Electrochimica Acta 256 (2017) 337–347

Contents lists available at ScienceDirect

Electrochimica Acta

journal homepage: www.elsevier.com/locate/electacta

Research paper

# Analysis of electrochemical noise data by use of recurrence quantification analysis and machine learning methods

Y. Hou[a], C. Aldrich[a], K. Lepkova[b,*], L.L. Machuca[b], B. Kinsella[b]

[a] Department of Mining Engineering and Metallurgical Engineering, Western Australian School of Mines Curtin University, GPO Box U1987, Waterford, WA6152, Australia
[b] Curtin Corrosion Engineering Industry Centre, School of Chemical and Petroleum Engineering, Curtin University, Bentley, WA, 6102, Australia

A R T I C L E I N F O

*Article history:*
Received 21 July 2017
Received in revised form 25 September 2017
Accepted 27 September 2017
Available online 28 September 2017

*Keywords:*
Electrochemical noise
Recurrence quantification analysis
Linear discriminant analysis
Random forest
Corrosion type identification

A B S T R A C T

By use of recurrence quantification analysis (RQA), twelve features were extracted from the electrochemical noise signals generated by three types of corrosion: uniform, pitting and passivation. Machine learning methods, i.e. linear discriminant analysis (LDA) and random forests (RF), were used to identify the different corrosion types from those features. Both models gave satisfactory performance, but the RF model showed better prediction accuracy of 93% than the LDA model (88%). Furthermore, an estimation of the importance of the variables by use of the RF model suggested the RQA variables laminarity (LAM) and determinism (DET) played the most significant role with regard to identification of corrosion types. In addition, the comparison of noise resistance with the resistance obtained from EIS measurement showed that the noise resistance can be used for monitoring corrosion rate variations not only for uniform corrosion and passivation, but also for pitting.



## 1. Introduction

The concept of electrochemical noise (EN) was first introduced by Iverson [1] and Tyagai et al. [2] decades ago. Back then, the research focus was merely on potential fluctuations. Subsequently, in 1986, Eden et al. [3] further developed the noise measurement setup, which included two identical working electrodes (WE) connected by a zero resistance ammeter (ZRA), a reference electrode (RE) and a potentiometer, allowing the recording of current and potential noise simultaneously. Since then, EN with ZRA has been widely studied in the corrosion field, owing to its ease of setting up, non-destructiveness, non-intrusiveness and the ability to provide information on both the corrosion rate and type which other electrochemical techniques failed to offer [4–9].

Among various research areas, the identification of different types of corrosion has always attracted considerable interest from corrosion researchers and engineers. Numerous efforts have been made to extract discriminative features from collected EN data to indicate corrosion types. These features can be sourced from three kinds of analytical domains, namely the time domain, frequency domain and time-frequency domain. The primary feature obtained from frequency domain analysis is the roll-off slope of the power spectral density (PSD) plot [10]. A large number of indicators have been extracted from the time domain analysis of the EN data, including:

(i) Statistics, such as the standard deviation, kurtosis, skewness, and localization index of measurements [11];
(ii) The cumulative probability of corrosion events and Weibull probability plots from transient analysis [12];
(iii) Largest Lyapunov exponent and correlation dimension from chaotic analysis [13,14];
(iv) Recurrence rate, determinism, maxline, etc. from recurrence quantification analysis [15];
(v) Hausdorff exponent, Hurst exponent and spectral-power exponents from fractal analysis [16];
(vi) Energy distribution plot from wavelet analysis [17].
(vii) In addition, Homborg et al. [18] have published several papers on the time-frequency joint analysis by Hilbert-Huang transformation which showed good application prospect.

Recently, owing to the rapid development in machine learning techniques, new ideas have been proposed for the interpretation of

* Corresponding author at: Curtin Corrosion Engineering Industry Centre, School of Chemical and Petroleum Engineering, Curtin University, Bentley, WA, 6102, Australia.
*E-mail address:* k.lepkova@curtin.edu.au (K. Lepkova).

https://doi.org/10.1016/j.electacta.2017.09.169
0013-4686/

electrochemical noise. For example, Huang et al. [19–21] made use of cluster analysis of current and potential signals and LDA models to identify different pitting states in low carbon steel exposed to $NaHCO_3$ + NaCl solutions. Li et al. [22] have used features extracted from EN signals generated by uniform corrosion, pitting and passivation in 304 stainless steel as predictors in artificial neural network models designed to distinguish between different types of corrosion.

More recently, the authors [23] have proposed a formal methodology for the identification of localized corrosion in low carbon steel from uniform corrosion based on the use of recurrence quantification analysis (RQA). This statistical process control approach could be used to monitor corrosion in a continuous way with high reliability, although the discrimination between localized corrosion and passivation needs to be further improved. In the previous study, only four RQA variables were used to set up the corrosion monitoring scheme and the number of data in the passivation system was limited. In this study, to improve the separability of pitting and passivation, the EN signals were collected from the same corrosion systems with different sizes of electrodes, creating a larger database for the following model development. Meanwhile, an extended set of feature variables were extracted by recurrence quantification analysis (RQA) of the EN signals. LDA and RF models were subsequently developed with the extracted RQA variables as predictors to identify different corrosion types. In addition, noise resistance was compared with the corrosion resistance obtained from electrochemical impedance spectroscopy (EIS) to explore its usefulness as a corrosion rate indicator.

## 2. Methodology

### 2.1. Recurrence quantification analysis

The recurrence plot (RP) is a graphical representation of a square matrix as represented by Eq. (1):

$$R_{i,j} = H(\varepsilon - \parallel \boldsymbol{x}_i - \boldsymbol{x}_j \parallel), i,j = 1,2,\ldots,N \quad (1)$$

where $R_{i,j}$ is the $(i,j)^{th}$ point in the recurrence plot, $N$ is the number of points in the dataset, ε is a predefined threshold radius, $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ are the measured EN values at times $i$and $j$, and &tpcheck; refers to Euclidean distance between this pair of data points, and H represents the Heaviside function, which gives values of either zero or one, i.e. if the distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ falls within the threshold radius, then $R_{i,j} = 1$, otherwise, $R_{i,j} = 0$.

Recurrence quantification analysis (RQA) was used to extract twelve variables from the RPs, as shown in Table 1. An open source Matlab toolbox – Cross Recurrence Plot Toolbox – was used to do the calculations [24]. The threshold value $\varepsilon$ was determined as 0.02σ (σ is the standard deviation of the linear-detrended data segment), since it could best reveal the differences of the RQA variables in different corrosion systems. A detailed definition and description of the variables can be found in [23–26].

### 2.2. Linear discriminant analysis

Linear discriminant analysis (LDA) is commonly used for pattern recognition problems or simply used for dimensionality reduction before classification. The idea is similar to principal component analysis (PCA) in the sense of linear transformation.

**Table 1**
Recurrence quantification variables used in present work.

| Number | RQA variable | Equation |
|---|---|---|
| 1 | Recurrence rate | $RR = \frac{1}{N^2}\sum_{i,j=1}^{N} R_{i,j}(\varepsilon)$ |
| 2 | Determinism | $DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{i,j}^{N} R_{i,j}(\varepsilon)}, l_{min} = 2$<br>$P(l)$ – Histogram of the diagonal lines* of length $l$. |
| 3 | Averaged diagonal length | $L_{mean} = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=l_{min}}^{N} P(l)}$ |
| 4 | Length of longest diagonal line | $L_{max} = \max(\{l_i;\ i = 1,2,\ldots N_l\}$<br>$N_l$–Total number of diagonal lines. |
| 5 | Entropy of diagonal length (ENTR1) | $ENTR1 = -\sum_{l=l_{min}}^{N} p(l)\ln p(l)$<br>$p(l)$ – Probability distribution of diagonal lines. |
| 6 | Laminarity | $LAM = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)}, v_{min} = 2$<br>$P(v)$ – Histogram of vertical lines** of length $v$. |
| 7 | Trapping time | $TT = \frac{\sum_{v=vmin}^{N} vP(v)}{\sum_{v=v_{min}}^{N} P(v)}$ |
| 8 | Length of longest vertical line | $V_{max} = \max(\{v_i;\ i = 1,2,\ldots,N_v\}$<br>$Nv$ – Total number of vertical lines. |
| 9 | Recurrence times of 1st type | $\{RT1(i) = t_i - t_{i-1} \vert i = 1,2,\ldots\}$ |
| 10 | Recurrence times of 2nd type | $\{RT2(i) = t'_i - t'_{i-1} \vert i = 1,2,\ldots\}$ |
| 11 | Entropy of recurrence period density (ENTR2) | $ENTR2 = -\frac{1}{\ln(t_{max})}\cdot\sum_{t=1}^{t_{max}} P(t)\cdot\ln(P(t))$<br>$P(t) = R(t)/\sum_{k=1}^{t_{max}} R(k)$ – Recurrence time probability density.<br>$R(t)$ – The histogram of recurrence times.<br>$t_{max}$ – The maximum recurrence time. |
| 12 | Transitivity | $TRANS = \frac{\sum_{i,j,k=1}^{N} R_{i,j}.R_{j,k}.R_{k,j}}{\sum_{i,j,k=1}^{N} R_{i,j}\cdot R_{j,k}.R_{k,j}}$ |

Notes: * A diagonal line is the line formed with consecutive recurrent points and parallel to the 45° line of a square. ** A vertical line is the line parallel to the y axis and consisted of recurrent points. The length of a line is represented by the number of the recurrent points in the line.