$e = \frac{1}{N} \sum_1^N (y_k - y_{k,out})^2. \quad (9)$$

The network training updates the initial weights and bias of each layer of MLN to minimize the loss function and to optimize the network performance. In the batch mode, the inputs are processed as concurrent vectors and changes of network parameters are accumulated over an epoch before applied. Online training is slightly different since it constantly updates the network weights for each input sample. According to the

works of Principe et al. (2000) and Wilson and Martinez (2003), batch training enables faster calculations and yields smaller errors than online training if the training set is not too large. Thus, the batch training approach is chosen in this work.

The Levenberg–Marquardt algorithm is employed to minimize the value of e . Similar to the quasi-Newton methods, the Levenberg–Marquardt algorithm is designed to approach second-order training speed without computing the Hessian matrix (Hagan and Menhaj, 1994). Because the loss function is the form of a sum of squares, the Hessian matrix can be approximated as follow:

$$H = J^T J. \quad (10)$$

Then the gradient can be computed as:

$$g = J^T e. \quad (11)$$

Here J (cf. Eqs. (10)–(13)) is the Jacobian matrix that contains first derivatives of the MLN errors with respect to weights and bias (x stands for parameter in neural network like w or b).

$$J(x) = \begin{pmatrix} \frac{\partial e_1(x)}{\partial x_1} & \frac{\partial e_1(x)}{\partial x_2} & \dots & \frac{\partial e_1(x)}{\partial x_n} \\ \frac{\partial e_2(x)}{\partial x_1} & \frac{\partial e_2(x)}{\partial x_2} & \dots & \frac{\partial e_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_n(x)}{\partial x_1} & \frac{\partial e_n(x)}{\partial x_2} & \dots & \frac{\partial e_n(x)}{\partial x_n} \end{pmatrix}. \quad (12)$$

Thus, the parameters can be updated iteratively as follow:

$$x_{k+1} = x_k - [J^T J + tI]^{-1} J^T e. \quad (13)$$

In Eq. (13), t is a variable step-size. It decreases after a successful step that results in decreased e and increases after a tentative step when the e increases.

3.4. Hyperparameters tuning in the MLN

Despite its universal approximation capability, the MLN is indeed a “black box” and the tuning of the hyperparameters is a “black art”. Hyperparameter optimization is extremely intractable due to the expensive computational cost of objective function evaluation. Once hyperparameters are changed, such as number of units, type of activation function and etc., a new model on training set should be retrained. As Eq. (14) shows, different hyperparameters θ are constantly tested until the optimal hyperparameters θ^* are sought, which maximize the model performance $f(\theta)$.

$$\theta^* = \arg \max_{\theta \in \Theta} f(\theta). \quad (14)$$

Random search and grid search are slightly better than manual tuning because the training-prediction-evaluation cycle can be conducted automatically in a loop once grid of model hyperparameters is set up. Compared with Bayesian

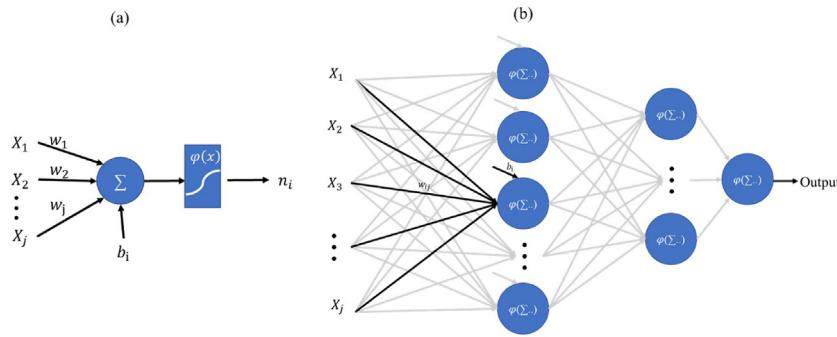


Fig. 4 – (a) The structure of a neuron, (b) the basic structure of MLN.

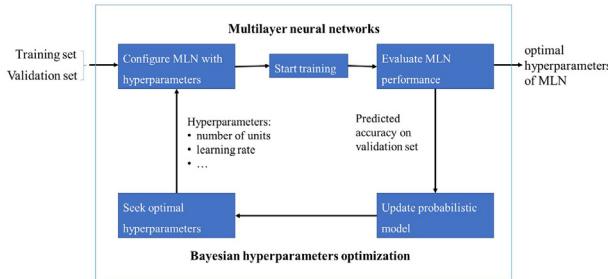


Fig. 5 – The scenario of MLN training.

approaches (Snoek et al., 2012), these methods are still relatively inefficient because they are uninformed by past evaluations. Unlike grid search or random search, the Bayesian approaches keep track of past evaluation results by forming a probabilistic model, which maps the hyperparameters to the probability of scores on objective function. The procedure of the Bayesian hyperparameter optimization is used here and shown below.

Algorithm 1. Bayesian hyperparameter estimation

```

Input: Objective function  $f(\theta)$ , hyperparameters  $\theta$ , number of iterations  $N$ 
Establish a surrogate probabilistic model  $P(y|\theta)$  of the objective function
for number of iterations do
    Search optimal hyperparameters on surrogate probabilistic model
    Apply the acquired hyperparameters to the objective function
    Select hyperparameters if better performance appears
    Update surrogate probabilistic model with new parameters
End
Output: Desired optimal hyperparameters  $\theta^*$ 

```

The mean square error (MSE) and relative mean error (RME) are used here to the test data to examine the generalization performance of MLN model. Further evaluation is conducted through “pairwise factor analyses” in Section 4. The general scenario of MLN training is shown in Fig. 5.

4. Simulation results and performance analysis

4.1. Hydrodynamics characteristics

For the study of the effect of different feed-spacer selection, hydrodynamics results of two representative structures, among all of 500 various geometries are analyzed with respect to equivalent inlet velocities. The simulated pressure distri-

butions have a significant difference with respect to various spacer structures. It is shown that the pressure of structure 1 (cf. Fig. 6(a)) is about six times of that in structure 2 (cf. Fig. 6(b)). Furthermore, we analyze the velocity vectors (cf. Fig. 7) and streamlines distribution (cf. Fig. 8) for various cross profile (yz slices) with respect to both structures. The cross profile is set at $x=1/2, 11/20, 3/5$ of the length in computational domain, respectively. Obviously, the latter one (cf. Fig. 7(b)) has more uniform velocity distribution than that of the former one (cf. Fig. 7(a)). It is shown in Fig. 8 that more vortexes appear in the case of structure 1 than that in the case of structure 2. The vortexes help suppress the CP phenomenon, even though they do not affect much the flux in BWRO (Luo et al., 2020). However, it might lead to more energy consumption.

4.2. Identification of the pressure-drop model

In this work, the data set is divided into the training sets (80%), the validation sets (10%) and the test sets (10%). The MLN training is conducted on MATLAB R2019a, using the ‘feed-forwardnet’ function in the ‘Deep Learning Toolbox’. In the hidden layers, the hyperbolic tangent transfer function (cf. Eq.

(15)) is chosen as the activation function due to its superior differentiable feature. It is in general more efficient because of its steeper derivative for faster grading. In the output layer, the linear transfer function is employed. The computational time for MLN training at a computing node is about one second that is negligible. Compared with CFD simulations, the MLN can in general reduce the time of solving PDE problems by two orders of magnitude.

$$\varphi(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1. \quad (15)$$

Fig. 9 compares the results by using three-layered feed-forward neural network, four-layered feed-forward neural

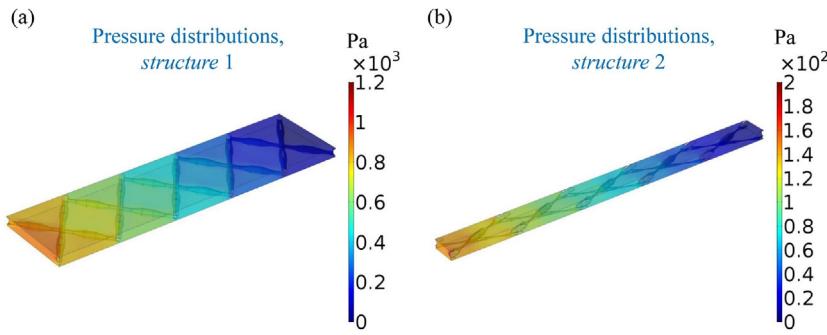


Fig. 6 – Simulated pressure distributions (Pa) for two representative (a) spacer structure 1 and (b) spacer structure 2 with a uniform inlet velocity.

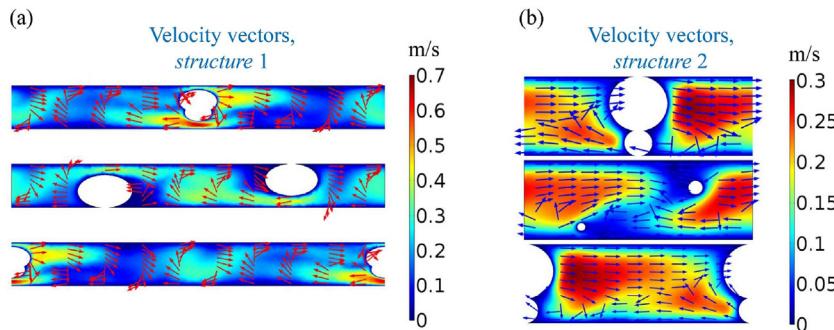


Fig. 7 – The distribution of the velocity vectors to both the magnitude and direction with respect to various yz slices that cut at $x = 1/2, 11/20, 3/5$ (from top to bottom) of the length for the computational domain for two representative (a) spacer structure 1 and (b) spacer structure 2.

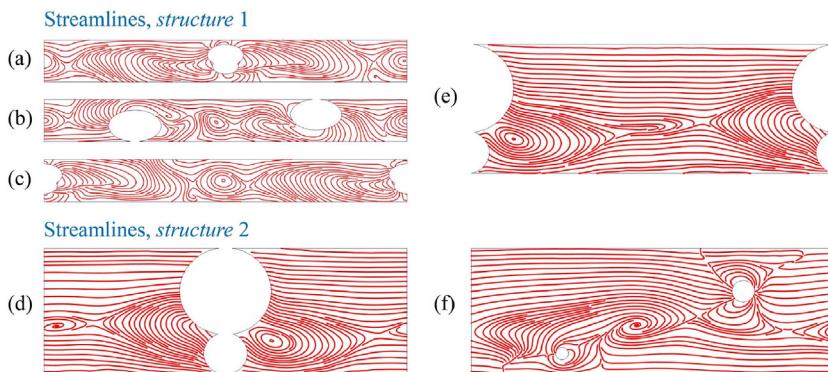


Fig. 8 – Streamlines on various yz slices that are cut at $x = 1/2, 11/20, 3/5$ of the length for the computational domain with respect to two representative structure 1 (a–c) and structure 2 (d–f).

network and five-layered feed-forward neural network. The use of one-layered neural network and two-layered neural networks lead to much worse results, thus they are not considered in this work. Each subfigure of Fig. 9 depicts the performance of neural network with regression coefficient R in terms of training sets and validation sets as well as test sets. The x-coordinate and the y-coordinate stand for actual $\frac{\Delta p}{\Delta L}$ (kPa/m)

and the forecast $\hat{\Delta p}$ (kPa/m) respectively. In Fig. 9(a) and (b), it is shown that all the regression coefficients in four-layered neural network are larger than that in three-layered network. So, the four-layered neural network is, in overall, more accurate for the prediction than the three-layered neural network is. In Fig. 9(b) and (c), it is shown that the performance of validation sets and test sets that uses the five-layered neural network is much worse than that uses the four-layered neural network. This indicates that the generalization performance is very poor. It implies that there is overfitting in the five-layered feed-forward neural network. Therefore, in this

work the four-layered feed-forward neural network is chosen as the surrogate model for pressure drop.

Based on the Bayesian hyperparameter estimation and the calculation, the number of neurons in each layer is 25, 25, 20, 1 correspondingly. The model of $\frac{\Delta \hat{p}}{\Delta L}$ is shown below in Eq. (16). Here, $IW^{1,1}, IW^{2,1}, IW^{3,1}, LW^{4,1} \in R^{25 \times 7}, R^{25 \times 25}, R^{20 \times 25}, R^{1 \times 20}$ and $b^1, b^2, b^3, b^4 \in R^{25 \times 1}, R^{25 \times 1}, R^{20 \times 1}, R^{1 \times 1}$ are the weight matrixes and biased vectors. The MSE on the training set and on the test set are 6.71×10^{-27} and 4.96×10^{-6} , respectively. The corresponding RME (%) is 0.226%.

$$\frac{\Delta \hat{p}}{\Delta L} = LW^{4,1}\varphi(IW^{3,1}\varphi(IW^{2,1}\varphi(IW^{1,1}X + b^1) + b^2) + b^3) + b^4. \quad (16)$$

4.3. Assessment of the identified pressure-drop model

To further evaluate the surrogate pressure-drop model, a series of pairwise factor analyses are con-

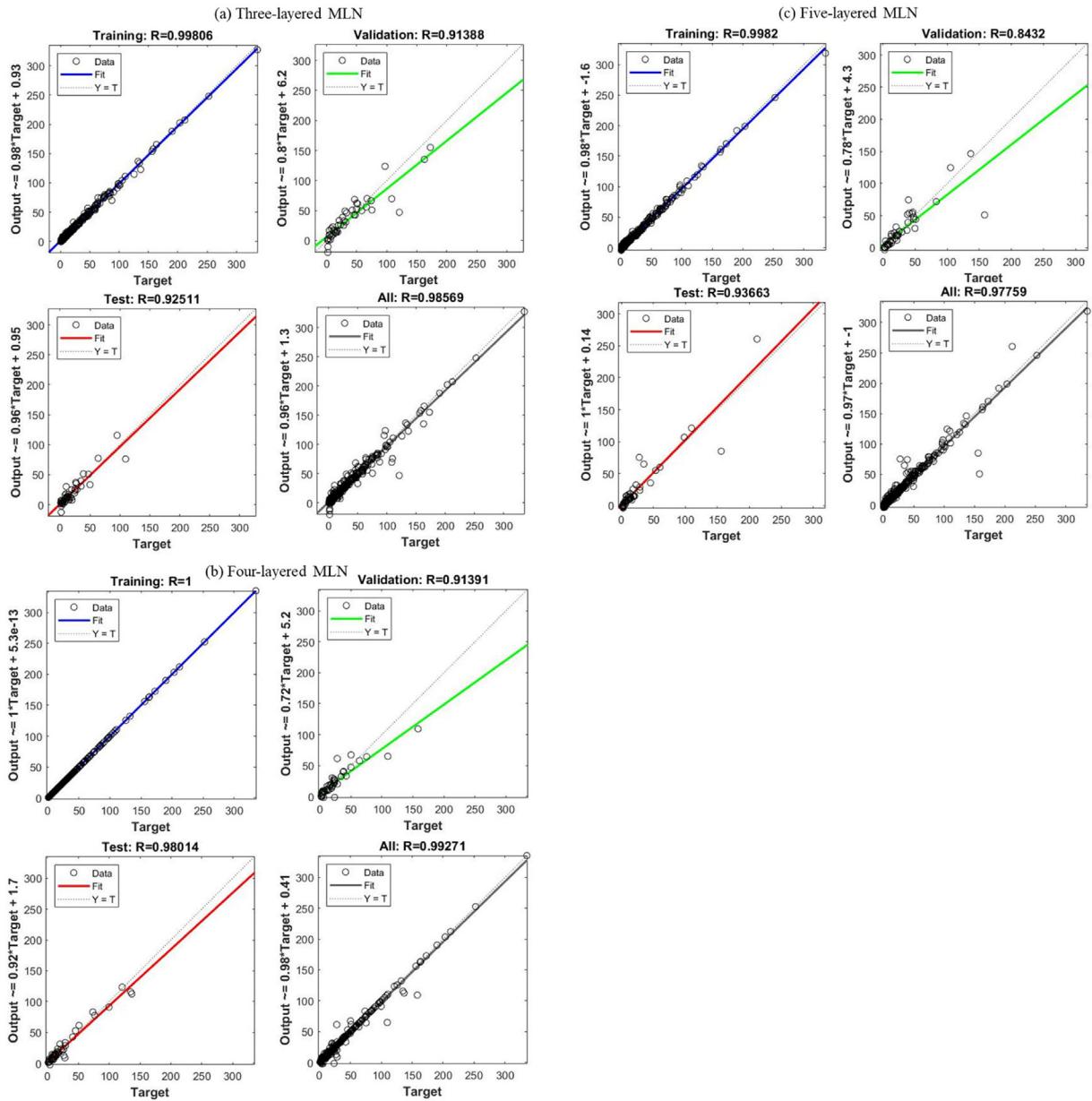


Fig. 9 – The performance analysis of (a) three-layered MLN, (b) four-layered MLN, (c) five-layered MLN.

ducted. As shown in Fig. 10, pairwise combinations of $\bar{D}_{1,A}$, $\bar{D}_{2,A}$, $\bar{D}_{1,B}$, $\bar{D}_{2,B}$, \bar{D}_{tot} , $\bar{\alpha}$, \bar{U}_0 correspond to the x -, y -coordinates in 2D plots, while the color scale is associated with the values of $\frac{\Delta\hat{p}}{\Delta L}$. Based on the work of Bucs et al. (2014), here we choose $D_{1,A}$, $D_{2,A}$, $D_{1,B}$, $D_{2,B}$, D_{tot} , α , U_0 as 354 μm , 221 μm , 449 μm , 289 μm , 711 μm , 90° and 0.075 m/s, which correspond to the normalized values 0.281, 0.131, 0.389, 0.210, 0.326, 0.500, 0.100. While $D_{2,A}$, $D_{2,B}$, $\bar{\alpha}$, $\bar{U}_0 \in [0, 1]$, the ranges of $D_{1,A}$, $D_{1,B}$, \bar{D}_{tot} are constrained by Eq. (4). The pairwise combinational parameters to be studied are free to adjust while fixing the other parameters. For example, by varying $D_{1,A}$, $D_{2,A}$ in their ranges with $D_{1,B}$, $D_{2,B}$, D_{tot} , $\bar{\alpha}$, \bar{U}_0 fixed as 0.389, 0.210, 0.326, 0.500, 0.100, a distribution of $\frac{\Delta\hat{p}}{\Delta L}$ can be computed over the $D_{1,A}$, $D_{2,A}$ coordinates.

It can be inferred from Fig. 10(a) that $\frac{\Delta\hat{p}}{\Delta L}$ has an overall increasing trend with bigger \bar{U}_0 . The similar trend is observed in subfigures pertaining to $\bar{\alpha}$ (cf. Fig. 10(b)). As shown in

Fig. 10(c), $\frac{\Delta\hat{p}}{\Delta L}$ decreases as \bar{D}_{tot} and the other parameter both increase. Fig. 10(d) shows that the impact on $\frac{\Delta\hat{p}}{\Delta L}$ is apparent with pairwise combinations of \bar{D}_{tot} , $\bar{\alpha}$, \bar{U}_0 , where the magnitude of the peak $\frac{\Delta\hat{p}}{\Delta L}$ is about three times than those in Fig. 10(a) and (b). These results are in accordance with our previous conclusions drawn from CFD calculations (Luo et al., 2020). The relation between $\bar{\alpha}$ and $\frac{\Delta\hat{p}}{\Delta L}$ also resembles the results in the work of Johannink et al. (2015). There, the relation between the mesh angle and the pressure drop is studied with a simplified non-woven spacer geometry, which showed that larger mesh angle resulted in larger pressure drop. In this work, the positive correlation of $\bar{\alpha}$ and $\frac{\Delta\hat{p}}{\Delta L}$ is acquired with a more complicated non-woven spacer geometry. Compared with D_{tot} , $\bar{\alpha}$, \bar{U}_0 , parameters $D_{1,A}$, $D_{2,A}$, $D_{1,B}$, $D_{2,B}$ have inconspicuous impact on $\frac{\Delta\hat{p}}{\Delta L}$ (see Fig. 10(e)).

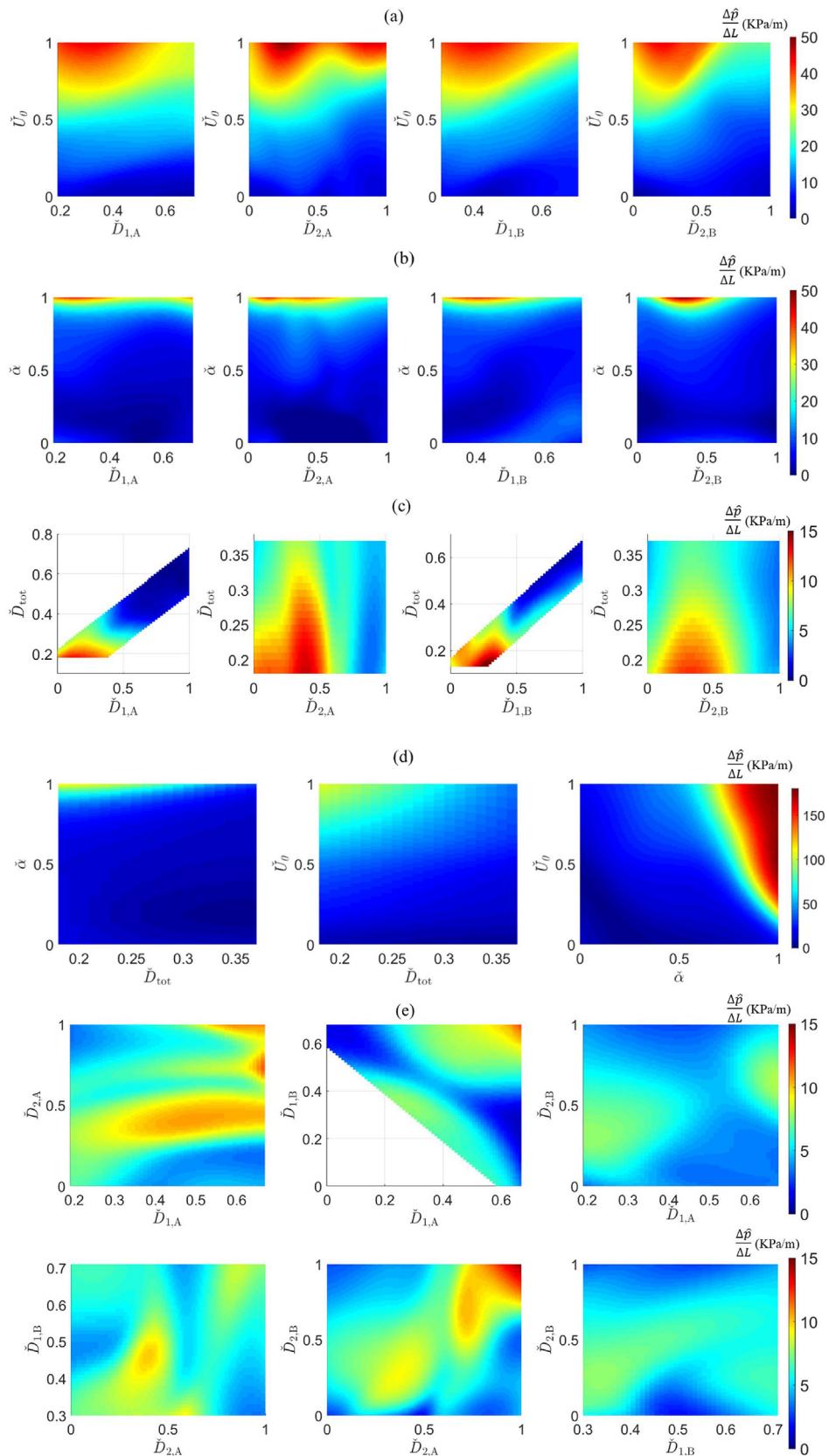


Fig. 10 – Correlation analysis of all pairwise combinational design parameters with $\frac{\Delta\hat{p}}{\Delta L}$ (kPa/m) in the cases of (a) pairwise factor analysis for \dot{U}_0 , (b) pairwise factor analysis for $\dot{\alpha}$, (c) pairwise factor analysis for \dot{D}_{tot} , (d) pairwise factor analysis for \dot{D}_{tot} , $\dot{\alpha}$, \dot{U}_0 , and (e) pairwise factor analysis for $\dot{D}_{1,A}$, $\dot{D}_{2,A}$, $\dot{D}_{1,B}$, $\dot{D}_{2,B}$.

5. Conclusions and outlook

In BWRO processes, flow resistance has a greater influence on water recovery than CP because of the sufficiently low

feed concentration. Therefore, only the hydrodynamic effect is taken into consideration in the CFD simulations in this work. Based on the CFD simulation data, a surrogate pressure-drop model incorporating 7 design parameters is established

with a self-designed MLN. The prediction accuracy of four-layered MLN result is 99.734%. The following conclusions are drawn:

- (1) HPC strategy is employed for the numerical simulations of 3D CFD models on Tianhe-2 supercomputer. The total computational cost is about 10,000 core hours for 50 RO CFD models with 500 parameter groups. Compared with sequential computing strategy, this method shortens the absolute computing time by nearly 50 times.
- (2) The identified pressure drop model provides a comprehensive quantitative description of the geometric structures and operation conditions, which is consistent with the conclusions drawn from previous literature (Johannink et al., 2015) and our recent work (Luo et al., 2020).
- (3) A typical iterative solution approach for tackling 3D PDE-constrained optimization problems often needs to solve the direct PDE problem, corresponding sensitivity and adjoint PDE problems at each iteration, which is computationally very expensive. In this work, the HPC strategy employed for the large-scale numerical simulations of 3D CFD models together with the proposed MLN method are developed to improve the overall computing efficiency for this type of problems.

Future work will be devoted to searching optimal design parameters based on the aforementioned predictive models for system BWRO processes. For seawater RO, both pressure drop and mass transfer will be taken into consideration in the CFD models.

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgements

The corresponding author Yi Heng acknowledges support provided by the “Young Overseas High-Level Talents Introduction Plan” funding of China and Zhuijiang Talent Program of Guangdong Province (No. 2017GC010576). This manuscript is prepared based on the extension of the work shortlisted for the 2019 “Tianhe Star” excellent supercomputing application award delivered by National Supercomputing Center in GuangZhou, China. Chunyan Huang is supported in part by National Natural Science Foundation of China (No. 11971503) and the Young Talents Program of CUFE (No. QYP1809).

References

- Bucs, S.S., Linares, R.V., Marston, J.O., Radu, A.I., Vrouwenvelder, J.S., Picioreanu, C., 2015. Experimental and numerical characterization of the water flow in spacer-filled channels of spiral-wound membranes. *Water Res.* 87, 299–310.
- Bucs, S.S., Radu, A.I., Lavric, V., Vrouwenvelder, J.S., Picioreanu, C., 2014. Effect of different commercial feed spacers on biofouling of reverse osmosis membrane systems: a numerical study. *Desalination* 343, 26–37.
- Carlberg, K.T., Jameson, A., Kochenderfer, M.J., Morton, J., Peng, L., Witherden, F.D., 2019. Recovering missing CFD data for high-order discretizations using deep neural networks and dynamics learning. *J. Comput. Phys.* 395, 105–124.
- Chen, S., Wu, Z., Christofides, P.D., 2020. A cyber-secure control-detector architecture for nonlinear processes. *AIChE J.*
- Crini, G., Lichtfouse, E.J.E.C.L., 2019. Advantages and disadvantages of techniques used for wastewater treatment. *Environ. Chem. Lett.* 17, 145–155.
- Csáji, B.C., 2001. Approximation with artificial neural networks. Faculty of Sciences, Etvs Lornd University, Hungary 24, 7.
- Djomehri, M.J., Biswas, R.J.P.C., 2003. Performance enhancement strategies for multi-block overset grid CFD applications. *Parallel Comput.* 29, 1791–1810.
- Farhat, N.M., Staal, M., Bucs, S.S., Van Loosdrecht, M.C.M., Vrouwenvelder, J.S., 2016. Spatial heterogeneity of biofouling under different cross-flow velocities in reverse osmosis membrane systems. *J. Membr. Sci.* 520, 964–971.
- Gao, Y., Haavisto, S., Li, W., Tang, C.Y., Salmela, J., Fane, A.G., 2014. Novel approach to characterizing the growth of a fouling layer during membrane filtration via optical coherence tomography. *Environ. Sci. Technol.* 48, 14273–14281.
- Gu, B., Adjiman, C.S., Xu, X.Y., 2017. The effect of feed spacer geometry on membrane performance and concentration polarisation based on 3D CFD simulations. *J. Membr. Sci.* 527, 78–91.
- Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* 5, 989–993.
- Haidari, A.H., Heijman, S.G.J., van der Meer, W.G.J., 2016. Visualization of hydraulic conditions inside the feed channel of reverse osmosis: a practical comparison of velocity between empty and spacer-filled channel. *Water Res.* 106, 232–241.
- Horstmeyer, N., Lippert, T., Schön, D., Schleiderer, F., Picioreanu, C., Achterhold, K., Pfeiffer, F., Drewes, J.E., 2018. CT scanning of membrane feed spacers – impact of spacer model accuracy on hydrodynamic and solute transport modeling in membrane feed channels. *J. Membr. Sci.* 564, 133–145.
- Johannink, M., Masilamani, K., Mhamdi, A., Roller, S., Marquardt, W., 2015. Predictive pressure drop models for membrane channels with non-woven and woven spacers. *Desalination* 376, 41–54.
- Kimaev, G., Ricardez-Sandoval, L.A., 2019. Nonlinear model predictive control of a multiscale thin film deposition process using artificial neural networks. *Chem. Eng. Sci.* 207, 1230–1245.
- Kim, Y.C., Lee, S., Park, S.J., 2016. Effects of membrane transport properties and structure parameter on pressure-retarded osmosis performance. *Desalin. Water Treat.* 57, 10111–10120.
- Kutz, J.N., 2017. Deep learning in fluid dynamics. *J. Fluid Mech.* 814, 1–4.
- Kwon, K., Park, B.H., Kim, D.H., Kim, D., 2017. Comparison of spacer-less and spacer-filled reverse electrodialysis. *J. Renew. Sustain. Energy Rev.* 9, 044502.
- Lee, K.S., Cho, Y., Choo, K.Y., Yang, S., Han, M.H., Kim, D.K., 2018. Membrane-spacer assembly for flow-electrode capacitive deionization. *Appl. Surf. Sci.* 433, 437–442.
- Li, F., Meindersma, W., De Haan, A., Reith, T., 2002. Optimization of commercial net spacers in spiral wound membrane modules. *J. Membr. Sci.* 208, 289–302.
- Li, M., 2019. Predictive modeling of a commercial spiral wound seawater reverse osmosis module. *Chem. Eng. Res. Des.* 148, 440–450.
- Li, M., Bui, T., Chao, S., 2016. Three-dimensional CFD analysis of hydrodynamics and concentration polarization in an industrial RO feed channel. *Desalination* 397, 194–204.
- Lu, Z., Pu, H., Wang, F., Hu, Z., Wang, L., 2017. The expressive power of neural networks: a view from the width. *Adv. Neural Inf. Process Syst.*, 6231–6239.
- Luo, J., Li, M., Heng, Y., 2020. A hybrid modeling approach for optimal design of non-woven membrane channels in brackish water reverse osmosis process with high-throughput computation. *Desalination* (accepted).
- Mansouri, N., Moghimi, M., Taherinejad, M., 2019. Investigation on hydrodynamics and mass transfer in a feed channel of a spiral-wound membrane element using response surface methodology. *Chem. Eng. Res. Des.* 149, 147–157.

- Principe, Jose, C., Neil, R., Euliano, Lefebvre, W.C., 2000. *Neural and Adaptive Systems: Fundamentals through Simulations*. Wiley, New York, pp. 672.
- Rees, T., Dollar, H.S., Wathen, A.J., 2010. Optimal solvers for PDE-constrained optimization. *SIAM J. Sci. Comput.* 32, 271–298.
- Sano, Y., Fukagawa, K., Kuwahara, F., 2019. Numerical estimation of limiting current density by focusing on mass transfer within porous spacers in an electro-dialysis. *Membranes* 9, 75.
- Siddiqui, A., Lehmann, S., Haaksman, V., Ogier, J., Schellenberg, C., van Loosdrecht, M.C.M., Kruithof, J.C., Vrouwenvelder, J.S., 2017. Porosity of spacer-filled channels in spiral-wound membrane systems: quantification methods and impact on hydraulic characterization. *Water Res.* 119, 304–311.
- Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. *Adv. Neural Inf. Process Syst.*, 2951–2959.
- Staal, M., Farhat, N., van Loosdrecht, M., Vrouwenvelder, J., 2017. Biofouling patterns in spacer-filled channels: high-resolution imaging for characterization of heterogeneous biofilms. *Desalin. Water Treat.* 80, 1–10.
- Taamneh, Y., Bataineh, K., 2017. Improving the performance of direct contact membrane distillation utilizing spacer-filled channel. *Desalination* 408, 25–35.
- Ulbrich, M., Bloemen Waanders, B.V., 2018. An introduction to partial differential equations constrained optimization. *Optim. Eng.* 19, 515–520.
- West, S., Wagner, M., Engelke, C., Horn, H., 2016. Optical coherence tomography for the in situ three-dimensional visualization and quantification of feed spacer channel fouling in reverse osmosis membrane modules. *J. Membr. Sci.* 498, 345–352.
- Wilson, D.R., Martinez, T.R., 2003. The general inefficiency of batch training for gradient descent learning. *Neural Netw.* 16, 1429–1451.
- Wu, Z., Tran, A., Ren, Y.M., Barnes, C.S., Chen, S., Christofides, P.D., 2019a. Model predictive control of phthalic anhydride synthesis in a fixed-bed catalytic reactor via machine learning modeling. *Chem. Eng. Res. Des.* 145, 173–183.
- Wu, Z., Tran, A., Rincon, D., Christofides, P.D., 2019b. Machine-learning-based predictive control of nonlinear processes. Part II: computational implementation. *AIChE J.* 65, e16734.
- Wu, Z., Tran, A., Rincon, D., Christofides, P.D., 2019c. Machine learning-based predictive control of nonlinear processes. Part I: theory. *AIChE J.* 65, e16729.