

Novel Data-Driven Methodologies for Parameter Estimation and Interpretation of Fuel Cells Performance

Vitor V. Lopes

Energy Systems Modeling and Optimization Unit
National Laboratory for Energy and Geology (LNEG)
Lisbon, Portugal
vitor.lopes@lneg.pt

Augusto Q. Novais

Energy Systems Modeling and Optimization Unit
National Laboratory for Energy and Geology (LNEG)
Lisbon, Portugal
augusto.novais@lneg.pt

Carmen M. Rangel

Hydrogen and Fuel Cell Unit
National Laboratory for Energy and Geology (LNEG)
Lisbon, Portugal
carmen.rangel@lneg.pt

Abstract— Fuel cell based power generation systems are expected to become more widespread in the near future. Stationary fuel cells may be used as an uninterruptible or back-up power supply, or to supply micro-grids. In particular, proton exchange membrane fuel cells (PEMFC) are an attractive technology due to its high energy density, rigid and simple structure, low operating temperature and fast start-up characteristics.

The power quality assessment of fuel cells as a viable power sources requires a good understanding of the fuel cell performance characteristics. This paper presents two novel data-driven methodologies for the identification of the main steady state (polarization curve) and the dynamic (impedance response) characteristics for fuel-cells allowing the development of rapid, accurate and empirical models based on the experimental data. M-NMF is a modified non-negative matrix factorization technique developed for the analysis of polarization curve data that allows to identify the three main contributions for the fuel-cell power degradation, while for impedance spectroscopy data, this paper proposes the use of fractional order transfer functions (FC-FOTC) to describe the main dynamic modes present in the fuel-cell.

A brief description of these two approaches is presented, together with the analysis of a real experimental dataset obtained from a 12W open cathode PEMFC stack to illustrate their potential and scope. While the former is instrumental for the deconvolution of the fuel cell polarization curves into its major components, the latter enables the estimation of the parameters related to the inherent transport and kinetic phenomena, thus opening the way, in both cases, for the interpretation of the effect of the operating conditions on the relative dominance and magnitude of these components and phenomena.

Fuel-cells performance; Data-driven modeling; Power quality.

I. INTRODUCTION

Polymer electrolyte membrane fuel cells (PEMFC) are regarded as a viable fuel cell technology alternative with extensive applications, such as stationary power plants and portable power sources. A PEMFC comprises anodic and cathodic regions and a polymer membrane electrolyte.

The characterization of the fuel cells steady-state performance is often accomplished by the use of polarization curves, which depict the fuel cell overall losses against the fuel cell current intensity, under given operation conditions [1]. Their interpretation is usually performed with the help of models, which are instrumental in extracting meaningful information about cell performance. However the simplest empirical models do not explicitly take into account the cell spatial dimensions and are based on direct fit of polarization experimental data to non-linear polynomial expressions [2], aiming at the modelling of the three major over-potential effects present in the fuel cell, i.e. activation, ohmic and concentration losses [3].

In addition, accurate dynamic models for the PEMFC electrical behaviour allow the development of efficient controllers for a quick load-following response, ensuring fuel economy and enhanced PEMFC lifetime.

Analyses employing mathematical factorization techniques, such as the proposed M-NMF, are commonly designated as exploratory data analyses (EDA) or data mining, and aim at the identification of significant patterns on given data-sets prior to the application of other more sophisticated data analysis techniques [4]. EDA are already used in electrochemistry and successful applications are described in the Brown and Bear review [5]. In a more recent review, Ni and Kokot [6] showed

that such well established techniques as polarography and voltammetry can also benefit significantly from their application.

On the other hand, data on dominant process dynamics present in the PEMFC such as interfacial reaction kinetics, electrolyte conductance, oxygen diffusion and membrane water balance [7,8], are commonly acquired through the application of Electrochemical Impedance Spectroscopy (EIS) techniques. Impedance variations versus frequency of the AC source [9] are thus obtained, which are normally represented in a Bode or Nyquist plot and correspond to the frequency response analysis of the PEMFC at given potential and operating conditions.

Frequency response data is usually modeled by adjusting a transfer function (TF) model to the observed input-output behavior. In such cases EIS data are commonly modeled by dynamic models based on transient equivalent electric circuits consisting of resistor and capacitor elements, where the electrode-electrolyte interface is represented by the Randles equivalent circuit [10]. A more accurate description of such phenomena is possible by introducing fractional-order elements, a.k.a. fractance energy storage device [11], which provided the motivation for the proposed FC-FOTF modeling approach.

Although fractional calculus is not new, its application to real world problems is very recent. A specific property of fractional derivatives, which makes them suitable for representing some physical and electrochemical phenomena, is that they do not depend only on the local properties of the function. It has been therefore successfully applied to capacitor theory [12], inhomogeneous porous materials [13], fractional order diffusion [14] and electrode-electrolyte interfaces [15]. A common practice to improve accuracy in the development of these models, which was applied in the development of the proposed FC-FOTF, is the use of modified dynamic elements, such as constant phase or the Warburg diffusional element [16].

These two approaches can find application in the design of power conditioning units (PCU) for the development of fuel cell stacks which require to be integrated in a power grid. Thus, the M-NMF data factorization technique apart from constituting a monitoring tool, also allows to compute performance losses and adjust actual fuel cell feed streams to meet power requirements. The FC-FOTF, on the other hand, can be used to predict the impact of the PCU ripple current on the power output reduction and, thus, furthering both an improved PCU design and power quality.

II. MODELING APPROACHES

A. M-NMF Outline: Polarization curve data (steady-state)

Data factorization methods are mathematical methods that aim at finding common patterns on given data-sets and reorganizing information therein contained according to such patterns. One of the most common methods is the principal component analysis (PCA), which computes the optimal base set that can be used to approximate high-dimensional data in a least squares sense [4]. Another widely used technique is the independent component analysis (ICA) which also aims at

finding a basis, but where the projected data statistical dependence is minimized [17]. ICA and PCA are two similar EDA techniques in the sense that both aim at extracting common patterns observed from a given data-set, but which differ in the base vectors properties. Unlike PCA, ICA base vectors are neither orthogonal nor ranked.

For some applications, these techniques lead to new entities and structures that are difficult to interpret in terms of physical meaning. To address this issue, several researchers suggest that the search for a representative basis should be more confined. Thus, new data factorization methods such as non-negative matrix factorization (NMF) were developed to facilitate a rapid qualitative and quantitative data analysis and interpretation [18,19].

Specifically, this work proposes a new NMF, M-NMF, to extract base vectors with physical meaning and relevant for the analysis and interpretation of polarization curve data-sets, not requiring the pre-specification of equation models [20].

Taking into account the following mathematical properties observed in the literature [2,3]: i) over-potential losses are divided only in three components and are all positive; ii) activation over-potential is modeled as an increasing concave function of the current; iii) concentration over-potential is modeled as an increasing convex function of the current; iv) ohmic over-potential has an increasing linear dependence of the current; the M-NMF is described by the following nonlinear optimization problem formulation:

$$\begin{aligned}
\min \quad & \|X - WH\|_F^2 + \lambda \|G\|_1^2 \quad (3) \\
\text{s.t.} \quad & W \geq \mathbf{0} \\
& H \geq \mathbf{0} \\
& G \geq \mathbf{0} \\
& \forall_{j,l=0\dots n-1,v=1} w_{l,v} \leq w_{j,v} + g_{j,v}(i_l - i_j) \\
& \forall_{j,l=0\dots n-1,v=2,3} w_{l,v} \geq w_{j,v} + g_{j,v}(i_l - i_j)
\end{aligned}$$

where W is the three dimensional rank basis matrix composed by the $w_{j,v}$ elements, G is composed by the $g_{j,v}$ elements and represents the sub-gradients of the matrix of W , and H is the three dimensional coefficients matrix. The n-by-m data-set matrix X is constructed by the column-wise concatenation of the different polarization curves $E(i)$:

$$X = \begin{bmatrix} E_0(i_0) & \cdots & E_{m-1}(i_0) \\ \vdots & \ddots & \vdots \\ E_0(i_{n-1}) & \cdots & E_{m-1}(i_{n-1}) \end{bmatrix} \quad (4)$$

where the $i_0=0$ to i_{n-1} are the n values for the current intensity common to all polarization curves and sorted by increasing order.

1) Numeric method for solving the nonlinear optimization

The numerical method used to solve this problem is based upon the alternating least squares method, i.e. it alternately fixes one matrix and improves the other:

$$\begin{aligned} \text{Find } \mathbf{W}_{k+1} : f(\mathbf{W}_{k+1}, \mathbf{H}_k) &\leq f(\mathbf{W}_k, \mathbf{H}_k) \\ \text{Find } \mathbf{H}_{k+1} : f(\mathbf{W}_{k+1}, \mathbf{H}_{k+1}) &\leq f(\mathbf{W}_{k+1}, \mathbf{H}_k) \end{aligned} \quad (5)$$

where f represents the objective function. The procedure is called the "block coordinate descent" method, where sequentially one block of variables is minimized under the respective constraints while the remaining variables are fixed. For this particular case, there are only two block variables (\mathbf{W} and \mathbf{H}) and, hence, when one variable block is fixed, the resultant constrained penalized least square sub-problem admits one single unique solution.

The convergence of the proposed algorithm is not a trivial result. However, for the case of two blocks any limit point of the sequence $\{W_k, H_k\}$ generated by the algorithm is a stationary point of the original optimization problem.

B. FC-FOTF Outline: Impedance spectroscopy data (dynamic response)

FC-FOTC represents a new model formulation based on FOTF, i.e. a generalization of FOTF for modelling PEMFC impedance data, consisting of a parameterization through the use of a modal form composed by elementary fractional order terms of 1st and 2nd kind and an identification method to estimate the model parameters based on a nonlinear least squares optimization method.

The fractional order differential equation for a given system, using the Laplace transform, can be represented by $Z(s)$ through a modal form based on a partial fraction expansion, composed by transfer functions of first and second kind:

$$Z(s) = R_0 + \sum_{k=1}^p \frac{R_k}{\tau_k s^{\alpha_k + 1}} + \sum_{k=p+1}^{p+q} \frac{R_k(1 + \gamma_k s^{\alpha_k})}{\tau_k s^{2\alpha_k + 2\zeta_k \tau_k s^{\alpha_k + 1}} \quad (6)$$

where p, q represents the number of terms of first and second kind, respectively; α_k is the k -th fractional order, R_k is the k -th gain, $\frac{1}{\tau_k} = w_k^{\alpha_k}$ in which w_k is the k -th cut-off frequency [12]. For the second kind terms, two additional parameters will be used: γ_k is related with the pole time-constant of the k -th element and ζ_k that can be considered as a pseudo-damping constant (similar to second-order systems but without the same meaning) [13].

1) Identification of the model parameters

The main goal in system identification is to estimate the (real valued) model parameters:

$$[R_0, \dots, R_{p+q}, \alpha_1, \dots, \alpha_{p+q}, \tau_1, \dots, \tau_{p+q}, \zeta_{p+1}, \dots, \zeta_{p+q}, \gamma_{p+1}, \dots, \gamma_{p+q}]$$

for the FOTF model (6) with a given number of terms p and q , respectively, of 1st and 2nd kind from a discrete set of measured frequency response data $Z_m(f_n)$, for $n = 1 \dots N$. The parameters are estimated by minimizing the following quadratic error:

$$\frac{1}{2} \sum_{n=1}^N (Z_m(f_n) - Z(f_n, \theta))^H (Z_m(f_n) - Z(f_n, \theta)) \quad (7)$$

i.e. the difference between the measured frequency response data and the model estimates, f_n is the frequency value, N the number of data points and E^H is the complex conjugate transpose of E . The Levenberg-Marquardt (LM) method is commonly used in this kind of optimization problems though it does not warrant good estimates for the model parameters. In particular, the existence of multiple minima due to the non-convexity of G and the possibility of outliers in the experimental data-set increases the risk of poor parameter estimates by the direct application of the method.

To minimize this risk, a new optimization strategy based on the LM method is proposed. This strategy can be seen as an extension of the random sample consensus algorithm (RANSAC) for nonlinear least squares. Broadly speaking, this strategy can cope with: a) data-sets contaminated with outliers, since it determines which samples are inliers and hence consistent with the model (consensus subset); and b) multiple minima by applying the LM method from multiple randomly selected starting points and to multiple (statistical) variations of the objective function.

Fig. 1 presents an outline of the proposed strategy. The algorithm starts by randomly picking a subset of the data (step 1), with a cardinality as small as possible to minimize the risk of selecting outliers, but high enough to allow the solution of the LM inner linear system.

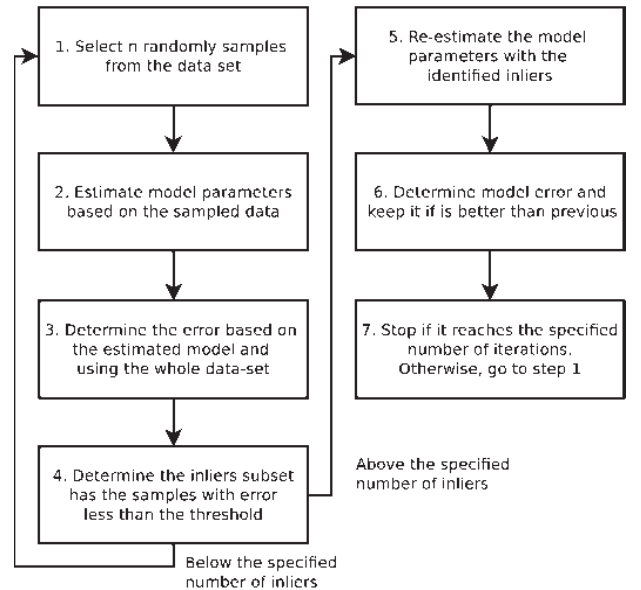


Figure 1: Modified random sampling consensus algorithm to the convergence performance of the Levenberg-Marquardt method.

In step (2), the model parameters are estimated by using the LM method with the parameters initiated randomly. Then, the errors are evaluated based on the estimated parameters and

using the whole data-set (step 3), followed by the determination of which samples are inliers by selecting those with an error smaller than a given threshold (step 4). This threshold is computed by using a robust statistic estimator for the error variance, based on the errors determined with the parameters computed in step 3. If the number of inliers is less than $0.75N$ then a return is made to step 1. Otherwise the model parameters are re-estimated using the LM method with the inliers subset (step 5). If the value of the objective function G is smaller than the previously known value, then the consensus subset is updated with the present one and the model parameters kept (step 6). Iteration proceeds until a stop criterion is reached.

III. RESULTS AND DISCUSSION

This section presents the main results from the application of the two methodologies to the analysis of experimental data collected from an open-air cathode PEMFC stack with a configuration and under experimental conditions described in [19]. Polarization curves were run in a step-by-step mode, starting at the open-circuit potential and taking the fuel cell to its maximum power. Hydrogen pressure was fixed at 0.5bar, being humidified at 40°C. The different polarization curves labels were recorded for all three hydrogen flowrates and the air was vented by a fan at four different levels (Table I).

TABLE I. EXPERIMENTAL DESIGN FLOW-RATES AND THE CORRESPONDING AIR TO H₂ FEED RATIO.

	Air [L min ⁻¹]			
	3.5	4.6	5.8	8.0
H ₂ [Lmin ⁻¹]	Feed Ratio			
0.2	17.6	23.2	28.8	40.0
0.3	11.7	15.4	19.2	26.7
0.4	8.7	11.6	14.4	20.0

The impedance spectroscopy data was collected under a constant hydrogen flowrate (0.4Lmin⁻¹ at 0.5 bar) and for three different operating conditions: a) high air flow (8.0 Lmin⁻¹) and low current (0.5A); b) high air flow and high current (1.0A); and c) low air flow (4.6 Lmin⁻¹) and low current.

A. Analysis of polarization data

Fig.2 shows the extracted M-NMF basis vectors (W) with the following properties: a) the three rank decomposition is enough to describe the original data variance; b) the M-NMF is able to find basis vectors with shapes similar to the over-

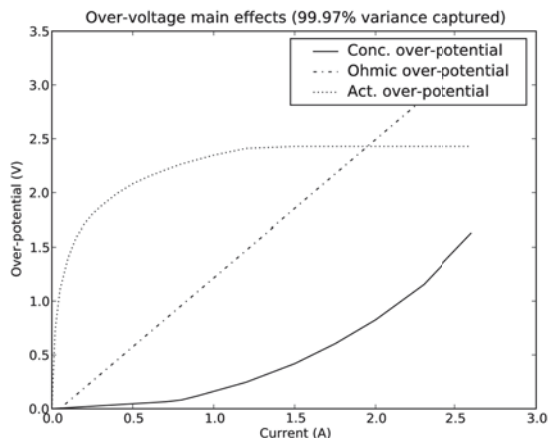


Figure 2: The M-NMF basis vectors (W) as function of current

potential contributions identified by empirical curve fitting methods, thus improving the interpretability of this data; and, c) the ohmic component has a constant slope, although this is not explicitly imposed on the M-NMF formulation.

Fig. 3 highlights the high variance distribution of the concentration over-potential over the different operating conditions.

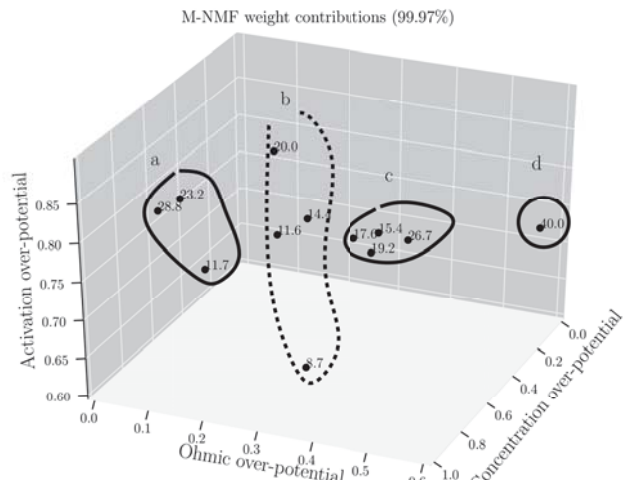


Figure 3: M-NMF weight contributions 3d plot (from H matrix) with the corresponding O₂ to H₂ feed ratio (labels).

A detailed analysis, reveals three major groups: i) cluster (a) corresponds to large concentration and small ohmic over-potential losses; ii) cluster (b) refers to the highest H₂ flow-rate and shows that with an increase of the feed ratio there are smaller ohmic and concentration over-potentials but a larger activation over-potential; and iii) cluster (c) that is related to small concentration and large ohmic over-potentials. The isolated point (case d) corresponds to the highest feed ratio conditions.

While the average over-potential components weight varies with the level of hydrogen flow, activation loss is the one that exhibits less variation and an overall higher order of magnitude, but with an irregular pattern.

The ohmic weight component has a general decreasing tendency with an increase in air flow rate, for intermediate and high values of hydrogen flow rates; this is associated to a decrease in the concentration over-potential. For the lowest hydrogen flow rate used in this work, 0.2 Lmin⁻¹, the ohmic component is quite variable, displaying a minimum at intermediate air flow rates; the effect is inverse for the corresponding concentration over-potential.

For a given hydrogen flow, the concentration over-potential loss is at its lowest for the higher air flow rates, with the smallest loss found at the highest air to H₂ feed ratio of 40. The concentration over-potential loss decreases with an increase in air flow. The case where both hydrogen and air flows are at their lowest, respectively 0.2 and 3.5 Lmin⁻¹, falls out of this common trend.

B. Analysis of impedance data

Fig. 4 presents the results obtained with a simple model composed by two terms ($p=1, q=1$), being noticeable the good agreement achieved for the low frequency range.

The parameter identification shows that there are two main principal modes extracted from the impedance spectroscopy data: corresponding to the anode and the cathode dynamics.

Table 1, shows the estimated model parameters. Note that, the table presents τ_k^{1/α_k} which can be regarded as the physical time-constant for the k -th modal term instead of τ_k ($\tau_k s^{\alpha_k} = (\tau_k^{1/\alpha_k} s)^{\alpha_k}$).

TABLE II. ESTIMATED MODEL PARAMETERS FOR THE THREE IMPEDANCE CURVES.

	R_0	R_1	τ_1^{1/α_1}	α_1	R_2	τ_2^{1/α_2}	α_2	ζ_2	γ_2
a	0.99	1.00	5.61E-3	0.64	5.13	0.14	0.42	0.09	2.27
b	0.71	0.94	7.89E-3	0.52	2.77	0.19	0.47	0.34	6.19
b	1.05	1.38	7.16E-3	0.60	5.31	0.18	0.59	0.63	8.31

The results analysis shows that the proposed model formulation can be used to estimate good quality parameters from experimental impedance data.

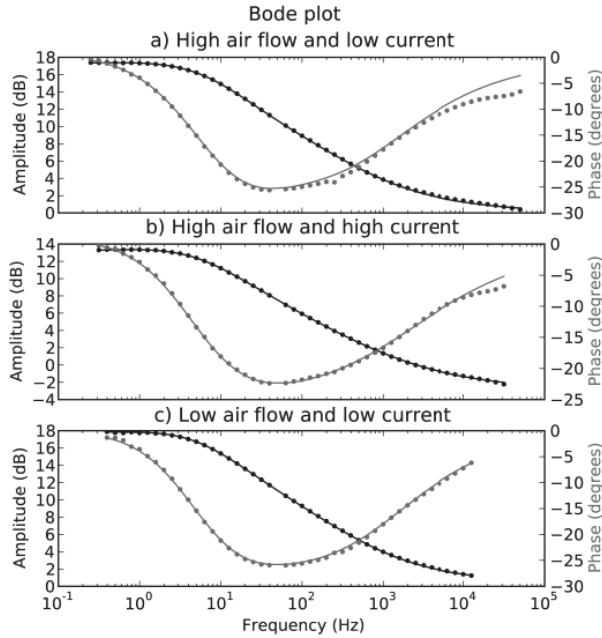


Figure 4: Bode plot showing the magnitude and phase components. The experimental values (dotted) and the model estimations (solid lines).

The knowledge of the main kinetic process in the PEMFC allows, in addition, to associate the R_0 with the membrane transport phenomena and the 1st and 2nd term with the anodic and cathodic processes, respectively. Results in table 1 show (τ_2^{1/α_2} vs. τ_1^{1/α_1}) that cathodic processes are responsible for the main dominant time-constant in all the analyzed cases.

This paper presented two novel data-driven methodologies for the analysis and interpretation of data measuring fuel-cells steady-state and dynamic performances, M-NMF and FC-FOTC respectively.

The deconvolution of fuel cell polarization curves into three major components was achieved with good results by the use of M-NMF, a modified NMF method supported by an algorithm addressing the associated nonlinear optimization problem, which proved also useful for the interpretation of experimental data.

FC-FOTC is a new model formulation based on FOTF written in the modal form with p terms of 1st kind and q terms of 2nd kind, which presents a new nonlinear optimization strategy for the estimation of the model parameters from datasets contaminated with outliers. It was shown that it is possible, with this methodology, to derive good model approximations from impedance data and estimate model parameters with a high degree of accuracy.

The application of these two methodologies was illustrated with real experimental data obtained from a 12W open cathode PEMFC series stack. It is thought that they can enhance the power quality of fuel cell stacks applications integrated in a power grid, since the impact of power conditioning unit design variables can be easily extracted.

ACKNOWLEDGMENT

The authors would like to acknowledge the partial financial support from the FAI-REIVE project.

REFERENCES

- [1] J. Wu, X. Yuan, M. Blanco, J. Martin, J. Zhang, H. Wang, Diagnostic tools in PEM fuel cell research: Part I Electrochemical techniques. *Int. J. Hydrogen*, 33, 2008, pp. 1735-1746
- [2] L. Pisani, G. Murgia, M. Valentini, B. D'Aguzzo, A new semi-empirical approach to performance curves of polymer electrolyte fuel cells, *J. Power Sources* 108, 2002, pp. 192-203
- [3] M. Hoffman D. Noren, Clarifying the Butler-Volmer equation and related approximations for calculating activation losses in solid oxide fuel cell models, *J. Power Sources*, 152, 2005, pp. 175-181
- [4] I. T. Jolliffe Principal Component Analysis, 2nd ed., Springer series in statistics. New York Springer-Verlag, 2002
- [5] S. Brown and R. Bear, Jr. Chemometric Techniques in Electrochemistry: A Critical Review, *CRC Crit. Rev. Analyt. Chem.*, 24, 1993, pp. 99-131
- [6] Y. Nia, S. Kokot Does chemometrics enhance the performance of electroanalysis? *Analytica Chimica Acta*, 626 (2), 2008, pp. 130-146
- [7] N. Wagner, W. Schnurnberger, B. Muller, M. Lang, Electrochemical impedance spectra of solid-oxide fuel cells and polymer membrane fuel cells, *Electrochimica Acta* 43, 1998, pp. 3785-3793
- [8] V. Paganin, C. Oliveira, E. Ticianelli, T. Springer, E. Gonzalez, Modelistic interpretation of the impedance response of a polymer electrolyte fuel cell, *Electrochimica Acta*, 43, 1998, pp. 3761-3766
- [9] X. Yuan, C. Song, H., J. Zhang, Electrochemical Impedance Spectroscopy in PEM Fuel Cells Fundamentals and Applications, Springer, 2010

- [10] N. Fouquet, C. Douleta, G. Nouillanta, B. Ould-Bouamama, Model based PEM fuel cell state-of-health monitoring via ac impedance measurement, *Journal of Power Sources*, 159(2), 2006, pp. 905-913
- [11] K. Saito, M. Sugi, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E76(2), 2003, pp. 205-209
- [12] S. Westerlund, Ekstam L., Capacitor theory, *IEEE Transactions Dielectrics and Electrical Insulation*, 1(5), 1994, pp. 826-839
- [13] M. Fellah, Z. Fellah, C. Depollier, Transient wave propagation in inhomogeneous porous materials: Application of fractional derivatives, *Signal Processing* 86(10), 2006, pp. 2658-2667
- [14] J. Bisquert, A. Compte, Theory of the electrochemical impedance of anomalous diffusion, *Journal of Electroanalytical Chemistry* 499(1), 2001, pp. 112-120
- [15] L. Gray, S. Liu, Self-affine fractal model for a metal-electrolyte interface, *Physical Review B*, 35(10), 1987, pp. 5379-5381
- [16] E. Barsoukov, J. Macdonald, *Impedance Spectroscopy: Theory, Experiment, and Applications*, 2nd edition, John Wiley and Sons, 2005
- [17] M. Berrya, M. Brownea, A. Langvilleb, V. Paucac, R. Plemmons, Algorithms and applications for approximate non-negative matrix factorization. *Comp. Stat. & Data Anal.*, 53, 2007, pp. 155-173.
- [18] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, 2nd edition, 1999.
- [19] Lopes, V. V., Novais, A.Q., Rangel, C. R.; Hydrogen PEMFC stack performance analysis: a data-driven approach. In *HYPOTHESIS VIII, Hydrogen - Power Theoretical and Engineering Solutions International Symposium*, 1-3 April, Lisbon, 2009
- [20] Lopes, V. V.; Novais, A.Q., Rangel, C. M.; Hydrogen PEMFC stack performance analysis: A data-driven approach, *International Journal of Hydrogen Energy*, 35 (18), 2010, pp. 9973-9982
- [21] Lopes, V. V.; Rangel, C.M.; Novais, A.Q.; "Fractional-order transfer functions applied to the modelling of hydrogen PEM fuel cells", 21st European Symposium on Computer Aided Process Engineering – ESCAPE 21, 29th May-1st June, Chalkidiki, Greece, 2011
- [22] T. Hartley, C. Lorenzo, A solution to the fundamental linear fractional order differential equation, *Critical Reviews in Biomedical Engineering*; 36(1), 2008, pp.23-38
- [23] R. Malti, X. Moreau, F. Khemane, *New Trends in Nanotechnology and Fractional Calculus Applications*, Editors: D. Baleanu, Z. Güvenc, J. Machado, Springer Netherlands, 2010