

CORROSION CONTROL APPROACH USING DATA MINING

Stephen Dapiap¹, Gregory Wajiga², Michael Egwurube², Musa Kadzai², Nathaniel Oye² and ThankGodAnazodo¹

Department of Strategic Information, Institute of Human Virology, Nigeria¹.
Federal University of Technology Yola, Nigeria².

ABSTRACT

In this work we have developed a decision support system that can determine corrosion and project the time for corrosion growth maintenance using probabilistic modelling approach. Historical data of the atmospheric industrial environment conditions for five years on three metals- zinc, iron and steel with known thicknesses were used. The environmental conditions included precipitation, wind speed, sulphur dioxide, relative humidity and temperature. For the five-year period, the percentiles of corrosion in the environment were determined to be from 63.1% to 69%. The corrosion rates of Zinc, iron and steel were 0.92µm/year, 0.9µm/year and 0.51µm/year respectively. At the end of the first year the expected time to initiate corrosion growth maintenance actions for zinc, iron and steel were 9 years, 10 years and 14 years in that order. The probable contributions of each of the environmental factors to the corrosiveness of the environment were determined.

KEYWORDS

Data Mining, Corrosion, Corrosion Control, Decision Support System, Probabilistic Modelling, Environmental Factors.

1. INTRODUCTION

With long exposure to atmosphere in the presence of water and oxygen, rust formed on metals surfaces diminishes them gradually to states of unusefulness if left untreated. With this exposure, all metals have the natural tendency to revert to the lesser energy state of ore. In simple terms, this natural phenomenon as stated by American Galvanizers Association is corrosion, and is most commonly seen as iron ore or rust [1]. Metallic corrosion includes both oxidation and exposure to oxygen in the environment and electrochemical processes meaning the metal forms corrosion cells on its surface that greatly accelerate the transformation of metal back to the ore state. Song *et al.*, [2] defined corrosion as “a gradual destruction of a material, a substance, or an entity, usually by solution or other means attributed to a chemical process”. Corrosion effects incur both direct and indirect costs. The corrosion effects have many serious economic, health, safety and technological consequences for the society. National Association of Corrosion Engineers International [3] estimated the costs (direct and indirect) of corrosion in the United States of America to have exceeded \$1 trillion. The association stated further that “though metals corrosion is a natural phenomenon, estimates show 25-30% could be stymied if proper corrosion protection methods were employed”. In a similar way, the World Corrosion Organisation [4] reported that the “annual direct cost of metallic corrosion worldwide is \$2.2 trillion United States Dollars which is over 3% of the world’s GDP”. Nigeria’s contribution to the world corrosion annual costs is not immediately known. Some organisations especially the oil companies should have

established corrosion control measures. In the Nigerian public sector, corrosion control measures might not be a priority. Awareness on the corrosion and corrosion control seems to be minimal.

To effectively tackle corrosion problem, scientists have attempted to investigate the corrosion of materials especially metals under different environmental conditions. The study of corrosion and corrosion science has been around for many years, yet the importance of learning more about them continues to increase. To decrease its presence in the society needs more understanding of corrosion. Understanding corrosion involve corrosion control which is either preventative or by maintenance. There have been existing methods of corrosion control through prevention using corrosion sensors and frequent inspections. A properly implemented corrosion control program is expected to disclose corrosion in the early stages. Such control program should include equipment design, corrosion monitoring, corrosion inspection and survey. The design entails minimizing the potential for corrosion. Corrosion monitoring involves series of surveys planned and organized to obtain comprehensive information on corrosion over time [5]. Bowles [6] expressed corrosion inspection and survey as a “snapshot” of corrosion at a specific point in time , while corrosion maintenance means fixing and repair of corrosion [7].Corrosion sensors and monitoring activities are important in the context of creating a “smart structure”, for obtaining information on actual condition. Citing example with pipelines explosions experienced in Nigeria in 1998 and 2005, [8]alluded to the fact that metal degradation and failure can occur anywhere in the world and at any time. The authors also tried to explained situation points at which corrosion tend to occur most and faster such as the production or processing, transportation and storage of crude oil or other substances especially liquids and natural gases.In this work “data mining” based corrosion control process is introduced. The process is aimed at preventive and predictive maintenance instead of the existing practice which uses fixed scheduled inspections and using corrosion sensors continuously to facilitate preemptive actions. The use of “data mining” process includes analyzinghistorical data collected over time aboutknown metal thickness previously exposed to environmental factors such as rainfall, temperature, humidity, and so forth. The results of the analysis provide support for decision making by taking proactive and knowledge-driven decisions on corrosion control.

2. DATA MINING AND ALGORITHM

Data mining is defined as a process of identifying and extracting hidden patterns, knowledge and information from databases and data warehouses [9]. Data mining, also popularly referred to as knowledge discovery from data(KDD), is the automated or convenient extraction of patterns representing knowledge implicitly stored or catchable in databases, data warehouses, the Web, other massive information repositories, or data streams. In McCarthy statement [10], data mining involves computer programs that infer relations among different kinds of data in databases. The goal has been to infer useful relations that might not have been noticed or at least could not have been confirmed among this data. Data mining is a multidisciplinary field, drawing work from areas including database technology, machine learning, statistics, pattern recognition, information retrieval, neural networks, knowledge-based systems, artificial intelligence, high-performance computing, and data visualization. The data-mining is the analysis component of KDD and it currently relies heavily on known techniques from machine learning, pattern recognition, and statistics to find patterns from data in the data-mining step of the KDD process[11].The essence of data mining to many according to [12] is the possibility of serendipitous discovery of unsuspected but valuable information. This means the process is essentially exploratory. The central aim of data mining is discovery. It is not concerned with those areas of statistics involving how best to collect the data in the first place so as to answer a specific question, such as experimental design and survey design. Essentially, data mining assures that the data have already been collected, and is concerned with how to discover its secret [12].

The knowledge discovery goals are defined by the intended use of the system. Two types of goals were identified as Verification and Discovery [13]. Verification is concerned with verifying the user's hypothesis while with discovery is concerned with finding new patterns. The discovery aspect is further subdivided into prediction, where the system finds patterns for predicting the future behaviour of some entities, and description, where the system finds patterns for presentation to a user in an understandable form. The discovery-oriented data mining appears to be more useful as it involves fitting models to, or determining patterns from, observed data. Data-mining methods can be viewed as consisting of three primary algorithmic components: model representation, model evaluation, and search. In data mining, algorithm plays a much more central role. Algorithms are more central in data mining given that data mining has mixed percentage of computer science and related disciplines. The mixed percentage of data mining means it has to also contend with other forms of data other than numerical data alone as in statistics. In an algorithm, there are possibly three components [13] :Model representation, Model-evaluation and Search method which consists of two components- parameter search and model search. Mannilain [14] tried to look at what formed data mining from probabilistic approach, data compression approach, microeconomic view and inductive databases view. The probabilistic approach view data mining as the task of finding the underlying joint distribution of the variables in the data. In this approach, the typical intention of the researcher is finding a short and understandable representation of the joint distribution, for example, Bayesian network [15]. Ayello and others in [16] used the Bayes theorem approach to model uncertainties in corrosion occurrence considering both knowledge uncertainties and data uncertainties to make informed decisions. The studies involved field measurements of internal corrosion and external corrosion of oil and gas pipeline. In a similar study, a supervised neural network method was used by Kamrunnahar and Urquidi-Macdonald in [17] as a data mining tool to predict corrosion behaviour of metal alloys. The neural network model learned the underlying laws that map the alloy's composition and environment to the corrosion rate. They used existing corrosion data on corrosion and in addition collected data on corrosion resistive alloys for both direct and alternate currents corrosion experiments. Certain parameters which include pH, temperature, time of exposure, electrolyte composition, metal composition, and so forth. were categorized and prioritized based on the results from data mining. The categorization helped in the understanding of the synergetic effects of the parameters and variables on electrochemical potentials and corrosion rates. The data mining process therefore starts with model development and secondly focusing on implementation and testing of the resulting methods. Data mining is an application-oriented research domain and currently focuses on domain-oriented decision delivery [18].

Hand and co-authors from the scientific and statistical approach [19], agreed to other authors statements on data mining as an analytical process of most often large observational data sets carried out with a view to obtain unsuspected relationships and also to summarize same in novel ways that are both understandable and useful to the data owner. In the authors words, "the relationships and summaries derived through a data mining exercise are often referred to as models or patterns" and these include linear equations, rules, clusters, graphs, tree structures, recurrent patterns in time series, and so forth. The statements confirmed the position on data mining as typically dealing with already collected data for some purpose as opposed to data from experiments with a particular interest in mind. It implies that no particular strategy is employed to collect data as the case in statistical studies and analysis.

3. PROBABILISTIC MODELLING OF ENVIRONMENTAL FACTORS

Model building in data mining is data-driven [19]. It is usually not driven by the notion of any underlying mechanism or "reality," but simply seeks to capture the relationships in the data. A model is characterized to describe the whole or a large part of the data set. A pattern on the other is considered as characterizing some local aspect of the data and a pattern can be considered to be a predicate that returns true for those objects or parts of objects in the data for which the pattern occurs, and false otherwise. The environmental factors under which metals are exposed influence the corrosivity of the environment with resultant corrosive effects on a metal's surface. A metal has some physical attributes such as thickness, weight and surface areas and an environment has parameters as temperature, wind, salinity, and so forth depending on the type of the environment. In an atmospheric environment thickness of the metal is the main characteristic of the metal considered for corrosion measurement because corrosion weakens a metal by attacking its thickness through the surface.

The probabilistic modelling shall be used to develop a data mining method that can be applied to obtain information from a database on metals previously exposed to an atmospheric environment. The metal thickness and environmental factors shall be denoted using oval shapes. The metal thickness and the environmental factors shall be represented as nodes and the relationship between them using directed arrows as edges. Probabilistic networks according to [20], "have become an increasingly popular paradigm for reasoning under uncertainty addressing such tasks as diagnosis, prediction, decision making, classification, and data mining". Interactions between the metal thickness and the environmental factors directly represent a cause-effect relationship. The long term relationship between the metal thickness and the environmental factors is perceived to have impact on the metal strength due to corrosive effects. Thus in the networks, links between the metal thickness and the environmental factors is directed to the "corrosion" node. The corrosion rate can be derived from the relationship thereby the time to carry out first maintenance is predicted. This concept is simply represented on a probabilistic network of Figure 1. The method is intended for information extraction for effective corrosion control and as a decision support system for on the time control action. The circular nodes represent predicting corrosion from the environmental factors interactions and resultant corrosion rate from which the time to scheduled first maintenance can be estimated.

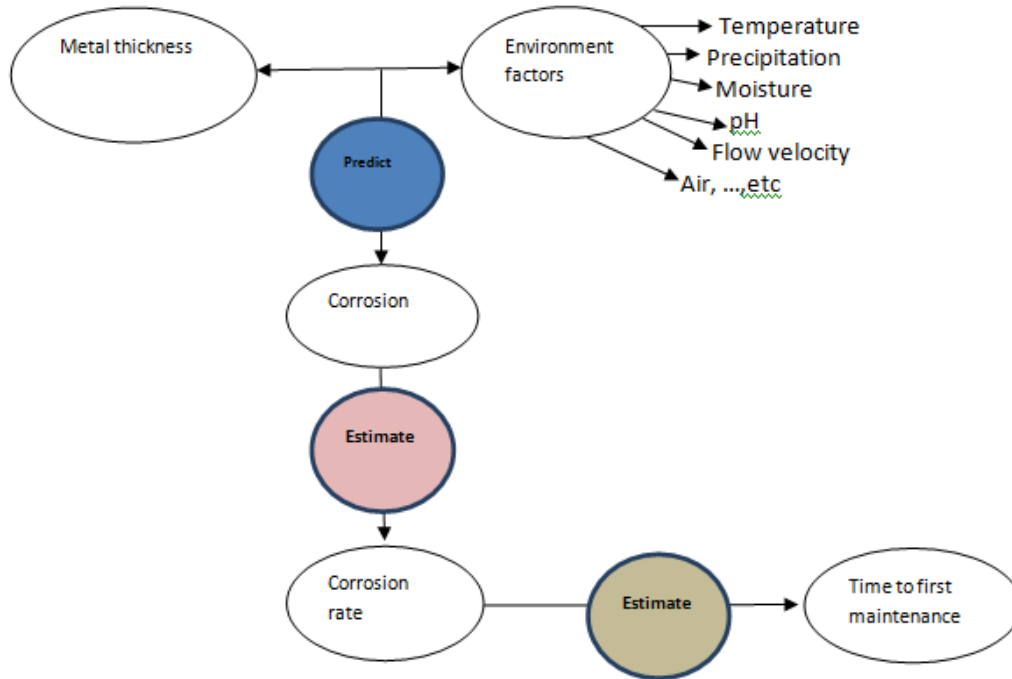


Figure 1. Conceptual Model of Metal Exposure to Environmental Factors

The conceptual model of Figure 1 can be represented using the probabilistic modelling format. A combined modelled qualitative and quantitative presentation for two environmental factors is indicated in Figure 2. Figure 2 represents the conditional probability of corrosion given the environmental factors denoted as

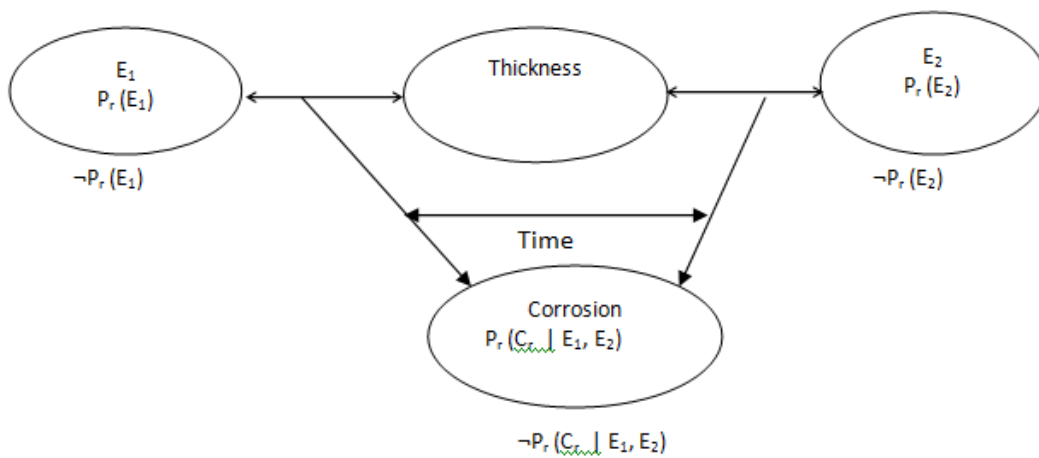


Figure 2. Conceptual Qualitative and Quantitative Representation of $P_r(C_r | E_1, E_2)$

$P_r(\text{Corrosion} | \text{Environmental factors})$. Consider these variables as C_r (Corrosion), E_1 (Environmental factor 1), and E_2 (Environmental factor 2). By using a conditional probability to specify the probabilistic relationship between these variables, then for each combination of outcomes (states) of E_1 and E_2 , we need to specify a probability distribution $P_r(C_r | E_1, E_2)$, given

in, Table 1. Table 1 is the conditional probability table reflecting the converging connection of Figure 2, where the child node is C_r and the parent nodes are E_1 and E_2 . In a similar fashion we can model a general relationship of more than two environmental factors as depicted in Figure 3. The conditional probability distribution for C_r given E specifies the probabilistic relations between these variables. Thus for each combination of outcomes or states of E , the probability distribution over the states of C_r needs to be specified. This conditional probability expresses the probability of corrosion if a relationship exists between the environmental factors.

Table1. Conditional Probability Table for Two Environmental Factors

	$E_1 = \text{no}$		$E_1 = \text{yes}$	
	$E_2 = \text{no}$	$E_2 = \text{yes}$	$E_2 = \text{no}$	$E_2 = \text{yes}$
$C_r = \text{no}$	$E_1, E_2, C_r = \text{no}$	$E_1, C_r = \text{no}, E_2 = \text{yes}$	$E_2, C_r = \text{no}, E_1 = \text{yes}$	$E_1, E_2 = \text{yes}, C_r = \text{no}$
$C_r = \text{yes}$	$E_1, E_2 = \text{no}, C_r = \text{yes}$	$E_2, C_r = \text{yes}, E_1 = \text{no}$	$E_1, C_r = \text{yes}, E_2 = \text{no}$	$E_1, E_2, C_r = \text{yes}$

The probability of occurrence of an environmental factor is mutually exclusive, so also the probability of corrosion occurrence.

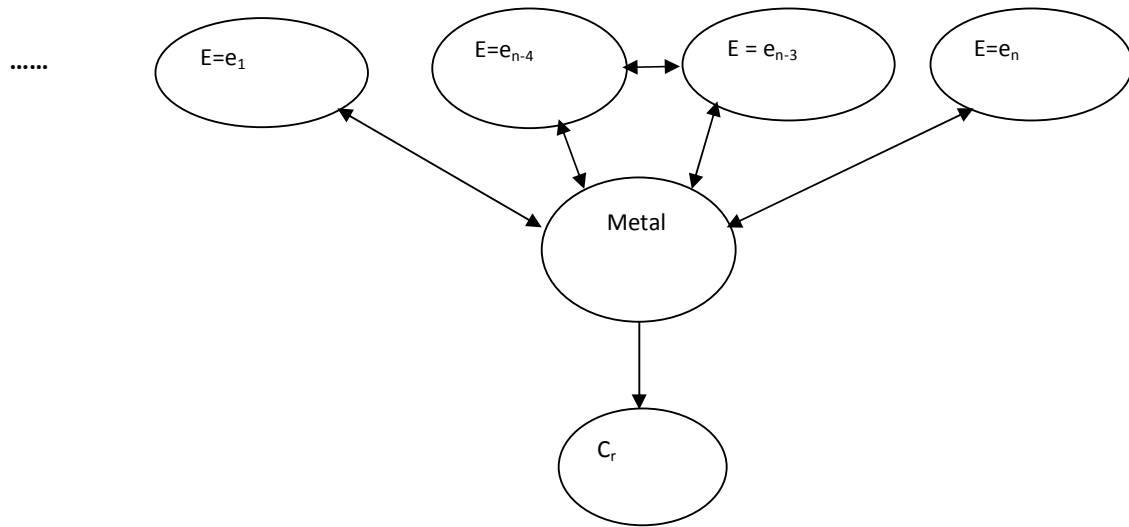


Figure 3. Qualitative Representation of Multiple Environmental Factors

A general structure of a conditional probability table for multiple environmental factors is possible drawing from Table 1 and is shown in Table 2.

Table 2. Probability Values for Multiple Environmental Factors

A	True								False							
B	True				False				True				False			
C	True		False		True		False		True		False		True		False	
D	True	False	True	False	True	False	True	False	True	False	True	False	True	False	True	False
.	.								.							
.	.								.							
.	.								.							
True	TTTT...TTTTF...F				.				.				FFF...F			
False	(TTTT...T)-(TTTTF...F)				.				.				(FFF...F)			

4. FORMULATING THE DATA MINING ALGORITHM

Based on the preceding models of Figures 2 and 3, the algorithm to mine for information from the database can be derived starting with a single environmental factor. The probability of corrosion with the existence or non-existence of say environmental factor 1 (E_1) is given from the quantitative relationship and the law of total probability:

(1) *Single Environmental Factor*

$$\begin{aligned}
 P_r(C_r) &= P_r(C_r | E_1) \\
 &= P_r(C_r | E_1) + P_r(C_r | \neg E_1) \\
 &= P_r(C_r | E_1) P_r(E_1) + P_r(C_r | \neg E_1) P_r(\neg E_1) \dots
 \end{aligned}
 \tag{1}$$

(2) *Two Environmental Factors*

$$\begin{aligned}
 P_r(C_r) &= P_r(C_r | E_1) + P_r(C_r | E_2) \\
 &= P_r(C_r | E_1) P_r(E_1) + P_r(C_r | \neg E_1) P_r(\neg E_1) + P_r(C_r | E_2) P_r(E_2) \\
 &\quad + P_r(C_r | \neg E_2) P_r(\neg E_2) \\
 \text{So } P_r(C_r) &= P_r(C_r | E_1, E_2) P_r(E_1) P_r(E_2) \\
 &= P_r(C_r | E_1) P_r(E_1) + P_r(C_r | E_2) P_r(E_2) \dots
 \end{aligned}
 \tag{2}$$

E_1 and E_2 are independent. Therefore,

$$P_r(C_r | E_1, E_2) = (P_r(C_r | E_2) * P_r(E_1 | C_r, E_2)) / P_r(E_1 | E_2) \dots
 \tag{3}$$

Equation (3) gives the probability of corrosion occurrence given that environmental factor 1 and environmental factor 2 are known. Equation (3) is the same as

$$(P_r(C_r | E_1, E_2) = ((P_r(C_r | E_2) * P_r(E_1 | C_r, E_2)) / (P_r(E_1 | C_r, E_2) * P_r(C_r | E_2)$$

$$+ P_r(E_1 | \neg C_r, E_2) * P_r(\neg C_r | E_2)) \dots \tag{4}$$

4.1. Estimating Corrosion Value

From the probabilistic network of Figure 2, a breakdown of the quantitative probability expressions for each factor and the possible relationships between each pair can be visualized in tabular form. Table 1 is now transformed using probabilities of the environmental factors due to converging conditional dependencies as presented in Table 3. As earlier mentioned, the probability of occurrence of an environmental factor is mutually exclusive, so also the probability of corrosion occurrence. In Table 3, we can compute the probability of corrosion in a given environment from the network probabilities. This value is obtained by summing the various combinations in which corrosion is true, and breaking those probabilities down into known probabilities:

Table 3. Probabilities of Environmental Factors

E ₁	True		False	
	True	False	True	False
E ₂				
C _r = True	P _r (C _r E ₁ , E ₂)	P _r (C _r E ₁ , ¬E ₂)	P _r (C _r ¬E ₁ , E ₂)	P _r (C _r ¬E ₁ , ¬E ₂)
C _r = False	P _r (¬C _r E ₁ , E ₂)	P _r (¬C _r E ₁ , ¬E ₂)	P _r (¬C _r ¬E ₁ , E ₂)	P _r (¬C _r ¬E ₁ , ¬E ₂)

$$\begin{aligned}
 P_r(C_r) &= P_r(C_r | E_1, E_2) + P_r(C_r | \neg E_1, E_2) + P_r(C_r | E_1, \neg E_2) + P_r(C_r | \neg E_1, \neg E_2) \\
 &= P_r(C_r | E_1, E_2) * P_r(E_1, E_2) + P_r(C_r | \neg E_1, E_2) * P_r(\neg E_1, E_2) + \\
 &\quad P_r(C_r | E_1, \neg E_2) * P_r(E_1, \neg E_2) + P_r(C_r | \neg E_1, \neg E_2) * P_r(\neg E_1, \neg E_2) \\
 &= P_r(C_r | E_1, E_2) * P_r(E_1) * P_r(E_2) + P_r(C_r | \neg E_1, E_2) * P_r(\neg E_1) * \\
 &\quad P_r(E_2) + P_r(C_r | E_1, \neg E_2) * P_r(E_1) * P_r(\neg E_2) + P_r(C_r | \neg E_1, \neg E_2) * \\
 &\quad P_r(\neg E_1) * P_r(\neg E_2) \dots \tag{5}
 \end{aligned}$$

4.2. Diagnosing of Environmental Factors

By way of inference, the probabilities of environmental factor 1 and environmental factor 2 can be obtained of being true and therefore the chances that each have caused corrosion, if it is known that corrosion (C_r) is true or did not contribute to corrosion effect. This may be possible by using the Bayes theory:

$$\begin{aligned}
 P_r(E_2 | C_r) &= (P_r(C_r | E_2) * P_r(E_2)) / P_r(C_r) \\
 &= ((P_r(C_r | E_1, E_2) * P_r(E_1) + P_r(C_r | \neg E_1, E_2) * P_r(\neg E_1)) * P_r(E_2)) / P_r(C_r) \\
 P_r(E_1 | C_r) &= (P_r(C_r | E_1) * P_r(E_1)) / P_r(C_r)
 \end{aligned}$$

$$= ((P_r(C_r|E_1, E_2) * P_r(E_2)) + P_r(C_r|E_1, \neg E_2) * P_r(\neg E_2)) * P_r(E_1))/P_r(C_r) \dots \quad (6)$$

4.3. Formulating Generic Algorithm with Multiple Environmental Factors

In multiple environmental factors, the derivation follows the same format from the previous sections and a general form in a tabular form was shown in Table 2. The generic models for the pattern extraction among the environmental factors are given as follows:

(1) Probability of corrosion Occurrence

$$\begin{aligned}
 P_r(C_r) = & P_r(C_r| x_1, \dots, x_n)(P_r(x_1) * P_r(x_2) * \dots * P_r(x_n)) + \\
 & P_r(C_r| x_1, x_2, \dots, x_{n-1}, \neg x_n)(P_r(x_1) * P_r(x_2) * \dots * P_r(x_{n-1}) * P_r(\neg x_n)) + \\
 & P_r(C_r| x_1, x_2, \dots, \neg x_{n-1}, x_n)(P_r(x_1) * P_r(x_2) * \dots * P_r(\neg x_{n-1}) * P_r(x_n)) + \\
 & P_r(C_r| x_1, x_2, \dots, \neg x_{n-1}, \neg x_n)(P_r(x_1) * P_r(x_2) * \dots * P_r(\neg x_{n-1}) * P_r(\neg x_n)) + \\
 & P_r(C_r| x_1, x_2, \dots, \neg x_{n-2}, x_{n-1}, x_n)(P_r(x_1) * P_r(x_2) * \dots * P_r(\neg x_{n-2}) * P_r(x_{n-1}) \\
 & \quad * P_r(x_n)) + \\
 & P_r(C_r| x_1, x_2, \dots, \neg x_{n-2}, x_{n-1}, \neg x_n)(P_r(x_1) * P_r(x_2) * \dots * P_r(\neg x_{n-2}) \\
 & \quad * P_r(x_{n-1}) * P_r(\neg x_n)) + \\
 & P_r(C_r| x_1, x_2, \dots, \neg x_{n-2}, \neg x_{n-1}, x_n)(P_r(x_1) * P_r(x_2) * \dots * P_r(\neg x_{n-2}) * \\
 & \quad P_r(\neg x_{n-1}) * P_r(x_n)) + \\
 & P_r(C_r| x_1, x_2, \dots, \neg x_{n-2}, \neg x_{n-1}, \neg x_n)(P_r(x_1) * P_r(x_2) * \dots * P_r(\neg x_{n-2}) \\
 & \quad * P_r(\neg x_{n-1}) * P_r(\neg x_n)) + \dots + \\
 & P_r(C_r| \neg x_1, \neg x_2, \dots, \neg x_{n-2}, \neg x_{n-1}, \neg x_n)(P_r(\neg x_1) * P_r(\neg x_2) * \dots * P_r(\neg x_{n-2}) \\
 & \quad * P_r(\neg x_{n-1}) * P_r(\neg x_n)) \dots \quad (7)
 \end{aligned}$$

(2) Diagnostic values:

$$\begin{aligned}
 P_r(x_1|C_r) = & (P_r(C_r|x_1) * P_r(x_1))/P_r(C_r) \\
 = & ((P_r(C_r|(x_2, \dots, x_n), x_1) * P_r(x_2) * P_r(x_3) * \dots * P_r(x_n) + P_r(C_r|(\neg x_2, \dots, \neg x_n), x_1) \\
 & \quad * P_r(\neg x_2) * P_r(\neg x_3) * \dots * P_r(\neg x_n)) * P_r(x_1))/P_r(C_r) \\
 & \quad \vdots \\
 & \quad \vdots \\
 & \quad \vdots \\
 = & ((P_r(C_r|(x_1, \dots, x_{n-1}), x_n) * P_r(x_1) * P_r(x_2) * \dots * P_r(x_{n-1}) + P_r(C_r|(\neg x_1, \dots, \neg x_{n-1}), x_n) * \\
 & \quad P_r(\neg x_1) * P_r(\neg x_2) * \dots * P_r(\neg x_{n-1})) * P_r(x_n))/P_r(C_r) \\
 & \quad \dots \quad (8)
 \end{aligned}$$

Where x_i is environmental factors and C_r is corrosion.

5. METAL DEGRADATION RATE AND THE TIME TO FIRST MAINTENANCE

The corrosion rate C_{rr} , according to [21] is defined as the “amount of corrosion loss in thickness per year”. The corrosion rate may depend on the environment. In the atmospheric environment, the corrosion rate is expressed as

$$C_{rr} = d/t \quad \dots \quad (9)$$

where d is the total amount of thickness loss due to corrosion in micrometres, t , is the time in years during which the loss occurred. The corrosion rate is most commonly expressed in the units of $\mu\text{m}/\text{y}$.

American Galvanizers Association (2012) gave the standard of 5% rust of a metal surface to mark the first time to institute maintenance on a metal when exposed to corrosive agents. Therefore,

$$T_m(T_h, x_i) = (.05 * T_h * y) / C_{rr} \quad \dots \quad (10)$$

Where T_m is the time to first maintenance, T_h is thickness of metal, x_i is the environment parameters, y is number of years of exposure and C_{rr} is the corrosion rate.

6. IMPLEMENTATION OF THE MODEL

Historical outdoor data for five years were obtained from an industrial environment and used to implement the formulated data mining model. The process of the data mining started with the design of the database to keep the data sets for all the five years. The data sets can then be selected by each year. Sample of the database interface is indicated in figure 4 . The five years data sets are presented in Tables 4-8.

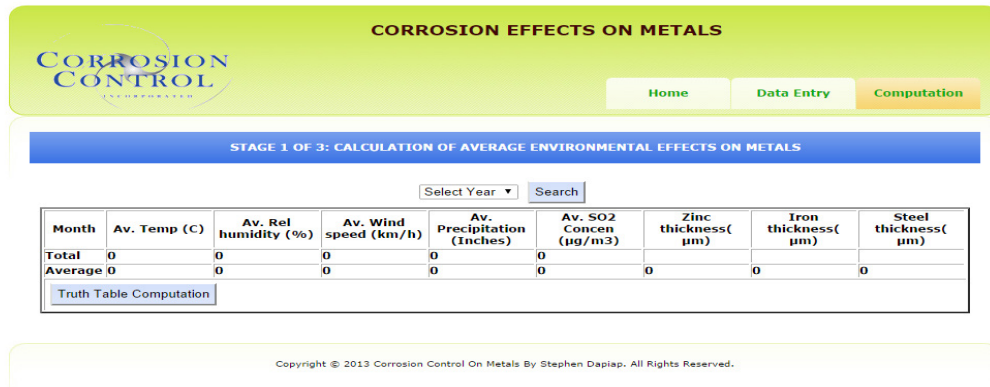


Figure 4. Database Selection Interface

Table 4. Historical Data on Metals in an Industrial Environment for 2009

Month	Average temperature (°C)	Average relative humidity (%)	Average wind speed (km/h)	Average precipitation (in)	Thickness(µm)			Average SO ₂ concentration (µg/m ³)
					Zinc	Iron	Steel	
Jan	21.3	32	25.00	0.00	164.0	170.0	147.7	42
Feb	22.9	27	22.50	0.00	163.0	170.0	147.0	72
Mar	24.9	30	15.40	0.80	163.0	169.9	147.0	29
Apr	25.7	34	10.80	5.00	162.8	169.9	146.8	45
May	24.5	54	6.06	7.15	161.8	169.7	146.8	80
Jun	22.7	65	18.30	7.90	161.6	169.6	146.6	71
Jul	21.4	80	12.20	11.98	160.0	168.8	145.7	62
Aug	20.9	85	15.10	11.18	159.1	167.8	145.0	76
Sep	21.9	92	22.40	10.50	156.9	165.5	144.8	56
Oct	22.9	48	25.91	5.20	154.3	163.0	143.1	55
Nov	22.7	36	28.20	0.20	153.0	161.8	142.3	59
Dec	21.5	15	30.00	0.00	153.0	160.0	141.6	58

Table 5. Data on Metals in an Industrial Environment for 2010

Month	Average temperature ($^{\circ}\text{C}$)	Average relative humidity (%)	Average wind speed (km/h)	Average precipitation (in)	Thickness(μm)			Average SO_2 concentration($\mu\text{g}/\text{m}^3$)
					Zinc	Iron	Steel	
Jan	20.0	30	24.50	0.00	152.9	159.6	141.6	65
Feb	21.9	22	26.00	0.04	152.7	159.5	141.6	92
Mar	24.1	28	17.20	0.67	152.2	158.8	141.6	66
Apr	19.3	31	12.10	3.03	152.1	158.6	141.5	79
May	23.1	44	10.00	7.09	151.4	158.6	141.3	85
Jun	21.9	76	12.30	7.64	151.0	158.4	141.1	90
Jul	20.8	86	12.00	11.73	149.1	157.5	140.8	56
Aug	20.4	92	14.30	11.02	146.8	156.1	139.9	74
Sep	21.4	95	22.60	8.58	144.2	154.4	138.9	68
Oct	22.2	48	24.90	2.09	142.6	153.3	137.9	39
Nov	21.5	32	30.00	0.16	142.2	151.8	136.0	70
Dec	20.2	17	32.00	0.08	142.0	149.8	135.7	72

Table 6. Data on Metals in an Industrial Environment for 2011

Month	Average temperature ($^{\circ}\text{C}$)	Average relative humidity (%)	Average wind speed (km/h)	Average precipitation (in)	Thickness(μm)			Average SO_2 concentration($\mu\text{g}/\text{m}^3$)
					Zinc	Iron	Steel	
Jan	20.5	29	30.10	0.00	141.8	149.5	135.7	70
Feb	21.8	22	25.20	0.08	141.8	149.5	135.7	56
Mar	24.4	31	16.00	3.04	141.6	149.4	134.8	62
Apr	25.0	34	12.00	5.03	141.6	149.3	134.5	92
May	24.3	52	8.10	8.10	141.6	149.1	134.5	80
Jun	22.5	78	11.00	10.00	139.0	149.1	133.4	65
Jul	21.0	82	15.00	13.03	138.2	148.0	133.1	45
Aug	20.9	89	15.40	13.02	137.7	146.2	132.3	77
Sep	21.4	94	20.10	10.52	135.1	144.4	131.5	29
Oct	22.6	42	25.80	4.00	133.0	142.6	130.2	34
Nov	21.7	36	26.00	0.46	132.2	141.7	130.0	18
Dec	20.2	12	28.00	0.06	131.0	139.8	129.9	66

Table 7.Data on Metals in an Industrial Environment for 2012

Month	Average temperature (°C)	Average relative humidity (%)	Average wind speed (km/h)	Average precipitation (in)	Thickness(μm)			Average SO ₂ concentration(μg/m ³)
					Zinc	Iron	Steel	
Jan	19.0	26	28.00	0.00	130.9	139.8	129.9	41
Feb	22.8	21	20.00	0.00	130.9	139.7	129.8	78
Mar	28.4	29	18.00	3.20	130.7	139.7	129.8	90
Apr	30.0	32	18.80	4.32	129.8	139.2	129.7	70
May	36.0	48	10.00	7.30	129.7	138.6	128.9	85
Jun	22.7	66	18.60	7.88	128.7	138.5	128.6	64
Jul	21.9	80	16.20	11.96	125.2	136.9	128.0	80
Aug	20.9	88	15.30	11.98	123.0	134.2	127.3	71
Sep	32.0	91	21.40	10.53	121.2	132.0	126.6	52
Oct	29.2	56	22.00	4.06	120.1	130.6	126.2	68
Nov	22.0	36	28.00	1.06	120.0	130.2	124.5	75
Dec	20.0	18	30.00	0.00	120.0	129.0	124.0	24

Table 8.Data on Metals in an Industrial Environment for 2013

Month	Average temperature (°C)	Average relative humidity (%)	Average wind speed (km/h)	Average precipitation (in)	Thickness(μm)			Average SO ₂ concentration(μg/m ³)
					Zinc	Iron	Steel	
Jan	18.0	30	15.02	0.00	120.0	129.0	124.0	40
Feb	28.0	28	18.52	0.08	120.0	128.0	124.0	12
Mar	34.0	15	12.37	2.67	118.0	128.0	124.0	56
Apr	38.0	16	1.85	3.06	118.0	128.0	123.0	90
May	40.0	56	4.06	8.06	117.0	126.0	123.0	85
Jun	38.0	78	6.32	8.67	114.0	125.0	123.0	63
Jul	20.0	86	8.23	13.13	113.0	125.0	121.0	51
Aug	25.0	92	2.12	13.02	111.0	123.0	120.0	76
Sep	42.0	95	1.42	10.58	110.8	122.0	120.0	35
Oct	36.0	48	13.91	4.09	110.2	121.0	118.0	55
Nov	20.0	36	18.21	0.17	109.0	121.0	118.0	29
Dec	19.0	15	17.30	0.09	109.0	119.0	118.0	23

7. RESULTS

The implementation of the models produced the results of the association between the environmental factors for all the years with examples of computer generated results in graphical formats for the first two years in Figures 5-8. Figures 7 and 8 capture further results of computation from the associations and the probability of corrosion, corrosion rate, the probable values of the environmental factors and the time to first maintenance. The results for the other three years are similarly obtained. In year 2009 the probabilities of temperature, humidity, wind speed, precipitation and sulphur dioxide occurrence from the average values were obtained as

0.23, 0.5, 0.19, 0.05 and 0.59 respectively. The results for year 2010 were 0.21, 0.5, 0.2, 0.04 and 0.71. Similarly the results for 2011, 2012 and 2013 were obtained with 0.22, 0.5, 0.19, 0.06 and 0.58 for 2011, 0.25, 0.49, 0.21, 0.05 and 0.67 for 2012 and 0.3, 0.5, 0.29, 0.05 and 0.51 for year 2013. The humidity maintained almost the same values for all the years with a slight difference of only 0.01 in the fourth year.

From the same figures, the average positive relationship amongst the environmental factors were obtained for each of the years as 0.49, 0.5, 0.5, 0.65, 0.46. These values show above average corrosive effect on the environment throughout all the five years as a result of the environmental factors interactions or association. The percentiles of materials corroding under these environmental factors in 2009, 2010, 2011, 2012 and 2013 were 64.8%, 69.1%, 64.7%, 64.3% and 63.1% respectively. The corrosion rates of zinc, iron and steel in 2009 under these factors were 0.92 $\mu\text{m}/\text{year}$, 0.83 $\mu\text{m}/\text{year}$ and 0.51 $\mu\text{m}/\text{year}$ respectively. Similarly, the corrosion rates for 2010 in that order were 0.91 $\mu\text{m}/\text{year}$, 0.82 $\mu\text{m}/\text{year}$ and 0.49 $\mu\text{m}/\text{year}$. The 2011 corrosion rates were 0.90 $\mu\text{m}/\text{year}$, 0.81 $\mu\text{m}/\text{year}$ and 0.48 $\mu\text{m}/\text{year}$; 0.91 $\mu\text{m}/\text{year}$, 0.90 $\mu\text{m}/\text{year}$, 0.49 $\mu\text{m}/\text{year}$ for 2012 and 0.92 $\mu\text{m}/\text{year}$, 0.83 $\mu\text{m}/\text{year}$, 0.50 $\mu\text{m}/\text{year}$ for 2013. Also the probable corrosion effects on any material exposed in this environment by either temperature, wind speed, humidity, precipitation or sulphur dioxide were 3.4, 2.7, 18.9, 13.2 and 0.9 in that order. Based on the respective corrosion rates of the metals, the estimated projected time to carry out first maintenance on the metals were 9 years, 14 years and 10 years for zinc, iron and steel respectively. In a similarly way, the values for years 2010 to 2013 were obtained. The results of the probable contributions of each of the environmental factors to the corrosivity of the environment are presented in Table 16. The probable values could not add up to 1 because there are other contributions due to the combination or fuzziness of interactions of the environmental factors not included in the results.

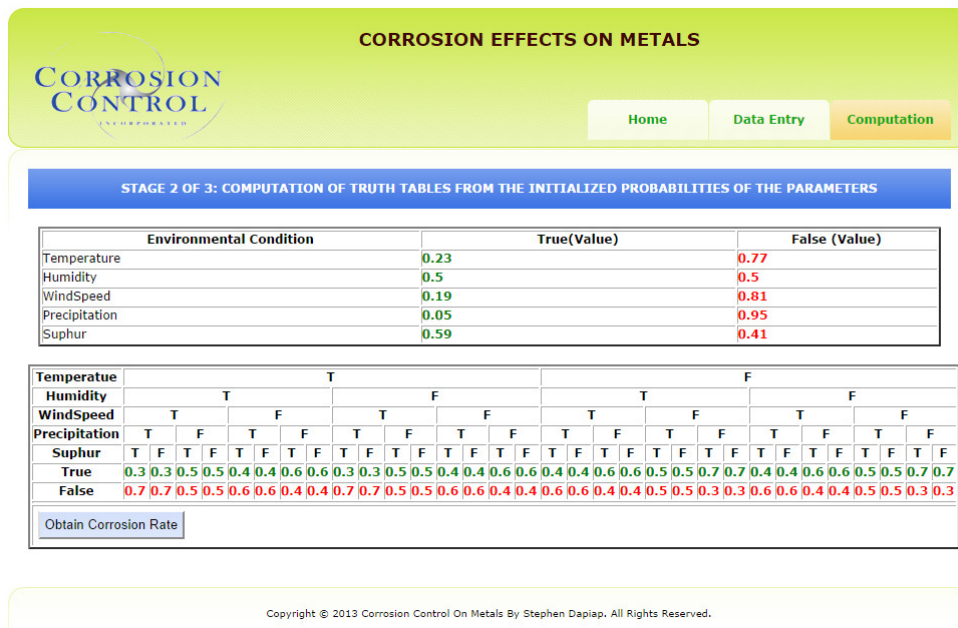


Figure 5. Results of Association for 2009

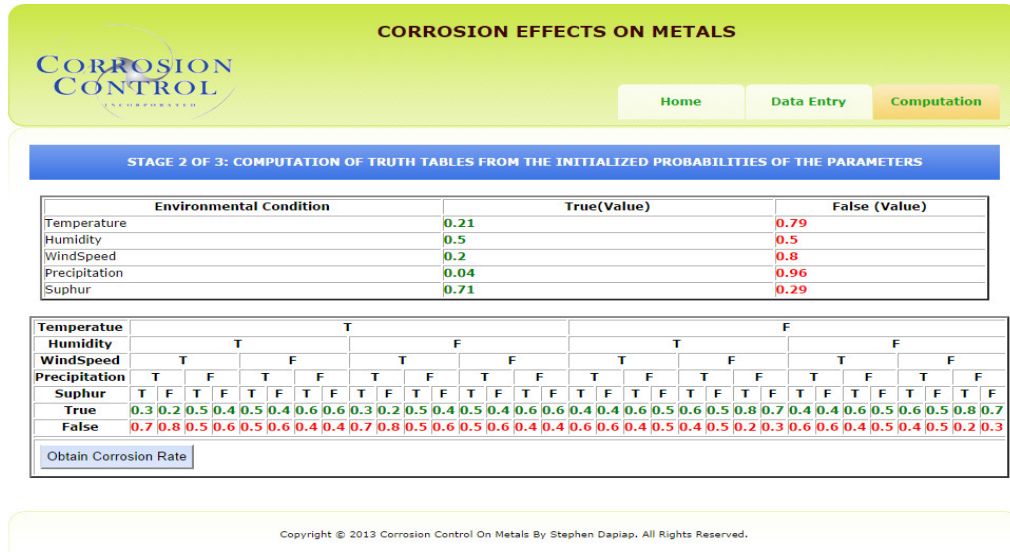


Figure 6. Results of Association for 2010

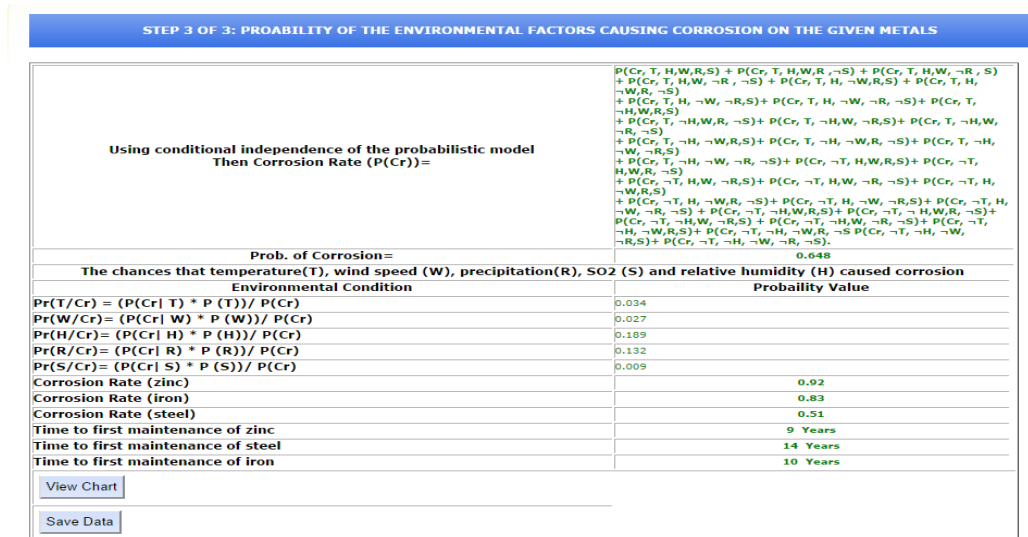


Figure 7. Computed Values for 2009

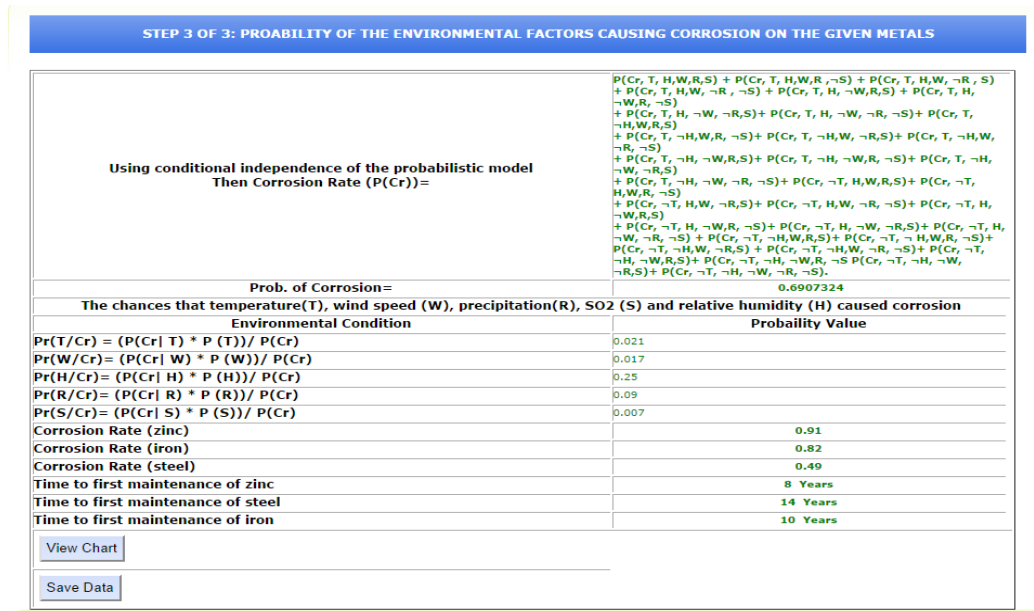


Figure 8. Computed Values for 2010

Table 9. Probable Contributions of Environmental Factors to Corrosion Levels

Environmental Factors	2009	2010	2011	2012	2013
Temperature	3.4	2.1	3.4	3	4.8
Wind	2.7	1.7	2.8	2.5	4.6
Humidity	18.9	25	18.7	21	13.4
Precipitation	13.2	9	13.3	10	12.9
SO ₂	0.1	0.7	1	0.9	1.1

The corrosion rates and the time to first maintenance of the metals due to the influence of the environmental factors are displayed in Figures 9 and 10.



Copyright © 2013 Corrosion Control On Metals By Stephen Dapiap. All Rights Reserved.

Figure 9. The Corrosion Rates

Figure 10. Time to First Maintenance

8. CONCLUSIONS

We have seen from this work that corrosion can be determined in a precise manner under certain environmental factors. It was noted that the corrosivity percentiles were distributed in the range 63.1% - 69.0%. The method in this work is more cost effective compared to continuous use of *expensive corrosion* sensors with frequent inspections and surveys to monitor and measure corrosion. Using this novel method will help in carrying out corrosion control maintenance at the right time and by this cut unnecessary maintenance activities that may incur huge avoidable costs since the consequences of corrosion are many and border on both safety and economic loss. The results also show that the effect of environmental factors can be determined; thus enabling us to use predictive analysis for information on timely control. Therefore future corrosion presence can be determined, corrosion rates computed and the right time to initiate corrosion maintenance can be estimated with historical environmental data bank thereby cutting unnecessary labour-induced costs due to multiple inspections and surveys activities for corrosion information. It may not necessarily have to be very large data bank before data mining can be applied. An area for further research here is on determining the corrosion rates of different metals of same thickness under the same environmental conditions for the same length of time. The interest here is, will the corrosion rates be different considering the different properties of the metals.

REFERENCES

- [1] American Galvanizers Association(2012). Hot-Dip Galvanizing for Corrosion Protection, a specifier's guide, pp 1-10. <http://galvanizeit.org/about-hot-dip-galvanizing/the-corrosion-problem>.
- [2] Song, W., Martin H. J., Hicks, A., Seely, D., Christopher A., Walton, C.A., Lawrimore II, B.W., Wang, T.P. &Horstemeyer, M.F. (2014). Corrosion Behaviour of Extruded AM30 Magnesium Alloy Under Salt-Spray and Immersion Environments. *Corrosion Science*.Volume 78, pp 353–368.
- [3] National Association of Corrosion Engineers International (2013). Study on Corrosion Costs and Preventative Strategies, pp1-2.
- [4] World Corrosion Organisation (2010). International Standardisation on Materials Selection for the Oil and Gas Industry.http://events.nace.org/euro/corrodia/Fall_2010/wco.asp
- [5] PRG Inc.(2002). Survey Vs Monitoring. Retrieved June 20, 2013 from: www.prginc.com.
- [6] Bowles, J.C. (1997). Some benefits of Long-term Thinking, *Pipeline & Gas Journal*, Published at www.undergroundinfor.com.
- [7] Fontana, M.G. (1986). *Corrosion Engineering*. McGraw-Hill, New York.
- [8] Samimi, Amir &Bagheri, Ali (2013). Investigation of Corrosion in Oil and Gas Pipelinesin Southwestern Iran. *World of Sciences Journal*; 2013[02], pp 114-127.
- [9] Silltow, J. (2014). Data Mining 101: Tools and Techniques. Security Control and Audit Ltd.<http://www.theiia.org/intAuditor/itaudit/archives/2014/august/data-mining-101-tools-and-techniques>.
- [10] McCarthy, J. (2000). Phenomenal Data Mining: From Data to Phenomena. *SIGKDD Explorations* 1(2): pp 24-29.
- [11] Belanche, L. &Vellido, A. (2013). Intelligent Data Analysis and Data Mining.http://www.lsi.upc.edu/~avellido/teaching/data_mining.html.
- [12] Hand, D. J. (1981). *Discrimination and Classification*.Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Chichester, U.K.
- [13] Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. (1997). From Data Mining to Knowledge Discovery in Databases, AAAI 97.Providence, Rhode Island.
- [14] Mannila, H. (2000). Theoretical frameworks for data mining.ACM SIGKDD Explorations Newsletter,Volume 1 Issue 2, January 2000, pp 30-32.
- [15] Heckerman, D. (1996). Bayesian Networks for Knowledge Discovery. In *Advances in Knowledge Discovery and Data Mining*, eds. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, 273–306. Menlo Park, Calif.: AAAI Press.
- [16] Ayello, F., Jain,S., Sridhar,N. & Koch, G.H. (2014). Quantitative Assessment of Corrosion Probability—A Bayesian Network Approach. *Corrosion: Vol. 70, No. 11*, pp. 1128-1147.

- [17] KamrunnaharMst&Urquidi-Macdonald Mirna (2009). Prediction of Corrosion Behavior Using Neural Network as a Data Mining Tool. Corrosion Science, Volume 52, Issue 3, March 2010, Pp 669–677.
- [18] Rao, V. S., Vidyavathi, S & Ramaswamy, G. (2010). Distributed Data Mining And Agent Mining Interaction And Integration: A Novel Approach.
- [19] Hand, David., Mannila, Heikki & Smyth, Padhraic. (2001). Principles of Data Mining. The MIT Press, Pp 2- 205
- [20] Kjaerulff, U.B. & Madsen, A. L. (2005). Probabilistic Networks-An Introduction to Bayesian Networks and Influence Diagrams. Aalborg University, pp 10-31.
- [21] Corrosionpedia (2014). Corrosion Rate. <http://www.corrosionpedia.com/definition/337/corrosion-rate>

Authors

Stephen Dapiap

The main author. Holds PhD in Computer Science and currently works in the department of Strategic Information of Institute of Human Virology, Nigeria.

ThankGodAnazodo

Holds BSc Computer Science and currently works in the department of Strategic Information of Institute of Human Virology, Nigeria.

Gregory Wajiga

A professor in the department of Computer Science, Federal University of Technology, Yola- Nigeria.

Michael Egwurube

A professor in the department of Mathematics, Federal University of Technology, Yola- Nigeria.

Musa Kadzai

Holds PhD in Mathematics and currently with the department of Mathematics, Federal University of Technology, Yola-Nigeria.

Nathaniel David Oye

Holds PhD in Information Technology and currently with the department of Computer Science, Federal University of Technology, Yola-Nigeria.