https://doi.org/10.15255/KUI.2019.010
KUI-39/2019
Original scientific paper
Received March 31, 2019
Accepted June 10, 2019

Comparative Study of Predicting the Molecular Diffusion Coefficient for Polar and Non-polar Binary Gas Using Neural Networks and Multiple Linear Regressions

N. Melzi,[a] *L. Khaouane*,[a] *Y. Ammi*,[a,b*] *S. Hanini*,[a] *M. Laidi*,[a] and *H. Zentou*[c]

[a] Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Médéa, 26 000 Médéa, Algeria
[b] Department of Chemical Engineering, University Center Ahmed Zabana Relizane, 48 000 Relizane, Algeria
[c] Department of Chemical and Environmental Engineering, Universiti Putra Malaysia, 43 400 Serdang, Malaysia

Abstract

In the current study, an *artificial neural network* (ANN) and *multiple linear regressions* (MLR) have been used to develop predictive models for the estimation of molecular diffusion coefficients of 1252 polar and non-polar binary gases at multiple pressures over a wide range of temperatures and substances. The quality and reliability of each method were estimated in terms of the *correlation coefficient* (*R*), *mean squared errors* (MSE), *root mean squared error* (RMSE), and in terms of *external validation coefficients* (Q^2_{ext}).

The comparison between the artificial neural network (ANN) and the multiple linear regressions (MLR) revealed that the neural network models showed a good predicting ability with lower errors (the roots of the mean squared errors in the total database were 0.1400 for ANN1 and 0.1300 for ANN2), and (root mean squared errors in the total databases were 0.5172 for MLR1 and 0.5000 for MLR2).

Keywords

Prediction, molecular diffusion, neural networks, multiple linear regressions

1 Introduction

Mass transfer is a term generally used to describe the transport of a substance (mass) in liquid and gaseous media. It commonly appears in different industrial applications, including the chemical reactions and petroleum industry, as well as ecological natural processes.[1] The mass transfer process mainly includes two aspects: molecular diffusion and convective mass transfer. Diffusion is a form of mass transfer by which molecules, ions, or other small particles spontaneously mix, moving from regions of relatively high concentration into regions of lower concentration. The knowledge of diffusion coefficient is essential to describe several processes and chemical reactions.[1–4] Diffusion is normally a slow process, and is a rate-determining factor in many cases of mass transfer. The hydrodynamic theory of diffusion mass transfer is well developed. However, there is no rigorous theory for estimation of the constituting parameters in this theory, the diffusion coefficients.[1] Motivated by the development in theoretical concepts such as the molecular kinetic theory, several studies have suggested various theoretical and semi-empirical models for the estimation of diffusion coefficients of binary gas mixture, such as Stefan-Maxwell (SM), Chapman–Enskog,[5] Gilliland,[6] Arnold,[7] Hirschfelder–Bird–Spotz (HBS),[8] Chen and Othmer,[9] Fuller–Schettler–Giddings,[10] Huang *et al.*[11] However, it is found that these empirical or semi-empirical ground models had a limited predictive capacity due to the lack of experimental data needed for the validation of the models.

More models have been developed based on HBS equation to estimate gases diffusion coefficients accurately using gas dynamic properties, but these models are frequently proposed for infrequent systems (high temperatures, big molecules, *etc.*), which does not reflect the real range of Gilliland and Arnold equations. The equation proposed by Fuller–Schettler–Giddings (FSG)[10] may represent a more practical and reliable concept, which shows more precision and accuracy. On the other hand, *artificial neural networks* (ANN) models can be considered a powerful tool for simulation, identification, design and control of processes. They have been widely used in a diverse range of fields, such as chemical engineering, biology, medical sciences, agriculture, and other applications. Therefore, ANN seems to be an effective and promising tool in developing models for predicting the molecular diffusion coefficients accurately. In parallel, *multiple linear regressions* (MLR) have been highly applied in process engineering recently, and shown good accuracy.

Zaefizadeh compared the artificial neural network (ANN) with multiple linear regressions (MLR) in predicting performance, and revealed that the neural network method worked better than linear regression.[12] In addition, *Soares* studied the applicability of artificial neural networks and

* Corresponding author: Dr Yamina Ammi
Email: ammi.yamina@yahoo.fr

the analysis of multiple linear regressions of major components. Their results also indicated that the use of ANN might offer more accuracy compared to the MLR.[13]

Although having a high accuracy with the simplicity of implantation, there is a lack of studies investigating the application of MLR and ANN for diffusion coefficient estimation. *Eslamloueyan and Khademi*[14] have developed a neural network-based model for prediction of binary gas diffusivity, at atmospheric pressure as a function of temperature. A set of 336 experimental data points for the binary diffusion coefficients of both organic and inorganic gases has been used for the network training stage, where the developed ANN model showed lower relative error (4.47 %) and good performance compared to the other models reported in the literature. However, further studies are needed to develop universal models capable of estimating diffusion coefficients of both polar and non-polar gases at a large range of conditions. In this context, the presented study has used artificial neural networks and multiple linear regressions for predicting the molecular diffusion coefficient for 311 polar and 941 non-polar binary gases over a wide range of temperatures and pressures with high accuracy.

2 Materials and methods

2.1 Empirical equations of the molecular diffusion coefficient

The mass diffusivity D_{AB} for a binary system is a function of temperature, pressure, and composition. The available data on D_{AB} for most binary mixtures are, moreover, quite limited in range and accuracy.[15] Low pressure of binary gas mixtures, D_{AB}, increases with increasing temperature, inversely proportional to the pressure.[16] These disparities are all depicted, with diverse degrees of accuracy, by the following empirical equations of the kinetic theory of gases, which are employed for the prediction of the D_{AB} values. Some of these correlations are presented in Table 1.

2.2 Neural networks

Artificial neural networks (ANNs) have become a popular tool in the analysis of non-linear relationships between process variables and product characteristics due to the simplicity of application and high accuracy. Several studies in the literature have recently used ANN models for

Table 1 – Empirical models for diffusion coefficients estimation

Equation	Equation formula		Details
Stefan–Maxwell (SM) equation[14]	$D_{AB} = \frac{a}{n\sigma_{AB}^2}\left[\frac{8RT}{\pi}\left(\frac{1}{M_A}+\frac{1}{M_B}\right)\right]^{\frac{1}{2}}$	(1)	D_{AB}: diffusion coefficient P: gas pressure / bar T: absolute temperature / K σ_{AB}: collision diameter between A and B (A)°. Ω_D: collision integral. M_A, M_B: molecular masses of gas A and gas B, respectively / g mol^{-1} V_A and V_B: molar volumes at boiling point / cm^3 mol^{-1} k: Boltzmann constant T_C, V_C: critical temperature and critical volume, respectively
Chapman–Enskog equation[16]	$D_{AB} = \frac{0.002667T^{\frac{3}{2}}}{PM_{AB}^{\frac{1}{2}}\sigma_{AB}^2\Omega_D}$ $\sigma_{AB} = (\sigma_A+\sigma_B)/2$ $M_{AB} = 2[(1/M_A)+(1/M_B)]^{-1}$	(2)	
Wilke and Lee[17]	$D_{AB} = \frac{\left[3.03-\left(\frac{0.98}{M_{AB}^{\frac{1}{2}}}\right)\right](10^{-3})T^{\frac{3}{2}}}{PM_{AB}^{\frac{1}{2}}\sigma_{AB}^2\Omega_D}$	(3)	
Gilliland equation[7]	$D_{AB} = \frac{0.0043T^{1.5}(1/M_A+1/M_B)^{1/2}}{P\left[(V)_A^{\frac{1}{3}}+(V)_B^{\frac{1}{3}}\right]^2}$	(4)	
Hirschfelder–Bird–Spotz (HBS)[8]	$D_{AB} = \frac{0.00816T^{1.5}(1/M_A+1/M_B)^{1/2}}{P\sigma_{AB}^2\Omega_D}$	(5)	
Chen and Othmer[9]	$D_{AB} = \frac{0.43(\frac{T}{100})^{1.81}(1/M_A+1/M_B)^{1/2}}{[(V_{CA}/100)0.4+(V_{CB}/100)0.4]^2P\left(\frac{T_{CB}T_{CA}}{10}^4\right)^{0.1405}}$	(6)	
Huang *et al.*[11]	$D_{AB} = \frac{5.06T^{1.75}(1/M_A+1/M_B)^{1/2}}{P^{1.286}\left[V_A^{\frac{1}{3}}+V_B^{\frac{1}{3}}\right]^2}$	(7)	

the prediction of physicochemical properties. The majority of these models were multi-layered feed-forward non-linear ANN, trained *via* the back-propagation rule to satisfy a function approximation[19] where non-linear *feed-forward neural networks* (FFNN) have shown efficacy for universal functional approximation.

A multi-layered feed-forward neural network is composed of three layers: the input layer, the output layer, and a hidden layer, as illustrated in Fig. 1. For the formal description of the neuron, the so-called mapping function *r* can be used, which assigns for each neuron *i* a subset μ_i. The subset μ_i consists of all predecessors of the given neuron *i*. Each neuron in a particular layer is connected with all neurons in the next layer. The connection between the i^{th} and j^{th} neuron is characterised by the weight coefficient w_{ij} (Fig. 1). The weight coefficient reflects the degree of importance of the given connection in the neural network.[19]

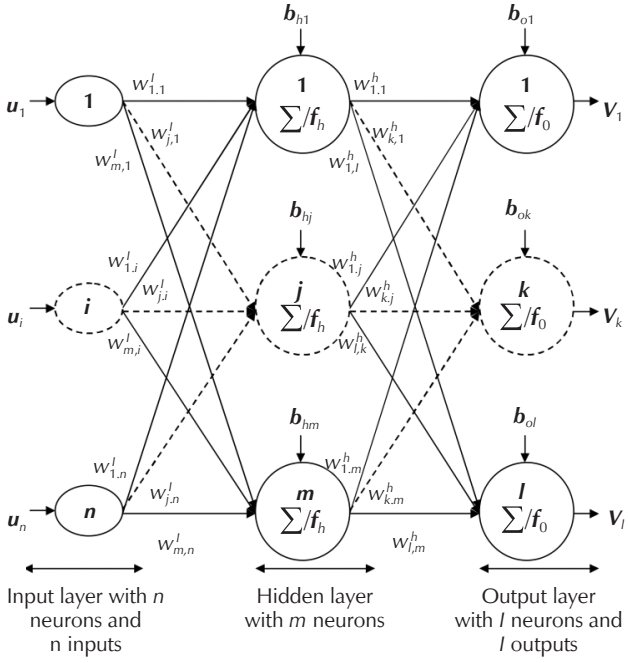

Fig. 1 – Three-layer feed-forward neural network

2.3 Multiple linear regressions

Multiple linear regressions (MLR) is a statistical technique used to study a linear relationship between a dependent variable Y_i and several independent variables X_i.[13,20] A multiple regression model can be represented by the following equation:

$$Y_i = A_0 + \sum_{i=1}^{n} A_i X_i \qquad (8)$$

where Y_i represents the output, X_i represents the inputs, A_i their coefficient, A_0 is a constant (called the intercept), and *n* is the number of inputs.[21]

3 Modelling procedure

3.1 Database collection

In the present research, a database "DB" accessible in the literature has been used.[11,14,15,18,19,20-33] The database collected from literature includes 1252 experimental data composed of 219 binary systems and 117 gases. The training database collected from the literature can be accessed in supplementary data.

The selection of the type and number of inputs was conducted after a comprehensive review of the models proposed in the literature.[14,22,23,29] Two distinct databases were used (DB1 with the critical volume) and (DB2 with the critical pressure) in order to develop two different neural networks and two multiple linear regression models. Table 2 shows the inputs and outputs used for each database: "DB1 for developed NN1 and MLR1", and "DB2 for developed NN2 and MLR2".

Table 2 – Inputs and output used for each database "DB1 and DB2"

	DB1 for NN1	DB2 for NN2
Inputs	T P $T_C^{(A)}$ $T_C^{(B)}$ $V_C^{(A)}$ $V_C^{(B)}$ M (A) M (B) μ(A) μ(B)	T P $T_C^{(A)}$ $T_C^{(B)}$ $P_C^{(A)}$ $P_C^{(B)}$ M (A) M (B) μ(A) μ(B)
Output	D_{AB}	D_{AB}

The statistical analyses for each database in terms of minimum, maximum, mean, and standard deviation are depicted in Table 3.

3.2 Modelling with neural networks

Details of the different steps conducted to develop the models presented in this study are shown in Fig. 2. In this work, the samples were split randomly into three subsets: seventy percent (70 %) of the data was used to train the model, 15 % for validation, and 15 % in the test case for both databases (DB1 and DB2).

The training algorithm used in this work was the BFGS quasi-Newton (***trainbfg***). Each NN contained three layers of neurons or nodes: one input layer with ten neurons of NN1 and neurons for NN2 in the input layer, one hidden layer with a number of active neurons optimized during training, and one output layer with one unit that generated the estimated value of diffusion coefficient D_{AB}. The number of hidden neurons differed 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.

The application of ANN for diffusion coefficient modelling was performed by STATISTICA software, which was used to optimize the architecture of the two neural networks models (ANN1 and ANN2).

Table 3 – Statistical analysis of parameters for each database (DB1 and DB2)

Parameters	Symbol	Unit	Minimum	Maximum	Mean	STD
temperature	T	K	77.2	1200	387.34	143.84
pressure	P	bar	0.78	449.88	9.73	32.23
critical temperature	$T_C^{(A)}$	K	5.19	761	238.49	245.97
critical temperature	$T_C^{(B)}$	K	5.19	819	219.1	186.64
critical volume	$V_C^{(A)}$	$cm^3\,mol^{-1}$	21.23	492	136.7	119.15
critical volume	$V_C^{(B)}$	$cm^3\,mol^{-1}$	41.7	492.1	123.83	84.93
molecular weight	M (A)	$g\,mol^{-1}$	4	235.9	41.15	43.99
molecular weight	M (B)	$g\,mol^{-1}$	2.01	253.8	40.91	44.29
dipole moment	μ(A)	D	0	4.22	0.25	0.7
dipole moment	μ(B)	D	0	4.22	0.24	0.62
critical pressure	$P_C^{(A)}$	bar	2.33	223.34	35.62	36.36
critical pressure	$P_C^{(B)}$	bar	2.33	223.34	37.85	26.08
diffusion coefficient between two gases A and B	D_{AB}	$cm^2\,s^{-1}$	$7.6000 \cdot 10^{-6}$	8.1	0.58	0.8

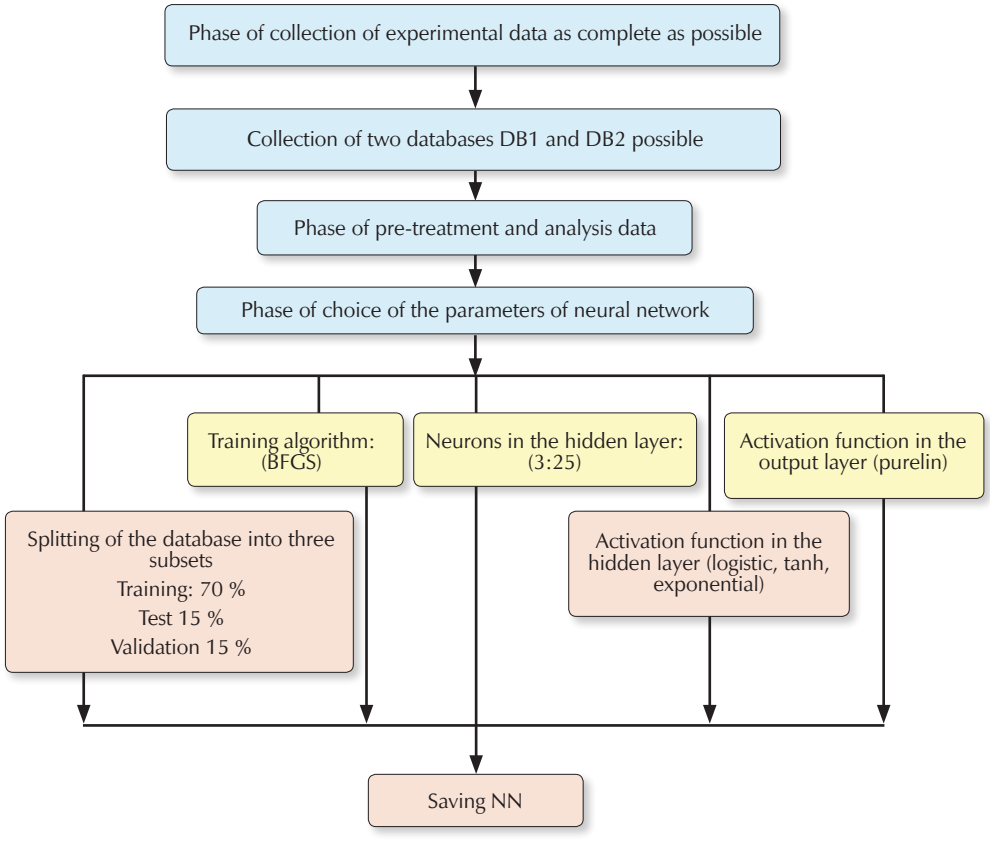

Fig. 2 – Flowchart of the neural networks models development procedure

Table 4 – Multi-layer feed-forward artificial neural networks with the best architectures

NN models	Input layer variables	Hidden layer		Output layer		Training method
		transfer function	number of neurons	transfer function	variable	
NN1	M (A) / g mol^{-1} M (B) / g mol^{-1} $V_C^{(A)}$ / cm^3 mol^{-1} $V_C^{(B)}$ / cm^3 mol^{-1} $T_C^{(A)}$ / K $T_C^{(B)}$ / K μ(A) / D μ / D T / K P (atm)	Hyperbolic tangent Sigmoid (tansig)	8	linear transfer function (purelin)	diffusion coefficient (D_{AB}) / cm^2 s^{-1}	BFGS: quasi-Newton back- propagation (***trainbfg***)
NN2	M (A) / g mol^{-1} M (B) / g mol^{-1} $P_C^{(A)}$ / cm^3 mol^{-1} $P_C^{(B)}$ / cm^3 mol^{-1} $T_C^{(A)}$ / K $T_C^{(B)}$ / K μ(A) / D μ(B) / D T / K P / bar	Hyperbolic tangent Sigmoid (tansig)	6	linear transfer function (purelin)	diffusion coefficient (D_{AB}) / cm^2 s^{-1}	BFGS:quasi-Newton back- propagation (***trainbfg***)

Table 4 shows the inputs, outputs, number of neurons, training methods, and adaptation learning functions. It depicts the formation of the architectures of NN models.

3.3 Modelling with multiple linear regressions

The same databases mentioned for ANN development were employed to develop MLR by using the "regress" MATLAB function.

4. Results and discussion

4.1 Neural networks

The plot and the specification of the linear regression were acquired directly with the use of MATLAB function "postreg".

Fig. 3(a) shows the curves of the linear regression of the diffusion coefficient rates calculated by the first neural model "ANN1" optimized with the rates of experimental diffusion coefficient with vectors of regression approaching the ideal [α, β, R] = [0.968, 0.03, 0.98]. The correlations coefficients for the total phase are generally considered as excellent in the range ($0.90 \le R \le 1.00$). The correlation coefficient of the developed ANN1 was $R = 0.968$ with interval of confidence IC95% = 0.005, which indicates good robustness of the first neural model "ANN1" established and the prediction possibility for the different parameters of the diffusion coefficient of gases.

To estimate the performance of the second neural model obtained "ANN2", the predicted values of the diffusion coefficient by the ANN2 model were compared and plotted to the experimental values of diffusion coefficient for the total phase, as shown in Fig. 3(b), which represents the curves of the linear regression of the second neural model with vectors of regression approaching the ideal [α, β, R] = [0.97, 0.01, 0.984]. The correlation coefficient of the ANN2 model was excellent ($R = 0.9845$) with an interval of confidence IC95% = 0.0482, which implies good robustness of the second neural model "ANN2" established and the possibility of predicting the diffusion coefficients of gases from the different parameters.

4.2 Multiple linear regressions

The models obtained for the prediction of the diffusion coefficient are linear models (Eqs. (9) and (10)):

$$\begin{aligned}\mathrm{MLR1} = {} & -0.229 + 0.0038T + 4.05\cdot10^{-4}P - \\ & - 8.179\cdot10^{-4}T_C^{(A)} - 0.002T_C^{(B)} - 5.44\cdot10^{-4}V_C^{(A)} + 7.8\cdot10^{-5}V_C^{(B)} \\ & -8.49\cdot10^{4}M(\mathrm{A}) + 5.51\cdot10^{4}M(\mathrm{B}) + 0.03\mu(\mathrm{A}) + 0.117\mu(\mathrm{B})\end{aligned} \quad (9)$$

$$\begin{aligned}\mathrm{MLR2} = {} & 0.214 + 0.003T + 7.484\cdot10^{-4}P - 9.1555\cdot10^{-4}T_C^{(A)} - \\ & - 0.0014T_C^{(B)} - 2.456\cdot10^{-4}P_C^{(A)} - 0.001P_C^{(B)} - \\ & -0.001M(\mathrm{A}) + 2.655\cdot10^{-4}M(\mathrm{B}) + 0.0405\mu(\mathrm{A}) + 0.106\mu(\mathrm{B})\end{aligned} \quad (10)$$

Fig. 4(a) shows the linear regression curves of model MLR1 by comparison between the experimental values and the calculated values for the first MLR model. Linear regression vectors, [α, β, R] = [0.76, 0.26, 0.58] of the total phase indicated a moderate MLR1 model predictive ability. The correlation coefficient of MLR1 was 0.7645 with an interval of confidence IC95% = 0.0717. The correlation coefficients are generally considered satisfactory ($0.50 \le R < 0.90$).[32]

The obtained results show that the MLR1 model had an acceptable predicting power to estimate the diffusion coefficient of gases.

Fig. 4(b) shows the linear regression curves of the diffusion coefficient rates calculated by the MLR2 model with the experimental diffusion coefficient rates with regression vectors approaching approval [α, β, R] = [0.5004, 37.4867, 0.7038] of the total phase. The correlation coefficient of MLR2 was 0.7511 in the satisfactory range ($0.50 \leq R < 0.90$); [34] with an interval of confidence equal to IC95% = 0.0717. Similarly to MLR1, the results indicated that the MLR2 model showed a good performance for the prediction of the diffusion coefficient of gases.

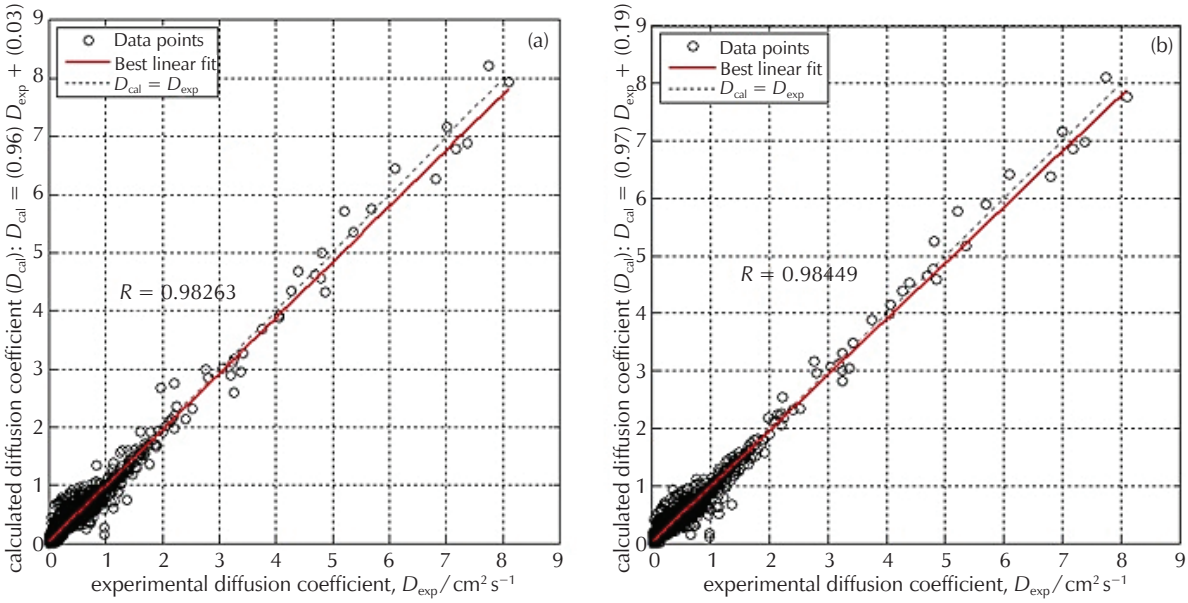

Fig. 3 – Comparison between experimental and calculated values for the whole data set: (a) NN1, (b) NN2

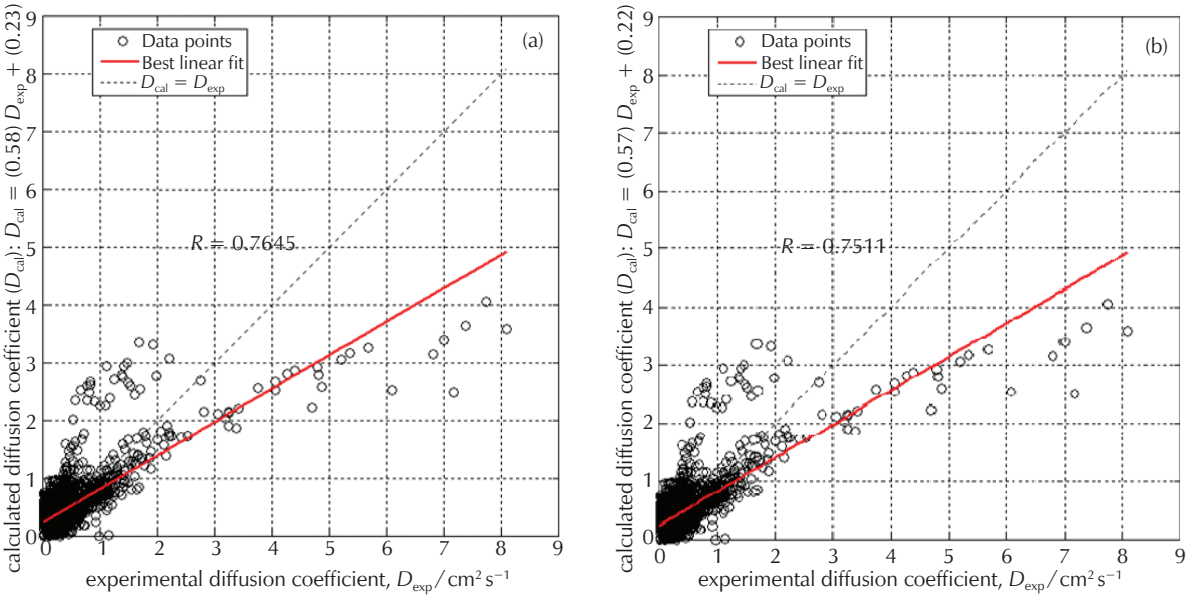

Fig. 4 – Comparison between experimental and calculated values for the whole data set: (a) MLR1, (b) MLR2

5. Comparison

5.1 Comparison between Neural networks and multiple linear regressions

The comparison between the multiple linear regressions (MLR) and the artificial neural network (ANN) models was estimated in terms of root mean squared error (RMSE), mean squared errors (MSE), external validation coefficient (Q^2_{ext}), and in terms of the correlation coefficient (R), for the entire data set.

- mean squared errors (MSE)

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(Y_{i,exp} - Y_{i,pred})^2 \quad (11)$$

- root mean squared error (RMSE)

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n}(Y_{i,exp} - Y_{i,pred})^2}}{n} \quad (12)$$

- external validation coefficient (Q^2_{ext})

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{pred}(Y_{i,exp} - Y_{i,pred})^2}{\sum_{i=1}^{pred}(Y_{i,exp} - Y_{TR})^2} \quad (13)$$

where N is the total number of data, $Y_{i,exp}$ is the experimental value, $Y_{i,cal}$ indicates the calculated value from the neural network models or multiple linear regressions, respectively, and Y_{TR} is the mean experimental diffusion coefficient value of the samples for the training set.

As shown in Fig. 5, the ANN models developed in this work gave lower errors than the MLR models. The mean squared errors for the total dataset were 0.02 for ANN1, 0.01 for ANN2, root mean squared errors were 0.14 for ANN1, 0.13 for ANN2, and the external validation coefficient were 0.50 for ANN1, 0.49 for ANN2. The mean squared errors for the total data set were 0.26 for MLR1 and 0.25 for MLR2, the root mean squared errors were 0.5172 for MLR1 and 0.5 for MLR2 and the external validation coefficients were 0.58 for MLR1 and 0.59 for MLR2. The results of this comparison showed superiority of the ANN models.

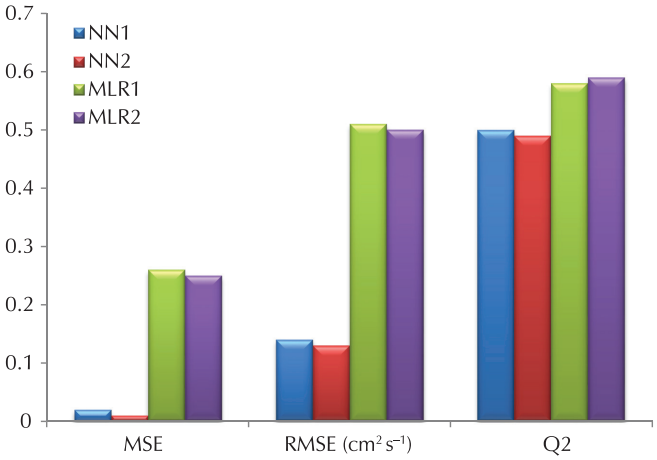

Fig. 5 – Comparison between neural networks and multiple linear regressions models

5.2 Comparison between neural networks and empirical models

To compare the neural network models developed in this work with other models, the diffusivity of numerous elements in the data set needed to be estimated using the developed networks models, as well as seven different correlations, which Chapman–Enskog, Gilliland proposed: Arnold, Hirschfelder–Bird–Spotz (HBS), Chen and Othmer, Fuller–Schettler–Giddings (FSG), and Huang.

In order to estimate the true predictive power of neural network models, one must compare the predicted and observed (experimental) activities of a fairly large set of external test compounds that have not been used in model development.[35] It is the accuracy of the performance on this test set that determines the true predictive power of

Table 5 – Comparison of the optimized ANN models with other models in the literature

Models	α	β	R	MSE	RMSE / cm² s⁻¹	Q^2
NN1	0.968	0.03	0.98	0.02	0.14	0.50
NN2	0.97	0.01	0.984	0.01	0.13	0.49
Wilke and Lee (1955)	0.712	0.101	0.95	0.09	0.30	0.89
Chapman–Enskog (1977)	0.734	0.101	0.95	0.07	0.282	0.9
Gilliland (1934)	0.4558	0.112	0.95	0.24	0.496	0.70
HBS (1949)	0.427	0.117	0.74	0.352	0.59	0.58
Chen and Othmer (1962)	0.001	$9.14 \cdot 10^{-4}$	0.74	$3.27 \cdot 10^{4}$	181.01	0.95

a neural network model. The predictive power of such a model can be estimated by the coefficient (Q^2_{ext}).[36]

Table 5 shows a comparison between the proposed ANN and various models reported in the literature for D_{AB} for prediction in terms of the mean squared errors (MSE), root mean squared error (RMSE), external validation coefficient (Q_2) and correlation coefficient (R), which were calculated using the training data. Table 5 clearly shows good performance of ANN developed models. Both ANN models show higher correlation coefficient (R), and their lower mean squared errors (MSE), lower root mean squared error (RMSE), and lower external validation coefficient (Q_2) compared to the previous models reported in the literature. This confirms the strength and the accuracy of ANN models for the prediction of diffusion coefficient.

6. Conclusions

The aim of the present study was to develop two feed-forward artificial neural networks (ANN) models and two multiple linear regression (MLR) models in order to predict the molecular diffusion coefficient of the binary gas mixture. A good agreement was observed between the experimental values and the predicted values for each neural model (the roots of the mean squared errors in the total database were 0.1400 for ANN1 and 0.1300 for ANN2). On the other hand, the developed MLR model for diffusion coefficient prediction showed acceptable agreement with experimental values (root mean squared errors in the total databases were 0.5172 for MLR1 and 0.5000 for MLR2), but it was less accurate with lower relation coefficient and higher errors compared to the ANN models. Moreover, the ANN models developed in this work showed lower errors compared to the other models reported previously in the literature.

Thus, the good predicting power and the high accuracy of ANN models in estimating the coefficient of diffusion of binary gas mixture is clear, which makes these models highly recommended over the other models for coefficient diffusion prediction in different chemical processes.

ACKNOWLEDGEMENTS

The authors acknowledge the team of LBMPT laboratory, University of Medea, the University Center Ahmed Zabana Relizane, and University Putra Malaysia for their help throughout this project.

List of abbreviations
Popis kratica

ANN – artificial neural network
– umjetna neuronska mreža

FFNN – feed-forward neural network
– aciklička neuronska mreža

FSG – Fuller–Schettler–Giddings

HBS – Hirschfelder–Bird–Spotz

MLR – multiple linear regressions
– model višestruke linearne regresije

MSE – mean squared error
– srednja kvadratna pogreška

R – correlation coefficient
– korelacijski koeficijent

RMSE – root mean squared error
– korijen srednje kvadratne pogreške

SM – Stefan-Maxwell

References
Literatura

1. *O. Medvedev*, Diffusion coefficients in multicomponent mixtures, 2004, Ph. D. thesis, Technical University of Denmark, Copenhagen, Denmark.
2. *L. Boardman, N. Wild*, The diffusion of pairs of gases with molecules of equal mass, Proc. R. Soc. Lond. A **162** (1937) 511–520, doi: https://doi.org/10.1098/rspa.1937.0199.
3. *J. Bosma, J. Wesselingh*, Estimation of diffusion coefficients in dilute liquid mixtures, Chem. Eng. Res. Des. **4** (1999) 325-328. doi: https://doi.org/ 10.1205/026387699526250.
4. *T. R. Marrero, E. A. Mason*, Gaseous diffusion coefficients, J. Phys. Chem. Ref. Data. **1** (1972) 3–118, doi: https://doi.org/10.1063/1.3253094.
5. *J. Giddings*, Dynamics of Chromatography, Marcel Decker, New York, 1965, NY, Part I, p. 340, doi: https://doi.org/10.1201/9781315275871.
6. *E. R. Gilliland*, Diffusion coefficients in gaseous systems, Ind. Eng. Chem. **26** (1934) 681–685, doi: https://doi.org/10.1021/ie50294a020.
7. *J. H. Arnold*, Studies in diffusion: Studies in diffusion. II. A kinetic theory of diffusion in liquid systems, J. Am. Chem. Soc. **52** (1930) 3937–3955, doi: https://doi.org/10.1021/ja01373a025.
8. *J. C. Hirschfelder, R. B. Bird, E. L. Spotz*, The transport properties of gases and gaseous mixtures, J. Chem. Phys. **44** (1930) 205–231, doi: https://doi.org/10.1021/cr60137a012.
9. *D. F. Othmer, H. T. Chen*, Correlating diffusion coefficients in binary gas systems, Ind. Eng. Chem. Process. Des. Dev. **1**

(1962) 249–254, doi: https://doi.org/10.1021/i260004a003.

10. *E. N. Fuller, P. D. Schettler, J. C. Giddings*, A new method for prediction of binary gas phase diffusion coefficient, Ind. Eng. Chem. **58** (1966) 19–27, doi: https://doi.org/0.1021/ie50677a007.
11. *T. C. Huang, J. F Young, C. J. Huang, C. H. Kuo*, Measurement of diffusion coefficients by the method of the gas chromatography, J. Chromatogr. A **70** (1972) 13–24, doi: https://doi.org/10.1016/s0021-9673(01)91051-9.
12. *M. Zaefizadeh*, Comparison of Multiple Linear Regressions (MLR) and Artificial Neural Network (ANN) in Predicting the Yield Using its Components in the Hulless Barley American-Eurasian, J. Agric. Environ. Sci. **33** (2011) 1828–1837.
13. *T. Santos, D. Mendes, R. Torres*, Artificial neural networks and multiple linear regression model using principal components to estimate rainfall over South America, Nonlin. Processes Geophys. **23** (2016) 13–20, doi: https://doi.org/10.1016/s0021-9673(01)91051-9.
14. *R. Eslamloueyan, M. Khademi*, A neural network-based method for estimation of binary gas diffusivity, Chemom. Intell. Lab. Syst. **104** (2010) 195–204, doi: https://doi.org/10.1016/j.chemolab.2010.08.009.
15. *D. Gavril, K. R. Atta, G. Karaiskakis*, Determination of collision cross-sectional parameters from experimentally measured gas diffusion coefficients, Fluid. Phase. Equilib. **218** (2004) 177–188, doi: https://doi.org/10.1016/j.fluid.2003.12.010.
16. *S. Chapman, T. G. Cowling*, The Mathematical Theory of Non-Uniform Gases, 3rd Ed. Cambridge Math, 1970, doi: https://doi.org/10.1002/zamm.19920721124.
17. *C. Wilke, C. Lee*, Estimation of diffusion coefficients for gases and vapors, Ind. Eng. Chem. **47** (1955) 1253–125747, doi: https://doi.org/10.1021/ie50546a056.
18. *J. C. Giddings, S. L. Seager*, Rapid determination of gaseous diffusion coefficients by means of gas chromatography apparatus, J. Chem. Phys. **33** (1960) 1579–1580, doi: https://doi.org/10.1063/1.1731448.
19. *C. Si-Moussa, S. Hanini*, Prediction of high-pressure vapor liquid equilibrium of six binary systems carbon dioxide with six esters using an artificial neural network model, J. Chem. Eng. **25** (2008) 183–199, doi: https://doi.org/10.1590/s0104-66322008000100019.
20. *N. Tebal, P. Pinang*, Comparison Between Multiple Linear Regression And Feed forward Back propagation Neural Network Models For Predicting PM 10 Concentration Level Based On Gaseous And Meteorological Parameters, IJAST. **1** (2011) 42–49.
21. *S. I. V. Sousa, F. G. Martins, M. C. Pereira*, Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, Environ. Model. Soft. **22** (2007) 97–103, doi: https://doi.org/10.1016/j.envsoft.2005.12.002.
22. *B. E. Poling, J. M. Prausnitz, J. P. O'Connel*, The properties of gases and liquids. McGraw-Hill., New York, 2001, doi: https://doi.org/10.1021/ja0048634.
23. *G. Karaiskakis, D. Gavril*, Determination of diffusion coefficients by gas Chromatography, J. Chromatogr. A **1037** (2004) 147–189, doi: https://doi.org/10.1002/chin.200435278.
24. *J. Knox, L. McLaren, L. Anal*, The Spreading of Air Peaks in Capillary and Packed Gas Chromatography Columns, An. Ch. **35** (1963) 449–454, doi: https://doi.org/10.1021/ac60197a033.
25. *E. Grushka. V. Maynard*, Measurements of gaseous diffusion coefficients by gas chromatography, J. Chem. Educ. **49** (1972) 565, doi: https://doi.org/10.1021/ed049p565.
26. *J. Line, S. Weissman*, Determination of the temperature dependence of gaseous diffusion coefficients using gas chromatographic apparatus, J. Phys. Chem. **56** (1972) 2288–2F0, doi: https://doi.org/10.1063/1.1677533.
27. *S. Gotoh, M. Manner, J. P. Sorensen, E. Stewart, J. Warren*, Binary diffusion coefficients of low-density gases, Chem. Eng. Data. **19** (1974) 169–171. doi: https://doi.org/10.1021/je60061a026.
28. *G. Staker, P. J. Dunlop*, The pressure dependence of the mutual diffusion coefficients of binary mixtures of helium and six other gases at 300 K: tests of Thorne's equation, Chem. Phys. Lett. **42** (1976) 419–422, doi: https://doi.org/10.1016/0009-2614(76)80643-4.
29. *T. K. Sherwood, R. C. Reid, J. M. Prausnitz*, The properties of gases and liquids, New York, McGraw-Hill, 1966, p. 688, doi: https://doi.org/10.1002/aic.690150602.
30. *H. Worth, W. Nüsse. J. Piiper*, Determination of binary diffusion coefficients of various gas species used in respiratory physiology, Respir. Physiol. **32** (1978) 15–26, doi: https://doi.org/10.1016/0034-5687(78)90097-x.
31. *N. Katsanos, G. Karaiskakis*, Measurement of diffusion coefficients by reversed-flow gas chromatography instrumentation, Chromatogr. Sci. **237** (1982) 1–14, doi: https://doi.org/10.1016/S0021-9673(00)88267-9.
32. *L. Fokin, A. Kalashnikov, A. F. Zolotukhina*, Transport properties of mixtures of rarefied gases. Hydrogen–methane system, J. Eng. Phys. Thermophys **84** (2011) 1408–1420, doi: https://doi.org/10.1007/s10891-011-0612-7.
33. *W. Lu, G. Huirong, I. M. Chou, R. C. Burruss, L. Li*, Determination of diffusion coefficients of carbon dioxide in water between 268 and 473K in a high-pressure capillary optical cell with in situ Raman spectroscopic measurements, Geochim. Cosmochim. Acta **115** (2013) 183–204, doi: https://doi.org/10.1016/j.gca.2013.04.010.
34. *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** (2015) 2300–2310, doi: https://doi.org/10.1007/s11814-015-0086-y.
35. *A. Golbraikh*, *A. Tropsha*, Beware of q^2!, J. Mol. Graph. Model. **2** (2002) 269–276, doi: https://doi.org/10.1016/s1093-3263(01)00123-1.
36. *M. Hamadache*, Modélisation en vue de la prédiction ou de la corrélation de l'activité toxicologique à partir de la structure moléculaire, 2004, Ph. D. thesis, University of Medea, Algeria.

SAŽETAK

Usporedna studija predviđanja koeficijenta molekularne difuzije za polarni i nepolarni binarni plin pomoću neuronskih mreža i višestrukih linearnih regresija

Naima Melzi,[a] *Latifa Khaouane*,[a] *Yamina Ammi*,[a,b*] *Salah Hanini*,[a] *Maamar Laidi*[a] i *Hamid Zentou*[c]

U ovoj studiji primijenjene su *umjetna neuronska mreža* (ANN) i *model višestruke linearne regresije* (MLR) za razvoj prediktivnih modela za procjenu koeficijenata molekularne difuzije 1252 polarnih i nepolarnih binarnih plinova pri višestrukim tlakovima u širokom rasponu temperatura i tvari. Kvaliteta i pouzdanost svake metode procijenjeni su pomoću *korelacijskog koeficijenta* (*R*), *srednjih kvadratnih pogrešaka* (MSE), *korijena srednje kvadratne pogreške* (RMSE) te *koeficijenata vanjske validacije* (Q^2_{ext}).

Usporedba između umjetne neuronske mreže (ANN) i višestrukih linearnih regresija (MLR) otkrila je da modeli neuronske mreže pokazuju dobru sposobnost predviđanja s nižim pogreškama (korijeni srednjih kvadratnih pogrešaka u ukupnoj bazi podataka bili su 0,1400 za ANN1 i 0 (1300 za ANN2 a pogreške korijena srednje vrijednosti u ukupnim bazama podataka bile su 0,5172 za MLR1 i 0,5000 za MLR2).

Ključne riječi
Predviđanje, molekularna difuzija, neuronske mreže, višestruke linearne regresije

[a] *Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Médéa, 26 000 Médéa, Alžir*
[b] *Department of Chemical Engineering, University Center Ahmed Zabana Relizane, 48 000 Relizane, Alžir*
[c] *Department of Chemical and Environmental Engineering, Universiti Putra Malaysia, 43 400 Serdang, Malezija*

Izvorni znanstveni rad
Prispjelo 31. ožujka 2019.
Prihvaćeno 10. lipnja 2019.