Paper No.
05554

CORROSION2005

Use of Artificial Neural Networks for Predicting Crude Oil Effect on CO_2 Corrosion of Carbon Steels

Hernández, S.; Nesic, S.
Ohio University
Institute for Corrosion and Multiphase Technology
342 West State St.
Athens, OH 45701
e-mail: hernande@ohio.edu

Weckman, G., Ghai, V.
Ohio University
Industrial and Manufacturing Systems Engineering
Stocker Center 280
Athens, OH 45701-2979

ABSTRACT

The role of crude oil on CO_2 corrosion has gained special attention in the last few years due to its significance when predicting corrosion rates. However, the complexity and variability of crude oils makes it hard to model its effects, which can influence not only wettability properties but also the corrosivity of the associated brine. This study evaluates the usefulness of Artificial Neural Networks (ANN) to predict the corrosion inhibition offered by crude oils as a function of several of their properties which have been related in previous studies to the protectiveness of crude oils, i.e. nitrogen and sulfur contents, resins and asphaltenes, TAN, nickel and vanadium content, etc. Results showed that neural networks are a powerful tool and that the validity of the results is closely linked to the amount of data available and the experience and knowledge that accompany the analysis.

Keywords: crude oil, artificial neural networks, corrosion prediction, CO_2 corrosion.

INTRODUCTION

Modeling the effect of crude oil in CO_2 corrosion is not an easy task. Even though many researchers have worked in the area, the complexity of the chemical nature of crude oil makes it difficult to generalize their behaviour or to

develop a mechanistic model. The majority of the work that has been done in the area is either experimental or based on field data. Even though plenty is yet to be understood, there is presently a better understanding of the effect of hydrocarbons than a decade ago and also there is a higher awareness of the relevance of including its effect into the available prediction models.

Nowadays, most authors agree that both wettability and changes in the brine chemistry are the major ways in which crude oil affects the corrosion of carbon steels in production environments. These two effects, in turn, are a consequence of the specific chemistry of the crude oil.

Even if no definite modelling or prediction was done, the discovery that both fluid dynamics and interfacial properties playing a role in CO_2 corrosion and the differentiation made among various crude oils with different origins played an important role in enhancing our knowledge base of the effect of crude oil in CO2 corrosion.

In 1991, Efird[1] stressed the importance of testing the effect of specific crude oils and including it in corrosion prediction and testing. He also introduced the definition of "Corrosion Rate Break" as the level of produced water in crude oil production where corrosion is accelerated and becomes a problem. He found that the corrosion rate/produced water content curves for different crude oils and commonly fall into one of three general types as shown in **Figure 1**. This shows that the onset of accelerated corrosion in crude oil production cannot be reliably predicted by using a predetermined produced water level, and also explain how different crude oils can show very different wettability properties and offer different degrees of protection.

In 1993 Smart[2] presented another important piece of work, relating petrophysical data and wettability properties to corrosion. Based on the work by Anderson in 1986[3,4] Smart indicated that wettability could be strongly affected by surface active compound present in the crude oil. These surface active compounds are believed to be polar compounds containing oxygen, nitrogen or sulphur which are more prevalent in the heaviest fractions of crude oils such as resins and asphaltenes.

The polarity of the crude oil and of the corrosion products such as iron oxide, iron carbonate or iron sulphide are considered a major factor in wettability also. The polarity of a crude oil is described as a synergistic action of its polar and polarizable molecules, such as resins and asphaltenes and its hetero/atomic compounds: nitrogen, sulphur and oxygen. The presence of these polar constituents could change wettability as related to corrosion by reducing the interfacial tension between water and oil and by changing the tendency of crude oil to wet the surface.

In an attempt to model the effect of crude oil, in 1993, de Waard et al[5] developed an empirical factor to be used in his widely known prediction correlation. To his expression of corrosion rate:

$$\log(Vcorr) = 5.8 - \frac{1710}{273 + t} + 0.67 \log(pCO_2) \qquad (1)$$

a multiplier (F_{oil}) equal to one will be added when water cuts are higher than 30% and flow velocities bigger than 1m/s. These critical flow rate and water percentages come from the work done by Wicks et al[6] and Lotz et al[7], respectively. When these conditions are not met it is assumed that the steel will be oil wetted and water will be entrained in the crude oil, then the multiplier will be zero, meaning no corrosion rate. This multiplier was improved in 2001 to include different modes of water entrainment, linking API gravity to the water-in-oil emulsion stability which, in turn, is linked to the effect of oil wetting of the metal. At this point the multiplier was written as:[8]

$$F_{oil} = 0.545 \frac{\alpha}{90} + 4.3 WU_{liq} \left(1 + \frac{\alpha}{90}\right) \qquad (2)$$

Where W is the average water fraction of the liquid measured at the wellhead, U_{liq} is the liquid velocity in m/s and α is the angle of deviation (in degrees) of the tubing form the vertical.

This factor, however, comes from fitting observed field data with corrosion rates predicted with a semi-empirical model for CO_2 corrosion[9], and was exclusively developed for very light oils. Later in the same year and since equation (2) can be expressed in three terms, that is:

$$F_{oil} = 4.3WU_{liq} + 0.545\frac{\alpha}{90} + 4.3WU_{liq}\frac{\alpha}{90} \qquad (F_{oil} \le 1) \qquad (3)$$

de Waard et al suggest that the influence of light crude oils on corrosion rates in near liquid full systems consists of three contributions, so they tried to link each of these contributions to a specific "mode" of entrainment[10]. They defined three modes of water entrainment: mode I being water in oil emulsion, modes II and III when water separates from the crude oil phase. In mode II conditions are such that the water phase can remain stationary at certain locations (with oil flowing around it) while in mode III water will move with the fluid, thus wetting the steel intermittently.

Based on the results from Craig [11] and using 10% water as an indicator of the emulsion breakpoint, they derived a relationship between the API gravity and the emulsion breakpoint:

$$W_{break} = -0.0166*API + 0.83 \qquad 50>API>20 \qquad (4)$$

W_{break} is regarded as an indication of the interfacial tension between the crude oil and the water: the lower this tension, the higher is the amount of water which can be present as an emulsion in the oil. When the interfacial tension between oil and water is low, the interfacial tension between metal and oil will also be lower resulting in a better wetting of the steel by the oil, thereby reducing the rate of corrosion. According to this, they generated an oil factor which changes as the watercut changes in relation to W_{break}, as follows:

$$F_{oil} = 0.071\frac{W}{W_{break}}U_{liq} \qquad W < W_{break},\ F_{oil} \le 1 \qquad (5)$$

$$F_{oil} = 0.071\frac{W}{W_{break}}U_{liq} + 0.545\frac{\alpha}{90} + 0.071\frac{\alpha}{90}\frac{W}{W_{break}}U_{liq} \qquad W \ge W_{break},\ F_{oil} \le 1 \qquad (6)$$

At higher water cuts, the validity of Equation 6 becomes limited and does not obey the boundary condition that F_{oil} approaches 1 when the water cut approaches 100%. So, at high water cuts an additional constraint is added to Equation 6:

$$F_{oil} \ge W \qquad (7)$$

Also, for very low flow velocities or near stagnant conditions:

$$\text{When} \quad U_{liq} < \frac{\alpha}{90}, F_{oil} = 1 \qquad (8)$$

This approach links oil wetting properties with API gravity of the oil. However it remains an empirical correlation built on limited field data and does not consider the influence of crude oil properties which can affect the stability of water in oil dispersions.

More recently, in 2004, a new approach was presented by Cai *et. al*. where a criterion for forming stable water-in-oil dispersed flow was proposed as the mean for calculating a critical velocity for water entrainment[12] in oil-water pipe flows. To account for the complexity of the hydrodynamics and to extract a valid criterion for water separation and entrainment Cai et al.[13] followed Brauner[14] and Barnea[15] proposal. Two main physical properties, maximum droplet size, d_{max}, related to breakup and coalescence and critical droplet size, d_{crit}, related to settling and separation are compared to deduce this criterion. Since water is entrained by the flowing oil phase in the form of droplets, it is essential to know the maximum droplet size d_{max} that can be sustained by the flow without further breakup. In dilute water-in-oil dispersion d_{max} evolves from a balance between the turbulent kinetic energy and the droplet surface energy. For the dilute dispersion Brauner[12] shows that:

$$\left(\frac{d_{\max}}{D}\right)_{dilute} = 1.88\left[\frac{\rho_o(1-\varepsilon_w)}{\rho_m}\right]^{-0.4} We_o^{-0.6}\,\mathrm{Re}_o^{0.08} \tag{9}$$

Where:

$$\mathrm{Re}_o = \frac{\rho_o D U_c^2}{\eta_o} \qquad We_o = \frac{\rho_o D U_c^2}{\sigma}$$

where D and d_{max} denote the pipe diameter and maximum droplet size, respectively, in m. ε_w presents the water cut. ρ denotes the density of liquid, in $kg\ m^{-3}$. The subscripts o, m and *dilute* present the oil phase, the oil-water mixture and dilute oil-water dispersion, respectively. η_o denotes the viscosity of oil phase, in $Pa.s$. σ presents the oil surface tension, in Nm^{-1}.

It is noted that this equation can be only used in the dilute dispersions i.e. as long as it satisfies the following condition:

$$(1-\varepsilon_w)\frac{\rho_o}{\rho_m} \cong 1 \tag{10}$$

In dense dispersions, droplet coalescence takes place. Under such conditions, the flowing oil phase disrupts the tendency of the water droplets to coalesce. Brauner[12] has shown that this leads to:

$$\left(\frac{d_{\max}}{D}\right)_{dense} = 2.22 C_H^{0.6}\left(\frac{\rho_o U_c^2 D}{\sigma}\right)^{-0.6}\left(\frac{\varepsilon_w}{1-\varepsilon_w}\right)^{0.6}\left[\frac{\rho_m}{\rho_o(1-\varepsilon_w)} f\right]^{-0.4} \tag{11}$$

where U_c denotes the velocity of continuous phase, in ms^{-1}. C_H is a constant with the order of $O(1)$. D and d_{max} denote the pipe diameter and the maximum droplet size, respectively, in m. ε_w presents the water cut. ρ denotes the density of liquid, in $kg\ m^{-3}$. The subscripts o, m and *dilute* present the oil phase, the oil-water mixture and dilute oil-water dispersion, respectively. f is the friction factor. σ presents the oil surface tension, in Nm^{-1}.

$$f = 0.046/\mathrm{Re}_o^{0.2}$$

Thus, given a water-oil fluid system and operational conditions, the maximum droplet size that can be sustained is the larger of the two values obtained via (9) and (11), which can be considered as the worst case for a given oil-water system:

$$\frac{d_{\max}}{D} = Max\left\{\left(\frac{d_{\max}}{D}\right)_{dilute}, \left(\frac{d_{\max}}{D}\right)_{dense}\right\} \tag{12}$$

Droplets larger than a critical droplet size d_{crit} separate out from the two-phase flow dispersion either due to gravity forces, predominant in horizontal flow, or due to deformation and "creaming" typical for vertical flow[13]. *Critical droplet diameter*, d_{cb}, above which separation of droplets due to gravity takes place can be found via a balance of gravity and turbulent forces as[13]

$$\left(\frac{d_{cb}}{D}\right) = \frac{3}{8}\frac{\rho_o}{|\Delta\rho|}\frac{fU_c^2}{Dg\cos(\theta)} = \frac{3}{8}f\frac{\rho_o}{\Delta\rho}Fr_o \tag{13}$$

Where Froude number is:

$$Fr_o = \frac{U_c^2}{Dg\cos(\theta)}$$

and

$$\Delta\rho = |\rho_o - \rho_w|$$

Where D and d_{cb} denote the pipe diameter and the *critical droplet size*, respectively, in *m*. ρ denotes the density of liquid, in *kg* m^{-3}. The subscripts *o and w* present the oil phase and water phase, respectively. θ denotes the inclination of pipe, in *degree*.

This effect is predominant at low pipe inclinations i.e. in horizontal and near-horizontal flows. Critical droplet diameter, $d_{c\sigma}$, above which drops are deformed and "creamed", leading to migration of the droplets towards the pipe walls in vertical and near-vertical flows, can be calculated with the equation proposed by Brodkey.[16]

$$\left(\frac{d_{c\sigma}}{D}\right) = \left[\frac{0.4\sigma}{|\rho_o - \rho_w| gD^2\cos(\beta)}\right]^{0.5} \tag{14}$$

$$\beta = \begin{cases} |\theta| & |\theta| < 45^o \\ 90 - |\theta| & |\theta| > 45^o \end{cases}$$

Where D and $d_{c\sigma}$ denote the pipe diameter and the *critical droplet size*, respectively, in *m*. ρ denotes the density of liquid, in *kg* m^{-3}. The subscripts *o and w* present the oil phase and water phase, respectively. f is the friction factor. θ denotes the inclination of pipe, in *degree*. σ presents the oil surface tension, in Nm^{-1}.

The critical droplet diameter, d_{crit}, can then be conservatively estimated for any pipe inclination according to the suggestion made by Barnea[13] (1987):

$$\frac{d_{crit}}{D} = Min\left\{\left(\frac{d_{cb}}{D}\right),\left(\frac{d_{c\sigma}}{D}\right)\right\} \tag{15}$$

At this point the final criterion for entrainment emerges. The transition from stratified flow to stable water-in-oil dispersion takes places when the oil phase turbulence is intense enough to maintain the water phase broken up into droplets not larger then d_{max} which has to be smaller than the a critical droplet size d_{crit} causing droplet separation. The transition criterion is then (Brauner[12], 2001):

$$d_{\max} \le d_{crit} \tag{16}$$

Equations (12) and (15) into (16) give us means to determine the critical velocity.

In this model oil density was found to affect the critical velocity significantly while surface tension and viscosity had smaller effects. The critical velocity increases as water cuts do.

The models described above primarily address the water wetting (entrainment and separation) issue. The question remains however, what is it that makes various crude oils affect corrosion differently even when tested at the same hydrodynamic conditions. Indeed there must be something going beyond the entrainment/separation issue.

In a previous work presented by Hernández *et al* in 2002[17] an insight was given into the variables in crude oil composition that could be playing a major role in the corrosion inhibition offered by crude oils. In that work, a statistical analysis was performed with several Venezuelan crude oils evaluated experimentally under the same conditions. Crude oils were separated in two groups: paraffinic and asphaltenic depending on their distribution of saturates, aromatics, resins and asphaltenes (SARA) and then the effect of basic chemical and physical properties of crude oils were evaluated by using multiple linear regression analyses. The variables evaluated included:

a) SARA Analysis
b) API density
c) Total nitrogen.
d) Sulfur content
e) Total Acid Number (TAN)
f) Concentrations of Vanadium and Nickel
g) % crude oil

The type of crude oil, either asphaltenic or paraffinic, and its concentration were found to be significant parameters when evaluating crude inhibiting effects. For asphaltenic crude oils, having more than 3% of asphaltenes in their composition, the parameters influencing the most at low crude oil concentrations were the sum of resins + asphaltenes, and sulphur content. At high concentrations, the effect of sulphur content was not significant and the sum of resins and asphaltenes dominates the inhibiting process.

In this case the inhibiting capacity, defined as the ratio between corrosion rates with and without crude oil present was described by the following equation:

$$Inhibiting\ Capacity\left(asphaltenic\right) = 0.48 + 0.004 \times \left(\%crude\,oil\right) + 0.0048 \times \left(re\sin + asphaltenes\right) + 0.0436 \times \left(S\right) \quad (17)$$

For those considered as paraffinic crude oils, having more than 50% of saturated components, the total nitrogen content, resins and asphaltenes where the variables found to influence the most. At low crude oil concentrations, nitrogen was found to be the most significant parameter, indicating that the inhibition process can be related to the adsorption of nitrogen based compounds at low concentrations. At higher oil concentrations, nitrogen content becomes less significant and the inhibiting capacity can be related to the formation of stable emulsions, due to the presence of polar functional groups found in the resin fraction.

The inhibiting capacity of paraffinic crude oils was described by:

$$Inhibiting\ Capacity\left(paraffinic\right) = -0.230 + 0.0026 \times \left(\%crude\,oil\right) + 0.0576 \times \left(re\sin\right) + 0.000844 \times \left(N\right) - 0.333 \times \left(asphaltenes\right) \quad (18)$$

The results from this model are plotted in **Figure 2.** The correlation coefficient (R-square) values, however, were in the order of 50% which suggested that the relationship between the predictor and the response was not really linear and that more sophisticated analysis tools should be used. This brought us to an idea of using an Artificial Neural Network (ANN) to analyze the complex nature of crude oils and its relationship to the inhibition.

An Artificial Neural Network (ANN) is an intelligent data-driven modeling tool that is able to capture and represent complex and non-linear input/output relationships. ANNs are massively parallel, distributed processing systems that can continuously improve their performance via dynamic learning. ANNs are used in many important applications, such as function approximation, pattern recognition and classification, memory recall, prediction, optimization and noise-filtering. ANNs are suited for applications involving complex systems. They are used in many commercial products such as modems, image-processing and recognition systems, speech recognition software, data mining, knowledge acquisition systems and medical instrumentation, etc [18,19,20,21]. A key advantage of the ANN is its ability to learn, recognize, generalize, classify and interpret incomplete and noisy inputs (data). The ANN can acquire information/knowledge about a given process through data in the training phase. This information is then stored in numerical form within the weights. ANNs have proven to be powerful tools for data exploration with capability to discover previously unknown dependencies and relationships in datasets. In cases where the input-output relations are mathematically complex, training data is noisy and attributes are real valued, neural network approaches have been reported to perform favorably to other approaches.[22,23]

The ANN is composed of several layers of processing elements or nodes. The Processing Elements (PE), which contains the transfer function, are linked by connections, with each connection having an associated weight, W_i. The weight of a connection expresses relative strength of the input data or transfer data from layer to layer and output. The ANN can appear in many configurations called architectures. These architectures can have many different transfer functions, different number of input PE's, Output PE's, Hidden PE's and hidden layers. **Figure 3** shows a simple PE having n weights, $\{w_1, w_2 \ldots w_n\}$ and **Figure 4** illustrates a typical ANN with 5 inputs, 1 hidden layer with 5 processing elements and 1 output node.

Neural networks have two distinct phases of operation: *training* and *testing*. Typically, a number of key design parameters need to be chosen before training the network such as:

1. System architecture (topology)
2. Training algorithm
3. Number of training cycles (epochs)

During the learning phase, the network learns by adjusting the weights so as to be able to correctly predict or classify the output target of a given set of input samples. With supervised learning, the network is able to learn from the input and the error (the difference between the output and the desired response). One distinguishing characteristic of an ANN is their adaptability, which requires a unique information flow design depicted in **Figure 5**. The performance feedback loop utilizes a cost function to provide a measure of deviation between the calculated output and the desired output. This performance feedback is utilized directly to adapt the parameters, weights and biases, so that the system output improves with respect to the desired goal.

Once a network is trained, it is ready for the testing phase. The task of the network in the testing phase is to produce an output, given an input, based on the model or hypothesis learned during training. It is important to note that unlike in the training phase, the network parameters remain unchanged during the testing phase.

METHODOLOGY

Corrosion Data

The detailed description of the corrosion tests and results used to train the neural networks were published in a previous paper[15]. Only a summary is presented below to get the reader a flavor of the procedure. Fifteen Venezuelan crude oils were evaluated (see Table 1). An analysis of saturated components, aromatics, resins and asphaltenes (SARA) was performed to each crude oil. API density (˚API), total nitrogen content (N_{TOTAL}), Total Acid Number (TAN), sulfur content (S%), vanadium (V), nickel (Ni) were measured according to ASTM standards.

Weight loss corrosion tests were performed on coupons exposed in autoclaves, with several crude oil-saline solution (3.5% NaCl) mixtures, simulating average conditions of well heads (72 psi CO_2 and 80°C). Water cuts were higher than 20% in all cases, and total volume inside the autoclave was always kept at 1.5 L. Test sequence was 20, 50, 80 and 99% of water (v/v), except on the cases of very heavy crude oils (Cerro Negro, Boscán y Zuata) that could only be tested at 80 and 99% water. A rotating speed of 500 rpm was kept in order to get a homogeneous mixture of crude oil and saline solution.

P-110 and also L-80 were the steels selected for this study, their compositions are listed on Table 2, but no significant differences were found between the two types of steel, and only P-110 results are shown in this paper. Three coupons were used for each set of testing conditions; two of them were used for corrosion rate calculations and the third for surface analysis and corrosion product characterization. Coupons were ground using a Silicon Carbide sand paper of grid 600, then cleaned with acetone, distilled water and dried. Their dimensions were taken and their weight determined using an analytical balance.

The coupons were accommodated in the autoclave using the Teflon holder, the solution was poured and the autoclave was then closed and introduced in the heater assembly. The autoclave was purged with CO_2 for 30 minutes, to remove the air that could be inside. After deaeration, the equipment was pressurized until a pressure of

72psi CO_2 was reached and maintained, and then temperature was raised until 80°C in a one-hour time period. Once the conditions were found, the test was performed with a total time of 120 hours. De-scaling of the coupons for the calculation of corrosion rates was made according to the standard ASTM G1-90.

After calculation of corrosion rates, these were translated into inhibiting capacity, by dividing the values of each test by the value obtained in blank tests, so that:

$$Inhibiting\ Capacity = \mathbf{1} - \frac{corrosionrate_{withcrude}}{corrosionrate_{blank}} \quad (19)$$

Neural Network Development

The methodology consists of statistical analysis of the data and development of the neural network model, and can be categorized into the following steps:

1. Preliminary Analysis of the data.
2. Statistical Analysis using commercially available statistical software.
3. Neural Network model construction.
4. Network Optimization using Genetic Algorithm.
5. Multiple tests runs on the selected model.
6. Sensitivity Analysis to find how input variables (chemical constituents) affect the output (percentage inhibition) and their correlation with the other variables.

Step 1: The Preliminary Analysis of the Data

In the preliminary analysis various graphs and scatter plots of the data were examined to find out obvious patterns that might exist within the data.

In this case no distinction was made between paraffinic and asphaltenic crude oils, as in previous work, hoping that the ANN would generate a model that could be indistinctively used for all types of crude oils.

Step 2: Statistical Analysis of the data

Statistical Analysis of the data was performed using MINITAB® [24]. Multiple regression analysis was performed to come up with a regression equation that would be able to explain how the model could be augmented by knowing any possible linear relationships among each of the input variables and the output.

Multiple Regression Analysis Results:

The regression equation is:

```
%Inhibition =  71.7 + 0.0231 API - 0.0695 S(%) + 0.000086 total + 0.0576 TAN - 0.722
               Saturates - 0.714 Aromatics - 0.700 resins - 0.727 asphaltenes +
               0.000227 V - 0.00472 Ni + 0.00365 %Crude oil
```

Predictor	Coef	SE Coef	T	P
Constant	71.72	34.64	2.07	0.041
API	0.023113	0.005457	4.24	0.000
S(%)	-0.06953	0.09460	-0.73	0.464
total	0.00008586	0.00006380	1.35	0.181
TAN	0.05759	0.01581	3.64	0.000
Saturates	-0.7223	0.3464	-2.09	0.039
Aromatics	-0.7136	0.3450	-2.07	0.041
resins	-0.7000	0.3446	-2.03	0.045
asphaltenes	-0.7273	0.3386	-2.15	0.034
V	0.0002274	0.0002177	1.04	0.299
Ni	-0.004718	0.003407	-1.38	0.169
%Crude oil	0.0036530	0.0004038	9.05	0.000

R-Square = 55.0%

Here, **Coef** is the regresión coefficient for a given variable, **SE Coef** is the standard error of the coefficient.

The **t-value (T)** is used to **c**ompare the t-value to the t-distribution to determine if a predictor is significant. The bigger the absolute value of the t-value, the more likely the predictor is significant. The formula is

$$\frac{\text{Estimated coefficient}}{\text{Standard error of the coefficient}}$$

The **p-value (P)** is the probability value and it is often used in hypothesis tests to help decide whether to reject or fail to reject a null hypothesis. The p-value is the probability of obtaining a test statistic that is at least as extreme as the actual calculated value, if the null hypothesis is true. The smaller the p-value, the smaller the probability is that one would be making a mistake by rejecting the null hypothesis. A commonly used cut-off value for the p-value is 0.05. For example, if the calculated p-value of a test statistic is less than 0.05, the null hypothesis is rejected.

The p-values for the estimated coefficients of API, TAN and Crude Oil are 0.000, indicating that they are significantly related to % Inhibition. The p-values for V, Ni, Total, S% are >0.05, indicating that these are not related to Inhibition at a-level of 0.05.

The R-Square value obtained was 55%, which is fairly low suggesting that the relationship between the predictor and response variables is not linear. The R-Square value of 55% implies that only 55% of the variability in the output could be captured and explained by this linear model. In addition, to see if the model could be improved to better explain the relationship, the model was modified to use power transformations, which are transformations that when applied to a data set can often yield a data set that does follow approximately a normal distribution. The BOX-COX and BOX-TIDWELL transformations[25] were performed on the original regression model. The results did not improve reinforcing the fact that the relationship between the predictor and response variables is not linear. Lastly, stepwise regression was performed to consider reducing the model size by

eliminating some of the input variables (within the scope of the analysis). Once again, there was no significant improvement in the R-Sq value.

Step 3: Neural Network Model Construction

A number of different architectures, such as Multilayer Perceptron[26], Generalized Feed Forward Network, Modular Neural Network and Radial Basis Function[26], were considered for the ANN model. The type of ANN architecture used for the analysis of Corrosion Inhibition for Crude Oil is the Multilayer Perceptron (MLP), so only this one will be described here. MLPs are layered feedforward networks typically trained with backpropagation (learning algorithm). MLPs have been proven to be universal approximators[27], capable of implementing any given function through the use of various non-linear transfer functions. One of the most commonly used functions are the hyperbolic tangent function. The hyperbolic tangent function compresses a unit's net input into an activation value in the range [-1, 1]. A number of variations were considered and tested by varying the quantity of hidden layers and PEs.

From the whole database, 65% of data points was used for training the neural network, 15% for cross validation, and 20% for testing.

Step 4: Network Optimization using Genetic Algorithm

Once the MLP structure was chosen as the best model, the next step was to optimize the network architecture to select the best number of processing elements (PE's). Using the Genetic Algorithm Module, within the Neuro-Solutions software (www.nd.com), the MLP with two hidden layer (6 PEs each) was shown to model the crude oil behavior resulting with the best accuracy in predicting inhibition rate (see **Figure 6**)

Step 5: Multiple tests run on the selected model

After network optimization, a number of runs on the selected ANN were performed by randomizing the data, to ensure that the network was able to understand, interpret and learn from the data. From those runs six models (NN Tests 1 to 6) were chosen as the best models generated.

Step 6: Sensitivity Analysis

Sensitivity analysis is a method for extracting the cause and effect relationship between the inputs and outputs of the network. The network learning is disabled during this operation such that the network weights are not affected. The basic idea is that the inputs to the network are shifted slightly and the corresponding change in the output is reported either as a percentage or a raw difference. The activation control component generates the input data for the sensitivity analysis by temporarily increasing the input by a small value (dither). The corresponding change in output is the sensitivity data. Each input channel to the network was varied between its mean ± 1 standard deviation, while all other inputs were fixed at their respective means values. Sensitivity Analysis was performed for the chosen MLP network and for all the test runs for that particular model. The sensitivity was computed based on the corresponding difference (delta) in the output(s) as graphed using the Max-Min criteria of the output (inhibition). A Cumulative Sensitivity graph was constructed by averaging the sensitivity values for all the selected NN test runs (Test 1 to 6).

In addition, the data was subdivided on the basis of the crude oil percentages for further analysis. The data was separated by Crude Oil Percentage into 4 different groups: 1%, 20%, 50% and 80%. A sensitivity analysis was

also performed on these groups. Based on the results, similar behavior patterns were noted between the 1% and 20% crude oil data and similarly between the 50% and 80% crude oil data analysis. The similar groups were combined (1% and 20%) & (50% and 80%) and another sensitivity analysis was performed to see if there were similarities between low or high crude oil concentrations.

RESULTS AND DISCUSSION

As described in the previous section, several NN tests were performed and the sensitivity about the mean for each case was calculated. **Figures 7 to 9** show an example of the results (NN Test 6). For this particular example the *R* value was 0.967 and as can be seen from **Figure 7** the prediction of the neural net is really accurate. From the sensitivity graph (see **Figure 8**) it can be inferred that the variables having the greatest influence in the response were crude oil percentage (%crude oil), Ni content, API gravity and total nitrogen, in that order. It is important to stress that the results shown here are representative of the data from these experiments, related to a sample of Venezuelan crude oils only. The results are not intended to be taken as universal until the model is completely developed, and additional data are used to calibrate it.

Figures 9 to 19 show the separate sensitivity for each variable. According to this NN test an increase in %crude oil will cause an increase in the inhibiting capacity (see **Figure 9**), and this trend was clearly demonstrated in the experimental results. Nickel content seems to diminish the inhibiting capacity, then showing a decreasing effect into inhibition as its content increases in oil (see **Figure 10**). API (**Figure 11**) and total Nitrogen (**Figure 12**) seem to improve corrosion inhibition. Nickel in crude oils can be seen as a measure of the metal/porphyrin complexes compounds containing either Ni or Vanadium and there has been evidence indicating that these compounds are surface active and could help increase the tendency of the crude oil to wet the surface of the metal. However, in practical application, Vanadium and Nickel are usually measured because these materials have serious deleterious effects on catalyst performance during refining by catalytic processes.

The effect of Vanadium (**Figure 13**) and Total Acid Number (**Figure 14**) resulted in an increase of the inhibiting capacity; however the effect is very small as can be seen for the values in the y-axis. The content of %S in the range tested (**Figure 15**) showed to decrease the inhibiting capacity.

In regards to the SARA components of the crude oil, none showed a significant effect, however saturates (**Figure 16**) showed to decrease the inhibiting capacity as their content increases, contrary to Aromatics (**Figure 17**), Resins (**Figure 18**) and Asphaltenes (**Figure 19**) which show to increase inhibition as their content increases.

Figure 20 shows the results of the neural networks in comparison with the data for the cumulative analysis from six test runs. The application of neural networks allows the user to construct the graph of crude oil vs. inhibition if all the parameters are known. As most crude oil properties are interrelated and in turn dependent of the origin where the crude oil comes from, it is important to know all the variables. The cumulative sensitivity, as shown in **Figure 21**, is the result of combining six test and their individual sensitivities. The parameters showing the highest influence were %crude oil, nickel content, API degree and Total Acid Number, in that order.

The tendency of % crude oil vs. inhibition was clear in both the data and the model. An increase in crude oil content increases the degree of corrosion protectiveness by the crude oil. With API density, even if the data is scattered, the model predicts an increase in inhibition as API increases, implying lighter crude oils providing higher values of inhibiting capacities. Regarding Nickel content the effect was persistent through all the analysis, while it was not obvious in the data. In the ranges tested, of 5-110 ppm Nickel, an increase in Nickel content decreases the ability of the crude oil to provide inhibition. This effect is contradictory to what would be expected

if one thinks of nickel as related to metallic complexes that could help changing interfacial properties between water and crude oil thus helping water to get entrained and increasing crude oil wettability.

In order to see if this effect was repeatable, separated sensitivity analyses were performed for the various crude oil contents evaluated: 1, 20, 50 and 80%.

- For 1% crude oil (see **Figure 22**) the model tends to predict a higher inhibiting capacity than the real measured values, but the R value is still considerably high, 0.963. Nickel appears to be most significant followed by API, sulfur content, TAN and asphaltenes. All but Nickel increase the inhibiting capacity when increased in number or concentration.
- For 20% crude oil (**Figure 23**, R=0.998) Nickel is not that critical and the variables influencing the most are API, total Nitrogen, resins and TAN. Saturates seem to have a detrimental effect.
- For 50% crude oil (**Figure 24**, R=0.960) the four variables with the highest sensitivity are Nickel, Vanadium, aromatics and sulfur. Nickel and aromatics decrease the value of inhibiting capacity as their content increases. If only positive effects are considered then V, %S, asphaltenes and resins show the highest influence.
- For 80% (**Figure 25**, R=0.990) Nickel and Vanadium showed the highest sensitivities, in both cases producing a decrease in the inhibiting capacity as their content increases. Asphaltenes follow and then aromatics, the latter also having an inverse relationship. Note that sensitivity values are a lot higher for the first two cases.

An interesting result from the model is that it was able to point out notably different behaviors when the crude oil concentration changes. By putting together the data for low concentrations (1 and 20%, see **Figure 26**) and the data for higher concentrations (50 and 80%, see **Figure 27**) and looking at the sensitivities it can be concluded that at higher concentrations the presence of crude oil per se (coverage) have the greatest influence and the effects of different variables is not as relevant. At low crude oil concentrations the sensitivities are a lot higher (up to 0.8) indicating that inhibition is not as much related to the amount of crude oil but to the presence of oil or a combination of the two.

According to the model, for most cases, Nickel content seems to have a great effect in decreasing the % inhibition provided by crude oil. From the scientific point of view, no explanation can be offered at this point regarding the effect of Nickel unless more experimentation is performed.

If we now return to the multiple regression analysis, it was seen that most of the input variables are highly correlated, which is actually expected based on knowledge about crude oil chemistry. The following relationships were found by analyzing the data on this paper (graphs are not shown for space limitations):

- Increased amounts of aromatic compound result in an increase in density (API density) whereas an increase in saturated compounds results in a decrease in API density.
- Lower API crude oils tend to have higher sulfur contents (%S), asphalt content (asphaltenes and resins), and are associated to higher nitrogen contents.
- As % sulfur increases so does nickel and both Ni and V tend to decrease as API increases.

These highly correlated variables made it hard for the neural network to pick up sensitivities, however by combining the multiple regression results with the neural network results it can be inferred that at low oil concentrations API could be used as the leading variable in the inhibition process. On the other hand all variables evaluated will be affecting inhibition to some degree so all of them would have to be measured and further analysis will have to be performed to see if these co-variations can be systematically predicted.

To illustrate this point **Figure 28** shows an example of the response of the neural network when all input variables are known while **Figure 29** shows an example of a crude oil where only four variables were known (resins,

asphaltenes, nickel content and sulfur) and the rest were deducted from rough approximations using the relationships found from the multiple regression analysis. When all variables are known the response is more accurate, as expected. However, the response in the second case is not bad indicating that by doing a profound analysis of these relationships the model could be improved and the number of inputs could be reduced.

CONCLUSIONS

- Based on experimental data from Venezuelan crude oils an effective neural network model was developed that can predict the ability of a crude oil to provide corrosion protectiveness in a CO_2 environment.

- Provided that the model is fed with enough data from crude oils with different origins, the model would have the capability of creating the curve of %crude oil vs. corrosion rate for a given crude oil by knowing some of its physical and chemical properties, most of them routinely measured.

- Sensitivity analyses were performed to extract the variables having the highest effect on the response. Most of the trends noticed in the experimental data were captured by the network. However, the indicated correlations between the variables could not always be explained and could lead to dubious interpretations. More research is needed in order to expand both the experimental database and the modeling capability with a goal to generate the knowledge necessary to determine the factors governing the effect of crude oil in CO_2 corrosion

FUTURE WORK WITH ANN

ANNs act like "black boxes" in the sense that relationships are encoded incomprehensibly as weight vectors within the trained network. Though artificial neural networks have proven to be empirically successful, they are generally treated as numerical enigmas, often generating incomprehensible and hard-to-understand models. They cannot easily support the generation of scientific theories unless these relationships can be explained in a comprehensible form to aid in the process of prediction. This is primarily because the processing in an ANN occurs at the sub-symbolic level as numerical estimation and manipulation of network parameters. Therefore, it is not always possible to directly translate these large sets of real valued parameters into symbols or concepts which have semantic significance. Thus, lack of transparency and inability to provide an end user explanation capability have generally been recognized as the most significant obstacles to the more widespread application of Artificial Neural Networks.

Rule-extraction can enhance the capabilities of an ANN by developing rules based on the assigned weights. The knowledge embedded within the trained ANN in the form of weights needs to be extracted and expressed as a set of rules. Rule extraction from a trained ANN originated with Gallant in1988, where the rules were used to explain the solutions reached by the network. Basically, knowledge is acquired during the training phase and is then encoded within the network architecture, the activation function associated with each layer, and the set of numerical weights. Rule extraction is based on the behavior of the neurons within the ANN. A significant amount of research has been expended in these recent years to develop techniques for extraction of rules from trained ANNs. Key methods include the search based subset algorithm[28], and the M-of-N algorithm.[29,16] Validity Interval Analysis (VIA)[15], the genetic algorithm approach to rule extraction[30], the connectionist scientist game[31], the rule-extraction as a learning approach [32] and other approaches which extract fuzzy rules like the Fuzzy ARTMAP system.[33]

For future research, the plan is to use various methods such as VIA analysis and Rule extraction using decision trees (C4.5 Algorithm) to further understand the relationships between the input variables and inhibition with crude oils.

REFERENCES

1 K.D. Efird "Preventive Corrosion Engineering in Crude Oil Production. Offshore Technology Conference (OTC) Paper# 6599. 1991

2 J.S. Smart, "Wettability – a Major Factor in Oils and Gas System Corrosion" NACE CORROSION/93 conference, paper # 70.

3 W. Anderson , "Wettability literature survey-Part 1: Rock/Oil/Brine Interactions and the effects of core handling on Wettability. Journal of Petroleum Tecnology, October 1986.

4 W. Anderson, "Wettability Literature Survey-Part 2: Wettability Measurement. Journal of Petroleum Tecnology, October 1986.

5 C.de Waard, U. Lotz; "Prediction of CO2 corrosion of carbon steel". NACE CORROSION/93 conference, Paper # 69.

6 M. Wicks, J.P. Fraser, "Entrainment of water by flowing oil", Materials Performance, 14, 5 (1975) p.9

7 U. Lotz, L. van Bodegom, and C. Ouwehand, "The effect of type of Oil and Gas Condensate on Carbonic Acid Corrosion" CORROSION/90 conference. Paper #41. National Association of Corrosion Engineers (1990).

8 C. de Waard, L. Smith, P. Bartlett, H. Cunningham. "Modelling Corrosion rates in Oil Production Tubing". EUROCORR 2001, Paper # 254. The European Corrosion Congress, European Federation of Corrosion, Italy, October 30, 2001.

9 C. De Waard, U. Lotz, A.Dugstad ; "Influence of liquid flow velocity on CO2 corrosion: a semi-empirical model", NACE CORROSION/95 conference, paper #128.

10 C de Waard, L. Smith, B.D. Craig, "The influence of crude oil on well tubing corrosion rates". EUROCORR 2001, Paper # 174. .The European Corrosion Congress, European Federation of Corrosion, Italy, October 30, 2001.

11 B. Craig, "Predicting the conductivity of Water-in-oil solutions as a means to estimate corrosiveness", Corrosion 54, 8 National Association of Corrosion Engineers (1998), p.657.

12 J. Cai; S. Nesic; C de Waard ; "Modeling of water wetting in oil-water pipe flow", CORROSION 2004 conference, paper # 04663. National Association of Corrosion Engineers (2004).

13 J. Cai; S. Nesic; C. de Waard, ; "Modeling of water wetting in oil-water pipe flow", CORROSION 2004 Conference, paper # 04663.National Association of Corrosion Engineers (2004).

14 N. Brauner, "The Prediction of Dispersed Flows Boundaries in Liquid-Liquid and Gas-liquid Systems", Int. J. of Multiphase Flow, Vol.27, pp. 885~910 (2001)

15 D. Barnea, "A Unified Model for Predicting Flow Pattern Transitions for the Whole Range of Pipe Inclinations", Int. J. of Multiphase Flow, Vol.11, pp. 1~12(1987).

16 R.S. Brodkey , " The Phenomena of Fluid Motions", Adison-Wesley, Reading M.A., 1969.

17 S.Hernández, S. Duplat, J.R. Vera, E. Barón, "A statistical approach for analyzing the inhibiting effects of differente types of crude oil in CO2 corrosion of carbon steel". CORROSION/2002 conference, Paper #02293. National Association of Corrosion Engineers (2002).

18 Efraim, T., Jay E. A., Liang T. P., & McCarthy, R. V. (2001). Decision support systems and intelligent systems. Upper Saddle River, NJ: Prentice Hall.

19 LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989), "Back propagation applied to handwritten zip code recognition", Neural Computation, 1, p.541-551.

20 Pomerleau, D. A. (1989), "ALVINN: an autonomous land vehicle in a neural network", Technical Report CMU-CS-89-107, Pittsburg, PA: Carnegie Mellon.

21 Waibel, A. H. (1989), "Modular construction of time delay neural networks for speech recognition", Neural Computation, 1, p.39-46.

22 Thurn, S. B. (1993), "Extracting provably correct rules from artificial neural networks", Technical Report IAI-TR-93-5, Bonn, Germany: University of Bonn.

23 Towell, G. G., & Shavlik, J. W. (1993), "Extracting refined rules from knowledge-based neural networks", Machine Learning, 13, p.71-101.

24 MINITAB, Statistical Software for Windows, registered trademark of Minitab Inc.

25 D.C. Montgomery, E. A Peck and G. Vining, Introduction to Linear Regression Analysis, Third Edition, John Wiley & Sons, New York (2001)

26 Principe, J.C., Euliano, E.R. &Lefebvre, W.C. neural and adaptive systems: Fundamentals through simulations with CD-ROM. New York, NY: John Wiley & Sons, Inc. (1999).

27 Reed, R. D., & Marks, R. J. (1998). Neural smithing: Supervised learning in feedforward artificial neural networks. Cambridge, MA: MIT Press.

28 Fu, L. M. (1991), "Rule learning by searching on adapted nets. Proceedings of the Ninth National Conference on Artificial Intelligence", Anaheim, CA: AAAI Press, p. 590-595.

29 Craven, M. W., & Shavlik, J. W. (1993), "Learning symbolic rules using artificial neural networks. Proceedings of the Tenth International Conference on Machine Learning", Amherst, MA: Morgan Kaufmann, p. 73-80.

30 Keedwell, E., Narayanan, A., & Savic, D. A. (1999), "Using genetic algorithms to extract rules from trained neural networks", Proceedings of the Genetic and Evolutionary Computing Conference, Orlando, FL: Morgan Kaufmann, p. 793.

31 McMillan, C., Mozer, M. C., & Smolensky, P. (1991), "The connectionist scientist game: Rule extraction and refinement in a neural network", Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, Hillsdale, NJ: Lawrence Erlbaum, p. 424-430.

32 Craven, M. W., & Shavlik, J. W. (1994), “Using sampling and queries to extract rules from trained neural networks”, Proceedings of the Eleventh International Conference on Machine Learning, New Brunswick, N.J., Morgan Kaufmann, p. 37-45.

33 Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J., & Rosen, D. (1992), “Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps”, IEEE Transactions on Neural Networks, 3, September 1992, p. 698-713.

Table 1 Characteristics of the crude oils evaluated.

ND = not detected

	PIC4-99	PIC4-98	SBC52	VLA83	SBC36	Santa Bárbara	*FUL-16*	*Boscán*
Saturated (%)	75.4	53.8	62.2	60.6	64.6	60.5	43.2	12.4
Aromatics (%)	21.9	41	34.8	29.5	32.2	34.1	32.6	46.1
Resins (%)	2.7	5.2	3	9	3.2	5.4	17.6	29.9
Asphaltenes (%)	ND	ND	ND	0.9	ND	ND	6.6	11.6
Sulfur (%p/p)	0.58	0.67	0.566	0.9	0.57	0.585	1.1	4.4
V (ppm)	<10	-	< 10	130	-	5	125	1059
Ni (ppm)	<10	-	< 10	10	-	5	25	90
V/Ni	-	-	-	13	-	1	5	12
Total nitrogen (ppm)	819	-	34	990	628	860	3276	6603
TAN (KOH/g)	0.059	0.0843	0.24	0.046	0.023	0.02	0.11	1.02
°API @ 60°F	35.8	31.9	42.9	33.8	30	36.2	20.7	10
Kerogen Type	II (Marine)	II (Marine)	II (Marine)	II(Marine)	II(Marine)	II(Marine)	II (Marine)	II(Marine)
Maturity	Very mature	Very mature	Very mature	Mature	Very Mature	Very Mature	Very mature	Very mature

	Bachaquero	Menemota	Lagomedio	Pacón-Mara	Zuata	Cerro Negro	Ful29
Saturated (%)	21.1	30	47.4	41.7	13.3	13.3	41.7
Aromatics (%)	50.1	45.1	39.9	41.3	47.8	50.9	36.8
Resins (%)	19.7	16.4	10.9	11.5	28.8	23.3	18.6
Asphaltenes (%)	9.2	8.6	1.9	5.4	10.1	12.4	3
Sulfur (% p/p)	2.6	2.42	1.45	2.54	3.7	4.1	0.5
V (ppm)	409	362	232	336	458	474	48
Ni (ppm)	44	44	30	28	95	102	< 10
V/Ni	9	8	8	12	5	5	-
Total nitrogen (ppm)	4340	3450	1693	2334	6948	6424	1910
TAN (KOH/g)	4.78	0.74	0.14	0.40	3.35	3.67	0.49
°API@60°F	11	20.4	30.9	27.5	8.5	8.1	15.3
Kerogen type	II(Marine)	II(Marine)	II(Marine)	II(Marine)	II(Marine)	II(Marine)	II(Marine)
Maturity	Very mature	Very mature	Very Mature	Very Mature	Not mature	Not mature	Very mature

Table 2 Chemical Composition of carbon steels used (%w/w)

	C	Mn	Cr	Ni	Si	Cu	S	P
API L80	0.30	1.20	-	0.25	0.45	0.35	0.030	0.030
API P110	0.32	1.24	0.5	-	0.22	-	0.010	-

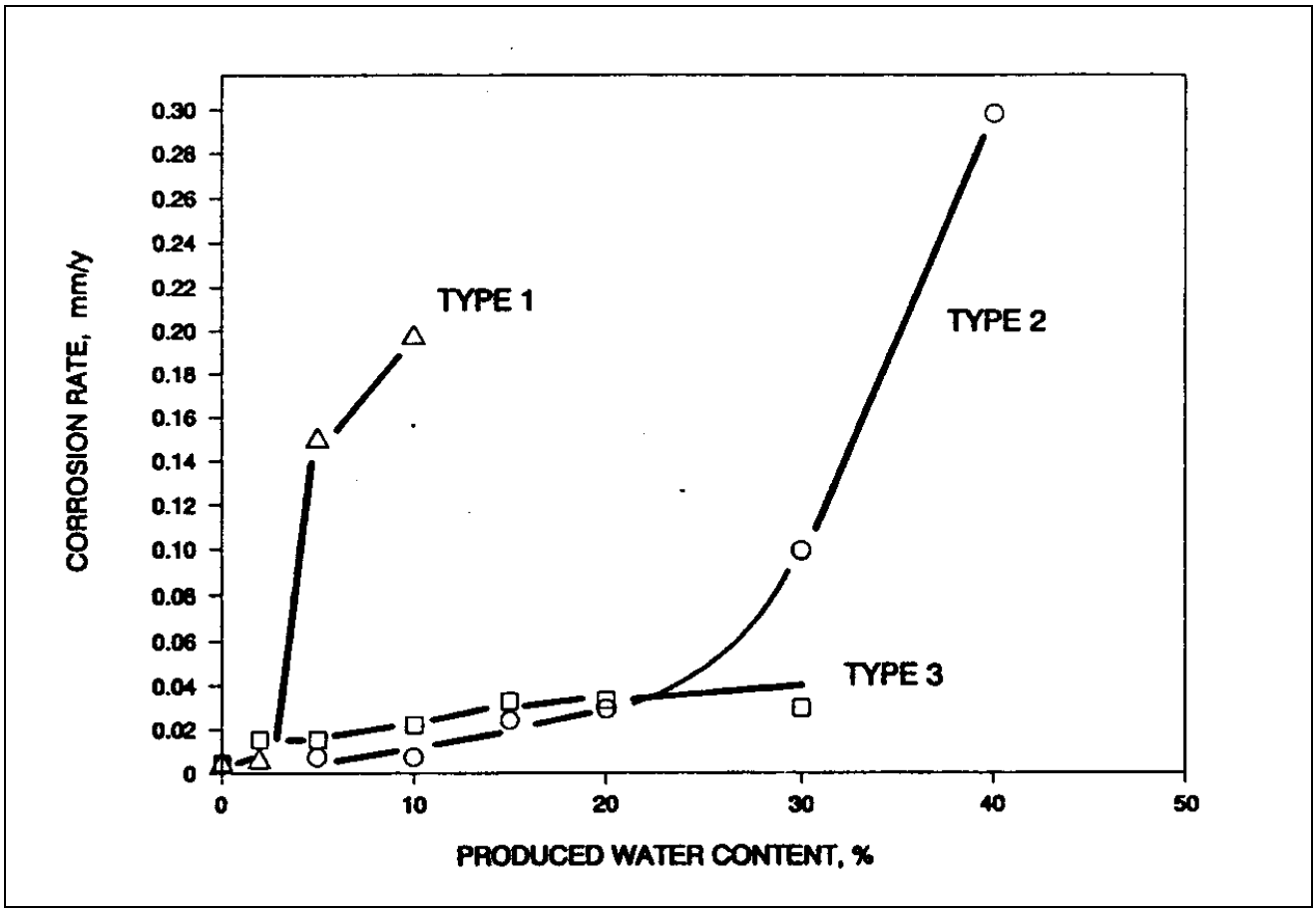

Figure 1 The change of corrosion rate of steel in crude oil/produced water mixtures with increasing produced water showing the three types of behaviour observed.[1]

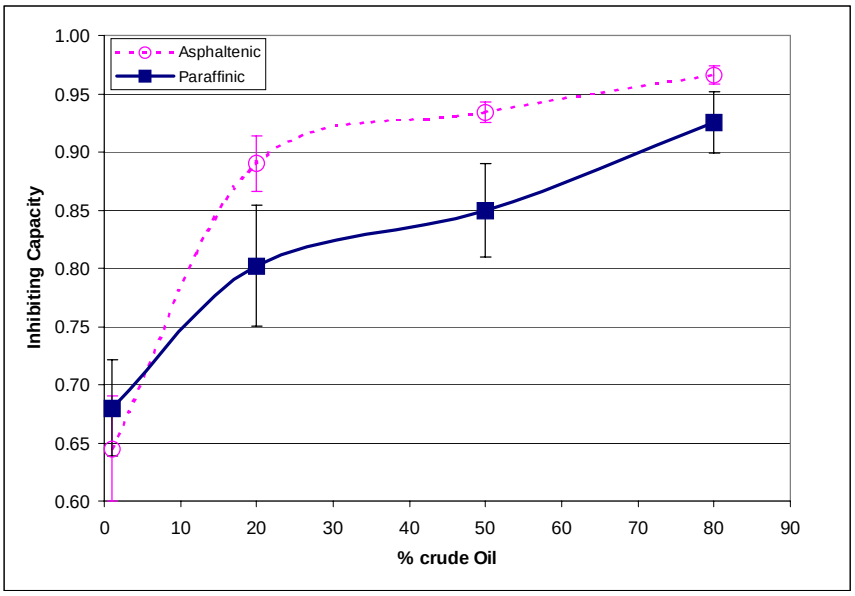

Figure 2 Variation of the inhibiting capacity with crude oil content for asphaltenic and paraffinic crude oils.[15]

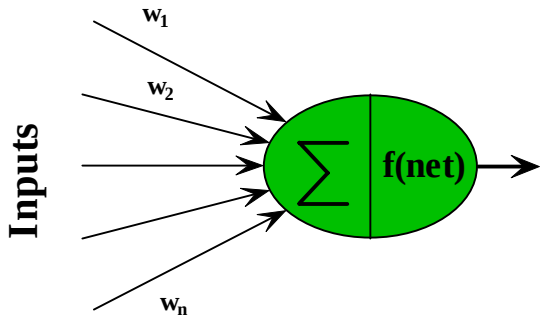

Figure 3 A typical processing element (PE) in an Artificial Neural Network showing multiple inputs, one output node.[26]

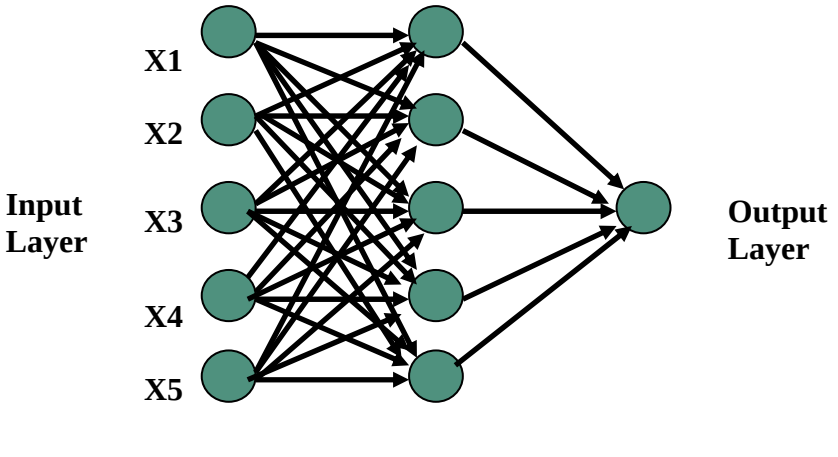

Figure 4 A typical ANN with 5 inputs, 1 hidden layer with 5 processing elements and 1 output node.

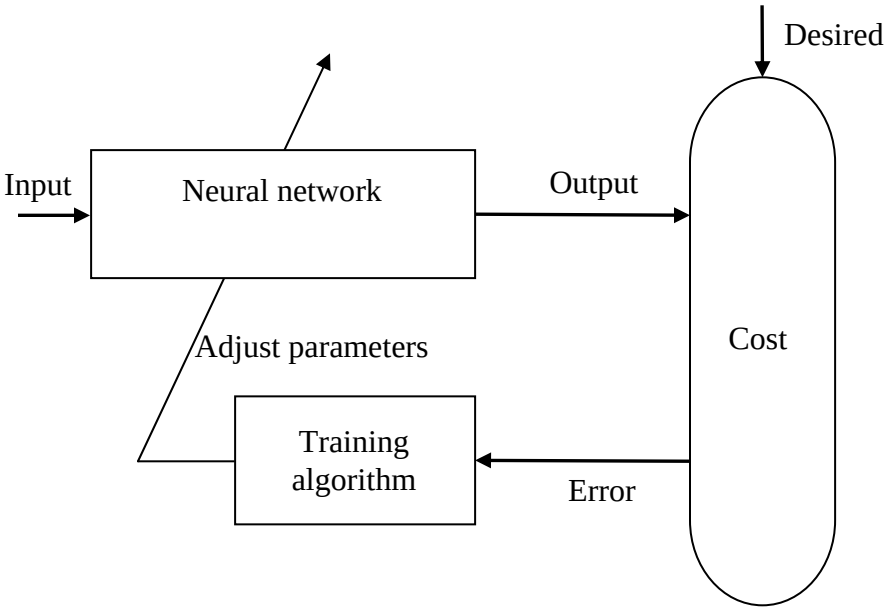

Figure 5 Information flow for training phase
Adapted from Principe et al. [26]

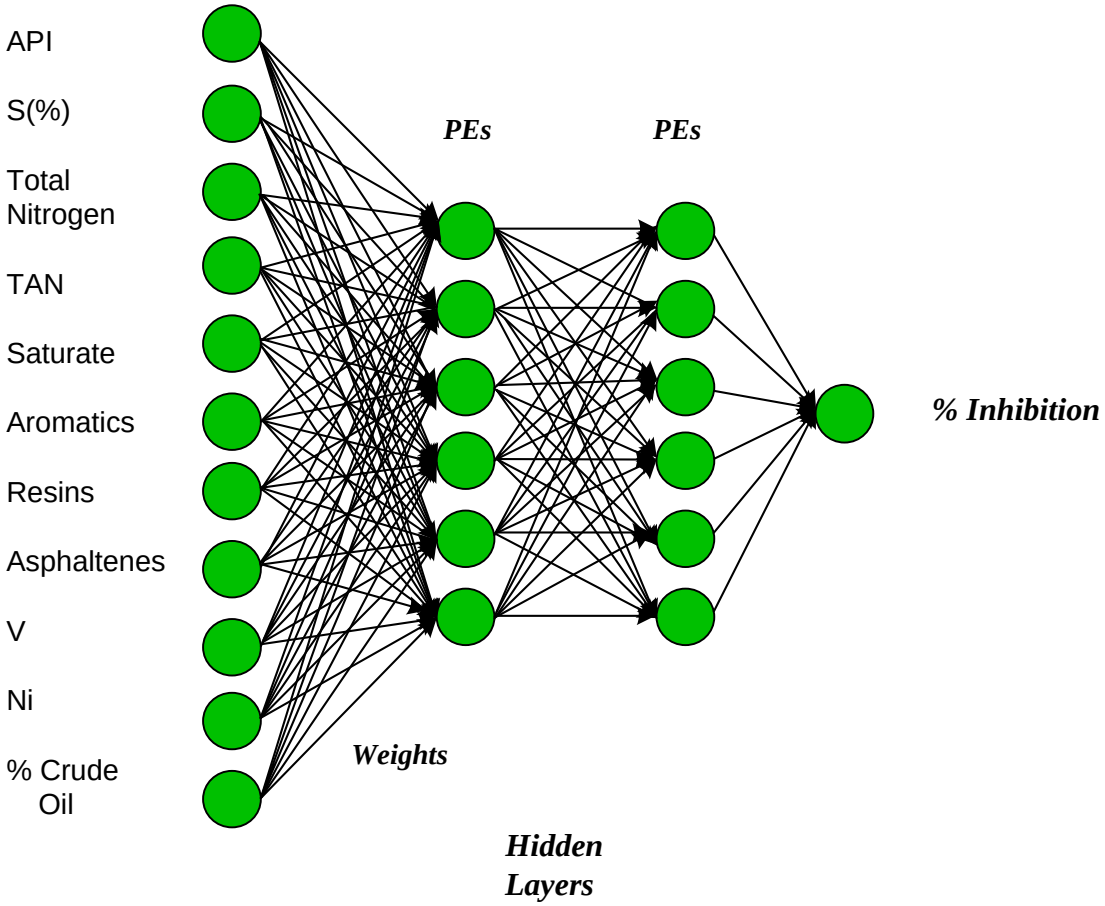

Figure 6: MLP Architecture for Crude Oil Inhibition

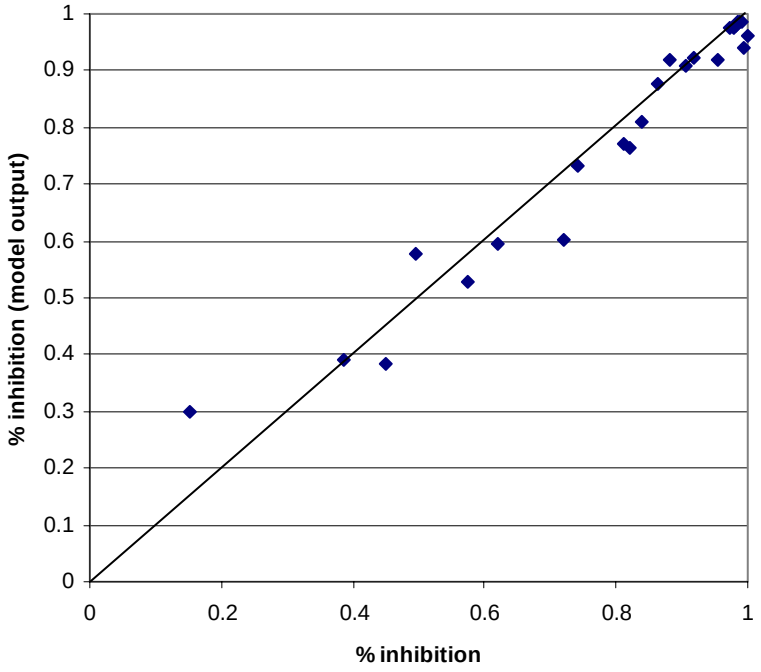

Figure 7 Actual vs. predicted Inhibition for NN test # 6

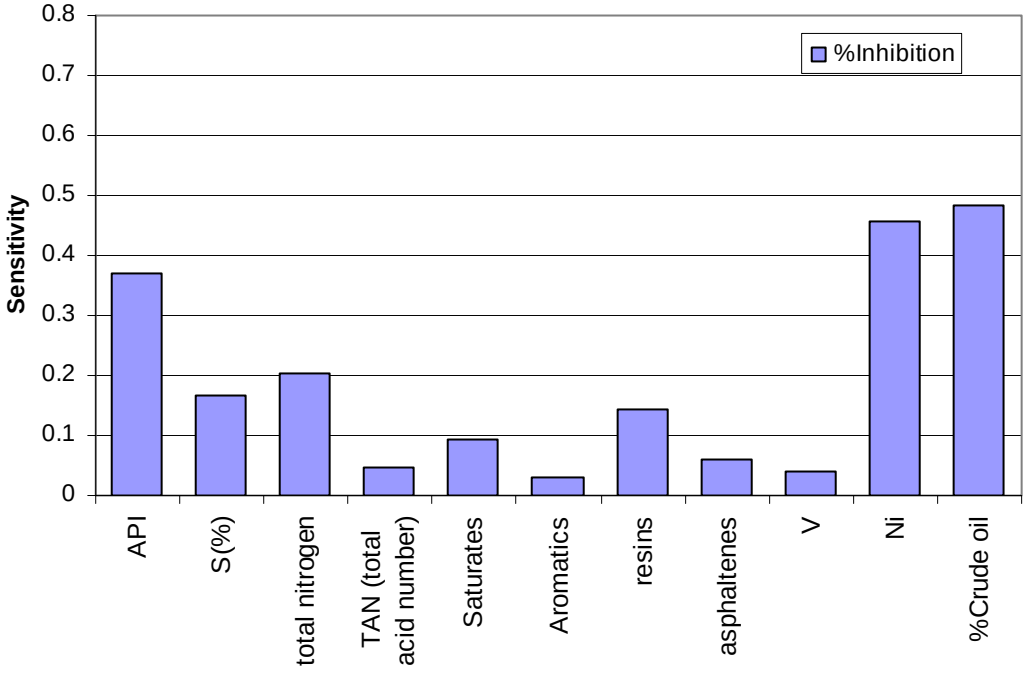

Figure 8 Sensitivity about the mean for test # 6

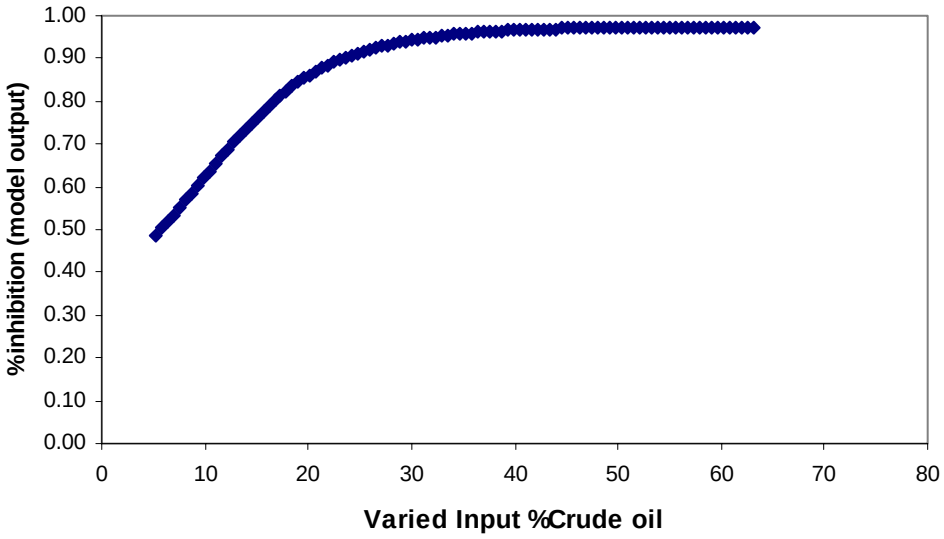

Figure 9 Separate sensitivity for %Crude oil

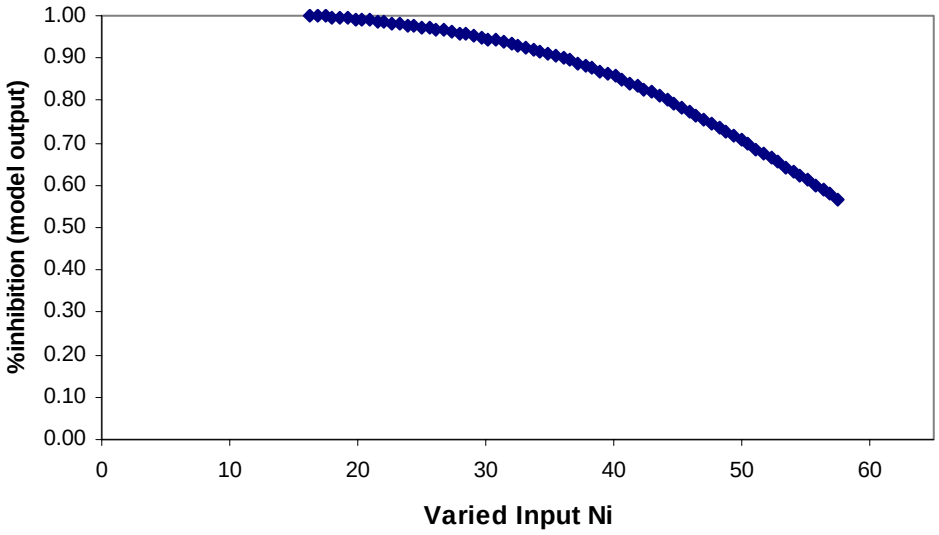

Figure 10 Separate sensitivity for Nickel, test # 6

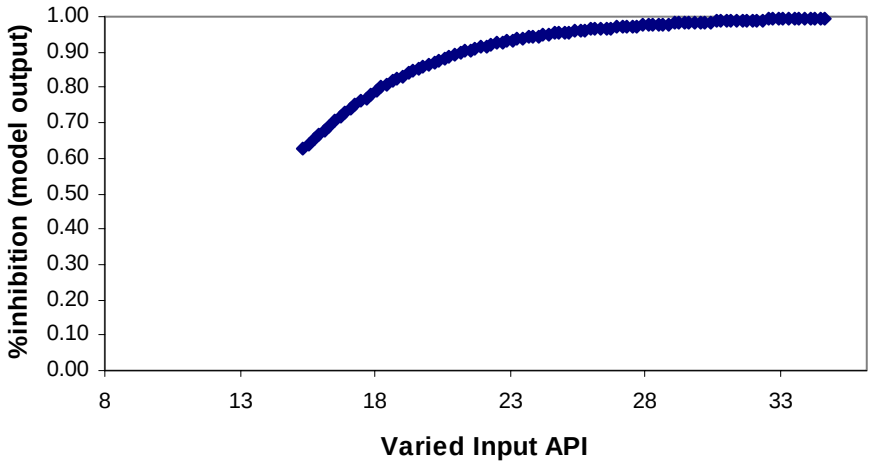

Figure 11 Separate sensitivity for API, test # 6

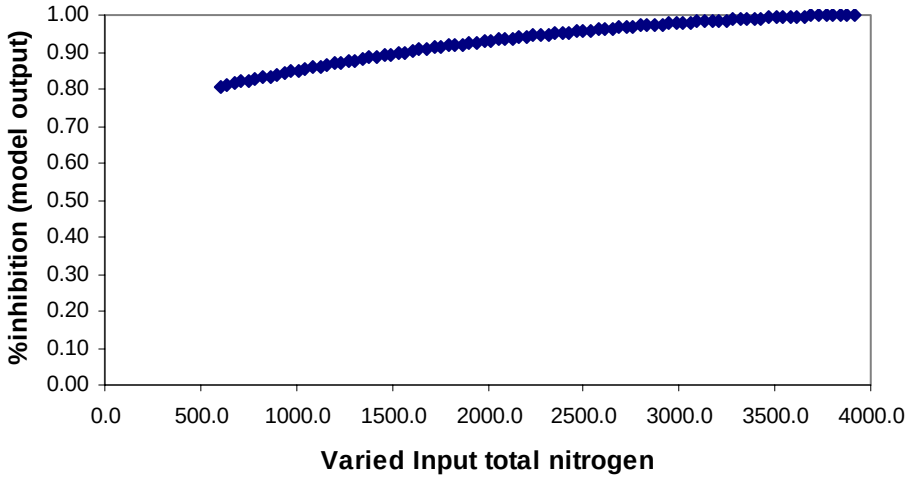

Figure 12 Separate sensitivity for total Nitrogen, test # 6

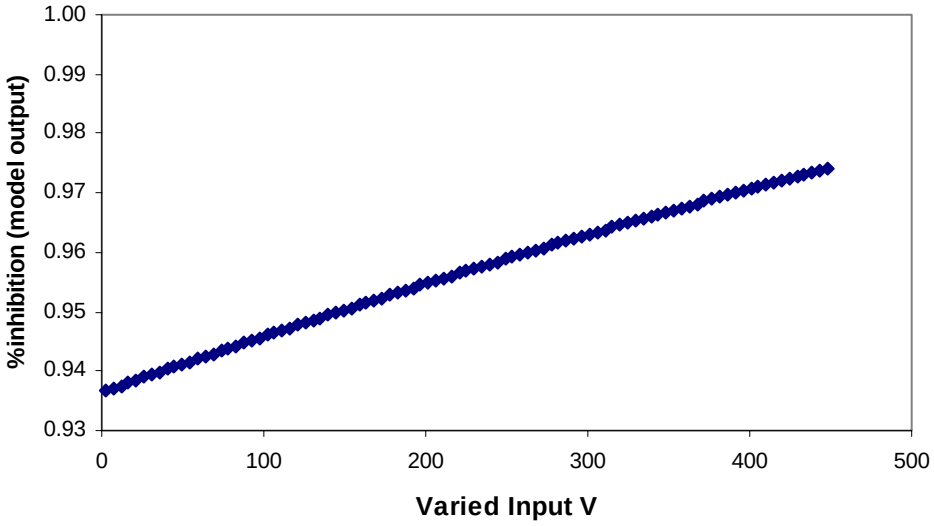

Figure 13 Separate sensitivity for Vanadium, test # 6

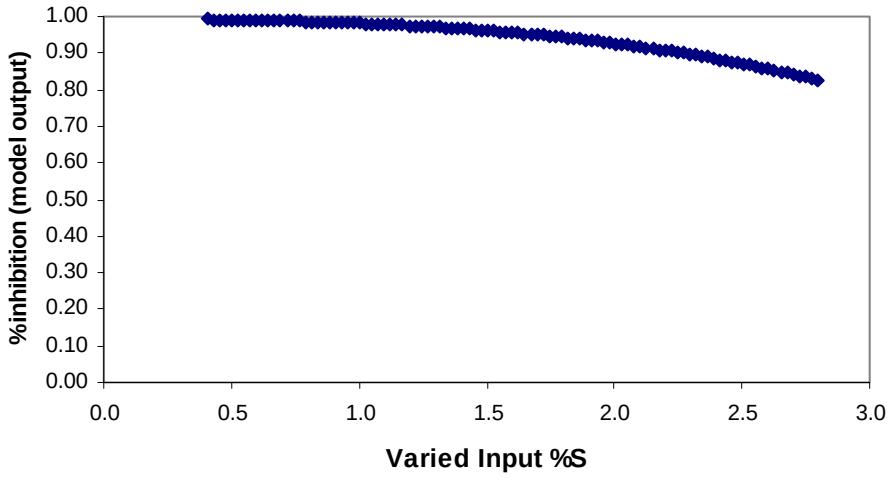

Figure 14 Separate sensitivity for %S, test # 6

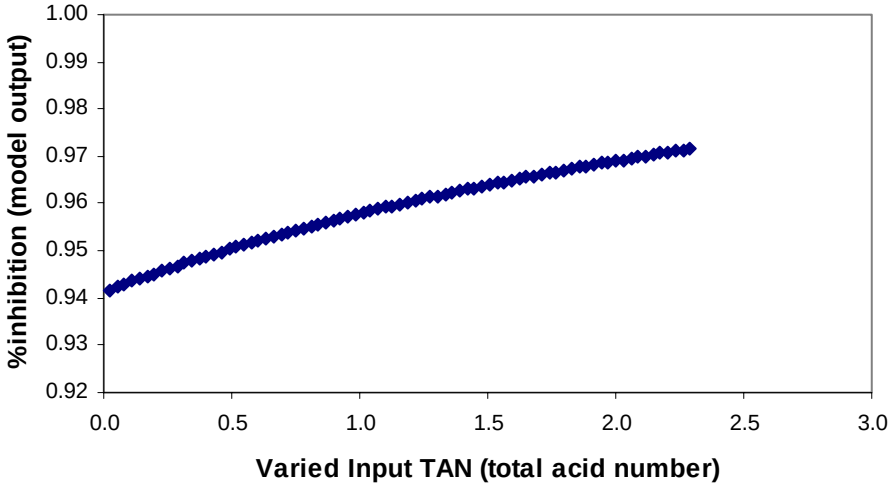

Figure 15 Separate sensitivity for TAN, test # 6

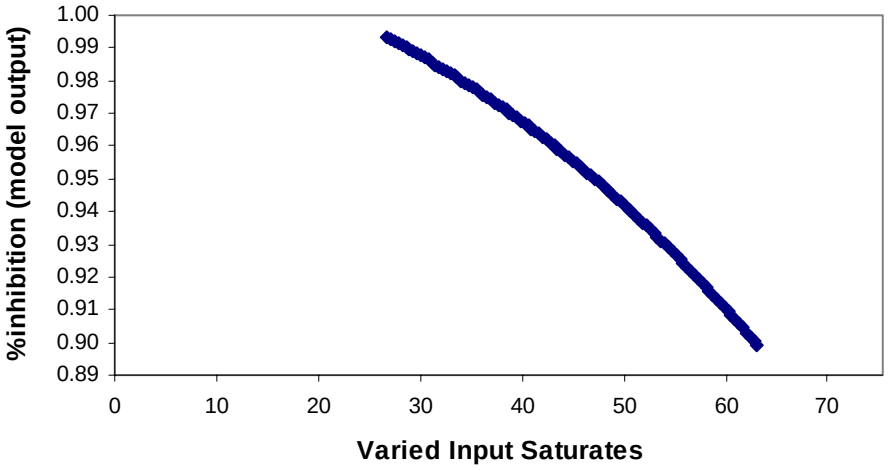

Figure 16 Separate sensitivity for Saturates, test # 6

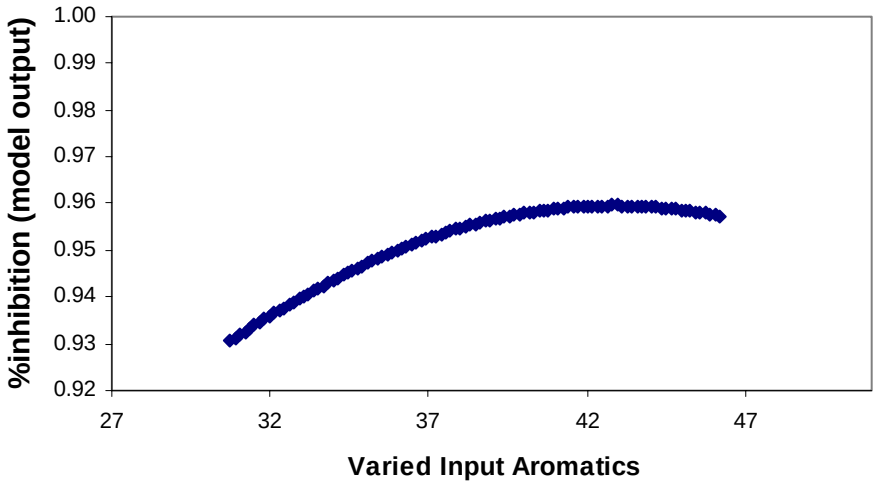

Figure 17 Separate sensitivity for Aromatics, test # 6

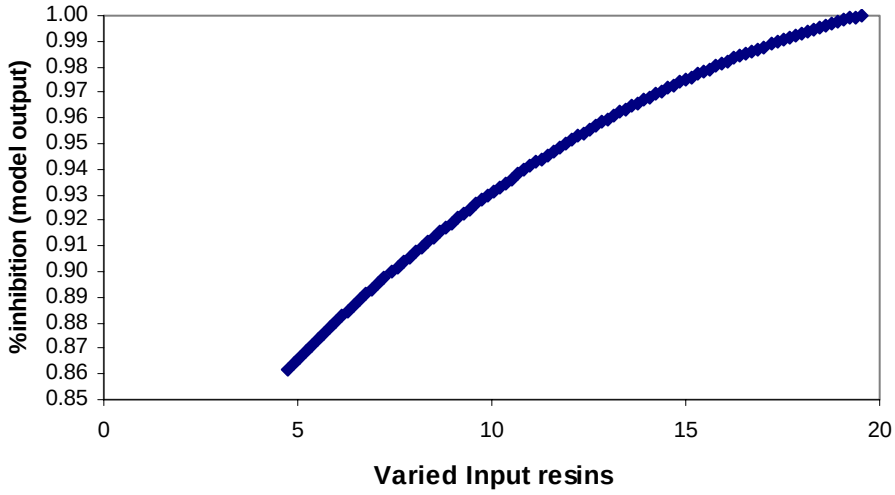

Figure 18 Separate sensitivity for Resins, test # 6

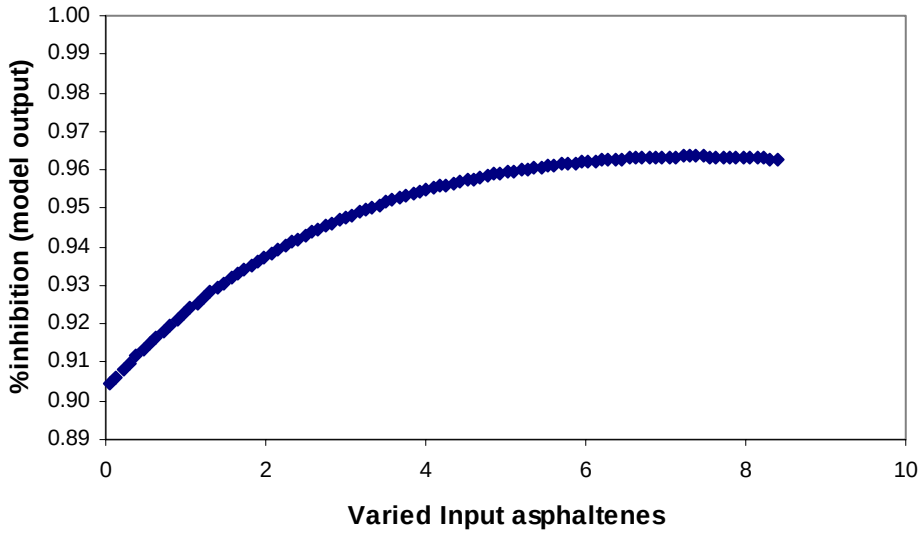

Figure 19 Separate sensitivity for Asphaltenes, test # 6

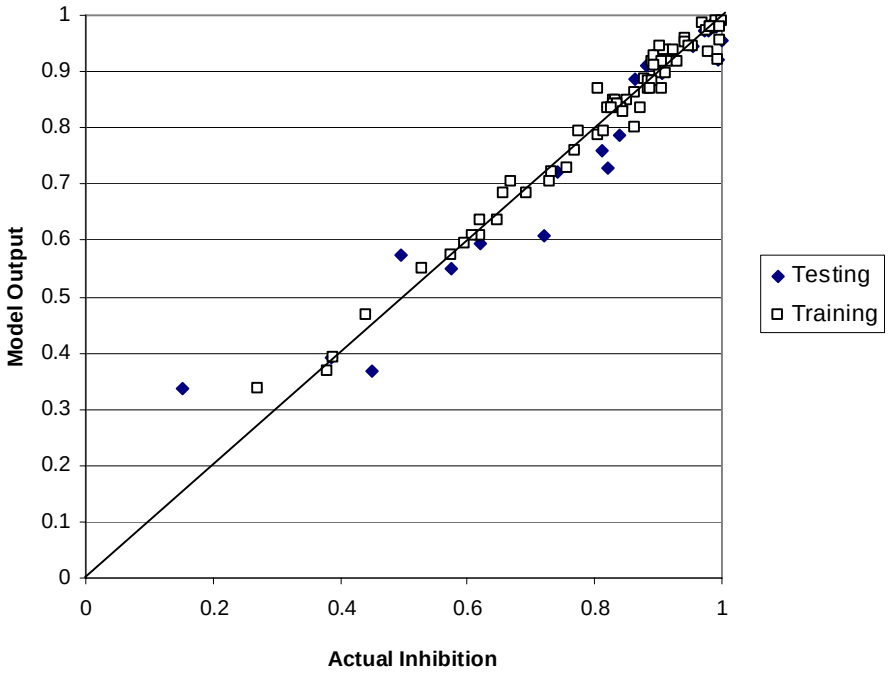

Figure 20 Neural network Model results vs. Actual data

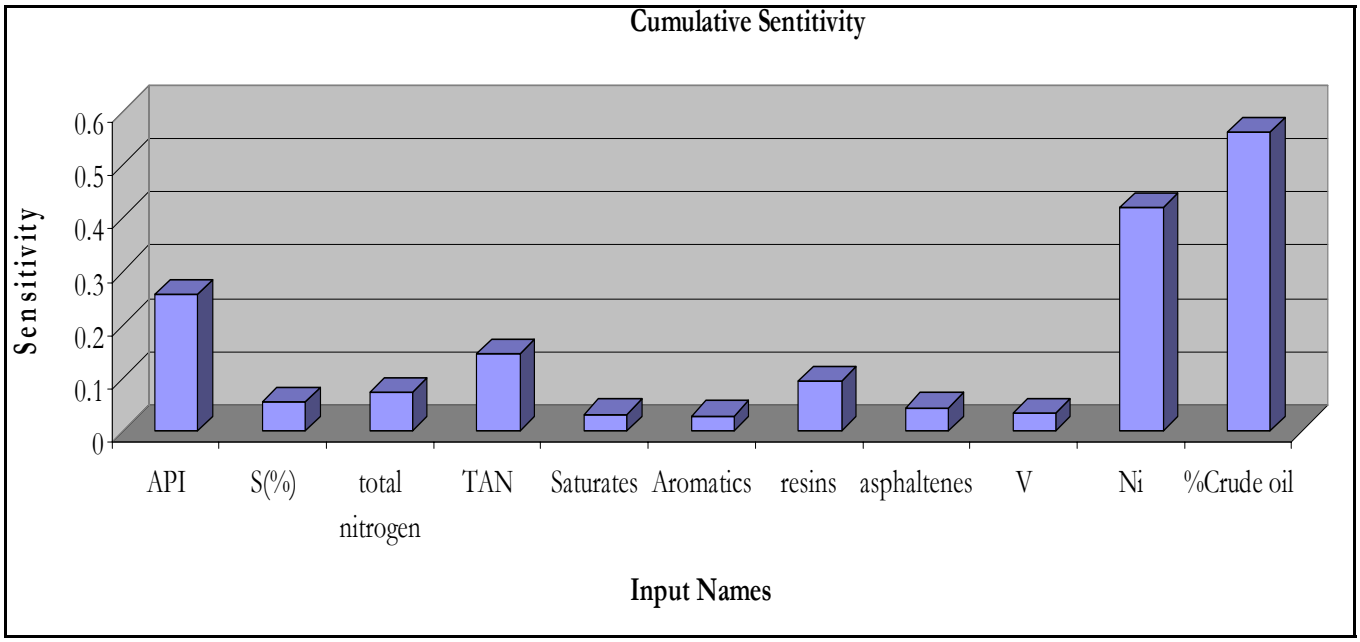

Figure 21Cumulative sensitivity Graph

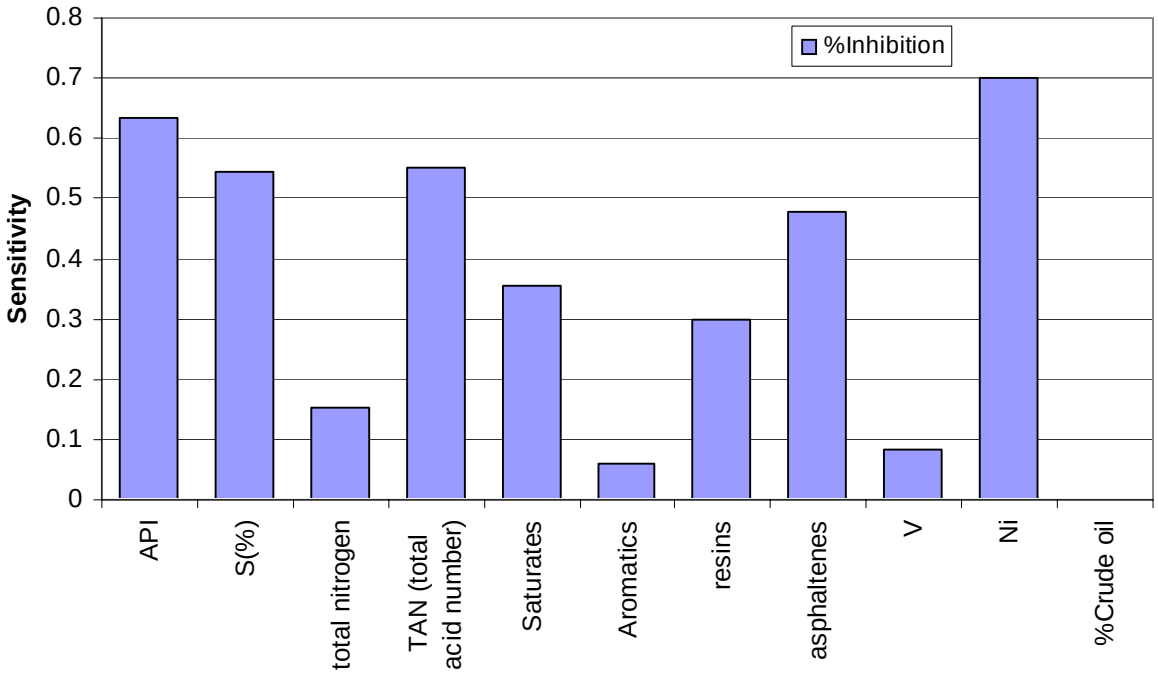

Figure 22 Sensitivity about the mean for 1% crude oil concentration

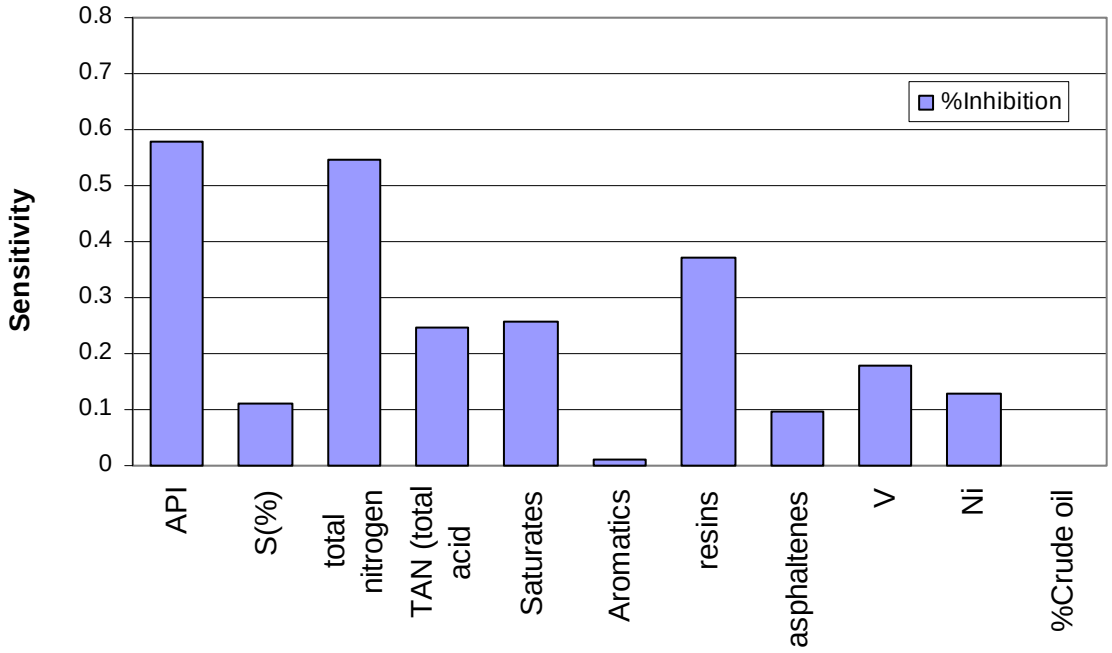

Figure 23 Sensitivity about the mean for 20% crude oil concentration

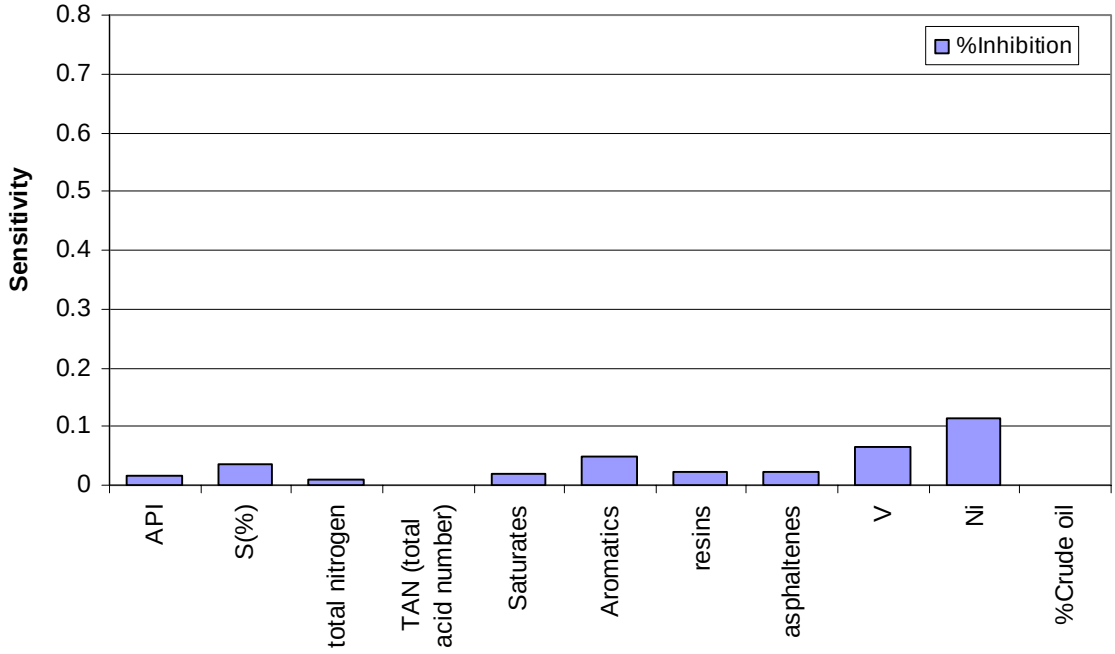

FIGURE 24 Sensitivity about the mean for 50% crude oil concentration

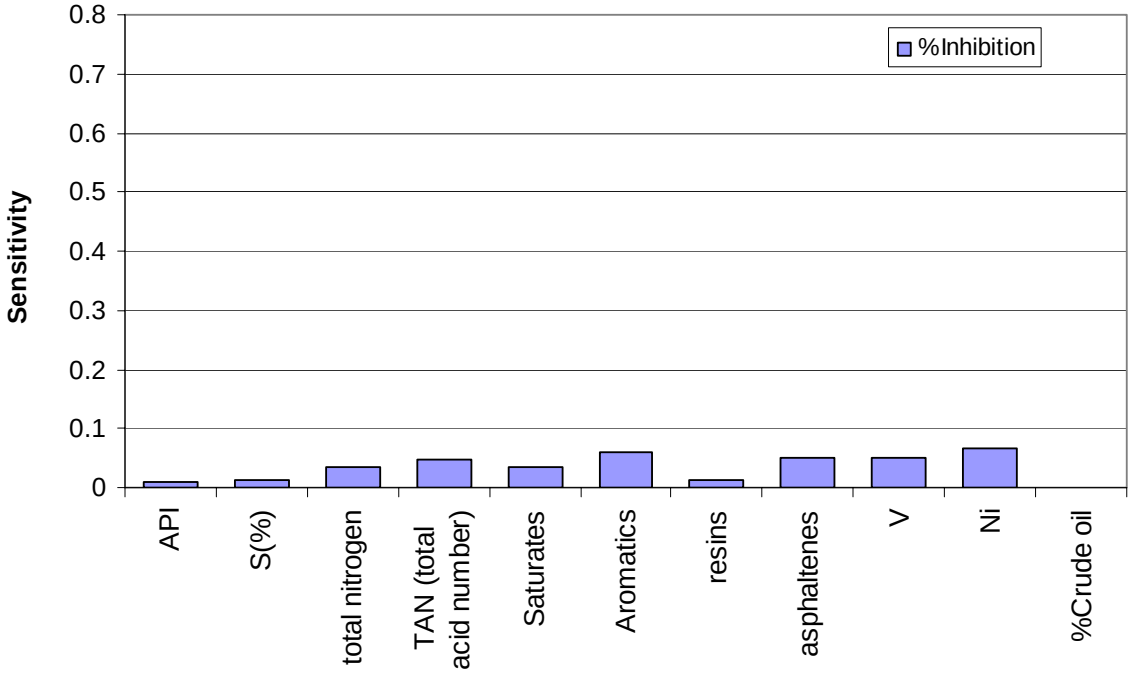

Figure 25 Sensitivity about the mean for 80% crude oil concentration

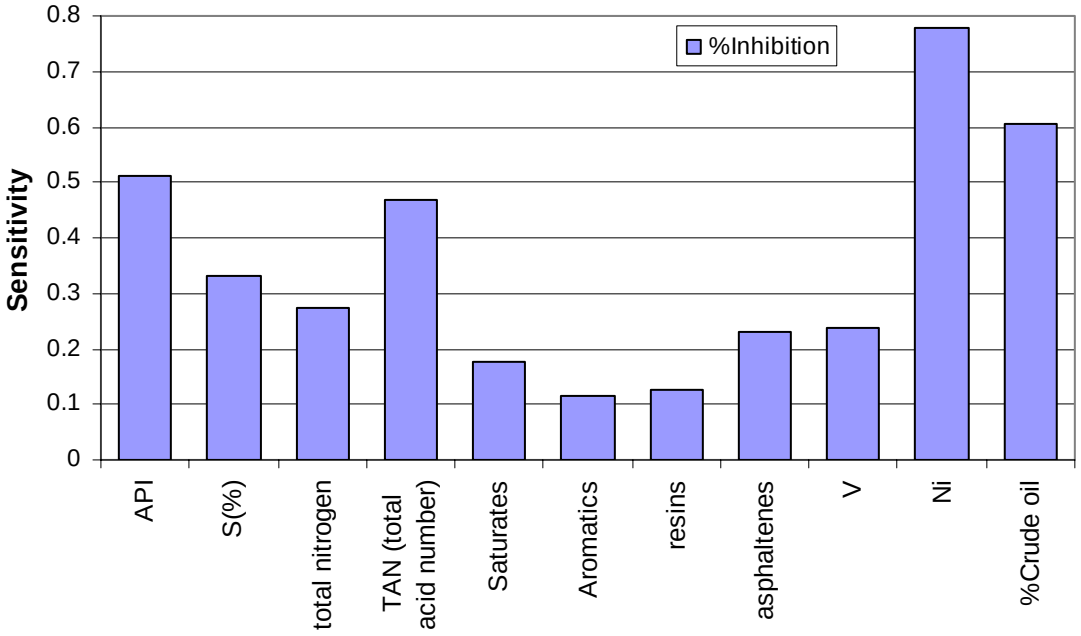

Figure 26 Sensitivity about the mean for 1% and 20% combined

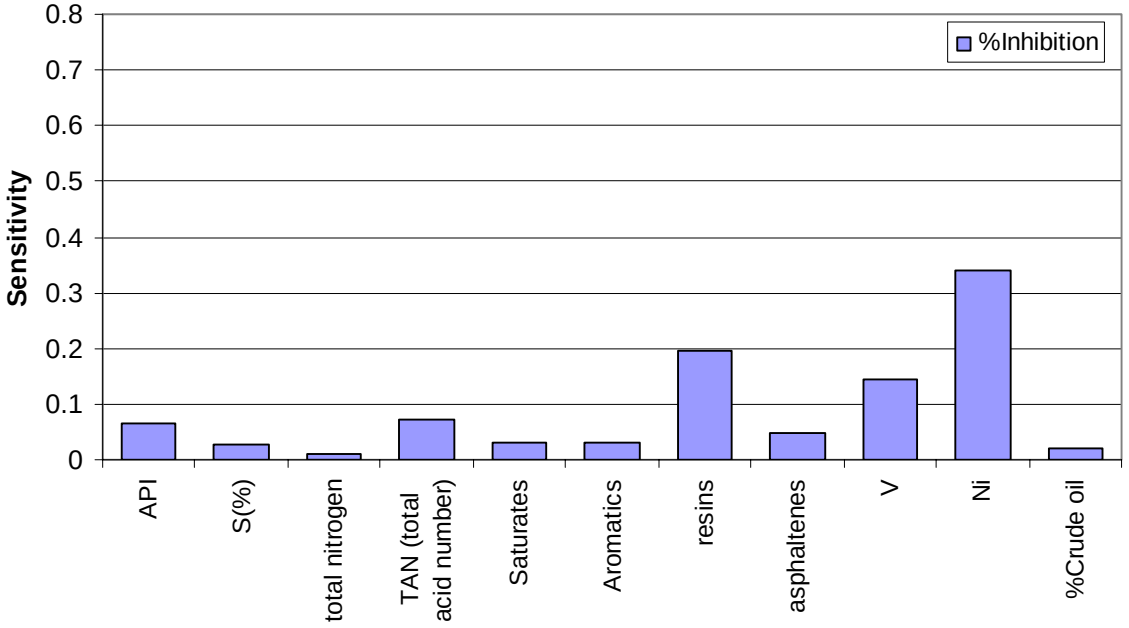

Figure 27 Sensitivity about the mean for 50% and 80% combined

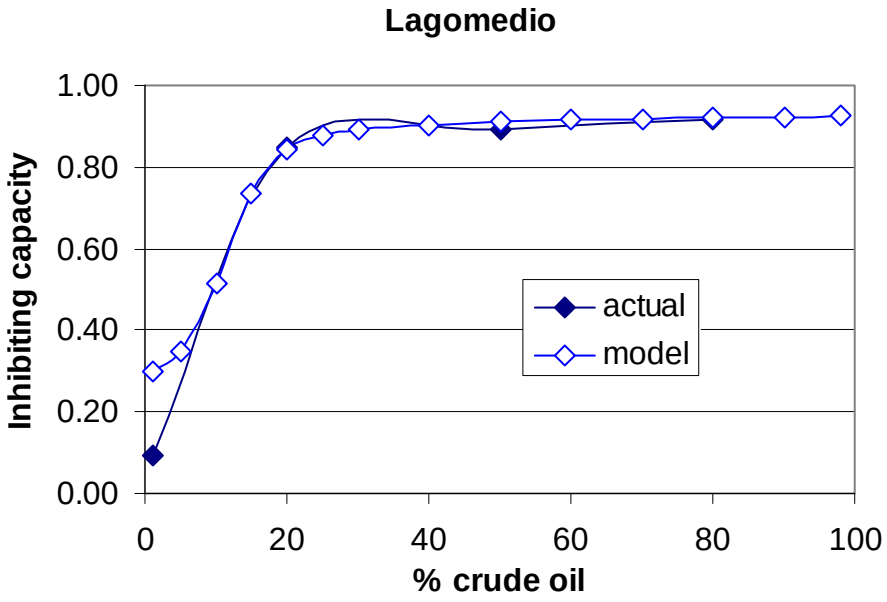

Figure 28 Prediction of the neural network when all variables are known.

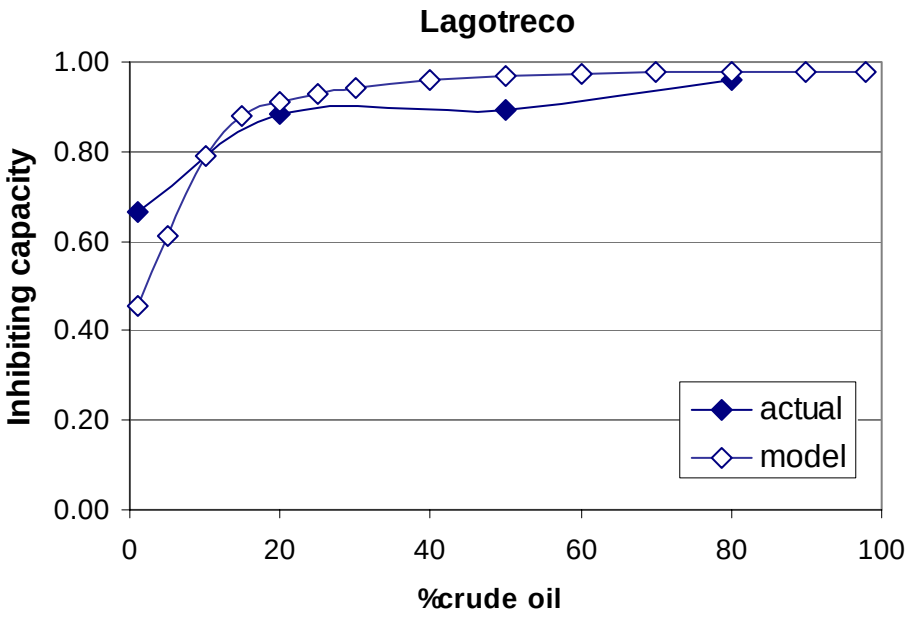

Figure 29 Prediction of the neural network by knowing four variables and deducting the others from the linear relationships found with the multiple regression analysis