

## Research Article

# Corrosion Predictive Model in Hot-Dip Galvanized Steel Buried in Soil

**Lorena-De Arriba-Rodríguez, Francisco Ortega-Fernández, Joaquín M. Villanueva-Balsera , and Vicente Rodríguez-Montequín **

*Project Engineering Area, University of Oviedo, C/Independencia 13, 33004 Oviedo, Spain*

Correspondence should be addressed to Vicente Rodríguez-Montequín; [montequi@uniovi.es](mailto:montequi@uniovi.es)

Received 23 April 2021; Revised 5 September 2021; Accepted 16 September 2021; Published 29 September 2021

Academic Editor: Atila Bueno

Copyright © 2021 Lorena-De Arriba-Rodríguez et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Corrosion is one of the main concerns in the field of structural engineering due to its effect on steel buried in soil. Currently, there is no clearly established method that allows its calculation with precision and ensures the durability of this type of structures. Qualitative methods are commonly used rather than quantitative methods. The objective of this research is the development of a multivariate quantitative predictive model for estimating the loss of thickness that will occur in buried hot-dip galvanized steel as a function of time. The technique used in the modelling is the Adaptive Regression of Multivariate Splines (MARS). The main drawback of this kind of studies is the lack of data since it is not possible to have a priori the corrosive behaviour that the buried material will have as a function of time. To solve this issue, a solid and reliable database was built from the analysis and treatment of the existing literature and with the results obtained from a predictive model to estimate the thickness loss of ungalvanized steel. The input variables of the model are 5 characteristics of the soil, the useful life of the structure, and the loss of corroded ungalvanized steel in the soil. This last data is the output variable of another previous predictive model to estimate the loss of thickness of bare steel in a soil. The objective variable of the model is the loss of thickness that hot-dip galvanized steel will experience buried in the ground and expressed in  $\text{g/m}^2$ . To evaluate the performance and applicability of the proposed model, the statistical metrics RMSE,  $R^2$ , MAE, and RAE and the graphs of standardized residuals were used. The results indicated that the model offers a very high prediction performance. Specifically, the mean square error was  $290.6 \text{ g/m}^2$  (range of the objective variable is from  $51.787 \text{ g/m}^2$  to  $5950.5 \text{ g/m}^2$ ),  $R^2$  was 0.96, and from a relative error of 0.14, the success of the estimate was 100%. Therefore, the use of the proposed predictive model optimizes the relationship between the amount of hot-dip galvanized steel and the useful life of the buried metal structure.

## 1. Introduction

Soil is a complex and highly heterogeneous environment whose local characteristics damage buried or semiburied steel structures [1]. The main cause of failure of this type of structure is the corrosion [2, 3]. Corrosion is a mechanism of steel degradation that deteriorates the metal until it ends up causing the failure of the infrastructure [4, 5]. Throughout history, there have been numerous environmental disasters and risks to human health derived from this type of structural failure [6, 7]. In addition to its consequences on the resistance of structures, it also has an economic impact

[8]. The corrosion that occurs in buried steel represents a serious economic and environmental problem [9]. In 2016, the National Association of Corrosion Engineers International revealed that the global cost triggered by corrosion in 2013 represented 3.4 of world Gross Domestic Product [10].

In the design of buried steel structures, it is intended to define the appropriate amount of steel in accordance with the required guarantees that ensure its useful life. Currently, many infrastructure projects are being undertaken that require a buried structure for which a long-term guarantee is required, such as solar plants. These kinds of projects are usually performed under the form of EPC (engineering,

procurement, and construction) contracts. The contractor usually has to offer a long-life guaranty of the facility (e.g., solar plants could be more than 25 years). Under this context, it is not enough following the existing standards for the design of the structure, but they need to determine the thickness of coating to have enough chance of accomplishing the guaranty avoiding the expensive penalties. To achieve the optimal trade-off between safety and cost of buried steel structures, it is vital to provide designers with the optimal tools [11]. The problem is that there is no clearly established method that provides quantitative values for the loss of thickness that galvanized steel will experience over time [12, 13]. The engineer/designer has to study in depth the effect of soil corrosion to size this type of infrastructure and then look at the reference guides published by lobby of the relevant industry sector and select the option that best fits the calculation.

The most accepted methodology for the evaluation of the corrosive nature consists of using qualitative guidelines to predict the aggressiveness of the soil on the metal and using extrathicknesses to ensure the durability and safety of the structure [14]. These methodologies are tables that relate the fundamental variables that intervene in the loss of thickness of the material with points. Depending on the result obtained, they classify the soils according to their aggressiveness into generic groups, without distinguishing the specific characteristics of each place or determining quantitative data that facilitate the design. Based on these guidelines, the responsible engineer must decide the thickness that will guarantee the useful life of the infrastructure and that will not cause its collapse and its serious consequences for humans, the environment, and the economy.

The most widespread qualitative techniques are the standards developed by the American Society for Testing and Materials (ASTM G187-18), the National Association of Corrosion Engineers (NACE), and the American Association of Works of Water (ANSI/AWWA C105) [15–17]. The NACE and ASTM methods are tables that relate soil resistivity with the severity of corrosion. Depending on the values presented by the soil, five corrosive degrees are differentiated: negligible, mild, moderate, severe, and extreme. Their main drawback is that they only consider one parameter. Therefore, its applicability is very limited since the final corrosion is due to the combined effect of several factors [18, 19]. The AWWA standard is the most widely used method to predict the corrosive behaviour of buried galvanized steel [20]. It consists of a point system that qualitatively classifies corrosive behaviour. Each type of soil is obtaining different qualifications depending on the resistivity, the pH, and the redox power. Total amount of these variables is classified into four corrosive categories: mild, moderate, appreciable, and severe. Although it considers more soil characteristics than the NACE and ASTM methods, they are still few compared to the variables involved in corrosion.

Although the most widely used methodologies are those cited above, there are more studies that develop this qualitative approach of classifying the degree of soil corrosion

based on the factors that they consider to be the most influential in corrosion [21–23]. The main advantage of all of them is that they are generally simple formatting and handling tables or diagrams. This simplicity that allows an agile use of the tool is also associated with its main disadvantage since it does not provide a numerical value. Therefore, the engineer who uses this type of technique obtains information on the rate of corrosivity of the soil that will affect the buried metallic structure, but it is the designer who decides the thickness of the galvanized steel. They are methods strongly subject to interpretation and do not help to decide optimal galvanized steel thicknesses based on the lifetime of the structure.

The work developed by Romanoff for the United States National Standards Office (NBS) is one of the most outstanding studies in the field of quantifying the loss of thickness experienced by metals buried in the soil. In 1910, this revolutionary scientist led the most comprehensive and exemplary quantitative study on the corrosion experienced by galvanized steel buried in the soil. The project consisted of burying thousands of galvanized steel strips in soils with different corrosive characteristics and analysing their evolution over approximately 20 years. As a result, in 1957 Romanoff published a series of tables detailing the characteristics of the soils and the evolution of the thickness loss that had occurred in metals [24]. Later, he carried out another series of works that supplemented his initial research [25]. The reason that it is considered the most important study in the field of engineering related to corrosion is due to the duration of the tests, the large number of samples, the number of soils with different characteristics, and that it was an experiment carried out in a real environment. However, the project had the drawback that it focused especially on bare steel and the number of samples dedicated to galvanized steel being much smaller.

Subsequent quantitative studies have been carried out based on burying galvanized steel samples in the soil and periodically digging them up to calculate their mass loss [26, 27]. The main disadvantages that they present in comparison to Romanoff tables are short duration of the test, use of simulated environments in the laboratory, small number of samples, small number of soils, or study of few corrosive factors.

In recent years, a new quantitative approach has been initiated in the field of soil corrosion, which consists of modelling the results obtained in real or laboratory tests of the evolution of corrosion over time. Several methods have been used for modelling, mainly neural networks, genetic algorithms, and fuzzy adaptive systems [28–32]. The problem is that these methods require very large databases, and in most studies, the raw data tend to respond to tests of short duration, with few soils or in simulated environments that do not allow its use to be extrapolated to other locations. In this study, the Multivariate Adaptive Regression Splines (MARS) method has been used. It is a nonparametric regression technique that provides optimal results when there is a limited number of cases.

When comparing quantitative and qualitative methods, it is evident that the first are a more precise tool in the design

of buried steel structures. However, the use of the qualitative approach is much more extended. The main reason is that there is no clearly established quantitative method and those that do exist are complex and require more time to use. In addition, the most important quantitative research led by Romanoff are data tables published in 1952 that describe the effect of corrosion over time of steels and steel alloys that were used at that time in the construction of buried metal structures.

In this sense, the predictive model developed in this research allows to automatically obtain the loss of thickness of the buried material. It is a model fed by a database created from the tests that Romanoff carried out in a real environment and for more than 20 years with galvanized steel. The database includes the physical, chemical, granulometric, and climatological characteristics of the soil that influence the rate of corrosion and the real effect that they caused on the metal. To complete this information, the output variable of another predictive model on the loss of thickness of bare steel in the soil due to corrosion has been introduced as an input variable, for which there is a more extensive database [33]. Once these data have been processed, the duration and number of soils analysed allows the results to be extrapolated to any type of soil.

The material used in the development of this work is a reference steel in construction models: hot galvanized steel. Currently, it is used in partially or fully buried in soil in power transmission and distribution structures, solar projects, or storage of different chemicals [34–36]. Metallic coatings of zinc and their alloys are widely used to protect the base steel against corrosion phenomena that can take place in the steel/environment interaction [37, 38]. In 2003, the International Lead Zinc Research Organization led a project that concluded that zinc coatings reduce corrosion rates, promote uniform attack, and ultimately increase the life of structures in the ground [39]. The hot galvanizing technique consists of immersing a suitably conditioned steel strip in a bath of molten zinc [40]. Thus, a rapid reaction is generated between the molten zinc and the steel, producing an intimate union (intermetallic layer) between both materials. Control of the thickness of the intermetallic layer is done by adding aluminium to the bath.

A fact that Romanoff himself was able to conclude in his research was the existence of similar behaviours in terms of corrosion experienced by galvanized steel in the different groups of soils (reducing, oxidizing, alkaline, acidic, etc.), being the behaviours between groups much more diverse than in the case of bare steel, while within each group they are more homogeneous. For instance, on acid-oxidizing inorganic soils, the galvanized coating remained virtually intact even after 13 years of exposure. On the contrary, acid-reducing organic soils presented a more heterogeneous behaviour in terms of their corrosion rate, although in all cases, the galvanized coating used in the tests (with a thickness of 3 oz/ft<sup>2</sup>, which is equivalent to about 915 g/m<sup>2</sup>) ends up being consumed at the end of the tests.

These behaviours show that the oxidizing or reducing, acidic to alkaline, and organic or inorganic nature of the soil where the structure is buried plays a very important role.

Therefore, it is possible to draw conclusions by classifying soils based on three main variables: if the soil was organic or inorganic, if the soil was acidic or alkaline, and if the environment was oxidizing or reducing. Based on these three characteristics, the soils have been classified as follows: Cinders (Cin), Acid Oxidizing Inorganic (AOI) soils, Inorganic Oxidizing Alkaline (IOA) soils, Acid Reducing Inorganic (ARI) soils, Inorganic Alkaline Reducing (IAR) soils, and Acid Reducing Organic (ARO) soils. It should be noted that, for the rest of the possible combinations of values of these variables, there was no soil among those studied by Romanoff that fulfilled them. To determine the group to which the case under study belongs, the factors, degree of soil aeration, organic or inorganic nature, pH value (acidic or basic nature), and whether the water table is reached must be studied.

The objective of this study is the construction of a quantitative predictive model that estimates the thickness of hot-dip galvanized steel buried in the soil that will be lost over time due to the effect of corrosion. In this way, the engineer in charge of the design of this type of the structure will have a quantitative value of the loss of the material that the buried galvanized steel will experience as a function of time.

The paper is organized as follows. First, the composition of the studied galvanized steel is presented. Then, the steps for the construction of the database were described and how the model was built using the MARS technique. Finally, the results are presented and discussed, and the conclusions obtained in the investigation are detailed.

## 2. Materials and Methods

Starting from the data collected in the literature, a solid and robust database of hot-dip galvanized steel samples has been analysed, transformed, and prepared. Finally, a model has been developed to quantitatively estimate the loss of thickness that will occur in buried galvanized steel. The process is detailed below.

*2.1. Hot-Dip Galvanized Steel.* The material studied in this research is hot-dip galvanized steel. The approximate composition of steel (wt. %) is 0.09% carbon, 0.39% manganese, 0.08% phosphorus, and 0.04% sulphur, and its density is greater than that attributed to iron, reaching 7850 kg/m<sup>3</sup>. The galvanized coating used has a thickness of 3 oz/ft<sup>2</sup> equivalent to about 915 g/m<sup>2</sup> and consists of a continuous zinc coating with a density of 7140 kg/m<sup>3</sup>.

*2.2. Database.* The database that feeds the models has been constituted from two different sources: on the one hand, from the values compiled in the tables prepared by Romanoff and Denison [24, 25]; on the other hand, from the results of a predictive model that estimates the corrosion of buried bare steel. To continue, it is detailed how the data collected in the literature has been analysed and refined. Then, the introduction of the output variable of the other model in the database is explained.

Romanoff led a project that studied the corrosion of galvanized steel in 62 different soils. In 47 of them, only a point value of the loss of thickness of galvanized steel is available, which was taken around 10 years after burial. In the remaining 15 soils, various corrosion measurements are available at different times between the start of the experiment and its completion. The collection of these data provides much information on the evolution of corrosion over time and provides a detailed characterization of the soil variables involved in corrosion.

Once the raw data was configured, the next step was to assess the quality of the information contained in that set. After an exhaustive analysis, it was determined that there were deficiencies in the set that had to be solved for the database to be solid and reliable. These limitations had already been previously detected in other publications [41]. All of them are summarized in two groups: lack of data for some variables in the original trials and identification of unusual values.

To detect unusual values, the dispersion of each variable in each soil was studied. For instance, Figure 1 shows the scatter plot of the total acidity variable where the 62 soils studied by Romanoff are located on the  $x$ -axis and the  $y$ -axis showing its total acidity value. As can be seen, there is a value that is far to the rest of the values that the variable usually takes. These results that come out of the range of usual values of the variables have been defined as outliers.

After its identification, each specific outlier was studied, grouping its causes into two types: on the one hand, strange values due to errors in measurement, annotation, or transcription in the original tests; on the other hand, rare data due to real measurements in soils with special characteristics and, therefore, not caused by measurement or transcription errors.

The solution that has been adopted for the problem of the absence of data is not to fill in empty values by cases. The reason is that entering new data would distort the initial information. For unusual values, only cases where they have been shown with certainty to be errors have been removed. The selection of the MARS algorithm as a prediction technique has allowed working with a limited set of samples since this technique adjusts to this type of dataset better than other modelling techniques that require a more abundant dataset.

In summary, the treatment of the Romanoff data that has been carried out to achieve a reliable and representative database is described in Table 1.

Once the limitations of Romanoff's research were resolved, the way to solve the problem of data scarcity was studied. After multiple tests, it was concluded that the only possible way to generate a reliable model to predict the corrosion of galvanized steel was to add to the database the parameter thickness loss in  $\text{g}/\text{m}^2$  experienced by bare steel buried in the ground. The problem is that when a soil is to be studied to bury a galvanized steel structure, the loss of thickness that bare steel will experience in that same soil is not available. The decision taken to compensate for the lack of information was to enter into the database the information on the estimate of the loss of thickness of bare steel

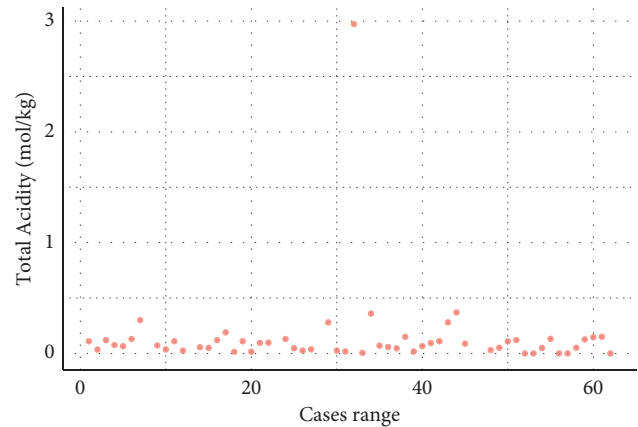


FIGURE 1: Scatter plot of the total acidity variable.

TABLE 1: Treatment of the database.

Database problems	Solution
Empty values	Not to falsify data
Outliers:	
Errors	Elimination
Soil with special characteristics	Not worked

provided by the basic predictive model of a previous study [33].

Once the modifications explained have been made, a reliable dataset is available which constitutes a firm foundation to feed the predictive models.

**2.3. Predictive Model.** The technique used for the development of the predictive corrosion estimation model is the MARS algorithm. This algorithm was designed by Friedman in 1991 to solve difficult, multivariate regression problems with complex nonlinear relationships. Regression problems are those in which a model must predict a numerical value. Its selection is based on its predictive capacity, its multi-dimensional adjustment, its robustness, and that once trained, it can be easily implemented in an application [42].

The MARS algorithm consists of finding a set of simple linear functions (splines) that characterize the data and uses them together to obtain the best predictive performance. The procedure consists of partitioning the definition domain of the function into different regions and adjusting each of them to a spline-type function. The MARS algorithm generates many of these simple linear functions, called base functions for one or more input variables.

The construction of the MARS model to fit the data is carried out in two phases [43]. In the first, an algorithm is used to select the base functions and the nodes. In the second, an algorithm is used to eliminate base functions, until the best set of these is found. At each step, the base function is suppressed, the elimination of which improves the degree of fit or causes the least loss of information. The purpose of this last stage is to reduce the degree of complexity of the model and significantly reduce the high dimensionality of the problem.

MARS is a multidimensional fitting method, and its objective is to approximate an unknown function  $f$  defined in a domain  $D$  contained in the vector space  $R^n$ , to a domain  $D'$  defined in  $R$ , from the projection of a sample of data from the input space onto the space of exit. For this, the algorithm looks for the approximation function  $f''(x_1, x_2, \dots, x_n)$ , by means of the linear combination of a set of base functions  $B_i(x)$  parameterized by the position of the nodes.

The final approximation provided by MARS is in the form of the following equation:

$$f'(x_1, x_2, \dots, x_n) = \sum_{i=1}^M a_i B_i(x), \quad (1)$$

where  $x$  belongs to  $D'$  contained in  $Rm$  and  $1 \leq m \leq n$  is the number of variables involved in the construction of the base function  $B_i$ .

The coefficients of the base functions  $B_i$  are  $a_i$  variables, and  $M$  is the number of base functions of the model.

The base functions used by MARS are built as products of base functions:

$$b_k(x) = (x - t_k)_+ = \text{Max}(0, x - t_k), \quad (2)$$

where  $t_k$  is the position of node  $k$  and the subscript “+” denotes the positive part of the affected parentheses.

The maximum number of base functions involved in the construction (number of products) is called the number of interactions between the variables. The value of this parameter is set by the user, and it remains fixed during algorithm training.

Thus, the base functions used by the MARS algorithm have the following form:

$$B_i(x) = \begin{cases} 1, & i = 1, \\ \prod_{j=1}^{J_i} [S_{ji} \cdot (x_{v(j,i)} - t_{ji})]_+, & i = 2, 3, \dots, \end{cases} \quad (3)$$

where  $J_i$  is the degree of interaction of the  $B_i$  functions,  $S_{ji}$  is the sign indicator ( $\pm 1$ ),  $v(j, i)$  is the index of the independent variable that is being divided, and  $t_{ji}$  the position of the division and is called the node.

In this paper, the input variables of the algorithm are 7, and they are detailed in Table 2. The output of the model is the loss of mass due to corrosion that galvanized steel will experience on the soil, expressed in  $\text{g/cm}^2$ .

The performance of this tool for the design of a galvanized steel structure on a soil consists in introducing in the model the 7 input variables of Table 2. Then, the developed model provides a quantitative value in  $\text{g/m}^2$  of the loss of thickness that the material will experience.

### 3. Results and Discussion

A key phase in the development of a predictive model is the evaluation of the reliability of its predictions. The aim is to compare the estimates proposed by the model with the real corrosion values in the training and test phases of the model.

TABLE 2: Input variables for the predictive model of galvanized steel.

Input variables	Unit
Resistivity	Ohm m
Temperature	$^{\circ}\text{C}$
Precipitation	mm/year
Soil moisture	%
pH	—
Time	years
Loss of mass due to corrosion of bare steel	$\text{g/m}^2$

The problem is that it is not possible to have a priori the corrosive behaviour that galvanized steel buried in soil will suffer over time. Therefore, the only way to evaluate the accuracy of the model is by dividing the database created with the actual values of thickness loss of galvanized steel due to corrosion in training and test patterns.

It was decided to allocate 90% of the cases to training patterns and 10% to test patterns. The reason for choosing a 90/10 ratio is based on the fact that removing more than 10% of the information that feeds the model in the training phase could distort the results due to the loss of relevant information.

The model has been trained with the degree of interaction between variables equal to 2 and maximum number of model terms before pruning, i.e., the maximum number of terms created by the forward pass equal to 35.

The following sections detail the evaluation of the predictive model. In addition, an analysis of the results according to the type of soil has been carried out to study the homogeneity of the corrosive behaviour in soils with the same typology.

**3.1. Predictive Model.** The statistical metrics used to evaluate the goodness of the model are Root Mean Square Error (RMSE), Coefficient of Determination ( $R^2$ ), Mean Absolute Error (MAE), and Relative Absolute Error (RAE).

In addition, a graphic study of standardized residuals has been carried out. The residual concept refers to the difference between the real value and the result predicted by the model and is therefore the estimation error. The graphs that have been used in this paper were regression error characteristic curve (REC), observed vs predicted, density of residuals, normal Q-Q, and scale location.

**3.1.1. RMSE and  $R^2$ .** It is important to highlight that the object variable has a variation interval between 51.87 and 5950.5  $\text{g/m}^2$ . The RMSE and  $R^2$  results obtained in the training and the model test are shown in Table 3.

An RMSE (train) = 280.43 and RMSE (test) = 290.60 have been obtained. As the variation range of the object variable is from 51.87 to 5950.5, the errors generated by the model in both cases are low for the slack experienced by the variable. The behaviour in both phases is very similar, so we can conclude that the model is stable.

The  $R^2$  value is very close to 1 in the model in the training phase and in the test phase. In the latter case, the result is slightly higher due to the smaller number of cases.

TABLE 3: RMSE and  $R^2$  of the model in the training and analysis phase.

Phase	RMSE ( $\text{g}/\text{m}^2$ )	$R^2$
Train analysis	280.43	0.93
Test analysis	290.60	0.96

3.1.2. *REC.* The purpose of Regression Error Characteristic curves is to compare different regression models. To do this, they plot the tolerance for error on the  $x$ -axis versus the precision of a regression model on the  $y$ -axis [44]. This accuracy value is calculated as the percentage of cases that fit each relative error tolerance.

Figure 2 shows the REC curve that results from comparing the training model (green line) and the test model (blue line). In addition, a control model (red line) has been considered that corresponds to the mean value of the corroded thickness. The quick rising of the curve to the top entails that the model provides correct predictions.

Table 4 shows the relationship between the RAE and the precision of the estimate in a quantitative way. In addition, it adds the MAE metric that expresses the loss of thickness in  $\text{g}/\text{m}^2$  that each percentage of relative error supposes.

As we can see in Table 4, the test phase and the training have a similar behaviour, which makes the model stable. Furthermore, for a relative error of 0.1, a precision of 0.95 over 1 is reached, and from a relative error of 0.14, the precision is maximum. In addition, it provides a significant improvement over the defined control.

3.1.3. *Predicted vs Predicted Plot.* Figure 3 reflects the results obtained when comparing the real corrosion values of galvanized steel with the values estimated by the model. In Figure 3(a), we can see the result of the training model, and in 3(b), the behaviour in the test phase. The blue line corresponds to the theoretical evolution of the model, and the points are the estimates it provides. The further the predictions are from the blue reference line, the points are represented by colours ranging from green (optimal) to red (maximum error).

The only difference between the behaviour in the training phase and in the test phase is the amount of data. In both cases, the points are adjusted to the theoretical evolution of the model, indicating an optimal behaviour of the predictive model.

3.1.4. *Density of Residuals' Plot.* The density plot detects the incorrect behaviour of the residuals and shows us the shape of the distribution of the relative error of the model. The peaks of the function reflect where the error values are concentrated.

Figure 4 shows the density plot of the predictive model in the training phase and in the test phase. As we can see, in both cases, the model of the relative error distribution is adjusted to the normal distribution.

The curves are highly centred on the mean errors, indicating that there are few modelling failures.

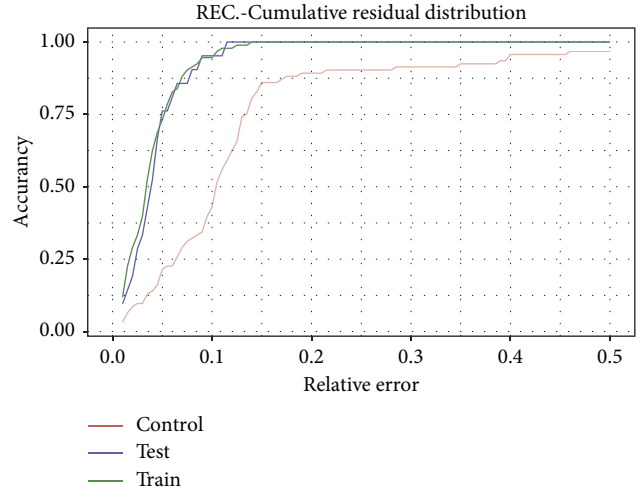


FIGURE 2: REC plot of the predictive model.

TABLE 4: RAE, MAE, and model accuracy training results.

RAE (%)	MAE ( $\text{g}/\text{m}^2$ )	Accuracy
0.01	59	0.11
0.05	295	0.75
0.10	590	0.95
0.14	826	1
0.20	1180	1
0.25	1475	1

3.1.5. *Normal Q-Q.* A Q-Q graph has been made to know the distribution of the error in the model. As we can see in Figure 5, both in the case of training and in the test, the points fall on the diagonal. This means that the model error follows a normal distribution.

The training phase and the test phase indicate that the predictive model follows a normal error distribution. Therefore, for errors that are associated with randomness, there is no cause of failure in the model.

3.1.6. *Scale-Location Plot.* Scale location plot is used to assess the independence of the model residuals. For a correct functioning of the model, they must respond to a cloud of points without any pattern. In Figure 6 we see that, in both the training and test models, there is no pattern in the placement of the points.

In this graph, the model in the training phase and the test phase do not follow any pattern; therefore, no anomalies are detected.

The statistical values RMSE,  $R^2$ , MAE, and RAE that have been obtained in the training and test models are very similar and reflect a stable and precise behaviour of the model. The standardized residual plots confirm that there are no anomalies in the performance and that the distribution of the errors adjusts to the normal distribution, which show the success of the estimates.

Therefore, according to the comparative evaluation of the training and testing phase carried out in this study, the results obtained show that the predictive model developed in

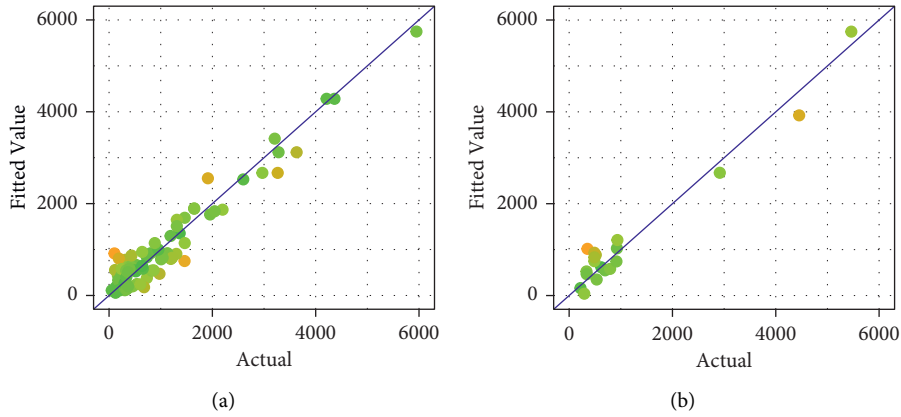


FIGURE 3: Real value vs estimate in the training phase (a) and in the test phase (b).

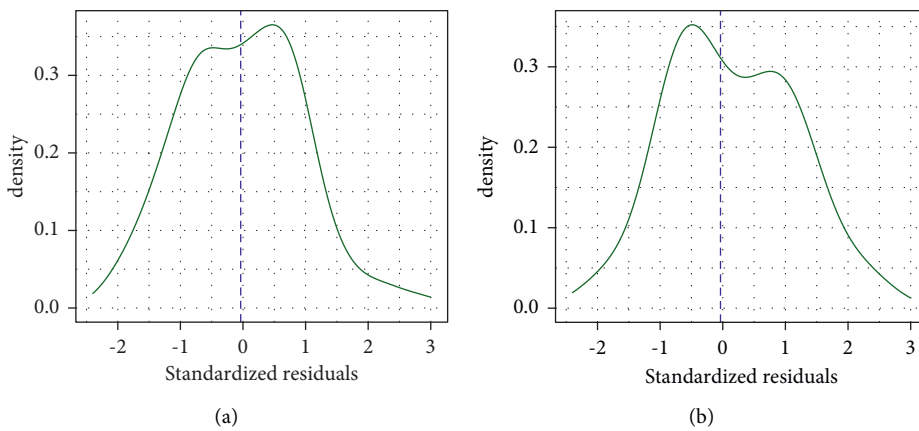


FIGURE 4: Density plot of the model in the training phase (a) and in the test phase (b).

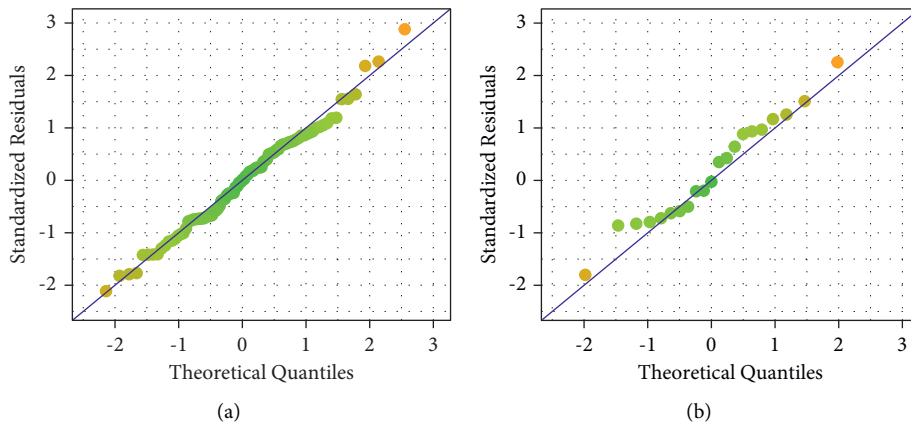


FIGURE 5: Normal Q-Q plot of the predictive model in training (a) and in test (b).

this paper allows to successfully estimate the loss of thickness of hot-dip galvanized steel buried in a soil. The main advantage of the developed model is that it provides the quantitative value of the thickness that will be lost due to the corrosion of the steel in the useful life of the structure. In this way, the quantity of the material is optimized with its service time. In addition, it eliminates the subjectivity in the

dimensioning of the structure, avoiding an excess of thickness that increases the cost of the structures.

3.2. Study of the Influence of the Type of Soil. As already mentioned in this paper, the existing literature in the field of corrosion corroborates similar behaviours of galvanized



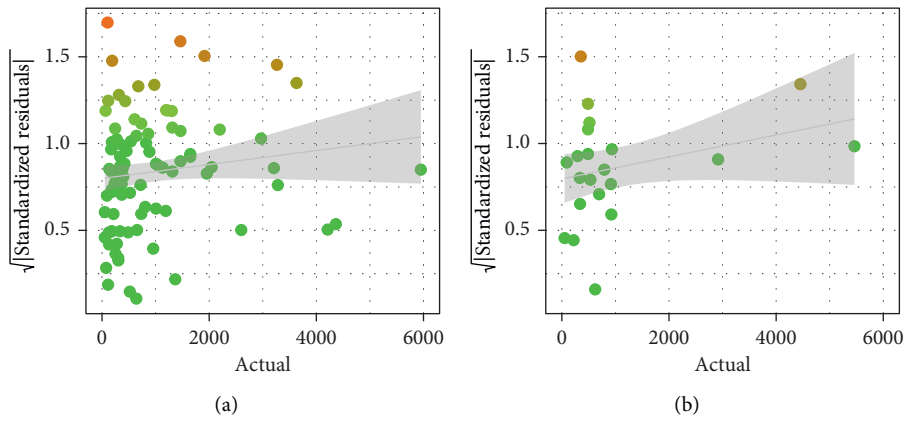


FIGURE 6: Scale-location plot in the training phase (a) and in the test (b).

TABLE 5: Analysis of the number of cases and the average thickness loss by soil type.

Soil group	Number of cases (%)	Average thickness loss
Cin	4.39	2704.31
AOI	28.95	247.07
IOA	9.65	469.66
ARI	18.42	664.98
IAR	18.42	1114.37
ARO	20.17	1918.90

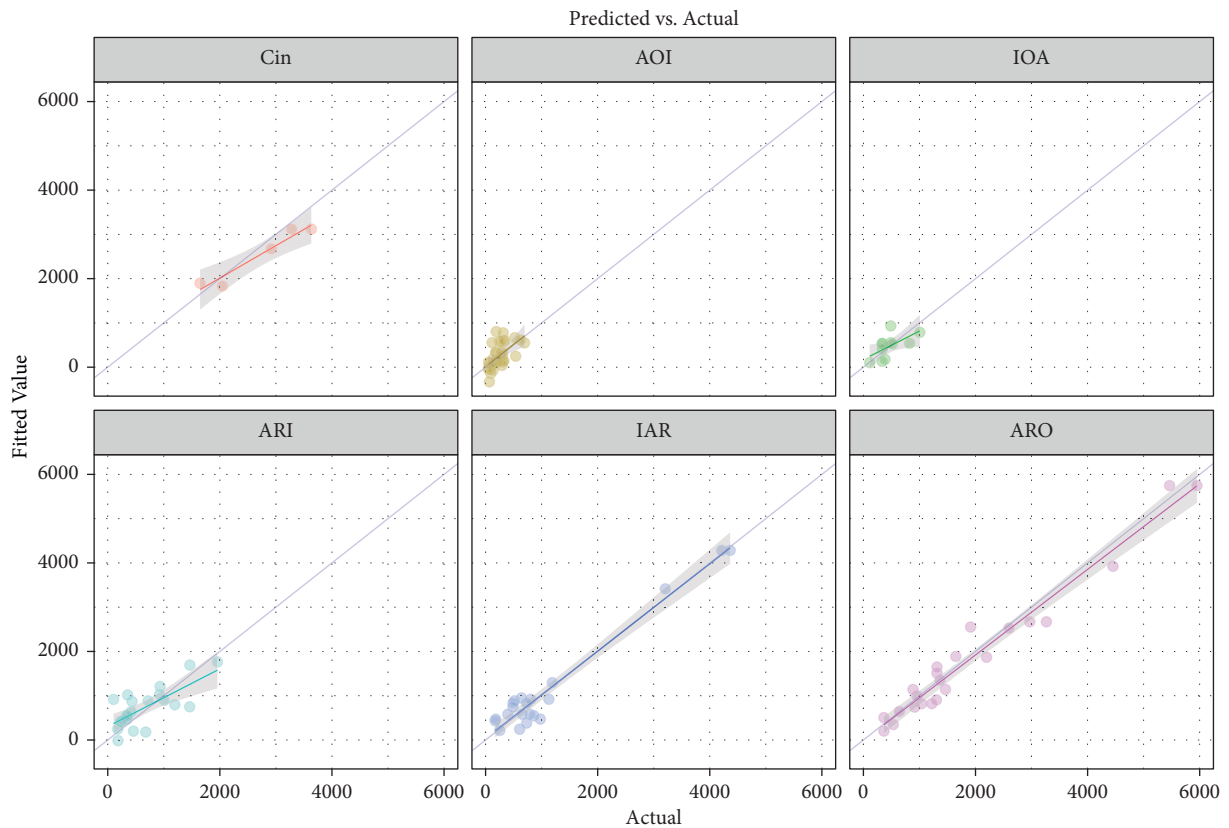


FIGURE 7: Predicted vs actual plot for each soil type.



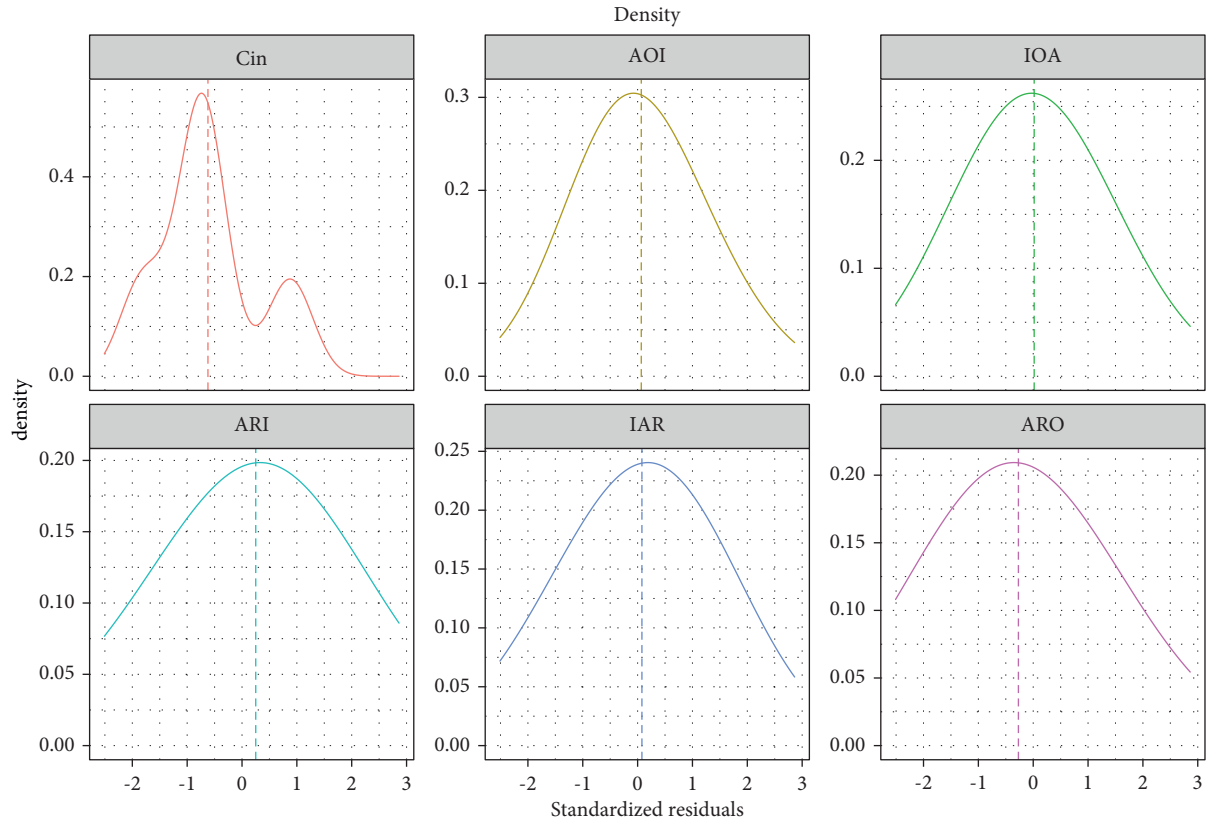


FIGURE 8: Density plot for each type of soil.

steel buried in soils of the same typology. For this reason, the study of the object variable, metal thickness loss, has been considered relevant, depending on the soil groups. Remember that the soils collected in the database are divided into Cin, AOI, IOA, ARI, IAR, and ARO.

Table 5 shows the amount of data available for each of them and the average loss of thickness of galvanized steel buried in each type of soil.

Figure 7 shows the progression of the model results in the optimal case (blue line) and the predictions made by the model for each of the soil types (points).

In all types of soil, the trend of the points is close to the main diagonal that represents the theoretical behaviour of the model; therefore, the behaviour is stable.

The distribution of the results for each type of soil is grouped in different areas of the main diagonal; accordingly, it is evident that the loss of thickness is similar to each type of soil.

The density graph has been constructed for each one of the soils to detect if there is any anomalous behaviour of the errors (Figure 8).

The functions are highly centred on the mean of the errors, which indicates the goodness of the model. There is an anomalous behaviour in the case of the Cin soil type, and the reason is that, of this soil, we have very few cases in the database (less than 5%).

The analysis of results by type of soil shows similar behaviours for soils belonging to the same typology, as shown in the existing bibliography.

## 4. Conclusions

The most widely used methods to estimate the thickness loss suffered by steel buried in a soil are qualitative guidelines that provide information on the aggressiveness of the corrosion. Therefore, the dimensioning of the structure depends on the responsible engineer. The predictive model developed in this paper consists of a multivariate quantitative model that provides the loss of thickness in  $\text{g/m}^2$  as a function of the useful life of the structure. The model is fed from a database consisting of real tests that Romanoff carried out for more than 20 years and is completed with the results of a multivariate predictive model that estimates the loss of ungalvanized steel in the soil due to corrosion.

One of the most remarkable points of the model is that we are introducing as input of the model the output given by other model that predicts the loss of thickness that a bare steel will experience in the same conditions. So, with this information, the model is able to learn the inherent relations and extrapolate all this information to forecast the loss of thickness of the galvanized steel.

The use of the model makes the dimensioning of the structure independent of the engineer, and an efficient relationship between the amount of material and durability is achieved. The results were evaluated through a comparative statistical analysis between the model in the training and testing phase, which confirmed the reliability of the estimates and the good performance of the model. In addition, it

shows the homogeneity of the corrosion behaviour in soils of similar typologies.

The major limitation of this study is the lack of data. Accordingly, the recommendations for future research are to carry out new tests in the areas of the multidimensional space that present lower density of points in order to achieve a more representative dataset. Due to the large time frames required to carry out representative tests on the evolution of corrosion over time, the database can be supplemented with accelerated corrosion tests to study the effect of specific conditions on the metal. Furthermore, a limitation of the research is that it responds to a single construction material. Therefore, the introduction of other reference materials in the field of structural engineering is proposed as a future line of research.

## Data Availability

The data supporting this research are from previously reported studies and datasets, which have been cited.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] O. Oudbashi, "A methodological approach to estimate soil corrosivity for archaeological copper alloy artefacts," *Heritage Science*, vol. 6, no. 1, p. 2, 2018.
- [2] K. M. Usher, A. H. Kaksonen, I. Cole, and D. Marney, "Critical review: microbially influenced corrosion of buried carbon steel pipes," *International Biodeterioration & Biodegradation*, vol. 93, pp. 84–106, 2014.
- [3] A. Chen, X. Ruan, and D. M. Frangopol, "Life-cycle civil engineering: innovation, theory and practice," in *Proceedings of the 7th International Symposium on Life-Cycle Civil Engineering*, CRC Press, Shanghai, China, February 2021.
- [4] O. V. Hembara and O. E. Andreikiv, "Effect of hydrogenation of the walls of oil-and-gas pipelines on their soil corrosion and service life," *Materials Science*, vol. 47, no. 5, pp. 598–607, 2012.
- [5] R. Pierre and P. D. Roberge, *Handbook of Corrosion Engineering*, ISBN 978-260-7, McGraw-Hill Education, New York, NY, USA, Third edition, 2019.
- [6] K. Siler-Evans, A. Hanson, C. Sunday, N. Leonard, and M. Tumminello, "Analysis of pipeline accidents in the United States from 1968 to 2009," *International Journal of Critical Infrastructure Protection*, vol. 7, no. 4, pp. 257–269, 2014.
- [7] G. A. Papadakis, "Major hazard pipelines: a comparative study of onshore transmission accidents," *Journal of Loss Prevention in the Process Industries*, vol. 12, no. 1, pp. 91–107, 1999.
- [8] B. Hou, X. Li, X. Ma et al., "The cost of corrosion in China," *Npj Materials Degradation*, vol. 1, no. 1, pp. 1–10, 2017.
- [9] G. H. Koch, M. P. H. Brongers, N. G. Thompson, Y. P. Virmani, and J. H. Payer, "Corrosion Cost and Preventive Strategies in the United States (No. FHWA-RD-01-156)," *CC Technologies Laboratories*, U.S. Federal Highway Administration, Washington, DC, USA, 2002.
- [10] G. Koch, J. Varney, N. Thompson, O. Moghissi, M. Gould, and J. Payer, "International measures of prevention, application, and economics of corrosion technologies study," *NACE International*, vol. 216, 2016.
- [11] A. Yajima, H. Wang, R. Y. Liang, and H. Castaneda, "A clustering based method to evaluate soil corrosivity for pipeline external integrity management," *International Journal of Pressure Vessels and Piping*, vol. 126–127, pp. 37–47, 2015.
- [12] L.-d. Arriba-Rodriguez, J. Villanueva-Balsera, F. Ortega-Fernandez, and F. Rodriguez-Perez, "Methods to evaluate corrosion in buried steel structures: a review," *Metals*, vol. 8, no. 5, p. 334, 2018.
- [13] Y. Beauregard and A. Mah, "Assessing soil corrosivity for buried structural steel," in *Proceedings of the International Pipeline Conference, American Society of Mechanical Engineers Digital Collection*, vol. 84447, p. V001T03A005, Virtual conference, 2021.
- [14] A. Poudel, K. Prasad Dahal, K. C. Dinesh, and J. Bhattarai, "A classification approach for corrosion rating of soil to buried water pipelines: a case study in budhanilkantha-maharajganj roadway areas of Nepal," *World Journal of Applied Chemistry*, vol. 5, no. 3, p. 47, 2020.
- [15] ASTM G187 - 18, Standard Test Method for Measurement of Soil Resistivity Using the Two-Electrode Soil Box Method, 2005.
- [16] American Water Works Association ANSI/AWWA C105, *Ductil-Iron Pipe Standard*, American Water Works Association, Denver, CO, USA, 2012.
- [17] P. R. Roberge, *Corrosion Basics: An Introduction*, NACE international, Houston, TX, USA, 2006.
- [18] M. N. Norhazilan, Y. Nordin, K. S. Lim, R. O. Siti, A. R. A. Safuan, and M. H. Norhamimi, "Relationship between soil properties and corrosion of carbon steel," *Journal of Applied Sciences Research*, vol. 8, pp. 1739–1747, 2012.
- [19] R. Amaya-Gómez, E. Bastidas-Arteaga, F. Muñoz, and M. Sánchez-Silva, "Statistical soil characterization of an underground corroded pipeline using in-line inspections," *Metals*, vol. 11, no. 2, p. 292, 2021.
- [20] H. Najjaran, R. Sadiq, and B. Rajani, "Modeling pipe deterioration using soil properties - an application of fuzzy logic expert system," *Pipeline Engineering and Construction*, vol. 2004, Article ID 14673, 10 pages, 2004.
- [21] M. Taghipour, G. R. Lashkaripour, M. Ghafoori, and N. Hafezimeghaddas, "Evaluating the soil corrosion of bushehr, Iran, based on a new classification system for corrosive soils," *Anti-corrosion Methods & Materials*, vol. 63, no. 5, pp. 347–354, 2016.
- [22] European Standard Store EN 12501-2, *Protection of Metallic Materials against Corrosion*, European Standard Store, Pilsen, Czech Republic, 2003.
- [23] German Technical and Scientific Association for Gas and Water DVGW GW 9, Evaluation of Soils in View of Their Corrosion Behaviour towards Buried Pipelines and Vessels of Non-alloyed Iron Materials, 2011.
- [24] M. Romanoff, *Underground Corrosion*, US Government Printing Office, Washington, DC, USA, 1957.
- [25] I. A. Denison and M. Romanoff, "Corrosion of galvanized steel in soils," *Journal of Research of the National Bureau of Standards*, vol. 49, no. 5, pp. 299–316, 1952.
- [26] D. Nguyen Dang, L. Lanarde, M. Jeannin, R. Sabot, and P. Refait, "Influence of soil moisture on the residual corrosion rates of buried carbon steel structures under cathodic protection," *Electrochimica Acta*, vol. 176, pp. 1410–1419, 2015.
- [27] T. Wu, J. Xu, M. Yan, C. Sun, C. Yu, and W. Ke, "Synergistic effect of sulfate-reducing bacteria and elastic stress on

- corrosion of X80 steel in soil solution,” *Corrosion Science*, vol. 83, pp. 38–47, 2014.
- [28] J. L. Alamilla, M. A. Espinosa-Medina, and E. Sosa, “Modelling steel corrosion damage in soil environment,” *Corrosion Science*, vol. 51, no. 11, pp. 2628–2638, 2009.
- [29] F. A. V. Bazán and A. T. Beck, “Stochastic process corrosion growth models for pipeline reliability,” *Corrosion Science*, vol. 74, pp. 50–58, 2013.
- [30] C. Yang, “Parallel-series multiobjective genetic algorithm for optimal tests selection with multiple constraints,” *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 8, pp. 1859–1876, 2018.
- [31] J. Zhang, H. Sun, Z. Sun, W. Dong, and Y. Dong, “Fault diagnosis of wind turbine power converter considering wavelet transform, feature analysis, judgment and BP neural network,” *IEEE Access*, vol. 7, pp. 179799–179809, 2019.
- [32] A. Pourdayaei, H. Mokhlis, H. A. Illias, S. H. A. Kaboli, and S. Ahmad, “Short-term electricity price forecasting via hybrid backtracking search algorithm and ANFIS approach,” *IEEE Access*, vol. 7, pp. 77674–77691, 2019.
- [33] L.-D. Arriba-Rodríguez, V. Rodríguez-Montequín, J. Villanueva-Balsera, and F. Ortega-Fernández, “Design of predictive models to estimate corrosion in buried steel structures,” *Sustainability*, vol. 12, no. 23, p. 9879, 2020.
- [34] D. Nakhaie, A. Kosari, J. M. C. Mol, and E. Asselin, “Corrosion resistance of hot-dip galvanized steel in simulated soil solution: a factorial design and pit chemistry study,” *Corrosion Science*, vol. 164, Article ID 108310, 2020.
- [35] L. Hanson, “Increase steel service life using hot-dip galvanizing,” *Building blocks*, vol. 11, 2015.
- [36] American Galvanizers Association, *Performance of Hot-Dip Galvanized Steel Products*, American Galvanizers Association, Centennial, CO, USA, 2010.
- [37] Robinson, J; Ltd, P., *Predicting the In-Ground Performance of Galvanized Steel*, Robinson, J; Ltd, P, Rock Hill, South Carolina, 2005.
- [38] R. F. D. C. Pereira, E. S. D. D. Oliveira, M. A. G. D. A. Lima, and S. L. D. C. Brasil, “Corrosion of galvanized steel under different soil moisture contents,” *Materials Research*, vol. 18, no. 3, pp. 563–568, 2015.
- [39] International Lead Zinc Research Organization Review of Data Available on the Corrosion Rate of Galvanized Steel in Soil, 2003.
- [40] J. Bian, Y. Zhu, X.-H. Liu, and G.-D. Wang, “Development of hot dip galvanized steel strip and its application in automobile industry,” *Journal of Iron and Steel Research International*, vol. 13, no. 3, pp. 47–50, 2006.
- [41] R. E. Ricker, “Analysis of pipeline steel corrosion data from NBS (NIST) studies conducted between 1922-1940 and relevance to pipeline management,” *Journal of research of the National Institute of Standards and Technology*, vol. 115, no. 5, p. 373, 2010.
- [42] J. H. Friedman, “Multivariate adaptive regression splines,” *Annals of Statistics*, vol. 19, pp. 1–67, 1991.
- [43] M. Kuhn and K. Johnson, *Applied Predictive Modelling*, vol. 26, Springer, Berlin, Germany, 2003.
- [44] B. Aşıkçıl and A. Erar, “Regression error characteristic curves based on the choice of best estimation method,” *Selçuk Journal of Applied Mathematics*, vol. 14, pp. 1–9, 2013.