

# Monitoring of Carbon Steel Corrosion by Use of Electrochemical Noise and Recurrence Quantification Analysis

Y. Hou<sup>1</sup>, C. Aldrich<sup>1\*</sup>, K. Lepkova<sup>2</sup>, L.L. Machuca<sup>2</sup>, B. Kinsella<sup>2</sup>

<sup>1</sup>Department of Mining Engineering and Metallurgical Engineering, Western Australian School of Mines, Curtin University, PO Box U1987, Waterford, WA 6152, Australia

<sup>2</sup>Curtin Corrosion Engineering Industry Centre, School of Chemical and Petroleum Engineering, Curtin University, Bentley, WA 6102, Australia

<sup>1,\*</sup>Corresponding author, [chris.aldrich@curtin.edu.au](mailto:chris.aldrich@curtin.edu.au)

## Abstract

*The corrosion of carbon steel in aqueous media resulting in uniform corrosion, pitting corrosion and passivation was investigated on a laboratory scale. Recurrence quantification analysis was applied to short segments of electrochemical current noise measurements. These segments were converted to recurrence variables, which could be used as reliable predictors in a multilayer perceptron neural network model to identify the type of corrosion. In addition, an automated corrosion monitoring scheme is proposed, based on the principal component scores of the recurrence variables. This approach used the uniform corrosion measurements as reference data and could differentiate between uniform and non-uniform corrosion.*

**Keywords:** *Carbon steel, Electrochemical noise, Pitting corrosion, Recurrence quantification analysis, Process monitoring, Neural networks*

## 1. Introduction

It is well-established that corrosion is a major global problem causing damage in the order of hundreds of billions of dollars in the USA alone [1, 2] and damage on a similar scale in other developed economies. The most dangerous form of corrosion is localized, where unexpected, rapid damage to local metal structures can lead to catastrophic failure [3]. These failures are difficult to prevent, since localized corrosion is difficult to measure reliably, despite extensive investigation over many years. For example, although visual observation of weight-loss coupons used widely in industry can give an indication of the incidence of localized corrosion, these methods are very slow and cumbersome. Likewise, electrical resistance methods are also not particularly useful, as localized corrosion is often associated with negligible change on the electrical resistance of the metal. In contrast, methods based on electrochemical noise (EN) measurement have shown more promise as reliable indicators of localized corrosion [3-6].

Electrochemical noise can be ascribed to the formation of microcells on the surfaces of metals subject to corrosion. These microcells give rise to oscillating current and potentials that contain important information on the dynamics of the corrosion process [7, 8]. As a consequence, the detection of localized corrosion based on electrochemical noise measurements has been studied based on a number of different analytical approaches. These include statistical analysis of the data [9-11], Fast Fourier transforms, maximum entropy methods [12, 13] with power spectral density analysis, wavelet transforms [14] with transient analysis and energy distribution, phase space methods with correlation dimension [7, 15] and recurrence quantification analysis [8, 16-22].

Recurrence quantification analysis (RQA) in particular, is an emerging approach to the analysis of time series data. The RQA approach allows characterization of data by a similarity matrix, typically containing the Euclidean distances between subsequent measurements in the time series. A number of variables such as recurrence ratio, determinism, entropy and average diagonal line length etc., can be derived from a binary or thresholded version of the similarity matrix. In previous studies [16-22], authors implied that recurrence ratio and determinism were related to the initiation rate and interaction time of microcells on the metal respectively. The dynamics of uniform corrosion tends to be associated with higher recurrence ratio and low determinism value, while localized corrosion is characterised with low recurrence ratio and higher determinism. In spite of all the efforts of applying RQA to study corrosion dynamics based on EN data, recurrence variables have not been used for corrosion type identification and process monitoring.

In this study, an automated monitoring scheme for aqueous corrosion systems based on the use of recurrence quantification analysis of electrochemical noise is proposed. A multivariate approach is taken, where four recurrence variables are used as predictors in a neural network model to predict different corrosion behaviour in the steel. In addition, a corrosion monitoring scheme based on principal component model is proposed, aiming at detecting different corrosion mechanisms online.

## 2. Methodology

### 2.1. Recurrence quantification analysis

Recurrence is a fundamental feature of many nonlinear dynamic systems. The graphical way to visualize this feature is called recurrence plots (RPs). RPs with its quantification analysis (RQA) is an advanced tool designed for the analysis of dynamic systems [16, 23].

In previous studies, recurrence plots were based on phase space reconstruction theory, which requires embedding of a time series in a so-called ‘phase space’ [17]. They are graphical representations of an  $N \times N$  matrix, which can be denoted as indicated by eq. (1), where  $N$  is the number of states in phase space,  $\varepsilon$  is a predefined threshold radius,  $\mathbf{x}_i, \mathbf{x}_j$  are the points in phase space occurring at times  $i$  and  $j$ ,  $\|\cdot\|$  denotes the Euclidean norm of the vectors, and  $H$  represents Heaviside function. If the distance between  $\mathbf{x}_i$  and  $\mathbf{x}_j$  falls within the threshold radius, then  $R_{i,j} = 1$ , otherwise,  $R_{i,j} = 0$ .

$$R_{i,j} = H(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), i, j = 1, 2, \dots, N \quad (1)$$

In this paper,  $R_{i,j}$  is calculated on the basis of original data space instead of phase space, i.e. the data was taken as measured, while generating the RPs. Therefore,  $\mathbf{x}_i, \mathbf{x}_j$  represent the measured current values at times  $i$  and  $j$ , and  $\|\cdot\|$  refers to Euclidean distance between these two data points.

The threshold value ( $\varepsilon$ ) is a critical parameter influencing recurrence analysis. If  $\varepsilon$  is too small, there may be too few recurrence points and the recurrence structure of the underlying system could not be revealed. On the other hand, if  $\varepsilon$  is too large, almost every point would be identified as a recurrence point, resulting in many artefacts [24]. Consequently, care should be taken when choosing the threshold value. There are several criteria proposed in the literature regarding the selection of  $\varepsilon$  [25-27]. In this study, the threshold value was fixed as 20% of the standard deviation of the original data segment.

Several methods were proposed to represent recurrence plots quantitatively [24]. Among them, variables derived from the recurrence matrices, such as the recurrence rate (R), determinism (D), entropy (E) and average diagonal line length (L) are commonly used [28], and have also been considered in this investigation. More formally, given a recurrence matrix  $\mathbf{R}(\varepsilon) \in \mathbb{R}^{N \times N}$ , derived from a time series  $\mathbf{y} \in \mathbb{R}^{N \times 1}$  with  $(i,j)$ th element  $R_{i,j}(\varepsilon)$ ,  $i, j = 1, 2, \dots, N$ , then

- $R$  represents the density of the recurrence points in the entire recurrence plot, it can be calculated according to eq. (2):

$$R = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}(\varepsilon) \quad (2)$$

- Likewise,  $D$  represents the ratio of the number of recurrence points which form diagonal line segments to the total number of recurrence points in the superior triangle of the recurrence plot without counting the number of points that constitute the bisector (Fig. 6). The diagonal line segment is defined as two or more adjacent points that form a line that is parallel to the bisector of the square.  $D$  is a measure of the predictability of a system and is calculated based on eq.(3):

$$D = \frac{\sum_{l=l_{min}}^N l P(l)}{\sum_{i,j}^N R_{i,j}(\varepsilon)} \quad (3)$$

In eq. (3),  $l_{min}$  is a fixed value, which defines the minimum number of points in the diagonal line segment. A value of  $l_{min} = 2$ , was used in this investigation. Finally,  $P(l)$  is the frequency of lengths  $l$  of the diagonal lines.

- $E$  is the Shannon entropy of the probability  $p(l) = P(l)/N_l$  to find the diagonal line with length  $l$ , where  $P(l)$  is the frequency of diagonal lines of length  $l$ .  $E$  reflects the complexity of the recurrence matrix with respect to the diagonal lines.  $E$  is computed from eq.(4):

$$E = - \sum_{l=l_{min}}^N p(l) \ln p(l) \quad (4)$$

○  $L$  represents the average length of the diagonal lines in the recurrence plot, and can be interpreted as the mean prediction time. It is computed from eq.(5):

$$L = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=l_{min}}^N P(l)} \quad (5)$$

All the above calculations were implemented using free RQA software described elsewhere [28].

Table 1 gives an overview of corrosion systems studied via RQA of electrochemical noise measurements. On the one hand, a large proportion of scholarly work has focused on the study of the corrosion of stainless steel using RQA. This is the first time a study applies RQA to corrosion processes at carbon steel. On the other hand, the authors mostly used the physical meanings of RQA variables  $D$  and/or  $R$  to indicate the dynamics of different corrosion stages or regimes. In this study, four variables, i.e.  $D, E, R, L$  are used simultaneously from the statistical point of view, rather than exploring the relationship of particular parameter(s) with corrosion processes.

Table 1. Recurrence quantification analysis of electrochemical noise in corrosion systems.

Variables	Corrosion System	References
$D$	Analysis of electrochemical oscillations generated from pitting corrosion of copper in different saline solutions at various potentials	[8]
$D$	Changes in the dynamics of corrosion fatigue in UNS S31603 stainless steel immersed in natural seawater	[16]
$D, R$	Protective effect of nanostructured films on austenitic stainless steel 304	[17]
$D, E, R$	Initiation, steady state progress and decay of pitting in austenitic stainless steel 316 in 300 mg/L of NaCl in deionized water	[18]
$D, E, R$	Intergranular corrosion of sensitized stainless steel UNS S30400 immersed in 0.5 M H <sub>2</sub> SO <sub>4</sub> & 0.01 M KSCN solution	[19]
$D, R$	Effect of hydrostatic pressure on corrosion of Ni-Cr-Mo-V stainless steel in a 3.5%NaCl solution	[20]
$D, R$	Effect of ultrasonic vibrational solidification treatment on corrosion of AZ80 Mg alloy	[21]
$D, R$	Carbon reinforcement steel in concrete exposed to a 3% NaCl solution	[22]

The extraction of recurrence variables from electrochemical noise in this investigation is shown diagrammatically in Fig. 1. The electrochemical noise signal is segmented into  $n$  non-overlapping time series each containing  $N$  measurements. The distance matrix  $\mathbf{D}_j$  contains the Euclidean distances between each pair of points in the time series segment  $\mathbf{y}_j$  and  $\mathbf{R}_j$  the recurrence matrix, is simply the thresholded version of  $\mathbf{D}_j$ . Vector  $\mathbf{q}_j$  is derived from  $\mathbf{D}_j$  and contains  $m$  recurrence variables. The matrix  $\mathbf{Q}$  contains the  $n$  recurrence variable vectors extracted from all the  $n$  segments of the electrochemical noise signal.

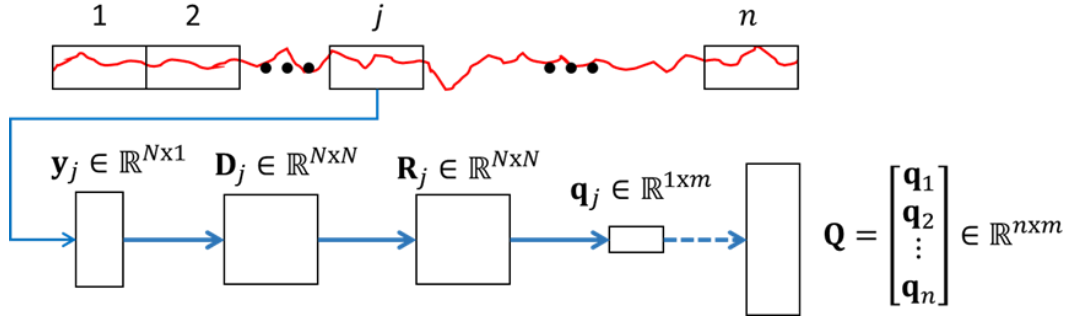


Fig. 1. Derivation of recurrence variables from  $n$  non-overlapping segments of a time series.

## 2.2. Principal component model for corrosion monitoring

Monitoring of the corrosion in the steel was based on the use of a principal component model derived from data matrix  $\mathbf{Q} \in \mathbb{R}^{n \times m}$ , with  $n$  samples and  $m$  variables. Monitoring is based on off-line calibration of the model, and online application of the model.

### 2.2.1. Off-line calibration of the principal component model

If  $\mathbf{Q} \in \mathbb{R}^{n \times m}$  comprises the data matrix representative of the process with  $m$  variables and  $n$  measurements,  $\mathbf{S}$  is the covariance matrix of the recurrence variables typically scaled to zero mean and unit variance,  $\mathbf{P}_k$  is the loading matrix of the first  $k < m$  principal components,  $\mathbf{\Lambda}_k$  is a diagonal matrix containing the  $k$  eigenvalues of the decomposition,  $\tilde{\mathbf{P}}_k$  is the loading matrix of the  $m-k$  remaining principal components and  $\tilde{\mathbf{\Lambda}}_k$  is a diagonal matrix containing the  $m-k$  remaining eigenvalues of the decomposition. Off-line model calibration is done with data ( $\mathbf{Q}$ ) associated with uniform corrosion.

$$\mathbf{S} = \frac{\mathbf{Q}^T \mathbf{Q}}{n-1} = \mathbf{P}_k \mathbf{\Lambda}_k \mathbf{P}_k^T + \tilde{\mathbf{P}}_k \tilde{\mathbf{\Lambda}}_k \tilde{\mathbf{P}}_k^T \quad (6)$$

$$\mathbf{T}_k = \mathbf{Q} \mathbf{P}_k \quad (7)$$

In eq. (7),  $\mathbf{T}_k \in \mathbb{R}^{n \times k}$  is the score matrix of the principal component model consisting of the first  $k$  principal components. These scores that are associated with uniform corrosion dominated operating conditions, are enclosed by a control limit. In this investigation, the first  $k = 2$  principal components were retained, as they collectively explained approximately 84% of the variance of the variables. Moreover, no assumptions were made with regard to the distribution of the data, as the control limit was fitted by means of a Gaussian mixture model.

### 2.2.2. On-line application of the model

During process monitoring, newly measured data consisting of a time series of electrochemical noise measurements is segmented to yield  $\mathbf{y}_{new,i}$  from which a distance matrix  $\mathbf{D}_{new,i}$  is calculated. Through thresholding  $\mathbf{D}_{new,i}$ , the recurrence matrix  $\mathbf{R}_{new,i}$  is extracted. Then  $\mathbf{R}_{new,i}$  is converted into vector  $\mathbf{q}_{new,i}$ . This vector of recurrence quantification variables is then projected onto the principal component model, to generate the principal component scores,  $\mathbf{t}_{new,i}$ , which can be computed from eq. (8)

$$\mathbf{t}_{new,i} = \mathbf{q}_{new,i} \mathbf{P}_k \quad (8)$$

The model error is calculated from eq. (9)

$$\mathbf{e}_{new,i} = \mathbf{q}_{new,i} - \hat{\mathbf{q}}_{new,i} \quad (9)$$

If this observation falls outside the control limits placed on uniform corrosion ( $\mathcal{L}_T$ ), it is flagged as pitting or passivation.

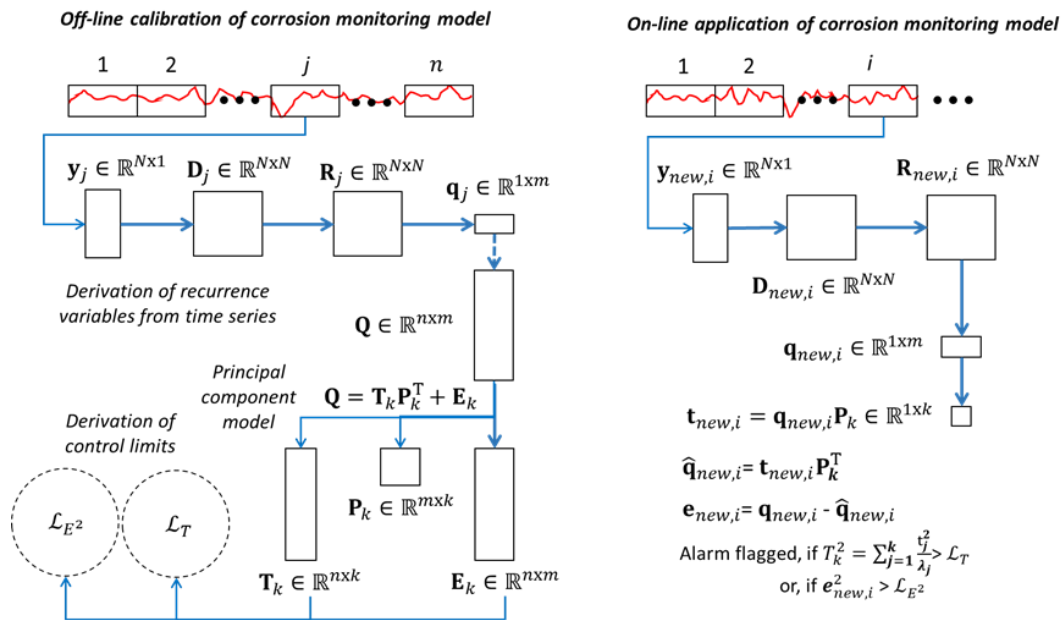


Fig. 2. Off-line calibration (left) and on-line monitoring of corrosion from electrochemical noise (right).

The process monitoring approach is shown diagrammatically in Fig. 2. The general approach is well established in chemical processing engineering (see for example [29]).

### 3. Experimental work

The chemical composition of the carbon steel (1030) used in the experiments is as follows (wt.%): C (0.37), Mn (0.80), Si (0.282), P (0.012), S (0.001), Cr (0.089), Ni (0.012), Mo (0.004), Sn (0.004), Al (0.01), and Fe (balance). Two rectangular samples with surface area of 1.4 cm × 1.5 cm were made from the same material and soldered with a conducting wire to each for electrical connection. Subsequently, the samples were electro-coated (Powercron 6000CX) and embedded in epoxy resin (Epofix), leaving 2.1 cm<sup>2</sup> of the steel surface exposed. The final shape of the sample is illustrated in Fig. 3. Before tests, the exposed surfaces of the steel samples were abraded on silicon carbide paper up to 1200 grit, rinsed with ultra-pure water and ethanol and dried with nitrogen. This procedure was

conducted prior to each run of the tests. The same set of steel samples with identical grinding procedures prior to each run were used in all the tests. After tests, the surfaces of the samples were examined for localized corrosion and the maximum pit depth was measured using optical 3D microscopy (Infinite focus microscope, Alicona Instruments, Austria).

Three test solutions were used, viz. 0.1 M sodium chloride (NaCl; Merck, 99.7%), 0.5 M sodium hydrogen carbonate (NaHCO<sub>3</sub>; Merck, 99.7%), and a solution containing 0.45 M sodium hydrogen carbonate and 0.1 M sodium chloride (0.45 M NaHCO<sub>3</sub> + 0.1 M NaCl). These solutions were used to set up uniform corrosion, passivation and pitting systems, respectively. All the solutions were prepared with ultra-pure water (Milli-Q system, resistivity 18.2 MΩ cm) and analytical reagents.

Electrochemical noise (EN) measurement were carried out in a configuration which is schematically shown in **Error! Reference source not found.** Two nominally identical aforementioned carbon steel samples were placed parallel to each other as working electrodes (WE 1 and WE 2), facing the operator. The inter-electrode spacing was fixed at 2.4 cm (centre-to-centre). In addition, a commercial Ag/AgCl electrode was used as reference electrode (single junction electrode placed in a capillary with a porous ceramic tip and filled with 3 M KCl solution designed to give a potential of 0.210 V against S.H.E). The current flowing through the two working electrodes was measured using a zero resistance ammeter mode (ZRA) of Gamry Reference 600 and the potential of the short-circuited WEs was measured with regard to the commercial reference electrode. ESA410 data acquisition software was employed to collect the EN data. The sampling rate was 2 Hz.

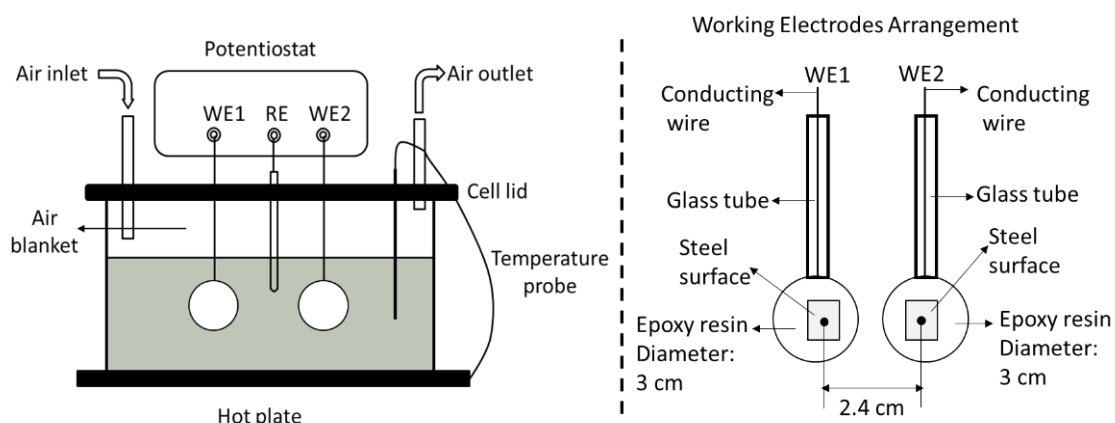


Fig. 3. Experimental setup.

All the test solutions were air-saturated in a separate cell before pumping into the test cell equipped with the electrodes. During the tests, the temperature of the solution was controlled at  $30 \pm 1$  °C by a hot plate with a temperature probe. The oxygen level in the solution was kept constant by continuously pumping air through the cell, creating an air blanket on top of the solution.

For uniform and passivation tests, the EN measurement started immediately after immersion of the test samples. After exposure to the test solutions, the samples were taken out of the test cell and rinsed immediately with ultra-pure water and ethanol, followed by drying with nitrogen gas. Afterwards, the surfaces of the two working electrodes were directly observed under optical microscope.

Table 2. Experimental conditions.

Corrosion Type	Solution	pH	Test Duration
Uniform	0.1 M NaCl	6.6	18 h
Pitting	0.45 M NaHCO <sub>3</sub> + 0.1 M NaCl	8.6	19 h
Passivation	0.5 M NaHCO <sub>3</sub>	8.4	2.5 h

For pitting corrosion tests, the specimens were prepassivated in 0.5 M NaHCO<sub>3</sub> solution for four hours (without EN recording), before NaCl was added to yield a solution of 0.45 M NaHCO<sub>3</sub> + 0.1 M NaCl. The EN signals were recorded immediately after the NaCl solution was added. The experimental conditions are summarized in Table 2. All the tests were done in duplicate. Potential and current noise were recorded simultaneously and the electrochemical current noise (ECN) was subjected to the analytical process described above. The electrochemical potential noise (EPN) signals were used as an indicator of the repeatability of the tests.

## 4. Results and discussion

### 4.1. Electrochemical noise measurements

The electrochemical noise (EN) data (after linear detrending) associated with the three types of corrosion in Table 2 are shown in Fig. 4(a)-(c). The microscopic appearance of the same sample surface associated with three different types of corrosion is shown in Fig. 4(d)-(f). Amplified images of current and potential noise signals corresponding to the time intervals starting from vertical dashed lines in Fig. 4 are displayed in Fig. 5.

It can be seen from Figs. 4 and 5 that the electrochemical current and potential signals show different behaviours for different corrosion systems. During the test period for uniform corrosion, high frequency fluctuations were observed for both potential and current noise. The maximum amplitude for potential was 12 mV while that for current was around 15  $\mu$ A. The enlarged portion of EN signals presented in Fig. 5(a) shows no distinctive peaks. In the case of pitting corrosion, current noise fluctuated rapidly around zero with very small amplitude (approx. 0.02  $\mu$ A) for about seven hours. Afterwards, the current signal showed typical peaks for metastable pitting, i.e., sudden rise followed by exponential decline, as indicated in Fig. 5(b). The amplitude was increased to approximately 0.03  $\mu$ A at this stage. After 10 h of EN recording, the amplitude grew larger with a maximum value of 0.63  $\mu$ A. Potential noise showed similar patterns. It also should be noted that after pitting test, only one of the working electrodes was pitted. The maximum pit depth measured was 55  $\mu$ m. But the other working electrode had a passive-like appearance. In contrast, for passivation system, the EN signals fluctuated with a relatively lower frequency with amplitudes no more than 0.02  $\mu$ A and 2 mV, respectively.



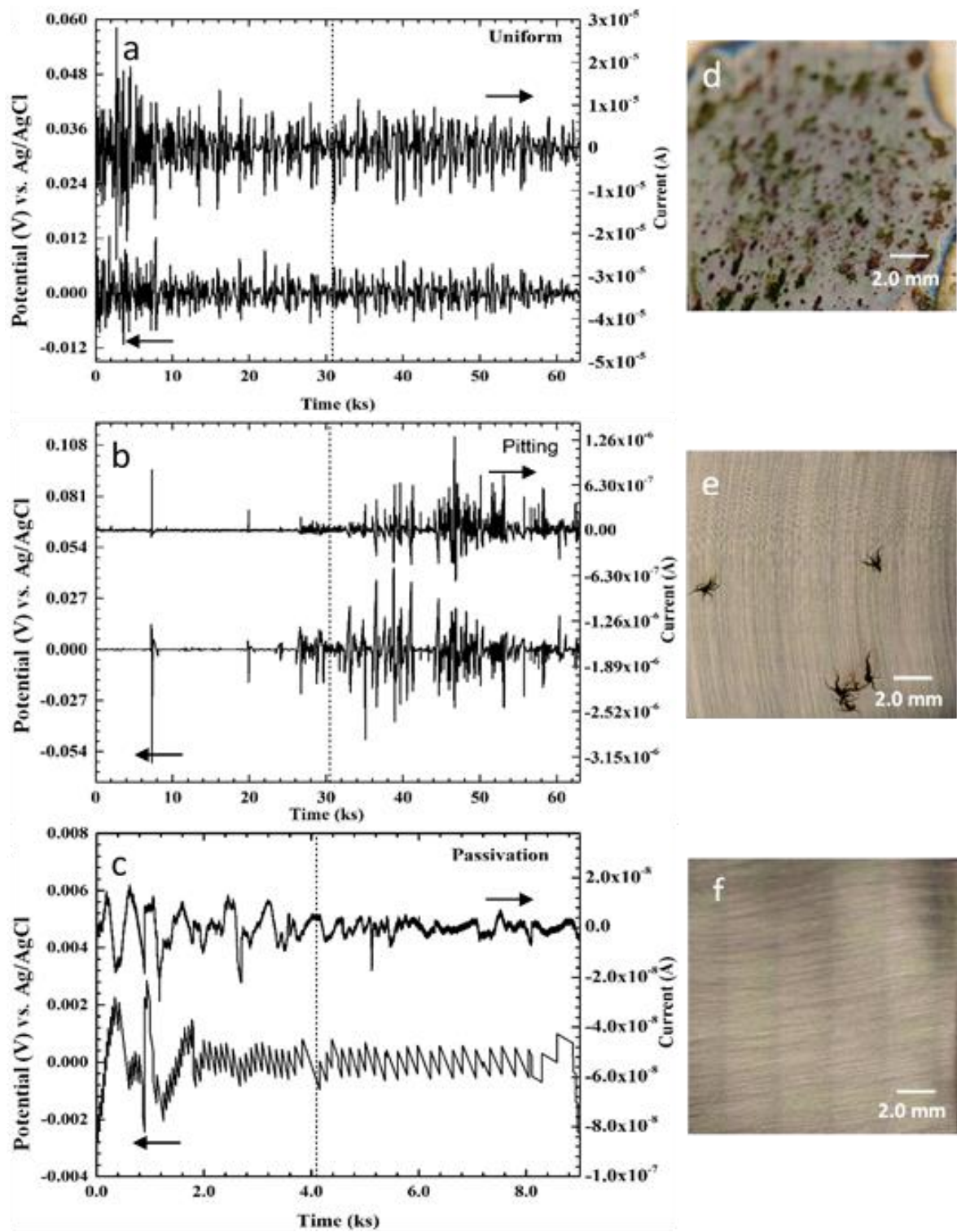


Fig. 4. Electrochemical potential (bottom line) and current (top line) noise signals and associated appearance of steel surfaces for three types of corrosion (a, d) uniform (b, e) pitting (c, f) passivation.

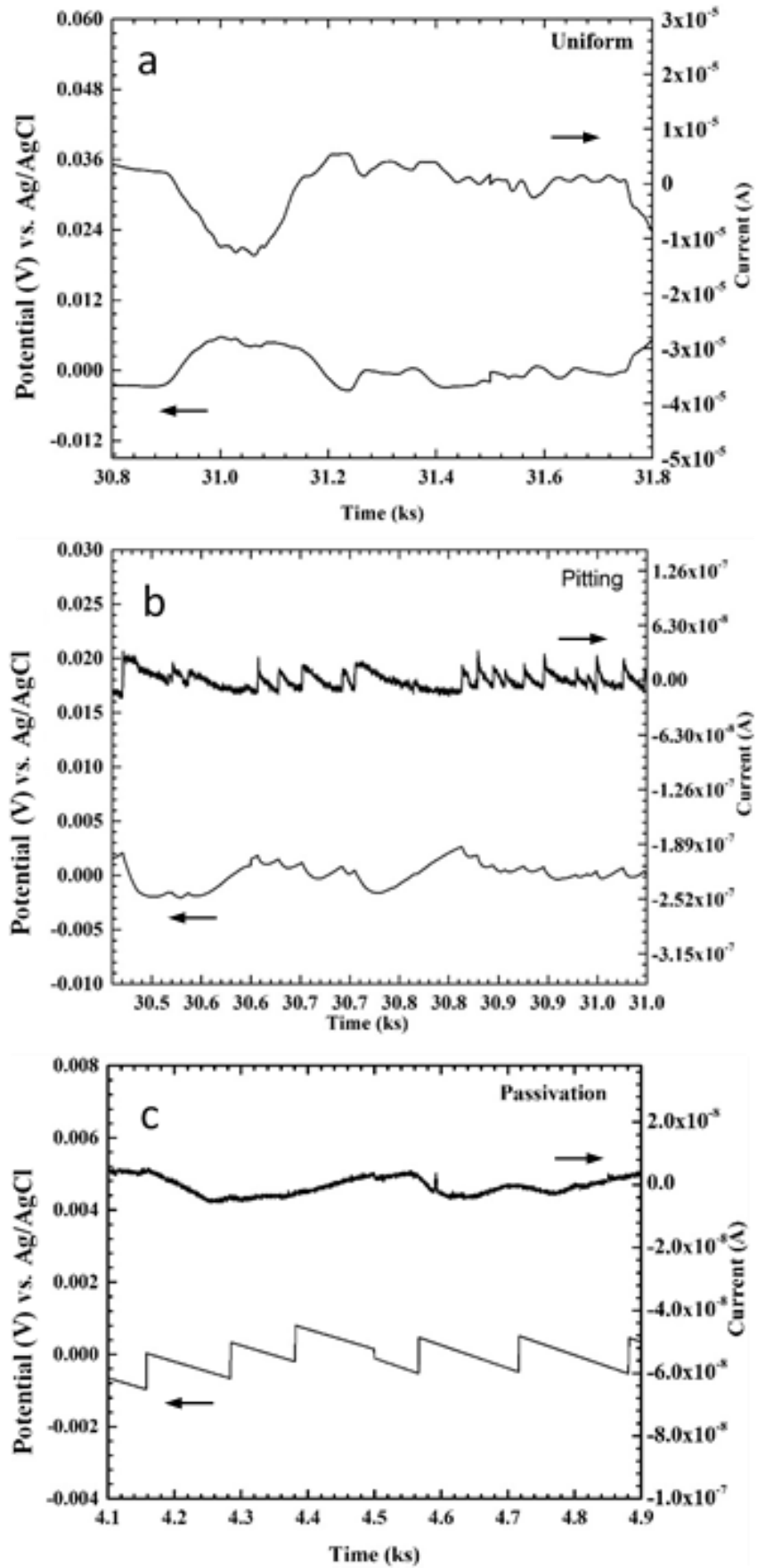


Fig. 5 Enlarged portions of marked areas in Fig. 4 (a) Uniform corrosion (b) Pitting (c) Passivation.

The recurrence plots of the signals generated according to the scheme presented in Fig. 1, are shown in Fig. 6. The ECN data for uniform corrosion, pitting and passivation is chopped into 70, 75 and 10 nonoverlapping time series segments respectively. Each of the segments contains 1800 data points. Four randomly selected data segments of the uniform corrosion process are given in the top row of the figure, to give an indication of the recurrence structures and their variation. Likewise, the images in the second and third rows represent the recurrence plots for pitting corrosion and passivation respectively.

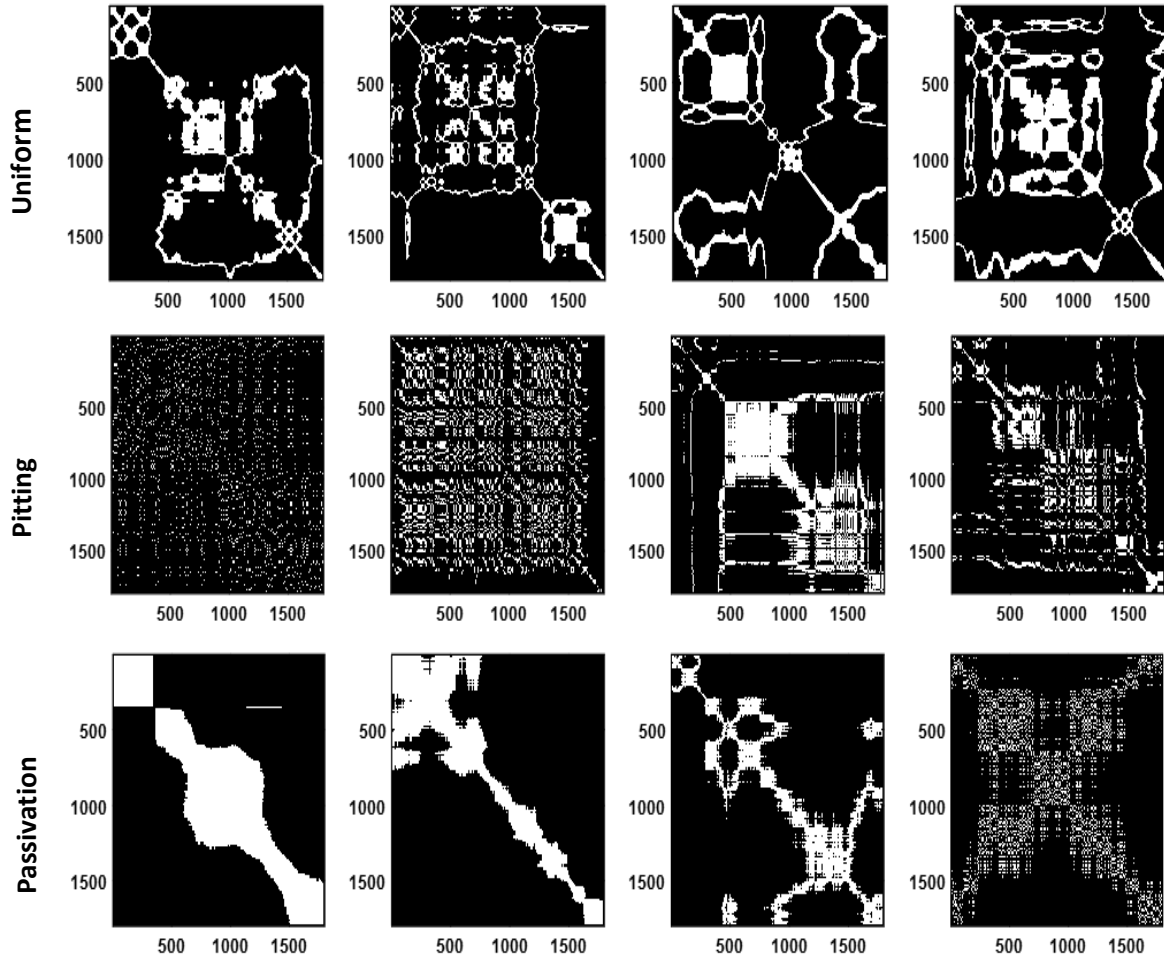


Fig. 6. Recurrence plots for different current data segments of three corrosion types, i.e., uniform (first row), pitting (second row) and passivation (third row).

The calculated RQA variables (eqs (2)-(5)) for the different corrosion systems are shown in Fig. 7. The vertical lines separate the three different corrosion systems, i.e. uniform corrosion, pitting corrosion and passivation, from left to right. It is indicated that, in the studied systems,  $D$  and  $E$  were capable of distinguishing different corrosion types to some extent. For example, the  $E$  values fluctuate between 0 and 2 for uniform corrosion, while that for pitting corrosion are located in the range of -2 to 0. It is difficult, however, to distinguish passivation from pitting. Apparently,  $R$  had no capability to discriminate different corrosion systems since all the values fluctuated at around zero.  $L$  is similar to  $R$ , but to some extent,  $L$  performs better in the separation of uniform and pitting corrosion.

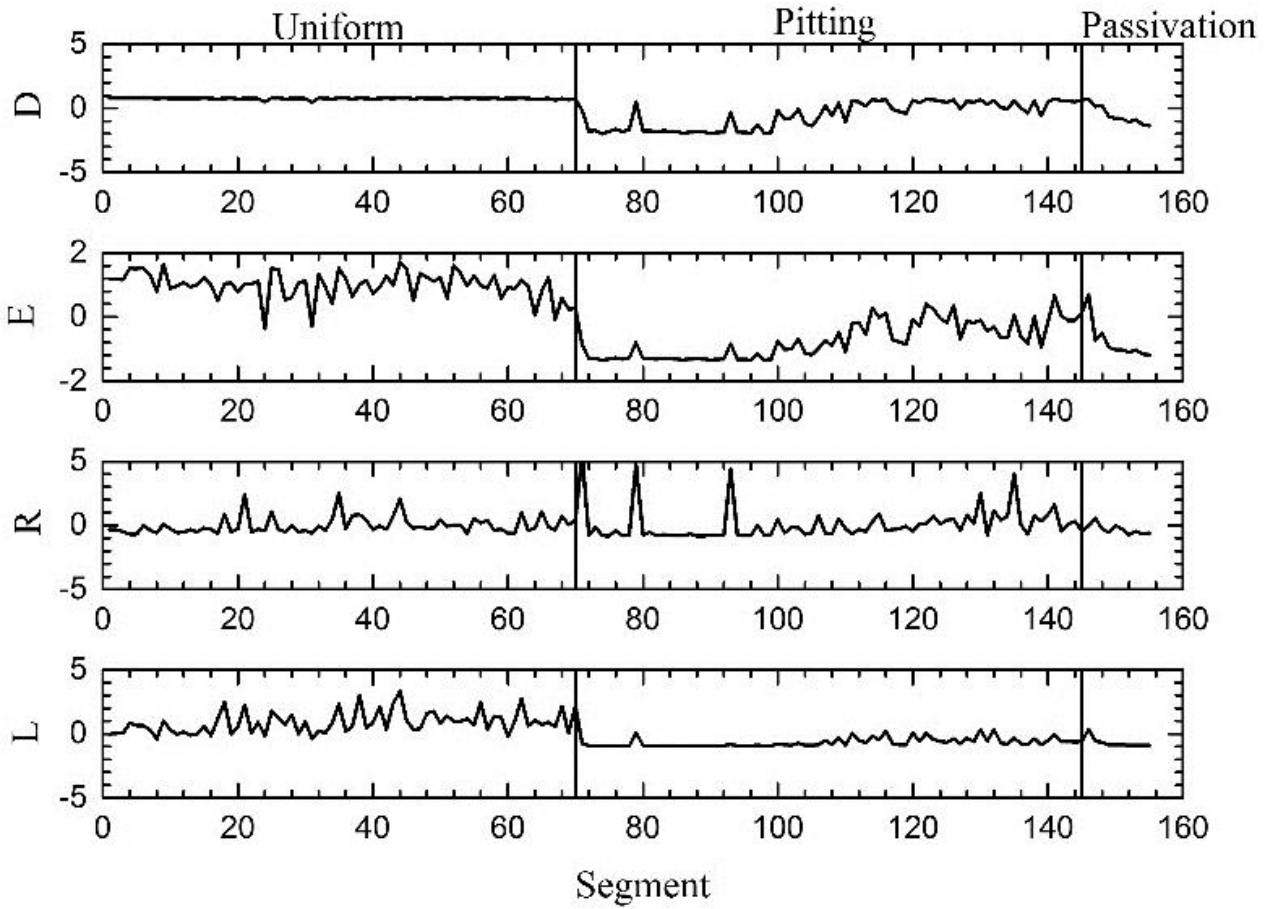


Fig. 7. Line plots of the zero mean, unit variance scaled values  $D$ ,  $E$ ,  $R$  and  $L$  (from top to bottom). The vertical lines separate uniform corrosion, pitting corrosion and passivation, from left to right.

The correlational structure of the four variables, determinism, entropy, recurrence ratio and length, is shown in Table. The values in the table represent the linear correlation coefficients between any two variables. This can be calculated from eq. (10).

$$r = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \quad (10)$$

Where  $X$  and  $Y$  represent any two RQA variable vectors,  $cov(\cdot)$  calculates the covariance between these two vectors,  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of  $X$  and  $Y$ , and  $r$  is the correlation coefficient of these two RQA variables. Since it has been mentioned before that the four RQA variables formed a data matrix  $\mathbf{Q}$ , the actual calculation was conducted on  $\mathbf{Q}$  using a simple Matlab code.

It can be seen from Table,  $D$  is strongly correlated with  $E$ , indicating that the information contained in  $\mathbf{Q}$  is redundant. In order to reduce the redundancy, principal component analysis (PCA) was conducted on  $\mathbf{Q}$  to extract its principal components (PC). Finally, the first two scores of the corresponding PCs were obtained.

Table 3. Correlation among four RQA variables.

Variables	<i>D</i>	<i>E</i>	<i>R</i>	<i>L</i>
<i>D</i>	1	0.816	0.339	0.550
<i>E</i>	0.816	1	0.126	0.638
<i>R</i>	0.339	0.126	1	0.319
<i>L</i>	0.550	0.638	0.319	1

Fig. 8 shows the scatter plot of the principal component scores for each of the three corrosion systems. Where  $t_1$  and  $t_2$  are the scores of the first two PCs. The percentage of variance explained by each PC is 62.0% and 22.8% respectively. Uniform corrosion appears to be clearly distinguishable from the other two systems, yet it is not possible to discriminate between passivation and pitting corrosion.

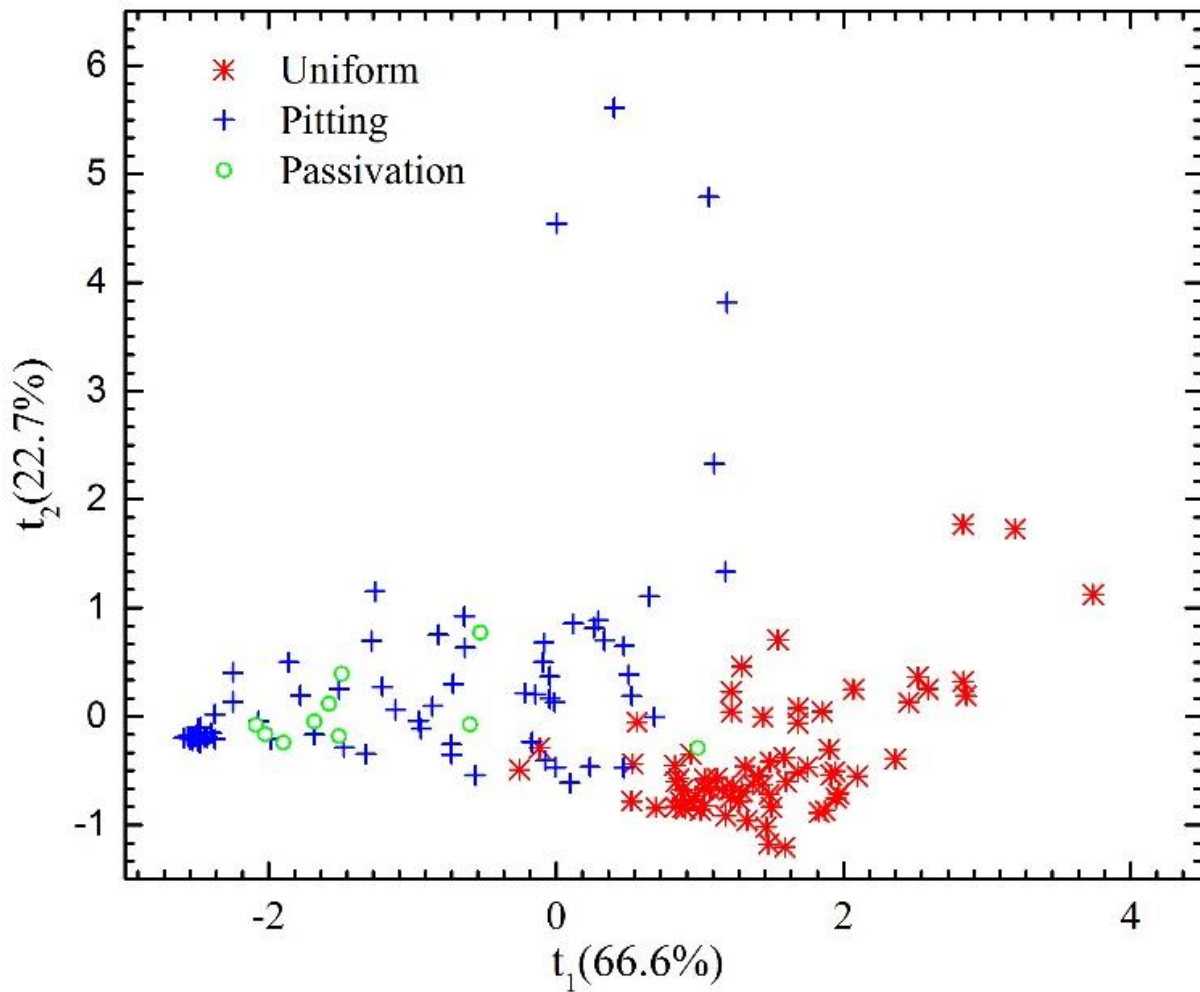


Fig. 8. Principal component score plot of three corrosion types.

Although Fig. 7 and Fig. 8 give some indication of the ability of the variables to discriminate between the different types of corrosion, further quantitative analysis was done by constructing a multilayer neural network model to classify the data, as shown in Fig. 9. The input layer of the network consisted of four nodes associated with four predictor variables. The hidden layer consisted of 8 nodes, while the output layer had three nodes, since the three different output classes (corrosion types), were coded in a [1 0 0; 0 1 0; 0 0 1] format. Both the hidden and output nodes had bipolar sigmoidal activation functions (*tansig*) and the Levenberg-Marquardt algorithm was used to train the network in Matlab 14b with a cross-entropy error function.

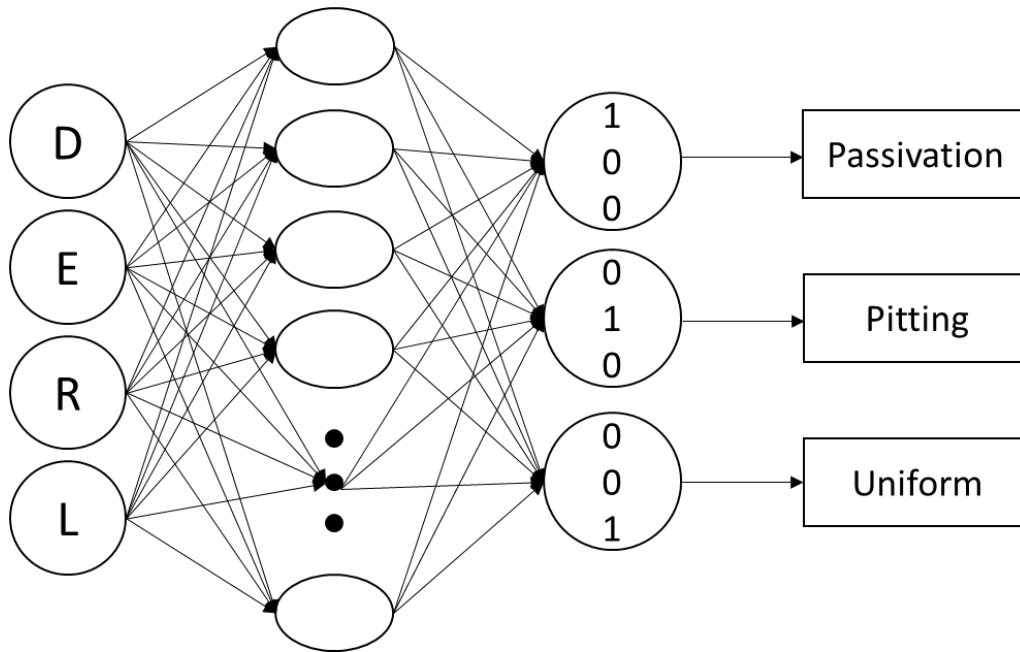


Fig. 9 Architecture of two-layer feed-forward neural network

The training data (70 samples from the uniform corrosion system, 75 samples from the pitting corrosion system and 10 samples from the passivation system) were randomly divided between a training data set (60% of the samples), and a validation and test data set, each containing 20% of the samples. The training and validation data sets were used during the construction of the classification models, while the test data served as an independent data set to assess the performance of the trained models.

The results are shown in the so-called confusion matrix in Fig. 10, where classes 1, 2 and 3 correspond with passivation, pitting corrosion and uniform corrosion, respectively. The test data were not used in the development of the model and is used as an indication of the generalized performance of the trained neural network. From Fig. 10, it can be seen that there were 31 samples in the test set, of which 3 belonged to Class 1, 13 to Class 2 and 15 to Class 3. The neural network was unable to identify Class 1 reliably and misallocated 2 of the 3 samples to Class 2. These results are consistent with the information shown in Fig. 8, where it is clear that it would be difficult to discriminate between Classes 1 and 2. The overall accuracy of the model was 93.5%. However, since the data sets used in this model were relatively small, especially the data set associated with passivation, these results should be interpreted as indicative only.

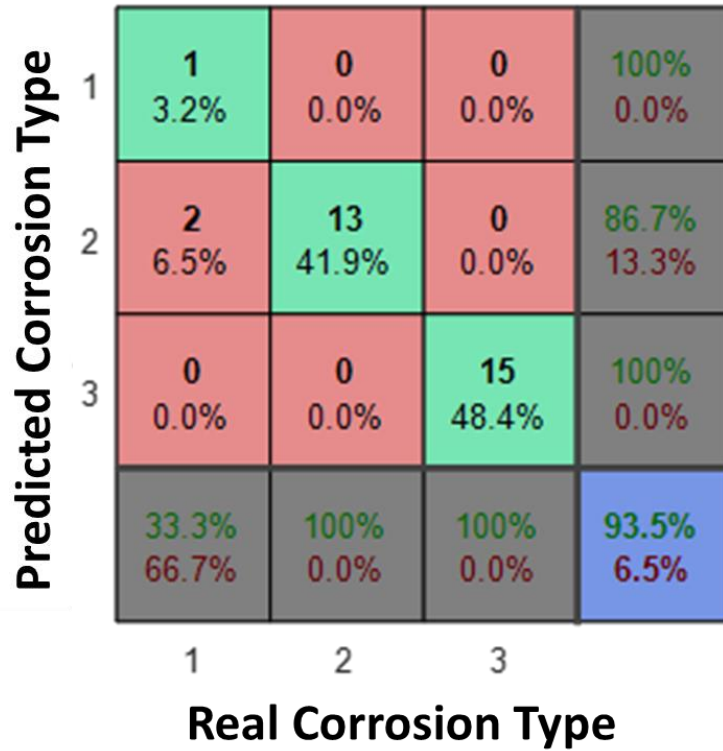


Fig. 10. Confusion matrix for the test data set consisting of 31 samples. Class 1 = passivation, Class 2 = pitting and Class 3 = uniform corrosion

The significance of the predictor variables in the classification model could also be assessed. This was done by training a multilayer perceptron as before, but with five predictor variables. Four of these were the recurrence variables, *D*, *E*, *R* and *L* and the fifth was a random variable with no predictive power. This variable was used to assess the statistical significance of the recurrence variables, as discussed in [29-31].

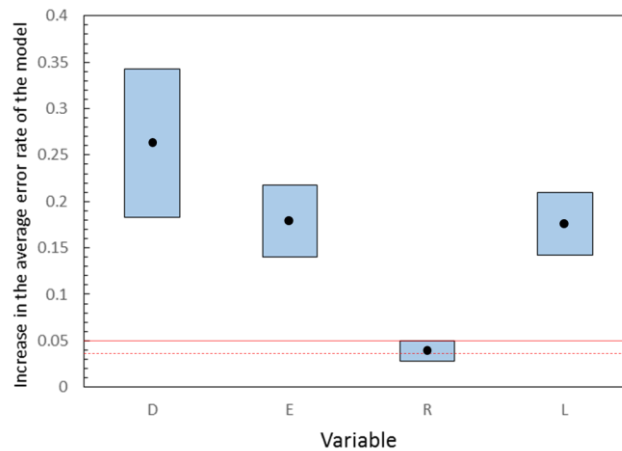


Fig. 11. Variable importance plot of the recurrence predictor variables in the three corrosion systems. The black dots indicate mean values, while the top and bottom of each box indicate the mean  $\pm$  three standard deviations. The horizontal dashed and solid lines indicate the 95% and 99% significance levels.

In the analysis, the classification model was interrogated by repeatedly randomizing each of the predictor variables, one at a time and observing the deterioration of the performance of the model in each case. The results are shown in the variable importance plot in Fig. 11.

The black dots in the figure show the average increase in the error rates of the model (in this case based on 30 runs or variable permutations). The top of each box shows the mean value plus three standard deviations, while the bottom of each box indicates the mean value minus three standard deviations. The broken and solid horizontal lines indicate the 95% and 99% significance levels of the increases in the errors, as determined by the dummy (random) variable that was included with the model and discussed in [31]. According to this plot, determinism ( $D$ ) played the most important role in discriminating between the three corrosion systems, while the recurrence rate ( $R$ ) did not play a statistically significant role.

#### 4.2. Application of corrosion monitoring scheme

The efficacy of the corrosion monitoring scheme is illustrated by the data in Fig. 12. It is essentially the same as Fig. 8, except that the data representing uniform corrosion or normal operating conditions have been delineated with a 95% control limit based on the use of a five-kernel Gaussian mixture model.

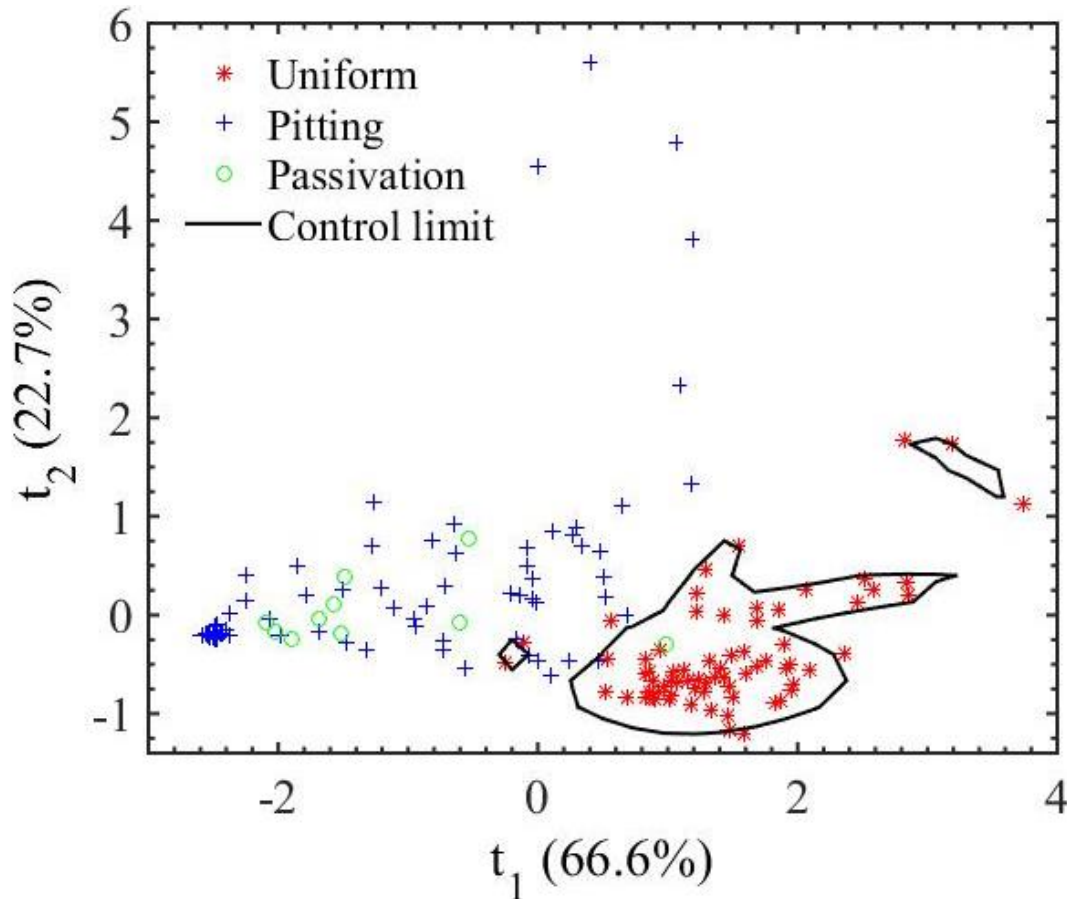


Fig. 12. Principal component score plot showing uniform corrosion (\*) as normal operating conditions and pitting corrosion (+) and passivation (o) as fault conditions. The 95% control limits enclosing the uniform corrosion data have been fitted with a 3-component Gaussian mixture model.

The first two principal components accounted for 89.3% of the variance of the variables and as indicated, this map can be used in the online monitoring of the corrosion of the system.



Some suggestions for other interested researchers towards using the proposed techniques:

- Firstly, the EN data should be recorded continuously under normal operation conditions, i.e. non-localized corrosion conditions, for a sufficient time depending on the actual corrosion system and sampling rate. This may require some experience from the user or familiarity with the corrosion system. Ideally, it is better to collect as many data as possible in this condition for computation of a reliable control limit.
- There are no strict rules regarding the choice of sampling rate. Users could customise their own sampling rates as they deem appropriate for the corrosion process under study. However, it is recommended to use a sampling rate of more than 1 Hz.
- After collection of the EN data, users need to decide how many data points there should be contained in each segment. In principle, the length of the time series segment should be sufficient to capture characteristic fluctuations or periodic behaviour in the time series. This can be estimated from the autocorrelation function or the average mutual information of the time series. The point where the autocorrelation function reaches its first minimum value or point of decorrelation, can be used as an indication of the minimum length of the time series segment (see for example [32-33], where the same principle is applied to the phase space embedding of time series data).
- After calculating the control limit for normal operation conditions according to the procedures proposed in this paper, users can then record EN data under different test conditions either continuously or intermittently. The sampling rate and the number of points in the new segment should be identical to those used for the normal operating conditions. If the data projected onto the map falls outside the normal operation condition control limit, a potential fault condition is flagged.

## 5. Conclusions

In this paper, the corrosion tests with carbon steel in aqueous media under three different conditions, viz., uniform corrosion, pitting and passivation, have indicated that the corrosion monitoring map generated based on the use of recurrence quantification and principal component analysis appears to be a promising approach for continuous corrosion monitoring. New electrochemical noise measurements can be projected onto the map continuously and will be flagged as potential fault conditions, if these projections are located outside control limits imposed on acceptable corrosion conditions, such as uniform corrosion.

It should be pointed out that the proposed method is not based on any fundamental physical or chemical models at work. The corrosive effects of the different environments to which the steel samples were exposed could not be studied quantitatively and more reliable corrosion data are needed.

The present monitoring map could not distinguish between passivation and pitting, and this needs to be improved, for example, by considering the use of additional recurrence quantification variables. However, these are comparatively minor issues, and the proposed method provides a principled approach to the automation of corrosion monitoring in industry.

## References

- [1] P. Bai, H. Zhao, S. Zheng, C. Chen, Initiation and developmental stages of steel corrosion in wet H<sub>2</sub>S environments, *Corros Sci*, 93 (2015) 109-119.

- [2] N.A. Hoog, M.J.J. Mayer, H. Miedema, R.M. Wagterveld, M. Saakes, J. Tuinstra, W. Olthuis, A. van den Berg, Stub resonators for online monitoring early stages of corrosion, *Sensors and Actuators B: Chemical*, 202 (2014) 1117-1136.
- [3] Y.J. Tan, Sensing localised corrosion by means of electrochemical noise detection and analysis, *Sensor Actuat B-Chem*, 139 (2009) 688-698.
- [4] A.M. Homborg, T. Tinga, X. Zhang, E.P.M. van Westing, P.J. Oonincx, G.M. Ferrari, J.H.W. de Wit, J.M.C. Mol, Transient analysis through Hilbert spectra of electrochemical noise signals for the identification of localized corrosion of stainless steel, *Electrochimica Acta*, 104 (2013) 84-93.
- [5] E.C. Rios, A.M. Zimer, P.C.D. Mendes, M.B.J. Freitas, E.V.R. de Castro, L.H. Mascaro, E.C. Pereira, Corrosion of AISI 1020 steel in crude oil studied by the electrochemical noise measurements, *Fuel*, 150 (2015) 325-333.
- [6] A.M. Homborg, E.P.M. van Westing, T. Tinga, X. Zhang, P.J. Oonincx, G.M. Ferrari, J.H.W. de Wit, J.M.C. Mol, Novel time-frequency characterization of electrochemical noise data in corrosion studies using Hilbert spectra, *Corros Sci*, 66 (2013) 97-110.
- [7] A. Legat, J. Osredkar, V. Kuhar, M. Leban, Detection of various types of corrosion processes by the chaotic analysis of electrochemical noise, *Mater Sci Forum*, 289-2 (1998) 807-811.
- [8] E. Cazares-Ibanez, G.A. Vazquez-Coutino, E. Garcia-Ochoa, Application of recurrence plots as a new tool in the analysis of electrochemical oscillations of copper, *J Electroanal Chem*, 583 (2005) 17-33.
- [9] Y.F. Cheng, M. Wilmott, J.L. Luo, The role of chloride ions in pitting of carbon steel studied by the statistical analysis of electrochemical noise, *Appl Surf Sci*, 152 (1999) 161-168.
- [10] G. Suresh, U.K. Mudali, Electrochemical noise analysis of pitting corrosion of type 304L stainless Steel, *Corrosion*, 70 (2014).
- [11] N. Upadhyay, M.G. Pujar, C.R. Das, C. Mallika, U.K. Mudali, Pitting corrosion studies on solution-annealed borated type 304L stainless steel using electrochemical noise technique, *Corrosion*, 70 (2014) 781-795.
- [12] C.A. Loto, Electrochemical noise measurement technique in corrosion research, *Int J Electrochem Sc*, 7 (2012) 9248-9270.
- [13] R. Zhao, S.Z. Song, Electrochemical potential noise analysis of 304 stainless steel weld zone exposed to alkaline solutions by different parameters, *Appl Mech Mater*, 331 (2013) 469-473.
- [14] A. Aballe, M. Bethencourt, F.J. Botana, M. Marcos, J.M. Sanchez-Amaya, Use of wavelets to study electrochemical noise transients, *Electrochim Acta*, 46 (2001) 2353-2361.
- [15] D.H. Xia, S.Z. Song, J.H. Wang, J.B. Shi, H.C. Bi, Z.M. Gao, Determination of corrosion types from electrochemical noise by phase space reconstruction theory, *Electrochem Commun*, 15 (2012) 88-92.
- [16] N. Acuna-Gonzalez, E. Garcia-Ochoa, J. Gonzalez-Sanchez, Assessment of the dynamics of corrosion fatigue crack initiation applying recurrence plots to the analysis of electrochemical noise data, *Int J Fatigue*, 30 (2008) 1211-1219.

- [17] C. Lopez-Melendez, E.M. Garcia-Ochoa, M.I. Flores-Zamora, R.G. Bautista-Margulis, C. Carreno-Gallardo, C.P.C. Morquecho, J.G. Chacon-Nava, A. Martinez-Villafane, Dynamic study of current fluctuations of nanostructured films, *Int J Electrochem Sc*, 7 (2012) 1160-1169.
- [18] L.S. Montalban, P. Henttu, R. Piche, Recurrence quantification analysis of electrochemical noise data during pit development, *Int J Bifurcat Chaos*, 17 (2007) 3725-3728.
- [19] E. Garcia-Ochoa, J. Gonzalez-Sanchez, N. Acuna, J. Euan, Analysis of the dynamics of Intergranular corrosion process of sensitised 304 stainless steel using recurrence plots, *J Appl Electrochem*, 39 (2009) 637-645.
- [20] Y.E. Yang, T. Zhang, Y.W. Shao, G.Z. Meng, F.H. Wang, Effect of hydrostatic pressure on the corrosion behaviour of Ni-Cr-Mo-V high strength steel, *Corros Sci*, 52 (2010) 2697-2706.
- [21] T. Sun, Z.Y. Wang, J. Li, T. Zhang, Effect of ultrasonic vibration solidification treatment on the corrosion behavior of AZ80 magnesium alloy, *Int J Electrochem Sc*, 8 (2013) 7298-7319.
- [22] E. Garcia-Ochoa, F. Corvo, Using recurrence plot to study the dynamics of reinforcement steel corrosion, *Prot Met Phys Chem+*, 51 (2015) 716-724.
- [23] J.B. Gao, H.Q. Cai, On the structures and quantification of recurrence plots, *Phys Lett A*, 270 (2000) 75-87.
- [24] N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems, *Phys Rep*, 438 (2007) 237-329.
- [25] G.B. Mindlin, R. Gilmore, Topological analysis and synthesis of chaotic time-series, *Physica D*, 58 (1992) 229-242.
- [26] J.P. Zbilut, C.L. Webber, Embeddings and delays as derived from quantification of recurrence plots, *Phys Lett A*, 171 (1992) 199-203.
- [27] J.P. Zbilut, J.M. Zaldivar-Comenges, F. Strozzi, Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data, *Phys Lett A*, 297 (2002) 173-181.
- [28] X. Li, G. Ouyang, X. Yao, X. Guan, Dynamical characteristics of pre-epileptic seizures in rats with recurrence quantification analysis, *Physics Letters A*, 333 (2004) 164-171.
- [29] C. Aldrich, L. Auret, Fault detection and diagnosis with random forest feature extraction and variable importance methods, in: *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 2010, pp. 79-86.
- [30] L. Auret, C. Aldrich, Empirical comparison of tree ensemble variable importance measures, *Chemometrics and Intelligent Laboratory Systems*, 105 (2011) 157-170.
- [31] L. Auret, C. Aldrich, Interpretation of nonlinear relationships between process variables by use of random forests, *Miner Eng*, 35 (2012) 27-42.
- [32] J.P. Barnard, C. Aldrich, M. Gerber, Identification of dynamic process systems with surrogate data methods. *AIChE J*, 47(9) (2001) 2064-2075.
- [33] C. Aldrich, B.C. Qi, P.J. Botha, Analysis of electrochemical noise with phase space methods. *Miner Eng*, 19(14) (2006) 1402-1409.