# Corrosion inhibition of mild steel by some sulfur containing compounds: Artificial neural network modeling

## K.F. Khaled[1,2]*, N. S. Abdel-Shafi[2,4]

[1]*Materials and Corrosion Laboratory, Chemistry Department, Faculty of Science, Taif University, Saudi Arabia*
[2]*Electrochemistry Research Laboratory, Chemistry Department, Faculty of Education, Ain Shams Univ., Roxy, Cairo, Egypt*
[3]*Chemistry Department, Faculty of Science, Hail University, Saudi Arabia*

## Abstract

Advances in corrosion inhibition studies revealed that the relationship between structure of the inhibitor molecules and their inhibition efficiencies is essentially a non-linear and highly complex process particularly out of reach of classical statistical modeling techniques. The non-linearity of the corrosion process forced us to look for other solutions to track this complex process. Application of Artificial Neural Networks (ANNs) may provide better and more comprehensive results. In this work ANNs were used to predict the inhibition efficiencies of ten sulfur containing compounds on the corrosion of mild steel in hydrochloric acid solutions. A (6-3-1) network was adopted to predict the corrosion inhibition efficiencies of the sulfur containing compounds. The descriptors (inputs) were obtained using quantum chemical calculations. Highest occupied molecular orbitals, $E_{HOMO}$, lowest unoccupied molecular orbitals, $E_{LUMO}$, energy gap, ($E_{LUMO}$-$E_{HOMO}$), molecular area, molecular volume and total dipole moments were selected as the ANN inputs to predict the corrosion inhibition efficiencies (output).

*Keywords***:** Artificial Neural network; Corrosion inhibitor; Quantum chemical descriptors

## 1- Introduction

The reliable prediction of corrosion behavior is a fundamental requirement for the effective control of corrosion. At first sight, this is relatively easy. There is an extensive corrosion literature, and we can simply look up the relevant information. Unfortunately, real world corrosion never seems to involve quite the same conditions as have been tested before and there is also the difficult question of the inherent variability of the corrosion process[1].

Several attempts to use neural network modeling in the corrosion science have been made in the last three decades. An early published attempt to apply a neural network to a corrosion problem was that of Smets and Bogaerts [2]. They developed a series of neural networks to predict the SCC of type 304 stainless steel in near-neutral solutions as a function of chloride content, oxygen content and temperature [3]. They conclude that the neural networks are adept at many classification tasks such as data analysis, which require the ability to match large amounts of input information simultaneously and generate categorical or generalized output. They successfully applied neural networks for the prediction of the SCC risk of austenitic stainless steels. They firstly developed a neural network describing the SCC risk of steel type 304 as a function of the temperature and the chloride concentration. Secondly, they used the neural networks to estimate the SCC risk of the same material as a function of the oxygen and chloride concentration. Finally, they combined the effect of the three parameters (temperature, chloride concentration, and oxygen content) on the occurrence of SCC for austenitic chromium-nickel stainless steels in high-temperature water.

Bassam et al[4] proposed a neural network model that can predict the corrosion type in a steel pipeline as a function of four input parameters (inhibitor concentration, experimental time, real experimental impedance component, and imaginary impedance experimental component) and exhibit a good ability for generalization. The neural network model was successfully trained with experimental database and validated with a fresh database (in the specified range of key operating conditions)

Urquidi-Macdonald et al[5] developed a neural network model for predicting the number and depth of pits in heat exchangers. No information about the network size or number of training point was given, even though there appear

to be rather little. Hence there must be some doubt about the consistency of the predictions obtained. On the other hand the predicted evolution of pit depth and number does appear plausible[3].

Smets and Bogaerts[6] used "artificial intelligence" approach (i.e. combination of different knowledge representations, such as neural networks and expert systems) to provide better (phenomenological) models of the SCC problem in boiling water reactor environments. Such models can help engineers and utilities to predict and manage the potential risks of SCC in boiling water nuclear reactors. Silverman and Rosen[7] jointed artificial neural networks with an expert system in order to foretell the type of corrosion from polarization curves. Inputs to the networks included the passive current density, the pitting potential and the re-passivation potential, while outputs were the risks of crevice, pitting and general corrosion. Two approaches were used: independent networks for each type of corrosion, and a single combined network producing all three outputs. The expert system was used to understand the outputs produced by the two approaches.

This combination of neural network with an expert system 'supervisor' offers interesting possibilities; although in this particular case it seems probable that the available training set (87 examples) was not really adequate to describe the dependence of the type of corrosion on the seven input variables [3]. Hence a larger training set could have lead to neural networks that were sufficient by themselves[3]. Again, Silverman [8] also applied neural network methods to the prediction of service behaviour of polymeric linings from experimental data. Inputs to the network were parameters obtained from a short-term solvent uptake test, while the output was a pass-fail description of service behaviour[3]. Taking into account the inherent complexity of this problem the results obtained seem hopeful. For example, the neural network was able to produce a correct prediction for three out of four examples of a lining material that was not included in the training data [3].

Khaled et al. [9, 10] used artificial neural network model to predict the inhibition efficiency of thiophene and its derivatives. The predictions were validated and were reliable. The neural network analysis produced instantaneous results of corrosion inhibition efficiency.

Artificial neural networks (ANNs) have been used to model tribological processes and have provided promising results [12]. The main functions performed by ANNs were predictions (model) and classifications of the process. Prediction may be used for diagnosis, accelerated life-time testing, on-line control of manufacturing processes that involve wear and prediction of the main properties of the mechanical systems, during the conceptual design stage. Both supervised and unsupervised models were used successfully to model the process. P. Srinivasa Pai et al. [12], used neural networks to model the microabrasion–corrosion processes of two tribological couples, a polymer/mild steel (as a function of load and potentials and in a carbonate/bicarbonate solution of pH 9.8) and ceramic/lasercarb coating (as a function of load and pHs of different solutions. The results clearly demonstrate that artificial neural networks, ANNs can be effectively used to model tribocorrosion processes and there is a need for conducting further experiments and data acquisition in order to model the process in a more effective manner. Ben-Hain and Macdonald [13] used the neural network model to forecast the influence of various parameters on the acidity of simulated brines. The solutions were based on sodium and magnesium chlorides; the network inputs were sodium and magnesium cations concentration and the temperature, and the output was the predicted pH value. A training set of more than hundred points were used for training, and a network with only two hidden nodes in a single layer was used, and the network achieved a good result. The prediction error was of the same order as the experimental error.

Cottis [1] presented a very good survey about applications of artificial neural networks in corrosion science. He concluded that artificial neural networks have generally been claimed to be successful in modeling various types of corrosion behavior. However, it is difficult to evaluate the quality of the model that is provided by a neural network with many input variables, and many studies have only validated the performance of the neural network to a limited extent [1]. Also, in this study; Cottis mentioned that there is a need to provide an indication of the confidence that can be placed in the prediction of a neural network. This requires further work towards developing the necessary techniques.
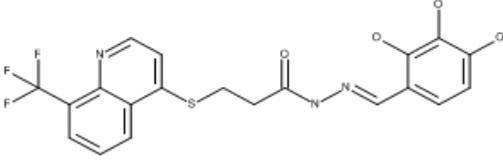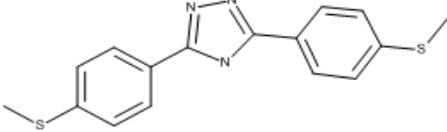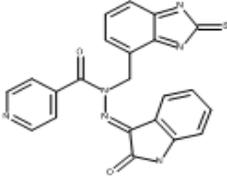
On continuation of our works on using neural networks to model the corrosion inhibition efficiency[9, 10], The aim of this work is to present a predictive model for the experimental corrosion inhibition data of mild steel by ten sulfur containing in hydrochloric acid solutions [11]. The molecular structures of the ten thio (sulfur) containing compounds namely: 6-thioguanine (TG), 2,6-dithiopurine (DTP), 4-amino-5-phenyl-4$H$-1, 2, 4-trizole-3-thiol (APTT), $N'$-(3,4-dihydroxybenzylidene)- 3-{[8-(trifluoromethyl)quinolin-4-yl]thio} propane hydrazide (DHBTPH), 3-{[8-(trifluoromethyl)quinolin-4- yl]thio}-$N'$-(2,3,4-trihydroxybenzylidene)propane hydrazide (TQTHBH), 5-(1$H$-1,2,4-triazole-1-alkyl)-1,3,4-oxadiazole-2-thiol (TAOT), $N'$-[4-(diethylamino) benzylidine]-3-{[8-(trifluoromethyl) quinolin-4-yl]thio}propane hydrazide (DEQTPH), Quinolin- 5-ylmethylene-3-{[8-(trifluoromethyl)quinolin-4-

yl]thio}propanohydrazide (QMQTPH), 3,5-bis(4-methylthio phenyl)-4H- 1,2,4-triazole (4-MTHT), *N*-(2-thio benzimidazolyl methyl) isatin- 3-isonicotinoyl hydrazone (TBIH) are presented in Table 1.

The selected sulfur containing compounds were previously reported as corrosion inhibitors for mild steel in 1.0 M HCl solution at 30 °C [11].

**Table 1:** Inhibition efficiencies and molecular structures of the studied inhibitor series

| | Inhibitor Name | Structure | Inhibition Efficiency [11] |
|---|---|---|---|
| 1 | 6-thioguanine (TG) |  | 77.9 |
| 2 | 2,6-dithiopurine (DTP) |  | 85.9 |
| 3 | amino-5-phenyl-4H-1,2,4–**4** triazole-3-thiol (APTT) |  | 86.1 |
| 4 | N'-(3,4-dihydroxylbenzylidene)-3-{[8-(trifluoromethyl)quinolin-4-yl]thio}propanohydrazide (DHBTPH) |  | 88.4 |
| 5 | 5-((1H-1,2,4-triazol-1-yl)methyl)-4H-pyrazole-3-thiol (TAOT) |  | 89.1 |
| 6 | (E)-N'-(4-(diethylamino)benzylidene)-3-((8-(trifluoromethyl)quinolin-4-yl)thio)propanehydrazide (DEQTPH) |  | 91.8 |
| 7 | (E)-N'-(naphthalen-1-ylmethylene)-3-((8-(trifluoromethyl)quinolin-4-yl)thio)propanehydrazide (QMQTPH) |  | 93.1 |

| | **Inhibitor Name** | **Structure** | **Inhibition Efficiency [11]** |
|---|---|---|---|
| 8 | 3-{[8-(trifluoromethyl)quinolin-4-yl]thio}-N'-(2,3,4-trihydroxybenzylidene)propane hydrazide (TQTHBH) | | 88.9 |
| 9 | 3,5-bis(4-methylthiophenyl)-4H-1,2,4-triazole (4-MTHT) | | 96.8 |
| 10 | N-(2-thio benzimidazolyl methyl) isatin-3-isonicotinoyl hydrazone (TBIH) | | 98.04 |

## 2. Computation details

Geometrical parameters of all stationary points for the investigated sulfur containing compounds are optimized by employing analytical energy gradients. The generalized gradient approximation (GGA) within the density functional theory was conducted with the software package DMol$^3$ in Materials Studio of Accelrys Inc. [14]. All calculations were performed using the Becke-Lee-Yang-Parr (BLYP) exchange correlation functional and the double numerical with polarization (DNP) basis set [15-17], since this was the best set available in DMol$^3$. A Fermi smearing of 0.005 hartree and a real space cutoff of 3.7 Å was chosen to improve the computational performance. All computations were performed with spin polarization.

The phenomenon of electrochemical corrosion takes place in the liquid phase, so it is relevant to include the effect of solvent in the computations. The Self-Consistent Reaction Field (SCRF) theory [18], with Tomasi's polarized continuum model (PCM) was used to perform the calculations in solution. These methods model the solvent as a continuum of uniform dielectric constant ($\varepsilon =78.5$) and define the cavity where the solute is placed as a uniform series of interlocking atomic spheres. Frontier orbital distribution was obtained, at the same basis set level, to analyze the reactivity of inhibitor molecules.

### 2.1 Artificial Neural Networks

Neural networks are inspired by the way the human brain works. The brain consists of billions of neurons, which are linked together into a complex network. A neuron communicates with another by sending an electrical signal along an axon, which is a long nerve fiber that connects to the second neuron at a synapse. Each neuron acts as an information processing element because the electrical signals sent out by one neuron depend on the strength of the incoming signals at its synapses[19]. An artificial neural network consists of an interconnected network of processing elements called nodes. A number of simplifying assumptions are made compared with the way that neurons are connected within the brain. It is assumed that the nodes are arranged in layers and that the input connections to each node come only from nodes in the layer directly below it. Hence, the structure of an artificial neural network looks something like the illustration below.
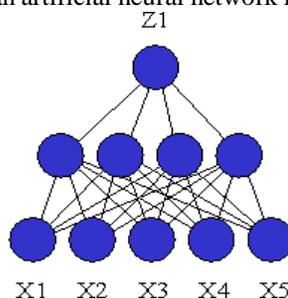

**Figure 1:** Structure of an artificial neural network[19]

The lower layer represents the input layer, in this case, with five inputs labeled X1 through X5. The input layer is used to introduce the input (predictor) variables to the network. The upper layer is the output layer. The outputs of the nodes in this layer represent the predictions made by the network for the response variables. In this case, there is a single response, labeled Z1. This network also contains a single hidden layer with four nodes[19]. In this work six descriptors include (molecular orbitals; HOMO and LUMO energy, energy gap, dipole moment, molecular area and molecular volume) the single response was the predicted inhibition efficiencies.

It has been found empirically that a single hidden layer is sufficient for modeling most data sets. Additional hidden layers allow the neural network to model more complex functions. As corrosion is a complex and nonlinear phenomenon that is too complex to be described by analytical methods or empirical rules, which make it an ideal pheromone to be studied using more than one hidden layer in artificial neural networks [10]. Details about the kolmogorov theorem that explain how a network with more hidden layers might be able to find a good model more easily are described elsewhere [9].

Training is the process whereby the connection weights and biases are set so as to minimize the prediction error for the network. For a particular set of weights and biases, each of the training cases are introduced to the network and an error function is used to determine how well the calculated outputs match the expected output values. As it is, in general, not possible to analytically determine where the global minimum of the error function is, an iterative method is used. In each training cycle, the value and gradient of the error function are calculated, and the weights and biases are adjusted according to the algorithm being used[19].

### 3. Inhibitors
Corrosion inhibition experimental studies for ten sulfur containing compounds namely, 6-thioguanine (TG), 2,6-dithio purine (DTP), 4-amino-5-phenyl-*4H*-1, 2, 4-trizole-3-thiol (APTT), *N'*-(3,4-dihydroxybenzylidene)- 3-{[8-(trifluoro methyl) quinolin-4-yl]thio}propanohydrazide (DHBTPH), 3-{[8-(trifluoromethyl)quinolin-4- yl]thio}-*N'*-(2,3,4-trihydroxy benzylidene)propane hydrazide (TQTHBH), 5-(1*H*-1,2,4-triazole-1-alkyl)-1,3,4-oxadiazole-2-thiol (TAOT), *N'*-[4-(diethyl lamino) benzylidine]-3-{[8-(trifluoromethyl) quinolin-4-yl]thio}propane hydrazide (DEQTPH), Quinolin- 5-ylmethylene-3-{[8-(trifluoromethyl)quinolin-4-yl]thio}propanohydrazide (QMQTPH), 3,5-bis(4-methylthiophenyl)-4*H*- 1,2,4-triazole (4-MTHT), *N*-(2-thio benzimidazolyl methyl) isatin- 3-isonicotinoyl hydrazone (TBIH) are presented in Table 1.The experimental corrosion inhibition efficiencies of these inhibitor candidates have been collected from literature [11].

### 4. Results and discussion
*4.1 QSAR study using the artificial neural network*
Artificial neural networks in this work are presented from the perspective of their potential use as modeling tools in quantitative structure–activity relationships (QSAR) research. Neural networks are characterized by topology, computational characteristics of their elements, and training rules [20].

The quality of the inhibition efficiencies data was assessed and the presented in Table 2. The distribution of the experimental data in Table 1 shows accepted normal distribution which enables us to start building a correlation matrix. The normal distribution behavior of the studied data in Table 1, was confirmed by the values of standard deviation, mean absolute deviation, variance, skewness and kurtosis presented in Table 2 after applying a univeriate analysis. Univeriate analysis is a technique used for generating statistics independently for the experimental inhibition efficiencies in Table 1. The skewness and kurtosis are the most important parameters . Skewness is the third moment of the distribution, which indicates the symmetry of the distribution. As the skewness is negative, the distribution of data values within the column is skewed toward negative values. For a symmetrical distribution, the skewness is zero. Kurtosis is the fourth moment of the distribution, which indicates the profile of the column of data relative to a normal distribution. Univariate analysis calculates Fisher kurtosis, which subtracts 3.0 from the definition above. For a normally distributed data set, it gives a value of 0.0. If the kurtosis is positive, the distribution of data in the column is more sharply peaked than a normal distribution. If the kurtosis is negative, the distribution is flatter than a normal distribution [21, 22].

Descriptors for the ten sulfur containing compounds are presented in Table 3. Quantum chemical parameters calculated in presence of solvent molecules as presented in computational details section are depicted in Table 3. These compound's descriptors and properties are used for developing quantitative structure activity relationships and property prediction.

Table 3 presents the quantum chemical descriptors that will be used to build a correlation matrix. These descriptors include the HOMOs, LUMOs, energy gap, total dipole moment, molecular area and volume.

**Table 2:**Univariate analysis of the experimental inhibition efficiencies data

|  | B : Experimental Inhibition Efficiencies |
|---|---|
| Number of sample points | 10 |
| Range | 20.14000000 |
| Maximum | 98.04000000 |
| Minimum | 77.90000000 |
| Mean | 89.60400000 |
| Median | 89 |
| Variance | 30.51730000 |
| Standard deviation | 5.82307000 |
| Mean absolute deviation | 4.26480000 |
| Skewness | -0.34085500 |
| Kurtosis | -0.64955400 |

**Table 3:** Descriptors for sulfur containing compounds calculated using quantum chemical method

| Inhibitor Abbreviation [11] | Experimental Inhibition Efficiencies[11] | HOMO (eV) | LUMO (eV) | Energy gap (LUMO-HOMO), eV | Total Dipole Moment (e Å) | Surface Area (Å$^2$) | Molecular Volume (Å$^3$) | Neural Network Prediction for Experimental Inhibition Efficiencies |
|---|---|---|---|---|---|---|---|---|
| DTP | 85.90 | -9.28 | -4.23 | 5.05 | 2.53 | 176.41 | 140.91 | 90.04 |
| 4-MTHT | 96.80 | -8.13 | -1.17 | 6.96 | 6.63 | 364.40 | 295.52 | 96.79 |
| TG | 77.90 | -8.20 | -1.24 | 6.97 | 2.67 | 180.27 | 140.58 | 83.85 |
| APTT | 86.10 | -8.41 | -0.86 | 7.55 | 5.18 | 213.85 | 174.26 | 86.10 |
| DEQTPH | 91.80 | -8.33 | -1.44 | 6.89 | 9.57 | 531.97 | 441.11 | 91.80 |
| DHBTPH | 88.40 | -8.50 | -2.02 | 6.48 | 10.08 | 478.26 | 392.00 | 93.75 |
| QMQTPH | 93.10 | -8.39 | -1.48 | 6.91 | 11.17 | 487.91 | 405.20 | 93.10 |
| TAOT | 89.10 | -9.48 | -1.44 | 8.04 | 4.71 | 188.99 | 149.89 | 89.10 |
| TBIH | 98.04 | -8.90 | -3.62 | 5.27 | 3.54 | 431.31 | 377.17 | 98.01 |
| TQTHBH | 88.90 | -7.91 | -1.23 | 6.68 | 5.79 | 459.73 | 374.30 | 88.91 |

Correlation matrix presented in Table 4 illustrates all possible pairwise correlation coefficients for a set of descriptors. In Table 4, we identify the highly correlated pairs of variables, and therefore redundancy in the data set will be identified [10]. Each cell of the matrix corresponds to the correlation between two columns of Table 3. The correlation coefficients lie between -1.0 and +1.0. A value approaching +1.0 indicates that the two columns are highly correlated and a value approaching -1.0 also indicates a high degree of correlation, except that the data changes values in opposite directions. A correlation coefficient close to 0.0 indicates very little correlation between the two columns. The diagonal of the matrix always has the value of 1.0. To aid in visualizing the results, the cells in the correlation matrix grid are colored according to the correlation value in each cell. A standard color scheme is used when the correlation matrix is generated: +0.9 ≤X≤+1.0 (orange), +0.7≤ X<+0.9 (yellow), -0.7<x>+0.7 (white), -0.9 <x>-0.7 (yellow) and -1.0 ≤x≤-0.9 (orange) [10, 22].

A regression analysis of the descriptor variables presented in the correlation matrix in Table 4 compared against the measured corrosion inhibition values in Table 1 [23, 24]. There are many more descriptor variables than measured inhibition values, so we have reduced a number of descriptors that did not highly correlates with each other. Typically, a ratio between two and five measured values for every descriptor should be sought to prevent over fitting.

*J. Mater. Environ. Sci. 5 (4) (2014) 1288-1297*                    *K.F.Khaled et al.*

*ISSN : 2028-2508*

*CODEN: JMESCN*

**Table 4:** Correlation matrix of the studied variables

|  | Experimental Inhibition Efficiencies[11] | HOMO (eV) | LUMO (eV) | Energy gap eV | Total Dipole Moment (e Å) | Surface Area (Å²) | Molecular Volume (Å³) |
|---|---|---|---|---|---|---|---|
| Experimental Inhibition Efficiencies[11] | 1.00 | -0.03 | -0.17 | -0.20 | 0.37 | 0.61 | 0.63 |
| HOMO (eV) | -0.03 | 1.00 | 0.61 | 0.20 | 0.37 | 0.47 | 0.45 |
| LUMO (eV) | -0.17 | 0.61 | 1.00 | 0.90 | 0.38 | 0.10 | 0.06 |
| Energy gap, eV | -0.20 | 0.20 | 0.90 | 1.00 | 0.27 | -0.14 | -0.17 |
| Total Dipole Moment (e Å) | 0.37 | 0.37 | 0.38 | 0.27 | 1.00 | 0.77 | 0.75 |
| Surface Area (Å²) | 0.61 | 0.47 | 0.10 | -0.14 | 0.77 | 1.00 | 0.99 |
| Molecular Volume (Å³) | 0.63 | 0.45 | 0.06 | -0.17 | 0.75 | 0.99 | 1.00 |

**Table 5:** Summary of input data for neural network training

| | |
|---|---|
| Number of rows requested | 10 |
| Number of rows used | 10 |
| Number of rows omitted due to invalid row description | 0 |
| Number of rows omitted due to invalid data | 0 |
| Number of columns requested | 7 |
| Number of columns used | 7 |
| Number of columns omitted due to invalid column description | 0 |
| Number of columns omitted due to invalid data | 0 |
| Number of cells omitted due to invalid data | 0 |
| Number of cells replaced by default value | 0 |

The major drawback of regression analysis is the danger of overfitting. This is the risk that an apparently good regression equation will be found which is based on a chance numerical relationship between the y variable and one or more of the x variables, rather than a genuine predictive relationship. When an overfitted model is used predictively, the predicted values for untested compounds can turn out to be very different from the true values, when these are eventually determined, even though the predicted values for the original tested compounds used to derive the regression equation were close to the true values. Such a regression equation has no predictive power. Cross validation is used to estimate the true predictive power of every regression model to reduce the risk of chance overfitting going undetected. It is important to check the results of cross validation as well as the other diagnostics to ensure that a regression model has true predictive ability[19].

**Table 6:** Cross validation of the input data for neural network training

| | |
|---|---|
| r^2 | 0.999994 |
| r^2(CV) | -0.134644 |
| Residual Sum of Squares | 5.11E-05 |
| Predictive Sum of Squares | 10.2118 |
| Analysis type | Neural Network Training |
| Network type | Predictor |
| Number of hidden layers | 1 |
| Network configuration | 6 (Input) - 3 - 1 (Output) |

Tables 5 and 6 show a summary of the input data for the neural network training set and the cross validation of the descriptor values used in the neural network analysis. Table 5 shows that all the studied sulfur containing compounds corrosion inhibitor's data was used in building the QSAR model and these data are validated against the descriptor values calculated from quantum chemical calculations [10, 22].

The cross validation data for the neural network model (Table 6) operates by repeating the calculation several times using subset of the original data to obtain a prediction model and then comparing the predicted values with the actual values for the omitted data.

Applying the neural network prediction model generates a model containing predictions corresponding to each output of the neural network. The neural network model adds a new column containing a calculation of the model (Table 2). Also, residual values of the predictions correspond to each output of the neural network.

The correlation coefficient $r^2$ is the key measure of the predictive power of the suggested model. As the closer the value is approaching 1.0, the better the predictive power of the model. For a good model, the $r^2$ value should be fairly close to 1.0. The correlation coefficient $r^2$ for this study is equal to 0.99999, which is reasonably high and indicates the predictive power of the model [10, 22]. A closer look at the last column in Table 3 (the predicted inhibition efficiency using the neural network model) and comparing the predicted values with the experimental values (column 2) proves the efficiency of the suggested model.

Investigation of the neural network analysis (Table 6) in the QSAR study shows that the network configuration is six inputs, three hidden layers and one output layer. It has too many degrees of freedom (usually the number of network connections between nodes) for the number of observations (rows of data) for which the network is being trained. In this study there is one hidden layer with three nodes [10, 22].
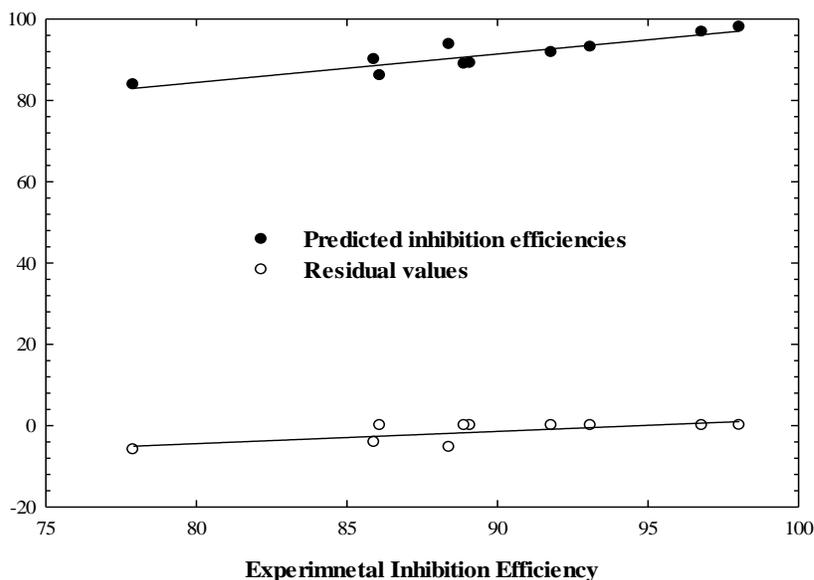


**Figure 2:** Plot of predicted inhibition and residuals vs. measured corrosion inhibition [22] using NNA.

Figure 2 shows a relation between the predicted values, residual values and the experimental data in Table 1. A residual can be defined as the difference between the predicted value in the generated model and the measured value for corrosion inhibition. To test the constructed QSAR model, potential outliers have been identified in Figure 3. An outlier can be defined as a data point whose residual value is not within cross validated $r^2$ values, is also high, even though the regression is significant according to the F-test.

Figure 3 contains two charts. One contains the residual values plotted against the corrosion inhibition measurements and the other displays the residual values plotted against Table 5 row number. Each chart contains a dotted line that indicates the critical threshold of two standard deviations beyond which a value may be considered to an outlier. Inspection of Figure 4 shows that there are no points outside the dotted lines, which make the QSAR model acceptable.
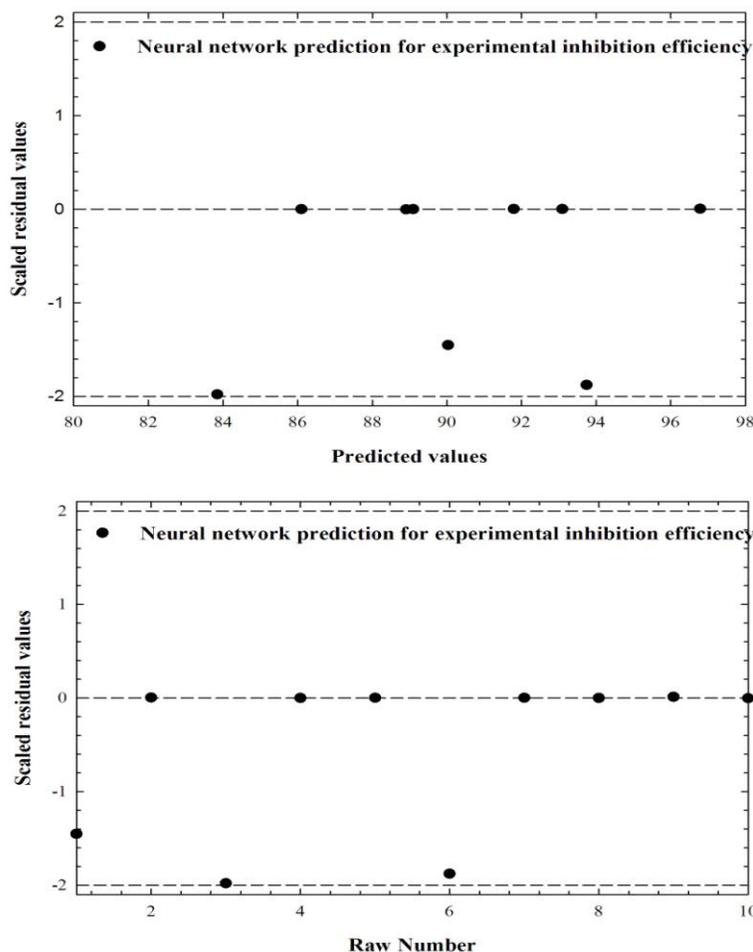
**Figure 3:** Outlier analysis for inhibition efficiency.

## 5. Conclusions

Advances in corrosion inhibition studies revealed that the relationship between structure of the inhibitor molecules and their inhibition efficiencies is essentially a non-linear and highly complex; particularly out of reach of classical statistical modeling techniques. Corrosion chemists were forced to look for other solutions. The application of neural networks may provide better and more comprehensive results. The Artificial Neural Networks (ANNs) were used to predict the inhibition efficiencies of sulfur containing compounds on the corrosion of mild steel in hydrochloric acid solutions. A network configuration consists of 6 inputs (quantum chemical descriptors), three hidden layers and one output (predicted inhibition efficiency was adopted to predict the corrosion inhibition efficiencies of ten sulfur containing compounds. The descriptors (input) were obtained using quantum chemical calculations include,:highest occupied molecular orbitals, $E_{HOMO}$, lowest unoccupied molecular orbitals $E_{LUMO}$, Energy gap ($E_{LUMO}$-$E_{HOMO}$), molecular area, molecular volume and total dipole moments were selected as the ANN input to predict the corrosion inhibition efficiencies. The predictions were reliable and the output values are within the range used in the training set. The neural network produced almost instantaneous results of corrosion inhibitor efficiency.

## References

1. Cottis R.A., 2.40 - Neural Network Methods for Corrosion Data Reduction, in: B. Cottis, M. Graham, R. Lindsay, S. Lyon, T. Richardson, D. Scantlebury, H. Stott (Eds.) Shreir's Corrosion, Elsevier, Oxford, 2010, pp. 1680-1692.
2. Smets H.M.G., Bogaert W.F.L., *Corrosion*, 48 (1992) 618-623.

3.   Cottis R.A., Qing L., Owen G., Gartland S.J., Helliwell I.A., Turega M., *Mater. Des.*, 20 (1999) 169-178.
4.   Bassam A., Ortega-Toledo D., Hernandez J.A., Gonzalez-Rodriguez J.G., Uruchurtu J., *J. Solid State Electrochem.*, 13 (2009) 773-780.
5.   Urquidi-Macdonald M., Eiden M.N., Macdonald D.D., Proc. Modifications of Passive Films, Paris, 1993.
6.   Smets H.M.G., Bogaerts W.F.L., *Fuzzy Set. Syst.*, 74 (1995) 153-162.
7.   Silverman D.C., Rosen E.M., *Corrosion*, 48 (1992) 734-745.
8.   Silverman D.C., *Corrosion*, 50 (1994) 411-418.
9.   Khaled K.F., El-Sherik A.M., *Int. J. Electrochem. Sci*, 8 (2013) 9918 - 9935.
10.  Khaled K., Al-Mobarak N., *Int. J. Electrochem. Sci*, 7 (2012) 1045-1059.
11.  Musa A.Y., Mohamad A.B., Kadhum A.A.H., Takriff M.S., Ahmoda W., *J. Ind. Eng. Chem.*, 18 (2012) 551-555.
12.  Srinivasa Pai P., Mathew M.T., Stack M.M., Rocha L.A., *Tribol. Int.*, 41 (2008) 672-681.
13.  Ben-Haim M., Macdonald D.D., *Corros. Sci.*, 36 (1994) 385-393.
14.  Zhang J., Qiao G., Hu S., Yan Y., Ren Z., Yu L., *Corros. Sci.*, 53 (2011) 147-152.
15.  Mohallem J.R., De Coura T.O., Diniz L.G., De Castro G., Assafrão D., Heine T., *J. Phys. Chem. A*, 112 (2008) 8896-8901.
16.  Ciezak J.A., Leão J.B., *J. Phys. Chem. A*, 110 (2006) 3759-3769.
17.  Ciezak J.A., Trevino S.F., *J. Phys. Chem. A*, 110 (2006) 5149-5155.
18.  Wong M.W., Frisch M.J., Wiberg K.B., *J. Am. Chem. Soc.*, 113 (1991) 4776-4782.
19.  Materials Studio 6.0 Manual, Accelrys, (2009).
20.  Niculescu S.P., *J. Mol. Struct. :THEOCHEM*, 622 (2003) 71-83.
21.  Khaled K.F., *Corros. Sci.*, 53 (2011) 3457-3465.
22.  Accelrys, Materials Studio Manual, in, Accelrys, USA, 2013.
23.  Hluchan V., Wheeler L., Hackerman N., *Werkst. Korros.*, 39 (1988) 512-517.
24.  Babić-Samardžija K., Lupu C., Hackerman N., Barron A.R., Luttge A., *Langmuir*, 21 (2003) 12187-12196.

(2014); http://www.jmaterenvironsci.com