

Identification of Corrosive Substances through Electrochemical Noise using Wavelet and Recurrence Quantification Analysis

Lorraine Marques Alves¹, Romulo A. Cotta¹, Adilson Ribeiro Prado² and Patrick Marques Ciarelli¹

¹Federal University of Espírito Santo, Av. Fernando Ferrari, 514, Vitória-ES, Brazil

²Federal Institute of Espírito Santo, Rodovia ES 010, km 6, 5, Serra-ES, Brazil

lorraine_ma@hotmail.com, rcotta@gmail.com, adilsonnp@ifes.edu.br, patrick.ciarelli@ufes.br

Keywords: Corrosion, Electrochemical Noise, Wavelet Transform, Recurrence Quantification Analysis.

Abstract: There are many types of corrosive substances that are used in industrial processes or that are the result of chemical reactions and, over time or due to process failures, these substances can damage, through corrosion, machines, structures and a lot of equipment. As consequence, this can cause financial losses and accidents. Such consequences can be reduced considerably with the use of methods of identification of corrosive substances, which can provide useful information to maintenance planning and accident prevention. In this paper, we analyze two methods using electrochemical noise signal to identify corrosive substances that is acting on the metal surface and causing corrosion. The first method is based on Wavelet Transform, and the second one is based on Recurrence Quantification Analysis. Both methods were applied on a data set with six types of substances, and experimental results shown that both methods achieved, for some classification techniques, an average accuracy above 90%. The obtained results indicate the both methods are promising.

1 INTRODUCTION

Corrosive substances are substances that by chemical action cause severe damage on contact with living tissue or, in case of leakage, damage materials or even destroy structures, means of transport and may cause many hazardous situations (Javaherdashti et al., 2013). Corrosive materials include acids, anhydrides, alkalis, halogens salts, organic halides and other substances that are widely used. Sulfuric acid, for instance, is widely used in manufacturing, for many chemical processes, and in automotive and industrial truck batteries. Sodium hydroxide is another corrosive material that is used in the purification of petroleum products, and in the manufacture of soap, pulp and paper (Allegri, 2012).

The health effects of corrosive substances are worrying factors in the industrial environment. Effects of direct contact vary from irritation causing inflammation to a corrosive effect causing ulceration and, in severe cases, chemical burns. Ignition of combustible materials may occur because some corrosive materials are oxidizers and some corrosives are unstable and tend to decompose when heated (Allegri, 2012). Therefore, the detection and monitoring of corrosive substances are of great importance for the preservation of health and prevention of industrial accidents.

The corrosion effect of these substances can be a source of unplanned costs. The global cost of corrosion is estimated around US\$ 2.5 trillion, equivalent to 3.4% of world GDP (Gross National Product) (Koch et al., 2016). This factor added to probability of accidents highlight the importance of researches and developing of technology in this field. Fortunately, due to the simultaneous occurrence of oxidation and reduction reactions during the corrosion process, it is possible to measure the current and electrical potential fluctuations on the surfaces that are suffering this process. These measured signals are called electrochemical noise (ECN) (Fofano and Jambo, 2007).

ECN signals have been used in corrosion monitoring processes for many years. But, only in the last decade that the real potential of these signals, combined with methods of useful features extraction, has become clearer (Al-Mazeedi and Cottis, 2004). An example of an application of ECN signals is the identification of corrosive substances, that can be useful to troubleshoot faults in industrial processes, assist in maintenance planning and even avoid accidents.

In this paper, we compare two techniques for features extraction of the ECN signals based on the wavelet transform and RQA (Recurrence Quantification Analysis), associated to machine learning techniques in order to create an intelligent system capable of de-

test different types of corrosive substances in aqueous solution. The results obtained in the experiments indicate that both approaches are promising.

2 ELECTROCHEMICAL NOISE DATA ANALYSIS

One of the biggest challenges in the analysis of electrochemical noise is related to the stochastic nature of the corrosion process, which result in most cases in nonstationary signals. The nonstationarity of electrochemical noise signals can be observed in two primary ways: by fluctuations in the variation of the potential or current and by the variation of statistical properties of the signal over time. One approach that has been used for ECN analysis is the wavelet transform. This method overcomes the limitations of the Fourier transform, since it enables the decomposition of the signal into diferente frequency components for different time intervals (Cottis et al., 2015). RQA is another approach to the analysis of ECN data that allows characterization of data by similarity matrix, containing the distances between subsequent measurements in the time series. Many variables can be derived from the similarity matrix and has been used to identify corrosion type and process monitoring (Hou et al., 2016).

2.1 Wavelet Transform Analysis

In conventional Fourier analysis is not possible to find in what period of time certain frequency band of a signal occurred, because this information is lost during the transform. A way to overcome this problem is to use the wavelet transform. Wavelet can distinguish the local characteristics of a signal on different scales and, by translations, they cover all the region in which the signal is studied. This locality property of wavelets is an advantage over the Fourier Transform in the analysis of nonstationary signals, being a more efficient tool, and applicable to the study of ECN signals (Aballe et al., 1999; Cottis et al., 2015).

For the analysis of discrete signals from sampled corrosive processes, it is conventionally used the Discrete Wavelet Transform (DWT) to obtain the coefficients values of different frequency bands for each time interval. These values are obtained by convolution of the sampled signal by functions that are displaced and dilated versions of a wavelet function (or mother wavelet). Thus, the original signal can be written as a sum of wavelet functions ($\phi_{J,n}(t)$ e $\psi_{J,n}(t)$) weighted by their corresponding coefficients, called detail ($d_{J,n}$) and smooth coefficients ($s_{J,n}$). These

coefficients indicate the correlation between the wavelet function and the corresponding signal segment (Aballe et al., 1999), as shown in Equations 1 to 3:

$$x(t) \approx \sum_n s_{J,n} * \phi_{J,n}(t) + \sum_n d_{J,n} * \psi_{J,n}(t) + \sum_n d_{J-1,n} * \psi_{J-1,n}(t) + \dots + \sum_n d_{1,n} * \psi_{1,n}(t), \quad (1)$$

$$s_{J,n} = \int x(t) \phi_{J,n}^*(t) dt, \quad (2)$$

$$d_{J,n} = \int x(t) \psi_{J,n}^*(t) dt, \quad (3)$$

where $n = 1 \dots N$, N is the length of the discrete signal and J stands for the decomposition level of DWT.

The coefficient matrix generated by DWT can be difficult to interpret for some ECN signals. A more useful way to represent the results of the wavelet transform in the analysis of electrochemical noise is through the concept of coefficient energy distribution. Thus, the contribution of each energy level of decomposition is calculated regarding the total energy of the signal. In this context, the signal energy may be calculated by (Aballe et al., 1999):

$$E = \sum_{n=1}^N x_n^2, \quad (4)$$

where E is the total energy of signal, x_n is the signal values in the instants $n = 1, 2, 3, \dots, N$ and N is the length of the discrete signal.

From the total energy E , the fraction of energy of each detail coefficient (E_j^d) and of smooth coefficient (E_j^s) can be calculated, respectively, according to Equations 5 and 6, where J are the levels used in the decomposition of the signal through the DWT.

$$E_j^d = 1/E \sum_{n=1}^{N/2^j} d_{j,n}^2, \quad (5)$$

$$E_j^s = 1/E \sum_{n=1}^{N/2^j} s_{j,n}^2. \quad (6)$$

Another recently developed ECN analysis tool is the concept of entropy associated with wavelet transform (Moshrefi et al., 2014). While the transform coefficients indicate the transient behavior of the signal, the concept of entropy is used to measure this degree of variability. Thus, the concept of entropy based on wavelet analysis reveals the degree of order/disorder of ECN signals, which will vary according to the conditions of the corrosion process. The entropy of a discrete random variable x with probability $p(x_i)$ can be defined by:

$$H(x) = - \sum_{i=1}^n p(x_i) \log(p(x_i)), \quad (7)$$

where $p(x_i)$ is estimated as the kernel density.

As the energy, entropy of the wavelet transform decomposition levels provides information to analyze the ECN signals that cannot be obtained through temporal analysis of the signals.

2.2 Recurrence Quantification Analysis

RQA is a developed approach for the analysis of dynamic systems and is based on the Recurrence Plots (RP) study. Recurrence matrix is the starting point for the discussion of the RQA theory. The formal concept of recurrence was introduced by Henri Poincaré in 1890 and, in a simplistic way, it states that an initial state or configuration of a mechanical system, subjected to conserved forces, will reoccur again in the course of the time evolution of the system (Bergelson, 2000). The RP method was developed for the visualization of the dynamic's trajectories in the phase space of dynamic systems. A recurrence plot is a graphical representation of a $N \times N$ matrix, whose elements are given by Equation 8:

$$RM_{i,j} = H(\epsilon - \|x_i - x_j\|), i, j = 1, 2, \dots, N, \quad (8)$$

where N is the number of states in phase space, ϵ is a predefined threshold radius, x_i and x_j are the points in phase space occurring at time i and j , $\|\cdot\|$ denotes the Euclidean norm of the vectors, and H represents Heaviside function. If the distance between x_i and x_j falls within the threshold radius, then $RM_{i,j} = 1$, otherwise, $RM_{i,j} = 0$ (Hou et al., 2016). In this paper, the matrix will be obtained on the time series of electrochemical noise data, similarly at (Hou et al., 2016).

The threshold value ϵ must be chosen correctly, since this value influences directly in the recurrence analysis. If ϵ is too large, almost all points will be identified as a recurrence point. On the other hand, if ϵ is too small, there may be too few recurrence points impairing the disclosure of recurrence structure (Marwan et al., 2007). In this paper, the ϵ value was fixed as 20% of the standard deviation of the original data segment, like used in (Hou et al., 2016).

Variables derived from the recurrence matrices, such as the recurrence rate (R), determinism (D), entropy (E) and average diagonal line length (L) are used to represent quantitatively recurrence plot (Marwan et al., 2007).

Given a $N \times N$ recurrence matrix $RM_{i,j}(\epsilon)$, $i, j = 1, 2, \dots, N$, then recurrence rate R (Equation 9) is the measure concerning the density of recurrence points and corresponds to the correlation definition for cases where the number of points is very large.

$$R = \frac{1}{N^2} \sum_{i,j=1}^N RM_{i,j}(\epsilon). \quad (9)$$

Determinism D is a measure of system predictability. According to Equation 10, $P(l)$ is the number of diagonals with length l in RP, and l_{min} is the smallest size for a row to be considered a diagonal (usually $l_{min} = 2$). In other words, the value of D is the reason between the number of points belonging to diagonals and the number of recurrence points.

$$D = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{i,j=1}^N RM_{i,j}(\epsilon)}. \quad (10)$$

The average length L of the diagonal lines is the number of points belonging to diagonals divided by the number of diagonals in RP, and it can be computed from Equation 11.

$$L = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=l_{min}}^N P(l)}. \quad (11)$$

Finally, E (Equation 12) measures the Shannon entropy of the probability $p(l) = P(l)/N_l$ to find a diagonal line with length l and reflects the complexity of the recurrence matrix with respect to diagonal lines.

$$E = - \sum_{l=l_{min}}^N p(l) \ln p(l). \quad (12)$$

In previous studies, authors suggest that corrosive events can be distinguished by the values of these features. For example, uniform corrosion can be associated with high values of R and low values of D , whereas localized corrosion is associated with low value of R and high value of D (Montalban et al., 2007; Garcia-Ochoa and Corvo, 2015).

3 MATERIALS AND METHOD TO COLLECT THE DATA

Corrosion analysis, through signal processing, consists in the mounting of an experimental apparatus, called electrochemical cell, and it is used an A/D (analog/digital) converter for the measurements of electrochemical noise. In this work, potential signals were measured and stored. Electrochemical cell is an experimental apparatus consisting of an inert metal immersed in an aqueous solution containing ions in different oxidation states. The cell used in this study consists of two steel electrodes AISI 1020 used as working electrodes and counter electrodes. These electrodes are nominally identical and coated with termocontract, and they have exposed area to solution equal to $18mm^2$.

According to the American Institute of Iron and Steel and International Society of Automotive Engineers, 1020 steel consists of about 0.18 to 0.23%

carbon (C), 0.3 to 0.6% manganese (Mn), at most 0.040% phosphorus (P) and 0.050% sulfur (S). Carbon steel 1020 is among the most used metals in the industry. The reference electrode used to collect data was silver/silver chloride ($Ag/AgCl$). The electrochemical cell is connected to the potentiostat interface, and this is connected to the computer, where the data can be stored. Figure 1 shows a diagram with the instruments used for data collection.

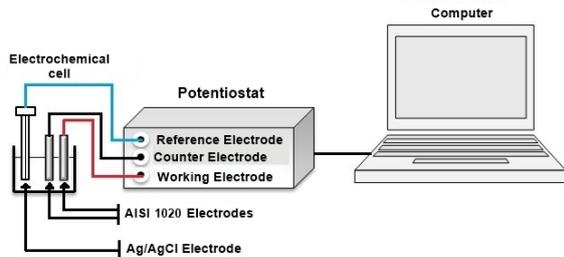


Figure 1: Experimental apparatus configuration.

Table 1 shows the substances used in this work and their applications. We have chosen substances which are common in industrial environments.

Table 1: Used reagents and their concentrations (values in mol/L) in aqueous solutions.

Substance	Concentration	Application
KCl	0.2, 0.4 and 0.6	Fertilizer production
$NaOH$	0.1, 0.2 and 0.3	Boilers
KOH	0.1, 0.2 and 0.3	Petrochemical industry
$NaCl$	0.2, 0.6 and 1.0	Cooling water
$FeCl_3$	0.1, 0.2 and 0.3	Coagulant in water treatment
H_2SO_4	0.2, 0.3 and 0.4	Fertilizer production

The acquisition of the signals was obtained using a potentiostat AUTOLAB[®] PGSTAT 101 Metrohm model. This instrument is equipped with three connections: working electrode, counter electrode and reference electrode. The reference electrode used was $Ag/AgCl$. To analyze the ECN measurements, for each reagent was performed three different concentrations, totaling 18 measurements with 60 minutes each. The sampling frequency used was 4Hz, such that each measurement has 14400 points. Figure 2 shows an example of measured signal.

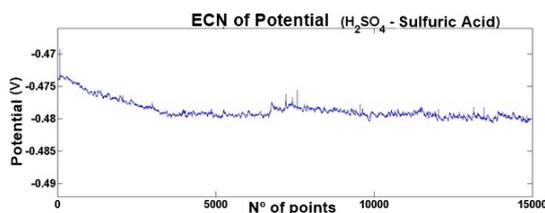


Figure 2: Potential ECN signal.

4 EXPERIMENTS AND DISCUSSION

The experiments were divided into two phases. In the first phase was defined the size of data segments, that is, the number of points of each sample before extracting the features. In the second phase, the classification algorithms are trained with the features extracted by wavelet and RQA to identify the type of corrosive substance. In both phases, the accuracy metric was used as quality measure. The value of the accuracy is calculated by the ratio of the number of samples correctly classified by the total number of samples, multiplied by 100%.

4.1 Size of Data Segments

The features used to identify the type of substance are extracted of several data segments. Therefore, it is important to define its size, because few points per segment supply little information, but, if is used many points per segment, it will be impossible to evaluate properly the proposed method. Thus, we evaluated non-overlapping data segments in the range of 144 to 2880 points, with increments of 144 points. For this experiment was used a total of 14400 points equally distributed in all six classes, where 70% of the points were used to train and the others 30% to test.

The first step of wavelet analysis method is to remove the mean of each time series and to define the corresponding wavelet family (father and mother) (Aballe et al., 1999). The features used in this experiment were computed from signal of potential with Wavelet Transform of Daubechies (db4) with decomposition at 8 levels. The main property of the Daubechies function is that it is a wavelet highly localized in time, which is good for electrochemical noise studies, where short time duration events are the norm (Bertocci et al., 1997). The features extracted of each non-overlapping data packets of ECN signal of potential were: energy, entropy, detail and smooth coefficients, as described in Section 2.1, resulting in a feature vector with 27 elements. These features were selected with SBS algorithm (Sequential Backward Feature Selection), which is a search algorithm that starts with a complete set of features and for each iteration removes the feature with the least impact on the accuracy. Thus, only the most significant features are kept (Dutra, 1999). After computing the features, they were normalized by the mean and standard deviation.

To classification, we used a MLP (Multi-Layer Perceptron) with 20 neurons in the hidden layer, learning rate equal to 0.001 and 1000 training epochs using Levenberg-Marquardt algorithm (Marquardt,

1963) for training. Figure 3 shows the accuracy values obtained for different numbers of points using wavelet transform features. It is observed that the greater the number of points per segment, the better is the accuracy.

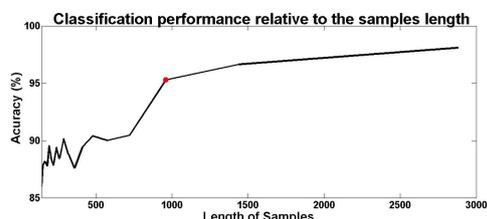


Figure 3: Accuracy \times sample length.

Ideally, a good choice is a large number of points per segment (sample). However, the higher number of points is, the smaller is the number of samples available for training and testing. A compromise between the two quantities is 960 points per segment, as indicated in Figure 3. Therefore, we can have a reasonable number of samples per class.

4.2 Corrosive Substances Identification

For these experiments, the features were extracted from non-overlapping data segments (samples) of 960 points each, from electrochemical noise of potential signals. Wavelet Transform of Daubechies (db4) with decomposition at 8 levels was used to compute the energy and entropy, and detail and smooth coefficients, resulting in a feature vector with 27 elements. Similarly, RQA was also used to extract features. The measures described in Section 2.2 were extracted through the RQA software 13.1 routines package, available at <http://homepages.luc.edu/~cwebber/>. Therefore, for each sample was obtain a vector with 4 elements.

After this step, the samples were stratified into 3 folds of data, each one with 45 samples to each class. Then, for each classifier were obtained 3 results from 3 tests, and each result was achieved using 2 folds for training/validation and 1 fold for testing. For each result was used a different fold for testing. The features were normalized by the mean and standard deviation computed in the training folds. In this work we used the following classifiers: MLP, PNN (Probabilistic Neural Network) (Masters, 1995), kNN (k Nearest Neighbor) (Duda et al., 2000), Decision Tree (Quinlan, 1988) and SVM (Support Vector Machines) (Theodoridis and Koutroumbas, 2008) with linear (SVM-L) and radial basis function (SVM-RBF) kernel.

The following configurations were tested for each technique in order to maximize the accuracy on the training set and, possibly, also on the test set. MLP

was trained using Levenberg-Marquardt algorithm, learning rate of 0.001, 1000 epochs and evaluated different numbers of neurons in the hidden layer (1 to 50 neurons). Different values of standard deviation were tested for PNN (0.1 to 1.0, with steps of 0.1). For kNN, we employed Euclidean distance and we varied the value of k (1 to 10). For SVM-L and SVM-RBF, different values of cost c were evaluated (0 to 10, with steps of 0.5). Moreover, different values of gamma (0.001 to 0.025, with steps of 0.005) were evaluated for SVM-RBF. For the decision tree was used standard Matlab implementation, which does not have parameters to tune. For each test, the parameters of each technique were adjusted in order to maximize the accuracy on the training folds, and these parameters were used to classify the samples of the test fold.

Table 2 shows the values of the accuracy obtained by each technique in each test fold when using wavelet transform to feature extraction. The best result for each fold is highlighted. As can be seen, MLP achieved mean accuracy slightly better than those obtained by the other classifiers, followed by SVM with linear kernel.

Table 2: Classification results of each technique when using wavelet to feature extraction (accuracy values are in percent).

Test	MLP	PNN	kNN	Tree	SVM-L	SVM-RBF
1	95.56	71.97	90.48	84.76	87.66	87.66
2	97.14	75.00	94.29	89.52	94.28	92.38
3	95.87	68.94	86.67	81.90	91.42	90.47
Mean	96.19	71.97	90.48	85.40	91.12	90.17
Std.	0.008	0.030	0.030	0.038	0.033	0.023

Similarly, Table 3 shows the classification results of each technique when using the features extracted by RQA. As can be seen, decision tree achieved mean accuracy slightly better than those obtained by the other classifiers.

Table 3: Classification results of each technique when using RQA to feature extraction (accuracy values are in percent).

Test	MLP	PNN	kNN	Tree	SVM-L	SVM-RBF
1	91.11	94.44	94.44	97.78	83.33	86.67
2	92.22	88.89	90.00	92.22	88.88	91.11
3	78.89	86.67	88.89	87.78	81.11	78.89
Mean	87.41	90.00	91.11	92.59	84.44	85.56
Std.	0.074	0.040	0.029	0.050	0.040	0.061

The results in Tables 2 and 3 indicate that both approaches, wavelet transform and RQA, present a high accuracy to perform detection of corrosive substances, but the best mean accuracy of wavelet transform was slightly better than the results obtained by RQA. Furthermore, we observed that MLP performance was better when using features extracted by wavelet transform, while decision tree was more effective when

using features of RQA. This indicates some classifiers are more appropriate to some type of features than others. The authors did not find works with similar objectives, but in the task of identifying types of corrosion, as in (Jian et al., 2013) and (Hou et al., 2016), that also use ECN signal, the accuracy is in the range of 90% to 100%.

5 CONCLUSIONS

This paper presented an approach to identify some types of reagents on metal surface through electrochemical noise signals using wavelet transform and recurrence quantification analysis. Comparing the two techniques, we noticed that both had a similar performance, but wavelet transform was able to provide a slightly higher average accuracy. For the classifiers evaluated, we noted that MLP achieved an average accuracy above 95% to perform the task. In relation to the classification stage, the feature vector obtained from the RQA is smaller, and requires less processing capacity.

Therefore, the results of this study highlight the importance of using wavelet transform and RQA for electrochemical noise analysis. In future work, we intend to analyze the combination of these methods, other algorithms and other features in order to improve the results.

ACKNOWLEDGEMENTS

Lorraine Marques Alves would like to thank CAPES for the scholarship granted.

REFERENCES

- Aballe, A., Bethencourt, M., Botana, F., and Marcos, M. (1999). Using wavelets transform in the analysis of electrochemical noise data. *Electrochimica Acta*, 44(26):4805–4816.
- Al-Mazeedi, H. and Cottis, R. (2004). A practical evaluation of electrochemical noise parameters as indicators of corrosion type. *Electrochimica Acta*, 49(17–18):2787–2793.
- Allegrì, T. (2012). *Handling and Management of Hazardous Materials and Waste*. Springer Science & Business Media.
- Bergelson, V. (2000). The multifarious poincaré recurrence theorem. *Descriptive set theory and dynamical systems*, 1:31–57.
- Bertocci, U., Gabrielli, C., Huet, F., and Keddan, M. (1997). Noise resistance applied to corrosion measurements. *Electrochemical Society Journal*, 144:31–37.
- Cottis, R. A., Homborg, A., and Mol, J. (2015). The relationship between spectral and wavelet techniques for noise analysis. *Electrochimica Acta*, pages 277–287.
- Duda, R., Hart, D. G., and Stork, P. (2000). *Pattern Recognition*. Wiley Interscience, New York, 1nd edition.
- Dutra, L. (1999). Feature extraction and selection for ers-1/2 insar classification. *International Journal of Remote Sensing*, pages 993–1016.
- Fofano, S. and Jambo, H. (2007). *Corrosion: fundamentals, monitoring and control*. Ciencia Moderna, Brazil, 1nd edition.
- Garcia-Ochoa, E. and Corvo, F. (2015). Using recurrence plot to study the dynamics of reinforcement steel corrosion. *Protection of Metals and Physical Chemistry of Surfaces*, 51(4):716–724.
- Hou, Y., Aldrich, C., Lepkova, K., Suarez, L., and Kinsella, B. (2016). Monitoring of carbon steel corrosion by use of electrochemical noise and recurrence quantification analysis. *Corrosion Science*, pages 63–72.
- Javaherdashti, R., Nwaoha, C., and Tan, H. (2013). *Corrosion and Materials in the Oil and Gas Industries*. CRC Press.
- Jian, L., Weikang, K., Jiangbo, S., Ke, W., Weikui, W., Z. Weipu, and Zhoumo, Z. (2013). Determination of corrosion types from electrochemical noise by artificial neural networks. *International Journal of Electrochemical Science*, 8:2365–2377.
- Koch, G., Varney, J., Thompson, N., Moghissi, O., Gould, M., and Payer, J. (2016). International measures of prevention, application and economics of corrosion technologies study (impact). Technical report.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.
- Marwan, N., Romano, M., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics reports*, 438:237–329.
- Masters, T. (1995). *Advanced algorithms for Neural Networks: A C++ Sourcebook*. John Wiley and Sons, New York, 1nd edition.
- Montalban, L., Henttu, P., and Piche, R. (2007). Recurrence quantification analysis of electrochemical noise data during pit development. *International Journal of Bifurcation and Chaos*, 17:3725–3728.
- Moshrefi, R., Mahjani, M., and Jafarian, M. (2014). Application of wavelet entropy in analysis of electrochemical noise for corrosion type identification. *Electrochemistry Communications*, 48:49–51.
- Quinlan, J. (1988). *Decision trees and multivalued attributes*. Oxford University Press, New York, 1nd edition.
- Theodoridis, S. and Koutroumbas, K. (2008). *Pattern Recognition*. Elsevier, San Diego, 4nd edition.