Iranian Journal of Chemical Engineering
Vol. 7, No. 4 (Autumn), 2010, IAChE

Research note

Product Yields Prediction of Tehran Refinery Hydrocracking Unit Using Artificial Neural Networks

M. Bahmani[1], Kh. Sharifi[2], M. Shirvani[2]*

1- Department of Chemistry, Applied Chemistry Group, Tarbiat Moalem University, Dr. Mofatteh Street, Tehran, Iran.
2- Department of Chemical Engineering, Iran University of Science and Technology, Narmak Street, Tehran, Iran.

Abstract

In this contribution Artificial Neural Network (ANN) modeling of the hydrocracking process is presented. The input–output data for the training and simulation phases of the network were obtained from the Tehran refinery ISOMAX unit. Different network designs were developed and their abilities were compared. Backpropagation, Elman and RBF networks were used for modeling and simulation of the hydrocracking unit. The residual error (root mean squared difference), correlation coefficient and run time were used as the criteria for judging the best network. The Backpropagation model proved to be the best amongst the models considered. The trained networks predicted the yields of products of the ISOMAX unit (diesel, kerosene, light naphtha and heavy naphtha) with good accuracy. The residual error (root mean squared difference) between the model predictions and plant data indicated that the validated model could be reliably used to simulate the ISOMAX unit. A four-lumped kinetic model was also developed and the kinetic parameters were optimized utilizing the plant data. The result of the best ANN model was compared to the result of the kinetic model. The root mean square values for the kinetic model were slightly better than the ANN model but the ANN models are more versatile and more practical tools in such applications as fault diagnosis and pattern recognition.

Keywords: *Artificial Neural Networks, Hydrocracking Process, Product Yields*

1- Introduction

1.1- Hydrocracking unit

The growing demand for middle distillates and the increasing use of heavy crude oils as feedstocks have made hydrocracking one of the most importance upgrading petroleum refinery processes. During hydrocracking, high molecular weight compounds are broken down to form low molecular weight compounds. The reaction takes place over a catalyst in a hydrogen–rich atmosphere and other reactions such as hydrodesulfurization and hydrodemetallization occur simultaneously. [1]

* Corresponding author: bahmani@tmu.ac.ir

A simplified flow diagram for this configuration is shown in Fig. 1. The fresh feed combined with the recycled unconverted fraction is passed downward through the reactor in the presence of hydrogen. Because hydrogen is consumed during hydrocracking, make-up hydrogen has to be added to the recycled hydrogen. The effluent of the hydrocracking reactor is passed through high- and low-pressure separators where the hydrogen is converted and recycled. The liquid product, denoted in Fig. 1 as a single-stage product, is sent to the fractionator where the final products are separated from the unconverted fraction.

Ideally modeling of the hydrocracking process should take into account various reactions and mechanisms that are involved with all the components in the petroleum reaction mixture. However, in actual practice, it is very difficult to keep track of individual components and the reactions involved. This leads to an approximate and quite sophisticated model that cannot be directly used by the field engineer to predict the product yields at different operation conditions. Fig. 2 shows one such reaction network consisting of 10 lumps. This reaction network requires determination of a large number of rate constants namely, 32. alage Moreover, the models prepared have a number of physical parameters inherent to the process that are usually unknown and have to be determined through a parameter estimation methodology. [2, 3]

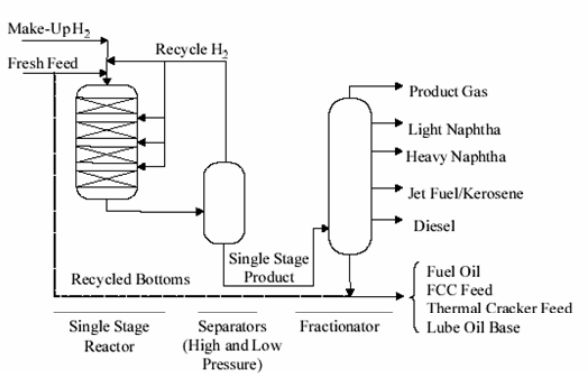

Figure 1. Simplified flow diagram of hydrocracking unit

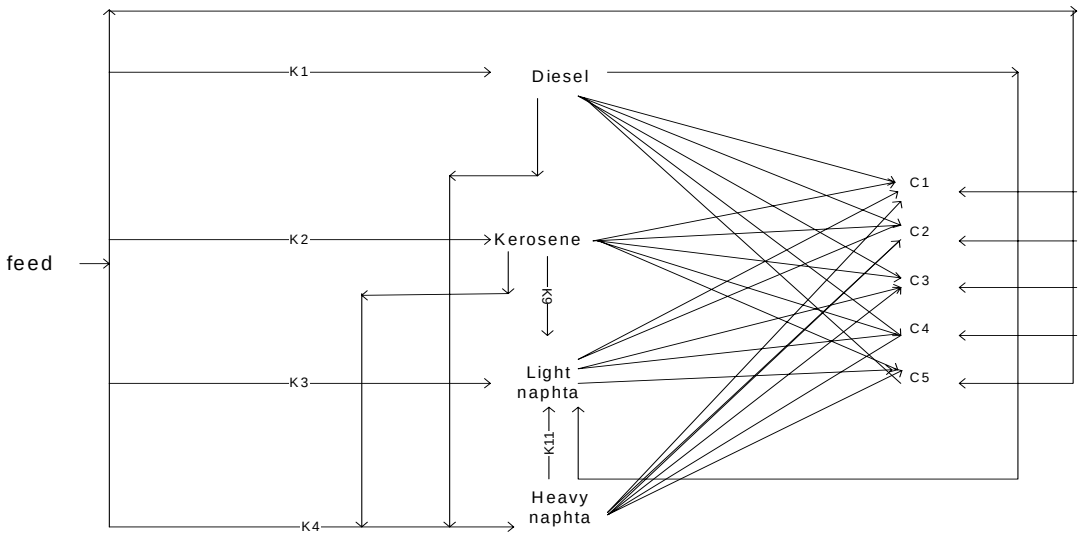

Figure 2. Hydrocracking reaction network based on discrete lumping approach

Kinetic modeling of heavy fractions hydrocracking has been investigated based on two main approaches; lumped models [1, 4-8] and detailed molecular models [9]. The lumped models have been based on discrete and continuous lumped models. In the discrete lumping approach, the individual components in the reaction mixture are divided into discrete pseudo components (lumps) based on the physical properties such as; true boiling point (TBP), carbon number (CN), and/or molecular weight (MW). The success of this approach depends on the number of lumps. A large number of lumps would result in more accurate model predictions, but this increases the number of kinetic parameters that should be determined [10].

1.2-Artificial neural networks

In the cases where a deterministic model cannot adequately describe a system, then the use of a neural network, which is actually a 'black box', can give better results. Deriving information from the biological counterparts, neural networks are based on the collaboration of highly interconnected simple functional units that express a complicated cross-correlation between dependent and independent variables [11]. Artificial neural networks (ANNs) offer a promising alternative to modeling for a number of reasons. ANNs are able to capture the nonlinearities in system behavior very effectively. They are based on plant data that is usually available from the operation of the unit in previous times. Additionally, a neural network model of the hydrocracking unit does not need complex analysis of the feedstock, is tolerant to partial and noisy data input, and can be improved continuously as more operational data are made available. [12]. Backpropagation, Elman and Radial Basis Function (RBF) are networks that are usually used in the chemical process [13].

The objective of this work was to explore the use of Backpropagation, Elman and RBF neural network architectures in creating a model of the hydrocracking process. Our approach is based on: (1) formulation of an artificial neural network (ANN) with the input and output variable values obtained from a hydrocracking unit of an operating refinery, (2) preprocessing the data including: (a) elimination of the out of range data, (b) normalization of input data and (c) principal component analysis to reduce the dimension of the input vectors, (3) dividing the input data randomly to the training, validation and the test sets, (4) training and testing the ANN with the refinery data and selecting the best architecture and parameters for these networks.

Backpropagation is the generalization of the Widrow-Hoff learning rule for multiple-layer networks and nonlinear differentiable transfer functions. Input vectors and the corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as defined by the user. Networks with biases, a sigmoid layer, and a linear output layer are capable of approximating any function with a finite number of discontinuities. Standard Backpropagation is a gradient descent algorithm, as is the Widrow-Hoff learning rule, in which the network weights are moved along the negative of the gradient of

the performance function. Properly trained Backpropagation networks tend to give reasonable answers when presented with inputs that they have never seen [14, 15].

RBF networks can require more neurons than standard feedforward Backpropagation networks, but often they can be designed in a fraction of the time it takes to train standard feedforward networks. They work best when many training vectors are available [16].

Elman networks are two-layer Back-propagation networks, with the addition of a feedback connection from the output of the hidden layer to its input. This feedback path allows Elman networks to learn, recognize and generate temporal patterns, as well as spatial patterns. The Hopfield network is used to store one or more stable target vectors. These stable vectors can be viewed as memories that the network recalls when provided with similar vectors that act as a cue to the network memory [17].

In this study, two different backpropagation networks were initially designed and their results were compared to actual plant data. The first model (model 1) is based on six input and four output variables and the second model (model 2) has nine input and four output variables. Finally, the best backpropagation network was then compared with the RBF and Elman networks.

2- Back propagation neural network models

2.1-Backpropagation neural network model development

In the present work extensive plant data from the ISOMAX unit of Tehran refinery has been collected. The data used in the model development is representative of plant operation, spanning the range of operating conditions that may be encountered during routine operating conditions. The data must be pre-processed in order for the ANN to effectively learn from it.

Due to the large number of data sets collected, some statistical analysis was applied to the data, and data values that appeared to be scattered far away from the majority of values were considered as outliers and were thus excluded from the data set. After removing the outliers the number of data records in the database consisted of 1408 data points. Then the data were normalized linearly between -1 and 1, considering the minimum and maximum values. This normalization of variable values was carried out in order to avoid using data spanning over different orders of magnitude and therefore they have zero mean and unit variance.

An effective procedure for reducing the dimension of the input vectors is principal component analysis. For this purpose using the neural network toolbox in Matlab those components that contribute to less than 2% of the overall variation in the data set were discarded.

The refined database was then randomly divided into three distinct subsets. 60% of the database was used to train the model. Training a neural network is an iterative process in which a nonlinear optimization algorithm, typically a gradient descent method, is used to obtain the optimal values of the weights and biases.

The second subset, which was a 20% cut from the original database, was used for validating the ANN. The error on the validation set is monitored during the

training process. The validation error normally decreases during the initial phase of training, as does the training set error. However, when the network begins to over fit the data, the error on the validation set typically begins to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are returned.

The third subset was 20% of the database and was used to test the ANN. This data set was in fact used to compare different models. In this study the physical properties and operating conditions of the reactor and distillation section of the hydrocracking unit were taken as the inputs and the yields of the products were taken as the outputs of the networks.

2.2- Designs of the backpropagation networks

The performance of an ANN is very dependent on the number of hidden layers and the number of hidden nodes in each layer. The feedforward networks often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. Determination of the number of hidden nodes is one of main aspects of designing the neural network architecture. Too small a number of hidden nodes limits the ability of the ANN to model the problem and, furthermore, such a network may not train well to an acceptable error. On the other hand, too many hidden nodes would result in the network memorizing the data rather than learning them for generalization. The number of hidden nodes is determined by varying the number of nodes, starting with only a few hidden nodes and then adding up more nodes as the Mean Squared Error (MSE) computed on the training patterns comes to a minimum. The number of hidden nodes at that point is taken as the optimum.

In this work, for model 1 each neuron in the input layer corresponds to a particular feed property or reactor operating conditions of the unit, and the output layer corresponds to the product yields. Fig. 3 shows the neural network architectures for model 1, consisting of an input layer, a hidden layer, and an output layer. Subsequently, several ANN structures with one hidden layer of sigmoid neurons and one output layer of linear neurons were tested to identify the most appropriate number of hidden nodes. Fig. 4 shows the effect of number of hidden nodes on the MSE. The best results were obtained for the network architecture of 6×36×4, composed of 6 input nodes, 36 hidden nodes, and 4 nodes in the output layer.

In model 1 the operating conditions of the distillation section of the ISOMAX unit was not utilized. Whilst, in model 2 the operating conditions of the distillation section such as the temperature of the middle tray, pressure of the column and the amount of used steam were also included as input data.

Fig. 5 shows the neural network architectures for model 2. In Fig. 6 the effect of number of hidden nodes on the MSE for model 2 can be seen. It can be concluded from this figure that the best results were obtained for the network architecture consisting of 9 input nodes, 43 hidden nodes, and 4 nodes in the output layer. The number of neurons in the hidden layer in this model corresponded to the lowest prediction error as was obtained for model 1. It can be seen that model 2 contains more hidden nodes than model 1

and therefore, it is a more complex network. Moreover, prediction error in model 2 is less than model 1, since in model 2 the effect of the operating conditions of the distillation section of the hydrocracking process is also considered. Prediction error was defined as MSE differences between measured values and those computed by the ANN model.

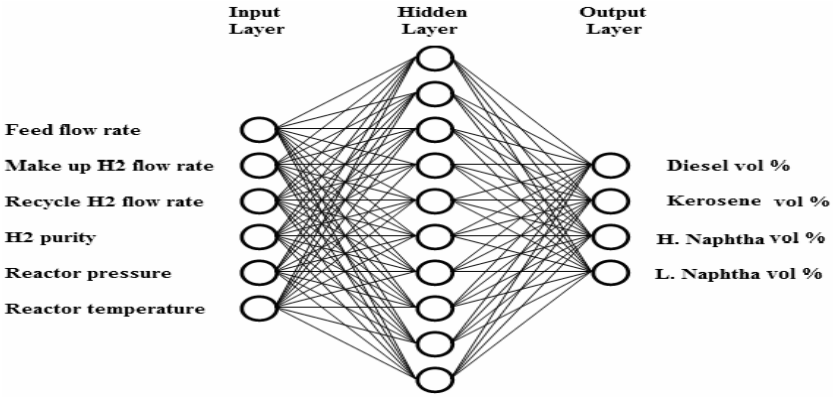

Figure 3. Network architecture for model 1

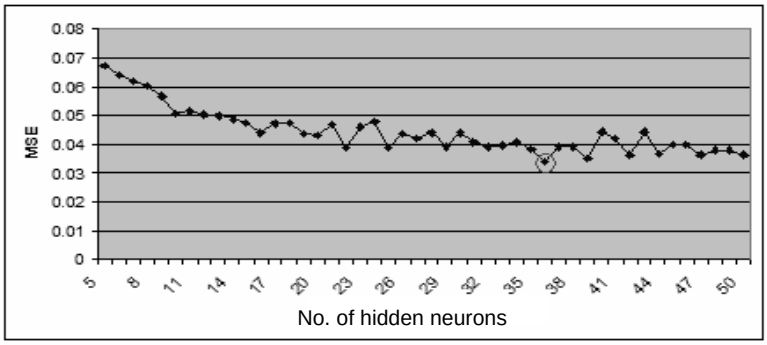

Figure 4. Mean square error as a function of number of neurons in the hidden layer for model 1

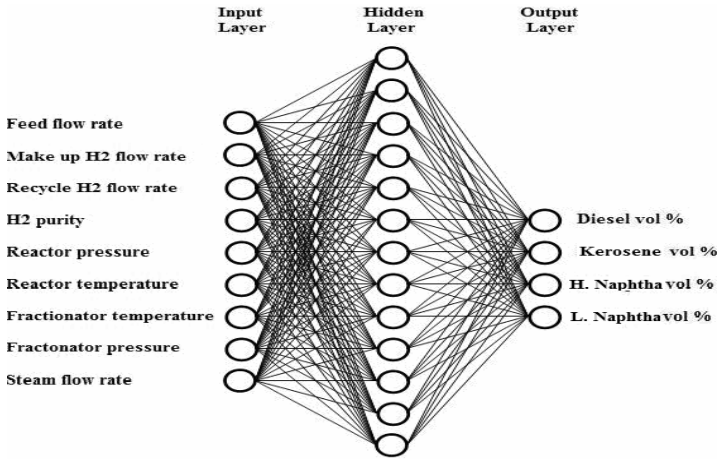

Figure 5. Network architecture for model 2

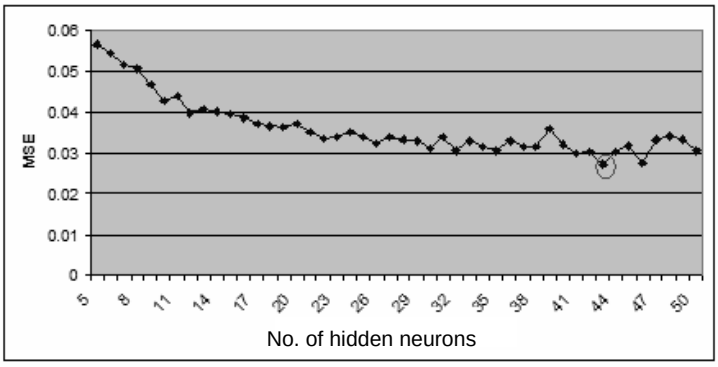

Figure 6. Root mean square as a function of neurons in the hidden layer for model 2.

2.3- Discussion of the performance of the backpropagation ANN models

In order to judge the effectiveness of the neural networks obtained, the results obtained were compared with actual measurements obtained from the hydrocracking plant. Fig. 7 shows the actual and ANN predicted diesel yields for model 1 and Fig. 8 shows the actual and ANN predicted diesel yields for model 2. The continuous line represents the daily predictions of the diesel yield by the ANN, whereas the actual values are indicated by symbols. It can be seen that the ANN predictions are in good agreement with actual refinery daily measurements, except for some points scattered around the main trend of the mean measured values.

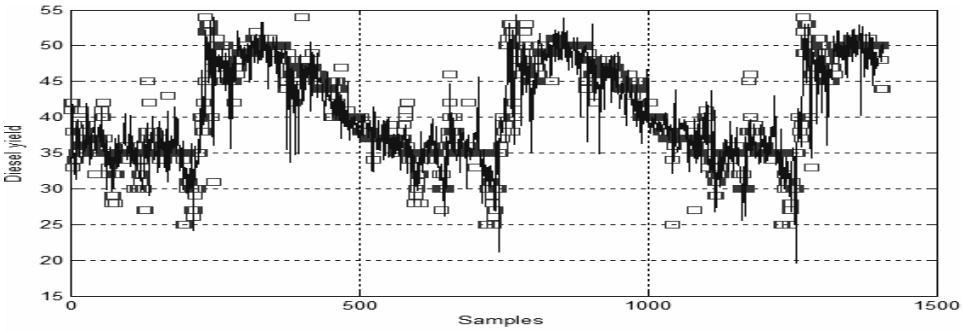

Figure 7. Comparison of actual and backpropagation ANN predicted diesel yield for model 1

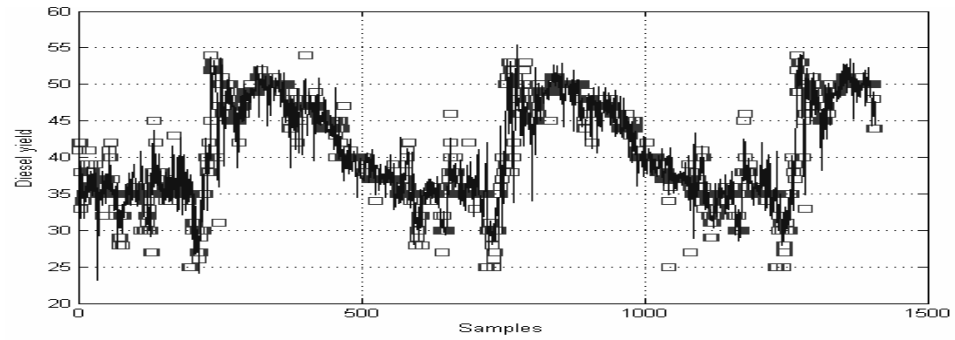

Figure 8. Comparison of actual and backpropagation ANN predicted diesel yield for model 2

Prediction performance of the ANN for the kerosene yields for model 1 and model 2 are depicted in Figs. 9 and 10, when applied to the test data set. There is a close correspondence between the measured and predicted values. The equations of these correlations are shown in the figures. In these equations Output represents the predicted and Target the measured values, respectively. The correlation coefficient R is computed and shown for each line; it is used to determine the strength of the correlation between the measured and predicted values. While values of the slope of the fitted line show how much of the variation in the predicted values is accounted for by the model, obviously a slope close to one indicates high model prediction accuracy. Ideally the slope and the Y-axis intercept of the fit should be 1 and 0 respectively. It can be concluded from these figures that a good fit was observed for the data set, with better predictions for model 2.

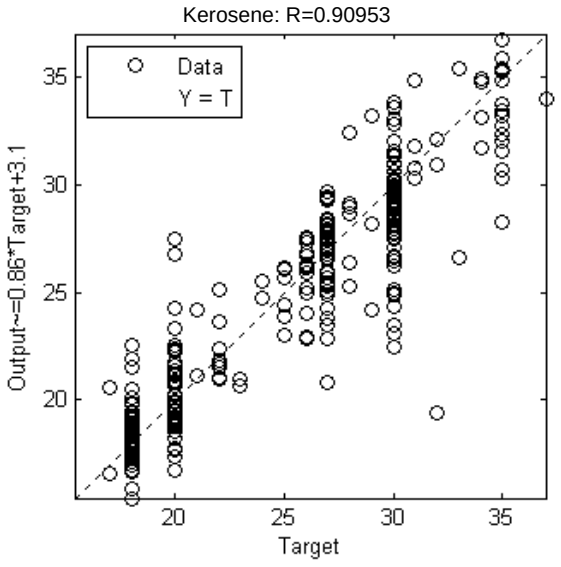

Figure 9. Correlation between measured and predicted values for model 1.

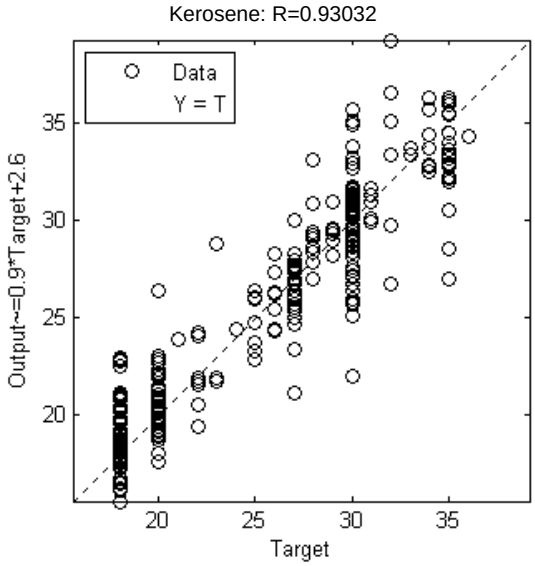

Figure 10. Correlation between measured and predicted values for model2.

Due to the large number of data sets (typically 280 data points) used in the testing phase of ANN, a number of outliers can be seen spread around the mean value. These outliers may be justified by reasoning that the actual plant data could inherently contain some errors.

3- Elman and RBF neural network models

3.1- Elman and RBF neural network model development

Similar to model 2 of the Backpropagation network, the Elman networks were designed with three layers consisting of an input layer, a hidden layer, and an output layer. The best architecture for these networks was achieved by varying the number of nodes in the hidden layer. Simplified network architecture of Elman networks is presented in Fig. 11. But in the RBF network the best results can be obtained by varying the value of the parameter of SPREAD. SPREAD determines the width of an area in the input space to which each neuron responds. SPREAD

should be large enough so that the neurons can respond strongly to the overlapping regions of the input space.

3.2- Comparison of Backpropagation, Elman and RBF networks

To evaluate the efficiency and abilities of Backpropagation and Elman networks, variations of MSE with a different number of hidden neurons for these networks are shown in Fig. 12, and additionally, Table 1 shows the variation of run time with the number of hidden neurons for these two networks. The MSE values for the backpropagation network are slightly lower than the Elman network, but it can be seen from Table 1 that for the same number of hidden nodes the backpropagation network converges in seconds, whereas the Elman network takes 30 minutes to converge.

For the RBF network, the variations of MSE and run time with increments in the amount of SPREAD are shown in Table 2. Fig. 13 shows that the backpropagation network has lower MSE values. The lowest error for the model 2 of the Backpropagation network was obtained for a network with 9 input nodes, 43 nodes in the hidden layer and 4 nodes in the output layer. This result was also obtained for the Elman network. When the parameter of SPREAD is equal to 45, the lowest error for the RBF network was also achieved. The run time for the RBF network is less than the two other networks, but the error for this network is more than the two other networks.

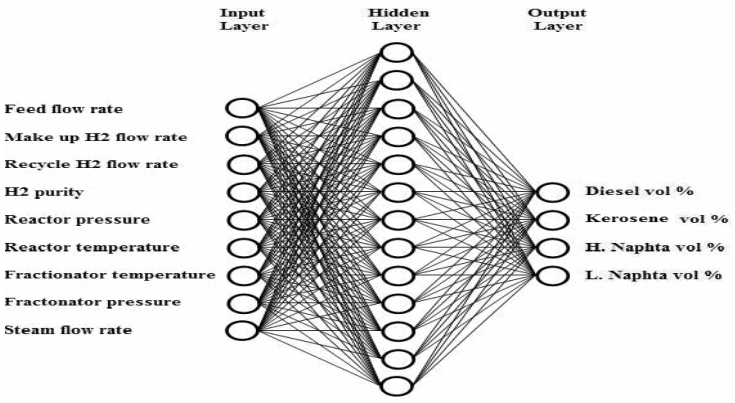

Figure 11. Network architecture for Elman networks

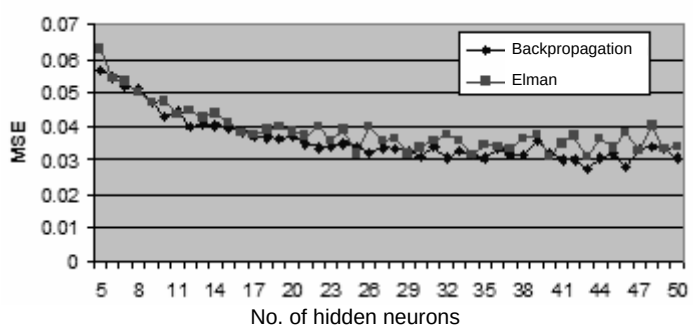

Figure 12. Comparison of MSE of the model 2 of the backpropagation and the Elman network

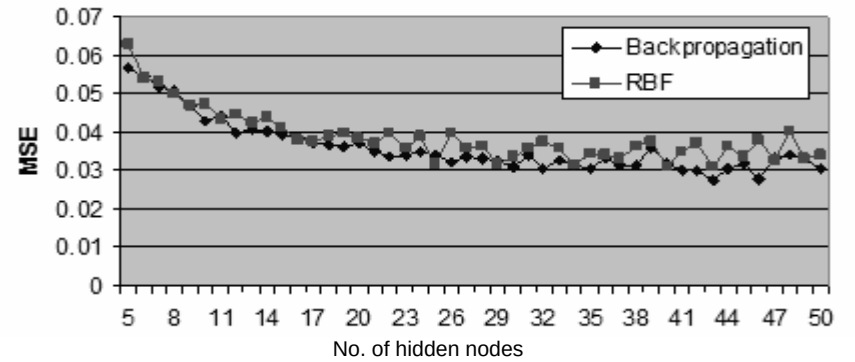

Figure 13. Root mean square as a function of hidden neurons for Backpropagation and RBF networks

Table 1. Run Time variation with number of hidden nodes for Backpropagation and Elman networks

No. of hidden neurons	Run Time	
	Backpropagation	RBF
5	0:00:03	0:00:53
6	0:00:04	0:01:42
7	0:00:03	0:01:58
8	0:00:02	0:02:04
9	0:00:03	0:04:11
10	0:00:07	0:02:31
40	0:00:11	0:29:56
41	0:00:09	0:16:07
42	0:00:08	0:21:35
43	0:00:11	0:22:04
44	0:00:12	0:17:33
45	0:00:12	0:20:51
46	0:00:13	0:18:14
47	0:00:11	0:24:17
48	0:00:17	0:18:41
49	0:00:13	0:30:05
50	0:00:12	0:28:15

Overall, amongst these networks the backpropagation network proved to be the best network for modeling the hydrocracking unit. Elman network is a historical network and it takes more CPU time and, furthermore, its error is more than the backpropagation network. The drawback to the RBF network is that it produces a network with as many hidden neurons as there are input vectors. For this reason, RBF does not return an acceptable solution in cases that many input vectors are needed to properly define a network, as is typically the case.

Table 2. MSE and Run Time variation with SPREAD for RBF networks

SPREAD	MSE	Run Time
0.1	0.2	5-6 sec
0.5	9.72	5-6 sec
1	58	5-6 sec
2	168.2	5-6 sec
3	205.5	5-6 sec
4	29.5	5-6 sec
5	6.9	5-6 sec
10	0.425	5-6 sec
100	0.049	5-6 sec
110	0.052	5-6 sec
120	0.054	5-6 sec
130	0.054	5-6 sec
140	0.054	5-6 sec
150	0.054	5-6 sec
160	0.054	5-6 sec
170	0.054	5-6 sec
180	0.054	5-6 sec
190	0.054	5-6 sec
200	0.054	5-6 sec

Correlation between measured and predicted diesel yield are shown in Figs. 14-16 for Backpropagation, Elman and RBF networks respectively. The equations of these correlations are shown in these figures. In these equations Output represents the predicted and Target the measured values

respectively. The correlation coefficient, R, for each line is computed and shown. It is used to determine the strength of the relationship between the measured and predicted values. There is a close correspondence between the measured and predicted values for each of the three networks, but the best correlation coefficient was obtained for the Backpropagation network.

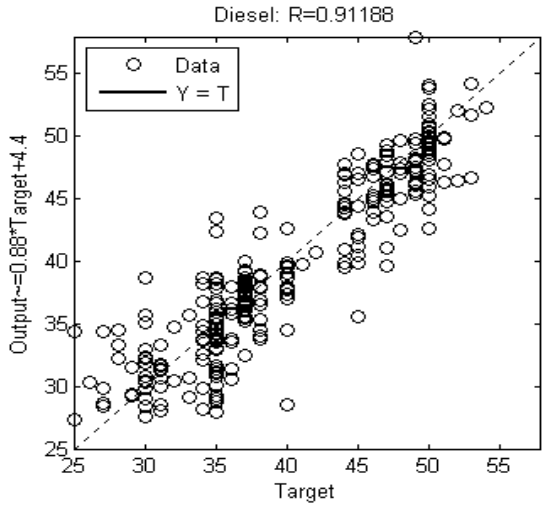

Figure 14. Correlation between measured and predicted diesel yield for Backpropagation network.

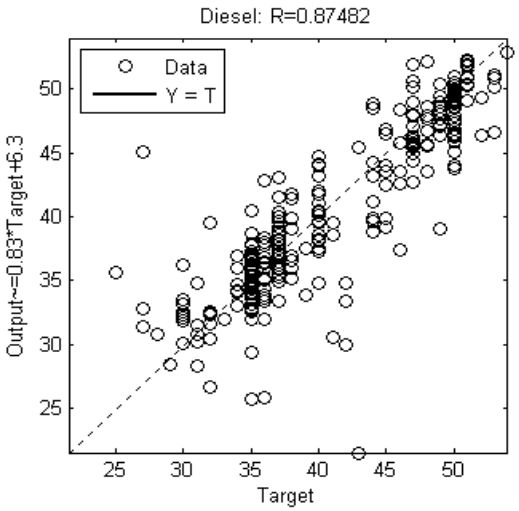

Figure 15. Correlation between measured and predicted diesel yield for Elman network.

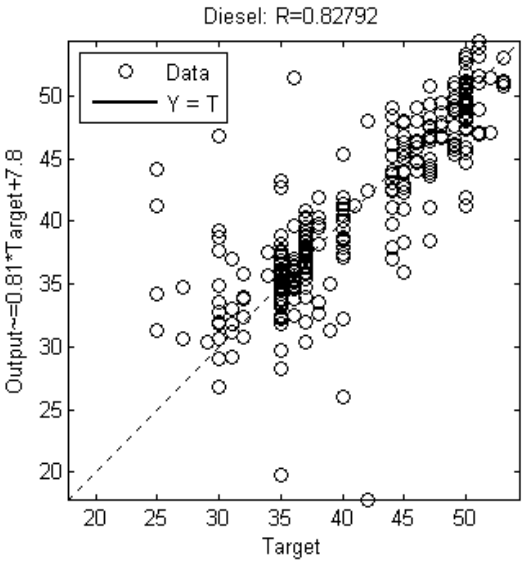

Figure 16. Correlation between measured and predicted diesel yield for RBF network.

4- Comparison of the ANN model with classical hydrocracking reaction models

A reaction network involving four lumps as shown in Fig. 17 was utilized to model the plant data.

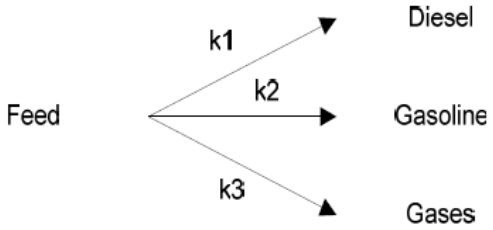

Figure 17. Reaction network for hydrocracking 4-lumped kinetic model

Mass balance involving first order kinetics was used to model the hydrocracking process as follows:

$$\frac{dy_i}{d\left(\frac{1}{LHSV}\right)} = r_i = K_i y_i$$

Where y_i is the weight fraction of the lumps, ri is the reaction rate for each lump and LHSV is the liquid Hourly Space Velocity. The reactions were assumed to be first order as recommended by Quader [4] and the rate constants were assumed to follow an Arrhenius type dependency on temperature $K = K_0 EXP\left(\frac{-E}{RT}\right)$ in which K_0 is the frequency factor and E is the activation energy. The minimization of the objective function, based on the sum of square errors between the experimental and calculated product compositions, was applied to find the best set of kinetic parameters. This objective function was solved using the least-squares criterion with a nonlinear regression procedure based on Marquardt's algorithm available in MATLAB. The optimized kinetic parameters are shown in Table 3.

Table 3. Optimized kinetic parameter values for the four lumped kinetic model

	K_0 (1/hr)	E (KJ/mol)
K_1 (rate constants for production diesel)	1.3618e10	19664.38
K_2 (rate constants for production of gasoline)	3.068e11	52822.036
K_3 (rate constants for production of gases)	1.225e11	20277.88

Figs. 18 and 19 show the parity plot for the kinetic model developed and, as it can be seen, there is a good comparison between the model and plant data. The MSE values for the kinetic model are shown in Table 4. Table 5 shows the MSE values for the ANN models developed in this study. Apart from the gas yields, the kinetic model produces better MSE values compared to the ANN models. Although this clearly shows the accuracy of the kinetic model in predicting the yields, nonetheless this finding has no direct impact on the usefulness of ANN models. The ANN's are widely accepted as practical tools for simulations in circumstances in which the complex interconnection of many input variables strongly correlate with many output variables. In this study the best ANN model is developed i.e., model 2 relates various input parameters as shown in Fig. 4. The effect of various operational parameters such as fractionator pressure and temperature, steam flowrate, H_2 purity, recycle rate and reactor temperature and pressure on the product yield can be easily studied by means of model 2, whereas to model such effects by means of classical chemical engineering techniques would require rigorously modeling the whole hydrocracking plant. Additionally, it is widely accepted that pattern recognition and fault diagnosis studies in chemical engineering can be easily carried out by means of ANN derived models, whereas similar studies utilizing classical chemical engineering techniques would pose a formidable task.

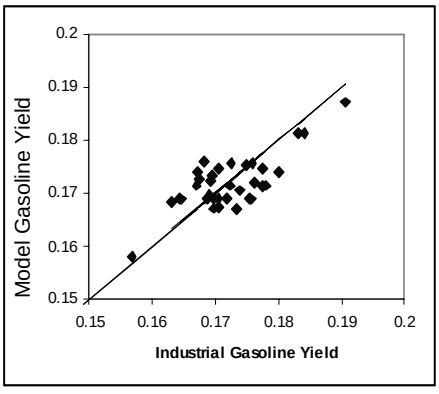

Figure 18. Parity plot for the gasoline yield of the 4-lumped kinetic model

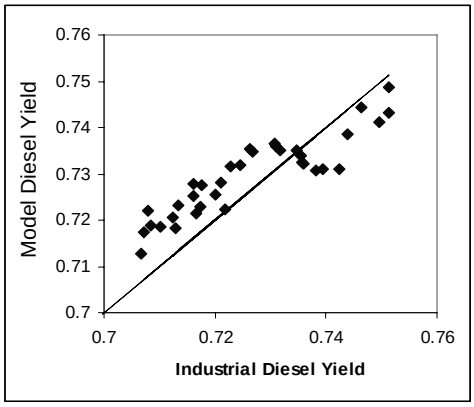

Figure 19. Parity plot for the diesel yield of the 4-lumped kinetic model

Table 4. Mean squared error values obtained for the 4-lumped model developed

MSE	Product
0.0252	Gases
0.0086	Gasoline
0.0091	Diesel

Table 5. Mean squared error values obtained for the 4- lumped model developed

ANN models developed in this study	**MSE values**
Backpropagation model1	0.035
Backpropagation model2	0.027
Elman	0.03
RBF	0.054

5- Conclusions

This work describes the methodology of developing artificial neural networks for predicting actual industrial outputs. The methodology was used to develop and validate neural networks, which were then used to predict the yields of hydrocracking unit products. Backpropagation, Elman and RBF networks were used for this purpose. Initially two different backpropagation networks were studied. The neural network models were trained using data from the Tehran refinery ISOMAX unit. The model predictions were in excellent agreement with industrial refinery data. The error in model 2 was less than in model 1, but the number of neurons in the hidden layer for model 2 was more than in model 1. Accuracy of the predictions of the trained neural networks to the refinery data was determined by the root mean squared error and the coefficient of correlation. Backpropagation, Elman and RBF networks were tested by comparing their MSE, Run Time and correlation coefficient between measured and predicted values. Among these networks Backpropagation network has the lowest value of MSE and the best correlation coefficient, and the RBF network has the minimum Run Time. Overall, Backpropagation network is the best model to predict the complex behavior of the multivariate hydrocracking unit.

Finally, a four-lumped kinetic model was developed and its kinetic parameters were optimized utilizing the plant data. Although the kinetic model predicted the product yields better than the ANN model, the ANN models relate input variables from different parts of the ISOMAX plant to the product yields and therefore the ANN models are more versatile than the kinetic model.

References

[1] Ancheyta, J., Sánchez, S., Rodríguez, M. A. "Kinetic modeling of hydrocracking of heavy oil fractions: A review", Catalysis Today, 109, 76–92 (2005).

[2] Balasubramanian, P., Pushpavanam, S. "Model discrimination in hydrocracking of vacuum gas oil using discrete lumped kinetics", Fuel, 87, 1660–1672 (2008).

[3] Chavarría-Hernandez, J.C., Ramírez, J. and Baltanás, M. A. "Single-event-lumped-parameter hybrid (SELPH) model for non-ideal hydrocracking of n-octane", Catalysis Today, 130, 455–461 (2008).

[4] Quader, S. A. and Hill, G. R. "Hydrocracking of gas oils". Ind. Eng. Chem. Process Des. DeV., 8 (1), 98, (1969).

[5] Laxminarasimahn, C.S., Verma, R.P. and Ramachandaran, P.A., "Continuous lumping model for simulation of hydrocracking", AIChEJ, Sept., 43, No.9, 2645-2653, (1996).

[6] Callegas, M. A. and Martinez, M. T. "Hydrocracking of a Maya residue: kinetics and products yield distribution". Ind. Eng. Chem. Res., 38, 3285, (1999).

[7] Mohanty, S., Saraf, D. N. and Kunzru, D., "Modeling of a hydrocracking reactor". Fuel Process. Technol., 29, 1, (1991).

[8] Baltanas, M. A. and Froment, G. F., "Computer generation of reaction networks and calculation of product distributions in the hydroisomerisation and hydrocracking of paraffins on Pt-containing bifunctional catalysts". Comput. Chem. Eng., 9, 77, (1985).

[9] Liguras, D. K. and Allen, D. T., "Structural models for catalytic cracking: 1. Model compounds reactions". Ind. Eng. Chem. Res., 28, 655, (1989).

[10] Bhutani, N., Ray, A. K., and Rangaiah, G. P., "Modeling, simulation, and multi-objective optimization of an industrial hydrocracking unit", Ind. Eng. Chem. Res., 45, 1354-1372, (2006).

[11]. Bellos, G.D., Kallinikos, L.E., Gounaris, C.E. and Papayannakos, N.G. "Modeling of the performance of industrial HDS reactors using a hybrid neural network approach", Chemical Engineering and processing, 44, 505-515 (2005).

[12]. Baghat, P. "An introduction to neural nets", Chem. Eng. Prog., 55-61 (1990)

[13]. Rallo, R., ferré-Giné, J., Arenas, A. and Giralt, F. "Neural virtual sevsor for the inferential prediction of product quality from process variables", Compute. Chem. Eng, 26, 1735-1754, (2002).

[14]. Rumelhart, D., Mcclelland, J. L. and Williams, R. J. "Learning representations by back-propagation", Nature 323, 533-536 (1986).

[15]. Hagiwara, K. and Kuno, K. "Regularization learning and early stopping in linear networks", International Joint Conference on Neural Networks, IJCNN2000, pp. 511-516 (2000).

[16]. Chen, S., Cowan, C.F.N. and Grant, P.M. "Orthogonal least squares learning algorithm for radial basis function networks", IEEE Transactions on Neural networks, Vol. 2, No. 2, March, pp. 302-309, (1991).

[17]. Elman, J. L., "Finding structure in time", Cognitive Science, Vol. 14, pp. 179-211, (1990).