From the Research Laboratories
В изследователските лаборатории

OPTIMIZATION OF ENGINE OIL FORMULATION USING RESPONSE SURFACE METHODOLOGY AND GENETIC ALGORITHM: A COMPARATIVE STUDY

[1]Behnaz Azmoon, [2]Abolfazl Semnani, [3]Ramin Jaberzadeh Ansari, [1]Hamid Shakoori Langeroodi, [4]Mahboube Shirani, [2]Shima Ghanavati Nasab
[1]Petrokimia Sepahan Company – Iran
[2]Shahrekord University – Iran
[3]Amirkabir University of Technology – Iran
[4]University of Jiroft – Iran

Abstract. Two potent mathematical and statistical methods of response surface methodology (RSM) and genetic algorithm (GA) based on artificial neural network (ANN) were employed for prediction and optimization of three-constituent synthetic engine oil. Polyalpha olephin-4 (PAO4) Hitec 5780 (HI 5780), and Hitec 11100 (HI 11100) were used as base oil and the additives of the engine oil, respectively. The models were applied for the percentage of oil constituents and viscosity at 40 ˚C (Vis at 40 ˚C), viscosity at 100 ˚C (Vis at 100 ˚C), Viscosity index (VI), flash point (FP) and Noack of the finished oil. The range of the viscosity at 40 ºC and 100 ºC were selected according to ISO viscosity grade for engine oil. The optimization includes maximization of FP and VI and minimization of Noack. The obtained results showed that ANN has higher potential and capability and more accuracy for prediction and optimization of the process.

Keywords: engine oil formulation; PAO_4; modeling; mixture design; artificial neural network; genetic algorithm

Introduction

The main function of lubricants is the reduction of wear and friction (Richardson, 2000). Automobile engine lubricating oil mainly contains base oil and some additives. Generally, the properties of lubricant are enhanced by engine oil additives. The significant roles of additives are to control deposition of lacquer and sludge, reduce corrosion, wear and oxidation, alter physical properties including pour point, flash point and modify chemical properties (Totten, 2006). The base oils are classified into five groups. Groups I-III are obtained from mineral sources, group IV oils are all polyalphaolephin (PAO) and group V oils are all other base stocks which are not included in group I-IV (Korcek & Jensen, 1976; Mensah-Brown, 2013; Murray et al., 1982). Synthetic base stocks which are used in motor oil formulation include esters, polyisobutylenes (PIB), polyalphaolephin, and polyalkalene glycols (PAG). PAO is one of the most widely used synthetic base oils which comparing to other synthetic oils and mineral oils, has the advantages such as better performance, higher viscosity index, better thermal oxidation stability and higher flash points. However, PAOs are more expensive than mineral oils (Mang & Dresel, 2007). The society of automotive engineers (SAE) classified motor oils according to the oil viscosity. Furthermore, American Petroleum Institute (API) classified motor oils according to the type of fuel into two general groups of Gasoline oil (with symbols (S)) and diesel oil (with symbol (C)) identified (Khalilian et al., 2016). RSM is a set of mathematical techniques which describes the relation between several independent variables (Tung & McMillan, 2004). The techniques were developed by Box and Willson (Myers et al., 2016) Mixture design is a special class of RSM techniques which is considered as the technology of the quality to reach the best of a product and response depending only on the amount of components present in the mixtures (Box & Wilson, 1951; Cafaggi et al., 2003). These components are dependent to each other. The sum of total parameters is exactly 100%. ANN is useful for identifying the relation between the components and the properties of the oil. These models have been applied to different systems (Brandvik & Daling, 1998; Dutta et al., 2010; Kumar & Porkodi, 2009). The main advantages of ANN models are their ability to identify both linear and nonlinear models, disregard of the nature of the original process. In addition, they are easy to implement and deliver good performance. After establishing the neural network-based model, the optimization is conducted using genetic algorithm. Genetic algorithm is a stochastic technique based on evolutionary algorithms. This algorithm search for all possible solutions, measure their fitness and reproduce new solutions in order to achieve global optima (Bezerra et al., 2008). The principal advantage of genetic algorithm is its independency on the error surface. Therefore, multi-dimensional, non-differential, non-continuous, and even non-parametrical problems can be solved using genetic algorithm.

In this work, the best formulation of the engine oil 5W40 SN containing of PAO_4 as base oil as well as, HI 5780 and Hi11100 as additives was obtained. Vis at 40 ˚C and at 100 ˚C, VI, F.P., and Noack were optimized to achieve the best formulation. RSM-mixture design and GA based on ANN are used for modeling and optimization purpose. Finally, the efficiencies of both models were compared by a set of experimental data provided by extreme vertices designs.

Experimental

Engine oil constituents

The following components were used: PAO4, HI 5780 and HI 11100 from ELA Company. The role of HI 1100 and HI 5780 is VI improver. In each experiment, the constituents were blended and stirred for 30 min.

Instrumental

VI, Vis at 40 ˚C, and Vis at 100 ˚C were performed by viscometer Anton bar model SVM 3000. Flash points were determined by flash point tester Herzog model HC852 and Noack was determined by Noack tester Tajhiz Gostar Apadana.

Test method

Different tests were executed according to ASTM D-445 for viscosity at 100 ˚C, 40 ˚C, ASTM D-2270 for viscosity index, ASTM D- 92 for flash point, ASTM D-97 for pour point and ASTM D-5800 for Noack.

Mathematical and statistical modeling

In this study, two different methods were applied to find the best formulation for the engine oil (5W40SN) production. The first method was implemented using Minitab17.0 software. The results were analyzed by executing following steps: (1) establishing the surface model: As mixture design and RSM were applied on the process the sequential F-test and other adequacy were obtained. The criterion for finding best fitting function are the evaluations of several statistical parameters such as the multiple correlation coefficient (R^2), the coefficient of variation (CV) and adjusted multiple correlation coefficient(adjusted R^2) (Goldberg, 1989). According to this information quadratic and linear model were selected. The selected model for a mixture is shown by Eq. (1) (Lu & Anderson-Cook 2012);

$$Y_n(x) = \sum_{1 \le i \le q}^{q} \beta_i X_i + \sum_{1 \le i \le j \le q}^{q} \beta_{ij} X_i X_j + \beta_{123} X_1 X_2 X_3 \tag{1}$$

where Y_n represents the response function of the experimental data (viscosities at 100 ˚C, 40 ˚C, VI, FP and Noack), x_1, x_2, x_3, x_4, x_5, and x_6 are independent variables and β_{ij} represents the coefficients of the interaction parameters; (2) analysis of variance (ANOVA). After analyzing each response, multi responses optimization was employed with numerical tools using Minitab 17.0.

In the second method, a neural network was firstly trained to use as a function for linking the components and the properties of the engine oil. Then using the same fitness function as the first method, genetic algorithm was applied to find the best formulation of the engine oil.

Procedure

The first method is categorized under RSM which includes approximation of responses using low-order polynomial with interaction and optimum level using methods like steepest ascent. The first step is to find a relation between the components and the properties using experimental data. Then, an appropriate criterion for measuring the quality of responses has to be established. Finally, the optimal mixture is found by using an optimization approach.

ANN is a calculation model that endeavors to simulate the functionalities and structure of biological neural networks. It is a parallel broadcast processing system combined of neurons (node) and connection (weights). In this study suggested the use of RSM and a genetic algorithm based on ANN to find optimizing of engine oil formulation (Orives et al., 2014).

Results and discussion

Experimental design by RSM

Experimental design was generated with the Minitab 17.0 at the interest concentration range. An extreme vertices design was based on a lower and upper bound on their component amount (Maran et al., 2014; Maran et al., 2013) (Eq. (2):

$$\text{a} \sum_j X_j = 1 \quad \text{and} \quad L_j \leq X_j \leq U_j \tag{2}$$

where Xj = component proportion; L_j = lower constraint; U_j = upper constraint.

According to ISO viscosity grade the following constraints were imposed on the components:

$$x_1 + x_2 + x_3 = 100, 57 \leq x_1 \leq 77, 13 \leq x_2 \leq 33 \text{ and } 9.5 \leq x_3 \leq 11.0$$

where x_1, x_2, and x_3 are the percentages of PAO_4, 5780, and HI 11100 respectively.

The responses for the dependent variables, Vis at 100˚C, Vis at 40˚C, VI, FP, and Noack were obtained by mixing the three components of the engine oil according to the suggested conditions (Table 1). A quadratic and linear model were selected and fitted to the obtained results (Mourabet et al., 2014).

Table 1. Extreme vertices experimental design and obtained responses for the dependent variables

Run	Suggested formulation			Obtained responses				
	PAO4	HI 5780	HI-11100	Vis at40	Vis at100	VI	FP	Noack
1	62.2	27.9	9.8	112.0	19.1	193.0	224.0	4.9
2	76.0	13.0	11.0	49.7	9.5	180.6	229.0	6.8
3	57.0	32.0	11.0	143.0	23.3	194.5	220.0	4.4
4	71.9	18.2	9.8	64.1	12.0	186.8	228.0	6.1
5	71.4	17.9	10.5	64.4	12.0	186.1	229.0	6.1
6	57.0	33.0	10.0	147.8	24.0	195.4	220.0	4.2

7	61.9	27.4	10.5	108.6	18.6	192.2	222.0	5.0
8	77.0	13.5	9.5	49.0	9.4	180.9	230.0	6.8
9	61.9	27.9	10.0	107.9	18.5	192.6	221.0	5.0
10	77.0	13.0	10.0	49.0	9.4	180.9	232.0	6.8
11	57.5	33.0	9.5	146.1	23.8	195.0	217.0	4.3
12	66.9	22.9	10.1	84.5	15.1	190.3	226.0	5.4
13	71.9	17.9	10.0	64.2	12.0	187.3	230.0	6.1

Modeling by RSM

The assessment of every dependent variable was carried out by using quadratic and linear model which contained quadratic and linear terms. The ANOVA results displayed the linear, interactive and quadratic relationship between the effects of independent variables on dependent variable. The significance of each term was assessed according to their corresponding p-values. The p-value reveals the significance of each variable. The p-value less than 0.05 shows that variable is significant and the p-value greater than 0.05 indicates that variable is insignificant (Tschoegl, 2012). The Fisher's F-test (F values) were measured and found to be high with very low probability value ($p < 0.05$), which displays a high degree of adequacy of model and also indicate that the process variable combinations were highly significant (Maran et al., 2014). Determination coefficient (R^2), correlation coefficient (R) and adjusted determination coefficient (R^2_a) and coefficient of variance (CV) were also determined to evaluate the suitability of the model. The determination coefficient (R^2) value of the regression model indicated that only the values of the total variations were not explained by the proposed model (Maran et al., 2013). The large value of the adjusted determination coefficient (R^2_a) indicated that the model was highly significant. The regression equations were obtained after the analysis of variance (ANOVA). These equations represented the level of Vis at 100˚C, Vis 40 ˚C, VI, FP, PP, and Noack as a function of PAO_4, HI 5780 and HI 11100. Eq. (3) was adjusted to fit experimental data and Y_1 represents the viscosity at 40 ˚C (cSt).

$$Y_1 = Vis at\, 40^{\circ}\mathrm{C} = +0.88378*X_1+12.62910*X_2+30.79787*X_3-0.14319*X_1*X_2-0.37837*X_1*X_3-0.43370*X_2*X_3 \quad (3)$$

The model F-value of 1069.8 implies that the model is significant. Values of $p < 0.05$ indicate that the model terms are significant according to ANOVA (Table 2) yielding an experimental R^2 of 99.87 % and the predicted R^2 of 99.43 % which are in reasonable agreement with the adjusted R^2 of 99.78 %.

Table 2. ANOVA for response of the Viscosity at 40 ˚C

Source	Sum of squares	DOF*	Mean square	F-value	p-value	
Model	17401.1	5	3480.2	1069.8	<0.05	significant
Residual	22.7	7	3.25			
Corrected Total	17423.8	12				

Eq.(4) was adjusted to the experimental data to show the model, where Y_2 is viscosity at 100 ˚C (cSt).

$$Y_2 = Vis at 100^{\circ}C = +0.08986*X_1 + 1.62015*X_2 + 2.18304*X_3 - 0.016322X_1*X_2 - 0.025479*X_1*X_3 - 0.034224*X_2*X_3 \quad (4)$$

The proposed model was analyzed by using analysis of variance (ANOVA). As the results indicated in Table 3, the F-value of the model equals to 1182.9 which implies that the model is significant and the values of $p < 0.05$ indicate the model terms are significant. The predicted (R^2) was 99.47 % and the adjusted R^2 was 99.80 % as shown in Table 3.

Table 3. ANOVA for response of the Viscosity at 100 ˚C

Source	Sum of squares	DOF*	Mean square	F-value	p-value	
Model	378.3	5	75.6	1182.98	<0.05	significant
Residual	0.45	7	0.064			
Corrected Total	378.2	12				

VI is a measure of the change of viscosity with variations in temperature. The higher VI value shows that oil viscosity changes with temperature. VI is an important factor for quality of engine oils (Maran et al., 2014). Table 4 shows ANOVA response for VI.

Table 4. ANOVA for response of the VI

Source	Sum of squares	DOF	Mean square	F-value	p-value	
Model	363.2	5	72.6	197.5	<0.05	significant
Residual	2.58	7	0.37			
Corrected Total	365.8	12				

Consequently, Eq. (5) was presented,

$$Y_3 = VI = +0.65384*X_1+0.34048*X_2-71.69262*X_3+0.021404*X_1*X_2+0.91217*X_1*X_3+0.91935*X_2*X_3 \quad (5)$$

where Y_4 represents the dependent variable of VI. Moreover, x_1 and x_2 influence the response positively and x_3 negatively. The F-value of the model equals 197.5 which imply that the model is significant and the values of $p < 0.05$ indicate the model terms are significant; the predicted R_2 of 96.22% is in reasonable agreement with the adjusted R_2 of 98.79%.

The flash point is a measure of the tendency of oil to form a flammable mixture with air under controlled laboratory conditions (Aleme & Barbeira, 2012). The minimum FP, according to the ISO 5W 40 SN, is 200 °C in engine oil. Based on the obtained results, the predictive Eq. (6) was produced and (Y_4) represents the dependent variable of flash point.

$$Y_4 = F.P = 2.4637x_1 + 2.5037x_2 + 0.3694x_3 \quad (6)$$

As analysis of variance shows in Table 5, this model is significant. F-value 61.7 and P-value (less than 0.05) of the model imply that the model is significant. The predicted R^2 of 84.4% is in reasonable agreement with the adjusted R^2 of 91.01%. The Noack test is employed to measure the evaporative loss of an oil at 250 °C after 1 h (Waddoups et al., 2001).

Table 5. ANOVA for response of the F.P

Source	Sum of squares	DOF	Mean square	F-value	p-value	
Model	259.3	2	129.6	61.7	0.4806	significant
Residual	21.0	10	2.1			
Corrected Total	280.3	12				

The Noack is measured traditionally by ASTM D5800. The predictive Eq. (7) represents the response of the dependent variable Y_5 (Noack). According to the obtained ANOVA results in Table 6, the model F-value is 1161.7 and the p-value is <0.05 which reveal that the model is significant. The predicted (R^2) was 99.35 % and the adjusted R^2 was 99.49 %.

$$Y_5 = Noack = 0.0858x_1 - 0.0399x_2 + 0.0686x_3 \quad (7)$$

Table 6. ANOVA for response of Noack

Source	Sum of squares	DOF	Mean square	F-value	p-value	
Model	11.3	2	5.65	1161.7	<0.05	significant
Residual	0.049	10	4.863E-003			
Corrected Total	11.35	12				

Optimization by RSM

According to the responses (Y_1, Y_2, Y_3, Y_4, and Y_5) of the obtained multiresponse optimization were evident. In the case of large number of responses, desirability function is the most popular to be used. In 1980 Suich and Derringer found the desirability function to find the best value for all response (Kim et al., 2002; Myers et al., 2016).

$$D = (d_1 \times d_2 \times \ldots \times d_n)^{\frac{1}{n}} = \left(\prod_{i=1}^{n} d_i\right)^{\frac{1}{n}} \tag{8}$$

where n is the number of responses. Desirability always possess values between 0 and 1, where di=0 for an undesirable response, and d_i=1 a thoroughly desirable response (Das et al., 2009).

Final results of RSM

In order to optimize the process, FP and VI were maximized. Moreover, for Vis at 100 ˚C and 40 ˚C the restriction of (12.6-16) cSt and Vis (60-100) were set respectively for engine oil. Noack was minimized. The optimum formulation displayed a Viscosity at 100 ˚C, 40 ˚C, VI, FP, Noack and PP of 85.0 cSt, 15.0 cSt, 190.6, 225 ˚C, 5.0, respectively when the formulation was composed of 66.9% PAO_4, 22.9% HI 5780, and 10.1% HI 11100. The obtained optimum conditions were used to validate the model prediction. Validation was performed in triplicate and the average values for Viscosity at 100 ˚C, 40 ˚C , VI, FP , Noack were 84.5 cSt, 15.1 cSt, 190.3, 224 ˚C , 5.4, and -39 , respectively.

Modeling by ANN: prediction by ANN

In this work, a feed-forward neural network (FF ANN) was employed to establish the relation between components and outputs. The general structure of FF ANN is shown in Fig. 1.

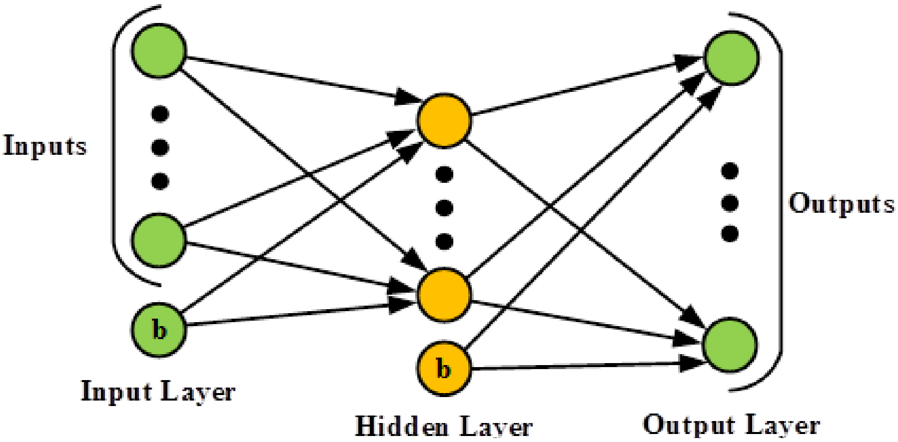

Fig. 1. Structure of feed-forward network

The first step is to determine the number of neurons in the hidden layer. There are several methods for calculating the number of neurons (Gómez et al., 2009). However, there is not any determinate method to use. Considering the available methods, we selected the number of neurons using trial and error. The next step is to select the learning algorithm, in order to reach the best weights for the neural network. The choice of a suitable learning method is a crucial part of modeling by ANN, because the best training of network is achieved by minimizing the error function. The back-propagation algorithm was used for training the understudy feed-forward ANN. To this aim, the Levenberg– Marquardt optimization algorithm is employed for its training. The last step is to validate and to verify the prediction model on the basis of error function. In this work, the mean square error (MSE) was employed as the error function. Also, the correlation coefficient (r) was selected as a parameter to show the predictive ability of the network. The same experimental data (Table 1) were used for training the artificial neural network. The data were randomly dispersed into three groups, 70% in the training set, 15% in the test set and 15% in the validation set. After repetition of the trails, a network with 5 hidden neurons showed the best performance. The result is shown in Fig. 2. The MSE value was obtained to be 0.84 which is shown in Fig. 3. The weights which are obtained from the final trained network are shown in Tables 7 and 8. The ANN prediction results are indicated in Table 9.

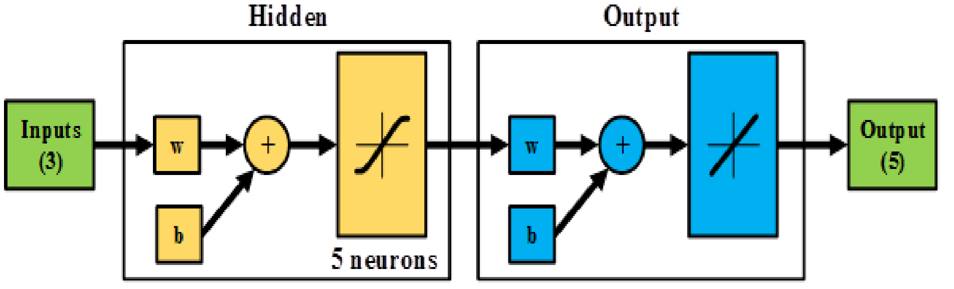

Fig. 2. Optimal architecture of ANN model

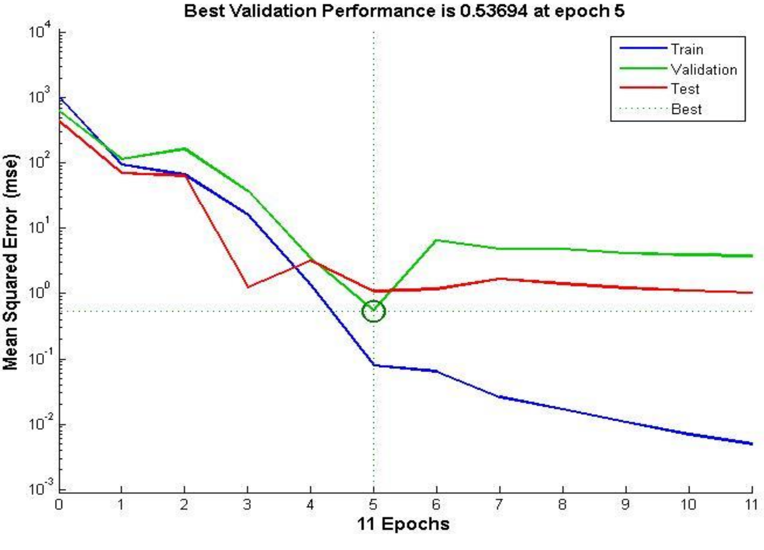

Fig. 3. Evolution of network performance (MSE) during training phase using Levenberg-Marquardt

Table 7. Optimal values of the first layer of the network weights

	HN* 1	HN 2	HN 3	HN 4	HN 5
Input1	-0.0545	1.1646	1.3667	0.4580	1.0161
Input2	2.3945	-1.7114	-0.9790	-1.7227	-1.5985
Input3	1.3787	0.1334	-2.7552	0.3063	-2.3289
Bias	-1.2436	-1.3973	1.8797	1.4195	1.7555

* Hidden Neuron

Table 8. Optimal values of the output layer of the network weights

	Output1	Output2	Output3	Output4	Output5
HN 1	-0.1467	-0.0968	-0.1545	-2.5956	-0.8856
HN 2	-0.3489	-0.3887	-0.7107	0.1757	0.5220
HN 3	-0.1095	-0.2013	0.0992	1.6376	0.2603
HN 4	-0.7940	-0.7375	-0.3786	-0.2704	0.1387
HN 5	-0.1508	-0.0074	-0.2504	-3.5385	-0.8456
Bias	0.2211	0.1997	0.0647	0.1590	0.0846

Table 9. Input data and predicted response which have been obtained by ANN

Run	Input data suggested by experimental design			Output data obtained by ANN				
	PAO4	HI 5780	HI-11100	Vis at40	Vis at100	VI	FP	Noack
1	62.2	27.9	9.8	112.1	19.0	192.9	223.8	5.0
2	76.0	13.0	11.0	49.8	9.6	180.9	229.1	6.7
3	57.0	32.0	11.0	143.0	23.3	194.5	220.0	4.5
4	71.9	18.2	9.8	66.5	12.3	186.1	228.7	6.1
5	71.4	17.9	10.5	64.4	11.9	185.4	228.3	6.1
6	57.0	33.0	10.0	147.6	23.7	195.4	220.8	4.3
7	61.9	27.4	10.5	109.2	18.6	192.3	221.9	4.9
8	77.0	13.5	9.5	49.9	9.5	181.2	230.0	6.7
9	61.9	27.9	10.0	108.3	18.5	192.6	220.9	4.8
10	77.0	13.0	10.0	49.1	9.4	181.0	230.0	6.8
11	57.5	33.0	9.5	146.3	23.8	195.1	217.3	4.3
12	66.9	22.9	10.1	83.9	15.0	190.6	226.3	5.4
13	71.9	17.9	10.0	64.8	12.0	185.6	228.8	6.2

Optimization by genetic algorithm

In this work, the genetic algorithm was used to obtain an optimal condition to minimize a number of preliminary experiments. GA-ANN was employed to find the properties of different mixtures. The GA with the following properties is used to determine the controllers' parameters: chromosome population = 40; number of generation = 100; crossover fraction = 0.8; elite count = 5 %; migration fraction = 0.2; migration interval = 20.

The optimum formulation using the genetic algorithm displayed a Vis at 100 ˚C, Vis at 40 ˚C, VI, FP, Noack and PP of 82.1 cSt, 14.6 cSt, 188.4, 225 ˚C , 5.6, and -39 respectively when the formulation was composed of 67.0 .% PAO4, 22.1% 5780, and 10.8% HI 11100. Validation was performed in triplicate and the average values for Viscosity at 100 ˚C, 40 ˚C , VI, FP and Noack were 82.1 cSt, 14.6 cSt, 188.5, 226 ˚C ,5.6, and -39 , respectively. The results of optimization are similar to the experimental data.

Therefore, the combination of ANN and GA (i.e. GA-ANN) find the solution with minimum error from experimental data. Also, the mean fitness value in each generation is shown in Fig 4.

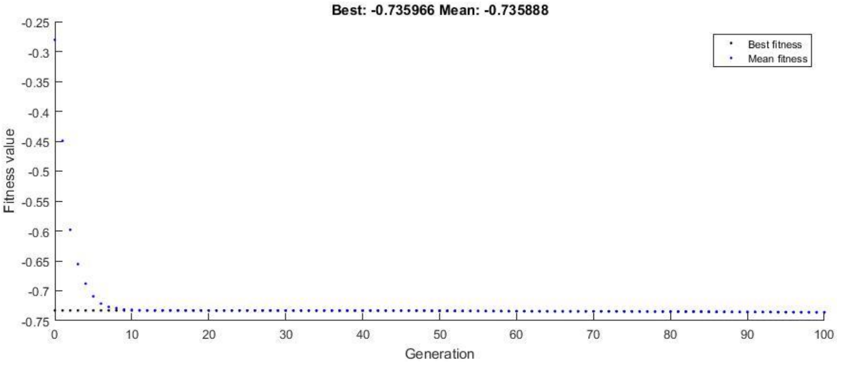

Fig. 4. Mean fitness value for generations

Comparison mixture design and ANN

The predicted data obtained from RSM-mixture design and ANN and the experimental data were compared. To this aim, a new set of 4 experiments was considered in the acceptable range, which does not belong to the training data set (Table 10). Table 11 shows the predicted data of RSM-mixture design and ANN.

Table 10. Experimental range for predict

Run	PAO4	HI 5780	HI11100
A	60.0000	29.0000	11.0000
B	64.0000	26.5000	9.5000
C	74.0000	15.0000	11.0000
D	69.0000	20.5000	10.5000

Table 11. Comparison of modeling powers of RSM and ANN

Property		Run			
		A	B	C	D
Vis at 40 °C	Actual values	116.0	105.0	53.9	75.5
	Prediction of ANN	118.5	106.3	54.1	76.4
	Error of ANN	**-2.5**	**-1.3**	**-0.2**	**-0.9**
	Prediction of RSM	120.8	101.7	55.1	73.2
	Error of RSM	**-4.8**	**3.3**	**-1.2**	**2.3**
Vis at 100 °C	Actual values	19.6	18.0	10.2	13.8
	Prediction of ANN	19.9	18.2	10.3	13.9
	Error of ANN	**-0.3**	**-0.2**	**-0.1**	**-0.1**
	Prediction of RSM	20.2	17.6	10.5	13.4
	Error of RSM	**-0.6**	**0.4**	**-0.3**	**0.4**
VI	Actual values	192.0	192.7	182.0	189.2
	Prediction of ANN	192.8	192.6	181.8	188.9
	Error of ANN	**-0.8**	**0.1**	**0.2**	**0.3**
	Prediction of RSM	193.0	192.1	182.8	188.3
	Error of RSM	**-1.0**	**0.6**	**-0.8**	**0.9**
FP	Actual values	220.0	227.0	222.0	224.0
	Prediction of ANN	219.5	227.9	228.1	226.8
	Error of ANN	**0.5**	**-0.9**	**-6.1**	**-2.8**
	Prediction of RSM	221.4	223.1	229.9	226.6
	Error of RSM	**-1.4**	**3.9**	**-7.9**	**-2.6**
Noack	Actual values	4.9	5.4	6.5	5.6
	Prediction of ANN	4.7	5.3	6.6	5.7
	Error of ANN	**0.2**	**0.1**	**-0.1**	**-0.1**
	Prediction of RSM	4.7	5.0	6.5	5.8
	Error of RSM	**0.2**	**0.4**	**0**	**-0.2**

Obviously ANN was more powerful in prediction of the process. In order to evaluate the precision and accuracy of both applied models, the mean squared error (MSE) was calculated for RSM and ANN. MSE can be defined by Eq. (9):

$$MSE = \frac{1}{N}\sum_{1}^{N}(y_{i,exp} - y_{i,pred})^2 \tag{9}$$

where MSE is the mean squared error; N is the number of experimental data points; yi,exp is the experimental value of training sample i and yi,pred is the predicted value from the neural network for training sample i.

The MSE was 130.82 and 55.9 for RSM and ANN respectively. Moreover, a regression analysis for predicted data by ANN and experimental data was done. As the results show in Fig. 5, all data scatter around the 45º line which is an indicative for high suitability of ANN prediction for the process. According to the obtained results, ANN had higher predictive accuracy than RSM-mixture design even with limited number of experiments.

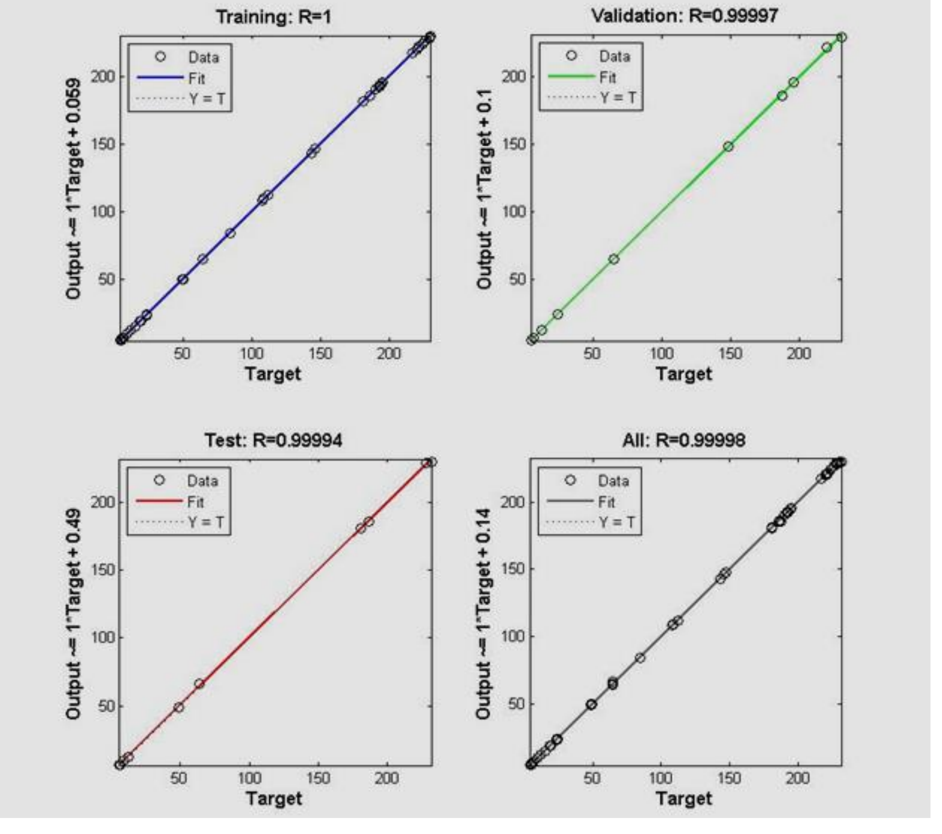

Fig. 5. Network model with training, validation, test and all predictions

Comparison of optimization by RSM and GA-ANN

RSM and GA ANN were used to predict and optimize the proposed process. Table 12 and Table 13 show the comparison of the results obtained by two methods for multi-objective which indicate the suitability of both RSM and GA ANN for this study.

Table 12. Comparison of optimization powers of GA-ANN and RSM

Method	Component	Concentration
	PAO4	69.3
Optimization by GA-ANN	HI5780	21.1
	HI11100	9.5
	PAO4	67.0
Optimization by RSM	HI5780	22.0
	HI11100	111.0

Table 13. Comparison of result by GA-ANN and RSM

Property		Run
Vis at 40ºC	Result by GA-ANN	80.5
	Result by RSM	80.8
Vis at 100ºC	Result by GA-ANN	14.5
	Result by RSM	14.5
VI	Result by GA-ANN	189.8
	Result by RSM	188.9
FP	Result by GA-ANN	228.0
	Result by RSM	225.0
Noack	Result by GA-ANN	5.6
	Result by RSM	5.6

Conclusions

In this paper, RSM Mixture design with six independent variables and ANN with five neurons in hidden layer were used for formulation of engine oil. RSM and ANN models were successfully employed for prediction and optimization of the process. The comparison of the obtained results from two statistical methods of RSM- mixture design and ANN confirmed that ANN model had better prediction capability of target values. In this study, the genetic algorithm was used as the optimization algorithm. The results indicate that there was no significant difference between the experimental and predicted data. Achieving optimized components by application of genetic algorithm leads to an economical formulation due to the lower consumption of expensive additives HI 5780 (preserving the properties of oil at their optimal values based on ISO-grade for engine oil).

Acknowledgments: The financial support of this project by Shahrekord University is appreciated. The authors were also partially supported by the Center of Excellence for Mathematics, Shahrekord University.

REFERENCES

Aleme, H.G. & Barbeira, P.J.S. (2012). Determination of flash point and cetane index in diesel using distillation curves and multivariate calibration. *Fuel, 102*, 129 – 134.

Bezerra, M.A., Santelli, R.E., Oliveira, E.P,. Villar, L.S. & Escaleira, L.A. (2008). Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta, 76,* 965 – 977.

Box, G.E.P. & Wilson, K.B. (1951). On the experimental attainment of optimum conditions. *J. Roy. Stat. Soc. B, 13*, 1 – 45.

Brandvik, P.J. & Daling, P.S. (1998). Optimisation of oil spill dispersant comp osition by mixture design and response surface methods. *Chemometrics & Intelligent Laboratory Systems, 42,* 63 – 72.

Cafaggi, S., Leardi, R., Parodi, B., Caviglioli, G. & Bignardi, G. (2003). An example of application of a mixture design with constraints to a pharmaceutical formulation*. Chemometrics & Intelligent Laboratory Systems, 65*, 139 – 147.

Das, P., Mukherjee, S., Ganguly, S., Bhattacharyay, B.K. & Datta, S. (2009). Genetic algorithm based optimization for multi-physical properties of HSLA steel through hybridization of neural network and desirability function. *Comp. Materials Sci., 45,* 104 – 110.

Dutta, S., Parsons, S.A., Bhattacharjee, C., Bandhyopadhyay, S. & Datta, S. (2010). Development of an artificial neural network model for adsorption and photocatalysis of reactive dye on TiO_2 surface. *Expert Systems with Applications, 37*, 8634 – 8638.

Goldberg, D.E. (1989). *Genetic algorithms in search, optimization, and machine learning.* Reading: Addison-Wesley.

Gómez, I., Franco, L. & Jerez, J.M. (2009). Neural network architecture selection: can function complexity help. *Neural Processing Lett., 30*, 71 – 87.

Khalilian, H., Semnani, A., Haddadi, H. & Nekoeinia, M. (2016). Multiresponse optimization of a hydraulic oil formulation by mixture design and response surface methods. *J. Tribology, 138*, art. no. 021801.

Kim, D., Rhee, S. & Park, H. (2002). Modelling and optimization of a GMA welding process by genetic algorithm and response surface methodology. *Int. J. Production Res., 40,* 1699 – 1711.

Korcek, S. & Jensen, R.K. (1976). Relation between base oil composition and oxidation stability at increased temperatures. *ASLE Transactions, 19*, 83 – 94.

Kumar, K.V. & Porkodi, K. (2009). Modelling the solid–liquid adsorption processes using artificial neural networks trained by pseudo second order kinetics. *Chem. Eng. J., 148*, 20 – 25.

Lu, L. & Anderson-Cook, C.M. (2012). Rethinking the optimal response surface design for a first-order model with two-factor interactions, when protecting against curvature. *Quality Engineering, 24*, 404 – 422.

Mang, T. & Dresel, W (2007). *Lubricants and lubrication.* Weinheim: John Wiley & Sons.

Maran, J.P., Manikandan, S. & Mekala, V. (2013). Modeling and optimization of betalain extraction from Opuntia ficus-indica using Box–Behnken design with desirability function. *Ind. Crops & Products. 49*, 304 – 311.

Maran, J.P., Priya, B. & Manikandan, S. (2014). Modeling and optimization of supercritical fluid extraction of anthocyanin and phenolic compounds from Syzygium cumini fruit pulp. *J. Food Sci. & Tech., 51*, 1938 – 1946.

Mensah-Brown, H. (2013). Optimization of the production of lubricating oil from rerefined used lubricating oil using response surface methodology. *ARPN J. Eng. & Appl. Sci., 8,* 749 – 756.

Mourabet, M., El Rhilassi, A., Bennani-Ziatni, M. & Taitai, A. (2014). Comparative study of artificial neural network and response surface methodology for modelling and optimization the adsorption capacity of fluoride onto apatitic tricalcium phosphate. *Universal J. Appl. Math., 2,* 84 – 91.

Murray, D., Clarke, C., MacAlpine, G. & Wright, P. (1982). The effect of base stock composition on lubricant oxidation performance. *SAE Technical Paper;* art. no. 821236.

Myers, R.H., Montgomery, D.C. & Anderson-Cook, C.M. (2016). *Response surface methodology: process and product optimization using designed experiments*. Hoboken: John Wiley & Sons.

Orives, J.R., Galvan, D., Coppo, R.L., Rodrigues, C.H.F., Angilelli, K.G. & Borsato, D. (2014). Multiresponse optimisation on biodiesel obtained through a ternary mixture of vegetable *oil and animal fat: simplex-*

centroid mixture design application. Energy Conversion & Management, 79, 398 – 404.

Richardson, D.E. (2000). Review of power cylinder friction for diesel engines. *J. Eng. Gas Turbines & Power, 122*, 506 – 519.

Totten, G.E. (2006). *Handbook of lubrication and tribology: application and maintenance*. Boca Raton: CRC Press.

Tschoegl, N.W. (2012). *The phenomenological theory of linear viscoelastic behavior: an introduction*. Berlin: Springer.

Tung, S.C. & McMillan, M.L. (2004). Automotive tribology overview of current advances and challenges for the future. *Tribology Inter., 37*, 517 – 536.

Waddoups, M., Hartley, R.J. & Miyoshi, T. (2001). Molybdenum-free low volatility lubricating oil composition. *US Patents* 6333298 B1.

✉ **Dr. A. Semnani (corresponding author)**
Department of Chemistry
Shahrekord University
Shahrekord, P.O. Box 115, Iran
E-mail: a.semnani1341@gmail.com