PERGAMON

Corrosion Science 42 (2000) 35–52

CORROSION SCIENCE

Artificial neural network modeling of atmospheric corrosion in the MICAT project

Salvador Pintos[a], Nestor V. Queipo[a], Oladis Troconis de Rincón[b], Alvaro Rincón[b], Manuel Morcillo[c,*]

[a]*Applied Computing Institute, College of Engineering, University of Zulia, Maracaibo, Zulia, 4005, Venezuela*

[b]*Center for Corrosion Studies, College of Engineering, University of Zulia, Maracaibo, Zulia, 4005, Venezuela*

[c]*National Center for Metallurgic Research, Gregorio del Amo Avenue, Madrid, 28040, Spain*

Received 15 June 1998; accepted 15 December 1998

Abstract

This paper presents an Artificial Neural Network(ANN)-based solution methodology for modeling atmospheric corrosion processes from observed experimental values, and an ANN model developed using the cited methodology for the prediction of the corrosion rate of carbon steel in the context of the Iberoamerican Corrosion Map (MICAT) Project, which includes seventy-two test sites in fourteen countries throughout Iberoamerica. The ANN model exhibited superior performance in terms of goodness of fit (sum of square errors) and residual distributions when compared against a classical regression model also developed in the context of this study, and is expected to provide reasonable corrosion rates for a variety of climatological and pollution conditions. Furthermore, the proposed methodology holds promise to be an effective and efficient tool for the construction of analytical models associated with corrosion processes of other metals in the context of the MICAT project, and, in general, in the modeling of corrosion phenomena from experimental data. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: A. Steel; B. Modeling studies; C. Atmospheric corrosion

* Corresponding author. Fax: +349-341-534-7425.
E-mail address: grupom@cenim.csic.es (M. Morcillo)

0010-938X/00/$ - see front matter
PII: S0010-938X(99)00054-2

1. Introduction

The Iberoamerican Atmospheric Corrosion Map Project (MICAT-Mapa Iberoamericano de Corrosión Atmosférica) was established as a resolution of the Board of Directors of the V Centennial Science and Technology Program for Development (CYTED) during a meeting in La Habana, Cuba in 1988. The project, originally conceived at the Iberoamerican Corrosion Congress in Maracaibo, Venezuela in 1986, includes seventy-two test sites exposed to the atmosphere in fourteen countries throughout Iberoamerica, namely: Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, Portugal, Peru, Mexico, Venezuela, Panama, Spain, and Uruguay. Fig. 1 displays the network of test sites. Along the lines of the ISOCORRAG, and ICP/UNECE projects, the MICAT project has three main objectives: (i) construct the corrosion map for Iberoamerica; (ii) provide a better understanding of atmospheric corrosion phenomena; and (iii) identify mathematical models that could predict the corrosion rate of metals in the atmosphere as a function of meteorological and pollution variables for Iberoamerica. Regarding objective (iii), a significant amount of work has been done in order to establish analytical expressions or models for the behavior of metals in the atmosphere as a function of easily determined variables (instead of long lasting experiments). The task at hand has shown to be difficult, mainly because of the complexity (non-linearities) associated with the physicochemical processes responsible for atmospheric corrosion phenomena.

Most of the predictive models used to date, are regression models that fit the data available such that their mean square error is minimized. However, these models have been shown to be, effective only in very restrictive areas, and limited to capture the nonlinear nature of the corrosion process. Hence, the continuous search for mathematical models that could predict the atmospheric corrosion rate for rather general climatic and contamination level conditions [1–3].

Artificial Neural Network (ANN) Modeling, in particular the one based on a class of ANN called multilayer perceptron, has recently emerged as a promising area in corrosion research [4], because of the potential of the ANN to predict any complex process with arbitrary precision provided its architecture and a set of parameters are properly set [5]. This study presents the development of an ANN based model for the prediction of the corrosion rate of carbon steel (Fe) in the context of the MICAT project (hence, considering a broad spectrum of climatological and pollution conditions), and evaluate its performance (mean square error and residual distribution) against a classical regression model.

The remainder of the paper is structured as follows. Section 2 provides a detailed description of the problem under consideration, while section 3 presents the adopted solution methodology including an introduction to ANN and the ANN modeling process. The analysis and discussion of the results obtained using the developed ANN model and its relative performance when compared against a regression model, are the subject of section 4. Conclusions and recommendations

Fig. 1. Network of test sites in the MICAT project.

Table 1
Learning data set (Part I) with the prefixes A, B, CO, CR, CU, and CH, denoting test sites in Argentina, Brazil, Columbia, Costa Rica, Cuba and Chile, respectively

Test site	T	RH	TOW	P	SO_2	Cl^-	Fe
A011	14.1	79.2	0.681	817	5.0	30.0	24.9
A012	13.9	78.8	0.707	805	5.0	40.2	54.8
A013	14.5	80.3	0.736	1226	5.0	70.2	66.0
A022	17.1	71.5	0.482	983	5.0	10.0	16.1
A023	17.0	73.6	0.494	1420	5.0	9.0	12.4
A031	20.6	75.8	0.666	2158	5.0	1.5	5.7
A033	22.1	74.7	0.633	1720	5.0	1.5	5.5
A042	20.0	49.3	0.097	111	5.0	1.5	4.5
A043	18.3	50.8	0.099	93	5.0	1.5	5.6
A051	−2.0	83.8	0.307	114	5.0	30.2	37.4
A061	17.0	77.5	0.593	1178	6.2	1.5	25.3
B011	21.5	73.5	0.482	847	0.8	8.9	8.6
B012	20.9	74.9	0.482	1167	1.3	7.4	11.5
B013	21.2	74.6	0.482	996	1.7	1.6	13.1
B021	23.8	89.3	0.482	1122	23.8	8.6	52.5
B022	22.9	91.1	0.482	1471	20.7	6.7	47.2
B031	24.8	77.2	0.582	605	9.5	359.8	159.8
B032	24.5	78.6	0.582	985	5.2	174.7	194.6
B041	22.7	72.7	0.579	960	40.4	4.4	98.7
B042	22.5	71.0	0.579	876	59.1	10.7	161.2
B061	19.8	74.9	0.648	1409	28.0	1.5	14.6
B062	19.6	76.0	0.648	1910	28.0	1.5	23.4
B063	19.4	75.2	0.648	1034	28.0	1.5	24.2
B071	20.1	79.1	0.598	1305	46.5	13.9	127.1
B073	20.2	83.6	0.598	1229	26.6	13.7	73.1
B081	26.1	87.7	0.682	2395	5.0	1.5	19.4
B101	20.4	72.0	0.442	1440	5.0	1.5	12.9
CO12	27.6	87.0	0.966	940	14.2	36.4	30.6
CO13	28.2	87.1	0.976	940	8.9	69.5	54.0
CO22	11.4	89.5	0.800	1800	5.0	1.5	17.7
CO23	14.2	73.4	0.800	1800	5.0	1.5	19.6
CR31	22.9	88.3	0.838	3677	8.2	20.5	69.3
CR41	18.9	83.4	0.695	1845	4.2	15.5	16.6
CU11	25.2	79.5	0.468	1591	37.1	15.8	36.0
CU12	25.4	79.4	0.468	1303	36.5	10.9	26.4
CU13	25.3	79.4	0.468	1447	19.8	10.9	29.0
CU31	23.9	81.0	0.571	1488	16.0	12.3	32.5
CU33	23.9	81.0	0.571	1488	15.3	4.2	34.1
CH11	14.2	71.0	0.130	355	21.0	4.4	29.3
CH12	14.2	68.3	0.372	367	17.9	4.2	44.1
CH41	12.2	82.0	0.686	1294	63.9	14.3	167.2
CH51	−2.3	85.0	0.264	114	4.5	14.2	24.1

Table 2
Learning data set (Part II) with the prefixes EC, E, M, and PA, denoting test sites in Ecuador, Spain, Mexico, and Panama, respectively

Test site	T	RH	TOW	P	SO_2	Cl^-	Fe
EC12	26.9	82.1	0.661	635	2.7	1.3	22.6
EC21	12.9	66.0	0.409	554	1.0	0.4	7.7
EC31	23.0	81.0	0.776	631	3.0	58.4	66.1
E011	12.0	68.8	0.384	652	16.2	1.5	19.3
E013	11.1	62.7	0.241	334	16.2	1.5	19.7
E021	13.0	62.8	0.416	359	3.8	3.9	12.6
E022	13.0	62.8	0.416	359	8.9	3.9	10.8
E023	13.0	62.8	0.416	359	6.3	3.9	10.6
E031	16.8	65.1	0.170	443	18.6	3.8	17.9
E033	15.7	65.5	0.156	655	14.8	5.1	18.1
E041	18.1	65.2	0.390	554	8.3	1.5	20.3
E042	17.0	62.8	0.302	521	5.7	1.5	19.4
E043	17.2	61.9	0.316	374	1.9	1.5	21.0
E051	16.3	59.3	0.151	416	10.3	1.5	12.3
E053	15.6	57.5	0.274	266	2.8	1.5	6.4
E061	16.9	71.5	0.369	1828	38.1	11.6	27.0
E062	15.8	68.1	0.314	1704	34.7	10.3	29.6
E063	15.1	69.6	0.249	1312	45.7	28.2	28.1
E071	14.3	69.0	0.339	271	4.1	10.4	14.1
E072	11.0	77.0	0.450	510	3.7	10.2	12.8
E073	10.8	74.0	0.447	451	3.0	8.7	18.2
E081	8.8	72.0	0.100	738	9.1	1.8	3.3
E082	6.9	72.0	0.100	624	8.9	1.6	3.6
E083	7.8	72.0	0.100	681	9.0	1.7	4.8
M011	16.0	62.0	0.265	747	15.6	1.5	15.4
M012	15.2	64.5	0.288	747	5.6	1.5	8.6
M013	15.6	63.2	0.277	747	17.6	1.5	5.1
M022	21.0	56.0	0.212	1724	9.9	1.5	11.4
M023	21.0	56.0	0.200	1372	7.1	1.5	13.7
M041	28.0	78.0	0.564	834	6.5	34.4	22.6
M042	27.0	77.0	0.581	792	15.2	22.9	24.8
M043	28.0	73.3	0.570	985	7.2	14.0	18.7
PA11	26.5	69.2	0.558	1686	38.5	4.8	25.6
PA12	27.2	70.3	0.562	1739	20.0	8.5	23.0
PA13	27.1	74.0	0.588	1387	11.0	9.6	28.3
PA21	27.1	75.5	0.620	3815	63.0	11.0	125.0
PA22	27.3	76.5	0.658	4656	50.5	18.7	93.3
PA23	26.9	75.5	0.691	3622	28.7	20.1	113.7
PA32	27.2	68.7	0.568	2082	20.3	12.7	14.2
PA33	27.3	71.6	0.580	2171	9.1	24.4	20.1
PA41	27.1	65.3	0.531	2427	13.7	4.5	23.0

of areas worthy of future research efforts in the modeling of corrosion processes using ANN are also included.

2. Problem definition

The problem of interest here corresponds to the identification of a model for the prediction of atmospheric corrosion of a given metal in terms of metereochemical variables in the context of the MICAT project. Specifically, the model is expected to provide a prediction of the corrosion rate of carbon steel (Fe in μm/year) as a function of the time of wetness (TOW), chloride deposition rate (Cl^- in mg Cl^- d), sulfate deposition rate (SO_2 in mg SO_2/m^2 d), relative humidity (RH in %), precipitation (P in mm), and temperature (T in °C). Furthermore, the model must be constructed considering the actual values of the corrosion rate of the carbon steel and the corresponding metereochemical variables at test stations throughout Iberoamerica (Tables 1–5).

Mathematically, the problem under investigation is one of function estimation and may be stated as follows: given a set of n observations of metereochemical variables and the corresponding observed corrosion rate values, denoted by $\bar{x}_i$, $\hat{f}_i$,

Table 3
Learning data set (Part III) with the prefixes PA, PE, PO, U, and V, denoting test sites in Panama, Peru, Portugal, Uruguay and Venezuela, respectively

Test site	T	RH	TOW	P	SO_2	Cl^-	Fe
PA42	26.9	69.1	0.560	2210	6.9	9.8	20.9
PA43	27.4	70.7	0.579	2123	3.9	11.8	23.2
PE24	19.1	86.0	0.837	14	17.6	39.4	35.2
PE32	18.8	83.5	0.761	13	28.9	19.8	34.0
PE41	16.4	37.0	0.003	17	5.0	1.5	15.0
PE42	17.2	33.3	0.003	89	5.0	1.5	16.5
PE61	25.4	84.0	0.500	1523	5.0	1.5	15.7
PE62	25.8	82.9	0.500	1656	5.0	1.5	12.9
PO11	15.9	62.1	0.379	1129	68.2	161.8	78.9
PO12	16.9	56.4	0.315	1129	67.2	82.6	106.4
PO31	17.2	71.3	0.351	685	9.1	6.9	27.1
PO32	16.1	66.0	0.350	685	7.2	5.1	24.5
PO33	17.4	72.4	0.400	685	7.1	6.1	29.9
U011	16.8	73.5	0.586	1185	1.0	2.2	8.2
U021	17.1	76.2	0.479	1036	16.3	7.6	37.6
U041	−2.3	90.0	0.354	473	3.2	30.2	56.0
V011	27.7	78.0	0.419	809	6.6	73.1	30.1
V021	28.0	74.5	0.462	311	2.7	24.3	20.9
V031	26.6	77.1	0.571	362	3.9	32.0	16.6
V032	28.1	74.2	0.427	344	2.6	25.7	15.1
V041	27.7	75.0	0.513	983	19.0	15.8	23.0
V061	26.5	84.0	0.684	807	1.4	12.8	37.3

Table 4
Validation data set

Test site	T	RH	TOW	P	SO_2	Cl^-	Fe
A041	18.0	50.6	0.114	35	5.0	1.5	4.6
A052	−3.1	84.0	0.276	114	5.0	30.2	35.9
B023	23.0	90.4	0.482	1444	24.5	5.2	48.5
B072	23.2	80.0	0.598	1152	33.9	14.2	61.3
CO11	27.6	87.0	0.971	940	7.7	43.7	15.9
CR11	27.6	80.2	0.562	1598	5.3	45.2	61.6
CU32	23.9	81.0	0.571	1488	8.9	5.9	28.6
EC11	26.1	71.4	0.554	936	4.2	1.5	19.5
E032	15.9	63.0	0.168	705	16.6	4.3	16.1
E052	15.8	57.7	0.126	239	5.4	1.5	6.8
M021	21.0	56.0	0.190	1352	6.7	1.5	15.2
PA31	27.2	68.0	0.559	2593	20.9	7.4	25.7
PE31	19.2	85.0	0.761	13	28.9	19.8	27.5
PE52	12.2	67.0	0.325	792	0.0	0.0	1.7
U031	17.3	80.1	0.608	1515	0.7	5.0	14.0
V051	26.5	77.0	0.570	608	1.6	23.1	53.0

with $i = 1 \ldots n$: find $f(\vec{x})$ such that $\Sigma_{i=1}^{n}(f(\vec{x}_i) - \hat{f}_i)^2$ is minimized wherein, $f(\vec{x})$ is a function (model) providing the corrosion rate of carbon steel (Fe), and $\vec{x}$ is a vector of the metereochemical variables: TOW, Cl^-, SO_2, RH, P, T.

3. Solution methodology

The proposed solution methodology is based on the use of Artificial Neural Networks (ANN), specifically the so called multilayer perceptron model. This model is an universal approximator that is able to predict any sampled process with arbitrary precision provided its architecture and a set of parameters are properly established. This section, after presenting an introduction to ANN,

Table 5
Test data set

Test site	T	RH	TOW	P	SO_2	Cl^-	Fe
A032	20.9	73.8	0.631	2624	5.0	1.5	5.8
A053	−2.9	84.5	0.295	240	5.0	30.2	41.1
B033	24.2	77.0	0.582	716	4.4	167.7	128.4
CO21	14.1	81.4	0.800	1800	5.0	1.5	13.6
CR21	25.3	88.4	0.763	3531	4.7	138.1	371.5
CH21	14.0	82.0	0.762	463	18.7	8.4	35.5
E012	10.6	64.5	0.271	495	16.2	1.5	22.5
PE51	12.2	67.0	0.325	632	0.0	0.0	1.0
V021	27.1	76.6	0.496	263	4.2	38.7	23.9

discusses the ANN-based modeling process carried out in this study and the construction, for comparison purposes, of a linear regression model along the lines of those reported in the literature for the MICAT project [7].

3.1. Introduction to artificial neural networks

These biologically inspired models have found applications in many different areas such as, classification of data, pattern recognition, clustering analysis, and function approximation. This section provides an introduction to an ANN model called multilayer (three layer) perceptron used in function approximation tasks such as the one faced in this study. A more detailed discussion on other ANN models and applications may be found elsewhere [6]. The introduction to the ANN model under consideration (Fig. 3), is given in terms of basic definitions (artificial neurons, links, layers, weights, threshold, activation function), architecture, and learning process.

3.1.1. Basic elements

The basic processing elements are the so-called artificial neurons. With reference to Fig. 2, an *artificial neuron* basically evaluates an *activation function* taking as argument a weighted sum of its inputs. Commonly used activation functions are the logistic functions, an example of a logistic function is given by Eq. (1); note that the output is limited to values in the interval [0, 1] and in general, the output data must be subject to a normalization process.

$$g(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

The neurons are arranged in *layers* (Fig. 3), and there are three different layers, namely, input, hidden, and output layers. There are some 'special' neurons, the ones at the input layer and the bias neurons, that have as output their corresponding inputs. The *bias neurons* are introduced to facilitate the consideration of the *threshold* value associated with each neuron. Furthermore, the neurons are connected among them through channels called *links*, with the

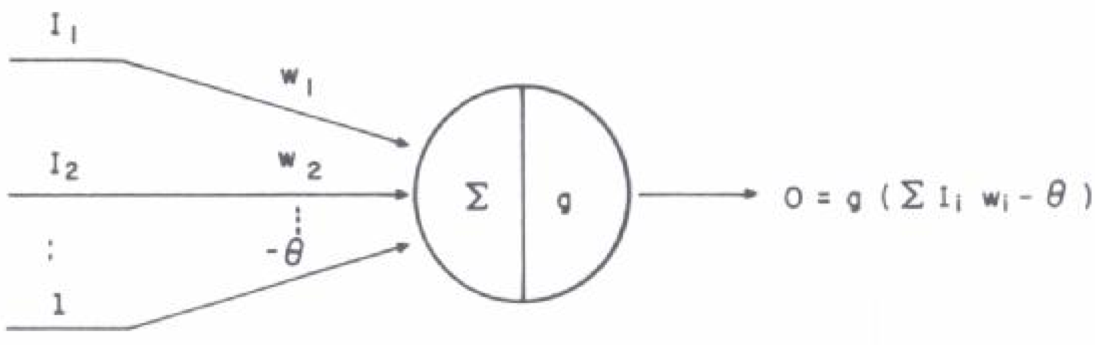

Fig. 2. Illustration of the information processing ability of an artificial neuron.

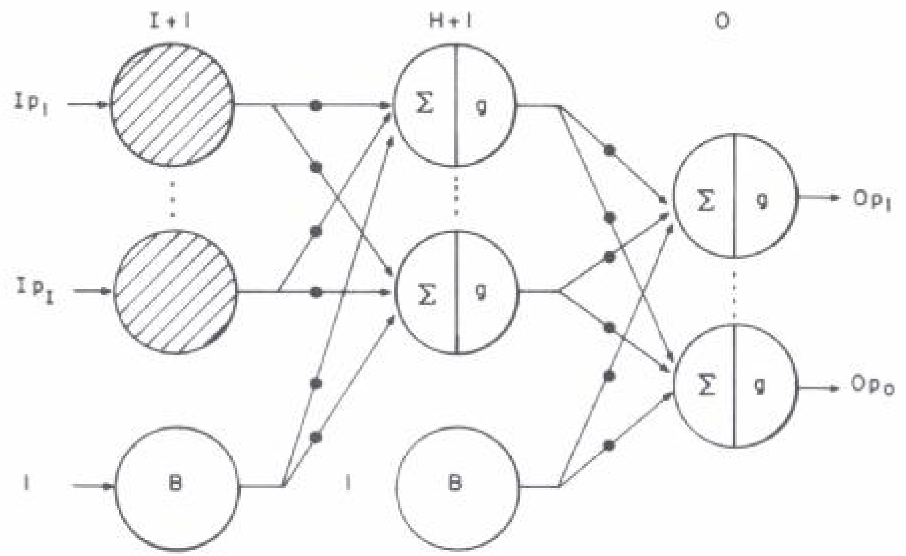

Fig. 3. Schematic of the three-layer perceptron ANN model.

strength of the connections specified as real numbers called *weights*. The links usually connect neurons in consecutive layers from left to right.

The output of the ANN model can be written as:

$$O_p(I_p) = G[W \times (G[V \times I_p])] \tag{2}$$

where I_p denotes the input values, V and W the weight matrices between the input layer and hidden layer, and the hidden layer and output layer, respectively. The symbol G represents the application of the logistic function to each of the elements of its argument.

3.1.2. Architecture

It can be shown that a three-layer perceptron (the one used in this investigation) is, for function approximation purposes, as capable as ANN with higher number of hidden layers, and can estimate a function with arbitrary precision provided the number of neurons in the hidden layers and the weights are properly set [5]. As a result, our attention will be restricted to three-layer perceptrons.

The number of neurons in the input and output layers are established by the number of input, and output variable(s) of the function(s) to be estimated. The number of neurons in the hidden layer can be settled through different criteria; for example, limiting the number of parameters of the model (weights) to be a fraction of the total number of data points available to the *learning* process (to be discussed next).

3.1.3. Learning

This process makes reference to the identification of a set of weights that, for a given architecture, minimize the sum of the square of the model errors (Eq. (3)); the errors represent the difference between the ANN model output values and the observed function values.

$$\min \text{Error}(V, W) = \sum_{i=1}^{n} \|O_{pi} - \hat{f}_i\|^2 \tag{3}$$

The nonlinear nature of the aforementioned error function makes it nonconvex and prone to have local optima. The most common ANN learning algorithm is a gradient based optimization procedure called Backpropagation [6], which essentially, modifies the weights (from a starting set of weights) by moving in the direction contrary to the error function gradient thus ensuring a lower error value (Eq. (3)). The cited algorithm, as any other gradient-based optimization procedure, is limited by the fact that it is sensitive to the set of initial weights, and it may get trapped in local optima.

3.2. ANN modeling

The neural network modeling process used in this study may be described in four stages: (i) preprocessing of the original data set (identification of outliers); (ii) partitioning of the preprocessed data set into learning, validation and test sets; (iii) ANN model architecture setting, learning and testing; and (iv) implementation.

3.2.1. Preprocessing of the original data set

Under the assumption of the continuous nature of the corrosion process, two test sites with similar metereochemical values should have similar corrosion rates. In order to detect data values violating this premise, a cluster analysis was conducted on the metereochemical variable values of the *original* data set. In general, observations within a cluster, with corrosion rates atypical for a given cluster, were eliminated. In addition, since the input variables were so different in ranges, in order to facilitate the learning process (to be explained shortly) the input variables were scaled using the following transformation: $\hat{x}=(x-\mu)/\sigma$, with

Table 6
Fundamental statistical information associated with the data set used for the construction of the ANN model

Variable	Minimum	Maximum	Mean	S.D.
T	−3.100	28.200	19.202	7.024
RH	33.300	91.100	73.238	10.267
TOW	0.003	0.976	0.479	0.208
P	13.000	4656.000	1093.423	831.236
SO_2	0.000	68.200	13.930	14.928
Cl^-	0.000	359.800	19.927	42.623
Fe	1.000	371.500	36.212	46.304

x, and $\hat{x}$ denoting, an input variable and scaled input variable, respectively, and the symbols μ and σ representing the mean and standard deviation among the observed values for the metereochemical variables (Table 6).

3.2.2. Partitioning the data set

The resulting data from the preprocessing process conducted in the previous step, are divided into three disjoint data sets: learning, validation and test data sets. The *learning* data set is the one used to train (specify the weights) of the ANN. The *validation* data set is the one used in conjunction with the learning data set to identify when to stop the learning process so that the resulting ANN exhibits good generalization properties. The *test* data set allows the assessment of the prediction capabilities of the ANN model. The ANN is evaluated using as performance criteria mean square error and residual distributions over learning and test data sets. The data set was sorted by country and test site and the learning, validation, and test data were constructed using an stratified sampling procedure. The selected sampling procedure aims to data sets that include observations of all countries and test sites. The number of observations per data set were assigned to be one hundred-and-five, sixteen and nine, for the learning, validation and test sets, respectively.

3.2.3. ANN model architecture setting

Since the nodes for the input and output layer are set by the number of metereochemical variables (six) and the number of variables to be predicted (one), only the number of hidden neurons needs to be establish before the ANN model architecture is completed. In this work the number of hidden neurons was established so that the number of parameters (weights) during the learning process is a fraction (approximately 40%) of the number of observations available in the learning data set. Hence avoiding the possibility of the ANN model overfitting the data.

3.2.4. Learning and testing

The learning process was conducted using as optimization procedure the standard Backpropagation algorithm, with weight updates each time the complete learning data set was considered. The initial set of weights were obtained using Montecarlo optimization, hence reducing the possibility of getting trapped in local optima. For a given set of initial weights, the learning process was stopped when the mean square error in the *validation data set* started to grow for additional number of iterations in the optimization procedure. The 'best' set of weights was selected as the one with the lowest possible mean square error and relative good performance on the validation data test. After the conclusion of the ANN learning process the resulting model was evaluated in terms of its mean square error and residual distributions over the *test data set*, and its relative performance when compared against a regression model with the same structure of one reported in the literature [7] created using the learning and validation data sets.

3.2.5. Implementation

The solution methodology is implemented using a combination of both commercial and academic software. The Statistical Analysis System (SAS) [8] was used, for the cluster analysis (PROC CLUSTER) in the preprocessing of the original data set, for the stratified sampling conducted in the partitioning of the data set, and in the construction of the regression model (PROC REG) in the testing phase of the ANN modeling process. The Artificial Neural Network models were created with the assistance of the Stuttgart Neural Network Simulator (SNNS) [9].

4. Results and discussion

This section presents an Artificial Neural Network (ANN) model for the prediction of the corrosion rate of carbon steel (Fe) as a function of relevant metereochemical variables (TOW, Cl^-, SO_2, RH, P, and T) in the context of the MICAT project, and discusses the performance exhibited by the ANN model in terms of goodness of fit (mean square error), and residual distributions for training and testing data sets. In addition, the corresponding results for a linear regression model, also developed in the context of this study, are included for comparison purposes.

The ANN model obtained following the methodology described in section 3 is illustrated in Fig. 4, and its output can be expressed in matrix form as specified by Eq. (2), with $V[H \times (I + 1)]$ and $W[O \times (H + 1)]$ given by the following weight

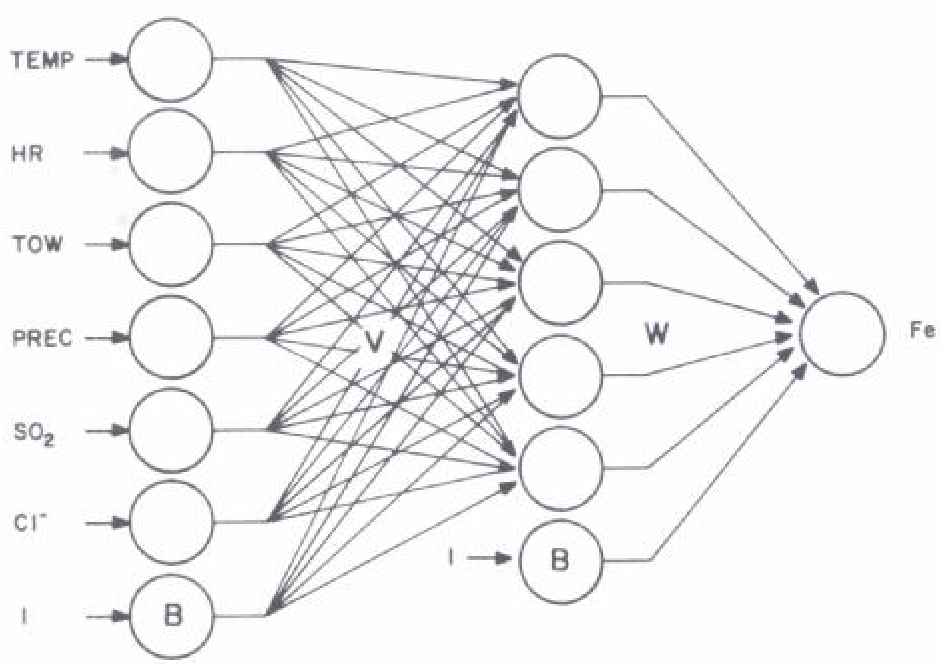

Fig. 4. Artificial Neural Network model.

matrices

$$V=\begin{pmatrix} 0.0745 & -0.9768 & 0.9852 & -0.4833 & -0.7987 & -1.0064 & -3.0598 \\ -3.3410 & -2.4293 & 1.8049 & 2.3564 & 3.1194 & -4.6681 & -3.2227 \\ 0.4417 & -1.9811 & 1.5018 & -0.7849 & -0.0004 & -2.4218 & 2.0833 \\ -0.4767 & 1.9384 & 1.2480 & 0.1868 & 1.1675 & 2.6001 & 1.8650 \\ -2.4737 & -1.1378 & -0.1231 & 1.6081 & 0.6292 & -0.1397 & 3.2680 \end{pmatrix}$$

$$W=[-0.3466 \quad 3.1603 \quad 0.0620 \quad -1.4238 \quad 2.9312 \quad -3.1031].$$

The ANN model captures well-known nonlinear interactions between the corrosion rate and some of the most significant metereochemical variables such as TOW, SO_2 and Cl^-, as it is illustrated in Figs. 5–8. In general, each figure depicts the corrosion behavior as a function of one or more of the metereochemical variables; the rest of the variables are considered either constant (at their mean values) or, if they are highly correlated with the variable(s) under consideration, at the values specified by simple linear models. The climatological variables are highly correlated as shown in Table 7, which displays the correlation value and the probability of a null correlation for the different pairs of climatological variables.

Figs. 5 and 6 shows the behavior of the corrosion rate with respect to chloride deposition rate (Cl^-), and sulfate deposition rate (SO_2), respectively. In both cases, as expected, the corrosion rate increases with higher values of Cl^- and SO_2, respectively, with the former exhibiting a higher nonlinear interaction with the

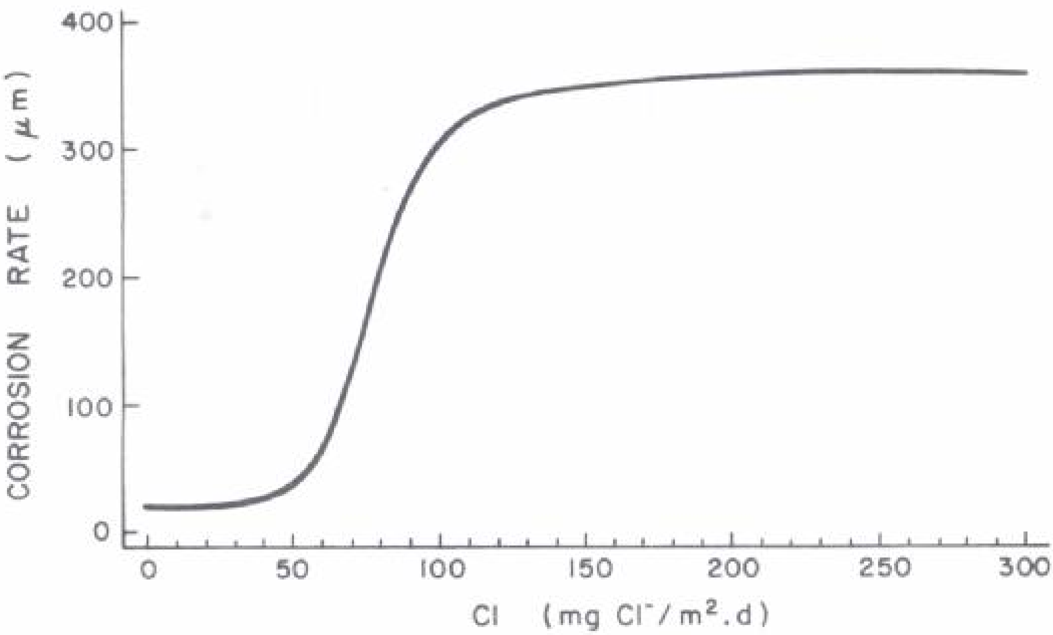

Fig. 5. Corrosion rate of carbon steel (Fe) vs chloride deposition rate Cl^-.

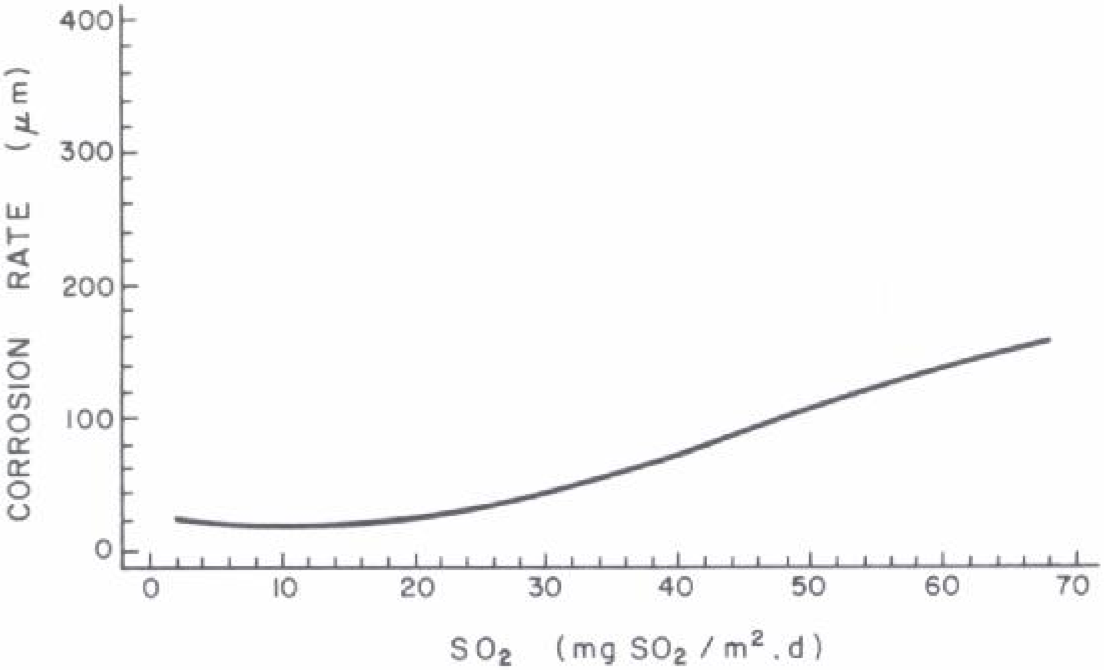

Fig. 6. Corrosion rate of carbon steel (Fe) vs sulfate deposition rate SO_2.

variable of interest than the latter. Incidently, the behavior of the corrosion rate with respect to chloride deposition rate (Cl^-) is consistent with worldwide data sets associated with pure marine atmospheres (not contaminated with SO_2) [10].

Furthermore, Fig. 7 depicts the ANN model response surface as a function of Cl^- and time of wetness TOW having a constant SO_2 (mean value) and simple linear regression models for the rest of the variables; specifically, $T = 12.45 + 15.59$

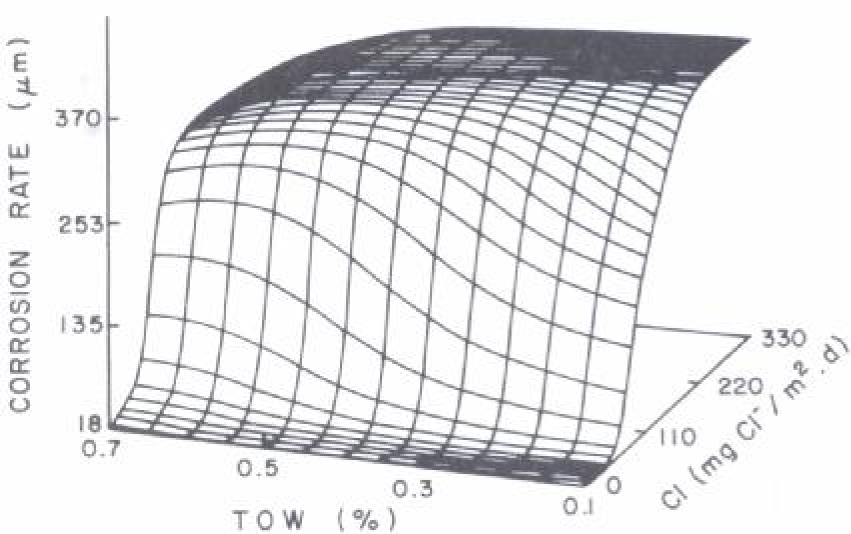

Fig. 7. Corrosion rate of carbon steel (Fe) as a function of chloride deposition rate Cl^- and time of wetness (TOW).

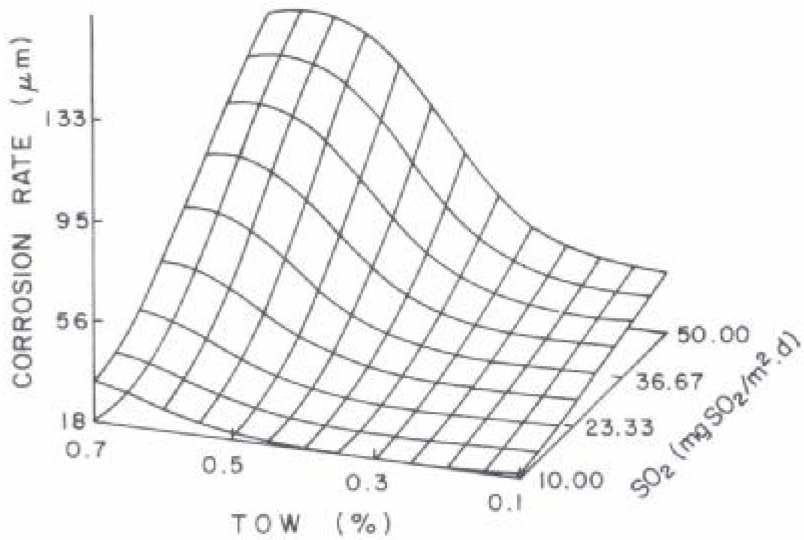

Fig. 8. Corrosion rate of carbon steel (Fe) as a function of sulfate deposition rate SO_2 and time of wetness (TOW).

$\times$ TOW, RH = 55.73 + 35.88 $\times$ TOW, and P = 241.00 + 1813.00 $\times$ TOW . Similarly, the ANN model response surface as a function of TOW and SO_2 is shown in Fig. 8, with Cl^- assumed constant (mean value) and the aforementioned linear regression models establishing the values for the variables T, RH, and P. In both instances the figures show significant nonlinear interactions among the considered variables.

The ANN model shows superior fitting (learning and validation data sets) capabilities exhibiting a mean square error (MSE) and (R^2) coefficient equal to 140.45 and 0.90, respectively, which represents a 65% (27%) lower (higher) value than those provided by a quadratic regression model constructed as part of this study (similar to the one reported by Morcillo [7], with a 95% degree of

Table 7
Correlation values (top) and probability of a null correlation (bottom) among the climatological variables

	T	RH	TOW	P
T	1.00	0.18	0.47	0.44
	0.00	0.05	0.00	0.00
RH	0.18	1.00	0.69	0.32
	0.05	0.00	0.00	0.00
TOW	0.47	0.69	1.00	0.44
	0.00	0.00	0.00	0.00
P	0.44	0.32	0.45	1.00
	0.00	0.00	0.00	0.00

significance level regarding the linear coefficients. The linear regression model obtained is shown in Eq. (4).

$$\text{Fe} = b_0 + \text{Cl}^-(b_1 + b_2 \times P + b_3 \times \text{RH}) + b_4 \times \text{TOW} \times \text{SO}_2 \tag{4}$$

with $b_0 = 6.8124$, $b_1 = -1.6907$, $b_2 = 0.0004$, $b_3 = 0.0242$, and $b_4 = 2.2817$.

In addition, the residuals of the ANN model are smaller and more symmetrically distributed when compared with its regression model counterpart (Fig. 9).

The prediction performance of the ANN model over a set of data not used during the training process was also superior to the one exhibited by the regression analysis model. The MSE over the validation data provided by the ANN model was approximately four times smaller that the one provided by the linear regression model (220 vs 1082). Within this data set, the minimum and maximum error for the ANN model (Regression Model) were −11.9 and 32.9 (−9.5 and 88.7), respectively; note again the more symmetric distribution of the errors for the ANN model.

5. Conclusions

1. An Artificial Neural Network-based methodology for the modeling of atmospheric corrosion is described. The methodology involves preprocessing and partitioning of the experimental data set, ANN model architecture setting, construction and testing of the ANN model, and implementation.
2. Using the proposed solution methodology an ANN model was constructed and evaluated using experimental data associated with the corrosion rate of carbon

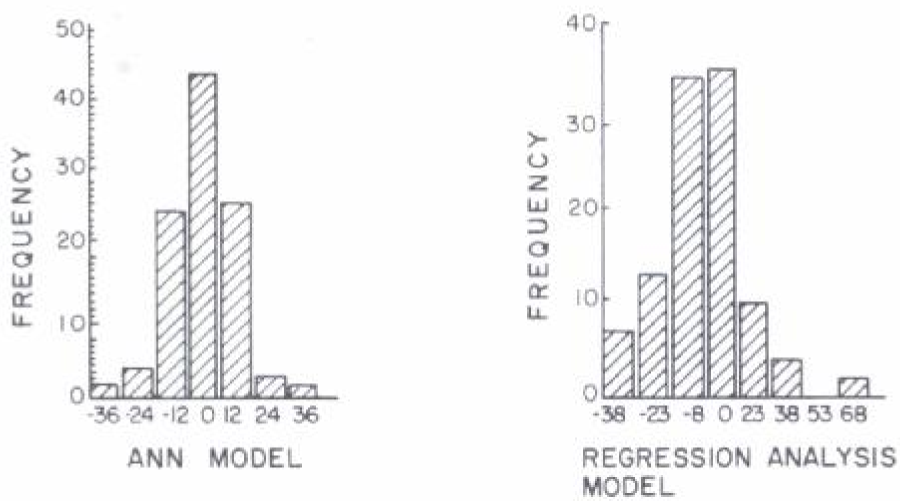

Fig. 9. Residual distributions for the ANN and regression analysis models.

steel (Fe) as a function of the time of wetness (TOW), chloride deposition rate (Cl^-), sulfate deposition rate (SO_2), relative humidity (RH), precipitation (P), and temperature (T), in the context of the MICAT project. The ANN model (i) reproduced some well-known nonlinear interactions among the variables of interest, and, (ii) provided excellent results regarding goodness of fit (MSE) and residual distributions on training and testing data sets.

3. Specifically, regarding item (ii) the application of the ANN model on the training and validation (testing) data sets resulted in a MSE equal to 140 (220), which, was 63% (80%) lower than those obtained using a linear regression model developed from the same data. In addition, the residual distributions reported by the ANN model when using training and testing data sets were significantly smaller and more symmetric than the corresponding to the aforementioned regression model.
4. The ANN model developed as part of the study holds promise to be useful in the prediction of carbon steel corrosion rates under a wide spectrum of climatological and pollution conditions in Iberoamerica. The effectiveness and efficiency of the proposed solution methodology should be evaluated in the modeling of atmospheric corrosion processes associated with other metals in the MICAT project, and, in general, in the modeling of a wider range of corrosion processes from experimental data.

Acknowledgements

The authors would like to thank the Consejo de Desarrollo Científico y Humanístico, of the University of Zulia, and the Consejo Nacional de Investigaciones Cientificas, Venezuela, for prividing financial support to conduct this work. Also special thanks are given to the MICAT Iberoamerican Group for sharing their experimental data.

References

[1] R. Legault, P. Pearson, Atmospheric corrosion in marine environments, Corrosion 34 (12) (1978) 433.
[2] A. Sanchez, C. Perez, P. Merino, L. Espada, Modelo teorico para la corrosión atmosférica del Cinc, in: Fourth Iberoamerican Congress of Corrosion and Protection, Argentina, 1992. I. p. 61.
[3] S. Feliu, M. Morcillo, J. Feliu, The prediction of atmospheric corrosion from meteorological and pollution parameters, Corrosion Science 34 (3) (1993) 403.
[4] I. Helliwell, M. Turega, R. Cottis, Neural networks for corrosion data reduction, Corrosion 96 (1996) 739.
[5] R. Hecht-Nielsen, Neurocomputing, Addison Wesley, 1989.
[6] D. Rumelhart, J. McClelland, Parallel Distribution Processing: Explorations in the Microstructure of Cognition, vol. 1 & 2, MIT Press, 1986.
[7] M. Morcillo, Atmospheric corrosion in Iberoamerica: The MICAT Project, in: W.W. Kirk, H.H.

Lawson (Eds.), Atmospheric Corrosion, American Society for Testing and Materials, Philadelphia, 1994 ASTM STP 1239.

[8] SAS Institute Inc., SAS Quality Control Reference Manual, Ver. 6.0, 1st ed., Cary: NC, USA, 1990.

[9] University of Stuttgart, Stuttgart Neural Network Simulator (SNNS) User Manual, Ver. 4.1, Report No. 6/95, available through anonymous ftp at ftp.informatik.uni-stuttgart.de in the directory /pub/SNNS, Germany, 1995.

[10] M. Morcillo, B. Chico, E. Otero, L. Mariaca, Effect of marine aerosol on atmospheric corrosion, Materials Performance 38 (1999) 72.