Modeling of Oxidative Coupling of Methane over Mn/Na$_2$WO$_4$/SiO$_2$ Catalyst Using Artificial Neural Network

Ehsani, Mohammad Reza*+; Bateni, Hamed; Razi Parchikolaei, Ghazal

Department of Chemical Engineering, Isfahan University of Technology,
P.O. Box 84156-83111 Isfahan, I.R. IRAN

ABSTRACT: *In this article, the effect of operating conditions, such as temperature, Gas Hourly Space Velocity (GHSV), CH_4/O_2 ratio and diluents gas (mol% N_2) on ethylene production by Oxidative Coupling of Methane (OCM) in a fixed bed reactor at atmospheric pressure was studied over Mn/Na$_2$WO$_4$/SiO$_2$ catalyst. Based on the properties of neural networks, an artificial neural network was used for model development from experimental data. In order to prevent network complexity and effective data input to network, principal component analysis method was used and the numbers of output parameters were reduced from 4 to 2. A feed-forward back-propagation network was used for simulating the relations between process operating conditions and aspects of catalytic performance, which include conversion of methane, C_2^+ products selectivity, yield of C_2^+ and C_2H_4/C_2H_6 ratio. Levenberg– Marquardt method is presented to train the network. For first output, optimum network with 4-9-1 topology and for second output, optimum network with 4-6-1 topology was prepared.*

INTRODUCTION

Oxidative Coupling of Methane (OCM) Process

Oxidative Coupling of Methane (OCM) is one of the promising routes for direct conversion of natural gas to higher hydrocarbons, especially ethylene that can be used for future production of basic petrochemicals (ethylene and ethane) or liquid fuels [1,2]. OCM is particularly important because ethylene is precursor to many other reaction products [3].

The key to commercial success of OCM is the identification of a suitable catalyst capable of high C$_2$ selectivity and high yield at significant level of methane conversion especially at low temperature conditions [2,4]. With the aim of improving ethane and ethylene's yield, many researches have been focused on the catalysts used in the OCM process, their properties and influential factors on their performance [5]. *Fang et al.* have identified Mn/Na$_2$WO$_4$/SiO$_2$ as a promising OCM catalyst. Malekzade et al. represented that the performance of MO$_x$/Na$_2$WO$_4$/SiO$_2$ catalysts for the oxidative coupling of methane is well correlated with the electrical conductivity

of the catalysts under OCM conditions [6]. *Mahmoodi et al.* have indicated that the sodium salts and different oxo anions largely influence the structures, reducibility, and catalytic performances of the M–Na–Mn/SiO$_2$-based nanocatalysts (M = V, Cr, Nb, W, Mo) in the oxidative coupling of methane reaction. Among these synthesized nanocatalysts, Na$_2$WO$_4$–Mn/SiO$_2$ as well as Na$_2$MoO$_4$–Mn/SiO$_2$ shows the best catalytic performance at the OCM reaction conditions [7]. In this work Mn/Na$_2$WO$_4$/SiO$_2$ has been applied to OCM reaction because of its high catalytic performance.

Artificial Neural Networks

Neural networks, or more precisely Artificial Neural Networks (ANNs), are a branch of artificial intelligence [8]. ANNs were inspired by biological neural networks [9,10]. They were made of a large number of simple computing elements, called nodes or neurons that arranged to form an input layer, one or more hidden layers and an output layer [11]. They further include interconnections between the nodes of successive layers through the so-called weights [12]. The role of the weights is to modify the signal carried from one node to the other and either enhance or diminish the influence of the specific connection [13]. Each neuron in the hidden layer receives weighted inputs from each neuron in the previous layer plus one bias, as given by Eq. (1).

$$Z_i = \left(\sum_{k=1}^{N_{j-1}} X_k^{j-1} W_{k,i} + b_i \right) \tag{1}$$

where X_k^{j-1} denotes the input from k-th node in the j-th layer, $W_{k,j}$ is the weight of the link between node k and all the nodes in the previous layers, and b_i is the bias to the node, N_{j-1} is the number of nodes in the layer j-1. This sum is passed along to an activation function, to produce the output of the node, calculation as: $Y_i = f(Z_i)$ [10]. Activation or transfer function can be any type of mathematical function, but sigmoid function (Eq. (2)) is the most commonly used [9,10].

$$f(Z_i) = \frac{1}{1 + e^{-Z_i}} \tag{1}$$

The internal weights of the network are adjusted in the course of an iterative process termed training [12] and the algorithm used for this purpose so-called training algorithm. The error Back-Propagation (BP) algorithm is the most common form of learning, utilized today in artificial neural networks. There exist many network architectures such as Multilayer Perceptron, Radial Basis Function, Probabilistic Neural Networks and several others. Among them, Multilayer Perceptron is the most popular [10]. The number of nodes in the feed forward neural network input layer is equal to the number of inputs in the process, whereas the number of output nodes is equal to the number of process output.

Basically, the back-propagation training procedure is intended to obtain an optimal set of the network weight, which minimizes an error function. The commonly employed error function is the Mean Squared Error (MSE) as defined by

$$MSE = \frac{1}{N_P K} \sum_{i=1}^{N_P} \sum_{k=1}^{K} \left(t_{i,k} - y_{i,k} \right)^2 \tag{3}$$

Where N_P and K denote the number of patterns and output nodes used in the training, i denotes the index of the output pattern (vector), and k denote the index of the output node. Meanwhile, $t_{i,k}$ and $y_{i,k}$ express the desired (target) and predicted values of the k-th output at i-th input pattern, respectively [14].

For NN training, available data is divided into three parts: training set, test set and validation set. The parameters $W_{k,i}$ and b_i in Eq. (1) are calculated from training set. Training has been continued whenever error on the validation set starts to increase. Test set is used to evaluate neural network performance [15].

ANNs have been used for many chemical engineering applications such as steady state and dynamic process modeling, process identification, yield maximization, nonlinear control and fault detection and diagnosis [16].

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a classical data analysis method that provides a sequence of the best linear approximations to a given high-dimensional data set. It is one of the most popular techniques for dimensionality reduction [17]. There are different routes for selection of a subset (consist of m principal component) from variables of the main set such as scree graph and cumulative percentage of total variation that in both route, covariance matrix, its eigenvalues and eigenvectors must be calculate. [18].

The rule for choosing m (the number of principal components) is to select a cumulative percentage of total variation to which it is desired that the selected Principal Components (PCs) should contribute (for example 80% or 90%). The required number of PCs is then the smallest value of m for which this chosen percentage is exceeded. The obvious definition of percentage of variation accounted for the first m PCs is as follows

$$t_m = 100 \frac{\sum\limits_{k=1}^{m} \lambda_k}{\sum\limits_{k=1}^{p} \lambda_k} \tag{4}$$

Where λ_k denotes the eigenvalue (variance) of the k-th PC.

Choosing a cut-off t* somewhere between 70% and 90% and retaining m PCs, where m is the smallest integer for which $t_m > t^*$, provides a rule which in practice preserves in the first m PCs most of the information [18].

The scree graph, which was suggested by *Cattell* (1966) but which was already in common use, is even more subjective in its usual form, as it involves looking at a plot of eigenvalues (λ_k) against the number of components (k) the slope of lines joining the plotted points are steep to the left of k, and not steep to the right. This value of k is then taken to be the number of components m to be retained [18,19].

All steps for PCA method is presented in Fig. 1.

Among all modeling methods, ANN presents more acceptable and accurate models. For this reason, it is the objective of this work to develop a model using ANN to predict the performance of $Mn/Na_2WO_4/SiO_2$ catalyst in OCM reaction which is occurred in the fixed bed reactor at atmospheric pressure.

EXPERIMENTAL SECTION
Catalyst preparation

The $Mn/Na_2WO_4/SiO_2$ Catalyst was prepared by incipient wetness impregnation method. The syntheses of catalyst was carried out by impregnating of silica support with aqueous solutions containing 2 wt.% Manganese nitrate and 5 wt.% sodium tungstate. Then mixture was evaporated to dryness by using vacuum and sample was calsinated at 850 °C. The catalyst was then palletized, crushed and sieved to 20–25 mesh.
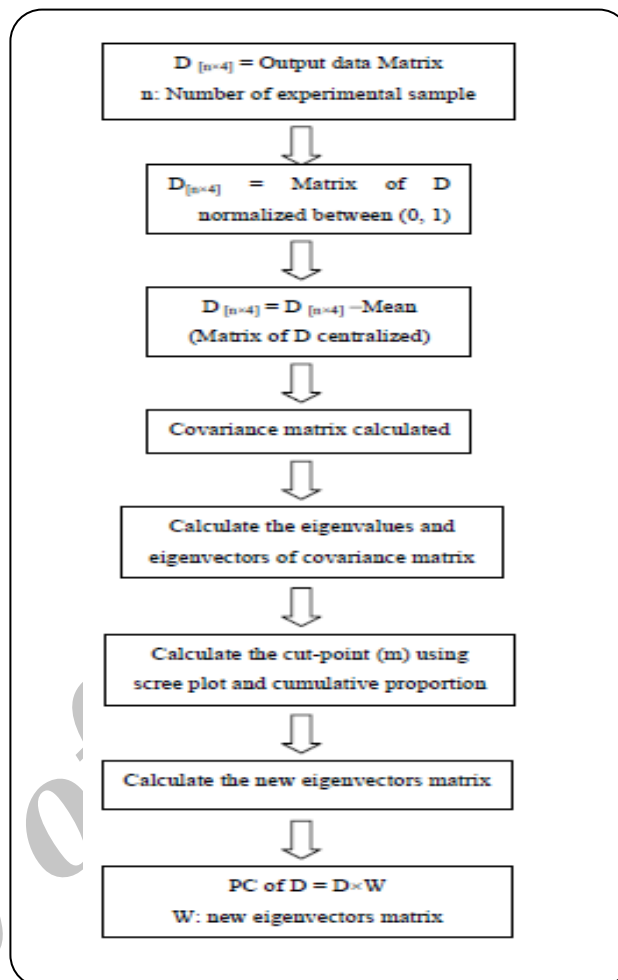

Fig. 1: flowchart showing the steps of PCA method.

Reactor system

The catalytic reaction was carried out in a tubular fixed bed quartz micro reactor with internal diameter 12mm shown in Fig. 2. The amount of 0.5 g catalyst was loaded in the quartz reactor filled with quartz granules in the test space of the reactor so far to minimize the contribution from any gas-phase reactions. The remaining space of the reactor below the catalyst bed was filled with quartz wool. The reactor was heated by an electrical furnace and the reaction temperature was measured with a K-type thermocouple placed in a well in the catalyst bed and controlled with T.I.C (Jumo Dicon 5.1) as well as mass flow controlled with electrical controller (Brooks 5580). Reactant gases was included CH_4 (99.996%), O_2 (99.99%) and N_2 (99.99% purity). Reaction was done at different conditions but in constant atmospheric pressure. The operating parameters and their ranges were shown in Table 1.

Table 1: Operating parameter and their ranges.

Parameters	Range
Temperature	750 - 850 °C
Gas Hourly Space Velocity (GHSV)	12000 - 24000 cm³/gh
Mol Percent of N_2	20 – 60
Ratio of CH_4/O_2	2 – 6
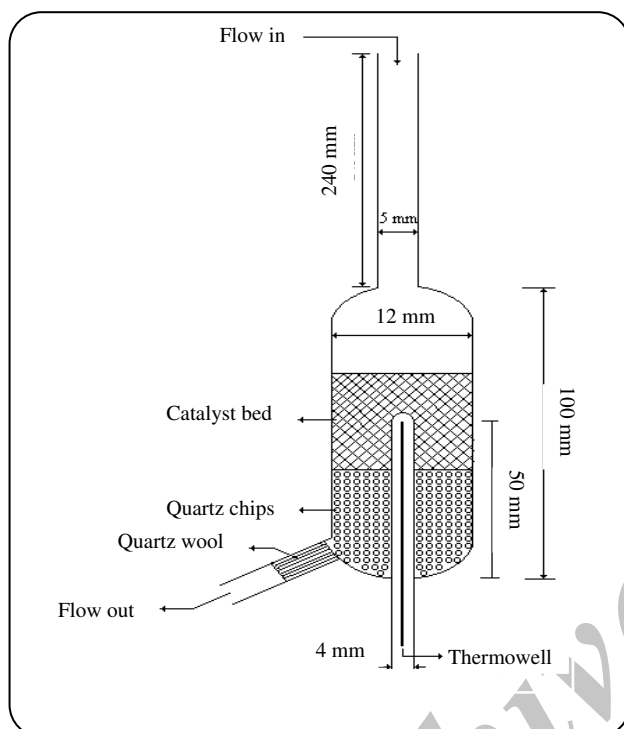
Fig. 2: Schematic drawing of reactor.

The reaction products were then analyzed with an on-line gas chromatography (Chrompack) equipped with FID with helium for identification of compounds except methane and TCD for methane, using a Porapak Q column for the separation of ethane, ethylene and carbon monoxide and a molecular sieve column for the separation of oxygen, nitrogen, methane and carbon monoxide (Fig. 3). Results obtained from products analysis represented aspects of catalytic performance, which include conversion of methane, C_2^+ products selectivity, yield of C_2^+ and C_2H_4/C_2H_6 ratio.

Some of the important parameters are defined as:

Methane conversion:

$$x_{CH_4}(\%) = \frac{\text{moles of } CH_4 \text{ converted}}{\text{moles of } CH_4 \text{ in feed}} \times 100$$

Selectivity of C_2^+ product:

$$S_{C_2^+}(\%) = \frac{\sum(2 \times \text{moles of } C_2)}{\text{moles of } CH_4 \text{ converted to all products}}$$

Yield of C_2^+ product:

$$Y_{C_2^+}(\%) = \text{Methane conversion} \times C_2^+ \text{ Selectivity}$$

Gas hourly space velocity:

$$GHSV = \frac{\text{feed flow rate}}{\text{catalyst weight}} \left(cm^3 g^{-1} h\right)$$

NEURAL NETWORK MODELING

In the present work the effect of operating conditions, such as temperature, Gas Hourly Space Velocity (GHSV), CH_4/O_2 ratio and diluents gas (mol% N_2) on ethylene production by Oxidative Coupling of Methane (OCM) in a fixed bed reactor at atmospheric pressure was studied over $Mn/Na_2WO_4/SiO_2$ catalyst and use of neural networks.

Training data generation

For modeling the OCM process data acquired from reactor test was used. The data must be pre-processed so that the ANN becomes able to effectively learn from it. Some statistical analysis is applied to the data, all values that appear to be scattered far away from the majority of values, are considered as outliers and were excluded from the data set. After removing the outliers the number of data records in the database results in 100 data points. The database was divided into three distinct subsets. A first subset of 60 patterns is used to train the model, as previously discussed. The second subset of 20 patterns is the validation set which is used to determine when to stop the training stage. The third data set of 20 patterns is used to test the ANN and to evaluate the efficiency of the ANN predictions. All data sets are mutually exclusive sets of vectors selected from the same measured data from the reactor tests.

Selection of ANN architecture

A feed-forward back-propagation network, widely known as MultiLayer Perceptron (MLP), was used for simulating the relations between process operating conditions (the above-mentioned parameters) and aspects of catalytic performance, which include conversion of methane,
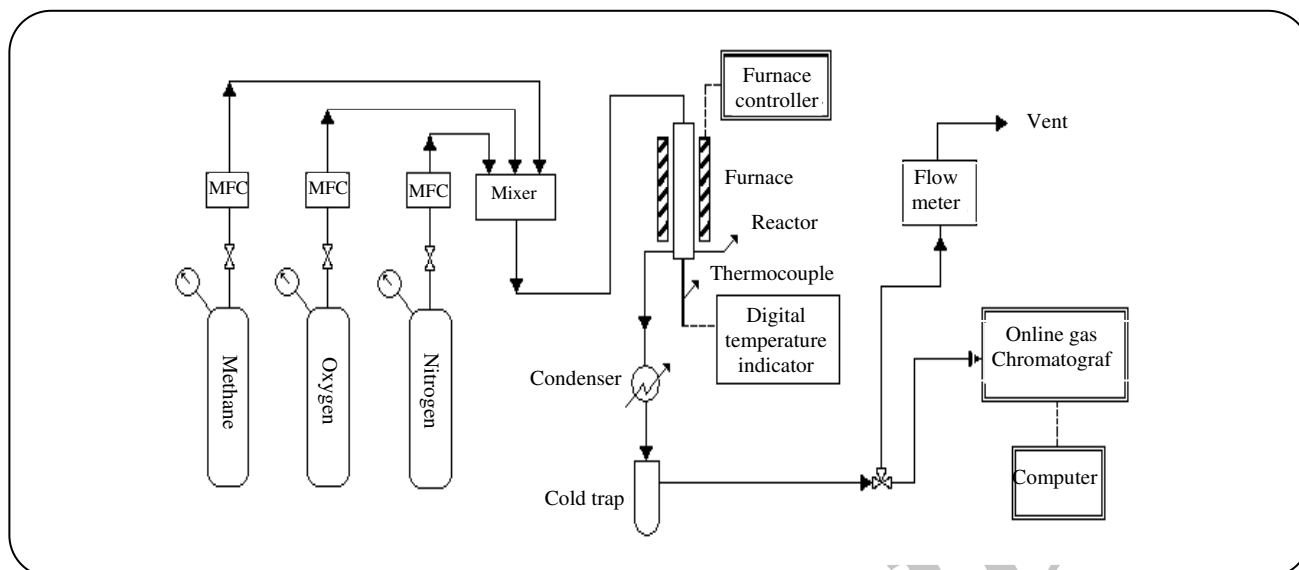
110

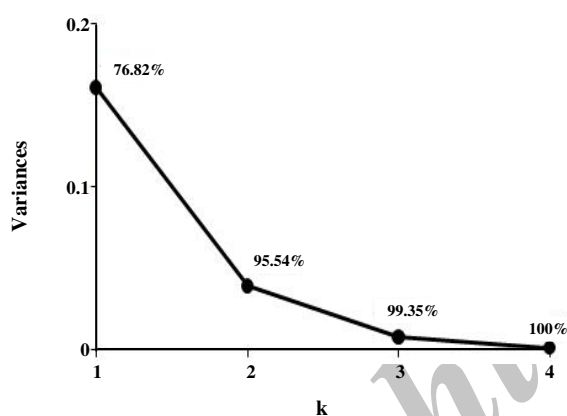**Fig. 3: Schematic diagram of experimental test rig system.**

Fig. 4: Scree graph for output data (value on the diagram denote the cumulative percentage).

$Cov(D) =$

$$\begin{bmatrix} 0.0602 & 0.0034 & 0.0588 & 0.0382 \\ 0.0034 & 0.0375 & 0.0254 & 0.0114 \\ 0.0588 & 0.0254 & 0.0739 & 0.0447 \\ 0.0382 & 0.0114 & 0.0447 & 0.0372 \end{bmatrix}$$

$Eigenvalues =$

$$\begin{bmatrix} 0.0014 & 0 & 0 & 0 \\ 0 & 0.0079 & 0 & 0 \\ 0 & 0 & 0.0391 & 0 \\ 0 & 0 & 0 & 0.1603 \end{bmatrix}$$

$Eigenvevtors =$

$$\begin{bmatrix} 0.0618 & 0.3489 & -0.4268 & 0.5673 \\ 0.3997 & 0.0743 & 0.8926 & 0.1947 \\ -0.6791 & 0.2674 & 0.1357 & 0.6700 \\ 0.0688 & -0.8951 & -0.0517 & 0.4375 \end{bmatrix}$$

C_2^+ products selectivity, yield of C_2^+ and C_2H_4/C_2H_6 ratio. In order to prevent network complexity and effective data input to network, Principal Component Analysis (PCA) method was used.

In this paper, both route of cumulative percentage of total variation and scree graph has been used to determine the principal components. Cumulative percentage of total variation has been considered 90%. Scree graph for our experiment is presented in Fig. 4. Also values of cumulative percentage of total variations have been reported in this figure.

Covariance matrix, the eigenvalues and eigenvectors of covariance matrix has been obtained as following:

According to Fig. 4, first two variables have been chosen as the principal components by preserve 95.54% of variation (m=2). Therefore with multiplying the output data matrix (D) in the third and fourth column of the eigenvectors matrix (W), the new matrix consists of two columns is obtained (New D) and dimensions of data and subsequently the numbers of neural network output were reduced from 4 to 2.

New D = D×W

$$W = \begin{bmatrix} -0.4268 & 0.5673 \\ 0.8926 & 0.1947 \\ 0.1357 & 0.6700 \\ -0.0517 & 0.4375 \end{bmatrix}$$
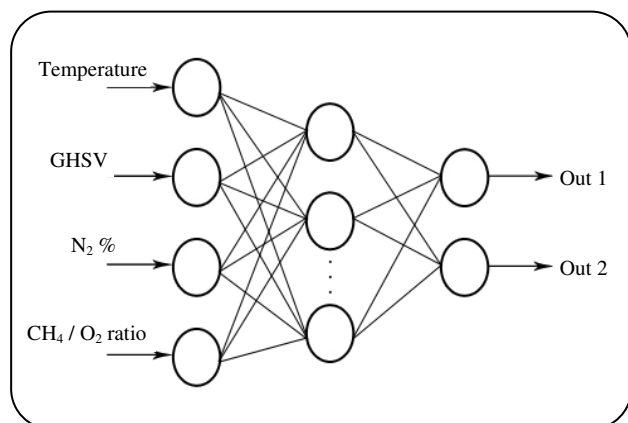
111

Fig. 5: Initial ANN structure.

Fig. 6: Final structure of networks.

Then, initial ANN structure consists of 4 nodes in input layer and 2 nodes in output layer as shown in Fig. 5.

Because of independent operation of each 2 nodes available at output layer, in order to avoid network complexity two specific neural networks were designed for each output as it shown in Fig. 6 (network 1 for out 1 and network 2 for out 2). Levenberg-Marquardt method was presented to train the networks. The Tangent-sigmoid function (Eq. (3)) was used as the activation function in both, the hidden and the output layers.

$$\tan\text{sig}(n) = \frac{e^n - e^{-n}}{e^n + e^{-n}} \tag{5}$$

For supplementation of ANN structure, the number of nodes in hidden layer must be determined and the number of optimum nodes is selected based on minimum of Mean Squared Error (MSE) on test data set.

MSE (error function) of different NN models for Network_out1 and Network_out2, against the number of neurons in the hidden layer, respectively are plotted as shown in Figs. 7 and 8.

According to aforesaid figures for first output, optimum network with 4-9-1 topology (one hidden layer which includes 9 neurons) and for second output, optimum network with 4-6-1 topology (one hidden layer including 6 neurons) was prepared.

Selection of training procedure and training precision

As mentioned in Training data generation section after removing the outliers the number of data records in the database results in 100 data points. To avoid overtraining, the database was divided into three distinct subsets: training set, test set and validation set.
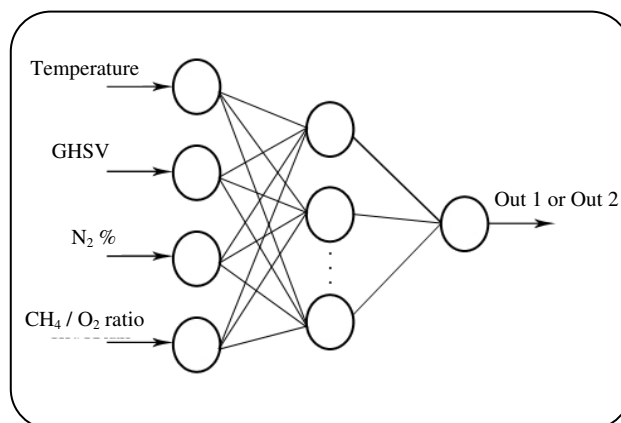
The error function (MSE) against the epoch for training, test and validation data set has been plotted in Figs. 9 and 10 for network1 and network 2, respectively.

The error function for training patterns usually decreases with the progress of training, while that for validation patterns decreases at the initial stage of training and then increases with training further on, as shown in Figs. 9 and 10. If the validation error increases with continued training, the training is terminated due to potential for overtraining. If the validation error remains the same for more than 10 successive epochs, it is assumed that the network has converged. It can be seen from above mentioned figures that the training stopped after only 16 iteration for network1 and 25 iterations for network2 because the validation error increased which means the training above this will impede the generalization capability of network.

RESULTS AND DISCUSSION

The obtained topologies for both neural networks must be evaluated. For this purpose the predicted values of network1 (neural networks output when test data set is used as input to the networks) for different time of measurement were plotted against the experimentally measured values for the corresponding times of measurement (acquired data) and were shown in Fig. 11. The correlation coefficient was also calculated for this network and found to be 0.977. Also a same work was done for network2 and the neural network output (predicted values) for this network is observed to fit well with the experimental data. This is also illustrated in Fig. 12. The fit is observed to be linear with correlation coefficient of 0.972.
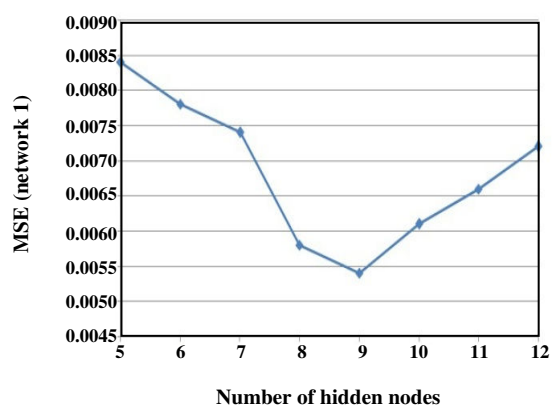
112

Fig. 7 Comparison of actual and ANN predicted output1 for Network1.
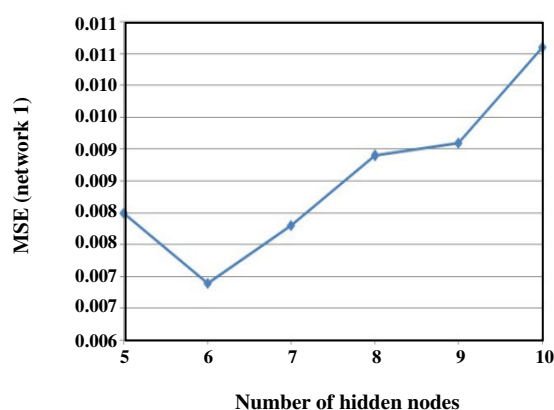

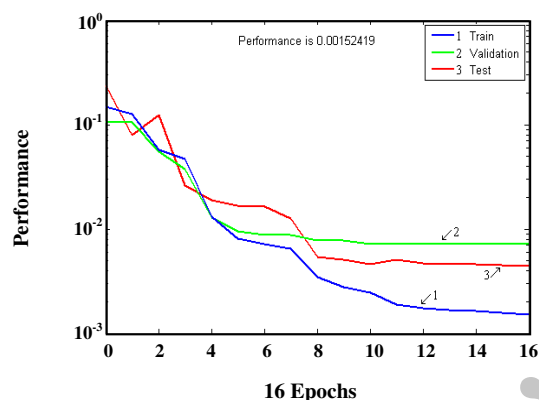*Fig. 8: Comparison of actual and ANN predicted output1 for Network2.*
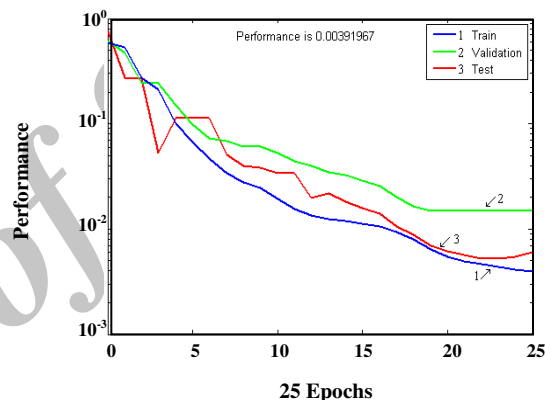

Fig. 9: Training procedure of network1.

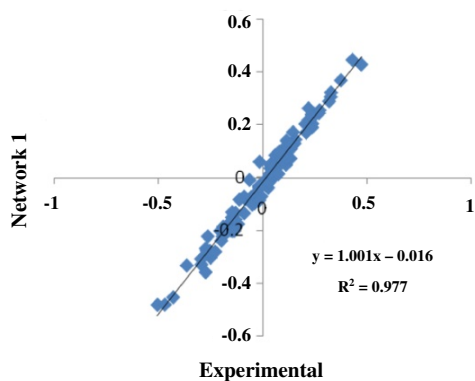*Fig. 10: Training procedure of network2.*

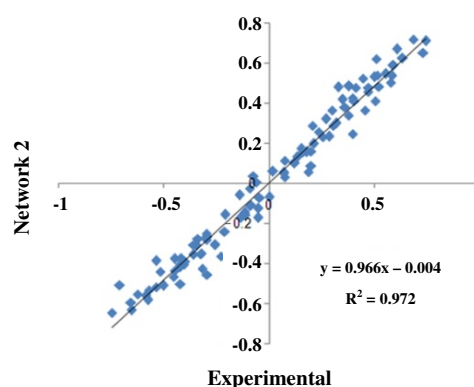*Fig. 11: Comparison of actual and ANN predicted output1 for Network1.*

Fig. 12: Comparison of actual and ANN predicted output2 for Network2.

Both two mentioned figures exhibit excellent agreement between the experimental and calculated data. So the obtained ANN models are adequate for modeling the OCM process over Mn/Na$_2$WO$_4$ /SiO$_2$ catalyst and predicting aspects of catalytic performance in different operating condition for this process.

113

CONCLUSIONS

In this paper, a feed-forward back-propagation network has been developed as a predicting model for OCM process over the Mn/Na$_2$WO$_4$/SiO$_2$ catalyst that occurs in the fixed bed reactor at atmospheric pressure. It is proved that the trained network could well simulate the relations between operation conditions and catalytic performances. Finally it is suggested to use this model in optimum operating conditions to find out the best performance of this catalyst.

REFERENCES

[1] Farsi A., Moradi A., Ghader S., Shadravan V., Manan Z.A., Kinetics Investigation of Direct Natural Gas Conversion by Oxidative Coupling of Methane, *J. Nat. Gas Sci. and Eng.*, **2**, p. 270 (2010).

[2] Zheng W., Cheng D., Zhu N., Chen F., Zhan X., Studies on the Structure and Catalytic Performance of S and P Promoted Na-W-Mn-Zr/SiO2 Catalyst for Oxidative Coupling of Methane, *J. Nat. Gas Chem.*, **19**, p. 15 (2010).

[3] Prodip K. Kundu, Yan Zhang, Ajay K. Ray, Modeling and Simulation of Simulated Countercurrent Moving Bed Chromatographic Reactor for Oxidative Coupling of Methane, *Chem. Eng. Sci.*, **64**, p. 5143 (2009).

[4] Talebizadeh A., Mortazavi Y., Khodadadi A.A., Comparative Study of the Two-Zone Fluidized-Bed Reactor and the Fluidized-Bed Reactor for Oxidative Coupling of Methane over Mn/Na$_2$WO$_4$/SiO$_2$ Catalyst, *Fuel Process. Technol.*, **90**, p. 1319 (2009).

[5] Nouralishahi A., Pahlavanzadeh H., Daryan J.T., Determination of Optimal Temperature Profile in an OCM Plug Flow Reactor for the Maximizing of Ethylene Production, *Fuel Process. Technol.*, **89**, p. 667 (2008).

[6] Malekzadeh A., Khodadadi A., Abedini M., Amini M., Bahramian A., Dalali A.K., Correlation of Electrical Properties and Performance of OCM MOx/Na$_2$WO$_4$/SiO$_2$ Catalysts, *Catal. Commun.*, **2**, p. 241 (2001).

[7] Mahmoodi S., Ehsani M.R., Ghoreishi S.M., Effect of Promoter in the Oxidative Cupling of Methane over Synthesized Mn/SiO$_2$ Nanocatalysts via Incipient Wetness Impregnation, *J. Ind. Eng. Chem.*, **16**, p. 923 (2010).

[8] Gardner M.W., Dorling S.R., Artificial Neural Network (the Multilayer Perceptron) - A Review of Applications in the Atmospheric Sciences, *Atmos. Environ.*, **32**, p. 2627 (1998).

[9] Himmelblau D.M., Applications of Artificial Neural Networks in Chemical Engineering, *Korean J. Chem. Eng.*, **17** (4), p. 373 (2000).

[10] Medina E.A., Paredes J.I.P., Artificial Neural Network Modeling Techniques Applied to the Hydrodesulfurization Process, *Math. Comput. Modell.*, **49**, p. 207 (2009).

[11] Mousavi M., Avami A., Modeling and Simulation of Water Softening by Nanofiltration Using Artificial Neural Network, *Iran. J. Chem. Chem. Eng.*, **25** (4), p. 37 (2006).

[12] Blasco J.A., Fueyo N., Larroya J.C., Dopazo C., Chen Y.J., A Single-Step Time-Integrator of a Methane-Air Chemical System Using Artificial Neural Networks, *Comput. Chem. Eng.*, **23**, p. 1127 (1999).

[13] Michalopoulos J., Papadokonstadakis S., Arampatzis G., Lygeros A., Modelling of an Industrial Fluid Catalytic Cracking Unit Using Neural Networks, *Trans I Chem E*, **79**, p. 137 (2001).

[14] Istadi I., Amin N.A.S., Modelling and Optimization of Catalytic-Dielectric Barrier Discharge Plasma Reactor for Methane and Carbon Dioxide Conversion Using Nybrid Artificial Neural Network-Genetic Algorithm Technique, *Chem. Eng. Sci.*, **62**, p. 6568 (2007).

[15] Priddy K.L., Keller P.E., "Artificial Neural Networks: An Introduction", The Society of Photo-Optical Instrumentation Engineers (SPIE), Washington (2005).

[16] Lahiri S.K., Ghanta K.C., Artificial Neural Network Model with the Parameter Tuning Assisted by a Differential Evolution Technique: the Study of the Hold Up of the Slurry Flow in a Pipeline, *Chem. Ind. Chem. Eng. Q.*, **15** (2), p. 103 (2009).

[17] Barshan E., Ghodsi A., Azimifar Z., Jahromi M.Z., Supervised Principal Component Analysis: Visualization, Classification and Regression on Subspaces and Submanifolds, *Pattern Recognit.*, **44**, p. 1357 (2011).

[18] Jolliffe I.T., "Principal Component Analysis", 2nd ed, Springer-Verlage, New York (2002).

[19] Lhabitant F.S., "Hedge Funds: Quantitative Insights", John Wily & Sons Ltd, Chichester, UK (2004).