Corrosion Rate Models for Oil and Gas Pipeline Systems: *A Numerical Approach*

Constance Obiuto Nwankwo
Department of Industrial & Production Engineering,
Faculty of Engineering,
Nnamdi Azikiwe University, Awka.

Professor Christopher C. Ihueze
Department of Industrial & Production Engineering,
Faculty of Engineering,
Nnamdi Azikiwe University, Awka.

Abstract— **This Paper focuses on the use of artificial neural networks and polynomial regression to model corrosion rate responses in oil and gas pipeline. A critical analysis of operating parameters; mean PH (X_1), mean Temp (X_2), mean pressure (X_3) and mean aqueous CO_2 partial pressure (X_4) was undertaken with statistical tools in matlab version 8.0. Of the 3 models applied, while artificial neural network has R- square value of 0.9655 the quadratic and linear models have R-square values of 0.9283 and 0.5010 respectively, indicating that the ANN model has the best fit for the responses. Sensitivity analysis carried out on the model showed that X_2 has the highest effect in increasing corrosion rate in oil and gas pipe line. It is therefore recommended that pipeline management team must take a concerted effort in controlling the effects of temperature in oil and gas pipeline installations.**

Keywords— Oil & Gas pipeline, Corrosion rate, Artificial neural networks, Multi-Layer Perceptron, Linear model, Quadratic model.

I. BACKGROUND OF STUDY

Developing a predictive model in oil and gas pipeline systems can never be over-emphasized. This is because Pipelines, like other structures in nature, deteriorate over time. Mild steel corrodes easily because all common structural metals form surface oxide films when exposed to pure air, but oxide formed on mild steel is readily broken down and in the presence of moisture it is not repaired [1]. This deterioration in metallic pipeline usually occurs as a result of the damaging effects of the surrounding environment. For mild steel, one of the most dominant forms of such deterioration is corrosion [2]. Hence, pipeline operators throughout the world are confronted with the expensive and risky task of operating aged pipelines as this is very grievous. This makes the ability to predict the rate of corrosion in pipeline systems very paramount. It will improve pipeline reliability and integrity. If subsequent activation of due action at the right time succeeds the prediction results, the catastrophic health, safety and environmental effects due to pipeline failure will be eliminated.

Several models for predicting corrosion rate exits, this study focused on polynomial and neural network models. Shriram et al [3] used polynomial approximation for predicting the cycling performance of lithium ion in batteries. Escobedo-Trujillo et al [4] applied neural network model and polynomial model to improve the coefficient of performance prediction for solar intermittent refrigeration system. Romanova et al [5] developed an evolutionary polynomial regression model for prediction of H_2S induced corrosion in concrete sewer pipes. Smets and Bogaerts [6] applied neural networks to a corrosion problem and developed networks to predict the SCC of type

304 stainless steel in near neutral solutions as a function of chloride content, oxygen content and temperature. They found that the neural network approach out-performed traditional regression techniques. Urquidi-Macdonald [7] developed a neural network model used for predicting the number and depth of pits in heat exchangers. Ben-Hain and MacDonald [8] described the use of neural network models to predict the influence of various parameters on the acidity of simulated geological brines. Several other researchers [9,10] in their work found out that the neural network prediction method out performed traditional methods of prediction. Silverman and Rosen [11] combined artificial neural networks with an expert system in order to predict the type of corrosion from a polarization curve. Inputs to the networks included the passive current density, the pitting potential and the 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. An expert system was used to interpret the outputs produced by the two approaches. The relatively small size of the training data set was one of the major concerns regarding the reliability of the model.

Most of the models found in literature are mainly based on laboratory data and, in some cases, are validated with field data. Moreover, they can be classified as mechanistic models, semi-empirical models and empirical models. They may give remarkably different corrosion rate predictions for the same field case, and which models are most successful in their prediction vary considerably from case to case. A review of polynomial models in prediction problems shows that the literature is quite extensive; but none however focuses on prediction of corrosion rate in a steel pipeline. Most of the literature dealt with corrosion inhibition.

In this study, the polynomial modelling is dissected into two, the linear and quadratic, so as to display that there are interactions between the variables being modelled.

In recent times, a nondeterministic artificial intelligence model has been applied by researchers with the aim to increase the accuracy of prediction of occurrence of corrosion and of corrosion rate, this method is excellent for modeling non-linear and complex systems and can interpolate from past experience/data. These models are based on neural network technique and they outperform deterministic models. However, application of these existing models requires advanced modelling techniques that are not readily available to the corrosion engineer who may be interested in creating his/her own neural network model for the corrosion problem at hand, hence the need for this particular research. Also

considering the real conditions affecting corrosion rate in Nigerian oil and gas pipelines, it is not same as obtained in other countries were similar research have taken place as seen in literature. This slight deviation will make such existing models not suitable in a Nigerian oil and gas context. In this study, a critical analysis of field data is done using artificial neural networks, and a predictive corrosion rate model is achieved. The neural network model can be reliably applied to predict rate of corrosion in any oil and gas pipeline facility within the shores of Nigeria.

II. METHODOLOGY

The aim is to model the corrosion data (field data from an oil and gas pipeline facility) as a function of the four independent variables. The data is first modelled as linear and quadratic multiple regression problem. The aim of the quadratic model in polynomial analysis is to reveal the effects of interaction in predictive accuracy. Thereafter the corrosion will be modelled with feed-forward artificial neural network (ANN), the multilayer perceptron analysis precisely. The results from multiple regression will then be compared with that of artificial intelligence for mutual validity and also to demonstrate the efficiency of the artificial intelligent model.

A. Linear Model

The general linear regression equation for four variable problem is given as:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \quad (1)$$

Where y = rate of corrosion, x_1 = mean pH,
x2 = mean temperature, x3 = mean pressure and
x4 = mean aqueous CO2 pressure.
in compact form, equation 1 becomes;

$$y = a_0 + \sum_{i=1}^{n} a_i x_i \quad (2)$$

B. Quadratic Model

The general quadratic equation for a four-variable problem is given as;

$$y = a_0 + \sum_{i=1}^{d} a_i x_i + \sum_{i=d+1}^{2d} a_{i,i} x_i^2 + \sum_{i \neq k} a_{i,k} x_i x_k$$

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_{1,2}x_1x_2$$
$$+ a_{1,3}x_1x_3 + a_{1,4}x_1x_4 + a_{2,3}x_2x_3 + a_{2,4}x_2x_4$$
$$+ a_{3,4}x_3x_4 + a_{1,1}x_1^2 + a_{2,2}x_2^2 + a_{3,3}x_3^2 + a_{4,4}x_4^2 \quad (3)$$

Where y = rate of corrosion, x_1 = mean pH,
x2 = mean temperature, x3 = mean pressure and
x4 = mean aqueous CO2 pressure.

III. SENSITIVITY ANALYSIS

Following the method of Dieter (1999), the sensitivity of corrosion rate expressed in equation (3) can be established by taking partial derivatives of both sides of the equation with respect to the four unknown variables.

$$\frac{\delta y}{\delta x_1} = a_1 + a_{1,2}x_2 + a_{1,3}x_3 + a_{1,4}x_4 + 2a_{1,1}x_1 \quad (4)$$

$$\frac{\delta y}{\delta x_2} = a_2 + a_{1,2}x_1 + a_{2,3}x_3 + a_{2,4}x_4 + 2a_{2,2}x_2 \quad (5)$$

$$\frac{\delta y}{\delta x_3} = a_3 + a_{1,3}x_1 + a_{2,3}x_2 + a_{3,4}x_4 + 2a_{3,3}x_3 \quad (6)$$

$$\frac{\delta y}{\delta x_4} = a_4 + a_{1,4}x_1 + a_{2,4}x_2 + a_{3,4}x_3 + 2a_{4,4}x_4 \quad (7)$$

To put the results of equation 4-7 into a meaningful basis, we determine the relative sensitivities. The relative sensitivity of the variables $x_1, x_2\ x_3\ and\ x_4$ is given by:

$$S_{x_1} = \frac{\delta y/y}{\delta x_1/x_1}, \quad S_{x_2} = \frac{\delta y/y}{\delta x_2/x_2}, \quad S_{x_3} = \frac{\delta y/y}{\delta x_3/x_3}, \quad \text{and}$$

$$S_{x_4} = \frac{\delta y/y}{\delta x_4/x_4}$$

IV. ARTIFICIAL NEURAL NETWORKS

Neural networks are inspired by biological nervous systems. They are trained, so that a particular input leads to a specific target output. Based on a comparison of the output and the target, A neural network is trained iteratively until the network output matches the target. Neural networks are applied in training complex functions in various fields, including pattern recognition, identification, classification, speech, vision, and control systems. An ANN can be 'trained' to model relationships between input and output parameters from examples of the known inputs and their corresponding outputs. These input and output are presented to the network using neurons located in input and output layers respectively (figure 1). Neurons are connected where the strength of connections is indicated by "weight". In the training process, a set of examples of input-output pairs are passed through the model and the weights are adjusted in order to minimize error between the answers from the network and the desired outputs. This weight alteration procedure is controlled by the learning algorithm.

Multilayer Layer Perceptron (MLP) and Radial Basis Function (RBF) networks are the most neural network function approximation approaches. The MLP which is a feed forward propagation analysis was chosen for this study. MLP is made of an input layer, one or more hidden layers, and output layer (fig 1).
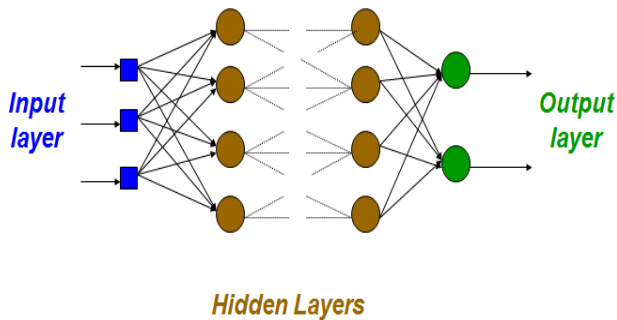
Figure 1. Simple neural network diagram (http://cs231n.github.io/neural-networks-1/)

The nodes are the biological analogy of neurons. A general Node i in a MLP network weights all inputs and sums the linear combination with a bias in the summer. A nonlinear activation function then transforms the sum to the output (fig 2)

where x_k, $k = 1, 2 \ldots \ldots K$ are the inputs, $w_{ki}^{(p)}$ are the weights, $\theta_i^{(p)}$ is the bias, $n_i^{(p)}$ is the weighted sum from the summer, g is the activation function and $y_i^{(p)}$ is the output. The relationships hold

$$n_i^{(p)} = \theta_i^{(p)} + \sum_{k=1}^{K} w_{ki}^{(p)} x_k \tag{8}$$

$$y_i^{(p)} = g\left(n_i^{(p)}\right) \tag{9}$$
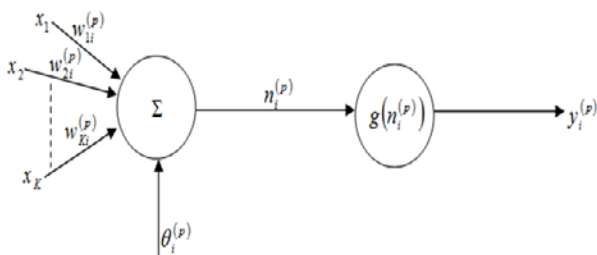
Figure 2: Neural network computation at a node (Jose, C. et al, 1999)

In real network, there could be many layers of nodes between the input and output. A network architecture with single hidden layer ($P = 2$) of i nodes and single output is developed and presented in figure 3.

A. The Network Design
Multilayer Layer Perceptron networks with different architectures were designed and solved with Levenberg-Marquardt algorithm. A three-layered topology network was chosen, consisting of: input, hidden and output layers respectively. A one hidden layer network is recommended for training efficiency, more hidden layers lead to complex networks.
The 'Neural Network Toolbox' for MATLAB is used in the solution. Different network architectures with single hidden layer and the number of nodes was varied from 5 to 30 in steps

of five to ascertain the sensitivity of the method accuracy to number of nodes. The tangent sigmoid (tansig) and linear (purelin) transfer functions were used in the hidden layer(s) and output layer, respectively. The Error tolerance or stopping criterion maintained throughout is 0.000001. The maximum number of iterations or epochs allowed in all the architectures is 300. A default learning rate of 0.05 was used in all the structures. The results are indicated at every 50th epoch.

V. RESULTS AND DISCUSSION

A. Curve Fitting and Modelling
Substituting the n measured data points from field data into equation (1), we obtain a set of n equations for solution of coefficients of equation (2), so that by substitution into equation (1),

$$y = 0.2693 + 0.0039x_1 - 0.0001x_2 - 0.0006x_3 - 0.0074x_4 \tag{10}$$

A solution of a normal equation generated from equation (3) resulted to

$$y = -0.9331 + 0.0876x_1 + 0.0425x_2 + 0.0139x_3 + 0.2339x_4$$

$$-0.0031x_1x_2 - 0.0015x_1x_3 - 0.1197x_1x_4 - 0.0001x_2x_3 + 0.0239x_2x_4 + 0.0028x_3x_4 + 0.0123x_1^2 - 0.0010x_2^2 - 0.0001x_3^2 + 0.0051x_4^2 \tag{11}$$

The R^2-value and correlation of the linear model are 0.5010 and 0.7078 respectively. These statistical indices are not satisfactory as also evident from unacceptable level of scatter evident in figure 4. The ideal line is also shown; the more accurate the model the less deviated from the ideal line is the predicted data. The poor performance of the linear model suggested the importance of trying the quadratic model since it houses all possible second order interaction of variables.

The R^2-value and correlation (r) of the quadratic model are 0.9283 and 0.9635 respectively. These statistical indices are satisfactory. This is evident in figure 5 from which it is seen that the scatter of the model from the ideal line has been reduced. This improvement in accuracy is due to consideration of interactions in the quadratic model.

B. Neural Network Analysis and Results
Multilayer Layer Perceptron networks with different architectures were designed and solved with Levenberg-Marquardt algorithm. The 'Neural Network Toolbox' for MATLAB is used in the solution. There are four input parameters into the network, which consist of x_1 (*Mean PH*), x_2 (*Mean Temp*), x_3 (*Mean Pressure*) and x_4(*Mean Aqueous CO₂ Partial Pressure*). There is one output parameter corresponding to the y (*Mean Corrosion Rate within the Period*) measured in (*mm/yr*). There were different network architectures with single hidden layer and the number of nodes was varied from 5 to 30 in steps of five to ascertain the sensitivity of the method's accuracy to number of nodes. The

tangent sigmoid (tansig) and linear (purelin) transfer functions were used in the hidden layer and output layer, respectively. The Error tolerance or stopping criterion maintained throughout is 0.000001. The maximum number of iterations or epochs allowed in all the architectures is 300. A default learning rate of 0.05 was used in all the structures. The results are indicated at every 50th epoch.

The results from the ANN are presented in table 1 alongside the results of the polynomial models. The y_{EXP} represents experimental data while the y_{LPM} and y_{LPM} respectively represent the results from linear and quadratic models, that is equations (6 and 7). The $y_{ANN,N5}$, $y_{ANN,N10}$, $y_{ANN,N15}$, $y_{ANN,N20}$, $y_{ANN,N25}$, and $y_{ANN,N30}$ are respectively the results from the ANN model of nodes 5 to 30 in steps of five. It is seen from table 1 that the number of epochs and consequently the time needed for ANN modeling of the studied corrosion data reduced with rise in the number of nodes. It is also seen from R^2-values and Correlation that accuracy of ANN modeling of the studied corrosion data improved with rise in number of nodes. The meaning is that from both stand points of speed and accuracy, it is better to use higher number of nodes and less number of iterations than use lower number of nodes and higher number of iterations. It is clear that ANN modelling of the studied corrosion is more accurate than the polynomial regression models at all the number of nodes considered. The two different approaches can be view as mutual validation. It must be noted that the importance of the polynomial regression models is not in any way degraded as they are veritable means of ascertaining the statistical significance of interactions. The results of the ANN modelling of the studied corrosion data is presented in figures 6a-f. A visual comparison of the developed models is presented in figure 7 where it is seen that the ANN model with thirty nodes agreed most with the ideal line followed by the quadratic polynomial model then the linear polynomial model. Table 1 is the summary of accuracies and efficiencies of all the developed models.

C. Sensitivity Results

Taking the partial differentiation of the quadratic model which has a reliable statistical index, with respect to each of the four unknown variables yield the four simultaneous equations (12-15)

$$0.0246x_1 - 0.0031x_2 - 0.0015x_3 - 0.1197x_4 = -0.0876 \quad (12)$$
$$-0.003x_1 - 0.0020x_2 - 0.0001x_3 - 0.0239x_4 = -0.0425 \quad (13)$$
$$-0.0015x_1 - 0.0001x_2 - 0.00002x_3 + 0.0028x_4 = -0.0139 \quad (14)$$
$$-0.1197x_1 + 0.0239x_2 + 0.0028x_3 + 0.0102x_4 = -0.2339 \quad (15)$$

The solution of the resulting $4*4$ matrix from equations 12-15 is solved using the Scilab 5.4.1 software to obtain $x_1 = 3.8334, x_2 = 3.1582, x_3 = 50.4496 \ and \ x_4 = 0.8057$ respectively; Substituting the values of $x_1, x_2, x_3 \ and \ x_4$ into equation (11), we obtain $y = -0.1924$

For the relative sensitivities;

$$S_{x_1} = \frac{\delta y/y}{\delta x_1/x_1} = 0.1888,$$
$$S_{x_2} = \frac{\delta y/y}{\delta x_2/x_2} = 1.1788, \quad S_{x_3} = \frac{\delta y/y}{\delta x_3/x_3} = 0.0582,$$
$$S_{x_4} = \frac{\delta y/y}{\delta x_4/x_4} = -0.0753$$

Therefore, it can be deduced that changes in the variable x_2 has the greatest influence on the response y.

VI. CONCLUSION

This study established that;
1. Artificial neural network is an appropriate numerical technique for modelling corrosion rate in oil and gas pipeline systems when compared to the polynomial models that equally fits the corrosion rate data. It has displayed the best fit amongst the studied models.
2. The quadratic model considered all second order interactions between the variables and equally had a good fit.
3. The quadratic model has helped to reveal the variable with the greatest effect on response, it was used to determine the relative sensitivities of the variables.
4. The variable x_2 ,which is the temperature, has the highest effect in increasing corrosion rate, in oil and gas pipelines.

REFERENCES

[1] Badmos A.Y; Ajimotokam H.A and Emmanuel E.O(2009). "Corrosion in Petroleum Pipelines", New York Science Journal, 2(5), 36-40.

[2] Callister W.D. Jr (1997). Material Science and Engineering, An Introduction, Fourth Edition, John Willey & Sons USA.

[3] Shriram S., Qingzhi G., Premanand R., Ralph E.W (2006). "Prediction of cycling performance of lithium ion in batteries" Journal Power of Sources 156 p. 620-628.

[4] Escobedo-Trujillo BA; Colorado D and Riv-Era W. (2016). "Neural network and polynomial model to improve the coefficient of performance prediction for solar intermittent refrigeration system" Solar Energy, 129:28-37.

[5] Romanova, A; Faramarzi, A; Mahmoodian A and Alani, M.A (2014). "An evolutionary polynomial regression (EPR) model for prediction of H2S induced corrosion in concrete sewer pipes", 11th International Conference on Hydro-informatics, New York City, USA.

[6] Smets, H. M. and Bogaerts, W. F. (1992). "Neural network prediction of stress-corrosion cracking," Mater Perform, Vol. 31, 1992, p. 64-67.

[7] Urquidi-Macdonald, M; Eiden, M. N. and Macdonald D. D (1993). "Development of a neural network model for predicting damage functions for pitting corrosion in condensing heat exchangers," Modifications of Passive Films, Paris, pp. 336-343.

[8] Ben-Hain M. and Macdonald D. D (1994). "Modeling geological brines in salt-dome high level nuclear waste isolation repositories by artificial neural networks" Corrosion Science, Vol. 36, 1994, p. 385-393.

[9] Trasatti S. P. and Mazza F. (1996). "Crevice corrosion: a neural network approach" Corrosion Journal, Vol. 31, pp.105-112.

[10] Pidaparti R. M; Jayanti S and Palakal M. J. (2002). "Residual Strength and Corrosion Rate Predictions of Aging Aircraft Panels: Neural Network Study" Journal of Aircraft, No. 1, Vol. 39, pp. 175-180.

Table 1. Performance (speed and accuracy) of the developed models

Model	y_{LPM}	y_{QPM}	$y_{ANN,N5}$	$y_{ANN,N10}$	$y_{ANN,N15}$	$y_{ANN,N20}$	$y_{ANN,N25}$	$y_{ANN,N30}$
Time (sec)	-	-	7	3	2	1	<1	1
Epochs	-	-	227	132	71	29	21	21
R^2-value	0.5010	0.9283	0.9644	0.9633	0.9629	0.9635	0.9630	0.9655
R	0.7078	0.9635	0.9822	0.9815	0.9813	0.9818	0.9814	0.9827
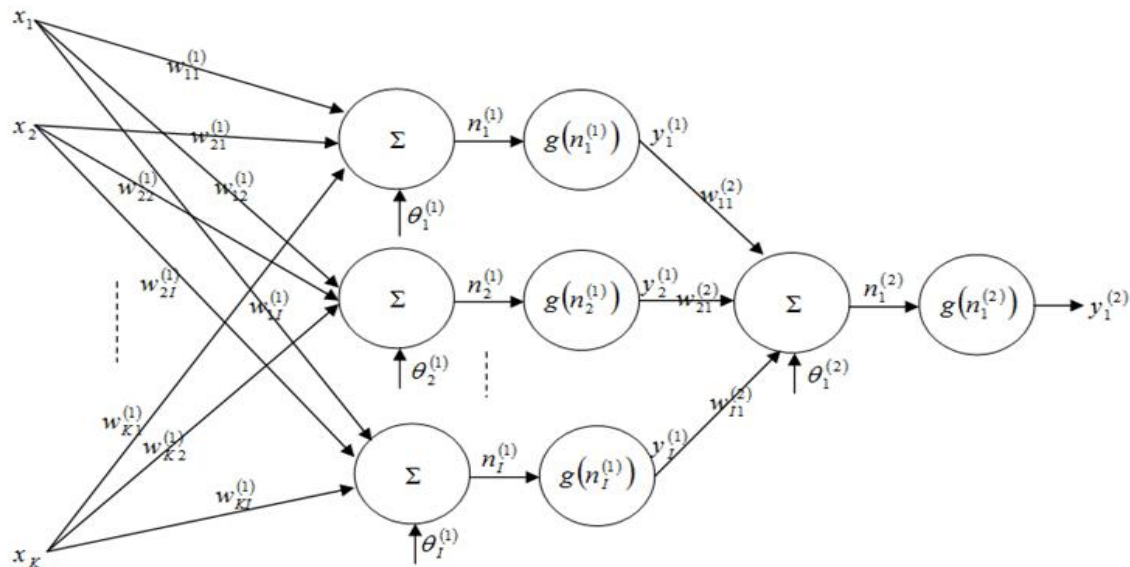
Figure 3: Neural network Diagram
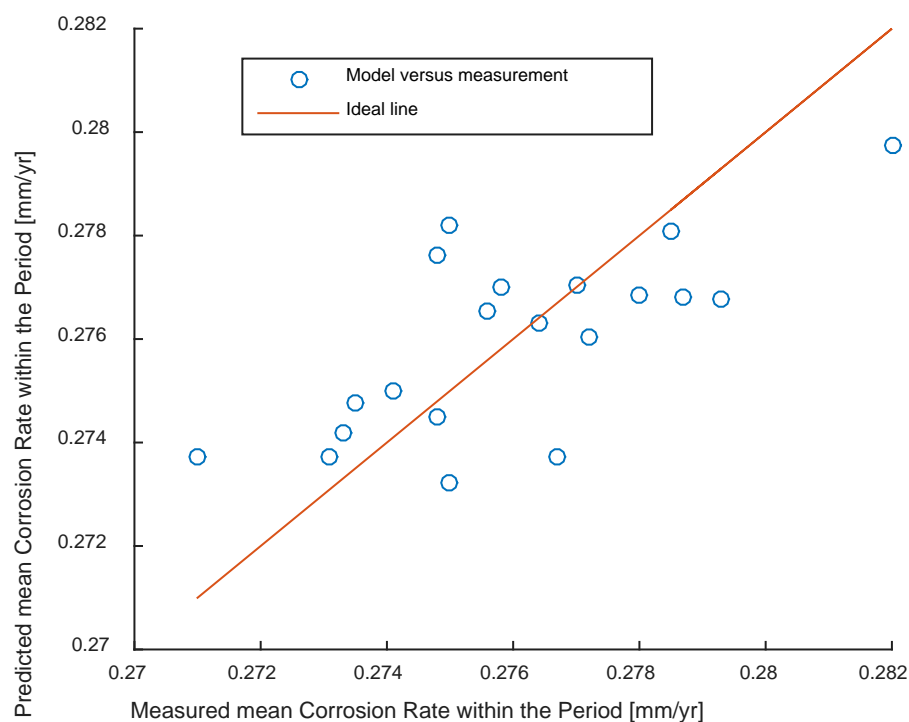
Figure 4. The scatter plot of predicted data around the ideal line for the linear polynomial model.

Published by :

http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 08, August-2018
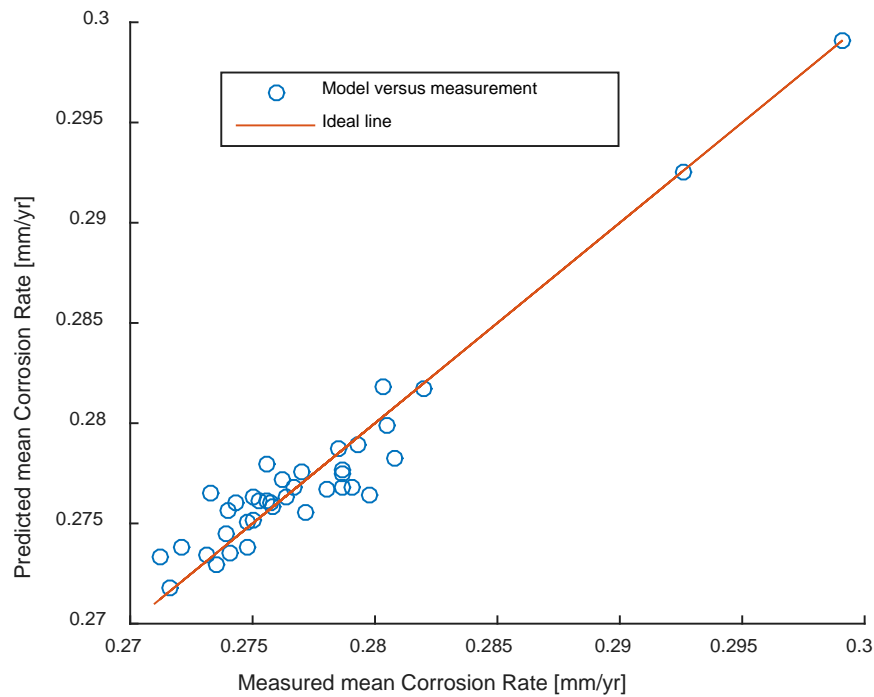
Figure 5. The scatter plot of predicted data around the ideal line for the quadratic polynomial model.
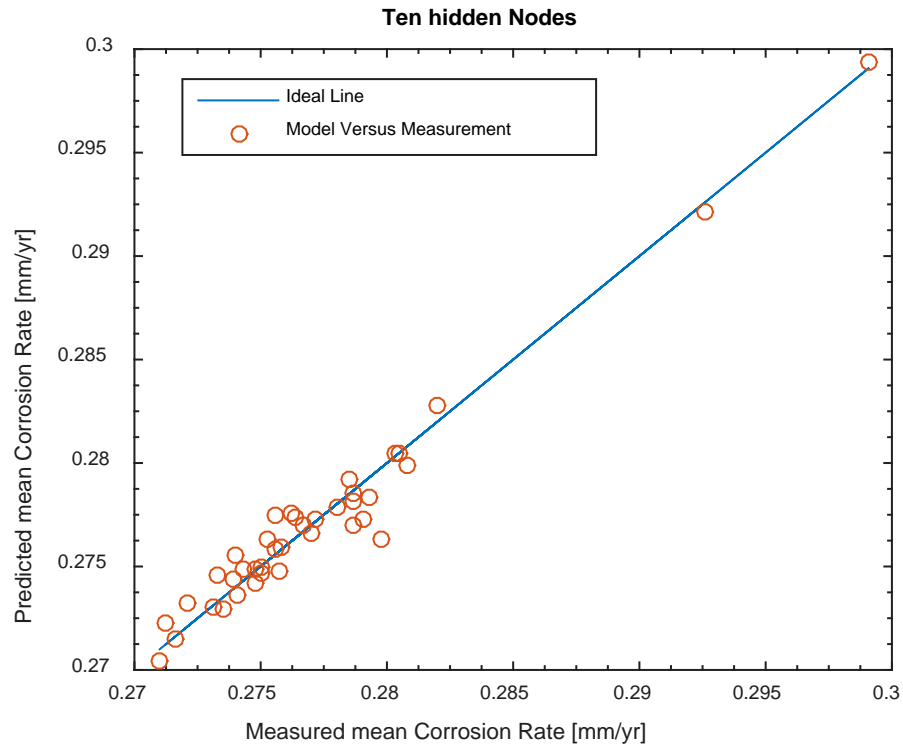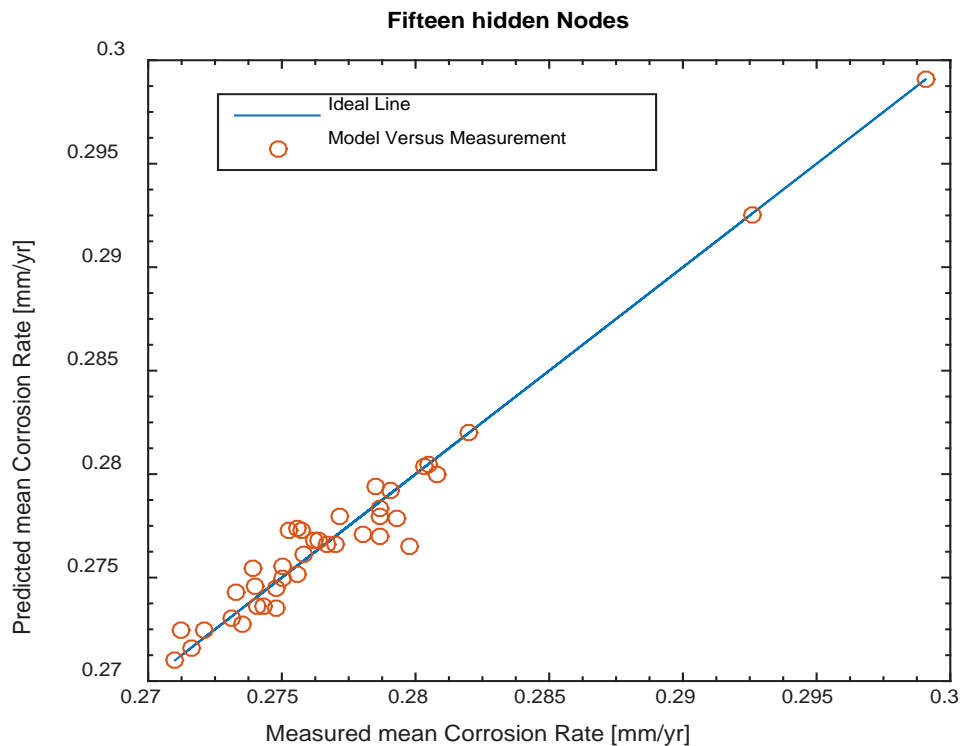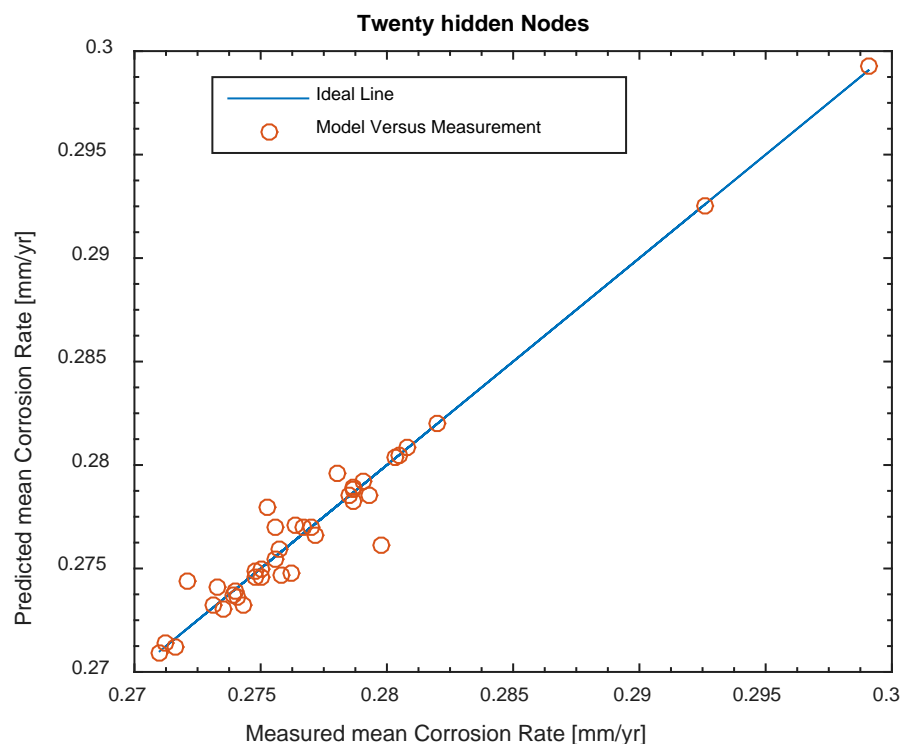
Five hidden Nodes
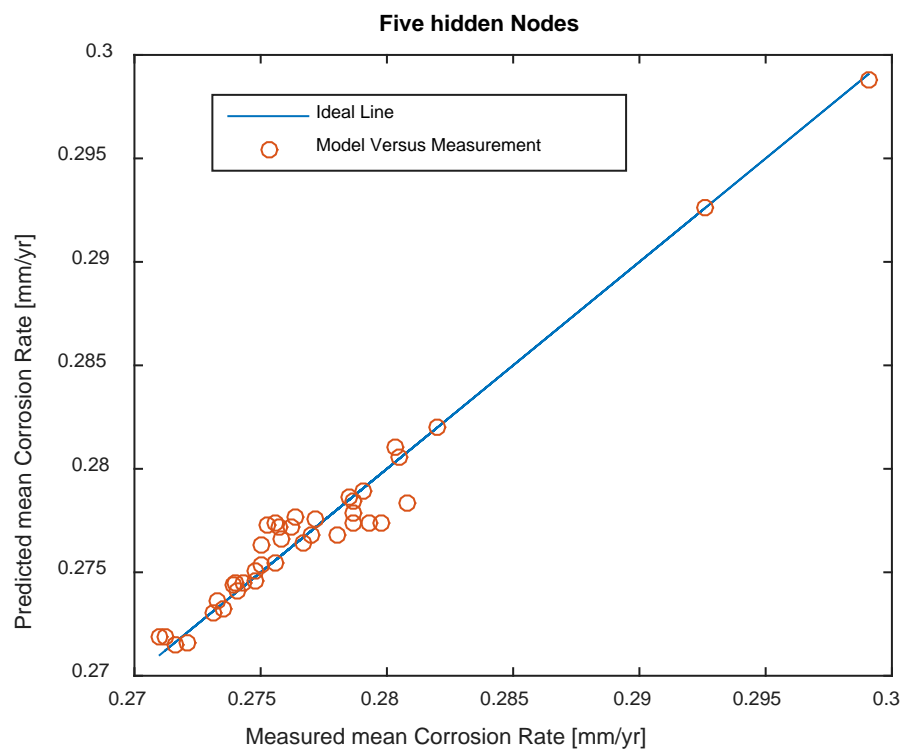
Figure 6a: The scatter plot of predicted data around the ideal line for the ANN modeling with 5 nodes

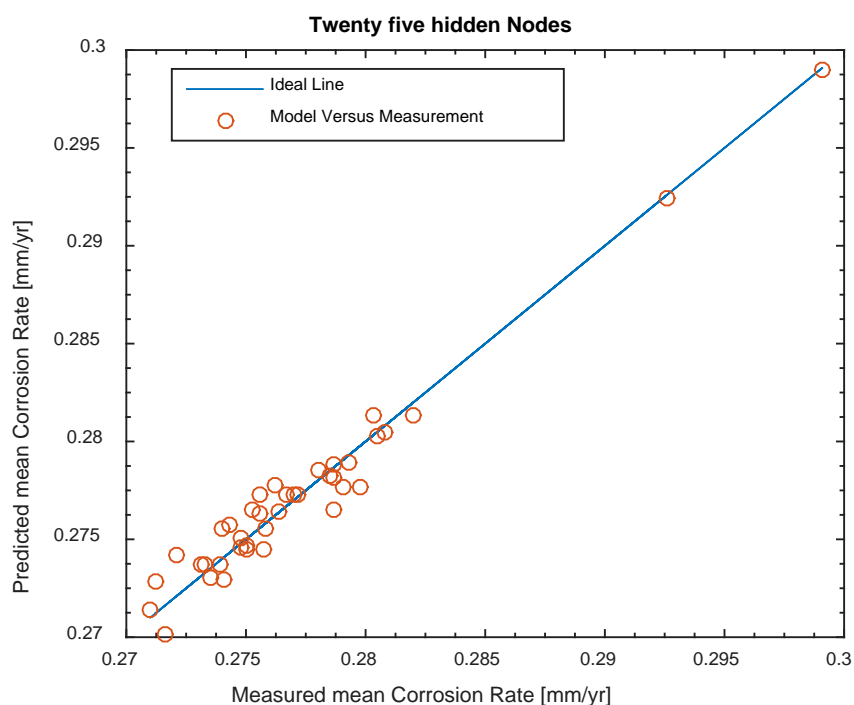Figure 6b: The scatter plot of predicted data around the ideal line for the ANN modelling with 10 nodes

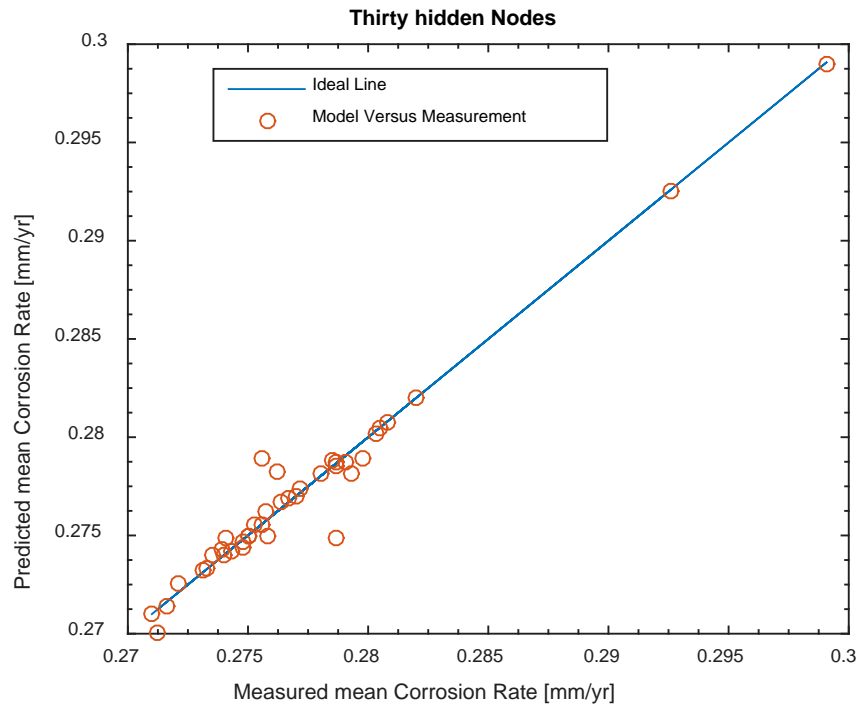Figure 6c: The scatter plot of predicted data around the ideal line for the ANN modelling with 15 nodes

Figure 6d: The scatter plot of predicted data around the ideal line for the ANN modelling with 20 nodes

Figure 6e: The scatter plot of predicted data around the ideal line for the ANN modelling with 25 nodes

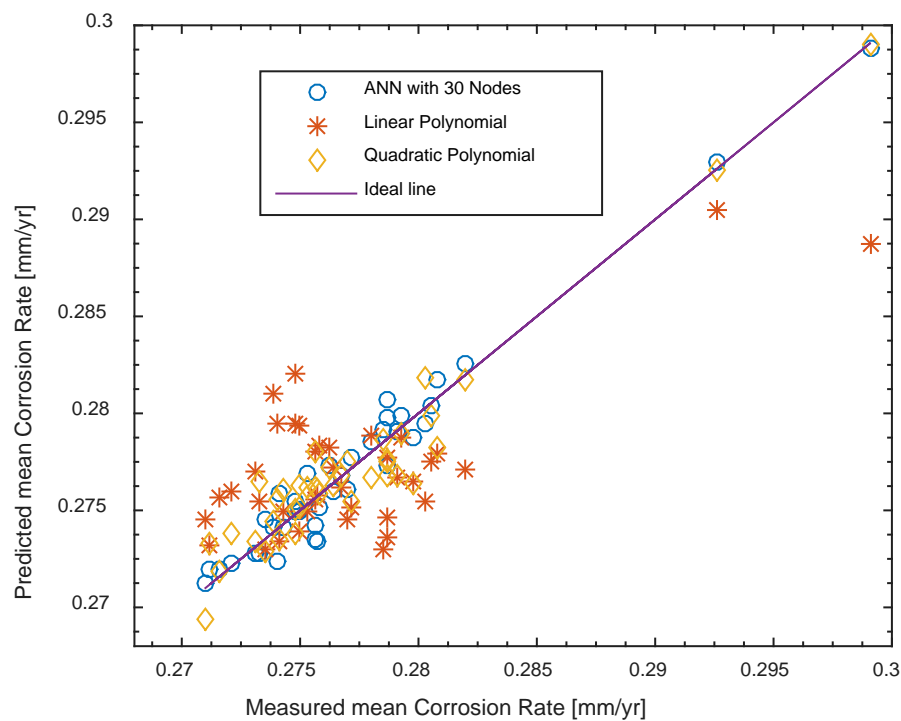Figure 6f: The scatter plot of predicted data around the ideal line for the ANN modelling with 30 nodes
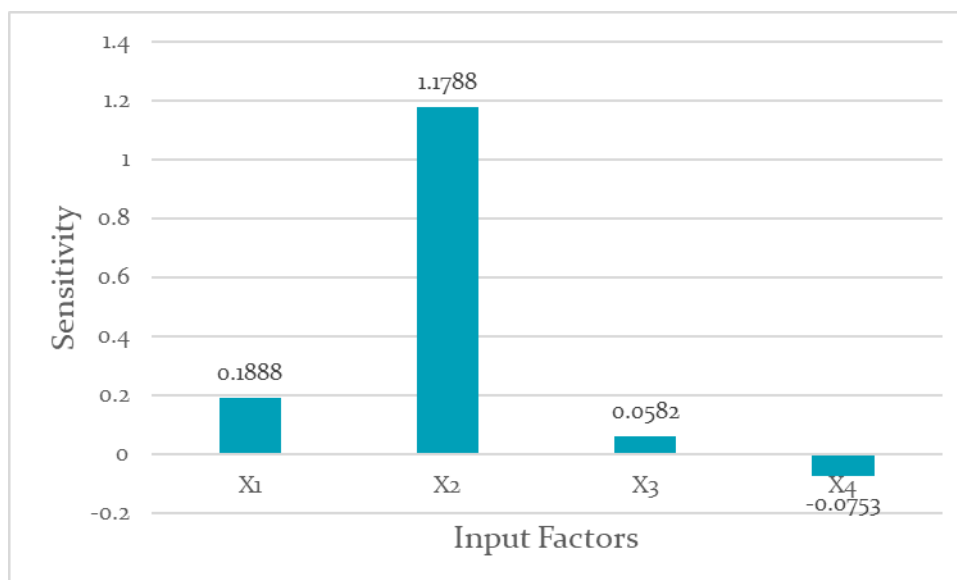

Figure 7. A visual comparison of the developed models

Figure 8: Graph of relative sensitivity