Desalination 355 (2015) 56–67

Contents lists available at ScienceDirect

Desalination

journal homepage: www.elsevier.com/locate/desal

Big data analysis of hollow fiber direct contact membrane distillation (HFDCMD) for simulation-based empirical analysis

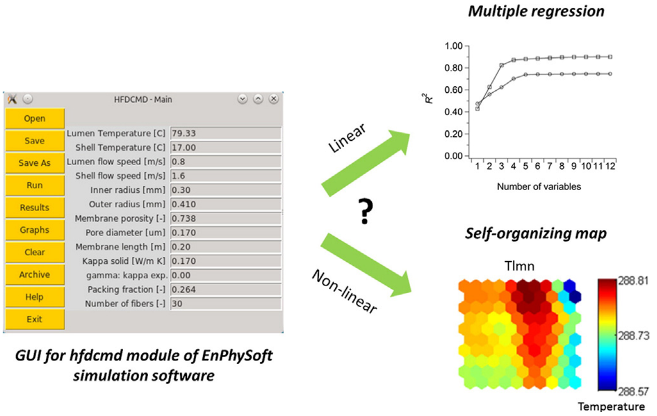
Seo Jin Ki [a], Hyeon-Ju Kim [b], Albert S. Kim [a,*]

[a] Civil and Environmental Engineering, University of Hawaii at Manoa, 2540 Dole Street Holmes 383, Honolulu, HI 96822, USA
[b] Seawater Utilization Plant Research Center, Korea Research Institute of Ships and Ocean Engineering, Goseong-gun, Gangwon-do 219-822, Republic of Korea

HIGHLIGHTS

- Simulations are done for hollow fiber direct contact membrane distillation (HFDCMD).
- Self-organizing map illustrated physical complexity of coupled mass/heat transfer.
- Multiple regression ranked the primary input parameters from big simulation data.
- Real and dimensionless data sets provide distinct insights of HFDCMD process.
- Performance of HFDCMD is most significantly influenced by membrane geometry.

GRAPHICAL ABSTRACT


HFDCMD - Main

Lumen Temperature [C] 79.33
Shell Temperature [C] 17.00
Lumen flow speed [m/s] 0.8
Shell flow speed [m/s] 1.6
Inner radius [mm] 0.30
Outer radius [mm] 0.410
Membrane porosity [-] 0.738
Pore diameter [um] 0.170
Membrane length [m] 0.20
Kappa solid [W/m K] 0.170
gamma: kappa exp. 0.00
Packing fraction [-] 0.264
Number of fibers [-] 30

GUI for hfdcmd module of EnPhySoft simulation software

ARTICLE INFO

Article history:
Received 10 August 2014
Received in revised form 5 October 2014
Accepted 9 October 2014
Available online xxxx

Keywords:
Hollow fiber direct contact membrane distillation
EnPhySoft simulation software
Self-organizing map
Multiple linear regression
Big data analysis
Multi-physical phenomena

ABSTRACT

A large number of hollow fiber direct contact membrane distillation (HFDCMD) are simulated using previously developed software, hfdcmd, i.e., a module of an environmental software package (EnPhySoft). Of 11,059,200 cases, 7,453,717 cases are found to be physically meaningful for practical applications. The self-organizing map (SOM) and multiple linear regression (MLR) methods were used to statistically analyze the big data. Using the raw data, physical and dimensionless data sets were prepared with specific formats: the former identifies the most significant parameters, and the latter compares relative importance between input parameters. SOM analysis did not provide transparent dependencies between inputs and/or outputs of HFDCMD: instead, it helped categorize parameters into groups of similar characteristics. Using MLR, we found that macroscopic quantities such as temperature and radii of lumen and shell sides were more influential on the MD performance than microscopic quantities such as pore size and membrane length. A rough (order-of-magnitude) prediction for heat and mass fluxes requires only four key input parameters.

1. Introduction

Membrane distillation (MD) is a non-isothermal process in which the partial pressure gradient of evaporated water vapor derives mass transfer with an accompanied transformation between liquid and gas phases [1]. MD received close attention due to its capability to produce

---

* Corresponding author. Tel.: +1 808 956 3718; fax: +1 808 956 5014.
E-mail address: albertsk@hawaii.edu (A.S. Kim).

http://dx.doi.org/10.1016/j.desal.2014.10.008
0011-9164/

clean (distillated) water using various energy resources including industrial waste, low-grade heat, and alternative methods. Rejection of solutes in MD processes is based on selective evaporation of water molecules while solutes stay in the liquid feed phase. In principle, the rejection ratio of MD is as high as that of reverse osmosis (RO), and sometimes even superior for the removal of arsenic and boron [2,3]. The current bottleneck of MD technology includes lower distillate (permeate) flux and membrane wetting. In cases of RO, applied pressure can be flexibly adjusted for required performance, but the driving force of MD is bound to the vapor pressure difference between hot feed and permeate streams. The theoretical maximum of the vapor pressure difference is the atmospheric pressure. To minimize heat transfer across the membrane, fast stream flow tangent to the membrane surface should be established, but high hydraulic pressure in membrane channels may cause the membrane wetting, i.e., liquid flow through membrane pores. Therefore, the liquid entry pressure needs to be carefully estimated based on physical and chemical properties of the membrane, under which hydraulic pressure will be applied to generate stable tangential flows.

Membrane distillation (MD), characterized by condensation phases and methods, includes direct contact membrane distillation (DCMD), vacuum membrane distillation (VMD), sweep gas membrane distillation (SGMD), and air gap membrane distillation (AGMD). DCMD has several advantages over the others: easy setup and operation, low resistance to mass and heat transfer, and no need to use an external condenser [1]. VMD usually provides higher distillate flux than that of DCMD and SGMD because the driving force is enhanced by low vacuum pressure ($1-3\%$ of atmospheric pressure) on the distillate side, and, more importantly, dry air is absent in pore spaces. SGMD collects evaporated water using a sweep gas (typically air) and condenses the mixed gas using external condensers. AGMD utilizes cold wall surfaces on which evaporated water condenses and forms dew. As water vapor molecules stay near the cold wall until the gas-to-liquid phase transformation is completed, the mass transfer resistance of AGMD is the highest compared to the other methods.

Hydrophobic MD membranes are often prepared as flat sheets. However, hollow fiber (HF) membranes are preferred in industrial applications because of high packing density, i.e., the total membrane surface area available per packed volume. HF membranes are geometrically characterized by pore size, (outer and inner) diameters, and length, which are of orders of $O(0.1)$ μm, $O(0.1)$ mm, and $O(0.1)$ m, respectively. Therefore, both ratios of the fiber length to diameter and diameter to pore size are on the order of $O(10^3)$. For example, when HFs (less than 1 mm diameter) are packed in a vessel of inner diameter 10 cm, the number of HFs can easily exceed 5000. These multiscale properties of hollow fiber direct contact membrane distillation (HFDCMD) processes make fundamental simulation studies of concurrent mass/heat transfer phenomena formidable.

Dominant transfer mechanisms of HFDCMD are heat/momentum transfer in the lumen and shell sides and heat/mass transfer in solid and void membrane regions. Characteristic parameters of HFDCMD processes can be categorized into three groups: (1) operational conditions (temperatures and flow rates (or speeds) of the lumen and shell inlets), (2) intrinsic membrane properties (pore size, porosity, tortuosity, and thermal conductivity), and (3) membrane and module geometry (inner and outer radii, length, and packing density). To overcome the above-mentioned fundamental difficulties based on multiple length scales of HFDCMD, a small number (of an order of $O(10)$) fibers were used for experimental [4] and modeling [5,6] studies.

A cylindrical cell model was developed for a large number of densely packed HFs to investigate correlations between vapor mass transfer and heat transfer [7]. It is a difficult task to thoroughly model MD phenomena using a finite number of HFs because of the geometrical complexity. In MD, mass transfer phenomena is strongly correlated with conductive and convective heat transfer through the solid and void parts of the membrane, respectively. The prediction of vapor flux (i.e., distillate production rate) requires an estimation of the heat transfer coefficient in the lumen and shell interfaces of temperature gradient. Although empirical correlations were often employed to predict the amount of trans-membrane heat transfer, these equations were developed for pure heat transfer along impermeable ducts using experimental and theoretical results, and then (recently) applied to coupled heat and mass transfer in HFDCMD (see Section 2.1 for details).

When a model is developed and verified with experimental data, a number of cases can be tested to collect physically meaningful results and to find optimal operation conditions. This approach can noticeably reduce the number of (costly) experiments in a variety of science and engineering disciplines. When a large amount of data is produced using a series of simulations, analysis of the big data is another major task. The National Big Data Research and Development Initiative was established by the US White House to help solve some of the Nation's most pressing challenges for scientific discovery, environmental and biomedical research, education, and national security [8]. It aims to make the most of the fast-growing volume of digital data and to improve knowledge and insights from the large and complex collections of digital data. Specific engineering processes such as MD have a variety of operation variables and conditions. Controlling parameters of DCMD easily exceeds 10 (see Section 2.2.2). If each parameter has 5 candidate values, the total number of operational cases is almost 10 million, i.e., $5^{10} = 9{,}765{,}625$. It is practically impossible to conduct this number of experiments within a limited amount of time to find optimal conditions. If a rigorous physical model that can predict engineering MD phenomena is available, key governing parameters of the MD performance and their interrelationships can be readily identified by analyzing the big data of MD simulations. Along these lines, we use our previous modeling tool, specifically developed for HFDCMD [7], to generate simulation outputs of an order of 10 million cases and analyze the big data using advanced statistical approaches, i.e., a self-organizing map (SOM) and a conventional multiple linear regression (MLR). In this paper, we aim to understand the relative significance of core parameters and their importance on mass/heat transfer phenomena, link conventional dimensionless numbers to the direct simulation results, provide phenomenological correlations for fast estimation of mass and heat transfer, and finally identify key controlling parameters.

2. Theory and simulations

2.1. System configuration

We consider a cylindrical vessel containing a large number of hollow fibers of an order of at least a few thousand. The fiber number and geometrical complexity make it difficult to simulate coupled heat and mass transfer phenomena occurring through individual fibers. We previously developed a model to analyze the performance of HFDCMD by considering a cylindrical unit cell, which was assumed to contain average physical and engineering characteristics [7].

We consider the hot-out/cold-in (HOCI) operation mode of HFDCMD, where the hot and cold streams flow in the shell and lumen regions, respectively. The opposite operational scheme is the hot-in/cold-out (HICO) mode. HOCI maximizes the flux with properly available heat sources, and HICO fully utilizes any available thermal energy in the feed stream. The hot feed stream flows with shell velocity v carrying thermal energy of temperature T_f (typically 40–80 °C). A cold permeate stream of temperature T_p (5–25 °C) flows in the opposite direction of the shell flow. This counter-current flow maintains the trans-membrane temperature gradient pseudo-constant along the fiber.

The thermal conductivity of membrane materials for MD is of an order of $O(0.1)$ W/m·K. The heat transfer from the hot feed to the cold permeate consists of conduction though the solid part of the membrane and convection accompanied by the vapor flux through the void membrane pores. Water evaporates near pore inlets at the feed-membrane interface and migrates to pore outlets on the permeate side. This evaporation precedes the transformation of hot water from its liquid to gas (vapor) phase, consuming the ambient thermal energy of the shell feed stream as much as the latent heat. Therefore, the vapor temperature at the pore inlet is close to the feed temperature in the local thermodynamic equilibrium. As vaporized water molecules diffuse through (tortuous) membrane pores, they carry thermal energy from the feed-membrane to the membrane-permeate interfaces. This increases the permeate temperature along the fiber. Viscous gas transfer is negligible in DCMD because gaseous water and air molecules are confined in pore spaces.

The counter-current cold stream obtains (i) the conducted heat at the solid membrane surface and (ii) the convected heat plus carried mass at the pore outlets. The transfer rate of the water vapor determines the MD performance with predetermined operation conditions. The hot feed stream loses as much heat as the cold stream receives, and therefore the feed/permeate stream temperatures decrease/increase along the fiber in the counter-current scheme. Because the lumen volume is usually smaller than the shell volume (per fiber), the lumen temperature varies more rapidly in the longitudinal direction than the shell temperature when lumen and shell flow speeds are similar. In this case, the dominant transfer mechanisms in each region are as follows: momentum and heat transfer in the lumen and shell regions, and heat and mass transfer in the void and solid membrane parts. The membrane mediates between lumen and shell regions for water evaporation influenced by stream temperatures. These mechanisms can be characterized quantitatively using representative dimensionless numbers (see Appendix A for details).

If N_f hollow fibers of inner and outer radii of a and b, respectively, are packed in a cylindrical vessel of radius R_v, then the packing fraction is calculated as $\phi = N_f b^2/R_v^2$, and average volume per each fiber is $\pi c^2 L$ where c $(=R_v/\sqrt{N_f})$ is the cell radius (see Fig. 1(a)). A conduct consisting of a solid rod in a cylinder has the same packing ratio and boundary conditions for heat transfer as shown in Fig. 1(b). Throughout this paper, the inner, shell, and cell surfaces are represented by $r = a, b,$ and c, respectively. Variables for the lumen and shell regions are also indicated using subscripts, lmn and shl, respectively. Fluid velocities are assumed to be zero on the inner and shell surfaces of the HF membrane, and tangential stress and radial temperature gradient are set to be zero on the cell surface. Mathematical details can be found in our previous work [7].

2.2. Simulations

2.2.1. HFDCMD simulation software

The simulation software is named hfdcmd as a part of EnPhySoft (Environmental Physics Software), which is written in FORTRAN90 using intrinsic object-oriented programming features. Note that we use uppercase (HFDCMD) and lowercase letters (hfdcmd) to indicate processes of hollow fiber DCMD and its simulation software (command name), respectively. Specific programming modules include functions and subroutines for Brownian, Knudsen, and effective diffusion coefficients with physical properties (including reference constants and temperature-dependent functions) of liquid water and water vapor. GNU FORTRAN (gfortran version 4.8.2) was used to compile the developed code set using the GNU make utility. The software hfdcmd uses an input file to read 11 parameters, as listed in Table 1. The EnPhySoft package is GNU licensed and open to the public. GNU Octave (version 3.8.1) was used for post-processing, 2D visualization, and image conversion. Among several releases of Linux distribution, Ubuntu (the latest version 14.04 named trusty) was selected for long-term, stable development.

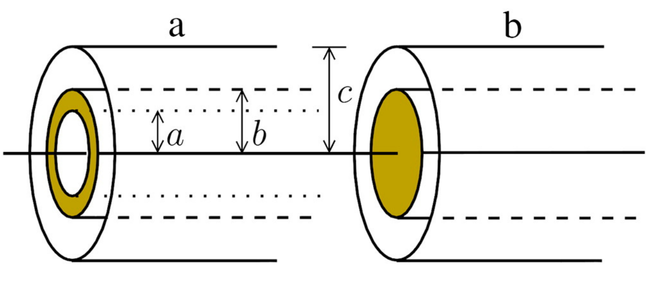


Fig. 1. Schematic of (a) a hollow fiber located in a imaginary cylindrical cell, which represents a number of densely packed fibers in a vessel and (b) conduct of heat transfer consisting of a solid rod inside a impermeable cylindrical pipe. Each has length L.

For easy access and use of hfdcmd, graphical user interface (GUI) is developed and included in the EnPhySoft package, as shown in Fig. 2(a). Tcl/Tk script is used to develop the hfdcmd GUI, which can open, save, and save as an input file, along with running, showing, archiving and clearing numerical and graphical results in an output window. A help file explains physical meanings of input and output variables. To expand the accessibility of this EnPhySoft package, Oracle VirtualBox was used to run an Ubuntu OS in Windows, Mac, and (other) Linux OSs. In Linux, EnPhySoft can be directly installed, but the use of VirtualBox is also possible. Fig. 2(b) shows the EnPhySoft VirtualBox installed on a Mac OSX. The general performance of the Ubuntu OS strongly depends on a (graphical) desktop environment due to the limited partial resources available from the main OS. Popular desktop environments such as GNOME and KDE are excellent, but their full implementation on a VirtualBox may significantly hamper the performance of operating systems. Instead, we use Lubuntu, which uses a fast and lightweight operating system with a minimal desktop, called LXDE. The hfdcmd software (as an independent application of EnPhySoft) is available by accessing the corresponding author's personal package archive.

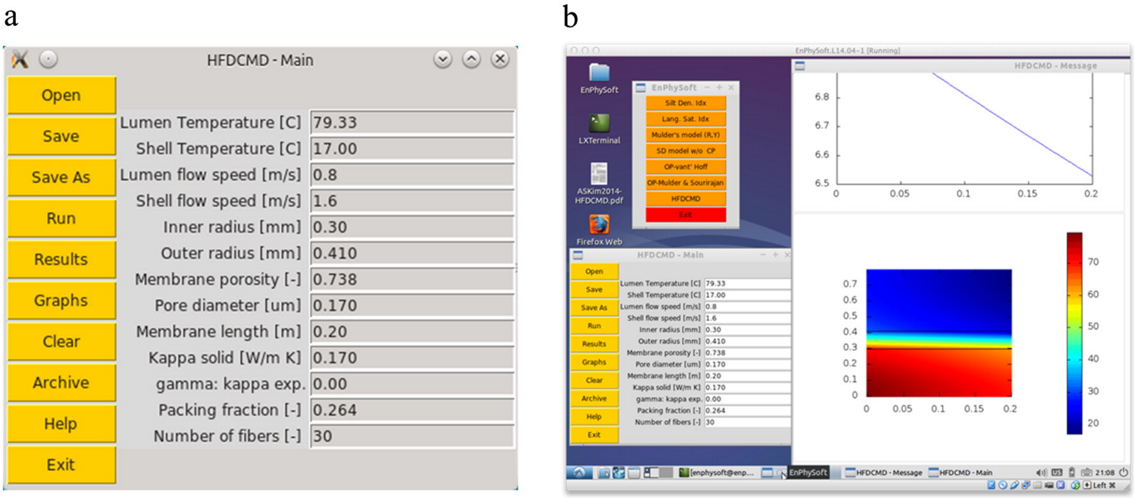
2.2.2. Cases and result selection criteria

A single run of hfdcmd generates 20 files (8 image and 12 data files), which are used by the Octave script to summarize and visualize simulation results. The total size of all 20 files is 1.2 MB. Two dimensional temperature profiles in the lumen, membrane, and shell regions are stored in 3 different files (each having 300 − 400 KB) as big as image files. To run a hfdcmd simulation, 11 parameters are required, as listed in Table 1. These are temperatures and fluid velocities of the hot feed and cold permeate streams (4), inner radius, thickness, pore diameter, porosity, length, and thermal conductivity of the HF membrane (6), and (fiber) packing fraction (1). The numbers within parentheses indicate the number of parameters in each category. We set 4–6 representative cases of the first 10 parameters, and 3 cases for the packing fraction. $\langle F_w \rangle$ and $\langle S_q \rangle$ are calculated using the simulation program and converted later to their dimensionless forms of Pe and Nu, respectively.

The total number of simulations is 11,059,200, which theoretically requires as large as 13.2 TB to store all simulation results. This is out of the typical hard drive size in modern PCs. Only core results can

Table 1
A list of 11 parameters for hfdcmd simulation. The number of simulations is a multiple of each number of cases, totaling 11,059,200.

Parameters	Values	Number of cases
Temperature in lumen [°C], T_{lmn}	5.0, 12.50, 20.0, and 25.0	4
Temperature in shell [°C], T_{shl}	40.0, 50.0, 60.0, 70.0, 80.0, and 90.0	6
Velocity in lumen [m/s], u	0.40, 0.80, 1.20, 1.60, and 2.00	5
Velocity in shell [m/s], v	0.40, 0.80, 1.20, 1.60, and 2.00	5
Inner radius of HF [mm], a	0.15, 0.25, 0.35, and 0.45	4
Thickness of HF [mm], δ_m	0.05, 0.15, 0.25, and 0.35	4
Pore diameter [μm], d_{p}	0.05, 0.15, 0.25, and 0.35	4
Porosity [ ], ε	0.70, 0.75, 0.80, and 0.90	4
Length of HF [m], L	0.20, 0.30, 0.40, 0.60, 0.80, 1.00	6
Thermal conductivity [W/m · K], κ	0.10, 0.17, 0.24, and 0.30	4
Packing fraction [ ], ϕ	0.42, 0.48, and 0.54	3



Lumen Temperature [C] 79.33
Shell Temperature [C] 17.00
Lumen flow speed [m/s] 0.8
Shell flow speed [m/s] 1.6
Inner radius [mm] 0.30
Outer radius [mm] 0.410
Membrane porosity [-] 0.738
Pore diameter [um] 0.170
Membrane length [m] 0.20
Kappa solid [W/m K] 0.170
gamma: kappa exp. 0.00
Packing fraction [-] 0.264
Number of fibers [-] 30

Fig. 2. Snapshots of (a) hfdcmd GUI and (b) Oracle VirtualBox.

be saved in a single output file, the size of which is about 3 KB. If only the core files are saved, then the 311 GB is required. Although this is a manageable disk size, the large number of files significantly slows down the OS performance. NTFS and ext2/ext3 (of Windows and Linux, respectively) can hold up to 4.3 billion files, but Linux Shell performance is noticeably degraded by handling a number of files, especially to read all files sequentially for statistical analyses. After each run of hfdcmd, a special script was used to extract core numerical results, stored in a line of a minimum file size. This extracted information was appended to a file that contains results of the previous runs. The size of this single file was reduced to 1.5 GB for the 7,453,717 physically meaningful cases.

To finish all the runs quickly, we used the basic parallel computing schemes for optimal job allocation. No specific parallel libraries (such as message passing interface (MPI) and open multiple processing (openMP)) were used, but the large number of jobs were instead divided into three groups based on the packing fraction values. Each group of simulations was run in different computing nodes, which in turn had two quad-cores (8 cores) of Intel (R) Xeon (R) CPU E5345 2.33 GHz and 16 GB of random access memory. The total of 3 identical nodes (in terms of hardware configuration) were available and so 24 ($= 3 \times 8$) hfdcmd simulations were simultaneously conducted at a time. Each node took 17.5 hours with 8 simultaneous jobs (i.e., one simulation job per core). This is equivalent to an elapsed time as long as $17.5 \times 3 \times 8 = 420$ hours (=17.5 days) of sequential runs using a single core.

Fig. 3 shows a flow chart of sequential simulation processes of hfdcmd, followed by SOM and MLR analyses. Using a combination of the 11 parameter values (see Table 1), hfdcmd simulation determines whether the assigned parameter set is physically acceptable in terms of the membrane length. If the input length is longer than the theoretical maximum, the case is discarded; otherwise, the simulation continues. This is because when a membrane is longer than the theoretical maximum (plus fluid speeds are slow), temperature polarization becomes significant so that a zone of zero transmembrane temperature gradient can be located in the middle of a hollow fiber. Although membrane distillation can continue, the performance is lower than expected because membrane surface areas are not fully utilized. Therefore, hfdcmd simulation program sorts out the physically less meaningful and practically inefficient cases before calculating mass and heat transfer phenomena. If the case is accepted, simulation continues and results are added to database files. Out of 11,059,200 cases, 32.6% was discarded and 67.4% was used for statistical analyses. The selected data were introduced to statistical tools after transformed into forms appropriate for each analysis (i.e., data proprocessing). Using these data sets, SOM and MLR provide graphical and numerical analyses, respectively. Based on map patterns of each parameter, SOM method allows us to group the input parameters with respect to their thermal, geometrical, and fluid dynamic influences on mass and heat transfer. MLR method not only quantifies the relative significance of different input parameters, but provides empirical equations to predict output variables. Details of each step are described in Section 2.3.

2.3. Big data analysis

2.3.1. Input data preparation

Two data sets were prepared from such massive simulations for statistical analyses: one using real variables of physical units (see Table 1) and another using reduced, dimensionless variables (see Appendix A). In the physical set, the tortuosity (τ) was calculated using Eq. (A.28) as a function of the membrane porosity (ε). The tortuosity solely

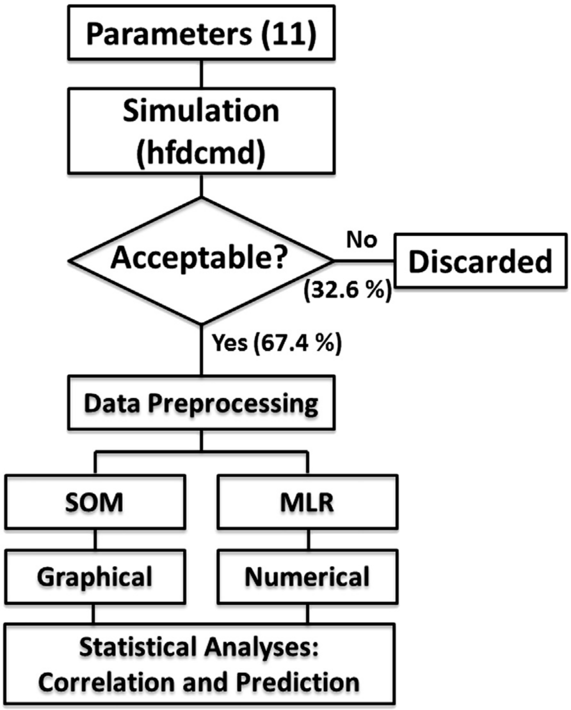


Fig. 3. Conceptual flow chart of hfdcmd and statistical simulations.

depends on the porosity so that their correlation must be apparently strong. Because $\langle F_w \rangle$ is strongly influenced by the porosity and tortuosity (more specifically, their ratio ε/τ), we intentionally included the tortuosity as the 12th input parameter. Two output parameters of this physical set were mass and heat transfer rate per fiber length, $\langle F_w \rangle$ and $\langle S_q \rangle$, respectively. In the dimensionless set, 12 new reduced variables were proposed: reduced shell temperature ($T_{\mathrm{shl}}/T_{\mathrm{lmn}}$), temperature polarization ($T_{\mathrm{shl}}/T_{\mathrm{lmn}} - 1$), effective porosity ($\varepsilon_{\mathrm{eff}} = \varepsilon/\tau$), structural factor (SF), and dimensionless numbers (Re and Pr of lumen and shell regions, Pe of all regions, and Nu of the membrane region only). Among them, Pe and Nu of the membrane region (denoted $\mathrm{Pe}_{\mathrm{mbr}}$ and $\mathrm{Nu}_{\mathrm{mbr}}$, respectively) were set as two output variables, corresponding to $\langle F_w \rangle$ and $\langle S_q \rangle$ of the physical set, respectively.

In summary, the total numbers of input variables were 12 and 10 in the physical and dimensionless data sets, respectively, which were used to assess correlation levels between variables and to predict corresponding output variables (i.e., transmembrane mass and heat transfer rates). The SOM and MLR were used to analyze these two big data sets generated from a series of hfdcmd numerical simulations. MATLAB (version 7.6) and IBM SPSS Statistics (version 21) were used for SOM and MLR analyses, respectively.

2.3.2. Self-organizing map (SOM)

SOM effectively identifies patterns and relationships between variables using an unsupervised learning algorithm and extracts information from sophisticated multi-dimensional data sets by creating noise-reducing data maps [9–11]. The SOM approximates the probability density function of the input data and represents the data in a more comprehensive fashion of fewer dimensions (2D or 3D). As the system stability is maintained through convergence to an equilibrium map, the SOM provides great flexibility in data presentation and comprehension [12], and can be applied to various engineering and science fields. Comprehensive reviews of SOM and its wide applications can be found elsewhere [13–17].

During the unsupervised learning, the input data are used to train a neural network without any prior knowledge of data structure. The data are naturally partitioned into different numbers of clusters depending on their own characteristics. Suppose that a data set contains n factors (i.e., n dimensions), where the density factor i is expressed as a vector x_i, i.e., an input layer to the SOM. This input node i is connected to other nodes ($j \neq i$). The connection weights, $w_{ij}(t)$, are updated adaptively at each iteration step t. This iterative process continues until the difference between input data x_i and the weight w_{ij}, defined as

$$d_j(t) = \sum_{i=0}^{N-1} \left(x_i - w_{ij}(t)\right)^2 \tag{1}$$

is fully minimized and so convergence is reached. Initial weight values are randomly assigned with small magnitudes. The neuron having the shortest distance to the input vector is chosen as a winner. The winner and neighboring neurons are allowed to change their weights to keep minimizing the distances. At each iteration the weighted vector is updated as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t) \cdot \lfloor x_i - w_{ij}(t) \rfloor Z_j \tag{2}$$

where $Z_j = 1$ for the winning and neighboring vectors, $Z_j = 0$ for other remaining vectors, and $\eta(t)$ is a fraction for corrected learning. In SOM analysis, all input data (or variables) were normalized into the range $[-1, 1]$ so that heterogeneous orders of input variable magnitude do not affect the neuron structure in a two-dimensional hexagonal grid. After preprocessing, data were processed by linear initialization and batch training algorithm, respectively [18]. The profile of representative samples of a variable (i.e., its distribution pattern) was finally visualized in individual component planes. The map size of SOM was determined by 10×10 neurons rather than the default size ($\approx 5\sqrt{n}$) to quickly investigate data patterns without loosing any generality (see Section 3.1 for details).

2.3.3. Multiple linear regression (MLR)

MLR is a statistical method to find the linear relationships between a dependent variable (y) and multiple independent variables (x_1, x_2, etc). MLR can generate an equation that best describes linear combinations between inputs and an output [16]. In this study, a logarithmic transformation was used to normalize the dependent variable to satisfy the homoscedasticity requirement, i.e., the variances of the output variable are normally distributed across the range of predictor variables. The stepwise regression method was used to select important predictor variables in the regression process using probabilities of 0.05 and 0.1 for entry and removal, respectively. The average inflation factor of the physical and dimensionless data sets was calculated and the values were close to 1.0. This indicates that multicollinearity (i.e., high correlation between variables) do not affect the validity of the regression equation and coefficients [16].

3. Results and discussion

3.1. Non-linear data analysis with large simulation data sets

3.1.1. SOM analysis for the physical data set

Fig. 4 shows spatial patterns of the 14 parameters of the physical data set, analyzed using the SOM method. The 12 data inputs include the calculated tortuosity, while the two outputs are the heat and mass transfer rates per finger length, $<F_w>$ and $<S_q>$, denoted Fw_bar and Sq_bar in Fig. 4, respectively. Because fancy mathematical notations are not available in SOM result visualizations, no subscripts and superscripts are used in Figs. 3 and 4. Each parameter map (i.e., component plane) represents a distribution of 100 samples, visually shown in the 10 by 10 hexagonal grid. This grid is reduced from the total 7,453,717 original samples. Color bars indicate ranges of parameter values and similar map patterns in general imply strong correlations. Here we first discuss the data partitioning. As shown in the fifteenth panel (at the bottom row, second to the right), the samples were partitioned into 6 different groups based on the characteristic similarity of the input parameters. The sixteenth panel shows the Davies-Bouldin index versus the number of clusters tested. This index evaluates clustering algorithms, and is minimized for the best clustering. In our study, six clusters will best explain the variation patterns of the HFDCMD parameters. Some SOM component panels have interesting patterns of parameter variables. Most of 12 input and 2 output parameter maps show gradient structures of vertical zigzag stripes.

In the first group, the shell temperature (Tshl), inner radius (Rinr), outer radius (Rotr), pore diameter (d_pore), membrane length, and thermal conductivity of membrane solid material (KappaSolid) have a structural pattern of blue-left to red-right, which has close similarity to $< S_q >$ (Sq_bar). In the second group, the lumen temperature (Tlmn) and packing fraction (pacfrac) have the vertical stripe structures, but color variations look disordered and oscillating, which indicates that they are less correlated with others. The third group consists of average lumen velocity (u_bar) and shell velocity (v_bar), which have panel maps very similar to each other, having a red-left to blue-right gradient of vertical zigzag stripes. The color trend of this third group is opposite to that of the first group. The fourth group has porosity (as the input parameter) and tortuosity (as the calculated parameter). The porosity map pattern is unique in having the horizontal gradient of blue-top to red-bottom stripes. As the tortuosity is inversely proportional to the porosity, its pattern follows red-top to blue-bottom horizontal stripes. These two parameters are the least correlated to others because they are most basic (rudimentary) physical characteristics of membrane material. As expected, the inverse relationship between the porosity and tortuosity is clearly shown in the map

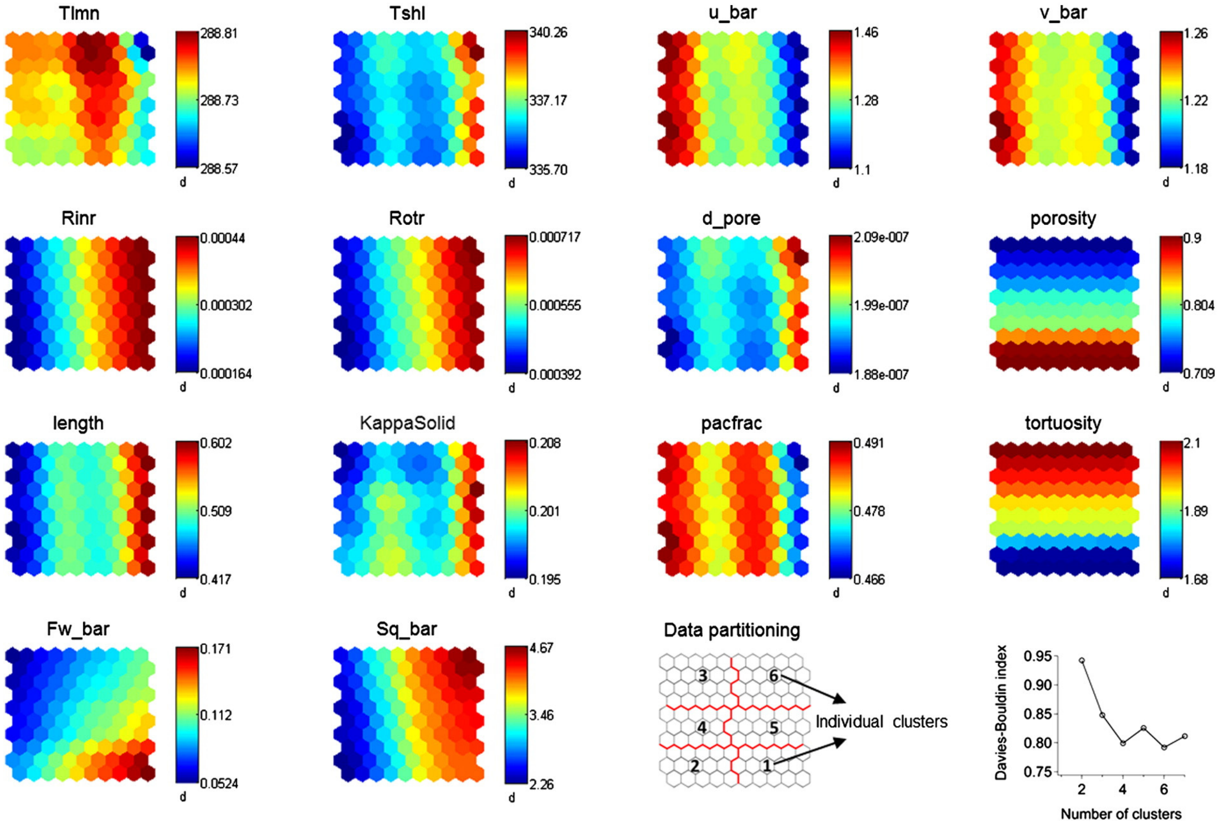


Fig. 4. Variation of variables (i.e., 14 component planes) of representative samples analyzed by a self-organizing map in a 10-by-10 hexagonal grid. In the figure, data partitioning (shown in the fifteenth panel) indicates that individual clusters of similar characteristics are classified into 6 groups based on the minimum Davies–Bouldin index (shown in the sixteenth panel).

patterns. The fifth group is the output group having Fw_bar and Sq_bar. Fw_bar has a unique map pattern, which looks similar to the vertically flipped Sq_bar. No apparent relationship is found between Fw_bar and other parameters; this implies that mass transfer rate cannot be simply predicted using only a few or even all the parameters, and combined influences of several input parameters nonlinearly determine Fw_bar. Even if a correlation is made, its accuracy must be limited. Although both Fw_bar and Sq_bar have the similar trend of blue-left to red-right vertical stripes, peak points are at vertically opposite locations, i.e., cluster 1 and 6 of Fw_bar and Sq_bar, respectively. Two maps of Rinr and Rotr resemble that of Sq_bar, indicating very strong correlations. There is not any specific map of the diagonal gradient from blue left-top to red right-bottom, which characterizes the map of Fw_bar. This implies that certain sets of parameter values to maximize Fw_bar and Sq_bar are not entirely consistent with each other.

On summary, the first group parameters include thermal (Tshl and KappaSolid) and geometrical (Rinr, Rotr, d_pore, and length) characteristics, which are closely related to the heat transfer rate, Sq_bar. The third group parameters, u_bar and v_var, represent fluid dynamic characteristics, strongly correlated to each other. However, this group shows an opposite pattern to Sq_bar, as compared to the first group. As Fw_bar is influenced by Sq_bar, it seems to be correlated, to some extent, with the lumen temperature (Tlmn) as a heat sink and packing fraction (pacfrac) representing compactness of heat source space. Increasing higher heat transfer rate, however, does not always guarantee monotonously linear increase in mass transfer rate. This illustrates the complexity of HFDCMD phenomena, so the current nonlinear SOM analysis seems limited in either identifying interactions between parameters or correlations to final outputs directly.

3.1.2. SOM analysis for the dimensionless data set

Fig. 5 shows results of SOM's analysis using the dimensionless data set of 12 total variables (10 inputs and 2 outputs). Similar to Fig. 4, this large data set is reduced to 100 representative samples presented in component planes with color bars. We include the reduced shell temperature (Tshl_red), temperature polarization (Tem_pol), and shell Prandtl number (Prandtl_shl) in the first analysis group of Fig. 5. Tshl_red and Tem_pol have the same gradient of the horizontal stripes: blue-top to red-bottom. Tshl_red represents the relative significance of the feed temperature over the permeate temperature, whereas Tem_pol implies temperature gradient as well as polarization across the hollow fiber membrane. Prandtl_shl also has the horizontal characteristics of reverse color scheme. The horizontal map patterns of these three parameters better resemble that of the membrane Peclet number (Peclet_mbr) in the positive or negative direction, which represents the mass transfer rate across the HF membrane per unit length. Next, vertical gradients are clearly shown in maps of the lumen Reynolds number (Reynolds_lmn), lumen Peclet number (Peclet_lmn), and shell Peclet number (Peclet_shl), all having the red-left to blue-right gradient. Reynolds_shl looks similar to the above three parameters with a slight diagonal pattern. This trend is opposite to the membrane Nusselt number (Nusselt_mbr) representing the heat transfer rate across the membrane.

The map of effective porosity (porosity_eff), i.e., porosity divided by the tortuosity, has an almost identical map pattern to that of Nusselt_mbr, indicating their strong correlation. The lumen Prandtl number (Prnadtl_lmn) also has a vertical stripe pattern but the gradient has a peak region in the right middle. Unlike other physical dimensionless numbers, the shell Prandtl number (Prandtl_shl) has the gradient structure of horizontal stripes. We believe this influences the semi-diagonal stripe structure of the shell Reynolds number (Reynolds_shl). The structural factor has the vertical stripe pattern, but the peak region is located near the center of the map. Compared to the effective porosity, SF is less correlated with Peclet_mbr and Nusselt_mbr. Two other less correlated parameters are Prandtl_lmn and Reynolds_shl. Note that Peclet_mbr is best correlated with Tshl_red and Tem_pol, and

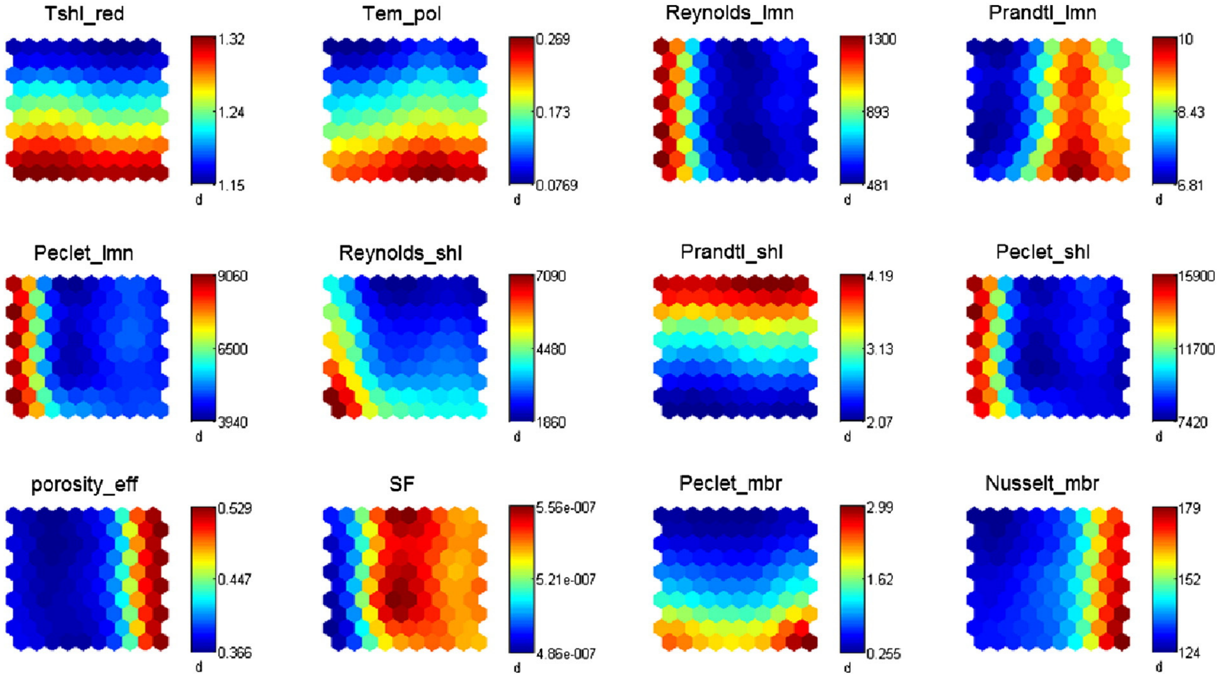


Fig. 5. Variation of dimensionless variables (i.e., 12 component planes) of representative samples analyzed by a self-organizing map in a 10-by-10 hexagonal grid.

Nusselt_mbr shows a specific similarity only to the effective porosity in the positive direction.

On summary, temperature difference or polarization, denoted as Tshl_red and Tem_pol, respectively, primarily governs the dimensionless mass transfer across the membrane, Peclet_mbr. Reynolds and Peclet numbers of lumen and shells are closely correlated, as they are proportional to u_bar and v_bar of the lumen and shell regions, respectively (see Fig. 4). On the other hand, the dimensionless heat transfer rate, Nusselt_mbr, is most significantly influenced by geometrical factors such as porosity_eff and structural factor, SF, in the positive way. We conclude that this SOM analysis of the dimensionless data set shows more compact and less-scattered map patterns in comparison to maps of the real data set (see Fig. 4), but barely provides new specific information in addition to those from the physical data set. We discuss this issue in Section 3.2 using the MLR method.

3.2. Linear data analysis with large simulation data sets

3.2.1. Multiple regression analysis for the physical data set

The physical and dimensionless data sets were reanalyzed using the MLR method to predict their output variables. In the physical data set, Fw_bar and Sq_bar were regressed using varying number of parameters from 1 to 12, whereas Peclet_mbr and Nusselt_mbr were regressed using 1 to 10 dimensionless variables. Fig. 6 shows predictive capabilities of the MLR method in terms of R^2 values of each predicted variable, i.e., Fw_bar, Sq_bar, Peclet_mbr and Nusselt_mbr. The R^2 values of the four dependent variables increase as the number of independent variables increases from 1 to 4. Using more than 5 dependent variables does not significantly enhance the predictability of the MLR method. At least four variables are required to explain noticeable variations of Fw_bar, Sq_bar, and Peclet_mbr, whereas three parameters are enough to predict Nusselt_mbr.

Table 2 shows coefficients of four predicted variables in the form of

$$\log_{10}(10+y) = B_0 + \sum_{i=1}^{4} B_i x_i \tag{3}$$

where B_0 is a constant and B_i for $i = 1$–4 are (unstandardized) coefficients of independent variables shown in Table 2 in sequence. In the physical data set, Fw_bar is reasonably well predicted using four independent variables: Tshl, d_pore, Rinr, and Rotr with $R^2 = 0.7$. Increasing the number of independent variables does not noticeably increase the prediction accuracy, as shown in Fig. 6. Sq_bar has key parameters of Tshl, Rinr, Rotr, and Tlmn with $R^2 = 0.88$, which looks almost saturated.

The unstandardized coefficients for individual variables in regression models describe the strength of the relationship between independent and dependent variables (for example, as the shell temperature Tshl increases by one, (log-transformed) Fw_bar increased by 1.80×10^{-4} units). Validity of this prediction using unstandardized coefficients are quantified using standard error (SE) values, which are much smaller than unstandardized coefficients, B_i. On the other hand, the standardized coefficients for individual variables provide standard units of measure, i.e., their significances in the models as predictors.

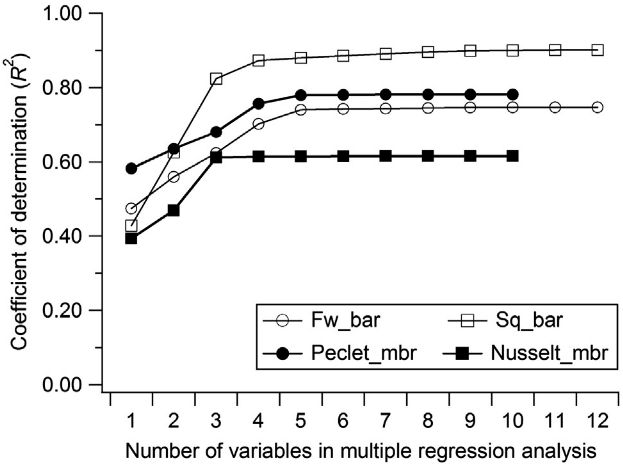


Fig. 6. The fraction of variance (i.e., coefficient of determination R^2) obtained by the regression equation with different numbers of predictor variables.

Table 2
Results of multiple regression analysis with four predictor variables.

Models[a]	Variables	Unstandardized coefficients		Standardized coefficients
		B	SE[b]	β[c]
Fw_bar	Constant	0.94	1.77×10^{-05}	
	Tshl	1.80×10^{-04}	5.15×10^{-08}	0.70^{*}
	Rinr	23.78	1.26×10^{-02}	0.59^{*}
	Rotr	-14.04	1.00×10^{-02}	-0.44^{*}
	d_pore	11704.18	7.83	0.30^{*}
Sq_bar	Constant	0.81	3.76×10^{-04}	
	Rinr	600.31	0.13	0.98^{*}
	Rotr	-339.03	0.10	-0.70^{*}
	Tshl	2.62×10^{-03}	5.13×10^{-07}	0.67^{*}
	Tlmn	-1.94×10^{-03}	1.15×10^{-06}	-0.22^{*}
Peclet_mbr	Constant	0.40	1.51×10^{-04}	
	Tshl_red	0.48	1.17×10^{-04}	0.74^{*}
	Peclet_lmn	7.59×10^{-06}	3.94×10^{-09}	0.59^{*}
	Peclet_shl	-3.61×10^{-06}	2.36×10^{-09}	-0.47^{*}
	porosity_eff	0.12	9.39×10^{-05}	0.23^{*}
Nusselt_mbr	Constant	1.84	4.05×10^{-04}	
	Peclet_shl	-1.28×10^{-05}	6.31×10^{-09}	-0.79^{*}
	Peclet_lmn	1.72×10^{-05}	1.05×10^{-08}	0.63^{*}
	porosity_eff	0.68	2.51×10^{-04}	0.62^{*}
	Tshl_red	0.07	3.14×10^{-04}	0.05^{*}

[a] Dependent variable y was transformed using a logarithmic function, $\log_{10}(y + 10)$, when performing the regression analysis.
[b] SE indicates the standard error of unstandardized coefficients B.
[c] Asterisk(*) indicates the p-value is less than 0.001.

Although the unstandardized coefficients of Tshl for Fw_bar and Sq_bar are small (1.80×10^{-4} and 2.62×10^{-3}, respectively), its standardized coefficients are the highest (0.70) for Fw_bar and the second highest (0.67) for Sq_bar. Signs of the unstandardized and standardized coefficients are identical to each other. The negative sign indicates an inverse relationship between dependent and independent variables. For example, as Rotr increases, both Fw_bar and Sq_bar decrease, keeping other independent variables fixed. This is because a thicker HF membrane reduces both heat and mass transfer rates.

In the prediction of Fw_bar of the physical set, the most significant parameters among all 12 are ranked Tshl, Rinr, Rotr, and d_pore, in order. Because the operation mode is HOCI, the shell temperature is higher than the lumen temperature in a wide range of 15–85 °C. A higher Tshl definitely increases both mass and heat transfer rates with similar significances, characterized by the standardized coefficients, 0.70 and 0.67, respectively. The inner radius (Rinr) has stronger influence than the outer radius (Rotr) in predicting both Fw_bar and Sq_bar, and the strongest to Sq_bar with $\beta = 0.98$. β value differences between Rinr and Rotr (in magnitude) must stem from the cylindrical geometry of the azimuthal symmetry. The least influencing parameters on Fw_bar and Sq_bar are d_pore and Tlmn, respectively. Mass transfer rate per unit length, Fw_bar, is primarily restricted by the vapor evaporation rate at the interface between hot feed and (outer) membrane surface; and it is inversely proportional to the membrane thickness, i.e., Rotr subtracted by Rinr. Given that, pore spaces of an order of $d_{\text{pore}}^{2} \times (\text{Rotr} - \text{Rinr})$ geometrically determine the vapor transfer rate. d_pore is ranked the fourth in the MLR of Fw_bar. Interestingly, this indicates that the lumen temperature is not included in Fw_bar prediction and so is less important than the pore size of HF membrane. In addition, the least influencing parameter of Sq_bar is the lumen temperature (Tlmn) with the negative coefficient. The heat transfer rate is directly proportional to the temperature difference, i.e., Tshl – Tlmn. The radial geometry of HFs is more important in Sq_bar, followed by temperatures. d_pore in Fw_bar disappears in Sq_bar, as replaced by Tlmn. Comparing the fourth parameters (i.e., d_pore and Tlmn in predictions of Fw_bar and Sq_bar, respectively) shows physical insight that mass transfer is primarily initiated by the hot feed temperature and the heat transfer is mostly governed by, as expected, the temperature difference.

3.2.2. Multiple regression analysis for the dimensionless data set

As indicated above, Peclet_mbr and Nusselt_mbr represent dimensionless forms of mass and heat transfer rates, respectively. The dimensionless data set has four common, most significant parameters, indirectly implying that the dimensionless analysis can provide the compact forms of physically meaningful results. These parameters are ranked in order: Tshl_red, porosity_eff, Peclet_lmn, and Peclet_shl for predicting Peclet_mbr and porosity_eff, Peclet_shl, Peclet_lmn, and Tshl_red for regressing Nusselt_mbr. Among them, Peclet_shl always has the negative sign, ranked the first and third in terms of β of Nusselt_mbr and Peclet_mbr, respectively. Note that the lumen and shell Peclet numbers are related to thermal diffusion, whereas the membrane Peclet number is based on vapor diffusion through membrane pores.

In predicting Peclet_mbr, the reduced shell (feed) temperature Tshl_red is found to be the most significant parameter with $\beta = 0.74$, which is highly analogous to $\beta = 0.70$ of Tshl in Fw_bar regression. Tshl_red has the least impact on the Nusselt_mbr prediction as expected from the fact that Tshl and Tlmn in Sq_bar regression are ranked the third and fourth, respectively. The noticeably small β value of Tshl_red explains the plateau trend of Nusselt_mbr plot (in Fig. 6) after the number of independent variables is more than three. It is interesting that Tem_pol does not show up in predicting any dimensionless numbers for membrane transport. The effective porosity appears in both regression equations, and is ranked third and fourth of membrane Nusselt and Peclet numbers. Note that it was absent in Fw_bar and Sq_bar predictions. This effective porosity is as important as Peclet_lmn ($\beta = 0.63$) in the Nusselt_mbr prediction. The structural factor does not show up as an important input parameter in the analysis of the dimensionless set. In both physical and dimensionless sets, microstructure of the membrane is found to be secondarily important as compared to macroscopic radial geometry and operational conditions.

The lumen and shell Peclet (thermal) numbers appear to be key controlling parameters of both membrane Peclet and Nusselt numbers. The lumen Peclet number has a stronger influence in transmembrane mass transfer than the shell Peclet number, but has less impact on heat transfer than the shell Peclet number. Given conditions in the shell side, as the lumen Peclet number increases, mass and heat transfer rates increase because the lumen velocity and hydraulic diameter are linearly proportional to the lumen Peclet number (see Eq. (A.18)). As the (only) lumen temperature increases, the kinematic viscosity decreases so that the lumen Peclet number increases. In most empirical correlations used for heat transfer phenomena of DCMD processes, viscosity and density are often assumed to be fixed at specific temperatures. As these temperature-dependent fluid mechanics properties are included in the dimensionless analysis of the current study, implications easily become complex to understand. The lumen Peclet number increases as the lumen velocity and lumen temperature increase: the former reduces the temperature polarization on the lumen surface at $r = a$, while the latter reduces the temperature gradient followed by the vapor concentration gradient across the membrane. Contributions of the lumen velocity and temperature to the mass and heat transfer through the lumen Peclet number are not consistent with physical insight due to the temperature dependent kinematic viscosity.

The same analysis can be applied to the effect of the shell Peclet number on mass/heat transfer phenomena (see Eq. (A.19)). Given the shell velocity, a higher HF packing reduces volumetric flow rate in the shell side, allowing more temperature polarization in both tangential and normal directions near the outer membrane surface. Variation of the shell temperatures (40–80 °C) is much more than that of the lumen temperatures (5–25 °C). In this range of shell temperatures, the kinematic viscosity decreases below 50 percent of the standard value. Similar to the lumen Peclet number, fast velocity, low packing ratio, and high temperature in the shell side will increase the shell Peclet number. This is because lower kinematic viscosity (due to higher shell temperature) is equivalent to faster shell velocity. A lower packing ratio indicates a higher volumetric flow rate in the shell region, which diminishes the temperature polarization on the outer membrane surface at $r = b$. In general, increasing the shell Peclet number should result in higher mass and heat transfer rates. In this aspect, it looks appropriate that signs of β of shell and lumen Peclet numbers are exchanged.

This seemingly untenable analysis can be resolved by explicitly considering the reduced shell temperature (Tshll_red), which includes differences between the shell and lumen temperatures, i.e., the driving force of the transfer phenomena. There is no doubt that feed temperature and feed-distillate temperature differences play a key role in mass and heat transfer, respectively. Dimensionless numbers such as the Reynolds, Prandtl, and Peclet numbers characterize fluid properties, which vary with temperature. Variation of fluid density, viscosity, and thermal diffusivity and conductivity influence the lumen and shell dimensionless numbers. These must not be as important as lumen and shell temperatures to transfer phenomena, but their trends are counterintuitive.

We indicate that HFDCMD is purely multi-physical phenomena in which mass/heat/momentum transfer are strongly inter-correlated in a sophisticated manner. Variable reduction and functional mapping might not be straightforward to quantitatively predict the performance of DCMD. In the physical data analysis, d_pore (used for the Fw_bar prediction) is the only parameter that microscopically characterizes the membrane. In the dimensionless analysis, effective porosity is commonly included for membrane Peclet and Nusselt numbers. More importantly, dimensionless analysis uses the same set of four independent parameters. From these established equations, one can rapidly compare influences of the four dimensionless input parameters to mass and heat transfer rates using their standardized coefficients.

4. Concluding remarks

Using the previously developed software, we generated big data to test the HFDCMD performance of various cases. Of the total 11,059,200 cases, 3,605,483 cases were dropped due to the partial availability of the hollow fiber membrane.

In the SOM analysis, the physical data set was classified into six subgroups based on their data similarity, and the correlation degrees of each input parameter to the heat and mass fluxes were easily identified by graphically comparing map patterns. No single parameter is found to closely resemble a map of the mass transfer rate, which confirms the complexity of the HFDCMD phenomena. Instead, heat transfer is better characterized, having similar map patterns of temperatures and velocities of lumen and shell sides. The dimensionless group shows more homogeneous map patterns than those of the physical set using the dimensionless variables constructed by the same set of important input parameters. The mass and heat fluxes are represented as the membrane (thermal) Peclet and Nusselt numbers, respectively. The Peclet number map resembles those of reduced temperature and temperature polarization, whereas the Nusselt number map has similarity to that of the effective porosity.

The MLR method showed that only three or four input parameters were enough to predict the performance of HFDCMD within a certain degree of accuracy. In a practical sense, this selects the most significant input parameters, providing a straightforward approach to understanding the complex performance characteristics. Using the physical data set, we found that macroscopic parameters such as temperature and radii of lumen and shell sides mostly controlled the MD performance. The only exception is the (microscopic) pore diameter influencing the mass flux. The dimensionless data set has four common parameters for both mass and heat flux prediction. A large temperature variation in the shell side changes fluid properties such as mass density and viscosity, competing with effects of fluid velocities. Although the lumen and shell Peclet numbers are significant parameters, reduced shell temperature and effective porosity have the highest beta values for mass and heat fluxes, respectively.

Overall, mass transfer is primarily controlled by the feed temperature, whereas the heat transfer is modulated by the temperature difference between the feed and distillate streams. When the temperature-dependent fluid properties are used to calculate the conventional dimensionless numbers, the increasing temperature provides equivalent effect of increasing fluid velocity. This indicates that the conventional dimensionless analysis may require fundamental revisions to include not only fluid- and thermo-dynamic effects but also temperature-dependent fluid properties.

Acknowledgement

This research was supported by the National R&D project of “Development of Energy Utilization Technology of Deep Sea Water Resource” supported by the Ministry of Oceans and Fisheries of the Republic of Korea.

Appendix A. Dimensionless number analysis for HFDCMD

A.1. Overview

In MD processes, the heat transfer rate can be expressed as

$$Q_i = h_i \Delta T_i \tag{A.1}$$

where $i = f$, m, and p indicate feed, membrane, and permeate sides, respectively. On the same line, ΔT_f, ΔT_m, and ΔT_p are temperature differences between the bulk feed and feed-membrane interface, across the membrane of thickness $\delta_m(=b-a)$, and between the permeate-membrane interface and bulk permeate stream, respectively. As temperature is a point function in the cylindrical cell, h_i and ΔT_i are often implied as length-averaged quantities. Within the HF membrane, we have $h_m = h_{m0} + J_w H_w / \Delta T_m$, where h_{m0} is the pure heat transfer coefficient of the porous membrane and H_w is the evaporation enthalpy of water. Phenomenologically, the vapor flux is proportional to the vapor pressure difference between the hot feed and cold permeate:

$$J_w \sim C_m \left[P_f - P_p\right] \tag{A.2}$$

where C_m is the membrane mass transfer coefficient and P is the partial pressure of the vapor gas. The complexity comes from the coupling of Q_i and J_w through h_m. In the steady state, Q_i in each region should be equal to each other because there is neither heat source nor sink in the vessel. Then, one writes

$$Q = Q_i \cong \frac{T_f - T_p}{h_f^{-1} + h_m^{-1} + h_p^{-1}} \tag{A.3}$$

Estimation of h_f and h_p is a crucial step for the performance analysis because they vary with geometrical and fluid dynamic properties. Empirical correlation previously developed using heat transfer experiments (or simulations) is well accepted in MD literature. These correlations represent the Nusselt number (Nu) as a function of the Reynolds (Re), Prandtl (Pr), and Grashoff (Gr) numbers as well as the ratio of hydraulic diameter (d_h) to channel length (L).

Gryta et al. [19] summarized 13 correlations commonly used for the estimation of the Nusselt number in non-porous cylindrical conducts under laminar flow (see Fig. 1(b)) and concluded that heat transfer for MD carried out in plate-and-frame modules could be best quantified by

$$\mathrm{Nu} = 0.74\mathrm{Re}^{0.2}(\mathrm{Gr\,Pr})^{0.1}\,\mathrm{Pr}^{0.2} \tag{A.4}$$

For laminar flow ($\mathrm{Re} < 2100$) of hollow fiber and capillary modules, Gryta and Tomaszewska [20] suggested two correlations:

$$\mathrm{Nu} = 4.36 + \frac{0.023\ \mathrm{Pe}(d_h/L)}{1 + 0.0012\ \mathrm{Pe}(d_h/L)} \tag{A.5}$$

and

$$\mathrm{Nu} = 4.36 + \frac{0.036\ \mathrm{Pe}(d_h/L)}{1 + 0.0011[\mathrm{Pe}(d_h/L)]^{0.8}} \tag{A.6}$$

and reported that Eq. (A.6) provided a better match than Eq. (A.5) with experimental observation. Some researchers used two different correlations to estimate separately heat transfer coefficients in feed and permeate sides. Kim et al. [21] used Eq. (A.6) for the tube (lumen) side stream and used Groehn's correlation for the shell side of a hollow fiber module [22]:

$$\mathrm{Nu} = 0.206(\mathrm{Re}\cos\theta)^{0.63}\,\mathrm{Pr}^{0.36} \tag{A.7}$$

where θ is the yaw angle. Similarly, Edwie and Chung [23] used two different empirical correlations for heat transfer coefficients of the hot feed and cold distillate streams for HFDCMD. Phattaranawik et al. [24] and Ho et al. [25] used Eq. (A.6) for the flat-sheet MD modules. As mentioned above, these correlations were interchangeably used for the flat-sheet and hollow-fiber modules by employing the specific hydraulic diameter. Selection of specific correlations seems to be subjective to best explain experimental observations. A recent summary of empirical correlations for heat transfer used for MD can be found elsewhere [26].

The inner and outer radii of hollow fiber membranes (i.e., a and b, respectively) are of an order of $O(0.1)$ mm and the thickness δ_m has a comparable size to a. This implies that the curvature effect of the hollow fiber needs to be carefully considered in solving governing equations. In a steady state ($\partial/\partial t = 0$), heat and mass flux, q and J_w, across the membrane satisfies

$$\nabla\cdot q = 0 \ \text{ and } \ \nabla\cdot J_w = 0 \tag{A.8}$$

respectively. Because there is neither source nor sink of heat and mass in the vessel, divergences of q and J_w must vanish. One can assume that heat transfer across the HF membrane is predominant in the radial direction (indicated using the subscript r):

$$q \approx q_r = -\kappa\frac{\partial T}{\partial r} + \overline{H}J_w \tag{A.9}$$

where T is the absolute temperature within membrane, κ is the thermal conductivity of the porous membrane and $\overline{H}(=\overline{H}_0 - \overline{c}_pT)$ is the effective molar enthalpy using $\overline{H}_0 = 24.527$ kJ/mol and $\overline{c}_p = 5.2434 \times 10^{-2}$ kJ/mol·K. Fick's law may represent the vapor flux through pore spaces characterized using porosity ε and tortuosity τ of the membrane:

$$J_w = -\frac{\varepsilon}{\tau}D_{\mathrm{eff}}\frac{\partial n_w}{\partial r} \tag{A.10}$$

where n_w is the molar concentration of water vapor in pore spaces [mol/m^3] and D_{eff} is the effective diffusion coefficient, obtained by combining Brownian and Knudsen diffusivity using Bosanquet's relationship [7,27,28]. General solutions for the coupled Eqs. (A.9) and (A.10) were analytically obtained using perturbation theory, and an iterative method was used to solve for local and length-averaged profiles of heat and mass fluxes along the hollow fiber [7].

A.2. Lumen and shell regions

Fluid mechanics properties in the lumen and shell regions include stream speeds (i.e., $\boldsymbol{u}$ and $\boldsymbol{v}$ for the lumen and shell, respectively) and hydraulic diameter (d_h). Thermodynamic properties of water are density (ρ), dynamic viscosity (η), kinematic viscosity (ν), specific heat at constant pressure (c_p), and thermal conductivity (k_w) as functions of temperature.

A.2.1. Hydraulic diameter

As discussed above, heat transfer correlations used for MD were developed using non-porous conducts, as shown in Fig. 1(b). A cylindrical rod of radius b is inside a thermally insulating impermeable pipe of radius c. The hydraulic diameter is defined as

$$d_h = 4\frac{V}{S_w} \tag{A.11}$$

where $V\ (=\pi(c^2 - b^2)L)$ is the liquid volume and $S_w\ (=2\pi(b + c)L)$ is the wetted surface area by the liquid. Then, we calculate

$$d_h = 2(c-b) = (2b)\left(\frac{1}{\sqrt{\phi}} - 1\right) \tag{A.12}$$

where $\varphi = b^2/c^2$ is the volume (or packing) fraction of the rod inside the pipe. Fig. 1(a) and (b) have similar characteristics at $r = c$. Our previous work assumed $\partial T/\partial r|_{r=c} = 0$ and $\partial v/\partial r|_{r=c} = 0$ (in Fig. 1(a)), of which the former indicated thermally insulated surface of the rod-containing pipe (in Fig. 1(b)). Therefore, the effective diameter for heat transfer must be calculated using conducting surface only such as $S_c = 2\pi bL$ (instead of wetted surface). Then we have

$$d_c = (2b)\left(\frac{1}{\phi} - 1\right) \tag{A.13}$$

which is greater than d_h because $S_w > S_c$ from $\varphi < 1.0$. In summary, the characteristic diameters of the shell and lumen regions are

$$d_{\mathrm{shl}} = d_c \tag{A.14}$$

and

$$d_{\mathrm{lmn}} = 2a \tag{A.15}$$

respectively.

A.2.2. Reynolds number

The momentum transfer is characterized using Reynolds number, defined as a ratio of inertial and viscous forces:

$$\mathrm{Re} = \frac{Ud_h\rho}{\eta} = \frac{Ud_h}{\nu} \tag{A.16}$$

where the kinematic viscosity of water ν varies with fluid temperature: 1.307 (at 10 °C), 0.553 (at 50 °C), and 0.326 (at 90 °C) in 10^{-6} m^2/s. This noticeable dependence of ν on stream temperatures significantly influences the performance of MD in terms of mass and heat fluxes (see Appendix B for functional forms of ρ and η with respect to temperature T). In Eq. (A.16), U is the (cross-section averaged) flow speed and d_h is the hydraulic diameter. The Reynolds number includes hydrodynamic characteristics in the longitudinal direction of the fiber, and is calculated for lumen and shell regions, denoted $\mathrm{Re}_{\mathrm{lmn}}$ and $\mathrm{Re}_{\mathrm{shl}}$, respectively. In the lumen region, $U = \boldsymbol{u}$, $d_h = 2a$, and $\boldsymbol{\nu} = \boldsymbol{\nu}(T_{\mathrm{lmn}})$; in the shell region, $U = \boldsymbol{v}$, $d_h = d_c$, and $\boldsymbol{\nu} = \boldsymbol{\nu}(T_{\mathrm{shl}})$, where $\mathrm{T}_{\mathrm{lmn}}$ and $\mathrm{T}_{\mathrm{shl}}$ are the lumen and shell temperatures, respectively.

A.2.3. Prandtl and Peclet numbers

The ratio of viscous to thermal diffusion rate is defined as the Prandtl number:

$$\mathrm{Pr} = \frac{\nu}{\alpha} = \frac{c_p\eta}{k_w} \tag{A.17}$$

where $\alpha = k_w/\rho c_p$ [m^2/s] is the thermal diffusivity, c_p [J/kg·K] is the specific heat at constant pressure, and k_w [W/m·K] is the thermal conductivity of water [29] (see Appendix B for their dependence on temperature). The Prandtl number only includes viscous and thermal properties of water, which are independent of module geometry and fluid dynamics. A product of the Prandtl and Reynolds numbers is thermal Peclet number, i.e., a ratio of advective transport to thermal diffusion: Pe = Re Pr. In lumen and shell regions, the Peclet numbers are denoted as

$$\mathrm{Pe}_{\mathrm{lmn}} = \frac{u \cdot 2a}{\nu(T_{\mathrm{lmn}})} \tag{A.18}$$

and

$$\mathrm{Pe}_{\mathrm{shl}} = \frac{v \cdot 2b\left(\phi^{-1}-1\right)}{\nu(T_{\mathrm{shl}})} \tag{A.19}$$

respectively, where the kinematic viscosity ν is represented as a function of temperature.

A.3. Solid and void membrane regions

A.3.1. Nusselt number

The ratio of convective to conductive heat transfer is represented as the Nusselt number:

$$\mathrm{Nu} = \frac{hL_c}{k_m} \tag{A.20}$$

where (by definition) h [W/m^2·K] is the convective heat transfer coefficient, L_c [m] is the characteristic length, and k_m [W/m·K] is the membrane heat conductivity. Here, h can be interpreted as the heat flux at the feed-membrane interface per unit temperature and L_c is the outer diameter of the fiber, $2b$. The absence of the inner diameter $2a$ in the Nusselt number is due to empirical correlations specifically developed for conducts as shown in Fig. 1(b). In a cylindrical coordinate system (as applied to HFDCMD in this paper), heat/mass fluxes defined as heat/mass transfer rates per unit surface area do not provide consistent understanding and measures of transfer phenomena across the hollow fiber membrane. The product hL_c as a whole (instead of h and L_c separately) better explains characteristics of HFDCMD heat transfer. The reason is as follows. In the steady state, the heat equation in the radial direction simplifies to

$$\frac{1}{r}\frac{\partial}{\partial r}(rq) = 0 \tag{A.21}$$

which indicates that rq = constant. Conversely, q decreases with respect to r from the inner ($r = a$) to outer ($r = b$) surfaces. Therefore, we let

$$hL_c = \frac{2\left\langle S_q\right\rangle}{\Delta T_f} \tag{A.22}$$

where $\langle S_q\rangle$ is the longitudinal average of S_q ($=rq$), which represents the amount of heat lost in the feed stream per unit fiber length per unit time. The coefficient 2 in Eq. (A.22) doubles the radial distance, i.e., diameter as corresponding to $L_c = 2b$. In the steady state, $\langle S_q\rangle$ is equal to the heat obtained by the cold stream per unit fiber length per unit time. Here we define the Nusselt number for HFDMCD as

$$\mathrm{Nu} = \frac{2\left\langle S_q\right\rangle}{k_m \Delta T_f} \tag{A.23}$$

A.3.2. Peclet number

The vapor molecules diffuse through tortuous membrane pores from the feed-membrane to the permeate-membrane interfaces. As they carry thermal energy, the Peclet number is defined as

$$\mathrm{Pe} = \frac{J_w L_c}{D} \tag{A.24}$$

where D is the diffusion coefficient of vapor molecules in the membrane pore. Similar to hL_c of Eq. (A.22), $J_w r$ is radially constant because of

$$\frac{1}{r}\frac{\partial}{\partial r}(rJ_w) = 0 \tag{A.25}$$

Then, we can rewrite the Peclet number as

$$\mathrm{Pe} = \frac{\langle F_w\rangle}{D_{\mathrm{eff}}} \tag{A.26}$$

where $\langle F_w\rangle = \langle J_w r\rangle$ indicates the amount of vapor mass transferred across the membrane per unit fiber length per unit time.

A.3.3. Structural factor

A hollow fiber membrane used for DCMD can be physically characterized using pore diameter (d_p), thickness (δ_m), porosity (ε), tortuosity (τ), and fiber length (L). A thin HF membrane with high porosity, low tortuosity, and large diameter must provide high flux. Then, the structural factor can be defined as

$$\mathrm{SF} = \frac{\varepsilon}{\tau}\frac{d_p}{\delta_m} \tag{A.27}$$

where the tortuosity is usually a decreasing function of porosity [30]. We use Beeckman's tortuosity expression [31]:

$$\tau = \frac{\varepsilon}{1-(1-\varepsilon)^{1/3}} \tag{A.28}$$

which is specifically developed for highly interconnected pores. Details of tortuosity discussions can be found elsewhere [7,30].

In summary, the Reynolds and Prandtl numbers characterize physical and fluid dynamic properties in the lumen ($\mathrm{Re}_{\mathrm{lmn}}$ and $\mathrm{Pr}_{\mathrm{lmn}}$) and shell ($\mathrm{Re}_{\mathrm{shl}}$ and $\mathrm{Pr}_{\mathrm{shl}}$) regions. The Nusselt and Peclet numbers with the structural factor are used only for the membrane region.

Appendix B. Thermal properties of water

B.1. Density (ρ)

Water density as a function of temperature and salinity can be calculated using empirical formula [32]:

$$\rho(T)\left[\mathrm{kg/m^3}\right] = 1000\left(1-\frac{(T+288.9414)(T-3.9863)^2}{508929.2(T+68.12963)}\right) \tag{B.1}$$

where T is in Celsius.

B.2. Dynamic and kinematic viscosities

Vogel equation predicts the absolute water viscosity as a function of temperature:

$$\eta(T)[\mathrm{Pa\cdot s}] = \exp\left(-10.6266+\frac{578.919}{T-137.546}\right) \tag{B.2}$$

where temperature T in Kelvin. The kinematic viscosity $\nu(T)$ [m^2/s] $= \eta(T)/\rho(T)$ can be easily calculated using Eqs. (B.1) and (B.2).

B.3. Thermal conductivity

The expression of thermal conductivity of water [29] is

$$k_w = 0.6065 \times \left(-1.48445 + 4.12292T^* + 1.63866T^{*2}\right) \tag{B.3}$$

where $T^* = T\,[\mathrm{K}]/298.15\,[\mathrm{K}]$ for $274 < T < 370$ [K].

B.4. Specific heat at constant pressure

The specific heat of water c_p has insensitive values of 4.204 (at 5 °C), 4.183 (at 20 °C), 4.178 (lowest, at 30 °C), 4.183(at 55 °C), and 4.208 (at 90 °C) with respect to temperature. On average, we used $c_p = 4.190$ [J/kg · K].

References

[1] M. Khayet, T. Matsuura, Membrane distillation: principles and applications, Elsevier, Amsterdam, The Netherlands, 2011. 512.
[2] D. Qu, J. Wang, D.Y. Hou, Z.K. Luan, B. Fan, C.W. Zhao, Experimental study of arsenic removal by direct contact membrane distillation, J. Hazard. Mater. 163 (2009) 874–879.
[3] D.Y. Hou, J. Wang, X.C. Sun, Z.K. Luan, C.W. Zhao, X.J. Ren, Boron removal from aqueous solution by direct contact membrane distillation, J. Hazard. Mater. 177 (2010) 613–619.
[4] K.Y. Wang, T.S. Chung, M. Gryta, Hydrophobic PVDF hollow fiber membranes with narrow pore size distribution and ultra-thin skin for the fresh water production through membrane distillation, Chem. Eng. Sci. 63 (2008) 2587–2594.
[5] J.G. Lee, W.S. Kim, Numerical modeling of the vacuum membrane distillation process, Desalination 331 (2013) 46–55.
[6] G.Q. Guan, X. Yang, R. Wang, R. Field, A.G. Fane, Evaluation of hollow fiber-based direct contact and vacuum membrane distillation systems using aspen process simulation, J. Membr. Sci. 464 (2014) 127–139.
[7] A.S. Kim, Cylindrical cell model for direct contact membrane distillation (DCMD) of densely packed hollow fibers, J. Membr. Sci. 455 (2014) 168–186.
[8] Executive Office of the President (EOP), Obama administration unveils “Big data” initiative: announces $200 million in New R&D investments, EOP, Office of Science and Technology Policy, Washington, DC, 2012. 4 (Available online at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf).
[9] T. Kohonen, Self-organized formation of topologically correct feature maps, Biol. Cybern. 43 (1982) 59–69.
[10] T. Kohonen, Analysis of a simple self-organizing process, Biol. Cybern. 44 (1982) 135–140.
[11] T. Kohonen, Self-organizing maps, 3rd edition, Springer Series in Information Sciences, vol. 30, Springer, Berlin, Germany, 2001. 502.
[12] H. Ritter, K. Schulten, Convergence properties of Kohonen’s topology conserving maps: fluctuations, stability, and dimension selection, Biol. Cybern. 60 (1988) 59–71.
[13] H. Ritter, T. Martinetz, K. Schulten, Neural computation and self-organizing maps – an introduction, Addison-Wesley, New York, 1992. 306.
[14] S. Kaski, J. Kangas, T. Kohonen, Bibliography of self-organizing map (SOM) papers: 1981–1997, Neural Comput. Surv. 1 (1998) 1–176.
[15] S.J. Ki, J.H. Kang, S.W. Lee, Y.S. Lee, K.H. Cho, K.G. An, J.H. Kim, Advancing assessment and design of stormwater monitoring programs using a self-organizing map: Characterization of trace metal concentration profiles in stormwater runoff, Water Res. 45 (2011) 4183–4197.
[16] A. Field, Discovering statistics using SPSS, 2nd edition Sage Publications, Inc., Thousand Oaks, CA, 2005. 779.
[17] A.M. Kalteh, P. Hiorth, R. Bemdtsson, Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application, Environ. Model Softw. 23 (2008) 835–845.
[18] J. Vesanto, J. Himberg, E. Alhoniemi, J. Parhankangas, SOM Toolbox for MATLAB 5, Report A57, SOM Toolbox Team, Helsinki University of Technology, Espoo, Finland, 2000, p. 59 (Available online at: http://www.cis.hut.fi/somtoolbox/package/papers/techrep.pdf.).
[19] M. Gryta, M. Tomaszewska, A.W. Morawski, Membrane distillation with laminar flow, Sep. Purif. Technol. 11 (1997) 93–101.
[20] M. Gryta, M. Tomaszewska, Heat transport in the membrane distillation process, J. Membr. Sci. 144 (1998) 211–222.
[21] Y.D. Kim, K. Thu, N. Ghaffour, K. C. Ng, Performance investigation of a solar-assisted direct contact membrane distillation system, J. Membr. Sci. 427 (2013) 345–364.
[22] H.G. Groehn, Influence of the yaw angle on heat transfer and pressure drop of tube bundle heat exchangers, Proceedings of the 7th International Heat Transfer Conference, Munich, Germany, 1982 (Paper HX8).
[23] F. Edwie, T.S. Chung, Development of hollow fiber membranes for water and salt recovery from highly concentrated brine via direct contact membrane distillation and crystallization, J. Membr. Sci. 421–422 (2012) 111–123.
[24] J. Phattaranawik, R. Jiraratananon, A.G. Fane, Effect of pore size distribution and air flux on mass transport in direct contact membrane distillation, J. Membr. Sci. 215 (2003) 75–85.
[25] C.D. Ho, C.H. Huang, F.C. Tsai, W.T. Chen, Performance improvement on distillate flux of countercurrent-flow direct contact membrane distillation systems, Desalination 338 (2014) 26–32.
[26] M. Khayet, Membranes and theoretical modeling of membrane distillation: A review, Adv. Colloid Interf. 164 (2011) 56–88.
[27] A.S. Kim, A two-interface transport model with pore-size distribution for predicting the performance of direct contact membrane distillation (DCMD), J. Membr. Sci. 428 (2013) 410–424.
[28] C.H. Bosanquet, The optimum pressure for a diffusion separation plant, British TA Report, 1944 (BR/507).
[29] M.L.V. Ramires, C.A.N. Decastro, Y. Nagasaka, A. Nagashima, M.J. Assael, W.A. Wakeham, Standard reference data for the thermal-conductivity of water, J. Phys. Chem. Ref. Data 24 (1995) 1377–1381.
[30] B. Ghanbarian, A.G. Hunt, R.P. Ewing, M. Sahimi, Tortuosity in porous media: a critical review, Soil Sci. Soc. Am. J. 77 (2013) 1461–1477.
[31] J.W. Beeckman, Mathematical description of heterogeneous materials, Chem. Eng. Sci. 45 (1990) 2603–2610.
[32] S.C. McCutcheon, J.L. Martin, T.O. Barnwell Jr., Water quality, in: D.R. Maidment (Ed.), Handbook of Hydrology, McGraw-Hill, New York, NY, 1993 (Chapter 11).