

<https://doi.org/10.15255/KUI.2019.024>

KUI-7/2020

Original scientific paper

Received June 22, 2019

Accepted October 19, 2019

# A Comparison of "Neural Networks and Multiple Linear Regressions" Models to Describe the Rejection of Micropollutants by Membranes

Y. Ammi,<sup>a,b\*</sup> L. Khaouane,<sup>b</sup> and S. Hanini<sup>b</sup><sup>a</sup> University Center Ahmed Zabana Relizane, 48 000, Algeria<sup>b</sup> Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Médéa, 26 000, Algeria

This work is licensed under a Creative Commons Attribution 4.0 International License



## Abstract

A rejection process of organic compounds by nanofiltration and reverse osmosis membranes was modelled using the artificial neural networks. Three feed-forward neural networks based on quantitative structure-activity relationship (QSAR-NN models) characterised by a similar structure (twelve neurons for QSAR-NN<sub>1</sub>, QSAR-NN<sub>2</sub>, and QSAR-NN<sub>3</sub> in the input layer, one hidden layer and one neuron in the output layer), were constructed with the aim of predicting the rejection of organic compounds. A set of 1394 data points for QSAR-NN<sub>1</sub>, 980 data points for QSAR-NN<sub>2</sub>, and 436 data points for QSAR-NN<sub>3</sub> were used to construct the neural networks. Good agreements between the predicted and experimental rejections were obtained by QSAR-NN models (the correlation coefficient for the total dataset were 0.9191 for QSAR-NN<sub>1</sub>, 0.9338 for QSAR-NN<sub>2</sub>, and 0.9709 for QSAR-NN<sub>3</sub>). Comparison between the feed-forward neural networks and multiple linear regressions based on quantitative structure-activity relationship "QSAR-MLR" revealed the superiority of the QSAR-NN models (the root mean squared errors for the total dataset for the QSAR-NN models were 10.6517 % for QSAR-NN<sub>1</sub>, 9.1991 % for QSAR-NN<sub>2</sub>, and 5.8869 % for QSAR-NN<sub>3</sub>, and for QSAR-MLR models they were 20.1865 % for QSAR-MLR<sub>1</sub>, 19.3815 % for QSAR-MLR<sub>2</sub>, and 16.2062 % for QSAR-MLR<sub>3</sub>).

## Keywords

Neural networks, multiple linear regressions, models, rejection, micropollutants, membranes

## 1 Introduction

Nanofiltration (NF) and reverse osmosis (RO) are key processes widely used for the removal of ions and organics from contaminated feedwater in process water, drinking water, and wastewater treatment.<sup>1,2</sup> The occurrence of trace organic contaminants in treated and untreated wastewater appears to be a significant environmental health concern.<sup>3,4</sup> Hence, the long-term effects of consumption of water containing low concentrations of micropollutants will remain an unanswered question for the foreseeable future.<sup>5</sup> Meanwhile, water treatment facilities are implementing monitoring programs, research organizations dealing with water reuse have published reports, and studies have addressed the topic.<sup>6</sup>

Fewer studies have examined micropollutant rejection as a function of the interactions between compounds and membrane surfaces.<sup>6–8</sup> In general, three major interactions primarily affecting solute-membrane rejection can be pointed out as follows: steric hindrance (sieving effect), electrostatic repulsion (charge effect), and hydrophobic/adsorptive interactions.<sup>9–11</sup> An important aspect to deal with the problem has been the identification of compound physicochemical properties and membrane characteristics to explain transport and removal of micropollutants by different mechanisms, explicitly size/steric exclusion, hy-

drophobic adsorption and partitioning, and electrostatic repulsion.<sup>1,12,13–18</sup>

Nanofiltration can be characterised by the molecular weight cut-off (MWCO) or ionic retention of salts, such as the magnesium sulphate salt rejection (SR(MgSO<sub>4</sub>)) or the sodium chloride salt rejection (SR(NaCl)); reverse osmosis membranes being dense are characterised also by salt retention.<sup>19</sup> The use of MWCO was acceptable for modelling purposes; however, NF membranes with a broad range of MWCO (pore size and distribution) may make it difficult to estimate the rejection of contaminants, thus magnesium sulphate salt rejection SR(MgSO<sub>4</sub>) may be more appropriate.<sup>5</sup> Moreover, MWCO alone is frequently unable to predict the rejection of charged and uncharged organic compounds by NF/RO membranes.<sup>5,19,20</sup> Hence, in this work, sodium chloride salt rejection "SR (NaCl)" is considered as a factor of input because it is the first parameter giving an indication of the "porosity" of the membrane, measured by the manufacturer under standard conditions.<sup>21</sup> Agenson *et al.* showed, in general, that the strongest retention rates are obtained for membranes having the highest retention rate in NaCl.<sup>22</sup>

A number of articles have proposed a mechanistic understanding of the interaction between membranes and organic compounds; others have tried to apply fitting parameter models to model rejection.<sup>8,23,24</sup> However, there have been few models to "predict" the rejection of compounds.<sup>5</sup> In recent years, there have been various attempts

\* Corresponding author: Dr Yamina Ammi  
e-mail: [ammi.yamina@yahoo.fr](mailto:ammi.yamina@yahoo.fr)

to advance the use of QSAR as a viable approach to develop data-driven models to describe the performance of membrane processes.<sup>5,25–28</sup> In 2009, *Yangali Quintanilla et al.* evaluated a MLR based on QSAR to model the 161 rejections of 50 neutral organic compounds by six NF and nine RO membranes. They demonstrated that the magnesium sulphate salt rejection ( $SR(MgSO_4)$ ) may be a possible lump parameter that defines size exclusion capability of neutral organic compounds by NF/RO membrane. This model (Eq. 1) correlates the retention with the properties of neutral organic molecules (size and hydrophobicity) and properties of the NF/RO membrane (porosity):<sup>28</sup>

$$\begin{aligned} \text{Rejection} = & 183.920\text{Eqwidth} + 31.830\text{Length} - \\ & - 0.549\text{Log}k_{ow} + 883.294\text{SR}(MgSO_4) - \\ & - 945.13 \end{aligned} \quad (1)$$

In 2010, *Yangali Quintanilla et al.* had modelled the rejection of emerging contaminants (pharmaceuticals, endocrine disruptors) by NF membranes. They used principally partial least squares, and multiple linear regression analysis to find QSAR models that combine interactions between membrane characteristics, filtration operating conditions, and compound properties to predict the rejection. They based their study on a database of 106 rejections of 66 neutral and ionic organic molecules by NF membranes. They concluded that MWCO was acceptable for modelling purposes, but  $SR(MgSO_4)$  may be more appropriate. They developed two MLR-QSAR models (Eqs. 2 and 3). These models correlate the rejection with the properties of neutral and ionic organic molecules (size and hydrophobicity) and properties of the NF membrane (porosity):<sup>5</sup>

$$\begin{aligned} \text{Rejection} = & 265.150\text{Eqwidth} - 117.356\text{depth} + \\ & + 81.662\text{length} - 5.229\text{logD} + \\ & + 1358.090\text{SR}(MgSO_4) - 144.817 \end{aligned} \quad (2)$$

$$\begin{aligned} \text{Rejection} = & 265.150\text{Eqwidth} - 117.356\text{depth} + \\ & + 81.662\text{length} - 5.229\text{logD} - \\ & - 0.272\text{MWCO} - 62.565 \end{aligned} \quad (3)$$

In this study, we developed feed-forward neural networks (FFNN) models and multiple linear regressions "QSAR-MLR", to predict the rejection of organic compounds by NF/RO membranes. To the best of our knowledge, no studies have been reported in the literature that have used feed-forward neural networks for modelling the rejection of organic compounds by NF/RO membranes, taking into account the molecular weight cut-off and salt rejection "magnesium sulphate and sodium chloride". The feed-forward neural networks "FFNN" also compared with multiple linear regressions "QSAR-MLR".

The aim of our study was the superiority of the QSAR-NN or QSAR-MLR models, and therefrom to confirm that the rejection of organic compounds by NF/RO membrane is a linear or nonlinear phenomenon, and determine the best porosity indicator (MWCO,  $SR(MgSO_4)$ , and  $SR(NaCl)$ ) for retention of neutral and ionic organic compounds by NF/RO membrane. The paper is organised as follows. *Section 2* presents a database collection; modelling the rejection of charged and uncharged organic compounds by NF/RO membranes using QSAR-NN and QSAR-MLR. *Section 3*

presents results and discussions. *Section 4* presents comparisons between QSAR-NN, QSAR-MLR, and other models. *Section 5* presents selected organic compounds rejection at varying feedwater pH values.

## 2 Materials and Methods

### 2.1 Data Base Collection

In this work, we have used data available in the literature.<sup>4–8,14,17–19,29–46</sup> The first database (DB<sub>1</sub>) consists of 1394 rejections of 116 organic compounds, the second database (DB<sub>2</sub>) consists of 980 rejections of 102 organic compounds, and the third database (DB<sub>3</sub>) consists of 436 rejections of 42 organic compounds. The list of the 116 organic compounds is shown in Supplementary Data A.

The selection of the input and output variables was based on interactions between organic compound properties, membrane characteristics, and filtration operating conditions for the removal of organic compounds by NF/RO membranes. These solute–membrane interactions are determined by organic compound solute properties, membrane properties, and operating conditions.<sup>11,28,46</sup> The inputs considered in this work were: compound descriptors (molecular weight, compound hydrophobicity "logD", dipole moment, molecular length, and equivalent molecular width), membrane characteristics (molecular weight cut-off (MWCO), magnesium sulphate salt rejection "SR ( $MgSO_4$ )", sodium chloride salt rejection "SR ( $NaCl$ )", surface membrane charge "zeta potential", and membrane hydrophobicity "contact angle"), and operating conditions (pH, pressure, recovery, and temperature).

Molecular weight (MW) and logD were obtained from Chemspider,<sup>47</sup> the Dipole moment, another fundamental descriptor, was calculated by Chem3D Ultra 12.0, and ChemAxon<sup>48</sup> used for computing molecular descriptors of size, such as molecular length, molecular width, and molecular depth. The equivalent molecular width is defined as the geometric mean of width and depth.<sup>5,49</sup>

The values of the minimum (min), mean (mean), maximum (max), and standard deviations (STD) for the inputs and output data are presented in Table 1.

The samples were randomly split into three subsets: 80 % for training set, 10 % for validation set, and 10 % for test set. Fig. 1 shows the total data as function of the molecular weight for each database. The rejections according to the molecular weight are indicated by pink circles for DB1 "1394 points", blue circles for DB2 "980 points", and green circles for BD3 "436 points". When the molecular weight varied from 150 to 300  $\text{mg mol}^{-1}$ , the experimental rejections, which had percentages greater than 60.5 %, agglomerated with each other.

### 2.2 Modelling with neural networks

A neural network (NN) is a massively parallel-distributed processor made up of simple processing units known as

Table 1 – Statistical analysis of inputs and output

	DB <sub>1</sub>				DB <sub>2</sub>				DB <sub>3</sub>			
	Min	Mean	Max	STD	Min	Mean	Max	STD	Min	Mean	Max	STD
$M_w/g\ mol^{-1}$	30.0200	244.5202	446.4800	68.7297	30.0200	247.4053	446.4800	6837297	60.0520	269.2295	446.4800	88.5031
logD	-7.0800	1.1272	6.1300	1.9384	-7.0800	0.9968	6.1300	1.9384	-5.3900	0.9855	5.1300	2.1062
Dipole moment /Debye	0.0000	4.0478	15.0300	2.9158	0.0000	3.9423	15.0300	2.9158	0.0000	4.2638	10.9100	3.2755
Length/nm	0.4410	1.2749	2.1120	0.2731	0.4410	1.2781	2.1120	0.2731	0.5660	1.2877	1.9970	0.2806
Eqwidth/nm	0.1676	0.3598	0.4753	0.04473	0.1676	0.3593	0.4753	0.0447	0.2264	0.3671	0.4753	0.0574
MWCO/Dalton	100.0000	217.5574	400.0000	79.0327	-	-	-	-	-	-	-	-
SR/MgSO <sub>4</sub>	-	-	-	-	0.1000	0.9335	0.9970	79.0327	-	-	-	-
SR/NaCl	-	-	-	-	-	-	-	-	0.1210	0.7792	0.9960	0.2728
Zeta potential/mv	-87.0000	-28.5868	3.1000	24.4887	-48.0400	-20.4601	3.1000	24.4887	-41.3000	-22.0564	3.1000	12.2077
Contact angle/°	20.1000	47.3673	94.7000	13.4829	20.1000	46.1141	70.7000	13.4829	20.1000	44.5829	70.7200	14.2688
pH	2.5000	7.1522	10.0100	1.1795	2.5000	7.0831	10.0100	1.1795	3.0000	6.6557	9.0000	1.2737
Pressure/kPa	240.0000	1136.9480	4100	794.0021	240.0000	1316.5873	4100.0000	794.0021	240.0000	969.6330	4100.0000	581.3423
Recovery/%	0.0300	33.5154	98.4800	31.3667	0.0400	25.7517	92.2600	31.3667	0.0300	26.7278	98.4800	27.0375
Temperature/°C	13.0000	21.4795	25.0000	2.8883	13.0000	21.2949	25.0000	2.8888	20.0000	23.2626	25.0000	2.1322
Rejection/%	0.0000	77.7059	100.0000	27.0116	0.0000	79.6160	100.0000	27.01165	0.0000	83.3365	100.0000	24.6214

nodes that perform certain mathematical functions, usually nonlinear. This kind of non-algorithmic computation is characterised by a system resembling the human brain structure. One of the great advantages of these models is their ability to learn (store experimental knowledge), generalise (make the knowledge available) or automatically extract rules from complex data.<sup>50</sup> The role of the QSAR-NN is to transform the inputs information into output information. During the training process, the weights are corrected to produce output values as close as possible to the target values.<sup>51</sup>

The procedure based on the design and optimisation of the architecture of neural network "QSAR-NN" is described in Fig. 2.

### 2.3 Modelling with multiple linear regressions

Multiple linear regressions "QSAR-MLR" are the most widely used and known modelling methods, and are used as the basis for a number of multivariate methods.<sup>52</sup> Multiple linear regressions consist of a quantitative relationship

between a group of predictor variables ( $X$ ) and a response  $y_i$ , as shown in Eq. (4):

$$y_i = A_0 + \sum_{k=1}^{12} A_k X_k \quad (4)$$

where  $y_i$  is the output of QSAR-MLR,  $X_k$  represents the inputs,  $A_k$  represents the coefficients, and  $A_0$  is the intercept of the equation.

## 3 Results and discussion

### 3.1 Neural networks

The database was divided into three different sets. Training phase (80 %) is the major part of the data, because from this data set, the connection weights of the neurons are adjusted during the training to acquire the knowledge of the network. On the other hand, the validation phase (10 %) is used for verifying the generalisation ability of the network. The test phase (10 %) is used to evaluate the NN

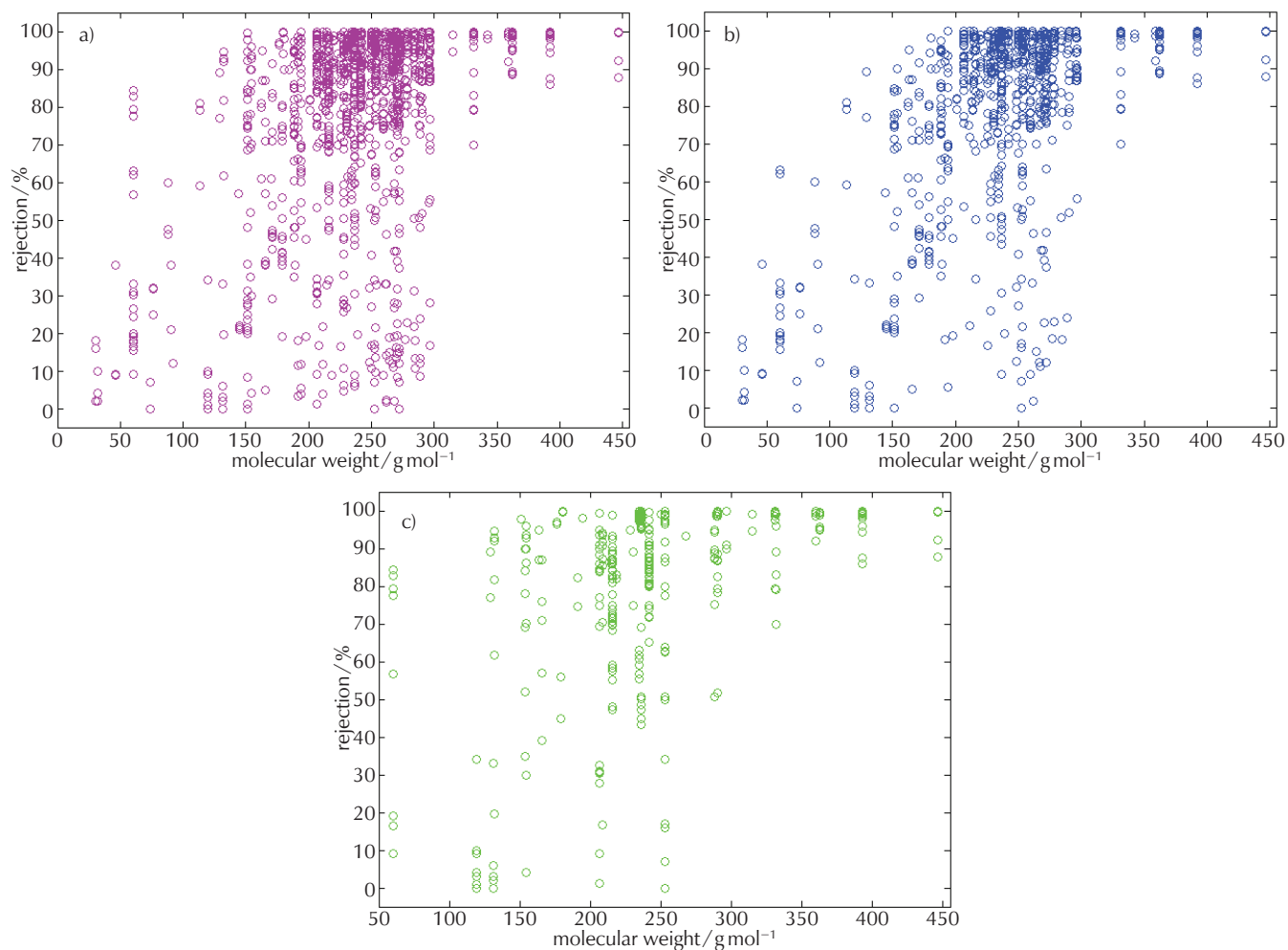


Fig. 1 – Three databases as a function of the molecular weight: (a) DB<sub>1</sub>, (b) DB<sub>2</sub>, (c) DB<sub>3</sub>

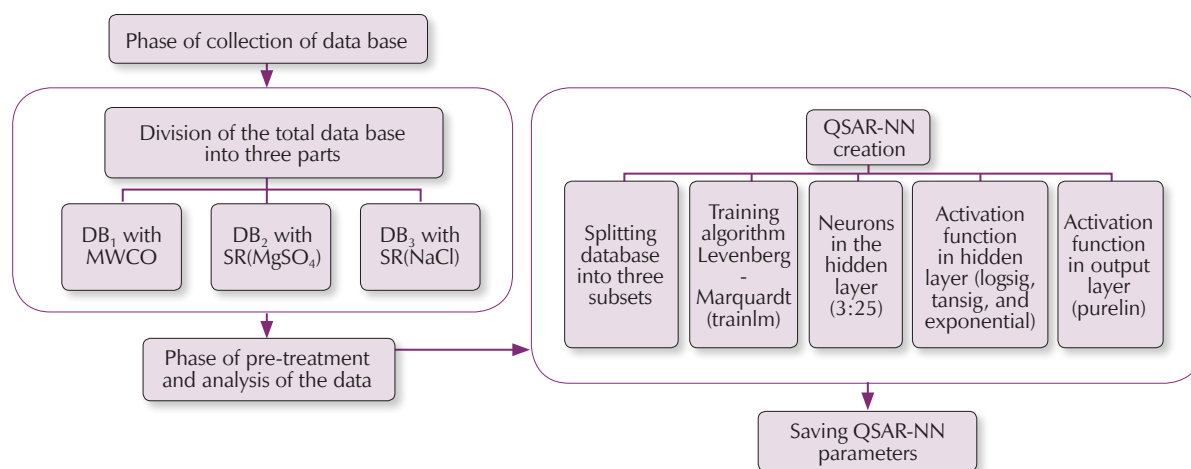


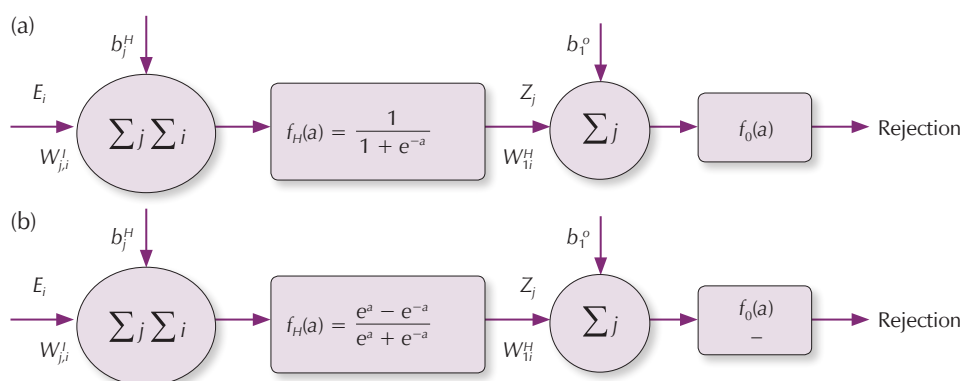
Fig. 2 – Procedure of the design and optimisation of the architecture of neural network

performance in real situations. The training algorithm used in this work was the BFGS quasi-Newton (*trainbfg*). Each QSAR-NN had three layers of neurons or nodes: one input layer with twelve neurons for QSAR-NN<sub>1</sub>, QSAR-NN<sub>2</sub>, and QSAR-NN<sub>3</sub> in the input layer, one hidden layer with a number of active neurons optimised during training, and

one output layer with one unit that generated the estimated value of rejection. The number of hidden neurons varied from 3 to 25 neurons. The tangent sigmoid (*tansig*), the log sigmoid (*logsig*), and the exponential transfer functions were used in the hidden layer. The pure-linear (*purelin*) transfer function was used in the output layer.

Table 2 – Structures of the optimized QSAR-NN models

QSAR-NN models	Training algorithm	Input layer	Hidden layer		Output layer	
		number of neurons	number of neurons	activation function	number of neurons	activation function
QSAR-NN <sub>1</sub>	BFGS quasi-Newton ( <i>trainbfg</i> )	12	25	logsig	1	purelin
QSAR-NN <sub>2</sub>			23	tansig		
QSAR-NN <sub>3</sub>			19	tansig		

Fig. 3 – Schematic representation of the optimised QSAR-NN: (a) QSAR-NN<sub>1</sub>, (b) QSAR-NN<sub>2</sub>, and QSAR-NN<sub>3</sub>

The application of QSAR-NN modelling of the rejection of charged and uncharged organic compounds by NF/RO membranes was performed using *STATISTICA 8* software. Table 2 shows the structure of the optimised QSAR-NN models.

The weight matrices and bias vectors of the optimised QSAR-NN models are listing in Supplementary Data B, where  $w_i$  is the input-hidden layer connection weight matrix (25 rows  $\times$  12 columns for NN<sub>1</sub>, 23 rows  $\times$  12 columns for QSAR-NN<sub>2</sub>, and 19 rows  $\times$  12 columns for QSAR-NN<sub>3</sub>),  $w_h$  is the hidden layer output connection weight matrix (25 rows  $\times$  1 column for QSAR-NN<sub>1</sub>, 23 rows  $\times$  1 column for QSAR-NN<sub>2</sub>, and 19 rows  $\times$  1 column for QSAR-NN<sub>3</sub>),  $b_h$  is the hidden neurons bias column vector (25 rows for QSAR-NN<sub>1</sub>, 23 rows for QSAR-NN<sub>2</sub>, and 19 rows for QSAR-NN<sub>3</sub>), and  $b_o$  is the output neurons bias column vector (1 row for QSAR-NN<sub>1</sub>, QSAR-NN<sub>2</sub>, and QSAR-NN<sub>3</sub>).

From the optimised neural networks (QSAR-NN) shown in Fig. 3, assimilation of rejection of organic compounds by NF/RO membranes can be expressed by a mathematical model incorporating all inputs  $E_i$  ( $M_w$ , logD, dipole moment, length, Eqwidth, MWCO, SR(MgSO<sub>4</sub>), SR(NaCl), zeta potential, contact angle, pH, pressure, recovery, and temperature) within it, as follows:

The instance outputs  $Z_j$  of the hidden layer:

$$Z_j = f_H \left[ \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right] = \frac{1}{1 + \exp \left( - \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right)} \quad (5)$$

$$j = 1, 2, \dots, 25$$

The output "Rejection":

$$\text{Rejection}(\text{NN}_1) = f_0 \left[ \sum_{j=1}^{25} w_{1j}^H Z_j + b_1^O \right] = \sum_{j=1}^{25} w_{1j}^H Z_j + b_1^O \quad (6)$$

The combination of Eqs. (2) and (3) leads to the following mathematical formula, which describes the retention assimilation by taking into account all inputs  $E_i$  ( $M_w$ , logD, dipole moment, length, Eqwidth, MWCO, SR(MgSO<sub>4</sub>), SR(NaCl), zeta potential, contact angle, pH, pressure, recovery, and temperature):

$$\text{Rejection}(\text{QSAR-NN}_1) = \sum_{j=1}^{25} w_{1j}^H \frac{1}{1 + \exp \left( - \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right)} + b_1^O \quad (7)$$

Similarly, rejection "SR(MgSO<sub>4</sub>)" and rejection "SR(NaCl)" can be expressed by mathematical equations extracted from QSAR-NN<sub>2</sub> and QSAR-NN<sub>3</sub>, optimised as follows (with:

$$Z_j = f_H \left[ \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right] = \frac{\exp \left( \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right) - \exp \left( - \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right)}{\exp \left( \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right) + \exp \left( - \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right)}$$

for QSAR-NN<sub>2</sub> and QSAR-NN<sub>3</sub>):

$$\text{Rejection}(\text{QSAR-NN}_2) = \sum_{j=1}^{23} w_{1j}^H \frac{\exp \left( \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right) - \exp \left( - \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right)}{\exp \left( \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right) + \exp \left( - \sum_{i=1}^{12} w_{ji}^H E_i + b_j^H \right)} + b_1^O \quad (8)$$



$$\text{Rejection}(\text{QSAR-NN}_3) = \sum_{j=1}^{19} W_{1j}^H \frac{\exp\left(\sum_{i=1}^{12} w_{ji}' E_i + b_j^H\right) - \exp\left(-\sum_{i=1}^{12} w_{ji}' E_i + b_j^H\right)}{\exp\left(\sum_{i=1}^{12} w_{ji}' E_i + b_j^H\right) + \exp\left(-\sum_{i=1}^{12} w_{ji}' E_i + b_j^H\right)} + b_1^o \quad (9)$$

Fig. 4 shows a comparison between experimental and calculated rejections (total data set) for the present models with an agreement vector approaching the ideal in the adjustment of the profiles of the neural networks,  $[\alpha, \beta, R] = [0.8608, 10.8711, 0.9191]$  for QSAR-NN<sub>1</sub>;  $[\alpha, \beta, R] = [0.8896, 8.7206, 0.9338]$  for QSAR-NN<sub>2</sub>, and  $[\alpha, \beta, R] = [0.9481, 4.3043, 0.9709]$  for QSAR-NN<sub>3</sub>. Correlation coefficients are generally considered to be excellent ( $0.90 \leq R \leq 1.00$ ) for these neural networks (QSAR-NN<sub>1</sub>,

QSAR-NN<sub>2</sub>, and QSAR-NN<sub>3</sub>), which shows good robustness of the third established QSAR-NN<sub>3</sub> neural model, and the possibility of predicting the different parameters that characterise the rejection of organic molecules during NF/RO.

The performance of each model was evaluated in terms of correlation coefficient ( $R$ ) and Root Mean Squared Error (RMSE). RMSE was used to determine the modelling error between the modelled and measured values of responses. RMSE is defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Y_{i,\text{exp}} - Y_{i,\text{cal}})^2}{n}} \quad (10)$$

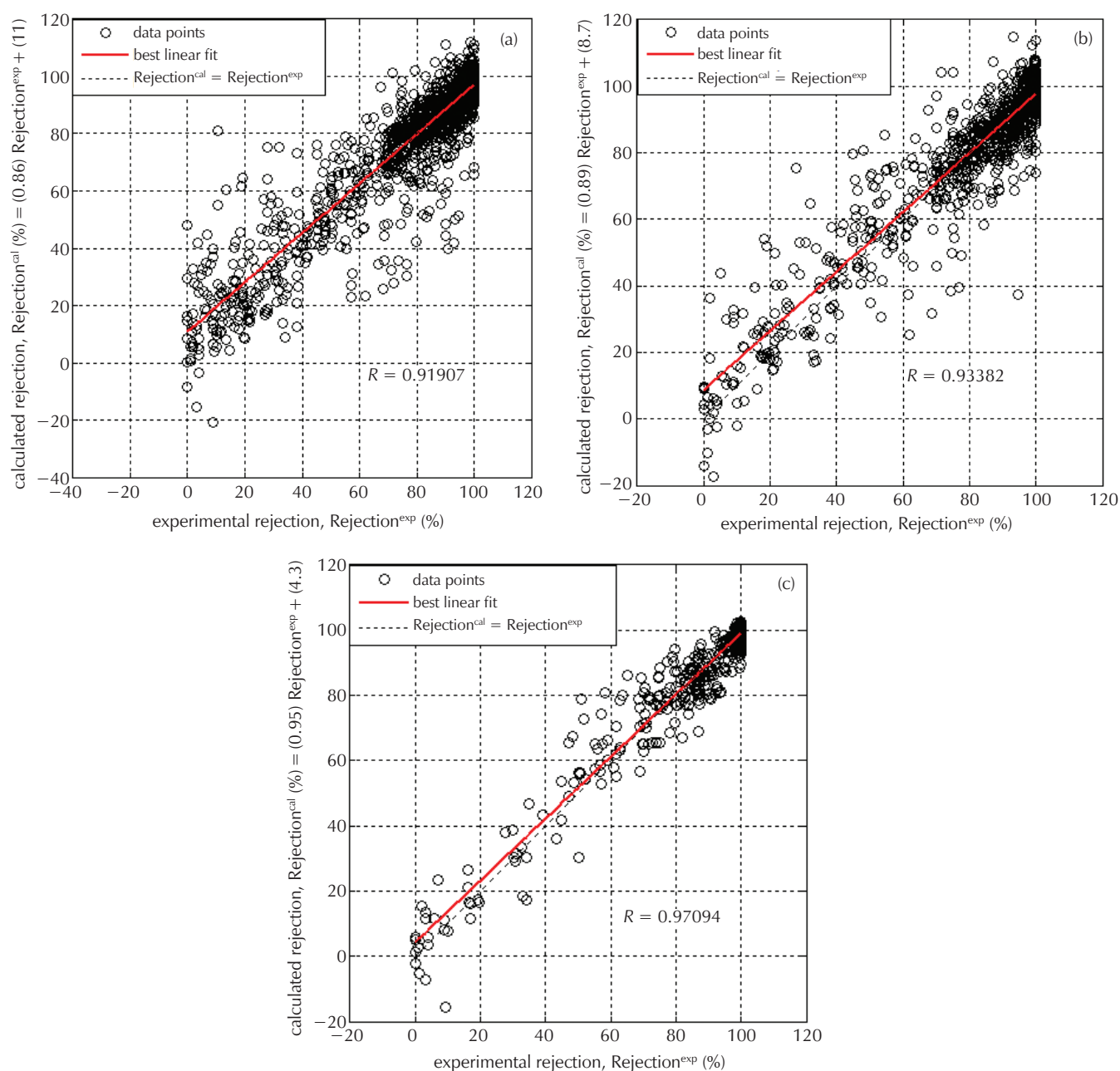


Fig. 4 – Comparison between experimental and calculated values for the total data set: (a) QSAR-NN<sub>1</sub>, (b) QSAR-NN<sub>2</sub>, (c) QSAR-NN<sub>3</sub>

Table 3 – Linear regression vectors [linear equation:  $y^{\text{cal}} = \alpha y^{\text{exp}} + \beta$ , with  $\alpha$  = slope,  $\beta$  = intercept,  $R$  = correlation coefficient], and RMSE (Root Mean Squared Error)

		$\alpha$	$\beta$	$R$	RMSE
QSAR-NN <sub>1</sub>	training phase	0.8657	10.5293	0.9303	9.8773
	validation phase	0.7833	17.3876	0.8586	14.8824
	test phase	0.9165	5.5148	0.9013	11.5353
	total	0.8608	10.8711	0.9191	10.6517
QSAR-NN <sub>2</sub>	training phase	0.8960	8.2197	0.9467	8.2870
	validation phase	0.8661	9.1982	0.8247	13.8828
	test phase	0.8665	11.8542	0.9324	10.2032
	total	0.8896	8.7206	0.9338	9.1991
QSAR-NN <sub>3</sub>	training phase	0.9444	4.6665	0.9722	5.6524
	validation phase	1.0000	-0.6412	0.9289	8.574
	test phase	0.9428	4.7299	0.9908	4.2211
	total	0.9481	4.3043	0.9709	5.8869

Table 3 shows the vectors of linear regression for the three neural models (QSAR-NN<sub>1</sub>, QSAR-NN<sub>2</sub>, and QSAR-NN<sub>3</sub>). This table shows the parameters of the least favourable regression in QSAR-NN<sub>1</sub>. It is obvious that the proposed approach gives satisfactory results with regression vector values approaching the ideal [i.e.  $\alpha = 1$  (slope),  $\beta = 0$  (intercept),  $R = 1$  (correlation coefficient)] in the adjustment of the profiles of QSAR-NN<sub>1</sub>, QSAR-NN<sub>2</sub>, and QSAR-NN<sub>3</sub>.

### 3.2 Multiple linear regressions

Multiple linear regressions model relates a set of "predictor" variables to the potency of the response variable (Rejection). It was used to develop the models using the organic compounds properties, membrane characteristics, and filtration operating conditions to predict the rejection of charged and uncharged organic compounds by NF/RO membranes. The same datasets used for developing QSAR-NN were used to develop QSAR-MLR. They were implemented using MATLAB function "regress". The general QSAR-MLR linear equations for rejection were as follows:

QSAR-MLR<sub>1</sub>

$$\begin{aligned} \text{Rejection} = & -15.0389 + 0.0606\text{MW} - 1.1476\log\text{D} + \\ & + 0.4326\text{dipole} + 9.0773\text{length} + 194.8995\text{eqwidth} - \\ & - 0.1228\text{MWCO} - 0.1378\text{zetapotential} + \\ & + 0.3086\text{contactangle} + 2.5678\text{pH} + 0.0022\text{pressure} - \\ & - 0.1995\text{recovery} - 0.4797\text{temperature} \end{aligned} \quad (11)$$

QSAR-MLR<sub>2</sub>

$$\begin{aligned} \text{Rejection} = & -8.0195 + 0.1033\text{MW} + 0.2602\log\text{D} + \\ & + 0.6966\text{dipole} + 4.6426\text{length} + 163.2833\text{eqwidth} - \\ & - 1.1655\text{SR}(\text{MgSO}_4) - 0.1180\text{zetapotential} + \\ & + 0.0146\text{contactangle} + 1.2827\text{pH} + 0.0026\text{pressure} - \\ & - 0.1935\text{recovery} - 0.7092\text{temperature} \end{aligned} \quad (12)$$

QSAR-MLR<sub>3</sub>

$$\begin{aligned} \text{Rejection} = & -90.0110 - 0.0338\text{MW} - 2.4398\log\text{D} + \\ & + 0.7469\text{dipole} - 13.3283\text{length} + 451.7653\text{Eqwidth} + \\ & + 26.6623\text{SR}(\text{NaCl}) - 0.0507\text{zetapotential} + \\ & + 0.0118\text{contactangle} + 4.1022\text{pH} + 0.0005\text{pressure} - \\ & - 0.1851\text{recovery} - 0.5264\text{temperature} \end{aligned} \quad (13)$$

The prediction of the rejection processes of charged and uncharged organic compounds by NF/RO membranes according to the QSAR-MLR models are presented in Fig. 5. The predicted values for the rejection were in moderate agreement with the observed values "total datasets" ( $R = 0.6642$  for QSAR-MLR<sub>1</sub>,  $R = 0.6560$  for QSAR-MLR<sub>2</sub>, and  $R = 0.7522$  for QSAR-MLR<sub>3</sub>). These correlation coefficients are generally considered satisfactory ( $0.50 \leq R < 0.90$ ) for QSAR-MLR<sub>1</sub>, QSAR-MLR<sub>2</sub>, and QSAR-MLR<sub>3</sub> models, which implies acceptable robustness of the multiple linear regressions models, and the possibility of predicting the different parameters that characterise the rejection of organic compounds by NF/RO.

## 4 Comparisons

### 4.1 Comparison between QSAR-NN and QSAR-MLR models

In order to establish the developed QSAR-NN models as a plausible alternative to the QSAR models, a comparison between the two approaches was made in terms of the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Standard Error of Prediction (SEP), Residual Predictive Deviation (RPD), and Range Error Ratio (RER). Equations of those parameters are given below:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n \left| (y_{i,\text{exp}} - y_{i,\text{cal}}) \right| \quad (14)$$

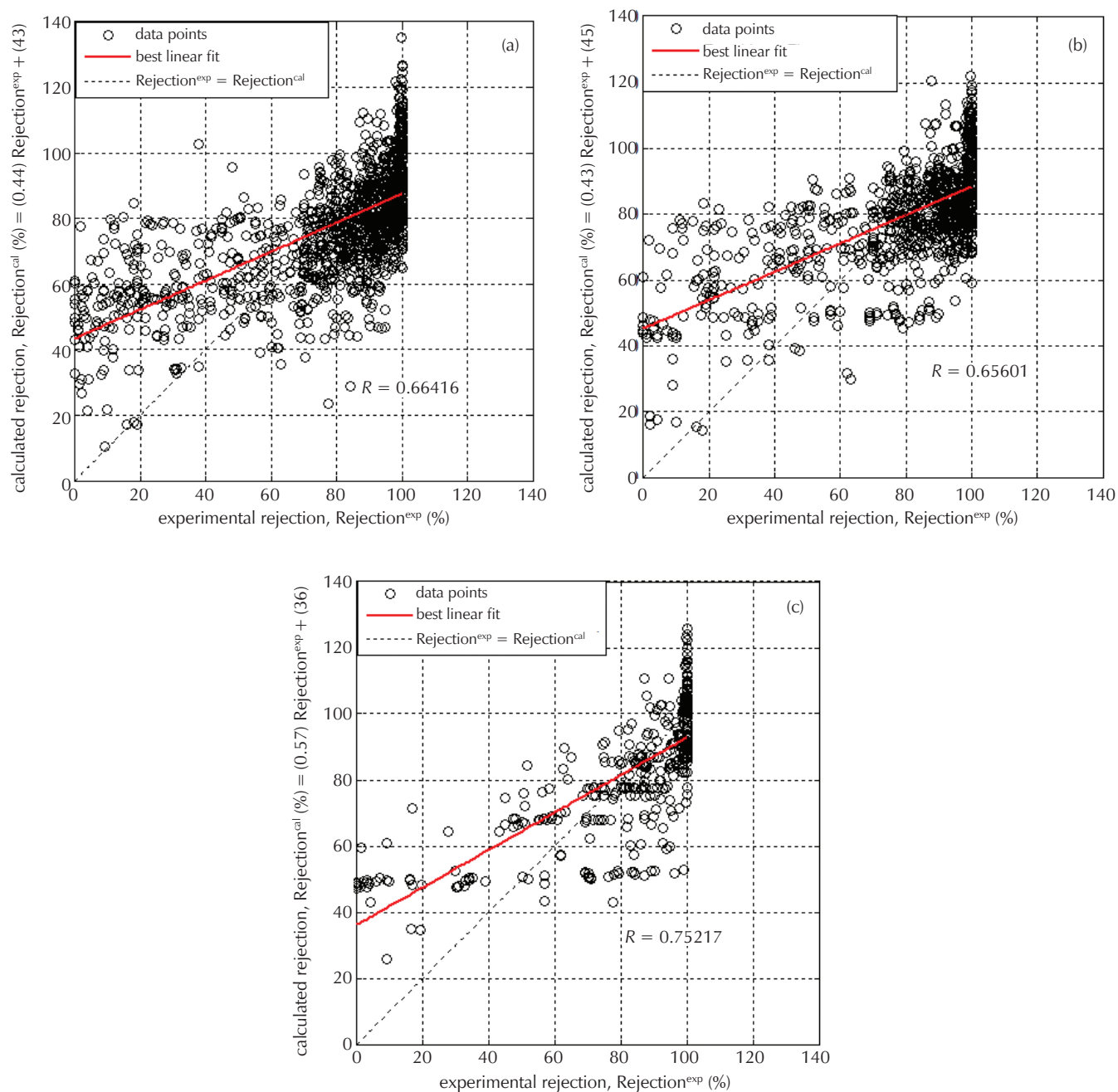


Fig. 5 – QSAR-MLR models: predicted vs. observed plot for total database: (a) QSAR-MLR<sub>1</sub>, (b) QSAR-MLR<sub>2</sub>, (c) QSAR-MLR<sub>3</sub>

$$SEP(\%) = \frac{RMSE}{Y_e} \cdot 100 \quad (15)$$

$$RPD = \frac{SD}{RMSE} \quad (16)$$

$$RER = \frac{Max - Min}{RMSE} \quad (17)$$

where  $n$  is the total number of data;  $Y_{i,exp}$  is the experimental value,  $Y_{i,cal}$  represents the calculated value from the neural network model,  $Y_e$  is the mean value of experimental data,  $SD$  is the standard deviation of experimental data,  $Min$  is the minimum of experimental data, and  $Max$  is the maximum of experimental data.

Fig. 6 shows the comparison between the QSAR-MLR and the QSAR-NN models. The performance of each model was evaluated in terms of root mean squared error, mean absolute error, standard error of prediction, residual predictive deviation, and range error ratio. As expected, QSAR-NN<sub>1</sub> model had the lowest accuracy. Another conclusion, which may be extracted from Fig. 6, is that QSAR-NN<sub>2</sub> model and QSAR-NN<sub>3</sub> model, which use experimental rejection, could be more robust and more accurate than the QSAR-NN<sub>1</sub> model. It shows that QSAR-NN<sub>3</sub> model is more effective and more accurate than other developed QSAR-NN models in this study, while the QSAR-NN models gave lower errors than QSAR-MLR models. The errors of QSAR-NN<sub>3</sub> model were lower than the errors of QSAR-NN<sub>1</sub> and QSAR-NN<sub>2</sub> models, and the errors of QSAR-



MLR<sub>3</sub> model was lower than the errors of QSAR-MLR<sub>1</sub> and QSAR-MLR<sub>2</sub> models. These results confirm that chlorides sodium salt rejection may be more appropriate for modelling the rejection of charged and uncharged organic compounds by NF/RO membranes using neural networks.

We adopted the five-level interpretations of RPD provided by Viscarra Rossel: excellent predictions (RPD > 2.5); good (RPD of 2.0–2.5); approximate quantitative predictions (RPD of 1.8–2.0); possibility to distinguish high and low values (RPD of 1.4–1.8); and unsuccessful (RPD < 1.40).<sup>53</sup> Since the RPD of QSAR-NN<sub>1</sub>, QSAR-NN<sub>2</sub>, and QSAR-NN<sub>3</sub> were 2.54, 2.79, and 4.18 respectively, and they were higher than 2.5 for all neural networks models. Therefore, we can say that QSAR-NN models had excellent levels of prediction for the rejection of micropollutants by NF/RO membranes.

On the other hand, the RPD of QSAR-MLR<sub>1</sub> and QSAR-MLR<sub>2</sub> were 1.34 and 1.33 respectively, being lower than 1.40, thus these models were unsuccessful in predicting the rejection of charged and uncharged organic compounds by NF/RO membranes. However, the RPD of QSAR-MLR<sub>3</sub> was 1.52, *i.e.*, between 1.40 and 1.8, and as a result, these models showed a possibility to distinguish high and low values. These results prove that QSAR-NN has a higher capacity of prediction than QSAR-MLR.

Furthermore, the values of RER of QSAR-NN<sub>2</sub> and QSAR-NN<sub>3</sub> were 10.78 and 16.99 respectively, being higher than the recommended (RER > 10), which implies that these

models showed a good predicting ability of the rejection. Additionally, the RER values of QSAR-NN<sub>1</sub>, QSAR-MLR<sub>1</sub>, QSAR-MLR<sub>2</sub>, and QSAR-MLR<sub>3</sub> (ranged from 4.95 to 9.39) in this study, did not exceed the recommended threshold values. The obtained results show that these models have an acceptable predicting power.

Thus, the good predicting power and the high accuracy of QSAR-NN models to estimate the rejection of micropollutants by NF/RO membranes is quite obvious, which makes these models highly recommendable over the other models (QSAR-MLR).<sup>28</sup>

#### 4.2 Comparison between our QSAR-NN and other models

Table 4 shows a comparison between our QSAR-NN models and other models. In our current study, we used the three QSAR-NN for the prediction of the rejection of micropollutants by NF/RO membranes, taking into account three porosity indicators (MWCO, SR(MgSO<sub>4</sub>), and SR(NaCl)). However, in the work of *Yangali-Quintanilla et al.*,<sup>28</sup> they used only one porosity indicator SR(MgSO<sub>4</sub>) for the neutral organic compounds, and in our earlier study, we were using two porosity indicators (MWCO and SR(MgSO<sub>4</sub>)) for the neutral and ionic organic compounds. Our QSAR-NN<sub>3</sub> models with porosity indicators SR(NaCl) give a more accurate and more robust prediction of the rejection of organic compounds (neutral and ionic) by NF/RO membranes than other models.

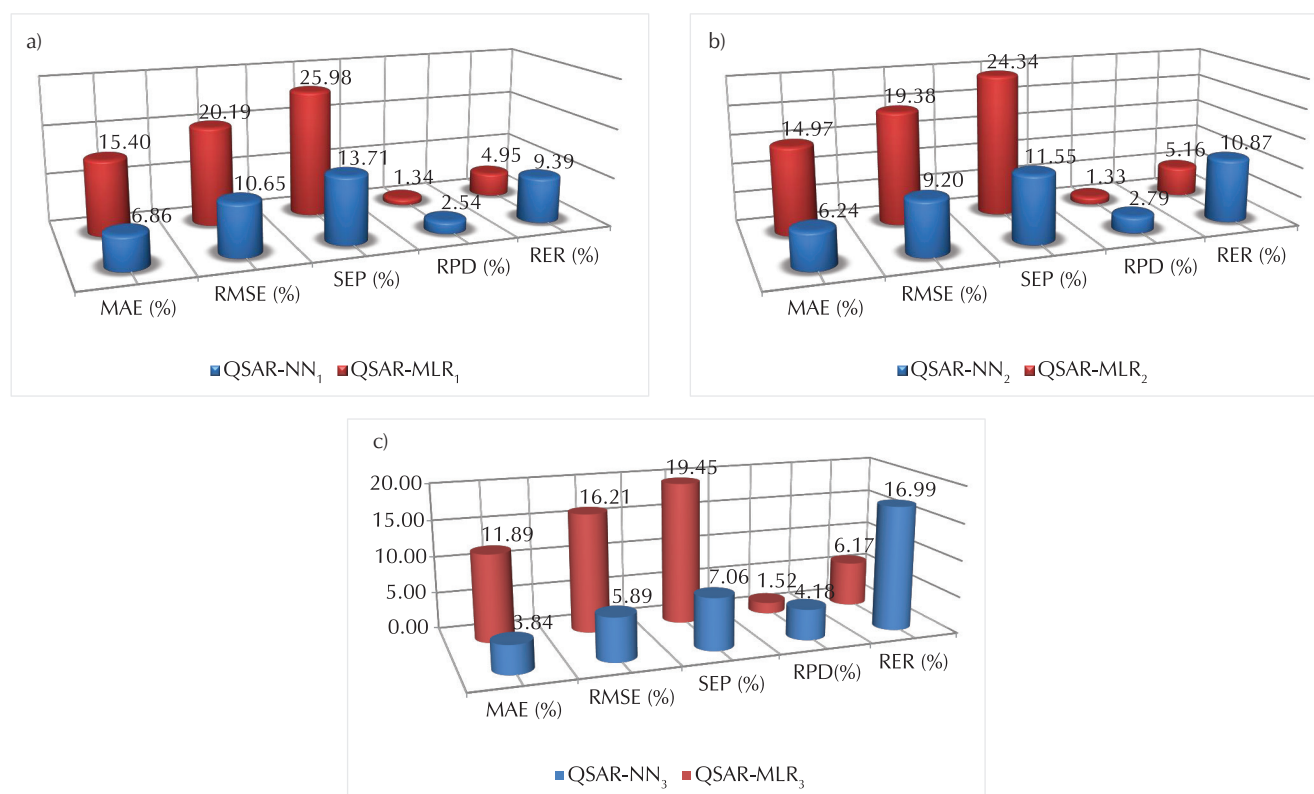


Fig. 6 – Comparison between the QSAR-MLR and QSAR-NN models: (a) models with MWCO, (b) models with SR(MgSO<sub>4</sub>), (c) models with SR(NaCl)

Table 4 – Overview of various works on models QSAR-NN for predicting the rejection of micropollutants by NF/RO

Name of model	Porosity indicators	Database size	Number and nature of organic compounds	R	MAE/%	RMSE/%	SEP/%
Yangali-Quintanilla et al. <sup>28</sup>	SR(MgSO <sub>4</sub> )	161	50 neutral	0.9848	–	5.0000	–
Our work in 2015 <sup>29</sup>	MWCO	965	82 neutral and ionic	0.9128	7.9138	11.2430	15.0730
	SR(MgSO <sub>4</sub> )	701	81 neutral and ionic	0.9419	6.6125	8.2047	12.1087
Our work in 2018 <sup>29</sup>	MWCO	1394	116 neutral and ionic	0.9191	6.8647	10.6517	13.7077
	SR(MgSO <sub>4</sub> )	980	102 neutral and ionic	0.9338	6.2388	9.1991	11.5543
	SR(NaCl)	436	42 neutral and ionic	0.9709	3.8403	5.8869	7.0640

## 5 Selected organic compounds rejection at varying feedwater pH values

The experimental and calculated rejection values of the selected organic compound (2,4-dihydroxybenzoic acid) at different pH are shown in Fig. 7 for both NF-90 and NF-200 membranes for QSAR-NN<sub>3</sub> model.

The calculated rejection values are in accordance with the experimental rejection values for the organic compounds (2,4-dihydroxybenzoic acid) (Fig. 7). These results suggested an excellent validation of the QSAR-NN<sub>3</sub> model. The increase in experimental rejection observed at a pH between 3 and 5 for both nanofiltration membranes can likely be attributed to an increase in the negative effective membrane surface charge (as quantified by both zeta potential measurements), resulting in an increased degree of electrostatic repulsion. At pH values of 5, 7, and 9, the rejection of 2,4-dihydroxybenzoic acid remained at about 90 % for both nanofiltration membranes tested, although the negative surface charge continued to increase for both nanofiltration membranes in this pH range.<sup>17</sup> The QSAR-NN<sub>3</sub> model can be used to predict rejection of organic compounds in a wide range of pH.

## 6 Conclusions

The present paper illustrates the use of three neural network models that were developed with the aim of predicting the rejection processes of charged and uncharged organic compounds by NF/RO membranes. The purpose of the current study was to develop three feed-forward neural network models able to summarize interactions between membrane characteristics, filtration operating conditions, and physicochemical properties of charged and uncharged organic compounds.

Comparison between neural network (QSAR-NN) and multiple linear regressions (QSAR-MLR) revealed that QSAR-NN model gave the best performance: The root mean squared errors for the total dataset for the QSAR-NN models were 10.6517 % for QSAR-NN<sub>1</sub>, 9.1991 % for QSAR-NN<sub>2</sub>, and 5.8869 % for QSAR-NN<sub>3</sub>, while for QSAR-MLR models they were 20.1865 % for

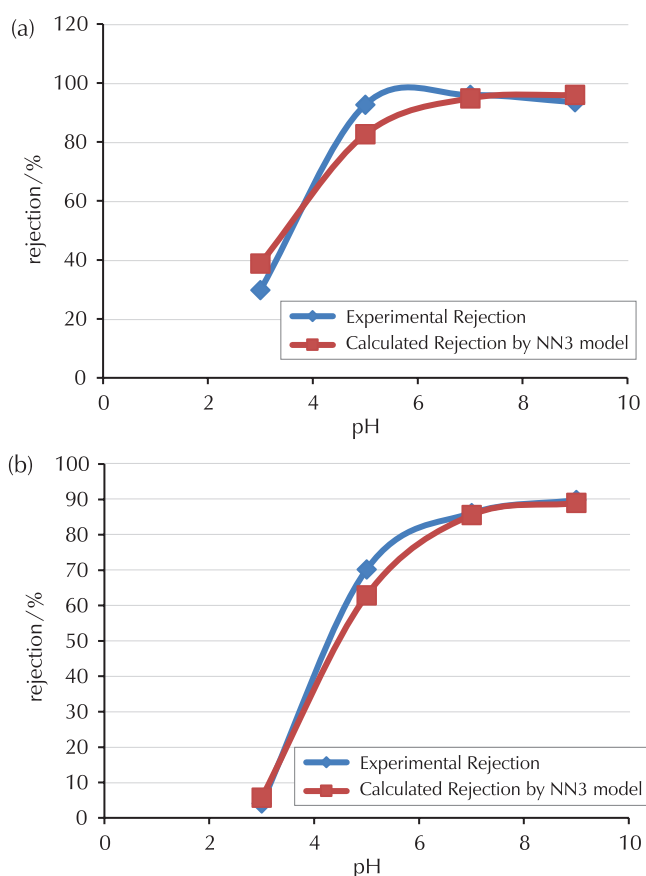


Fig. 7 – Comparison between the experimental and calculated rejection (QSAR-NN<sub>3</sub> model) of 2,4-dihydroxybenzoic acid by NF-90 (a) and NF-200 (b) vs. feedwater pH

QSAR-MLR<sub>1</sub>, 19.3815 % for QSAR-MLR<sub>2</sub>, and 16.2062 % for QSAR-MLR<sub>3</sub>. This study has demonstrated the superiority of neural networks in predicting the rejection processes of charged and uncharged organic compounds by NF/RO membranes.

Chloride sodium salt rejection may be a possible lump parameter for modelling rejection of charged and uncharged organic compounds by NF/RO membranes.

## ACKNOWLEDGEMENTS

Authors gratefully acknowledge the team of the Laboratory of Biomaterials and Transport Phenomena, the University of Médéa, and University Center of Relizane for their help throughout this project. The authors also thank the anonymous reviewers for their constructive comments, which helped to improve the quality and presentation of this paper.

## References

## Literatura

1. B. Van der Bruggen, C. Vandecasteele, Modelling of the retention of uncharged molecules with nanofiltration, *Water Res.* **36** (5) (2002) 1360–1368, doi: [https://doi.org/10.1016/S0043-1354\(01\)00318-9](https://doi.org/10.1016/S0043-1354(01)00318-9).
2. T. O. Mahlangu, E. M. V. Hoek, B. B. Mamba, A. R. D. Verliefde, Influence of organic, colloidal and combined fouling on NF rejection of NaCl and carbamazepine: Role of solute–foulant–membrane interactions and cake-enhanced concentration polarisation, *J. Membr. Sci.* **471** (0) (2014) 35–46, doi: <https://doi.org/10.1016/j.memsci.2014.07.065>.
3. N. Bolong, A. F. Ismail, M. R. Salim, T. Matsuura, A review of the effects of emerging contaminants in wastewater and optional for their removal, *Desalination* **239** (2009) 229–246, doi: <https://doi.org/10.1016/j.desal.2008.03.020>.
4. D. Dolar, T. Ignjatić Zokić, K. Košutić, D. Ašperger, D. Mutavdžić Pavlović, RO/NF membrane treatment of veterinary pharmaceutical wastewater: comparison of results obtained on a laboratory and a pilot scale, *Environ. Sci. Pollut. Res. Int.* **19** (4) (2012) 1033–1042, doi: <https://doi.org/10.1007/s11356-012-0782-7>.
5. V. Yangali-Quintanilla, A. Sadmani, M. McConville, M. Kennedy, G. Amy, A QSAR model for predicting rejection of emerging contaminants (pharmaceuticals, endocrine disruptors) by nanofiltration membranes, *Water Res.* **44** (2) (2010) 373–384, doi: <https://doi.org/10.1016/j.watres.2009.06.054>.
6. A. R. Verliefde, S. G. Heijman, E. R. Cornelissen, G. Amy, B. Van der Bruggen, J. C. van Dijk, Influence of electrostatic interactions on the rejection with NF and assessment of the removal efficiency during NF/GAC treatment of pharmaceutically active compounds in surface water, *Water Res.* **41** (15) (2007) 3227–3240, doi: <https://doi.org/10.1016/j.watres.2007.05.022>.
7. L. D. Nghiem, A. I. Schäfer, M. Elimelech, Role of electrostatic interactions in the retention of pharmaceutically active contaminants by a loose nanofiltration membrane, *J. Membr. Sci.* **286** (1-2) (2006) 52–59, doi: <https://doi.org/10.1016/j.memsci.2006.09.011>.
8. A. R. D. Verliefde, E. R. Cornelissen, S. G. J. Heijman, J. Q. J. C. Verberk, G. L. Amy, B. Van der Bruggen, J. C. van Dijk, The role of electrostatic interactions on the rejection of organic solutes in aqueous solutions with nanofiltration, *J. Membr. Sci.* **322** (1) (2008) 52–66, doi: <https://doi.org/10.1016/j.memsci.2008.05.022>.
9. C. Bellona, J. E. Drewes, P. Xu, G. Amy, Factors affecting the rejection of organic solutes during NF/RO treatment – a literature review, *Water Res.* **38** (12) (2004) 2795–2809, doi: <https://doi.org/10.1016/j.watres.2004.03.034>.
10. S. G. J. Heijman, A. R. D. Verliefde, E. R. Cornelissen, G. Amy, J. C. Van Dijk, Influence of natural organic matter (NOM) fouling on the removal of pharmaceuticals by nanofiltration and activated carbon filtration, *Water Sci. Technol.* **7** (4) (2007) 17–24, doi: <https://doi.org/10.2166/ws.2007.131>.
11. S. Gur-Reznik, I. Koren-Menashe, L. Heller-Grossman, O. Rufel, C. G. Dosoretz, Influence of seasonal and operating conditions on the rejection of pharmaceutical active compounds by RO and NF membranes, *Desalination* **277** (1-3) (2011) 250–256, doi: <https://doi.org/10.1016/j.desal.2011.04.029>.
12. Y. Kiso, Y. Sugiura, T. Kitao, K. Nishimura, Effects of hydrophobicity and molecular size on rejection of aromatic pesticides with nanofiltration membranes, *J. Membr. Sci.* **192** (1-2) (2001) 1–10, doi: [https://doi.org/10.1016/S0376-7388\(01\)00411-2](https://doi.org/10.1016/S0376-7388(01)00411-2).
13. A. I. Schäfer, L. D. Nghiem, T. D. Waite, Removal of the Natural Hormone Estrone from Aqueous Solutions Using Nanofiltration and Reverse Osmosis, *Environ. Sci. Technol.* **37** (1) (2003) 182–188, doi: <https://doi.org/10.1021/es0102336>.
14. K. Kimura, G. Amy, J. E. Drewes, T. Heberer, T.-U. Kim, Y. Watanabe, Rejection of organic micropollutants (disinfection by-products, endocrine disrupting compounds, and pharmaceutically active compounds) by NF/RO membranes, *J. Membr. Sci.* **227** (1–2) (2003) 113–121, doi: <https://doi.org/10.1016/j.memsci.2003.09.005>.
15. L. D. Nghiem, A. I. Schäfer, M. Elimelech, Removal of natural hormones by nanofiltration membranes: measurement, modeling, and mechanisms, *Environ. Sci. Technol.* **38** (6) (2004) 1888–1896, doi: <https://doi.org/10.1021/es034952r>.
16. H. Ozaki, H. Li, Rejection of organic compounds by ultra-low pressure reverse osmosis membrane, *Water Res.* **36** (1) (2002) 123–130, doi: [https://doi.org/10.1016/S0043-1354\(01\)00197-X](https://doi.org/10.1016/S0043-1354(01)00197-X).
17. C. Bellona, J. E. Drewes, The role of membrane surface charge and solute physico-chemical properties in the rejection of organic acids by NF membrane, *J. Membr. Sci.* **249** (1-2) (2005) 227–234, doi: <https://doi.org/10.1016/j.memsci.2004.09.041>.
18. P. Xu, J. E. Drewes, C. Bellona, G. Amy, T. U. Kim, M. Adam, T. Heberer, Rejection of emerging organic micropollutants in nanofiltration-reverse osmosis membrane applications, *Water Environ. Res.* **77** (1) (2005) 40–48, doi: <https://doi.org/10.2175/106143005X41609>.
19. K. Kimura, S. Toshima, G. Amy, Y. Watanabe, Rejection of neutral endocrine disrupting compounds (EDCs) and pharmaceutical active compounds (PhACs) by RO membranes, *J. Membr. Sci.* **245** (1-2) (2004) 71–78, doi: <https://doi.org/10.1016/j.memsci.2004.07.018>.
20. B. Van Der Bruggen, J. Schaep, D. Wilms, C. Vandecasteele, A Comparison of Models to Describe the Maximal Retention of Organic Molecules in Nanofiltration, *Separat. Sci. Technol. Separat. Sci. Technol.* **35** (2) (2000) 169–182, doi: <https://doi.org/10.1081/SS-100100150>.
21. Y. Kiso, Y. Nishimura, T. Kitao, K. Nishimura, Rejection properties of non-phenylic pesticides with nanofiltration membranes, *J. Membr. Sci.* **171** (2) (2000) 229–237, doi: [https://doi.org/10.1016/S0376-7388\(00\)00305-7](https://doi.org/10.1016/S0376-7388(00)00305-7).
22. K. Agenson, J. Oh, T. Urase, Retention of a wide variety of organic pollutants by different nanofiltration/reverse osmosis membranes: controlling parameters of process, *J. Membr. Sci.* **225** (1-2) (2003) 91–103, doi: <https://doi.org/10.1016/j.memsci.2003.08.006>.
23. E. R. Cornelissen, J. Verdouw, A. J. Gijsbertsen-Abrahamse, J. A. M. H. Hofman, A nanofiltration retention model for trace contaminants in drinking water sources, *Desalination* **178** (1-3) (2005) 179–192, doi: <https://doi.org/10.1016/j.desal.2004.11.047>.
24. T.-U. Kim, J. E. Drewes, R. Scott Summers, G. L. Amy, Solute transport model for trace organic neutral and charged compounds through nanofiltration and reverse osmosis mem-



- branes, *Water Res.* **41** (17) (2007) 3977–3988, doi: <https://doi.org/10.1016/j.watres.2007.05.055>.
25. M. Dornier, M. Decloux, G. Trystram, G. Lebert, Dynamic modeling of crossflow microfiltration using neural networks, *J. Membr. Sci.* **98** (3) (1995) 263–273, doi: [https://doi.org/10.1016/0376-7388\(94\)00195-5](https://doi.org/10.1016/0376-7388(94)00195-5).
  26. G. B. Sahoo, C. Ray, Predicting flux decline in crossflow membranes using artificial neural networks and genetic algorithms, *J. Membr. Sci.* **283** (1-2) (2006) 147–157, doi: <https://doi.org/10.1016/j.memsci.2006.06.019>.
  27. D. Libotean, J. Giralt, F. Giralt, R. Rallo, T. Wolfe, Y. Cohen, Neural network approach for modeling the performance of reverse osmosis membrane desalting, *J. Membr. Sci.* **326** (2) (2009) 408–419, doi: <https://doi.org/10.1016/j.memsci.2008.10.028>.
  28. V. Yangali-Quintanilla, A. Verliefe, T. U. Kim, A. Sadmani, M. Kennedy, G. Amy, Artificial neural network models based on QSAR for predicting rejection of neutral organic compounds by polyamide nanofiltration and reverse osmosis membranes, *J. Membr. Sci.* **342** (1–2) (2009) 251–262, <https://doi.org/10.1016/j.memsci.2009.06.048>.
  29. Y. Ammi, L. Khaouane, S. Hanini, Prediction of the rejection of organic compounds (neutral and ionic) by nanofiltration and reverse osmosis membranes using neural networks. *Korean J. Chem. Eng.* **32** (11) (2015) 2300–2310. <http://doi:10.1007/s11814-015-0086-y>.
  30. Y. Ammi, L. Khaouane, S. Hanini, A Model Based on Bootstrapped Neural Networks for Modeling the Removal of Organic Compounds by Nanofiltration and Reverse Osmosis Membranes, *Arab. J. Sci. Eng.* **43** (11) (2018) 6271–6284, doi: <https://doi.org/10.1007/s13369-018-3484-8>.
  31. Y. Yoon, R. M. Lueptow, Removal of organic contaminants by RO and NF membranes, *J. Membr. Sci.* **261** (1-2) (2005) 76–86, doi: <https://doi.org/10.1016/j.memsci.2005.03.038>.
  32. A. R. D. Verliefe, E. R. Cornelissen, S. G. J. Heijman, I. Petrinic, T. Luxbacher, G. L. Amy, B. Van der Bruggen, J. C. van Dijk, Influence of membrane fouling by (pretreated) surface water on rejection of pharmaceutically active compounds (PhACs) by nanofiltration membranes, *J. Membr. Sci.* **330** (1-2) (2009) 90–103, doi: <https://doi.org/10.1016/j.memsci.2008.12.039>.
  33. A. R. D. Verliefe, E. R. Cornelissen, S. G. J. Heijman, J. Q. J. C. Verberk, G. L. Amy, B. Van der Bruggen, J. C. van Dijk, Construction and validation of a full-scale model for rejection of organic micropollutants by NF membranes, *J. Membr. Sci.* **339** (1-2) (2009) 10–20, doi: <https://doi.org/10.1016/j.memsci.2009.03.038>.
  34. A. M. Comerton, R. C. Andrews, D.M. Bagley, C. Hao, The rejection of endocrine disrupting and pharmaceutically active compounds by NF and RO membranes as a function of compound and water matrix properties, *J. Membr. Sci.* **313** (1-2) (2008) 323–335, doi: <https://doi.org/10.1016/j.memsci.2008.01.021>.
  35. A.M. Comerton, R.C. Andrews, D.M. Bagley, The influence of natural organic matter and cations on the rejection of endocrine disrupting and pharmaceutically active compounds by nanofiltration, *Water Res.* **43** (3) (2009) 613–622, doi: <https://doi.org/10.1016/j.watres.2008.11.003>.
  36. V. Yangali-Quintanilla, M. Kennedy, G. Amy, T. U. Kim, Modeling of RO/NF membrane rejections of PhACs and organic compounds: A statistical analysis, *Drink. Water Eng. Sci.* **1** (1) (2008) 7–15, doi: <https://doi.org/10.5194/dwes-1-7-2008>.
  37. V. Yangali-Quintanilla, A. Sadmani, M. McConville, M. Kennedy, G. Amy, Rejection of pharmaceutically active compounds and endocrine disrupting compounds by clean and fouled nanofiltration membranes, *Water Res.* **43** (9) (2009) 2349–2362, doi: <https://doi.org/10.1016/j.watres.2009.02.027>.
  38. V. A. Yangali Quintanilla, Rejection of emerging organic contaminants by nanofiltration and reverse osmosis membranes effects of fouling, modelling and water reuse, Doctorate Thesis, 2010, University of Delft of Technology (Netherlands), doi: <https://doi.org/10.1201/b10832>.
  39. C. Y. Tang, Y.-N. Kwon, J. O. Leckie, Effect of membrane chemistry and coating layer on physicochemical properties of thin film composite polyamide RO and NF membranes: II. Membrane physicochemical properties and their dependence on polyamide and coating layers, *Desalination* **242** (1-3) (2009) 168–182, doi: <https://doi.org/10.1016/j.desal.2008.04.004>.
  40. H. Huang, H. Cho, K. Schwab, J. G. Jacangelo, Effects of feed-water pretreatment on the removal of organic microconstituents by a low fouling reverse osmosis membrane, *Desalination* **281** (0) (2011) 446–454, doi: <https://doi.org/10.1016/j.desal.2011.08.018>.
  41. K. V. Plakas, A. J. Karabelas, A systematic study on triazine retention by fouled with humic substances NF/ULPRO membranes, *Separat. Purif. Technol.* **80** (2) (2011) 246–261, doi: <https://doi.org/10.1016/j.seppur.2011.05.003>.
  42. D. Dolar, A. Vuković, D. Ašperger, K. Košutić, Effect of water matrices on removal of veterinary pharmaceuticals by nanofiltration and reverse osmosis membranes, *J. Environ. Sci.* **23** (8) (2011) 1299–1307, doi: [https://doi.org/10.1016/S1001-0742\(10\)60545-1](https://doi.org/10.1016/S1001-0742(10)60545-1).
  43. D. Dolar, K. Košutić, D. Ašperger, Influence of Adsorption of Pharmaceuticals onto RO/NF Membranes on Their Removal from Water, *Water Air Soil Pollut. Water* **224** (1) (2013) 1–13, doi: <https://doi.org/10.1007/s11270-012-1377-0>.
  44. D. Dolar, K. Košutić, M. Periša, S. Babić, Photolysis of enrofloxacin and removal of its photodegradation products from water by reverse osmosis and nanofiltration membranes, *Separat. Purif. Technol.* **115** (13) (2013) 1–8, doi: <https://doi.org/10.1016/j.seppur.2013.04.042>.
  45. A. H. M. A. Sadmani, R. C. Andrews, D. M. Bagley, Impact of natural water colloids and cations on the rejection of pharmaceutically active and endocrine disrupting compounds by nanofiltration, *J. Membr. Sci.* **450** (0) (2014) 272–281, doi: <https://doi.org/10.1016/j.memsci.2013.09.017>.
  46. A. H. M. A. Sadmani, R. C. Andrews, D. M. Bagley, Nanofiltration of pharmaceutically active and endocrine disrupting compounds as a function of compound interactions with DOM fractions and cations in natural water, *Separat. Purif. Technol.* **122** (0) (2014) 462–471, doi: <https://doi.org/10.1016/j.seppur.2013.12.003>.
  47. <http://www.chemspider.com> (20. 3. 2015).
  48. <http://www.chemicalize.org> (20. 3. 2015).
  49. J. L. C. Santos, P. de Beukelaar, I. F. J. Vankelecom, S. Velizarov, J. G. Crespo, Effect of solute geometry and orientation on the rejection of uncharged compounds by nanofiltration, *Separat. Purif. Technol.* **50** (1) (2006) 122–131, doi: <https://doi.org/10.1016/j.seppur.2005.11.015>.
  50. A. Rezaei, S. Hanini, M. Laidi, An optimisation methodology of artificial neural network models for predicting solar radiation: a case study, *Theoret. Appl. Climatol.* **123** (3) (2016) 769–783, doi: <https://doi.org/10.1007/s00704-015-1398-x>.
  51. D. Kruzlicova, J. Mocak, B. Balla, J. Petka, M. Farkova, J. Havel, Classification of Slovak white wines using artificial neu-

- ral networks and discriminant techniques, Food Chem. **112** (4) (2009) 1046-1052, doi: <https://doi.org/10.1016/j.foodchem.2008.06.047>.
52. M. Hamadache, L. Khaouane, O. Benkortbi, C. Si Moussa, S. Hanini, Prediction of Acute Herbicide Toxicity in Rats from Quantitative Structure-Activity Relationship Modeling, Environ. Eng. Sci. **31** (5) (2014), doi: <https://doi.org/10.1089/ees.2013.0466>.
53. R. V. Rossel, R. N. McGlynn, A. B. McBratney, Determining the composition of mineral-organic mixes using UV-VIS-NIR diffuse reflectance spectroscopy, Geoderma **137** (2006) 70-82, doi: <https://doi.org/10.1016/j.geoderma.2006.07.004>.

## SAŽETAK

### Usporedba modela "neuronskih mreža i višestrukih linearnih regresija" za opisivanje odbacivanja mikroonečišćivala membranama

Yamina Ammi,<sup>a,b\*</sup> Latifa Khaouane<sup>b</sup> i Salah Hanini<sup>b</sup>

Postupak odbacivanja organskih spojeva nanofiltracijom i membranama reverzne osmoze modeliran je umjetnim neuronskim mrežama. Konstruirane su tri neuronske mreže zasnovane na kvantitativnom odnosu strukture-aktivnosti (QSAR-NN modeli) karakterizirane sličnom strukturom (dvanaest neurona za QSAR-NN<sub>1</sub>, QSAR-NN<sub>2</sub> i QSAR-NN<sub>3</sub> u ulaznom sloju, jedan skriveni sloj i jedan neuron u izlaznom sloju), s ciljem predviđanja odbacivanja organskih spojeva. Za izgradnju neuronskih mreža upotrijebljeni su skupovi od 1394 podatkovnih točaka za QSAR-NN<sub>1</sub>, 980 podatkovnih točaka za QSAR-NN<sub>2</sub> i 436 podatkovnih točaka za QSAR-NN<sub>3</sub>. Dobre usklađenosti između predviđenih i eksperimentalnih odbacivanja dobivene su modelima QSAR-NN (korelacijski koeficijent za ukupni skup podataka bio je 0,9191 za QSAR-NN<sub>1</sub>, 0,9338 za QSAR-NN<sub>2</sub> i 0,9709 za QSAR-NN<sub>3</sub>). Usporedba neuronskih mreža i višestrukih linearnih regresija zasnovanih na kvantitativnom odnosu struktura-aktivnost "QSAR-MLR" otkrila je superiornost modela QSAR-NN (korigirane srednje kvadratne pogreške za ukupni skup podataka za modele QSAR-NN bile su 10,6517 % za QSAR-NN<sub>1</sub>, 9,1991 % za QSAR-NN<sub>2</sub>, i 5,8869 % za QSAR-NN<sub>3</sub>, a za modele QSAR-MLR 20,1865 % za QSAR-MLR<sub>1</sub>, 19,3815 % za QSAR-MLR<sub>2</sub>, i 16,2062 % za QSAR-MLR<sub>3</sub>).

#### Ključne riječi

Neuronske mreže, višestruke linearne regresije, modeli, odbacivanje, mikroonečišćivala, membrane

<sup>a</sup> University Center Ahmed Zabana Relizane, 48 000, Algeria

<sup>b</sup> Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Médéa, 26 000, Algeria

Izvorni znanstveni rad  
Prispjelo 22. lipnja 2019.  
Prihvaćeno 19. listopada 2019.



## Supplementary data

**ANNEX A.** List of the 116 organic compounds used for developing neural networks based on the quantitative structure-activity relationship (QSAR-NN) and multiple linear regressions based on the quantitative structure-activity relationship (QSAR-MLR) models.

#	Name	Molecular formula	$M_w/g\ mol^{-1}$	CAS Number #
01	1,3,5-benzenetricarboxylic acid	$C_9H_6O_6$	210.14	554-95-0
02	1,4-dioxane	$C_4H_8O_2$	133.41	71-55-6
03	1,5-naphthalenedisulfonic acid	$C_{10}H_8O_6S_2$	288.3	81-04-9
04	17 $\alpha$ -ethinyl estradiol	$C_{20}H_{24}O_2$	296.40	57-63-6
05	17 $\alpha$ -ethynylestradiol	$C_{20}H_{24}O_3$	312.40	4717-40-2
06	17 $\alpha$ -estradiol	$C_{18}H_{24}O_2$	272.38	57-91-0
07	17 $\beta$ -estradiol	$C_{18}H_{24}O_2$	272.39	50-28-2
08	2-(1H)-quinoline	$C_9H_7NO$	145.16	59-31-4
09	2-(2-butoxyethoxy)ethanol	$C_8H_{18}O_3$	162.23	112-34-5
10	2,4-dihydroxybenzoic acid	$C_7H_6O_4$	154.02	89-86-1
11	2,4-dinitrophenol	$C_6H_4N_2O_5$	184.11	1326-82-5
12	2-ethoxyethanol	$C_4H_{10}O_2$	90.12	110-80-5
13	2-methoxyethanol	$C_3H_8O_2$	76.11	109-86-4
14	2-naphthalenesulfonic acid	$C_{10}H_8O_3S$	208.23	120-18-3
15	2-naphthol	$C_{10}H_8O$	144.17	135-19-3
16	4-phenylphenol	$C_{12}H_{10}O$	170.21	92-69-3
17	acetaminophen	$C_8H_9NO_2$	151.17	103-90-2
18	acetic acid	$C_2H_4O_2$	60.052	64-19-7
19	alachlor	$C_{14}H_{20}ClNO_2$	269.77	15972-60-8
20	aminopyrine	$C_{13}H_{17}N_3O$	231.30	58-15-1
21	antipyrine	$C_{11}H_{12}N_2O$	188.23	60-80-0
22	atenolol	$C_{14}H_{22}N_2O_3$	266.34	29122-68-7
23	atraton	$C_9H_{17}N_5O$	211.26	1610-17-9
24	atrazin	$C_8H_{14}ClN_5$	215.69	1912-24-9
25	atropine	$C_{17}H_{23}NO_3$	289.38	51-55-8
26	bentazon	$C_{10}H_{12}N_2O_3S$	240.28	25057-89-0
27	bezafibrate	$C_{19}H_{20}ClNO_4$	361.82	41859-67-0
28	bisphenol A	$C_{15}H_{16}O_2$	228.29	80-05-7
29	bromoform	$CHBr_3$	252.73	75-25-2
30	caffeine	$C_8H_{10}N_4O_2$	194.14	58-08-2
31	caprolactam	$C_6H_{11}NO$	113.16	105-60-2
32	carbadox	$C_{11}H_{10}N_4O_4$	262.22	41859-67-0
33	carbamazepine	$C_{15}H_{12}N_2O$	236.28	298-46-4
34	carbontetrabromide	$CBr_4$	131.63	558-13-4
35	carbontetrachloride	$CCl_4$	153.82	56-23-5
36	chloroform	$CHCl_3$	119.38	67-66-3
37	chlorotoluron	$C_{10}H_{13}ClN_2O$	212.68	15545-48-9
38	ciprofloxacin	$C_{17}H_{18}FN_3O_3$	331.35	85721-33-1
39	clenbuterol	$C_{12}H_{18}Cl_2N_2O$	277.20	37148-27-9
40	clofibrac acid	$C_{10}H_{11}ClO_3$	214.65	882-09-7
41	cotinine	$C_{10}H_{12}N_2O$	176.215	486-56-6
42	cyclophosphamide	$C_7H_{15}Cl_2N_2O_2P$	261.09	50-18-0
43	DEET	$C_{12}H_{17}NO$	191.27	134-62-3
44	dexamethasone	$C_{22}H_{29}FO_5$	392.46	50-02-2
45	dichloroacetic acid	$C_2H_2Cl_2O_2$	128.94	79-43-6
46	diclofenac	$C_{14}H_9Cl_2NO$	296.16	15307-86-5
47	diethylstilbesterol	$C_{18}H_{20}O_2$	268.35	56-53-1
48	dikegulac	$C_{12}H_{18}O_7$	274.27	18467-77-1
49	diuron	$C_9H_{10}Cl_2N_2O$	233.10	330-54-1
50	enrofloxacin	$C_{19}H_{22}FN_3O_3$	359.40	93106-60-6
51	equilin	$C_{18}H_{20}O_2$	268.36	474-86-2
52	estriol	$C_{18}H_{24}O_3$	288.39	50-27-1
53	estrone	$C_{18}H_{22}O_2$	270.37	53-16-7
54	ethanol	$C_2H_6O$	46.07	64-17-5

#	Name	Molecular formula	M <sub>w</sub> /g mol <sup>-1</sup>	CAS Number #
55	febantel	C <sub>20</sub> H <sub>22</sub> N <sub>4</sub> O <sub>6</sub> S	446.48	58306-30-2
56	fenoprofen	C <sub>15</sub> H <sub>14</sub> O <sub>3</sub>	242.28	31879-05-7
57	formaldehyde	CH <sub>2</sub> O	30.03	50-00-0
58	gemfibrozil	C <sub>15</sub> H <sub>22</sub> O <sub>3</sub>	250.34	25812-30-0
59	glucose	C <sub>6</sub> H <sub>12</sub> O <sub>6</sub>	180.16	50-99-7
60	glutaric acid	C <sub>5</sub> H <sub>8</sub> O <sub>4</sub>	132.115	110-94-1
61	glycerol	C <sub>3</sub> H <sub>8</sub> O <sub>3</sub>	92.10	56-81-5
62	hydrocortisone	C <sub>21</sub> H <sub>30</sub> O <sub>5</sub>	362.46	50-23-7
63	ibuprofen	C <sub>13</sub> H <sub>18</sub> O <sub>2</sub>	206.29	15687-27-1
64	isopropanol	C <sub>3</sub> H <sub>8</sub> O	60.10	67-63-0
65	isopropylantipyrine	C <sub>14</sub> H <sub>18</sub> N <sub>2</sub> O	230.30	479-92-5
66	isoproturon	C <sub>12</sub> H <sub>18</sub> N <sub>2</sub> O	206.29	34123-59-6
67	ketoprofen	C <sub>16</sub> H <sub>14</sub> O <sub>3</sub>	254.29	22071-15-4
68	lidocaine	C <sub>14</sub> H <sub>22</sub> N <sub>2</sub> O	234.4	137-58-6
69	lindane	C <sub>6</sub> H <sub>6</sub> Cl <sub>6</sub>	290.83	58-89-9
70	linuron	C <sub>9</sub> H <sub>10</sub> C <sub>12</sub> N <sub>2</sub> O <sub>2</sub>	249.10	330-55-2
71	lopressor	C <sub>15</sub> H <sub>25</sub> NO <sub>3</sub>	267.364	37350-58-6
72	mecoprop	C <sub>10</sub> H <sub>11</sub> ClO <sub>3</sub>	214.64	7085-19-0
73	methacetin	C <sub>9</sub> H <sub>11</sub> NO <sub>2</sub>	165.19	51-66-1
74	methanol	CH <sub>4</sub> O	32.04	67-56-1
75	metobromuron	C <sub>9</sub> H <sub>11</sub> BrN <sub>2</sub> O <sub>2</sub>	259.10	3060-89-7
76	metolachlor	C <sub>15</sub> H <sub>22</sub> ClNO <sub>2</sub>	283.79	51218-45-2
77	metoprolol	C <sub>15</sub> H <sub>25</sub> NO <sub>3</sub>	267.37	37350-58-6
78	metoxuron	C <sub>10</sub> H <sub>13</sub> ClN <sub>2</sub> O <sub>2</sub>	228.68	19937-59-8
79	metribuzin	C <sub>8</sub> H <sub>14</sub> N <sub>4</sub> OS	214.29	21087-64-9
80	metronidazole	C <sub>6</sub> H <sub>9</sub> N <sub>3</sub> O <sub>3</sub>	171.16	443-48-1
81	monolinuron	C <sub>9</sub> H <sub>11</sub> ClN <sub>2</sub> O <sub>2</sub>	214.65	1746-81-2
82	monomethylphthalate	C <sub>9</sub> H <sub>8</sub> O <sub>4</sub>	180.16	4376-18-5
83	MTBE (methyl tert-butyl ether)	C <sub>5</sub> H <sub>12</sub> O	88.15	1634-04-4
84	NAC (standard) carbaryl (1-naphthyl-N-methylcarbamate)	C <sub>12</sub> H <sub>11</sub> NO <sub>2</sub>	201.23	63-25-2
85	N-acetyl-L-tyrosine	C <sub>11</sub> H <sub>13</sub> NO <sub>4</sub>	223.23	537-55-3
86	naproxen	C <sub>14</sub> H <sub>14</sub> O <sub>3</sub>	230.27	22204-53-1
87	NDMA (N-nitrosodimethylamine)	C <sub>2</sub> H <sub>6</sub> N <sub>2</sub> O	74.08	62-75-9
88	NDPA (diphenylnitrosamine)	C <sub>12</sub> H <sub>10</sub> N <sub>2</sub> O	198.23	86-30-6
89	nonylphenol	C <sub>15</sub> H <sub>24</sub> O	220.36	25154-52-3
90	oxybenzone	C <sub>14</sub> H <sub>12</sub> O <sub>3</sub>	228.08	131-57-7
91	pentoxifylline	C <sub>13</sub> H <sub>18</sub> N <sub>4</sub> O <sub>3</sub>	278.31	006493-05-6
92	perchloroethene	C <sub>2</sub> Cl <sub>4</sub>	165.83	127-18-4
93	phenacetin	C <sub>10</sub> H <sub>13</sub> NO <sub>2</sub>	179.22	62-44-2
94	phenazon	C <sub>11</sub> H <sub>12</sub> N <sub>2</sub> O	188.23	60-80-0
95	pindolol	C <sub>14</sub> H <sub>20</sub> N <sub>2</sub> O <sub>2</sub>	248.33	13523-86-9
96	primidone	C <sub>12</sub> H <sub>14</sub> N <sub>2</sub> O <sub>2</sub>	218.26	125-33-7
97	procaine	C <sub>13</sub> H <sub>20</sub> N <sub>2</sub> O <sub>2</sub>	236.31	59-46-1
98	progesterone	C <sub>21</sub> H <sub>30</sub> O <sub>2</sub>	314.46	57-83-0
99	prometryn	C <sub>10</sub> H <sub>19</sub> N <sub>3</sub> S	241.36	7287-19-6
100	propranolol	C <sub>16</sub> H <sub>21</sub> NO <sub>2</sub>	259.34	525-66-6
101	propylparaben	C <sub>21</sub> H <sub>30</sub> O <sub>2</sub>	314.46	57-83-0
102	salbutamol	C <sub>13</sub> H <sub>21</sub> NO <sub>3</sub>	239.32	18559-94-9
103	simazin	C <sub>7</sub> H <sub>12</sub> ClN <sub>5</sub>	201.66	122-34-9
104	sotalol	C <sub>12</sub> H <sub>20</sub> N <sub>2</sub> O <sub>3</sub> S	272.37	3930-20-9
105	sucrose	C <sub>12</sub> H <sub>22</sub> O <sub>11</sub>	342.30	57-50-1
106	sulfachloropyridazine	C <sub>10</sub> H <sub>9</sub> ClN <sub>4</sub> O <sub>2</sub> S	284.73	80-32-0
107	sulfamerazine	C <sub>11</sub> H <sub>12</sub> N <sub>4</sub> O <sub>2</sub> S	264.3	127-79-7
108	sulfamethizole	C <sub>9</sub> H <sub>10</sub> N <sub>4</sub> O <sub>2</sub> S <sub>2</sub>	270.33	144-82-1
109	sulfamethoxazole	C <sub>10</sub> H <sub>11</sub> N <sub>3</sub> O <sub>3</sub> S	253.28	723-46-6
110	terbutaline	C <sub>12</sub> H <sub>19</sub> NO <sub>3</sub>	225.29	23031-25-6
111	tetraethylenepentamine	C <sub>8</sub> H <sub>23</sub> N <sub>5</sub>	189.3	112-57-2

#	Name	Molecular formula	$M_w/g\ mol^{-1}$	CAS Number #
112	trichloroacetic acid	$C_2HCl_3O_2$	163.39	76-03-9
113	trichloroethene	$C_2HCl_3$	131.39	79-01-6
114	triclosan	$C_{12}H_7Cl_3O_2$	287.95	3380-34-5
115	trimethoprim	$C_{14}H_{18}N_4O_3$	290.32	738-70-5
116	urea	$CH_4N_2O$	60.06	57-13-6

ANNEX B. Weights and bias of the optimised QSAR-NN<sub>1</sub> model

Input-hidden layer connections												Hidden layer – output connections		
Weights												Bias	Weights	Bias
$w'_{j,1}$	$w'_{j,2}$	$w'_{j,3}$	$w'_{j,4}$	$w'_{j,5}$	$w'_{j,6}$	$w'_{j,7}$	$w'_{j,8}$	$w'_{j,9}$	$w'_{j,10}$	$w'_{j,11}$	$w'_{j,12}$	$b_j^H$	$w_{i,j}^H$	$b_i^0$
7.2543	9.6510	-1.1509	-0.5229	-0.5781	-2.3239	-5.1358	1.9045	-1.9102	3.3482	1.0421	-0.6314	0.7091	-0.2936	
-6.5995	0.9372	-0.4893	-4.8986	-0.2104	-4.1948	0.2977	1.1115	-0.8228	-1.8299	-0.5010	3.5292	4.4415	1.1671	
2.2851	-0.9396	-1.2465	-2.7195	-3.1700	0.0667	-5.2542	0.2258	-2.6779	3.9133	0.2007	-2.6003	2.9282	0.6458	
3.4212	-4.4453	-4.5290	2.6873	-0.5741	6.2931	13.4276	-3.8687	-2.2673	4.2925	-2.1050	-2.4805	-1.5072	-1.7295	
3.5162	-4.0280	1.2348	-1.5229	0.2798	7.3793	-5.8274	-1.1807	18.5372	8.6676	0.0518	2.4391	2.6164	5.9830	
-3.2315	-0.0412	1.5080	4.4308	-0.3078	-5.1318	7.2885	0.3269	-15.4471	2.4560	2.4688	3.6898	5.3947	-1.8723	
-1.8967	-5.2757	8.7229	3.8204	1.0474	4.7165	3.1082	4.3083	5.5364	4.0459	-1.3650	1.5793	-0.3415	0.2906	
-3.1544	-1.6553	5.6860	-1.7981	-0.7112	3.5766	-5.2059	-0.0090	16.2916	4.7436	-1.2570	-2.2510	-2.1667	-0.9304	
-4.2949	4.5187	-3.5941	-0.7832	3.3981	1.0363	-2.0857	-0.1512	2.6421	-1.9945	1.4527	1.6570	-1.4498	0.5376	
-1.5734	3.8419	3.7600	-3.3667	-0.5008	2.0405	-0.5350	5.6774	-4.1561	3.5703	-1.6669	-1.2820	-1.5209	0.1783	
-0.8150	-8.2746	4.8433	-2.4298	1.1923	4.1521	1.2509	-4.0717	-4.8640	-0.5552	-1.9774	2.2937	-1.4779	2.2032	
3.3108	1.8365	-12.7519	-2.1037	-0.1508	-5.2124	-4.4505	-1.4271	0.9729	-2.3982	-0.0287	4.7968	2.2196	-0.5093	
2.2983	-2.1892	-0.5345	-1.9585	2.0102	-3.0123	0.024	-0.1343	1.5658	4.4758	-1.7246	1.3355	2.5823	-1.4197	-1.4376
-6.6487	0.4504	2.9469	6.9690	-0.7828	2.3599	6.2060	-2.9179	-4.4266	3.1603	-0.8510	1.2432	0.9139	-2.6608	
2.8514	-0.2683	2.4687	0.9783	3.0966	1.8261	7.1206	2.8130	-3.9001	-5.4748	0.6965	-11.5741	-1.4881	-0.6670	
6.4656	2.7255	0.3597	-2.1250	-3.5123	-2.5432	6.8401	2.4607	-2.7771	2.7860	-0.1308	-1.2049	-1.3966	0.3363	
0.8676	-2.7205	-0.6273	-1.8534	-1.5976	3.6903	-3.0195	0.2212	3.3234	-4.1697	-0.1142	-1.7504	3.0841	-1.5187	
-0.2444	-2.9867	0.4978	1.7975	1.3171	-1.8520	2.4446	0.8166	-1.3038	-0.1150	-1.5505	12.6289	2.3071	0.5562	
-1.0021	0.7352	5.0196	1.0473	-5.1052	-4.8710	-0.0237	-1.1215	0.8992	-0.1000	-0.9903	-4.7788	-1.6246	0.7929	
-0.1102	3.6438	-0.5097	1.6882	1.3007	2.6868	-0.6678	5.8430	0.0210	1.8481	-2.9467	1.6446	2.9073	-3.6755	
-3.7598	2.2326	-1.6593	0.2437	0.8194	-0.7571	-1.8180	7.5073	3.3581	-3.0696	-0.6818	1.3513	-0.0023	0.8457	
3.8736	2.0510	-1.2680	2.6267	-0.2528	-0.1513	0.5941	1.3325	-4.1390	-0.7053	-0.5612	4.0974	0.6258	0.1114	
-0.3657	1.3691	-0.7722	-1.3598	0.0542	2.1856	-1.3801	-1.6382	-8.6009	3.5202	0.1744	-3.9514	-1.6425	1.2202	
-0.8189	0.0523	-0.9155	-0.3628	1.8324	0.8170	-2.1191	-1.8835	-0.2996	0.4072	-0.4494	-0.4818	1.0514	2.4631	
-4.1804	-0.2979	1.1133	2.2501	1.1804	-0.6994	1.4683	1.1702	5.9891	1.8899	-2.7033	-1.2098	3.0810	1.4745	

Weights and bias of the optimised QSAR-NN<sub>2</sub> model

Input-hidden layer connections												Hidden layer – output connections		
Weights												Bias	Weights	Bias
$w'_{j,1}$	$w'_{j,2}$	$w'_{j,3}$	$w'_{j,4}$	$w'_{j,5}$	$w'_{j,6}$	$w'_{j,7}$	$w'_{j,8}$	$w'_{j,9}$	$w'_{j,10}$	$w'_{j,11}$	$w'_{j,12}$	$b_j^H$	$w_{i,j}^H$	$b_i^0$
1.0276	0.2913	0.1141	-1.3298	3.8821	1.1528	-0.9293	0.8573	-1.4048	-2.0065	-0.3955	1.1566	0.1485	-0.6935	
-2.4368	0.3397	0.0791	1.1424	4.9953	0.1884	0.9219	-0.2742	1.9750	0.2266	0.1285	-0.6974	-1.1737	1.3253	
1.1539	-0.4112	2.9210	1.2905	-2.4450	0.7094	0.4124	0.8246	0.9047	0.5492	-0.6756	0.4843	-0.1478	0.2531	
0.5759	1.0119	-0.6287	0.3849	2.2353	-0.0030	0.2477	1.9484	2.2233	-0.1833	-0.7349	-4.3617	-0.6342	0.5375	
0.6512	-0.0862	-0.1147	0.6349	0.9008	-0.6737	0.5368	-0.5513	-1.0648	-0.2085	1.3585	-0.0687	1.4942	-0.4888	
0.1597	0.5851	-2.9995	-0.0764	0.2079	-0.8582	-1.3071	1.1880	-0.0174	-0.4312	-0.4836	0.1704	-2.6063	0.4198	
-0.9765	0.2169	2.2693	0.5513	0.2022	1.5001	-0.2959	-2.4265	-0.6991	0.3910	1.6911	0.3070	0.2871	0.1457	
-0.2410	-0.3252	-1.8598	-0.3677	0.1234	-2.5025	-1.4022	-0.1521	0.8134	0.1126	-2.6637	-0.4669	-2.4816	1.4298	
-0.5857	-0.2305	-0.1666	-0.1853	0.4360	0.3197	-1.9207	0.2580	5.8024	-0.3919	-0.2349	1.1660	-0.7967	0.0738	
-0.7060	-0.0200	-0.1932	-0.5368	0.8133	-1.1797	0.5348	0.4683	0.8783	-1.2372	0.1939	-1.8746	-0.5750	-0.4137	
0.2179	0.1756	-0.3219	0.2748	0.3966	-0.9596	0.4062	0.6318	0.6170	0.7764	0.4754	1.5558	0.7906	-0.6913	
1.3783	0.2801	0.2287	-0.0712	0.1579	0.7008	0.4095	0.2564	-1.4748	0.5697	0.4289	-0.0576	0.8198	0.8173	0.46911
-0.7400	-0.9527	-0.1205	0.4997	-0.5487	-0.3947	1.7889	-2.0579	-2.3960	-0.7351	0.7526	-0.1836	-1.2491	0.5410	
2.2611	-0.3632	-0.0426	0.8739	0.0710	0.0389	4.0457	0.7007	-0.0678	-0.4055	0.9944	-0.7565	-0.3110	-0.6217	
-4.1073	1.1132	2.5261	-0.6984	0.5087	0.6321	-0.8560	0.3006	0.5312	0.9444	1.5728	0.9620	0.5131	-0.8833	
-1.3122	0.0797	-0.1931	-0.9787	-0.2008	-0.6517	0.3189	1.9572	0.6820	-0.4264	0.1956	-0.0149	-2.0063	-0.7632	
-1.6292	0.2920	0.5366	-0.9281	-3.9048	0.0172	-0.0701	-2.9349	0.0539	1.4814	-0.3577	-0.5332	0.8862	0.6560	
0.0716	-0.5625	-0.1783	-0.5926	3.4220	-1.2733	0.4067	1.4425	-1.2804	2.4695	-0.3562	0.3333	0.6639	-0.5061	
1.0179	0.3289	0.0117	0.6153	1.8531	-0.8524	-1.7223	0.0241	-1.6524	0.4046	0.8923	1.5450	-0.9593	0.3486	
-0.3010	-0.8579	0.0350	3.1753	-2.5622	0.9244	0.5416	0.2890	-0.1150	-0.0546	-1.0558	-1.2148	0.2070	0.4896	
-0.5319	1.9492	0.4186	0.5489	-2.1636	-0.9798	0.4933	-0.2267	0.4882	-3.0500	-0.3559	1.7955	-0.2303	-0.9640	
0.5186	-0.5216	-0.4101	-0.9363	0.4508	0.5588	-0.1101	-0.9494	0.4633	1.2330	-1.5887	0.0577	-0.3946	-1.3540	
0.0240	-1.1579	-0.4503	-0.6708	0.4096	-0.8691	-0.3171	0.7185	0.5704	-0.1554	-0.2157	-0.1018	0.0180	-0.5364	

Weights and bias of the optimised QSAR-NN<sub>3</sub> model

Input-hidden layer connections												Hidden layer – output connections		
Weights												Bias	Weights	Bias
$w'_{j,1}$	$w'_{j,2}$	$w'_{j,3}$	$w'_{j,4}$	$w'_{j,5}$	$w'_{j,6}$	$w'_{j,7}$	$w'_{j,8}$	$w'_{j,9}$	$w'_{j,10}$	$w'_{j,11}$	$w'_{j,12}$	$b_j^H$	$w_{i,j}^H$	$b_1^0$
0.1448	0.1578	1.9750	0.4483	1.6285	-1.2787	-0.2686	0.4935	0.1809	-0.5208	0.3453	-0.8274	-0.1221	0.0464	0.53805
0.1228	0.4795	-0.6204	0.3354	-0.5533	1.2689	0.4305	-0.8396	0.2279	0.8751	-0.1106	-1.4491	-0.1321	-0.1390	
-0.1180	0.2851	0.0385	-0.6574	-0.3492	0.9271	0.7121	-0.4769	0.0981	0.2075	1.7279	-1.3591	0.3129	-0.0633	
0.1793	0.1565	-0.5301	1.7168	0.0504	0.0338	-0.2141	-1.3306	0.2964	0.7789	0.4665	0.9241	0.9490	0.2319	
0.0382	0.1031	-0.1081	0.5137	0.8054	1.8265	0.3080	-0.5879	-1.8438	-0.5235	-0.1964	-1.0728	-0.7796	0.3540	
-0.6889	-0.1985	-0.5820	0.4941	0.1807	-7.0927	0.4860	-0.5010	-1.5558	-0.7827	0.4366	0.6420	-0.7803	0.1816	
-0.1957	0.2915	0.5058	-0.0401	0.0853	-0.5961	-0.8871	0.9960	-0.8511	1.0239	0.3939	-0.2971	-0.4469	-0.5354	
-0.4302	-0.5385	1.4104	-0.5526	-0.3615	0.1707	1.2238	1.3598	-0.5235	-0.4995	1.3465	0.8840	-2.2442	0.9123	
0.6249	-0.2645	0.0782	0.5875	0.3671	-0.0712	1.5470	-0.0031	-1.7095	-0.6461	-0.1239	1.0709	1.2034	-0.9369	
0.0306	-0.5211	0.7136	0.1989	2.7820	-0.2269	1.3788	0.0847	0.2210	-2.5850	0.8794	-0.5286	-0.0304	-0.0439	
0.2038	0.8360	1.9488	0.4410	-3.0801	-3.6899	3.9191	-0.0409	-0.9584	-0.1352	0.0688	1.1743	-1.3213	0.6031	
0.3258	1.1175	0.4003	-1.0335	-2.5214	1.9446	-3.7659	-0.2070	-0.4314	-0.3351	-0.8232	0.1524	-0.2033	0.2495	
0.0173	0.0210	0.5928	0.0735	-0.6484	-0.8134	0.2781	-0.4745	-0.7071	-0.6538	0.0676	0.3848	-0.8763	-0.2656	
-0.3286	-0.6484	-0.4071	1.1221	2.231	-0.0474	3.2888	-0.1270	-0.9831	-1.4872	-0.2098	0.1014	2.7020	1.4251	
1.0146	-0.1890	0.7943	-0.1390	0.0370	0.1037	3.2905	0.4522	-2.7759	1.3361	-0.8124	2.1464	-0.5208	0.0107	
0.0261	0.0726	1.4528	-1.9859	-0.6039	-1.3232	-0.9582	-0.3947	-1.1498	0.0902	0.5011	-0.1169	1.2519	0.1397	
0.4182	0.2585	0.6159	1.3488	-0.7792	0.2131	0.2963	-0.5371	-0.4225	-1.7779	1.5339	-2.0049	1.0956	0.3106	
0.4739	1.0353	-0.7093	0.1830	-2.6293	-0.0851	0.3781	-0.1677	0.4158	0.1060	-0.2043	-0.6044	-0.8645	0.4418	
-0.7744	-0.8086	1.0409	-2.5323	-0.9557	-0.0332	-0.5665	-0.2739	0.7469	1.3376	-1.4922	-0.0718	0.7492	0.4206	