

## EVALUATING THE ABILITY OF R FOR MODELING A COMMERCIAL SCALE VGO HYDROCRACKING PLANT USING ARTIFICIAL NEURAL NETWORK (ANN)

Sepehr Sadighi\*, Seyed Reza Seif Mohaddecy

Research Institute of Petroleum Industry (RIPI), Catalysis Technologies Development Division, P.O. Box 14665137, Tehran, Iran

Received January 21, 2018; Accepted April 23, 2018

---

### Abstract

This research is devoted to evaluating the ability of R software to model an industrial scale vacuum gas oil (VGO) hydrocracking plant using ANN (artificial neural network) modeling approach. R is an open source program for statistical computation and graphics that consists a language plus a run-time environment, and access to certain system functions. The input layer of ANN model developed for VGO hydrocracking plant consists age of catalyst, flow rates of fresh VGO and recycle streams and inlet temperatures of catalytic beds. Moreover, neurons in the output layer of ANN are yield of products including light gases (Gas), liquefied petroleum gas (LPG), light naphtha (LN), heavy naphtha (HN), kerosene (Ker) and diesel (Dis). Networks are trained and tested based on 68 data points gathered from the target commercial scale plant using 'neuralnet' package obtained from the CRAN repository. From results, it is concluded that R can predict product yields of all unseen data with high accuracy; therefore, it is an alternative for modeling chemical engineering processes without requiring any license or charges for an activation code.

**Keywords:** Modeling; Artificial Neural Network; Vacuum gas oil; Hydrocracking; R software.

---

## 1. Introduction

Process modeling is a computed based procedure that generally comprises using software to determine a system of interconnected elements which should be solved simultaneously so that the steady-state or dynamic nature of the system can be explained. These models are strictly demanded in favor of effective design, perfect control and predict product yields and qualities versus variables such as space velocity and temperature [1].

Within the field of chemical process modeling and simulation, the use of artificial neural networks (ANN) has evolved for a diverse range of engineering applications such as fitting experimental data, machine diagnostics, pattern recognition, quality control, signal processing, and process control, all topics of interest to chemists and chemical engineers [2]. ANNs represent the massive parallel interconnections of a simple neuron that functions as a collection system, and they are designed in an attempt to mimic the human brain in order to emulate human performance, i.e., to function intelligently [3]. ANNs have the advantage that they employ non-linear basis functions, and therefore are able to approximate far more complex non-linear behavior. Furthermore, they are fast and have very small requirements in terms of storage space and retrieval time, as all that needs to be stored is the coefficients of the approximation model [4]. In this technique, based on information compiled from the process and without any needs for identifying the physical and chemical characteristics of feed and products, desired outputs such as yield or conversion can be simulated [5].

ANNs have been extensively applied for modeling various chemical engineering processes [6-19]. In most of these studies, Neural Network (NN) Toolbox of Matlab software (Mathworks, Inc.) has been used to create, train, test, validate and visualize neural networks. The validity, reliability, robustness, and resiliency of NN Toolbox have been reported previously in many

studies, and there is no doubt about it. However, MathWorks products are regulated to use in support of academic and instructive activities, and the right to utilize these products for commercial purposes is absolutely prohibited. Hence, having a substitution for this powerful software, especially when acquiring the license for Matlab (international sanctions or financial problem for purchasing the license), is so beneficial, encouraging and promising.

As an option, R is free and open source software, authorizing anyone to apply and modify it. R can be seen as a dialect of S language (developed at AT&T) for which John Chambers was awarded the 1998 Association for Computing Machinery (ACM) software [20]. This software is licensed under the GNU General Public License, with copyright held by The R Foundation for Statistical Computing. It has about 5000 packages accessible from assorted repositories that study intensively in topics such as an artificial neural network (ANN), the adaptive neuro-fuzzy system (ANFIS), stochastic boosting, data mining, nonlinear optimization, and bio-informatics.

To evaluate and examine the validity and reliability of ANN models implemented in R, the present study uses ANN modeling techniques to predict the yield of products in an industrial scale VGO hydrocracking plant. Based on the actual data obtained from the target plant, feed-forward structures are composed of 7, 3 and 1 neurons in the input, hidden and output layers, respectively. After training the constructed ANN model by accessible R packages, the capability of the developed network to predict yields of hydrocracking products is tested.

## 2. Process description

Table 1. Specifications of the fresh VGO in the target process [19]

Variable	Unit	Value
Fresh feed flow rate	m <sup>3</sup> /h	163.7
Density	kg/m <sup>3</sup>	911
Total sulfur	wt%	1.4
ASTM Distillation, D1160		
IBP	°C	333
10%	°C	389
30%	°C	414
50%	°C	434
70%	°C	453
90%	°C	481
FBP	°C	500

In the target hydrocracking process, the VGO feed (see Table 1) obtained from the vacuum distillation tower is mixed with unconverted oil (offtest), and then is mixed with hydrogen. This flow is passed through heat exchangers and heaters (Figure 1), and it is introduced to three parallel reactors which each of them have four catalytic fixed beds. They contain 15%, 25%, 30% and 30% of the total catalyst weight percentage, respectively. In the first bed, hydrotreating reactions are carried out to remove sulfur, nitrogen and metal compounds. The second, third and fourth catalytic beds often perform the hydrocracking reactions.

Along the hydrocracking reactor, quenching systems are provided for the purpose of controlling exothermic reactions (hydrotreating and hydrogenation). This facility injects hydrogen into the output stream leaving 1st, 2nd and 3rd beds to reduce its temperature, and therefore preserve the catalyst from excessive coke formation.

Furthermore, at the top of the reactor, ceramic balls are used to distribute the feed, and also prevent from the entrance of physical impurities into the reactor. The product, after passing through heat exchangers and air coolers, enters the high pressure separator (HPS). The hydrogen gas leaving HPS with the purity of 86.3 mol% is mixed with the make-up hydrogen (purity of 96.5 mol%), and then its major part is added to the feed of the reactor. The other fraction of this stream is injected between the catalytic beds to control the reaction temperature.

The effluent of HPS enters the low pressure separator (LPS) and is separated into two streams, i.e., sour gas and liquid product. The former enters the desulfurization unit, and the latter is introduced to the debutanizer tower. The liquid product leaving this tower enters the main fractionation column to obtain the hydrocracking products, i.e., light gas, LPG, light and heavy naphtha, kerosene, and diesel.

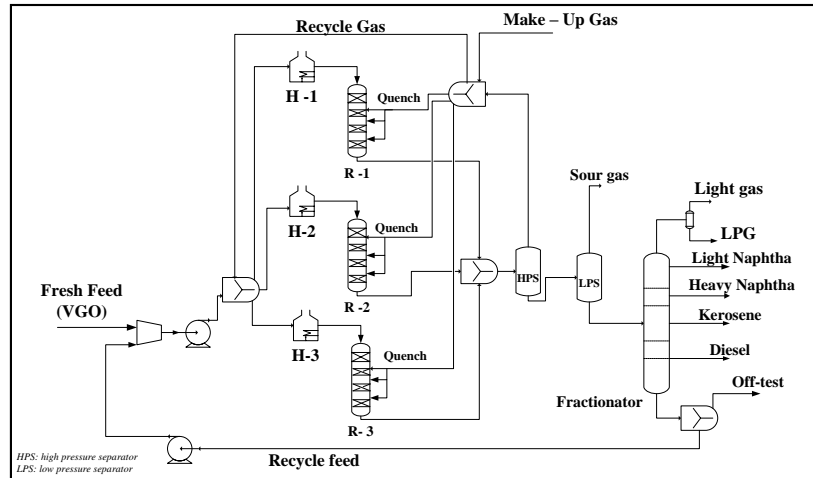


Figure 1. Block flow diagram (BFD) of the target VGO hydrocracking plant [19]

### 3. Developing the artificial neural network model in the R environment

R is a language and environment for statistical computing and graphics which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, and clustering) and graphical techniques, and is highly extensible. The advantage of using this software is its availability as free software under the terms of the Free Software Foundation’s GNU General Public License in source code form.

In this research, neural networks for predicting yields of VGO hydrocracking products including light gases (Gas), liquefied petroleum gases (LPG), light naphtha (LN), heavy naphtha (HN), kerosene (Ker) and diesel (Dis) are implemented in R software (Version 3.4.1, Copyright 2017). The package ‘neuralnet’ (Version 1.33) obtained from the CRAN repository is applied to develop and construct the corresponding ANN models. This package permits resilient settings through custom-choice of error and activation function, and also the calculation of generalized weights [21] is implemented. The globally convergent algorithm of ‘neuralnet’ is based on the resilient back-propagation without weight back-tracking and additionally modifies one learning rate, either the learning rate associated with the smallest absolute gradient [22].

The most widely employed networks have one hidden layer only. Each node within a given layer is connected to all of the nodes of the previous layer. The node sums up the weighted inputs and bias, and it passes the result through a linear function as follows [23]:

$$a_j = \sum_{i=0}^m w_{ji} y_i + b_j \quad (1)$$

where  $w_{ji}$  is the weight that goes from the input ( $i$ ) to the hidden neuron ( $j$ );  $b$  is the bias to the node, and  $y_i$  is the input unit of the neuron. By utilizing an activation function ( $f$ ), the output of the neuron can be written as follows:

$$Z_j = f(a_j) \quad (2)$$

By using ‘neuralnet’ package, a feed-forward ANN with one hidden layer is developed to simulate the target industrial scale VGO hydrocracking plant. The input layer of the ANN model consists age of catalyst (Age), volume flow rates of fresh VGO (Vflow) and recycle stream (Rflow) and temperature of the catalytic beds (i.e.,  $T_{b1}$ ,  $T_{b2}$ ,  $T_{b3}$ , and  $T_{b4}$ ). The output layer has one node which is the yield of the intended hydrocracking products. For each of them, an individual neural network is constructed. It should be mentioned that to prevent from memorizing instead of learning, only 3 neurons are selected for the hidden layer such that the number of coefficients is less than the number of actual data. To evaluate the accuracy

of the model, the absolute average deviation (AAD%) between the actual and predicted data is calculated as follows:

$$AAD\% = \frac{\sum_{n=1}^{N_t} \sqrt{\frac{(y_n^{actual} - y_n^{model})^2}{y_n^{actual^2}}}}{N_t} \times 100 \quad (3)$$

where  $y$  and  $N_t$  are output variables (i.e. yield of hydrocracking products) and number of data points, resp.; superscripts actual and model show the actual (or real) and the predicted values, resp.

#### 4. Results and discussions

During the period of data collection (about 557 days), 69 valid points are obtained which is randomly divided into two categories including training and testing datasets by using the 'sample' command of the R software. These groups consist 41 and 28 data points (60% and 40% of total data), respectively. All data points are scaled based on the maximum and minimum values of the actual data. In the next step, up to reaching the minimum relative mean squared error (RMSE), networks are trained by using 'neuralnet' package installed in the R environment. The number of iterations, AAD% of training and testing datasets for a yield of hydrocracking products, i.e. Gas, LPG, LN, HN, Ker, and Dis are presented in Table 2, respectively. From this table, it can be concluded that R can learn the pattern of hydrocracking yields satisfactory with an acceptable number of iterations (less than 200). The total CPU time for training each network is less than 1 sec which can be an advantage for developing ANNs using R software.

Table 2. Number of iterations and AAD% of training and testing data for VGO hydrocracking neural networks

Product	AAD%_train	Steps	AAD%_test
Gas	0.0672	192	0.0923
LPG	0.172	174	0.302
LN	0.175	183	0.366
HN	0.183	134	0.295
Ker	0.114	82	0.164
Dis	0.184	189	0.315

Structures (topology) of ANNs which are constructed by R for Gas, LPG, LN, HN, Ker, and Dis are illustrated in Figures 2 to 7, respectively. As seen, each network has 24 weights and 4 biases, totally 28 coefficients. As mentioned before, for training a network, there are 41 data points for the yield of each hydrocracking product. Hence, the degree of freedom (DOF) for each ANN is equal to 17. This positive DOF fairly prevents from memorizing (over learning) instead of training. However, increasing the number of data points enhances the reliability

of training data, and resiliency of the ANN model.

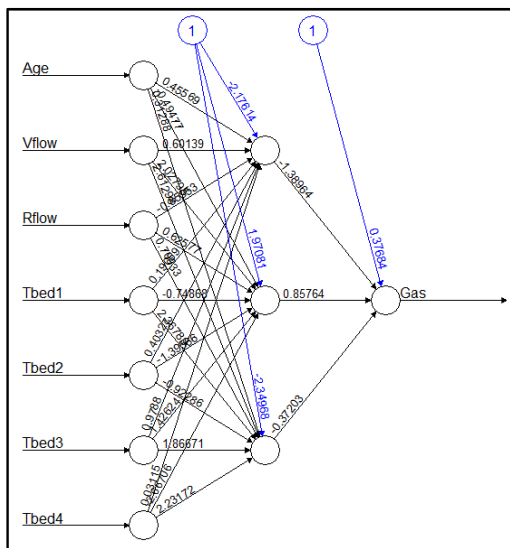


Figure 2. Structure of ANN constructed by R for predicting the yield of gas

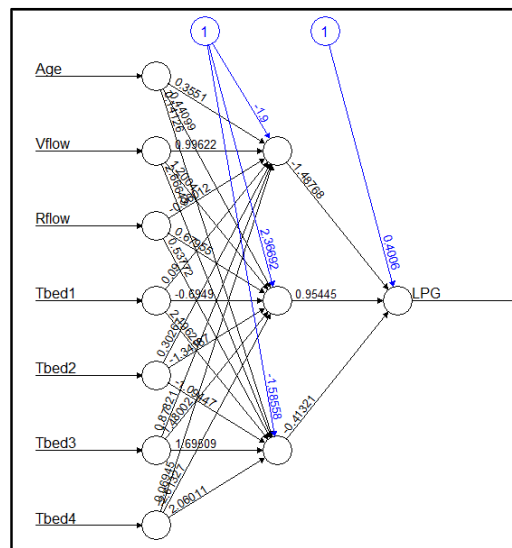


Figure 3. Structure of ANN constructed by R for predicting the yield of LPG

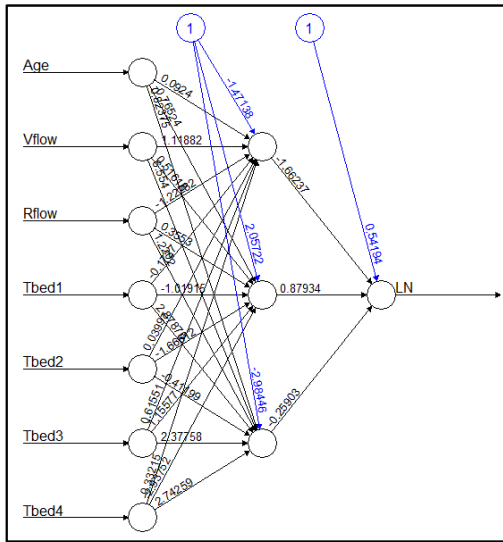


Figure 4. Structure of ANN constructed by R for predicting the yield of LN

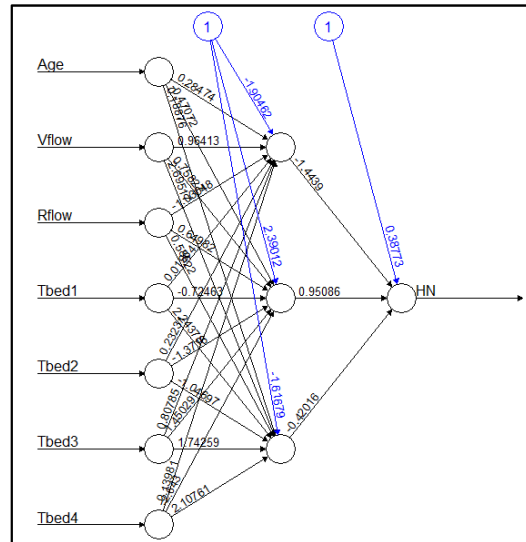


Figure 5. Structure of ANN constructed by R for predicting the yield of HN

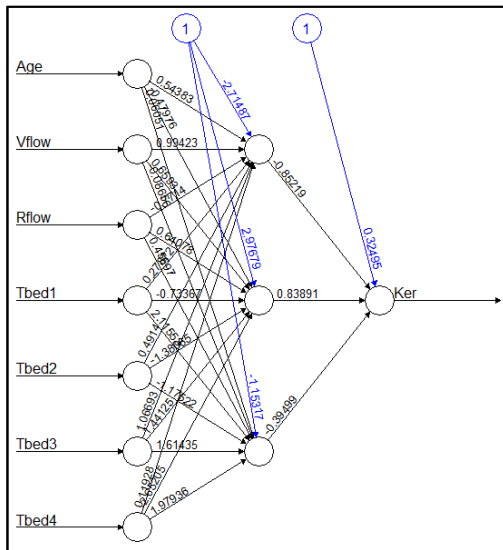


Figure 6. Structure of ANN constructed by R for predicting the yield of kerosene

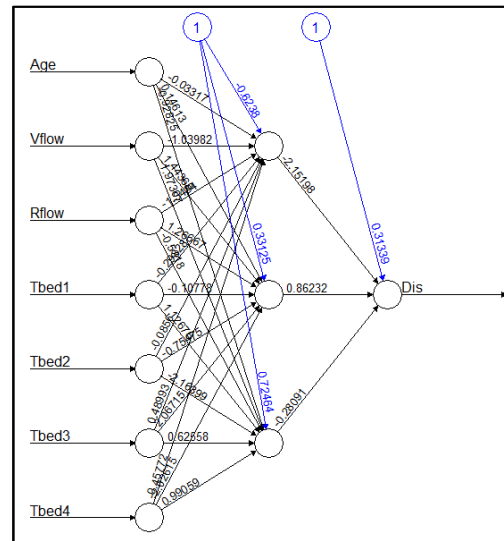


Figure 7. Structure of ANN constructed by R for predicting the yield of diesel

After training the ANNs, the unforeseen (predicting) data set is entered into the model, and the corresponding yield for each product is estimated and de-scaled using the 'compute' command in the R. As seen from Table 2, the AADs% of predicting data for all products are higher than the AAD% of testing data points, and also it is close to the AAD% of trained data. In order to further examine the reliability of the model, actual yields of Gas, LPG, LN, HN, Ker, and Dis are sketched versus the predicted values as parity plots in Figures 8 to 12, respectively. All these figures are generated by graphical tools provided in the R software. As seen, ANNs implemented in the R environment has high accuracy, and therefore they can be utilized for the purpose of sensitivity analysis and optimizing the target commercial scale VGO hydrocracking process or monitoring the activity of the catalyst. We suppose that the observed deviations can be due to random errors of measurement, and some abnormal conditions in the process, such as decreasing pressure of pumps and compressors, swinging in the quench flow rate or reaching the flood condition in the separation towers.

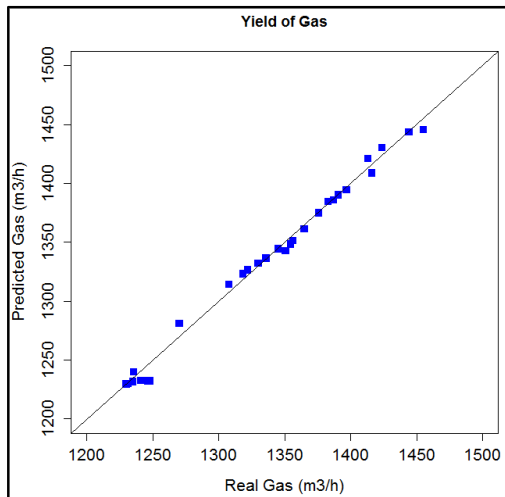


Figure 8. Comparison between actual and predicted values of gas yield predicted by R

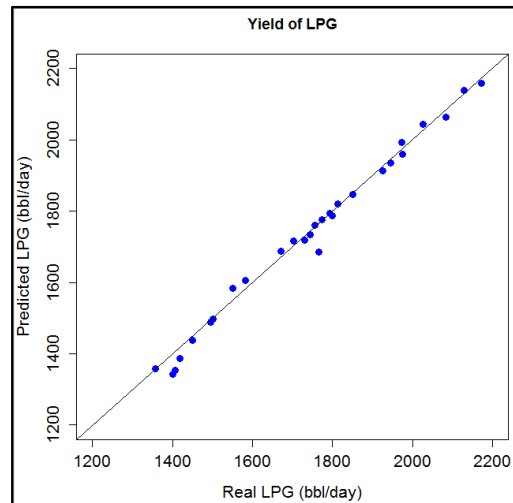


Figure 9. Comparison between actual and predicted values of LPG yield predicted by R

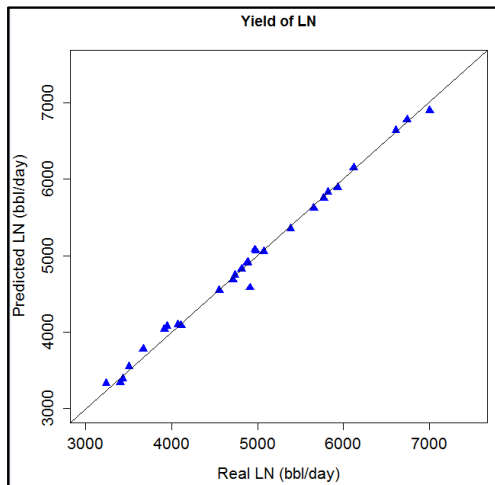


Figure 10. Comparison between actual and predicted values of LN yield predicted by R

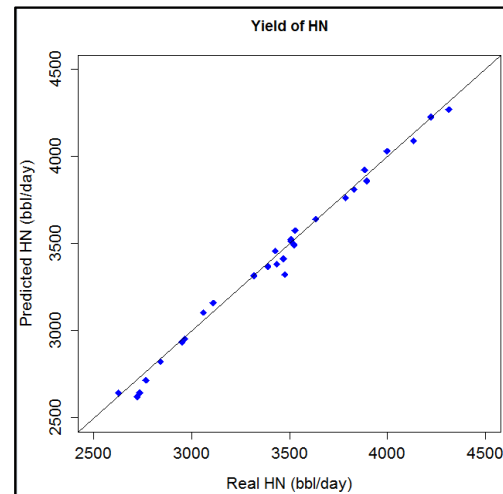


Figure 11. Comparison between actual and predicted values of HN yield predicted by R

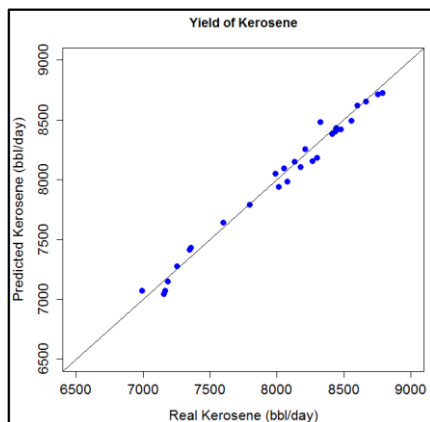


Figure 12. Comparison between actual and predicted values of kerosene yield predicted by R

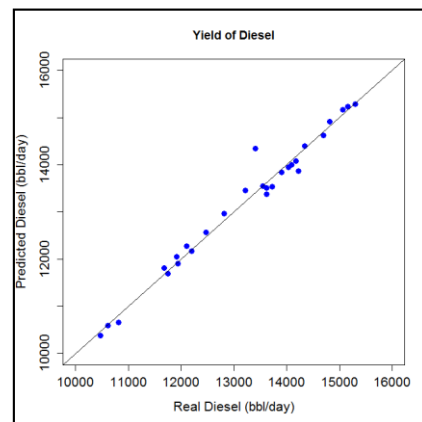


Figure 13. Comparison between actual and predicted values of diesel yield predicted by R

## 5. Conclusions

This research discussed the accuracy and reliability of R software and 'neuralnet' package for modeling yield of products in a commercial scale VGO hydrocracking plant using ANN technique. ANN models were trained and tested on the basis of 69 actual data points gathered from the target plant. For training, 60% of data points were randomly selected, and training procedure was carried out applying the back-propagation method. Other points were utilized to investigate predicting ability of ANN models constructed in R environment. These models were designed based on a feed-forward neural network with 7 neurons in the input layer and 3 neurons in the hidden layer. Input neurons were the age of catalyst, fresh and recycle feed flow rates and bed temperatures. The output neurons were yields of VGO hydrocracking products including light gases, LPG, light and heavy naphtha, kerosene and diesel. For each product, an individual network was implemented in the R. Results showed that R could predict the yield of products for testing points with the AAD% of 0.092%, 0.302%, 0.366%, 0.294%, 0.164% and 0.315%, respectively. The high accuracy of observed predictions can confirm the reliability of R to be used as free software with no license restrictions in modeling chemical engineering processes when ANN methodology can be executed as a modeling strategy.

## References

- [1] Sadighi S. J. Chem. Eng. Jap., 2016; 49(12): 979.
- [2] Himmelblau DM. Ind. Eng. Chem. Res., 2008; 47: 5782.
- [3] Kusumoputro B, Sutarya D, Faqih A. Int. J. Technol., 2016; 4: 709.
- [4] Franke LLC, Chatzopoulos K, Rigopoulos S. Combust. Flame, 2017; 185: 245.
- [5] Sadigh, S, Zahedi S, Hayati R, Bayat M. Ener. Technol., 2013; 1: 743.
- [6] Taskin H, Kubat C, Uygun O, Arslankaya S. Comput. Chem. Eng., 2006; 30: 850.
- [7] Bas D, Dudak FC, Boyaci IH. J. Food Eng., 2007; 79: 622.
- [8] Riverol, C.: Pilipovik, M. V., Chem. Eng. J., 2007;133: 133.
- [9] Bhutani N, Rangaiah GP, Ray AK. Ind. Eng. Chem. Res., 2006; 45: 7807.
- [10] Niaei A, Towfighi J, Khataee AR. Rostamizadeh, Petrol. Sci. Technol, 2007; 25: 967.
- [11] Istadi I, Amin NAS. Ind. Eng. Chem. Res. Catal., 2006; 45:6655.
- [12] Istadi I, Amin NAS. Bull. Chem. React. Eng. Catal., 2007; 2: 37.
- [13] Zahedi G, Mohammadzadeh S, Moradi M. Energy&Fuels, 2008; 22: 2671.
- [14] Alhajree I, Zahedi GR, Manan ZA, Mohammad Zadeh S. J. Pet. Sci. Eng., 2011; 78: 627.
- [15] Vasseghian Y, Ahmadi M. J. Chem. Pet. Eng., 2014; 48(2): 125.
- [16] Sadighi S, Mohaddecy RS. Int. J. Technol., 2013; 2: 102-111.
- [17] Sadighi S, Mohaddecy R, Norouzian A. Bull. Chem. React. Eng. Catal., 2015; 10: 210.
- [18] Sadighi S, Ghorbani Y. Petrol. Coal, 2017; 59(6): 777.
- [19] Sadighi S, Mohaddecy SR, Abbasi A. Int. J. Technol., 2018, under publishing.
- [20] Torgo L. Data Mining with R, 1<sup>st</sup> ed.; CRC Press, Taylor & Francis Group; London/New York, 2011.
- [21] Intrator O. Intrator, N.: Using Neural Nets for Interpretation of Nonlinear Models. Proceedings of the Statistical Computing Section, 1993, 244-249, San Francisco: American Statistical Society.
- [22] Fritsch S, Guenther F, Suling M, Mueller SM. Package 'neuralnet': Training of Neural Networks, 2016, <https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf>.
- [23] Haykin S, Hamilton O. Neural Networks. 2nd ed., Prentice Hall International, Upper Saddle River, NJ, 1998.

*To whom correspondence should be addressed: Dr. Sepehr Sadighi, Research Institute of Petroleum Industry (RIPI), Catalysis Technologies Development Division, P.O. Box 14665137, Tehran, Iran*