Petroleum and Coal

Data Modeling and Analysis of a Reflux Splitter Column (RSC) Using Artificial Neural Network Methodology (ANN) Implemented in R

Sepehr Sadighi[1]*, Davood Faraji[2]

[1] Research Institute of Petroleum Industry (RIPI), Catalysis Technologies Development Division, P.O. Box 14665137, Tehran, Iran
[2] Hydrocracking Unit (HCU) of Imam Khomeini Shazand Refinery, Arak, Iran

Abstract

This paper deals with developing an artificial neural network (ANN) model for the reflux splitter column (RSC) of a commercial vacuum gas oil hydrocracking unit. RSC separates light product, light naphtha, heavy naphtha, kerosene and diesel from the feed, and the residue is pumped back to the feed line of the hydrocracking reactors. The input-output data required for training and testing ANN is obtained from an Iranian refinery. The ANN model is implemented in R program which is an open source program and has access to certain system functions. The input layer of ANN model developed for RSC includes volume flow rate and temperature of rich oil, pressure of overhead, volume flow rate and temperature of super heat steam, flow rate and temperature of the reflux, flow rates of diesel, kerosene and heavy naphtha splitter vapors, temperature of fifth tray of RSC, and the value of pressure between trays 5 and 6. The networks are arranged to predict flow rates of LPG, light naphtha, heavy naphtha, kerosene, and diesel products. They are trained and cross validated based on 153 data points gathered from the target RSC unit by applying 'neuralnet' package obtained from the CRAN repository. The networks compared against the testing data, and it is shown that AADs% of unseen data for the heavy naphtha, kerosene, and diesel are 5.86%, 3.42% and 3.95%, respectively.

Keywords: Artificial Neural Network; Reflux Stripping Column; Distillation, R software.

1. Introduction

Mathematical modeling is well appreciated in the field of chemical engineering for the sake of designing, operating, controlling and optimizing the process plant. For any target process, a model should be developed to predict product yields and qualities versus variables such as pressure, flow rate and temperature [1]. Process models are computer-based programs that commonly comprise a system of interconnected elements which should be solved synchronously for disclosing the nature of the system.

Distillation is one of the most imperative process equipment in petroleum refining, and it is indispensable to be controlled close to the optimum operating conditions. The task of a distillation tower is to separate a feed consisting different compounds into two or more other streams, based on the difference in the volatility of those components. To develop a reliable, rigorous model for a distillation column working in a crude oil refinery, the complexity of the feed mixture makes it highly difficult to describe its components at a molecular level. For this reason, in refinery applications, any oil stream is typically characterized based on a distillation assay [2]. Existing commercial simulators like Aspen plus or Hysys from Aspen Technology and PetroSim from KBC do not have such limitations for the number of species, and it is possible to use a unique set of pseudo components for petroleum assay streams; but, this method increases the calculation time, and following reports become unnecessarily sophisticated.

Recently, soft computing methods mainly artificial neural networks (ANN) and fuzzy logic were widely applied for modeling, controlling and optimizing many chemical engineering processes [3-19]. ANN is an information processing paradigm that is inspired by the way of the biological nervous, and it is in an excursion to simulate the human brain for emulating human action, i.e., to function intelligently [20]. This modeling strategy has the following benefits: 1. it is highly nonlinear; thus, its architecture can be more rigorous and more representative than most other fundamental models; 2. its topology does not have to be pre-specified; therefore, they are completely resilient models [21].

However, studies focus on applying ANN for modeling a multi-cut distillation column in a crude oil refinery are scarce [22-24]. In order to discus on the idea of this research, using data points obtained from an industrial scale reflux splitter column (RSC) located in a VGO hydrocracking plant, a feed-forward ANN is developed in R program to predict the side products of the target distillation tower. These products are light petroleum gas (LPG), light naphtha (LN), heavy naphtha (HN), kerosene (Ker) and diesel (Dis). After training the ANN model by available tools in R program (i.e., nueralnet), the capability of developed networks to simulate the flow rate of RSC are validated against the unseen data.

2. Process description

In the Isomax process studied in this research, VGO hydrocracking is carried out at elevated temperature and pressure in a hydrogen atmosphere. The pressure at reactor inlet is about 180 barg, and the temperature at the inlet is started from 370°C to about 410°C at the end of the run (EOR) [25]. This single stage VGO hydrocracking process executes both hydrocracking and hydrotreating in a single catalytic reactor with three parallel trains. After the hydrocracking reactors, and high and low pressure separators, the feed to the fractionation section still have everything from hydrogen and hydrogen sulfide to the heaviest fractions, and it is essential to divide this stream into the favorite products such as liquefied petroleum gas (LPG), gasoline, kerosene, and diesel which are accepted by the market. The fractionation unit includes the following sections:

1. Sponge absorber tower
2. Recycle splitter column (RSC) with side cut strippers (Figure 1)
3. Diesel and kerosene strippers
4. Naphtha stripper
5. Gasoline stabilizer

The sponge absorber is designed to separate LPG fraction from the vapor of the splitter feed flash drum. In the RSC which is the target of this study, the light product and light naphtha are seized overhead, and the other fractions, i.e. heavy naphtha, kerosene, and diesel are taken as side cuts. The recycle splitter overhead then routed to the stabilizer column to recover light naphtha, LPG, and light gases.

The feed of the recycle splitter enters the lower portion of the column and is warmed up to the required flash zone temperature. This distillation column is a steam stripped column with several side cut draw offs. The overhead from the column is condensed and routed to the reflux accumulator. Heavy naphtha and kerosene as the recycle splitter side cut products are conducted to the corresponding strippers, and they are bared with vapor. The lowest side cut product is a diesel with a TBP range of about 260°C-360°C. This commodity is regularly a high quality, low pour point stream. Due to control its pour point, the end point of that must be adjusted meticulously to pass the required specifications. Finally, down the 370°C, it is removed from the bottom of the distillation tower and returned as the recycle feed to the reactor.

3. Developing the ANN for the RSC

R is a language and environment for statistical computing and graphics 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 modeling, 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. The ability and accuracy of R for modeling chemical engineering processes using its ANN package has been validated and reported in the previous work [26].
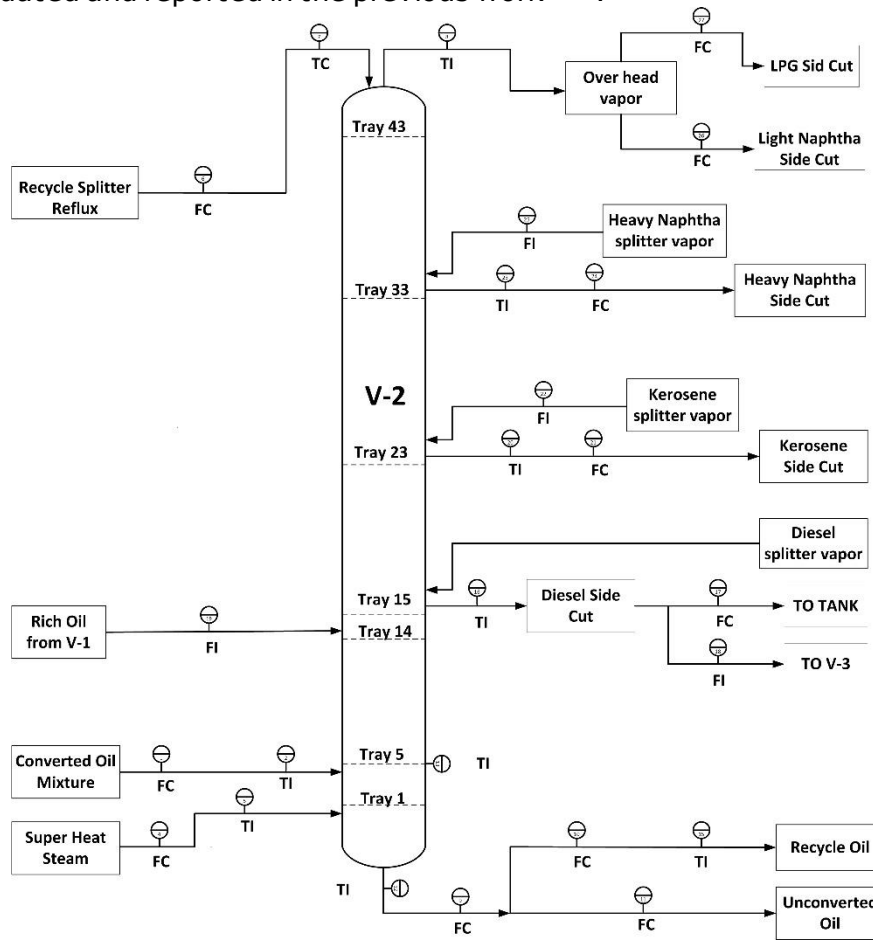

Figure 1. Block flow diagram (BFD) of the main fractionation tower of Isomax unit FC: flow controller; TI: temperature indicator

In this study, neural networks are developed in R software (Version 3.4.1, Copyright 2017) for predicting flow rates of output streams (i.e., LPG, LN, HN, Ker, and Dis) coming from an RSC hydrocracking plant. 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 [27] 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 [28].

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 [29]:

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

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 x_i is the input unit of the neuron. By utilizing an activation function (f), the output of the neuron can be written as follows:

$$Z_j = f(a_j) \tag{2}$$

By using 'neuralnet' package, a feed-forward ANN with one hidden layer is developed to simulate the target RSC plant. The input layer of the ANN model consists volume flow rate and temperature of rich oil (Fin and Tin), pressure of overhead (PT), volume flow rate and temperature of super heat steam (Fsin and TS1), flow rate of reflux (FR1), temperature of the reflux (TR1), flow rates of diesel, kerosene and heavy naphtha splitter vapors (R14, FRK24 and FRHN34, respectively), temperature of fifth tray (T5), and the value of pressure between trays 5 and 6 (P56). It should be mentioned that to prevent 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.

4. Results and discussions

During the period of data collection (about 300 days), 153 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 91 and 62 data points (60% and 40% of total data), respectively. All data points are scaled based on the maximum and mini-mum 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 the flow rates of output streams from the RSC, i.e. LPG, LN, HN, Ker and Dis are presented in Table 1, respectively. From this table, it can be concluded that R can learn the pattern of distillation products. The total CPU time for training each network is less than 3 sec which can be an advantage for developing ANNs using R software. However, this table shows that the iteration is stopped for LPG and LN in less than 100 steps, and training is terminated. Moreover, it results that the AAD% of training for LPG and LN are higher than 5%, in consequence with raising the AAD% of prediction for unseen data (testing). It is supposed that due to the high volatility and vapor pressure of these streams, measuring their flow rate has the abnormal gross error. Moreover, our observations confirm that setting the flowrates of these streams by adjusting the manipulated variables of the RSC is tremendously difficult, and any nuance changes create an enormous deviation in the set point volume flow rates of LPG and LN. Therefore, the collected data for these outputs are not certain which are clearly justified by results of developed ANNs. Additionally, the feasibility of some faults such as signal transmis-sion, calibration, and power fluctuation should not be neglected. Thus, the ANN model cannot test and train flow rates of LPG and LN based on the adjusted variables obtained from the target RSC.

Table 1. Number of iterations and AAD% of training and testing data for VGO hydrocracking neural networks

Product	AAD%_train	Steps	AAD%_test
LPG	7.07	92	17.4
LN	25.66	46	56.28
HN	2.56	798	5.86
Ker	1.55	1874	3.42
Dis	2.1	3009	3.95

The topology of ANNs which are constructed by R for flow rates of LPG, LN, HN, Ker, and Dis are illustrated in Figures 2 to 6, respectively. As seen, each network consist of 42 weights and 4 biases, totally 46 coefficients. As mentioned before, for training a network, there are 91 data points for the yield of each hydrocracking product. Hence, the degree of freedom (DOF) for each ANN is equal to 45. This positive DOF slightly prevents memorizing (over learning) instead of training. It is obvious that increasing the number of data points prevents the over learning, and enhances the resiliency of the ANN model.

After training ANNs, the unforeseen (predicting) data set are injected into the model, and the corresponding flow rate for each distillation stream 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 AADs% of testing data points, and except to those of LPG and

LN, they are near to the AAD% of trained data. In order to further investigation the reliability of the model, actual yields of LPG, LN, HN, Ker, and Dis are sketched versus the predicted values as parity plots in Figures 7 to 11, respectively. All these figures are achieved by graphical tools provided in the R software. As seen, ANNs for predicting HN, Ker, and diesel flow rates have high certainty; thus, they can be utilized for the purpose of planning, sensitivity analysis and optimizing the target RSC plant.
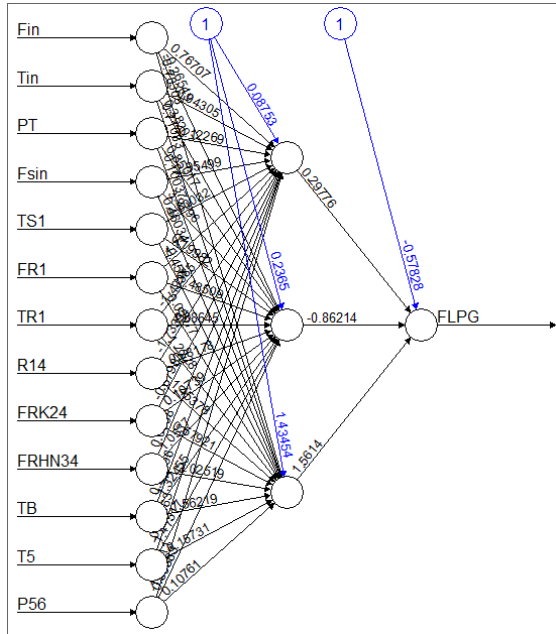
Figure 2. Structure of ANN constructed by R for predicting the yield of LPG
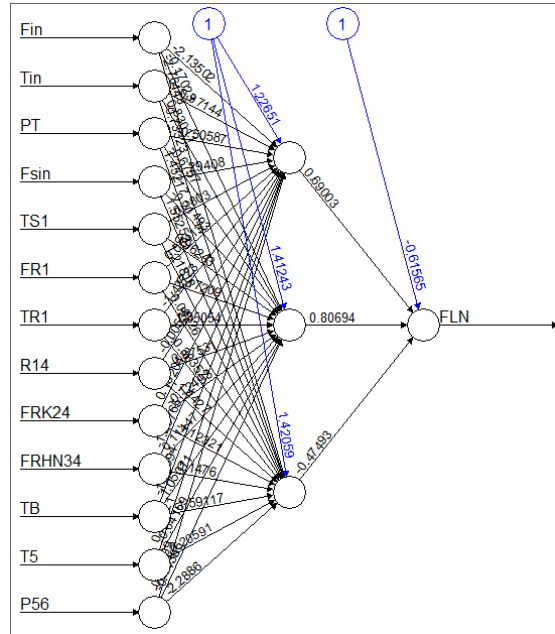
Figure 3. Structure of ANN constructed by R for predicting the yield of LN
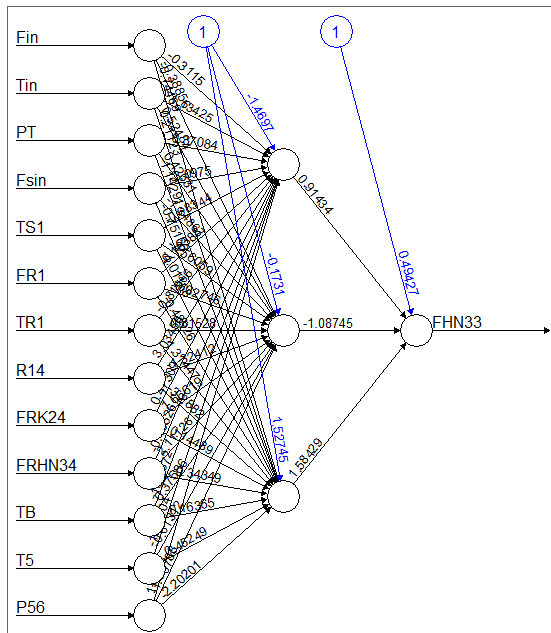
Figure 4. Structure of ANN constructed by R for predicting the yield of HN
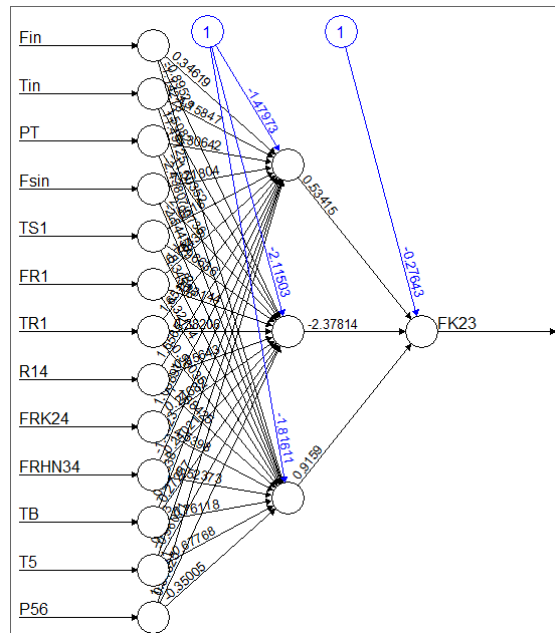
Figure 5. structure of ANN constructed by R for predicting the yield of kerosene
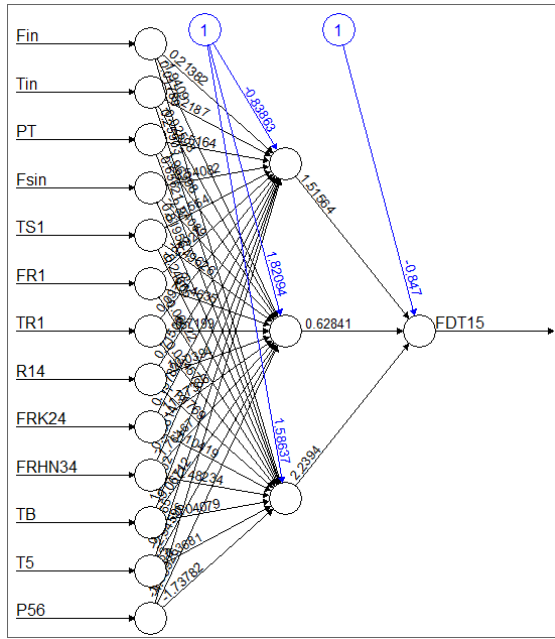
Figure 6. Structure of ANN constructed by R for predicting flow rate of diesel
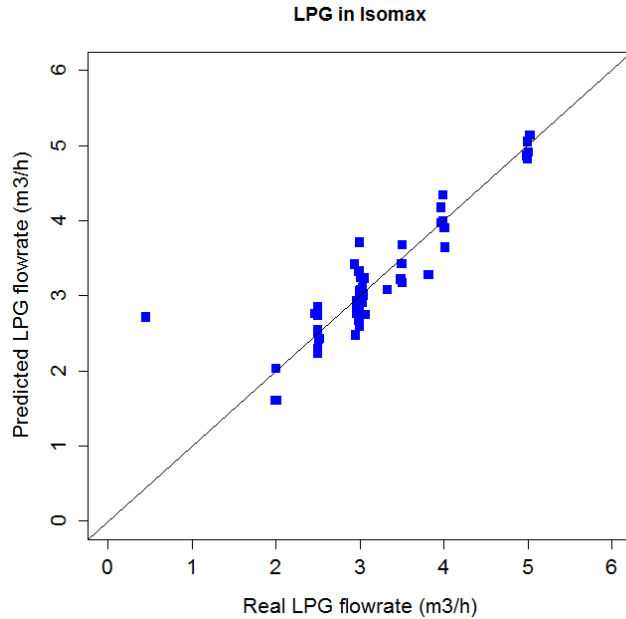

Figure 7. Comparison between actual and predicted values of LPG flow rate predicted by R
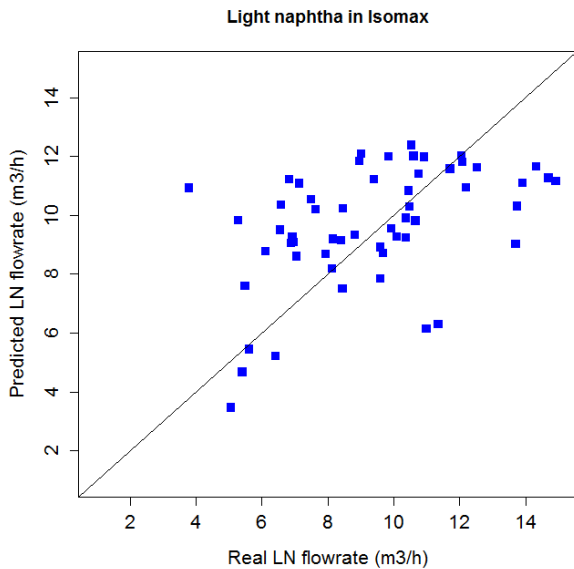

Figure 8. Comparison between actual and predicted values of LN flow rate predicted by R
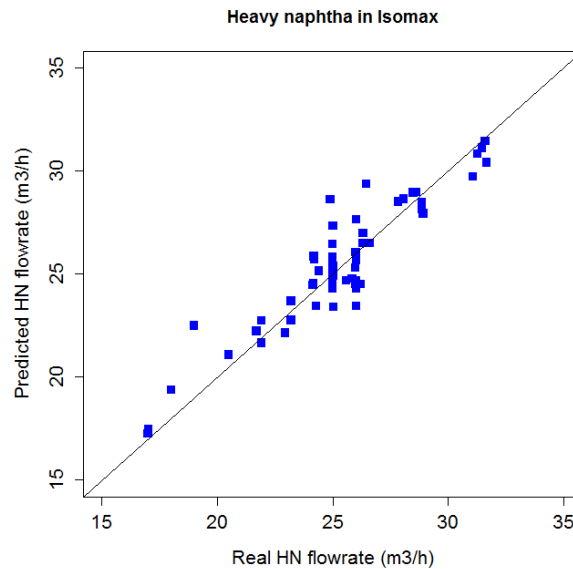

Figure 9. Comparison between actual and predicted values of HN flow rate predicted by R

5. Conclusions

This research demonstrated that the reflux splitter column (RSC) of a commercial scale VGO hydrocracking plant could be molded by using a feed forward artificial neural network. This model was implemented in R software by using 'neuralnet' package. Neural networks could learn about the connection among manipulating variables of the distillation column and flow rates of LPG, light naphtha, heavy naphtha, kerosene and diesel cuts.

To develop the ANN model, 153 actual data points gathered from the target RSC unit. The architecture of ANN models was designed based on a feed-forward neural network with 13 neurons in the input layer and 3 neurons in the hidden layer. For training step, 91 of data points were randomly chosen, and it was performed by using the back-propagation method. Other unseen data points (i.e., 62 data points) were utilized to investigate predicting ability of ANN models constructed in R environment. Results showed that these models could predict flow rates of heavy naphtha, kerosene, and diesel with appreciable accuracy; however, for that of light naphtha and LPG, the predictions were not reliable. This methodology did not need the huge amount of efforts and time such as the mechanistic approaches. Therefore, it can be applied for optimizing and planning the performance of the target RSC unit.
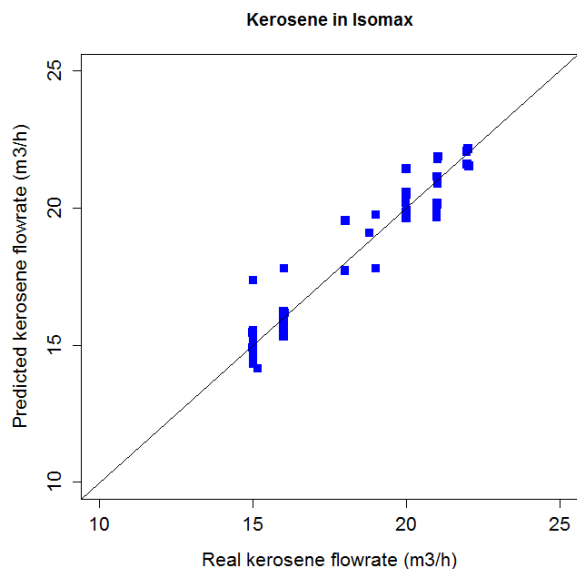

Figure 10. Comparison between actual and predicted values of kerosene flow rate predicted by R
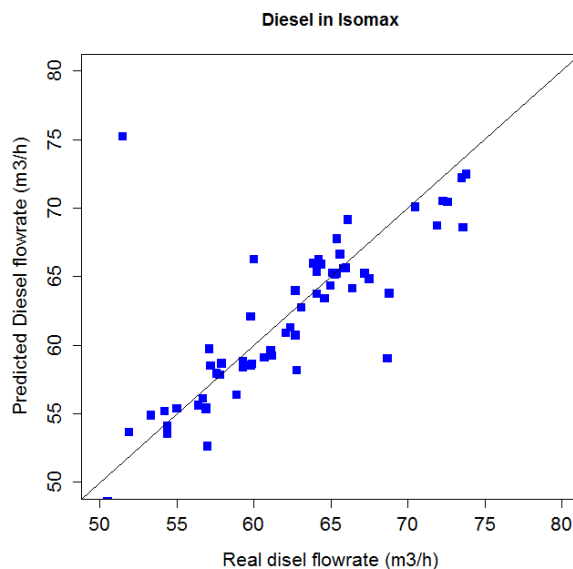
Figure 11. Comparison between actual and predicted values of diesel flow rate predicted by R

Acknowledgements

References

[1]   Sadighi S. J. Chem. Eng. Jap., 2016; 49(12), 979.
[2]   Nikolaychuk E, Stratiev D, Velkov I, Veli A, Sotirov S, Mitkova M, Pet. Coal, 2015; 57(3): 266-279.
[3]   Sadighi S, Zahedi S, Hayat, R, Bayat M. Ener. Technol., 2013; 1: 743.
[4]   Sadighi S, Mohaddecy R, Norouzian A. Bull. Chem. React. Eng. Catal., 2015; 10: 210.
[5]   Sadighi S, Mohaddecy R, Pet. Coal, 2017; 59(2): 195-201.
[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 MV. Chem. Eng. J., 2007; 133: 133.
[9]   Bhutani, N.: Rangaiah, G.P.: Ray, A.K., Ind. Eng. Chem. Res., 2006; 45: 7807.
[10]   Niae A, Towfighi J, Khataee AR. Rostamizadeh K. 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, Pet Coal, 2017; 59(6): 777.
[19]   Sadighi S, Mohaddecy SR, Abbasi A. Int. J. Technol., 2018; 9(1): 99.
[20]   Kusumoputro B, Sutarya D, Faqih A. Int. J. Technol., 2016; 4: 709.
[21]   Himmelblau DM. Korean J. Chem. Eng., 2000; 17(4): 373.
[22]   Biyanto R Widjiantoro BL, Jabal AA, Budiati T. Int. J. Eng. Sci. Technol., 2010; 2(6): 177-188
[23]   Ahmed DF, Khalaf AH. Journal of Chem. Eng. Process Technol., 2016; 7: 272.
[24]   Pipoola LT, Babagana G, Susu AA. Int. J. Recent Res. Appl. Studies, 2013; 15(3): 337.
[25]   Sadighi S, Ahmad A. Can. J. Chem. Eng., 2013; 91: 1077.
[26]   Sadighi S, Mohaddecy SR. Pet. Coal, 2018; 60(3): 358.
[27]   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.
[28]   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.
[29]   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, sadighi_sepehr@yahoo.com