Statistical and Principal Component Analysis in the Design of Alkaline Methanol Fuel Cells

Tanja Clees
University of Applied Sciences
Bonn-Rhein-Sieg and Fraunhofer Institute
for Algorithms and Scientific Computing
Sankt Augustin, Germany
email: Tanja.Clees@scai.fraunhofer.de

Bernhard Klaassen
Fraunhofer Institute for Algorithms
and Scientific Computing
Sankt Augustin, Germany
email: Bernhard.Klaassen@scai.fraunhofer.de

Igor Nikitin
Fraunhofer Institute for Algorithms
and Scientific Computing
Sankt Augustin, Germany
email: Igor.Nikitin@scai.fraunhofer.de

Lialia Nikitina
Fraunhofer Institute for Algorithms
and Scientific Computing
Sankt Augustin, Germany
email: Lialia.Nikitina@scai.fraunhofer.de

Sabine Pott
Fraunhofer Institute for Algorithms
and Scientific Computing
Sankt Augustin, Germany
email: Sabine.Pott@scai.fraunhofer.de

***Abstract*—In this paper, the electrochemical alkaline methanol oxidation process, which is relevant for the design of efficient fuel cells, is considered. An algorithm for reconstructing the reaction constants for this process from the experimentally measured polarization curve is presented. The approach combines statistical and principal component analysis and determination of the trust region for a linearized model. It is shown that this experiment does not allow one to determine accurately the reaction constants, but only some of their linear combinations. The possibilities of extending the method to additional experiments, including dynamic cyclic voltammetry and variations in the concentration of the main reagents, are discussed.**

***Index Terms*—modeling of complex systems; observational data and simulations; advanced applications; mathematical chemistry.**

I. Introduction

Fuel cells are environmentally friendly portable energy sources based on obtaining electricity as a result of electrochemical reactions. They are similar to galvanic cells; the difference is that the main reagents in fuel cells can be replenished many times. Among fuel cells, a special group is formed by the so-called direct fuel cells, in which the intermediate stage of the production of gaseous hydrogen is omitted and the oxidative reaction proceeds directly. In fact, this is the same combustion reaction, but here the energy is released not in the form of heat or mechanical pressure of gas, but in the form of electric power. Among the various fuels in such cells, the most common are alcohols, methanol and ethanol. We consider the methanol oxidation reaction to reduce the number of intermediate reagents for modeling. Although acidic reactions are easier to model and have been well studied, they rely on the use of expensive noble metal electrodes. The alkaline environment allows the use of cheaper materials for the production of electrodes.

The main problem for the analysis of electrochemical alkaline methanol oxidation is the large number of intermediate reagents and reactions, as well as the fact that the elementary reaction constants are not known a priori, and they must be reconstructed from the experiment. Moreover, such quantities, as surface coverages of the electrode, are experimentally not measurable and require mathematical modeling. The challenge here is to consider the systems of differential-algebraic equations of high dimension. Additional complication comes from the stiffness of the system: some components evolve much faster than others. There are also slow reactions that inhibit the entire process and lead to the unattainability of a stationary state and hysteresis effects in the current-voltage dependence.

In this paper, we continue our research on mathematical modeling of alkaline methanol oxidation in the context of design of efficient fuel cells. A detailed description of the mathematical model is given in our previous paper [1]. In our other papers, a model reduction [2] of the system from Ordinary Differential Equations (ODE) to Differential-Algebraic Equations (DAE) was performed; a chemical interpretation for hysteresis effect [3] in dynamics of the system is presented; a footprint of the dynamics in the form of electrochemical impedance spectrum [4] was analyzed. Our papers extend the approaches of [5]–[8] for analysis of this and analogous electrochemical processes, paying more attention to mathematical aspects of the problems. We use the electrochemical measurement methodology from [9]. We also use general methods of model data analysis [10], most of which are implemented as ready-to-use procedures in the system Mathematica [11].

The main goal of the experiments under consideration is the reconstruction of the reaction constants describing the underlying electrochemical processes. Note that the reconstructed values of the constants for alkaline methanol oxidation are given in [1]. The main purpose of this work is to estimate *the reconstruction accuracy* of the reaction constants.

In Section II, an overview of the mathematical model of the reactions in alkaline methanol oxidation is given. In

 ISBN: 978-1-61208-887-7

Section III, the method of statistical and principal component analysis for estimating the reconstruction accuracy of the reaction constants in the considered process is presented. Section IV summarizes the obtained results.

The experimental measurements used in this paper were carried out by our colleagues at INES / TU Braunschweig [1]–[4].

II. The model

Figure 1a shows the chemical reaction network under consideration. Figure 1b shows an experimental setup containing an evacuated teflon cell with a rotating electrode. A detailed description of reactions and setup can be found in [1]. In the experiments, the voltage η in the cell changes along a certain profile, and the current I_{cell} through the cell is measured. In particular, Cyclic Voltammetry (CV) experiments use a saw-like voltage profile shown in Figure 1c.

The resulting cell response is shown in Figure 2a. Note that at low voltage values, the profiles for the increasing and decreasing half-cycles of the saw run approximately along the same curve. In this zone, the reagents are in equilibrium, their concentrations depend only on the voltage, following the so-called Polarization Curve (PC). At higher voltages, a new reagent is formed. Its reactions are slow, as a result, the equilibrium is disturbed, a delay appears in the response of the system, and a characteristic hysteresis is formed on the CV-plot.

The evolution of the system is described by ODE of the form

$$d\theta_i/dt = F_i(\theta,\eta),\ i = 1..6,\ I_{cell} = F_7(\theta,\eta), \qquad (1)$$

where $\theta_i \in [0,1]$ are surface coverages for 6 reagents. The right-hand sides of the equations are polynomials in θ_i, the exact form of which is given in [1]. Some monomials corresponding to electron exchange reactions have an exponential voltage dependence. The purpose of the experiments is to reconstruct the reaction constants k_i, which are the coefficients of the monomials in the given model. In total, there are 14 reaction constants for the considered system of reactions. In the normalization used in [1], the reaction constants are measured in $[mol/(m^2 s)]$, while their numerical values vary over a wide range. Therefore, it is convenient to use the decimal logarithms of the reaction constants as model parameters: $p_i = \log_{10} k_i$.

III. Statistical and principal component analysis

In this work, we will consider the PC-part of the CV-curve, shown in detail in Figure 2b. In this part, the state of the system can be considered stationary, and the terms with derivatives in (1) can be omitted. As a result, a closed polynomial system on θ_i is formed, which can be solved. The obtained θ_i as a function of voltage can be substituted into the expression for the current, resulting in a model response function $I_{cell} = f(\eta,p)$. In Figure 2b, the red curve corresponds to the model, and the blue dots to the measured values. It can be seen that the model reproduces the experiment very well.

For reconstruction in [1], a fitting procedure is used that minimizes the sum of the squares of the deviations of the model from the values measured in the experiment:

$$L_2^2 = \textstyle\sum_i (f(\eta_i,p) - f_{exp,i})^2. \qquad (2)$$

The p-values obtained at the minimum point represent the reconstructed reaction constants. The accuracy of the reconstruction is determined as follows [10]. At the minimum, the sensitivity matrix X and related matrices are determined:

$$X_{ij} = \partial f(\eta_i,p)/\partial p_j,\ cov = \epsilon^2 (X^T X)^{-1} \qquad (3)$$

$$\sigma_i = (cov_{ii})^{1/2},\ corr_{ij} = \sigma_i^{-1} cov_{ij}\, \sigma_j^{-1}, \qquad (4)$$

the covariance matrix cov, the diagonal values of which determine the standard deviations of the parameters σ_i, and the correlation matrix $corr$. The value ϵ appearing in the definitions represents the estimate of the experimental data error $f_{exp,i}$, calculated by the formula $\epsilon^2 = L_2^2/N_{dof}$, $N_{dof} = N_{pt} - N_{par}$. Here, N_{pt} is the number of experimental points, N_{par} is the number of model parameters, N_{dof} is the number of degrees of freedom for a fit, in our case:

$$N_{pt} = 21,\ N_{par} = 14,\ N_{dof} = 7, \qquad (5)$$

$$L_2^2 = 2.08 \cdot 10^{-9} A^2,\ \epsilon = 1.72 \cdot 10^{-5} A. \qquad (6)$$

The obtained small value of ϵ corresponds to the good quality of the fit. In fact, this value corresponds to the size of the dot on the graph Figure 2b.

In the case when the parameter errors are large and highly correlated with each other, Principal Component Analysis (PCA) must be performed to interpret the result. The confidence region in parameter space is an ellipsoid that can be stretched in some directions and compressed in others. In stretched directions, measurements have a large error, in compressed directions – small one. The values and directions of the semi-axes of the ellipsoid can be determined using Singular Value Decomposition (SVD) of the sensitivity matrix:

$$X = u\lambda v^T,\ u^T u = 1,\ v^T v = v v^T = 1,\ a_k = \epsilon/\lambda_k. \qquad (7)$$

Here, the matrix X is $N_{pt} \times N_{par}$ rectangular, u is $N_{pt} \times N_{par}$ semi-orthogonal, λ is $N_{par} \times N_{par}$ diagonal, v is $N_{par} \times N_{par}$ orthogonal. The a_k values represent the semi-axes of the error ellipsoid. The columns of the v matrix (or the rows of the v^T matrix) represent the directions of the axes of the ellipsoid in the parameter space, while the columns of the u matrix represent the profiles of the principal components in the space of experiments. Figure 2d shows such profiles for the first four components, in red-green-blue-cyan order for $u_{i,1-4}$. These profiles show the variation of PC curve when the parameters are displaced along the axes of the error ellipsoid.

Note: addressing the question, why we choose PCA for the analysis, and not another factorizing method, e.g., Independent Component Analysis (ICA) or Curvilinear Component Analysis (CCA), etc. These methods are very close to PCA, and ICA even uses SVD in the main phase of the computation, so called signal whitening. However, these methods belong to different fields of application; ICA is used

 ISBN: 978-1-61208-887-7

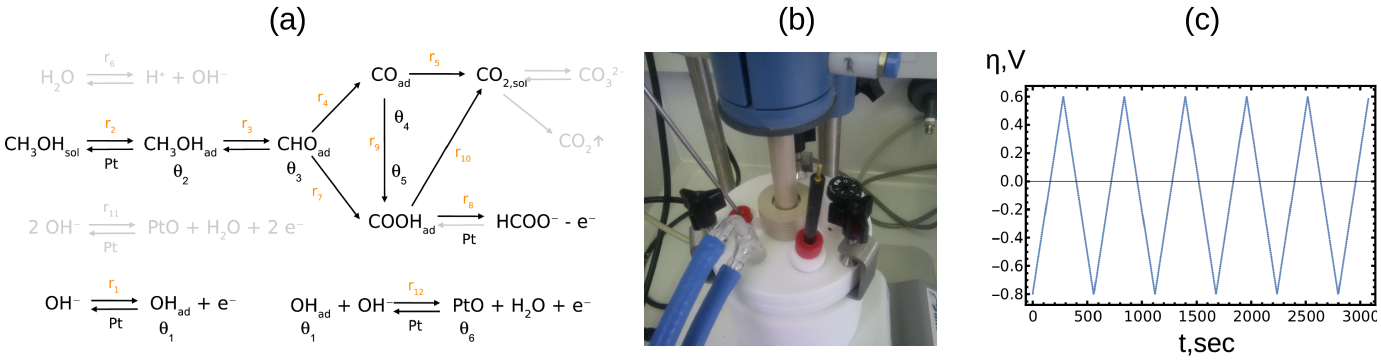

Fig. 1. (a) network of chemical reactions, (b) experimental setup, (c) saw-like voltage profile. Images from [1], [2].

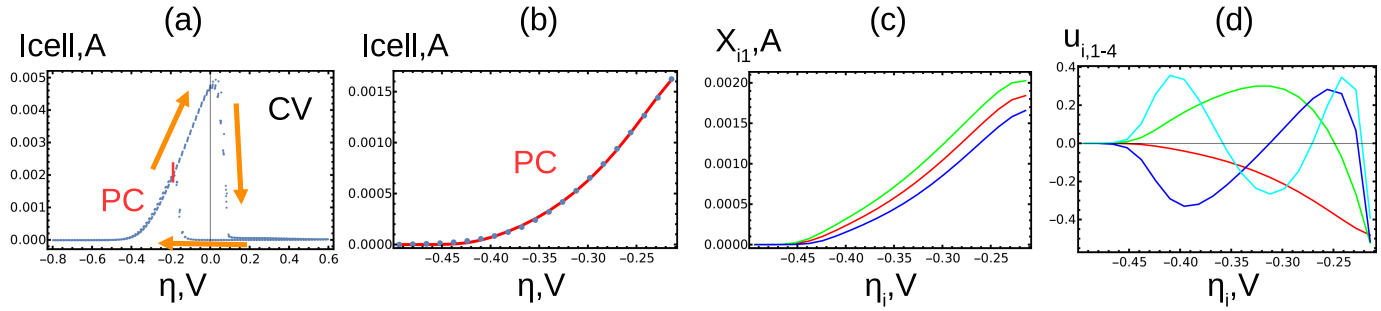

Fig. 2. (a) CV plot with selected PC part, (b) fit of PC curve by the model, (c) sensitivity analysis, (c) principal component analysis (for color coding see main text).

TABLE I
PARAMETER CENTRAL VALUES AND TRUST REGION OF LINEAR MODEL

p_j	0.949	-4.5	0.398	-0.563	4.72	-3.46	0.352	-0.101	1.2	-8.66	1.89	-1.08	-1.72	-7.82
dp_j	0.3	0.1	0.3	0.4	1.5	0.1	0.1	0.06	0.1	0.1	0.08	0.2	0.15	0.15

TABLE II
RESULTS OF PRINCIPAL COMPONENT ANALYSIS

λ_k, A	a_k	v_{jk}													
$8.12 \cdot 10^{-3}$	$2.12 \cdot 10^{-3}$	−0.531	0.014	0.461	−0.459	0.140	−0.013	−0.007	0	0	0	0	−0.522	0.032	−0.032
$1.21 \cdot 10^{-3}$	$1.43 \cdot 10^{-2}$	0.396	−0.159	−0.251	0.250	0.214	−0.010	−0.018	0	0	0.002	0	−0.769	0.167	−0.167
$2.71 \cdot 10^{-4}$	$6.36 \cdot 10^{-2}$	−0.078	0.639	−0.052	0.058	−0.402	0.065	−0.065	0	0	0.004	0	−0.052	0.451	−0.451
$1.24 \cdot 10^{-4}$	0.139	−0.066	−0.711	0.091	−0.074	−0.191	−0.058	0.066	0	0	−0.014	0	0.197	0.443	−0.443
$7.79 \cdot 10^{-6}$	2.21	0.305	−0.063	0.211	−0.123	−0.728	−0.290	0.277	0	−0.001	−0.113	0	−0.234	−0.206	0.206
$5.1 \cdot 10^{-6}$	3.38	0.546	0.214	0.540	−0.152	0.408	−0.204	0.205	0	0	−0.056	0	0.192	0.160	−0.160
$8.51 \cdot 10^{-7}$	$2.02 \cdot 10^{1}$	0.315	−0.066	0.148	−0.264	−0.163	0.336	−0.449	0.001	0.032	0.679	−0.003	−0.010	−0.033	0.033
$1.69 \cdot 10^{-7}$	$1.02 \cdot 10^{2}$	0.033	−0.083	0.508	0.459	−0.114	0.441	−0.349	−0.001	−0.029	−0.434	0.003	−0.033	−0.046	0.046
$2.52 \cdot 10^{-8}$	$6.84 \cdot 10^{2}$	0.240	−0.003	−0.304	−0.617	−0.003	0.439	−0.030	−0.003	−0.080	−0.518	0.012	0.021	0.013	−0.013
$7.81 \cdot 10^{-9}$	$2.21 \cdot 10^{3}$	0.047	−0.001	−0.075	−0.135	0.004	−0.582	−0.722	0.008	0.212	−0.257	−0.039	0.031	0.004	−0.004
$5.92 \cdot 10^{-10}$	$2.91 \cdot 10^{4}$	0	0	−0.002	−0.002	0	−0.165	−0.159	−0.037	−0.972	0.022	0.002	0.007	0	0
$9.35 \cdot 10^{-13}$	$1.84 \cdot 10^{7}$	0	0	0	0	0	0.024	0.024	0.485	−0.028	0	−0.874	0	0	0
$3.38 \cdot 10^{-13}$	$5.09 \cdot 10^{7}$	0	0	0	0	0	0.014	0.014	−0.874	0.028	0	−0.485	0	0	0
0	∞	0	0	0	0	0	0	0	0	0	0	0	0.707	0.707	

 ISBN: 978-1-61208-887-7

to separate independent sources in signal processing, while CCA is a method of non-linear dimensionality reduction. PCA and the underlying spectral decomposition are general statistical methods for data analysis. Here, we only need to perform a statistical analysis of the measurement errors, for which the standard PCA method is best suited.

Details of implementation: it is recommended in [10] to use the formula (3) when there is no independent estimate of measurement errors. In this case, the measurement error is obtained from the χ^2-criterion. Assuming the same error for all measurements, one has $\chi^2 = L_2^2/\epsilon^2$, in case of a good fit, $\chi^2 = N_{dof}$ is fulfilled; from here one can find ϵ. The geometric meaning of this definition is that for a good fit the deviations of the experimental points from the model can be considered appearing due to a random measurement error, and the standard deviation of such a variation characterizes this error. The term $N_{dof} = N_{pt} - N_{par}$ in the denominator of this definition, instead of the traditional $N_{pt} - 1$, takes into account the influence of the fit parameters, which take on a part of the experimental scatter.

The sensitivity matrix is determined using finite difference schemes of the form

$$X_{ij}^{+} = (f(\eta_i, p + dp_{(j)}) - f(\eta_i, p))/dp_j, \quad (8)$$

$$X_{ij}^{-} = (f(\eta_i, p) - f(\eta_i, p - dp_{(j)}))/dp_j, \quad (9)$$

$$X_{ij} = (f(\eta_i, p + dp_{(j)}) - f(\eta_i, p - dp_{(j)}))/(2dp_j), \quad (10)$$

where $X_{ij}^{\pm}$ – forward/backward, $X_{ij} = (X_{ij}^{+} + X_{ij}^{-})/2$ – central difference scheme. Here, $dp_{(j)} = (0, ..., dp_j, ..., 0)$ represents a vector with non-zero entry on j-th place. Such derivatives, for $j = 1$, are shown in Figure 2c, red line for central, green/blue lines for forward/backward schemes, respectively. More precisely, up to the next order, is the central difference scheme, which should be taken as the final answer. Other profiles are needed to evaluate the non-linearity of the function. Indeed, for linear functions, all these three profiles coincide, and their deviation from each other is a measure of non-linearity. In practice, we adjust the dp_j variation so that the curves deviate from each other by $\sim 20\%$. This analysis is performed for all j; the results are shown in Table I. Such variations define a box-like trust region where a linear model can be applied.

Another detail is the presence of failures in the definition of the model function. Since this definition uses the solution of a high degree polynomial system, the applied numerical procedures can lose solutions sometimes. Such failures are rare, estimated in $\sim 0.5\%$ cases. However, this is sufficient to destabilize the automatic minimization procedures, with the result that achieving the true minimum is not guaranteed. The methods described in [1] help in this situation, including finding the starting point manually and applying random search with discarding the failed cases. Automatic minimization algorithms were applied to the resulting starting points, which slightly improved the objective function. Finally, we made sure visually and by formal ϵ-criterion that the fit has a good quality. Since the failures also occur when using finite-difference schemes, some of the dp_j parameters required fine tuning to completely eliminate the failed cases.

The next surprise was the degeneration of the matrix $X^T X$, which makes the usage of the formulas (3), (4) impossible. The degeneration is seen in Table II, where the last row has a zero eigenvalue. The eigenvector for this value corresponds to the simultaneous variation of the parameters $\delta p_{13} = \delta p_{14}$. Formally, this direction corresponds to the infinite scatter a_k, meaning that it is impossible to measure the corresponding linear combination of constants. Indeed, a detailed analysis of the model given in [1] shows that such a variation corresponds to the exact symmetry of the system and does not change the observed values. In fact, the system depends only on the difference between the logarithmic parameters $p_{13} - p_{14}$, or, in the original notation, on the ratio of the corresponding reaction constants. This symmetry takes place only for a stationary system on the PC-part of the CV-curve. Dynamics can break this symmetry and make these constants individually measurable.

Further, in Table II, the first three lines correspond to the variation a_k that is located within the previously defined trust region of the linear model. The fourth line is also located in the trust region, with a tension. The following directions are outside of the trust region. Thus, the first four directions appear to be measured more or less accurately, while the rest of the directions have too large scatter. This conclusion is the main result of the PCA.

Note that the determination of the covariance and correlation matrices in this problem becomes meaningless. The point is not only that the sensitivity matrix is formally degenerate. There are many directions along which the error ellipsoid is strongly stretched, in the projections on axes of the initial parameters giving very large errors, strongly correlated between different parameters. It is PCA/SVD method that sheds light on the structure of solutions in the problem under consideration.

To improve the obtained result, in addition to considering the complete dynamic problem, other experiments can be included in the analysis. In particular, the experiments with variations of the volume concentrations of the main reagents $c_{1,2}$ can be used. These concentrations enter polynomially in the model [1], and accordingly extend the model response $I_{cell} = f(\eta, c_1, c_2, p)$ and the sensitivity matrix X. The reconstruction accuracy for such an extension can be analyzed using the general methodology described here.

Implementation in Mathematica: we use Mathematica [11] for the calculations described above. The `NSolve` method is used to solve a stationary polynomial system; for the obtained set of solutions, real roots are selected from the interval $\theta_i \in [0, 1]$. The system also has a spurious solution ($\theta_i = \delta_{i4}$, $I_{cell} = 0$), which also needs to be removed. As a result of numerical instabilities, with these selections, the roots sometimes disappear, leading to the aforementioned failures of the algorithm. To solve the dynamic system, `NDSolve` method is used, able to integrate both ODE and the partially reduced DAE system [2]. Systems are highly stiff and the integration algorithm also fails sometimes. In

 ISBN: 978-1-61208-887-7

the fitting algorithm [1], to select the starting point, the interactive configuration tool `Manipulate` is used. After that, we used automatic methods for minimizing the L_2-norm, local `FindMinimum` and global `NMinimize`. In this work, we have used `NonlinearModelFit`, which provides a convenient interface to the same optimization methods. The differentiation algorithm requires setting up finite-difference schemes, which can be passed to the fitting method via the `Gradient` option. Further, the calculation of the covariance and correlation matrix turns out to be impossible due to the degenerations described above. At this point, the standard computation should be replaced with PCA using the available method `SingularValueDecomposition`.

IV. Conclusion and Future Work

We considered an algorithm for reconstructing the reaction constants from the experimentally measured polarization curve for the electrochemical alkaline methanol oxidation process. Our approach combines statistical and principal component analysis. We define formal criteria for reconstruction accuracy based on the estimate of the trust region for the linearized model. As a result of the analysis, it turned out that the described experiment does not make it possible to determine precisely all 14 reaction constants, but only 4 their certain linear combinations. Of the remaining orthogonal combinations, one corresponds to the symmetry of the stationary system and is fundamentally indeterminate in the described experiment. The remaining 9 combinations have insufficient reconstruction accuracy. To improve this result, other experiments should be involved in the analysis, including fully dynamic cyclic voltammetry and variations in the concentration of the main reagents. We are going to expand the developed methodology for additional experiments elsewhere.

References

[1] T. Clees et al., “Mathematical modeling of alkaline methanol oxidation for design of efficient fuel cells”, Advances in Intelligent Systems and Computing, vol 947, 2020, pp. 181-195.

[2] T. Clees et al., “Parameter identification and model reduction in the design of alkaline methanol fuel cells”, Int. J. On Advances in Systems and Measurements, vol. 13, 2020, pp. 94-106.

[3] T. Haisch et al., “The origin of the hysteresis in cyclic voltammetric response of alkaline methanol electrooxidation”, Physical Chemistry Chemical Physics, vol. 22, 2020, pp. 16648-16654.

[4] T. Clees et al., “Electrochemical impedance spectroscopy of alkaline methanol oxidation”, in Proc. INFOCOMP 2017, pp. 46-51, Pub. IARIA 2017.

[5] U. Krewer, T. Vidakovic-Koch, and L. Rihko-Struckmann, “Electrochemical oxidation of carbon-containing fuels and their dynamics in low-temperature fuel cells”, ChemPhysChem, vol. 12, 2011, pp. 2518-2544.

[6] U. Krewer, M. Christov, T. Vidakovic, and K. Sundmacher, “Impedance spectroscopic analysis of the electrochemical methanol oxidation kinetics”, J. Electroanalytical Chem., vol. 589, 2006, pp. 148-159.

[7] B. Beden, F. Kardigan, C. Lamy, and J. M. Leger, “Oxidation of methanol on a platinum electrode in alkaline medium: effect of metal ad-atoms on the electrocatalytic activity”, J. Electroanalytical Chem., vol. 142, 1982, pp. 171-190.

[8] F. Ciucci, “Revisiting parameter identification in electrochemical impedance spectroscopy: Weighted least squares and optimal experimental design”, Electrochimica Acta, vol. 87, 2013, pp. 532-545.

[9] A. J. Bard and L. R. Faulkner, Electrochemical Methods: Fundamentals and Applications, Wiley 2000.

[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in C, Cambridge University Press 1992.

[11] Mathematica 12, Reference Manual, http://reference.wolfram.com, [retrieved: August, 2021]

 ISBN: 978-1-61208-887-7