International Journal of Industrial Engineering, 17(1), 69-79, 2010.

EXTRACTING KNOWLEDGE FROM CARBON DIOXIDE CORROSION INHIBITION WITH ARTIFICIAL NEURAL NETWORKS

G. Weckman[1], W. Young[1], S. Hernández[2], M. Rangwala[1] and V. Ghai[1]

[1]Industrial and Systems Engineering, Ohio University, Athens, OH 45701

[2]BP America Inc., 501 Westlake Park Blvd., Houston, TX 77079

The artificial neural network (ANN) has a proven reputation of accurately modeling the interacting relationships in a complex non-linear system. However, an ANN model is often considered a "black-box" in the sense that its estimates appear incomprehensible. This limitation is alleviated by using knowledge extraction techniques and algorithms. Better understanding of these relationships is significantly important to the oil industry, where the factors that affect corrosion are not well understood. To provide insight, this paper presents a number of different techniques to extract knowledge from an ANN trained with a CO_2 corrosion dataset. These techniques include Network Interpretation Diagrams, Garson's Algorithm, Sensitivity Analysis, Family of Curves and Surfaces, and TREPAN-Plus. From a knowledge-based perspective, these methods can provide the oil industry with the ability to determine the role of input variables in predicting corrosion inhibition. The limitations and advantages of each of these techniques are also discussed.

Significance: In order for the oil industry to understand the causes of corrosion in gas pipelines, this paper reviews several methods to extract knowledge from an artificial neural network (ANN), a machine-learning scheme. Carbon dioxide (CO_2) corrosion is a complex process involving several factors and through modeling, insights can be gained into variable interrelationships of the underlying process.

Keywords: Artificial neural networks, knowledge extraction, variable relationships, optimization

(Received 23 June 2008; Accepted in revised form 11 August 2009)

1. INTRODUCTION

The majority of oil and gas pipelines are made of carbon steel. Carbon steel, like other materials in nature, deteriorates over time. In a metallic pipeline, this deterioration usually occurs due to damaging effects from the surrounding environment. For carbon steel, one of the most dominant forms of deterioration is corrosion. Corrosion reduces the amount of metal in the pipe's cross-sectional wall thickness. As the pipe ages, it becomes less safe and less reliable. It is extremely important for oil and gas companies to monitor the quality of the pipe before potentially hazardous affects arise from a defective pipe. Throughout the world, companies are confronted with the expensive, time consuming, and potentially dangerous process of repairing or replacing damaged pipelines. Thus, it is important for pipeline operators to have an accurate and comprehensible model that predicts the remaining life of the section of pipe.

Carbon steel is vulnerable to corrosion from exposure to CO_2 due to electrochemical processes where ions dissolve at anodes and hydrogen evolves at the cathode (Nesic et al., 1995). This chemical reaction results in the formation of solid $FeCO_3$ films; these films can be protective or non-protective, depending on formation conditions. The presence of CO_2 acts as a catalyst increasing the hydrogen evolution, thereby increasing the corrosion rate of carbon steel in aqueous solution. Some researchers assume that H_2CO_3 either serves as an extra source of H^+ ions or is reduced directly (de Waard and Milliams, 1975, Gray et al., 1989). It is also assumed that both these reactions are independent of each other (Nesic et al., 1995). Particular attention is drawn to the recent reviews of the main design considerations European Federation of Corrosion, 1997) and prediction techniques related to CO_2 corrosion compiled by the (European Federation of Corrosion (European Federation of Corrosion, 1994). The role of crude oil in CO_2 corrosion has gained special attention in the last few years due to its significance when predicting or modeling corrosion rates. Efird (1991) was one of the first to identify the importance of testing the effect of crude oils in order to predict corrosion rates. Hernández et al., (2002) presented how certain variables in the composition of crude oil would influence the inhibition of corrosion.

Repairing or replacing a section of pipeline in the oil and gas industry is prohibitive due to the expense and time required. Therefore, a model that indicates the significance of factors related to the corrosion of a steel pipeline is paramount to the oil industry. A model should account for the particular type of oil flowing into the pipeline. This research is devoted to using an artificial neural network (ANN) as an artificial intelligence approach for modeling the corrosion rate of carbon steel and extends the analysis to yield comprehensible models by using rule extraction procedures.

2. LITERATURE REVIEW

For years, researchers have presented various approaches detailing the process of corrosion. Corrosion prediction has been identified as a key approach in utilizing the knowledge of the corrosion process and applying it to industrial corrosion related problems. Many corrosion models have been developed over the years, which are briefly outlined in this section. These models can be categorized into three main categories: empirical, semi-empirical and mechanistic models, based upon how firmly they are grounded in theory. It is important to note that some of these models are analytically complex and one would need a thorough understanding of the thermodynamic and electro-chemical processes occurring in corrosion in order to comprehend their meaning.

It has been observed through empirical models that the CO_2 corrosion rates in operational crude oil pipelines are much lower than those obtained under laboratory conditions (where crude oil was not used or where synthetic crude oils were used) (Nesic et al., 1995). The semi-empirical models, typically based on the mechanistic approach, are the "worst-case" models because they do not take into consideration the presence of protective surface films, corrosion inhibitors, hydrocarbons, different steel types, high pressure and other realistic conditions found in the oil and gas industry (Nesic and Vrhovac 1997). In mechanistic models, CO_2 corrosion is considered as a complex phenomenon where electrochemical, transport, and chemical processes occur simultaneously. Due to the complexity of the process, there is no single mechanistic approach that is able to model all of these complexities. Although some significant advances in corrosion control, the best practice is still to blend or neutralize crude oil. Better understanding of the corrosion mechanism in the presence of naphthenic acids is necessary.

Sinha and Pandey (2002) proposed a fuzzy-logic based ANN for reliability assessment of oil and gas pipelines. This model was trained with field observation data collected using magnetic flux leakage (MFL) tools in order to characterize the actual condition of aging pipelines vulnerable to metal loss corrosion. The object of this work was to develop a simulation-based probabilistic neural network model to estimate the probability of failure of aging pipelines vulnerable to corrosion. An expert system of a crude oil distillation unit (CDU) was developed to carry out the process of optimization on maximizing the oil production rate under the required oil product qualities (Liau et al., 2004). The expert system was established using the expertise of a practical CDU operating system provided by a group of experienced engineers. The input operating variables of the CDU system were properties of crude oil and manipulated variables while the system output variables were defined as oil product qualities.

Crude oil blending is an important aspect in the petroleum refining industry. Many blend automation systems use real-time optimizer (RTO), which apply current process information in order to update the model and predict the optimal operating policy based on the on-line analyzers. In certain situations, oil fields cannot apply these analyzers. Yu and Morales (2005) proposed an off-line optimization technique to overcome the main drawback of RTO. Historical data was used to approximate the output of the on-line analyzers with an ANN, and then the desired optimal inlet flow rates were calculated by the optimization technique via the neural model. After off-line optimization, the inlet flow rates are used for on-line control.

In recent years, the field of artificial intelligence has been explored for modeling the corrosion process. This leads to the necessity of developing a more robust model that is able to predict the corrosion rate with a high rate of accuracy even in the presence of a limited noisy data set. ANNs are being recognized as a powerful and general technique for machine learning because of their non-linear modeling abilities (Reed and Marks, 1998). ANNs have been one of the most promising approaches to the corrosion modeling process as demonstrated by Hernández et al., (2006). This study evaluated the usefulness of predicting corrosion inhibition by using an ANN.

3. INDUSTRY APPLICATION: CO_2 CORROSION MODEL

The corrosion dataset is a collection of the characteristic composition of fifteen Venezuelan crude oils used to predict the ability of a crude oil to offer corrosion inhibition in a CO_2 environment (Hernández et al., 2002). These attributes include measurements of American Petroleum Institute (API) density, sulphur, total nitrogen, total acid number (TAN), saturates, aromatics, resins, asphaltenes, vanadium, nickel and percentage of crude oil. Small sample sizes make it challenging to derive any mathematical model based on experimental data. This is especially true for traditional statistical models, which are typically based on least squares regression.

The API employs many of the attributes found in the Venezuelan crude oil dataset to grade crude oil in its scale. Crude oil is graded based upon the amount of impurities found in samples, with sulphur and nitrogen content in addition to the content of heavy metals, such as vanadium and nickel. The total acid number (TAN) indicates how much oxidation has taken place in a fluid. An in depth description of the test conditions used to determine these oil characteristics can be found in Hernández et al., (2002). These attributes are used to predict the inhibiting capacity and hence the corrosion rate given by Equation 1.

$$Inhibiting\ capacity = 1 - \frac{Corrosion\ rate_{crude\ oil}}{Corrosion\ rate_{blank}} \quad \ldots \quad (1)$$

NeuroSolutions, software developed by NeuroDimensions Inc. (2006), was used to experiment with different neural network models for the datasets described above. Various network types such as multi-layer perceptron and generalized feed forward networks were tested to obtain the best neural network model. Due to its limited size, the corrosion dataset was divided into 70%, 15%, and 15% sections for training, cross-validation, and testing respectively in order to provide for more training data. An 11-5-3-1 MLP network utilizing a hyperbolic tangent function achieved the best accuracy for the corrosion model. This model was trained for 20,000 epochs, which resulted in a correlation co-efficient (r) value of 0.942.

In addition to the regression problem with continuous output data, a different model based on grouping the output into classes was needed to create decision trees. The continuous output ranges were transformed into a classification problem. The output (% inhibition) was divided into five classes as shown in Table 1.

Table 1. Corrosion Class labels

%Inhibition	Class
0.75-0.849	Cl1
0.85-0.889	Cl2
0.89-0.949	Cl3
0.95-0.979	Cl4
0.98 and above	Cl5

4. ANN KNOWLEDGE EXTRACTION

A key shortcoming in modeling complex systems by ANN is the lack of transparency in their estimates, causing the networks to act like 'black-boxes' by not providing information on the formation of an estimate. In addition, no indication is provided on the importance of the variables being used to model the system. However, the network itself contains valuable information such as the relationship between several interrelated indicators. By exploring these relationships through machine learning techniques, it is possible to capture predictive knowledge regarding network relationships to one another and the output. This is essential in understanding the basis of the decisions as this type of computer support system is often used in critical applications and verifications of decisions a requirement. Many important knowledge based systems have been developed and successfully applied to diverse databases, such as speech recognition, game playing, medical diagnosis, financial forecasting, and industrial control (Mitchell, 1997). Usually these models are difficult to understand because processing in a neural network occurs at the sub-symbolic level as numerical estimation and manipulation of network parameters. An ANN captures task-relevant knowledge as part of its training regimen. The non-linear mapping of an ANN's knowledge is encoded in terms of the architecture, transfer functions, weights, and biases. The knowledge represents the hypothesis learned by the network.

4.1. Network Interpretation Diagram

A visual representation method for the network and its connection weights was developed by Özesmi & Özesmi (1999). This visual method is termed Neural Interpretation Diagrams (NID), due to the underlying methodology representing the connection weights in the form of line joining neurons in each of the layers in the network. Studying of magnitude of connection weights helps in predicting the variable contribution as well as understanding the interactions between the input variables (Aoki and Komatsu, 1999, Chen and Ware, 1999). Olden and Jackson (2002) explain the concept as when positive (or negative) connection weights transfer from input-hidden to hidden-output layer in an MLP, a resultant positive/excitatory effect of input variables is seen on the network output. Whereas a negative/inhibitory effect is projected when opposing connection, resulting in a weights flow from input-hidden to hidden-output layer. The product of the two connection weights subsequently passing between the layers of the MLP determines the final effect of the input variable on the output. The results of NID on the corrosion system are shown in Figure 1.

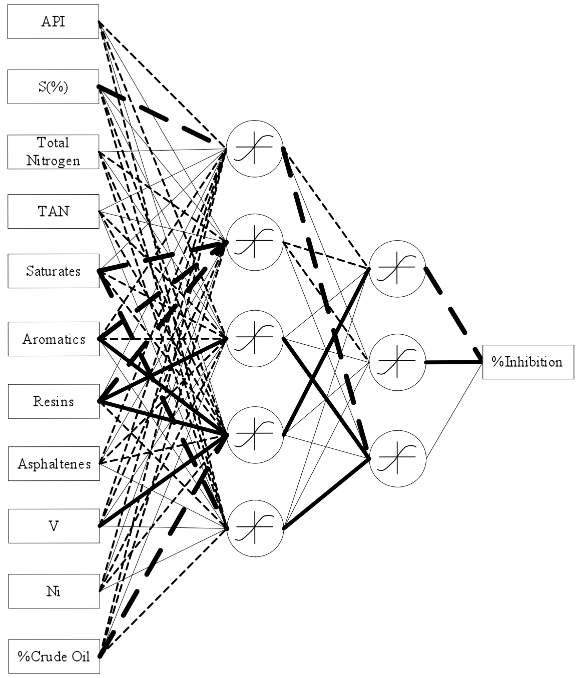

Figure 1. NID for 11-5-3-1 MLP Network

The NID indicates that there are four significant variables in the model, percentage crude oil (%Crude Oil), V, Resins and Aromatics. These four attributes have large positive or negative weights to/from the first and second hidden layer of the network. If a bold line cannot be traced from an input to an output, it is difficult to identify the actual significance of the input, for example Ni. The advantage of this method is in it being very easy to implement in order to acquire some basic information about the modeled system. Although an input's value is normalized before weights in the system are assigned, the process elements are non-linear. Since users determine values of weights that indicate the line thicknesses, the results can be very subjective.

4.2. Garson's Algorithm

Garson's Algorithm is a neural network process which provides information in the form of connection weights. The input from each variable is encoded into the network model as a weight, with the contribution of each of these variables to the output mainly dependent on the magnitude and direction of these connection weights (Olden & Jackson 2000) The mapping between the input variables and the predicted response generated in the case of a MLP, is a bi-level process of information flow involving weight transfer from input to hidden and then from hidden to output layer.

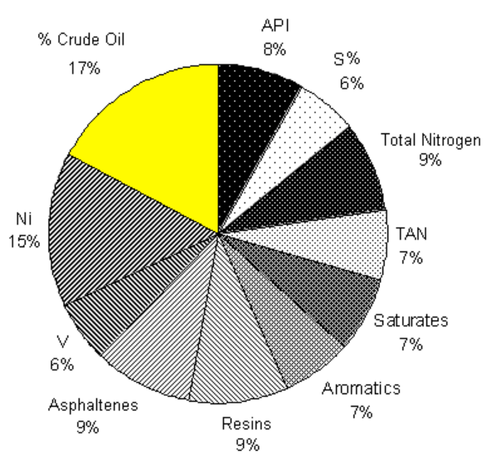

Figure 2. Results of Garson's Algorithm Showing Relative Importance of Input Variables.

Garson (1991) formulated an algorithm that calculated the relative importance of each of the input variable in a given network. Goh (1995) later proposed further enhancements to this algorithm. The result of Garson's Algorithm applied to the corrosion system is shown in Figure 2.

In this method, the %Crude Oil and Ni are considerably important in the corrosion system. However, several other factors appear significant including; Total Nitrogen, Resins, and Asphaltenes, which are all approximately equal in weight. Dynamically, this methodology does not offer any indication as to what happens to the output of the network when different levels of the input parameters change. It doesn't indicate if the effect of one parameter is beneficial or detrimental. However, it does reduce the amount of user bias that the NID introduces to the result. Again, it does not tell much about the inhibition other than the fact that %crude oil and Ni are important.

4.3. Sensitivity Analysis

Another approach used in extracting knowledge from ANNs is Sensitivity Analysis, which attempts to model the interaction of various input factors (Recknagel et al., 1997). Sensitivity analysis is a method used to extract cause and affect relationships between input and output variables. Sensitivity analysis also provides feedback as to which input variables are the most significant relative to other input variables. Based on this analysis, insignificant variables can be removed from the ANN, which would reduce the size, complexity, and training times. However, this would remove the impact and relationships that the input variable has to the output and other input variables. Figure 3a displays the sample results of sensitivity analysis to the corrosion system. Figure 3b displays an individual response to the input variable, %Inhibition, to an input variable, %Crude Oil. According to this test an increase in %crude oil will cause an increase in the inhibiting capacity (see Figure 3b), and this trend was clearly demonstrated in the experimental results.

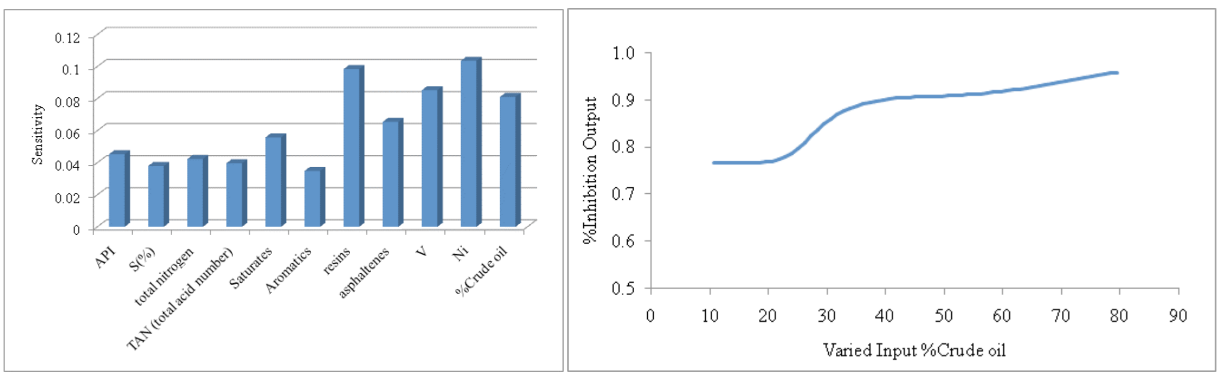

Figure 3. a) Input Sensitivity Results b) Individual Sensitivity Response.

The results of the sensitivity analysis identified four key input attributes that are important to the corrosion system such as Ni, resins, %Crude oil and V. In addition, the variables asphaltenes and aromatics are close behind the top four sensitive variables. It should be noted that these results differ from the NID and Garson's Algorithm. The NID does not necessarily identify Ni as a significant variable. In comparison, the Garson's algorithm does not indicate resins as being as significant in comparison to other input variables. This may not be a misleading result since NID and Garson's Algorithm do not encompass interactions with other input variables. However, sensitivity analysis requires that all input parameters, aside from the ones under analysis, be locked to their mean value. This result might not correspond with the result from the physical system due to the impracticality of holding operational values to their means.

4.4. Family of Surfaces

Sensitivity analysis can be extended to investigate a family of surfaces based upon knowledge obtained by training a neural network (Young & Weckman, 2007). A combination of input variables can be changed over a range, and the response of the output, %Inhibition, can be observed. To generate a surface response of the %Inhibition, each input value was held at its average value except the values of Ni and resins. Ni and resins inputs were then varied from their sample minimum to their sample maximum. Figure 4 illustrates that as resins decrease in the system the total %Inhibition also decreases while as Ni decrease, the % Inhibition will increase.

There are two forms of family of surfaces in discussed in this paper. The first is shown in Figure 4. This type of surface is essentially the same as the relationships that a sensitivity analysis would show with one exception. In this 3d surface, 2 inputs are varied and the change of a single variable is observed in output, which differs from the sensitivity results where only 1 input was varied, while the change in output was observed. This is a more accurate representation of the relationships that the ANN is using for its estimation. This is because it is highly unlikely that attributes can co-exist at their sample average. For example, input attributes are usually correlated. If one attribute increase, it is very likely that another attribute will change as well. Thus, the surface explains the output characteristics of the model when multiple inputs are varied. If a highly accurate model can be obtained through an ANN, one can assume that the output surfaces are mimicking the true behavior of the system, which could be used to better understand the underlying system behavior.

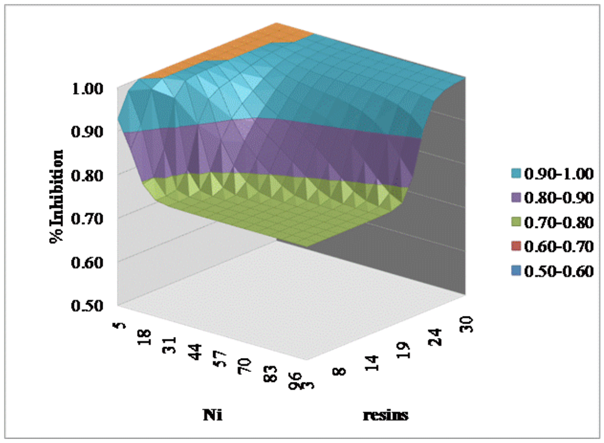

Figure 4. 3D Surface Relationship for Ni, resins, and %Inhibition

Although this technique is very useful to understanding the neural network models, it is not limited to only investigating a system's response when input values are varied. Output values can be varied over a specified range. Statistical optimization techniques, such as the SIMPLEX method, can be utilized to generate additional surfaces by solving for a constant input value. Therefore, the optimization condition is to minimize error while not allowing input parameters to exceed the normalization limits.

Figure 5a shows the effects of varying sensitive values (resins) and the output parameter %Inhibition. The idea of varying an output value in order to solve for an input variable often leads to the development of system optimization. In this particular case, if a high %Inhibition is desired, the interaction of Ni and resins can be determined over a practical range. Figure 5b is also a point of emphasis in this case because it is important to look at the error produced from the solved input value. The particular portion of the surface where the error is significant, this region cannot be trusted (high values or resins and low values of %inhibition. This implies that the value of resins could not be adjusted in such a way where the absolute percent error could be reduced to zero. In addition, this result insinuates that the neural network should be trained with more data in the failing regions (does not understand the system) or the combination of inputs are not feasible at this boundary. If errors of this nature occur, the range of the varied parameters can be adjusted to a level where the surface can be trusted.

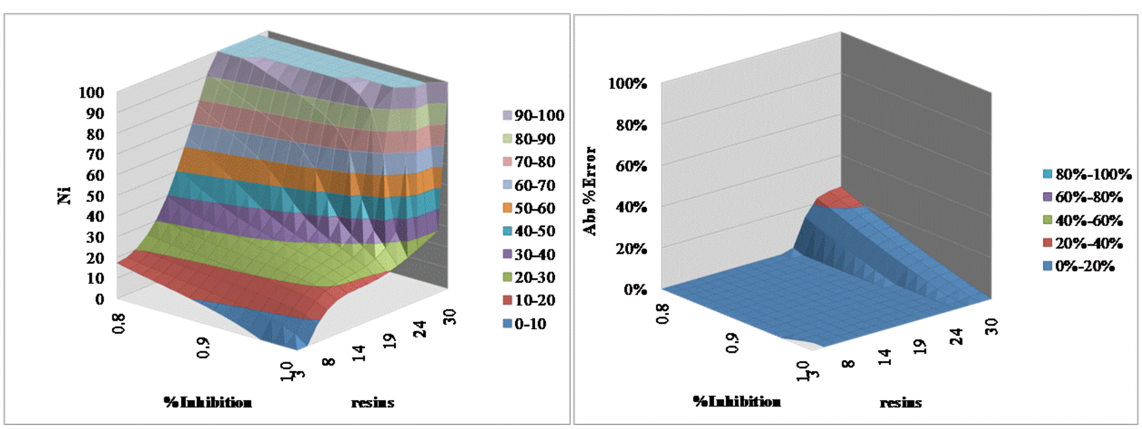

Figure 5. a) 3D Surface Relationship for %Inhibition, Ni, and Resins b) Abs. %Error of Surface

Generating surface relationships has a benefit over the standard sensitivity analysis method. This is because it provides the user with more flexibility and power to determine how the relationships are determined. Users can view complex system interaction and can determine the range in which attributes should be varied. Sensitivity analysis also requires non-changing variables to be held to their sample averages. In some cases, the relationship determined through this method may not represent the system accurately. This is because it may not be feasible for a system to exist where all of its attributes are held to the sample mean. Therefore, when users generate surface relationships through this method, they have full control of how the response is generated. This type of analysis is not limited to merely investigating three variables, but can be employed to investigate a single response over two or more variables. Surfaces can be generated to understand the physical phenomena being modeled. If some of the relations modeled by a neural network are known, this type of analysis can be used to provide evidence of correct system modeling.

This is an example of the second kind of surface. This surface is very different from the previous surface. This surface can be thought of as a solution space, where the solution space exists inside the surface. For example, a high inhibition is desired. If one desires an inhibition level of 0.8 to 1.0, one might ask what values other sensitive inputs must be in order to obtain this high inhibition. Therefore, to generate this surface, the %inhibition is varied for high desired values as well as one other input (resin). Once this range is determined, an optimization technique is used to calculate the necessary value that NI must be in order to produce the high level of inhibition. Thus, any value underneath the surface, or solution space, will give you a resulting %inhibition of 0.8 to 1.0. In other words, if a combination of Ni and resins are picked and plotted on the surface, if the points are inside the solution space, the desired inhibition level will be achieved. In this case, these surfaces can be used as design considerations. The other note here is the second figure showing abs %error. Since an optimization strategy is being utilized, a combination of inputs might not be feasible to generate the desired inhibition. The model itself is bounded by “allowable” combinations of input values based on the sample data. A constraint was placed on the normalization values. Thus, since the normalization used a large portion of sample values, it is assumed that an input could not lie far outside the limits of +/- 1.2. Thus, a 20% tolerance is allowed for the solved values. The error graph would reflect how much confidence could be placed on the solution space. If there is a high error, the solution space could result in a lower or high value of inhibition than what was desired. So now the question becomes, could these surfaces help you better understand the underlying system behavior, or could it be used as a potential tool for design?

4.5. Decision Trees: Trepan

Decision trees classify data through recursive partitioning of the data set into mutually exclusive subsets which best explain the variation in the dependent variable under observation (Biggs et al., 1991, Liepins et al., 1990). The TREPAN algorithm developed by Craven (1996), is a novel rule-extraction algorithms that mimics the behavior of a neural network. Given a trained neural network, TREPAN extracts decision trees that provide a close approximation to the function represented by the network, however it could also be applied to a wide variety of non-neural network based learning models (Craven and Shavlik, 1996). An extension of the original algorithm was performed to enhance the TREPAN software in order to handle a wider variety of problems. TREPAN currently works with multi-class classification problems and two class regression problems. The aim is to be able to work with classification as well as regression problems with multiple classes. The corrosion dataset was analyzed with multiple classes using the modified TREPAN algorithm called TREPAN-Plus. The result for a TREPAN run is illustrated in Figures 6.

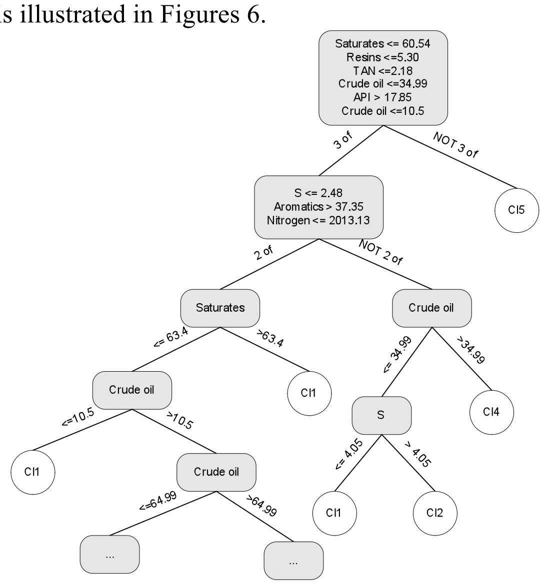

Figure 6. Corrosion: TREPAN Decision Tree

Table 2. Rules Extracted From the Trepan Decision Tree

Rule No.	Rule Text	Class Label
1	If ANY 3 of {Saturates <= 0.476, resins <= -0.728, TAN <= -0.0815, Crude oil <= -0.126, API > -0.276, Crude oil <= -0.684} and ANY 2 of {S <= 0.0139, Aromatics > 0.0589, nitrogen <= -0.506} and Saturates <= 0.557 and Crude oil <= -0.684	CL1
2	IF NOT 2 of {S <= 0.0139, Aromatics > 0.0589, nitrogen <= -0.506} and Crude oil <= -0.126 and S <= 0.739	CL1

The first rule implies that "If ANY 3 of {Saturates <= 0.476 or resins <= -0.728 or TAN <= -0.0815 or Crude oil <= -0.126 or API > -0.276 or Crude oil <= -0.684} and ANY 2 of {S <= 0.0139 or Aromatics > 0.0589 or nitrogen <= -0.506} and Saturates <= 0.557 and Crude oil <= -0.684 the class label is CL1" i.e. the predicted inhibition rate would be in range (0.75 - 0.849) as per Table 1. Some of the disadvantages for creating decision trees with the TREPAN algorithm include a potential loss of forecast accuracy. Some reasons for this reduction can be attributable to the bias nature of constructing class ranges. In this analysis, the best decision tree model produced an overall accuracy in predicting the correct class of 85.71%.

A possible advantage from the use of a decision tree on the corrosion study case is that it could be used to define risk levels, defined here as class levels, so that the user can understand the degree of protection that could be expected from a crude sample and also what could make it more or less corrosive.

From a knowledge-based perspective, these methods can provide the oil industry with a tool to predict corrosion inhibition. However the ability to determine the role of each variable in predicting corrosion inhibition needs careful analysis as different methodologies conduce to different conclusions. While the effect of increased %crude oil in reducing inhibition is well understood and clearly demonstrated in experiments, the effect of variables such as resins and Vanadium is still not well understood, and while most of these techniques indicated an effect, the degree and nature (increasing or decreasing % inhibition) of the effect vary significantly. This may be a consequence of the tight relationship that exists among some of these variables, some of which were identified by the authors in a previous paper (reference to NACE 05554)

- 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.

Carbon dioxide corrosion is a complex mechanism, and even more complex is the nature of crude oils. These highly correlated variables make it hard for the neural network to pick up sensitivities 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.

5. CONCLUSIONS

This paper explored a number of knowledge extraction techniques for ANNs, from basic to complex algorithms and the quality of information gained from these techniques. It was demonstrated how these techniques can be used to extract various levels of knowledge from various ANNs that were used to model a complex system such as corrosion. Methods were explored that would allow the investigator to go beyond the ANN's limitations typically referred to as a "black box." By extracting knowledge to a comprehensible form, further relationships between input and output variables can be explored and can be used to support a generation of a more usable prediction tool while enhancing the understanding of the actual system. Future research in this area will take the next step in the development of a knowledge-based model in order to form the initial mechanistic model.

In addition, this paper introduced a new technique named 'Family of Surfaces'. This technique allows a combination of input variables to be changed over a range, and the response of the output can be observed. This technique is not limited to only investigating a system's response but the output values can also be varied over a specified range. Statistical optimization techniques can be utilized in order to generate additional surfaces by solving for input values. If some of the relations being modeled by a neural network are known, this type of analysis can be used to provide evidence that the network is modeling the system correctly. This type of analysis could be run over several random samples while recording the impact of values with respect other values being a sample value rather than a mean value. This methodology could generate a more reliable and realistic physical sensitivity analysis, currently under development by the authors.

Lastly, this paper also introduced an extension of the decision tree algorithm 'TREPAN'. This extension allows the analyst the ability to convert continuous output problems into a multiclass decision tree. The decision tree then establishes knowledge extraction in the form of rules. An ANN model with knowledge extraction can better understand complex systems and relationships while presenting them in a more comprehensible state. The combination of ANN and Knowledge extraction can be an extremely powerful tool, in both predicting and gaining insight on how a complex systems behaves in this research have the potential to increase knowledge of the physical relationships and prediction accuracy.

6. REFERENCES

1. Aoki, I., and Komatsu, T. (1999). Analysis and prediction of the fluctuation of sardine abundance using a neural network, Oceanol. Acta, 20, 81–88.
2. Biggs, D., de Ville, B. and Suen, E. (1991) A method of choosing multiway partitions for classification and decision tree. Journal of Applied Statistics 18(1): 49-62.
3. Chen, D., and Ware, D. (1999). A neural network model for forecasting fish stock recruitment, Can. J. Fish, Aquatic. Sci, 56, 2385–2396.
4. Craven, M. (1996) Extracting Comprehensible models from trained Neural Networks, PhD Thesis, Computer Science Department, University of Wisconsin, Madison, WI.
5. Craven, M., and Shavlik, J. (1996). Extracting tree-structured representations of trained networks. In Advances in Neural Information Processing Systems, Denver, CO, MIT Press, 8, 24–30.
6. de Waard, C., and Milliams, D. (1975). Carbonic Acid Corrosion of Steel, Corrosion, 31(5), 131-177.
7. Efird, K. (1991). Preventive corrosion engineering in crude oil production, Offshore Technology Conference (OTC), 6599.
8. European Federation of Corrosion (1994). CO_2 corrosion control in oil and gas industry, A Working Party Report, 13, The Institute of Materials, London, England.
9. European Federation of Corrosion, (1997). CO_2 corrosion control in oil and gas production, A Working Party Report, 23, The Institute of Materials, London, England.
10. Garson, G. (1991). Interpreting neural network connection weights, Artificial Intelligence Expert, 6, 47-51.
11. Goh, A. (1995). Back-propagation neural networks for modeling complex systems, Artificial Intelligence in 1. Engineering, 9, 143-151.
12. Gray, L., Anderson, B., Danysh, M, and Tremaine, P. (1989). Mechanism of carbon steel corrosion in brines containing dissolved CO_2 at pH4, 464, Corrosion, Houston, TX, NACE International.
13. Hernández, S., Duplat, S., Vera, J., and Barón, E. (2002). A statistical approach for analyzing the inhibiting effects of different types of crude oil in CO_2 corrosion of carbon steel. Corrosion, 02293. National Association of Corrosion Engineers.
14. Hernández, S., Nesic, S., Weckman, G. and Ghai, V. (2006). Use of artificial neural networks for predicting crude oil effect on CO_2 corrosion of carbon steels, Corrosion, 62(6), 467-482.
15. Liau, L., Yang, T. and Tsai, M. (2004). Expert system of a crude oil distillation unit for process optimization using neural networks, Expert Systems with Applications, 26(2), 247-255.
16. Liepins, G., Goeltz, R. and Rush, R. (1990) Machine learning techniques for natural resource data analysis. AI Applications 4(3): 9-18.
17. Mitchell, T. (1997). Machine learning. 1st edition. Computer Science Series. Boston, MA: WCB McGraw-Hill.
18. Nesic, S., Postlethwaite, J., and Olsen, S. (1995). An electrochemical model for prediction of CO_2 corrosion, Corrosion/95, 131, Houston, TX, NACE International.
19. Nesic, S. and Vrhovac, M. (1997). A Neural Network Model for CO_2 Corrosion of Carbon Steel, Journal of Corrosion Science & Engineering, 1(6). 1995-2000.
20. NeuroDimension (2006). Neural Network Software, http://www.nd.com, Retrieved on January 01, 2006.
21. Olden, J. (2000). An artificial neural network approach for studying phytoplankton succession, Hydrobiology, 436, 131–143.
22. Olden, J., and Jackson, D. (2001). Fish-habitat relationships in lakes: Gaining predictive and explanatory insight by using artificial neural networks, Transactions of the American Fisheries Society, 130, 878-897.
23. Olden, J., and Jackson, D. (2002). Illuminating the "black box": a randomization approach understanding variable contributions in artificial neural networks, Ecological Modeling, 154, 135-150.
24. Özesmi, S. and Özesmi, U. (1999), An artificial neural network approach to spatial habitat modeling with interspecific interaction, Ecological Modeling, 116, 15–31.
25. Recknagel, F., French, M., Harkonen, P., and Yabunake, K. (1997), Ecological Modeling, 96, 11-28.

26. Reed, R., and Marks, R., (1998). Neural smithing: Supervised learning in feed forward artificial neural networks," MIT Press, Cambridge, MA.
27. Sinha, S. and Pandey, M. (2002). Probabilistic Neural Network for Reliability Assessment of Oil and Gas Pipelines, Computer-Aided Civil & Infrastructure Engineering, 17(5), 320-329.
28. Thorburn, W. (1915). Occam's razor, Mind, 24, pp. 287-288.
29. Young, W. and Weckman, G. (2007). Output- and Input-Response Surfaces Generated from an Artificial Neural Network for an Imperial to Semi-Mechanistic Model: A Heuristic Approach, Artificial Neural Networks in Industrial Engineering (ANNIE) Conference, St. Louis, Missouri
30. Yu, W. and Morales, A. (2005). Neural networks for the optimization of crude oil blending, International Journal of Neural Systems, 15(5), 377-389.

BIOGRAPHICAL SKETCH

Gary Weckman is an Associate Professor in the Industrial and Systems Engineering Department at Ohio University. Dr. Weckman's primary research focus has been multidisciplinary applications utilizing knowledge extraction techniques with artificial neural networks (ANN). He has used ANNs to model complex systems such as large scale telecommunication network reliability, ecological relationships, stock market behavior and industrial process scheduling. Before joining the Ohio University faculty in 2002, he was a faculty member at Texas A&M University-Kingsville for six years. He has also practiced industrial engineering for over 12 years at such firms as General Electric Aircraft Engines, Kenner Products and The Trane Company.

William Young is a doctoral candidate in the Integrated Engineering program at Ohio University. His dissertation is focused on developing a team-compatibility decision support system. Young has worked on projects that were funded by the General Electric Aviation, National Science Foundation (GK-12 Fellow), and the Ohio Department of Labor. William received his bachelor's (BSEE) and master's (MSEE) degrees in Electrical Engineering at Ohio in 2002 and 2005 respectively. Young's research is focused on utilizing statistical and machine learning methodologies for topics that include cost modeling, team-compatibility, digital technologies for education, sports modeling, and ecological and environmental monitoring systems.

Sandra is a Materials Engineer with a Ph.D. in Material Sciences. She is currently working for BP exploration in Alaska. The last five years she worked for BP in Houston as a corrosion/materials engineer as part of the Integrity Management group and the Inherently Reliable Facilities (IRF) research program. Previous to that she worked for 13 years for PDVSA INTEVEP (Research Center for the Venezuelan National Oil Company) as a corrosion specialist and research scientist; and for a year as an Assistant Research Professor at the Institute for Corrosion and Multiphase Technology at Ohio University.

Maimuna Rangwala is a APICS Certified Supply Chain Professional (CSCP) working in the 3PL industry. Maimuna received her Bachelors degree in Production Engineering (BE) from Mumbai University in 2002 and Master of Science (MS) in Industrial engineering and Master of Financial Economics (MFE) from Ohio University in 2006 and 2007 respectively. Maimuna's experience ranges from SAP implementations, labor management implementations and value engineering-based design changes. She is currently working on continuous process improvement and distribution network optimization models.