Physica A 395 (2014) 332–339

Contents lists available at ScienceDirect

Physica A

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

Multi-*q* pattern classification of polarization curves

Ricardo Fabbri [a,*], Ivan N. Bastos [a], Francisco D. Moura Neto [a], Francisco J.P. Lopes [b], Wesley N. Gonçalves [c], Odemir M. Bruno [c]

[a] *Instituto Politécnico, Universidade do Estado do Rio de Janeiro, 28625-570 - Nova Friburgo, RJ, Brazil*

[b] *Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Sala G1-019 - Cidade Universitária, 21941-902 - Rio de Janeiro, RJ, Brazil*

[c] *Instituto de Física de São Carlos (IFSC), Universidade de São Paulo (USP), Av. Trabalhador São Carlense, 400, 13560-970 - São Carlos, SP, Brazil*

HIGHLIGHTS

- Multiple polarization curves of two stainless steels are experimentally acquired.
- The multi-*q* approach automatically classifies the profiles using Tsallis entropy.
- A success rate of 80–90% was achieved using only 2% of the original data.

ARTICLE INFO

Article history:
Received 11 May 2013
Received in revised form 30 July 2013
Available online 9 October 2013

Keywords:
Profile pattern classification
Polarization curve
Tsallis entropy
Multi-*q* pattern analysis

ABSTRACT

Several experimental measurements are expressed in the form of one-dimensional profiles, for which there is a scarcity of methodologies able to classify the pertinence of a given result to a specific group. The polarization curves that evaluate the corrosion kinetics of electrodes in corrosive media are applications where the behavior is chiefly analyzed from profiles. Polarization curves are indeed a classic method to determine the global kinetics of metallic electrodes, but the strong nonlinearity from different metals and alloys can overlap and the discrimination becomes a challenging problem. Moreover, even finding a typical curve from replicated tests requires subjective judgment. In this paper, we used the so-called multi-*q* approach based on the Tsallis statistics in a classification engine to separate the multiple polarization curve profiles of two stainless steels. We collected 48 experimental polarization curves in an aqueous chloride medium of two stainless steel types, with different resistance against localized corrosion. Multi-*q* pattern analysis was then carried out on a wide potential range, from cathodic up to anodic regions. An excellent classification rate was obtained, at a success rate of 90%, 80%, and 83% for low (cathodic), high (anodic), and both potential ranges, respectively, using only 2% of the original profile data. These results show the potential of the proposed approach towards efficient, robust, systematic and automatic classification of highly nonlinear profile curves.

1. Introduction

Several results of experimental techniques in materials science and engineering are available in the form of profiles. Despite the crucial importance of these types of data, the use of statistics on profiles is rare, as the adequate statistical

* Corresponding author. Tel.: +55 2182028624.
E-mail addresses: rfabbri@lems.brown.edu, rfabbri@iprj.uerj.br (R. Fabbri), inbastos@iprj.uerj.br (I.N. Bastos), fmoura@iprj.uerj.br (F.D.M. Neto), flopes@ufrj.br (F.J.P. Lopes), wnunes@ursa.ifsc.usp.br (W.N. Gonçalves), bruno@ifsc.usp.br (O.M. Bruno).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.09.048

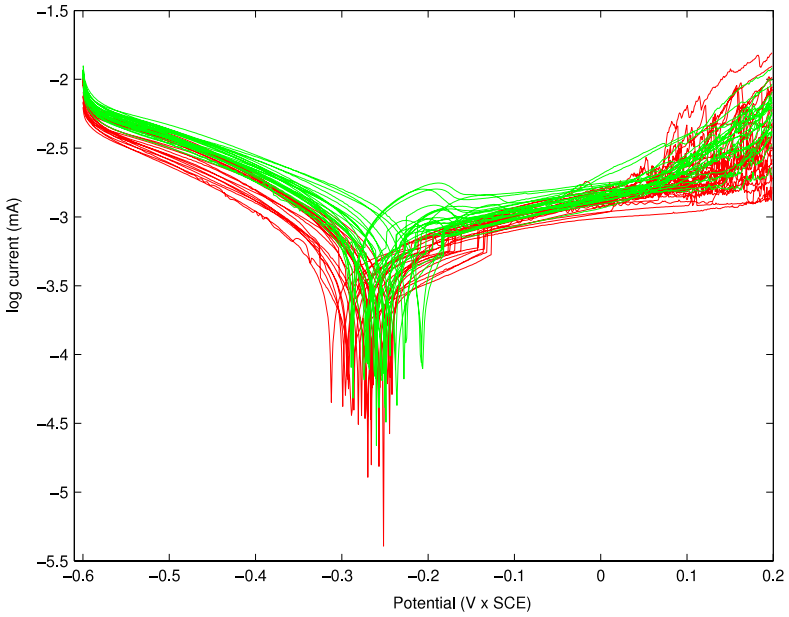

Fig. 1. 48 experimental polarization curves of stainless steels 304 (red) and 316 (green). Each curve consists of 800 sample points. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

techniques themselves are under development [1–3]. Useful applications of profile analysis include outlier detection and classification/clustering of groups of results, taking into account the inherent stochasticity of the experimental data.

Since the use of statistics for realistic profile datasets is currently somewhat scarce, the evaluation and comparison of experimental results is ordinarily performed by subjective criteria, e.g., after performing three to five runs of an experiment to identify representatives. In the case of profiles, a physical model with a related adjustment of parameters is a good indication of pertinence to a target group. In this situation, however, the analysis is mainly focused on the parameter values instead of profiles themselves. We here explore systematic data analysis techniques suited for *complex* or *highly nonlinear* profiles, i.e., when there is no simple model description or parameters to represent the entire profile curve.

In the fields of corrosion and materials science, there are numerous examples of data with complex profile shapes: the aforementioned polarization curves, impedance diagrams, voltammograms, electrochemical noise records and so on. Despite this profusion of profile data, the work of Strutt et al. [4] is a rare example of profile analysis from corrosion, albeit the studied profiles were related to the geometric shape of corroded interfaces, and not electrochemical data as in the present work.

Polarization curves, or, more generally, current–potential plots, have crucial importance in corrosion studies as well as in electrochemistry. They are essential for measuring the global kinetics of electrodes. A number of electrochemical parameters can be obtained from these tests, from charge transfer, passivation, corrosion current density, to mass transport properties. As expected, for any set of experimental results, dispersion always occurs. Even a stable reaction performed under stringent control of experimental procedures, such as the ferri/ferrocyanide redox reaction, can present significant scattering [5]. Moreover, the level of scattering in polarization curves naturally increases as the localized corrosion takes place, which exhibits random nature. Another challenging case is that the surface of commercial grades of steels, which has wide industrial interest, has inherent heterogeneity of their metallurgical aspects. Furthermore, certain stochastic phenomena such as pitting produce electrochemical current variations that produce overlap in the experimental curves, rendering them almost inseparable. However, these fluctuations can also be useful to describe the type and the intensity of corrosion attack [6,7]. Thus, for some corrosion resistant alloys (CRA) such as stainless steels, the evaluation of parameters from polarizations curves near pitting potential is an important procedure to classify their corrosion resistance.

Having had success in employing Tsallis statistics [8] to the classification of image patterns as well as other applications [9], we devise a similar multi-q approach based on such statistics to propose a method to classify highly nonlinear profiles from corrosion tests, such as those in Fig. 1. The primary goal of any classification method is to assign a class label (e.g., material type) to a given data sample (e.g., image or profile). In our case, the profiles are the polarization curves and the output label is one of two stainless steels. Despite having slightly different chemical compositions, the polarization of both materials is very similar, especially in the cathodic region and in the potential immediately past the corrosion potential region. This similarity produces overlapping curves which require a systematic approach to classify, an analysis that has not been routinely carried out in this field.

1.1. Formulation and notation

In statistical mechanics, the concept of entropy is used to characterize a system based on its distribution of states, which can be determined by the system energy level. If one has less information about the system then at many states the system can still be characterized. This idea produces a direct link between the concept of entropy and the lack of information of a system. From an information-theoretical point of view, the entropy represents how close a given probability distribution is to the uniform distribution. Since the kinetic aspects of polarization curves are related to the allowed microstates of the system, an approach based on the concept of entropy is then a natural choice. In addition, the random nature of this kind of experimental data is caused, among other factors, by the joint diversity of the microstates, which is directly related to the different chemical composition of the materials and the superficial conditions, chiefly near the pitting potential. Because such chemical compositions are expected to produce both short- and long-range interaction in the material [10], using the Tsallis entropy, which introduces a new parameter q to help model these interactions across different scales, motivated our approach.

Formally, the Tsallis entropy S_q for a given profile curve and parameter q is defined as

$$S_q = \frac{1 - \sum p_i^q}{q - 1}, \tag{1}$$

where p_i, $i = 1, \ldots, N$ is an estimated probability distribution of desired pattern properties of a signal. In this paper we use the histogram of measurement values in the quantized profile to represent pattern properties, where the profile amplitude was quantized in N levels of the same size. The multi-q analysis consists in expressing the statistical aspects of this histogram as a vector of entropy values for $n < N$ different values of q, e.g., 0.1, 0.2, $\ldots$, q_n. This is called the multi-q vector of the profile, which is much like an entropic spectrum. In practice, best-performing values for n are typically very small (e.g., 10), two to three orders of magnitude less than the number of raw data (profile) sample points (e.g., 800 in the present paper).

2. Materials and methods

2.1. Setup for generating polarization curve data

Two commercial austenitic stainless steels (UNS S30400 and UNS S31600) were used, hereafter named 304 and 316 for simplicity. Bar samples were coated with Teflon™ and an area of 0.20 cm^2 was kept accessible to the electrolyte. Before each test, the samples were grit with emery paper up to #600, and subsequently washed with distilled water and dried with hot air. Additionally, after the polarization tests, the surfaces were analyzed with optical microscopy to detect the occurrence of crevice. All the presented curves were free of observed crevice. These two stainless steels have austenitic microstructure, but their main difference being the molybdenum content. The 316 alloy has *circa* 2.5% Mo that improves its pitting resistance in chloride media. The 304 has only a residual content of molybdenum. The aerated aqueous electrolyte was 3.5% in mass of NaCl with temperature kept within 25.0 ± 0.1 °C.

All experiments were performed inside a Faraday cage in order to avoid spurious electric noise. After the immersion of the sample in the solution, the assumed steady-state condition is attained after an hour at open-circuit status. After this step, potentiodynamic polarization was applied from -0.6 up to 0.2 V $\times$ SCE under a potential rate of 1.0 mV s^{-1}. All values of the potential were measured with a saturated calomel electrode (SCE) reference. A platinum wire was used as the counter-electrode. In order to obtain sufficient data, 24 curves were performed for each alloy. Each polarization curve consists of a profile of potential versus current (or, equivalently for a constant area, current density) as a sequence of 800 data points.

2.2. Classification approach

We performed profile pattern classification experiments whose basic goal is to assign a steel class label to any given profile curve for which the steel type is previously unknown. The classification approach is data-driven and supervised, in that the model is not specific for stainless steel polarization curves, but it is instead learned and adjusted/trained from previously seen data [9]. In supervised classification, the classifier is trained from a set of samples that are known to belong to the classes (*a priori* knowledge) [11]. The classifier is then validated by a different set of so-called 'test' samples for which we know the ground-truth steel classes but have not been used to train the classifier. For each test sample the classifier outputs a likely steel type and we generate a success rate based on how well these match to ground-truth.

We employed a number of minor variants of our so-called multi-q classification approach to classify the signals, which consists in using Tsallis statistics [9] on the profile data followed by the naive Bayes classifier [11]. We have also used classical principal component analysis (PCA) [12] as an optional step. PCA consists in expressing each original profile (800 points) in a rotated and translated coordinate frame such that the entire dataset variance along each direction is maximum. The desired dimension may be reduced to less than 800 by keeping only the coordinates of the highest variance. The goal of this reduction is threefold: representational efficiency, visualization purposes, and for automatically filtering out the low-variance aspects of data that may otherwise add up to hinder the classification performance.

Table 1
Classification rates for a number of different methods with variable q and also classifying directly over the raw data. The bold line highlights our best proposed result, which differs from the first line in that it is automatic and less dependent on the application domain, as explained in the text.

Method	Classification rate (%)		
	Full potential	Low potential	High potential
Tsallis $q = 1$	83	83	65
Tsallis $q = 0.1$	73	65	60
Multi-q, $q = 0.1, 0.2, \ldots, 1.0$	73	69	69
Multi-q, $\boldsymbol{q} = \mathbf{0.1}, \mathbf{0.2}, \ldots, \mathbf{2.0}$	**83**	**90**	**80**
Naïve Bayes on all 800 points	81	75	73

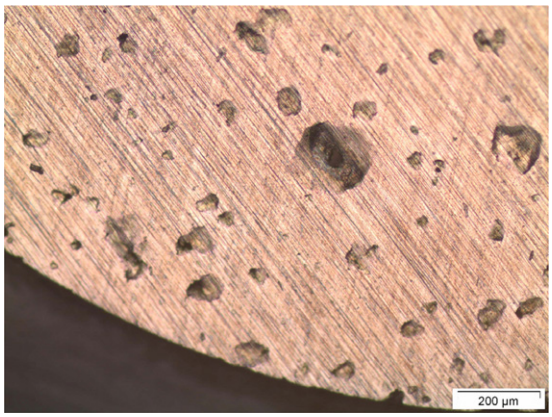

Fig. 2. Pitting corrosion after a polarization curve for the 304 stainless steel with original magnification of 100×.

In order to objectively evaluate the performance of the classification approach, we use the stratified 10-fold cross-validation scheme [13]. In this scheme, the profile dataset is randomly divided into 10 folds, considering that each fold contains two profiles, one for each steel class. At each run of this scheme, the classifier is trained using all but one fold and then is evaluated on how it classifies the samples from the separated fold. This process is repeated such that each fold is used once as validation. The performance is averaged, generating a single number for classification rate which represents the overall proportion of success over all runs.

3. Results and discussion

Fig. 1 shows the 48 polarization curves of stainless steels 304 and 316 swept from cathodic to anodic potential regions. At an actual corrosion situation, cathodic and anodic processes occur on the same metallic surface, under equal intensity. Thus, to obtain a global view of the electrode kinectics we normally apply a potential range that covers both regions in a three-electrode cell. Under cathodic potential, there is a predominance of cathodic process on work-electrode, that is a fixation of electrons, which in the present case is the reduction of dissolved oxygen gas. On the other hand, under anodic polarization, the net current provokes the corrosion of alloy, an oxidation process. At low anodic potential the corrosion process tends to be uniform, and at high potential it can present a localized corrosion such as pitting that has essentially a stochastic nature. Between these regions, the net current approaches null value, thus the logarithm of current attains the lowest level, defining the corrosion potential. The cathodic process occurs to the left and the anodic to the right side of the corrosion potential. The polarization of two stainless steels in aerated solution bears certain similarities, at least in the cathodic region and for potentials just beyond the corrosion potential. Therefore, an overlapping of curves can be observed, despite their different chemical composition. However, close to potential where there is a higher susceptibility to pitting, the difference of resistance appears together with some natural dispersion. Nevertheless, as in Fig. 1, we can see a significant overlap, when a large number of curves is considered. Therefore, at first sight, one cannot easily assign a given curve to the related steel type. Moreover, 304 exhibits a higher scattering of current above *circa* 0.0 V × SCE. In that region, a highly stochastic event occurs: the pitting attack, as in Fig. 2. The 316 steel type is specially designed to resist this localized corrosion.

Mild evidence of the high resistance of 316 can be observed at a corrosion potentials, with mean value *circa* −250 mV × SCE, being slightly higher for 316 than 304, which fares close to −280 mV × SCE. However, the main stochastic phenomenon occurs for higher potentials, where the randomness aspect is strong due to pitting. The main interest in stainless steel corrosion behavior is limited to this high potential range. In this case, for a small number of experiment runs, as in ordinary practice, it is possible to classify erroneously one steel instead of the other as the most pitting resistant, in a false-positive, or, similarly, false-negative evaluation, or to not obtain the true representative result. Thus, a judgment bias can take place due to the small number of observations. There is no generally accepted theory on how exactly these stochastic fluctuations

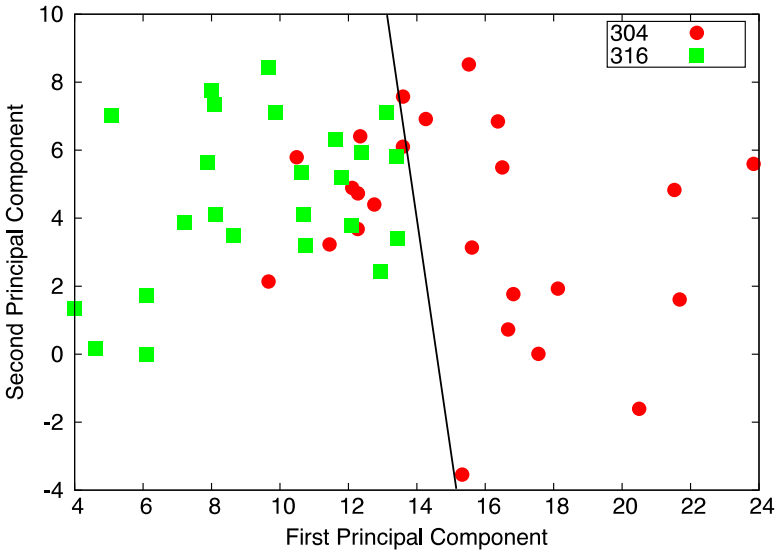

Fig. 3. *Results from a direct classical approach, full potential range.* First and second principal components for a classic principal component analysis on raw profile data, together with a plausible classification boundary.

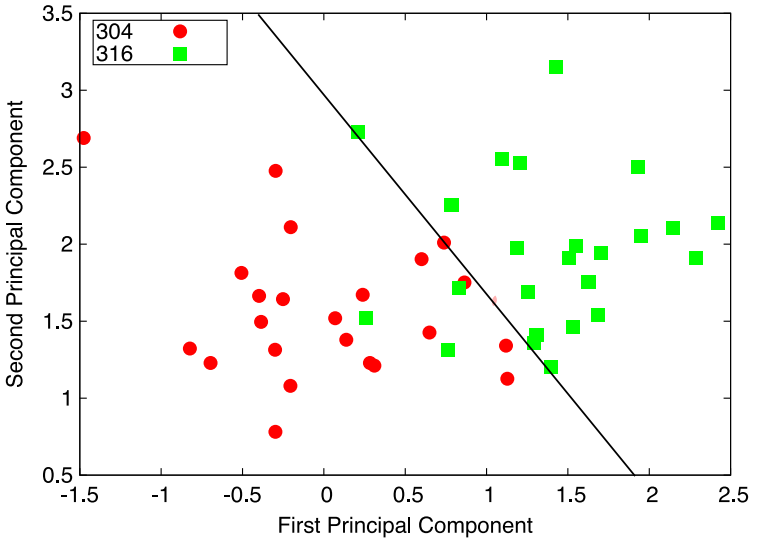

Fig. 4. *Results from the proposed approach, full potential range.* The first and second principal components for the proposed Tsallis entropy approach with q ranging from 0.1 to 2.0 in steps of 0.1, together with a plausible classification boundary. Even with 2 dimensions per profile for ease of visualization, only three 316 curves (green squares) are confused with the 304 ones (red circles).

occur [14], despite their huge industrial importance. To properly study this fact, the advisable number of test runs is far beyond the usual standard of subjective procedures which rely on only 3–4 runs.

In addition to the aforementioned stochastic events, in the region of interest (pitting) the current is of central importance. In this case, higher currents mean more severe corrosion. At high potential, *circa* 0.2 V × SCE, the so-called random localized attack occurs, and the possibility of wrong interpretation among the curves is further aggravated.

Fig. 3 depicts the first and second principal component of the obtained polarization curves by using classical principal component analysis directly on raw profile data. A reasonable boundary separating the two sets of polarization curves was obtained, with, however, some degree of misclassification which may be aggravated if a larger number of experiments were available. Moreover, 316 shows a smaller variance (closer principal components) than 304. The relatively small cluster extent of 316 suggests less susceptibility to stochastic corrosion of pits, which is consistent with the literature [4]. A plausible classification boundary is drawn with the sole purpose of guiding the eye to the clusters.

As an attempt to improve this separation, we performed so-called multi-q pattern analysis, as shown in Fig. 4 for only two principal components for visualization purposes (i.e., using two classification dimensions per profile). In this approach, with q varying from 0.1 to 1.0, a very good separation was found.

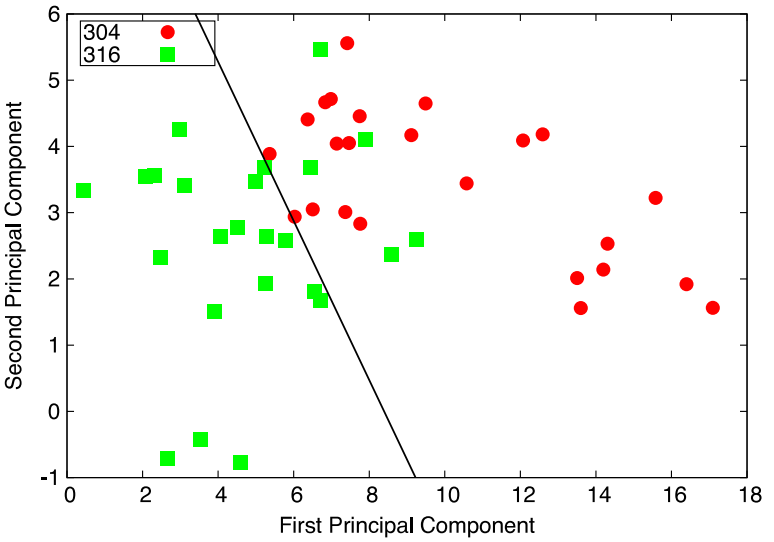

Fig. 5. *Results from a direct classical approach, low potential range.* First and second principal components for a classic principal component analysis on raw profile data, together with a plausible classification boundary.

Table 1 depicts a more comprehensive summary on the score of several classification tests. The Tsallis entropy approach, even for as few as two principal components, enabled classifying most polarization curves correctly. Moreover, the profile data can be efficiently represented by at most 20 Tsallis entropy values, as opposed to all 800 points. As the table shows, it was possible to attain the best classification rate using only a single value of $q = 1$. However, using the multi-q vector (bold line) the optimal selection of this parameter is fully automatic for a given dataset with no observed decrease in the classification rate.

3.1. Classifying low and high potential ranges separately

We also performed classification experiments to observe the behavior of the proposed approach on different sections of the profiles which have distinct statistical properties. Classification results using only the first half of the profiles, with potentials inferior than -0.2 V $\times$ SCE, are shown in Figs. 5 and 6. These potentials are roughly related to the predominance of cathodic (-0.6 to -0.2 V $\times$ SCE) and anodic (-0.2 to 0.2 V $\times$ SCE) processes on both steels.

Similar results for potentials above -0.2 V $\times$ SCE are shown in Figs. 7 and 8. More comprehensive results are shown in the last two columns of Table 1. For both low and high potentials, the proposed method achieved very good classification rates, having an edge of 7%–15% over the best success rate without using Tsallis statistics. The results also indicate that the profiles are significantly more challenging to distinguish in the high potential range, and can be confirmed by the visual inspection of Fig. 1, just where its behavior is more meaningful for technological use.

3.2. Further discussion

With the present methodology, the classification of strongly nonlinear curves with a systematic procedure is made possible. The success in classifying a profile that originally contains 800 points using only a couple or so dimensions could be understood as a signal expressed in a more suitable statistical representation, compressed and filtered with a low sampling frequency. In this case, a natural question is the possibility of aliasing [15] in the data. However, the first and second principal components are not subsampled data from the set, but parameters derived from the statistics, specifically the multi-q vector of Tsallis entropies, optionally followed by the projection of maximum variance given by PCA.

It is clear that the best classification is attained with a wider multi-q vector with q ranging from 0.1 to 2.0, instead of a fixed q. Moreover, even though the corrosion behavior at high potential is of great interest, the best classification rate is obtained with the entire polarization curves, and not only at high potentials (anodic process). In this case, from Table 1, the values are 83% for the entire curve against 80% for the anodic half of the potential. Thus, for the purpose of a reliable classification, polarization curves performed from cathodic to anodic potentials convey more information than a single range.

4. Conclusion

A relatively high number of polarization curves were generated through the experiments performed for two austenitic stainless steels UNS S30400 and UNS S31600 in chloride medium at 25 °C. For the entire potential range of the polarization curves a degree of overlap of the experimental curves was observed. The data samples were subsequently tested with a

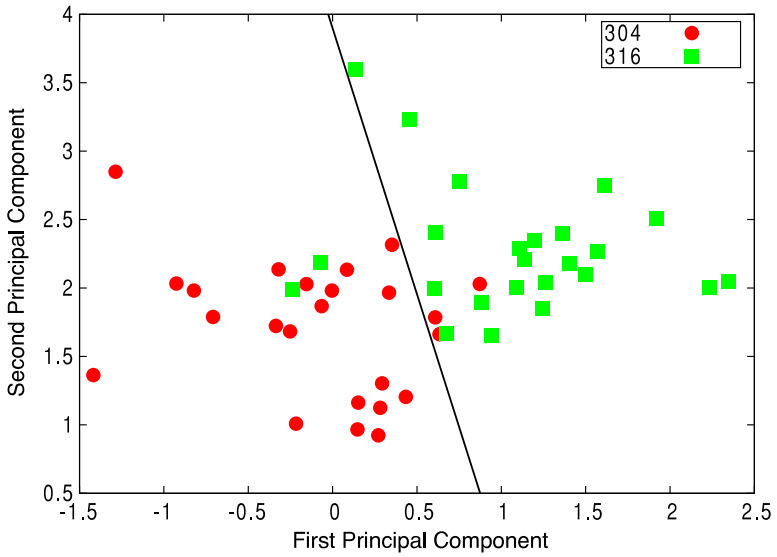

Fig. 6. *Results from the proposed approach, low potential range.* The first and second principal components for the proposed Tsallis entropy approach with q ranging from 0.1 to 2.0 in steps of 0.1, together with a plausible classification boundary.

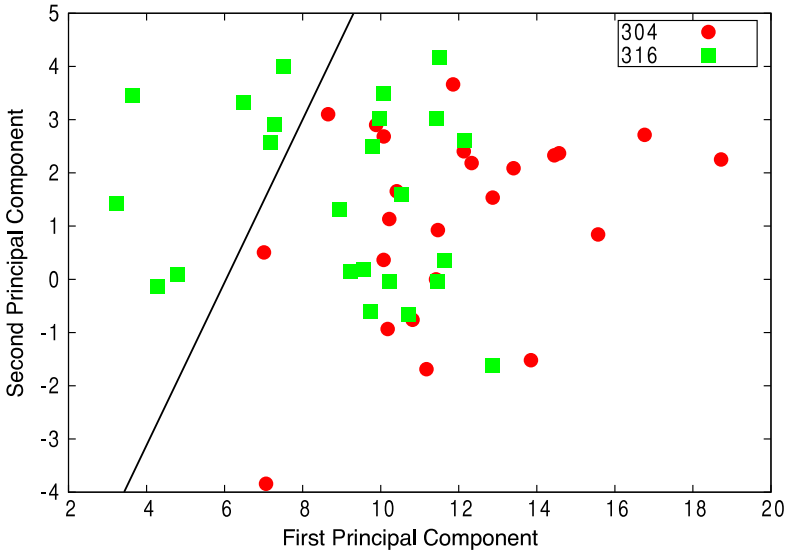

Fig. 7. *Results from a direct classical approach, high potential range.* First and second principal components for a classic principal component analysis on the raw profile data, together with a plausible classification boundary.

Tsallis-based statistics in order to classify the profiles for each steel. For the case where a multi-q pattern analysis was carried out on a complete potential range, from cathodic up to anodic regions, an excellent classification rate was obtained, using only 2% of the original profile data. When automatically classifying only the low potential range, the approach attained 90% correct classification, as opposed to 80% at high potentials. However, the industrial interest for these steels lies on the high potential range. These results together with the inspection of a series of scatter plots using principal component analysis demonstrate the capacity of the proposed approach towards efficient, robust, systematic and automatic classification of highly nonlinear profile curves, which can be used for analyzing the pattern properties as well as for outlier detection.

With the promising results obtained in the classification of polarization curves, an extension to study further types of signals such as time records of electrochemical noise and comparison with related procedures already available in the literature is ongoing in our research group. In these aspects, the Tsallis statistics can be very useful to identify the corrosion status from a continuous recording of on-line monitoring. In this scenario, the noisy potential and current related to the corrosion process would be recorded in order to detect changes that affect the corrosion attack.

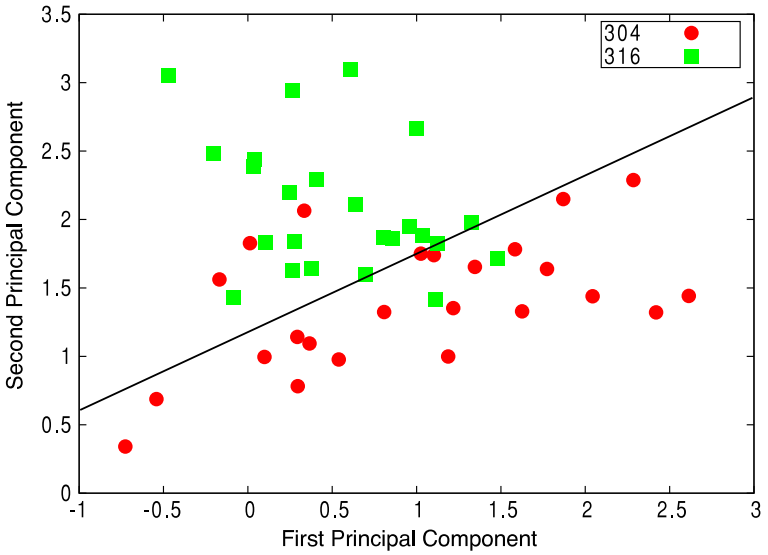

Fig. 8. *Results from the proposed approach, high potential range.* The first and second principal components for the proposed Tsallis entropy approach with q ranging from 0.1 to 2.0 in steps of 0.1, together with a plausible classification boundary.

Acknowledgments

The authors thank Mr. Denisar Ismério for the electrochemical tests. The authors acknowledge the financial support of the Brazilian agencies CNPq, FAPERJ and FAPESP. R.F. acknowledges the support from FAPERJ/Brazil (Grant #111.852/2012), W.N.G. from FAPESP (Grant #2010/08614-0), O.M.B. from CNPq (Grant #308449/2010-0 and #473893/2010-0) and FAPESP (Grant #2011/01523-1).

References

[1] H. Zhang, S. Albin, Detecting outliers in complex profiles using a χ^2 control chart method, IIE Transactions 41 (2009) 335–345.
[2] M. De Magalhães, A. Costa, F. Moura Neto, A hierarchy of adaptive control charts, International Journal of Production Economics 119 (2) (2009) 271–283.
[3] M. De Magalhães, F. Moura Neto, A Laplacian spectral method in phase I analysis of profiles, Applied Stochastic Models in Business and Industry 28 (2012) 251–263.
[4] J. Strutt, J. Nicholls, B. Barbier, The prediction of corrosion by statistical analysis of corrosion profiles, Corrosion Science 25 (5) (1985) 305–315.
[5] E. Tourwé, T. Breugelmans, R. Pintelon, A. Hubin, Extraction of a quantitative reaction mechanism from linear sweep voltammograms obtained on a rotating disk electrode. Part II: application to the redoxcouple, Journal of Electroanalytical Chemistry 609 (1) (2007) 1–7.
[6] I.N. Bastos, R.P. Nogueira, Electrochemical noise characterization of heat-treated superduplex stainless steel, Materials Chemistry and Physics 112 (2) (2008) 645–650.
[7] H.S. Klapper, J. Goellner, A. Heyn, The influence of the cathodic process on the interpretation of electrochemical noise signals arising from pitting corrosion of stainless steels, Corrosion Science 52 (4) (2010) 1362–1372.
[8] C. Tsallis, Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World, Springer, 2009, URL http://books.google.com.br/books?id=K7xOhNeGk6kC.
[9] R. Fabbri, W. Gonçalves, F. Lopes, O. Bruno, Multi-q pattern analysis: a case study in image classification, Physica A. Statistical Mechanics and its Applications 391 (19) (2012) 4487–4496. http://dx.doi.org/10.1016/j.physa.2012.05.001.
[10] C. Punckt, M. Bölscher, H. Rotermund, A. Mikhailov, L. Organ, N. Budiansky, J. Scully, J. Hudson, Sudden onset of pitting corrosion on stainless steel as a critical phenomenon, Science 305 (5687) (2004) 1133–1136.
[11] R. Duda, P. Hart, D. Stork, Pattern Classification, Vol. 2, Wiley, New York, 2001.
[12] L. da Fontoura Costa, R.M. Cesar Jr., Shape Analysis and Classification: Theory and Practice, CRC Press, 2010.
[13] I. Witten, E. Frank, M. Hall, Data Mining: Practical Machine Learning Tools and Techniques, third ed., Morgan Kaufmann, 2011.
[14] L. Balazs, J. Gouyet, Two-dimensional pitting corrosion of aluminium thin layers, Physica A. Statistical Mechanics and its Applications 217 (3) (1995) 319–338.
[15] I. Bastos, F. Huet, R. Nogueira, P. Rousseau, Influence of aliasing in time and frequency electrochemical noise measurements, Journal of the Electrochemical Society 147 (2) (2000) 671–677.