



DATA ACQUISITION AND DATA PROCESSING CHALLENGES IN HEAVY METAL MEASUREMENTS

José Miguel Dias Pereira^{1, 2)}, Ricardo Manuel Nunes Salgado^{1, 3)}

¹⁾ LabIM, ESTSetúbal/IPS, Escola Superior de Tecnologia de Setúbal do Instituto Politécnico de Setúbal, Rua Vale de Chaves, Campus do IPS, Estefanilha, 2910-761 Setúbal, Portugal

E-mail: dias.pereira@estsetubal.ips.pt

²⁾ Instituto de Telecomunicações, Av. Rovisco Pais, 1049-001, Lisboa, Portugal.

³⁾ REQUIMTE/CQFB, Chemistry Department, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal.

Abstract: Water quality is a key factor to preserve human life quality, as well as, environmental and biological ecosystems. This paper highlights specific issues related with the acquisition and data processing of heavy metal measurement systems. Between the main challenges related with this kind of measurements, such as the ones related with the very low signal amplitudes to be measured, since very low concentrations, in the order of tens of p.p.b., of some heavy metals, can be very dangerous for human life and for ecosystems sustainability. Additional challenges, that are associated with online heavy metals measurements, are related with the capability to obtain accurate results using a low number of measurement points. Thus, the main goal of this paper is a comparison of different data processing algorithms that can be used to improve heavy metal measurement accuracy when a segmented voltammetric voltage scan is performed with a low number of measurement points. Regarding data processing of the measurement data, B-Spline, Gaussian and artificial neural network based techniques are compared with traditional least mean square techniques based on polynomial curve fitting. The performance of each technique is evaluated in terms of the required number of measurement points, for a given root mean square deviation between curve fitted and experimental data. A brief comparison of the different techniques, in terms of insensitivity to errors caused by measurement data outliers, is also presented. *Copyright © Research Institute for Intelligent Computer Systems, 2015. All rights reserved.*

Keywords: Heavy metals, curve fitting methods, measurement data outliers, measurement accuracy, sensitivity and celerity.

1. INTRODUCTION

Heavy metals (HM) are chemical elements with a specific gravity higher than 5, such as arsenic, cadmium, iron, lead and mercury, among others, and, some of them, can be very toxic for living organisms. In some cases, the toxic effect on the living organisms can be observed even when small concentrations of these elements are present in the environment. However, small amounts of HM can be an important micronutrient for many living organisms, as long as their concentrations is lower than specific thresholds. Above these thresholds, defined by international environmental organizations [1, 2], HM are considered as pollutant and poisoned for the organisms. Acute and chronic diseases, caused by excessive HM concentrations, can affect the organisms depending on the exposure time and the concentration of these compounds. The organs more sensitive to this toxic effect are the lung, kidney and liver but effects on the blood composition and in the nervous system can be also

observed due to the contact with HM in the environment.

Moreover, human activities can also contribute to increase the concentration of the HM and change the natural balance between the elements and compounds found in surface, groundwater and the coastal waters [3]. Thus, the measurement of concentration variability in the natural environment is a challenge for online monitoring HM concentrations that can affect the human life and the natural ecosystems. The water quality measurements are important to get information about retention time of the water, aquifers age, flows and recharge area and also in areas with irrigation or aquaculture systems. The monitoring plan and the environment measurements can help us to decide prohibitive areas and non-prohibitive areas for some specific activities due the contamination and pollution risk involved. The renewable of the estuary water in the aquaculture tanks should not have HM, or at least, they should be present in very small concentrations

for bivalve aquaculture production (e.g. oysters, bivalve in general and fishes).

Concerning HM measurement technologies, it is important to refer that sensors and micro devices (e.g. microelectrodes) had an extraordinary evolution in the last years and there exist promising solutions to perform heavy metals measurements based on micro-electromechanical-systems and bulk acoustic wave sensors [4, 5]. Recent developments of the previous systems and sensors enabled the appearance of new solutions to measure a large number of chemical and biological quantities. Nevertheless, even if these measurement methods are very attractive in terms of hardware compactness, they exhibit several limitations, namely in terms of reproducibility, selectivity, cross-sensitivity and response time.

In the last years, the authors developed several research activities focusing their attention in water quality assessment systems [6, 7]. Low-cost solutions and prototypes for the measurement of water temperature, conductivity, dissolved oxygen, turbidity, pH and HM, have been presented and successfully tested in real measuring environments. This paper highlights specific measurement issues related with online HM measurements, in terms of measurement celerity and data processing.

Different curve fitting techniques of the measurement data are used to compare their relative performance, in terms of the minimization of the number of measurement points that are required to achieve a given measurement accuracy.

It is important to underline that there are a large number of curve fitting techniques that can be used to evaluate HM concentrations in water solutions [8-14]. However, comparative studies of the different techniques are not frequent and, above all, those comparative studies generally considers a continuous, and linear, voltammetric voltage scan and not a segmented voltammetric voltage scan, as the one proposed by the authors to improve online HM measurement celerity.

The remaining part of this paper is organized as follows: section 2 includes a survey of several HM measurement methods underlining the main characteristics of electrochemical methods; section 3 focuses some specific issues related with the acquisition and data processing of HM measurement signals and section 4 performs a comparative analysis between the traditional polynomial, B-Spline, Gaussian and ANN based curve methods. The last part, section 5, is dedicated to conclusions.

2. HEAVY MEASUREMENT METHODS

This section presents a brief summary of different heavy metal measurement methods underlining

electrochemical methods, based on voltammetric techniques, that are a promising solution to perform online measurements.

2.1 ATOMIC ABSORPTION METHODS

The atomic absorption methods and atomic emission methods can be carried out by atomic absorption spectrometry (AAS) or atomic emission spectrometry, respectively. In the AAS, the sample is aspirated through the atomization chamber where it is nebulized by a flow gas oxidant to aerosol form and the atomization reaction occur in the chamber giving a signal of the atoms formed, ionized to cations and electrons. The excited ions and molecules can produce emission spectra that can be measured in a detector. The fuel, the oxidant and the maximum burning velocity used for exciting the molecules depend on the necessary temperature to measure a specific metal.

The excellent limits of detection and linear dynamic range, capability of measuring different HM, low chemical interference and a stable and reproducible signal are some of the advantages of the AAS techniques. The disadvantages of this technique are spectral interferences, investment cost of the equipment and the operating costs and the fact that samples typically must be in solution and cannot be applied on online measurements in the field. Another limitation of this technique is the application to solid matrices. In some cases, a previous extraction is needed. For this cases, the spark or arc atomic emission spectroscopy can be used for HM analysis in solid samples.

Other kind of equipments can be used for HM monitoring environmental samples can include the X-ray fluorescence and ultraviolet-visible spectroscopy. Due to the high investment costs and maintenance, the X-ray fluorescence are not so commonly used in many laboratories and the low levels of sensitivity and limits of detection and quantification, makes the ultraviolet-visible spectroscopy (UV-Vis) not so used for monitoring HM in environment. In X-ray fluorescence, the samples are bombarding with a high energy X-ray or gamma rays in order to get excited and to obtain the emission spectrum characteristics or a fluorescence X-ray spectrum [15]. The use of UV-Vis spectroscopy for monitoring HM needs in most of the cases to use a chemical reagent that reacts in complexation reaction with the metal and produce a color that can be measured by the absorption spectrum. The solutions of the HM ions can be colored (i.e., absorb visible light) because d electrons within the metal atoms can be excited from one electronic state to another. The presence of other species, such as certain anions or ligands color can

strongly affect the analysis of HM by this technique. In other cases, the colors are often too intense to use this method for quantitative measurement.

2.2 ELECTROCHEMICAL METHODS

The electrochemical analytical methods can be used to provide reliable results in short periods of time, that makes them an attractive solution for online measurements, the results are consistent throughout the development cycle of the compounds and, most of the times, these methods and data can be transferable to laboratories for validation and comparison with the classical measurements in AAS.

The electroanalytical methods can easily be used to measure HM with high accuracy, precision, sensitivity and selectivity and often reproducible. Due to the simplicity of the equipment, it is often selected for online monitoring of environmental samples. This is one of the major advantages of using this technique. The electrochemical methods associated to gold microelectrodes [16] or carbon vitreous microelectrodes [17] are sensitive, selective, rapid and easy techniques applicable to analysis HM in fields for monitoring the environment.

The sensitivity of the electrochemical techniques is determinant for the use of this technique in the field for monitoring HM. It is important to determine very small concentration in the range of the $\mu\text{g/l}$, ng/l or pg/l in the environmental samples in order to study toxicity effects on the living organisms or evaluate, in general, the environmental impact of the HM.

Voltammetric techniques are useful for measuring continuously very low concentration of HM in drinking water, wastewater, groundwater, surface water, soil and sediment extracts.

The voltammetric instruments are composed by a cell with three electrodes immersed in a solution containing the analyte and an excess of electrolyte (nonreactive) called supporting electrolyte. One of the three electrode is the working electrode of vitreous carbon (VC), or mercury electrodes such as dropping mercury electrode, static mercury electrode and hanging mercury drop electrode, whose potential varied linearly with time, the other electrode is the reference electrode (commonly a saturated calomel, or a silver/silver chloride electrode (Ag/AgCl/KCl(sat.))) and the third electrode is the counter electrode, which is often a coil of platinum wire or a pool of mercury that simply serves to conduct electricity from the signal source through the solution to the electrode [15].

The potential signals used in voltammetry include the linear scan voltammetry, normal pulse voltammetry, the differential pulse polarography, the square wave voltammetry (SWV) and the cyclic voltammetry (CV) for the analysis of HM in the different matrices. Square wave voltammetric (SWV) technique is the most sensitive for direct measurements of HM concentrations. The current is measured at the end of each half cycle. The wave form can be applied to a stationary electrode or static mercury drop electrode. In this case, the time interval is arranged to allow the drop to grow to a pre-determined size. Discrete current-potential points separated by the potential increment ΔE can be obtained as a response. The number of current-potential points is determined according to a certain potential range ΔE in the scan rate. The currents increase proportionally to these scan rate. The response can sometimes be distorted by the electronic noise. A smoothing procedure is necessary to correct the interpretation of the response curve. The use of a ΔE as small as possible should be considered in order to avoid this response distortion. In this technique, the net current is generally compared with theoretical predictions of a dimensionless current. The experimental and dimensionless currents are related by the Cottrell factor for the characteristic time:

$$i = n \cdot F \cdot A \cdot C \cdot \sqrt{\frac{D}{\pi t_p}} \cdot \Psi \quad (1)$$

where Ψ is the dimensionless current, $t_p = \Gamma/2$ pulse width and the other symbols have their usual meaning: i is the current density (I/A); I is the current intensity; A is the area; D is the diffusion coefficient; F is the Faraday constant; n is the number of electrons transferred; and C is the concentration.

3. SPECIFIC ISSUES RELATED WITH ACQUISITION AND DATA PROCESSING OF HM MEASUREMENT SIGNALS

In the following paragraphs some concerns that are associated with HM measurements are highlighted. The main concerns are related with the measurement of very low signal amplitudes, that, for example in voltammetric HM measurement methods, can correspond to the measurement of current with amplitudes lower than some tens of nA, as well as, the requirement of a minimal number of measurement points when online measurements are performed.

3.1 CALIBRATION CURVES, LIMITS OF DETECTION AND QUANTIFICATION

Calibration is an essential operation to improve measurement system accuracy. In which concerns HM measurements, a relationship between measurement results and standard HM concentration values must be established in order to improve accuracy and also to evaluate measurement linearity and sensitivity.

Standards are obtained using stock or buffer solutions with well-known and accurate HM concentrations. Calibration curve are normally obtained by the analysis of at least 6 standard solutions used for each compound, ranging within a specific measurement range that is directly related with the expected concentration ranges under measurement. The calibration curves should have regression coefficients higher than 0.95 in orders to achieve acceptable results for quantitative analysis of HM.

To determine the limits of detection (LOD) and the limits of quantification (LOQ) of the heavy metal in the selected methods, ten blank samples are analysed to determine the lowest signal/noise ratio of each HM. The limits of detection (LOD) can be calculated by $3 \times SD/m$, where SD is the standard deviation of the lowest signal/noise ratio of the HM and m is the slope of the calibration curve. In the same way, the limits of quantification (LOQ) are calculated by using the factor of 10 instead of 3 in the same formula ($10 \times SD/m$).

3.2 MEASUREMENT UNCERTAINTY

All measurements are subjected to uncertainties that result from different error sources. The errors sources can be classified as gross, systematic or random, depending on their origin. Gross errors occur when a measurement is invalidated by a major event, such as the failure of equipment and systematic errors occur when the same experiment is repeated several times and the individual measurements cluster around a mean value. The systematic errors can be cancelled, or at least deeply reduced, using proper calibration procedures. By its turn, random errors are associated with inaccuracies that can be associated with an operator, equipment or measuring method, and their quantization is based on statistical parameters of measurement data, namely, the standard deviation of a set of measurement records. Regarding HM measurements, systematic and random errors occur due to improper sampling techniques and handling of sample, mistakes by operators and inadequate knowledge of the experimental procedure, incorrect use, calibration of the equipment and faulty

instruments, erroneous preparation of solutions, and use of contaminated glassware material and reagents. Random errors can be caused by the experimental techniques that are used, including measuring equipments, by the variations in the measuring conditions, and by external interferences and noise, among others.

3.3 IMPROVING THE CELERITY OF HM MEASUREMENTS

One important demand in a large number of online measurement systems is related with the capability to provide accurate measuring results using a minimal number of measurement points and acquisition rate. In this context, to improve the celerity of voltammetric based HM measurements, a segmented square wave voltammetry (SSWV) voltage scanning method, already proposed by the authors in previous works [6,7], can be considered instead of the traditional SWV continuous scanning method. The SSWV proposes a variable voltage increment between adjacent sweep points (ΔV), being these increments much lower around specific voltages ranges (ΔV_M) for which higher measurement resolution or accuracy are required. These voltage ranges are usually associated with the redox potentials of the specific heavy metals that are under analysis or with the metals that exhibit higher concentration levels in a given place.

Taking as an example the measurement of the concentration of three different HM, Fig. 1 is a pictorial representation of the counter-working electrode voltage (V_{CW}) variation for a continuous voltage scan (SWV) and for a segmented voltage scan (SSWV) that includes, in this case, a set of three voltage intervals being each one associated with the redox potential of each metal under analysis.

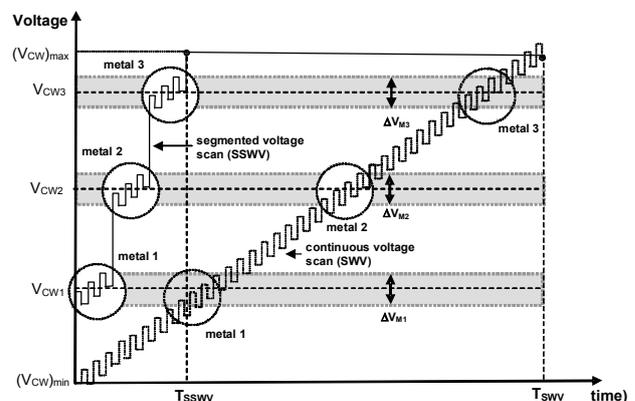


Fig. 1 – Voltammetric sweep voltage variation for a measurement solution that contains three different heavy metals (SWV- dotted line, SSWV- continuous line).

The ratio between SSWV and SWV voltage sweep durations, T_{SSWV} and T_{SWV} , respectively, is given by:

$$\frac{T_{SSWV}}{T_{SWV}} = \frac{\sum_{i=1}^N \Delta V_{CW_i}}{(V_{CW})_{\max} - (V_{CW})_{\min}} \cdot \frac{(S.R.)_{SSWV}}{(S.R.)_{SWV}} \quad (2)$$

where $(S.R.)_{SSWV}$ and $(S.R.)_{SWV}$ are the voltage sweep rate for SSWV and SWV, respectively, ΔV_{CW_i} is the voltage sweep range amplitude that is associated with the redox potential of each metal under analysis, $(V_{CW})_{\max}$ and $(V_{CW})_{\min}$ are the minimum and maximum values of the overall voltage sweep range, respectively. From (2), it is possible to conclude that for equal values of sweep rates, a substantial time reduction in the measurement can be obtained if the sum of the segmented voltage sweep range amplitudes is much lower than the overall voltage sweep range. This is an important advantage of the SSWV voltage scan that can be crucial to improve the performance of the voltammetric HM measurement systems and that can dictate the success, or failure, of online HM measurements. It is important to refer that an initial continuous voltage scanning can be required to identify the main metals contained in the water under analysis, if they are not a priori known.

3.4 LOW LEVEL CURRENT MEASUREMENTS

In voltammetric measurement techniques of HM concentrations it is important to underline that the measurement of very low HM concentrations implies very low current measurement capabilities. Typically, currents amplitudes that must be measured can be lower than some tens of nA [18] and a particular attention must be taken to the signal conditioning circuits characteristics, namely to the operational amplifiers specifications. Operational amplifiers based on field effect transistors (FET) with very high input impedances are typically used because it is essential to obtain reference electrode (RE) current amplitudes (I_{RE}) as low as possible relatively to the working electrode current amplitudes (I_{WE}). If this condition is not verified, accuracy measurements are strongly affected. Some of the current that comes from the counter electrode (CE) is diverted to the RE and consequently the working electrode current amplitude is lower than the current generated by ions dissolved in the solution. As an example, the FET input LM356 operational amplifier has typical values of bias and offset currents equal to 30 pA and 3 pA, respectively. These current amplitudes assure HM

concentration measurement errors much lower than 1 ng/l. Assuming an I_{CW} current amplitude equal to 100 nA, in this case, the relative measurement error caused by bias and offset currents is lower than 0.03 % of I_{CW} . Operational amplifiers voltage offsets errors are not so critical because they only affect the voltage scanning amplitude that is associated with each voltammetric current peak. These errors are very small and easily compensated by the measuring calibration system.

3.5 NOISE

Concerning noise effects, different analog and digital filtering techniques must be used to improve the signal to noise ratio and to reduce measurement uncertainty. The immunity of the measurement system to noise interferences is usually quantified by the normal mode rejection ratio (NMRR) and by the common mode rejection ratio (CMRR). The NMRR represents the capability of the measurement system to reject signals of a given frequency or within a given frequency range and is defined as:

$$NMRR = 20 \cdot \log_{10} \left(\frac{V_m}{V_{nmd}} \right) \quad (3)$$

where V_m represents the measurement signal amplitude and V_{nmd} represents the signal amplitude deviation caused by the disturbing signal that is present in the input channel of the measurement system.

This rejection ratio, usually expressed in dB, is particularly important when making low-level DC measurements because the measurement accuracy is substantially affected by the noise of the mains frequency of 50 and 60 Hz which is always superimposed with the signals under measurement. Fig. 2 represents the signal to noise ratio (SNR) of a measuring channel that is affected by a noise signal with a root mean square (R.M.S) amplitude equal to 100 mV when the amplitude of the signal under measurement varies from 10 to 50 mV and NNRR of the signal conditioning circuit varies between 40 and 70 dB with 10 dB increments.

By its turn, the common mode rejection rate (CMRR) represents the capability of a differential input amplifier, or measurement system, to reject a signal that is common to both inputs of an amplifier and is defined by:

$$CMRR = 20 \cdot \log_{10} \left(\frac{G_{diff}}{G_{comm}} \right) \quad (4)$$

where G_{diff} and G_{comm} represent the differential gain and common mode gain of the amplifier, respectively.

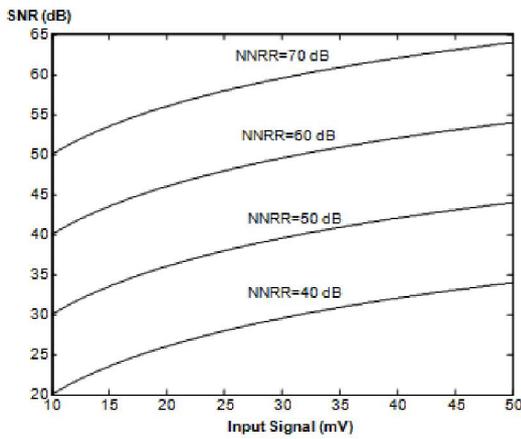


Fig. 2 – Signal to noise ratio of a measuring channel that is affected by a noise signal with a root mean square (R.M.S.) amplitude equal to 100 mV (NNRR-normal mode rejection ratio).

Fig. 3 represents the signal to noise ratio (SNR) of a measuring channel that is affected by a common voltage amplitude equal to 1 V when the amplitude of the signal under measurement varies between 10 to 50 mV and CMRR of the signal conditioning circuit varies between 60 and 120 dB with 20 dB increments.

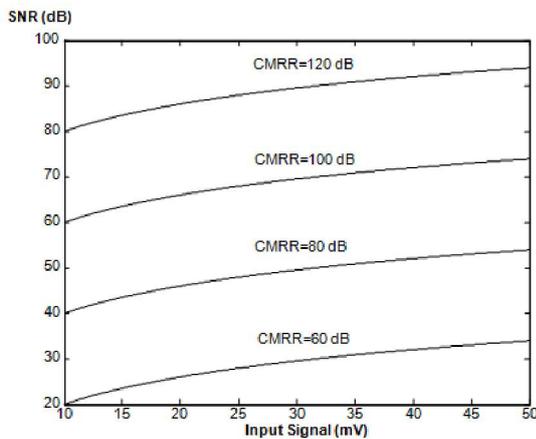


Fig. 3 – Signal to noise ratio of a measuring channel that is affected by common voltage amplitude equal to 1V.

3.6 SELF-DITHERING

Regarding the analog to digital (A/D) conversion of the measurement signals, it is essential to take advantage of signal processing techniques to improve the signal to noise ratio of the digitized signal. These techniques include dithering, oversampling, decimation, noise shaping and digital filtering [19, 20]. Basically using averaging techniques and oversampling the effective number of bits of an A/D converter can increase substantially at the expenses of the reduction of bandwidth of the acquisition system. For an

oversampling factor equal to N, it is possible to demonstrate that the effective number of bits gain is, for low self-dither amplitudes, relatively to the A/D quantization step, approximately given by:

$$ENOB_{gain} = \frac{20 \cdot \log_{10}(\sqrt{N})}{6.02} \quad (5)$$

In oversampling acquisition mode, an additional resolution bit can be obtained at the expenses of a factor four reduction in the input signal of the bandwidth. It is important to underline that this bandwidth reduction does not raises any problems, in terms of the HM measurement system performance, because the data acquisition rate of a generic A/D converters is much higher than the voltage sweep rate that is required for voltammetric purposes.

Concerning the implementation of dithering techniques, it is possible to take advantage of the noise that is contained in the A/D converter input signal without need of any external dither generator since the implementation of self-dithering, without any additional hardware requirement, is effective.

4. CURVE FITTING OF MEASUREMENT DATA

In the following paragraphs, a comparative study between B-Spline [21, 22], Gaussian, artificial neural network [23, 24] and polynomial curve fitting techniques, of HM measurement data is performed using the root mean square (RMS) deviation, between curve fitted and experimental data, as a relative performance factor. For the same set of voltammetric measurement data, the performance is analyzed using a variable number of measurement points that is a subset of the complete measurement data set. Sensitivity of each method to measurement data outliers is also analyzed.

4.1. EXPERIMENTAL DATA

A laboratory prepared solution, with well known concentrations of three HM metals, was used to perform a comparative analysis between the different curve fitting methods.

The experimental voltammogram data of the water sample with 80µg/l of Cu²⁺, 50µg/l of Pb²⁺ and 2µg/l of Cd²⁺, that was used for comparative analysis, is represented in Fig. 4. The voltage scan range, associated with the voltammogram, varies between -800 and 0 mV with 1 mV voltage increment between adjacent measurement points. The temperature (T) of the water sample, that affects the measurement results, was set to 20°C.

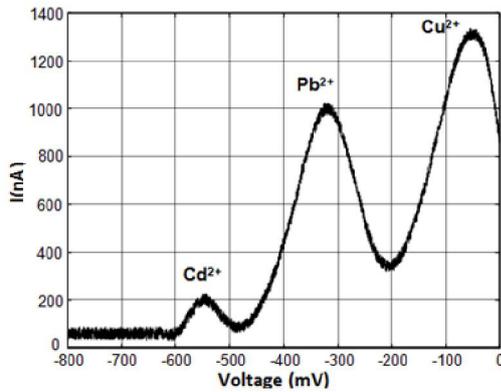


Fig. 4 – Experimental voltammogram for a water sample with 80µg/l of Cu²⁺, 50µg/l of Pb²⁺ and 2µg/l of Cd²⁺ (T=20°C).

To evaluate the performance of the different methods, in an objective way, a decimation factor (DF) was applied to the total number of measurement points. The decimation factor was defined by:

$$DF = \frac{N_S}{N_I} \quad (6)$$

where N_S represents the total number of measurements points and N_I represents the number of measurements that are used for curve fitting purposes.

The performance of each method was evaluated by using the R.M.S. error between measurement and curve fitting data. According to this metrics, for a specific DF value, the best method is the one that gives a lower root mean square deviation around the voltage ranges (ΔV_M) associated with the voltammogram current peaks.

4.2. DATA PROCESSING ALGORITHMS

Different curve fitting methods were tested for data processing of the measurement data. Fig. 5 represents the flow diagram that describes the main processing blocks associated with the evaluation of the performance of different algorithms used for curve fitting purposes. The measurement data array contains two columns with the measurement points values. Each measurement point includes a voltage value, included in the voltammetric voltage range (VR), and the associated current value. In order to test the capability of each algorithm to use a minimum number of measurement points, a data decimation block extract a subset of measurement values from the complete measurement data array. The coefficients DF and N_SEG are used to adjust the decimation rate and, if necessary, the number of voltages sub-ranges, that are extracted from the

complete measurement data set respectively. An optional outlier removal data processing block can be used to remove outliers from the decimated measurement data array. Chauvenet's criterion [25] was used to identify measurement outliers. For curve fitting purposes, four different algorithms were considered, namely, polynomial algorithm, Gaussian algorithm, Radial Basis Function based algorithm and B-Spline algorithm.

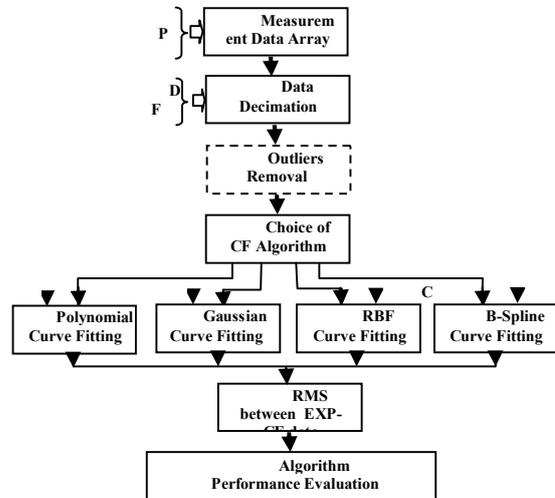


Fig. 5 – Flow diagram of data processing of measurement data (NP- number of measurement points; VR- voltage range; ΔV - voltage scan increment; DF- decimation factor; N_SEG- number of segments; CFP- curve fitting parameters; RMS- root mean square; EXP- experimental; CF- curve fitted).

Regarding curve fitting parameters: polynomial curve fitting parameters (CFP_P) includes the polynomial degree and the root mean square deviation between experimental and curve fitted data (RMSE); Gaussian curve fitting parameters (CFP_G) includes the number of Gaussian functions to be considered, the mean and standard values of each Gaussian function and the RMSE parameter; the Radial Basis Function Parameters (CFP_{RBF}) includes the artificial neural network architecture, the maximum number of training epochs, the maximum number of hidden layer neurons and the RMSE parameter; the B-Spline curve fitting parameters (CFP_S) includes the B-Spline order and the RMSE parameter. The algorithm performance evaluation is based on the RMS deviation between experimental and curve fitted data, defined by:

$$RMS = \sqrt{\frac{1}{NP} \cdot \sum_{i=1}^{NP} (EXP(i) - CF(i))^2} \quad (7)$$

where NP represents the number of measurement points, EXP(i) and CF(i) represent a generic

experiment measurement value and the associated curve fitted value, respectively.

Concerning software, the MATLAB program was used for data processing purposes. The main functions used for curve fitting purposes were selected from the optimization, artificial neural network and statistical toolboxes.

4.3. POLYNOMIAL CURVE FITTING

Regarding polynomial curve fitting, one of the more important issue, when using this method, is the choice of the polynomial degree that minimizes the least mean square (LMS) error between the curve fitted and the experimental data. To avoid numerical oscillations that occur if an excessive polynomial degree is used, it is important that the LMS error is evaluated relatively to a set of data, validation data set, during the selection of the best polynomial degree. It is important to remember that a polynomial function can pass through all the points that are used for curve fitting purposes, that mean the usage of a polynomial degree with $N-1$ order for a set of N measurement points, but the interpolation errors between those points are usually very large. By the other hand, underfitting occurs if a very low order polynomial is selected and the polynomial function is not sufficient to fit the measurement set, being the curve fitting errors very large. The errors that were previously mentioned are even larger if it is needed to perform extrapolation of measurement data. As an example, Fig. 6 represents the polynomial curve fitting results when the voltammetric measurement data is approximated by an 8 degree polynomial and the measurement data is decimated by a factor equal to 16. The vertical axis values are normalized against the maximum current value. The large error values that are obtained, probably caused by overfitting, assumes values higher than 85 % in terms of HM concentrations (Pb^{2+}) and, unfortunately, the minimum values of the errors are not verified around the redox potentials of the metals contained in the water samples. This means that the curve fitting results are very poor not only in terms of HM concentration evaluation but also in terms of its identification.

Regarding the polynomial results for different values of the decimation factors (DF), Fig. 7 represents the R.M.S. error as a function of the DF when 4th, 6th and 8th polynomial degrees are used. In this case for higher DF values the better results are obtained with a 4th degree (P_{degree}) polynomial. As it is clearly shown, as expected, for an 8th degree polynomial overfitting errors cause very large error for DF factors above 30.

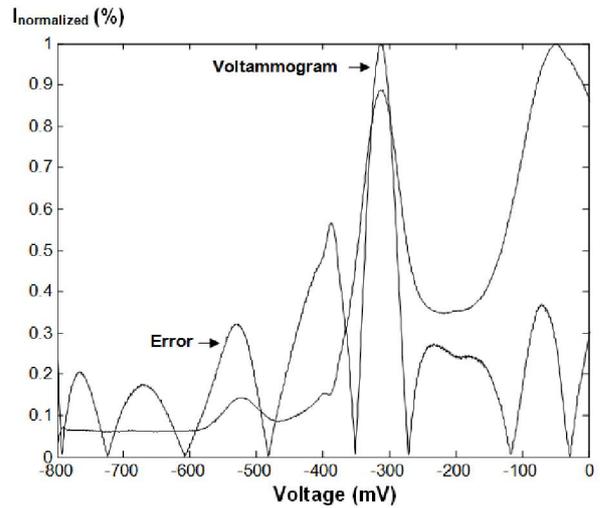


Fig. 6 – Polynomial curve fitting errors when the voltammetric measurement data is approximated by an 8 degree polynomial (DF=16).

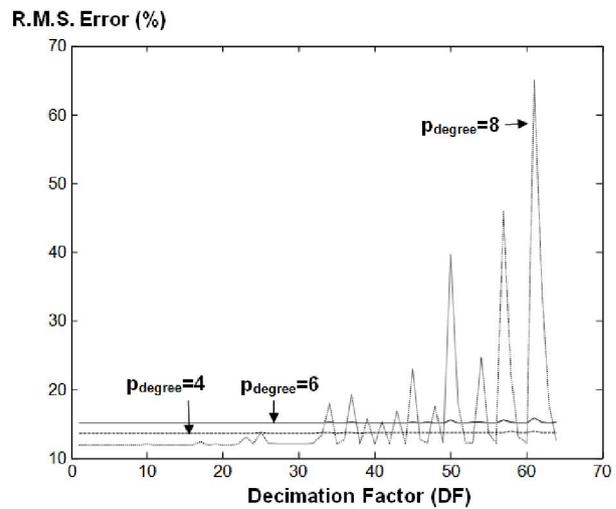


Fig. 7 – R.M.S. error as a function of the decimation factor (DF) for a curve fitting performed with a 4th, 6th and 8th degree polynomials.

4.4. ANN, B-SPLINE AND GAUSSIAN CURVE FITTING

In order to compare the relative performance of Gaussian, B-Spline, and Radial Basis Function ANN (RBF_{NN}) curve fitting methods, the same set of DF coefficients were used and the RMS deviation between experimental and curve fitted data was evaluated. The Gaussian curve fitting algorithm used a modified version of the Nelder-Mead algorithm [26] and the number of Gaussian functions (n) is equal to the number of the current peaks (different heavy metals) contained in the voltammogram data. Each primary Gaussian function (GF_i) is defined by the following relationship:

$$GF_i(V) = k_i \cdot e^{-\frac{(V-\mu_i)^2}{2\cdot\sigma_i^2}}$$

$$GF(V) = \sum_{i=1}^n GF_i(V) \tag{8}$$

where k_i represents the amplitude, μ_i the mean and σ_i the standard deviation of each primary Gaussian function (GF_i).

The Gaussian curve fitting parameters, associated with the means and amplitudes of each primary Gaussian function (GF_i), are used to identify the HM types and their concentrations, respectively. The standard deviations are also important to evaluate the overlapping degree between the Gaussian curves and to validate the interpolation results.

Radial Basis Function ANN curve fitting was performed using the following main parameters: maximum number of training epochs equal 200; RMS performance goal equal to 0.1%; spread constant equal to 1; maximum number of hidden layer neurons equal to 64 and a maximum number of validation failures equal 5. The Levenberg-Marquardt back-propagation algorithm [27] was used for training the ANN.

Fig. 8 represents the curve fitting results that was obtained for each curve fitting method when the DF coefficient varies between 1 and 64 and the number of measurement points (N_s) is, as previously, equal to 800.

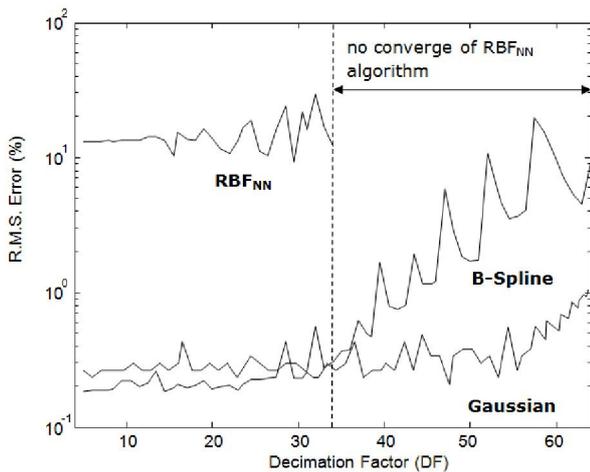


Fig. 8 – Comparative performance of Gaussian, B-Spline and RBF_{NN} curve fitting methods as a function of the decimation factor.

The results show clearly that the Gaussian curve fitting method outperforms the B-Spline curve fitting method for high values of the decimation factor. It can also be observed that, for a decimation factor higher than 35, the Gaussian curving method

outperforms de B-Spline method and the number of measurement points that are required to perform curve fitting of the voltammogram data can be reduced by a factor of 1/64 with a R.M.S. interpolation error increase less than 1%.

These results were confirmed with a large number of voltammetric curve profiles, with different HM types and concentrations, and the qualitative results that were obtained, remained valid.

4.5. SENSITIVITY TO MEASUREMENT DATA OUTLIERS

Several tests were performed to evaluate the sensitivity of each method to outliers of measurement data. A variable number of outliers, between 0 and 5 % of the total number of measurement points (N_s), were inserted in the measurement data records and the root mean square error, between experimental data with outliers and curve fitted data, was evaluated. In all the simulated cases, the Gaussian curve fitting method is less affected by outliers than the other methods. Fig. 9 represents the relative increment of the R.M.S. error for the Gaussian and the polynomial interpolation methods.

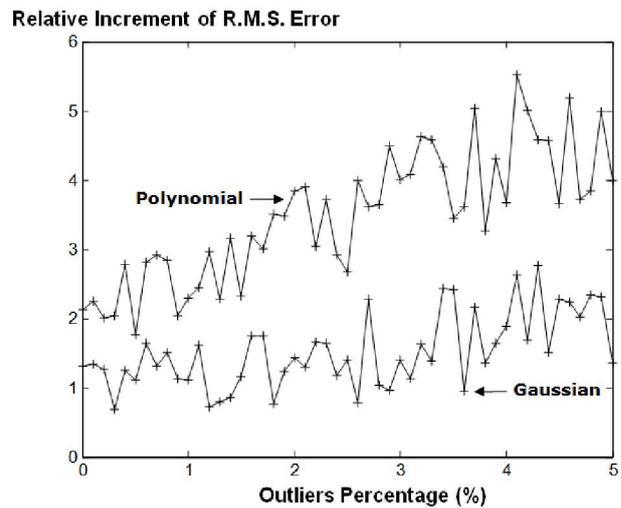


Fig. 9 – Relative increment of the R.M.S. error for the Gaussian and the polynomial interpolation methods as a function of the measurement outliers percentage.

As it is clearly shown, the Gaussian curve fitting method outperforms the polynomial curve fitting method, particularly when high values of DF are considered. Additional comparative results with the others curve fitting methods confirmed that Gaussian curve fitting minimizes the errors caused by outliers, as long as they are not concentrated around the redox potential of the HM under analysis.

5. CONCLUSIONS

This paper highlights specific issues related with the acquisition and data processing of heavy metal measurement signals. Regarding online measurements of heavy metals, two real challenges are related with the choice of the measurement method and the choice of the curve fitting method that can give the best accuracy using a minimum number of measurement points. Concerning curve fitting of the measurement data, a comparative analysis between polynomial, Gaussian, B-Spline, and Artificial Neural Network curve fitting methods was performed. The performance of each method is evaluated in terms of the required number of measurement points, for a given root mean square deviation between curve fitted and experimental data, and in terms of sensitivity to interpolations errors caused by measurement data outliers. It was clearly shown, that the Gaussian curve fitting method outperforms the others methods that were used for comparative analysis. For example, the usage of the polynomial curve fitting method gives always a root mean square interpolation error higher than 10 % and the Gaussian curve fitting error, using the same number of measurement points, for curve fitting purposes, is always more than 10 times lower.

6. REFERENCES

- [1] European Environment Agency, Water EU policies, available on line <http://www.eea.europa.eu/themes/water/intro>, accessed November 2015.
- [2] US Environmental Protection Agency (EPA), Regulatory Information, available online on <http://www2.epa.gov/regulatory-information-topic/>, accessed November 2015.
- [3] O. Malm et al., Mercury and Methylmercury in Fish and Human Hair from the Tapajós River Basin-Brazil, *Science of the Total Environment*, (175) 2, (1995), pp. 141-150.
- [4] S. Yurish, M. Gomes, *Smart Sensors and MEMS in Mathematics, Physics and Chemistry*, Kluwer Academic Publishers, 2003, 181 p.
- [5] M. Gomes, A. Duarte, J. Oliveira, Comparison of two methods for coating piezoelectrical crystals, *Anal. Chim. Acta*, (300) (1995), pp. 329-334.
- [6] O. Postolache, J. M. Dias Pereira, P. S. Girão, Water Quality Monitoring and Associated Distributed Measurement Systems: An Overview in Series: "Water Quality and Assessment", *Society for Instrumentation, Systems, and Automation Society*, In-Tech Publishing, Rijeka, Croatia, 2012, pp. 25-64.
- [7] M. Dias Pereira, O. Postolache, P. S. Girão, Heavy Metals Measurement: A Suitable Solution to Improve Online Measurement Celerity, *Instrumentation Science & Technology*, (40) 4, (2012), pp. 355-371.
- [8] Zou Xiaoyong, Mo Jinyuan, Spline wavelet overlapped peaks analysis, *Science Bulletin*, (44) 10 (1999), pp. 901-904.
- [9] M. A. Tarighat, Analytical Methods: Orthogonal Projection Approach and Continuous Wavelet Transform-feed Forward Neural Networks for Simultaneous Spectrophotometric Determination of some Heavy Metals in Diet Samples, *Food Chemistry*, (192) (2016), pp. 548-556.
- [10] A. Scozzari, N. Acito, G. Corsini, Signal Analysis of Voltammetric Data Series for Water Quality Tests and Classification, in Proceedings of the Instrumentation and Measurement Technology Conference, Pisa, Italy (May, 2005), Vol. 1, pp. 89-92.
- [11] K. Kano, T. Konse and T. Kubota, The Curve Fitting Analysis of D.C. and A.C. Voltammograms of a Two-step Surface-redox Reaction. The Application to the Surface-redox System of Adriamycin Adsorbed on a Pyrolytic Graphite Electrode, *Bull. Chem. Soc. of Japan*, (58) 7 (1985), pp. 1879-1885.
- [12] B. D. Fleming, N. L. Barlow, J. Zhang, A.M. Bond, F.A. Armstrong, Application of power spectra patterns in Fourier transform square wave voltammetry to evaluate electrode kinetics of surface-confined proteins, *Anal. Chem.*, (78) 9 (2006), pp. 2948-2956.
- [13] Zheng Xiaoping, Mo Jinyuan and Cai Peixiang, Spline wavelet in the resolution of overlapping voltammetric peaks, *Journal of Science in China (Series B): Chemistry*, Publisher Science China Press, co-published with Springer, (42) 2 (1999), pp. 145-152.
- [14] W. Huang, T. L. E. Henderson, A. M. Bond, and K. B. Oldham, Curve fitting to resolve overlapping voltammetric peaks: model and examples, *Analytica Chimica Acta*, (304) 1 (1995), pp. 1-15.
- [15] D. A. Skoog, F. J. Holler, T. A. Nieman, *Principles of Instrumental Analysis*, 5th edition, Saunders College publishing, Orlando, Florida, Chap. 25, 1998, pp. 639-670.
- [16] J. Zhuang, L. Zhang, W. Lu, D. Shen, R. Zhu, D. Pan, Determination of trace copper in water samples by anodic stripping voltammetry at gold microelectrode, *Int. J. Electrochem. Sci.*, (6) (2011), pp. 4690-4699.
- [17] J. M. Friedrich, C. Ponce-de-León, G. W. Reade, F. C. Walsh, Reticulated vitreous carbon as an electrode material, *J. Electroanal. Chem.*, (561) (2004), pp. 203-217.
- [18] R. L. McCreery, K. K. Kline, *Laboratory Techniques in Electroanalytical Chemistry*, 2nd Edition, P. T. Kissinger and W. R. Heineman eds., Dekker, New York, 1995, Chap. 4 and 10.

- [19] Analog Devices, Application Note AN-283, Sigma-Delta ADCs and DACs, available online on <http://www.analog.com/>, accessed July 2015.
- [20] J. M. Dias Pereira, P. S. Girão, O. Postolache, Adaptive Analog-to-Digital Conversion Using Self-Dithering in Data Acquisition Systems, in *Proceedings of the 11th IEEE International Conference on Electronics, Circuits and Systems, (ICECS'2004)*, Tel Aviv, Israel, (December 2004), Vol. 1, pp. 627-630.
- [21] M. Unser, A. Aldroubi, M. Eden, B-Spline Signal Processing: Part I-Theory, *IEEE Transactions on Signal Processing*, (41) 2 (1993), pp. 821-833.
- [22] S. A. Dyer, J. S. Dyer, Cubic-Spline Interpolation: Part 1, *IEEE Instrumentation Measurement Magazine*, (4) 1 (2001), pp. 44-46.
- [23] L. Yingwei, N. Sundararajan, P. Saratchandran, Performance evaluation of a sequential minimal radial basis function (RBF) neural network learning algorithm, *IEEE Transactions on Neural Networks*, (9) 2 (1998), pp. 308-318.
- [24] S. Haykin, *Neural Networks – A Comprehensive Foundations*, 2nd edition, Prentice Hall, NJ, USA, 1999, 842 p.
- [25] J. R. Taylor, *An Introduction to Error Analysis*, 2nd edition, University Science Books, Sausalito, California, USA, 1997, 327 p.
- [26] J. A. Nelder, R. Mead, A Simplex Method for Function Minimization, *Computer Journal*, (7) 4 (1965), pp. 308-313.
- [27] F. M. Silva, L. B. Almeida, *Speeding up Backpropagation*, in: *Advanced Neural Computers*,

R. Eckmiller edition, North Holland, 1990, pp. 151-160.



José Miguel Dias Pereira, received his degree in Electrical Engineering from the Instituto Superior Técnico (IST) of the Technical University of Lisbon (UTL). He is, at present, a Principal Coordinator Professor in the Polytechnic Institute of Setúbal. His main research areas, developed as a Researcher member of the Instituto de Tele-

comunicações, include smart sensing, data sensor fusion, A/D and D/A conversion techniques, instrumentation and transducers for industrial applications and advanced data processing techniques for biomedical and environmental applications.



Ricardo Manuel Nunes Salgado, PhD in Chemical and Biochemical Engineering. Assistant Professor in Setúbal Polytechnic Institute (Portugal) and member of REQUIMTE/LAQV research center. Researcher in the biodegradation and advanced oxidation technologies applied to water and waste-

water treatment; ecology and environment impact assessment of organic compounds and heavy metals; and reuse of waste materials for biodiesel production.