Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2020.25156

195 (2020) 8–18
August

Effect of data preprocessing on the artificial neural network (ANN) analysis of fouling rate in microfiltration membranes

Bomin Kim[a], Yoonjin Kim[b], Jaehyun Ju[a], Yongjun Choi[a], Sangho Lee[a,*]

[a]*School of Civil and Environmental Engineering, Kookmin University, 77 Jeongneung-ro, Seoungbukgu, Seoul 02707, Korea, Tel. +82-2-910-4529; Fax: +82-62-910-4939; email: sangholee@kookmin.ac.kr (S. Lee)*
[b]*Bukang Tech Co., Ltd., Yuseong Daero 1184-25, Yuseong-gu, Dajeon, Korea*

Received 1 July 2019; Accepted 18 November 2019

ABSTRACT

Hollow fiber microfiltration membranes have been broadly adopted for drinking water production, wastewater treatment, and pretreatment of reverse osmosis (RO). Nevertheless, membrane fouling is a critical issue for the operation and maintenance of hollow fiber microfiltration processes and its prediction is still challenging. In this study, fouling of hollow fiber microfiltration membrane was investigated by applying artificial neural network (ANN) technique. A single fiber filtration unit was used to evaluate the rate of fouling resistance change with time (dR_f/dt) in constant flux operation. Results showed that the ANN models failed to fit the dR_f/dt values from the experimental data obtained in 1 min interval. Since this is attributed to large variations of the collected data, several techniques for data preprocessing were applied to improve the data quality. First, the removal of the outliers from the collected data was attempted but failed to increase the accuracy of the model fit. Further trimming of the data values that fall below the 25th percentile and above the 75th percentile could improve the model fit. However, excessive filtering of the data resulted in a poor model fit due to the oversimplification of the data. On the other hand, an increase in the data collection interval was found to be effective to improve the model fit accuracy. These results strongly suggest that proper preprocessing of the data is essential for the analysis and prediction of membrane fouling by ANN models.

Keywords: Hollow fiber; Microfiltration; Membrane fouling; Artificial neural network; Model fit; Data quality; Data preprocessing

1. Introduction

Microfiltration (MF) using hollow fiber membranes has been widely accepted as an efficient method to produce high-quality drinking water treatment [1–4]. This offers an attractive means to remove particulate contaminants such as clay, algae, bacteria, and pathogenic protozoa from drinking water [5–7]. Another major advantage of hollow fiber membranes over other configurations of membranes is the high membrane surface area to footprint ratio achieved by the low aspect ratio (diameter-to-length ratio) of fibers [8,9]. Since the late 1990s, a lot of full-scale membrane plants have been operated to the increasingly stringent regulatory requirements [6,7,10].

However, membrane fouling is still a critical problem for the hollow fiber microfiltration process [11]. The contaminants deposit or attach on the membranes on which they are retained during drinking water treatment [12]. As a result, the productivity of drinking water is significantly reduced with time and the lifespan of the membrane is shortened [9]. Fouling behavior is influenced by membrane surface properties, the nature of the particulate or dissolved foulant, water solution properties, etc. [7]. Inadequate management of the hydrodynamics is also a major factor that

* Corresponding author.

aggravates membrane fouling owing to the non-uniform profile of local fluxes along the hollow fiber membrane [13].

In the constant flux operation, where the applied pressure is adjusted to maintain the same flux with time, membrane fouling causes an increase in transmembrane pressure (TMP), as illustrated in Fig. 1. Accordingly, the value of $d(\mathrm{TMP})/dt$ is a measure to quantify the propensity of fouling. According to the resistance-in-series model, the TMP(ΔP) is given as a function of fouling resistance (R_f) [4]:

$$\mathrm{TMP}(t) = \Delta P(t) = J\eta\left(R_m + R_f(t)\right) \tag{1}$$

where t is the time (s), J is the permeate flux (L/m^2 h or m/s), η is the viscosity (Pa-s), and R_m is the intrinsic membrane resistance (m^{-1}). Accordingly, the fouling propensity is expressed as dR_f/dt.

$$\frac{d(\mathrm{TMP})}{dt} = \frac{d(\Delta P(t))}{dt} = \eta J \frac{dR_f(t)}{dt} \tag{2}$$

Understanding the fouling behavior of hollow fibers is important in order to improve the operation and design of the hollow fiber system [11,14,15]. This has led to a substantial number of studies on fouling mechanisms and its control [6,16–23]. Nevertheless, it is still challenging to predict MF fouling in most cases due to its complex characteristics [24–26]. Mathematical models have been proposed for its theoretical interpretation and prediction [4,7,20,27–29]. Recently, a handful of studies on the application of statistical models including artificial neural network (ANN) [25,30] and support vector machine [24,31] have also been carried out. But the potential of these techniques has not been adequately demonstrated [14,15].

In this study, fouling behaviors of hollow fiber microfiltration membrane were assessed using ANN technique. A special focus was on the effect of data quality on the accuracy of ANN models. Several methods to process the data obtained from laboratory-scale experiments were attempted and compared. The novelty of this work lies on the understanding of the capability and limitation of ANN models in connection with the issues on the quality of the collected data. This will provide sights on the fouling analysis and prediction in hollow fiber MF processes.

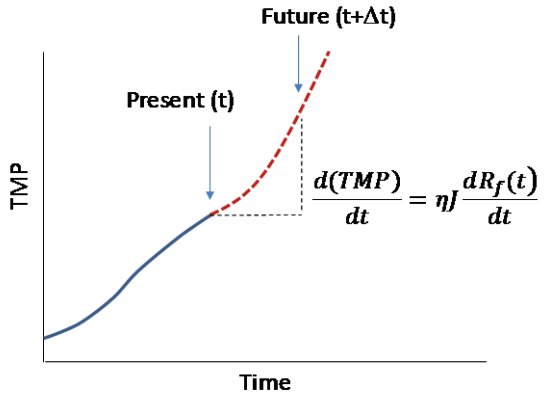

Fig. 1. Basic concept for the prediction of transmembrane pressure (TMP) using the rate of fouling resistance (dR_f/dt).

2. Material and methods

2.1. Experimental setup

A schematic diagram of a laboratory-scale, submerged hollow fiber membrane system for accelerated fouling test is shown in Fig. 2. The system consisted of multiple filtration tanks, allowing the simultaneous testing of MF fibers at the same time. Each tank had a working volume of 1 L and MF fiber was immersed vertically in the reactor. A magnetic stirrer was positioned just below the membrane and the stirring speed was controlled by a magnetic stirrer plate.

Table 1 summarizes the properties of the membrane used in this study. The MF fibers were made of polyvinylidene fluoride (PVDF) with the nominal pore size of 0.1 μm. They had an internal diameter of 0.7 mm and an external diameter of 1.2 mm. The length of the fiber was 200 mm. Since the fiber was relatively short, the pressure drop along the fiber was neglected.

Permeate from the membrane module was pulled by a peristaltic pump (EW-07551-00, Cole-Parmer, USA). A permeate volume was frequently measured by collecting permeate volume using amass cylinder. The TMP was continuously measured using a pressure transducer (ISE40A-01-R, SMC, Japan) and a data logger (USB-6008, NI, U.S.A.), which were connected to a computer. The temperature of solution was kept constant at 20°C. Total recycle mode, where both the retentate from the MF loop and permeate were recycled into the tank, was adopted to keep the reactor volume constant during the operation time. The conditions for the MF experiments are listed in Table 2.

2.2. Model foulants and test conditions

Model foulant used in this study was alginate (Sigma-Aldrich, Korea, alginic acid sodium salt from brown algae), which intends to simulate MF fouling by algae in raw water. The concentrations of the foulants were 2, 5, 10, 20 mg/L, respectively. The flux was adjusted between 50 and 200 L/m^2 h. Prior to each filtration test, all membranes were stabilized using deionized water during 500 min.

2.3. Development of ANN model

In this study, an ANN model with a multi-layer perceptron network that had a back-propagation training algorithm was used to fit the rate of fouling resistance change with time (dR_f/dt). ANN are computational models able to simulate the processing and learning functions of a human brain [30,32]. In accordance with the human brain, an ANN model is composed of simple elements operating in parallel [14]. Neurons in a certain layer of the ANN are connected to those from the previous layer by a number of weighted connections. In addition, there is an extra weight, named bias, which is summed to the rest of the input weights [30]. The neuralnet package in the R software was used to develop ANN models using the experimental data. RPro+ algorithm was used with the maximum step of 1,000,000.

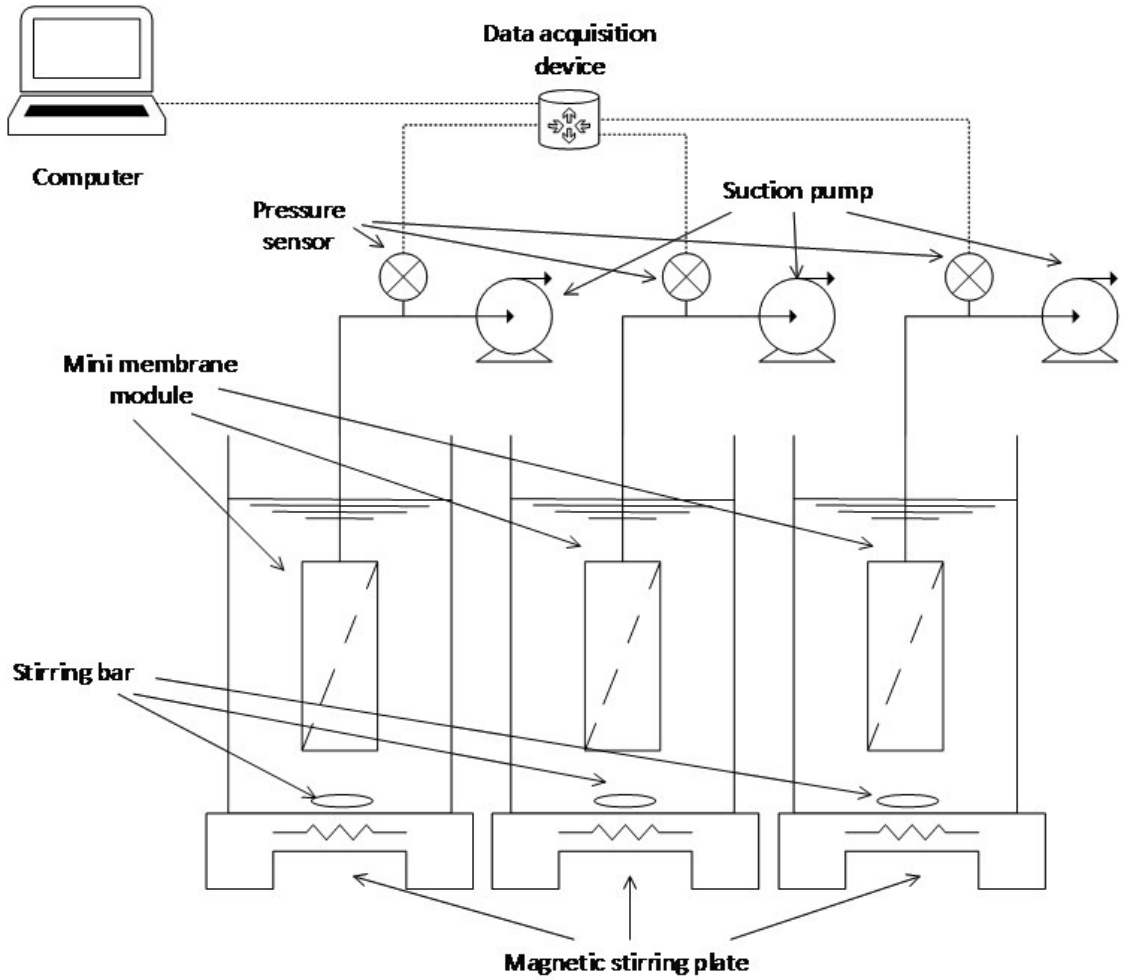

Fig. 2. Experimental set-up for submerged microfiltration (MF).

Table 1
Properties of MF membrane

	Properties
Membrane type	Hollow fiber
Membrane material	Polyvinylidene difluoride (PVDF)
Pore size (μm)	0.1
Length (mm)	200
Effective area (mm^2)	753.98
Outer diameter (mm)	1.2

The training of ANN model was carried out by adjusting the connection value among elements in order to minimize its performance factor defined as the mean squared error.

Foulant concentration, flux, time, TMP were selected as independent variables to predict dR_f/dt. It should be noted that dR_f/dt and TMP are independent. Since the purpose of this ANN model is to analyze the dynamic behavior of MF fouling, which is expressed as dR_f/dt. On the other hand, TMP simply represents a status of MF membrane at a given time. Nevertheless, TMP should be also considered as the independent variable because it may affect fouling rate due to foulant compaction. Input data were presented to the network through the input layer.

3. Results and discussions

3.1. Experimental data

A series of MF experiments were carried out by varying the foulant concentration and flux as shown in Table 2. The filtration results are illustrated in Fig. 3. The interval for the data collection was 1 min. When the foulant concentrations were low (Figs. 3a and b), the increase in TMP with time was not significant. As the foulant concentration increased up to 20 mg/L (Fig. 3c), the TMP rapidly increases. In addition to the foulant concentration, the flux also affected the rate of TMP increase. As expected, the TMP increased faster at higher flux than at lower flux, which is attributed to an increase in driving force for foulant deposition.

Using the data in Fig. 3, the dR_f/dt values were calculated and the results are shown in Fig. 4. Unlike the TMP, there seems to be no tendency in the dR_f/dt. At low foulant concentration (Fig. 4a), the variations of the data were high. They seem to be reduced with an increase in the foulant

Table 2
Experimental conditions

Concentration (mg/L)	Flux (L/m² h)	Operation time (min)
2	50	360
	100	
	150	
	200	
5	50	
	100	
	150	
	200	
10	50	
	100	
	150	
	200	
20	50	
	100	
	150	
	200	

concentration. The effect of flux on the dR_f/dt was not clearly found. The dR_f/dt should increase with an increased flux. Nevertheless, the flux effect was hidden by large variations of the data.

3.2. Development of ANN model

The ANN model was attempted to be derived using the data in Fig. 4. However, it failed to fit the experimental data as illustrated in Fig. 5. Although the experimental dR_f/dt changes from -164.5×10^7 1/m s to 157.8×10^7 1/m s, the model prediction was almost constant, suggesting that the training of ANN was not successful. Many attempts were carried out by changing the model parameters (i.e., interaction/layer, neurons per layer, threshold, maximum steps, etc.) for ANN training but all of them were unsuccessful. This suggests that the ANN technique was not suitable to deal with the data from our MF experiments.

This is attributed to the properties of the input data. As mentioned earlier, the input data has large variations. The fitting of the normal distribution to the data indicates that the mean and standard deviation are 1.107×10^7 1/m s and 29.7×10^7 1/m s, respectively (Fig. 6a). Compared with

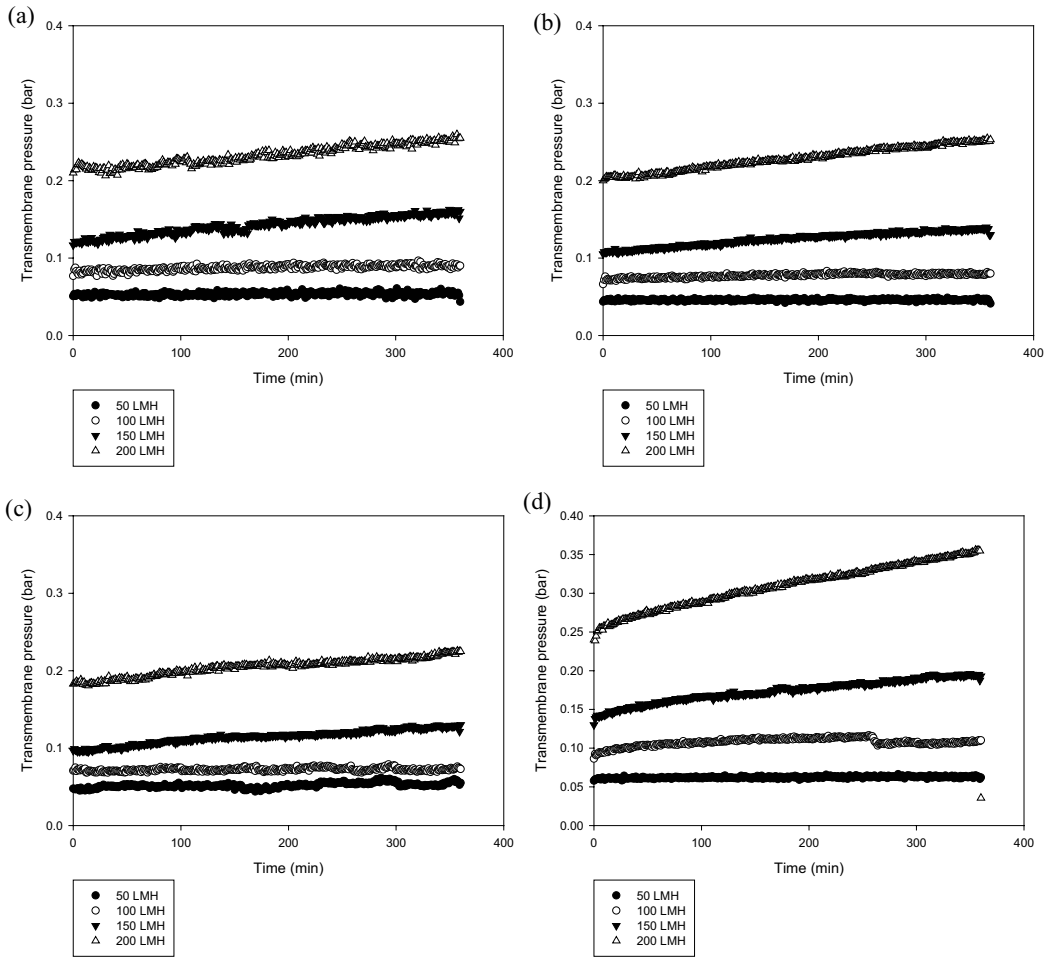

Fig. 3. Dependence of TMP on time during MF treatment of alginate solution with different initial concentrations. (a) 2 mg/L, (b) 5 mg/L, (c) 10 mg/L, and (d) 20 mg/L.

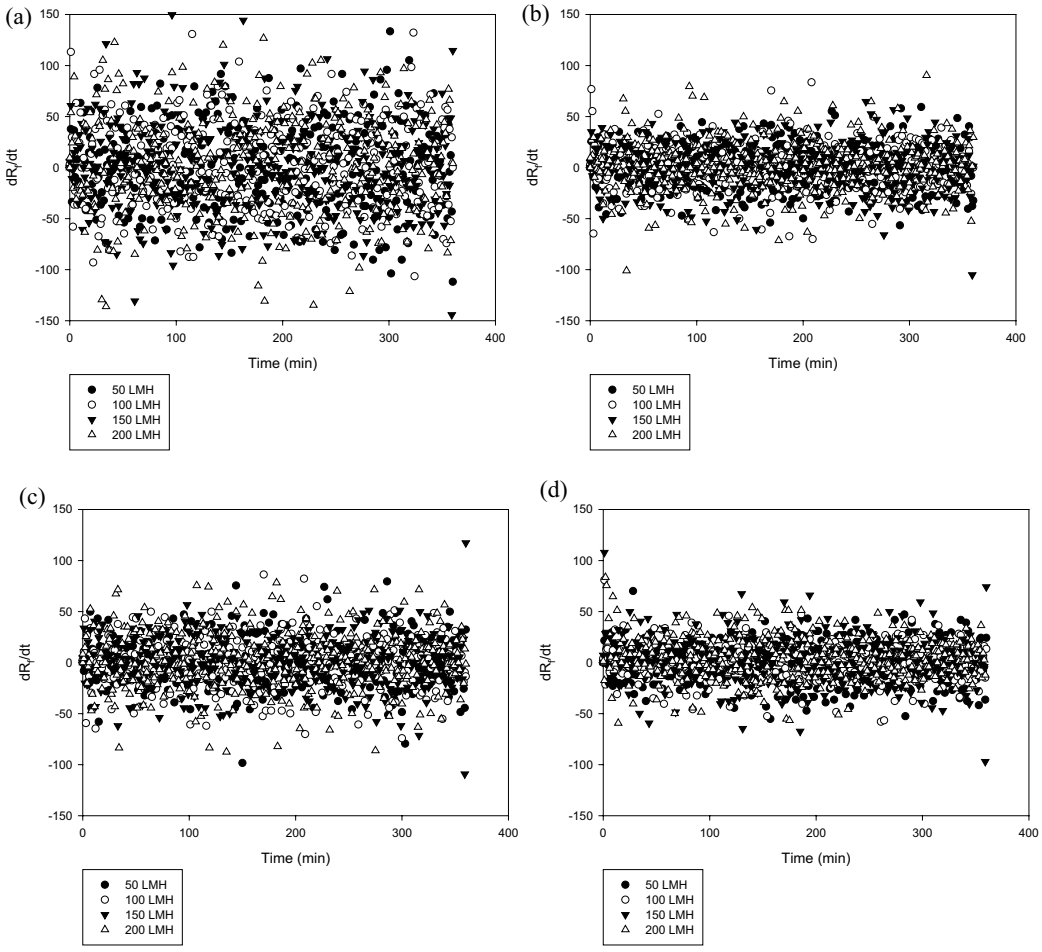

Fig. 4. Variations of dR_f/dt with time during MF treatment of alginate solution with different initial concentrations. (a) 2 mg/L, (b) 5 mg/L, (c) 10 mg/L, and (d) 20 mg/L.

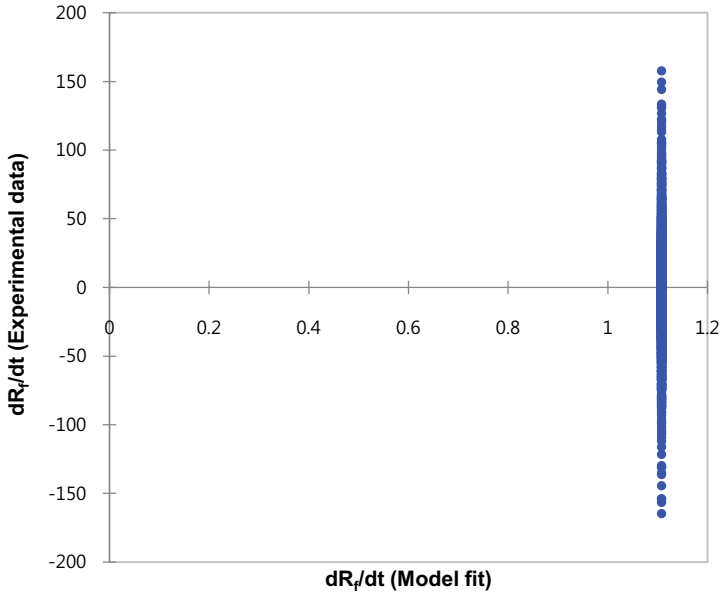

Fig. 5. Comparison of ANN model prediction with experimental data for dR_f/dt without data preprocessing.

the mean, the standard deviation is too large, resulting in a problem in the ANN model fit. The input data was also presented using a box plot. It seems that there are a few outliers that have large deviations from the center of the data. Since they are clearly noises, it was recommended that they should be properly eliminated prior to the ANN model fit.

3.3. Effect of data reprocessing: "Trimming"

To reduce the variations of the data, the data falling outside the proper range was eliminated to prepare the input data for the ANN model fit. First, the outliers were taken out from the raw data. Before the data processing, the number of raw data was 5,776. After the removal of the outliers by selecting the data between 25% and 75% percentiles from the original ones, the number of data was reduced to 2,746. Nevertheless, the ANN model fit failed again as illustrated in Fig. 7a. The box plot in Fig. 7b presents that the variations were reduced by the data trimming. In fact, the standard deviation of the data was reduced from 29.7×10^7 1/m s to 8.453×10^7 1/m s. But it does not seem to be sufficient to have a reasonable fit by the ANN model.

Accordingly, a more aggressive data trimming was tried to further decrease the variations of the data. By selecting the data between 35% and 65% percentiles, the number of data points was reduced to 291. In this case, the ANN model showed a better fit to the processed data (Fig. 8a). The range of the data was narrowed as shown in Fig. 8b and the standard deviation of the data was

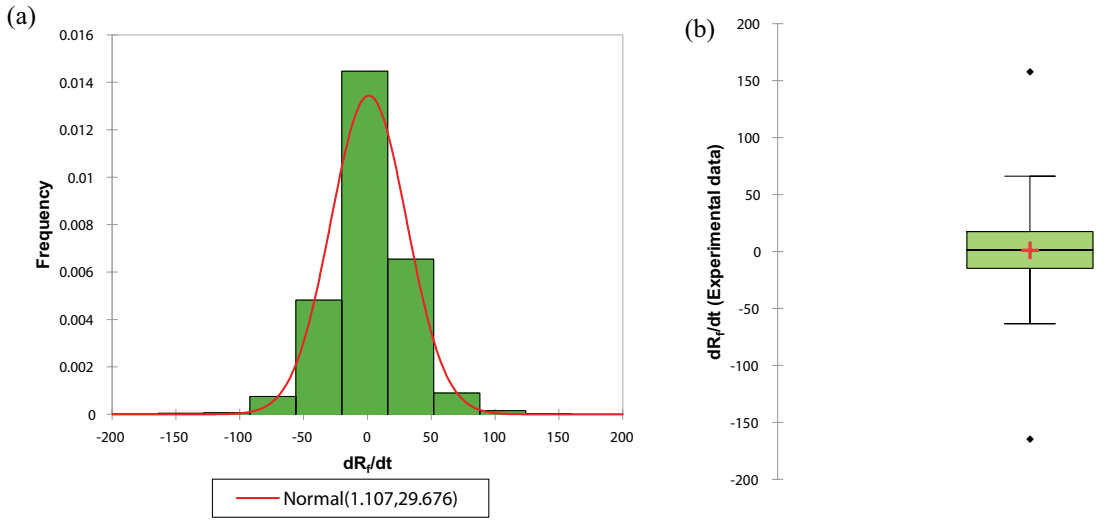

Fig. 6. (a) Probability distribution of experimental data for dR_f/dt and (b) box plot.

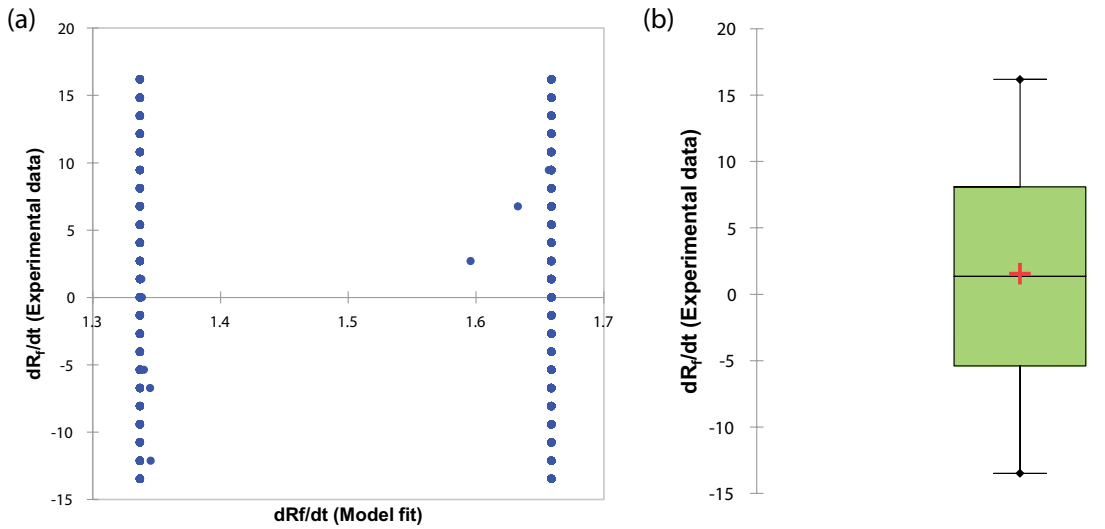

Fig. 7. Effect of data preprocessing on the accuracy of model fit and the distribution of the processed data. Data processing method: removal of outliners (a) comparison of ANN model prediction with experimental data for dR_f/dt after data preprocessing and (b) box plot for the processed data.

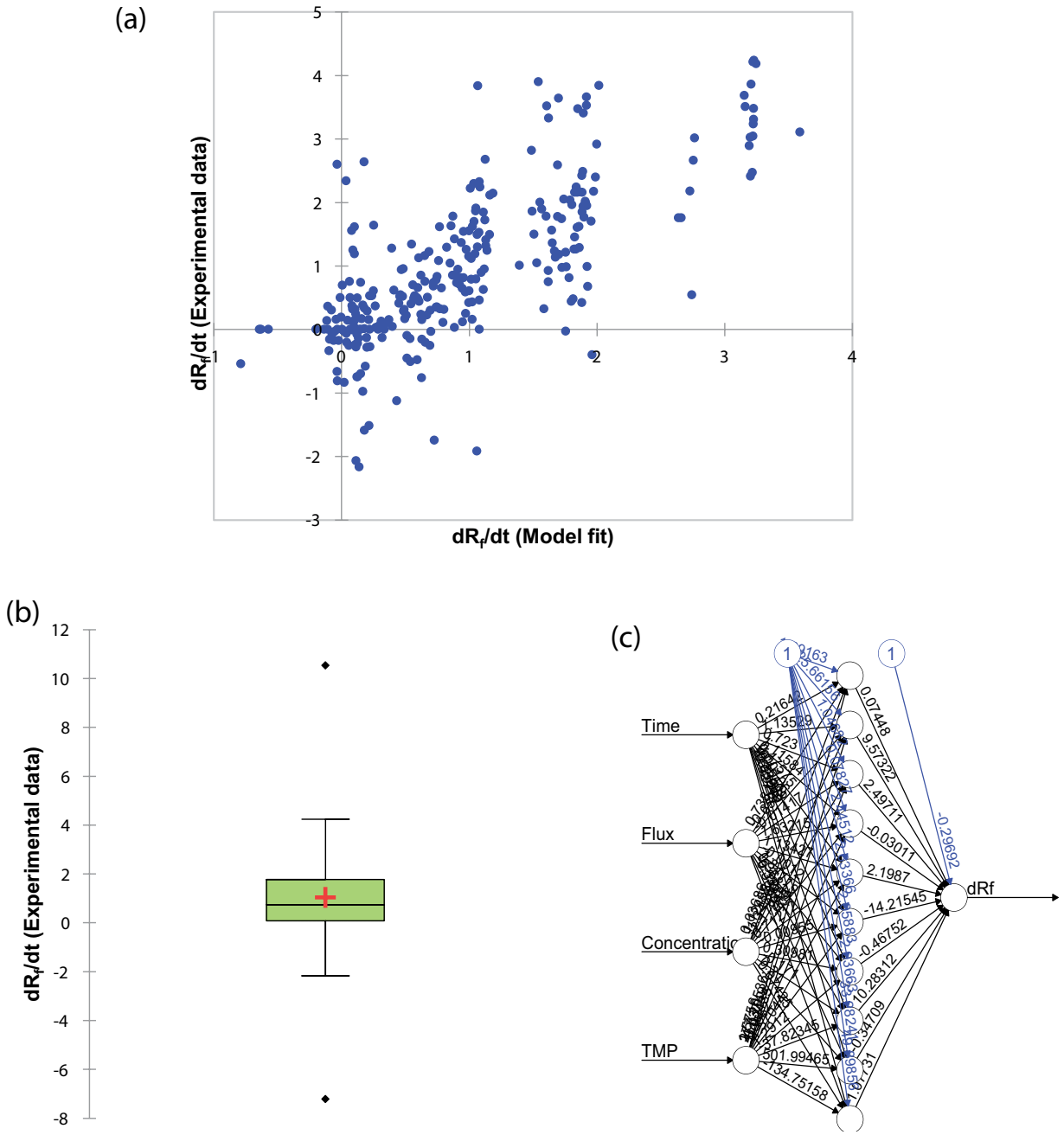

Fig. 8. Effect of data preprocessing on the accuracy of model fit and the distribution of the processed data. Data processing method: use of the data between 25% and 25% (a) comparison of ANN model prediction with experimental data for dR_f/dt after data preprocessing and (b) box plot for the processed data.

1.198×10^7 1/m s. As a result of this model fit, an ANN model was obtained, which structure is illustrated in Fig. 8c.

From the data in Fig. 8b, the data between 25% and 75% percentiles were selected to further reduce the standard deviation and used for the ANN model fit. In this case, however, the ANN model fit failed instead as shown in Fig. 9a. It is interesting to note that the accuracy of the ANN model fit was not improved even with a decreased standard deviation. The distribution of the data in Fig. 9b became narrower than that in Fig. 8b. The number of the data and the standard deviation were 216 and 0.632×10^7 1/m s, respectively. However, the characteristics of the data also seem to be altered by this data preprocessing. The mean of the original data was 1.107×10^7 1/m s but the mean of the processed data became 0.517×10^7 1/m s. This indicates that the center of the data was significantly changed. Fig. 9c shows the structure of the ANN model, which is not meaningful due to low accuracy and deterioration of input data by the processing.

3.4. Effect of data reprocessing: adjustment of data collection interval

Another approach for the reduction of the variations in the data was attempted by adjusting the interval for the data collection. The raw data was collected with 1 min interval. It was adjusted to 20 min and the ANN model fit was carried out. As shown in Fig. 10a, the model matched the experimental data well, suggesting that the adjustment of the data collection interval is more effective than the trimming of extreme data. A narrow distribution of the data was obtained as shown in Fig. 10b. The mean was 1.042×10^7 1/m s and the standard deviation was 1.693×10^7 1/m s. The structure of the ANN model is presented in Fig. 10c.

As a next step, the interval for data collection was set to 60 min. A more reasonable model fit to the experimental data was observed in Fig. 11a. Again, the adjustment of the interval helps to improve the accuracy of the model fit. The box plot in Fig. 11b demonstrates the distribution of the data after the processing. In this case, the mean was 0.911×10^7 1/m s and the standard deviation was 1.10×10^7 1/m s. The result of the model fit is also presented as a form of the ANN model structure in Fig. 11c.

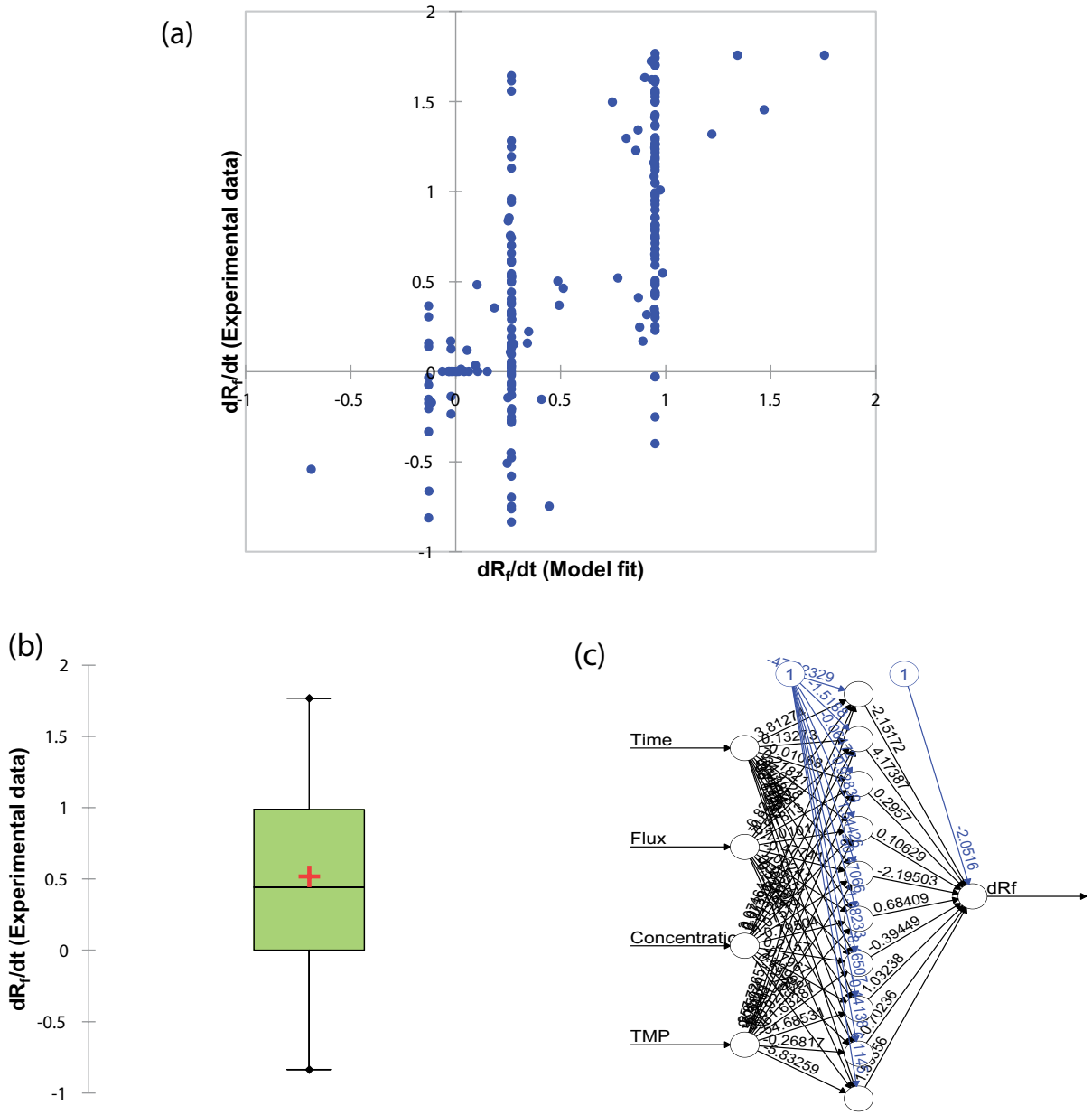

Fig. 9. Effect of data preprocessing on the accuracy of model fit and the distribution of the processed data. Data processing method: use of the data between 25% and 25% and removal of outliners (a) comparison of ANN model prediction with experimental data for dR_f/dt after data preprocessing and (b) box plot for the processed data.

Based on these results, it can be concluded that the selection of the data collection interval is an important factor affecting the accuracy of the ANN model fit. Although the number of data decreases with an increase in the interval, it poses a positive effect on the ANN model fit. Of course, a further increase in the interval may result in a negative effect on the ANN model due to an insufficient number of data point. Thus, the optimum interval for data collection should exist.

4. Conclusions

This paper investigated the effect of data preprocessing on the ANN analysis of the fouling rate in MF membranes. Based on the results, the following conclusions were withdrawn:

- The dR_f/dt calculated from TMP exhibited large variation in 1 min interval of the data collection. The resulting distribution of the data for dR_f/dt was very broad.

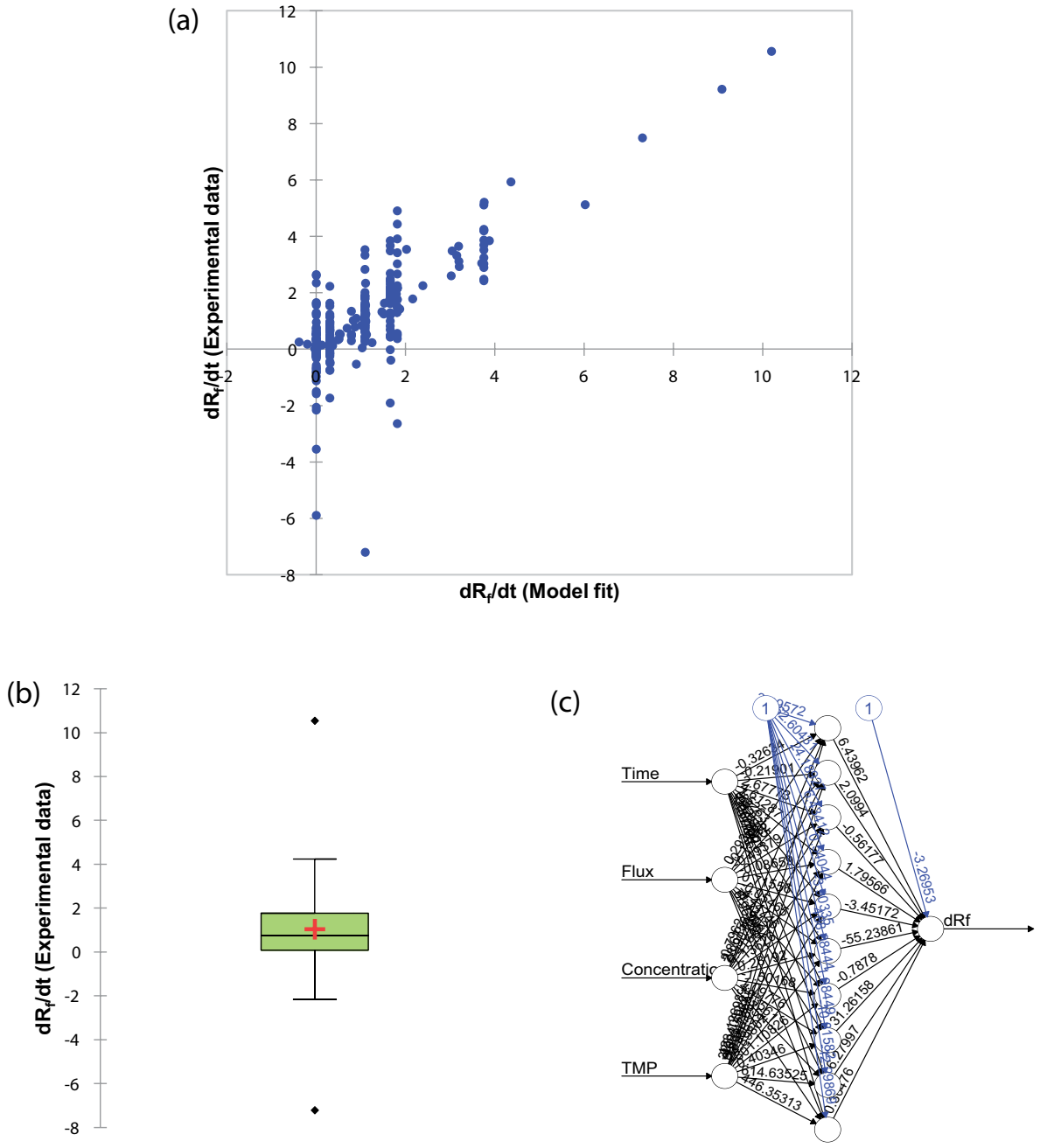

Fig. 10. Effect of data preprocessing on the accuracy of model fit and the distribution of the processed data. Data processing method: adjustment of the sampling interval to 20 min (original sampling interval: 1 min) (a) comparison of ANN model prediction with experimental data for dR_f/dt after data preprocessing and (b) box plot for the processed data.

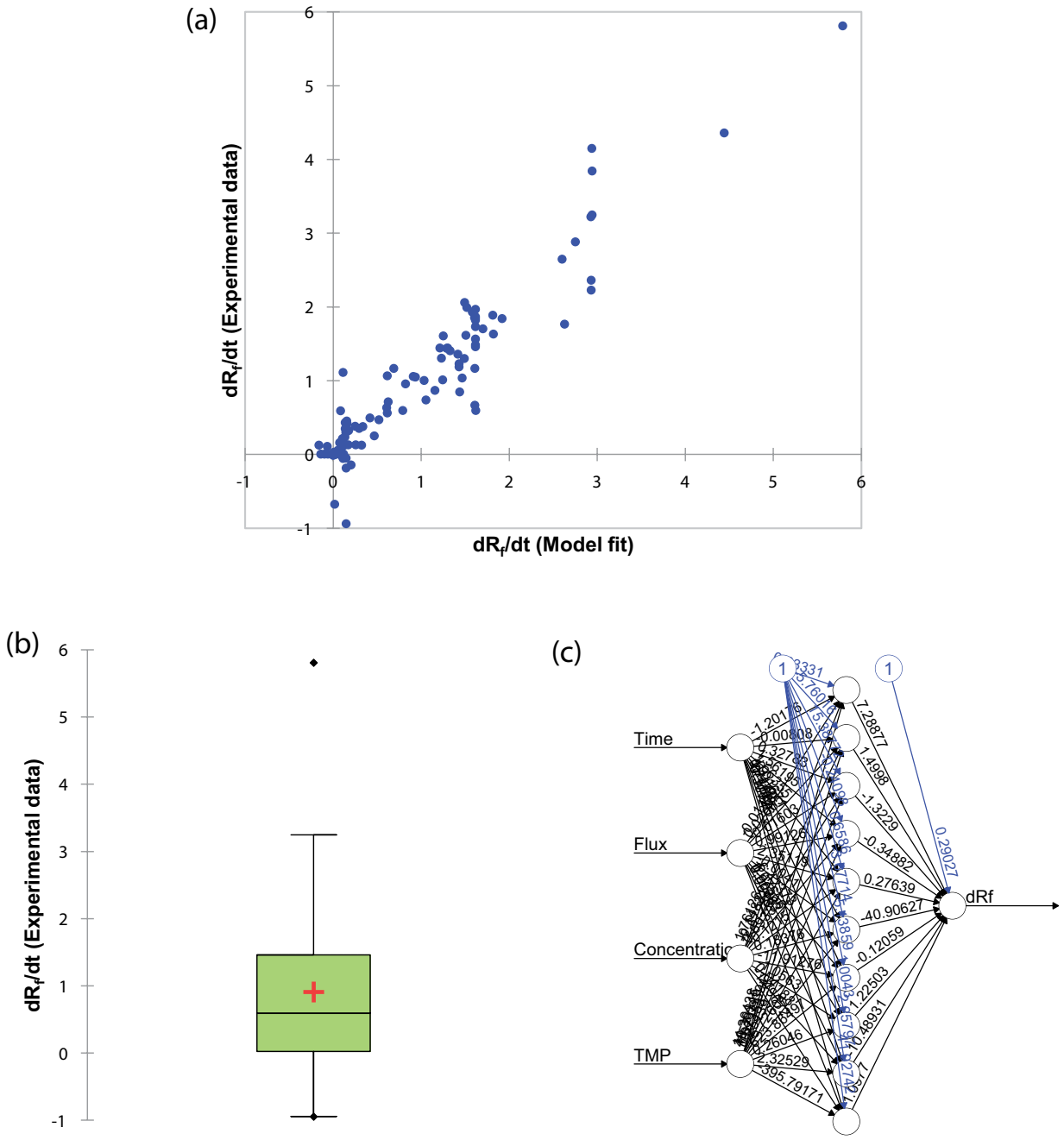

Fig. 11. Effect of data preprocessing on the accuracy of model fit and the distribution of the processed data. Data processing method: adjustment of the sampling interval to 60 min (original sampling interval: 1 min). (a) comparison of ANN model prediction with experimental data for dR_f/dt after data preprocessing and (b) box plot for the processed data.

- An ANN model with the four independent variables (foulant concentration, time, flux, and TMP) was trained using the dR_f/dt data. However, the model failed to fit the experimental results due to the large variations in the input data. It does not seem to be possible to use this data directly for the ANN model fit.
- The trimming of the data that falls outside the desired ranges was attempted and the results were analyzed. When a proper range was selected, the elimination of the outliers led to a better fit to the experimental results by the ANN model. Nevertheless, the characteristics of the data were affected by this preprocessing, which led to a failure of the ANN model fit.
- As an alternative technique, the interval for the data collection was adjusted to decrease the standard deviation of the data. Results showed that the ANN model matched the experimental data well after this processing. It is evident from the results that the adjustment of data

collection interval is more effective than the trimming of extreme data.

Acknowledgment

This subject is supported by Korea Ministry of Environment as "Global Top Project (2017002100001)."

References

[1] T. Turken, R. Sengur-Tasdemir, E. Ates-Genceli, V.V. Tarabara, I. Koyuncu, Progress on reinforced braided hollow fiber membranes in separation technologies: a review, J. Water Process Eng., 32 (2019) 100938.
[2] M. Bodzek, K. Konieczny, A. Kwiecińska, Application of membrane processes in drinking water treatment–state of art, Desal. Water Treat., 35 (2011) 164–184.
[3] A.G. Fane, S. Chang, E. Chardon, Submerged hollow fibre membrane module—design options and operational considerations, Desalination, 146 (2002) 231–236.
[4] S. Lee, P.-K. Park, J.-H. Kim, K.-M. Yeon, C.-H. Lee, Analysis of filtration characteristics in submerged microfiltration for drinking water treatment, Water Res., 42 (2008) 3109–3121.
[5] J.G. Jacangelo, R. Rhodes Trussell, M. Watson, Role of membrane technology in drinking water treatment in the United States, Desalination, 113 (1997) 119–127.
[6] R. Bogati, C. Goodwin, K. Marshall, K.T. Leung, B.Q. Liao, Optimization of chemical cleaning for improvement of membrane performance and fouling control in drinking water treatment, Sep. Sci. Technol., 50 (2015) 1835–1845.
[7] S.F. Anis, R. Hashaikeh, N. Hilal, Microfiltration membrane processes: a review of research trends over the past decade, J. Water Process Eng., 32 (2019) 100941.
[8] K. Glucina, Q. Derekx, C. Langlais, J.M. Laine, Use of advanced CFD tool to characterize hydrodynamic of commercial UF membrane module, Desal. Water Treat., 9 (2009) 253–258.
[9] Y.-J. Kim, J.-W. Jung, S. Lee, Comparison of fouling rates for pressurized and submerged ultrafiltration membranes, Desal. Water Treat., 54 (2015) 3610–3615.
[10] V. Mavrov, H. Chmiel, J. Kluth, J. Meier, F. Heinrich, P. Ames, K. Backes, P. Usner, Comparative study of different MF and UF membranes for drinking water production, Desalination, 117 (1998) 189–196.
[11] X. Li, J. Li, Z. Cui, Y. Yao, Modeling of filtration characteristics during submerged hollow fiber membrane microfiltration of yeast suspension under aeration condition, J. Membr. Sci., 510 (2016) 455–465.
[12] J. Jung, J. Ryu, S.Y. Choi, K.Y. Park, W.J. Song, Y. Yu, Y.-s. Jang, J. Park, J. Kweon, Autopsy study of irreversible foulants on polyvinylidene fluoride hollow-fiber membranes in an immersed microfiltration system operated for five years, Sep. Purif. Technol., 199 (2018) 1–8.
[13] S. Chang, A.G. Fane, The effect of fibre diameter on filtration and flux distribution—relevance to submerged hollow fibre modules, J. Membr. Sci., 184 (2001) 221–231.
[14] M. Bagheri, A. Akbari, S.A. Mirbagheri, Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: a critical review, Process Saf. Environ. Prot., 123 (2019) 229–252.
[15] Y.-J. Kim, T. Yun, S. Lee, D. Kim, J. Kim, Accelerated testing for fouling of microfiltration membranes using model foulants, Desalination, 343 (2014) 113–119.
[16] M. Liu, C. Xiao, X. Hu, Fouling characteristics of polyurethane-based hollow fiber membrane in microfiltration process, Desalination, 298 (2012) 59–66.
[17] Y. Yang, S. Qiao, M. Zheng, J. Zhou, X. Quan, Enhanced permeability, contaminants removal and antifouling ability of CNTs-based hollow fiber membranes under electrochemical assistance, J. Membr. Sci., 582 (2019) 335–341.
[18] B.K. Thakur, S. De, A novel method for spinning hollow fiber membrane and its application for treatment of turbid water, Sep. Purif. Technol., 93 (2012) 67–74.
[19] B.K. Pramanik, S.K. Pramanik, D.C. Sarker, F. Suja, Impact of ozonation, anion exchange resin and UV/H_2O_2 pre-treatments to control fouling of ultrafiltration membrane for drinking water treatment, Environ. Technol., 38 (2017) 1383–1389.
[20] Q. Han, T.A. Trinh, J.W. Chew, Cake formation of bidisperse suspensions in dead-end microfiltration, J. Membr. Sci., 577 (2019) 31–40.
[21] B.-B. Choi, Y.-J. Choi, J.-S. Choi, S. Lee, H.-J. Oh, Energy management in submerged microfiltration systems by optimum control of aeration, Desalination, 247 (2009) 233–238.
[22] D.-C. Choi, S.-Y. Jung, Y.-J. Won, J.H. Jang, J. Lee, H.-R. Chae, K.H. Ahn, S. Lee, P.-K. Park, C.-H. Lee, Three-dimensional hydraulic modeling of particle deposition on the patterned isopore membrane in crossflow microfiltration, J. Membr. Sci., 492 (2015) 156–163.
[23] Y. Ko, Y. Choi, H. Cho, Y. Shin, S. Lee, Comparison of fouling behaviors of hydrophobic microporous membranes in pressure- and temperature-driven separation processes, Desalination, 428 (2018) 264–271.
[24] K. Gao, X. Xi, Z. Wang, Y. Ma, S. Chen, X. Ye, Y. Li, Use of support vector machine model to predict membrane permeate flux, Desal. Water Treat., 57 (2016) 16810–16821.
[25] Q.-F. Liu, S.-H. Kim, S. Lee, Prediction of microfiltration membrane fouling using artificial neural network models, Sep. Purif. Technol., 70 (2009) 96–102.
[26] S. Ghandehari, M.M. Montazer-Rahmati, M. Asghari, A comparison between semi-theoretical and empirical modeling of cross-flow microfiltration using ANN, Desalination, 277 (2011) 348–355.
[27] S. Chang, A.G. Fane, S. Vigneswaran, Modeling and optimizing submerged hollow fiber membrane modules, AIChE J., 48 (2002) 2203–2212.
[28] C. Duclos-Orsello, W. Li, C.-C. Ho, A three mechanism model to describe fouling of microfiltration membranes, J. Membr. Sci., 280 (2006) 856–866.
[29] K.-J. Hwang, C.-Y. Liao, K.-L. Tung, Analysis of particle fouling during microfiltration by use of blocking models, J. Membr. Sci., 287 (2007) 287–293.
[30] T.-M. Hwang, Y. Choi, S.-H. Nam, S. Lee, H. Oh, K. Hyun, Y.-K. Choung, Prediction of membrane fouling rate by neural network modeling, Desal. Water Treat., 15 (2010) 134–140.
[31] S.A. Aya, T. Ormanci Acar, N. Tufekci, Modeling of membrane fouling in a submerged membrane reactor using support vector regression, Desal. Water Treat., 57 (2016) 24132–24145.
[32] E. Piron, E. Latrille, F. René, Application of artificial neural networks for crossflow microfiltration modelling: "black-box" and semi-physical approaches, Comput. Chem. Eng., 21 (1997) 1021–1030.