

Article

Design of Predictive Models to Estimate Corrosion in Buried Steel Structures

Lorena-de Arriba-Rodríguez, Vicente Rodríguez-Montequín *, Joaquín Villanueva-Balsera 
and Francisco Ortega-Fernández

Department of Mining Exploitation and Prospecting, University of Oviedo, 33004 Oviedo, Spain;
lorena.dearriba@api.uniovi.es (L.-d.A.-R.); jmvillanueva@uniovi.es (J.V.-B.); fdeasis@uniovi.es (F.O.-F.)

* Correspondence: montequi@uniovi.es; Tel.: +34-985-104-272

Received: 16 October 2020; Accepted: 24 November 2020; Published: 26 November 2020



Abstract: Corrosion is the main mechanism of the degradation of steel structures buried in the soil. Due to its aggressiveness, the material gradually loses thickness until the structure fails, which may cause serious environmental problems. The lack of a clearly established method in the design leads to the need for conservative excess thicknesses to ensure their useful life. This implies inefficient use of steel and an increase in the cost of the structure. In this paper, four quantitative and multivariate models were created to predict the loss of buried steel as a function of time. We developed a basic model, as well as a physical and an electrochemical one, based on multivariate adaptive regression spline (MARS), and a simpler model for comparative purposes based on clusters with Euclidean distance. The modeling was synthesized in a computer tool where the inputs were the characteristics of the soil and the time and the outputs were the loss of thickness of each predictive model and the description of the most similar real tests. The results showed that in all models, for relative errors of 10%, over 90% of predictions were correct. In addition, a real example of the operation of the tool was defined, where it was found that the estimates of the models allow the necessary optimization of steel to fulfill its useful life.

Keywords: steel corrosion; computer tool; predictive models; quantitative techniques; optimization; sustainable engineering

1. Introduction

The greatest enemy of metal structures buried in the ground is corrosion, which is a mechanism of degradation of materials produced by an electrochemical reaction with the soil in which they are immersed [1].

The underground environment is characterized by great corrosive complexity [2,3]. The cause of this instability is the fact that the soil is constantly changing and presents great variability in its local conditions [4,5]. Two geographically close soils may have different characteristics and, therefore, degrade steel with different aggressiveness.

Corrosion is an essential factor in the design phase of metal structures due to its effects on structural resistance [6,7]. If the steel does not satisfy the useful life of the structure, it may create serious risks for human health and the environment. Such corrosion also has an impact on the economic calculation related to the structures and increases the cost of their maintenance [8,9].

The existing literature on the corrosion of steel in soil shows a lack of a clearly established method in the dimensioning of buried metal structures [10]. The engineer responsible for the design of the structure estimates the amount of steel required for each project. In general, the design applies excess thickness to the material to avoid the failure of the structure.

On many occasions, engineering projects require the use of metallic structures partially or totally buried in the soil. An instance of this type of structure is the elements that constitute solar plants, which are introduced into the ground without applying cathodic protections, which, in these cases, are expensive and not very efficient. The structures are cut and welded in situ. Nonmetallic coatings in these applications are reserved only for extremely corrosive soils, for example, those containing gypsum. The models developed in this research are of special interest for these types of installations. To achieve the most adequate compensation between the quantity of steel and safety, it is necessary to adjust the tolerances and achieve the optimum thickness in the design that guarantees the useful life of the steel. In this way, safer and more sustainable structures may be defined in terms of efficiency and demand for materials.

The main objective of this paper was the prediction of corrosion in buried metal structures for a specified period of useful life. To perform this, data mining techniques were applied to obtain predictive models of corrosion in steel in the underground environment. These are multivariable models that, based on the characteristics of the soil, calculate the necessary thicknesses of metal to guarantee a specific useful life period. The parameters used to feed the modeling are physical, electrochemical, soil grain size, and climatological variables of the terrain.

Once the models were developed, they were materialized in a software tool where the characteristics of the soil and the agreed operating time are entered. The engineer responsible for the design of buried metal structures obtains a quantitative value of the behavior of steel in the soil. The output variable is the loss of thickness due to corrosion that the steel will have according to its useful life.

The paper starts with a deep analysis of the state of the art and the definition of the applied methodology. Then, it explains the creation of the database through the analysis, processing, and preparation of the available information. Next, modeling techniques are defined, followed by a description of how models were developed and materialized in a simple tool. Finally, the results are discussed, and the conclusions obtained in the research are detailed.

2. Literature Review

Since the introduction of metals in construction, the existing related literature shows a low number of scientific projects aimed at analyzing corrosion in the underground environment. The reason lies in the variability of soil conditions, which makes it difficult to study [11].

Soil is highly complex due to its local properties and the influence of external factors on its characteristics. For instance, rain or artificial manipulation can significantly change the behavior of soil on steel [12]. Therefore, all the variables involved must be analyzed to adequately determine the corrosion that the material will suffer.

The methods used to predict corrosion in buried metal structures may be divided into two main typologies: qualitative and quantitative methods. The former are usually tables of scores that provide guidance on the possible effects of corrosion but do not permit establishing a value to aid in sizing. The American Water Works Association (AWWA) developed one of the most widespread qualitative methods described in the AWWA C-105 Standard, which consists of assigning points to different soil variables involved in corrosion [13]. With the sum of all the factors, the method informs about if the corrosion is going to be light, moderate, appreciable, or severe, or if it requires protective measures. In contrast, quantitative methods establish a specific measure or recommended numerical values of metal loss in the soil due to this degradation mechanism. In this field, the most relevant method is the Romanoff tables. This study stands out for its duration and for the number of soils studied under real conditions [10].

In recent years, a new trend has emerged in the field of corrosion, which consists of the generation of predictive models with computer tools. Most of them focus on the period in which existing metallic structures were built and buried [14–16]. With these methods, it is possible to estimate the corrosion from measurements in situ or simulated in the laboratory, being very useful for establishing inspection

and maintenance policies. The drawback is that they do not help in the design of the structures because they do not predict the loss of thickness that the material will have; they only evaluate its condition.

Other models aim to predict the corrosion that the structure will suffer in an underground environment. For these models, statistical approaches are widely used [17]. For instance, Alamilla led the creation of a mathematical model to simulate the spread of corrosion damage over time [18]. They generated a unique model that provides a series of graphs that relate time to material thickness loss using interpolation equations. In this study, four parameters were considered: pH, resistivity, redox potential, and electric potential. In 2015, a project based on the Markov hidden random field theory proposed an inspection strategy to evaluate the corrosion of a buried pipeline [19]. The biggest drawback of these studies is the difficulty of handling and interpreting the graphs where the results are presented.

Other studies on corrosion assessment and modeling have used very small databases. An example is the paper based on three Qatar oil pipelines or a multivariate statistical model fed with information provided by a single pipeline [20,21]. In these cases, the biggest problem was that its generalizability to other locations is very limited.

Other examples create their databases from simulated conditions in a laboratory [22] or feed the model by combining real tests with simulated results [23]. The fundamental limitation of this research is that extrapolating the conditions of a laboratory to a real environment causes errors due to the great complexity of the soil.

It is complex to obtain a representative predictive model that relates the characteristics of the soil with the loss of material from the steel buried and that is easy and fast to manage. In this paper, a database fed by tests carried out in real environments has been generated. It is made up of quantitative information on all the variables that influence soil corrosion. Therefore, it is a multivariate approach that considers the damage caused to steel by corrosive parameters and the interaction between them.

The database generates four different models. On the one hand, it generates three predictive models: basic, physical, and electrochemical. Depending on the soil where the structure is going to be designed, the model informs which is the most appropriate estimate. On the other hand, it generates a distance or clustering model that provides the most similar real cases existing in the database.

The benefit of having more than one predictive model is that it allows the estimation of corrosion to be adapted to the characteristics of the soil. In addition, the clustering model details all the parameters of the soils most similar to the one under study. Thus, the user has complete information on these samples and their real behavior due to corrosion over time.

In addition, the models were synthesized in a computer tool. The input variables are the parameters of the soil where the structure is going to be designed and its useful life. It automatically provides the material loss that the steel will experience in each model and the thickness losses experienced by the same material buried in similar conditions.

The engineer responsible for the project does not have to handle graphs or tables of values. The output variables are numerical values of the thickness that the buried steel will lose during the useful life of the structure.

3. Materials and Methods

To predict the corrosion behavior of steel, a robust and representative database was first constructed. Subsequently, the models and the computer tool were developed and evaluated.

3.1. Data Collection

The severity of the corrosion is a function of the combination of a large number of parameters. The study of its effects considering the individual behavior of a single factor may cause very serious errors in the design of the structures and in the guarantees of their useful life [24]. For this reason, the present research has developed multivariate models that consider the characteristics of the soil that most influence the process of degradation of metals [25].

In the literature on buried steel structures, the work done by scientist M. Romanoff is considered the most relevant and extensive source of information [10]. It is a quantitative and multivariate method that has not yet been replaced by any other methodology. The reason is that it is a real experiment that lasted 20 years and consisted of burying a large number of metal samples in different types of soil [26]. All the results were compiled in numerous tables where the characteristics of the soils and materials were noted.

From the data collected from the Romanoff tables, the information has been analyzed and prepared to generate a reliable and meaningful database of the models.

3.1.1. Data Analysis

The metal used in this study was Bessemer steel, a reference steel in the field of construction of this type of buried structure. This material presents an approximate composition of 0.09% carbon, 0.39% manganese, 0.08% phosphorus, and 0.04% sulfur, and its density is greater than that attributed to iron, reaching 7850 kg/m³.

The database of the predictive models developed has been configured with the results of the most complete scientific publication on steels published by Romanoff [27], and his work has been complemented that of I.A. Denison [28]. The combination of the information provided by both studies established a data set on the corrosion of Bessemer steel in 62 different soils. The locations of the samples included different places in the United States.

The tables from these tests are the result of periodically unearthing samples over 15 years to study the evolution of corrosion in steel. They define practically all the characteristics of the soil that may influence corrosion. The complete list of the physical, electrochemical, granulometric, and climatological variables of the soil for which there is available information is presented in Table 1. It indicates, for each parameter, the original unit used in the Romanoff studies and its equivalent in the international system.

Table 1. Description of the variables studied by Romanoff.

Variable	Original Unit	Unit S.I.
Site number	-	-
Floor	-	-
Location	-	-
Internal drain	-	-
Burial depth	ft	m
Content in sand	%	%
Slime content	%	%
Clay content	%	%
Colloids	%	%
Suspension	%	%
Resistivity	ohm-cm	ohm-m
Temperature	°F	°C
Annual rainfall	in/yr	mm/yr
Equivalent humidity	%	%
Porosity	%	%
Density (specific gravity)	g/cm ³	kg/m ³
Change of volume by retraction	%	%
pH	-log[mol/L]	-log[mol/L]
Total acidity	mg-eq/100 g	mol/kg
[Na + K]	mg-eq/100 g	mol/kg
[AC]	mg-eq/100 g	mol/kg
[Mg]	mg-eq/100 g	mol/kg
[CO ₃]	mg-eq/100 g	mol/kg
[HCO ₃]	mg-eq/100 g	mol/kg
[Cl]	mg-eq/100 g	mol/kg
[SO ₄]	mg-eq/100 g	mol/kg
Mass loss due to corrosion	oz/ft ²	g/m ²

3.1.2. Data Preparation

Despite the relevance of Romanoff's research in the field of corrosion, it presents a series of limitations, mainly due to its 1957 publication. In 2007, the organization "National Institute of Standards and Technology" (NIST) published a comprehensive analysis of Romanoff's work and revealed a number of shortcomings in the performance and documentation of the original trials [29]. However, the study concluded that Romanoff's results could be used for the development of predictive models if their restrictions are always considered.

The deficiencies that the NIST described in 2007 and that had to be resolved to create the database of the models are the following:

- Absence of data in practically all the variables involved. Table 2 defines the empty values for the main study variables. As can be seen, they are a big problem in the configuration of the model database, especially in some parameters, such as those related to chemical composition.
- Identification of unusual values. These are the data that remain outside the limits presented by the variable in most cases, which we have summarized in Table 3.

Table 2. Description of the variables studied by Romanoff.

Variable	Empty Values
Burial depth	15
Content in sand	28
Slime content	28
Clay content	28
Colloids	29
Suspension	28
Resistivity	0
Temperature	0
Annual rainfall	0
Equivalent humidity	0
Porosity	17
Density (specific gravity)	25
Change of volume by retraction	16
pH	0
Total acidity	6
[Na + K]	24
[AC]	24
[Mg]	25
[CO ₃]	24
[HCO ₃]	25
[Cl]	23
[SO ₄]	21

These outliers may originate from errors in measurement or data entry, or they may not be incorrect data. For instance, Table 3 shows the maximum values of resistivity and total acidity well above the average. In this case, the NIST analysis ruled that these outliers are due to real measurements in soils with special characteristics, and therefore are not trial failures. However, the analysis of the loss of mass parameter of the steel detected corrosion values lower than the measurements made in previous periods. Obviously, the loss of mass due to corrosion must always be equal or greater than the measurement at a previous time. Therefore, the data corresponding to these situations are considered unusual values and would affect the reliability of the models.

The solutions that have been established to overcome the two limitations found in the available information are as follows:

- No empty values have been filled that would falsify the starting information.

- Incomplete patterns have been maintained whenever possible.
- Unusual values have been studied to determine their origin and have only been eliminated when they have been checked to be errors. In the specific case of the mass loss variable, the anomalous values have been eliminated and replaced by a logarithmic function, as advised by NIST. The reason is that the corrosion of the steel initially grows quite fast and gradually stabilizes, which conforms to a logarithmic function (1).

$$y = a \ln x + b. \quad (1)$$

Table 3. Description of the variables studied by Romanoff.

Variable	Minimum Value	Average Value	Maximum Value
Burial depth	0.56	0.9	2.41
Content in sand	1.2	20.3	97.3
Slime content	2.1	37.8	78.4
Clay content	0.6	33.3	87.8
Colloids	2.7	26.1	71.5
Suspension	1.8	22.2	45.5
Resistivity	0.6	17.2	451
Temperature	3.9	12.9	20.9
Annual rainfall	142	941	1565
Equivalent humidity	2.8	28.4	72.8
Porosity	1.1	10.9	40.6
Density (specific gravity)	1260	1720	2080
Change of volume by retraction	0	4.3	39.8
pH	2.6	6.6	9.5
Total acidity	0	0.071	2.974
[Na + K]	0.0023	0.0107	0.45
[AC]	0	0.007	0.1555
[Mg]	0.0012	0.0043	0.095
[CO ₃]	0	0	0.0002
[HCO ₃]	0	0.0071	0.02
[Cl]	0	0.0009	0.43
[SO ₄]	0	0.0083	0.567

3.2. Modelling Process

This research performed two types of modeling to study the corrosion: a distance model and three predictive models.

The reason for building a distance model was to make it easier for the engineer to search the existing literature for the real case most similar to the one being studied. As already explained in this paper, the largest real experiment on buried steel structures is that of Romanoff, but his tables are complicated to work with. Thanks to this model, engineers may enter the soil parameters and obtain the descriptions of the soils most similar to theirs and what their corrosive behavior has been over time.

Furthermore, the distance calculation has also been used to estimate the quality of the predictions of all the models. It consists of evaluating how close the case studied is to the multidimensional space of some point by which the real evolution of corrosion is known.

It was decided to develop three predictive models instead of just one to achieve a better adaptation to the soil characteristics and the information available. Depending on the input variables, it will be more convenient to use the prediction of one or the other. All models consider a burial depth of approximately 2 m, which is the depth that is usually used in this type of buried metal structures.

3.2.1. Distance Model

The technique used to group the most similar cases is based on the concept of distance between points in multidimensional spaces. It is a nonlinear mathematical model that allows predicting metal

corrosion from interpolation, seeking the best solutions in the areas of the space between the known points or data.

The Euclidean distance between 2 points is the length of the segment that joins them [30]. In a two-dimensional space between any two points, P_1 and P_2 , of Cartesian coordinates (x_1, y_1) and (x_2, y_2) , respectively, it corresponds to the following (2):

$$d_E = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \quad (2)$$

When extrapolating the two-dimensional solution to an n-dimensional Euclidean space, the Euclidean distance between the points $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$ is defined as follows (3) [31]:

$$d_E = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}. \quad (3)$$

This technique has presented difficulties due to the quality of the database: the Euclidean distance may be affected by the different units in which the input variables have been defined. This problem has been solved by normalizing between 0 and 1 all the parameters used in the model. This ensures that all input variables have the same weight in calculating the distance between two points.

Another relevant aspect in the generation of the distance model is that the database has presented empty values for which information is not available. Therefore, the model will sometimes have to compare spaces with different dimensions. Consequently, a greater number of dimensions causes the points to be at a greater maximum distance. To solve this problem, the distances have been normalized, dividing the Euclidean distance by the maximum distance in space, which corresponds to the square root of the number of dimensions.

3.2.2. Predictive Model

Process modeling may be considered as a functional approximation problem, which is a matter of finding a mathematical expression that can reproduce the injective relationship that exists between the inputs and outputs of the process.

In this paper, the multivariate adaptive regression spline (MARS) algorithm has been used to approximate an unknown function by projecting a sample of data from the input space onto the output space.

MARS algorithm is a calculation process that was developed in 1991 by Jerome Friedman and is characterized by high estimation efficiency [32]. In 1993, De Veaux led a comparative analysis between the predictive capacity of the MARS algorithm and the neural networks, and the result showed that in numerous cases, the one presented by the algorithm was superior to that of the neurons [33].

MARS is an algorithm widely described in other sources, and its creator Friedman defined it as a flexible modeling method for high-dimensional data [34]. It consists of a flexible, nonparametric regression technique that extends the linear model by incorporating nonlinearities and interactions of variables [35].

Furthermore, the MARS algorithm is robust against collinearities of the input variables and has the ability to automatically select the most relevant factors as well as their interactions [36]. Therefore, in this work, MARS has selected the input variables and has analyzed the most important parameters to predict the corrosion that will occur in the buried steel.

The objective of the algorithm is to approximate an unknown function f that represents the relationship between the soil variables that influence corrosion and the loss of thickness that will occur in the steel. To do this, MARS searches for the approximation function f' by creating subregions (nodes) and carrying out linear combinations of sets of base functions B_i parameterized by the position of said nodes (4).

$$f' = \sum_{i=1}^M a_i B_i. \quad (4)$$

here, a_i is the coefficient of the base function B_i and M is the number of base functions of the model. The base functions involved in the construction of the function are the interactions of the variables. Two parameters intervene in the fit of the model: the maximum number of base functions and the maximum degree of interaction. This approach to the solution through decomposition into base functions facilitates the interpretability of the results to build a continuous function.

3.3. Generation of the Models

The MARS algorithm performed an exhaustive search to determine the relevant parameters of the process through adaptive decomposition. In this work, we used a modified version of MARS algorithm developed by the research group, named API-MARS [36]. The best results were obtained considering 35 basis functions and 3 as the maximum degree of interactions. Below, the input variables of each of the predictive models are described.

The basic model was performed for the five most influential factors in corrosion and the least empty values in the database. The following table qualifies the input variables of the model and their measurement units (Table 4).

Table 4. Input variables for the basic model.

Input Variable	Unit
Resistivity	Ohm·m
Temperature	°C
Precipitation	mm/year
Soil moisture	%
pH	-
Time	years

These parameters have practically no empty values and allow the use of information from 61 different soils. As a result, this is the most representative model of the Romanoff data and, therefore, the most stable.

The physical model includes the parameters of the basic model and adds variables related to the granulometric composition of the soil: clay and silt content. The reason for this choice is that in order to know the size of the particles that make up the terrain, it is sufficient to define two of the three components of the ternary diagram, since the third one depends on the other two. The complete list of input variables is described in Table 5:

Table 5. Input variables for the physical model.

Input Variable	Unit
Resistivity	Ohm·m
Temperature	°C
Precipitation	mm/year
Soil moisture	%
pH	-
Slime content	%
Clay content	%
Time	years

The inconvenience is that the information on the texture of the terrain in the tests carried out by NIST is only detailed in 34 different soils. Therefore, this model is less representative than the basic model.

Finally, the electrochemical model has considered the variables of the basic model and includes the electrochemical variables that most affect corrosion: the presence of chlorides and sulfates. Table 6 shows all the input parameters of the model and their units of measurement.

Table 6. Input variables for the electrochemical model.

Input Variable	Unit
Resistivity	Ohm·m
Temperature	°C
Precipitation	mm/year
Soil moisture	%
pH	-
Chlorine (Cl)	mol/kg
Sulfates (SO ₄)	-
Time	years

The composition of sulfates and chlorine only is available in 38 soils of the prepared database, which makes this predictive model the least representative of the three created in this paper.

The distance model needs at least two input variables and at most 14 of the complete list defined above in Table 2 and has two main objectives. On the one hand, it must report on the corrosion loss of mass that the real test has experienced with the characteristics most similar to the case study. On the other hand, it is an evaluation tool for the three predictive models, since the results obtained in the estimation may not be too far from the real similar cases.

To evaluate the quality of the prediction of the models, it is not possible to have the corrosion that structures are going to suffer in a soil over time. Therefore, the only way to assess the precision of the developed models is by comparing the predictions with the actual values in the data set. To achieve this objective, the MARS algorithm was used. This calculation process has configured a set of training patterns and test patterns based on the total existing data. In this way, the true value of corrosion in the test patterns has been compared with the value estimated by the models in the training set.

Once the patterns were constituted, the distance model was used to assess how close the case studied is to the multidimensional space of some point by which the real evolution of corrosion is known. A value close to 100% in the quality index indicates that the prediction has very high reliability since the case studied is very close in multidimensional space to some point for which the real evolution of corrosion is known. On the contrary, when the quality index takes values lower than 80%, the prediction given by the model must always be considered insufficient to produce with a sufficient level of confidence.

The purpose of this study was to predict the thickness loss that buried steel structures will experience in order to help engineers design them optimally and sustainably. For this reason, the results of the models were synthesized in a computer tool.

To facilitate the use of the tool in the design of buried steel structures, it was decided that the output variable of the program would be the loss of thickness due to corrosion, expressed in microns. This conversion was carried out taking into account the densities of the study materials, namely 7850 kg/m³ for steel and 7140 kg/m³ for zinc.

4. Results and Discussion

The precision of the results presented was studied using the concepts of absolute and relative error. The first is the difference between the actual and the estimated value, and its unit is g/m². As for the relative error, it is the quotient between the absolute error and the real value multiplied by 100. It is important to note that the relative error indicates the quality of the prediction, since the lower its percentage, the more reliable the estimated measure will be [37]. In addition, the mean square error of each model is identified to know the average of the squared errors and to know the difference between the estimator and what is estimated [38].

Table 7 shows a description of the results obtained during the training stages of the basic model. It was observed that for a relative error level of 5% over the corrosion range, the system obtained 64.5%

correct answers and the root-mean-square deviation was 460.9 g/cm². Therefore, the corrosion value in buried steel estimated by the basic model showed good behavior.

Table 7. Results of the quality system of the basic model.

Absolute Error (g/m ²)	Relative Error (%)	Model Success (%)
84	1	17.9
421	5	64.5
842	10	93.1
1179	14	98.8
1684	20	99.7
2106	25	99.7

In the physical model (Table 8), the absolute error reached lower values than in the basic model. This means that the difference between the actual and the estimated value was smaller. However, for a relative error of 5%, the physical model has presented a lower percentage of success. The reason is that the most corrosive soils in the database do not have values for their granulometric composition, so they may not be analyzed by this model. The root-mean-square deviation was 220.4; therefore, it was substantially less than the basic model. In short, the model is more reliable than the basic one, but due to the low number of cases used in training, its functionality is less representative.

Table 8. Results of the quality system of the physical model.

Absolute Error (g/m ²)	Relative Error (%)	Model Success (%)
29	1	13.2
146	5	59.9
293	10	79.2
410	14	92.4
586	20	98.5
732	25	100

The same happened in the electrochemical model, for which substantially fewer data are used than in the basic model. Only complete electrochemical analysis of 38 soils is available in the database.

The results of the electrochemical model are presented in Table 9, showing that in 76% of cases the errors were less than 420 g/m². Again, it was lower than for the basic model but higher than for the physical model. This is the model with the greatest success, but the fact that it was developed with less information indicates that its generalizability is limited.

Table 9. Results of the quality system of the electrochemical model.

Absolute Error (g/m ²)	Relative Error (%)	Model Success (%)
84	1	23.8
420	5	76.7
839	10	96.1
1175	14	99
1678	20	100
2098	25	100

To highlight the use of the models and their contribution to the research community, we assumed that a steel structure is intended to be designed on a soil with the characteristics described in Table 10 for a useful life of 15 years. Next, the corrosion behavior of steel was studied as if the research described in this research had not been developed, and then, the loss of mass due to corrosion was determined through the computer tool created.

Table 10. Input variables of the study soil.

Variable	Values
Contend in sand (%)	50
Slime content (%)	20
Resistivity (ohm-m)	50
Temperature (°C)	12
Annual rainfall (mm/year)	1000
Humidity (%)	22
pH	6

In the case of not having the tool developed in this investigation, the usual procedure would be to go to the tables prepared by NIST and to try to find the apparently most similar cases. Therefore, according to Romanoff's work, the three soils that we have considered to be most similar to the one proposed are those shown in Table 11 (Soil 14, Soil 16, and Soil 36).

Table 11. Soils proposed by Romanoff tables.

Variable	Study Soil	Soil 14	Soil 16	Soil 36
Sand content (%)	50	56.6	50.4	60.6
Slime content (%)	20	29.5	23.1	21.8
Resistivity (ohm-m)	50	35.2	82.9	112
Temperature (°C)	12	6.8	19.6	17.8
Annual rainfall (mm/year)	1000	691	1565	1346
Humidity (%)	22	17.2	22.2	13.8
pH	6	6.2	4.4	4.5
Steel corrosion (gr/m ²)		1465	2350	702

Table 11 presents large differences between the three material losses due to corrosion that steel experienced in each of the three soils. Choosing one of the three selected soils leads to differences of more than 300% in the estimation of corrosion. Consequently, the engineer in charge of the design of the structure should opt for the proposal of one of them based on their experience and knowledge.

The new tool developed allows one to resolve the subjectivity in the decision since the results obtained are independent of who the user of the tool is. In the case of the soil described in Table 10, the tool provides the following results (Table 12).

Table 12. Prediction of the computer tool.

Models	Steel Corrosion (Microns)	Quality (%)
Clustering		
- Soil 34	175.52	85
- Soil 9	190.83	78
- Soil 55	165.44	75.8
Basic	158	85
Physical	141	57
Electrochemical	-	

With the available information, the electrochemical model could not be executed, but the basic, physical, and distance models could be executed. In addition, the system only provides the loss of mass due to corrosion, which facilitates design work.

Notably, the distance model determined that none of the cases that had been chosen in the tables are the closest to the one studied. Given the quality percentage of the prediction, the most similar soil is the one defined by soil 34. Therefore, the engineer could directly use the corrosion information in that

soil for the design of the buried steel structure. Furthermore, if the steel thickness is to be optimized with the useful life of the structure, the solution proposed by the basic model should be used. In this way, the consumption of materials would be reduced, improving the economic cost and sustainability of the infrastructure.

In any case, the use of the tool allows decisions to be made in a more objective way, regardless of the person in charge of the design. In addition, it provides a lot of information automatically when entering the input variables and in a manageable format.

5. Conclusions

In this paper, we developed a clustering model and three predictive models to estimate the thickness loss that steel experiences when it is buried in the ground. The construction of a reliable and representative database that feeds the modeling was carried out through the analysis and preparation of Romanoff tables.

The clustering model is based on the concept of Euclidean distance and allows us to know the loss of corrosion that has occurred in the steel in the most similar soils in the database. The basic, physical, and electrochemical predictive models were performed by using the MARS algorithm and estimating the loss of material based on the characteristics of the soil where the structure will be buried. In addition, the quality of the predictions was evaluated to discuss the effectiveness of the estimates.

The results obtained were materialized in a computer application that provides the loss of thickness that the steel will suffer according to a certain period. In this way, it provides the evolution of corrosion over time and guarantees the agreed useful life of the structure. In addition to providing the loss of metal thickness, the tool describes the parameters of the real cases that the distance model detects as being more similar. In addition, below each estimate, the percentage of the prediction's quality is indicated.

The results show that this project may contribute to optimizing the design of buried steel structures. It is able to estimate the loss of thickness that the structure will suffer due to corrosion and avoid using conservative and subjective excess thickness.

The models adjust tolerances and optimize the amount of metal needed to meet its useful life. This allows for the efficient and sustainable use of materials and a reduction in the total cost of the structure.

Although this research was done for Bessemer steel, both the modeling methodology and techniques could be applied to other types of materials, such as galvanized steel. The only modification in the development would be the database, since the information related to each type of metal would have to be added. Future research could make a comparison of how corrosion in the soil affects each of them. In this way, the responsible engineer would choose the most efficient type of steel for each soil and structure.

Author Contributions: Conceptualization, V.R.-M. and F.O.-F.; formal analysis, L.-d.A.-R.; funding acquisition, F.O.-F.; investigation, L.-d.A.-R.; methodology, V.R.-M.; software, J.V.-B.; supervision, F.O.-F.; validation, J.V.-B.; writing—original draft, L.-d.A.-R.; writing—review and editing, V.R.-M. and F.O.-F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was subsidized through the Plan of Science, Technology, and Innovation of the Principality of Asturias (Ref: FC-GRUPIN-IDI/2018/000225).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of; in the writing of the manuscript; or in the decision to publish the results.

References

1. Schmitt, G. Global needs for knowledge dissemination, research, and development in materials deterioration and corrosion control. *World Corros. Organ.* **2009**, *38*, 14.

2. Parker, M.; Peattie, E.G. *Pipeline Corrosion and Cathodic Protection: A Practical Manual for Corrosion Engineers, Technicians, and Field Personnel*; Gulf Professional Publishing: Houston, TX, USA, 1984.
3. Hudson, J.C.; Banfield, T.A.; Holden, H.A. Tests on the corrosion of buried ferrous metals. *J. Iron Steel Inst.* **1942**, *146*, 107.
4. Logan, K.H.; Brush, W.W. The bureau of standards soil-corrosion investigation [with discussion]. *J. Am. Water Works Assoc.* **1929**, *21*, 311–317. [[CrossRef](#)]
5. Sivakumar, S.; Ramaswamy, J. Corrosion of Mild Steel Buried Underground for 3 Years in Different Soils of Varying Textures. *J. Mater. Eng. Perform.* **2019**, *28*, 863–875. [[CrossRef](#)]
6. Hembara, O.V.; Andreikiv, O.E. Effect of hydrogenation of the walls of oil-and-gas pipelines on their soil corrosion and service life. *Mater. Sci.* **2012**, *47*, 598–607. [[CrossRef](#)]
7. Wasim, M.; Li, C.-Q.; Robert, D.; Mahmoodian, M. Effect of Soil's Acidity and Saturation on Degradation of Fracture Toughness of Buried Cast Iron. *J. Mater. Civil Eng.* **2020**, *32*, 04020180. [[CrossRef](#)]
8. Schweitzer, P.A. *Fundamentals of Metallic Corrosion: Atmospheric and Media Corrosion of Metals*; CRC press: Boca Raton, FL, USA, 2006.
9. De Baere, K.; Verstraelen, H.; Rigo, P.; Van Passel, S.; Lenaerts, S.; Potters, G. Study on alternative approaches to corrosion protection of ballast tanks using an economic model. *Mar. Struct.* **2013**, *32*, 1–17. [[CrossRef](#)]
10. Arriba-Rodriguez, L.; Villanueva-Balsera, J.; Ortega-Fernandez, F.; Rodriguez-Perez, F. Methods to Evaluate Corrosion in Buried Steel Structures: A Review. *Metals* **2018**, *8*, 334. [[CrossRef](#)]
11. Wu, Y.; Liu, T.; Luo, S.; Sun, C. Corrosion characteristics of Q235 steel in simulated Yingtan soil solutions. Korrosionsverhalten von Q235 Stahl in simulierter Bodenlösung. *Mater. Und Werkst.* **2010**, *41*, 142–146. [[CrossRef](#)]
12. Akkouche, R.; Remazeilles, C.; Jeannin, M.; Barbalat, M.; Sabot, R.; Refait, P. Influence of soil moisture on the corrosion processes of carbon steel in artificial soil: Active area and differential aeration cells. *Electrochim. Acta* **2016**, *213*, 698–708. [[CrossRef](#)]
13. Najjaran, H.; Sadiq, R.; Rajani, B. Modeling Pipe Deterioration using Soil Properties—An Application of Fuzzy Logic Expert System. *Pipeline Eng. Constr.* **2004**, 1–10. [[CrossRef](#)]
14. Barbalat, M.; Caron, D.; Lanarde, L.; Meyer, M.; Fontaine, S.; Castillon, F.; Vittonato, J.; Refait, P. Estimation of residual corrosion rates of steel under cathodic protection in soils via voltammetry. *Corros. Sci.* **2013**, *73*, 222–229. [[CrossRef](#)]
15. Papavinasam, S. *Corrosion Control in the Oil and Gas Industry*; Elsevier: Amsterdam, The Netherlands, 2013; ISBN 978-0-12-397306-1.
16. Allahkaram, S.R.; Isakhani-Zakaria, M.; Derakhshani, M.; Samadian, M.; Sharifi-Rasaey, H.; Razmjoo, A. Investigation on corrosion rate and a novel corrosion criterion for gas pipelines affected by dynamic stray current. *J. Nat. Gas Sci. Eng.* **2015**, *26*, 453–460. [[CrossRef](#)]
17. Velázquez, J.C.; Caleyó, F.; Valor, A.; Hallen, J.M. Predictive Model for Pitting Corrosion in Buried Oil and Gas Pipelines. *Corrosion* **2009**, *65*, 332–342. [[CrossRef](#)]
18. Alamilla, J.L.; Espinosa-Medina, M.A.; Sosa, E. Modelling steel corrosion damage in soil environment. *Corros. Sci.* **2009**, *51*, 2628–2638. [[CrossRef](#)]
19. Wang, H.; Yajima, A.; Liang, R.Y.; Castaneda, H. A clustering approach for assessing external corrosion in a buried pipeline based on hidden Markov random field model. *Struct. Saf.* **2015**, *56*, 18–29. [[CrossRef](#)]
20. Wang, X.; Wang, H.; Tang, F.; Castaneda, H.; Liang, R. Statistical analysis of spatial distribution of external corrosion defects in buried pipelines using a multivariate Poisson-lognormal model. *Struct. Infrastruct. Eng.* **2020**, 1–16. [[CrossRef](#)]
21. El-Abbasy, M.S.; Senouci, A.; Zayed, T.; Mirahadi, F.; Parvizsedghy, L. Artificial neural network models for predicting condition of offshore oil and gas pipelines. *Autom. Constr.* **2014**, *45*, 50–65. [[CrossRef](#)]
22. He, B.; Han, P.; Hou, L.; Zhang, D.; Bai, X.-H. Understanding the effect of soil particle size on corrosion behavior of natural gas pipeline via modelling and corrosion micromorphology. *Eng. Fail. Anal.* **2017**, *80*, 325–340. [[CrossRef](#)]
23. Gadala, I.M.; Abdel Wahab, M.; Alfantazi, A. Numerical simulations of soil physicochemistry and aeration influences on the external corrosion and cathodic protection design of buried pipeline steels. *Mater. Des.* **2016**, *97*, 287–299. [[CrossRef](#)]
24. Lins, V.D.F.C.; Ferreira, M.L.M.; Saliba, P.A. Corrosion resistance of API X52 carbon steel in soil environment. *J. Mater. Res. Technol.* **2012**, *1*, 161–166. [[CrossRef](#)]

25. Schaschl, E.; GA, M. Some new views on soil corrosion. *Corros. Eng. Dig.* **1964**, *13*, 168–173. [[CrossRef](#)]
26. Cadle, M. NACE International Cathodic Protection Task Group Update. Available online: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.708.6329&rep=rep1&type=pdf> (accessed on 1 October 2020).
27. Romanoff, M. *Underground Corrosion, National Bureau of Standards Circular 579*; US Government Printing Office: Washington, DC, USA, 1957; Volume 25.
28. Denison, I.A.; Romanoff, M. Corrosion of galvanized steel in soils. *J. Res. Natl. Bur. Stand.* **1952**, *49*, 299–316. [[CrossRef](#)]
29. Ricker, R.E. Analysis of pipeline steel corrosion data from NBS (NIST) studies conducted between 1922–1940 and relevance to pipeline management. *J. Res. Natl. Inst. Stand. Technol.* **2010**, *115*, 373. [[CrossRef](#)] [[PubMed](#)]
30. Treves, F. *Topological Vector Spaces, Distributions and Kernels: Pure and Applied Mathematics*; Elsevier: Amsterdam, The Netherlands, 2016; Volume 25, ISBN 978-1-4832-2362-9.
31. Bourbaki, N. *Topological Vector Spaces: Chapters 1–5*; Springer Science & Business Media: Berlin, Germany, 2013.
32. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [[CrossRef](#)]
33. De Veaux, R.D.; Psychogios, D.C.; Ungar, L.H. A comparison of two nonparametric estimation schemes: MARS and neural networks. *Comput. Chem. Eng.* **1993**, *17*, 819–837. [[CrossRef](#)]
34. Friedman, J.H.; Roosen, C.B. An introduction to multivariate adaptive regression splines. *Stat Methods Med. Res.* **1995**, *4*, 197–217. [[CrossRef](#)]
35. Vicente, R.M.; Fernando, R.P.; Francisco, O.F.; Joaquín, V.B. Modelling of lateral flow in a Hot Strip Mill (HSM) using adaptive techniques. In Proceedings of the 2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 12–14 December 2015; pp. 44–49.
36. Montequín, M.T.R. Modelado Evolutivo Mediante Técnicas Adaptativas Aplicado al Control de Inclusiones en Bandas Laminadas en Caliente (Evolutionary Modeling Using Adaptive Techniques Applied to the Control of Inclusions in Hot Rolled Strips). Ph.D. Thesis, Universidad de Oviedo, Oviedo, Spain, 2003.
37. Wackerly, D.; Mendenhall, W.; Scheaffer, R.L. *Mathematical Statistics with Applications*; Cengage Learning: Boston, MA, USA, 2014.
38. Lehmann, E.L.; Casella, G. *Theory of Point Estimation*; Springer Science & Business Media: Berlin, Germany, 2006.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).